

# Identification of novel mammalian growth regulatory factors by genome-scale quantitative image analysis

Josephine N. Harada,<sup>1,5</sup> Kristen E. Bower,<sup>2,5</sup> Anthony P. Orth,<sup>1,5</sup> Scott Callaway,<sup>3</sup> Christian G. Nelson,<sup>1</sup> Casey Laris,<sup>4</sup> John B. Hogenesch,<sup>1</sup> Peter K. Vogt,<sup>2</sup> and Sumit K. Chanda<sup>1,6</sup>

<sup>1</sup>Genomics Institute of the Novartis Research Foundation, San Diego, California 92121, USA; <sup>2</sup>Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, California 92037, USA; <sup>3</sup>Vala Sciences, Inc., La Jolla, California 92037, USA; <sup>4</sup>Biomedical Research Division, Beckman Coulter, Inc., San Diego, California 92121, USA

Functional profiling technologies using arrayed collections of genome-scale siRNA and cDNA arrayed libraries enable the comprehensive global analysis of gene function. However, the current repertoire of high-throughput detection methodologies has limited the scope of cellular phenotypes that can be studied. In this report, we describe the systematic identification of mammalian growth-regulatory factors achieved through the integration of automated microscopy, pattern recognition analysis, and cell-based functional genomics. The effects of 7364 human and mouse proteins, encoded by individually arrayed cDNAs, upon proliferation and viability in U2OS osteosarcoma cells were evaluated in a live-cell, kinetic assay using quantitative image analysis. Overexpression of more than 86 cDNAs (1.15%) conferred dramatic increases in the proliferation, as determined cell enumeration. These included several known growth regulators, as well as previously uncharacterized ones (LRRK1, Ankrd25). In addition, novel functional roles for two genes (5033414D02Rik, 2810429O05Rik), now termed *Gatp1* and *Gatp2*, respectively, were identified. Further analysis demonstrated that these encoded proteins promoted cellular proliferation and transformation in primary cells. Conversely, cells depleted for *Gatp1* underwent apoptosis upon serum reduction, suggesting that *Gatp1* is essential for cell survival under growth-factor-restricted conditions. Taken together, our findings offer new insight into the regulation of cellular growth and proliferation, and demonstrate the value and feasibility of assessing cellular phenotypes through genome-level computational image analysis.

[Supplemental material is available online at [www.genome.org](http://www.genome.org).]

Over the last decade, the rapid expansion in genome sequence has led to the identification of genes at a rate that far exceeds the capacity to understand their function (Adams et al. 2000; Lander et al. 2001; Venter et al. 2001; Waterston et al. 2002; Gibbs et al. 2004). Numerous large-scale gene profiling strategies permit the study of RNA expression dynamics at whole-genome levels (Schena et al. 1995; Lockhart and Winzler 2000; Su et al. 2004) and protein-protein interactions within entire proteomes (Pandey and Mann 2000; Gerlt 2002; MacBeath 2002). Furthermore, several tools aimed at directly assessing gene function in large-scale cell-based assays have recently been developed (Ziauddin and Sabatini 2001; Michiels et al. 2002; Albayrak and Grimm 2003; Arts et al. 2003; Aza-Blanc et al. 2003; Chanda et al. 2003; Chen et al. 2003; Fiscella et al. 2003; Iourgenko et al. 2003; Kiger et al. 2003; Berns et al. 2004; Boutros et al. 2004; Huang et al. 2004; Paddison et al. 2004; Zheng et al. 2004). These collections, comprised of nucleic acid and viral libraries arrayed in an addressable format, apply RNA interference- (RNAi) and overexpression-based methods of probing gene function (Paddison and

Hannon 2002; Carpenter and Sabatini 2004; Dorsett and Tuschl 2004; Grimm 2004; Steinmetz and Davis 2004). Although several genome-scale screens using these methodologies have been described, they have largely been restricted to single end-point reporter gene assays that measure whole-well luminescence and fluorescence (Aza-Blanc et al. 2003; Chanda et al. 2003; Chen et al. 2003; Fiscella et al. 2003; Iourgenko et al. 2003; Boutros et al. 2004; Huang et al. 2004; Paddison et al. 2004; Zheng et al. 2004) or visual inspection of image data (Kiger et al. 2003).

The high-throughput acquisition and analysis of cellular image data, or high-content screening (HCS), can enable the analysis of complex cellular events not readily measurable with current high-throughput detection methodologies. HCS systems provide an integrated approach whereby image acquisition, processing, and data analysis may be performed in an automated and high-throughput fashion to create a dynamic cellular database (Price et al. 2002; Abraham et al. 2004). This database may then be mined according to specified metrics that represent the cellular phenotype of interest. Data may further be selectively extracted (or "gated") from individual cells or subpopulations of cells, facilitating the analysis of heterogeneous cell populations. Although this platform presents several potential advantages over current detection methodologies for highly parallel assessments of cellular phenotypes, its utility in cell-based functional genomics applications has not been fully assessed.

<sup>5</sup>These authors contributed equally to this work.

<sup>6</sup>Corresponding author.

E-mail [schanda@gnf.org](mailto:schanda@gnf.org); fax (858) 812-1920.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.3889305>. Article published online before print in July 2005. Freely available online through the *Genome Research* Immediate Open Access Option.

Toward this end, we used HCS analysis to individually assess the gain-of-function phenotypes conferred by ~7000 genes in cellular proliferation and survival processes through overexpression of their respective cDNAs. This genetic approach yielded a spectrum of activities that affected cellular growth or viability, as measured by cell enumeration, under serum-depleted conditions. We further characterized a subset of these identified clones (*LRRK1*, *Ankrd25*, *Gatp1*, *Gatp2*, *Ptges*) in a series of validation assays and confirmed that they positively affect cell growth in multiple cellular backgrounds and induce oncogenic transformation in primary cells. Lastly, we used RNAi methodologies to show that *Gatp1*, which encodes a highly conserved protein of unknown function, is required for cell survival under reduced-serum conditions. This study demonstrates a successful integration of HCS technologies with high-throughput analyses of gene function, and suggests that these methodologies can facilitate the genetic analysis of cellular phenotypes on a global scale.

