
**Mechanisms of Signal Transduction:
Identification of Transcription Factor
Binding Sites Upstream of Human Genes
Regulated by the Phosphatidylinositol
3-Kinase and MEK/ERK Signaling
Pathways**

John W. Tullai, Michael E. Schaffer, Steven
Mullenbrock, Simon Kasif and Geoffrey M.
Cooper

J. Biol. Chem. 2004, 279:20167-20177.

doi: 10.1074/jbc.M309260200 originally published online February 9, 2004

Access the most updated version of this article at doi: [10.1074/jbc.M309260200](https://doi.org/10.1074/jbc.M309260200)

Find articles, minireviews, Reflections and Classics on similar topics on the [JBC Affinity Sites](#).

Alerts:

- [When this article is cited](#)
- [When a correction for this article is posted](#)

[Click here](#) to choose from all of JBC's e-mail alerts

Supplemental material:

<http://www.jbc.org/content/suppl/2004/02/17/M309260200.DC1.html>

This article cites 71 references, 44 of which can be accessed free at
<http://www.jbc.org/content/279/19/20167.full.html#ref-list-1>

Identification of Transcription Factor Binding Sites Upstream of Human Genes Regulated by the Phosphatidylinositol 3-Kinase and MEK/ERK Signaling Pathways*[§]

Received for publication, August 20, 2003, and in revised form, February 2, 2004
Published, JBC Papers in Press, February 9, 2004, DOI 10.1074/jbc.M309260200

John W. Tullai^{‡§}, Michael E. Schaffer^{§¶}, Steven Mullenbrock[‡], Simon Kasif^{¶||},
and Geoffrey M. Cooper^{‡¶**}

From the [‡]Department of Biology, [¶]Bioinformatics Program, and the ^{||}Department of Biomedical Engineering,
Boston University, Boston, Massachusetts 02215

We have taken an integrated approach in which expression profiling has been combined with the use of small molecule inhibitors and computational analysis of transcription factor binding sites to characterize regulatory sequences of genes that are targets of specific signaling pathways in growth factor-stimulated human cells. T98G cells were stimulated with platelet-derived growth factor (PDGF) and analyzed by DNA microarrays, which identified 74 immediate-early gene transcripts. Cells were then treated with inhibitors to identify subsets of genes that are targets of the phosphatidylinositol 3-kinase (PI3K) and MEK/ERK signaling pathways. Four groups of PDGF-induced genes were defined: independent of PI3K and MEK/ERK signaling, dependent on PI3K signaling, dependent on MEK/ERK signaling, and dependent on both pathways. The upstream regions of all genes in the four groups were scanned using TRANSFAC for putative *cis*-elements as compared with a background set of non-induced genes. Binding sites for 18 computationally predicted transcription factors were over-represented in the four groups of co-expressed genes compared with the background sequences ($p < 0.01$). Many of the *cis*-elements identified were conserved in orthologous mouse genes, and many of the predicted elements and their cognate transcription factors were consistent with previous experimental data. In addition, chromatin immunoprecipitation assays experimentally verified nine predicted SRF binding sites in T98G cells, including a previously unknown SRF site upstream of *DUSP5*. These results indicate that groups of human genes regulated by discrete intracellular signaling pathways share common *cis*-regulatory elements.

The identification of regulatory elements that control gene expression is one of the paramount problems in genomics and

systems biology. However, computational identification of transcription factor binding sites is difficult because they consist of short, degenerate sequences that occur frequently by chance (1–4). One approach to this problem is to search for genes that share clusters of transcription factor binding sites, for example, upstream of developmentally regulated genes (5–7). An alternative strategy limits searches for these elements to the upstream regions of genes that might be expected to be regulated by common transcription factors because they are functionally related (8) or coordinately expressed. Studies of coordinate expression have included analyses of yeast sporulation and metabolic responses (9–11), cell cycle progression in yeast and human cells (12–16), and circadian rhythmicity (17).

In the present study, we have taken an integrated approach in which microarray expression profiling has been combined with the use of small molecule inhibitors to identify candidate transcription factor binding sites in groups of genes that are regulated by specific signaling pathways in growth factor-stimulated human cells. Many growth factors stimulate receptor-protein tyrosine kinases, leading to activation of intracellular signaling pathways that modulate gene expression by altering the activity of transcription factors (18). A primary response to growth factor stimulation of mammalian cells is the transcriptional induction of ~100 immediate-early genes, whose induction results directly from the post-translational modification of pre-existing transcription factors (19). As many immediate-early genes themselves encode transcription factors, their induction results in further downstream alterations in programs of gene expression.

Growth factor receptors stimulate a variety of downstream signaling pathways, including the cAMP, JAK/STAT, MEK¹/ERK, and phosphatidylinositol 3-kinase (PI3K) pathways. We used microarray analysis to identify immediate-early genes induced by the MEK/ERK and PI3K pathways, which play critical roles in cell proliferation and survival. Activation of the MEK/ERK pathway is mediated by the Raf protein kinases, which are coupled to growth factor receptors by Ras proteins (20). Once activated, ERK phosphorylates a variety of targets, including transcription factors and the protein kinase Rsk. Stimulation of growth factor receptors also results in activation of PI3K, leading to formation of the membrane phospholipid PIP₃. PIP₃ activates several downstream targets, including the protein kinase Akt, which plays a critical role in cell survival

* This work was supported by Grants R01 CA18689 and P20 GM66401 and fellowship F32 GM067392 (to J. W. T.) from the National Institutes of Health, and D90-9870710 and KDI-9980088 from the National Science Foundation. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[§] The on-line version of this article (available at <http://www.jbc.org>) contains Supplementary Data.

The microarray gene expression data from this study has been submitted to GEO (Gene Expression Omnibus) under accession number GSE1128.

[§] These authors contributed equally to this study.

** To whom correspondence should be addressed: Boston University, Dept. of Biology, 5 Cummington St., Boston, MA 02215. Tel.: 617-353-8735; Fax: 617-353-8484; E-mail: gmcooper@bu.edu.

¹ The abbreviations used are: MEK, mitogen-activated protein kinase/extracellular signal-regulated kinase; ERK, extracellular signal-regulated kinase; PI3K, phosphatidylinositol 3-kinase; PDGF, platelet-derived growth factor; PIP₃, phosphatidylinositol 3,4,5-trisphosphate; SRE, serum response element; SRF, serum response factor; RT-PCR, reverse transcription polymerase chain reaction.

