

## **Repression of interrupted and intact rDNA by the SUMO pathway in *Drosophila melanogaster***

Yicheng Luo<sup>1</sup>, Elena Fefelova<sup>1,2</sup>, Maria Ninova<sup>1</sup>, Yung-Chia Ariel Chen<sup>1</sup> and  
Alexei A. Aravin<sup>1&</sup>

<sup>1</sup>California Institute of Technology,  
Division of Biology and Biological Engineering  
Pasadena, CA 91125, USA

<sup>2</sup>Institute of Molecular Genetics,  
Russian Academy of Sciences,  
Moscow 123182, Russia

&To whom correspondence should be addressed:  
Alexei Aravin [aaa@caltech.edu](mailto:aaa@caltech.edu)

## Abstract

Ribosomal RNAs (rRNAs) are essential components of the ribosome and are among the most abundant macromolecules in the cell. To ensure high rRNA level, eukaryotic genomes contain dozens to hundreds of rDNA genes, however, only a fraction of the rRNA genes seems to be active, while others are transcriptionally silent. We found that individual rDNA genes have high level of cell-to-cell heterogeneity in their expression in *Drosophila melanogaster*. Insertion of heterologous sequences into rDNA leads to repression associated with reduced expression in individual cells and decreased number of cells expressing rDNA with insertions. We found that SUMO (Small Ubiquitin-like Modifier) and SUMO ligase Ubc9 are required for efficient repression of interrupted rDNA units and variable expression of intact rDNA. Disruption of the SUMO pathway abolishes discrimination of interrupted and intact rDNAs and removes cell-to-cell heterogeneity leading to uniformly high expression of individual rDNA in single cells. Our results suggest that the SUMO pathway is responsible for both repression of interrupted units and control of intact rDNA expression.

## Introduction

Ribosomal RNAs (rRNAs) are the main structural and central enzymatic components of ribosomes. The constant production of ribosomal RNA is essential for cell growth and division. The high demand for rRNA production is addressed by a two-prong strategy. First, cells use a dedicated transcription machinery composed of RNA polymerase I and associated unique transcription factors to transcribe rDNA genes. Second, genomes contain multiple identical rDNA genes that are transcribed simultaneously.

In most eukaryotes rRNA genes encoding the 18S, 28S and 5.8S rRNAs are transcribed under the control of a single promoter producing pre-rRNA (Mandal, 1984). After co-transcriptional processing and modification, the pre-RNA is converted to mature 18S, 28S and 5.8S rRNAs which together with the additional 5S rRNA and ribosomal proteins are assembled into pre-ribosomal subunits (Henras et al., 2015). Actively transcribed rDNA genes and pre-ribosomal subunits at different steps of maturity form the nucleolus – a membraneless compartment in the nucleus characterized by tripartite morphological ultrastructure.

In eukaryotic genomes dozens to hundreds of nearly identical rDNA units are arranged head-to-tail forming one or several clusters. Studies in such diverse organisms as different plants, yeast, fruit flies, mice and humans showed that only fraction of available rDNA genes is transcriptionally active (Bird et al., 1981; Coffman et al., 2005; Conconi et al., 1989; Dammann et al., 1993; Flavell et al., 1988). In organisms that have several rDNA clusters located on different chromosomes, whole clusters might be transcriptionally inactive and positioned outside of the nucleolus. For example, in *Drosophila*, where rDNA clusters are located on the X and Y chromosomes, in certain genotypes only the Y chromosome cluster is active and forms the nucleolus (Greil and Ahmad, 2012; Zhou et al., 2012). In mammals which carry several ribosomal loci on different chromosomes some clusters are constantly active, while the activity of others depends on the tissue type, presence of nutrients and stress (Ali et al., 2008, 2012; de Capoa et al., 1985a; de Capoa et al., 1985b; Tseng et al., 2008; Young et al.,

2007; Zhang et al., 2007). Direct observation of rDNA transcription by electron microscopy showed that individual rDNA units within a single cluster might also have different transcriptional activity and are subject to all-or-none regulation: any given rDNA unit is either actively transcribed with multiple RNA pol I associated with nascent pre-rRNA positioned along the body of the unit, or is inactive. (Foe, 1978; McKnight et al., 1976).

Due to recombination-driven homogenization, individual rDNA units are similar in sequence (Eickbush et al., 2007), arguing against difference in sequence leading to differential expression. In fact, repressed rDNA units might become active when the genetic environment or demand for rRNA production changes. For example, in *Drosophila* X-linked rDNA genes that are inactive in males are active in females (Zhou et al., 2012). The molecular mechanism for differential activity of rDNA units was studied in several organisms. In mammals repressed rDNA units are enriched in DNA methylation, repressive histone marks (Coffman et al., 2005; Earley et al., 2006; Li et al., 2006; Santoro et al., 2001; Santoro et al., 2002; Zhou et al., 2002). The chromatin remodeling complexes NoRC and NuRD, in cooperation with noncoding RNAs derived from rDNA loci were shown to be involved in rDNA repression (Santoro et al., 2002; Strohner et al., 2001; Xie et al., 2012; Zhou et al., 2002; Bierhoff et al., 2014; Mayer et al., 2006; Santoro et al., 2010). However, a deep understanding of the molecular mechanisms establishing and maintaining rDNA repression is lacking.

Genomes of many arthropod species including *Drosophila melanogaster* harbor transposable elements that integrate into rDNA, creating distinct rDNA units. R1 and R2 belong to the non-long terminal repeat (non-LTR) retrotransposons and encode a sequence-specific endonuclease that is responsible for integration of these elements into 28S rDNA (Burke et al., 1987; Eickbush et al., 2015; Jakubczak et al., 1990; Xiong et al., 1988; Yang et al., 1999). It is believed that R1 and R2 lack their own promoters and instead rely on the rDNA transcription machinery for their expression by transcribing as part of pre-rRNA followed by excision mediated by the transposon-encoded ribozyme (Eickbush et al., 2010; Eickbush et al., 2013).

The fraction of rDNA units with R1 and R2 insertions varies between different *Drosophila melanogaster* strains, but was estimated to reach up to 80% of units in some strains (Jakubczak et al., 1992). Despite the abundance of R1 and R2, their expression level is usually low. Indeed, electron microscopy and run-on analysis of nascent transcripts revealed that rDNA units with transposon insertions are transcriptionally silent, indicating a mechanism that can distinguish damaged rDNA copies and repress their transcription (Jamrich et al., 1984; Ye and Eickbush, 2006). Interestingly, R1 and R2 expression is increased if the total number of rDNA units in the genome is low, suggesting that repression is sensitive to cellular rRNA demand (Eickbush et al., 2003; Long et al., 1981; Terracol, 1986). The molecular mechanism for repression of interrupted rDNA units and its relationship to the general mechanism of rDNA silencing remained unknown.

SUMO (Small Ubiquitin-like Modifier) is a small protein related to ubiquitin that is covalently attached to other proteins in sequential reactions mediated by E1, E2 and, for some targets, E3 SUMO ligases. Unlike ubiquitination, SUMOylation typically is not linked to protein degradation and instead changes protein-protein interactions by facilitating recruitment of new binding partners or masking existing binding sites. The majority of SUMOylated proteins reside in the nucleus and SUMOylation was shown to be involved in regulation of transcription, mRNA processing, chromatin organization, replication and repair (Geiss-Friedlander et al., 2007) as well as rRNA maturation and nucleolus function (Finkbeiner et al., 2011; Haindl et al., 2008; Yun et al., 2008; Westman et al., 2010). SUMOylation was shown to be essential for growth and viability of *Saccharomyces cerevisiae* (Johnson et al., 1997), *Drosophila melanogaster* (Lehembre et al., 2000) and mice (Nacerddine et al., 2005). Mammalian genomes encode four distinct SUMO genes with partially redundant functions. In contrast, only one SUMO gene is present in the *Drosophila melanogaster* genome, making it a good model to understand the diverse functions of SUMOylation (Guo et al., 2004; Melchior, 2000; Yang et al., 1999).

Here we have established a new model to study regulation of rDNA expression using a molecularly marked single-unit rDNA transgene. We found that this model

faithfully recapitulates repression of interrupted rDNA units. We discovered that repression of rDNA units damaged by transposon insertions, as well as intact rDNA units is controlled by the SUMO pathway, indicating that the same molecular mechanism is responsible for epigenetic inactivation of intact and interrupted rDNA.

## Results

### **A single-unit rDNA transgene is expressed from a heterologous genomic site**

The *D. melanogaster* genome contains numerous (50-200) rDNA units arranged in tandem repeats in the heterochromatin of the X and Y chromosomes (Figure 1A). Many rDNA units contain insertions of R1 and R2 transposons that lack their own promoters and co-transcribe with pre-rRNA (Ye and Eickbush, 2006). However, expression of R1 and R2 elements is low in the majority of *Drosophila* strains, indicating that units with insertions are specifically repressed (Ye and Eickbush, 2006). The large number of identical rDNA units organized in large arrays makes it impossible to study expression of individual rDNA units by conventional methods. To circumvent this problem and study regulation of rDNA expression, we created flies with a transgene that carries a molecularly marked single rDNA unit inserted into a heterologous genomic locus (Figure 1A). It harbors the non-transcribed spacer (NTS), which includes elements that regulate RNA polymerase I (Pol I) transcription and the complete transcribed portion, which generates the 47S pre-rRNA. The pre-rRNA contains a 5' external transcribed sequence (5' ETS) and internal transcribed spacers (ITS), which are removed from the pre-rRNA during its processing into mature 18S, 5.8S, 2S and 28S rRNAs. To discriminate transcripts generated from the transgene from endogenous rRNA we inserted a 21bp unique identification sequence (UID) into the external transcribed spacer (ETS) downstream of the transcription start site (Figure 1A). Along with the rest of the ETS, the UID sequence is removed from pre-rRNA during its processing to mature rRNAs in the nucleolus, however, it allowed us to monitor transgene

transcription. In contrast to endogenous rDNA arrays, which are located in heterochromatin of the X and Y chromosomes, the single-unit rDNA transgene was integrated into a euchromatic site on 2nd chromosome (chr 2L: 1,582,820) using site-specific integration.

First, we tested if the rDNA transgene is expressed in the heterologous location. RT-qPCR with one primer specific to UID sequence detected expression of the rDNA transgene in transgenic flies, while no PCR product was produced in the parental strain used for transgenesis (Figure 1B). Thus, RT-qPCR is able to discriminate expression of the marked rDNA transgene from endogenous rDNA and insertion of UID sequence does not disrupt the Pol I promoter and enhancer elements.

Next, we employed fluorescent *in situ* hybridization (FISH) to detect expression of the rDNA transgene in individual cells. Nascent transgene transcripts differ from more abundant endogenous pre-rRNA by a 21 nt UID sequence, however, the standard FISH protocol using short probe against UID failed to detect expression of the transgene. To circumvent low sensitivity of standard FISH, we used a method based on the mechanism of the hybridization chain reaction (HCR-FISH) that offers a combination of high sensitivity and quantitation (Choi et al., 2018). HCR-FISH allowed specific detection of nascent transgene transcripts and revealed that transgene RNA is present exclusively in the nucleus (Figure 1C) consistent with the ETS portion being processed out of pre-rRNA and degraded soon after transcription. Thus, the marked rDNA transgene enables expression analysis of an individual rDNA unit inserted in a heterologous genomic location.

### **Insertion of a heterologous sequences into rDNA leads to decreased expression**

In *Drosophila* significant fraction of rDNA units contain insertions of R1 and R2 retrotransposons, however, units damaged by transposon insertions are usually silent (Long and Dawid, 1979; Jamrich et al., 1984). R1 and R2 each integrate into single unique site positioned in 28S RNA at 2711 nt and 2648 nt of 28S (Jakubczak

et al., 1990), respectively. To understand how insertion of transposons into rDNA influences its expression, we generated a series of additional transgenes. The full-length R1 (5.3 Kb) and R2 (3.6 Kb) transposon sequences were inserted in their respective integration sites within the 28S rDNA (Figure 2A). We also generated transgenes that contain insertion of an unrelated sequence (promoterless CFP) in the same positions as R1 and R2 (Figure 2A). In R1' and R2' transgenes, respective transposon sequences were flanked by additional UIDs. All transgenes were integrated into the same genomic location, which allows us to directly compare their expression.

