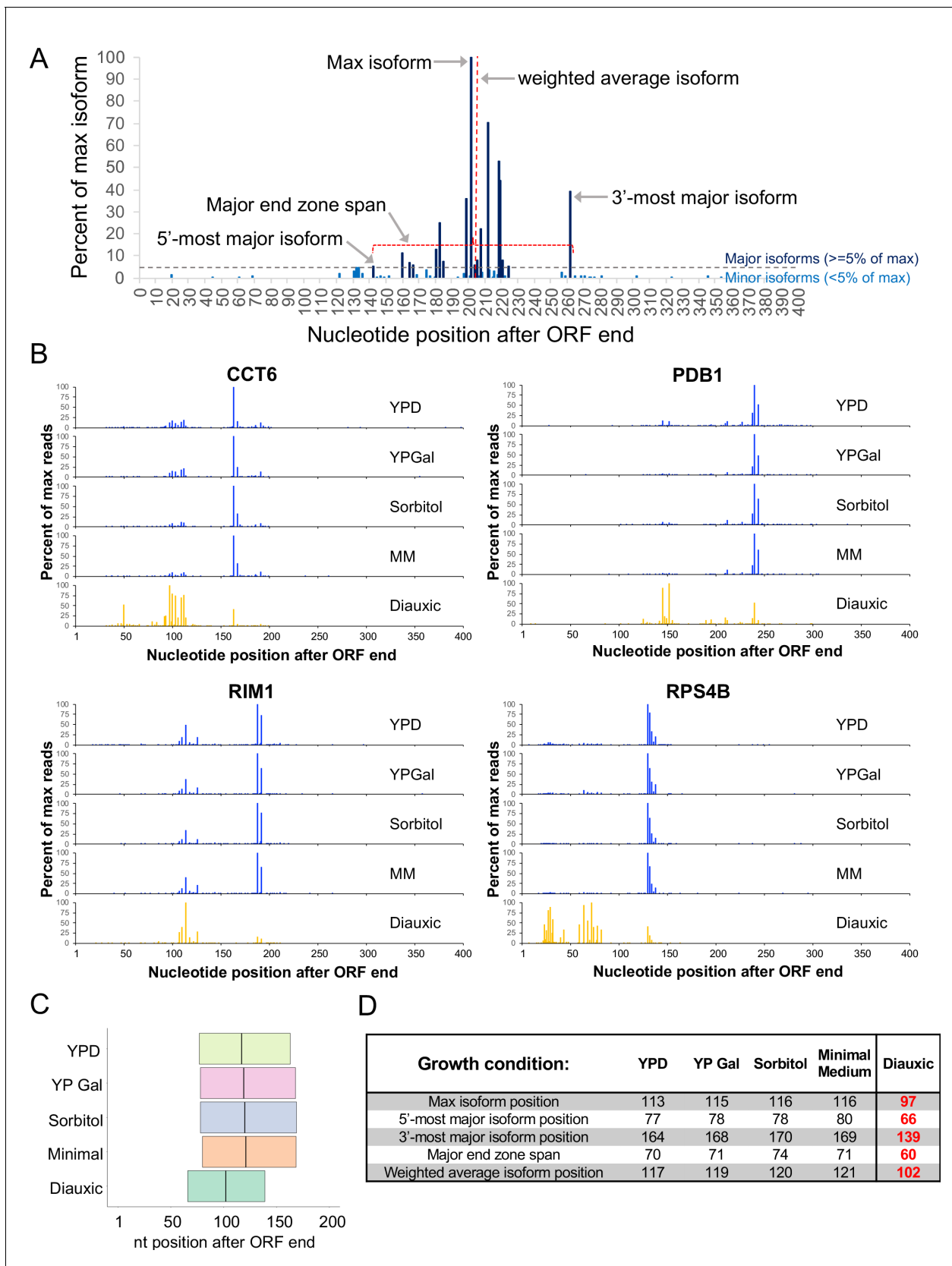# Figures and figure supplements

The transcriptional elongation rate regulates alternative polyadenylation in yeast

**Joseph V Geisberg** *et al*

Figure 1. Poly(A) sites are shifted upstream in diauxic cells. (**A**) Representative end zone profile (histogram of isoform frequencies) with key landmarks indicated. (**B**) End zone profiles for four genes under five growth conditions. (**C**) Major end zone under five growth conditions. Boundaries represent

Figure 1 continued on next page
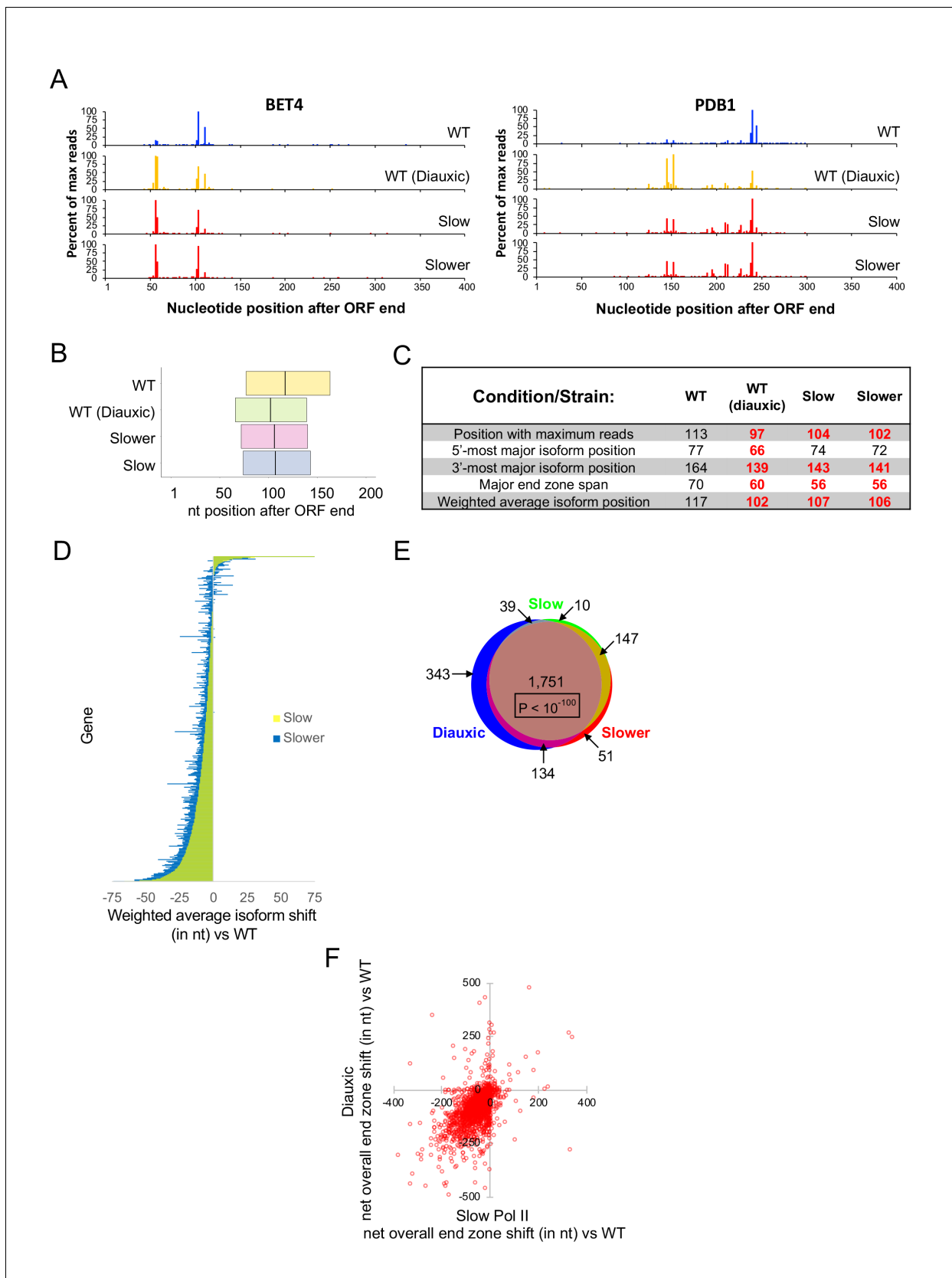
Figure 1 continued