TABLE I
Genes induced by PDGF

T98G cells were rendered quiescent and then stimulated by treatment with human PDGF-BB for 30 minutes. The values for each gene represent the mean average \log_2 ratio and standard deviation for dye-swap normalized determinations (*N*) comparing five independent cultures of PDGF-stimulated versus non-stimulated cells. Some genes were represented more than once on the array and thus have more than five determinations. Only genes induced >2-fold are presented and were used in subsequent analysis. Each gene is represented by the Unigene gene name and GenBank™ accession number provided with the microarrays.

| ID ^a | Gene symbol ^b | Gene name | Acc. no. ^c | Log ₂ avg. fold induction \pm S.D. | N | ID | Gene symbol | Gene name | Acc. no. | Log ₂ avg. fold induction \pm S.D. | N |
|-----------------|--------------------------|--|-----------------------|---|----|----|-----------------|--|-----------|---|----|
| 1 | <i>FOS</i> | v-Fos FBJ osteosarcoma viral oncogene homolog | V01512 | 6.4 \pm 0.48 | 5 | 38 | <i>ADRB2</i> | Adrenergic, β 2-, receptor, surface | M15169 | 1.7 \pm 0.21 | 5 |
| 2 | <i>UNG2</i> | Uracil-DNA glycosylase 2 | AA291356 | 5.8 \pm 0.68 | 5 | 39 | <i>TIEG</i> | TGF β inducible early growth response | AF050110 | 1.6 \pm 0.16 | 9 |
| 3 | <i>NR4A1</i> | Nuclear receptor subfamily 4, group A, member 1 | NM_002135 | 5.6 \pm 0.92 | 5 | 40 | | Human CpG island DNA genomic MseI fragment | Z63118 | 1.6 \pm 0.44 | 5 |
| 4 | <i>DUSP1</i> | Dual specificity phosphatase 1 | X68277 | 5.1 \pm 0.67 | 5 | 41 | <i>RXRG</i> | Retinoid X receptor, γ | U38480 | 1.6 \pm 0.51 | 2 |
| 5 | <i>ZFP36</i> | Zinc finger protein homologous to Zfp-36 in mouse | M92844 | 4.6 \pm 0.42 | 10 | 42 | <i>DUSP6</i> | Dual specificity phosphatase 6 | AB013382 | 1.6 \pm 0.50 | 5 |
| 6 | <i>NR4A3</i> | Nuclear receptor subfamily 4, group A, member 3 | X89894 | 4.4 \pm 1.14 | 4 | 43 | <i>CD44</i> | Human cell surface glycoprotein CD44 | L05411 | 1.6 \pm n/a | 1 |
| 7 | <i>EGR2</i> | Early growth response 2 | J04076 | 4.2 \pm 0.63 | 5 | 44 | | Human proto-oncogene Bcl | U51869 | 1.5 \pm 0.24 | 5 |
| 8 | <i>NR4A2</i> | Nuclear receptor subfamily 4, group A, member 2 | X75918 | 4.2 \pm 0.30 | 5 | 45 | <i>PLAU</i> | Plasminogen activator, urokinase | M15476 | 1.5 \pm 0.28 | 9 |
| 9 | <i>EGR3</i> | Early growth response 3 | X63741 | 4.1 \pm 0.85 | 10 | 46 | <i>SOCS3</i> | STAT induced STAT inhibitor 3 | AB006967 | 1.5 \pm 0.43 | 3 |
| 10 | <i>FOSB</i> | FBJ murine osteosarcoma viral oncogene homolog B | L49169 | 3.9 \pm 0.38 | 3 | 47 | <i>PIM1</i> | Pim-1 oncogene | M24779 | 1.5 \pm 0.28 | 5 |
| 11 | <i>JUNB</i> | Jun B proto-oncogene | U20734 | 3.8 \pm 0.33 | 5 | 48 | <i>MCL1</i> | Myeloid cell leukemia sequence 1 | L08246 | 1.4 \pm 0.10 | 5 |
| 12 | <i>ATF3</i> | Activating transcription factor 3 | L19871 | 3.7 \pm 0.62 | 10 | 49 | <i>EBI2</i> | Epstein-Barr virus induced gene 2 | L08177 | 1.4 \pm 0.49 | 3 |
| 13 | <i>ETR101</i> | Immediate early protein | AA194084 | 3.4 \pm 0.40 | 5 | 50 | <i>CCL2</i> | Human gene for JE protein | X60001 | 1.4 \pm 0.32 | 4 |
| 14 | <i>CTGF</i> | Connective tissue growth factor | U14750 | 3.3 \pm 0.36 | 5 | 51 | <i>ARHE</i> | Ras homolog gene family, member E | W03441 | 1.4 \pm 0.38 | 5 |
| 15 | <i>BRAP</i> | BRCA1-associated protein | AW804509 | 3.2 \pm 0.35 | 5 | 52 | | Human nuclear lamin A and nuclear lamin C gene | L12401 | 1.4 \pm n/a | 1 |
| 16 | <i>CYR61</i> | Cysteine-rich, angiogenic inducer, 61 | Y12084 | 3.0 \pm 0.37 | 5 | 53 | | ESTs | AI023436 | 1.3 \pm 0.42 | 5 |
| 17 | <i>IL6</i> | Interleukin 6 (interferon, β 2) | X04430 | 3.0 \pm 0.30 | 5 | 54 | <i>RGS1</i> | Regulator of G-protein signalling 1 | S59049 | 1.3 \pm 1.26 | 3 |
| 18 | <i>DUSP5</i> | Dual specificity phosphatase 5 | U15932 | 2.8 \pm 0.16 | 10 | 55 | <i>RGS2</i> | Regulator of G-protein signalling 2, 24kD | AI652515 | 1.3 \pm 0.28 | 10 |
| 19 | <i>C8FW</i> | Phosphoprotein regulated by mitogenic pathways | AJ000480 | 2.8 \pm 0.15 | 4 | 56 | <i>COPEB</i> | Core promoter element binding protein | AL037844 | 1.3 \pm 0.28 | 5 |
| 20 | <i>SYK</i> | Spleen tyrosine kinase | Z29630 | 2.7 \pm 1.04 | 5 | 57 | <i>KIAA0469</i> | KIAA0469 gene product | AB007938 | 1.2 \pm 0.18 | 8 |
| 21 | <i>EDN1</i> | Endothelin 1 | S56805 | 2.6 \pm 0.07 | 2 | 58 | <i>CBX4</i> | Chromobox homolog 4 (Drosophila Pc class) | AF013956 | 1.2 \pm 0.15 | 4 |
| 22 | | Sequence 49 from Patent WO9951727. | AX015384 | 2.5 \pm n/a | 1 | 59 | <i>SNK</i> | Serum-inducible kinase | NM_006622 | 1.2 \pm 0.25 | 5 |
| 23 | <i>PBEF</i> | pre-B-cell colony-enhancing factor | U02020 | 2.4 \pm 0.28 | 5 | 60 | <i>PHLDA1</i> | Pleckstrin homology-like domain, family A, member 1 | AF220656 | 1.2 \pm 0.46 | 5 |
| 24 | <i>TNFAIP3</i> | Tumor necrosis factor, α -induced protein 3 | AL157444 | 2.4 \pm 0.58 | 10 | 61 | <i>SLC21A3</i> | Solute carrier family 21, member 3 | U21943 | 1.2 \pm 0.55 | 5 |
| 25 | | Sequence 4 from Patent WO0017232 | AX024732 | 2.4 \pm 0.24 | 4 | 62 | <i>GADD45A</i> | Growth arrest and DNA-damage-inducible, alpha | AW025439 | 1.1 \pm 0.45 | 10 |
| 26 | <i>JUN</i> | v-Jun avian sarcoma virus 17 oncogene homolog | AI885769 | 2.4 \pm 0.34 | 10 | 63 | <i>MYC</i> | v-Myc viral oncogene homolog | J00120 | 1.1 \pm 0.36 | 14 |
| 27 | <i>SGK</i> | Serum/glucocorticoid regulated kinase | Y10032 | 2.3 \pm 0.26 | 4 | 64 | <i>F3</i> | Coagulation factor III (thromboplastin, tissue factor) | AI085165 | 1.1 \pm 0.28 | 15 |
| 28 | <i>IER3</i> | Immediate early response 3 | AI022951 | 2.3 \pm 0.28 | 5 | 65 | | Human thrombospondin-1 gene, partial cds | U12471 | 1.1 \pm 0.70 | 4 |

