CAMIO for deletion analysis of endogenous DNA sequences in multicellular organisms

Hui-Min Chen, Jorge Garcia Marques, Ken Sugino, Dingjun Wei, Rosa Linda Miyares and Tzumin Lee*

Howard Hughes Medical Institute, Janelia Research Campus, 19700 Helix Drive, Ashburn, VA 20147, USA

* To whom correspondence should be addressed.

Tel: 571-209-4198

Email: leet@janelia.hhmi.org

Abstract

The genome is the blueprint for an organism. Interrogating the genome, especially locating critical cis-regulatory elements, requires deletion analysis. This is conventionally performed using synthetic constructs, making it cumbersome and non-physiological. Thus, we created Cas9-mediated Arrayed Mutagenesis of Individual Offspring (CAMIO) to achieve high-throughput analysis of native DNA. CAMIO utilizes CRISPR that is spatially restricted to generate independent deletions. Controlled by recombination, a single guide RNA is stochastically chosen from a set targeting a specific DNA region. Combining two sets increases variability, leading to either indels at 1-2 target sites or inter-target deletions. Cas9 restriction to male germ cells elicits autonomous double-strand-break repair, consequently creating offspring with diverse mutations. Thus, from a single population cross, we can obtain a deletion matrix covering a large expanse of DNA at both coarse and fine resolution. We demonstrate the ease and power of CAMIO by mapping 5'UTR sequences crucial for *chinmo's* post-transcriptional regulation.

Introduction

It never ceases to amaze that genomes, sequences composed of a seemingly simple four-letter code, serve as the fundamental blueprints for life. From these

four nucleotides emanate layers upon layers of complexity. The central dogma depicts the basic information flow from DNA to RNA to protein. However, numerous mechanisms increase the complexity at each step and can feed backwards. Some examples include epigenetic modifications, promoter/enhancer usage, alternative splicing, post-transcriptional and post-translational modifications. Understanding how complex biology unfolds, advances, and evolves from genomic sequences requires development of precise and efficient tools to interrogate the genome. While spontaneous or induced mutations once served as the basis of our understanding about many biological processes, we now exploit sophisticated methods to directly manipulate genomes for systematic mechanistic studies.

Pioneering geneticists, like Dr. Barbara McClintock, enlightened us with fundamental understanding of genetic information and regulation at the level of genes and chromosomes. More recently, to understand gene regulation, molecular biologists have relied heavily on *in vitro* or cell culture-based assays to assess the function of DNA fragments. One example of this is promoter bashing used to pinpoint critical or regulatory sequence elements in a promoter[1,2]. Also, in enhancer screening, larger DNA fragments upstream or downstream of a target gene are fused to reporter genes[3-5]. This synthetic approach often fails to recapitulate full endogenous patterns, and thus requires second-round rescue experiments to draw solid conclusions. Such indirect methods are inefficient and

63 non-physiological, prompting the search for tailor-made sequence-specific DNA
64 nucleases that permit targeted mutagenesis of native sequences.
65
66 In the early 2000s, zinc-finger nucleases (ZFNs)[6], the first of the tailored
67 nucleases, were heavily adopted[7,8]. Although it seemed a promising strategy for
68 genome editing, major setbacks, especially off-target toxicity limited its utility[9].
69 Therefore, the enthusiasm for ZFNs declined after the arrival of transcription
70 activator-like effector nucleases (TALENs)[10-12]. The higher specificity of TALENs
71 led to fewer off-target disruptions, and hence less cytotoxicity[13]. However,
72 constructing a TALEN is technically challenging because the homologous
73 sequences encoding the TALEN repeats are prone to recombine with each
74 other[9].
75
76 The newest technology, CRISPR (clustered regularly interspaced short
77 palindromic repeats), was first exploited for targeted mutagenesis[14] and then
78 swiftly adopted to edit genomes of diverse organisms[15-20]. CRISPR's popularity
79 lies in its simplicity: a Cas9 nuclease plus an easily made guide RNA (gRNA)
80 induces a double-strand break (DSB) targeted by DNA-RNA base pairing[14,16].
81 Cells repair the DSBs either through nonhomologous end joining (NHEJ)[21],
82 leading to insertion or deletion (indel), or through homology-directed repair
83 (HDR)[22] which replaces discontinuous sequences with an available template.
84 The fact that CRISPR can produce targeted DNA modifications with near
85 unlimited specificity has ensured its rapid expansion throughout the biological

and biomedical fields[23]. Indeed, CRISPR technology has already transformed studies from stem cell research and cancer biology to food production and pest control[24].

Drosophila melanogaster has been a powerful model organism for decoding the genome to understand complex biology[25]. However, even with existing genetic tools, it remains quite challenging to interrogate the entire fly genome, especially non-coding regions. Enhancing fly genetics with CRISPR is particularly needed for large-scale genome-wide screens as well as focused, detailed sequence analyses[26].

Here we describe a new high-throughput technology for deletion analysis called CAMIO (Cas9-mediated Arrayed Mutagenesis of Individual Offspring). We built a CRISPR-based mutagenesis pipeline in *Drosophila* male germ cells, to achieve massive production of independent indels in targeted loci with germline transmission[27,28]. We further created a transgenic system for simultaneously targeting multiple sites with an array of gRNAs. This way, we can readily generate a huge collection of organisms harboring either diverse, small, localized indels or large, defined deletions (inter-target deletions). This will enable efficient deletion analysis of sizable genomic regions *in vivo*. We include here an example that demonstrates the power of CAMIO. In 2006 our lab discovered that the expression of temporal protein Chinmo was regulated via its 5'UTR[29]. Using CAMIO, we are able to rapidly ascribe a critical aspect of this temporal control to

a 300-bp sequence of the 2kb UTR. In conclusion, CAMIO enables high-throughput organism-level genome structure-functional studies.

Results

Independent targeted mutagenesis in individual male germ cells

To mutate the genome in a high-throughput manner requires a germline pipeline for independent mutagenesis in individual germ cells rather than germline stem cells (GSCs). We have shown that the *bam* promoter can effectively and specifically drive flippase induction in female germ cells, and not in GSCs[28]. *bam*'s restriction to germ cells in both the male and female germline prompted us to explore whether we could conduct independent mutagenesis in individual germ cells.

To this end, we made *bamP(898)-Cas9* and tested its ability to mutate gRNA-targeted sites in male as well as female germline. For proof of principle, we chose the *ebony* gene for targeted mutagenesis. Loss-of-*ebony* mutations are easy to detect and there was a readily available transgenic gRNA targeting *ebony* (Fig. 1a)[19]. We determined the mutagenesis efficiency in male versus female founders (Fig. 1b). Surprisingly, over 25% of male gametes as opposed to only ~3% of female gametes carried loss-of-function *ebony* mutations. This result

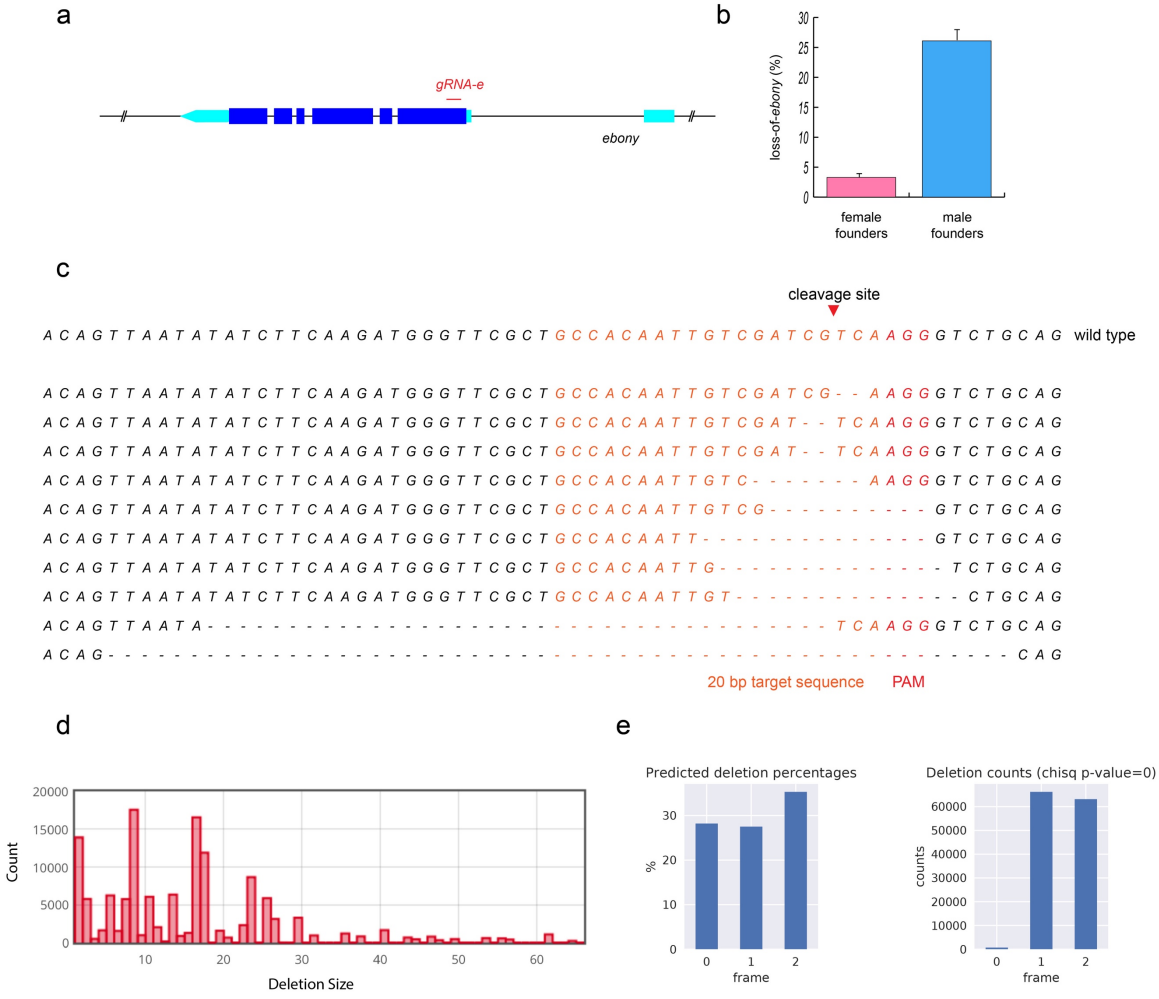


Figure 1. *bamP-Cas9* induces efficient CRISPR targeted mutagenesis in male germ cells. (A) Ebony transcript shown with UTRs in turquoise. U6 drives guide RNA (*gRNA-e*), which targets 5' end of *ebony* coding sequence. (B) Percent of progeny with ebony loss-of-function (LOF) mutations from female or male founders. Mean ± SEM (n=9). (C) Sequences of 10 randomly selected *ebony* LOF progeny from the same male founder. (D) Deletion profile of *UAS-shibire* transgene following CRISPR targeting collected from 919 phenotypically wildtype progeny. NGS data was analyzed and presented by Cas-analyzer, www.rgenome.net[30]. (E) Left, percentage of predicted indels (calculated by FORECasT[31] using *gRNA-shi* and *UAS-shibire* sequences) grouped into 3 reading frames, 0 represents the in-frame indels. Right) Actual reading frame percentages from phenotypically wildtype progeny, calculated from mapping results (obtained with CAS-analyzer). Chi-square test assuming equal distribution was used to assess the significance. The result is below machine precision and thus set to zero.

