SUMM4 complex biochemically couples insulator function and DNA replication timing control

Evgeniya N. Andreyeva[1]†, Alexander V. Emelyanov[1]†, Markus Nevil[2], Lu Sun[3], Elena Vershilova[1], Christina A. Hill[4], Michael C. Keogh[3], Robert J. Duronio[4-7], Arthur I. Skoultchi[1], Dmitry V. Fyodorov[1]*

[1]Department of Cell Biology, Albert Einstein College of Medicine, Bronx, NY 10461, USA

[2]UNC-SPIRE, University of North Carolina, Durham, NC 275999, USA

[3]Epicypher, Inc., Durham, NC 27709, USA

[4]Integrative Program for Biological and Genome Sciences, University of North Carolina, Chapel Hill, NC, 27599 USA

[5]Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC, 27599 USA

[6]Department of Biology, University of North Carolina, Chapel Hill, NC, 27599 USA

[7]Department of Genetics, University of North Carolina, Chapel Hill, NC, 27599 USA

†These authors contributed equally to this work

*Corresponding author. Email: dmitry.fyodorov@einsteinmed.org

Abstract:

The mechanisms that establish DNA replication timing programs in eukaryotes remain incompletely understood. *Drosophila* SNF2-related factor SUUR imparts under-replication (UR) of late-replicating intercalary heterochromatin (IH) in polytene chromosomes. We developed a proteomics technique termed MERCI to isolate a native complex SUMM4 comprising SUUR and chromatin boundary protein Mod(Mdg4)-67.2. Mod(Mdg4) stimulates the ATPase activity of SUUR and is required for its normal spatiotemporal distribution *in vivo*. Both SUMM4 subunits mediate the activities of *gypsy* insulator disrupting enhancer-promoter interactions and establishing chromatin barriers. Furthermore, *SuUR* or *mod(mdg4)* mutations reverse UR of IH. Our findings uncover a critical role for architectural proteins in attenuating replication fork progression and suggest an alternative mechanism for DNA replication timing that does not depend on an asynchronous firing of replication origins.

One-Sentence Summary:

A stable protein complex comprising an insulator factor and a SNF2-like ATPase imparts late replication of heterochromatin.

Main Text:

Replication of metazoan genomes occurs according to a highly coordinated spatiotemporal program, where discrete chromosomal regions replicate at distinct times during S phase (*1*). The replication program follows the spatial organization of the genome in Megabase-long constant timing regions interspersed by timing transition regions (*2*). The spatiotemporal replication program exhibits correlations with the genetic activity, epigenetic marks and features of 3D genome architecture and sub-nuclear localization. Yet the reasons for these correlations remain obscure. Interestingly, the timing of firing for any individual origin of replication is established during G1 before pre-replicative complexes (pre-RC) are assembled and licensed at origins (*3*).

Most larval tissues of *Drosophila melanogaster* grow via G-S endoreplication cycles that duplicate DNA without cell division resulting in polyploidy (*4*). Endreplicated DNA molecules frequently align in register to form giant polytene chromosomes (*5*). Importantly, genomic domains corresponding to the latest replicated regions of dividing cells, specifically pericentric (PH) and intercalary (IH) heterochromatin, fail to complete endoreplication resulting in under-replication (UR). In both dividing and endoreplicating cells, these regions are devoid of sites for binding the Origin of Replication Complex (ORC) and thus, their replication must rely on forks progressing from external origins (*6*). Although cell cycle programs are dissimilar between endoreplicating and mitotically dividing cells, they share biochemically identical DNA replication machinery (*4*). Thus, UR provides a facile readout for late replication initiation and delayed fork progression. The *Suppressor of UR* (*SuUR*) gene is essential for polytene chromosome UR in IH and PH (*7*). It encodes a protein (SUUR) containing a helicase domain with a homology to that of the SNF2/SWI2 family. It has been shown that the occupancy of ORC is not increased in *SuUR* mutants (*6*). Rather, SUUR negatively regulates the rate of replication (*8*) by an unknown mechanism. It has been proposed (*9*) that the retardation of replisome by SUUR takes place via simultaneous physical association with the components of the fork (*e.g.*, CDC45 and PCNA) (*8, 10*) and repressive chromatin proteins, such as HP1a (*11*).

To determine how SUUR functions in replication control we sought to identify its native complex. Previous attempts to characterize the native form of SUUR by co-IP or tag-affinity purification gave rise to multiple putative binding partners (*8, 10-12*). However, evaluating whether any of these proteins are present in a native SUUR complex is problematic because of the low abundance of SUUR. It also precludes its purification by conventional chromatography. Therefore, we developed a novel biochemical approach that relies on partial purification by multi-step FPLC and shotgun proteomics of chromatographic fractions by quantitative LCMS (**Fig. 1A**). We term this technology MERCI for MS-Enabled Rapid protein Complex Identification (*Materials and Methods*).

The depth of proteomic quantification is limited by the range of peptides identified in the information-dependent acquisition (IDA), dubbed "ion library" (IL). Unfortunately, SUUR-specific peptides could not be found in ILs obtained from acquisitions of crude nuclear extracts or fractions from the first, phosphocellulose, step (IL1, **fig. S1A**, **Data S1**). Thus, to quantify SUUR in phosphocellulose fractions, we augmented IL1 with the IL obtained by acquisition of recombinant SUUR (ILR, **Figs. 1A**, **B**). In ILs from subsequent chromatographic steps, peptides derived from native SUUR were detected **fig. S1A**, **Data S1**) and used for quantification of the cognate data-independent acquisitions (DIA/SWATH) (**Fig. 1C**).