TABLE I—continued

| ID ^a | Gene symbol ^b | Gene name | Acc. no. ^c | Log ₂ avg. fold induction ±S.D. | N | ID | Gene symbol | Gene name | Acc. no. | Log ₂ avg. fold induction ±S.D. | N |
|-----------------|--------------------------|--|-----------------------|---|----|----|-------------|--|----------|---|----|
| 29 | | Sequence 12 from Patent WO9954460 | AX013690 | 2.2 ± 0.34 | 5 | 66 | FOXC2 | Forkhead box C2 | Y08223 | 1.1 ± 0.18 | 5 |
| 30 | BMP6 | Bone morphogenetic protein 6 | AA426586 | 2.0 ± 0.15 | 4 | 67 | SRF | Serum response factor | J03161 | 1.1 ± 0.14 | 10 |
| 31 | BHLHB2 | Basic helix-loop-helix domain containing, class B, 2 | AB004066 | 1.9 ± 0.59 | 5 | 68 | TOB1 | Transducer of ERBB2, 1 | D38305 | 1.1 ± 0.21 | 5 |
| 32 | GEM | GTP-binding protein overexpressed in skeletal muscle | AW297828 | 1.9 ± 0.53 | 10 | 69 | CCL8 | Small inducible cytokine subfamily A, member 8 | AI590222 | 1.1 ± 0.28 | 9 |
| 33 | BCL3 | Homologous to members of the IκB family | AAC51348 | 1.9 ± 0.87 | 4 | 70 | | Human calcium transporting ATPase (ATP2B1) | L14562 | 1.1 ± 0.63 | 2 |
| 34 | SLC2A3 | Human glucose transporter 3 | AF274889 | 1.8 ± 0.25 | 5 | 71 | MGC3101 | Homo sapiens cDNA FLJ12582 fis | AI042427 | 1.1 ± 0.28 | 4 |
| 35 | IL8 | Interleukin 8 | M17017 | 1.8 ± 0.14 | 8 | 72 | LIF | Leukemia inhibitory factor | X13967 | 1.0 ± 0.49 | 5 |
| 36 | LOC57018 | Cyclin L ania-6a | AK022974 | 1.7 ± 0.12 | 5 | 73 | PPP1R15A | Growth arrest and DNA-damage-inducible 34 | AK001361 | 1.0 ± 0.18 | 4 |
| 37 | EGR1 | Early growth response 1 | AA399119 | 1.7 ± 0.68 | 9 | 74 | CEBPB | CCAAT/enhancer binding protein (C/EBP), β | W93514 | 1.0 ± 0.20 | 5 |

^a ID, gene identification number used in Fig. 1.

^b Gene symbol, LocusLink gene symbol.

^c Acc. no., GenBank™ accession number.

(21). Like ERK, Akt and other targets of PI3K signaling phosphorylate and activate transcription factors, leading to the rapid induction of immediate early genes.

Since induction of immediate-early genes is directly linked to signaling pathways that target transcription factors, genes that are responsive to a common signaling pathway might be expected to share transcription factor binding sites. We therefore sought to identify regulatory elements of genes induced by PI3K and MEK/ERK signaling, using a statistical analysis to identify transcription factor binding sites that were over-represented in the genomic regions upstream of groups of co-expressed genes. This approach identified binding sites for a limited number of transcription factors that were present at a high frequency upstream of genes regulated by specific signaling pathways. Many of the transcription factors predicted as regulators of immediate-early genes were established targets of the appropriate signaling pathways, and many of the predicted transcription factor binding sites were consistent with published experimental data and/or conserved in orthologous mouse genes. In addition, predicted binding sites for serum response factor (SRF) were confirmed directly by chromatin immunoprecipitation. It thus appears that biologically relevant transcription factor binding sites can be identified in groups of genes regulated by common signaling pathways in mammalian cells.

EXPERIMENTAL PROCEDURES

Cell Culture and Treatments—T98G human glioblastoma cells were grown in Minimal Essential Medium (Invitrogen) supplemented with fetal calf serum (10%). For growth factor/inhibitor treatments, cells were incubated in serum-free medium for 72 h, and either left unstimulated, or stimulated for 30 min with human PDGF-BB (50 ng/ml) (Sigma). U0126 (10 μM) (Cell Signaling Technology) and LY294002 (50 μM) (BioMol) were added 60 min prior to PDGF addition.

Immunoblots—In parallel to all microarray experiments, the activities of PI3K and MEK/ERK signaling pathways were assessed by immunoblotting cell lysates. Proteins were separated by electrophoresis in 8% SDS-polyacrylamide gels, electroblotted to nitrocellulose membranes, and probed with anti-phospho-Akt or anti-phospho-ERK antibodies (Cell Signaling Technologies) as recommended by the manufacturer. Blots were visualized using horseradish peroxidase-linked secondary antibody, and chemiluminescence (Amersham Biosciences).

RNA Preparations and Microarray Processing—Agilent Human I cDNA microarrays, containing PCR-amplified cDNA clones, were pro-

cessed per manufacturer's guidelines. Briefly, RNA was isolated from multiple harvests of unstimulated and stimulated cells using TRIzol (Invitrogen) and RNeasy (Qiagen) protocols. Total RNA was oligo(dT) primed and reverse-transcribed in the presence of cyanine-coupled dCTP (PerkinElmer Life Sciences). Cyanine 3-dCTP and cyanine 5-dCTP dye-swap hybridizations were performed. Dye-swap determinations compared PDGF-stimulated cells in the presence or absence of inhibitor *versus* unstimulated cells. Arrays were scanned with a GenePix 4000B scanner (Axon Instruments) with photomultiplier tube settings adjusted to eliminate signal saturation and provide an average Cyanine 3/Cyanine 5 intensity ratio of 1 across each array. GenePix Pro software (version 3.0) (Axon Instruments) was used to determine the Cyanine 3 and Cyanine 5 intensities for each array feature and the surrounding background. Following local background subtraction, the median intensities for each dye-swap pair were used to calculate the average log₂ ratio for each feature (22).

