The nanoCUT&RUN technique visualizes telomeric chromatin in Drosophila

Tao Chen[1,5,§], Xiaolu Wei[2,§], Cécile Courret[3,§], Min Cui[1,5], Lin Cheng[1,5], Jing Wu[5], Kami Ahmad[4,¶], Amanda M. Larracuente[3,¶], Yikang S. Rong[5,¶]

1: Sun Yat-sen University, Guangzhou, China

2: Department of Biomedical Genetics, University of Rochester Medical Center, Rochester, NY, USA

3: Department of Biology, University of Rochester, Rochester, NY, USA

4: Fred Hutchinson Cancer Research Center, Seattle, WA, USA

5: University of South China, Hengyang, China

§: contributed equally

¶: corresponding authors, kahmad@fredhutch.org (KA); alarracu@bio.rochester.edu (AML); zdqr03@yahoo.com (YSR).

ORCID XW 0000-0001-9952-3757
ORCID CC 0000-0001-5849-8014
ORCID AL 0000-0001-5944-5686
ORCID KA 0000-0001-8572-6182
ORCID YR 0000-0002-9787-9669

Abstract

Advances in genomic technology led to a more focused pattern for the distribution of chromosomal proteins and a better understanding of their functions. The recent development of the CUT&RUN technique marks one of the important such advances. Here we develop a modified CUT&RUN technique that we termed nanoCUT&RUN, in which a high affinity nanobody to GFP is used to bring micrococcal nuclease to the binding sites of GFP-tagged chromatin proteins. Subsequent activation of the nuclease cleaves the chromatin, and sequencing of released DNA identifies binding sites. We show that nanoCUT&RUN efficiently produces high quality data for the TRL transcription factor in Drosophila embryos, and distinguishes binding sites specific between two TRL isoforms. We further show that nanoCUT&RUN dissects the distributions of the HipHop and HOAP telomere capping proteins, and uncovers unexpected binding of telomeric proteins at centromeres. nanoCUT&RUN can be readily applied to any system in which a chromatin protein of interest, or its isoforms, carries the GFP tag.

Introduction

Telomeres protect the natural ends of linear chromosomes from being recognized as DNA breaks. In most organisms studied, chromosome ends are elongated by the enzyme telomerase using an RNA template. Telomerase-synthesized DNA repeats serve as binding sites for sequence specific binding proteins essential for end protection [reviewed in Fulcher et al. 2014]. However, in many organisms such as the model Drosophila and particularly Dipteran insects, either telomerase is missing, or it is missing a conserved domain necessary for high processivity [Fujiwara et al. 2005; Osanai et al. 2006; Mason et al. 2016]. In these organisms, retrotransposons or other sequences populate the ends of chromosomes. Despite possessing vastly different end sequences, at least some of these "telomerase-less" systems rely on a reverse transcription-based mechanism for end elongation. In addition, telomere-specific binding proteins have been identified, at least in Drosophila, that serve similar end protection functions as the sequence-specific binding proteins in the telomerase-containing systems [e.g., Cenci et al. 2003; Gao et al. 2010; Raffa et al. 2010; Zhang et al. 2016]. Among telomere-binding proteins is a class of so-called capping proteins that, when missing, renders chromosomes susceptible to end-to-end fusion. Drosophila capping proteins have been collectively called "Terminin" [Raffa et al. 2010], similar to the concept of "Shelterin" proposed for telomerase-maintained systems [de Lange 2005]. How capping proteins protect chromosome ends remains one of the major research topics in the field of telomere and genome maintenance.

In Drosophila, how capping proteins are recruited to telomeres remains obscure. It has been known for over 30 years that a Drosophila chromosome can be stably inherited for generations without the presence of telomeric retrotransposons [Beaucher et al. 2012 and references therein]. More recent genomic analyses uncovered surprisingly frequent

events in which the entire transposon array is lost from one or more telomeres in natural populations [Kern and Begun 2008; McGurk et al. 2021]. Some Drosophila species appear to have lost the telomeric retrotransposons [Saint-Leandre et al. 2019]. Moreover, we showed that a DNA fragment from the non-telomeric *white* locus is occupied by the HipHop and HOAP capping proteins only when the gene is situated at the very end of a chromosome [Gao et al. 2010]. These results suggest that capping protein binding does not require a sequence component from Drosophila telomeres. However, the natural binding partners of Drosophila capping proteins remain the three classes of non-LTR retrotransposons that are specifically enriched at chromosome ends. Therefore, one cannot rule out the hypothesis that there are “preferred” binding sites on the transposons that the capping proteins rely on for proper localization. In addition, physical interaction between these transposons and the proteins that bind them have been proposed to drive the rapid evolution of Drosophila telomeres [Villasante et al. 2008; Lee et al. 2017]. Therefore, characterization of telomeric protein binding on endogenous chromosomes in a telomerase-less system are important for a better understanding of the biology and evolution of telomeres and their functions. Here we profile these binding sites for the first time in the *Drosophila melanogaster* model. We chose the recently developed “CUT&RUN” technique [Skene and Henikoff 2017], but with our own modifications.

In 2004, Laemmli and colleagues [Schmid et al. 2004] developed the Chromatin ImmunoCleavage (ChIC) technique in which the Micrococcal Nuclease (MNase) is brought to the vicinity of a target protein by an interaction between Protein A and the target bound antibody. Bound MNase, which had been purified from bacteria as a fusion protein with Protein A, is activated by the addition of calcium and cleaves DNA around the site of target protein binding. This principle of targeted cleavage was further explored by Skene and

Henikoff [2017] to achieve efficient separation of the cleaved fragments from the bulk of uncut chromatin, and when combined with second generation sequencing led to the Cleavage Under Targets and Release Using Nuclease (CUT&RUN) method.

In CUT&RUN, similar to other ChIP methods, the principal target specificity is determined by antigen-antibody interactions. Therefore, antibodies might have to be developed for every target protein, and in special cases for every isoform of interest from loci encoding multiple ones, such as the *mod(mdg4)* locus in Drosophila that encodes as many as 31 different isoforms [flybase.org]. These limitations could be overcome by using epitope tagging so that a single anti-Tag antibody could be used to profile different targets or isoforms of a single target. An added advantage of using a common anti-Tag antibody is that profiles of different targets/isoforms could be directly compared as long as their relative expression levels have been taken into consideration.

In our studies of Drosophila chromatin proteins, we often employed tagging with a Green Fluorescent Protein (GFP) [e.g., Schneiderman et al. 2012; Wesolowska et al. 2013; Tang et al. 2017; Huang et al. 2021] so that live imaging of the target proteins could be achieved as long as the GFP-tagged target had been proven functional in genetic rescuing experiments. In theory, an anti-GFP antibody could be used to conduct CUT&RUN profiling of any such target. However, our collective experiences with monoclonal anti-GFP antibodies have been unsatisfactory. Since single-domain anti-GFP nanobodies offer consistent performance in IP-related experiments [e.g., Cicconi et al. 2017; Donohoe et al. 2018], and can be readily purified from bacteria, we were prompted to develop a method generally applicable to profiling GFP-tagged chromatin proteins . In our modified scheme called “nanoCUT&RUN”, the nuclease recruitment is accomplished by the binding of a GFP-specific nanobody [Saerens et al. 2005], similarly expressed as a

113 fusion protein, to a GFP-tagged target protein produced *in vivo*. Using nanoCUT&RUN, we
114 were able to profile the binding of the known TRL transcription factor. We were also able to
115 reveal the distribution of telomere capping proteins on retro-elements from Drosophila
116 telomeres and unexpectedly at some of the centromeric regions.
117

Results and Discussion

Designing the nanoCUT&RUN method

In the original ChIC and CUT&RUN schemes, a bacterially expressed Protein A (ProA) fused to MNase was used to tether the nuclease at antibody-bound sites of specific chromatin proteins [Schmid et al, 2004; Skene and Henikoff, 2017]. In our nanoCUT&RUN design, a single domain antibody recognizing the GFP motif replaces the ProA moiety (Figure 1A). We chose the Vhh4 clone of nanobody, which has been widely used in studies as a way to accomplish specific protein-protein interactions *in vivo* [Saerens et al. 2005; Caussinus et al. 2011].

