
Figures and figure supplements

The landscape of transcriptional and translational changes over 22 years of bacterial adaptation

John S Favate *et al*

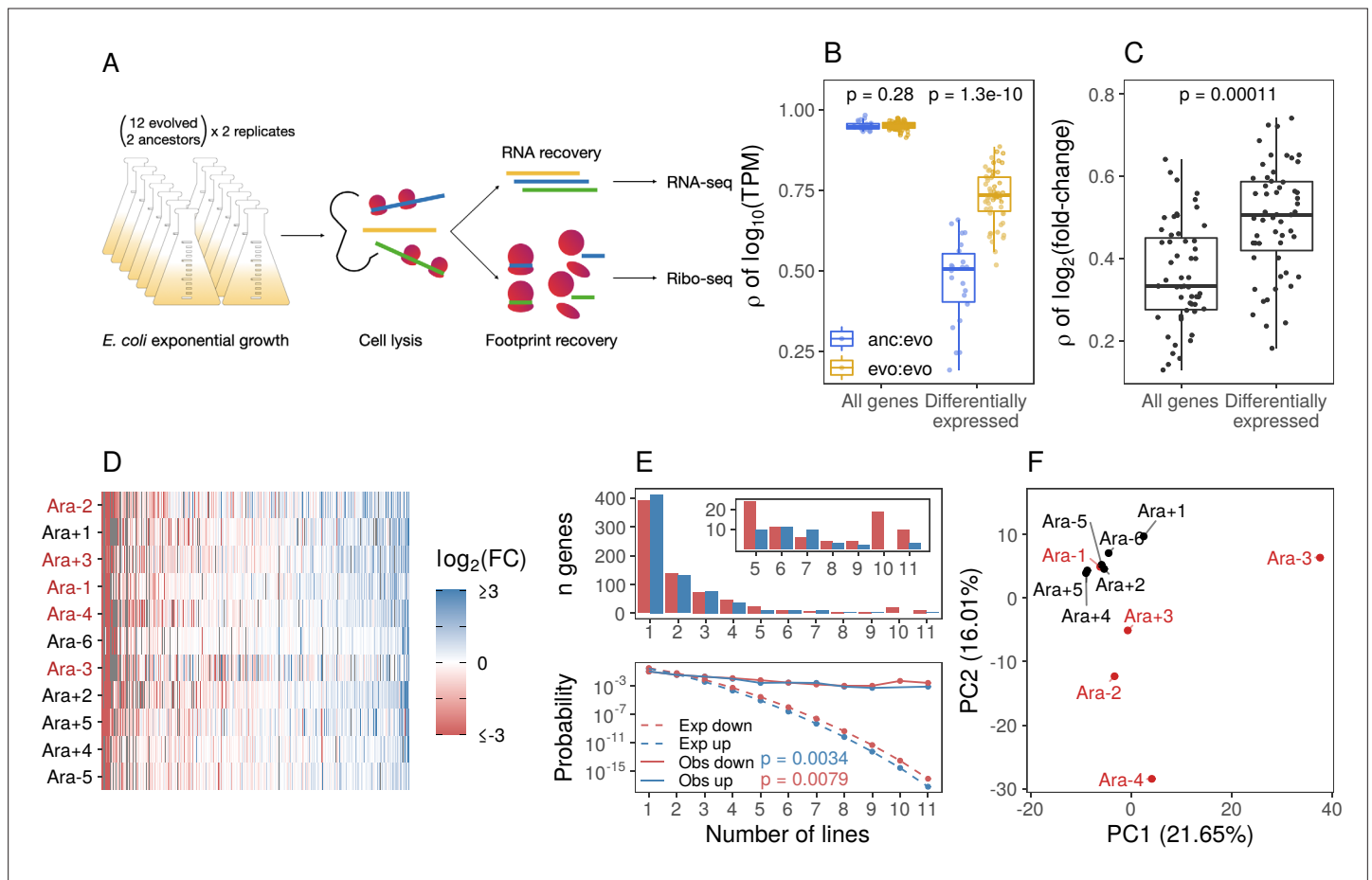


Figure 1. Parallel changes in mRNA abundances. **(A)** Schematic diagram of the experimental setup. **(B)** Pairwise Pearson correlations based on $\log_{10}(TPM)$ (where transcripts per million [TPM] is the mean from replicates) separated by comparisons between evolved lines or from ancestors to evolved lines. p-Values indicate the results of a Kolmogorov-Smirnov (KS) test. For differentially expressed genes (DESeq2 $q \leq 0.01$), evolved line were compared using the union of the significant genes from each line. When comparisons were between an evolved line and an ancestor, the significant genes from that evolved line were used. **(C)** Pairwise Spearman's correlations based on fold-changes from all genes, and the union of the significant genes between two evolved lines (differentially expressed). **(D)** Fold-changes of differentially expressed genes that were significantly altered in at least one line. Genes are ordered left to right in increasing mean fold-change across all evolved lines. Genes containing deletions are not assigned a fold-change and are represented as gray spaces. Lines with a mutator phenotype are in red. **(E)** The upper panel shows the number of genes (y-axis) that were both statistically significant and had a fold-change in the same direction in a particular number of lines (x-axis). The bottom panel shows the expected (dashed) and observed (solid) probability of observing a particular result. p-Values are the result of a KS test between the observed and expected distributions. **(F)** Principal component analysis (PCA) based on all fold-changes. In this case, genes with some form of deletion (complete or indel) are assigned a fold-change of -10 to indicate severe downregulation because they are either completely absent from the genome or not expected to produce functional proteins.



Figure 1—figure supplement 1. Sequencing data statistics. **(A)** The average number of reads aligned per gene using Kallisto for each library. The color scheme remains the same in panels B and D. **(B)** Distributions of mapped and deduplicated read counts per gene in each sample. **(C)** Correlations between the replicates based on rounded counts or transcripts per million (TPM). **(D)** The periodicity of the Ribo-seq datasets determined using a fast Fourier transform (see Codon-specific positioning of Ribo-seq data in Materials and methods).

