Genome-wide localization of exosome components to active promoters and chromatin insulators in *Drosophila*

Su Jun Lim, Patrick J. Boyle, Madoka Chinen, Ryan K. Dale and Elissa P. Lei*

Laboratory of Cellular and Developmental Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, 9000 Rockville Pike, Bethesda, MD 20892, USA

ABSTRACT

Chromatin insulators are functionally conserved DNA–protein complexes situated throughout the genome that organize independent transcriptional domains. Previous work implicated RNA as an important cofactor in chromatin insulator activity, although the precise mechanisms are not yet understood. Here we identify the exosome, the highly conserved major cellular 3′ to 5′ RNA degradation machinery, as a physical interactor of CP190-dependent chromatin insulator complexes in *Drosophila.* Genome-wide profiling of exosome by ChIP-seq in two different embryonic cell lines reveals extensive and specific overlap with the CP190, BEAF-32 and CTCF insulator proteins. Colocalization occurs mainly at promoters but also boundary elements such as *Mcp*, *Fab-8, scs* and *scs′*, which overlaps with a promoter. Surprisingly, exosome associates primarily with promoters but not gene bodies of active genes, arguing against simple cotranscriptional recruitment to RNA substrates. Similar to insulator proteins, exosome is also significantly enriched at divergently transcribed promoters. Directed ChIP of exosome in cell lines depleted of insulator proteins shows that CTCF is required specifically for exosome association at *Mcp* and *Fab-8* but not other sites, suggesting that alternate mechanisms must also contribute to exosome chromatin recruitment. Taken together, our results reveal a novel positive relationship between exosome and chromatin insulators throughout the genome.

INTRODUCTION

The exosome is a multisubunit complex conserved from archaea to humans that is the major cellular 3′ to 5′ RNA degradation machinery. Involved in the turnover of normal as well as aberrant RNAs, the exosome additionally plays a major role in RNA processing and maturation [reviewed in (1)]. The exosome consists of a core complex including a hexameric ring of RNase PH homology domain-containing subunits (Ski6/Rrp41, Rrp42, Rrp43, Rrp45, Rrp46 and Mtr3) capped by a trimer of S1/KH domain-containing subunits (Csl4, Rrp4 and Rrp40) [reviewed in (2)]. It has been shown that yeast exosomes channel RNA through the center of the core complex (3), but it is the association of either of the hydrolytic RNases Dis3/Rrp44 and Rrp6 with the yeast and human core exosome that provides enzymatic activity of the complex. A contrasting view in *Drosophila* suggests that exosome subunits can function independently or form a continuum of various functional complexes (4).

Although the core exosome and Dis3 localize to both the nucleus and the cytoplasm, the Rrp6 component is predominantly nuclear, suggesting specialized activities for the exosome in the nucleus. Rrp6 alone or the entire exosome have been implicated in several nuclear RNA quality control and surveillance pathways [reviewed in (5)]. Depletion of exosome levels or mutation of exosome components leads to stabilization of cryptic unstable transcripts (CUTs) in yeast, antisense promoter transcripts in mammals (6,7), as well as other aberrant RNAs. In yeast, Nrd1-dependent transcription termination of certain non-coding genes and CUTs from intergenic regions also involves recruitment of the exosome to promote transcript degradation (8–10), raising the possibility of chromatin proximal exosome

---

activity. In these cases, it is not known whether the exosome associates with chromatin in order to carry out its surveillance activities. Toward this end, an over-expressed tagged version of Rrp6 was shown to associate with chromatin of yeast protein coding genes using whole open reading frame cDNA mircroarrays (11). In *Drosophila*, it was demonstrated that certain exosome subunits associate with at least some actively transcribed genes (12,13) and may be recruited to chromatin through interaction with RNA polymerase II (Pol II) elongation factors Spt5 and Spt6 (14). Nevertheless, a high-resolution genome-wide study of exosome chromatin association has yet to be performed in any organism.

Distributed throughout the genome, chromatin insulators are DNA–protein complexes that organize chromatin into independent transcriptional domains. In *Drosophila*, there are at least five chromatin insulator families that can be categorized by the association of particular DNA-binding proteins. These classes include the C CCTC-binding factor (CTCF), Suppressor of Hairy-wing (Su(Hw)), GAGA factor (GAF), Zeste-white 5 (Zw5) and boundary element-associated factor (BEAF-32) [reviewed in (15,16)]. For example, the *hsp70* genes at 87A7 are located between the *scs* and *scs′* boundary elements, bound by Zw5 and BEAF-32, respectively, (17,18). Zw5 and BEAF-32 interact with each other, and this interaction may promote chromatin looping observed between *scs* and *scs′* (19). In addition, the *Fab-8* insulator is a well-characterized *cis*-regulatory region of the *Abdominal-B* (*Abd-B*) locus in the bithorax complex that harbors binding sites for the zinc-finger DNA-binding protein CTCF (20,21). CTCF is required for *Fab-8* insulator function and looping interactions among insulators, enhancers and promoters at *Abd-B* (22–25). Multiple insulator complexes share a common component, Centrosomal protein 190 (CP190), a zinc-finger and BTB/POZ domain-containing protein that may play a global role in chromatin organization (22,26). Since enhancers must often activate their target promoters from long distances, it is likely that insulators act as tethering sites for chromosomal loops that can constrain enhancer–promoter interactions and possibly protect a region from its surrounding chromatin environment.

Depending on their context, chromatin insulators could either repress or promote transcription based on the nature of the higher order chromatin interactions to which they contribute. Recent studies show that several insulator proteins, particularly CP190 and BEAF-32, associate with certain transcriptionally active promoters (27–29). In fact, BEAF-32 appears to be required for transcription at a number of the promoters to which it binds (28). How insulator proteins are targeted to specific promoters is unknown; however, insulator protein recruitment correlates with specific transcription initiation patterns (30). The precise role of insulator proteins at the promoter is still unclear, but these findings suggest that insulator proteins may regulate transcription in a more direct manner than previously suspected.

Earlier work suggests that the catalog of insulator-associated factors and potential regulators is not yet complete. The CP190/CTCF class of insulators interacts with Argonaute2, which promotes *Fab-8* activity as well as insulator-dependent looping at this site (25). Moreover, the *gypsy* class of insulator proteins has been shown to physically interact with the ubiquitin and SUMO ligase Topors and Lamin, a major component of the nuclear matrix (31). Recent work also showed the association of Top2 with *gypsy* insulator complexes, and this factor appears to be important for the stability of the Mod(mdg4)2.2 protein (32). Of particular interest to this study is the finding that the Rm62 helicase interacts with CP190 in an RNA-dependent manner, suggesting that RNA may be a component of the *gypsy* insulator complex (33). Finally, the RNA-binding protein Shep was recently identified as the first known tissue-specific regulator of *gypsy* insulator activity (34). Here, we sought to identify additional RNA-related chromatin insulator factors and examine their potential functional significance.

In this study, we identified the exosome as a novel interactor of CP190 insulator complexes. Genome-wide profiling of exosome components compared with insulator proteins by staining of polytene chromosomes as well as high-resolution ChIP-seq in two different cell lines reveals highly overlapping genome-wide association profiles of the exosome with CP190, BEAF-32 and CTCF insulator proteins. Unexpectedly, we find that the majority of exosome binding sites are situated at the promoters of transcribed genes but not throughout gene bodies or 3′ ends. Exosome chromatin association correlates with active transcription; however, transcript levels at exosome binding sites are not significantly changed in response to exosome depletion. Finally, we found that depletion of BEAF-32 or CP190 has no effect on exosome chromatin association, but reduction of CTCF levels reduces exosome association with the CTCF-dependent *Mcp* and *Fab-8* insulators. These results reveal a previously unknown association between exosome and chromatin insulators throughout the genome.

MATERIALS AND METHODS

Chromatin immunoprecipitation and ChIP-seq

Cells were fixed by addition of 1% formaldehyde to cell media for 10 min at R.T. with gentle agitation. Formaldehyde was quenched by addition of glycine to 0.125 M with gentle agitation for 5 min at R.T. Cells were pelleted at 400 rcf, washed twice in PBS, and resuspended in 1 ml ice–cold cell lysis buffer (5 mM PIPES, pH 8, 85 mM KCl, 0.5% NP-40) supplemented with Complete protease inhibitors (Roche). Nuclei were released by Dounce homogenization with pestle B and pelleted by centrifugation at 9190 rcf for 5 min at 4°C. Nuclei and chromatin were further processed as described previously (35). Immunoprecipitations were performed with 3 ul rabbit α-Rrp6 (14), rabbit α-Rrp40 (36), rabbit α-BEAF-32 (37), rabbit α-CP190 (25), rabbit α-CTCF (22), rabbit α-Su(Hw) (37) or rabbit IgG (Santa Cruz Biotechnology) coupled to rProtein A agarose beads (GE Healthcare). Primers used are listed in Supplementary Table S9.

Samples for ChIP-seq from input DNA, Rrp6 ChIP and Rrp40 ChIP were prepared according to the manufacturer's protocol with standard or TruSeq adapters (Illumina). DNA was sequenced on an Illumina Genome Analyzer II or HiSeq 2000 at the NIDDK Genomics Core Facility. Exosome ChIP-seq data are available at Gene Expression Omnibus (GSE41950).

dsRNA knockdowns

Amplicons used for dsRNA knockdowns were designed based on recommendations from the *Drosophila* RNAi Screening Center. Templates were PCR amplified from genomic DNA using primers containing the T7 promoter sequence (Supplementary Table S9). DsRNAs were produced by *in vitro* transcription of PCR templates using the MEGAscript T7 kit (Ambion) and purified using NucAway Spin Columns (Ambion). S2 and S3 cells were grown at 25°C in Shields and Sang M3 Insect medium (Sigma) supplemented with 0.1% yeast extract, 0.25% bactopeptone and 10% fetal bovine serum (HyClone). Transfections of 1×10^7 cells using 2 µg of dsRNA were performed using Cell Line Nucleofector Kit V (Amaxa Biosystems) transfection reagent using the G-30 protocol. On day 2, cells were diluted with normal media ∼1:5. Five days after transfection, cells were collected, and knockdown efficiency was confirmed by western blotting.

RNA purification and RNA-seq

Total RNA was isolated from S2 and S3 cells using Trizol Reagent (Invitrogen) using the recommended protocol. Polyadenylated RNA and ribosomal (rRNA) depleted RNA was purified from total RNA using the Poly(A)Purist MAG or MicroPoly(A)Purist Kits (Ambion) and the RiboMinus Eukaryote Kit for RNA-Seq (Invitrogen), respectively. Depletion of rRNA was ∼99% by qPCR measurement of 5S rRNA. Sequencing libraries were prepared from Poly(A)$^+$ and rRNA depleted RNA samples according to the manufacturer's protocol (Illumina), and sequencing was performed on an Illumina Genome Analyzer II or HiSeq 2000 at the NIDDK Genomics Core Facility. Exosome knockdown RNA-seq data are available at Gene Expression Omnibus (GSE41950).

CP190 immunoaffinity purification

CP190 complex immunoaffinity purification was performed from nuclear extracts from 0–24 h OR embryos as described previously (33). Columns were used only once for western blotting analyses but reused multiple times for RNase activity assays to conserve material, which led to slightly altered elution profiles.

RNase activity assay

RNA substrate was synthesized by *in vitro* transcription using the pAWG-*su(Hw)* vector containing the *su(Hw)* cDNA. Primers used are listed in Supplementary Table S9. The resulting 99 nt RNA was end labeled with [5'-^{32}P]pCp by incubation with T4 RNA ligase at 4°C O/N and purified by running on a 12% polyacrylamide gel (8 M urea, 1x TAE) followed by extraction of RNA bands by crushing with a motorized pestle and soaking in 0.3 M NaOAc pH 5.2. RNA was ethanol precipitated with 20 µg glycogen. RNA substrate was incubated with buffer alone, Csl4–Flag exosome complexes, or immunoaffinity column fractions in 10 mM Tris-HCl pH 8.0, 50 mM KCl, 5 mM MgCl$_2$ and 10 mM DTT for 2 h at 37°C. Csl4–Flag complexes were purified from transiently transfected S2 cells as described previously (36). Samples were then heated to 65°C for 15 min, immediately cooled on ice, and separated on a 15% polyacrylamide gel (8 M urea, 1x TAE) at 15 W for 45 min. Gels were then exposed to phosphorimager plates overnight and scanned on a Fuji FLA5000 imager.

