Available online at www.isas.org.in/jisas

JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 64(1) 2010 1-10

Differential Meta-analysis for Testing the Relative Importance of Two Competing Null Hypotheses over Multiple Experiments

Mehmet Kocak[1,2], Gaolin Zheng[3], Giri Narasimhan[4], E. Olusegun George[2] and Saumyadipta Pyne[5*]

[1]*Department of Biostatistics, St. Jude Children's Research Hospital, Memphis, TN 38103, USA*

[2]*Department of Mathematical Sciences, University of Memphis, Memphis, TN 38152, USA*

[3]*Department of Mathematics and Computer Science, North Carolina Central University, Durham, NC 27707, USA*

[4]*Bioinformatics Research Group (BioRG), School of Computing and Information Sciences, Florida International University, Miami, FL 33199, USA*

[5]*Department of Medical Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA 02115, USA*

SUMMARY

Gene expression experiments conducted under a variety of conditions can allow us to concurrently test more than one hypothesis for the same gene. For instance, if a particular gene has alternative modes of regulation, then it might be interesting to test the relative significance of each of those alternatives based on the gene's expressions under different conditions. In particular, if the significance values for two hypotheses about a gene appear to differ consistently in favor of a particular hypothesis in multiple independent experiments, then our new differential meta-analysis method can summarize the differences to test whether that hypothesis is overall more significant than the other hypothesis for the gene. Alternatively, one could first obtain differentially expressed gene sets with traditional meta-analysis of individual hypotheses, and then compare the sets, but such an approach addresses the problem post hoc and lacks sensitivity and statistical power. Our method, in contrast, addresses the problem directly and rigorously based on a novel statistic designed for testing the relative importance of two competing null hypotheses over multiple experiments. We also specify an analytical distribution for this combined statistic. We applied our method to genome-wide fission yeast cell cycle expression data and discovered interesting gene sets based on two interacting hypotheses.

Keywords : Meta-analysis, Logit, Gene Expression, Cell Cycle, Fission Yeast.

1. INTRODUCTION

Multiple laboratories conduct, often over a short time span, similar independent experiments to address common or related biological questions. The results from these studies could then be combined with meta-analytic procedures. Such procedures are useful for increasing power since experimental noise could lead to poor overlap among the results from individual experiments. In recent years, high-throughput platforms such as gene expression microarrays have introduced new computational issues in the context of meta-analysis. New methods were developed for assessment of platform-specific noise, calculating the false discovery rate of a combined hypothesis, incorporating collateral information from gene ontology, weighting the genes' *p*-values to increase the power of multiple testing, etc. (Kang *et al*. 2004; Parmigiani *et al.* 2004; Hu *et al.* 2005; Pyne *et al.* 2006). Such new challenges involving complex large-scale biological systems need to be addressed systematically by constructing new meta-analytic procedures. Genome-wide cell cycle

[*] *Corresponding author* : Saumyadipta Pyne
E-mail address : Saumyadipta_Pyne@dfci.harvard.edu

regulation is a classic example of such a problem that has been widely studied by different labs (e.g. Spellman *et al.* 1998; Whitfield *et al.* 2002) and which is complex enough to allow various parallel hypotheses about cell cycle regulation (e.g. Orlando *et al.* 2008).

In this paper, we present a procedure for solving a new issue in meta-analysis of microarray data that we call *differential meta-analysis*. Traditional meta-analysis tests a joint null hypothesis that is a conjunction of single null hypotheses, one from each individual experiment. For example, the null hypothesis of a gene's cell cycle regulation is commonly tested by combining the p-values (or z-scores) of its periodic expression in multiple cell cycle experiments (de Lichtenberg *et al.* 2005a; Oliva *et al.* 2005; Marguerat *et al.* 2006; Zheng *et al.* 2006). However complex experiments may present us with scenarios where a joint null hypothesis could also be a conjunction of more than one null hypothesis from each experiment. In particular, we are concerned with how the relative differences between a pair of measurements, each addressing a different hypothesis about a phenotype of the same biological entity (a gene, in this study), can be statistically combined across experiments to support or invalidate a conjecture on the relationship between the phenotypes. One possible approach could be to first combine the two hypotheses separately with traditional meta-analyses of the experiments, identify the two sets of differentially expressed genes, and then compute their set difference post hoc. However, this is unsatisfactory because it does not involve the experiment-wise relative differences, which are important to consider if the two hypotheses might interact with each other in an experiment.

Our method, in contrast, detects and accumulates the relative differences in the significance of alternative hypotheses about each gene tested in each experiment. For example, consider a complex disease that has both genetic and environmental factors. Our approach can help to address the question whether an environmental factor has a relatively larger statistically significant impact than a genetic factor on the disease over a patient population. Thus if there is a persistent discordance between the two hypotheses across multiple independent experiments, then our analytical approach allows us to summarize the differences between the hypotheses and test in a rigorous manner interesting questions such as: are these hypotheses competing with each other in the case of certain genes? Towards this, we extended the classical and most popular approach of meta-analysis, via combination of *p*-values (e.g. George and Mudholkar 1983), to design a novel differential statistic based on the logits of p-values, and then describe the analytical distribution of this statistic.

