The chromatin factor ROW cooperates with BEAF-32 in regulating long-range genes

Neta Herman[1], Sebastian Kadener[2,*] and Sagiv Shifman[1,*]

[1] Department of Genetics, The Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel

[2] Biology Department, Brandeis University, Waltham, MA, USA

[*]Corresponding authors

Abstract

Chromatin insulators are involved in transcription regulation and long-range chromatin contacts. In *Drosophila* the insulator protein BEAF-32 binds to promoters of housekeeping genes and the boundaries of Topological Associated Domains (TAD) and is involved in the regulation of long-range targets by mechanisms that remain incompletely understood. Here we show that Relative-Of-WOC (ROW), which is part of the Heterochromatin Protein 1c (HP1c) complex, is essential for the long-range transcription regulation mediated by the Boundary Element-Associated Factor Of 32kD (BEAF-32). We found that ROW physically interacts with the insulator complex proteins BEAF-32 and Chromator, as well as with HP1c, HP1b, and Without Children (WOC). Moreover, we found that ROW binds AT-rich sequences flanked by BEAF-32 motifs, which are located at promoters of housekeeping genes and in TAD boundaries. In agreement with these results, the genome distribution of ROW strongly correlated with BEAF-32 and WOC, but at a lower level with HP1c and HP1b. Knockdown of *row* resulted in downregulation of genes that are long-range targets of BEAF-32, are indirectly bound by BEAF-32 and ROW, and enriched with factors associated with RNA polymerase II pausing. These results suggest that ROW and BEAF-32 are involved in the recruitment of protein complexes required for transcription activation and long-range contacts between promoters of housekeeping and inducible genes.

Introduction

The genome can be divided at a large scale into A and B compartments, which correspond to open, transcriptionally active chromatin or to closed transcriptionally inactive chromatin, respectively[1]. Chromosomes are further organized at the sub-megabase scale into domains with a high level of interactions within them, known as topologically associating domains (TADs), separated by sharp boundaries with a lower level of interactions between domains [2,3]. TADs are frequently linked with a specific chromatin state and can be divided into active, PcG, HP1, and inactive TADs [3]. The majority of TAD boundaries in *Drosophila* are located at active promoters, however, it is not clear whether genome organization is a result of transcription activity or a mechanism that regulates transcription[4].

The *Drosophila* insulator protein BEAF-32 is bound to TAD boundaries located at promoters of housekeeping genes[5]. Those promoters are characterized by multiple BEAF-32 binding motifs surrounded by AT-rich spacers[6]. BEAF-32 has been implicated in both transcription and genome organization. For example, knockout of BEAF-32 was found to cause defects in the male X polytene chromosome morphology, suggesting that BEAF-32 is involved in chromatin structure or dynamics[7]. However, a previous study showed that depletion of *BEAF-32* does not affect TAD boundaries[8]. BEAF-32 was found to be associated with transcription regulation through multiple mechanisms, including regulation of enhancer-promoter interactions[9], through long-range contacts between promoters bound directly and indirectly by BEAF-32 [10], and by restricting the deposition of methylated H3K9 and the spread of heterochromatin to actively transcribed promoters[6].

A key component in the formation and maintenance of heterochromatin is HP1 – a highly conserved protein that was first identified in *Drosophila* (now called HP1a) [11,12]. *Drosophila* has five HP1 paralogs, three that are ubiquitously expressed (HP1a, HP1b, and HP1c) and two which are specific for the germline (HP1d and HP1e) [13]. Most of what we know about the HP1 family is from studies focused on *Drosophila* HP1a. HP1a is involved in heterochromatin formation and transcription silencing, and bind to H3K9me2/3 [14,15]. The potential functions of HP1b and HP1c are less known. HP1b is distributed at both euchromatin and heterochromatin, whereas HP1c is found mainly at euchromatin[16–20]. Although HP1c can bind H3K9me2/3 *in vitro*, it was shown to be localized at active promoters with poised RNA polymerase II (RNA pol II)[19]. The molecular mechanisms that determine the different genomic distribution of the HP1 paralogs remain mostly unknown.

The *Drosophila* HP1c is in a complex together with two zinc finger proteins, Without Children (WOC), and Relative Of WOC (ROW), as well as with the ubiquitin receptor protein Dsk2[19,21,22]. Recently it was suggested that the localization of the HP1c complex at euchromatin is dependent on ROW[23]. ROW contains protein domains that implicate it in transcription regulation, including multiple zinc-finger (ZNF) motifs, an AT-hook, and a glutamine-rich domain in the C-terminal, that resemble activation domains found in transcription factors [19]. Strikingly, the knockdown of *row*, *woc,* and *HP1c* leads to changes in the expression of a common set of genes [19]. Moreover, several lines of evidence suggest that the complex containing ROW, WOC, and HP1c is involved in transcription activation. First, HP1c interacts with the Facilitates Chromatin Transcription Complex (FACT), to recruits FACT to active genes, and the active form of RNA pol II [24]. Second, HP1c interacts with the ubiquitin receptor protein Dsk2 which is

involved in the positive regulation of transcription [21]. Third, results of chromatin immunoprecipitation (ChIP) in *Drosophila* S2 cells followed by high-throughput sequencing (ChIP-seq) of ROW, WOC, and HP1c revealed localization of the complex around the transcription start sites (TSSs) of actively transcribed genes [21]. Finally, depletion of *row* in S2 cells leads to downregulation of approximately 80% of the genes that are both differentially expressed and are targets of the complex [21]. However, *in vivo* expression analysis with RNAi lines of *row*, *woc,* and *HP1c* (RNA from whole larvae) showed similar numbers of upregulated and downregulated genes [19].

ROW is also an ortholog of *POGZ* - a human risk gene for neurodevelopmental disorders[25] that also interacts with heterochromatin proteins[26]. Although *row* is expressed throughout most *Drosophila* tissues and developmental stages, it displays the highest expression level in the larval central nervous system[25]. Moreover, neuron-specific knockdown of *row* in adult flies affects non-associative learning[25]. Thus, *row* is similar to *POGZ* as both interact with heterochromatin proteins and are involved in neurodevelopment and learning[27].

Here, we comprehensively studied and characterized *row* and its binding partners in *Drosophila*. We found that knockdown of *row* using constitutive promoter results in reduced viability and fertility. Interestingly, we found that in addition to WOC, HP1b, and HP1c, ROW binds to components of the insulator complex, BEAF-32, and Chromator. Moreover, ChIP-seq experiments showed that ROW binds AT-rich sequences located upstream to the transcription start sites of housekeeping genes and flanked by binding motifs of BEAF-32. The genome distribution of ROW was highly correlated with BEAF-32 and significantly enriched at TAD boundaries. Depletion of *row* and *BEAF-32* in S2 cells resulted in a correlated change in gene expression. The

differential expressed genes were more likely to be downregulated, to be indirect targets of ROW and BEAF-32 (without binding sequences), and to be regulated through long-range contacts.

Results

Knockdown of row causes a decrease in survival and fertility

To determine the functions of *row in vivo* we utilized two publicly available UAS-*row*^{RNAi} transgenic fly lines, hereby referred to as *row*$^{RNAi-1}$ and *row*$^{RNAi-2}$. When combined with the ubiquitous *actin5C-GAL4* driver, the progenies carrying the Gal4 driver and *row*^{RNAi} construct show a significant decrease in ROW protein levels in fly heads, which was more substantial in *row*$^{RNAi-1}$ relative to *row*$^{RNAi-2}$ (94% and 87%, respectively; P <0.05) (Fig 1A).

We examined the viability of *row*^{RNAi} flies (Fig 1B). Relative to the expected proportion of 50% (the Gal4 driver line is heterozygous for the insertion), there was a significant reduction in the progeny carrying both the Gal4 driver and expressing *row*$^{RNAi-1}$, (36.4%; $P < 1.1 \times 10^{-4}$), but the reduction in viability was not significant for *row*$^{RNAi-2}$ (46.2%; P =0.15) (Fig 1B). We observed that the lethality occurred at the pupal stage as *row*$^{RNAi-1}$ pupal eclosion was significantly reduced (13% compared to 33% of *row*$^{RNAi-1}$ control, P= 0.0058) (Fig 1C). In addition to development, knockdown of *row* also provoked shorted life span of the flies reaching adulthood for both *row*^{RNAi} lines (P < 0.001, Fig 1D). In males, the effect on life span was very pronounced relative to the controls, while in females we observed a smaller, yet still significant life span reduction (Fig 1D). To test the fertility of the flies lacking *row*, we collected virgin

row^{RNAi} flies, crossed them with WT flies, and counted the number of offspring (Fig 1E). Females and males *row*^{RNAi} flies showed a significant reduction in offspring number compared to control flies ($P < 0.001$ for both *row*^{RNAi}).

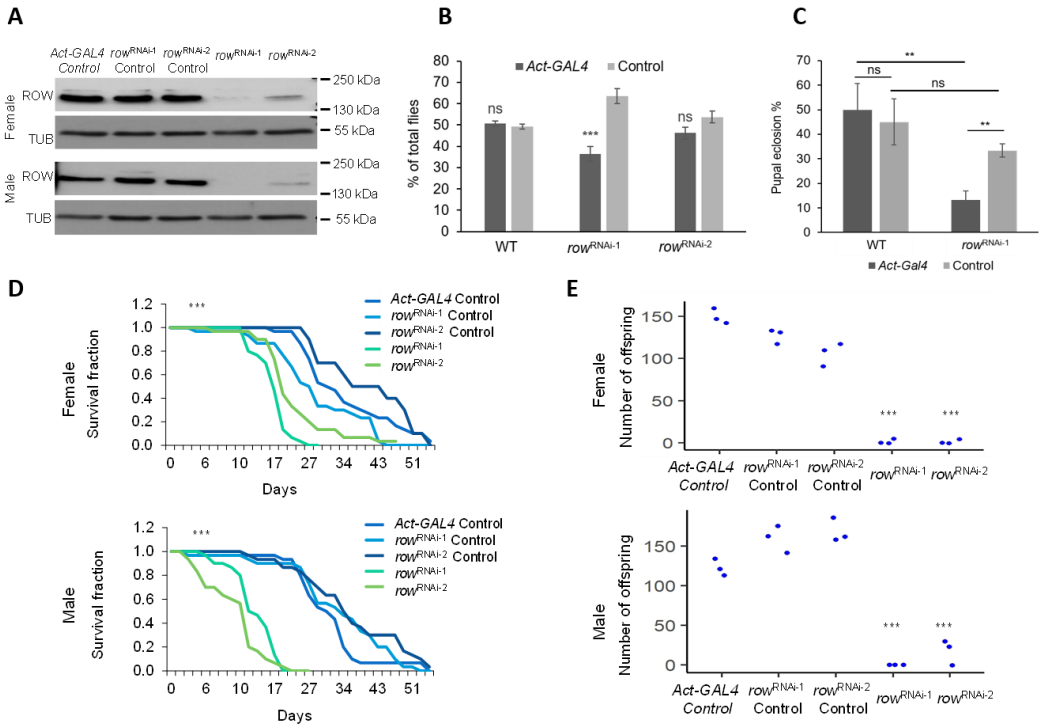


Figure 1. Effect of *row* knockdown on survival and fertility of *Drosophila* flies.

