Functional Profiling: From Microarrays via Cell-Based Assays to Novel Tumor Relevant Modulators of the Cell Cycle

Dorit Arlt,[1] Wolfgang Huber,[1] Urban Liebel,[2] Christian Schmidt,[1] Meher Majety,[1] Mamatha Sauermann,[1] Heiko Rosenfelder,[1] Stephanie Bechtel,[1] Alexander Mehrle,[1] Detlev Bannasch,[1] Ingo Schupp,[1] Markus Seiler,[1] Jeremy C. Simpson,[2] Florian Hahne,[1] Petra Moosmayer,[1] Markus Ruschhaupt,[1] Birgit Guilleaume,[1] Ruth Wellenreuther,[1] Rainer Pepperkok,[2] Holger Sültmann,[1] Annemarie Poustka,[1] and Stefan Wiemann[1]

[1]Division of Molecular Genome Analysis, German Cancer Research Center and [2]Cell Biology and Cell Biophysics Programme, European Molecular Biology Laboratory, Heidelberg, Germany

Abstract

Cancer transcription microarray studies commonly deliver long lists of "candidate" genes that are putatively associated with the respective disease. For many of these genes, no functional information, even less their relevance in pathologic conditions, is established as they were identified in large-scale genomics approaches. Strategies and tools are thus needed to distinguish genes and proteins with mere tumor association from those causally related to cancer. Here, we describe a functional profiling approach, where we analyzed 103 previously uncharacterized genes in cancer relevant assays that probed their effects on DNA replication (cell proliferation). The genes had previously been identified as differentially expressed in genome-wide microarray studies of tumors. Using an automated high-throughput assay with single-cell resolution, we discovered seven activators and nine repressors of DNA replication. These were further characterized for effects on extracellular signal-regulated kinase 1/2 (ERK1/2) signaling (G_1-S transition) and anchorage-independent growth (tumorigenicity). One activator and one inhibitor protein of ERK1/2 activation and three repressors of anchorage-independent growth were identified. Data from tumor and functional profiling make these proteins novel prime candidates for further in-depth study of their roles in cancer development and progression. We have established a novel functional profiling strategy that links genomics to cell biology and showed its potential for discerning cancer relevant modulators of the cell cycle in the candidate lists from microarray studies. (Cancer Res 2005; 65(17): 7733-42)

Introduction

Proliferation of mammalian cells requires that DNA replication is strictly regulated and done with high fidelity. The commitment of cells to enter the S phase is the primary decision point at which they control their division. Alterations in expression and activity of proteins that are involved in this process can cause unrestrained growth and are prerequisite for the development and progression of cancer (1). In this line, tumor-associated changes in mRNA transcription are frequently examined in tumor microarray studies. Whereas such analyses typically identify hundreds or thousands of genes (2–4), often, the function of the encoded proteins is unknown and their role in the disease process is not well understood (5). Long lists of differentially expressed genes reflect the complex interplay of biological processes that are involved in cancer (e.g., inflammation, angiogenesis, metabolism, as well as cell cycle control, and apoptosis; refs. 6–9). The distinction between causative changes in gene expression from secondary effects remains a key challenge.

We adopted the idea of functional profiling (10), developed a cancer relevant automated high-throughput assay, and applied it to genes that were associated with cancer by whole genome microarray analysis (11, 12). This identified novel proteins that modified the rate of DNA replication in mammalian cells, this being the first direct indication of a causative relation to oncogenic processes. We refined the functional profiles for 16 candidate genes and proteins with assays that probed other cell cycle– and cancer-related processes, extracellular signal-regulated kinase 1/2 (ERK1/2) signaling, and anchorage-independent growth. Functional profiling distinguished causative from secondary effects and is therefore able to bridge the gap between whole genome tumor microarray studies and single candidate analysis.

Materials and Methods

Selection of novel full-length cDNAs. Tumors of the brain, breast, kidney, and gastrointestinal stroma were screened for cancer-associated changes in mRNA transcription using a 31,500-element whole transcriptome cDNA array (11, 12). Genes with differential expression between normal and tumor samples or between tumor types were selected by a two-sample t test. From these, we selected 103 open

reading frames (ORF) based on the availability of full-length cDNAs for the cell-based bromodeoxyuridine (BrdUrd) incorporation assay. The cDNAs had been isolated from libraries of human fetal brain (library DKFZp564), human fetal kidney (DKFZp566), human amygdala (DKFZp761), human uterus (DKFZp586), and human testis (DKFZp434; ref. 13). ORFs were cloned and sequence validated in Gateway (14) vectors, then shuttled into NH$_2$-terminal cyan fluorescent protein (CFP) and COOH-terminal yellow fluorescent protein (YFP) to produce mammalian expression constructs under control of the cytomegalovirus promoter (15). Subcellular localization of proteins was determined as described (15). Initial annotation of protein sequences was done with automated tools (16).

Automated cell-based assay to assess DNA replication. Expression plasmids were isolated in a 96-well format using a MultiProbe IIex-Robot (Perkin-Elmer, Wellesley, MA) and Montage Plasmid Miniprep96 Kits (Millipore, Bedford, MA). DNA masterplates were generated with a final concentration of 27.5 ng/μL and a total volume of 80 μL per well. Cells (10,000; NIH 3T3 cells: ATCC CRL-1658) were transfected (Effectene, Qiagen, Hilden, Germany) using 100 ng plasmid DNA per well in an array of 12 chamber slides (MWG Biotech, Ebersberg, Germany) to mimic a 96-well plate. Each CFP and YFP expression construct was included in duplicate. Control cDNAs (CFP-cyclin A, cyclin A-YFP as positive controls for activating, CFP subunit A of phosphatase 2, subunit A of phosphatase 2-YFP as positive controls for repressing effects on DNA replication; CFP and YFP served as negative controls) were placed at fixed positions within the plate. After 24 hours, cells were incubated in presence of 1 mmol/L BrdUrd for 6 hours before fixation with 4% paraformaldehyde. Transfection, BrdUrd immunostaining with a mouse anti-BrdUrd primary antibody (Roche Diagnostics, Penzberg, Germany), incubation with a Cy5-labeled goat anti-mouse secondary antibody (Molecular Probes, Karlsruhe, Germany), and 4′,6-diamidino-2-phenyindole (DAPI) staining were done using a Multi-Probe robot.

Image acquisition and analysis. A fully automated high-content screening microscope (17) was used to measure on average 19 frames from each well. The image frames consisted of three color channels (CFP/YFP, Cy5, and DAPI) each at 1,280 × 1,024 pixels at 12-bit resolution. Images were analyzed by an automated segmentation and quantification algorithm that we implemented in Labview (National Instruments, http://ni.com).