Immunofluorescence of polytene chromosomes

Polytene chromosome spreads were prepared essentially as described previously (38). Rabbit α-Rrp6, rabbit α-Rrp40, mouse α-BEAF (19) (Developmental Studies Hybridoma Bank), guinea pig α-CP190 (25), guinea pig α-CTCF [generated similarly as in (22)], guinea pig α-Su(Hw) (35) and guinea pig α-Mod(mdg4)2.2 (35) were used for staining.

Western blotting

Western blotting was performed with guinea pig α-CP190, rabbit α-CTCF, guinea pig α-Mod(mdg4)2.2, mouse α-Lamin (39) (ADL67.10, Developmental Studies Hybridoma Bank), mouse α-BEAF, rabbit α-Rrp6, rabbit α-Rrp40, guinea pig α-Rrp4 (36), guinea pig α-Rrp46 (36), guinea pig α-Csl4 (36) and rat α-Ski6 (14).

Computational analyses

ChIP-seq

Reads were mapped with Bowtie 0.12.7 (40) to the dm3 assembly, excluding chrUextra. Only the best uniquely mapping reads allowing two mismatches were kept (parameters –best –strata -m1 -n2 –tryhard -k1). Libraries from technical replicates were merged. Duplicate reads were collapsed with Picard's MarkDuplicates (http://picard.sourceforge.net/command-line-overview.shtml), and reads falling in repetitive regions were removed. Peaks were called with SPP using z.thr = 3, window.size = 1000, an FDR of 1% and otherwise default parameters. Similar results were obtained using the MACS algorithm (41).

Binary heat maps

Supervised hierarchical clustering of overlap by at least 1 bp was performed as in (42), using the multi-intersect program from BEDTools (43) (with the cluster option) and the binary_heatmap() function from pybedtools (44).

Overlap enrichment heat maps

Data files containing called peaks were downloaded and converted to BED files (25,45–55). Endogenous siRNA cluster coordinates were used directly as reported in supplementary tables, or raw reads from Czech *et al.* and Fagegaltier *et al.* were clustered *de novo* as previously defined (25).

Enrichment scores were calculated as described previously (25). The full enrichment matrix was hierarchically clustered using correlation as a distance metric and complete linkage clustering as implemented in SciPy, with rows clustered identically as columns. Selected rows from the full clustered matrix in Supplementary Figure S1 are shown in Figure 4B. For the active regions heatmap, active regions were first defined as any region bound by both Pol II and H3K4me3 using data for these factors in S2 cells from modENCODE (http://intermine. modencode.org/). Genome-wide features used in the full heat map were then subsetted so that only those features overlapping an active region by at least 1 bp were considered. Enrichment scores were generated similarly to the full heat map, except that when features were shuffled, the new random locations were required to fall within an active region. The same row/column ordering in Figure 4B was applied in order to facilitate comparison.

Pie charts

Feature classes [TSSs (1 bp transcript start position), CDSs, introns, 5′UTRs and 3′UTRs] were extracted from all annotated isoforms of all annotated genes in FlyBase release 5.33. Intergenic regions were defined as the remainder of dm3. Since a ChIP-seq peak can fall in more than one class, we classified a peak by its highest priority annotation class, where the priorities from highest to lowest are TSS, CDS, intron, 5'UTR, 3'UTR and intergenic.

TSS matrices

Each row in the matrix in Figure 5B corresponds to the genomic region $+/- 1$ kb around each TSS split into 20 bp bins. Reads were extended 3' to a total length of 200 bp to represent the fragment size. For each column i in the matrix representing a 20 bp-wide genomic location, the input-normalized value in reads per million mapped reads (RPMMR) was calculated as $(IP_i / IP_{total}) - (input_i / input_{total})$ where IP_i and $input_i$ are the numbers of reads overlapping that region in the IP and input, respectively, and IP_{total} and $input_{total}$ are the total number of mapped reads, in millions, in the libraries. Active genes were defined as those with both Pol II and H3K4me3 within a window extending 250 bp upstream and 750 bp downstream of the TSS. We defined ambiguous genes as having either Pol II or H3K4me3 but not both in this window and inactive genes as having neither Pol II nor H3K4me3 in this window. Within each of the three categories, rows are sorted by row means. Line plots show the column sums of the TSS matrix (Figure 5B) or the column sums of a similarly constructed poly (A) site matrix in which each row is centered on the 3'-most co-ordinate of each isoform.

Note that modENCODE supplies several data sets for Pol II (accessions 3295 and 329), CTCF (283, 3281 and 913), CP190 (925, 280) and BEAF-32 (922, 274). In the enrichment analysis heatmap, these data sets were treated separately. In all other analyses, data sets have been combined by first concatenating all data sets for a factor and then merging any features that overlap by at least 1 bp into a single feature.

Hypergeometric overlap tests

To assess the enrichment of insulators and exosome specifically at active TSSs, we used the hypergeometric overlap test (56) where n is the total number of non-redundant annotated isoforms of all active genes; $n1$ is the number of isoforms with an exosome peak at an active TSS, $n2$ is the number of isoforms with an insulator peak at an active TSS and m is the number of isoforms with both exosome and insulator peak at an active TSS. Active TSSs were identified as those with both Pol II and H3K4me3 overlapping the TSS. For each factor, the bound active TSSs consisted of the set of active TSSs that had at least 1 bp overlap with at least one binding site for that factor. Similar results were obtained when considering the full set of TSSs, regardless of transcriptional status.

Divergent promoters

Divergent promoters were identified from coding and non-coding genes and were defined as a minus-strand TSS followed by a plus-strand TSS separated by no more than threshold N bp with no intervening genic sequence between them. Several different values of N (100, 250, 500 bp) were used for analyses.

GO analysis

Gene ontology was assessed using DAVID (57,58) with the genome as the background gene set and 'high' stringency setting for clustering. We considered all genes with a Rrp6 or Rrp40 peak anywhere in the gene. DAVID scaled enrichment values are reported.

RNA-seq and differential expression analysis

Libraries were sequenced on both GAII and HiSeq 2000 machines in either single-end and paired-end mode. Due to small insert sizes, only the first end of the PE libraries was used. Adapters were clipped using cutadapt 1.0 (59). Clipped reads were mapped with TopHat 1.4.1 (60) to the dm3 assembly, excluding chrUextra. Mapped reads were counted in all annotated coding and non-coding genes in FlyBase r5.33 using HTSeq in 'union' mode, which only includes reads that map unambiguously to a single gene. Counts from technical replicates were summed. Counts from replicates were used in DESeq using 'local' fit type and 'fit-only' sharing mode (61). Comparisons used the 'per-condition' method for estimating dispersions. Differentially expressed genes were considered those with adjusted P-values < 0.05.

Analysis of intergenic transcription, snoRNA and snoRNA precursors

To evaluate changes in upstream regions upon knockdown, reads were counted with HTSeq in 0–500 bp upstream of gene TSSs, all intergenic exosome peaks, and selected *cis*-regulatory regions of *Abd-B*. Subregions that also overlapped annotated genes were subtracted in order to focus on changes in intergenic RNA. Regions <100 bp were also removed. Bound upstream regions were defined as gene-level TSS intersecting an exosome peak. Upstream and intergenic peak regions were merged such that overlapping features were considered a single feature.

Annotated snoRNAs from FlyBase r5.33 were considered and classified as being chromatin-associated based on (62). Putative precursor regions for snoRNA were defined as the largest intron overlapping each snoRNA or ca-snoRNA. Ten intergenic 'orphan' snoRNAs, which do not have an obvious precursor, were not considered. Precursors were then merged such that each would be represented only once in the analysis even though multiple snoRNAs could be in one precursor, or both ca-snoRNAs and snoRNAs could come from the same precursor. For each precursor, the number of processed snoRNA-specific reads in the region was subtracted from the total number of reads in the region in order to isolate precursor-specific reads. These final precursor read counts, as well as counts for all mature snoRNAs including the 'orphan' snoRNAs, were examined.

Regions containing zero reads in all RNAseq libraries were removed. For each rRNA depletion knockdown experiment, DESeq was run as described above for RNA-seq, and all regions were reported.

RESULTS

Exosome and RNase activity copurify with chromatin insulator proteins

Given previous evidence that RNA and RNA-binding proteins associate with chromatin insulators, we sought to determine whether the exosome physically associates with CP190 chromatin insulator complexes. Using a well-established protocol to isolate CP190 insulator components and associated factors such as CTCF (22,25,33), embryonic nuclear extracts were immunoaffinity purified over control preimmune or α-CP190 antibody columns by washing and subsequent elution with increasing $MgCl_2$ concentrations (Figure 1A). Eluates were examined for the presence of insulator proteins as well as exosome subunits. Western blotting verifies the presence of CP190 and Mod(mdg4)2.2 insulator proteins in the α-CP190 eluates as well as the exosome components Rrp6, Rrp4, Rrp46, Csl4 and Ski6 but not in preimmune eluates. In contrast, a negative control protein, HP1a, is not identified as a specific interactor of CP190 (data not shown). Therefore, multiple exosome components copurify with CP190 insulator complexes, suggesting that the intact exosome may be stably associated with CP190 chromatin insulator complexes in the nucleus.

In order to assess whether physical association between the exosome and CP190 chromatin insulator complexes is functionally relevant, we examined whether immunopurified CP190 complexes harbor ribonuclease activity. To this end, a radiolabeled 99 nt *in vitro* transcribed RNA was used as a substrate in *in vitro* degradation assays. As a control, RNA was incubated with exosome complexes immunopurified from S2 cells stably transfected with Csl4-Flag using α-Flag (Figure 1B). Examination of the RNA on a high percentage urea polyacrylamide gel indicates extensive degradation of the RNA as result of incubation with exosome complexes compared with buffer alone. Incubation of the RNA substrate with preimmune or α-CP190 immunoaffinity
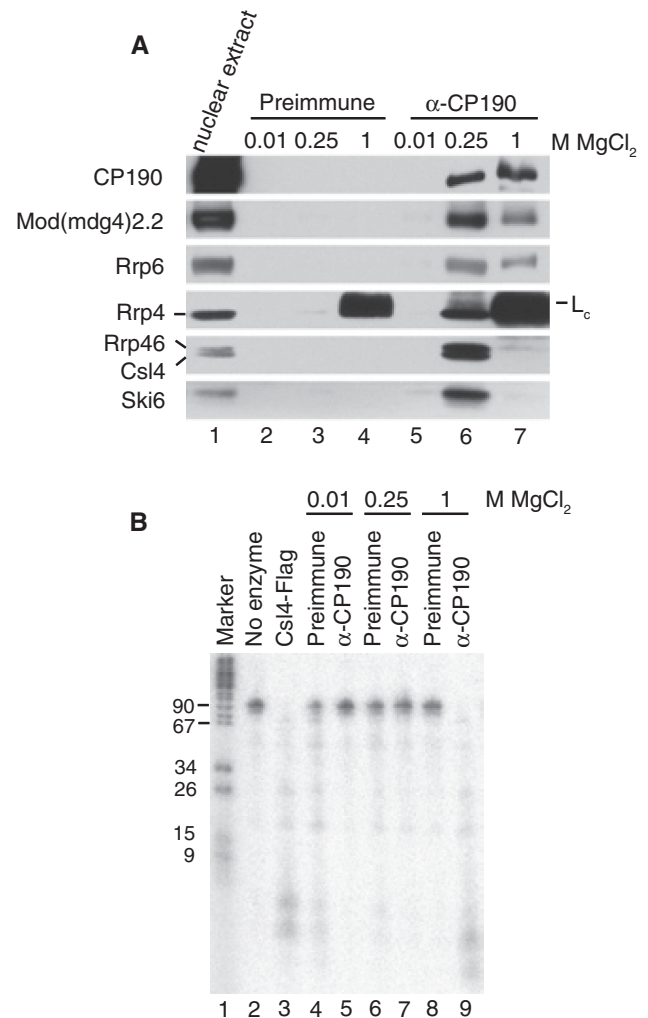

