

Drug Target Discovery by Gene Expression Analysis: Cell Cycle Genes

M. G. Walker*

Incyte Genomics and Department of Medicine, Stanford University, 1475 Flamingo Way, Sunnyvale, CA, USA

Abstract: Gene expression microarrays and gene expression databases provide new opportunities for the discovery of drug targets and for determination of a drug's mode of action. We review gene expression analysis methods and describe studies that have identified cell cycle genes using differential expression analysis and co-expression analysis. We present an example of the identification of previously-unrecognized human cell cycle genes, CDCA1 through CDCA8, that are co-expressed with known cell cycle genes including CDC2, CDC7, CDC23, cyclin, MCAK, mki67a, topoisomerase II, and others.

GENE EXPRESSION DATABASES AND ANALYSIS TOOLS

Gene expression microarrays and gene expression databases are widely available. These databases provide new opportunities for the discovery of drug targets and for determination of a drug's mode of action. However, we need computational tools to extract this information. The two most widely-used expression analyses methods are differential expression and co-expression. We will examine applications of these methods to find genes involved in the cell cycle.

DIFFERENTIAL EXPRESSION

Differential expression is based on a comparison of gene expression levels before and after a treatment or on differences in expression between different cell types. A common approach is to apply a compound known to induce or halt cell cycle to a cell culture and then to observe what genes change expression. The genes that change expression significantly can indicate the mode of action of a compound and in some cases we can identify genes not previously recognized as participants in the cell cycle. Several recent publications, which we examine next, describe applications of this method to find or characterize cell cycle genes.

Coller and colleagues studied the effects of c-MYC on gene expression in fibroblasts [1]. MYC affects normal and neoplastic cell proliferation, but the mechanism is not understood. They studied the effects of c-MYC activation in primary human fibroblasts on 6,416 genes, and found it affected the expression of 38 genes previously known to participate in cell growth, cell cycle, adhesion, and cytoskeletal organization.

Shaffer and colleagues studied changes in gene expression by BCL-6, a transcriptional repressor involved in B cell differentiation and inflammation [2]. They found that inhibition of BCL-6 decreased expression of c-MYC and increased expression of the cell cycle inhibitor p27kip1.

Chang and Laimins studied gene expression changes induced by human papillomaviruses, which infect keratinocytes and induce proliferation [3]. They compared the expression of 7,075 known genes and EST's in papillomavirus type 31 (HPV31) cells compared to normal human keratinocytes, and found that the downregulated genes included several involved in the regulation of cell growth and several whose expression increases in response to interferon.

Mariadason and colleagues studied the effects of butyrate on gene expression in cell cycle arrest, differentiation, and apoptosis [4]. Butyrate is a regulator of colonic epithelial cell maturation. They treated SW620 colonic epithelial cells with

*Address correspondence to this author at the 1475 Flamingo Way, Sunnyvale, CA, USA, 94087-3405; e-mail: mwalker@stanfordalumni.org; Phone: 408 736-5506

butyrate and/or trichostatin A, sulindac, and curcumin, and examined changes in expression of 8,063 gene sequences. They found that the expression pattern of butyrate-treated cells was most similar to that of a Caco-2 cell line that had spontaneously undergone a G0-G1 arrest and least similar to the G2-M arrest induced by curcumin.

Han and colleagues studied the potential oncogene hepatitis B viral protein (HBx) [5]. They compared the expression levels of 588 genes in a HepG2 cell line expressing HBx versus untransformed control cells. Their study identified a variety of genes with differential expression, among which was p53cdc.

Simbulan-Rosenthal and colleagues studied a gene implicated in maintenance of genomic integrity, Poly(ADP-ribose) polymerase (PARP) [6]. They compared the expression levels of 11,000 genes in PARP(-/-) mice versus their wild-type littermates. Their study showed that loss of PARP downregulates several genes involved in cell cycle progression or mitosis, among other functions.

Differential expression analysis requires treatments that alter the cell cycle, or cells that a priori differ in expression of cell cycle genes. Its use is usually restricted to cell lines or model organisms such as yeast.

CO-EXPRESSION

In co-expression analysis we look for previously-uncharacterized genes that mimic the expression patterns of known cell-cycle genes (with or without the types of treatment used in differential expression). The assumption in co-expression is that if the expression of one gene is very similar to the expression of another gene, then it is likely that they are related in their function. Highly similar expression, like highly similar sequence, suggests similar function. In the literature, this approach is called correlation analysis or co-expression analysis.

Spellman and colleagues examined changes in yeast gene expression after synchronizing cultures using alpha factor arrest, elutriation, or arrest of a cdc15 temperature sensitive mutant. They followed changes in gene expression at several time

points, and identified those genes whose expression patterns were most similar to the patterns of genes known to be expressed at G1, S, G2, M, and M/G1, using standard Pearson linear correlation. They defined a threshold score that was exceeded by 91% of known cell-cycle regulated genes, and found 800 genes that met or exceeded that score. These researchers also examined the promoter regions of co-expressed genes (as identified by cluster analysis), looking for conserved regulatory motifs. They found known cell-cycle regulator binding sites and novel candidate regulatory motifs. Other groups have performed similar analyses on the same data using different statistical methods, and have identified additional candidate regulatory motifs [7, 8].

