

Genomewide gain-of-function genetic screen identifies functionally active genes in mouse embryonic stem cells

Moshe Pritsker*, Nicole R. Ford, Harry T. Jenq, and Ihor R. Lemischka†

Department of Molecular Biology, Princeton University, Princeton, NJ 08544

Edited by Rudolf Jaenisch, Whitehead Institute for Biomedical Research, Cambridge, MA, and approved March 23, 2006 (received for review November 11, 2005)

Embryonic stem (ES) cells hold great promise for the future of medicine. To elucidate the molecular mechanisms that control ES cell self-renewal and differentiation, a comprehensive knowledge of the molecules involved in these processes is required. Here we describe an effective approach for genomewide identification of functionally active genes in ES cells. This approach combines genetic screens based on cDNA libraries with microarray detection methods to permit high-throughput functional analyses. We implement this strategy to identify genes whose overexpression can maintain phenotypic properties of undifferentiated mouse ES cells under differentiation-inducing conditions, specifically in the absence of leukemia inhibitory factor. The identified genes encode a variety of regulatory proteins whose function in ES cells was previously unknown. Moreover, our approach is capable of detecting genes whose overexpression promote differentiation or cell death. Overall, our studies establish a methodology for highly sensitive identification of genes that confer particular phenotypes on ES cells.

cDNA library | differentiation | microarray | phenotype | self-renewal

Mouse and human ES cells can be propagated extensively in culture as homogeneous self-renewing populations (1). These cells are widely used for studies of developmental processes and thought to provide a system for design of novel transplantation therapies (2, 3). *In vitro*, murine ES cells are maintained in the presence of leukemia inhibitory factor (LIF), a growth factor activating the gp130/Stat3 signaling pathway (4). Undifferentiated ES cells rapidly proliferate, form tight colonies with a characteristic morphology, possess high alkaline phosphatase activity levels, and preserve their multilineage differentiation capacities. Upon removal of LIF, ES cells undergo differentiation associated with decreased proliferation rates and morphological changes. Previous studies uncovered various aspects of ES self-renewal and differentiation (5–11). In addition, genomewide analyses have identified specific sets of transcripts and splicing variants generated in ES and other stem cell populations (12–14). However, the current understanding of molecular mechanisms regulating self-renewal and differentiation is far from complete (15). For example, only two genes, the transcription factors Nanog and Myc, have been shown to be sufficient for maintaining undifferentiated ES cells in the absence of LIF (5, 6, 16).

Gain-of-function genetic screens are a powerful method to identify genes sufficient to confer a particular cellular phenotype (17). Such screens have a long history in studies of microorganisms. Gain-of-function screens generally begin with an introduction of cDNA expression libraries into cells, followed by specific selective regimens. In mammalian cells, these screens usually require iterative rounds of selection and yield only one or few functional gene products that are identified by clone sequencing. To facilitate the isolation of small numbers of gene products amenable to direct sequencing, harsh selection conditions are used to minimize background noise. As a result, most gain-of-

function screens identify only the “most potent” phenotype-conferring genes. For example, the only currently described gain-of-function screen in mouse ES cells identified a single gene encoding the transcription factor Nanog (5). Moreover, the clone sequencing method is incapable of negative detection, complicating the identification of gene products that promote differentiation or apoptosis (18). These limitations prevent the comprehensive identification of functional genes in mammalian systems.

In this study, we applied the microarray technology as a method of large-scale parallel analysis to conduct comprehensive gain-of-function screens in ES cells (Fig. 1A). Our approach was designed to simultaneously monitor the activity of all gene products in a cDNA library as they function to mediate a given phenotype. We implemented this approach to identify genes whose overexpression is sufficient to maintain undifferentiated mouse ES cells in the face of differentiation-inducing conditions, specifically in the absence of LIF. We also show that our approach is capable of negative detection and can identify gene products that promote differentiation or cell death.