Quantitative RT-PCR—Total RNA preparations for the microarray hybridizations were used in quantitative reverse transcription polymerase chain reactions (RT-PCR). Reverse transcription of 0.25 μg of total RNA was performed in 20 μl using SYBR green RT-PCR reagents and random hexamer primers (Applied Biosystems) as recommended by the manufacturer. Following a 95 °C incubation for 10 min, forty cycles of PCR (95 °C/15 s; 60 °C/1 m), were then performed on an ABI Prism 7900HT Sequence Detection System with 1 μl of the RT reaction, 100 nM PCR primers (see Supplementary Table I for primer sequences), and SYBR Green PCR Master Mix in 10-μl reactions. Threshold cycles (C_T) for four replicate reactions were determined using Sequence Detection System software (version 2.0, release 4) and relative transcript abundance calculated following normalization with an 18 S ribosomal PCR amplicon. Amplification of only a single species was verified by a dissociation curve for each reaction.

Identification of Upstream Sequences—Transcription start sites relative to the human genome sequence were obtained for 64 of the 74 PDGF-induced genes from the LocusLink data base (www.ncbi.nlm.nih.gov/LocusLink/). The 5' annotations for 13 of these transcripts were extended an average of 124 bases using the Data base of Transcription Start Sites (March 11, 2002 release) (23). Human genomic BLAST (www.ncbi.nlm.nih.gov/BLAST/) was then used to verify the position of each transcript in the genome and 1-kb upstream sequences were extracted from the corresponding GenBank™ contig records (www.ncbi.nlm.nih.gov/Entrez/). This work was based on build 29 of the human genome assembly maintained by the National Center for Biotechnology Information.

Identification of Transcription Factor Binding Sites—The computer program Match (version 1.4.1), distributed with the TRANSFAC Professional data base (BioBase Biological Databases), was used to identify putative transcription factor binding sites within each upstream se-

quence (24). The 400 vertebrate position weight matrices in TRANSFAC (version 6.1) were used to score every position along each promoter sequence. In order to identify the maximum number of candidate transcription factor binding sites, all positions with scores greater than predefined Match thresholds that minimize false negatives (minFN14.prf; false negative rate of 10%) were considered matches in the subsequent analysis. To prevent a bias introduced by palindromic or internally repetitive *cis*-regulatory elements, overlapping matches, including on opposite DNA strands, were defined as a single match.

Statistical Analysis of the Site Frequencies—The statistical significance of the frequency of a *cis*-regulatory element in each of the four groups of co-expressed genes was assessed by comparison against the average frequency in 194 genes expressed in both PDGF-treated samples and controls. This background set of upstream regions consisted of genes not induced by PDGF, with average \log_2 ratios limited to between -0.005 and 0.005 and standard deviations less than 0.25 following PDGF treatment. The upstream sequences for each gene were obtained in the same manner as the induced genes. To identify statistically over-represented binding sites in the PDGF-induced co-expressed gene groups, the mean number of sites identified per upstream region in each co-expressed gene group was compared with the mean per upstream region in the background group with a one-tailed two-sample Student's *t* test. In addition, a non-parametric permutation test, which does not assume a normal distribution, was used to ensure the validity of the Student's *t* test for the analysis. For each matrix, a permutation test was employed by randomly permuting the group labels of the background and promoter upstream sequences, and a *t*-value generated from the mean number of sites identified in the shuffled groups (25). After 10,000 permutations, the *t*-values were sorted, and a *p* value determined based on relative rank of the unpermuted *t*-value among the ordered list of *t*-values from the permuted groups.

Comparison with Orthologous Mouse Sequences—We identified mouse orthologs for 65 PDGF-induced genes using the mouse homology map information found in LocusLink. A 1-kb nucleotide sequence upstream of the reported mouse transcription start site was used as input to the previously described Match program. The human and mouse sequences were then aligned using the Needleman-Wunsch global alignment tool found in version 2.5.0 of The European Molecular Biology Open Software Suite (26). The gap open and extension penalties were set at 50.0 and 3.0 , respectively, and the nucleotide-scoring scheme of match 10 , mismatch -9 was used. The positions of each site identified in the human sequence were mapped to positions in the aligned mouse sequence, and sites occurring in both organisms at the same alignment position were recorded.

Chromatin Immunoprecipitation—Chromatin immunoprecipitations were performed as described (27), with the following modifications. T98G cells were scraped and formaldehyde fixed at 37°C for 10 min. Shearing was performed to yield 500–1500 bp chromatin fragments with a Branson Sonifier 250, using four 30-s pulses at 25% output. Samples were pre-cleared with sonicated salmon sperm DNA/Protein A agarose (50% slurry) and immunoprecipitated overnight at 4°C using $4\ \mu\text{g/ml}$ anti-SRF antibody (Santa Cruz Biotechnology, sc-335) (28). Complexes were then washed successively in low salt wash (0.01% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, 150 mM NaCl, pH 8.1), high salt wash (0.01% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, 500 mM NaCl, pH 8.1), LiCl wash (0.25 M LiCl, 1% IGEPAL-Ca 630, 1% deoxycholic acid, 1 mM EDTA, 10 mM Tris-HCl, pH 8.1), and twice in 10 mM Tris-HCl, 1 mM EDTA pH 8.0. Cross-links were reversed for 6 h at 65°C , and samples were proteinase K treated for 2 h at 45°C , followed by purification using a Qiagen Gel Extraction kit (Qiagen). Immunoprecipitated chromatin was quantified with real-time PCR as described above, using primers that either flanked the predicted site or amplified a fragment within 134 bp of the predicted site (see Supplementary Table I for primer sequences). Each PCR reaction was carried out in quadruplicate and results for each promoter region are derived from at least two independent chromatin immunoprecipitations. Data were normalized to input and are presented as fold increase over *GAPDH*, a standard negative control for SRF chromatin immunoprecipitations (28).

