1 Supplementary Methods

1.1 Experimental Protocol

- Details of the media used can be found at `http://www.princeton.edu/genomics/botstein/protocols/Dex_lim.htm`.
- Details of the protocol for harvesting RNA from the chemostat can be found at `http://dunham.gs.washington.edu/MDchemostat.pdf`.
- The R64 release of the *S. cerevisiae* genome is available at `http://downloads.yeastgenome.org/sequence/S288C_reference/genome_releases/`.

1.2 Percent of Variance Explained

Percent of variance explained by each level was computed by using ANOVA to compare three nested linear models. The three methods represented, respectively, treatment, treament×biological, and treatment×biological×preparation. Each takes the form

$$Y = \beta {X_{1:p+1}}^T + E$$

where

- p is the number of degrees of freedom used by the parameters in the model (respectively 1, 3, and 7 for the three models)
- g is the number of genes
- Y is the $g \times 16$ matrix of red/green fold changes in microarray data or the log-transformed counts of RNA-Seq data
- $X_{1:p+1}$ is the first $p+1$ columns of the experimental design matrix (Figure S9), which includes the intercept and the p terms fit in the model
- β is the $g \times (p+1)$ matrix of terms
- E is the $g \times 16$ matrix of normally distributed residuals

An adjusted R^2 was found for each of the models $(1 - \frac{RSS(16-1)}{TSS(16-p-1)})$, as shown:

Model	p	Adjusted R^2
Treatment	1	R^2_T
Treatment×Biological	3	R^2_B
Treatment×Biological×Preparation	7	R^2_P

The percent of variation explained by each level was reported for each gene as R^2_T for the treatment variation, $R^2_B - R^2_T$ for the biological variation, $R^2_P - R^2_B$ for the preparation variation, and $1 - R^2_P$ for the residual variation that occurs on the chip or the lane. This metric is meant to break down the amount that the quantification of expression, and therefore the differential expression estimation, is affected by each level.

Since RNA-Seq data consists of counts rather than continuous measurements, we considered the possibility that the linear model, and therefore the R^2 measure, might not capture the distribution of variation correctly. We thus also calculated two alternative R^2 generalized across exponential families: $R^2_{DEV,P}$ which assumes a Poisson model for the residual noise, and $R^2_{DEV,NB2(ML)}$, which assumes a negative binomial model [1]. This metric is defined as the decrease in Kullback-Leibler divergence between the observations and the expectation under the full model versus the divergence between the observations and the expectation under a null model. The standard R^2 used in linear models with normal residuals can be viewed as a special case of this metric, so this divergence is a useful choice for comparing the count values to the traditional R^2 used in the microarray [2].

We used the edgeR package to perform a negative binomial fit on each gene and estimate the dispersion parameter. We define $y_{i,j}$ as the unnormalized count of gene i in sample j, $\hat{\mu_{i,j}}$ as the expected value of $y_{i,j}$ under the full model (using `fitted` on an edgeR fit), $\bar{y}_{i,j}$ as the expected value using only an intercept term and the library size normalization, $\hat{\alpha_i}$ is the dispersion of gene i estimated using edgeR, n is the number of samples (16) and p is the number of parameters estimated by the model (1 for the condition effect, 3 for condition×biological, and 7 for condition×biological×preparation). The alternative R^2 estimates for the Poisson and negative binomial models can then be calculated as

$$R^2_{DEV,P,i} = 1 - \frac{\sum_{j=1}^N y_{i,j}\log(y_{i,j}/\hat{\mu}_{i,j}) - (y_{i,j} - \bar{\mu}_{i,j})}{\sum_{j=1}^N y_{i,j}\log(y_{i,j}/\bar{y}_{i,j}) - (y_{i,j} - \bar{y}_{i,j})}\frac{n-1}{n-p-1},$$

$$R^2_{DEV,NB2(ML),i} = 1 - \frac{\sum_{j=1}^N y_{i,j}\log(\hat{y_{i,j}}/\hat{\mu}_{i,j}) - (y_i + \hat{\phi}_i^{-1})\log(\frac{y_{i,j}+\hat{\phi}_i^{-1}}{\hat{\mu}_{i,j}+\hat{\phi}_i^{-1}}))}{\sum_{j=1}^N y_{i,j}\log(\hat{y_{i,j}}/\bar{y}_{i,j}) - (y_i + \hat{\phi}_i^{-1})\log(\frac{y_{i,j}+\hat{\phi}_i^{-1}}{\bar{y}_{i,j}+\hat{\phi}_i^{-1}}))}\frac{n-1}{n-p-1}.$$