median values genome-wide for 5'-most and 3'-most major isoforms, and the vertical line within the major end zone represents the genome-wide median of the weighted average isoform position. (D) Table of statistics for landmark positions under five growth conditions. Numbers are the median values across genes with a combined read count of at least 1000 in both replicates in every condition. Numbers in bold red are shifted upstream from WT in a statistically meaningful way (p < 0.01).
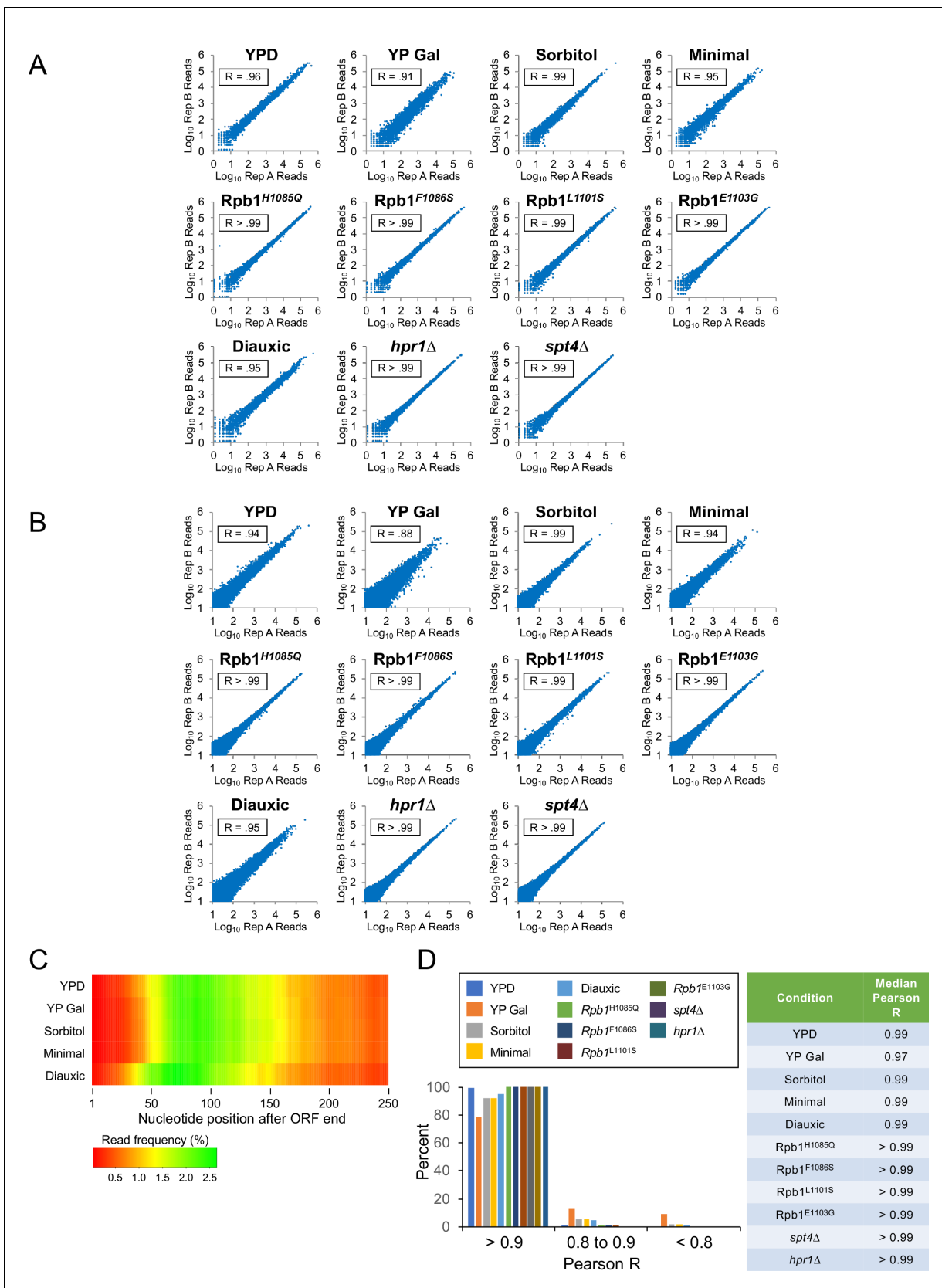
Figure 1—figure supplement 1. Correlation of biological replicates. (**A**) Correlation of reads per gene in two biological replicates. Each plot shows all genes with ≥1000 reads in the two conditions (>5000 genes/plot). (**B**) Correlation of reads per isoform in two biological replicates. The plot for each
Figure 1—figure supplement 1 continued on next page

Figure 1—figure supplement 1 continued

condition shows 25,000–75,000 isoforms. Isoforms with fewer than 10 reads were omitted. (C) Correlations of overall isoform profiles across biological replicates. Pearson correlations were determined for isoform profiles of the 2790 genes with $\geq$1000 total reads (in both replicates combined) in every experimental condition. Left: Distribution of isoform profile correlations for genes in every condition. Right: Genome-wide median correlations of isoform profiles across biological replicates. (D) Heat map of 3'UTR coordinate use in five growth conditions. For every condition, the heat map depicts the fraction of genes in the genome for which each 3' UTR position is used as a poly(A) site. In this plot, the relative use of poly(A) sites at individual genes is not considered; rather, the result for a given position at any gene's 3' UTR is binary. A position is counted as positive for the gene if any isoform endpoints map there.