We used our technique to combine ten independent experiments measuring genome-wide time course expression over the cell division cycle of the fission yeast *Schizosaccharomyces pombe* (Rustici *et al.* 2004; Oliva *et al.* 2005; Peng *et al.* 2005). Two different null hypotheses were considered for each gene *g*: (a) gene *g* is not cell-cycle regulated (i.e., not periodically expressed), and (b) gene *g* is not significantly regulated (i.e., has little deviation from its mean expression over time). The proposed approach allows us to study the difference in the significance levels of (a) the periodicity of oscillation and (b) the expression regulation for every gene over the course of *S. pombe* cell cycle. This led to a comprehensive identification of two statistically significant gene sets showing markedly opposite patterns of expression: (1) highly periodic but weakly regulated genes, and (2) highly regulated but not periodic genes. In particular, based on the latter set, we identified a new regulatory network of genes many of which are known for their response to environmental stress. Interestingly, this network is distinct from the core stress pathway in *S. pombe* and is possibly induced by specific stress responses to the cell cycle arrest mechanisms employed by the different phase synchronization protocols used by the experiments. We also validated the genome-wide ranking based on our differential meta-analysis with the help of different global enrichment patterns for well known functionally characterized gene sets in *S. pombe*. Supplementary results are available from the authors upon request.

2. METHODS

It is a common practice among biomedical researchers to report the significance of their empirical observations in the form of *p*-values. Given multiple independent experiments where the significance of a test statistic in each is given by a corresponding *p*-value, there is a long history of research to test the conjunction null hypothesis by combining the *p*-values into a unified measure of significance (Hedges and Olkin 1985). For

combining cell cycle periodicity *p*-values, both the sum of z-scores and the product of *p*-values have been used (de Lichtenberg *et al.* 2005a; Oliva *et al.* 2005; Marguerat *et al.* 2006; Zheng *et al.* 2006). The following methodology was originally developed for the purpose of detecting genomic changes in cancer tissue grafting (K. Mehmet, E.O. George, and S. Pyne, manuscript in preparation). However it is generic enough to be considered for the present application as described below.

2.1 Difference of Two Logit Statistics: A Proposed Statistic

We used the differential meta-analysis statistic of Mehmet, George and Pyne described as follows. For every gene g, we compute the difference of two logit statistics based on two different null hypotheses about g. The logits are derived from the p-values corresponding to the independent tests of these two null hypotheses. Specifically, for gene g in experiment j, let $-l_{1jg} = \log(P_{jg}/(1 - P_{jg}))$ and $-l_{2jg} = \log(Q_{jg}/(1 - Q_{jg}))$ be the logits of the p-values P_{jg} and Q_{jg} for testing respective pairs of hypotheses H^0_{1jg} versus H^1_{1jg}, and H^0_{2jg} versus H^1_{2jg}. Then their difference is given by $l_{1jg} - l_{2jg}$ which is equal to $-\log(((P_{jg}/(1 - P_{jg}))/(Q_{jg}/(1 - Q_{jg}))))$, the log of odds ratio for evaluating the amount by which the significance of one factor acting on g is changed relative to another factor in experiment j. Let $-L_{ig} = \sum_{j=1}^{N_i} l_{ijg}$ for $i = 1, 2$ denote the sum of the logits over all experiments performed. Then under $\bigcap_j H^0_{ijg}$, L_{ig} (and $-L_{ig}$) is a convolution of logistic random variables. George and Mudholkar (1983) computed the exact distribution of L_{ig}, and showed that this distribution can be accurately approximated using a t-distribution. Specifically, it is distributed as $L_{ig} \sim \pi\sqrt{N_i(5N_i+2)/3(5N_i+4)}(t_{5N_i+4})$. Consequently, under null hypotheses $\bigcap_j H^0_{1jg}$ and $\bigcap_j H^0_{2jg}$, $L_{1g} - L_{2g}$ is a convolution of $N_1 + N_2$ logistic random variables. Thus if $N_1 = N_2 = N$, then under $\bigcap_j H^0_{ijg}$, $i = 1, 2$, the combined statistic $L_{1g} - L_{2g}$ for gene g is accurately approximated by $L_{1g} - L_{2g} \sim \pi\sqrt{2N(5N+1)/3(5N+2)}(t_{10N+4})$.

We are interested in comparing the periodicity of oscillation of every gene in *S. pombe* against its expression regulation over the course of cell cycle in all $N = 10$ experiments. These hypotheses were previously tested for benchmarking purposes (de Lichtenberg *et al.* 2005a) for *S. cerevisiae* but were never compared against each other. Based on our differential meta-analysis for testing these two hypotheses against each other, we produced a final genome-wide ranking of all genes in *S. pombe*. The upper tail of the distribution of the combined statistic $L_{1g} - L_{2g}$ provides a measure of significance for the composite hypothesis that a given gene g is "relatively more periodic than regulated", while the lower tail provides the same for the converse hypothesis, i.e., "relatively more regulated than periodic."

2.2 Experimental Data

We used ten time course microarray experiments on *S. pombe* cell cycle based on two synchronization protocols – elutriation (Elu) and Cdc25 block-release (Cdc25) – see the published studies (Rustici *et al.* 2004; Oliva *et al.* 2005; Peng *et al.* 2005) for more details. The data were normalized by the original experimenters. The peak phase angles of all genes in all experiments are due to Marguerat *et al.* 2006.

2.3 *p*-values for Periodicity and Expression Regulation

Following previous work, we evaluated the p-value of periodicity P_{jg} for each gene g's time course in each experiment j based on the Fourier sum and a permutation test with 10^4 random permutations (see de Lichtenberg *et al.* 2005a for details). For computing the p-value of expression regulation, we used a test for the significance of peak expression in a centered time course in which the original (uncentered) expression profiles are not required. For a given experiment j, the population variance σ_j^2 over all the time courses was determined. Then, for a given gene g, 10^4 bootstrap samples were generated using only the expression values for gene g, and the sample variance of the bootstrap samples were compared with σ_j^2 to obtain the p-value of expression regulation Q_{jg} for g.