147 demonstrates that the male germline is particularly susceptible to *bamP-Cas9*-
148 mediated genome editing.
149
150 To address if the Cas9-mediated editing events occurred independently, we
151 sequenced a part of the *ebony* locus in individual mutants carrying *ebony*
152 deficiency. From a single male founder, we analyzed ten progeny with *ebony*
153 loss-of-function phenotypes. We uncovered seven different indels around the
154 Cas9 cut site (Fig. 1c). One identical *ebony* mutation occurred in three siblings,
155 possibly resulting from either differential deletion of GTC repeats or from
156 microhomology-mediated repair[32]. The recovery of many distinct indels from a
157 single founder argues for independent Cas9 actions in individual germ cells. This
158 result encouraged us to establish *bamP* induced CRISPR in the male germline
159 as a 'targeted mutagenesis pipeline'.
160
161 A mutagenesis pipeline could be useful for producing novel alleles of a protein of
162 interest. To explore this idea, we tested CRISPR's usefulness to produce novel
163 alleles of a well characterized gene, *shibire. shibire* encodes *Drosophila*
164 Dynamin, a motor protein crucial for synaptic vesicle endocytosis[33]. shi^{ts1}, a
165 temperature sensitive allele containing a missense mutation, is widely used in
166 behavioral assays to temporarily shut off neuronal activity[34]. We therefore
167 designed a gRNA against the *UAS-shibire* transgene at the region where the
168 temperature-sensitive shi^{ts1} point mutation is located, presuming that we could
169 produce additional temperature-sensitive or dominant negative alleles.

We tested our mutagenized transgene by expressing it in the eye with GMR-Gal4 and screening at 29 degrees. Unfortunately, the rough eye phenotype was not confined to temperature-sensitive or dominant negative alleles, but was also the result of high transgene expression in the eye. Frameshift mutations in the beginning of *shibire* would lead to premature stop codons, and the resultant small truncated proteins are likely non-functional. By contrast, in-frame mutations would create essentially full-size proteins. However, loss of critical amino acids could disrupt key catalytic functions but preserve the protein's ability to polymerize, thus creating a dominant negative allele. We therefore surveyed phenotypically wildtype offspring to see if they lacked in-frame mutations. We pooled around 1000 phenotypically normal progeny collected from 20 male founders for amplicon analysis with next generation sequencing (NGS). We obtained a large collection of diverse indels, with the majority of deletions smaller than 30 bps (Fig. 1d). Notably, there is a clear under-representation of in-frame mutations (Fig. 1e). The selective loss of such in-frame mutations is noteworthy, and supports the feasibility of making 'novel' useful proteins via deleting various amino acids of interest in a high-throughput manner.

Taken together, our data demonstrate that *bamP(898)* effectively restricts Cas9-mediated mutagenesis to germ cells. There is no evidence that clonal expansion contributes to the exceptionally high mutation efficiency in male founders.

Therefore, transgenic CRISPR, induced by *bamP*, can effectively serve as a pipeline for mass production of targeted mutations.

CAMIO: Cas9-mediated arrayed mutagenesis of individual offspring

Despite independent mutagenesis in each germ cell, using only one gRNA limits the offspring variation, as all indels are anchored around the same Cas9 cut site. To expand the diversity of deletions that one can recover from a single population cross, we next explored the possibility of multiplexing gRNA-targeted mutagenesis. Our vision for multiplexing gRNAs is to have a collection of gRNAs from which one is stochastically selected, rather than simultaneously expressing multiple gRNAs. Incorporating this multiplexed design into the male germline in combination with *bamP(898)*-Cas9 would enable both stochastically chosen gRNAs and offspring independent mutations. Supplying one gRNA at a time prevents contamination of rare deletions by much more frequent second-site mutations. This way, discrete clusters of simple deletions can be recovered from a single population cross. Therefore, we can tile a sizable DNA region with diverse small deletions with a repertoire of evenly spaced gRNAs.

To examine the feasibility of our multiplexing design, we targeted a *UAS-mCD8::GFP* transgene with four independent gRNAs. To stochastically activate only one out of the four gRNAs, we made a conditional U6-gRNA(x4) transgene that is dependent on PhiC31-mediated recombination (Fig 2a). Using the *nos*

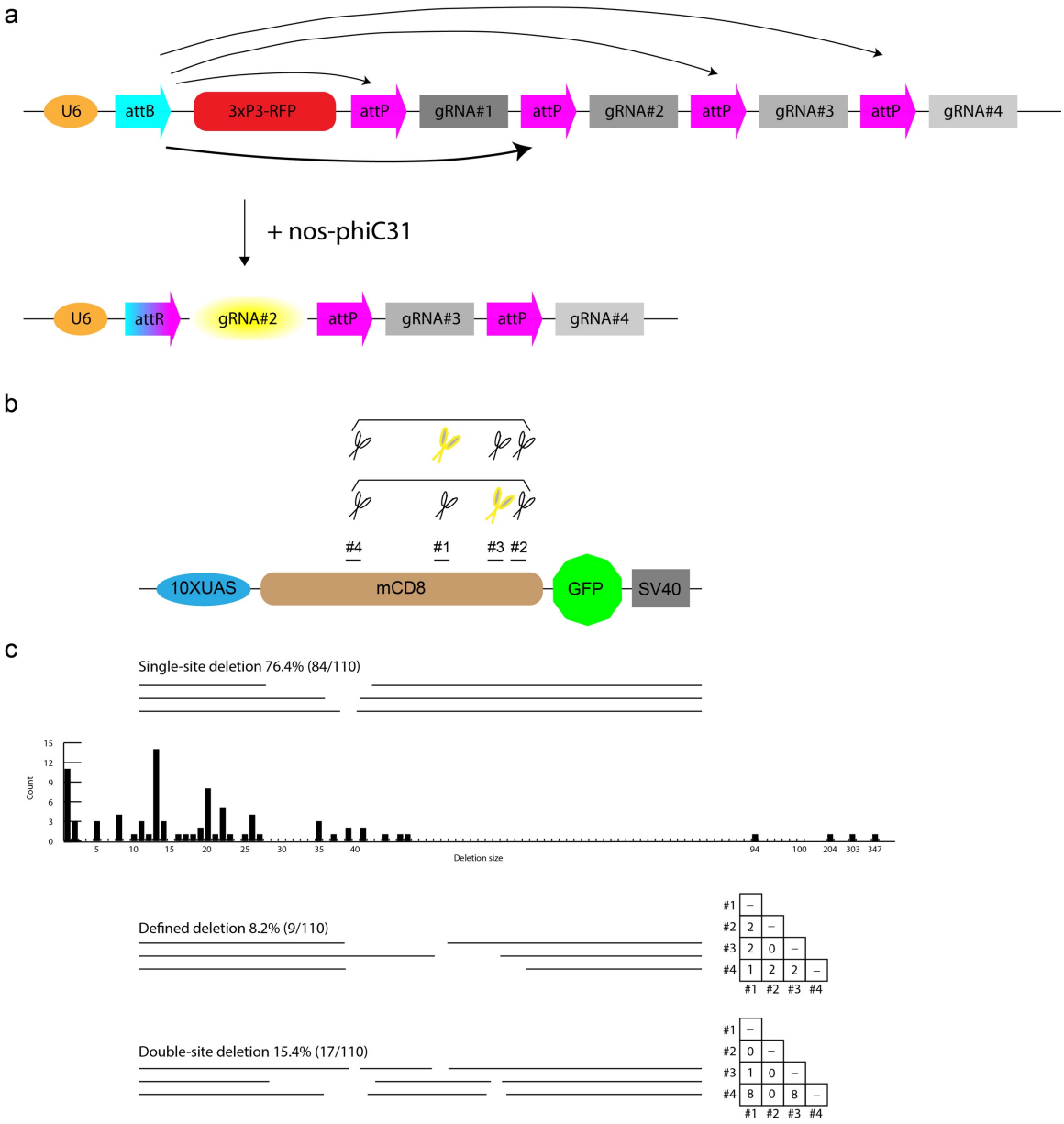


Figure 2. CAMIO produces diverse arrayed mutations around selected gRNA target sites. (A) Schematic of a conditional U6 gRNA transgene: pCis-{4gRNAs_mCD8}. The U6 promoter is separated from the gRNAs by a large fragment containing a 3xP3-RFP marker. With phiC31 recombination, the attachment site, attB can recombine stochastically with any of the four attP sites (black arrows), creating attR. The phiC31 recombinase is expressed in GSCs, controlled by *nos.* After recombination, a single gRNA is under the control of the U6 promoter (yellow oval) (B) 4 gRNA target sites along the mCD8 coding sequence were selected for pCis-{4gRNAs_mCD8} to disrupt mCD8::GFP expression. Simple or combined multiplexing mutagenesis can be achieved by incorporating one or more pCis-{4gRNAs_mCD8} transgenes. Here, we show two transgenes, each represented by four pair of scissors. Stochastically chosen gRNAs in each set are further marked in yellow. (C) Three categories of deletions arose from applying two copies of pCis-{4gRNAs_mCD8}: single-site deletion, defined deletion, and double-site deletion. Deletion size and count are depicted for single-site deletions. Right, use matrix of gRNA choices, deduced from sequencing GFP negative progeny.

231 promoter, we control the induction of the phiC31 recombinase in GSCs. Thus, in
232 each of the 12-24 GSCs per male fly founder[35], the transgene is irreversibly
233 recombined to express a single gRNA. Recombination occurs between a single
234 attB site downstream of the U6 promoter and a choice of attP sites upstream of
235 each gRNA; once reconstituted, the ubiquitous U6 promoter drives expression of
236 only one of the gRNAs. The intra-chromosomal recombination excises an
237 intervening 3xP3-RFP. Given the rather small size of each gRNA as compared to
238 the large 3xP3-RFP, the differences in length between the attB site and the
239 choice of any one attP site is therefore relatively trivial. Based on a previous,
240 similar construct for multicolor imaging[36], we expect that each gRNA should be
241 expressed at comparable frequencies. For brevity, we name the conditional U6-
242 gRNA transgene pCis, and then, in braces, add the number of gRNAs and the
243 name of the targeted DNA. For example, to target the UAS-mCD8::GFP
244 transgene, we created pCis-{4gRNAs_mCD8}. Also, when we describe the
245 individual gRNAs, we number them in sequence from 5' to 3'.
246
247 For the multiplexed targeted mutagenesis of mCD8::GFP, we established male
248 founders carrying *UAS-mCD8::GFP*, *bamP(898)-Cas9*, *nos-phiC31*, and *pCis-*
249 *{4gRNAs_mCD8}*, and crossed them to *act5C-Gal4* females for easy scoring of
250 GFP fluorescence in the progeny. Overall, approximately 35% of the progeny lost
251 GFP expression. We collected 30 GFP-negative offspring from two founder
252 males. Sequencing the mCD8-coding region revealed that each GFP-negative
253 offspring carried an indel corresponding to a single gRNA (Supplementary Fig.

1). Encouragingly, we recovered various deletions resulting from activation of each of the four gRNAs. However, the frequency of mutations at each target site varied. Both founders yielded many more deletions around the gRNA#1/#4 targets than the gRNA#2/#3 targets, possibly reflecting their different on-target potencies.

