

- heterozygosity and fitness through extra-pair matings. *Nature* **425**, 714–717 (2003).
26. Rico, C., Kuhnlein, U. & Fitzgerald, G. J. Male reproductive tactics in the threespine stickleback – an evaluation by DNA fingerprinting. *Mol. Ecol.* **1**, 79–87 (1992).
 27. Jones, A. G., Östlund-Nilsson, S. & Avise, J. C. A microsatellite assessment of sneaked fertilizations and egg thievery in the fifteen-spine stickleback. *Evolution* **52**, 848–858 (1998).
 28. Vos, C. C., De Jong, A. G., Goedhart, P. W. & Smulders, M. J. M. Genetic similarity as a measure for connectivity between fragmented populations of the moor frog (*Rana arvalis*). *Heredity* **86**, 598–608 (2001).
 29. Marshall, T. C., Slate, J., Kruuk, L. & Pemberton, J. M. Statistical confidence for likelihood-based paternity inference in natural populations. *Mol. Ecol.* **7**, 639–655 (1998).
 30. Neff, B. D. & Pitcher, T. E. Assessing the statistical power of genetic analyses to detect multiple mating in fishes. *J. Fish Biol.* **61**, 739–750 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank T. J. C. Beebee, E. Hespeler and M. Fuerst for help with microsatellite techniques. The team of the Reserva del Refugio (Ursi, Dolores, David, Juan and Javier Abajo), X. González, J. Palanca and N. Palanca provided help and facilities. The Dirección General del Medio Natural de la Diputación General de Aragón issued research permits for the Circo de Piedrafita area. D.R.V. was supported by grants of the University of Vigo, Xunta de Galicia and Deutscher Akademischer Austauschdienst. Grants of the Deutsche Forschungsgemeinschaft to M.V. and A.M. and of the Instituto de Estudios Altoaragoneses to M.V. supported the laboratory work.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.R.V. (vieites@berkeley.edu).

Genomic analysis of regulatory network dynamics reveals large topological changes

Nicholas M. Luscombe^{1*}, M. Madan Babu^{4*}, Haiyuan Yu¹, Michael Snyder², Sarah A. Teichmann⁴ & Mark Gerstein^{1,3}

¹Department of Molecular Biophysics and Biochemistry, ²Department of Molecular, Cellular and Developmental Biology, ³Department of Computer Science, Yale University, PO Box 208114, New Haven, Connecticut 06520-8114, USA

⁴Division of Structural Studies, MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

* These authors contributed equally to this work

Network analysis has been applied widely, providing a unifying language to describe disparate systems ranging from social interactions to power grids. It has recently been used in molecular biology, but so far the resulting networks have only been analysed statically^{1–8}. Here we present the dynamics of a biological network on a genomic scale, by integrating transcriptional regulatory information^{9–11} and gene-expression data^{12–16} for multiple conditions in *Saccharomyces cerevisiae*. We develop an approach for the statistical analysis of network dynamics, called SANDY, combining well-known global topological measures, local motifs and newly derived statistics. We uncover large changes in underlying network architecture that are unexpected given current viewpoints and random simulations. In response to diverse stimuli, transcription factors alter their interactions to varying degrees, thereby rewiring the network. A few transcription factors serve as permanent hubs, but most act transiently only during certain conditions. By studying sub-network structures, we show that environmental responses facilitate fast signal propagation (for example, with short regulatory cascades), whereas the cell cycle and sporulation direct temporal progression through multiple stages (for example, with highly inter-connected transcription factors). Indeed, to drive the latter

processes forward, phase-specific transcription factors inter-regulate serially, and ubiquitously active transcription factors layer above them in a two-tiered hierarchy. We anticipate that many of the concepts presented here—particularly the large-scale topological changes and hub transience—will apply to other biological networks, including complex sub-systems in higher eukaryotes.

We began by assembling a static representation of known regulatory interactions from the results of genetic, biochemical and ChIP (chromatin immunoprecipitation)–chip experiments. Figure 1 illustrates the complexity of the resultant network, which contains 7,074 regulatory interactions between 142 transcription factors and 3,420 target genes (interactions can be between transcription factors and non-transcription factor targets, or two transcription factors). To get a dynamic perspective, we integrated gene-expression data for the following five conditions: cell cycle¹³, sporulation¹⁴, diauxic shift¹², DNA damage¹⁶ and stress response¹⁵. From these data, we traced paths in the regulatory network that are active in each condition using a trace-back algorithm (see Methods).

Figure 1b presents the sub-networks active under different cellular conditions, and gross changes are apparent in the distinct sections of the network that are highlighted. Recent functional genomics studies have analysed the dynamics of a few transcription factors^{17,18}; however, Fig. 1 represents the first dynamic view of a genome-scale network.

Half of the targets are uniquely expressed in only one condition; in contrast, most transcription factors are used across multiple processes. The active sub-networks maintain or rewire regulatory interactions, and over half of the active interactions (1,476 of 2,476 total) are completely supplanted by new ones between conditions. Only 66 interactions are retained across four or more conditions; these comprise ‘hot links’⁶ that are always on (compared with the rest of the network) and mostly regulate house-keeping functions.

The large number of changing interactions makes rigorous comparison of active sub-networks impossible visually. Consequently, we introduce SANDY, an approach that combines: standard measures of network connectivity (involving global topological statistics⁶ and local network motifs⁴), newly derived follow-on statistics, and comparisons against simulated controls to assess the significance of each observation.

Overall, our calculations divide the five condition-specific sub-networks into two categories: endogenous and exogenous (Fig. 1). This allows us to rationalize the different sub-network structures in terms of the biological requirements of each condition. Endogenous processes (cell cycle and sporulation) are multi-stage and operate with an internal transcriptional programme, whereas exogenous states (diauxic shift, DNA damage and stress response) constitute binary events that react to external stimuli with a rapid turnover of expressed genes.

We begin SANDY by examining global topological measures that quantify network architecture (Fig. 1c)⁶. The view from recent studies is that these statistics are remarkably constant across many biological networks (including regulatory systems)^{1,5,6,19,20}. Moreover, most of them remain invariant between randomly simulated sub-graphs of different sizes (Methods).

In fact, we show that topological measures change considerably between the endogenous and exogenous sub-networks. (The probability, P , that their topological measures originate from the same population is $<10^{-4}$; Supplementary Information.) Furthermore, most of the observed measurements differ significantly from random expectation and are insensitive to addition of noise in the underlying network (Methods). The ‘in-degree’ (k_{in}) is the number of incoming edges per node (that is, the number of transcription factors regulating a target). Its average across each sub-network decreases by 20% from endogenous to exogenous conditions ($P < 5 \times 10^{-3}$). The ‘out-degree’ (k_{out}) represents the number of

outgoing edges per node (that is, the number of target genes for each transcription factor). Average values double from endogenous to exogenous conditions ($P < 0.338$; Supplementary Information). The ‘path length’ (l) is the shortest distance between two nodes (here, it is the number of intermediate regulators between a transcription factor and a terminating target gene). Its average halves from endogenous to exogenous conditions ($P < 10^{-3}$). Finally, the ‘clustering coefficient’ (c) gauges the level of inter-connectivity around a node (that is, the level of transcription factor inter-regulation). Values range from 0 for totally dispersed nodes to 1 for fully connected ones. Average coefficients nearly halve from endogenous to exogenous conditions ($P < 10^{-3}$).

In biological terms, the small in-degrees for target genes in exogenous conditions indicate that transcription factors are regulating in simpler combinations, and the large out-degrees signify that each transcription factor has greater regulatory influence by targeting more genes simultaneously. The short paths imply faster propagation of the regulatory signal. Conversely, long paths in the multi-stage, endogenous conditions suggest slower action arising from the formation of regulatory chains to control intermediate phases. Finally, high clustering coefficients in endogenous conditions signify greater inter-regulation between transcription factors. In summary, sub-networks have evolved to produce rapid, large-scale responses in exogenous states, and carefully coordinated processes in endogenous conditions (Fig. 1a).

SANDY also examines the sub-networks locally by calculating the occurrence of network motifs⁴, which are compact, specific patterns of inter-connection between transcription factors and targets. We show the occurrence of the most common motifs in Fig. 1c: single-input, multiple-input, and feed-forward-loop motifs (SIMs, MIMs, and FFLs). In SIMs a single transcription factor targets many genes;

in MIMs multiple transcription factors co-regulate sets of genes; and in FFLs a primary transcription factor regulates a secondary one, and both target a final gene. Motifs appear at similar relative frequencies across regulatory networks of diverse organisms (although individual motifs are not conserved)⁷, and this is also true for the randomly simulated sub-graphs. Therefore, constancy in motif usage is expected across conditions.

However, Fig. 1c shows that the relative occurrence of motifs varies considerably between endogenous and exogenous conditions ($P < 10^{-3}$). SIMs are favoured in exogenous sub-networks where they comprise $>55\%$ of regulatory interactions in motifs. But the frequency drops to $\sim 35\%$ in endogenous processes. Instead, these states favour FFLs ($\sim 44\%$). MIMs do not significantly change in their frequency.

Previous studies defined precise regulatory properties and information processing tasks for motifs⁴. SIMs and MIMs are implicated in conferring similar regulation over groups of genes, so they are ideal for directing the large-scale gene activation found in exogenous conditions. FFLs are buffers that respond only to persistent input signals. They are suited for endogenous conditions, as cells cannot initiate a new stage until the previous one has stabilized. Though used sparingly in exogenous processes they may be important in filtering spurious external stimuli.

Having quantified global and local changes with standard topological measures, we then moved to the follow-on statistics in SANDY (Fig. 2). Like many large-scale networks, the regulatory system displays scale-free characteristics (the probability P_k that a transcription factor targets k genes is proportional to $k^{-\gamma}$ for constant γ). This behaviour (maintained across all active sub-networks) signifies the presence of regulatory hubs targeting disproportionately large numbers of genes. Hubs are of general interest

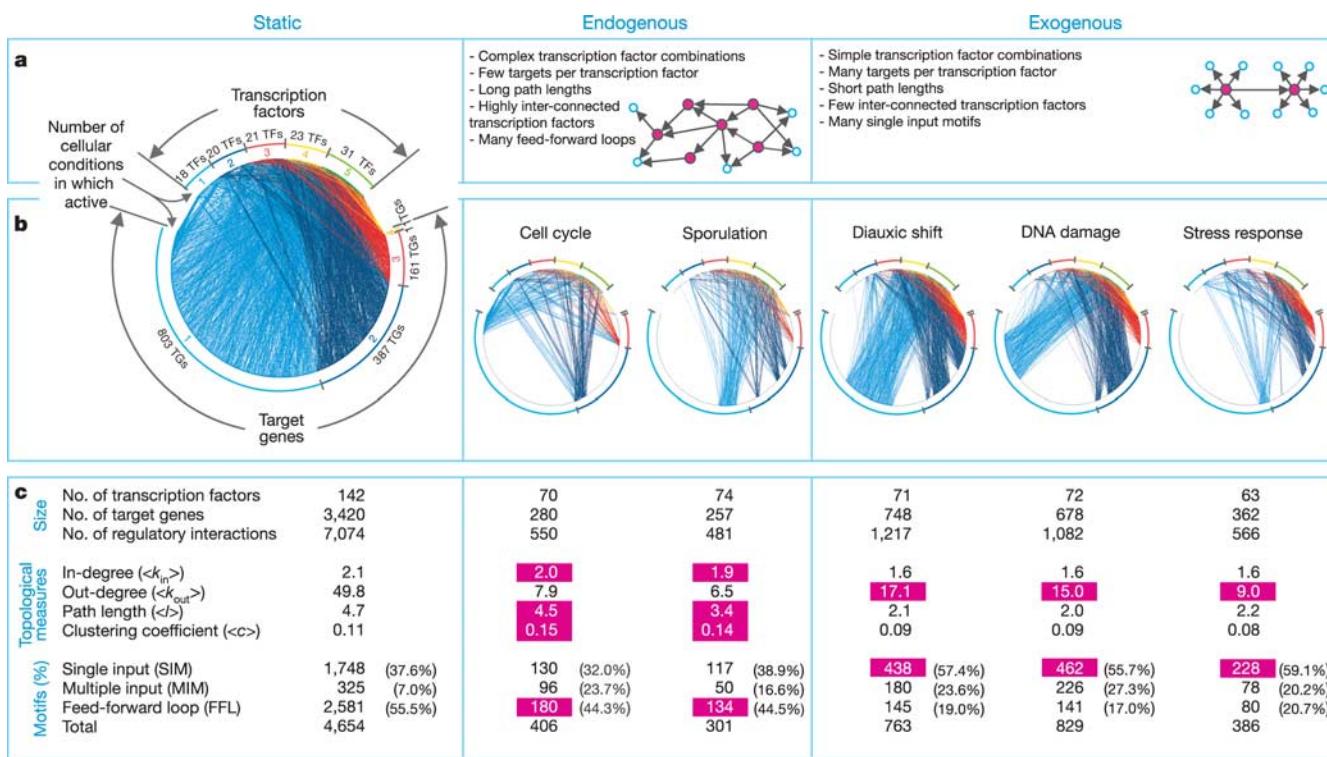


Figure 1 Dynamic representation of the transcriptional regulatory network and standard statistics. **a**, Schematics and summary of properties for the endogenous and exogenous sub-networks. **b**, Graphs of the static and condition-specific networks. Transcription factors and target genes are shown as nodes in the upper and lower sections of each graph respectively, and regulatory interactions are drawn as edges; they are coloured by the number of conditions in which they are active. Different conditions use distinct

sections of the network. **c**, Standard statistics (global topological measures and local network motifs) describing network structures. These vary between endogenous and exogenous conditions; those that are high compared with other conditions are shaded. (Note, the graph for the static state displays only sections that are active in at least one condition, but the table provides statistics for the entire network including inactive regions.)

as they represent the most influential components of a network⁶ and, accordingly, tend to be essential²¹. They are thought to target a broad spectrum of gene functions^{4,11,22} and are commonly located upstream in the network² to expand their influence via secondary transcription factors²³. These observations suggest that hubs would be invariant features of the network across conditions, and this expectation is supported by the random simulations that converge on similar sets of transcription factor hubs.

Figure 2a shows the observed regulatory hubs in each of the five conditions (Methods). They divide into two groups. The smaller one represents permanent hubs, which, in line with expectation, are important regardless of cellular state. They mainly comprise multi-functional transcription factors (such as Abf1) and house-keeping regulators (such as Mig1/2), and are responsible for maintaining the

hot links. However, contrary to expectation, most hubs (78%) are transient; that is, they are influential in one condition, but less so in others. Exogenous conditions have fewer transient hubs, suggesting a more centralized command structure. (This is reflected in different γ ; Supplementary Information.) About half of these hubs are known to be important for their respective conditions (for example, Swi4 in the cell cycle; Methods). For the remainder with sparse annotations, their transient-hub status in a particular condition considerably augments their functional annotation (for example, Sok2 in the cell cycle). These hubs may also relate to condition-dependent lethality, and this has clear implications for identifying specific drug targets.