The same three models fit in the traditional linear model were fit using Poisson or negative binomial regression, using edgeR's implementation:

$$Y_{i,\cdot} \sim \text{Pois}(\exp(\beta_i {X_{i,1:p+1}}^T))$$

$$Y_{i,\cdot} \sim \text{NegBin}(\exp(\beta_i {X_{i,1:p+1}}^T)), \alpha_i)$$

then their respective R^2_T, R^2_B and R^2_P were calculated for each gene based on the difference in adjusted alternative R^2 between each model. Figure S4 compares the Poisson and negative binomial models to the traditional R^2 on log TMM-normalized counts. TMM was chosen since it is the default used for edgeR normalization, and Figure S3 shows that TMM produces similar results to other normalization methods.

Supplemental References

1. A C Cameron and Frank Windmeijer. An R-squared measure of goodness of fit for some common nonlinear regression models. *J Econometrics*, 77(2):329–342, 1997.

2. A Colin Cameron and Frank Windmeijer. R-Squared Measures for Count Data Regression Models With Applications to Health-Care Utilization. *J Bus Econ Stat*, 14(2):209–220, 1996.

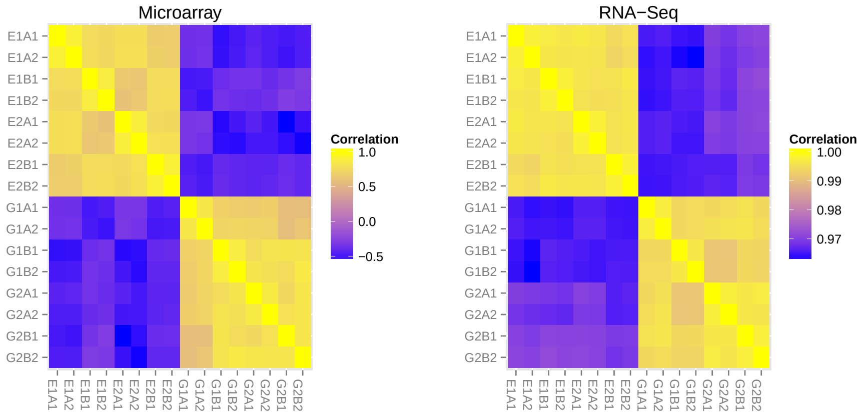

Figure S1: Spearman correlation matrices of microarray and RNA-Seq experiments. The microarray correlation matrix uses the log(cy5/cy3) fold change, which means the correlation can be negative, while the RNA-Seq matrix uses the raw counts, which means the correlation is always close to 1.