We found that the majority (83.3%) of deletions removed 20 or fewer bp and that the largest one eliminated 85 bp. To tile a sizable DNA region with such small deletions would require many gRNAs bombarding the region of interest at a density of around one gRNA per 100 bp. Alternatively, we should be able to create larger deletions spanning two Cas9 cuts elicited by two gRNAs acting at a distance. To explore co-employment of two gRNAs, we provided two copies of *pCis-{4gRNAs_mCD8}* for multiplexed dual mutagenesis of *UAS-mCD8::GFP* (Fig. 2b). We obtained a comparable loss-of-GFP mutation rate at ~35% despite co-expressing two identical or distinct U6-gRNAs. This phenomenon implicates that Cas9 activity (either level or duration) limits the efficiency of gRNA-directed mutagenesis in germ cells. Nonetheless, we could recover various diverse mutations from the dual gRNA-derived GFP-negative progeny, including many single-site deletions (76.4%) and quite a few double-site deletions (two target sites with independent indels; 15.4%) as well as some large deletions spanning two gRNA target sites (inter-target deletions; 8.2%) (Fig. 2c). Notably, the single-site deletions greatly outnumbered those involving two sites. This outcome is

favored in large-scale deletion analysis, as it increases the chance of recovering deletions without second-site contamination.

The above results demonstrate that using dual gRNA sets enables us to tile a region of interest not only with indels, but also with defined deletions. Random selection of a single gRNA from each of the two identical sets which contain four gRNAs will yield six possible defined deletions. Encouragingly, from a collection of only nine inter-target deletions, we recovered five of the six anticipated defined deletions. Nevertheless, physical hindrance may prevent two Cas9 complexes from acting simultaneously on very close gRNA targets. This may explain why we failed to recover the smallest defined deletion of 37 bp between the Cas9 cut sites of gRNA#2 and #3 targets. These results suggest that inter-target deletions utilizing two gRNAs can support rapid systematic DNA deletion analysis.

In sum, we established an effective strategy to express various permutations of two gRNAs in male GSCs. In combination with restricting Cas9 to male germ cells, we built a germline pipeline for multiplex targeted mutagenesis. We dub this genetic system CAMIO (Cas9-mediated arrayed mutagenesis of individual offspring), which can derive from a single population cross a matrix of variable deletions. This strategy enables deletion analysis of a substantial DNA region with both coarse (inter-target deletions) and fine (a variety of single- or double-site deletions) resolution. Below, we prove the power of CAMIO in structure-functional analysis of a 2.2kb-long genomic fragment.

299

300 **Structure-functional analysis of *chinmo* 5'UTR**

301

302 The Chinmo BTB-zinc finger nuclear protein is dynamically expressed in intricate
303 spatiotemporal patterns in the developing *Drosophila* central nervous system.
304 Such dynamic Chinmo expression governs various aspects of temporally
305 patterned neurogenesis, including age-dependent neural stem cell
306 proliferation[37,38] and birth order-dependent neuronal cell fate[29,39,40]. Notably,
307 *chinmo* transcripts exist much more broadly than Chinmo proteins, indicating
308 involvement of negative post-transcriptional regulation[29,38]. Consistent with this
309 notion, *chinmo* transcripts have long UTRs, including a 2.2kb 5'UTR and an
310 8.5kb 3'UTR[41]. To locate the involved regulatory elements in such long UTRs by
311 conventional structure-functional analysis would be a daunting task.

312

313 Nonetheless, we started by making reporter transgenes carrying *chinmo* 5'
314 and/or 3' UTR(s). For a functional readout, we utilized the development of the
315 *Drosophila* mushroom body (MB), which involves an orderly production of γ, α'β',
316 and αβ neurons. We first determined the roles of the 5' vs. 3'UTR in
317 downregulation of Chinmo expression along MB neurogenesis[29], by examining
318 the change in expression of reporter transgenes (containing either or both UTRs)
319 from early to late larval stages (Supplementary Fig. 2). Notably, presence of the
320 *chinmo* 5'UTR drastically suppressed the reporter expression. Interestingly, only
321 in the absence of the 3'UTR did we detect an enhanced 5'UTR-dependent

322 suppression at the late larval stage. These phenomena ascribe the *chinmo*
323 downregulatory mechanism(s) to the 5'UTR, and unexpectedly revealed some
324 upregulation by the 3'UTR. This upregulation could potentially be a transgene-
325 specific artifact, arguing for the importance of performing assessments in the
326 native environment. We thus turned to CAMIO to carry out structure-functional
327 analysis of the native *chinmo* 5'UTR.
328
329 *chinmo*'s 5'UTR is separated into 3 exons; the first two exons are neighboring
330 and the distant 3rd exon is separated from the 2nd by 36kb (Fig. 3a). A gRNA set
331 was designed to target each exon for CRISPR mutagenesis (Fig. 3b). We
332 provided two copies of the same set for induction of both indels and inter-target
333 deletions within an exon. Additionally, we paired gRNA sets for exon 1 and 2 to
334 create larger deletions that span the exon1/2 junction. Thus, we could delete
335 various parts of *chinmo* 5'UTR in its endogenous locus with simple fly pushing.
336
337 Based on the previously observed deletion rate of around 35%, we collected 300
338 male progeny from each CAMIO genetic cross. We hoped to saturate each
339 5'UTR exon with ~100 different deletions. In total, 1200 CAMIO males were
340 harvested from the four different gRNA array combinations (exons 1, 2, 3, and
341 1+2). We mapped potential indels by sequencing indexed PCR products in a
342 high-throughput manner.
343

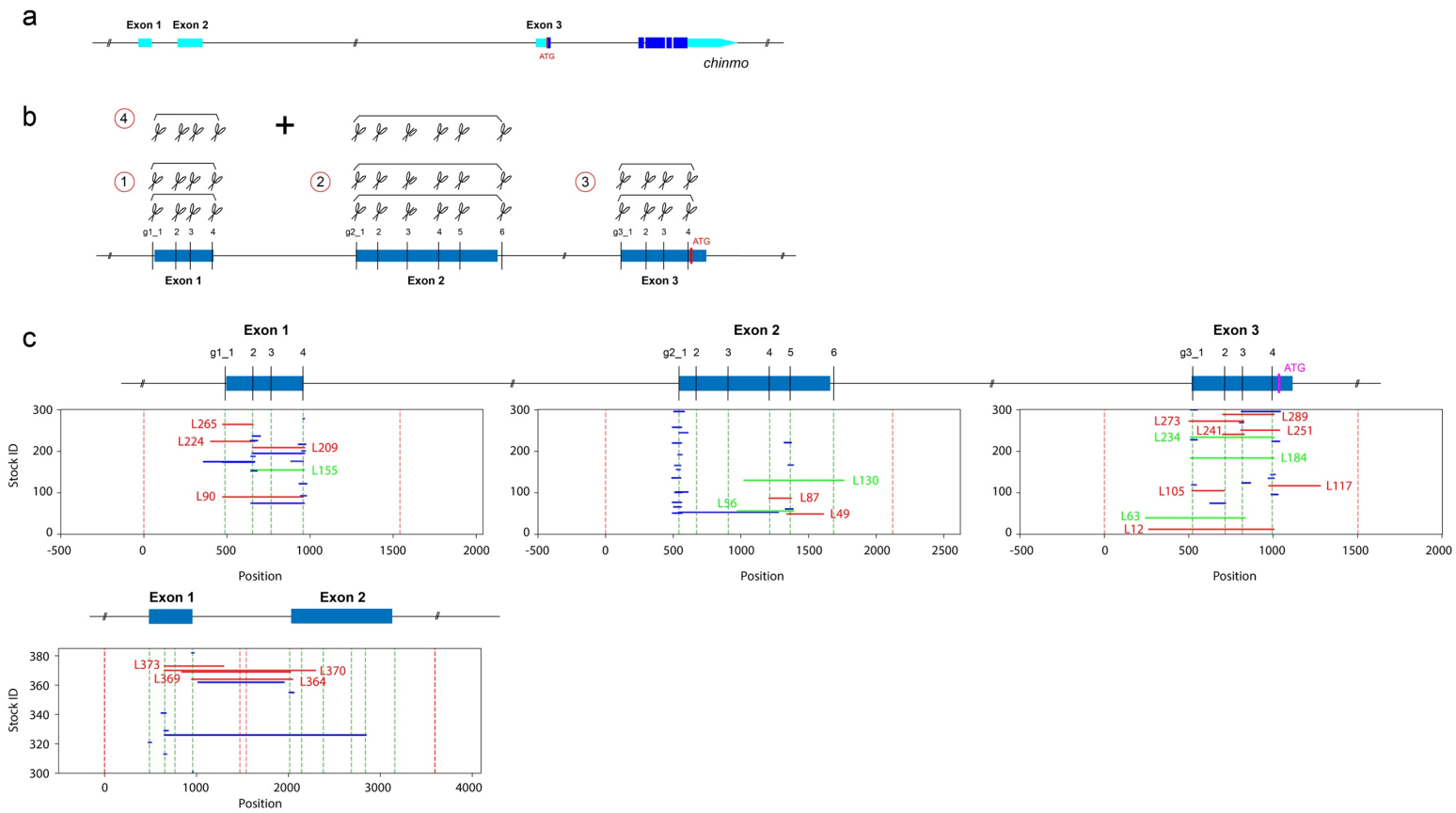


Figure 3. Applying CAMIO on *chinmo* 5' UTR. (A) An illustration of the *chinmo* gene, UTRs are depicted in turquoise. (B) We selected 4 target sites on *chinmo* exon 1 (455 bps), 6 target sites on exon 2 (1061 bps), and 4 target sites on exon 3 (639 bps). Each exon was dissected by CAMIO with a pair of gRNA sets, depicted as a set of scissors. Additionally, we combined gRNA sets from exon1 and exon2. (C) Representation of larger deletions predicted by NGS of 984 offspring (300 each for individual exons and 84 for Ex1-Ex2). Predicted deletions over 100 bps were subject to Sanger sequencing for confirmation. Confirmed >100 bps deletions are marked in red and given an ID#. Selected homozygous viable deletions (marked in green) that cover larger proportions of each exons were selected for MB development studies.

We detected numerous indels around each gRNA target site (Supplementary Fig. 3) and also recovered many inter-target deletions that together allow efficient coverage of the entire 5'UTR (Fig. 3c). We made organisms homozygous for the large inter-target deletions and examined MB morphology. Markedly, we found similar aberrant MB morphology with two exon 2 inter-target deletions, *chinmo*^{Ex2L56} and *chinmo*^{Ex2L130} (Fig. 4a). These deletions overlap by ~300 bp. Variable defects in the perpendicular projection of the bifurcated αβ axon lobes appeared at comparable frequencies (~30-50%) in homozygous as well as transheterozygous brains. Further, the penetrance of this phenotype is sensitive to Chinmo dosage, as a *chinmo* deficiency line effectively suppressed the phenotype (Fig. 4b). Marchetti and Tavosanis recently proposed that Chinmo downregulation plays a role in promoting α'β' to αβ MB neuron fate transition at the prepupal stage[42]. Therefore, we assessed Chinmo levels, and observed aberrantly elevated Chinmo in young MB neurons around pupa formation in both *chinmo*^{Ex2L56} and *chinmo*^{Ex2L130} homozygous mutants (Fig. 4c). This is consistent with the notion that these overlapping deletions have uncovered the essential region for this prepupal downregulation of Chinmo. In summary, a single round of CAMIO allowed us to identify a 300 bp locus in the 2.2kb 5'UTR critical for proper MB development.

