Drosophila SUMM4 complex couples insulator function and DNA replication 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

*Corresponding author: dmitry.fyodorov@einsteinmed.org

†These authors contributed equally to this work

Competing interest statement

Lu Sun and Michael-C Keogh 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.

Abstract Asynchronous replication of chromosome domains during S phase is essential for eukaryotic genome function, but the mechanisms establishing which domains replicate early versus late in different cell types remain incompletely understood. Intercalary heterochromatin domains replicate very late in both diploid chromosomes of dividing cells and in endoreplicating polytene chromosomes where they are also underrelicated. *Drosophila* SNF2-related factor SUUR imparts locus-specific underreplication of polytene chromosomes. SUUR negatively regulates DNA replication fork progression; however, its mechanism of action remains obscure. Here we developed a novel method termed MS-Enabled Rapid protein Complex Identification (MERCI) to isolate a stable stoichiometric native complex SUMM4 that comprises SUUR and a chromatin boundary protein Mod(Mdg4)-67.2. Mod(Mdg4) stimulates SUUR ATPase activity and is required for a normal spatiotemporal distribution of SUUR *in vivo*. SUUR and Mod(Mdg4)-67.2 together mediate the activities of *gypsy* insulator that prevent certain enhancer-promoter interactions and establish euchromatin-heterochromatin barriers in the genome. Furthermore, *SuUR* or *mod(mdg4)* mutations reverse underreplication of intercalary heterochromatin. Thus, SUMM4 can impart late replication of intercalary heterochromatin by attenuating the progression of replication forks through euchromatin/heterochromatin boundaries. Our findings implicate a SNF2 family ATP-dependent motor protein SUUR in the insulator function, reveal that DNA replication can be delayed by a chromatin barrier and uncover a critical role for architectural proteins in replication control. They suggest a mechanism for the establishment of late replication that does not depend on an asynchronous firing of late replication origins.

Introduction

Replication of metazoan genomes occurs according to a highly coordinated spatiotemporal program, where discrete chromosomal regions replicate at distinct times during S phase (Rhind & Gilbert, 2013). The replication program follows the spatial organization of the genome in Megabase-long constant timing regions interspersed by timing transition regions (Marchal, Sima, & Gilbert, 2019). The spatiotemporal replication program exhibits correlations with 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 at origins (Dimitrova & Gilbert, 1999), suggesting a mechanism that involves factors other than the core replication machinery.

Most larval tissues of *Drosophila melanogaster* grow via G-S endoreplication cycles that duplicate DNA without cell division resulting in polyploidy (Zielke, Edgar, & DePamphilis, 2013). Endoreplicated DNA molecules frequently align in register to form giant polytene chromosomes (Zhimulev et al., 2004). Importantly, in some cell types, genomic domains corresponding to the latest replicated regions of dividing cells, specifically pericentric (PH) and intercalary (IH) heterochromatin, fail to fully replicate during each endocycle resulting in underreplication (UR). These regions are depleted of sites for binding the Origin of Replication Complex (ORC) and thus their replication primarily relies on forks progressing from external origins (Sher et al., 2012) in both dividing and endoreplicating cells, which suggests that both cell types utilize related mechanisms of regulation of late replication. Although cell cycle programs are dissimilar between endoreplicating and mitotically dividing cells (Zielke et al., 2013), they likely share the components of core biochemical machinery

for DNA replication. Thus, underreplication provides a facile readout for late replication initiation and delayed fork progression.

The *Suppressor of UnderReplication* (*SuUR*) gene is essential for polytene chromosome underreplication in intercalary and pericentric heterochromatin (Belyaeva et al., 1998). In *SuUR* mutants, the DNA copy number in underreplicated regions is partially restored to almost reach those for fully polyploidized regions of the genome. *SuUR* encodes a protein (SUUR) containing a helicase domain with homology to that of the SNF2/SWI2 family. The occupancy of ORC in intercalary and pericentric heterochromatin is not increased in *SuUR* mutants (Sher et al., 2012), and thus the increased replication of underreplicated regions is likely not due to the firing of additional origins. Rather, SUUR negatively regulates the rate of replication fork progression (Nordman et al., 2014) by an unknown mechanism. It has been proposed (Posukh, Maksimov, Skvortsova, Koryakov, & Belyakin, 2015) that retardation of the replisome by SUUR takes place via simultaneous physical association with the components of the fork (*e.g.*, CDC45 and PCNA) (Kolesnikova et al., 2013; Nordman et al., 2014) and repressive chromatin proteins, such as HP1a (Pindyurin et al., 2008).

Using a newly developed proteomics approach, we discovered that SUUR forms a stable complex stoichiometric with a chromatin boundary protein Mod(Mdg4)-67.2. We demonstrate that SUUR and Mod(Mdg4)-67.2 together are required for maximal underreplication of intercalary heterochromatin and full activity of the *gypsy* insulator, thereby implicating insulators in obstructing replisome progression and the control of late DNA replication.

Results

Identification of SUMM4, the native form of SUUR in *Drosophila* embryos

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 (Kolesnikova et al., 2013; Munden et al., 2018; Nordman et al., 2014; Pindyurin et al., 2008). However, evaluating whether any of these proteins are present in a native SUUR complex is problematic because of the low abundance of SUUR, which also precludes its purification by conventional chromatography. Therefore, we developed a novel biochemical approach using embryonic extracts (which can be obtained in large quantities) that relies on partial purification by multi-step FPLC (***Figure 1A***) and shotgun proteomics of chromatographic fractions by quantitative LCMS. We term this technology MERCI for MS-Enabled Rapid protein Complex Identification (*Materials and Methods*).

Shotgun quantification of complex mixtures of polypeptides by LCMS is performed in two steps. First, the composition of the mixture is examined by Information-Dependent Acquisitions (IDA) that establish protein identities based on MS1 and MS2 spectra of detected tryptic peptides. This information is used to compile a so-called "ion library" (IL), which is then utilized to quantify spectral information obtained from the same samples by unbiased, Data-Independent Acquisitions (DIA), sometimes termed Sequential Window Acquisitions of All Theoretical Mass Spectra (SWATH-MS/SWATH). Importantly, the depth of proteomic quantification is limited by the range of peptides in the ion library (IL) originally built by IDA.

SUUR-specific peptides could not be found in ILs obtained from acquisitions of crude nuclear extracts or any fractions from the first, phosphocellulose, step (IL1, ***Figure 1B***, ***Supplementary File***

110 ***1***), and therefore, SUUR could not be quantified in SWATH acquisitions of phosphocellulose fractions
111 when IL1 alone is used as a reference. Thus, to measure the relative abundance of SUUR in
112 phosphocellulose fractions, we augmented IL1 with the ion library obtained by IDA of recombinant
113 SUUR (ILR, ***Figure 1B&C***). In ion libraries from subsequent chromatographic steps (IL2-IL5),
114 peptides derived from native SUUR were detected (***Figure 1B***, ***Supplementary File 1***) and used for
115 quantification of cognate DIA/SWATH acquisitions (***Figure 1D-H***).
116
117 The final aspect of the MERCI algorithm calls for re-quantification of FPLC fraction SWATH
118 acquisitions with an ion library from the last step (IL5) that is enriched for peptides derived from
119 SUUR and co-purifying polypeptides (***Figure 1A***) and includes only 140 proteins (***Figure 1B***,
120 ***Supplementary File 1***). In this fashion, scarce polypeptides (including SUUR and, potentially, SUUR-
121 binding partners) that may not be detectable in earlier steps will not evade quantification. Purification
122 profiles of proteins quantified in all five FPLC steps (132) were then artificially stitched into 83-point
123 arrays of Z-scores (***Figure 1I, Supplementary File 2***). These profiles were Pearson-correlated with
124 that of SUUR and ranked down from the highest Pearson coefficient, PCC (***Figure 2A***). Whereas the
125 PCC numbers for the bottom 130 proteins lay on a smooth curve, the top two proteins, SUUR (PCC =
126 1.000) and Mod(Mdg4) (PCC = 0.939) fell above the extrapolated (by polynomial regression) curve
127 (***Figure 2B***). Consistently, SUUR and Mod(Mdg4) exhibited nearly identical purification profiles in
128 all five FPLC steps (***Figure 2C***), unlike the next two top-scoring proteins, EGG (PCC = 0.881) and
129 CG6700 (PCC = 0.874) (***Figure 2—figure supplement 1A&B***). Also, HP1a (PCC = 0.503), which had
130 been proposed to form a complex with SUUR (Pindyurin et al., 2008) did not co-purify with SUUR in
131 any FPLC steps (***Figure 2—figure supplement 1C***).
132
133 Mod(Mdg4) is a BTB/POZ domain protein that functions as an adaptor for architectural proteins
134 that promote various aspects of genome organization (Georgiev & Gerasimova, 1989; Gerasimova,

135 Gdula, Gerasimov, Simonova, & Corces, 1995). It is expressed as 26 distinct polypeptides generated
136 by splicing *in trans* of a common 5'-end precursor RNA with 26 unique 3'-end precursors (Buchner et
137 al., 2000). IL5 contained seven peptides derived from Mod(Mdg4) (99% confidence). Whereas four of
138 them mapped to the common N-terminal 402 residues, three were specific to the C-terminus of a
139 particular form, Mod(Mdg4)-67.2 (***Figure 2—figure supplement 2***). Peptides specific to other splice
140 forms were not detected. We raised an antibody to the C-terminus of Mod(Mdg4)-67.2, designated
141 ModT antibody, and analyzed size exclusion column fractions by immunoblotting. Consistent with
142 SWATH analyses (***Figure 1G&2C***), SUUR and Mod(Mdg4)-67.2 polypeptides copurified as a
143 complex with an apparent molecular mass of ~250 kDa (***Figure 2D***). Finally, we confirmed that
144 SUUR specifically co-immunoprecipitated with Mod(Mdg4)-67.2 from embryonic nuclear extracts
145 (***Figure 2E***). As a control, XNP co-immunoprecipitated with HP1a as shown previously (Emelyanov,
146 Konev, Vershilova, & Fyodorov, 2010), but did not with SUUR or Mod(Mdg4) (***Figure 2E***). We
147 conclude that SUUR and Mod(Mdg4) form a stable stoichiometric complex that we term SUMM4
148 (Suppressor of Underreplication – Modifier of Mdg4).
149

150 **Biochemical activities of recombinant SUMM4 *in vitro***

151 We reconstituted recombinant SUMM4 complex by co-expressing FLAG-SUUR with
152 Mod(Mdg4)-67.2-His_6 in Sf9 cells and purified it by FLAG affinity chromatography (***Figure 3A***).
153 Mod(Mdg4)-67.2 is the predominant form of Mod(Mdg4) expressed in embryos (*e.g.*, ***Figure 2E***, left
154 panel). Thus, minor Mod(Mdg4) forms may have failed to be identified by IDA in IL5 (***Figure***
155 ***2—figure supplement 2A***). We discovered that FLAG-SUUR did not co-purify with another splice
156 form, Mod(Mdg4)-59.1 (***Figure 2—figure supplement 2C***, ***Figure 3A***). Whereas the identity of an
157 ~100-kDa Mod(Mdg4)-67.2-His_6 band copurifying with FLAG-SUUR was confirmed by mass-spec
158 sequencing, the FLAG-purified material from Sf9 cells expressing FLAG-SUUR and Mod(Mdg4)-

59.1 did not contain Mod(Mdg4)-specific peptides. 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; ***Figure 2B***) did not form a complex with FLAG-SUUR (***Figure 3—figure supplement 1A***), although it associated with its known partner WDE, an ortholog of hATF7IP/mAM (Wang et al., 2003).

The N-terminus of SUUR contains a region homologous with SNF2-like DEAD/H helicase domains. Although SUUR requires its N-terminal domain to function *in vivo* (Munden et al., 2018), it has been hypothesized to be inactive as an ATPase (Nordman & Orr-Weaver, 2015). We analyzed the ability of recombinant SUUR and SUMM4 (***Figure 3A***) to hydrolyze ATP *in vitro* in comparison to recombinant *Drosophila* ISWI (***Figure 3—figure supplement 1B***). Purified recombinant Mod(Mdg4)-67.2 (***Figure 3A***) and a variant SUUR protein with a point mutation in the putative Walker A motif (K59A) were used as negative controls (***Figure 3A***, ***Figure 3—figure supplement 1B***). Contrary to the prediction, both SUUR and SUMM4 exhibited strong ATPase activities (***Figure 3B***). 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 (***Figure 3—figure supplement 1C-E***). Linker histone H1-containing chromatin was also used as a substrate/cofactor, because SUUR has been demonstrated to physically interact with H1 (Andreyeva et al., 2017). In contrast to ISWI, SUUR was not stimulated by addition of DNA or nucleosomes and moderately (by about 70%) activated by H1-containing oligonucleosomes (***Figure 3C***) consistent with its reported direct physical interaction with H1 (Andreyeva et al., 2017).

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 ***Figure 3—figure supplement 2A***). Centrally or terminally positioned mononucleosomes were efficiently mobilized by ISWI and human BRG1 in a concentration- and time-dependent manner (***Figure 3—figure supplement 2B-E***). In contrast, SUUR and SUMM4 did not reposition either nucleosome (***Figure 3D***). Thus, SUUR and SUMM4 do not possess a detectable remodeling activity and may resemble certain other SNF2-like enzymes (*e.g.*, RAD54) that utilize the energy of ATP hydrolysis to mediate alternate DNA translocation reactions (Jaskelioff, Van Komen, Krebs, Sung, & Peterson, 2003).

The distribution of SUMM4 complex *in vivo*

We examined the positions of SUUR and Mod(Mdg4)-67.2 within polytene chromosomes by indirect immunofluorescence (IF) and discovered that they overlap at numerous locations (***Figure 4A***, ***Figure 4—figure supplement 1A&B***). In late endo-S phase, when SUUR exhibited a characteristic distribution, it co-localized with Mod(Mdg4)-67.2 at numerous (hundreds of) loci along the chromosome arms (***Figure 4—figure supplement 1B***). Mod(Mdg4)-67.2 was present at classical regions of SUUR enrichment, such as underreplicated domains in 75C and 89E (***Figure 4—figure supplement 1A***). The chromocenter, which consists of underreplicated pericentric heterochromatin, contains SUUR but did not show occupancy by Mod(Mdg4)-67.2 (***Figure 4—figure supplement 1A***). Conversely, there were multiple sites of Mod(Mdg4)-67.2 localization that were free of SUUR (***Figure 4—figure supplement 1A&B***). Individual pixel intensities of IF signals for SUUR and Mod(Mdg4)-67.2 were plotted as a 2D scatter plot (***Figure 4—figure supplement 1C***) and were found to exhibit a weak positive correlation (R^2=0.278). Consistent with the possible multi-phasic relative

distribution of SUUR and Mod(Mdg4)-67.2 (***Figure 4—figure supplement 1B***), the 2D plot encompassed four distinct areas, where SUUR and Mod(Mdg4)-67.2 were co-localized, enriched separately or absent (***Figure 4—figure supplement 1D***). When regions of SUUR-alone and Mod(mdg4)-67.2-alone enrichment were excluded, and only the regions of their apparent colocalization were considered, the anti-SUUR and anti-ModT signals exhibited a strong positive correlation (R^2=0.568, ***Figure 4—figure supplement 1D***).

The existence of chromosome loci heavily enriched for Mod(Mdg4)-67.2 but devoid of SUUR 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 (***Figure 4—figure supplement 2A***), we found that Mod(Mdg4)-67.2 can form a megadalton-sized complex that did not contain SUUR (***Figure 4—figure supplement 2B-D***). Therefore, a more intricate pattern of Mod(Mdg4)-67.2 distribution likely reflects loading of both SUMM4 and an alternative Mod(Mdg4)-67.2-containing 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* (Makunin et al., 2002). $mod(mdg4)^{m9}$ is a null allele with a deficiency that removes gene regions of the shared 5'-end precursor and eight specific 3'-precursors (Savitsky, Kim, Kravchuk, & Schwartz, 2016). $mod(mdg4)^{u1}$ contains an insertion of a *Stalker* element in the last coding exon of Mod(Mdg4)-67.2 3'-precursor (Gerasimova et al., 1995), and thus is predicted only to disrupt expression of this isoform. $SuUR^{ES}$ and $mod(mdg4)^{u1}$ are homozygous viable, and $mod(mdg4)^{m9}$ is recessive adult pharate lethal. Although homozygous $mod(mdg4)^{m9}$ animals die after the pupal stage, they survive until late third instar larvae (L3). Therefore, this allele cannot be used to study adult phenotypes, but it is possible to

analyze its effects in L3, such as on polytene chromosome structure. Importantly however, since the homozygous progeny is produced by heterozygous parents, the recessive phenotypes would not reveal themselves until the maternally loaded protein and RNA are exhausted (diluted and/or degraded) by late larval stages, as frequently occurs for other *Drosophila* mutants.

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, ***Figure 4—figure supplement 3A***). The truncated 70-kDa polypeptide failed to load into polytene chromosomes (***Figure 4B***, ***Figure 4—figure supplement 3B***). 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 (***Figure 4B***, ***Figure 4—figure supplement 3B***). The normal distribution pattern of SUUR in polytene chromosomes is highly dynamic (Andreyeva et al., 2017; Kolesnikova et al., 2013). SUUR is initially loaded in chromosomes at the onset of endo-S phase and then re-distributes through very late endo-S, when it accumulates in underreplicated domains and pericentric heterochromatin. In both *mod(mdg4)* mutants, we observed a striking absence of SUUR in euchromatic arms of polytene chromosomes during early endo-S (***Figure 4B***, ***Figure 4—figure supplement 3B***), which indicates that the initial deposition of SUUR 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 control. Potentially, in the absence of Mod(Mdg4), SUUR may be tethered to intercalary and pericentric heterochromatin loci by direct binding with linker histone H1 as shown previously (Andreyeva et al.,

2017). 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 (***Figure 4C***). This result indicates that the truncation of Mod(Mdg4) in *mod(mdg4)*^{u1} may have an antimorphic effect by mis-localization and deficient chromatin loading of interacting polypeptides, including SUUR (***Figure 4C***) and others (***Figure 4—figure supplement 2B-D***).

The role of SUMM4 as an effector of the insulator/chromatin barrier function

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) (Gause, Morcillo, & Dorsett, 2001). 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 (Adryan et al., 2007). Mod(Mdg4)-67.2 was previously shown to be essential for the insulator activity of *gypsy* (Gerasimova et al., 1995), which functions *in vivo* to prevent enhancer-promoter interactions and establish a barrier to the propagation of chromatin forms (Cai & Levine, 1995; Roseman, Pirrotta, & Geyer, 1993). We therefore tested whether SUMM4 contributes to the *gypsy* insulator functions.

The *ct*^{6} allele of *Drosophila* contains a *gypsy* element inserted between the wing enhancer and promoter of the gene *cut*. The insertion inactivates *cut* expression and results in abnormal wing development (***Figure 5A***). We discovered that both *mod(mdg4)*^{u1} and *SuUR*^{ES} mutations partially suppressed this phenotype (***Figure 5A***) and significantly increased the wing size compared to *ct*^{6} allele alone (***Figure 5B***). Thus, both subunits of SUMM4 are required to mediate the full enhancer-blocking activity of *gypsy*. Interestingly, the double, *SuUR*^{ES} and *mod(mdg4)*^{u1}, mutant produced an additional suppression of the *ct*^{6} phenotype compared to that by *mod(mdg4)*^{u1} alone (***Figure 5A***, red arrowhead),

which suggests that SUUR may contribute to the insulator function in the absence of Mod(Mdg4)-67.2.

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 (***Figure 5C***, top). When *P{SUPor-P}* is inserted in heterochromatin, *white* is protected from silencing resulting in red eyes (Roseman et al., 1995). Both *mod(mdg4)*^{u1} and *SuUR*^{ES} relieved the chromatin barrier function of Su(Hw) sites, causing repression of *white* (***Figure 5C***). We conclude that SUMM4 is an insulator complex that contributes to the enhancer-blocking and chromatin boundary functions of *gypsy* by a mechanism schematized in ***Figure 6A&B***.

The role of SUMM4 in regulation of DNA replication in polytene chromosomes

A similar, chromatin partitioning-related mechanism may direct the function of SUUR in the establishment of underreplication in late-replicating intercalary heterochromatin domains of polytene chromosomes (***Figure 6C***). It has been long known that 3D chromosome partitioning maps show an "uncanny alignment" with replication timing maps (Rhind & Gilbert, 2013). To examine the possible roles of SUMM4 in underreplication, 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 (***Figure 7A***), similar to previous reports (Andreyeva et al., 2017; Sher et al., 2012; Yarosh & Spradling, 2014). Excluding pericentric and sub-telomeric heterochromatin, we called 70 underreplicated regions (***Table 1***) in euchromatic arms, as described in *Materials and Methods*.

In both *SuUR* and *mod(mdg4)*^{m9} null larvae, we observed statistically significant suppression of underreplication in intercalary heterochromatin (***Figure 7B***, ***Figure 7—figure supplement 1A***, ***Table***

304 ***1***). In line with its lack of accumulation within the chromocenter of polytene chromosomes (***Figure***
305 ***4A***), Mod(Mdg4) was largely dispensable for underreplication in pericentric heterochromatin. The
306 NGS data strongly correlated with qPCR measurements of DNA copy numbers (***Figure 7C&D***).
307 Furthermore, cytological evidence in the 75C region supported the molecular analyses in that both
308 mutants exhibited a brighter DAPI staining of the 75C1-2 band than that in w^{1118}, indicative of higher
309 DNA content (***Figure 7D***). Importantly, consistent with the role of Mod(Mdg4)-dependent insulators
310 in the establishment of underreplication, the boundaries of underreplicated domains frequently
311 encompass multiple clustered Su(Hw) binding sites (***Figure 7C&D***).
312
313 Uniformly, *SuUR* mutation gave rise to a stronger relief of underreplication than that produced by
314 the $mod(mdg4)^{m9}$ null allele (***Table 1***). This result can be explained by embryonic deposition of
315 functional Mod(Mdg4) proteins and RNA by heterozygous mothers, unlike the complete absence of
316 SUUR throughout the life cycle of the homozygous viable and fertile $SuUR^{ES}$ animals. Although third
317 instar larvae are >1,000-fold larger, volume-wise, than the embryos, persistent Mod(Mdg4)-67.2 can
318 still be detected in polytene chromosomes of these larvae by IF despite its dilution and degradation
319 (***Figure 4B***, ***Figure 4—figure supplement 3B***). In contrast, unlike L3, first instar larvae (L1) are
320 nearly identical in size to the embryos. Therefore, since the endoreplication cycles initiate in embryos
321 and L1, in $mod(mdg4)^{m9}$ animals the first few out of 10-11 rounds of chromosome polytenization take
322 place with an almost normal amount of Mod(Mdg4) present, which may substantially limit the effect
323 of $mod(mdg4)^{m9}$ mutation on underreplication as measured in L3.
324
325 Seemingly, there is a contradiction between a strong effect that *mod(mdg4)* null mutation has on
326 the loading of SUUR in polytene chromosomes (***Figure 4B***) and a weaker effect on underreplication
327 (***Figure 7B-D***, ***Figure 7—figure supplement 1A&B***, ***Table 1***). However, the SUUR occupancy is
328 examined in L3 after the maternal *mod(mdg4)* product is nearly eliminated (***Figure 4B***). On the other

329 hand, the DNA copy number, although also measured in L3 (***Figure 7B-D***, ***Figure 7—figure***
330 ***supplement 1A&B***, ***Table 1***), is a product of multiple rounds of endoreplication that initiate before
331 Mod(Mdg4) is exhausted. To validate the putative effect of maternally contributed SUMM4 on the
332 establishment of underreplication, we performed qPCR measurements of DNA copy numbers in
333 salivary glands of homozygous *SuUR* animals produced by *inter se* crosses of heterozygous $SuUR^{ES}$/+
334 parents (***Figure 7C&D***, zygotic $SuUR^{ES}$). Similar to the maternal Mod(Mdg4), the initial maternal
335 contribution of SUUR partially limited the reversal of underreplication in cytological regions 4D and
336 75C. Thus, when the *SuUR* and *mod(mdg4)* null mutant animals are similarly derived from
337 heterozygous mothers that deposit wild-type gene product into their progeny, the mutant
338 underreplication phenotypes in the third instar larval salivary gland are essentially indistinguishable.
339 Finally, we analyzed the effect of homozygous $mod(mdg4)^{u1}$ mutation, which is viable and fertile, on
340 DNA copy numbers in the 75C underreplicated domain by qPCR and cytologically (***Figure 7D***). We
341 observed a substantially stronger suppression of underreplication than that in $mod(mdg4)^{m9}$,
342 presumably due to the absence of maternal contribution of full-length Mod(Mdg4)-67.2.
343
344 We conclude that SUUR and Mod(Mdg4)-67.2 act together as subunits of stable SUMM4 complex,
345 which is required for the establishment of underreplication in the intercalary heterochromatin domains of
346 *Drosophila* polytene chromosome.

Discussion

MERCI is a powerful new approach to characterize stable stoichiometric protein complexes

We present here a facile method, termed MERCI, to rapidly identify subunits of stable native complexes by only partial chromatographic purification. It allows one to circumvent the conventional, rate-limiting approach to purify proteins to apparent homogeneity. Since a multi-step FPLC scheme invariably leads to an exponential loss of material, reducing the number of purification steps in the MERCI protocol allows identification of rare complexes, such as SUMM4, which may be present in trace amounts in native sources. On the other hand, MERCI obviates introduction of false-positives frequently associated with tag purification of ectopically expressed targets that render results less reliable. Notably, MERCI is not limited to analyses of known polypeptides, since it is readily amenable to fractionation of native factors based on a correlation with their biochemical activities *in vitro*.

The dissection of protein interactome by extract fractionation on orthogonal FPLC columns and MS-based approaches has been previously attempted (Havugimana et al., 2012; Shatsky et al., 2016). However, unlike the newly developed MERCI approach, these studies were aimed at comprehensive, proteome-wide analyses, which managed to only yield data for the most abundant complexes. The major distinction of the MERCI protocol is that it is targeted towards a particular protein (SUUR in this study). The crucial final stage of the MERCI algorithm is re-quantification of all acquired SWATH data using a library acquired from fractions of the last column (IL5, ***Figure 1A***, ***B&I***). The target protein and co-purifying polypeptides are substantially enriched after several chromatographic steps and thus, yield a greater number of detected peptides, which helps a more precise quantification. Although SWATH allows reliable measurement of picogram amounts of proteins (***Figure 1—figure supplement 1A&B***), the range of quantified polypeptides is always limited by those present in IDA acquisitions (ion libraries). For low-abundance proteins, such as SUUR and Mod(Mdg4), specific peptides are not

371 detectable by IDA in earlier chromatographic steps (***Supplementary File 1***). Consequently, SWATH
372 quantification using only the cognate ion libraries would not discern the near perfect co-fractionation of
373 SUUR and Mod(Mdg4) in all five steps (***Figure 2C***), precluding identification of the SUUR-
374 Mod(Mdg4) complex (***Figure 2B&C***).
375
376 One limitation of the MERCI protocol is its failure to measure the absolute amounts of identified
377 polypeptides. For instance, quantification of SWATH data (***Figure 1D-H***) measures the relative (to
378 reference proteins and each other) amounts of SUUR across fractions. To measure the absolute levels of
379 SUUR, a semi-quantitative approach was used by building a titration curve from SWATH acquisitions
380 of known amounts of recombinant SUUR (***Figure 1—figure supplement 1A&B***). We estimated the
381 amount of SUUR in the nuclear extract (~140 pg in 25 µg total protein, ***Figure 1—figure supplement***
382 ***1B***) and in individual fractions from all chromatographic steps (***Figure 1—figure supplement 1C***).
383 Although in five FPLC steps we achieved >3,000-fold purification of SUUR, it remained only ~2% pure
384 (***Figure 1—figure supplement 1D***). A progressive loss of material precludes further purification (300 ng
385 of SUUR in 16 µg total protein). Thus, the SUMM4 complex would be nearly impossible to purify to
386 homogeneity from a substantial amount of starting material (~1 kg *Drosophila* embryos, ~2.5 g protein),
387 suggesting that SUMM4 could not be identified by the classical FPLC approach.
388
389 **SUMM4 regulates the function of *gypsy* insulator elements**
390 Both subunits of SUMM4 contribute to the known functions of *gypsy* insulator (***Figure 5A-C***).
391 Although a *SuUR* mutation decreased the insulator activity, the suppression was universally weaker than
392 that by *mod(mdg4)*^{u1}. It is possible that SUUR is not absolutely required for the establishment of the
393 insulator. For instance, the loss of SUMM4 may be compensated by the alternative complex of
394 Mod(Mdg4)-67.2 (***Figure 4—figure supplement 2***). Furthermore, the *mod(mdg4)*^{u1} allele is expected to

have an antimorphic function, since it can mis-localize interacting partner proteins, including SUUR itself (***Figure 4C***). Interestingly, *SuUR* has been previously characterized as a weak suppressor of variegation of the *white*^{m4h} X chromosome inversion allele, which places the *white* gene near pericentric heterochromatin (Belyaeva et al., 2003). In contrast, *SuUR* mutation enhances variegation in the context of insulated, heterochromatin-positioned *white* (***Figure 5C***). Therefore, this phenotype is unrelated to the putative *Su(var)* function of *SuUR* but, rather, is insulator-dependent.

