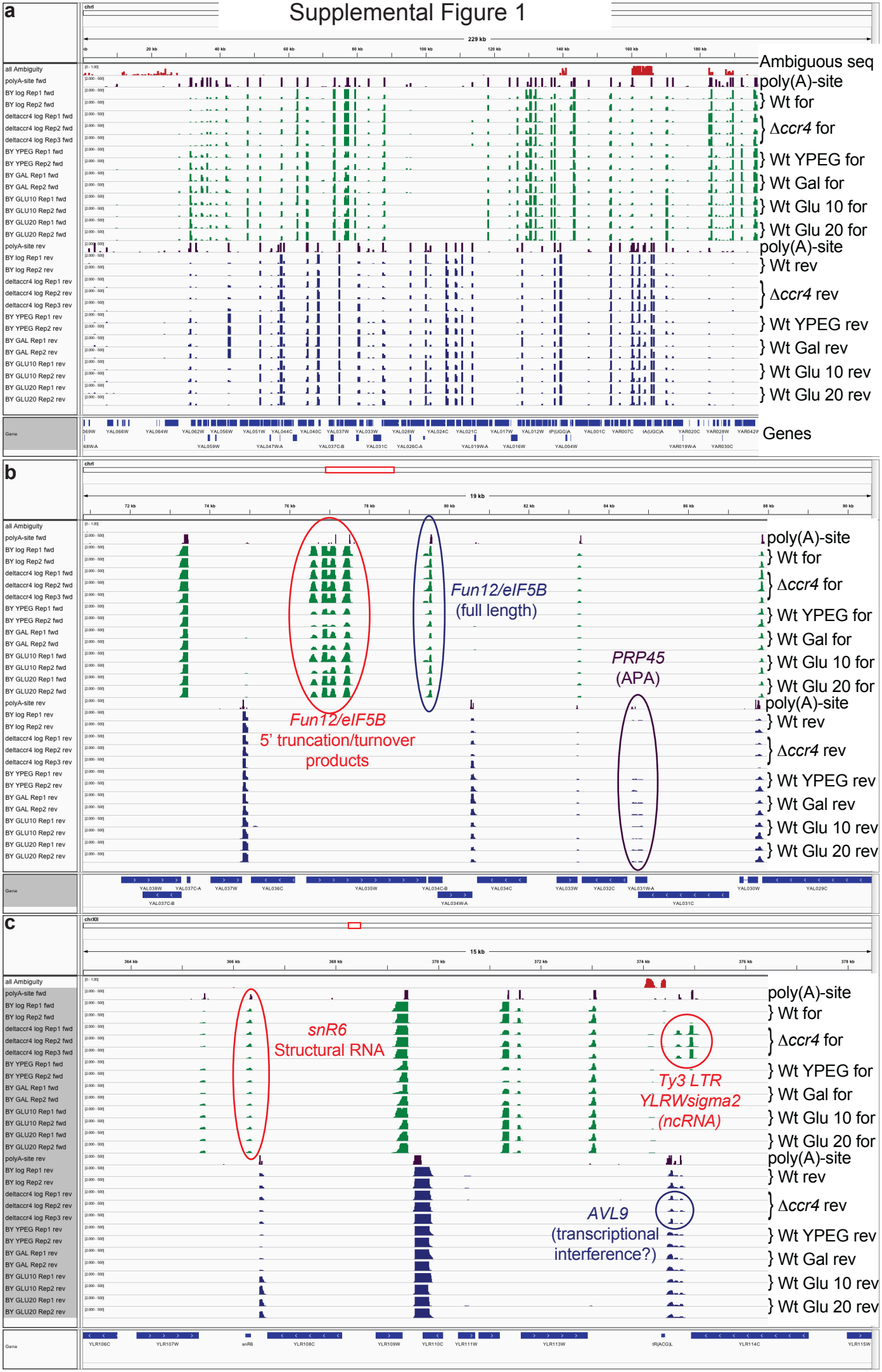
Supplemental Figure 1: Screen shots from the Integrated Genome Browser (IGV) set to display all 13 tracks of PAT-seq data aligned to the Sac/cer3 genome. All samples are in replicate, except Δ*ccr4,* which was sequenced in biological triplicate. In each of the figures the track height is fixed to show a coverage of between 2-500 reads, meaning that the dynamic range is curtailed to the low end of the spectrum. A change in peak height indicates a change in gene expression. Green and blue peaks represent adenylated 3' ends on the Watson and Crick strand respectively. Purple peaks are sites of adenylation calculated based on the presence of at least four non-templated A-residues at the end of a read. When these peaks form doublets or multiples, they represent Alternative PolyAdenylation (APA) or micro-heterogeneity of the polyadenylation site. Finally the red track of ambiguity shows regions to which mapping is not unique, reads mapping such sites are distributed between ambiguous sites.