Coller and colleagues in the research cited previously examined genes co-expressed with c-MYC, and found significant overlap with the genes identified by differential expression, but with differences between the methods [1]. While Pearson linear correlation and Spearman rank correlation are often successful, we have found cases in which they perform poorly. When applied to gene expression data from the Inctye LifeSeq database, these methods identified some known relationships, particularly among stoichiometrically-related proteins, such as components of immunoglobulins or ribosomes, but often failed to detect known relationships. We speculate that the reason for their frequent failure is that many, and perhaps most, genes do not show the linear or monotonic correlation assumed by these methods, but rather have more complex (non-linear) associations and respond to multiple interactions among regulatory proteins. In addition, these correlation methods are relatively sensitive to errors in measurement; expression measurements are not as reproducible as we might wish, and may contribute to their relative difficulty. Finally, most genes are not expressed in most tissues; when two genes have no measured expression in hundreds of libraries, and non-zero values in a handful of libraries, the genes can show high correlation regardless of the lack of true biological association. One then faces the problem of distinguishing spurious correlation from biologically meaningful correlation. In the next section, we present an alternative method that is complementary to linear and rank correlation methods.

GUILTY-BY-ASSOCIATION (GBA) METHOD FOR CO-EXPRESSION ANALYSIS

A common statistical approach to avoiding the assumptions and weaknesses of linear statistics is to use a non-parametric statistical analysis. One non-parametric method of co-expression analysis is Guilt-by-Association (GBA); we have used GBA previously to identify genes that are co-expressed with prostate cancer genes [9] and with neurotransmitter genes [10]. We describe here the GBA method of co-expression analysis and give an example of its use to identify previously-unrecognized human cell cycle genes. For GBA analysis, we consider a gene to be present (expressed) in a library if cDNA corresponding to that gene is detected in the sample taken from that library. We consider a gene to be absent (not expressed) in a library when no cDNA for that gene is detected in the library. Table 1 shows an example of the occurrences of two hypothetical genes, A and B, in 30 cDNA libraries. A “1” indicates that the gene was detected in the library; a “0” indicates that it was not detected. For a given pair of genes, the expression data in Table 1 can be summarized in a two-by-two contingency table. Table 2 presents such a co-expression contingency table for the hypothetical genes A and B in 30 libraries; Table 3 presents the same data as variables that we will use shortly.

Table 1. Expression of Two Hypothetical Genes A and B

A hypothetical example of the occurrences of genes A and B in 30 libraries. A “1” indicates that the gene was detected in the library; a “0” indicates that it was not detected

Gene	Library 1	Library 2	...	Library 30
A	1	1	...	0
B	1	0	...	0

We determine the probability that the co-expression shown in Table 2 occurs by chance using a Fisher Exact test [11]. The chi-square test, which is an approximation to the Fisher Exact test, is commonly used to analyze contingency tables such as Table 2. However, the chi-squared approximation is unsuitable when the expected number of counts in any cell in the table is small (less than 10). For many genes in our database, the expected number in at least one cell is less than 10;

hence we use the more computationally expensive but more accurate Fisher Exact test.

Table 2. Contingency Table Summary of co-Expression of Genes A and B
This contingency table summarizes the occurrence data from Table 1 as counts of the number of libraries in which the two genes are present or absent

Number of libraries	Gene A present	Gene A absent	Total
Gene B present	8	2	10
Gene B Absent	2	18	20
Total	10	20	30

In the Fisher Exact test, we take as our null hypothesis that there is no association between gene A and gene B. Under the null hypothesis, the marginal counts in Tables 2 and 3 are fixed, the expected count in each cell is a function of the marginals, and deviations from the expected count are random. The number of ways that k occurrences of a gene can be distributed in r libraries is $\binom{r}{k}$, that is, the combinatoric choose function. From Table 3, we can calculate the probability of observing n11 counts count in the cell {Gene A present and Gene B present} using the hypergeometric distribution. From the hypergeometric distribution, the probability of observing exactly n11 counts is $p(n11) = \frac{\binom{n1}{n11} \times \binom{n2}{n21}}{\binom{n.}{n.1}}$.

Table 3. Variables Representing Counts of Gene Occurrences

In this table, we represent the counts from Table 2 as variables to be used in the Fisher Exact test to calculate the probability that co-expression is due to chance

Number of libraries	Gene A present	Gene A absent	Total
Gene B present	n11	n12	n1.
Gene B absent	n21	n22	n2.
Total	n.1	n.2	n.?

To determine if there is association (lack of independence) between the genes, we calculate the sum of all the (hypergeometric) probabilities for

outcomes at least as extreme as the observed outcome. As a concrete example, consider in Table 2 the n_{11} count of 8 co-occurrences of gene A and gene B. We calculate the probability of observing a count of exactly 8 using the hypergeometric distribution, that is, $p(n_{11} \text{ is } 8) = (10 \text{ C } 8) \times (20 \text{ C } 2) / (30 \text{ C } 10)$. To test the null hypothesis, we are interested not only in the case in which we observe a count of exactly 8 in the cell, but also the cases in which we observe more extreme values of n_{11} , subject to the constraints of the marginals. Hence, we sum the probability of the observed count and of the more extreme possible counts ($n_{11} = 8, 9,$ and 10) to determine the total probability of counts at least as extreme as those observed. In the case of Table 2, the probability that the observed co-expression is due to chance is $p = 0.0003$.

HUMAN CELL CYCLE GENES IDENTIFIED BY GBA

To find cell-cycle gene using GBA, we examined the expression of genes in 1176 human cDNA libraries. These libraries were from diverse anatomic and pathologic states, mainly from surgery, biopsy, or post-mortem samples, or were prepared from cell lines, and include all the libraries that were in the LifeSeq database at the time of the analysis. Approximately 5000 cDNA's from each library were sequenced by gel electrophoresis, assembled, and aligned against known genes. Sequences that were significantly different from known gene sequences were assigned new identification numbers. All genes that were detected in at least five of the 1176 libraries were included in the analysis described here, which yielded 37,071 known and novel genes, gene fragments, or splice variants.