Expression analysis of transgenes by RT-qPCR in ovary and carcass (Figure 2B) showed that transposon insertions lead to a 3- to 5-fold reduction in pre-rRNA level compared to the intact transgene. R1 insertion caused a slightly stronger reduction in expression compared to R2 insertion. Importantly, insertions of a non-transposon CFP sequence lead to comparable decrease in expression, indicating that the repressive effect is not dependent on a specific transposon sequence and can be triggered by heterologous sequence. A 29 bp insertion caused a 2-fold decrease in transgene expression, demonstrating that short insertions can affect rDNA expression, though to a lesser degree. Thus, expression of the single-copy rDNA transgene recapitulates the behavior of endogenous rDNA and demonstrates repression of units damaged by transposon insertions.

### **Cell-to-cell variability in rDNA transgenes expression**

To monitor expression of rDNA transgenes in individual cells we employed HCR-FISH. Nurse cells of the fly ovary are an ideal cell type to study transgene expression as each egg chamber is composed of 15 identical sister nurse cells with large nuclei. The chromosomes of nurse cells forming one egg chamber undergo several rounds of endoreplication. Due to chromosome polytenization the HCR-FISH signals of rDNA transgenes and the endogenous control gene, *vasa*, are localized to several distinct loci on chromatin that likely represent sites of nascent transcription (Figure 2C). *Vasa* gene was expressed at similar levels in 15

nurse cells forming egg chamber. Expression of endogenous pre-rRNA detected by FISH probe against ETS (which detects rRNAs coming from all rDNA copies) was also similar in the different nuclei. In contrast, the rDNA transgenes have variable levels of expression in individual nuclei within one egg chamber with some nuclei showing no signal at all, while others had variable signal intensity. Double FISH demonstrated that nascent transgenic transcripts often co-localize with endogenous pre-rRNA (Figure 2C) indicating that transgene locus is recruited to nucleolus.

To analyze the cell-to-cell heterogeneity we first measured the fraction of nurse cells in each chamber that expressed the transgene. The intact rDNA transgene is expressed in the majority ( $10 \pm 1$  out of 15, 67%), but not all cells of the egg chamber (Figure 2D). The fraction of individual nuclei with detectable FISH signal drops to 27% for transgenes with R1 and R2 insertions (Figure 2D). Next, we measured the level of HCR FISH signal intensity in individual nuclei that had detectable signal. The intact rDNA transgene showed high level of cell-to-cell variability in expression (Figure 2E). The signal intensity was ~ 30-fold lower for transgenes containing R1 and R2 insertions compared to intact rDNA transgene (Figure 2E). These experiments revealed an unexpected cell-to-cell variability in expression of rDNA transgenes that was not present for control protein-coding gene *vasa*. rDNA units with insertions are expressed in less than one-third of all nuclei and even the intact unit is not expressed in all cells. Thus, repression of rDNA transgenes with insertions differs from cell to cell: in the majority of cells repression is complete and in the remaining cells expression is strongly reduced.

### **The SUMO pathway is required for repression of transposons integrated into native rDNA loci**

In the *Drosophila* germline many transposon families are repressed by the piRNA pathway. We recently found that SUMO, encoded by the single *smt3* gene in *Drosophila*, and the E3 SUMO-ligase Su(var)2-10/dPIAS are involved in piRNA-mediated transcriptional silencing of transposons in germline (Ninova et al.,

2020a). SUMO is covalently attached to many nuclear proteins in a conserved pathway that consists of the E1 (Uba2/Aos1) and E2 (Ubc9) SUMO ligases (Figure 3A). SUMOylation of target proteins requires transfer of SUMO from Ubc9, which may occur directly or be aided by specific E3 SUMO ligases.

In agreement with the function of *smt3* and *Su(var)2-10* in piRNA-guided transposon repression, germline-specific knockdown (GLKD) of either gene causes similar activation of many TE families (Ninova et al., 2020a; Ninova et al., 2020b). Notably, GLKD of *Su(var)2-10* had a modest effect on R1 and R2 compared to many other TE families that were upregulated more strongly. However, knockdown of *smt3* has a disproportionately strong effect on expression of R1 and R2 transposons. RNA-seq analysis showed that the levels of R1 and R2 transcripts increased ~1000 and ~300 fold, respectively, and they become among the top 20 most abundant cellular mRNA (an average of 3164 and 3421 TPM (transcripts per million) in *smt3* GLKD, compared to 2.76 and 10.84 TPMs in control ovaries) (Fig 3B,C, Figure 3 - figure supplement 1). The extremely strong R1 and R2 activation upon depletion of SUMO was also confirmed by RT-qPCR (Figure 3D). Analysis of RNA-seq data demonstrated that such strong up-regulation is exclusive for R1 and R2 and is not seen for any other protein-coding or ncRNA genes: the vast majority of other transposons and genes whose expression was affected by *smt3* GLKD changed less than 10-fold. Thus, SUMO seems to be involved in specific process of R1/R2 repression which is independent of *Su(var)2-10* SUMO ligase.

To further explore the role of SUMO and piRNA in repression of R1 and R2 we knocked down *smt3* and the single E2 SUMO ligase, *Ubc9*, in S2 cells that lack an active piRNA pathway. RT-qPCR showed strong and specific upregulation of R1 and R2 upon both *smt3* and *Ubc9* knock-downs in S2 cells (Figure 3E). In contrast, knock-down of *Su(var)2-10* in S2 cells did not cause activation of R1 and R2 (Figure 3 - figure supplement 1). Thus, SUMO is required for the potent and specific repression of R1 and R2 in both the soma and the germline, employing a mechanism that is independent of piRNA and *Su(var)2-10*. R1 and R2 are embedded into rDNA and use the rRNA transcription machinery for their

expression suggesting that SUMO controls expression of rDNA units damaged by transposon insertions.

### **SUMO depletion releases repression of interrupted and intact rDNA to equalize their expression**

To test if SUMO regulates expression of the single-copy rDNA transgenes we studied their expression upon Smt3 knock-down by RT-qPCR. Transgenes with R1 and R2 insertions showed 54-fold and 50-fold increase in pre-rRNA expression level upon Smt3 KD (Fig 4A). Different transgenes with R1 and R2 as well as CFP insertions showed similar 41- to 54-fold increase in pre-rRNA level. Thus, the function of SUMO in repression of damaged endogenous rDNA units with insertion of R1 and R2 transposons is recapitulated using rDNA transgenes. Surprisingly, expression of the intact rDNA transgene that lacks any insertion was also up-regulated ~12 fold upon SUMO KD (Figure 4A). While in wild-type flies interrupted rDNA copies are repressed compared to intact units (Figure 2), intact and interrupted copies have similar expression levels in the absence of SUMO. Thus, SUMO is required for differential expression of intact and interrupted rDNA. SUMO pathway participates in repression of both intact and damaged rRNA, but does so with different efficiency.

To get further insights into expression of intact and damaged rDNA transgenes upon SUMO depletion we employed HCR-FISH to study expression in individual nurse cells (Figure 4B). Confirming the RT-qPCR results, Smt3 GLKD lead to a marked increase in expression of both intact and damaged rDNA transgenes. First, in contrast to wild-type flies, expression of intact and interrupted rDNA was detected in almost all nuclei upon Smt3 KD (Figure 4C). Second, depletion of SUMO increased the signal intensity per nucleus: after Smt3 KD the signal increased 35~37-fold for damaged transgenes and ~ 2.5-fold for the intact transgene (Figure 4D). Consistent with the RT-qPCR results, intact and damaged transgenes showed similarly high expression level upon SUMO KD. Thus, while

damaged units are preferentially silenced in wild-type cells, loss of SUMO eliminates differential repression.

To test the role of SUMO in expression of endogenous rDNA units we measured pre-rRNA levels by RT-qPCR using primers that amplify the external transcribed spacer (EST) portion. Depletion of Smt3 led to 2.5-fold increase in pre-rRNA levels both in ovarian germ cells and in S2 cells (Figure 4E), demonstrating that SUMO controls expression of endogenous rDNA arrays.

### **The role of heterochromatin in SUMO-dependent rDNA repression**

SUMO-dependent repression of rRNA might be caused by co-transcriptional degradation of pre-rRNA or by suppression of Pol I transcription. To discriminate between these possibilities, we employed ChIP to measure the presence of RNA pol I on intact and damaged transgenes. We generated transgenic flies that express GFP-tagged Rpl135, an essential subunit of Pol I. As expected, GFP-tagged Rpl135 localized to the nucleolus (Figure 5 - figure supplement 1A). Rpl135 ChIP-qPCR using a GFP antibody showed enrichment of Pol I on the ETS portion and on 18S of native rDNA units (Figure 5 - figure supplement 1B). Enrichment of Pol I was also detected on promoters of the rDNA transgenes. Importantly, depletion of SUMO caused 2.5 to 4-fold increase in the levels of Pol I on intact and damaged transgenes (Figure 5A), indicating that the SUMO-dependent pathway regulates rDNA transcription.

rDNA arrays are embedded in heterochromatin on the X and Y chromosomes, a genomic compartment characterized by high level of the H3K9me3 mark, which is linked to repression of protein-coding genes transcribed by RNA pol II. H3K9me3 was proposed to play a role in rDNA silencing (Santoro et al., 2002). To study whether the SUMO pathway regulates rDNA expression through deposition of H3K9me3 mark, we performed H3K9me3 ChIP-seq in the fly ovary upon Smt3 knockdown. We found that endogenous rDNA units and R1 and R2 sequences are indeed enriched in H3K9me3 mark in wild-type flies, however, no change in H3K9me3 occupancy was observed upon SUMO depletion (Figure 5B). We used

independent CHIP-qPCR to compare the levels of heterochromatin mark on R1 and R2 transposons to other hetero- and euchromatin regions. These experiments confirmed that R1 and R2 sequences are heterochromatic, however, the level of H3K9me3 was not affected by Smt3 KD on R1 and only decreased by 36.7% on R2 (Figure 5C).

The level of H3K9me3 mark was ~ 4-fold lower on transgenic compared to endogenous rDNA copies and only slightly higher compared to control euchromatin region (Figure 5D). Furthermore, heterochromatin mark levels were similar on intact and damaged rDNA transgenes. Finally, H3K9me3 levels on intact or damaged transgenes did not decrease upon SUMO knockdown, despite the strong activation of their expression. Taken together, these results indicate that the level of H3K9me3 mark on endogenous and transgenic rDNA does not correlate with silencing.

To further explore the role of heterochromatin marks in repression of rDNA we used RNAi to knock-down three histone methyltransferases, *SetDB1*, *Su(var)3-9* and *G9a*, that are together responsible for all mono-, di- and trimethylation of H3K9 in *Drosophila* (Supplementary File 4). Depletion of *SetDB1* and *G9a* had no effect on R1 and R2 expression, while depletion of *Su(var)3-9* caused very mild ~2.5-fold change (compared to ~1,000 fold derepression upon *smt3* and *Ubc9* RNAi) (Figure 5E). Simultaneous knockdown of all three histone methyltransferases had the similar effect as depletion of *Su(var)3-9* alone indicating no redundancy. Thus, our results suggest that H3K9me1/2/3 marks are not essential for silencing and SUMO represses rDNA by an H3K9me-independent mechanism.

### **SUMOylation of nucleolar proteins involved in rDNA expression**

Proteomic studies in *Drosophila* identified several hundred SUMOylated proteins (Handu et al., 2015; Nie et al., 2009). However, the list is likely to be even longer considering frequency of the SUMOylation consensus ( $\Psi$ KxE/D) in the fly proteome and technical difficulties of detecting SUMOylation that often affects only a small fraction of target protein molecules (Hay, 2005). Many chromatin proteins,

including histones, are substrates of SUMOylation (Shiio and Eisenman, 2003; Nathan et al., 2003) suggesting a possibility that rDNA repression might be caused by SUMOylation of chromatin proteins on rDNA sequences. To explore this possibility, we analyzed previously published SUMO ChIP-seq data (Gonzales et al. 2014). This analysis revealed enrichment of SUMO at rDNA unit and R2 sequences as well as at the 5' regions of the R1 transposon (Figure 6A) indicating that SUMOylated proteins are indeed enriched on rDNA chromatin.

To find if specific proteins involved in rDNA transcription and nucleolar function are SUMOylated we employed sensitive SUMOylation assay. GFP-tagged candidate proteins were co-expressed together with FLAG-SUMO in S2 cells followed by their immunoprecipitation and detection of SUMOylation. The assay successfully detected SUMOylation of CTCF, conserved Zn-finger protein involved in high order chromatin organization that is known target of SUMOylation (MacPherson, et al., 2009; Guerrero, Maggert, 2011). Out of 7 other candidates we detected SUMOylation of three proteins including Udd and CG3756, two proteins involved in RNA pol I dependent transcription of rDNA (Figure 6B). As expected from previous observations (Hay, 2005), only small fraction of target proteins was modified. Thus, the limited screen of selected nucleolar proteins suggests that potentially many more proteins involved in rDNA transcription might be SUMOylated.

