SUPPLEMENT

SUPPLEMENTAL RESULTS.

Previously excluded ORFs.

Several open reading frames were previously excluded from the original *Halobacterium* genome annotation but were considered here for verification in the proteomics experiments. Four of these ORFs were detected in this proteomics analysis: Chr_ORF0320, Chr_ORF0503, Chr_ORF0960, and Chr_ORF1290. The predicted pI's were within the peak range for *Halobacterium* (pI~4.5). The first three of these were also detected in a previous proteomics experiment in which *Halobacterium* response to gamma radiation was tested (Whitehead et al. 2006). Chr_ORF1290 has been added to the *Halobacterium* genome (baliga.systemsbiology.net) as VNG1390a, which has been annotated as a predicted nucleotidyl transferase due to its match to *Haloarcula marismortui* ADP transferase (PSI-BLAST 68% identity at the amino acid level). The protein quantitation ratios for these proteins are listed in Supplemental Table 3.

Physiological characterization of transcriptome responses to oxygen perturbations.

Two paralagous pathways with opposing responses to oxygen.

Gene duplication is a primary mechanism for the evolution of new functions and for buffering detrimental mutations. There are several such duplicated gene families in *H. salinarum NRC-1*, for example, two paralagous copies of operons encoding ~13 gas vesicle protein genes (*gvp*). At least one of these clusters has been characterized in the

context of the biogenesis of flotation devices called gas vesicles, which provide vertical mobility toward the air-water interfaces (Offner et al. 2000). Gvp cluster 2 was previously thought to be silent (Jones et al. 1989) and expressed only in cluster 1 deletion strains in stationary phase cultures (Englert et al. 1990; Horne and Pfeifer 1989; Offner et al. 1996; Pfeifer et al. 1989). Interestingly, we observed that both gene clusters are transcribed and that their transcription profiles were anticorrelated to each other (Fig. 4A and B, Supplemental Fig. 3). Specifically, *gvp* cluster 1 transcripts were strongly correlated with changes in oxygen (cluster 19, oxic score = 9.2), an observation that coincides with previous studies showing that *gvp1* is expressed primarily during exponential growth phase (Offner et al. 1996; Yang and DasSarma 1990). In contrast, the transcription of *gvp* cluster 2 genes and *gvpD1*, a known repressor of *gvp* cluster 1 transcription (Hofacker et al. 2004), was repressed by oxygen (Fig. 4A and E) (oxic score = -4.4 and -3.3, respectively). We confirmed the translation of both sets of transcripts by observing unique peptides from proteins encoded by both clusters (see Supplemental Table 3). These results suggest differential regulation and possible functional divergence of these clusters.

Possible transcriptional regulatory circuitry governing energy metabolism during oxygen shifts. Our transcriptome data suggest that *H. salinarum NRC-1* appears to anticipate an alternate growth substrate when oxygen is limiting. Transcriptional regulatory circuitry may govern such an anticipatory feature, with likely regulatory protein candidates bearing both light- and oxygen-sensing domains, including Bat (VNG1464G) and Ark (VNG0916G) (Baliga et al., 2002). In this scenario, regulation would link oxygen

availability changes with metabolism via phototrophy, a strategy in keeping with the halobacterial natural environment, where intense sunlight leads to evaporation and reduction in oxygen tension (Robb et al. 1995). On the other hand, although DmsR (VNG0826C) is the most likely candidate for regulation of DMSO reduction (Muller and DasSarma 2005), the link between oxygen availability and the regulation of this process will have to be investigated further.

The translational response of *H. salinarum NRC-1* to changes in oxygen tension.

In previous studies using the ICAT method, coverage was ~15% (Baliga et al. 2002), and previous iTRAQ studies with fewer time points under gamma irradiation conditions detected a total of 43% of the *H. salinarum NRC-1* predicted proteome (Whitehead et al. 2006). Thus, the overall coverage for this study (54%) is the best recorded to date for quantitative microbial shotgun proteomics experiments

For codon adaptation index (CAI) calculations, we used two different methods. First, CAI was calculated using the *H. salinarum NRC-1* ribosomal proteins as the training set (see main text for details). To avoid potential biases inherent in using these genes (they were part of the oxygen-responsive gene set), we also calculated CAI using predicted highly expressed genes for *Halobacterium salinarum NRC-1* obtained from the Optimizer database (Puigbo et al. 2007; http://genomes.urv.es/OPTIMIZER/). Using the codon usage table for *H. salinarum NRC-1* from this database, we obtained similar results to those with the ribosomal proteins as the training set (0.77 for aerobic proteins vs. 0.70 for anaerobic).

It is notable that no known function was found for a significantly higher proportion of proteins encoded by oxygen-anticorrelated transcripts relative to those encoded by oxygen-correlated transcripts (37% vs. 5%) (Fig. 4A and B, Supplemental Table 1). We observed a similar trend of poor detection of the anaerobic proteins in all of the previous *H. salinarum* proteomics studies (P. Van, unpublished data). Approximately 30% of the total (oxygen-correlated and –anticorrelated) unknown function proteins have been detected in the current and in previous *H. salinarum* proteomics studies (Baliga et al. 2002; Goo et al. 2004; Goo et al. 2003; Whitehead et al. 2006), indicating that these are encoded by *bona fide* expressed genes.

