

Large-scale cDNA transfection screening for genes related to cancer development and progression

Dafang Wan^{*†}, Yi Gong^{†‡}, Wenxin Qin^{*†}, Pingping Zhang^{*†}, Jinjun Li^{*†}, Lin Wei^{*†}, Xiaomei Zhou^{*}, Hongnian Li^{*}, Xiaokun Qiu^{*}, Fei Zhong[§], Liping He^{*}, Jian Yu^{*}, Genfu Yao^{*}, Huiqiu Jiang^{*}, Lianfang Qian^{*}, Ye Yu^{*}, Huiqun Shu^{*}, Xianlian Chen^{*}, Huili Xu^{*}, Minglei Guo^{*}, Zhimei Pan^{*}, Yan Chen[‡], Chao Ge^{*}, Shengli Yang^{*†¶}, and Jianren Gu^{*¶}

^{*}National Laboratory for Oncogenes and Related Genes, Cancer Institute of Shanghai Jiao Tong University, Shanghai 200032, People's Republic of China; [†]Shanghai Research Center of Biotechnology, Chinese Academy of Sciences, Shanghai 200233, People's Republic of China; and [§]BioInfo Bridge, 16905 George Washington Drive, Rockville, MD 20853

Edited by Peter K. Vogt, The Scripps Research Institute, La Jolla, CA, and approved September 9, 2004 (received for review June 8, 2004)

A large-scale assay was performed by transfecting 29,910 individual cDNA clones derived from human placenta, fetus, and normal liver tissues into human hepatoma cells and 22,926 cDNA clones into mouse NIH 3T3 cells. Based on the results of colony formation in hepatoma cells and foci formation in NIH 3T3 cells, 3,806 cDNA species (8,237 clones) were found to possess the ability of either stimulating or inhibiting cell growth. Among them, 2,836 (6,958 clones) were known genes, 372 (384 clones) were previously unrecognized genes, and 598 (895 clones) were unigenes of uncharacterized structure and function. A comprehensive analysis of the genes and the potential mechanisms for their involvement in the regulation of cell growth is provided. The genes were classified into four categories: I, genes related to the basic cellular mechanism for growth and survival; II, genes related to the cellular microenvironment; III, genes related to host-cell systemic regulation; and IV, genes of miscellaneous function. The extensive growth-regulatory activity of genes with such highly diversified functions suggests that cancer may be related to multiple levels of cellular and systemic controls. The present assay provides a direct genomewide functional screening method. It offers a better understanding of the basic machinery of oncogenesis, including previously undescribed systemic regulatory mechanisms, and also provides a tool for gene discovery with potential clinical applications.

In the last decade, rapid progress in human genome research and advancements in technology have made important contributions to cancer genetics and to the identification of genes involved in carcinogenesis and cancer progression (1, 2). Based on sequencing and gene annotation at the genome scale, >30,000 genes have been described (3). Transcription profiling of cancers carried out by cDNA array (4, 5) and serial analysis of gene expression (SAGE) (6, 7) surveys also have provided extensive information on genes related to cancer. These commonly used technologies in cancer genomics have detected many changes in gene transcription and expression, but they are unable to identify those genes that are functionally relevant for cancer development, particularly genes involved in cancer cell growth. For example, the gene expression profile in cancer could identify genes differentially expressed, but these genes may include (i) genes directly participating in the regulation of cell proliferation and growth or (ii) genes merely existing for the specific function of particular parent tissues or organs, with no relation to cell growth.

Based on the above-mentioned concerns, we initiated a large-scale screening approach at the genomic level to search for genes directly related to cell proliferation and survival based on cDNA transfection into cancer cells and NIH 3T3 mouse fibroblasts. The large-scale screening has examined \approx 30,000 cDNA clones, selected from the 150,000 clones from human liver, placenta, and fetal tissues after removal of redundant species. Clones that showed substantial inhibitory or stimulatory effects on the colony-formation rate of cancer cells and transforming activity

in NIH 3T3 cells were selected. These cDNA clones were sequenced and analyzed for their identity.

The present study is based on large-scale cDNA transfection and provides a direct genome-wide functional screening method. It offers a broad and comprehensive database on genes that may affect cancer formation and progression, including aspects of systemic regulation that have not been recognized previously. This work also provides a tool for the search of additional target genes that may become valuable in clinical applications.

Materials and Methods

Cell Lines and Culture. SMMC7721 is a human hepatocellular carcinoma cell line, established in 1977 (a gift from the Shanghai Second University of Military Medicine). NIH 3T3 is a mouse embryonic fibroblast cell line obtained from American Type Culture Collection. Both cells were cultured in DMEM supplemented with 10% calf serum (GIBCO/BRL).

Construction of cDNA Libraries. Three separate cDNA libraries were constructed with mRNAs isolated from normal human liver, placenta, and fetal tissues. The human liver tissue was from a healthy individual who had died in an accident. The tissue was obtained according to consent regulations of the Chinese government. Placenta and fetal tissues were collected in accord with government regulations by the Hospital of Obstetrics and Gynecology (Shanghai Medical College, Fudan University, Shanghai, People's Republic of China). Fetal tissues were collected from an aborted fetus at the gestation age of 24 weeks in a case requiring termination of pregnancy because of maternal health conditions (primary hypertension). All sample collections were under consensus agreements, and the procedures were approved by the Ethical Review Committee of the World Health Organization Collaborating Center for Research in Human Reproduction. Further details regarding the construction of cDNA libraries can be found in *Supporting Materials and Methods*, which is published as supporting information on the PNAS web site.

Plating and Selection of cDNA Clones. It is important to select cDNA clones with relatively low abundance from the cDNA library before transfection. We developed a "negative selection" method designated as "plating and self-hybridization" to choose clones with a low hybridization signal after plating the cDNA

This paper was submitted directly (Track II) to the PNAS office.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. can be found in Tables 6–9, which are published as supporting information on the PNAS web site).

[†]D.W., Y.G., W.Q., P.Z., J.L., and L.W. contributed equally to this work.

