1

2

CAMIO for deletion analysis of endogenous DNA sequences in multicellular organisms

5

6

7 Hui-Min Chen, Jorge Garcia Marques, Ken Sugino, Dingjun Wei, Rosa Linda

8 Miyares and Tzumin Lee*

9

10

11 Howard Hughes Medical Institute, Janelia Research Campus, 19700 Helix Drive,

12 Ashburn, VA 20147, USA

13

14

15 * To whom correspondence should be addressed.

16 Tel: 571-209-4198

17 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 genes³⁻⁵. 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

non-physiological, prompting the search for tailor-made sequence-specific DNA nucleases that permit targeted mutagenesis of native sequences.

In the early 2000s, zinc-finger nucleases (ZFNs)⁶, the first of the tailored nucleases, were heavily adopted^{7,8}. Although it seemed a promising strategy for genome editing, major setbacks, especially off-target toxicity limited its utility⁹. Therefore, the enthusiasm for ZFNs declined after the arrival of transcription activator-like effector nucleases (TALENs)¹⁰⁻¹². The higher specificity of TALENs led to fewer off-target disruptions, and hence less cytotoxicity¹³. However, constructing a TALEN is technically challenging because the homologous sequences encoding the TALEN repeats are prone to recombine with each other⁹.

The newest technology, CRISPR (clustered regularly interspaced short palindromic repeats), was first exploited for targeted mutagenesis¹⁴ and then swiftly adopted to edit genomes of diverse organisms¹⁵⁻²⁰. CRISPR's popularity lies in its simplicity: a Cas9 nuclease plus an easily made guide RNA (gRNA) induces a double-strand break (DSB) targeted by DNA-RNA base pairing^{14,16}. Cells repair the DSBs either through nonhomologous end joining (NHEJ)²¹, leading to insertion or deletion (indel), or through homology-directed repair (HDR)²² which replaces discontinuous sequences with an available template. The fact that CRISPR can produce targeted DNA modifications with near unlimited specificity has ensured its rapid expansion throughout the biological

and biomedical fields²³. Indeed, CRISPR technology has already transformed studies from stem cell research and cancer biology to food production and pest control²⁴.

Drosophila melanogaster has been a powerful model organism for decoding the genome to understand complex biology²⁵. 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 analyses²⁶.

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'UTR²⁹. 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 GSCs²⁸. *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 $\pm$ 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.net³⁰. (E) Left, percentage of predicted indels (calculated by FORECasT³¹ 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.

demonstrates that the male germline is particularly susceptible to *bamP*-Cas9-mediated genome editing.

To address if the Cas9-mediated editing events occurred independently, we sequenced a part of the *ebony* locus in individual mutants carrying *ebony* deficiency. From a single male founder, we analyzed ten progeny with *ebony* loss-of-function phenotypes. We uncovered seven different indels around the Cas9 cut site (Fig. 1c). One identical *ebony* mutation occurred in three siblings, possibly resulting from either differential deletion of GTC repeats or from microhomology-mediated repair³². The recovery of many distinct indels from a single founder argues for independent Cas9 actions in individual germ cells. This result encouraged us to establish *bamP* induced CRISPR in the male germline as a ‘targeted mutagenesis pipeline’.

A mutagenesis pipeline could be useful for producing novel alleles of a protein of interest. To explore this idea, we tested CRISPR’s usefulness to produce novel alleles of a well characterized gene, *shibire*. *shibire* encodes *Drosophila* Dynamin, a motor protein crucial for synaptic vesicle endocytosis³³. *shi^{ts1}*, a temperature sensitive allele containing a missense mutation, is widely used in behavioral assays to temporarily shut off neuronal activity³⁴. We therefore designed a gRNA against the *UAS-shibire* transgene at the region where the temperature-sensitive *shi^{ts1}* point mutation is located, presuming that we could produce additional temperature-sensitive or dominant negative alleles.