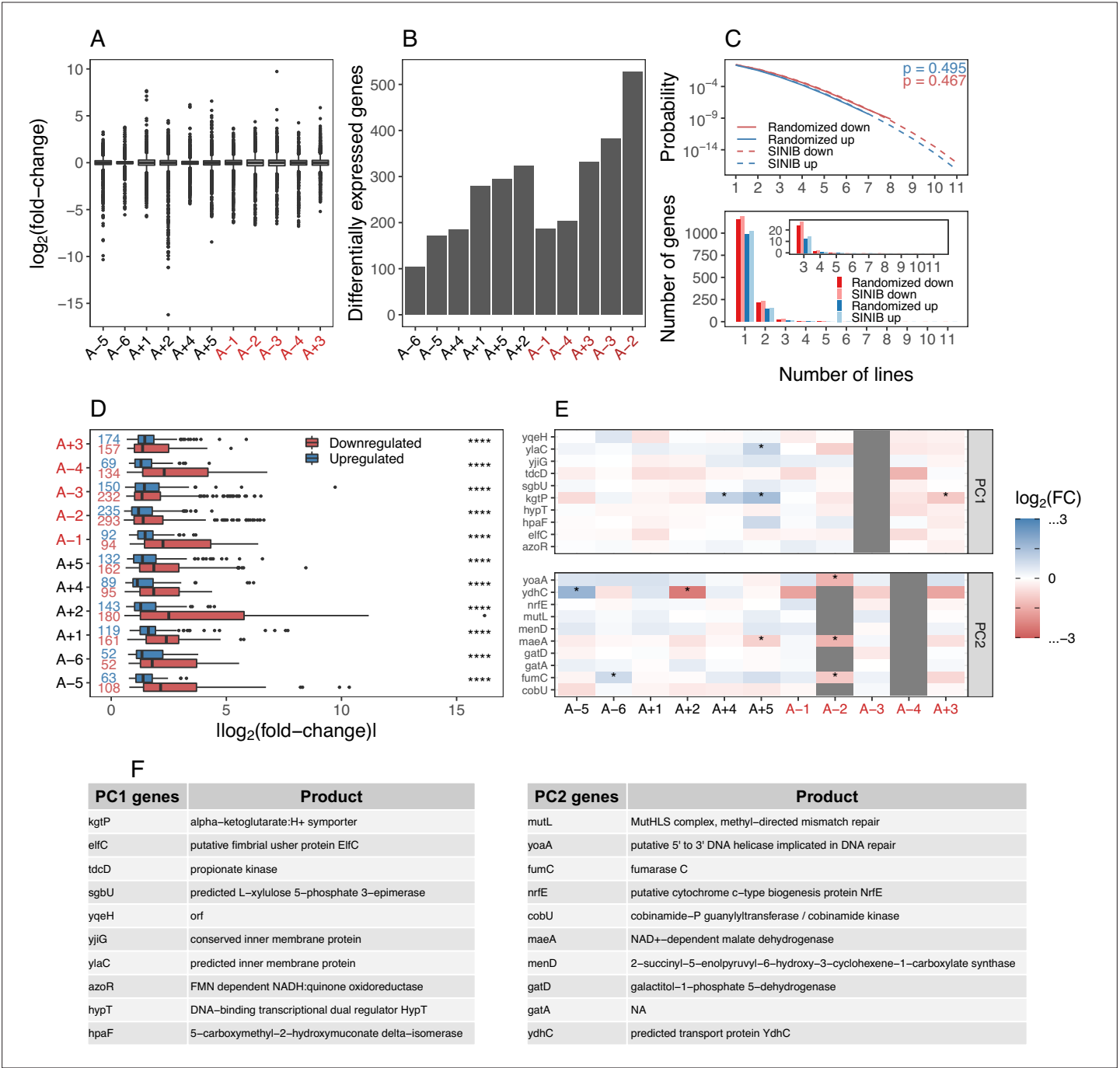


Figure 1—figure supplement 2. Magnitude and variation in mRNA fold-changes across evolved lines. **(A)** Distributions of all mRNA fold-changes (using DESeq2) in each line. Lines with a mutator phenotype are in red. **(B)** The number of differentially expressed genes (DEGs) (DESeq2 $q \leq 0.01$) in each line. **(C)** Upper panel shows the probabilities of observing a gene that was differentially expressed and altered in the same direction in a given number of lines (x-axis). The solid lines represent mean probabilities derived from randomizing the fold-changes of genes in each line 1 million times and the dashed lines represent the probabilities calculated using the sum of independent non-identical binomial random variables (SINIB) method as shown in **Figure 1E**. p-Values show the result of a Kolmogorov-Smirnov (KS) test comparing the randomized to the SINIB distributions. The lower panel shows the expected number of DEGs that are shared and altered in the same direction in a given number of lines (x-axis) based on the above probabilities. **(D)** Distributions of absolute fold-changes of DEGs in each line. The number of DEGs in each evolved line is indicated. Asterisks indicate the results of a KS test comparing distributions of the magnitudes of positive and negative fold-changes in each line NS: $p > 0.05$, *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$ ****: $p \leq 0.0001$. **(E)** The list of top 10 genes contributing to variation in each principle component, gray spaces represent deletions which were encoded as having a $\log_2(\text{fold} - \text{change}) = -10$. **(F)** The genes and descriptions of genes contributing to first two principal components retrieved from EcoCyc (Keseler et al., 2005).

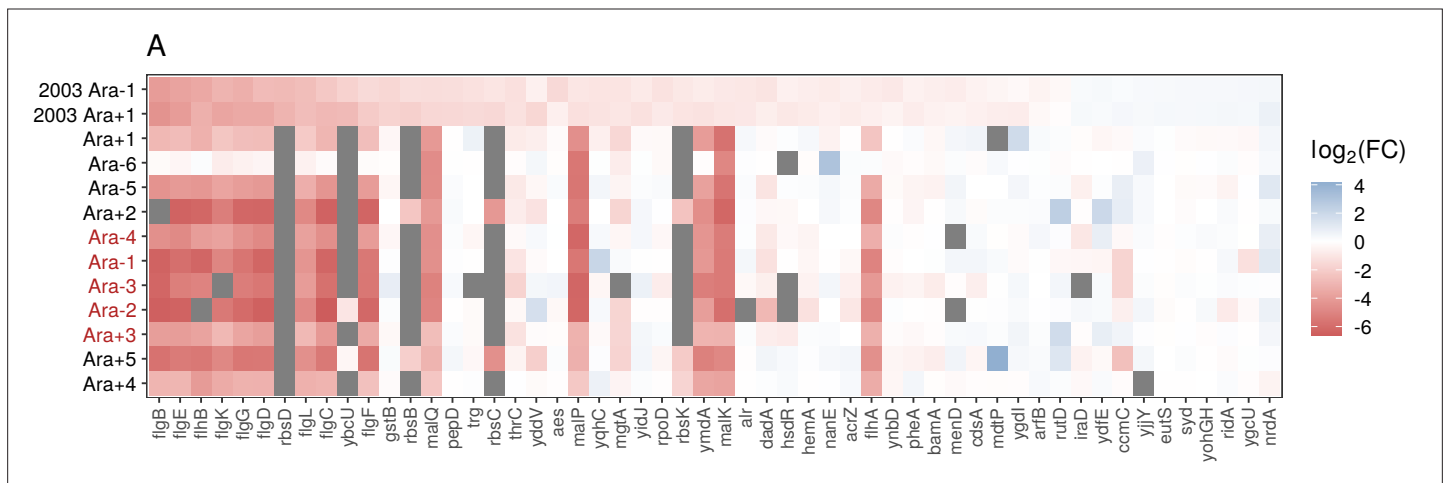


Figure 1—figure supplement 3. Comparison of expression changes between this study and *Cooper et al., 2003*. **(A)** The direction and magnitude of expression changes in genes identified as differentially expressed in *Cooper et al., 2003*, study and the direction of changes for those genes in our dataset. While the two datasets share a color scale for fold-change, the data underlying the *Cooper et al., 2003*, study was generated using a microarray compared to RNA-seq data in the current study.