The defining feature of transient hubs is their capacity to change interactions between conditions. We attempted to quantify this

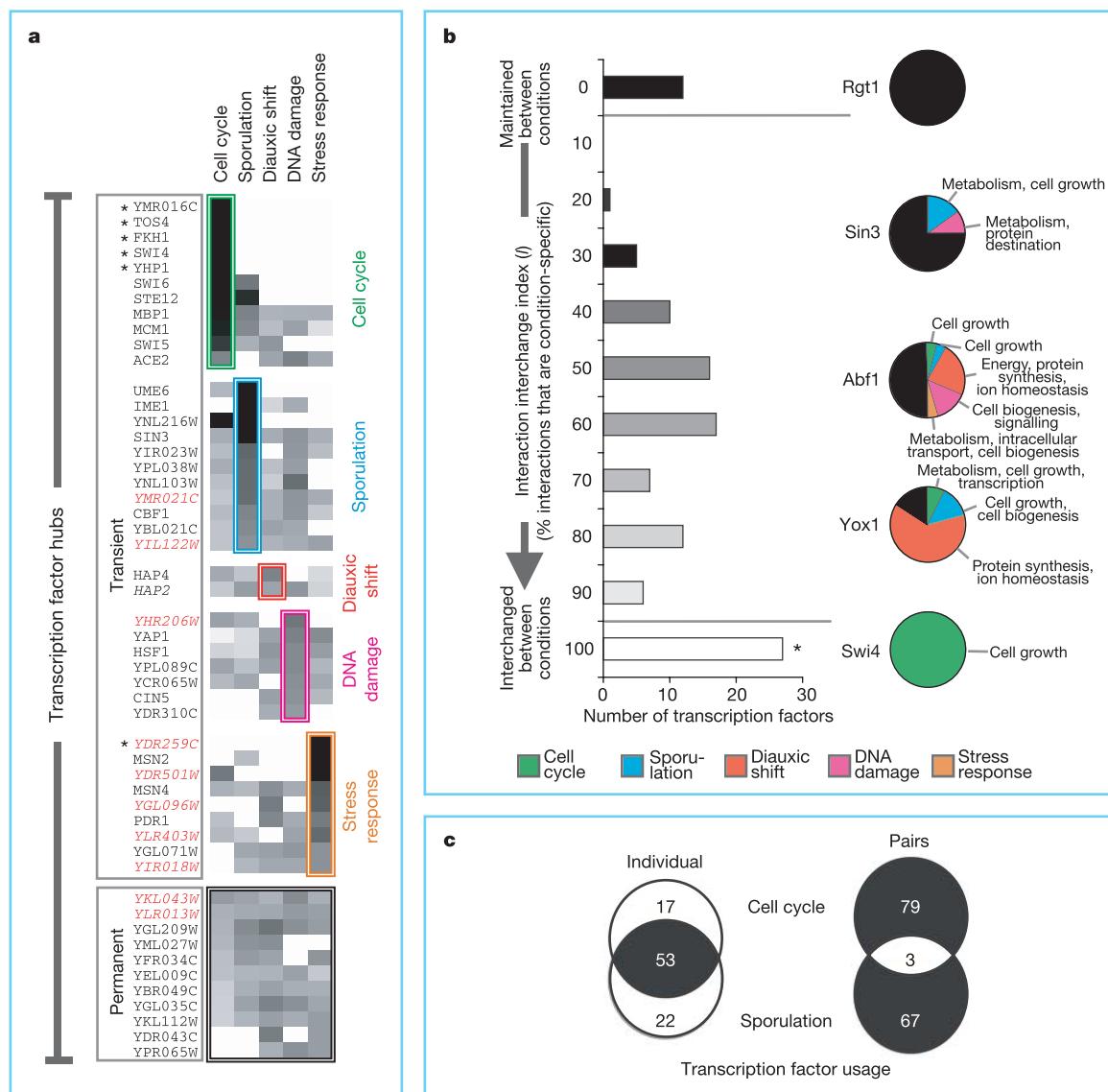


Figure 2 Newly derived 'follow-on' statistics for network structures. **a**, Transcription factor hub usage in different cellular conditions. The cluster diagram shades cells by the normalized number of genes targeted by transcription factor hubs in each condition. One cluster represents permanent hubs and the others condition-specific transient hubs. Genes are labelled with four-letter names when they have an obvious functional role in the condition, and seven-letter open reading frame names when there is no obvious role. Of the latter, gene names are red and italicised when functions are poorly characterized. Starred hubs show extreme interchange index values, $I = 1$. **b**, Interaction interchange index (I) of transcription factors between conditions. A histogram of I for all active transcription

factors shows a uni-modal distribution with two extremes. Pie charts show five example transcription factors with different proportions of interchanged interactions. We list the main functions of the distinct target genes regulated by each example transcription factor. Note how the transcription factors' regulatory functions change between conditions. **c**, Overlap in transcription factor usage between conditions. Venn diagrams show the numbers of individual transcription factors (large intersection) and pair-wise transcription factor combinations (small intersection) that overlap between the two endogenous conditions.

rewiring more broadly for every transcription factor in the network with the interchange index, I . This is defined so that higher values associate with transcription factors replacing a larger fraction of their interactions. Its histogram reveals a uni-modal central distribution with two groups of extreme outliers (Fig. 2b). At one extreme ($I \leq 10\%$), 12 transcription factors retain all interactions across multiple states. At the other end ($I \geq 90\%$), 27 transcription factors replace all interactions in switching conditions. Many of these are so extreme that they only regulate genes in a single condition and are inactive otherwise. These include six transient hubs of known importance for the cell cycle and stress response. Most transcription factors interchange only part of their interactions ($10\% < I < 90\%$). This group comprises most of the hubs; surprisingly, permanent hubs interchange interactions as often as transient ones, but over a larger number of conditions. Furthermore, transcription factors in this group often regulate genes of distinct functions in different conditions, thus shifting regulatory roles. For example, the permanent hub Abf1 regulates cell growth during endogenous conditions, but refocuses to intracellular transport during stress response (in addition to its maintained core functions).

The rewiring highlighted by the interchange index allows transcription factors to be active in many conditions. Indeed, 95 of the 142 transcription factors are used in more than one process (Fig. 1b). Specifically, within endogenous conditions 53 of 92 transcription

factors overlap between cell cycle and sporulation (Fig. 2c), and there is a similar overlap for exogenous conditions (Supplementary Information). With so much intersection in the repertoire of active transcription factors, the precise regulation of a condition cannot arise from the specificity of individual transcription factors. As others have observed²⁴, combinatorial transcription factor usage seems to be the key. We calculated that there are 360 unique pairwise transcription factor combinations (that is, two transcription factors regulating the same target) used in at least one condition. In contrast to individual transcription factors, only a minor proportion of pairs (51 of 360) participate in multiple processes and just 3 of 149 pairs overlap between endogenous conditions (Fig. 2c).

Thus far we have focused on the large dynamic changes occurring between different cellular conditions. However, dynamic transitions also take place within individual processes. Earlier, SANDY defined endogenous sub-networks by their long paths and high clustering. We can study the source of these observations by looking at the full scope of inter-regulation between transcription factors during the cell cycle (Fig. 3). This is possible because in a previous study¹³ expression-level measurements were made throughout the cell cycle and differentially expressed genes were assigned to one of five phases (early G1, late G1, S, G2 and M). We then traced back from the classified genes to identify active sub-networks during each phase (Methods).

A cluster diagram (Fig. 3a) shows that most transcription factors that are active in the cell cycle operate only in a particular phase (for example, Swi4 in late G1). Additionally, a sizeable minority of transcription factors is ubiquitously active throughout the whole cycle. We uncover two major forms of transcription factor inter-regulation. In serial inter-regulation²⁵ (Fig. 3b), the phase-specific transcription factors regulate each other in a sequential manner to drive the cell cycle forward. In fact, we detect complete loops of interactions within the complex circuitry, and the resulting regulatory cascades undoubtedly create the long paths. We also introduce the concept of parallel inter-regulation (Fig. 3c), in which the ubiquitous transcription factors control the phase-specific ones in a two-tiered system. This effectively provides a stable signal to aid the transition between phases. Furthermore, because about a third of ubiquitous transcription factors comprise permanent hubs, they may provide a channel of communication to relate the cell-cycle progression with house-keeping functions. Similar observations apply to sporulation (Supplementary Information).

SANDY presents an approach to examine biological network dynamics. In applying it to the yeast regulatory system, it becomes apparent that many observations made in the static state are not applicable to the condition-specific sub-networks. However, in refocusing to a dynamic perspective, we uncover substantial topological changes in network structure, and we capture the essence of the transcriptional regulatory data in a new way. Because of limitations in current data sets, we can examine this only through integrating gene-expression information. However, we anticipate future experiments to determine condition-specific interactions directly. Given the robustness of the observations to large perturbations (Methods), we expect our approach and findings to remain valid for these new data sets. Furthermore, we anticipate that many of the concepts we have introduced could be readily transferred to other types of biological networks and complex sub-systems in multicellular organisms, such as those directing the circadian cycle²⁶ and cellular development. □

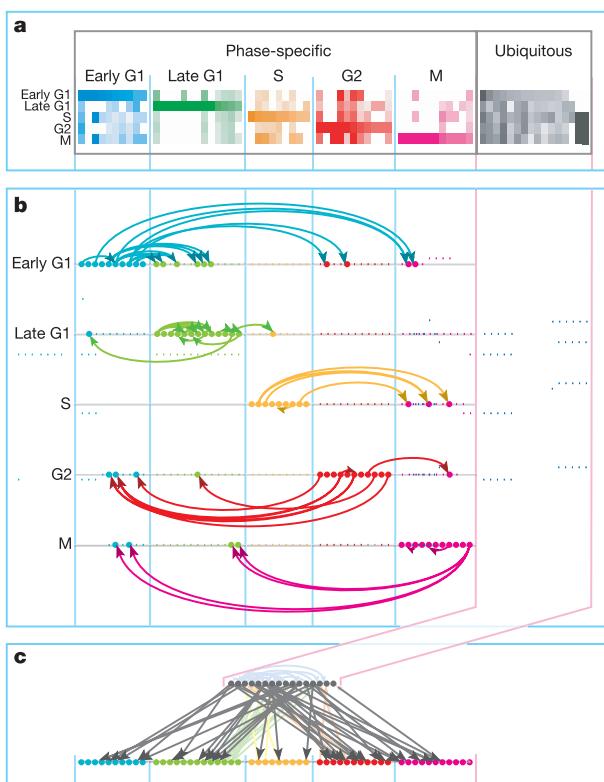


Figure 3 Transcription factor inter-regulation during the cell cycle time-course. **a**, The 70 transcription factors active in the cell cycle. The diagram shades each cell by the normalized number of genes targeted by each transcription factor in a phase. Five clusters represent phase-specific transcription factors and one cluster is for ubiquitously active transcription factors. Note, both hub and non-hub transcription factors are included. Transcription factor names are given in the Supplementary Information. **b**, Serial inter-regulation between phase-specific transcription factors. Network diagrams show transcription factors that are active in one phase regulate transcription factors in subsequent phases. In the late phases, transcription factors apparently regulate those in the next cycle. **c**, Parallel inter-regulation between phase-specific and ubiquitous transcription factors in a two-tiered hierarchy. Serial and parallel inter-regulation operate in tandem to drive the cell cycle while balancing it with basic house-keeping processes.

Methods

Detailed descriptions of the methods are in the Supplementary Information and at <http://sandy.topnet.gersteinlab.org>.

Data sets

The transcriptional regulatory network was assembled from the results of genetic, biochemical and ChIP-chip experiments^{9–11}. The gene-expression data were compiled from 240 microarray experiments for five conditions^{12–16}. We identified the following

numbers of genes with differential expression: cell cycle, 455; sporulation, 477; diauxic shift, 1,823; DNA damage, 1,718; and stress response, 866.

Trace-back algorithm

We used the following algorithm to define sections of the regulatory network used in each condition: (1) identify transcription factors as being 'present' in a condition if they have sufficiently high expression levels; (2) flag differentially expressed genes that appear in the regulatory network; (3) mark as 'active' the regulatory links between present transcription factors and differentially expressed genes; and (4) search for any other present transcription factors that are linked to a transcription factor with an already active link and make this connection active. The last step is repeated until no more links are made active. The same procedure identifies sub-networks that are active in particular phases of cell cycle and sporulation.

SANDY

This extends the methodology used by the TopNet software tool²⁷ and it evaluates each sub-network with the following: (1) standard statistics, including global measures of topology (k_{in} , k_{out} , l , and c)⁶ and local motif occurrence (SIM, MIM and FFL)¹; (2) follow-on statistics, including permanent and transient hub identification, interchange index (I) and counting the overlap in transcription factor usage (individual and pairs) across multiple conditions. (Hubs are transcription factors in the top 30% by number of target genes, in at least one condition. The number of target genes is normalized to measure the relative influence of a transcription factor hub in a particular process.) In all cases, regulatory functions are obtained from the Saccharomyces Genome Database²⁸ and are current as of June 2004; and (3) a comparison of observations and random expectation by simulating sub-graphs that are similar in size to each sub-network, and calculating standard and follow-on statistics for them. Simulated sub-graphs sample the same number of differentially expressed genes and trace through the static network. We also tested the sensitivity of our observations to noise by randomly perturbing the static networks by 30% (random addition, deletion and replacement of interactions), tracing back from the original differentially expressed genes and then recalculating the statistics.

Received 15 January; accepted 24 June 2004; doi:10.1038/nature02782.

1. Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabasi, A. L. The large-scale organization of metabolic networks. *Nature* **407**, 651–654 (2000).
2. Guelzim, N., Bottani, S., Bourgine, P. & Kepes, F. Topological and causal structure of the yeast transcriptional regulatory network. *Nature Genet.* **31**, 60–63 (2002).
3. Milo, R. *et al.* Network motifs: simple building blocks of complex networks. *Science* **298**, 824–827 (2002).
4. Shen-Orr, S. S., Milo, R., Mangan, S. & Alon, U. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genet.* **31**, 64–68 (2002).
5. Oltvai, Z. N. & Barabasi, A. L. Systems biology. Life's complexity pyramid. *Science* **298**, 763–764 (2002).
6. Barabasi, A. L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nature Rev. Genet.* **5**, 101–113 (2004).
7. Milo, R. *et al.* Superfamilies of evolved and designed networks. *Science* **303**, 1538–1542 (2004).
8. Teichmann, S. A. & Babu, M. M. Gene regulatory network growth by duplication. *Nature Genet.* **36**, 492–496 (2004).
9. Svetlov, V. V. & Cooper, T. G. Review: compilation and characteristics of dedicated transcription factors in *Saccharomyces cerevisiae*. *Yeast* **11**, 1439–1484 (1995).
10. Horak, C. E. *et al.* Complex transcriptional circuitry at the G1/S transition in *Saccharomyces cerevisiae*. *Genes Dev.* **16**, 3017–3032 (2002).
11. Lee, T. I. *et al.* Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298**, 799–804 (2002).
12. DeRisi, J. L., Iyer, V. R. & Brown, P. O. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680–686 (1997).
13. Cho, R. J. *et al.* A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol. Cell* **2**, 65–73 (1998).
14. Chu, S. *et al.* The transcriptional program of sporulation in budding yeast. *Science* **282**, 699–705 (1998).
15. Gasch, A. P. *et al.* Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell* **11**, 4241–4257 (2000).
16. Gasch, A. P. *et al.* Genomic expression responses to DNA-damaging agents and the regulatory role of the yeast ATR homolog Mec1p. *Mol. Biol. Cell* **12**, 2987–3003 (2001).
17. Odom, D. T. *et al.* Control of pancreas and liver gene expression by HNF transcription factors. *Science* **303**, 1378–1381 (2004).
18. Zeitlinger, J. *et al.* Program-specific distribution of a transcription factor dependent on partner transcription factor and MAPK signaling. *Cell* **113**, 395–404 (2003).
19. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442 (1998).
20. Wagner, A. & Fell, D. A. The small world inside large metabolic networks. *Proc. R. Soc. Lond. B* **268**, 1803–1810 (2001).
21. Yu, H., Greenbaum, D., Xin Lu, H., Zhu, X. & Gerstein, M. Genomic analysis of essentiality within protein networks. *Trends Genet.* **20**, 227–231 (2004).
22. Martinez-Antonio, A. & Collado-Vides, J. Identifying global regulators in transcriptional regulatory networks in bacteria. *Curr. Opin. Microbiol.* **6**, 82–489 (2003).
23. Madan Babu, M. & Teichmann, S. A. Evolution of transcription factors and the gene regulatory network in *Escherichia coli*. *Nucleic Acids Res.* **31**, 1234–1244 (2003).
24. Pilpel, Y., Sudarsanam, P. & Church, G. M. Identifying regulatory networks by combinatorial analysis of promoter elements. *Nature Genet.* **29**, 153–159 (2001).
25. Simon, I. *et al.* Serial regulation of transcriptional regulators in the yeast cell cycle. *Cell* **106**, 697–708 (2001).
26. Ueda, H. R. *et al.* A transcription factor response element for gene expression during circadian night. *Nature* **418**, 534–539 (2002).