RESULTS AND DISCUSSION

Identification of Immediate-Early Genes Induced by the PI3K and MEK/ERK Pathways—Microarray analysis was used to identify immediate-early genes induced by platelet-derived growth factor (PDGF) stimulation of quiescent T98G human glioblastoma cells, which were chosen for these experiments because they undergo reversible cell cycle arrest upon serum deprivation (29, 30). Seventy-four genes were reproducibly in-

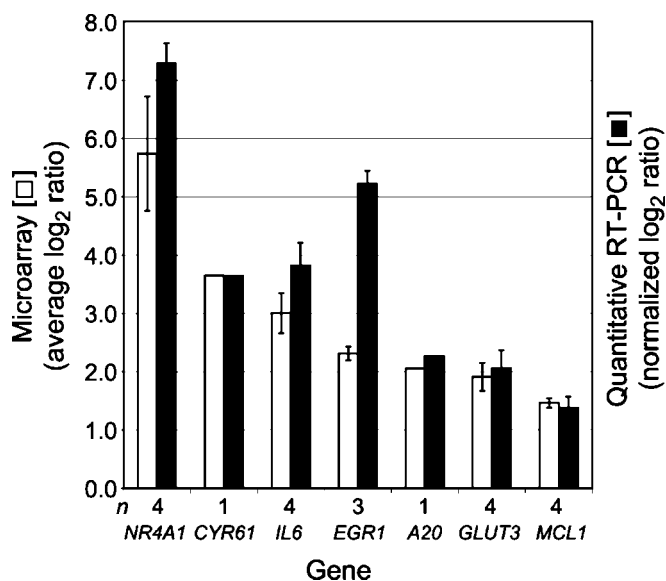


FIG. 1. Quantitative RT-PCR validation of microarray data. Total RNA samples for microarray hybridizations were tested in parallel with quantitative RT-PCR. PCR primers for 18 S ribosomal RNA were used to normalize the amount of RNA and the relative mRNA transcript levels between PDGF-treated and untreated samples were determined for 7 genes. The mean \pm S.D. is shown for both methods, where *n* indicates the number of RNA samples. Average \log_2 ratios from the microarrays and from real-time PCR are comparable units.

duced >2 -fold following 30 min of PDGF stimulation, the optimal time for induction of the immediate-early genes *fos* and *jun* (Table I). Gene inductions ranged from 2-fold to more than 80-fold ($2^{6.4}$) upon growth factor treatment, and were highly reproducible as evidenced by the standard deviations. Further, analysis of several representative genes by quantitative RT-PCR confirmed the array data (Fig. 1). The number of genes induced was in good agreement with other studies examining immediate early gene induction, and included expected genes such as *fos*, *jun*, *myc*, and *mcl1* (19, 31, 32).

Genes induced specifically by the PI3K and MEK/ERK pathways were determined using small molecule inhibitors of PI3K (LY294002) (33) and MEK (U0126) (34). As expected, LY294002 inhibited phosphorylation of Akt, whereas U0126 inhibited phosphorylation of ERK (Fig. 2A). Furthermore, U0126 did not affect Akt phosphorylation and LY294002 had no effect on ERK phosphorylation, demonstrating the specificity of the inhibitors for each pathway. Gene expression profiles were then determined by analysis of PDGF-stimulated cells pretreated with inhibitors. Representative gene targets in inhibitor-treated cells (and appropriate vehicle controls) were validated using quantitative RT-PCR (data not shown).

Some genes were primarily inhibited by LY294002 or U0126, indicating that they were induced principally by either PI3K or MEK/ERK signaling, respectively, whereas others were affected by both of these pathways (Fig. 2B). In contrast, the induction of some genes was not significantly inhibited by either LY294002 or U0126 alone. Although these genes could be induced by a distinct PDGF-stimulated pathway, it is also possible that they could be responsive to both PI3K and MEK/ERK signaling, with either pathway alone being sufficient to induce gene expression. These alternatives were distinguished by treatment of cells with both LY294002 and U0126 in combination (Fig. 2C), which identified seven genes that were significantly inhibited (>2 -fold) by both inhibitors in combination but not by either inhibitor alone. Induction of these genes can therefore be interpreted as being controlled by both PI3K and MEK/ERK signaling, with either pathway alone being sufficient for transcriptional activation.

