SUPPLEMENTARY MATERIAL FOR LIM ET AL.

Summary of Supplementary Figures and Tables

Figure S1. Full heat map of pairwise comparisons of binding sites for a particular factor and hierarchical clustering.

Figure S2. Binary heat map of genomic sites corresponding to active TSS, inactive/ambiguous TSS, non-TSS sites, exosome, and insulator proteins in S2 cells.

Figure S3. 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 in actively transcribed regions.

Figure S4. Enrichment of exosome, insulator proteins, and RNA pol II at divergently transcribed TSSs.

Figure S5. Expression levels of divergent versus non-divergent genes.

Figure S6. Western blotting to confirm knockdown of cells transfected with exosome dsRNA.

Figure S7. Histograms of distribution of RNA-seq fold changes observed in upstream TSS regions of exosome depleted cells.

Figure S8. Screenshots of snoRNA precursor accumulation in control versus exosome depleted S3 cells.

Figure S9. Scatterplot of transcript expression levels and fold change in control versus exosome depleted S3 cells.

Figure S10. Screenshot of aberrant transcription at the *Abd-B* locus in Rrp6 and Rrp40 depleted S3 cells.

Table S1. Percent overlap of exosome and insulator protein chromatin binding sites and transcriptionally active TSSs in S2 cells.

Table S2. Summary of gene ontology categories corresponding to exosome TSS binding with a minimum DAVID enrichment score of 3.5 in at least one cell type.

Table S3. DESeq output for differential expression in intergenic regions and snoRNA in rRNA depleted RNA-seq libraries of control versus Rrp6 depleted S3 cells.

Table S4. DESeq output for differential expression in intergenic regions and snoRNA in rRNA depleted RNA-seq libraries of control versus Rrp40 depleted S3 cells.

Table S5. DESeq output for differential expression of annotated gene models in oligo-dT selected RNA-seq libraries of control versus Rrp6-depleted S2 cells.

Table S6. DESeq output for differential expression of annotated gene models in oligo-dT selected RNA-seq libraries of control versus Rrp6-depleted S3 cells.

Table S7. DESeq output for differential expression of annotated gene models in oligo-dT selected RNA-seq libraries of control versus Rrp40-depleted S3 cells.

Table S8. Fisher's exact tests comparing exosome ChIP peaks versus genes changed in expression in control versus *Rrp6* or *Rrp40* knockdown cells.

Table S9. Primers used.

Supplementary Figure Legends

Figure S1. Full heat map of pairwise comparisons of binding sites for a particular factor and hierarchical clustering.

Pairwise comparisons of binding sites for a particular factor as in Figure 4B with hierarchical clustering. Rows are clustered by complete linkage using correlation as the distance metric, and columns are sorted identically to rows.

Figure S2. Binary heat map of genomic sites corresponding to active TSS, inactive/ambiguous TSS, non-TSS sites, exosome, and insulator proteins in S2 cells.

Each column represents a single genomic location, and a black mark in a row represents the presence of a particular factor at that site. Active TSS was determined based on the presence of Pol II and H3K4me3.

Figure S3. 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 in actively transcribed regions.

Similar to Figure 4B except enrichment scores were calculated considering only actively transcribed regions for both actual intersections and random shuffling within the same chromosome. Actively transcribed regions are defined as those being bound by both Pol II and H3K4me3.

Figure S4. Enrichment of exosome, insulator proteins, and RNA pol II at divergently transcribed TSSs.

Phred scores for *p*-values of Fisher's Exact tests for enrichment of the indicated factor at divergent genes separated by 100, 250, or 500 bp in S2 cells. Negative enrichment is indicated by negative values. Significant positive or negative enrichment ($p < 0.05$) corresponds to above or below dashed lines respectively. Overlapping symbols were shifted horizontally for improved clarity.

Figure S5. Expression levels of divergent versus non-divergent genes.

Cumulative histogram of frequency of total genes (y-axis) corresponding to a particular threshold of DESeq scaled gene expression (x-axis) in S2 cells.

Figure S6. Western blotting to confirm knockdown of cells transfected with exosome dsRNA.

S2 or S3 cells transfected with indicated dsRNA were blotted for Rrp6, Lamin, and Rrp40.

Figure S7. Histograms of distribution of RNA-seq fold changes observed in upstream TSS regions of exosome depleted cells.

Genome-wide distribution of RNA fold change in 500 bp regions upstream of exosome bound (top) compared to unbound (bottom) genes for rRNA depleted (A) *Rrp6* or (B) *Rrp40* S3 knockdown libraries compared to controls are shown. Number of sites examined (N), mean and

median $\log_2$ fold changes for each set, and Mann-Whitney *p* values are reported (top) for significance of difference between bound and unbound data sets for a particular knockdown. The value 0.0001 was added to mean scaled read counts of both control and knockdown libraries before ratios were calculated to avoid reporting otherwise infinite values. These appear as extreme outliers on either side of the normal distribution. A total of 9932 regions were examined in this analysis, 3532 and 4079 of which had a total of zero reads in all Rrp6 and Rrp40 depleted versus control libraries respectively and are not shown.

Figure S8. Screenshots of snoRNA precursor accumulation in control versus exosome depleted S3 cells.

S3 rRNA depleted RNA-seq profiles of Rrp6 or Rrp40 knockdown cells are shown (top). S3 ChIP-seq profiles of Rrp6, Rrp40, and input (bottom) are also shown for (A) *snoRNA:Me28S-A1322* and (B) *snoRNA:Me28S-A774b* including surrounding regions of their respective host genes. RNA-seq and ChIP-seq scales are in reads per million unique mapped reads. The bottom of each scale bar indicates 0.

