bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

SUMM4 complex couples insulator function and DNA replication timing control

Evgeniya N. Andreyeva,[1]† Alexander V. Emelyanov,[1]† Markus Nevil,[2] Lu Sun,[3] Elena Vershilova,[1] Christina A. Hill,[4] Michael C. Keogh,[3] Robert J. Duronio,[4,5,6,7] Arthur I. Skoultchi,[1]* & Dmitry V. Fyodorov[1]*

† equally contributing authors

* corresponding authors

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

The asynchronous timing of replication of different chromosome domains is essential for eukaryotic genome stability, but the mechanisms establishing replication timing programs remain incompletely understood. *Drosophila* SNF2-related factor SUUR imparts under-replication (UR) of late-replicating intercalary heterochromatin (IH) domains in polytene chromosomes[1]. SUUR negatively regulates DNA replication fork progression across IH; however, its mechanism of action remains obscure[2,3]. 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[4,5]. *In vitro*, Mod(Mdg4) stimulates the ATPase activity of SUUR, although neither SUUR nor SUMM4 can remodel nucleosomes. Mod(Mdg4)-67.2 and SUUR distribution patterns *in vivo* partially overlap, and Mod(Mdg4) is required for a normal spatiotemporal distribution of SUUR in chromosomes. SUUR and Mod(Mdg4)-67.2 mediate insulator activities of the *gypsy* mobile element that disrupt enhancer-promoter interactions and establish euchromatin-heterochromatin barriers in the genome. Furthermore, mutations of *SuUR* or *mod(mdg4)* reverse the locus-specific UR. These findings reveal that DNA replication can be delayed by a chromatin barrier and thus, uncover a critical role for architectural proteins in replication timing control. They also provide a biochemical link between ATP-dependent motor factors and the activity of insulators in regulation of gene expression and chromatin partitioning.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Main

Replication of metazoan genomes occurs according to a highly coordinated spatiotemporal program, where discrete chromosomal regions replicate at distinct times during S phase of the cell cycle[6]. In general, the replication program follows the spatial organization of the genome in Megabase-long constant timing regions interspersed by timing transition regions[7]. The spatiotemporal replication program was noted to exhibit correlations with the genetic activity, epigenetic marks and features of 3D genome architecture. Yet the mechanisms for these correlations remain obscure. It has been speculated that the structural features of DNA, epigenetic regulation of chromatin and sub-nuclear localization of chromosome domains can coordinately impart the asynchronous, locus-specific replication of the genome[7]. Interestingly, the timing of firing for any individual origin of replication is established during G1 before pre-replicative complexes (pre-RC) are assembled and licensed at origins[8].

Most larval tissues of *Drosophila melanogaster* grow via G-S endoreplication cycles that duplicate DNA without cell division resulting in polyploidy[9]. In some polyploid tissues like the salivary gland, all of the >1,000 duplicated DNA molecules align in register along their lengths to form giant polytene chromosomes[10]. Importantly, large genomic domains corresponding to the latest replicated regions of dividing cells, specifically pericentric (PH) and intercalary (IH) heterochromatin, occasionally fail to complete endoreplication resulting in under-replication (UR). As a rule, in both dividing and endoreplicating cells, these regions are devoid of discernable sites for binding the Origin of Replication Complex (ORC) and thus, their replication must rely on forks progressing from external distal origins[11]. Although cell cycle programs are drastically dissimilar between endoreplicating and mitotically dividing cells, they share biochemically identical DNA replication machinery[9]. Consequently, UR provides a facile

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

readout for late-S phase replication initiation and fork progression, which play a central role in establishing the DNA replication timing program.

The *Suppressor of UR* (*SuUR*) gene is essential for polytene chromosome UR in IH and PH[1]. It encodes a protein (SUUR) containing a helicase domain with a homology to that of the SNF2/SWI2 family. Its contribution to regulation of DNA replication remains enigmatic, although it has been shown that the occupancy of ORC is not increased in *SuUR* mutant UR regions[11]. Rather than inhibiting the pre-RC assembly or origin firing, SUUR negatively regulates the rate of replication[2] by an unknown mechanism. It has been proposed[3] that the retardation of replisome by SUUR takes place via simultaneous physical association with the components of the fork (*e.g.*, CDC45 and PCNA)[12] and repressive chromatin proteins, such as HP1a[13]. The deceleration of SUUR-bound replication forks was also invoked as an explanation for the apparent role of SUUR in the establishment of epigenetic marking of IH[3].

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

To determine how SUUR functions in replication control we sought to identify the native complex in which it acts. Previous attempts to characterize the native form of SUUR by co-IP or tag-affinity purification of ectopically expressed protein gave rise to multiple putative binding partners[2,12-14]. However, evaluating whether any of these proteins are present in a native SUUR complex is problematic because of the low abundance of SUUR, which is difficult to detect by immunoblotting of native material. It also precludes its purification by conventional chromatography. Therefore, we developed a novel biochemical approach that relies on partial purification by multi-step FPLC and shotgun proteomics of chromatographic fractions by

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

quantitative LCMS (**Fig. 1a**). We term this technology MERCI for MS-enabled Rapid protein complex Identification (see *Methods*).

The depth of proteomic quantification is limited by the range of peptides identified in the information-dependent acquisition (IDA), dubbed "ion library" (IL). Unfortunately, SUUR-specific peptides could not be found in ion libraries obtained from acquisitions of either crude nuclear extracts from embryos or the collection of fractions from the first, phosphocellulose step (IL1, **Extended Data Fig. E1a**, **Supplementary Table S1**). Thus, to quantify SUUR in phosphocellulose fractions, we augmented IL1 with the ion library obtained by acquisition of recombinant FLAG-SUUR (ILR, **Figs. 1a**, **b**) expressed in the baculovirus system. In ion libraries from all subsequent chromatographic steps, peptides derived from native SUUR were detected (**Extended Data Fig. E1a**, **Supplementary Table S1**) and used for quantification of the cognate data-independent acquisitions (DIA/SWATH) (**Fig. 1c**). Quantification of SWATH data measures only the relative (to the reference protein) amounts of SUUR across chromatographic fractions. To measure absolute levels of SUUR, a semi-quantitative approach can be used by building a titration curve from SWATH acquisitions of known amounts of recombinant SUUR (**Extended Data Figs. E1b**, **c**). In this manner, we estimated the amount of SUUR in the nuclear extract (~140 pg in 25 µg total protein, **Extended Data Fig. E1c**) and in individual fractions from all chromatographic steps (**Extended Data Fig. E1d**). Although in five FPLC steps we achieved >3,000-fold purification of SUUR, it remained only ~2% pure (**Extended Data Fig. E1e**). A progressive loss of material precludes further purification (300 ng of SUUR out of a total 16 µg protein). Thus, the putative SUUR complex cannot be purified to an apparent homogeneity from a substantial amount of starting material (~1 kg *Drosophila* embryos, equivalent to 2.5 g protein) by conventional chromatography.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The final aspect of the MERCI algorithm calls for re-quantification of all FPLC fraction SWATH acquisitions with an ion library from the last step (IL5), which is enriched for peptides derived from SUUR and co-purifying polypeptides (**Fig. 1b**) and includes only 140 proteins (**Extended Data Fig. E1a**, **Supplementary Table S1**). In this fashion, under-represented polypeptides (including SUUR and, potentially, subunits of a putative SUUR complex) that may not be IDA-detectable in earlier steps will not evade quantification. Purification profiles of proteins that could be quantified in all five FPLC steps (132) were then artificially stitched together into 83-point arrays of Z-scores (**Fig. 1d**, **Supplementary Table S2**). These profiles were analyzed for Pearson correlation with that of SUUR and ranked down from the highest Pearson coefficient, PCC (**Fig. 1e**). Whereas the PCC numbers for the bottom 130 proteins lay on a smooth curve, the top two proteins, SUUR itself (PCC = 1.000) and Mod(Mdg4) (PCC = 0.939) fell above the extrapolated (by polynomial regression) curve (**Fig. 1f**) and were thus distinct from the rest of proteins. Consistently, SUUR and Mod(Mdg4) exhibited nearly identical purification profiles in all five FPLC steps (**Fig. 1d**), unlike the next two top-scoring proteins, EGG (PCC = 0.881) and CG6700 (PCC = 0.874) (**Extended Data Figs. E1f**, **g**). Also, HP1a (PCC = 0.503), which had been previously hypothesized to form a stable complex with SUUR[13] did not co-purify with SUUR in any of the FPLC steps (**Extended Data Fig. E1h**).

Mod(Mdg4) is a BTB/POZ domain protein that functions as an adaptor for architectural proteins that promote various aspects of genome organization[4,5]. It is expressed as 26 distinct polypeptides generated by splicing *in trans* of a common 5'-end precursor RNA with 26 unique 3'-end precursors[15]. IL5 contained seven peptides derived from Mod(Mdg4) (99% confidence). Whereas four of them mapped to the common N-terminal 402 residues, three were specific to the C-terminus of a particular form, Mod(Mdg4)-67.2 (**Extended Data Figs. E1i**, **j**). No peptides

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

specific to other splice forms were detected. We raised an antibody to the C-terminus of Mod(Mdg4)-67.2, designated ModT antibody, and analyzed size exclusion column fractions by immunoblotting with SUUR and ModT antibodies. Consistent with our SWATH analyses (**Figs. 1c**, **d**) ~100-kDa SUUR and Mod(Mdg4)-67.2 polypeptides copurified as a complex with an apparent molecular mass of ~250 kDa (**Fig. 1g**). Finally, we confirmed that SUUR is specifically co-immunoprecipitated with Mod(Mdg4)-67.2 from crude extracts (**Fig. 1h**). As a control, XNP co-immunoprecipitated with HP1a as shown previously[16], but did not with SUUR or Mod(Mdg4). We conclude that *in vivo* SUUR and Mod(Mdg4) form a stable stoichiometric complex that we term SUMM4.

Biochemical activities of recombinant SUMM4 *in vitro*

We reconstituted recombinant SUMM4 complex by co-expressing FLAG-SUUR with Mod(Mdg4)-67.2-His_6 in Sf9 cells (**Fig. 2a**). As expected, they co-purified by FLAG affinity chromatography. Mod(Mdg4)-67.2 is the predominant splice form of Mod(Mdg4) expressed in embryos (*e.g.*, **Fig. 1h**). Thus, minor SUUR-Mod(Mdg4) complexes may have failed to be identified by IDA in IL5 (**Extended Data Fig. E1i**). We discovered that FLAG-SUUR did not co-purify with another splice form, Mod(Mdg4)-59.1 (**Extended Data Figs. E1k**, **Fig. 2a**). Therefore, the shared N-terminus of Mod(Mdg4) (1-402) is not sufficient for interactions with SUUR. However, this result does not exclude a possibility that SUUR may form complex(es) with some of the other, low-abundance 24 splice forms of Mod(Mdg4). The SUUR-Mod(Mdg4)-67.2 interaction is specific, as the second-best candidate from our correlation analyses (*Drosophila* SetDB1 ortholog EGG; **Fig. 1f**) did not form a complex with FLAG-SUUR

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(**Extended Data Fig. E2a**), although it associated with its known partner WDE, an ortholog of hATF7IP/mAM[17].

The N-terminus of SUUR contains a region homologous with SNF2-like DEAD/H helicase domains. Although SUUR requires its N-terminal domain to function *in vivo*[14], it has been hypothesized to be inactive as an ATPase[18]. We directly measured the ATPase activities of recombinant SUUR and SUMM4 (**Fig. 2a**) in comparison with that of recombinant *Drosophila* ISWI (**Fig. 2b**). Purified recombinant Mod(Mdg4)-67.2 and a variant SUUR protein with a point mutation in the putative Walker A motif (K59A) were used as negative controls (**Fig. 2a**, **Extended Data Fig. E2b**). Contrary to the prediction[18], both SUUR and SUMM4 efficiently hydrolyzed ATP. Notably, SUMM4 was 1.4- to 2-fold more active than SUUR alone, indicating that Mod(Mdg4)-67.2 stimulates SUUR enzymatic activity. We also examined whether SNF2 ATPase cofactors (DNA and nucleosomes) can stimulate the activity of SUUR. To this end, we reconstituted oligonucleosomes on plasmid DNA (**Extended Data Figs. E2bc-e**). Linker histone H1-containing chromatin was also used as a substrate/cofactor, because SUUR has been demonstrated to physically interact with H1[19]. In contrast to ISWI, SUUR was not stimulated by addition of DNA or nucleosomes and moderately (by about 70%) activated by H1-containing oligonucleosomes (**Fig. 2c**). SUMM4 was weakly activated by nucleosomes, whereas both ISWI and SUMM4 were inhibited by H1 assembled into chromatin.

We next examined the nucleosome remodeling ability of SUUR and SUMM4; specifically, their ability to expose a positioned DNA motif in the EpiDyne®-PicoGreen™ assay (*Methods* and **Extended Data Fig. E2f**). Centrally or terminally positioned mononucleosomes were efficiently mobilized by ISWI and human BRG1 in a concentration- and time-dependent manner (**Extended Data Figs. E2g-j**). In contrast, SUUR and SUMM4 did not reposition either

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

nucleosome (**Fig. 2d**). Thus, SUUR and SUMM4 do not possess a detectable remodeling activity. They resemble certain other SNF2-like enzymes (*e.g.*, RAD54) that instead utilize the energy of ATP hydrolysis to mediate alternate DNA translocation reactions[20].

The distribution of SUMM4 complex *in vivo*

We examined the localization patterns of SUMM4 subunits in polytene chromosomes by indirect immunofluorescence (IF) and discovered their strong overlap (**Fig. 3a**). In late endo-S phase when SUUR exhibited a characteristic pattern of distribution, enriched at IH and PH domains, it invariably co-localized with Mod(Mdg4)-67.2, except for the chromocenter that did not show appreciable occupancy by Mod(Mdg4)-67.2. In fact, Mod(Mdg4)-67.2 was present at classical regions of SUUR enrichment, such as UR domains in 75C and 89E (**Extended Data Fig. E3a**). In contrast, there were multiple sites of Mod(Mdg4)-67.2 localization that were free of SUUR. This finding suggests that there are additional native form(s) of Mod(Mdg4)-67.2, either as an individual polypeptide or in complex(es) other than SUMM4. When we fractionated *Drosophila* nuclear extract using a different progression of FPLC steps (**Extended Data Fig. E3b**), we found that Mod(Mdg4)-67.2 can form a megadalton-size complex that did not contain SUUR (**Extended Data Figs. E3c-e**). Therefore, a more intricate pattern of Mod(Mdg4)-67.2 distribution in polytene chromosomes likely reflects loading of both SUMM4 and the alternative complex.

We tested whether SUUR and Mod(Mdg4) loading into polytene chromosomes were mutually dependent using mutant alleles of *SuUR* and *mod(mdg4)*. $SuUR^{ES}$ is a null allele of *SuUR*[21]. $mod(mdg4)^{m9}$ is a null allele encompassing a 12,016-bp deficiency that removes gene regions of the shared 5'-end precursor of all Mod(Mdg4) splice variants in addition to eight

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

specific 3'-precursors[22]. *mod(mdg4)^{u1}* contains an insertion of a *Stalker* mobile element in the last coding exon of Mod(Mdg4)-67.2 3'-precursor and can only affect the expression of Mod(Mdg4)-67.2[5]. *SuUR^{ES}* and *mod(mdg4)^{u1}* are homozygous viable, and *mod(mdg4)^{m9}* is recessive adult pharate lethal. Consistently, we could not detect Mod(Mdg4)-67.2 expression in homozygous *mod(mdg4)^{m9}* L3 salivary glands, whereas *mod(mdg4)^{u1}* expressed a truncated polypeptide (*cf*, ~70 kDa and ~100 kDa, **Extended Data Fig. E3f**). The truncated 70-kDa polypeptide failed to load into polytene chromosomes (**Fig. 3b**, **Extended Data Fig. E3g**). Also, as shown previously, SUUR cannot be detected in *SuUR^{ES}* polytene 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, although diluted >1,000 fold during development and possibly degraded.

Importantly, the absence (or drastic decrease) of Mod(Mdg4)-67.2 also strongly reduced the loading of SUUR (**Fig. 3b**, **Extended Data Fig. 3g**). The normal distribution pattern of SUUR in polytene chromosomes is highly dynamic[12,19]. SUUR is initially loaded in chromosomes at the onset of endo-S phase and then re-distributes, while constantly remaining partially co-localized with the replication fork (PCNA), through very late endo-S when SUUR becomes accumulated in UR domains and PH. The deposition of SUUR and its co-localization with PCNA has been proposed to depend on direct physical interaction with components of the replisome[12]. In both *mod(mdg4)* mutants, we observed a striking absence of SUUR in polytene chromosomes during early endo-S, which indicates that the initial deposition is dependent on its interactions with Mod(Mdg4). Although SUUR deposition slightly recovered by late endo-S, it was still several fold weaker than that in wild type. Potentially, in the absence of Mod(Mdg4), SUUR may be

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

tethered to IH and PH loci by direct binding with linker histone H1 as shown previously[19], albeit with a lower efficiency. Finally, the gross subcellular distribution of SUUR also strongly correlated with that of Mod(Mdg4): a mis-localization of truncated Mod(Mdg4)-67.2 from nuclear to partially cytoplasmic was accompanied by a similar mis-localization of SUUR (**Fig. 3c**). This result indicates that the truncation of Mod(Mdg4) in *mod(mdg4)*^{u1} may likely have an antimorphic effect by mis-localization and deficient chromatin binding of interacting polypeptides, including SUUR and others (**Fig, 3c**, **Extended Data Figs. 3b-e**).

