Drosophila Problem Set

1) You have generated P element transformants of a construct that contains the mini-white gene. This gene confers an orange eye color in a homozygous white mutant background. For each line you have performed the following crosses:

   i) Initial orange eyed transformant is female in all cases.
   ii) Cross each individual transformant to w; CyO/Sp; TM3/Dp males.
   iii) Select orange eyed male that carries CyO and TM3 and cross to females homozygous for white.
   iv) Score resulting progeny for the presence/absence of the mini-white gene and the two balancer chromosomes (results of these crosses are listed below for each line).

For each Stock in which chromosome is P element located?

a) Stock 2: 220 flies total: 114 are P[w+]: 56 carry TM3; none carry CyO
   106 are w-: 32 carry TM3; all carry CyO.

b) Stock 3: 216 flies total: 150 are P[w+]: None carry TM3; 80 carry CyO
   66 are w-: all carry TM3; 36 carry CyO.

c) Stock 1: 102 total flies 76 are P[w+]: 38 are TM3; 24 are CyO; and 50 are female.
   26 are w-: all are CyO; 14 are TM3; all are male.

d) Stock 4: 200 flies total 148 flies are P[w+]: 52 are neither CyO nor TM3; 48 are TM3 and 48 are CyO.
   52 are w-: All are CyO and TM3
2) In a genetic screen in Drosophila you identified 4 lethal alleles of a third chromosomal gene, Gene X, required for nervous system development. In embryos homozygous mutant for Gene X the aCC neuron fails to develop on average 50% of the time.

Part I) Provide three distinct models to explain why these alleles appear to exhibit 50% expressivity of the mutant phenotype (loss of aCC). Explain each choice.

Part II) Due to the mutant phenotype you decide to clone Gene X. All four alleles were generated in a multiply marked third chromosome (see below) and are balanced over TM3 Sb e. To map the allele you carried out the following crosses and obtained the below listed results.

ru h st cu sr e ca - m[1]/TM3 Sb x wild-type
ru h st cu sr e ca - m[1]/+ virgins x ru h st cu sr e ca [m2]/TM3 Sb males

(Score non-Stubble progeny for presence of markers)

Genetic Map positions of the corresponding genes
ru (0.0) ------ h (26) ------ st (44) ------ cu (50) ------ sr (62) ------ e (71) ------ ca (100)

Resulting progeny:

<table>
<thead>
<tr>
<th>Genotype</th>
<th># of flies with indicated phenotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>ru</td>
<td>1250</td>
</tr>
<tr>
<td>ru h</td>
<td>950</td>
</tr>
<tr>
<td>ru h st</td>
<td>350</td>
</tr>
<tr>
<td>ru h st cu</td>
<td>200</td>
</tr>
<tr>
<td>ru h st cu sr</td>
<td>0</td>
</tr>
<tr>
<td>ru h st cu sr e</td>
<td>0</td>
</tr>
<tr>
<td>ru h st cu sr e ca</td>
<td>0</td>
</tr>
<tr>
<td>h st cu sr e ca</td>
<td>0</td>
</tr>
<tr>
<td>st cu sr e ca</td>
<td>0</td>
</tr>
<tr>
<td>cu sr e ca</td>
<td>0</td>
</tr>
<tr>
<td>sr e ca</td>
<td>400</td>
</tr>
<tr>
<td>e ca</td>
<td>450</td>
</tr>
<tr>
<td>ca</td>
<td>1500</td>
</tr>
</tbody>
</table>

a) Between which two markers does gene X map?

b) What is the relative distance between gene X and these two markers?

c) The physical distance between the two markers that flank gene X is 8 Mb. What is the average distance between crossovers in this interval given the number of recombinants obtained in this interval in the above experiment.
Part III) To map gene X to a small, defined molecular region you used genetic mapping with SNPs. You identified a panel of 10 SNPs between the ru cu ca – m[1] chromosome and the wild-type chromosome used in the above mapping experiments. These SNPs are spaced at ~100 kb intervals in the presumed genetic map position of Gene X. The SNP is defined as the sequence of the wild-type chromosome. You then determined for most of the recombinant chromosomes in this interval whether they contained the SNP of the mapping chromosome or the sequence from the original mutant chromosome.

Below are the results of these studies for the SNPs found to be relevant to map the left and right limit of Gene X.

