Supplementary Experimental Procedures

S1-DRIP detailed protocol

$3x10^{10}$ cells of exponentially growing cells (in 1.2 L of YEP+ 2% glucose) were pelleted at 5500 rpm in a Sorvall RC-5B centrifuge, and pellets were either frozen at -80ºC or used directly for genomic DNA extraction. DNA extractions were done with Qiagen tip 500/G following standard protocol, with three exceptions: 30ul of 100T 20mg/ml zymolyase was used for spheroplasting, RNase A was not added to the G2 buffer, and only 200ul of Proteinase K was used in the deproteinization step. DNA is resuspended in 1.5ml nuclease-free TE buffer and nutated overnight at 4ºC. $3x10^{10}$ cells should yield ~200- 300ug of DNA, all of which is used for an IP. Note that despite the omission of RNase A treatment, no RNA contamination is detectable in the genomic preps as assayed by ethidium gel and 260/280 ratio. For the S1 nuclease treatment, S1 nuclease is diluted fresh 1000x in dilution buffer (to 1U/ul), and 2.5ml reaction set up with final 1X S1 Buffer, 0.3M NaCl, and 300U S1 nuclease (Thermo Fisher, Waltham, MA). Reactions were conducted at 30ºC for 50 minutes, and stopped with an equal volume of TE-50 (0.2M Tris pH 8.0, 50mM EDTA pH 8.0), and 670ul of 3M NaCl. Following ethanol precipitation DNA was resuspended in 500ul TE, and sonicated with Covaris™ (9 cycles with settings at duty cycle: 20%, intensity: 10, cycles/burst: 200, and 30s rest in between). For the IP, 35ug of S9.6 antibody was pre-bound to 160ul of magnetic Protein A Dynabeads (Thermo Fisher) for 1 hr at 4ºC. Sonicated DNA in final 1X FA (1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS, 50 mM HEPES, 150 mM NaCl, 1 mM EDTA) was incubated with the S9.6—protein A magnetic beads for 2 hrs at 4ºC. Beads were then washed and the DNA eluted according to standard ChIP protocols: 5 successive washes with 1X FA, 1X FA with 0.5M NaCl, Lithium chloride detergent (0.25M LiCl, 0.5% NP-40, 0.5% Sodium deoxycholate, 1mM EDTA, and 10mM Tris-HCl), and finally twice with TE buffer. DNA was then eluted from the beads with 300ul of Elution buffer (1% SDS, 0.1M $NAHCO_3$). To ensure antibody is fully removed prior to library prep, samples are treated with 1ul proteinase K (20mg/ml) for 1 hour at 42° while shaking at 1000 rpm, extracted with phenol:chloroform, and ethanol precipitated. Final DNA pellet was resuspended in 50ul TE buffer. For RNase H treated controls, S1 nuclease-treated gDNA was treated with 50 units of RNase H (NEB, Ipswich, MA) overnight at 37ºC and re-precipitated prior to sonication and IP.

Spike-in

Components of the spike-in (listed in Supplemental Table 8) were mixed at equimolar ratio in annealing buffer (50mM NaCl, 10mM Tris ph8, 1mM EDTA). The oligos were annealed by incubating the mix at 95° for 2 minutes and slowly decreasing to room temperature (~ 2 hours). $1.2x10^9$ spike-in molecules were used in each S1-DRIP-seq reaction.

Wild-type specific peaks. The wild-type only hybrid-prone features listed in Supplemental Table 4 were identified by cross referencing the hybrid-prone features common to at least three of the four wild-type replicates with the hybrid-prone features identified in each individual *rnh1*Δ*rnh201*Δ replicate using the intersectBed tool. Hybrid-prone features only found in wild-type samples, along with coordinates of the hybrid-

prone region in the feature are reported. If multiple hybrid-prone regions existed within one feature, coordinates of only one region was reported.

Supplementary Figure Legends

Figure S1.1. Development of S1-DRIP.

(A) Representative α-DNA:RNA hybrid dot blot of gDNA from wild-type and *rnh1*Δ*rnh201*Δ at 30s and 2min of exposure.
(B) Quantitation of dot blot in Fig. 1B. Signals at each spot were quantified with ImageJ, and represented as % recovered relative to the untreated control.
(C) Effect of S1 amount on DNA:RNA hybrids. Excess use of S1 nuclease (10U/ug) led to destruction of R-loops in gDNA prior to sonication.

Figure S1.2. Spike-in sequencing.

(A) Expected spike-in sequencing results based on library prep protocol. During End repair, the combined action of Klenow and T4 DNA polymerase allows for strand displacement/filling in of the RNA region in the coding strand. This results in a full length dsDNA molecule for which reads can be obtained from both the coding and template strands.
(B) Schematic of the spike-in and sequencing results obtained. Reads mapped to the spike-in were distributed 40%-60% to the coding and template strands, respectively. The start sites of the reads, and the proportion starting at the 5'-most ends of the coding and template strands (the expected start site) are indicated.
(C) Models explaining the spike-in reads mapping to unexpected start site. Approximately 34% of total reads mapped to unexpected start sites. Potential explanations are: breakage at the site of DNA:RNA hybrid resulting in two separate dsDNA molecules for sequencing (left panel), and variability of 5' end start site as a result of incomplete oligonucleotide synthesis. Expected percent of full-length products was obtained from IDT website.

Figure S1.3. DRIP results and sample correlations.

(A) Average read counts recovered from various samples.
(B) Spearman correlation coefficients of the 4 wild-type, 4 *rnh1*Δ*rnh201*Δ IP samples, and 2 totals (input DNA).

Figure S1.4. Expected sequencing results of genomic DNA:RNA hybrids.

(A) Expected results based on library prep protocol, when hybrid size is smaller than fragmentation size (<200-300 bp), indicated by grey dotted lines. Fragments resulting from DRIP of sonicated DNA, and individual reads mapping to the plus (pink) and minus (blue) strands are depicted. For smaller hybrids, all fragments recovered after sonication should contain a hybrid and hence be recognized by the S9.6 antibody, as well as a dsDNA region that can be extended. Note that based on the specificity of the antibody (Hu et al., 2006) fragments with longer hybrids would be preferentially bound.
(B) Expected results when hybrid size is larger than fragmentation size (>200-300 bp), indicated by grey dotted lines. If hybrid region was greater than fragmentation size, then a subset of fragments IP'ed are expected to not contain a dsDNA region and hence will be poor substrates for library construction. This would lead to a different distribution of

sequenced tags than in (A), with a large internal region of the hybrid-prone region with poor coverage.

Figure S1.5. S1-DRIP-seq results.
(A) Comparison of S1-DRIP-seq, ChIP-seq, and DRIP-chip. Chromosomal snapshots of Chr X (top) and Chr III (bottom) of normalized coverage tracks from S1-DRIP-seq and ChIP-seq (El Hage et al., 2014), and BED tracks of called hybrid-prone regions from Chan et al.'s DRIP-chip study (2014).
(B) Quantitation of hybrid-prone regions in S1-DRIP-seq. For all 781 hybrid prone regions mapped in *rnh1*Δ*rnh201*Δ*,* the read counts from wild-type, *rnh1*Δ*rnh201*Δ*,* and RNase H-treated *rnh1*Δ*rnh201*Δ normalized to spike-in are shown. Hybrid prone regions are ranked along the x-axis by read counts in the *rnh1*Δ*rnh201*Δ in descending order.

Figure S1.6. S1-DRIP-qPCR validation.
(A) S1-DRIP-qPCR of *RPL15A* in wild-type, *rnh1*Δ*rnh201*Δ, and RNase-H treated *rnh1*Δ*rnh201*Δ*.* X-axis coordinates are plotted relative to *RPL15A* ORF coordinates.
(B) S1-DRIP-qPCR of hybrid-prone regions with strong, mid- and low level peaks, and control regions in wild-type and *rnh1*Δ*rnh201*Δ. To the right of the graph, representative snapshots of strong, mid-, low and no peaks from *rnh1*Δ*rnh201*Δ are shown.

Figure S2.1. S1-DRIP-seq peaks distribution among features.

Figure S2.2. Snapshots of hybrid-enriched loci.
(A) Representaive snapshots from *rnh1*Δ*rnh201*Δ and wild-type show relative R-loop levels in the two strains. The loci are from left to right, top to bottom: rDNA (RDN37-1, chr XII 450k-460k), Ty element (YDRCTy1-1, chr IV 645k-652k), telomere (TEL01L-TR, chr I 0-400bp), snoRNA (SNR128 and SNR190, chr X 139k-140.5k), tRNA (tA(AGC)F, chr VI 204.5k-205.5k), and protein-coding ORF (TPI1, chr IV 555k-557k).
(B) Blow up of the RDN37-1 locus from *rnh1*Δ*rnh201*Δ. Dotted lines indicate the borders of the spacer regions that are in the primary transcript, but are rapidly eliminated for the production of the three final mature transcripts.

Figure S2.3. S1-DRIP-qPCR of DNA:RNA hybrids at Ty1 elements.
(A) qPCR of all Ty1 elements in wild-type, *rnh1*Δ*rnh201*Δ, and RNase-H treated *rnh1*Δ*rnh201*Δ*.* X-axis coordinates are plotted relative to TSS of Ty1 elements (bp).
(B) qPCR to detect hybrids only at genomic Ty repeats and not in Ty1 cDNA. Specific Ty1 elements in wild-type and *rnh1*Δ*rnh201*Δ using qPCR primer sets spanning the 5' (left panel) and 3' (right panel) junctions of unique DNA and Ty1 elements.

Figure S2.4. RE-DRIP-qPCR of DNA:RNA hybrids at Ty1 elements.
(A) qPCR of all Ty1 elements in wild-type and *rnh1*Δ*rnh201*Δ. X-axis coordinates are plotted relative to TSS of Ty1 elements, and restriction sites are indicated below Ty1 schematic.
(B) qPCR of specific Ty1 elements using qPCR primer sets spanning the 5' (left panel) and 3' (right panel) junctions of unique DNA and Ty1 elements.

Figure S2.5. S1-DRIP-qPCR of rDNA locus. qPCR of the rDNA locus in wild-type, *rnh1*Δ*rnh201*Δ, and RNase-H treated *rnh1*Δ*rnh201*Δ. X-axis coordinates are plotted relative to 35S rRNA TSS.

Figure S2.6 Distribution of hybrids at tRNAs and sn/snoRNAs. tRNAs and sn/snoRNAs are aligned from TSS to TES and plotted ±1 kb. Metagene plots display the median read counts for each feature from an *rnh1*Δ*rnh201*Δ DRIP library (top) as well as control *rnh1*Δ*rnh201*Δ input DNA library (bottom). The heatmaps display the signals obtained from individual features, sorted according to total signal strength.

Figure S2.7. Mitochondrial S1-DRIP-seq. Snapshot of mitochondrial genome showing hybrid formation pattern in wild-type, *rnh1*Δ*rnh201*Δ and the corresponding RNase-H treated samples.

Figure S3. RNA induction in galactose. The relative change in RNA levels for GAL7 and SMC3 after Induction of expression with galactose was measured using quantitative reverse-transcriptase PCR.

Figure S4.1. Hybrids at ORFs.
(A) Metagene plot and heatmap of signals at hybrid-prone ORFs from control *rnh1*Δ*rnh201*Δ input DNA.
(B) Size distribution of the ORF-associated hybrid-prone regions.
(C) Overlap distribution of ORF-associated hybrid-prone regions.
(D) Comparison of the size of ORFs with the % of ORF hybrid-prone. For hybrid-prone regions largely overlapping ORFs (80-100%), ORFs size tend to smaller, while for hybrid-prone regions that only minorly overlap ORFs (1-20%) ORFs size tend to larger.

Figure S4.2. Expression-binned hybrid signal. Metagene plot and heatmap of signals at hybrid-prone ORFs ±1 kb from *rnh1*Δ*rnh201*Δ, binned by expression. Expression categories are: 1, 2, 3-4, 5-8, 9-12, 13-16 and 17-20.

Figure S5.1. Analysis of sequence features of hybrid-prone ORFs.
(A) GC content of hybrid-prone and non-hybrid prone regions. Only hybrid-prone genes in the top 5% of expression are significantly more GC-rich than non-hybrid prone genes. P-values were calculated with Wilcoxon rank-sum test, and significant p-values indicated.
(B) GC and AT skew of telomeric hybrids. Both GC [(G-C)/(G+C)] and AT [(A-T)/(A+T)] skew were high at telomeric hybrid regions.
(C) GC skew of hybrid-prone regions. No significant GC skew was observed at hybrid-prone regions in the various expression categories.

Figure S5.2. Inverse correlation of positive AT skew and expression level at yeast ORFs. Metagene plots and heatmaps of ORFs ±1 kb of AT skew binned by expression level pointed to the increased positive AT skew across ORFs at lower expressed ORFs. AT skew was calculated over a 50bp sliding window.

Figure S5.3. AT skew at hybrid-prone ORFs. The AT skew over hybrid-prone ORFs is more positive than over non-hybrid prone ORFs in low expression categories (Bins 9-20). P-values were calculated with Wilcoxon rank-sum test, with one asterisks indicating $p=0.05$ and two asterisks $p< 0.005$.

Figure S5.4. AT skew of hybrid and non-hybrid regions. Within hybrid-prone ORFs, the regions with hybrids had significantly more positive AT skew than non-hybrid regions in expression bins 5-20. P-values were calculated with Wilcoxon rank-sum test.

Figure S5.5. PolyA tracts and hybrid formation.
(A) Fraction of polyA tracts overlapping hybrid regions. Imperfect polyA tracts also correlated with hybrid formation but at longer lengths: 25 bp for 1 mismatch, and 27 bp for 2 mismatches.
(B) The absolute number of polyA tracts used for the analysis in S5.5A.
(C) AT skew of hybrid regions with and without polyA tracts. Hybrid-prone regions that contained a polyA tract ≥ 21 bp had significantly lower positive AT skew than hybrid-prone regions without long polyA tracts. P-values were calculated with Wilcoxon rank-sum test, with one asterisks indicating $p<0.015$ and two asterisks $p< 0.01$.

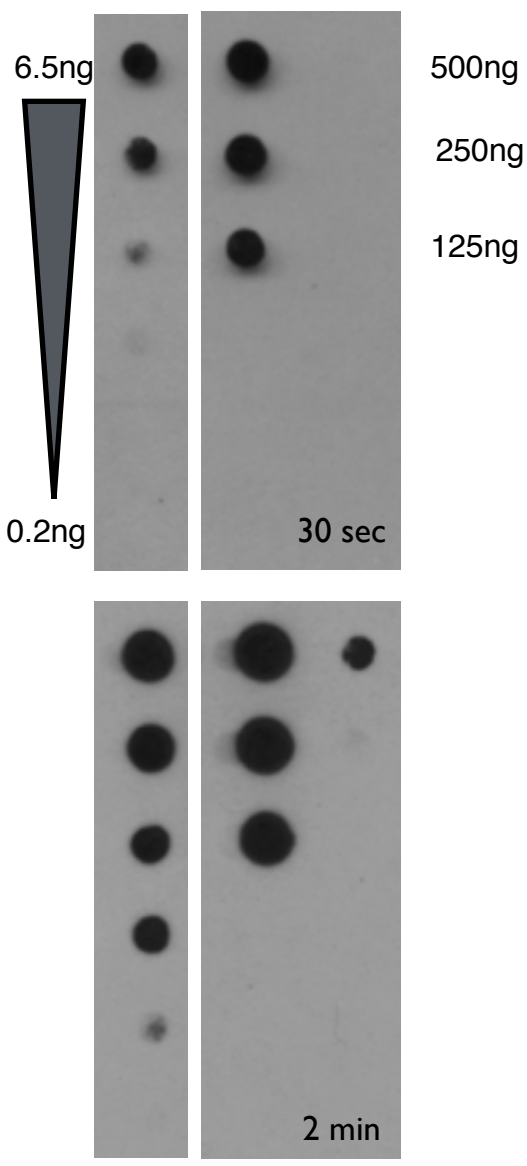
Figure S6.1. PolyA-associated hybrid regions.
(A) Snapshot of representative ORF with a polyA and hybrid prone region from *rnh1*Δ *rnh201*Δ. The site of the polyA, and distribution of uniquely mapped reads in the region are shown. Based on the reads distribution, the predicted site of the hybrid is indicated, with the darker color corresponding to the strongest hybrid formation.
(B) PolyA-associated hybrid regions are similar in size distribution to non-polyA hybrids.

Figure S6.2. Effect of PolyA deletions on hybrid formation.
(A) Outline of the deletion of polyA. PolyA tracts are deleted using PCR-based seamless gene editing technique (Horecka and Davis, 2014) in both the wild-type and *rnh1*Δ*rnh201*Δ background. The cassette used for knockout of polyA contains the 50 bp upstream of a polyA followed by the 50 bp downstream of the polyA, the URA3 selectable marker, and finally a second copy of the 50 bp downstream of the polyA. Following selection of integrants, recombination of the two copies of the 50 bp downstream of the polyA pops out the marker and leaves behind a scar-free deletion (pAΔ). All deletions were validated with Sanger sequencing.
(B) % hybrid signal in polyA (solid bars) and pAΔ (hatched bars) in wild-type (light green) and *rnh1*Δ*rnh201*Δ (dark green) strains. Deletion of polyAs led to significant decrease in hybrid signal in both wt and ΔΔ in 3 of 4 regions tested.
(C) Effect of pAΔ on mRNA expression levels. No significant decrease in RNA levels were detected in either wt (light green) or *rnh1*Δ*rnh201*Δ (dark green). RNA levels were quantified with quantitative reverse-transcriptase PCR. Data presented is the average of two experiments, and error bars represent their standard deviation.

Supplemental Sheet S1. Coordinates of hybrid-prone regions in *rnh1*Δ*rnh201*Δ.
Supplemental Sheet S2. Coordinates of hybrid-prone regions in wild-type.
Supplemental Sheet S3. Features overlapped with hybrid-prone regions in *rnh1*Δ*rnh201*Δ.

Supplemental Sheet S4. Features with an overlapped hybrid-prone region specific to wild-type.
Supplemental Sheet S5. ORFs in expression bin 1 and 2 with hybrid-prone regions identified in 2 of 4 *rnh1*△*rnh201*△ replicates.
Supplemental Sheet S6. MEME analysis of hybrid-prone regions.
Supplemental Sheet S7. Perfect polyA tracts ≥21 bp, overlapping hybrid-prone region and features associated with the hybrid-prone region.
Supplemental Sheet S8. Primers and strains used in this study.

