Accurate Recovery of Ribosome Positions Reveals Slow Translation of Wobble-Pairing Codons in Yeast

Hao Wang[1], Joel McManus[1,2], and Carl Kingsford[1] ⋆

[1] Computational Biology Department, School of Computer Science,
Carnegie Mellon University, Pittsburgh, Pennsylvania, United States of America
[2] Department of Biological Sciences, Carnegie Mellon University,
Pittsburgh, Pennsylvania, United States of America

⋆ To whom correspondence should be addressed. Email: carlk@cs.cmu.edu

Appendix

S1 One example of the complex digestion pattern observed in ribo-seq data

Ribosome footprint reads are not always highly concentrated on a single reading frame for all read lengths. For example, in a recently published yeast ribosome profiling data with deep coverage [1], although 96% of the reads with length 28 are skewed towards one frame, the highest frame portion for reads with a length not equal to 28 vary from 60% to 80% (Figure S1). Further, many ribosome reads do not have a typical footprint size. For the experiment above, fewer than 60% of uniquely mapped reads have length 28. While it is possible for reads from the other two frames to be caused by frameshifts, frameshifts are generally rare and thus cannot explain all of the off-frame reads. It appears that the simplified A-site offset assumption is unable to explain the observed reading frame patterns.

S2 Ribosome profiles are well explained by the blur model

We test whether the estimated blur vectors can truthfully characterize the observed read locations on the meta-profiles. Here, different blur vectors are convolved with the initial guess of the meta consensus — the in-frame values of $M_{obs}(28)$. Our modeled meta-profiles agree well with the observed meta-profiles. The blur process results in a good correlation and a small deviation between the modeled meta-profiles and the observed meta-profiles across all read lengths (Figure S3).

To test whether a single blur vector is sufficient to model all profiles for a given length, we train the blur vector on subsets of locations and on subsets of transcripts, and we get similar results compared to training the blur vector on the entire set (Figure S3). This indicates that the blur vectors are transcript and location independent.

In addition, allowing the length-specific clear position signals to be slightly deviated from the consensus further improves our model fitting. Instead of enforcing an identical consensus across all read lengths, these deviations result in both better modeled meta-profiles (Figure S4), and better modeled transcript profiles (Figure S5): the inconsistencies between the observed profiles and the modeled profiles are reduced by an average of 66% compared to not allowing such deviations.