## Results

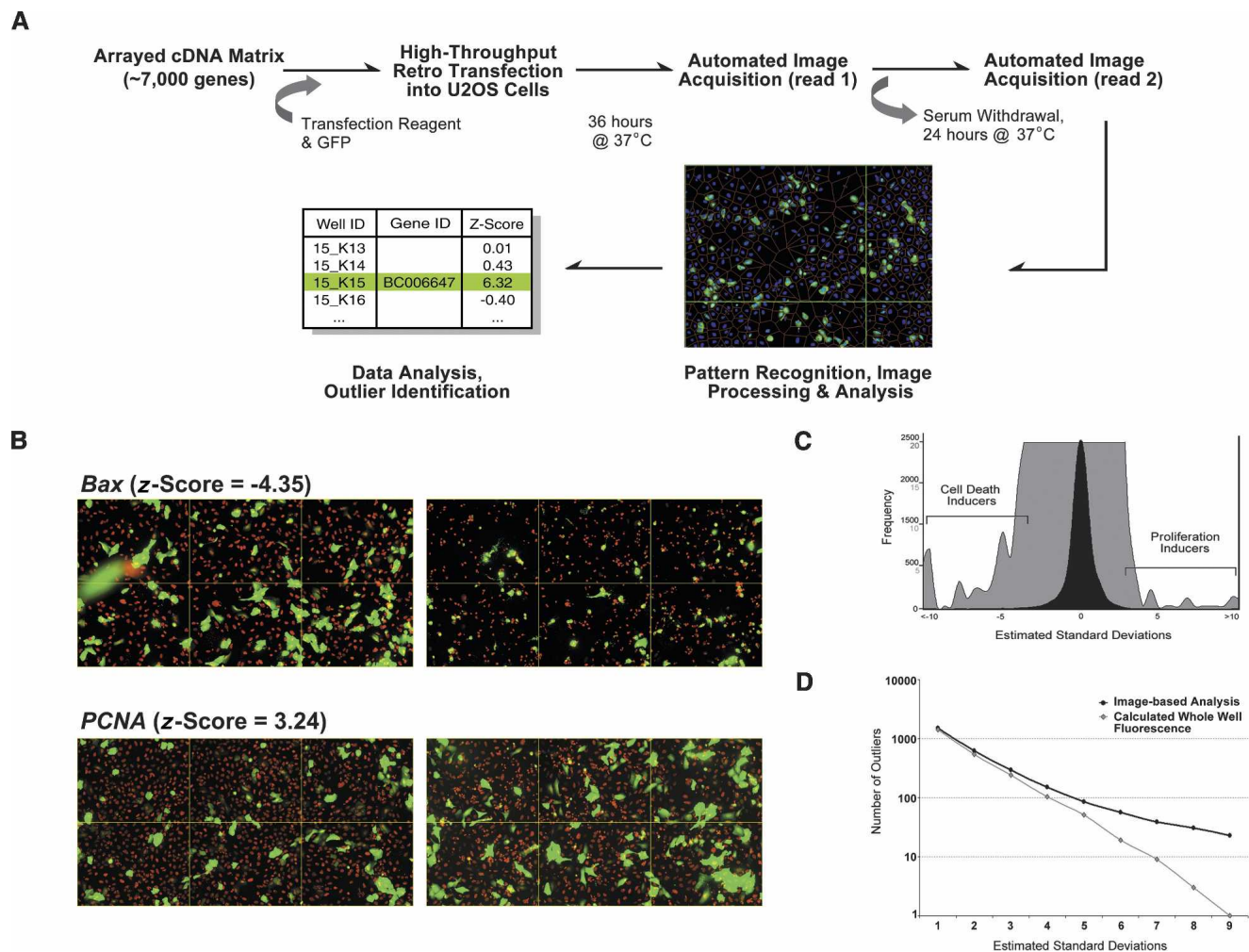
The precise control of cell proliferation and survival is a fundamental process required during mammalian development and is often deregulated in human disease (Brunner et al. 2003; Vermeulen et al. 2003; Alenzi et al. 2004). To identify novel gene activities that may participate in these cellular activities, we performed a genome-scale gain-of-function screen examining cell proliferation and viability under low-serum conditions. An arrayed cDNA expression library comprised of ~7000 full-length mammalian cDNAs was introduced, together with a green fluorescent protein (GFP) expression construct, into U2OS human osteosarcoma cells by means of a high-throughput transfection process (Fig. 1A). In contrast to loss-of-function siRNA screens that are used to identify endogenously expressed genes that are critical for a specific cellular process, this approach can be used to elucidate proteins that, when hyperactivated or ectopically expressed, are sufficient to modulate a particular phenotype. Thus, this assay was designed to identify genes that, when overexpressed, provide instructive cues directing the alteration of cellular proliferation and viability states.

To determine baseline transfection efficiencies for individual wells, all cells were marked with a DNA intercalating dye (Hoechst 33342) after 36 h, and image data for total cell counts and GFP-positive cells were collected (read 1) using a high-throughput fluorescence microscopy system (Q3DM EIDAQ100). To facilitate detection of proteins that would effect viability or growth, cells were then transferred to low-serum conditions. This resulted in a reduction of background cellular proliferation, and simultaneous sensitization to pro-apoptotic stimuli (data not shown). Finally, a subsequent imaging step (read 2) was performed to monitor alterations in cell number conferred by the overexpression of these individual cDNAs.

Image data were evaluated through the manipulation of a set of software modules (Cytoshop) that enabled the reduction of acquired image data to numerical descriptions of this cellular assay. Initially, nuclear images were corrected for shade distortion and background fluorescence through normalizing spatially segregated intensities and eliminating pixels that fell below the first (significant) mode of the signal distribution, respectively. These “corrected images” were then used to identify individual nuclei in a three-step process. First, to create a binary mask, a linear filter across all corrected nuclear images was applied (Price

et al. 1996). This was followed by an automatic histogram-based thresholding step that allowed for the production of binary images (Price et al. 1996). Lastly, a connected components algorithm was then used on these binary mask images to establish separate identifiers for individual nuclei (Price et al. 1996). Approximately 80% of unthresholded objects were successfully classified as cells using this analysis [Supplemental Table 2, see “% identified cells (IC)”], while the remaining 20% represented spatially overlapping sets of nuclei, cell debris, and DNA precipitate. Visual analysis typically resulted in cell numbers within 15% of the results derived from the imaging algorithm (data not shown). Each of the nuclei, also referred to as areas of interest (AOI), and their associated masks, were then inserted into an image database called an “image table.” These tables also stored several properties of each AOI, including corrected pixel values, characteristics of the nuclear masks, spatial location, the coordinates of the AOI’s bounding box, as well as others. Since the GFP signal was a cotransfection marker, the corrected GFP-channel (488-nm filter) AOI image pixels were selected relative to the nuclear AOI’s centroid coordinates ( $x$ ,  $y$ ). These nuclear location values were used to define an expanded bounding box of pixels from the corresponding coordinates ( $x$ ,  $y$ ) in the GFP-channel images. The 488-nm AOIs were then formed by clipping the appropriate image pixels from the corrected GFP-channel image and inserting these pixels into the image table in association with their respective nuclear AOIs. Finally, positive transfection events (GFP+ cells) were established in AOIs where the summation of pixels within the GFP channel exceeded a calculated threshold.

Thus, cell-by-cell measurements were extracted from dual-emission images corresponding to cell nuclei and GFP expression, and used to identify transfected and untransfected populations in each well. Gene activities that perturbed cellular proliferation kinetics were quantified by recording changes observed in the GFP cotransfected cell population at read 2 relative to read 1. The results included a distribution of growth phenotypes as depicted by a significance factor, or z-score (Fig. 1C). z-scores are a function of standard deviations from the plate average (z-score = 0), and reflect the variance in cellular proliferation rates conferred by the introduced cDNA. Therefore, transfected cDNAs that enhance cellular growth rates relative to plate mean values have z-scores >0, while those that severely retard cell division or reduce cell viability have z-scores <0. For example, the pro-apoptotic proteins Bad and Bax induced z-scores of  $-5.47$  and  $-4.35$  (Fig. 1B; Table 1; Kaufmann and Hengartner 2001; Cory and Adams 2002). Conversely, cotransfection of *PCNA*, previously shown to disrupt growth arrest under serum-starved conditions (Fukami-Kobayashi and Mitsui 1999), yielded a z-score of 3.24 (Fig. 1B; Table 1). Additional known modulators of cell growth and cell death identified in our analyses include *EZH2*, *CDK9*, *Mcl1*, and *Bik1*, which produced z-scores of 5.16, 3.56, 3.09, and  $-7.93$ , respectively (Table 1; Hegde et al. 1998; Craig 2002; Varambally et al. 2002; Vrana et al. 2002; Bracken et al. 2003). In all, we identified 86 cDNAs (1.2%) that positively affected proliferative activities under reduced-serum conditions to a level more than three standard deviations from the trimmed experimental mean, and 84 cDNAs (1.15%) that restricted growth or induced cell death by a z-score of  $<-3.25$  (Supplemental Table 2; data not shown). Analysis of Gene Ontology (GO) assignments and IntroPro domains contained within these identified proteins suggest that several of these molecules modulate cellular activities through assembly in dimeric or multimeric protein complexes and also at the level of transcriptional regulation (Supplemental Table 3).



