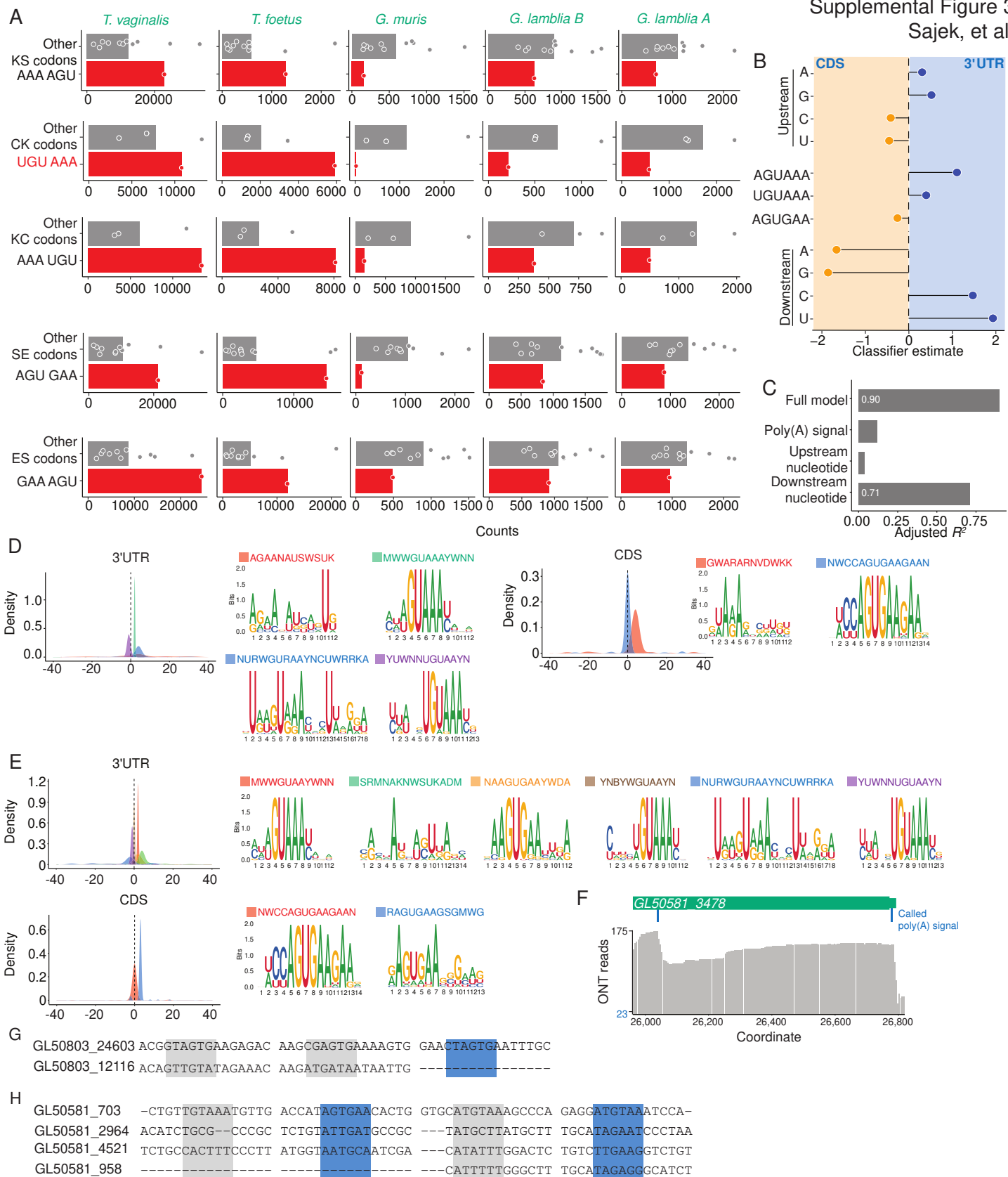


Supplemental Figure 3  
Sajek, et al.



**Supplemental Figure S3. Role of flanking nucleotides in WGURAA hexamers recognition in *G. lamblia A* and *G. lamblia B* (related to Figure 4).** **A.** Occurrence of AAA AGU, UGU AAA, AAA UGU, AGU GAA and GAA UGU di-codons in comparison to other Lys-Ser, Cys-Lys, Lys-Cys, Ser-Glu and Glu-Ser encoding combinations. **B.** Beta-coefficient values from the linear model applied to WGURAA-containing *k*-mers scores from gkmSVM classifier, corresponding to the upstream nucleotide, poly(A) signal, and downstream nucleotide in *G. lamblia B*. **C.** Variance explained by the linear model applied to WGURAA-containing *k*-mers scores from gkmSVM classifier by the full model, upstream nucleotide, poly(A) signal, and downstream nucleotide in *G. lamblia B*. **D,E.** The position and identity of sequence motifs obtained by merging the most predictive *k*-mers from trained gkmSVM models for *G. lamblia A* (D) and *G. lamblia B*. Motifs are shown as sequence logos. density plots represent their distribution across 3'UTR (upper panel) or CDS (lower panel) 80-mers. **F.** Example of *G. lamblia B* gene (GL50581\_3478) where a hexamer in the coding sequence was misclassified by the gkmSVM model. Premature cleavage after the hexamer in the CDS indicated as coverage drop. **G,H.** Example of gene duplication events in *G. lamblia A* (G) and *G. lamblia B*, where one of the genes contain WGURAA hexamer(s) in coding sequence and is prematurely cleaved. WGURAA hexamers preceding premature cleavage sites are highlighted blue.