A)



B)

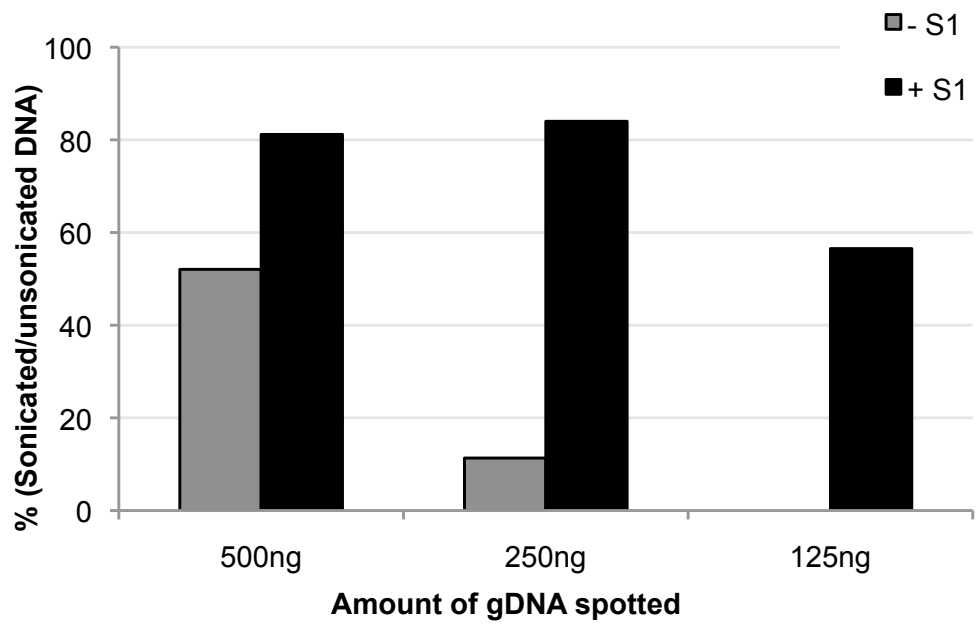
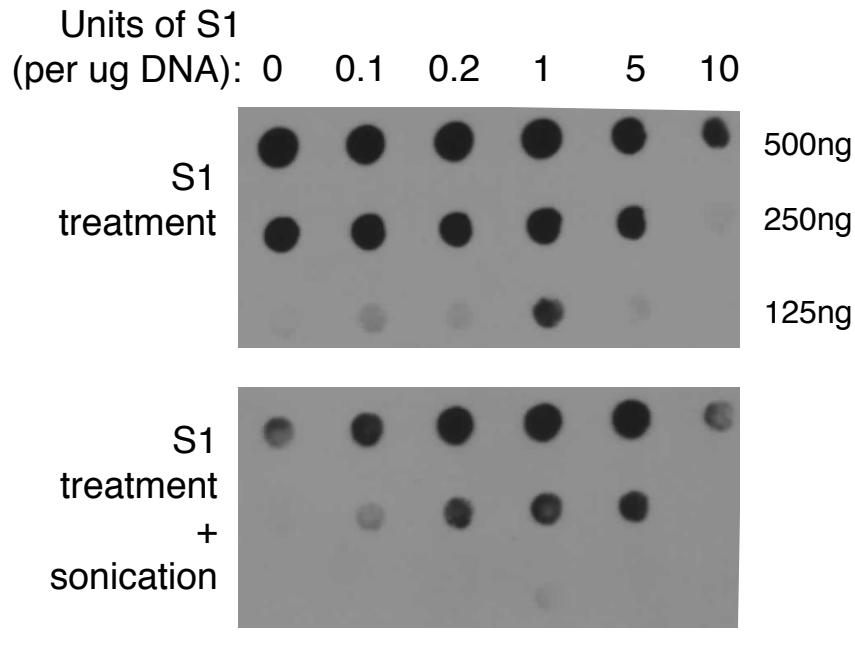
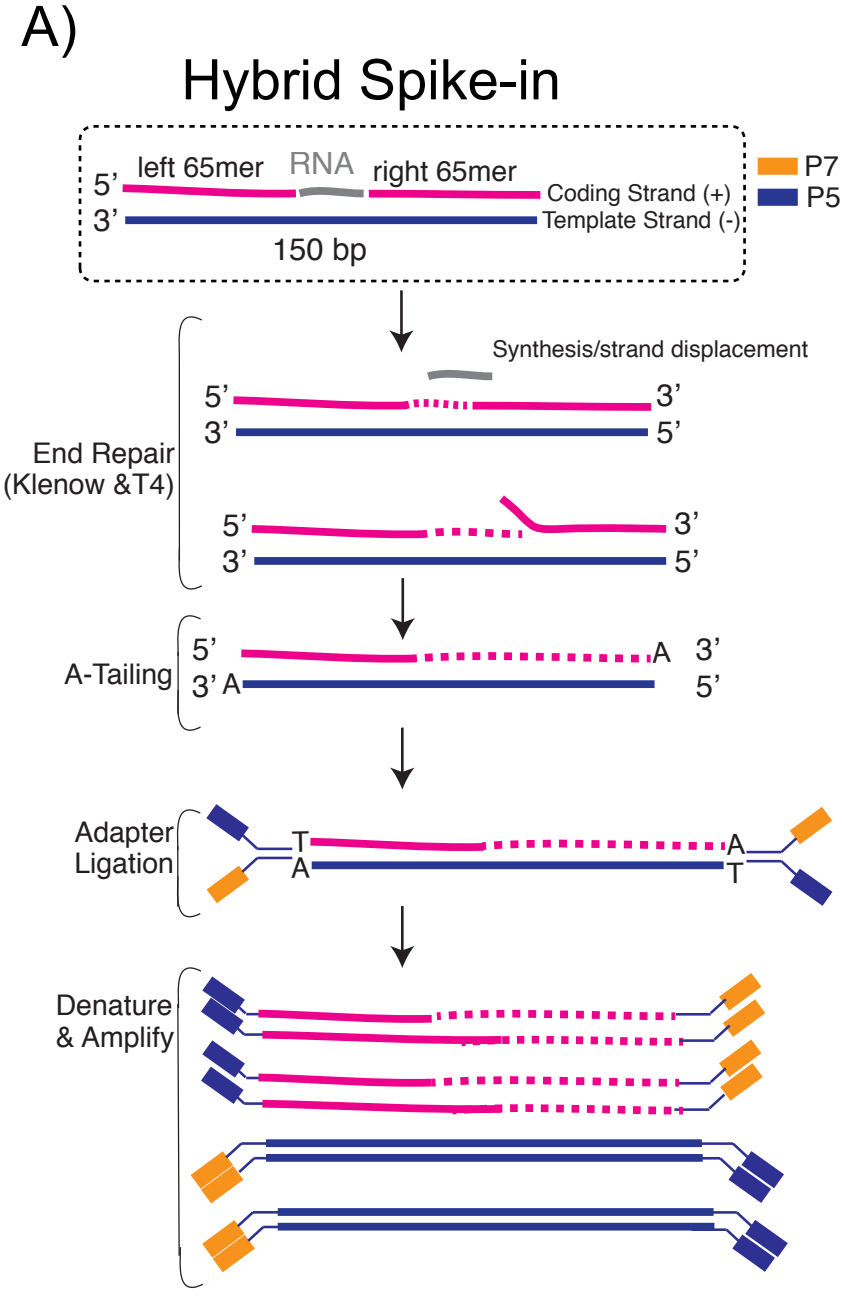


Figure S1.1
Dot Blot and S1 Dot blot

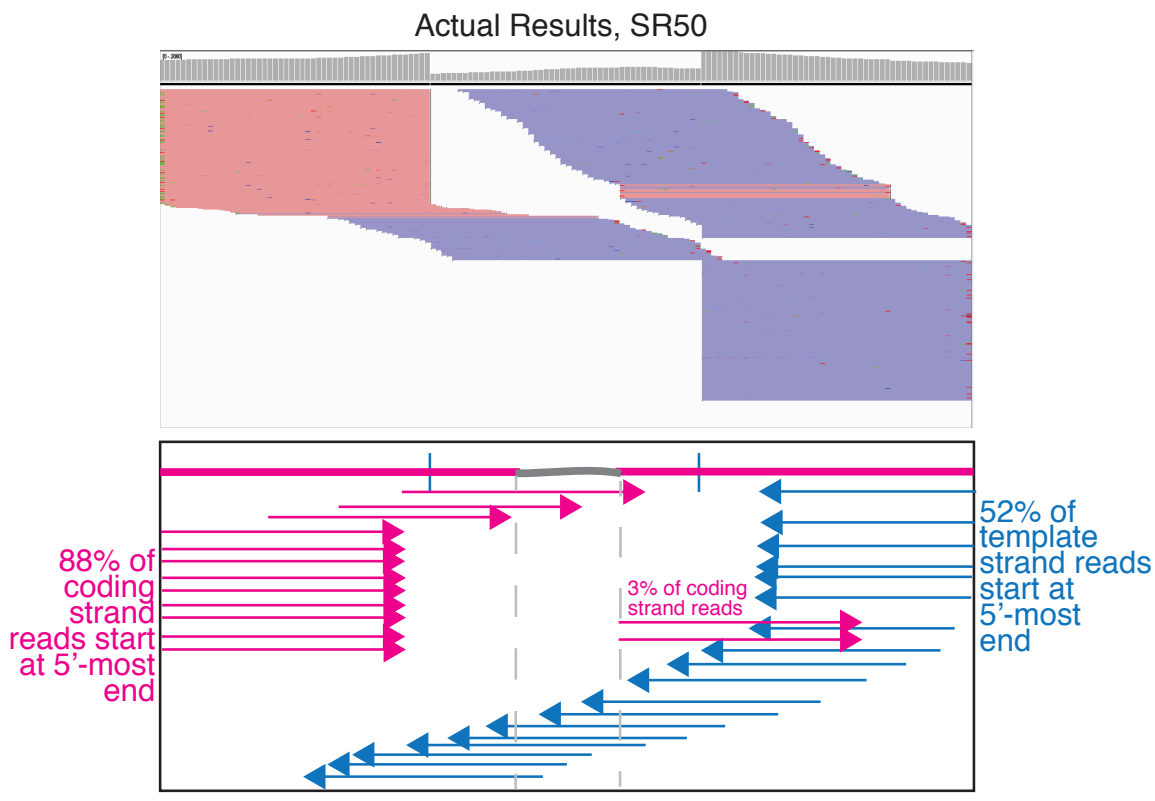
C)



A)



B)



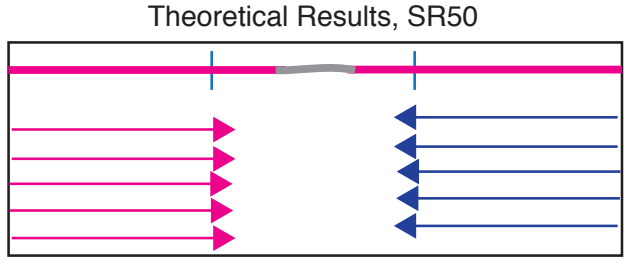
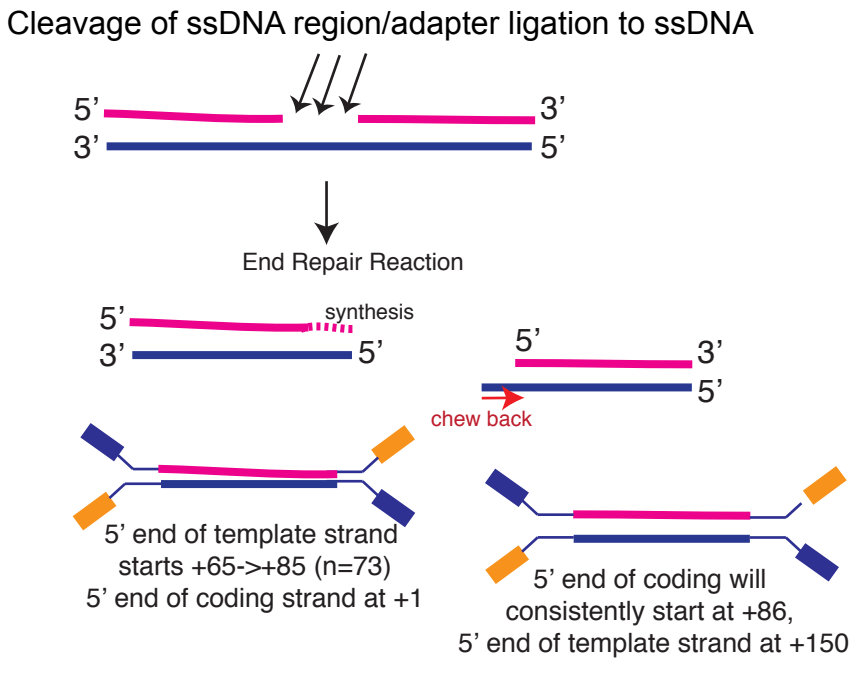
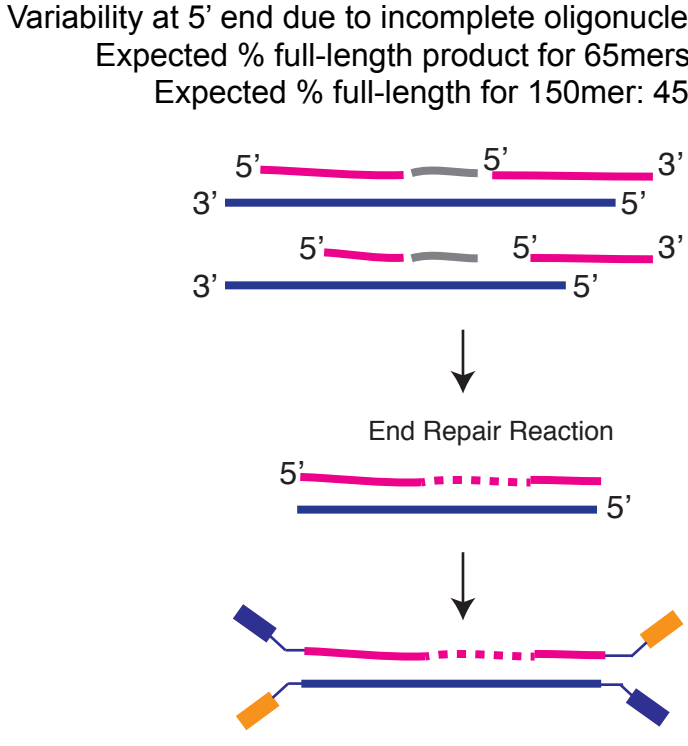


Figure S1.2

Spike-in sequencing

C)

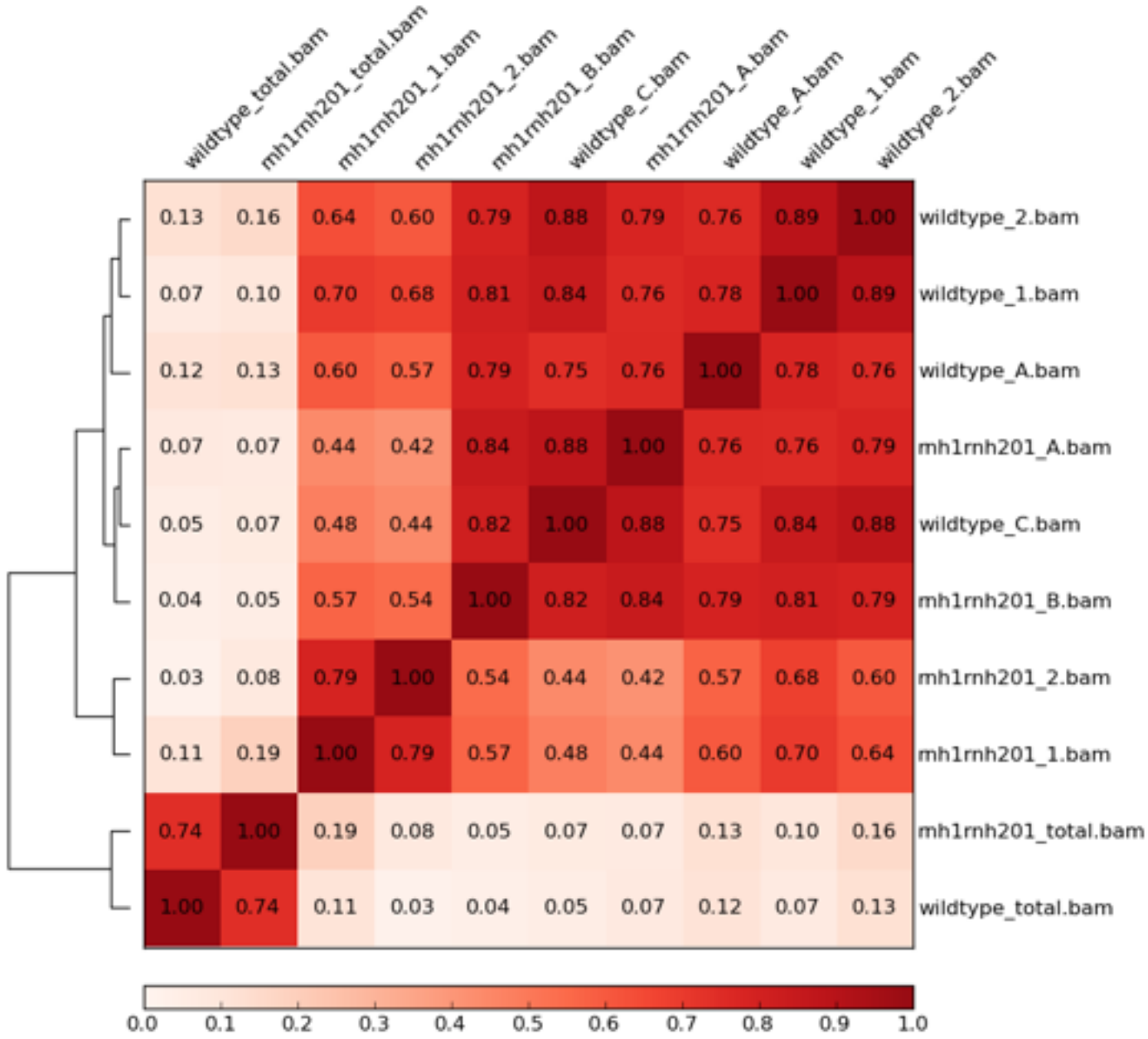




A)

	wildtype			rnh1Δrnh201Δ		
	TOTAL	DRIP	stdev	TOTAL	DRIP	stdev
Total (million)	25.4	11.8	1.0	21.1	11.2	1.6
Nuclear	87.8%	69.6%	11.0%	88.6%	72.1%	4.0%
(a) Ty/Delta	2.5%	13.0%	5.1%	3.2%	41.9%	15.2%
(b) rDNA	9.4%	45.5%	10.2%	7.5%	34.2%	5.1%
(c) Peaks	2.7%	4.8%	0.9%	1.5%	3.1%	0.5%
Mitochondrial	6.0%	17.2%	8.5%	9.1%	14.6%	4.8%

B)