[¶]To whom correspondence may be addressed at: National Laboratory for Oncogenes and Related Genes, Cancer Institute of Shanghai Jiao Tong University, 25/Ln 2200 Xie-Tu Road, Shanghai 200032, People's Republic of China. E-mail: nlorg@sh163.net or slyang@srcb.ac.cn.

© 2004 by The National Academy of Sciences of the USA

clones on an agarose disk and hybridization with [α - 32 P]dCTP-labeled total cDNAs from the relevant library. For a more detailed description, see *Supporting Materials and Methods*.

cDNA Transfection Assay. DNA was prepared from each cDNA clone after incubation in 2 ml of LB medium at 37°C overnight and purified by using the 96-well Qiagen (Valencia, CA) apparatus according to the manufacturer's protocol. It then was dissolved in 50 μ l of TE buffer (10 mM Tris/1 mM EDTA, pH 8.0) and quantified by UV absorbance measurement. Before transfection, 100 ng of DNA (for three wells) was precipitated in 70% ethanol. The pellet was air-dried and dissolved in 6 μ l of sterile, deionized water.

DNA transfections were performed in 96-well plates (GIBCO/BRL). Into each well, 2×10^3 SMMC7721 cells were seeded and grown at 37°C until reaching 60–70% confluence. For each triplet of transfection, the DNA (100 ng in 6 μ l) was mixed with 0.74 μ l of Lipofectamine reagent (GIBCO/BRL) and 3.3 μ l of serum-free DMEM and kept at room temperature for 15 min. The mixture then was diluted with 150 μ l of serum-free DMEM. Fifty microliters of aliquot was added into each cell-culture well in triplicate. DNA per well was ≈ 33 ng in 50 μ l of medium (0.66 μ g/ml). Three to four hours later, another 50 μ l of DMEM with 20% calf serum was added, and the cells were incubated at 37°C for 24 h. On day 2, the medium was replaced with fresh DMEM containing 10% calf serum and G418 (GIBCO/BRL), and the medium was replaced afterward every 2–3 days. The concentration of G418 was maintained between 300 and 1,000 μ g/ml for SMMC7721 and 200 and 500 μ g/ml for NIH 3T3 cells.

Measurement of Colony Formation of SMMC7721 Cells and Focus Formation of NIH 3T3 Cells. Because the pCMV-Script vector contains the *neo* gene encoding G418 resistance, DNA transfection of the vector itself into cancer cells resulted in a number of colonies formed in the presence of G418. cDNA clones that had growth-suppressing or -stimulatory effects showed a decrease or increase in the number of colonies formed, respectively, compared with the pCMV-Script empty vector control. In each 96-well plate of SMMC7721 cells, both pCMV-Script empty vector and pCMV-Script-p53 were added in triplicate as controls. Because pCMV-Script-p53 transfection usually reduced the colony formation rate of hepatoma cells to 0 to $\approx 10\%$ as compared with that of pCMV-Script empty vector transfection, we used pCMV-Script empty vector and pCMV-Script-p53 as controls in each 96-well plate.

In NIH 3T3 cells, transfection of the pCMV-Script empty vector alone resulted in focus formation attributed to spontaneous transformation. Transfections with the pCMV-Script vector control and with a vector containing the protooncogene JCL-1/bacillus Calmette–Guérin 1 served as references (8, 9) for zero and positive signals, respectively, in each 96-well plate.

In all of the transfection experiments, the colonies formed by SMMC7721 hepatoma cells and the foci formed by NIH 3T3 cells were counted under the microscope ($\times 100$). The mean values of each triplicate determination were calculated. The colony-formation rate was calculated as (no. of colonies by test cDNA – no. of colonies by vector)/no. of colonies by vector $\times 100\%$, and the rate of focus formation was calculated as (no. of foci by test cDNA – no. of foci by vector DNA)/no. of foci by vector DNA $\times 100\%$. As a control, the p53 transfection in SMMC7721 cells showed a consistent inhibitory effect of -82 to -100% ($-97 \pm 2.7\%$, $n = 140$) on colony formation, and JCL-1/bacillus Calmette–Guérin 1 gave a stimulatory effect of $+50$ to $+693\%$ ($252 \pm 129\%$, $n = 119$) on focus formation in NIH 3T3 cells. Based on these data, when the rate of a test cDNA was less than or equal to -50% or more than or equal to $+50\%$, we considered the cDNA clone as inhibitory or stimulatory for cell growth, respectively. Only these clones were selected for sequence analysis for further evaluation.

Table 1. Overall results of large-scale cDNA transfection screening

Clones	No. of cDNA clones
Overall plated	>150,000
Selected after self-annealing to remove abundant species	30,132
Transfected into human hepatoma SMMC7721 cells	29,910
Transfected into mouse NIH 3T3 cells	22,926
With stimulatory or inhibitory effects on hepatoma or NIH 3T3 cells*	8,237 (3,806)
Known genes [†]	6,958 (2,836)
Previously unrecognized full-length cDNAs [‡]	384 (372)
Uncharacterized unigenes [§]	895 (598)

*Total number of cDNA clones with stimulatory or inhibitory effects on hepatoma or NIH 3T3 cells. The numbers in parentheses indicate the unigene numbers.

[†]Number of cDNA clones of known genes with stimulatory or inhibitory effects on hepatoma and/or NIH 3T3 cells.

[‡]Number of cDNA clones of previously unrecognized genes with stimulatory or inhibitory effects on hepatoma and/or NIH 3T3 cells.

[§]Number of cDNA clones of unclassified unigenes with stimulatory or inhibitory effects on hepatoma and/or NIH 3T3 cells.

Sequencing of cDNA Clones. cDNA clones were sequenced by using the ABI BigDye Terminator chemistry and dGTP BigDye Terminator Kit v2 (Applied Biosystems) on an ABI PRISM 377 DNA sequencer and an ABI PRISM 3700 DNA analyzer (Applied Biosystems) with custom-designed primers. The sequencing was carried out by Shanghai GeneCore Biotechnologies (Shanghai, People's Republic of China).