The final aspect of the MERCI algorithm calls for re-quantification of FPLC fraction SWATH acquisitions with an IL from the last step (IL5) enriched for peptides derived from SUUR and co-purifying polypeptides (**Fig. 1A**) and includes only 140 proteins (**fig. S1A**, **Data S1**). In this fashion, scarce polypeptides (including SUUR and, potentially, subunits of its putative complex) that may not be detectable in earlier steps will not evade quantification. Purification profiles of proteins quantified in all five FPLC steps (132) were then artificially stitched into 83-point arrays of Z-scores (**Fig. 1D**, **Data S2**). These profiles were Pearson-correlated with that of SUUR and ranked down from the highest Pearson coefficient, PCC (**Fig. 1E**). Whereas the PCC numbers for the bottom 130 proteins lay on a smooth

curve, the top two proteins, SUUR (PCC = 1.000) and Mod(Mdg4) (PCC = 0.939) fell above the extrapolated (by polynomial regression) curve (**Fig. 1F**). Consistently, SUUR and Mod(Mdg4) exhibited nearly identical purification profiles in all five FPLC steps (**Fig. 1D**), unlike the next two top-scoring proteins, EGG (PCC = 0.881) and CG6700 (PCC = 0.874) (**figs. S1B**, **C**). Also, HP1a (PCC = 0.503), which had been proposed to form a complex with SUUR (*11*) did not co-purify with it in any FPLC steps (**fig. S1D**).

Mod(Mdg4) is a BTB/POZ domain protein that functions as an adaptor for architectural proteins that promote various aspects of genome organization (*13, 14*). It is expressed as 26 distinct polypeptides generated by splicing *in trans* of a common 5'-end precursor RNA with 26 unique 3'-end precursors (*15*). IL5 contained seven peptides derived from Mod(Mdg4) (99% confidence). Whereas four of them mapped to the common N-terminal 402 residues, three were specific to the C-terminus of a particular form, Mod(Mdg4)-67.2 (**figs. S1E-G**). Peptides specific to other splice forms were not detected. We raised an antibody to the C-terminus of Mod(Mdg4)-67.2, designated ModT antibody, and analyzed size exclusion column fractions by immunoblotting. Consistent with SWATH analyses (**Figs. 1C**, **D**) ~100-kDa SUUR and Mod(Mdg4)-67.2 polypeptides copurified as a complex with an apparent molecular mass of ~250 kDa (**Fig. 1G**). Finally, we confirmed that SUUR is specifically co-immunoprecipitated with Mod(Mdg4)-67.2 from crude extracts (**Fig. 1H**). As a control, XNP co-immunoprecipitated with HP1a as shown previously (*16*), but did not – with SUUR or Mod(Mdg4). We conclude that SUUR and Mod(Mdg4) form a stable stoichiometric complex that we term SUMM4.

We reconstituted recombinant SUMM4 complex by co-expressing FLAG-SUUR with Mod(Mdg4)-67.2-His_6 in Sf9 cells and co-purified them by FLAG affinity chromatography (**Fig. 2A**). Mod(Mdg4)-67.2 is the predominant form of Mod(Mdg4) expressed in embryos (*e.g.*, **Fig. 1H**). Thus, minor Mod(Mdg4) forms may have failed to be identified by IDA in IL5 (**fig. S1E**). We discovered that FLAG-SUUR did not co-purify with another splice form, Mod(Mdg4)-59.1 (**fig. S1G**, **Fig. 2A**).

Therefore, the shared N-terminus of Mod(Mdg4) (1-402) is not sufficient for interactions with SUUR. However, this result does not exclude a possibility that SUUR may form complex(es) with some of the other, low-abundance 24 splice forms of Mod(Mdg4). The SUUR- Mod(Mdg4)-67.2 interaction is specific, as the second-best candidate from our correlation analyses (*Drosophila* SetDB1 ortholog EGG; **Fig. 1F**) did not form a complex with FLAG-SUUR (**fig. S2A**), although it associated with its known partner WDE, an ortholog of hATF7IP/mAM (*17*).

The N-terminus of SUUR contains a region homologous with SNF2-like DEAD/H helicase domains. We analyzed the ability of recombinant SUUR and SUMM4 to hydrolyze ATP *in vitro* (**Fig. 2A**) in comparison to recombinant *Drosophila* ISWI (**fig. S2B**). Purified recombinant Mod(Mdg4)-67.2 and a variant SUUR protein with a point mutation in the putative Walker A motif (K59A) were used as negative controls (**Fig. 2A**, **fig. S2B**). Both SUUR and SUMM4 exhibited strong ATPase activities (**Fig. 2B**). SUMM4 was 1.4- to 2-fold more active than SUUR alone, indicating that Mod(Mdg4)-67.2 stimulates SUUR enzymatic activity. We then examined whether DNA and nucleosomes can stimulate the activity of SUUR. To this end, we reconstituted oligonucleosomes on plasmid DNA (**figs. S2B-E**). Linker histone H1-containing chromatin was also used as a substrate/cofactor, because SUUR has been demonstrated to physically interact with H1 (*18*). In contrast to ISWI, SUUR was not stimulated by addition of DNA or nucleosomes and moderately (by about 70%) activated by H1-containing oligonucleosomes (**Fig. 2C**).

We examined the nucleosome remodeling activities of SUUR and SUMM4; specifically, their ability to expose a positioned DNA motif in the EpiDyne®-PicoGreen™ assay (*Materials and Methods* and **fig. S2F**). Centrally or terminally positioned mononucleosomes were efficiently mobilized by ISWI and human BRG1 in a concentration- and time-dependent manner (**figs. S2G-J**). In contrast, SUUR and SUMM4 did not reposition either nucleosome (**Fig. 2D**). Thus, SUUR and SUMM4 do not possess a detectable remodeling activity.

We examined the localization patterns of SUMM4 subunits in polytene chromosomes by indirect immunofluorescence (IF) and discovered their strong overlap (**Fig. 3A**). In late endo-S phase when SUUR exhibited a characteristic distribution, it co-localized with Mod(Mdg4)-67.2, except for the chromocenter that did not show occupancy by Mod(Mdg4)-67.2. Mod(Mdg4)-67.2 was present at classical regions of SUUR enrichment, such as UR domains in 75C and 89E (**fig. S3A**). In contrast, there were multiple sites of Mod(Mdg4)-67.2 localization that were free of SUUR. This finding suggests that there are additional native form(s) of Mod(Mdg4)-67.2, either as an individual polypeptide or in complex(es) other than SUMM4. When we fractionated *Drosophila* nuclear extract using a different progression of FPLC steps (**fig. S3B**), we found that Mod(Mdg4)-67.2 can form a megadalton-size complex that did not contain SUUR (**figs. S3C-E**). Therefore, a more intricate pattern of Mod(Mdg4)-67.2 distribution likely reflects loading of both SUMM4 and the alternative complex.

