The Gene Search System: A Method for Efficient Detection and Rapid Molecular Identification of Genes in *Drosophila melanogaster*

Gakuta Toba,* Takashi Ohsako,† Naomasa Miyata,* Tsuyoshi Ohtsuka,* Ki-Hyeon Seong* and Toshiro Aigaki*,†

**Department of Biological Sciences, Tokyo Metropolitan University, Tokyo 192-0397 and †Precursory Research for Embryonic Science and Technology, Japan Science and Technology Corporation, Kyoto 619-0237, Japan*

Manuscript received September 4, 1998
Accepted for publication November 3, 1998

ABSTRACT

We have constructed a *P*-element-based gene search vector for efficient detection of genes in *Drosophila melanogaster*. The vector contains two copies of the upstream activating sequence (UAS) enhancer adjacent to a core promoter, one copy near the terminal inverted repeats at each end of the vector, and oriented to direct transcription outward. Genes were detected on the basis of phenotypic changes caused by GAL4-dependent forced expression of vector-flanking DNA, and the transcripts were identified with reverse transcriptase PCR (RT-PCR) using the vector-specific primer and followed by direct sequencing. The system had a greater sensitivity than those already in use for gain-of-function screening: 64% of the vector insertion lines (394/613) showed phenotypes with forced expression of vector-flanking DNA, such as lethality or defects in adult structure. Molecular analysis of 170 randomly selected insertions with forced expression phenotypes revealed that 21% matched the sequences of cloned genes, and 18% matched reported expressed sequence tags (ESTs). Of the insertions in cloned genes, 83% were upstream of the protein-coding region. We discovered two new genes that showed sequence similarity to human genes, *Ras-related protein 2* and *microsomal glutathione S-transferase*. The system can be useful as a tool for the functional mapping of the Drosophila genome.

GENOME sequencing and expressed sequence tag (EST) projects are rapidly progressing in various organisms. The next step in exploiting genomics requires an efficient method to detect and identify genes for functional mapping of the genome. Genetic approaches in *Drosophila melanogaster* have defined many genes that have been informative for understanding the function of their counterparts in vertebrates, including humans (Sidow and Thomas 1994; Banfi *et al.* 1996). It is desirable to develop a method for efficient detection of genes and for mapping them on the genome on the basis of their function. In Drosophila, *P*-element insertional mutagenesis appears to be the most suitable method for this purpose, because the sequence information of vector-flanking DNA can be obtained relatively easily (Cooley *et al.* 1988; Bellen *et al.* 1989; Bier *et al.* 1989). In fact, the gene disruption project with *P* elements is an integral component of the Berkeley Drosophila Genome Project (BDGP); the insertion lines contribute by serving as the materials for obtaining DNA markers and for studying gene function (Spradling *et al.* 1995).

Mutations caused by *P*-element insertion are principally loss of function. One possible limitation of a loss-of-function screen is the sensitivity of phenotype detection. Genes that are not essential for normal development would not be detectable on the basis of easily scorable phenotypes, such as viability or visible phenotypes (Miklos and Rubin 1996). Another potential problem with *P*-element mutagenesis is that a significant fraction of *P*-element-induced mutations are not associated with the insertions (Déak *et al.* 1997; Salzberg *et al.* 1997). In other words, molecular information obtained from *P*-element-flanking DNA is not always relevant to the phenotype because it might be caused by background mutations.

Gain-of-function mutagenesis is an alternative approach to identifying genes. Misexpression of genes using transgenic technology has been widely used to assess gene functions, especially since the GAL4-UAS system was introduced into Drosophila (Brand and Perrimon 1993). There are genes whose loss-of-function mutations do not show any obvious phenotype, but whose misexpression causes obvious phenotypes that are suggestive of the genes' normal functions (Chiba *et al.* 1995; Chen *et al.* 1997). Thus, gene detection based on a gain-of-function phenotype can identify genes that are not uncovered by loss-of-function phenotypes. Genetic screens will be more efficient with a *P* element that is capable of inducing forced expression of a gene, in addition to having the potential to disrupt a gene. The GAL4-UAS system that allows conditional gene expression appears to be suitable for a systematic generation of gain-of-

Corresponding author: Toshiro Aigaki, Department of Biological Sciences, Tokyo Metropolitan University, 1-1 Minami-osawa, Hachioji-shi, Tokyo 192-0397, Japan. E-mail: aigaki-toshiro@c.metro-u.ac.jp

function mutations. Insertion of a *P*-element vector containing upstream activating sequence (UAS) into the Drosophila genome allows GAL4-dependent forced expression of genes flanking the inserted *P* element. Although a gain-of-function phenotype alone is not sufficient to define the normal function of a gene, the method can be a powerful tool for functional mapping of the genome, along with providing rapid molecular identification of affected genes; genes would be detected based on an easily scorable phenotype, and identified molecularly through sequencing of misexpressed transcripts derived from an insert. Because background mutations occurring during *P*-element mobilization would not affect the GAL4-dependent phenotype, molecular information derived from misexpressed transcripts is closely associated with phenotypic information caused by forced expression of a gene. To obtain maximum efficiency, it is necessary to have a *P*-element vector that induces GAL4-dependent phenotypes at high frequency.

Two versions of *P* elements containing UAS for gain-of-function mutagenesis have been reported already (Rørth 1996; Crisp and Merriam 1997). Here we report a new version of the *P*-element vector containing UAS, which has a greater sensitivity in detecting genes than those previously reported. We also established a method for molecular identification of induced transcripts using reverse transcriptase PCR (RT-PCR) followed by direct sequencing. The system can be a useful tool for the functional mapping of the Drosophila genome.

MATERIALS AND METHODS

Construction of gene search vector: pCaSpeR3 *P*-element transformation vector (Thummel *et al.* 1988) was modified as follows: 5′*P* end of pCaSpeR3 was PCR mutagenized to create an *Eco*RI site using a primer corresponding to the 5′*P* end (5′GCCGAAGCTTACCGAAGTATACACTTAAAT) and an *Eco*RI site-flanked primer corresponding to the *P* transposase coding region (5′GCAGAATTCGACTAGTTTCATTTTTTTTTATTC CACGTAAGGG); the mutagenized product was digested with *Eco*RI and *Hin*dIII; the 140-bp fragment was used to replace a 4.7-kb *Eco*RI-*Hin*dIII fragment of pCaSpeR3, resulting in a 3.2-kb plasmid (pT1) containing the *P*-element ends and a multiple cloning site. The pCaSpeR3 was cleaved with *Eco*RI and *Eco*T22I, and the 4.0-kb fragment containing the *mini-white* gene and a multiple cloning site was subcloned into the *Eco*RI/*Pst*I site of pUC19 (TOYOBO, Osaka, Japan), generating a 6.6-kb plasmid (pT2). Five tandem repeats of UAS for GAL4 with a core promoter derived from the *Hsp70Bb* gene were excised from pUAST (Brand and Perrimon 1993) with *Bam*HI and *Eco*RI to obtain a 400-bp fragment, and with *Hin*dIII and *Xho*I to obtain a 440-bp fragment. The 400-bp fragment was subcloned into the *Bgl*II/*Eco*RI site located upstream of the *mini-white* gene in the pT2, producing a 7.0-kb plasmid (pT3). A 4.4-kb *Eco*RI-*Hin*dIII fragment (UAS/core promoter/*mini-white* gene) from the pT3 and the 440-bp *Hin*dIII-*Xho*I fragment (UAS/core promoter) from the pUAST were inserted together into the *Eco*RI/*Sal*I site of the pT1. The resulting construct contains a *P*-element-based gene search (GS) vector composed of 5254 bp, which has a *mini-white* gene as a marker in the middle and two copies of UAS adjacent to a core promoter, one copy near the terminal inverted repeats at each end of the vector, and oriented to direct transcription outward. The GS vector was introduced into flies using *P*-element-mediated transformation (Rubin and Spradling 1982) with the *Df(1)w* stock as a recipient. For genetic nomenclature, refer to FlyBase (http://flybase.bio.indiana.edu).

Generation and screening of GS vector insertion lines: The GS vector inserted on the second chromosome of the *Df(1)w* stock was mobilized onto a *CyO* chromosome using *Delta2-3* transposase (Robertson *et al.* 1988). Then the vector was further mobilized onto various chromosomes using *Delta2-3* transposase to generate a collection of new insertion lines. The 613 GS lines obtained include 147 on the X, 226 on the second, 237 on the third, and 3 on the fourth chromosomes. Insertions were kept in homozygous state or balanced with *Binsinscy*, *SM1*, or *TM3* for the X, second, and third chromosome, respectively. To induce forced expression of vector-flanking DNA, GS lines were crossed to GAL4-expressing lines and the F_1 individuals carrying both a GAL4 transgene and a GS insert were screened for lethality and visible phenotypes. The following GAL4 lines were used: *29BD-GAL4* (*P{GawB}29BD*; A. Brand and N. Perrimon, personal communication) obtained from A. Brand; *c355-GAL4* (*P{GawB}c355*; Harrison *et al.* 1995); *dpp-GAL4* (*P{GAL4-dpp.blk1}40C.6*; Staehling-Hampton *et al.* 1994); and *sev-GAL4* (*P{GAL4-Hsp70.sev}2*; Brunner *et al.* 1994) obtained from the Bloomington Drosophila Stock Center (Indiana). The GAL4 expression pattern was examined using *UAS-GFP* (*P{UAS-GFP.S65T}T2*; constructed by B. Dickson and obtained through the Bloomington Drosophila Stock Center) as a reporter. Flies were reared at 25° using standard fly culture medium.