Figure S9. Scatterplot of transcript expression levels and fold change in control versus exosome depleted S3 cells.

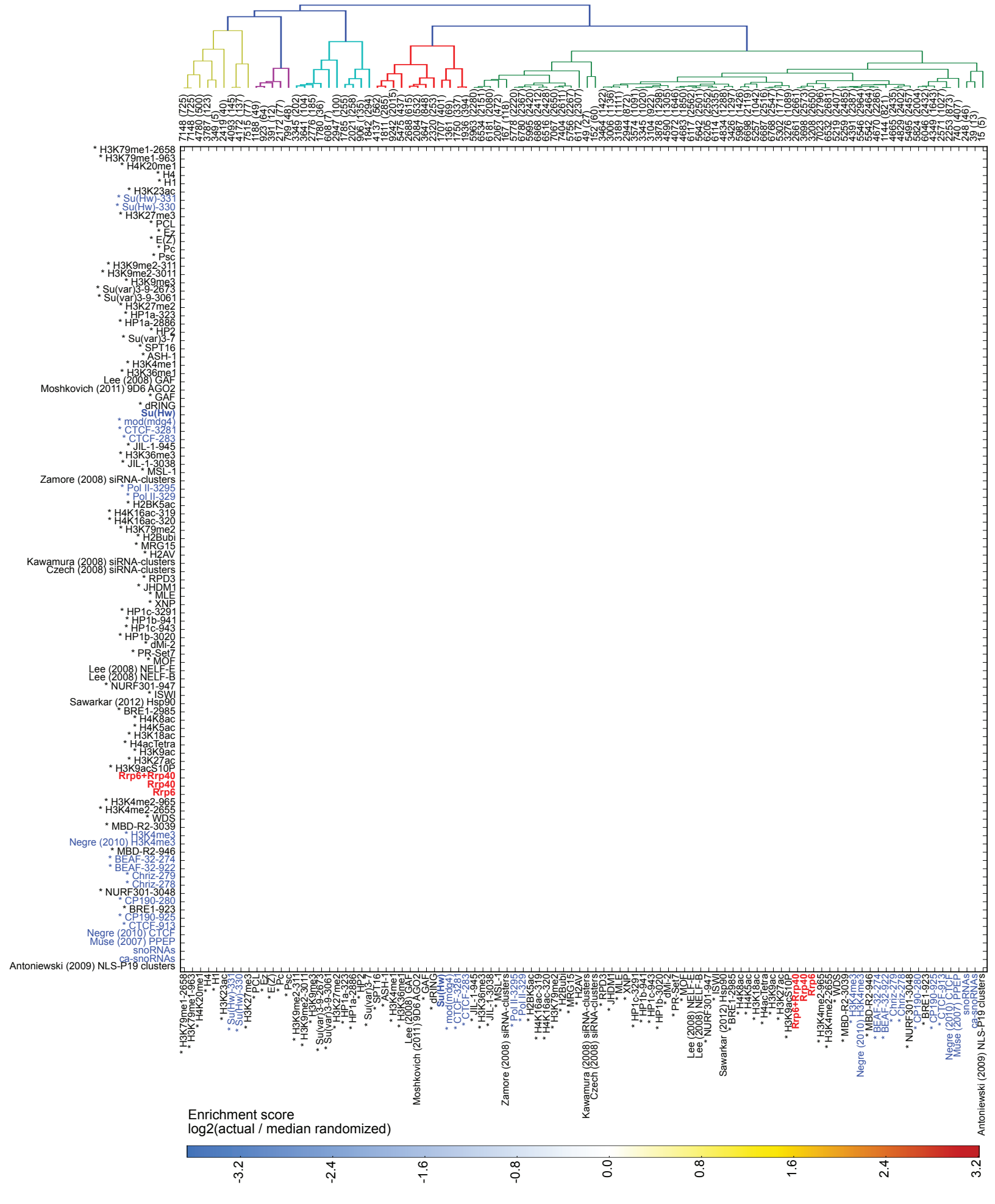
All transcripts (blue), with an exosome ChIP peak (red), significantly upregulated by knockdown (black ring) or significantly downregulated (green ring) are indicated. Corresponding histograms for mean expression in GFP knockdown samples (x-axis) and mean fold change in Rrp6

knockdown samples relative to GFP knockdown control (y-axis) are shown for (A) S2 and (B) S3 cells.

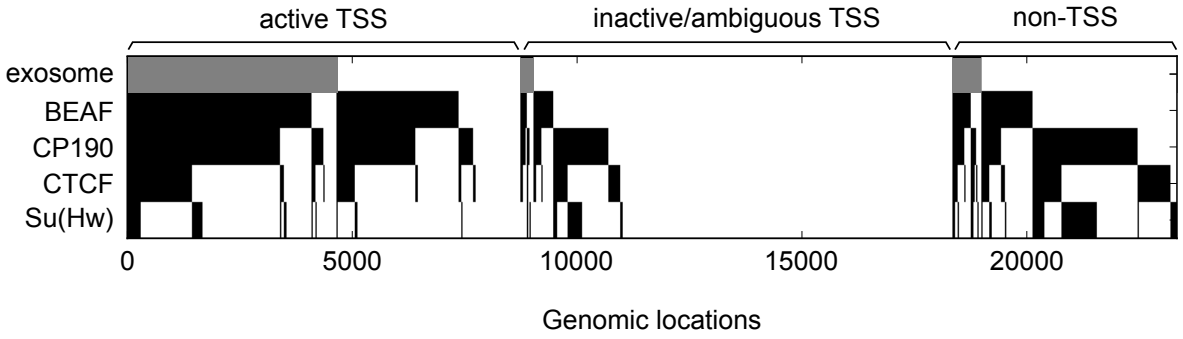
Figure S10. Screenshot of aberrant transcription at the *Abd-B* locus in Rrp6 and Rrp40 depleted S3 cells.