In this dataset, we observed the co-expression of many known cell cycle genes which are listed in Table 4. Table 5 shows the co-expression of these known genes with each other. Other known cell cycle-associated genes were also co-expressed with this set, but not so closely. In addition, we observed eight previously-uncharacterized genes (CDCA1, CDCA2, CDCA3, CDCA4, CDCA5, CDCA6, CDCA7, CDCA8) that are the most closely co-expressed with these known cell-cycle genes. The co-expression of the eight novel genes with the known cell cycle genes is shown in Table 6. Each of the eight novel genes is co-expressed

with one or more of the known genes with a p -value of 1.0×10^{-8} or better, and has p -values for association comparable to the p -values among the known genes. The eight genes all show regions of sequence similarity to uncharacterized EST and genomic sequences in Genbank, but do not show significant similarity to genes with known function.

Several aspects of the library selection and preparation in this GBA study are likely to affect the results of the analyses, because they violate one or more assumptions. Specifically, because more than one library may be obtained from a single patient (for example, multiple organs, or matched tumor and non-tumor tissue, or normalized and non-normalized), libraries are not completely independent. Libraries were, in many cases, normalized or subtracted to enrich the proportion of genes expressed at low levels; this alteration will make more difficult the detection of associations between genes expressed at different levels. Some libraries in which many novel genes were found were sampled to greater depth, leading to an inconsistent measure of presence or absence. Because of the random sampling, genes that are expressed at low levels may not be detected. The cDNA libraries used in this analysis were prepared at different times and with different methods. The effect of different cDNA library samples, different normalization, different preparation methods, or preparation at different times is most likely to be to obscure true relationships. Such differences will make the calculated probability of association less accurate. However, it is unlikely that a pattern that is consistent across 1176 libraries, has good p -values, and is consistent with known biological relationships would be introduced by the random effects of such differences. With a Bonferroni correction for multiple comparisons, the genes identified here show significant association with many, but not all, of the known cell cycle-associated genes.

GBA is better able to detect nonlinear relationships and is more robust against errors in expression measurement than is linear correlation. However, for genes whose expression is linearly or monotonically associated with other genes, correlation analysis may provide good results, possibly with smaller sample sizes than are required for GBA.

Table 4. Co-Expressed Cell-cycle Genes

These cell cycle genes are co-expressed with each other with extreme *p*-values (Fisher Exact test) in 1176 cDNA libraries

Gene name	Description and Citations
CDC2	CDC2 Cell division cycle protein 2 / Cyclin B1 Activation of the mitotic kinase CDC2 triggers entry into mitosis. CDC2 binds chromatin prior to S-phase, and is displaced during DNA replication [12, 13].
CDC7	CDC7 Cell division cycle protein 7 CDC7 kinase is conserved in eukaryotes from yeast to humans. It is essential for initiation of DNA replication and entry into S-phase[14-16].
CDC23	CDC23 Cell division cycle protein 23 CDC23 is a component of the anaphase-promoting complex (APC) that regulates mitosis by catalyzing the formation of cyclin B-ubiquitin conjugates, targeting cyclin B for degradation [17-19].
Cyclin B	Cyclin B Cyclin B is a subunit of the cyclin-dependent kinase 1 (Cdk1). Degradation of cyclin B by the anaphase-promoting complex (APC) is required for inactivation of the kinase and exit from mitosis. Cyclin-dependent kinases (CDKs) are regulators of cell cycle progression, and alterations and deregulation of CDK activity are characteristic of neoplasia. CDK inhibitors and modulators alter cell cycle and induce apoptosis and tumor regression [19-21].
hBub1	hBub1 mitotic checkpoint kinase hBub1 is a kinetochore protein that monitors chromosome attachment to the spindle in mitotic cells and controls exit from mitosis and chromosome segregation. The mitotic checkpoint ensure proper chromosome segregation by delaying anaphase until chromosomes are aligned on the spindle. Following spindle damage, cells exit mitosis and undergo apoptosis. hBub1 is required for the checkpoint response to spindle damage; mutations in hBub1 disrupt the mitotic checkpoint allowing cells to escape apoptosis and continue cell cycle progression, despite spindle damage, potentially leading to aneuploidy and contributing to neoplasia [22-27].
hKSP	hKSP kinesin-like spindle protein (alternate name HsEg5) A spindle-associated protein found with centrosomal microtubules during prophase and prometaphase centrosome separation, and associated with post-mitotic centrosome movement[28].
hp55cdc	hp55cdc cell division cycle protein / CDC20 p55cdc is a kinetochore and spindle microtubule-associated protein that mediates association of the spindle checkpoint protein Mad2 with the cyclosome/anaphase promoting complex, and appears to be essential for cell division. Over expression of p55cdc induces apoptosis. hp55cdc is also associated with the mitotic spindle protein kinase Aik (q.v.) [29-34].
MCAK	MCAK mitotic centromere-associated kinesin MCAK is a microtubule motor protein recruited to the centromere at prophase that participates in anaphase chromosome segregation[33, 35-38].
mitosin	mitosin (CENP-F kinetochore protein) Mitosin is a nuclear protein that associates with centromeres and spindle poles during M phase. Overexpression of N-terminally truncated mitosin blocks cell cycle progression. Mitosin is correlated with clinical outcome in node-negative breast cancer [39-41].
mki67a	mki67a (MIB-1) cell proliferation marker Expression of mKi67a is strictly associated with cell proliferation. It is widely use in pathology as a cell proliferation marker to measure the growth fraction of cells in human tumors [42-45].
MKLP-1	MKLP1 mitotic kinesin-like protein 1 (CHO1) MKLP1 is a spindle-associated protein required for mitotic progression [46-48].
myb	b-myb B-myb is a member of the myb family of cell-cycle regulated transcription factors, expressed in G1 and S phase. Activity of b-myb is stimulated by cyclin A/Cdk2-dependent phosphorylation [49-52].