2.4 Bioinformatic Analysis

The Dynamic Bayesian Network tool Banjo (Yu *et al.* 2004) was used to construct the gene regulatory

networks involving the P^-R^+ gene set (see Discussion). Four data sets based on Cdc25 synchronization protocol – one each from Oliva *et al.* and Peng *et al.* and two from Rustici *et al.* – were used to generate a Cdc25 regulation network. Separately, 4 data sets based on elutriation protocol – one from Peng *et al.* and three from Rustici *et al.* – were used to generate a Elu network. Only the nodes with positive degree and the edges which are present in at least two experiments were retained. The final network (Fig. 3) was constructed by union of the Cdc25 and the Elu networks. To indicate protocol specific regulation, when an edge was represented in more Cdc25 experiments than Elu experiments, it is depicted as a thin edge, and in bold otherwise. The dominant GO biological process was determined with hypergeometric test using Genecodis software (Carmona-Saez *et al.* 2007). For Gene Set Enrichment Analysis, we used the pre-ranked option (which allowed us to use our global ranking of genes based on differential meta-analysis) of the GSEA tool (Subramanian *et al.* 2007).

3. RESULTS AND DISCUSSION

We performed differential meta-analysis on data from 4940 genes in *S. pombe* (see Supplementary information). We computed the differential meta-analysis statistic (L_{1g}–L_{2g}; see Methods) for each gene g using its p-values for periodicity (P_{jg}) and expression regulation (Q_{jg}) from each experiment j, and then computed its combined p-value based on the t distribution described in the Methods. From the right and the left tails of the distribution (see Fig. 1), we determined, using a Benjamini-Hochberg based FDR cutoff of 0.01, the sets of genes that are relatively more significant for Periodicity (P) and expression Regulation (R) respectively. That is, we obtain a set (denoted by P^+R^-) containing relatively highly periodic but weakly regulated genes, and a second set (denoted by P^-R^+) containing genes that are highly regulated but which do not follow a cyclic pattern.

We report 48 genes in P^-R^+ (see Table 1) that are significantly more regulated than periodic. On the other hand, we identified 864 genes in P^+R^- to be relatively more periodic than regulated. Time course expression profiles of representative genes from either set are shown in Fig. 2. The genome-wide ranked list is given in the supplementary information.

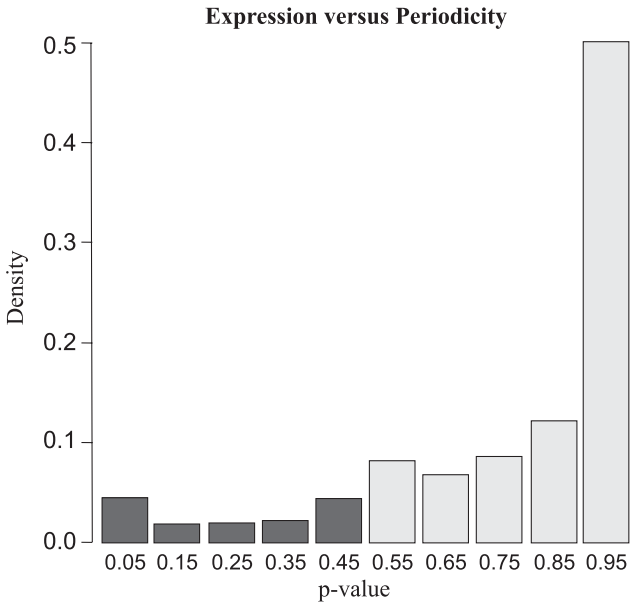

Fig. 1. Tails of the differential meta-analysis *p*-value distribution represent two competing hypotheses. The light grey bins in the right half of the plot represent the part of the *p*-value distribution for which the genes had relatively more significant periodicity than expression regulation, whereas the dark grey bins in the left half represent the converse. The left and the right tails, after applying FDR thresholds of 0.01, represent the significant populations of the two competing hypotheses. The bump in the middle of the distribution marks a collection of genes with similar levels of periodicity and expression regulation.

The expressions of weakly cell cycle regulated genes have been observed experimentally in the past in *S. cerevisiae* (de Lichtenberg *et al.* 2005b). However, we are unaware of any systematic identification of such expression profiles for any species. Known modules of weakly cell cycle regulated genes in *S. pombe*, such as the early-mid G2 phase ribosome biogenesis cluster (Oliva *et al.* 2005), are represented in P^+R^- (see below). Indeed based on the periodicity of expression, it was observed by (Oliva *et al.* 2005) that as many as 2000 genes might be cell cycle regulated in *S. pombe*, due to reasons that are adaptive or otherwise; and more than two-thirds of these are weakly regulated, which is supported by the findings of our method (864 genes in P^+R^-).

The other significant set P^-R^+ of 48 genes (see Table 1) with highly regulated non-periodic profiles is perhaps more interesting than the P^+R^- set. Given the different environmental stress factors associated with

Table 1. Significant genes by differential meta-analysis. The 48 most significant genes in *S. pombe* with relatively higher expression than periodicity during cell cycle are listed with the *S. cerevisiae* orthologs. The core stress responders in both species are marked in italic boldface whereas the specific stress responders in fission yeast appear in non-italic boldface. As shown, 20 of the listed genes are annotated as stress activated in *S. pombe*.

