4ZEV-SNF1 Microarray Analysis

Danielle Carpenter

August 27, 2012

These arrays use RNA isolated from samples collected on August 8, 2012 from chemostats growing at a dilution rate of $0.18hr^{-1}$ in both phosphate- and carbon-limited media. DBY12416 is the zinc-finger control strain which contains no zinc-finger binding sites, and DBY12424 is derived from DBY12416 but contains the zinc-finger binding sites inserted in front of the SNF1 transcription start site.

Visualizing the Array Data

The array data was downloaded from PUMA using the filters "Final Processed Red Intensity ¿= 350" and "Final Processed Green Intensity ¿= 350" and imported into R. The file contains 5157 rows of expression data. Several genes (YFL007W, YGR271C-A, YJL019W, YFR024C-A, YJL012C, —— , YBR074W, YOR087W, —— , —— , —— , YJL016W) appeared to be repeated multiple times in the data and were removed completely, resulting in expression data for 5142 genes.

Missing values were computed using the knn imputation function provided in the impute package from Bioconductor.

The data are then zero-normalized, and any genes with less than 2-fold change are dropped.

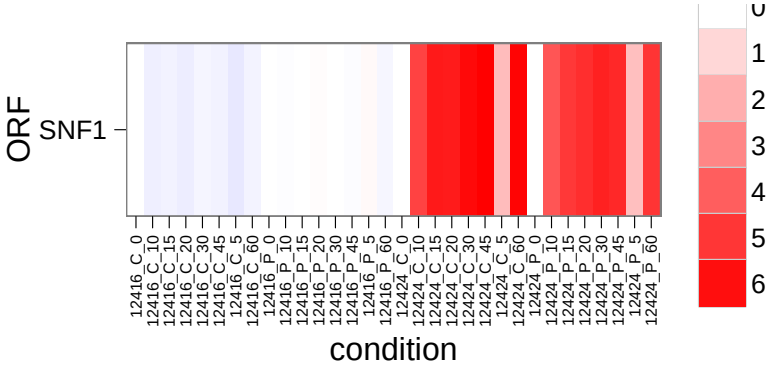
There are 1039 genes which change by more than 2-fold across all four conditions. Of these, 81 change by two-fold in the phosphate-limited conditions, and 999 in the carbon limited conditions. The genes which change more than two-fold in both conditions are listed below:

```
two.fold = data.frame(ORF = intersect(rownames(diff.p.dat), rownames(diff.c.dat)))
two.fold$name = as.character(orf2name(two.fold$ORF))
two.fold$name[which(two.fold$name == "NULL")] = "-"
two.fold$desc = orf2desc(two.fold$ORF)
tab = xtable(two.fold, label = "two.fold", caption = "Genes whose expression changes by
at least two-fold in both carbon and phosphate-limiting conditions.")
print(tab, tabular.environment = "longtable", floating = FALSE, include.rownames = FALSE)
```

ORF	name	desc
YNR034W-A	35	Putative protein of unknown function; expression is regulated by Msn2p/Msn4p
YFR053C	HXK1	Hexokinase isoenzyme 1, a cytosolic protein that catalyzes phosphorylation of glucose during g
YER015W	FAA2	Medium chain fatty acyl-CoA synthetase, activates imported fatty acids; accepts a wide range
YLL018C-A	COX19	Protein required for cytochrome c oxidase assembly, located in the cytosol and mitochondrial i
YGL208W	SIP2	One of three beta subunits of the Snf1 serine/threonine protein kinase complex involved in the
YCR020C-A	MAK31	Non-catalytic subunit of N-terminal acetyltransferase of the NatC type; required for replicatio
YCR041W	5	Dubious open reading frame unlikely to encode a protein, based on available experimental and
YER053C-A	15	Putative protein of unknown function; green fluorescent protein (GFP)-fusion protein localizes
YLR327C	TMA10	Protein of unknown function that associates with ribosomes; putative homolog of the F1F0-AT
YOL052C-A	DDR2	Multistress response protein, expression is activated by a variety of xenobiotic agents and envi
Q0130	OLI1	F0-ATP synthase subunit c (ATPase-associated proteolipid), encoded on the mitochondrial ge

YMR105C	PGM2	Phosphoglucomutase, catalyzes the conversion from glucose-1-phosphate to glucose-6-phosphat
YKL053W	26	Dubious open reading frame unlikely to encode a protein, based on available experimental and
YMR303C	ADH2	Glucose-repressible alcohol dehydrogenase II, catalyzes the conversion of ethanol to acetaldehyd
YIL136W	OM45	Protein of unknown function, major constituent of the mitochondrial outer membrane; located
YDR343C	HXT6	High-affinity glucose transporter of the major facilitator superfamily, nearly identical to Hxt7p
YDR536W	STL1	Glycerol proton symporter of the plasma membrane, subject to glucose-induced inactivation, st
YDL184C	RPL41A	Ribosomal protein L47 of the large (60S) ribosomal subunit, identical to Rpl41Bp and has simi
YBL106C	SRO77	Protein with roles in exocytosis and cation homeostasis; functions in docking and fusion of pos
YJL161W	FMP33	Putative protein of unknown function; the authentic, non-tagged protein is detected in highly p
YJL188C	BUD19	Dubious open reading frame, unlikely to encode a protein; not conserved in closely related Sacc
YDR342C	HXT7	High-affinity glucose transporter of the major facilitator superfamily, nearly identical to Hxt6p
YDR216W	ADR1	Carbon source-responsive zinc-finger transcription factor, required for transcription of the gluc
YFR015C	GSY1	Glycogen synthase with similarity to Gsy2p, the more highly expressed yeast homolog; expressi
YHL024W	RIM4	Putative RNA-binding protein required for the expression of early and middle sporulation gene
YDL222C	FMP45	Integral membrane protein localized to mitochondria (untagged protein); required for sporulati
YKL150W	MCR1	Mitochondrial NADH-cytochrome b5 reductase, involved in ergosterol biosynthesis
YGL088W	19	Dubious open reading frame unlikely to encode a protein, based on available experimental and
YDL233W	8	Putative protein of unknown function; green fluorescent protein (GFP)-fusion protein localizes
YLR109W	AHP1	Thiol-specific peroxiredoxin, reduces hydroperoxides to protect against oxidative damage; funct
YOR028C	CIN5	Basic leucine zipper (bZIP) transcription factor of the yAP-1 family; physically interacts with
YER067W	RGI1	Protein of unknown function involved in energy metabolism under respiratory conditions; prote
YHR127W	22	Protein of unknown function; localizes to the nucleus; required for asymmetric localization of K
YOR143C	THI80	Thiamine pyrophosphokinase, phosphorylates thiamine to produce the coenzyme thiamine pyro
YDR477W	SNF1	AMP-activated serine/threonine protein kinase found in a complex containing Snf4p and memb
YOR004W	UTP23	Component of the small subunit processome, involved in 40S ribosomal subunit biogenesis; inte
YBL001C	ECM15	Non-essential protein of unknown function, likely exists as tetramer, may be regulated by the b
YNL014W	HEF3	Translational elongation factor EF-3; paralog of YEF3 and member of the ABC superfamily; st
YPR160W	GPH1	Non-essential glycogen phosphorylase required for the mobilization of glycogen, activity is regu
YLR162W	30	Putative protein of unknown function; overexpression confers resistance to the antimicrobial pe
YPL223C	GRE1	Hydrophilin of unknown function; stress induced (osmotic, ionic, oxidative, heat shock and hea