Potential underpinnings of the time lagged phenomena observed in mRNA/protein correlations. In order to decipher possible regulatory mechanisms underlying potential posttranscriptional phenomena observed (i.e. time-lagged correlations between mRNA and protein), we examined the sequences of predicted 5' untranslated regions (UTR) of transcripts (Greenbaum et al. 2003) which were considered in the final time lagging analysis (Fig. 5, main text). In these UTR's, we investigated the relative contributions of transcript stability and structure, translation initiation, efficiency, and protein stability (Beyer et al. 2004; Unwin and Whetton 2006) by looking for the following signatures: (i) Shine-Dalgarno ribosome binding consensus sequences, since efficient translation initiation in archaea is known to be promoted by sequences close to the consensus of GGAGGUGA (Londei 2005; Sartorius-Neef and Pfeifer 2004); (ii) riboswitch consensus sequences such as vitamin B12-binding posttranscriptional regulatory structures (Nudler and Mironov 2004); (iii) potentially novel signatures using

the MEME and MAST algorithms (Bailey and Elkan 1994). It is important to note that ~30% of transcripts in *H. salinarum NRC-1* are predicted to be leaderless (Torarinsson et al. 2005). However, even considering the remaining 70% of transcripts with 5'-UTRs, we were surprised that we did not find any evidence of these specific types of posttranscriptional control (data not shown), suggesting that, if there is a regulated mechanism at work, the underlying mechanism is novel, not obvious at the primary sequence level, and will require more sophisticated investigation including experimental measurements of global transcript stability (Selinger et al. 2003; Wang et al. 2002) and protein stability (Pratt et al. 2002).

SUPPLEMENTAL MATERIALS AND METHODS.

Oxygen saturation calculation from www.seabird.com.

xsat(T,S) = exp (A(Ta) + S * B(Ta))= exp([A1 + A2 * (100/Ta) + A3 * ln(Ta/100) + A4 * (Ta/100)+ S * [(B1 + B2 * (Ta/100) + B3 * (Ta/100)2)])

Proteomics data analysis.

Documentation for using Libra within the Trans-Proteomic Pipeline (TPP) can be found at: http://sashimi.cvs.sourceforge.net/sashimi/trans_proteomic_pipeline/src/Quantitation/Libra/docs/libra_info.html. Spearman rank correlations for replicate analysis represented in Supplemental Figure 5 are as follows: 360 minute time point = 0.425; 477 minute time point = 0.382. R^2 values reported on the charts are the square of the Spearman correlations, which are robust to outliers.

mRNA data analysis. The "oxic score" described in the main text and in Supplemental Figure 2 is derived from the correlation between the position of a cluster's expression profile along PC1 with the rate of change in the oxygen profile with rate of change in the expression profiles (the Inferelator function; Bonneau et al. 2006). i.e. We observed that the closer a cluster's position to the extreme ends of the PC1 axis, the better its dynamics were predicted by the change in oxygen. The dynamics of the rate of change in a cluster's expression profile was given by the function: y+tau*dy/dt; where y is the initial expression ratio of the given cluster; dy/dt is the rate of change in y over the time interval, tau. Tau is is a constant based partly on the experimentally measured rate of mRNA degradation for prokaryotes and partly on a complex modeling algorithm (Bernstein et al. 2002; Bonneau et al. 2006). We observed that PC1 correlates well with this function (Supplemental Fig. 2). Thus, the position of the cluster along PC1 was used as the "oxic score", which is therefore a strong indicator of oxygen-responsiveness.

mRNA/mRNA time lagging analysis.

To assess the degree of temporal coherence in the transcriptome data (Supplemental Fig. 4), genes were grouped into operons, and an mRNA/mRNA time-lagged correlation analysis over time lags between -40 and +40 minutes was performed on these operons in a similar fashion as described in the main manuscript for mRNA/protein correlations. We found that for pairs of operons which had high time-lagged correlations (> 0.7), the time lags had a mean absolute deviation of ± 1.5 minutes, indicating that they have very similar temporal responses.

SUPPLEMENTAL FIGURE LEGENDS.

Supplemental Figure 1. Plots of principal components in comparison to oxygen profiles demonstrate that ~65% of the variance (principal components 1 and 3) can be explained by oxygen. Eigenvalues for the 10 most significant components are shown in the upper left. Projections corresponding to the first 8 principal components are shown in the ensuing graphs. The projection of the transcriptome data at each of the 61 time points in the three chemostat experiments is represented by the black line. The red line in each graph depicts the oxygen tension change at each of the 61 time points (not drawn to scale).