ATP-dependent motor proteins are required for the establishment of chromatin barrier and chromosome partitioning

Our discovery and analyses of SUMM4 provide a biochemical link between ATP-dependent motor factors and the activity of insulators in regulation of gene expression and chromatin partitioning. Insulator elements organize the genome into chromatin loops (Gerasimova et al., 1995) that are involved in the formation of topologically associating domains, TADs (Peterson, Samuelson, & Hanlon, 2021; Rowley et al., 2017; Szabo, Bantignies, & Cavalli, 2019). In mammals, CTCF-dependent loop formation requires ATP-driven motor activity of SMC complex cohesin (Davidson et al., 2019). In contrast, CTCF and cohesin are thought to be dispensable for chromatin 3D partitioning in *Drosophila* (Matthews & White, 2019). Instead, the larger, transcriptionally inactive domains (canonical TADs) are interspersed with smaller active compartmental domains, which themselves represent TAD boundaries (Rowley et al., 2017). It has been proposed that in *Drosophila*, domain organization does not rely on architectural proteins but is established by transcription-dependent, A-A compartmental (gene-to-gene) interactions (Rowley et al., 2017). However, *Drosophila* TAD boundaries are enriched for architectural proteins other than CTCF (Van Bortle et al., 2014), and their roles have not been tested in loss-of-function models. Thus, it is possible that in *Drosophila*, instead of CTCF, the 3D partitioning of the genome is facilitated by another group of insulator proteins, such as Su(Hw) and SUMM4 that together associate with class 3 insulators (Schwartz et al., 2012).

420

421 Moreover, SUUR may provide the DNA motor function to promote a physical separation of active
422 and inactive loci and help establish chromosome contact domains (***Figure 6A-C***). We propose that
423 within the SUMM4 complex, SUUR utilizes its putative ATP-dependent motor activity to translocate
424 along chromatin strands, thus facilitating the establishment of higher-order structures that isolate
425 promoters from enhancers (***Figure 6A***) and stabilize DNA loops/domains to prevent unrestricted
426 heterochromatin encroachment (***Figure 6B***) and penetration of replication forks (***Figure 6C***). The
427 translocation model is consistent with observations of an asymmetric, selective occupancy of SUUR
428 away from its initial sites of deposition via Su(Hw)-Mod(Mdg4) binding towards inside of intercalary
429 heterochromatin regions but not outside (***Figure 7—figure supplement 1C***) (Filion et al., 2010), which
430 may be facilitated by physical interactions between SUUR and linker histone H1 enriched in intercalary
431 heterochromatin (Andreyeva et al., 2017). It has been reported that another *Drosophila* BTB/POZ
432 domain insulator protein CP190 forms a complex with a DEAD-box helicase Rm62 that contributes to
433 the insulator activity (Lei & Corces, 2006). Thus, ATP-dependent motor proteins may represent an
434 obligatory component of the insulator complex machinery.

435

436 **SUMM4 mediates known biological functions of SUUR**

437 Our discovery explains previous observations about biological functions of SUUR. For instance, the
438 initial deposition of SUUR and its co-localization with PCNA has been proposed to depend on direct
439 physical interaction with components of the replisome (Kolesnikova et al., 2013). Our model indicates
440 that instead, the apparent co-localization of SUUR with PCNA throughout endo-S phase (***Figure***
441 ***4—figure supplement 3B***) may be caused by a replication fork retardation at insulator sites. SUUR is
442 deposited in chromosomes as a subunit of SUMM4 complex at thousands of loci by tethering via
443 Mod(Mdg4)-Su(Hw) interactions. As replication forks progress through the genome, they encounter
444 insulator complexes where replication machinery pauses for various periods of time before resolving the

445 obstacle. Thus, the increased co-residence time of PCNA and SUUR manifests cytologically as their
446 partial co-localization. With the progression of endo-S phase, some of the SUMM4 insulator complexes
447 are evicted and thus, the number of SUUR-positive loci is decreased, until eventually, the replication
448 fork encounters nearly completely impenetrable insulators demarcating the underreplicated domain
449 boundaries.
450
451 This mechanism is especially plausible given that boundaries of intercalary heterochromatin loci
452 very frequently encompass multiple, densely clustered Su(Hw) binding sites (*e.g.*, ***Figure 7C&D***). We
453 examined the data from genome-wide proteomic analyses for Su(Hw) and SUUR performed by DamID
454 in Kc167 cells (Filion et al., 2010). Strikingly, Su(Hw) DamID-measured occupancy does not exhibit a
455 discrete pattern expected of a DNA-binding factor. Instead, it appears broadly dispersed, together with
456 SUUR, up to tens of kbp away from mapped Su(Hw) binding sites (***Figure 7—figure supplement 1C***).
457 Interestingly, when Hidden Markov modeling was applied to the DamID data, Su(Hw), Mod(Mdg4)-
458 67.2 and SUUR occupancies were found to strongly correlate genome-wide in a novel chromatin form
459 ("malachite") that frequently demarcates the boundaries of intercalary heterochromatin (Khoroshko et
460 al., 2016). These observations strongly corroborate the translocation model for the mechanism of action
461 of SUMM4. According to this model, upon tethering to DNA-bound Su(Hw), SUMM4 traverses the
462 underreplicated region, which helps to separate it in a contact domain. As DNA within the
463 underreplicated region is tracked by SUUR (***Figure 6C***), it is brought into a transient close proximity
464 with both SUMM4 and the associated Su(Hw) protein, which is detected by DamID (or ChIP) as an
465 expanded occupancy pattern.
466
467 The deceleration of SUUR-bound replication forks was also invoked as an explanation for the
468 apparent role of SUUR in the establishment of epigenetic marking of intercalary heterochromatin
469 (Posukh et al., 2015). We propose that global epigenetic modifications observed in the *SuUR* mutant

likely do not directly arise from derepression of the replisome as suggested but, rather, result from the coordinate insulator-dependent regulatory functions of SUUR in both the establishment of a chromatin barrier and DNA replication control (*cf* ***Figure 6B&C***).

Architectural proteins can attenuate replication forks and regulate replication timing

Our work demonstrates for the first time that insulator complexes assembled on chromatin can attenuate the extent of replication in discrete regions of the salivary gland polylploid genome. Despite distinct cell cycle programs in dividing and endoreplicating cells (Zielke et al., 2013), the core biochemical composition of replisomes in both cell types is likely similar. Although the putative relationship is limited by a paucity of comparative biochemical analyses of replication factors in different cell types, related insulator-driven control mechanisms for DNA replication may be conserved in endoreplicating and mitotically dividing diploid cells. Our data thus implicates insulator/chromatin boundary elements as a critical attribute of DNA replication control. Our model suggests that delayed replication of repressed chromatin (*e.g.*, intercalary heterochromatin) during very late S phase can be imposed in a simple, two-component mechanism (***Figure 6C***). First, it requires that an extended genomic domain be 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 is 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.

An important frequent feature of the partially suppressed underreplication in *mod(mdg4)* animals is its asymmetry (***Figure 7D***, ***Figure 7—figure supplement 1B***), which is consistent with a unidirectional penetration of the underreplicated domain by a replication fork firing from the nearest external origin (***Figure 6C***). The SUMM4-dependent barrier may be created as a direct physical obstacle to MCM2-7

495 DNA-unwinding helicase or other enzymatic activities of the replisome. Alternatively, SUMM4 may
496 inhibit the replication machinery indirectly by assembling at the insulator a DNA/chromatin structure
497 that is incompatible with replisome translocation. This putative inhibitory structure may involve
498 epigenetic modifications of chromatin as proposed earlier (Gaszner & Felsenfeld, 2006), linker histone
499 H1 as shown previously (Andreyeva et al., 2017) and may also be dependent on Rif1, a negative DNA
500 replication regulator that acts downstream of SUUR (Munden et al., 2018).
501
502 In conclusion, we used a newly developed MERCI approach to identify a stable stoichiometric
503 complex termed SUMM4 that comprises SUUR, a previously known negative effector of replication,
504 and Mod(Mdg4), an insulator protein. SUMM4 subunits cooperate to mediate transcriptional
505 repression and chromatin boundary functions of *gypsy*-like (class 3) insulators (Schwartz et al., 2012)
506 and inhibit DNA replication likely by slowing down replication fork progression through the boundary
507 element. Thus, SUMM4 is required for coordinate regulation of gene expression, chromatin
508 partitioning and DNA replication timing. The insulator-dependent regulation of DNA replication
509 offers a novel mechanism for the establishment of replication timing in addition to the currently
510 accepted paradigm of variable timing of replication origin firing.

511 # Materials and Methods

512 **Recombinant proteins**

513 Recombinant proteins were expressed in Sf9 cells using baculovirus system (SUUR, Mod(Mdg4),
514 EGG and WDE), in *E. coli* (ISWI, ModT antigen and LCMS reference proteins) or obtained from
515 EpiCypher Inc. (human BRG1/SMARCA4).

516

517 *Sf9 cells*

518 All baculovirus constructs were cloned by PCR with Q5 DNA polymerase (New England Biolabs)
519 and ligation or Gibson assembly with NEBuilder HiFi DNA Assembly Cloning kit (New England
520 Biolabs) into pFastBac vector (Thermo Fisher) under control of polyhedrin promoter. All constructs
521 were validated by Sanger sequencing. Baculoviruses were generated according to the protocol by
522 Thermo Fisher. The baculoviruses were isolated by plaque purification, amplified three times, and their
523 titers were measured by plaque assay. FLAG-SUUR construct was cloned from *SuUR-RA* cDNA
524 (LD13959, DGRC). The following open reading frame (ORF) was expressed: M**DYKDDDDK**H-SUUR-
525 PA(1..962)-VEACGTKLVEKY*. To generate ATPase-dead mutant, SUUR-PA(K59) codon was replaced
526 with an alanine codon by PCR and Gibson cloning. Mod(Mdg4)-67.2-V5-His_6 and Mod(Mdg4)-59.1-
527 V5-His_6 constructs were cloned from cDNAs *mod(mdg4)-RT* and *mod(mdg4)-RI* synthesized as gBlocks
528 by IDT, Inc. The following ORFs were expressed: Mod(Mdg4)-67.2(1..610)-
529 GIL**EGKPIPNPLLGLDST**GASVEHHHHHH* and Mod(Mdg4)-59.1(1..541)-
530 GIL**EGKPIPNPLLGLDST**GASVEHHHHHH*. EGG-FLAG and EGG (untagged) were cloned by PCR
531 from *egg-RA* cDNA (IP14531). The following ORF was expressed: EGG-PA(1..1262)-**DYKDDDDK***
532 and EGG-PA(1..1262)-*. FLAG-WDE was cloned by PCR from *wde-RA* cDNA (LD26050). The
533 following ORF was expressed: M**DYKDDDDK**-WDE-PA(2..1420)-*. The sequences of FLAG and V5
534 tags are highlighted in bold typeface.

535

536 Cells, $2 \cdot 10^6$/ml in Sf-900 II SFM medium (Gibco), were infected at multiplicity of infection (MOI)
537 of ~10 in PETG shaker flasks (Celltreat, Inc.). After infection for 48-72 hours at 27°C, cells were
538 harvested, and recombinant proteins were purified by FLAG or Ni-NTA affinity chromatography
539 (Fyodorov & Kadonaga, 2003). Whereas, typically, amplified baculovirus stocks had titers above $5 \cdot 10^9$
540 pfu/ml, FLAG-SUUR viruses reached no more than $2\text{-}4 \cdot 10^8$ pfu/ml, presumably, due to the inhibitory
541 effect of over-expressed protein on viral DNA replication. Accordingly, whereas typical yields of
542 purified recombinant proteins were >100 µg from 1 L Sf9 cell culture, SUUR polypeptides were
543 produced at no more than 2 µg from 1 L culture, which also adversely affected the protein purity
544 (***Figure 1C&3A***, ***Figure 3—figure supplement 1A&B***).

545
546 *E. coli*

547 The expression construct for untagged recombinant *Drosophila* ISWI was prepared from a full-
548 length ISWI cDNA (Ito et al., 1999). Human TXNRD1 sequence was cloned from a cDNA provided by
549 Addgene (#38863), and TXNRD2 was synthesized as a gBlock gene fragment by IDT, Inc. The ORFs
550 were inserted by Gibson cloning in a pET backbone vector in frame with a C-terminal intein-CBD
551 (chitin-binding domain) tag. Protein expression was induced by IPTG in Rosetta 2 cells, and proteins
552 were purified in non-denaturing conditions by chitin affinity chromatography and intein self-cleavage as
553 described (Emelyanov et al., 2014), followed by anion exchange chromatography (Source 15Q) on
554 FPLC (see below). Note that the cloned human thioredoxin reductase ORFs do not express the C-
555 terminal selenocysteines. They were thus presumed catalytically inactive (Arner, Sarioglu, Lottspeich,
556 Holmgren, & Bock, 1999; Cheng & Arner, 2017) and designated hTXNRD1ci and hTXNRD2ci. They
557 were used exclusively as spike-in mass standards in LCMS acquisitions of *Drosophila* proteins.

558
559 Polypeptide corresponding to the C-terminal specific region of Mod(Mdg4)-67.2 was cloned in
560 pET24b vector in frame with a C-terminal His_6 tag. `M`-Mod(Mdg4)-67.2(403..610)-`GILEHHHHHH*` was

561 expressed in Rosetta 2 and purified by Ni-NTA affinity chromatography in non-denaturing conditions.
562 The polypeptide (ModT) was dialyzed into PBS (137 mM NaCl, 3 mM KCl, 8 mM NaH_2PO_4, 2 mM
563 KH_2PO_4) and used as an antigen for immunizations (see below). All recombinant proteins were
564 examined by SDS-PAGE along with Pierce BSA mass standards (Thermo Fisher), and their
565 concentrations were calculated from infrared scanning of Coomassie-stained gels (Odyssey Fc Imaging
566 System, LI-COR Biosciences). Detailed cloning and purification methods are provided below.
567

568 **Molecular cloning**

569 pFastBac-FLAG-SUUR

570 The coding sequence was amplified from LD13959 by PCR using the following primers: NdeI-
571 SUURf, TCCATATGTATCACTTTGTATCCGAGCAAAC and Sal1-SUURr,
572 AAGTCGACCTTGAACAGTTCCAATCGCTTTC (NdeI and SalI restriction sites are underlined). The PCR
573 product was digested with NdeI and SalI and ligated with the vector produced by NdeI-XhoI digestion
574 of pFastBac-Flag-ATRX construct (Emelyanov, Konev, Vershilova, & Fyodorov, 2010).
575

576 pFastBac-FLAG-SUUR(K59A)

577 The complete pFastBac-FLAG-SUUR construct was amplified by PCR using the following primers:
578 SUUR-KAf, CTTGGGCAGGTCGCTACGGTGGCGG and SUUR-KAr,
579 GTAGCGACCTGCCCAAGGCCACTCTCATCATTCAGG (mutated residues are underlined). The linear PCR
580 product was re-circularized by Gibson assembly.
581

582 pFastBac-Mod(Mdg4)-67.2-V5-His_6

583 The following gBlock (MMD4-RT) was synthesized by IDT, Inc.:
584 CGAAGCGCGCGGAATTCAT**ATG**GCCGATGACGAACAGTTTTCGCTGTGCTGGAACAACTTTAACACAAAT
585 TTGTCGGCAGGATTTCACGAGAGTCTCTGTCGGGGCGACTTGGTAGACGTCTCCTTGGCAGCAGAGGGAC

586 AAATTGTCAAGGCCCATCGTCTGGTACTCTCCGTCTGCAGCCCATTTTTTCGGAAAATGTTCACTCAGAT

587 GCCAAGCAACACTCACGCCATAGTATTTCTGAACAATGTTAGTCACAGCGCTTTGAAAGATCTGATCCAA

588 TTTATGTATTGTGGCGAAGTGAACGTTAAGCAAGACGCATTGCCGGCATTTATCTCCACTGCAGAAAGTC

589 TGCAAATTAAAGGATTGACCGATAACGACCCAGCTCCGCAACCCCCACAAGAGAGCTCGCCACCTCCCGC

590 TGCGCCTCATGTGCAGCAACAGCAAATCCCAGCCCAGCGGGTGCAACGACAACAGCCGCGTGCTAGCGCC

591 CGCTATAAAATTGAGACTGTGGATGATGGACTGGGCGACGAAAAACAAAGTACCACTCAGATTGTTATCC

592 AAACAACAGCTGCCCCGCAAGCAACTATTGTTCAACAACAACAGCCTCAACAAGCTGCACAACAAATACA

593 GTCGCAACAGTTGCAGACAGGTACAACAACAACTGCAACATTGGTAAGTACTAATAAGAGGAGTGCTCAG

594 CGCTCGTCCCTGACGCCGGCGTCCAGTAGTGCGGGTGTTAAAAGGAGTAAGACAAGCACTAGCGCAAACG

595 TGATGGATCCGCTGGATTCGACTACGGAGACAGGCGCAACTACAACGGCTCAACTGGTACCTCAGCAAAT

596 CACTGTACAAACATCCGTTGTCAGCGCTGCTGAGGCGAAGCTCCATCAGCAGAGTCCCCAACAGGTTCGC

597 CAGGAAGAGGCGGAGTATATAGATCTGCCTATGGAGCTGCCGACCAAGTCGGAACCGGATTACTCGGAAG

598 ATCATGGCGACGCGGCCGGTGACGCTGAGGGTACGTATGTCGAGGATGATACGTACGGTGACATGCGATA

599 CGACGATTCCTATTTTACAGAAAATGAGGACGCAGGCAACCAGACGGCCGCCAATACAAGCGGAGGTGGC

600 GTGACAGCGACCACTAGCAAAGCTGTTGTGAAACAACAGTCGCAGAACTATTCGGAGAGTAGTTTCGTAG

601 ATACCAGTGGCGACCAAGGTAACACCGAGGCACAGGCAGCCACAAGTGCTTCGGCGACCAAGATTCCGCC

602 CCGGAAACGGGGTCGACCGAAAACAAAAGTTGAGGACCAGACCCCTAAACCTAAATTGCTtGAGAAGTTG

603 CAGGCCGCAACACTGAACGAGGAAGCAAGTGAACCGGCCGTATATGCGTCGACCACGAAAGGCGGTGTTA

604 AACTGATATTTAACGGCCATTTGTTTAAATTCTCGTTTAGGAAAGCGGATTACAGTGTCTTCCAGTGTTG

605 TTATAGGGAGCATGGTGAAGAGTGCAAGGTCAGGGTCGTCTGCGATCAAAAGCGTGTATTTCCTTACGAG

606 GGTGAACACGTGCACTTCATGCAAGCTTCCGATAAGTCCTGCCTCCCTAGTCAGTTCATGCCAGGTGAGT

607 CCGGTGTCATTTCCAGTTTGAGCCCATCGAAAGAGCTCTTGATGAAGAATACCACTAAGCTCGAAGAGGC

608 GGATGATAAGGAAGACGAAGATTTCGAAGAGTTTGAGATCCAAGAAATAGACGAGATAGAATTGGACGAA

609 CCGGAGAAGACCCCCGCAAAGGAAGAAGAAGTTGACCCGAACGACTTTCGGGAGAAGATTAAGCGACGGC

610 TCCAGAAGGCCTTGCAAAACAAAAAG**AAA**GGAATTCTCGAGGGTAAGCCTATCCCTAACCCTCTCCTCGG

611 TCTCGATTCTACCGGTGCTAGCGTCGAGCACCACCACCACCACCAC**TGA**GATCCGGCTGCTAAC

612 (sequence coding for V5 tag is underlined; translation initiation/termination codons and codon 610 of

613 *mod(mdg4)-RT* are shown in bold). The vector fragment was amplified by PCR from pFastBac by using

614 the following primers: His-Stop-Vf, CAC**TGA**GATCCGGCTGCTAAC and NdeI-Vr,

615 **CAT**ATGAATTCCGCGCGCTTC. The expression construct was assembled by Gibson cloning.

616

617 pFastBac-Mod(Mdg4)-59.1-V5-His_6

618 The following gBlock (MMD4-RI) was synthesized by IDT, Inc.:

619 GGTAACACCGAGGCACAG**GTATGTGATGATCTCGATGACATGAAAGGCGCTATTAAGCATAGCCTGTTGA**

620 **CTTTTATTCGCGGTCAGCGCGGCTGCAAACTGCTGGCTTTTAACGGTCATAATTATGTTCGTAACAGGCG**

621 **TTCCAATCTCAAGACGTATTGGATATGCAGCAAAAAAGGCAGCACTAAATGCAACGCTCGTGTTGTTACA**

622 **AACGTAGTTGAGGGTGTTCACAAGATAGTTCTGGAAAGTTGCCATCATACGTGTCTGAACACCGAGAGGA**

623 **AGAAAAGGCTCTCGGTGACTAATGTAGTAGGAAAAGCGCGGTCGAAGTCCGAAAAAAGTGTATCCACGGG**

624 **CTTTATTAAAGAAGAAGGAGACGAGGACCTCACGTTGGAATTGCGGACCCTCAACCTGTCGATTGAGGAT**

625 **CTGAATAACCTCCAG**GGAATTCTCGAGGGTAAGCC (sequence corresponding to V5 tag is underlined;

626 variant-specific codons 403-541 of *mod(mdg4)-RI* are shown in bold). The vector fragment additionally

627 encompassing *mod(mdg4)* codons 1-402 were amplified by PCR from pFastBac-Mod(Mdg4)-67.2-V5-

628 His_6 by using the following primers: GIL-V5f, GGAATTCTCGAGGGTAAGCC and MMD397-402r,

629 **C**CTGTGCCTCGGTGTTACC. The expression construct was assembled by Gibson cloning.

630

631 pFastBac-EGG (untagged)

632 pFastBac-ATRX (untagged) construct (Emelyanov, Konev, Vershilova, & Fyodorov, 2010) was

633 digested with EcoRI and XhoI. The vector fragment (4.7 kbp) was ligated with a 4-kbp EcoRI-XhoI

634 fragment of *egg-RA* cDNA (IP14531).

635

636 pFastBac-EGG-FLAG

637 Double-stranded oligonucleotide was produced by annealing ApaI-FLAG-AflII-f,

638 CCCAATTGCCGCCTTCGTCTGCTC**GATTACAAGGATGATGATGACAAA*TAA***C and AflII-FLAG-ApaI-r,

639 TTAAG***TTA*TTTGTCATCATCATCCTTGTAATC**GAGCAGACGAAGGCGGCAATTGGGGGCC (sticky ends
640 are underlined; sequences corresponding to FLAG tag are shown in bold; stop codon is in bold and
641 italics) was cloned into ApaI-AflII-digested IP14531 by ligation. The resulting construct was digested
642 with EcoRI and XhoI, and the 4-kbp EGG-FLAG fragment was cloned into pFastBac as described
643 above.
644
645 pFastBac-FLAG-WDE
646 pFastBac-ATRX (untagged) construct (Emelyanov, Konev, Vershilova, & Fyodorov, 2010) was
647 digested with NdeI and NcoI. The vector fragment additionally encompassing 1.1 kbp of ATRX cDNA
648 sequence with a XhoI site (5.8 kbp total) was ligated with a double-stranded oligonucleotide produced
649 by annealing NdeI-FLAG-NcoI-f, T**ATGGATTACAAGGATGATGATGACAAA**ATGGGAGTAAACCAGAC
650 and NcoI-FLAG-NdeI-r, CATGGTCTGGTTTACTCCCAT**TTTGTCATCATCATCCTTGTAATCCA** (sticky
651 ends are underlined; sequences corresponding to FLAG tag are shown in bold). A 4.6-kbp NcoI-XhoI
652 fragment of *wde-RA* cDNA (LD26050) was cloned in the resulting construct by restriction digest and
653 ligation.
654
655 pET24-ISWI-intein-CBD
656 ISWI cDNA was amplified from pFastBac-ISWI construct (Ito et al., 1999) by PCR using the
657 following primers: NdeI-ISWIf, GTTTCAT**ATGGCTAGCAAAACAGATAC** and XhoI-ISWIr,
658 GGAAGGTACCCTTGGCAAAGCA**CCCCTTCTTCTTCTTTTTC** (NdeI and XhoI sites are underlined;
659 sequences corresponding to the ISWI ORF are shown in bold). The 3.1-kbp PCR fragment was digested
660 with NdeI and XhoI and cloned into pET24-intein-CBD construct in place of Protamin B (Emelyanov et
661 al., 2014) by ligation.
662
663 pET24-hTXNRD1ci-intein-CBD

664 Human TXNRD1 cDNA (Addgene #38863) was amplified by PCR using the following primers:
665 NdeI- hTXNRD1f, AACAT**ATGAACGGCCCTGAAGATCTTC** and SalI- hTXNRD1r,
666 TAGTCGACG**CAGCCAGCCTGGAGG** (NdeI and SalI sites are underlined; sequences corresponding to the
667 TXNRD1 ORF are shown in bold). The 1.5-kbp PCR fragment was digested with NdeI and SalI and
668 cloned into NdeI and XhoI sites of pET24-intein-CBD construct in place of Protamin B (Emelyanov et
669 al., 2014) by ligation.
670
671 pET24-hTXNRD2ci- intein-CBD
672 The following gBlock (TXNRD2) was synthesized by IDT, Inc.:
673 TTTTCAT**ATG**GAAGATCAGGCGGGCCAGCGCGATTATGATCTGCTGGTGGTGGGCGGCGGCAGCGGCGGC
674 CTGGCGTGCGCGAAAGAAGCGGCGCAGCTGGGCCGCAAAGTGGCGGTGGTGGATTATGTGGAACCGAGCC
675 CGCAGGGCACCCGCTGGGGCCTGGGCGGCACCTGCGTGAACGTGGGCTGCATTCCGAAAAAACTGATGCA
676 TCAGGCGGCGCTGCTGGGCGGCCTGATTCAGGATGCGCCGAACTATGGCTGGGAAGTGGCGCAGCCGGTG
677 CCGCATGATTGGCGCAAAATGGCGGAAGCGGTGCAGAACCATGTGAAAAGCCTGAACTGGGGCCATCGCG
678 TGCAGCTGCAGGATCGCAAAGTGAAATATTTTAACATTAAAGCGAGCTTTGTGGATGAACATACCGTGTG
679 CGGCGTGGCGAAAGGCGGCAAAGAAATTCTGCTGAGCGCGGATCATATTATTATTGCGACCGGCGGCCGC
680 CCGCGCTATCCGACCCATATTGAAGGCGCGCTGGAATATGGCATTACCAGCGATGATATTTTTTGGCTGA
681 AAGAAAGCCCGGGCAAAACCCTGGTGGTGGGCGCGAGCTATGTGGCGCTGGAATGCGCGGGCTTTCTGAC
682 CGGCATTGGCCTGGATACCACCATTATGATGCGCAGCATTCCGCTGCGCGGCTTTGATCAGCAGATGAGC
683 AGCATGGTGATTGAACATATGGCGAGCCATGGCACCCGCTTTCTGCGCGGCTGCGCGCCGAGCCGCGTGC
684 GCCGCCTGCCGGATGGCCAGCTGCAGGTGACCTGGGAAGATAGCACCACCGGCAAAGAAGATACCGGCAC
685 CTTTGATACCGTGCTGTGGGCGATTGGCCGCGTGCCGGATACCCGCAGCCTGAACCTGGAAAAAGCGGGC
686 GTGGATACCAGCCCGGATACCCAGAAAATTCTGGTGGATAGCCGCGAAGCGACCAGCGTGCCGCATATTT
687 ATGCGATTGGCGATGTGGTGGAAGGCCGCCCGGAACTGACCCCGACCGCGATTATGGCGGGCCGCCTGCT
688 GGTGCAGCGCCTGTTTGGCGGCAGCAGCGATCTGATGGATTATGATAACGTGCCGACCACCGTGTTTACC
689 CCGCTGGAATATGGCTGCGTGGGCCTGAGCGAAGAAGAAGCGGTGGCGCGCCATGGCCAGGAACATGTGG
690 AAGTGTATCATGCGCATTATAAACCGCTGGAATTTACCGTGGCGGGCCGCGATGCGAGCCAGTGCTATGT

691 GAAAATGGTGTGCCTGCGCGAACCGCCGCAGCTGGTGCTGGGCCTGCATTTTCTGGGCCCGAACGCGGGC
692 GAAGTGACCCAGGGCTTTGCGCTGGGCATTAAATGCGGCGCGAGCTATGCGCAGGTGATGCGCACCGTGG
693 GCATTCATCCGACCTGCAGCGAAGAAGTGGTGAAACTGCGCATTAGCAAACGCAGCGGCCTGGATCCGAC
694 CGTGACCGGC**TGC**CTCGAGTTTTTTTTTT (NdeI and XhoI sites are underlined; translation initiation
695 codon and codon 492 of hTXNRD2 are shown in bold). The DNA fragment was digested with NdeI and
696 XhoI and cloned by ligation in pET24-intein-CBD as described above.
697
698 pET24-ModT-His_6
699 Mod(Mdg4)-67.2-specific fragment of *mod(mdg4)-RT* cDNA was amplified from pFastBac-
700 Mod(Mdg4)-67.2-V5-His_6 by PCR using the following primers: NdeI-ModTf,
701 CCGAGCATATGGCAGCCACAAGTGCTTC and XhoI-ModTr, GGGTAGGCTTACCCTCGAGAATTCCTTTC
702 (NdeI and XhoI sites are underlined). The 0.6-kbp PCR fragment was digested with NdeI and XhoI and
703 cloned in pET24b (Millipore/Sigma) by ligation.
704
705 **FPLC purification of recombinant ISWI, hTXNRD1ci and hTXNRD2ci**
706 Protein samples eluted from the chitin resin (1-5 ml total sample volume) were diluted 3-fold with
707 chromatographic Buffer A (***Figure 1—source data 1***) and injected on a 0.5-ml Source 15Q equilibrated
708 to 5% Buffer B (***Figure 1—source data 1***) + 95% Buffer A. The column was washed with 20 *cv*
709 (column volumes) of 5% Buffer B, and proteins were eluted with a 20 *cv* linear gradient of 5-100%
710 Buffer B. 200-µl fractions were collected and analyzed by SDS-PAGE. Three to five peak fractions
711 were pooled, aliquoted, flash-frozen in liquid nitrogen and stored at -80°C.
712
713 **Crude cell extracts**
714 *Nuclear extract from* Drosophila *embryos*

~1 kg or ~200 g wild-type (Oregon R) *Drosophila* embryos were collected 0-12 h after egg deposition (AED) from population cages. The embryos were dechorionated, and nuclear extracts were prepared as described (Kamakaka, Tyree, & Kadonaga, 1991). Protein concentration was measured by Pierce BCA assay (Thermo Fisher). The extracts were fractionated by FPLC (***Figure 1A*** and ***Figure 4—figure supplement 2A***) on AKTA PURE system (Cytiva Life Sciences). Aliquots of chromatographic fractions were examined by quantitative shotgun proteomics or western blot analyses as described below. Peak SUUR or Mod(Mdg4) fractions were diluted to an appropriate ionic strength (if applicable) and used as a starting material for the next chromatographic step. Details on FPLC column sizes and run parameters are shown in ***Figure 1—source data 1***, ***Figure 4—figure supplement 2—source data 1***.