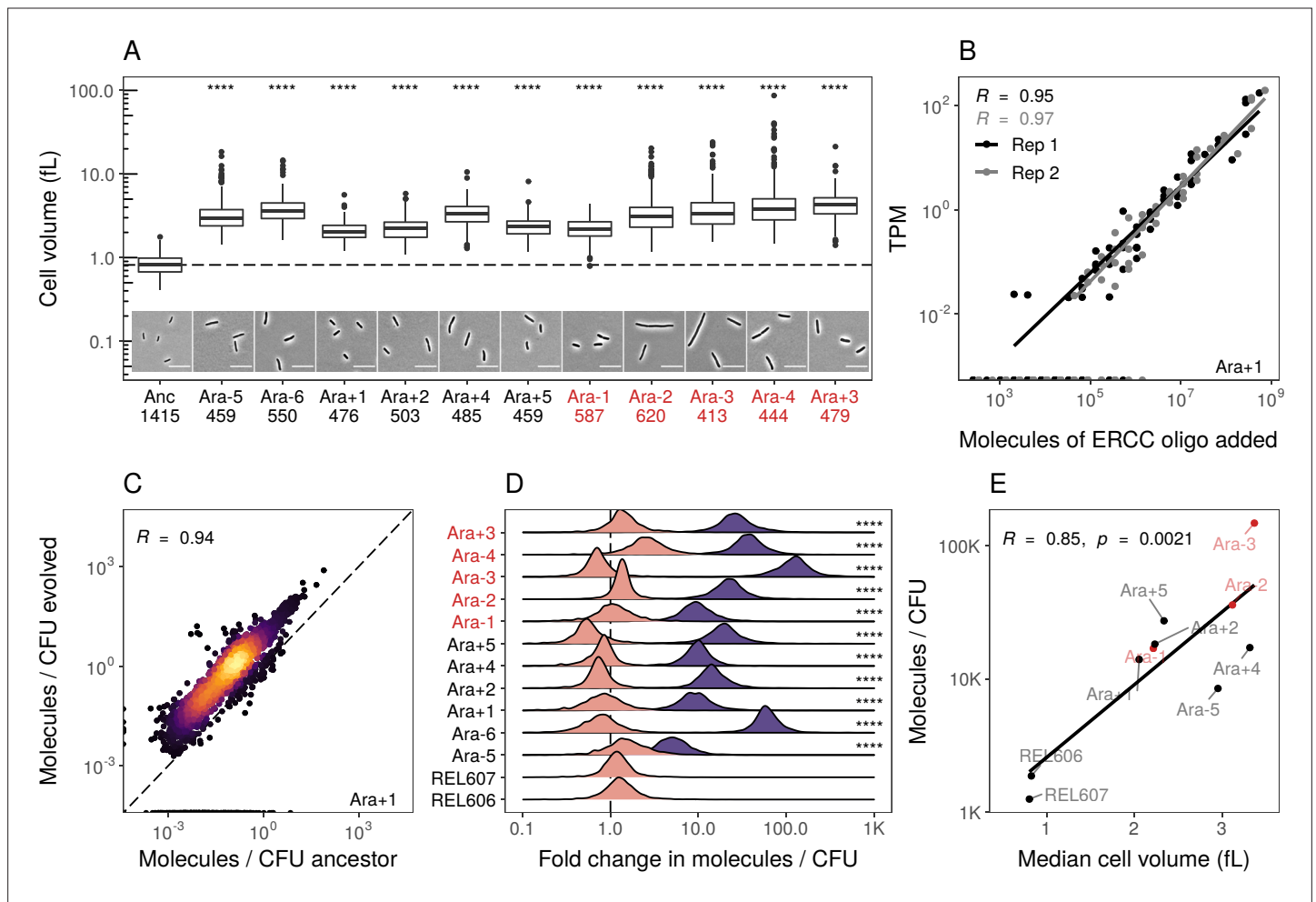


Figure 2. Evolved lines are larger in cell size and carry more mRNAs. **(A)** All evolved lines are larger than the ancestral strain. Distributions of cellular volume as determined by phase-contrast microscopy and assuming spherocylindrical shape of *Escherichia coli* along with representative images for each line. Numbers underneath a line's name indicate the total number of cells imaged (scale bar is 10 μ m). The dashed line indicates the ancestral median, p-values indicate the results of a t-test when each line is compared to the ancestor, **** $p \leq 0.0001$. Lines listed in red have mutator phenotypes. **(B)** Abundances of spike-in RNA control oligos are correlated with their estimates in sequencing data. Linear models relating the number of molecules of each ERCC control sequence added to their RNA-seq TPM (transcripts per million) in Ara+1 RNA-seq sample (see **Figure 2—figure supplement 2** for data for all lines). **(C)** Most genes have a higher absolute expression in evolved lines. Changes in the absolute number of mRNA molecules per CFU (colony-forming unit) in the 50,000th generation of Ara+1 relative to the ancestor. The values plotted are the averages between two replicates of the evolved lines and both replicates from two ancestors (REL606 and REL607; see **Figure 2—figure supplement 2** for all lines). **(D)** Absolute changes in mRNA abundances of genes in evolved lines are significantly larger than the variation between biological replicates (KS test, $p < 0.0001$ in all cases). Pink distributions indicate gene-specific fold-changes between biological replicates for each line (centered around 1). Purple distributions show the absolute fold-changes in molecules of RNA per CFU from the ancestor to each evolved line. Fold-changes are calculated in the same manner as in C. **(E)** Larger evolved lines have more mRNA per CFU. Relationship between the median cellular volume for each line and the total number of RNA molecules per CFU. Total molecules of RNA are calculated as the sum of the average number of molecules for each gene between replicates.