170

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

188

189 Taken together, our data demonstrate that *bamP(898)* effectively restricts Cas9-
190 mediated mutagenesis to germ cells. There is no evidence that clonal expansion
191 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.

promoter, we control the induction of the phiC31 recombinase in GSCs. Thus, in each of the 12-24 GSCs per male fly founder³⁵, the transgene is irreversibly recombined to express a single gRNA. Recombination occurs between a single attB site downstream of the U6 promoter and a choice of attP sites upstream of each gRNA; once reconstituted, the ubiquitous U6 promoter drives expression of only one of the gRNAs. The intra-chromosomal recombination excises an intervening 3xP3-RFP. Given the rather small size of each gRNA as compared to the large 3xP3-RFP, the differences in length between the attB site and the choice of any one attP site is therefore relatively trivial. Based on a previous, similar construct for multicolor imaging³⁶, we expect that each gRNA should be expressed at comparable frequencies. For brevity, we name the conditional U6-gRNA transgene pCis, and then, in braces, add the number of gRNAs and the name of the targeted DNA. For example, to target the UAS-mCD8::GFP transgene, we created pCis-{4gRNAs_mCD8}. Also, when we describe the individual gRNAs, we number them in sequence from 5' to 3'.

For the multiplexed targeted mutagenesis of mCD8::GFP, we established male founders carrying *UAS-mCD8::GFP*, *bamP(898)-Cas9*, *nos-phiC31*, and *pCis-{4gRNAs_mCD8}*, and crossed them to *act5C-Gal4* females for easy scoring of GFP fluorescence in the progeny. Overall, approximately 35% of the progeny lost GFP expression. We collected 30 GFP-negative offspring from two founder males. Sequencing the mCD8-coding region revealed that each GFP-negative 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

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

278
279 The above results demonstrate that using dual gRNA sets enables us to tile a
280 region of interest not only with indels, but also with defined deletions. Random
281 selection of a single gRNA from each of the two identical sets which contain four
282 gRNAs will yield six possible defined deletions. Encouragingly, from a collection
283 of only nine inter-target deletions, we recovered five of the six anticipated defined
284 deletions. Nevertheless, physical hindrance may prevent two Cas9 complexes
285 from acting simultaneously on very close gRNA targets. This may explain why we
286 failed to recover the smallest defined deletion of 37 bp between the Cas9 cut
287 sites of gRNA#2 and #3 targets. These results suggest that inter-target deletions
288 utilizing two gRNAs can support rapid systematic DNA deletion analysis.

289
290 In sum, we established an effective strategy to express various permutations of
291 two gRNAs in male GSCs. In combination with restricting Cas9 to male germ
292 cells, we built a germline pipeline for multiplex targeted mutagenesis. We dub
293 this genetic system CAMIO (Cas9-mediated arrayed mutagenesis of individual
294 offspring), which can derive from a single population cross a matrix of variable
295 deletions. This strategy enables deletion analysis of a substantial DNA region
296 with both coarse (inter-target deletions) and fine (a variety of single- or double-
297 site deletions) resolution. Below, we prove the power of CAMIO in structure-
298 functional analysis of a 2.2kb-long genomic fragment.

Structure-functional analysis of *chinmo* 5'UTR

The Chinmo BTB-zinc finger nuclear protein is dynamically expressed in intricate spatiotemporal patterns in the developing *Drosophila* central nervous system. Such dynamic Chinmo expression governs various aspects of temporally patterned neurogenesis, including age-dependent neural stem cell proliferation^{37,38} and birth order-dependent neuronal cell fate^{29,39,40}. Notably, *chinmo* transcripts exist much more broadly than Chinmo proteins, indicating involvement of negative post-transcriptional regulation^{29,38}. Consistent with this notion, *chinmo* transcripts have long UTRs, including a 2.2kb 5'UTR and an 8.5kb 3'UTR⁴¹. To locate the involved regulatory elements in such long UTRs by conventional structure-functional analysis would be a daunting task.