Significant gene	Ortholog	Significant gene	Ortholog
STR3	ENB1	***SPCC330.06C***	***AHP1***
SPCC132.04C	GDH2	SPBC354.08C	
SPBP8B7.05C	***NCE103***	**WIS2**	***CPR7***
SPCC1450.16C	TGL3	***SPAC27D7.11C***	
SPCC18B5.02C		***HSP16***	***HSP42***
ISP4	OPT2	***SPAC9E9.09C***	ALD6
GST2	URE2	FIP1	FTR1
HSP90	***HSP82***	SSA2	***SSA2***
SPAC26H5.09C		**SPBC19C7.04C**	IBI2
VHT1	VHT1	**STI1**	STI1
SPBP4G3.03		**SPBPB10D8.02C**	
SPBC36.03C	TPO1	***SPBC21C3.19***	RTC3
SSA1	***SSA2***	**ABC3**	YBT1
PGK1	PGK1	***SPAC27D7.10C***	
SPAC1786.04		SPBC23G7.13C	DUR3
SPAC5H10.01		ADG2	
FRP1	FRE5	***SPBPB2B2.12C***	GAL10
HSP9	***HSP12***	SPAC30D11.01C	
SPAC1786.01C	TGL5	SPAC30D11.11	IZH3
SPAC977.02		MFM2	
SPAC3G6.05	SYM1	***SPCC70.08C***	
SPBPB2B2.13	GAL3	LCF2	FAA1
SPCC1281.06C	OLE1	MEI2	
HTA2	***HTA1***	SPAC4A8.10	ROG1

the cell growth arrest mechanisms of different phase synchronization protocols (Futcher 1999), it is possible that the regulation is primarily due to the corresponding stress response of many of these genes. Indeed the dominant GO category for $P^{-}R^{+}$ is "cellular response to stress" (GO:0033554) with enrichment p-value = 9.07e-06. For example, some of the genes encode known heat shock proteins such as Hsp9, Hsp16 and Hsp90. Therefore to understand the impact of different synchronization protocols (i.e., of Cdc25 and Elu), we used multi-experiment Dynamic Bayesian Network analysis to identify the corresponding regulatory networks based on genes in the set $P^{-}R^{+}$ under the two different protocols (see Methods). Intriguingly, only few of the regulatory connections observed for multiple experiments following one protocol are also significant in the other, suggesting that the stress responses are likely to be protocol-specific. The combined multi-experiment network is shown in Fig. 3 in which the regulatory links dominant in the two different protocols are shown in distinct colours.

The genes in our regulatory network (Fig. 3) have two interesting properties. Many respond to multiple stress conditions in *S. pombe* (i.e. CESR genes in Chen *et al.* 2003; marked with double-rimmed boxes in Fig. 3), while some respond to specific stresses (SESR in Chen *et al.* 2003; single-rimmed boxes in Fig. 3). Interestingly, however, the network also includes several nodes that are neither CESR nor SESR genes (these are plotted as ellipses in Fig. 3). These genes, in particular such hub nodes as *fip1* or *ssa2*, may be considered as new candidates for stress regulation induced by cell cycle experimental protocols in *S. pombe*. Notably, for the most prominent new hub Ssa2 (encircled in Fig. 3), all the *S. cerevisiae* homologs are 70kDa heat shock proteins (Hsp70) (Penkett *et al.* 2006).

Another interesting aspect of the network is that it is distinct from the key stress-activated MAP kinase Sty1 (a.k.a. spc1) pathway in *S. pombe* (analogous to human p38 and *S. cerevisiae* Hog1; Gasch 2007). Such distinct pathways are known to occur, e.g., the Mec1 pathway in *S. cerevisiae* is activated in cell cycle arrest due to DNA damage but not due to heat shock treatment (Gasch *et al.* 2001). Studies of relationships among yeast pathways (Shiozaki and Russell 1995; Petersen and Hagan 2005; Brauer *et al.* 2008) suggest the possibility of subnetworks that overlap both the stress response and the cell cycle processes, which may be regulated according to the suitability of the metabolic environment within the cell to proceed, in this case, with mitosis (Futcher 2006). Thus it is possible that the genes which respond to the metabolic conditions of

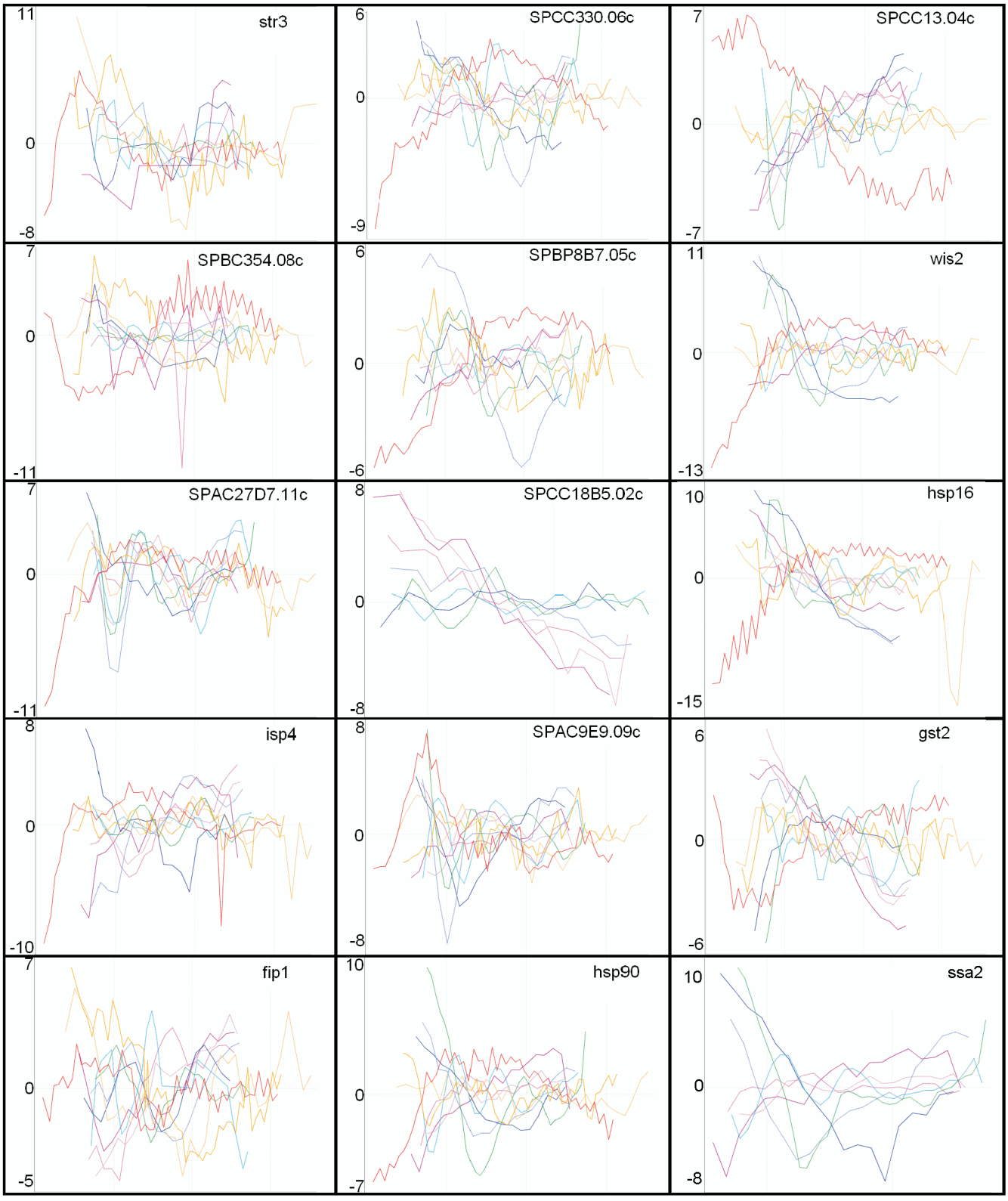
str3
SPCC330.06c
SPCC13.04c
SPBC354.08c
SPBP8B7.05c
wis2
SPAC27D7.11c
SPCC18B5.02c
hsp16
isp4
SPAC9E9.09c
gst2
fip1
hsp90
ssa2