To further explore this possibility, we retrieved all *D. melanogaster* proteins associated with Gene Ontology term “nucleolus” (GO:0005730 and child terms, n=243 Flybase-annotated genes) and searched for the SUMOylation consensus ( $\psi$ KxE/D) in their sequences. This analysis showed that 77% of nucleolar genes have products harboring a SUMOylation consensus. In addition, 68% (19 out of 28) genes associated with the biological process “transcription by RNA polymerase I” (GO:0006360) have a SUMOylation consensus. Proteomic approach was previously used to comprehensively identify SUMOylated proteins in *Drosophila* (Handu et al., 2015). Ontology analysis revealed that 27 out of 823 SUMOylated proteins are associated with the GO term “nucleolus” demonstrating significant enrichment of nucleolar proteins among SUMOylated proteins (BH=adjusted p-

value  $<0.01$ ) (Supplementary File 3). Taken together, the results of experimental and computational analyses indicate that many proteins involved in nucleolar function and rDNA transcription are SUMOylated.

To find proteins involved in SUMO-dependent rDNA repression we used RNAi to knock-down candidate genes and monitor expression of R1 and R2 transposons in S2 cells. We composed a list of proteins that have nucleolar function or involved in rDNA transcription and are SUMOylated according to published (Handu et al., 2015) and our own (Ninova, unpublished) mass-spec data (Supplementary File 4). We also selected proteins involved in SUMO pathway such as E3 SUMO ligases and SUMO isopeptidases. Out of 25 tested genes, knockdown of five (*Ulp1*, *CG13773*, *CG3756*, *Fib*, *mbm*) caused increase in expression of R1 and R2, including SUMO isopeptidase Ulp1 (Figure 6C). None of tested E3 SUMO ligases scored positive suggesting that SUMO-dependent rDNA repression requires yet unknown E3 ligase or involves direct transfer of SUMO by E2 ligase Ubc9 that is indispensable for the repression. The RNAi screen also revealed that repression requires two proteins that contribute to RNA polymerase I activity: *CG13773* and *CG3756*. However, the magnitude of R1/R2 upregulation was rather modest (3 to 30-fold) compared to dramatic ( $>1,000$ -fold) derepression observed upon SUMO knockdown. Overall, the results of RNAi screen further support the role of SUMO pathway in rDNA regulation and suggest that multiple proteins including several SUMOylated components of RNA pol I complex are involved in this process.

To directly test if local SUMOylation of chromatin proteins in vicinity of rDNA promoter lead to repression we recruited E2 SUMO ligase Ubc9 to rDNA transgene. Ubc9 enzyme catalyzes the transfer of SUMO that is covalently linked to it to many proteins that harbor simple SUMOylation motif (Johnson and Blobel, 1997) (Figure 3A). We generated transgenic flies with intact rDNA harboring 16-nt hairpin sequence that can be irreversibly bound by inactive version of bacterial Csy4 nuclease (Lee et al., 2013) (Figure 6D). Expression of Ubc9 fused to Csy4 lead to 4.7-fold decrease in transgene expression compared to control flies suggesting that increase in local SUMOylation lead to further rDNA repression (Figure 6D).

## Discussion

To satisfy the high demand for rRNAs – essential components of ribosomes - genomes of most organisms contain multiple identical rDNA genes. However, studies in many eukaryotic species paradoxically demonstrated that only a fraction of available rDNA genes are expressed, while other rDNA units with apparently identical sequence are inactive (Conconi et al., 1989; Morgan et al., 1983; Sogo et al., 1984). rDNA repression was linked to rDNA stability, prevention of recombination and preserving nucleolar structure (Espada et al., 2007; Sinclair et al., 1997). Differential expression of ribosomal RNA genes represents an ultimate case of epigenetic regulation: identical DNA sequences have drastically different expression levels within a single cell and these expression states are propagated through multiple cellular divisions.

Studies of rDNA repression are hampered by the fact that hundreds of rDNA units are present in the genome with almost identical sequence, which cannot be reliably discriminated (Ganley and Kobayashi, 2007). To circumvent this problem and understand regulation of rDNA expression, we used molecularly tagged rDNA transgenes inserted in a heterologous locus in the *D. melanogaster* genome. Insertion of a short unique sequence into the 5'-external transcribed spacer (ETS) of rDNA did not interfere with its expression and allowed us to monitor expression and chromatin structure of the rDNA transgenes and to discriminate it from the native rDNA units.

In many organisms, including *Drosophila*, native rDNA clusters are located within constitutive heterochromatin, repeat-rich and gene-poor regions that are transcriptionally silent (Németh and Längst, 2011; Yoshida, 1998). In agreement with previous studies (Karpen et al., 1988), our results show that individual ribosomal RNA units inserted in a heterologous genomic position outside of the native rDNA cluster can be transcribed by Pol I and processed (Figure 1C, 2C, Figure 5 - figure supplement 1). Insertions of an individual rDNA unit in a euchromatic region in *Drosophila* was shown to be transcribed and to recruit nucleolar proteins to form nucleoli that are morphologically and functionally similar

to endogenous nucleoli (Karpen et al., 1988). Furthermore, individual rDNA units in heterologous genomic locations are able to partially rescue the bobbed phenotype caused by decreased number of endogenous rDNA units, indicating that transgenic rRNA is functional and incorporated into ribosomes. Similarly, individual rDNA insertions into different genomic sites in *Saccharomyces cerevisiae* were shown to be competent in assembly of the RNA Polymerase I complex and production of ribosomal RNA (Oakes et al., 2006). Thus, our results corroborate previous studies that suggest that the repeated structure and genomic location of rDNA loci are not required for their transcription.

Previous studies suggested that rDNA units that are damaged by insertion of R1 and R2 retrotransposons, which integrate into specific sites in 28S rDNA in *Drosophila* are selectively repressed (Eickbush and Eickbush, 2003; Jolly and Thomas, 1980; Kidd and Glover, 1981; Long et al., 1981a, 1981b). Variable levels of repression were also observed upon integration of a non-transposon sequence in 28S rDNA in a native rDNA cluster (Eickbush and Eickbush, 2003). The molecular mechanism by which damaged rDNA units are repressed and its link to silencing of intact rDNA remained poorly understood. Our results indicate that integration of a sequence other than transposons can induce rDNA repression. Interestingly, extremely truncated R2 copies were shown to be actively expressed, indicating that rDNA with very short insertions can escape silencing (Eickbush and Eickbush, 2003). However, although the length of CFP (720 bp) is shorter than the full-length R1 and R2 transposons (5.4 and 3.6 kb, respectively), it is sufficient to induce repression. Integrating a heterologous sequence into rDNA also induced silencing in mammals, which lack transposons that specifically integrate into rDNA. Integration of the human growth hormone gene into rat ribosomal locus caused deletion of the ribosomal sequence and silencing of rDNA (Henderson and Robins, 1982). Thus, repression of rDNA copies damaged by insertions of heterologous sequences seems to be an evolutionary conserved process that prevents production of aberrant rRNA, which might interfere with proper ribosome assembly. Detection of transcripts from intact and damaged rDNA transgenes in individual cells revealed large cell-to-cell variability in their expression. Although intact

transgenes are expressed in a larger fraction of cells than damaged transgenes, they are still not expressed in all cells, indicating that intact rDNA units also undergo silencing, albeit less frequently than damaged units. In search of the molecular mechanism of rDNA repression we found that SUMO depletion lead to dramatic increase in the expression of R1 and R2 transposons, which integrate in native rDNA clusters (Figure 3B-E). Potent derepression of R1 and R2 was observed only upon SUMO depletion, but not in piRNA pathway mutants (Ninova et al., 2020a) or upon depletion of E3 SUMO ligase Su(var)2-10 (Ninova et al., 2020a) indicating that SUMO plays a special role in repression of rDNA-targeting transposons that is independent of the piRNA pathway. Furthermore, depletion of SUMO also activates rDNA transgenes indicating that they are repressed by the same mechanism that silences native rDNA units (Figure 4A). Surprisingly, SUMO KD increased expression of intact transgenes and released their repression in individual cells indicating that the same SUMO-dependent pathway is responsible for repression of both damaged and intact rDNA.

### **Mechanism of SUMO-dependent rDNA repression**

SUMO-dependent repression correlates with the levels of nascent pre-rRNA (Figure 4B-D) and decreased Pol I occupancy at rDNA promoters (Figure 2C) suggesting that repression acts at the level of transcription. How the presence of an insertion within the body of the rDNA unit lead to decreased transcription at the promoter remains to be understood. In mammals repression of rDNA units was shown to correlate with the presence of CpG DNA methylation and repressive histone marks near the rDNA promoter (Bird et al., 1981; Coffman et al., 2005; Conconi et al., 1989; Dammann et al., 1993; Flavell et al., 1988; Earley et al., 2006; Li et al., 2006; Santoro and Grummt, 2001; Santoro et al., 2002; Zhou et al., 2002). Furthermore, the chromatin remodeling complexes NoRC and NuRD were shown to be involved in rDNA repression probably by altering accessibility of the rDNA promoter to chromatin repressors such as DNA methyltransferases, histone methyltransferases and deacetylases (Santoro et al., 2002; Strohner et al., 2001;

Xie et al., 2012; Zhou et al., 2002). Unlike mammals *Drosophila* lack DNA methylation (Urieli-Shoval et al., 1982), so this mechanism plays no role in rDNA repression in flies. Our results indicate that native (but not transgenic) rDNA units are enriched in H3K9me3, a histone mark associated with repression of genes transcribed by Pol II. However, the level of H3K9me3 remained high on rDNA units upon SUMO depletion (Figure 5), indicating that the presence of this mark is not sufficient for repression. Furthermore, rDNA transgenes inserted in heterologous genomic locus are repressed by SUMO-dependent mechanism despite the fact that they lack H3K9me3 mark.

Large number of nuclear proteins are SUMOylated and SUMO is required for many nuclear processes (Nacerddine et al., 2005; Zhao and Blobel, 2005) including different stages of ribosome maturation (Heun, 2007; Takahashi et al., 2008; Finkbeiner et al., 2011) and transcriptional repression of protein-coding genes (Gill, 2005; Verger et al., 2003). Therefore, SUMO might play two different – though not necessarily mutually exclusive – functions in rDNA silencing. First, repression might depend on SUMOylation of proteins directly involved in rDNA silencing. Alternatively, depletion of SUMO might influence rDNA expression indirectly by changing expression of genes involved in repression. In agreement with known role of SUMO in Pol II transcription, its depletion in female germline caused changes in gene expression. We observed statistically significant >2-fold up-regulation of ~4% (323) and down-regulation of ~1.5% (138) of the analyzed genes (qval<0.05, likelihood ratio test). However, the vast majority of affected genes changed within 2-10 fold, in stark contrast to dramatic ~1000- and ~300-fold upregulation of R1 and R2 elements upon SUMO depletion (Fig 3B). We did not find any significantly enriched GO terms associated with the up- and down-regulated gene sets upon SUMO KD related to nucleolar function and/or rRNA transcription (BH-adjusted p-value cutoff 0.05). Manual inspection showed that the only gene associated with the cellular component GO term ‘nucleolus’ ("GO:0005730") and offspring terms among down-regulated genes was *smt3* itself. Only two out of the 323 up-regulated genes - *ph-p* and *CG9123* - are associated with the ‘nucleolus’ GO term, however, neither of them has known function in rDNA

expression. We did not find any matches to the biological process GO:0006360 (transcription by RNA polymerase I) and offspring. Therefore, while we cannot completely exclude a possibility of secondary effects of SUMO depletion, there is no direct indication that expression of genes involved in rDNA transcription and nucleolar function is affected by SUMO depletion.