Nonetheless, we started by making reporter transgenes carrying *chinmo* 5' and/or 3' UTR(s). For a functional readout, we utilized the development of the *Drosophila* mushroom body (MB), which involves an orderly production of γ , $\alpha'\beta'$, and $\alpha\beta$ neurons. We first determined the roles of the 5' vs. 3'UTR in downregulation of Chinmo expression along MB neurogenesis²⁹, by examining the change in expression of reporter transgenes (containing either or both UTRs) from early to late larval stages (Supplementary Fig. 2). Notably, presence of the *chinmo* 5'UTR drastically suppressed the reporter expression. Interestingly, only in the absence of the 3'UTR did we detect an enhanced 5'UTR-dependent

suppression at the late larval stage. These phenomena ascribe the *chinmo* downregulatory mechanism(s) to the 5'UTR, and unexpectedly revealed some upregulation by the 3'UTR. This upregulation could potentially be a transgene-specific artifact, arguing for the importance of performing assessments in the native environment. We thus turned to CAMIO to carry out structure-functional analysis of the native *chinmo* 5'UTR.

chinmo's 5'UTR is separated into 3 exons; the first two exons are neighboring and the distant 3rd exon is separated from the 2nd by 36kb (Fig. 3a). A gRNA set was designed to target each exon for CRISPR mutagenesis (Fig. 3b). We provided two copies of the same set for induction of both indels and inter-target deletions within an exon. Additionally, we paired gRNA sets for exon 1 and 2 to create larger deletions that span the exon1/2 junction. Thus, we could delete various parts of *chinmo* 5'UTR in its endogenous locus with simple fly pushing.

Based on the previously observed deletion rate of around 35%, we collected 300 male progeny from each CAMIO genetic cross. We hoped to saturate each 5'UTR exon with ~100 different deletions. In total, 1200 CAMIO males were harvested from the four different gRNA array combinations (exons 1, 2, 3, and 1+2). We mapped potential indels by sequencing indexed PCR products in a high-throughput manner.

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 $\alpha\beta$ 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 $\alpha'\beta'$ to $\alpha\beta$ MB neuron fate transition at the prepupal stage⁴². 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, $\alpha\beta$ and γ lobes) and Trio (red, $\alpha'\beta'$ and γ lobes). Missing or misshapen $\alpha\beta$ lobes (arrows). Scale bars: 50 μ m. (B) Percent of flies with abnormal $\alpha\beta$ lobes. Homozygous and transheterozygous deletions have strong MB $\alpha\beta$ 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 μ m.

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

Two innovations synergistically enable CAMIO, a germline pipeline for arrayed CRISPR mutagenesis. First, the *bamP* promoter can specifically limit Cas9 endonuclease activity to individual male germ cells—thus individual offspring receive independent mutations. Second, the random-choice gRNA arrays provide extensive coverage for deletion analysis, with both small indels and large deletions. Hence, the combination of *bamP*-Cas9 and gRNA arrays used in CAMIO enables *in vivo* targeted deletion analysis with both minimal molecular biology and minimal fly pushing.

We were happily surprised to discover a much higher CRISPR mutagenesis rate in male compared to female germ cells using *bamP*. This sex difference was also observed in CRISPR-induced gene targeting in our effort to improve Golic+ (manuscript in preparation). Currently, we do not know what leads to this phenomenon. *bamP* has a striking similar expression pattern in both the female and male germline: absence in GSCs and an early onset of expression during the four incomplete mitoses that produce the 16-cell germline cysts⁴³. One possibility is that the *bamP* activity is higher in the male than the female germline. Another possibility has to do with the sex differences in meiotic recombination—meiotic recombination does not occur in male *Drosophila*. Perhaps reduced access to homologous chromosomes as templates for homology-mediated repair favors indels. In any case, this feature allows us to utilize *bamP* to build a high-efficiency pipeline for targeted CRISPR mutagenesis in male germ cells.

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 MARCM⁴⁴. 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 testes⁴⁵, 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.

