Figure S1. SA-1 yeasts show increased production of ethanol and lower biomass yield followed by statistically significant alterations in the transcriptome. (A) Regression plot showing the relationship (in Pearson's R²) between biological replicates under the same condition. Each point represents the expression of a gene in log2(CPM), measured in the first biological replicate (X-axis) and the second

replicate (Y-axis). **(B)** PCA plot showing the explained variance in the first two components based on normalized gene expression for control (grey) and treated (black) samples. **(C)** Kernel density estimation (KDE) plot showing the distribution of gene expression ratios across the four samples. The Y-axis represents the density of ratio values in that sample, while the X-axis represents the ration between the expression of the gene in that sample when compared to the average expression across all samples in the dataset. Under the assumption that most genes are NOT differentially expressed, it is expected that the center of the distribution is close to 0 with slight skewness observed towards either the right or left caused by differential gene expression between conditions. **(D)** Distribution plots showing the relationship between observed p-values for differential gene expression testing when compared with the distribution of variance (Top) and mean expression (Bottom). The Y-axis shows the number of p-values observed in each bucket of the histogram, the X-axis shows the p-values thresholds for each bucket and the colors represent the quantile scores for each class (variance and expression). Under well controlled experiments that induce alterations in the transcriptome, it is expected to see a flat distribution of values with a peak in the [0, 0.1] set. Additionally, under the assumption that there is no expression/variance bias in the dataset, it is expected to see a near-even distribution of quantiles in each bucket. **(E)** Scatterplot showing the relationship between median expression value (X-axis) and standard deviation from expression (Y-axis) for each gene (gray) in the dataset. Under the assumption that these two variables are not correlated in the dataset, it is expected to see a random distribution of differentially expressed genes (red) inside the expected correlation range (defined by the dashed lines).