Several lines of evidence suggest that SUMO plays direct role in rDNA silencing through modification of one or several proteins involved in rDNA expression. First, many chromatin proteins, including histones, are substrates of SUMOylation (Shiio and Eisenman, 2003; Nathan et al., 2003) and ChIP-seq analysis revealed enrichment of SUMO at R1/R2 and rDNA sequences (Figure 6A). Second, our analysis of SUMOylated proteome using both targeted and unbiased approaches indicates that many nucleolar proteins, including several components of RNA pol I machinery and Udd protein previously implicated in rDNA expression (Zhang et al., 2014), are SUMOylated *in vivo* (Figure 6B, Supplementary File 4). Third, recruitment of SUMO ligase Ubc9 to rDNA transgene led to its repression revealing direct effect of local SUMOylation on rDNA expression (Figure 6D). Finally, local SUMOylation in proximity of gene promoters was shown to induce repression of Pol II driven transcription in *Drosophila* and mammals (Ninova et al., 2020a; Ninova et al., 2020b; Stielow et al., 2008; Smith et al., 2011; Lehembre et al., 2000; Yang et al., 2003; Yang and Sharrocks, 2004). Particularly, SUMOylation at promoters is required for repression induced by KRAB-ZFPs, the largest family of mammalian transcriptional silencers (Li et al., 2007). Pol I-driven rRNA promoters were found to be one of the most prominent sites of active sumoylation in human cells (Neyret-Kahn et al, 2013). Remarkably, inhibition of sumoylation caused upregulation of rRNA expression indicating that the role of SUMO in rDNA regulation is conserved between insects and mammals (Neyret-Kahn et al, 2013; Peng et al., 2019).

Our results suggest that SUMOylation of chromatin-associated proteins might act as a molecular mark for rDNA repression. The targets of SUMOylation involved in rDNA repression remained to be identified. Unlike SUMO depletion, knock-down of several proteins that are both SUMOylated and have known functions in rDNA

expression cause relatively mild increase in R1/R2 expression (Figure 6C) arguing against a possibility that any single protein from this list is solely responsible for SUMO-dependent repression. However, modest R1/R2 upregulation observed upon depletion of several proteins suggest that repression might require modification of several chromatin-bound proteins rather than one specific target. Promiscuous modification of multiple proteins is in line with the findings that SUMOylation consensus is very simple and quite common in the fly proteome (Nie et al., 2009) and our results that show that multiple nucleolar proteins are SUMOylated. The proposal that rDNA repression depend on cumulative SUMOylation of several proteins localized on rDNA corroborated by studies in yeast that led to the “SUMO spray” hypothesis (Psakhye and Jentsch, 2012) which suggests that SUMOylation of multiple proteins localized in physical proximity is required to induce downstream response in SUMO-dependent pathway. Importantly, multiple interactions between SUMOylated proteins and proteins that bind SUMO act synergistically and thus SUMOylation of any single protein is neither necessary nor sufficient to trigger downstream processes (Jentsch and Psakhye, 2013; Psakhye and Jentsch, 2012). It will be interesting to see if SUMO promiscuously ‘sprayed’ on rDNA units to induce their repression.

## **Acknowledgements**

We thank Katalin Fejes Toth and members of the Aravin lab for discussion and comments. We appreciate the help of Maayan Schwarzkopf and Niles Pierce with HCR FISH experiments. We thank Lynn Yi for help with bioinformatics analysis. We are grateful to Michael Buszczak and the Bloomington Stock Center for providing fly stocks, Igor Dawid for providing rDNA unit constructs. We thank Igor Antoshechkin (Caltech) for help with sequencing. M.N. is supported by NIH/NICHD grant (K99HD099316). This work was supported by grants from the National Institutes of Health (R01 GM097363) and by the HHMI Faculty Scholar Award to A.A.A.

## Figure Legends

### Figure 1. Single-copy rDNA transgene allows study of rDNA expression

#### (A) Scheme of endogenous rDNA units and the rDNA transgene.

Native rDNA units are composed of non-transcribed spacer (NTS) and a transcribed portion that produces the 47S pre-rRNA. The pre-rRNA contains a 5' external transcribed sequence (ETS) and internal transcribed spacers (ITSs), which are removed from the pre-rRNA to generate mature 18S, 5.8S, 2S and 28S rRNAs. Positions of R1 and R2 transposon integration sites in the 28S rRNA are indicated. The 9 Kb transgene contains one complete rDNA unit together with a 28bp of the 28S rDNA from the upstream unit. The transgene is marked by insertion of a 21bp unique identification sequence (UID, red bar) into the ETS and inserted using  $\Phi$ C31-mediated recombination into a common *att* site on 2<sup>nd</sup> chromosome (chr 2L: 1,582,820).

#### (B) rDNA transgene expression is detected by RT-PCR in fly ovaries.

RT-PCR amplicon of the UID ETS region is only detected in transgenic but not in wild type flies or in the absence of reverse transcriptase (-RT). rDNA transgene expression in ovaries was measured by RT-qPCR and normalized to rp49 mRNA. Error bars indicate standard deviation of three biological replicas. Statistical significance is estimated by two-tailed Student's t-test; \*\*\*  $p < 0.001$ .

#### (C) rDNA transgene expression is detected by HCR FISH in fly ovaries.

Nascent transcripts of the pre-rRNA transgene (arrowhead) were detected in nurse cell nuclei using a probe against the UID sequence (red). Control wild type flies lack the UID sequence. Scale bar: 5 $\mu$ m.

### Figure 2. Insertions into 28S rRNA lead to decreased transgene expression

#### (A) The scheme of rDNA transgenes.

The sequences of R1 and R2 transposons were inserted into their natural integration sites within the 28S rRNA. Transgenes R1' and R2' are identical to R1

and R2 transgenes, respectively, but contain second UID flanking retrotransposon sequence. The promoterless CFP sequences was inserted into the same R1 (CFP-1 transgene) and R2 (CFP-1 transgene) integration sites. 29nt sequence was inserted into R2 site. All constructs were integrated into the same genomic *att* site (chr 2L: 1,582,820) on the 2<sup>nd</sup> chromosome using  $\Phi$ C31-mediated recombination.

**(B) Transgenic rRNA expression is decreased upon insertion of foreign sequence.**

Expression of rDNA transgenes in ovary (top) and carcasses (bottom) measured by RT-qPCR and normalized to rp49 mRNA. Error bars indicate standard deviation of three biological replicas. Statistical significance is estimated by two-tailed Student's t-test; \*\*\*  $p < 0.001$ .

**(C) R1 or R2 insertion decreases transgenic rRNA transcripts as detected by HCR FISH**

Endogenous pre-rDNA is detected in stage 7-8 nurse cell nuclei using a probe against ETS (green), while transgenes are detected with probe against the UID (Red). Vasa mRNA is detected as a control (green). Scale bar: 10 $\mu$ m.

**(D) Transgenic rDNA is expressed in fewer nuclei upon R1/R2 insertion**

Shown is the number of nurse cell nuclei with positive HCR FISH signal per egg chamber (out of 15 nuclei per chamber). Each circle represents data from one egg chamber. Error bars indicate standard deviation from 5 egg chambers. Statistical significance is estimated by two-tailed Student's t-test; \*\*\*  $p < 0.001$ .

**(E) Transgenic rDNA HCR FISH signal is reduced upon R1/R2 insertion**

The total intensity of HCR FISH signal was measured in individual nuclei that have positive signal. Error bars indicate standard deviation from 5 nuclei. Statistical significance is estimated by two-tailed Student's t-test; \*\*\*  $p < 0.001$ .

**Figure 3. SUMO pathway is required for repression of transposons integrated into rDNA.**

### **(A) Scheme of SUMO pathway.**

SUMO is activated by E1 enzyme composed of Uba2/Aos1 dimer. Next SUMO is transferred from Aos1 to E2 Ligase Ubc9. Ubc9 can directly transfer SUMO to target proteins. SUMOylation of some (but not all) targets require help of the one of several E3 ligases that bring target protein and Ubc9-SUMO into proximity facilitating the transfer of SUMO.

### **(B) Change in genes and transposons expression upon SUMO KD in the ovary.**

(Left) Gene and transposon expression in RNA-seq data from *smt3* KD and control ovaries. Germline-specific knockdown of *smt3* was induced by small hairpin driven by maternal-tubulin-Gal4 driver; shRNA against the *white* gene was used as a control. Genes that change significantly ( $q_{val} < 0.05$ , LRT test, sleuth (Pimentel et al., 2017)) 2-fold and above are highlighted. R1, R2 transposons and *smt3* gene are labelled. (Right) Boxplot shows the distribution of fold changes in *smt3* KD versus control ovaries of genes (RefSeq) and transposons (RepBase). Genes with infinite fold change values (zero counts in control ovaries) are not shown. 1 is added to log<sub>2</sub>-transformed TPM values to enable the visual display of genes that are not expressed in a given condition.

### **(C-E) SUMO depletion leads to upregulation of R1 and R2 transposons**

**(C)** RNA-seq signal coverage along the R1 and R2 consensus sequences (RepBase) from control (shW) and SUMO-depleted (shSmt3) ovaries (r1 and r2 indicate two biological replicates). Data is normalized to total reads mapping to the genome.

**(D)** RT-qPCR analysis of R1 and R2 expression normalized to rp49 mRNA from SUMO-depleted (shSmt3) and control (shW) ovaries. PCR amplicons are within the R1 and R2 sequence (R1 and R2) or spans the junction between the 28S rRNA and the transposon (rRNA-R1 and rRNA-R2). Error bars indicate standard deviation of three biological replicates. Statistical significance is estimated by two-tailed Student's t-test; \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

**(E)** RT-qPCR analysis of R1 and R2 expression normalized to rp49 mRNA in S2 cells upon knockdown of SUMO (dsSmt3), the E2 SUMO ligase Ubc9 (dsUbc9) or control (dsGFP) by double strand RNA. Error bars indicate standard deviation of three biological replicates. Statistical significance is estimated by two-tailed Student's t-test; \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

**Figure 3 - Figure supplement 1. R1 and R2 are strongly up-regulated upon SUMO but not Su(var)2-10 depletion.**

Heat maps show RPKM-normalized expression of TE consensus (RepBase) in SUMO and Su(var)2-10 depleted ovaries (shSmt3 and shSv210) and respective controls (shW) (left) and in SUMO and Su(var)2-10 depleted S2 cells (dsSmt3 and dsSv210) and respective controls (dsGFP) (right). Data is average from two biological replicates. TEs with less than 10 reads in all conditions are excluded. Data is clustered by the maximum distance method.

**Figure 4. SUMO depletion leads to derepression of both damaged and intact rDNA transgenes**

**(A) SUMO knockdown depresses rDNA transgenes.**

Expression of rDNA transgenes in ovaries as measured by RT-qPCR upon germline knockdown of SUMO (shSmt3) and in control (shW). Data shows three biological replicates normalized to rp49 mRNA. Error bars indicate standard deviation of xxx biological replicates. Statistical significance is estimated by two-tailed Student's t-test; \*\*\*  $p < 0.001$ .

**(B) Transgenic pre-rRNA expression is repressed by SUMO as detected by HCR FISH**

Expression of rDNA transgenes in nurse cells is detected upon germline-specific knockdown of SUMO (shSmt3) or control (shWhite) gene using a probe against the UID sequence (Red). Vasa mRNA (green) is detected in parallel. Scale bar: 10 $\mu$ m. Bottom row shows magnification of individual boxed nuclei.

### **(C) rDNA transgenes are expressed in more nuclei upon SUMO KD as detected by HCR FISH**

Shown is the number of nurse cell nuclei with positive HCR FISH signal per egg chamber (out of 15 nuclei per stage 7/8 chamber). Each circle represents data from one egg chamber. Both intact and damaged transgenes are expressed in all nuclei upon SUMO knockdown. Error bars indicate the standard deviation of 5 egg chambers. Statistical significance is estimated by two-tailed Student's t-test; \*\*\*  $p < 0.001$ .

### **(D) Transgenic rDNA HCR FISH signal is increased upon SUMO KD**

The total intensity of HCR FISH signal was measured in individual nuclei that have positive signal. Error bars indicate the standard deviation of 5 nuclei. Statistical significance is estimated by two-tailed Student's t-test; \*\*\*  $p < 0.001$ .

### **(E) SUMO knockdown increases pre-rRNA expression.**

Expression of pre-rRNAs were measured by RT-qPCR using primers that target ETS region normalized to *rp49* mRNA. Germline-specific knockdown of SUMO (shSmt3) or control (shW) gene was induced by small hairpin driven by maternal-tubulin-Gal4 driver. In S2 cells, knockdown of SUMO (dsSmt3) or control (dsGFP) was induced by double stranded RNA. Error bars indicate the standard deviation of three biological replicates. Statistical significance is estimated by two-tailed Student's t-test; \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

## **Figure 5. SUMO KD does not affect H3K9me3 enrichment over rDNA and R1/R2**

### **(A) Pol I occupancy increases over rDNA transgenes upon SUMO KD**

Germline-specific knockdown of SUMO (shSmt3) or control (shW) gene was induced by small hairpin driven by maternal-tubulin-Gal4 driver. ChIP of Rpl135-GFP using a GFP antibody for pull-down was followed by qPCR analysis using UID-specific primers. Data were normalized using a sequence mapping to a gene-poor region. Error bars indicate standard deviation of three biological replicates.