We therefore expressed and purified from bacteria a new fusion protein in which MNase was fused to the GFP nanobody (nGFPMNase). As shown in Figure S1A, we were able to achieve a reasonable purification of this reagent. We tested the function of this fusion protein in the following two ways. First, we showed that nGFPMNase binds GFP *in vitro*. This was done by loading non-denaturing gels with protein samples that contain both GFP and nGFPMNase and visualizing the running position of GFP under a UV light. As shown in Figure S1B, the combination of GFP and nGFPMNase retards the migration of GFP, indicating binding between the two proteins. A 1:1 molar ratio of GFP and nGFPMNase was sufficient to retard most if not all of the GFP molecules on a native gel (Figure S1B). We next tested the ability of nGFPMNase to digest DNA, and importantly whether such nuclease activity can be activated by the presence of calcium similar to the activity of the original MNase. We mixed purified plasmid DNA with nGFPMNase with or without calcium (Figure S1C). Plasmid DNA was digested to completion in the presence of calcium demonstrating that the nuclease activity is enhanced by calcium. Therefore, our bacterially produced nGFPMNase effectively binds GFP and cleaves DNA, providing a

suitable reagent for CUT&RUN profiling.

Chromatin profiling TRL isoforms with nanoCUT&RUN

As a proof of principle for nanoCUT&RUN, we chose the well-characterized GAGA factor encoded by the *Trithorax-like* (*Trl)* gene in Drosophila [Farkas et al. 1994]. We used two transgenic lines generated by the modERN project in which TRL is epitope-tagged at its C-terminus (flybase.net). In both lines, tagged TRL proteins are expressed from a Bac transgene carrying endogenous regulatory elements of *trl*. The '804' line produces a tagged 519 aa 'TRL-short' isoform, while the '811' line produces a tagged 567 and 611 aa 'TRL-long' isoforms. Both long and short isoforms carry a zinc-finger DNA binding domain and a BTB/POZ homodimerization domain, but differ by the length of a poly-glutamine-rich segment, which serves as the transcriptional activator [Vaquero et al. 2000]. TRL-short is expressed continuously, while expression of TRL-long begins in mid-embryogenesis [Benyajati et al 1997].

We collected 0-12 hour old embryos, isolated nuclei, and performed nanoCUT&RUN profiling on duplicate samples, including samples lacking any GFP tag as a specificity control (no-tag control). About 12-22 million paired end reads were sequenced for each sample and mapped to the dm6 assembly of the Drosophila genome (Table S1). Replicate chromatin profiles for TRL-short were highly correlated (Spearman's correlation 0.9), while those for TRL-long had a more moderate correlation (Spearman's correlation 0.66). All of these profiles show peaks at previously described TRL binding sites [O'Brien et al. 1995], for example at the promoters of the *Hsp70* genes (Figure 1B). Both TRL-short and TRL-long isoforms coincide at these promoters. Peak calling by MACS2 identified 8,332 for the TRL-long isoform (Table S2), and the TRL-short isoform was also present at these sites.

However, peak calling of TRL-short isoform profiles identified substantially more (11,438) binding sites (Table S3), suggesting that some sites preferentially bind the TRL-short isoform. Indeed, differential peak analysis identified 3,663 peaks significantly enriched for the TRL-short isoform relative to TRL-long binding (Table S4). Strikingly, many TRL-short-enriched peaks fall at the promoters of developmentally regulated genes that are expressed in embryos, such as *N, wg, aop, sog, hid, wor,* and *sna* (Figure 1C). TRL-short is more abundant than TRL-long in the early embryo [Benyajati et al. 1997], but other neighboring peaks show similar relative magnitudes for both isoforms, again implying that many sites preferentially bind TRL-short. While differential analysis scored 837 sites as specifically enriched for TRL-long, almost all of these entries fall in highly repetitive unmapped contigs (Table S4). We did not consider these sites further.

The two TRL isoforms both contain a common zinc-finger DNA binding domain that recognizes a 'GAGA' motif [Pedone et al. 1996], so why does TRL-short preferentially bind some sites? The consensus motif for TRL-short enriched sites is an extended 'GA' repeat (Figure 1D), consistent with oligomer binding of TRL [Espinas et al. 1999; van Steensel et al. 2003], and these are found precisely at many sites of TRL-short signal. In contrast, sites enriched for TRL-long tend to fall in heterochromatic regions and in transposon repeats, and show extended smears of signal across repetitive sequences. Motif analysis of these regions is dominated by common repeat sequences, but embedded consensus GA motifs are present. It is intriguing that TRL shifts from euchromatic binding sites to heterochromatic sites during mitosis [Platero et al. 1998], perhaps related to the preference of TRL-long for a distinct set of sequence contexts. These differences highlight the utility of epitope-tagging protein isoforms for chromatin profiling where isoform specific antibodies may not be available.

Profiling of telomere capping proteins with nanoCUT&RUN

We were encouraged by our initial success in profiling TRL with nanoCUT&RUN and proceeded to apply the method to telomeric factors that we have been studying. The HOAP and HipHop proteins function as a complex that is specifically enriched at all telomeres in Drosophila [Cenci et al. 2003; Gao et al. 2010]. Previously using a single telomere with defined sequences from the *white* gene, we showed that HipHop and HOAP, along with their interacting HP1 protein occupy a large telomeric domain from the very end of the chromosome [Gao et al. 2010]. This suggests that these proteins maintain a specialized chromatin structure at telomeres.

Although our prior study provided the first detailed picture of how these important capping proteins are distributed on telomeres, it nevertheless suffers two drawbacks. First, the previous results were derived from a traditional ChIP plus qPCR assay in which a limited number of primer pairs from the telomeric region (about 1 kb apart and covering 11 kb in total) were used, thus greatly limiting the resolution. Second, the natural binding sites of these proteins are telomeric transposons. It remains possible that they distribute differently at native telomeres. Furthermore, the specific enrichment of these proteins at telomeres was established based on immunostaining results. It is possible that they have minor but important localization at non-telomeric positions. Therefore, we set out to use nanoCUT&RUN to profile binding of HipHop and HOAP, taking advantage of two fly lines in which the proteins of interest are expressed with a GFP tag. For HOAP, we used a knock-in line in which the endogenous *cav* locus was tagged [Gao et al. 2011; Wesolowska et al. 2013]. For HipHop, we constructed a transgene in which the *hiphop* locus was tagged at the N terminus, and expressed from its endogenous regulatory

214 elements. We showed that this transgene is able to rescue early larval lethality of a *hiphop*
215 deletion mutation previously generated [Gao et al. 2011] suggesting that the GFP-HipHop
216 protein is functional. Similar to profiling of TRL we used 0-12 hr old embryos. We
217 performed nanoCUT&RUN profiling on GFP-HipHop, GFP-HOAP and a no-tag control with
218 two different digestion durations (2 and 15 mins). The different digestions yielded highly
219 similar results (Wild type Spearman's rho=0.94, P<10-16; HOAP is Spearman's rho=0.92,
220 P<10-16; and HipHop Spearman's rho=0.95, P<10-16), therefore we generated two
221 additional biological replicates for each protein with a 2-minute digestion duration. About
222 25-36 million paired end reads were sequenced for each sample and mapped to the
223 heterochromatin-enriched *D. melanogaster* genome assembly [Chang and Larracuente
224 2019].
225
226 *HipHop and HOAP binding sites are enriched with telomeric elements*
227 When reads from nanoCUT&RUN were plotted on genome assembly zoomed in on
228 telomeres, we clearly see an enrichment of both GFP-HipHop and GFP-HOAP at the
229 telomeric elements. For each protein, nanoCUT&RUN profiling between the replicates
230 generated consistent results (Figure S2 and Tables S5 and S6). Figure 2 shows such a
231 zoomed-in view of the telomeric region from chromosome *3R*. Similar enrichment was
232 observed for telomeres from all major chromosomes (Figure S3, results are consistent
233 between 2- and 15-minute digestion; Figure S4). Because the fly strains may differ in their
234 organization and sequence of telomeric regions, and we map to a reference genome
235 assembly, we sequenced a genomic DNA control for each fly strain expressing GFP-
236 HipHop or GFP-HOAP, and the no-tag control. The read distribution is shown in Figure S5.
237 We do see some variation in genomic background between the strains suggesting that

differences between HOAP and HipHop on telomere 3L are driven by structural variation (Figure S5B). However, the overall pattern of HipHop and HOAP enrichment at the telomeres is the same. When we specifically examined the top 20% most abundant and enriched transposable elements presented in HipHop and HOAP nanoCUT&RUN reads relative to the no-tag control samples, the three classes of telomeric elements [*HeT-A,* *TART* and *TAHRE* (*HTT*)] are the most enriched elements (Figure S6 and Table S7). Besides transposons, some simple repeats are also enriched in HipHop and HOAP reads (Table S8).