Supplemental Figure 2. Projection of cluster profiles on the first principal component correlates strongly with a model for the influence of oxygen levels on dynamics. The influence of oxygen on the rate of change of each transcript profile (See Methods, main text, and Supplemental Methods) (X axis) is plotted against the relative position of each cluster along the 1st component (oxygen) of the PCA plot (Y axis, "oxic" score). Clusters which are significantly anticorrelated with oxygen appear in the lower left corner (negative correlation between rate of change of oxygen and rate of transcription change) and those significantly correlated with oxygen appear in the upper right corner (positive correlation). Cluster numbers 5, 10, 18, 19, 24, and 28; depicted in color; were found to be most significantly affected by oxygen and were pursued for further analysis. These clusters are plotted in the same colors in the main text Figure 3.

Supplemental Figure 3. Gas vesicle biogenesis gene clusters 1 and 2 are anticorrelated with each other, suggesting novel functions for the protein complexes encoded by these clusters. Transcript profiles for each gene in the two clusters are color coded: see *legend*. Time scales in minutes for all three chemostat experiments are plotted on the X-axis. The change in oxygen tension during the three experiments is depicted by the black line (see *legend*).

Supplemental Figure 4. mRNA expression profiles for all clusters included in the K-means analysis (main text Figure 3). The upper left graph depicts the oxygen tension changes that occurred during the three chemostat experiments. In each cluster graph, time points for each of the three chemostat experiments are plotted on the X-axis and log10 expression ratio on the Y-axis. Genes contained within each cluster are listed in Supplemental Table 4.

Supplemental Figure 5. Assessment of the reproducibility of iTRAQ proteomics quantitation on a large scale. (A) 360 minute time point. Top: Venn diagram depicting the number of proteins detected in set 1 (replicate 1 labeled with 117 Da isobaric tag, represented in the Venn diagram by a red circle), vs. set 10 (labeled with 115 Da isobaric tag, represented in the diagram by a green circle). 1019 proteins total proteins were detected in both sets of the 360 minute time point combined. The 588 proteins that were commonly detected in both replicates are depicted at bottom as a scatter plot comparing the quantitation ratio for each replicate. Each protein log10 ratio quantitation is

represented by a blue dot on the scatter plot. The black line represents the R^2 value (Spearman correlation coefficient squared). **(B)** 477 minute time point, which is organized similarly to (A). This analysis of replicates showed that the reproducibility of the iTRAQ method [Coefficient of variation (CV) of 18%] on this unprecedented large scale was comparable to other established quantitative methods such as ICAT (CV~20%) (Molloy et al. 2005).

Supplemental Figure 6. Proteins encoded by aerobically induced transcripts were significantly overrepresented in the proteomics dataset. (A) *Percent of predicted peptides detected for the proteins encoded by genes in the oxygen-responsive list (215 oxygen-correlated and –anticorrelated transcripts) from each iTRAQ multiplex set (1-10).* Predicted peptides are based on the annotated *H. salinarum NRC-1* genome sequence (baliga.systemsbiology.net). Peptides from proteins whose transcripts were correlated with oxygen (aerobic) are represented by the dark grey bars, whereas peptides from proteins whose transcripts were anticorrelated with oxygen (anaerobic) are represented by the white bars. X-axis numbers 1 through 10 correspond to each of the iTRAQ multiplex sets (see also main text Table II). (B) *Oxygen level in the first chemostat experiment (in which both mRNA and protein samples were extracted).* Grey and white striping demarcates each set (comprised of three conditions each; Table II). Each thin black bar represents the oxygen concentration at each time point in which protein was extracted.

Supplemental Figure 7. Time lags between transcription and translation were calculated for individual genes. Seven examples for genes subjected to peak time lagged correlation (PTLC) analysis were selected on the basis of *p*-value in permutation tests ($p < 0.01$). Two plots are shown for each gene example: quantitative iTRAQ protein data ($\log_{10}$ ratios; filled circles) and mRNA levels (empty circles) are shown on the left (e.g. A), and permutation test results to the right (e.g. B). Cubic spline interpolated profiles for the protein data (orange line), and mRNA in (blue line) are overlaid; see *legend*. The identity of genes in each example is depicted in the legend for the left mRNA/protein comparison graph and above each permutation test graph.

Supplemental Figure 8. Time lags between transcription and translation were calculated for operons. Selected examples for operons subjected to PTLC and permutation tests. mRNA/protein profiles and permutation test graphs are laid out as in Supplemental Figure 7. This analysis was performed similarly to that for individual genes. See legends for gene identities and color coding.

SUPPLEMENTAL REFERENCES

Bailey, T.L. and Elkan, C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2:** 28-36.

Baliga, N.S., Pan, M., Goo, Y.A., Yi, E.C., Goodlett, D.R., Dimitrov, K., Shannon, P., Aebersold, R., Ng, W.V. and Hood, L. 2002. Coordinate regulation of energy transduction modules in Halobacterium sp. analyzed by a global systems approach. *Proc. Natl. Acad. Sci. U S A* **99:** 14913-14918.

Bernstein, J.A., Khodursky, A.B., Lin, P.H., Lin-Chao, S. and Cohen, S.N. 2002. Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proc. Natl. Acad. Sci. U S A* **99:** 9697-9702.

Beyer, A., Hollunder, J., Nasheuer, H.P. and Wilhelm, T. 2004. Post-transcriptional expression regulation in the yeast *Saccharomyces cerevisiae* on a genomic scale. *Mol. Cell Proteomics* **3:** 1083-1092.