ChIP-chip profiles of CTCF-N and CP190 in embryos (Negre et al. 2010) and S3 ChIP-seq profiles of Rrp6, Rrp40, and input are shown for the *Abd-B* locus (top). S3 RNA-seq profiles of Rrp6 or Rrp40 depleted cells, either oligo-dT selected or depleted for rRNA are shown (middle). RNA-seq data from embryos at indicated stages from modENCODE are also shown (bottom). Coding sequences, promoters, and *cis*-regulatory regions are shown. ChIP–chip data are expressed as log2 (IP/input). RNA-seq and ChIP-seq scales are in reads per million unique mapped reads. The bottom of each scale bar indicates 0.

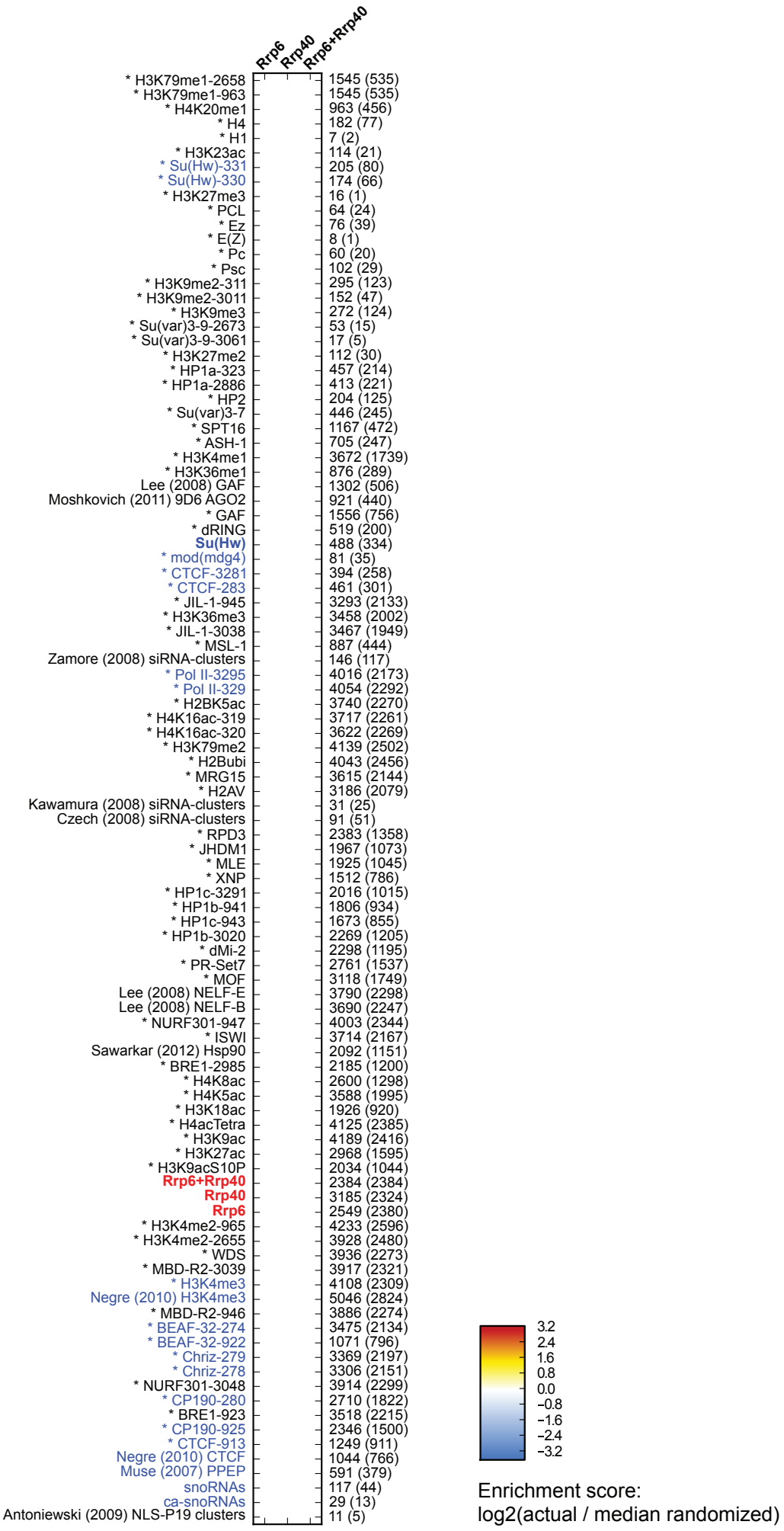
Lim_Supplementary_Fig_S1