**Figure 1.** Screen for mammalian growth regulatory factors by high-content imaging. (A) Screen workflow. Here, 7364 individually arrayed mammalian cDNAs were cotransfected with a GFP expression construct into U2OS cells. Cells were imaged in an automated manner using the Q3DM EIDAQ100 automated fluorescent microscope at 36 h post-transfection (read 1). Then, 24 h after exchanging normal culturing media for serum-depleted media, the image acquisition procedure was repeated (read 2). Image sets were analyzed using Cytoshop software. (B) Representative image table montages. Images from wells cotransfected with *GFP* and *Bax* or *PCNA* are shown together with their z-scores. Image sets in *left* panels were taken at time of serum-withdrawal. *Right* panel images were taken after 24 h of continued growth. Hoechst 33342 DNA stain is shown in red. (C) Frequency distribution of observed gene activities. Changes observed in GFP-positive cell growth in each well are presented as estimated standard deviations, and reflect the variance in well-to-well proliferation rates induced by a gene activity, relative to the plate mean. To enable the visualization of outliers, an enlargement of the total screen distribution (black) is shown in gray. Frequencies for each representation are indicated on the y-axis, respectively. (D) The relative sensitivity of whole-well fluorescence and image-based detection methodologies was assessed by determining the cumulative number of statistically significant outlier events (i.e., apoptosis and proliferation) as a function of estimated standard deviations from respective experimental medians (see Methods).

To provide an approximation of false-negative rates in this assay, we analyzed the activities of control wells arrayed in each plate containing known regulators of apoptosis (Bad and Bax, Supplemental Table 2, wells O21–O24 and P21–P24). Of these controls, 71% (113/159) fell outside the cut-off range ( $z < -3.25$ ) we used in the elucidation of our hits. However, if this standard is relaxed to a threshold of  $z < -2.5$ , >90% (144/159) of these control wells would be identified by our methodologies. These results suggest that, although our selection criteria were stringent, the overall performance of the assay and subsequent image analysis is robust and possesses a marginal false-negative rate.

Transfection efficiencies at read 1 were observed to be 30.2% on average, but varied widely from well to well, ranging from <math><1\%</math> to >80%. This distribution is a component of within screen

variability, which may be derived from several factors in cell-based assay systems. These include heterogeneous cellular responses, plate variations, inconsistencies in library preparation, and other inherent methodological limitations (Lundholt et al. 2003). An important advantage of using image-based methods in functional genomics studies is the ability to differentiate affected and unaffected cells through *in silico* segregation of mixed cell populations. This enables the precise measurement of signal present only in individual or distinct subpopulations of cells (i.e., transfected cells), and simultaneously comparing their activities to nonresponders (i.e., nontransfected cells) within the same well. To better understand the potential advantages of this analysis methodology, we compared this approach to data generated through solely calculating whole-well fluorescent output, which

**Table 1.** Growth regulators identified in primary screen analysis

MGC clone ID	Gene name	Gene description	z-score
*	Bad	<i>Mus musculus</i> Bcl-associated death promoter	-5.47
*	Bax	<i>Mus musculus</i> Bcl2-associated X protein	-4.35
BC003732	Biklk	<i>Mus musculus</i> Bcl-2 interacting killer-like protein	-7.93
BC005778	Pcna	<i>Mus musculus</i> proliferating cell nuclear antigen	3.24
BC010858	EZH2	<i>Homo sapiens</i> enhancer of Zeste homolog 2 ( <i>Drosophila</i> )	5.16
BC005633	Mcl1	<i>Mus musculus</i> myeloid cell leukemia sequence 1	3.09
BC001968	CDK9	<i>Homo sapiens</i> cyclin-dependent kinase 9 (CDC2-related kinase)	3.56
BC005408	LRRK1	<i>Homo sapiens</i> leucine-rich repeat kinase 1	5.30
BC023148	2810429O05Rik	<i>Mus musculus</i> RIKEN cDNA 2810429O05 gene	3.64
BC024960	Ptges	<i>Mus musculus</i> prostaglandin E synthase	3.68
BC024953	5033414D02Rik	<i>Mus musculus</i> RIKEN cDNA 5033414D02 gene	3.02
BC006647	Ankrd25	<i>Mus musculus</i> ankyrin repeat domain 25	6.32

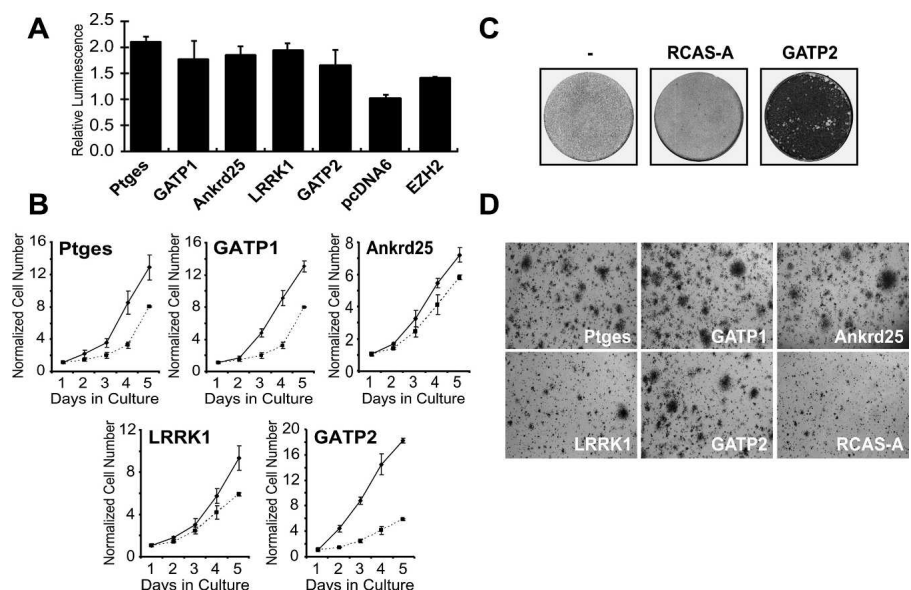
\*Control wells.