Bonneau, R., Reiss, D.J., Shannon, P., Facciotti, M., Hood, L., Baliga, N.S. and Thorsson, V. 2006. The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. *Genome Biol.* **7:** R36.

Englert, C., Horne, M. and Pfeifer, F. 1990. Expression of the major gas vesicle protein gene in the halophilic archaebacterium *Haloferax mediterranei* is modulated by salt. *Mol. Gen. Genet.* **222:** 225-232.

Goo, Y.A., Roach, J., Glusman, G., Baliga, N.S., Deutsch, K., Pan, M., Kennedy, S., DasSarma, S., Ng, W.V. and Hood, L. 2004. Low-pass sequencing for microbial comparative genomics. *BMC Genomics* **5:** 3.

Goo, Y.A., Yi, E.C., Baliga, N.S., Tao, W.A., Pan, M., Aebersold, R., Goodlett, D.R., Hood, L. and Ng, W.V. 2003. Proteomic analysis of an extreme halophilic archaeon, *Halobacterium sp. NRC-1. Mol. Cell Proteomics* **2:** 506-524.

Greenbaum, D., Colangelo, C., Williams, K. and Gerstein, M. 2003. Comparing protein abundance and mRNA expression levels on a genomic scale. *Genome Biol.* **4:** 117.

Hofacker, A., Schmitz, K.M., Cichonczyk, A., Sartorius-Neef, S. and Pfeifer, F. 2004. GvpE- and GvpD-mediated transcription regulation of the *p-gvp* genes encoding gas vesicles in *Halobacterium salinarum*. *Microbiology* **150:** 1829-1838.

Horne, M. and Pfeifer, F. 1989. Expression of two gas vacuole protein genes in *Halobacterium halobium* and other related species. *Mol. Gen. Genet.* **218:** 437-444.

Jones, J.G., Hackett, N.R., Halladay, J.T., Scothorn, D.J., Yang, C.F., Ng, W.L. and DasSarma, S. 1989. Analysis of insertion mutants reveals two new genes in the pNRC100 gas vesicle gene cluster of *Halobacterium halobium*. *Nucleic Acids Res.* **17:** 7785-7793.

Londei, P. 2005. Evolution of translational initiation: new insights from the archaea. *FEMS Microbiol. Rev.* **29:** 185-200.

Molloy, M.P., Donohoe, S., Brzezinski, E.E., Kilby, G.W., Stevenson, T.I., Baker, J.D., Goodlett, D.R., Gage, D.A. 2005. Large-scale evaluation of quantitative reproducibility and proteome coverage using acid cleavable isotope coded affinity tag mass spectrometry for proteomic profiling. *Proteomics* **5:** 1204-1208.

Nudler, E. and Mironov, A.S. 2004. The riboswitch control of bacterial metabolism. *Trends Biochem. Sci.* **29:** 11-17.

Offner, S., Hofacker, A., Wanner, G. and Pfeifer, F. 2000. Eight of fourteen gvp genes are sufficient for formation of gas vesicles in halophilic archaea. *J. Bacteriol.* **182:** 4328-4336.

Offner, S., Wanner, G. and Pfeifer, F. 1996. Functional studies of the *gvpACNO* operon of *Halobacterium salinarium* reveal that the GvpC protein shapes gas vesicles. *J. Bacteriol.* **178:** 2071-2078.

Pfeifer, F., Blaseio, U. and Horne, M. 1989. Genome structure of *Halobacterium halobium*: plasmid dynamics in gas vacuole deficient mutants. *Can J. Microbiol.* **35:** 96-100.

Pratt, J.M., Petty, J., Riba-Garcia, I., Robertson, D.H., Gaskell, S.J., Oliver, S.G. and Beynon, R.J. 2002. Dynamics of protein turnover, a missing dimension in proteomics. *Mol. Cell Proteomics* **1:** 579-591.

Sartorius-Neef, S. and Pfeifer, F. 2004. In vivo studies on putative Shine-Dalgarno sequences of the halophilic archaeon *Halobacterium salinarum*. *Mol. Microbiol.* **51:** 579-588.

Selinger, D.W., Saxena, R.M., Cheung, K.J., Church, G.M. and Rosenow, C. 2003. Global RNA half-life analysis in *Escherichia coli* reveals positional patterns of transcript degradation. *Genome Res.* **13:** 216-223.

Torarinsson, E., Klenk, H.P. and Garrett, R.A. 2005. Divergent transcriptional and translational signals in Archaea. *Environ. Microbiol.* **7:** 47-54.