Bioinformatics Analysis. Sequences of cDNAs were analyzed by using the programs DNASIS (Hitachi Software, Tokyo), VECTOR NTI (Invitrogen), and ORF FINDER (available at www.ncbi.nlm.nih.gov/gorf/gorf.html). Homology and identity analysis were performed with the WISCONSIN software package (GCG, Accelrys, San Diego), DATAEXTENDED (Accelrys), and BLAST (National Center for Biotechnology Information).

Results and Discussion

Overall Features of Large-Scale cDNA Transfection Screening. Three cDNA libraries were constructed in a eukaryotic expression vector pCMV-Script XR from normal human liver, placenta, and mixed fetal tissues. Overall, we plated 150,000 cDNA clones with sizes >2 kb and selected 30,000 cDNA clones with relatively low abundance as judged by self-hybridization. Transfection assays were performed with 29,910 individual cDNA clones into SMMC7721 human hepatoma cells and 22,926 clones into mouse NIH 3T3 cells, each in triplicate on 96-well plates. We found 8,237 clones that have either stimulatory or inhibitory effects on the growth of SMMC7721 cells, NIH 3T3 cells, or both. Among these, we identified 3,806 unique gene identities. In summary, 2,836 known genes (6,958 clones), 372 previously unrecognized genes (384 clones), and 598 uncharacterized unigenes (895 clones) were selected by our assay (Table 1). The percentage of unigenes that had a stimulatory or inhibitory effect on both or either SMMC7721 or NIH 3T3 cells is shown in Fig. 1, which is published as supporting information on the PNAS web site.

From the above data, 2,836 known genes had definitive effects on cell growth. Some were known for their roles in cancer development, but others had never been reported to be related to cancer or cell growth (Tables 2–6). The various functional groups of genes and their possible mechanisms of action suggest that cancer growth is a highly complex process and may be governed by multiple hierarchical levels of cellular and systemic regulation.

Based on the possible regulatory mechanism of cell growth and survival, we classified the various gene groups into the following four categories: I, genes related to the basic cellular machineries for

Table 2. Examples of known Category I genes related to cell growth revealed by large-scale cDNA transfection

Description	Unigenes	Clones	Examples
Cell-cycle regulators	59	103	Cyclins (cyclin I, cyclin D3, cyclin G1, cyclin L2); cell cycle kinases [CDK10, CDK2, CLK2, CDK activating kinase (CDK7), CDK9, cell cycle-related kinase (CCRK), Wee 1, CDC2L1]; CDK inhibitors [p15 INK4B, p19 (SKP1A), p21 (Cip1/WAF-1), p57 (Kip2)]; mitosis/spindle check point [Anaphase-promoting complex subunit 5(ANAPC5); CDC16, Kinesin light chain 2 (KLC2)]; others [cyclin G associated kinase (GAK), cell-cycle progression 2 protein (CPR2), Quiescin (Q6)]
Apoptosis	49	92	Proapoptotic genes [lipopolysaccharide-induced TNF factor (LITAF); TNFR superfamily member 1A, 1B, 3, 6b, 19, 21, 25; TNFR-associated protein (TRAF2); death-associated protein (DAP), death-associated protein 3 (DAP3); Bcl-xS; caspase 9]. Antiapoptotic genes [Bcl-xL; Survivin; BCL2-associated athanogene (BAG1); Optineurin (OPTN)]
Growth factors and growth inhibitors	30	389	Ferritin H; ferritin L; transferrin; bone morphogenetic protein 1–4 (BMP1–4), bone morphogenetic protein 1, 7 (BMP 1, 7); prostate differentiation factor (PLAB); epithelin (granulin); fibroblast growth factor 11 (FGF11); hepatoma-derived growth factor (HDGF); hepatocyte growth factor (HGF) activator precursor; hepatocyte growth factor (HGF) activator inhibitor type II (serine protease inhibitor, Kunitz type 2); HGF-like protein (macrophage stimulating 1); insulin-like growth factor (IGF II); TGFB1
Receptors	79	148	Growth factor receptors [insulin receptor, insulin-like growth factor 2 receptor (IGF2R); fibroblast growth factor 2 receptor (FGFR2); TRK E (DDR1); Eph A5, B3, B4, B6; nerve growth factor receptor (NGFR); activin A receptor type II-like 1 (ACVRL1); angiotensin II receptor]. Other receptors or receptor-like molecules [integrin α 3, α 5, α 7, α V; Integrin β 1, β 4; G protein-coupled R (GPR56); G protein-coupled R (GPCR5C); asialoglycoprotein R 1; nuclear receptor subfamily 1H2 (NR1H2); retinoid X receptor (RXRB)]
Signal transduction	309	671	Protein kinases [integrin-linked kinase (ILK); AMP-activated protein kinase (PRKAB1); protein kinase C like 1 (PRKCL1); Ca ²⁺ /calmodulin-dependent protein kinase kinase β (CAMKK2); Casein kinase 1 δ , 1 ϵ ; MEKK1 (MAP3K1); MEK2 (MAP2K2); JNKK2 (MAP2K7); I κ B kinase γ subunit (IKBK γ); ERK1 (MAPK3); CDC42-binding protein kinase β (CDC42BPB)]. Protein phosphatases [protein phosphatase 1 (PPP1CA); protein phosphatase 2A (PPP2R1A); protein phosphatase 1G (PPM1G); protein phosphatase 5 (PPP5C); dual-specificity protein phosphatases 6, 9 (DUSP6, DUSP9); protein tyrosine phosphatase receptor type A (PTPRA)]. Other enzymes and related molecules [phospholipase C δ 1; adenylate cyclase 3 (ADCY30); diacylglycerol kinase (DGKA); PI3K (PIK3CD, PIK3R2); inositol 1,4,5-trisphosphate 3-kinase C; Inositol polyphosphate-5-phosphatase]. Other signal molecules [Arrestin β 2; EGF R substrate (EPS15R); Ezrin (Villin 2); hepatocyte growth factor regulated tyrosine kinase substrate (HGS); insulin receptor substrate 1 (IRS1); PKC substrate 80K-H; FK506 binding protein 4 (FKBP4); GTPase activation protein 1 (RAP1GA1); rho guanine nucleotide exchange factor (GEF) 1; GTP binding protein 1, 3 (GTPBP1, GTPBP3); p53 binding protein 1 (TP53BP1); Calmodulin 1, 3; frizzled related protein (FRZB); NF-B transcription inhibitor RELB (I-Rel); I- κ B-interacting Ras-like protein 2 (KBRAS2); TGF induced gene product (TIEG); latent TGF β binding protein 1, 3, 4 (LTBP1, LTBP3, LTBP4); Dickkopf 3 (DKK3); epidermal growth factor receptor-bound protein 2 (GRB2); IGF binding protein 1, 2, 3, 4, 5 (IGFBP1, IGFBP2, IGFBP3, IGFBP4, IGFBP5); Ras-related associated with diabetes (RRAD)]