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 to chromatin by a physical association with zinc finger factor Suppressor of Hairy Wing, Su(Hw). Su(Hw) directly binds to consensus sequences that are present in *gypsy* transposable elements and are also widely distributed across the *Drosophila* genome in thousands of copies[23]. Mod(Mdg4)-67.2 was previously shown to be essential for the insulator activity of *gypsy*[5], which functions *in vivo* to disrupt enhancer-promoter interactions and establish a barrier to the propagation of chromatin forms[24,25]. We therefore tested whether SUMM4 contributes to *gypsy* insulator function. The *ct*^{6} allele of *Drosophila* contains a *gypsy* element inserted between the wing enhancer and promoter of the gene *cut* that inactivates *cut* expression and results in abnormal wing development (**Fig. 4a**). We discovered that both *mod(mdg4)*^{u1} and *SuUR*^{ES} mutations partially suppressed this phenotype (**Fig. 4a**). Thus, both subunits of SUMM4 are required to mediate the full enhancer-blocking activity of *gypsy*. Another *Drosophila in vivo* insulator assay makes use of a collection of *P{SUPor-P}* insertions that contain the *white* reporter gene flanked by 12 copies of *gypsy* Su(Hw)-binding sites. When *P{SUPor-P}* is inserted in heterochromatin, *white* is protected from

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

silencing resulting in red eyes[26]. Again, both *mod(mdg4)*^{u1} and *SuUR*^{ES} relieved the chromatin barrier function of Su(Hw) sites, causing repression of *white* (**Fig. 4b**). *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 PH[27]. In contrast, *SuUR* mutation enhances variegation in the context of insulated, PH-positioned *white*. We conclude that SUMM4 is an insulator complex that mediates the chromatin boundary function of *gypsy*.

Although a *SuUR* mutation decreased the activity of the *gypsy* insulator, the suppression was universally weaker than that by *mod(mdg4)*^{u1} (**Figs. 4a-b**). It is possible that SUUR is not absolutely required for the establishment of the insulator. For instance, the loss of SUMM4 may be compensated by the alternative complex of Mod(Mdg4)-67.2 (**Extended Data Figs. 3b-e**). Furthermore, the *mod(mdg4)*^{u1} allele is expected to be particularly strong due to its likely antimorphic function, since it can mis-localize interacting partner proteins, including SUUR itself (**Figs. 3b**, **c**).

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 UR in late-replicating IH domains of polytene chromosomes (**Fig. 4e**). It has been long known that 3D chromosome partitioning maps show an "uncanny alignment" with replication timing maps[6]. To examine the possible roles of SUMM4 in UR, we first analyzed DNA copy numbers in salivary glands of *w*^{1118} control, *SuUR* and *mod(mdg4)* mutant L3 larvae by qPCR at two known SUUR target loci (cytological regions 75C and 4D). As shown before, we observed an almost complete reversal of UR by *SuUR* mutation, whereas in *mod(mdg4)* mutants there was a moderate recovery of DNA copy number (**Fig. 4f**, **Extended Data Fig. 4a**).

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Furthermore, cytological evidence in 75C region supported the molecular analyses in that all mutants universally exhibited a brighter DAPI staining of the 75C1-2 band than that in w^{1118}, indicative of higher DNA content (**Fig. 4f**). We thus measured DNA copy number genome-wide by next generation sequencing (NGS). In w^{1118} salivary glands, the DNA copy profile revealed large (>100-kbp) domains of reduced ploidy (**Extended Data Fig. E4b**), similar to a previous report[19]. Excluding pericentric and sub-telomeric heterochromatin, we called 70 UR regions (**Extended Data Table E1**) in euchromatic arms, as described in *Methods*. In both *SuUR* and *mod(mdg4)* larvae, we observed suppression of UR in IH (**Fig. 4f**, **Extended Data Fig. E4b**, **Extended Data Table E1**). Consistent with its distribution *in vivo* (**Fig. 3a**), Mod(Mdg4)-67.2 was dispensable for UR in PH. The NGS data also strongly correlated with qPCR measurements of DNA copy numbers in 75C and 4D.

Uniformly, *SuUR* mutation gave rise to a stronger relief of UR than that produced by even the null mutation *mod(mdg4)*^{m9} (**Extended Data Table E1**). This result can be explained by considering embryonic deposition of functional Mod(Mdg4) proteins and RNA by heterozygous mothers (see above). Although third instar larvae are ~1,000-fold larger, volume-wise, than the embryos, persistent Mod(Mdg4)-67.2 can still be detected in these larvae by IF despite its dilution and degradation (**Extended Data Fig. E3g**). Unlike L3, first instar larvae are nearly identical in size to the embryos. Therefore, since the endoreplication cycles initiate in embryos, in *mod(mdg4)*^{m9} animals, the first few out of 10-11 rounds of chromosome polytenization take place with an almost normal amount of Mod(Mdg4) present, which may substantially limit the effect of *mod(mdg4)*^{m9} mutation on UR as measured in L3. Some UR loci exhibited a stronger recovery in the *mod(mdg4)*^{u1} mutant, which is homozygous viable and thus not subject to the maternal loading effect (**Fig. 4g**, **Extended Data Fig. E4d**). Interestingly, the UR of yet another

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

class of regions exhibited strong suppression by the null allele but was completely unresponsive to *mod(mdg4)*^{u1} (**Extended Data Fig. E4d**), suggesting a possibility that SUUR forms SUMM4-like complex(es) with Mod(Mdg4) splice forms other than Mod(Mdg4)-67.2, the only one affected by *mod(mdg4)*^{u1}. These additional complexes may be required for UR in IH domains that are suppressible by *mod(mdg4)*^{m9} but not *mod(mdg4)*^{u1}. Finally, an important frequent feature of the partially suppressed UR was its asymmetry (**Fig. 4f**, **Extended Data Figs. E4a-d**), which is consistent with a unidirectional penetration of the UR domain by a replication fork that fires from the nearest external origin (**Fig. 4e**).

We conclude that the SUMM4 complex is required for the establishment of UR in the IH domains of *Drosophila* polytene chromosome. SUMM4 likely causes UR by forming a barrier to replication fork progression. The barrier may be created as a direct physical obstacle to MCM2-7 DNA-unwinding helicase or other enzymatic activities of the replisome. Alternatively, SUMM4 may inhibit the replication machinery indirectly by assembling at the insulator a DNA/chromatin structure that is incompatible with replisome translocation. This putative inhibitory structure may involve epigenetic modifications of chromatin as proposed earlier[28] and may also be dependent on Rif1, a negative DNA replication regulator that acts downstream of SUUR[14] and/or linker histone H1 as shown previously[19].

Discussion

Our work demonstrates for the first time that insulator complexes assembled on chromatin can attenuate the progression of replication forks in salivary glands *in vivo*. Despite distinct cell cycle programs in dividing and endoreplicating cells, the biochemical composition of replication forks in both cell types is identical[9]. Therefore, similar insulator-driven control mechanisms for

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

DNA replication are likely conserved in mitotically dividing diploid cells. Our data thus implicate insulator/chromatin boundary elements as a critical component of DNA replication control. Our model suggests that delayed replication of repressed chromatin (*e.g.*, IH) during very late S phase can be imposed in a simple, two-stroke mechanism (**Fig. 4e**). (1) It requires that an extended genomic domain is completely devoid of functional origins of replication. The assembly and licensing of proximal pre-RC complexes can be repressed epigenetically or at the level of DNA sequence. And (2), this domain has to be separated from flanking chromatin by a barrier element associated with an insulator complex, such as SUMM4. This structural organization is capable of preventing or delaying the entry of external forks fired from distal origins. The current paradigm of replication timing largely focuses on the existence of "early" and "late" origins that are ordained during early S or G1 phase[8]. Our model offers an additional mechanism to establish a locus-specific late replication program. Insulator complexes and other genome architectural proteins are highly abundant in the genome and assembled in a sequence-specific manner throughout interphase. Even if all origins fire simultaneously (or stochastically, depending on availability of limiting factors) at the onset of S phase, any given DNA replication fork is forced to operate in a dense milieu of abundant, pre-positioned insulator complexes, continuously negotiating with them to get cleared for passage. When these barriers are eventually cleared, the outcome would be indistinguishable from that of an RC firing at "late origin" if analyzed cytologically (by incorporation of labeled nucleotides) or at the level of ChIP.

It is interesting to consider the possible role for SUMM4 in the functions of insulator. Insulator elements organize the genome into chromatin loops[5] that are involved in the formation of topologically associating domains, TADs[29-31] . In mammals, CTCF-dependent loop formation requires ATP-driven motor activity of SMC complex cohesin[32]. In contrast, CTCF and cohesin

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

are thought to be dispensable for chromatin 3D partitioning in *Drosophila*[33]. Instead, the larger, transcriptionally inactive domains (canonical TADs) are interspersed with smaller active compartmental domains, which themselves represent TAD boundaries[31]. 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[31]. On the other hand, *Drosophila* TAD boundaries are enriched for architectural proteins other than CTCF[34], and their roles have not been tested directly 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[35]. Moreover, SUUR may provide the DNA motor function to help establish chromosome contact domains (**Figs. 4c-e**). Given the noted high-affinity interactions between SUUR and linker histone H1[19], it is also possible that SUUR favors unidirectional translocation from its initial sites of deposition towards H1-enriched, repressed chromatin, which would promote a physical separation of active and inactive compartmental domains (**Figs 4d-e**). In fact, this hypothesis is supported by observations of an asymmetric, selective propagation of SUUR towards inside of IH regions (heavily enriched for H1[19]) but not outside (**Extended Data Fig. E4e**)[36].

Our model explains previous observations about other biological functions of SUUR. For instance, global epigenetic modifications observed in the *SuUR* mutant likely do not directly arise from de-repression of the replisome as suggested[3] 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* **Figs. 4d**, **e**). Also, the apparent co-localization of SUUR with PCNA throughout endo-S phase (**Extended Data Fig. E3g**) may be caused by a replication

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

fork retardation at insulator sites. Initially, SUUR is deposited in chromosomes as a subunit of SUMM4 complex at thousands of loci by tethering via MMD4-Su(Hw) interactions. As replication forks progress through the genome, they encounter insulator complexes where replication machinery pauses for various periods of time before resolving the obstacle. Thus, the increased co-residence time of PCNA and SUUR manifests cytologically as their partial co-localization. With the progression of endo-S phase, some of the SUMM4 insulator complexes are evicted and thus, the number of SUUR-positive loci is decreased, until eventually, the replication fork encounters nearly completely impenetrable insulators demarcating UR domain boundaries. This mechanism is especially plausible given that boundaries of IH loci very frequently encompass multiple, densely clustered Su(Hw) binding sites (*e.g.*, **Fig. 4f**, **Extended Data Fig. E4c**). We examined the data from genome-wide proteomic analyses for Su(Hw) and SUUR performed by DamID in Kc167 cells[36]. Strikingly, Su(Hw) DamID-measured occupancy does not exhibit a discrete pattern expected of a DNA-binding factor. Instead, it appears broadly dispersed, together with SUUR, up to tens of kbp away from mapped Su(Hw) binding sites (**Extended Data Fig. E4e**). This observation strongly corroborates the translocation model for SUMM4 mechanism of action. According to this model, upon tethering to DNA-bound Su(Hw), SUMM4 traverses the UR region, which helps to separate it in a contact domain. As DNA within the UR region is tracked by SUUR (**Figs. 4e**), it is brought into a transient close proximity with both SUMM4 and the associated Su(Hw) protein, which is detected by DamID (or ChIP) as an expanded occupancy pattern.

In a recent paper[37], Su(Hw) binding sites in S2 cells and *in vivo* in pupae were found to strongly correlate (within a 5-kbp window) with the occupancy of pre-RC components, ORC2/3, MCM2-7 and CDC45, biased towards origins in repressive “blue” and “black” chromatin. The

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

correlation has been interpreted as Su(Hw) creating a "landing platform" for the assembly of pre-RC. Interestingly, the strongest (by orders of magnitude) co-occupancy was observed for MCM2-7, a hexameric unwinding helicase that is a component of both the licensed pre-RC and active replisome. Incidentally, MCM2-7 would be the first replisome component to encounter any block to replication elongation. Moreover, the correlation between Su(Hw) and CDC45 does not fit the original conclusion, because CDC45 is specific to an activated RC, rather than pre-RC. Thus, an alternative interpretation of the co-occupancy of Su(Hw) with MCM2-7 and CDC45 is replisome slowing down at Su(Hw)-bound insulators, which increases their co-residence time. An apparent but much weaker correlation with ORC proteins may stem from juxtaposition of origins to a small subset of paused forks that failed to relocate >5 kbp after firing and prior to slowing down by an insulator.

Finally, 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 an apparent homogeneity. Since a multi-step FPLC scheme invariably leads to an exponential loss of material (**Extended Data Fig. 1e**), reducing the number of purification steps 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, which renders its 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*.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

In conclusion, we used a newly developed MERCI approach to identify a stable stoichiometric complex termed SUMM4 that comprises SUUR, a previously known effector of replication control, and Mod(Mdg4), an insulator protein. SUMM4 subunits cooperate to mediate transcriptional repression and chromatin boundary functions of *gypsy*-like (class 3) insulators and regulate DNA replication by slowing down replication fork progression through the boundary element. Thus, SUMM4 is required for coordinate regulation of gene expression, chromatin partitioning and DNA replication timing. The insulator-dependent regulation of DNA replication offers a novel mechanism for the establishment of replication timing in addition to the currently accepted paradigm of variable timing of replication origin firing.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Recombinant proteins

Sf9 cells

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

WDE-PA(2..1420)-*. The sequences of FLAG and V5 tags are highlighted in bold typeface. Whereas, typically, amplified virus stocks had titers above $5 \cdot 10^9$ pfu/ml, FLAG-SUUR viruses reached no more than $2\text{-}4 \cdot 10^8$ pfu/ml, presumably, due to the inhibitory effect of over-expressed protein on viral DNA replication.

Cells, $2 \cdot 10^6$/ml in Sf-900 II SFM medium (Gibco), were infected at multiplicity of infection (MOI) of ~10 in PETG shaker flasks (Celltreat, Inc.). After infection for 48-72 hours at 27°C, cells were harvested, and recombinant proteins were purified by FLAG or Ni-NTA affinity chromatography[38]. Recombinant FLAG-ISWI was prepared as described[39]. Typical yields were >100 µg purified recombinant protein from 1 L Sf9 cell culture with the exception of SUUR polypeptides that produced no more than 2 µg from 1 L culture, which also adversely affected the protein purity. Full-length recombinant human BRG1 (SMARCA4, EpiCypher) was produced in Sf9 cells and obtained from EpiCypher, Inc.

E. coli

Human TXNRD1 sequence was cloned from a cDNA provided by Addgene (#38863), and TXNRD2 was synthesized as a gBlock gene fragment by IDT, Inc. Both full-length ORFs were inserted in a pET backbone vector in frame with a C-terminal intein-CBD (chitin-binding domain) tag. Protein expression was induced by IPTG in Rosetta 2 cells, and proteins were purified in non-denaturing conditions by chitin affinity chromatography and intein self-cleavage[40] followed by anion exchange chromatography (Source 15Q) on FPLC. Detailed cloning and purification methods are available upon request. Note that the cloned human thioredoxin reductase ORFs do not express the C-terminal selenocysteines. They were thus

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

presumed catalytically inactive[41,42] and designated hTXNRD1ci and hTXNRD2ci. They were used exclusively as spike-in mass standards in LCMS acquisitions of *Drosophila* proteins.

Polypeptide corresponding to the C-terminal specific region of Mod(Mdg4)-67.2 was cloned in pET24b vector in frame with a C-terminal His_6 tag. `M`-Mod(Mdg4)-67.2(403..610)-`GILEHHHHHH*` was expressed in Rosetta 2 in non-denaturing conditions and purified by Ni-NTA affinity chromatography. The polypeptide (ModT) was dialyzed into PBS (137 mM NaCl, 3 mM KCl, 8 mM NaH_2PO_4, 2 mM KH_2PO_4) and used as an antigen for immunizations (see below). All recombinant proteins were examined by SDS-PAGE along with Pierce BSA mass standards (Thermo Fisher), and their concentrations were calculated from infrared scanning of Coomassie-stained gels (Odyssey Fc Imaging System, LI-COR Biosciences).

Crude cell extracts

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[43]. Protein concentration was measured by Pierce BCA assay (Thermo Fisher).

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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.

Fractionation of *Drosophila* embryonic nuclear extracts

Drosophila nuclear extracts were fractionated by FPLC (**Fig. 1b** and **Extended Data Fig. E3b**) 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 MMD4-PT fractions were diluted to 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 available upon request.