a) A zoomed-out view of yeast chromosome 3.

b) IGV screen shot to highlight 5' truncation or turnover products (circled in red) associated with the locus encoding the eIF5B (FUN12) translation initiation factor. Note, that these truncation products appear more abundant than the full-length product (circled in blue). Circled in purple is an example of the APA associated with the nearby *PRP45* gene.

c) Structural RNA can become adenylated in by the exosome during decay and thus snR6 the U6 spliceosomal RNA shows adenylated products. The aberrant expression of the Ty3 LTR (YLRWsigma2a) in Δ*ccr4* mutant cells might be directly responsible for the reduced expression (blue circle) of the secretory regulator *AVL9,* the 3'UTR of which is overlapped by the Ty3 LTR expressed on the opposite strand.

Supplemental Figure 1



Supplemental Figure 2:

a) Pearson's correlation between Wilkening et al., (2013) data for 3'T-fill versus RNA-seq. Filtered for a minimum of one read/gene.

Note: Wilkening et, al., (2013) used Spearman's rho, a rank-based correlation in their work. Comparing their and our data using spearman's rho; for 3'T-fill vs RNA-seq we get rho=0.778, for PAT-Seq versus RNA-seq we get rho=0.807. Between 3'T-fill and PAT-Seq we get rho=0.815. The value for rho in Wilkening et, al., (2013) figure 3c (3'T-fill vs RNA-seq) is reported as 0.859 (as compared to our calculation of rho=0.778 using their data). We assume this discrepancy is due to differences in data filtering.

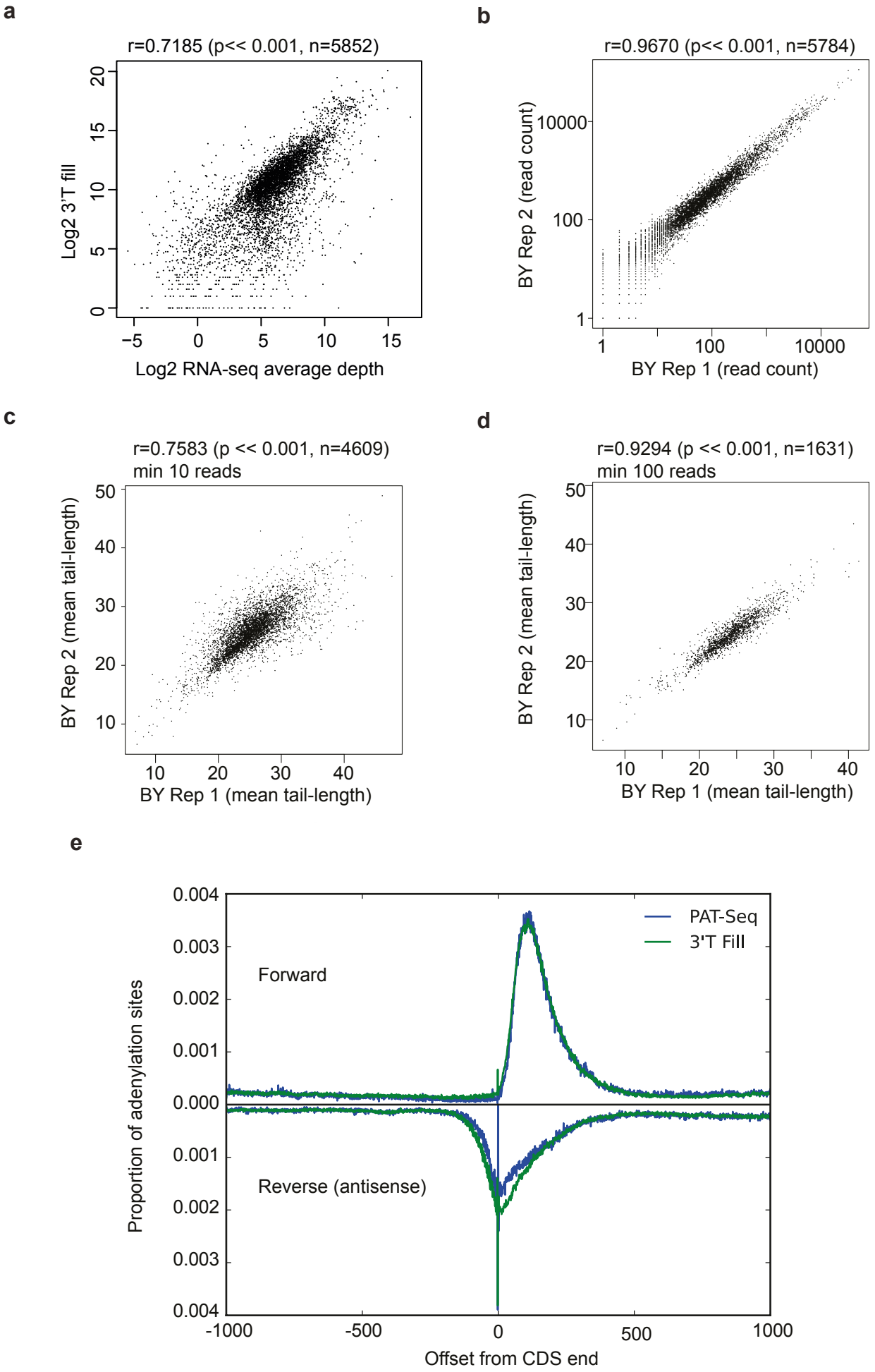
b) The reproducibility of digital gene expression by PAT-Seq is confirmed by the high correlation between BY4741 biological replicates.