27. Yu, H., Zhu, X., Greenbaum, D., Karro, J. & Gerstein, M. TopNet: a tool for comparing biological sub-networks, correlating protein properties with topological statistics. *Nucleic Acids Res.* **32**, 328–337 (2004).
28. Christie, K. R. *et al.* Saccharomyces Genome Database (SGD) provides tools to identify and analyze sequences from *Saccharomyces cerevisiae* and related sequences from other organisms. *Nucleic Acids Res.* **32**, D311–D314 (2004).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank P. Bertone, N. Domedel-Puig, E. Hovig, R. Jansen, K. Kleivi, G. Koentges, E. Koonin, B. Lenhard, A. Paccanaro, J. Rozowsky, J. Tegnér, V. Trifunovic, A. Todd, Y. Xia and H. Zao for comments on the paper. N.M.L. thanks the Anna Fuller Fund and the MRC LMB Visitor's Program. M.M.B. acknowledges financial support from the Cambridge Commonwealth Trust, Trinity College, Cambridge and the MRC LMB. M.G. is supported by the NSF and NIH.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to N.M.L., M.M.B., S.A.T. & M.G. (sandy@bioinfo.mbb.yale.edu).

Long-lasting self-inhibition of neocortical interneurons mediated by endocannabinoids

Alberto Bacci, John R. Huguenard & David A. Prince

Department of Neurology and Neurological Sciences, Stanford University School of Medicine, Stanford, California 94305, USA

Neocortical GABA-containing interneurons form complex functional networks responsible for feedforward and feedback inhibition and for the generation of cortical oscillations associated with several behavioural functions^{1,2}. We previously reported that fast-spiking (FS), but not low-threshold-spiking (LTS), neocortical interneurons from rats generate a fast and precise self-inhibition mediated by inhibitory autaptic transmission³. Here we show that LTS cells possess a different form of self-inhibition. LTS, but not FS, interneurons undergo a prominent hyperpolarization mediated by an increased K^+ -channel conductance. This self-induced inhibition lasts for many minutes, is dependent on an increase in intracellular $[Ca^{2+}]$ and is blocked by the cannabinoid receptor antagonist AM251, indicating that it is mediated by the autocrine release of endogenous cannabinoids. Endocannabinoid-mediated slow self-inhibition represents a powerful and long-lasting mechanism that alters the intrinsic excitability of LTS neurons, which selectively target the major site of excitatory connections onto pyramidal neurons; that is, their dendrites^{4–7}. Thus, modulation of LTS networks after their sustained firing will lead to long-lasting changes of glutamate-mediated synaptic strength in pyramidal neurons, with consequences during normal and pathophysiological cortical network activities.

We obtained whole-cell recordings from fast-spiking and LTS interneurons in layer V of rat somatosensory neocortical slices. Cells were characterized electrophysiologically by their firing behaviours as described previously⁸ (Fig. 1a, c, insets; see also Methods). All biocytin-filled LTS interneurons contained cholecystokinin (CCK) ($n = 7$; Supplementary Fig. S1), whereas FS cells were CCK-negative ($n = 4$; Supplementary Fig. S1), indicating that we sampled a homogeneous non-FS cell population.

Although LTS interneurons lack GABAergic autaptic transmission³, we found a unique form of slow self-inhibition (SSI) in

Genomic analysis of regulatory network dynamics reveals large topological changes

Supplementary Material

Nicholas M Luscombe^{1*+}, M Madan Babu^{2*+}, Haiyuan Yu¹,
Michael Snyder³, Sarah A Teichmann²⁺ and Mark Gerstein^{1,4+}

Department of Molecular Biophysics and Biochemistry¹,
Department of Molecular, Cellular and Developmental Biology³,
Department of Computer Science⁴,
Yale University
PO Box 208114, New Haven, CT 06520-8114, USA

Division of Structural Studies²,
MRC Laboratory of Molecular Biology,
Hills Road, Cambridge CB2 2QH, UK

+ Correspondence should be addressed to:
Email: sandy@bioinfo.mbb.yale.edu;
Tel: +1 203 432 6105 Fax: +1 360 838 7861

Contents

CONTENTS.....	2
ABBREVIATIONS	4
1. SUPPLEMENTARY WEBSITE.....	4
METHODS	
2. DATASETS.....	5
2.1 TRANSCRIPTIONAL REGULATORY NETWORK DATA.....	5
<i>2.1.1 Compilation of dataset</i>	5
<i>2.1.2 Contents of dataset</i>	5
2.2 GENE EXPRESSION DATA.....	5
<i>2.2.1 Compilation of dataset</i>	5
<i>2.2.2 Classification of endogenous and exogenous conditions</i>	5
<i>2.2.3 Contents of dataset</i>	6
2.3 AVAILABILITY	6
3. TRACE-BACK ALGORITHM TO IDENTIFY ACTIVE SUB-NETWORKS OF CELLULAR CONDITIONS	7
3.1 OUTLINE OF ALGORITHM	7
3.2 DEFINING DIFFERENTIALLY EXPRESSED GENES.....	7
3.3 IDENTIFYING PRESENT TFS	7
3.4 TRACING A CONDITION-SPECIFIC SUB-NETWORK	8
3.5 AVAILABILITY	8
4. STATISTICAL ANALYSIS OF NETWORK DYNAMICS (SANDY)	9
4.1 WELL-KNOWN STATISTICS	9
<i>4.1.1 Global topological measures</i>	9
<i>4.1.2 Local network motifs</i>	11
4.2 FOLLOW-ON STATISTICS	12
<i>4.2.1 TF hub usage</i>	12
<i>4.2.2 Interaction interchange index</i>	12
<i>4.2.3 TF usage across multiple conditions</i>	12
4.3 STATISTICAL VALIDATION WITH RANDOMLY SIMULATED NETWORKS	13
<i>4.3.1 Glossary of mathematical notations</i>	13
<i>4.3.2 Randomly simulated networks</i>	14
<i>4.3.3 Comparison between two observed condition-specific sub-networks</i>	15
<i>4.3.4 Comparison between one observed condition-specific sub-network and the randomly simulated networks</i>	18
<i>4.3.5 Comparisons between two observed condition-specific sub-networks with respect to expectations from random simulations</i>	20
<i>4.3.6 Sensitivity analysis</i>	26
4.4 AVAILABILITY	27
5. IDENTIFYING ACTIVE SUB-NETWORKS OF ENDOGENOUS CONDITION TIMES-COURSES	28
5.1 DEFINING DIFFERENTIALLY EXPRESSED GENES DURING ENDOGENOUS CONDITION PHASES	28
5.2 IDENTIFYING PRESENT TFS	28
5.3 TRACING PHASE-SPECIFIC SUB-NETWORKS	29
5.4 CLUSTERING TIME-DEPENDENT TF USAGE.....	29
5.4 AVAILABILITY	29

RESULTS

6. SANDY REVEALS LARGE TOPOLOGICAL CHANGES.....	30
6.1 WELL-KNOWN STATISTICS	30
<i>6.1.1 Global topological measures</i>	<i>30</i>
<i>6.1.2 Local network motifs.....</i>	<i>32</i>
6.2 FOLLOW-ON STATISTICS	34
<i>6.2.1 TF hub usage.....</i>	<i>34</i>
<i>6.2.2 Interaction interchange index.....</i>	<i>36</i>
<i>6.2.3 TF usage across multiple conditions.....</i>	<i>38</i>
6.3 STATISTICAL VALIDATION WITH RANDOMLY SIMULATED NETWORKS	40
<i>6.3.1 Comparison between two observed condition-specific sub-networks</i>	<i>40</i>
<i>6.3.2 Comparison between one observed condition-specific sub-network and the randomly simulated networks.....</i>	<i>42</i>
<i>6.3.3 Comparisons between two observed condition-specific sub-networks with respect to expectations from random simulations</i>	<i>43</i>
<i>6.3.4 Sensitivity analysis</i>	<i>47</i>
6.4 AVAILABILITY	47
7. NETWORK DYNAMICS DURING THE TIME-COURSES OF ENDOGENOUS CONDITIONS.....	48
7.1 PHASE-SPECIFIC AND UBIQUITOUS TFs	48
7.2 SERIAL AND PARALLEL INTER-REGULATION BETWEEN TFs	49
7.3 NETWORK DIAGRAM OF ENDOGENOUS CONDITION TIME-COURSES	51
7.4 AVAILABILITY	53
8. REFERENCES	54

Abbreviations

Network terms

TF transcription factor

Cellular condition terms

cc cell cycle
sp sporulation
ds diauxic shift
dd DNA damage
sr stress response

Network motif terms

SIM single input motif
MIM multiple input motif
FFL feed-forward loop

Methodological terms

TBA trace-back algorithm
SANDY statistical analysis of network dynamics

1. Supplementary website

Additional material accompanying the paper is available online at <http://sandy.topnet.gersteinlab.org>.

2. Datasets

2.1 Transcriptional regulatory network data

2.1.1 Compilation of dataset

We compiled the transcriptional regulatory data in *Saccharomyces cerevisiae* from the results of genetic, biochemical and ChIP-chip experiments (1-4).

We removed non-DNA-binding TFs that do not regulate through promoter binding. This was done by a sequence search of 156 known DNA-binding motifs from Pfam (current as of June 2004) (5) against the protein sequences of the TFs in the dataset. TFs without significant matches were removed.

2.1.2 Contents of dataset

The final dataset used in the paper contained the following:

network component	number
TFs	142
non-TF targets	3,420
regulatory interactions	7,074

Regulatory interactions exist between TFs and non-TF targets, and among TFs themselves.

2.2 Gene expression data

2.2.1 Compilation of dataset

We compiled the gene expression data from 240 published microarray experiments for five cellular conditions: cell cycle (6), sporulation (7), diauxic shift (8), DNA damage (9), and stress response (10).

We obtained lists of genes displaying significant gene expression changes during each condition from the original publications. Genes can be up- or down-regulated, but we do not differentiate between the two here.

2.2.2 Classification of endogenous and exogenous conditions

We classified cellular conditions as being endogenous (cell cycle, sporulation) or exogenous (diauxic shift, DNA damage, stress response). These are defined by whether conditions have an internal, multi-stage program of regulation (endogenous), or a binary, large-scale turnover of genes triggered by external stimuli (exogenous). Although

sporulation is initiated by an environmental stimulus (*ie* nitrogen depletion), we have classified it as endogenous because of its multi-stage transcriptional program.

2.2.3 Contents of dataset

The dataset used in the paper contained the following numbers of differentially expressed genes:

classification	condition	number of differentially expressed genes	reference
endogenous	cell cycle	455	(6)
	sporulation	477	(7)
exogenous	diauxic shift	1,823	(8)
	DNA damage	1,718	(9)
	stress response	866	(10)

2.3 Availability

A partial dataset of the transcriptional regulatory network (excluding licensed data) is available from the Supplementary Website. The original gene expression data are only available from the respective websites accompanying the original publications, however we provide lists of differentially expressed genes for each condition on the Supplementary Website.

3. Trace-back algorithm to identify active sub-networks of cellular conditions

3.1 Outline of algorithm

We traced paths in the regulatory network that are active during each cellular condition using a trace-back algorithm (TBA). The method assumes that genes displaying differential expression are regulated by TFs linked to them in the regulatory network. The full procedure involved the steps outlined below.

3.2 Defining differentially expressed genes

The description of how we defined differentially expressed genes is in §2.2.1.

3.3 Identifying present TFs

We identified TFs in the static regulatory network that can potentially participate in regulating a condition by determining their presence or absence in the cell from their expression levels.

We define a starting expression level for each TF using a reference dataset describing their protein and mRNA abundance (11); this dataset scales many mRNA and protein measurements over the cell cycle. From these data, we grouped TFs into those showing high, medium or low abundance:

relative cell cycle abundance	number of TFs
High	17
Medium	62
Low	63

For each condition, we then determine the presence or absence of each TF by assessing expression level changes relative to the starting abundance. This is a reasonable assumption as the original microarray experiments for all cellular conditions were conducted with the cell cycle as the start point.

We determined the presence or absence of TFs using the following rules:

relative cell cycle abundance	expression level change during condition	present or absent
high	up	present
	constant	present
	down	present
medium	up	present
	constant	present
	down	absent
low	up	present
	constant	absent
	down	absent

The numbers of present TFs in each condition are as follows:

cellular condition	number of present TFs
cell cycle	88
sporulation	85
diauxic shift	76
DNA damage	75
diauxic shift	85

3.4 Tracing a condition-specific sub-network

We traced the active regions of the regulatory network for each cellular condition as follows: (i) We flagged differentially expressed genes that appear in the regulatory network. (ii) We marked as active the regulatory interactions between present TFs and differentially expressed genes. (iii) We searched for any other present TFs that are linked to a TF with interactions that are already active, and made this connection active. (iv) We repeated step (iii) until no more interactions were made active. (v) The collection of all active interactions and inter-connected TFs and target genes comprise the condition-specific sub-network.

3.5 Availability

Lists of present TFs, a Perl script for the TBA and data files describing the condition-specific sub-networks are available from the Supplementary Website.

4. Statistical Analysis of Network Dynamics (SANDY)

Although it is clear that distinct regions of the regulatory network are used under different cellular conditions, it is difficult to discern the differences between them visually.

SANDY provides a framework to quantify the topological changes that condition-specific sub-networks undergo. The analysis is broadly separated into three parts:

- i. **Well-known statistics** (global topological measures, local network motifs).
- ii. **Newly-derived follow-on statistics** (hub usage, interchange index, TF usage).
- iii. **Statistical validation** with randomly simulated networks.