We tested whether SUUR and Mod(Mdg4) loading into polytene chromosomes were mutually dependent using mutant alleles of *SuUR* and *mod(mdg4)*. $SuUR^{ES}$ is a null allele of *SuUR* (*19*). $mod(mdg4)^{m9}$ is a null allele with a deficiency that removes gene regions of the shared 5'-end precursor and eight specific 3'-precursors (*20*). $mod(mdg4)^{u1}$ contains an insertion of a *Stalker* element in the last coding exon of Mod(Mdg4)-67.2 3'-precursor (*14*). $SuUR^{ES}$ and $mod(mdg4)^{u1}$ are homozygous viable, and $mod(mdg4)^{m9}$ is recessive adult pharate lethal. We could not detect Mod(Mdg4)-67.2 expression in homozygous $mod(mdg4)^{m9}$ L3 salivary glands by immunoblotting, whereas $mod(mdg4)^{u1}$ expressed a truncated polypeptide (*cf*, ~70 kDa and ~100 kDa, **fig. S3F**). The truncated 70-kDa polypeptide failed to load into polytene chromosomes (**Fig. 3B**, **fig. S3G**). As shown previously, SUUR could not be detected in $SuUR^{ES}$ chromosomes. Since homozygous $mod(mdg4)^{m9}$ L3 larvae were produced by *inter se* crosses of heterozygous parents, the very low amounts of Mod(Mdg4)-67.2 in $mod(mdg4)^{m9}$ polytene chromosomes (barely above the detection limit) were presumably maternally contributed.

The absence (or drastic decrease) of Mod(Mdg4)-67.2 also strongly reduced the loading of SUUR (**Fig. 3B**, **fig. S3G**). The normal distribution pattern of SUUR in polytene chromosomes is highly

dynamic (*10, 18*). SUUR is initially loaded in chromosomes at the onset of endo-S phase and then re-distributes through very late endo-S, when it becomes accumulated in UR domains and PH. In both *mod(mdg4)* mutants, we observed a striking absence of SUUR in polytene chromosomes during early endo-S, which indicates that the initial deposition is dependent on its interactions with Mod(Mdg4). Although SUUR deposition slightly recovered by late endo-S, it was still several fold weaker than that in wild type. Potentially, in the absence of Mod(Mdg4), SUUR may be tethered to IH and PH loci by direct binding with linker histone H1 as shown previously (*18*). Finally, the gross subcellular distribution of SUUR also strongly correlated with that of Mod(Mdg4): a mis-localization of truncated Mod(Mdg4)-67.2 from nuclear to partially cytoplasmic was accompanied by a similar mis-localization of SUUR (**Fig. 3C**). This result indicates that the truncation of Mod(Mdg4) in *mod(mdg4)*^{u1} may have an antimorphic effect by mis-localization and deficient chromatin binding of interacting polypeptides, including SUUR (**Fig. 3C**) and others (**figs. S3B-E**).

Mod(Mdg4)-67.2 does not directly bind DNA but instead, is tethered by a physical association with zinc finger factor Suppressor of Hairy Wing, Su(Hw). Su(Hw) directly binds to consensus sequences that are present in *gypsy* transposable elements and are also widely distributed across the *Drosophila* genome in thousands of copies (*21*). Mod(Mdg4)-67.2 was previously shown to be essential for the insulator activity of *gypsy* (*14*), which functions *in vivo* to disrupt enhancer-promoter interactions and establish a barrier to the propagation of chromatin forms (*22, 23*). We therefore tested whether SUMM4 contributes to *gypsy* insulator functions. The *ct*^{6} allele of *Drosophila* contains a *gypsy* element inserted between the wing enhancer and promoter of the gene *cut* that inactivates *cut* expression and results in abnormal wing development (**Fig. 4A**). We discovered that both *mod(mdg4)*^{u1} and *SuUR*^{ES} mutations partially suppressed this phenotype (**Fig. 4A**) and significantly increased the wing size compared to *ct*^{6} allele alone (**Fig. 4B**). Thus, both subunits of SUMM4 are required to mediate the full enhancer-blocking activity of *gypsy*. Another insulator assay makes use of a collection of *P{SUPor-P}* insertions

that contain the *white* reporter flanked by 12 copies of *gypsy* Su(Hw)-binding sites. When *P{SUPor-P}* is inserted in heterochromatin, *white* is protected from silencing resulting in red eyes (*24*). Both *mod(mdg4)*^{u1} and *SuUR*^{ES} relieved the chromatin barrier function of Su(Hw) sites, causing repression of *white* (**Fig. 4C**). We conclude that SUMM4 is an insulator complex that mediates the chromatin boundary function of *gypsy* by a mechanism schematized in **Figs. 4D**, **E**.

A similar, chromatin partitioning-related mechanism may direct the function of SUUR in the establishment of UR in late-replicating IH domains of polytene chromosomes (**Fig. 4F**). It has been long known that 3D chromosome partitioning maps show an "uncanny alignment" with replication timing maps (*1*). To examine the possible roles of SUMM4 in UR, we measured DNA copy number genome-wide in salivary glands of L3 larvae by next generation sequencing (NGS). In w^{1118} control salivary glands, the DNA copy profile revealed large (>100-kbp) domains of reduced ploidy (**fig. S4A**), similar to a previous report (*18*). Excluding pericentric and sub-telomeric heterochromatin, we called 70 UR regions (**table S1**) in euchromatic arms, as described in *Materials and Methods*. In both *SuUR* and *mod(mdg4)*^{m9} null larvae, we observed suppression of UR in IH (**Fig. 4G**, **fig. S4B**, **table S1**). Consistent with its distribution *in vivo* (**Fig. 3A**), Mod(Mdg4) was dispensable for UR in PH. The NGS data strongly correlated with qPCR measurements of DNA copy numbers (**Fig. 4H**, **fig. S4C**). Furthermore, cytological evidence in 75C region supported the molecular analyses in that both mutants exhibited a brighter DAPI staining of the 75C1-2 band than that in w^{1118}, indicative of higher DNA content (**fig. S4C**). We conclude that the SUMM4 complex is required for the establishment of UR in the IH domains of *Drosophila* polytene chromosome. SUMM4 likely causes UR by forming a barrier to replication fork progression.