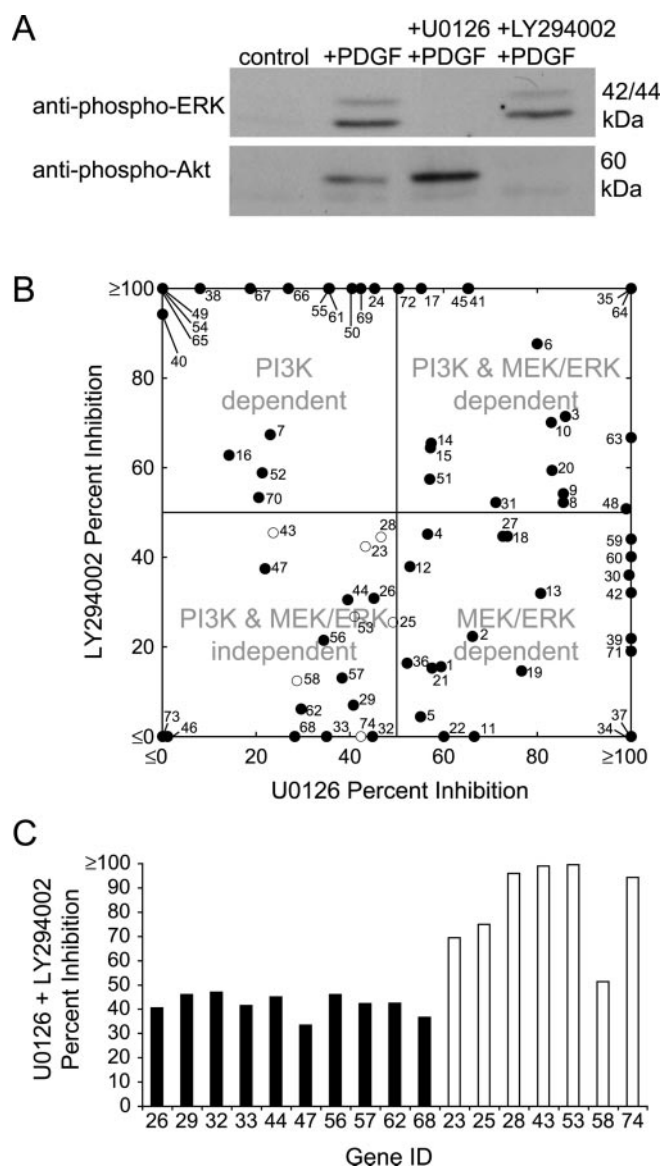


FIG. 2. Effect of PI3K and MEK inhibitors on gene induction. Cells were treated with PDGF as described in Table I. U0126 and LY294002 were added 60 min prior to PDGF addition. Dye swap determinations compared PDGF-stimulated cells in the presence and absence of inhibitor *versus* unstimulated cells. **A**, anti-phospho-Akt and anti-phospho-ERK immunoblots. Blots were stripped and reprobbed with anti-Akt and anti-ERK antibodies (Cell Signaling Technologies) to confirm equal loading of lanes. Vehicle controls for U0126 and LY294002 (Me₂SO and ethanol, respectively) were performed and had no effect when applied alone. **B**, percent inhibition of each of the 74 PDGF-induced transcripts by LY294002 or U0126 compared with their induction with PDGF alone. Data are averages of duplicate microarray analyses with independent sets of cultures. Data plotted on the *top* and *right* axes of the graph are $\geq 100\%$; data plotted on the *bottom* and *left* axes are $\leq 0\%$ (percent inhibitions are relative to the PDGF-induced value, and thus can be greater than 100% or less than 0%). Four groups of genes were defined, based on 50% inhibition with each inhibitor. Seven genes were inhibited more than 50% by both inhibitors in combination, but not by either inhibitor alone. These genes are indicated with open circles and were included in the PI3K- and MEK/ERK-dependent group for subsequent analysis. **C**, percent inhibition for genes in the PI3K- and MEK/ERK-independent group by LY294002 and U0126 in combination. White bars indicate the 7 genes with percent inhibition greater than 50% in the presence of both inhibitors. Two genes (46 and 73) are not shown because microarray data was not available for these genes for the double-inhibitor sample.

Identification of Transcription Factor Binding Sites in PDGF-induced Genes—To test for common transcription factor binding sites, the PDGF-induced genes were divided into four

groups (quadrants of Fig. 2B): PI3K- and MEK/ERK-independent (12 genes), PI3K-dependent (16 genes), MEK/ERK-dependent (21 genes), and dependent on both pathways (25 genes). Assignment was based on 50% inhibition by the appropriate inhibitors, which correlated with significant inhibition ($p < 0.05$). The seven genes that were not inhibited by LY294002 or U0126 alone, but were inhibited by both in combination, were classified as dependent on both pathways.

Sequences upstream of each transcription start site were obtained for 64 of 74 PDGF-induced genes from GenBank™ (PI3K- and MEK/ERK-independent, 10 genes; PI3K-dependent, 11 genes; MEK/ERK-dependent, 20 genes; dependent on both pathways, 23 genes), and each group of genes was analyzed using 400 vertebrate transcription factor binding site matrices from TRANSFAC (24). We limited the analysis to 1 kb to reduce detection of randomly occurring sequences. Although *cis*-regulatory elements are widely distributed throughout mammalian genomes, high concentrations of these elements often occur in proximal promoter regions. Based on published data in TRANSFAC, 82% of *cis*-regulatory elements that have been identified upstream of human genes occur within this 1-kb window.

To determine whether a transcription factor binding site was over-represented within a group of genes induced by a specific pathway (PI3K- and MEK/ERK-independent, PI3K-dependent, MEK/ERK-dependent, and PI3K- and MEK/ERK-dependent), we compared the frequency of sites within each group of upstream sequences to the background frequency in upstream sequences of 194 genes that were expressed in T98G cells, but were not induced by PDGF. The analysis was restricted to 230 matrices that detected no more than one site per kilobase in these background sequences, in order to focus on the most informative matrices. To identify a collection of sites that were statistically over-represented in the groups of PDGF-induced genes, the mean number of sites for each matrix per upstream region in each of the 4 groups of co-expressed genes was compared with the mean number of sites per upstream region in the background set of non-induced genes. The distribution of predicted transcription factor binding sites in the background set of upstream regions was approximately normal (see Supplementary Fig. 1), so a one-tailed two-sample Student's *t* test was used to identify transcription factor binding sites that occurred more frequently on average in each set of co-expressed genes compared with the background ($p \leq 0.01$). To independently validate the results of the *t* test, the analysis was compared with a more stringent non-parametric statistical method based on permutation testing. Following 10,000 iterations, ranked results from a permutation test revealed a set of statistically significant matrices that were similar to the Student's *t* test results. A comparison of the transcription factors identified by these two tests is discussed below (see Table II).

The distribution of the transcription factor binding sites identified in each group of co-expressed genes is presented in Fig. 3. For each matrix, the average frequency of sites identified relative to background is plotted on the x-axis, and the percentage of genes containing at least one site on the y-axis. For most matrices, the average frequency of sites in the induced genes did not differ significantly from background. However, some matrices identified sites with high frequencies above background, generally in a substantial fraction of genes. The average frequency of sites identified by 40 matrices indicated statistical over-representation ($p \leq 0.01$) in one or more groups (14 in the PI3K- and MEK/ERK-independent group, 25 in the PI3K-dependent group, 8 in the MEK/ERK-dependent group, and 13 in the PI3K- and MEK/ERK-dependent group). With a Student's *t* test *p* value threshold of 0.01, we expect one