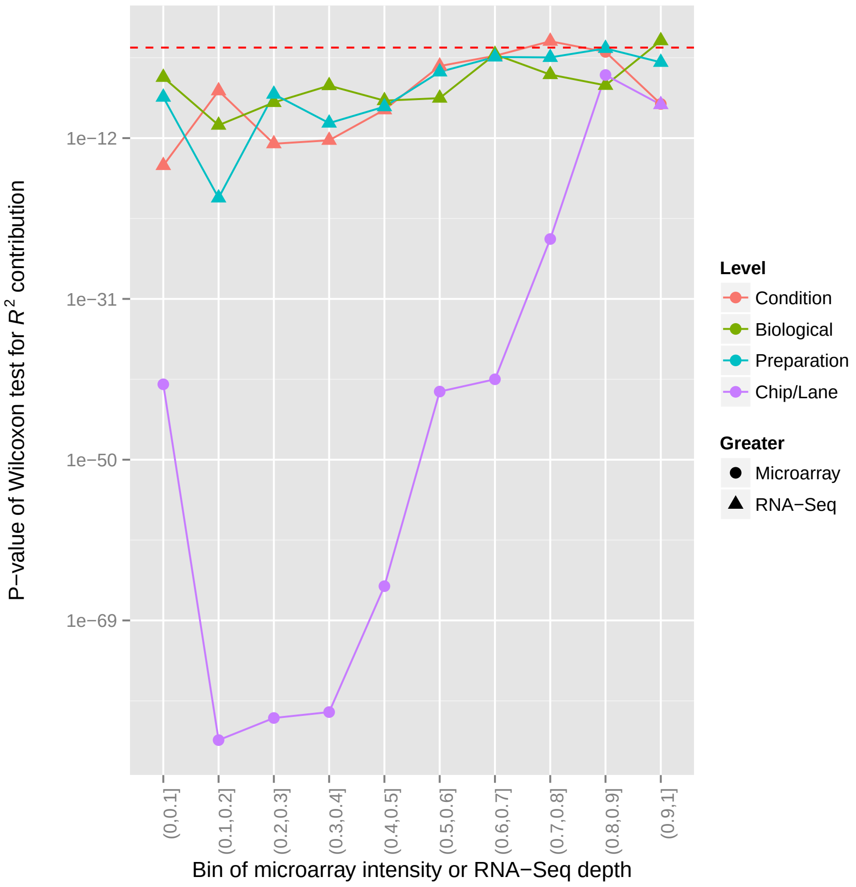

Figure S2: Wilcoxon rank-sum test p-values comparing the percent of variance explained at each level between RNA-Seq and microarrays, in each of 10 bins based on the intensity quantile. The shape of each point indicates whether RNA-Seq or microarrays showed higher R^2 at that stage. The p-values are shown on a log scale, with p = 0.05 shown as a horizontal dashed line.

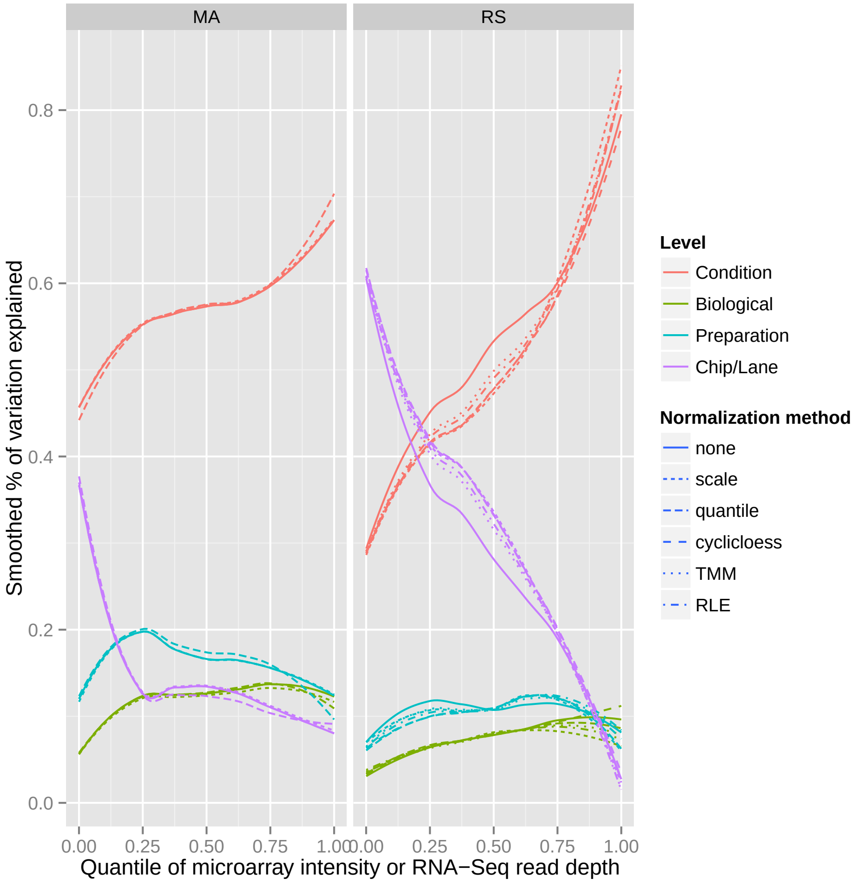

Figure S3: Percent of variance explained by each nested level of the experiment, comparing several normalization methods. “None”, “scale”, “quantile” and “cyclicloess” denote options used in `normalizeBetweenArrays` in the limma R package, TMM denotes the trimmed mean of M-values method of Robinson and Oshlack 2010, and RLE denotes the “relative log expression” method of Anders and Huber 2010. TMM and RLE were applied only to the RNA-Seq data.