(A) Levels of ROW protein in *row*^{RNAi} lines were decreased in both male and female fly heads. Western blot for the two *row*^{RNAi} lines compared to controls was performed using rat polyclonal αROW and rat αTUBULIN (TUB). Genotype description: *row*$^{RNAi-1}$(*Act-GAL4*/+; UAS-*row*$^{RNAi-1}$/+), *row*$^{RNAi-2}$(*Act-GAL4*/+; UAS-*row*$^{RNAi-2}$/+), *row*$^{RNAi-1}$ Control (Cyo/+; UAS-*row*$^{RNAi-1}$/+), *row*$^{RNAi-2}$ Control (Cyo/+; UAS- *row*$^{RNAi-2}$) and *Act-GAL4* Control (*Act-GAL4* /+).

(B) Decrease in the viability of *row*^{RNAi} lines. To examine the viability of *row*^{RNAi} flies, we counted the offspring generated from the cross between the heterozygous *Act-GAL4*/*Cyo* driver line with the homozygotes *row*^{RNAi} or wild-type (WT: w1118) flies as control. Values are the percentages from the total progeny ± standard error of the mean (SEM).

(C) Decrease in pupal eclosion of *row*^{RNAi} line. Pupae were collected from the cross between the heterozygous *Act-GAL4*/*Cyo* driver line with the homozygotes *row*$^{RNAi-1}$ or wild-type flies as control. Values are the percentages for each genotype of pupal eclosion ± SEM.

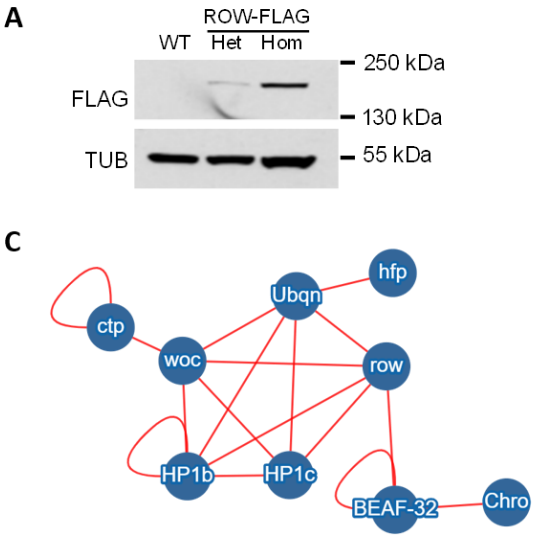
(D) Survival curve for *row*^{RNAi} flies showing reduced life span compared to controls.

(E) Crossing of females and males *row*^{RNAi} flies with WT (w1118) flies resulted in a strong reduction in offspring number. Values are the number of offspring that resulted from crosses between WT flies and flies with different genotypes.
Significance is represented by: **, $P < 0.01$; ***, $P < 0.001$; ns, non-significant.

ROW binds in vivo to HP1b/HP1c proteins and components of insulator complexes, BEAF-32 and Chromator

To study the interactors of ROW *in vivo*, we utilized CRISPR technology to FLAG-tag the endogenous ROW protein in flies. We validated the tagging with western blot (Fig 2A) and sequencing (Fig S1A). The tagging does not affect the levels of ROW protein (Fig S1B). We then utilized these flies to identify ROW interacting proteins by performing affinity-purification of ROW-containing protein complexes from fly heads followed by mass spectrometry analysis. We found five proteins that were co-purified with ROW in all the three independent experiments but were not detected in the control experiments (Fig 2B), and another 32 proteins that were co-purified with ROW in two out of the three experiments (Fig S1C). The five most significant proteins were HP1c, HP1b the zinc-finger finger protein WOC, the extraproteasomal ubiquitin receptor Dsk2 (known also as Ubqn), and the subunit of the cytoplasmic Dynein, Ctp (Fig 2B). We used the molecular interaction search tool (MIST) [28] to identify protein interactions that are supported by additional evidence from previous studies. The analysis provided evidence for two highly connected complexes (Fig S1D). The main complex that included ROW was composed of the five proteins that we identified as high-significant interactors (HP1c, HP1b, WOC, Dsk2, and Ctp), together with the transcription regulator *hfp*, and two components of an insulator complex: the boundary element BEAF-32 and the chromodomain protein, Chromator (Fig 2B-C). The

interaction between ROW and BEAF-32 is supported by two previous studies [29,30]; one of them is a two-hybrid experiment, which suggests a direct interaction between ROW and BEAF-32.



B

Protein name	Control IBAQ	ROW-FLAG IBAQ	Peptides number
ROW	0	7.93E+07	48
WOC	0	7.73E+07	62
HP1c	0	2.44E+08	9
HP1b	0	1.69E+08	11
Ubqn	0	2.54E+07	11
Ctp	0	3.06E+07	2
BEAF-32	0	2.73E+06	3
Chro	0	2.21E+05	2
Hfp	0	1.30E+05	1

Figure 2. Characterization of the interactome of ROW

(A) Western blot for flies expressing endogenous FLAG-tagged ROW (ROW-FLAG), with αFlag tag antibody to validate the tagging. αTubulin (TUB) antibody was used as a reference. Het, Heterozygotes; Hom, Homozygotes for tagged ROW.

(B) Table summarizing the affinity purification–mass spectrometry data. IBAQ (intensity-based absolute quantification) reflects the protein abundance in the sample. Peptides number is the number of Razor and unique peptides. The data are the mean of control (W1118, n=3) and ROW-FLAG flies (n=3) samples. The proteins presented in the table were co-purified with ROW in at least two out of the three experiments and none in the control experiments and are supported by additional evidence from previous studies.

(C) Network of previously identified protein-protein interactions associated with ROW.

ROW binds upstream to the transcription start sites of housekeeping genes

To examine the relationship between the different proteins of the ROW complex, we performed ChIP-seq for ROW, WOC, HP1c, and HP1b utilizing fly heads. We then calculated the pairwise correlation in ChIP-seq signals for the different proteins in non-

overlapping bins across the genome. We found a very strong and significant correlation between the binding sites of all the proteins in the ROW complex ($P < 2.2\times10^{-16}$) (Fig 3A), but the most significant correlation was between ROW and WOC ($r = 0.95$) and between HP1c and HP1b ($r = 0.94$). We also examined the overlap between the binding sites that were identified for each protein (MACS2 peak caller, *q-value* < 0.05). Similar to the quantitative analysis, we identified the strongest overlap between ROW and WOC (82%) and between HP1c and HP1b (76%) (Fig 3B).

We followed by examining the overlap between genomic annotations and ROW binding sites in the ChIP-seq. We found that the majority of ROW binding sites (69.8%) overlap promoter regions (Fig 3C). We observed a slightly lower overlap with promoter regions for WOC (63.8%), and a substantially lower overlap for HP1c and HP1b (43.0% and 35.1%, respectively) (Fig 3C). The binding profile of all four proteins showed similar enrichment of approximately 150 bases upstream of the TSS (Fig 3D). ChIP-seq profile of the four proteins for a representative region is shown in Fig 3E. We then determined the overlap between the ChIP-seq signal of the four proteins and histone marks (modENCODE data). We found that the histone marks of the active transcription, H3K4me3 and H3K27ac, showed a very significant overlap with all the four proteins in the complex (Fig 3F). However, other histone marks that have been linked to active promoters (H3K4me2, H3K9ac, H4K16ac, and H4K12ac) showed very significant overlap only with ROW and WOC (Fig 3F).

Given the strong overlap between ROW and active chromatin marks, we performed ChIP-seq from fly heads using an anti-H3K4me3 antibody. We compared the distribution of H3K4me3 near the TSS of genes bound by ROW to genes unbound by ROW and found a strong H3K4me3 signal flanking the peak of ROW binding site (Fig

3G). Promoters of constitutively expressed genes (housekeeping genes) generally show dispersed transcription from multiple TSSs in a relatively broad region (50-100 bases) and are typical with periodic arrays of nucleosomes. Promoters of regulated genes are with focused transcription and lack of organized chromatin structure [31–33]. To characterize the types of promoters bound by ROW we performed micrococcal nuclease followed by sequencing (MNase-seq). The nucleosome organization of promoters bound by ROW was typical for active, housekeeping genes, with a wide nucleosome-free region (NFR) near the TSS and a regular array of nucleosomes upstream to the TSS (Fig 3H). To pursue the link between ROW and transcription further, we defined ROW-associated genes based on the binding of ROW to their promoter region (Table S1). Based on RNA-seq we performed from fly heads, we compared the relative expression levels of ROW-associated and not-associated genes. We found that ROW-associated genes show a significantly higher expression level ($P < 2.2 \times 10^{-16}$) (Fig 3I). Furthermore, we found that ROW-associated genes tend to be housekeeping genes that are constitutively expressed, as they show high expression with low variation across developmental stages ($P < 2.2 \times 10^{-16}$) (Fig 3J).

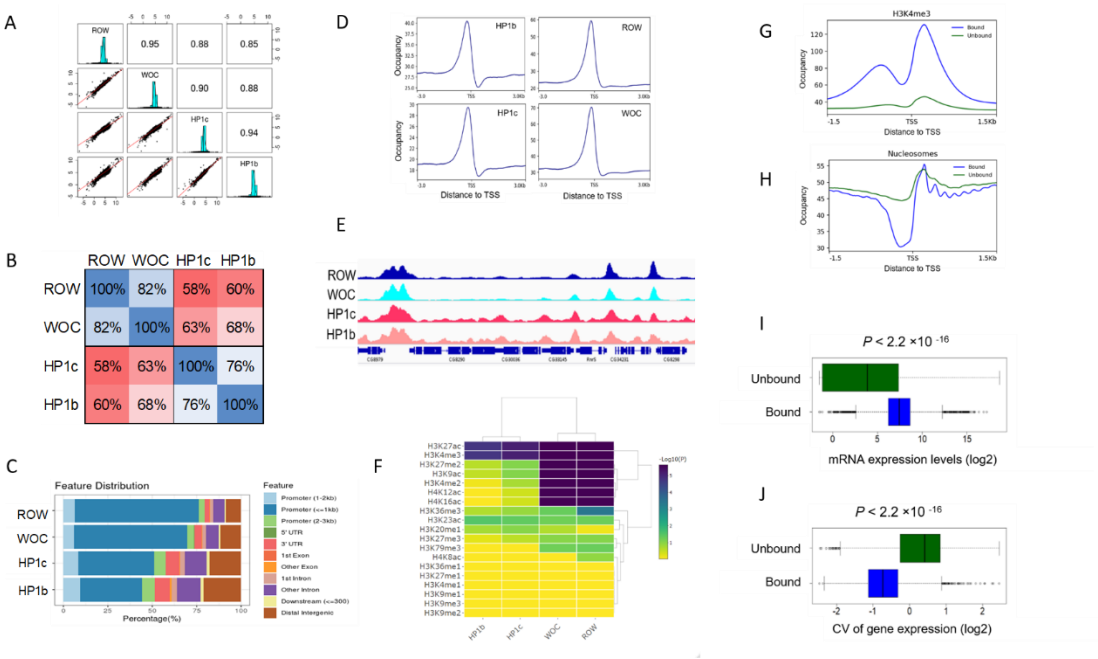


Figure 3. ROW binds to promoters of housekeeping genes

(A) Pairwise correlation between the ChIP-seq signals of ROW, WOC, HP1c, and HP1b in non-overlapping bins of 2000 bases across the genome.

(B). Percentage of overlap between the binding sites of ROW, WOC, HP1c, and HP1b.

(C) Genomic annotation of ROW, WOC, HP1c, and HP1b binding sites.

(D) Average signal profiles (metagene plot) of ROW, WOC, HP1c, and HP1b over a 6-kb window around TSSs.

(E) ChIP-seq signals for ROW, WOC, HP1c, and HP1b at example genomic region.