There is significantly more signal in the 4ZEV control strain than expected, in both conditions. My first thought was that the samples were mislabeled, and I had accidentally labeled the experimental strain twice; however, looking at the expression data for *SNF1*, you can see clearly that *SNF1* expression changes only in the experimental strains and not in the control strains.



```
## Loading required package: proxy
## Attaching package: 'proxy'
## The following object(s) are masked from 'package:stats':
##
## as.dist, dist
```

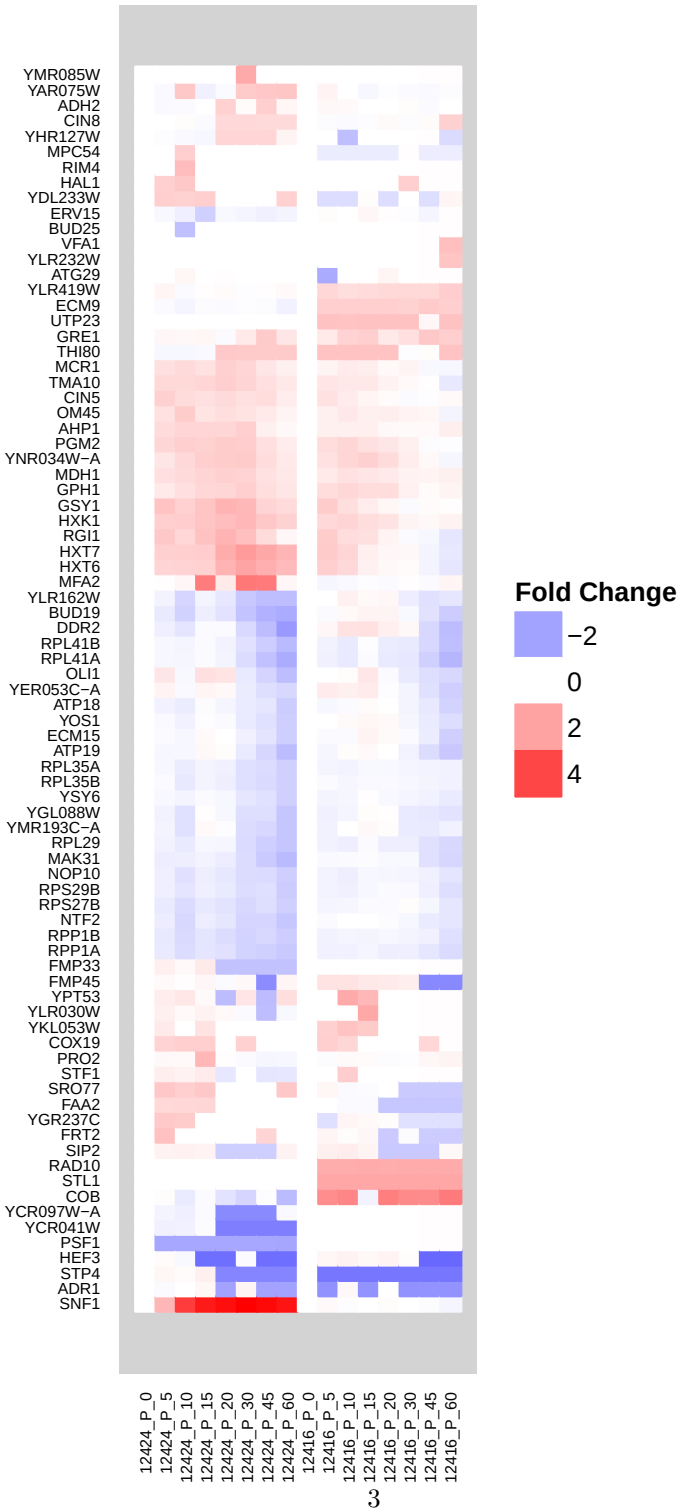


Figure 1: Phosphate-limited expression data. DBY12424 on left, DBY12416 (4ZEV control) on right.