Figure 2. Slow Pol II and diauxic end zones are highly similar. (**A**) End zone profiles for *PDB1* and *BET4* in strains harboring wild-type Rpb1 (in exponential and diauxic growth conditions), Rbp1 H1085Q ('slower'), and Rpb1 F1086S ('slow'). (**B**) Major end zones of these strains. Boundaries
*Figure 2 continued on next page*

Figure 2 continued

represent median values genome-wide for 5'-most and 3'-most major isoforms, and the vertical line within the major end zone represents the genome-wide median of the weighted average isoform position. (C) Table of statistics for landmark positions. Numbers are the median values across genes with a total of at least 1000 sequence reads in both replicates in every condition. Bold red numbers are shifted upstream vs WT in a statistically meaningful way (p < 0.01). (D) Bar graph representation of each gene's net shift in weighted average isoform position in strains with slow vs wild-type Rpb1. Each horizontal line represents one gene, ordered by shift values in the 'slow' strain; the graph includes 3497 genes with a combined read count of at least 1000 for both replicates in all three strains. Yellow bars represent the 'slow' strain, and blue is for the 'slower' strain; overlapping bars appear green. To obtain net shift values for every gene in each mutant strain, the average shift vs WT in two replicates was diminished by the absolute value of the average shift of the WT and mutant biological replicates. The net shift was set to zero if the absolute value of the shift vs WT was less than the absolute value of the shift between biological replicates. (E) Venn diagram overlap of genes categorized as upshifted in the diauxic condition, slower Rpb1 (H1085Q), or slow Rpb1 (F1086S) strains. (F) Correlation of end zone shifts in diauxic and slow Pol II strains. The average net overall end zone shift in slow Pol II strains (x-axis; see Materials and methods) is plotted against the net overall end zone shift in diauxic cells (y-axis). Negative values represent upstream shifts, and positive values indicate downstream end zone shifts.