Crossovers to the right of gene X (from a total of 150 crossovers tested):
SNP-B: 147/150 recombinant chromosomes contain SNP B
SNP-C: 125/150 recombinant chromosomes contain SNP C
SNP-A: 101/150 recombinant chromosomes contain SNP A

Crossovers to the left of gene X (from a total of 150 crossovers tested):
SNP-X: 22/150 recombinant chromosomes contain SNP X
SNP-R: 49/150 recombinant chromosomes contain SNP R
SNP-Z: 67/150 recombinant chromosomes contain SNP Z
SNP-T: 89/150 recombinant chromosomes contain SNP T
SNP-S: 108/150 recombinant chromosomes contain SNP S
SNP-V: 131/150 recombinant chromosomes contain SNP V
SNP-Y: 147/150 recombinant chromosomes contain SNP Y

Draw a molecular map that indicates the relative position of each SNP, the location of Gene X and the number of crossovers between Gene X and the closest flanking SNPs.

Part IV) Database searches identify 12 genes in the molecular interval that contains gene X.

a) Please outline your strategy for identifying which gene is most likely to encode gene X? Explain your logic
b) Also, outline your strategy for providing definitive evidence that one of these genes encodes gene X? Explain your logic
The hippo/warts signaling pathway is a relatively newly discovered signaling pathway that regulates cell proliferation and cell death in flies. Mutations in hippo, salvador, or warts lead to increased cell proliferation, decreased cell death, and cause at the molecular level a cell-autonomous upregulation of the levels of the Expanded protein (red in figure). The molecular signal that triggers activation of this pathway remains elusive, as hippo, salvador and warts all function in the signal-receiving cell to mediate signal transduction.

In a genetic screen you identify mutations in two new genes (genes A and B) that yield mutant phenotypes identical to those of hippo, salvador and warts. These mutations complement each other as well as mutations in all known members of the hippo/warts signaling pathway. To determine if either gene functions in the signal-sending cell, you use FLP/FRT-based methods to create mitotic clones homozygous mutant for each gene. You use GFP as your cell-autonomous marker to identify mutant clones unambiguously by their lack of GFP expression (a transgene is used to drive GFP under the control of a ubiquitously expressed promoter), and you assess Expanded protein levels within the mutant clones. Being a careful scientist you also create mitotic clones of a wild-type chromosome as your control. The results of the clonal analysis are shown in the accompanying figure. Please answer the following questions.

A) Which gene functions in a cell autonomous manner and why?

B) Which gene functions in a cell non-autonomous manner and why?

C) Which gene is the better candidate to encode the sought after signaling molecule that activates the hippo pathway.

D) 1. What is the genotype of the cells (dark green) marked by arrow #1.
   2. What is the genotype of the cells (light green) marked by arrow #2.

(Assume gene B is in the left arm of chromosome 2; please include all possible relevant information: e.g. FRTs, GFP, centromere...).
4) Loss of one copy of the *warts* gene partially suppresses the phenotype of overexpressing *expanded*, or of overexpressing *hippo*. These data are inferred to mean that Wts likely acts downstream of Expanded and Hippo in the same pathway. Please explain the logic as to why the ability of a 50% reduction in the function of gene A to suppress the overexpression phenotype of gene B is taken to suggest that gene A acts downstream, but not upstream, of gene B.

5) Following the protocol of Zhai et al (attached paper) you mapped your favorite complementation group using four P elements (as per Fig. 2). The schematic and list below indicate the molecular locations of each P element and the number of flies scored for each P element.

P[1]: 85/4297 (# of white flies/total)
P[2]: 58/10,000
P[3]: 31/9419
P[4]: 53/8426

Given the above information, please –

A) Determine the projected molecular distance between the different pairs of P element and the mutation (see Figure 2 in the Zhai paper for examples).

B) Determine the recombination rate between each P element in cM/MB.

C) Does the recombination rates vary significantly within this ~900 kb region?
   a. If so, provide a molecular explanation for why recombination rates are not equivalent within this region.
Expanded protein: red
(darker red = more protein)

GFP: green;
Mutant clone: no GFP

Merge of GFP and Expanded

Wild type control

Gene A

Gene B
Mapping *Drosophila* mutations with molecularly defined *P* element insertions


*Howard Hughes Medical Institute, †Program in Developmental Biology, ‡Department of Molecular and Human Genetics, §Texas Children’s Cancer Center, and ¶Division of Neuroscience, Baylor College of Medicine, Houston, TX 77030

Edited by Michael S. Levine, University of California, Berkeley, CA, and approved July 11, 2003 (received for review May 7, 2003)