525

526 **Methods**

527

528 **Fly strains.**

529 We used the following fly strains in this work: (1) *bamP*(898)-Cas9 in *attP2*; (2)
 530 *U6:3-gRNA-e*¹⁹; (3) *Df(3R)ED10838/TM2* (BDSC #9485); (4) *dU6-3-gRNA-shi* in
 531 *attP40*; (5) *UAS-shibire* in *attP2*; (6) *GMR-Gal4*; (7) *nos-phiC31-nls* #12; (8)
 532 *10XUAS-mCD8::GFP* in *attP40*⁴⁷; (9) *act5C-Gal4/TM6B*; (10) *pCis-*
 533 *{4gRNAs_mCD8}* P element insertion line #3, #10 on II, and #9, #12 on III; (11)
 534 *13XLexAop2-5' UTR-smGFP-OLLAS* in VK00027, *13XLexAop2-smGFP-cMyc-*
 535 *3' UTR* in *attP40*, and *13XLexAop2-5' UTR-smGFP-V5-3' UTR* in *su(Hw)attP5*;
 536 (12) *41A10-KD*; (13) *DpnEE-KO-LexAp65*; (14) *pCis-{4gRNAs_chinmo Exon1}* P
 537 element insertion line #23, #25 on II, and #4, #5 on III; *pCis-{6gRNAs_chinmo*
 538 *Exon2}* #7, #11 on II, and #14 on III; *pCis-{4gRNAs_chinmo Exon3}* #5 on II, and
 539 #2, #3 on III; (15) *FRT40A,UAS-mCD8::GFP,UAS-Cd2-Mir/CyO,Y*⁴⁴; (16)
 540 *OK107-Gal4*; (17) *UAS-Cas9*¹⁹; (18) *dU6_g2+3* in VK00027; (19) *Dpn-Cas9*
 541 (unpublished reagent); (20) *act5C-Cas9*¹⁹; (21) *13XLexAop2-53UTR-smGFP-V5-*
 542 *d*, *13XLexAop2-53UTR-smGFP-V5-bD*, and *13XLexAop2-53UTR-smGFP-V5-*
 543 *D23* in *attP40*.

544

545 **Molecular biology.**

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

codon-optimized *shibire* coding sequence that carries the gRNA-shi target site was ordered from GeneArt gene synthesis, and then cloned into pJFRC28⁴⁸. To generate dU6-3-gRNAs, we replaced 10XUAS-IVS-GFP-p10 of pJFRC28 with dU6-3 promoter and gRNA scaffold fragment from pTL2²⁸. For dU6-3-gRNA-shi, GTATGGGGTATCAAGCCGAT was selected as the spacer. To create dU6_g2+3, we first generated dU6-3-g2_2 and dU6-3-g2_3 separately, and cloned dU6-3-g2_2 into the backbone of dU6-3-g2_3.

To create conditional U6-gRNA set construct, pCis-{4gRNAs_mCD8}, a U6 promoter-AttB fragment was synthesized by PCR amplification from pCFD3¹⁹ and cloned into pCaST-elav-VP16AD, which contained the p-Element inverted repeats (Addgene, #15308). Then, we inserted a DNA fragment (Genscript) containing 4 different gRNAs targeting the mCD8 protein tag. These gRNAs were selected based on their ON and OFF target scores (Benchling). Each of these gRNAs was preceded by an AttP site and a HammerHead ribozyme⁴⁹. Finally, a 3Xp3-RFP-polyA(α -tubulin) fragment was synthesized by PCR amplification, using pure genomic DNA from a fly line in which this cassette was used as a marker (Bloomington, #54590). Then, this fragment was inserted upstream of this gRNA region. In the final construct, the AttB and AttP sites were separated by a 3.7 Kb region containing an ampicillin resistance gene, an origin of replication in bacteria and the 3Xp3-RFP-polyA marker.

pCis-{gRNAs_chinmo-Exon1/Exon2/Exon3}: following the same design described above, a DNA fragment was synthesized (Genscript), which contained 4 gRNAs (6 for Exon2) either targeting the corresponding exon or the exon-intron junction. This fragment was then inserted into pCis-{4gRNAs_mCD8}, thus removing the previous gRNAs cassette.