Fig. 2(a)

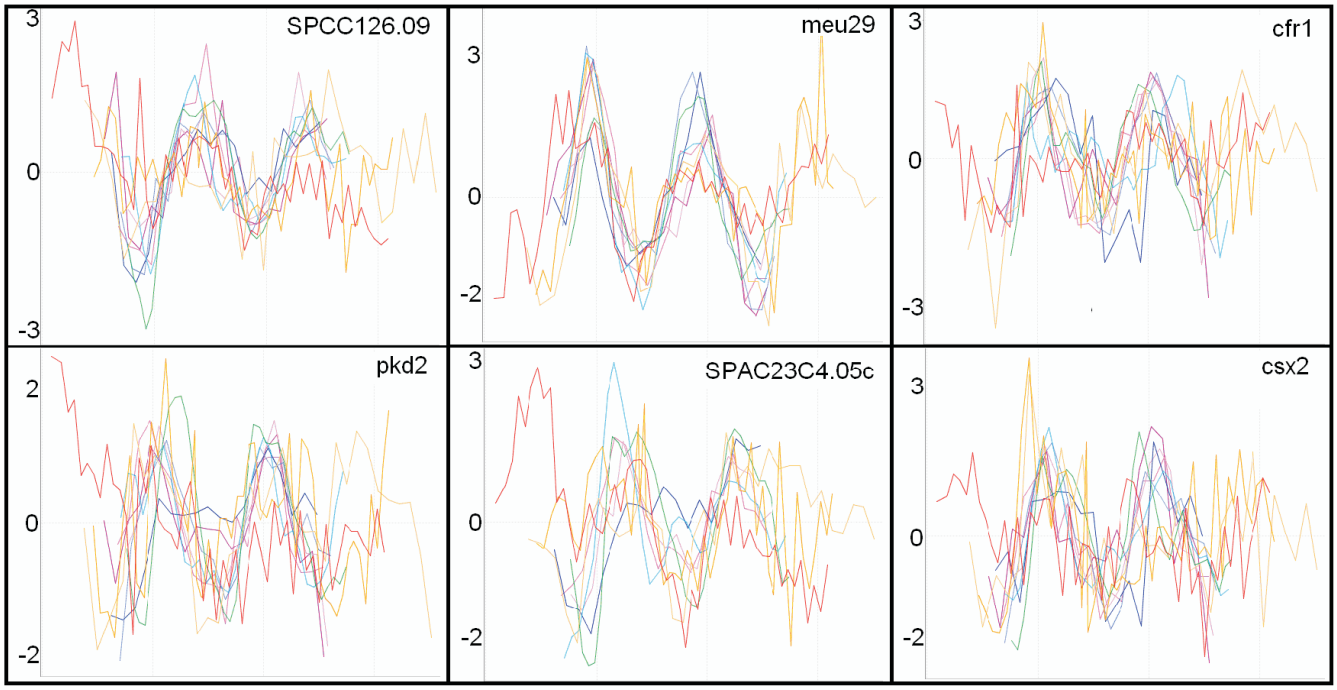

Fig. 2(b)

Fig. 2. Expression profiles in 10 experiments for some of the most significant genes in differential meta-analysis. The *y*-axis marks the peak expression for each gene, the range of which is markedly higher in plot (a) than in (b). The *x*-axis marks time points. Plot (a) shows the top 15 genes from the set $P^{-}R^{+}$ with high expression and low periodicity, while plot (b) shows 6 weakly regulated cyclic genes from the set $P^{+}R^{-}$. Both gene sets are based on differential meta-analysis. The time course profiles are generated with the help of the online resource Cyclebase.org (Gauthier *et al*. 2008).