E. coli *lysate*

40-ml Rosetta 2 overnight culture was harvested by centrifugation, resuspended in 20 ml HEG (25 mM HEPES, pH 7.6, 0.1 mM EDTA, 10% glycerol) supplemented with 0.1 M KCl, 1 mM DTT and 2 mM $CaCl_2$. Cells were disrupted by sonication and centrifuged to remove insoluble material. Nucleic acids were digested with 15 units micrococcal nuclease (Sigma Aldrich) for 20 min at 37°C, and the proteins were precipitated with 2 M ammonium sulfate. The pellet was resuspended in 10 ml HEG + 0.1 M KCl + 1 mM DTT with protease inhibitors (0.5 mM benzamidine, 0.2 mM PMSF) and dialyzed against the same buffer. After centrifugation, the concentration of soluble protein was measured by BCA assay, the *E. coli* lysate was diluted to 1 mg/ml using 100 mM ammonium bicarbonate (ABC) and stored at -80°C.

Mass-spectroscopy samples

Column fractions

739 For each chromatographic step, 14 to 20 fractions were selected based on the protein fractionation
740 profile according to the UV (A_{280}) absorbances measurements. 50-100 µl aliquots of chromatographic
741 fractions, starting material (SM) and column flow-through (FT, if applicable) were saved, and protein
742 concentrations were estimated based on their UV absorbances (1,000 mU A_{280} was considered to be
743 equivalent to 5 mg/ml total protein). Equal volumes of each fraction, SM and FT were used for MS
744 acquisitions, so that no more than 40 µg total protein was processed in each reaction. As a reference, the
745 reactions were supplemented with 1.5 µg each of purified recombinant human thioredoxin reductases 1
746 and 2 (hTXNRD1ci and hTXNRD2ci, catalytically inactive) expressed in *E. coli*. Dithiotreitol (DTT)
747 was added to the protein samples to 10 mM and NP-40 – to 0.02%. Reaction volumes were brought to
748 85 µl with 50 mM ammonium bicarbonate (ABC). All reagents, including water, were HPLC/MS grade.
749 The proteins were reduced for 1 h at 37°C and then alkylated with 30 mM iodoacetamide (IAA, 15 µl
750 200 mM IAA in water) for 45 min at room temperature in the dark. Alkylated proteins were desalted
751 into 50 mM ABC using ZebaSpin columns (40 kDa MWCO) and digested with 1 µg trypsin for 2 h at
752 37°C. 1 µg more trypsin was added, and the digestion progressed at 37°C overnight. Tryptic peptides
753 were lyophilized for 2 h on SpeedVac with heat and resuspended in 100 µl Sample Buffer: 1%
754 acetonitrile (ACN) and 0.1% formic acid (FA) in water. Equal volumes (23 µl) of samples were used for
755 IDA and SWATH acquisitions (in triplicate) as described below.

756

757 *Recombinant SUUR*

758 To generate the recombinant SUUR reference spectral library (ILR), ~0.5 µg purified recombinant
759 FLAG-SUUR (both 130 and 65 kDa bands, ***Figure 1C***) was mixed with 1.5 µg each of hTXNRD1ci and
760 hTXNRD2ci and processed for an IDA acquisition as described above, except for 0.5 µg trypsin was
761 used in each cleavage step, and the peptide sample was resuspended in 30 µl Sample Buffer. For
762 SWATH titration of SUUR (***Figure 1—figure supplement 1B***), 1 µg recombinant FLAG-SUUR was
763 mixed with 25 µg *E. coli* lysate protein and 1.5 µg each of hTXNRD1ci and hTXNRD2ci. 10-fold serial

764 dilutions down to 10 fg SUUR were also prepared using the mixture of *E. coli* lysate with reference
765 proteins. The samples were processed for SWATH acquisitions in triplicate as described above, 30 µl of
766 sample per injection.
767
768 *In-gel digestion of recombinant proteins for LCMS identification*
769 Recombinant SUUR or SUMM4 purified by FLAG immunoaffinity chromatography was resolved
770 on SDS-PAGE, stained with Coommassie Blue (***Figure 1C&3A, Figure 3—figure supplement 1A***), and
771 up to eight most prominent protein bands were excised. The gel slices were transferred to 1.5-ml
772 Eppendorf tubes, gently crushed with a RotoDounce pestle and destained with 25 mM ABC in 50%
773 methanol and then with 25 mM ABC in 50% ACN (30 min each at room temperature). The proteins
774 were reduced in 50 µl 10 mM DTT for 1 h at 55°C and alkylated with 30 mM IAA for 45 min at room
775 temperature in the dark. The gel fragments were washed with 25 mM ABC in 50% ACN, dehydrated
776 with 100% ACN, dried in a SpeedVac, rehydrated by addition of 50 µl 50 mM ABC and digested with
777 0.25 µg trypsin overnight at 37°C. The peptides were extracted once with 50 µl 10% FA and once with
778 100 µl 3% FA in 60% ACN, both extracts were combined, dried in a SpeedVac and resuspended in 50
779 µl Sample Buffer. 40 µl of each sample was injected for IDA acquisitions as described below.
780
781 **Mass-spectroscopy acquisition methods**
782 LC-MS/MS analyses were performed on a TripleTOF 5600+ mass spectrometer (AB SCIEX)
783 coupled with M5 MicroLC system (AB SCIEX/Eksigent) and PAL3 autosampler.
784
785 *Instrument settings*
786 LC separation was performed in a trap-elute configuration, which consists of a trapping column
787 (LUNA C18(2), 100 Å, 5 µm, 20 × 0.3 mm cartridge, Phenomenex) and an analytical column (Kinetex
788 2.6 µm XB-C18, 100 Å, 50 × 0.3 mm microflow column, Phenomenex). The mobile phase consisted of

water with 0.1% FA (phase A) and 100% ACN containing 0.1% FA (phase B). 200 ng to 10 µg total protein was injected for each acquisition. Peptides in Sample Buffer were injected into a 50-µl sample loop, trapped and cleaned on the trapping column with 3% mobile phase B at a flow rate of 25 µl/min for 4 min before being separated on the analytical column with a gradient elution at a flow rate of 5 µl/min. The gradient was set as follows: 0 to 48 min: 3% to 35% phase B, 48 to 54 min: 35% to 80% phase B, 54 to 59 min: 80% phase B, 59 to 60 min: 80% to 3% phase B, and 60 to 65 min at 3% phase B. An equal volume of each sample (23 µl) was injected four times, once for information-dependent acquisition (IDA), immediately followed by data-independent acquisition (DIA/SWATH) in triplicate. Acquisitions of distinct samples were separated by a blank injection to prevent sample carry-over. The mass spectrometer was operated in positive ion mode with EIS voltage at 5,200 V, Source Gas 1 at 30 psi, Source Gas 2 at 20 psi, Curtain Gas at 25 psi and source temperature at 200°C.

Information-dependent acquisitions (IDA) and data analyses

IDA was performed to generate reference spectral libraries for SWATH data quantification. The IDA method was set up with a 250-ms TOF-MS scan from 400 to 1250 Da, followed by MS/MS scans in a high sensitivity mode from 100 to 1500 Da of the top 30 precursor ions above 100 cps threshold (100 ms accumulation time, 100 ppm mass tolerance, rolling collision energy and dynamic accumulation) for charge states (*z*) from +2 to +5. IDA files were searched using ProteinPilot (version 5.0.2, ABSciex) with a default setting for tryptic digest and IAA alkylation against a protein sequence database. The *Drosophila* proteome FASTA file (21,970 protein entries, UniProt UP000000803, 3/21/2020) augmented with sequences for common contaminants as well as hTXNRD1 and hTXNRD2 was used as a reference for the search. Up to two missed cleavage sites were allowed. Mass tolerance for precursor and fragment ions was set to 100 ppm. A false discovery rate (FDR) of 5% was used as the cutoff for peptide identification.

SWATH acquisitions and data analyses

For SWATH (SWATH-MS, Sequential Window Acquisition of All Theoretical Mass Spectra) acquisitions (Zhu, Chen, & Subramanian, 2014), one 50-ms TOF-MS scan from 400 to 1250 Da was performed, followed by MS/MS scans in a high sensitivity mode from 100 to 1500 Da (15-ms accumulation time, 100 ppm mass tolerance, +2 to +5 z, rolling collision energy) with a variable-width SWATH window (Zhang et al., 2015). DIA data were quantified using PeakView (version 2.2.0.11391, ABSciex) with SWATH Acquisition MicroApp (version 2.0.1.2133, ABSciex) against selected spectral libraries generated in ProteinPilot. Retention times for individual SWATH acquisitions were calibrated using 20 or more peptides for hTXNRD1ci and hTXNRD2ci. The following software settings were utilized: up to 25 peptides per protein, 6 transitions per peptide, 95% peptide confidence threshold, 5% FDR for peptides, XIC extraction window 20 minutes, XIC width 100 ppm. Protein peak areas were exported as Excel files (***Supplementary File 2***) and processed as described below.

MERCI

MERCI is a novel approach for rapid identification of native protein complexes. It combines enrichment for a target subunit of a putative complex by consecutive FPLC steps and quantitative shotgun proteomics of chromatographic fractions. Crude nuclear extract from *Drosophila* embryos was fractionated as in ***Figure 1A***, ***Figure 1—source data 1***. At every step, 40 µg or less total protein from each of 10-20 fractions (equal volumes) was supplemented with a fixed amount (1.5 µg each) of exogenous reference proteins (human thioredoxin reductases), reduced, alkylated and digested with trypsin (see above). MS1 and MS2 spectra of tryptic peptides were acquired by IDA, and relative SUUR abundance in fractions was measured by data-independent acquisition (DIA/SWATH) in triplicate. SWATH data were quantified using cognate IDA-derived ion libraries. Protein areas for all quantified proteins were normalized to the sum of those for reference proteins. The relative numbers were averaged across triplicates, with standard deviations calculated. The average numbers for all quantified proteins

839 were further normalized by converting them to Z-scores (see ***Supplementary File 2*** for an example of
840 calculations). Peak SUUR fractions (one to five) were then subjected to the next FPLC/MERCI step.
841 After five column steps, the ion library from the ultimate FPLC step (IL5) was used to re-quantify
842 SWATH data from all steps. Z-scores for all purification steps were stitched together, and the large array
843 encompassing all data points for every protein was analyzed by Pearson correlation with SUUR
844 (***Supplementary File 2***). The most closely correlated purification profiles served as an indication for
845 protein co-purification, potentially, as subunits of a stable complex.
846
847 **Biochemical assays with recombinant proteins**
848 *Oligonucleosome substrates*
849 Oligonucleosomes were reconstituted *in vitro* as described (Lu et al., 2013) from supercoiled
850 plasmid DNA (3.2 kb, pGIE-0), native core histones and H1 prepared from *Drosophila* embryos
851 (Fyodorov & Levenstein, 2002) by gradient salt dialysis in the presence of 0.2 mg/ml nuclease-free
852 bovine serum albumin (BSA, New England Biolabs). Quality of reconstitution was assessed by SDS-
853 PAGE (***Figure 3—figure supplement 1C***), MNase (***Figure 3—figure supplement 1D***) and
854 chromatosome stop assays (***Figure 3—figure supplement 1E***).
855
856 *ATPase assay*
857 40 nM recombinant proteins were incubated in 25 µl reaction buffer containing 20 mM HEPES, pH
858 7.6, 0.15 M NaCl, 4 mM $MgCl_2$, 1 mM ATP, 0.1 mM EDTA, 0.02% (v/v) NP-40 and 0.1 mg/ml
859 nuclease-free BSA for 60 min at 27°C. Some reactions additionally contained 10 nM pGIE-0 plasmid
860 DNA or equivalent amounts of oligonucleosomes ±H1. ATPase assays were performed using ADP-Glo
861 Max kit (Promega). All reactions were performed in triplicate, the results were normalized to the ADP-
862 ATP titration curve according to the kit manual and converted to enzymatic rates (molecules of ATP

863 hydrolyzed per molecule of enzyme per minute). Averages and standard deviations were calculated.
864 Statistical differences were calculated by Mann-Whitney test.
865
866 *EpiDyne®-PicoGreen™ nucleosome remodeling assay*
867 EpiDyne®-PicoGreen™ is a restriction enzyme accessibility assay modified for increased throughput
868 and sensitivity (***Figure 3—figure supplement 2A***). Briefly, a recombinant ATPase over a concentration
869 range (***Figure 3—figure supplement 2B-E***) was mixed with 10 nM EpiDyne biotinylated nucleosome
870 remodeling substrate (EpiCypher), terminally positioned 6-N-66 (219 bp fragment) or centrally
871 positioned 50-N-66 (263 bp) and 1 mM ATP in 20 µL remodeling buffer, 20 mM Tris-HCl, pH 7.5, 50
872 mM KCl, 3 mM $MgCl_2$, 0.01% (v/v) Tween-20, 0.01% (w/v) BSA. The remodeling reactions were
873 incubated at 23˚C in 384-well format. At indicated time points, the reactions were quenched, and
874 nucleosome substrates were immobilized on an equal volume of streptavidin-coated magnetic beads
875 (NEB), pre-washed and resuspended in 2x quench buffer, 20 mM Tris-HCl, pH 7.5, 600 mM KCl,
876 0.01% (v/v) Tween-20 and 0.01% (w/v) BSA. Beads were successively washed by collection on a
877 magnet (three times with wash buffer, 20 mM Tris-HCl, pH 7.5, 300 mM KCl, 0.01% (v/v) Tween-20)
878 and buffer replacement (once with RE buffer, 20 mM Tris-HCl, pH 7.5, 50 mM KCl, 3 mM $MgCl_2$,
879 0.01% (v/v) Tween-20). Beads were resuspended in 20 µl restriction enzyme mix, 50 units/ml Dpn II
880 (NEB) in RE buffer, and incubated at 23˚C for 30 min, collected on a magnet, and supernatants from all
881 wells were transferred to a new plate. They were mixed with an equal volume of Quant-iT™
882 PicoGreen™ dsDNA reagent (ThermoFisher, Component A) and 1 unit/ml thermolabile proteinase K
883 (NEB) in TE and incubated at 23˚C for 1 hr. Fluorescence intensity was detected on an Envision
884 microplate reader with excitation at 480 nm and emission at 531 nm, and data expressed as relative
885 fluorescence units (RFU) through the EnVision Workstation (version 1.13.3009.1409).
886
887 ***Drosophila* population culture, mutant stocks and genetics**

Wild-type (Oregon R) flies were maintained in population cages on agar-grape juice and yeast paste plates at 26°C, 60% humidity with 12-h dark-light cycle. Mutant flies were reared, and crosses were performed at 26°C on standard cornmeal/molasses medium with dry yeast added to the surface. *SuUR*^{ES} was a gift of Igor Zhimulev, and *mod(mdg4)*^{m9} was a gift of Yuri Schwartz. All other alleles were obtained from the Bloomington Stock Center, Indiana. Combinations of alleles were produced either by crosses with appropriate balancers and segregation of markers or by female germline meiotic recombination. Intra-chromosomal recombination events were confirmed by PCR of genomic DNA. To genotype *SuUR*^{ES}, *mod(mdg4)*^{u1} recombined chromosomes, the following PCR primers were used: SUUR-Fwd: `CCTCAAAGAACAGCCAGAGC`; SUUR-Rev: `TTTGCTACTTCTGGGCGTTT`; diver-Rev: `TCAGTTTGAACTCGCACCAG`; Mod-Fwd: `CAGGGCCACACGCACTTAC`; Mod-Rev: `GTGAAGCCCTTAGGCAGCTC`; and Stalker-Rev: `GCTTGCAGCACAGTTAGCAC`. SUUR-Fwd/SUUR-Rev combination of primers produced a 770-bp PCR product for wild-type *SuUR*. SUUR-Fwd/diver-Rev combination produced an ~850-bp PCR product for *SuUR*^{ES}. Mod-Fwd/Mod-Rev combination produced a 1,532-bp PCR product for wild-type *mod(mdg4)*. Mod-Fwd/Stalker-Rev combination produced an ~1,700-bp PCR product for *mod(mdg4)*^{u1}.

Fly wings were dissected from ~5 days old adult males and transferred to a drop of PBS + 0.1% Triton X-100 (PBST). The wings were soaked in 80% glycerol in PBST and photographed using Zeiss AxioVert 200M microscope with EC Plan-Neofluar 2.5X/0.075 lens in bright field and CCD monochrome camera AxioCam MRm. For wing area measurements, images were processed using Fiji/ImageJ2 software package. Statistical differences were calculated by two-tailed t-test, assuming unequal variances. Adult fly eye images were taken on live, CO_2-anesthetized 2-day-old females on Zeiss stereomicroscope Discovery.V12 using CCD color camera AxioCam MRc.

Antibodies, immunoblots and immunoprecipitation (IP)

913 Polyclonal antibody (anti-ModT) was raised in Guinea pigs by Pocono Rabbit Farm & Lab. Rabbit
914 polyclonal antibody to the C-terminus of *Drosophila* XNP/ATRX (anti-XNP) was described previously
915 (Emelyanov, Konev, Vershilova, & Fyodorov, 2010). Rabbit and Guinea pig polyclonal antibodies to
916 *Drosophila* SUUR were a gift of Alexey Pindyurin (Nordman et al., 2014) and Igor Zhimulev
917 (Pindyurin et al., 2008). Rabbit polyclonal Mod(Mdg4)-FL antibody to full-length Mod(Mdg4)-67.2 that
918 recognizes all splice forms of Mod(Mdg4) was a gift of Jordan Rowley and Victor Corces. Mouse
919 monoclonal anti-FLAG (M2, Sigma Aldrich), anti-PCNA (PC10, Cell Signaling), anti-β-tubulin and
920 anti-HP1a (E7 and C1A9, Developmental Studies Hybridoma Bank) were obtained commercially.
921
922 Western blotting was performed using standard techniques. For FPLC column fraction analyses, 5-
923 10 µl of starting material and flow-through (if applicable) and 5-15 µl of column fractions were loaded
924 per lane. For expression analyses in salivary glands, 10 salivary glands from L3 larvae of indicated
925 genotype were frozen and thawed, boiled extensively in 40 µl 2x SDS-PAGE loading buffer,
926 centrifuged, and the material equivalent to four salivary glands was loaded per lane. The following
927 dilutions were used: 1:200,000 anti-ModT, 1:1,000 anti-Mod(Mdg4)-FL, 1:1,000 Guinea pig and rabbit
928 anti-SUUR, 1:1,000 anti-HP1a, 1:1,000 anti-β-tubulin and 1:2,000 anti-FLAG. Infrared-labeled
929 secondary antibodies: donkey anti-Guinea pig IRDye 800CW, goat anti-mouse IRDye 800CW, goat
930 anti-rabbit IRDye 800CW, goat anti-rabbit IRDye 680CW and goat anti-mouse IRDye 680RD – were
931 obtained from Li-COR Biosciences and used at 1:10,000. The blots were scanned on Odyssey Fc
932 Imaging System (LI-COR Biosciences).
933
934 Immunoprecipitation experiments were performed as described (Emelyanov et al., 2012). 400 µl
935 *Drosophila* embryonic nuclear extracts (~10 mg total protein) were incubated with 10 µl Guinea pig
936 anti-ModT, 30 µl rabbit anti-SUUR or 20 µl rabbit anti-XNP antibodies for 3 h at 4°C.
937 Immunocomplexes were collected by addition of 25 µl protein A-agarose plus (Thermo Fisher) for 2 h

938 at 4°C. After washing four times with 1 ml of buffer HEG (25 mM HEPES, pH 7.6, 0.1 mM EDTA,
939 10% glycerol) + 0.15 M NaCl, the immunoprecipitated proteins were eluted with 80 µl 2x SDS-PAGE
940 loading buffer and analyzed by SDS-PAGE and western blot using Guinea pig or rabbit anti-SUUR and
941 anti-Mod(Mdg4) and mouse anti-HP1a antibodies. For Mod(Mdg4) and HP1a, 8 µl of
942 immunoprecipitated material (equivalent to 1 mg nuclear extract proteins) and 5% input (2 µl nuclear
943 extract, 50 µg total protein) were analyzed. For SUUR, 20 µl of immunoprecipitated material
944 (equivalent to 2.5 mg nuclear extract proteins) and 10% input (10 µl nuclear extract, 250 µg total
945 protein) were analyzed.
946
947 **Polytene chromosomes and indirect immunofluorescence (IF) analyses**
948 For all cytological experiments, larvae were reared and collected at 18°C. Polytene chromosomes
949 and whole-mount salivary glands were prepared and analyzed as described previously (Andreyeva et al.,
950 2017). Briefly, salivary glands from wandering third instar larvae were dissected in PBS. Glands were
951 transferred into a formaldehyde-based fixative (one ~15-µl drop of 3% lactic acid, 45% acetic acid,
952 3.7% formaldehyde on a coverslip) for 2 min, squashed, and frozen in liquid N_2. The coverslips were
953 removed, and slides were placed in 70% ethanol for 20 min and stored at −20°C. The slides were
954 washed three times for 5 min in PBST. Primary antibodies were incubated overnight at 4°C in PBST +
955 0.1% BSA and washed three times for 5 min each with PBST. Secondary antibodies were incubated for
956 2 h at room temperature in PBST + 0.1% BSA and washed three times for 5 min each with PBST.
957
958 DNA was stained with 0.1 µg/ml DAPI in PBST for 3 min, and squashes were mounted in Prolong
959 Glass anti-fade mountant (Molecular Probes). Primary and secondary antibodies were used at the
960 following dilutions: Guinea pig anti-ModT, 1:50,000; rabbit anti-SUUR, 1:100; mouse anti-PCNA,
961 1:1,000; mouse anti-FLAG, 1:100; Alexa Fluor 488 highly cross-absorbed (HCA) goat anti-mouse,
962 Alexa Fluor 568 HCA goat anti-Guinea pig and Alexa Fluor 647 plus HCA goat anti-rabbit (all Thermo

963 Fisher), all 1:800. Indirect immunofluorescence (IF) images were obtained with Zeiss AxioVERT 200M
964 microscope and AxioCam MRm mono microscopy camera using a 40x/1.3 Plan-Neofluar or 63x/1.40
965 Plan-Apochromat lenses with oil immersion. Images were acquired using AxioVision software.
966
967 For whole-mount IF staining, L3 larvae were reared at 26°C, and salivary glands were dissected in
968 PBS and fixed in 3.7% formaldehyde (Sigma Aldrich) for 20 min at room temperature. The glands were
969 washed in PBS + 0.3% Triton X-100 and permeabilized for 30 min at 37°C in PBS + 1% Triton X-100.
970 Blocking was performed for 30 min at room temperature in PBS+ 0.3% Triton X-100 supplemented
971 with 10% fetal calf serum and 1% BSA. The glands were incubated with primary antibodies diluted in
972 blocking solution for 48 h at 4°C, washed three times with PBS + 0.3% Triton X-100 for 30 min, and
973 incubated with secondary antibodies in blocking solution overnight at 4°C. The stained glands were
974 washed three times with PBS + 0.3% Triton X-100 for 30 min, stained with DAPI (0.1 µg/ml) for 30
975 min, and mounted in Prolong Gold anti-fade (Invitrogen). IF images were obtained on a Leica SP8
976 confocal microscope using a 20X/0.75 PLAPO lens and processed using Fiji/ImageJ software.
977
978 To quantify the putative colocalization of SUUR and Mod(Mdg4)-67.2 in polytene chromosomes
979 (***Figure 4A***), the image resolution was reduced to 1,388 by 1,040. Pixel intensities (1,443,520) for
980 SUUR and ModT channels were extracted from Bitmap files (ImageJ), normalized to Z-scores and
981 plotted as an X-Y scatter plot (***Figure 4—figure supplement 1C***). For colocalization analyses, the plot
982 regions (Z_{ModT}>1 AND Z_{SUUR}<3, green) and (Z_{ModT}<1 AND Z_{SUUR}>3, red) were excluded from
983 consideration (***Figure 4—figure supplement 1D***).
984
985 **Next generation sequencing analyses (NGS)**
986 Salivary glands from female wandering third-instar larvae were isolated and flash-frozen in liquid N_2
987 until all samples were collected. Genomic DNA for sequencing was prepared from 25 L3 salivary gland

988 pairs or 10 mg embryos (0-6 h AED) using DNeasy Blood and Tissue kit (Qiagen). Each sample was
989 prepared in triplicate. The tissues were soaked in 180 µl buffer ATL + 20 µl proteinase K (15 mg/ml)
990 and lysed for 2-3 h at 55°C. The reactions were cooled to room temperature, supplemented with 4 µl
991 RNase A, ~40 mg/ml (Sigma Aldrich), and RNA was digested for 10 min. The genomic DNA was
992 fragmented with 0.002 units DNase I (Thermo Fisher) in 100-µl reactions containing 10 mM Tris-HCl,
993 pH 7.5, 10 mM $MnCl_2$, 0.1 mM $CaCl_2$, 0.1 mg/ml RNase A and 0.2 mg/ml nuclease-free BSA (1x
994 reaction buffer) for 15 min at 37°C. (DNAse I dilutions were prepared using 1x reaction buffer.)
995 Reactions were stopped by adding 5 µl 0.5 M EDTA, and DNase I was inactivated for 20 min at 65°C.
996 The fragmented DNA was purified on QiaQuick columns using PCR purification kit (Qiagen) and eluted
997 in 40 µl 10 mM Tris-HCl, pH 8.0. The size distribution of DNA fragments (200-600 bp, average ~400
998 bp) was confirmed and DNA concentration was measured on 2100 BioAnalyzer (Agilent). Libraries
999 were prepared from 20 ng of fragmented genomic DNA with the ThruPLEX DNA-seq kit using
1000 SMARTer® DNA Unique Dual Indexes (TakaraBio) and sequenced 150-bp paired-end reads on an
1001 NovaSeq 6000 (Novagene).
1002
1003 The sequencing quality of each sample was assessed using FASTQC version 0.11.7 (Andrews,
1004 2010). Raw paired-end reads were trimmed of adapters using BBDuk from the BBTools software
1005 version 38.71 using the parameters: `ktrim=r ref=adapters rcomp=t tpe=t tbo=t`
1006 `hdist=1 mink=11` (Bushnell, 2014). Reads were aligned to the BDGP Release 6 of the *Drosophila*
1007 *melanogaster* genome (dm6) (dos Santos et al., 2015) using Bowtie2 version 2.3.4.1 (Langmead &
1008 Salzberg, 2012) and parameters `-q --local --very-sensitive-local --no-unal --`
1009 `no-mixed --no-discordant --phred33 -I 10 -X 700`. Duplicate reads were marked
1010 using Picard 2.2.4 (BroadInstitute) and SAM files were converted to BAM format, filtered for quality (`-`
1011 `bq 5`), and removed of duplicates (`-bF 0x400`) using Samtools version 1.9 (Danecek et al., 2021). To
1012 examine replicate concordance, a principal component analysis (PCA) was performed using the

deepTools package. Replicates clustered indicating high genome-wide similarity within genotypes (*not shown*). For visualization, replicates were merged (`samtools merge`) and coverage was calculated across 50-bp bins and normalized to counts per million (CPM) using deeptools version 3.2.0: `bamCoverage -bs 50 –normalizeUsing CPM` (Ramirez et al., 2016). Each genotype was scaled to the diploid Oregon R embryo signal in 5-kb bins: `bigWigCompare --operation first -bs 5000`. DamID-chip data for SUUR and Su(Hw) were retrieved from GSE22069 (Filion et al., 2010). ChIP-chip data for Su(Hw) insulator elements were also used (Negre et al., 2010). underreplicated domains were called using a custom R script to identify regions at least 100 kb in length that fell below the average chromosomal read count as described (Andreyeva et al., 2017). Visualization of all data was performed on the UCSC Genome browser using the dm6 release of the *Drosophila* genome (Kent et al., 2002). Each data set was auto-scaled to its own min and maximum and the data were windowed by mean with 16-pixel smoothing applied.