Results

Microarray-Assisted Gain-of-Function Screens. We introduced an episomal cDNA library constructed from ES cells into the E14/T mouse ES cell line (Fig. 1A; for complete experimental details, see *Materials and Methods*; see also *Supporting Materials and Methods*, which is published as supporting information on the PNAS web site). This cell line constitutively expresses the polyoma virus large T protein and can stably maintain episomal vectors carrying the polyoma origin of replication (4, 19). Transfected cells were divided into two pools and propagated for 18 days in the presence or absence of LIF: permissive and selective conditions, respectively. The decreased proliferation rate associated with differentiation served as a selection criterion in our genetic screens. We reasoned that ES cells harboring gene products that maintain rapid proliferation rates would progressively increase their relative representation in the cell population propagated without LIF. Similarly, we expected that cells harboring differentiation-inducing gene products would undergo progressive representational decreases during culture in the presence of LIF. To simultaneously identify all gene products that confer these phenotypes, we used microarrays. Episomal libraries were isolated from transfected ES cells after selection and used to prepare material for hybridization to microarrays.

Conflict of interest statement: No conflicts declared.

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

Abbreviation: LIF, leukemia inhibitory factor.

*To whom correspondence may be sent at the present address: Harvard Medical School, Massachusetts General Hospital, 50 Blossom Street, EDR 410, Boston, MA 02114. E-mail: pritsker.moshe@mgh.harvard.edu.

†To whom correspondence may be addressed. E-mail: ilemischka@molbio.princeton.edu.

© 2006 by The National Academy of Sciences of the USA

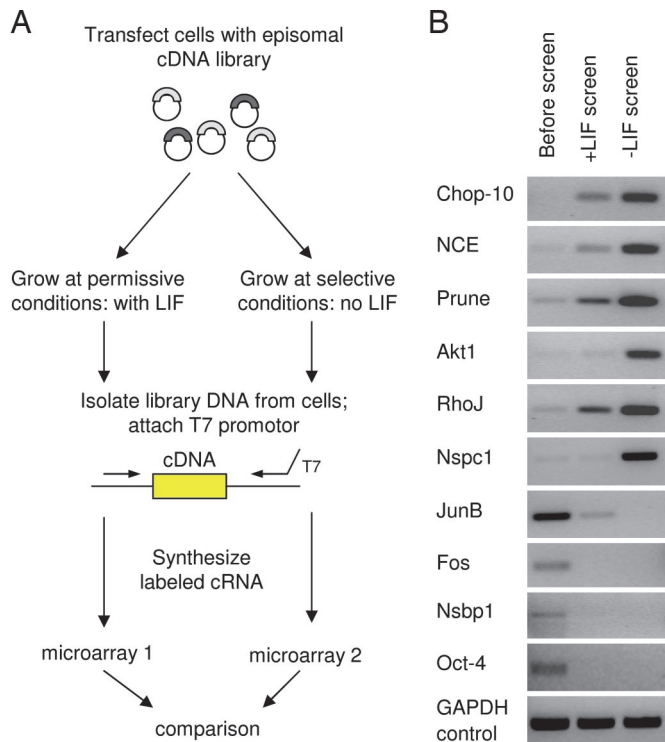


Fig. 1. Microarray-assisted gain-of-function genetic screen in ES cells. (*A*) Schematic diagram for comprehensive functional analysis. The ES cells were transfected with a cDNA library, selected and propagated at permissive or selective conditions in the presence or absence of LIF, respectively. The cDNA insert populations before and after the screen were labeled and analyzed with microarrays. (*B*) Confirmation of microarray results. Changes in cDNA representation detected with microarrays were confirmed by using PCR with gene-specific primers and libraries before and after the screen as templates. GAPDH primers were used as a control.

For analysis of obtained microarray data, we applied a 5-fold threshold to ensure identification of genes that undergo significant representational changes during selection. This threshold is higher than the 2-fold value generally applied in the microarray experiments and considered biologically significant. It is also justified statistically: Only $\approx 1\%$ of probes were identified as “increased” by using this value. We identified >560 probes (representing 499 genes) whose intensity increased at least 5-fold on microarrays hybridized with material from the “-LIF” conditions, relative to “+LIF” conditions (Table 2, which is published as supporting information on the PNAS web site). Probes representing Nanog and Oct4, whose overexpression is known to promote self-renewal and differentiation (5, 7), respectively, showed expected changes in their signal intensity (Tables 3 and 4, which are published as supporting information on the PNAS web site; microarray data). Intensity values for housekeeping control genes such as Gapdh did not show significant changes. Analysis of replicate experiments indicated reproducibility of screens conducted at the same conditions (Fig. 4, which is published as supporting information on the PNAS web site). Correlation coefficients of all microarray values obtained at the same (both permissive) or different (permissive and selective) conditions were 0.96 and 0.67, respectively. We used PCR with gene-specific primers designed toward 3' of coding regions to confirm microarray-detected representational changes above the selected 5-fold threshold. All of the tested genes were confirmed, and the examples are shown in Fig. 1*B*.