Similar distributions of HipHop and HOAP over telomeric elements

To uncover any preferred binding sites along the telomeric elements for the two proteins, we piled nanoCUT&RUN reads from either HipHop or HOAP on consensus retrotransposons from each of the three classes (*HeT-A*, *TART* and *TAHRE*) that we built based on previous annotations [Chang et al. 2019; McGurk et al. 2021]. As shown in Figure 3 and Figures S7 and S8, HipHop reads are distributed along the entire lengths of all three elements with a possible exception at a region about 6 kb in size in *TART-A* (Figure 3D). This region lies in the 5' UTR of *TART-A* [Abad et al. 2004a], just upstream of *orf1* that encodes the Gag protein. As shown in Figure S8H, this might be related to the fact that most of the *TART* elements are 5' truncated in this genetic background. Similarly, we observed a loss of HipHop and HOAP enrichment at around 9 kb of the *TART-B* element (Figure 3C), which is due to a deletion of this region in most of the copies in this genetic background (Figure S8G). On the most abundant telomeric element, *HeT-A*, HipHop reads are more or less evenly distributed along the entire length (Figure 3A) and this pattern is consistent among all *HeT-A* subfamilies (Figure S7). This suggests that

HipHop might not have a strongly preferred DNA sequence within each *HTT* element for binding. Consistently, we did not uncover any particularly strong motif(s) enriched among HipHop reads under the high confidence (irreproducible discovery rate<0.05; Table S5; see Methods) peaks.

The distribution of HOAP reads generally tracks those of HipHop. This pattern is also evident looking at the distribution of HipHop and HOAP along the parts of the telomere represented in our genome assembly (Figure 2 and Figures S3-S5). The enrichment extends beyond the HTT domain and into the sub-telomeric region >100 kb from the distal end of the chromosome assembly. HOAP is generally less enriched than HipHop at telomeric repeats (Figure 3 and Figures S7), with the exception of *TART-B* and *Het-A5*. We do not know if this reduction of HOAP occupancy is related to the two genes having different expression levels. Interestingly, along the consensus sequence of *TART-B,* HOAP appears as enriched as HipHop (Figure 3C). In addition, on the consensus *TAHRE* element, significant HOAP enrichment is limited to the very 3' end (Figure 3B). Given the fact that *TAHRE* is the least abundant of the three retrotransposons [Abda et al. 2004a, 2004b; George et al. 2006], it is possible that *TAHRE* polymorphisms among different strains alone could account for this observation, as most *TAHRE* elements are 5' truncated in the cav^{gfp} strain (Figure S8F). Therefore, we conclude that HipHop and HOAP share similar distribution patterns at the telomeres, and that they bind indiscriminately along the HTT elements without preferred binding sites.

Enrichment of telomeric proteins at centric heterochromatin

Unexpectedly, we detected an enrichment of telomere proteins on islands of repeats that correspond to the centromeres (Figure S2, S6, S9). The primary enrichment is of

HipHop on the 4th and *X* chromosome centromere (Figure 4). This pattern does not seem to be driven by any particular sequence within the centromere. Prior immunostaining studies localizing HipHop in relationship to telomeres were performed on (1) polytene cells from third instar larvae, or (2) mitotic cells with condensed chromosomes [Gao et al. 2010, 2011]. The centromeric regions are under-replicated in polytene cells [Rudkin 1969; Gall et al. 1971]. Mitotic centromeres might have poor accessibility to antibodies making it difficult to detect weaker signals than those at telomeres. These factors might have prevented us from detecting centromeric HipHop cytologically.

The *X* and 4th chromosomes are both acrocentric, having a very short arm so that centric heterochromatin is relatively close to a telomere. Whether this common feature leads to the enrichment of HipHop at their centromeres in particular requires further investigations. One could imagine that the spreading of HipHop-enriched telomeric chromatin might encroach the centromeres of acrocentric chromosomes. However, we consider “spreading” an unlikely mechanism for the appearance of HipHop on centromeric regions of acrocentric chromosomes. Although the centromeres and telomeres of chromosomes *X* and *4* appear “close” in cytological images, the physical distances are in the megabases range for the *X* chromosome. It is difficult to envision that HipHop-enriched chromatin could spread for thousands of kilobases from the end of the *X* chromosome.

Interestingly, our recent results showed that a hypomorphic *hiphop* mutation behaves as a recessive suppressor of heterochromatin-induced Position-Effect-Variation (PEV) [Cui et al. 2021]. Moreover, the specific case of PEV used in that study involves the *X* centric heterochromatin. These earlier results seem to be consistent with the presence of HipHop in centric heterochromatin as revealed by this study, and with a potential role of HipHop in silencing that is not limited to chromosome ends.

Concluding remarks

Here we developed the nanoCUT&RUN technique that could be a powerful addition to the series of improvements/extensions to the original CUT&RUN design. An advantage of nanoCUT&RUN is that it is readily applicable to any protein tagged with GFP. In model organisms with facile genetics, the normal function of the GFP-tagged proteins can be routinely verified by testing their effects on rescuing mutant phenotypes, thus providing additional confidence for the nanoCUT&RUN data. In addition, the nanoCUT&RUN method is advantageous when one's goal involves comparing different target proteins or isoforms of a single target, since all profiling using our method is based on the same nanobody-GFP interaction. While we were preparing our manuscript, Koidl and Timmers [2021] reported the "greenCut&Run" method in mammalian cells, which is based on an identical principle as nanoCUT&RUN.

We confirmed the feasibility of this approach using the well-studied TRL transcription factor, and demonstrated a useful application of this approach by profiling two telomere capping proteins in Drosophila for the first time. Our results confirm that telomeric capping in Drosophila is unlikely to require a specific DNA sequence at chromosome ends. In addition, we uncovered an enrichment of HipHop at centromeric regions, which seems consistent with prior phenotypic analyses of *hiphop* mutants.

Acknowledgements

The research has been supported by grants from the National Key R&D Program of China (2018YFA0107000) and NSFC (31371364) to YSR, National Institutes of Health R01HG010492 to KA and NIH R35GM119515 and NSF MCB 1844693 to AML, and an

334 Agnes M. Messersmith and George Messersmith Dissertation Fellowship to XW.

Materials and Methods

Drosophila stocks and genetics

Drosophila stocks were raised on standard cornmeal-based food and kept at a 25℃ incubator with constant lighting. Two transgenic lines: BL64804 (flybase genotype: w[1118]; P{y[+7.7] w[+mC]=Trl.BCDEH-GFP.FPTB}attP40) and BL64811 (flybase genotype: y[1] w[*]; P{y[+7.7] w[+mC]=Trl.IJ-GFP.FPTB}attP40) were obtained from the Bloomington Drosophila Stock Center. Each carries an insertion of a BAC clone in which one of the isoforms of the TRL protein is tagged with GFP (flybase.net). The cav^{gfp} allele encoding GFP-tagged HOAP proteins was described and characterized in Gao et al. 2011 and Wesolowska et al. 2013, and it was used in nanoCUT&RUN profiling of HOAP. The $hiphop^{L41}$ deletion allele of the *hiphop* gene was described in Gao et al. 2011. To generate a *gfp*-tagged *hiphop* transgene, a *hiphop* genomic clone previously used by Gao el al. 2011 was modified by inserting a *gfp* gene at the N-terminus of *hiphop* by recombineering [Zhang et al. 2014]. This clone was inserted into the genome at the *attP* site carried by the *P{CaryP}attP40* element at position of 25C by phiC31 mediated site-specific integration. A stock that is homozygous for both the $hiphop^{gfp}$ transgene on chromosome *2* and the $hiphop^{L41}$ mutation on chromosome *3* was used in nanoCUT&RUN profiling of HipHop. The w^{1118} stock was used as the no-tag control.

Purification and characterization of the nGFPMNase protein

The coding regions for the MNase nuclease domain and an anti-GFP nanobody were codon-optimized for expression in *E.coli* and synthesized by IGE Biotech (Guangzhou, China). They were cloned into the pET28a vector so that nGFPMNase has a N-terminal

His6 tag. A map is included in Figure S1D, and sequence of the plasmid is available upon request. Bacterial expression and purification were performed using standard protocols with an IPTG concentration of 0.5mM for induction and an Imidazole concentration of 250mM for elution.

