Transcriptional regulation and function during the human cell cycle

Raymond J. Cho[1,6]*, Mingxia Huang[2]*, Michael J. Campbell[3,7]*, Helin Dong[4], Lars Steinmetz[1], Lisa Sapinoso[8], Garret Hampton[8], Stephen J. Elledge[2], Ronald W. Davis[1,5] & David J. Lockhart[4,9]

*These authors contributed equally to this work.

We report here the transcriptional profiling of the cell cycle on a genome-wide scale in human fibroblasts. We identified approximately 700 genes that display transcriptional fluctuation with a periodicity consistent with that of the cell cycle. Systematic analysis of these genes revealed functional organization within groups of coregulated transcripts. A diverse set of cytoskeletal reorganization genes exhibit cell-cycle–dependent regulation, indicating that biological pathways are redirected for the execution of cell division. Many genes involved in cell motility and remodeling of the extracellular matrix are expressed predominantly in M phase, indicating a mechanism for balancing proliferative and invasive cellular behavior. Transcripts upregulated during S phase displayed extensive overlap with genes induced by DNA damage; cell-cycle–regulated transcripts may therefore constitute coherent programs used in response to external stimuli. Our data also provide clues to biological function for hundreds of previously uncharacterized human genes.

Introduction

Cell division represents one of the most fundamental biological activities, with an essential role in events ranging from gametogenesis and wound-healing to multicellular development[1–3]. Proper regulation of gene activity during the cell cycle is likely to govern critical aspects of diverse processes. Indeed, disruption of cell-cycle–dependent mRNA regulation results in profound phenotypic consequences. For example, defects in the retinoblastoma (RB) and E2F proteins, which control transcription at the G_1/S transition, precipitate developmental defects, aberrant cell morphologies and uncontrolled proliferation[4–6]. The pleiotropic effects of these mutations suggest that a considerable proportion of all genes experience differential regulation during the cell cycle. Although these transcripts have been comprehensively catalogued in yeast, few targets of cell-cycle–dependent transcriptional machinery are known in higher organisms[7,8]. A global perspective on gene regulation during eukaryotic mitosis might therefore expand our understanding of both normal cell division and the abnormal phenotypes observed in certain pathological processes.

Results

Transcriptional profiling of cell division and DNA damage

We set out to identify cell-cycle–regulated transcripts in human cells on a genome-wide scale using high-density oligonucleotide arrays. Primary fibroblasts prepared from human foreskin were grown to approximately 30% confluence and synchronized in late G_1 using a double thymidine-block protocol[9]. Cultures were then released from arrest, and cells were collected every 2 hours for 24 hours, covering nearly 2 complete cell cycles.

Messenger RNA was isolated, labeled and hybridized to sets of arrays containing probes for approximately 40,000 human genes and non-overlapping ESTs (ref. 10). We carried out the entire synchronization experiment in duplicate under identical conditions for 6,800 genes. The two data sets were averaged and analyzed using both supervised and unsupervised clustering of expression patterns. We identified 12 patterns that displayed a periodicity consistent with the length of the cell cycle in both data sets, with peaks corresponding to late G_1, S, G_2 and M phases. Based on comparison with these patterns, 731 transcripts from the initial set of 40,000, including 344 previously uncharacterized ESTs, were assigned to cell-cycle–regulated expression clusters (Fig. 1). These assignments were made for only transcripts that oscillated in obvious patterns, and the conservative nature of this analysis makes it possible that some cell-cycle–regulated transcripts were not selected. Consistent with previous studies, multiple cyclin family members and CDC family members were identified, as well as numerous genes involved in DNA replication and chromatin structure[4]. The primary data and clustering results are available (http://www.salk.edu/docs/labs/chipdata/).

We also sought to better understand the relationship between transcriptional changes that normally occur during mitosis and those observed during events that affect progression through the cell cycle. For the purposes of this assessment, we quantitated transcript levels in fibroblasts after exposure to ultraviolet light and methyl methane sulfonate (MMS), both of which cause DNA damage and consequent cell-cycle arrest at multiple checkpoints.

DNA-damage experiments were performed in replicate and results were compared between treated and untreated cells to identify consistent differences. We used stringent criteria to iden-

[1]Department of Genetics, Stanford University School of Medicine, Stanford, California, USA. [2]Department of Biochemistry and Molecular Biology, Department of Molecular and Human Genetics, Howard Hughes Medical Institute, Baylor College of Medicine, Houston, Texas, USA. [3]Molecular Applications Group, Palo Alto, California, USA. [4]Affymetrix, Inc., Santa Clara, California, USA. [5]Department of Biochemistry, Stanford University School of Medicine, Stanford, California, USA. [6]Present address: InGenuity Systems, Inc., Alviso, California, USA. [7]Present address: Celera Genomics, Foster City, California, USA. [8]Genomics Institute of the Novartis Research Foundation, San Diego, California, USA. [9]Salk Institute for Biological Studies, Laboratory of Genetics, La Jolla, California, USA. Correspondence should be addressed to M.J.C. (e-mail: michael.campbell@fc.celera.com).

tify these changes. To qualify, differences had to be called in at least two independent experiments and more than a twofold change was required. Of the more than 6,000 genes assayed, 36 were consistently induced and 6 were repressed after ultraviolet exposure, and 53 were induced and 7 were repressed after MMS exposure (see http://www.salk.edu/docs/labs/chipdata/).

Systematic functional analysis of expression clusters

Previous large-scale expression studies have relied heavily on subjective evaluations of the functions of differentially expressed genes[7,11,12]. But the rapidly increasing number of these data sets has created an urgent need for more standardized methods for assessing the coordinate regulation of biological pathways. Over-representation of genes from a common functional class within an expression cluster might reveal global pathway activation[13]. We implemented a statistical analysis based on this hypothesis. More than 3,000 characterized human genes represented on the arrays were classified into 160 functional categories (Table 1). Using the binomial distribution function, we calculated *P* values based on the frequency of occurrence of genes with common biological function in each expression cluster. Statistically significant enrichments for specific functions were detected in 3 of the 4 groups of cell-cycle–regulated transcripts, with *P* values ranging from 2.7×10^{-5} to 6.6×10^{-8} (Fig. 2).

Transcriptional regulation of central mitotic processes

The most basic test of our systematic analysis is the detection of biological processes known to have a central role in cell division. We found coordinate upregulation of transcripts involved in cell-cycle control, DNA replication and chromosome segregation, providing a basic validation of the approach (Figs. 2 and 3).