The *Chinmo* UTRs were amplified from *Drosophila* genomic DNA. smGFP⁵⁰ fused to V5, cmc or ollas were amplified from previously existing plasmids. Standard molecular biology techniques were used to clone the smGFP fusions containing one or two *Chinmo* UTRs into 13XLexAop2 (pJFRC15,⁴⁷). If the construct ended with *Chinmo* 3'UTR, the SV40 signal was removed from the vector backbone. 13XLexAop2-5' UTR-GFP-3' UTR was further modified to create d, bD, and D23 reporters containing various deletions in the 5' UTR.

***Drosophila* genetics.**

***ebony* and UAS-shibire mutagenesis by *bamP(898)*-Cas9.** Female or male founders (*bamP(898)*-Cas9/U6:3-gRNA-e) were mated to *Df(3R)ED10838/TM6B*, and chromosomes over *Df(3R)ED10838* were scored for *ebony* loss-of-function phenotype. In total, 709 progeny from 5 female founders and 1574 progeny from 9 male founders were collected and phenotype was assessed. Male *UAS-shibire* mutagenesis founders were crossed with *GMR-Gal4*, and the wildtype-eyed 919 progeny were sacrificed for next generation sequencing (NGS).

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 Bowtie2⁵² 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).

686 18 Gratz, S. J. *et al.* Genome engineering of *Drosophila* with the CRISPR RNA-
687 guided Cas9 nuclease. *Genetics* **194**, 1029-1035 (2013).

688 19 Port, F., Chen, H. M., Lee, T. & Bullock, S. L. Optimized CRISPR/Cas tools for
689 efficient germline and somatic genome engineering in *Drosophila*.
690 *Proceedings of the National Academy of Sciences of the United States of*
691 *America* **111**, E2967-2976, doi:10.1073/pnas.1405500111 (2014).

692 20 Komor, A. C., Badran, A. H. & Liu, D. R. CRISPR-Based Technologies for the
693 Manipulation of Eukaryotic Genomes. *Cell* **169**, 559,
694 doi:10.1016/j.cell.2017.04.005 (2017).

695 21 Lieber, M. R. The mechanism of double-strand DNA break repair by the
696 nonhomologous DNA end-joining pathway. *Annual review of biochemistry* **79**,
697 181-211, doi:10.1146/annurev.biochem.052308.093131 (2010).

698 22 San Filippo, J., Sung, P. & Klein, H. Mechanism of eukaryotic homologous
699 recombination. *Annual review of biochemistry* **77**, 229-257,
700 doi:10.1146/annurev.biochem.77.061306.125255 (2008).

701 23 Fellmann, C., Gowen, B. G., Lin, P. C., Doudna, J. A. & Corn, J. E. Cornerstones of
702 CRISPR-Cas in drug discovery and therapy. *Nature reviews. Drug discovery*
703 **16**, 89-100, doi:10.1038/nrd.2016.238 (2017).

704 24 Wang, H., La Russa, M. & Qi, L. S. CRISPR/Cas9 in Genome Editing and
705 Beyond. *Annual review of biochemistry* **85**, 227-264, doi:10.1146/annurev-
706 biochem-060815-014607 (2016).

707 25 Hales, K. G., Korey, C. A., Larracuente, A. M. & Roberts, D. M. Genetics on the
708 Fly: A Primer on the *Drosophila* Model System. *Genetics* **201**, 815-842,
709 doi:10.1534/genetics.115.183392 (2015).

710 26 Bier, E., Harrison, M. M., O'Connor-Giles, K. M. & Wildonger, J. Advances in
711 Engineering the Fly Genome with the CRISPR-Cas System. *Genetics* **208**, 1-18,
712 doi:10.1534/genetics.117.1113 (2018).

713 27 Chen, D. & McKearin, D. M. A discrete transcriptional silencer in the bam gene
714 determines asymmetric division of the *Drosophila* germline stem cell.
715 *Development* **130**, 1159-1170 (2003).