Embryo collection and nanoCUT&RUN

Drosophila strains were cultured at 25℃ on corn meal medium. Overnight (0-12h old) embryos were collected. They were washed off of grape juice-agar plates with Embryo Washing Buffer (EWB, 0.7% NaCl, 0.04% Triton-X100), dechorionated with 50% bleach, washed twice with EWB, and stored at -80℃ before use. To purify nuclei from embryos, about 30µl of embryos were suspended in 500µl of Buffer B (pH7.5, 15mM Tris-HCl, 15mM NaCl, 60mM KCl, 0.34M Sucrose, 0.5mM Spermidine, 0.1% β-mercaptoethanol, 0.25mM PMSF, 2mM EDTA, 0.5mM EGTA), and grinded with a pestle on ice. The grinded mixture was transferred to a 1.5ml Eppendorf tube and spun for 5min at 5000G and 4℃. The pellet was resuspended and washed with 500µl of Buffer A (pH7.5, 15mM Tris-HCl, 15mM NaCl, 60mM KCl, 0.34M Sucrose, 0.5mM Spermidine, 0.1% β-mercaptoethanol, 0.25mM PMSF). The wash was repeated twice, and the nuclei were resuspended in 600µl of WBSED buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.1% BSA, 2 mM EDTA, 0.5mM

376 Spermidine, 0.05% digitonin, 1X complete EDTA-free protease inhibitor from Roche).

377 NanoCUT&RUN was performed based a protocol for CUT&RUN
378 (dx.doi.org/10.17504/protocols.io.zcpf2vn) with the following modification: a 30µL volume
379 of starting embryos resulted in 600µL of nuclei suspended in WBSED to which 0.6µL of a
380 0.4mg/mL stock of nGFPMNase recombinant protein was added. The MNase released
381 DNA was suspended in 20µL of dH2O. A detailed nanoCUT&RUN protocol is provided in
382 Supplemental Methods.
383
384 **Sequencing and data analyses**
385 Libraries were sequenced for 150 cycles in paired-end mode on the Illumina Hiseq
386 X10 /Nova seq platform by AceGen (Guangzhou, China). Sequencing data have been
387 submitted to the NCBI short reads archive with the accession numbers of SUB9499573
388 and SUB11148972 (Table S1).

389 *Peak calling*

390 For telomere protein profiling, we trimmed the reads with Trim Galore
391 (https://github.com/FelixKrueger/TrimGalore/) (paired end default settings), and then
392 mapped the trimmed reads to a heterochromatin-enriched genome assembly [Chang and
393 Larracuente 2019] using bowtie2 [Langmead and Salzberg 2012]. We defined uniquely
394 mapped reads using samtools (v1.5) (-q 10). We called peaks using MACS version
395 2.1.1.20160309 [Zhang et al. 2008] (-q 0.01; hereafter referred to as MACS2 peaks).
396 For TRL profiling, we trimmed paired-end reads with Trim Galore within Galaxy
397 (usegalaxy.org) with default settings except hard-clipping 3 bp off 3' ends of reads.
398 Trimmed reads were mapped to the dm6 assembly with bowtie2 (-I 20 -X 1000, mate

399 dovetailing, one mate alignment to contain another, very sensitive end-to-end). We called
400 peaks using MACS2 Galaxy version 2.1.1.20160309.6 and differential enrichment of
401 peaks between TRL-long and TRL-short datasets using DiffBind Galaxy version 2.10.0.
402 Peak calls are provided in Tables S2 and S3.

403 *Irreproducible Discovery Rate analyses*

404 We performed an irreproducible discovery rate (IDR) analysis
405 (https://github.com/nboley/idr) to identify high confidence peaks that overlap between
406 replicates (IDR<0.05, represented by black dot in Figure S2A and C). We considered 2-
407 min and 15-min samples as replicates and ran the IDR analysis on the MACS2 peaks. The
408 localization of those high confidence peaks (Figure S2B and D) confirmed that the majority
409 of the telomere proteins are localized in telomeric regions. We further used STREME
410 (https://meme-suite.org/meme/tools/streme) to perform motif analysis with the fragment
411 sequences under the high confident peaks (IDR<0.05) of HipHop and HOAP (no-tag
412 samples as control, and 2-min and 15-min samples as replicates), and discovered no
413 specific motif enriched for HipHop or HOAP.

414 *Repeat analysis*

415 We performed analyses to determine the repeats enriched for HipHop and HOAP.
416 For complex repeats (e.g., complex satellite DNAs with repeat unit > 100bp, TEs), we
417 mapped trimmed reads to a heterochromatin-enriched genome assembly [Chang and
418 Larracuente 2019] using bowtie2 [Langmead and Salzberg 2012] (default settings), and
419 summarized read counts for each complex repeat using custom python scripts. We
420 normalized read count to the number of mapped reads and report RPM (Reads Per
421 Million). We calculated enrichment values as IP(RPM)/control(RPM), and considered a

repeat to be enriched only if it is in the top 20% of the IP RPM and also top 20% of the IP/control enrichment. For the HTT elements, we analyzed each subfamily [McGurk et al. 2021] separately (Table S7). Because the enrichment for subfamilies of *HeT-A* elements were similar, we combined onto a single consensus (Figure 3, Figure S6). For *TART* elements, *TART-A* and *TART-B* show different enrichment patterns, therefore we show both subfamilies (Figure 3, Figure S6). We also calculated the enrichment for the centromere islands as described previously (Figure S2) [Chang et al. 2019]. To determine which parts of HTT are represented in the enrichment, we examined the read pileup patterns along their consensus sequences. We used BLAST (v2.6.0) to map either reads matching HTT or genomic HTT variants (as a control) to the consensus dimer of the HTT, and then converted coordinates along a dimer to coordinates along a monomer consensus sequence.

For simple tandem repeats, we summarized overrepresented k-mers in the trimmed reads using kseek [Wei et al. 2014; https://github.com/weikevinhc/k-seek], and normalized the k-mer count to the number of mapped reads to the assembly and report the RPM value. We calculated the enrichment values as IP(RPM)/control(RPM), and considered the k-mers to be enriched if RPM>10 and enrichment value >1 in both replicates. Due to the repetitive nature of the elements enriched at telomeres, we used all the mapped reads including reads that have multiple mapping locations in the genome.

Data availability statement: Strains and plasmids are available upon request. Sequence reads are available on the NCBI short reads archive with the accession numbers SUB9499573 and SUB11148972, and code and data files for figures are https://github.com/LarracuenteLab/nanoCUTandRUN.github.repo. The authors affirm that

446 all data necessary for confirming the conclusions of the article are present within the

447 article, figures, and tables.

Figure legends

Figure 1. Chromatin profiling of TRL by nanoCUT&RUN

A: schematic of the method. Relative size of the nGFPMNase (magenta) to an antibody (brown) is shown. nGFPMNase binds to the GFP (green) tag of the protein of interest (red). In the presence of $Ca2^{+}$, MNase digests DNA (black line) that is not protected by the nucleosomes (yellow). **B**: landscapes of TRL isoform binding at three Hsp70 promoters. Genomic coordinates (in nt) and the scale of the *hsp70* region on *3R* are shown at the top. **C**: landscapes of preferential binding of TRL-short at the promoters of the *worniu* (*wor*) and *snail* (*sn*) genes, while an intergenic site binds both TRL isoforms (shown at the right end of the profile). Genomic coordinates and the scale of the region are shown at the top. **D**: consensus motifs of TRL-short sites.

Figure 2. Distribution of HipHop and HOAP on the telomere of chromosome *3R*

The y-axis represents the normalized enrichment of the target protein (HipHop or HOAP) or the no-tag control (WT) for replicate 1 in RPM. The gray lines correspond to multi-mapped reads, the black lines correspond to the uniquely mapped reads. The orange bars below each plot correspond to MACS2 peaks based on the uniquely mapping reads. The colored cytoband at the bottom of the plot shows the repeat organization. The color code is shown in the legend. The distribution of the two proteins on other telomeres are shown in Figures S3 and S4.