c) Replication of tail length measure is also strong, but note that a lower read number can reduce the correlation statistic.

d) Limiting analysis to only genes with >100 adenylated reads, significantly improves confidence in tail-length estimates**.**

d) The positions of adenylation as measured by PAT-Seq (in blue) and 3'T Fill sites (in green) plotted relative to the end of the coding sequence (0). The positions in the forward direction peak at ~100 bases after the stop reflecting the most common length of the yeast 3' UTR. Where coding or non-coding RNA run antisense (reverse) to these positions they are offset by ~100 bases peaking near the stop codon. This figure is equivalent and entirely in accord with Wilkening *et, al.,* (2013) Figure 4b part 2.

Supplemental Figure 2

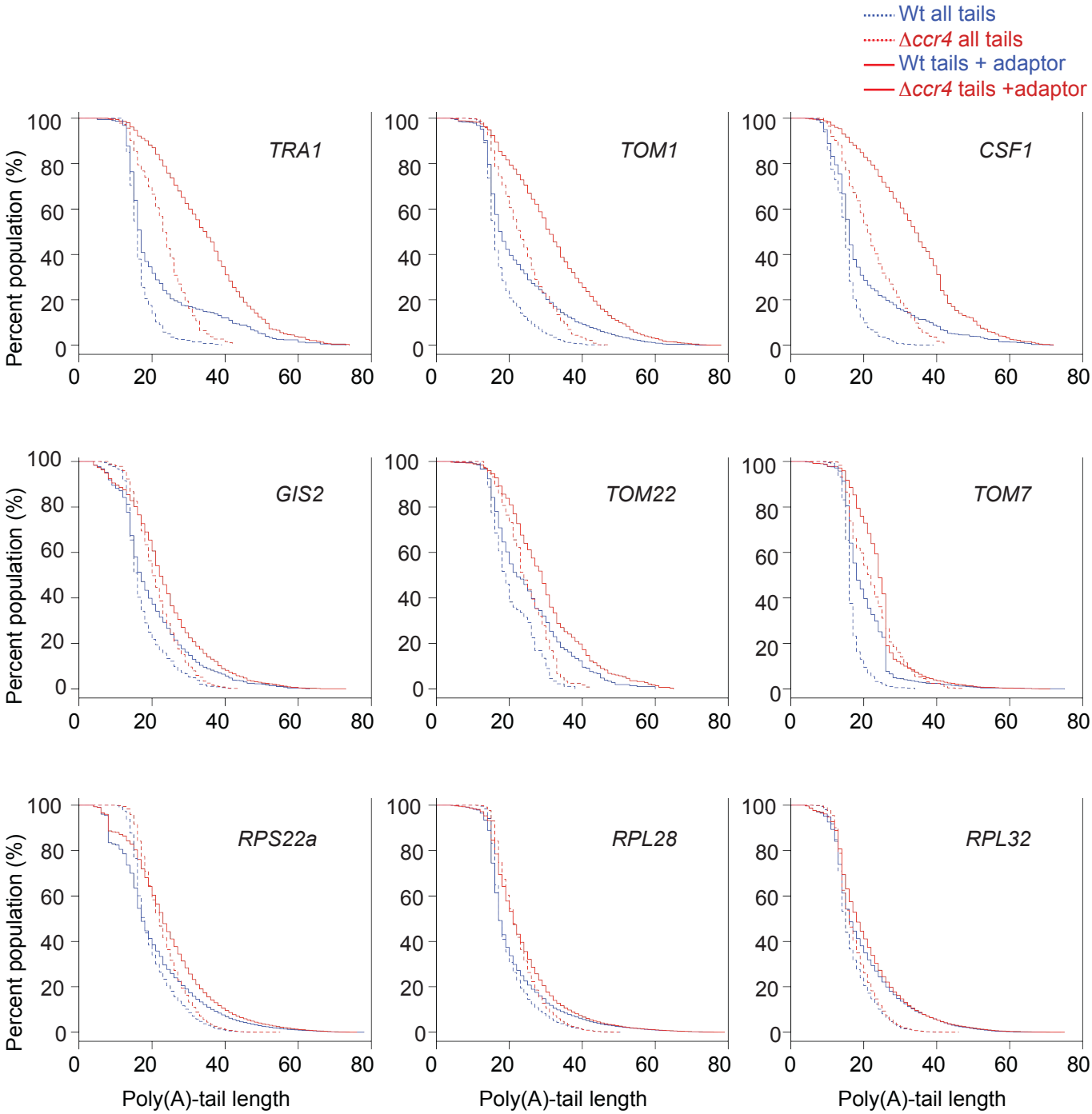


Supplemental Figure 3:

The PAT-seq approach samples poly(A)-distribution but may not always detect the absolute maximal native length. Where the poly(A)-tail is sequenced in its entirety in nearly all reads, the method can be taken as quantitative of the native length-distribution . Where the read-length of the NGS sequencing platform limits the sequencing to less the than the maximal poly(A)-tail the measure of poly(A)-distribution is a qualitative reflection of native length. Shown are inverse cumulative histograms of sequenced poly(A)-tail distributions for a selection of genes as indicated. The proportion of the reads having 1, 2, 3…. 80, terminating A residues are plotted such that the distribution of lengths in the *Δccr4* sample (red) tend to be longer then the distribution for wild-type (blue). Solid lines are used when all data were included. Dashed lines indicate further sub-sampling to those poly(A)-measurements that include the final 3'linker (ie the full native poly(A)-tail was sequenced). Such reads tend to show shorter overall poly(A)-distributions because loss of sequencing register can interfere with detection of the adapter sequence in longer reads (<40 A). However, meaningful adenylation-state