(F) Heatmap showing the significance of the overlap between different histone modifications and ROW, WOC, HP1c, and HP1b.

(G) A metagene plot of H3K4me3 over a 3-kb window around TSSs for genes bound and unbound by ROW.

(H) Average nucleosome profile over a 3-kb window around TSSs for genes bound and unbound by ROW.

(I) mRNA expression levels for all genes bound and unbound by ROW. Values are log2 of the normalized reads count based on RNA-seq from control flies.

(J) Coefficient of variance (CV) of mRNA expression levels for genes bound and unbound by ROW across 30 different developmental stages.

ROW binds AT-rich sequences flanked by motifs and binding sites of BEAF-32 at TAD boundaries

To identify the DNA sequences responsible for ROW binding, we searched for enriched DNA motifs within the regions surrounding the ROW binding sites. The most significant enrichment was for AT-rich sequences (MEME-ChIP analysis: *E*-value = 3×10^{-254}), which are located at the center of the ROW peak summit (Fig 4A and S2A). We found a similar enrichment of AT-rich sequences for all the other proteins in the complex (WOC, HP1c, and HP1b) (Fig S2A). We used an available tool for predicting DNA-binding specificities for Cys2His2 zinc fingers [34]; however, the predicted sequences for ROW (Fig S2B) did not resemble any motif that was significantly enriched in the binding sites of ROW.

In addition to the AT-rich motifs, we found significant enrichment for a sequence motif (TATCGA) approximately 100bp from the peak summit of ROW (*E*-value = 8.4×10^{-21}) (Fig 4B). Interestingly, BEAF-32 is known to bind this motif. To test whether ROW and BEAF-32 share binding sites across the genome, we compared the list of genes bound by ROW with those bound by BEAF-32 (identified previously in Kc167 cells [35]) (Table S1). Indeed, 76.6% of the ROW-associated genes are also bound by BEAF-32 (Fig 4C). The association between BEAF-32 and ROW is not only strong but highly specific, as an unbiased search for overlap between ROW-associated genes and genes targets of 84 transcription factors (FlyEnrichr) found that the most significantly enriched factor was BEAF-32 (Table S2). These results suggest a strong interaction between the binding of BEAF-32 and ROW. However, a metagene plot showed that BEAF-32 binding displaying a more pronounced peak closer to the TSS, suggesting that the two proteins act in proximity but not exactly in the same place (Fig 4D).

Since TAD boundaries in *Drosophila* are occupied by BEAF-32, we tested the association of ROW with TAD boundaries (identified in Kc167 cells [8]). Indeed, we found that ROW is enriched at the boundaries of TADs (enrichment coefficient $\hat{\beta}$ = 2.54, p <

1×10^{-20}), and is located on top of the TAD boundaries, similar to what is obtained for BEAF-32 (Fig 4E-F and S2C). The other partners of the ROW complex showed enrichment at TAD boundaries at a lower level ($\hat{\beta}$ = 2.41, 1.73, 1.24 for WOC, HP1c, HP1b, respectively) suggesting a specific and independent role for ROW at the TAD boundaries (Fig 4F). Since the proteins in the complex show correlation in their genome distribution (Fig 3A-B), and this correlation may explain the enrichment at TAD boundaries (Fig S2D), we used multiple logistic regression to test for the independent influence of each protein. We found a very strong decrease in enrichment for all the proteins except ROW (Fig 4F), which suggests that the enrichment of WOC, HP1c, and HP1b at TAD boundaries is due to their correlations with ROW.

The data presented above strongly suggest an important functional relationship between ROW and BEAF-32 at TAD boundaries. However, this analysis may be influenced by the fact that our ChIP-seq of ROW was performed with fly heads, while the BEAF-32 ChIP-seq data was obtained from Kc167 cells. To solve this issue, we performed ChIP-seq for ROW and BEAF-32 in S2 cells. We replicated in this experiment the enrichment of AT-rich sequences and BEAF-32 consensus motifs in the binding sites of both proteins (Fig S2E-F), but with different positions for ROW and BEAF-32 relative to the peak's summit (Fig S2E). We also observed a strong overlap between the genes bound by the two proteins (87.5% of genes bound by BEAF-32 were bound by ROW and 73.8% vice versa) (Fig S2G, Table S3). The ChIP-seq signals of ROW and BEAF-32 were both centered on TAD boundaries but were slightly shifted relative to TSSs (Fig S2H-I). The enrichment at TAD boundaries was higher for BEAF-32 ($\hat{\beta}$ = 3.02, $p < 1\times10^{-20}$) relative to ROW ($\hat{\beta}$ = 2.36 $p < 1\times10^{-20}$) (Fig 4G). Multiple logistic

regression showed a proportional reduction of the beta enrichment coefficient for both BEAF-32 ($\hat{\beta}$ = 1.97, $p < 1\times10^{-20}$) and ROW ($\hat{\beta}$ = 1.45, $p < 1\times10^{-20}$) (Fig 4G). This indicates a similar independent association of BEAF-32 and ROW at TAD boundaries and TSSs of active genes.

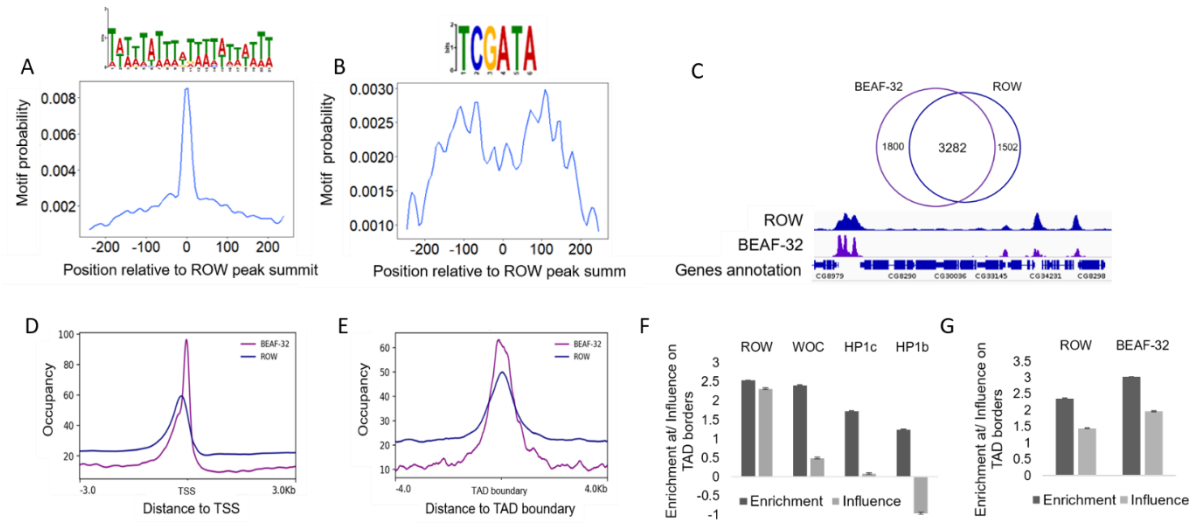


Figure 4. ROW and BEAF-32 ChIP-seq profiles relative to TSSs and TADs

(A) Central enrichment of AT-rich sequence in ROW binding regions. The logo shows the most significant enriched sequence (based on MEME-ChIP analysis). The plot shows the probability to have the motif relative to ROW ChIP-seq peaks (calculated with centriMo).

(B) BEAF-32 consensus motifs (logo) were found to be enriched within ROW binding sites (MEME-ChIP analysis). The plot (calculated with centriMo) shows that the probability to have the BEAF-32 motif is highest around 100 bases from the center of the ROW ChIP-seq peaks.

(C) Top: overlap between genes bound by ROW and genes bound by BEAF-32. Bottom: an example of ChIP-seq signal of ROW and BEAF-32 at a representative region with annotation of genes.

(D) The distribution of ROW and BEAF-32 binding relative to the positions of TSSs.

(E) The distribution of ROW and BEAF-32 binding relative to the positions of TAD boundaries.

(F-G) Results of multiple logistic regression used to compare the enrichment and independent influence at TAD boundaries of (F) ROW, WOC, HP1c, and HP1b (in fly heads) and (G) between BEAF-32 and ROW (in S2 cells). Values are the enrichment and influence beta coefficients ± standard error (calculated by HiCfeat R package [36]).

Genes differentially expressed in row knockdown flies are associated with metabolism, and most are not ROW targets

ROW binds upstream to TSS of active genes but also to TAD boundaries. If ROW is involved in transcription activation, we would expect to see downregulation of genes that are bound by ROW. To test it, we generated and sequenced RNA-seq libraries from control and *row* knockdown flies. As expected, the expression levels of *row* mRNA were significantly decreased in the row^{RNAi} lines by 40-60% ($P < 0.01$ for both row^{RNAi}) (Fig S3A). The differential expression analysis showed that the effect of *row* knockdown on gene expression was consistent between the RNAi lines (the square of the correlation between the fold changes (r= 0.79, $P < 2.2 \times 10^{-16}$) (Fig 5A), but the change in expression was more substantial for row^{RNAi-1} (on average 1.5 times larger) (Fig S3B-C). When analyzing the two row^{RNAi} lines together, we found 2035 genes with significant differential expression relative to the control (False Discovery Rate (FDR) < 0.05) (Table S4). Despite the suspected role of ROW in transcriptional activations, the number of genes upregulated and downregulated was nearly equal (53% and 51% were upregulated in row^{RNAi-1} and row^{RNAi-2}, respectively) (Fig 5B and Fig S3B-C), and this was similarly true when only genes bound by ROW were examined (53% were upregulated in row^{RNAi-1}, $P = 0.12$) (Fig 5B). Out of the differentially expressed genes, only 713 genes (35%) were bound by ROW, which is significantly lower than what is expected by chance (959 are expected by chance, $P < 2.2\times10^{-16}$) (Fig 5C). Thus, our analysis suggests that the majority of genes bound by ROW are not significantly affected by *row* knockdown, but instead most differentially expressed genes represent indirect effects.

To examine what type of genes are differentially expressed in the *row*^{RNAi} lines, we tested for enrichment of biological processes and transcription factors that are associated with them. The most significant biological processes were related to metabolism (Fig 5D and Table S5). Transcription factors enrichment analysis showed that the differentially expressed genes were modestly associated with a subunit of the polymerase negative elongation factor (NELF) and the suppressor of Hairy-wing insulator protein (Su(Hw)) (Odds Ratio (OR) = 1.21-1.31) (Fig 5E). Unlike ROW, the Su(Hw) insulator protein was previously shown to be enriched inside TADs and not at TAD boundaries [5]. Consistent with previous studies showing different localization of Su(Hw) and BEAF-32 [37], the most significant depletion (negative association) of DNA-binding protein at promoters of the differentially expressed genes was BEAF-32 (OR = 0.41 $P = 6.7\times10^{-69}$) (Fig 5E). Similarly, we found that TSSs of differentially expressed genes were less likely to be near TAD boundaries relative to non-differentially expressed genes ($P = 0.011$) or housekeeping genes ($P = 8.0\times10^{-11}$) (Fig S3D). Thus, this analysis reinforces the observation that knockdown of *row* results in changes in expression of genes mostly not directly bound by ROW.