Identification of Genes that Rescue ES Cell Phenotype in the Absence of LIF. From the set of probes whose intensity values increased in the absence of LIF, we selected 25 genes for individual confir-

mation. No previous knowledge was used in selection of these genes. Full-length cDNAs derived from these genes were introduced individually into the episomal vector, and each was tested for its ability to maintain growth of ES cells in the absence of LIF. We further tested whether the identified genes can rescue other phenotypic properties of undifferentiated ES cells such as colony formation at a clonal density, characteristic morphology, alkaline phosphatase activity, and the expression of pluripotency markers. Cells expressing Nanog and EGFP served as positive and negative controls. In 11 of 25 tested cases ($\approx 40\%$), the identified genes were sufficient to at least partially “rescue” ES cell proliferation rates in the absence of LIF (Fig. 2 and Table 1). Overexpression of these genes maintained at least a 10-fold higher number of cells after 9 days without LIF, in comparison with the EGFP control. In comparison, these genes do not render a significant growth advantage in the presence of LIF. The identified genes encode an array of regulatory proteins and show different levels of rescue activity in the various assays. For example, the transcription factor Cited2 and signaling protein Prune were sufficient to maintain all tested phenotypic properties at levels observed with the positive Nanog control. On the other hand, the transcription regulator Chop10 maintained robust cell proliferation rates; however, a large fraction of transfected cells displayed differentiated morphologies. The weakly characterized Polycomb factor Nspc1 and intracellular transport regulator Snx6 partially rescued cell growth; however, their overexpression was not sufficient to allow colony formation at a clonal density. The identified genes also were capable of promoting the continued expression of the ES cell pluripotency markers Rex1, Dppa4, and Zfp57 (20, 21). Expression levels of these markers decrease in the absence of LIF.

To compare the efficiency and sensitivity of our microarray-based approach with the traditional individual clone sequencing methods, we randomly picked and sequenced 50 clones from the cDNA library isolated from ES cells cultured without LIF. None of the 11 functionally confirmed genes was found among these sequenced clones. These observations clearly demonstrate that our microarray-based approach can identify functional genes at a level of sensitivity not possible by using traditional clone-by-clone sequencing approaches.

To test the functional specificity of the identified genes in the ES cell context, we analyzed the activities of Bmi1 and Cited1. These genes are closely related homologues of Nspc1 and Cited2, respectively. Neither Bmi1 nor Cited1 was capable of rescuing the properties of undifferentiated ES cells in the absence of LIF. In fact, the overexpression of Cited1 and Bmi1 induced an opposite phenotypic effect, differentiation, both with and without LIF (Fig. 5, which is published as supporting information on the PNAS web site). These data suggest that the observed ES cell functional activity is specific for individual members of multi-gene families.

Transcriptional Regulation of the Identified Genes. It is reasonable to expect that at least some of the genes that confer LIF independence are regulated by LIF signaling at the transcriptional level. To test this hypothesis, we measured changes in their expression levels in response to withdrawal of LIF during a 5-day time period (Fig. 6, which is published as supporting information on the PNAS web site). At the used conditions, expression levels of Socs-3, a known downstream target of the LIF/Stat3 signaling pathway, rapidly decreased (22). In comparison, the expression levels of the functionally identified genes remained unchanged. These data indicate that the identified genes are not activated by LIF at the transcriptional level. Transcription of Nanog also was shown to be independent on LIF (5).