The comprehensive identification of genes that experience cell-cycle–dependent transcriptional control in known pathways may provide insight into how regulatory disruption exerts biological effects. We first examined transcripts for genes that govern cell-cycle progression ($P=2.7\times10^{-6}$). *CCNA*, *CCNB* and *CCNF* family members were coordinately induced at the G_2-M transition, at which point they are known to regulate the entry into mitosis. We also found, however, that some genes were highly expressed during transitions they are believed to negatively regulate (such as *CDKN2D* and *MYT1*, during G_1-S). Coordination was observed for genes implicated in proteolytic control of cell-cycle progression, including *CCNA2*, *CDC34* (encoding protein kinase I), *CDC28* (protein kinase II), the *CDC5*-like gene *PLK*, *UBCH10* (cyclin-specific ubiquitin ligase) and *E2-EPF* (ubiquitin carrier protein), all of which are induced during M phase. Notably, a second set of ubiquitin-associated transcripts was highly expressed during the G_1-S transition, including the *NEDD4*-like gene *WWP2* and *CUL4A*.

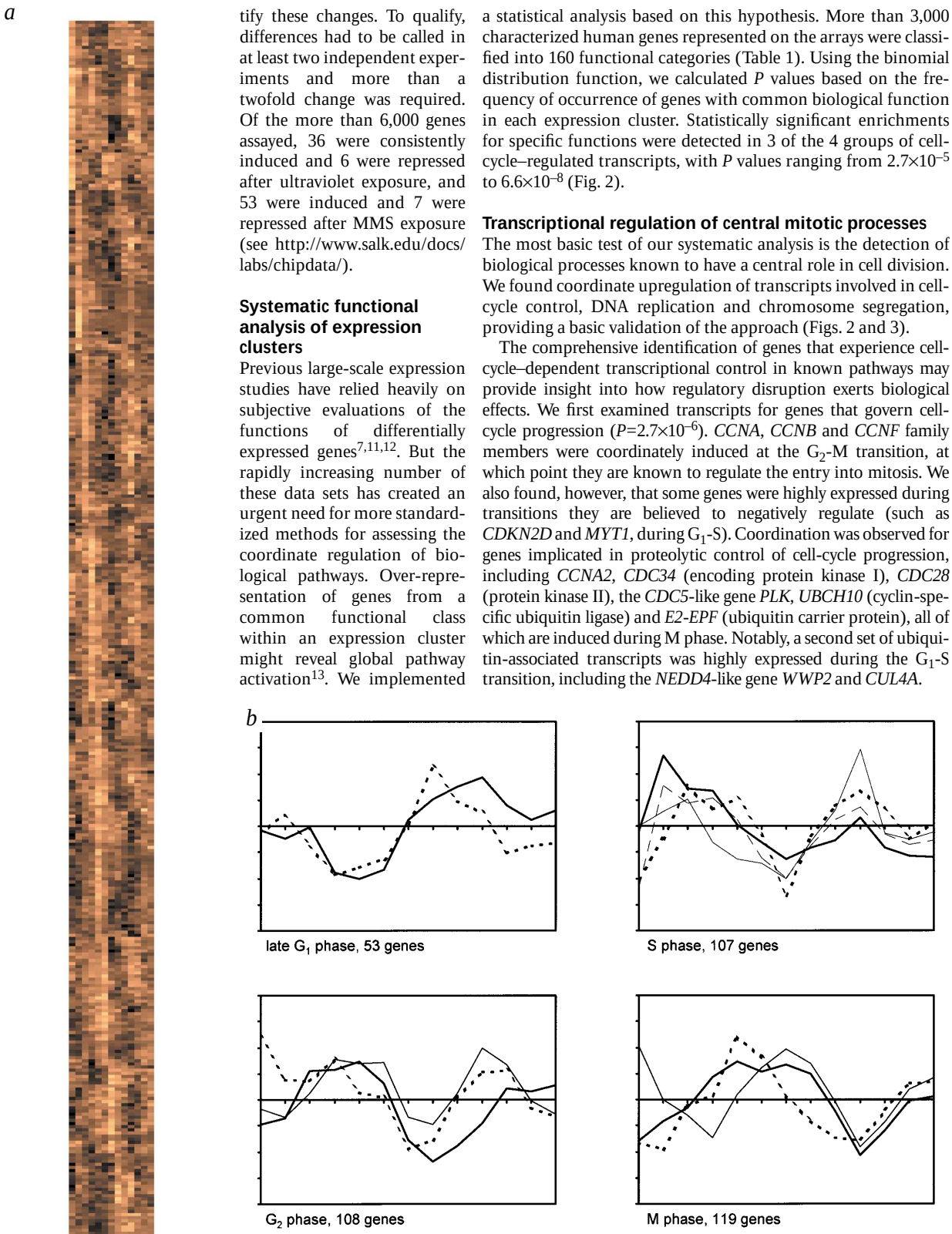

Fig. 1 Clustering analysis of cell-cycle–regulated transcripts. ***a***, Temporal behavior of 387 human cell-cycle–regulated transcripts with known function. Each horizontal line displays the expression data for one gene, where relative mRNA abundance, averaged between two experiments, is represented by color intensity. Twelve consecutive time points, each separated by two hours, are displayed for each gene. ***b***, Averaged mean centers of each of the 12 cell-cycle–regulated clusters identified here. Fluctuation of the variance with respect to the mean (units of 0.5, *y*-axis markers) is shown for 13 consecutive time points, each separated by 2 h (*x*-axis markers).

Table 1 • Biological function enrichments in cell-cycle–regulated expression clusters