(Table 4). contd.....

Gene name	Description and Citations
NLK1	NLK1 NIMA-like protein kinase 1 NLK1 is a human mitotic kinase, similar to the NIMA cell-cycle regulatory protein kinase in Aspergillus that is essential for entry into and progression through mitosis[53-55].
P1-CDC21	P1-CDC21 Member of the family of minichromosome maintenance proteins essential for DNA replication [56, 57].
PRC1	PRC1 protein regulating cytokinesis 1 PRC1 is a human mitotic-spindle associated CDK substrate protein required for cytokinesis [58].
prkAik2	protein kinase Aik2 / Aurora2 Localized to mitotic spindle poles; involved in regulating chromosome segregation and maintaining genomic stability. Associated with p55cdc/cdc20 [34, 59, 60].
survivin	survivin apoptosis inhibitor gene Survivin is expressed in the G2/M phase of the cell cycle. At the beginning of mitosis it associates with microtubules of the mitotic spindle. It inhibits apoptosis, and is hypothesized to overcome the apoptotic checkpoint, allowing cancer cells to survive [61, 62].
topo II	topoisomerase II Topo II is required for chromosome condensation and segregation during DNA replication. It's expression is cell cycle dependent, with protein levels and catalytic activity peaking in G2/M. It is hypothesized to be part of regulatory checkpoints at the entry and progression of mitosis and thus to regulate apoptosis. Topo II poisons induce carcinogenic chromosomal alterations [63-68].
UbcH10	UbcH10 cyclin-selective ubiquitin carrier protein Destruction of mitotic cyclins by ubiquitin-mediated proteolysis is required for cells to complete mitosis and enter anaphase of the next cell cycle. This process is catalyzed by UbcH10/E2-C. Mutant UbcH10 inhibits the destruction of cyclins, arrests cells in M phase, and inhibits the onset of anaphase [69, 70].

Table 5. Co-expression of Known Cell-cycle Genes

This table shows the probability, for each pair of genes, that their observed co-expression in 1176 cDNA libraries is due to chance. The *p*-values are expressed as the negative log (- log *p*) from the Fisher Exact Test.

	CDC2	CDC7	CDC23	Cyclin B	cyclin	hBub1	HKSP	hp55cdc	MCAK	mitosin	mki67a	MKLP-1	NLK1	CDC21	PRC1	Aik2	survivin	topo II	UbcH10
CDC2		13	3	12	8	6	6	5	10	12	7	0	0	11	7	7	11	23	6
CDC7	13		6	7	8	6	5	7	5	6	5	6	4	7	5	8	7	11	6
CDC23	3	6		10	0	5	6	9	7	8	5	8	0	7	6	3	9	13	4
Cyclin B	12	7	10		5	10	12	14	14	7	4	13	0	10	15	11	11	17	12
cyclin	8	8	0	5		5	6	6	12	7	7	6	10	6	10	5	9	11	8
hBub1	6	6	5	10	5		7	8	7	6	5	6	4	6	12	7	9	15	12
HKSP	6	5	6	12	6	7		11	13	8	9	19	4	14	16	9	18	19	15
hp55cdc	5	7	9	14	6	8	11		12	11	4	18	5	9	12	7	11	14	9
MCAK	10	5	7	14	12	7	13	12		13	7	13	8	12	13	8	18	24	12
mitosin	12	6	8	7	7	6	8	11	13		7	20	4	21	13	11	9	15	12
mki67a	7	5	5	4	7	5	9	4	7	7		7	4	6	5	5	10	6	6
MKLP-1	0	6	8	13	6	6	19	18	13	20	7		0	10	9	8	11	18	16
NLK1	0	4	0	0	10	4	4	5	8	4	4	0		6	7	4	6	8	6
CDC21	11	7	7	10	6	6	14	9	12	21	6	10	6		15	7	11	19	9
PRC1	7	5	6	15	10	12	16	12	13	13	5	9	7	15		10	8	18	8
Aik2	7	8	3	11	5	7	9	7	8	11	5	8	4	7	10		10	11	15
survivin	11	7	9	11	9	9	18	11	18	9	10	11	6	11	8	10		23	16
topo II	23	11	13	17	11	15	19	14	24	15	6	18	8	19	18	11	23		11
UbcH10	6	6	4	12	8	12	15	9	12	12	6	16	6	9	8	15	16	11	

Table 6. Co-expression of CDCA1 to CDCA8 Genes with Known Cell-cycle Genes

This table shows the probability, for each known gene versus each novel gene, that their observed co-expression in 1176 cDNA libraries is due to chance. The p -values are expressed as the negative log ($-\log p$) from the Fisher Exact test.