Identification of induced transcripts: To analyze the GAL4-induced transcripts, GS vector insertion lines were crossed to *hs-GAL4* stock (*P{GAL4-Hsp70.PB}89-2-1*; constructed by A. Brand and obtained through the Bloomington Drosophila Stock Center). The F_1 third instar larvae were transferred into a 1.5-ml microfuge tube (20–30 individuals/tube), and heat-shocked at 37° for 1 hr. Poly(A)$^+$ RNA was isolated from the larvae using the QuickPrep Micro mRNA purification kit (Amersham Pharmacia Biotech, Arlington, IL). mRNA was reverse-transcribed using the first-strand cDNA synthesis kit (Amersham Pharmacia Biotech) with a *Not*I site-flanked oligo(dT) primer (5′AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTT TTTTTTTTTT, Amersham Pharmacia Biotech). A total of 1 μl of the reaction was used to amplify both 5′*P* and 3′*P* transcripts by PCR using ELONGASE enzyme mix (GIBCO BRL, Gaithersburg, MD) in a total volume of 50 μl with the upstream common primer (5′CTGAATAGGGAATTGGGAA TTCG) and the *Not*I site-flanked oligo(dT) primer. To amplify the transcripts of the 5′*P* or 3′*P* element ends separately, 1 μl of the first PCR reaction was reamplified in a total volume of 50 μl using either the 5′*P*-specific primer (5′GTGTATACTT CGGTAAGCTTCG) or the 3′*P*-specific primer (5′ATTGCAAG CATACGTTAAGTGGA) as an upstream primer together with a downstream primer (5′AGAACTGGAAGAATTCGCGG). PCR was carried out using a Perkin-Elmer (Norwalk, CT) gene amp PCR system 2400 or 9700 with the following thermal cycling program: 94° (60 sec), 16 cycles of 94° (15 sec) –65° (10 min), 12 cycles of 94° (15 sec) –65° (10 min with 15-sec increment for every cycle), and 72° for 10 min, then held at 4°. The resulting PCR products were electrophoresed on a 1.0% agarose (Type II; Sigma, St. Louis) gel, and the amplified bands were excised with a razor blade and subsequently purified using the QIAEX II gel extraction kit (QIAGEN, Chatsworth, CA). The purified DNA fragments were used as a template for sequencing reactions with the dRhodamine terminator cycle sequencing FS ready reaction kit (Perkin-

Elmer) using the 5′3′*P* common primer (5′CGACGGGACCACC TTATGTTA). Sequencing was carried out using a Perkin-Elmer ABI PRISM genetic analyzer 310. Sequence similarity searches were performed using the BLASTN or BLASTX program (Altschul *et al.* 1997) with the NCBI nonredundant nucleic acid database and dbest, or NCBI nonredundant protein database.

Subcloning: For subcloning of cDNA derived from the misexpressed transcripts of *Rap2l* and *Mgstl*, RT-PCR products obtained were blunt-ended using T4 DNA polymerase (TOYOBO), digested with *Not*I, and ligated into the *Not*I/*Eco*RV site of pBluescript SK+ (Stratagene, La Jolla, CA). At least three clones were sequenced to determine the structure of the cDNAs.

RACE: The 5′ end structure of wild-type transcripts of *Rap2l* was determined using the 5′ rapid amplification of cDNA ends (RACE) system, version 2.0 (GIBCO BRL), according to the manufacturer's protocol. Poly(A)$^+$ RNA was isolated from wild-type (Canton-S) larvae as described above. First-strand cDNA for *Rap2l* was synthesized using a gene-specific primer (R-1: 5′CTATAAAAGCGTACAACAA). A poly (C) tail was added to the 3′ ends of the cDNA using terminal deoxynucleotidyl transferase (GIBCO BRL) and dCTP (GIBCO BRL). Tailed cDNA was amplified by PCR using a nested, gene-specific primer (R-2: 5′CGAACGATGGTGGCGAATACTT) and a poly (G)-containing anchor primer (GIBCO BRL). The cDNA was reamplified using a nested, gene-specific primer (R-3: 5′GG GTGCTGGCTGACTTCCTTT) and the anchor primer. R-3 was used as a primer for direct sequencing. Similarly, the 5′ end structure of the *Mgstl* transcript was determined using gene-specific primers M-1 (5′AAGGTCTAGACCTATGTGCTC) for reverse transcription, M-2 (5′CGTTCGGATCGTCGAACTT) for the first PCR, and M-3 (5′CCTCTAGAAGACGGGATTG GAG) for the second PCR and for direct sequencing.

The 3′ end structure of *Rap2l* transcript was determined by 3′ RACE. The first strand cDNA was synthesized from poly(A)$^+$ RNA from wild-type larvae using the *Not*I site-flanked oligo(dT) primer. The cDNA was amplified by PCR using a gene-specific primer (R-4: 5′TCGTCTCGGGATGCTTTATTGA) and the downstream primer used for the amplification of misexpressed transcripts described above. The cDNA was reamplified using a nested, gene-specific primer (R-5: 5′GCACA GAGCAATTCGCATCCAT). R-5 was used as a primer for direct sequencing. Similarly, 3′ RACE for the *Mgstl* transcript was carried out using M-4 (5′GCGAATTCAAACACATACAATGG CC) as a gene-specific primer, which was also used for direct sequencing.

Analysis of *Mgstl* gene structure: The genomic region containing *Mgstl* was amplified by PCR using primers M-1 and M-4. The amplified fragments were directly sequenced as described above using the PCR primers as a sequencing primer. The 5′- and 3′-flanking regions were obtained by inverse PCR; genomic DNA isolated from wild-type (Canton-S) flies was digested with *Hin*dIII, self-ligated, and PCR amplified using primers M-3 and M-5 (5′CTCGAATTCTTCGTGGCCTTGG). The amplified products were blunt-ended by T4 DNA polymerase (TOYOBO) and digested with *Hin*dIII. The resulting two fragments were subcloned into the *Hin*cII/*Hin*dIII site of pBluescript SK+, and at least three clones were sequenced. All restriction enzymes used in this study were purchased from TOYOBO.

RESULTS

Scheme of the GS system: The GS system consists of three steps: (1) the generation of fly lines with single inserts that allow conditional forced transcription of genomic sequence, (2) induction of forced expression and screening for lines with a detectable phenotype, and (3) the molecular identification of transcribed sequences (Figure 1). We constructed a *P*-element-based GS vector utilizing the GAL4-UAS ectopic expression system (Brand and Perrimon 1993). The GS vector is similar to the EP element constructed by Rørth (1996), but it contains two copies of the UAS adjacent to the core promoter from the *Hsp70Bb* gene, one copy near the terminal inverted repeats at each end of the GS vector and oriented to direct transcription outward. Thus, upon GAL4 activation the vector integrated into the genome will induce transcription toward the flanking DNA on both sides of its integration site. The induced transcripts may cause phenotypic changes if the products affect biological pathways operating in the organism. An insertion upstream of the protein-coding region will cause over- or ectopic expression of the gene, while an insertion downstream of the protein-coding sequence produces antisense RNA, which may interfere with translation of the wild-type mRNA (McGarry and Lindquist 1986; Nicole and Tanguay 1987). Both sense and antisense strands will be transcribed if the GS vector is inserted in the middle of a gene. Observation of a phenotypic change for a given GS insert would indicate that the insertion site is near or within a gene that is capable of altering phenotype; the transcripts are readily and rapidly identifiable using RT-PCR followed by direct sequencing.

Generation and screening of GS vector insertion lines: We generated a total of 613 GS vector insertion lines (GS lines for short) and screened for dominant synthetic phenotypes, such as lethality, semilethality (<50% viability), or visible phenotype in the adult structure using four GAL4-expressing lines as drivers. The frequency of producing any phenotype with forced expression of flanking DNA depends on the extent and level of GAL4 expression. Two *P{GawB}* enhancer-trap lines, *29BD-GAL4* and *c355-GAL4*, express GAL4 in all imaginal discs at high level (Figure 2). These were selected because they are likely to produce phenotypes at high frequency. The two other GAL4 drivers have more specific expression patterns: *dpp-GAL4* is expressed along the anterior/posterior compartment boundary of each imaginal disc, and *sev-GAL4* is expressed mainly in the eye imaginal discs (Figure 2). *dpp-GAL4* allows us to assess the variability of effects caused by forced expression of flanking DNA in different body parts, while with *sev-GAL4*, the effects of forced expression are assessed in the eye, in which subtle perturbations of gene regulatory networks are detectable (Xu and Rubin 1993).