Meanwhile, notably absent were indels or large deletions involving the target sites g2_2 and g2_3. This area in exon 2 may carry essential sequences for Chinmo regulation that is critical for organism viability. Alternative explanations

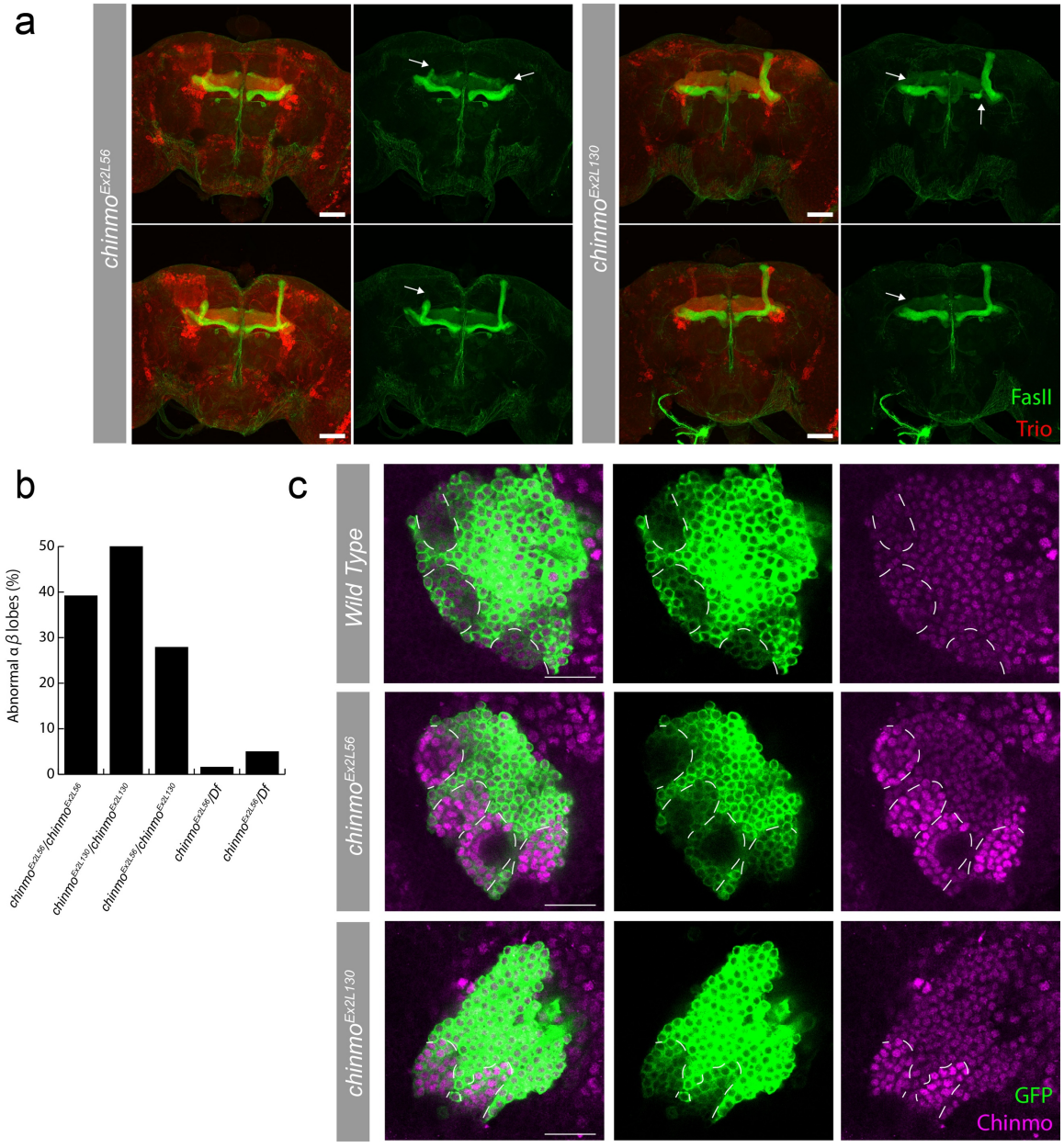


Figure 4. Overlapping deletions in 5'UTR alter Chinmo expression and MB morphology. (A) Stacked confocal images of adult MBs stained for FasII (green, αβ and γ lobes) and Trio (red, α'β' and γ lobes). Missing or misshapen αβ lobes (arrows). Scale bars: 50 μm. (B) Percent of flies with abnormal αβ lobes. Homozygous and transheterozygous deletions have strong MB αβ lobe defects, whereas hemizygous deletions over a *chinmo* deficiency line (*Df*) are much less penetrant. (C) Single optical sections of MB lineages (*OK107-Gal4*, *UAS-mCD8::GFP*) immunostained for GFP and Chinmo at white pupa stage. Chinmo staining is elevated in newly derived MB neurons (weaker GFP signal, outlined by white dashed lines) in homozygous *chinmo*^{Ex2L56} and *chinmo*^{Ex2L130}. Green: GFP; Magenta: Chinmo. Scale bars: 20 μm.

for the failure in recovering indels from that region include: a relatively shallow sequencing depth of the exon 2 region, our small sample size, unexpectedly low gRNA on-target strength for these gRNAs, or flawed design in the exon 2 gRNA set pCis-{6gRNAs_chinmo Exon2}. To address the last concern that bothered us most, we assessed the usage of specific gRNAs for the exon 2 set in progeny that do not contain Cas9 (Supplementary Fig. 4). While g2_2 and g2_3 were not recruited as frequently as others, they each still emerged 6-7% of the time. A rate of 6-7% should be sufficient for us to recover some indels, as gRNA#1 was selected ~15% of the time and produced multiple indels in our CAMIO experiment.

To examine whether this region of the 5'UTR is indeed critical, we exploited mosaic analysis to create somatic mutations in different tissues. A transgene, *dU6_g2+3*, was hence assembled to ubiquitously express both g2_2 and g2_3. We began with MB-specific CRISPR mutagenesis by inducing *UAS-Cas9* specifically in the MB lineage using a MB specific Gal4 (*OK107-Gal4*). We saw no temporal fate changes in the MB, the classic phenotype of *chinmo* misregulation in the MB. We next elicited CRISPR mutagenesis in all neural stem cells (neuroblast: NB) with NB-restricted Cas9 (*dpn-Cas9*) and *dU6_g2+3*. These animals were viable and showed no abnormal tumor-like NBs in larval or adult brains (typical of Chinmo overexpression)[38]. These data failed to provide evidence in support of presence of critical brain regulatory elements in the region targeted by g2_2 and g2_3. We strengthened this negative conclusion by

directly removing various small-to-large fragments around g2_2 and g2_3 targets from the above *chinmo* UTR-containing GFP transgene. In developing MBs, we observed indistinguishable GFP expression profiles between wild-type and modified 5'UTRs (Supplementary Fig. 5). In contrast to our negative findings in the brain, we found severe embryonic or early larval lethality when we induced early ubiquitous somatic mutations with *act5C-Cas9* and *dU6_g2+3*. This dominant lethality provides a direct explanation for our failure in recovering viable organisms carrying g2_2 or g2_3 induced indels. Together, these data suggest an essential role for *chinmo* 5'UTR outside of the brain.

In conclusion, a single round of CAMIO successfully led us to uncover two critical regions of the *chinmo* 5' UTR. The first critical region lies around the g2_2 and g2_3 targets and carries essential sequences for organism viability, unrelated to Chinmo's functions in the brain. The second critical region of the chinmo 5'UTR was uncovered by the overlapping $chinmo^{Ex2L56}$ and $chinmo^{Ex2L130}$ deletions. We determined that this ~300bp region is essential to down-regulate Chinmo expression, ensuring proper MB development. This fruitful case-study exemplifies the power of CAMIO in high-throughput unbiased deletion analysis of the genome.

Discussion

433 Two innovations synergistically enable CAMIO, a germline pipeline for arrayed
434 CRISPR mutagenesis. First, the *bamP* promoter can specifically limit Cas9
435 endonuclease activity to individual male germ cells—thus individual offspring
436 receive independent mutations. Second, the random-choice gRNA arrays provide
437 extensive coverage for deletion analysis, with both small indels and large
438 deletions. Hence, the combination of bamP-Cas9 and gRNA arrays used in
439 CAMIO enables *in vivo* targeted deletion analysis with both minimal molecular
440 biology and minimal fly pushing.
441
442 We were happily surprised to discover a much higher CRISPR mutagenesis rate
443 in male compared to female germ cells using *bamP*. This sex difference was also
444 observed in CRISPR-induced gene targeting in our effort to improve Golic+
445 (manuscript in preparation). Currently, we do not know what leads to this
446 phenomenon. *bamP* has a striking similar expression pattern in both the female
447 and male germline: absence in GSCs and an early onset of expression during
448 the four incomplete mitoses that produce the 16-cell germline cysts[43]. One
449 possibility is that the *bamP* activity is higher in the male than the female germline.
450 Another possibility has to do with the sex differences in meiotic recombination—
451 meiotic recombination does not occur in male *Drosophila*. Perhaps reduced
452 access to homologous chromosomes as templates for homology-mediated repair
453 favors indels. In any case, this feature allows us to utilize *bamP* to build a high-
454 efficiency pipeline for targeted CRISPR mutagenesis in male germ cells.
455

Conventional gRNA multiplexing provides all gRNAs at once as a cocktail, which expands indel diversity but inevitably creates complex and often biased deletion patterns. The off-target effects of a gRNA cocktail also accumulate in an additive manner. By contrast, CAMIO selects a single gRNA from each set and complexity can be added by increasing the number of sets. Thus, CAMIO confers every gRNA with some autonomy while achieving multiplexed mutagenesis as a whole. Off-target concerns in CAMIO can be adequately addressed by examining multiple independent mutations of similar kinds. Also, arrays of targeted mutations can be introduced into specific genetic backgrounds with CAMIO. For example, when performing CAMIO on the *chinmo* 5' UTR, we purposefully targeted a 2L chromosome arm that also carries transgenes for twin-spot MARCM [44]. Hence, all the CAMIO *chinmo* indels were immediately ready for mosaic analysis. Independent indels can be directly screened for visible phenotypes in mosaic organisms. However, we favor mapping the indels first by NGS, which can be conducted in a high-throughput manner via combinatorial sample indexing. We have also reduced the costs by pooling distinct amplicons for co-indexing.

In general, we recovered similar indel spectrums to what has been commonly described. For gRNAs that are inherently potent, like g1_4 for *chinmo* 5' UTR, we obtained many indels around the cut site. Despite recovering numerous single-site deletions, we rarely see single-site deletions exceeding 30 bps in length. Therefore, the observed ease in creating diverse inter-target deletions by CAMIO

is particularly valuable for systematic DNA dissection. In the case of CAMIO on *chinmo* 5'UTR, we recovered most of the predicted inter-target deletions with the exception of the 'toxic' g2_2 and g2_3. Notably, the largest inter-target deletion we have identified exceeds 2 kb in length. These observations suggest that we can be more aggressive in choosing more disperse gRNA targets to cover larger genomic regions.

After all, the capacity of CAMIO is mainly determined by how many gRNAs one can pack into a single set. Given the small size of gRNAs, we expect no problem in packing six or even more gRNAs without compromising the system. This intuition was largely supported by seeing reasonable recruitment frequencies for all six tandem gRNAs in the *chinmo* exon 2 set. Further, there is still room for improvement on the gRNA selection process. For instance, increasing the distance between the U6 promoter and the gRNA set would make the selection more impartial. In sum, we have shown that the CAMIO system holds great promise for *in vivo* deletion analysis. Yet, our demonstrations have not nearly reached the limitations of CAMIO as far as the number of targets and size of DNA that can be evaluated in a single experiment.