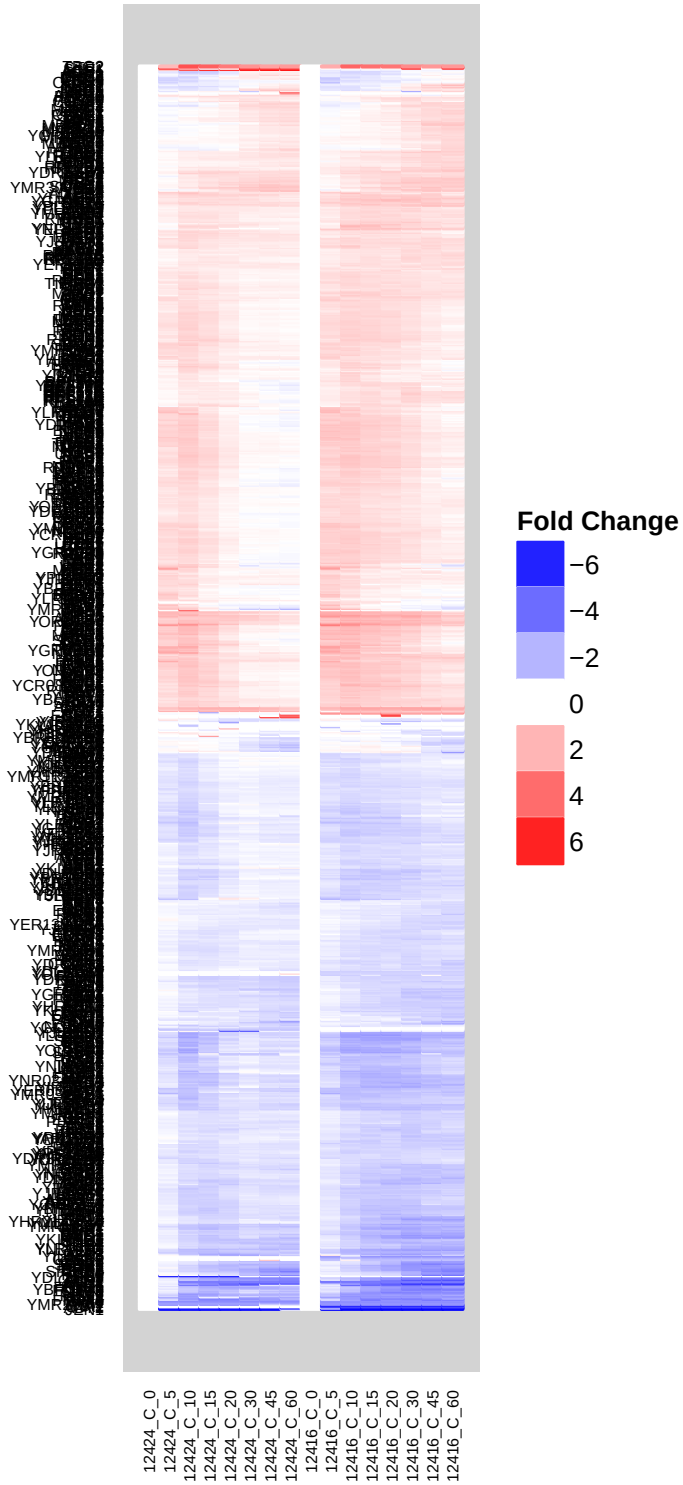


Figure 2: Carbon-limited expression data. 4ZEVpr-SNF1 on left, 4ZEV control on right.

Analysis of the 4-ZEV Arrays

I'm not sure what to make of the control expression data. Obviously, someaspect of my experimental conditions is causing a significant gene expression response in both my control and experimental data. Because this response is more significant in my carbon-limited data (i.e., many more genes respond), I'm inclined to say that this expression change is due to some change in carbon-source availability. My preliminary guess is that the process of sampling (which I had tried to minimize by taking only 5 mL of culture for each timepoint rather than the standard 10 mL) is increasing the concentration of glucose available in the vessel, causing this gene expression response. This effect may be more significant at slower growth rates.

I decided to look at the zero timepoints for each chemostat, to see which genes were very highly differentially expressed across the conditions. The genes which make up the most significantly repressed cluster in the carbon-limited chemostats are *PHO5*, *PHM6*, *SPL2*, *PHO11*, *PHO12*, and *PHO84*. They are expressed at a much lower level than my reference RNA, a phosphate-limited chemostat at $0.15hr^{-1}$. A cluster of genes which are highly-expressed in carbon-limited chemostats when compared to phosphate-limited chemostats includes *HXT7*, *HXT6*, *ALD4*, and *ADH2*, among other genes. There is not a significant difference in there expression in 4ZEVpr-SNF1 when compared to the 4ZEV control, indicating that these genes can be upregulated even in the absence of Snf1p. A quick GO-slim search on the cluster shows enrichment for transmembrane transport, carbohydrate metabolic process, generation of precursor metabolites and energy, and cofactor metabolic process.

I decided to see what happened to this cluster of 35 genes that were already highly expressed at time zero in my carbon-limited samples. (Figure 4) Over the course of the experiment, many of these genes are highly repressed in the carbon-limited vessels (with very similar expression patterns in both experimental and control), while there is not significant response in the phosphate-limited chemostats (with the exception of *HXT7* and *HXT6*). This repression of the cluster strongly resembles the typical response to a pulse of glucose, lending support to my idea that sampling the carbon-limited chemostats is increasing the available glucose, triggering a gene expression response.

SVD of 4-ZEV arrays

I decided to see if subtracting the high control background would produce anything useful. There are three main eigengenes for both carbon and phosphate-limited control arrays, and they appear to be the reverse of each other (e.g., repression in carbon-limitation becomes activation in phosphate-limitation). It should also be noted that these three eigengenes account for a much smaller portion of the total expression fraction in the phosphate-limited chemostats, supporting the idea that the effect of sampling is much larger in carbon-limited chemostats than phosphate-limited chemostats. (Figures 5 and 6)

Correlation of gene expression with eigengenes.

I would like to know which genes in each of the 4ZEV arrays correspond to each of the top three eigengenes. To do so, I will correlate each gene's expression pattern with each of the top three eigengenes.

There appear to be two eigengenes that explain the majority of the variance in the carbon-limited 4ZEV chemostat time course. The main is a strong repression signal, and the second appears to be brief induction followed by lasting repression. When I identified genes whose signal was correlated with each eigengene (using a p-value of 0.05 calculated by permutation testing), eigenvector 2 appears to be actually made up of two different clusters: one cluster consistening of genes who appear to be undergoing a delayed repression, and one of a rapid and transient induction.

I used SGD's GOSlim tool to test for GO term enrichment in each group, and found that many of the genes in clustering with eigengene 1 are of biological process unknown, carbohydrate metabolic process(AMS1, ARA1, ATH1, CDC19, CIT2, ENO1, ENO2, FBA1, GCY1, GDB1, GID7, GLC3, GLG1, GLK1, GND1, GND2, GPH1, GPM1, GRE3, GSY1, GSY2, GUT1, GUT2, HSP104, HXK1, IGD1, IMA1, MDH3, NDE2, NTH2, OPI10, PDC1, PFK1, PFK26, PGI1, PGK1, PGM2, PIG2, PSK1, SHC1, SOL3, SOL4, TAL1, TDH1, TDH2, TDH3, TKL2, TPI1, TPS1, TPS2, TPS3, TSL1, UBC8, UGP1, VID28, VID30, YJR096W,

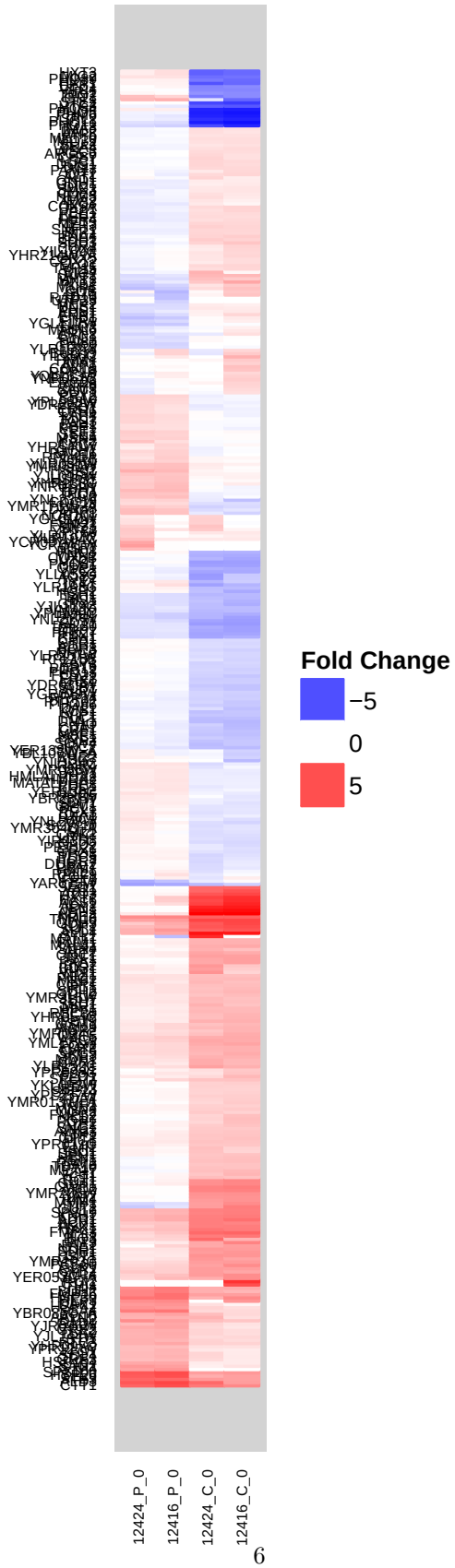


Figure 3: Expression of top 10% of genes at time zero.