Figure 3 shows the frequency of phenotypes obtained with four distinct GAL4 lines. As expected from the expression pattern, a high frequency of phenotypes was obtained with the ubiquitously expressing *GAL4* transgenes, 57% for *29BD-GAL4* and 48% for *c355-GAL4*. Frequency of lethality was correlated with the total frequency of phenotype (33, 29, 14, and 7% for *29BD-GAL4*, *c355-GAL4*,

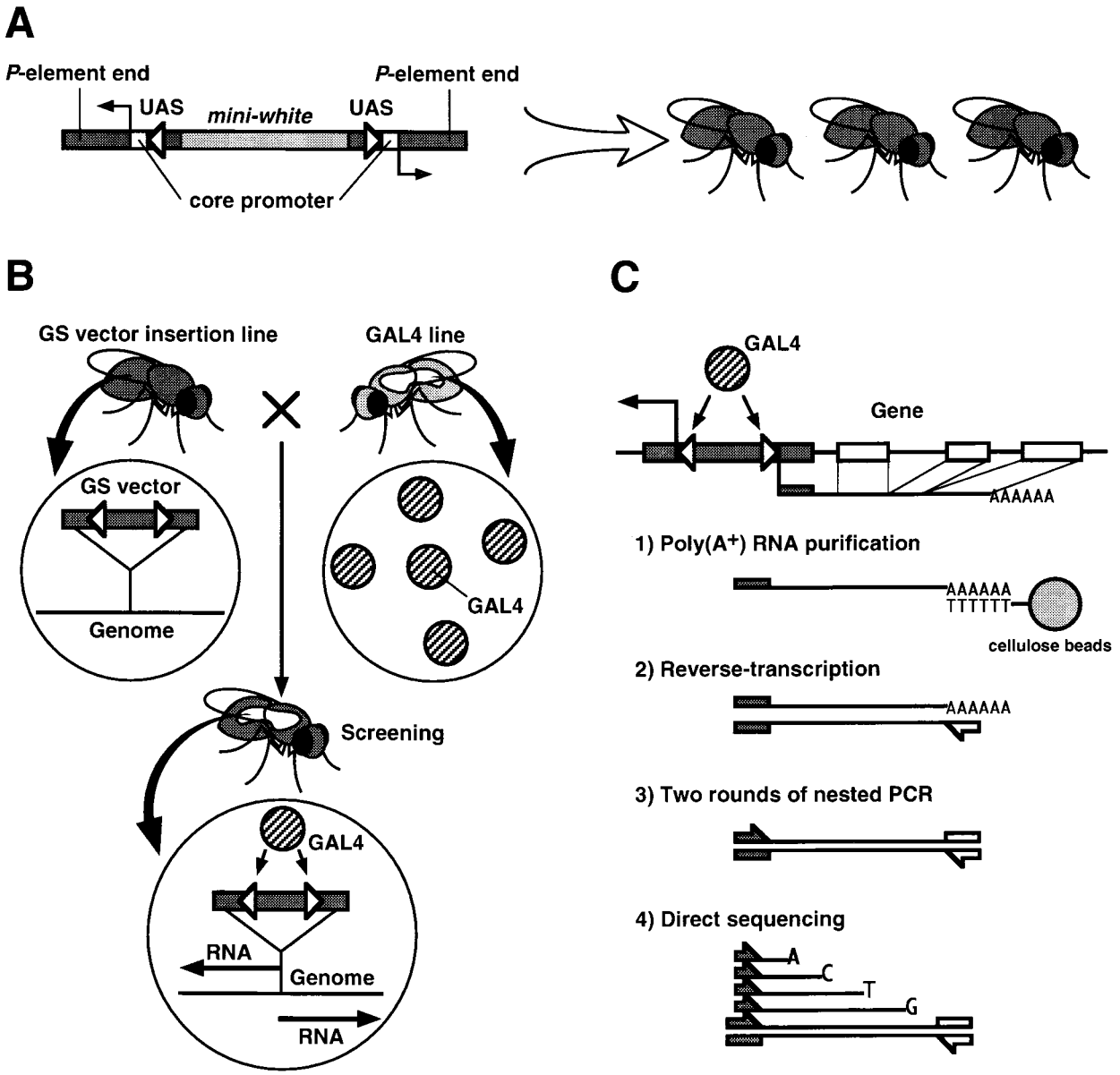


Figure 1.—Schematic representation of the GS system. (A) Structure of the GS vector. The GS vector contains UAS and a core promoter derived from the *Hsp70Bb* gene near the inverted terminal repeats at both *P*-element ends. The *mini-white* gene is included as a marker. A collection of transgenic flies, each with a single insertion of the GS vector, were generated by mobilizing the vector in the genome using Δ*2-3* transposase (see materials and methods). (B) Screening of GS vector insertion lines (GS lines) for phenotypes. Flies from GS lines were crossed to flies bearing GAL4 drivers to induce forced expression of the vector-flanking sequences in the F_1. The F_1 were screened for lethality and visible phenotypes. (C) Molecular analysis of induced transcripts. Upon GAL4 activation, transcription occurs toward the flanking genomic sequences through the *P*-element ends. (1) GS lines were crossed to *hs-GAL4* and poly(A)$^+$ RNA was prepared from heat-shocked F_1 larvae and (2) reverse-transcribed using an oligo(dT) primer. (3) cDNAs corresponding to the induced transcripts were amplified by two rounds of nested PCR using the vector-specific primers. (4) Finally, 5′ end sequences of the cDNAs were determined by direct sequencing.

sev-GAL4, and *dpp-GAL4*, respectively), while visible phenotypes were obtained approximately at the same frequency (20%), except for that with *c355-GAL4* (13%). Figure 4 represents the number of lines that showed visible phenotypes in each body part, which roughly corresponds to where GAL4 is expressed. *29BD-GAL4*, *c355-GAL4*, and *dpp-GAL4* induced visible phenotypes in various body parts, because these express GAL4 in all imaginal discs. Phenotypes caused by *sev-GAL4* are seen principally in the eye, as expected on the basis of expression of this driver. Overall, 394 lines (64%) showed a detectable phenotype in combination with at least one GAL4 line. *29BD-GAL4* appeared to be the most efficient driver to detect genes on the basis of a misexpression phenotype; it induced phenotypic changes in 88% of the GS lines that showed phenotypes with any of the GAL4 drivers.

Molecular analysis of forced-expression transcripts:

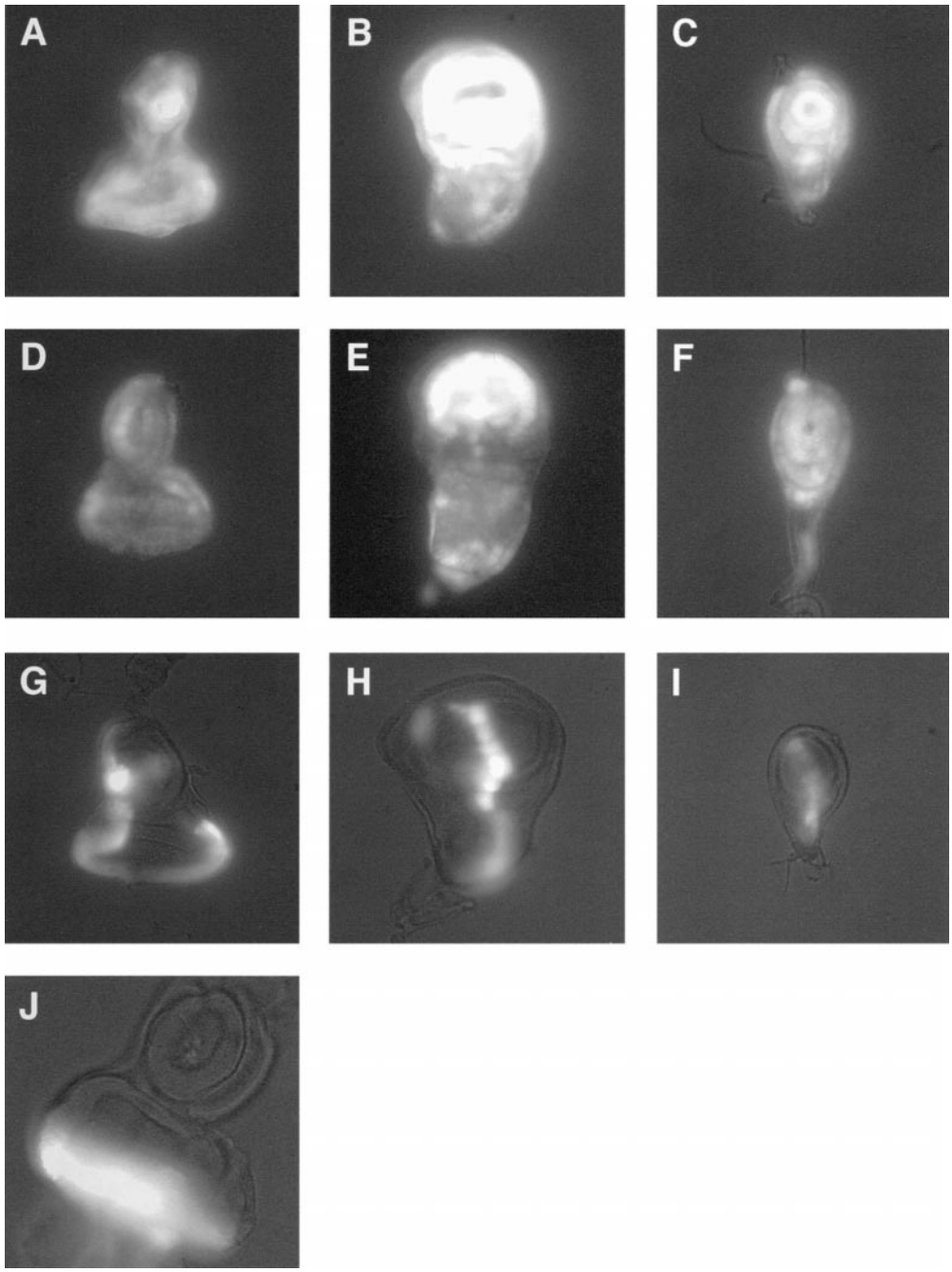


Figure 2.—Expression pattern in the imaginal discs of GAL4 lines used as drivers. (A–C) *29BD-GAL4*. (D–F) *c355-GAL4*. (G–I) *dpp-GAL4*. (J) *sev-GAL4*. (A, D, G, J) Eye-antennal disc. (B, E, H) Wing disc. (C, F, I) Leg disc. *29BD-GAL4*, a *P{GawB}* enhancer trap-line, expresses GAL4 in all imaginal discs at high level more or less ubiquitously. *c355-GAL4* also expresses GAL4 in all imaginal discs, with a high level of expression in the wing pouch. *dpp-GAL4* is expressed along the anterior side of anterior/posterior boundary of each imaginal disc. *sev-GAL4* is expressed mainly in the eye-antennal discs. Driver expression pattern was examined using *UAS-GFP* as a reporter.

To identify genes whose expression was forced in GS lines, we performed molecular analysis of the induced transcripts for 170 insertions of randomly selected GS lines among those that showed a phenotype upon forced transcription of flanking DNA. GS lines were crossed to the *hs-GAL4* line, poly(A)$^+$ RNA was isolated from heat-shocked F_1 larvae, and the transcripts derived from the vector insertion site were amplified with RT-PCR using vector-specific and oligo(dT) primers. The amplified cDNA fragments were subjected to single-pass sequencing using a vector-specific primer corresponding to the *P*-element end (see materials and methods). In most of the cases (146 of 170 inserts), we obtained two distinct transcripts derived from a single insert, indicating that the GS vector was indeed capable of inducing transcription bidirectionally. Database searches of the obtained sequences revealed that 47% of insertions were in known sequences (Table 1). A total of 21% showed similarity to sequences of cloned genes, 18% matched reported EST sequences (BDGP / Howard Hughes Medical Institute Drosophila EST Project; D. Harvey, L. Hong, M. Evans-Holm, J. Pendleton, C. Su, P. Brokstein, S. Lewis and G. M. Rubin, unpublished results), and 4% matched reported sequence tagged site (STS) sequences (BDGP; G. M. Rubin, unpublished results; European Drosophila mapping Consortium; M. Ashburner, unpub-

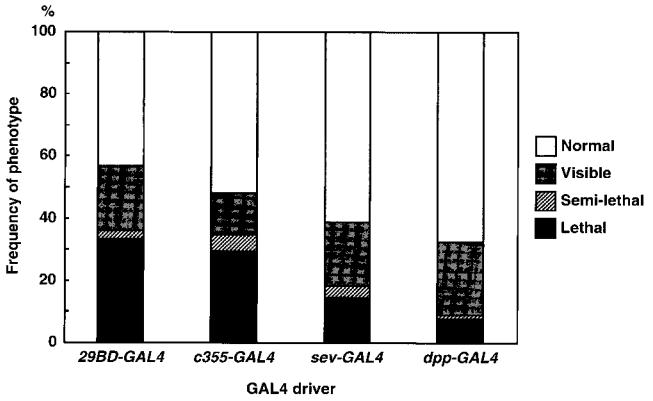


Figure 3.—Frequency of phenotypes for 613 GS lines crossed to four different GAL4 drivers. Flies carrying a GS insert were crossed to flies carrying each of the four GAL4 drivers, and the F_1 progenies carrying both transgenes were screened for lethality, semilethality, and visible adult phenotypes. Semilethal: viability was <50%. Visible phenotypes were scored if the penetrance was >50%. Lines showing both semilethality and visible phenotypes were included in the visible category. Flies were reared at 25°.

TABLE 1

Summary of molecular analysis

	No. of insertions	%
Known sequences[a]		
Cloned genes	35	21
ESTs	31	18
STSs	6	4
Transposons or repetitive sequences	7	4
Novel sequences		
Homolog	2	1
Novel	89	52
Total	170	100

[a] Similarity scores for cloned genes and ESTs are shown in Table 2.

lished results). We investigated the insertion sites relative to the transcription start site for those that matched cloned genes or ESTs (Table 2). The 5′-most ends of mRNA reported so far were defined as +1. More than 50% of insertions were found between −150 and +100, most frequently in between −100 and −1 (Figure 5). With respect to the insertions in cloned genes, 83% were upstream of the protein-coding region (data not shown), suggesting that most of the phenotypes detected in this screen were caused by over- or ectopic expression of full-length products.

Identification of a gene similar to human *Rap2*: Molecular analysis of misexpressed transcripts revealed that two of the GS insertions were in new genes that showed sequence similarity to human genes (Table 1). We subcloned the RT-PCR products into a plasmid and sequenced the entire cDNA. The sequence of the misexpressed transcripts derived from the insert in line GS2069 showed a similarity to human *Ras-related protein 2* (*Rap2*; Pizon *et al.* 1988); thus we named it *Ras-related protein 2-like* (*Rap2l*). *Rap2l* was detected as an insertion that showed a lethal phenotype when combined with

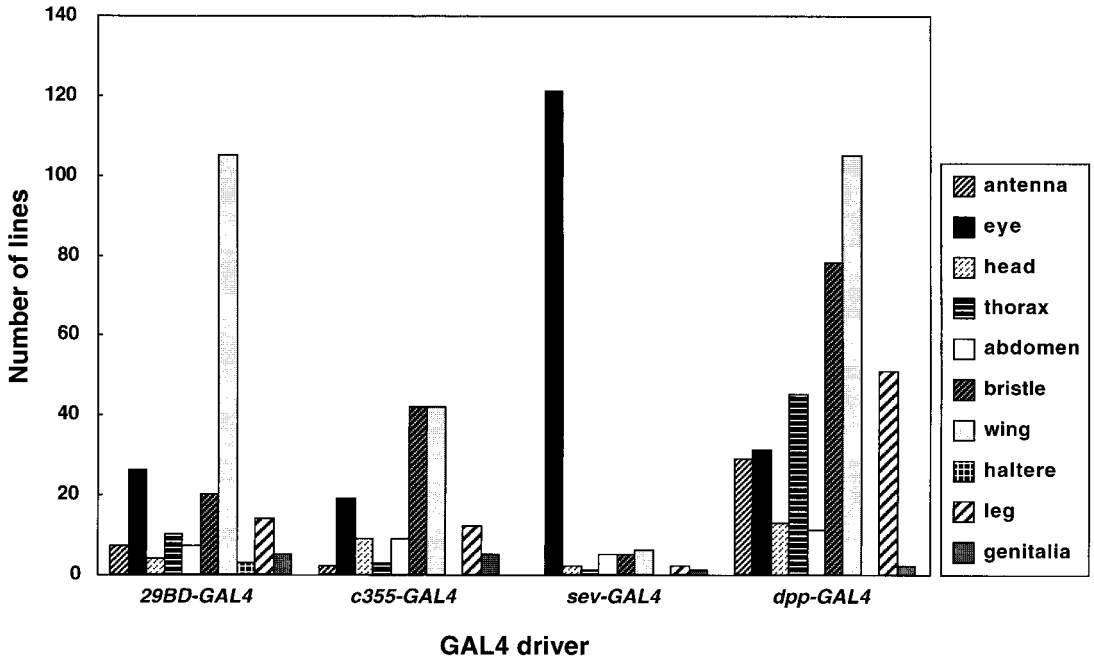


Figure 4.—Summary of visible phenotypes by body part. The number of lines showing any visible phenotypes in each body part is indicated for each GAL4 driver. Note that the phenotype frequency is correlated roughly with the extent and level of GAL4 expression in each body part; phenotypes appeared in various body parts in *29BD-GAL4*, *c355-GAL4*, and *dpp-GAL4*, while *sev-GAL4* induced phenotypes mainly in the eyes.