The isolation of chemically induced mutations in forward genetic screens is one of the hallmarks of *Drosophila* genetics. However, mapping the cost common trait and is identifying the molecular lesions associated with these mutations are often difficult and labor-intensive. Two mapping methods are most often used in flies: meiotic recombination mapping with marked chromosomes and deficiency mapping. The availability of the fly genome sequence allows the establishment and usage of molecular markers. Single-nucleotide polymorphisms have therefore recently been used to map several genes. Here we show that thousands of molecularly mapped *P* element insertions in fly strains that are publicly available provide a powerful alternative method to single-nucleotide polymorphism mapping. We present a strategy that allows mapping of lethal mutations, as well as viable mutations with visible phenotypes, with minimal resources. The most important unknown in using recombination rates to map at high resolution is how accurately recombination data correlate with molecular maps in small intervals. We therefore surveyed distortions of recombination rates in intervals <500 kb. We document the extent of distortions between the recombination and molecular maps and describe the required steps to map with an accuracy of <50 kb. Finally, we describe a recently developed method to determine molecular lesions in 50-kb intervals by using a heteroduplex DNA mutation detection system. Our data show that this mapping approach is inexpensive, efficient, and precise, and that it significantly broadens the application of *P* elements in *Drosophila*.

Forward genetic screens in *Drosophila* using chemical mutagens or γ-ray irradiation have uncovered thousands of complementation groups (1, 2). However, identifying the molecular lesions that underlie the phenotypes has traditionally been labor-intensive and time-consuming. All mapping strategies depend on a detectable phenotype caused by the mutation (e.g., lethality) to establish its genetic linkage with known markers or failure to complement characterized deficiencies. One of the most common strategies is meiotic recombination mapping, and resolution and accuracy primarily depend on the availability of markers and the number of meiotic events that are scored.

The publication of the fly genome (3) made it possible to generate many more precisely mapped markers on the basis of single-nucleotide polymorphisms (SNPs). SNPs are single base-pair differences between homologous chromosomes. The use of SNPs as molecular markers is a two-step process. First, SNP maps have to be established or confirmed; second, SNPs have to be detected by either sequencing or restriction enzyme digestion. Several teams have recently shown that SNPs are present in sufficient numbers in the *Drosophila* genome to theoretically allow the mapping of every gene. They have also demonstrated the successful application of the method in a few cases (4–6). An alternative to SNPs are *P* elements with molecularly defined insertion sites. They are transposable elements that often have been engineered with a visible marker such as the "white" gene. The Berkeley *Drosophila* Genome Group Gene Disruption Project has generated >6,500 publicly available fly strains that carry molecularly mapped *P* elements. This implies that *P* elements are available, on average, every 20–30 kb. Here we test the applicability of these molecularly mapped *P* insertions as tools to map lethal complementation groups as an alternative to SNP mapping.

We have recently isolated numerous lethal complementation groups in a genetic screen by using ethyl methanesulfonate as a mutagen. To map these complementation groups in parallel, we compared the cost and effort of several mapping strategies: meiotic recombination with visible markers or molecularly mapped *P* insertions, male recombination, deficiency mapping, and SNP mapping. One of the major problems associated with deficiency mapping strategies and recombination mapping experiments by using visual markers is that they rarely provide a resolution of <300 kb in flies. This limitation is mostly due to a paucity of visible markers as well as to the incomplete coverage of the genome by deficiencies or the lack of information concerning the molecular breakpoints of the deficiencies. The male recombination technique (7) allows mapping to an interval between molecularly mapped *P* insertions and can map mutations to a small interval as close as *P* insertions are used. However, on the basis of our experience, male recombination has several major drawbacks that hamper a large-scale approach: (i) the necessity to recombine markers onto the chromosomes that carry the mutations to be mapped; (ii) the inability to define an approximate mapping position in each round (i.e., the mutation maps either to the left or the right of the *P* insertion, but no finer estimate is obtained); and (iii) the need to perform at least three mapping rounds. These drawbacks have led several teams to focus on SNP mapping approaches, especially for high-resolution mapping (4–6). However, SNP mapping requires costly molecular biology reagents and analyses. This restriction is especially an issue when one needs to establish high-resolution SNP maps. We therefore explored an alternative high-throughput and high-resolution mapping strategy by using the set of molecularly mapped *P* insertions that are publicly available (ref. 8; http://flypush.imgen bcm tmc edu/pscreen).

Theoretically, by using two molecularly mapped *P* insertions, a mutation can be mapped to a single nucleotide. The accuracy of this approach depends on the co-linearity of the physical and recombinational maps and the number of progeny scored. Recombination rates (RRs) vary greatly over the length of the chromosome; they are several-fold lower close to the centromeres and telemores than in the middle of the chromosome arm.
flies from P insertion 1 crosses × 100, and RD2 (cM) is the percentage of white-eyed flies from P insertion 2 crosses × 100.