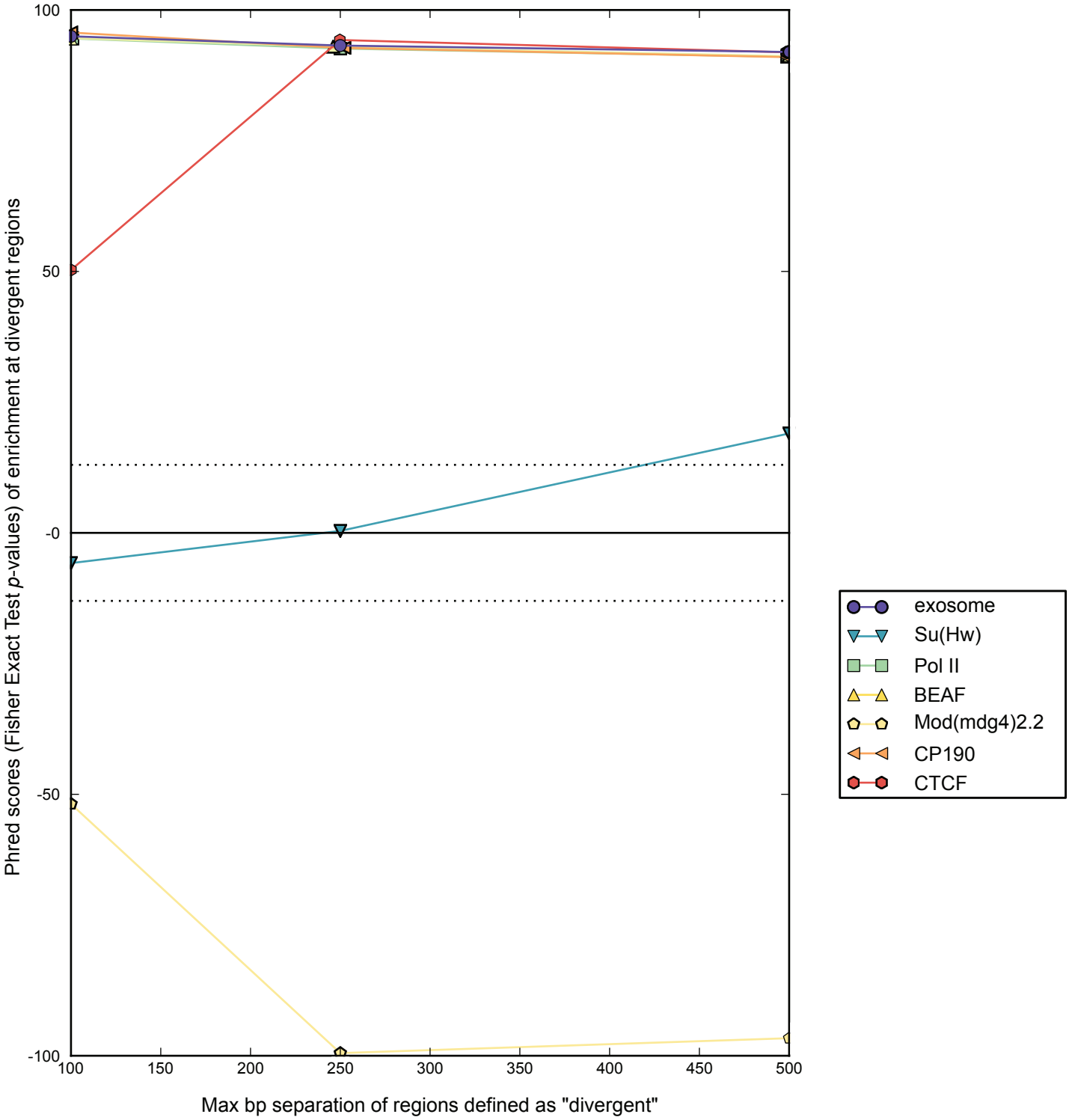
Lim_Supplementary_Fig_S2



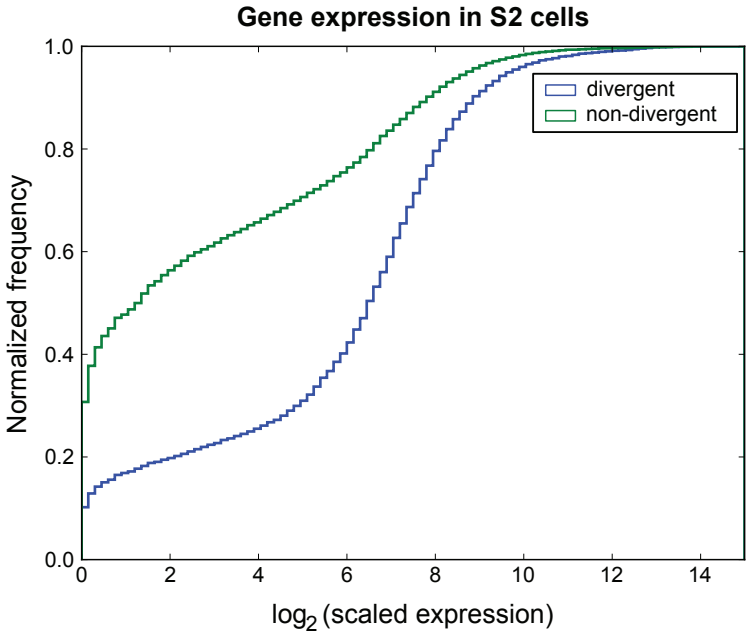
Lim_Supplementary_Fig_S3



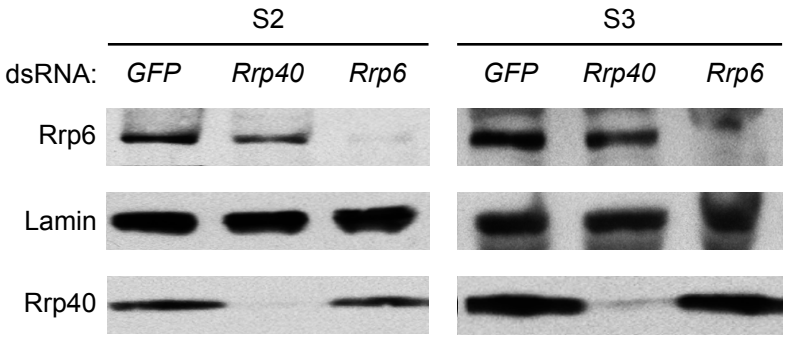
Lim_Supplementary_Fig_S4



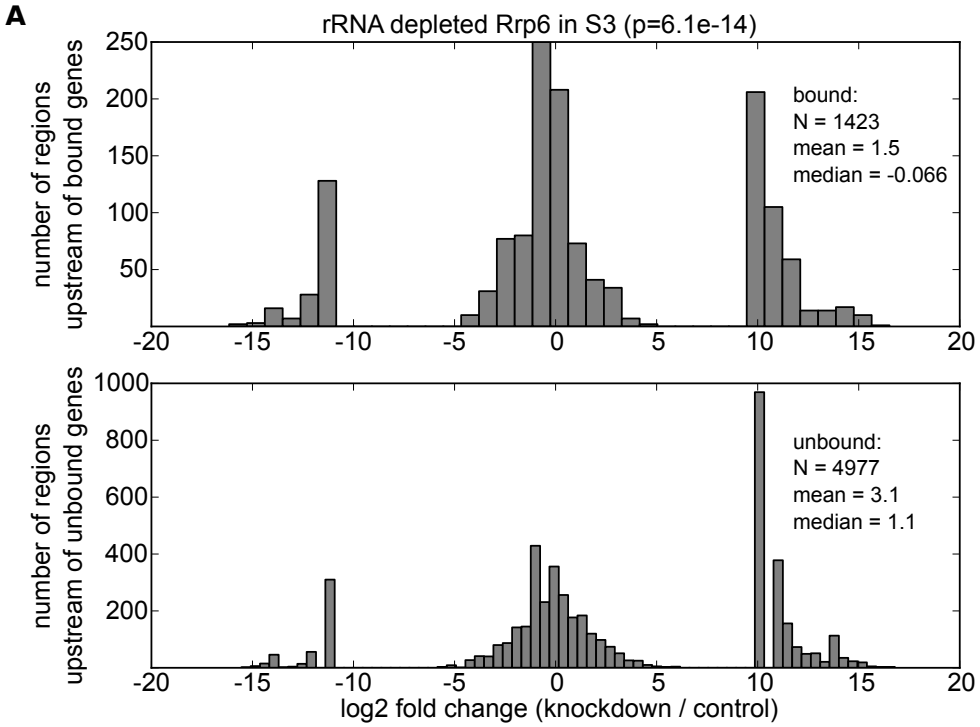
Lim_Supplementary_Fig_S5

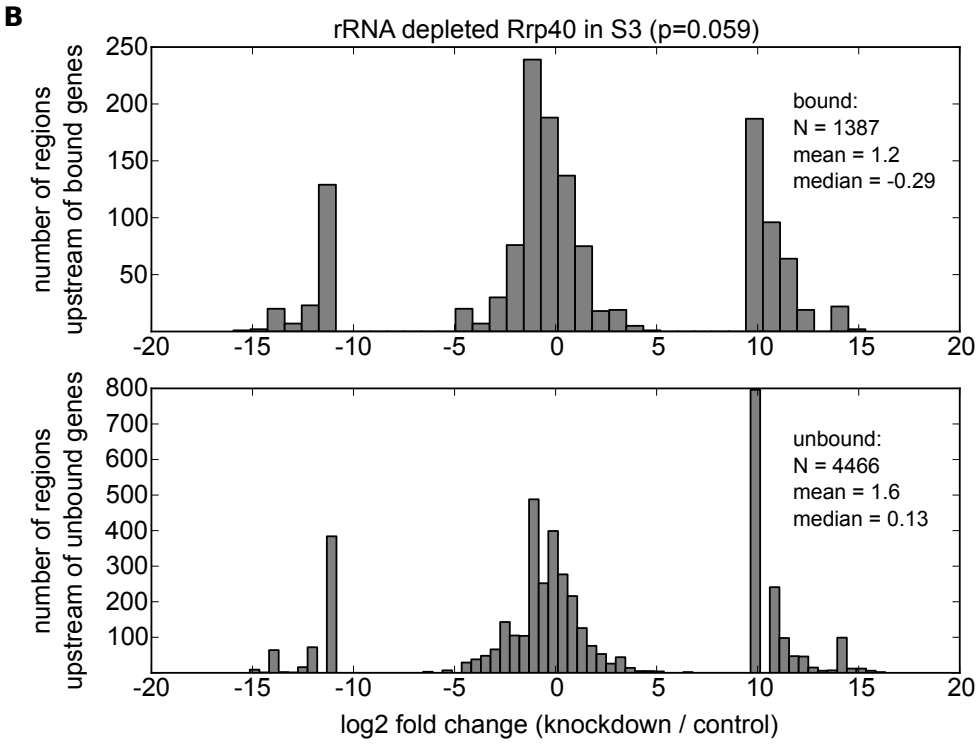


Lim_Supplementary_Fig_S6

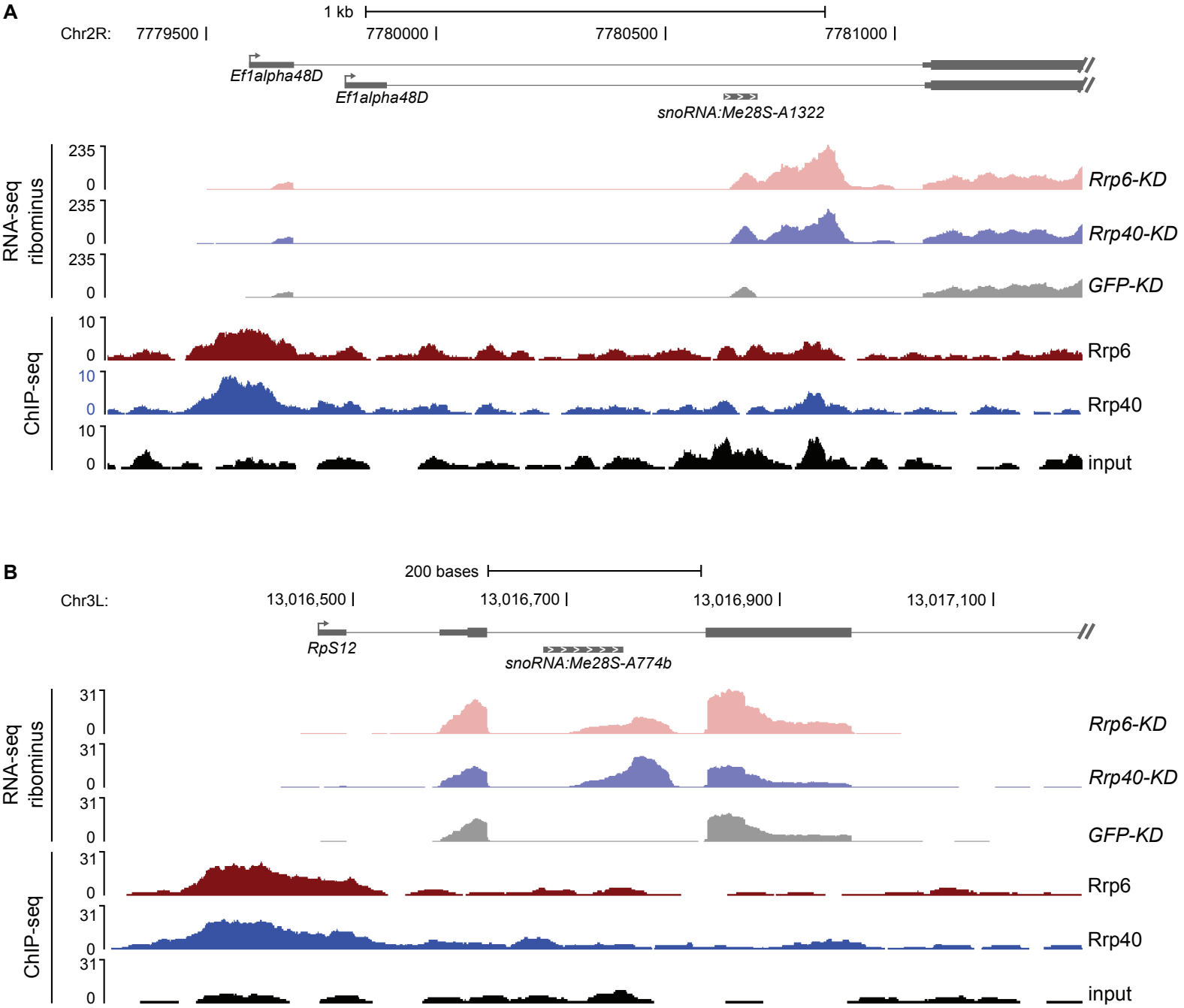


Lim_Supplementary_Fig_S7

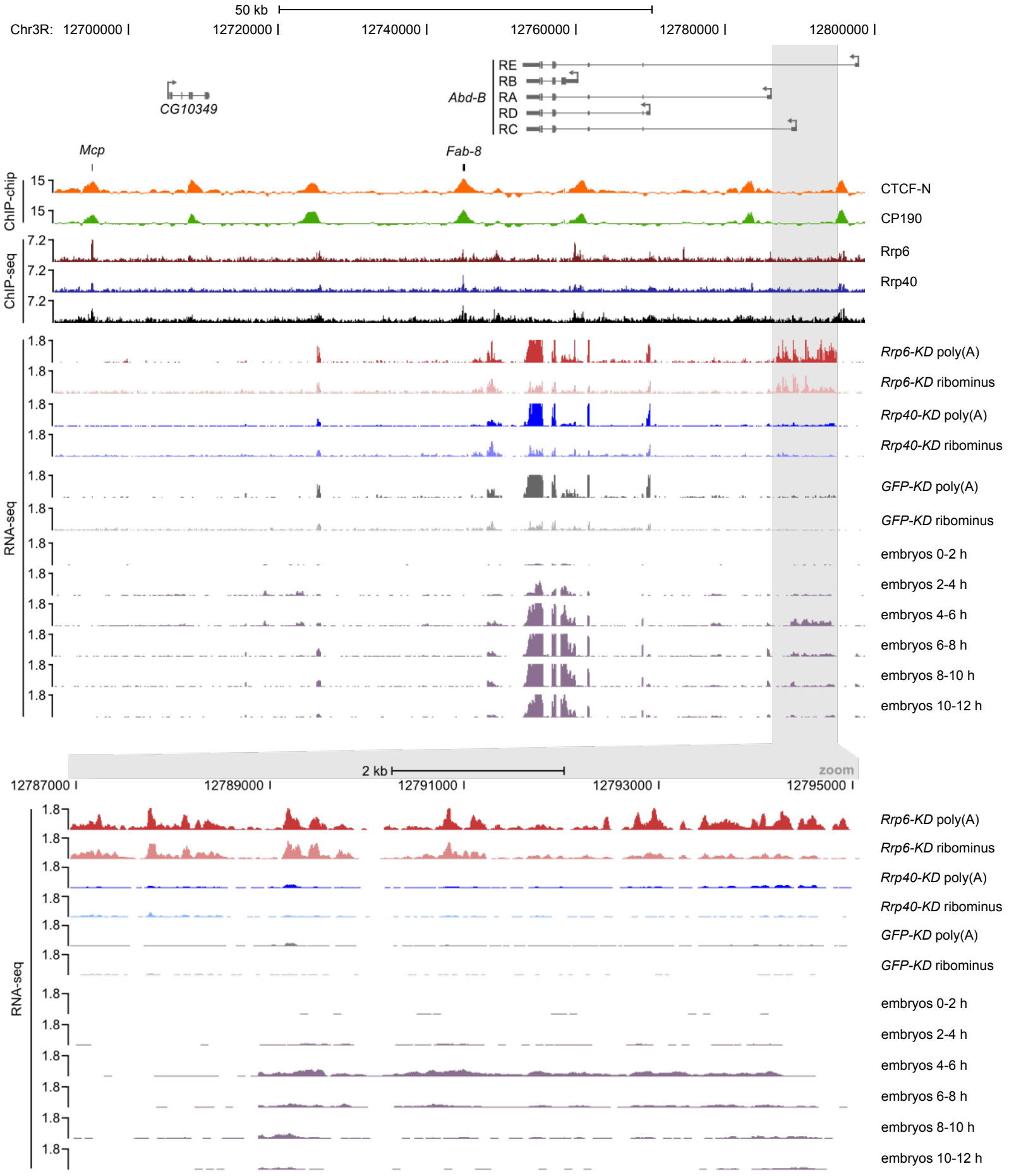




Lim_Supplementary_Fig_S8



Lim_Supplementary_Fig_S10



Feature set	% sites overlapping exosome	% sites overlapping Rrp6	% sites overlapping Rrp40	% sites overlapping BEAF-32	% sites overlapping CP190	% sites overlapping CTCF	% sites overlapping active TSS