Figure 3. HipHop and HOAP distributions on telomeric retro-elements

Distribution of HipHop and HOAP on the consensus sequences of *TART-A, TART-B1, HeT-A,* and *TAHRE* elements. The y-axis represents the mean normalized enrichment

(in RPM) of the two replicates for each target protein (HipHop or HOAP) over the no-tag control.

Figure 4: Distributions of HipHop and HOAP on centromeres of the *X* and 4th chromosomes.

The y-axis represents the normalized enrichment of target protein (HipHop or HOAP) or the no-tag control (WT) for replicate 1 in RPM. The gray lines correspond to multi-mapping reads, the black lines correspond to the uniquely mapping reads. The orange bars below each plot correspond to MACS2 peaks based on the uniquely mapping reads. The colored cytoband at the bottom of the plot shows the repeat organization. The color code is shown in the legend.

References

Abad JP, De Pablos B, Osoegawa K, De Jong PJ, Martín-Gallardo A, Villasante A. Genomic analysis of Drosophila melanogaster telomeres: full-length copies of HeT-A and TART elements at telomeres. Mol Biol Evol. 2004a Sep;21(9):1613-9.

Abad JP, De Pablos B, Osoegawa K, De Jong PJ, Martín-Gallardo A, Villasante A. TAHRE, a novel telomeric retrotransposon from Drosophila melanogaster, reveals the origin of Drosophila telomeres. Mol Biol Evol. 2004b Sep;21(9):1620-4.

Beaucher M, Zheng XF, Amariei F, Rong YS. Multiple pathways suppress telomere addition to DNA breaks in the Drosophila germline. Genetics. 2012 Jun;191(2):407-17.

Benyajati C, Mueller L, Xu N, Pappano M, Gao J, Mosammaparast M, Conklin D, Granok H, Craig C, Elgin S. Multiple isoforms of GAGA factor, a critical component of chromatin structure. Nucleic Acids Res. 1997 Aug 15;25(16):3345-53.

Caussinus E, Kanca O, Affolter M. Fluorescent fusion protein knockout mediated by anti-GFP nanobody. Nat Struct Mol Biol. 2011 Dec 11;19(1):117-21.

Cenci G, Siriaco G, Raffa GD, Kellum R, Gatti M. The Drosophila HOAP protein is required for telomere capping. Nat Cell Biol. 2003 Jan;5(1):82-4.

Chang CH, Larracuente AM. Heterochromatin-Enriched Assemblies Reveal the Sequence and Organization of the Drosophila melanogaster Y Chromosome. Genetics. 2019 Jan;211(1):333-348.

Chang CH, Chavan A, Palladino J, Wei X, Martins NMC, Santinello B, Chen CC, Erceg J, Beliveau BJ, Wu CT, Larracuente AM, Mellone BG. Islands of retroelements are major components of Drosophila centromeres. PLoS Biol. 2019 May 14;17(5):e3000241.

Cicconi A, Micheli E, Vernì F, Jackson A, Gradilla AC, Cipressa F, Raimondo D, Bosso G, Wakefield JG, Ciapponi L, Cenci G, Gatti M, Cacchione S, Raffa GD. The Drosophila telomere-capping protein Verrocchio binds single-stranded DNA and protects telomeres from DNA damage response. Nucleic Acids Res. 2017 Apr 7;45(6):3068-3085.

Cui M, Bai Y, Li K, Rong YS. Taming active transposons at Drosophila telomeres: The interconnection between HipHop's roles in capping and transcriptional silencing. PLoS Genet. 2021 Nov 23;17(11):e1009925. doi: 10.1371/journal.pgen.1009925.

de Lange T. Shelterin: the protein complex that shapes and safeguards human telomeres. Genes Dev. 2005 Sep 15;19(18):2100-10. doi: 10.1101/gad.1346005.

Donohoe CD, Csordás G, Correia A, Jindra M, Klein C, Habermann B, Uhlirova M. Atf3 links loss of epithelial polarity to defects in cell differentiation and cytoarchitecture. PLoS Genet. 2018 Mar 1;14(3):e1007241.

Espinás ML, Jiménez-García E, Vaquero A, Canudas S, Bernués J, Azorín F. The N-terminal POZ domain of GAGA mediates the formation of oligomers that bind DNA with high affinity and specificity. J Biol Chem. 1999 Jun 4;274(23):16461-9.

Farkas G, Gausz J, Galloni M, Reuter G, Gyurkovics H, Karch F. The Trithorax-like gene encodes the Drosophila GAGA factor. Nature. 1994 Oct 27;371(6500):806-8.

Fujiwara H, Osanai M, Matsumoto T, Kojima KK. Telomere-specific non-LTR retrotransposons and telomere maintenance in the silkworm, Bombyx mori. Chromosome Res. 2005;13(5):455-67.

Fulcher N, Derboven E, Valuchova S, Riha K. If the cap fits, wear it: an overview of telomeric structures over evolution. Cell Mol Life Sci. 2014 Mar;71(5):847-65.

Gall JG, Cohen EH, Polan ML. Reptitive DNA sequences in drosophila. Chromosoma. 1971;33(3):319-44.

Gao G, Walser JC, Beaucher ML, Morciano P, Wesolowska N, Chen J, Rong YS. HipHop interacts with HOAP and HP1 to protect Drosophila telomeres in a sequence-independent manner. EMBO J. 2010 Feb 17;29(4):819-29.

Gao G, Cheng Y, Wesolowska N, Rong YS. Paternal imprint essential for the inheritance of telomere identity in Drosophila. Proc Natl Acad Sci U S A. 2011 Mar 22;108(12):4932-7.

George JA, DeBaryshe PG, Traverse KL, Celniker SE, Pardue ML. Genomic organization of the Drosophila telomere retrotransposable elements. Genome Res. 2006 Oct;16(10):1231-40.

Huang W, Liu Z, Rong YS. Dynamic localization of DNA topoisomerase I and its functional relevance during Drosophila development. G3 (Bethesda). 2021 Sep 6;11(9):jkab202.

Kern AD, Begun DJ. Recurrent deletion and gene presence/absence polymorphism: telomere dynamics dominate evolution at the tip of 3L in Drosophila melanogaster and D. simulans. Genetics. 2008 Jun;179(2):1021-7.

Koidl S, Timmers HTM. greenCUT&RUN: Efficient Genomic Profiling of GFP-Tagged Transcription Factors and Chromatin Regulators. Curr Protoc. 2021 Oct;1(10):e266.

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012 Mar 4;9(4):357-9. doi: 10.1038/nmeth.1923.

Lee YC, Leek C, Levine MT. Recurrent Innovation at Genes Required for Telomere Integrity in Drosophila. Mol Biol Evol. 2017 Feb 1;34(2):467-482.