Changes in cellular growth rates conferred by cDNAs at serum-depleted conditions were recorded by high content image analysis. Results are presented as z-scores. Gene names, descriptions and MGC accession numbers are also provided.

is comparable to detection by a fluorescent plate reader. We find that image-based analysis of individual cells provides an overall increase in sensitivity, as assessed through a comparison of statistically significant outliers detected by each method (Fig. 1D). Single-cell analysis was dramatically more robust in extrapolating data from wells with low transfection efficiencies (Supplemental Fig. 4A), suggesting that this approach may be especially suitable for rare-event detection and high-variability assays. To further investigate the basis for this increased sensitivity of outlier detection, we compared the distribution of activities deduced from each analysis method (Supplemental Fig. 4B). Image-based analysis resulted in an ~10-fold compression of standard deviation across the assay when compared to whole-well fluorescent output (data not shown). When we assessed activities from this quantitative imaging assay to a parallel cDNA screen using a whole-well luminescence detection methodology, we observed a similar (13-fold) increase in the screen-wide standard deviation in the latter assay (Supplemental Fig. 4C). Since reduction in variability results in a higher confidence (smaller confidence interval) for probabilistic identification of outliers ("hits"), it is likely that the observed advantage in sensitivity for image-based analysis reflects an increase in the precision of acquisition and quantification of reporter events. Thus, we conclude the use of single-cell microscopic analysis results in superior data extraction and outlier detection in comparison to currently available high-throughput fluorescent detection protocols.

To ascertain the robustness of the screen and image analysis, and to determine false-positive rates, we next conducted a secondary analysis of proteins

that were most proficient in inducing cell death or proliferation under serum-depleted conditions. In all, 92.5% (37/40) of sequence-verified clones demonstrated similar activities to those observed in the initial assay (Supplemental Table 4; data not shown). We postulate that the observed 7.5% false-positive rate results from inherent variabilities associated with high-throughput cellular genomics methodologies. We selected 15 cDNAs that, when over-expressed, exhibited robust proliferation-inductive phenotypes for further investigation (Supplemental Table 5). These included several gene products of unknown function including LRRK1, Ankrd25, Gatp1, and Gatp2, as well as prostaglandin E synthase (Ptges), an inducible enzyme that functions downstream of cyclooxygenase-2 (COX-2) in the prostaglandin E2 (PGE2) biosynthetic pathway (Levy 1997; Williams et al. 1999). To further confirm their proliferative potential under serum-depleted conditions, the activities of these encoded proteins were also examined in a human embryonic kidney cell line,



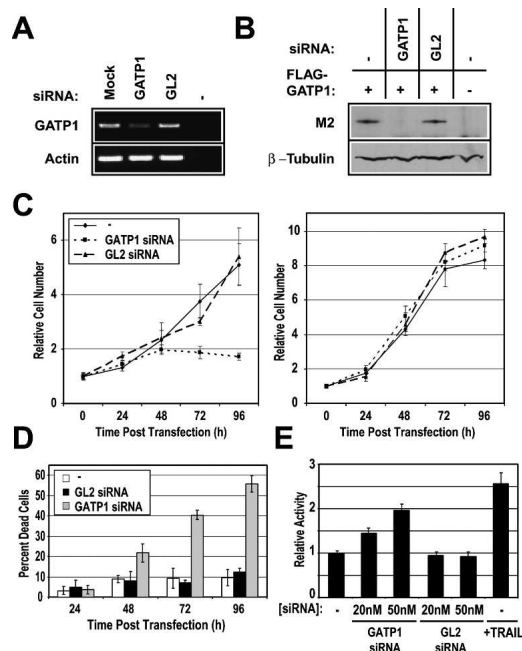
**Figure 2.** Validation of growth-activating and transforming abilities of identified proteins. (A) Over-expression of *LRRK1*, *Ankrd25*, *Gatp1*, *Gatp2*, and *Ptges* in HEK293 cells induces cell proliferation under serum-depleted conditions. Changes in luminescence, indicative of intrawell ATP levels, are presented relative to pcDNA6-transfected control. (B) Ectopic *LRRK1*, *Ankrd25*, *Gatp1*, *Gatp2*, and *Ptges* expression (solid lines, respectively) confers increased growth kinetics in primary cells, as compared to an empty vector control (dotted lines). Primary CEFs expressing growth activators were seeded into 384-well plates at a density of 500 cells per well. Cell proliferation was monitored in culture for 5 d by quantitative fluorescence imaging with the Q3DM EIDAQ100. Images were analyzed using Cytoshop, and cell numbers were determined through Hoechst 33342 staining. (C) CEFs infected with *Gatp2*-expressing RCAS vector grown under serum rich agar show increased saturation density compared to uninfected (-) or vector-infected (RCAS-A) controls. CEF monolayers were infected with activator-expressing virus, or the parental RCAS control virus, and subsequently overlaid with agar medium (Bister et al. 1977; Bos et al. 1990). Foci were visualized after 3–4 wk in culture by crystal violet staining. (D) CEFs ectopically expressing *LRRK1*, *Ankrd25*, *Gatp1*, *Gatp2*, and *Ptges* were embedded in soft agar, and colony growth was monitored over 4 wk (Bister et al. 1977). Infection with the parental viral vector RCAS-A was used as a negative control.

HEK293. Ectopic expression of 11/15 clones reproducibly enhanced growth rates >1.5-fold, as measured by monitoring cellular metabolic activity (Fig. 2A; Supplemental Table 5). These findings validate the pro-proliferative activities of these molecules, and demonstrate that these gene products participate in regulating cell cycle activities in multiple cell types.

The abilities of these molecules to increase proliferation rates and induce oncogenic phenotypes were further assessed in primary cells. The retrovirus-mediated transformation of chicken embryo fibroblasts (CEFs) by oncogenes has been shown to result in cell immortalization, anchorage-independent growth, loss of contact inhibition, and increases in cell density at saturation (Bos et al. 1990; Vogt 2001). Avian retroviruses expressing these selected cDNAs were thus used to transduce these factors into primary fibroblasts. Analysis of CEFs stably expressing these genes demonstrated that several of these molecules can confer a growth advantage in this primary cell type (Fig. 2B; Supplemental Table 5), consistent with their putative role in cellular proliferation. Furthermore, ectopic expression of *Gatp2* reproducibly potentiated CEF growth to saturation densities well beyond that observed with controls, resulting in large areas of multilayered growth (Fig. 2C). The introduction of *Gatp1*, *LRRK1*, and *Ptges* similarly led to the formation of localized focal lesions (data not shown; Supplemental Table 5).

Expression of *LRRK1*, *Ankrd25*, *Gatp1*, *Gatp2*, and *Ptges* was also demonstrated to promote anchorage-independent growth in CEF cells, underscoring their oncogenic potential. While these five clones induced this transformation phenotype to various degrees (Fig. 2D), *Gatp1*, *Gatp2*, and *Ankrd25* directed substantially more aggressive colony formation in soft agar culture, as compared to infection with the empty RCASA parental virus. Taken together, these data confirm the proliferative activities of the putative growth-activators identified in our primary screen analysis, and establish their oncogenic potential in primary cells.