Mass-spectroscopy samples

Column fractions

For each chromatographic step, 14 to 20 fractions were selected based on the protein fractionation profile according to the UV (A_{280}) absorbances measurements. 50-100 µl aliquots of chromatographic fractions, starting material (SM) and column flow-through (FT, if applicable) were saved, and protein concentrations were estimated based on their UV absorbances (1,000 mU A_{280} was considered to be equivalent to 5 mg/ml total protein). Equal volumes of each fraction, SM and FT were used for MS acquisitions, so that no more than 40 µg total protein was processed in each reaction. As a reference, the reactions were supplemented

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

with 1.5 µg each of purified recombinant human thioredoxin reductases 1 and 2 (hTXNRD1ci and hTXNRD2ci, catalytically inactive) expressed in *E. coli*. Dithiotreitol (DTT) was added to the protein samples to 10 mM and NP-40 – to 0.02%. Reaction volumes were brought to 85 µl with 50 mM ammonium bicarbonate (ABC). All reagents, including water, were HPLC/MS grade. The proteins were reduced for 1 h at 37°C and then alkylated with 30 mM iodoacetamide (IAA, 15 µl 200 mM IAA in water) for 45 min at room temperature in the dark. Alkylated proteins were desalted into 50 mM ABC using ZebaSpin columns (40 kDa MWCO) and digested with 1 µg trypsin for 2 h at 37°C. 1 µg more trypsin was added, and the digestion progressed at 37°C overnight. Tryptic peptides were lyophilized for 2 h on SpeedVac with heat and resuspended in 100 µl Sample Buffer: 1% acetonitrile (ACN) and 0.1% formic acid (FA) in water. Equal volumes (23 µl) of samples were used for IDA and SWATH acquisitions (in triplicate) as described below.

Recombinant SUUR

To generate the recombinant SUUR reference spectral library (ILR), ~0.5 µg purified recombinant FLAG-SUUR (both 130 and 65 kDa bands, **Fig. 1a**) was mixed with 1.5 µg each of hTXNRD1ci and hTXNRD2ci and processed for an IDA acquisition as described above, except for 0.5 µg trypsin was used in each cleavage step, and the peptide sample was resuspended in 30 µl Sample Buffer. For SWATH titration of SUUR (**Extended Data Figs. E1b, c**), 1 µg recombinant FLAG-SUUR was mixed with 25 µg *E. coli* lysate protein and 1.5 µg each of hTXNRD1ci and hTXNRD2ci. 10-fold serial dilutions down to 10 fg SUUR were also prepared using the mixture of *E. coli* lysate with reference proteins. The samples were processed for SWATH acquisitions in triplicate as described above, 30 µl of sample per injection.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

In-gel digestion of recombinant proteins for LCMS identification

Recombinant SUUR purified by FLAG immunoaffinity chromatography was resolved on SDS-PAGE, stained with Coommassie Blue (**Fig. 1a**), and eight most prominent protein bands were excised. The gel slices were transferred to 1.5-ml Eppendorf tubes, gently crushed with a RotoDounce pestle and destained with 25 mM ABC in 50% methanol and then with 25 mM ABC in 50% ACN (30 min each at room temperature). The proteins were reduced by 10 mM DTT for 1 h at 55°C and alkylated with 30 mM IAA for 45 min at room temperature in the dark. The gel fragments were washed with 25 mM ABC in 50% ACN, dehydrated with 100% ACN, dried in a SpeedVac, rehydrated by addition of 50 µl 50 mM ABC and digested with 0.25 µg trypsin overnight at 37°C. The peptides were extracted once with 50 µl 10% FA and once with 100 µl 3% FA in 60% ACN, both extracts were combined, dried in a SpeedVac and resuspended in 50 µl Sample Buffer. 40 µl of each sample was injected for IDA acquisitions as described below.

Mass-spectroscopy acquisition methods

Instrument settings

LC-MS/MS analyses were performed on a TripleTOF 5600+ mass spectrometer (AB SCIEX) coupled with M5 MicroLC system (AB SCIEX/Eksigent) and PAL3 autosampler. LC separation was performed in a trap-elute configuration, which consists of a trapping column (LUNA C18(2), 100 Å, 5 µm, 20 × 0.3 mm cartridge, Phenomenex) and an analytical column (Kinetex 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 the 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 data 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 with 21,970 protein entries downloaded from UniProt on 3/21/2020 (UP000000803) 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 acquisitions[44], 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[45]. 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 Table S2**) and processed as described below.

MERCI

MERCI is a novel approach for rapid identification of native protein complexes. It combines gradual enrichment for a target subunit of a putative complex by consecutive FPLC steps with parallel quantitative shotgun proteomics of chromatographic fractions. Crude nuclear extract from *Drosophila* embryos was fractionated as in **Fig. 1a**. 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 two exogenous reference proteins (human thioredoxin reductases for isolation of

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Drosophila complexes), 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 were further normalized by converting them to Z-scores (see **Supplementary Table S2** for an example of calculations). Peak SUUR fractions (typically, 1-5) were then subjected to the next FPLC/MERCI step. After five column steps, the ion library from the ultimate FPLC step (IL5) was used to re-quantify SWATH data from all steps. Z-scores for all purification steps were stitched together, and the large array encompassing all data points for every protein was analyzed by Pearson correlation with SUUR (**Supplementary Table S2**). The most closely correlated purification profiles served as an indication for protein co-purification, potentially, as subunits of a stable complex.

The dissection of protein interactome by extract fractionation on orthogonal FPLC columns and MS-based approaches has been previously attempted[46,47]. 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, **Figs. 1a**, **d**). The target protein and co-purifying polypeptides are substantially enriched after several chromatographic steps (**Extended Data Fig. E1e**) and thus, yield a greater number of detected

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

peptides, which helps a more precise quantification. Although SWATH allows reliable measurement of picogram amounts of proteins (**Extended Data Fig. E1c**), the range of quantified polypeptides is limited by those present in IDA acquisitions. For low-abundance proteins, such as SUUR and Mod(Mdg4), specific peptides are not detectable by IDA in earlier chromatographic steps (**Supplementary Table S1**). Consequently, SWATH quantification using only the cognate ion libraries would not discern the near perfect co-fractionation of SUUR and Mod(Mdg4) in all five steps (**Fig. 1d**), precluding identification of the SUUR- Mod(Mdg4) complex (**Figs. 1d-f**).

Biochemical assays with recombinant proteins

Oligonucleosome substrates

Oligonucleosomes were reconstituted *in vitro* as described[48] from supercoiled plasmid DNA (3.2 kb, pGIE-0), native core histones and H1 prepared from *Drosophila* embryos[49] by gradient salt dialysis in the presence of 0.2 mg/ml nuclease-free bovine serum albumin (BSA, New England Biolabs). Quality of reconstitution was assessed by MNase and chromatosome stop assays (**Extended Data Figs. E2d**, **e**).

ATPase assay

40 nM recombinant proteins were incubated in 25 µl reaction buffer containing 20 mM HEPES, pH 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 nuclease-free BSA for 60 min at 27°C. Some reactions additionally contained 10 nM pGIE-0 plasmid DNA or equivalent amounts of oligonucleosomes ±H1. ATPase assays were performed using ADP-Glo Max kit (Promega). All reactions were performed in triplicate, the

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

results were normalized to the ADP-ATP titration curve according to the kit manual and converted to enzymatic rates (molecules of ATP hydrolyzed per molecule of enzyme per minute). Averages and standard deviations were calculated. Statistical differences were calculated by Mann-Whitney test.

EpiDyne®-PicoGreen™ nucleosome remodeling assay

EpiDyne®-PicoGreen™ is a restriction-enzyme accessibility assay format modified for increased throughput and sensitivity (**Extended Data Fig. 2f**). In brief, a recombinant ATPase over a concentration range (**Extended Data Figs. E2g-j**) was mixed with 10 nM EpiDyne biotinylated nucleosome remodeling substrate (EpiCypher) , terminally positioned 6-N-66 (219 bp DNA fragment) or centrally positioned 50-N-66 (263 bp) and 1 mM ATP in 20 µL remodeling buffer, 20 mM Tris-HCl, pH 7.5, 50 mM KCl, 3 mM $MgCl_2$, 0.01% (v/v) Tween-20, 0.01% (w/v) BSA. The remodeling reactions were incubated at 23˚C in 384-well format. At indicated time points, the reactions were quenched, and nucleosome substrates were immobilized on an equal volume of streptavidin-coated magnetic beads (NEB), pre-washed and resuspended in 2x quench buffer, 20 mM Tris-HCl, pH 7.5, 600 mM KCl, 0.01% (v/v) Tween-20 and 0.01% (w/v) BSA. Beads were successively washed by collection on a magnet (three times with wash buffer, 20 mM Tris-HCl, pH 7.5, 300 mM KCl, 0.01% (v/v) Tween-20) and buffer replacement (once with RE buffer, 20 mM Tris-HCl, pH 7.5, 50 mM KCl, 3 mM $MgCl_2$, 0.01% (v/v) Tween-20). Beads were resuspended in 20 µl restriction enzyme mix, 50 units/ml Dpn II (NEB) in RE buffer, and incubated at 23˚C for 30 min, collected on a magnet, and supernatants from all wells were transferred to a new plate. They were mixed with an equal volume of Quant-iT™ PicoGreen™ dsDNA reagent (ThermoFisher, Component A) and 1 unit/ml thermolabile

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

proteinase K (NEB) in TE and incubated at 23˚C for 1 hr. Fluorescence intensity was detected on an Envision microplate reader with excitation at 480 nm and emission at 531 nm, and data expressed as relative fluorescence units (RFU) through the EnVision Workstation (version 1.13.3009.1409).

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. Details and PCR primer sequences are available upon request.

Fly wings were dissected from ~5 days old adult males and transferred to the 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. 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)

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Polyclonal antibody (anti-ModT) was raised in Guinea pigs by Pocono Rabbit Farm & Lab. Rabbit polyclonal antibody to the C-terminus of *Drosophila* XNP/ATRX (anti-XNP) was described previously[16]. Rabbit and Guinea pig polyclonal antibodies to *Drosophila* SUUR were a gift of Alexey Pindyurin and Igor Zhimulev. Rabbit polyclonal MMD4(FL) antibody to full-length Mod(Mdg4)-67.2 that recognizes all splice forms of Mod(Mdg4) was a gift of Jordan Rowley and Victor Corces. Mouse monoclonal anti-FLAG (M2, Sigma Aldrich), anti-PCNA (PC10, Cell Signaling), anti-β-tubulin and anti-HP1a (AA12.1 and C1A9, Developmental Studies Hybridoma Bank) were obtained commercially.

Western blotting was performed using standard techniques. For FPLC column fraction analyses, 5-10 µl of starting material and flow-through (if applicable) and 5-15 µl of column fractions were loaded per lane. For expression analyses in salivary glands, 10 salivary glands from L3 larvae of indicated genotype were frozen and thawed, boiled extensively in 40 µl 2x SDS-PAGE loading buffer, centrifuged, and the material equivalent to four salivary glands was loaded per lane. The following dilutions were used: 1:200,000 anti-ModT, 1:1,000 anti-MMD4(FL), 1:1,000 Guinea pig and rabbit anti-SUUR, 1:1,000 anti-HP1a, 1:1,000 anti-β-tubulin and 1:2,000 anti-FLAG. Infrared-labeled secondary antibodies: donkey anti-Guinea pig IRDye 800CW, goat anti-mouse IRDye 800CW, goat anti-rabbit IRDye 800CW, goat anti-rabbit IRDye 680CW and goat anti-mouse IRDye 680RD – were obtained from Li-COR Biosciences and used at 1:10,000. The blots were scanned on Odyssey Fc Imaging System (LI-COR Biosciences).

Immunoprecipitation experiments were performed as described[50]. 400 µl *Drosophila* embryonic nuclear extracts (~10 mg total protein) were incubated with 10 µl Guinea pig anti-ModT, 30 µl rabbit anti-SUUR or 20 µl rabbit anti-XNP antibodies for 3 h at 4°C.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Immunocomplexes were collected by addition of 25 µl protein A-agarose plus (Thermo Fisher) for 2 h at 4°C. After washing four times with 1 ml of buffer HEG (25 mM HEPES, pH 7.6, 0.1 mM EDTA, 10% glycerol) + 0.15 M NaCl, the immunoprecipitated proteins were eluted with 80 µl 2x SDS-PAGE loading buffer and analyzed by SDS-PAGE and western blot using Guinea pig or rabbit anti-SUUR and anti-Mod(Mdg4) and mouse anti-HP1a antibodies. For Mod(Mdg4) and HP1a, 8 µl of immunoprecipitated material (equivalent to 1 mg nuclear extract proteins) and 5% input (2 µl nuclear extract, 50 µg total protein) were analyzed. For SUUR, 20 µl of immunoprecipitated material (equivalent to 2.5 mg nuclear extract proteins) and 10% input (10 µl nuclear extract, 250 µg total protein) were analyzed.

Polytene chromosomes and indirect immunofluorescence (IF) analyses

Polytene chromosomes were prepared and analyzed as described previously[19]. For all cytological experiments, larvae were reared and collected at 18°C. Salivary glands from wandering third instar larvae were dissected in PBS. Glands were transferred into a formaldehyde-based fixative (one ~15-µl drop of 3% lactic acid, 45% acetic acid, 3.7% formaldehyde on a coverslip) for 2 min[51], squashed, and frozen in liquid N_2. The coverslips were removed, and slides were placed in 70% ethanol for 20 min and stored at −20°C. The slides were washed three times for 5 min in PBST. Primary antibodies were incubated overnight at 4°C in PBST + 0.1% BSA and washed three times for 5 min each with PBST. Secondary antibodies were incubated for 2 h at room temperature in PBST + 0.1% BSA and washed three times for 5 min each with PBST.

DNA was stained with 0.1 µg/ml DAPI in PBST for 3 min, and squashes were mounted in Prolong Glass anti-fade mountant (Molecular Probes). Primary and secondary antibodies were

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

used at the following dilutions: Guinea pig anti-ModT, 1:50,000; rabbit anti-SUUR, 1:100; mouse anti-PCNA, 1:1,000; mouse anti-FLAG, 1:100; Alexa Fluor 488 highly cross-absorbed (HCA) goat anti-mouse, Alexa Fluor 568 HCA goat anti-Guinea pig and Alexa Fluor 647 plus HCA goat anti-rabbit, all 1:800. Indirect immunofluorescence (IF) images were obtained with Zeiss AxioVERT 200M microscope and AxioCam MRm mono microscopy camera using a 40x/1.3 Plan-Neofluar or 63x/1.40 Plan-Apochromat lenses with oil immersion. Images were acquired using AxioVision software.

For whole-mount IF staining, L3 larvae were reared at 18°C, and salivary glands were dissected in PBS and fixed in 3.7% formaldehyde (Sigma Aldrich) for 20 min at room temperature. The glands were washed in PBS + 0.3% Triton X-100 and permeabilized for 30 min at 37°C in PBS + 1% Triton X-100. Blocking was performed for 30 min at room temperature in PBS+ 0.3% Triton X-100 supplemented with 10% fetal calf serum and 1% BSA. The glands were incubated with primary antibodies diluted in blocking solution for 48 h at 4°C, washed three times with PBS + 0.3% Triton X-100 for 30 min, and incubated with secondary antibodies in blocking solution overnight at 4°C. The stained glands were washed three times with PBS + 0.3% Triton X-100 for 30 min, stained with DAPI (0.1 μg/ml) for 30 min, and mounted in Prolong Gold anti-fade (Invitrogen). IF images were obtained on a Leica SP8 confocal microscope using a 20X/0.75 PLAPO lens and processed using ImageJ software (Fiji).

Next generation sequencing analyses (NGS)

Salivary glands from female wandering third-instar larvae were isolated and flash-frozen in liquid N_2 until all samples were collected. Genomic DNA for sequencing was prepared from 25 L3 salivary gland pairs or 10 mg embryos (0-6 h AED) using DNeasy Blood and Tissue kit

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(Qiagen). Each sample was prepared in triplicate. The tissues were soaked in 180 µl buffer ATL + 20 µl proteinase K (15 mg/ml) and lysed for 2-3 h at 55°C. The reactions were cooled to room temperature, supplemented with 4 µl RNase A, ~40 mg/ml (Sigma Aldrich), and RNA was digested for 10 min at room temperature. The genomic DNA was fragmented with 0.002 units DNase I (Thermo Fisher) in 100-µl reactions containing 10 mM Tris-HCl, 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 reaction buffer) for 15 min at 37°C. (DNAse I dilutions were prepared using 1x reaction buffer.) Reactions were stopped by adding 5 µl 0.5 M EDTA, and DNase I was inactivated for 20 min at 65°C. The fragmented DNA was purified on QiaQuick columns using PCR purification kit (Qiagen) and eluted in 40 µl 10 mM Tris-HCl, pH 8.0. The size distribution of DNA fragments (200-600 bp, average ~400 bp) was confirmed and DNA concentration was measured on 2100 BioAnalyzer (Agilent). Libraries were prepared from 20 ng of fragmented genomic DNA with the ThruPLEX DNA-seq kit using SMARTer® DNA Unique Dual Indexes (TakaraBio) and sequenced 150-bp paired-end reads on an NovaSeq 6000 (Novagene).