Spearman Correlation Coefficients

Figure S1.3
DRIP results
and
correlation

Figure S1.4 genome hybrid sequencing

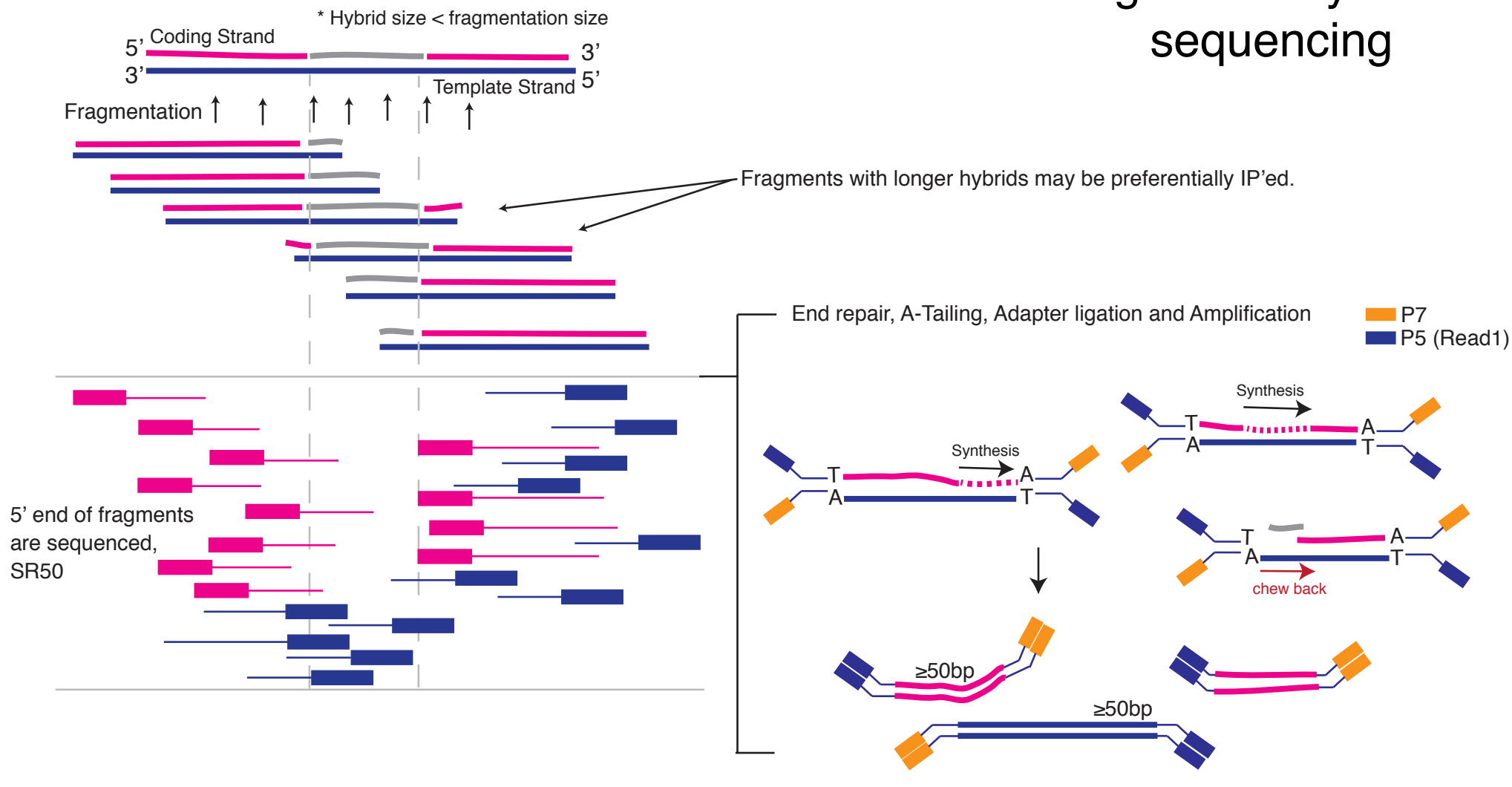
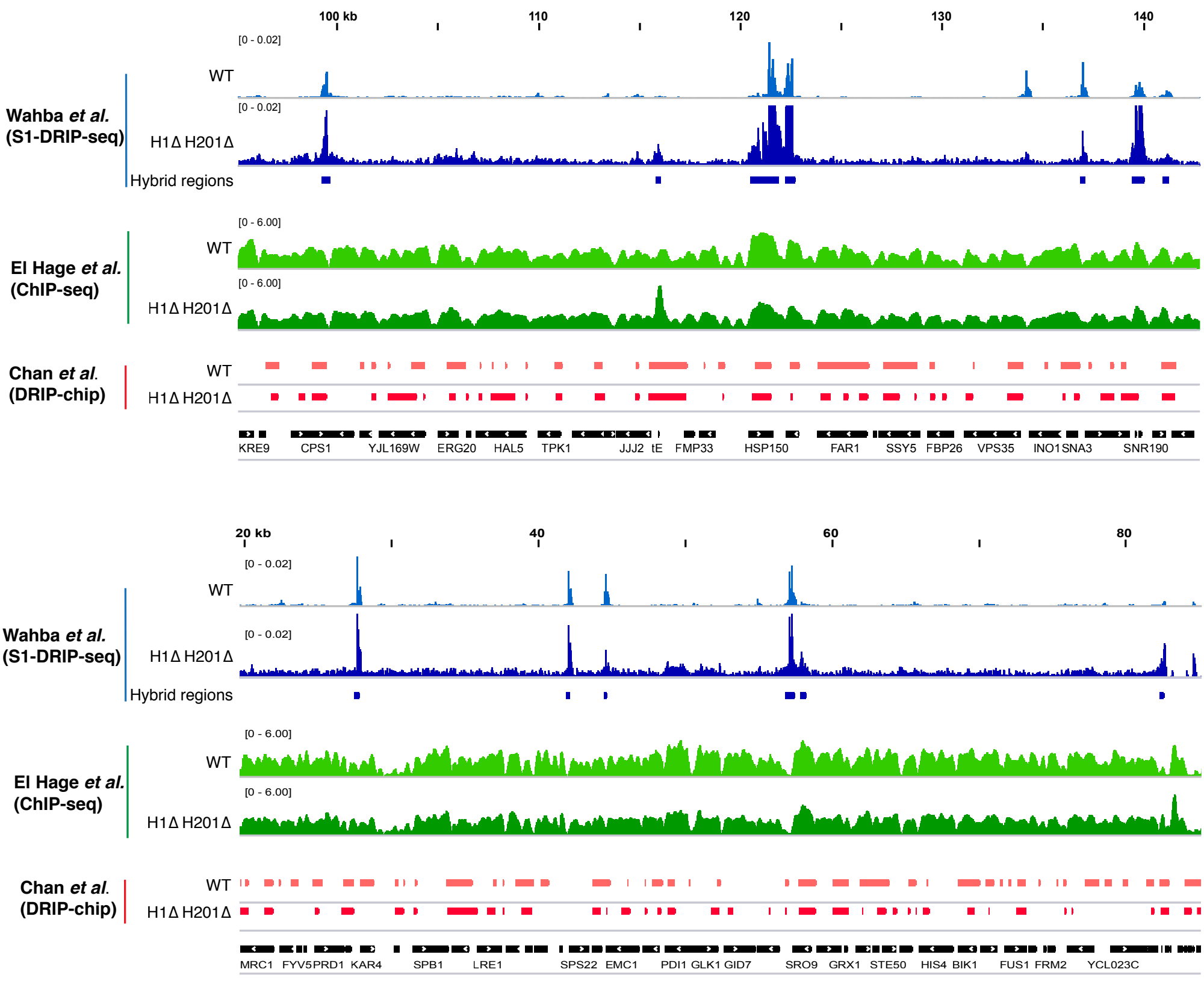


Figure S1.5 comparison with other methods

A)



B)

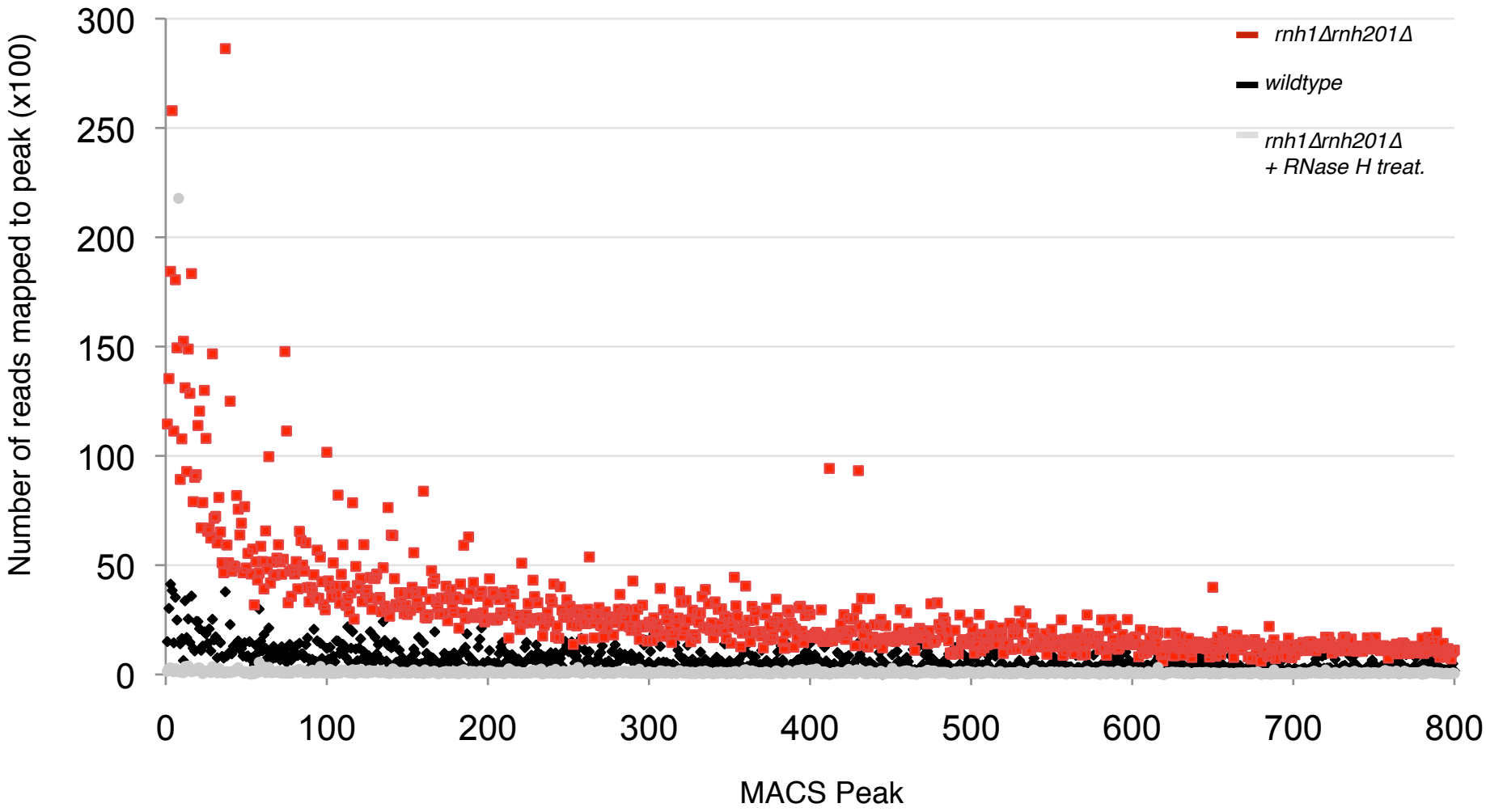
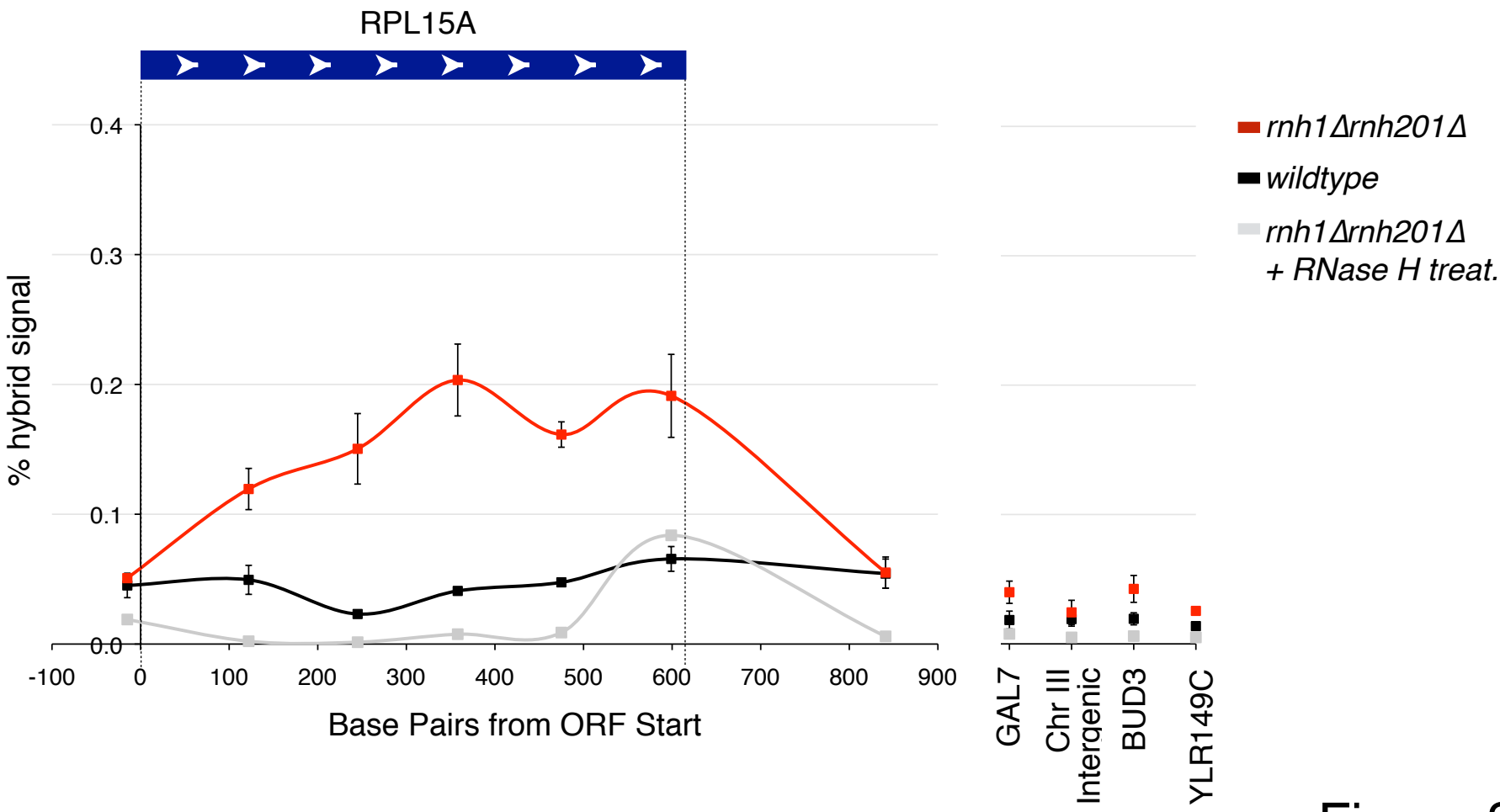


Figure S1.5B
S1-DRIP-seq
signal

A)



B)

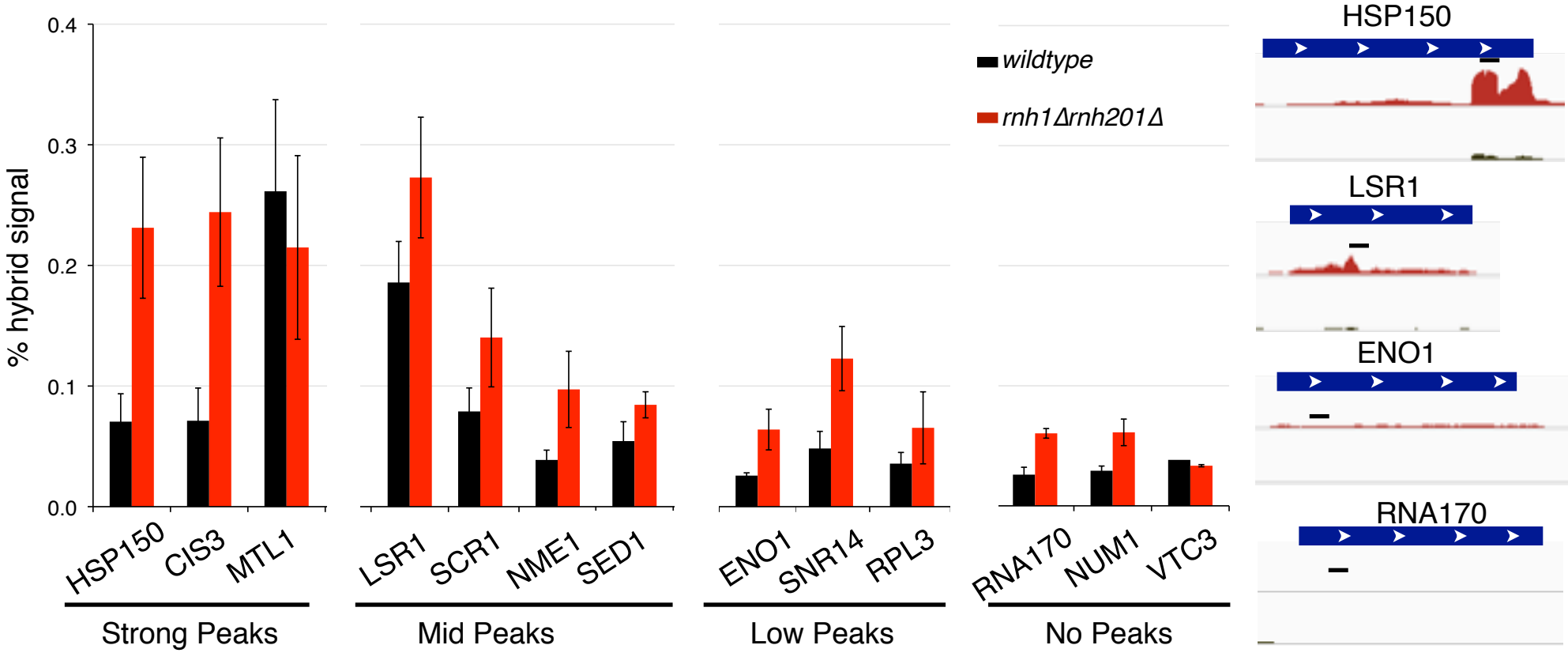


Figure S1.6
DRIP-qPCR
validation

A)

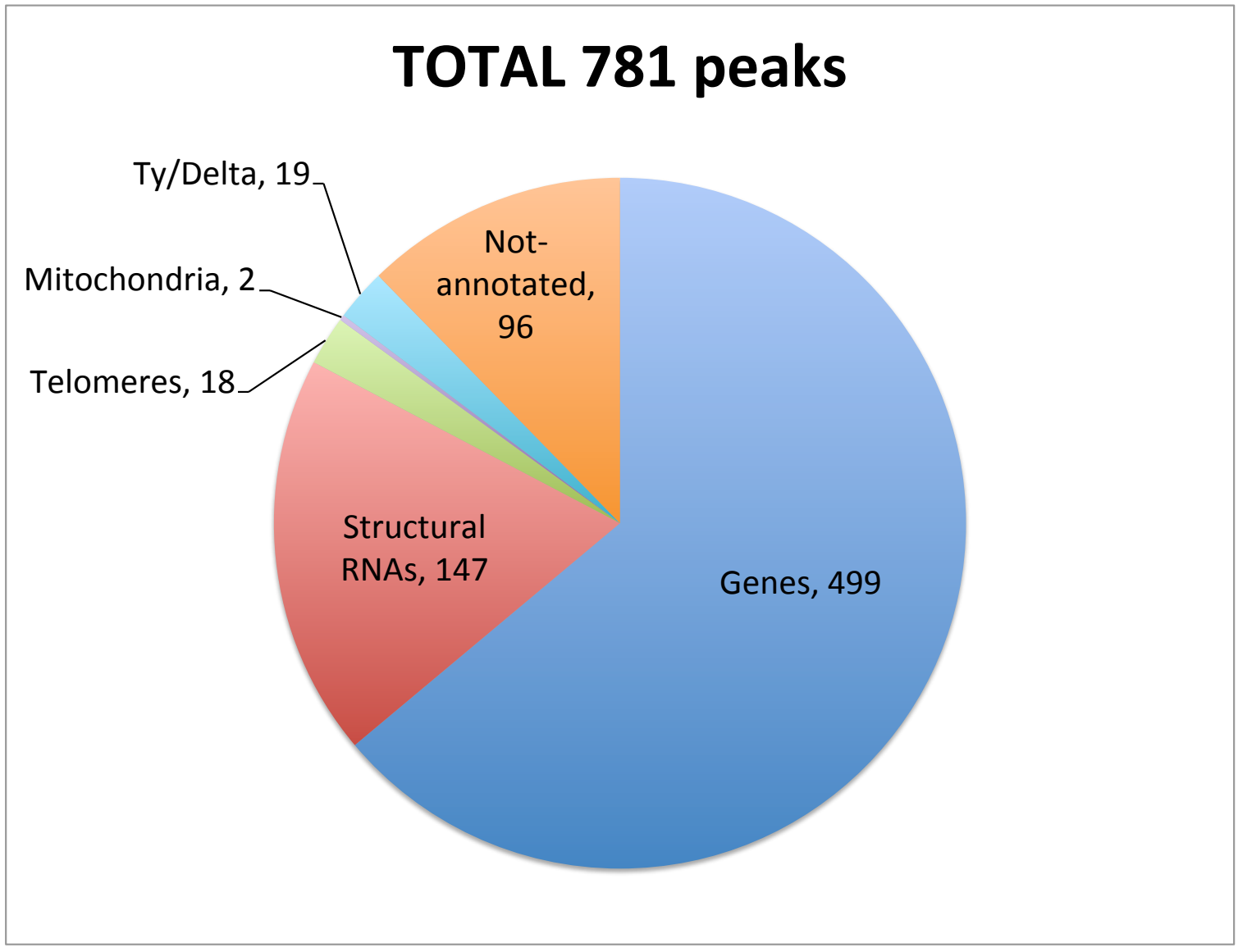


Figure S2.1
DRIP
distribution

A)