Statistical significance is estimated by two-tailed Student's t-test; \*  $p < 0.05$ , \*\*\*  $p < 0.001$ .

**(B-C) H3K9me3 enrichment over native rDNA and the R1 and R2 transposons sequences is unaffected by SUMO KD.**

Germline knockdown was induced by small hairpin driven by maternal-tubulin-Gal4 driver (B) H3K9me3 ChIP-seq signal and corresponding input coverage across the R1 and R2 consensus sequences (RepBase), and the rDNA unit (Stage et al., 2007) from control (shW) and SUMO-depleted (shSmt3) ovaries. Data is normalized to total reads mapping to the genome. (C) H3K9me3 ChIP-qPCR using primers to native 28S rDNA interrupted with R1 and R2 insertions as in Figure 3D. ChIP to input enrichment is normalized to the region that has high level of H3K9me3 mark (chr2R: 4,141,405 - 4,141,502). Error bars indicate the standard deviation of three biological replicates. Statistical significance is estimated by two-tailed Student's t-test; \*\*  $p < 0.01$ .

**(D) H3K9me3 mark measured on rDNA transgenes by ChIP-qPCR.**

H3K9me3 ChIP-qPCR analysis of rDNA transgenes in control (shW) and SUMO-depleted (shSmt3) ovaries using primers to the UID sequence and RP49 (control region), normalized to H3K9me3-enriched region (chr2R: 4,141,405~4,141,502). Error bars indicate the standard deviation of three biological replicates. Statistical significance is estimated by two-tailed Student's t-test; \*\*  $p < 0.01$ .

**Figure 5 - Figure supplement 1.**

**(A) Rpl135-GFP localizes to the nucleolus.**

Rpl135-GFP (green) and nucleolus marker Fibrillarin IF (red) show co-localized in nurse cells of the fly ovary. Scale bar: 10 $\mu$ m.

**(B) Rpl135-GFP ChIP-qPCR.**

Fold enrichment of Rpl135-GFP as measured on ETS and 18S of endogenous rDNA in ovaries by ChIP-qPCR normalized using a sequence mapping to a

gene-poor region kalahari. Error bars indicate standard deviation of three biological replicates. Statistical significance is estimated by two-tailed Student's t-test; \*  $p < 0.05$ , \*\*  $p < 0.01$ .

## **Figure 6. SUMOylated proteins**

### **(A) SUMO ChIP-seq profile on rDNA and R1 and R2 consensus sequences.**

SUMO ChIP-seq signal over the R1 and R2 consensus and the rDNA unit as in (4A). ChIP-seq data is from (Gonzales et al. 2014).

### **(B) SUMOylation of proteins involved in rDNA transcription and nucleolar function.**

GFP-tagged proteins were co-expressed with FLAG-SUMO in S2 cells. SUMOylation was detected after immunoprecipitation with anti-GFP antibodies. Several proteins were tagged at either N- or C-terminus. In case of CG3756, only C-terminally tagged protein appears to be SUMOylated.

### **(C) RNAi screen for genes involved in R1 and R2 repression in S2 cells.**

After knock-down of gene expression using RNAi expression of R1 and R2 transposons was measured by RT-qPCR and normalized to *rp49* mRNA. Shown is fold-upregulation of R1 and R2 expression upon knock-down compared to control (double-stranded RNA against eGFP gene). Expression levels were measured in three biological replicates. Statistical significance is estimated by two-tailed Student's t-test; \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ . Information about genes and efficiency of RNAi KD is shown in Supplementary File 4.

### **(D) Tethering of E2 SUMO ligase Ubc9 suppresses rDNA transgene.**

(Left) The scheme of tethering experiment. The rDNA transgene was modified by insertion of binding site for Csy4 RNA-binding protein into ETS. E2 SUMO ligase Ubc9 was fused with Csy4 protein and co-expressed with rDNA transgene. (Right) The results of RT-qPCR analysis of rDNA transgene expression upon Ubc9 tethering.

## **Supplementary File 1. *Drosophila melanogaster* stocks**

## **Supplementary File 2. Primers**

**Supplementary File 3. Cellular component (CC) GO analysis of SUMOylated proteins identified in S2 cells.** The table shows results from GO enrichment analysis for the cellular component on SUMOylated proteins identified by Handu et al. 2015, performed using the enrichGO function of the clusterProfiler package (Yu et al., 2012)

## **Supplementary File 4. RNAi screen for genes involved in R1 and R2 repression in S2 cells**

For each gene the information column describes its known or predicted function and nucleolus localization. Association with GO term 'nucleolus' (GO:0005730) and its child terms is shown according to Gene Ontology analysis. The presence of SUMOylation motif(s) indicates the presence of consensus SUMOylation sequence ( $\psi$ KxE/D) in protein sequence. Expression of R1 and R2 transposons were measured by RT-qPCR in S2 cells after knock-down using double-stranded RNA using primers listed in Supplementary File 2. Shown is fold-change in R1 and R2 levels upon knock-down of corresponding gene compared to control KD (double-stranded RNA against eGFP gene). Knockdown efficiency shows the depletion of target gene as measured by RT-qPCR. Expression levels were measured in three biological replicates and normalized to rp49 mRNA. Statistical significance is estimated by two-tailed Student's t-test; \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

## **Supplementary File 5. Numerical data from RNA-seq analysis by sleuth**

## Materials and methods

### Fly stocks

All flies were raised at 25°C. The stock with shRNA against *white* was obtained from Bloomington (BDSC #33623), shRNA against *Smt3* Su(var)2-10 were described in (Ninova et al., 2020a). shRNAs were driven by the maternal alpha-tubulin67C-Gal4 driver (BDSC #7063 or #7062).

### Construction of rDNA transgenes and generation of transgenic flies

To make an rDNA unit marked with a unique sequence (UID), the full-length rDNA unit including the IGS (10.5 kb) was amplified by overlapping PCR from the DmrY22 plasmid (Long and Dawid, 1979) (a gift from the Dawid Lab). The UID sequence 5'-GACTCGAGTCGACATCGATGC-3' was inserted into the EST region at position +139 by overlap PCR. Overlap PCR was also used to insert R1, R2 and CFP at the R1 and R2 positions (2711 and 2648 nt) of the 28S rDNA. rDNA units were ligated into a vector that contains the *mini-white* gene and the  $\Phi$ C31 attB site using Gibson Assembly (NEB). All rDNA units were inserted into genomic site 22A3 (y1 w1118; PBac{y+-attP-3B} VK00037, BDSC stock #9752) using  $\Phi$ C31-mediated recombination.

To generate UASp-Rpl135-GFP transgenic flies cDNA of the Rpl135 gene was cloned into the pENTR™/DTOPO® (Invitrogen) entry vector and recombined into a Gateway® destination vector containing the  $\Phi$ C31 attB site, the *mini-white* marker, the UASp promoter and the eGFP C-terminal gateway cassette. The construct was integrated into the attP40 landing site at 25C6 (y1w67c23; P{CaryP}attP40). Transgenic flies used in this study are listed in Supplemental Supplementary File 1.

### RNA HCR FISH and image analysis

The previously described HCR FISH protocol (Choi et al., 2018) was used with the following modifications. Fly ovaries were dissected in cold PBS and fixed in 300  $\mu$ l fixation solution (4% paraformaldehyde, 0.15% Triton X-100 in PBS) at room temperature followed by three washes with PBX (PBS, 0.3% Triton X-100) for 5 min each at room temperature. Samples were dehydrated in 500  $\mu$ l 70% ethanol and permeabilized overnight at 4°C on a nutator. Samples were rehydrated in 500  $\mu$ l wash buffer (2 $\times$  SSC, 10% [v/v] formamide) for 5 min at room temperature and pre-hybridized in 500  $\mu$ l hybridization buffer (50% formamide, 5 $\times$  SSC, 9 mM citric acid pH 6.0, 0.1% Tween 20, 50  $\mu$ g/ml heparin, 1 $\times$  Denhardt's solution, 10% dextran sulfate) for 30 min at 37°C. Following pre-hybridization, the hybridization solution containing 2 pmol of each probe was added and samples were incubated 12-16 hours at 37°C. Samples were washed four times with 500  $\mu$ l probe wash buffer (50% formamide, 5 $\times$  SSC, 9 mM citric acid pH 6.0, 0.1% Tween 20, 50  $\mu$ g/ml heparin) for 15 min each at 37°C and three times with 5  $\times$  SSCT for 5 min each at room temperature. Samples were incubated in 500  $\mu$ l amplification buffer (5 $\times$  SSC, 0.1% Tween 20, 10% dextran sulfate) for 30 min at room temperature. 30 pmol each of hairpin H1 hairpin H2 were prepared separately by incubating at 95°C for 90 seconds and letting them cool to room temperature in the dark for 30 min. Samples were incubated with the hairpin solution for 12-16 h in the dark at room temperature followed by washing with 500  $\mu$ l 5 $\times$  SSCT at room temperature in following order: 2 $\times$  5 min, 2 $\times$  30 min and 1 $\times$  5 min. Samples were preserved on glass slides with mounting medium and imaged using a ZEISS LSM880 microscope. Probe sequences are listed in Supplementary File 2.

FISH signal intensity was analyzed using Fiji. For each nucleus the z-stacks of images were taken at same interval distance and the total FISH signal was calculated as a sum of the signals from each z-stack image.

### **Immunofluorescence microscopy**

Immunofluorescence microscopy was performed as previously described (Hur et al., 2013). Anti-Fibrillarin antibody (Abcam ab5821) was added to fixed ovaries at

1:500 dilution and incubated at 4°C overnight followed by incubation with 1:500 dilution of secondary Alexa fluor568 antibody. Images were acquired using the ZEISS LSM880.

### **RNAi in S2 cells**

The RNAi protocol was described previously (Rogers and Rogers, 2008). For RNAi screen in S2 cells, we used the protocol from Drosophila Genomics Resource Center (DGRC) cell-based RNAi (<https://fgr.hms.harvard.edu/fly-cell-RNAi-6-well-format>) scaled down to 24-well plate with subsequent changes. Primers to generate templates for dsRNA transcription are listed in Supplementary File 2. Double-stranded RNA were synthesized using the MEGAscript™ T7 Transcription Kit. Around  $0.5 \times 10^6$  cells in serum-free media were plated to wells containing 5 µg of dsRNA. After bathing cells with dsRNA on room temperature for 30 minutes, complete media with 10% FBS was added. Cells were grown for 3 days until harvesting.

### **RT-qPCR**

About 20 fly ovaries were dissected, homogenized in 1 ml TRIzol (Invitrogen) and RNA was extracted and precipitated according to the manual. Reverse transcription was performed using SuperScript III (Invitrogen) with random hexamer. RT-qPCR target expression was normalized to rp49 mRNA expression. All qPCR primers are listed in Supplementary File 2.

### **ChIP-seq and ChIP-qPCR**

All ChIP experiments followed the protocol described previously (Chen et al., 2016) using anti-H3K9me3 antibody from Abcam (ab8898) and anti-GFP antibody from ThermoFisher (A11122) for Rpl135-GFP. qPCR signal using primers to the respective regions was normalized to Kalahari as described previously (Sienski et al., 2015). All qPCR primers are listed in Supplementary File 2. ChIP-seq libraries

were generated using the NEBNext ChIP-Seq Library Prep Master Mix Set. All libraries were sequenced on the Illumina HiSeq 2500 platform.

## RNA-seq

Total RNA was extracted from fly ovaries by using TRIzol according to the manual. Ribosomal RNA depletion was performed with the Ribo-Zero™ rRNA Removal Kit (Illumina). RNA-seq libraries were made using the NEBNext® Ultra™ Directional RNA Library Prep Kit. Libraries were sequenced on the Illumina HiSeq 2500 platform.

Raw RNA-seq and ChIP-seq data for *smt3* KD in *D. melanogaster* ovaries is available on the GEO database, GSE115277. RNA-seq data from S2 cells is available on the GEO database, GSE141068.