The sequencing quality of each sample was assessed using FASTQC version 0.11.7[52]. Raw paired-end reads were trimmed of adapters using BBDuk from the BBTools software version 38.71 using the parameters: `ktrim=r ref=adapters rcomp=t tpe=t tbo=t hdist=1 mink=11`[53]. Reads were aligned to the BDGP Release 6 of the *Drosophila melanogaster* genome (dm6)[54] using Bowtie2 version 2.3.4.1[55] and parameters `-q --local --very-sensitive-local --no-unal --no-mixed --no-discordant --phred33 -I 10 -X 700`. Duplicate reads were marked using Picard 2.2.4[56] and SAM files were converted to BAM format, filtered for quality (`-bq 5`), and removed of duplicates (`-bF 0x400`) using Samtools version 1.9[57]. For visualization, replicates were merged (`samtools`

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

`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`[58]. 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[36]. Su(Hw) insulator class II elements were also used[59]. UR 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[19]. Visualization of all data was performed on the UCSC Genome browser using the dm6 release of the *Drosophila* genome[60]. 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 **Extended Data Table E2**. 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: $\langle\Delta\Delta Ct\rangle = (\langle Ct_{target}\rangle - \langle Ct_{intergenic86D}\rangle)_{SG} - (\langle Ct_{target}\rangle - \langle Ct_{intergenic86D}\rangle)_{embryo}$. Standard deviations for

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

$<\Delta\Delta Ct>$ were calculated as $\sigma_{\Delta\Delta Ct}$ = square root of (σ^2_{target} + $\sigma^2_{intergenic86D}$)/2. $\Delta\Delta Ct$'s were converted to DNA copy numbers as $2^{-<\Delta\Delta Ct>}$. The confidence interval was calculated in the range of $2^{-<\Delta\Delta Ct>-\sigma}$ and $2^{-<\Delta\Delta Ct>+\sigma}$.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

1 Belyaeva, E. S. *et al.* Su(UR)ES: a gene suppressing DNA underreplication in intercalary and pericentric heterochromatin of Drosophila melanogaster polytene chromosomes. *Proc Natl Acad Sci U S A* **95**, 7532-7537 (1998).

2 Nordman, J. T. *et al.* DNA copy-number control through inhibition of replication fork progression. *Cell reports* **9**, 841-849, doi:10.1016/j.celrep.2014.10.005 (2014).

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

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

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

6 Rhind, N. & Gilbert, D. M. DNA replication timing. *Cold Spring Harbor perspectives in biology* **5**, a010132, doi:10.1101/cshperspect.a010132 (2013).

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

9 Zielke, N., Edgar, B. A. & DePamphilis, M. L. Endoreplication. *Cold Spring Harbor perspectives in biology* **5**, a012948, doi:10.1101/cshperspect.a012948 (2013).

10 Zhimulev, I. F. *et al.* Polytene chromosomes: 70 years of genetic research. *International review of cytology* **241**, 203-275, doi:10.1016/S0074-7696(04)41004-3 (2004).

11 Sher, N. *et al.* Developmental control of gene copy number by repression of replication initiation and fork progression. *Genome research* **22**, 64-75, doi:10.1101/gr.126003.111 (2012).

12 Kolesnikova, T. D. *et al.* Drosophila SUUR protein associates with PCNA and binds chromatin in a cell cycle-dependent manner. *Chromosoma* **122**, 55-66, doi:10.1007/s00412-012-0390-9 (2013).

13 Pindyurin, A. V. *et al.* Interaction between the Drosophila heterochromatin proteins SUUR and HP1. *Journal of cell science* **121**, 1693-1703, doi:10.1242/jcs.018655 (2008).

14 Munden, A. *et al.* Rif1 inhibits replication fork progression and controls DNA copy number in Drosophila. *eLife* **7**, doi:10.7554/eLife.39140 (2018).

15 Buchner, K. *et al.* Genetic and molecular complexity of the position effect variegation modifier mod(mdg4) in Drosophila. *Genetics* **155**, 141-157, doi:10.1093/genetics/155.1.141 (2000).

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

17 Wang, H. *et al.* mAM facilitates conversion by ESET of dimethyl to trimethyl lysine 9 of histone H3 to cause transcriptional repression. *Mol Cell* **12**, 475-487, doi:10.1016/j.molcel.2003.08.007 (2003).

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

19 Andreyeva, E. N. *et al.* Regulatory functions and chromatin loading dynamics of linker histone H1 during endoreplication in Drosophila. *Genes Dev* **31**, 603-616, doi:10.1101/gad.295717.116 (2017).

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

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

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

23 Adryan, B. *et al.* Genomic mapping of Suppressor of Hairy-wing binding sites in Drosophila. *Genome biology* **8**, R167, doi:10.1186/gb-2007-8-8-r167 (2007).

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

26 Roseman, R. R. *et al.* A P element containing suppressor of hairy-wing binding regions has novel properties for mutagenesis in Drosophila melanogaster. *Genetics* **141**, 1061-1074, doi:10.1093/genetics/141.3.1061 (1995).

27 Belyaeva, E. S. *et al.* Effect of the Suppressor of Underreplication (SuUR) gene on position-effect variegation silencing in Drosophila melanogaster. *Genetics* **165**, 1209-1220, doi:10.1093/genetics/165.3.1209 (2003).

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

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

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

31 Rowley, M. J. *et al.* Evolutionarily Conserved Principles Predict 3D Chromatin Organization. *Mol Cell* **67**, 837-852 e837, doi:10.1016/j.molcel.2017.07.022 (2017).

32 Davidson, I. F. *et al.* DNA loop extrusion by human cohesin. *Science* **366**, 1338-1345, doi:10.1126/science.aaz3418 (2019).

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

34 Van Bortle, K. *et al.* Insulator function and topological domain border strength scale with architectural protein occupancy. *Genome biology* **15**, R82, doi:10.1186/gb-2014-15-5-r82 (2014).

35 Schwartz, Y. B. *et al.* Nature and function of insulator protein binding sites in the Drosophila genome. *Genome research* **22**, 2188-2198, doi:10.1101/gr.138156.112 (2012).

36 Filion, G. J. *et al.* Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. *Cell* **143**, 212-224, doi:10.1016/j.cell.2010.09.009 (2010).

37 Vorobyeva, N. E. *et al.* Insulator protein Su(Hw) recruits SAGA and Brahma complexes and constitutes part of Origin Recognition Complex-binding sites in the Drosophila genome. *Nucleic Acids Res* **41**, 5717-5730, doi:10.1093/nar/gkt297 (2013).

38 Fyodorov, D. V. & Kadonaga, J. T. Chromatin assembly in vitro with purified recombinant ACF and NAP-1. *Methods Enzymol* **371**, 499-515 (2003).

39 Ito, T. *et al.* ACF consists of two subunits, Acf1 and ISWI, that function cooperatively in the ATP-dependent catalysis of chromatin assembly. *Genes Dev* **13**, 1529-1539 (1999).

40 Emelyanov, A. V. *et al.* 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**, 2027-2040, doi:10.1101/gad.248583.114 (2014).

41 Arner, E. S., Sarioglu, H., Lottspeich, F., Holmgren, A. & Bock, A. 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. *Journal of molecular biology* **292**, 1003-1016, doi:10.1006/jmbi.1999.3085 (1999).

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

42 Cheng, Q. & Arner, E. S. 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**, 5476-5487, doi:10.1074/jbc.M117.776310 (2017).

43 Kamakaka, R. T., Tyree, C. M. & Kadonaga, J. T. 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**, 1024-1028 (1991).

44 Zhu, X., Chen, Y. & Subramanian, R. 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**, 1202-1209, doi:10.1021/ac403385y (2014).

45 Zhang, Y. *et al.* The Use of Variable Q1 Isolation Windows Improves Selectivity in LC-SWATH-MS Acquisition. *Journal of proteome research* **14**, 4359-4371, doi:10.1021/acs.jproteome.5b00543 (2015).

46 Havugimana, P. C. *et al.* A census of human soluble protein complexes. *Cell* **150**, 1068-1081, doi:10.1016/j.cell.2012.08.011 (2012).

47 Shatsky, M. *et al.* Quantitative Tagless Copurification: A Method to Validate and Identify Protein-Protein Interactions. *Molecular & cellular proteomics : MCP* **15**, 2186-2202, doi:10.1074/mcp.M115.057117 (2016).

48 Lu, X. *et al.* Drosophila H1 regulates the genetic activity of heterochromatin by recruitment of Su(var)3-9. *Science* **340**, 78-81, doi:10.1126/science.1234654 (2013).

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

49 Fyodorov, D. V. & Levenstein, M. E. in *Current Protocols in Molecular Biology* 21.27.21-21.27.27 (Wiley & Sons, 2002).

50 Emelyanov, A. V. *et al.* Identification and characterization of ToRC, a novel ISWI-containing ATP-dependent chromatin assembly complex. *Genes Dev* **26**, 603-614, doi:10.1101/gad.180604.111 (2012).

51 Novikov, D. V., Kireev, I. & Belmont, A. S. High-pressure treatment of polytene chromosomes improves structural resolution. *Nature methods* **4**, 483-485, doi:10.1038/nmeth1049 (2007).

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

53 Bushnell, B. *BBTools software package.*, <http://bbtools.jgi.doe.gov> (2014).

54 dos Santos, G. *et al.* FlyBase: introduction of the Drosophila melanogaster Release 6 reference genome assembly and large-scale migration of genome annotations. *Nucleic Acids Res* **43**, D690-697, doi:10.1093/nar/gku1099 (2015).

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

56 BroadInstitute. *"Picard Tools." Broad Institute, GitHub repository.* , <http://broadinstitute.github.io/picard/> (

57 Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10**, doi:10.1093/gigascience/giab008 (2021).

58 Ramirez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160-165, doi:10.1093/nar/gkw257 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

59 Negre, N. *et al.* A comprehensive map of insulator elements for the Drosophila genome. *PLoS Genet* **6**, e1000814, doi:10.1371/journal.pgen.1000814 (2010).

60 Kent, W. J. *et al.* The human genome browser at UCSC. *Genome research* **12**, 996-1006, doi:10.1101/gr.229102 (2002).

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

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

63 Lefevre, G. in *The Genetics and Biology of Drosophila.* Vol. Ia (ed Ashburner & Novitski) 31-66 (Academic Press, 1976).

64 Hoskins, R. A. *et al.* The Release 6 reference sequence of the Drosophila melanogaster genome. *Genome research* **25**, 445-458, doi:10.1101/gr.185579.114 (2015).

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure legends

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Asterisks, IgG heavy and light chains detected due to antibody cross-reactivity. Mod(Mdg4)-67.2(FL) antibody recognizes all splice forms of Mod(Mdg4).

Fig. 2. Biochemical activities of SUMM4. a, Recombinant SUMM4. Mod(Mdg4)-His_6, 67.2 (p100, cyan arrowhead) and 59.1 (p75, green arrowhead) splice forms were co-expressed with FLAG-SUUR (red arrowheads, p130 and p65) or separately in Sf9 cells and purified by FLAG or Ni-NTA affinity chromatography. Mod(Mdg4)-67.2 forms a specific complex with SUUR. **b**, ATPase activities of recombinant FLAG-ISWI (green bars), FLAG-SUUR (red bars) and SUMM4 (FLAG-SUUR + Mod(Mdg4)-67.2-His_6, purple bars). Equimolar proteins were analyzed in reactions in the absence or presence of plasmid DNA or equivalent amounts of reconstituted oligonucleosomes, ±H1. rSUUR(KA) and rMMD4, ATPases activities of K59A mutant of SUUR (gray bars) and Mod(Mdg4)-67.2-His_6 (cyan bars). Hydrolysis rates were converted to moles ATP per mole protein per minute. All reactions were performed in triplicate, error bars represent standard deviations. *p*-values for statistically significant differences are indicated (Mann-Whitney test). **c**, DNA- and nucleosome-dependent stimulation or inhibition of ATPase. The activities were analyzed as in (**b**). Statistically significant differences are shown (Mann-Whitney test). **d**, Nucleosome sliding activities by EpiDyne®-PicoGreen™ assay (see *Methods*) with 5 nM of recombinant ISWI, SUUR or SUMM4. Reaction time courses are shown for terminally (6-N-66) and centrally (50-N-66) positioned mononucleosomes (**Extended Data Figs. E2g-j**). RFU, relative fluorescence units produced by PicoGreen fluorescence.

Fig. 3. Temporal and spatial 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 (**Extended Data Fig. E3a**). ModT (green) and SUUR (red) signals overlap extensively in euchromatic arms. The additional strong ModT IF loci that are SUUR-free and Mod(Mdg4)-67.2-free SUUR in the chromocenter are obvious. DAPI staining shows the overall chromosome morphology. **b**, SUUR loading into chromosomes during early endo-S phase is compromised in *mod(mdg4)* mutants. *SuUR* mutation does not appreciably change the distribution of Mod(Mdg4)-67.2. Polytene chromosomes were prepared from *wild-type* and mutant salivary glands and stained with ModT and SUUR antibodies as in (**a**). Endo-S timing was established by PCNA staining (**Extended Data Fig. E3g**). **c**, Abnormal subcellular distribution of SUMM4 subunits in *mod(mdg4)* mutants. L3 salivary glands were fixed and whole-mount-stained with DAPI, ModT and SUUR antibodies. Whereas both polypeptides are mostly nuclear in the wild type, they are partially mis-localized to cytoplasm in *mod(mdg4)*^{u1} mutant.

Fig. 4. Biological functions of SUMM4 in regulation of gene expression and DNA replication. a, SUMM4 subunits are required for the enhancer-blocking activity at *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* [61]. 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**, SUMM4 subunits are required for the chromatin barrier activity of Su(Hw) binding sites. Top: schematic diagram of

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the *P{SUPor-P}* reporter system[62]; clustered 12 copies of *gypsy* Su(Hw) binding sites flanks the transcription unit of *white*. *KV00015* and *KV00138* are *P{SUPor-P}* insertions in pericentric heterochromatin of 2L. $SuUR^{ES}$ and $mod(mdg4)^{u1}$ alleles are recessive suppressors of the boundary that insulates *white* from heterochromatin encroaching. **c**, Schematic model for the function of SUMM4 in blocking enhancer-promoter interactions in the ct^6 locus. **d**, Schematic model for the function of SUMM4 in establishing a chromatin barrier in heterochromatin-inserted *P{SUPor-P}* elements. **e**, Schematic model for a putative function of SUMM4 in blocking/retardation of replication fork progression in IH domains. **f**, DNA copy numbers were determined by real-time qPCR. Copy numbers were calculated relative to embryonic DNA and normalized to a control intergenic region. The X-axis shows chromosome positions (in megabase pairs) of target amplicons. Error bars represent the confidence interval (see *Methods*). Black arrowheads, positions of mapped Su(Hw) binding sites[59]. Yellow boxes show approximate boundaries of cytogenetic bands. The right column shows representative images of DAPI-stained polytene chromosome fragments corresponding to this genomic region; genotypes of the animals are as indicated, locations of cytological bands 75C1 and C2 are shown in yellow. The disk thickness and DAPI staining of 75C1-2 is strongly increased in mutants. Close-up view of DNA copy numbers in 75C (3L) from high-throughput sequencing data are presented as in (**g**). **g**, Analyses of DNA copy numbers in *Drosophila* salivary gland cells. DNA from L3 salivary glands was subjected to high-throughput sequencing. DNA copy numbers (normalized to diploid embryonic DNA) are shown across the 3L chromosome arm. Genomic coordinates in Megabase pairs are indicated at the bottom. The control trace (w^{1118} allele) is shown as semitransparent light gray in the foreground; $SuUR^{ES}$, $mod(mdg4)^{u1}$ and $mod(mdg4)^{m9}$ traces are shown in the

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

background in red, cyan and green, respectively; their overlaps with w^{1118} traces appear as lighter shades of color. Black box, 75C cytological region shown in (**g**).

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Extended data figures and tables

Extended Data Fig. E1. a, Representation of SUUR in ion libraries ILR and IL1-5 (**Fig. 1b**, **Suppl. Table S1**). Total number of identified proteins and the confidence rank of SUUR among them as well as the total number of detected peptides (95% confidence) and the number of SUUR-specific peptides are shown. **b**, Schematic of SWATH quantification of recombinant SUUR, nuclear extract (starting material) and FPLC fractions for SUUR using ion library ILR. **c**, 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). **d**, 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. **e**, SUUR purification by FPLC. Total protein was measured by BCA assay, and SUUR was measured as in (**d**). Relative purity, purification factor in each step and cumulative purification factor are shown. **f-h**, SWATH quantitation profiles of SUUR (red), EGG (green), CG6700 (blue) and HP1a (black) fractionation across five FPLC steps as in **Fig. 1d**. Pearson correlation coefficients are shown (**Fig. 1e**). **i**, Mod(Mdg4)-specific peptides from ion library IL5 (**Suppl. Table S1**). 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. **j**, 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 **i**) are highlighted in bold red. **k**, Mod(Mdg4)-59.1 polypeptide sequence. The common part is shaded in gray, splice form-specific part is shaded in green.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Extended Data Fig. E2. 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. **b**, Recombinant FLAG-SUUR(K59A) and FLAG-ISWI expressed in Sf9 cells. See legend to **Fig. 1b**. **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; molecular mass markers (kDa) are shown on the left. **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. **f**, 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™

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(dsDNA detection reagent) fluorescence. **g**, 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 indicate linear ranges. RFU, relative fluorescence units. **h**, Early remodeling rates for ISWI were calculated by linear regression analyses of data in respective linear ranges. ISWI exhibits a stronger remodeling activity with a centrally positioned nucleosome substrate. **i**, Titration of human BRG1 remodeling activity. Data are presented as in (**g**). **j**, Early remodeling rates for BRG1 were calculated and plotted as in (**h**). BRG1 does not exhibit a bias towards remodeling centrally or terminally positioned nucleosomes.