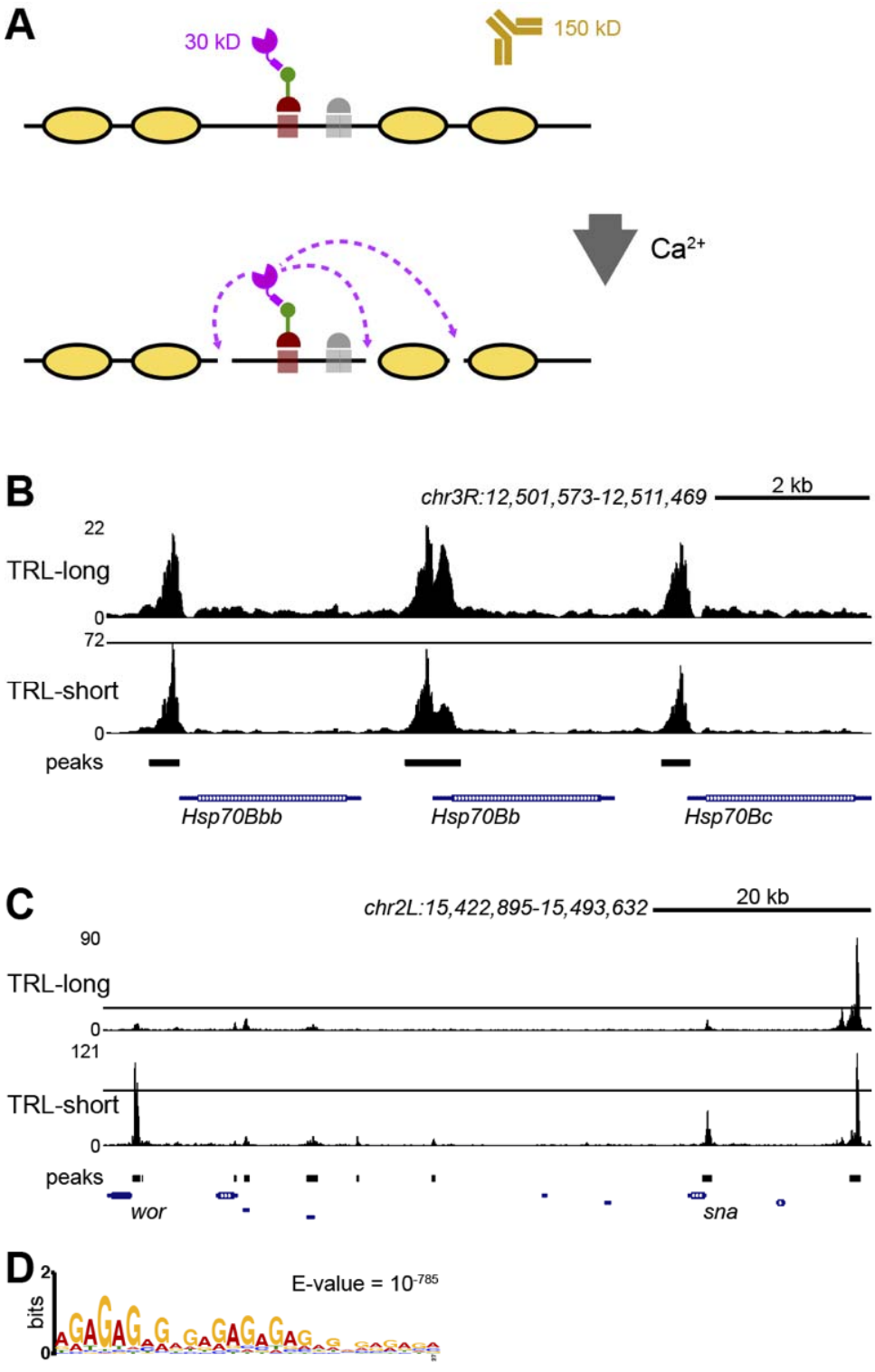
Li H. and Durbin R. (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics, 25:1754-60. [PMID: 19451168]

Mason JM, Randall TA, Capkova Frydrychova R. Telomerase lost? Chromosoma. 2016

579 Mar;125(1):65-73.
580
581 McGurk MP, Dion-Côté AM, Barbash DA. Rapid evolution at the Drosophila telomere:
582 transposable element dynamics at an intrinsically unstable locus. Genetics. 2021 Feb
583 9;217(2):iyaa027. doi: 10.1093/genetics/iyaa027.
584
585 Meers, M.P., Tenenbaum, D. & Henikoff, S. Peak calling by Sparse Enrichment Analysis
586 for CUT&RUN chromatin profiling. Epigenetics & Chromatin 12, 42 (2019).
587
588 O'Brien T, Wilkins RC, Giardina C, Lis JT. Distribution of GAGA protein on Drosophila
589 genes in vivo. Genes Dev. 1995 May 1;9(9):1098-110.
590
591 Osanai M, Kojima KK, Futahashi R, Yaguchi S, Fujiwara H. Identification and
592 characterization of the telomerase reverse transcriptase of Bombyx mori (silkworm) and
593 Tribolium castaneum (flour beetle). Gene. 2006 Jul 19;376(2):281-9.
594
595 Pedone PV, Ghirlando R, Clore GM, Gronenborn AM, Felsenfeld G, Omichinski JG. The
596 single Cys2-His2 zinc finger domain of the GAGA protein flanked by basic residues is
597 sufficient for high-affinity specific DNA binding. Proc Natl Acad Sci U S A. 1996 Apr
598 2;93(7):2822-6.
599
600 Platero JS, Csink AK, Quintanilla A, Henikoff S. Changes in chromosomal localization of
601 heterochromatin-binding proteins during the cell cycle in Drosophila. J Cell Biol. 1998 Mar
602 23;140(6):1297-306.
603
604 Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F,
605 Manke T. deepTools2: a next generation web server for deep-sequencing data analysis.
606 Nucleic Acids Research. 2016 Apr 13: gkw257.
607
608 Raffa GD, Raimondo D, Sorino C, Cugusi S, Cenci G, Cacchione S, et al. Verrocchio, a
609 Drosophila OB fold-containing protein, is a component of the terminin telomere-capping
610 complex. Genes Dev. 2010 Aug 1;24(15):1596-601. doi: 10.1101/gad.574810.
611
612 Rudkin GT. Non replicating DNA in Drosophila. Genetics. 1969;61(1):Suppl:227-38.
613
614 Saerens D, Pellis M, Loris R, Pardon E, Dumoulin M, Matagne A, Wyns L, Muyldermans
615 S, Conrath K. Identification of a universal VHH framework to graft non-canonical antigen-
616 binding loops of camel single-domain antibodies. J Mol Biol. 2005 Sep 23;352(3):597-607.
617
618 Saint-Leandre B, Nguyen SC, Levine MT. Diversification and collapse of a telomere
619 elongation mechanism. Genome Res. 2019 Jun;29(6):920-931. doi:
620 10.1101/gr.245001.118.
621
622 Schmid M, Durussel T, Laemmli UK. ChIC and ChEC; genomic mapping of chromatin
623 proteins. Mol Cell. 2004 Oct 8;16(1):147-57.
624
625 Schneiderman JI, Orsi GA, Hughes KT, Loppin B, Ahmad K. Nucleosome-depleted
626 chromatin gaps recruit assembly factors for the H3.3 histone variant. Proc Natl Acad Sci U

627 S A. 2012 Nov 27;109(48):19721-6.
628
629 Skene PJ, Henikoff S. An efficient targeted nuclease strategy for high-resolution mapping
630 of DNA binding sites. Elife. 2017 Jan 16;6:e21856.
631
632 Sproul JS, Khost DE, Eickbush DG, Negm S, Wei X, Wong I, Larracuente AM. Dynamic
633 Evolution of Euchromatic Satellites on the X Chromosome in Drosophila melanogaster and
634 the simulans Clade. Mol Biol Evol. 2020 Aug 1;37(8):2241-2256. doi:
635 10.1093/molbev/msaa078.
636
637 Tang X, Cao J, Zhang L, Huang Y, Zhang Q, Rong YS. Maternal Haploid, a
638 Metalloprotease Enriched at the Largest Satellite Repeat and Essential for Genome
639 Integrity in Drosophila Embryos. Genetics. 2017 Aug;206(4):1829-1839.
640
641 van Steensel B, Delrow J, Bussemaker HJ. Genomewide analysis of Drosophila GAGA
642 factor target genes reveals context-dependent DNA binding. Proc Natl Acad Sci U S A.
643 2003 Mar 4;100(5):2580-5.
644
645 Vaquero A, Espinás ML, Azorin F, Bernueś J. Functional mapping of the GAGA factor
646 assigns its transcriptional activity to the C-terminal glutamine-rich domain. J Biol Chem.
647 2000 Jun 30;275(26):19461-8.
648
649 Villasante A, de Pablos B, Méndez-Lago M, Abad JP. Telomere maintenance in
650 Drosophila: rapid transposon evolution at chromosome ends. Cell Cycle. 2008 Jul
651 15;7(14):2134-8.
652
653 Wei KH, Grenier JK, Barbash DA, Clark AG. Correlated variation and population
654 differentiation in satellite DNA abundance among lines of Drosophila melanogaster. Proc
655 Natl Acad Sci U S A. 2014 Dec 30;111(52):18793-8.
656
657 Wesolowska N, Amariei FL, Rong YS. Clustering and protein dynamics of Drosophila
658 melanogaster telomeres. Genetics. 2013 Oct;195(2):381-91.
659
660 Zhang Y, Schreiner W, Rong YS. Genome manipulations with bacterial recombineering
661 and site-specific integration in Drosophila. Methods Mol Biol. 2014;1114:11-24.
662
663 Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers
664 RM, Brown M, Li W, Liu XS. Model-based analysis of ChIP-Seq (MACS). Genome Biol.
665 2008;9(9):R137. doi: 10.1186/gb-2008-9-9-r137.
666
667 Zhang Y, Zhang L, Tang X, Bhardwaj SR, Ji J, Rong YS. MTV, an ssDNA Protecting
668 Complex Essential for Transposon-Based Telomere Maintenance in Drosophila. PLoS
669 Genet. 2016 Nov 11;12(11):e1006435. doi: 10.1371/journal.pgen.1006435.
670