Quantitative real-time PCR

Genomic DNA samples prior to DNase I fragmentation (see above) were diluted to ~0.25 ng/µl. Real-time PCR was performed using 0.5 ng genomic DNA on a ViiA7 thermocycler (Applied Biosystems) with a three-step protocol (95°C 15 sec, 60°C 30 sec, 68°C 60 sec) and iTaq Universal SYBR Green Supermix (Bio-Rad). Primer sequences are provided in ***Figure 7—source data 1***. Each reaction was performed in three technical replicates for each of the three biological samples (N=9). For each amplicon, the average Ct value (<Ct>) was calculated and normalized to the average Ct value for a random intergenic genomic sequence as a loading control. Further, for each template, the ΔCt was normalized to the average Ct value for embryonic DNA (diploid control). Standard deviation (σ_{Ct}) for each reaction in triplicate was also calculated. The following ΔΔCt formula was used: $\langle\bullet\bullet Ct\rangle = (\langle Ct_{target}\rangle - \langle Ct_{intergenic86D}\rangle)_{SG} - (\langle Ct_{target}\rangle - \langle Ct_{intergenic86D}\rangle)_{embryo}$. Standard deviations for

1037 <ΔΔCt> were calculated as $\sigma_{\bullet\bullet Ct}$ = square root of (σ^2_{target} + $\sigma^2_{intergenic86D}$)/2. ΔΔCt's were

1038 converted to DNA copy numbers as $2^{-<\Delta\Delta Ct>}$. The confidence interval was calculated in the range

1039 between $2^{-<\Delta\Delta Ct>-\sigma}$ and $2^{-<\Delta\Delta Ct>+\sigma}$.

1040

1041 To examine the putative zygotic function(s) of *SuUR*, heterozygous $SuUR^{ES}$ parents were produced

1042 by balancing with *TM6B, Tb* and crossed *inter se*. L3 salivary glands were dissected from homozygous

1043 *SuUR* mutant progeny, and DNA copy numbers were measured by qPCR as described above.

Acknowledgments

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.

Additional Information

Funding

NIH Office of the Director	NIGMS R01 GM074233	Dmitry V Fyodorov
NIH Office of the Director	NIGMS R01 GM129244	Arthur I Skoultchi
NIH Office of the Director	NIGMS R01 GM124201	Robert J Duronio
NIH Office of the Director	NIGMS R44 GM123869	Michael-C Keogh (EpiCypher)
NIH Office of the Director	NCI T32 CA217824 Training Grant	Markus Nevil
NIH Office of the Director	NIGMS K12 GM000678 Training Grant	Markus Nevil
NIH Office of the Director	NIGMS Equipment Supplement	Dmitry V Fyodorov, Arthur I Skoultchi, and Charles C Query

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Additional Files

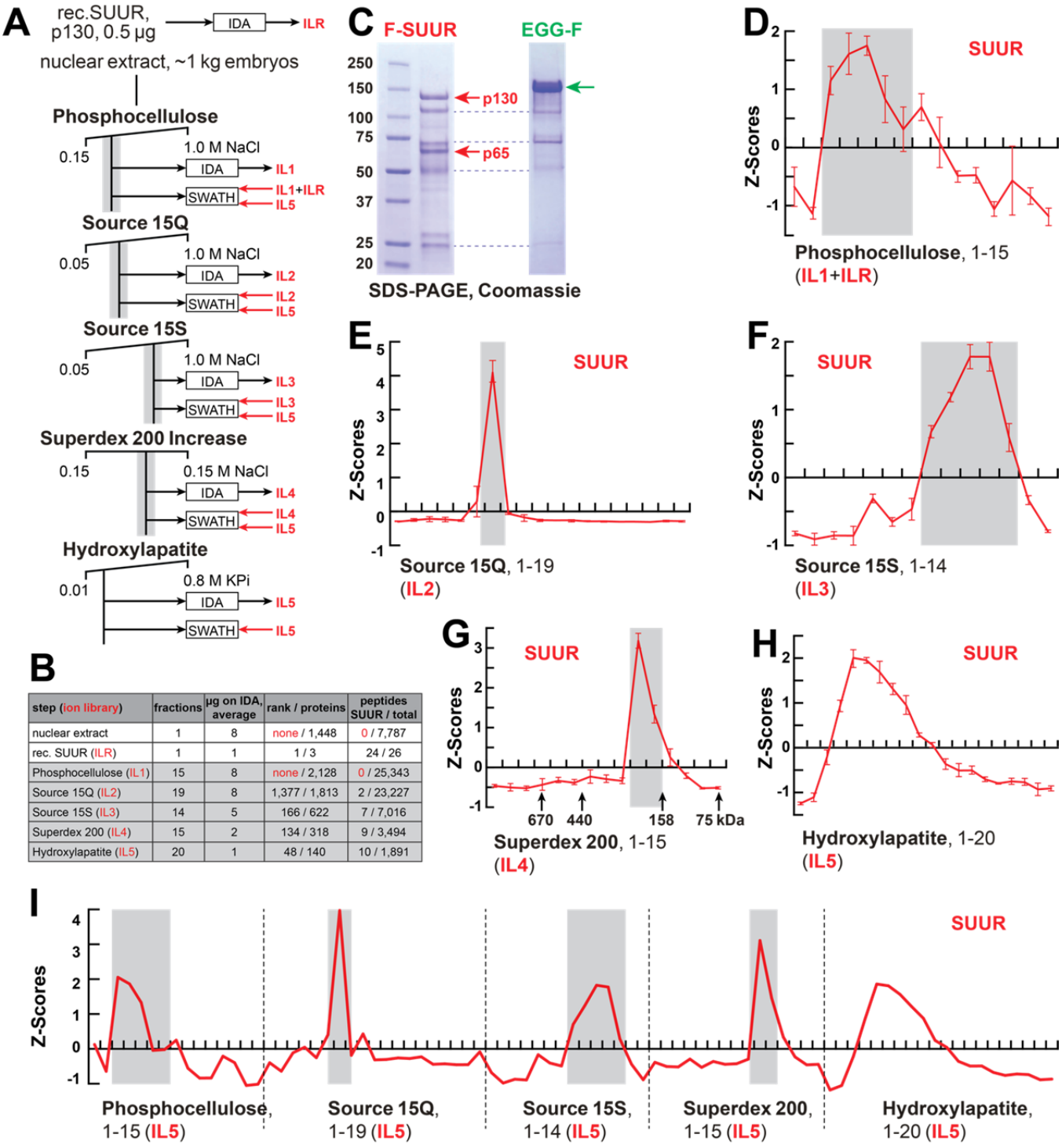
Supplementary Files

Supplementary File 1. Protein identities and peptide spectral data (ion libraries) obtained by IDA acquisitions for FPLC fractions (IL1-5, ***Figure 1A***) and recombinant SUUR (ILR, ***Figure 1C***).

Supplementary File 2. Raw data of SWATH acquisitions for FPLC fractions (***Figure 1A***) quantified using ion library IL5 (***Supplementary File 1***) and an example of protein purification profile analyses (hydroxylapatite step, ***Figure 1A&H***).

Data Availability

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



step (ion library)	fractions	µg on IDA, average	rank / proteins	peptides SUUR / total
nuclear extract	1	8	none / 1,448	0 / 7,787
rec. SUUR (ILR)	1	1	1 / 3	24 / 26
Phosphocellulose (IL1)	15	8	none / 2,128	0 / 25,343
Source 15Q (IL2)	19	8	1,377 / 1,813	2 / 23,227
Source 15S (IL3)	14	5	166 / 622	7 / 7,016
Superdex 200 (IL4)	15	2	134 / 318	9 / 3,494
Hydroxylapatite (IL5)	20	1	48 / 140	10 / 1,891

Figure 1. FPLC fractionation and MERCI quantification of native SUUR. (***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***) Representation of SUUR in ion

1088 libraries ILR and IL1-5 (***Supplementary File 1***). Total number of identified proteins and the confidence
1089 rank of SUUR among them as well as the total number of detected peptides (95% confidence) and the
1090 number of SUUR-specific peptides are shown. (***C***) Recombinant FLAG-SUUR expressed in Sf9 cells.
1091 Identities of eight most prominent bands were determined by mass-spectroscopy. p130 and p65
1092 correspond to full-length and C-terminally truncated FLAG-SUUR, respectively (red arrows). Other
1093 bands represent common Sf9-specific contaminants purified by FLAG chromatography (blue dashed
1094 lines), *cf* purified EGG-F (green arrow). Molecular mass marker bands are indicated (kDa). (***D-H***)
1095 SWATH quantitation profiles of SUUR fractionation across individual FPLC steps. Ion libraries (IL)
1096 used for SWATH quantitation are shown at the bottom of each panel. Z-scores across indicated column
1097 fractions are plotted; error bars, standard deviations (*N*=3). Gray rectangles, fraction ranges used for the
1098 next FPLC step; in (***G***), black arrows, expected peaks of globular proteins with indicated molecular
1099 masses in kDa. (***I***) SWATH quantitation profiles of SUUR fractionation across five FPLC steps. IL5 ion
1100 library was used for SWATH quantification.
1101
1102 ***Figure 1—source data 1.*** FPLC column parameters (***Figure 1A***). The following FPLC column
1103 parameters were used for partial purification of native SUMM4. HEG: 25 mM HEPES, pH 7.6, 0.1 mM
1104 EDTA, 10% glycerol, 0.02% NP-40, 1 mM DTT, 1 mM benzamidine, 0.4 mM PMSF; 10 mM KPi: 10
1105 mM potassium phosphate, pH 7.6, 10% glycerol, 1 mM DTT, 1 mM benzamidine, 0.4 mM PMSF; 0.8
1106 M KPi: 800 mM potassium phosphate, pH 7.6, 10% glycerol, 1 mM DTT, 1 mM benzamidine, 0.4 mM
1107 PMSF; *cv*, column volume.
1108
1109 ***Figure 1—source data 2.*** Recombinant proteins expressed in Sf9 cells and purified by FLAG affinity
1110 chromatography. Lane 1, protein size marker; lane 2, FLAG-SUUR, 72-hour infection of Sf9 cells; lane
1111 3, FLAG-SUUR, 60-hour infection of Sf9 cells; lane 4, XNP-FLAG (Emelyanov, Konev, Vershilova, &
1112 Fyodorov, 2010), 72-hour infection of Sf9 cells; lane 5, XNP-FLAG, 60-hour infection of Sf9 cells; lane

6, EGG-FLAG, 72-hour infection of Sf9 cells; lane 7, EGG-FLAG, 60-hour infection of Sf9 cells. Prep amounts equivalent to ~20 ml Sf9 culture were loaded in each lane. Cropped images encompassing lanes 1-2 and 6 (open boxes, dashed red line) were used for ***Figure 1C***.

Figure 1—figure supplement 1. Quantification of SUUR in chromatographic fractions. (***A***) Schematic of SWATH quantification of recombinant SUUR, nuclear extract (starting material) and FPLC fractions for SUUR using ion library ILR. (***B***) SUUR titration curve obtained by SWATH quantitation of 10 fg – 1 µg recombinant FLAG-SUUR in the presence of 25 µg *E. coli* lysate; both axes are logarithmic ($\log_{10}$). Red rectangle, SUUR quantification in 25 µg nuclear extract; error bars, standard deviations (*N*=3). (***C***) SWATH quantitation profiles of SUUR fractionation across individual FPLC steps. Ion library ILR was used for SWATH quantification, and relative amounts were converted to estimated ng SUUR per fraction. Error bars, standard deviations (*N*=3); colored boxes, peak fractions of SUUR. (***D***) SUUR purification by FPLC. Total protein was measured by BCA assay, and SUUR was measured as in (***C***). Relative purity, purification factor in each step and cumulative purification factor are shown.

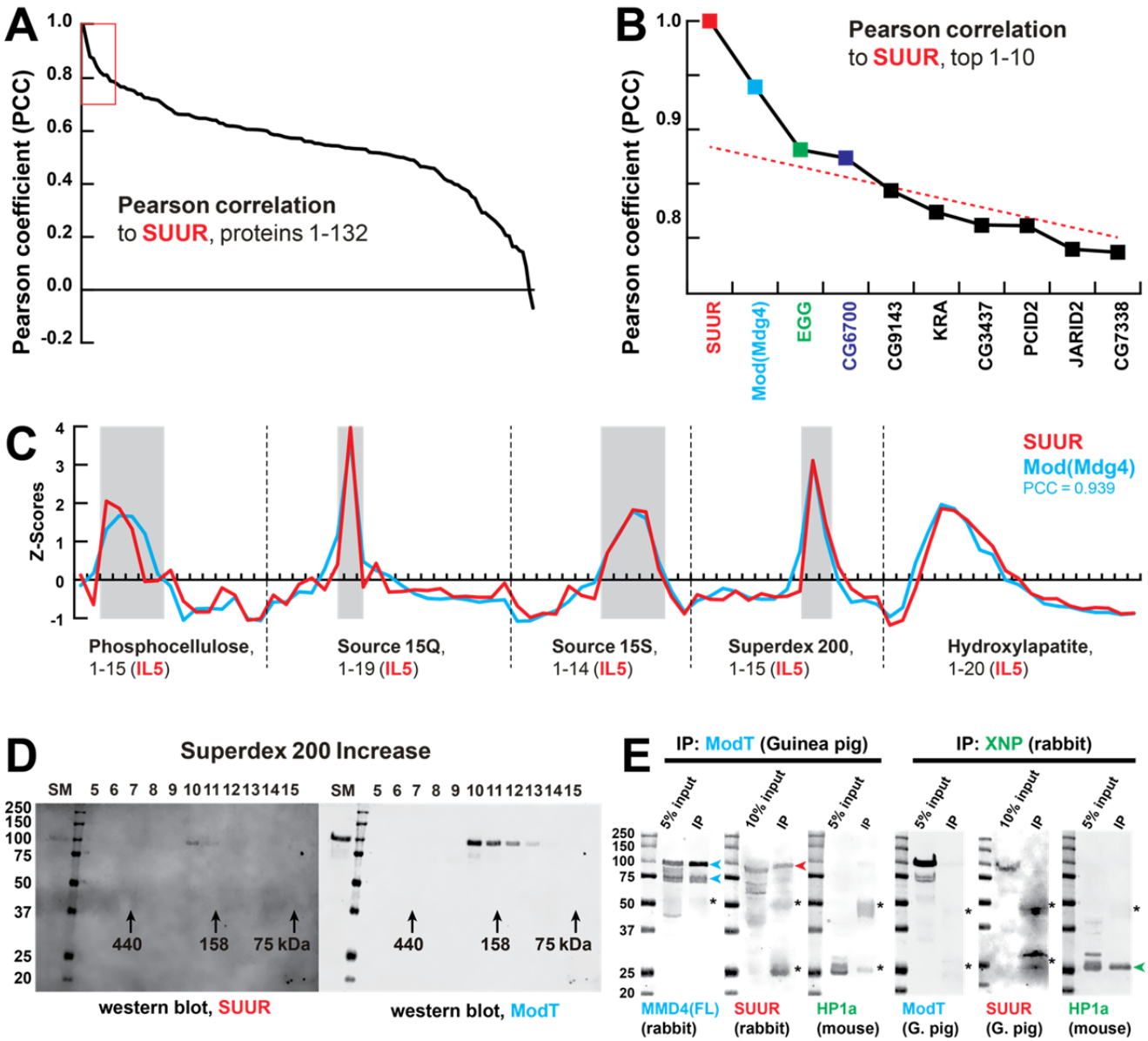


Figure 2. Identification of the SUMM4 complex by MERCI. (***A***) Pearson correlation of fractionation profiles for individual 132 proteins to that of SUUR, sorted from largest to smallest. Red box, the graph portion shown in (***B***). (***B***) Top ten candidate proteins with the highest Pearson correlation to SUUR. Red dashed line, trend line extrapolated by polynomial regression (n = 5) from the bottom 130 proteins. (***C***) SWATH quantitation profiles of SUUR (red) and Mod(Mdg4) (cyan) fractionation across five FPLC steps, *cf* ***Figure 1I***. IL5 ion library was used for SWATH quantification. (***D***) Western blot analyses of Superdex 200 fractions with SUUR and ModT antibodies, *cf* ***Figure 1G***. Molecular mass markers are shown on the left (kDa). (***E***) Co-IP experiments. SUUR (red arrowhead) co-purifies from nuclear

1136 extracts with Mod(Mdg4)-67.2 (cyan arrowheads) but not HP1a (green arrowhead). Anti-XNP co-IPs
1137 HP1a but not SUUR of Mod(Mdg4)-67.2. Asterisks, IgG heavy and light chains detected due to
1138 antibody cross-reactivity. Mod(Mdg4)-67.2(FL) antibody recognizes all splice forms of Mod(Mdg4).
1139

1140 ***Figure 2—source data 1.*** Western blots of chromatographic fractions. Left panels, 700 nm channel
1141 (Odyssey Fc), rabbit anti-SUUR antibody and protein size marker; right panels, 800 nm channel
1142 (Odyssey Fc), Guinea pig ModT antibody; top panels, hydroxylapatite fractions: starting material, flow-
1143 through, marker, fractions 1-12 (***Figure 1H***); bottom panels, Superdex 200 Increase fractions: starting
1144 material, marker, fractions 5-15 (***Figure 1G***). Cropped images from bottom panels (open boxes, dashed
1145 red line) were used for ***Figure 2D***.
1146

1147 ***Figure 2—source data 1.*** Co-IP of SUMM4 subunits. Panels A and E, westerns, 700 nm channel
1148 (Odyssey Fc), mouse anti-HP1a and protein size marker; panel B, western, 800 nm channel (Odyssey
1149 Fc), rabbit anti-Mod(Mdg4)-FL; panels C and G, westerns, 700 nm channel (Odyssey Fc), protein size
1150 marker only; panel D, western, 800 nm channel (Odyssey Fc), rabbit anti-SUUR; panel F, western, 800
1151 nm channel (Odyssey Fc), Guinea pig ModT; panel H, western, 800 nm channel (Odyssey Fc), Guinea
1152 pig anti-SUUR. Lanes 1, 5, 9, 12, 15 and 18, protein size marker; lanes 2, 6, 10, 13, 16 and 19, input
1153 (nuclear extract), 5 or 10%; lanes 3 and 7, IP with Guinea pig ModT antibody #1; lanes 4 and 8, IP with
1154 Guinea pig ModT antibody #2; lanes 11 and 17, IP with rabbit preimmune serum; lanes 14 and 20, IP
1155 with rabbit anti-XNP. Cropped images encompassing lanes 1-3, 5-7, 12-14 and 18-20 (open boxes,
1156 dashed red line) were used for ***Figure 2E***.
1157

1158 ***Figure 2—figure supplement 1.*** Comparisons of SWATH quantification profiles for protein
1159 fractionation. (***A-C***) SWATH quantitation of SUUR (red), EGG (***A***, green), CG6700 (***B***, blue) and HP1a

(***C***, black) fractionation profiles across five FPLC steps as in ***Figure 1I&2C***. Pearson coefficients (PCC) are shown (***Figure 2A&B***).

Figure 2—figure supplement 2. Identification of Mod(Mdg4)-67.2 as a subunit of the SUMM4 complex. (***A***) Mod(Mdg4)-specific peptides from ion library IL5 (***Supplementary File 1***). Gray shading, peptides specific to the common part (coding exons 2-4) of Mod(Mdg4); cyan shading, peptides specific to polypeptide Mod(Mdg4)-67.2 encoded by *pre-mod(mdg4)-T*, exons 2-3. Peptide sequences, confidence levels, charges (*z*), theoretical and observed *m/z*, column retention times (RT) and total MS2 ion counts are shown. (***B***) Mod(Mdg4)-67.2 polypeptide sequence. The common part is shaded in gray, splice form-specific part is shaded in cyan. Peptides from ion library IL5 (as in E) are highlighted in bold red. (***C***) Mod(Mdg4)-59.1 polypeptide sequence. The common part is shaded in gray, splice form-specific part is shaded in light green.

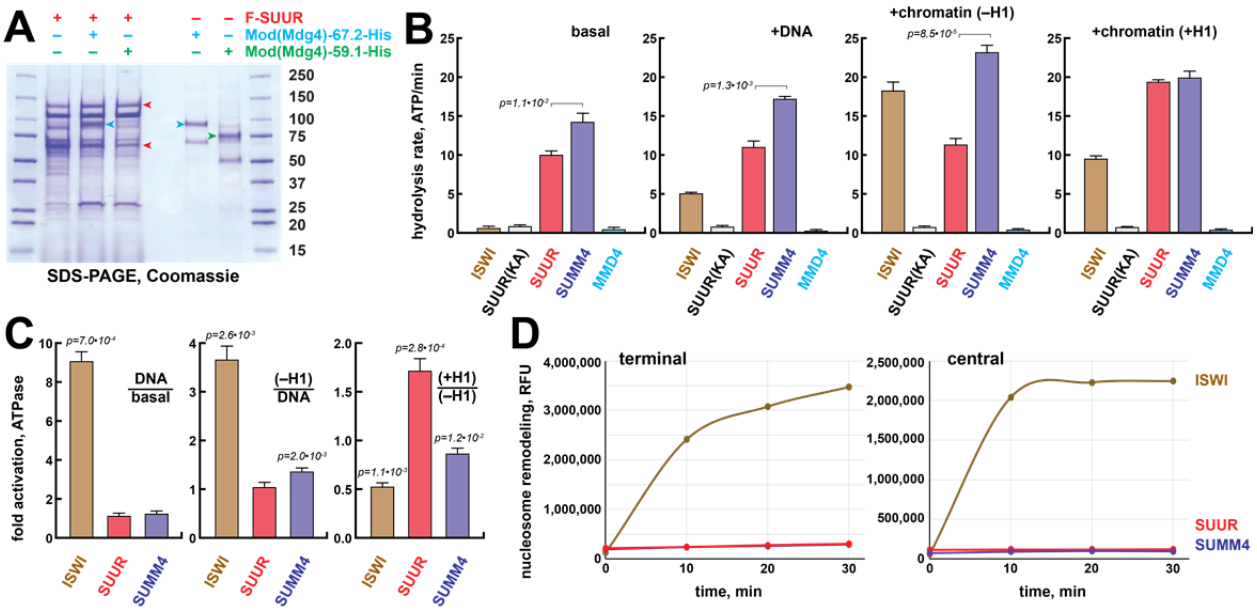


Figure 3. Biochemical activities of recombinant SUMM4. (***A***) Recombinant SUMM4. Mod(Mdg4)-His$_6$, 67.2 (p100, cyan arrowhead) and 59.1 (p75, green arrowhead) splice forms were co-expressed with FLAG-SUUR (red arrowheads, p130 and p65) or separately in Sf9 cells and purified by FLAG or Ni-NTA affinity chromatography. Mod(Mdg4)-67.2 forms a specific complex with SUUR. Identities of the 130-, 100-, 75- and 65-kDa protein bands from FLAG- and Ni-NTA-purified material were determined by mass-spectroscopy. (***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 amounts of 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 activity. 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

1189 centrally (50-N-66) positioned mononucleosomes (***Figure 3—figure supplement 2B-E***). RFU, relative
1190 fluorescence units produced by PicoGreen fluorescence.
1191
1192 ***Figure 3—source data 1.*** Recombinant proteins expressed in Sf9 cells and purified by FLAG or Ni-
1193 NTA affinity chromatography. Lanes 1 and 7, protein size marker; lane 2, FLAG-SUUR, FLAG-
1194 purified; lane 3, FLAG-SUUR + Mod(Mdg4)-67.2-His6, FLAG-purified; lane 4, FLAG-SUUR +
1195 Mod(Mdg4)-59.1-His6, FLAG-purified; lane 5, Mod(Mdg4)-67.2-His6, Ni-NTA-purified; lane 6,
1196 Mod(Mdg4)-59.1-His6, Ni-NTA-purified. All proteins were purified 72 hours post-infection. Prep
1197 amounts equivalent to ~20 ml (FLAG-purified, lanes 2-4) or ~1 ml (Ni-NTA-purified, lanes 5 and 6) Sf9
1198 cultures were loaded in each lane. Cropped image encompassing all lanes (open box, dashed red line)
1199 was used for ***Figure 3A***.
1200
1201 ***Figure 3—figure supplement 1.*** Recombinant proteins and biochemical substrates. (***A***) Physical
1202 interactions of recombinant EGG, SUUR and WDE. Untagged EGG (green arrowhead) was co-
1203 expressed with FLAG-SUUR (red arrowheads, p130 and p65) or WDE-FLAG (purple arrowhead) in Sf9
1204 cells and purified by FLAG affinity chromatography. EGG forms a specific complex with WDE but not
1205 SUUR. Molecular mass markers (kDa) are shown on the left. (***B***) Recombinant FLAG-SUUR(K59A)
1206 expressed in Sf9 cells and ISWI expressed in *E. coli*. See legend to ***Figure 1C***. (***C***) Protein composition
1207 of *in vitro* reconstituted chromatin. Oligonucleosomes prepared from plasmid DNA and core histones
1208 with (+H1) or without H1 (–H1) were analyzed by SDS-PAGE and Coomassie staining. Positions of
1209 BSA, H1 and core histone bands are indicated on the right. (***D***) Micrococcal nuclease (MNase) analysis
1210 of reconstituted chromatin. Partial digestion with five different dilutions of MNase was performed on
1211 H1-free (–H1) and H1-containing (+H1) oligonucleosomes. Deproteinated DNA fragments were
1212 analyzed by agarose gel electrophoresis and stained with ethidium. Note the increased nucleosome

1213 repeat length in (+H1) lanes consistent with H1 incorporation. Triangles at the top indicate increasing
1214 MNase concentrations; 123 bp ladder was used as a molecular mass marker. (***E***) Chromatosome stop
1215 assay. Oligonucleosomes assembled with or without H1 were subjected to partial MNase digestion, and
1216 DNA was analyzed by agarose gel electrophoresis and ethidium bromide staining. Positions of the core
1217 particle and chromatosome DNA are indicated by arrowheads. DNA fragment sizes in the 20-bp DNA
1218 ladder marker are shown.
1219
1220 ***Figure 3—figure supplement 2.*** EpiDyne®-PicoGreen™ biochemical assay. (***A***) EpiCypher®
1221 EpiDyne®-PicoGreen™ assay design. EpiDyne nucleosomes encompass a restriction site shielded by the
1222 initial nucleosome position but exposed for Dpn II cleavage upon remodeling (sliding or displacement).
1223 Biotinylated substrates are immobilized on streptavidin magnetic beads. Digest by Dpn II releases the
1224 substrates from beads, and supernatant is quantified by PicoGreen™ (dsDNA detection reagent)
1225 fluorescence. (***B***) Titration of *Drosophila* ISWI remodeling activity using terminally (6-N-66) or
1226 centrally (50-N-66) positioned mononucleosomes. Early reaction time points were separately plotted to
1227 indicate linear ranges. RFU, relative fluorescence units. (***C***) Early remodeling rates for ISWI were
1228 calculated by linear regression analyses of data in respective linear ranges. ISWI exhibits a stronger
1229 remodeling activity with a centrally positioned nucleosome substrate. (***D***) Titration of human BRG1
1230 remodeling activity. Data are presented as in (***B***). (***E***) Early remodeling rates for BRG1 were calculated
1231 and plotted as in (***D***). BRG1 does not exhibit a bias towards remodeling centrally or terminally
1232 positioned nucleosomes.