Extended Data Fig. E3. a, Colocalization of SUUR and Mod(Mdg4)-67.2 in *wild-type* polytene chromosomes. See legend to **Fig. 3a**. Positions of 3L and 3R telomeres are indicated; approximate boundaries of cytological regions are shown according to[63]; positions of IH regions 75C and 89E that are under-replicated and responsive to *SuUR* mutation are marked by circles. **b**, Schematic of partial FPLC purification of an alternative complex of Mod(Mdg4)-67.2. Cyan boxes, fraction ranges used for the next chromatographic step. **c**, 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 **Fig. 1g**. **d**, Western blot analysis of Source 15S fractions with the ModT antibody. **e**, Western blot analyses of Superose 6 fractions with the ModT antibody. Black arrows, expected peaks of globular proteins with indicated molecular masses in kDa. **f**, 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

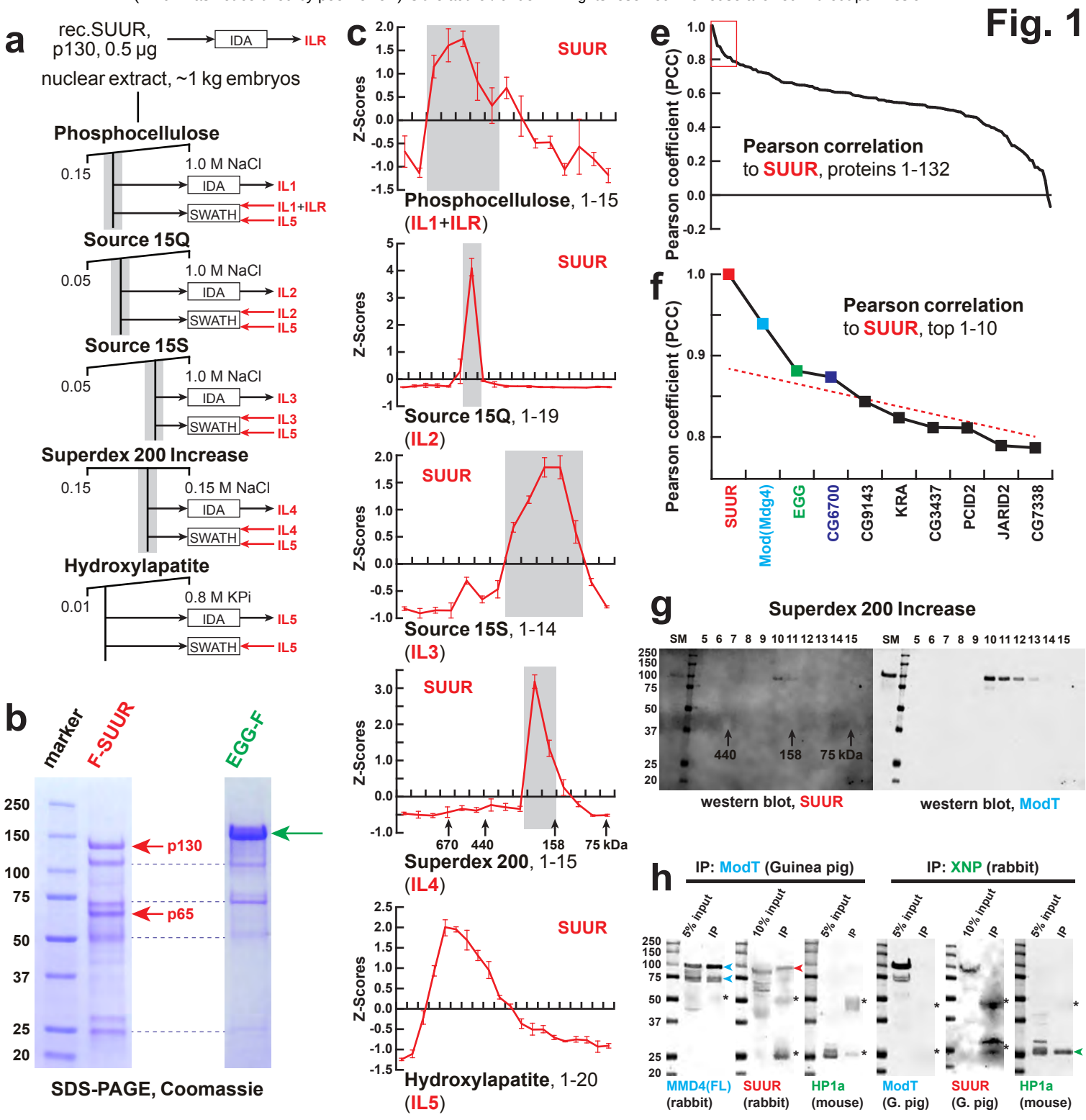
bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

control). Mass marker sizes (kDa) are shown on the left. **g**, Abnormal spatiotemporal distribution of SUUR in polytene chromosomes of *mod(mdg4)* mutants. See legend to **Fig. 3b**. 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.

Extended Data Fig. E4. a, qPCR- and NGS-measured DNA copy numbers in region 4D. The data were obtained and presented as for the 75C region (**Fig. 4f**). **b**, Genome-wide analyses of DNA copy numbers in *Drosophila* salivary gland cells (w^{1118} control). Chromosome arms are shown on the left. Brown- and green-shades boxes, mapped pericentric and telomeric heterochromatin regions[64], respectively. Asterisks, positions of UR domains (**Extended Data Table E1**). See legend to **Fig. 4g** for other designations. **c**, Genome-wide analyses of DNA copy numbers in *Drosophila* salivary gland cells in chromosomes X, 2L, 2R and 3R. The data were obtained and presented as for the 3L arm (**Fig. 4g**). **d**, Close-up views 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[59]. See (**c**) for other designations. **e**, Sample plots of DamID profiles for SUUR (red) and Su(Hw) (purple), $\log_2$ enrichment over Dam-only control[36]. Positive values are plotted in dark colors and negative values in light colors for contrast. DNA copy numbers in salivary gland cells (black) indicate UR IH domains. Vertical bars, Su(Hw) binding sites[59].

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fig. 1



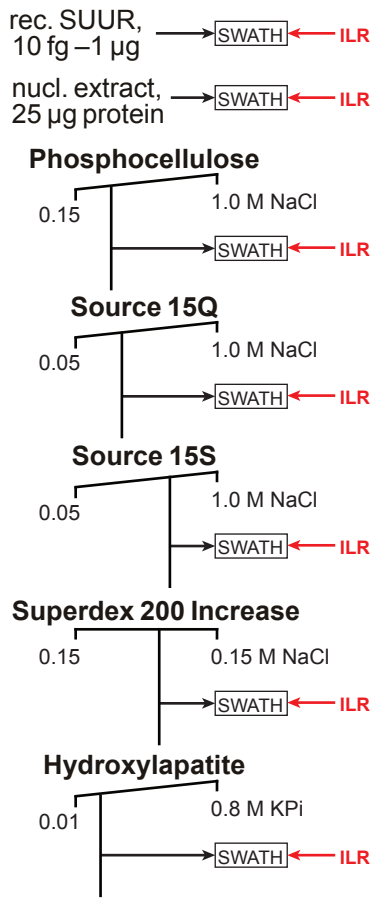
bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fig. E1

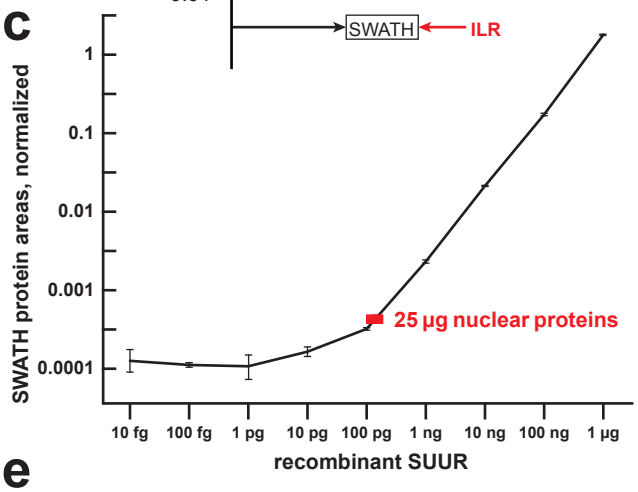
a

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

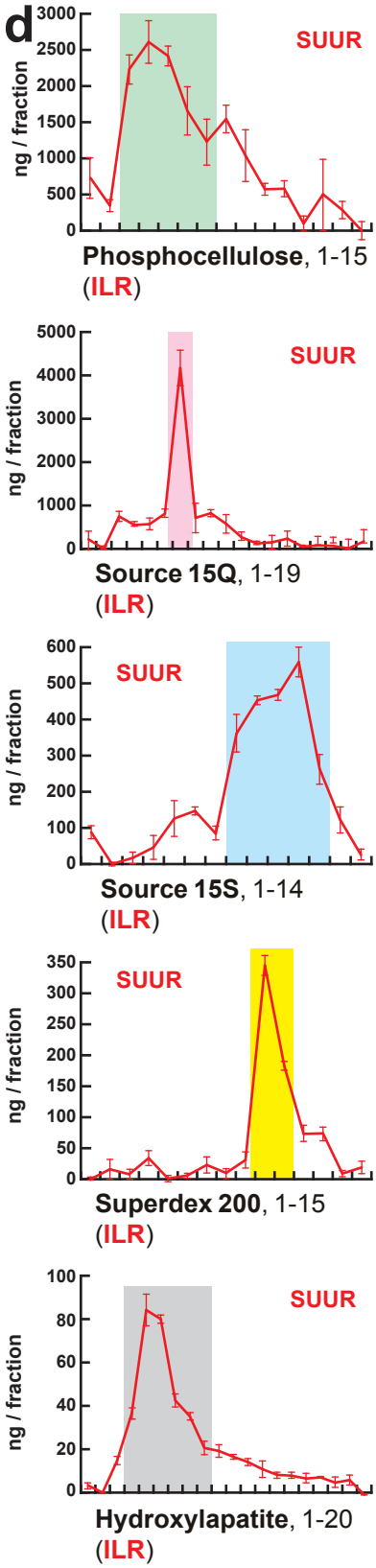
b



c



d

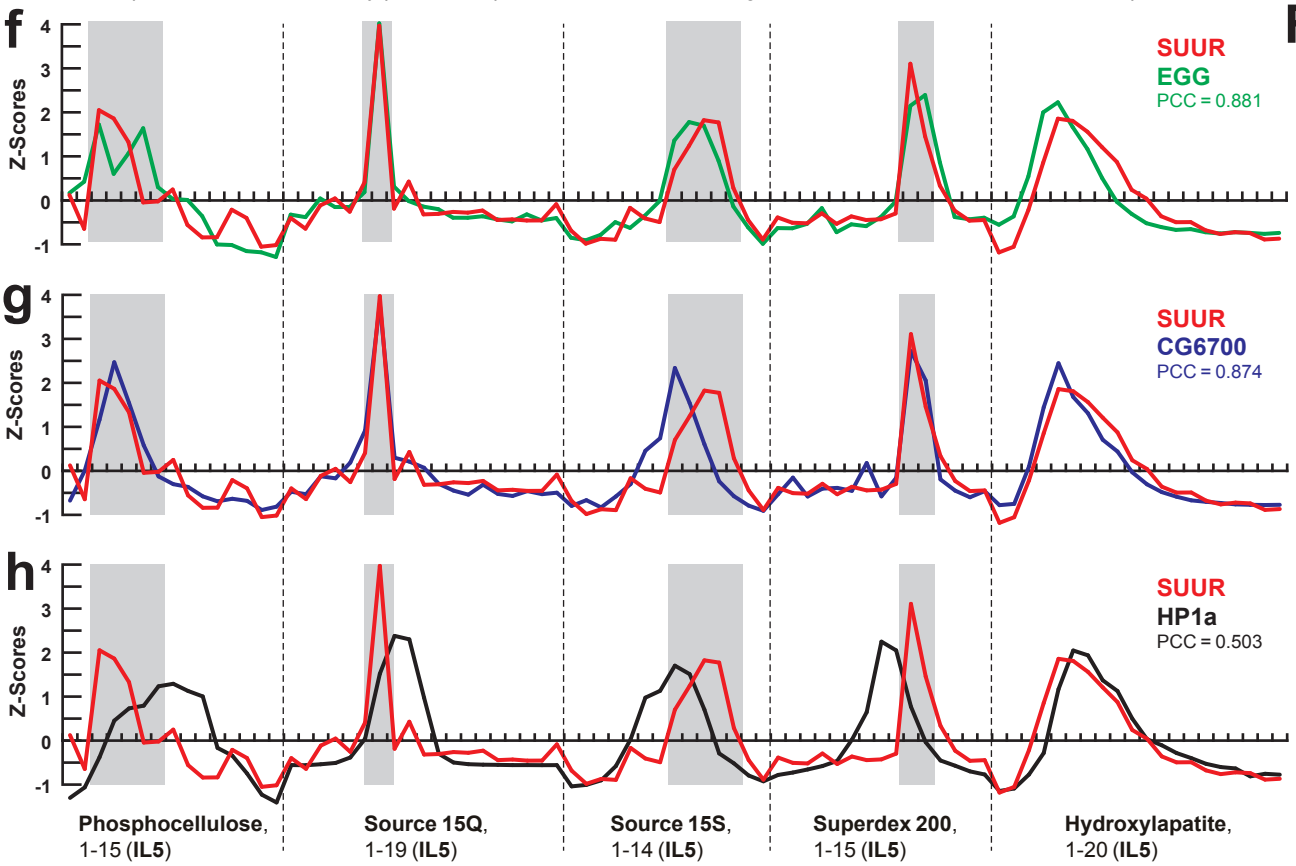


e

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fig. E1



i

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

j **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
```

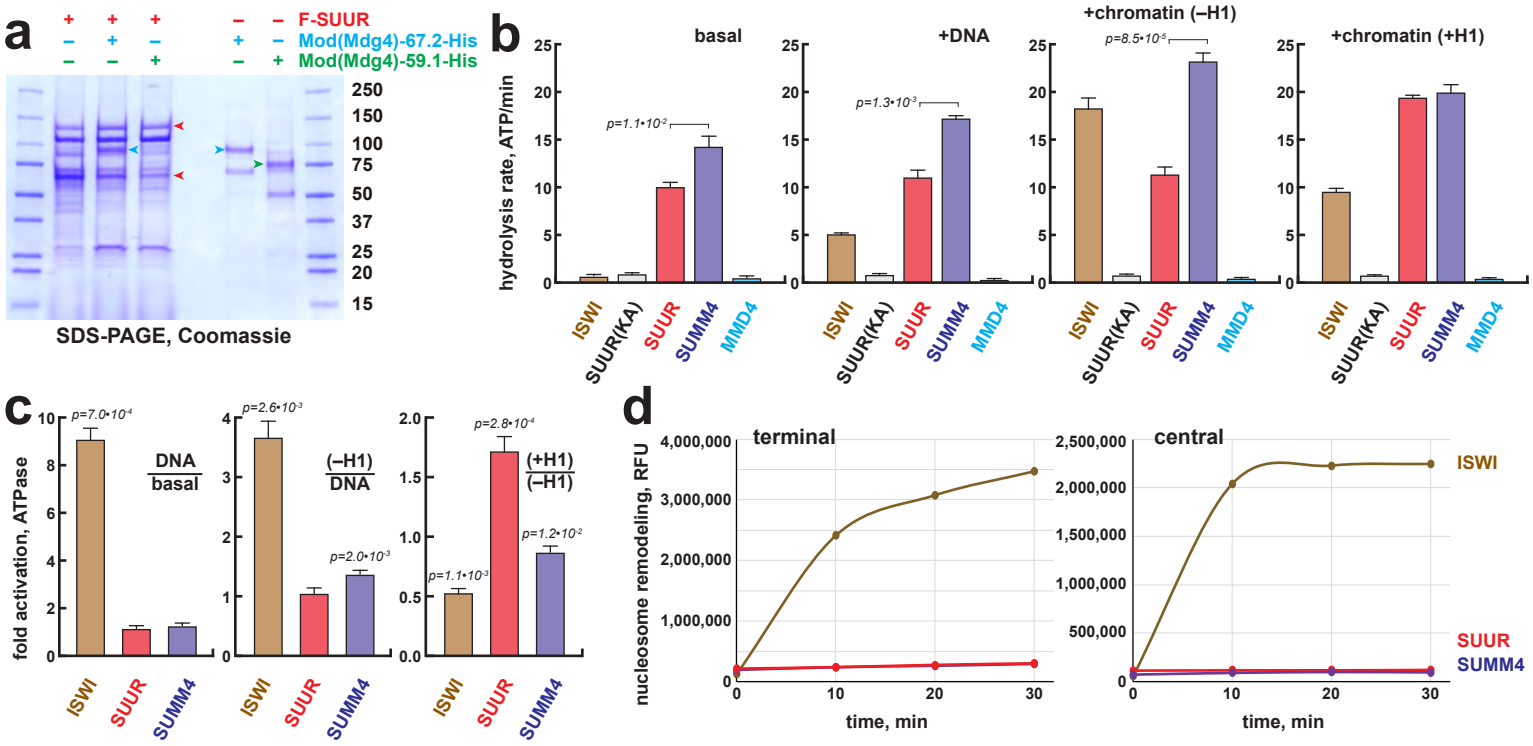
k **Mod(Mdg4)-PI (59.1)**

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
```