For the region between P insertions that do not flank the mutation, the RR was calculated as the following. For example, if three P insertions were used and the mutation was between P insertions 1 and 2, \( RR_{(1,2)} \) was calculated from \( RR_{(1,3)} \) and \( RR_{(1,1)} \):

\[
RR_{(1,2)} = \frac{RR_{(1,3)} \times MD_{(1,2)} - RR_{(1,2)} \times MD_{(1,2)}}{MD_{(1,3)}}
\]

DNA Isolation, Primer Design, and PCR. Genomic DNA was prepared by homogenizing 10 heterozygous flies (mutant chromosome over isogenized chromosome) in 500 μl of squishing buffer (10 mM Tris, pH 8.0/1 mM EDTA/25 mM NaCl/0.2 mg/ml proteinase K), incubating at 55°C for 30–60 min, and inactivating proteinase K at 95°C for 4–5 min.

Primers were designed by using a modified version of the PRIMER3 algorithm from the Whitehead Institute, Massachusetts Institute of Technology (11) (www-genome.wi.mit.edu/genome_software/other/primer3.html). Our implementation for automatic tiled primer pair design for larger regions is publicly available at http://flypush.imagen.bcm.tmc.edu/primer. This algorithm is run in a loop to generate primer pairs covering 600 bp with a 100-bp overlap. Hence, 100 primer pairs with highly homogenous characteristics such as melting temperature and GC content are generated to cover 50 kb.

PCR was performed in a 20-μl volume including 1 μl of DNA and 10 pmol of each primer in a 96-well format. For the regions screened so far, 95–100% of the PCRs successfully amplified the DNA in the first round.

Results
Meiotic Recombination by Using Molecularly Mapped P Insertions.

The speed and efficiency of a recombinational-based mapping approach largely depend on the ease with which the marker(s) can be scored. Most P insertions of the Berkeley Drosophila Genome Group Gene Disruption Project are marked with white*, arguably the easiest marker to score. Hence, molecularly defined P insertions with a visible marker allow molecular mapping by scoring eye color alone. Fig. 1 summarizes the overall strategy. It is based on meiotic recombination mapping for the first two mapping steps: rough and fine (Fig. 1A). The crossing scheme is illustrated in Fig. 1B. We use two independently isolated alleles (1 and 2) that fail to complement each other to avoid mapping errors due to second-site hits. The issue of second-site hits has to be kept in mind when mapping complementation groups consisting of single alleles. Indeed, when 15 to 25 mM ethyl methanesulfonate is used, ∼50% of all chromosome arms bearing lethals carry a single lethal hit if a Poisson distribution is assumed (9).

In brief, mutant allele 1 of an essential complementation group is crossed to a homozygous viable w*-marked P insertion strain. Recombination occurs in F1 females heterozygous for allele 1 and the P insertion-bearing chromosome. These females

Materials and Methods
Drosophila Stocks.

All P insertion lines used in this study are listed at http://flypush.imagen.bcm.tmc.edu/pscreen and are made available by the Bloomington Stock Center.

Calculation of Projected Molecular Position (PMP) and RRs.

The PMP was calculated by using each pair of flanking P elements as in Fig. 1C. The RR of the region between a pair of P insertions that flank the mutation was calculated from the RDs:

\[
RR_{(1,2)} = \frac{RD1 + RD2}{MD_{(1,2)}} \times 10^{-8}
\]

where \( RR_{(1,2)} \) (cM/megabases) is the RR of the region between P elements 1 and 2, RD1 (cM) is the percentage of white-eyed offspring. Shown in the yellow box are the possible recombination events, which are color-coded and labeled A, B, and C, corresponding to the F1 female, where the three types of recombination events are marked likewise. (C) Calculation of mapping positions. PMP, projected molecular distance in base pairs; MD, molecular distance in base pairs; RD, recombination distance in cM. Note that the PMP can be calculated by using \( P_1 \) (as shown) or \( P_2 \) (PMDc).
are crossed to males heterozygous for allele 2 (or a deficiency strain that fails to complement allele 1) balanced over an *hs-hid* balancer (13). A single heat shock at 38°C for 1 h 4 d after setting up the cross kills all of the progeny that carry the balancer. Hence, in the *F*₂ offspring, the only progeny that survive if no crossing-over occurs have the genotype mutant allele 2/ *P{w¹¹m}*. These progeny are red-eyed (Fig. 1B, green box). White-eyed flies are the result of recombination events between the *P* insertion and mutant allele 1 (Fig. 1B, yellow box). The proportion of white-eyed flies in the *F*₂ progeny therefore represents the RD in cM between the mutation and *P* insertion (Fig. 1C). Because the *P* insertions are molecularly mapped, the molecular distance (MD, in base pairs) between any two *P* insertions can easily be calculated from their insertion sites. Hence, for every pair of *P* element insertions, a PMP of the mutation can be defined (Fig. 1C). Viable mutations can be mapped similarly if the phenotype of the homozygous mutants can easily be distinguished from the recombinant white-eyed flies.