We used CAMIO to perform deletion analysis on the 5'UTR of *chinmo,* which has important roles in governing Chinmo protein levels. There evidently exist multiple mechanisms governing *chinmo* expression throughout development. We successfully identified a region responsible for Chinmo downregulation in the MB

around pupa formation. The resulting elevated Chinmo expression affected MB morphogenesis, possibly due to abnormal neuronal fate transition. In addition, we found a large region critical for embryo viability. While roles for Chinmo have been described in the brain and as a downstream target of JAK/STAT in the testes[45], our data suggest another essential role for Chinmo in embryonic development. The identification of discrete non-coding regions regulating different biological processes within a single UTR has exemplified the utility of CAMIO in resolving complex UTR functions. Given its multiplex and combinatorial power, CAMIO should also greatly aid the dissection of promoters, enhancers, long non-coding RNAs, DNA repeats and more.

In theory, CAMIO should work for any organism where a *bamP*-like promoter exists. Particularly, if a founder parent (possibly a father) can produce a large number of targeted mutants, CAMIO may become a desirable genetic screening platform for interrogating the genome. A mouse gene, *Gm114*, was identified as a putative ortholog of *Drosophila bam*[46]. Encouragingly, strikingly similar to *bam, Gm114* is greatly enriched in undifferentiated spermatocytes and spermatids but absent or extremely low in undifferentiated spermatogonia. Orthologs of *bam* and *Gm114* were also found in zebrafish, chicken, macaque, and others. We have pioneered CAMIO as a germline pipeline for arrayed CRISPR mutagenesis in *Drosophila*. CAMIO can expedite systematic structure-functional analysis of the genome across diverse model organisms.

Methods

Fly strains.

We used the following fly strains in this work: (1) *bamP(898)-Cas9* in *attP2*; (2) *U6:3-gRNA-e*[19]; (3) *Df(3R)ED10838/TM2* (BDSC #9485); (4) *dU6-3-gRNA-shi* in *attP40*; (5) *UAS-shibire* in *attP2*; (6) *GMR-Gal4*; (7) *nos-phiC31-nls #12*; (8) *10XUAS-mCD8::GFP* in *attP40*[47]; (9) *act5C-Gal4/TM6B*; (10) *pCis-{4gRNAs_mCD8}* P element insertion line #3, #10 on II, and #9, #12 on III; (11) *13XLexAop2-5' UTR-smGFP-OLLAS* in *VK00027*, *13XLexAop2 –smGFP-cMyc-3' UTR* in *attP40*, and *13XLexAop2-5' UTR-smGFP-V5-3' UTR* in *su(Hw)attP5*; (12) *41A10-KD*; (13) *DpnEE-KO-LexAp65*; (14) *pCis-{4gRNAs_chinmo Exon1}* P element insertion line #23, #25 on II, and #4, #5 on III; *pCis-{6gRNAs_chinmo Exon2}* #7, #11 on II, and #14 on III; *pCis-{4gRNAs_chinmo Exon3}* #5 on II, and #2, #3 on III; (15) *FRT40A,UAS–mCD8::GFP,UAS–Cd2-Mir/CyO,Y*[44]; (16) *OK107-Gal4*; (17) *UAS-Cas9*[19]; (18) *dU6_g2+3* in *VK00027*; (19) *Dpn-Cas9* (unpublished reagent); (20) *act5C-Cas9*[19]; (21) *13XLexAop2-53UTR-smGFP-V5-d*, *13XLexAop2-53UTR-smGFP-V5-bD*, and *13XLexAop2-53UTR-smGFP-V5-D23* in *attP40*.

Molecular biology.

To create *bamP(898)-Cas9*, the full *bam* promoter (-898)[27] was ordered from gBlocks, IDT, and Cas9 was also flanked by *bam* 3' UTR. To create UAS-shibire,

548 codon-optimized *shibire* coding sequence that carries the gRNA-shi target site
549 was ordered from GeneArt gene synthesis, and then cloned into pJFRC28[48]. To
550 generate dU6-3-gRNAs, we replaced 10XUAS-IVS-GFP-p10 of pJFRC28 with
551 dU6-3 promoter and gRNA scaffold fragment from pTL2[28]. For dU6-3-gRNA-shi,
552 GTATGGGGTATCAAGCCGAT was selected as the spacer. To create
553 dU6_g2+3, we first generated dU6-3-g2_2 and dU6-3-g2_3 separately, and
554 cloned dU6-3-g2_2 into the backbone of dU6-3-g2_3.
555
556 To create conditional U6-gRNA set construct, pCis-{4gRNAs_mCD8}, a U6
557 promoter-AttB fragment was synthetized by PCR amplification from pCFD3[19] and
558 cloned into pCaST-elav-VP16AD, which contained the p-Element inverted repeats
559 (Addgene, #15308). Then, we inserted a DNA fragment (Genscript) containing 4
560 different gRNAs targeting the mCD8 protein tag. These gRNAs were selected
561 based on their ON and OFF target scores (Benchling). Each of these gRNAs was
562 preceded by an AttP site and a HammerHead ribozyme[49]. Finally, a 3Xp3-RFP-
563 polyA(α-tubulin) fragment was synthetized by PCR amplification, using pure
564 genomic DNA from a fly line in which this cassette was used as a marker
565 (Bloomington, #54590). Then, this fragment was inserted upstream of this gRNA
566 region. In the final construct, the AttB and AttP sites were separated by a 3.7 Kb
567 region containing an ampicillin resistance gene, an origin of replication in bacteria
568 and the 3Xp3-RFP-polyA marker.
569

570 pCis-{gRNAs_chinmo-Exon1/Exon2/Exon3}: following the same design
571 described above, a DNA fragment was synthetized (Genscript), which contained
572 4 gRNAs (6 for Exon2) either targeting the corresponding exon or the exon-intron
573 junction. This fragment was then inserted into pCis-{4gRNAs_mCD8}, thus
574 removing the previous gRNAs cassette.
575
576 The *Chinmo* UTRs were amplified from Drosophila genomic DNA. smGFP[50]
577 fused to V5, cmyc or ollas were amplified from previously existing plasmids.
578 Standard molecular biology techniques were used to clone the smGFP fusions
579 containing one or two *Chinmo* UTRs into 13XLexAop2 (pJFRC15,[47]). If the
580 construct ended with *Chinmo* 3'UTR, the SV40 signal was removed from the
581 vector backbone. *13XLexAop2-5' UTR-GFP-3' UTR* was further modified to
582 create d, bD, and D23 reporters containing various deletions in the 5' UTR.
583
584 ***Drosophila* genetics.**
585 ***ebony* and UAS-shibire mutagenesis by *bamP(898)-Cas9*.** Female or male
586 founders (*bamP(898)-Cas9/U6:3-gRNA-e*) were mated to *Df(3R)ED10838/TM6B*,
587 and chromosomes over *Df(3R)ED10838* were scored for ebony loss-of-function
588 phenotype. In total, 709 progeny from 5 female founders and 1574 progeny from
589 9 male founders were collected and phenotype was assessed. Male *UAS-shibire*
590 mutagenesis founders were crossed with *GMR-Gal4*, and the wildtype-eyed 919
591 progeny were sacrificed for next generation sequencing (NGS).
592

CAMIO on UAS-mCD8::GFP. Male founders (*UAS-mCD8::GFP*, *bamP(898)-Cas9*, *nos-phiC31*, and one or two copies of *pCis-{4gRNAs_mCD8}* were mated with *act5C-Gal4* females for scoring of loss of green fluorescence in the progeny. For one copy of *pCis-{4gRNAs_mCD8}*, we screened 807 progeny from 20 founder males. 30 loss-of-GFP progeny from two founders were further subject to sequence analysis. For two copies of *pCis-{4gRNAs_mCD8}*, we screened 570 progeny from 16 founders. 110 loss-of-GFP progeny were analyzed and grouped into three deletion categories.

CAMIO on *chinmo* 5' UTR. After mating with females carrying a second chromosome balancer for stock keeping, 1200 male progeny from 4 CAMIO on *chinmo* 5' UTR gRNA set combinations were sacrificed for genomic study. We designed primer sets that produce amplicons covering exon 1, 2, 3, and exons 1+2. Males from combination 1 were intentionally numbered 1-300, and their genomic amplicons were carefully matched and mixed with counterparts from combination 2 and 3. Finally, 300 DNA mixtures plus 84 amplicons from combination 4 were tagmented (Nextera XT DNA Library Prep Kit, illumina) and barcoded (Nextera XT Index Kit v2) for NGS.

Immunohistochemistry and confocal imaging.

Fly brains at indicated larval, pupal, and adult stages were dissected, fixed, and immunostained as described previously[29,44]. The following primary antibodies were used in this study: chicken GFP polyclonal antibody (1:500, Invitrogen,

A10262); rat anti-Deadpan (1:100, abcam, ab195173); mouse 1D4 anti-Fasciclin II (1:50, DSHB); rabbit polyclonal anti-Trio (1:1000)[51]; rat anti-Chinmo (1:500, a gift from the Sokol lab)[41]. Secondary antibodies from Molecular Probes were all used in a 1:200 dilution. The immunofluorescent signals were collected using Zeiss LSM 880 confocal microscope and processed using Fiji and Adobe Illustrator.

Bioinformatics.

Sequence reads (FASTQ data) were first processed with cutadapt [https://github.com/marcelm/cutadapt] to remove adapter sequences with options: --overlap=7 --minimum-length=30 -a "CTGTCTCTTATACACATCTCTGAGCGGGCTGGCAAGGCAGACCG". Then they were mapped to the genomic sequence corresponding to the *chinmo* region using Bowtie2[52] with the following options: --local --score-min G,20,0 -D 20 -R 3 -L 20 -i S,1,0.50 --no-unal. Resulting SAM files were parsed with Pysam [https://github.com/pysam-developers/pysam] to extract deletion/insertion information from the Cigar strings. When the cigar string contained 'D' or 'N', we extracted mapped sequences as deletions and designated them as type D. When the cigar string contained 'I', we extracted them as insertions and designated them as type I. While parsing the SAM file, soft clipped reads (reads partially mapped to the genome) were detected and clipped (unmapped) portions were set aside in a FASTA file. The sequences in this FASTA file were then re-mapped using bowtie2 to the genomic sequence encompassing the *chinmo*

639 gene. Then, mapped portions of the first mapping and that of the second
640 mapping (when the second one existed) were merged to form a deletion event
641 which we denoted as type L (large gap). This type of gap often contained
642 inserted sequences in the middle. We discarded any events with less than 10
643 reads.

References

1 Chalfie, M. & Kain, S. R. *Green fluorescent protein: properties, applications and protocols*. Vol. 47 (John Wiley & Sons, Inc. , 2005).

2 Boulin, T., Etchberger, J. F. & Hobert, O. *Reporter gene fusions*. (WormBook, ed. The C. elegans Research Community, WormBook, 2006).

3 Kvon, E. Z., Stampfel, G., Yanez-Cuna, J. O., Dickson, B. J. & Stark, A. HOT regions function as patterned developmental enhancers and have a distinct cis-regulatory signature. *Genes Dev* **26**, 908-913, doi:10.1101/gad.188052.112 (2012).