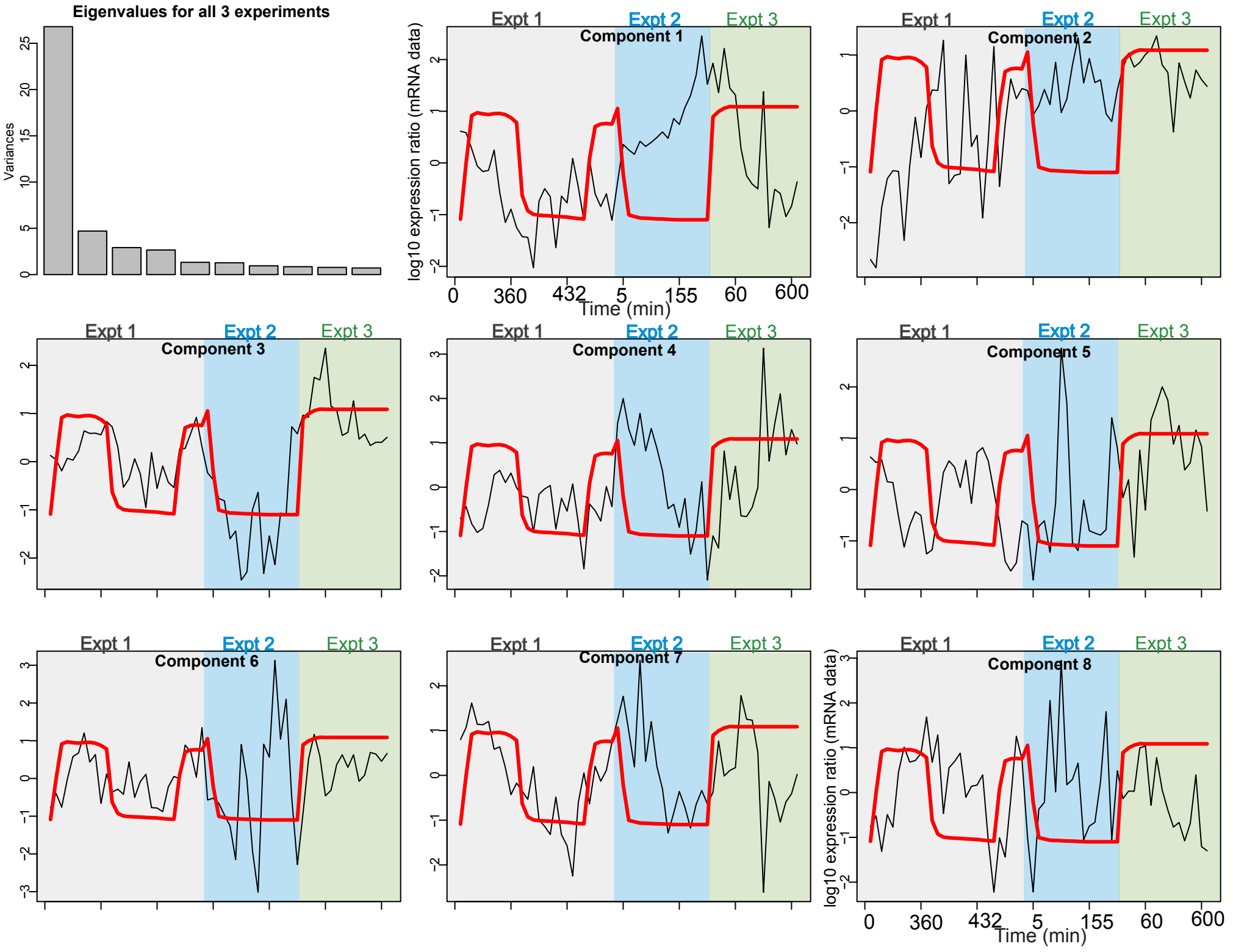
Unwin, R.D. and Whetton, A.D. 2006. Systematic Proteome and Transcriptome Analysis of Stem Cell Populations. *Cell Cycle* **5**.

Wang, Y., Liu, C.L., Storey, J.D., Tibshirani, R.J., Herschlag, D. and Brown, P.O. 2002. Precision and functional specificity in mRNA decay. *Proc. Natl. Acad. Sci. U S A* **99:** 5860-5865.

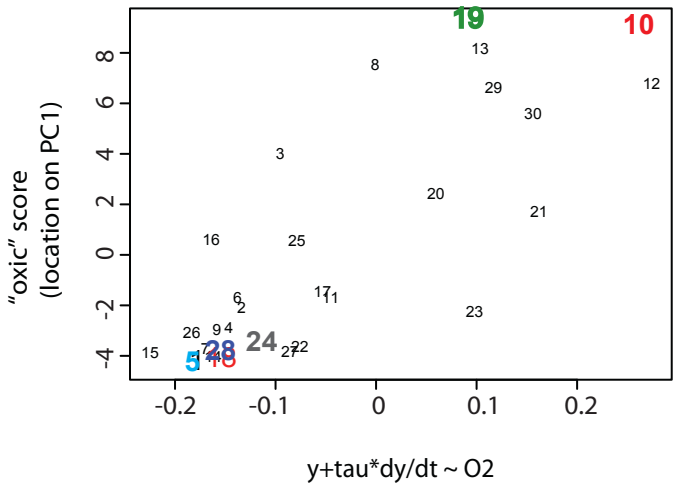
Whitehead, K., Kish, A., Pan, M., Kaur, A., Reiss, D.J., King, N., Hohmann, L., Diruggiero, J. and Baliga, N.S. 2006. An integrated systems approach for understanding cellular responses to gamma radiation. *Mol. Syst. Biol.* **2:** 47.