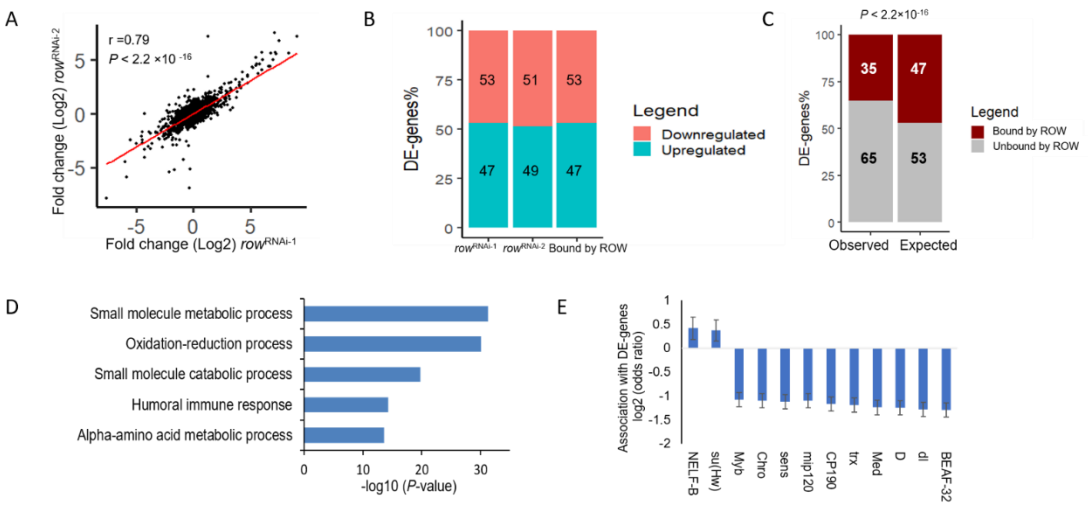


Figure 5. Knockdown of *row* cause misregulation of genes involved in metabolism, more likely to be non-ROW targets

(A) Correlation between the fold changes (log2) in *row*$^{RNAi-1}$ and *row*$^{RNAi-2}$. Fold changes were calculated relative to *Act-GAL4* control (using edgeR).

(B) The percentage of differentially expressed genes (DE-genes) which were upregulated and downregulated in the two *row*^{RNAi} lines for all genes and genes bound by ROW in *row*$^{RNAi-1}$.

(C) The percentages of differentially expressed genes that were bound or unbound by ROW and the expected values under an independence assumption.

(D) The top 5 most significantly enriched biological processes for differentially expressed genes after removing redundant terms.

(E) A significant association between genes bound by transcription factors (FlyEnrichr database) and differentially expressed genes. The odds ratios (Log_2) are shown for all the transcription factors found to be significantly positively associated with differentially expressed genes and the top 10 most significant factors showing negative association with differentially expressed genes. Significance was calculated by Fisher's exact test.

ROW and BEAF-32 regulate the expression of a common set of genes which are long-range targets of BEAF-32

The expression analysis in fly heads suggests that the majority of differentially expressed genes represent indirect or secondary effects of the decreased levels of *row* (ROW is depleted from the beginning of development). To further examine these results in a more homogeneous model that is not exposed to developmental effects

we used transient knockdown in S2 cells. We treated the S2 cells with dsRNA against *row*, *BEAF-32,* or both genes and analyzed gene expression in treated and untreated cells with RNA-seq. The reduction in protein levels was verified using western blot (Fig 6A). We found a significant correlation between the changes in expression in the cells with *BEAF-32* knockdown and *row* knockdown (r = 0.44, $P < 2.2\times10^{-16}$) (Fig 6B), suggesting that ROW and BEAF-32 influence the expression of a common set of genes. Changes in gene expression in cells with knockdown of both *row* and *BEAF-32* were significantly correlated with the changes in cells with knockdown of *row*, knockdown of *BEAF-32*, and with the additive effect of the two genes (the sum of the fold changes in *row* and *BEAF-32* separate knockdowns) (Fig S4).

Comparing gene expression in the treated and untreated cells, we found 51 differentially expressed genes (FDR < 0.1) (Table S6). The 51 genes had a significant overlap with a previously reported list of differentially expressed genes upon *row* knockdown in S2 cells[21] (OR = 24.1; $P = 3.2\times10^{-13}$). The majority of the differential expressed genes in S2 cells were downregulated (*row* knockdown: 66%, $P = 0.025$; *BEAF-32* knockdown: 60%, $P = 0.26$; knockdown of both: 70%, $P = 0.011$) (Fig 6C), consistent with the role of ROW and BEAF-32 in transcription activation. Transcription factors enrichment analysis showed that the differentially expressed genes are enriched with proteins known to regulate transcriptional pausing by RNA pol II [38–40]: NELF-B (OR=3.7), NELF-E (OR = 2.6), and Trl (known also as GAF) (OR = 1.9) (Fig 6D). We used our ChIP-seq results to test if the differentially expressed genes are bound by ROW and BEAF-32. Indeed, we found a positive association between the differentially expressed genes and genes bound by ROW (88% bound by ROW vs. 75% expected by chance, $P = 0.023$) (Fig 6E), while for BEAF-32 the association was

significantly lower than what is expected by chance (45% bound by BEAF-32 vs. 65% expected, P=0.0029) (Fig 6F).

A previous study showed that BEAF-32 regulates the expression of genes associated with RNA pol II pausing by long-range interactions [10]. The long-range contacts were between direct ChIP peaks of BEAF-32 containing DNA binding motifs and indirect low-intensity peaks without BEAF-32 motifs [10]. To examine the possibility that the differentially expressed genes in our experiment are also long-range targets of BEAF-32 and ROW through indirect binding, we studied the existence of BEAF-32 motifs (TCGATA) and ROW AT-rich motif (15 repeats of A or T) in their promoters. We found that the differentially expressed genes are depleted of BEAF-32 motifs (22% have the motif vs. 37% expected by chance; $P = 0.019$) (Fig 6G). Furthermore, out of the differentially expressed genes which were bound by ROW, only 24% have the ROW motif (42% are expected by chance; $P = 0.015$) (Fig 6H), suggesting that the binding of ROW to the promoter of the differentially expressed genes is the outcome of indirect binding. Indirect peaks of BEAF-32 were found previously to be enriched with RNA pol II pausing[10]. Consistent with this observation, we found that genes bound by ROW were enriched with the binding of factors associated with RNA pol II pausing (NELF-B, NELF-E, and Trl; $P < 2.2\times10^{-16}$), but the enrichment was significantly stronger for indirect peaks of ROW without the predicted binding sequences (NELF-B: $P = 2.1\text{x}10^{-6}$, NELF-E: $P = 3.7\text{x}10^{-6}$ and Trl: $P < 2.2\times10^{-16}$) (Fig 6I). Finally, we compared the differentially expressed genes with a published list of BEAF-32 long-range targets that were differentially expressed in S2 cells with a mutation that impairs the interaction of BEAF-32 with CP190 and the binding of BEAF-32 to the indirect peaks[10]. We found that the differentially expressed genes are positively associated with those long-range

targets of BEAF-32 (OR = 2.6, P = 0.009) (Fig 6J), especially when testing the association of genes downregulated in both data sets (OR = 6.4, P = 0.00013) (Fig 6J). Altogether, these results uncover a new mechanism of transcriptional activation by ROW and BEAF-32 of indirect targets.

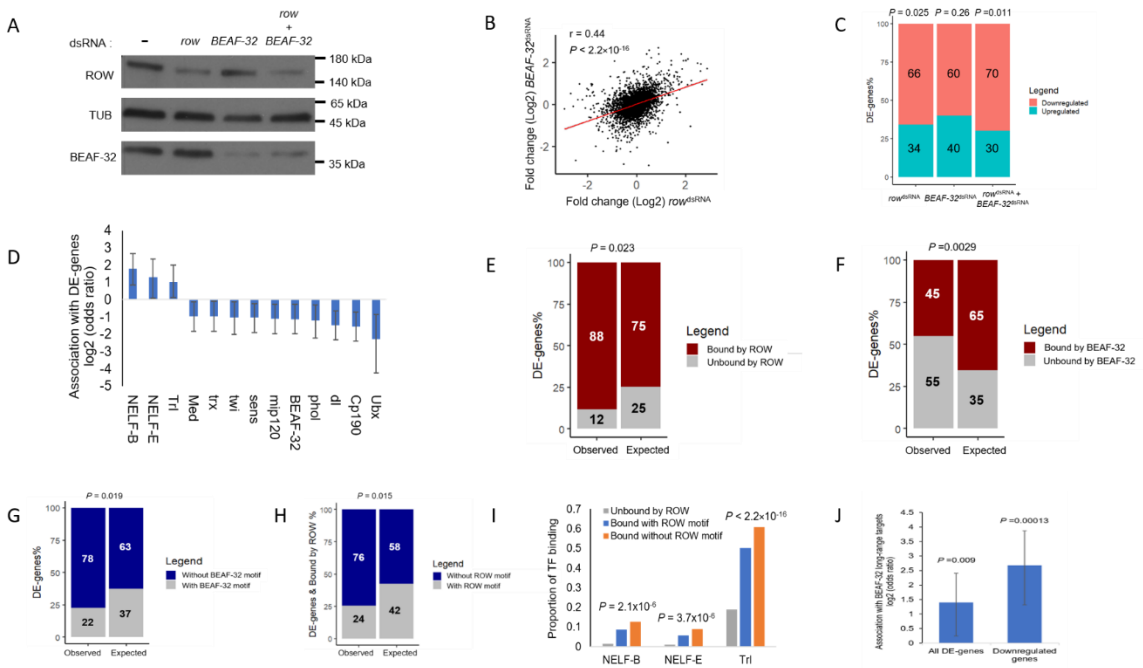


Figure 6. Knockdown of *row or BEAF-32* in S2 cells causes downregulation of a common set of long-range targets.

(A) Protein levels of ROW, BEAF-32, and Tubulin (TUB) in non-treated S2 cells and cells treated with dsRNA against *row*, *BEAF-32,* or both.

(B) Correlation between the gene expression fold change (log2) in cells treated with *row*^{dsRNA} and cells treated with *BEAF-32*^{dsRNA}. The fold change was calculated relative to control cells.

(C) The percentage of differentially expressed genes (DE-genes) which were upregulated and downregulated in *row*^{dsRNA}, *BEAF-32*^{dsRNA,} and *row*^{dsRNA} + *BEAF-32*^{dsRNA} treated cells.

(D) Association between transcription factors binding (FlyEnrichr database) and the differentially expressed genes. The odds ratios (Log_2) are shown for all the transcription factors found to be positively associated and the top ten most significant factors showing negative association with the differentially expressed genes.

(E) The percentages of differentially expressed genes bound or unbound by ROW and the expected values under independence assumption.

(F) The percentages of differentially expressed genes that were bound or unbound by BEAF-32 and the expected values under independence assumption.

(G) The percentages of differentially expressed genes with or without a BEAF-32 motif (TCGATA) at the promoter region and the expected values under independence assumption.

(H) The percentages of differentially expressed genes bound by ROW with or without a ROW motif (15 repeats of A or T) at the promoter region and the expected values under an independent assumption.

(I) Proportion of genes bound by the Transcription Factors (TF): NELF-B, NELF-E, and Trl out of genes unbound by ROW, bound by ROW with ROW motif, and bound by ROW without ROW motif.

(J) Association between differentially expressed genes with previously published long-range targets of BEAF-32. Values are the odds ratio (Log_2) of the association between all genes or restricting the analysis to downregulated genes.