Within-well analysis. The data from one well were modeled as

$$b_i = m(t_i) + \theta_{f(i)} + \kappa d_i + \varepsilon_i$$

where b_i, t_i, and d_i are the BrdUrd, transfection, and DAPI intensities, respectively, for cell i; m is a smooth regression function, $f(i)$ is the frame in which cell i was found; θ_f is the background for the fth frame; κ is a spectral crosstalk coefficient; and ε_i is a noise term. We introduced the correction terms θ_f and κd_i to address empirically observed systematic effects: effects: θ_f accounts for the different baseline levels of the fluorescence intensities in different frames and κ for spectral overlap from the DAPI signal into the CFP detection channel. A separate coefficient κ was fitted for each 96-well plate to account for apparent long-time drift of the equipment. Values for κ were between 0 and 0.06. Spectral overlap between DAPI and YFP was negligible. The smooth local regression function m, its derivative m', and confidence bands for m' were fitted using the R package locfit (18). A constant bandwidth $\alpha = 1$ was used throughout. To determine the intensity t° that corresponds to very weakly expressing cells we calculated the midpoint of the shorth, which is the shortest interval that contains half of the t_i. As a measure of the effect, we used the z score, which is the ratio of the estimated slope and its SE.

$$z = \frac{\hat{m}'\,(t^*)}{\hat{\sigma}_{m'}\,(t^*)}$$

Between-well analysis. To obtain a ranking of the proteins, we compared the set of z scores from the replicate wells for each ORF to the overall distribution of z scores via the two-sided Wilcoxon test. We considered an effect significant when $P < 0.05$. Multiple testing adjustment for control of the false discovery rate was done following Benjamini and Hochberg (19).

We provide software for statistical procedures and visualizations through the package *prada*, which is part of the bioinformatics platform *Bioconductor* (http://www.bioconductor.org).

Assay for quantification of anchorage-independent growth. Flat-bottomed 96-well plates were coated with poly(2-hydroxylethyl methacrylate) [poly(HEMA), 120 mg/mL of 95% ethanol] as described (20). Kidney clear cell carcinoma cells Caki-1 (ATCC HTB-46) and Caki-2 (ATCC HTB-47) cells were cultured in RPMI 1640 with 1% penicillin/streptomycin, 1% L-glutamine, 1% nonessential amino acids, and 10% fetal bovine serum (Life Technologies, Gaithersburg, MD) and transfected with Lipofectamine 2000 (Invitrogen, Karlsruhe, Germany) according to the manufacturer's instructions. Proteins found to activate DNA replication were transfected into Caki-2 cells to measure their potential to stimulate anchorage-independent growth of this cell line. Repressors of DNA replication were analyzed in Caki-1 cells for their potential to reduce anchorage-independent growth. Twenty-four hours after transfection, 3,500 Caki-1 or Caki-2 cells were spread in each well of the poly(HEMA)-coated plates in a total volume of 200 μL RPMI 1640. Medium was exchanged after
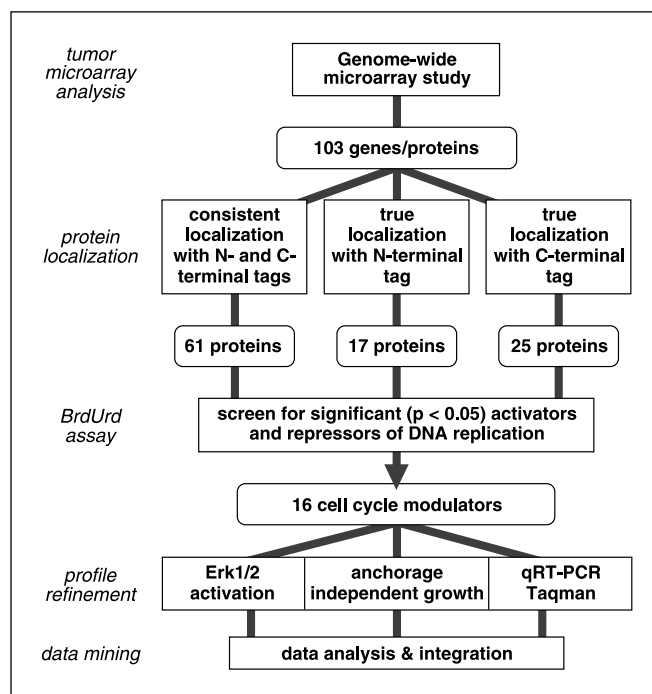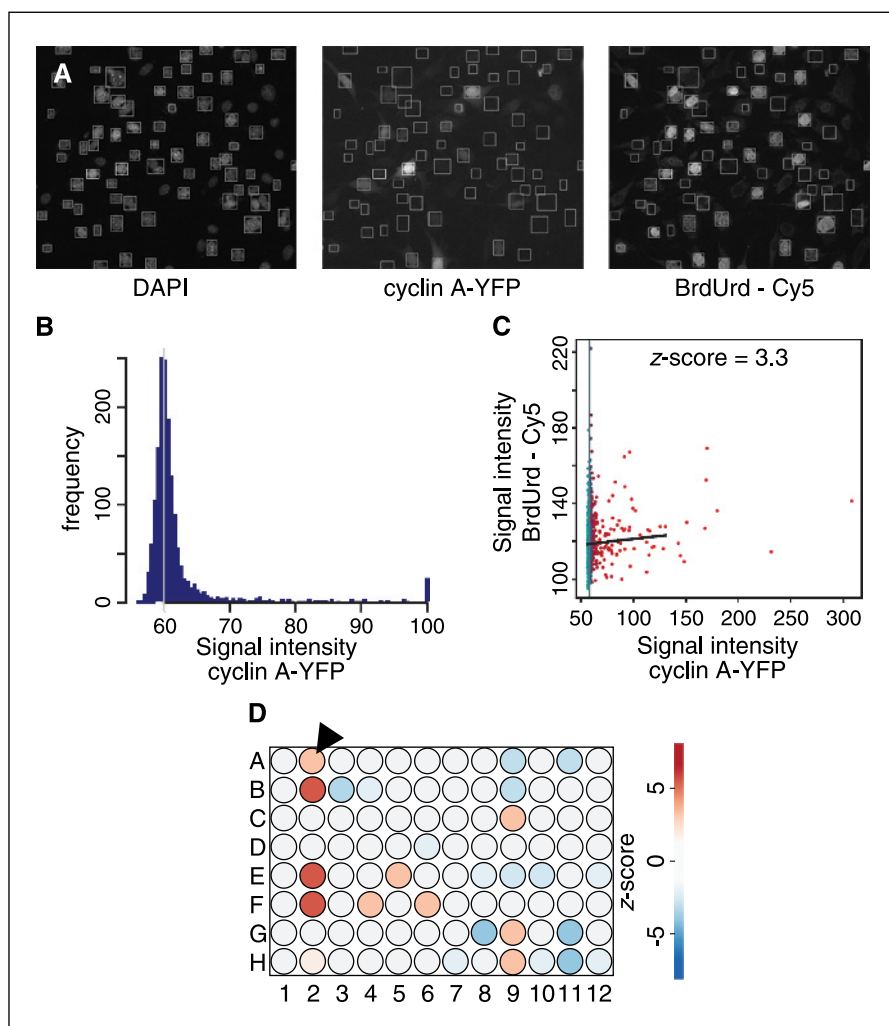