## Bioinformatics analysis

For differential gene expression analysis, RNA-seq was pseudoaligned to the *Drosophila melanogaster* transcriptome (RefSeq) and transposon consensus (from RepBase, Jurka et al. 2005), using kallisto (Bray et al., 2016) with the following parameters: ‘–single -t 4 -l 200 -s 50 -b 30 –rf-stranded’. Subsequent differential expression analysis was performed with sleuth using the gene analysis option (Pimentel et al., 2017). Fold changes in gene expression were calculated from the average TPM between replicates in knockdown versus control ovaries. The R code for this analysis is available on github at [https://github.com/mninova/smt3\\_KD](https://github.com/mninova/smt3_KD). Gene ontology (GO) analysis was performed on the genes that were significantly up- and down- regulated upon *smt3* KD ( $\log_2FC > 1$  and  $qval < 0.05$  (likelihood ratio test in sleuth)), using the clusterProfiler R package (Yu et al., 2012) as well as the FlyMine web server, using all ovary-expressed genes that were not filtered out by sleuth, or all genes as background. In either case, no enrichment for rRNA- or nucleolar- related function was found (BH-adjusted p-value 0.05).

For genomic enrichment histograms and heatmaps in Supplemental Figure 1, RNA-seq and ChIP-seq data were aligned to the dm3 assembly using Bowtie 0.12.17 (Langmead et al., 2009) with `-a -best -strata -m 1 -v 2` settings, and to TE consensus sequences from RepBase (Jurka et al. 2005) allowing up to 3 mismatches. TE consensus expression was calculated using an in-house python script as RPKM of total mapped reads to the genome. ChIP-seq data was further aligned to the rDNA unit (Stage et al. 2007) allowing zero mismatches. Per-base coverage of consensus sequences was calculated with bedtools (Quinlan and Hall, 2010) using total mapped reads as a scaling factor. Heatmaps were generated with the 'pheatmap' R package.

Gene products of *D. melanogaster* associated with the GO terms "nucleolus" (GO:0005730) and transcription by RNA polymerase I ("GO:0006360") and respective offspring terms were retrieved using the EBI QuickGO tool, and their aminoacid sequences were retrieved from UniProt. Sequences were searched with the SUMOylation consensus pattern  $\psi Kx[E/D]$  where  $\psi$  is a hydrophobic aminoacid and x is any aminoacid (Rodriguez et al., 2001; Sampson et al., 2001). Experimentally identified SUMOylated proteins in S2 cells (n=923) were retrieved from the list provided by Handu et al. GO enrichment analysis was performed using the clusterProfiler R package, `enrichGO` function (Yu et al., 2012) using all *Drosophila melanogaster* proteins as background, and BH-adjusted p-value of 0.05.

## References

- Ali, S. A., Zaidi, S. K., Dacwag, C. S., Salma, N., Young, D. W., Shakoori, A. R., Montecino, M. A., Lian, J. B., van Wijnen, A. J., Imbalzano, A. N., Stein, G. S., & Stein, J. L. (2008). Phenotypic transcription factors epigenetically mediate cell growth control. *Proceedings of the National Academy of Sciences of the United States of America*, 105(18), 6632–6637. <https://doi.org/10.1073/pnas.0800970105>
- Ali, S. A., Dobson, J. R., Lian, J. B., Stein, J. L., van Wijnen, A. J., Zaidi, S. K., & Stein, G. S. (2012). A RUNX2-HDAC1 co-repressor complex regulates rRNA gene expression by modulating UBF acetylation. *Journal of cell science*, 125(Pt 11), 2732–2739. <https://doi.org/10.1242/jcs.100909>
- Bierhoff, H., Dammert, M. A., Brocks, D., Dambacher, S., Schotta, G., & Grummt, I. (2014). Quiescence-induced LncRNAs trigger H4K20 trimethylation and transcriptional silencing. *Molecular cell*, 54(4), 675–682. <https://doi.org/10.1016/j.molcel.2014.03.032>
- Bird, A., Taggart, M., & Macleod, D. (1981). Loss of rDNA methylation accompanies the onset of ribosomal gene activity in early development of *X. laevis*. *Cell*, 26(3 Pt 1), 381–390. [https://doi.org/10.1016/0092-8674\(81\)90207-5](https://doi.org/10.1016/0092-8674(81)90207-5)
- Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature biotechnology*, 34(5), 525–527. <https://doi.org/10.1038/nbt.3519>
- Burke, W. D., Calalang, C. C., & Eickbush, T. H. (1987). The site-specific ribosomal insertion element type II of *Bombyx mori* (R2Bm) contains the coding sequence for a reverse transcriptase-like enzyme. *Molecular and cellular biology*, 7(6), 2221–2230. <https://doi.org/10.1128/mcb.7.6.2221>
- de Capoa, A., Marlekaj, P., Baldini, A., Archidiacono, N., & Rocchi, M. (1985a). The transcriptional activity of individual ribosomal DNA gene clusters is modulated by serum concentration. *Journal of cell science*, 74, 21–35.
- de Capoa, A., Marlekaj, P., Baldini, A., Rocchi, M., & Archidiacono, N. (1985b). Cytologic demonstration of differential activity of rRNA gene clusters in different human tissues. *Human genetics*, 69(3), 212–217. <https://doi.org/10.1007/BF00293027>
- Chen, Y. A., Stuwe, E., Luo, Y., Ninova, M., Le Thomas, A., Rozhavskaia, E., Li, S., Vempati, S., Laver, J. D., Patel, D. J., Smibert, C. A., Lipshitz, H. D., Toth, K. F., & Aravin, A. A. (2016). Cutoff Suppresses RNA Polymerase II Termination to Ensure Expression of piRNA Precursors. *Molecular cell*, 63(1), 97–109. <https://doi.org/10.1016/j.molcel.2016.05.010>
- Choi, H., Schwarzkopf, M., Fornace, M. E., Acharya, A., Artavanis, G., Stegmaier, J., Cunha, A., & Pierce, N. A. (2018). Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. *Development (Cambridge, England)*, 145(12), dev165753. <https://doi.org/10.1242/dev.165753>

Coffman, F. D., He, M., Diaz, M. L., & Cohen, S. (2005). DNA replication initiates at different sites in early and late S phase within human ribosomal RNA genes. *Cell cycle (Georgetown, Tex.)*, 4(9), 1223–1226. <https://doi.org/10.4161/cc.4.9.1961>

Conconi, A., Widmer, R. M., Koller, T., & Sogo, J. M. (1989). Two different chromatin structures coexist in ribosomal RNA genes throughout the cell cycle. *Cell*, 57(5), 753–761. [https://doi.org/10.1016/0092-8674\(89\)90790-3](https://doi.org/10.1016/0092-8674(89)90790-3)

Dammann, R., Lucchini, R., Koller, T., & Sogo, J. M. (1993). Chromatin structures and transcription of rDNA in yeast *Saccharomyces cerevisiae*. *Nucleic acids research*, 21(10), 2331–2338. <https://doi.org/10.1093/nar/21.10.2331>

Earley, K., Lawrence, R. J., Pontes, O., Reuther, R., Enciso, A. J., Silva, M., Neves, N., Gross, M., Viegas, W., & Pikaard, C. S. (2006). Erasure of histone acetylation by *Arabidopsis* HDA6 mediates large-scale gene silencing in nucleolar dominance. *Genes & development*, 20(10), 1283–1293. <https://doi.org/10.1101/gad.1417706>

Eickbush, D. G., & Eickbush, T. H. (2003). Transcription of endogenous and exogenous R2 elements in the rRNA gene locus of *Drosophila melanogaster*. *Molecular and cellular biology*, 23(11), 3825–3836. <https://doi.org/10.1128/mcb.23.11.3825-3836.2003>

Eickbush, D. G., & Eickbush, T. H. (2010). R2 retrotransposons encode a self-cleaving ribozyme for processing from an rRNA cotranscript. *Molecular and cellular biology*, 30(13), 3142–3150. <https://doi.org/10.1128/MCB.00300-10>

Eickbush, T. H., & Eickbush, D. G. (2007). Finely orchestrated movements: evolution of the ribosomal RNA genes. *Genetics*, 175(2), 477–485. <https://doi.org/10.1534/genetics.107.071399>

Eickbush, T. H., & Eickbush, D. G. (2015). Integration, Regulation, and Long-Term Stability of R2 Retrotransposons. *Microbiology spectrum*, 3(2), MDNA3–2014. <https://doi.org/10.1128/microbiolspec.MDNA3-0011-2014>

Eickbush, D. G., Burke, W. D., & Eickbush, T. H. (2013). Evolution of the R2 retrotransposon ribozyme and its self-cleavage site. *PloS one*, 8(9), e66441. <https://doi.org/10.1371/journal.pone.0066441>

Finkbeiner, E., Haindl, M., Raman, N., & Muller, S. (2011). SUMO routes ribosome maturation. *Nucleus (Austin, Tex.)*, 2(6), 527–532. <https://doi.org/10.4161/nucl.2.6.17604>

Flavell, R. B., O'Dell, M., & Thompson, W. F. (1988). Regulation of cytosine methylation in ribosomal DNA and nucleolus organizer expression in wheat. *Journal of molecular biology*, 204(3), 523–534. [https://doi.org/10.1016/0022-2836\(88\)90352-x](https://doi.org/10.1016/0022-2836(88)90352-x)

Foe V. E. (1978). Modulation of ribosomal RNA synthesis in *Oncopeltus fasciatus*: an electron microscopic study of the relationship between changes in chromatin structure and transcriptional activity. *Cold Spring Harbor symposia on quantitative biology*, 42 Pt 2, 723–740. <https://doi.org/10.1101/sqb.1978.042.01.074>

Ganley, A. R., & Kobayashi, T. (2007). Highly efficient concerted evolution in the ribosomal DNA repeats: total rDNA repeat variation revealed by whole-genome shotgun sequence data. *Genome research*, 17(2), 184–191. <https://doi.org/10.1101/gr.5457707>

Geiss-Friedlander, R., & Melchior, F. (2007). Concepts in sumoylation: a decade on. *Nature reviews. Molecular cell biology*, 8(12), 947–956. <https://doi.org/10.1038/nrm2293>

Gill G. (2005). Something about SUMO inhibits transcription. *Current opinion in genetics & development*, 15(5), 536–541. <https://doi.org/10.1016/j.gde.2005.07.004>

Gonzalez, I., Mateos-Langerak, J., Thomas, A., Cheutin, T., & Cavalli, G. (2014). Identification of regulators of the three-dimensional polycomb organization by a microscopy-based genome-wide RNAi screen. *Molecular cell*, 54(3), 485–499. <https://doi.org/10.1016/j.molcel.2014.03.004>

Greil, F., & Ahmad, K. (2012). Nucleolar dominance of the Y chromosome in *Drosophila melanogaster*. *Genetics*, 191(4), 1119–1128. <https://doi.org/10.1534/genetics.112.141242>

Guerrero, P. A., & Maggert, K. A. (2011). The CCCTC-binding factor (CTCF) of *Drosophila* contributes to the regulation of the ribosomal DNA and nucleolar stability. *PLoS one*, 6(1), e16401. <https://doi.org/10.1371/journal.pone.0016401>

Guo, D., Li, M., Zhang, Y., Yang, P., Eckenrode, S., Hopkins, D., Zheng, W., Purohit, S., Podolsky, R. H., Muir, A., Wang, J., Dong, Z., Brusko, T., Atkinson, M., Pozzilli, P., Zeidler, A., Raffel, L. J., Jacob, C. O., Park, Y., Serrano-Rios, M., ... Wang, C. Y. (2004). A functional variant of SUMO4, a new I kappa B alpha modifier, is associated with type 1 diabetes. *Nature genetics*, 36(8), 837–841. <https://doi.org/10.1038/ng1391>

Haindl, M., Harasim, T., Eick, D., & Muller, S. (2008). The nucleolar SUMO-specific protease SENP3 reverses SUMO modification of nucleophosmin and is required for rRNA processing. *EMBO reports*, 9(3), 273–279. <https://doi.org/10.1038/embo.2008.3>

Handu, M., Kaduskar, B., Ravindranathan, R., Soory, A., Giri, R., Elango, V. B., Gowda, H., & Ratnaparkhi, G. S. (2015). SUMO-Enriched Proteome for *Drosophila* Innate Immune Response. *G3 (Bethesda, Md.)*, 5(10), 2137–2154. <https://doi.org/10.1534/g3.115.020958>