Discussion

Our study uncovers new crosstalk between the insulator proteins BEAF-32 and the chromatin binding protein ROW. Both proteins bind promoters of housekeeping genes located at the boundaries of TADs, at sites that are AT-rich and flanked by motifs and binding of BEAF-32. The association of BEAF-32 and ROW with TAD boundaries may be an indirect effect of the localization of the two proteins at TSSs of active genes. The ChIP-seq analysis shows that the enrichment at those sites is highest for ROW, following by WOC, HP1c, and HP1b. We also found the knockdown of *row* and *BEAF-32* affects a common set of genes, most are downregulated and are long-range targets, with an indirect binding of ROW and BEAF-32. Our results together with previously published studies suggest the existence of hierarchical recruitment of the protein complex that includes HP1c and WOC to TAD boundaries by ROW and the cooperation with BEAF-32 that recruits other insulator proteins to regulate genes through long-range interactions (Figure 6).

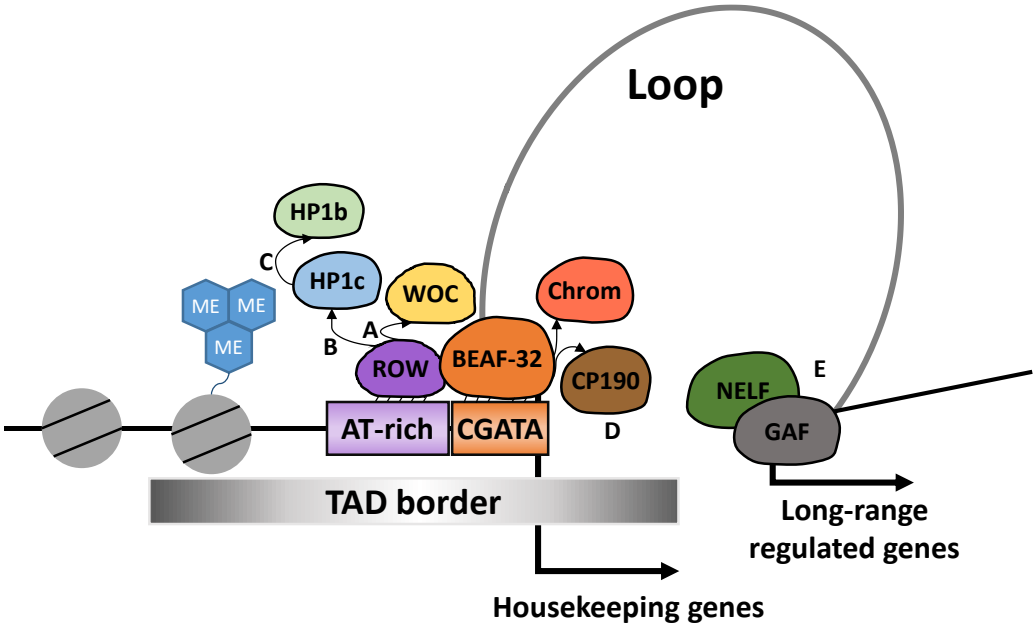


Figure 7. A model for the cooperation of ROW with BEAF-32 and other chromatin-bound proteins in transcription and nuclear organization. Our proposed model is that ROW and BEAF-32 provide DNA specificity. ROW recruits to the sites the HP1c complex that was shown previously to be involved in gene activation[21]. BEAF-32 interacts with CP190 and Chromator that are responsible for the physical interactions required for long-range contacts[41]. The model relies on the interaction between ROW and BEAF-32, the localization of both proteins at most of the same promoters, and the presence of AT-rich sequences and CGATA motifs in the promoters of housekeeping genes, characterized by a broad nucleosome-free region and H3K4me3. The recruitment of proteins such as WOC and HP1c to the site by ROW is proposed to be essential for transcription activation of genes that have long-range contact with the active promoters.

(A) WOC shows the highest overlap in binding with ROW. A previous study showed that the ROW is required for the binding of WOC to the chromatin [23]. It was also shown that HP1c, WOC, and ROW regulate a common set of genes[19].

(B) HP1c shows a lower correlation with ROW, relative to WOC. HP1c binding to the chromatin was previously found to be dependent on ROW [19,23].

(C) The correlation in genome distribution is highest for HP1b with HP1c. Among the proteins tested, HP1b shows the lowest correlation with ROW and with the genomic annotations associated with ROW (active promoters and TAD boundaries). A previous study showed that HP1c and HP1b form heterodimers *in vitro* and *in vivo* [42].

(D) Previously published studies showed that BEAF-32 regulates the expression of genes not directly bound by BEAF-32 through long-range contacts that depend on the interaction of BEAF-32 with CP190[10,43].

(E) Based on our study, ROW and BEAF-32 regulate a common set of genes that are not direct targets of the two proteins (lack of DNA binding motifs) but instead are long-range contacts bound by GAF and NELF proteins.

Based on our proposed model, the specificity of the binding of the multiple proteins that are involved in the long-range regulation is due to the cooperation of two sequence-specific DNA-binding proteins, BEAF-32, and ROW. BEAF-32 binds to its motifs (CGATA) [37] located near AT-rich sequences, that are bound by ROW (probably through its AT hooks). Based on our analysis, those binding sites are promoters of housekeeping genes located at TAD boundaries. The evidence that those are promoters of housekeeping genes is supported by the H3K4me3 marks, the broad nucleosome-free region (NFR) near the TSS bordered by a regular array of nucleosomes, and the high and consistent expression of the genes across developmental stages. After ROW and BEAF-32 bind DNA, they recruit other proteins, like WOC and HP1c by ROW, and CP190 or Chromator by BEAF-32. The notion that the sequence-specific binding of ROW and BEAF-32 directs the localization of the HP1c complex is consistent with the strongest enrichment of ROW to promoters (relative to WOC, HP1c, and HP1b), and a larger independent enrichment of BEAF-32 and ROW to TAD boundaries. Previous studies showed that the recruitment of WOC and HP1c to the chromatin is dependent on ROW [19,23], but our results show a much higher concordance in the binding of WOC and ROW and a much lower correlation between ROW and HP1c. The recruitment of HP1b may be dependent on the heterodimerization with HP1c, as evident by the high correlation we observed in the

genomic distribution between HP1c and HP1b, the lowest correlation of HP1b with ROW, and previous studies showing that HP1c/HP1b heterodimers are formed both *in vitro* and *in vivo* [42].

It was previously found that ROW complex bounds developmental genes that are associated with RNA pol II pausing[21]. We suggest that the binding of ROW to genes enriched with RNA pol II pausing (developmentally regulated genes and inducible genes) is established mostly indirectly, through long-range contacts with direct binding sites at promoters of housekeeping genes. Our work defines the complex of ROW as essential for transcriptional activation of the long-range genes. The mechanism that explains how the complex of ROW promotes transcriptional activation of long-range genes is first through stabilizing the NELF complex and stalled Pol II at these genes. Such a mechanism was suggested by a previous study which showed that depletion of Dsk2 (binding partner of ROW) causes a decrease in NELF-E and Pol II pausing at TSSs[21]. Stalled Pol II is known to enhance the expression of genes by maintaining accessible chromatin structure[44,45]. In addition, the recruitment of the complex FACT can facilitate RNA Pol II elongation. A previous study showed that HP1c recruits FACT to active genes and active forms of RNA polymerase II, and in the absence of HP1c, the recruitment of FACT into heat-shock genes was altered and the expression levels were reduced[24].

In summary, we provide here a molecular mechanism that clarifies how regulated genes can be transcriptionally activated by the HP1c complex which is recruited to the vicinity of the BEAF-32 complex and promoters of housekeeping genes. Thus, we

expand the knowledge regarding the crosstalk between the two complexes and their role in transcription regulatory mechanisms of housekeeping and regulated genes.

Materials and Methods

Fly stocks and maintenance

The *row*$^{RNAi-1}$ and *row*$^{RNAi-2}$ are 25971 and v28196 respectively from the Bloomington *Drosophila* Stock Center and VDRC Stock Center. *Act-GAL4*/Cyo flies (3953) were ordered from the Bloomington Stock Center. All flies were raised at 25°C in 12-hour light: 12-hour dark cycles on standard diets. We generated an isogenic background for the two *Drosophila row*^{RNAi} strains by crossing females 7 times with W1118 male flies. In each cross, female offspring carrying the transgene were selected by the red-eye phenotype or by genotyping.

Antibodies

Polyclonal rat αROW [19] (WB 1:10000), polyclonal rabbit αROW [21] (ChIP-seq: 1µl), αHP1c[21] (ChIP-seq: 5µl), αHP1b [21] (ChIP-seq: 5µl) and Rabbit αWOC [46] (ChIP-seq: 1µl) were previously described in the indicated references. Mouse monoclonal αFlag is SIGMA ALDRICH (F1804, WB 1:500). Rat monoclonal αTubulin (WB: 1:10000) and rabbit αH3K4me3 (ChIP-seq: 5µl) are Abcam (ab6160 and ab8580 respectively). Rabbit igG is Santa Cruz (sc-2027, ChIP-seq: 5µl). Mouse αBEAF-32 is Developmental Studies Hybridoma Bank (AB_1553420, WB: 1:200, ChIP-seq: 8µl)

Western blot

20 fly heads per sample were collected on dry ice and homogenized in 200µl RIPA lysis buffer (50 mM Tris-HCl at pH 7.4, 150 mM NaCl, 1 mM EDTA, 1% NP-40, 0.5% sodium

deoxycholate, 0.1% sodium dodecyl sulfate, and 1 mM DTT, and protease inhibitor tablets [Roche]) by motorized pestle. For S2 cells, 10^6 cells were homogenized in 100 µl of RIPA buffer. The lysates were kept on ice for 20 min and then centrifuged at max speed for 30 minutes. The supernatants were collected and 20µl per sample was boiled with protein sample buffer (Bio-Rad). Criterion XT Bis-Tris gels (Bio-Rad) were used for gel electrophoresis.

Viability assay

The UAS-*row*^{RNAi} flies or W1118 control flies were crossed with *Act-GAL4*/CyO flies. Then, the number of CyO and non-CyO offspring were counted. The percentage of the specific genotype from the total progeny was calculated. Each cross was performed three times and the significance of the result was tested using a two-sided binomial test.

Pupal eclosion

The UAS-*row*$^{RNAi-1}$ flies or control W1118 flies were crossed with *Act-GAL4*/CyO flies (n=3 for each cross-type). Seven days after the cross was made, 40 pupae were randomly collected, each pupa to a separate tube with food and a hole for fresh air. The number and the genotype of newly enclosed flies were recorded. The Statistical test performed was a two-sided t-test.

Survival assay

From each genotype 3 vials containing 20 flies (0-3 days old, 10 males and 10 females) were kept on standard food at 25°C. Once in three days, the number of flies that died was counted and their gender was recorded. The experiment was continued until all flies died. Tests for the difference between the survival curves were performed with the OIsurv R package using the survdiff function.

Fertility assay

From each genotype, 8 males or 12 females were crossed with w1118 flies and the number of offspring was counted. Three crosses were performed for each genotype. Statistical tests were performed with ANOVA followed by Tukey's test.

Co-immunoprecipitation (co-IP) assay

For each sample, 100µl of fly heads were homogenized in 500µl lysis buffer (50mM TRIS HCL 7.4, 150mM NaCl, 1mM EDTA, 1% TRITON X100 and protease inhibitor tablets [Roche]). Homogenate was kept on ice for 30 min and then sonicated using a Bioruptor sonication device (Diagenode) for 2 min (10 sec on 10 sec off). Sonicated lysate was centrifuged at 21,000g for 10 min and the supernatant was then collected, 35µl were removed for input.

Pre-clean: protein A/G PLUS- Agarose beads (Santa Cruz) were washed 3 times with 150µl of lysis buffer, resuspended with 100µl lysis buffer, and then added to the lysate. The samples were rotated for 30 min at 4°C and centrifuged at 3000 rpm for 1 min. The lysate was collected to a new tube and the beads were discarded.