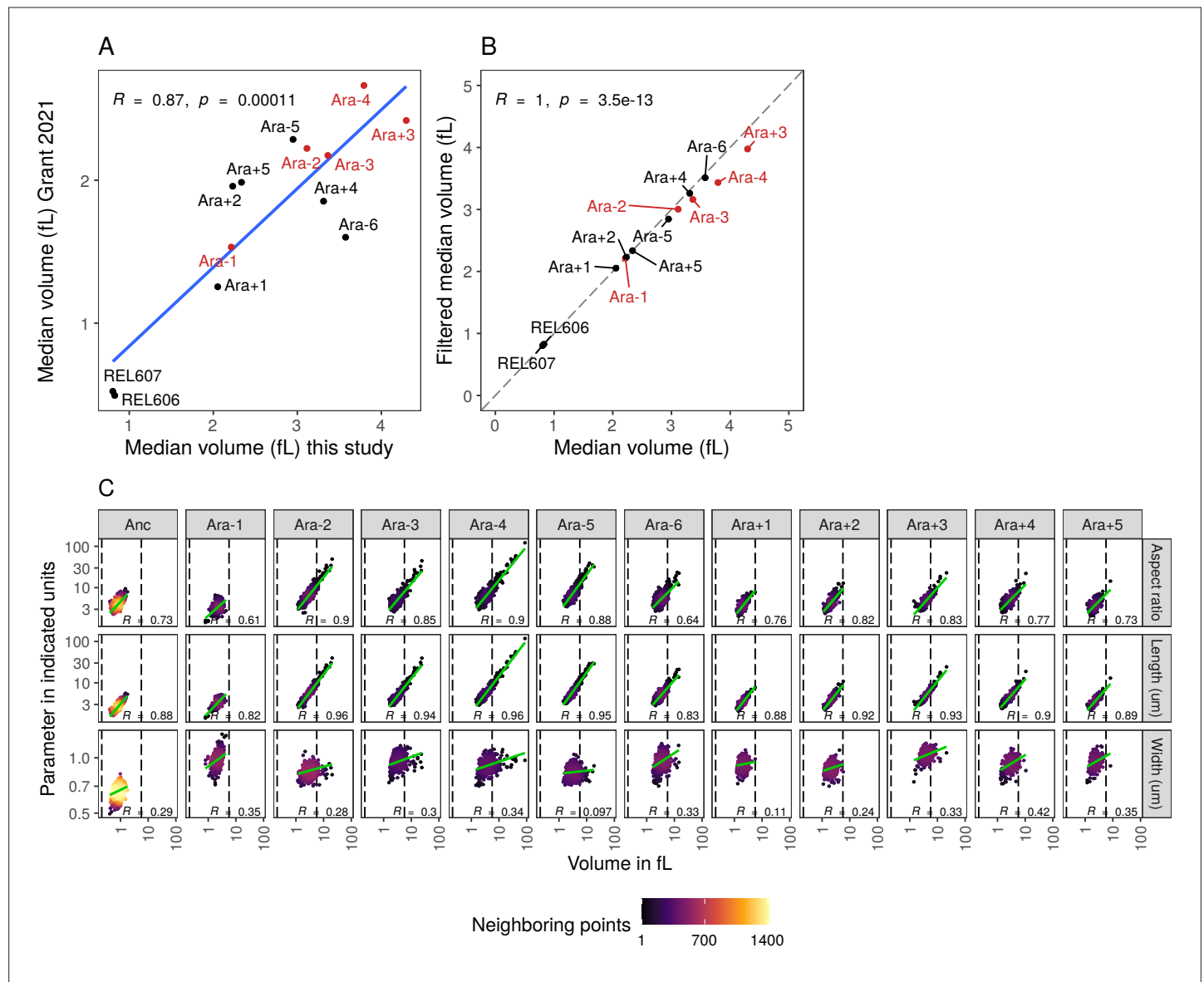


Figure 2—figure supplement 1. Relationship between cellular features and cell volume. (A) Comparison of median volumes of each evolved line from this manuscript to estimates of cellular volumes from *Grant et al., 2021*. (B) Relationship between median cell volumes of all cells compared to median cell volume of filtered cells between 0.21 and 5.66 fL used in *Grant et al., 2021*. (C) Increase in cell volume is more strongly correlated with cell length compared to cell width. The dotted lines indicate volumes of 0.21 and 5.66 fL.

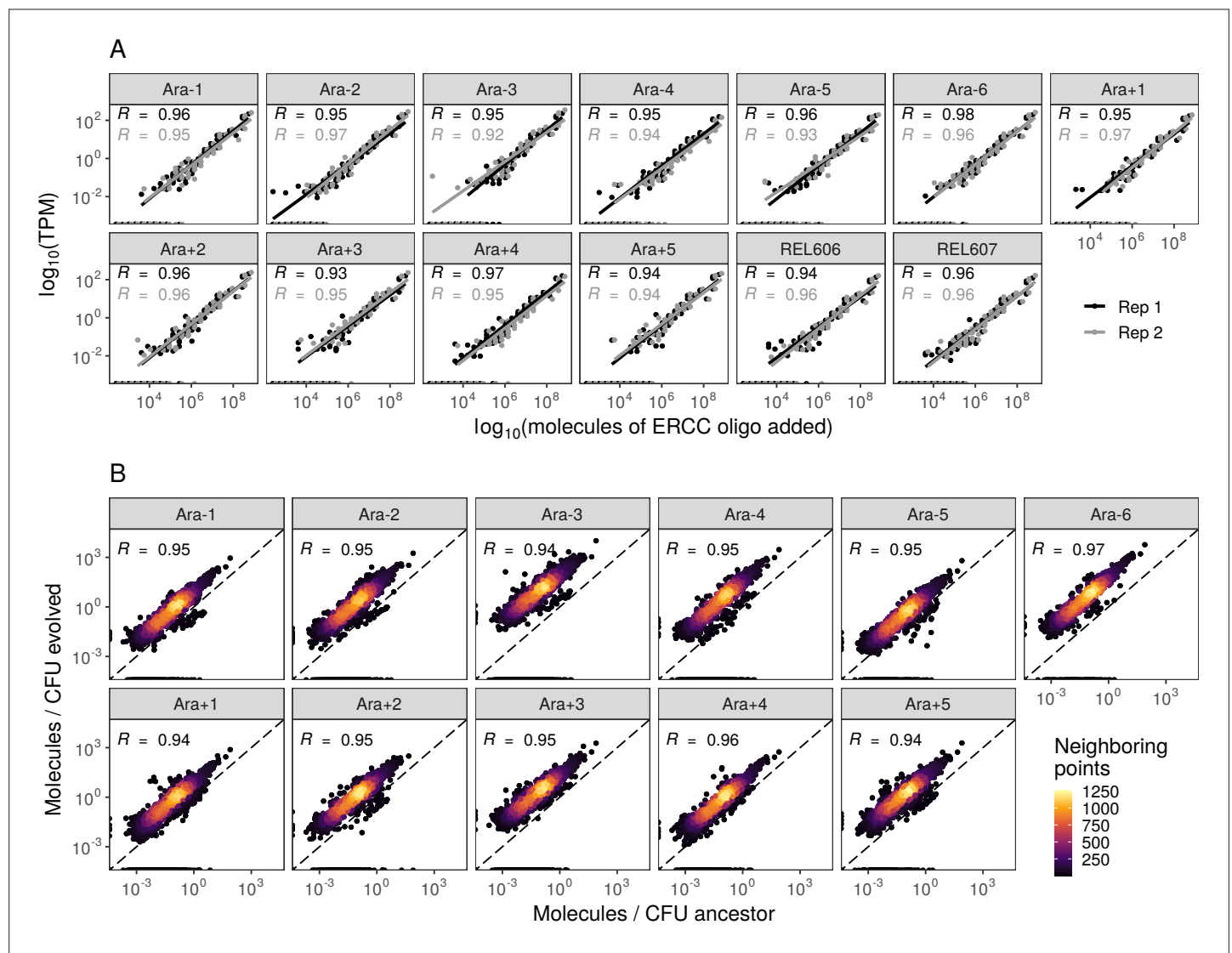


Figure 2—figure supplement 2. Absolute changes in mRNA abundances per CFU across all evolved lines. **(A)** Abundances of spike-in RNA control oligos are correlated with their estimates in sequencing data. Linear models relating the number of molecules of each ERCC control sequence added to their RNA-seq TPM (transcripts per million). **(B)** Most genes have a higher absolute expression in evolved lines. Changes in the absolute number of mRNA molecules per CFU (colony-forming unit) in the 50,000th generation of each line relative to the ancestor. The values plotted are the average between two replicates of the evolved lines and both replicates from both ancestors. REL606 and REL607 are ancestral strains.