716 28 Chen, H. M., Huang, Y., Pfeiffer, B. D., Yao, X. & Lee, T. An enhanced gene
717 targeting toolkit for *Drosophila*: Golic+. *Genetics* **199**, 683-694,
718 doi:10.1534/genetics.114.173716 (2015).

719 29 Zhu, S. *et al.* Gradients of the *Drosophila* Chinmo BTB-zinc finger protein
720 govern neuronal temporal identity. *Cell* **127**, 409-422,
721 doi:10.1016/j.cell.2006.08.045 (2006).

722 30 Park, J., Lim, K., Kim, J. S. & Bae, S. Cas-analyzer: an online tool for assessing
723 genome editing results using NGS data. *Bioinformatics (Oxford, England)* **33**,
724 286-288, doi:10.1093/bioinformatics/btw561 (2017).

725 31 Allen, F. *et al.* Predicting the mutations generated by repair of Cas9-induced
726 double-strand breaks. *Nat Biotechnol*, doi:10.1038/nbt.4317 (2018).

727 32 Sfeir, A. & Symington, L. S. Microhomology-Mediated End Joining: A Back-up
728 Survival Mechanism or Dedicated Pathway? *Trends in biochemical sciences*
729 **40**, 701-714, doi:10.1016/j.tibs.2015.08.006 (2015).

730 33 Ferguson, S. M. & De Camilli, P. Dynamin, a membrane-remodelling GTPase.
731 *Nature reviews. Molecular cell biology* **13**, 75-88, doi:10.1038/nrm3266
732 (2012).

733 34 Kitamoto, T. Conditional modification of behavior in *Drosophila* by targeted
734 expression of a temperature-sensitive shibire allele in defined neurons.
735 *Journal of neurobiology* **47**, 81-92 (2001).

736 35 Lehmann, R. Germline stem cells: origin and destiny. *Cell stem cell* **10**, 729-
737 739, doi:10.1016/j.stem.2012.05.016 (2012).

738 36 Kanca, O., Caussinus, E., Denes, A. S., Percival-Smith, A. & Affolter, M. Raeppli:
739 a whole-tissue labeling tool for live imaging of *Drosophila* development.
740 *Development* **141**, 472-480, doi:10.1242/dev.102913 (2014).

741 37 Narbonne-Reveau, K. *et al.* Neural stem cell-encoded temporal patterning
742 delineates an early window of malignant susceptibility in *Drosophila*. *eLife* **5**,
743 doi:10.7554/eLife.13463 (2016).

744 38 Dillard, C., Narbonne-Reveau, K., Foppolo, S., Lanet, E. & Maurange, C. Two
745 distinct mechanisms silence chinmo in *Drosophila* neuroblasts and
746 neuroepithelial cells to limit their self-renewal. *Development* **145**,
747 doi:10.1242/dev.154534 (2018).

748 39 Kao, C. F., Yu, H. H., He, Y., Kao, J. C. & Lee, T. Hierarchical deployment of
749 factors regulating temporal fate in a diverse neuronal lineage of the
750 *Drosophila* central brain. *Neuron* **73**, 677-684,
751 doi:10.1016/j.neuron.2011.12.018 (2012).

752 40 Ren, Q. *et al.* Stem Cell-Intrinsic, Seven-up-Triggered Temporal Factor
753 Gradients Diversify Intermediate Neural Progenitors. *Current biology : CB* **27**,
754 1303-1313, doi:10.1016/j.cub.2017.03.047 (2017).

755 41 Wu, Y. C., Chen, C. H., Mercer, A. & Sokol, N. S. Let-7-complex microRNAs
756 regulate the temporal identity of *Drosophila* mushroom body neurons via
757 chinmo. *Dev Cell* **23**, 202-209, doi:10.1016/j.devcel.2012.05.013 (2012).

758 42 Marchetti, G. & Tavosan, G. Steroid Hormone Ecdysone Signaling Specifies
759 Mushroom Body Neuron Sequential Fate via Chinmo. *Current biology : CB* **27**,
760 3017-3024.e3014, doi:10.1016/j.cub.2017.08.037 (2017).