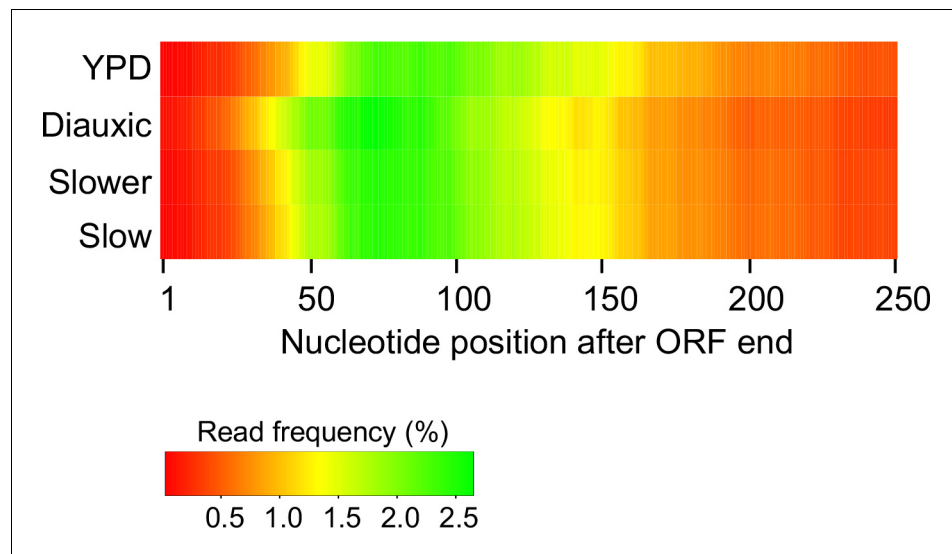
Figure 2—figure supplement 1. Heat map of percent coordinate utilization in 3'UTRs. Heat map of 3'UTR coordinate use in wild-type Pol II (in exponential and diauxic conditions), and two slow-Pol II strains (Rpb1 H1085Q, 'slower' and Rpb1 F1086, 'slow'). For every condition, the heat map depicts the fraction of genes in the genome for which each 3' UTR position is used as a poly(A) site.
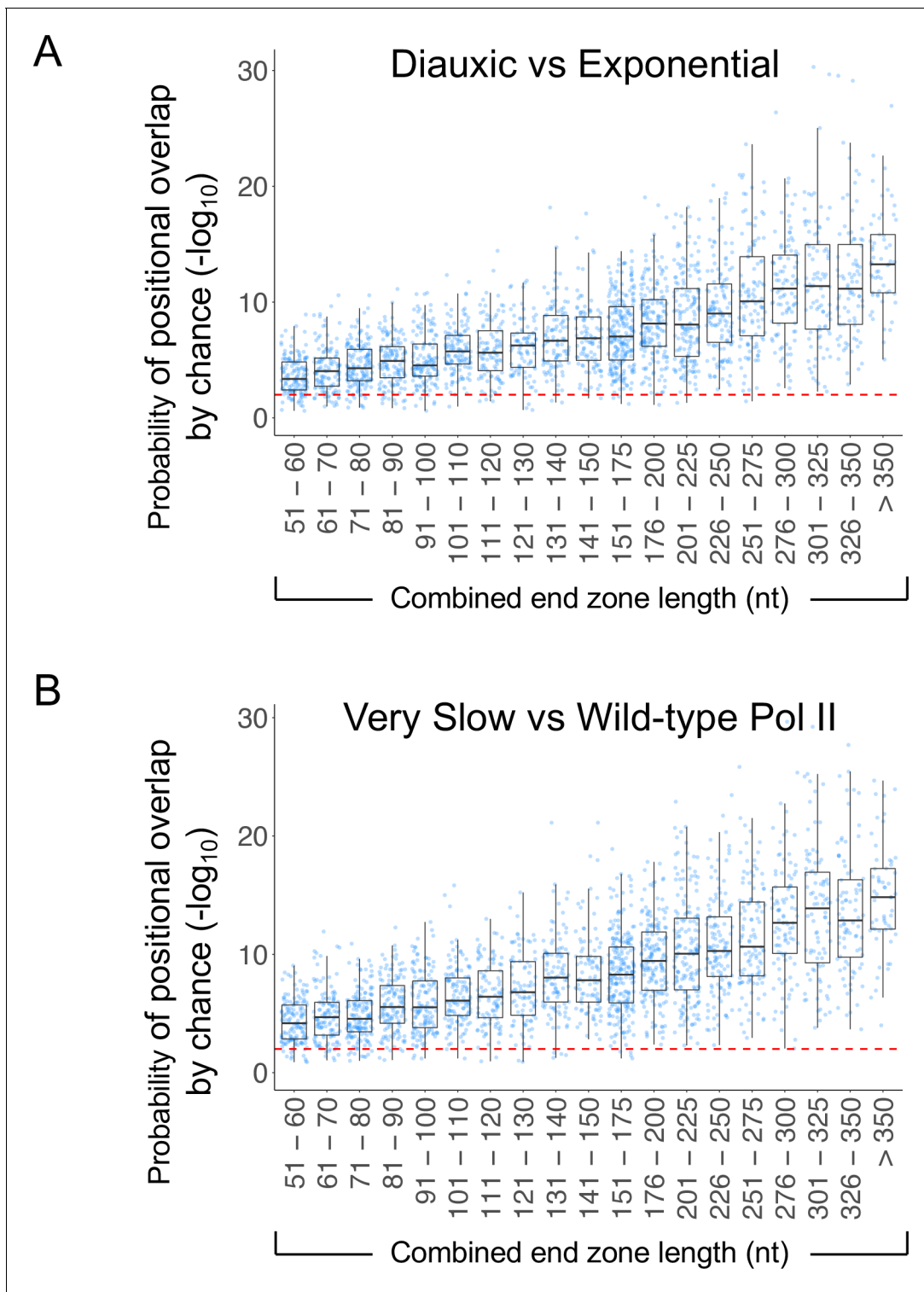
Figure 3. High overlap in poly(A) sites used in diauxic and slow-Pol II strains. (**A**) Probability of overlap in isoform distribution by chance as a function of combined end zone length in strains with very slow (H1085Q) or wild-type Rpb1. (**B**) Probability of overlap in isoform distribution by chance as a function of combined end zone length in exponential growth and diauxic conditions.
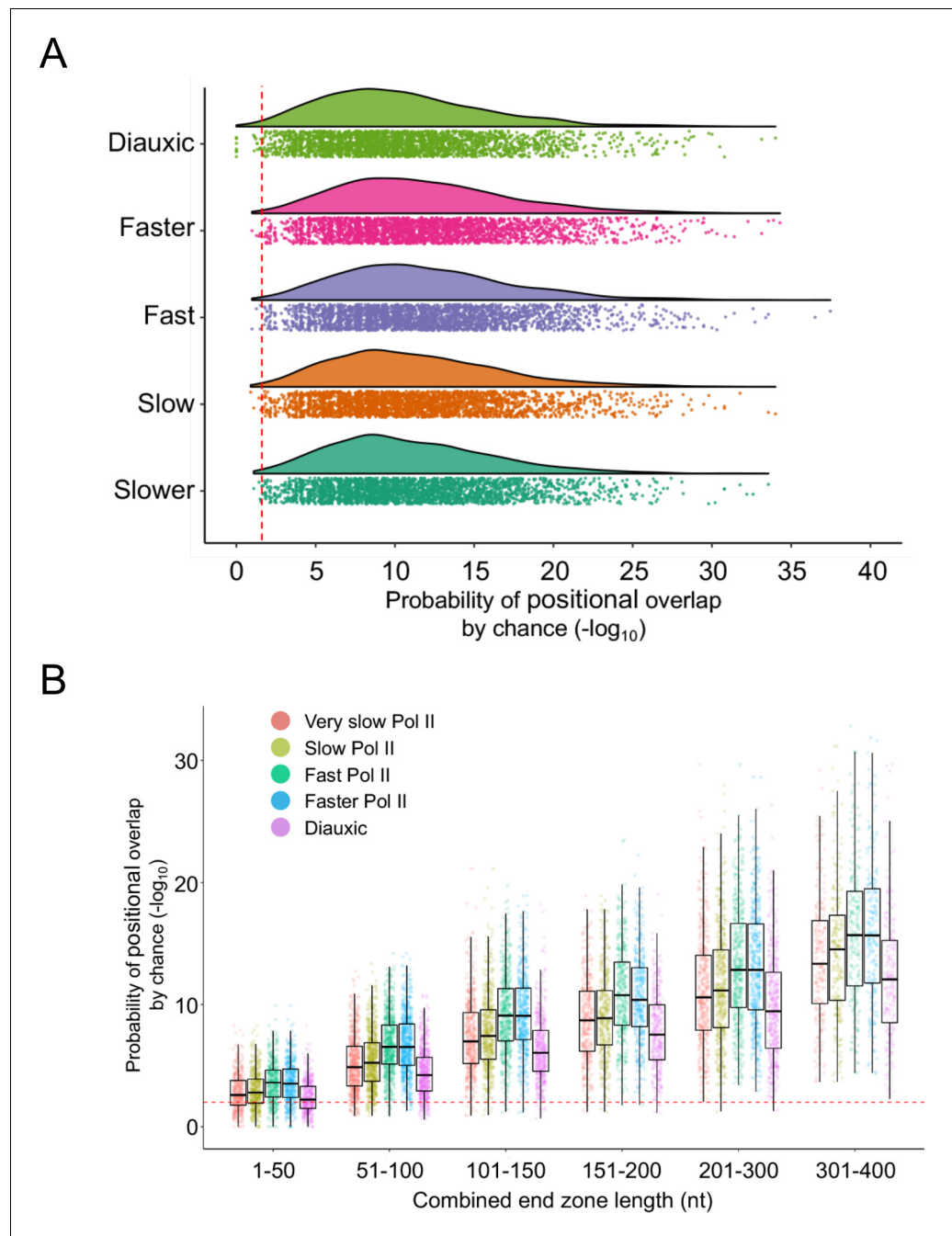
Figure 3—figure supplement 1. High poly(A) site overlap across strains/conditions despite differences in relative levels. (**A**) Conservation of isoform endpoints across strains/conditions. The probabilities of positional overlap by chance were computed using the assumption that all non-A positions in the entire region from 1 to 400 nt after the ORF end are eligible to be isoform endpoints. (**B**) Conservation of isoform endpoints across strains/conditions with various combined end zone lengths. The probabilities of positional overlap by chance were computed using the assumption that possible poly(A) sites are confined to all non-A positions within each gene's combined end zone length.
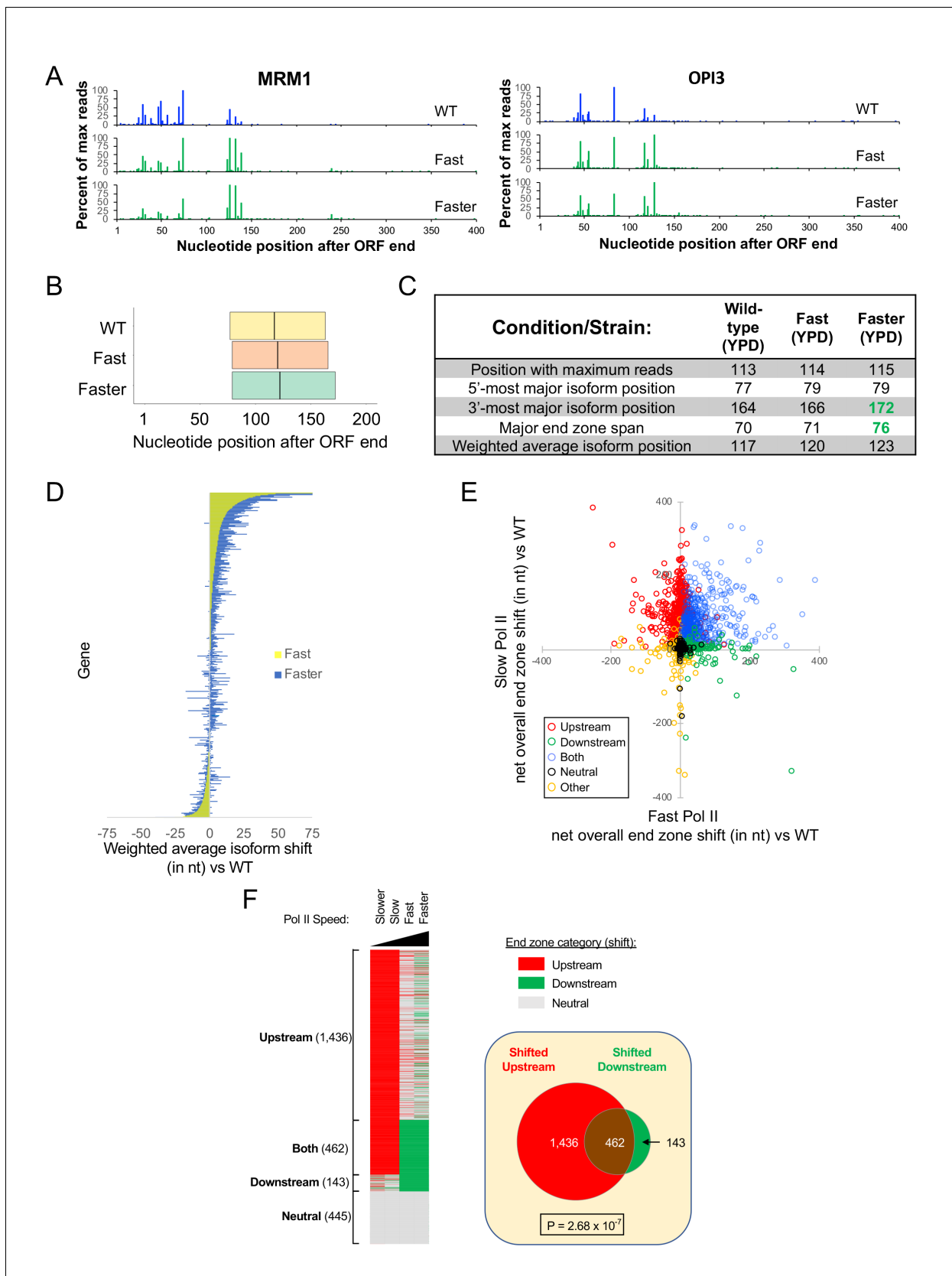
Figure 4. Increased usage of downstream poly(A) sites in fast Pol II strains. (**A**) End zone profiles for *MRM1* and *OPI3* in strains with wild-type, L1101S ('fast'), and E1103G ('faster') Rpb1. (**B**) Major end zones of these strains. Boundaries represent median values genome-wide for 5'-most and 3'-most