Biological function		late G_1 (53)		S (107)		G_2 (108)		M (119)	
amino acid metabolism	(35)	0.1	0/0	0.5	1.1	0.5	1/1	0.3	0/1
cell-to-cell adhesion	(137)	0.5	0/1	0.7	3/2	0.6	1/2	**4.7**	**11/2**
chromosome segregation	(17)	0.1	0/0	0.1	0/0	2.7	3/0	**5.0**	**5/0**
cytokine signaling	(151)	0.5	0/1	0.9	4/2	0.6	2/2	1.1	0/3
cytoskeletal reorganization	(33)	0.1	0/0	1.9	3/0	**6.1**	**7/1**	0.5	1/1
DNA replication	(47)	**7.2**	**7/0**	0.3	0/1	0.3	0/1	0.8	2/1
glycolysis	(31)	0.1	0/0	0.2	0/0	0.2	0/0	0.2	0/1
G-protein signaling	(223)	0.7	0/2	0.7	3/3	1.0	1/3	0.7	4/4
immune regulation	(274)	0.6	2/2	1.2	1/4	1.2	1/4	1.0	2/5
intracellular transport	(114)	0.4	1/1	1.2	4/2	0.5	1/2	0.8	3/2
ionic homeostasis	(47)	0.6	1/0	0.3	0/1	0.3	0/1	0.3	0/1
mitosis and cell-cycle control	(89)	0.3	0/1	0.5	1/1	**5.0**	**9/1**	**5.7**	**10/1**
mRNA regulation	(553)	1.3	7/4	0.9	10/8	1.0	6/8	0.9	8/9
muscular contraction	(82)	0.3	0/0	2.1	5/1	**5.3**	**9/1**	0.3	1/1
neurotransmitter signaling	(69)	0.2	0/1	0.5	0/1	0.4	1/1	0.5	0/1
PIP signaling	(51)	0.2	0/0	0.3	0/1	0.4	1/1	0.4	1/1
protein phosphorylation	(292)	0.6	2/2	0.7	4/4	1.1	7/4	1.0	7/5
translation	(102)	1.5	3/1	0.5	1/2	0.7	0/2	0.8	0/2

Shown are representative biological functional categories and their enrichment, as calculated based on the binomial distribution function, in sets of cell-cycle–regulated expression clusters. Of 160 functional categories analyzed, 18 are shown. The total number of genes in each functional category and in each temporal expression group is shown in parentheses. Negative logarithms of *P* values are listed. The superscript values are the number of genes from a given functional category observed (numerator) and those expected to be found by chance (denominator) in a given expression group. *P* values less than than 1×10^{-4} are in bold.

Consistent with findings in yeast, numerous DNA replication genes were induced in late G_1 and S phase, including those encoding some E2F targets, such as DNA polymerase-δ, DNA primase and replication factors, PCNA, uracil DNA glycosylase *UNG2* and multiple MCM proteins ($P=6.6\times10^{-8}$). We also observed over-representation of regulators of actin-based cytoskeletal reorganization in the G_2 clusters ($P=8.4\times10^{-7}$). These genes include isoforms of *RHO*, *ACTN1*, *ARF6*, *ARPC1B*, *ARNO* and *CDK5*. Other induced transcripts encode gelsolin-like actin polymerization domains, such as *CAPG*, but have not been assigned a more specific biological role. Enrichment during G_2 was also detected for transcripts involved in muscular contraction such as *TPM1* and *MYRL2* ($P=4.7\times10^{-6}$).

Genes that regulate chromosome segregation were coordinately upregulated during M phase, such as *RMSA1* (regulator of mitotic spindle assembly) and *NEK2* (centrosomal kinase; $P=1.1\times10^{-5}$). Perhaps the most striking global upregulation during M phase was observed for genes encoding microtubule-dependent motor proteins that physically effect chromosome segregation, including mitosin and virtually every kinesin known to interact directly with the centromere[14,15]. Cell-cycle regulation was not observed for dynein genes, indicating restriction of transcriptional control to a subset of motor proteins.

Cell-cycle control of extracellular matrix activity and apoptosis

These analyses reveal that a large proportion of the genes induced during M phase facilitate cell-to-cell adhesion and perform other roles in the extracellular matrix ($P=1.9\times10^{-5}$). Genes found in these clusters encode molecules such as the catenin CTNND2, integrin α-6, collagens COL6A1 and COL6A3, and junction plakoglobin. Other genes highly expressed during G_2 and M have diverse roles in positively regulating cell movement through the extracellular matrix: for example, *VASP* (vasodilator-stimulated phosphoprotein), *HAS2* (hyaluron synthase), *RHAMM* (hyaluron receptor) and *PLAT* (tissue plasminogen activator; Figs. 2 and 3; refs. 16–19).

Enrichment was also detected in the G_2 clusters for genes encoding effectors of programmed cell death ($P=2.7\times10^{-5}$), including the Bcl-2 homolog BAK1, caspase-3, and numerous caspase substrates such as α-fodrin, CDKN1, CAS, HEF1 and plectin. But a number of negative regulators of apoptosis were induced during other phases, including *MCL1* and *API3*.

DNA-damage response involves S-phase transcripts

It is possible that cell-cycle–regulated transcripts constitute organized programs of biological activities that are required outside of cell division. DNA damage in *Saccharomyces cerevisiae*, for example, is known to induce a significant proportion of the transcripts that are induced during late G_1 phase (M.J.C. and S. Tavazoie, unpublished data). The observation of such common targets might be indicative of responses that involve cell-cycle–dependent transcriptional machinery.

Of the 6,800 transcripts surveyed here, 42 were reproducibly induced in response to either ultraviolet light or MMS in human fibroblasts. These groups included many known damage-induced transcripts such as *JUN-B*, *GADD45* and *IEX1*. We found that genes upregulated by these DNA-damaging agents overlapped with those in the S-phase cluster ($P=1.9\times10^{-5}$ for ultraviolet-induced transcripts and 1.1×10^{-13} for MMS-induced transcripts), but not with any other group of cell-cycle–regulated genes. FACS analyses (data not shown) indicate that cells were not arrested in S phase after DNA damage, suggesting that the observed induction of these transcripts was not simply the result of arrest at this stage.

Discussion

These systematic expression analyses revealed both known and unexpected results. DNA-replication genes were strongly upregulated during late G_1, as were numerous cell-cycle control and chromosome-segregation genes during M phase (Table 1). However, the observed upregulation of genes related to cytoskeletal reorganization and programmed cell death during G_2, and induction of extracellular matrix-associated genes during M, indicates that a broader set of biological pathways is transcriptionally activated during cell division than was previously known. Notably, whereas six biological functional classes were found to be strongly over-represented in cell-cycle–regulated expression clusters, none were significantly under-represented. One interpretation of this observation is that acquisition of cell-cycle–specific regulatory elements occurs stochastically in the genome, but is retained by genes for which mitotic regulation confers a biological advantage.