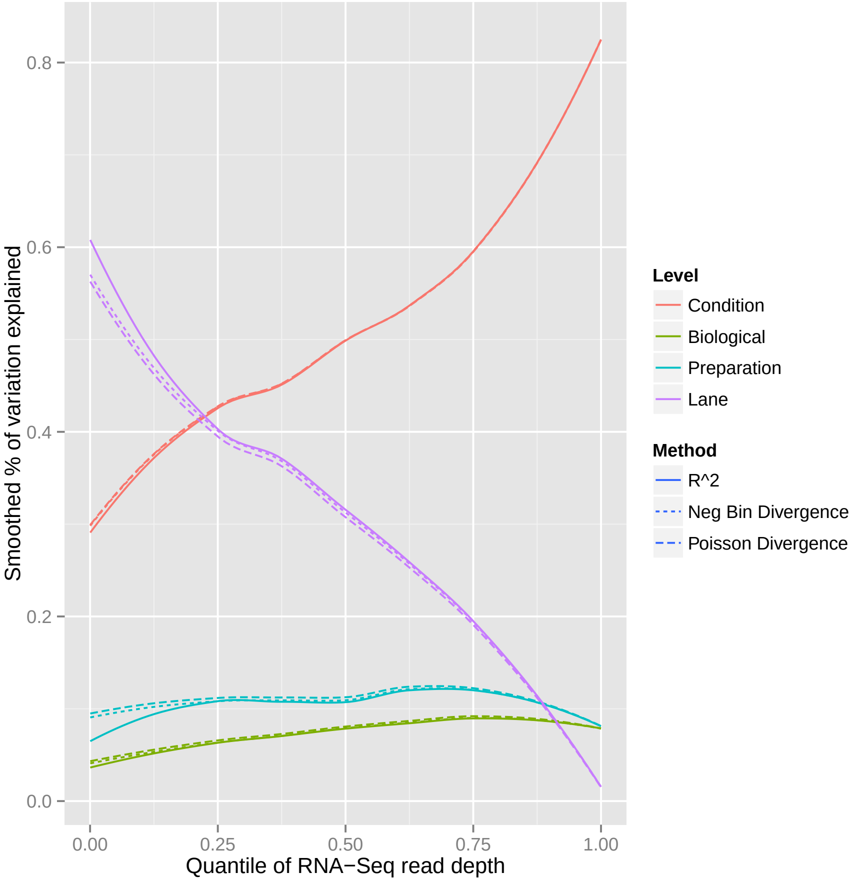

Figure S4: Comparison of an R^2 from ANOVA computed for transformed RNA-Seq, using TMM normalization, to a pseudo-R^2 metric making use of either a Poisson or a negative binomial model.