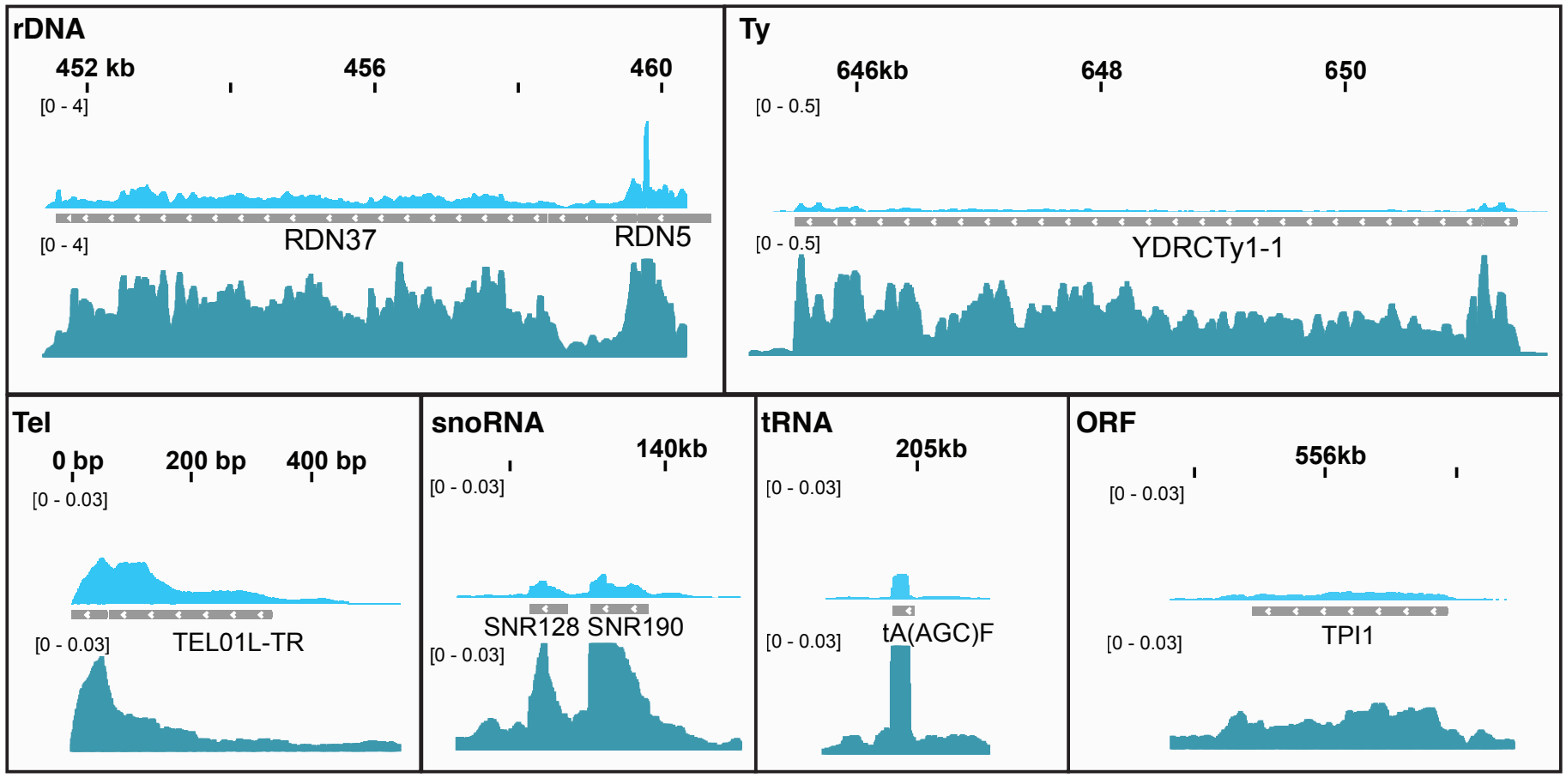
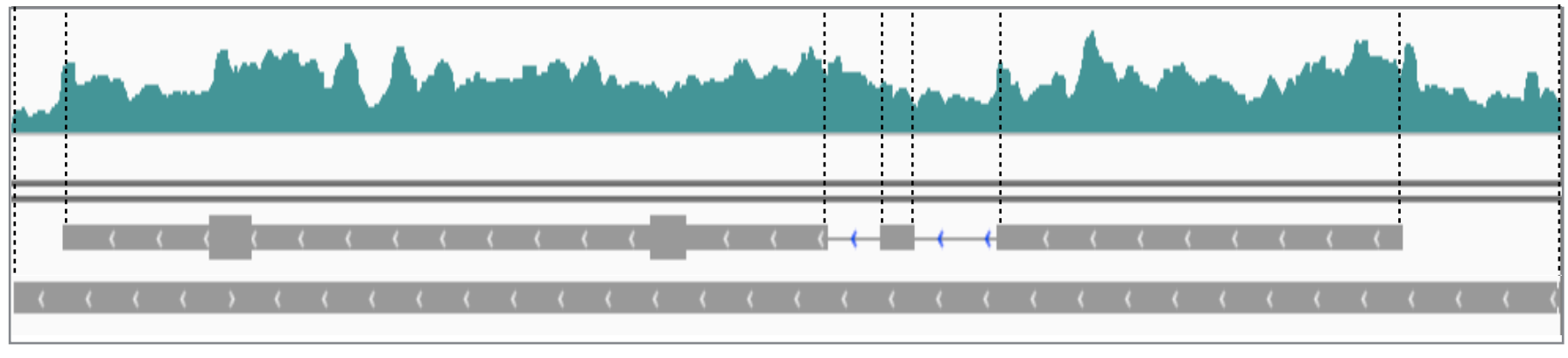
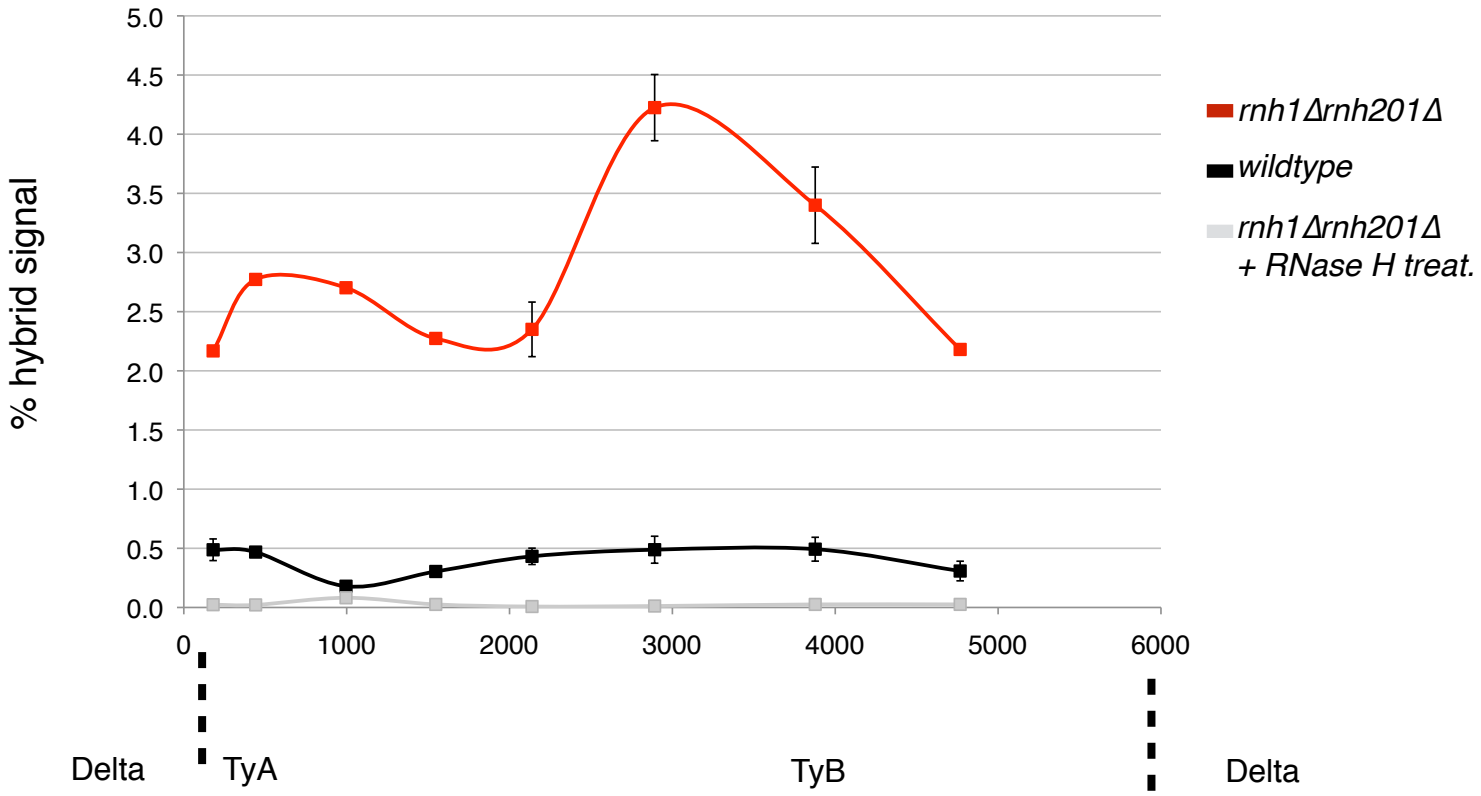


Figure S2.2
Hybrid-forming regions WT vs H1ΔH201Δ

B)



A)



B)

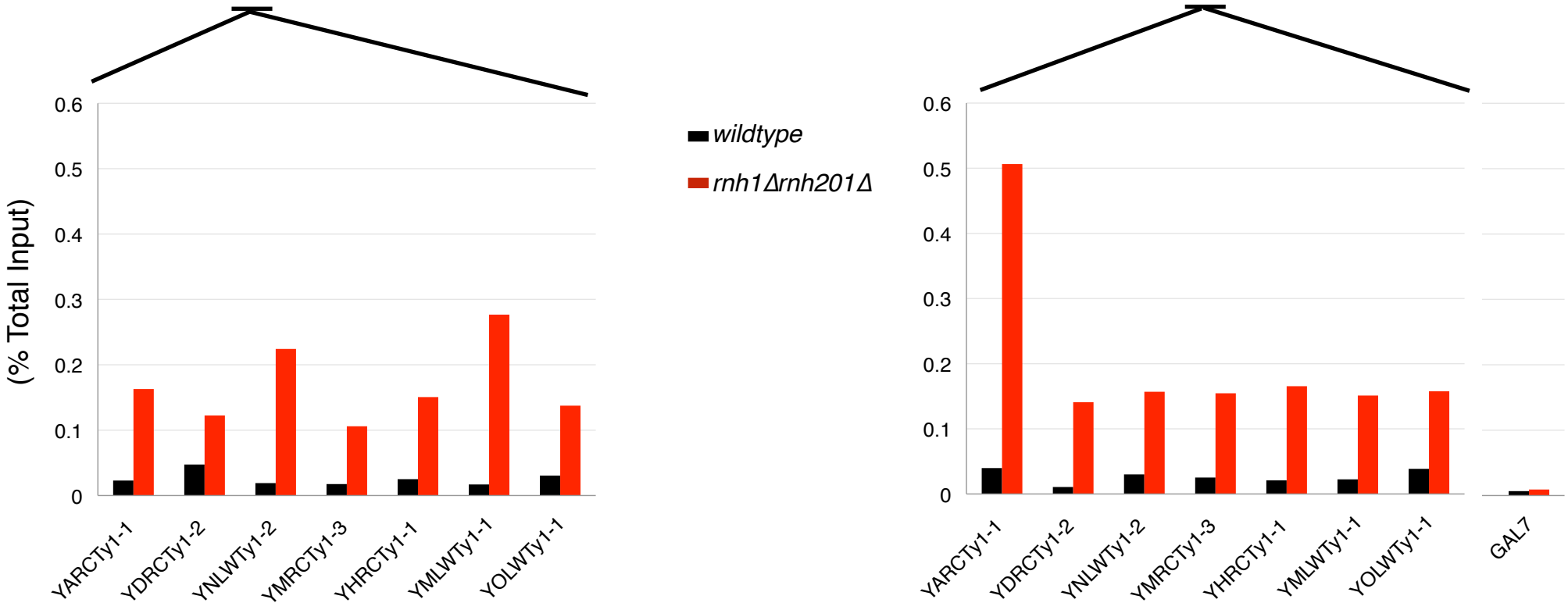


Figure S2.3
sonication and
DRIP-qPCR

A)

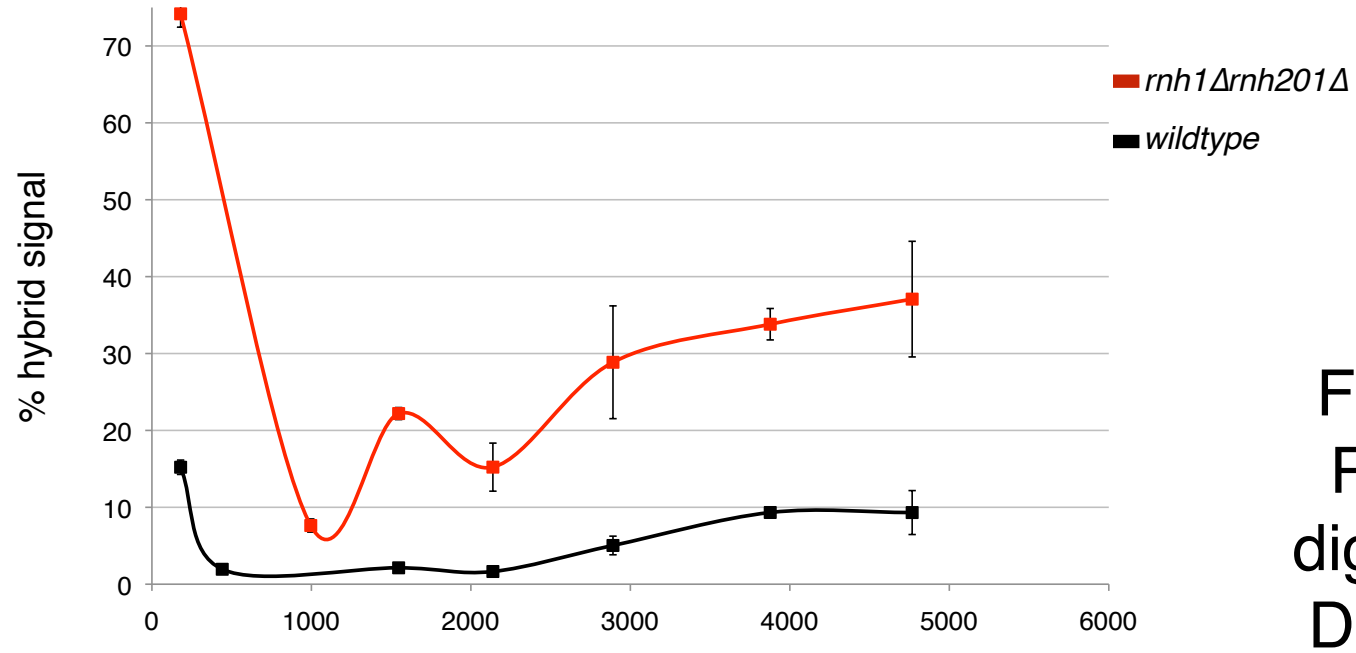
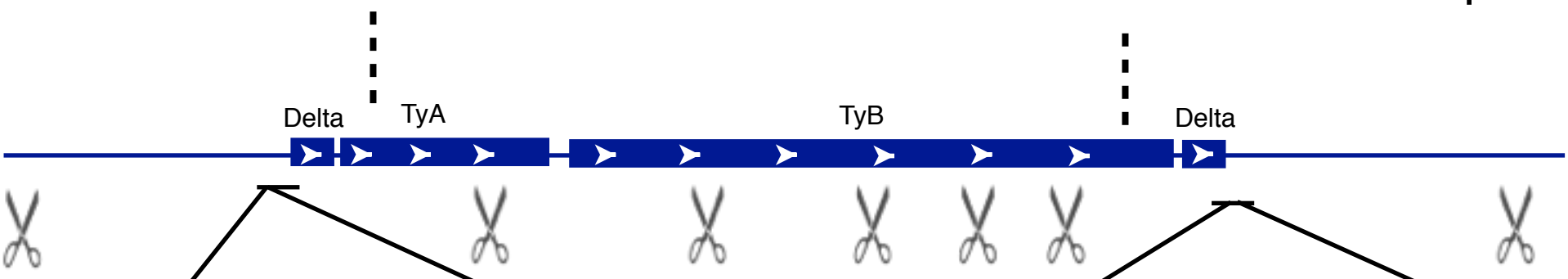
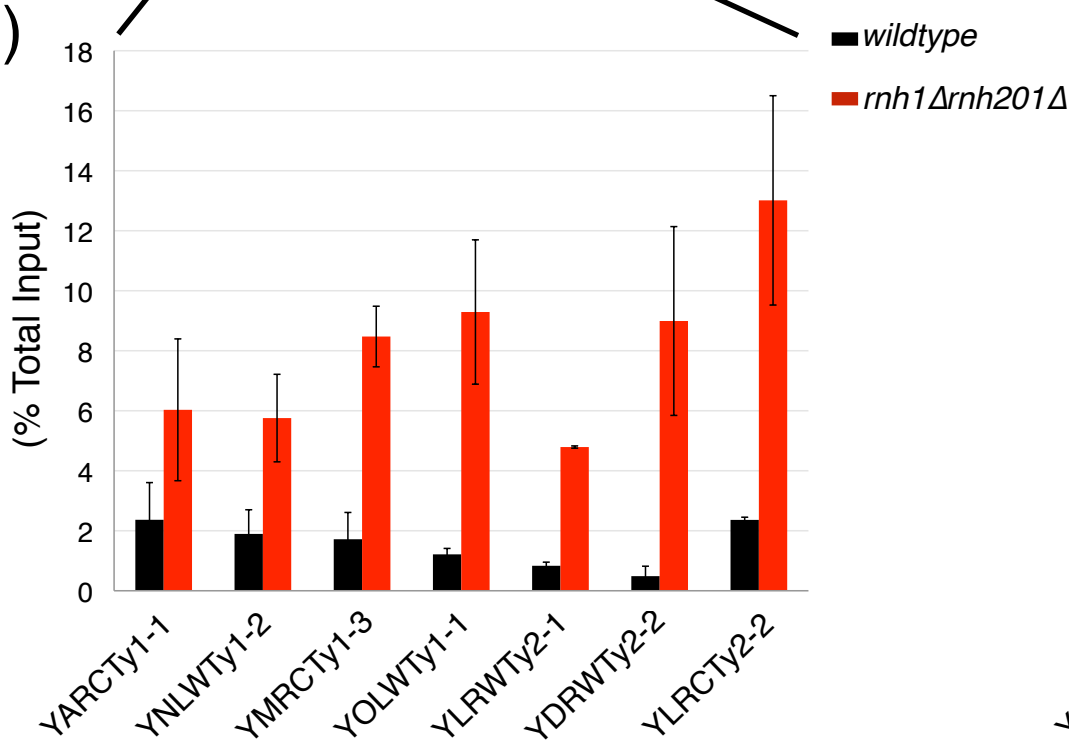
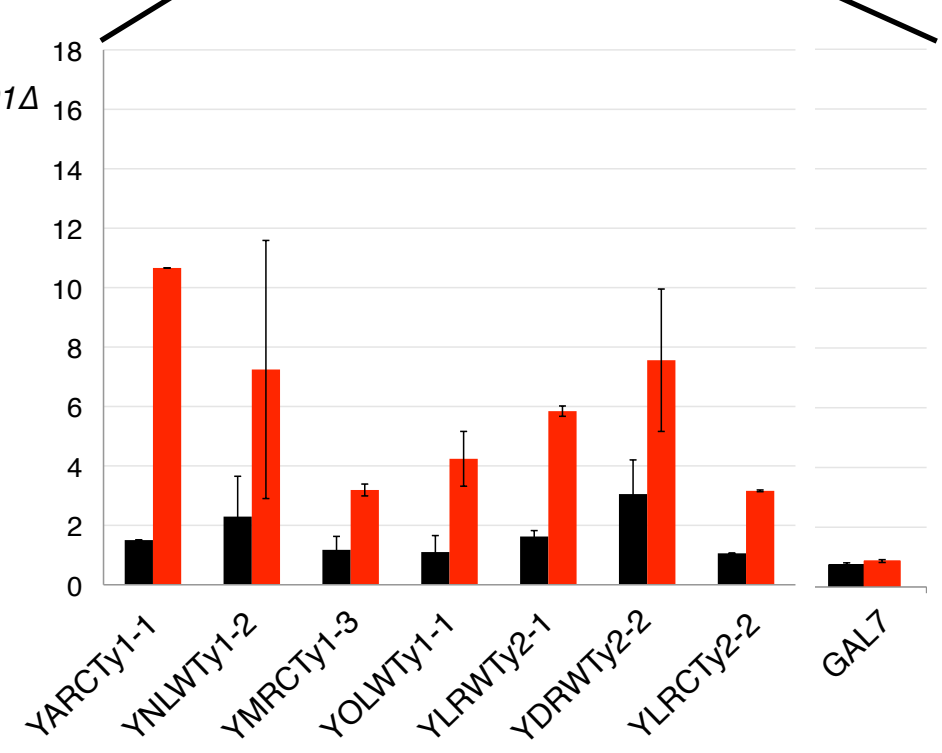