These data provide a useful point of reference for examining established mechanisms of cell-cycle regulation. For example, the coordinate induction of cell-cycle-control genes associated with ubiquitin during M phase suggests a requirement for restricting

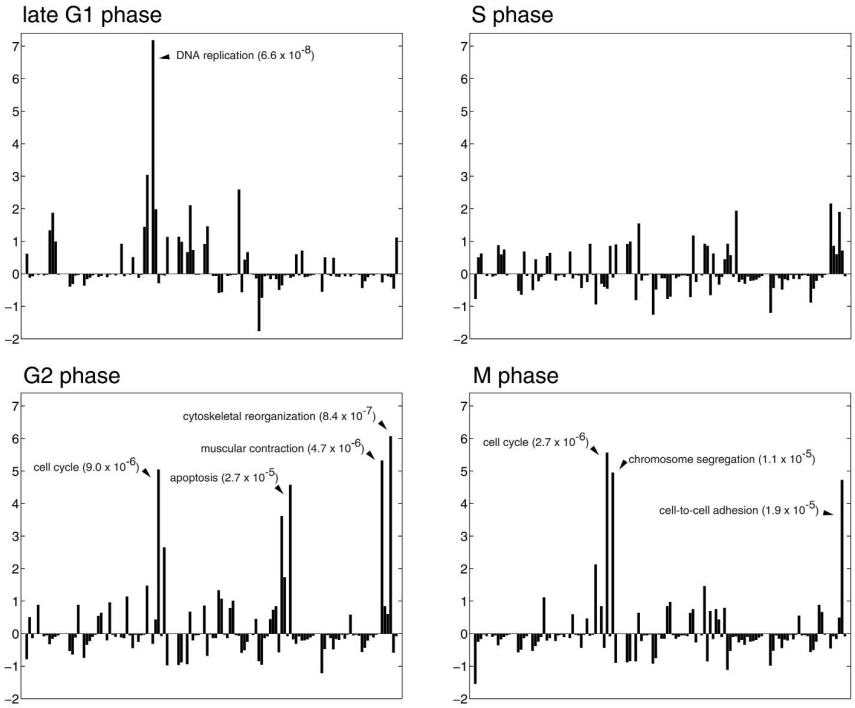

Fig. 2 Over- and under-representation of genes in biological functional categories in cell-cycle–regulated expression clusters, as calculated by the binomial distribution function. More than 160 functional categories were used (numbered along the *x* axis), and the negative logarithm of the *P* value is plotted on the *y* axis. Functional over-representations with a *P* value less than 1×10^{-4} are labeled. Results are shown for late G_1, S, G_2 and M phases. *P* values for under-represented functional categories were arbitrarily given negative values so that they could be visually distinguished from over-represented groups.

the time of activity of these genes, and might indicate another means by which aberrant transcriptional regulation interferes with normal cell-cycle progression. These results are in direct contrast to observations made in *Saccharomyces cerevisiae*, in which transcriptional control of proteolytic regulators is observed only rarely[7]. Evidence has implicated cullin family members in the proteolytic degradation of G_1-S regulators, supporting the possibility that the ubiquitin-associated genes induced at this time might participate in checkpoint regulation[20].

These data also link previously characterized genes to specific mitotic activities. For example, the nucleotide excision repair genes *RAD2* and *ERCC1* are induced during S phase, suggesting that they might function in the correction of misincorporation in newly synthesized DNA. We also observed that the DNA damage-responsive gene *GADD45* (increased levels of which have been shown to block passage through a G_2-M checkpoint) is transcriptionally repressed during this period[21]. Naturally low levels of *GADD45* transcript during G_2-M might permit cell-cycle progression in the absence of DNA damage-dependent induction, providing a sensitive means of checkpoint regulation[22].

The transcripts induced during S phase that are involved in the regulation of actin-based cytoskeletal reorganization include several genes previously implicated in cytokinesis. *RHO*-mediated reorganization of actin stress fibers is essential for mitosis, and its deregulation has been implicated in tumor-specific abnormalities in cell separation[23–26]. α-Actinin and moesin localize to the cleavage furrow during cell division and are believed to link actin filaments to the plasma membrane[27,28]. Coordinate transcriptional upregulation of this set of genes occurs relatively early in the cell cycle, well before the onset of mitosis.

A broader set of actin-related transcripts not linked directly to cytokinesis are also induced at this time. Their gene products might be responsible for more global cytoskeletal changes that accompany mitosis. For example, ARF6 and ARNO cooperate to direct reorganization of cortical actin at both endosomes and the plasma membrane and have recently been implicated in cell migratory activities[29–31]. CDK5 modulates RAC activity and is required for the proper development of actin-based outgrowths such as neurites[32]. Other genes encode gelsolin-like actin polymerization domains and are hypothesized to function in cellular motility, but have not been assigned a specific role in cell division[33]. These results indicate that diverse cytoskeletal reorganization pathways are activated to facilitate the gross morphological changes required in dividing cells. It remains to be determined whether specific temporal expression is sufficient to delineate such pleiotropic functions for this set of genes.

The systematic analysis applied here is perhaps most valuable in detecting unexpected coordinate regulation of functional pathways, which is likely to prove difficult based solely on subjective surveys of genome-scale experimental data. Using this approach, we identified an over-representation of transcripts with cell-to-cell adhesion and extracellular-matrix–associated functions in the G_2 cluster. At least some of these gene products, such as integrin α-6, plakoglobin, desmoyokin and laminin B2, represent components of structures (such as desmosomes and hemidesmosomes) that facilitate intercellular adhesion and communication[34]. Transcriptional upregulation of these genes might prepare dividing cells for re-establishment of contact and communication with the extracellular environment.

A number of matrix-associated transcripts in the G_2 and M clusters are known to positively regulate cellular motility, however, in some cases through signaling to the actin cytoskeleton. This group of genes includes *PDGFA*, *HAS2*, *RHAMM*, *VASP* and *PLAT*. PDGFA functions as powerful mitogen and chemoattractant, whereas the hyaluronate-rich environment generated by enzymes such as HAS2 appears to facilitate cell movement. Both VASP and RHAMM are believed to increase cellular motility through direct physical interactions with the cytoskeleton[35,36]. It is possible that these genes have a role in the separation and migration of daughter cells following mitosis. Many of these gene products have been characterized in terms of their ability to increase the motility of tumor cells. For example, overexpression of either HAS2 or RHAMM is sufficient to promote development of invasive cellular phenotypes[17,19]. Therefore, restriction of the activity of these genes to defined cell-cycle intervals might represent a general mechanism for enabling activities required for proliferation, while limiting tumorigenic cellular behavior.