survival and growth; II, genes related to the microenvironment; III, genes related to host–cell systemic regulation; and IV, genes of miscellaneous function.

Category I: Genes Related to Basic Cellular Machineries for Survival and Growth. Category I (Table 2) includes genes that are closely related to basic cellular machineries. This group is broadly defined, containing genes that code for regulators of the cell cycle and of apoptosis, gene expression at multiple levels, protein transportation and degradation, DNA replication and repair (10), acetylation, and methylation (11), as well as protooncogenes and tumor suppressors. Some of the genes reported here have not been linked previously to cancer development and progression. But the fact that they all had specific effects on cell growth in single-gene transfections suggests that cellular mechanisms for cancer development and progression are much more complex than commonly accepted.

Category II: Genes Related to Microenvironment. In addition to genes coding for proteins of basic cellular machineries, we also identified several groups of genes coding for proteins involved in cell–cell and cell–matrix interactions. Category II (Table 3) contains cell surface and membrane proteins, glycosylation-related proteins, matrix-cell adhesion and cytoskeleton-related proteins, proteases, and proteins related to angiogenesis or vasculogenic mimicry (12–14). Cancer tissues are not simple ensembles of cancer cells. Instead, they are composed of cancer cells packaged in a defined, complex matrix structure along with other cells including fibroblasts, vascular cells, and various types of immune cells. These cells with their interactions and communications provide a microenvironment for cancer cells. We therefore defined the genes functioning in cell–cell interactions as microenvironment-related genes. These genes affect the growth, progression, and behavior of cancer cells (15).

Category III: Genes Related to Host–Cell Interaction and Systemic Regulation. In category III (Table 4), we listed the following groups of genes: genes related to the host responses to environment, nutrition, and redox activities (16); genes involved in the immune response, such as cytokine/chemokine receptors; other genes related to systemic responses, including ion channels (17, 18), transporters, neurotransmitter, and small-molecule receptors; and the circadian rhythm-related genes, such as *PER1* and *CLOCK* (19, 20). The finding of these genes with definitive effects in cancer cell proliferation is extraordinary, indicating that although cells react to conditions of the host system in many delicate ways, such conditions also can affect growth behavior profoundly. The corresponding genes represent the molecular basis for the systemic host-dependent regulation of cells from different tissues or organs. Disruption of such host–cell homeostasis may induce abnormalities in individual cell growth.

Category IV: Genes of Uncharacterized Functions Related to Cell Growth. In the miscellaneous group, category IV (Table 5), there were a number of genes with functions that are not yet well characterized. They include some cancer-related proteins and RNA genes, i.e., proteins related to leukemia, Wilms tumor (21), and colon and small-cell lung cancer (22); the H19 RNA gene (23) (regulated by p53); development-related genes; transcripts from repeated DNA sequences (LINE 1 and SINE-R11) and transposon-related genes; virus-related proteins; RNA helicases and RNA-binding proteins; and other proteins. Because the functions of these genes are not well characterized, their specific roles in cell growth must await additional investigations.

Previously Unrecognized Genes and Unclassified Unigenes. In this report, we also identified 372 previously unrecognized genes with full-length sequences acting as suppressors or stimulators of cell

Table 2. (continued)