exosome	100.0	100.0	100.0	83.5	78.1	40.6	84.3
Rrp6	81.8	100.0	81.8	73.2	70.7	36.6	73.4
Rrp40	69.5	69.5	100.0	80.2	75.3	37.7	82.0
BEAF-32	48.9	52.1	67.4	100.0	66.3	32.0	68.2
CP190	33.7	37.3	47.0	50.1	100.0	35.1	46.0
CTCF	34.8	37.8	45.4	47.0	68.5	100.0	39.2
active TSS	50.4	52.8	69.9	71.7	62.9	25.4	100.0

Table S1. Percent overlap of exosome and insulator protein chromatin binding sites and transcriptionally active TSSs in S2 cells.

Ontology classes	Enrichment scores	
	S2	**S3**
mitotic cell cycle	12.5	11.1
nuclear lumen	9.5	7.5
purine nucleotide binding	7.8	9.2
proteolysis	7.6	6.4
mitotic chromatid segregation	6.4	3.9
ribosomal protein	5.5	5.9
DNA repair	4.6	4.2
macromolecular complex assembly	4.3	5.3
protein transport	4.2	4.5
endocytosis	4.1	3.5
zinc finger, PHD-type	3.9	2.0
actin cytoskeleton organization	3.8	4.2
trasncription coactivator activity	3.7	2.5
RNA binding	3.6	1.9
chromatin modification	3.5	3.3
ubiquitin mediated activity	2.6	4.9
cell adhesion	0.7	3.8

Table S2. Summary of gene ontology categories corresponding to exosome TSS binding with a minimum DAVID enrichment score of 3.5 in at least one cell type.

Rrp6-S2-poly(A), up vs unchanged genes

	with peak	without peak	total
up	2	51	53
unchanged	4060	10948	15008
total	4062	10999	15061

p-value(left tail=1.185e-05, right tail=1, two tail=1.685e-05)

Rrp6-S2-poly(A), down vs unchanged genes

	with peak	without peak	total
down	0	4	4
unchanged	4060	10948	15008
total	4060	10952	15012

p-value(left tail=0.2832, right tail=1, two tail=0.5799)

Rrp6-S2-poly(A), changed vs unchanged genes

	with peak	without peak	total
changed	2	55	57
unchanged	4060	10948	15008
total	4062	11003	15065

p-value(left tail=3.876e-06, right tail=1, two tail=6.956e-06)

Rrp6-S3-poly(A), up vs unchanged genes

	with peak	without peak	total
up	31	216	247
unchanged	3656	11121	14777
total	3687	11337	15024

p-value(left tail=1.776e-06, right tail=1, two tail=3.177e-06)