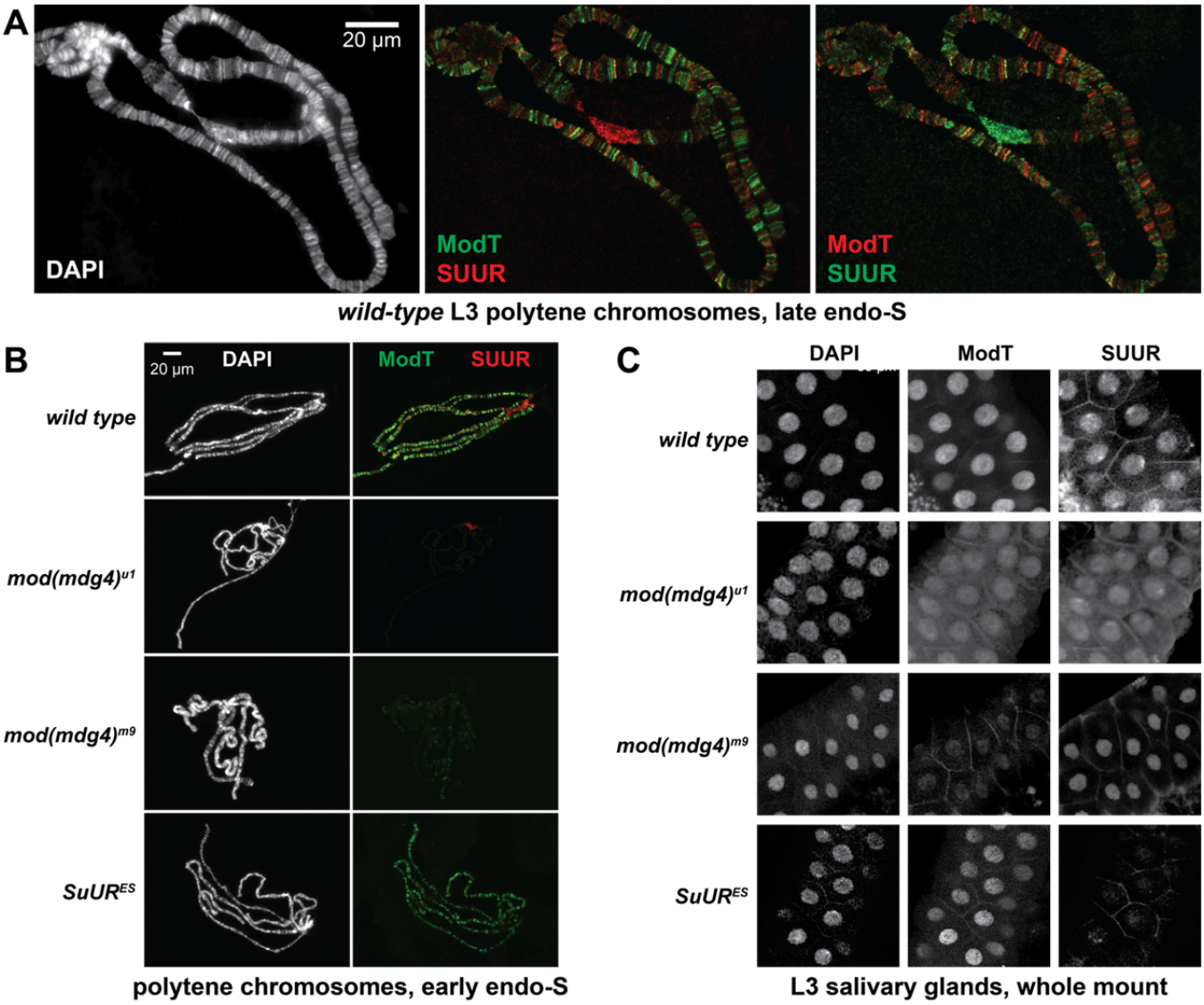


Figure 4. Spatiotemporal distribution of SUMM4 *in vivo*. (***A***) Colocalization of SUUR and Mod(Mdg4)-67.2 in *wild-type* polytene chromosomes. Localization patterns of Mod(Mdg4)-67.2 and SUUR in L3 polytene chromosomes were analyzed by indirect IF staining. The polytene spread fragment (3L and 3R arms) corresponds to a nucleus in late endo-S phase, according to PCNA staining (***Figure 4—figure supplement 1A***). Left panel, DAPI staining shows the overall chromosome morphology. Middle panel, ModT (green) and SUUR (red) signals overlap extensively in euchromatic arms. Right panel, a colocalization image with swapped red (ModT) and green (SUUR) channels is shown for comparison. Note the additional strong ModT IF loci that are SUUR-free as well as Mod(Mdg4)-67.2-free SUUR in pericentric 3LR. (***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 (***Figure 4—figure supplement 3B***). (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 the cytoplasm in *mod(mdg4)*^{u1} mutant.

Figure 4—figure supplement 1. Spatial distribution of SUUR and Mod(Mdg4)-67.2 in polytene chromosomes and analyses of their colocalization. (***A***) Colocalization of SUUR and Mod(Mdg4)-67.2 in *wild-type* polytene chromosomes. See legend to ***Figure 4A***. 3L and 3R telomeres are marked; approximate boundaries of cytological regions are shown according to (Lefevre, 1976); positions of intercalary heterochromatin regions 75C and 89E that are underreplicated and responsive to *SuUR* mutation are marked by circles. (***B***) The patterns of colocalization and independent loading of SUUR and Mod(Mdg4)-67.2 in *wild-type* polytene chromosomes. Subtracted and overlapping images were produced in ImageJ (*Materials and Methods*). Green, enriched Mod(Mdg4)-67.2 and low SUUR; red, enriched SUUR and low Mod(Mdg4)-67.2; magenta or cyan, overlapping enriched Mod(Mdg4)-67.2 and SUUR. (***C***) Quantification of the overlap between SUUR and Mod(Mdg4)-67.2 in *wild-type* polytene chromosomes (***Figure 4A***). Individual pixel intensities of anti-SUUR and anti-ModT IF signals are normalized to Z-scores and plotted on *x*- and *y*-axes, respectively (*Materials and Methods*); they exhibit a weak positive correlation (R^2>0.2). (***D***) Visually, the 2D plot (***C***) is split in four separate areas demarcated by Z_{ModT}= 1 and Z_{SUUR}= 3. When pixels representing ModT-only and SUUR-only areas (green and red, respectively) are removed, the remaining pixels that are simultaneously enriched for Mod(Mdg4)-67.2 and SUUR (blue) exhibit a strong positive correlation (R^2>0.5).

1267 ***Figure 4—figure supplement 2.*** Alternative complex(es) of Mod(Mdg4)-67.2. (***A***) Schematic of partial
1268 FPLC purification of an alternative complex of Mod(Mdg4)-67.2. Cyan boxes, fraction ranges used for
1269 the next chromatographic step. (***B***) Western blot analyses of Q Sepharose FF fractions with SUUR and
1270 ModT antibodies. SUUR and ~25% total Mod(Mdg4)-67.2 present in the starting material (SM)
1271 fractionate in the flow-through (FT, arrows), whereas Mod(Mdg4)-67.2 also fractionates as an
1272 additional, SUUR-free peak (cyan box). Molecular mass markers are as in ***Figure 2D***. (***C***) Western blot
1273 analysis of Source 15S fractions with the ModT antibody. (***D***) Western blot analyses of Superose 6
1274 fractions with the ModT antibody. Black arrows, expected peaks of globular proteins with indicated
1275 molecular masses in kDa.
1276
1277 ***Figure 4—figure supplement 3.*** Spatiotemporal distribution of SUMM4 subunits in polytene
1278 chromosomes of *mod(mdg4)* and *SuUR* mutant alleles. (***A***) Western blot analyses of lysates of whole
1279 salivary glands. L3 salivary glands from homozygous animals of indicated genotypes were probed with
1280 ModT (green) and β-tubulin antibodies (red, loading control). Mass marker sizes (kDa) are shown on the
1281 left. (***B***) Spatiotemporal distribution of SUUR in polytene chromosomes. See legend to ***Figure 4B***.
1282 Although SUUR is not properly loaded into *mod(mdg4)* chromosomes during early endo-S phase (as in
1283 wild type), its deposition partially recovers during late endo-S.

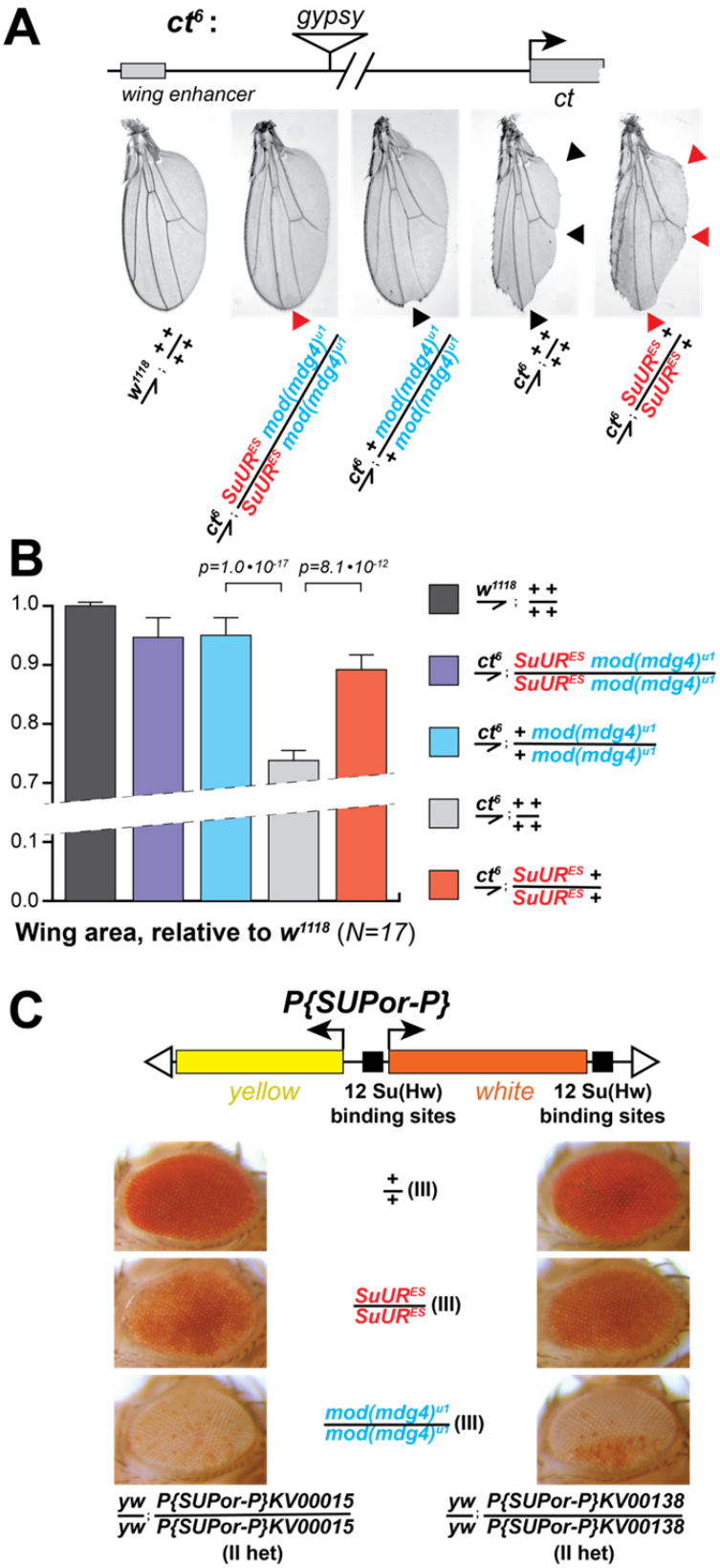


Figure 5. Biological functions of SUMM4 in regulation of gene expression. (***A***) SUMM4 subunits are required for the enhancer-blocking activity in ct^6. Top: schematic diagram of the ct^6 reporter system;

the *gypsy* retrotransposon is inserted in between the wing enhancer and promoter of *cut* (Bag, Dale, Palmer, & Lei, 2019). 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 the 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 (Bellen et al., 2004); 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 encroachment.

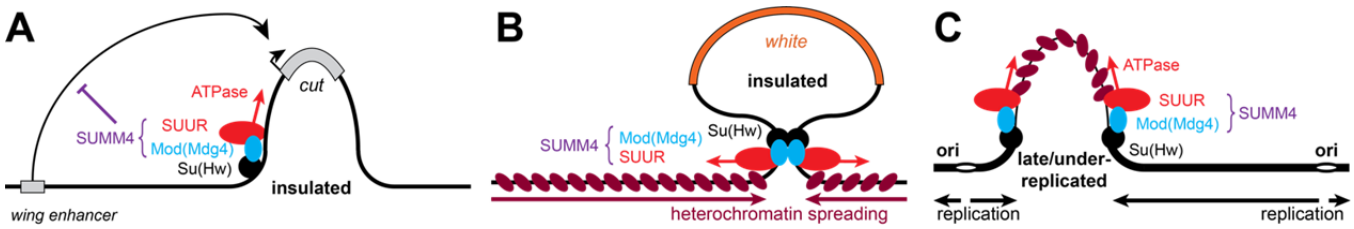


Figure 6. Schematic models for the biological functions of SUMM4 in regulation of gene expression and DNA replication. (***A***) Schematic model for the function of SUMM4 in blocking enhancer-promoter interactions in the ct^6 locus. A *gypsy* mobile element inserted between wing enhancer and gene *cut* encompasses multiple Su(Hw) binding sites. (***B***) Schematic model for the function of SUMM4 in establishing a chromatin barrier in heterochromatin-inserted *P{SUPor-P}* elements. The reporter gene *white* is flanked on both sides by 12 copies of *gypsy* insulator element. (***C***) Schematic model for a putative function of SUMM4 in blocking/retardation of replication fork progression in intercalary heterochromatin domains. Black oval, Su(Hw) protein bound to a *gypsy* insulator element(s); cyan oval, Mod(Mdg4)-67.2 protein tethered to Su(Hw); red oval, SUUR protein associated with Mod(Mdg4)-67.2 in SUMM4 complex; brown ovals represent heterochromatin components; gray rectangles, gene *cut* and its upstream wing enhancer; orange rectangle, gene *white*.

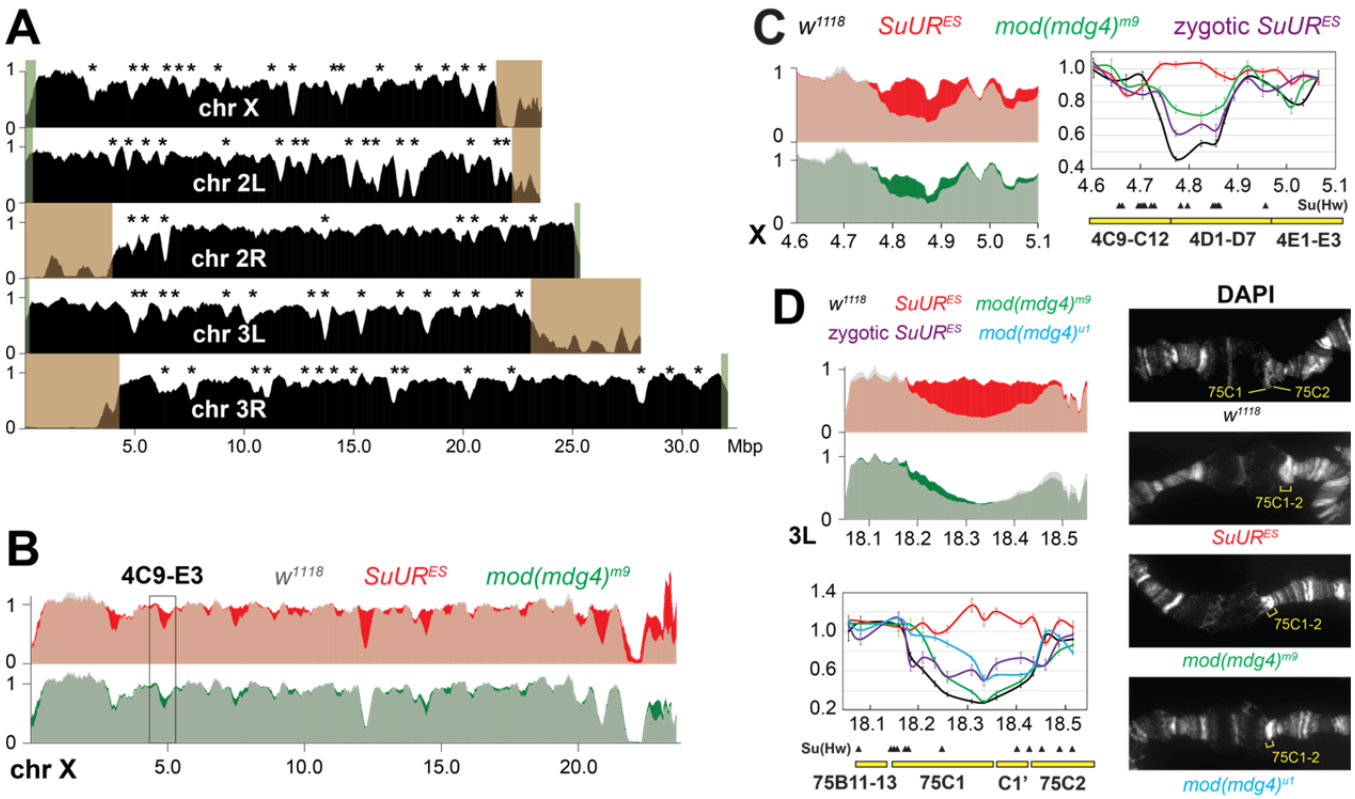


Figure 7. Biological functions of SUMM4 in regulation of DNA replication. (***A***) Genome-wide analyses of DNA copy numbers in *Drosophila* salivary gland cells (w^{1118} control). DNA from L3 salivary glands was subjected to high-throughput sequencing. DNA copy numbers (normalized to diploid embryonic DNA) are shown for chromosomes X, II and III. Chromosome arms are indicated in white. Brown- and green-shades boxes, mapped pericentric and telomeric heterochromatin regions (Hoskins et al., 2015), respectively. Asterisks, positions of underreplicated domains (***Table 1***). Genomic coordinates in Megabase pairs are indicated at the bottom. (***B***) Analyses of DNA copy numbers in *Drosophila* salivary gland cells from wild-type and mutant alleles. Normalized DNA copy numbers are shown across the X chromosome. The control trace (w^{1118} allele) is shown as semitransparent light gray in the foreground; $SuUR^{ES}$ (homozygous null) and $mod(mdg4)^{m9}$ (zygotic null from crosses of heterozygous parents) 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. (***C***) Close-up view of DNA copy numbers in region 4C9-E3 from high-throughput

1325 sequencing data are presented as in (***B***). DNA copy numbers were also measured independently by
1326 real-time qPCR. The numbers were calculated relative to embryonic DNA and normalized to a control
1327 intergenic region. The X-axis shows chromosome positions (in Megabase pairs) of target amplicons.
1328 Black, w^{1118}; red, $SuUR^{ES}$ (homozygous null); green, $mod(mdg4)^{m9}$ (zygotic null from crosses of
1329 heterozygous parents); purple, $SuUR^{ES}$ (zygotic null from crosses of heterozygous parents). Error bars
1330 represent the confidence interval (see *Materials and Methods*). Black arrowheads, positions of mapped
1331 Su(Hw) binding sites (Negre et al., 2010). Yellow boxes show approximate boundaries of cytogenetic
1332 bands. (***D***) Close-up view of DNA copy numbers by high-throughput sequencing and by qPCR for
1333 region 75B11-C2 and DAPI-stained polytene chromosome segments around cytological regions 75B-
1334 75C. Yellow lines or brackets in DAPI images indicate positions of 75C1 and 75C2 bands (w^{1118}
1335 control) or fused 75C1-2 band (mutants); cyan, $mod(mdg4)^{u1}$ (homozygous null); for other
1336 designations see (***C***).
1337
1338 ***Figure 7—source data 1.*** Primer sequences used for qPCR. Genomic coordinates indicate full
1339 amplicons, including the length of each primer. Coordinates refer to the BDGP R6/dm3 assembly.
1340
1341 ***Figure 7—figure supplement 1.*** Biological functions of SUMM4 in regulation of underreplication. (***A***)
1342 Genome-wide analyses of DNA copy numbers in *Drosophila* salivary gland cells in chromosome arms
1343 2L, 2R, 3L and 3R. The data were obtained and presented as for the X chromosome (***Figure 7B***). Black
1344 box, 75B11-C2 cytological region. (***B***) Close-up view of DNA copy numbers by high-throughput
1345 sequencing for additional genomic regions. Approximate cytogenetic locations are indicated at the top of
1346 each panel. Short vertical bars at the bottom, positions of mapped Su(Hw) binding sites (Negre et al.,
1347 2010). See legend to ***Figure 7C&D*** for other designations. (***C***) Sample plots of DamID profiles for
1348 SUUR (red) and Su(Hw) (purple), $\log_2$ enrichment over Dam-only control (Filion et al., 2010). Positive
1349 values are plotted in dark colors and negative values in light colors for contrast. DNA copy numbers in

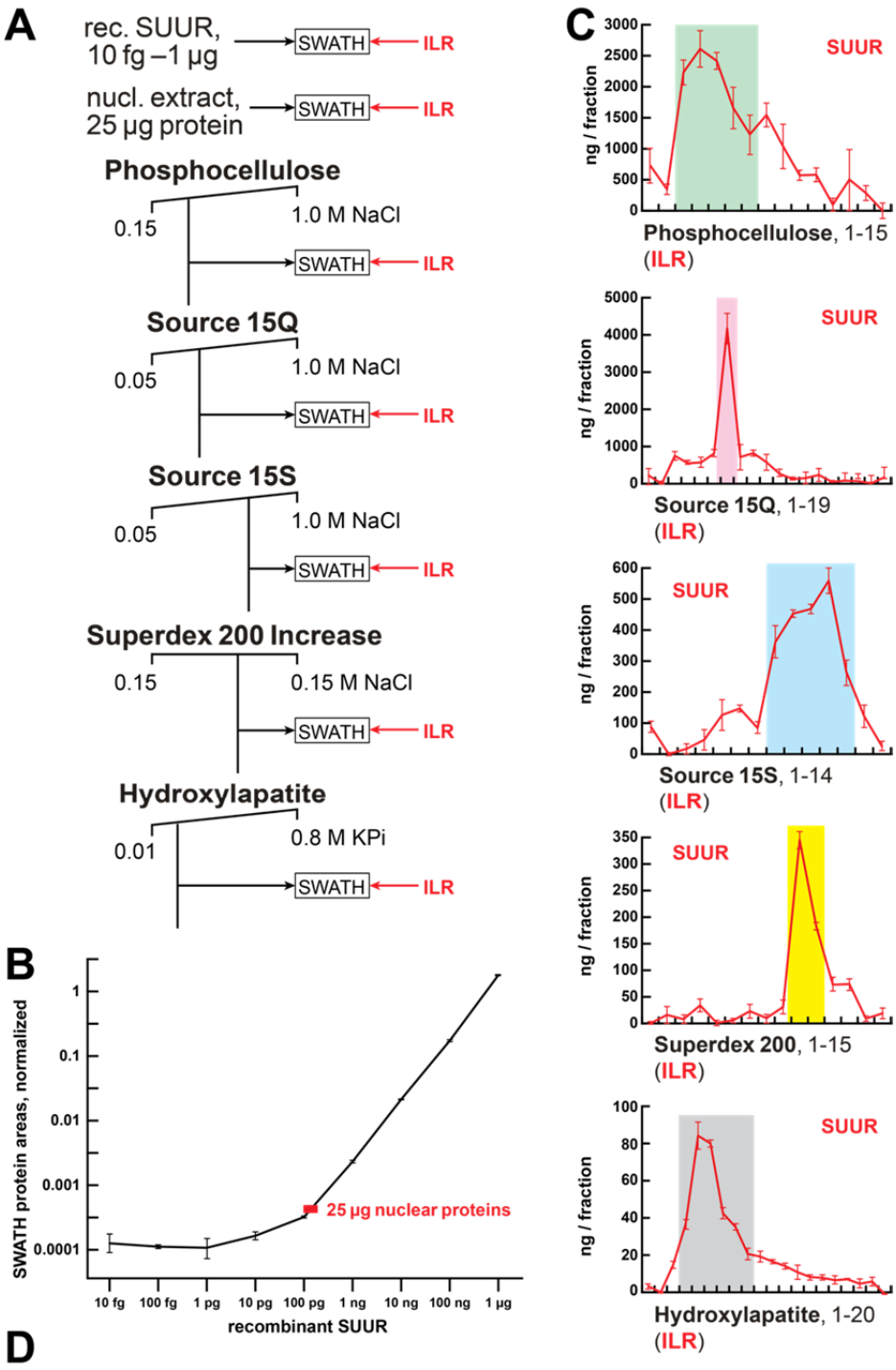
1350 salivary gland cells (black) indicate underreplicated intercalary heterochromatin domains. Vertical bars,

1351 Su(Hw) binding sites (Negre et al., 2010).

1352 ***Figure 1—source data 1.*** FPLC column parameters (***Figure 1A***). The following FPLC column
1353 parameters were used for partial purification of native SUMM4. HEG: 25 mM HEPES, pH 7.6, 0.1 mM
1354 EDTA, 10% glycerol, 0.02% NP-40, 1 mM DTT, 1 mM benzamidine, 0.4 mM PMSF; 10 mM KPi: 10
1355 mM potassium phosphate, pH 7.6, 10% glycerol, 1 mM DTT, 1 mM benzamidine, 0.4 mM PMSF; 0.8
1356 M KPi: 800 mM potassium phosphate, pH 7.6, 10% glycerol, 1 mM DTT, 1 mM benzamidine, 0.4 mM
1357 PMSF; *cv*, column volume.