Figure 1. Schematic representation of the experimental and analysis workflow in functional profiling. Lists from whole genome tumor microarray studies (11, 12), and a full-length ORF collection (13, 15) were intersected to select 103 previously uncharacterized proteins for the functional profiling assays. Data were analyzed only from those expression constructs (NH$_2$-terminal CFP or COOH-terminal YFP) annotated as expressing correctly localizing proteins. Statistical analysis determined 16 proteins with significant effect on DNA replication, which were analyzed in further assays (ERK1/2 activation and anchorage-independent growth). Quantitative reverse transcription-PCR (*qRT-PCR*) analysis (Taqman) compared gene expression between two cancer cell lines which differ in anchorage-independent growth. Finally, data were integrated to annotate the novel cancer-relevant candidate genes and proteins.

Figure 2. Within-well analysis of protein overexpression and the effects on DNA replication. *A,* cells were fixed and stained with DAPI (*left*) to allow for automatic identification (*nuclei boxed*). Fluorescence intensities were recorded for each cell in the YFP channel and indicate the level of cyclin A-YFP expression (*middle*). Cy5 fluorescence in the nucleus is proportional to the level of BrdUrd incorporation (*right*). *B,* histogram of the YFP fluorescence intensities (arbitrary units) from 1,025 cells. The vertical grey line indicates the midpoint of the shorth t^* that is an estimator of the mode of the distribution. *C,* scatterplot of expression level versus BrdUrd incorporation. The control population of nonexpressing cells (*left*), showing a considerable intrinsic variation in their BrdUrd-Cy5 signal. A systematic trend was observed for the population of expressing cells towards the right end of the plot: the higher the expression level, the more cells had incorporated BrdUrd. This is visualized by the local regression curve, which was calculated based on the data of nonexpressing and weakly overexpressing cells only, not regarding cells with signal intensities >100. As expected, the slope was indicative of cyclin A being an activator of DNA replication. *D,* visual overview of the protein effects from a whole 96-well plate. *Z* scores for positive (activators) and negative (repressors) slopes are indicated in red and blue color, respectively. Data from well A2 (*arrowhead*) are shown in (*A-C*).

4 days of culture. Measurement of cell growth was done as described by Fukazawa et al. (20).

Results and Discussion

Assay automation. We designed a functional profiling strategy that took genes that had been associated with cancer by microarray studies and screened the encoded proteins in tumor relevant assays (Fig. 1). ORFs of previously uncharacterized genes were subcloned from publicly available full-length cDNA resources (13, 21) into expression vectors (15) and transfected into NIH 3T3 cells to overexpress the encoded proteins. All ORFs were fused to NH$_2$-terminal CFP or COOH-terminal YFP variants of the green fluorescent protein (GFP), and their expression was monitored by automated microscopy. The protein localization at the subcellular level was visually inspected for all constructs (15, 22). This information was used in the subsequent assays, only data obtained with the relevant expression constructs were considered (Fig. 1).

A DNA replication assay was fully automated with help of liquid handling robots and customized statistics software. Transient trans-fections were carried out in 96-well plates; analysis of protein overexpression and induced effects was 24 hours later. We applied automated high-content screening microscopy (17, 23) to achieve a single-cell resolution (Fig. 2) and traced individual cells within their microenvironment of the well. We restricted our analysis to cells having small levels of protein expression to minimize the perturbation of the cellular system. Whereas we cannot completely rule out nonphysio-logic effects that might be induced by even minor increases in protein content, our strategy is superior to other approaches, like plate reader assays and RNA interference, where a correlation of protein levels with effects is not possible at a single-cell resolution. Additionally, we could simultaneously identify activators and repressors of DNA replication (i.e., potential oncogenes and tumor suppressor genes; Figs. 2 and 3). Cyclin A and phosphatase 2A were used in the assay set up and in every screening plate as positive controls for activation and repression, respectively.

DNA replication was quantified by monitoring the rate of BrdUrd incorporation with help of a Cy5-labeled antibody. In addition, cells were stained with DAPI to facilitate autofocus of the microscope and to aid the image segmentation algorithm with the cell finding. Thus, three fluorescence intensities (GFP/YFP, Cy5,
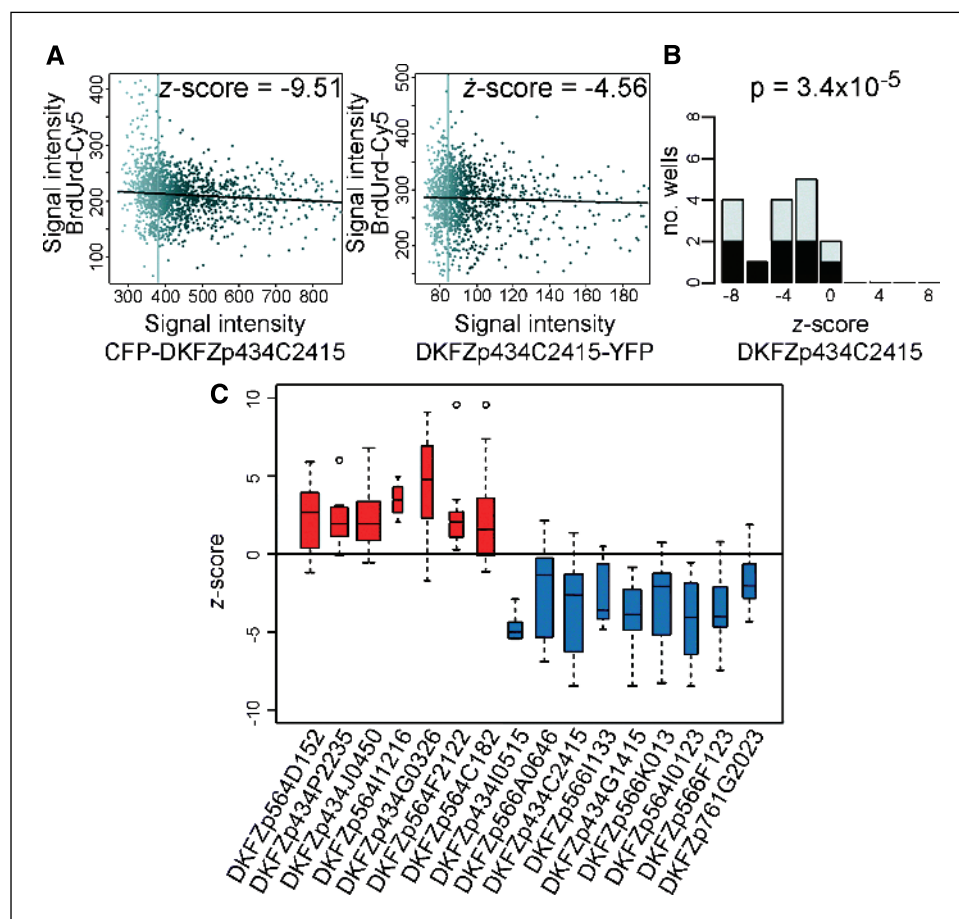
Figure 3. Between-well analysis to validate activating and repressing effects on DNA replication through replicate experiments. *A*, scatterplots from individual wells were produced to show activating or repressing effects. The *z* scores of −9.51 and −4.56 indicate the protein DKFZp434C2415 to be a repressor. *B*, we summarized and compared the results from multiple wells for the same protein through histograms of the *z* scores for both tag orientations. The plot shows the *z* scores calculated for individual wells, separately for NH_2-terminal CFP-fusions (*gray*) and for COOH-terminal YFP-fusions (*black*). $P = 3.4 \times 10^{-5}$ is indication that the measured effect is significant. *C*, box plot of the *z* score values were produced for the seven activators (*red*) and nine repressors of DNA replication (*blue*) identified in the screen. *Columns,* means; *bars,* SE. Outliers (○).