4 Pennacchio, L. A. *et al.* In vivo enhancer analysis of human conserved non-coding sequences. *Nature* **444**, 499-502, doi:10.1038/nature05295 (2006).

5 Dupuy, D. *et al.* A first version of the Caenorhabditis elegans Promoterome. *Genome research* **14**, 2169-2175, doi:10.1101/gr.2497604 (2004).

6 Kim, Y.-G., Cha, J. & Chandrasegaran, S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences* **93**, 1156-1160 (1996).

7 Bibikova, M., Golic, M., Golic, K. G. & Carroll, D. Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. *Genetics* **161**, 1169-1175 (2002).

8 Beumer, K. J. *et al.* Efficient gene targeting in Drosophila by direct embryo injection with zinc-finger nucleases. *Proceedings of the National Academy of Sciences* **105**, 19821-19826 (2008).

9 Kim, H. & Kim, J.-S. A guide to genome engineering with programmable nucleases. *Nature Reviews Genetics* (2014).

10 Miller, J. C. *et al.* A TALE nuclease architecture for efficient genome editing. *Nature biotechnology* **29**, 143-148 (2011).

11 Boch, J. *et al.* Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326**, 1509-1512 (2009).

12 Moscou, M. J. & Bogdanove, A. J. A simple cipher governs DNA recognition by TAL effectors. *Science* **326**, 1501-1501 (2009).

13 Mussolino, C. *et al.* TALENs facilitate targeted genome editing in human cells with high specificity and low cytotoxicity. *Nucleic acids research* **42**, 6762-6773 (2014).

14 Jinek, M. *et al.* A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821 (2012).

15 Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823 (2013).

16 Hwang, W. Y. *et al.* Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nature biotechnology* **31**, 227-229 (2013).

17 Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823-826 (2013).

18 Gratz, S. J. *et al.* Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. *Genetics* **194**, 1029-1035 (2013).

19 Port, F., Chen, H. M., Lee, T. & Bullock, S. L. Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America* **111**, E2967-2976, doi:10.1073/pnas.1405500111 (2014).

20 Komor, A. C., Badran, A. H. & Liu, D. R. CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. *Cell* **169**, 559, doi:10.1016/j.cell.2017.04.005 (2017).

21 Lieber, M. R. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annual review of biochemistry* **79**, 181-211, doi:10.1146/annurev.biochem.052308.093131 (2010).

22 San Filippo, J., Sung, P. & Klein, H. Mechanism of eukaryotic homologous recombination. *Annual review of biochemistry* **77**, 229-257, doi:10.1146/annurev.biochem.77.061306.125255 (2008).

23 Fellmann, C., Gowen, B. G., Lin, P. C., Doudna, J. A. & Corn, J. E. Cornerstones of CRISPR-Cas in drug discovery and therapy. *Nature reviews. Drug discovery* **16**, 89-100, doi:10.1038/nrd.2016.238 (2017).

24 Wang, H., La Russa, M. & Qi, L. S. CRISPR/Cas9 in Genome Editing and Beyond. *Annual review of biochemistry* **85**, 227-264, doi:10.1146/annurev-biochem-060815-014607 (2016).

25 Hales, K. G., Korey, C. A., Larracuente, A. M. & Roberts, D. M. Genetics on the Fly: A Primer on the Drosophila Model System. *Genetics* **201**, 815-842, doi:10.1534/genetics.115.183392 (2015).

26 Bier, E., Harrison, M. M., O'Connor-Giles, K. M. & Wildonger, J. Advances in Engineering the Fly Genome with the CRISPR-Cas System. *Genetics* **208**, 1-18, doi:10.1534/genetics.117.1113 (2018).

27 Chen, D. & McKearin, D. M. A discrete transcriptional silencer in the bam gene determines asymmetric division of the Drosophila germline stem cell. *Development* **130**, 1159-1170 (2003).

28 Chen, H. M., Huang, Y., Pfeiffer, B. D., Yao, X. & Lee, T. An enhanced gene targeting toolkit for Drosophila: Golic+. *Genetics* **199**, 683-694, doi:10.1534/genetics.114.173716 (2015).

29 Zhu, S. *et al.* Gradients of the Drosophila Chinmo BTB-zinc finger protein govern neuronal temporal identity. *Cell* **127**, 409-422, doi:10.1016/j.cell.2006.08.045 (2006).

30 Park, J., Lim, K., Kim, J. S. & Bae, S. Cas-analyzer: an online tool for assessing genome editing results using NGS data. *Bioinformatics (Oxford, England)* **33**, 286-288, doi:10.1093/bioinformatics/btw561 (2017).

31 Allen, F. *et al.* Predicting the mutations generated by repair of Cas9-induced double-strand breaks. *Nat Biotechnol*, doi:10.1038/nbt.4317 (2018).

32 Sfeir, A. & Symington, L. S. Microhomology-Mediated End Joining: A Back-up Survival Mechanism or Dedicated Pathway? *Trends in biochemical sciences* **40**, 701-714, doi:10.1016/j.tibs.2015.08.006 (2015).

33 Ferguson, S. M. & De Camilli, P. Dynamin, a membrane-remodelling GTPase. *Nature reviews. Molecular cell biology* **13**, 75-88, doi:10.1038/nrm3266 (2012).

34 Kitamoto, T. Conditional modification of behavior in Drosophila by targeted expression of a temperature-sensitive shibire allele in defined neurons. *Journal of neurobiology* **47**, 81-92 (2001).

35 Lehmann, R. Germline stem cells: origin and destiny. *Cell stem cell* **10**, 729-739, doi:10.1016/j.stem.2012.05.016 (2012).

36 Kanca, O., Caussinus, E., Denes, A. S., Percival-Smith, A. & Affolter, M. Raeppli: a whole-tissue labeling tool for live imaging of Drosophila development. *Development* **141**, 472-480, doi:10.1242/dev.102913 (2014).

37 Narbonne-Reveau, K. *et al.* Neural stem cell-encoded temporal patterning delineates an early window of malignant susceptibility in Drosophila. *eLife* **5**, doi:10.7554/eLife.13463 (2016).

38 Dillard, C., Narbonne-Reveau, K., Foppolo, S., Lanet, E. & Maurange, C. Two distinct mechanisms silence chinmo in Drosophila neuroblasts and neuroepithelial cells to limit their self-renewal. *Development* **145**, doi:10.1242/dev.154534 (2018).

39 Kao, C. F., Yu, H. H., He, Y., Kao, J. C. & Lee, T. Hierarchical deployment of factors regulating temporal fate in a diverse neuronal lineage of the Drosophila central brain. *Neuron* **73**, 677-684, doi:10.1016/j.neuron.2011.12.018 (2012).

40 Ren, Q. *et al.* Stem Cell-Intrinsic, Seven-up-Triggered Temporal Factor Gradients Diversify Intermediate Neural Progenitors. *Current biology : CB* **27**, 1303-1313, doi:10.1016/j.cub.2017.03.047 (2017).

41 Wu, Y. C., Chen, C. H., Mercer, A. & Sokol, N. S. Let-7-complex microRNAs regulate the temporal identity of Drosophila mushroom body neurons via chinmo. *Dev Cell* **23**, 202-209, doi:10.1016/j.devcel.2012.05.013 (2012).

42 Marchetti, G. & Tavosanis, G. Steroid Hormone Ecdysone Signaling Specifies Mushroom Body Neuron Sequential Fate via Chinmo. *Current biology : CB* **27**, 3017-3024.e3014, doi:10.1016/j.cub.2017.08.037 (2017).

43 Fuller, M. T. & Spradling, A. C. Male and female Drosophila germline stem cells: two versions of immortality. *Science* **316**, 402-404, doi:10.1126/science.1140861 (2007).

44 Yu, H.-H., Chen, C.-H., Shi, L., Huang, Y. & Lee, T. Twin-spot MARCM to reveal the developmental origin and identity of neurons. *Nature neuroscience* **12**, 947-953 (2009).

45 Flaherty, M. S. *et al.* chinmo is a functional effector of the JAK/STAT pathway that regulates eye development, tumor formation, and stem cell self-renewal in Drosophila. *Dev Cell* **18**, 556-568, doi:10.1016/j.devcel.2010.02.006 (2010).

46 Tang, H., Ross, A. & Capel, B. Expression and functional analysis of Gm114, a putative mammalian ortholog of Drosophila bam. *Dev Biol* **318**, 73-81, doi:10.1016/j.ydbio.2008.03.001 (2008).

47 Pfeiffer, B. D. *et al.* Refinement of tools for targeted gene expression in Drosophila. *Genetics* **186**, 735-755 (2010).

48 Pfeiffer, B. D., Truman, J. W. & Rubin, G. M. Using translational enhancers to increase transgene expression in Drosophila. *Proceedings of the National Academy of Sciences* **109**, 6626-6631 (2012).

49 He, Y. *et al.* Self-cleaving ribozymes enable the production of guide RNAs from unlimited choices of promoters for CRISPR/Cas9 mediated genome editing. *Journal of genetics and genomics = Yi chuan xue bao* **44**, 469-472, doi:10.1016/j.jgg.2017.08.003 (2017).

50 Viswanathan, S. *et al.* High-performance probes for light and electron microscopy. *Nat Methods* **12**, 568-576, doi:10.1038/nmeth.3365 (2015).

51 Awasaki, T. *et al.* The Drosophila trio plays an essential role in patterning of axons by regulating their directional extension. *Neuron* **26**, 119-131 (2000).

52 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).

Acknowledgements

We thank Janelia Fly Core and Quantitative Genomics for technical support. We thank Crystal Di Pietro and Kathryn Miller for administrative support. This work was supported by Howard Hughes Medical Institute.

Author contributions

H.-M.C. and T.L. conceived the project. H.-M.C. performed the CAMIO experiments. J.G.M. conceptualized the design of CAMIO gRNA multiplexing and generated the CAMIO gRNA sets. K.S., H.-M.C., and T.L. contributed to the design, execution, and analyses of the CAMIO NGS project. D.W. contributed to the sample preparation for CAMIO NGS. R.L.M. designed the chinmo-UTRs reporter constructs. K.S. provided statistical data analyses. H.-M.C., R.L.M. and T.L. wrote the manuscript. T.L. supervised the project.

Competing interests

The authors declare no competing interests.

Figure legends

Figure 1. *bamP-Cas9* induces efficient CRISPR targeted mutagenesis in male germ cells. (A) Ebony transcript shown with UTRs in turquoise. U6 drives guide RNA (*gRNA-e*), which targets 5' end of *ebony* coding sequence. (B) Percent of progeny with ebony loss-of-function (LOF) mutations from female or male founders. Mean ± SEM (n=9). (C) Sequences of 10 randomly selected *ebony* LOF progeny from the same male founder. (D) Deletion profile of *UAS-shibire* transgene following CRISPR targeting collected from 919 phenotypically wildtype progeny. NGS data was analyzed and presented by Cas-analyzer, www.rgenome.net[30]. (E) Left, percentage of predicted indels (calculated by FORECasT[31] using *gRNA-shi* and *UAS-shibire* sequences) grouped into 3 reading frames, 0 represents the in-frame indels. Right) Actual reading frame percentages from phenotypically wildtype progeny, calculated from mapping results (obtained with CAS-analyzer). Chi-square test assuming equal distribution was used to assess the significance. The result is below machine precision and thus set to zero.