For rough mapping, we have selected sets of 7–10 *P* insertions spanning each autosome arm shown in Table 2, which is published as supporting information on the PNAS web site, www.pnas.org. Mutant allele 1 is crossed to all *P* insertions in parallel, and at least 500 and, if possible, >1,000 *F*₂ progeny from each *P* insertion cross are scored to achieve an accuracy of <1 cM, which corresponds to an average interval of 500 kb (~25 megabases/55 cM per chromosome arm). Given the variable RRs in different regions of the chromosomes and the presence of double crossover events, we consider the PMP calculated from the two closest-flanking *P* insertions to be the most accurate. The rough mapping interval is determined as PMP ±1 cM. Lower RRs at centromeric and telomeric regions and higher RRs in the middle of the chromosome arms are known from low-resolution maps (9, 10). Correspondingly, we found that 1 cM in the telomeric cytological division 21 (2L) corresponds to 1,550 kb, whereas 1 cM in division 95 (3R) corresponds to 102 kb. The variability in RR over the entire chromosome does not, however, significantly affect the accuracy of mapping in cM but does affect the size of the mapping interval.

Once the rough mapping interval (~1 cM) is established, several strategies can be pursued. Deficiencies can be used to confirm and further define the mapping position. However, we favor a fine mapping strategy that uses the same principle as for the rough mapping strategy, because it provides highly consistent results. We use four to five *P* insertions spanning the interval defined by the rough mapping interval and score 10,000 *F*₂ progeny for each *P* insertion to achieve an accuracy of ~0.1 cM. The PMP of the mutation can then be determined as illustrated above (Fig. 1C). We find this strategy more accurate than deficiency mapping due to higher fidelity in scoring and not depending on deficiency coverage.

**RRs Vary Greatly in Small Genomic Intervals.** The accuracy of the mapping of a point mutation site depends on the RR. In other words, the PMP will be accurate only if the RRs in the interval between the mutation to be mapped and the two *P* insertions are identical. Unfortunately, very little is known about the variation of RRs in regions of 500 kb or fewer. We have mapped the lethality of a total of 15 complementation groups on chromosome arms 2L, 2R, and 3R to a 250- to 1,600-kb interval. We then carried out fine mapping (Fig. 2). Each diagram represents a mapping experiment, where PMPs obtained from each pair of flanking *P* insertions are indicated. We then used this data set of 15 genomic regions to survey the variability of RRs. Our analysis shows highly variable RRs (~2-fold difference within the region) in ~50% of the cases (Fig. 2B, F–J, N, and O). For example, Fig. 2H and I show ~6.6- and 12-fold differences in RRs in proximity of the mapped gene. Highly variable RRs in adjacent intervals have a strong effect on the accuracy of fine mapping. In contrast, when the RR is similar across the region spanned by the *P* insertions, the PMP becomes restricted to a smaller interval (PMP interval), as observed for groups A and I. In A, the RRs over a 493-kb region are very similar; therefore, all PMPs using different pairs of *P* insertions are within a 4-kb interval. In I, all PMPs are within 26 kb of an 876-kb region. In this example, the variation of RRs was never observed to be >2.08-fold. As the RR variability increases, so does the PMP interval. In our survey, we analyzed the RR variation of segments between pairs of neighboring *P* insertions in 15 fine mapping regions (eight examples are shown in Table 1). We found that, on average, the RR of a segment between a pair of *P* insertions varies by 34.7% when compared with the mean RR of the whole interval (n = 47). This variability did not show significant differences for varying lengths of segments or cytological locations. The variation in recombination is clearly due to the presence of recombinational “hot spots” and “cold spots” throughout the genome and is inherent to all methods based on meiotic recombination. If only the flanking *P* insertions of a mutation are used, however, the differences in RRs compared with surrounding areas become less relevant, and the distortion becomes much smaller the closer the flanking *P* elements are.

When RRs vary greatly and the PMP intervals are large, several strategies can be pursued. First, small deficiencies and mutant alleles of candidate genes often allowed us to narrow the region significantly or even identify the mutant gene (Fig. 2B, C, and E). Second, an additional round of fine mapping can be carried out with two or more *P* insertions around the PMP interval. This strategy has proven to be rather effective, and the density of *P* insertion is sufficient in most genomic regions (8). Finally, a high-resolution SNP map could be established to use recombinants from the *P* recombination mapping for SNP mapping.