Gene name	CDCA1	CDCA2	CDCA3	CDCA4	CDCA5	CDCA6	CDCA7	CDCA8
CDC2	8	7	6	5	7	7	9	0
CDC7	4	5	12	5	7	0	10	3
CDC23	6	5	5	5	10	6	4	10
Cyclin B	5	8	10	13	7	10	7	8
cyclin	9	7	5	5	6	0	8	4
hBub1	7	8	5	5	6	4	9	5
HKSP	6	9	8	8	12	8	9	17
hp55cdc	8	7	10	10	11	8	5	10
MCAK	8	5	7	11	9	5	9	8
mitosin	9	7	10	8	12	5	9	4
mki67a	8	9	3	4	0	4	6	4
MKLP-1	7	4	10	7	11	7	10	0
NLK1	9	4	0	4	4	0	5	4
CDC21	9	10	11	6	15	4	9	8
PRC1	9	12	9	6	16	6	10	8
Aik2	6	7	11	11	8	5	8	6
survivin	6	5	8	8	7	8	9	9
topo II	16	11	11	7	14	9	13	15
UbcH10	10	7	11	13	11	6	12	6

DISCUSSION

A difficulty in expression studies is that a given stimulus often has quite different effects in different cell lines. Genes that change expression in one cell line in response to a stimulus are often unchanged in another, closely related cell line, dependent, presumably, on the differences in receptors, transcription factors, and other proteins present in each cell type. Thus, it is difficult to determine how much of a regulatory pathway we may have found, and what the exact, quantitative connections are among the genes. In the current state of the art, it is almost impossible to predict what effects a compound will have on gene expression in a previously-untested cell type. Pursuit of a better understanding of alterations of gene expression and the connections among regulatory genes is likely to remain an active research area for some time. However, the methods available today are sufficient to identify many

genes that participate in the cell cycle, and thus can provide candidate targets for new cancer drugs.

ACKNOWLEDGEMENTS

Colleagues at Incyte Genomics contributed in many ways to the research reported here, and their help is gratefully acknowledged.

ACCESSION NUMBERS

Gene name	Genbank Accession #	DbEST identifier
CDCA1	BG354574	8402074
CDCA2	BG354575	8402075
CDCA3	BG354576	8402076
CDCA4	BG354577	8402077
CDCA5	BG354578	8402078
CDCA6	BG354579	8402079
CDCA7	BG354580	8402080
CDCA8	BG354581	8402081

REFERENCES

- [1] Collier, H. A.; Grandori, C.; Tamayo, P.; Colbert, T.; Lander, E. S.; Eisenman, R. N.; Golub, T. R. Expression Analysis with Oligonucleotide Microarrays Reveals that MYC Regulates Genes Involved in Growth, Cell Cycle, Signaling, and Adhesion. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 3260-3265.
- [2] Shaffer, A. L.; Yu, X.; He, Y.; Boldrick, J.; Chan, E. P.; Staudt, L. M. BCL-6 Represses Genes that Function in Lymphocyte Differentiation, Inflammation, and Cell Cycle Control. *Immunity* **2000**, *13*, 199-212.
- [3] Chang, Y. E.; Laimins, L. A. Microarray Analysis Identifies Interferon-Inducible Genes and Stat-1 as Major Transcriptional Targets of Human Papillomavirus Type 31. *J. Virol.* **2000**, *74*, 4174-4182.
- [4] Mariadason, J. M.; Corner, G. A.; Augenlicht, L. H. Genetic Reprogramming in Pathways of Colonic Cell Maturation Induced by Short Chain Fatty Acids: Comparison with Trichostatin A, Sulindac, and Curcumin and Implications for Chemoprevention of Colon Cancer. *Cancer Res.* **2000**, *60*, 4561-4572.
- [5] Han, J.; Yoo, H. Y.; Choi, B. H.; Rho, H. M. Selective Transcriptional Regulations in the Human Liver Cell by Hepatitis B Viral X Protein. *Biochem. Biophys. Res. Commun.* **2000**, *272*, 525-530.