Figure 1. CP190 insulator complexes associate with the nuclear exosome and copurify with RNase activity. (**A**) Exosome subunits physically associate with CP190 insulator complexes. Western blotting of embryonic nuclear extracts (lane 1, 0.4% of input) bound to a control preimmune column (lanes 2–4) or α-CP190 column (lanes 5–7) and step-eluted with increasing $MgCl_2$ concentrations as indicated. Note the presence of the IgG light chain (L_c) in both 1 M $MgCl_2$ elutions in the α-Rrp4 blot. For CP190 blots, 4% of the eluates were probed. For all other blots, 96% of the eluates were probed. (**B**) RNase activity copurifies with CP190 insulator complexes. Shown are DNA marker (lane 1), intact *in vitro* transcribed RNA substrate (lane 2), RNA substrate after incubation with Csl4–Flag purified control complex (lane 3), or RNA substrate after incubation with complexes step-eluted from either preimmune antibody column (lanes 4, 6, 8) or α-CP190 column (lanes 5, 7, 9) at $MgCl_2$ concentrations as indicated. Samples were run on a 15% PAGE urea denaturing gel and exposed to a phosphorimager plate.

column fractions results in extensive degradation specifically in the 1 M α-CP190 elution fraction, while the majority of RNA substrate remains intact in other fractions tested. Lack of activity in the 250 mM CP190 fraction was unexpected but may be due to absence of a critical factor, presence of a repressor or minor differences in purification procedures used for western blotting and RNase activity assays (see Materials and Methods). These results suggest that CP190 chromatin insulators may associate with active exosome complexes *in vivo*.

Exosome subunits colocalize with certain chromatin insulator proteins on polytene chromosomes

Given physical interaction between CP190 and the exosome, we wished to examine the genome-wide binding profiles of the exosome compared with insulator proteins. To this end, we performed immunostaining of highly replicated interphase polytene chromosomes from third instar larval salivary glands. Wildtype chromosomes were stained with either α-Rrp6 or α-Rrp40 and costained for α-CP190, α-BEAF-32, α-CTCF, α-Mod(mdg4)2.2 or α-Su(Hw) insulator proteins (Figures 2–3). Both Rrp6 and Rrp40 localize to DAPI interbands, indicative of localization to transcribed regions of the genome, similar to a previous study of the exosome component Ski6 (14). Since both exosome antibodies were generated in the same organism, double staining of Rrp6 and Rrp40 was not possible, but the localization patterns of both proteins are highly similar. For example, both antibodies intensely label the actively transcribed 'gooseneck' region at cytological position 31. Extensive colocalization is observed for Rrp6 compared with CP190 (Figure 2A) and to a higher extent with BEAF-32 (Figure 2B). In contrast, at this level of resolution, Rrp6 does not appear to colocalize considerably with the insulator proteins CTCF or Mod(mdg4)2.2 (Figure 2C and D). Furthermore, no appreciable overlap was observed between Rrp40 and Su(Hw) (Figure 3). The same results were obtained with both α-Rrp6 and α-Rrp40 antibodies (data not shown). These results suggest that the exosome colocalizes with actively transcribed regions and specific classes of insulator proteins throughout the genome.

High-resolution genome-wide mapping of exosome subunits by ChIP-seq

In order to obtain a high-resolution map of exosome genome-wide chromatin association, we performed ChIP-seq of exosome components in two embryonic cell lines, S2 and S3 cells. We determined Rrp6 and Rrp40 ChIP-seq profiles using Illumina sequencing and peak calling using the SPP algorithm at 1% FDR (63). Greater than 2700 exosome peaks were observed for Rrp6 or Rrp40 in either S2 or S3 cells (Figure 4A). High correspondence was observed between Rrp6 and Rrp40 in the same cell type; 82% of Rrp6 in S2 overlaps with Rrp40, and 84% of Rrp6 in S3 overlaps with Rrp40. Similar profiles obtained with antibodies to two distinct exosome components suggests that the intact exosome may associate with chromatin at these sites and furthermore verifies the specificity of the antibodies. Antibodies directed against other exosome components did not successfully detect signal on either polytene chromosomes or by ChIP or were not tested due to limited availability (data not shown). We subsequently refer to either specific subunit profiles or a common set of 'exosome' binding sites for a particular cell type at which both Rrp6 and Rrp40 are bound.

We next examined the overlap of exosome components with that of insulator proteins and other chromatin-associated factors or other features previously profiled in S2 cells. We calculated enrichment scores for two-way overlaps between all factors based on comparison with random shuffling of sites within the same chromosome. As expected, Rrp6 and Rrp40 profiles display among the highest levels of enrichment between one another compared with all other tested factors (Figure 4B). Consistent with polytene chromosome staining, statistically significant levels of enrichment are observed for both exosome factors with CP190 and BEAF-32 (78% and 84% of exosome sites, respectively, Supplementary Table S1). Substantial enrichment is also observed with CTCF (41%), whereas little to no enrichment is observed between either exosome component compared with Su(Hw) and Mod(mdg4)2.2 (20% and 3%, respectively). Unsupervised hierarchical clustering was performed in order to group factors based on overall similarity of enrichment profiles (Supplementary Figure S1). This analysis further indicates correspondence of exosome components with the CP190 and BEAF-32 insulator proteins. Notably, we identified high levels of overlap enrichment as well as hierarchical clustering with Chromator/Chriz, a factor recently implicated in boundary formation in a genome-wide chromatin conformation study (64). In summary, exosome colocalizes significantly with specific insulator proteins on a genome-wide level.

Exosome recruitment correlates with active transcription

Our comparative analyses also revealed extensive overlap of exosome components with chromatin-associated factors and marks indicative of active transcription. Strong enrichment between Rrp6 and Rrp40 is observed with factors such as RNA Pol II and H3K4me3 (Figure 4B). These results were not unexpected given previous work indicating physical interaction between exosome and transcription elongation factors (14). Based on this work as well as another study suggesting that the Rrp4 exosome component associates with gene bodies (12), we anticipated that exosome peaks would be observed at the middle and/or 3'-end of genes. In striking contrast, we found that the distribution of exosome peaks with respect to genes is heavily biased to the TSS and not gene bodies (Figures 4C and D). Plots of average Rrp6 or Rrp40 signal in S2 cells normalized to input within 1 kb of an annotated TSS reveal a strong enrichment of exosome components slightly upstream of TSSs, whereas little to no enrichment is observed in the vicinity of polyadenylation sites (Figure 4E). Similar TSS enrichment is also observed for exosome binding in S3 cells (data not shown).

Based on preferential binding to the TSS and the fact that the exosome is the major RNA degradation machinery, we hypothesized that exosome recruitment to chromatin may occur in a transcription-dependent manner. When exosome signal in S2 cells is examined at TSSs separated into transcriptionally active or inactive based on the presence of Pol II and H3K4me3, enrichment is observed only for active TSSs (Figure 5A). In order to explore the relationship of binding with gene expression level in more detail, matrices were generated for input normalized ChIP signal in 100 bins +/− 1 kb of annotated TSSs (Figure 5B). As expected, a clear bias toward
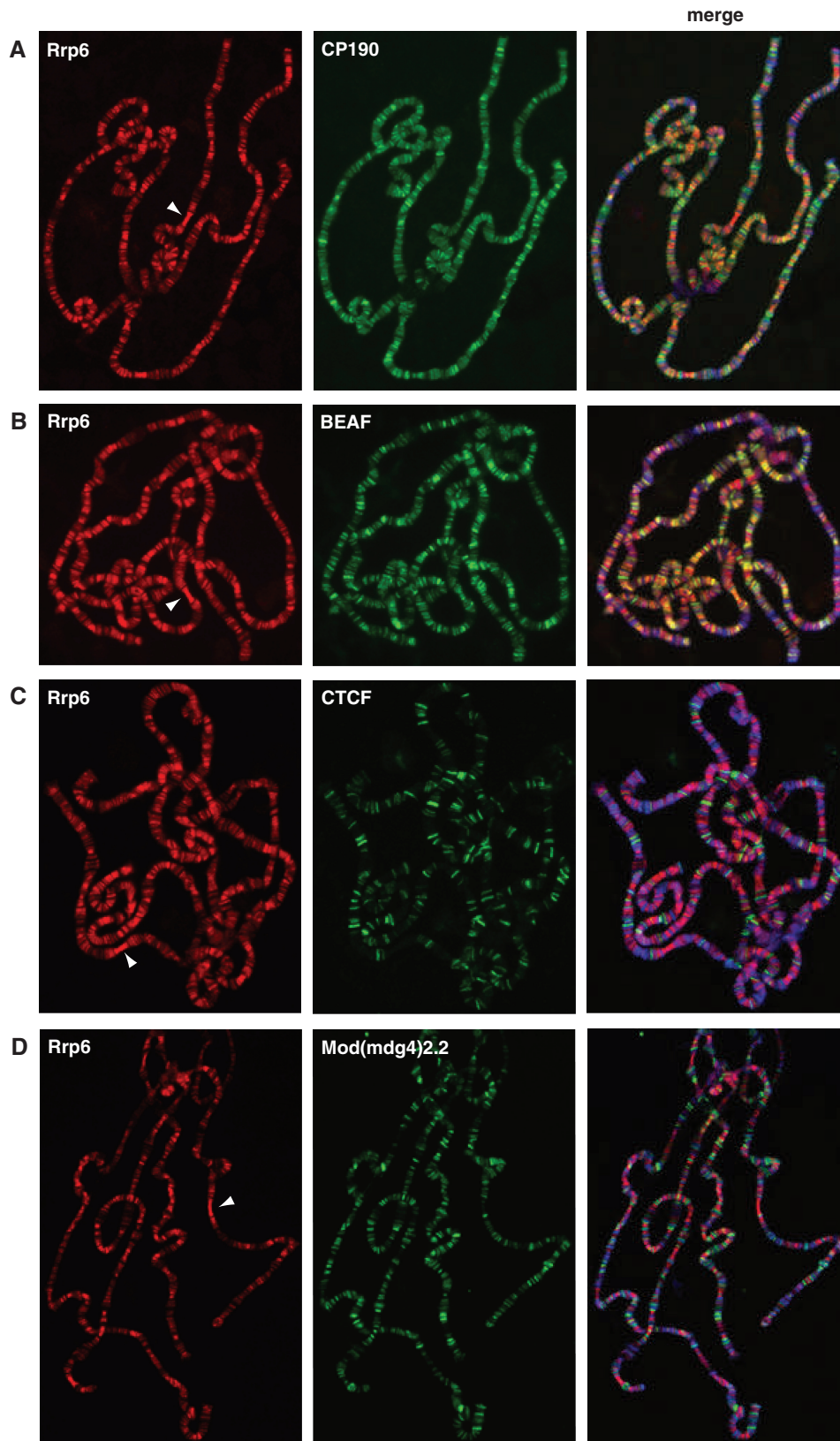
Figure 2. Exosome colocalizes extensively with CP190 and BEAF-32 insulator proteins on polytene chromosomes. Rabbit α-Rrp6 was detected with α-rabbit conjugated Alexa-594 (red). Guinea pig α-CP190 (**A**), mouse α-BEAF-32 (**B**), guinea pig α-CTCF (**C**) or guinea pig α-Mod(mdg4)2.2 (**D**) were detected with α-guinea pig or α-mouse conjugated Alexa-488 (green) as appropriate. DAPI-stained DNA is shown in blue in the merged image (right). White arrowheads point to the transcriptionally active 'gooseneck' region at cytological position 31. Yellow signal in the merged images indicates a high degree of overlap of Rrp6 with CP190 and BEAF-32.
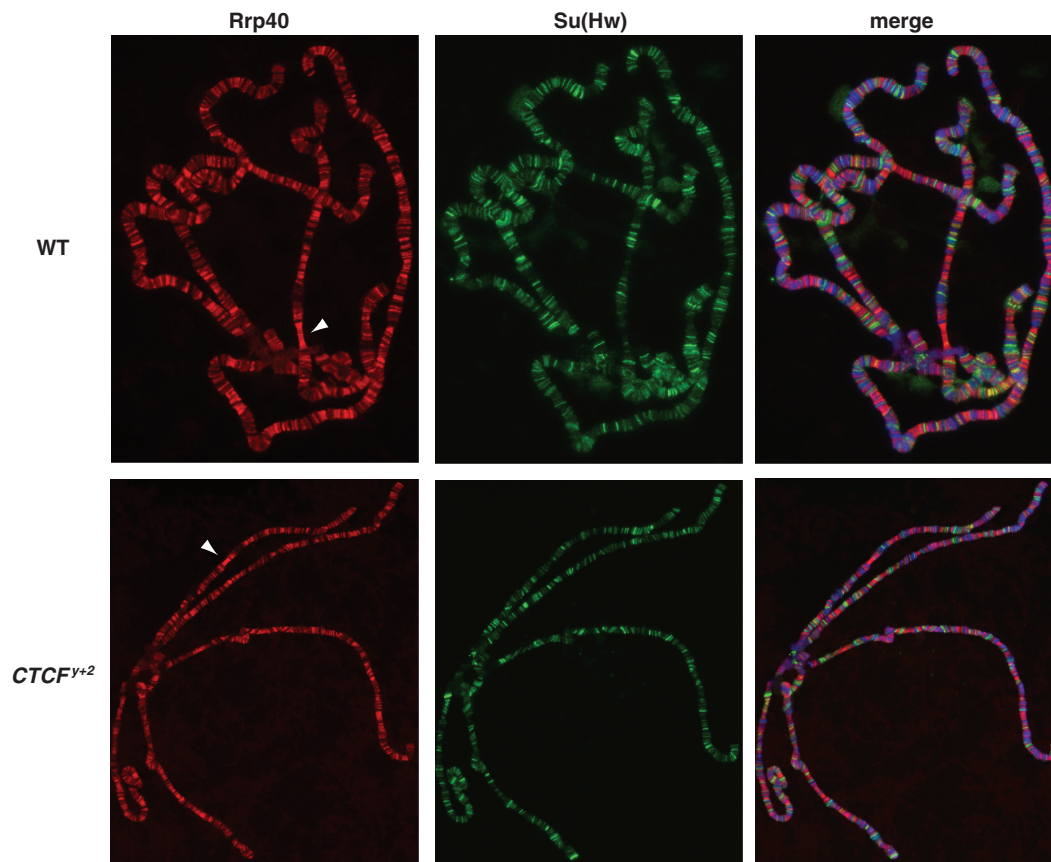
Figure 3. Exosome chromatin association is not globally affected in *CTCF^{y+2}* mutants. Localization of Rrp40 (red) and the Su(Hw) insulator protein (green) on wildtype (top) or *CTCF^{y+2}* mutant (bottom) larval salivary gland polytene chromosomes. Rabbit α-Rrp40 was detected with α-rabbit conjugated Alexa-594. Guinea pig α-Su(Hw) was detected with α-guinea pig conjugated Alexa-488. DAPI-stained DNA is shown in blue in the merged image (right). White arrowheads point to the transcriptionally active 'gooseneck' region at cytological position 31.