4.1 Well-known statistics

4.1.1 Global topological measures

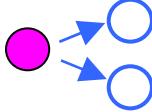
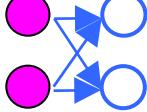
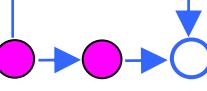
Global topological measures quantify the large-scale structural architecture of the network (12,13). The measures we calculated are listed on the following page.

topological measure	symbol	schematic	description
<i>in-degree</i>	k_{in}		The number of direct incoming edges per node (<i>i.e.</i> number of TFs per target gene). Its average is the mean across all nodes in the network
<i>out-degree</i>	k_{out}		The number of direct outgoing edges per node (<i>i.e.</i> number of target genes per TF). Its average is the mean across all TFs in the network.
<i>degree distribution exponents</i>	β, γ		The most suitable distributions for the in- and out-degrees was found by calculating the best-fitting exponential ($P_k \sim e^{-\beta k}$) and power-law ($P_k \sim k^\gamma$) distributions, where P_k is the probability that a randomly picked node has k interactions. We found the best-fitting distribution by minimising the root mean squared difference between the actual and fitted distributions. In all networks, the in-degree showed exponential and the out-degree power-law behaviours.
<i>path length</i>	l		The shortest distance (in number of intermediate nodes) between two nodes. Here, it is calculated as the shortest distance between each TF and its terminating target genes. Its average is the mean across all paths in the network.
<i>diameter</i>	d		The maximum path length in the network.
<i>clustering coefficient</i>	c		The ratio of the number of edges between a node's neighbours and the maximum number of possible edges; c_i for a node i is calculated as $\frac{2E_i}{k_i(k_i - 1)}$. Here, values of c are calculated using only the TF nodes in the network and the directionality of edges is ignored. Its average is the mean across all TFs in the network.

4.1.2 Local network motifs

Local network motifs are precise patterns of inter-connections between a small number of TFs and target genes (14). We search for three of the most common motifs described below.

In order to identify the motifs, we constructed a pair of affinity matrices A and B . Matrix A contained binary entries A_{ij} where a 1 indicated a regulatory interaction from TF j to target gene i . Matrix B was a sub-matrix of A , containing only the rows corresponding to target genes that are TFs themselves. Nodes and edges can be part of more than one motif.

network motif	abbreviation	schematic	description
<i>single input motif</i>	<i>SIM</i>		We identify the subset of rows in B , such that the sum of each row is 1. For each TF column, we then find non-zero entries.
<i>multiple input motif</i>	<i>MIM</i>		We identify the subset of rows in A so that the sum of each row is greater than 1. Then for each row, we identify other rows regulated by the same set of TFs. The collection of rows represents a motif.
<i>feed-forward loop</i>	<i>FFL</i>		For each primary TF we identify non-zero entries in B , which correspond to regulated secondary TFs. For each primary and secondary TF pair, we then identify all rows in A regulated by both TFs.

4.2 Follow-on statistics

4.2.1 TF hub usage

Hubs were defined as TFs in the top 30%, by number of targets, in at least one of the five cellular conditions.

We normalized the number of target genes of each TF hub using $P_{ij} = \frac{n_{ij}}{\sum_i n_{ij}} \sqrt{\frac{\sum_j n_{ij}}{\sum_i \sum_j n_{ij}}}$,

where P_{ij} is the propensity that a given TF i regulates n genes in condition j relative to the total number of genes it regulates under all conditions. This provides a measure of the relative influence of a TF as a regulatory hub in a particular cellular condition.

Hubs were clustered according to their propensity values using a k -means clustering algorithm (15) with $k = 6$, where k is the pre-defined number of clusters. The value of k was chosen so as to group the five cellular conditions separately (transient hubs), and also provide an additional cluster for the permanent hubs. Tests using $k = 4 - 8$ resulted in similar clusters with a few outlying TFs.

4.2.2 Interaction interchange index

The interchange index (I) measures the degree to which TFs maintain or replace regulatory interactions between cellular conditions. It is calculated as $I = \frac{\sum_i n_i}{N} \times 100\%$

where n_i is the number of regulatory interactions unique to condition i and N is the total number of regulatory interactions active in all conditions. It effectively measures the fraction of regulatory interactions that are unique to a particular condition. Values of I range from 0% to 100%; lower values indicate greater proportions of interactions being maintained across multiple conditions, and high values indicate that most interactions are interchanged, so are unique to particular conditions.

4.2.3 TF usage across multiple conditions

Individual TF usage counts the number of TFs that are used in single conditions or across multiple conditions. Pair-wise TF combination usage counts the number of TF pairs that target the same gene together in single conditions or across multiple conditions.

For the paper, we count only the usage of individual TFs and pair-wise combinations targeting genes that are uniquely expressed in a single condition. This allows us to examine the source of regulatory specificity for each condition.

4.3 Statistical validation with randomly simulated networks

4.3.1 Glossary of mathematical notations

Here we describe the statistical tests that we employed to determine whether the observations that we make for the condition-specific sub-networks are significant, when compared to random expectation.

Below we provide a glossary of the mathematical notations that we use in describing the tests.

notation	description
o	Observed value (typically for condition-specific sub-networks).
e	Expected value (typically from random simulations).
m, n	Networks of a given size typically corresponding to a particular cellular condition.
k	Subscript of a single network from the set of simulated networks of a given size.
T	Collection of all global topological measures (<i>i.e.</i> k_{in} , k_{out} , l and c).
t	A single global topological measure (<i>i.e.</i> k_{in} , k_{out} , l or c).
F	Collection of all local network motifs (<i>i.e.</i> SIMs, MIMs and FFLs).
f	A single local network motif (<i>i.e.</i> SIMs, MIMs or FFLs).
N	Occurrence of local network motifs in a network.
d	Difference between two networks of the means of a single global topological measure ($\langle t \rangle$).
Z	Normalised difference between two networks of: <ul style="list-style-type: none"> the means of a single global topological measure ($\langle t \rangle$). the occurrence of all local network motifs (F).
S	Compiled difference between two networks of the means of all global topological measures ($\langle T \rangle$).
R	Pooled difference between endogenous and exogenous sub-networks of the means of a single global topological measure ($\langle t \rangle$).
Q	Overall difference between endogenous and exogenous sub-networks: <ul style="list-style-type: none"> the means of all global topological measures ($\langle T \rangle$). the occurrence of all local network motifs (F).

4.3.2 Randomly simulated networks

We generated 1,000 simulated networks as controls for each cellular condition. For each simulated network, we:

- i. Sampled the same number of “differentially expressed” genes out of all yeast genes for the given condition.
- ii. Sampled the same number of “present” TFs from the list of all TFs.
- iii. Traced through the static network using the sampled genes and TFs.

The randomly simulated networks range in size between 232 and 746 nodes. For each network we calculated both the well-known and follow-on statistics to quantify the topologies of the random controls.

4.3.3 Comparison between two observed condition-specific sub-networks

We used the following methods to determine if topological differences observed between two (or more) condition-specific sub-networks are statistically significant. The resulting statistics are discussed in §6.3.1 (we do not quote these in the main paper).

network statistic	significance test
global topological measures	<ol style="list-style-type: none"> 1. <i>Significance test between two observed condition-specific sub-networks.</i> For a single topological measure, we calculated the statistical significance of the difference between a pair of observed condition-specific sub-networks. <ol style="list-style-type: none"> i. We performed a Mann-Whitney U-test to compare the distributions of each topological measure t (k_{in}, k_{out}, l, c) separately between two sub-networks m and n (eg cell cycle and sporulation). ii. We rejected the null hypothesis H_0 (that the distributions of the topological measure in the two observed sub-networks originate from the same population) if $p \leq 0.01$. 2. <i>Pooled significance test between observed endogenous and exogenous sub-networks</i> For a single topological measure, we calculated the pooled statistical significance of the difference between the observed endogenous and exogenous categories of condition-specific sub-networks. <ol style="list-style-type: none"> i. In order to group the observed condition-specific sub-networks that display similar topologies, we examined the pair-wise p-values between them (calculated above). We “linked” sub-networks if their pair-wise $p > 0.01$ (ie the difference in the topological measure is non-significant). We then clustered all of the sub-networks based on these links using complete linkage clustering (ie every member of a cluster must be linked together). In practise for every topological measure, this grouped the two endogenous conditions (cell cycle, sporulation) and three exogenous conditions (diauxic shift, DNA damage, stress response). ii. Within each cluster (ie endogenous, exogenous), we pooled (ie concatenated) the distributions of the topological measure t from the member sub-networks. We then performed a Mann-Whitney U-test (in a similar fashion to above) to compare the pooled distributions of the topological measure in the observed endogenous and exogenous sub-networks. iii. We rejected the null hypothesis H_0 (that the distributions of topological measures in the observed endogenous and exogenous sub-networks originate from the same population) if $p \leq 0.01$.

1. *Significance test between two observed condition-specific sub-networks.*
 We calculated the statistical significance of the difference in motif usage (*ie* number of SIMs, MIMs, FFLs) between a pair of condition-specific sub-networks.
 - i. For two observed sub-networks, we produced a contingency table (three rows for motifs; two columns for sub-networks) in which each cell contained the observed occurrence (N_{fm}^o) of motif f in sub-network m . We produced another table containing the expected motif distribution for the two sub-networks; the expected occurrence (N_{fm}^e)

$$N_m^{fe} = \frac{\sum_f N_{fo}^o \sum_m N_m^o}{\sum_{fm} N_{fm}^o}.$$
 - ii. We performed a χ^2 -test with two degrees of freedom between the distributions of observed and expected motif occurrences.
 - iii. We rejected the null hypothesis H_0 (that the motif usage in the two observed sub-networks originate from the same population) if $p \leq 0.01$.
2. *Pooled significance test between observed endogenous and exogenous sub-networks.*
 We calculated the pooled statistical significance of the difference in motif usage between the endogenous and exogenous categories of condition-specific sub-networks.
 - i. We clustered the observed condition-specific sub-networks by examining their pair-wise p -values from above (in a similar fashion to above). In practise this grouped the two endogenous conditions and three exogenous conditions together.
 - ii. Within each cluster, we summed the occurrences of each motif type from the member sub-networks. We produced contingency tables describing the observed and expected motif occurrences (in a similar to above) for the pooled endogenous and exogenous sub-networks.
 - iii. We performed a χ^2 -test with two degrees of freedom (in a similar fashion to above).
 - iv. We rejected the null hypothesis H_0 (that the motif usage in the observed endogenous and exogenous sub-networks originate from the same population) if $p \leq 0.01$.

hub usage

1. *Overlap in hubs between two observed condition-specific sub-networks.*
We calculated the percentage overlap in TFs that are classified as hubs between each pair of observed condition-specific sub-networks.
2. *Correlation of numbers of target genes between two observed condition-specific sub-networks.*
For TF hubs that overlap between observed sub-networks, we calculated the Pearson correlation coefficients of the number of genes that they target.

4.3.4 Comparison between one observed condition-specific sub-network and the randomly simulated networks

We used the following methods to determine if the topologies of the observed condition-specific sub-networks differ from those of randomly simulated networks of similar sizes. The resulting statistics are discussed in §6.3.2 (we do not quote these in the main paper).

network statistic	significance test
<i>global topological measures</i>	<ul style="list-style-type: none"> • <i>Significance test between one observed condition-specific sub-network and the corresponding randomly simulated networks.</i> <p>For a single topological measure, we calculated the statistical significance of the difference in the observed mean for a condition-specific sub-network compared with the expected (<i>i.e.</i> random) means in simulated networks of similar size.</p> <ol style="list-style-type: none"> We calculated an expected distribution of the mean topological measure ($\langle t_m^e(k) \rangle$) using the set of simulated networks of size m. Simulated networks are indicated by the subscript k which typically runs from 1 to 1,000 (<i>i.e.</i> 1,000 simulated networks of size m). We counted the number of these expected means ($\langle t_m^e(k) \rangle$) that are greater than the observed mean $\langle t_m^o \rangle$ in the corresponding condition-specific sub-network (<i>i.e.</i> $\langle t_m^e(k) \rangle > \langle t_m^o \rangle$). We calculated a p-value as the fraction of these expected means that satisfied this requirement with respect to the total number of comparisons (<i>i.e.</i> 1,000 comparisons). We rejected the null hypothesis H_0 (that the topological measure in an observed condition-specific sub-network originates from the same population as that of the simulated networks) if $p \leq 0.01$ or $p \geq 0.99$.
<i>local network motifs</i>	<ul style="list-style-type: none"> • <i>Significance test between one observed condition-specific sub-network and the corresponding randomly simulated networks.</i> <p>For a single motif type, we calculated the statistical significance of the difference in motif usage observed for a condition-specific sub-network compared with the expected usage in simulated networks of similar size (in a similar fashion to above).</p> <ol style="list-style-type: none"> We calculated an expected distribution of the occurrence of motif f ($N_m^{fe}(k)$) using the set of simulated networks of size m (in a similar fashion to above). We counted the number of these expected motif occurrences ($N_m^{fe}(k)$) that are greater than the observed occurrence (N_m^{fo}) in the corresponding condition-specific sub-network (<i>i.e.</i> $N_m^{fe}(k) > N_m^{fo}$).

<p><i>local network motifs</i></p>	<ul style="list-style-type: none"> iii. We calculated a p-value as the fraction of these expected occurrences that satisfied this requirement with respect to the total number of comparisons (<i>ie</i> 1,000 comparisons). iv. We rejected the null hypothesis H_0 (that the occurrence of a motif type in an observed condition-specific sub-network originates from the same population as that of the simulated networks) if $p \leq 0.01$ or $p \geq 0.99$.
------------------------------------	---

4.3.5 Comparisons between two observed condition-specific sub-networks with respect to expectations from random simulations

We used the following methods to determine whether the topological differences of two (or more) observed condition-specific sub-networks: (a) represent a real shift in network topologies between the endogenous and exogenous conditions, or (b) can be simply explained by the differences in sub-network sizes. We calculated if the observed differences between two condition-specific sub-networks are larger than the differences between the corresponding sets randomly simulated networks. The resulting statistics are discussed in §6.3.3 and we quote some of the major p -values in the main paper.

network statistic	significance test
global topological measures	<p>1. <i>Difference between two observed condition-specific networks.</i> For a single topological measure, we calculated the difference between a pair of observed condition-specific sub-networks.</p> <ul style="list-style-type: none"> i. We calculated the observed difference ($d_{mn}^{to} = \langle t_m^o \rangle - \langle t_n^o \rangle$) in mean values between a pair of condition-specific sub-networks m and n. $\langle t_m^o \rangle$ is the mean of the topological measure in an observed sub-network m. ii. We could have easily quantified the difference between sub-networks using a U-statistic (in a similar fashion to §4.3.3). However the difference d provides an intuitive measure that is computationally inexpensive. <p>2. <i>Difference between two sets of randomly simulated networks.</i> For a single topological measure, we calculated the difference between two sets of simulated networks that are of similar sizes as the pair of observed condition-specific sub-networks above.</p> <ul style="list-style-type: none"> i. We calculated the expected differences $d_{mn}^{te}(k)$ between matching pairs of simulated networks of size m and n. The pairs of networks are indicated by the subscript k. This provides an expected distribution of 1,000 differences d. ii. We normalised these expected differences by calculating a difference Z-score ($Z_{mn}^{te}(k) = \frac{d_{mn}^{te}(k) - \langle d_{mn}^{te} \rangle}{\sigma_{mn}^{te}}$), where $\langle d_{mn}^{te} \rangle$ and σ_{mn}^{te} are the mean and standard deviation of the expected distribution of d. This provides an expected distribution of 1,000 Z-scores (in a similar fashion to above). Note the Z-score normalises the differences d between every pair of network sizes and for all topological measures. The importance of doing this will become apparent below. iii. We also used the mean and standard deviation (of the expected distribution of difference d) to calculate a Z-score for the observed d between the corresponding pair of condition-specific sub-networks.