- [6] Simbulan-Rosenthal, C. M.; Ly, D. H.; Rosenthal, D. S.; Konopka, G.; Luo, R.; Wang, Z. Q.; Schultz, P. G.; Smulson, M. E. Misregulation of Gene Expression in Primary Fibroblasts Lacking Poly(ADP-ribose) Polymerase. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 11274-11279.
- [7] Zhang, M. Q. Promoter Analysis of Co-Regulated Genes in the Yeast Genome. *Comput. Chem.* **1999**, *23*, 233-250.
- [8] Wolfsberg, T. G.; Gabrielian, A. E.; Campbell, M. J.; Cho, R. J.; Spouge, J. L.; Landsman, D. Candidate Regulatory Sequence Elements for Cell Cycle-Dependent Transcription in *Saccharomyces cerevisiae*. *Genome Res.* **1999**, *9*, 775-792.
- [9] Walker, M. G.; Volkmuth, W.; Sprinzak, E.; Hodgson, D.; Klingler, T. Prediction of Gene Function by Genome-Scale Expression Analysis: Prostate Cancer-Associated Genes. *Genome Res.* **1999**, *9*, 1198-1203.
- [10] Walker, M. G.; Volkmuth, W.; Klingler, T. M. Pharmaceutical Target Discovery using Guilt-by-Association: Schizophrenia and Parkinson's Disease Genes. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **1999**, *146*, 282-286.
- [11] Agresti, A. *Categorical Data Analysis*. **1990**, New York: Wiley.
- [12] De Souza, C. P.; Ellem, K. A.; Gabrielli, B. G. Centrosomal and Cytoplasmic Cdc2/cyclin B1 Activation Precedes Nuclear Mitotic Events. *Exp. Cell Res.* **2000**, *257*, 11-21.
- [13] Krude, T.; Musahl, C.; Laskey, R. A.; Knippers, R. Human Replication Proteins hCdc21, hCdc46 and P1Mcm3 Bind Chromatin Uniformly before S-Phase and are Displaced Locally during DNA Replication. *J. Cell Sci.* **1996**, *109*, 309-318.
- [14] Jiang, W.; McDonald, D.; Hope, T. J.; Hunter, T. Mammalian Cdc7-Dbf4 Protein Kinase Complex is Essential for Initiation of DNA Replication. *EMBO J.* **1999**, *18*, 5703-5713.
- [15] Masai, H.; Sato, N.; Takeda, T., and Arai, K. CDC7 Kinase Complex as a Molecular Switch for DNA Replication. *Front. Biosci.* **1999**, *4*, D834-D840.
- [16] Donaldson, A. D.; Fangman, W. L.; Brewer, B. J. Cdc7 is Required Throughout the Yeast S Phase to Activate Replication Origins. *Genes Dev.* **1998**, *12*, 491-501.
- [17] Prinz, S.; Hwang, E. S.; Visintin, R., and Amon, A. The regulation of Cdc20 proteolysis reveals a role for APC components Cdc23 and Cdc27 during S phase and early mitosis. *Curr. Biol.* **1998**, *8*, 750-760.
- [18] Zhao, N.; Lai, F.; Fernald, A. A.; Eisenbart, J. D.; Espinosa, R.; Wang, P. W.; Le Beau, M. M. Human CDC23: cDNA Cloning, Mapping to 5q31, Genomic Structure, and Evaluation as a Candidate Tumor Suppressor Gene in Myeloid Leukemias. *Genomics* **1998**, *53*, 184-190.
- [19] Hershko, A. Mechanisms and Regulation of the Degradation of Cyclin B. *Philos. Trans. R. Soc. Lond B Biol. Sci.* **1999**, *354*, 1571-1575; discussion 1575-1576.
- [20] Hajdуч, M.; Havlieek, L.; Vesely, J.; Novotny, R.; Mihal, V.; Strnad, M. Synthetic Cyclin Dependent Kinase Inhibitors. New Generation of Potent Anti-Cancer Drugs. *Adv. Exp. Med. Biol.* **1999**, *457*, 341-353.
- [21] Sausville, E. A.; Zaharevitz, D.; Gussio, R.; Meijer, L.; Louarn-Leost, M.; Kunick, C.; Schultz, R.; Lahusen, T.; Headlee, D.; Stinson, S.; Arbuck, S. G.; Senderowicz, A. Cyclin-Dependent Kinases: Initial Approaches to Exploit a Novel Therapeutic Target. *Pharmacol. Ther.* **1999**, *82*, 285-292.
- [22] Seeley, T. W.; Wang, L.; Zhen, J. Y. Phosphorylation of Human MAD1 by the BUB1 Kinase In Vitro. *Biochem. Biophys. Res. Commun.* **1999**, *257*, 589-595.