Figure 4 continued on next page

Figure 4 continued

major isoforms, and the vertical line within the major end zone represents the genome-wide median of the weighted average isoform position. (**C**) Table of statistics for landmark positions. Numbers aremedian values across genes with a total of at least 1000 sequence reads in both replicates in every condition. Numbers in bold green are significantly shifted downstream from WT (p < 0.01). (**D**) Bar graph representation of each gene's net shift in weighted average isoform position in strains with fast vs wild-type Rpb1. Each horizontal line represents one gene, ordered by shift values in the 'fast' strain; the graph includes 3627 genes with a combined read count of at least 1000 for both replicates in all three strains. Yellow represents the 'fast' strain and blue the 'faster' strain, with the overlap appearing green. To obtain net shift values for every gene in each mutant strain, the average shift vs WT in two replicates was diminished by the absolute value of the average shift of the WT and mutant biological replicates. The net shift was set to zero if the absolute value of the shift vs WT was less than the absolute value of the shift between biological replicates. (**E**) 2790 genes are plotted as a function of the average overall net end zone shift (see Materials and methods) in either catalytically fast (x-axis) or slow (y-axis) Pol II mutants. Genes were classified into Upstream (red), Downstream (green), Both (blue), Neutral (black) and Other (orange) on the basis of each gene's net end zone shift (see text). The upper right-hand quadrant comprises genes shifted upstream in slow Pol II mutants and downstream in fast Pol II mutants, while genes in the upper left-hand quadrant are shifted upstream in both fast and slow Pol II mutant strains. The bottom right quadrant contains genes that are shifted downstream in both slow and fast Pol II mutants, while the few genes whose end zones are shifted downstream in slow Pol II and upstream in fast Pol II strains are found in the bottom left quadrant. (**F**) Left: Classification of genes by category. The