expressed genes is observed for exosome ChIP signal at the TSS. Strong enrichment is seen for exosome at a set of promoters that display promoter-proximal enriched Pol II (PPEP, Figure 4B), suggesting that exosome preferentially associates with genes with paused Pol II. Taken together, we conclude that exosome chromatin association correlates with active transcription.

Overlap of exosome with insulator proteins at specific transcriptionally active sites

Given that exosome and certain insulator proteins are enriched at transcriptionally active regions and TSSs, we wanted to verify that enrichment of overlap observed throughout the genome is not due simply to a general preference for active regions. We first assessed the overlap of exosome with insulator proteins at genomic sites corresponding to active TSS, inactive/ambiguous TSS and non-TSS sites in S2 cells (Supplementary Figure S2, Supplementary Table S1). As expected, the majority of sites at which exosome and insulator proteins overlap corresponds to active TSSs. However, it is apparent that many active TSSs are not co-occupied by exosome and insulator proteins. Therefore, we calculated the statistical probabilities of two-way overlap between exosome and CP190, BEAF-32 or CTCF sites by

hypergeometric tests considering only active TSSs of all annotated gene isoforms and indeed observed highly significant overlap over expectation ($P < 3.7 \times 10^{-110}$ for each comparison). We also determined that enrichment scores for two-way overlaps between exosome and insulator proteins are similar when considering only actively transcribed regions of the genome (Supplementary Figure S3) compared to when transcriptionally inert regions of the genome are also considered (Figure 4B). We conclude that colocalization of exosome and CP190, BEAF-32 and CTCF occurs at specific regions of the genome and is not solely due to active transcription status.

Exosome preferentially associates with divergently transcribed genes as well as specific functional gene classes similar to BEAF-32

Like BEAF-32, as well as CP190 and CTCF, exosome binding is enriched at TSS of divergently transcribed genes. Previous work showed that BEAF-32 preferentially associates with the promoters of 'head-to-head' or divergent gene pairs (28). Thus, we performed a series of Fisher's Exact tests to examine whether exosome-bound and insulator-bound regions associate preferentially with divergent versus non-divergent TSSs. We confirmed the enrichment of BEAF-32 at divergent genes using different
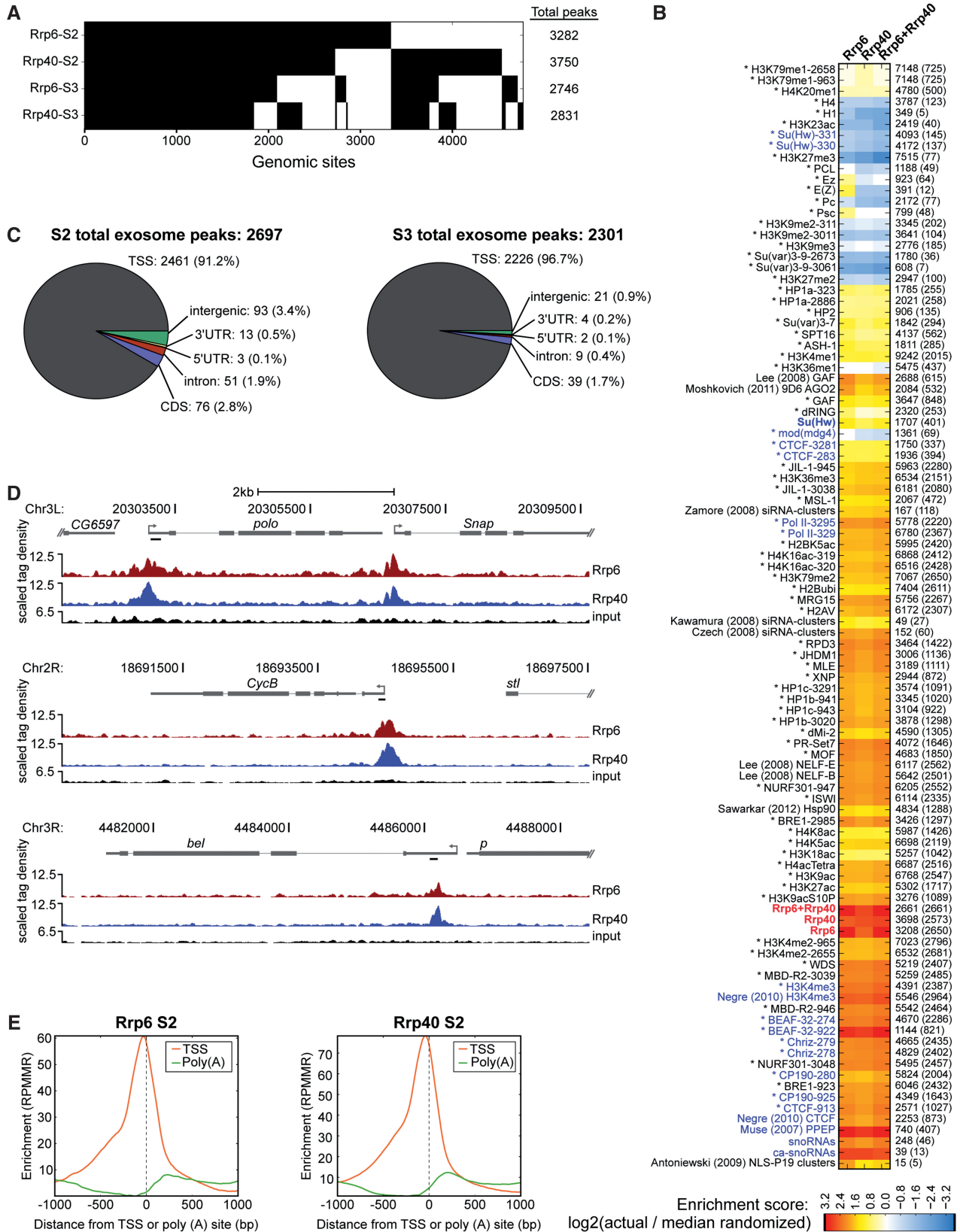
Figure 4. ChIP-seq analysis of exosome components in two embryonic cell lines. (**A**) Binary heat map of Rrp6 and Rrp40 binding sites in S2 and S3 cells ordered by supervised hierarchical clustering. Each column represents a single genomic location, and a black mark in a row represents the presence of a particular factor at that site. Total peak counts are indicated for each factor (right). (**B**) Heat map of log$_2$ enrichment scores for pairwise comparisons of binding sites for Rrp6, Rrp40 and exosome with additional data sets in S2 cells. Color scale corresponding to enrichment value is indicated (bottom). Positive values indicate significant enrichment whereas negative values indicate significant negative enrichment. Self–self
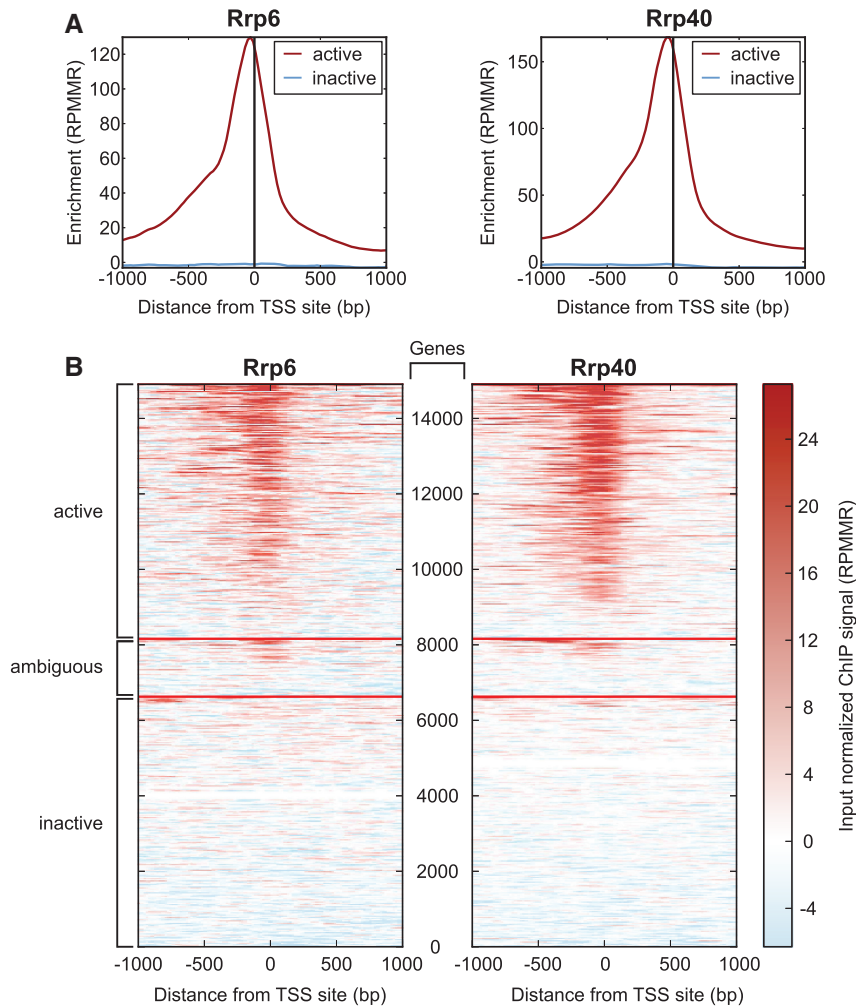
(continued)