Description	Unigenes	Clones	Examples
Transactivators and transcription factors	129	215	Transactivators [AP2 (TFAP2A); ATF3; E74-like factor 1, 3 (Ets domain transcription factor ELF1, ELF3); Forkhead box M1 (FOXM1); MAD-3 (NFKBIA); Myb-binding protein 1A (MYBBP1A); nuclear factor of κ light polypeptide gene enhancer in B-cells 2 (p49/p100) (NF κ B2); interleukin enhancer binding factor 2, 45kDa (ILF2); nuclear factor NF-IL6 (CEBPB); E2F-related transcription factor (TFDP1); HOXA11; HOXC10; STAT3; STAT5A]. Transcription factors [TAF6; BRF2; transcription elongation factor TFIIB (BRF1); general transcription factor II B (GTF2B)]. Other transcription-related genes [RNA polymerase II (POLR2J); Tripartite motif-containing 28 (TRIM28); interleukin enhancer binding factor 3, 90 kDa (ILF3)]
Splicing related	25	38	Splicing factor 3a (SF3A2, SF3A3), PRP8 pre-mRNA processing factor 8, splicing factor, arginine/serine-rich 2 (SFRS2), SF1, SFRS2IP, NMP200, SF3B2, SF3B3
Translation related	95	347	EIF1AY, EIF2B2, EIF2C2, EIF3S2, EIF4A1, EIF4B, EIF4G1, EIF5A, EIF2AK4
DNA replication and repair	61	92	DNA Replication [MCM3; MCM5; MCM7; Topoisomerase III β (TOP3B)]. DNA Repair [Tankyrase (TNKS); damage-specific DNA-binding protein 1 (DDB1); DNA damage inducible RNA-binding protein (CIRBP); ERCC3; xeroderma pigmentosum, complementation group C (XPC); POLM; XRCC1]
Acetylation, deacetylation, and methylation	20	42	Methyl CpG binding protein 1, 3 (MBD1, MBD3); methyltransferase like 3 (METTL3); histone deacetylase 1, 2, 3, 5, 6, 10; MYST histone acetyltransferase 4 (MYST4)
Protein processing and transportation	107	276	Protein processing [Calreticulin; CCT2; CCT3; CCT5; CCT7; DNAJA1; DNAJB1; DNAJB2; PDIR; heat shock protein (Hsp) 27, 47, 60, 70, 75, 90; thioredoxin-like (TXNL)]. Protein transportation and vesicle formation [adaptin (GGA1, GGA2); ADP-ribosylation factor 1 (ARF1); ARF like 1, 3 (ARL1, ARL3); COPE; caveolin 1; flotillin 1, 2; vesicle-associated membrane protein 8 (VAMP8); vacuolar protein sorting 28, 33B; Golgi autoantigen (GOLGA2); SEC61A1, SCAMP 2, 3, 4; cargo selection protein (TIP47)]
Proteasome and protein degradation	54	120	SUMO-1 activating enzyme subunit 1 (SAE1); ubiquitin C (UBC); proteasome 26S subunit ATPase 1, 3, 4, 5 (PSMC1, PSMC3, PSMC4, PSMC5); proteasome 26S subunit non-ATPase 2 (PSMD2); proteasome subunit β type 2, 10 (PSMB2, PSMB10); proteasome activator subunit 2 (PSME2); proteasome inhibitor subunit 1 (PSMF1); ubiquitin specific protease 3 (USP3); ubiquitin specific protease 9 (USP9X); ubiquitin-conjugating enzyme E2E 3, E2L 6 (UBE2E3, UBE2L6); ubiquitin-activating enzyme E1 (UBE1)
Protooncogenes and putative protooncogenes	56	91	c-fes (FES); c-fms (CSF1R); c-myc binding protein (MYCBP); c-kit; spleen tyrosine kinase (SYK); v-myb myeloblastosis viral oncogene homolog (avian)-like 2 (MYBL2); ARAF1; JUN; JUNB; JUND; ARHA, ARHB; MAFF; BCR; MLN51; ELK1; ubiquitin specific protease 4 (USP4); smoothened homolog (SMO)
Tumor suppressors	37	89	Tumor suppressors [p53; BRCA2; BRCA1 associated protein (BRAP); ATM; MEN1; NF2; NOTCH1; tumor suppressor TSSC3 (PHLDA2); tumor suppressor NOC2 (RPH3AL)]. Metastasis suppressors (KISS1; KAI1). Tumor suppressor candidates [CUTL1; OVCA2; glioma tumor suppressor candidate region gene 2 (GLTSCR2); human homolog of <i>Drosophila</i> neuralized gene (NEURL); member RAS oncogene family-like (RAB2L), tumor suppressor candidate 2 (TUSC2)]

All categories and subgroups were classified according to gene function described in literature and information from Gene Ontology. For Tables 2–5, the total number of unigenes is 2,836, and the total number of clones is 6,958.

growth in cDNA transfection (see Table 7). The sequences of these previously unrecognized genes have been deposited in the GenBank database. The molecular mechanisms of several previously unrecognized genes, such as BNIPL (24–26), LASS2 (27), and HCCS1 (28) (a gene previously isolated from chromosome 17p13.3 by positional cloning), that inhibited or stimulated cell growth have been elucidated by our laboratory. These results provide additional support for the feasibility and reliability of the transfection screening. The functional characterization of the remaining previously unrecognized genes is under way.

In addition to the 372 genes with full-length sequences, we also obtained 598 previously unrecognized unigenes by the cDNA transfection assay (Table 8). Because the full-length genes of these sequences have not been characterized and nothing is known about their function, we designated them as unclassified. Further analysis of these sequences is needed.

The Effects of cDNA Clones with Full-Length or Partial Gene Sequences. An important concern in using cDNA libraries is the intactness of the cDNA insert. In constructing the cDNA libraries, we used agarose gel electrophoresis to select cDNAs with a size >2.0 kb. Also, we used self-hybridization to reduce the prevalence of highly abundant clones. Retrospective sequencing indicated that \approx 42% of all clones (6,958) from known genes were full-length cDNAs. Among them, 1,465 of 2,836 (51.7%) known unigenes were full-length cDNAs and 1,371 of 2,836 (48.3%) were partial sequences, including 3' UTR sequences (175 of 2,836, 6.17%). To examine the problem of partial vs. complete cDNA sequences, we selected 25 known genes of which both complete and partial sequences were found to have specific effects on cell growth. A

detailed analysis of the similarities and differences between complete and partial sequences is given in Table 9. Clones with incomplete 5' ends either behaved the same as full-length clones or gave effects opposite to those seen with full-length parent cDNA, i.e., reversed stimulatory to inhibitory action or *vice versa*. In either case, such clones could be selected by the present assay. The transfection results with partial cDNAs probably depend on the domains that remain intact in the expressed protein. If the essential domains remain intact, the effect would be similar to that of the full-length molecule. If only some domains are retained, the truncated protein could compete with parent molecules or other proteins, resulting in an opposite effect. Similar results have been reported by Chu *et al.* (29) in their study of a cDNA library in retroviral vectors screening for genes that activate T cells. They also found dominant negative effects of many partial cDNA sequences (29). Partial sequences that give rise to gene products lacking essential domains may fail to score in the transfection assay. Thus, a negative reading from a truncated sequence cannot exclude involvement of the full-length gene in cell growth. For example, we have detected p53, BRCA2, BRCA1-associated protein 1 (BAP1), BRCA1-associated protein (BRAP), MEN-1, and NF2 as tumor suppressors, but we failed to detect BRCA1 or RB and its related genes. Recently, a paper by Ota *et al.* (30) described the complete sequencing and characterization of 21,243 full-length human cDNAs. Further studies with these full-length cDNAs will enable a complete functional analysis of the genes involved.