The induction of pro-apoptotic genes such as *BAK1* during

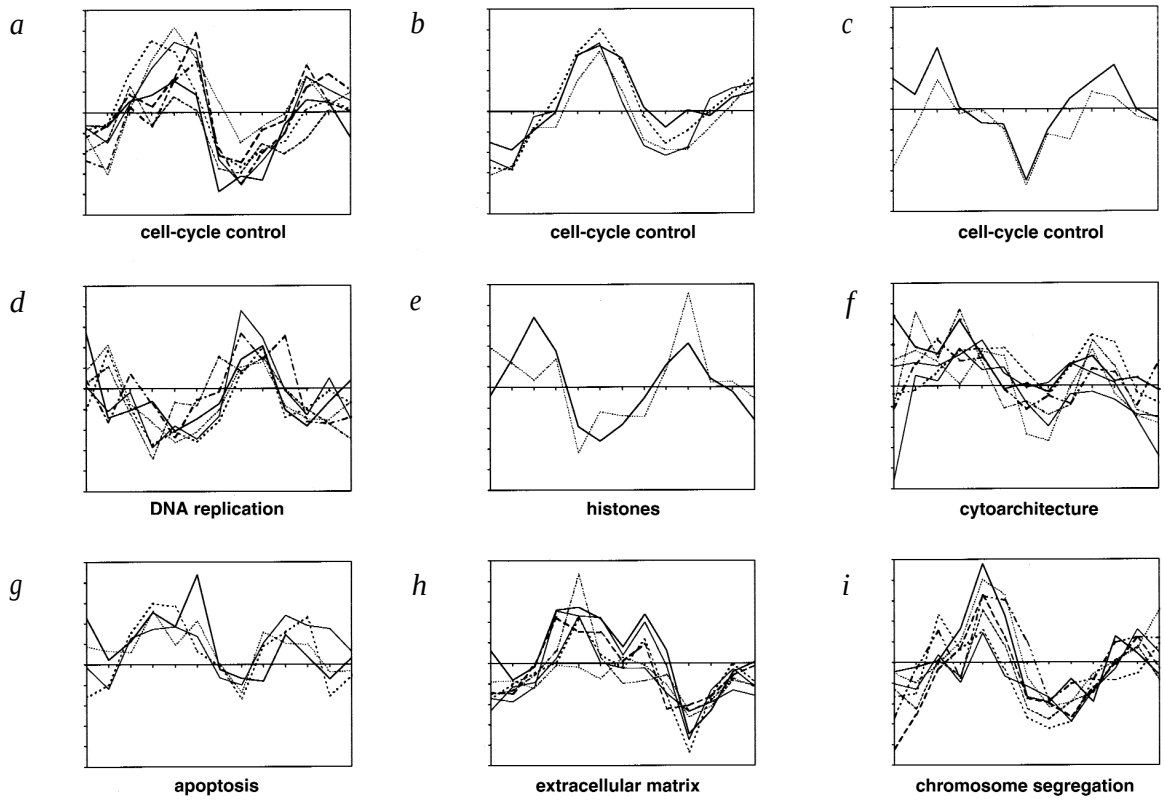

Fig. 3 Cell-cycle regulation of human genes with related biological function. The normalized abundance of mRNA levels, averaged from two experiments, is displayed for each gene. Fluctuation of variance with respect to the mean (units of 0.5, *y*-axis markers) is shown for 13 consecutive time points, each separated by 2 h (*x*-axis markers). Results are shown for genes involved in cell-cycle control (*a*–*c*), nucleosome assembly and DNA replication (*d*,*e*), cytoskeletal reorganization (*f*), apoptosis (*g*), extracellular matrix remodeling (*h*) and chromosome segregation (*i*). The genes shown are as follows: ***a***, *CDC2*, *CDKN2D*, *CDC28*, *CDC28*, *CCNA2*, *CCNF* and *UBCH10*; ***b***, *CCNB1*, *CDC20*, *E2-EPF* and *PLK*; ***c***, *MYT1* and *WEE1*; ***d***, *CHAF1B*, *PCNA*, *RAD2*, replication factors *C3* and *C4*, and *UNG2*; ***e***, *H1F4* and *H4*; ***f***, *TPM1*, *ARF6*, *MYOIC*, *MYRL2*, *SPTN1* and *RHOB*; ***g***, *BAK1*, *CASP3*, *SPTAN1* and *PLEC1*; ***h***, *COL6A1*, *HAS2*, *ITGA6*, *MSN*, *PLAT*, *PRG1* and *VASP*; ***i***, *CENPA*, *HKSP*, *KID*, *KNSL2*, *KNSL5*, *KNSL6*, *TOP2A* and *TTK1*.

G_2 is suggestive of increased apoptotic activity during this period. BAK1 has been proposed to induce caspase activity by means of oligomerization-based pore formation in mitochondria and consequent cytochrome *c* release, leading to degradation of a broad range of target proteins[37]. We found that many of these apoptotic substrates also display G_2 induction, including effectors of actin-based cytoskeletal reorganization such as α-fodrin, CDKN1 and HEF1 (ref. 31). The decreased activity of Bcl-related inhibitors of apoptosis during G_2/M has been proposed to be partially responsible for the observed susceptibility to programmed cell death resulting from arrest at this point[38]. These results suggest that a limited set of pro-apoptotic pathways are available during G_2 and might contribute to these effects.

The induction of anti-apoptotic genes such as *MCL1* during S phase might serve to counteract the pro-apoptotic effects of the E2F1 transcription factor, which positively regulates the G_1/S transition[39]. Evidence has supported a broad role for apoptosis-related genes in cell-cycle progression, possibly through direct regulation of the E2F class of transcription factors or other activities downstream of cyclin-dependent kinase-2 (Cdk2; refs. 40–42). Further experiments are required to evaluate the full functional significance of these observations.

The significant overlap between DNA-damage–induced and S-phase transcripts indicates that common subsets of genes might be required for response to apparently unrelated cellular conditions. This phenomenon has been observed in *S. cerevisiae*, in which widely divergent biological processes have been linked by their dependence on the proper function of a common set of genes[43]. One explanation might be that many genes that function directly in DNA damage repair are required for the proper execution of chromosome replication. But few of the overlapping genes participate directly in DNA synthesis or repair. Instead, these gene products function predominantly in the stress response, either through intracellular signaling (the transcription factors JUN-B, ATF3, ATF4 and EGR, and the dual specificity phosphatase DUSP5) or extracellular communication (ligands including IL-6 and IL-11). It is possible that the general regulatory response to cellular stress is governed partially by transcriptional mechanisms that also coordinate gene activity during the cell cycle. These data offer an opportunity to identify other transcriptional responses that display significant similarity with cell-cycle–dependent gene regulation.