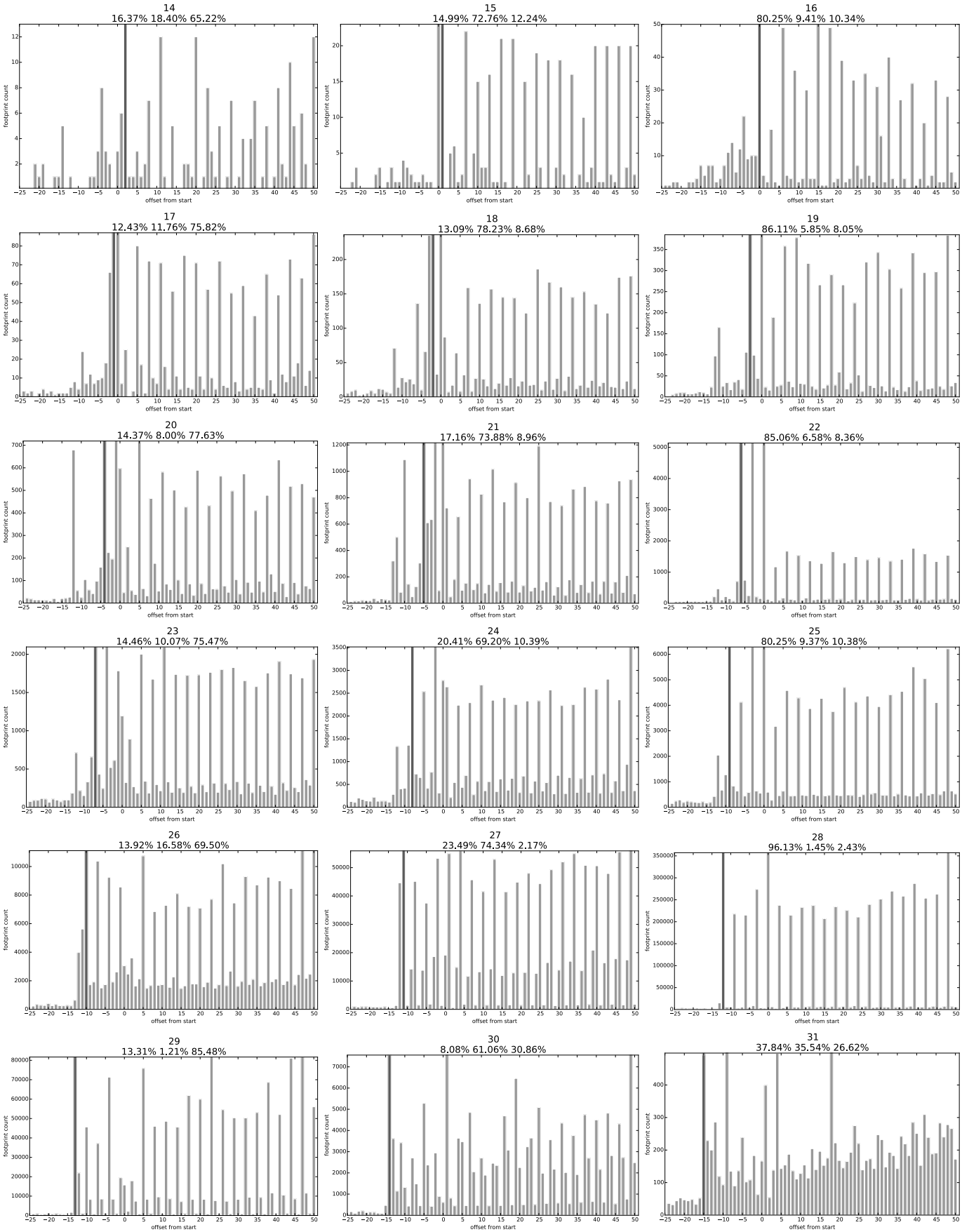
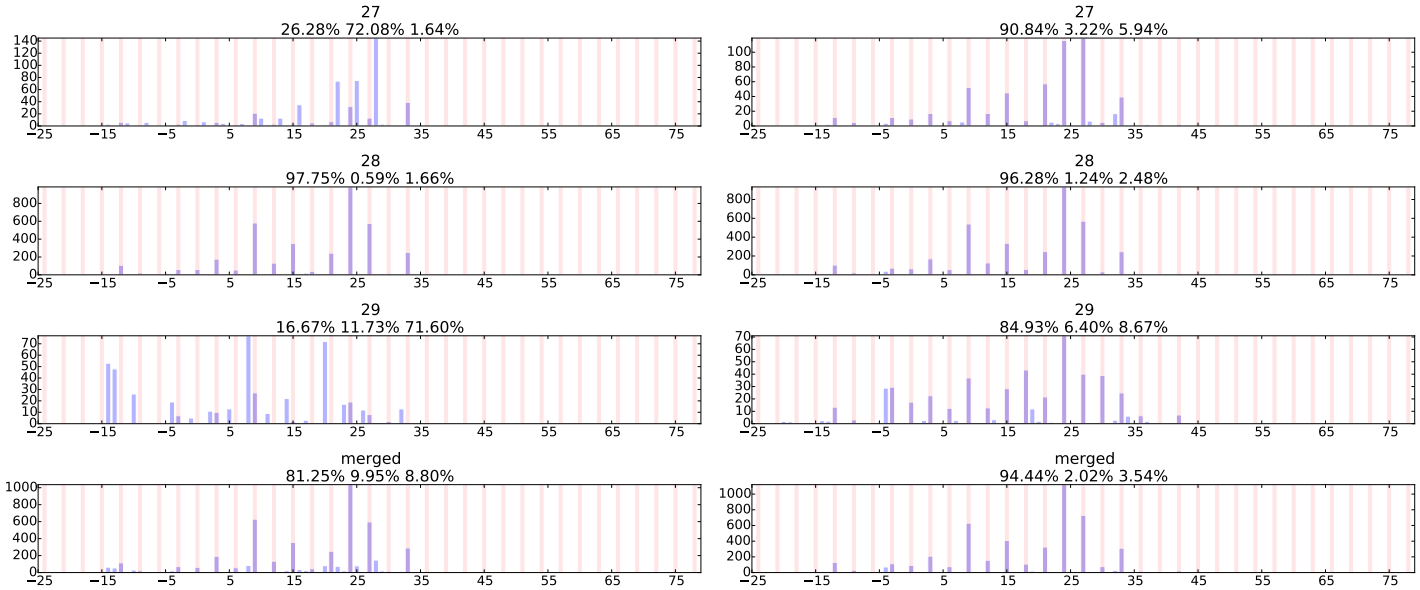