Since forced expression of *Gatp1* resulted in increased proliferation rates in both immortalized and primary cells, we further investigated its necessity for cellular growth using RNA interference methodologies. Inhibition of *Gatp1* expression was achieved in NIH3T3 cells using synthetic small-interfering RNA (siRNA) oligonucleotides, which reduced target mRNA and protein levels, respectively (Fig. 3A,B). Importantly, *Gatp1* transcript and protein levels were unaffected by *GL2* control siRNAs. Cells depleted for *Gatp1* under standard growth conditions demonstrated normal growth kinetics comparable to mock- or control-siRNA-transfected cells (Fig. 3C, right panel). In contrast, the inspection of *Gatp1* siRNA-transfected cultures under reduced-serum conditions (1% FCS) revealed a severe reduction of cellular proliferation (Fig. 3C, left panel). Cells transfected with the *GL2* control siRNAs, as well as siRNAs directed against *Gatp2*, *LRRK1*, and *Ptges*, under these growth conditions displayed proliferation rates similar to mock-transfected cells (Fig. 3C; Supplemental Table 5; data not shown). To understand the basis for this diminished proliferative capacity, cell cycle profiles of NIH3T3 cells, transfected with *Gatp1* or control siRNA, were measured by high-content image analysis at 24-h intervals post-transfection (Wan et al. 2004). Under reduced-serum conditions, treatment with *Gatp1* siRNA led to the induction of a prominent sub-G<sub>1</sub> fraction (<2N DNA content) of cells, which appeared at later times post-transfection (Fig. 3D), indicative of apoptotic induction. This is consistent with the demonstration that siRNA-mediated abrogation of *Gatp1* expression activates primary effector caspase-3 and caspase-7 (Fig. 3E; Kaufmann and Hengartner 2001; Cory and



**Figure 3.** *Gatp1* is required for cell growth and viability under reduced serum conditions. Target mRNA and protein knockdown by *Gatp1* siRNA was confirmed by (A) semiquantitative RT-PCR and (B) immunoblot analysis. RNAs harvested from NIH3T3 cells transfected with the indicated siRNA were analyzed using *actin* and *Gatp1* gene-specific primers. Lysates prepared from NIH3T3 cells transfected with *Gatp1* or control (*GL2*) siRNAs, along with a cDNA encoding *Gatp1*-Flag fusion protein, as indicated, were immunoblotted with M2 anti-Flag (B, top panel) and  $\beta$ -tubulin antibodies (B, bottom panel). (C) *Gatp1* expression is required for proliferation under low-serum conditions. Approximately 1000 NIH3T3 cells seeded into 384-well plates at low (left panel) and normal (right panel) serum were mock-transfected (—●—) or transfected with *Gatp1* (—■—) or *GL2* (—▲—) siRNAs. Cell numbers at each condition were evaluated by imaging over the indicated time period. Cells were visualized by staining with Hoechst 33342. (D) Abrogation of *Gatp1* expression at low-serum conditions increases cell death. Images from C were analyzed to evaluate the effects of *Gatp1* and *GL2* (control) siRNAs on cell cycle distribution (Wan et al. 2004). Histograms plotting the number of events versus total nuclear fluorescence for sample wells were used to define boundaries for the sub-G<sub>1</sub> population. Results are presented as percentages of total cells in each well. Samples were prepared in quadruplicate for each condition, and replicate wells were averaged. (E) siRNA-mediated depletion of *Gatp1* activates primary effector caspase-3 and caspase-7. NIH3T3 cells transfected with the indicated siRNA at 1% serum were analyzed using the Apo-ONE Homogeneous Caspase-3/7 Assay kit. Untransfected cells treated with recombinant human TRAIL were analyzed as positive control.

Adams 2002). Thus, we conclude that suppression of *Gatp1* activity under growth-factor-limited conditions results in the induction of an apoptotic cascade.

Inspection of *Gatp1* peptide sequence showed that this protein contains an uncharacterized conserved domain (KOG4544), and expression analysis of this gene and its homolog across mouse and human tissues, respectively, revealed elevated mRNA levels in both lymphoid and epithelial compartments (Supplemental Fig. 5A,B). Interestingly, these tissues types are associated with cellular proliferation and turnover. Taken together, these findings suggest that this protein, which is highly conserved between human, mouse, *Caenorhabditis elegans*, and *Drosophila melanogaster* (Supplemental Fig. 5C), may function as a critical mediator of cellular proliferation and apoptosis in response to levels of extracellular mitotic stimuli.

## Discussion

To date, the functions of approximately one-half of human and one-third of mouse genes have been described in Medline (Su et al. 2004). Many of these ascribed functions are determined based on primary amino acid sequence or expression analysis, and provide little insight into the cellular roles of these genes (Su et al. 2004). Technological advances in the automation of cellular imaging and quantitative image analysis have enabled the application of fluorescence-microscopy-based methods toward high-throughput biological inquiries (Price et al. 2002; Abraham et al. 2004). In comparison to single-measurement well-based assays, HCS systems provide added dimensionalities to the large-scale analyses of gene function, including the ability to record measurements from multiple assay endpoints, from either live or fixed cells, and from both individual cells and cell populations. In the current study, we have used these unique capabilities in a genome-scale functional screen to identify previously unreported modulators of cellular proliferation and transformation. However, several potential biological artifacts and limitations may be associated with this method of analysis. These include potential false negatives derived from inaccuracies in library manipulation, inconsistencies in the expression of the full-length sequenced clones, and possible inactivity of proteins in orthologous biological backgrounds (i.e., mouse cDNAs expressed in human cells). To address some of these concerns and to evaluate a subset of potent growth-inducing proteins (LRRK1, Ankrd25, Gatp1, Gatp2, and Ptges) in additional cellular assays, we used an RCAS retroviral expression system to ectopically express these genes at lower and more physiological levels. Activities of these five encoded proteins were found to augment cellular proliferation in primary cells, and promote phenotypes reflective of oncogenic transformation, raising the possibility that these genes may contribute to the pathogenesis of certain cancers. Further investigation demonstrated that Gatp1 is required for cell survival under serum-restricted conditions. Gatp1 contains an uncharacterized protein domain (KOG4544), and is highly conserved in mammalian species. We have also demonstrated that the pro-proliferative activity of Gatp1 can be recapitulated in both human and avian cell types. Thus, we postulate that Gatp1 is an evolutionarily conserved sensor of extracellular growth cues, and is a critical mediator of the cellular response to these external signals.

Our findings further demonstrate the feasibility of conducting high-throughput genomics using fully automated single-cell analysis. We have shown that this image-based approach has considerable advantages over “whole-well” detection methodologies, including significant augmentations in sensitivity, as well as rare-event detection. Furthermore, this technology can uniquely enable the quantitative analysis of certain discrete molecular phenotypes at the level of an individual cell, such as changes in subcellular localization, morphological alterations, cellular migration, or cytoskeletal rearrangements, on a global scale. Although there exists a considerable number of other cellular phenotypes currently addressable by HCS (Price et al. 2002; Abraham et al. 2004), novel assays will likely require the generation of additional image analysis algorithms, or modifications of existing ones. Taken together with the implementation of functional genomics strategies in cell-based assays, these developments should considerably accelerate the discovery of new gene activities, and facilitate the systematic and comprehensive molecular description of biological systems at the cellular level.

## Methods

### High-throughput (retro) transfection and image acquisition

High-throughput transfections of the expression-ready subset of the Mammalian Gene Collection were performed essentially as described (Chanda et al. 2003; Huang et al. 2004; <http://function.gnf.org>). This library, consisting of 7364 unique full-length mammalian cDNAs in the mammalian expression vector pCMV-Sport 6, was spotted into 384-well plates such that each well contained 62.5 ng of an individual cDNA of known identity. In a semiautomated process, cDNAs were incubated with a non-liposomal transfection reagent (Fugene6; Roche Applied Sciences) and 20 ng of pZs-Green-C1, a construct that contains the cytomegalovirus immediate-early promoter driving a human codon-optimized variant of the *Zoanthus* green fluorescent protein (GFP; Clontech). Then ~4000 U2OS human osteosarcoma cells (ATCC) were introduced into each well to complete the transfection process. pZs-Green-C1 was used to mark the cotransfected population.