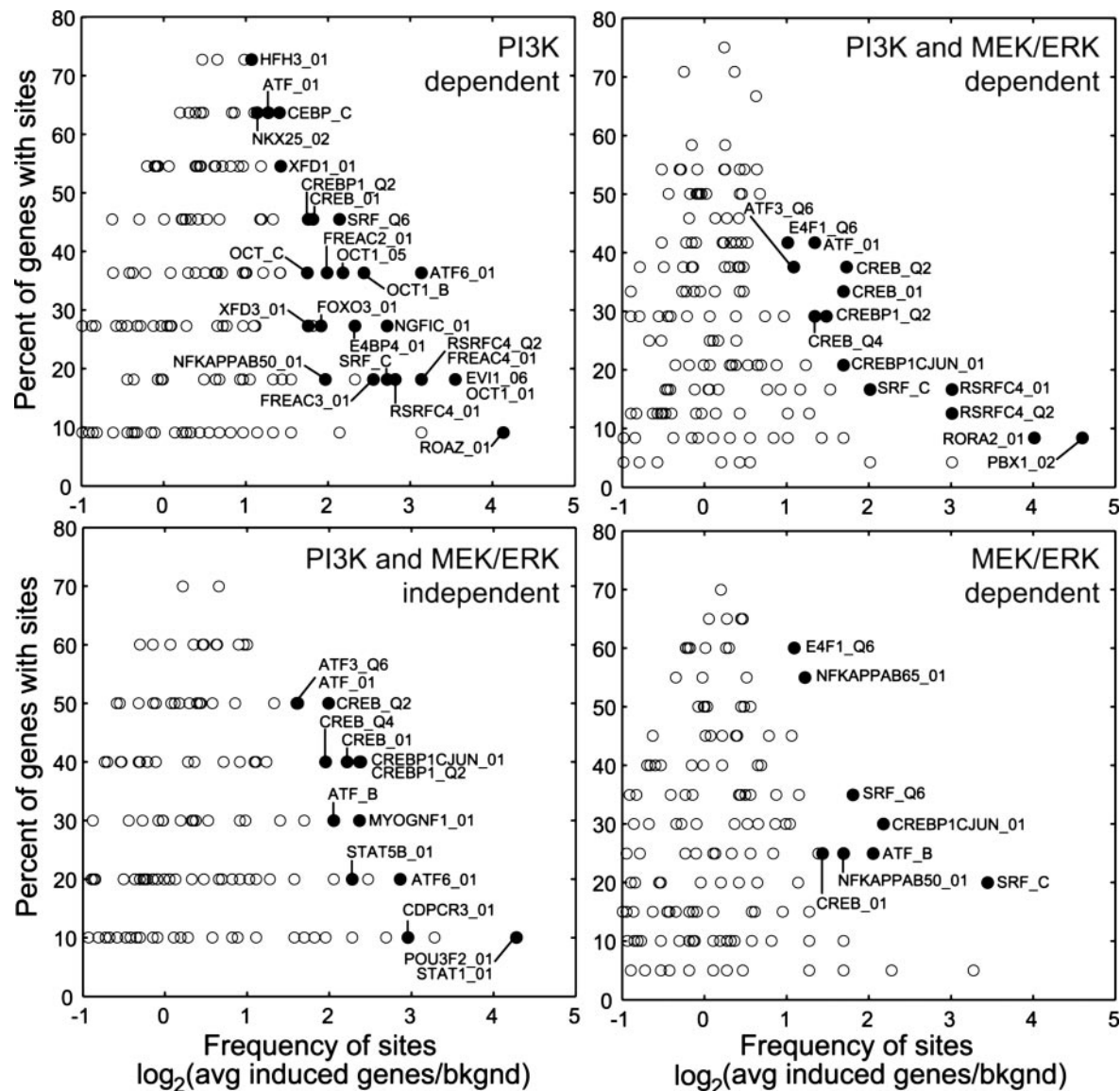


FIG. 3. **Distribution of transcription factor binding sites within groups of co-expressed genes.** Each graph represents the distribution of binding sites identified by vertebrate matrices from the TRANSFAC data base within one of the four sets of co-expressed genes. The frequency of binding sites detected by each matrix within the upstream regions of each group of induced genes relative to a background set of non-induced genes is presented as a \log_2 ratio (x-axis). The percentage of genes in each group with at least one identified binding site for each matrix is plotted on the y-axis. Matrices that were significantly over-represented ($p \leq 0.01$) in each group of induced genes are indicated by filled circles and labeled with the corresponding TRANSFAC identifier (without V\$ prefix). All matrices plotted with p values and background scores can be found in Supplementary Table II.

false positive (Type I) error in 100 such tests. Multiple hypothesis testing with the 230 matrices used in our analysis would thus be expected to yield 2.3 false-positives in the statistically significant matrices from each group of co-expressed genes. Therefore, the number of matrices identified in each group of co-expressed genes is substantially greater than would be expected by chance.

Confirmation of SRF Binding Sites—Several approaches were used to assess the validity of the computational predictions. First, the predicted transcription factor binding sites were compared with published experimental data. Second, predicted sites were analyzed for conservation in the mouse, as physiologically relevant transcription factor binding sites are frequently conserved in the non-coding regions of orthologous genes (35). Next, we asked whether the transcription factor(s) deduced from the predicted binding sites were known to be regulated by the relevant signaling pathway. In addition, the sites predicted by matrices that represent serum response factor (SRF) binding sites were further tested by chromatin immunoprecipitation.

A detailed example of the verifications for the well-studied transcription factor, SRF, is presented in Fig. 4. Consistent with activation of SRF by both PI3K and MEK/ERK pathways (36), the SRF matrix, V\$SRF_C, detected a significant number of sites in genes induced by these pathways. Sixteen SRF binding sites (serum response elements, or SREs) were found in 10 promoter regions. Thirteen of these had previously been identified, verifying the computational predictions (Fig. 4A). In addition, there were 3 genes (*CYR61*, *JUNB*, and *ETR101*) reportedly regulated by SRF for which we did not identify a SRE. This was because the SRE for *CYR61* occurs immediately outside the 1-kb window used for our analysis, while the SRE for *JUNB* is downstream of the gene (37, 38). The SRE in the third gene, *ETR101*, was previously described in the mouse ortholog, *pip92* (39); this site also occurs outside the 1-kb analysis window in both the mouse and human sequences.

Sites identified by the SRF matrix were further evaluated by comparison with orthologous mouse sequences. These sequences were available for seven of the ten SRE-containing