Figure 5. Genome-wide exosome chromatin association correlates strongly with active transcription. (**A**) Exosome associates preferentially with active promoters. Profile of S2 Rrp6 and Rrp40 input normalized ChIP-seq tag density around TSSs associated (red) or not associated (blue) with H3K4me3 and Pol II 250 bp upstream or 750 bp downstream of the TSS. (**B**) Matrices of Rrp6 and Rrp40 S2 ChIP signal across all annotated TSSs +/− 1 kb. Each row corresponds to an annotated gene, and red lines indicate grouping of rows into active, ambiguous or inactive TSSs. Rows are ordered by mean signal from highest to lowest within each subgroup. Active genes were defined as genes with Pol II and H3K4me3 within 250 bp upstream or 750 bp downstream of the TSS, ambiguous genes had one of these marks within this window, and inactive genes had neither mark in the same window. Signal intensity is shown in input-normalized enriched RPMMR. Color scale corresponding to enrichment is indicated (right).

threshold distances and also observed enrichment for CP190, CTCF and exosome (Supplementary Figure S4). In contrast, Su(Hw) and Mod(mdg4)2.2 display minimal or negative enrichment for divergent promoters, respectively. Interestingly, we found that divergent genes display notably higher expression than non-divergent genes (Supplementary Figure S5). Indeed, Pol II is also highly enriched at divergent TSSs (Supplementary Figure S4),

raising the possibility that enrichment at divergent promoters may be at least partially due to elevated transcriptional activity at these regions.

In an attempt to better understand the biological significance of exosome chromatin association, we performed gene ontology analysis to identify functional classes of genes enriched for exosome binding. DAVID analysis (57) on exosome binding sites reveals extensive

Figure 4. Continued
comparisons are indicated in gray, and pairwise comparisons that are not statistically significant ($P > 0.001$) are indicated in white. Numbers to the right of each row indicate the total number of features in each data set, and the number of sites overlapping with exosome are indicated in parentheses. Data sets in bold were generated in this study. Data generated by modENCODE consortium is marked with an asterisk. Other data are from studies as indicated (25,48–55). The ca-snoRNA data set was defined in embryos (62). Full heat map with hierarchical clustering is shown in Supplementary Figure S1. (**C**) Classification of Rrp6 and Rrp40 ChIP-seq peaks in S2 and S3 cells with respect to annotated genes. Number of sites and percentage of total in parentheses corresponding to TSS, transcription start site; CDS, coding sequence; 5′ UTR, 5′ untranslated region; 3′ UTR, 3′ untranslated region. See methods for classification hierarchy of overlapping categories. (**D**) Screenshots of input, Rrp6 and Rrp40 ChIP-seq signals at the *polo*, *CycB* and *bel* loci in S3 cells. The bottom of each scale bar indicates 0. Black bars indicate primer sets used for directed ChIP in Figure 6B. (**E**) Genome-wide profiles of Rrp6 and Rrp40 S2 input normalized ChIP-seq tag density around transcription start sites (orange) or poly (A) sites (green) for all genes.

overrepresentation of cell cycle and ribosomal protein genes in both S2 and S3 cells (Supplementary Table S2). Consistent with our finding that exosome colocalizes with BEAF-32 genome-wide, BEAF-32 has previously been shown to associate with cell cycle genes (29). Overall, these results demonstrate that the genome-wide exosome chromatin binding profile shares features with that of CP190, BEAF-32 and CTCF insulator proteins.

Lack of stabilized transcripts at exosome-bound regions

We wished to address whether chromatin regions bound by the exosome correspond directly to sites of active RNA degradation. A previous study in human cells revealed the stabilization of an ~1.5 kb polyadenlyated divergent transcript in the opposite orientation at promoters of many genes (6). In order to determine whether a similar phenomenon occurs in *Drosophila*, we performed dsRNA knockdown of GFP as a control, as well as *Rrp6* and *Rrp40* in S3 cells, harvested total RNA, and performed rRNA depletion followed by non-directional RNA-seq. Efficient knockdown of exosome components was confirmed by western blotting (Supplementary Figure S6), although residual protein could be detected by ChIP (data not shown). Visual inspection of our RNA-seq data did not uncover obvious evidence of upstream divergent transcription in either control or exosome knockdown libraries (data not shown). RNA-seq libraries sequenced to a depth of over 20 M reads from two biological replicates each of control versus knockdown cell lines were analysed for differential expression specifically in the 500 bp intergenic region upstream of exosome-bound TSSs using the DESeq algorithm, and no statistically significant upregulation of these exosome-bound upstream regions was observed (Supplementary Tables S3 and S4).

As an alternate approach, we also examined distributions of genome-wide RNA fold changes upon exosome subunit knockdown at sites upstream of TSSs either bound or unbound by exosome. For exosome-bound upstream regions, an approximately equal number of sites show an increase or decrease of RNA-seq reads in Rrp6 or Rrp40 depleted cells compared with control cells (Supplementary Figure S7). In contrast, regions upstream of TSSs unbound by exosome actually display slightly increased median expression in Rrp6 depleted cells. Furthermore, no statistically significant changes in expression were detected at intergenic exosome peaks by DESeq analysis (Supplementary Tables S3 and S4). Similar results were obtained with oligo-dT selected RNA-seq libraries from either S2 or S3 cells (data not shown). Therefore, we conclude that exosome chromatin recruitment does not necessarily correspond to regions of transcript degradation, and we find no clear evidence for extensive divergent transcription even when the major cellular RNA degradation machinery is depleted.

Given that exosome is implicated in snoRNA maturation in yeast, we examined whether depletion of exosome subunits leads to stabilization of snoRNA or putative precursors. In *Drosophila*, the majority of snoRNAs are derived from introns of protein coding

or non-coding genes (65). We used DESeq expression analysis to examine mature snoRNAs and intron regions harboring snoRNAs in rRNA depleted exosome knockdown versus control S3 libraries. We were unable to detect evidence for genome-wide stabilization of mature snoRNA or precursors (Supplementary Tables S3 and S4); however, we did observe accumulation of two likely snoRNA precursors with 3′ extensions when either Rrp6 or Rrp40 is knocked down (Supplementary Figure S8). Both of these snoRNA genes are bound by exosome, in addition to 61 of 249 annotated snoRNAs genome-wide considering either S2 or S3 cells. Two-way overlap between exosome and snoRNA genes is enriched over random expectation (Figure 4B), and interestingly, exosome preferentially associates with a class of chromatin-associated snoRNAs (ca-snoRNAs, 21 of 40) suggested to be involved in maintaining open chromatin structure (62). These results suggest that the *Drosophila* exosome machinery affects processing of at least a subset of snoRNA transcripts and might perform this function in association with chromatin.

Exosome binding at the TSS does not correlate with exosome-dependent transcript regulation genome-wide

We wondered whether exosome may be recruited to genes to enhance transcript degradation. In order to address this possibility, we performed mRNA expression profiling using control, *Rrp6*, and/or *Rrp40* knockdown RNA-seq libraries from S2 and S3 cells using the DESeq algorithm, this time at the gene-level for annotated genes. Similar to a previous study (66), we identified hundreds of transcripts with altered expression profiles upon knockdown, the majority of which are stabilized upon exosome depletion (Supplementary Tables S5–S7). Strikingly, we found that transcripts that increase in exosome knockdowns are actually less likely to harbor an exosome peak at the TSS compared with transcripts unaffected by knockdown (Supplementary Table S8). Negative enrichment is likely due to the fact that exosome associates with higher expression genes whereas transcript stabilization is biased toward lower expression transcripts (Supplementary Figure S9). Hence, no obvious genome-wide correlation is apparent between exosome-dependent effects on gene expression and exosome chromatin association.

Curiously, depletion of exosome in S3 cells led to stabilization of an aberrant transcript at the *Abd-B* locus. The *Abd-B* gene is expressed in posterior cells later in embryonic development and is regulated by a host of downstream *cis*-regulatory elements including the CTCF-dependent boundary elements *Mcp* and *Fab-8*. Upon close visual inspection, we noted stabilization of an ~20 kb polyadenylated transcript extending between downstream of the RE isoform promoter to beyond the RA isoform promoter particularly in *Rrp6* but also in *Rrp40* depleted cells (Supplementary Figure S10). We verified by directional RT-PCR that the transcript is transcribed in the same orientation as the *Abd-B* gene (data not shown), which is normally expressed in S3 but not S2 cells. Transcription in this region was not observed in exosome depleted S2 cells, suggesting that the transcript

Table 1. Summary of phenotypes observed for exosome RNAi hairpin lines using indicated Gal4 drivers

VDRC stock no.	Gene	*Act5C*::Gal4	*vgM*::Gal4	*GMR*::Gal4
35 090	*Dis3*	Lethal	Blistered wings, scutellar bristles missing	50% show necrosis of presumptive head tissue at pupal stage
26 038	*Mtr3*	Lethal	Blistered wings, scutellar bristles missing	Unaffected
35 199	*Rrp6*	Lethal	Blistered wings, scutellar bristles missing	20% show necrosis of presumptive head tissue at pupal stage
24 305	*Rrp42*	Lethal	Blistered wings, scutellar bristles missing	Not tested
32 612	*Ski6*	Lethal	Blistered wings, scutellar bristles missing	Unaffected

requires the same factors needed for normal *Abd-B* expression. We also observed evidence for expression of a similar transcript during the 4–10 h window of embryonic development based on modEncode oligo-dT selected RNA-seq data from whole embryos (Supplementary Figure S10). Aberrant expression of this nature was not observed extensively throughout the genome, and the functional or mechanistic significance of this stabilized transcript is not known.

Exosome subunits are essential in *Drosophila*

To date, specific mutations in exosome subunit genes have not been reported in *Drosophila*, precluding *in vivo* analyses of exosome function in the context of organismal development. We obtained a panel of UAS-inducible transgenic double-stranded RNA (dsRNA) hairpin lines directed against exosome subunit genes from the Vienna *Drosophila* RNAi Center and tested the effects of knockdowns using various Gal4-drivers on general development and adult viability. As expected, expression of RNAi hairpins against *Dis3*, *Mtr3*, *Rrp6*, *Rrp42* or *Ski6* using the strong ubiquitous *Act5C*::Gal4 driver results in adult lethality (Table 1), indicating that individual exosome subunits are essential for viability and are non-redundant. We additionally tested RNAi hairpins in limited subsets of tissue and observed necrosis of presumptive head tissue in pupae when driving with the *GMR*::Gal4 eye-restricted driver, although this phenotype was not fully penetrant or observed with all knockdown lines. Furthermore, driving of all hairpins using the *vestigial* wing margin enhancer *vgM*::Gal4 resulted in necrosis and severe wing blistering as well as loss or disrupted arrangement of scutellar bristles. Based on these results, we were unable to utilize exosome hairpin lines to assay insulator-dependent phenotypes *in vivo*. Nevertheless, these findings confirm the essential role of the exosome in general fly development, similar to the insulator proteins CP190 (26) and CTCF (22,23).