Figure S2.4 Restriction digestion and DRIP-qPCR



B)





A)

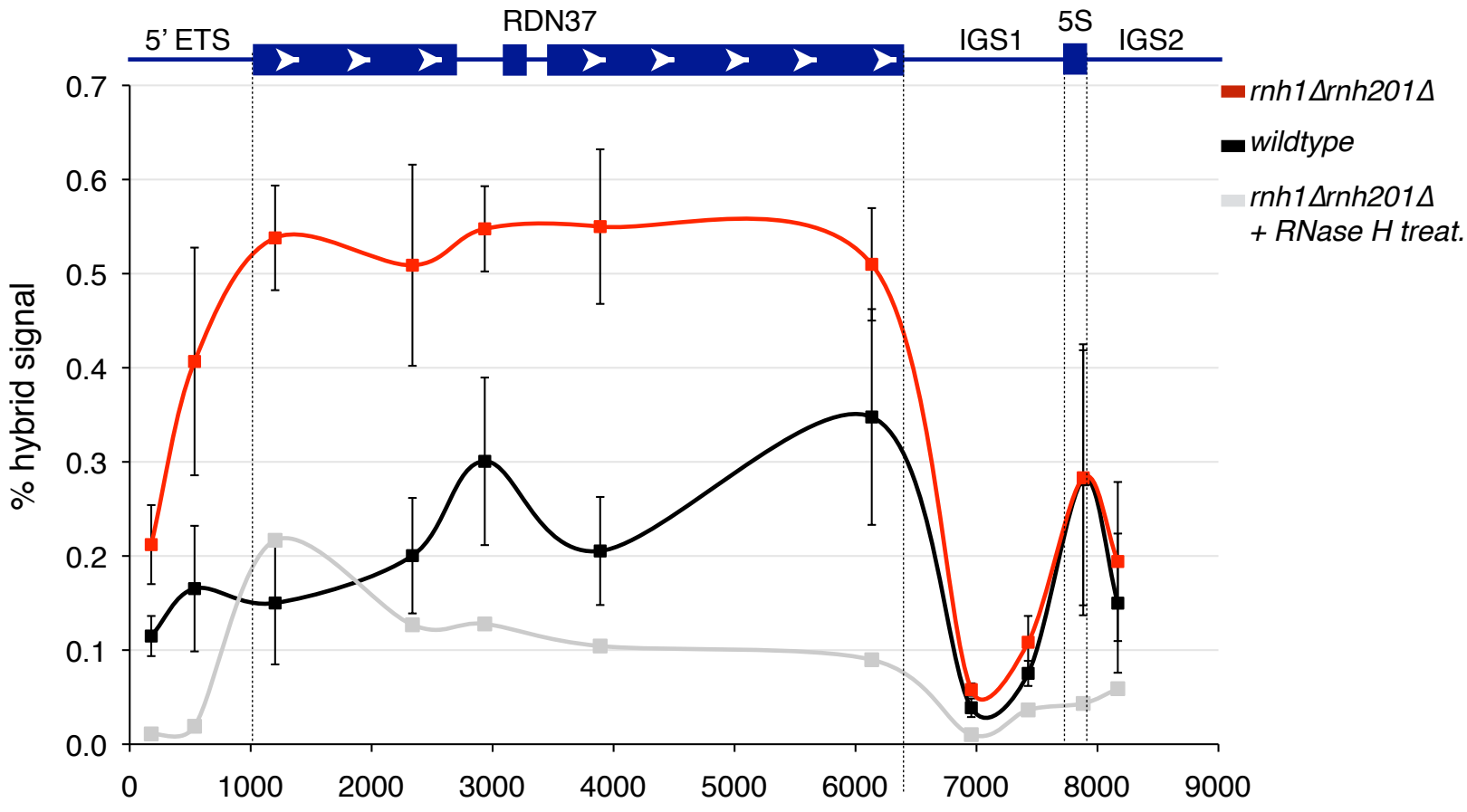
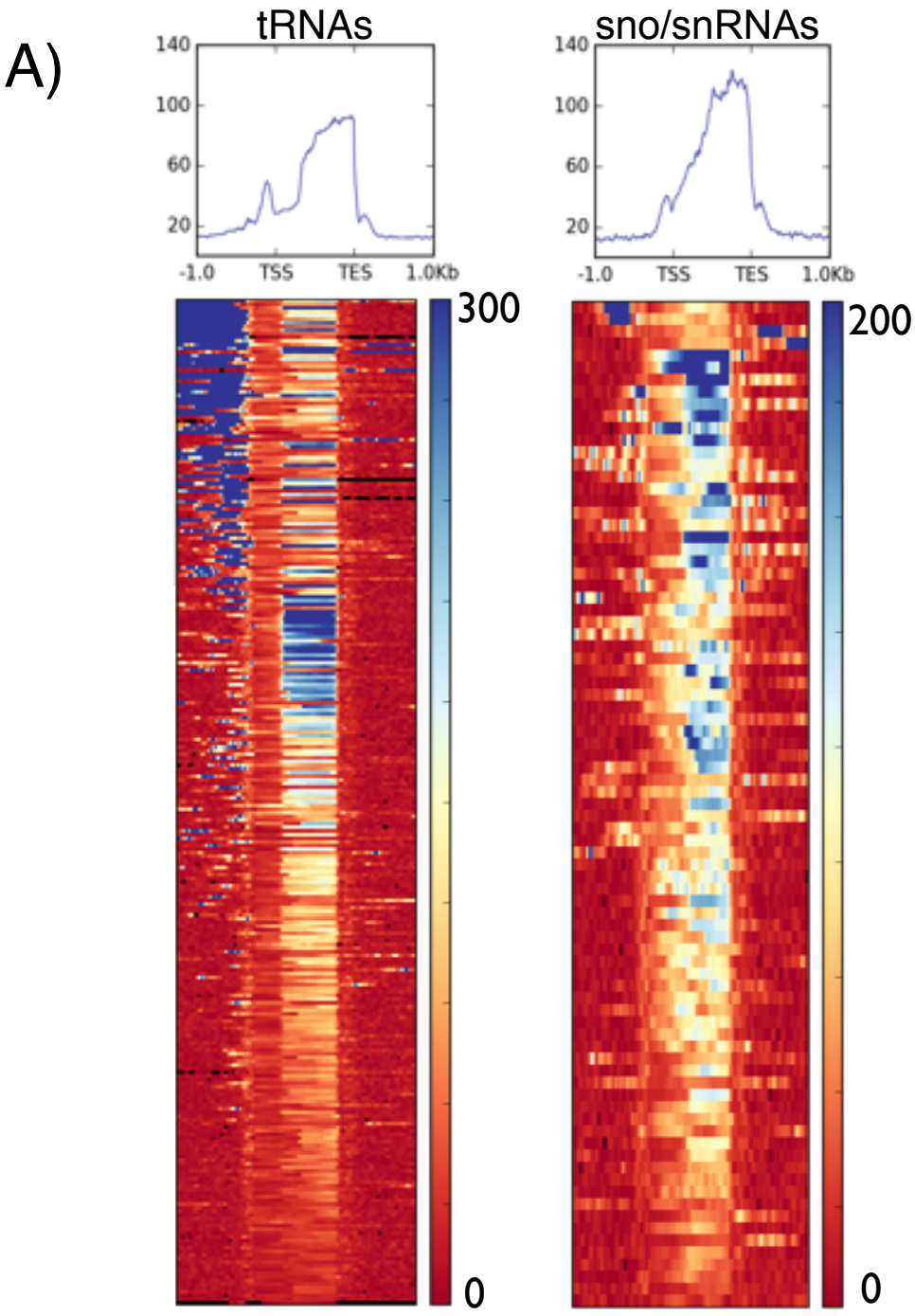
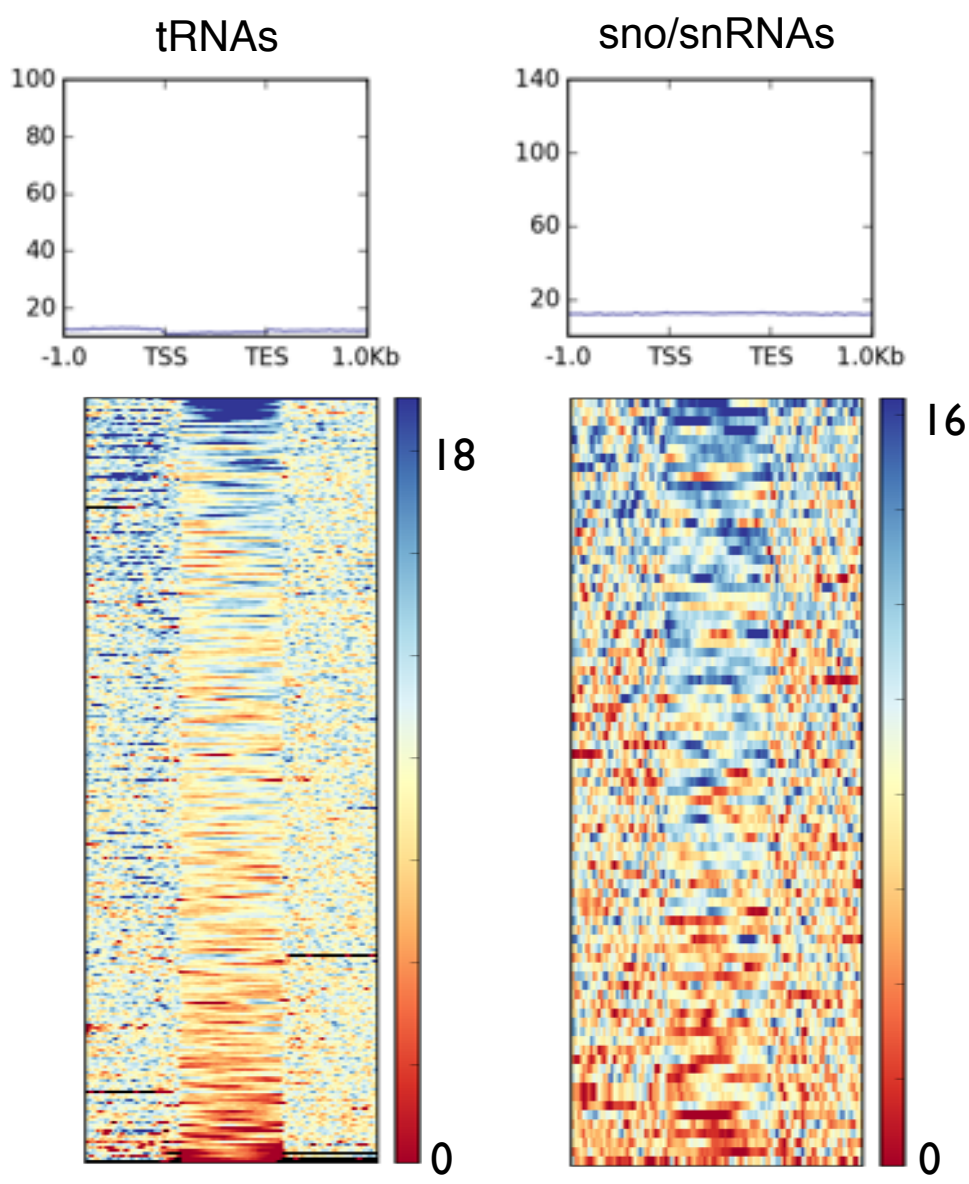


Figure S2.5
rDNA DRIP-qPCR

A)



Signal from DRIP



Signal from TOTAL

TOTALS

Figure S2.6 DRIP and

Figure S2.7
mitochondria
DRIP

A)

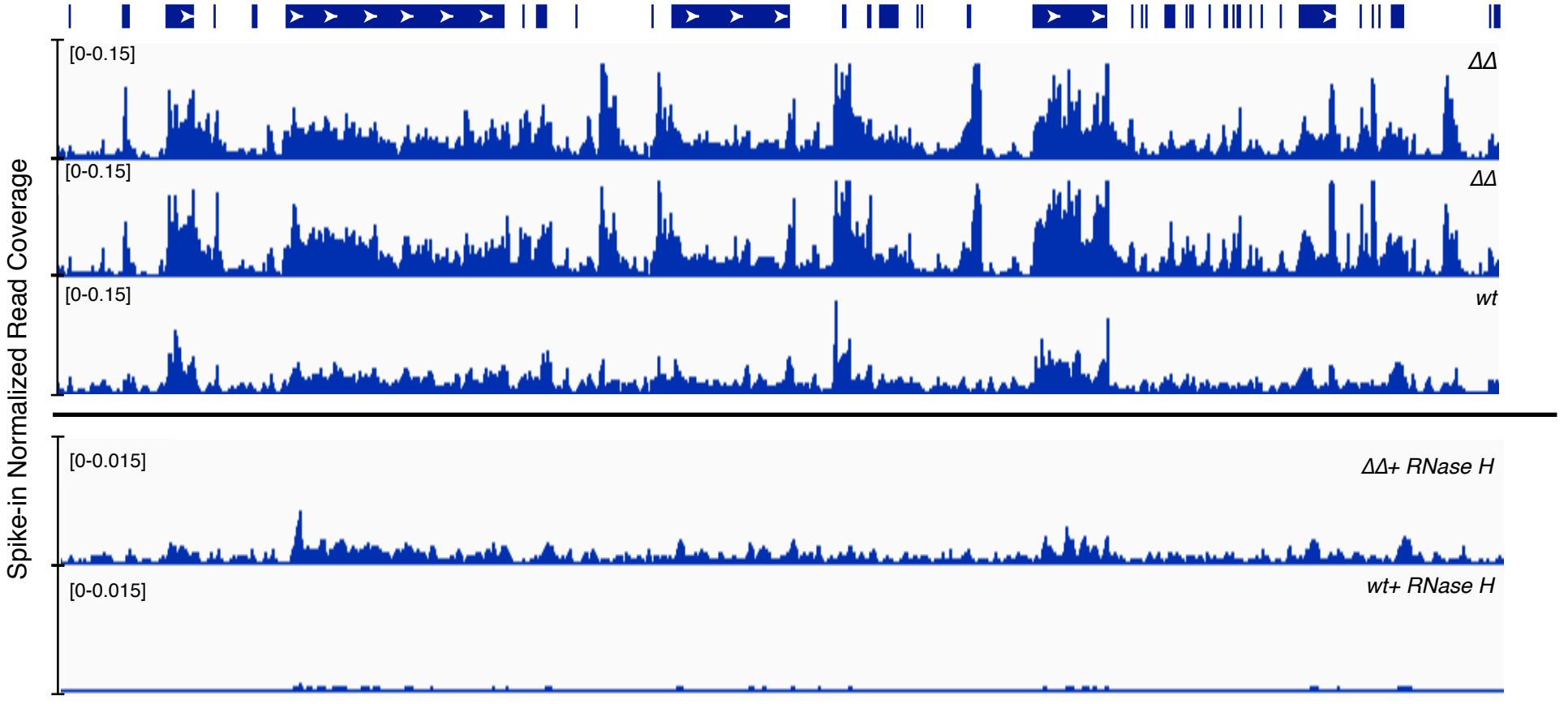
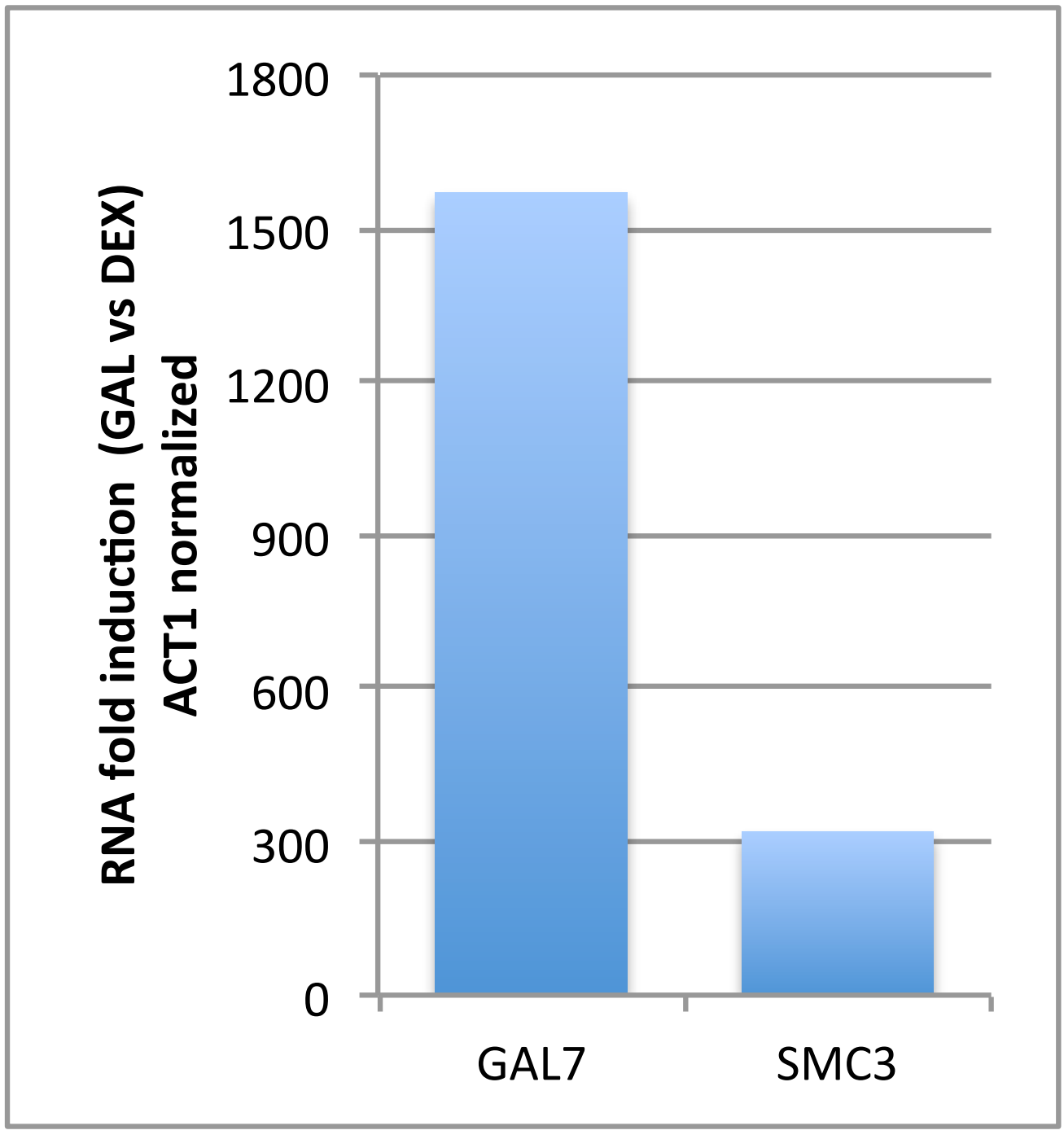
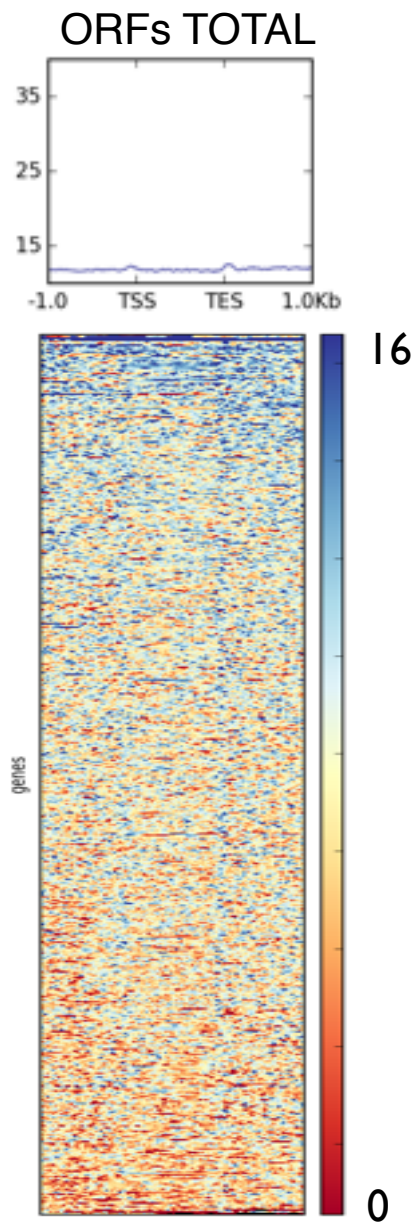