Identification of Genes that Induce Differentiation. The microarray analysis also has identified $>3,000$ probes (representing 2,953

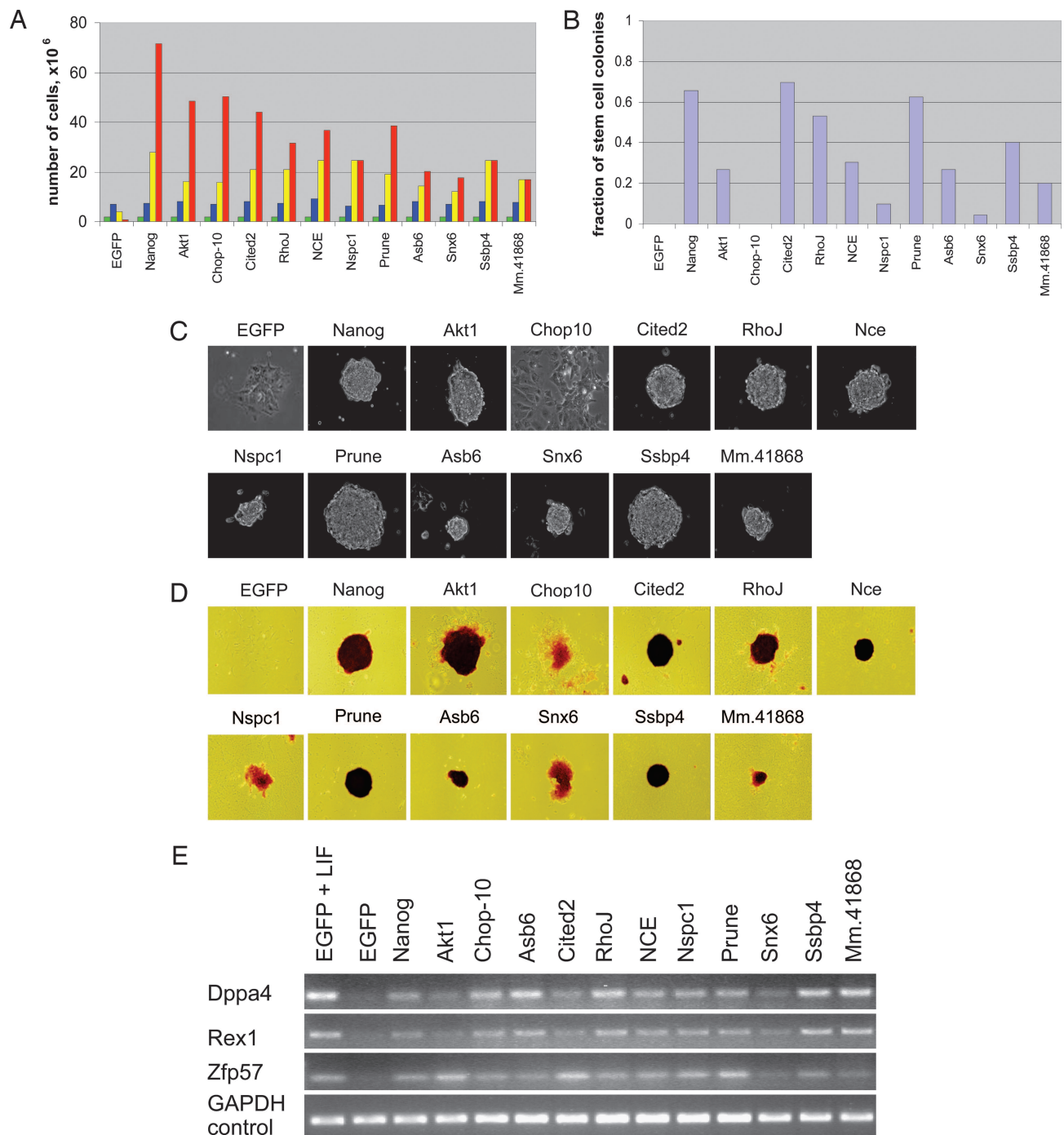


Fig. 2. Confirmation of function for identified gene products. (A) Cumulative number of cells was determined for genes that rescue ES cell proliferation in the absence of LIF. After transfection, ES cells were propagated in the absence of LIF, passaged, and counted every 3 days. Nanog and EGFP served as positive and negative controls, respectively. Colored bars indicate the following: green, day 0; blue, day 3; yellow, day 6; red, day 9. (B) Fraction of stem cell colonies formed by transfected cells at a clonal density after 7 days in the absence of LIF. (C and D) Morphology (C) and alkaline phosphatase staining (D) of colonies formed at a clonal density after 7 days in the absence of LIF. (E) The functionally identified genes rescue expression of ES cell molecular markers after 4 days in the absence of LIF, as determined by RT-PCR. Cells transfected with EGFP and Nanog served as a negative and positive control, respectively.