Hay R. T. (2005). SUMO: a history of modification. *Molecular cell*, 18(1), 1–12. <https://doi.org/10.1016/j.molcel.2005.03.012>

Heun P. (2007). SUMO Organization of the nucleus. *Current opinion in cell biology*, 19(3), 350–355. <https://doi.org/10.1016/j.ceb.2007.04.014>

Henras, A. K., Plisson-Chastang, C., O'Donohue, M. F., Chakraborty, A., & Gleizes, P. E. (2015). An overview of pre-ribosomal RNA processing in eukaryotes. *Wiley interdisciplinary reviews. RNA*, 6(2), 225–242. <https://doi.org/10.1002/wrna.1269>

Hur, J. K., Luo, Y., Moon, S., Ninova, M., Marinov, G. K., Chung, Y. D., & Aravin, A. A. (2016). Splicing-independent loading of TREX on nascent RNA is required for efficient expression of dual-strand piRNA clusters in *Drosophila*. *Genes & development*, 30(7), 840–855. <https://doi.org/10.1101/gad.276030.115>

Impens, F., Radoshevich, L., Cossart, P., & Ribet, D. (2014). Mapping of SUMO sites and analysis of SUMOylation changes induced by external stimuli. *Proceedings of the National Academy of Sciences of the United States of America*, 111(34), 12432–12437. <https://doi.org/10.1073/pnas.1413825111>

Jakubczak, J. L., Burke, W. D., & Eickbush, T. H. (1991). Retrotransposable elements R1 and R2 interrupt the rRNA genes of most insects. *Proceedings of the National Academy of Sciences of the United States of America*, 88(8), 3295–3299. <https://doi.org/10.1073/pnas.88.8.3295>

Jakubczak, J. L., Xiong, Y., & Eickbush, T. H. (1990). Type I (R1) and type II (R2) ribosomal DNA insertions of *Drosophila melanogaster* are retrotransposable elements closely related to those of *Bombyx mori*. *Journal of molecular biology*, 212(1), 37–52. [https://doi.org/10.1016/0022-2836\(90\)90303-4](https://doi.org/10.1016/0022-2836(90)90303-4)

Jakubczak, J. L., Zenni, M. K., Woodruff, R. C., & Eickbush, T. H. (1992). Turnover of R1 (type I) and R2 (type II) retrotransposable elements in the ribosomal DNA of *Drosophila melanogaster*. *Genetics*, 131(1), 129–142.

Jamrich, M., & Miller, O. L., Jr (1984). The rare transcripts of interrupted rRNA genes in *Drosophila melanogaster* are processed or degraded during synthesis. *The EMBO journal*, 3(7), 1541–1545.

Jentsch, S., & Psakhye, I. (2013). Control of nuclear activities by substrate-selective and protein-group SUMOylation. *Annual review of genetics*, 47, 167–186. <https://doi.org/10.1146/annurev-genet-111212-133453>

Johnson, E. S., & Blobel, G. (1997). Ubc9p is the conjugating enzyme for the ubiquitin-like protein Smt3p. *The Journal of biological chemistry*, 272(43), 26799–26802. <https://doi.org/10.1074/jbc.272.43.26799>

Johnson, E. S., Schwienhorst, I., Dohmen, R. J., & Blobel, G. (1997). The ubiquitin-like protein Smt3p is activated for conjugation to other proteins by an Aos1p/Uba2p heterodimer. *The EMBO journal*, 16(18), 5509–5519. <https://doi.org/10.1093/emboj/16.18.5509>

Karpen, G. H., Schaefer, J. E., & Laird, C. D. (1988). A *Drosophila* rRNA gene located in euchromatin is active in transcription and nucleolus formation. *Genes & development*, 2(12B), 1745–1763. <https://doi.org/10.1101/gad.2.12b.1745>

Lee, H. Y., Haurwitz, R. E., Apffel, A., Zhou, K., Smart, B., Wenger, C. D., Laderman, S., Bruhn, L., & Doudna, J. A. (2013). RNA-protein analysis using a conditional CRISPR nuclease. *Proceedings of the National Academy of Sciences of the United States of America*, 110(14), 5416–5421. <https://doi.org/10.1073/pnas.1302807110>

Lehembre, F., Badenhorst, P., Müller, S., Travers, A., Schweisguth, F., & Dejean, A. (2000). Covalent modification of the transcriptional repressor tramtrack by the ubiquitin-related protein Smt3 in *Drosophila* flies. *Molecular and cellular biology*, 20(3), 1072–1082. <https://doi.org/10.1128/mcb.20.3.1072-1082.2000>

Li, J., Längst, G., & Grummt, I. (2006). NoRC-dependent nucleosome positioning silences rRNA genes. *The EMBO journal*, 25(24), 5735–5741. <https://doi.org/10.1038/sj.emboj.7601454>

Li, X., Lee, Y. K., Jeng, J. C., Yen, Y., Schultz, D. C., Shih, H. M., & Ann, D. K. (2007). Role for KAP1 serine 824 phosphorylation and sumoylation/desumoylation switch in

regulating KAP1-mediated transcriptional repression. *The Journal of biological chemistry*, 282(50), 36177–36189. <https://doi.org/10.1074/jbc.M706912200>

Long, E. O., Collins, M., Kiefer, B. I., & Dawid, I. B. (1981). Expression of the ribosomal DNA insertions in bobbed mutants of *Drosophila melanogaster*. *Molecular & general genetics : MGG*, 182(3), 377–384. <https://doi.org/10.1007/BF00293925>

Long, E. O., & Dawid, I. B. (1979). Expression of ribosomal DNA insertions in *Drosophila melanogaster*. *Cell*, 18(4), 1185–1196. [https://doi.org/10.1016/0092-8674\(79\)90231-9](https://doi.org/10.1016/0092-8674(79)90231-9)

Mandal R. K. (1984). The organization and transcription of eukaryotic ribosomal RNA genes. *Progress in nucleic acid research and molecular biology*, 31, 115–160. [https://doi.org/10.1016/s0079-6603\(08\)60376-1](https://doi.org/10.1016/s0079-6603(08)60376-1)

MacPherson, M. J., Beatty, L. G., Zhou, W., Du, M., & Sadowski, P. D. (2009). The CTCF insulator protein is posttranslationally modified by SUMO. *Molecular and cellular biology*, 29(3), 714–725. <https://doi.org/10.1128/MCB.00825-08>

Neyret-Kahn, H., Benhamed, M., Ye, T., Le Gras, S., Cossec, J. C., Lapaquette, P., Bischof, O., Ouspenskaia, M., Dasso, M., Seeler, J., Davidson, I., & Dejean, A. (2013). Sumoylation at chromatin governs coordinated repression of a transcriptional program essential for cell growth and proliferation. *Genome research*, 23(10), 1563–1579. <https://doi.org/10.1101/gr.154872.113>

Nie, M., Xie, Y., Loo, J. A., & Courey, A. J. (2009). Genetic and proteomic evidence for roles of *Drosophila* SUMO in cell cycle control, Ras signaling, and early pattern formation. *PloS one*, 4(6), e5905. <https://doi.org/10.1371/journal.pone.0005905>

Ninova, M., Chen, Y. A., Godneeva, B., Rogers, A. K., Luo, Y., Fejes Tóth, K., & Aravin, A. A. (2020a). Su(var)2-10 and the SUMO Pathway Link piRNA-Guided Target Recognition to Chromatin Silencing. *Molecular cell*, 77(3), 556–570.e6. <https://doi.org/10.1016/j.molcel.2019.11.012>

Ninova, M., Godneeva, B., Chen, Y. A., Luo, Y., Prakash, S. J., Jankovics, F., Erdélyi, M., Aravin, A. A., & Fejes Tóth, K. (2020b). The SUMO Ligase Su(var)2-10 Controls Hetero- and Euchromatic Gene Expression via Establishing H3K9 Trimethylation and Negative Feedback Regulation. *Molecular cell*, 77(3), 571–585.e4. <https://doi.org/10.1016/j.molcel.2019.09.033>

Mayer, C., Schmitz, K. M., Li, J., Grummt, I., & Santoro, R. (2006). Intergenic transcripts regulate the epigenetic state of rRNA genes. *Molecular cell*, 22(3), 351–361. <https://doi.org/10.1016/j.molcel.2006.03.028>

McKnight, S. L., & Miller, O. L., Jr (1976). Ultrastructural patterns of RNA synthesis during early embryogenesis of *Drosophila melanogaster*. *Cell*, 8(2), 305–319. [https://doi.org/10.1016/0092-8674\(76\)90014-3](https://doi.org/10.1016/0092-8674(76)90014-3)

Melchior F. (2000). SUMO--nonclassical ubiquitin. *Annual review of cell and developmental biology*, 16, 591–626. <https://doi.org/10.1146/annurev.cellbio.16.1.591>

Morgan, G. T., Reeder, R. H., & Bakken, A. H. (1983). Transcription in cloned spacers of *Xenopus laevis* ribosomal DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 80(21), 6490–6494. <https://doi.org/10.1073/pnas.80.21.6490>

Nathan, D., Sterner, D. E., & Berger, S. L. (2003). Histone modifications: Now summoning sumoylation. *Proceedings of the National Academy of Sciences of the United States of America*, 100(23), 13118–13120. <https://doi.org/10.1073/pnas.2436173100>

Nacerddine, K., Lehembre, F., Bhaumik, M., Artus, J., Cohen-Tannoudji, M., Babinet, C., Pandolfi, P. P., & Dejean, A. (2005). The SUMO pathway is essential for nuclear integrity and chromosome segregation in mice. *Developmental cell*, 9(6), 769–779. <https://doi.org/10.1016/j.devcel.2005.10.007>

Németh, A., & Längst, G. (2011). Genome organization in and around the nucleolus. *Trends in genetics : TIG*, 27(4), 149–156. <https://doi.org/10.1016/j.tig.2011.01.002>

Nie, M., Xie, Y., Loo, J. A., & Courey, A. J. (2009). Genetic and proteomic evidence for roles of *Drosophila* SUMO in cell cycle control, Ras signaling, and early pattern formation. *PloS one*, 4(6), e5905. <https://doi.org/10.1371/journal.pone.0005905>

Peng, Y., Wang, Z., Wang, Z., Yu, F., Li, J., & Wong, J. (2019). SUMOylation down-regulates rDNA transcription by repressing expression of upstream-binding factor and proto-oncogene c-Myc. *The Journal of biological chemistry*, 294(50), 19155–19166. <https://doi.org/10.1074/jbc.RA119.010624>

Pimentel, H., Bray, N. L., Puente, S., Melsted, P., & Pachter, L. (2017). Differential analysis of RNA-seq incorporating quantification uncertainty. *Nature methods*, 14(7), 687–690. <https://doi.org/10.1038/nmeth.4324>

Psakhye, I., & Jentsch, S. (2012). Protein group modification and synergy in the SUMO pathway as exemplified in DNA repair. *Cell*, 151(4), 807–820. <https://doi.org/10.1016/j.cell.2012.10.021>

Rodriguez, M. S., Dargemont, C., & Hay, R. T. (2001). SUMO-1 conjugation in vivo requires both a consensus modification motif and nuclear targeting. *The Journal of biological chemistry*, 276(16), 12654–12659. <https://doi.org/10.1074/jbc.M009476200>

Rogers, S. L., & Rogers, G. C. (2008). Culture of *Drosophila* S2 cells and their use for RNAi-mediated loss-of-function studies and immunofluorescence microscopy. *Nature protocols*, 3(4), 606–611. <https://doi.org/10.1038/nprot.2008.18>

Sampson, D. A., Wang, M., & Matunis, M. J. (2001). The small ubiquitin-like modifier-1 (SUMO-1) consensus sequence mediates Ubc9 binding and is essential for SUMO-1 modification. *The Journal of biological chemistry*, 276(24), 21664–21669. <https://doi.org/10.1074/jbc.M100006200>

Santoro, R., & Grummt, I. (2001). Molecular mechanisms mediating methylation-dependent silencing of ribosomal gene transcription. *Molecular cell*, 8(3), 719–725. [https://doi.org/10.1016/s1097-2765\(01\)00317-3](https://doi.org/10.1016/s1097-2765(01)00317-3)

Santoro, R., Li, J., & Grummt, I. (2002). The nucleolar remodeling complex NoRC mediates heterochromatin formation and silencing of ribosomal gene transcription. *Nature genetics*, 32(3), 393–396. <https://doi.org/10.1038/ng1010>