Figure S3
RNA levels
upon GAL
induction

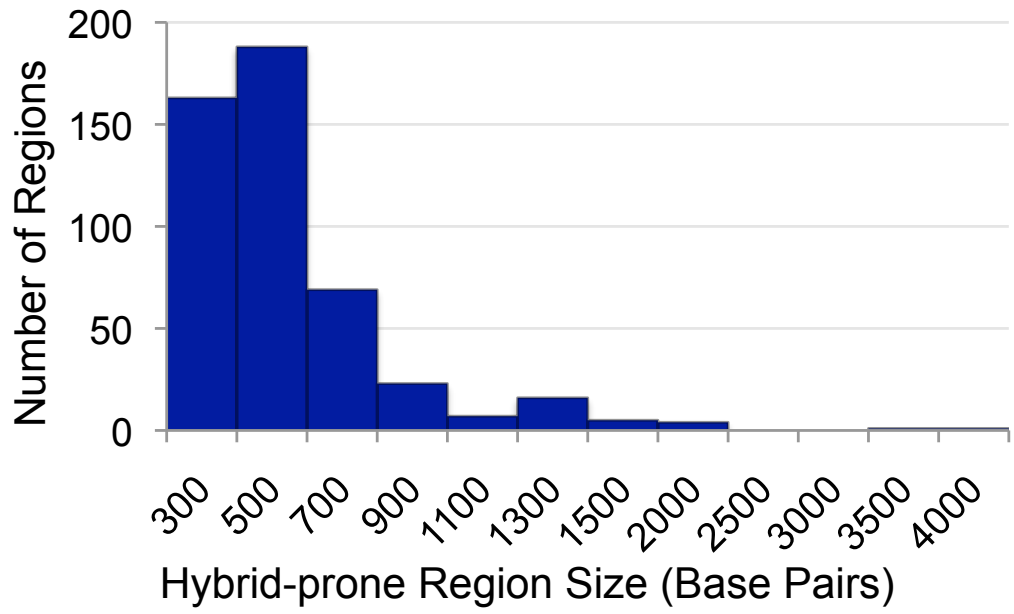
A)



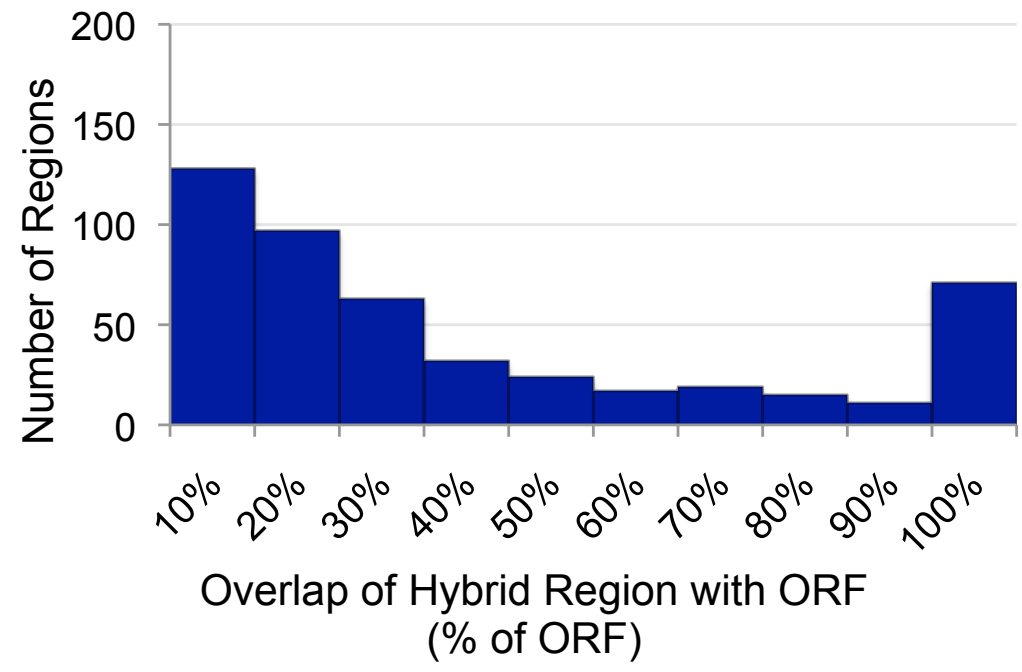
A)



B)



C)



D)

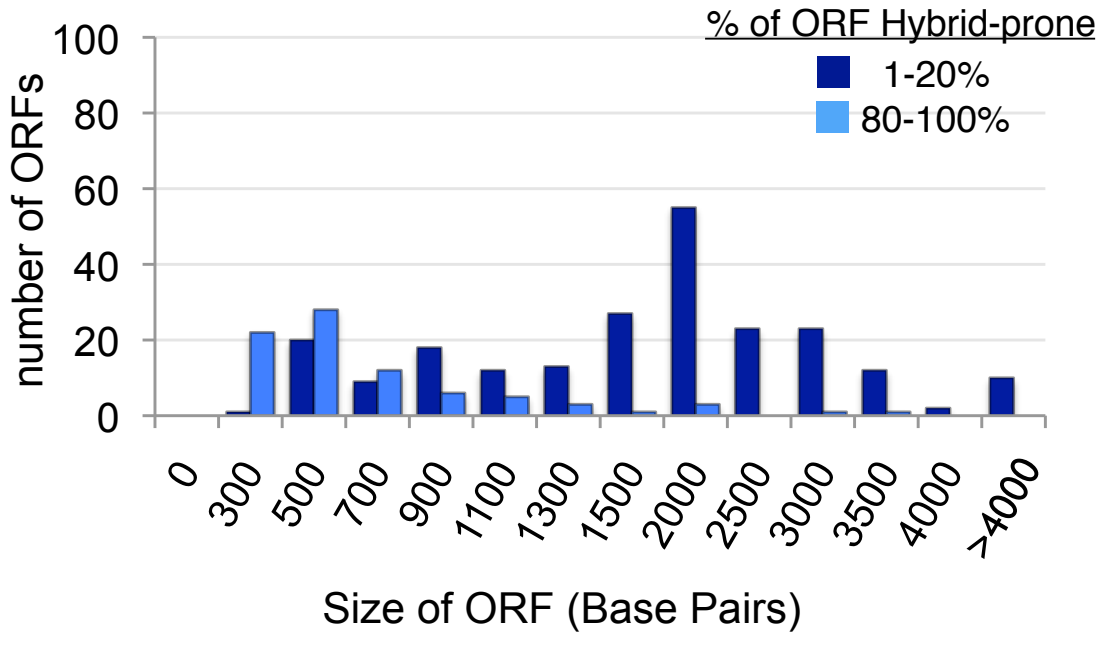
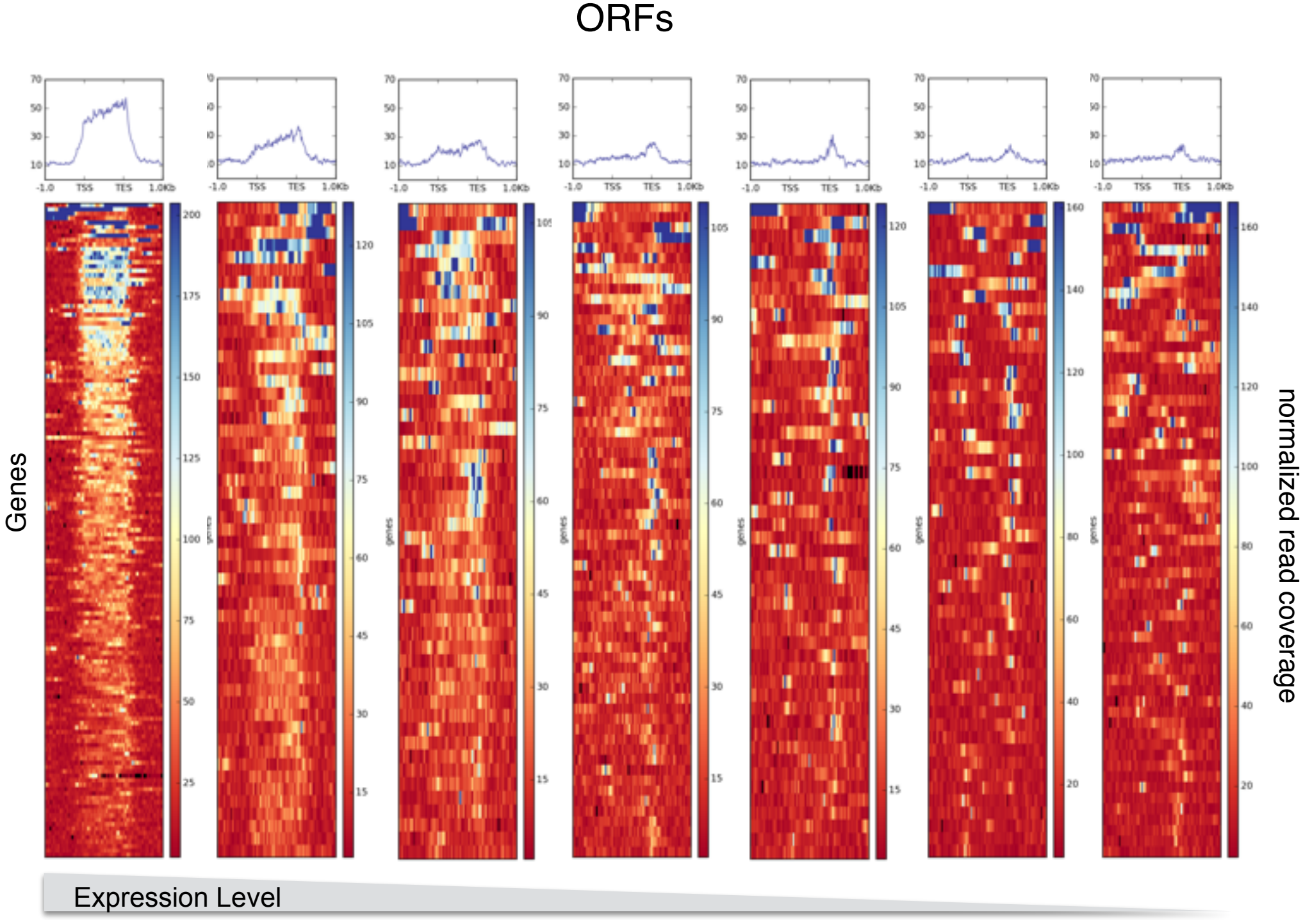


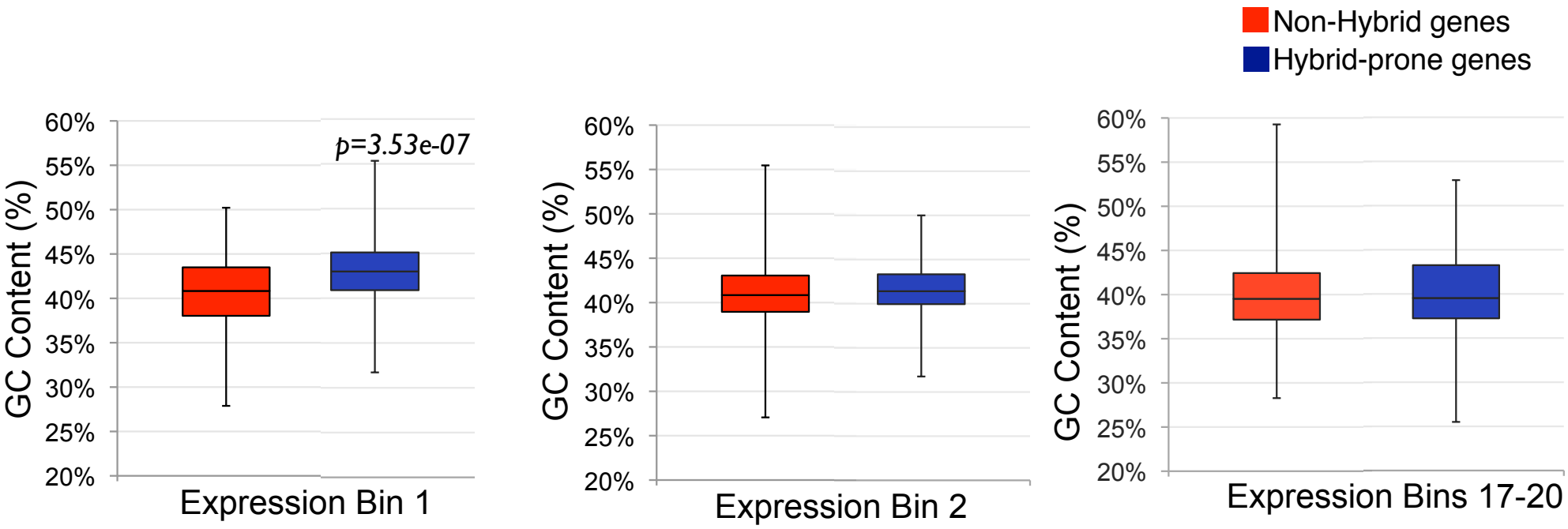
Figure S4.1
ORF and
hybrids

Figure S4.2 expression levels heat maps

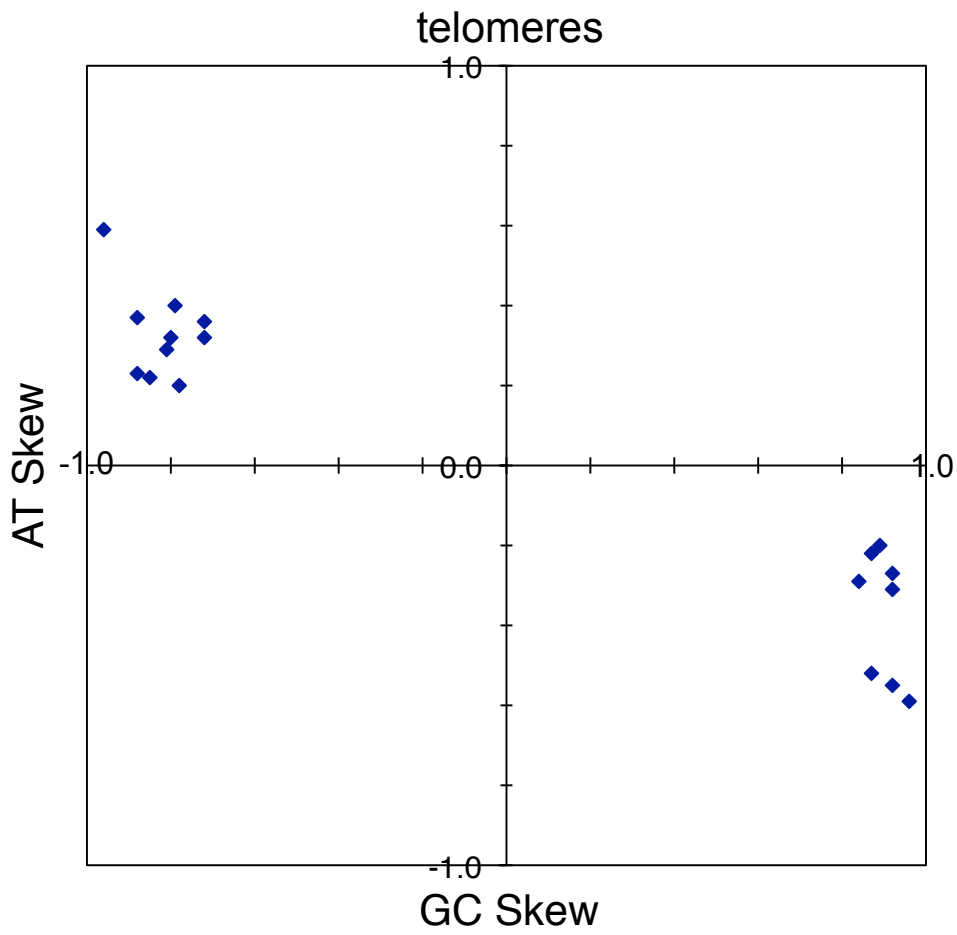
A)



A)



B)



C)

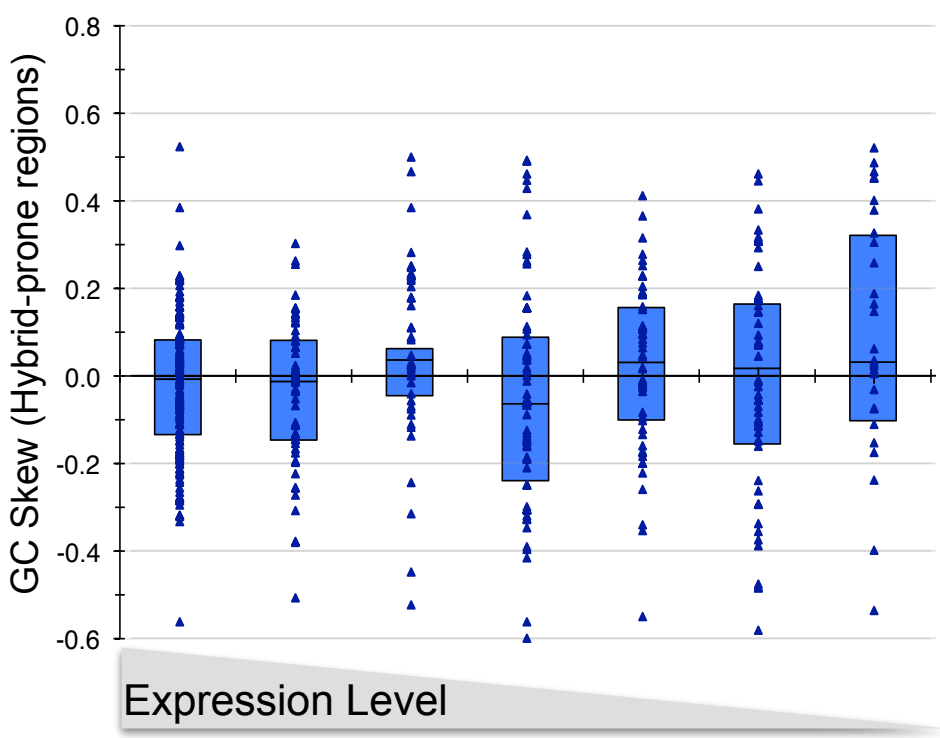


Figure S5.1
GC content,
GC skew,
AT skew

A)