Column	Phosphocellulose	Source 15Q	Source 15S	Superdex 200 Increase	Hydroxylapatite
Column volume, ml	48	4	0.5	24	0.5
Buffer A	HEG	HEG	HEG	HEG + 0.15 M NaCl	10 mM KPi
Buffer B	HEG + 1 M NaCl	HEG + 1 M NaCl	HEG + 1 M NaCl	N/A	0.8 M KPi
Starting material (SM)	nuclear extract	fxns 3-7 (Ph-Cell)	fxn 7 (15Q)	fxns 8-12 (15S)	fxns 10-11 (Superdex)
SM volume, ml	100	38	1.2	0.6	0.9
Diluted with	Buffer A	Buffer A	Buffer A	N/A	Buffer A
Dilution volume, ml	50	85	5	N/A	1.8
Equilibrate to, %B	10%	5%	5%	0%	0%
Column wash, cv	3	10	10	N/A	10
Elution gradient	10-100%	5-100%	5-100%	N/A	0-100%
Elution volume, cv	8	10	16	1.2	10
Fraction volume, ml	12	1.4	0.15	0.5	0.15

1358



D

step	SUUR	total protein	purity	purification factor	cumulative
nuclear extract	14 ± 1 µg	2.5 g	0.0006%	N/A	N/A
Phosphocellulose	7.9 ± 0.3 µg	80 mg	0.010%	17.6	18
Source 15Q	4.2 ± 0.4 µg	6.3 mg	0.07%	6.8	119
Source 15S	2.1 ± 0.1 µg	1.4 mg	0.15%	2.3	268
Superdex 200	530 ± 12 ng	70 µg	0.8%	5.0	1,352
Hydroxylapatite	300 ± 4 ng	16 µg	1.9%	2.5	3,348

Figure 1—figure supplement 1. Quantification of SUUR in chromatographic fractions. (***A***) Schematic of SWATH quantification of recombinant SUUR, nuclear extract (starting material) and FPLC fractions

1362 for SUUR using ion library ILR. (***B***) SUUR titration curve obtained by SWATH quantitation of 10 fg –
1363 1 µg recombinant FLAG-SUUR in the presence of 25 µg *E. coli* lysate; both axes are logarithmic
1364 ($\log_{10}$). Red rectangle, SUUR quantification in 25 µg nuclear extract; error bars, standard deviations
1365 (*N*=3). (***C***) SWATH quantitation profiles of SUUR fractionation across individual FPLC steps. Ion
1366 library ILR was used for SWATH quantification, and relative amounts were converted to estimated ng
1367 SUUR per fraction. Error bars, standard deviations (*N*=3); colored boxes, peak fractions of SUUR. (***D***)
1368 SUUR purification by FPLC. Total protein was measured by BCA assay, and SUUR was measured as in
1369 (***C***). Relative purity, purification factor in each step and cumulative purification factor are shown.

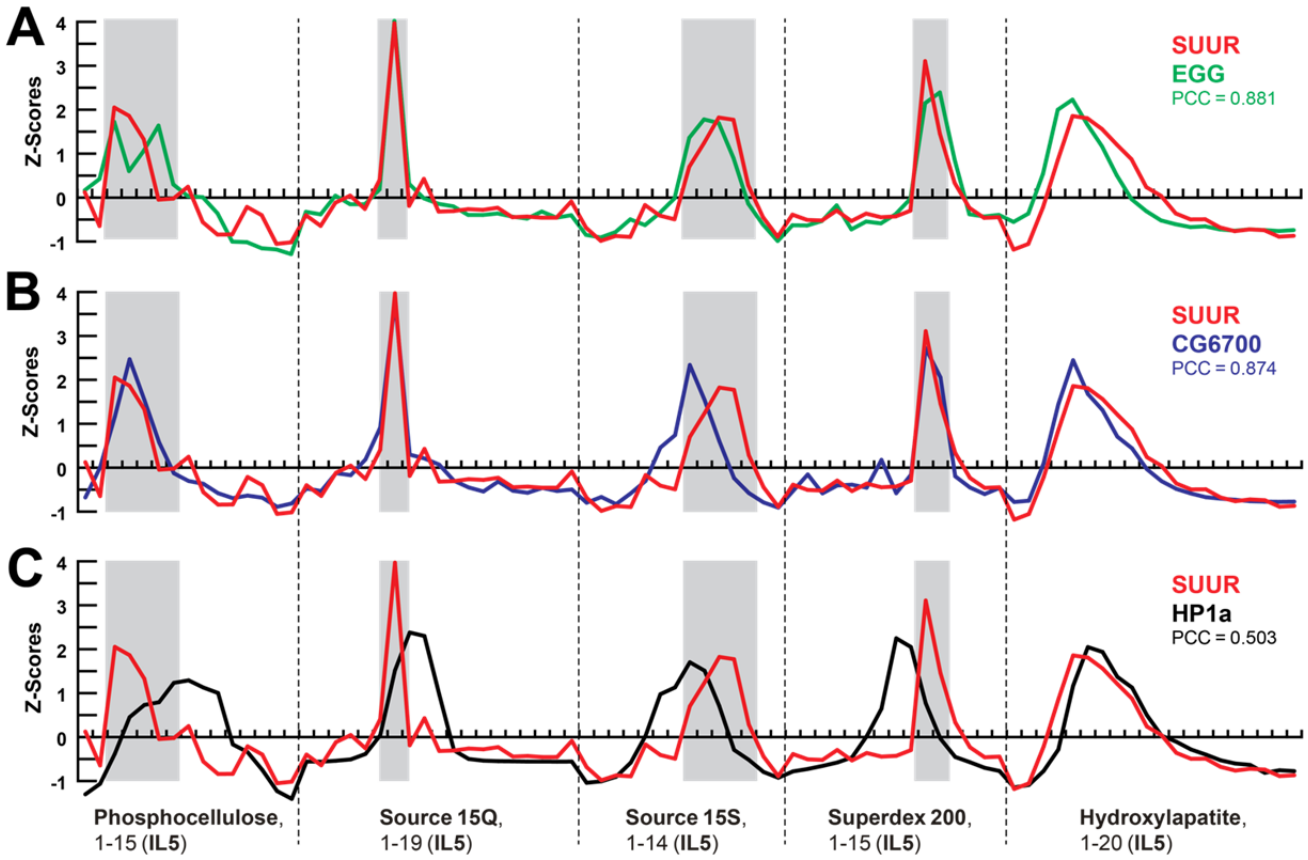


Figure 2—figure supplement 1. Comparisons of SWATH quantification profiles for protein fractionation. (***A-C***) SWATH quantitation of SUUR (red), EGG (***A***, green), CG6700 (***B***, blue) and HP1a (***C***, black) fractionation profiles across five FPLC steps as in ***Figure 1I&2C***. Pearson coefficients (PCC) are shown (***Figure 2A&B***).

A

sequence	confidence	theor. *m/z*	obs. *m/z*	*z*	RT, min	MS2 counts
SSLTPASSSAGVK	>99%	596.314	596.315	2	21.4	2,444.44
TSVVSAAEAK	>99%	481.761	481.762	2	22.1	786.02
LHQQSPQQVR	>99%	407.555	407.555	3	14.7	3,335.51
QEEAEYIDLPMELPTK	>99%	953.461	953.460	2	45.3	1,522.16
LQAATLNEEASEPAVY	>99%	853.418	853.418	2	42.1	633.45
VFPYEGEHVHFMQASDK	>99%	505.986	505.986	4	41.8	2,216.17
SCLPSQFMPGESGVISSLSPSK	>99%	765.705	765.703	3	45.2	1,011.73

B **Mod(Mdg4)-PT (67.2)**

mod(mdg4): exons 2-4 *pre-mod(mdg4)-T*: exons 2-3

```
  1 MADDEQFSLCWNNFNTNLSAGFHESLCRGDLVDVSLAAEGQIVKAHRLVLSVCSPFFRKMFTQMPSNTHAIVFLNNVSHSALKDLIQFMYCG  92
 93 EVNVKQDALPAFISTAESLQIKGLTDNDPAPQPPQESSPPPAAPHVQQQQIPAQRVQRQQPRASARYKIETVDDGLGDEKQSTTQIVIQTTA 184
185 APQATIVQQQQPQQAAQQIQSQQLQTGTTTTATLVSTNKRSAQRSSLTPASSSAGVKRSKTSTSANVMDPLDSTTETGATTTAQLVPQQITV 276
277 QTSVVSAAEAKLHQQSPQQVRQEEAEYIDLPMELPTKSEPDYSEDHGDAAGDAEGTYVEDDTYGDMRYDDSYFTENEDAGNQTAANTSGGGV 368
369 TATTSKAVVKQQSQNYSESSFVDTSGDQGNTEAQAATSASATKIPPRKRGRPKTKVEDQTPKPKLLEKLQAATLNEEASEPAVYASTTKGGV 460
461 KLIFNGHLFKFSFRKADYSVFQCCYREHGEECKVRVVCDQKRVFPYEGEHVHFMQASDKSCLPSQFMPGESGVISSLSPSKELLMKNTTKLE 552
553 EADDKEDEDFEEFEIQEIDEIELDEPEKTPAKEEEVDPNDFREKIKRRLQKALQNKKK                                  610
```

C **Mod(Mdg4)-PI (59.1)**

mod(mdg4): exons 2-4 *pre-mod(mdg4)-I*: exons 4-5

```
  1 MADDEQFSLCWNNFNTNLSAGFHESLCRGDLVDVSLAAEGQIVKAHRLVLSVCSPFFRKMFTQMPSNTHAIVFLNNVSHSALKDLIQFMYCG  92
 93 EVNVKQDALPAFISTAESLQIKGLTDNDPAPQPPQESSPPPAAPHVQQQQIPAQRVQRQQPRASARYKIETVDDGLGDEKQSTTQIVIQTTA 184
185 APQATIVQQQQPQQAAQQIQSQQLQTGTTTTATLVSTNKRSAQRSSLTPASSSAGVKRSKTSTSANVMDPLDSTTETGATTTAQLVPQQITV 276
277 QTSVVSAAEAKLHQQSPQQVRQEEAEYIDLPMELPTKSEPDYSEDHGDAAGDAEGTYVEDDTYGDMRYDDSYFTENEDAGNQTAANTSGGGV 368
369 TATTSKAVVKQQSQNYSESSFVDTSGDQGNTEAQVCDDLDDMKGAIKHSLLTFIRGQRGCKLLAFNGHNYVRNRRSNLKTYWICSKKGSTKC 460
461 NARVVTNVVEGVHKIVLESCHHTCLNTERKKRLSVTNVVGKARSKSEKSVSTGFIKEEGDEDLTLELRTLNLSIEDLNNLQ           541
```

Figure 2—figure supplement 2. Identification of Mod(Mdg4)-67.2 as a subunit of the SUMM4 complex. (***A***) Mod(Mdg4)-specific peptides from ion library IL5 (***Supplementary File 1***). Gray shading, peptides specific to the common part (coding exons 2-4) of Mod(Mdg4); cyan shading, peptides specific to polypeptide Mod(Mdg4)-67.2 encoded by *pre-mod(mdg4)-T*, exons 2-3. Peptide sequences, confidence levels, charges (*z*), theoretical and observed *m/z*, column retention times (RT) and total MS2 ion counts are shown. (***B***) Mod(Mdg4)-67.2 polypeptide sequence. The common part is shaded in gray, splice form-specific part is shaded in cyan. Peptides from ion library IL5 (as in E) are highlighted in bold red. (***C***) Mod(Mdg4)-59.1 polypeptide sequence. The common part is shaded in gray, splice form-specific part is shaded in light green.

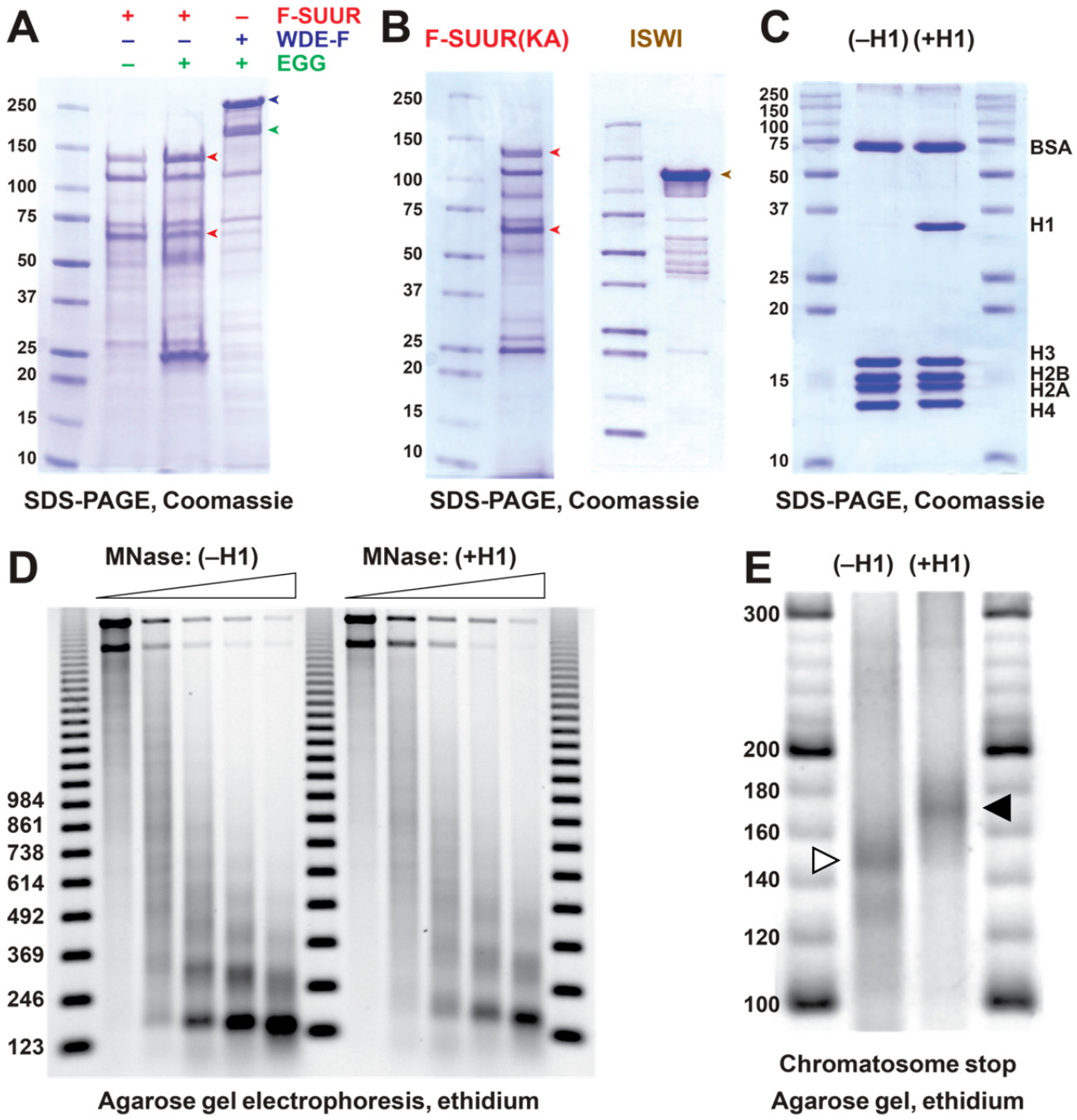


Figure 3—figure supplement 1. Recombinant proteins and biochemical substrates. (***A***) Physical interactions of recombinant EGG, SUUR and WDE. Untagged EGG (green arrowhead) was co-expressed with FLAG-SUUR (red arrowheads, p130 and p65) or WDE-FLAG (purple arrowhead) in Sf9 cells and purified by FLAG affinity chromatography. EGG forms a specific complex with WDE but not SUUR. Molecular mass markers (kDa) are shown on the left. (***B***) Recombinant FLAG-SUUR(K59A) expressed in Sf9 cells and ISWI expressed in *E. coli*. See legend to ***Figure 1C***. (***C***) Protein composition

of *in vitro* reconstituted chromatin. Oligonucleosomes prepared from plasmid DNA and core histones with (+H1) or without H1 (–H1) were analyzed by SDS-PAGE and Coomassie staining. Positions of BSA, H1 and core histone bands are indicated on the right. (***D***) Micrococcal nuclease (MNase) analysis of reconstituted chromatin. Partial digestion with five different dilutions of MNase was performed on H1-free (–H1) and H1-containing (+H1) oligonucleosomes. Deproteinated DNA fragments were analyzed by agarose gel electrophoresis and stained with ethidium. Note the increased nucleosome repeat length in (+H1) lanes consistent with H1 incorporation. Triangles at the top indicate increasing MNase concentrations; 123 bp ladder was used as a molecular mass marker. (***E***) Chromatosome stop assay. Oligonucleosomes assembled with or without H1 were subjected to partial MNase digestion, and DNA was analyzed by agarose gel electrophoresis and ethidium bromide staining. Positions of the core particle and chromatosome DNA are indicated by arrowheads. DNA fragment sizes in the 20-bp DNA ladder marker are shown.

Figure 3—figure supplement 1—source data 1. Recombinant proteins expressed in Sf9 cells and purified
by FLAG or Ni-NTA affinity chromatography. Lane 1, protein size marker; lane 2, FLAG-SUUR, FLAG-
purified; lane 3, FLAG-SUUR + EGG untagged, FLAG-purified; lane 4, WDE-FLAG + EGG untagged, FLAG-purified; lane 5, EGG-FLAG, FLAG-purified; lane 6, EGG-FLAG + His6-SUUR, FLAG-purified;
lane 7, WDE-FLAG + EGG untagged + His6-SUUR, FLAG-purified; lane 8, His6-SUUR, Ni-NTA-purified; lane 9, EGG-FLAG + His6-SUUR, Ni-NTA-purified; lane 10, WDE-FLAG + EGG untagged + His6-SUUR, Ni-NTA-purified. All proteins were purified 72 hours post-infection. Prep amounts equivalent to ~20 ml Sf9 cultures were loaded in each lane. Cropped image encompassing lanes 1-4 (open

box, dashed red line) was used for ***Figure 3—figure supplement 1A***.

Figure 3—figure supplement 1—source data 2. Recombinant proteins expressed in Sf9 or *E. coli* cells and purified by FLAG, Ni-NTA or chitin affinity chromatography. Lanes 1, 7 and 10, protein size marker;
lane 2, FLAG-SUUR(K59A), FLAG-purified; lane 3, FLAG-SUUR(K59R), FLAG-purified; lane 4, FLAG-SUUR wild-type, FLAG-purified; lane 5, His6-SUUR wild-type, Ni-NTA-purified. All proteins were purified 72 hours post-infection. Prep amounts equivalent to ~20 ml Sf9 cultures were loaded in each
lane. Lane 6, FLAG-ISWI (Sf9 ells), FLAG-purified, 2 µg; lane 8, ISWI untagged (*E. coli*), chitin-purified, 0.5 µg; lane 9, ISWI untagged (*E. coli*), chitin-purified, 1 µg; lane 11, ISWI untagged (*E. coli*), chitin-purified, 2 µg. Cropped images encompassing lanes 1-2 and 10-11 (open boxes, dashed red line) were used for ***Figure 3—figure supplement 1B***.

Figure 3—figure supplement 1—source data 3. SDS-PAGE of salt dialyzed chromatin ±H1. Lanes 1 and 4, protein size marker; lane 2, oligonucleosomes, 2 µg DNA; lane 3, oligonucleosomes +H1, 2 µg DNA. Cropped image encompassing all lanes (open box, dashed red line) was used for ***Figure 3—figure supplement 1C***.

Figure 3—figure supplement 1—source data 4. Micrococcal nuclease (MNase) digest of salt-dialyzed chromatin ±H1, 1.25% agarose gel, ethidium-stained. Lanes 1, 7 and 13, 123-bp DNA ladder; lanes 2-6, oligonucleosomes, 250 ng DNA; lanes 8-12, oligonucleosomes +H1, 250 ng DNA; lanes 2 and 8, digested with 10^{-4} units MNase for 15 min at room temperature (RT); lanes 3 and 9, digested with $3 \cdot 10^{-4}$ units MNase for 15 min at RT; lanes 4 and 10, digested with 10^{-3} units MNase for 15 min at RT; lanes 5

1441 and 11, digested with $3 \cdot 10^{-3}$ units MNase for 15 min at RT; lanes 6 and 12, digested with 10^{-2} units

1442 MNase for 15 min at RT. Cropped image encompassing all lanes (open box, dashed red line) was used

1443 for ***Figure 3—figure supplement 1D***.

1444

1445 ***Figure 3—figure supplement 1—source data 5.*** Micrococcal nuclease (MNase) digest of salt-dialyzed

1446 chromatin ±H1, 3% agarose gel, ethidium-stained (chromatosome stop assay). Lanes 1, 4, 5, 8, 9, 12, 13

1447 and 16, 20-bp DNA ladder; lanes 2, 6, 10 and 14, oligonucleosomes; lanes 3, 7, 11 and 15,

1448 oligonucleosomes +H1; lanes 2 and 3, digested with $5 \cdot 10^{-3}$ units MNase for 15 min at RT; lanes 6 and

1449 7, digested with $1.5 \cdot 10^{-2}$ units MNase for 15 min at RT; lanes 10 and 11, digested with 10^{-2} units

1450 MNase for 15 min at RT; lanes 14 and 15, digested with $3 \cdot 10^{-2}$ units MNase for 15 min at RT; lanes 2,

1451 3, 6 and 7, 125 ng DNA; lanes 10, 11, 14 and 15, 250 ng DNA. Cropped image (open box, dashed red

1452 line) was used for ***Figure 3—figure supplement 1E***.

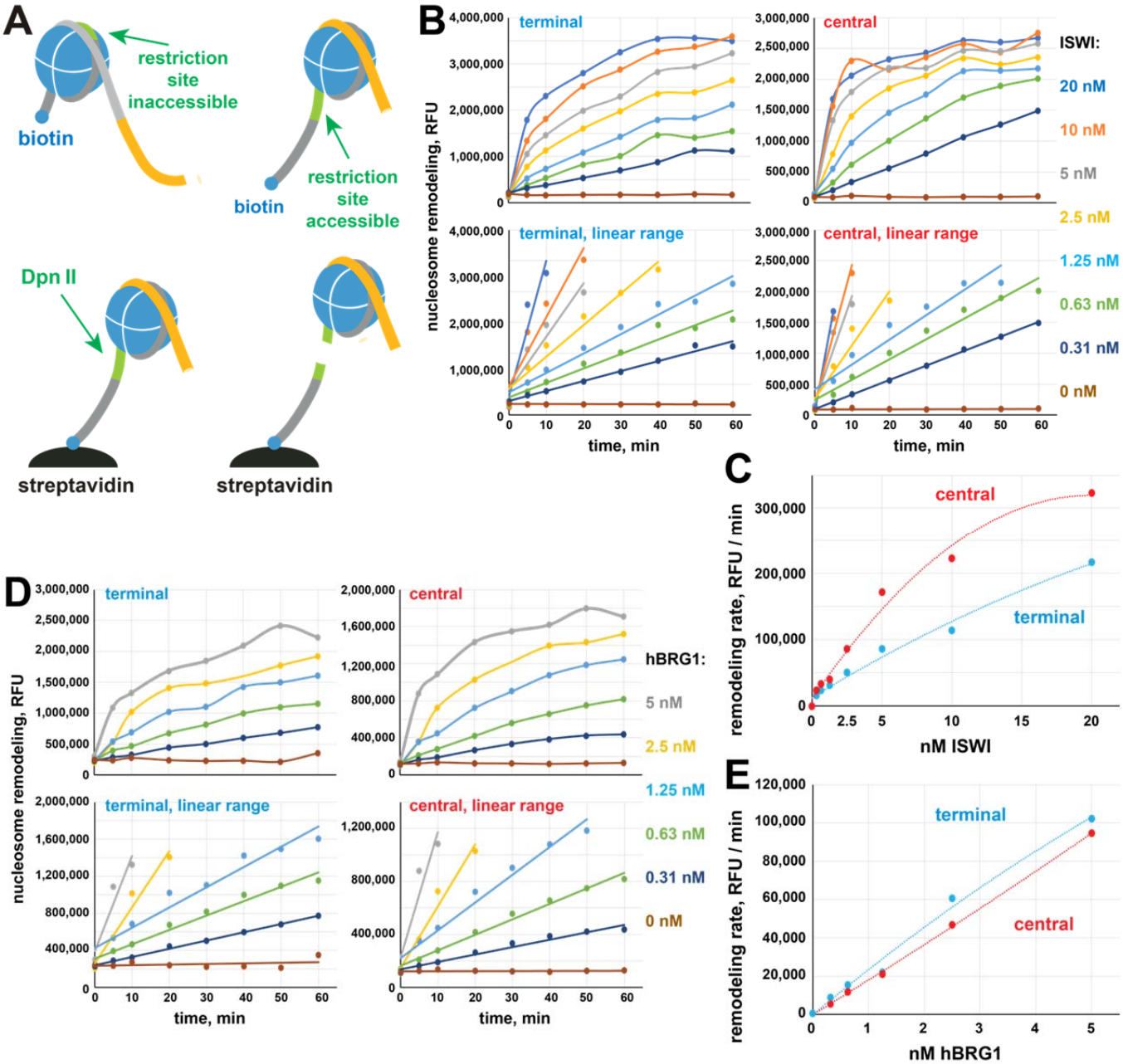


Figure 3—figure supplement 2. EpiDyne®-PicoGreen™ biochemical assay. (***A***) EpiCypher® EpiDyne®-PicoGreen™ assay design. EpiDyne nucleosomes encompass a restriction site shielded by the initial nucleosome position but exposed for Dpn II cleavage upon remodeling (sliding or displacement). Biotinylated substrates are immobilized on streptavidin magnetic beads. Digest by Dpn II releases the substrates from beads, and supernatant is quantified by PicoGreen™ (dsDNA detection reagent) fluorescence. (***B***) Titration of *Drosophila* ISWI remodeling activity using terminally (6-N-66) or centrally (50-N-66) positioned mononucleosomes. Early reaction time points were separately plotted to

1461 indicate linear ranges. RFU, relative fluorescence units. (***C***) Early remodeling rates for ISWI were
1462 calculated by linear regression analyses of data in respective linear ranges. ISWI exhibits a stronger
1463 remodeling activity with a centrally positioned nucleosome substrate. (***D***) Titration of human BRG1
1464 remodeling activity. Data are presented as in (***B***). (***E***) Early remodeling rates for BRG1 were calculated
1465 and plotted as in (***D***). BRG1 does not exhibit a bias towards remodeling centrally or terminally
1466 positioned nucleosomes.