categories are: 'Upstream,' genes whose poly (A) sites were upshifted in both slow-Pol II strains; 'Downstream,' genes whose end zone profiles were downshifted in both fast-Pol II strains; 'Neutral,' genes with no end zone shift in any slow or fast Pol II-containing strain; and 'Other,' genes with any other combination of properties (see Materials and methods). Right: Venn diagram illustrating the 'Both' sub-category of genes (see Materials and Methods), i.e. the intersection of the set of genes shifted upstream in slow Pol II (Upstream category) with the set of genes shifted downstream in the presence of fast Pol II (Downstream category).
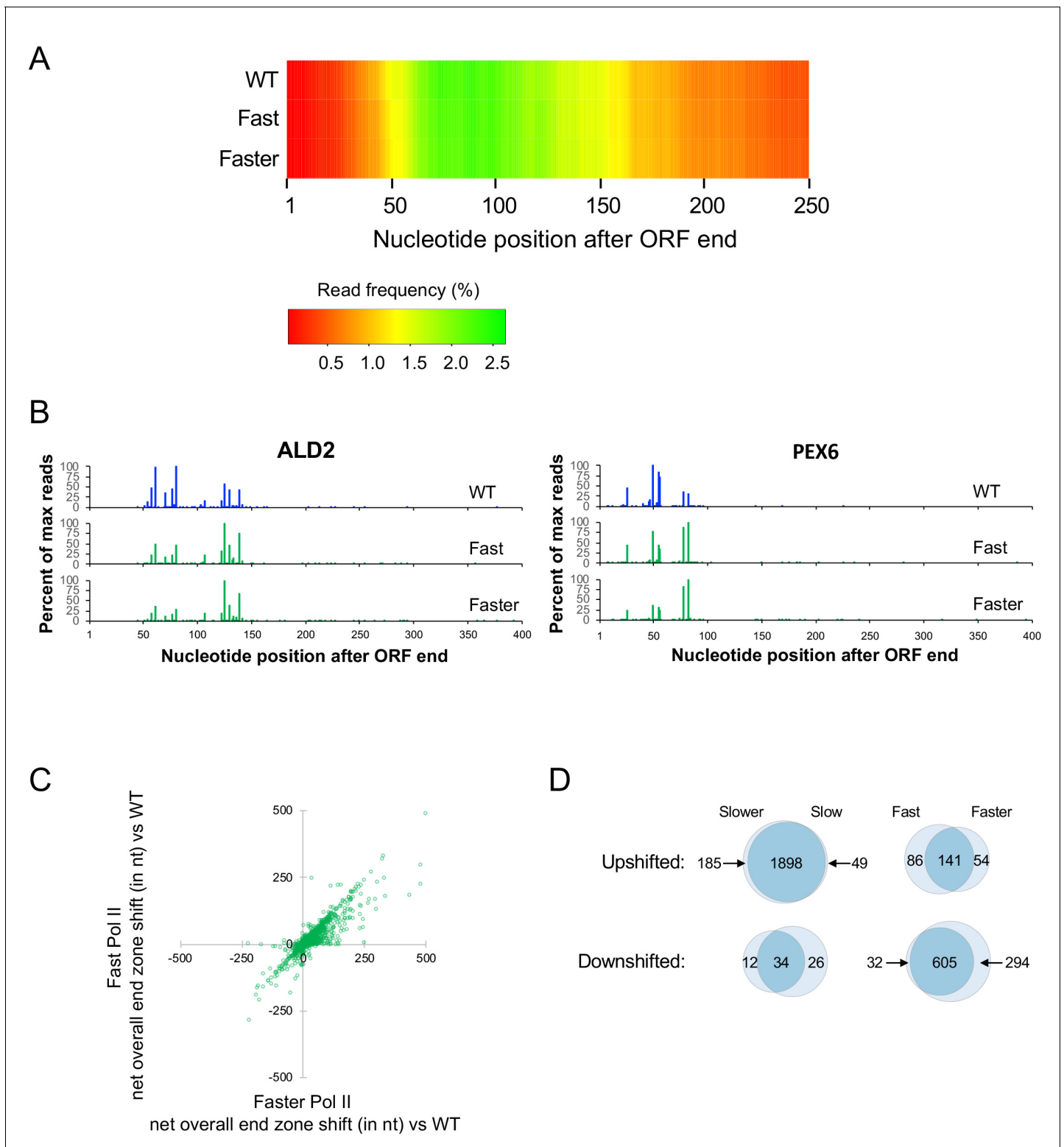
Figure 4—figure supplement 1. Downstream end zone shift in fast Pol II strains. (**A**) Heat map of 3'UTR coordinate use in wild-type Pol II and two fast-Pol II strains (Rpb1 L1101S, 'fast,' and Rpb1 E1103G, 'faster'). For every strain, the heat map depicts the fraction of genes in the genome for which each 3' UTR position is used as a poly(A) site. (**B**) End zone profiles of *ALD2* and *PEX6* in strains with wild-type, 'fast', and 'faster'Pol II. (**C**) Scatter plot of the average net overall zone shifts in 'fast'and 'faster' strains. (**D**) Venn diagrams depicting the high degree of overlap in upshifted and downshifted gene categories in strains with altered Pol II elongation rates.
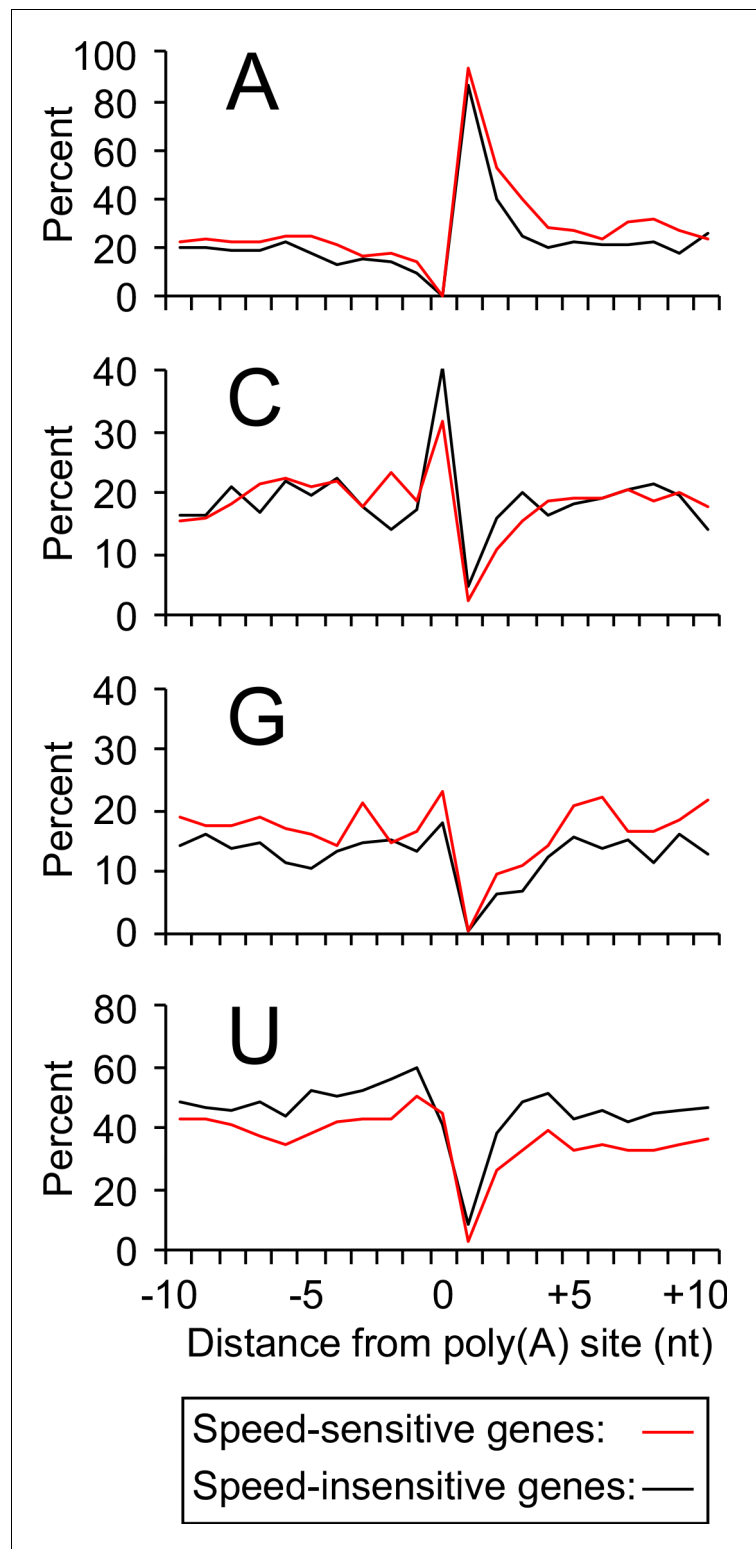
Figure 5. Increased purine content in sequences flanking poly(A) sites of genes sensitive to Pol II speed. Nucleotide frequencies near max isoform poly(A) sites in the wild-type strain (exponential culture) for speed-sensitive genes (462 genes belonging to the 'Both' category; red lines) and speed-insensitive genes (445 genes belonging to the 'Neutral' category; black lines).
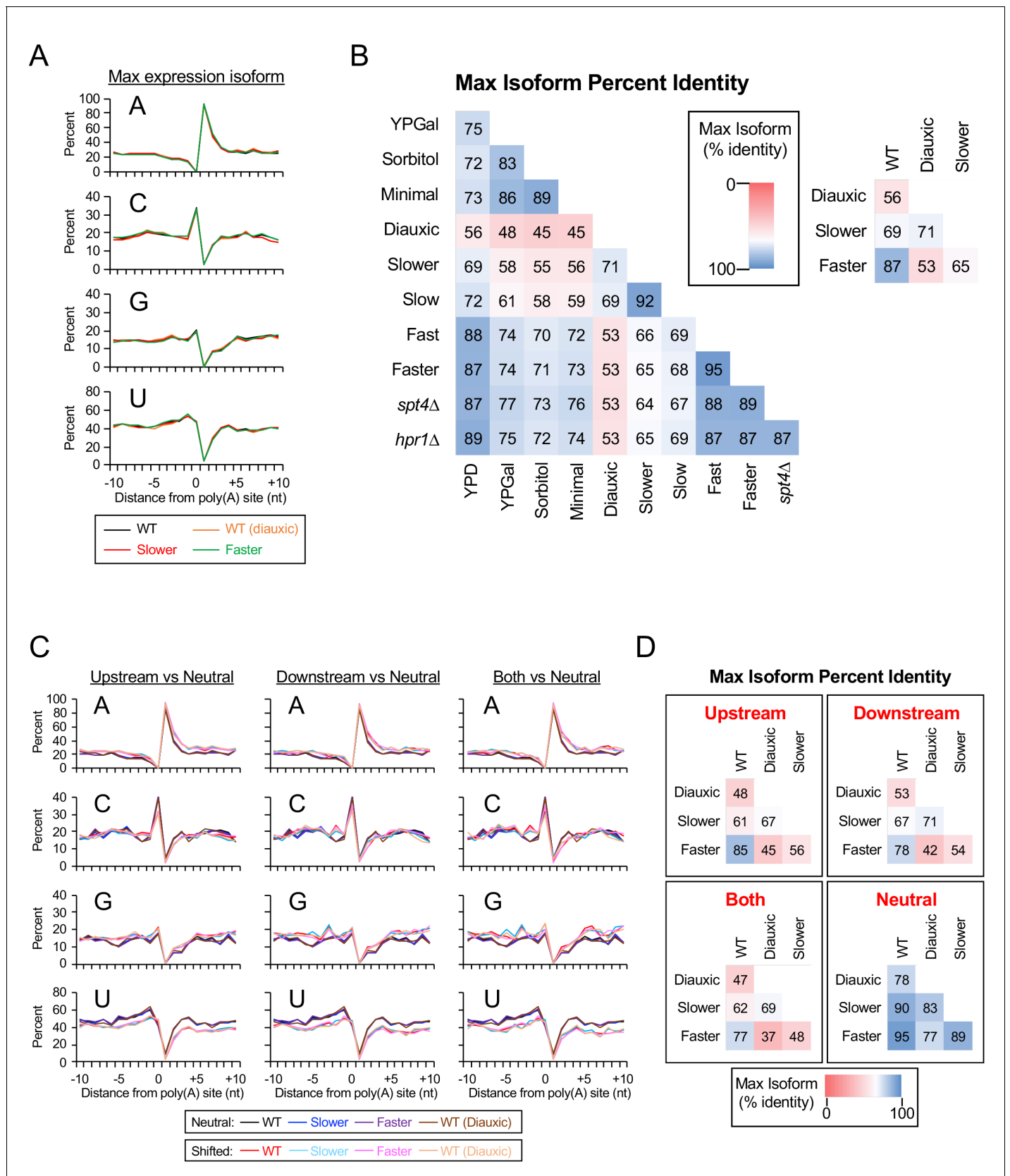
Figure 5—figure supplement 1. Percent identity of max isoform positions by condition, strain, and category. (**A**) Nucleotide frequencies near max isoform poly(A) sites in strains with WT Rpb1 (exponential and diauxic cultures), Rpb1 H1085Q ('slower'), and Rpb1 E1103G ('faster'). Frequencies are