- [23] Ouyang, B.; Lan, Z.; Meadows, J.; Pan, H.; Fukasawa, K.; Li, W.; Dai, W. Human Bub1: A Putative Spindle Checkpoint Kinase Closely Linked to Cell Proliferation. *Cell Growth Differ.* **1998**, *9*, 877-885.
- [24] Taylor, S. S.; McKeon, F. Kinetochores: Localization of Murine Bub1 is Required for Normal Mitotic Timing and Checkpoint Response to Spindle Damage. *Cell* **1997**, *89*, 727-735.
- [25] Myrie, K. A.; Percy, M. J.; Azim, J. N.; Neeley, C. K.; Petty, E. M. Mutation and Expression Analysis of Human BUB1 and BUB1B in Aneuploid Breast Cancer Cell Lines. *Cancer Lett.* **2000**, *152*, 193-199.
- [26] Imai, Y.; Shiratori, Y.; Kato, N.; Inoue, T.; Omata, M. Mutational Inactivation of Mitotic Checkpoint Genes, hsMAD2 and hBUB1, is Rare in Sporadic Digestive Tract Cancers. *Jpn. J. Cancer Res.* **1999**, *90*, 837-840.
- [27] Cahill, D. P.; Lengauer, C.; Yu, J.; Riggins, G. J.; Willson, J. K.; Markowitz, S. D.; Kinzler, K. W.; Vogelstein, B. Mutations of Mitotic Checkpoint Genes in Human Cancers. *Nature* **1998**, *392*, 300-303.
- [28] Whitehead, C. M.; Winkfein, R. J.; Rattner, J. B. The Relationship of HsEg5 and the Actin Cytoskeleton to Centrosome Separation. *Cell Motil. Cytoskeleton* **1996**, *35*, 298-308.
- [29] Kao, C. T.; Lin, M.; O'Shea-Greenfield, A.; Weinstein, J., and Sakamoto, K. M. Over-expression of p53Cdc Inhibits Granulocyte Differentiation and Accelerates Apoptosis in Myeloid Cells. *Oncogene* **1996**, *13*, 1221-1229.
- [30] Kramer, E. R.; Gieffers, C.; Holzl, G.; Hengstschlager, M.; Peters, J. M. Activation of the Human Anaphase-Promoting Complex by Proteins of the CDC20/Fizzy Family. *Curr. Biol.* **1998**, *8*, 1207-1210.
- [31] Weinstein, J.; Jacobsen, F. W.; Hsu-Chen, J.; Wu, T.; Baum, L. G. A Novel Mammalian Protein, p53CDC, Present in Dividing Cells is Associated with Protein Kinase Activity and has Homology to the *Saccharomyces cerevisiae* Cell Division Cycle Proteins Cdc20 and Cdc4. *Mol. Cell Biol.* **1994**, *14*, 3350-3363.
- [32] Kallio, M.; Weinstein, J.; Daum, J. R.; Burke, D. J.; Gorbsky, G. J. Mammalian p53CDC Mediates Association of the Spindle Checkpoint Protein Mad2 with the Cyclosome/Anaphase-Promoting Complex, and is Involved in Regulating Anaphase Onset and Late Mitotic Events. *J. Cell Biol.* **1998**, *141*, 1393-406.
- [33] Saffery, R.; Irvine, D. V.; Griffiths, B.; Kalitsis, P.; Wordeman, L.; Choo, K. H. Human Centromeres and Neocentromeres Show Identical Distribution Patterns of >20 Functionally Important Kinetochores-Associated Proteins. *Hum. Mol. Genet.* **2000**, *9*, 175-185.
- [34] Farruggio, D. C.; Townsley, F. M.; Ruderman, J. V. Cdc20 Associates with the Kinase Aurora2/Aik. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 7306-7611.
- [35] Maney, T.; Ginkel, L. M.; Hunter, A. W.; Wordeman, L. The Kinetochores of Higher Eucaryotes: A Molecular View. *Int. Rev. Cytol.* **2000**, *194*, 67-131.
- [36] Wordeman, L.; Wagenbach, M., and Maney, T. Mutations in the ATP-Binding Domain Affect the Subcellular Distribution of Mitotic Centromere-Associated Kinesin (MCAK). *Cell Biol. Int.* **1999**, *23*, 275-286.
- [37] Maney, T.; Hunter, A. W.; Wagenbach, M.; Wordeman, L. Mitotic Centromere-Associated Kinesin is Important for Anaphase Chromosome Segregation. *J. Cell Biol.* **1998**, *142*, 787-801.
- [38] Kim, I. G.; Jun, D. Y.; Sohn, U.; Kim, Y. H. Cloning and Expression of Human Mitotic Centromere-Associated Kinesin Gene. *Biochim. Biophys. Acta* **1997**, *1359*, 181-186.
- [39] Clark, G. M.; Allred, D. C.; Hilsenbeck, S. G.; Chamness, G. C.; Osborne, C. K.; Jones, D.; Lee, W. H. Mitosin (a New Proliferation Marker) Correlates with Clinical Outcome in Node-Negative Breast Cancer. *Cancer Res.* **1997**, *57*, 5505-5508.
- [40] Zhu, X.; Ding, L.; Pei, G. Carboxyl Terminus of Mitosin is Sufficient to Confer Spindle Pole Localization. *J. Cell. Biochem.* **1997**, *66*, 441-449.
- [41] Zhu, X. Structural Requirements and Dynamics of Mitosin-Kinetochores Interaction in M phase. *Mol. Cell. Biol.* **1999**, *19*, 1016-1024.
- [42] Schluter, C.; Duchrow, M.; Wohlenberg, C.; Becker, M. H.; Key, G.; Flad, H. D.; Gerdes, J. The Cell Proliferation-Associated Antigen of Antibody Ki-67: A Very Large, Ubiquitous Nuclear Protein with Numerous Repeated Elements, Representing a New Kind of Cell Cycle-Maintaining Proteins. *J. Cell Biol.* **1993**, *123*, 513-522.
- [43] Duchrow, M.; Schluter, C.; Key, G.; Kubbutat, M. H.; Wohlenberg, C.; Flad, H. D.; Gerdes, J. Cell Proliferation-Associated Nuclear Antigen Defined by Antibody Ki-67: A New Kind of Cell Cycle-Maintaining Proteins. *Arch. Immunol. Ther. Exp.* **1995**, *43*, 117-121.
- [44] Dalquen, P.; Baschiera, B.; Chaffard, R.; Dieterich, H.; Feichter, G. E.; Krmer, K.; Torhorst, J. MIB-1 (Ki-67) Immunostaining of Breast Cancer Cells in Cytologic Smears. *Acta Cytol.* **1997**, *41*, 229-237.