After incubating for 36 h at 37°C and 5% CO<sub>2</sub>, normal culturing medium (DMEM + 10% fetal calf serum) was exchanged for phenol-red-free DMEM supplemented with 0.5% FCS using the Molecular Devices EMBLA 384-well plate washer. Medium was supplemented with an approximate concentration of 1.0 µg/mL Hoechst 33342 (Molecular Probes) to permit visualization of nuclei.

Cell images were then acquired at this time (read 1) using the Beckman-Coulter/Q3DM EIDAQ100 (Q3DM) automated high-content imaging system affixed with a Nikon Super Fluor 10×, 0.5 numerical aperture (NA) objective, and CoHu 640 × 480 pixel charge-coupled device (CCD) camera with 9.9 µm<sup>2</sup> pixels. Six images from adjacent fields were acquired in each channel (Hoechst 33342 and GFP), representing 1.81 mm<sup>2</sup>, or ~17% of the total well area. These conditions typically resulted in the analysis of ~1000 cells per well. For each field, autofocus was first performed and then the Hoechst channel was acquired. Subsequently, upon filter wheel rotation, the GFP channel was also acquired. From the approximate middle of each well, this acquisition sequence was repeated until two rows of three adjacent fields were acquired from both channels. Then, 24 h post-serum withdrawal, this image acquisition procedure was repeated (read 2).

### Image and data analysis

Image analysis was performed using Cytoshop, an integrated image analysis and data visualization tool that generates numerical descriptions of cellular assay data from image sets. We used this software to locate each cell in an image by segmenting images into regions of individual cells. These regions of pixels were then used for the extraction of an assortment of cellular measurements including cell number, size, intensity, and distribution of labeled molecules. Measurements are subsequently collated into a relational database (image table), which may then be analyzed according to desired measurements. Additional details on Cytoshop algorithms may be found in Price et al. (2002) and references therein (Abraham et al. 2004).

Image tables were gated using nuclear morphology filters, based on fluorescent intensity and object area and other features, to first identify single cells (vs. cell clusters or cellular debris) in each well (see “Identified Cells” in Supplemental Table 2). This population was subsequently separated into the transfected and untransfected population by uniformly setting GFP intensity thresholds (see “GFP+ Cells” in Supplemental Table 2). The proportion of GFP-positive cells imaged in each well at both reads 1 and 2 was thus derived, and subsequently used to calculate the

fractional change in GFP signal in each well [fractional change = (fraction of GFP-positive cells at read 2)/(fraction of GFP-positive cells at read 1)]. Fractional changes were then log-transformed ( $\log_{10}$ ) and divided by a constant ( $\log_{10}2$ ) to reflect the number of population doublings (alternatively halvings) observed from read 1 to read 2. These values were used to derive z-scores by calculating the variance (in standard deviations) from the trimmed plate mean for each value in the analysis set. cDNAs that enabled continued cell proliferation under low-serum conditions with an arbitrarily selected z-score threshold ( $>3.0$ ) were selected for further analysis. Those that reproducibly elevated fractional change relative to vector control were retained for additional validation studies.

Fractional changes of GFP fluorescence (Fig. 1D; Supplemental Fig. 4A,B) between kinetic reads (READ2/READ1) were extrapolated based on the percentage of GFP+ cells in a given well (imaged-based analysis) or calculated fluorescent output of a well (whole-well fluorescence). The latter was determined by simply multiplying average GFP intensity within a well by the number of GFP+ cells in the same well. GFP intensity was noted to be at saturating levels in certain wells (data not shown); thus, in some cases, this calculation becomes a function of cell number, and may result in the overestimation of whole-well fluorescence detection sensitivity. The values from each methodology were median-normalized by plate, and then log-transformed ( $\log_{10}$ ) and divided by a constant ( $\log_{10}2$ ). We then computed absolute deviation from the median for each value in the analysis set, and the median of all these absolute deviations (MAD) was used as a robust estimate of the standard deviation. Multiples of this estimate were used to establish outlier thresholds.

#### Validation of screen hits using the CellTiter-Glo Luminescent Cell Viability Assay

To provide additional validation, the effect of primary screen hits on cell proliferation was assessed in an alternate cellular background using the CellTiter-Glo Luminescent Cell Viability Assay (Promega). Briefly, 62.5 ng of each screen hit was spotted into 384-well plate format and retrotransfected into ~2000 HEK293 cells using Fugene6. Transfections were performed at low (DMEM, 0.5% FCS) serum conditions, and incubated for 48 h at 37°C. CellTiter-Glo reagent was then added to cells according to the manufacturer's specifications, and luminescence output was measured using the Acquest Plate Reader (LJL Biosystems). The relative luminescence recorded in each well, which correlates with the number of metabolically active cells, was normalized to pcDNA6 vector-transfected (Invitrogen) control wells.

#### Construction of growth activator clone-expressing chicken embryo fibroblast cell lines and growth kinetic analysis

Transfer of cDNA clones into Gateway-compatible, subtype A replication-competent avian sarcoma leucosis virus vectors was performed using Gateway LR clonase (Invitrogen). Upon sequence verification, recombinant viral vectors were transfected into primary chicken embryo fibroblasts using DMSO shock as described (Bister et al. 1977) to generate stable growth activator-expressing cell lines.

Viral supernatants were harvested from stable lines and used to infect fresh CEFs (Bister et al. 1977; Bos et al. 1990). After serially passaging infected cells three times, growth kinetic analyses were performed using the Beckman-Coulter/Q3DM EIDAQ100 quantitative high-throughput fluorescence microscopy system. Briefly, cells were seeded at ~500 cells per well in 384-well plate format, and maintained for 5 d in culture at 37°C in Ham F10 cloning media (Bister et al. 1977). Cells were visual-

ized by Hoechst 33342 staining, and growth was monitored by imaging daily using a 10 $\times$ , 0.5 NA objective, and Cohu CCD camera. Effects of stably expressing select growth-activator cDNA clones were quantitated using Cytoshop by first gating image data with a nuclear morphology filter to identify individual cells, and subsequently setting nuclear staining intensity thresholds.

#### Focus formation and soft agar assays in primary chicken embryo fibroblasts

Focus formation assays were performed as described (Bister et al. 1977; Bos et al. 1990). CEFs were seeded at a density of  $5 \times 10^5$  cells/well of a six-well plate and infected with activator-expressing viruses. Cells were then overlaid with agar medium and maintained in culture for 3–4 wk. Foci were subsequently visualized by staining with crystal violet. Soft agar assays were performed as described (Bister et al. 1977).

#### Transfection and validation of siRNAs

A synthetic siRNA SMARTpool, comprised of a pool of four independent SMARTselected siRNA duplexes targeting *Gatp1*, was purchased from Dharmacon. The sequences of individual siRNAs in the pool that demonstrated efficacy are 5'-AACAAACAGGA GUUCAUGGUUA-3' and 5'-AACCAACAGGAGUUCAUGGUU-3'. All siRNA transfections were performed at 20 or 50 nM concentration using *TransIT*-TKO following the manufacturer's specifications (Mirus). Transfection efficiencies were determined by performing parallel transfections with a Cy3-conjugated siRNA targeting the *GL2* firefly luciferase gene (Dharmacon), and estimated to be >90% by fluorescence microscopy.