CTCF is required for exosome recruitment to certain sites including *Fab-8*

Given physical association between the exosome and CP190 insulator complexes as well as extensive genome-wide overlap, we next examined whether insulator proteins are required for exosome chromatin association. In order to deplete insulator proteins, S3 cells were transfected with dsRNAs to *GFP* as a control, *BEAF-32*, *CP190* or *CTCF*. Western blotting verified highly efficient knockdown of insulator proteins but no change in levels of the exosome components Rrp6 or Rrp40 (Figure 6A).

Directed ChIP followed by quantitative PCR was performed for a variety of well-characterized chromatin insulator sites, such as *scs*, *scs′*, *Mcp* and *Fab-8* as well as several non-divergent cell cycle gene promoters strongly enriched for exosome binding. We also examined the highly transcribed *RpL32* coding region, as a control site.

We first examined the chromatin association profiles of Rrp6 and Rrp40 compared with that of insulator proteins in control cells. In *GFP*-treated cells, Rrp6 associates with all regions except *RpL32* substantially over the IgG control (Figure 6B). For Rrp40, signal is strongest at promoters, lower at *Fab-8* and *scs′*, and essentially at background levels at *scs* and *Mcp*. The BEAF-32 insulator protein is also very highly enriched at these promoter regions as well as its characterized binding site *scs′*. We note that the *scs′* insulator overlaps the *aurora* promoter, which is also part of a divergently transcribed gene pair. However, BEAF does not associate with *scs*, *Mcp* or *Fab-8* insulators. In contrast, CP190 and CTCF insulator proteins preferentially associate with characterized binding sites *Mcp* and *Fab-8* and to a lesser extent with certain promoters. Therefore, exosome subunit binding partially mirrors that of insulator proteins at these insulator and promoter sites.

As a control for knockdown efficiency, we next verified that BEAF-32, CP190 and CTCF knockdowns reduce chromatin association of the respective target protein at all binding sites. We thus examined interdependence of insulator proteins to their respective binding sites. BEAF-32 is not significantly affected by knockdown of either CP190 or CTCF at *scs′* or *scs* insulator sites or at the *polo*, *CycB* or *bel* promoters (Figure 6B). Furthermore, CP190 chromatin association at these sites is unaffected by BEAF-32 depletion, consistent with a recent study showing independence of BEAF-32 and CP190 binding despite extensive genome-wide overlap (67). Confirming previous work, CP190 ChIP is reduced in CTCF depleted cells at the insulator sites *Mcp* and *Fab-8*, (25) but is mainly unaffected at the *polo*, *CycB* and *bel* promoters (Figure 6B). Similar to a previous study (67), CTCF binding is unaffected by either BEAF-32 or CP190 depletion at all sites tested including the *Fab-8* insulator, which contrasts with dependence on CP190 previously identified in S2 cells (25). This result may be due to differences in cell type or expression state of the *Abd-B* locus. Interestingly, Rrp6 chromatin association is substantially reduced at *Fab-8* and at *Mcp* and weakly reduced at the *CycB* promoter uniquely in CTCF depleted cells, whereas binding to all other loci tested remains constant under all knockdown conditions
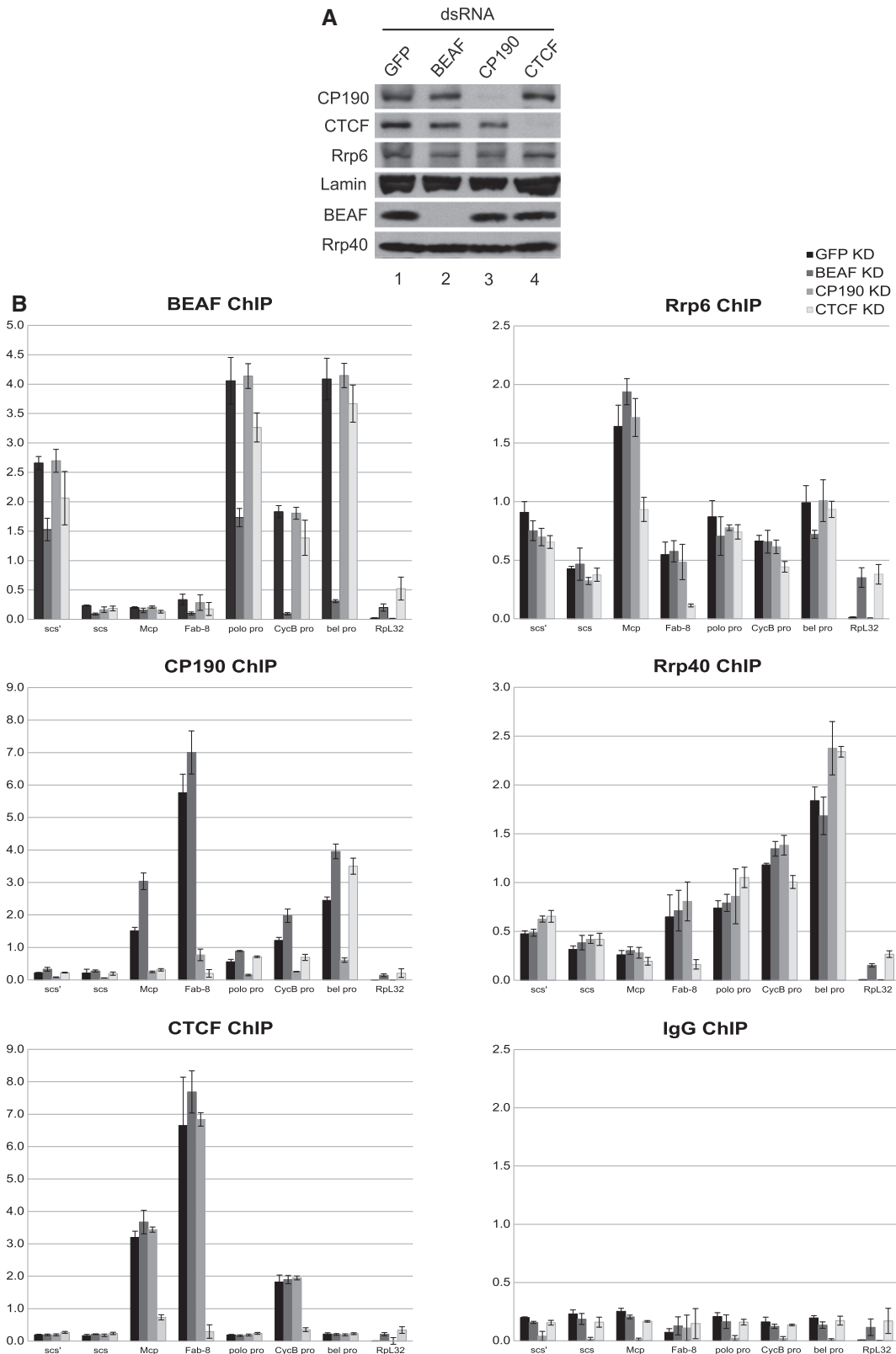
Figure 6. CTCF is specifically required for exosome chromatin association at the *Fab-8* insulator. (**A**) Western blotting of lysates from S3 cells transfected with GFP (lane 1), BEAF-32 (lane 2), CP190 (lane 3) or CTCF (lane 4) dsRNA. (**B**) S3 cells transfected with GFP (black), BEAF-32 (dark gray), CP190 (gray) or CTCF (light gray) dsRNA were subjected to ChIP using α-BEAF-32, α-CP190, α-CTCF, α-Rrp6, α-Rrp40 or control rabbit IgG as indicated. Percent input DNA immunoprecipitated is shown for each primer set, and error bars indicate standard deviation of quadruplicate PCR measurements. Similar results were obtained for independent biological replicates.

(Figure 6B). Similarly, Rrp40 ChIP signal is specifically reduced at *Fab-8* in CTCF depleted cells.

In order to determine whether CTCF plays a genome-wide role in recruitment of the exosome to chromatin, we examined Rrp40 localization in polytene chromosomes of wildtype compared with $CTCF^{y+2}$ mutants. This mutant produces little or no CTCF protein (22) and displays strongly reduced *Abd-B* looping interactions (25). We found no significant differences at this level of resolution (Figure 3), and Rrp6 was similarly unaffected (data not shown). These results indicate dependence on CTCF for exosome recruitment to the *Mcp* and *Fab-8* insulators and possibly the *CycB* promoter or other untested sites. We were unable to assess whether insulator protein chromatin association is dependent on exosome components using the depletion strategy because sufficient levels of knockdown could not be achieved in order to fully remove Rrp6 or Rrp40 from chromatin (data not shown). Our results indicate that the exosome is at least partially dependent on CTCF for chromatin association.

DISCUSSION

Here, we present the first high-resolution map of exosome chromatin association in any organism. In contrast to expectations based on the prior literature, we find that exosome binds mainly promoters but not other genic regions of actively transcribed genes. Our results support the conclusion that exosome associates with chromatin in a transcription-dependent manner; however, exosome is not recruited to all actively transcribed genes. At the genome-wide level, exosome associates extensively with the CP190, BEAF-32 and CTCF insulator proteins. In support of this finding, our results uncover a previously unknown physical association between exosome and chromatin insulators. At the *Mcp* and *Fab-8* insulators, exosome recruitment is reduced by CTCF depletion. However, at most other sites tested, insulator proteins are dispensable for exosome chromatin association. Our results provide new insights into potential functions of the nuclear exosome on chromatin.

Genome-wide profiling of exosome chromatin association

Our findings demonstrate that exosome associates physically with CP190 insulator complexes, and both exosome and specific insulator factors colocalize extensively throughout the genome. We found that exosome copurifies with CP190 insulator complexes isolated from embryonic nuclear extracts. Since CP190 is a component of multiple chromatin insulator complexes, we relied on parallel genome-wide colocalization approaches in a variety of cell types in order to ascertain the specificity of the interaction. Comparison of genome-wide profiles of exosome and insulator proteins on larval salivary gland polytene chromosomes revealed extensive overlap between exosome and the insulator proteins CP190 and BEAF-32, but low correspondence with CTCF, Su(Hw) and Mod(mdg4)2.2. Similarly, high-resolution ChIP-seq profiling in S2 cells confirmed the specific overlap of exosome with CP190 and BEAF-32 and was also able to

detect significant overlap with CTCF that was not apparent in polytene chromosomes. Differential detection of the positive correlation between exosome and CTCF using the two techniques could either be due to differences in cell type or developmental stage examined, or alternatively, the higher resolution of the ChIP-seq method compared with the cytological approach. We analysed our ChIP-seq data in greater detail in order to take advantage of genome-wide profiling data of chromatin-associated factors in the same cell type compiled by other studies and the modENCODE consortium.

We found that exosome colocalizes with CP190, BEAF-32 and CTCF insulator proteins at a specific subset of active promoters. It is currently unclear how this specificity of recruitment is achieved. Previous work in *Drosophila* showed that exosome interacts physically with transcription elongation factors and colocalizes extensively with these factors on polytene chromosomes (14). Another study demonstrated that transfected epitope tagged Rrp4 and Rrp6 can crosslink to the 3′-end of several genes tested, in support of a cotranscriptional RNA surveillance model (12). However, our high-resolution ChIP-seq data clearly shows that exosome recruitment to genes occurs in sharp peaks preferentially at promoters compared with coding regions. Likewise, localized binding is also observed for CP190, BEAF-32 and CTCF to promoters.

Importantly, overlap between exosome and insulator proteins is also observed at well-characterized insulator sequences. Exosome ChIP signal is apparent at many known CP190, CTCF or BEAF-32-dependent insulator sites such as *scs*, *scs′*, *Mcp* and *Fab-8*, and these were further confirmed by directed ChIP. Despite extensive overlap with CP190, BEAF-32 and CTCF insulator proteins, we found that only CTCF is required for exosome recruitment to any sites examined in this study. Depletion of CTCF reduces exosome recruitment to both *Mcp* and *Fab-8* whereas other insulator proteins are dispensable. Since reduction of CTCF levels reduces *Abd-B* expression (23,25), it is possible that reduced exosome chromatin association is a consequence of a change in transcription state in the general vicinity. Finally, considering that insulator proteins do not appear to be required for exosome association with other sites tested including co-occupied promoters, additional alternate mechanisms must exist to recruit exosome to chromatin.