Our work demonstrates for the first time that insulator complexes assembled on chromatin can attenuate the progression of replication forks in salivary glands *in vivo*. Despite distinct cell cycle

programs in dividing and endoreplicating cells, the biochemical composition of replisomes in both cell types is identical (*4*). Therefore, similar insulator-driven control mechanisms for DNA replication are likely conserved in mitotically dividing diploid cells. Our data thus implicates insulator/chromatin boundary elements as a critical component of DNA replication control. Our model suggests that delayed replication of repressed chromatin (*e.g*., IH) during very late S phase can be imposed in a simple, two-stroke mechanism (**Fig. 4E**). First, it requires that an extended genomic domain is completely devoid of functional origins of replication. The assembly and licensing of proximal pre-RC complexes can be repressed epigenetically or at the level of DNA sequence. And second, this domain has to be separated from flanking chromatin by a barrier element associated with an insulator complex, such as SUMM4. This structural organization is capable of preventing or delaying the entry of external forks fired from distal origins. The current paradigm of replication timing largely focuses on the existence of "early" and "late" origins that are ordained during early S or G1 phase (*3*). Our model offers an additional mechanism to establish a locus-specific late replication program without a reliance on the variable timing of replication fork firing. Insulator complexes and other genome architectural proteins are highly abundant in the genome and assembled in a sequence-specific manner in G1. Even if all origins fire simultaneously (or stochastically, depending on availability of limiting factors) at the onset of S phase, any given DNA replication fork is forced to operate in a dense milieu of abundant, pre-positioned insulator complexes, continuously negotiating with them to get cleared for passage. After a replisome eventually escapes the barrier, the outcome would be indistinguishable from that of an RC firing at a "late origin" if analyzed cytologically (by incorporation of labeled nucleotides) or at the level of ChIP.

In conclusion, we used a newly developed MERCI approach to identify a stable stoichiometric complex termed SUMM4 that comprises SUUR, a previously known effector of replication control, and Mod(Mdg4), an insulator protein. SUMM4 subunits cooperate to mediate transcriptional repression and chromatin boundary functions of *gypsy*-like (class 3) insulators and regulate DNA replication by slowing down replication fork progression through the boundary element. Thus, SUMM4 is required for

coordinate regulation of gene expression, chromatin partitioning and DNA replication timing. The insulator-dependent regulation of DNA replication offers a novel mechanism for the establishment of replication timing in addition to the currently accepted paradigm of variable timing of replication origin firing.

5

References and notes:

1. N. Rhind, D. M. Gilbert, DNA replication timing. *Cold Spring Harbor perspectives in biology* **5**, a010132 (2013).
2. C. Marchal, J. Sima, D. M. Gilbert, Control of DNA replication timing in the 3D genome. *Nat Rev Mol Cell Biol* **20**, 721-737 (2019).
3. D. S. Dimitrova, D. M. Gilbert, The spatial position and replication timing of chromosomal domains are both established in early G1 phase. *Mol Cell* **4**, 983-993 (1999).
4. N. Zielke, B. A. Edgar, M. L. DePamphilis, Endoreplication. *Cold Spring Harbor perspectives in biology* **5**, a012948 (2013).
5. I. F. Zhimulev *et al.*, Polytene chromosomes: 70 years of genetic research. *International review of cytology* **241**, 203-275 (2004).
6. N. Sher *et al.*, Developmental control of gene copy number by repression of replication initiation and fork progression. *Genome research* **22**, 64-75 (2012).
7. E. S. Belyaeva *et al.*, Su(UR)ES: a gene suppressing DNA underreplication in intercalary and pericentric heterochromatin of Drosophila melanogaster polytene chromosomes. *Proc Natl Acad Sci U S A* **95**, 7532-7537 (1998).
8. J. T. Nordman *et al.*, DNA copy-number control through inhibition of replication fork progression. *Cell reports* **9**, 841-849 (2014).
9. O. V. Posukh, D. A. Maksimov, K. N. Skvortsova, D. E. Koryakov, S. N. Belyakin, The effects of SUUR protein suggest its role in repressive chromatin renewal during replication in Drosophila. *Nucleus* **6**, 249-253 (2015).
10. T. D. Kolesnikova *et al.*, Drosophila SUUR protein associates with PCNA and binds chromatin in a cell cycle-dependent manner. *Chromosoma* **122**, 55-66 (2013).
11. A. V. Pindyurin *et al.*, Interaction between the Drosophila heterochromatin proteins SUUR and HP1. *Journal of cell science* **121**, 1693-1703 (2008).

12. A. Munden *et al.*, Rif1 inhibits replication fork progression and controls DNA copy number in Drosophila. *eLife* **7**, (2018).

13. P. G. Georgiev, T. I. Gerasimova, Novel genes influencing the expression of the yellow locus and mdg4 (gypsy) in Drosophila melanogaster. *Molecular & general genetics : MGG* **220**, 121-126 (1989).

14. T. I. Gerasimova, D. A. Gdula, D. V. Gerasimov, O. Simonova, V. G. Corces, A Drosophila protein that imparts directionality on a chromatin insulator is an enhancer of position-effect variegation. *Cell* **82**, 587-597 (1995).

15. K. Buchner *et al.*, Genetic and molecular complexity of the position effect variegation modifier mod(mdg4) in Drosophila. *Genetics* **155**, 141-157 (2000).

16. A. V. Emelyanov, A. Y. Konev, E. Vershilova, D. V. Fyodorov, Protein complex of Drosophila ATRX/XNP and HP1a is required for the formation of pericentric beta-heterochromatin in vivo. *J Biol Chem* **285**, 15027-15037 (2010).

17. H. Wang *et al.*, mAM facilitates conversion by ESET of dimethyl to trimethyl lysine 9 of histone H3 to cause transcriptional repression. *Mol Cell* **12**, 475-487 (2003).

18. E. N. Andreyeva *et al.*, Regulatory functions and chromatin loading dynamics of linker histone H1 during endoreplication in Drosophila. *Genes Dev* **31**, 603-616 (2017).

19. I. V. Makunin *et al.*, The Drosophila suppressor of underreplication protein binds to late-replicating regions of polytene chromosomes. *Genetics* **160**, 1023-1034 (2002).

20. M. Savitsky, M. Kim, O. Kravchuk, Y. B. Schwartz, Distinct Roles of Chromatin Insulator Proteins in Control of the Drosophila Bithorax Complex. *Genetics* **202**, 601-617 (2016).