YLR345W, YPR1), response to chemical stimulus (ACT1, ADR1, AHP1, BXI1, CCP1, CCW12, CDC48, CTA1, CTT1, ETP1, GAD1, GCY1, GND1, GRE3, GRX2, HBT1, HSP104, HSP12, HYR1, KIN82, LAP3, MCR1, MDG1, PDR10, PDR15, PRX1, ROD1, SNQ2, SOD1, SPI1, TRR2, TRX2, TRX3, TSA2, UGA2, USV1, YDL124W, YJR096W, YPR1), and generation of precursor metabolites and energy (AAC1, ACS1, CDC19, CIT1, ENO1, ENO2, ETR1, FBA1, GDB1, GLC3, GLG1, GLK1, GPH1, GPM1, GSY1, GSY2, HXK1, IGD1, ISF1, LSC1, NDE2, PDC1, PET10, PFK1, PFK26, PGI1, PGK1, PGM2, PIG2, PSK1, RGI1, RGI2, SHY1, TDH1, TDH2, TDH3, TPI1, UGP1, YLR345W).

For eigengene 2, the cluster which is repressed is enriched in cellular amino acid metabolic process (ADH2, ALD2, ALD3, CAR2, GAD1, GDH3, GLT1, GTT1, LAP3, PDC6, YAT1), nucleobase containing small molecule metabolic process (ADH2, ALD4, ATP6, GND2, IRA2, OLI1, RNR3, SOL3, TKL2), response to chemical stimulus (GAD1, HBT1, HSP12, HYR1, LAP3, SPI1, TRX3, TSA2), and carbohydrate metabolic process (CDC19, GND2, SHC1, SOL3, TDH2, TDH3, TKL2).

The cluster which is transiently activated is highly enriched in rRNA processing, various terms associated with ribosomal biogenesis.

I am also interested in seeing how my data correspond to the 2006 Ronen paper, so I will subset the data associated with each gene expression pattern she observed.

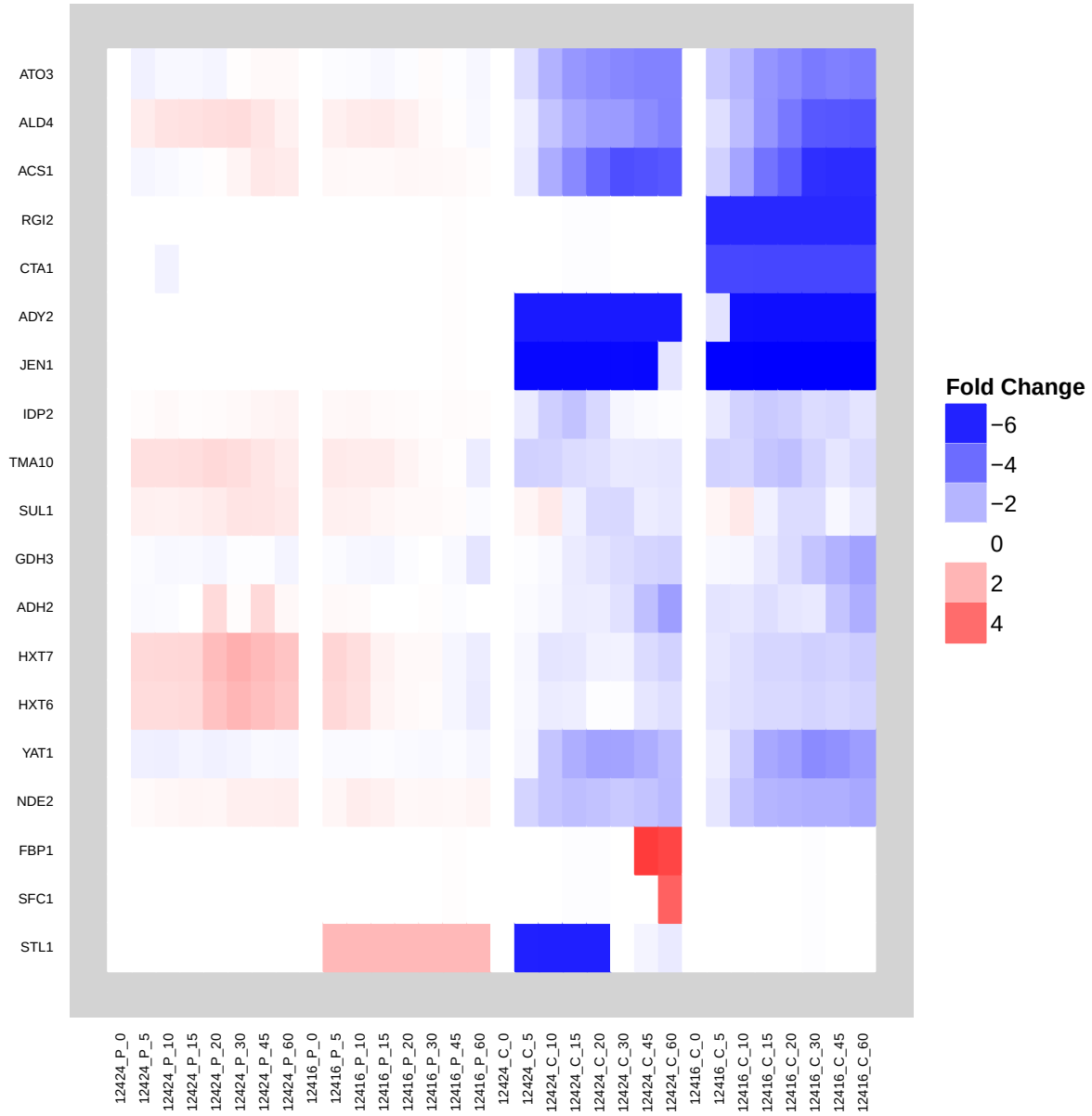


Figure 4: Expression time-course for genes which had basal high expression levels in the carbon-limited chemostats.

```
## Using eigengene as id variables
```