genes) whose intensity decreased >5 -fold in the library harboring ES cells propagated at the permissive, LIF-containing conditions. From this set, we have selected 20 genes for functional confirmation. Our functional analyses found that nine of the selected genes induced the expected phenotypic effect, differentiation, when overexpressed in ES cells (Fig. 3 and Table 4). Cells transfected with these genes decreased their proliferation rate and underwent typical morphological changes associated

with differentiation. The identified genes encode various regulatory proteins, including transcription factors, RNA-binding factors, and signaling molecules. Among the transcription factors, we found the well known protooncogenes JunB and c-Fos, transcription repressor Nab1, and ETS family member SpiC. We also identified RNA-binding proteins, the ATP-dependent helicase Ddx3, and translation regulator Pumilio 1. Other identified genes included Catenin α 1, nucleosome-binding protein

Table 1. Genes that rescue ES cell phenotypic properties

Genes	Rescued ES properties		
	Rapid growth	Morphology	Alkaline phosphatase
<i>Akt1/PKB, Mm.6645</i> , protein kinase	+	+	+
<i>Chop-10, Mm.110220</i> , DNA damage-inducible transcript 3 (<i>Ddit3</i>)	+	-	+/-
<i>Cited2, Mm.272321</i> , Cbp/p300-interacting transactivator 2	+	+	+
<i>RhoJ, Mm.27467</i> , Ras homolog gene family, member J	+	+	+
<i>NCE, Mm.337238</i> , NEDD8-conjugating enzyme	+	+	+/-
<i>Nspc1, Mm.12261</i> , nervous system polycomb 1	+	+/-	+/-
<i>Prune, Mm.14155</i> , Prune homolog (<i>Drosophila</i>)	+	+	+
<i>Asb6, Mm.27656</i> , ankyrin repeat and SOCS box-containing protein 6	+/-	+/-	+/-
<i>Snx6, Mm.28240</i> , sortin nexin 6	+	+/-	+/-
<i>Ssbp4, Mm.6667</i> , single-stranded DNA-binding protein 4	+	+	+
<i>Mm.41868</i> , hypothetical protein	+/-	+/-	+/-

Nsbp1, and ubiquitination regulator Fbxo 30. These results demonstrate that our microarray-assisted approach also is capable of negative detection and can identify genes that induce differentiation.

Sources of Noise. Although PCR analyses using gene-specific primers confirmed microarray-detected changes (Fig. 1B), the rate of functional confirmation for individual genes currently achieved in our experiments is $\approx 40\%$ (Fig. 2). The following observations suggest that this experimental "noise" may be, at least partially, due to differences between cDNAs representing identified genes in the library used in the screen and cDNAs whose function was tested individually in our confirmation experiments. For example, the microarray data showed a significant increase in representation of the IFN Response Factor 1 (IRF1) gene, which was further confirmed by PCR with specific primers. A 5' truncated cDNA corresponding to an amino terminus-deficient version of this protein was identified by

sequencing of clones isolated from cells after the screen. Surprisingly, overexpression of a full-length IRF1-encoding cDNA induced cell death, opposite to the expected phenotype, whereas cells carrying the 5' truncated construct proliferated normally. Similarly, truncated transcripts whose functions strongly differ from the full-length genes were detected in gain-of-function screen studies previously published by other groups (23). As another example, constitutive expression of a Nanog cDNA with a truncated 3' UTR induced higher rates of proliferation than the full-length version. A possible explanation is the presence of repetitive B2 elements in the 3' UTR that can alter transcription from our vector (5).

Because of incomplete cDNA synthesis reactions, cDNA libraries usually contain a significant number of clones representing truncated transcripts. Approximately 50% of the clones in the particular library used in our studies are full-length, and a similar percentage was detected among clones sequenced after selection. Alternative splicing and alternative polyadenylation affect $>50\%$ of mammalian genes, including those expressed in ES cells (14, 24). Because microarray probes generally cover the 3' portions of mRNAs and, therefore, cannot effectively distinguish between full-length, truncated, or alternative isoforms, these phenomena may contribute to false positives observed in our studies. Additional possible sources of false positives also may include, for example, artificial fluctuations in the clone representation during propagation of transfected cells in culture. However, to preserve the clone representation, the number of transfected cells used in our studies was 100-fold higher than the number of independent recombinants present in the starting cDNA library.