- [45] Scholzen, T.; Gerdes, J. The Ki-67 Protein: from the Known and the Unknown. *J. Cell Physiol.* **2000**, *182*, 311-322.
- [46] Nislow, C.; Lombillo, V. A.; Kuriyama, R.; McIntosh, J. R. A Plus-End-Directed Motor Enzyme that Moves Antiparallel Microtubules in vitro Localizes to the Interzone of Mitotic Spindles. *Nature* **1992**, *359*, 543-547.
- [47] Kobayashi, N.; Reiser, J.; Kriz, W.; Kuriyama, R.; Mundel, P. Nonuniform Microtubular Polarity Established by CHO1/MKLP1 Motor Protein is Necessary for Process Formation of Podocytes. *J. Cell Biol.* **1998**, *143*, 1961-1970.
- [48] Sharp, D. J.; Yu, W.; Ferhat, L.; Kuriyama, R.; Rueger, D. C.; Baas, P. W. Identification of a Microtubule-Associated Motor Protein Essential for Dendritic Differentiation. *J. Cell Biol.* **1997**, *138*, 833-843.
- [49] Robinson, C.; Light, Y.; Groves, R.; Mann, D.; Marias, R.; Watson, R. Cell-Cycle Regulation of B-Myb Protein Expression: Specific Phosphorylation during the S Phase of the Cell Cycle. *Oncogene* **1996**, *12*, 1855-1864.
- [50] Saville, M. K.; Watson, R. J. The Cell-Cycle Regulated Transcription Factor B-Myb is Phosphorylated by Cyclin A/Cdk2 at Sites that Enhance its Transactivation Properties. *Oncogene* **1998**, *17*, 2679-2689.
- [51] Saville, M. K.; Watson, R. J. B-Myb: a Key Regulator of the Cell Cycle. *Adv. Cancer Res.* **1998**, *72*, 109-140.
- [52] Horstmann, S.; Ferrari, S.; Klempnauer, K. H. Regulation of B-Myb Activity by Cyclin D1. *Oncogene* **2000**, *19*, 298-306.
- [53] Lu, K. P.; Hunter, T. The NIMA Kinase: a Mitotic Regulator in *Aspergillus nidulans* and Vertebrate Cells. *Prog. Cell Cycle Res.* **1995**, *1*, 187-205.
- [54] Lu, K. P.; Hunter, T. Evidence for a NIMA-like Mitotic Pathway in Vertebrate Cells. *Cell* **1995**, *81*, 413-424.
- [55] Shen, M.; Haggblom, C.; Vogt, M.; Hunter, T.; Lu, K. P. Characterization and Cell Cycle Regulation of the Related Human Telomeric Proteins Pin2 and TRF1 Suggest a Role in Mitosis. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 13618-13623.
- [56] Ishimi, Y.; Ichinose, S.; Omori, A.; Sato, K.; Kimura, H. Binding of Human Minichromosome Maintenance Proteins with Histone H3. *J. Biol. Chem.* **1996**, *271*, 24115-24122.
- [57] Hu, B.; Burkhardt, R.; Schulte, D.; Musahl, C.; Knippers, R. The P1 Family: A New Class of Nuclear Mammalian Proteins Related to the Yeast Mcm Replication Proteins. *Nucleic Acids Res.* **1993**, *21*, 5289-5293.
- [58] Jiang, W.; Jimenez, G.; Wells, N. J.; Hope, T. J.; Wahl, G. M.; Hunter, T., et al. PRC1: A Human Mitotic Spindle-Associated CDK Substrate Protein Required for Cytokinesis. *Mol. Cell.* **1998**, *2*, 877-885.
- [59] Kimura, M.; Matsuda, Y.; Yoshioka, T.; Sumi, N.; Okano, Y. Identification and Characterization of STK12/Aik2: A Human Gene Related to Aurora of Drosophila and Yeast IPL1. *Cytogenet. Cell. Genet.* **1998**, *82*, 147-152.
- [60] Kimura, M.; Matsuda, Y.; Yoshioka, T.; Okano, Y. Cell Cycle-Dependent Expression and Centrosome Localization of a Third Human Aurora/Ipl1-related Protein kinase, AIK3. *J. Biol. Chem.* **1999**, *274*, 7334-7340.
- [61] Li, F.; Ambrosini, G.; Chu, E. Y.; Plescia, J.; Tognin, S.; Marchisio, P. C., et al. Control of Apoptosis and Mitotic Spindle Checkpoint by Survivin. *Nature* **1998**, *396*, 580-584.
- [62] Verdecia, M. A.; Huang, H.; Dutil, E.; Kaiser, D. A.; Hunter, T.; Noel, J. P. Structure of the Human Anti-apoptotic Protein Survivin Reveals a Dimeric Arrangement. *Nat. Struct. Biol.* **2000**, *7*, 602-608.
- [63] Larsen, A. K.; Skladanowski, A.; Bojanowski, K. The Roles of DNA Topoisomerase II during the Cell Cycle. *Prog. Cell Cycle Res.* **1996**, *2*, 229-239.
- [64] Kaufmann, W. K. Human Topoisomerase II Function, Tyrosine Phosphorylation and Cell Cycle Checkpoints. *Proc. Soc. Exp. Biol. Med.* **1998**, *217*, 327-334.
- [65] Holm, C.; Stearns, T.; Botstein, D. DNA topoisomerase II must act at mitosis to prevent nondisjunction and chromosome breakage. *Mol. Cell. Biol.* **1989**, *9*, 159-168.
- [66] Sumner, A. T. Inhibitors of Topoisomerase II Delay Progress Through Mitosis and Induce a Doubling of the DNA Content in CHO Cells. *Exp. Cell Res.* **1995**, *217*, 440-447.
- [67] Cimini, D.; Antocchia, A.; Tanzarella, C.; Degraasi, F. Topoisomerase II Inhibition in Mitosis Produces Numerical and Structural Chromosomal Aberrations in Human Fibroblasts. *Cytogenet. Cell Genet.* **1997**, *76*, 61-67.
- [68] Anderson, H.; Roberge, M. Topoisomerase II Inhibitors Affect Entry into Mitosis and Chromosome Condensation in BHK cells. *Cell Growth Differ.* **1996**, *7*, 83-90.
- [69] Bastians, H.; Topper, L. M.; Gorbsky, G. L.; Ruderman, J. V. Cell Cycle-Regulated Proteolysis of

Mitotic Target Proteins. *Mol. Biol. Cell.* **1999**, *10*, 3927-3941.

[70] Townsley, F. M.; Aristarkhov, A.; Beck, S.; Hershko, A.; Ruderman, J. V. Dominant-Negative

Cyclin-Selective Ubiquitin Carrier Protein E2-C/UbcH10 Blocks Cells in Metaphase. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 2362-2367.