21. B. Adryan *et al.*, Genomic mapping of Suppressor of Hairy-wing binding sites in Drosophila. *Genome biology* **8**, R167 (2007).

22. R. R. Roseman, V. Pirrotta, P. K. Geyer, The su(Hw) protein insulates expression of the Drosophila melanogaster white gene from chromosomal position-effects. *EMBO J* **12**, 435-442 (1993).

23. H. Cai, M. Levine, Modulation of enhancer-promoter interactions by insulators in the Drosophila embryo. *Nature* **376**, 533-536 (1995).

24. R. R. Roseman *et al.*, A P element containing suppressor of hairy-wing binding regions has novel properties for mutagenesis in Drosophila melanogaster. *Genetics* **141**, 1061-1074 (1995).

25. I. Bag, R. K. Dale, C. Palmer, E. P. Lei, The zinc-finger protein CLAMP promotes gypsy chromatin insulator function in Drosophila. *Journal of cell science* **132**, (2019).

26. H. J. Bellen *et al.*, The BDGP gene disruption project: single transposon insertions associated with 40% of *Drosophila* genes. *Genetics* **167**, 761-781 (2004).

27. N. Negre *et al.*, A comprehensive map of insulator elements for the Drosophila genome. *PLoS Genet* **6**, e1000814 (2010).

Acknowledgments:

General

This paper is dedicated to the memory of Jonathan R. Warner who participated in initial discussions that have led to the development of MERCI technology. We thank B. Bartholdy, K. Beirit, B. Birshtein, V. Elagin, M. Gamble, T. Kolesnikova, A. Lusser, A. Pindyurin, C. Schildkraut, Y. Schwartz, S. Sidoli and I. Zhimulev for helpful discussions and critical reading of the manuscript. We thank Y. Schwartz, I. Zhimulev and Bloomington Stock Center for fly stocks and V. Corces, A. Pindyurin, J. Rowley and I. Zhimulev for antibodies. We are grateful to A. Aravin and B. Godneeva for cloning EGG and WDE baculovirus constructs. We thank A. Kumar and N. Baker for help with confocal microscopy, and M. Rogers and J. Secombe for the use of Zeiss Discovery.V12. Confocal images were obtained at the Analytical Imaging Facility (Einstein). We thank P. Schultes for help with maintaining the LCMS instrument.

Funding

National Institutes of Health grant R01GM074233 (DVF)

National Institutes of Health grant R01GM129244 (AIS, DVF)

National Institutes of Health grant R01GM124201 (RJD)

National Institutes of Health grant R44GM123869 (MCK/EpiCypher)

National Institutes of Health training grant K12GM000678 (MN)

National Institutes of Health training grant T32CA217824 (MN)

National Institute of General Medical Sciences equipment supplement (DVF, AIS, C. Query)

Author contributions

Conceptualization: AVE, DVF

Methodology, visualization and interpretation: ENA, AVE, MN, LS, MCK, RJD, AIS, DVF

Investigation: ENA, AVE, LS, EV, CAH, DVF

Funding acquisition: MCK, RJD, AIS, DVF

Supervision and project administration: MCK, RJD, AIS, DVF

Writing – original draft: RJD, AIS, DVF

Writing – review & editing: ENA, AVE, MN, LS, MCK, RJD, AIS, DVF

Competing interests

LS and MCK are employed by Epicypher, Inc., a commercial developer and supplier of the EpiDyne® nucleosomes and associated remodeling assay platforms used in this study. The remaining authors declare no competing interests.

Data and materials availability

NGS data has been submitted to Gene Expression Omnibus (GEO, accession number GSE189421).

Reviewers may access the data via: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE189421 using the token upirgumcxngljul

Supplementary materials:

Materials and Methods

Supplementary Text

Figs. S1 to S4

Tables S1 to S2

References (*28–64*)