General Consideration of Large-Scale cDNA Transfection as a Functional Genomic Screening Assay. The present cDNA screening system is a cell-proliferation and growth-based assay. It therefore

Table 3. Examples of known Category II genes related to cell growth revealed by large-scale cDNA transfection

Description	Unigenes	Clones	Examples
Cell surface and membrane protein	113	282	Cell junction-related proteins [Plakophilin 1, 3; Symplekin; Connexin 43 (GJA1); Connexin 40 (GJA5); Connexin 26 (GJB2); tight junction protein (TJP1, TJP3, TJP4); junction plakoglobin (JUP); desmoplakin; dermatopontin]. Membranous glycoproteins [chondroitin sulfate proteoglycan 2, 4 (CSPG2, CSPG4); galectin 3 (LGALS3); Annexin A1, A2, A11; α -2-HS-glycoprotein (AH5G); heparan sulfate proteoglycan 2 (HSPG2)]
Glycosylation	45	91	Aminidases [α -N-acetylgalactosaminidase (NAGA); hexosaminidase B (HEXB)]. Glycosidases [sialidase 1, 4; galactosidase α]. Glycosyltransferases [UDP-GlcNAc: β Gal β -1,3-N-acetylglucosaminyl transferase 3 (B3GNT3); UDP-Gal: β GlcNAc β -1,4-galactosyltransferase polypeptide 2 (B4GALT2); N-acetyl glucosaminyltransferase III (mannosyl β -1,4)-glycoprotein β -1,4-N-acetylglucosaminyltransferase); Sialyltransferase 4B (SIAT4B); Mannosyl (α -1,3)-glycoprotein β -1,2-N-acetylglucosaminyltransferase (MGAT1); Mannosyl (α -1,3)-glycoprotein β -1,4-N-acetylglucosaminyltransferase isoenzyme B (MGAT4B); Dolichyl-phosphate N-acetylglucosaminephosphotransferase 1 (DPAGT1)]
Matrix, cell adhesion, and cytoskeleton	168	722	Cell adhesion molecules [intracellular adhesion molecule 3 (ICAM3); CDH23, CDH5; laminin receptor 1 (LAMR1); platelet/endothelial CAM-1 (PECAM1); protocadherin 1, 16]. Matrix proteins [extracellular matrix protein 1 (ECM1); Desmin; Fibrillin 1, 2; Laminin α 5, β 1 and 2, γ 1 and 3; Osteonectin (SPARC)]. Collagen-related proteins [Collagen COL2A1, COL6A1, COL6A2]. Cytoskeleton and related proteins [Actin α 1 (ACTN1); Actin β (ACTB); Actin γ 1 (ACTG1); actin-binding protein (CORO1A); Gelsolin (GSN); Adducin 2 (ADD2); Dystrobrevin- β ; Plectin 1 (PLEC1); Profilin 2; Zyxin]
Proteases	62	207	Metalloproteinases [MMP2; MMP9; MMP11 (stromelysin 3); MMP15; MMP24; Matargidin (ADAM15)]. ATP-dependent metalloproteinases [YME1L1; Procollagen (type III) N-endopeptidase (PCOLN3)]. Serine proteases [plasminogen activator s (PLAT, PLAU); plasminogen; TIMP2; TIMP3; serine (or cysteine) proteinase inhibitor member 1, 2 (SERPINE1, SERPINE2); pigment epithelium differentiation factors (SERPINF1, SERPINF2)]. Other proteases [calcium-activated neutral protease (CAPN1); Ca ²⁺ -dependent protease (CAPNS1); Cathepsin B, C, D, F, L]. Unclassified related proteins (ADAM12, ADAM15)
Angiogenesis and vasculature	16	38	VEGFB; VEGF R1 (FLT1); eNOS (NOS3); tumor endothelial marker TEM1 (CD164 sialomucin-like 1); placenta growth factor (PGF)

All categories and subgroups were classified according to gene function described in literature and information from Gene Ontology. For Tables 2–5, the total number of unigenes is 2,836, and the total number of clones is 6,958.

provides direct information about individual cDNA clones regardless of whether the sequence is known or previously unrecognized. The known sequences inform us about the types of genes that are involved in cell proliferation and growth, cancer formation and progression, and possibly invasion and metastasis. Our findings establish the critical role of systemic regulation,

mediated by such proteins as neurotransmitter receptors, ion channels, and small-molecule receptors in growth regulation. Large-scale transfection is a valuable tool for functional genomics. It can be extended to a search for diagnostic markers, therapeutic genes and polypeptides, and target genes for drug discovery.

Table 4. Examples of known Category III genes related to cell growth revealed by large-scale cDNA transfection