TABLE 2

Insertions in genes with sequence similarity to cloned genes and ESTs

Inserted locus					Phenotype[b]			
GS line	Gene or EST	Accession	Score	Insertion site[a]	*29BD-GAL4*	*c355-GAL4*	*sev-GAL4*	*dpp-GAL4*
97	*TfIIA-S*	X83271	<1e −100	−677	SL, rough eye	Rough eye	ND	ND
1029	*NetB*	U60317	1e −24	−95	Notched wings	SL	—	—
1032	*ph-p*	M64750	<1e −100	−353	L	L	SL, rough eye	SL
1038	*ras*	L14847	<1e −100	−666	L	L	L	—
1053	*amn*	U22825	<1e −100	+2597	L	SL, int. veins	Rough eye	Rough eye
1069	*Fas2*	M77165	4e −69	−63	SL, spread wings	M	Rough eye	ACV shortened
1073	*wapl*	U40214	2e −84	−117	SL, tergite ab.	Extra veins	SL, rough eye	Notched wings
1091	*elav*	M21153	<1e −100	−450	L	L	L	L
1115	*bss*	X89811	<1e −100	+364	L	L	L	L
1131	*Act5C*	X15730	<1e −100	−923	SL, aristae missing	SL, hair # reduced	SL, rough eye	SL, ACV missing
1141	*arm*	X54468	<1e −100	Intron	L	L	L	SL, fused tarsal seg.
1144	*ovo*	X59772	<1e −100	−58	L	L	L	L
2011	*HmgD*	M77023	<1e −100	−66	L	L	L	L
2042	*HmgD*	M77023	<1e −100	−10	L	L	L	L
2115	*exu*	S72757	<1e −100	3′ FR	L	L	—	Fused veins
2120	*fy*	AF022891	<1e −100	−196	—	—	—	—
2137	*HmgD*	M77023	2e −11	Intron	L	L	L	L
2141	*sop*	U01335	9e −21	−35	SL, blistered wings	SL	L	ACV missing
2160	*Gst2*	M95198	<1e −100	+10	Bent wings	L	Rough eye	—
2163B	*guf*	AF038597	<1e −100	Intron	Spread wings	—	—	—
2220	*exu*	S72757	<1e −100	3′ FR	L	SL, fused veins	—	Fused tarsal seg.
2227	*aop*	M97694	<1e −100	−97	L	L	L	Def. tarsi
2228	*ex*	L14768	<1e −100	+617	ND	ND	ND	Fused veins
3026	*ttk*	Z11723	2e −35	−3	L	L	L	L
3029	*Dhod*	L00964	3e −37	−277	L	L	Rough eye	Aristae missing
3052	*His2AvD*	X07485	<1e −100	Intron	L	L	Rough eye	ACV shortened
3069	*ttk*	Z11723	1e −43	Intron	L	L	L	L
3089	*neur*	S62597	<1e −100	+24	SL, int. veins	Int. veins	Rough eye	Int. veins
3097	*mod(mdg4)*	U30905	<1e −100	Intron	L	L	SL, rough eye	Unextended wings
3120	*stg*	X57495	3e −92	−135	SL, def. wings	SL, extra bristles	—	—
3127	*His2AvD*	X07484	<1e −100	+2	L	L	SL, rough eye	Fused tarsal seg.
3129	*stwl*	U41367	<1e −100	−111	L	L	L	L
3165	*ttk*	X71626	1e −15	−416	L	L	L	L
3205	*neur*	X61617	<1e −100	−69	Int. veins	Int. veins	—	Int. veins
3230	*gro*	M20571	6e −76	Intron	L	L	L	L
1027	LD12308	AA438512	2e −63	−93	L	L	Rough eye	—

(*continued*)

TABLE 2

(Continued)

Inserted locus					Phenotype[b]			
GS line	gene or EST	Accession	Score	Insertion site[a]	*29BD-GAL4*	*c355-GAL4*	*sev-GAL4*	*dpp-GAL4*
1084	LD12308	AA438512	1e −63	−93	L	Extra bristles	SL, rough eye	Notched wings
1135	LD22118	AA817082	<1e −100	−74	L	L	L	SL, blistered wings
2002	LD29847	AA949818	<1e −100	−37	L	L	L	L
2003	LD27171	AA941860	<1e −100	+290	SL	SL, extra bristles	Rough eye	Extra bristles
2007	LD06340	AA263242	8e −25	−250	SL	—	—	—
2025	LD12957	AA438639	3e −88	+37	Rough eye	—	Rough eye	—
2038	LD01639	AA735228	<1e −100	−29	SL, rough eye	Bent bristles	Rough eye	—
2048A	LD03829	AA201147	2e −49	+101	L	L	L	SL, def, tarsi
2048B	LD07122	AA263935	3e −30	−19	L	L	L	SL, def. tarsi
2053	LD04728	AA201504	<1e −100	−3	Hair orientation ab.	Hair orientation ab.	—	—
2055	LD03274	AA390332	<1e −100	−51	SL	SL	SL, rough eye	Aristae transform to tarsi
2067	GM02209	AA567240	<1e −100	+56	L	L	L	L
2074	LD03274	AA390332	<1e −100	−51	L	L	—	Aristae transform to tarsi
2091	LD03274	AA390332	<1e −100	−51	L	L	Rough eye	Aristae transform to tarsi
2121	LD29214	AA952141	3e −49	+35	L	L	L	L
2138	LD04967	AA201761	6e −53	+362	—	—	SL	—
2163A	LD04971	AA201765	<1e −100	−6	Spread wings	—	—	—
2191	LD06340	AA263242	6e −82	−275	Rough eye	—	—	—
2202	LD14959	AA440376	<1e −100	Intron	SL	—	—	—
2207	GM09451	AA697215	1e −64	−275	L	Notched wings	Rough eye	—
2208	LD29743	AA541057	4e −62	−11	L	Def. antennae	—	Thick veins
2209	LD33989	AA979429	<1e −100	+6	SL	—	—	—
3005	LD09360	AA390491	<1e −100	−1	L	SL, extra veins	SL, rough eye	—
3028	GM10514	AA803288	<1e −100	+64	—	—	Rough eye	—
3082	LD14744	AA440145	<1e −100	Intron	SL, extra veins	—	SL, rough eye	—
3086	HL04053	AA698259	4e −92	Intron	SL, def. tarsi	Hair # increased	Rough eye	—
3087	LD32772	AA951892	2e −29	−34	L	L	L	Aristae missing
3130	LD02456	AA202301	6e −26	Intron	L	SL	—	Fused tarsal seg.
3199	LD21713	AA735667	<1e −100	−56	L	L	SL, rough eye	SL, def. tarsi
3219	LD07107	AA263927	<1e −100	566	L	L	L	L

ab., abnormalities; ACV, anterior crossvein; def., deformed; FR, flanking region; int., interrupted; L, lethal; ND, not determined; seg., segments; SL, semilethal; #, number; —, normal.

[a] Transcription start site (+1) of each cloned gene or EST was defined as the 5′-most ends of mRNA sequence reported so far. The vector insertion sites were determined based on the sequences of the 5′ end of induced transcripts.

[b] One of the most prominent visible phenotypes is shown for each GAL4/GS combination. The detailed phenotypic information is available upon request.

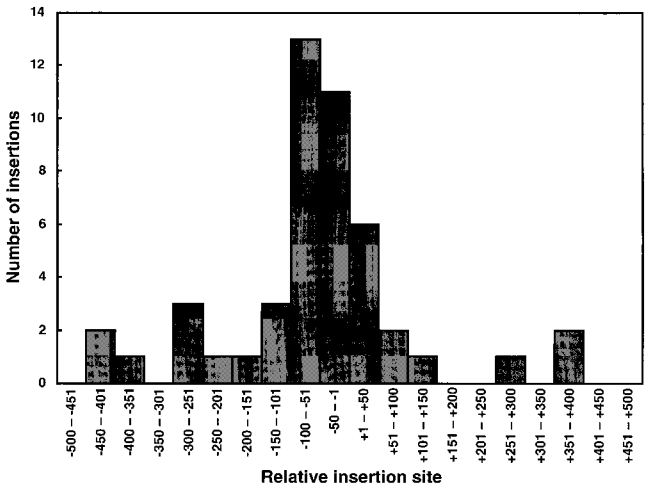


Figure 5.—Insertion sites of GS vector in 47 lines with a forced expression phenotype. The insertion sites relative to the transcription start site were investigated for the insertions in cloned genes or ESTs. The 5′ most ends of mRNA reported so far were defined as +1. Insertions mapped between −500 and +500 are indicated. Insertions in an intron were not included. More than 50% of insertions were found between −150 and +100, most frequently in between −100 and −1.

29BD-GAL4 or *c355-GAL4*, and a rough eye phenotype when driven by *sev-GAL4* (Figure 6). *Rap2l* gene was localized to cytological map 60B by chromosomal *in situ* hybridization. It is indeed an active gene in the wild-type flies, since a cDNA encoding RAP2L was obtained by RT-PCR using poly(A)$^+$ RNA prepared from the wild-type larvae as a template. A full-length cDNA encoding RAP2L was obtained by 5′ and 3′ RACE, which was 661 bases long with 5′- and 3′-untranslated regions of 47 and 65 bases, respectively. DNA sequence analysis revealed that the protein-coding region and 3′-untranslated region (UTR) were identical to those of the misexpressed transcript (Figure 7A). Based on the difference of 5′-UTR between the wild-type and the misexpressed transcripts, the insertion site of the GS vector was determined to be 339 bp upstream of the transcription start site (Figure 7A). Recently, the genome sequencing of this region (P1 clone DS00543) has been completed by BDGP (GenBank accession number AC004642), and has revealed that the *Rap2l* protein-coding region was interrupted by three introns. Sequence comparison between the genomic region and cDNA demonstrated that the misexpressed transcript was spliced and polyadenylated at exactly the same sites as the wild-type transcript (Figure 7A). Note that a single-pass sequencing read of 602 bases for the misexpressed transcript was long enough to reach the second exon that is separated from the first exon by 1160 bases of an intron.