*global
topological
measures*

3. *Significance test between two observed condition-specific sub-networks for a single topological measure.*

For a single topological measure, we calculated the statistical significance of the difference in topologies between the two observed condition-specific sub-networks. We compared this with the expected difference between the corresponding simulated networks.

- i. We counted the number of these expected difference Z-scores that are greater than the observed Z-score for the pair of condition-specific sub-networks (*ie* $Z_{mn}^{te}(k) > Z_{mn}^{to}$).
- ii. We calculated a p -value as the fraction of these expected Z-scores that satisfied this requirement with respect to the total number of comparisons (*ie* 1,000 Z-score comparisons).
- iii. We rejected the null hypothesis H_0 (that the difference in the topological measure between the two observed condition-specific sub-networks is as expected from random simulations) if $p \leq 0.01$ or $p \geq 0.99$.

4. *Combined significance test between two observed condition-specific sub-networks across all topological measures.*

We calculated the combined statistical significance for the difference across all topological measures between two observed condition-specific sub-networks. We compared this with the expected difference between the corresponding simulated networks.

- i. We calculated a combined difference score S that sums the squares of Z-scores (from above) across every topological measure ($S_{mn}^{To} = \sum_t (Z_{mn}^{to})^2$).
- ii. Each topological measure has equal weighting as the difference Z-scores are normalised. Note that this step is similar in spirit to calculating a χ^2 -score.
- iii. Likewise, we calculated the expected S -score from matching pairs, k , of simulated networks (in a similar fashion to the observed score; *ie* $S_{mn}^{Te}(k) = \sum_t Z_{mn}^{te}(k)^2$). This provided an expected score distribution for S .
- iv. We counted the number of these expected S -scores that are greater than the observed S -score.
- v. We calculated a combined p -value as the fraction of expected S -scores that satisfied this requirement with respect to the total number of comparisons (*ie* 1,000 S -score comparisons).
- vi. We rejected the null hypothesis H_0 (that the difference of all topological measures between the two observed condition-specific sub-networks is as expected from random simulation) if $p \leq 0.01$ or $p \geq 0.99$.

<i>global topological measures</i>	<p>5. <i>Pooled significance test between observed endogenous and exogenous sub-networks for a single topological measure.</i></p> <p>For a single topological measure, we calculated the statistical significance of the difference in topologies between the observed endogenous and exogenous sub-networks. We compared this with the expected difference between the corresponding simulated networks.</p> <ul style="list-style-type: none"> i. We clustered the observed and simulated networks based on the pair-wise p-values of the condition-specific sub-networks from above. As before (in a similar fashion to §4.3.3), this grouped the two endogenous and three exogenous conditions. ii. We calculated a pooled difference score R between the observed endogenous and exogenous conditions by summing the squares of Z-scores across all pairs of condition-specific sub-networks (n and m), where n and m belong to separate endogenous and exogenous categories (<i>ie</i> $R^{to} = (Z_{cc,ds}^{to})^2 + (Z_{cc,dd}^{to})^2 + (Z_{cc,sr}^{to})^2 + (Z_{sp,ds}^{to})^2 + (Z_{sp,dd}^{to})^2 + (Z_{sp,sr}^{to})^2$). iii. Likewise, we calculated the expected pooled R-score from matching pairs of simulated networks (in a similar fashion to the observed score; <i>ie</i> $R^{te} = Z_{cc,ds}^{te}(k)^2 + Z_{cc,dd}^{te}(k)^2 \dots + Z_{sp,sr}^{te}(k)^2$). This provided an expected score distribution for R. iv. We counted the number of these expected R-scores that are greater than the observed R-score. v. We calculated a pooled p-value as the fraction of expected R-scores that satisfied this requirement with respect to the total number of comparisons (<i>ie</i> 1,000 R-score comparisons). vi. We rejected the null hypothesis H_0 (that the difference in the topological measure between the observed endogenous and exogenous sub-networks is as expected from random simulation) if $p \leq 0.01$ or $p \geq 0.99$. <p>6. <i>Overall significance test between observed endogenous and exogenous sub-networks across all topological measures.</i></p> <p>We calculated the overall statistical significance of the difference across all topological measures between the observed endogenous and exogenous sub-networks. We compared this with the expected difference between the corresponding simulated networks. The overall p-value provided a single statistic that measures the significance of the global topological difference between all the observed condition-specific sub-networks.</p> <ul style="list-style-type: none"> i. We clustered the observed and simulated networks into the endogenous and exogenous categories as above. ii. We calculated an overall difference score Q between the observed endogenous and exogenous conditions by summing R^{to} across all topological measures ($Q^{to} = \sum_t R^{to}$). This provided a single Q-score to measure the difference between the endogenous and exogenous sub-networks while taking into account all of the topological measures.
--	---

<p><i>global topological measures</i></p>	<ul style="list-style-type: none"> iii. Likewise, we calculated the expected overall Q-score from matching pairs of simulated networks (in a similar fashion to above; ie $Q^{Te} = \sum_t R^{te}$). This provided an expected score distribution for Q. iv. We counted the number of these expected Q-scores that are greater than the observed Q-score. v. We calculated an overall p-value as the fraction of expected Q-scores that satisfied this requirement with respect to the total number of comparisons (ie 1,000 Q-score comparisons). vi. We rejected the null hypothesis H_0 (that the difference in overall topology between the observed endogenous and exogenous sub-networks is as expected from random simulation) if $p \leq 0.01$ or $p \geq 0.99$.
<p><i>local network motifs</i></p>	<ol style="list-style-type: none"> 1. <i>Difference between two observed condition-specific sub-networks.</i> We calculated the difference in overall motif usage between a pair of observed condition-specific sub-networks. <ul style="list-style-type: none"> • We calculated a χ^2-score $[\chi^2]_{mn}^{Fo}$ to measure the difference in overall motif (F) occurrences between a pair of observed condition-specific sub-networks m and n (in a similar fashion to §4.3.3). 2. <i>Difference between two sets of randomly simulated networks.</i> Similarly, we calculated the difference in overall motif usage between two sets of randomly simulated networks that are of similar sizes as the pair of observed condition-specific sub-networks above. <ol style="list-style-type: none"> i. We calculated the expected χ^2-scores $[\chi^2]_{mn}^{Fe}(k)$ between matching simulated networks to provide an expected score distribution (in a similar fashion to above). ii. We normalised the expected χ^2-scores by calculating a difference Z-score ($Z_{mn}^{Fe}(k) = \frac{[\chi^2]_{mn}^{Fe}(k)^2 - <[\chi^2]_{mn}^{Fe}>^2}{\sigma_{mn}^{Fe}}$) where $<[\chi^2]_{mn}^{Fe}>$ and σ_{mn}^{Fe} are the mean and standard deviation of the expected χ^2-distribution (in a similar fashion to above). This provided an expected score distribution for Z. iii. Likewise, we calculated a difference Z-score Z_{mn}^{Fo} between the corresponding condition-specific sub-networks. 3. <i>Significance test between two observed condition-specific sub-networks across all motif occurrences.</i> We calculated the statistical significance of the difference in overall motif usage between the two observed condition-specific sub-networks. We compared this with the expected difference between the corresponding simulated networks. <ol style="list-style-type: none"> i. We counted the number of these expected Z-scores that are greater than the observed Z-score. ii. We calculated a p-value as the fraction of expected Z-scores that satisfied

<p><i>local network motifs</i></p>	<p>this requirement with respect to the total number of comparisons (<i>ie</i> 1,000 Z-score comparisons).</p> <p>iii. We rejected the null hypothesis H_0 (that the difference in overall motif usage between the two observed condition-specific sub-networks is as expected from random simulations) if $p \leq 0.01$ or $p \geq 0.99$.</p> <p>4. <i>Overall significance test between observed endogenous and exogenous sub-networks across all motif occurrences.</i></p> <p>We calculated the statistical significance of the difference in motif usage between the observed endogenous and exogenous sub-networks. We compared this with the expected difference between the corresponding simulated networks. The overall p-value provided a single statistic that measures the significance of the difference in overall motif usage between all the observed condition-specific sub-networks.</p> <p>i. We clustered the observed and simulated networks into the endogenous and exogenous categories using the pair-wise p-values above.</p> <p>ii. We calculated an overall difference score Q between the observed endogenous and exogenous conditions by summing the squares of Z-scores across all pairs of condition-specific sub-networks belonging to separate categories (<i>ie</i> $Q^{Fo} = (Z_{cc,ds}^{Fo})^2 + (Z_{cc,dd}^{Fo})^2 \dots + (Z_{sp,sr}^{Fo})^2$). Note here we do not need to calculate the S- and R- scores (as with topological measures) as the occurrence of different motif types are already combined when calculating the χ^2- and Z-scores.</p> <p>iii. Likewise, we calculated the expected Q-score from matching pairs of simulated networks to (in a similar fashion to the observed score; <i>ie</i> $Q^{Fe} = Z_{cc,ds}^{Fe}(k)^2 + Z_{cc,dd}^{Fe}(k)^2 \dots + Z_{sp,sr}^{Fe}(k)^2$). This provides an expected score distribution for Q.</p> <p>iv. We counted the number of these expected Q-scores that are greater than the observed Q-score.</p> <p>v. We calculated an overall p-value as the fraction of expected Q-scores that satisfied this requirement with respect to the total number of comparisons (<i>ie</i> 1,000 Q-score comparisons).</p> <p>vi. We rejected the null hypothesis H_0 (that the difference in overall motif usage between the observed endogenous and exogenous sub-networks is as expected from random simulation) if $p \leq 0.01$ or $p \geq 0.99$.</p>
------------------------------------	---

	<p>1. <i>Overlap in hub usage between two randomly simulated networks.</i> We compared the usage of TF hubs in the observed condition-specific sub-networks compared with the expected usage in randomly simulated networks.</p> <ul style="list-style-type: none"> i. We calculated the percentage overlap in TFs being classified as hubs between matching pairs of simulated networks of size m and n (in a similar fashion to above). ii. We compared the overlap between the observed condition-specific sub-networks and the mean expected overlap between simulated networks of corresponding sizes. <p>2. <i>Correlation of numbers of target genes between two randomly simulated networks.</i> We compared the correlation in numbers of target genes of TF hubs in the observed condition-specific sub-networks compared with the expected correlation in randomly simulated networks.</p> <ul style="list-style-type: none"> i. For TF hubs that overlap between matching pairs of simulated networks of size m and n, we calculated the Pearson correlation coefficients of the number of genes that they target (in a similar fashion to above). ii. We compared the level of correlation between observed condition-specific sub-networks and the average correlation between randomly simulated networks of corresponding sizes (in a similar fashion to above).
<i>hub usage</i>	

4.3.6 Sensitivity analysis

We test the robustness of the observations for the condition-specific sub-networks by conducting a sensitivity analysis.

- i. We generated 1,000 static networks containing error-rates of 10%, 20% and 30%. The errors we introduced include random addition, deletion and exchange of regulatory interactions.
- ii. For each condition, we traced through each erroneous network using the correct list differentially expressed genes and present TFs. This provides 1,000 sub-networks containing random errors for each condition.
- iii. We calculate the topological measures and motif occurrence for each error-containing sub-network.

We used the following methods to determine if there are topological differences between condition-specific sub-networks and error-containing sub-networks. The resulting statistics are discussed in §6.3.4.

network statistic	significance test
<i>global topological measures</i>	<ul style="list-style-type: none"> • <i>Significance test between one observed condition-specific sub-network and the corresponding simulated error-containing networks.</i> For a single topological measure, we calculated the statistical significance of the difference in the observed mean for a condition-specific sub-network compared with the expected (<i>ie</i> random) means in the error-containing networks of similar size. v. We calculated an expected distribution of the mean topological measure ($\langle t_m^e(k) \rangle$) using the set of simulated networks of size m. vi. We counted the number of these expected means ($\langle t_m^e(k) \rangle$) that are greater than the observed mean $\langle t_m^o \rangle$ in the corresponding condition-specific sub-network (<i>ie</i> $\langle t_m^e(k) \rangle > \langle t_m^o \rangle$). vii. We calculated a p-value as the fraction of these expected means that satisfied this requirement with respect to the total number of comparisons (<i>ie</i> 1,000 comparisons). viii. We rejected the null hypothesis H_0 (that the topological measure in an observed condition-specific sub-network originates from the same population as that of the simulated error-containing networks) if $p \leq 0.01$ or $p \geq 0.99$.
<i>local network motifs</i>	<ul style="list-style-type: none"> • <i>Significance test between one observed condition-specific sub-network and the corresponding simulated error-containing networks.</i> For a single motif type, we calculated the statistical significance of the difference in motif usage observed for a condition-specific sub-network

compared with the expected usage in the error-containing networks of similar size (in a similar fashion to above).

- i. We calculated an expected distribution of the occurrence of motif f ($N_m^{fe}(k)$) using the set of simulated networks of size m (in a similar fashion to above).
- ii. We counted the number of these expected motif occurrences ($N_m^{fe}(k)$) that are greater than the observed occurrence (N_m^{fo}) in the corresponding condition-specific sub-network (*ie* $N_m^{fe}(k) > N_m^{fo}$).
- iii. We calculated a p -value as the fraction of these expected occurrences that satisfied this requirement with respect to the total number of comparisons (*ie* 1,000 comparisons).
- iv. We rejected the null hypothesis H_0 (that the occurrence of a motif type in an observed condition-specific sub-network originates from the same population as that of the simulated error-containing networks) if $p \leq 0.01$ or $p \geq 0.99$.

4.4 Availability

Perl scripts implementing SANDY, data files containing the randomly simulated networks and their topological statistics are available from the Supplementary Website.

5. Identifying active sub-networks of endogenous condition times-courses

We examined the dynamics of the regulatory network during the time-course of the endogenous conditions. We identified the active regulatory sub-networks for each phase of the cell cycle and sporulation using a similar procedure as described in §3.

5.1 Defining differentially expressed genes during endogenous condition phases

The studies of Cho *et al.* (6) and Chu *et al.* (7) measured gene expression during successive phases of the cell cycle and sporulation through a time-course. The two studies classified differentially expressed genes into the phases during which they displayed peak expression. We used these data to examine the dynamics of the regulatory network during the time-course of these conditions.

The datasets contained the following numbers of differentially expressed genes during each phase.

condition	phase	number of differentially expressed genes	reference
cell cycle	early G1	75	(6)
	late G1	143	
	S	97	
	G2	69	
	M	105	
<hr/>			
sporulation	metabolic	52	(7)
	early I	61	
	early II	45	
	early-mid	95	
	middle	158	
	mid-late	61	
	late	5	

5.2 Identifying present TFs

We used the same list of present TFs for the cell cycle and sporulation as described in §3.3.

5.3 Tracing phase-specific sub-networks

We defined the phase-specific sub-networks for the cell cycle and sporulation by applying the TBA to each set of phase-specific differentially expressed genes.

5.4 Clustering time-dependent TF usage

We clustered TFs that are actively used during the time-courses with a similar procedure we used to cluster TF hubs (§4.2.1); we calculated propensity values for all active TFs in the cell cycle and sporulation, and then clustered TFs using $k = 6$ for the cell cycle (five phase-specific and one ubiquitous cluster), and $k = 7$ for sporulation (six phase-specific and one ubiquitous).

5.4 Availability

Data files containing lists of differentially-expressed genes, the phase-specific sub-networks and TF clusters are available from the Supplementary Website.