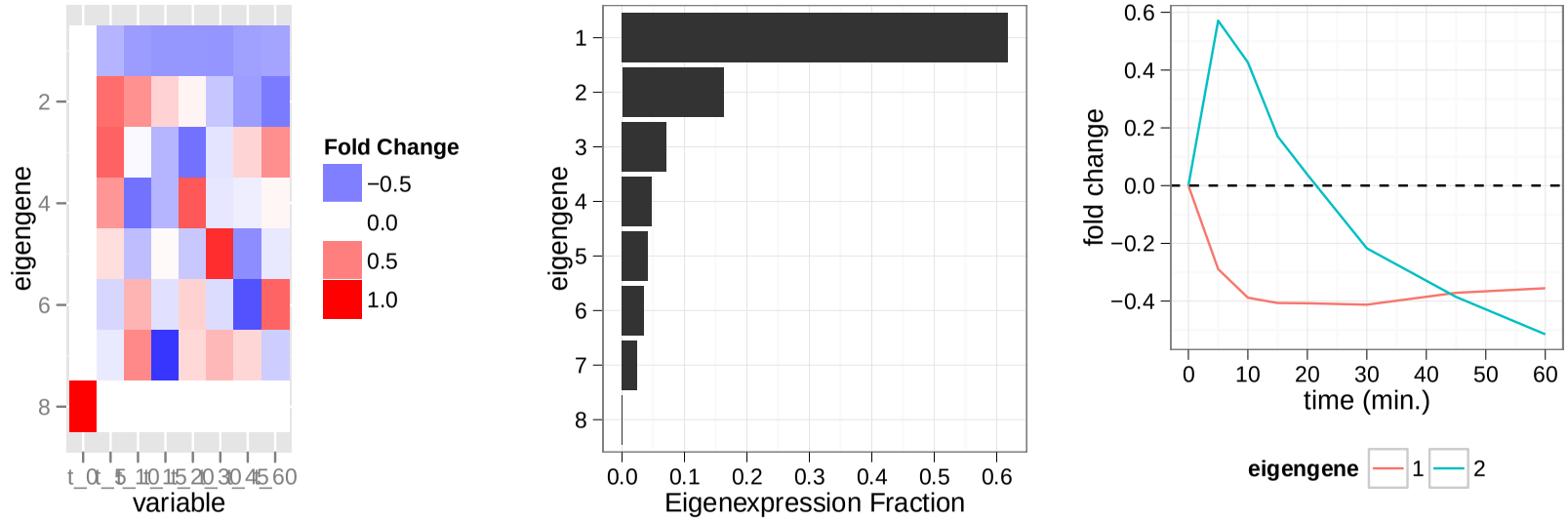


Figure 5: SVD on carbon-limited control arrays.

```
## Using eigengene as id variables
```