and DAPI) were recorded for each cell. The effect of the candidate proteins' expression on DNA replication was determined by local regression of the Cy5 signal (BrdUrd) on the GFP/YFP signal (expression). For each well, this resulted in an estimated regression slope and a *z* score that is the estimate standardized by its SE. Data from replicate wells per protein were summarized by the Wilcoxon statistic and *P* value for the comparison of their *z* scores to the overall distribution of *z* scores. This resulted in a ranking of the proteins for effect strength and a list of 16 proteins with *P*s < 0.05.

Assay performance. The results of all microscopic frames (19 on average) from within one well were combined, taking into the account disturbance factors as spectral overlaps and frame-specific background values. One such frame is shown in Fig. 2*A*. The intracellular amount of recombinant protein varied over a broad range. The majority of cells were either untransfected (cells left of the midpoint of the shorth in Fig. 2*B*) or expressed the protein at low levels (right of the midpoint of the shorth). Few cells strongly expressed the recombinant protein with signal intensities of >100 in the YFP channel. A scatterplot of expression level versus BrdUrd incorporation is shown in Fig. 2*C*. Even the population of nonexpressing cells showed a considerable intrinsic variation in DNA replication as reflected by varying BrdUrd-Cy5 signals. A systematic trend was observed, however, for the population of cyclin A–expressing cells towards the right end of

the plot: the higher the expression level, the more cells had incorporated BrdUrd during the experiment. This trend is visualized by the local regression curve in Fig. 1*C*. This curve was calculated based on the data of nonexpressing and weakly overexpressing cells. Strongly overexpressing cells (e.g., signal intensity >100 in Fig. 2*B-C*) were downweighed as in these the level of recombinant protein could induce nonphysiologic effects (e.g., apoptosis). A *z* score was calculated as a summary of the data for the cells within one well.

To obtain a visual overview of the protein effects from a 96-well experiment, we produced plate plots (Fig. 2*D*). *Z* scores for positive (activators) and negative (repressors) slopes are indicated in color. The red color for cyclin A is indicative of this protein to be an activator of DNA replication. Interactive versions of the plate plots are provided in the web supplement (http://www.dkfz.de/mga/home/fctassay/sphase). Scatterplots similar to and including the one shown in Fig. 2*C* can be viewed by mouse clicking on a well. To assess statistical significance, we compared the results from multiple wells for the same protein through histograms of the *z* scores for both tag orientations (Fig. 3).

Candidate selection. We selected 103 genes (Fig. 1; Table S1) for a cellular screen based on the following criteria: (i) their transcripts were contained in a set of 396 differentially expressed genes from genome-wide tumor microarray studies (11, 12); (ii)

the encoded proteins had no public functional annotation, other than from high-throughput projects such as microarray analysis; (iii) we had determined the subcellular localization of the proteins (15, 22). This provided an initial piece of information that connected the proteins to their respective cellular environment and thus limited their potential for protein interactions. In addition, use of full-length proteins ensured that all motifs determining, for example, protein localization, activity, and regulation were present.

DNA replication assay and functional profiling. Analyzing the 103 proteins in the BrdUrd incorporation assay, cells were detected in 2,462 of 2,880 wells. Fluorescence intensities were recorded for CFP/YFP, Cy5 (BrdUrd), and DAPI signals individually, resulting in $3 \times 46,221$ images. A total of 2,251,739 cells were analyzed, giving over 6.75 million data points. Sixteen of the 103 proteins showed a significant effect on the passage through the S phase (individual $P < 0.05$; false discovery rate <0.16). Seven of these were activators and nine were repressors of BrdUrd incorporation and thus of DNA replication (Fig. 3; Table 1). The 16 proteins were then examined for a possible effect on ERK1/2 activation, to measure effects on G_1-S transition (24). There we found one activating and one inhibiting protein (Fig. S1; Table 1). Repressors of DNA replication were further subjected to a poly(HEMA) growth assay to test their potential to modulate anchorage-independent cell proliferation. Three proteins were positive in this assay (Fig. 4; Table 1). We annotated all 16 genes and proteins using tumor microarray results from Oncomine (ref. 25; Table S1). Taqman RNA quantification was done in two kidney cancer cell lines (Caki-1 and Caki-2) to refine analysis of the poly(HEMA) assay and was included in the annotation as well as protein and domain classifications from Source (26) and Interpro (ref. 27; Table S1). Together, this integrated data set provided us with candidates for detailed and ongoing studies. In the following, four such genes will be described to show the synergy that is achieved with the functional profiling strategy.