*Figure 5—figure supplement 1 continued on next page*

Figure 5—figure supplement 1 continued

shown for the 2790 genes with ≥1000 reads in the 11 conditions described in the paper. (B) Percent identity of max isoform coordinates in the indicated conditions/strains. The analysis was performed on 2790 genes with over 1000 reads (combined from both replicates). (C) Comparison of nucleotide frequencies near max isoform poly(A) sites in distinct subsets of yeast genes. (D) Pairwise percent identities of max isoform coordinates for wild-type (exponentially-growing and diauxic), slower-Pol II (Rpb1 H1085Q), and faster-Pol II (Rpb1 E1103G). Max isoform identity percentages are shown for the three major groups (Upstream (1898 genes), Downstream (605 genes), and Neutral (445 genes)) as well as the Both sub-category (462 genes).
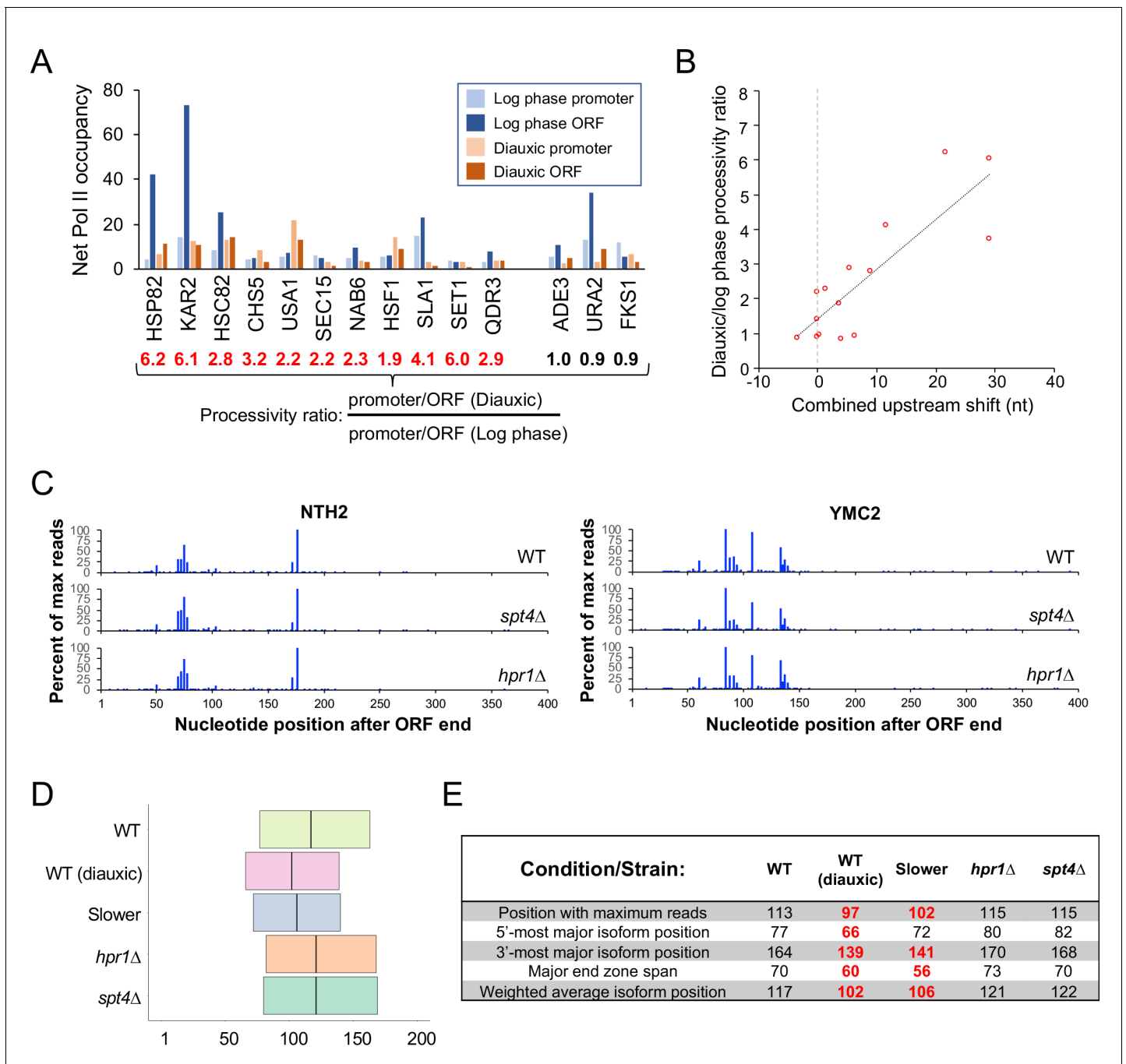
Figure 6. Pol II elongation rate is linked to shifted end zone profiles in diauxic conditions. (**A**) Pol II occupancy (background-subtracted ChIP signal) at promoters and ORFs of select genes in logarithmic growth and diauxic conditions. For every gene, the promoter/ORF occupancy ratio is determined for each condition, and the ratio of these ratios (diauxic/log phase), termed the processivity ratio, is given under the locus name. (**B**) Scatter plot of the diauxic/log phase processivity ratio vs upstream shift (see Materials and methods) in nt observed in diauxic conditions. (**C**) End zone profiles of *NTH1* and *YMC2* in