A number of genes that are cell-cycle–regulated, including *EDN1*, *EGR1*, *PLAT* and *PLAUR*, were previously reported to be transcriptionally activated effectors of the serum response[12]. Because one effect of serum deprivation is mitotic synchronization, the previously observed results might not be specific to the response to serum. Further analysis and comparison with the previous results will allow identification of serum-dependent transcriptional effects that are not the result of cell-cycle arrest. These data should provide a helpful general reference point for future human profiling studies that involve aspects of cell-cycle progression.

More than 300 previously uncharacterized EST sequences were also found to be cell-cycle–regulated at the mRNA level. Because many human ORFs have no significant similarity to characterized genes, prediction of biological function based entirely on sequence homology is limited. Identification of transcriptional behavior under different conditions provides an additional means for assessing the function of these genes. Approximately 25% of the previously characterized genes identified here as cell-cycle–regulated were known to have a mitotic-phase–specific role, compared with 6% of all previously characterized genes represented on the arrays. Therefore, observation of cell-cycle regulation for a transcript indicates a more than fourfold increase in the probability of function during cell division.

The cell-cycle cluster to which an uncharacterized EST is assigned provides additional clues to its biological role. For example, based on the calculations made about biological function here, genes upregulated during M phase have a more than twofold higher probability of participating in cell-to-cell adhesion. Moreover, assessing these transcriptional results in the context of sequence homology may provide additional clues to cellular roles[11]. For example, one uncharacterized EST induced during G_2/M shows similarity to the gene encoding katanin, an ATPase whose oligomerization leads to microtubule severing during cell division[44,45]. Another EST, induced in S phase, displays strong similarity to the gene encoding the RTEF1 transcription factor, which preferentially induces muscle-specific genes[46,47]. This gene might be partially responsible for the cell-cycle regulation of myosin transcripts. The increasing number of genome-scale transcriptional data sets will soon supply multiple lines of evidence concerning the cellular role of many newly sequenced or previously uncharacterized genes. These data should provide opportunities for more systematic comparisons between expression and DNA sequence data for the purposes of predicting gene function.

Although the perspective provided by our study is restricted to the transcriptional level, the detection of functions ranging from DNA replication to cytoskeletal reorganization indicates that these data constitute a representative survey of mitotic activities in higher eukaryotes. The observed regulation of processes such as extracellular matrix remodeling and programmed cell death indicates that cell-cycle–dependent transcriptional control influences the execution of biological programs not obviously related to mitosis. In contrast to the limited changes observed following deletion of specific transcription factors[22,48], the regulatory fluctuations observed here and in other expression profiles of global cellular processes involve broad functional classes of genes and multiple members of gene families[12,49]. In a manner analogous to quantitative genetic phenomena, certain biological activities might depend on the combination of numerous, subtle transcriptional changes whose individual effects are difficult to measure.

The functional classifications used here represent only a first attempt to formally structure public domain knowledge about human genes for the purpose of systematically analyzing experimental genomic data. Generation of increasingly refined representations of such information should enable a far more sophisticated understanding of this and other transcriptional data sets. Many biological activities might result from a small number of regulatory targets and therefore not meet the criteria for statistical over-representation. Given the number and size of expression data sets, however, biologically meaningful evaluations of findings represent the rate-limiting step in studies of this type. The methods demonstrated here provide an initial solution to this growing problem. Because of the scope of this data set, more conclusions remain to be drawn than can be described in this manuscript. We encourage the reader to directly access the data by means of our web site (http://www.salk.edu/docs/labs/chipdata/).

Methods

Synchronization of cells. Human cell lines used here were derived from human foreskin samples[50]. Cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum and antibiotics (all from Gibco). Cells were grown in 150 cm^2 flasks to 25–30% confluence and synchronized using a double-thymidine-block protocol[9], with a first block for 16 h, a 10 h release, and a second block for 15 h. The final concentration of thymidine used in the block medium was 2 mM. We collected ~2×10^7 cells at each time point. About 10% of cells from each time point were fixed in ethanol and stained with propidium iodide for flow cytometry analysis, and the remainder were immediately frozen at –80 °C for subsequent isolation of RNA.

Ultraviolet and MMS treatment. Cells were grown asynchronously and passaged before reaching 80% confluency. At the last passage, cells were combined and split equally into 18 150-cm^2 dishes and grown overnight (~16 h). Media was then removed from each dish and stored separately. All dishes were then rinsed once with 1×PBS. Six dishes were treated with ultraviolet irradiation (40 J/m^2) and six dishes were treated with MMS at a final concentration of 0.015% (v/v). The remaining six dishes were not subjected to any treatment. Media was then added back to all dishes and all samples were incubated another 4 h before collection. For collection, cells were trypsinized for removal from plates, immediately flash frozen in liquid nitrogen, and stored at –80 °C .

Isolation and hybridization of mRNA. Total RNA was isolated and biotin-labeled as described[10]. Samples were hybridized to a total of five arrays containing approximately 1.4×10^6 oligonucleotide probes to approximately 40,000 human genes and ESTs (Affymetrix). We carried out hybridizations at 42 °C for 14 to 16 h with mixing on a rotisserie at 60 r.p.m. Following hybridization, the solutions were removed, the arrays were rinsed with 6×SSPE-T (0.9 M NaCl, 60 mM NaH_2PO_4, 6 mM EDTA, 0.005% Triton X-100, pH 7.6), rinsed with 0.5×SSPE-T (75 mM NaCl, 5 mM NaH_2PO_4, 0.5 mM EDTA, 0.005% Triton X-100, pH 7.6), and incubated with 0.5×SSPE-T at 42 °C for 15 min. Following washing, hybridized biotinylated cRNA was fluorescently labeled by incubation with streptavidin-phycoerythrin (2 μg/ml; Molecular Probes) and acetylated BSA (1 mg/ml; Sigma) in 6×SSPE-T at 42 °C for 10 min. Unbound streptavidin-phycoerythrin was removed by rinsing at RT before scanning. The arrays were read at a resolution of 3 μm using a specially designed confocal scanner (Affymetrix) as described[10].

Expression data analysis. We normalized data between samples by setting the mean hybridization signal for each sample equal to 1,000 arbitrary units and the standard deviation to 1,500 arbitrary units. The hybridization signal for each gene or EST was then normalized to a unit standard deviation with a mean of zero. The values for the 2 time courses were then averaged, except for the zero- and 12-hour time points, for which duplicates were not available. The averaged time course was then subjected to clustering using the Pearson correlation coefficient as the metric of similarity.