differences are observed despite sampling only shorter tails within the distribution. That is, an extended tail is still observed in the *Δccr4* sample. Shown are a selection of genes that were identified as having a statistically significant difference in adenylation between wild-type and the *Δccr4* sample.

It is important to note that even when the method is not quantitative for absolute lengh, it has considerable power to detect differential tail-length between conditions. This can be compared to the trust we place in the ability of RNA-Seq to detect differential expression even though systematic biases in sequencing efficiency can occur, for example due to base composition or the exact manner of fragmentation and size selection.

Supplemental Figure 3



Supplemental Figure 4: a) Little correlation is observed between poly(A)-length and read count, or **b)** protein abundance when the average poly(A)-tail length measurement is used. Note previous work in our laboratory used either weighted lengths or ratio-lengths to show a positive correlation, however both such measures favour abundant transcripts and are biased to the longest reads within a population rather than the average.

c) Comparison of mean Poly(A)-tail lengths as measured by either PAL-seq (estimates for 3943 genes) or PAT-Seq (estimates for 4609 genes). Dashed lines indicate the likely noise introduced by both technologies. The lower limit is set by the requirement of an annealed oligo-dT primer in both methods, the upper limit to PAL-seq is based on the literature which shows an upper limit to absolute tail length of ~80 in yeast and thus the median for any gene is likely significantly below this. In PAT-seq, the measured mean poly(A)-tail length can be an under-representation of the true length in this experiment because the sequenced read length (100 bases) might not capture the longest tails in the population and additional contain enough unique sequence to map to the genome.

<< indicates much more than the stated significance value.

Supplemental Figure 4