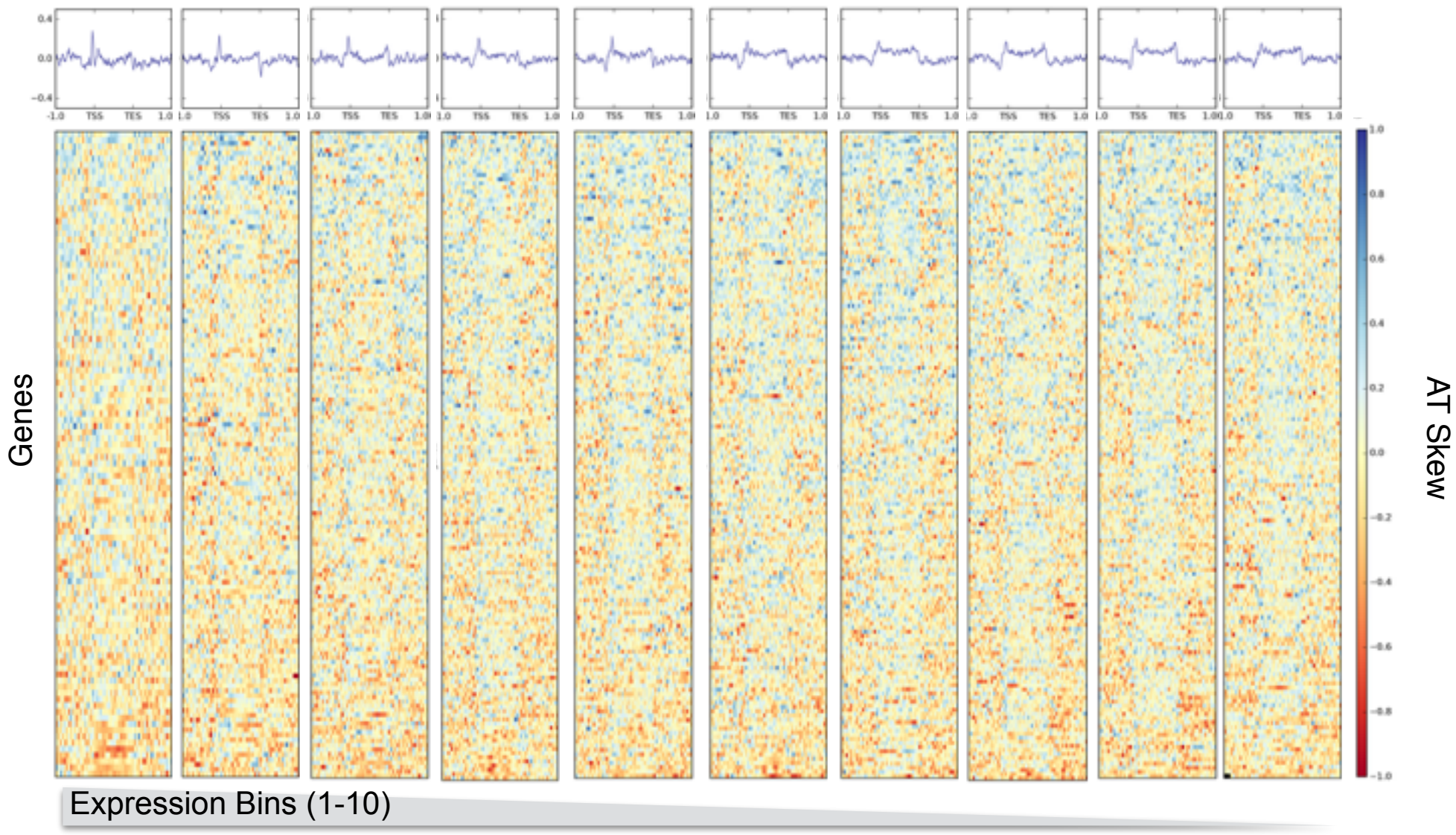
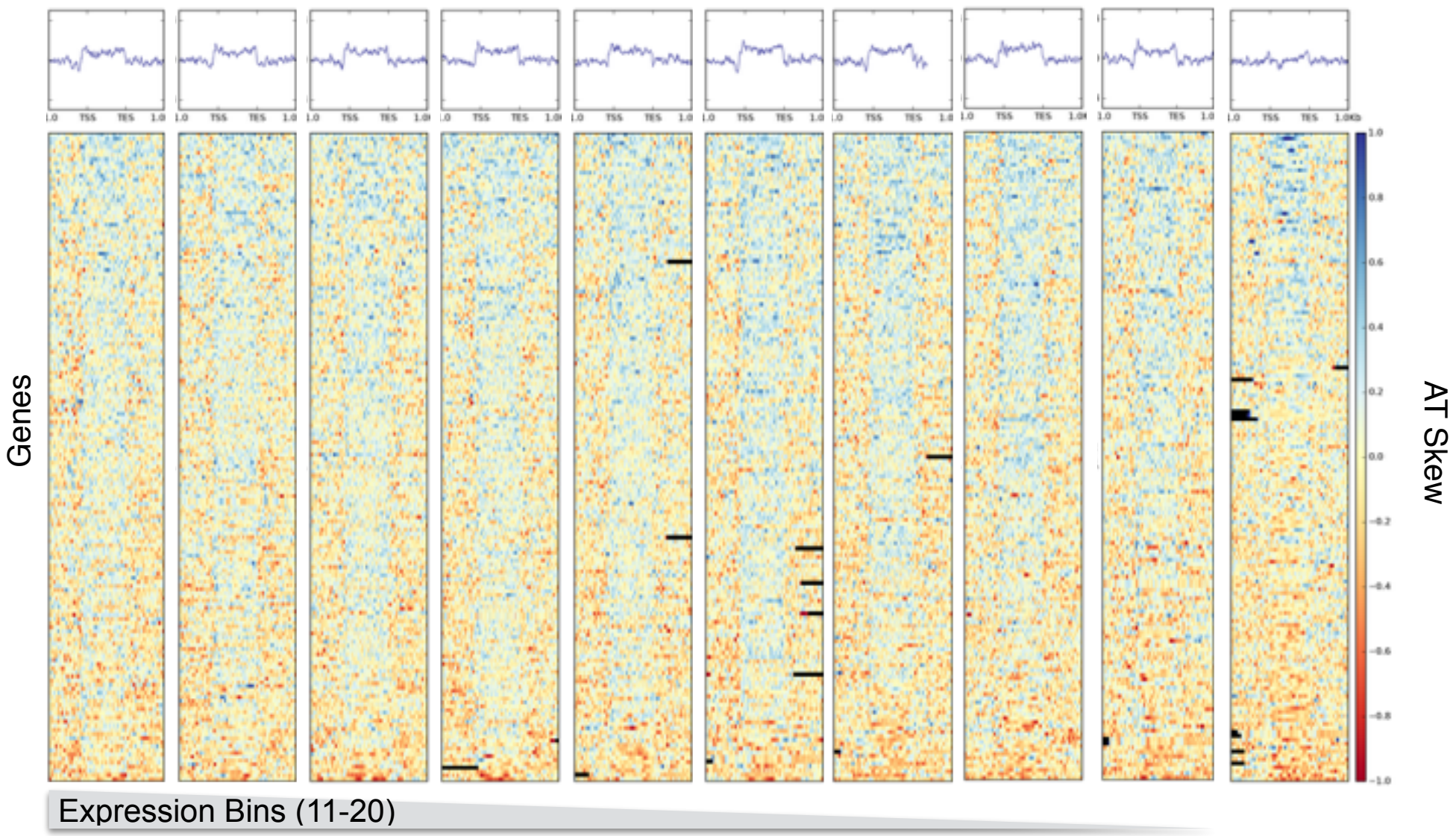


Figure S5.2

AT skew heat map



A)

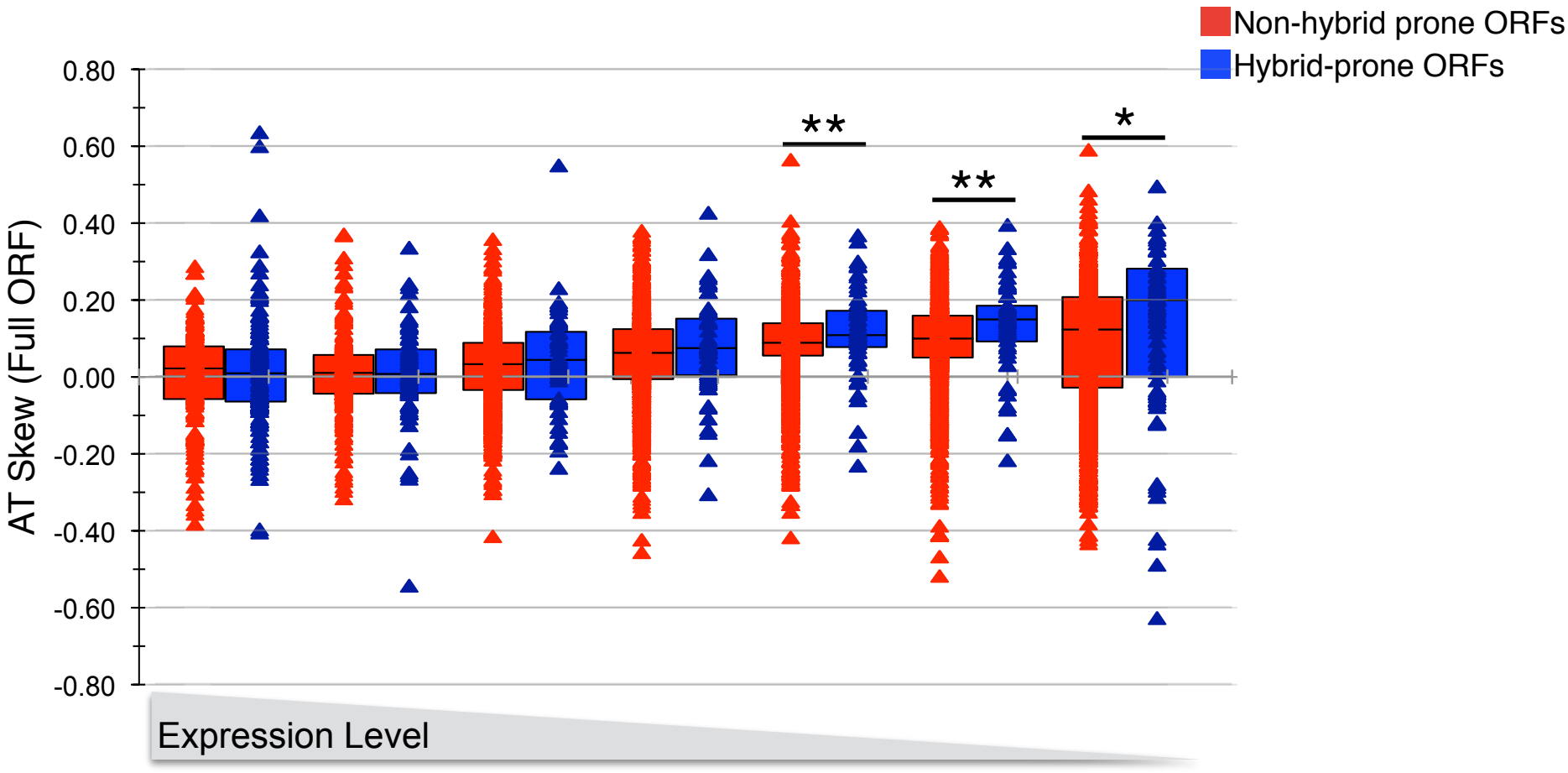
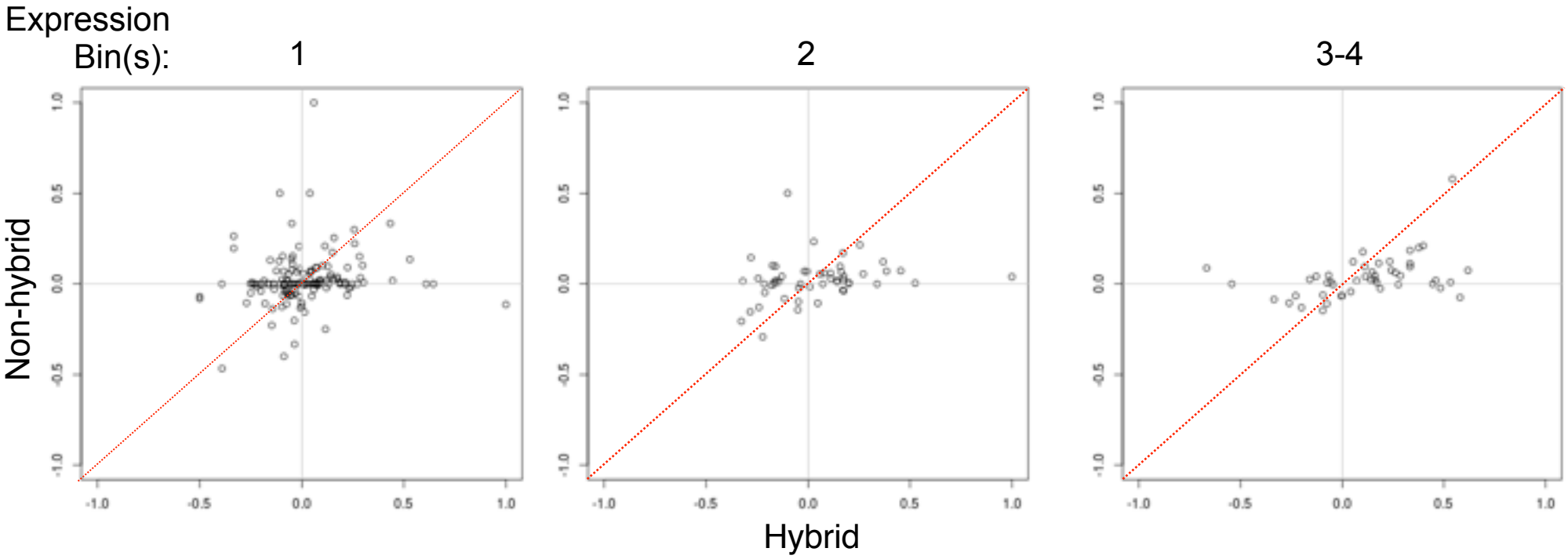
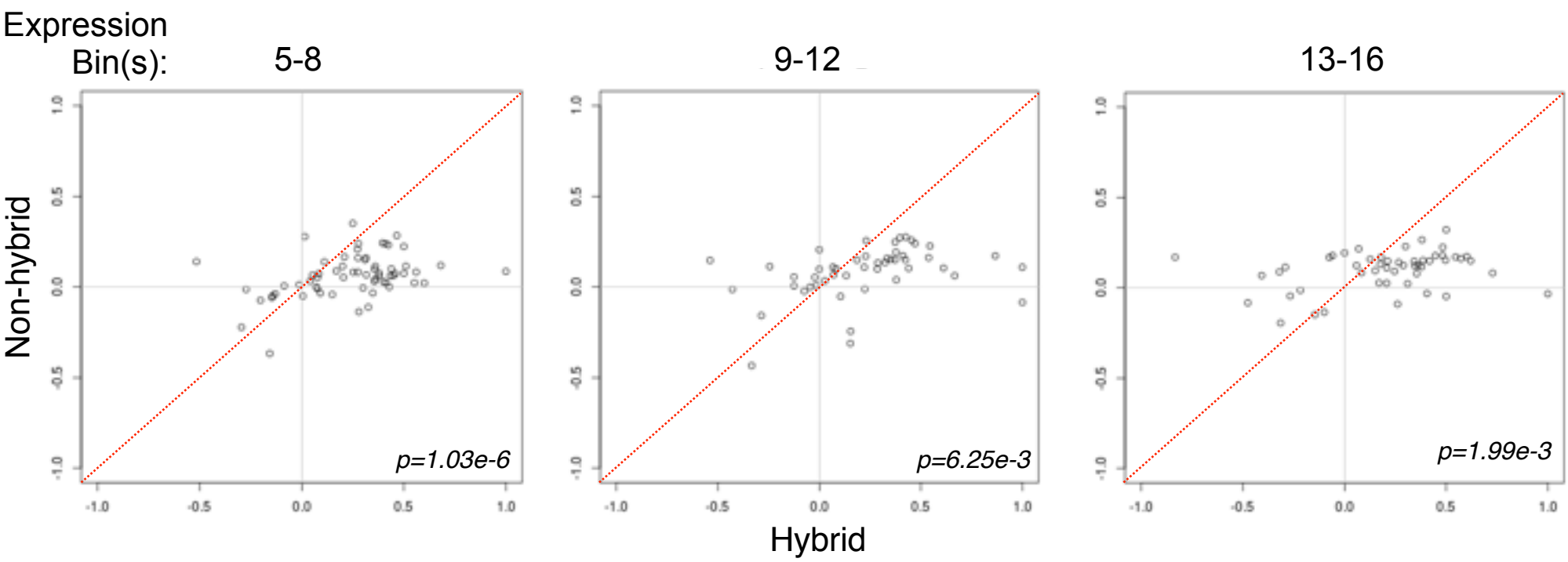


Figure S5.3
AT skew
hybrid ORFs
vs allORFs





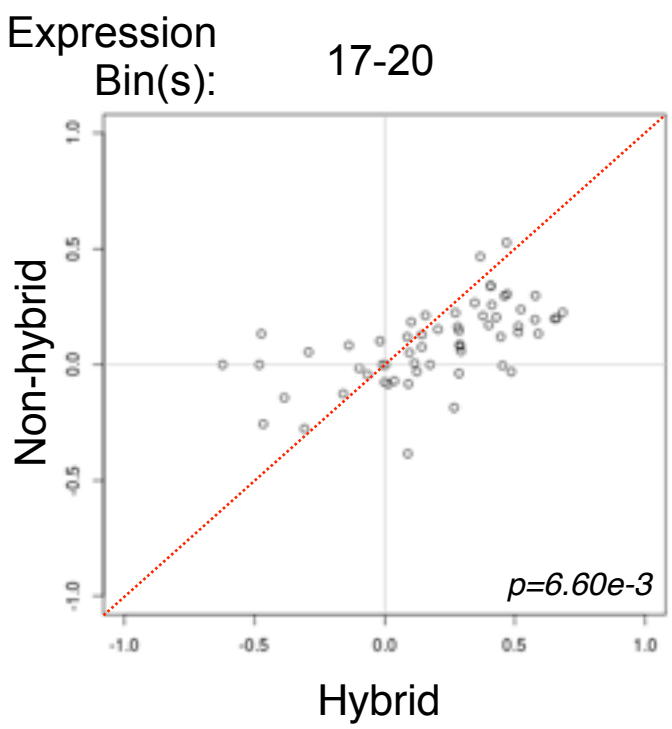
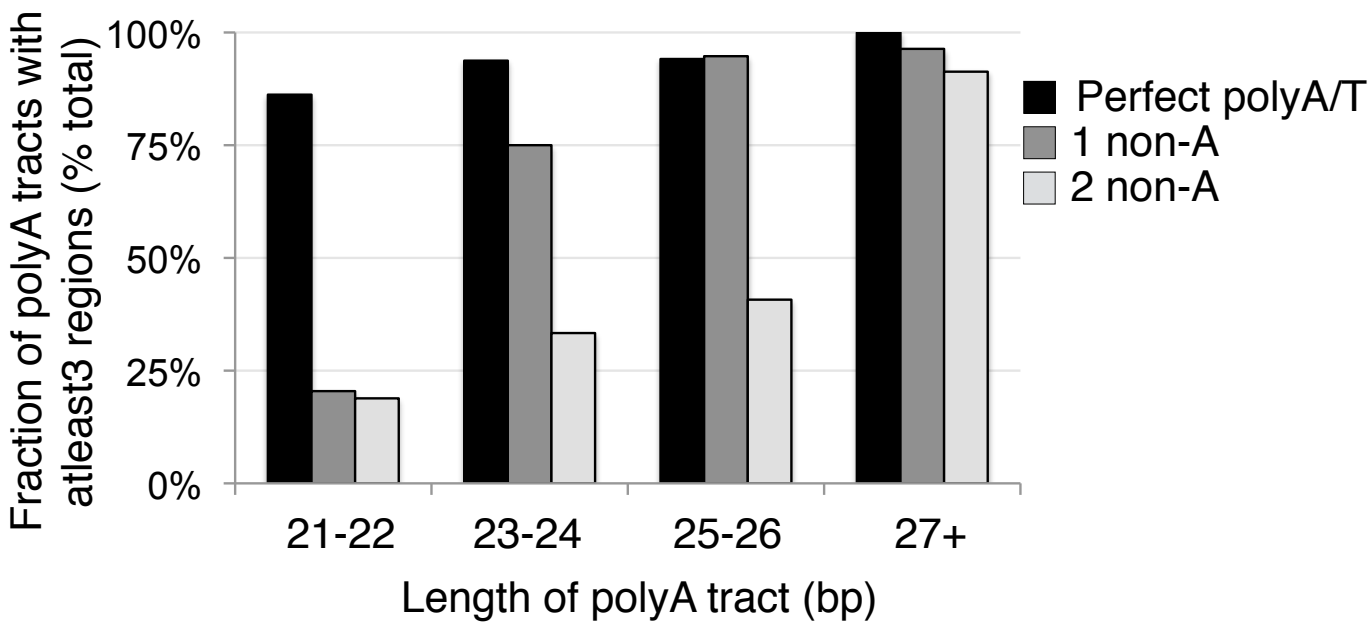