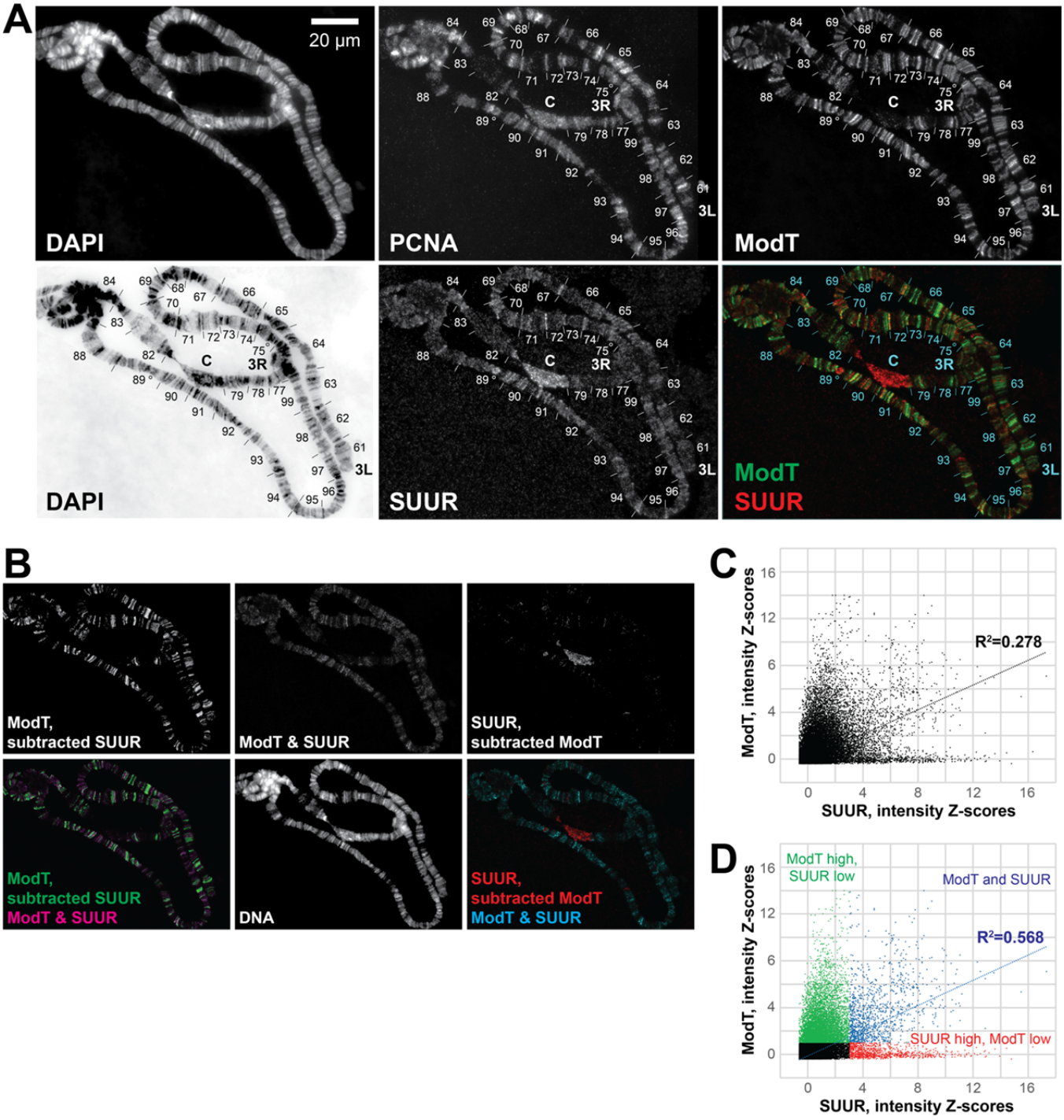


Figure 4—figure supplement 1. Spatial distribution of SUUR and Mod(Mdg4)-67.2 in polytene chromosomes and analyses of their colocalization. (***A***) Colocalization of SUUR and Mod(Mdg4)-67.2 in *wild-type* polytene chromosomes. See legend to ***Figure 4A***. 3L and 3R telomeres are marked; approximate boundaries of cytological regions are shown according to (Lefevre, 1976); positions of intercalary heterochromatin regions 75C and 89E that are underreplicated and responsive to *SuUR*

1473 mutation are marked by circles. (***B***) The patterns of colocalization and independent loading of SUUR
1474 and Mod(Mdg4)-67.2 in *wild-type* polytene chromosomes. Subtracted and overlapping images were
1475 produced in ImageJ (*Materials and Methods*). Green, enriched Mod(Mdg4)-67.2 and low SUUR; red,
1476 enriched SUUR and low Mod(Mdg4)-67.2; magenta or cyan, overlapping enriched Mod(Mdg4)-67.2
1477 and SUUR. (***C***) Quantification of the overlap between SUUR and Mod(Mdg4)-67.2 in *wild-type*
1478 polytene chromosomes (***Figure 4A***). Individual pixel intensities of anti-SUUR and anti-ModT IF signals
1479 are normalized to Z-scores and plotted on *x*- and *y*-axes, respectively (*Materials and Methods*); they
1480 exhibit a weak positive correlation (R^2>0.2). (***D***) Visually, the 2D plot (***C***) is split in four separate areas
1481 demarcated by Z_{ModT}= 1 and Z_{SUUR}= 3. When pixels representing ModT-only and SUUR-only areas
1482 (green and red, respectively) are removed, the remaining pixels that are simultaneously enriched for
1483 Mod(Mdg4)-67.2 and SUUR (blue) exhibit a strong positive correlation (R^2>0.5).

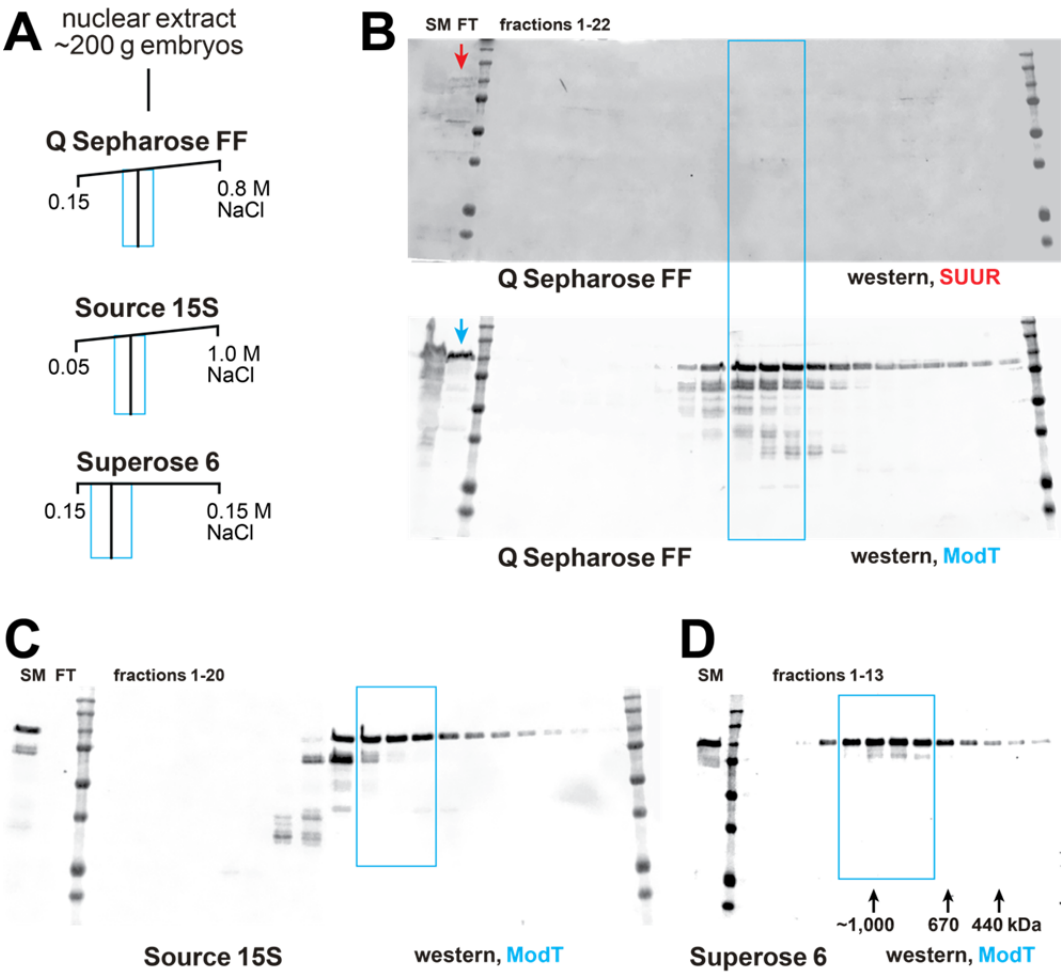


Figure 4—figure supplement 2. Alternative complex(es) of Mod(Mdg4)-67.2. (***A***) Schematic of partial FPLC purification of an alternative complex of Mod(Mdg4)-67.2. Cyan boxes, fraction ranges used for the next chromatographic step. (***B***) Western blot analyses of Q Sepharose FF fractions with SUUR and ModT antibodies. SUUR and ~25% total Mod(Mdg4)-67.2 present in the starting material (SM) fractionate in the flow-through (FT, arrows), whereas Mod(Mdg4)-67.2 also fractionates as an additional, SUUR-free peak (cyan box). Molecular mass markers are as in ***Figure 2D***. (***C***) Western blot analysis of Source 15S fractions with the ModT antibody. (***D***) Western blot analyses of Superose 6 fractions with the ModT antibody. Black arrows, expected peaks of globular proteins with indicated molecular masses in kDa.

1495 ***Figure 4—figure supplement 2—source data 1.*** FPLC column parameters (***Figure 4—figure***
1496 ***supplement 2A***). The following FPLC column parameters were used for partial purification of an
1497 alternative complex of Mod(Mdg4)-67.2. HEG: 25 mM HEPES, pH 7.6, 0.1 mM EDTA, 10% glycerol,
1498 0.02% NP-40, 1 mM DTT, 1 mM benzamidine, 0.4 mM PMSF; *cv*, column volume.
1499

1500 ***Figure 4—figure supplement 2—source data 2.*** Western blots of chromatographic fractions. Left
1501 panels, 700 nm channel (Odyssey Fc), rabbit anti-SUUR antibody and protein size marker; right panels,
1502 800 nm channel (Odyssey Fc), Guinea pig ModT antibody; top pannels, Q Sepharose FF fractions:
1503 starting material, flow-through, marker, fractions 1-10; bottom pannels, Q Sepharose FF fractions:
1504 fractions 11-22, marker. Cropped and spliced images encompassing all panels (open boxes, dashed red
1505 line) were used for ***Figure 4—figure supplement 2B***.
1506

1507 ***Figure 4—figure supplement 2—source data 3.*** Western blots of chromatographic fractions. Left
1508 panels, 700 nm channel (Odyssey Fc), rabbit anti-SUUR antibody and protein size marker; right panels,
1509 800 nm channel (Odyssey Fc), Guinea pig ModT antibody; top pannels, Source 15S fractions: starting
1510 material, flow-through, marker, fractions 1-9, empty, marker; bottom pannels, Source 15S fractions:
1511 marker, empty, fractions 10-20, marker. Cropped images encompassing all panels (open boxes, dashed
1512 red line) were used for ***Figure 4—figure supplement 2C***.
1513

1514 ***Figure 4—figure supplement 2—source data 4.*** Western blots of chromatographic fractions. Left
1515 panel, 700 nm channel (Odyssey Fc), protein size marker; right panels, 800 nm channel (Odyssey Fc),
1516 Guinea pig ModT antibody. Superose 6 fractions: starting material, marker, fractions 1-13. Cropped
1517 images from both panels (open boxes, dashed red line) were used for ***Figure 4—figure supplement 2D***.

1518 ***Figure 4—figure supplement 2—source data 1.*** FPLC column parameters (***Figure 4—figure***

1519 ***supplement 2A***). The following FPLC column parameters were used for partial purification of an

1520 alternative complex of Mod(Mdg4)-67.2. HEG: 25 mM HEPES, pH 7.6, 0.1 mM EDTA, 10% glycerol,

1521 0.02% NP-40, 1 mM DTT, 1 mM benzamidine, 0.4 mM PMSF; *cv*, column volume.

Column	**Q Sepharose FF**	**Source 15S**	**Superose 6**
Column volume, ml	10	1	24
Buffer A	HEG	HEG	HEG + 0.15 M NaCl
Buffer B	HEG + 1 M NaCl	HEG + 1 M NaCl	N/A
Starting material (SM)	nuclear extract	fxns 11-13 (Q)	fxn 11-13 (15S)
SM volume, ml	20	8.5	0.6
Diluted with	Buffer A	Buffer A	N/A
Dilution volume, ml	10	25	N/A
Equilibrate to, %B	10%	5%	0%
Column wash, cv	5	6	N/A
Elution gradient	10-100%	5-100%	N/A
Elution volume, cv	12	12	1.2
Fraction volume, ml	3	0.25	0.5

1522

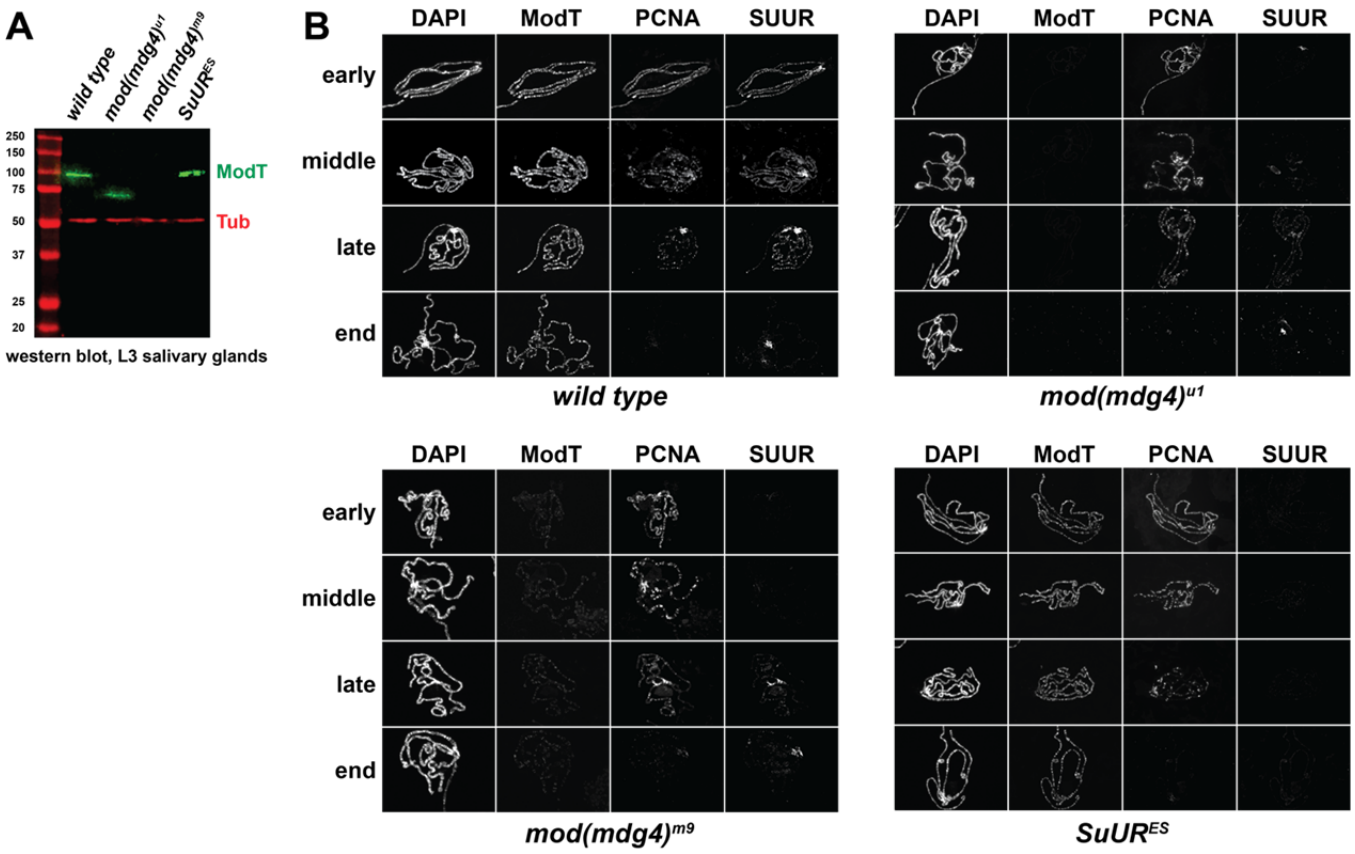


Figure 4—figure supplement 3. Spatiotemporal distribution of SUMM4 subunits in polytene chromosomes of *mod(mdg4)* and *SuUR* mutant alleles. (***A***) Western blot analyses of lysates of whole salivary glands. L3 salivary glands from homozygous animals of indicated genotypes were probed with ModT (green) and β-tubulin antibodies (red, loading control). Mass marker sizes (kDa) are shown on the left. (***B***) Spatiotemporal distribution of SUUR in polytene chromosomes. See legend to ***Figure 4B***. Although SUUR is not properly loaded into *mod(mdg4)* chromosomes during early endo-S phase (as in wild type), its deposition partially recovers during late endo-S.

Figure 4—figure supplement 3—source data 1. Western blots of salivary gland lysates. Left panel, 700 nm channel (Odyssey Fc), mouse anti-tubulin antibody and protein size marker; right panels, 800 nm channel (Odyssey Fc), Guinea pig ModT antibody. Lanes 1, protein size marker; lanes 2, L3 salivary glands, wild type; lanes 3, L3 salivary glands, *mod(mdg4)*^{u1}; lanes 4, L3 salivary glands, *mod(mdg4)*^{m9}; lanes 5, L3 salivary glands, *SuUR*^{ES}. Cropped images from both panels (open boxes, dashed red line)

overlayed in different colors (left panel, red; right panel, green) were used for ***Figure 4—figure supplement 3A***.

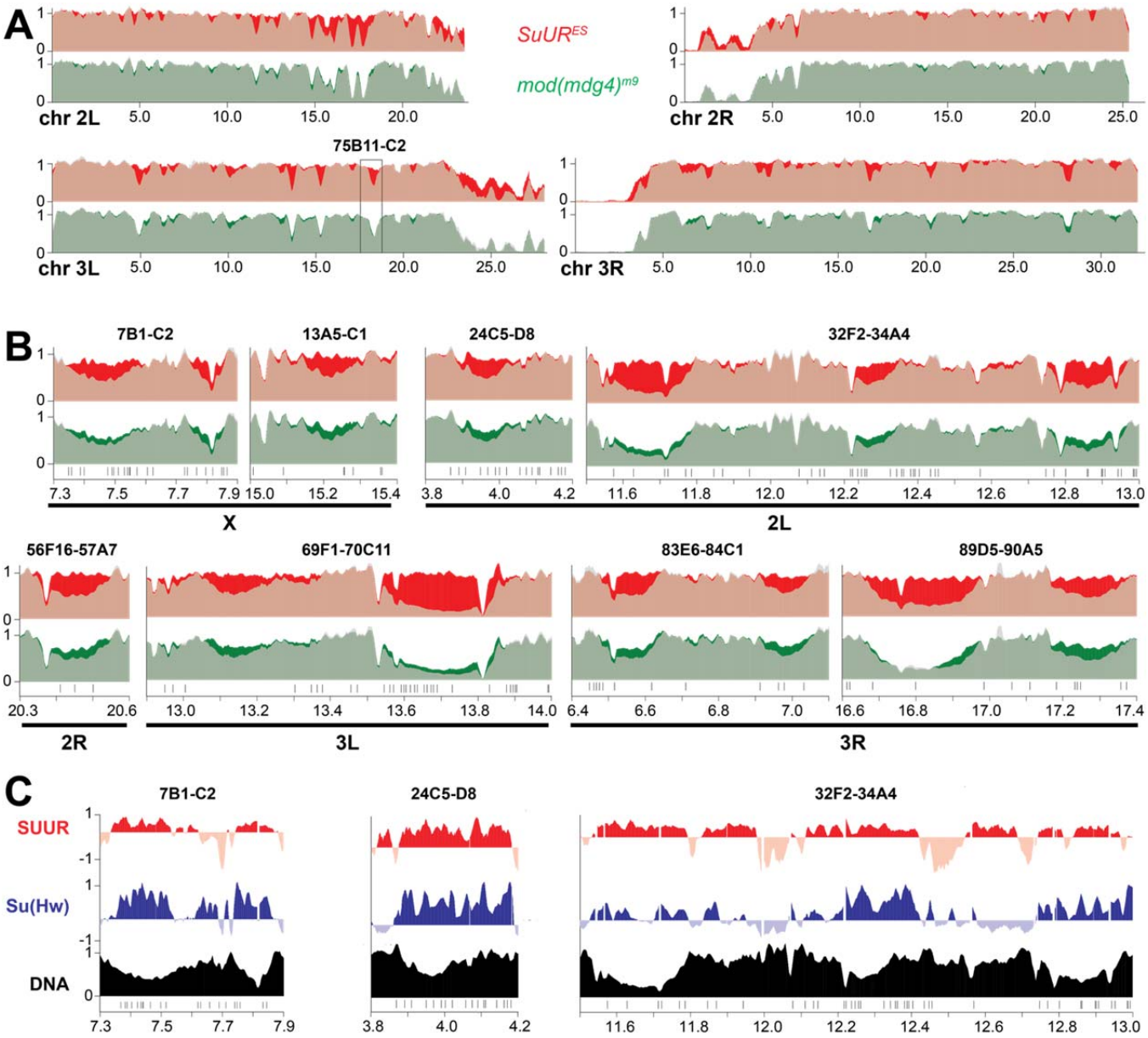


Figure 7—figure supplement 1. Biological functions of SUMM4 in regulation of underreplication. (***A***) Genome-wide analyses of DNA copy numbers in *Drosophila* salivary gland cells in chromosome arms 2L, 2R, 3L and 3R. The data were obtained and presented as for the X chromosome (***Figure 7B***). Black box, 75B11-C2 cytological region. (***B***) Close-up view of DNA copy numbers by high-throughput sequencing for additional genomic regions. Approximate cytogenetic locations are indicated at the top of each panel. Short vertical bars at the bottom, positions of mapped Su(Hw) binding sites (Negre et al., 2010). See legend to ***Figure 7C&D*** for other designations. (***C***) Sample plots of DamID profiles for SUUR (red) and Su(Hw) (purple), $\log_2$ enrichment over Dam-only control (Filion et al., 2010). Positive

1548 values are plotted in dark colors and negative values in light colors for contrast. DNA copy numbers in
1549 salivary gland cells (black) indicate underreplicated intercalary heterochromatin domains. Vertical bars,
1550 Su(Hw) binding sites (Negre et al., 2010).

1551 ***Figure 7—source data 1.*** Primer sequences used for qPCR. Genomic coordinates indicate full

1552 amplicons, including the length of each primer. Coordinates refer to the BDGP R6/dm3 assembly.

Cytological location	4C9-E3
Sequence	**Genomic coordinates**
CCTCGATCGGTTTACATTCG CCATAAACCCAAACGAGCTG	X:4,607,333..4,607,433
CACATGGTGTCCTTGCATTC GCCTAAACCAGCGATTCAAC	X:4,643,535..4,643,626
GGGATGTGCTGCCTTTTATG AGTTGCCACGACCAAAACTC	X:4,673,491..4,673,595
TGAAGGCCCTGGATGATAAG TGGCATAGATATCGGTGTGC	X:4,706,888..4,706,995
GGCTTGATTTTCGACTGCTC AAAGGAAACAGCTCCGTGTG	X:4,742,052..4,742,153
TTGCAGTGCCTCAAAGTCAG ACCGACCAAAATCGAGACTG	X:4,774,644..4,774,740
CCTATCACCTGCCCATTTTG TTACGTCCCTGGTTTCTTGC	X:4,826,332..4,826,430
AGCCATCCTGTTGCATCTTC GCGCCAACAAATTCTCTCAG	X:4,856,456..4,856,547
ACCTCGCCAACATTACCAAC AAACAACACGACGGCTCTTC	X:4,873,801..4,873,880
AACTGCCCAAAGTGAAGGTG GTTCAAGTGCAGCCAATGTG	X:4,893,272..4,893,370
CGGCAAACACGACTACAATG CAGTCGGATGCTGGTAGATATG	X:4,920,840..4,920,943
AGCATGGACCCATCGATTAC TTTCCCTGGGTAGCATTCAC	X:4,951,780..4,951,879
GAGATGCAAGATGCCACAAG CCTTAGAGCGCTTCAATTCG	X:4,982,299..4,982,391
AGGCAACCTGCAACTGAAAC ACAATTGCGTACGTGAGCTG	X:5,009,757..5,009,859
GTCTTGGAGTTGCCGTTTTG TGCGCTGATCTCGTTAGATG	X:5,033,854..5,033,945
CTAACCATCGCCAAATCCTC CGTCCACAATTAGCTTGCAG	X:5,064,863..5,064,959
TCCCTGCGACAACCTTTAAC CTCCGTGACATGCTTGATTC	X:5,097,851..5,097,941

Cytological location	75B11-C2
Sequence	**Genomic coordinates**
ATTTGGACTGGGGCAGTTTC CTGAAACACGGAAGTTGAGTCC	3L:18,062,106..18,062,230
AAAACACAAGCACATAGGCAAC AGTTTCTGGCGTTGTATCCG	3L:18,087,066..18,087,175
GTGCACGGACGCGTATAATC AAGTTAGCTCACGTGAGATGATG	3L:18,164,427..18,164,499
ACTATTATTTCTGGCTGGCTACG GCCGGCTGCTACTTATGGC	3L:18,188,845..18,188,948
ATACAGATACAGCTCGCACTGG AGTGGTGCCGATGGAAAAAC	3L:18,214,103..18,214,210
ACCACGCCCCTAAGCAAATAG ATCTCGCCAGCTAAAGATCTCG	3L:18,238,935..18,239,021
TGGGGCATTTTTGACGGTAG GCTTTTAGCCTCGAGAAACCG	3L:18,263,954..18,264,043
CTTGGCTCAGGTTTCCCTTC AAAGGACGCCACAACAATGC	3L:18,313,914..18,314,025
ATCTCTCTGGGGCATCCAAG CGCCAGCGCAGTTAAAAGTAAC	3L:18,338,911..18,339,046
TGCACCAAGCTACACAATGG CACAGGACTCCAAATTCTGCAC	3L:18,364,090..18,364,232
AGTGATAGCGGAGTAACAGTGG GTGGCGTGGATCCAACTTTATG	3L:18,414,106..18,414,187
TGCGCTAGTTCTCACCAACG ACCAACTTAAGCACCAACTAAGG	3L:18,439,417..18,439,489
ACGGGTGCCCTTAATGTTTAC GGTCGTTGCCCATGTCTTTG	3L:18,464,296..18,464,376
CAACCCTATCCATCCATCCATG CAATCGGCCTAATTCACCCATG	3L:18,491,978..18,492,057
ACATATTCGCCGACCAAGTG ACACTAACACGTGCCCCTAAC	3L:18,520,543..18,520,680

Cytological location	86D9
Sequence	**Genomic coordinates**
TGGCGCCGCTTTCTTATTAG AGAACAGGTTTGTGCGCTTG	3R:11,261,333..11,261,450

1553

1554 ***Table 1.*** Underreplicated domains and suppression of underreplication in SUMM4 subunit mutant alleles.

1555 Domains of underreplication (UR) in euchromatic arms of polytene chromosomes were called in w^{1118} as

1556 described in *Methods*. Their genomic coordinates, approximate cytological location ("Cyto band") and

1557 average DNA copy numbers ("<CN>") in homozygous w^{1118}, $SuUR^{ES}$ and $mod(mdg4)^{m9}$ L3 larvae are shown.

1558 <CN> numbers were normalized to the average DNA copy numbers across euchromatic genome.

1559 Underreplication percent recovery levels were calculated as $(\langle CN\rangle_{mut} - \langle CN\rangle_{w1118}) / (1 - \langle CN\rangle_{w1118})$;

1560 negative numbers indicate increased underreplication. Underreplication *p*-values were calculated using the

1561 DESeq2 package by averaging the Wald test *p*-values of each 5-kbp bin significantly different than

1562 the w^{1118} signal. Underreplication was called as suppressible by a mutant if $p < 0.01$; regions that do not

1563 exhibit a statistically significant recovery of underreplication are marked in red. Averages of <CN> across all

1564 called underreplicated domains and averages of percent Recovery across all suppressible underreplicated

1565 domains ("<Recovery>", bottom row) were adjusted for each underreplicated domain length; calculation

1566 errors = standard deviations.