Figure 2. CAMIO produces diverse arrayed mutations around selected gRNA target sites. (A) Schematic of a conditional U6 gRNA transgene: pCis-{4gRNAs_mCD8}. The U6 promoter is separated from the gRNAs by a large fragment containing a 3xP3-RFP marker. With phiC31 recombination, the

attachment site, attB can recombine stochastically with any of the four attP sites (black arrows), creating attR. The phiC31 recombinase is expressed in GSCs, controlled by *nos.* After recombination, a single gRNA is under the control of the U6 promoter (yellow oval) (B) 4 gRNA target sites along the mCD8 coding sequence were selected for pCis-{4gRNAs_mCD8} to disrupt mCD8::GFP expression. Simple or combined multiplexing mutagenesis can be achieved by incorporating one or more pCis-{4gRNAs_mCD8} transgenes. Here, we show two transgenes, each represented by four pair of scissors. Stochastically chosen gRNAs in each set are further marked in yellow. (C) Three categories of deletions arose from applying two copies of pCis-{4gRNAs_mCD8}: single-site deletion, defined deletion, and double-site deletion. Deletion size and count are depicted for single-site deletions. Right, use matrix of gRNA choices, deduced from sequencing GFP negative progeny.

Figure 3. Applying CAMIO on *chinmo* 5' UTR. (A) An illustration of the *chinmo* gene, UTRs are depicted in turquoise. (B) We selected 4 target sites on *chinmo* exon 1 (455 bps), 6 target sites on exon 2 (1061 bps), and 4 target sites on exon 3 (639 bps). Each exon was dissected by CAMIO with a pair of gRNA sets, depicted as a set of scissors. Additionally, we combined gRNA sets from exon1 and exon2. (C) Representation of larger deletions predicted by NGS of 984 offspring (300 each for individual exons and 84 for Ex1-Ex2). Predicted deletions over 100 bps were subject to Sanger sequencing for confirmation. Confirmed >100 bps deletions are marked in red and given an ID#. Selected

853 homozygous viable deletions (marked in green) that cover larger proportions of
854 each exons were selected for MB development studies.
855
856 **Figure 4. Overlapping deletions in 5'UTR alter Chinmo expression and MB**
857 **morphology.** (A) Stacked confocal images of adult MBs stained for FasII (green,
858 αβ and γ lobes) and Trio (red, α'β' and γ lobes). Missing or misshapen αβ lobes
859 (arrows). Scale bars: 50 μm. (B) Percent of flies with abnormal αβ lobes.
860 Homozygous and transheterozygous deletions have strong MB αβ lobe defects,
861 whereas hemizygous deletions over a *chinmo* deficiency line (*Df*) are much less
862 penetrant. (C) Single optical sections of MB lineages (*OK107-Gal4*, *UAS-*
863 *mCD8::GFP*) immunostained for GFP and Chinmo at white pupa stage. Chinmo
864 staining is elevated in newly derived MB neurons (weaker GFP signal, outlined
865 by white dashed lines) in homozygous $chinmo^{Ex2L56}$ and $chinmo^{Ex2L130}$. Green:
866 GFP; Magenta: Chinmo. Scale bars: 20 μm.
867

Supplementary information

Supplementary figure legends

Supplementary Figure 1. Targeted mutagenesis of mCD8::GFP with pCis-{4gRNAs_mCD8}. pCis-{4gRNAs_mCD8} was created to target four different sites along the mCD8 coding sequence of UAS-mCD8::GFP. We analyzed 30 GFP-negative progeny from two different male founders. Each GFP-negative progeny carried a indel located at one of the four target sites. The location, size, and overall counts of these 30 indels are summarized here.

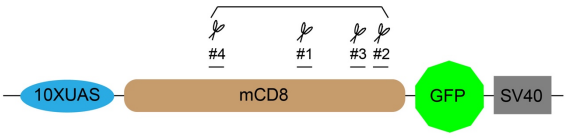
Supplementary Figure 2. Three *13XLexAop2-GFP* reporters for investigating *chinmo* 5' and 3' UTRs. Top: diagram of 13XLexAop2-GFP reporters flanked by both *chinmo* 5' and 3' UTRs, only the 5' UTR, or only the 3' UTR. Bottom: GFP transgene expression was induced and restricted in MB lineages by immortalizing a transient MB NB expression (*41A10-KD*) into a sustained MB NB production of these three GFPs, with this genetic setup: *DpnEE-KO-LexAp65; 13XLexAop2-GFP-UTRs; 41A10-KD*. Their expression profile is shown by immunostaining for GFP at two time points (48 and 72 hours after larval hatching, ALH). GFP with the 3' UTR is expressed at a much higher level than GFP with both 5' and 3' UTRs. Expression of GFP with the 5' UTR decreases over time. Green: GFP; Red: Deadpan. Scale bars: 20 μm.

Supplementary Figure 3. The small indels created by CAMIO in *chinmo* Exon 1, 2, and 3. We sequenced 984 progeny (300 each for individual exons, and additional 84 for Ex1-Ex2) from our CAMIO on *chinmo* 5' UTR experiment. The size, location, and corresponding stock ID of the predicted deletions (marked in red) and insertions (marked in blue) are depicted here.

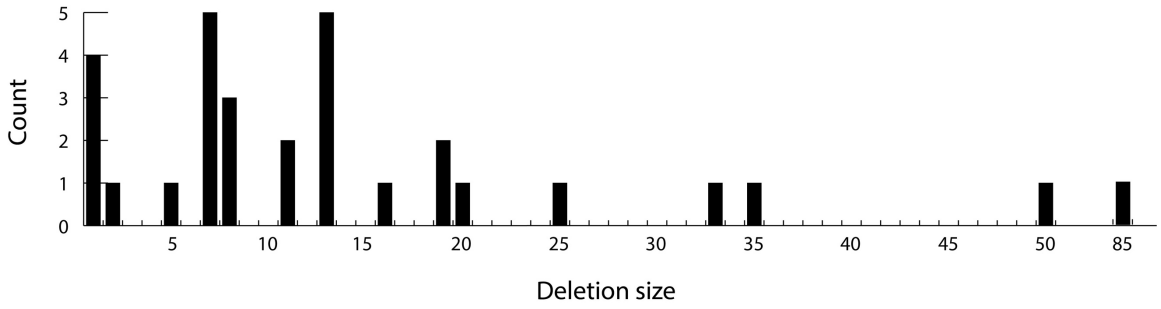
Supplementary Figure 4. Examining gRNA array selection of *pCis-(6gRNAs_chinmo Exon2)*. We studied how frequently each gRNA gets selected after nos-phiC31 mediated recombination by scoring progeny of male *nos-phiC31; pCis-(6gRNAs_chinmo Exon2*). (A) gRNA choices of 92 progeny from 9 male founders are presented in this heat map. (B) Counts for each gRNA from all the founders are aggregated and whether there is bias in selection is tested by Chi-square test assuming equal distribution. g2_2 (6.5%, 6/92) and g2_3 (7.6%, 7/92) are underrepresented.

Supplementary Figure 5. Three 5'UTR-GFP-3'UTR reporters carrying small to large *chinmo* Exon 2 deletions for studying the region around g2_2 and g2_3 target sites. We generated three additional chinmo 5' UTR-GFP-3' UTR reporters: d (50 bps deletion around g2_3), bD (larger 158 bps deletion), and D23 (removing all the Exon 2 sequence upstream of g2_3). GFPs were induced in the same *41A10-KD* immortalization strategy. We did not observe difference in expression among the four GFP UTR reporters at three development time points: 48h AEL (after egg laying), 96h AEL and, white pupa. Scale bars: 30 μm.