Figure 7B shows the deduced protein sequence of the RAP2L compared to those of human Rap2 and Drosophila RAP1, whose gain-of-function mutation is known as *Roughened* (Hariharan *et al.* 1991). RAP2L protein is 182-amino-acid residues long and contains a GTP-binding domain shared by the Ras family proteins.

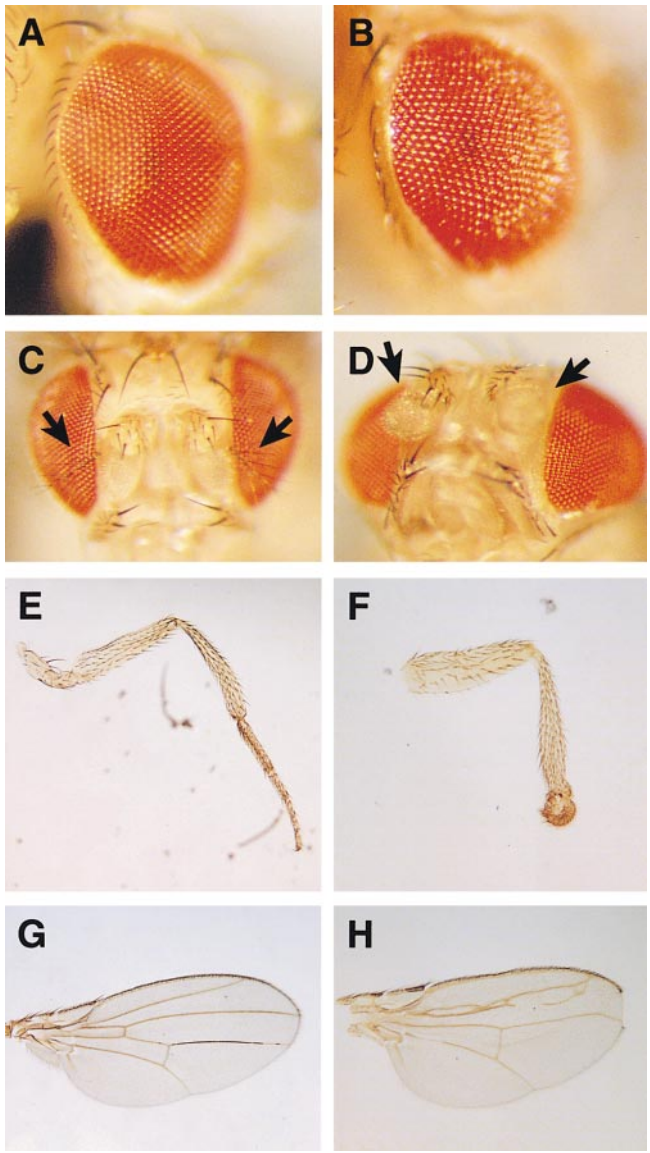
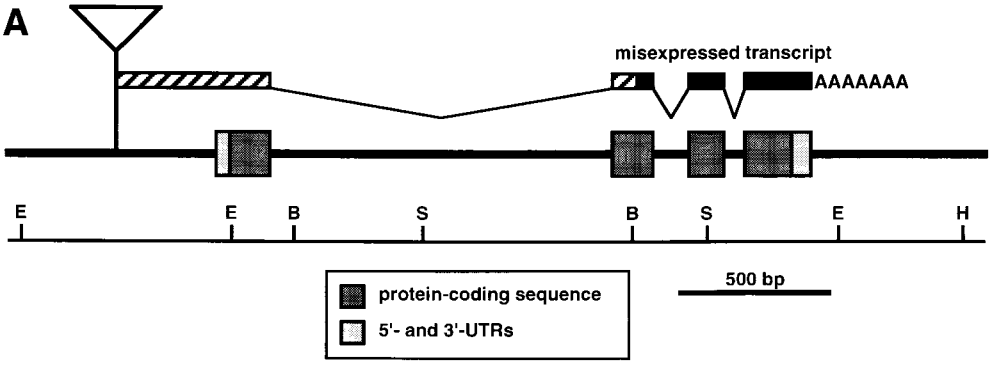


Figure 6.—Examples of phenotypes caused by forced expression of vector-flanking DNA. (A, C, E, G) Wild-type eye, aristae (indicated by arrows), leg, and wing, respectively. (B) A rough eye phenotype of *sev-GAL4*/*GS2069* fly. (D, F, H) Phenotypes observed in *GS1051*/+;*dpp-GAL4*/+ adults: (D) missing aristae (indicated by arrows), (F) leg with fused tarsal segments, and (H) wing with complex defects, such as fused, missing, or disrupted veins and notched margin.

The amino acid sequence was 68 and 55% identical to human Rap2 and Drosophila RAP1, respectively. The N-terminal residues including the GTP-binding domain were highly conserved among the three proteins, while the similarity was less for the carboxy termini.

Identification of a gene similar to human *mGST*: Line GS1051 had a GS vector insertion in a gene whose sequence is similar to human *microsomal glutathione S-transferase* (*mGST*), which encodes an enzyme involved in the detoxification defense system (DeJong *et al.* 1988). Misexpression of the vector-flanking DNA resulted in semilethality in combination with each of the four GAL4 drivers. All viable flies showed visible phenotypes in vari-



B

```
RAP2L          1 MREFKVVVLGSGGVGKSALTVQFVSGCFIEKYDPTIEDFYRKEIEVDSSPCVLEILDTAG  60
Human Rap2     1 MREYKVVVLGSGGVGKSALTVQFVTGTFIEKYDPTIEDFYRKEIEVDSSPSVLEILDTAG  60
Drosophila RAP1 1 MREYKIVVLGSGGVGKSALTVQFVQCIFVEKYDPTIEDSYRKQVEVDGQQCMLEILDTAG  60

RAP2L         61 TEQFASMRDLYIKNGHGFIVMYSLTNHQTFQDISSMKNVITRVKGSQPAPILLVANKFDL 120
Human Rap2    61 TEQFASMRDLYIKNGQGFILVYSLVNQQSFQDIKPMRDQIIRVKRYEKVPVILVGNKVDL 120
Drosophila RAP1 61 TEQFTAMRDLYMKNGQGFVLVYSITAQSTFNDLQDLREQILRVKDTDDVPMVLVGNKCDL 120

RAP2L        121 DCQREVSTAEGNALAQLWDCPFIEASAKDRINVNEVFATIVREMNLTQENRQKKNYCC-- 178
Human Rap2   121 ESEREVSSSEGRALAEEWGCPFMETSAKSKTMVDELFAEIVRQMNYAAQ-PDKDDPCCSA 179
Drosophila RAP1 121 EEERVVGKELGKNLATQFNCAFMETSAKAKVNVNDIFYDLVRQINKKSPEKKQKKPKKSL 180

RAP2L        179 CTLL 182
Human Rap2   180 CNIQ 183
Drosophila RAP1 181 CVLL 184
```

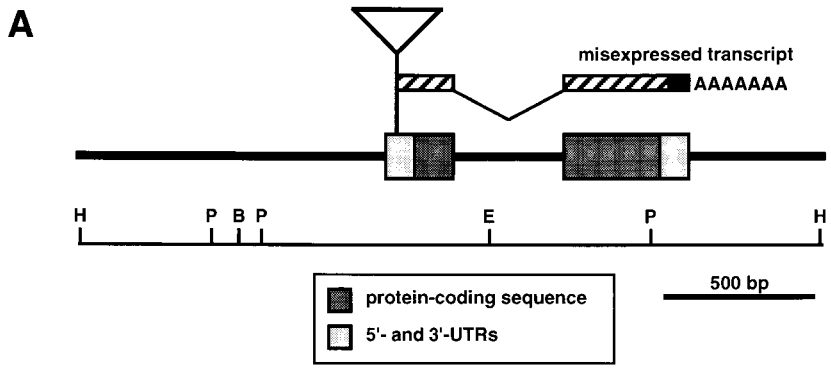
Figure 7.—(A) Diagram of genomic organization of *Rap2l* and the structure of misexpressed transcript. The sequence of cDNA encoding RAP2L was obtained by 5′ and 3′ RACE using poly(A)$^+$ RNA prepared from the wild-type larvae as a template. The exon-intron boundaries were determined on the basis of a comparison of sequences between a full-length cDNA and the genomic region containing *Rap2l* (P1 clone DS00543, GenBank accession number AC004642). Exons are represented by boxes with protein-coding sequence (dark gray) and UTR (light gray). Misexpressed transcript starting from the GS vector (indicated by a triangle) was spliced and polyadenylated as the wild-type transcript. The sequence corresponding to the hatched region was obtained by a single-pass sequencing, and used as a query for an initial search of databases. B, E, H, and S on the solid line at the bottom represent the restriction sites of *Bgl*II, *Eco*RI, *Hin*dIII, and *Spe*I, respectively. (B) Aligned amino acid sequences of RAP2L, human Rap2, and Drosophila RAP1. Identical residues for all three proteins are reverse-contrasted and those shared by two of them are highlighted. In residues of RAP2L, 68% (124/182) and 55% (100/182) were identical to human Rap2 and Drosophila RAP1, respectively.

ous body parts: a strong rough eye phenotype was induced by *29BD-GAL4* and *sev-GAL4*; a mild rough eye phenotype was produced by *c355-GAL4* and *dpp-GAL4*; and missing aristae, fused tarsal segments, reduced size of scutellum, some missing macrochaetae, and notched wings were induced by *29BD-GAL4*, *c355-GAL4*, and *dpp-GAL4*. Examples of such visible phenotypes caused by *dpp-GAL4* are shown in Figure 6.