Rrp6-S3-poly(A), down vs unchanged genes

	with peak	without peak	total
down	23	18	41
unchanged	3656	11121	14777
total	3679	11139	14818

p-value(left tail=1, right tail=1.842e-05, two tail=2.656e-05)

Rrp6-S3-poly(A), changed vs unchanged genes

	with peak	without peak	total
changed	54	234	288
unchanged	3656	11121	14777
total	3710	11355	15065

p-value(left tail=0.01012, right tail=0.9932, two tail=0.01881)

Rrp40-S3-poly(A), up vs unchanged genes

	with peak	without peak	total
up	7	198	205
unchanged	3703	11150	14853
total	3710	11348	15058

p-value(left tail=5.657e-17, right tail=1, two tail=1.401e-10)

Rrp40-S3-poly(A), down vs unchanged genes

	with peak	without peak	total
down	0	7	7
unchanged	3703	11150	14853
total	3703	11157	14860

p-value(left tail=0.1344, right tail=1, two tail=0.2042)

Rrp40-S3-poly(A), changed vs unchanged genes

	with peak	without peak	total
changed	7	205	212
unchanged	3703	11150	14853
total	3710	11355	15065

p-value(left tail=9.881e-18, right tail=1, two tail=1.629e-10)

Rrp6-S3-ribominus, up vs unchanged genes

	with peak	without peak	total
up	17	310	327
unchanged	3693	11029	14722
total	3710	11339	15049

p-value(left tail=2.107e-21, right tail=1, two tail=2.871e-10)

Rrp6-S3-ribominus, down vs unchanged genes

	with peak	without peak	total
down	0	16	16
unchanged	3693	11029	14722
total	3693	11045	14738

p-value(left tail=0.009873, right tail=1, two tail=0.01743)

Rrp6-S3-ribominus, changed vs unchanged genes

	with peak	without peak	total
changed	17	326	343

unchanged	3693	11029	14722
total	3710	11355	15065

p-value(left tail=5.192e-23, right tail=1, two tail=3.23e-10)

Rrp40-S3-ribominus, up vs unchanged genes

	with peak	without peak	total
up	4	18	22
unchanged	3706	11335	15041
total	3710	11353	15063

p-value(left tail=0.3378, right tail=0.828, two tail=0.624)

Rrp40-S3-ribominus, down vs unchanged genes

	with peak	without peak	total
down	0	2	2
unchanged	3706	11335	15041
total	3706	11337	15043

p-value(left tail=0.568, right tail=1, two tail=1)

Rrp40-S3-ribominus, changed vs unchanged genes

	with peak	without peak	total
changed	4	20	24
unchanged	3706	11335	15041
total	3710	11355	15065

p-value(left tail=0.2598, right tail=0.8771, two tail=0.4802)

Table S8. Fisher exact tests comparing exosome ChIP peaks versus genes changed in expression in control versus Rrp6 or Rrp40 knockdown cells.

Primers	Sequence
scs'	TAGTTGTCGAGTGCCTGTGC
	TGCACTGCAGCTGATTTTTC
scs	GCCGACGCAACTTATTTTTC
	CAGCAATAGCCCCACTTTTC
Mcp	TCATGTGTTAGTGCGTGAGAG
	CAATGTTGCCATCTTTGTCG
Fab-8	CATCTTCCGTTCATCGTTTC
	TGTTGGTGAGCAAGCGAAGA
polo pro	TCACGTCCAATTCATTTTGG
	GGAAACGGAGATCAGATCCA
CycB pro	TCGAGTTTTTGCACACGAAG
	CGTCAAACTCATCGGCATTT
bel pro	GCTTGGAAAATGCTGCTTCT
	ACCGCATCTATCGATTCCTC
RpL32	CTGCATGAGCAGGA
	ATGACCATCCGCCC
BEAF-32 KD	TAATACGACTCACTATAGGGAGAAGTGGCAAAACCGTGCGGATTACC
	TAATACGACTCACTATAGGGAGAAGTGGCCAAGCTAAGTGCAAAATAGT
CP190 KD	TAATACGACTCACTATAGGGAGAGTCTGCTCTGGTTCCTGCTC
	TAATACGACTCACTATAGGGAGAGTAAACGGACGACCCATTAGCATTC
CTCF KD	TAATACGACTCACTATAGGGAGAGAGCACTTGAAGGATGGCT
	TAATACGACTCACTATAGGGAGAGAGCCCGACATCAGTTCAAT
GFP KD	TAATACGACTCACTATAGGGAGAAGTGGGGATCCATGGTGAGCAAG
	TAATACGACTCACTATAGGGAGAAGTGGCTGCAGTTATTACTTGTACAG
Rrp6 KD	TAATACGACTCACTATAGGGAGAGCTGCCAAAGCCTTGAATATG
	TAATACGACTCACTATAGGGAGAGCATCTCCCTTGGAAGACTC
Rrp40 KD	TAATACGACTCACTATAGGGAGAAGTGGCCAGCCTCCATATCGTATCTC
	TAATACGACTCACTATAGGGAGAAGTGGCCGAGTTGACGCAGACCA
su(Hw)	TAATACGACTCACTATAGGGAGAAAACAACAACGGTCATCAATG
	TCG CAG GGT ATC GCT GCT GTCAAAGTTCTTATCGC

Table S9. Primers used. Sequences are listed 5' to 3'.