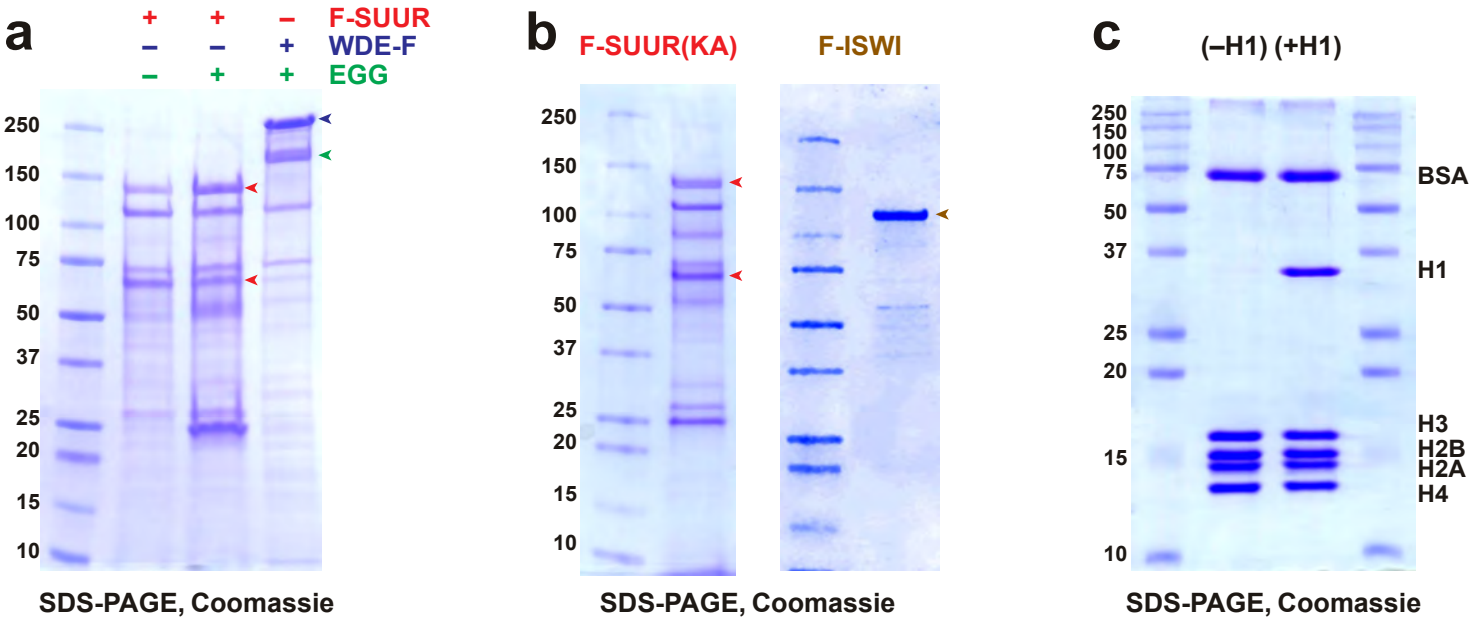
bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

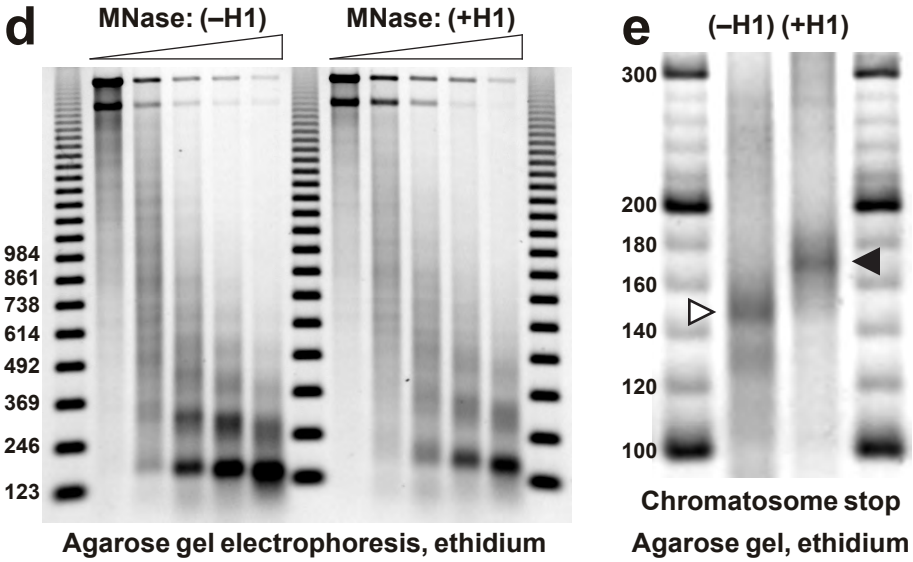
Fig. 2

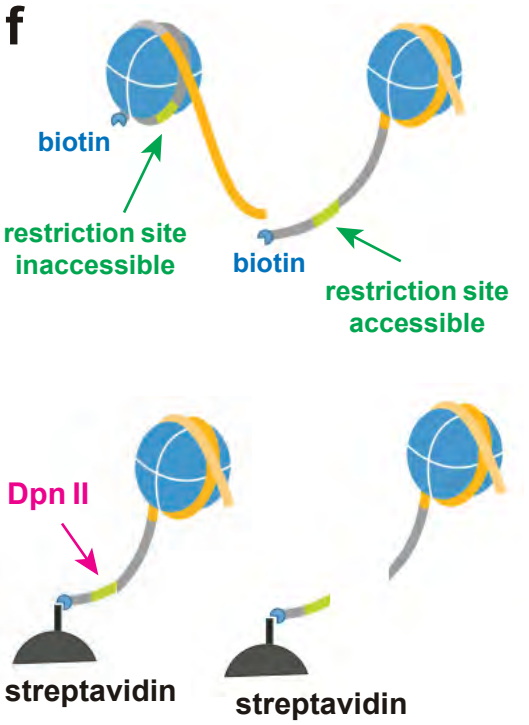


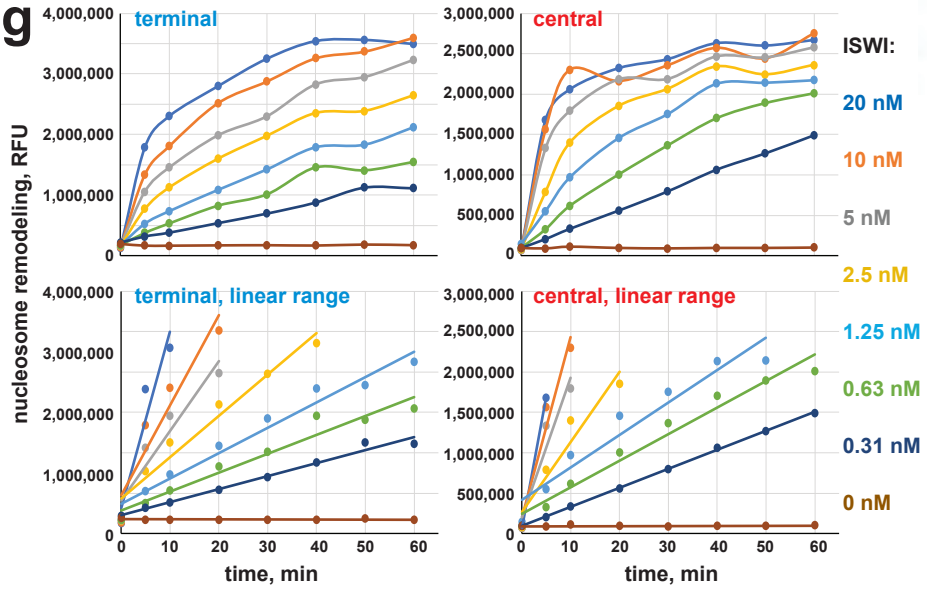
bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

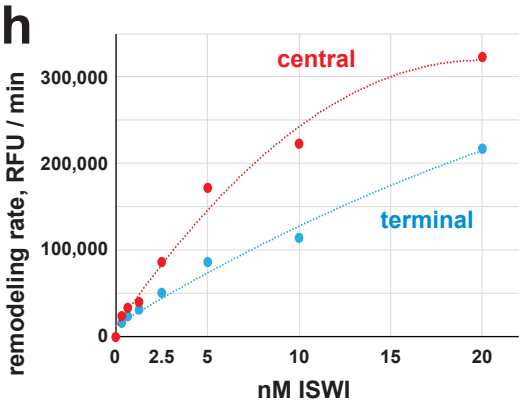
Fig. E2





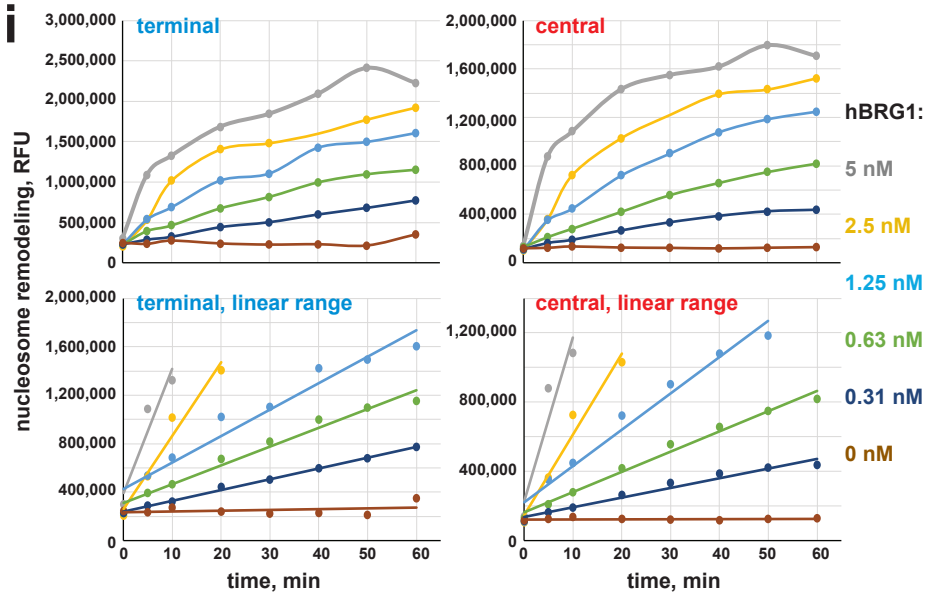


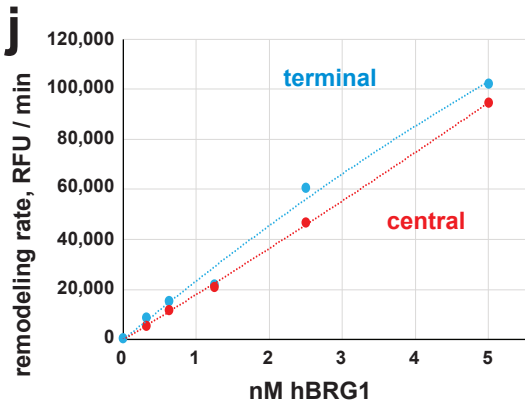




bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

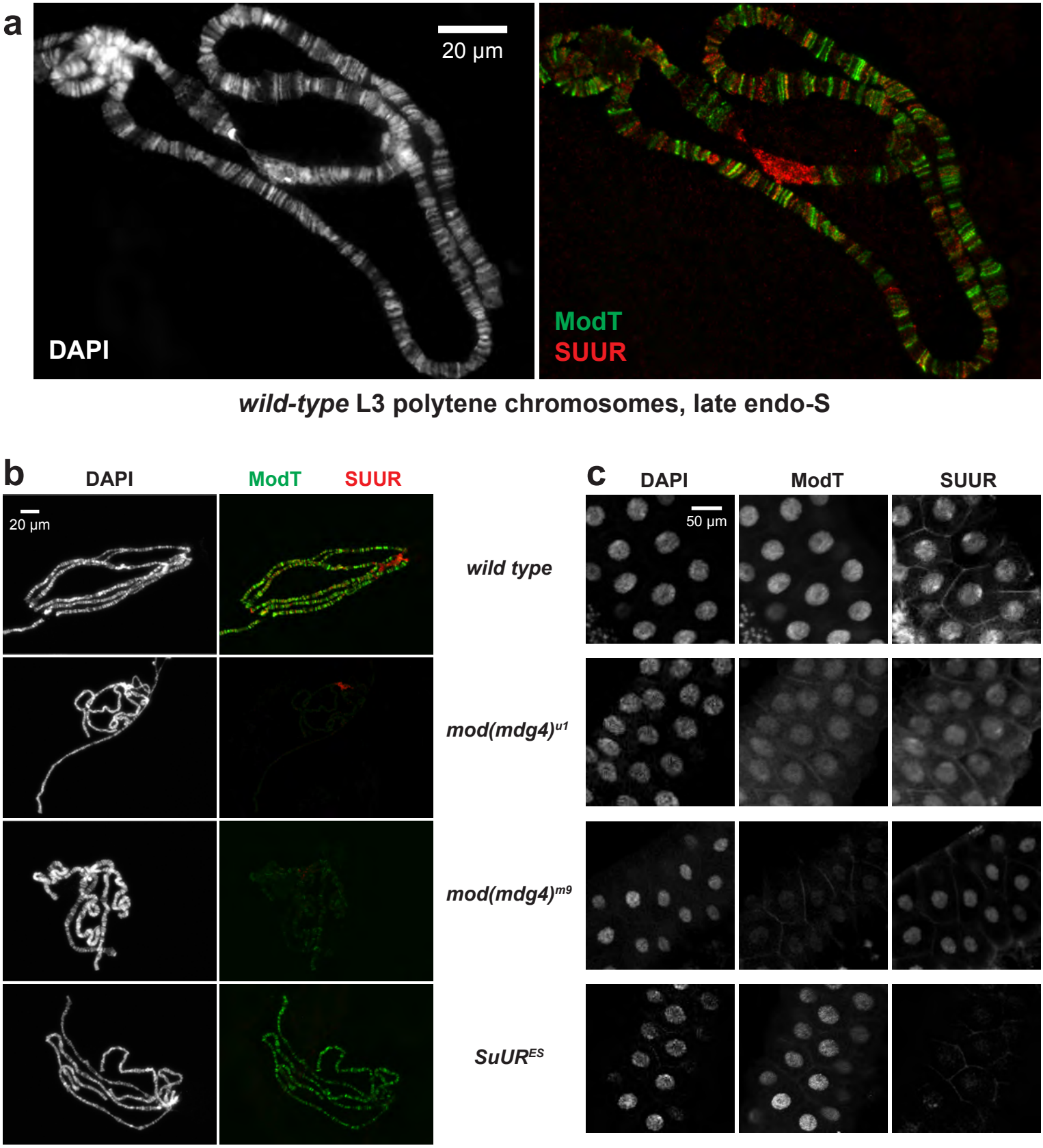
Fig. E2





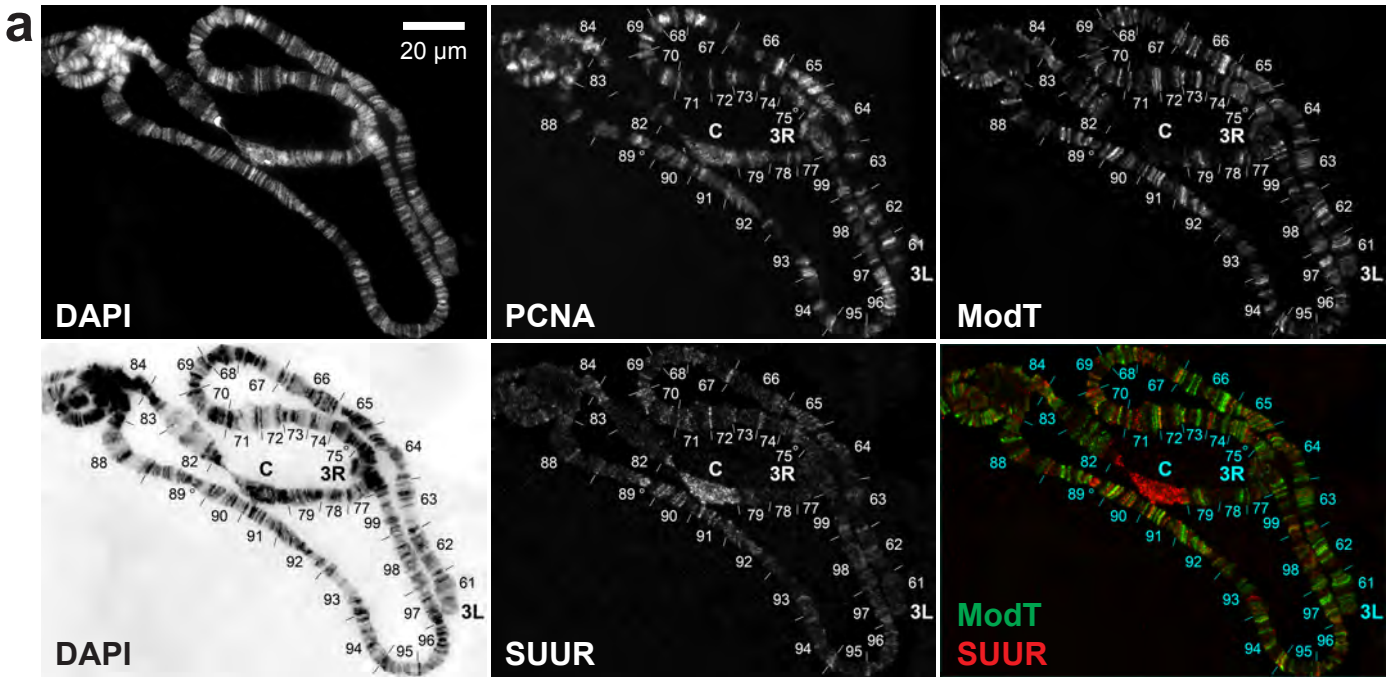
bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