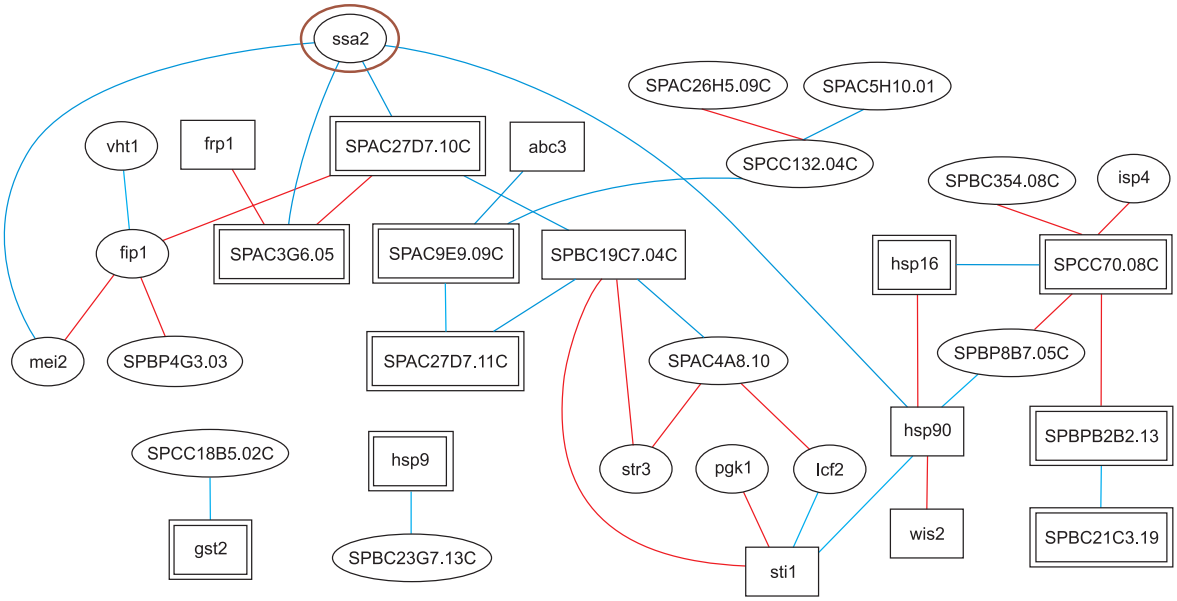

Fig. 3. The regulatory network based on the significant genes from differential meta-analysis. The rectangular nodes represent known stress response genes; the double-rimmed nodes are induced genes in the CESR list while single-rimmed ones belong to the SESR list. The ellipses represent genes that are in neither list. A blue edge represents regulation that is significant in more Elutriation experiments than Cdc25, and a red edge represents the converse situation. A prominent hub node, *ssa2*, which is listed neither as CESR nor as SESR gene, is encircled.

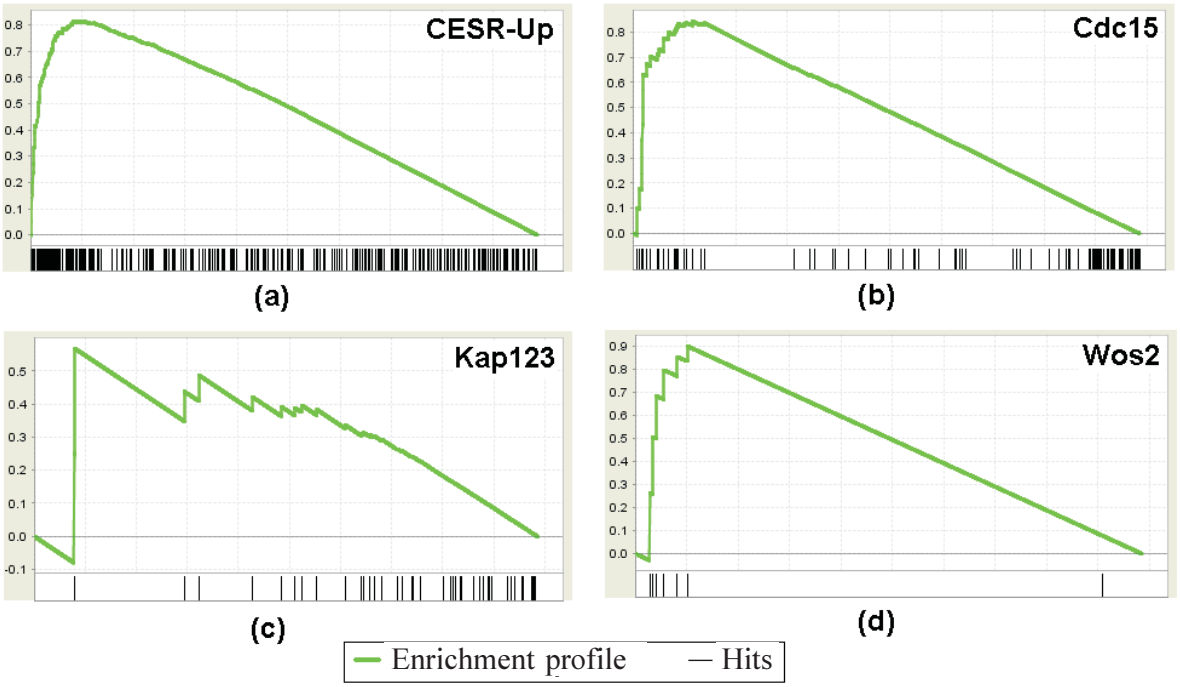

Fig. 4. Gene Set Enrichment Analysis for known gene sets in *S. pombe*. The enrichment scores (y-axis) are computed for four known gene clusters (named within each plot) with respect to the global list of 4940 genes (x-axis) ranked by their differential meta-analysis p-values. The list is depicted from the lowest p-value (i.e. for gene with high regulation and low periodicity) at the left end to the highest p-value (the converse scenario) at the right end, and a "hit" for a gene from the chosen cluster is recorded with a thin vertical mark at the corresponding position of the gene in our rank list. See text for discussion.

different cell cycle experimental protocols might assume protocol-specific regulatory roles and form distinct pathways.