761 43 Fuller, M. T. & Spradling, A. C. Male and female *Drosophila* germline stem
762 cells: two versions of immortality. *Science* **316**, 402-404,
763 doi:10.1126/science.1140861 (2007).

764 44 Yu, H.-H., Chen, C.-H., Shi, L., Huang, Y. & Lee, T. Twin-spot MARCM to reveal
765 the developmental origin and identity of neurons. *Nature neuroscience* **12**,
766 947-953 (2009).

767 45 Flaherty, M. S. *et al.* chinmo is a functional effector of the JAK/STAT pathway
768 that regulates eye development, tumor formation, and stem cell self-renewal
769 in *Drosophila*. *Dev Cell* **18**, 556-568, doi:10.1016/j.devcel.2010.02.006
770 (2010).

771 46 Tang, H., Ross, A. & Capel, B. Expression and functional analysis of Gm114, a
772 putative mammalian ortholog of *Drosophila* bam. *Dev Biol* **318**, 73-81,
773 doi:10.1016/j.ydbio.2008.03.001 (2008).

774 47 Pfeiffer, B. D. *et al.* Refinement of tools for targeted gene expression in
775 *Drosophila*. *Genetics* **186**, 735-755 (2010).

776 48 Pfeiffer, B. D., Truman, J. W. & Rubin, G. M. Using translational enhancers to
777 increase transgene expression in *Drosophila*. *Proceedings of the National*
778 *Academy of Sciences* **109**, 6626-6631 (2012).
779 49 He, Y. *et al.* Self-cleaving ribozymes enable the production of guide RNAs
780 from unlimited choices of promoters for CRISPR/Cas9 mediated genome
781 editing. *Journal of genetics and genomics = Yi chuan xue bao* **44**, 469-472,
782 doi:10.1016/j.jgg.2017.08.003 (2017).
783 50 Viswanathan, S. *et al.* High-performance probes for light and electron
784 microscopy. *Nat Methods* **12**, 568-576, doi:10.1038/nmeth.3365 (2015).
785 51 Awasaki, T. *et al.* The *Drosophila* trio plays an essential role in patterning of
786 axons by regulating their directional extension. *Neuron* **26**, 119-131 (2000).
787 52 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat*
788 *Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).
789

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 $\pm$ 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.net³⁰. (E) Left, percentage of predicted indels (calculated by FORECasT³¹ 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

homozygous viable deletions (marked in green) that cover larger proportions of each exons were selected for MB development studies.

Figure 4. Overlapping deletions in 5'UTR alter Chinmo expression and MB morphology. (A) Stacked confocal images of adult MBs stained for FasII (green, $\alpha\beta$ and γ lobes) and Trio (red, $\alpha'\beta'$ and γ lobes). Missing or misshapen $\alpha\beta$ lobes (arrows). Scale bars: 50 μm . (B) Percent of flies with abnormal $\alpha\beta$ lobes.

Homozygous and transheterozygous deletions have strong MB $\alpha\beta$ 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 .

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 μ m.

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 μ m.

914 **Supplementary figures**

915

916

917 **Supplementary Figure 1. Targeted mutagenesis of mCD8::GFP with pCis-**
918 **{4gRNAs_mCD8}.** pCis-{4gRNAs_mCD8} was created to target four different
919 sites along the mCD8 coding sequence of UAS-mCD8::GFP. We analyzed 30
920 GFP-negative progeny from two different male founders. Each GFP-negative
921 progeny carried a indel located at one of the four target sites. The location, size,
922 and overall counts of these 30 indels are summarized here.
923

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 μ m.

937

938 **Supplementary Figure 3. The small indels created by CAMIO in *chinmo***
 939 **Exon 1, 2, and 3.** We sequenced 984 progeny (300 each for individual exons,
 940 and additional 84 for Ex1-Ex2) from our CAMIO on *chinmo* 5' UTR experiment.
 941 The size, location, and corresponding stock ID of the predicted deletions (marked
 942 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 μ m.