Cancer-relevant cell cycle modulators. The protein encoded by cDNA DKFZp434P2235, an activator of passage through the S phase, localizes to the plasma membrane. The corresponding gene has been independently reported as *PRC17* oncogene (28). Pei et al. found it amplified in 15% of prostate cancers and highly overexpressed in about half of metastatic prostate tumors. We found the transcription of the *PRC17* (DKFZp434P2235) gene up-regulated in kidney tumors (Table 1). The gene is also expressed in the kidney carcinoma cell line Caki-1, which shows anchorage-independent growth but not in Caki-2 which does not (Figs. S2 and S3). The encoded protein interacts with Rab5 and activates its GTPase activity (28). Pei et al. discuss that expression of *PRC17* could affect the amount of inactive Rab5, which should inhibit receptor endocytosis and result in prolonged growth factor signaling. This hypothesis is consistent with our finding that the overexpressed DKFZp434P2235 protein promotes DNA synthesis in the S phase (Fig. 3). Whereas the *PRC17* oncogene takes one candidate from our list of truly novel genes, the discovery of this protein in the DNA replication assay shows the power of our integrated approach to identify proteins with cancer relevance.

The gene for DKFZp566A0646 was found down-regulated in kidney tumors (Table 1) and to have a lower expression in Caki-1 compared with Caki-2 cells (Fig. S2). The overexpressed protein inhibited DNA synthesis in the S phase (Fig. 3) and blocked

anchorage-independent growth of Caki-1 cells (Fig. 4). These findings suggest a repressing effect on tumorigenicity *in vitro*. The protein encoded by cDNA DKFZp566A0646 contains a domain common to the Rab subfamily of Ras small GTPases (IPR003579). The function of its orthologous protein p34 in rat is not established; however, in a yeast two-hybrid screen, it was shown to be a binding partner for the γ and α subunits of the AP1 and AP2 complexes, respectively (29). These adaptor complexes participate in intracellular clathrin-coated vesicle trafficking. We found that *Rab5* as well as its effector early endosome antigen-1 (*EEA-1*) are down-regulated in kidney tumors (11). Together with the up-regulation of the Rab5 GTPase-activating enzyme PRC17, this would imply reduced receptor internalization through endocytosis. In addition, Rab GDP-dissociation inhibitor α (*GDIα*) was found up-regulated and *Rab11b* down-regulated in kidney tumors, supporting the hypothesis that endocytosis is inhibited in kidney cancer. Changing expression levels of *Rab* genes have also been described for other cancer types (30–32).

The protein encoded by cDNA DKFZp564D152 has 36% identity over 186 residues with the SET protein of *Drosophila melanogaster* (SOURCE; ref. 26), which belongs to the nucleosome assembly protein (NAP) family. The NAP family is a group of histone chaperone-like proteins that are known to be regulators of transcriptional control (33). Nucleosome assembly and disassembly are essential processes in the S phase of the cell cycle. In proliferating cells, the synthesis of histones during the S phase is coupled to DNA synthesis. This ensures the packing of chromatin. The protein NAP-1 is thought to be a cell cycle regulator and interacts with Kap114p (34). That protein is a member of the karyopherin/importin family, which executes the nuclear import of histone 2A and histone 2B. The activating effect of the DKFZp564D152 protein on DNA replication (Fig. 3), in conjunction with its nuclear and cytoplasmic localization support the assumption that it might be involved in cytoplasmic-nuclear transfer of histones.

The protein encoded by DKFZp434C2415 was found to repress DNA replication in the S phase (Fig. 3). However, it had a significant activating effect on ERK1/2 activation in serum starved as well as in stimulated cells (Table 1; Fig. S1). The protein was recently described as UBA5, which is an E1-like enzyme in the ubiquitin degradation pathway (35). Ubiquitin activating enzymes regulate the rate of ubiquitin conjugation and ubiquitin mediated protein degradation, which are essential control mechanisms in the transition from G_1 to S phase. Several studies show that cell cycle regulating proteins are directly targeted for ubiquitin-mediated destruction through ERK1/2 (36, 37). Thus, DKFZp434C2415 could regulate ERK1/2 activity through ubiquitin mediated degradation of, for example, phosphatases, the activated ERK1/2 would then mark cell cycle–regulating proteins for degradation.

The DKFZp566F123 protein was identified as a BTB/POZ zinc finger protein (POK protein). The BTB/POZ domain (IPR000210) mediates protein/protein interaction, whereas the C2H2 zinc finger (IPR007087) is a DNA-binding domain. DKFZp566F123 shares closest homology to the transcriptional repressor BCL-6. The latter protein is involved in hematopoiesis, oncogenesis, and immune response (38) and multiple roles in cell survival and differentiation are described (39). Zhang

Table 1. Summary of the results for the 16 candidate proteins

Clone ID	Genbank accession no.	Localization NH$_2$-terminal tag	Localization COOH-terminal tag	S-phase effect	S-phase P	ERK1/2 effect	ERK1/2 P	HEMA effect
DKFZp564D152	AL136629	Cytoplasm and nucleus	Cytoplasm and nucleus	+	0.0012	NE		NA
DKFZp434P2235	AL136860	Golgi and plasma membrane	Golgi and plasma membrane	+	0.0021	NE		NA
DKFZp434J0450	AL136869	Nucleus	Nucleus	+	7.9e−5	NE		NA
DKFZp564I1216	AL136600	Cytoplasm and nucleus	Endoplasmic reticulum	+	0.014	NE		NA
DKFZp434G0326	AL136820	Cytoplasm	Cytoplasm	+	0.0025	ND		NA
DKFZp564F2122	AL136604	Cytoskeleton and microtubules	Cytoskeleton and microtubules	+	0.0088	NE		NA
DKFZp564C182	AL136628	Cytoplasm	Golgi	+	0.022	NE		NA
DKFZp434I0515	AL136761	Cytoplasm and nucleus	Cytoplasm and nucleus	−	6e−6	−	0.02	NE
DKFZp566A0646	AL136715	Cytoplasm	Cytoplasm	−	0.0038	NE		−
DKFZp434C2415	AL136757	Cytoplasm	Cytoplasm	−	3.4e−5	+	1e−4	NE
DKFZp566I133	AL136711	Endoplasmic reticulum	NA	−	0.015	NE		−
DKFZp434G1415	AL136759	Nucleus	Nucleus	−	7.9e−7	NE		−
DKFZp566K013	AL136712	Mitochondria	Mitochondria	−	8e−5	NE		NE
DKFZp564I0123	AL136615	Cytoplasm	Cytoplasm	−	3.4e−4	NE		NE
DKFZp566F123	AL050276	Nucleus	Nucleus	−	9.3e−4	NE		NE
DKFZp761G2023	AL136570	Nucleus	Nucleus	−	0.035	NE		NE

NOTE: Localization data were taken from http://www.LIFEdb.de (50) and are given for both orientations of the tag relative to the ORF. Wrongly localizing fusion proteins are indicated in gray. Measured effects on the S phase are given with direction (+, activators; −, repressors), and P value of the calculated z score All proteins were analyzed for effects on ERK1/2 activation. Activators of ERK1/2 phosphorylation are marked with +, repressors with −. Proteins having no effect on ERK1/2 phosphorylation are marked with NE. Proteins repressing DNA replication were tested for effects on anchorage-independent growth [poly(HEMA) assay] in the Caki-1 cell line (Fig. 4).