wild-type, *spt4Δ*, and *hpr1Δ* strains. (**D**) Plot of genome-wide median major end zones in wild-type (log phase and diauxic), slower-Pol II (Rpb1 H1085Q), *hpr1Δ*, and *spt4Δ* strains. (**E**) Landmark statistics table in these strains. (All genes with >1000 reads/condition). Bold red numbers represent statistically meaningful upstream shifts vs WT (p < 0.01).
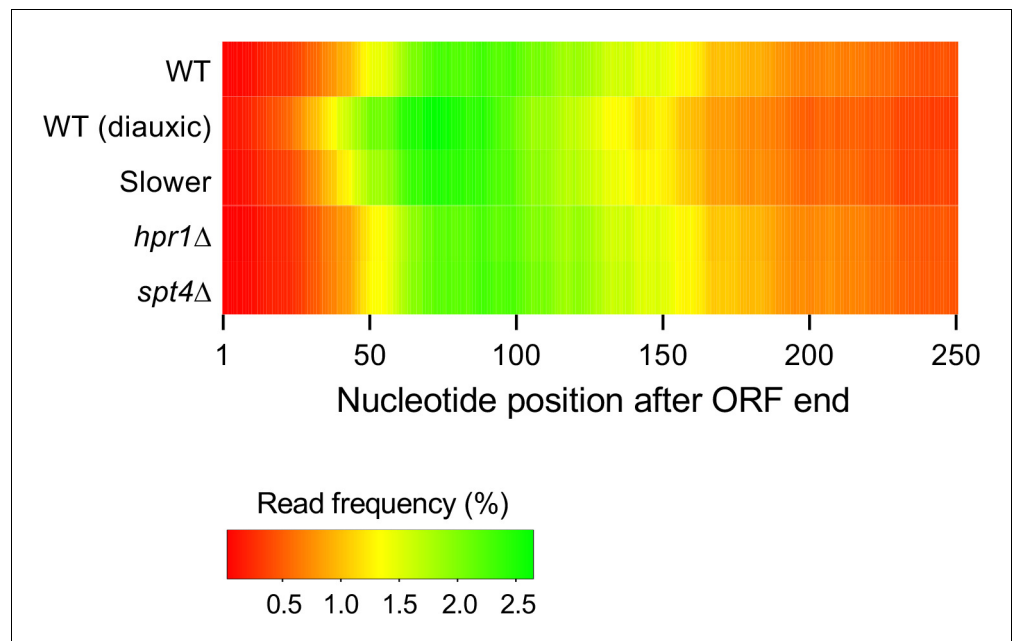
Figure 6—figure supplement 1. 3'UTR percent coordinate utilization for several strains/conditions. Heat map of 3'UTR coordinate use in wild-type Pol II (in exponential and diauxic conditions), 'slower' Pol II (Rpb1 H1085Q), *hpr1Δ*, and *spt4Δ* strains. Each row of the heat map depicts the fraction of genes in the genome for which each 3' UTR position is used as a poly(A) site.
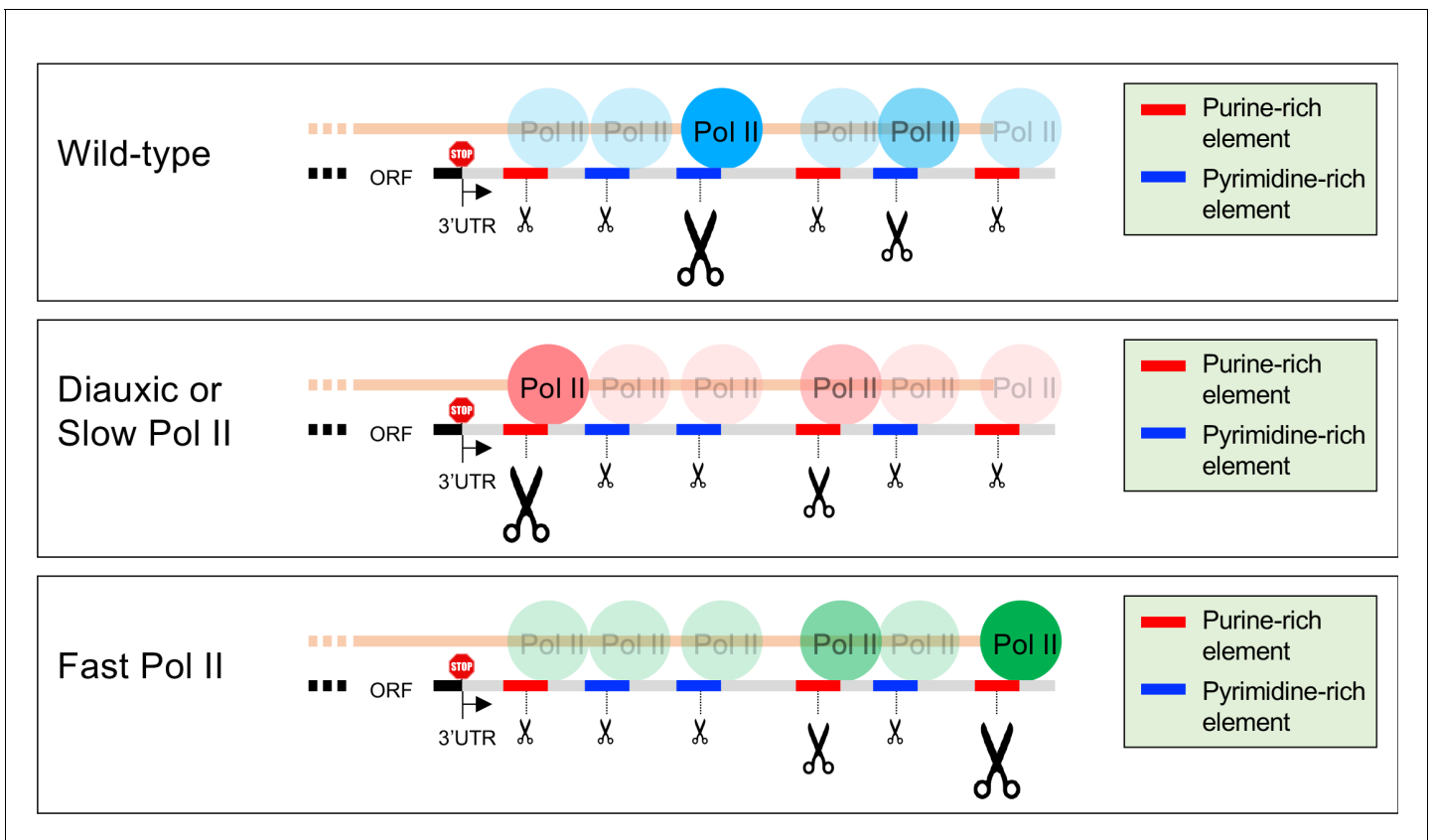
Figure 7. Model of poly(A) site shift in Pol II speed-sensitive genes. The 3'UTRs of speed-sensitive genes contain purine-rich elements (red line segments) and pyrimidine-rich elements (blue line segments) of varying strengths (small, medium or large scissors). Under normal conditions (exponentially-growing wild-type cells), cleavage and polyadenylation takes place predominantly at pyrimidine-rich elements. In diauxic conditions and in cells harboring slow Pol II, purine-rich elements drive an upstream shift in polyadenylation patterns, likely due to increased Pol II dwell time at those sequences. Conversely, fast Pol II shifts the poly(A) patterns to more distal purine rich sites.