To date, we have identified the molecular lesions in eight mutant genes of the 15 fine mapping examples shown in Fig. 2 (Table 1, examples A–H), either by failure to complement a mutation in a known gene (Table 1, examples B, C, and E) or by molecular mutation detection (Table 1, examples A, D, and F–H). Table 1 summarizes the mapping information of these genes, including the size of each gene, the distance between the gene and the PMP by using the closest *P* insertions, the length of the segment between the closest *P* insertions, and the recombination variation of the segment of the whole mapping region. The RRs over the region surrounding the mutation are relatively constant for A, F, and G, and, therefore, the PMP intervals are within 50 kb. The molecular lesions were identified by mutation detection (see below) and are point mutations located 10 kb (group A), <1 kb (group F), and 1.5 kb (group G) away from the PMP by using the closest-flanking *P* insertions, respectively. For groups C and E, the mutations were found to be outside the PMP interval. For group E, complementation tests with small deficiencies narrowed the region to the left of the PMP interval (data not shown). The candidate region of group C is close to the telomere (division 21), whereas RRs are very low. In this particular case, 1 cM corresponds to ~1.5 megabases. We therefore performed complementation tests with all of the available mutant alleles in the 150-kb region surrounding the PMP interval. This revealed that the mutation mapped 51 kb to the right of the PMP. We conclude that mapping positions with an accuracy of <50 kb can generally be achieved with one round of fine mapping in regions of low RR variation and two rounds of fine mapping in regions with high RR variation.
Fig. 2. Fine mapping results for 15 representative complementation groups. Panels represent mapping experiments, and cytological positions of each region are listed next to each panel label. The x axis is the distance between the P insertions in kilobases. The first P insertion site is normalized to 0, and the relative insertion sites of the other P elements are listed on the x axis. The number of flies counted is indicated above each P insertion (white-eyed flies/total). The red crosses mark the PMP by using pairs of P insertions, and their intervals (in kilobases) are indicated at the bottom. The y axis represents the RR (in cm/megabases) of each segment between two neighboring P insertions. Identified genes bearing the mutations are indicated by red bars on the x axis, and their relative genomic regions are listed in the red box above the gene in examples (A–H).
(14). Heteroduplex DNA therefore reaches its melting temperature (Tm) than its corresponding homoduplex DNA due to a base-pair mismatch. Heteroduplex DNA therefore reaches its Tm earlier in an increasing temperature gradient. This slight difference in time can be displayed by capillary electrophoresis, as shown in Fig. 3 A, C, E, and G. The optimal size of DNA fragments to be analyzed with this method is between 400 and 600 bp (12). To cover a 50-kb region with overlapping DNA fragments of this size, we developed a program for automatic generation of equidistant primer pairs that allow amplification of DNA by using the same conditions for all DNA fragments (see Materials and Methods). The PCR products are in 96-well plates and are loaded simultaneously on a SpectruMedix Reveal TGCE apparatus. The TGCE results are compared among alleles and against the homozygous unmutagenized chromosome as a control. A difference in the trace of only one allele, but not the control, indicates a unique mutation, which is subsequently sequenced. SpectruMedix Reveal TGCE traces of two example alleles and their corresponding point mutations in sequences are shown in Fig. 3.

**Discussion**

Here, we illustrate a high-resolution mapping approach based on classical meiotic recombination by using a unique set of 6,500 P insertions as molecularly defined markers. The mapping strategy presented here uses the same principle as classic recombination mapping methods with marked chromosomes but offers much higher resolution and greater accuracy due to the large collection of molecularly defined P element insertions and the availability of the sequence of the *Drosophila* genome. Our high-resolution mapping experiments demonstrate that recombination and physical maps of molecularly characterized regions can be difficult to align in regions spanning <500 kb. Although these distortions often preclude a theoretical resolution of <1 kb, a resolution of <50 kb can often be achieved, especially when an additional round of fine mapping is pursued. When compared with other mapping methods, we find that this method is very powerful, cheap, and quick.

**Comparison of Available High-Resolution Mapping Strategies: SNPs and P Insertions as Molecular Markers.** Compared with SNP mapping, the P insertion mapping strategy offers several advantages. For SNP mapping, a low-density SNP map has to be generated or confirmed for each isogenized chromosome that is mutagenized. Subsequently, for high-resolution SNP mapping, individual maps have to be generated for the region where the mutation is located. The P insertion mapping strategy, however, uses publicly available P element insertion lines that can be used in any white− genetic background. Second, the scoring of recombination events in SNP mapping requires molecular analysis, e.g., by sequencing or restriction enzyme digestion, to identify restriction fragment-length polymorphisms associated with SNPs. In contrast, recombination events in P insertion mapping are easily recognized by eye color. A fundamental difference in established SNP mapping techniques and P insertion mapping as presented in this article is the use of recombination information. If sufficient numbers of SNPs are identified (for example, 1 SNP per ~10 kb), mapping to either side of individual SNPs (similar to male recombination) can provide the resolution necessary to map to a single gene. If the SNPs are far apart, the binary result (either left or right) will not provide a high enough resolution to map to a single gene. In contrast, our strategy uses the recombination ratios to calculate an actual distance between the mutation and flanking P insertions. Furthermore, P insertion mapping is an extremely low-cost technique, because no molecular reagents are required; it can therefore be performed with the most basic laboratory setup. In addition, one round of rough mapping and one round of fine mapping can be accomplished in 8–10 wk. Finally, P insertion mapping is a highly flexible approach and can be combined with any of the other mapping methods, depending on expertise and reagents. In summary, the versatility, accuracy, low cost, and high speed of the P insertion mapping strategy make it highly efficient and effective.

