Figures and figure supplements

Resolving the prevalence of somatic transposition in Drosophila

Christoph D Treiber and Scott Waddell
Figure 1. FACS strategy to extract αβ-Kenyon cells (αβ-KCs) from fly brains. (a) Posterior projection view of a confocal microscope stack showing αβ-KC somata (arrow heads) in a female Mb008b-GAL4; UAS-mCherry fly brain. The general neuropil is stained with the anti-bruchpilot antibody nc82 (white) and red indicates αβ-KC mCherry expression. (b) Anterior projection from the same brain as in (a) showing axons of the mushroom body lobes of αβ-KCs in either brain hemisphere, which form distinct bilaterally symmetrical L-shaped projections into vertical and horizontal lobes. This lobe structure will be used in all schematic representations of the brain in the rest of this manuscript to indicate the source of each gDNA sample. Scale bar 100 μm. (c) Example plot from a FACS run of a single fly brain to illustrate the selection strategy used to sort mCherry-positive αβ-KCs from mCherry-negative cells in dissociated brain tissue. Sorting gates were hand-drawn, with the aim of selecting a pure proportion of mCherry αβ-KCs (right inset shows a simplified fly brain highlighting αβ-KCs) and a second population of similar size, from the mCherry negative cells in the rest of the brain (left inset). Single cells are represented as red points and areas of high-density are colored blue. The same gates were used for all samples in this study. (d) mCherry and FasII expression is elevated in mCherry positive cell fractions. Graph shows relative expression levels in FAC-sorted αβ-KCs as compared to that in unlabeled cells. Error bars denote standard error of the mean (SEM).

DOI: 10.7554/eLife.28297.002
Figure 2. Single fly αβ-KC sequencing suggests transposon hoovering. (a) Schematic of the experimental approach. Six individual flies were processed independently. The circular plot shows the gDNA sequencing coverage of mCherry positive αβ-KCs (red trace) and mCherry negative cells from the rest of the brain (blue trace), on chromosome 2R from one representative individual fly. The schematic (top right) depicts the 4 fruit fly chromosome pairs. Chromosome 2R, which is the source data in the circular plot, is highlighted in black. Schematic fly brain (bottom right) indicates the color scheme; αβ-KCs (red), the rest of the brain (blue). Sequencing read alignments on other regions of the gDNA exhibited a similar coverage (data not shown). (b) Plot of a representative example of a germline transposon insertion that was found on chromosome 2L in each of the 6 individual flies and that is absent in the Drosophila melanogaster reference genome (Release 5.57). Putative new insertions were found at loci, which were approximately 10 kb up- and downstream of the germline insertion site of the same transposon type. Dark red diamonds represent the germline insertion of the transposon Doc, which was found in each of the 12 samples (αβ-KCs and the rest of the brain), and light red diamonds represent putative somatic Doc insertions. The genomic location of the turtle gene is shown below in blue. Boxes indicate exons and lines intronic regions of the gene. Schematic fly brain represents the color code used for this panel.

Figure 2 continued on next page
Figure 2—figure supplement 1. Total number of putative non-reference somatic TE insertions from 12 samples with 1, 2, 3–9 and more than 10 diagnostic reads.
DOI: 10.7554/eLife.28297.004
Figure 3. Transposon copy number and putative insertion rates do not correlate with age or transposon expression levels. (a) The number of putative somatic insertion events does not differ between young (3–4 days) and old (30 days) flies (Mann-Whitney test, p=0.2). Error bars denote SEM. (b) Figure 3 continued on next page.
Figure 3 continued