Table 1. Summary of the results for the 16 candidate proteins (Cont'd)

Tumor profiling	Oncomine	Annotation (Source; InterPro)
Up in kidney tumors versus normal, up in GIST tumors versus normal	Up in lung (Bhattacharjee), down in liver (Chen), down in multiple tumors (Ramaswamy)	TSPY-like 1, IPR002164 NAP
Up in kidney tumors versus normal	Not differential in one study	TBC1 domain family, member 3, IPR000195 RabGAP/TBC domain
Down in kidney tumors versus normal	Up in lung (Bhattacharjee), up in endometrium (Risinger), up in breast cancer with bad prognosis (van't Veer), down in melanoma (Segal), down in prostate (Singh)	RecQ protein-like 5, no InterPro-hits
Down in kidney tumors versus normal	Up in liver (Chen), up in metastatic prostate cancers (Ramaswamy), up in breast cancer with bad prognosis (van't Veer)	Transmembrane protein 9; IPR004153 CXCXC repeat
Different in glioblastomas multiforme versus oligodendrogliomas	Up in breast adenocarcinoma: grade 3 versus grade 1 (van't Veer)	Hypothetical protein, IPR000379 esterase/lipase/thioesterase
NA	Down in breast adenocarcinoma: metastasis within 5 y of diagnosis versus no metastasis; grade 3 versus grade 1 (van't Veer)	Spermatogenesis associated 7, no InterPro-hits
ND	Not differentially expressed in four studies	Hypothetical protein, no InterPro-hits
Down in kidney tumors versus normal	Not probed in any of the studies	Radial spokehead-like 1, IPR001156 Peptidase S60, transferin lactoferrin, IPR006802 Radial spokehead-like protein
Down in kidney tumors versus normal	Up in liver (Chen), lung (Garber), down in kidney (Higgins), up in breast cancer with bad prognosis (van't Veer)	Hypothetical protein, IPR003579 Ras small GTPase, Rab type
Up in kidney tumors versus normal	Up in lung (Garber), down in pancreas (Iacobuzio)	Ubiquitin-activating enzyme estrone-domain containing 1, IPR000594 UBA/THIF-type NAD/FAD binding fold, IPR000205 NAD-binding site, IPR006140 D-isomer specific 2-hydroxyacid dehydrogenase, NAD-binding, IPR009036 Molybdenum cofactor biosynthesis
Up in kidney tumors versus normal	Up in liver (Chen), prostate (Luo, Dhanasekaran), lung (Garber), renal (Higgins), and in multiple tumors (Ramaswamy); down in pancreas (Iacobuzio)	Likely homologue of rat vacuole membrane protein 1, no InterPro-hits
Different in glioblastomas multiforme versus oligodendrogliomas	Not differentially expressed in three studies	Hypothetical protein, IPR001656 tRNA pseudouridine synthase D, TruD
Different in brain tumors stage III versus stage IV	Down in CNS (Khatua, astrocytoma; Nutt, glioma; Ramaswamy, Multicancer)	Dynamin-related protein DNM1, IPR000375 Dynamin central region/IPR001401 Dynamin/IPR001849 Pleckstrin-like/IPR003130 Dynamin GTPase effector
NA	Up in lung (Bhattacharjee), up in salivary adenoid cystic carcinoma (Frierson)	Protein activator of the IFN-induced protein kinase, IPR000999 Ribonuclease III family/IPR001159 double-stranded RNA binding (DsRBD) domain
Up in GIST tumors versus normal	Down in prostate in three different studies (La Tulippe, Luo JH, Welsh)	Zinc finger protein 288, IPR000210 BTB/POZ domain/IPR007087 Zn-finger, C2H2 type
Different in glioblastomas multiforme versus oligodendrogliomas	Down in high-grade breast tumors (van't Veer)	LIM homeobox 6, IPR001356 Homeobox/IPR001781 Zn-binding protein, LIM/IPR003350 Homeodomain protein CUT

NOTE: Proteins reducing growth of Caki-1 carcinoma cells are marked with −, proteins having no effect when overexpressed are marked with NE. Tumor profiling results were extracted from Boer et al. (11) and Sultmann et al. (12). The individual studies mentioned in the Oncomine (25) column are referenced in http://www.oncomine.org. Protein annotations are from the Source (26) and InterPro (27) databases. The complete data for the 103 proteins analyzed is in Table S1.
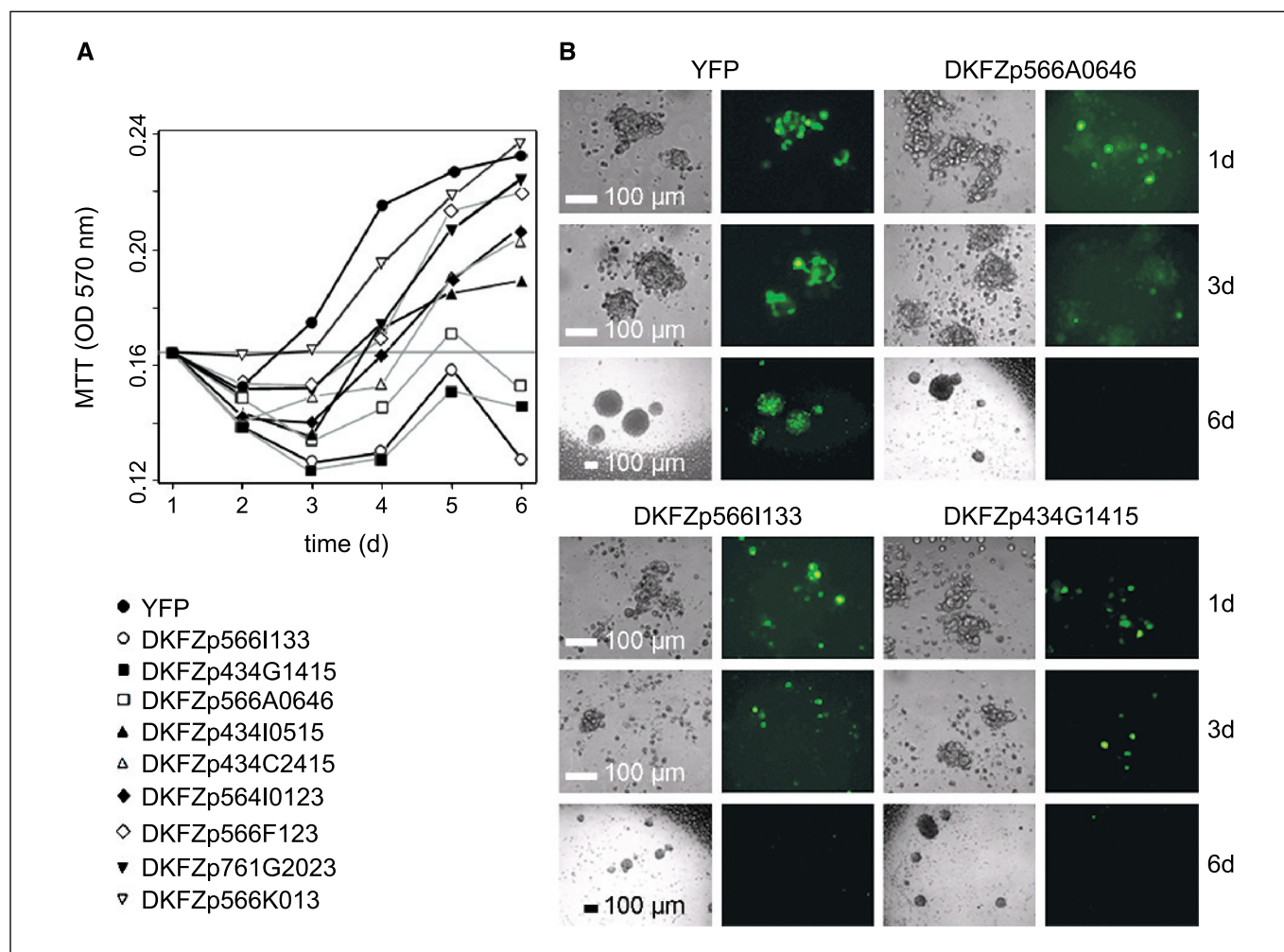Abbreviations: NA, not analyzed; ND, no detectable expression.