Description	Unigenes	Clones	Examples
Genes related to environment, nutrition, and redox	418	884	General (SOD1). Nutrition response and metabolism-related [hexokinase 1; high-glucose-regulated protein 8 (HGRG8); fatty acid desaturase 1, 2, 3 (FADS1, FADS2, FADS3); phosphoglycerate kinase 1 (PGK1); hydroxysteroid (11- β) dehydrogenase 2 (HSD11B2); cytochrome P450 family (2, 3, 4, 11, 17, 19, 20, 21); S-adenosylhomocysteine hydrolase (AHCY); methylenetetrahydrofolate dehydrogenase (MTHFD1); AK2, AK3; adenosine deaminase (ADAR); cytoskeleton-related vitamin A responsive protein (JWA)]. Hormone regulation [corticotropin-releasing hormone (CRH); glucocorticoid modulatory element binding protein 1 (GMEB1); Adrenomedullin (ADM); prolactin regulatory element-binding protein (PREB); cytosolic thyroid hormone-binding protein (pyruvate kinase, muscle); gastrin-binding protein (HADHA, HADHB)]. ATP synthesis and ATPase (ATP5A1; ATP5B; ATP5F1; ATP5G2). Oxidation and electron transfer [flavoprotein (ETFB); flavoprotein dehydrogenase (ETFDH); NADH dehydrogenase (NDUFB1); P450 cytochrome oxidoreductase]. Glutathione-related [glutathione S-transferase (GST) A2, π ; glutathione synthetase; glutathione peroxidase 1, 3, 4; γ -glutamyltransferase-like activity 1 (GGTLA1); glyceronephosphate O-acyltransferase (GNPAT)]
Immune response related	101	265	Cytokine/chemokines [interferon γ ; interferon β 2 (IL6); chemokine (C-X3-C motif) ligand 1; chemokine (C-X-C motif) ligand 2; thymosin β 10]. Cytokine/chemokine receptors (CSF2RB; CSF3R). Interleukin receptors (IL1RL1; IL2RB; IL4R; IL10RA; IL17R). Cytokine-induced or regulatory proteins [interferon γ inducible protein 30 (IFI30); interferon-induced tetratricopeptide IFI60 (IFIT4); interferon-induced cellular resistance mediator MxA (MX 1)]. Antigen-related proteins [B cell receptor-associated protein 31 (BCAP31); antigen NY-CO-43 (PRKWINK2); CD4; CD44; hepatocellular carcinoma-associated antigen 59,66; squamous cell carcinoma antigen (SART1, 3); MAGED 1, 2, 4; melanoma 1 antigen (CD63); melanoma-associated gene (D2S448)]. Immune response modulatory proteins [Defensin α 1; decay accelerating factor (DAF); NK cell activating factor (PRG2)]
Ion channels, transporters, and exchangers	111	222	Ion channels [chloride channel 7 (CLCN7); sodium channel (SCN5A); potassium channel (KCNK3); calcium channel β 3 subunit (CACNB3); vacuolar H ⁺ ATPase subunit V0 (ATP6V0D1)]. Transporters and related proteins [anion exchanger (SLC4A1); potassium/chloride transporters (SLC12A4); sodium/hydrogen exchanger regulator factor 1, 2 (SLC9A3R1, SLC9A3R2); organic cation transporter (SLC22A1L, SLC22A1); amino acid transporter 4, 7 (SLC7A4, SLC7A7); facilitated glucose transporter (SLC2A8)]
Other proteins related to systemic regulation	20	30	PER1; Timeless Clock; NMDA R 2C (GRIN2C); GABARAP; GABARAPL; nicotinic acetylcholine R (CHRN4); folate R 2 (FOLR2); opioid growth factor receptor (OGFR); ciliary neurotrophic factor receptor (CNTFR); GDNF family receptor α 1, 2 (GFRA1, 2); sphingolipid G-protein-coupled receptor 5 (EDG5)

All categories and subgroups were classified according to gene function described in literature and information from Gene Ontology. For Tables 2–5, the total number of unigenes is 2,836, and the total number of clones is 6,958.

Table 5. Examples of known Category IV genes related to cell growth revealed by large-scale cDNA transfection

Description	Unigenes	Clones	Examples
Cancer-related proteins	37	157	Myeloid leukemia-associated (SET); Myeloid cell leukemia sequence 1 (BCL2-related); Wilms tumor 1-associated protein; CD164 antigen sialomucin (CD164); SURF1, SURF4, SURF6; Semaphorin 6A, 3B; H19
Development related	99	285	Adipose differentiation-related protein (Adipophilin); Ob gene (Leptin); Drebrin 1 (DBN1); No arches (nar) (CPSF4); PBK 1 (Trophoblast related gene); Mesoderm-specific transcript homolog (MEST); delta-like 1 homolog (Dlk1)
Transcripts from repeated sequence and transposon-related genes	5	7	LINE; SINE-R11; Transposon-like element; Ac-like transposable element (ALTE)
Virus-related proteins	12	31	EBV-induced gene 3 (EBI3); Ubiquitin-specific protease 7 (Herpes virus-associated) (USP7); Vaccinia-related kinase 3 (VRK3); HBxAg transactivated protein 2 (XTP2)
RNA helicases and RNA-binding proteins	45	86	RNA binding motif protein 3, 5, 6; RNA helicase (DDX56); Staufen (STAU); hnRNP A0, A1, A/B, C, D, F, K, M, R, U
Other proteins	474	938	Amyloid β A4 precursor (APBA3); Calcyphosine; Calcizzarin (S100A11); capping protein α 1 (CAPZA1); Prosaposin (PSAP); putative translation initiation factor (SUI1); mitochondrial carrier homolog 1 (MTCH1)

All categories and subgroups were classified according to gene function described in literature and information from Gene Ontology. For Tables 2–5, the total number of unigenes is 2,836, and the total number of clones is 6,958.

In our assays, the number of cDNA clones for transfection was small, considering that the whole genome comprises >30,000 genes and gives rise to a far greater number of protein isoforms. Representation of genes could be biased by the sources of the cDNA libraries. We used libraries from fetus, placenta, and liver because they offer the most complex mRNA populations. However, if only the mutated version of a gene affects cell growth (e.g., *ras*), it would not score in our assay (31, 32).

The large number and wide variety of genes that affect cell growth is striking. We performed a genome-scale screen for genes related to cell growth and cancer development and progression by using a functional assay, which to our knowledge has not been performed previously. In a related study, a cDNA library expressed by a retroviral vector was transfected into T cells to search for genes affecting activation (29). As a method of functional genomics (33), large-scale transfection linked to functional screening is applicable to a wide variety of cells in a search for genes related to specific functions (34, 35).

Conclusions

The present assay, based on large-scale cDNA transfection, has provided a direct genome-wide functional screening method to

offer a comprehensive database on genes that affect cell growth and are related to cancer formation or progression. The involvement of genes with highly diversified functions suggests that cell growth may be a much more complex process than previously recognized and may be related to multiple levels of hierarchical regulation, including the cellular basic machineries, microenvironment, and host–cell homeostasis. Of particular interest is the systemic regulation; it includes ion channels, small-molecule transporters and receptors, and some neurotransmitter receptors that have not previously been connected to cell growth. The present assay provides an important tool in the search for genes of potential clinical relevance, such as diagnostic markers and drug targets.

We thank Dr. Jian Ni for his contribution to bioinformatics analysis and Dr. Yuhong Xu for her contribution to suggestions and revision in manuscript preparation. This work was supported by National Key Basic Research Project of China Grant 973 and grants from the Shanghai Municipal Commission for Science and Technology and Shanghai Neworgen, Inc.