Yang, C.F. and DasSarma, S. 1990. Transcriptional induction of purple membrane and gas vesicle synthesis in the archaebacterium *Halobacterium halobium* is blocked by a DNA gyrase inhibitor. *J. Bacteriol.* **172:** 4118-4121.

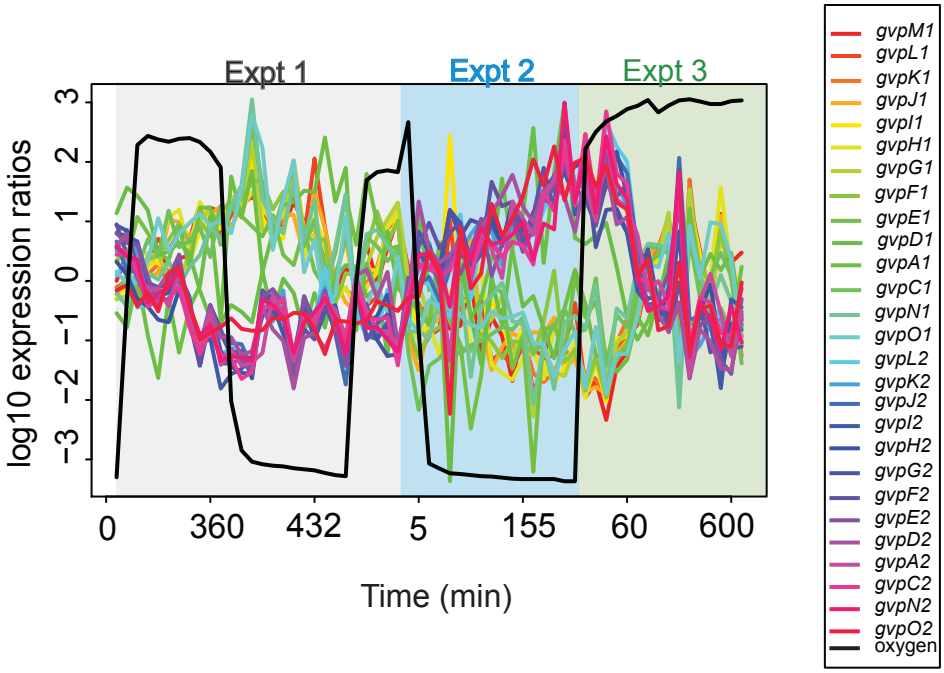
Supplemental Figure 1.



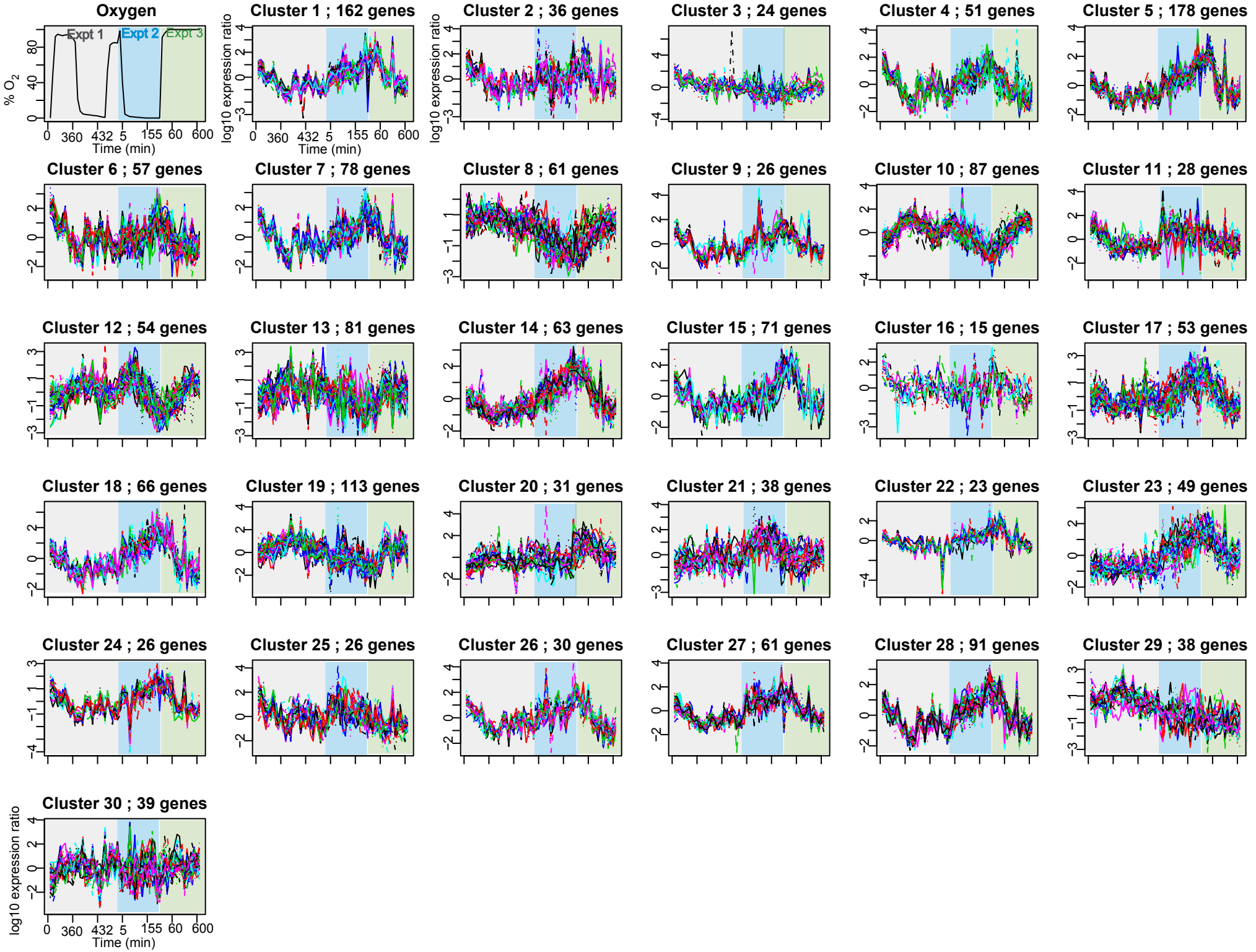
Supplemental Figure 2.