Exosome chromatin association is dependent on active transcription

Our results suggest that exosome is recruited to chromatin in a transcription-dependent manner. Previous work showed that exosome is recruited to heat shock genes upon transcriptional induction (14). Likewise, we found that exosome associates preferentially with active over inactive promoters. These findings would be consistent with an RNA-based mechanism of recruitment to chromatin. However, specific association of exosome with the promoter and not gene bodies implies that exosome is not simply cotranscriptionally recruited to and stably associated with an elongating transcript since its ChIP

signal would likely extend further 3′ into the gene as has been observed for a variety of RNA-binding proteins (68,69). If exosome is indeed recruited through direct RNA interaction, nascent RNAs may only be accessible to the exosome during a brief window, such as prior to capping, before assembly into protein-rich ribonucleo-particles. However, even if recruited to a transcript cotranscriptionally, exosome may not be able to load on the majority of transcripts for degradation since the 3′-end would not be accessible. Another possibility is that exosome specifically recognizes short RNAs associated with stalled Pol II. Finally, exosome recruitment could depend on assembly of the transcriptional apparatus but not the RNA itself. In this case, the full or partially formed exosome complex could be poised for degradation at promoters but not actively engaged.

Stabilization of transcripts as a result of exosome depletion

We were unable to obtain evidence for substantial transcript stabilization corresponding to regions to which exosome associates with chromatin. We did observe exosome at a subset of snoRNA genes, but we only observed a few cases of snoRNA precursor accumulation in exosome knockdowns. Given the strong enrichment of exosome at the TSS and previous reports that divergent transcription is stabilized at human and yeast promoters in exosome depleted cells (6,70,71), we looked for evidence of stabilized transcripts using both oligo-dT selected and rRNA depleted RNA-seq libraries of exosome knockdown cells. However, we found no evidence for divergent transcription upstream of annotated promoters regardless of exosome depletion, consistent with a previous study of uni-directional genome-wide transcription initiation in *Drosophila* (72). One possibility is that divergent upstream transcription does not occur in *Drosophila* and bidirectional transcription from a single promoter may be specific to certain species (73). Other purely technical possibilities are that our RNA-seq libraries were of insufficient depth or failed to capture aberrant transcription or altered stability since a conventional size selection step was used for sequencing, which in effect constrains the lower limit of captured RNA to 80 nt. Perhaps development and use of specialized cloning protocols could successfully identify aberrant transcripts or mild effects on transcript stabilization.

In order to test the possibility that exosome is recruited to chromatin in order to increase the efficiency of RNA surveillance, we also looked for evidence of stabilization of transcripts in exosome knockdown cells at genes at which exosome is associated at the promoter. We note that our knockdowns could not fully remove exosome from chromatin; therefore, sufficient depletion may not have been achieved in order to observe an effect on target RNAs. We did observe stabilization of many transcripts in exosome depleted cells; however these genes do not generally correspond to exosome binding sites. In fact, we observed a negative relationship between the two data sets likely because exosome tends to be recruited to more highly transcribed genes whereas the transcripts stabilized

in exosome knockdown cells tend to be lowly expressed in wildtype cells. Although we cannot rule out that cotranscriptional recruitment is a mechanism to promote efficient RNA surveillance on a subset of genes, this does not appear to be a widespread or efficient mechanism. Another possibility is that recruitment of exosome to chromatin poises the degradation machinery for action under specific yet unknown conditions or to degrade transcripts synthesized *in trans* to binding sites.

Although not a focus of our study, our expression analysis in exosome subunit depleted cell lines revealed mainly stabilization of low expression transcripts. Our results generally confirm a previous study, which methodically depleted each individual exosome subunit in S2 cells and performed gene expression profiling by microarray (66). In that study, substantially varying expression profiles were obtained with individual subunit knockdowns, supporting a previously suggested model that the *Drosophila* exosome functions not as a singular degradation machine but as independently functioning, partially redundant subunits (36). A genome-wide expression profiling of conditional exosome mutants in *Arabidopsis* also detected limited subunit-specific effects on transcript levels (74). We did observe markedly different expression profiles between *Rrp6* and *Rrp40* knockdowns; however, this was likely due to arrest of cell proliferation in Rrp6 but not Rrp40 depleted cells. Our results are consistent with cell cycle arrest specifically observed in Rrp6 knockdown cells as previously reported (36), and these effects substantially confound our expression analyses. Intriguingly, we did find that both exosome subunits preferentially associate with promoters of cell cycle genes, suggesting that exosome could play a direct role in regulation of cell cycle gene transcript regulation. It would be interesting to examine exosome chromatin association at different stages of the cell cycle given that expression of many cell cycle regulators are tightly controlled and regulated in a cell cycle-dependent manner.

Functional relevance of exosome–insulator complex association

Our findings raise the possibility that exosome contributes to insulator activity or regulation, perhaps through an RNA-based mechanism. RNA may contribute to higher order insulator-dependent interactions and has been postulated to be an important component of the *gypsy* insulator complex in *Drosophila* (33,34) and the CTCF/cohesin insulator complex in mammals (75). Thus association of exosome with insulator complexes could regulate the abundance or activity of insulator-associated RNAs, which in turn affect RNA-mediated interactions between insulator complexes and factors capable of regulating insulator activity. A feasible scenario is that exosome regulates expression of the *Abd-B* locus through CTCF-dependent recruitment to the *Mcp* and *Fab-8* insulators. Thus far, no specific RNA has yet been implicated in insulator activity at *Abd-B*. However, a complex array of intergenic transcripts arising from the *cis*-regulatory region of *Abd-B* has been shown to regulate expression

of the downstream homeotic gene *abd-A* (76–78). Therefore, exosome surveillance and degradation activity may be particularly important at this complicated gene region. Isolation of conditional or loss-of-function exosome mutants should help elucidate the precise role of exosome in chromatin insulator activity.

ACCESSION NUMBERS

SUPPLEMENTARY DATA

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Houseley,J., LaCava,J. and Tollervey,D. (2006) RNA-quality control by the exosome. *Nat. Rev. Mol. Cell Biol.*, **7**, 529–539.
2. Lorentzen,E., Basquin,J. and Conti,E. (2008) Structural organization of the RNA-degrading exosome. *Curr. Opin. Struct. Biol.*, **18**, 709–713.
3. Bonneau,F., Basquin,J., Ebert,J., Lorentzen,E. and Conti,E. (2009) The yeast exosome functions as a macromolecular cage to channel RNA substrates for degradation. *Cell*, **139**, 547–559.
4. Kiss,D.L. and Andrulis,E.D. (2011) The exozyme model: a continuum of functionally distinct complexes. *RNA*, **17**, 1–13.
5. Fasken,M.B. and Corbett,A.H. (2009) Mechanisms of nuclear mRNA quality control. *RNA Biol.*, **6**, 237–241.
6. Preker,P., Nielsen,J., Kammler,S., Lykke-Andersen,S., Christensen,M.S., Mapendano,C.K., Schierup,M.H. and Jensen,T.H. (2008) RNA exosome depletion reveals transcription upstream of active human promoters. *Science*, **322**, 1851–1854.
7. Flynn,R.A., Almada,A.E., Zamudio,J.R. and Sharp,P.A. (2011) Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. *Proc. Natl Acad. Sci. USA*, **108**, 10460–10465.
8. Vasiljeva,L. and Buratowski,S. (2006) Nrd1 interacts with the nuclear exosome for 3' processing of RNA polymerase II transcripts. *Mol. Cell*, **21**, 239–248.
9. Arigo,J.T., Eyler,D.E., Carroll,K.L. and Corden,J.L. (2006) Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. *Mol. Cell*, **23**, 841–851.
10. Thiebaut,M., Kisseleva-Romanova,E., Rougemaille,M., Boulay,J. and Libri,D. (2006) Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance. *Mol. Cell*, **23**, 853–864.
11. Hieronymus,H., Yu,M.C. and Silver,P.A. (2004) Genome-wide mRNA surveillance is coupled to mRNA export. *Genes Dev.*, **18**, 2652–2662.
12. Hessle,V., Bjork,P., Sokolowski,M., Gonzalez de Valdivia,E., Silverstein,R., Artemenko,K., Tyagi,A., Maddalo,G., Ilag,L., Helbig,R. *et al.* (2009) The exosome associates cotranscriptionally with the nascent pre-mRNP through interactions with heterogeneous nuclear ribonucleoproteins. *Mol. Biol. Cell*, **20**, 3459–3470.
13. Hessle,V., von Euler,A., Gonzalez de Valdivia,E. and Visa,N. (2012) Rrp6 is recruited to transcribed genes and accompanies the spliced mRNA to the nuclear pore. *RNA*, **18**, 1466–1474.
14. Andrulis,E.D., Werner,J., Nazarian,A., Erdjument-Bromage,H., Tempst,P. and Lis,J.T. (2002) The RNA processing exosome is linked to elongating RNA polymerase II in Drosophila. *Nature*, **420**, 837–841.
15. Gurudatta,B.V. and Corces,V.G. (2009) Chromatin insulators: lessons from the fly. *Brief. Funct. Genomic. Proteomic.*, **8**, 276–282.
16. Maeda,R.K. and Karch,F. (2007) Making connections: boundaries and insulators in Drosophila. *Curr. Opin. Genet. Dev.*, **17**, 394–399.
17. Gaszner,M., Vazquez,J. and Schedl,P. (1999) The Zw5 protein, a component of the scs chromatin domain boundary, is able to block enhancer-promoter interaction. *Genes Dev.*, **13**, 2098–2107.
18. Zhao,K., Hart,C.M. and Laemmli,U.K. (1995) Visualization of chromosomal domains with boundary element-associated factor BEAF-32. *Cell*, **81**, 879–889.
19. Blanton,J., Gaszner,M. and Schedl,P. (2003) Protein:protein interactions and the pairing of boundary elements in vivo. *Genes Dev.*, **17**, 664–675.
20. Barges,S., Mihaly,J., Galloni,M., Hagstrom,K., Muller,M., Shanower,G., Schedl,P., Gyurkovics,H. and Karch,F. (2000) The Fab-8 boundary defines the distal limit of the bithorax complex iab-7 domain and insulates iab-7 from initiation elements and a PRE in the adjacent iab-8 domain. *Development*, **127**, 779–790.
21. Moon,H., Filippova,G., Loukinov,D., Pugacheva,E., Chen,Q., Smith,S.T., Munhall,A., Grewe,B., Bartkuhn,M., Arnold,R. *et al.* (2005) CTCF is conserved from Drosophila to humans and confers enhancer blocking of the Fab-8 insulator. *EMBO Rep.*, **6**, 165–170.
22. Gerasimova,T.I., Lei,E.P., Bushey,A.M. and Corces,V.G. (2007) Coordinated control of dCTCF and gypsy chromatin insulators in Drosophila. *Mol. Cell*, **28**, 761–772.
23. Mohan,M., Bartkuhn,M., Herold,M., Philippen,A., Heinl,N., Bardenhagen,I., Leers,J., White,R.A., Renkawitz-Pohl,R., Saumweber,H. *et al.* (2007) The Drosophila insulator proteins CTCF and CP190 link enhancer blocking to body patterning. *EMBO J.*, **26**, 4203–4214.
24. Kyrchanova,O., Toshchakov,S., Podstreshnaya,Y., Parshikov,A. and Georgiev,P. (2008) Functional interaction between the Fab-7 and Fab-8 boundaries and the upstream promoter region in the Drosophila Abd-B gene. *Mol. Cell. Biol.*, **28**, 4188–4195.
25. Moshkovich,N., Nisha,P., Boyle,P.J., Thompson,B.A., Dale,R.K. and Lei,E.P. (2011) RNAi-independent role for Argonaute2 in CTCF/CP190 chromatin insulator function. *Genes Dev.*, **25**, 1686–1701.
26. Pai,C.Y., Lei,E.P., Ghosh,D. and Corces,V.G. (2004) The centrosomal protein CP190 is a component of the gypsy chromatin insulator. *Mol. Cell*, **16**, 737–748.
27. Bartkuhn,M., Straub,T., Herold,M., Herrmann,M., Rathke,C., Saumweber,H., Gilfillan,G.D., Becker,P.B. and Renkawitz,R. (2009) Active promoters and insulators are marked by the centrosomal protein 190. *EMBO J.*, **28**, 877–888.
28. Jiang,N., Emberly,E., Cuvier,O. and Hart,C.M. (2009) Genome-wide mapping of boundary element-associated factor (BEAF) binding sites in Drosophila melanogaster links BEAF to transcription. *Mol. Cell. Biol.*, **29**, 3556–3568.
29. Emberly,E., Blattes,R., Schuettengruber,B., Hennion,M., Jiang,N., Hart,C.M., Kas,E. and Cuvier,O. (2008) BEAF regulates cell-cycle genes through the controlled deposition of H3K9

methylation marks into its conserved dual-core binding sites. *PLoS Biol.*, **6**, 2896–2910.