- 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) *Nature* **409**, 860–921.
- 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) *Science* **291**, 1304–1351.
- Claverie, J. M. (2001) *Science* **291**, 1255–1257.
- Zembutsu, H., Ohnishi, Y., Tsunoda, T., Furukawa, Y., Katagiri, T., Ueyama, Y., Tamaoki, N., Nomura, T., Kitahara, O., Yanagawa, R., et al. (2002) *Cancer Res.* **62**, 518–527.
- Xu, X. R., Huang, J., Xu, Z. G., Qian, B. Z., Zhu, Z. D., Yan, Q., Cai, T., Zhang, X., Xiao, H. S., Qu, J., et al. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 15089–15094.
- Porter, D. A., Krop, I. E., Nasser, S., Sgroi, D., Kaclin, C. M., Marks, J. R., Riggins, G. & Polyak, K. (2001) *Cancer Res.* **61**, 5697–5702.
- Stollberg, J., Urschitz, J., Urban, Z. & Boyd, C. D. (2000) *Genome Res.* **10**, 1241–1248.
- Lucas, S., Brasseur, F. & Boon, T. (1999) *Cancer Res.* **59**, 4100–4103.
- Langnaese, K., Kloos, D. U., Wehnert, M., Seidel, B. & Wieacker, P. (2001) *Cytogenet. Cell Genet.* **94**, 233–240.
- Feng, Z., Hu, W., Komissarova, E., Pao, A., Hung, M. C., Adair, G. M. & Tang, M. S. (2002) *J. Biol. Chem.* **277**, 12777–12783.
- Esteller, M., Fraga, M. F., Paz, M. F., Campo, E., Colomer, D., Novo, F. J., Calasanz, M. J., Galm, O., Guo, M., Benitez, J., et al. (2002) *Science* **297**, 1807–1808.
- Thiery, J. P. (2002) *Nat. Rev. Cancer* **2**, 442–454.
- Friedl, P. & Wolf, K. (2003) *Nat. Rev. Cancer* **3**, 362–374.
- Hendrix, M. J. C., Seftor, E. A., Hess, A. R. & Seftor, R. E. B. (2003) *Nat. Rev. Cancer* **3**, 411–421.
- Liotta, L. A. & Kohn, E. C. (2001) *Nature* **11**, 375–379.
- Meyer, T. E., Liang, H. Q., Buckley, A. R., Buckley, D. J., Gout, P. W., Green, E. H. & Bode, A. M. (1998) *Int. J. Cancer* **77**, 55–63.
- Nishi, T. & Forgac, M. (2002) *Nat. Rev. Mol. Cell Biol.* **3**, 94–103.
- Wissenbach, U., Niemeier, B. A., Fixemer, T., Schneidewind, A., Trost, C., Cavalié, A., Reus, K., Meese, E., Bonkhoff, H. & Flockerzi, V. (2001) *J. Biol. Chem.* **276**, 19461–19468.
- Etchegaray, J. P., Lee, C., Wade, P. A. & Reppert, S. M. (2003) *Nature* **421**, 177–182.
- Cheng, M. Y., Bullock, C. M., Li, C., Lee, A. G., Bermak, J. C., Belluzzi, J., Weaver, D. R., Leslie, F. M. & Zhou, Q. Y. (2002) *Nature* **417**, 405–410.
- Overall, M. L., Parker, N. J., Scarcella, D. L., Smith, P. J. & Dziadek, M. (1998) *Mamm. Genome* **9**, 657–659.
- Kuroki, T., Trapasso, F., Yendamuri, S., Matsuyama, A., Alder, H., Williams, N. N., Kaiser, L. R. & Croce, C. M. (2003) *Cancer Res.* **63**, 3352–3355.
- Cui, H., Onyango, P., Brandenburg, S., Wu, Y., Hsieh, C. L. & Feinberg, A. P. (2002) *Cancer Res.* **62**, 6442–6446.
- Shen, L., Hu, J., Lu, H., Wu, M., Qin, W., Wan, D., Li, Y. Y. & Gu, J. (2003) *FEBS Lett.* **540**, 86–90.
- Qin, W., Hu, J., Guo, M., Xu, J., Li, J., Yao, G., Zhou, X., Jiang, H., Zhang, P., Shen, L., et al. (2003) *Biochem. Biophys. Res. Commun.* **308**, 379–385.
- Zhou, Y. T., Soh, U. J., Shang, X., Guy, G. R. & Low, B. C. (2000) *J. Biol. Chem.* **277**, 7483–7492.
- Pan, H., Qin, W. X., Huo, K. K., Wan, D. F., Yu, Y., Xu, Z. G., Hu, Q. D., Gu, K. T., Zhou, X. M., Jiang, H. Q., et al. (2001) *Genomics* **77**, 58–64.
- Zhao, X., Li, J., He, Y., Lan, F., Fu, L., Guo, J., Zhao, R., Ye, Y., He, M., Chong, W., et al. (2001) *Cancer Res.* **61**, 7383–7387.
- Chu, P., Pardo, J., Zhao, H., Li, C. C., Pali, E., Shen, M. M., Qu, K., Yu, S. X., Huang, B. C., Yu, P., et al. (2003) *J. Biol. Chem.* **278**, 1–16.
- Ota, T., Suzuki, Y., Nishikawa, T., Otsuki, T., Sugiyama, T., Irie, R., Wakamatsu, A., Hayashi, K., Sato, H., Nagai, K., et al. (2004) *Nat. Genet.* **36**, 40–45.
- Satomi, Y., Bu, P., Okuda, M., Tokuda, H. & Nishino, H. (2003) *Cancer Lett.* **196**, 17–22.
- Fujita, M., Norris, D. A., Yagi, H., Walsh, P., Morelli, J. G., Weston, W. L., Terada, N., Bennion, S. D., Robinson, W., Lemon, M., et al. (1999) *Melanoma Res.* **9**, 279–291.
- Brummelkamp, T. R. & Bernards, R. (2003) *Nat. Rev. Cancer* **3**, 781–789.
- 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) *Proc. Natl. Acad. Sci. USA* **100**, 12153–12158.
- 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) *Proc. Natl. Acad. Sci. USA* **100**, 12147–12152.