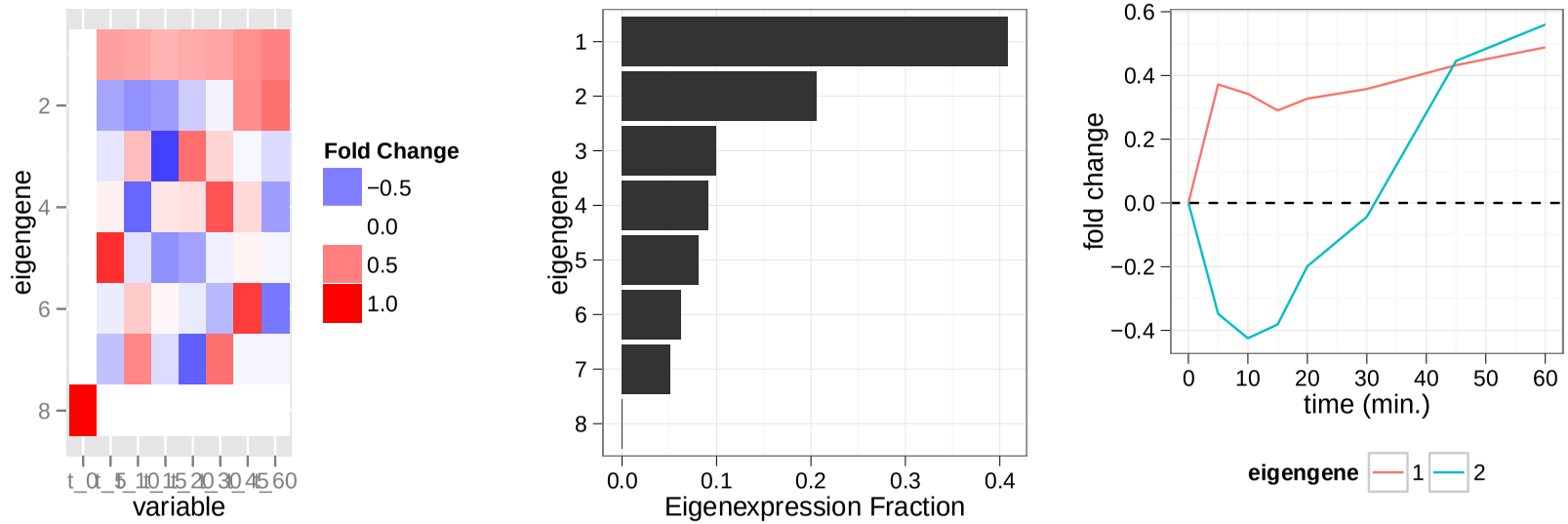


Figure 6: SVD on phosphate-limited control arrays.

```
## [1] 10
## [1] 10
```

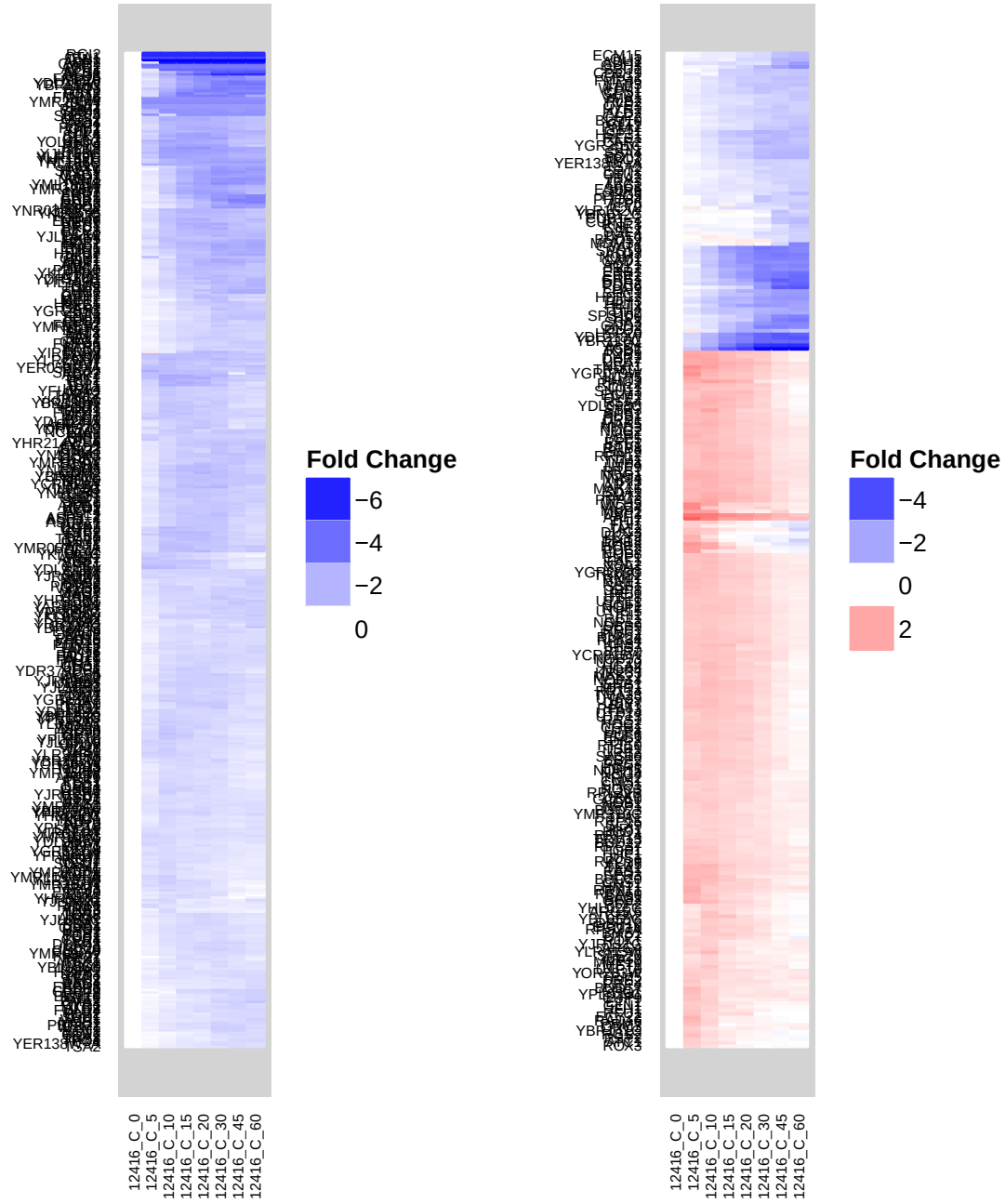


Figure 7: Expression patterns of genes correlated with eigenvector 1 (left) and eigenvector 2 (right) for carbon-limited 4ZEV strains.

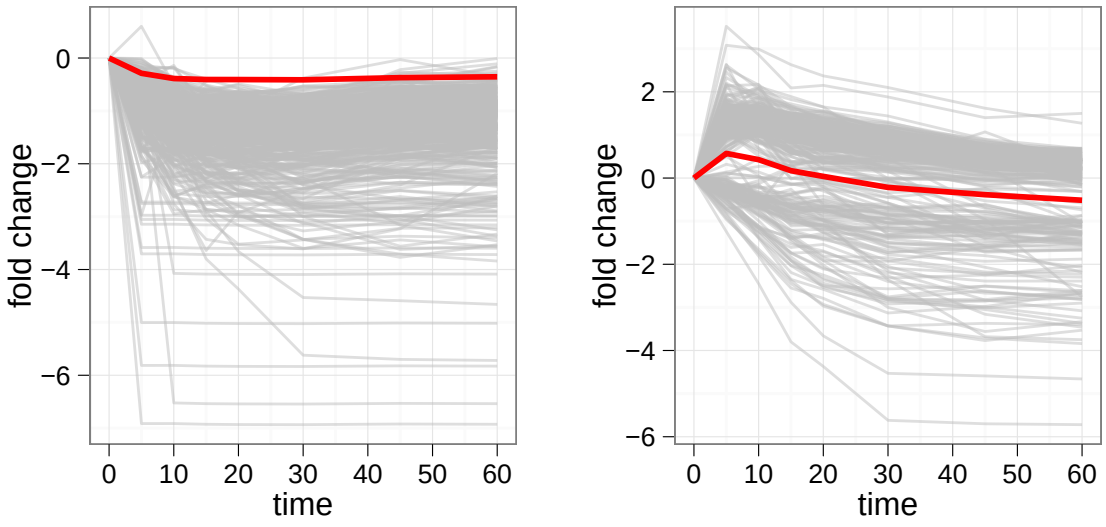
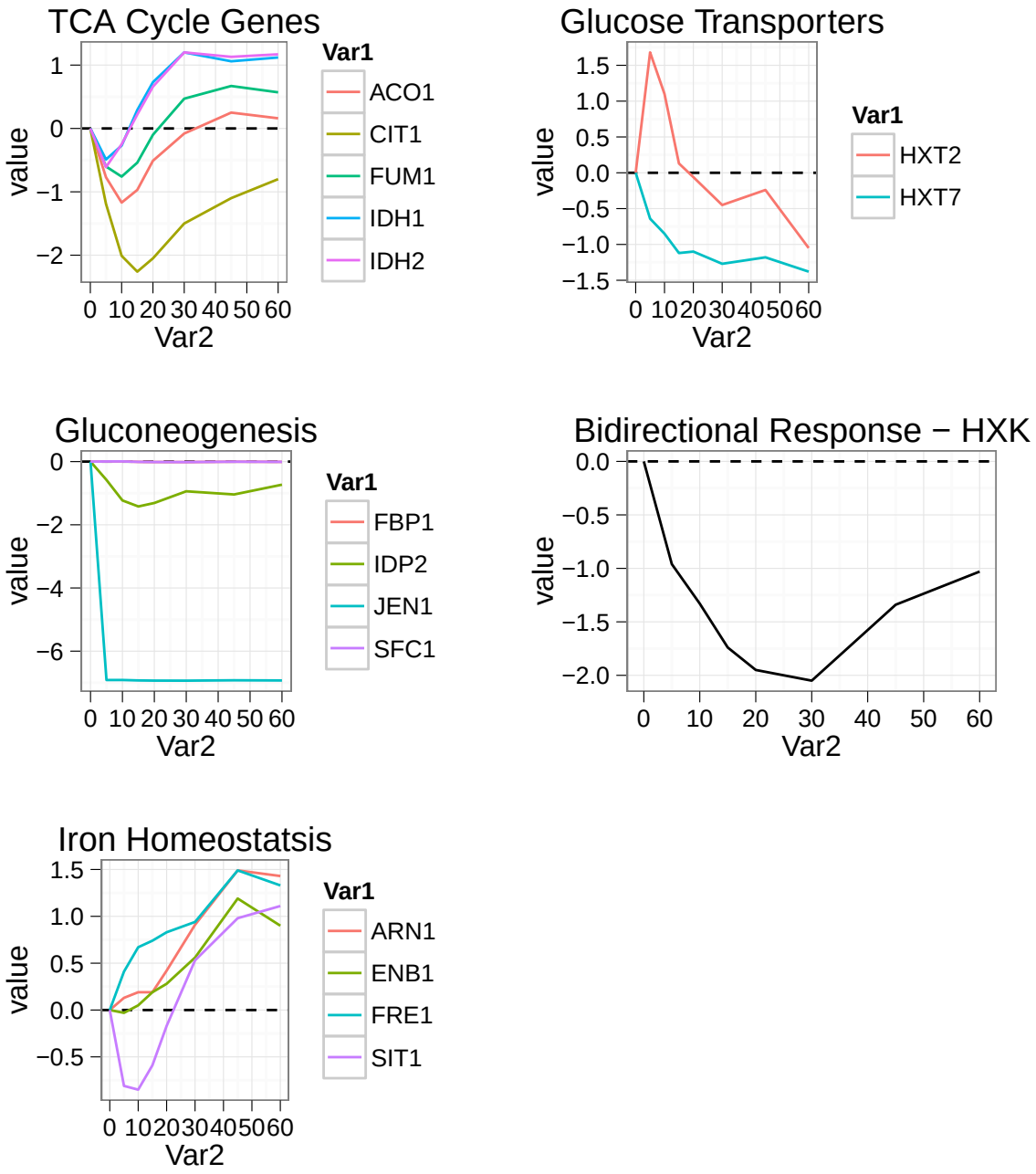