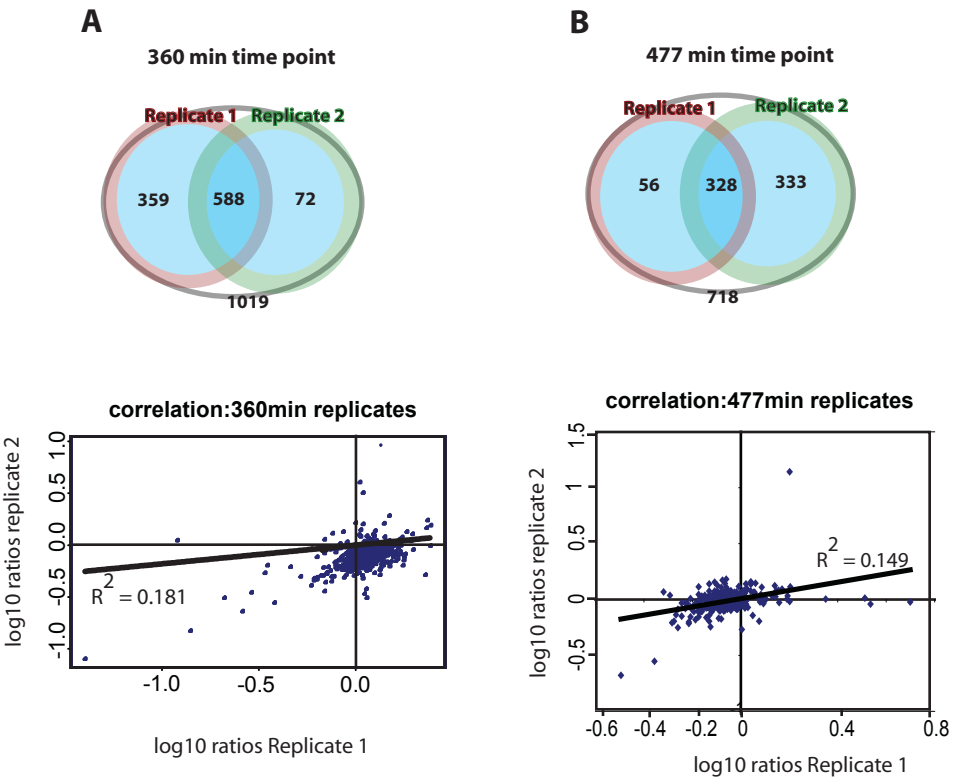
Supplemental Figure 3.



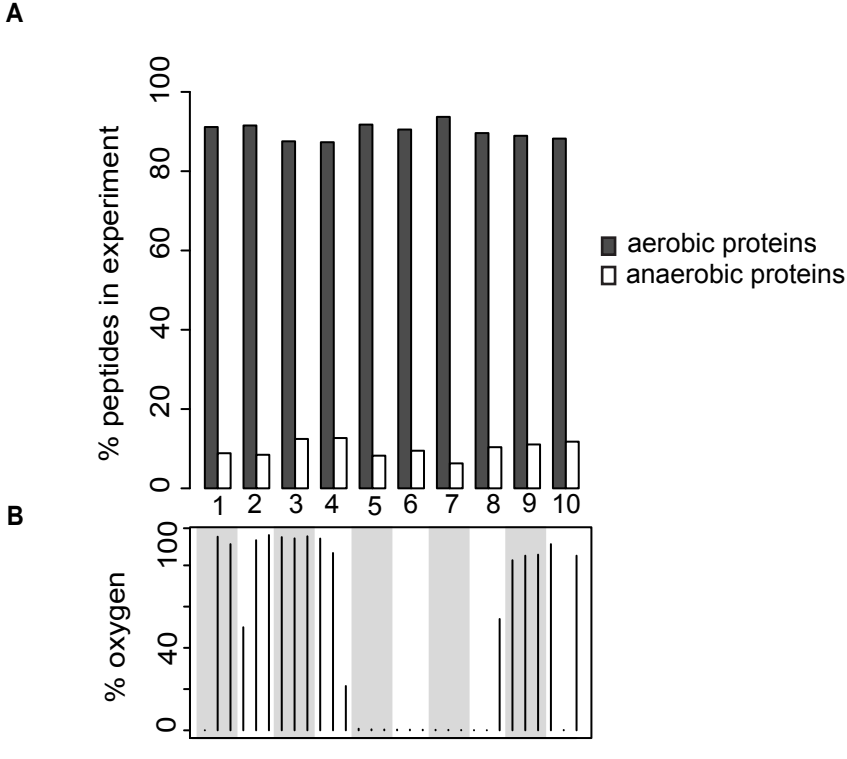
Supplemental Figure 4.