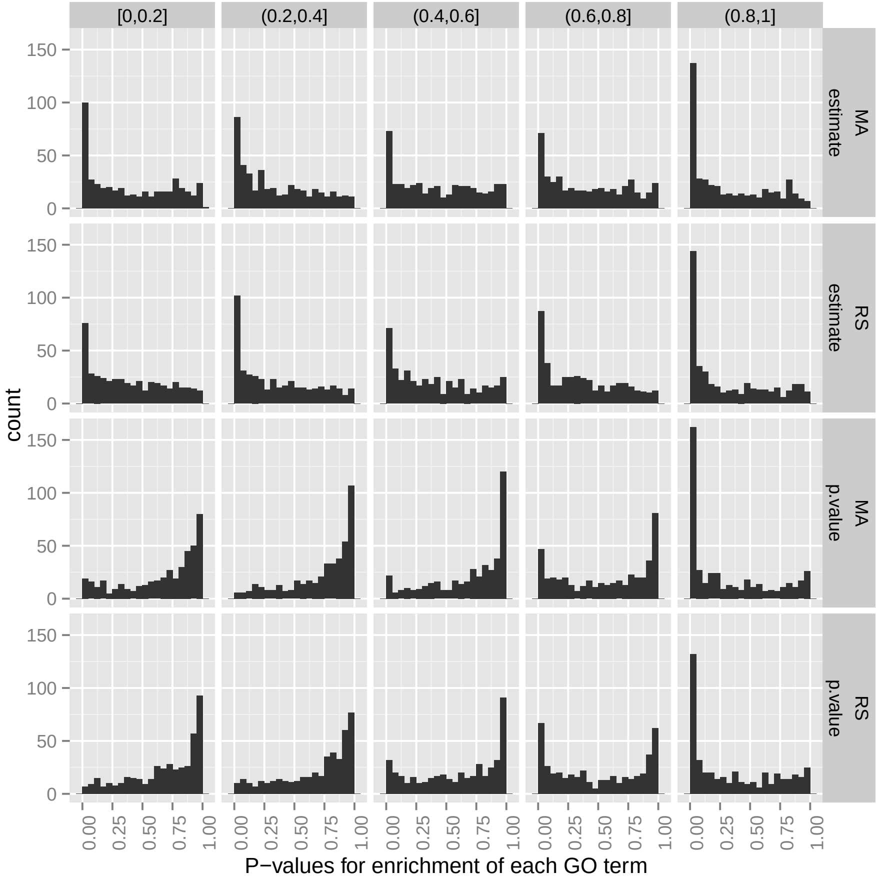

Figure S5: P-value histograms for Wilcoxon tests of gene set enrichment. The plot is divided vertically based on the technology (micrarray or RNA-Seq) and metric (log fold change estimate or p-value) used in the enrichment analysis, and horizontally based on the quantile of the average intensity within each gene.

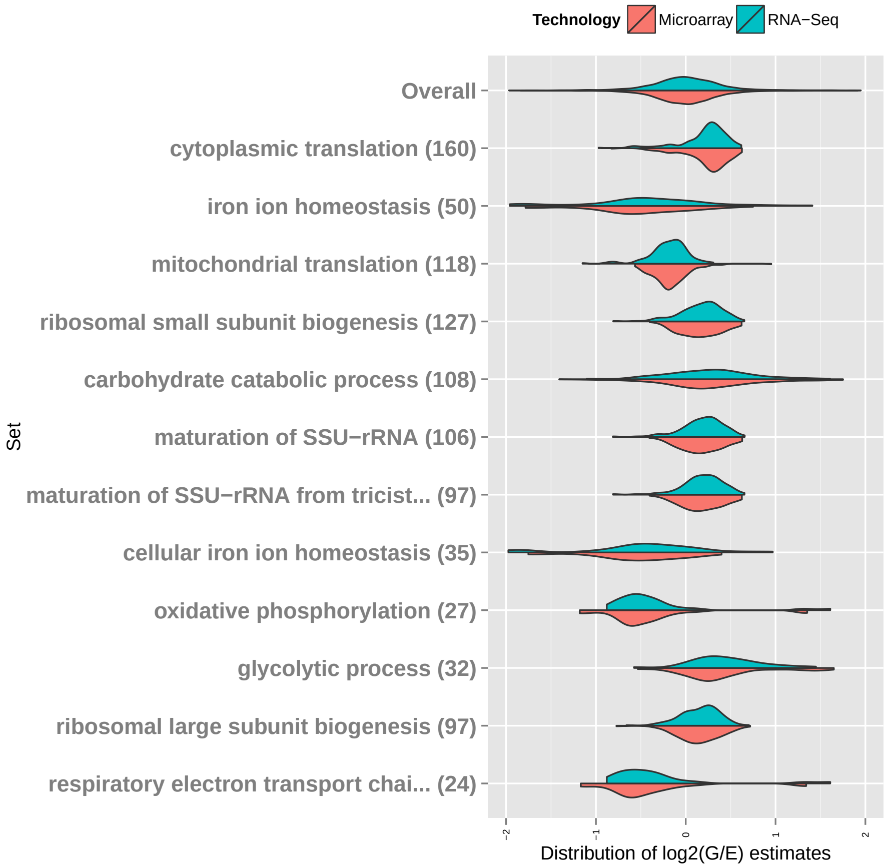

Figure S6: The 12 sets that were found most significantly differentially expressed in microarrays and RNA-Seq, within the Biological Process ontology.