Santoro, R., Schmitz, K. M., Sandoval, J., & Grummt, I. (2010). Intergenic transcripts originating from a subclass of ribosomal DNA repeats silence ribosomal RNA genes in trans. *EMBO reports*, 11(1), 52–58. <https://doi.org/10.1038/embor.2009.254>

Shiio, Y., & Eisenman, R. N. (2003). Histone sumoylation is associated with transcriptional repression. *Proceedings of the National Academy of Sciences of the United States of America*, 100(23), 13225–13230. <https://doi.org/10.1073/pnas.1735528100>

Sienski, G., Batki, J., Senti, K. A., Dönertas, D., Tirian, L., Meixner, K., & Brennecke, J. (2015). Silencio/CG9754 connects the Piwi-piRNA complex to the cellular heterochromatin machinery. *Genes & development*, 29(21), 2258–2271. <https://doi.org/10.1101/gad.271908.115>

Smith, M., Mallin, D. R., Simon, J. A., & Courey, A. J. (2011). Small ubiquitin-like modifier (SUMO) conjugation impedes transcriptional silencing by the polycomb group repressor Sex Comb on Midleg. *The Journal of biological chemistry*, 286(13), 11391–11400. <https://doi.org/10.1074/jbc.M110.214569>

Sogo, J. M., Ness, P. J., Widmer, R. M., Parish, R. W., & Koller, T. (1984). Psoralen-crosslinking of DNA as a probe for the structure of active nucleolar chromatin. *Journal of molecular biology*, 178(4), 897–919. [https://doi.org/10.1016/0022-2836\(84\)90318-8](https://doi.org/10.1016/0022-2836(84)90318-8)

Stielow, B., Sapetschnig, A., Krüger, I., Kunert, N., Brehm, A., Boutros, M., & Suske, G. (2008). Identification of SUMO-dependent chromatin-associated transcriptional repression components by a genome-wide RNAi screen. *Molecular cell*, 29(6), 742–754. <https://doi.org/10.1016/j.molcel.2007.12.032>

Strohner, R., Nemeth, A., Jansa, P., Hofmann-Rohrer, U., Santoro, R., Längst, G., & Grummt, I. (2001). NoRC--a novel member of mammalian ISWI-containing chromatin remodeling machines. *The EMBO journal*, 20(17), 4892–4900. <https://doi.org/10.1093/emboj/20.17.4892>

Takahashi, Y., Dulev, S., Liu, X., Hiller, N. J., Zhao, X., & Strunnikov, A. (2008). Cooperation of sumoylated chromosomal proteins in rDNA maintenance. *PLoS genetics*, 4(10), e1000215. <https://doi.org/10.1371/journal.pgen.1000215>

Terracol R. (1986). Transcription of rDNA insertions in bobbed mutants of *Drosophila melanogaster*. *Genetical research*, 48(3), 167–174. <https://doi.org/10.1017/s0016672300024964>

Tseng, H., Chou, W., Wang, J., Zhang, X., Zhang, S., & Schultz, R. M. (2008). Mouse ribosomal RNA genes contain multiple differentially regulated variants. *PloS one*, 3(3), e1843. <https://doi.org/10.1371/journal.pone.0001843>

Urieli-Shoval, S., Gruenbaum, Y., Sedat, J., & Razin, A. (1982). The absence of detectable methylated bases in *Drosophila melanogaster* DNA. *FEBS letters*, 146(1), 148–152. [https://doi.org/10.1016/0014-5793\(82\)80723-0](https://doi.org/10.1016/0014-5793(82)80723-0)

Verger, A., Perdomo, J., & Crossley, M. (2003). Modification with SUMO. A role in transcriptional regulation. *EMBO reports*, 4(2), 137–142. <https://doi.org/10.1038/sj.embor.embor738>

Westman, B. J., Verheggen, C., Hutten, S., Lam, Y. W., Bertrand, E., & Lamond, A. I. (2010). A proteomic screen for nucleolar SUMO targets shows SUMOylation modulates the function of Nop5/Nop58. *Molecular cell*, 39(4), 618–631. <https://doi.org/10.1016/j.molcel.2010.07.025>

Xie, W., Ling, T., Zhou, Y., Feng, W., Zhu, Q., Stunnenberg, H. G., Grummt, I., & Tao, W. (2012). The chromatin remodeling complex NuRD establishes the poised state of rRNA genes characterized by bivalent histone modifications and altered nucleosome positions. *Proceedings of the National Academy of Sciences of the United States of America*, 109(21), 8161–8166. <https://doi.org/10.1073/pnas.1201262109>

Xiong, Y., Burke, W. D., Jakubczak, J. L., & Eickbush, T. H. (1988). Ribosomal DNA insertion elements R1Bm and R2Bm can transpose in a sequence specific manner to locations outside the 28S genes. *Nucleic acids research*, 16(22), 10561–10573. <https://doi.org/10.1093/nar/16.22.10561>

Yang, S. H., Jaffray, E., Hay, R. T., & Sharrocks, A. D. (2003). Dynamic interplay of the SUMO and ERK pathways in regulating Elk-1 transcriptional activity. *Molecular cell*, 12(1), 63–74. [https://doi.org/10.1016/s1097-2765\(03\)00265-x](https://doi.org/10.1016/s1097-2765(03)00265-x)

Yang, S. H., & Sharrocks, A. D. (2004). SUMO promotes HDAC-mediated transcriptional repression. *Molecular cell*, 13(4), 611–617. [https://doi.org/10.1016/s1097-2765\(04\)00060-7](https://doi.org/10.1016/s1097-2765(04)00060-7)

Yang, J., Malik, H. S., & Eickbush, T. H. (1999). Identification of the endonuclease domain encoded by R2 and other site-specific, non-long terminal repeat retrotransposable elements. *Proceedings of the National Academy of Sciences of the United States of America*, 96(14), 7847–7852. <https://doi.org/10.1073/pnas.96.14.7847>

Ye, J., & Eickbush, T. H. (2006). Chromatin structure and transcription of the R1- and R2-inserted rRNA genes of *Drosophila melanogaster*. *Molecular and cellular biology*, 26(23), 8781–8790. <https://doi.org/10.1128/MCB.01409-06>

Fujiwara, A., Abe, S., Yamaha, E., Yamazaki, F., & Yoshida, M. C. (1998). Chromosomal localization and heterochromatin association of ribosomal RNA gene loci and silver-stained nucleolar organizer regions in salmonid fishes. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology*, 6(6), 463–471. <https://doi.org/10.1023/a:1009200428369>

Young, D. W., Hassan, M. Q., Pratap, J., Galindo, M., Zaidi, S. K., Lee, S. H., Yang, X., Xie, R., Javed, A., Underwood, J. M., Furcinitti, P., Imbalzano, A. N., Penman, S., Nickerson, J. A., Montecino, M. A., Lian, J. B., Stein, J. L., van Wijnen, A. J., & Stein, G. S. (2007). Mitotic occupancy and lineage-specific transcriptional control of rRNA genes by Runx2. *Nature*, 445(7126), 442–446. <https://doi.org/10.1038/nature05473>

Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics : a journal of integrative biology*, 16(5), 284–287. <https://doi.org/10.1089/omi.2011.0118>

Yun, C., Wang, Y., Mukhopadhyay, D., Backlund, P., Kolli, N., Yergey, A., Wilkinson, K. D., & Dasso, M. (2008). Nucleolar protein B23/nucleophosmin regulates the vertebrate SUMO pathway through SENP3 and SENP5 proteases. *The Journal of cell biology*, 183(4), 589–595. <https://doi.org/10.1083/jcb.200807185>

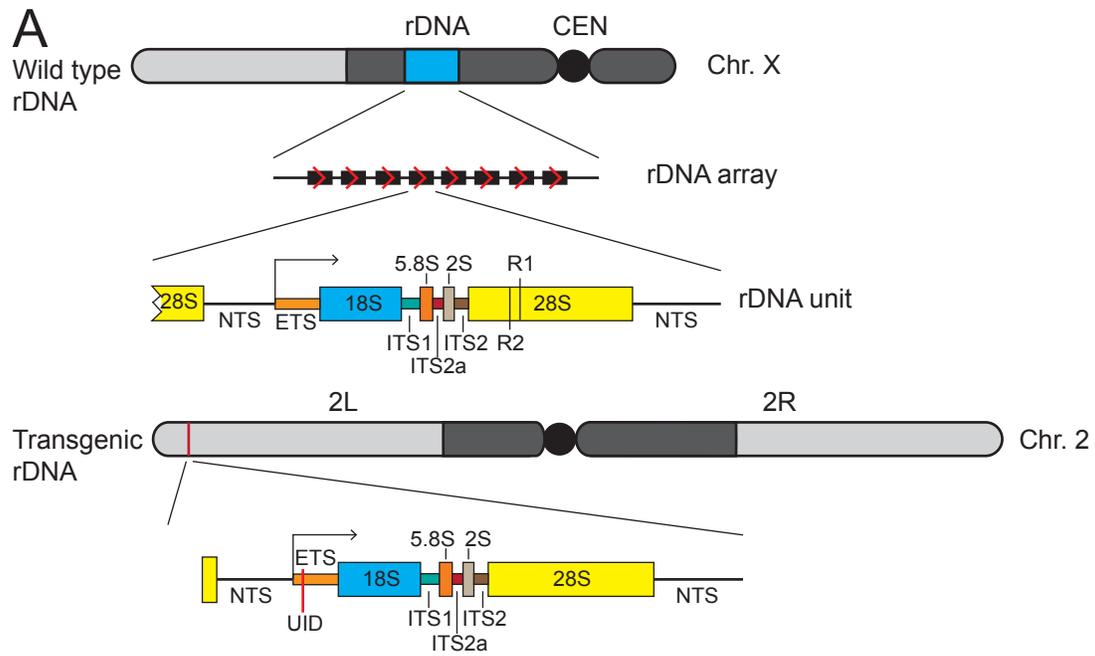
Zhang, Q., Shalaby, N. A., & Buszczak, M. (2014). Changes in rRNA transcription influence proliferation and cell fate within a stem cell lineage. *Science (New York, N.Y.)*, 343(6168), 298–301. <https://doi.org/10.1126/science.1246384>

Zhang, S., Wang, J., & Tseng, H. (2007). Basonuclin regulates a subset of ribosomal RNA genes in HaCaT cells. *PloS one*, 2(9), e902. <https://doi.org/10.1371/journal.pone.0000902>

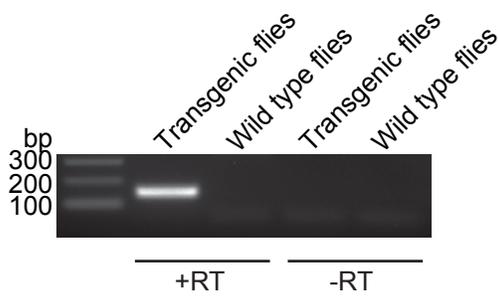
Zhao, X., & Blobel, G. (2005). A SUMO ligase is part of a nuclear multiprotein complex that affects DNA repair and chromosomal organization. *Proceedings of the National Academy of Sciences of the United States of America*, 102(13), 4777–4782. <https://doi.org/10.1073/pnas.0500537102>

Zhou, J., Sackton, T. B., Martinsen, L., Lemos, B., Eickbush, T. H., & Hartl, D. L. (2012). Y chromosome mediates ribosomal DNA silencing and modulates the chromatin state in *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America*, 109(25), 9941–9946. <https://doi.org/10.1073/pnas.1207367109>

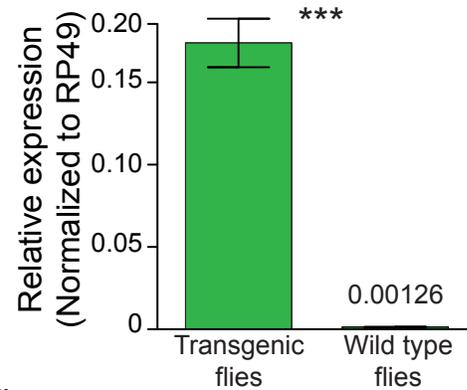
Zhou, Y., Santoro, R., & Grummt, I. (2002). The chromatin remodeling complex NoRC targets HDAC1 to the ribosomal gene promoter and represses RNA polymerase I transcription. *The EMBO journal*, 21(17), 4632–4640. <https://doi.org/10.1093/emboj/cdf460>



**B** UID Primers RT-PCR



**RT-qPCR**



**C** Transgenic flies Wild type flies