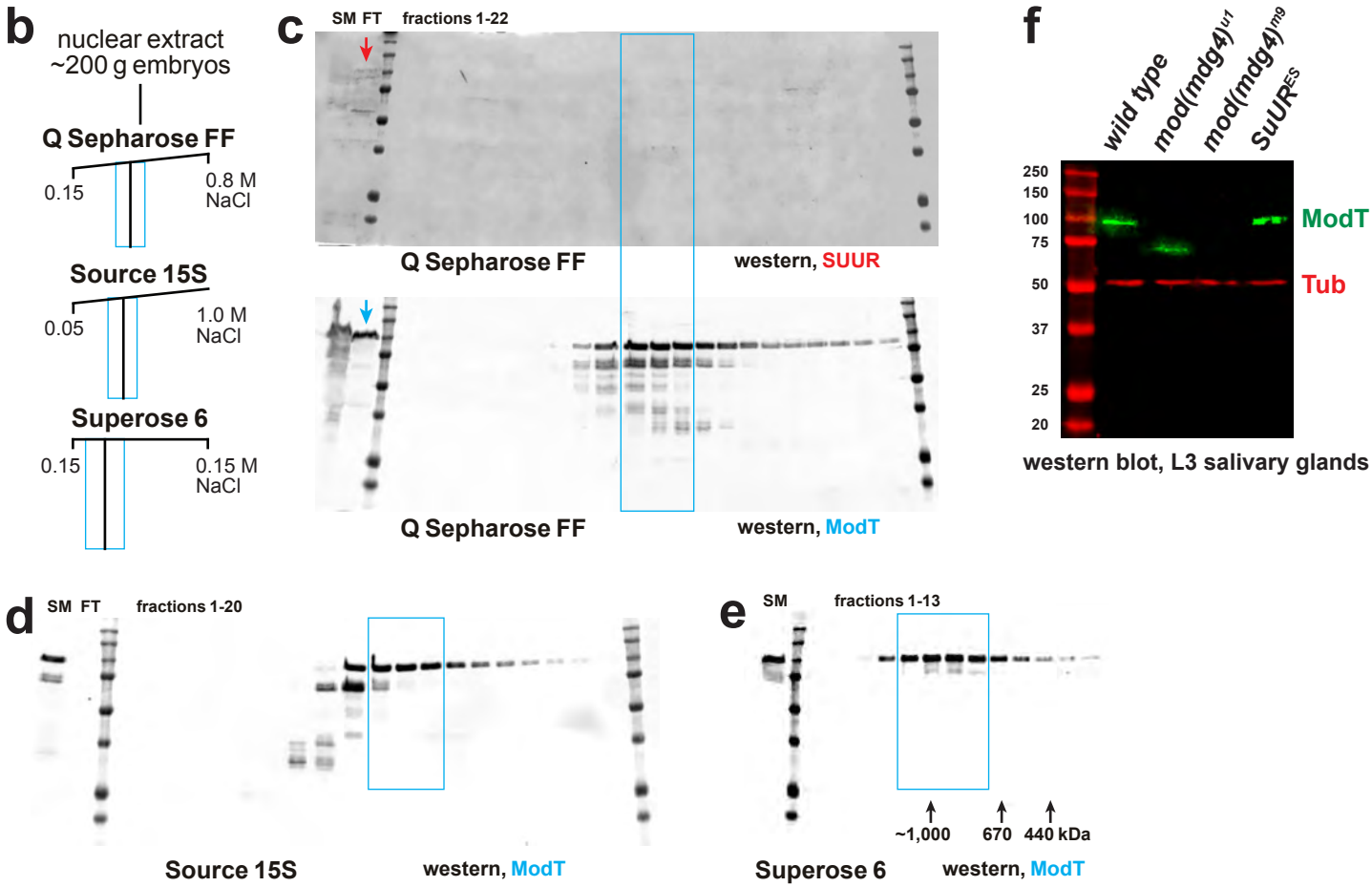
Fig. 3



bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

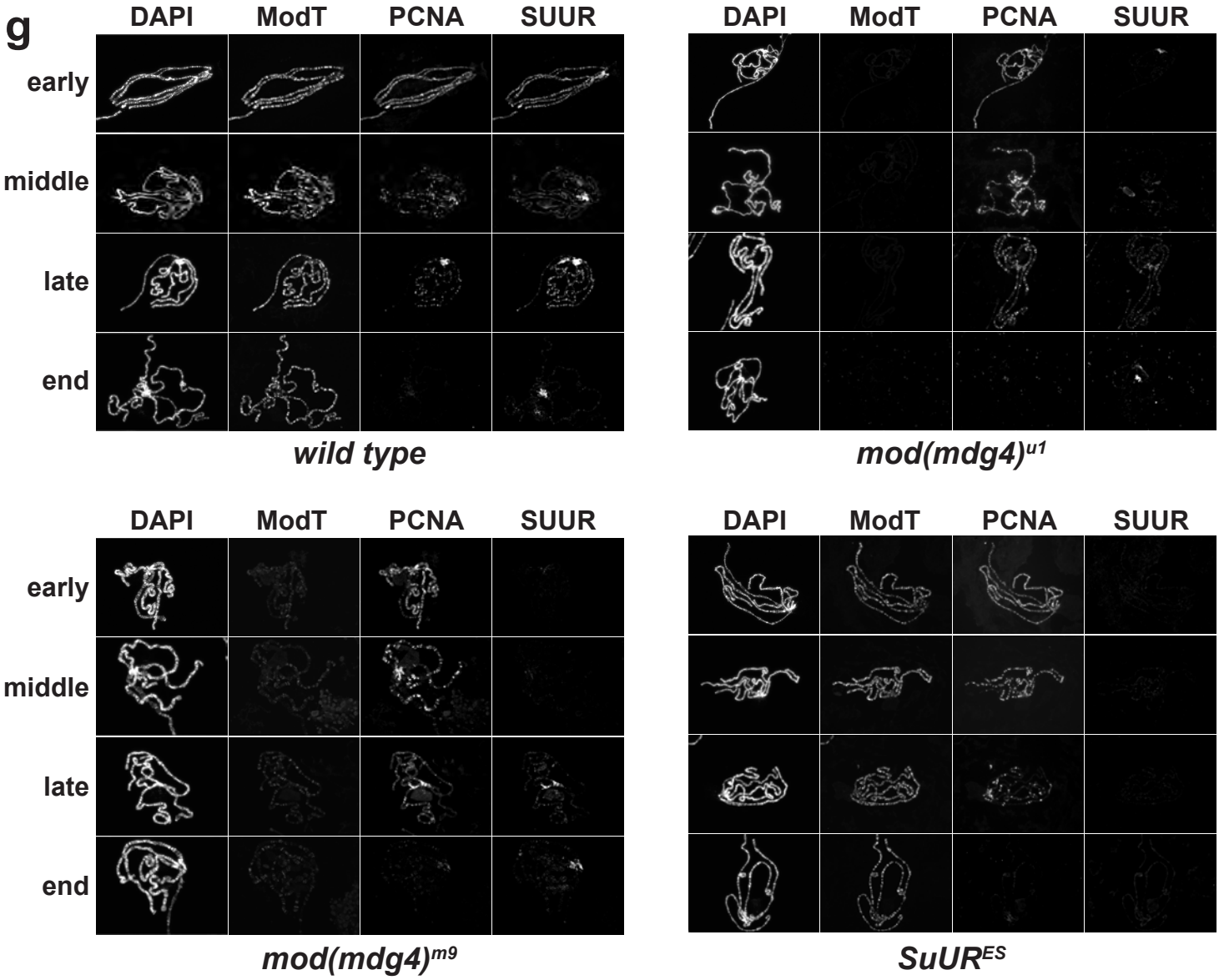
Fig. E3





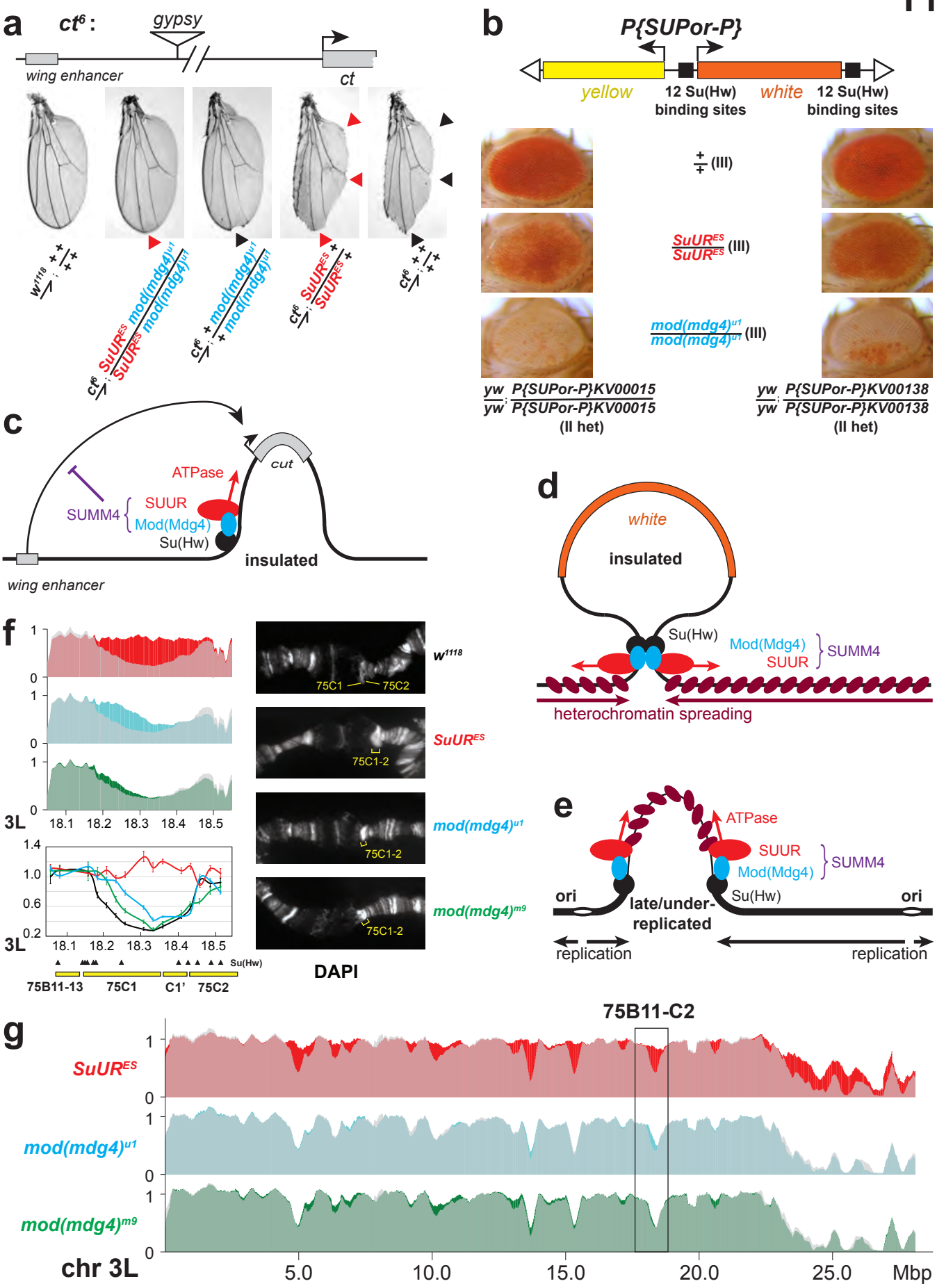
bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fig. E3



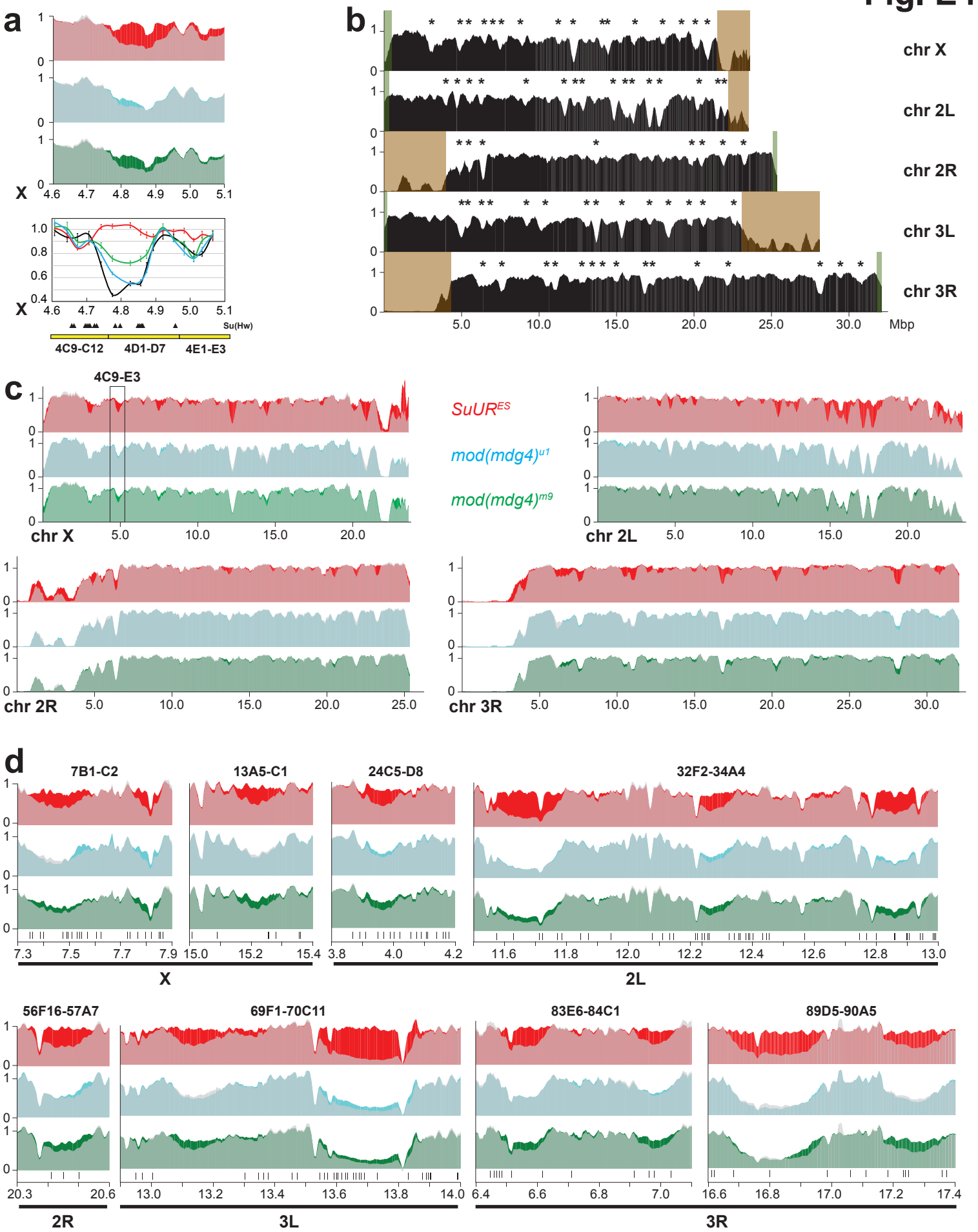
bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fig. 4