Finally, to validate our genome-wide ranked list based on differential meta-analysis, we examined global enrichment patterns (Fig. 4) for different functionally characterized gene sets in *S. pombe*. First, we used the core genes induced by stress in *S. pombe* which are denoted by 'CESR-Up' in Fig. 4(a) (Chen *et al.* 2003). While most genes in this set clearly had high ranks in our global list (as shown by vertical ticks at the bottom of the figure), the ranks of many were also evenly distributed, which supports our belief that the collection of significant genes in the present study is only partially overlapping with the core stress pathway in *S. pombe*. Second, a large M phase cluster 'Cdc15', consisting of periodic genes of which some are strongly and some weakly cell cycle regulated (Oliva *et al.* 2005), was found to have corresponding ranks in Fig. 4(b). Third, the weakly expressed (in early-mid G2 phase) but periodic cluster 'Kap123' of ribosomal biogenesis genes (Oliva *et al.* 2005) is also found to have matching low ranks in Fig. 4(c). Finally, we tested with the late S/early G2 phase cluster 'Wos2' (Oliva *et al.* 2005) which could represent either hypothesis. While these genes are periodic, their promoters contain typical stress response motifs (Oliva *et al.* 2005). Interestingly, the gene set weighed in favor of the stress hypothesis with high ranks in our global rank list (Fig. 4(d)). In fact, the representative gene for this cluster *wos2* encodes a chaperone activator that interacts with the heat shock protein Hsp90, which is one of the most significant genes by our meta-analysis (p-value = 1.77×10^{-8}, Table 1).

We understand that our present method has its limitations. For instance, it is not obvious how to extend the combined statistic to scenarios that simultaneously involve more than two competing hypotheses. Moreover, the tests for each of the two individual hypotheses may not be equally powerful. Finally, since our combined null hypothesis is based on conjunction over experiments, it could be rejected even if the null is rejected in at least one experiment. However a possible way to address the issue is with a partial

conjunction null hypothesis (Heller *et al.* 2007). Similarly, the effect of any positive dependency structure among the expression patterns of related genes may be addressed with a suitable method (e.g. Benjamini and Yekutieli, 2001). Nevertheless, differential meta-analysis could be a powerful tool for biological studies involving hypotheses that may apply competitively to the same target entity. For instance, comparison of gene expressions due to a transcription factor's deletion versus its binding to promoter DNA could help us identify the genes that are not regulated directly by that transcription factor (Tang *et al.* 2006). Similarly, the competitive scenario of transcription factor binding versus nucleosome occupancy at the same regulatory region of a gene could be tested with differential meta-analysis (Narlikar *et al.* 2007). Our method also has many potential applications in biomedical data analysis. For example, the efficacy of a new anti-malarial drug may depend on multiple factors such as the parasite strain and the drug dosage (Sidhu *et al.* 2002). Differential meta-analysis of DNA data (such as SNPs) versus mRNA data (gene expression) from the same cohort of subjects could reveal intermediate mechanisms that play key roles between a genetic signature and its actual expression. Similarly, differential meta-analysis at "omic" levels, say, of transcriptome and proteome data could reveal modules involved in post-transcriptional or other intermediate modes of regulation. As experiments of increasing size and complexity accumulate in data repositories, new methods of meta-analysis such as the present one will be very useful.

ACKNOWLEDGEMENTS

This work was supported by the National Institutes of Health [P01 DA15027-01 to GN and EOG]. The authors thank Prof. Kalai Mathee for useful discussions.

REFERENCES

Brauer, M.J., Huttenhower, C., Airoldi, E.M., Rosenstein, R., Matese, J.C., Gresham, D., Boer, V.M., Troyanskaya, O.G. and Botstein, D. (2008). Coordination of growth rate, cell cycle, stress response, and metabolic activity in yeast. *Mol. Biol. Cell*, **19(1)**, 352-367.

Carmona-Saez, P., Chagoyen, M., Tirado, F., Carazo, J.M. and Pascual-Montano, A. (2007). GENECODIS: A web-based tool for finding significant concurrent annotations in gene lists. *Genome Biol.*, **8(1)**, R3.

Chen, D., Toone, W.M., Mata, J., Lyne, R., Burns, G., Kivinen, K., Brazma, A., Jones, N. and Bähler, J. (2003). Global transcriptional responses of fission yeast to environmental stress. *Mol. Biol. Cell,* **14(1)**, 214-229.

de Lichtenberg, U., Jensen, L.J., Fausboll, A., Jensen, T.S., Bork, P. and Brunak, S. (2005a). Comparison of computational methods for the identification of cell cycle-regulated genes. *Bioinformatics*, **21(7)**, 1164-1171.

de Lichtenberg, U., Wernersson, R., Jensen, T.S., Nielgen, H.B., Fausboll, A., Schmidt, P., Hansen, F.B., Knudsen, S. and Brunak, S. (2005b). New weakly expressed cell cycle regulated genes in yeast. *Yeast*, **22(15)**, 1191-1201.

Futcher, B. (1999). Cell cycle synchronization. *Methods Cell Sci.*, **21(2-3)**, 79-86.

Futcher, B. (2006). Metabolic cycle, cell cycle, and the finishing kick to start. *Genome Biol.*, **7(4)**, 107.

Gasch, A.P., Huang, M., Metzner, S., Botstein, D., Elledge, S.J. and Brown, P.O. (2001). Genomic expression responses to DNA-damaging agents and the regulatory role of the yeast ATR homolog Mec1p. *Mol. Biol. Cell,* **12(10)**, 2987-3003.

Gasch, A.P. (2007). Comparative genomics of the environmental stress response in ascomycete fungi. *Yeast,* **24(11)**, 961-976.