On the basis of sequence similarity, the gene was named *Microsomal glutathione S-transferase-like* (*Mgstl*) and localized to 19E by chromosomal *in situ* hybridization. A cDNA encoding MGSTL was amplified by RT-PCR using mRNA prepared from the wild-type larvae, and a full-length cDNA sequence was determined by 5′ and 3′ RACE. Analysis by 5′ RACE revealed that the GS vector was inserted 39 bp downstream of the transcription start site, and 60 bp upstream of the first ATG codon for translation (Figure 8A). The 675-bases-long wild-type full-length cDNA contained an open reading frame encoding MGSTL. Sequences of the protein-coding region and 3′-UTR were identical to those of the misexpressed transcript (Figure 8A). The genomic region containing *Mgstl* was obtained by PCR and inverse PCR, and revealed that there is only one intron (378 bases) within this gene. Comparison of the sequences between the wild-type and the misexpressed transcript for *Mgstl* demonstrated that they were spliced and polyadenylated at exactly the same sites (Figure 8A). Figure 8B shows the deduced amino acid sequence of MGSTL consisting of 152 residues compared with that of human mGST consisting of 155 residues. The *Mgstl* intron position corresponded to the second intron in the human *mGST* gene, which contains three introns (Kelner *et al.* 1996). Although the sequence identity to human mGST was 45%, the hydrophobicity profile was very similar (data not shown). A single-pass sequencing was sufficient to obtain sequence information for the second exon, which contains a region with a high similarity to human mGST.

DISCUSSION

Gain-of-function screening based on misexpression phenotypes is an alternative to a loss-of-function screening approach to discover new genes (Miklos and Rubin 1996; Rørth 1996; Crisp and Merriam 1997; Rørth *et al.* 1998). In a loss-of-function screen, Cooley *et al.* (1988) showed that 15% of new *P*-element insertions caused phenotypes. This frequency may not correspond to the frequency of gene disruption by *P* elements. There must have been insertions that disrupted a gene, but phenotypic defects were not detected, because the



B

```
MGSTL        1 MASPVELLSLSNPVFKSFTFWVGVLVIKMLLMSLLTAIQRFNTKTFANPED-LMSPK---  56
Human mGST   1 --MVDLTQVMDDEVFMAFASYATIILSKMMLMSTATAFYRLTRKVFANPEDCVAFGKGEN  58

MGSTL       57 -LKVKFDDPNVERVRRAHRNDLENILPFFAIGLLYVLTDPAAFLAINLFRAVGIARIVHT 115
Human mGST  59 AKKYLRTDDRVERVRRAHLNDLENITPFLGIGLLYSLSGPDPSTAILHFRLFVGARIYHT 118

MGSTL      116 LVYAVVVVPQPSRALAFFVALGATVYMALQVIASAAF-                       152
Human mGST 119 IAY-LTPLPQPNRALSFFVGYGVTLSMAYRLLKSKLYL                       155
```

Figure 8.—(A) Diagram of genomic organization of *Mgstl* and the structure of misexpressed transcript. The sequence of cDNA encoding MGSTL was obtained by 5′ and 3′ RACE using poly(A)$^+$ RNA prepared from the wild-type larvae as a template. The genomic region was obtained by PCR and inverse PCR, subcloned into plasmid, and sequenced. Exons are represented by boxes with protein-coding sequence (dark gray) and UTR (light gray). Misexpressed transcript starting from the GS vector (indicated by a triangle) was spliced and polyadenylated as the wild-type transcript. The sequence of the hatched region was obtained by a single-pass sequencing, and used for an initial search of databases. B, E, H, and P on the solid line at the bottom represent the restriction sites of *Bgl*II, *Eco*RI, *Hin*dIII, and *Pst*I, respectively. (B) Aligned amino acid sequences of MGSTL and human mGST. In residues of MGSTL, 45% (68/152) were identical to human mGST, which is reverse-contrasted.

defects were compensated by functionally overlapping genes, or phenotypes were too subtle for detection (Miklos and Rubin 1996). Considering that the *P* elements are frequently inserted upstream of transcription start sites (Spradling *et al.* 1995), some *P*-element insertions might be near a gene but not disrupt their function. These two categories of insertions could be potentially detected with the gain-of-function mutagenesis; ectopic expression of genes in various tissues would increase the probability of producing a phenotype, and the site preference of *P*-element insertions is in favor of causing misexpression of a gene.

A high-efficiency vector in terms of inducing phenotypic changes would be valuable as a tool for functional mapping of the genome through discovery of genes on the basis of phenotypes and obtaining sequence information associated with them. The GS vector used in this study appeared to be very efficient in terms of phenotype frequency. The EP element constructed by Rørth (1996) was the first vector used for a systematic gain-of-function mutagenesis. Crisp and Merriam (1997) constructed another version of misexpression vector with the *yellow* gene as a marker, which is convenient for identification of flies carrying the misexpression vector and a GAL4 driver containing the *white* gene as a marker. We showed that the frequency of GAL4-dependent phenotypes was extremely high with the GS vector, 10-fold higher than those obtained with the EP element (Rørth 1996; Rørth *et al.* 1998). For instance, the frequency of phenotypes with *sev-GAL4* was 38% for GS inserts, while the same driver induced phenotypic changes in 4% of EP inserts (Rørth 1996). Likewise, 32% of GS inserts showed phenotypes with *dpp-GAL4*, while only 2% of EP inserts had phenotypes with the same driver (Rørth *et al.* 1998). The EP element has the UAS enhancer/core promoter near the 3′ end of the *P* element only, while the GS vector contains the UAS enhancer/core promoter at both ends of the vector. Although this modification increases by 2-fold the probability of inducing forced expression of genes, this cannot account for a 10-fold difference. The frequency of phenotypes also depends on the criteria of mutant phenotype, especially for visible phenotype. However, this is unlikely to be the cause of the 10-fold difference in the frequency of phenotypes, because the frequency of lethal phenotype (which is unambiguous) was extremely high for GS inserts compared to EP inserts (7.3 *vs.* 0.3% for *dpp-GAL4* and 14 *vs.* 0.3% for *sev-GAL4*, respectively). Comparable data for the same GAL4 drivers are not available for the vector constructed by Crisp and Merriam (1997). The high frequency of mutant phenotypes involving the GS vector must be attributed to its unique structure. The mechanism for the high efficiency is not clear, but it may be due to its insertion frequency near genes, or more likely due to the efficiency of forced expression of flanking DNA.

In the GS system, GAL4-dependent phenotypic changes simply indicate the presence of a gene near the vector insertion site, and this is sufficient for rapid detection and identification of new genes. For the functional mapping of the genome, it is important to obtain reliable molecular information from the insertion site that is

associated with a phenotype. We have established a procedure for obtaining the sequence of misexpressed transcripts derived from an insertion. We used RT-PCR using vector-specific and oligo(dT) primers, followed by single-pass sequencing. Although it requires more steps compared to inverse PCR using the genome DNA as a template, mRNA sequences are more informative than genomic sequences, which might contain noncoding sequences. In fact, it was indeed the case for *Rap2l* and *Mgstl* genes, which we characterized in this study. The misexpressed transcripts of these genes were spliced correctly, and a single-pass sequencing of the RT-PCR products was sufficient to reach the second exon of each gene.

On the basis of the sequence similarity to a human Rap2, we identified a new gene, *Rap2l*. The amino acid sequence of RAP2L was also similar to that of Drosophila RAP1, the only member of the Rap family known in Drosophila. Dominant mutations of *Rap1* have been shown to genetically interact with *fat facets* in eye development (Li *et al.* 1997). Human Rap2 has been thought to be a modifier of the Ras signaling pathway (Pizon *et al.* 1988), and it has been shown to have GTP-binding activity and a low intrinsic GTPase activity (Lerosey *et al.* 1991). However, the function of human Rap2 has not been clearly demonstrated. Overexpression of Rap2 in cultured cells had no effects on cellular proliferation or transformation induced by the *ras* oncogene (Jimenez *et al.* 1991). Nevertheless, a high conservation of amino acid sequences between humans and Drosophila suggests that Rap2 has some important functions. *In vivo* studies using Drosophila mutants should facilitate understanding the function of Rap2 proteins. Loss-of-function studies are especially necessary to define cellular function. The *P*-element insert in the *Rap2l* locus is useful for generating loss-of-function alleles by local transposition (Tower *et al.* 1993) or by excising the vector from the chromosome, which occasionally deletes flanking DNA (Salz *et al.* 1987). The same is true for *Mgstl*, which encodes a protein similar to human mGST. The identity was 45%, but the hydrophobicity profile was very similar, suggesting that they share a functional similarity. Studies on mGST using Drosophila mutants should provide evidence for its *in vivo* function.