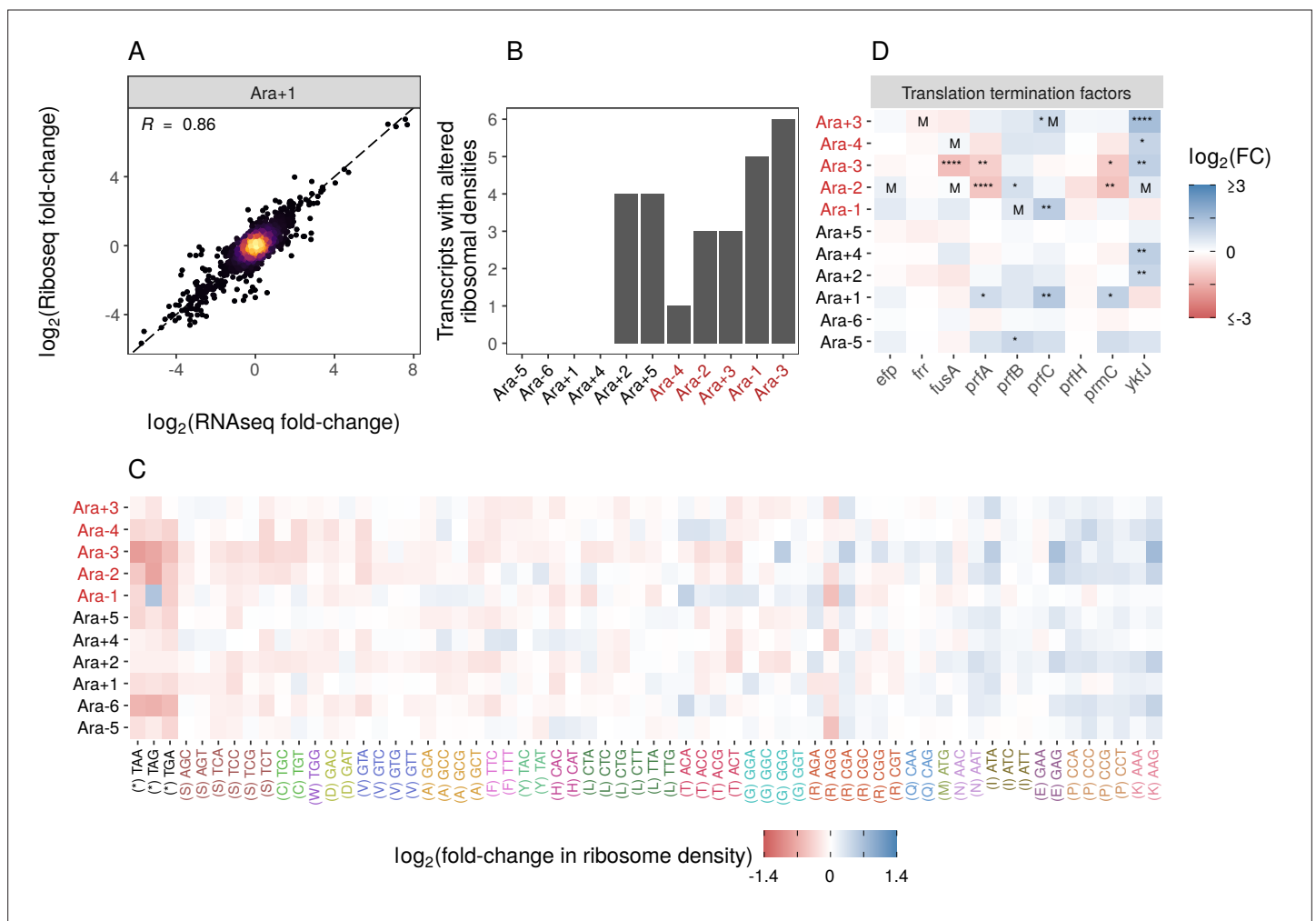


Figure 3. Changes in gene expression at the translational level. **(A)** Translational changes are correlated with transcriptional changes. The relationship between RNA-seq and Ribo-seq fold-changes in Ara+1 (see **Figure 3—figure supplement 1A** for all evolved lines). **(B)** The distribution of genes with significantly altered ribosome densities ($q \leq 0.01$) estimated using Riborex ($q \leq 0.01$). **(C)** Evolved lines have faster translation termination. Stop codons had lowered ribosome density compared to all sense codons. Changes in codon-specific ribosome densities in each of the evolved lines relative to the ancestor. Codons are colored according to the amino acid they code for. Amino acids are ordered left to right in order of mean fold-change across the lines. **(D)** Fold-changes in mRNA abundances of translation termination factors and related genes *ykfJ*, *prfH*, *prfA*, *prmC*, *prfB*, *fusA*, *efp*, *prfC*. RNA-seq fold-changes for termination factors, asterisks indicate DESeq2 q-values (blank: $p > 0.05$, *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$ ****: $p \leq 0.0001$ and an 'M' indicates an SNP in that gene).

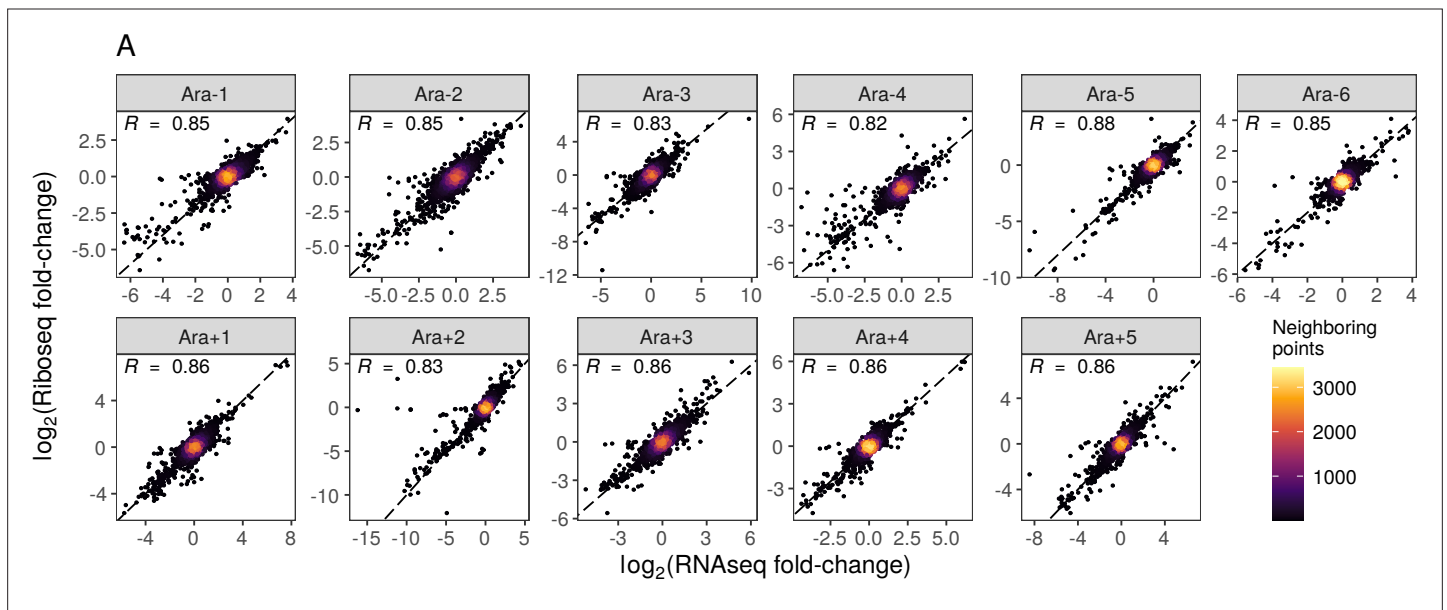


Figure 3—figure supplement 1. Relationship between RNAseq and riboseq fold-changes for all evolved lines. **(A)** The relationship between RNA-seq fold-changes and Ribo-seq fold-changes in evolved lines.