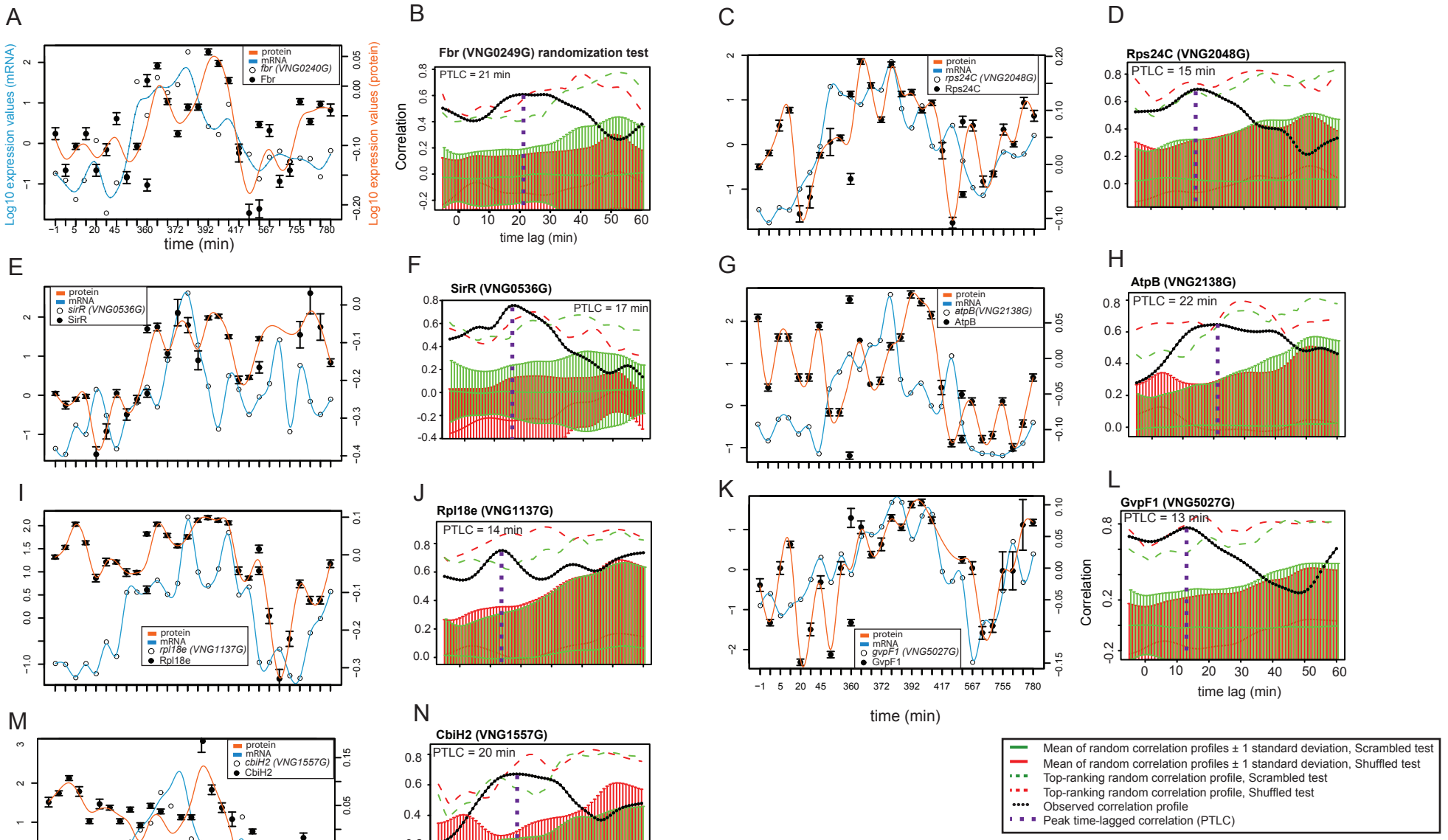
Supplemental Figure 5.



Supplemental Figure 6.



Supplemental Figure 7.



Supplemental Figure 8.