Data S1 to S2

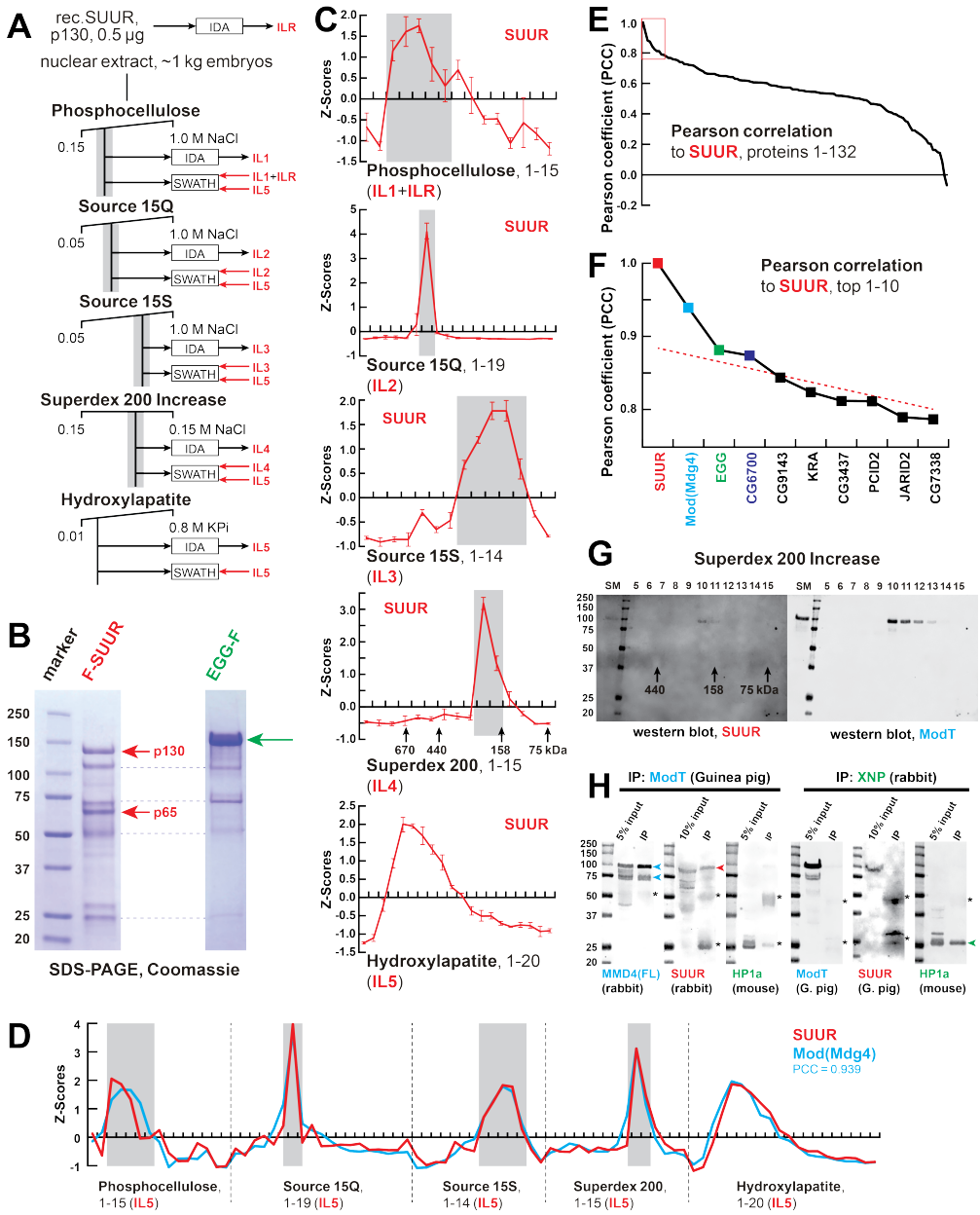

Fig. 1. Identification of the SUMM4 complex by MERCI. (**A**) Schematic of FPLC purification of the native form of SUUR using MERCI approach. ILR, ion library obtained by IDA of recombinant FLAG-SUUR; IL1-5, ion libraries obtained by IDA of FPLC fractions from chromatographic steps 1-5. KPi, potassium phosphate, pH 7.6. (B) Recombinant FLAG-SUUR expressed in Sf9 cells. Identities of eight most prominent bands were determined by mass-spectroscopy. p130 and p65 correspond to full-length and C-terminally truncated FLAG-SUUR, respectively (red arrows). Other bands represent common Sf9-specific contaminants purified by FLAG chromatography (blue dashed lines), *cf* purified EGG-F (green arrow). Molecular mass marker bands are indicated (kDa). (C) SWATH quantitation profiles of SUUR fractionation across individual FPLC steps. Ion libraries (IL) used for SWATH quantitation are shown at the bottom. Z-scores across indicated column fractions are plotted; error bars, standard deviations (N=3). Gray rectangles, fraction ranges used for the next FPLC step; in Superdex 200 step, black arrows, expected peaks of globular proteins with indicated molecular masses in kDa. (D) SWATH quantitation profiles of SUUR (red) and Mod(Mdg4) (cyan) fractionation across five FPLC steps. IL5 ion library was used for SWATH quantification. (E) Pearson correlation of fractionation profiles for individual 132 proteins to that of SUUR, sorted from largest to smallest. Red box, the graph portion shown in (F). (F) Top ten candidate proteins with the highest Pearson correlation to SUUR. Red dashed line, trend line extrapolated by polynomial regression (n = 5) from the bottom 130 proteins. (G) Western blot analyses of Superdex 200 fractions with SUUR and ModT antibodies. Molecular mass markers are shown on the left (kDa). (H) Co-IP experiments. SUUR (red arrowhead) co-purifies from nuclear extracts with Mod(Mdg4)-67.2 (cyan arrowheads) but not HP1a (green arrowhead). Anti-XNP co-IPs HP1a but not SUUR of Mod(Mdg4)-67.2. Asterisks, IgG heavy and light chains detected due to antibody cross-reactivity. Mod(Mdg4)-67.2(FL) antibody recognizes all splice forms of Mod(Mdg4).

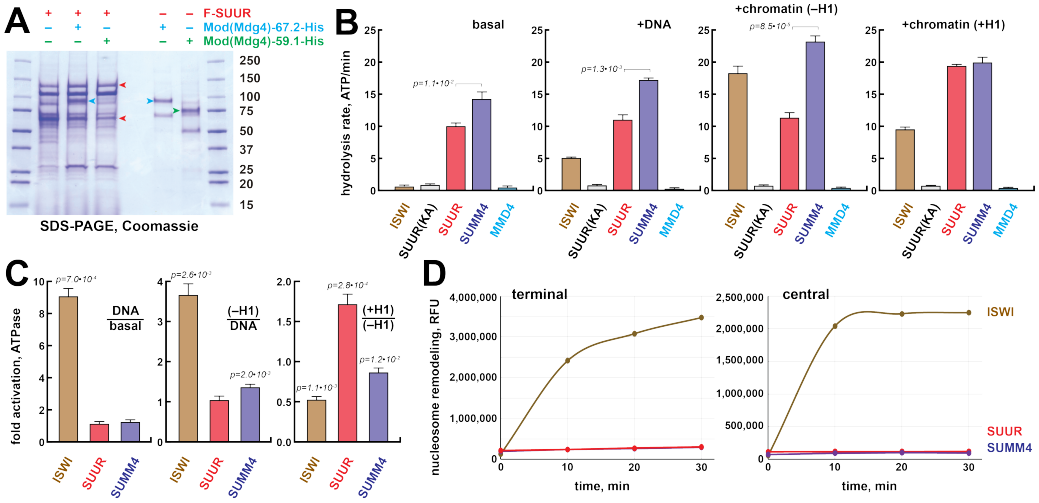

Fig. 2. Biochemical activities of SUMM4. (**A**) Recombinant SUMM4. Mod(Mdg4)-His$_6$, 67.2 (p100, cyan arrowhead) and 59.1 (p75, green arrowhead) splice forms were co-expressed with FLAG-SUUR (red arrowheads, p130 and p65) or separately in Sf9 cells and purified by FLAG or Ni-NTA affinity chromatography. Mod(Mdg4)-67.2 forms a specific complex with SUUR. (B) ATPase activities of recombinant ISWI (brown bars), FLAG-SUUR (red bars) and SUMM4 (FLAG-SUUR + Mod(Mdg4)-67.2-His$_6$, purple bars). Equimolar proteins were analyzed in reactions in the absence or presence of plasmid DNA or equivalent amounts of reconstituted oligonucleosomes, ±H1. SUUR(KA) and MMD4, ATPases activities of K59A mutant of SUUR (gray bars) and Mod(Mdg4)-67.2-His$_6$ (cyan bars). Hydrolysis rates were converted to moles ATP per mole protein per minute. All reactions were performed in triplicate, error bars represent standard deviations. *p*-values for statistically significant differences are indicated (Mann-Whitney test). (C) DNA- and nucleosome-dependent stimulation or inhibition of ATPase. The activities were analyzed as in (B). Statistically significant differences are shown (Mann-Whitney test). (D) Nucleosome sliding activities by EpiDyne®-PicoGreen™ assay (see *Materials and Methods*) with 5 nM of recombinant ISWI, SUUR or SUMM4. Reaction time courses are shown for terminally (6-N-66) and centrally (50-N-66) positioned mononucleosomes (**figs. S2G-J**). RFU, relative fluorescence units produced by PicoGreen fluorescence.