Fig. S1. Meta-profiles on reads with different lengths near the start codon. These profiles are generated by accumulating reads for a given length from all transcripts on a location. The darker line in each profile is the starting point of the 3-nt periodicity. Reads are divided into 3 frames corresponding to the $1^{st}, 2^{nd}$ and 3^{rd} nucleotide of codons, and read length and the distribution of these frames are noted above each profile. (Data is from GSM1335348 [1]). Reads with various lengths show a unique shape of 3-nt periodicity. That is, for each of the 3 bases within a codon, the order of the read abundance will often persist across codon locations. For example, if the first frame is the most abundant frame for the first codon, it is also likely to be the most abundant frame for subsequent codons.



(a) before deblur (b) after deblur

Fig. S2. One example of the ribosome profiles for different read lengths before and after debluring on transcript YOR302W. Read length and frame distribution are noted above each profile. In-frame (frame 0) loci are marked with light vertical lines, and read pileups are marked with dark vertical bars. The clear position signal is assumed to be consistent across read lengths, but slight shifts and deviations are allowed.

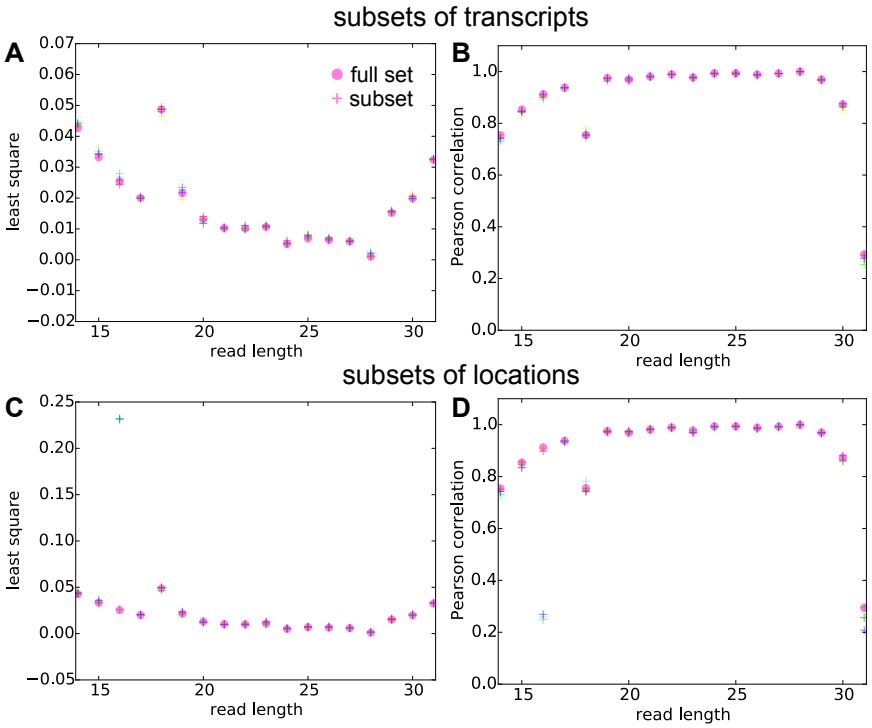


Fig. S3. The agreement between the observed meta-profiles and the modeled meta-profiles with blur vectors trained on subsamples of data. (A, B) The transcripts are randomly divided into five groups with equal size, and the blur vectors are trained on the subset of transcripts. (C, D) The profile locations are randomly divided into five groups with equal sizes, and the blur vectors are trained on the subset of locations. The least square and the Pearson correlation are between the the observed meta-profiles ($M_{obs}(l)$) and the modeled meta-profiles ($\widehat{M_{obs}}(l) = M_{init} * b(l)$) for each read lengths l. Results are qualitatively similar regardless of whether the blur vectors are trained on a subsample of the data. Read length 18 is modeled slightly worse primarily due to low coverage and significant outliers.