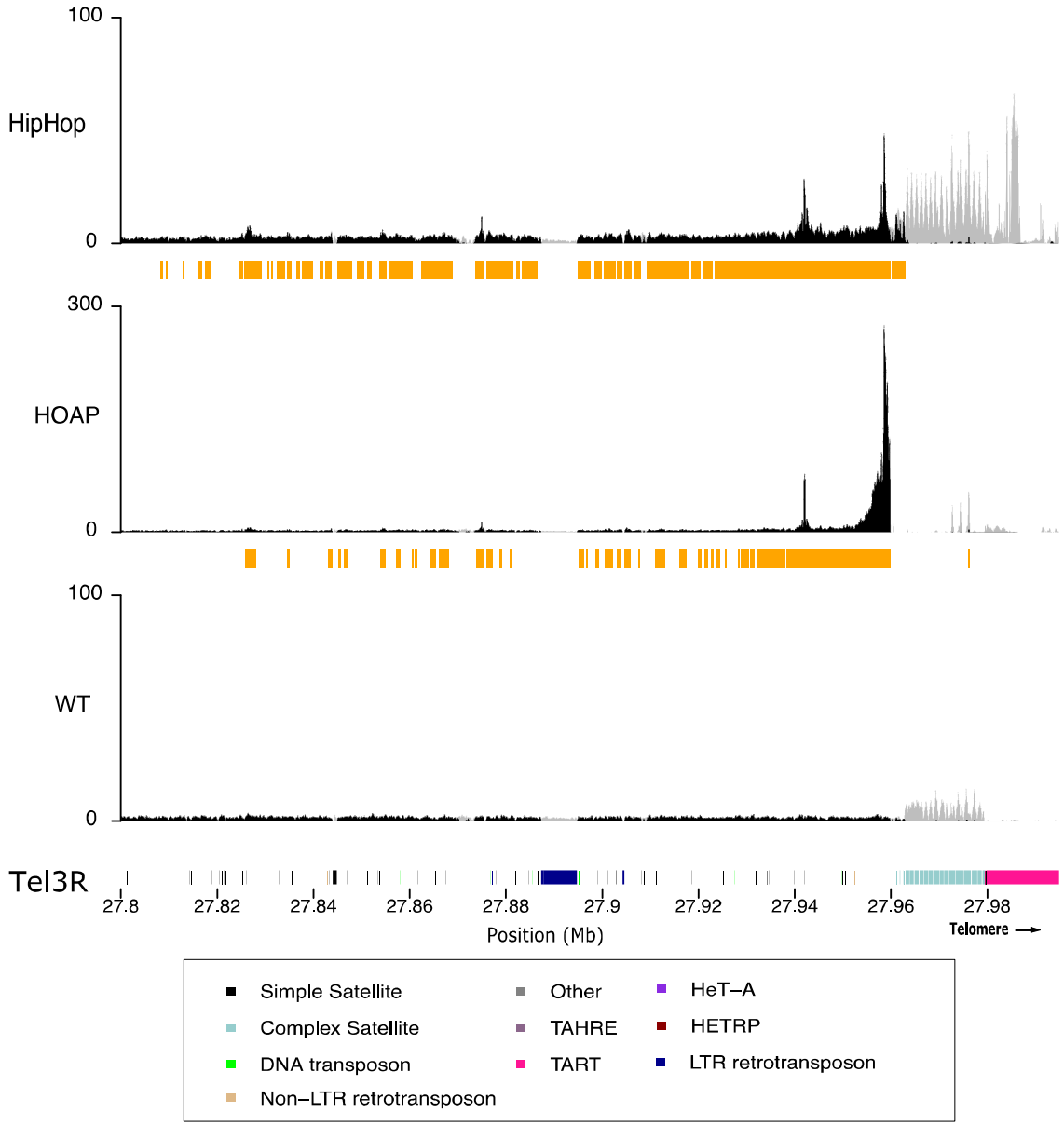
671 Figure 1



672

673

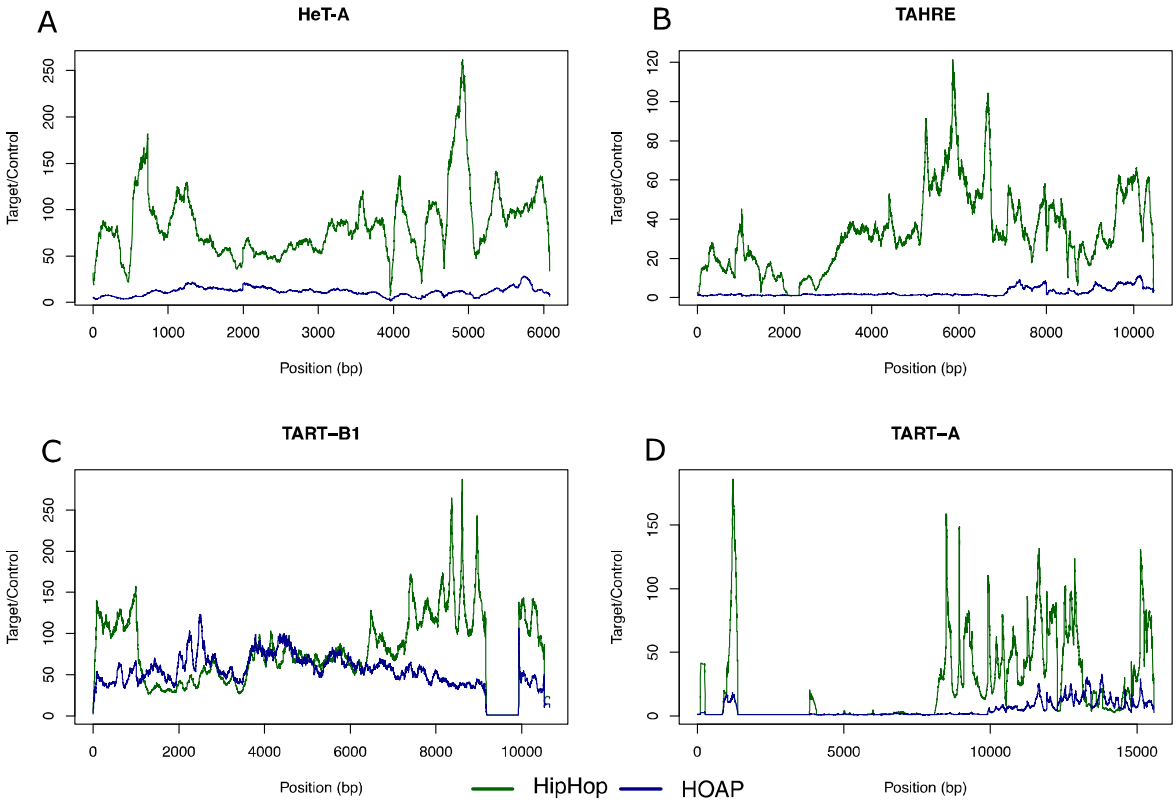
674 Figure 2



675

676

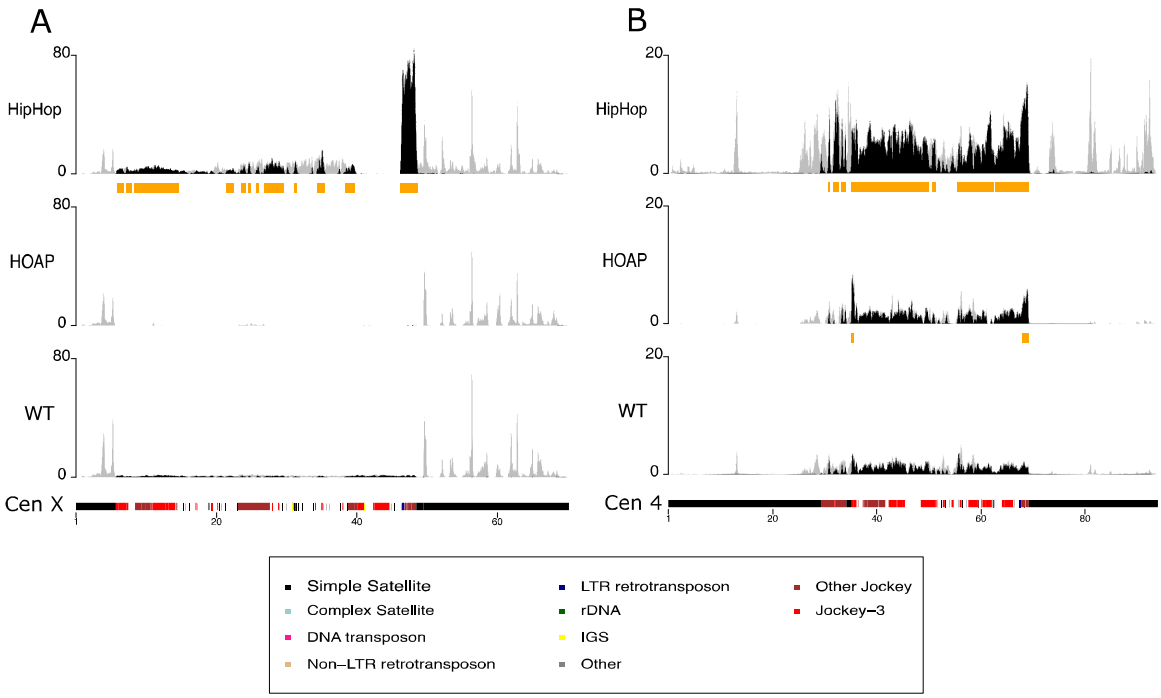
677 Figure 3



678

679

680 Figure 4



681

682

Supplemental information legends

Figure S1. Characterization of nGFPMNase

A: purification of the nGFPMNase fusion protein from bacteria. Extracts from different fractions were run on SDS-Page and stained with Coomassie Blue. Lanes 1: purified nGFPMNase; 2: insoluble fraction from bacteria overexpressing nGFPMNase; 3: soluble fraction; 4: total extract from overexpressing bacteria; 5: total extract from uninduced bacterial culture. “M” denotes protein markers with sizes in KD indicated to the right. The arrow marks the running position of nGFPMNase. **B**: nGFPMNase (nanoGM) binds GFP. GFP fluorescence from a native protein gel is shown with protein components loaded onto each lane shown at the top. Note that nGFPMNase alone does not emit fluorescence. The double star marks the running position of the complex between nGFPMNase and GFP. The single star marks the running position of GFP alone. **C**: nGFPMNase (nanoGM) digests DNA in the presence of calcium. Plasmid DNA was mixed with purified nGFPMNase in the nuclease digestion buffer with or without calcium. “M” denotes DNA markers with sizes indicated to the right. **D**: map of the nGFPMNase expression plasmid.

Figure S2. IDR analysis

Panels **A** and **C** represent the peak scores of replicate 1 *versus* replicate 2 on a log10 scale. The IDR analyses detected 1686 peaks in common between the two HipHop replicates but only 138 peaks passed the cutoff of IDR<0.05 (in black). The IDR analyses detected 307 peaks in common between the two HOAP replicates but only 58 peaks passed the cutoff of IDR<0.05 (in black). Panels **B** and **D** represent the localization of the peaks with an IDR <0.05. The majority of those peaks are localized on telomeres. We also detect a minority of peaks on the centromeres. All the peaks localized outside the

707 telomeres and centromeres are grouped in the category “Other”. However, in this category
708 most of the peaks actually localized on one Y-linked scaffold (Y_scaffold4), which is also
709 enriched in HTT, however this scaffold is unlikely to represent the Y telomere because of
710 its cytological location [Chang and Larracuente 2019].
711
712 **Figure S3. Distribution of HipHop and HOAP on the telomeres of chromosome *2R***
713 **(A), *2L* (B), *4* (C), *X* (D), and *3L* (E)**
714 The y-axis represents the normalized enrichment of target protein or the no-tag
715 control (WT) for replicate 1 in RPM. The gray lines correspond to multi-mapped reads, the
716 black lines correspond to the uniquely mapped reads. The orange bars below each plot
717 correspond to MACS2 peaks based on the uniquely mapping reads. The colored cytoband
718 at the bottom of the plot shows the repeat organization. The color code is shown in the
719 legend.
720
721 **Figure S4. Distribution of HipHop and HOAP on the telomeres of all chromosomes**
722 **after 2 and 15-minute digestion.**
723 The y-axis represents the normalized enrichment of target protein or the no-tag
724 control (WT) for two different digestion durations (2 or 15 min) in RPM . The gray lines
725 correspond to multi-mapped reads, the black lines correspond to the uniquely mapped
726 reads. The orange bars below each plot correspond to MACS2 peaks based on the
727 uniquely mapping reads. The colored cytoband at the bottom of the plot shows the repeat
728 organization. The color code is shown in the legend.
729 **Figure S5. Distribution of HipHop and HOAP and genomic DNA on all telomeres.**