IP: 25µl of Red ANTI-FLAG M2 Affinity Gel from SIGMA ALDRICH (F2426) were washed 3 times with TBS (50mM TRIS pH 7.5,150mM NaCl and protease inhibitor tablets [Roche]) and then added to samples. The samples were rotated for 3 hours at 4°C. Next, the samples were centrifuged at 3000 rpm for 1 min and the unbound lysate was collected to a new tube. Beads were washed 3 times using 400µl washing buffer (300mM NaCl, 0.1% NP 40,50mM TRIS HC1 7.5 and protease inhibitor tablets [Roche]) and twice using 400µl of TBS. For western blot, the beads were eluted with SDS-PAGE X1 sample buffer and boiled at 95°C for 5 min

Mass spectrometry (MS)

Sample preparation**:** The packed beads were resuspended in 100µl 8M urea, 10mM DTT, 25 mM Tris-HCl pH 8.0 and incubated for 30 min at 22°C. Next, Iodoacetamide (55mM) was added and beads were incubated for 30 min (22°C, in the dark), followed by the addition of DTT (20mM). The Urea was diluted by the addition of 6 volumes of 25mM Tris-HCl pH 8.0. Trypsin was added (0.3µg/ sample) and the beads were incubated overnight at 37°C with gentle agitation. The beads were spun down and the peptides were desalted on C18 Stage tips. Two-thirds of the eluted peptides were used for MS analysis.

NanoLC-MS/MS analysis: MS analysis was performed using a Q Exactive Plus mass spectrometer (Thermo Fisher Scientific, Waltham, MA USA) coupled on-line to a nanoflow HPLC instrument, Ultimate 3000 Dionex (Thermo Fisher Scientific, Waltham, MA USA). Peptides were separated over an acetonitrile gradient (3% to 32% for 45 min.; 32 to 50% for 15 min., 50 to 80% for 10 min.) run at a flow rate of 0.3µl/min on a reverse-phase 25-cm-long C18 column (75µm ID, 2µm, 100Å, Thermo PepMapRSLC). The survey scans (380–2,000 m/z, target value 3E6 charges, maximum ion injection times 50ms) were acquired and followed by higher-energy collisional dissociation (HCD) based fragmentation (normalized collision energy 25). A resolution of 70,000 was used for survey scans. The MS/MS scans were acquired at a resolution of 17,500 (target value 1E5 charges, maximum ion injection times 120ms). Dynamic exclusion was 60 sec. Data were acquired using Xcalibur software (Thermo Scientific). To avoid a carryover, the column was washed with 80% acetonitrile, 0.1% formic acid for 25 min between samples.

MS data analysis: Mass spectra data were processed using the MaxQuant computational platform[47] (version 1.5.3.12). Peak lists were searched against the

UniProt Fasta database of *Drosophila melanogaster*, using both annotated and predicted sequences. The search included cysteine carbamidomethylation as a fixed modification as well as oxidation of methionine as variable modifications and allowed up to two miscleavages. The “match-between-runs” option was used. Peptides with a length of at least seven amino-acids were considered and the required FDR was set to 1% at the peptide and protein level. Protein identification required at least 2 unique or razor peptides per protein. Relative protein quantification in MaxQuant was performed using the label-free quantification (LFQ) algorithm and with intensity-based absolute quantification (IBAQ). Protein contaminants and proteins identified by less than 2 peptides were excluded. Statistical analysis between control (n=3) and rowTAG samples (n=3) was performed using the Perseus statistical package[48] (computational platform for comprehensive analysis of (prote) omics data). LFQ values were used as the input for Perseus analysis. Ribosomal proteins were excluded from the results.

Tagging *row* endogenously by CRISPR/Cas9

Tagging *row* endogenously in *Drosophila Melanogaster* using the CRISPR/Cas9 method was performed based on the approach previously described[49] with some modifications. Two gRNAs (one targeting *row* and one targeting the *white* gene) was cloned into pCFD4d plasmid (Addgene plasmid #83954) (as described in the protocol “cloning two gRNAs into plasmid pCFD4”; http://www.crisprflydesign.org/) - (pCFD4d*w/row*)

Primers used for the cloning are: 5'-TATATAGGAAAGATATCCGGGTGAACTTCC CGTGGGGCTTGTATCATTGGTTTTAGAGCTAGAAATAGCAAG-3' (forward)

5'-CCAAAGAGCAGGAATGGTATATTTTAACTTGCTATTTCTAGCTCTAAAACCCAAAG

AGCAGGAATGGTATCGACGTTAAATTGAAAATAGGTC-3' (reverse).

To generate the donor plasmid for homologous recombination (HR) of the gene *row*, pUC57-white [coffee] plasmid (Addgene 84006) was digested with SacI and HindIII to exclude the donor template for HR of the white gene. We used w1118 flies genomic DNA to amplify by PCR both ~1kb upstream and ~1kb downstream from the stop codon of the gene *row*. To add the tag, the primers included the FLAG-TAG sequence next to the stop codon. As we used the Gibson Assembly kit, the primers included part of the plasmid backbone as well. The vector backbone and the two parts of the amplified homologous arms were assembled using Gibson Assembly kit (pUC57-*row*tag).

Primers used for the cloning:

Upstream part: 5'-ACGGCCAGTGAATTCGAGCTCGCGGGTTGAGGTTTATAAGTC-3' (F)

5'-TCACTTGTCATCGTCATCCTTGTAATCTTGCGGATGGTGATGGTGCT-3' (R)

Downstream part: 5'-GATTACAAGGATGACGATGACAAGTGATACAAGCCCCAC

GGAAA-3' (F)

5'-CTATGACCATGATTACGCCAGGAGCTATGCCTACCCCTTC -3' (R)

pCFD4d*w/row*, pUC57-white[coffee], and pUC57-*row*tag plasmids were injected into vas-Cas9 (y1, M{vas-Cas9ZH-2A) flies[49] by Rainbow Transgenic Flies, Inc. Individual injected G_0 flies were mated with second chromosome balancers flies, and non-red-eyed G_1-*Cyo* individual flies were crossed to 2nd chromosome balancers flies again. Screening for the desired *row*-FLAG tagged line and elimination of random integration in the G_1-Cyo individuals performed by PCR was as previously described[49]. For final validation, we sequenced the entire *row* locus and performed WB as shown in the results section.

Knockdown experiments in S2 cells

RNAi experiments in S2 cells were performed as previously described[21] with modifications. dsRNA to *row* and *BEAF-32* were generated using the MEGAscript kit (Ambion). Cells were diluted to 10^6 ml^{-1} and 4 µg of dsRNA per 1ml of cells were added. After 2 days the cells were diluted again to 10^6 ml^{-1} and 8 µg of dsRNA per 1ml of cells were added. After 2 days, the cells were washed twice with PBS, collected, and used for downstream experiments. The primers used for producing dsRNA to *row* were previously described[21]:

Forward *row* T7: TAATACGACTCACTATAGGGTGATACAGACGCTGAGTGATTG

Reverse *row* T7: TAATACGACTCACTATAGGGAGGAACCACATCCCAAGATG

Primers used for producing dsRNA to *BEAF-32*:

Forward BEAF-32 T7: TAATACGACTCACTATAGGGGCGAGGATCCACTGTGCTAT

Reverse BEAF-32 T7: TAATACGACTCACTATAGGGACGCTGATTTGCCCATTTAC

ChIP-seq using fly heads

The ChIP was performed based on a previous protocol[50] with modifications. 100µl of fly heads were collected using dry ice and homogenized in 1 mL of NEB buffer (10mM HEPES-Na at pH 8.0, 10mM NaCl, 0.1mM EGTA-Na at pH 8.0, 0.5mM EDTA-Na at pH 8.0, 1mM DTT, 0.5% Tergitol NP-10, 0.5mM Spermidine, 0.15mM Spermine and protease inhibitor tablets [Roche]) for a total of 2 min. Homogenate was poured into Bio-Spin chromatography columns (BIO-RAD) and centrifuged at 1000g for 4 min. The filtered homogenate was centrifuged at 6000g for 10 min and the supernatant was discarded. The pellet was resuspended in 1 mL of NEB and centrifuged at 11,000rpm for 20 min on a sucrose gradient (0.65 mL of 1.6 M sucrose in NEB, 0.35 mL of 0.8 M sucrose in NEB). The Nuclei pellet was then resuspended in 1 mL of NEB and 11%

formaldehyde was added for a final concentration of 1%. Nuclei were cross-linked for 10 min at room temperature and quenched by adding 1/10 vol of 1.375 M glycine. The nuclei were collected by centrifugation at 6000g for 5 min. Nuclei were washed twice in 1 mL of NEB and resuspended in 350mL of Sonication buffer (10mM HEPES-Na at pH 7.5, 2mM EDTA at pH 8.0, 1% SDS, 0.2% Triton X-100, 0.5mM Spermidine, 0.15mM Spermine). Nuclei were sonicated using a Bioruptor sonication device (Diagenode) for 3 cycles of 7 min (30 seconds on 30 second off). Sonicated nuclei were centrifuged at 15,000g for 10 min and the supernatant was frozen at 80°C in 150µl aliquots. 50µl were taken for checking the chromatin quality and the DNA fragment size, which was between 100 to 1000 bp. For each sample, we used 50µl from the aliquot diluted with 500µl of IP buffer (50 mM HEPES/KOH at pH 7.6,2 mM EDTA, 1% Triton, 0.1% Na Deoxycholate in PBS and protease inhibitor tablets [Roche]). Samples were rotated overnight at 4°C after adding antibodies. ChIP-seq was performed with rabbit αHP1c (n=3), αHP1b (n=2), αWOC (n=3), αROW (n=3), αH3K4me3 (n=3) and IgG antibodies (n=2). 20µl per sample of Dynabeads Protein G (Invitrogen) was washed twice with 1ml of IP buffer. Beads were resuspended in 25µl IP buffer per sample. The beads were added to the samples and rotated for 1 hour at 4°C. Chromatin immobilization and barcoding performed as previously described[51]

ChIP-seq using S2 cells

The ChIP was performed based on a previous protocol[21] with modifications. For each sample, a 25 cm flask with 8-10*10^6 S2 cells was collected into a 15 ml tube and formaldehyde was added to a final concentration of 1.8%. The cells were crosslink for 10 min at RT on a shaker. To stop the reaction, glycine was added to a final concentration of 125mM (from a stock solution of 2.5M in PBS). The cells were

centrifuged for 3 min at 1500g at 4°C, the supernatant was discarded and the cells were resuspended in 2ml PBS. Then, the cells were centrifuged again (3 min at 1500g at 4°C), the supernatant was discarded and the cells were resuspended in 1ml washing buffer A (10mM Hepes pH7.9, 10mM EDTA 0.5mM EGTA 0.25% Triton X100m, protease inhibitor). The 1ml suspension was transferred to an Eppendorf, incubated for 10 min at 4°C on a wheel, and then centrifuged (3 min at 1500g at 4°C). The supernatant was discarded, the pellet resuspended in 1ml washing buffer B (10mM Hepes pH7.9, 100mM NaCl, 1mM EDTA, 0.5mM EGTA, 0.01 % Triton X100) and incubated for 10 min at 4°C on a wheel. The suspension was centrifuged again (3 min at 1500g at 4°C), the supernatant was discarded and the pellet was resuspended in 450µl TE buffer. 50µl 10% SDS was added and the Eppendorf was inverted 5 times and centrifuged at 1500g at 4°C for 3min. The upper phase was carefully removed with a pipette, 500µl TE was added and the Eppendorf was inverted 5 times after which it was centrifuged. The last procedure was repeated once again. After the second centrifuge, the upper phase was removed, TE with 1mM PMSF was added to a final volume of 400µl and 4 µl of 10% SDS was added. The suspension was sonicated using Bioruptor plus sonication device (Diagenode) for 20 cycles (30 sec on 30 sec off).