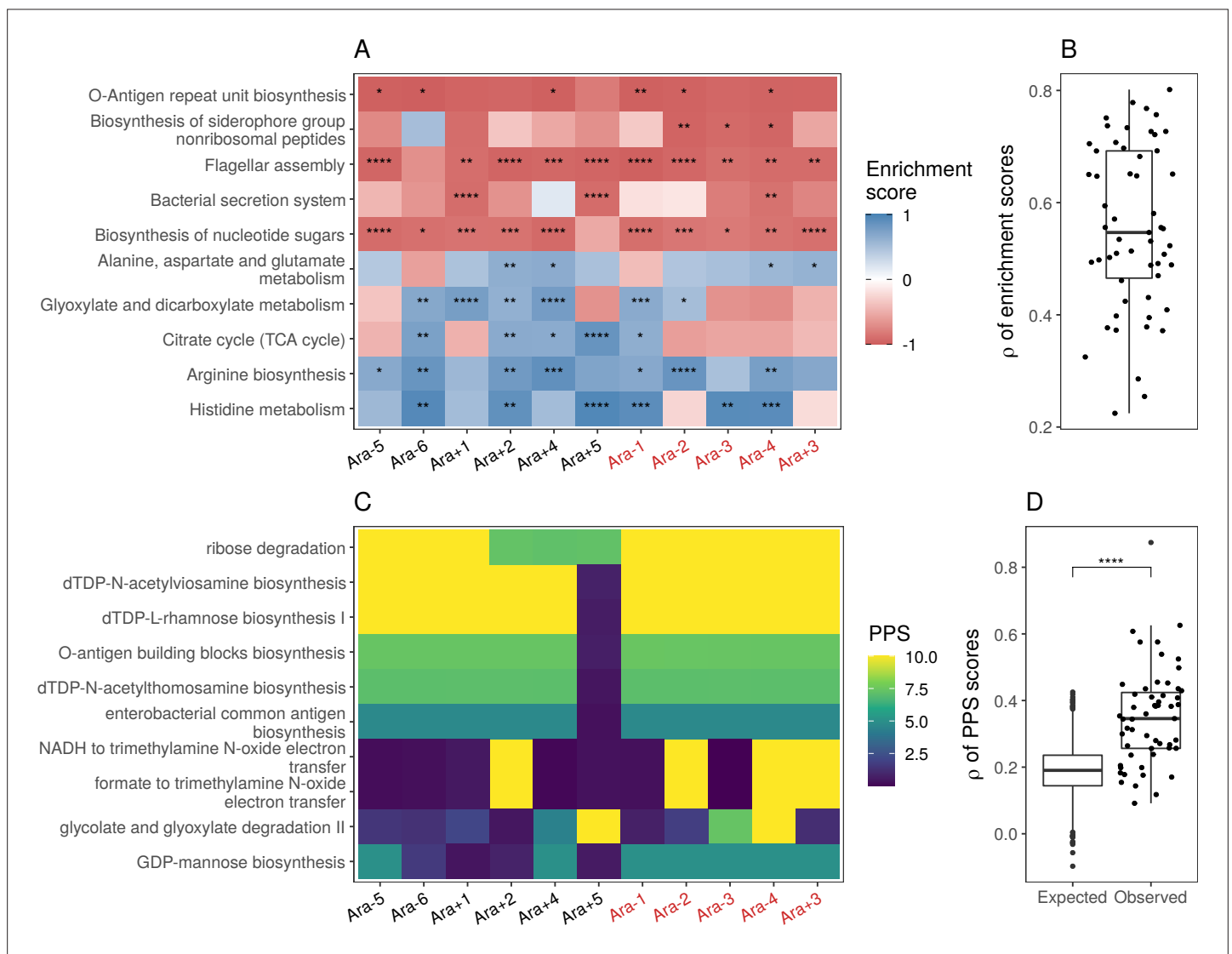


Figure 4. Parallel changes in biological processes and pathways. **(A)** Parallel changes in biological processes and pathways. The top 10 KEGG pathways that were significantly altered ($FDR \leq 0.05$) based on RNA-seq data. Enrichment score represents the degree to which a pathway was up- (positive) or downregulated (negative). Functional categories are ordered by increasing mean enrichment score across the lines. Enrichment score represents the degree to which a pathway was up- (positive) or downregulated (negative). **(B)** Distribution of pairwise Spearman's correlations of enrichment scores of all significantly altered functional categories ($FDR \leq 0.05$). **(C)** The top 10 pathways with the highest mean Pathway perturbation scores (PPS) calculated from RNA-seq fold-changes. Higher PPS indicates larger degrees of alteration but does not indicate directionality. **(D)** Distribution of pairwise Spearman's correlations based on all PPS (observed) compared to 1000 sets of correlations generated from PPS calculated after randomization of fold-changes (expected). The p-value is the result of a Kolmogorov-Smirnov test (blank: $p > 0.05$, *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$, ****: $p \leq 0.0001$).