Figure 8: Expression patterns of genes correlated with eigenvector 1 (left) and eigenvector 2 (right) for carbon-limited 4ZEV strains.




```
## NULL
```

Figure 9: Expression of genes in the Carbon-limited 12416 chemostats from clusters identified in Ronen 2006.

```
## [1] 10
## [1] 10
## [1] 10
```

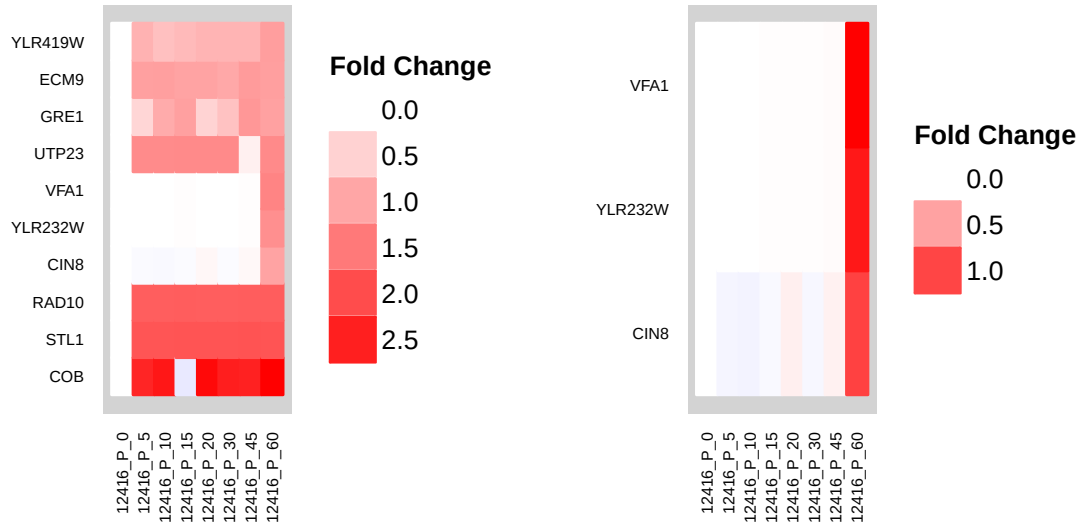


Figure 10: Expression patterns of genes correlated with eigenvector 1 (left) and eigenvector 2 (right) for phosphate-limited 4ZEV strains.

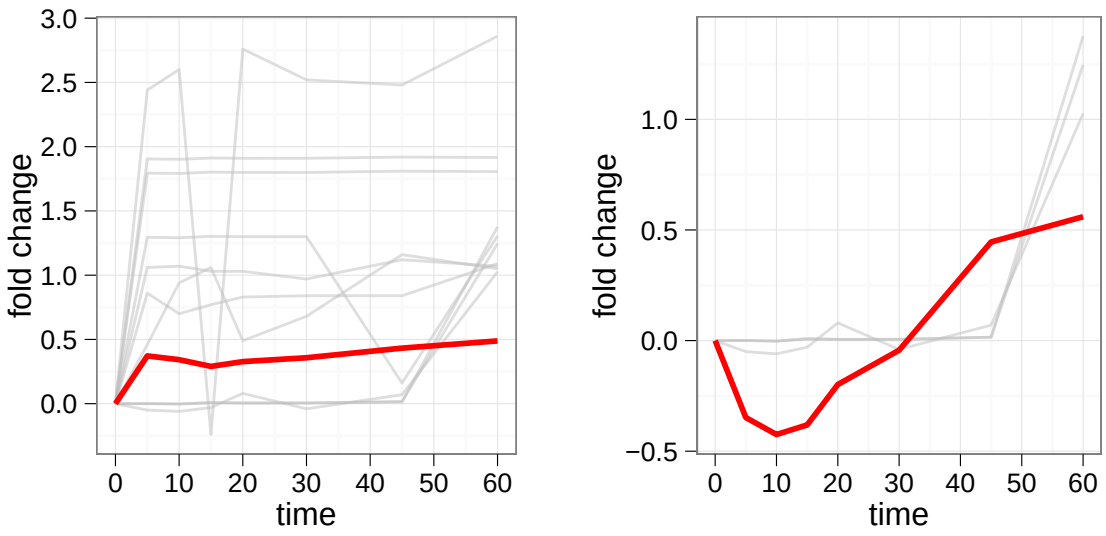


Figure 11: Expression patterns of genes correlated with eigenvector 1 (left) and eigenvector 2 (right) for phosphate-limited 4ZEV strains.

Analysis of 4ZEV$_{pr}$-SNF1 data with 4ZEV signal removed.