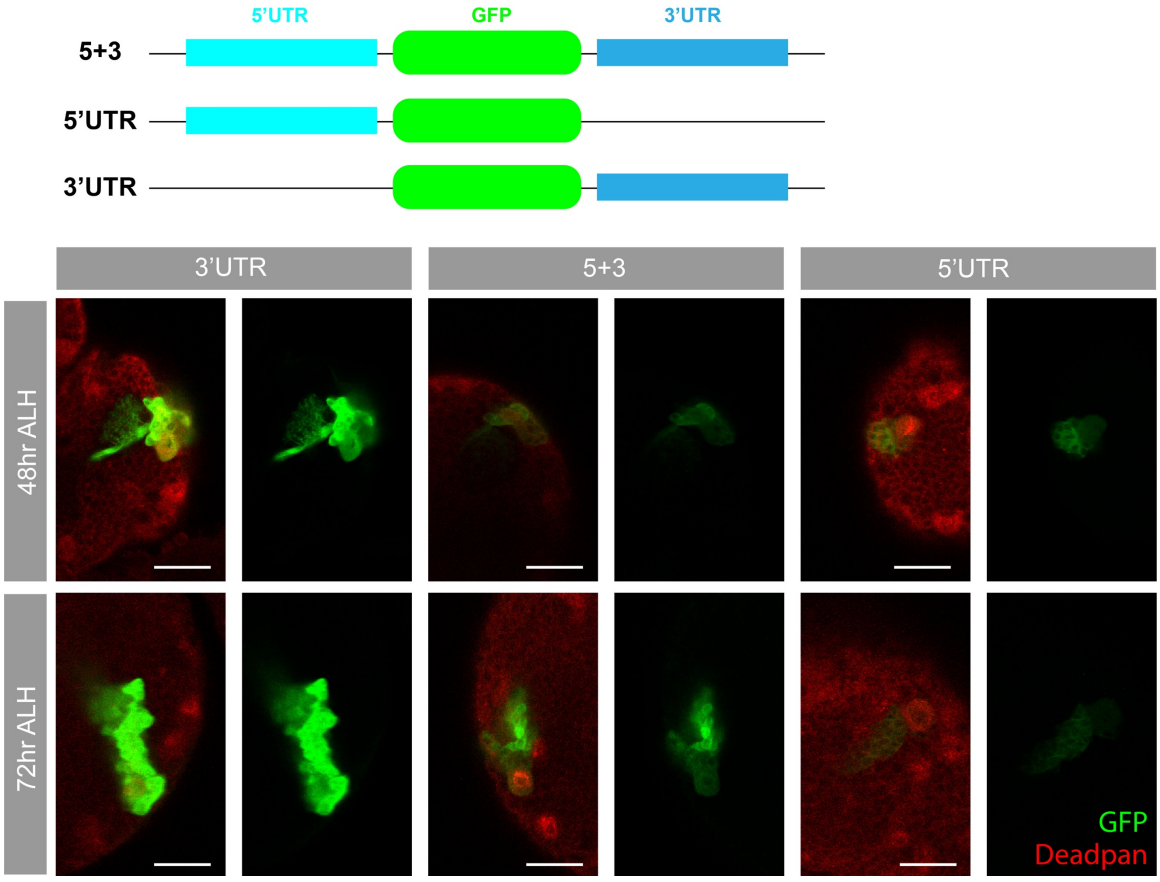
Supplementary figures



	#1	#2	#3	#4
Founder A	10	0	1	4
Founder B	9	2	1	3
Overall	62%	7%	7%	24%

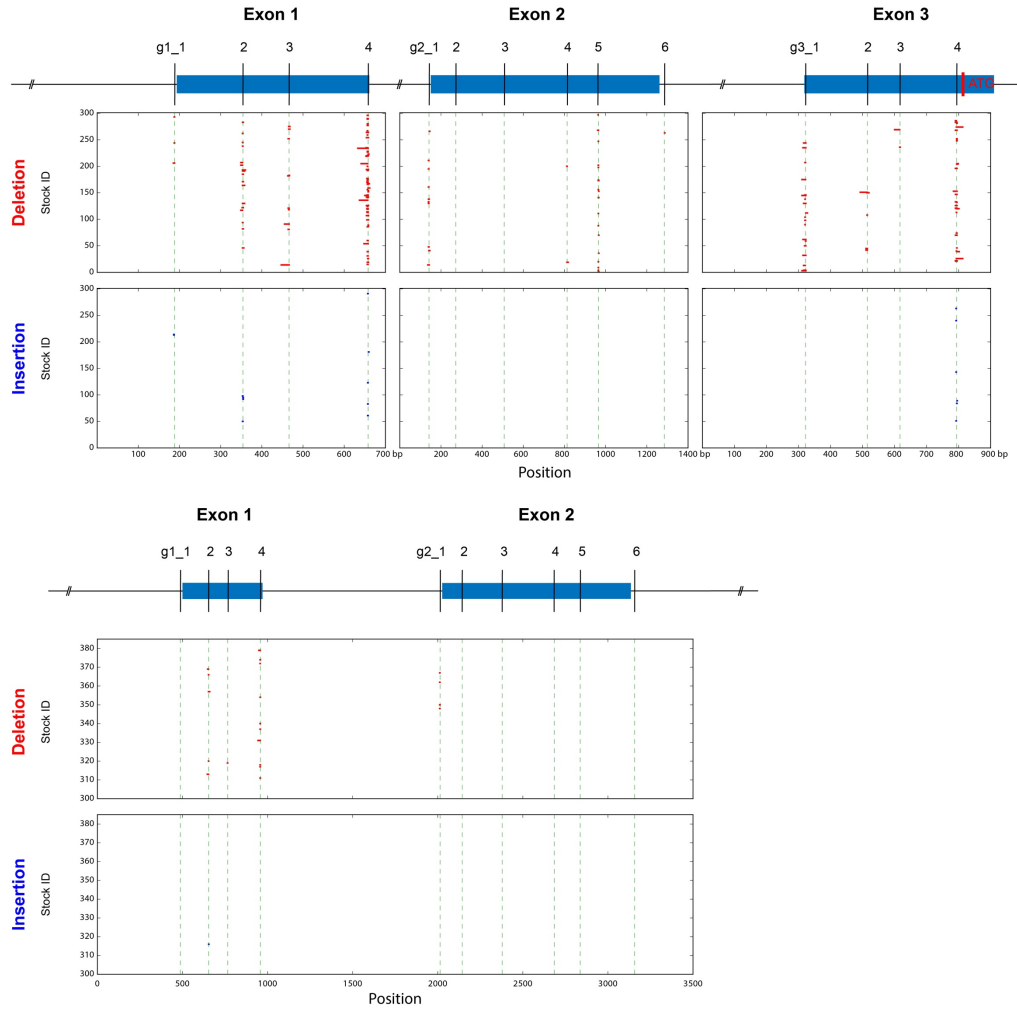


Supplementary Figure 1. Targeted mutagenesis of mCD8::GFP with pCis-{4gRNAs_mCD8}. pCis-{4gRNAs_mCD8} was created to target four different sites along the mCD8 coding sequence of UAS-mCD8::GFP. We analyzed 30 GFP-negative progeny from two different male founders. Each GFP-negative progeny carried a indel located at one of the four target sites. The location, size, and overall counts of these 30 indels are summarized here.

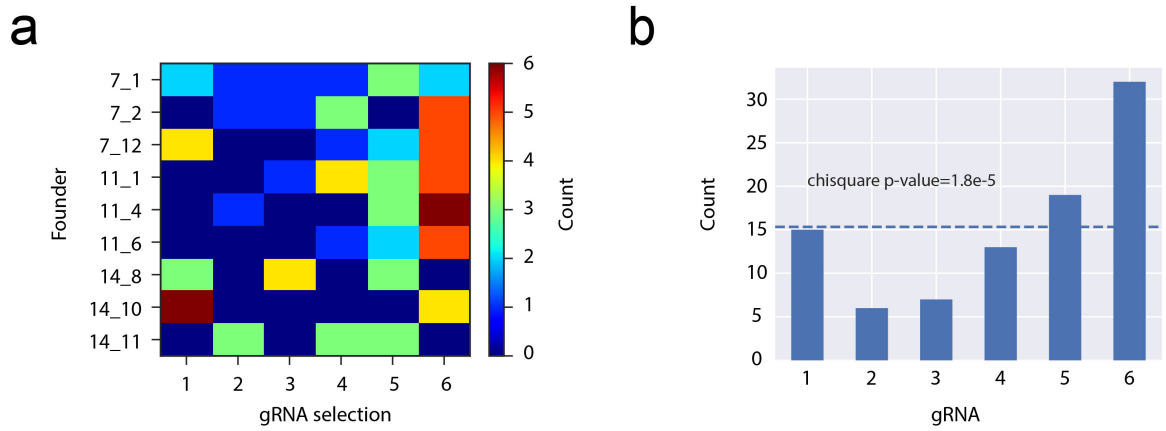


924

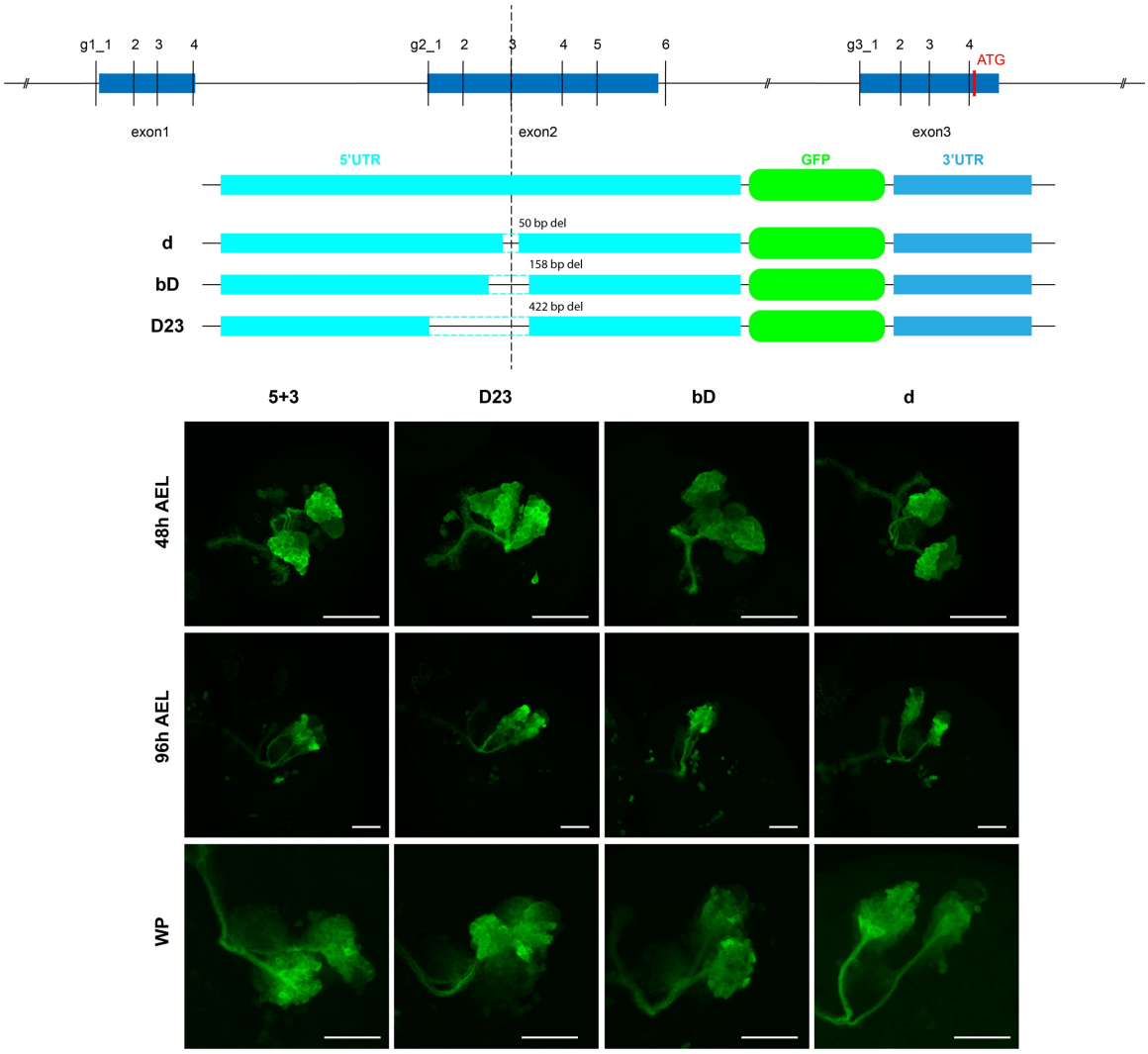
925 **Supplementary Figure 2. Three *13XLexAop2-GFP* reporters for**
926 **investigating *chinmo* 5' and 3' UTRs.** Top: diagram of 13XLexAop2-GFP
927 reporters flanked by both *chinmo* 5' and 3' UTRs, only the 5' UTR, or only the 3'
928 UTR. Bottom: GFP transgene expression was induced and restricted in MB
929 lineages by immortalizing a transient MB NB expression (*41A10-KD*) into a
930 sustained MB NB production of these three GFPs, with this genetic setup:
931 *DpnEE-KO-LexAp65; 13XLexAop2-GFP-UTRs; 41A10-KD*. Their expression
932 profile is shown by immunostaining for GFP at two time points (48 and 72 hours
933 after larval hatching, ALH). GFP with the 3' UTR is expressed at a much higher
934 level than GFP with both 5' and 3' UTRs. Expression of GFP with the 5' UTR
935 decreases over time. Green: GFP; Red: Deadpan. Scale bars: 20 μm.
936



Supplementary Figure 3. The small indels created by CAMIO in *chinmo* Exon 1, 2, and 3. We sequenced 984 progeny (300 each for individual exons, and additional 84 for Ex1-Ex2) from our CAMIO on *chinmo* 5' UTR experiment. The size, location, and corresponding stock ID of the predicted deletions (marked in red) and insertions (marked in blue) are depicted here.



Supplementary Figure 4. Examining gRNA array selection of *pCis-(6gRNAs_chinmo Exon2)*. We studied how frequently each gRNA gets selected after nos-phiC31 mediated recombination by scoring progeny of male *nos-phiC31; pCis-(6gRNAs_chinmo Exon2*). (A) gRNA choices of 92 progeny from 9 male founders are presented in this heat map. (B) Counts for each gRNA from all the founders are aggregated and whether there is bias in selection is tested by Chi-square test assuming equal distribution. g2_2 (6.5%, 6/92) and g2_3 (7.6%, 7/92) are underrepresented.



952

953 **Supplementary Figure 5. Three 5'UTR-GFP-3'UTR reporters carrying small**
954 **to large *chinmo* Exon 2 deletions for studying the region around g2_2 and**
955 **g2_3 target sites.** We generated three additional chinmo 5' UTR-GFP-3' UTR
956 reporters: d (50 bps deletion around g2_3), bD (larger 158 bps deletion), and
957 D23 (removing all the Exon 2 sequence upstream of g2_3). GFPs were induced
958 in the same *41A10-KD* immortalization strategy. We did not observe difference in
959 expression among the four GFP UTR reporters at three development time points:
960 48h AEL (after egg laying), 96h AEL and, white pupa. Scale bars: 30 μm.