The y-axis of the first plot (white background) represents the normalized enrichment (in RPM) for a second replicate of target protein (HipHop Rep2 and HOAP Rep2) or the no-tag control (WT Rep2). The y-axis of the second plot (yellow background) represents the normalized enrichment (in RPM) of the genomic DNA coverage of each strain (WG: whole genome). The gray lines correspond to multi-mapped reads, the black lines correspond to the uniquely mapped reads. The orange bars below each first plot correspond to MACS2 peaks based on the uniquely mapping reads. The colored cytoband at the bottom of the plot shows the repeat organization. The color code is shown in the legend.

Figure S6. Enrichment of HipHop and HOAP on complex repeats

The plot shows the normalized enrichment of target protein over the no-tag control (in RPM) for the top 10 repeats enriched in both HipHop and HOAP nanoCut&Run profiling. The full dataset is in Table S7.

Figure S7. HipHop and HOAP distributions on *HeT-A* subfamilies

Distribution of HipHop and HOAP on the individual subfamilies of *HeT-A* from [McGurk et al. 2021]. The y-axis represents the mean normalized enrichment (in RPM) of the two replicates for each target protein (HipHop or HOAP) over the no-tag control.

Figure S8. Genomic pileup on HTT elements

Genomic read coverage on HTT elements (*TART-A, TART-B, TAHRE, Het-A, Het-A 1D, Het-A 2, Het-A 3, Het-A 5*) of GFP-HipHop and GFP-HOAP strains. The y-axis represents the normalized reads coverage in RPM.

Figure S9. Distributions of HipHop and HOAP and genomic DNA on all centromeres.

The y-axis of the first plot (white background) represents the normalized enrichment (in RPM) for a second replicate of target protein (HipHop Rep2 and HOAP Rep2) or the no-tag control (WT Rep2). The y-axis of the second plot (yellow background) represents the normalized enrichment (in RPM) of the genomic DNA coverage of each strain (WG: whole genome). The gray lines correspond to multi-mapped reads, the black lines correspond to the uniquely mapped reads. The orange bars below each first plot correspond to MACS2 peaks based on the uniquely mapping reads. The colored cytoband at the bottom of the plot shows the repeat organization. The color code is shown in the legend.

Table S1. Summary of samples metrics

Summarized the total number of reads, the number of mapping reads and the accession number for each sample.

Table S2. TRL-long binding sites

Peaks defined by MACS2 for the TRL-long-GFP isoform.

Table S3. TRL-short binding sites

Peaks defined by MACS2 for the TRL-short-GFP isoform.

Table S4. Differentially bound TRL sites

Differential peaks defined by DiffBind with FDR<0.05.

Table S5. IDR peaks for HipHop

IDR output table based on MACS2 peak calling, for HipHop protein between the biological replicates. The score column represents the scaled IDR value and is used for IDR cutoff, peaks with IDR<0.05 have score>540.

Table S6. IDR peaks for HOAP

IDR output table based on MACS2 peak calling, for HOAP protein between the biological replicates. The score column represents the scaled IDR value and is used for IDR cutoff, peaks with IDR<0.05 have score>540.

Table S7. Enrichment scores for individual complex repeat and transposable element families

Normalized enrichment scores of the target protein over no-tag control (in RPM) for the complex satellite and transposable element enriched in HipHop and HOAP.

Table S8. Enrichment scores for simple repeats

Normalized enrichment scores of the target protein over no-tag control (in RPM) for the simple tandem repeats enriched in HipHop and HOAP.

Supplemental protocol

A detailed protocol for nanoCUT&RUN.