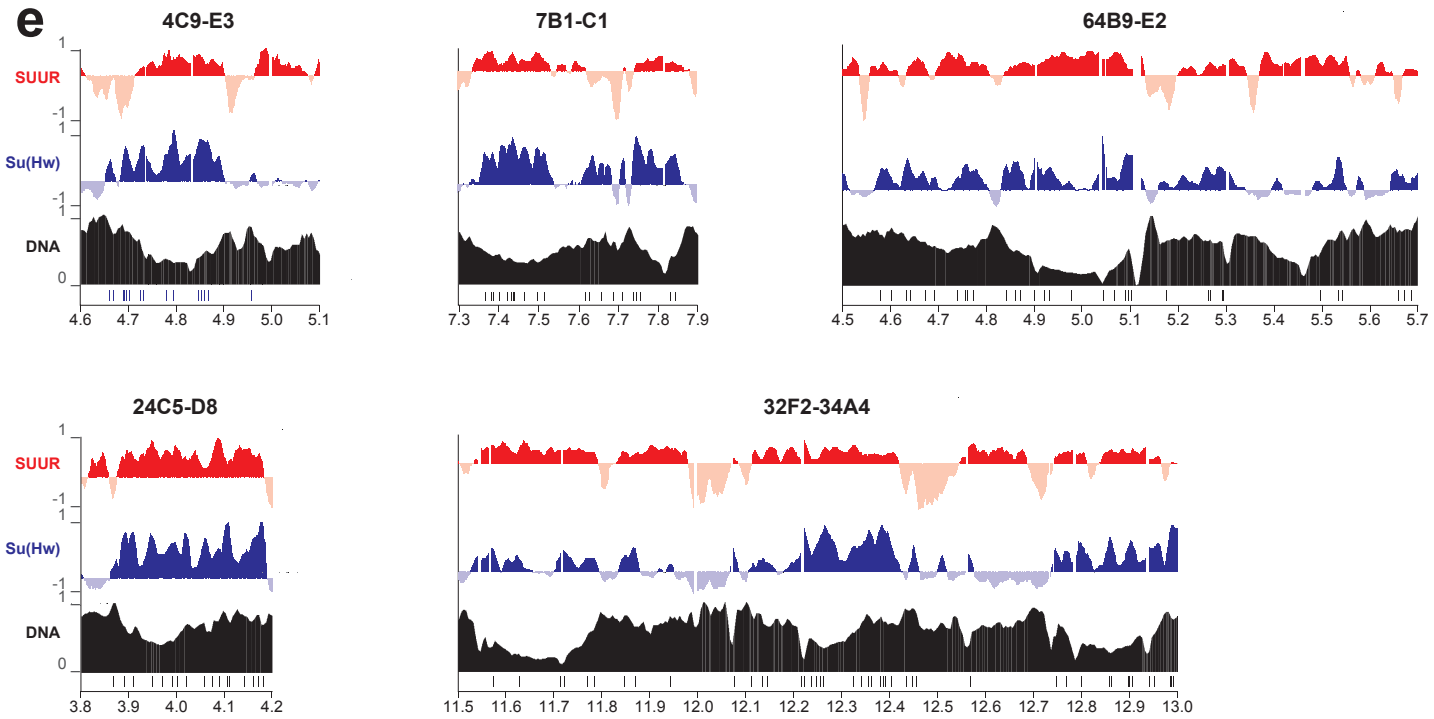
bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fig. E4



bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fig. E4



Extended Data Table E1. UR domains and UR suppression in SUMM4 subunit mutant alleles. Domains of under-replication (UR) in euchromatic arms of polytene chromosomes were called in w^{1118} as described in *Methods*. Their genomic coordinates, approximate cytological location ("Cyto band") and average DNA copy numbers ("<CN>") in homozygous w^{1118}, $SuUR^{ES}$, $mod(mdg4)^{m9}$ and $mod(mdg4)^{u1}$ L3 larvae are shown. <CN> numbers were normalized to the average DNA copy numbers across euchromatic genome. Percent UR recovery levels were calculated as $(\langle CN\rangle_{mut} - \langle CN\rangle_{w}) / (1 - \langle CN\rangle_{w})$. UR was called as suppressible by a mutant if Recovery > 12.5%. Averages of <CN> across all called UR domains and averages of percent Recovery across all suppressible UR domains ("<Recovery>", bottom row) were adjusted for each UR domain length. Calculation errors = standard deviations.

N	chromosome coordinates				Length	UR, w^{1118}	UR, $SuUR^{ES}$			UR, $mod(mdg4)^{m9}$			UR, $mod(mdg4)^{u1}$		
	arm	left	right	Cyto band		<CN>	<CN>	Recovery	suppressed	<CN>	Recovery	suppressed	<CN>	Recovery	suppressed
1	X	2,950,001	3,140,000	3C3-C7	190,000	0.51	0.93	86%	**YES**	0.58	14%	**YES**	0.52	0%	
2	X	4,710,001	4,900,000	4C15-D5	190,000	0.56	0.96	92%	**YES**	0.81	57%	**YES**	0.66	24%	**YES**
3	X	4,965,001	5,070,000	4E1-E2	105,000	0.72	0.86	50%	**YES**	0.80	28%	**YES**	0.77	18%	**YES**
4	X	6,415,001	6,525,000	6A1-B1	110,000	0.71	0.90	65%	**YES**	0.80	29%	**YES**	0.79	27%	**YES**
5	X	7,335,001	7,560,000	7B1-B4	225,000	0.65	0.98	95%	**YES**	0.79	40%	**YES**	0.68	8%	
6	X	7,750,001	7,865,000	7B7-C1	115,000	0.64	0.94	84%	**YES**	0.84	55%	**YES**	0.86	61%	**YES**
7	X	8,880,001	9,005,000	8B5-C2	125,000	0.73	0.86	50%	**YES**	0.76	9%		0.74	1%	
8	X	9,405,001	9,555,000	8D12-E7	150,000	0.72	0.91	67%	**YES**	0.85	47%	**YES**	0.61	-42%	
9	X	11,170,001	11,325,000	10A10-B3	155,000	0.67	0.84	53%	**YES**	0.78	35%	**YES**	0.71	13%	**YES**
10	X	12,040,001	12,430,000	11A2-A10	390,000	0.38	0.97	94%	**YES**	0.42	6%		0.30	-13%	
11	X	13,950,001	14,100,000	12D1-E1	150,000	0.69	0.72	10%		0.73	14%	**YES**	0.57	-37%	
12	X	14,290,001	14,565,000	12E7-F1	275,000	0.51	0.94	87%	**YES**	0.69	36%	**YES**	0.41	-21%	
13	X	17,925,001	18,030,000	16F3-F5	105,000	0.67	0.99	98%	**YES**	0.90	68%	**YES**	0.77	29%	**YES**
14	X	20,000,001	20,105,000	19A4-B1	105,000	0.79	1.12	157%	**YES**	0.82	12%		0.93	65%	**YES**
15	X	20,525,001	21,020,000	19D2-E7	495,000	0.50	0.97	93%	**YES**	0.51	2%		0.38	-25%	
16	X	21,630,001	22,450,000	20A5-C1	820,000	0.04	0.32	29%	**YES**	0.06	2%		0.05	1%	
17	X	22,550,001	22,995,000	20C2-F3	445,000	0.48	0.81	64%	**YES**	0.74	51%	**YES**	0.52	8%	
18	2L	3,920,001	4,025,000	24D1-D4	105,000	0.63	0.93	81%	**YES**	0.80	46%	**YES**	0.73	29%	**YES**
19	2L	4,585,001	4,790,000	25A2-A5	205,000	0.66	0.99	98%	**YES**	0.78	36%	**YES**	0.54	-33%	
20	2L	5,400,001	5,510,000	25E1-E4	110,000	0.82	0.99	95%	**YES**	0.90	45%	**YES**	0.86	24%	**YES**
21	2L	6,155,001	6,320,000	26B9-C2	165,000	0.74	1.08	130%	**YES**	0.88	54%	**YES**	0.75	4%	
22	2L	9,030,001	9,150,000	29F8-30A2	120,000	0.76	0.98	93%	**YES**	0.95	79%	**YES**	0.77	5%	
23	2L	11,535,001	11,795,000	32F2-33A1	260,000	0.44	0.90	83%	**YES**	0.57	24%	**YES**	0.45	2%	
24	2L	12,215,001	12,340,000	33D3-E1	125,000	0.58	0.86	66%	**YES**	0.75	40%	**YES**	0.71	32%	**YES**
25	2L	12,765,001	12,970,000	33F5-34A3	205,000	0.55	0.91	79%	**YES**	0.73	40%	**YES**	0.63	17%	**YES**
26	2L	14,685,001	15,010,000	35B4-B8	325,000	0.41	0.88	80%	**YES**	0.54	23%	**YES**	0.43	4%	
27	2L	15,295,001	15,735,000	35D1-D4	440,000	0.49	0.76	53%	**YES**	0.54	9%		0.44	-11%	
28	2L	15,770,001	15,900,000	35D4-D6	130,000	0.54	0.87	71%	**YES**	0.68	31%	**YES**	0.58	9%	
29	2L	15,925,001	16,240,000	35D6-F1	315,000	0.29	0.90	87%	**YES**	0.38	12%		0.31	2%	
30	2L	16,925,001	17,375,000	36B4-C7	450,000	0.23	0.89	85%	**YES**	0.26	4%		0.21	-3%	
31	2L	17,515,001	18,100,000	36C10-E4	585,000	0.34	0.87	80%	**YES**	0.36	2%		0.31	-5%	
32	2L	18,160,001	18,300,000	36E6-F2	140,000	0.67	0.99	97%	**YES**	0.90	69%	**YES**	0.76	28%	**YES**

33	2L	20,110,001	20,290,000	38C1-C4	180,000	0.48	0.69	41%	**YES**	0.46	-5%		0.47	-3%	
34	2L	20,485,001	20,620,000	38C8-D1	135,000	0.77	0.98	93%	**YES**	0.99	97%	**YES**	0.79	8%	
35	2L	21,400,001	21,550,000	39D3-E2	150,000	0.10	0.15	5%		0.14	3%		0.13	3%	
36	2L	21,805,001	22,125,000	40A4-E4	320,000	0.53	0.94	87%	**YES**	0.54	1%		0.35	-39%	
37	2R	4,875,001	5,050,000	41C4-D1	175,000	0.35	0.86	78%	**YES**	0.34	-1%		0.48	20%	**YES**
38	2R	5,410,001	5,535,000	41F1-F3	125,000	0.58	0.79	50%	**YES**	0.52	-13%		0.53	-12%	
39	2R	6,290,001	6,505,000	42A14-B1	215,000	0.13	0.50	42%	**YES**	0.14	1%		0.12	-1%	
40	2R	13,620,001	13,760,000	50B6-C3	140,000	0.63	0.95	88%	**YES**	0.78	41%	**YES**	0.61	-3%	
41	2R	20,355,001	20,540,000	56F17-57A5	185,000	0.56	0.92	83%	**YES**	0.71	35%	**YES**	0.60	9%	
42	2R	21,830,001	21,945,000	58A2-A4	115,000	0.72	0.95	83%	**YES**	0.71	-3%		0.55	-59%	
43	2R	23,145,001	23,320,000	59D1-D6	175,000	0.62	1.04	110%	**YES**	0.67	13%	**YES**	0.60	-5%	
44	3L	4,840,001	5,100,000	64C1-C5	260,000	0.38	0.92	87%	**YES**	0.40	3%		0.41	5%	
45	3L	5,385,001	5,510,000	64C15-D3	125,000	0.51	0.88	76%	**YES**	0.73	45%	**YES**	0.60	18%	**YES**
46	3L	6,290,001	6,485,000	65A11-B3	195,000	0.52	0.89	77%	**YES**	0.71	38%	**YES**	0.48	-8%	
47	3L	9,180,001	9,300,000	67A1-A7	120,000	0.67	0.97	90%	**YES**	0.73	20%	**YES**	0.47	-59%	
48	3L	10,000,001	10,195,000	67D3-D10	195,000	0.62	0.97	93%	**YES**	0.79	44%	**YES**	0.59	-8%	
49	3L	13,085,001	13,220,000	70A1-A2	135,000	0.66	1.01	104%	**YES**	0.89	66%	**YES**	0.58	-23%	
50	3L	13,550,001	13,855,000	70B6-C4	305,000	0.26	0.95	94%	**YES**	0.39	18%	**YES**	0.40	19%	**YES**
51	3L	15,175,001	15,500,000	71B7-D3	325,000	0.39	0.94	89%	**YES**	0.46	10%		0.35	-8%	
52	3L	17,115,001	17,240,000	73F1-74A1	125,000	0.71	1.02	106%	**YES**	0.84	45%	**YES**	0.77	20%	**YES**
53	3L	18,175,001	18,525,000	75B11-75D2	350,000	0.45	0.87	76%	**YES**	0.47	4%		0.63	32%	**YES**
54	3L	20,555,001	20,695,000	77D1-77E3	140,000	0.60	1.02	106%	**YES**	0.84	61%	**YES**	0.67	18%	**YES**
55	3R	6,060,001	6,310,000	83D2-E4	250,000	0.70	0.92	72%	**YES**	0.63	-22%		0.43	-91%	
56	3R	6,495,001	6,635,000	83F1-84A1	140,000	0.53	0.96	91%	**YES**	0.71	39%	**YES**	0.56	6%	
57	3R	6,915,001	7,055,000	84B1-B2	140,000	0.64	0.93	80%	**YES**	0.82	49%	**YES**	0.70	16%	**YES**
58	3R	7,550,001	7,785,000	84D9-84E2	235,000	0.44	0.80	65%	**YES**	0.51	12%		0.50	11%	
59	3R	10,450,001	10,660,000	86B6-C4	210,000	0.55	0.98	97%	**YES**	0.66	25%	**YES**	0.51	-10%	
60	3R	10,910,001	11,140,000	88C15-86D4	230,000	0.45	0.94	89%	**YES**	0.46	2%		0.37	-14%	
61	3R	12,050,001	12,165,000	87A5-B1	115,000	0.63	0.96	88%	**YES**	0.81	49%	**YES**	0.67	10%	
62	3R	12,745,001	12,935,000	87C8-D4	190,000	0.67	0.89	68%	**YES**	0.60	-21%		0.60	-22%	
63	3R	14,935,001	15,055,000	88D8-D10	120,000	0.70	0.88	61%	**YES**	0.84	47%	**YES**	0.75	17%	**YES**
64	3R	16,670,001	16,970,000	89D6-E5	300,000	0.40	0.92	87%	**YES**	0.47	10%		0.37	-5%	
65	3R	17,160,001	17,355,000	89F1-90A2	195,000	0.62	0.94	84%	**YES**	0.86	64%	**YES**	0.57	-12%	
66	3R	20,085,001	20,290,000	92C4-E1	205,000	0.61	0.81	53%	**YES**	0.71	26%	**YES**	0.62	3%	
67	3R	20,340,001	20,525,000	92E4-E12	185,000	0.58	0.96	91%	**YES**	0.79	50%	**YES**	0.70	29%	**YES**
68	3R	22,110,001	22,295,000	94A2-A4	185,000	0.61	0.93	83%	**YES**	0.76	39%	**YES**	0.67	16%	**YES**
69	3R	28,005,001	28,295,000	98B7-C3	290,000	0.40	0.91	85%	**YES**	0.60	32%	**YES**	0.53	21%	**YES**
70	3R	28,370,001	28,480,000	98C5-D2	110,000	0.73	0.98	94%	**YES**	0.91	66%	**YES**	0.79	22%	**YES**
UR domains: **70** <Length>: 216 ± 64 kbp Average <CN> across all UR domains: 0.49 ± 0.08							Suppressed UR domains: **68** <Length>: 218 ± 64 kbp <Recovery>: 79 ± 10%			Suppressed UR domains: **46** <Length>: 173 ± 35 kbp <Recovery>: 41 ± 9%			Suppressed UR domains: **24** <Length>: 159 ± 34 kbp <Recovery>: 25 ± 6%		

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Extended Data Table E2. Primer sequences used for qPCR. Genomic coordinates indicate full 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

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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, Y. Schwartz, S. Sidoli and I. Zhimulev for helpful discussions and critical reading of the manuscript. We thank Y. Schwartz 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). This research was supported by the National Institutes of Health (NIH) grants R01 GM074233 to D.V.F., R01 GM129244 to A.I.S, R01 GM124201 to R.J.D. and R44 GM123869 to M.C.K. (EpiCypher). M.N. was supported by postdoctoral training grants K12 GM000678 and T32 CA217824. LCMS instrument was purchased with an equipment supplement from the National Institute of General Medical Sciences (NIGMS) to D.V.F. and A.I.S. We thank P. Schultes for help with maintaining the instrument.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Author Information

†These authors contributed equally: Evgeniya N. Andreyeva, Alexander V. Emelyanov

Affiliations

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

Evgeniya N. Andreyeva, Alexander V. Emelyanov, Elena Vershilova, Arthur I. Skoultchi & Dmitry V. Fyodorov

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

Markus Nevil

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

Lu Sun & Michael C. Keogh

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

Christina A. Hill & Robert J. Duronio

[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**

Robert J. Duronio

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Contributions

E.N.A. designed and performed quantitative proteomic, genetic and cytological experiments, prepared DNA for NGS and analyzed the data. A.V.E. purified native proteins by FPLC, cloned, expressed and purified recombinant proteins, designed and performed biochemical assays *in vitro*, performed real-time PCR analyses and analyzed the data. M.N. and C.H. prepared DNA libraries for NGS and analyzed the data. L.S. performed biochemical assays with recombinant proteins *in vitro* and analyzed the data. E.V. prepared nuclear extracts and performed western blot analyses. M.C.K. supervised L.S., analyzed and interpreted the data. R.J. D. supervised M.N. and C.H., analyzed and interpreted the data and co-wrote the manuscript (with D.V.F.). A.I.S. supervised E.N.A., analyzed and interpreted the data and co-wrote the manuscript (with D.V.F.). D.V.F. conceptualized the study, supervised E.N.A. and A.V.E., analyzed and interpreted the data and wrote the manuscript. All authors contributed to editing of the final version of the manuscript.

Corresponding authors

*Correspondence to: arthur.skoultchi@einsteinmed.org & dmitry.fyodorov@einsteinmed.org

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Ethics declarations

Competing interests

L.S. and M.C.K. 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.

bioRxiv preprint doi: https://doi.org/10.1101/2021.10.02.462895; this version posted October 3, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary information

Supplementary Table S1.xlsx. Identities of proteins and peptide spectra from ion libraries used for MERCI quantification of SUMM4.

Supplementary Table S2.xlxs. Raw protein area data from SWATH acqusitions quantified by IL5 and MERCI correlation analyses.