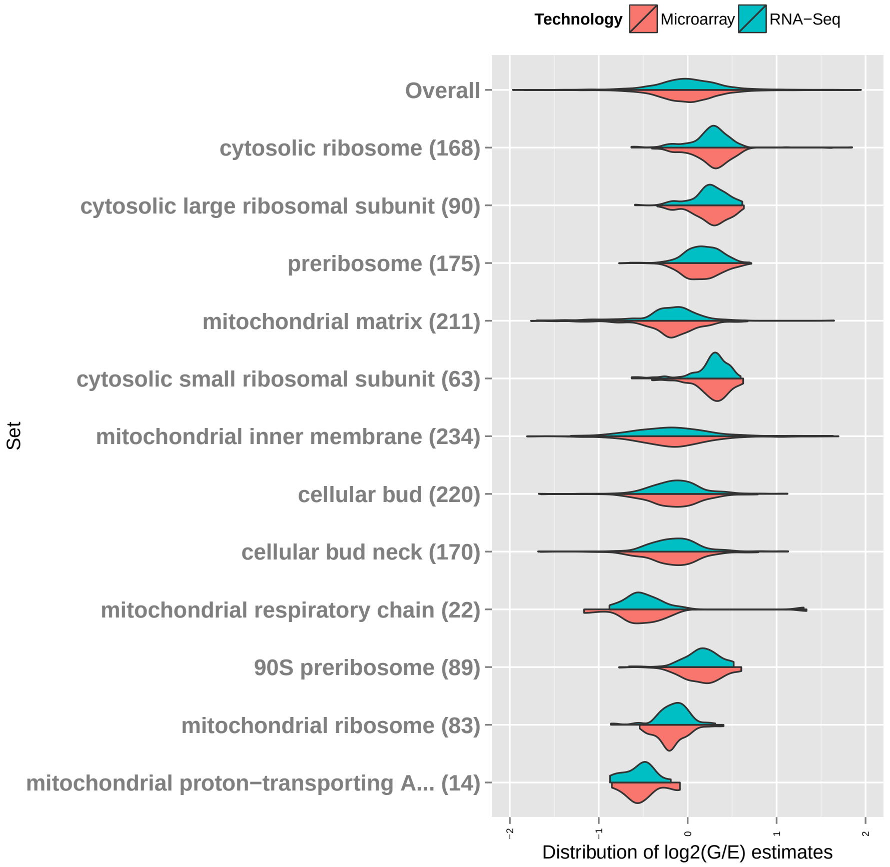

Figure S7: The 12 sets that were found most significantly differentially expressed in microarrays and RNA-Seq, within the Cellular Compartment ontology.