N	chromosome coordinates				Length	UR, w^{1118}	UR, $SuUR^{ES}$			UR, $mod(mdg4)^{m9}$		
	arm	left	right	Cyto band		<CN>	<CN>	Recovery	*p*-value	<CN>	Recovery	*p*-value
1	X	2,950,001	3,140,000	3C3-C7	190,000	0.51	0.93	86%	**7.3E-05**	0.58	14%	1.1E-02
2	X	4,710,001	4,900,000	4C15-D5	190,000	0.56	0.96	92%	**3.9E-04**	0.81	57%	**6.9E-05**
3	X	4,965,001	5,070,000	4E1-E2	105,000	0.72	0.86	50%	**5.6E-04**	0.80	28%	1.4E-02
4	X	6,415,001	6,525,000	6A1-B1	110,000	0.71	0.90	65%	**1.4E-03**	0.80	29%	**7.3E-03**
5	X	7,335,001	7,560,000	7B1-B4	225,000	0.65	0.98	95%	**1.2E-03**	0.79	40%	**2.8E-03**
6	X	7,750,001	7,865,000	7B7-C1	115,000	0.64	0.94	84%	**3.0E-09**	0.84	55%	**5.2E-07**
7	X	8,880,001	9,005,000	8B5-C2	125,000	0.73	0.86	50%	**5.5E-03**	0.76	9%	**4.6E-03**
8	X	9,405,001	9,555,000	8D12-E7	150,000	0.72	0.91	67%	**3.6E-04**	0.85	47%	**3.6E-03**
9	X	11,170,001	11,325,000	10A10-B3	155,000	0.67	0.84	53%	**3.2E-03**	0.78	35%	**2.6E-03**
10	X	12,040,001	12,430,000	11A2-A10	390,000	0.38	0.97	94%	**1.4E-08**	0.42	6%	**6.8E-03**
11	X	13,950,001	14,100,000	12D1-E1	150,000	0.69	0.72	10%	1.0E-02	0.73	14%	1.4E-02
12	X	14,290,001	14,565,000	12E7-F1	275,000	0.51	0.94	87%	**4.1E-04**	0.69	36%	**8.1E-04**
13	X	17,925,001	18,030,000	16F3-F5	105,000	0.67	0.99	98%	**1.7E-15**	0.90	68%	**3.4E-05**
14	X	20,000,001	20,105,000	19A4-B1	105,000	0.79	1.12	157%	**1.4E-13**	0.82	12%	**6.1E-03**
15	X	20,525,001	21,020,000	19D2-E7	495,000	0.50	0.97	93%	**1.3E-07**	0.51	2%	**4.9E-03**
16	X	21,630,001	22,450,000	20A5-C1	820,000	0.04	0.32	29%	**1.8E-03**	0.06	2%	**6.4E-03**
17	X	22,550,001	22,995,000	20C2-F3	445,000	0.48	0.81	64%	**7.8E-05**	0.74	51%	**3.5E-04**
18	2L	3,920,001	4,025,000	24D1-D4	105,000	0.63	0.93	81%	**7.9E-07**	0.80	46%	**5.9E-05**
19	2L	4,585,001	4,790,000	25A2-A5	205,000	0.66	0.99	98%	**1.9E-08**	0.78	36%	**1.3E-03**
20	2L	5,400,001	5,510,000	25E1-E4	110,000	0.82	0.99	95%	**4.0E-08**	0.90	45%	**8.3E-03**
21	2L	6,155,001	6,320,000	26B9-C2	165,000	0.74	1.08	130%	**7.3E-14**	0.88	54%	**4.7E-04**
22	2L	9,030,001	9,150,000	29F8-30A2	120,000	0.76	0.98	93%	**1.5E-04**	0.95	79%	**3.3E-03**
23	2L	11,535,001	11,795,000	32F2-33A1	260,000	0.44	0.90	83%	**2.9E-04**	0.57	24%	**1.5E-03**
24	2L	12,215,001	12,340,000	33D3-E1	125,000	0.58	0.86	66%	**3.6E-11**	0.75	40%	**1.1E-04**

25	2L	12,765,001	12,970,000	33F5-34A3	205,000	0.55	0.91	79%	**8.8E-04**	0.73	40%	**7.0E-05**
26	2L	14,685,001	15,010,000	35B4-B8	325,000	0.41	0.88	80%	**5.7E-04**	0.54	23%	**7.2E-04**
27	2L	15,295,001	15,735,000	35D1-D4	440,000	0.49	0.76	53%	**2.3E-05**	0.54	9%	**4.0E-03**
28	2L	15,770,001	15,900,000	35D4-D6	130,000	0.54	0.87	71%	**4.5E-08**	0.68	31%	**6.7E-04**
29	2L	15,925,001	16,240,000	35D6-F1	315,000	0.29	0.90	87%	**6.7E-07**	0.38	12%	**1.4E-05**
30	2L	16,925,001	17,375,000	36B4-C7	450,000	0.23	0.89	85%	**1.4E-04**	0.26	4%	**4.3E-03**
31	2L	17,515,001	18,100,000	36C10-E4	585,000	0.34	0.87	80%	**5.0E-06**	0.36	2%	**3.7E-03**
32	2L	18,160,001	18,300,000	36E6-F2	140,000	0.67	0.99	97%	**3.3E-06**	0.90	69%	**3.1E-06**
33	2L	20,110,001	20,290,000	38C1-C4	180,000	0.48	0.69	41%	**8.9E-04**	0.46	-5%	**1.8E-03**
34	2L	20,485,001	20,620,000	38C8-D1	135,000	0.77	0.98	93%	**1.0E-06**	0.99	97%	**2.1E-05**
35	2L	21,400,001	21,550,000	39D3-E2	150,000	0.10	0.15	5%	**3.2E-03**	0.14	3%	**4.4E-03**
36	2L	21,805,001	22,125,000	40A4-E4	320,000	0.53	0.94	87%	**6.9E-05**	0.54	1%	**9.5E-03**
37	2R	4,875,001	5,050,000	41C4-D1	175,000	0.35	0.86	78%	**2.3E-10**	0.34	-1%	**4.0E-03**
38	2R	5,410,001	5,535,000	41F1-F3	125,000	0.58	0.79	50%	**1.1E-03**	0.52	-13%	**2.2E-03**
39	2R	6,290,001	6,505,000	42A14-B1	215,000	0.13	0.50	42%	**9.3E-04**	0.14	1%	**2.7E-03**
40	2R	13,620,001	13,760,000	50B6-C3	140,000	0.63	0.95	88%	**4.1E-18**	0.78	41%	**1.3E-05**
41	2R	20,355,001	20,540,000	56F17-57A5	185,000	0.56	0.92	83%	**2.0E-06**	0.71	35%	**8.2E-04**
42	2R	21,830,001	21,945,000	58A2-A4	115,000	0.72	0.95	83%	**1.1E-05**	0.71	-3%	2.2E-02
43	2R	23,145,001	23,320,000	59D1-D6	175,000	0.62	1.04	110%	**1.3E-22**	0.67	13%	**7.7E-03**
44	3L	4,840,001	5,100,000	64C1-C5	260,000	0.38	0.92	87%	**3.5E-08**	0.40	3%	**6.6E-03**
45	3L	5,385,001	5,510,000	64C15-D3	125,000	0.51	0.88	76%	**1.9E-22**	0.73	45%	**6.0E-09**
46	3L	6,290,001	6,485,000	65A11-B3	195,000	0.52	0.89	77%	**4.9E-05**	0.71	38%	**1.2E-04**
47	3L	9,180,001	9,300,000	67A1-A7	120,000	0.67	0.97	90%	**6.5E-09**	0.73	20%	1.0E-02
48	3L	10,000,001	10,195,000	67D3-D10	195,000	0.62	0.97	93%	**4.4E-13**	0.79	44%	**5.7E-06**
49	3L	13,085,001	13,220,000	70A1-A2	135,000	0.66	1.01	104%	**3.6E-09**	0.89	66%	**2.9E-06**
50	3L	13,550,001	13,855,000	70B6-C4	305,000	0.26	0.95	94%	**1.8E-06**	0.39	18%	**7.3E-04**
51	3L	15,175,001	15,500,000	71B7-D3	325,000	0.39	0.94	89%	**5.6E-04**	0.46	10%	**3.7E-03**
52	3L	17,115,001	17,240,000	73F1-74A1	125,000	0.71	1.02	106%	**4.3E-05**	0.84	45%	**2.7E-03**
53	3L	18,175,001	18,525,000	75B11-75D2	350,000	0.45	0.87	76%	**6.8E-05**	0.47	4%	**4.6E-03**
54	3L	20,555,001	20,695,000	77D1-77E3	140,000	0.60	1.02	106%	**2.2E-22**	0.84	61%	**3.6E-11**
55	3R	6,060,001	6,310,000	83D2-E4	250,000	0.70	0.92	72%	**7.6E-04**	0.63	-22%	1.0E-02
56	3R	6,495,001	6,635,000	83F1-84A1	140,000	0.53	0.96	91%	**7.8E-08**	0.71	39%	**2.2E-04**
57	3R	6,915,001	7,055,000	84B1-B2	140,000	0.64	0.93	80%	**3.9E-04**	0.82	49%	**1.9E-05**
58	3R	7,550,001	7,785,000	84D9-84E2	235,000	0.44	0.80	65%	**8.0E-06**	0.51	12%	**4.2E-03**
59	3R	10,450,001	10,660,000	86B6-C4	210,000	0.55	0.98	97%	**8.1E-11**	0.66	25%	**7.6E-04**
60	3R	10,910,001	11,140,000	88C15-86D4	230,000	0.45	0.94	89%	**2.3E-10**	0.46	2%	**2.3E-03**
61	3R	12,050,001	12,165,000	87A5-B1	115,000	0.63	0.96	88%	**9.9E-24**	0.81	49%	**5.9E-09**
62	3R	12,745,001	12,935,000	87C8-D4	190,000	0.67	0.89	68%	**7.5E-05**	0.60	-21%	1.1E-02
63	3R	14,935,001	15,055,000	88D8-D10	120,000	0.70	0.88	61%	**7.6E-06**	0.84	47%	**1.0E-04**
64	3R	16,670,001	16,970,000	89D6-E5	300,000	0.40	0.92	87%	**2.7E-09**	0.47	10%	**3.2E-03**
65	3R	17,160,001	17,355,000	89F1-90A2	195,000	0.62	0.94	84%	**1.0E-03**	0.86	64%	**2.8E-04**
66	3R	20,085,001	20,290,000	92C4-E1	205,000	0.61	0.81	53%	**1.5E-03**	0.71	26%	**3.6E-03**
67	3R	20,340,001	20,525,000	92E4-E12	185,000	0.58	0.96	91%	**5.0E-05**	0.79	50%	**7.2E-04**
68	3R	22,110,001	22,295,000	94A2-A4	185,000	0.61	0.93	83%	**3.4E-11**	0.76	39%	**3.0E-04**
69	3R	28,005,001	28,295,000	98B7-C3	290,000	0.40	0.91	85%	**2.5E-05**	0.60	32%	**6.9E-04**
70	3R	28,370,001	28,480,000	98C5-D2	110,000	0.73	0.98	94%	**1.2E-09**	0.91	66%	**4.3E-07**
UR domains: **70** <Length>: 216 ± 64 kbp Average <CN> across all UR domains: 0.49 ± 0.08							Suppressed UR domains: **69** <Length>: 217 ± 64 kbp <Recovery>: 78 ± 11%			Suppressed UR domains: **60** <Length>: 225 ± 67 kbp <Recovery>: 26 ± 9%		

1567

References

Adryan, B., Woerfel, G., Birch-Machin, I., Gao, S., Quick, M., Meadows, L., . . . White, R. (2007). Genomic mapping of Suppressor of Hairy-wing binding sites in Drosophila. *Genome Biol, 8*(8), R167. doi:10.1186/gb-2007-8-8-r167

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. . Retrieved from http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Andreyeva, E. N., Bernardo, T. J., Kolesnikova, T. D., Lu, X., Yarinich, L. A., Bartholdy, B. A., . . . Fyodorov, D. V. (2017). Regulatory functions and chromatin loading dynamics of linker histone H1 during endoreplication in Drosophila. *Genes Dev, 31*(6), 603-616. doi:10.1101/gad.295717.116

Arner, E. S., Sarioglu, H., Lottspeich, F., Holmgren, A., & Bock, A. (1999). High-level expression in Escherichia coli of selenocysteine-containing rat thioredoxin reductase utilizing gene fusions with engineered bacterial-type SECIS elements and co-expression with the selA, selB and selC genes. *J Mol Biol, 292*(5), 1003-1016. doi:10.1006/jmbi.1999.3085

Bag, I., Dale, R. K., Palmer, C., & Lei, E. P. (2019). The zinc-finger protein CLAMP promotes gypsy chromatin insulator function in Drosophila. *J Cell Sci, 132*(5). doi:10.1242/jcs.226092

Bellen, H. J., Levis, R. W., Liao, G., He, Y., Carlson, J. W., Tsang, G., . . . Spradling, A. C. (2004). The BDGP gene disruption project: single transposon insertions associated with 40% of *Drosophila* genes. *Genetics, 167*(2), 761-781.

Belyaeva, E. S., Boldyreva, L. V., Volkova, E. I., Nanayev, R. A., Alekseyenko, A. A., & Zhimulev, I. F. (2003). Effect of the Suppressor of Underreplication (SuUR) gene on position-effect variegation silencing in Drosophila melanogaster. *Genetics, 165*(3), 1209-1220. doi:10.1093/genetics/165.3.1209

Belyaeva, E. S., Zhimulev, I. F., Volkova, E. I., Alekseyenko, A. A., Moshkin, Y. M., & Koryakov, D. E. (1998). 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*(13), 7532-7537. doi:10.1073/pnas.95.13.7532

BroadInstitute. "Picard Tools." Broad Institute, GitHub repository. . version 2.2.4. Retrieved from http://broadinstitute.github.io/picard/

Buchner, K., Roth, P., Schotta, G., Krauss, V., Saumweber, H., Reuter, G., & Dorn, R. (2000). Genetic and molecular complexity of the position effect variegation modifier mod(mdg4) in Drosophila. *Genetics, 155*(1), 141-157. doi:10.1093/genetics/155.1.141

Bushnell, B. (2014). BBTools software package. Retrieved from http://bbtools.jgi.doe.gov

Cai, H., & Levine, M. (1995). Modulation of enhancer-promoter interactions by insulators in the Drosophila embryo. *Nature, 376*(6540), 533-536. doi:10.1038/376533a0

Cheng, Q., & Arner, E. S. (2017). Selenocysteine Insertion at a Predefined UAG Codon in a Release Factor 1 (RF1)-depleted Escherichia coli Host Strain Bypasses Species Barriers in Recombinant Selenoprotein Translation. *J Biol Chem, 292*(13), 5476-5487. doi:10.1074/jbc.M117.776310

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., . . . Li, H. (2021). Twelve years of SAMtools and BCFtools. *Gigascience, 10*(2). doi:10.1093/gigascience/giab008

Davidson, I. F., Bauer, B., Goetz, D., Tang, W., Wutz, G., & Peters, J. M. (2019). DNA loop extrusion by human cohesin. *Science, 366*(6471), 1338-1345. doi:10.1126/science.aaz3418

Dimitrova, D. S., & Gilbert, D. M. (1999). The spatial position and replication timing of chromosomal domains are both established in early G1 phase. *Mol Cell, 4*(6), 983-993. doi:10.1016/s1097-2765(00)80227-0

dos Santos, G., Schroeder, A. J., Goodman, J. L., Strelets, V. B., Crosby, M. A., Thurmond, J., . . . FlyBase, Consortium. (2015). FlyBase: introduction of the Drosophila melanogaster Release 6 reference genome assembly and large-scale migration of genome annotations. *Nucleic Acids Res, 43*(Database issue), D690-697. doi:10.1093/nar/gku1099

Emelyanov, A. V., Konev, A. Y., Vershilova, E., & Fyodorov, D. V. (2010). Protein complex of Drosophila ATRX/XNP and HP1a is required for the formation of pericentric beta-heterochromatin in vivo. *J Biol Chem, 285*(20), 15027-15037. doi:M109.064790 [pii] 10.1074/jbc.M109.064790

Emelyanov, A. V., Rabbani, J., Mehta, M., Vershilova, E., Keogh, M. C., & Fyodorov, D. V. (2014). Drosophila TAP/p32 is a core histone chaperone that cooperates with NAP-1, NLP, and nucleophosmin in sperm chromatin remodeling during fertilization. *Genes Dev, 28*(18), 2027-2040. doi:10.1101/gad.248583.114

Emelyanov, A. V., Vershilova, E., Ignatyeva, M. A., Pokrovsky, D. K., Lu, X., Konev, A. Y., & Fyodorov, D. V. (2012). Identification and characterization of ToRC, a novel ISWI-containing ATP-dependent chromatin assembly complex. *Genes Dev, 26*(6), 603-614. doi:10.1101/gad.180604.111

Filion, G. J., van Bemmel, J. G., Braunschweig, U., Talhout, W., Kind, J., Ward, L. D., . . . van Steensel, B. (2010). Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. *Cell, 143*(2), 212-224. doi:10.1016/j.cell.2010.09.009

Fyodorov, D. V., & Kadonaga, J. T. (2003). Chromatin assembly in vitro with purified recombinant ACF and NAP-1. *Methods Enzymol, 371*, 499-515. doi:10.1016/S0076-6879(03)71037-4

Fyodorov, D. V., & Levenstein, M. E. (2002). Chromatin assembly in *Drosophila* systems. In *Current Protocols in Molecular Biology* (pp. 21.27.21-21.27.27). New York: Wiley & Sons.

Gaszner, M., & Felsenfeld, G. (2006). Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat Rev Genet, 7*(9), 703-713. doi:10.1038/nrg1925

Gause, M., Morcillo, P., & Dorsett, D. (2001). Insulation of enhancer-promoter communication by a gypsy transposon insert in the Drosophila cut gene: cooperation between suppressor of hairy-wing and modifier of mdg4 proteins. *Mol Cell Biol, 21*(14), 4807-4817. doi:10.1128/MCB.21.14.4807-4817.2001

Georgiev, P. G., & Gerasimova, T. I. (1989). Novel genes influencing the expression of the yellow locus and mdg4 (gypsy) in Drosophila melanogaster. *Mol Gen Genet, 220*(1), 121-126. doi:10.1007/BF00260865

Gerasimova, T. I., Gdula, D. A., Gerasimov, D. V., Simonova, O., & Corces, V. G. (1995). A Drosophila protein that imparts directionality on a chromatin insulator is an enhancer of position-effect variegation. *Cell, 82*(4), 587-597. doi:10.1016/0092-8674(95)90031-4

Havugimana, P. C., Hart, G. T., Nepusz, T., Yang, H., Turinsky, A. L., Li, Z., . . . Emili, A. (2012). A census of human soluble protein complexes. *Cell, 150*(5), 1068-1081. doi:10.1016/j.cell.2012.08.011

Hoskins, R. A., Carlson, J. W., Wan, K. H., Park, S., Mendez, I., Galle, S. E., . . . Celniker, S. E. (2015). The Release 6 reference sequence of the Drosophila melanogaster genome. *Genome Res, 25*(3), 445-458. doi:10.1101/gr.185579.114

Ito, T., Levenstein, M. E., Fyodorov, D. V., Kutach, A. K., Kobayashi, R., & Kadonaga, J. T. (1999). ACF consists of two subunits, Acf1 and ISWI, that function cooperatively in the ATP-dependent catalysis of chromatin assembly. *Genes Dev, 13*(12), 1529-1539.

Jaskelioff, M., Van Komen, S., Krebs, J. E., Sung, P., & Peterson, C. L. (2003). Rad54p is a chromatin remodeling enzyme required for heteroduplex DNA joint formation with chromatin. *Journal of Biological Chemistry, 278*(11), 9212-9218.

Kamakaka, R. T., Tyree, C. M., & Kadonaga, J. T. (1991). Accurate and efficient RNA polymerase II transcription with a soluble nuclear fraction derived from *Drosophila* embryos. *Proceedings of the National Academy of Sciences of the United States of America, 88*(3), 1024-1028.

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., & Haussler, D. (2002). The human genome browser at UCSC. *Genome Res, 12*(6), 996-1006. doi:10.1101/gr.229102

Khoroshko, V. A., Levitsky, V. G., Zykova, T. Y., Antonenko, O. V., Belyaeva, E. S., & Zhimulev, I. F. (2016). Chromatin Heterogeneity and Distribution of Regulatory Elements in the Late-Replicating

Intercalary Heterochromatin Domains of Drosophila melanogaster Chromosomes. *PLoS One, 11*(6), e0157147. doi:10.1371/journal.pone.0157147

Kolesnikova, T. D., Posukh, O. V., Andreyeva, E. N., Bebyakina, D. S., Ivankin, A. V., & Zhimulev, I. F. (2013). Drosophila SUUR protein associates with PCNA and binds chromatin in a cell cycle-dependent manner. *Chromosoma, 122*(1-2), 55-66. doi:10.1007/s00412-012-0390-9

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods, 9*(4), 357-359. doi:10.1038/nmeth.1923

Lefevre, G. (1976). A photographic representation and interpretation of the polytene chromosomes of Drosophila melanogaster salivary glands. In Ashburner & Novitski (Ed.), *The Genetics and Biology of Drosophila.* (Vol. Ia, pp. 31-66). London: Academic Press.

Lei, E. P., & Corces, V. G. (2006). RNA interference machinery influences the nuclear organization of a chromatin insulator. *Nature Genetics, 38*(8), 936-941.

Lu, X., Wontakal, S. N., Kavi, H., Kim, B. J., Guzzardo, P. M., Emelyanov, A. V., . . . Skoultchi, A. I. (2013). Drosophila H1 regulates the genetic activity of heterochromatin by recruitment of Su(var)3-9. *Science, 340*(6128), 78-81. doi:10.1126/science.1234654

Makunin, I. V., Volkova, E. I., Belyaeva, E. S., Nabirochkina, E. N., Pirrotta, V., & Zhimulev, I. F. (2002). The Drosophila suppressor of underreplication protein binds to late-replicating regions of polytene chromosomes. *Genetics, 160*(3), 1023-1034. doi:10.1093/genetics/160.3.1023

Marchal, C., Sima, J., & Gilbert, D. M. (2019). Control of DNA replication timing in the 3D genome. *Nat Rev Mol Cell Biol, 20*(12), 721-737. doi:10.1038/s41580-019-0162-y

Matthews, N. E., & White, R. (2019). Chromatin Architecture in the Fly: Living without CTCF/Cohesin Loop Extrusion?: Alternating Chromatin States Provide a Basis for Domain Architecture in Drosophila. *Bioessays, 41*(9), e1900048. doi:10.1002/bies.201900048

Munden, A., Rong, Z., Sun, A., Gangula, R., Mallal, S., & Nordman, J. T. (2018). Rif1 inhibits replication fork progression and controls DNA copy number in Drosophila. *Elife, 7*. doi:10.7554/eLife.39140

Negre, N., Brown, C. D., Shah, P. K., Kheradpour, P., Morrison, C. A., Henikoff, J. G., . . . White, K. P. (2010). A comprehensive map of insulator elements for the Drosophila genome. *PLoS Genet, 6*(1), e1000814. doi:10.1371/journal.pgen.1000814

Nordman, J. T., Kozhevnikova, E. N., Verrijzer, C. P., Pindyurin, A. V., Andreyeva, E. N., Shloma, V. V., . . . Orr-Weaver, T. L. (2014). DNA copy-number control through inhibition of replication fork progression. *Cell Rep, 9*(3), 841-849. doi:10.1016/j.celrep.2014.10.005

Nordman, J. T., & Orr-Weaver, T. L. (2015). Understanding replication fork progression, stability, and chromosome fragility by exploiting the Suppressor of Underreplication protein. *Bioessays, 37*(8), 856-861. doi:10.1002/bies.201500021

Peterson, S. C., Samuelson, K. B., & Hanlon, S. L. (2021). Multi-Scale Organization of the Drosophila melanogaster Genome. *Genes (Basel), 12*(6). doi:10.3390/genes12060817

Pindyurin, A. V., Boldyreva, L. V., Shloma, V. V., Kolesnikova, T. D., Pokholkova, G. V., Andreyeva, E. N., . . . Zhimulev, I. F. (2008). Interaction between the Drosophila heterochromatin proteins SUUR and HP1. *J Cell Sci, 121*(Pt 10), 1693-1703. doi:10.1242/jcs.018655

Posukh, O. V., Maksimov, D. A., Skvortsova, K. N., Koryakov, D. E., & Belyakin, S. N. (2015). The effects of SUUR protein suggest its role in repressive chromatin renewal during replication in Drosophila. *Nucleus, 6*(4), 249-253. doi:10.1080/19491034.2015.1074366

Ramirez, F., Ryan, D. P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., . . . Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res, 44*(W1), W160-165. doi:10.1093/nar/gkw257

Rhind, N., & Gilbert, D. M. (2013). DNA replication timing. *Cold Spring Harb Perspect Biol, 5*(8), a010132. doi:10.1101/cshperspect.a010132

Roseman, R. R., Johnson, E. A., Rodesch, C. K., Bjerke, M., Nagoshi, R. N., & Geyer, P. K. (1995). A P element containing suppressor of hairy-wing binding regions has novel properties for mutagenesis in Drosophila melanogaster. *Genetics, 141*(3), 1061-1074. doi:10.1093/genetics/141.3.1061

Roseman, R. R., Pirrotta, V., & Geyer, P. K. (1993). The su(Hw) protein insulates expression of the Drosophila melanogaster white gene from chromosomal position-effects. *EMBO J, 12*(2), 435-442. doi:10.1002/j.1460-2075.1993.tb05675.x

Rowley, M. J., Nichols, M. H., Lyu, X., Ando-Kuri, M., Rivera, I. S. M., Hermetz, K., . . . Corces, V. G. (2017). Evolutionarily Conserved Principles Predict 3D Chromatin Organization. *Mol Cell, 67*(5), 837-852 e837. doi:10.1016/j.molcel.2017.07.022

Savitsky, M., Kim, M., Kravchuk, O., & Schwartz, Y. B. (2016). Distinct Roles of Chromatin Insulator Proteins in Control of the Drosophila Bithorax Complex. *Genetics, 202*(2), 601-617. doi:10.1534/genetics.115.179309

Schwartz, Y. B., Linder-Basso, D., Kharchenko, P. V., Tolstorukov, M. Y., Kim, M., Li, H. B., . . . Pirrotta, V. (2012). Nature and function of insulator protein binding sites in the Drosophila genome. *Genome Res, 22*(11), 2188-2198. doi:10.1101/gr.138156.112

Shatsky, M., Dong, M., Liu, H., Yang, L. L., Choi, M., Singer, M. E., . . . Biggin, M. D. (2016). Quantitative Tagless Copurification: A Method to Validate and Identify Protein-Protein Interactions. *Mol Cell Proteomics, 15*(6), 2186-2202. doi:10.1074/mcp.M115.057117

Sher, N., Bell, G. W., Li, S., Nordman, J., Eng, T., Eaton, M. L., . . . Orr-Weaver, T. L. (2012). Developmental control of gene copy number by repression of replication initiation and fork progression. *Genome Res, 22*(1), 64-75. doi:10.1101/gr.126003.111

Szabo, Q., Bantignies, F., & Cavalli, G. (2019). Principles of genome folding into topologically associating domains. *Sci Adv, 5*(4), eaaw1668. doi:10.1126/sciadv.aaw1668

Van Bortle, K., Nichols, M. H., Li, L., Ong, C. T., Takenaka, N., Qin, Z. S., & Corces, V. G. (2014). Insulator function and topological domain border strength scale with architectural protein occupancy. *Genome Biol, 15*(6), R82. doi:10.1186/gb-2014-15-5-r82

Wang, H., An, W., Cao, R., Xia, L., Erdjument-Bromage, H., Chatton, B., . . . Zhang, Y. (2003). mAM facilitates conversion by ESET of dimethyl to trimethyl lysine 9 of histone H3 to cause transcriptional repression. *Mol Cell, 12*(2), 475-487. doi:10.1016/j.molcel.2003.08.007

Yarosh, W., & Spradling, A. C. (2014). Incomplete replication generates somatic DNA alterations within Drosophila polytene salivary gland cells. *Genes Dev, 28*(16), 1840-1855. doi:10.1101/gad.245811.114

Zhang, Y., Bilbao, A., Bruderer, T., Luban, J., Strambio-De-Castillia, C., Lisacek, F., . . . Varesio, E. (2015). The Use of Variable Q1 Isolation Windows Improves Selectivity in LC-SWATH-MS Acquisition. *J Proteome Res, 14*(10), 4359-4371. doi:10.1021/acs.jproteome.5b00543

Zhimulev, I. F., Belyaeva, E. S., Semeshin, V. F., Koryakov, D. E., Demakov, S. A., Demakova, O. V., . . . Andreyeva, E. N. (2004). Polytene chromosomes: 70 years of genetic research. *Int Rev Cytol, 241*, 203-275. doi:10.1016/S0074-7696(04)41004-3

Zhu, X., Chen, Y., & Subramanian, R. (2014). Comparison of information-dependent acquisition, SWATH, and MS(All) techniques in metabolite identification study employing ultrahigh-performance liquid chromatography-quadrupole time-of-flight mass spectrometry. *Anal Chem, 86*(2), 1202-1209. doi:10.1021/ac403385y

Zielke, N., Edgar, B. A., & DePamphilis, M. L. (2013). Endoreplication. *Cold Spring Harb Perspect Biol, 5*(1), a012948. doi:10.1101/cshperspect.a012948