Figure 4. Screening of repressors of DNA proliferation for their effect on anchorage-independent growth of Caki-1 kidney carcinoma cells. *A,* growth curves of transiently transfected Caki-1 cells growing in poly(HEMA)-coated plates (transfection efficiency, 90%). MTT reduction (*OD*) was measured from days 1 to 6. *B,* inhibition of anchorage independent growth. While YFP-transfected Caki-1 cells retained their potential to grow on poly(HEMA) substrate, cells overexpressing proteins from cDNAs DKFZp566A0646, DKFZp566I133, and DKFZp434G1415 did not survive in this assay.

et al. hypothesize that the protein encoded by DKFZp566F123 might also be involved in these biological processes and we found that overexpression of this protein has indeed a repressing effect on passage through the S phase (Fig. 3). Thus, we postulate that DKFZp566F123 is a transcriptional repressor that regulates the expression of proteins that are essential for control of cell cycle progression, and its down-regulation would lead to enhanced cell growth. Also for other POK proteins (e.g., the tumor suppressor HIC1; ref. 40), it has been shown that their inactivation is associated with cell transformation (39).

Conclusions

We have developed a fully automated cell-based assay that is designed to identify modulators of DNA replication. It has a single-cell resolution with exceptionally high sensitivity and specificity. The assay was applied to narrow down a list of cancer associated genes from microarray studies, identifying novel candidate modulators of the cell cycle. Here we selectively discerned proteins based on their property to affect a specific activity of cells (DNA replication); however, this transfection screen is adaptable for analyzing diverse cellular pathways and processes.

The endogenous expression of genes in tumor and normal cells in many cases correlated with the outcome of the DNA replication assay. For example, the gene encoding cDNA DKFZp434I0515 was found down-regulated in kidney tumor profiling, whereas it is expressed in normal kidney. Gene expression was not detectable in Caki-1 and Caki-2 cells. The encoded protein had repressing effects on DNA replication. Such correlation should however not be regarded obligatory (41), because cellular effects of endogenous and overexpressed proteins may be tissue

and cell type specific that are tuned by cellular regulatory networks.

Depending on the outcome of the initial DNA replication screen additional assays were carried out with protein subsets to further characterize the candidates. Whereas it is clear that the described high-throughput assays do not uncover all molecular and cellular functions of the respective proteins, they effectively bridge the gap between gene lists from tumor microarray studies and the capacity of single gene and protein analysis that is needed to comprehensively characterize a pathway with traditional biochemistry and cell biology methodologies.

This concept holds the potential for clustering genes based on their functional profiles (42) and epistatic analyses, to quantitatively elucidate complex genetic networks. Quantitative cell-based analysis offers an advantage by permitting the detection of gene functions that are associated with subtle, possibly buffered effects. These are basis for the complex interactions of multiple processes that are involved in the onset and progression of most diseases (43). Combined with the mapping of protein interaction networks (44) and other large-scale cancer relevant assays (45–49), this will lay the ground for the comprehensive exploitation of genomic resources and information and ultimately lead to an accelerated functional understanding of the cellular systems that control health and disease.

Acknowledgments

References

1. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell 2000;100:57–70.
2. van 't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002;415:530–6.
3. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. Nature 2000;406:747–52.
4. Alizadeh AA, Eisen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 2000;403:503–11.
5. Ramaswamy S, Tamayo P, Rifkin R, et al. Multiclass cancer diagnosis using tumor gene expression signatures. Proc Natl Acad Sci U S A 2001;98: 15149–54.
6. Bissell MJ, Radisky D. Putting tumours in context. Nat Rev Cancer 2001;1:46–54.
7. Besson A, Assoian RK, Roberts JM. Regulation of the cytoskeleton: an oncogenic function for CDK inhibitors? Nat Rev Cancer 2004;4:948–55.
8. Sherr CJ. Principles of tumor suppression. Cell 2004; 116:235–46.
9. Balkwill F. Cancer and the chemokine network. Nat Rev Cancer 2004;4:540–50.
10. Giaever G, Chu AM, Ni L, et al. Functional profiling of the *Saccharomyces cerevisiae* genome. Nature 2002;418: 387–91.
11. Boer JM, Huber WK, Sultmann H, et al. Identification and classification of differentially expressed genes in renal cell carcinoma by expression profiling on a global human 31,500-element cDNA array. Genome Res 2001; 11:1861–70.
12. Sültmann H, v. Heydebreck A, Huber W, et al. Gene expression in kidney cancer is associated with novel tumor subtypes, cytogenetic abnormalities and metastasis formation. Clin Cancer Res 2005;11:646–55.
13. Wiemann S, Weil B, Wellenreuther R, et al. Toward a catalog of human genes and proteins: sequencing and analysis of 500 novel complete protein coding human cDNAs. Genome Res 2001;11:422–35.
14. Hartley JL, Temple GF, Brasch MA. DNA cloning using *in vitro* site-specific recombination. Genome Res 2000;10:1788–95.
15. Simpson JC, Wellenreuther R, Poustka A, Pepperkok R, Wiemann S. Systematic subcellular localization of novel proteins identified by large scale cDNA sequencing. EMBO Rep 2000;1:287–92.