6. SANDY reveals large topological changes

6.1 Well-known statistics

6.1.1 Global topological measures

Below we list all of the global topological measures that we calculated for the static and condition-specific networks. In addition to those published in the main paper, we show three additional measures that were excluded:

- i. In-degree exponential distribution exponent (β).
- ii. Out-degree power-law distribution exponent (γ).
- iii. Network diameter (d).

topological measure	cellular conditions					
	static	endogenous		exogenous		
		cell cycle	sporulation	diauxic shift	DNA damage	stress response
# TFs	142	70	74	71	72	63
# target genes	3,420	280	257	748	678	362
# interactions	7,074	550	481	1,217	1,082	566
in-degree ($< k_{in} >$)	2.1	2.0	1.9	1.6	1.6	1.6
in-degree exponential exponent (β)	0.8	0.7	0.8	1.2	1.2	1.2
out-degree ($< k_{out} >$)	49.8	7.9	6.5	15.0	15.0	9.0
out-degree power-law exponent (γ)	0.6	1.5	1.5	0.8	0.8	0.9
path length (l)	4.7	4.5	3.4	2.0	2.0	2.2
diameter (d)	12	12	10	6	6	7
Clustering coefficient (c)	0.11	0.15	0.14	0.09	0.09	0.08

As discussed in §6, the p -values resulting from comparisons with randomly simulated networks indicate that: (i) endogenous and exogenous sub-networks display statistically significant differences in topological measures, and (ii) individual sub-networks have topologies that diverge significantly from random expectation.

Descriptions and interpretations of the major topological measures are provided in the main paper. Below we describe the observations for the three measures that were excluded.

topological measure	observation	interpretation
<i>in-degree exponential distribution exponent (β)</i>	<p>The distribution of incoming interactions per target gene is characterised by an exponential behaviour (probability for a given gene to be regulated by k TFs decreases proportionally to $e^{-\beta k}$); the exponent, β, is 0.8 in the static network.</p> <p>Though exponential behaviour is maintained in all conditions, exponents are larger in the exogenous conditions signifying a faster drop-off.</p>	<p>The exponential behaviour indicates a sharp decay in the in-degree distribution, and presumably reflects the molecular constraints on the number of TFs that can co-regulate at the same promoter (16).</p> <p>Along with the average in-degrees ($\langle k_{in} \rangle$), the changes in exponents suggest that TF combination usage is simpler in the exogenous conditions, which reflects the more direct-acting nature of these cellular states.</p>
<i>out-degree power-law distribution exponent (γ)</i>	<p>The distribution of outgoing interactions follows a power-law behaviour (probability that a given TF regulates k genes decreases proportionally to $k^{-\gamma}$); the exponent γ is 0.6 in the static network.</p> <p>Power-law behaviour is maintained in all conditions; however exponents double from 0.8 to 1.5 between the exogenous and endogenous sub-networks.</p>	<p>The power-law behaviour signifies a broader decay profile than the in-degree distribution and it is indicative of a hub-containing network structure. The exponent is less than what is observed for other molecular biological networks, signifying the distribution is not as polarized, or “hubby”.</p> <p>The changes in exponents between conditions suggest that the exogenous sub-networks contain fewer TF hubs.</p>
<i>diameter (d)</i>	<p>The diameter is 12 in the static network, showing that the longest path has 12 intermediate TFs. Diameters double from the exogenous ($d = 6-7$) to endogenous conditions ($d = 10-12$).</p>	<p>As with path lengths (l), the diameter measures the distance between a TF and its final target; it gauges the immediacy of a regulatory signal. Shorter distances in the exogenous sub-networks suggest that external stimuli propagate to the required targets quickly. Longer distances in endogenous conditions are due to formation of regulatory cascades.</p>

6.1.2 Local network motifs

Below we provide the numbers of regulatory interactions participating in network motifs (SIMs, MIMs, FFLs). Percentages are calculated as a fraction of the total number of interactions in motifs for each condition.

network motif	cellular conditions					
	static	endogenous		exogenous		
		cell cycle	sporulation	diauxic shift	DNA damage	stress response
SIMs	1,748 (37.6%)	130 (32.0%)	117 (38.9%)	438 (57.4%)	462 (55.7%)	228 (59.1%)
MIMs	325 (7.0%)	96 (23.7%)	50 (16.6%)	180 (23.6%)	226 (27.3%)	78 (20.2%)
FFLs	2,581 (55.5%)	180 (44.3%)	134 (44.5%)	145 (19.0%)	141 (17.0%)	80 (20.7%)
Total	4,654	406	301	763	829	386

As discussed in the main paper, there are large differences in motif occurrences between the endogenous and exogenous conditions. As discussed in §6, the *p*-values resulting from comparisons with randomly simulated networks indicate that: (i) endogenous and exogenous sub-networks display statistically significant differences in motif usage, and (ii) individual sub-networks have motif occurrences that diverge significantly from random expectation.

Studies have ascribed particular information processing tasks to motifs (14). Here we describe two example motifs found in the conditions-specific sub-networks.

network motif	description	example
<i>SIM</i> <i>MIM</i>	Simultaneous regulation of multiple genes such as those involved in the same pathway or macromolecular complex. They appear suited for controlling large-scale turnover of genes observed in exogenous conditions.	DNA damage. Rpn2 regulates three proteosomal subunits Rpt2, Rpt4, and Rpt6.
<i>FFL</i>	Regulatory buffer that respond only to persistent input signals from the primary TF, and allows for rapid shutdown when signal ceases. It appears suited for endogenous conditions as cells will only enter a new phase once the regulatory signal from the previous one has stabilised. The signal will also terminate quickly once the cell has entered a new phase.	Sporulation. Rim1 acts as the primary and Ime1 as the secondary TF to regulate Ime2 in the early phase. Ime2 is a kinase that stimulates about 20 further TFs in the middle and late phases; it ensures a quick shutdown of the regulatory cascade through phosphorylation of Ime1.

6.2 Follow-on statistics

6.2.1 TF hub usage

The power-law behaviour of the out-going degree distribution indicates the presence of TF hubs that target a disproportionately large number of genes. We identify a total of 51 TF hubs that separate into two main groups that we describe below:

hub type	observation	example
<i>permanent</i>	<p>There are 11 hubs that are permanent features of the regulatory network regardless of cellular condition. These mainly comprise TFs that regulate house-keeping functions or multiple functions.</p>	<p>Multifunctional regulator. Abf1 is a general transcriptional activator; it has 291 target genes in the static network and regulates an average on 55.2 genes across all condition-specific sub-networks.</p> <p>House-keeping regulator. Mig1 and Mig2 are TFs involved regulation of glucose metabolism.</p>
<i>transient</i>	<p>There are 40 transient hubs, which are influential in one cellular condition, but much less so in others. The hubs group in five sub-clusters representing each cellular condition.</p> <p>TF hubs with direct functional associations.</p> <p>Half of the hubs are important for the particular condition; it is harder to make direct functional associations for the remaining TFs as they have less complete annotations.</p>	<p>TF hubs with direct functional associations.</p> <p>Cell cycle. 10 out of 11 hubs are known cell cycle TFs, including the Swi4 and Mbp1 G1/S regulators (17-19).</p> <p>Sporulation. Includes Ime1, a key inducer of early meiotic genes (20-22), and Ume6 a co-activator (21-24). Ndt80, an important regulator of the middle stages of sporulation (21,22,25) is absent as it currently has only one assigned target gene in the dataset.</p> <p><i>Diauxic shift</i> Includes Hap2 and Hap4 global</p>

		regulators of respiratory gene expression and activator of cytochrome C (26-28).
<i>transient</i>	<p>TF hubs with unclear functional associations.</p> <p>Of great interest are TFs that are very sparsely annotated. As hubs, they are obviously important in the cellular state under consideration. We can augment these annotations by predicting key regulatory roles in their respective conditions, and this should provide a useful starting point for further experimental characterization. Such functional predictions are not trivial to do, and it is only through integrating gene expression data with the regulatory network that we are able to do this.</p>	<p>Stress response. Includes the Msn2 and Msn4 stress response regulators.</p> <p>TF hubs with unclear functional associations.</p> <p>Sporulation. There are three regulators of nitrogen utilization (YIR023W, YPL038W, YNL103W); these appear to be surprising inclusions, but as sporulation is initiated by nitrogen depletion, their inclusion is biologically meaningful.</p> <p>There are two poorly annotated TF hubs (YMR021C, YIL113W).</p> <p>Diauxic shift. Poorly annotated TF hubs (YHR206W).</p> <p>Stress response. Poorly annotated TF hubs (YDR259C, YDR501W, YGL096W, YLR403W, YIR018W).</p> <p>Permanent hubs. Poorly annotated hubs (YKL043W, YLR013W).</p>

6.2.2 Interaction interchange index

The interchange index (I) addresses the dynamic nature of the regulatory interactions between TFs and target genes. It quantifies the extent to which each TF contributes to the rewiring of the regulatory network.

i. Hot links

Just 66 interactions are retained across four or more conditions and we consider these to represent hot links (29,30) that are “always on” compared with the rest of the network. Many interactions originate from two types of TFs: metabolic regulators (eg Mig1 and Mig2), and general transcriptional regulators (eg Abf1 and Reb1). Therefore we associate hot links with the continual regulation of house-keeping functions in the cell. Many of the TFs making these links also comprise permanent hubs.

ii. Interchange index values

1,476 out of a total 2,479 regulatory interactions are unique to a particular condition, and the remainder is common to two or more. Thus over half of the interactions are exchanged completely between conditions, resulting in specific regulation of the respective cellular states.

Indices calculated for all TFs display a uni-modal distribution with two extreme outliers:

interchange index (I)	# TFs	example TF
$I \leq 10\%$	12	Rgt1
$10\% < I < 90\%$	73	Sin3, Abf1, Yox1
$I \geq 90\%$	27	Swi4

iii. Example TFs

Below we provide descriptions of example TFs that have different index values. These examples are depicted in the main paper. An interesting observation is that TFs exchanging regulatory interactions between conditions also shift their regulatory role.

example TF	interchange index (I)	description
Swi4	100%	The TF makes a total of 43 regulatory interactions. It has an extremely high interchange index; in fact it only targets genes during the cell cycle and none during other conditions. It is a regulator of the G1/S transition and predominantly targets genes involved in DNA synthesis and cell wall synthesis.

		<p>The TF makes a total of 120 regulatory interactions. It has a fairly high index and shows little overlap of interactions across cellular conditions.</p> <p>The TF has so far been implicated in control of the cell cycle, and DNA synthesis and repair, but information about its full range of regulatory functions is currently limited (31,32).</p>
<i>Yox1</i>	84%	<p>Consistent with its role in DNA synthesis, it makes 21 regulatory interactions during the cell cycle and 30 during sporulation. Surprisingly, it produces the most interactions – 91 – during diauxic shift, which suggests a previously unreported role in this process. The interchange of regulatory interactions brings about a dramatic shift in the function of Yox1; it is focused on controlling cell growth during the cell cycle and sporulation, whereas it redirects its attention to protein synthesis during diauxic shift.</p>
<i>Abf1</i>	51%	<p>The TF makes a total of 160 regulatory interactions. It has an intermediate exchange index and among its many cited functions as general transcriptional activator are the regulation of meiosis (33), metabolic activities (34), translation (35,36) and gene silencing (37)</p> <p>It preserves 79 interactions across multiple conditions, which corresponds to the consistent regulation of a core set of cellular functions such as glycolytic pathways, translation and cell biogenesis.</p> <p>The other half of interactions – 81 – is exchanged, and this results in a shift of regulatory focus on top of its core functions. In the cell cycle and sporulation Abf1 regulates cell growth, whereas in stress response it controls intracellular transport.</p>
<i>Sin3</i>	25%	<p>The TF makes a total of 20 regulatory interactions. It has a fairly low index indicating a high overlap in regulatory interactions between cellular conditions. It is a silencing component of the Rpd3-containing histone deacetylase complex (38), and many of its target genes are involved in metabolic processes and cell growth. The few condition-specific interactions are made during sporulation or DNA damage.</p>
<i>Rgt1</i>	0%	<p>The TF makes a total of six regulatory interactions. It has an extremely low index, and maintains all of its interactions across multiple cellular conditions. It is a house-keeping regulator controlling glucose metabolism and glucose import.</p>

6.2.3 TF usage across multiple conditions

We discuss that there is large overlap in individual TF usage but small overlap in TF pair usage; therefore regulatory specificity is achieved through combinatorial TF use (39).

i. Individual TF usage

There is large overlap in the use of individual TFs. 95 TFs are used in more than one condition. Between the endogenous conditions 53 out of 92 TFs overlap and between exogenous conditions 44 out of 100 TFs are used in all three states.

Below we show the overlap in individual TF usage between all the cellular conditions.

		cellular conditions				
cellular conditions		<i>cell cycle</i>	<i>sporulation</i>	<i>diauxic shift</i>	<i>DNA damage</i>	<i>stress response</i>
	<i>cell cycle</i>	70	53	52	56	37
	<i>sporulation</i>		75	62	62	45
	<i>diauxic shift</i>			79	70	49
	<i>DNA damage</i>				81	45
	<i>stress response</i>					63

ii. Pair-wise TF usage

There is small overlap in the use of pair-wise TF combinations between conditions. The static network has an average in-degree of 2.1, meaning that genes are typically regulated by pairs of TFs. In-degrees are larger in the endogenous sub-networks, combinatorial TF regulation is more prevalent in these conditions.

There are 360 distinct pair-wise TF combinations across the five conditions; of these 309 are unique to a single condition. Between endogenous conditions, just 3 out of 149 pairs overlap, and between exogenous conditions just 3 out of 233 pairs overlap.

Abf1 provides a nice example of combinatorial TF specificity. During sporulation, it combines with Ime1 to regulate Hop1, which is involved in chromosomal segregation. During diauxic shift, it acts in conjunction with the Hap2-Hap4 heteromeric complex to regulate the Aac2 major mitochondrial ADP/ATP carrier.

Below we show the overlap in pair-wise TF usage between all the cellular conditions.

		cellular conditions				
cellular conditions		<i>cell cycle</i>	<i>sporulation</i>	<i>diauxic shift</i>	<i>DNA damage</i>	<i>stress response</i>
	<i>cell cycle</i>	82	3	2	3	0
	<i>sporulation</i>		70	6	12	1
	<i>diauxic shift</i>			127	28	3
	<i>DNA damage</i>				125	3
	<i>stress response</i>					12

6.3 Statistical validation with randomly simulated networks

6.3.1 Comparison between two observed condition-specific sub-networks

We provide a full list of p -values obtained from this analysis on the Supplementary Website. See §4.3.3 for precise definitions and calculations of p -values.

i. Global topological measures and local network motifs

Briefly, the p -values indicate that the distribution of topological measures and occurrence of network motifs are similar within the endogenous and exogenous categories (*ie* differences are non-significant), whereas they are different between the two types of conditions (*ie* differences are significant).