The abundance of cell-cycle-phase–specific transcripts is expected to fluctuate periodically. We identified genes displaying periodic patterns using supervised clustering methods. We first compared transcriptional fluctuations with a series of sine waves offset by consecutive time intervals, similar to previous analyses of cell-cycle expression data[8]. Typically, we demanded that a gene be correlated to one of the sine wave patterns by a Pearson's correlation coefficient of at least 0.70 over 12 samples. However, many known cell cycle-regulated transcripts displayed patterns that did not conform to simple sine waves. Therefore, we used these genes to identify additional groups of periodically fluctuating transcripts in the hierarchically clustered data set. Based on these approaches, we identified 387 cell-cycle–regulated transcripts. We assigned each transcript to the most similar of 12 cell-cycle–regulated seed patterns, based on a correlation coefficient. These 12 clusters were then each assigned to a cell cycle phase, generating the clusters shown in Fig. 1. More than 90% of these transcripts were scored as detectable in more than one-quarter of the time points.

The database contains 160 controlled-vocabulary categories that are classified hierarchically. The database schema is similar to that of the MIPS yeast database (http://www.mips.org), but also contains functions relevant to multicellular organisms. More than 3,000 human genes were assigned to one or

more of these categories by reviewing information from Medline and other publicly available information sources. If the published information strongly supported the placement of a gene into a category, that assignment was made. More than one functional assignment could be made for a given gene.

The binomial distribution function was applied as follows. Briefly, if the number of genes in an expression cluster with a given biological function was less than or equal to the number expected by random chance, then the *P* value was calculated using the formula:

$$P=\left[\begin{matrix} n \\ x \end{matrix}\right]p^{x}(1-p)^{n-x}$$

where n is the number of genes in a given cluster, x is the observed number of genes found in both a given cluster and a given functional category, and p is the overall frequency at which genes of a given functional category are found in the genes represented on arrays. If the number of genes in an expression cluster with a given biological function was greater than the number expected by random chance, then the *P* value was calculated using the formula:

$$P=1-\sum_{i=0}^{x-1}\left[\begin{matrix} n \\ i \end{matrix}\right]p^{i}(1-p)^{n-i}$$

Note added in proof: The expression clustering results displayed close agreement with independent analysis using self-organizing maps[51]. Please see http://www.salk.edu/docs/labs/chipdata/.

Acknowledgments

We thank D. Galloway for providing the cell line; L. Wodicka for help with experiments; M. Ho for help with data analysis; M. Mittmann and D. Balaban for array design; Affymetrix Product Development for array testing and validation; F.F. Cai for help in construction of the human functional database; and M. Mindrinos, S. Chang, C. Chon, J. Yang and N. Thayer for discussions and a critical reading of the manuscript. This work was supported in part by NIH grants to S.J.E. and R.W.D.

Received 3 February; accepted 10 November 2000.