Figure S5.4
AT skew

A)



B)

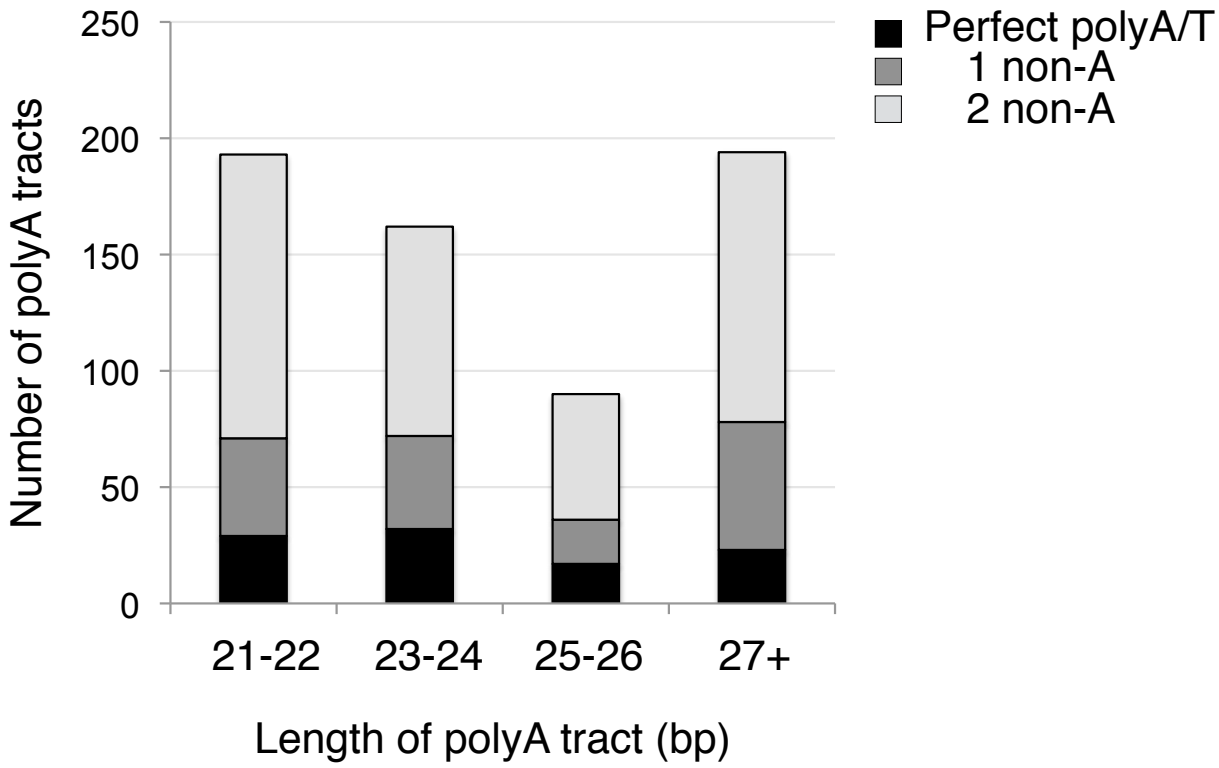
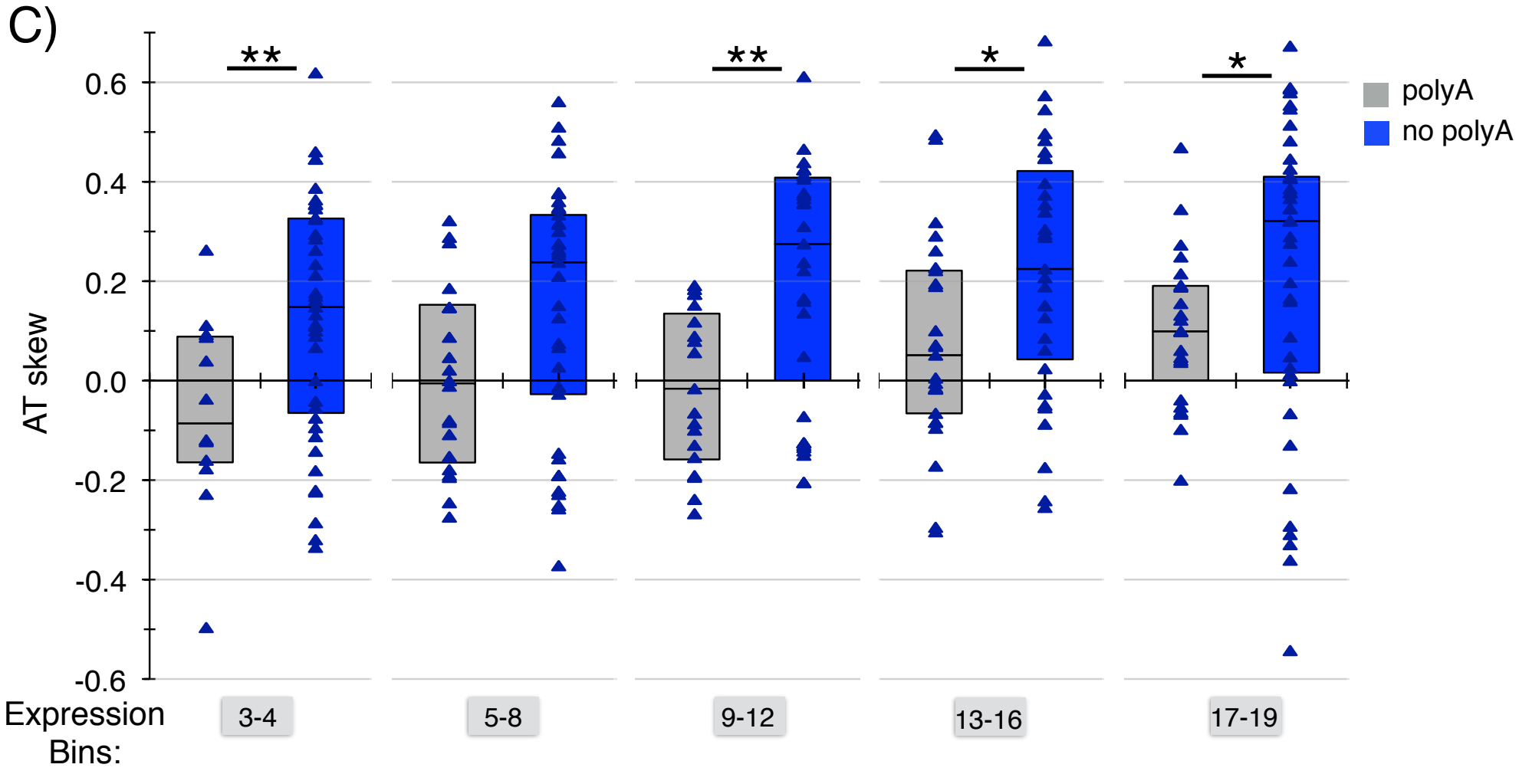


Figure S5.5 polyA mutations, polyA and ATskew

C)



A)

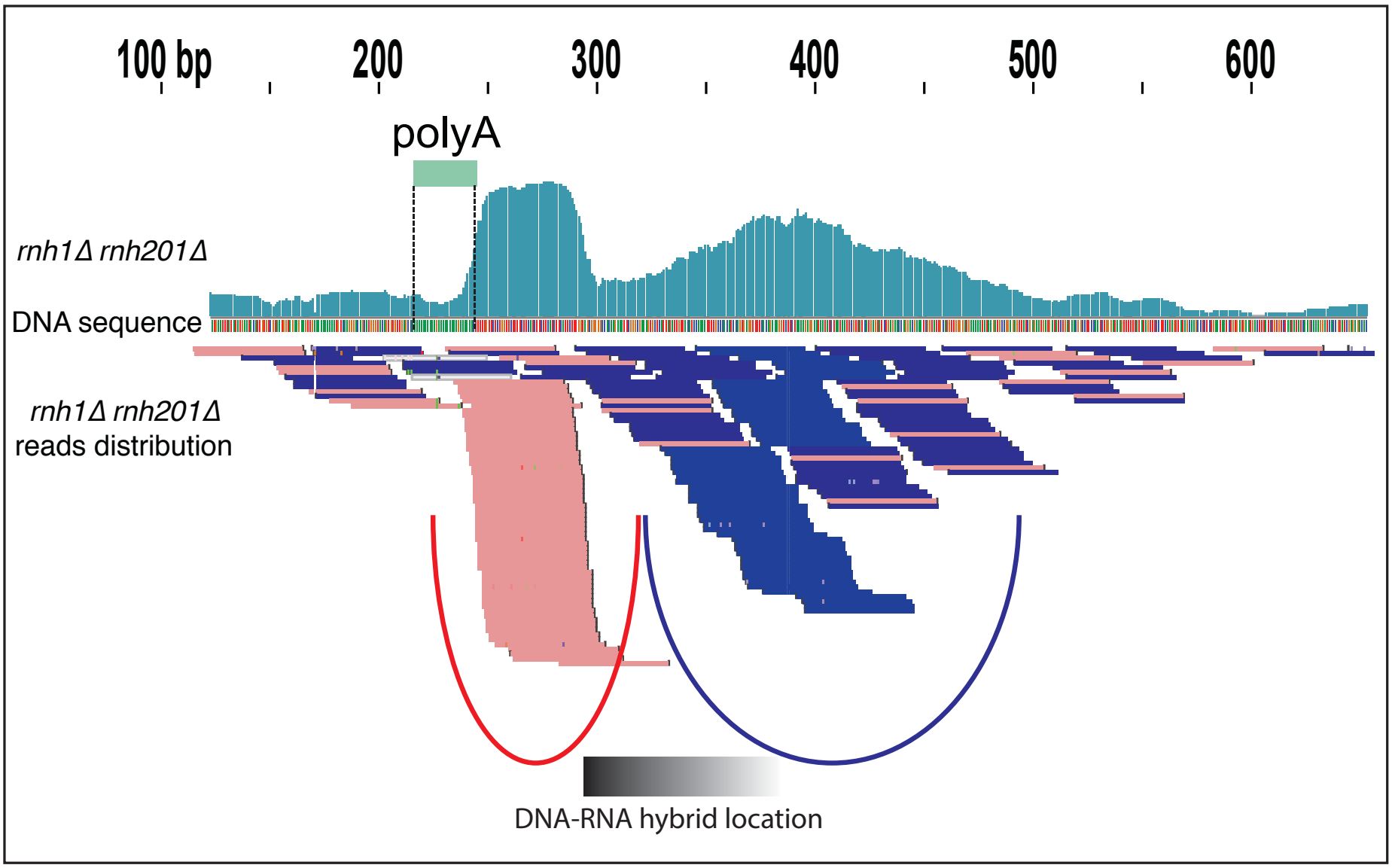


Figure S6.1
polyA
peaks

B)

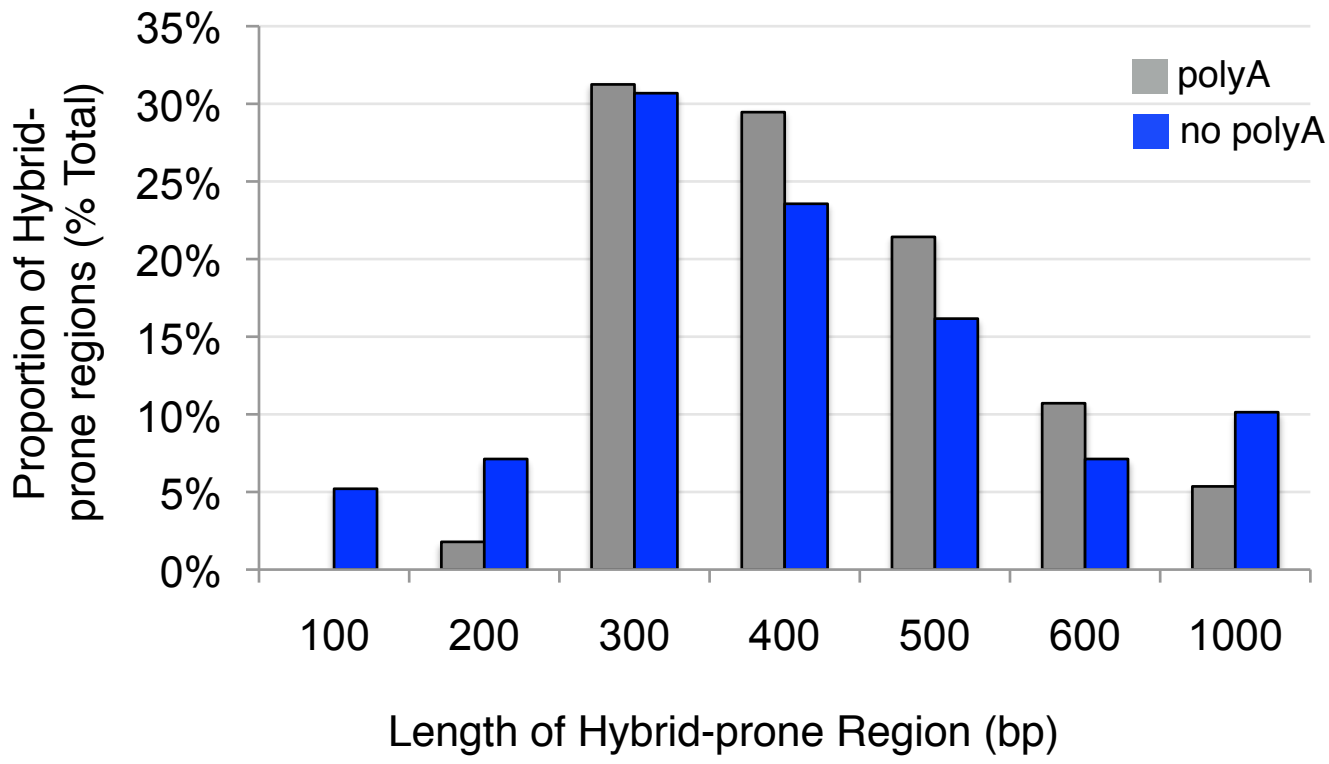
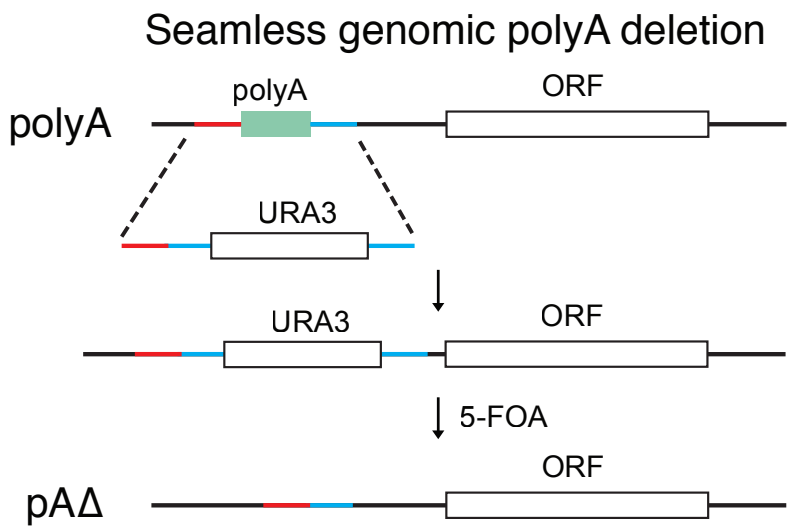
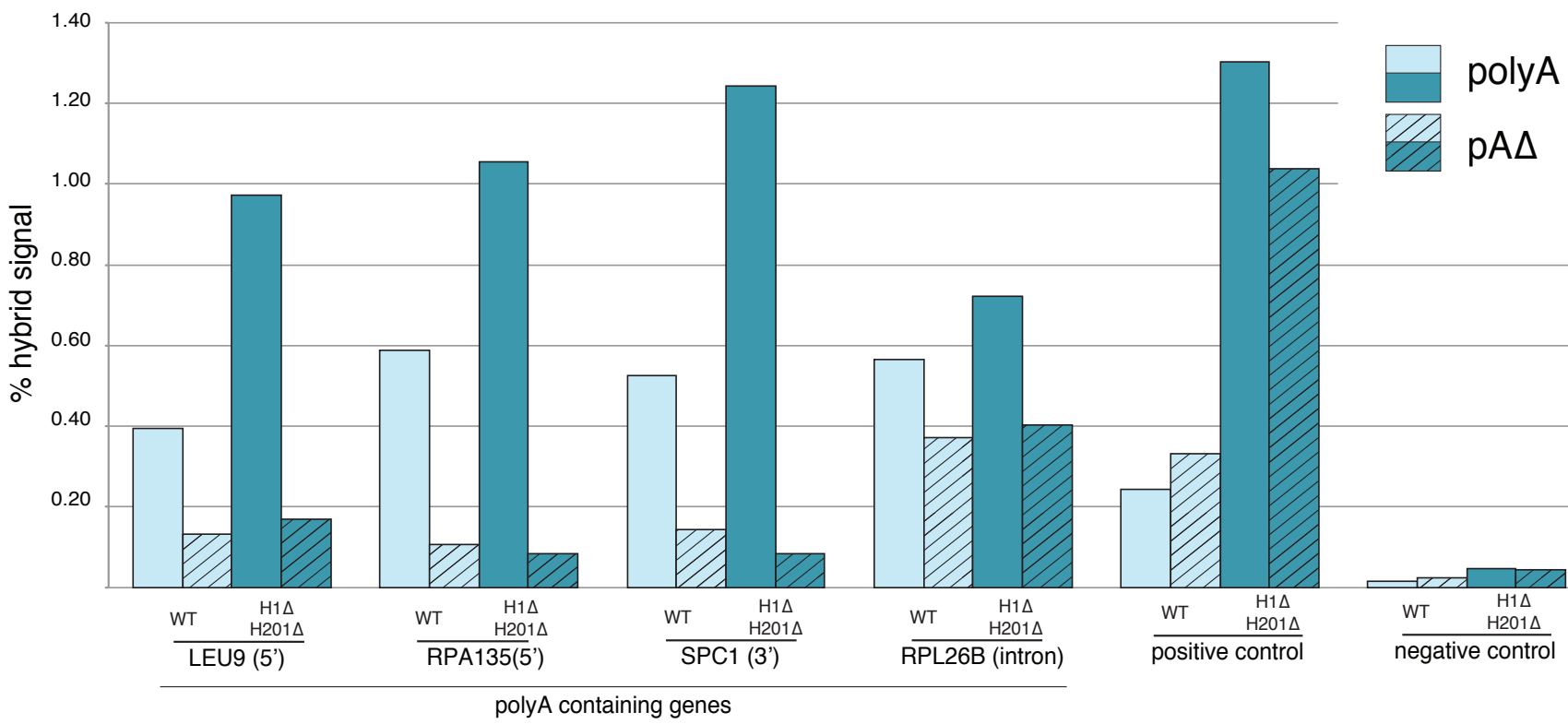


Figure S6.2 polyA deletion

A)



B)



C)