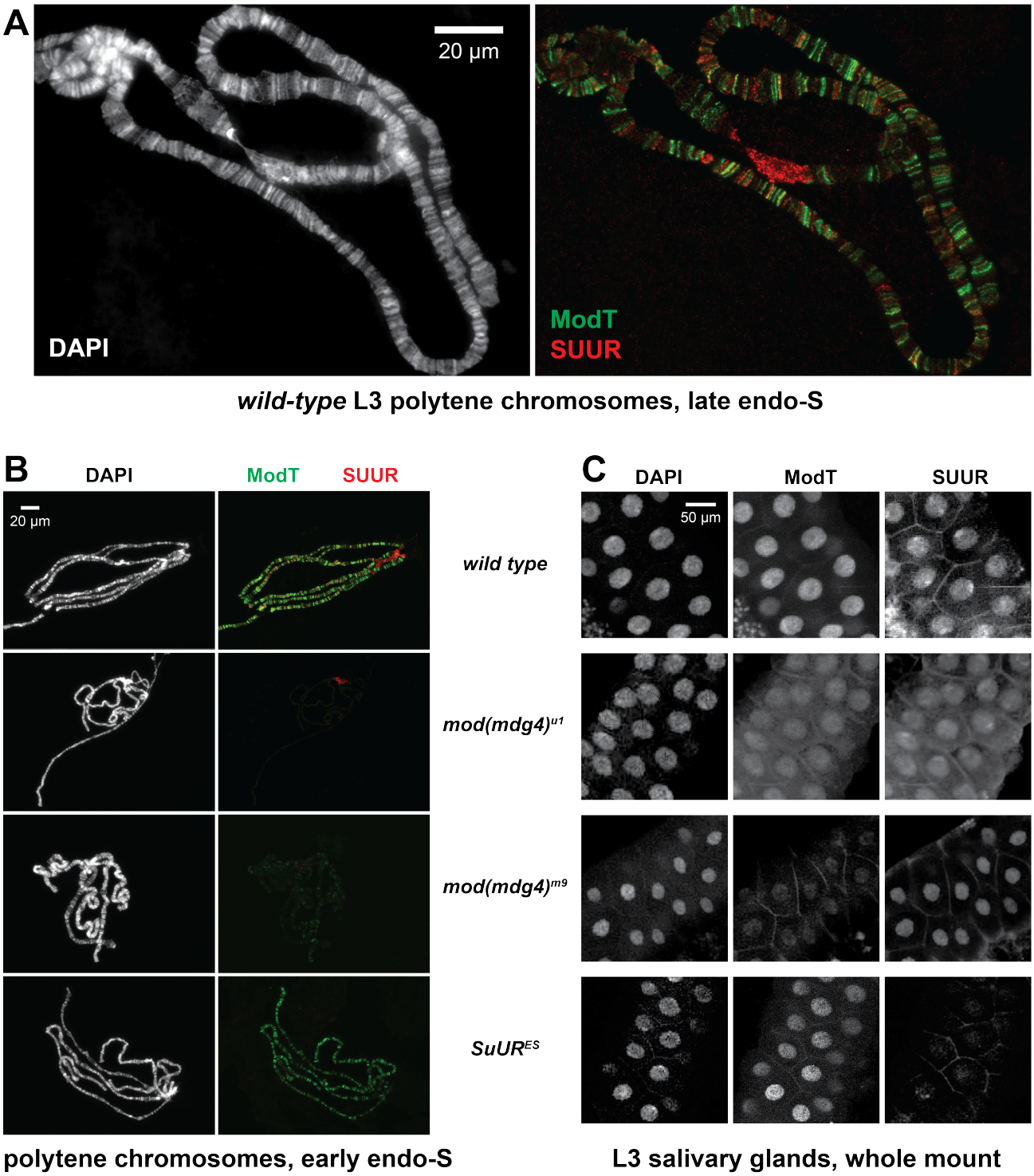

Fig. 3. Spatiotemporal distribution of SUMM4 *in vivo*. (**A**) Colocalization of SUUR and Mod(Mdg4)-67.2 in *wild-type* polytene chromosomes. Localization patterns of Mod(Mdg4)-67.2 and SUUR in L3 polytene chromosomes were analyzed by indirect IF staining. The polytene spread fragment (3L and 3R arms) corresponds to a nucleus in late endo-S phase, according to PCNA staining (**fig. S3A**). ModT (green) and SUUR (red) signals overlap extensively in euchromatic arms. The additional strong ModT IF loci that are SUUR-free and Mod(Mdg4)-67.2-free SUUR in pericentric 3LR are obvious. DAPI staining shows the overall chromosome morphology. (B) SUUR loading into chromosomes during early endo-S phase is compromised in *mod(mdg4)* mutants. *SuUR* mutation does not appreciably change the distribution of Mod(Mdg4)-67.2. Endo-S timing was established by PCNA staining (**fig. S3G**). (C) Abnormal subcellular distribution of SUMM4 subunits in *mod(mdg4)* and *SuUR* mutants. L3 salivary glands were fixed and whole-mount-stained with DAPI, ModT and SUUR antibodies. Whereas both polypeptides are mostly nuclear in wild type, they are partially mis-localized to cytoplasm in *mod(mdg4)*^{u1} mutant.