The progress of genome sequencing and the EST project is important for the functional mapping of the genome on the basis of gain-of-function screens. A partial sequence of cDNAs derived from misexpressed transcripts would be sufficient for identifying genomic DNA clones or cDNA clones that are available from BDGP through commercial vendors, which facilitates further analysis of individual genes. We found that 18% of the insertions with forced-expression phenotypes showed sequence similarity to ESTs. Since the EST data are rapidly growing, the GS system will identify many more genes corresponding to ESTs. Gain-of-function phenotypes obtained by forced expression of EST-corresponding genes might provide a clue as to their functions. The system may also identify genes that may not be found as an EST, such as those expressed normally at very low levels, expressed in a few cells, or expressed only transiently during development. The GS system should contribute to functional genomics as a method for easy detection and rapid molecular identification of genes in the Drosophila genome, and the obtained inserts will serve as materials to start loss-of-function studies on the new genes.

We thank S. Kawasaki, M. Matsuno, and T. Umemiya for contributing to the screening, A. Nose, Y. Fuyama, J. Merriam, M. Wolfner, and K. White for comments on the manuscript, and the Bloomington Stock Center for providing fly stocks. This work was supported in part by a Human Frontier Science Program (HFSP) grant (RG-377/93 B).

LITERATURE CITED

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang *et al.*, 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:** 3389–3402.

Banfi, S., G. Borsani, E. Rossi, L. Bernard, A. Guffanti *et al.*, 1996 Identification and mapping of human cDNAs homologous to Drosophila mutant genes through EST database searching. Nat. Genet. **13:** 167–174.

Bellen, H. J., C. J. O'Kane, C. Wilson, U. Grossniklaus, R. K. Pearson *et al.*, 1989 *P*-element-mediated enhancer detection: a versatile method to study development in Drosophila. Genes Dev. **3:** 1288–1300.

Bier, E., H. Vaessin, S. Shepherd, K. Lee, K. McCall *et al.*, 1989 Searching for pattern and mutation in the Drosophila genome with a P-*lacZ* vector. Genes Dev. **3:** 1273–1287.

Brand, A. H., and N. Perrimon, 1993 Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. Development **118:** 401–415.

Brunner, D., K. Dücker, N. Oellers, E. Hafen, H. Scholz *et al.*, 1994 The ETS domain protein pointed-P2 is a target of MAP kinase in the Sevenless signal transduction pathway. Nature **370:** 386–389.

Chen, F., M. Barkett, K. T. Ram, A. Quintanilla and I. K. Hariharan, 1997 Biological characterization of Drosophila Rapgap1, a GTPase activating protein for Rap1. Proc. Natl. Acad. Sci. USA **94:** 12485–12490.

Chiba, A., P. Snow, H. Keshishian and Y. Hotta, 1995 Fasciclin III as a synaptic target recognition molecule in Drosophila. Nature **374:** 166–168.

Cooley, L., R. Kelley and A. Spradling, 1988 Insertional mutagenesis of the Drosophila genome with single *P* elements. Science **239:** 1121–1128.

Crisp, J., and J. Merriam, 1997 Efficiency of an F1 selection screen in a pilot two-component mutagenesis involving *Drosophila melanogaster* misexpression phenotypes. Dros. Inf. Serv. **80:** 90–92.

Deák, P., M. M. Omar, R. D. Saunders, M. Pál, O. Komonyi *et al.*, 1997 *P*-element insertion alleles of essential genes on the third chromosome of *Drosophila melanogaster*: correlation of physical and cytogenetic maps in chromosomal region 86E-87F. Genetics **147:** 1697–1722.

DeJong, J. L., R. Morgenstern, H. Jornvall, J. W. DePierre and C. P. Tu, 1988 Gene expression of rat and human microsomal glutathione *S*-transferases. J. Biol. Chem. **263:** 8430–8436.

Hariharan, I. K., R. W. Carthew and G. M. Rubin, 1991 The Drosophila *Roughened* mutation: activation of a *rap* homolog disrupts eye development and interferes with cell determination. Cell **67:** 717–722.

Harrison, D. A., R. Binari, T. S. Nahreini, M. Gilman and N. Perrimon, 1995 Activation of a Drosophila Janus kinase (JAK) causes hematopoietic neoplasia and developmental defects. EMBO J. **14:** 2857–2865.

Jimenez, B., V. Pizon, I. Lerosey, F. Beranger, A. Tavitian *et al.*,

1991 Effects of the ras-related rap2 protein on cellular proliferation. Int. J. Cancer **49:** 471–479.

Kelner, M. J., M. N. Stokely, N. E. Stovall and M. A. Montoya, 1996 Structural organization of the human microsomal glutathione *S*-transferase gene (GST12). Genomics **36:** 100–103.

Lerosey, I., P. Chardin, J. De Gunzburg and A. Tavitian, 1991 The product of the *rap2* gene, member of the *ras* superfamily. Biochemical characterization and site-directed mutagenesis. J. Biol. Chem. **266:** 4315–4321.

Li, Q., I. K. Hariharan, F. Chen, Y. Huang and J. A. Fischer, 1997 Genetic interactions with *Rap1* and *Ras1* reveal a second function for the Fat facets deubiquitinating enzyme in Drosophila eye development. Proc. Natl. Acad. Sci. USA **94:** 12515–12520.

McGarry, T. J., and S. Lindquist, 1986 Inhibition of heat shock protein synthesis by heat-inducible antisense RNA. Proc. Natl. Acad. Sci. USA **83:** 399–403.

Miklos, G. L., and G. M. Rubin, 1996 The role of the genome project in determining gene function: insights from model organisms. Cell **86:** 521–529.

Nicole, L. M., and R. M. Tanguay, 1987 On the specificity of antisense RNA to arrest *in vitro* translation of mRNA coding for Drosophila hsp23. Biosci. Rep. **7:** 239–246.

Pizon, V., P. Chardin, I. Lerosey, B. Olofsson and A. Tavitian, 1988 Human cDNAs *rap1* and *rap2* homologous to the Drosophila gene *Dras3* encode proteins closely related to ras in the 'effector' region. Oncogene **3:** 201–204.

Robertson, H. M., C. R. Preston, R. W. Phillis, D. M. Johnson-Schlitz, W. K. Benz *et al.*, 1988 A stable genomic source of *P* element transposase in *Drosophila melanogaster*. Genetics **118:** 461–470.

Rørth, P., 1996 A modular misexpression screen in Drosophila detecting tissue-specific phenotypes. Proc. Natl. Acad. Sci. USA **93:** 12418–12422.

Rørth, P., K. Szabo, A. Bailey, T. Laverty, J. Rehm *et al.*, 1998 Systematic gain-of-function genetics in Drosophila. Development **125:** 1049–1057.

Rubin, G. M., and A. C. Spradling, 1982 Genetic transformation of Drosophila with transposable element vectors. Science **218:** 348–353.

Salz, H. K., T. W. Cline and P. Schedl, 1987 Functional changes associated with structural alterations induced by mobilization of a *P* element inserted in the *Sex-lethal* gene of Drosophila. Genetics **117:** 221–231.

Salzberg, A., S. N. Prokopenko, Y. He, P. Tsai, M. Pál *et al.*, 1997 *P*-element insertion alleles of essential genes on the third chromosome of *Drosophila melanogaster*: mutations affecting embryonic PNS development. Genetics **147:** 1723–1741.

Sidow, A., and W. K. Thomas, 1994 A molecular evolutionary framework for eukaryotic model organisms. Curr. Biol. **4:** 596–603.

Spradling, A. C., D. M. Stern, I. Kiss, J. Roote, T. Laverty *et al.*, 1995 Gene disruptions using *P* transposable elements: an integral component of the Drosophila genome project. Proc. Natl. Acad. Sci. USA **92:** 10824–10830.

Staehling-Hampton, K., P. D. Jackson, M. J. Clark, A. H. Brand and F. M. Hoffmann, 1994 Specificity of bone morphogenetic protein-related factors: cell fate and gene expression changes in Drosophila embryos induced by decapentaplegic but not 60A. Cell Growth Differ. **5:** 585–593.

Thummel, C. S., A. M. Boulet and H. D. Lipshitz, 1988 Vectors for Drosophila *P*-element-mediated transformation and tissue culture transfection. Gene **74:** 445–456.

Tower, J., G. H. Karpen, N. Craig and A. C. Spradling, 1993 Preferential transposition of Drosophila *P* elements to nearby chromosomal sites. Genetics **133:** 347–359.

Xu, T., and G. M. Rubin, 1993 Analysis of genetic mosaics in developing and adult Drosophila tissues. Development **117:** 1223–1237.

Communicating editor: T. C. Kaufman