1. Nicklas, R.B. How cells get the right chromosomes. *Science* **275**, 632–637 (1997).
2. Nasmyth, K. Viewpoint: putting the cell cycle in order. *Science* **274**, 1643–1645 (1996).
3. Stillman, B. Cell cycle control of DNA replication. *Science* **274**, 1659–1664 (1996).
4. Sanchez, I. & Dynlacht, B.D. Transcriptional control of the cell cycle. *Curr. Opin. Cell Biol.* **8**, 318–324 (1996).
5. Dyson, N. The regulation of E2F by pRB-family proteins. *Genes Dev.* **12**, 2245–2262 (1998).
6. Friend, S.H. *et al.* A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. *Nature* **323**, 643–646 (1986).
7. Cho, R.J. *et al.* A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol. Cell* **2**, 65–73 (1998).
8. Spellman, P.T. *et al.* Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell* **9**, 3273–3297 (1998).
9. Rao, P.N. & Johnson, R.T. Mammalian cell fusion: studies on the regulation of DNA synthesis and mitosis. *Nature* **225**, 159–164 (1970).
10. Lockhart, D.J. *et al.* Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotechnol.* **14**, 1675–1680 (1996).
11. Chu, S. *et al.* The transcriptional program of sporulation in budding yeast. *Science* **282**, 699–705 (1998); erratum: **282**, 1421 (1998).
12. Iyer, V.R. *et al.* The transcriptional program in the response of human fibroblasts to serum. *Science* **283**, 83–87 (1999).
13. Tavazoie, S., Hughes, J.D., Campbell, M.J., Cho, R.J. & Church, G.M. Systematic determination of genetic network architecture. *Nature Genet.* **22**, 281–285 (1999).
14. Chan, G.K., Schaar, B.T. & Yen, T.J. Characterization of the kinetochore binding domain of CENP-E reveals interactions with the kinetochore proteins CENP-F and hBUBR1. *J. Cell Biol.* **143**, 49–63 (1998).
15. Schaar, B.T., Chan, G.K., Maddox, P., Salmon, E.D. & Yen, T.J. CENP-E function at kinetochores is essential for chromosome alignment. *J. Cell Biol.* **139**, 1373–1382 (1997).
16. Liu, K. *et al.* Reversible tumorigenesis induced by deficiency of vasodilator-stimulated phosphoprotein. *Mol. Cell. Biol.* **19**, 3696–3703 (1999).
17. Kosaki, R., Watanabe, K. & Yamaguchi, Y. Overproduction of hyaluronan by expression of the hyaluronan synthase Has2 enhances anchorage-independent growth and tumorigenicity. *Cancer Res.* **59**, 1141–1145 (1999).
18. Entwistle, J., Hall, C.L. & Turley, E.A. HA receptors: regulators of signalling to the cytoskeleton. *J. Cell. Biochem.* **61**, 569–577 (1996).
19. Nguyen, D.H., Hussaini, I.M. & Gonias, S.L. Binding of urokinase-type plasminogen activator to its receptor in MCF-7 cells activates extracellular signal-regulated kinase 1 and 2 which is required for increased cellular motility. *J. Biol. Chem.* **273**, 8502–8507 (1998).
20. Feng, H. *et al.* CUL-2 is required for the G1-to-S-phase transition and mitotic chromosome condensation in *Caenorhabditis elegans*. *Nature Cell Biol.* **1**, 486–492 (1999).
21. Wang, X.W. *et al.* GADD45 induction of a G2/M cell cycle checkpoint. *Proc. Natl. Acad. Sci. USA* **96**, 3706–3711 (1999).
22. Harkin, D.P. *et al.* Induction of GADD45 and JNK/SAPK-dependent apoptosis following inducible expression of BRCA1. *Cell* **97**, 575–586 (1999).
23. Moorman, J.P., Bobak, D.A. & Hahn, C.S. Inactivation of the small GTP binding protein Rho induces multinucleate cell formation and apoptosis in murine T lymphoma EL4. *J. Immunol.* **156**, 4146–4153 (1996).
24. O'Connell, C.B., Wheatley, S.P., Ahmed, S. & Wang, Y.L. The small GTP-binding protein rho regulates cortical activities in cultured cells during division. *J. Cell Biol.* **144**, 305–313 (1999).
25. Madaule, P. *et al.* Role of citron kinase as a target of the small GTPase Rho in cytokinesis. *Nature* **394**, 491–494 (1998).
26. Guasch, R.M., Scambler, P., Jones, G.E. & Ridley, A.J. RhoE regulates actin cytoskeleton organization and cell migration. *Mol. Cell. Biol.* **18**, 4761–4771 (1998).
27. Vallenius, T., Luukko, K. & Makela, T.P. CLP-36 PDZ-LIM protein associates with nonmuscle α-actinin-1 and α-actinin-4. *J. Biol. Chem.* **275**, 11100–11105 (2000).
28. Sato, N., Funayama, N., Nagafuchi, A., Yonemura, S. & Tsukita, S. A gene family consisting of ezrin, radixin and moesin. Its specific localization at actin filament/plasma membrane association sites. *J. Cell Sci.* **103**, 131–143 (1992).
29. Radhakrishna, H., Al-Awar, O., Khachikian, Z. & Donaldson, J.G. ARF6 requirement for Rac ruffling suggests a role for membrane trafficking in cortical actin rearrangements. *J. Cell Sci.* **112**, 855–866 (1999).
30. Kondo, A. *et al.* A new paxillin-binding protein, PAG3/Papα/KIAA0400, bearing an ADP-ribosylation factor GTPase-activating protein activity, is involved in paxillin recruitment to focal adhesions and cell migration. *Mol. Biol. Cell* **11**, 1315–1327 (2000).
31. Law, S.F., O'Neill, G.M., Fashena, S.J., Einarson, M.B. & Golemis, E.A. The docking protein HEF1 is an apoptotic mediator at focal adhesion sites. *Mol. Cell. Biol.* **20**, 5184–5195 (2000).
32. Leeuwen, F.N. *et al.* The guanine nucleotide exchange factor Tiam1 affects neuronal morphology; opposing roles for the small GTPases Rac and Rho. *J. Cell Biol.* **139**, 797–807 (1997).
33. Arai, M. & Kwiatkowski, D.J. Differential developmentally regulated expression of gelsolin family members in the mouse. *Dev. Dyn.* **215**, 297–307 (1999).
34. North, A.J. *et al.* Molecular map of the desmosomal plaque. *J. Cell Sci.* **112**, 4325–4336 (1999).
35. Assmann, V., Jenkinson, D., Marshall, J.F. & Hart, I.R. The intracellular hyaluronan receptor RHAMM/IHABP interacts with microtubules and actin filaments. *J. Cell Sci.* **112**, 3943–3954 (1999).
36. Bachmann, C., Fischer, L., Walter, U. & Reinhard, M. The EVH2 domain of the vasodilator-stimulated phosphoprotein mediates tetramerization, F-actin binding, and actin bundle formation. *J. Biol. Chem.* **274**, 23549–23557 (1999).
37. Wei, M.C. *et al.* tBID, a membrane-targeted death ligand, oligomerizes BAK to release cytochrome c. *Genes Dev.* **14**, 2060–2071 (2000).
38. Yamamoto, K., Ichijo, H. & Korsmeyer, S.J. BCL-2 is phosphorylated and inactivated by an ASK1/Jun N-terminal protein kinase pathway normally activated at G(2)/M. *Mol. Cell. Biol.* **19**, 8469–8478 (1999).
39. Phillips, A.C., Ernst, M.K., Bates, S., Rice, N.R. & Vousden, K.H. E2F-1 potentiates cell death by blocking antiapoptotic signaling pathways. *Mol. Cell* **4**, 771–781 (1999).
40. Hakem, A., Sasaki, T., Kozieradzki, I. & Penninger, J.M. The cyclin-dependent kinase Cdk2 regulates thymocyte apoptosis. *J. Exp. Med.* **189**, 957–968 (1999).
41. Gil-Gomez, G., Berns, A. & Brady, H.J. A link between cell cycle and cell death: Bax and Bcl-2 modulate Cdk2 activation during thymocyte apoptosis. *EMBO J.* **17**, 7209–7218 (1998).
42. Lind, E.F. *et al.* Bcl-2-induced changes in E2F regulatory complexes reveal the potential for integrated cell cycle and cell death functions. *J. Immunol.* **162**, 5374–5379 (1999).
43. Ross-Macdonald, P. *et al.* Large-scale analysis of the yeast genome by transposon tagging and gene disruption. *Nature* **402**, 413–418 (1999).
44. McNally, F.J. & Thomas, S. Katanin is responsible for the M-phase microtubule-severing activity in Xenopus eggs. *Mol. Biol. Cell* **9**, 1847–1861 (1998).
45. Hartman, J.J. *et al.* Katanin, a microtubule-severing protein, is a novel AAA ATPase that targets to the centrosome using a WD40-containing subunit. *Cell* **93**, 277–287 (1998).
46. Farrance, I.K. & Ordahl, C.P. The role of transcription enhancer factor-1 (TEF-1) related proteins in the formation of M-CAT binding complexes in muscle and non-muscle tissues. *J. Biol. Chem.* **271**, 8266–8274 (1996).
47. Stewart, A.F., Suzow, J., Kubota, T., Ueyama, T. & Chen, H.H. Transcription factor RTEF-1 mediates α1-adrenergic reactivation of the fetal gene program in cardiac myocytes. *Circ. Res.* **83**, 43–49 (1998).
48. Lee, S.B. *et al.* The Wilms tumor suppressor WT1 encodes a transcriptional activator of amphiregulin. *Cell* **98**, 663–673 (1999).
49. White, K.P., Rifkin, S.A., Hurban, P. & Hogness, D.S. Microarray analysis of Drosophila development during metamorphosis. *Science* **286**, 2179–2184 (1999).
50. Martin, L.G., Demers, G.W. & Galloway, D.A. Disruption of the G1/S transition in human papillomavirus type 16 E7-expressing human cells is associated with altered regulation of cyclin E. *J. Virol.* **72**, 975–985 (1998).
51. Tomayo, P. *et al.* Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc. Natl. Acad. Sci. USA* **96**, 2907–2912 (1999).