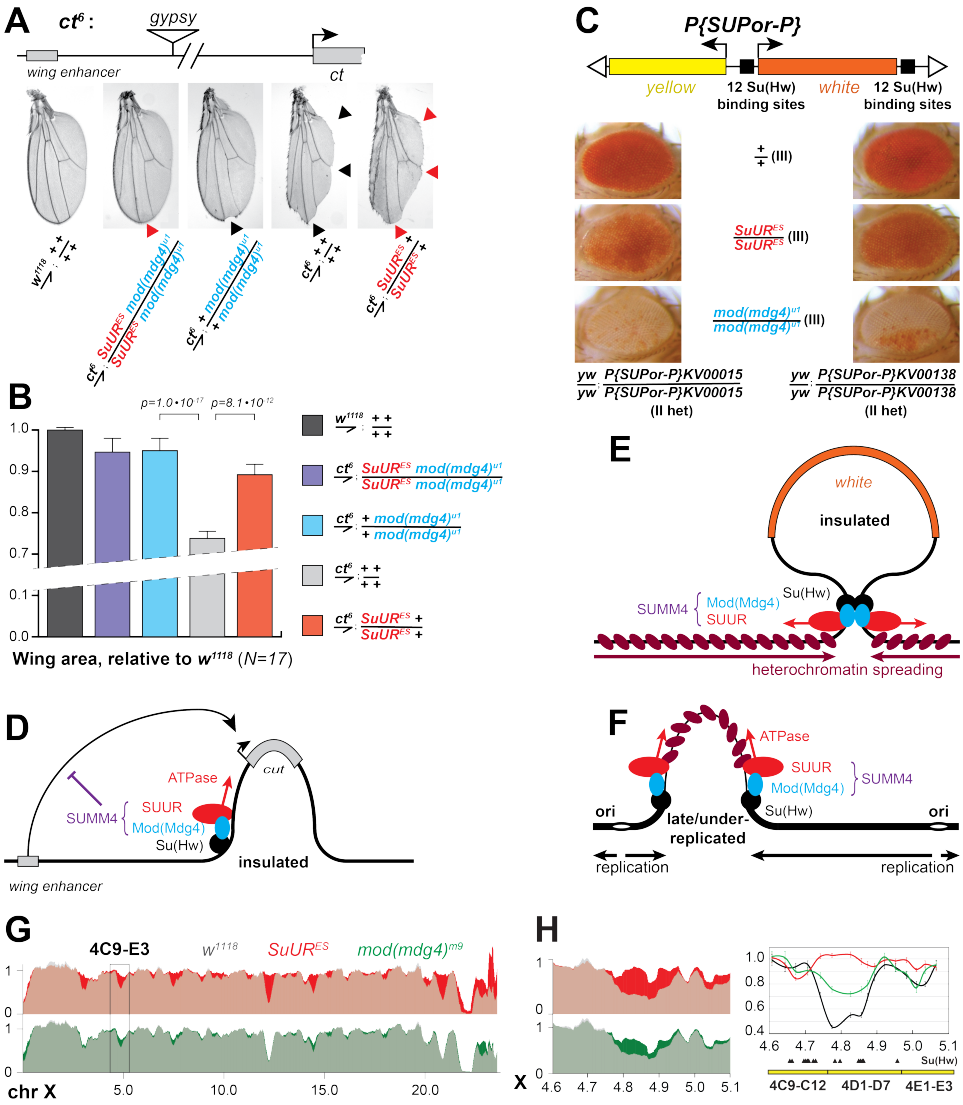

Fig. 4. Biological functions of SUMM4 in regulation of gene expression and DNA replication. (**A**) SUMM4 subunits are required for the enhancer-blocking activity in *ct*^{6}. Top: schematic diagram of the *ct*^{6} reporter system; the *gypsy* retrotransposon is inserted in between the wing enhancer and promoter of *cut* (*25*). Bottom left: the appearance of wild type adult wing; bottom right: the appearance of *ct*^{6} adult wing in the wild-type background. *SuUR*^{ES} and *mod(mdg4)*^{u1} alleles are recessive suppressors of the *ct*^{6} phenotype. Red and black arrowheads point to distinct anatomical features of the wing upon *SuUR* mutation. (B) Relative sizes (areas) of wings in adult male flies of indicated phenotypes were measured as described in *Materials and Methods*. *p*-values for statistically significant differences are indicated (t- test). (C) SUMM4 subunits are required for the chromatin barrier activity of Su(Hw) binding sites. Top: schematic diagram of the *P{SUPor-P}* reporter system (*26*); clustered 12 copies of *gypsy* Su(Hw) binding sites flanks the transcription unit of *white*. *KV00015* and *KV00138* are *P{SUPor-P}* insertions in pericentric heterochromatin of 2L. *SuUR*^{ES} and *mod(mdg4)*^{u1} alleles are recessive suppressors of the boundary that insulates *white* from heterochromatin encroaching. (D) Schematic model for the function of SUMM4 in blocking enhancer-promoter interactions in the *ct*^{6} locus. (E) Schematic model for the function of SUMM4 in establishing a chromatin barrier in heterochromatin-inserted *P{SUPor-P}* elements. (F) Schematic model for a putative function of SUMM4 in blocking/retardation of replication fork progression in IH domains. (G) Analyses of DNA copy numbers in *Drosophila* salivary gland cells. DNA from L3 salivary glands was subjected to high-throughput sequencing. DNA copy numbers (normalized to diploid embryonic DNA) are shown across the X chromosome. Genomic coordinates in Megabase pairs are indicated at the bottom. The control trace (*w*^{1118} allele) is shown as semitransparent light gray in the foreground; *SuUR*^{ES} and *mod(mdg4)*^{m9} traces are shown in the background in red and green, respectively; their overlaps with *w*^{1118} traces appear as lighter shades of colors. Black box, 4C9-E3 cytological region. (H) Close-up view of DNA copy numbers in region 4C9-E3 from high-throughput sequencing data are presented as in (G). DNA copy numbers were also measured independently by real-time qPCR. The numbers were calculated relative to embryonic DNA and normalized to a control intergenic region. The X-axis shows chromosome positions (in Megabase pairs) of target amplicons. Error bars represent the confidence interval (see *Materials and Methods*). Black arrowheads, positions of mapped Su(Hw) binding sites (*27*). Yellow boxes show approximate boundaries of cytogenetic bands.