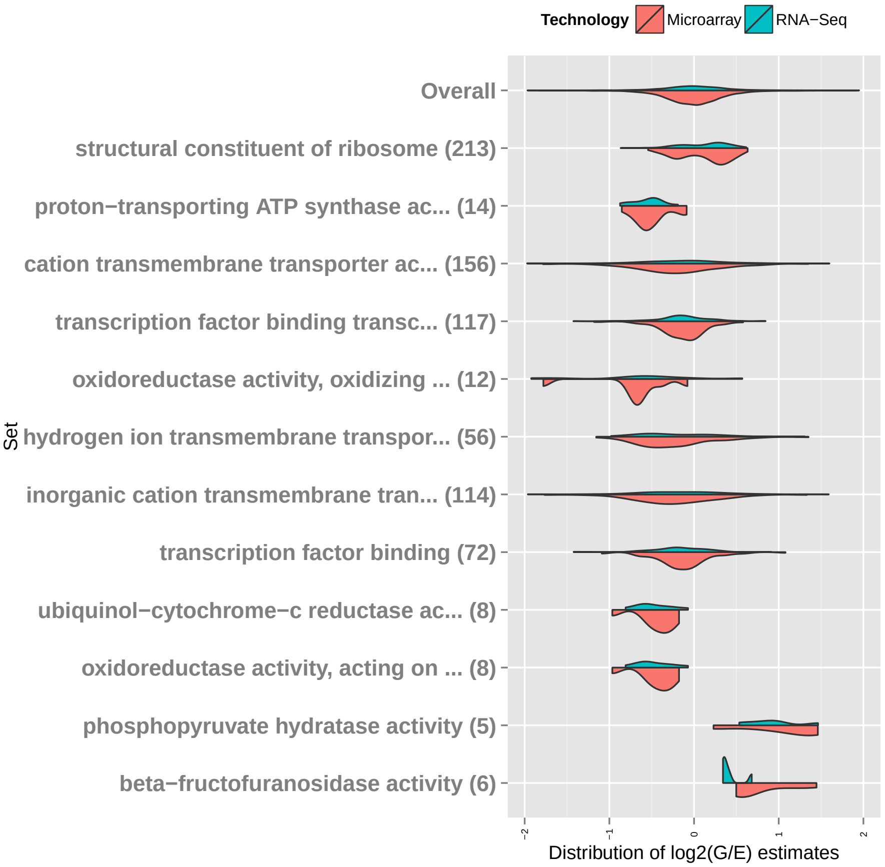

Figure S8: The 12 sets that were found most significantly differentially expressed in microarrays and RNA-Seq, within the Molecular Function ontology.

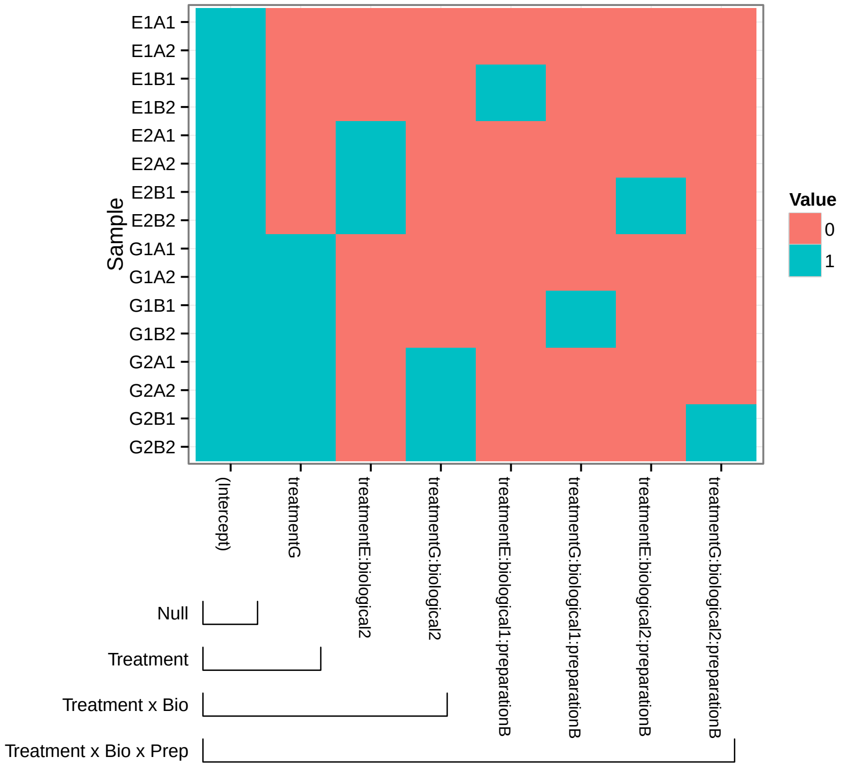

Figure S9: The design matrix X used to fit linear models to the nested experiment and determine R^2. The Null, Treatment, Treatment×Biological, and Treatment×Biological×Preparation models use 1, 2, 4, or 8 columns of this matrix, respectively, as shown.