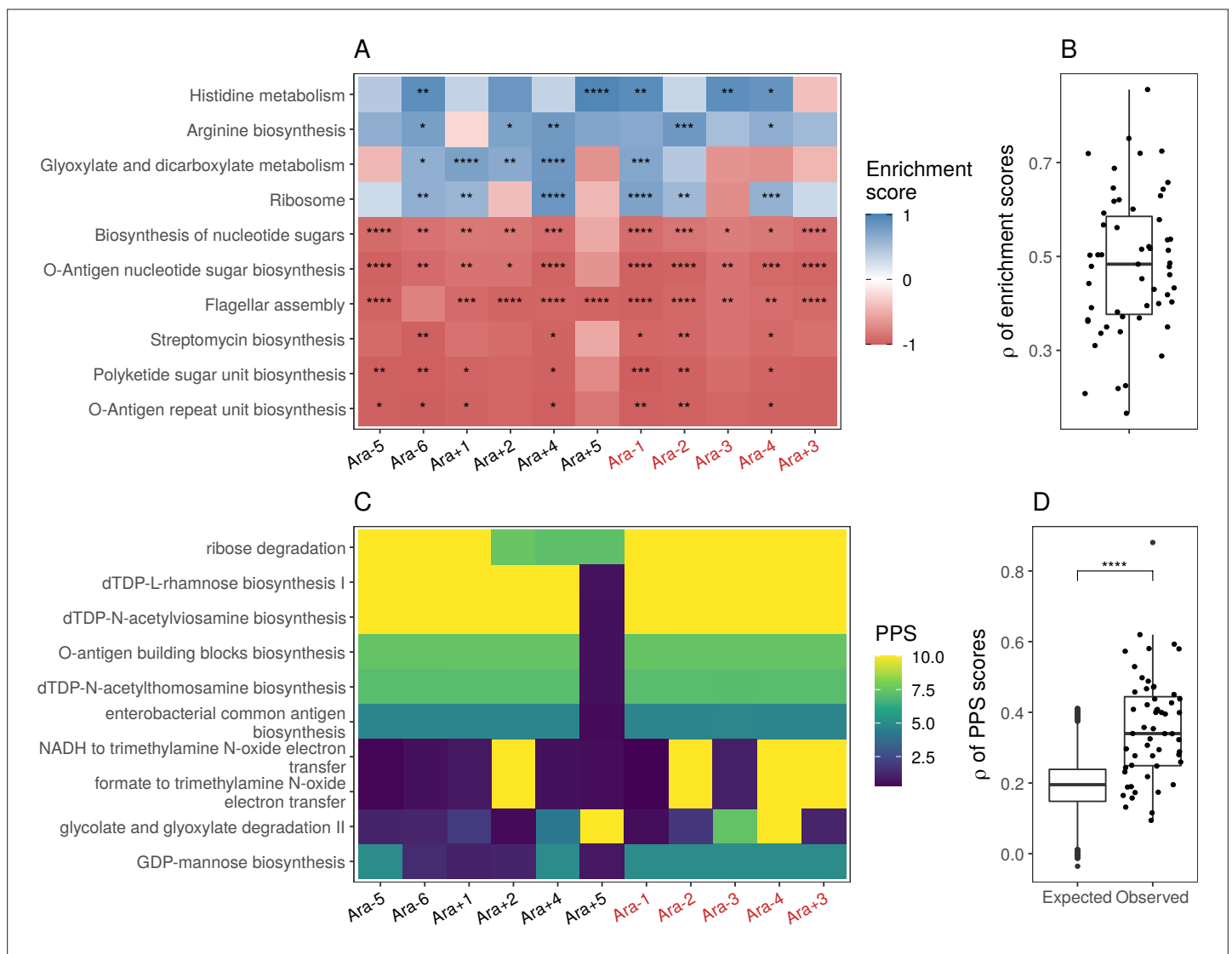


Figure 4—figure supplement 1. Parallel changes in biological processes and pathways based on Ribo-seq data. **(A)** Parallel changes in biological processes and pathways. The top 10 KEGG pathways that were significantly altered ($FDR \leq 0.05$) based on Ribo-seq data. Enrichment score represents the degree to which a pathway was up- (positive) or downregulated (negative). Functional categories are ordered by increasing mean enrichment score across the lines. Enrichment score represents the degree to which a pathway was up- (positive) or downregulated (negative). **(B)** Distribution of pairwise Spearman's correlations of enrichment scores of all significantly altered functional categories ($FDR \leq 0.05$). **(C)** The top 10 pathways with the highest mean pathway perturbation scores (PPS) calculated from Ribo-seq fold-changes. Higher PPS indicates larger degrees of alteration but does not indicate directionality. **(D)** Distribution of pairwise Spearman's correlations based on all PPS (observed) compared to 1000 sets of correlations generated from PPS calculated after randomization of fold-changes (expected). The p-value is the result of a Kolmogorov-Smirnov test (blank: $p > 0.05$, *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$ ****: $p \leq 0.0001$).

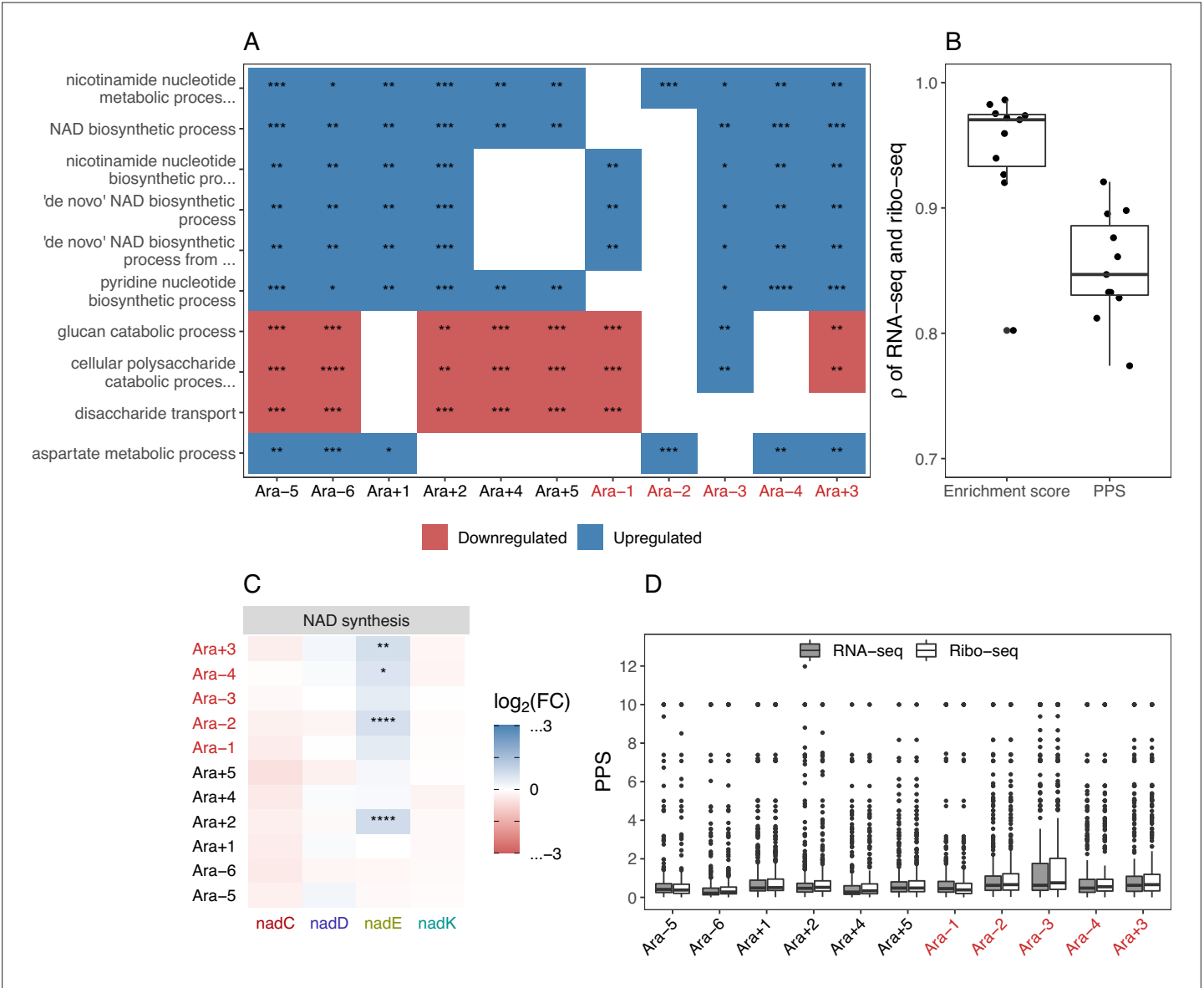


Figure 4—figure supplement 2. GO and other functional analyses of differentially expressed genes. **(A)** The top 10 GO biological process categories that were significantly altered (Fisher's exact test ≤ 0.05). White spaces indicate that the category was not significantly altered in that line. **(B)** Spearman's correlations between the RNA-seq and Ribo-seq enrichment scores within each line. **(C)** RNA-seq fold-changes and DESeq2 q-values for the remaining genes in the nicotinamide adenine dinucleotide (NAD) synthesis pathway shown in **Figure 5**. Gene names along the x-axis are colored based on operon membership (blank: $p > 0.05$, *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$ ****: $p \leq 0.0001$). **(D)** Distribution of RNA-seq and Ribo-seq pathway perturbation scores (PPS) for each line.