After subtracting the top-three eigengenes of the respective control data sets from the experimental data sets, there are still 194 genes which change by at least 2-fold in the carbon-limited data set, and 40 genes which change by at least 2-fold in both conditions. There are 16 genes which change by at least 2-fold in both conditions.

```
tab = xtable(data.frame(ORF = unlist(orf2name(intersect(rownames(p.424.new),
    rownames(c.424.new))))))
print(tab, tabular.environment = "longtable", floating = FALSE, include.rownames = FALSE)
```

ORF
HXK1
YCR041W
ADH2
HXT6
STL1
SRO77
FMP33
BUD19
ADR1
GSY1
YDL233W
RGI1
THI80
SNF1
HEF3
YLR162W

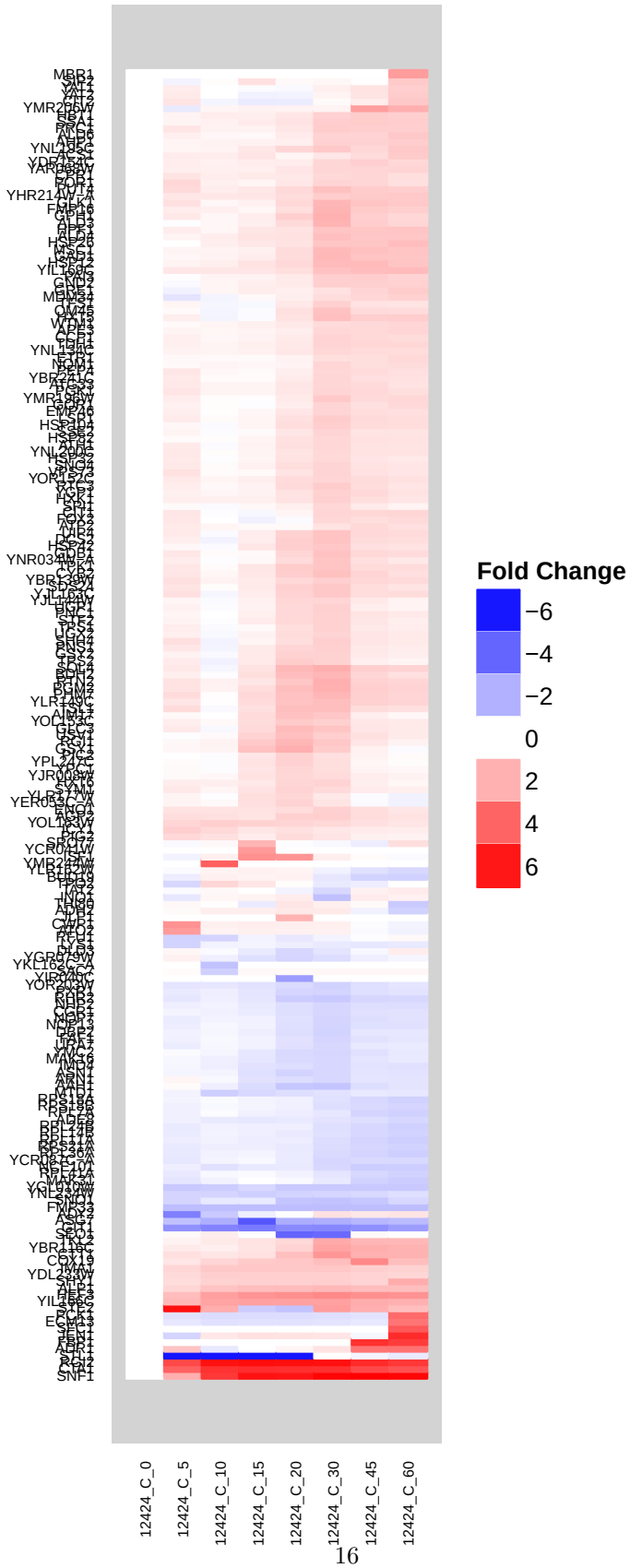


Figure 12: Carbon-limited 4ZEVpr-SNF1 with control top two eigengenes subtracted

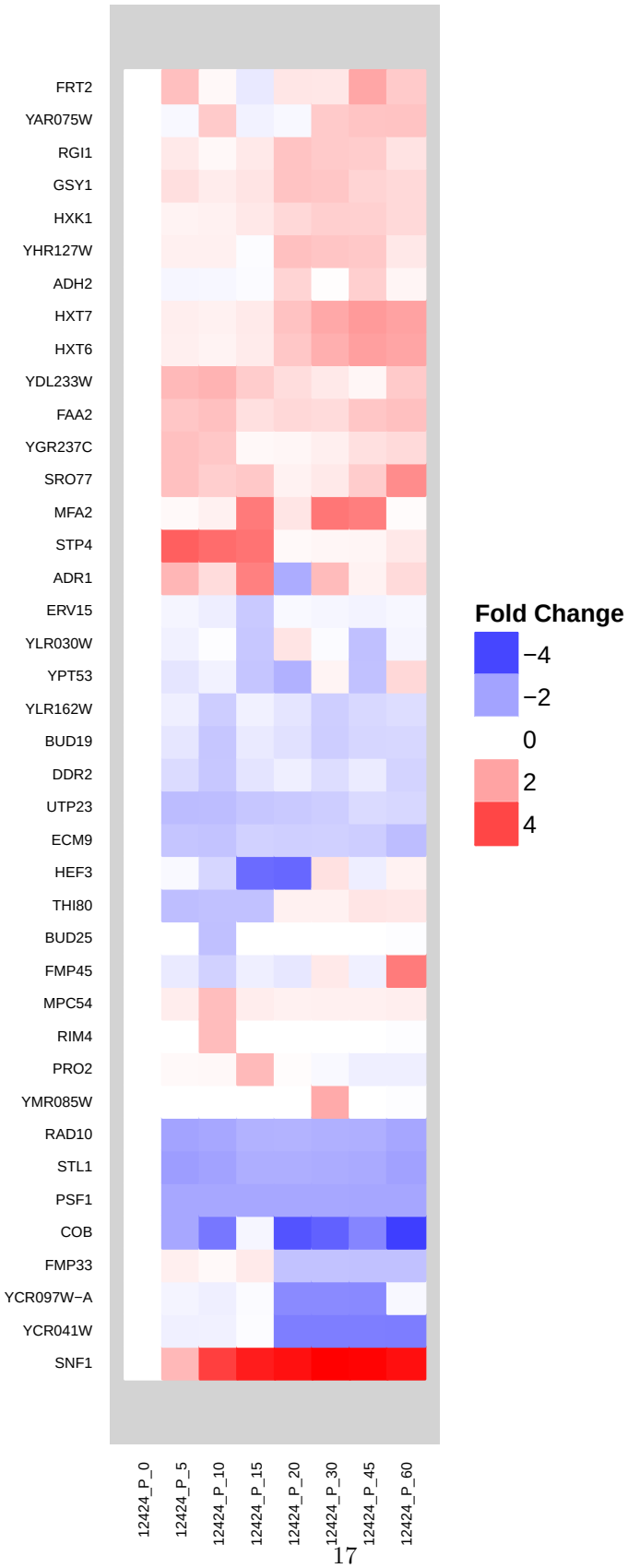


Figure 13: Phosphate-limited 4ZEVpr-Snf1 with control top two eigengenes subtracted