**Table 1. Summary of mapping results for eight genes**

<table>
<thead>
<tr>
<th>Gene label (as in Fig. 2)</th>
<th>Gene size, kb</th>
<th>Length of gene-bearing segment, kb</th>
<th>Distance between gene and PMP, * kb</th>
<th>Recombination variation, † %</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>97</td>
<td>10</td>
<td>6.92</td>
</tr>
<tr>
<td>B</td>
<td>85</td>
<td>172</td>
<td>0</td>
<td>10.12</td>
</tr>
<tr>
<td>C</td>
<td>17</td>
<td>182</td>
<td>51</td>
<td>20.91</td>
</tr>
<tr>
<td>D</td>
<td>4</td>
<td>254</td>
<td>9</td>
<td>2.81</td>
</tr>
<tr>
<td>E</td>
<td>6</td>
<td>362</td>
<td>42</td>
<td>30.25</td>
</tr>
<tr>
<td>F</td>
<td>7</td>
<td>148</td>
<td>0</td>
<td>41.11</td>
</tr>
<tr>
<td>G</td>
<td>18</td>
<td>83</td>
<td>0</td>
<td>8.14</td>
</tr>
<tr>
<td>H</td>
<td>4</td>
<td>338</td>
<td>23</td>
<td>9.16</td>
</tr>
</tbody>
</table>

*PMP is obtained with the two closest-flanking P insertions.

†Recombination variation is calculated by dividing the difference between the PMP interval calculated from different pairs of P elements by the PMP interval calculated from the same pair of P elements. Complementation tests with small deficiencies or in combination to further delimit the region or uncover the molecular lesion. Complementation tests with small deficiencies or alleles of candidate genes are usually the first tests to be performed.

When a 50-kb or smaller region is defined, we use TGCE to detect point mutations. This system is based on the principle that the heteroduplex DNA has a slightly lower melting temperature (Tm) than its corresponding homoduplex DNA due to a base-pair mismatch (14). Heteroduplex DNA therefore reaches its Tm earlier in an increasing temperature gradient. This slight difference in time can be displayed by capillary electrophoresis, as shown in Fig. 3 A, C, E, and G. The optimal size of DNA fragments to be analyzed with this method is between 400 and 600 bp (12). To cover a 50-kb region with overlapping DNA fragments of this size, we developed a program for automatic generation of equidistant primer pairs that allow amplification of DNA by using the same conditions for all DNA fragments (see Materials and Methods). The PCR products are in 96-well plates and are loaded simultaneously on a SpectruMedix Reveal TGCE apparatus. The TGCE results are compared among alleles and against the homozygous unmutagenized chromosome as a control. A difference in the trace of only one allele, but not the control, indicates a unique mutation, which is subsequently sequenced. SpectruMedix Reveal TGCE traces of two example alleles and their corresponding point mutations in sequences are shown in Fig. 3.

**Discussion**

Here, we illustrate a high-resolution mapping approach based on classical meiotic recombination by using a unique set of 6,500 P insertions as molecularly defined markers. The mapping strategy presented here uses the same principle as classic recombination mapping methods with marked chromosomes but offers much higher resolution and greater accuracy due to the large collection of molecularly defined P element insertions and the availability of the sequence of the *Drosophila* genome. Our high-resolution mapping experiments demonstrate that recombination and physical maps of molecularly characterized regions can be difficult to align in regions spanning <500 kb. Although these distortions often preclude a theoretical resolution of <1 kb, a resolution of <50 kb can often be achieved, especially when an additional round of fine mapping is pursued. When compared with other mapping methods, we find that this method is very powerful, cheap, and quick.