Heatmap showing the normalized number of sequencing reads that map onto each of the 111 reference transposon sequences that were analyzed in this study. None of the few visible differences in the amount of transposon sequences in the gDNA from αβ-KCs when compared to the rest of the brain of the same individual are statistically significant. Individuals #1 - #3 are young flies (3–4 days) and #4 - #6 are old flies (30 days). FPKM stands for fragments per kilobase of transposon sequence per million fragments mapped. (c) Boxplot showing the normalized number of reads that map onto each of the 111 reference transposon sequences per αβ-KC sample of young (3–4 days) and old (30 days) flies. Whiskers represent Min and Max and the box represents the first and third interquartile interval. No statistical difference was evident (Mann-Whitney test, p=0.9184). (d) Plot showing no linear correlation between the expression levels of 5 different transposons in αβ-KCs and the number of putative new insertions of each transposon identified in these cells. gDNA data was acquired from 6 independent biological replicates. Error bars denote SEM. (e) Plot showing the logarithmic number of reads that map to each transposon consensus sequence taken from the Drosophila genome on the x-axis, and the logarithmic average number of putative insertions in 6 flies. Each point represents one transposon type. The line depicts the linear regression ($R^2 = 0.7166$) and the 95% confidence interval.

DOI: 10.7554/eLife.28297.006
Immobile genetic elements appear to mobilize. (a) Representative example of a germline IGE insertion event that was found in each of the 6 individual flies and that is absent in the simulated Drosophila melanogaster reference genome (DMsim). Putative somatic insertion events of the same
IGE occurred at loci which are approximately 10 kb up- and downstream of the original germline insertion site. Dark purple diamonds represent the original germline insertion site, which, as expected, was present in each of the 12 samples (αβ-KCs and the rest of the brain). Light purple diamonds represent putative somatic IGE insertions in each of the samples. (b) Plot showing the number of putative new insertions in WGS data from oocytes (Khurana et al., 2011). Samples were normalized to a depth of 18.3-fold (as in Khurana et al., 2011) and the bars represent the putative insertions that were not detected in the parental strains. In addition, the number of IGE ‘mobilizations’ is shown. Note that the number of false positive IGE insertions is highest in the sample obtained from 21 day old dysgenic ovaries, and decreases in the F2 generation. (c) Graph illustrating the penetrance of a small selection of simulated IGE insertions. The actual penetrance for each locus should be 1. However, due to variations in local sequencing coverage, the analysis pipeline assigns varying frequencies to each IGE insertion. The penetrance of each IGE insertion shown apparently increases with age. (d) The average number of putative de novo insertions present in the three samples from oocytes (Khurana et al., 2011) correlates with the theoretical number of sequencing reads that map onto each transposon sequence in the Drosophila reference genome. The number of reads was based on the sequencing coverage (18.3-fold) and the number of 76nt fragments that overlap with each transposon reference sequence. Note, for example, Roo, R1 and FW the three endogenous transposons that contribute most frequently to ‘insertions’ are also the most abundant elements in the reference genome.

DOI: 10.7554/eLife.28297.007
Figure 5. Evidence for chimera formation during DNA amplification. (a) The DNA fragment sizes of long, overlapping sequencing read pairs were assessed by merging each read pair. The length of these merged fragments varied and peaked at 270 bp. (b) In silico assembled read pairs which are all 250 basepairs apart map at genomic locations which are further apart than the predicted size. Plot shows the calculated fragment size which is based on the distance between each of the two mapped paired-end read. (c) Transposon: chromosome breakpoints occur across the entire length of the genome.
Figure 5 continued

transposons. Plotted are the number of putative transposon insertions in each fly tested and grouped into bins based on the relative position of the breakpoint along the length of each transposon. (d) Graph illustrating the frequency of each size of complementary sequence spanning the junction of chimeric amplicons (only those above 1 are shown). (e) Schematic depicting how chimeric DNA, which was formed during gDNA amplification, can result in read-pairs which lead TEMP to predict the presence of a rare somatic transposon insertion down- or upstream of a germline insertion. For TEMP, gDNA is extracted, fragmented, sequenced and paired-end reads are aligned to a reference genome. According to data presented in panel (b), during gDNA amplification chimera are favorably formed between sections of gDNA that are between ~1 and 10,000 basepairs apart. This clustering mirrors the range of the apparent transposon and IGE hovering as predicted by TEMP (see Figures 2b and 4a).

DOI: 10.7554/eLife.28297.008