To assess the efficacy of target protein knockdown conferred by the SMARTpool, cDNA clone *Gatp1* was Flag-epitope tagged and transfected in the presence or absence of the indicated SMARTpool/siRNA into NIH3T3 mouse embryo fibroblasts (ATCC). Approximately  $2 \times 10^5$  NIH3T3 cells were seeded into six-well plates and either mock-transfected, or transfected with an siRNA targeting *Gatp1* or *GL2*. After incubating for 24 h, medium was replaced, and cells were transfected with Flag-*Gatp1* using *TransIT*-3T3 reagent (Mirus) according to the manufacturer's instructions. Then, 48 h post-transfection with DNA, cells were harvested, and the resultant extracts were resolved on 10% NuPAGE Novex Bis-Tris gels (Invitrogen). Immunoblot analysis was performed upon protein transfer to polyvinylidene difluoride membranes (Amersham Biosciences), using the M2 anti-Flag (Sigma-Aldrich) and D-10 anti- $\beta$ -tubulin monoclonal antibodies (Santa Cruz Biotechnology).

Target mRNA knockdown was further validated using semi-quantitative RT-PCR. RNAs from siRNA-transfected cells were harvested using the RNeasy mini-kit (Qiagen). cDNA was then produced using Superscript III reverse transcriptase (Invitrogen) with 1  $\mu$ g of total RNA as template. Ensuuing cDNA products were used in PCR reactions with primer sets 5'-ATATCGCTGCGCTG GTCGTC-3' and 5'-ACATAGGAGTCCTTCTGACC-3' to amplify cytoplasmic  $\beta$ -actin, and 5'-GGAGTTCATGGTTACGCATG-3' and 5'-TAGGAACGAGAAAGGCTGGC-3' to amplify the RIKEN cDNA 5033414D02 gene.

#### Requirement for activator clone *Gatp1* in cell growth and survival

The effect of gene knockdown on cellular growth rates was determined using imaging methods similar to those described above. Briefly, ~1000 NIH3T3 cells were seeded into 384-well tissue culture plates under normal (10%) and low (1%) serum conditions, and transfected with the indicated siRNA. Transfected cells were maintained in culture for 4 d at 37°C, and cell

growth was monitored daily by quantitative imaging. Cell cycle studies were additionally performed using these image sets in Cytoshop (Wan et al. 2004). Histograms evaluating total nuclear intensity and event frequency for representative wells were used to define fluorescence boundaries for sub-G<sub>1</sub>, G<sub>1</sub>, and S/G<sub>2</sub>/M populations. Fractions of cells in each group were determined for all wells, and replicate wells were averaged for each condition.

siRNA-transfected NIH3T3 cells were evaluated for the induction of apoptosis using the Apo-ONE Homogeneous Caspase-3/7 Assay kit (Promega) according to the manufacturer's instructions. Cells treated with 1 µg/mL recombinant TRAIL (Calbiochem) for 4 h were analyzed in parallel to serve as a positive control (Aza-Blanc et al. 2003).

## Acknowledgments

We thank Hilmar Lapp, Nicole Johnson, Abel Gutierrez, Paul DeJesus, Myleen Medina, and Brendan Smith for providing excellent technical assistance; and Trey Sato, Garrett Hampton, and Tim Moran for helpful comments on the manuscript and advice. This work is supported by funding from the Novartis Research Foundation and National Institutes of Health research grants CA42564, CA79616, and CA78230 to P.K.V.