16. del Val C, Mehrle A, Falkenhahn M, et al. High-throughput protein analysis integrating bioinformatics and experimental assays. Nucleic Acids Res 2004;32: 742–8.
17. Liebel U, Starkuviene V, Erfle H, et al. A microscope-based screening platform for large-scale functional protein analysis in intact cells. FEBS Lett 2003;554: 394–8.
18. Loader C. Local regression and likelihood. New York: Springer Verlag; 1999.
19. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society B 1995;57:289–300.
20. Fukazawa H, Mizuno S, Uehara Y. A microplate assay for quantitation of anchorage-independent growth of transformed cells. Anal Biochem 1995;228:83–90.
21. Strausberg RL, Feingold EA, Grouse LH, et al. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. Proc Natl Acad Sci U S A 2002;99:16899–903.
22. Wiemann S, Arlt DH, Huber W, et al. From ORFeome to biology: a functional genomics pipeline. Genome Res 2004;14:2136–44.
23. Starkuviene V, Liebel U, Simpson JC, et al. High-content screening microscopy identifies novel proteins with a putative role in secretory membrane traffic. Genome Res 2004;14:1948–56.
24. Chang F, Steelman LS, Shelton JG, et al. Regulation of cell cycle progression and apoptosis by the Ras/ Raf/MEK/ERK pathway (Review). Int J Oncol 2003; 22:469–80.
25. Rhodes DR, Yu J, Shanker K, et al. ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia 2004;6:1–6.
26. Diehn M, Sherlock G, Binkley G, et al. SOURCE: a unified genomic resource of functional annotations, ontologies, and gene expression data. Nucleic Acids Res 2003;31:219–23.
27. Mulder NJ, Apweiler R, Attwood TK, et al. InterPro, progress and status in 2005. Nucleic Acids Res 2005;33: D201–5.
28. Pei L, Peng Y, Yang Y, et al. PRC17, a novel oncogene encoding a Rab GTPase-activating protein, is amplified in prostate cancer. Cancer Res 2002;62:5420–4.
29. Page LJ, Sowerby PJ, Lui WW, Robinson MS. γ-Synergin: an EH domain-containing protein that interacts with γ-adaptin. J Cell Biol 1999;146:993–1004.

30. Yao R, Wang Y, Lubet RA, You M. Differentially expressed genes associated with mouse lung tumor progression. Oncogene 2002;21:5814–21.
31. He H, Dai F, Yu L, et al. Identification and characterization of nine novel human small GTPases showing variable expressions in liver cancer tissues. Gene Expr 2002;10:231–42.
32. Calvo A, Xiao N, Kang J, et al. Alterations in gene expression profiles during prostate cancer progression: functional correlations to tumorigenicity and down-regulation of selenoprotein-P in mouse and human tumors. Cancer Res 2002;62:5325–35.
33. Shikama N, Chan HM, Krstic-Demonacos M, et al. Functional interaction between nucleosome assembly proteins and p300/CREB-binding protein family coactivators. Mol Cell Biol 2000;20:8933–43.
34. Mosammaparast N, Ewart CS, Pemberton LF. A role for nucleosome assembly protein 1 in the nuclear transport of histones H2A and H2B. EMBO J 2002;21: 6527–38.
35. Komatsu M, Chiba T, Tatsumi K, et al. A novel protein-conjugating system for Ufm1, a ubiquitin-fold modifier. EMBO J 2004;23:1977–86.
36. Yehia G, Schlotter F, Razavi R, Alessandrini A, Molina CA. Mitogen-activated protein kinase phosphorylates and targets inducible cAMP early repressor to ubiquitin-mediated destruction. J Biol Chem 2001; 276:35272–9.
37. Niu H, Ye BH, Dalla-Favera R. Antigen receptor signaling induces MAP kinase-mediated phosphorylation and degradation of the BCL-6 transcription factor. Genes Dev 1998;12:1953–61.
38. Zhang W, Mi J, Li N, et al. Identification and characterization of DPZF, a novel human BTB/POZ zinc finger protein sharing homology to BCL-6. Biochem Biophys Res Commun 2001;282:1067–73.
39. Albagli-Curiel O. Ambivalent role of BCL6 in cell survival and transformation. Oncogene 2003;22: 507–16.
40. Pinte S, Stankovic-Valentin N, Deltour S, Rood BR, Guerardel C, Leprince D. The tumor suppressor gene HIC1 (hypermethylated in cancer 1) is a sequence-specific transcriptional repressor: definition of its consensus binding sequence and analysis of its DNA binding and prepressive properties. J Biol Chem 2004; 279:38313–24.
41. Nilsson JA, Cleveland JL. Myc pathways provoking cell suicide and cancer. Oncogene 2003;22:9007–21.

42. Piano F, Schetter AJ, Morton DG, et al. Gene clustering based on RNAi phenotypes of ovary-enriched genes in *C. elegans*. Curr Biol 2002;12:1959–64.

43. Hartman JLt, Garvik B, Hartwell L. Principles for the buffering of genetic variation. Science 2001;291: 1001–4.

44. Li S, Armstrong CM, Bertin N, et al. A map of the interactome network of the metazoan *C. elegans*. Science 2004;303:540–3.

45. Gelman MS, Ye XK, Stull R, et al. Identification of cell surface and secreted proteins essential for tumor cell survival using a genetic suppressor element screen. Oncogene 2004;23:8158–70.

46. Eustace BK, Sakurai T, Stewart JK, et al. Functional proteomic screens reveal an essential extracellular role for hsp90 α in cancer cell invasiveness. Nat Cell Biol 2004;6:507–14.

47. Kittler R, Putz G, Pelletier L, et al. An endoribonuclease-prepared siRNA screen in human cells identifies genes essential for cell division. Nature 2004;432:1036–40.

48. Wan D, Gong Y, Qin W, et al. Large-scale cDNA transfection screening for genes related to cancer development and progression. Proc Natl Acad Sci U S A 2004;101:15724–9.

49. Boutros M, Kiger AA, Armknecht S, et al. Genome-wide RNAi analysis of growth and viability in *Drosophila* cells. Science 2004;303:832–5.

50. Bannasch D, Mehrle A, Glatting K-H, Pepperkok R, Poustka A, Wiemann S. LIFEdb: a database for functional genomics experiments integrating information from external sources, and serving as a sample tracking system. Nucleic Acids Res 2004;32: D505–8.