We favour the p -values discussed in §6.3.3 as they consider the dependence of topology on network sizes, and so allow us to identify the topological changes that really differentiate between the endogenous and exogenous conditions.

ii. Hub usage

Below we show the percentage overlap in TFs that act as hubs in different cellular conditions.

cellular conditions						
cellular conditions		<i>cell cycle</i>	<i>sporulation</i>	<i>diauxic shift</i>	<i>DNA damage</i>	<i>stress response</i>
	<i>cell cycle</i>		61.9%	47.6%	42.9%	38.1%
	<i>sporulation</i>			45.5%	50.0%	36.4%
	<i>diauxic shift</i>				66.7%	66.7%
	<i>DNA damage</i>					61.9%
	<i>stress response</i>					

Below we show the Pearson correlation coefficients for the numbers of genes targeted by the overlapping TF hubs.

cellular conditions						
cellular conditions		<i>cell cycle</i>	<i>sporulation</i>	<i>diauxic shift</i>	<i>DNA damage</i>	<i>stress response</i>
	<i>cell cycle</i>		0.58	0.30	0.33	0.22
	<i>sporulation</i>			0.59	0.60	0.47
	<i>diauxic shift</i>				0.62	0.72
	<i>DNA damage</i>					0.74
	<i>stress response</i>					

6.3.2 Comparison between one observed condition-specific sub-network and the randomly simulated networks

We provide a full list of p -values obtained from this analysis on the Supplementary Website. See §4.3.4 for precise definitions and calculations of p -values.

- **Global topological measures and local network motifs**

Briefly, the p -values indicate that many of the topological measures and motif occurrences are significantly different to that expected from the randomly simulated networks. The most prominent differences are apparent for the features that stand out for the endogenous and exogenous sub-networks. For example, clustering coefficients are high compared with the simulated networks for cell cycle ($p < 0.004$) and sporulation ($p < 0.011$).

6.3.3 Comparisons between two observed condition-specific sub-networks with respect to expectations from random simulations

The p -values listed here indicate the significance of the difference in the topologies and motif usage between distinct condition-specific sub-networks. The statistics are calculated taking into account the dependence of topology on network size, and therefore help identify the topological changes that really differentiate between the endogenous and exogenous conditions.

We provide a full list of p -values obtained from this analysis on the Supplementary Website. See §4.3.5 for precise definitions and calculations.

i. Global topological measures

Below we show the overall^{*} and pooled⁺ p -values indicating the significance of the difference in global topologies between endogenous and exogenous sub-networks across all^{*} (*ie* combined) and individual⁺ topological measures.

exogenous conditions					
	<i>all</i>	<i>in-degree</i>	<i>out-degree</i>	<i>path length</i>	<i>clustering coefficient</i>
endogenous conditions	<0.001 [*]	<0.005 ⁺	<0.338 ⁺	<0.001 ⁺	<0.001 ⁺

The overall p -value for all topological measures demonstrates that the global topologies of endogenous and exogenous conditions differ to a greater extent than expected from the random simulations. Therefore, the topological changes represent a real shift in sub-network structures between the endogenous and exogenous conditions, and are not only due to differences in sub-network sizes.

The combined p -values for the individual topological measures are also significant, indicating that they change more than expected from the random simulations. The only exception is the out-degree whose p -value is non-significant. This is because as the average out-degree depends greatly on the size of the simulated network (where mean values increase with network size). This is not surprising as a relatively small pool of TFs (which are the only nodes with outgoing connectivities) must be shared among a large number of target genes.

Below we show the combined p -values indicating the significance of the difference in topologies between pairs of individual conditions. P -values calculated for the overall changes across all topological measures (*ie* k_{in} , k_{out} , l , c).

cellular conditions						
		<i>endogenous</i>		<i>exogenous</i>		
cellular conditions		<i>cell cycle</i>	<i>sporulation</i>	<i>diauxic shift</i>	<i>DNA damage</i>	<i>stress response</i>
	<i>cell cycle</i>		<0.170	<0.001	<0.001	<0.001
	<i>sporulation</i>			<0.004	<0.027	<0.021
	<i>diauxic shift</i>				<0.391	<0.442
	<i>DNA damage</i>					<0.559
	<i>stress response</i>					

The overall global topologies of sub-networks within the endogenous (cell cycle, sporulation) and exogenous (diauxic shift, DNA damage, stress response) categories of conditions are similar.

In contrast the overall topologies between the endogenous and exogenous sub-networks are significantly different. There are a few individual exceptions where p -values indicate a non-significant difference (*eg* sporulation and DNA damage). However the overall and pooled p -values shown above clearly emphasises that there are topological differences between the endogenous and exogenous sub-networks. (See also Supplementary Website for more p -values).

ii. Local network motifs

Below we show the overall p -value indicating the significance of the difference in motif usage between endogenous and exogenous sub-networks. P -values are calculated for the overall occurrence all motif types (*ie* MIMs, SIMs, FFLs).

	exogenous conditions
endogenous conditions	<0.001

The p -value clearly shows that the occurrences of network motifs in the endogenous and exogenous conditions are significantly different. This difference is greater than the expectations calculated from the randomly simulated networks. Therefore, the changes in motif usage represent a real shift in sub-network structures between the endogenous and exogenous conditions, and are not only due to differences in sub-network sizes.

Below we show the p -value indicated the significance of the difference in overall motif occurrences between pairs of individual conditions.

cellular conditions						
		endogenous		exogenous		
cellular conditions		<i>cell cycle</i>	<i>sporulation</i>	<i>diauxic shift</i>	<i>DNA damage</i>	<i>stress response</i>
	<i>cell cycle</i>		<0.485	<0.013	<0.001	<0.001
	<i>sporulation</i>			<0.032	<0.001	<0.016
	<i>diauxic shift</i>				<0.691	<0.955
	<i>DNA damage</i>					<0.740
	<i>stress response</i>					

The overall motif occurrences of sub-networks within the endogenous (cell cycle, sporulation) and exogenous (diauxic shift, DNA damage, stress response) categories of conditions are similar.

In contrast motif usages between endogenous and exogenous conditions are significantly different. There are a few individual exceptions where p -values indicate a non-significant difference (eg sporulation and diauxic shift). However the overall and pooled p -values shown above clearly emphasises that there are topological differences between the endogenous and exogenous sub-networks. (See also Supplementary Website for more p -values).

iii. Hub usage

Below we show the mean percentage overlap in TFs that are defined as hubs in the simulated networks.

cellular conditions						
cellular conditions		<i>cell cycle</i>	<i>sporulation</i>	<i>diauxic shift</i>	<i>DNA damage</i>	<i>stress response</i>
	<i>cell cycle</i>		87.2%	78.7%	78.7%	95.7%
	<i>sporulation</i>			89.4%	89.4%	83.0%
	<i>diauxic shift</i>				100%	76.6%
	<i>DNA damage</i>					76.6%
	<i>stress response</i>					

Below we show the mean Pearson correlation coefficients for the numbers of genes targeted by overlapping TF hubs in the simulated networks.

cellular conditions						
cellular conditions		<i>cell cycle</i>	<i>sporulation</i>	<i>diauxic shift</i>	<i>DNA damage</i>	<i>stress response</i>
	<i>cell cycle</i>		0.97	0.95	0.95	0.99
	<i>sporulation</i>			0.99	0.99	0.93
	<i>diauxic shift</i>				1.0	0.91
	<i>DNA damage</i>					0.91
	<i>stress response</i>					

The percentage overlap and correlation coefficients are much higher for the randomly simulated networks than for the corresponding condition-specific sub-networks. Therefore randomly simulated networks have more uniform hub usage compared with what is observed in the condition-specific sub-networks. This indicates that random expectation is for different size networks to converge on the same set of TF hubs regardless of the regions of the network being used.

6.3.4 Sensitivity analysis

We provide the *p*-values obtained from this analysis on the Supplementary Website. See §4.3.6 for precise definitions and calculations of *p*-values.

- **Global topological measures and local network motifs**

Briefly, the *p*-values indicate that many of observed topological measurements and motif occurrences are maintained even when random errors are introduced into the data. All of the trends that we observe in the comparisons between condition-specific sub-networks remain at all error rates (10%, 20%, 30%). Therefore, the results we report are robust against substantial errors in the underlying data.

At 10% and 20% error rates, most changes in topologies are non-significant. At 30% error rates, some observations are affected to a significant level; the most affected include the mean clustering coefficients, and the occurrence of MIMs and FFLs for each sub-network. This is perhaps unsurprising as these particular observations depend on precise or tight interactions between nodes; the simulations introduce errors by dispersing these interactions so lowering the level of clustering and occurrence of MIMs and FFLs. Nevertheless, we continue to observe the same trends when we compare their values and occurrences across different conditions.

6.4 Availability

Data generated by SANDY are available from the Supplementary Website. These include the global topological measures and local motif occurrence in condition-specific and simulated networks, TF hub propensity values and clustering, interchange indices for all TFs, TF usage overlap between conditions, and full list of *p*-values from the statistical analyses.

7. Network dynamics during the time-courses of endogenous conditions

7.1 Phase-specific and ubiquitous TFs

TFs used to regulate the cell cycle and sporulation group into two main clusters depending on their time-dependent activity:

i. Phase-specific TFs

Most TFs operate during a particular phase, which is highlighted by their phase-specific targeting of genes.

cellular condition	# TFs	description
<i>cell cycle</i>	54	Activity of major cell cycle regulators is in line with previous observations, so emphasizing the validity of the methods we use in the paper (40). TFs group into five distinct sub-clusters representing some of the major phases during the cell cycle. Swi4 and Mbp1 are clustered in the late G1 phase (19), Fkh1 is found in G2 (41-44), Mcm1 is in M (41-44), and Ace2 and Swi5 are in the early G1 phase (45,46).
<i>sporulation</i>	51	TFs group into six distinct sub-clusters that correspond to some of the phases during sporulation. Major meiotic regulators are shown to be active in previously reported phases. Rim1, Ime1 and Ume6 are clustered in the early phases and Ndt80 in the middle phase.

ii. Ubiquitous TFs

A sizeable minority of TFs is ubiquitously active throughout the cell cycle. These TFs regulate genes indiscriminately of the cellular phase. About a third of ubiquitous TFs comprise permanent hubs.

cellular condition	# TFs	description
<i>cell cycle</i>	16	TFs form a single cluster and there are no cell cycle-specific TFs present in this group. About a third of TFs comprises permanent hubs and includes the Abf1 general transcriptional regulator.
<i>sporulation</i>	23	TFs form a single cluster and there are no sporulation-specific TFs present in this group. Again about a third of TFs consists of permanent hubs and includes the Abf1 general transcriptional regulator.

7.2 Serial and parallel inter-regulation between TFs

Of interest is the pattern of inter-regulation between TFs to control the temporal progression of the cell cycle and sporulation. We uncover two types of inter-regulation:

i. Serial inter-regulation

Phase-specific TFs in one phase of the condition regulate further TFs in subsequent phases. This serves to drive the condition forward (47).

cellular condition	description
<i>cell cycle</i>	<p>TFs in the G1 and S phases target TFs in G2 and M. These in turn, target TFs in the G1 phase in preparation for the next cycle.</p> <p>We can identify complete loops of regulatory interactions among the complex circuitry. Swi4 and Mbp1 in late G1 target TFs in the G2 and M phases. Fkh1 and Mcm1 in these phases in turn regulate Swi5 and Ace2 in early G1 of the next cell cycle. Finally Mcm1 loops back to regulate Swi4, the original TF.</p> <p>There is also limited back-regulation of TFs in the previous phase, such as that of Swi4 targeting Hap1. Presumably, these comprise inhibitory interactions turning off the activity of TFs, but we cannot be certain of this without detailed knowledge of the regulatory signal (<i>ie</i> activating or repressing) of the TF-target relationships. Unfortunately this information is rarely available currently.</p>
<i>sporulation</i>	<p>TFs in the metabolic and early phases target those in the middle to late phases. TFs in the later phases on the other hand, appear to back-regulate those in the earlier phases. As sporulation is not a cyclic cellular condition, we anticipate that these latter interactions are repressive and shutdown the early-phase TFs.</p> <p>The metabolic phase was previously identified as a stage during which many metabolic processes are repressed in preparation for sporulation (7). The phase appears to provide much of the forward regulation between TFs also. For example, Mig2, a glucose repressor targets two TFs in the mid-late phase. Rap1 is a multi-functional regulator implicated in meiotic facilitation (48,49), and targets many TFs in the early to late phases.</p> <p>There is only limited TF inter-regulation involving the major meiotic TFs, Ime1, Ume6 and Ndt80; only Rpn1 from the early phase forms a FFL motif with Ime1. The main reason for this is that Ime1 and Ume6 target Ime2, a major meiotic kinase, which in turn regulates further TFs in the middle to late phases of the condition.</p>

ii. Parallel inter-regulation

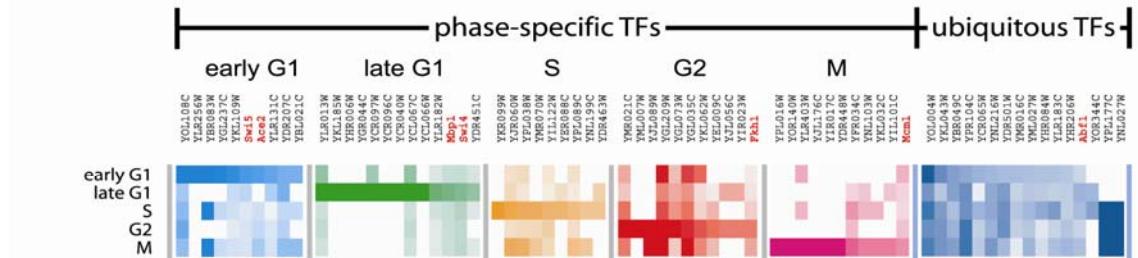
Ubiquitously active TFs inter-regulate with phase-specific TFs in a two-tier hierarchy. This method of regulation allows ubiquitously active TFs to be involved in the regulation of multiple cellular functions and aid the smooth transition between different phases. Interestingly, as many TFs are permanent hubs, parallel inter-regulation may provide a channel of communication between progression of the cellular conditions and control of the house-keeping processes.

cellular condition	description
<i>cell cycle</i>	<p>In parallel inter-regulation, ubiquitously active TFs inter-regulate with the phase-specific TFs from all five stages of the cell cycle. Most of the regulation is from the ubiquitous to phase-specific TFs, but there is also some reciprocal regulation.</p> <p><i>Abf1</i> is an example of a ubiquitously active TF, which targets five TFs in the early G1 to M phases. As described elsewhere, regulatory specificity is achieved through the combinatorial use of regulatory partners. In early G1 it combines with <i>Mac1</i>, <i>Sin3</i> and <i>Rox1</i> to regulate <i>Ume6</i> (acting as a mitotic repressor (50)). In the M phase it acts alone to regulate <i>Pho2</i>, a co-regulator of <i>Swi5</i> for homothallic switching (51).</p>
<i>sporulation</i>	<p>Again, ubiquitously active TFs inter-regulate with the phase-specific TFs in all stages of sporulation. Similarly to the cell cycle, most of the regulation is from the ubiquitous to phase-specific TFs, although there is some reciprocal regulation.</p> <p><i>Abf1</i> is again used as a ubiquitous TF, and targets four TFs in the early to middle stages. Regulatory specificity is again achieved by combinatorial use of regulatory partners. During early phase, it combines with <i>Sin3</i> to regulate <i>Ume6</i>, and with <i>Hap4</i> to target <i>HmlAlpha2</i>.</p>