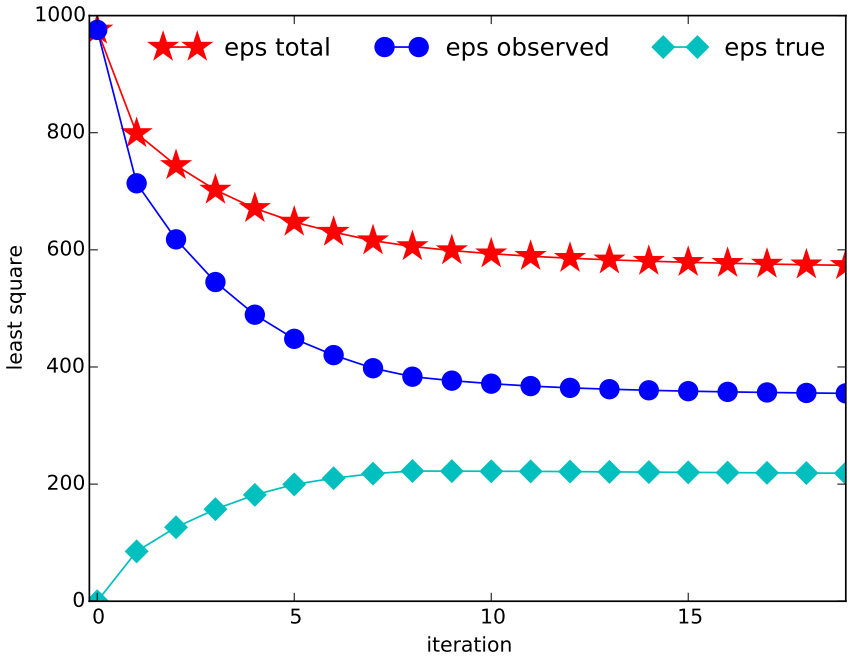


Fig. S4. The change of the objective function in equation (1) for deblurring the meta-profiles. The circle line is the overall inconsistency between the modeled meta-profile ($\widehat{M}_{obs}(l)$) and the observed meta-profile ($M_{obs}(l)$), The diamond line is the overall inconsistency between the consensus clear signal (M^{kl}_{true}) and the read-length-specific deblurred signal ($M_{true}(l)$), and the star line is the sum of the two. The overall inconsistency between the modeled meta-profiles and the observed meta-profiles successfully goes down during the optimization, and the observed profiles are better modeled by sacrificing the inconsistencies of clear signals across different lengths.