The following solutions were added in this order to the lysate and between each addition, the lysate was incubated on the wheel at 4°C for 2min: 42µl 10% TRITON (final 1%), 4.2µl 10% Deoxycholate (final 0.1%), and 15µl 4M NaCl (final 140 mM). The lysate was Incubated on the wheel at 4°C for 10min and centrifuged at full-speed at 4°C for 5min. The supernatant was divided into aliquots and was frozen at -80°C, 45µl were kept for input and for checking DNA fragment size.

For the ChIP-seq experiment, 50 µl of chromatin aliquot was used and diluted with 450 RIPA buffer (140 mM NaCl, 10 mM Tris-HCl pH, 1 mM EDTA, 1 % TritonX100, 0.1 % SDS, 0.1 % Deoxycholate and protease inhibitor). Samples were rotated overnight at 4°C after adding antibodies. Washes performed as described in the "ChIP-seq using fly heads" paragraph. Chromatin immobilization and barcoding performed as previously described[51]

ChIP-seq analysis

ChIP-seq reads were aligned by bowtie2 software to the full genome sequence of *Drosophila melanogaster* (UCSC version dm3). Duplicate reads were removed using the samtools rmdup function and uniquely mapped reads were kept (samtools view -q 10). Peak calling was performed with the callpeak function of the MACS2 tool[52] with a q-value<0.05. An IgG sample was used as a control. Peaks in ChrU and ChrUextra were excluded for subsequent analysis.

For visualization, bedGraph files were created by deepTools (a python tool) with bamCoverage function and normalization parameter set to RPKM. The bedGraph file with the average between replicas was generated and the corresponding bigWig file was created. The bigWig files were uploaded to the integrative genomic viewer (IGV). The BigWig file of BEAF-32 in Kc167 cells was download from NCBI (GSM762845) [35]. The ComputeMatrix and plotProfile functions of deepTools were used for plotting the distribution of the ChIP-seq targets relative to the positions of TSSs or TAD boundaries. For each target, the average across replicates was used. UCSC refGene annotation file was used to determine the TSS locations. Classified TADs locations were obtained from the Chorogenome Navigator[8].

The correlation of ChIP-seq signal across the genome was calculated based on non-overlapping bins of 2000bp. The deeptools pyBigWig python extension was used for calculating the average signal in each bin. A pairwise correlation was calculated with the psych R package. The locations of peaks within different genes annotation were determined using the annotatePeak function from the ChIPseeker R package[53]. The function enrichPeakOverlap from ChIPseeker was used to test the overlap of ChIP-seq peaks with different histone modifications. Histone modification data were from modENCODE ChIP-chip data. Multiple logistic regression implemented in HiCfeat R package[36] was used to calculate the enrichment and influence of proteins on TAD boundaries.

UCSC refGene annotation was used to define the genes that are bound by ROW or BEAF-32. Promoter regions (between - 250bp and +50 of TSS) of genes that overlapped with ROW/BEAF-32 peaks were considered to be bound by the proteins (Table S1 and Table S3). A BED file of BEAF-32 called peaks in Kc167 cells was download from NCBI (GSM762845) [35]. Expression levels of genes bound or unbound by ROW were calculated using RNA-seq data performed on control fly heads in this study. Normalized (based on library size) count data including all genes was used. The coefficient of variance (CV) of gene expression levels for genes bound or not bound by ROW was calculated using published *Drosophila* gene expression data from 30 developmental stages [54].

Motif enrichment analysis was performed using the MEME-ChIP motif analysis tool [58] with the default parameters. Input sequences were taken from an interval that included 250bp downstream and 250bp upstream of the peak's summit. CentriMo

option [59] with default parameters was used to calculate the probability to find the motif relative to the peak summit.

Predicting DNA-binding specificities for Cys2His2 zinc fingers:

ROW protein sequence (download from UCSC) was used as input in the online tool for predicting zinc finger DNA binding (http://zf.princeton.edu/logoMain.php) [34].

MNase-seq

100µl of fly heads per sample was homogenized for 5 min in a homogenization chamber with 300µl of ice-cold Nuc. Buffer-I (60mM KCl, 15mM NaCl, 5mM $MgCl_2$, 0.1mM EGTA, 15mM Tris-HCl pH=7.5, 0.5mM DTT, 0.1mM PMSF and 1x complete protease inhibitor (Roche)). Next, 200µl more of Nuc. Buffer-I was added. The homogenate liquid was poured into Bio-Spin chromatography columns (BIO-RAD) and centrifuged at 1000g for 1 min at 4°C. 500µl of ice-cold Nuc. Buffer-II (0.04% NP-40 and Nuc. buffer I) was added to the filtered homogenate, mixed by inverting the tube, and incubated for exactly 7 min on ice. During the 7 min incubation, 13ml tubes with 8ml ice-cold Nuc. Buffer-III (1.2M sucrose + Nuc. buffer I) was prepared. The 1ml filtered homogenate was poured into the 13 ml tube of Nuc. Buffer- III. The tube was centrifuged for 10 min at 11,000rpm. The supernatant was carefully removed so that only the pellet remained (containing the intact purified nuclei). The nuclei pellet was re-suspend in 500µl ice-cold MNase digestion buffer (0.32M sucrose, 50mM Tris-HCl pH=7.5, 4mM $MgCl_2$, 1mM $CaCl_2$ and 0.1mM PMSF). The 500µl nuclei were split into 2 Eppendorf with 250µl in each. 5 units (50 Gel Units) of Micrococcal Nuclease (NEB) were added to one tube and 20 units (200 Gel Units) to the other tube (for over digesting). The tubes were inverted for mixing and incubated at 37°C for 10 min. The reaction was stopped by the addition of 13µl of 20mM EDTA and 13µl of 10% SDS in

each tube, followed by inverting and cooling on ice. 2.8µl of 10mg/ml Proteinase K was added and mixed by inverting. The nuclei were incubated at 65°C overnight and then centrifuged for 1 min at 15,000g. The supernatant from each tube was taken and cleaned separately by PCR clean-up kit (Invitrogen) with 26µl of elution buffer at the final step. Next, 1µl of 0.3ug/µl RNase free DNase (Sigma), 0.75µl of 0.5M Tris-HCl pH=7.4, and 3.5µl 120mM NaCl were added to each tube and mixed by pipetting and spin-down. The tubes were then incubated for 1h at 37°C and the mix was cleaned again using a PCR clean-up kit (Invitrogen) with 12µl of elution buffer at the final step. Each tube volume was put to run on a 1.5% agarose gel alongside a 100bp DNA ladder. Mononucleosome bands (150bp-200bp) from the gel were excised and cleaned with a gel clean-up kit (Invitrogen) with 50µl of DDW as a final step. The purified DNA was quantified by nanodrop and an equal amount from the two digestion levels was taken for DNA library preparations. DNA libraries were made as previously described[60].

RNA purification and libraries preparation for RNA-seq using fly heads

We crossed *Actin GAL4*/CyO flies with *row*$^{RNAi-1/2}$ flies or with w1118 control flies and selected 0-3 days old *row*$^{RNAi-1}$ (*Act-GAL4*/+; UAS-*row*$^{RNAi-1}$/+), *row*$^{RNAi-2}$ (*Act-GAL4*/+; UAS- *row*$^{RNAi-2}$) and *Act-GAL4* Control (*Act-GAL4* /+) flies. The flies were raised for three more days to recover from CO_2 exposure and collected using dry ice. RNA was prepared using TRI Reagent (SIGMA) according to the manufacturer's protocol. 3' RNA-seq library was performed based on Guttman's lab "RNAtag Seq protocol"[61] with modifications. Fragmentation of RNA performed using FastAP Thermosensitive Alkaline Phosphatase buffer (Thermo Scientific) for 3 min at 94°C. Samples were placed on ice and then the FastAP enzyme was added for 30 min at 37°C. RNA was purified using 2.5× volume of SPRI beads (Agencourt) and then linker ligation was

performed with an internal sample-specific barcode using T4 RNA ligase I (NEB). All RNA samples were pooled into a single Eppendorf and purified using RNA Clean & Concentrator columns (Zymo Research). Poly A selection performed with Dynabeads Oligo(dt) beads according to the manufacture's protocol. RT was performed with a specific primer (5'-CCTACACGACGCTCTTCC-3') using AffinityScript Multiple Temperature cDNA Synthesis Kit (Agilent Technologies). RNA degradation was performed by incubating the RT mixture with 10% 1 M NaOH (2µl of RT mixture, 70°C, 12 min). The pH was then neutralized by AcOH (4 µl for 22 µl mixture). Next, the cDNA was cleaned using Silane beads (Life Technologies). The 3'-end of the cDNA was ligated to linker2 using T4 RNA ligase I. The sequences of linkers are partially complementary to the standard Illumina read1 and read2/barcode adapters, respectively. The cDNA was cleaned using Silane beads and PCR was performed using enrichment primers and Phusion HF Master Mix (NEB) for 12 cycles. The library was cleaned with 0.8× volume of SPRI beads.

RNA purification and libraries preparation for RNA-seq using S2 cells

RNA purification was performed with RNeasy Mini Kit (Qiagen) according to the manufacturer's protocol. 3' RNA-seq library generation was performed based on a previous protocol[62]. In brief, 100ng from each sample was incubated with oligo-dT reverse transcription (RT) primers containing 7bp barcode and 8bp UMI (Unique Molecular Identifier) at 72°C for 3 minutes and moved directly to ice. This was followed by a Reverse transcription reaction (SmartScribe kit Clontech) for 1 hour at 42°C and 70°C for 15 minutes. Next, samples were pooled and purified using SPRI beads X1.2 (Agencourt AMPure XP, Beckman Coulter). The pooled barcoded samples were then tagmented using Tn5 transposase (loaded with oligos Tn5MEDS-A) for 8

min at 55°C Followed by the addition of 0.2% SDS to strip off the Tn5 and SPRI X2 purification was performed. Finally, a PCR was performed with primers containing NGS sequences (KAPA HiFi HotStart ReadyMix, Kapa Biosystems, 12 cycles) and the library was cleaned using 0.8x SPRI beads.

Analysis of RNA-seq data

RNA-seq reads were aligned by STAR [63] to the full genome sequences of *Drosophila Melanogaster* (UCSC version dm3) and counted per annotated gene using ESET (End Sequencing analysis Toolkit) [64]. Differential expression analysis was performed using the edgeR R package. In the data from fly heads, we compared the two *row*^{RNAi} lines to the control [65]. Only genes with a count per million (CPM) > 1 in at least 3 different samples were included in the analysis. The ranked list of differentially expressed genes based on FDR (Table S4) was used as an input for the enriched GO terms tool –Gorilla [66] (Table S5). The REVIGO tool [57] was used to remove redundant GO terms.

In the data from S2 cells, we compared all treated samples to untreated samples. The FilterByExpr function in the edgeR R package is used to filter out non-expressed genes (Table S6).

The "Transcription_Factors_from_DroID_2015" file from FlyEnrichr [55,56] was used to calculate the odds ratio of DE-genes to be enriched or to be depleted within different transcription factor associated genes. Significance and odds-ratio were calculated using Fisher's exact test. All significantly enriched transcription factors and the 10 top significantly depleted factors were presented. The associations between genes that were significantly differentially expressed with other lists of genes (For example genes bound by ROW or BEAF-32 and BEAF-32 long-range targets) were tested using Fisher's exact test.