## References

- Abraham, V.C., Taylor, D.L., and Haskins, J.R. 2004. High content screening applied to large-scale cell biology. *Trends Biotechnol.* **22**: 15–22.
- Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., et al. 2000. The genome sequence of *Drosophila melanogaster*. *Science* **287**: 2185–2195.
- Albayrak, T. and Grimm, S. 2003. A high-throughput screen for single gene activities: Isolation of apoptosis inducers. *Biochem. Biophys. Res. Commun.* **304**: 772–776.
- Alenzi, F.Q. 2004. Links between apoptosis, proliferation and the cell cycle. *Br. J. Biomed. Sci.* **61**: 99–102.
- Arts, G.J., Langemeijer, E., Tissingh, R., Ma, L., Pavliska, H., Dokic, K., Dooijes, R., Mestic, E., Clasen, R., Michiels, F., et al. 2003. Adenoviral vectors expressing siRNAs for discovery and validation of gene function. *Genome Res.* **13**: 2325–2332.
- Aza-Blanc, P., Cooper, C.L., Wagner, K., Batalov, S., Deveraux, Q.L., and Cooke, M.P. 2003. Identification of modulators of TRAIL-induced apoptosis via RNAi-based phenotypic screening. *Mol. Cell* **12**: 627–637.
- Berns, K., Hijmans, E.M., Mullenders, J., Brummelkamp, T.R., Velds, A., Heimerikx, M., Kerkhoven, R.M., Madiredjo, M., Nijkamp, W., Weigelt, B., et al. 2004. A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428**: 431–437.
- Bister, K., Hayman, M.J., and Vogt, P.K. 1977. Defectiveness of avian myelocytomatosis virus MC29: Isolation of long-term nonproducer cultures and analysis of virus-specific polypeptide synthesis. *Virology* **82**: 431–448.
- Bos, T.J., Monteclaro, F.S., Mitsunobu, F., Ball Jr., A.R., Chang, C.H., Nishimura, T., and Vogt, P.K. 1990. Efficient transformation of chicken embryo fibroblasts by c-Jun requires structural modification in coding and noncoding sequences. *Genes & Dev.* **4**: 1677–1687.
- Boutros, M., Kiger, A.A., Armknecht, S., Kerr, K., Hild, M., Koch, B., Haas, S.A., Heidelberg Fly Array Consortium, Paro, R., and Perrimon, N. 2004. Genome-wide RNAi analysis of growth and viability in *Drosophila* cells. *Science* **303**: 832–835.
- Bracken, A.P., Pasini, D., Capra, M., Prosperini, E., Colli, E., and Helin, K. 2003. EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. *EMBO J.* **22**: 5323–5335.
- Brunner, T. and Mueller, C. 2003. Apoptosis in disease: About shortage and excess. *Essays Biochem.* **39**: 119–130.
- Carpenter, A.E. and Sabatini, D.M. 2004. Systematic genome-wide screens of gene function. *Nat. Rev. Genet.* **5**: 11–22.
- Chanda, S.K., White, S., Orth, A.P., Reisdorph, R., Miraglia, L., Thomas, R.S., DeJesus, P., Mason, D.E., Huang, Q., Vega, R., et al. 2003. Genome-scale functional profiling of the mammalian AP-1 signaling pathway. *Proc. Natl. Acad. Sci.* **100**: 12153–12158.
- Chen, C., Grzegorzewski, K.J., Barash, S., Zhao, Q., Schneider, H., Wang, Q., Singh, M., Pukac, L., Bell, A.C., Duan, R., et al. 2003. An integrated functional genomics screening program reveals a role for BMP-9 in glucose homeostasis. *Nat. Biotechnol.* **21**: 294–301.
- Cory, S. and Adams, J.M. 2002. The Bcl2 family: Regulators of the cellular life-or-death switch. *Nat. Rev. Cancer* **2**: 647–656.
- Craig, R.W. 2002. MCL1 provides a window on the role of the BCL2 family in cell proliferation, differentiation and tumorigenesis. *Leukemia* **16**: 444–454.
- Dorsett, Y. and Tuschl, T. 2004. siRNAs: Applications in functional genomics and potential as therapeutics. *Nat. Rev. Drug Discov.* **3**: 318–329.
- Fiscella, M., Perry, J.W., Teng, B., Bloom, M., Zhang, C., Leung, K., Pukac, L., Florence, K., Concepcion, A., Liu, B., et al. 2003. TIP, a T-cell factor identified using high-throughput screening increases survival in a graft-versus-host disease model. *Nat. Biotechnol.* **21**: 302–307.
- Fukami-Kobayashi, J. and Mitsui, Y. 1999. Overexpression of proliferating cell nuclear antigen in mammalian cells negates growth arrest by serum starvation and cell contact. *Jpn. J. Cancer Res.* **90**: 286–293.
- Gerlt, J.A. 2002. "Fishing" for the functional proteome. *Nat. Biotechnol.* **20**: 786–787.
- Gibbs, R.A., Weinstock, G.M., Metzker, M.L., Muzny, D.M., Sodergren, E.J., Scherer, S., Scott, G., Steffen, D., Worley, K.C., Burch, P.E., et al. 2004. Genome sequence of the brown Norway rat yields insights into mammalian evolution. *Nature* **428**: 493–521.
- Grimm, S. 2004. The art and design of genetic screens: Mammalian culture cells. *Nat. Rev. Genet.* **5**: 179–189.
- Hegde, R., Srinivasula, S.M., Ahmad, M., Fernandes-Alnemri, T., and Alnemri, E.S. 1998. Blk, a BH3-containing mouse protein that interacts with Bcl-2 and Bcl-xL, is a potent death agonist. *J. Biol. Chem.* **273**: 7783–7786.
- Huang, Q., Raya, A., DeJesus, P., Chao, S.H., Quon, K.C., Caldwell, J.S., Chanda, S.K., Izpisua-Belmonte, J.C., and Schultz, P.G. 2004. Identification of p53 regulators by genome-wide functional analysis. *Proc. Natl. Acad. Sci.* **101**: 3456–3461.
- Iourgenko, V., Zhang, W., Mickanin, C., Daly, I., Jiang, C., Hexham, J.M., Orth, A.P., Miraglia, L., Meltzer, J., Garza, D., et al. 2003. Identification of a family of cAMP response element-binding protein coactivators by genome-scale functional analysis in mammalian cells. *Proc. Natl. Acad. Sci.* **100**: 12147–12152.
- Kaufmann, S.H. and Hengartner, M.O. 2001. Programmed cell death: Alive and well in the new millennium. *Trends Cell Biol.* **11**: 526–534.
- Kiger, A., Baum, B., Jones, S., Jones, M., Coulson, A., Echeverri, C., and Perrimon, N. 2003. A functional genomic analysis of cell morphology using RNA interference. *J. Biol.* **2**: 27.
- Lander, E.S., Linton, L.M., Birren, B., Nussbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.
- Levy, G.N. 1997. Prostaglandin H synthases, nonsteroidal anti-inflammatory drugs, and colon cancer. *FASEB J.* **11**: 234–247.
- Lockhart, D.J. and Winzler, E.A. 2000. Genomics, gene expression and DNA arrays. *Nature* **405**: 827–836.
- Lundholt, B.K., Scudder, K.M., and Pagliaro, L. 2003. A simple technique for reducing edge effect in cell-based assays. *J. Biomol. Screen.* **8**: 566–570.
- MacBeath, G. 2002. Protein microarrays and proteomics. *Nat. Genet.* **32** **Suppl**: 526–532.
- Michiels, F., van Es, H., van Rompaey, L., Merchiers, P., Francken, B., Pittois, K., van der Schueren, J., Brys, R., Vandersmissen, J., Beirinckx, F., et al. 2002. Arrayed adenoviral expression libraries for functional screening. *Nat. Biotechnol.* **20**: 1154–1157.
- Paddison, P.J. and Hannon, G.J. 2002. RNA interference: The new somatic cell genetics? *Cancer Cell* **2**: 17–23.
- Paddison, P.J., Silva, J.M., Conklin, D.S., Schlabach, M., Li, M., Aruleba, S., Balija, V., O'Shaughnessy, A., Gnoj, L., Scobie, K., et al. 2004. A resource for large-scale RNA-interference-based screens in mammals. *Nature* **428**: 427–431.
- Pandey, A. and Mann, M. 2000. Proteomics to study genes and genomes. *Nature* **405**: 837–846.
- Price, J.H., Hunter, E.A., and Gough, D.A. 1996. Accuracy of least squares designed spatial FIR filters for segmentation of images of fluorescence stained cell nuclei. *Cytometry* **25**: 303–316.
- Price, J.H., Goodacre, A., Hahn, K., Hodgson, L., Hunter, E.A., Krajewski, S., Murphy, R.F., Rabinovich, A., Reed, J.C., and Heynen, S. 2002. Advances in molecular labeling, high throughput imaging and machine intelligence portend powerful functional cellular biochemistry tools. *J. Cell Biochem. Suppl* **39**: 194–210.
- Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. 1995. Quantitative

- monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**: 467–470.
- Steinmetz, L.M. and Davis, R.W. 2004. Maximizing the potential of functional genomics. *Nat. Rev. Genet.* **5**: 190–201.
- Su, A.I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G., et al. 2004. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci.* **101**: 6062–6067.
- Varambally, S., Dhanasekaran, S.M., Zhou, M., Barrette, T.R., Kumar-Sinha, C., Sanda, M.G., Ghosh, D., Pienta, K.J., Sewalt, R.G., Otte, A.P., et al. 2002. The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature* **419**: 624–629.
- Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al. 2001. The sequence of the human genome. *Science* **291**: 1304–1351.
- Vermeulen, K., Van Bockstaele, D.R., and Berneman, Z.N. 2003. The cell cycle: A review of regulation, deregulation and therapeutic targets in cancer. *Cell Prolif.* **36**: 131–149.
- Vogt, P.K. 2001. Jun, the oncoprotein. *Oncogene* **20**: 2365–2377.
- Vrana, J.A., Bieszczad, C.K., Cleaveland, E.S., Ma, Y., Park, J.P., Mohandas, T.K., and Craig, R.W. 2002. An MCL1-overexpressing Burkitt lymphoma subline exhibits enhanced survival on exposure to serum deprivation, topoisomerase inhibitors, or staurosporine but remains sensitive to 1- $\beta$ -D-arabinofuranosylcytosine. *Cancer Res.* **62**: 892–900.
- Wan, Y., Hur, W., Cho, C.Y., Liu, Y., Adrian, F.J., Lozach, O., Bach, S., Mayer, T., Fabbro, D., Meijer, L., et al. 2004. Synthesis and target identification of hymenialdisine analogs. *Chem. Biol.* **11**: 247–259.
- Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., et al. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**: 520–562.
- Williams, C.S., Mann, M., and DuBois, R.N. 1999. The role of cyclooxygenases in inflammation, cancer, and development. *Oncogene* **18**: 7908–7916.
- Zheng, L., Liu, J., Batalov, S., Zhou, D., Orth, A., Ding, S., and Schultz, P.G. 2004. An approach to genomewide screens of expressed small interfering RNAs in mammalian cells. *Proc. Natl. Acad. Sci.* **101**: 135–140.
- Ziauddin, J. and Sabatini, D.M. 2001. Microarrays of cells expressing defined cDNAs. *Nature* **411**: 107–110.

## Web site references

<http://function.gnf.org>; Genome-scale functional profiling resource site.

Received March 1, 2005; accepted in revised form June 1, 2005.