We expect that employing full-length and completely characterized cDNA libraries will significantly improve the confirmation rate in our approach, so that each microarray-detected change in cDNA representation could be considered as a significant indication of biological function. However, even in their current format, our studies provide a significant advancement for genetic analysis in mammalian cells, as shown by a large number of functionally identified genes.

Discussion

The development of large-scale robust methods for functional analyses is necessary to reduce the rapidly increasing gap between the vast amount of gene expression data and the knowledge of biological function. Such methods have been widely applied in studies of microorganisms, yeast, and *Escherichia coli* (25–28). In these studies, we developed a combination of cDNA

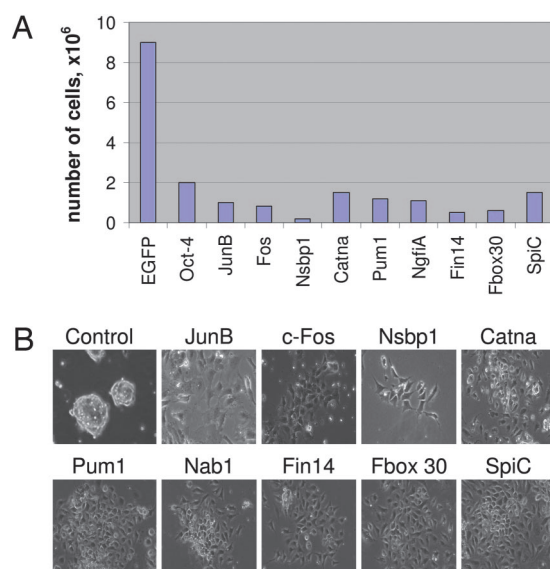


Fig. 3. Confirmation of function for genes promoting differentiation or cell death. (A) Number of cells was determined after transfected ES cells were propagated in the presence of LIF for 6 days. The EGFP construct served as a control. (B) Cell morphology of transfected cells.

library screens and microarray analyses to identify functional genes in mouse ES cells. Although previously applied in *E. coli* (28), this methodology is used here in a mammalian system. Other large-scale gain-of-function techniques developed for mammalian cells include transfections of arrayed cDNA collections and transfected-cell microarrays (29, 30). Although valuable and capable of providing genomewide functional information, these approaches require state-of-the-art procedures and special equipment such as robotic printers and analyzers. In comparison, our approach is cost-effective and easily adaptable because it can be applied with any commercially available microarrays and in any proliferating cell type after establishing moderate selective conditions. We expect this approach will provide information complementary to the previously described large-scale loss-of-function approaches based on the RNA interference technology (31, 32).

Using the developed approach, we were able to identify an array of genes that rescue the ES cell phenotypic properties in the absence of LIF and whose function in ES cells was previously unknown. The identified genes encode a wide range of regulatory proteins. We hypothesize that these genes inhibit differentiation processes induced at the withdrawal of LIF and they exert their activity through a variety of molecular pathways. For example, one of the identified genes encodes the Akt1 kinase, a known downstream component of the phosphatidylinositol 3-kinase (PI3-K) pathway (33). Inhibition of this pathway was shown to induce growth arrest and differentiation of ES cells in the presence of LIF (34, 35). Then, it is reasonable to speculate that the PI3-K pathway regulates genes that function downstream of the LIF/Stat3 pathway. The rescue activity of another gene, the Nedd8-conjugation enzyme NCE, suggests involvement of ubiquitin-mediated protein degradation in ES cell regulation. Previous studies have shown that NCE regulates levels of the cell cycle inhibitor p27 (36). Another identified gene, Prune, was shown to undergo amplification in human sarcomas and possess a cAMP phosphodiesterase activity (37). It also was shown to interact with nm23-H1, a tumor suppressor gene encoding a nucleoside diphosphate kinase. An additional gene with the rescue activity encodes the transcription factor Cited2 that was shown previously to enhance proliferation of embryonic fibroblasts (38). Interestingly, its close homologue, Cited1, induced an opposite effect, differentiation, which implies specificity of the identified functional activity across multigene families. We also monitored changes in transcriptional levels of the genes with the rescue activity to test whether any of them is directly regulated by LIF signaling. We did not detect significant changes in their expression as a result of LIF withdrawal. These results indicate that the identified genes are not downstream transcriptional targets of the LIF/Stat3 signaling pathway. A similar lack of correlation between transcriptional regulation and functional activity was observed in genomewide studies of yeast deletion mutants (27).