Acknowledgments

This work was supported by The Legacy Heritage Bio-Medical Program of the Israel Science Foundation [Grant No. 839/10 to SS and SK], and by the Israel Science Foundation (grant no. 575/17 to SS).

References

1. Szabo, Q., Bantignies, F. & Cavalli, G. Principles of genome folding into topologically associating domains. *Science Advances* (2019). doi:10.1126/sciadv.aaw1668
2. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
3. Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* **148**, 458–472 (2012).
4. Oudelaar, A. M. & Higgs, D. R. The relationship between genome structure and function. *Nature Reviews Genetics* 1–15 (2020). doi:10.1038/s41576-020-00303-x
5. Ulianov, S. V. *et al.* Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. *Genome Res.* **26**, 70–84 (2016).
6. Emberly, E. *et al.* BEAF regulates cell-cycle genes through the controlled deposition of H3K9 methylation marks into its conserved dual-core binding sites. *PLoS Biol.* **6**, e327 (2008).
7. Roy, S., Gilbert, M. K. & Hart, C. M. Characterization of BEAF mutations isolated by homologous recombination in Drosophila. *Genetics* (2007). doi:10.1534/genetics.106.068056
8. Ramírez, F. *et al.* High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat. Commun.* **9**, 1–15 (2018).
9. Cuvier, O., Hart, C. M., Käs, E. & Laemmli, U. K. Identification of a multicopy chromatin boundary element at the borders of silenced chromosomal domains. *Chromosoma* **110**, 519–531 (2002).
10. Liang, J. *et al.* Chromatin Immunoprecipitation Indirect Peaks Highlight Long-Range Interactions of Insulator Proteins and Pol II Pausing. *Mol. Cell* **53**, 672–681 (2014).
11. Vermaak, D. & Malik, H. S. Multiple Roles for Heterochromatin Protein 1 Genes in Drosophila. *Annu. Rev. Genet.* **43**, 467–492 (2009).
12. James, T. C. & Elgin, S. C. Identification of a nonhistone chromosomal protein

associated with heterochromatin in Drosophila melanogaster and its gene. *Mol. Cell. Biol.* **6**, 3862–3872 (1986).

13. Vermaak, D., Henikoff, S. & Malik, H. S. Positive selection drives the evolution of rhino, a member of the heterochromatin protein 1 family in drosophila. *PLoS Genet.* **1**, 0096–0108 (2005).
14. Eissenberg, J. C. & Elgin, S. C. R. HP1a: A structural chromosomal protein regulating transcription. *Trends in Genetics* **30**, 103–110 (2014).
15. Jacobs, S. A. & Khorasanizadeh, S. Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science (80-.).* **295**, 2080–2083 (2002).
16. Vakoc, C. R., Mandat, S. A., Olenchock, B. A. & Blobel, G. A. Histone H3 lysine 9 methylation and HP1γ are associated with transcription elongation through mammalian chromatin. *Mol. Cell* **19**, 381–391 (2005).
17. McNally, J. C., Müller, W. G., Walker, D., Wolford, R. & Hager, G. L. The glucocorticoid receptor: Rapid exchange with regulatory sites in living cells. *Science (80-.).* **287**, 1262–1265 (2000).
18. Mattout, A. *et al.* Heterochromatin Protein 1β (HP1β) has distinct functions and distinct nuclear distribution in pluripotent versus differentiated cells. *Genome Biol.* **16**, 213 (2015).
19. Font-Burgada, J., Rossell, D., Auer, H. & Azorín, F. Drosophila HP1c isoform interacts with the zinc-finger proteins WOC and Relative-of-WOC to regulate gene expression. *Genes Dev.* **22**, 3007–3023 (2008).
20. Zhang, D., Wang, D. & Sun, F. Drosophila melanogaster heterochromatin protein HP1b plays important roles in transcriptional activation and development. *Chromosoma* **120**, 97–108 (2011).
21. Kessler, R. *et al.* DDsk2 regulates H2Bub1 and RNA polymerase II pausing at dHP1c complex target genes. *Nat. Commun.* **6**, 7049 (2015).
22. Abel, J., Eskeland, R., Raffa, G. D., Kremmer, E. & Imhof, A. Drosophila HP1c is regulated by an auto-regulatory feedback loop through its binding partner Woc. *PLoS One* **4**, e5089 (2009).
23. Di Mauro, G., Carbonell, A., Escudero-Ferruz, P. & Azorín, F. The zinc-finger proteins WOC and ROW play distinct functions within the HP1c transcription complex. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1863**, 194492 (2020).

24. Kwon, S. H. *et al.* Heterochromatin protein 1 (HP1) connects the FACT histone chaperone complex to the phosphorylated CTD of RNA polymerase II. *Genes Dev.* **24**, 2133–2145 (2010).

25. Stessman, H. A. F. *et al.* Disruption of POGZ Is Associated with Intellectual Disability and Autism Spectrum Disorders. *Am. J. Hum. Genet.* **98**, 541–552 (2016).

26. Nozawa, R.-S. *et al.* Human POGZ modulates dissociation of HP1alpha from mitotic chromosome arms through Aurora B activation. *Nat. Cell Biol.* **12**, 719–27 (2010).

27. Suliman-Lavie, R. *et al.* Pogz deficiency leads to transcription dysregulation and impaired cerebellar activity underlying autism-like behavior in mice. *Nat. Commun.* **11**, 1–15 (2020).

28. Hu, Y. *et al.* Molecular Interaction Search Tool (MIST): An integrated resource for mining gene and protein interaction data. *Nucleic Acids Res.* **46**, D567–D574 (2018).

29. Giot, L. *et al.* A Protein Interaction Map of Drosophila melanogaster. *Science (80-.).* **302**, 1727–1736 (2003).

30. Guruharsha, K. G. *et al.* A protein complex network of Drosophila melanogaster. *Cell* **147**, 690–703 (2011).

31. Rach, E. A. *et al.* Transcription initiation patterns indicate divergent strategies for gene regulation at the chromatin level. *PLoS Genet.* **7**, e1001274 (2011).

32. Ngoc, L. V., Kassavetis, G. A. & Kadonaga, J. T. The RNA polymerase II core promoter in Drosophila. *Genetics* **212**, 13–24 (2019).

33. Juven-Gershon, T. & Kadonaga, J. T. Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Developmental Biology* **339**, 225–229 (2010).

34. Persikov, A. V. & Singh, M. De novo prediction of DNA-binding specificities for Cys2His 2 zinc finger proteins. *Nucleic Acids Res.* **42**, 97–108 (2014).

35. Wood, A. M. *et al.* Regulation of Chromatin Organization and Inducible Gene Expression by a Drosophila Insulator. *Mol. Cell* **44**, 29–38 (2011).

36. Mourad, R. & Cuvier, O. Computational Identification of Genomic Features That Influence 3D Chromatin Domain Formation. *PLoS Comput. Biol.* **12**,

e1004908 (2016).

37. Jiang, N., Emberly, E., Cuvier, O. & Hart, C. M. Genome-Wide Mapping of Boundary Element-Associated Factor (BEAF) Binding Sites in Drosophila melanogaster Links BEAF to Transcription. *Mol. Cell. Biol.* **29**, 3556–3568 (2009).
38. Wu, C. H. *et al.* NELF and DSIF cause promoter proximal pausing on the hsp70 promoter in Drosophila. *Genes Dev.* **17**, 1402–1414 (2003).
39. Lee, C. *et al.* NELF and GAGA Factor Are Linked to Promoter-Proximal Pausing at Many Genes in Drosophila. *Mol. Cell. Biol.* **28**, 3290–3300 (2008).
40. Tsai, S. Y., Chang, Y. L., Swamy, K. B. S., Chiang, R. L. & Huang, D. H. GAGA factor, a positive regulator of global gene expression, modulates transcriptional pausing and organization of upstream nucleosomes. *Epigenetics and Chromatin* **9**, (2016).
41. Vogelmann, J. *et al.* Chromatin Insulator Factors Involved in Long-Range DNA Interactions and Their Role in the Folding of the Drosophila Genome. *PLoS Genet.* (2014). doi:10.1371/journal.pgen.1004544
42. Lee, D. H., Ryu, H. W., Kim, G. W. & Kwon, S. H. Comparison of three heterochromatin protein 1 homologs in Drosophila. *J. Cell Sci.* **132**, jcs222729 (2019).
43. Heurteau, A. *et al.* Insulator-based loops mediate the spreading of H3K27me3 over distant micro-domains repressing euchromatin genes. *Genome Biol.* (2020). doi:10.1186/s13059-020-02106-z
44. Adelman, K., Gilchrist, D., Nechaev, S. & Dos Santos, G. Promoter-proximal pausing of Pol II enhances gene expression. *FASEB J.* (2009).
45. Gilchrist, D. A. *et al.* NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. *Genes Dev.* (2008). doi:10.1101/gad.1643208
46. Raffa, G. D., Cenci, G., Siriaco, G., Goldberg, M. L. & Gatti, M. The putative Drosophila transcription factor woc is required to prevent telomeric fusions. *Mol. Cell* **20**, 821–831 (2005).
47. Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **11**, 2301–2319

(2016).

48. Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods* **13**, 731–740 (2016).
49. Tianfang Ge, D., Tipping, C., Brodsky, M. H. & Zamore, P. D. Rapid screening for CRISPR-directed editing of th drosophila genome using white coconversion. *G3 Genes, Genomes, Genet.* **6**, 3197–3206 (2016).
50. Menet, J. S., Abruzzi, K. C., Desrochers, J., Rodriguez, J. & Rosbash, M. Dynamic PER repression mechanisms in the Drosophila circadian clock: From on-DNA to off-DNA. *Genes Dev.* **24**, 358–367 (2010).
51. Gutin, J. *et al.* Fine-Resolution Mapping of TF Binding and Chromatin Interactions. *Cell Rep.* **22**, 2797–2807 (2018).
52. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
53. Yu, G., Wang, L. G. & He, Q. Y. ChIP seeker: An R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015).
54. Brown, J. B. *et al.* Diversity and dynamics of the Drosophila transcriptome. *Nature* **512**, 393–399 (2014).
55. Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
56. Chen, E. Y. *et al.* Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).
57. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. Revigo summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**, e21800 (2011).
58. Bailey, T. L. *et al.* MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).
59. Bailey, T. L. & MacHanick, P. Inferring direct DNA binding from ChIP-seq. *Nucleic Acids Res.* **40**, (2012).
60. Blecher-Gonen, R. *et al.* High-throughput chromatin immunoprecipitation for genome-wide mapping of in vivo protein-DNA interactions and epigenomic states. *Nat. Protoc.* **8**, 539–554 (2013).
61. Shishkin, A. A. *et al.* Simultaneous generation of many RNA-seq libraries in a

single reaction. *Nat. Methods* **12**, 323–325 (2015).

62. Klein-Brill, A., Joseph-Strauss, D., Appleboim, A. & Friedman, N. Dynamics of Chromatin and Transcription during Transient Depletion of the RSC Chromatin Remodeling Complex. *Cell Rep.* (2019). doi:10.1016/j.celrep.2018.12.020
63. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
64. Derr, A. *et al.* End Sequence Analysis Toolkit (ESAT) expands the extractable information from single-cell RNA-seq data. *Genome Res.* **26**, 1397–1410 (2016).
65. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2009).
66. Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. GOrilla: A tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10**, 48 (2009).