30. Rach,E.A., Winter,D.R., Benjamin,A.M., Corcoran,D.L., Ni,T., Zhu,J. and Ohler,U. (2011) Transcription initiation patterns indicate divergent strategies for gene regulation at the chromatin level. *PLoS Genet.*, **7**, e1001274.

31. Capelson,M. and Corces,V.G. (2005) The ubiquitin ligase dTopors directs the nuclear organization of a chromatin insulator. *Mol. Cell*, **20**, 105–116.

32. Ramos,E., Torre,E.A., Bushey,A.M., Gurudatta,B.V. and Corces,V.G. (2011) DNA topoisomerase II modulates insulator function in Drosophila. *PLoS One*, **6**, e16562.

33. Lei,E.P. and Corces,V.G. (2006) RNA interference machinery influences the nuclear organization of a chromatin insulator. *Nat. Genet.*, **38**, 936–941.

34. Matzat,L.H., Dale,R.K., Moshkovich,N. and Lei,E.P. (2012) Tissue-specific regulation of chromatin insulator function. *PLoS Genet.*, **8**, e1003069.

35. Moshkovich,N. and Lei,E.P. (2010) HP1 recruitment in the absence of argonaute proteins in Drosophila. *PLoS Genet.*, **6**, e1000880.

36. Graham,A.C., Kiss,D.L. and Andrulis,E.D. (2006) Differential distribution of exosome subunits at the nuclear lamina and in cytoplasmic foci. *Mol. Biol. Cell*, **17**, 1399–1409.

37. Bushey,A.M., Ramos,E. and Corces,V.G. (2009) Three subclasses of a Drosophila insulator show distinct and cell type-specific genomic distributions. *Genes Dev.*, **23**, 1338–1350.

38. Gerasimova,T.I., Byrd,K. and Corces,V.G. (2000) A chromatin insulator determines the nuclear localization of DNA. *Mol. Cell*, **6**, 1025–1035.

39. Stuurman,N., Sasse,B. and Fisher,P.A. (1996) Intermediate filament protein polymerization: molecular analysis of Drosophila nuclear lamin head-to-tail binding. *J. Struct. Biol.*, **117**, 1–15.

40. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.

41. Zhang,Y., Liu,T., Meyer,C.A., Eeckhoute,J., Johnson,D.S., Bernstein,B.E., Nusbaum,C., Myers,R.M., Brown,M., Li,W. *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.*, **9**, R137.

42. Kim,J., Chu,J., Shen,X., Wang,J. and Orkin,S.H. (2008) An extended transcriptional network for pluripotency of embryonic stem cells. *Cell*, **132**, 1049–1061.

43. Quinlan,A.R. and Hall,I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.

44. Dale,R.K., Pedersen,B.S. and Quinlan,A.R. (2011) Pybedtools: a flexible Python library for manipulating genomic datasets and annotations. *Bioinformatics*, **27**, 3423–3424.

45. Celniker,S.E., Dillon,L.A., Gerstein,M.B., Gunsalus,K.C., Henikoff,S., Karpen,G.H., Kellis,M., Lai,E.C., Lieb,J.D., MacAlpine,D.M. *et al.* (2009) Unlocking the secrets of the genome. *Nature*, **459**, 927–930.

46. Schwartz,Y.B., Kahn,T.G., Stenberg,P., Ohno,K., Bourgon,R. and Pirrotta,V. (2010) Alternative epigenetic chromatin states of polycomb target genes. *PLoS Genet.*, **6**, e1000805.

47. Richter,C., Oktaba,K., Steinmann,J., Muller,J. and Knoblich,J.A. (2011) The tumour suppressor L(3)mbt inhibits neuroepithelial proliferation and acts on insulator elements. *Nat. Cell. Biol.*, **13**, 1029–1039.

48. Muse,G.W., Gilchrist,D.A., Nechaev,S., Shah,R., Parker,J.S., Grissom,S.F., Zeitlinger,J. and Adelman,K. (2007) RNA polymerase is poised for activation across the genome. *Nat. Genet.*, **39**, 1507–1511.

49. Czech,B., Malone,C.D., Zhou,R., Stark,A., Schlingeheyde,C., Dus,M., Perrimon,N., Kellis,M., Wohlschlegel,J.A., Sachidanandam,R. *et al.* (2008) An endogenous small interfering RNA pathway in Drosophila. *Nature*, **453**, 798–802.

50. Ghildiyal,M., Seitz,H., Horwich,M.D., Li,C., Du,T., Lee,S., Xu,J., Kittler,E.L., Zapp,M.L., Weng,Z. *et al.* (2008) Endogenous siRNAs derived from transposons and mRNAs in Drosophila somatic cells. *Science*, **320**, 1077–1081.

51. Kawamura,Y., Saito,K., Kin,T., Ono,Y., Asai,K., Sunohara,T., Okada,T.N., Siomi,M.C. and Siomi,H. (2008) Drosophila

52. Lee,C., Li,X., Hechmer,A., Eisen,M., Biggin,M.D., Venters,B.J., Jiang,C., Li,J., Pugh,B.F. and Gilmour,D.S. (2008) NELF and GAGA factor are linked to promoter-proximal pausing at many genes in Drosophila. *Mol. Cell. Biol.*, **28**, 3290–3300.

53. Fagegaltier,D., Bouge,A.L., Berry,B., Poisot,E., Sismeiro,O., Coppee,J.Y., Theodore,L., Voinnet,O. and Antoniewski,C. (2009) The endogenous siRNA pathway is involved in heterochromatin formation in Drosophila. *Proc. Natl Acad. Sci. USA*, **106**, 21258–21263.

54. Negre,N., Brown,C.D., Shah,P.K., Kheradpour,P., Morrison,C.A., Henikoff,J.G., Feng,X., Ahmad,K., Russell,S., White,R.A. *et al.* (2010) A comprehensive map of insulator elements for the Drosophila genome. *PLoS Genet.*, **6**, e1000814.

55. Sawarkar,R., Sievers,C. and Paro,R. (2012) Hsp90 globally targets paused RNA polymerase to regulate gene expression in response to environmental stimuli. *Cell*, **149**, 807–818.

56. Fury,W., Batliwalla,F., Gregersen,P.K. and Li,W. (2006) Overlapping probabilities of top ranking gene lists, hypergeometric distribution, and stringency of gene selection criterion. *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, **1**, 5531–5534.

57. Huang da,W., Sherman,B.T. and Lempicki,R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.*, **4**, 44–57.

58. Huang da,W., Sherman,B.T. and Lempicki,R.A. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.*, **37**, 1–13.

59. Martin,M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, **17**, 10–12.

60. Trapnell,C., Pachter,L. and Salzberg,S.L. (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, **25**, 1105–1111.

61. Anders,S. and Huber,W. (2010) Differential expression analysis for sequence count data. *Genome Biol.*, **11**, R106.

62. Schubert,T., Pusch,M.C., Diermeier,S., Benes,V., Kremmer,E., Imhof,A. and Langst,G. (2012) Df31 protein and snoRNAs maintain accessible higher-order structures of chromatin. *Mol. Cell*, **48**, 434–444.

63. Kharchenko,P.V., Tolstorukov,M.Y. and Park,P.J. (2008) Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat. Biotechnol.*, **26**, 1351–1359.

64. Sexton,T., Yaffe,E., Kenigsberg,E., Bantignies,F., Leblanc,B., Hoichman,M., Parrinello,H., Tanay,A. and Cavalli,G. (2012) Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell*, **148**, 458–472.

65. Huang,Z.P., Zhou,H., He,H.L., Chen,C.L., Liang,D. and Qu,L.H. (2005) Genome-wide analyses of two families of snoRNA genes from Drosophila melanogaster, demonstrating the extensive utilization of introns for coding of snoRNAs. *RNA*, **11**, 1303–1316.

66. Kiss,D.L. and Andrulis,E.D. (2010) Genome-wide analysis reveals distinct substrate specificities of Rrp6, Dis3, and core exosome subunits. *RNA*, **16**, 781–791.

67. Schwartz,Y.B., Linder-Basso,D., Kharchenko,P.V., Tolstorukov,M.Y., Kim,M., Li,H.B., Gorchakov,A.A., Minoda,A., Shanower,G., Alekseyenko,A.A. *et al.* (2012) Nature and function of insulator protein binding sites in the Drosophila genome. *Genome Res*, **22**, 2188–2198.

68. Lei,E.P., Krebber,H. and Silver,P.A. (2001) Messenger RNAs are recruited for nuclear export during transcription. *Genes Dev.*, **15**, 1771–1782.

69. Komarnitsky,P., Cho,E.J. and Buratowski,S. (2000) Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev.*, **14**, 2452–2460.

70. Neil,H., Malabat,C., d'Aubenton-Carafa,Y., Xu,Z., Steinmetz,L.M. and Jacquier,A. (2009) Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature*, **457**, 1038–1042.

71. Xu,Z., Wei,W., Gagneur,J., Perocchi,F., Clauder-Munster,S., Camblong,J., Guffanti,E., Stutz,F., Huber,W. and Steinmetz,L.M.

(2009) Bidirectional promoters generate pervasive transcription in yeast. *Nature*, **457**, 1033–1037.

72. Ni,T., Corcoran,D.L., Rach,E.A., Song,S., Spana,E.P., Gao,Y., Ohler,U. and Zhu,J. (2010) A paired-end sequencing strategy to map the complex landscape of transcription initiation. *Nat. Methods*, **7**, 521–527.

73. Core,L.J., Waterfall,J.J., Gilchrist,D.A., Fargo,D.C., Kwak,H., Adelman,K. and Lis,J.T. (2012) Defining the status of RNA polymerase at promoters. *Cell Rep.*, **2**, 1025–1035.

74. Chekanova,J.A., Gregory,B.D., Reverdatto,S.V., Chen,H., Kumar,R., Hooker,T., Yazaki,J., Li,P., Skiba,N., Peng,Q. *et al.* (2007) Genome-wide high-resolution mapping of exosome substrates reveals hidden features in the Arabidopsis transcriptome. *Cell*, **131**, 1340–1353.

75. Yao,H., Brick,K., Evrard,Y., Xiao,T., Camerini-Otero,R.D. and Felsenfeld,G. (2010) Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA. *Genes Dev.*, **24**, 2543–2555.

76. Gummalla,M., Maeda,R.K., Castro Alvarez,J.J., Gyurkovics,H., Singari,S., Edwards,K.A., Karch,F. and Bender,W. (2012) abd-A regulation by the iab-8 noncoding RNA. *PLoS Genet.*, **8**, e1002720.

77. Bender,W. (2008) MicroRNAs in the Drosophila bithorax complex. *Genes Dev.*, **22**, 14–19.

78. Tyler,D.M., Okamura,K., Chung,W.J., Hagen,J.W., Berezikov,E., Hannon,G.J. and Lai,E.C. (2008) Functionally distinct regulatory RNAs generated by bidirectional transcription and processing of microRNA loci. *Genes Dev.*, **22**, 26–36.