The set of genes inducing differentiation in ES cells included two well known protooncogenes, JunB and c-Fos, that are known to enhance proliferation and inhibit differentiation in other cell types (39). In this set, we also identified SpiC, a member of the ETS domain family of transcription factors activated in erythroleukemias (40). These data indicate that the functional activity of the identified genes depends on a specific context, which is an overall set of molecules available in a particular type of cells. Therefore, gene functional roles must be experimentally determined in each cell type of interest that further underlines the requirement for large-scale functional analyses such as those presented in these studies.

What is the role of the identified genes *in vivo*? Previous studies indicate that a deficiency in either the Cited2 or Akt1 gene is not detrimental to ES cell precursors in the mouse epiblast (41, 42). A possible explanation lies in the fundamental

difference between gain-of-function and loss-of-function analyses. Gain-of-function analysis identifies genes that are sufficient to confer a specific phenotype, whereas loss-of-function analysis identifies genes that are necessary. The former type of analysis cannot determine whether genes are essential, whereas the latter is affected by functional redundancy in a studied system. For example, cyclin genes (e.g., cyclin E), whose functional roles were firmly established in biochemical and gain-of-function experiments, were shown to be dispensable for mouse development (43). An additional possible explanation is that regulatory mechanisms in cultured ES cells do not accurately reflect regulatory pathways present in the early mouse embryo. For example, although LIF is required to maintain undifferentiated ES cells *in vitro*, mouse mutants that lack the LIF gene are viable (44). Further loss-of-function experiments are required to test whether the genes identified in our studies are essential for maintenance of the ES cell phenotype.

Ultimately, a complete list of genetic perturbations, gain and loss of function, and their phenotypic consequences will be necessary to understand complex molecular mechanisms regulating ES cell self-renewal and differentiation. Our studies establish an efficient approach to comprehensively determine gain-of-function genetic modifications that confer particular fates on ES cells. Using lineage-specific reporter lines, this approach also can be used to identify gene products inducing differentiation toward a particular cell type and, thus, is potentially useful for development of stem cell therapies.

Materials and Methods

ES Cell Culture. The E14/T ES cell line (gift of Austin Smith, University of Edinburgh) was used for most of the experiments. This cell line constitutively expresses the polyoma virus large T antigen and, therefore, can maintain plasmid vectors (episomes) carrying the polyoma virus origin of replication such as pPyCAGIP (4, 19). Cells were grown without feeder cells as described elsewhere, in the presence of 15% FBS and LIF (4).

cDNA Library. The cDNA library in pPyCAGIP (gift of Austin Smith) prepared from undifferentiated ES cells cultured on embryonic fibroblasts has been described in ref. 5. We measured the fraction of full-length cDNAs. Sequences of 11 of 20 randomly chosen clones included known initiation codons as compared with annotations at the National Center for Biotechnology Information database.

Gain-of-Function Screens in ES Cells. For the library screen, 5×10^7 E14/T cells were electroporated with 50 μg of the library DNA. The transfected cells were selected with puromycin (2 $\mu\text{g}/\text{ml}$) for 6 days in regular growth conditions. After selection, the cells were divided into two pools that were separately propagated in the presence or absence of LIF for 18 days. At the end of the screen, the library cDNA was extracted from cells, and cDNA inserts were amplified by using PCR with primers complementary to the regions just outside of the cloning site; the reverse primer was attached to the T7 promoter sequence. The products of the reaction were used for *in vitro* biotinylation transcription reaction, and the biotinylated cRNA was analyzed by using Mouse430 microarrays (Affymetrix), according to the standard procedure.

Analysis of Data from Genomewide Genetic Screens. To identify genes that increase their representation in the total cDNA population in the absence of LIF, we compared microarray probe intensity values at $-$ LIF versus $+$ LIF conditions. We identified 560 probes that were recognized by the Affymetrix system as “present” and “increased” >5 -fold (arbitrary threshold) at the $-$ LIF conditions. To identify these genes that induce differentiation or cell death and, therefore, decrease their