**Comparison of Available High-Resolution Mapping Strategies: SNPs and P Insertions as Molecular Markers.** Compared with SNP mapping, the P insertion mapping strategy offers several advantages. For SNP mapping, a low-density SNP map has to be generated or confirmed for each isogenized chromosome that is mutagenized. Subsequently, for high-resolution SNP mapping, individual maps have to be generated for the region where the mutation is located. The P insertion mapping strategy, however, uses publicly available P element insertion lines that can be used in any white− genetic background. Second, the scoring of recombination events in SNP mapping requires molecular analysis, e.g., by sequencing or restriction enzyme digestion, to identify restriction fragment-length polymorphisms associated with SNPs. In contrast, recombination events in P insertion mapping are easily recognized by eye color. A fundamental difference in established SNP mapping techniques and P insertion mapping as presented in this article is the use of recombination information. If sufficient numbers of SNPs are identified (for example, 1 SNP per ~10 kb), mapping to either side of individual SNPs (similar to male recombination) can provide the resolution necessary to map to a single gene. If the SNPs are far apart, the binary result (either left or right) will not provide a high enough resolution to map to a single gene. In contrast, our strategy uses the recombination ratios to calculate an actual distance between the mutation and flanking P insertions. Furthermore, P insertion mapping is an extremely low-cost technique, because no molecular reagents are required; it can therefore be performed with the most basic laboratory setup. In addition, one round of rough mapping and one round of fine mapping can be accomplished in 8–10 wk. Finally, P insertion mapping is a highly flexible approach and can be combined with any of the other mapping methods, depending on expertise and reagents. In summary, the versatility, accuracy, low cost, and high speed of the P insertion mapping strategy make it highly efficient and effective.
The Impact of Recombination Variation on the Accuracy of P Insertion Mapping. In theory, meiotic recombination mapping should allow mapping to a single nucleotide. Indeed, in a few early studies, different alleles of the same locus were ordered relative to each other on the basis of recombinational mapping without the knowledge of the molecular nature of the affected genes (15–17). Likewise, high accuracy can be achieved through recombinational mapping by using molecularly mapped P insertions. The primary obstacle is the translation of this recombinational map in cM into the physical map in base pairs, which primarily depends on the colinearity of the two maps over short distances.

In our survey of 15 sample regions, we have found considerable variations in RRs in some small genomic intervals. Although these variations are likely to be primarily due to recombination hot spots and cold spots in the genome, it is also possible that P element insertions affect the recombination frequency in their vicinity. The RRs we observed may thus differ from those between wild-type and mutant chromosomes. When highly variable RRs are observed, a second round of fine mapping with between wild-type and mutant chromosomes. When highly variable RRs are observed, a second round of fine mapping with P insertions around the PMP interval is, in our hands, more efficient and less expensive than other methods.

From Fine Mapping to Mutation Detection. In an effort to establish a standard method applicable in a high-throughput manner for 50-kb regions, we used a new mutation detection system based on TGCE. To screen a 50-kb genomic interval, the region has to be subdivided by using tiled primer pairs. We implemented a web-based system that automatically designs primer pairs covering large intervals of DNA. This system proved very reliable and is publicly available at http://flypush igen.bcm.tmc.edu/ primer. The actual mutation detection on a SpectruMedix Reveal TGCE system proved suitable for a large-scale approach in both handling and data evaluation. Although we did find point mutations with this system so far in four cases, we cannot yet quantify the accuracy of the system. One drawback is the rather large number of false positives (the heterozygote mutant fragment showed a difference in Tm from the control homozygote fragment, yet it failed to reveal a point mutation by sequencing). False positives are an inherent problem of TGCE that can, for example, be caused by repetitive DNA sequences or the high sensitivity of the system. Complementary techniques like single-stranded nucleic digestion should improve the ratio of correct to false calls significantly. However, the number of candidate fragments that have to be sequenced is reduced by more than an order of magnitude when TGCE as an intermediate step, resulting in significant savings. Obviously, the larger the number of available alleles in a complementation group, the higher the likelihood of detecting the point mutations. When TGCE is not available, sequencing of candidate genes is an option.

In summary, we believe that the mapping approach presented in this study provides very valuable tools to map chemically induced mutations with minimum resources and is applicable in a high-throughput manner.

We thank Kathy Matthews, Kevin Cook, Koen Venken, and Sunil Mehta for critical reading of this manuscript and helpful discussions. We thank Ulrich Tepass for helpful discussions of experimental procedures. We further thank Mengfei Huang, Adam Xu, and Sven Vilain for assistance in the mapping experiments. Stocks with hs-hid balancers were a gift from Ruth Lehmann (New York University, New York). R.G.Z., P.R.H., K.L.S., H.J.-N., and K.K.N. are supported by the Howard Hughes Medical Institute. P.R.H. was further supported by a European Molecular Biology Organization postdoctoral fellowship; K.K.N., H.P., and H.J.B. are supported by the Pediatric Tumor Foundation; and M.P.G. is supported by a National Research Service Award. This work was further supported by the National Institutes of Health. H.J.B. is an investigator of the Howard Hughes Medical Institute.