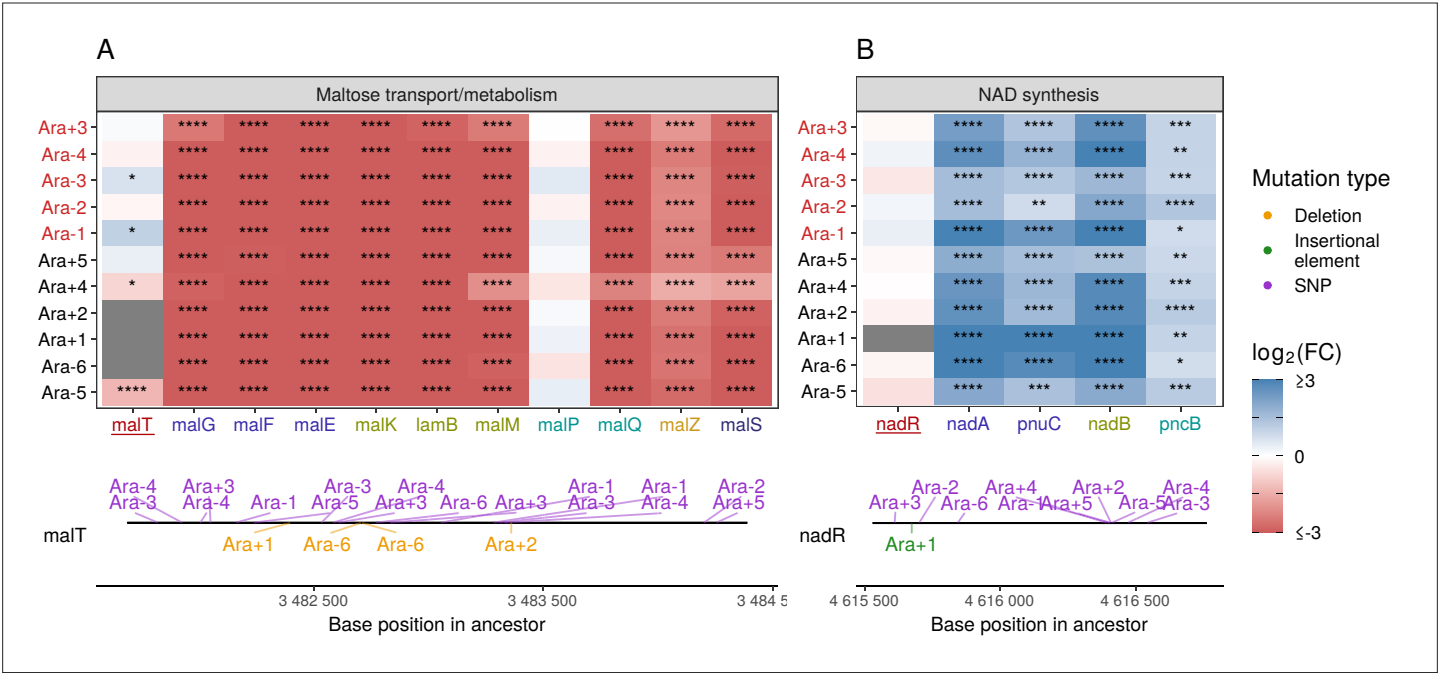


Figure 5. Mutations in transcriptional regulators lead to parallel changes in gene expression. RNA-seq fold-changes for genes belonging to (A) maltose-transport/metabolism and (B) nicotinamide adenine dinucleotide (NAD) biosynthesis. Gene names in each category are colored based on their operon membership. Mutations in transcriptional activator malT decrease expression of its downstream genes/operons. Mutations in transcriptional repressor nadR increase expression of its downstream genes/operons. Asterisks indicate statistical significance of fold-changes (blank: $q > 0.05$, *: $q \leq 0.05$, **: $q \leq 0.01$, ***: $q \leq 0.001$ ****: $q \leq 0.0001$). Gray panels in the heatmap indicate gene deletion. Lower panels show the type and location of mutations in each transcription factor.

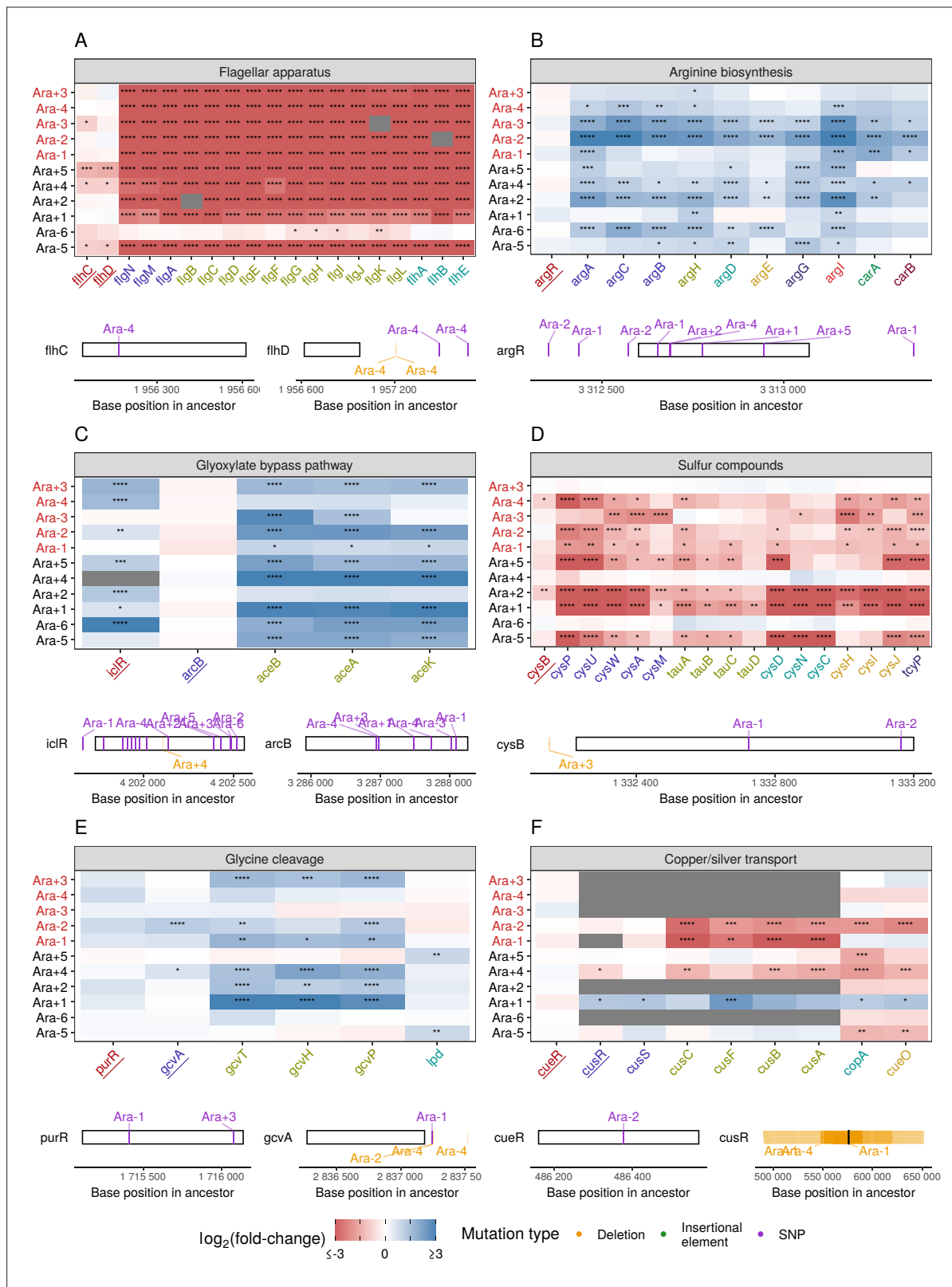


Figure 5—figure supplement 1. Link between mutations and expression changes for other gene sets. (A–F) Mutations in transcriptional regulators lead to parallel changes in gene expression (RNA-seq). Gene names in each category are colored based on their operon membership. Transcription factors for each class of genes are underlined. Asterisks indicate statistical significance of fold-changes (blank: $p > 0.05$, *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$, ****: $p \leq 0.0001$). Gray panels in the heatmap indicate gene deletion. Lower panels show the type and location of mutations in each transcription factor.