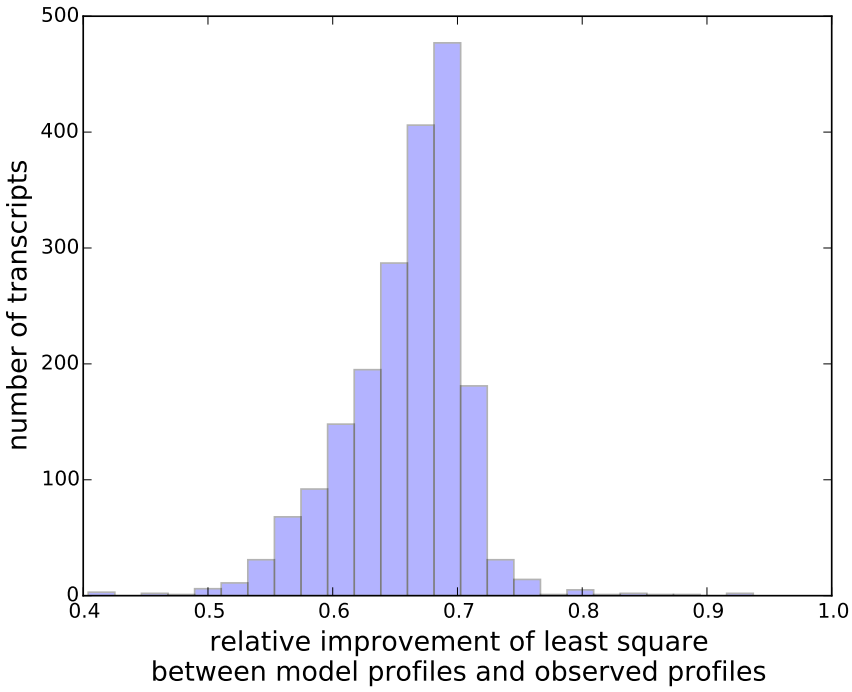


Fig. S5. Histogram of the relative improvement of the overall inconsistency between the observed profiles P_{obs} and the modeled profiles $\widehat{P}_{obs}$ after the deblur procedure. The relative improvement is defined as the change of the inconsistency between P_{obs} and $\widehat{P}_{obs}$, over the initial inconsistency between the two.

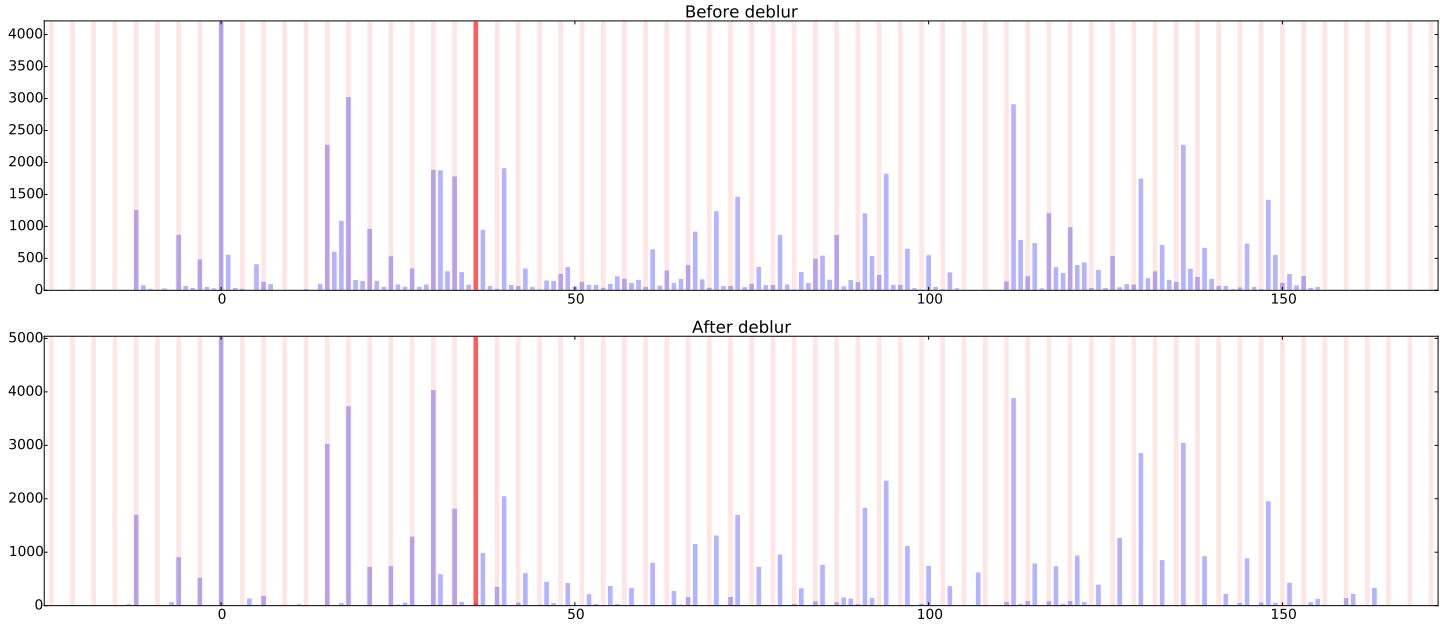


Fig. S6. An example of the ribosome profiles with artificially programmed frameshift on transcript YDL061C before and after deblur. Frame-0 loci are marked with pink vertical lines and the frameshift point is marked with a darker vertical line. The high frame-0 skewness before the frameshift point and the high frame-1 skewness after the frameshift point are well maintained and strengthened by the deblur process.