Gauthier, N.P., Larsen, M.E., Wernersson, R., de Lichtenberg, U., Jensen, L.J., Brunak, S. and Jensen, T.S. (2008). Cyclebase.org – A comprehensive multi-organism online database of cell-cycle experiments. *Nucleic Acids Res.,* **36**, D854-859.

George, E.O. and Mudholkar, G.S. (1983). On the convolution of logistic random variables. *Metrika,* **30***,* 1-14.

Hedges, L.V. and Olkin, I. (1985). *Statistical Methods for Meta-Analysis.* Academic Press, New York.

Hu, P., Greenwood, C.M. and Beyene, J. (2005). Integrative analysis of multiple gene expression profiles with quality-adjusted effect size models. *BMC Bioinformatics*, **6**, 128.

Kang, S., Wu, Y.F., An, N. and Ren, M. (2004). A systematic review and meta-analysis of the efficacy and safety of a fixed, low-dose perindopril-indapamide combination as first-line treatment of hypertension. *Clinical Therapeutics,* **26(2)**, 257-270.

Marguerat, S., Jensen, T.S., de Lichtenberg, U., Wilhelm, B.T., Jensen, L.J. and Bähler, J. (2006). The more the merrier: Comparative analysis of microarray studies on cell cycle-regulated genes in fission yeast. *Yeast*, **23(4)**, 261-277.

Narlikar, L., Gordân, R. and Hartemink, A.J. (2007). A nucleosome-guided map of transcription factor binding sites in yeast. *PLoS Comp. Biol.*, **3(11)**, 2199-2208.

Oliva, A., Rosebrock, A., Ferrezuelo, F., Pyne, S., Chen, H., Skiena, S., Futcher, B. and Leatherwood, J. (2005). The cell cycle regulated genes of *Schizosaccharomyces pombe*. *PLoS Biol.*, **3(7)**, e225.

Orlando, D.A., Lin, C.Y., Bernard, A., Wang, J.Y., Socolar, J.E., Iversen, E.S., Hartemink, A.J. and Haase, S.B. (2008). Global control of cell-cycle transcription by coupled CDK and network oscillators. *Nature,* **453(7197)**, 944-947.

Parmigiani, G., Garrett-Mayer, E.S., Anbazhagan, R. and Gabrielson, E. (2004). A cross-study comparison of gene expression studies for the molecular classification of lung cancer. *Clinical Cancer Res.*, **10**, 2922-2927.

Peng, X., Karuturi, R.K., Miller, L.D., Lin, K., Jia, Y., Kondu, P., Wang, L., Wong, L.S., Liu, E.T., Balasubramanian, M.K. and Liu, J. (2005). Identification of cell cycle-regulated genes in fission yeast. *Mol. Biol. Cell*, **16(3)**, 1026-1042.

Penkett, C.J., Morris, J.A., Wood, V. and Bähler, J. (2006). YOGY: A web-based, integrated database to retrieve protein orthologs and associated Gene Ontology terms. *Nucleic Acids Res.*, **34**, W330-334.

Petersen, J. and Hagan, I.M. (2005). Polo kinase links the stress pathway to cell cycle control and tip growth in fission yeast. *Nature,* **435**, 507-512.

Pyne, S., Futcher, B. and Skiena, S. (2006). Meta-analysis based on control of false discovery rate: Combining yeast ChIP-chip datasets. *Bioinformatics,* **22(20)**, 2516-2522.

Pyne, S., Skiena, S. and Futcher, B. *et al*. (2008). Meta-analysis of microarray data. In: *Bioinformatics Algorithms: Techniques and Applications,* Mandoiu I and Zelikovsky A. (eds.), Wiley-Interscience.

Rustici, G., Mata, J., Kivinen, K., Lió, P., Penkett, C.J., Burns, G., Hayles, J., Brazma, A., Nurse, P. and Bähler, J. (2004). Periodic gene expression program of the fission yeast cell cycle. *Nature Genetics*, **36(8)**, 809-817.

Shiozaki, K. and Russell, P. (1995). Cell-cycle control linked to extracellular environment by MAP kinase pathway in fission yeast. *Nature,* **378**, 739-743.

Sidhu, A.B., Verdier-Pinard, D. and Fidock, D.A. (2002). Chloroquine resistance in *Plasmodium falciparum* malaria parasites conferred by pfcrt mutations. *Science*, **298(5591)**, 210-213.

Spellman, P.T., Sherlock, G., Zhang, M.Q., Iyer, V.R., Anders, K., Eisen, M.B., Brown, P.O., Botstein, D. and Futcher, B. (1998). Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell,* **9(12)**, 3273-3297.

Subramanian, A., Kuehn, H., Gould, J., Tamayo, P. and Mesirov, J.P. (2007). GSEA-P: A desktop application for Gene Set Enrichment Analysis. *Bioinformatics,* **23(23)**, 3251-3253.

Tang, L., Liu, X. and Clarke, N.D. (2006). Inferring direct regulatory targets from expression and genome location analyses: A comparison of transcription factor deletion and overexpression. *BMC Genomics*, **7**, 215.

Whitfield, M.L., Sherlock, G., Saldanha, A.J., Murray, J.I., Ball, C.A., Alexander, K.E., Matese, J.C., Perou, C.M., Hurt, M.M., Brown, P.O. and Botstein, D. (2002). Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol. Biol. Cell*, **13(6)**, 1977-2000.

Yu, J., Smith, V.A., Wang, P.P., Hartemink, A.J. and Jarvis, E.D. (2004). Advances to Bayesian network inference for generating causal networks from observational biological data. *Bioinformatics***, 20**, 3594-3603.

Zheng, G., Milledge, T., George, E.O. and Narasimhan, G. (2006). Pooling evidence to identify cell cycle-regulated genes. *IWBRA* 2006.