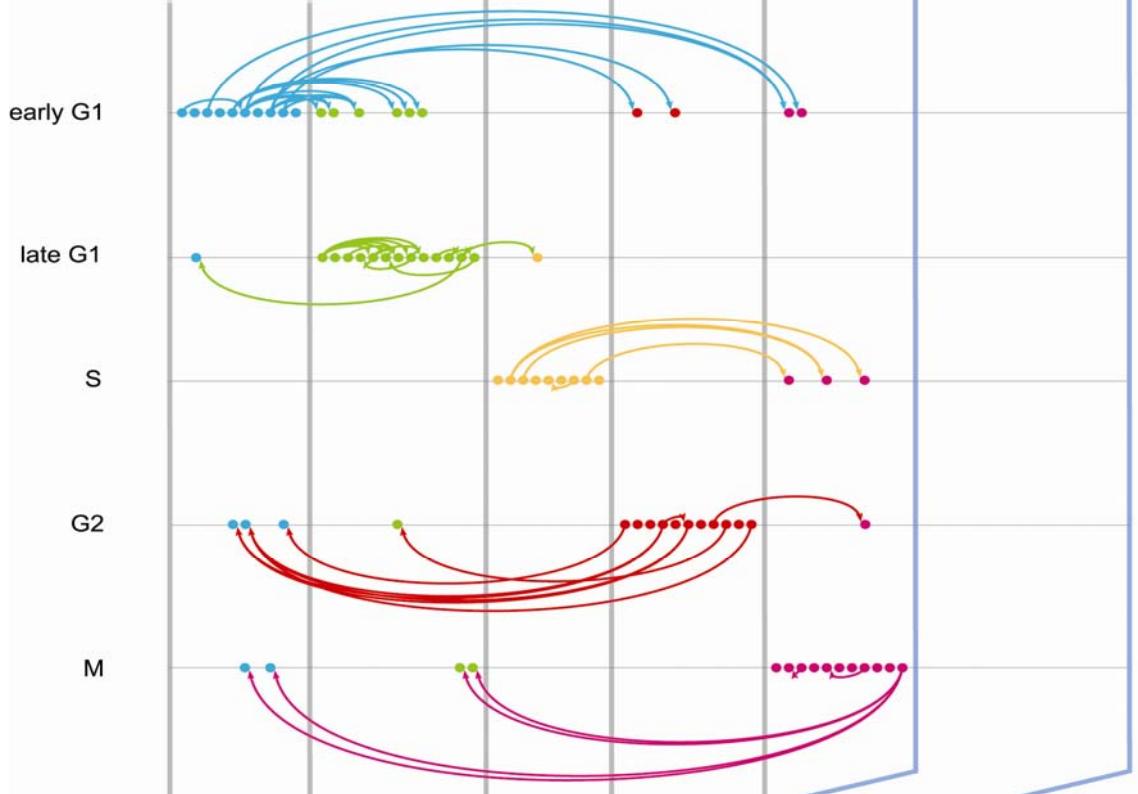
7.3 Network diagram of endogenous condition time-courses

Below we show a schematic of TF inter-regulation during the cell cycle time-course.

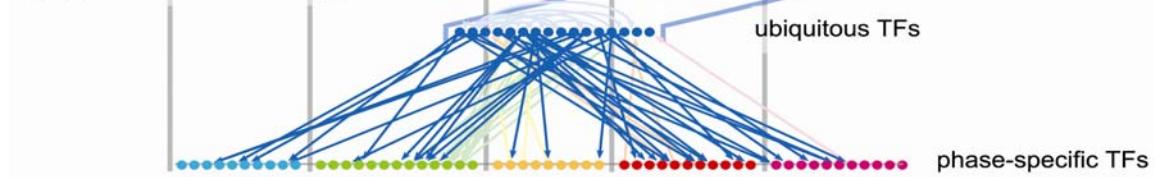
(a) phase-specific and ubiquitous transcription factors



(b) serial inter-regulation

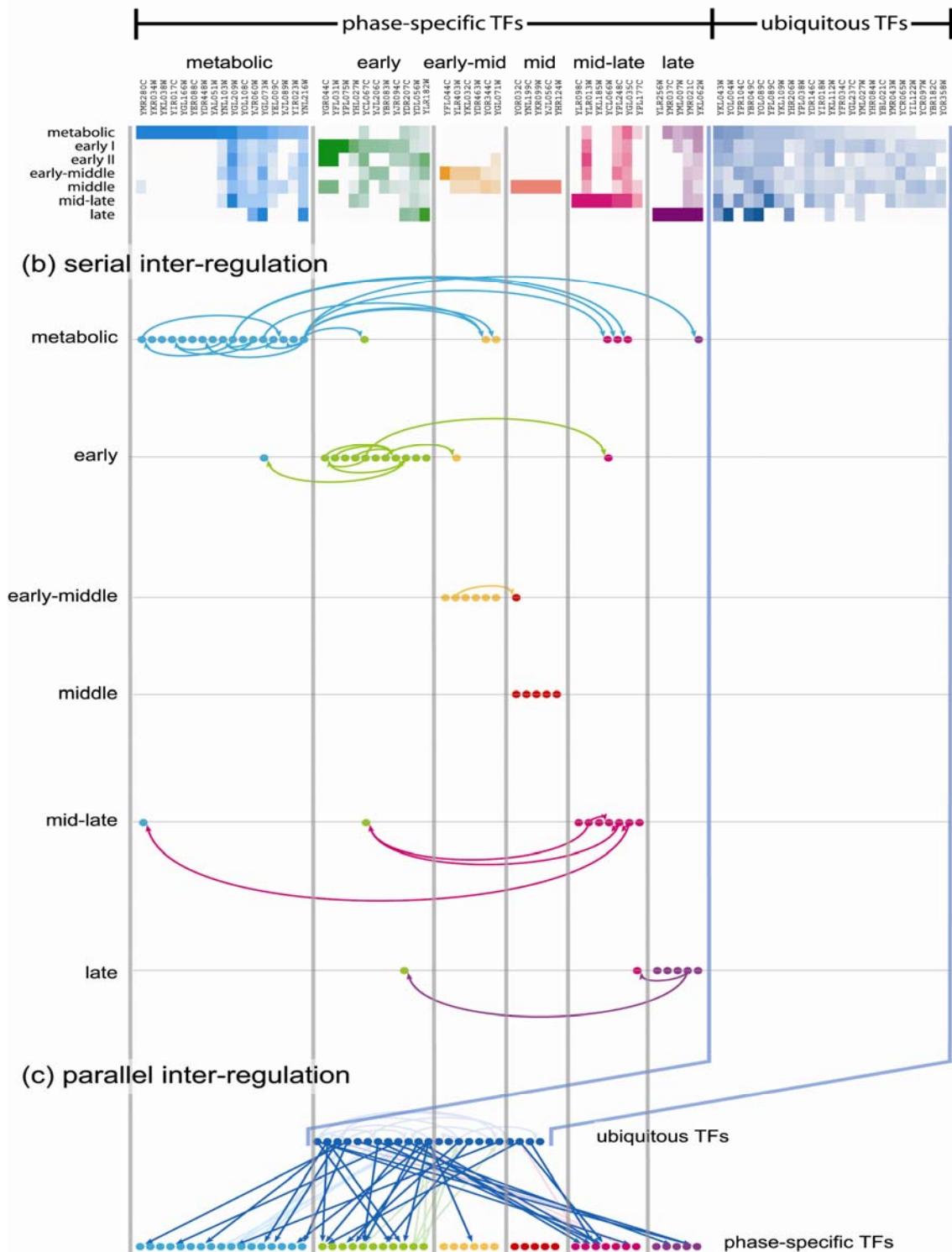


(c) parallel inter-regulation



Below we show a schematic of TF inter-regulation during the sporulation time-course.

(a) phase-specific and ubiquitous transcription factors



7.4 Availability

Data files containing the phase-specific sub-networks, details of TF clustering and the figures are available from the Supplementary Website.

8. References

1. Svetlov, V.V. and Cooper, T.G. (1995) *Yeast*, **11**, 1439-1484.
2. Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I. *et al.* (2002) *Science*, **298**, 799-804.
3. Horak, C.E., Luscombe, N.M., Qian, J., Bertone, P., Piccirillo, S., Gerstein, M. and Snyder, M. (2002) *Genes Dev*, **16**, 3017-3033.
4. Matys, V., Fricke, E., Geffers, R., Gossling, E., Haubrock, M., Hehl, R., Hornischer, K., Karas, D., Kel, A.E., Kel-Margoulis, O.V. *et al.* (2003) *Nucleic Acids Res*, **31**, 374-378.
5. Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E.L. *et al.* (2004) *Nucleic Acids Res*, **32**, D138-141.
6. Cho, R.J., Campbell, M.J., Winzeler, E.A., Steinmetz, L., Conway, A., Wodicka, L., Wolfsberg, T.G., Gabrielian, A.E., Landsman, D., Lockhart, D.J. *et al.* (1998) *Mol Cell*, **2**, 65-73.
7. Chu, S., DeRisi, J., Eisen, M., Mulholland, J., Botstein, D., Brown, P.O. and Herskowitz, I. (1998) *Science*, **282**, 699-705.
8. DeRisi, J.L., Iyer, V.R. and Brown, P.O. (1997) *Science*, **278**, 680-686.
9. Gasch, A.P., Huang, M., Metzner, S., Botstein, D., Elledge, S.J. and Brown, P.O. (2001) *Mol Biol Cell*, **12**, 2987-3003.
10. Gasch, A.P., Spellman, P.T., Kao, C.M., Carmel-Harel, O., Eisen, M.B., Storz, G., Botstein, D. and Brown, P.O. (2000) *Mol Biol Cell*, **11**, 4241-4257.
11. Jansen, R., Greenbaum, D. and Gerstein, M. (2002) *Genome Res*, **12**, 37-46.
12. Albert, R. and Barabasi, A.L. (2002) *Reviews of Modern Physics*, **74**, 47-111.
13. Barabasi, A.L. and Oltvai, Z.N. (2004) *Nat Rev Genet*, **5**, 101-113.
14. Shen-Orr, S.S., Milo, R., Mangan, S. and Alon, U. (2002) *Nat Genet*, **31**, 64-68.
15. Tavazoie, S., Hughes, J.D., Campbell, M.J., Cho, R.J. and Church, G.M. (1999) *Nat Genet*, **22**, 281-285.
16. Guelzim, N., Bottani, S., Bourgine, P. and Kepes, F. (2002) *Nat Genet*, **31**, 60-63.
17. Andrews, B.J. and Herskowitz, I. (1989) *Cell*, **57**, 21-29.
18. Breeden, L. and Mikesell, G.E. (1991) *Genes Dev*, **5**, 1183-1190.
19. Koch, C., Moll, T., Neuberg, M., Ahorn, H. and Nasmyth, K. (1993) *Science*, **261**, 1551-1557.
20. Kassir, Y., Granot, D. and Simchen, G. (1988) *Cell*, **52**, 853-862.
21. Vershon, A.K. and Pierce, M. (2000) *Curr Opin Cell Biol*, **12**, 334-339.
22. Kassir, Y., Adir, N., Boger-Nadjar, E., Raviv, N.G., Rubin-Bejerano, I., Sagee, S. and Shenharr, G. (2003) *Int Rev Cytol*, **224**, 111-171.
23. Strich, R., Suroskey, R.T., Steber, C., Dubois, E., Messenguy, F. and Esposito, R.E. (1994) *Genes Dev*, **8**, 796-810.
24. Bowdish, K.S., Yuan, H.E. and Mitchell, A.P. (1995) *Mol Cell Biol*, **15**, 2955-2961.
25. Xu, L., Ajimura, M., Padmore, R., Klein, C. and Kleckner, N. (1995) *Mol Cell Biol*, **15**, 6572-6581.
26. Olesen, J., Hahn, S. and Guarente, L. (1987) *Cell*, **51**, 953-961.

27. Hahn, S. and Guarente, L. (1988) *Science*, **240**, 317-321.
28. Forsburg, S.L. and Guarente, L. (1989) *Genes Dev*, **3**, 1166-1178.
29. Goh, K.I., Kahng, B. and Kim, D. (2002) *Phys Rev Lett*, **88**, 108701.
30. de Menezes, M.A. and Barabasi, A.L. (2004) *Phys Rev Lett*, **92**, 028701.
31. Kaufmann, E. (1993) *Chromosoma*, **102**, 174-179.
32. Pramila, T., Miles, S., GuhaThakurta, D., Jemielo, D. and Breeden, L.L. (2002) *Genes Dev*, **16**, 3034-3045.
33. Gailus-Durner, V., Xie, J., Chintamaneni, C. and Vershon, A.K. (1996) *Mol Cell Biol*, **16**, 2777-2786.
34. Nebohacova, M., Novakova, Z., Haviernik, P., Betina, S. and Kolarov, J. (1996) *Folia Microbiol (Praha)*, **41**, 115-116.
35. Mager, W.H. and Planta, R.J. (1990) *Biochim Biophys Acta*, **1050**, 351-355.
36. Della Seta, F., Ciafre, S.A., Marck, C., Santoro, B., Presutti, C., Sentenac, A. and Bozzoni, I. (1990) *Mol Cell Biol*, **10**, 2437-2441.
37. Diffley, J.F. and Stillman, B. (1989) *Science*, **246**, 1034-1038.
38. Loewith, R., Smith, J.S., Meijer, M., Williams, T.J., Bachman, N., Boeke, J.D. and Young, D. (2001) *J Biol Chem*, **276**, 24068-24074.
39. Pilpel, Y., Sudarsanam, P. and Church, G.M. (2001) *Nat Genet*, **29**, 153-159.
40. Futcher, B. (2002) *Curr Opin Cell Biol*, **14**, 676-683.
41. Koranda, M., Schleiffer, A., Endler, L. and Ammerer, G. (2000) *Nature*, **406**, 94-98.
42. Kumar, R., Reynolds, D.M., Shevchenko, A., Goldstone, S.D. and Dalton, S. (2000) *Curr Biol*, **10**, 896-906.
43. Pic, A., Lim, F.L., Ross, S.J., Veal, E.A., Johnson, A.L., Sultan, M.R., West, A.G., Johnston, L.H., Sharrocks, A.D. and Morgan, B.A. (2000) *Embo J*, **19**, 3750-3761.
44. Zhu, G., Spellman, P.T., Volpe, T., Brown, P.O., Botstein, D., Davis, T.N. and Futcher, B. (2000) *Nature*, **406**, 90-94.
45. McBride, H.J., Yu, Y. and Stillman, D.J. (1999) *J Biol Chem*, **274**, 21029-21036.
46. McInerny, C.J., Partridge, J.F., Mikesell, G.E., Creemer, D.P. and Breeden, L.L. (1997) *Genes Dev*, **11**, 1277-1288.
47. Simon, I., Barnett, J., Hannett, N., Harbison, C.T., Rinaldi, N.J., Volkert, T.L., Wyrick, J.J., Zeitlinger, J., Gifford, D.K., Jaakkola, T.S. et al. (2001) *Cell*, **106**, 697-708.
48. White, M.A., Dominska, M. and Petes, T.D. (1993) *Proc Natl Acad Sci U S A*, **90**, 6621-6625.
49. White, M.A., Wierdl, M., Detloff, P. and Petes, T.D. (1991) *Proc Natl Acad Sci U S A*, **88**, 9755-9759.
50. Kadosh, D. and Struhl, K. (1997) *Cell*, **89**, 365-371.
51. Bhoite, L.T., Allen, J.M., Garcia, E., Thomas, L.R., Gregory, I.D., Voth, W.P., Whelihan, K., Rolfes, R.J. and Stillman, D.J. (2002) *J Biol Chem*, **277**, 37612-37618.