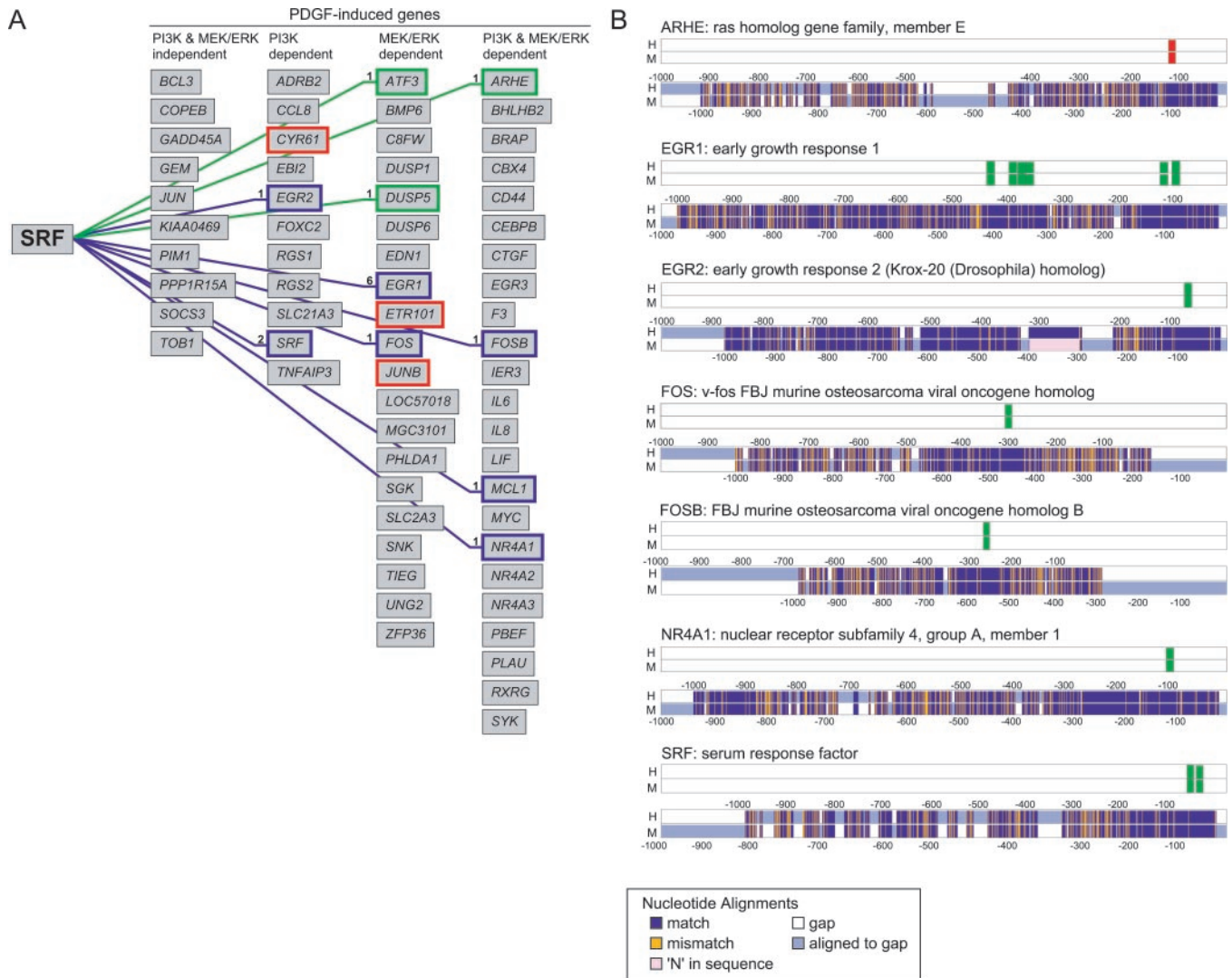


FIG. 4. Analysis of SRF target genes. A, diagram of SRF targets identified by the TRANSFAC matrix, V\$SRF_C. All of the PDGF-induced genes for which upstream sequences could be obtained are shown. Genes connected to SRF with green or purple lines have computationally identified SREs, with the number of SREs upstream of each gene indicated. Genes that have been previously shown to have the SRE are colored blue (53–57). Three genes previously shown to be regulated by SRF but in which we did not identify an SRE are highlighted in red. Each gene is represented by its LocusLink symbol. B, global alignments of upstream sequences for five human (H) and mouse (M) ortholog pairs illustrating conservation of computationally identified SRF binding sites. The SREs identified in both the human and mouse upstream sequences are shown above a graphical depiction of the alignments numbered relative to the transcription start sites. The sites colored green represent SREs that have been previously described in human (53–57) and in mouse (58–63).

human genes. Although these aligned sequences had low overall percent identities, 13 of the SREs were conserved: 8 were identical, 4 differed in a single matrix position, and one in two positions of low weight (Fig. 4B). In addition to 12 of 13 experimentally verified sites, we also identified an unreported, but conserved, SRE in the promoter of the RhoE (*ARHE*) gene.

In addition to these validations, SRF *cis*-element predictions were tested by chromatin immunoprecipitation to obtain direct experimental verification of the computational predictions within the cell system used. Chromatin from T98G cells was immunoprecipitated using an anti-SRF antibody, and quantitative PCR was used to detect enrichment for specific upstream regions (Fig. 5). *GAPDH*, a gene not regulated by SRF, was used as a negative control (28). In addition, four genes with no SRF binding sites detected were selected from the background set (not induced by PDGF) as predicted negatives.

The genes tested included those with *cis*-elements predicted by two SRF matrices, V\$SRF_C (shown in Fig. 4) as well as a second SRF matrix in the TRANSFAC data base, V\$SRF_Q6. It is noteworthy that V\$SRF_Q6 was a less stringent matrix, and

predicted SRF binding sites in the background set of promoter sequences at approximately a 5-fold greater frequency than V\$SRF_C (0.19 per kb for V\$SRF_Q6 compared with 0.04 per kb for V\$SRF_C; see Supplementary Table II).

SRF binding to 8 of the 10 genes predicted by the V\$SRF_C matrix was confirmed by the chromatin immunoprecipitation assays (Fig. 5). The promoter regions of each of these genes (*EGR1*, *EGR2*, *FOSB*, *FOS*, *MCL1*, *SRF*, *NR4A1*, and *DUSP5*) were significantly enriched (10- to >100-fold) in chromatin immunoprecipitates with anti-SRF antibody in comparison to *GAPDH*. As expected, the highest fold enrichment was obtained with *EGR1*, which contains 6 SRF binding sites. In contrast, the 4 predicted negative genes from the background set did not show any significant enrichment over *GAPDH* in anti-SRF chromatin immunoprecipitates. The genes for which SRF binding sites were demonstrated by this analysis in T98G cells included all 7 genes in which SRF binding sites had been previously observed in other systems (*EGR1*, *EGR2*, *FOSB*, *FOS*, *MCL1*, *SRF*, and *NR4A1*) as well as *DUSP5*, in which SRF binding had not been previously described. Despite the

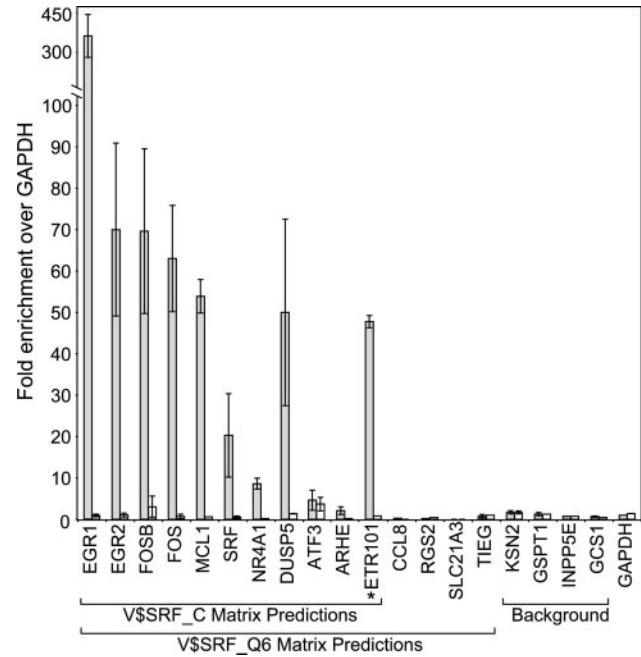


FIG. 5. Analysis of SRF binding sites by chromatin immunoprecipitation. Chromatin fragments of T98G cells were immunoprecipitated with anti-SRF antibody. Immunoprecipitation of each promoter region was quantitated by real time PCR and normalized to input. Each gene is represented by its LocusLink symbol, and data are presented for each gene as a fold increase over *GAPDH* in the presence (gray bars) or absence (white bars) of anti-SRF antibody. Data are the mean of at least two independent determinations \pm S.E. Genes are indicated as having SRF binding sites predicted by the V\$SRF_C or V\$SRF_Q6 matrices, or as having been derived from the background set of promoter sequences. *, the V\$SRF_Q6 matrix predicted a site within the 1-kb analysis window of *ETR101*. However, the mouse ortholog of this gene contains a previously identified SRF binding site immediately upstream of this 1 kb region. This site was detected in the human sequence with both the V\$SRF_C and V\$SRF_Q6 matrices. Chromatin immunoprecipitation of *ETR101* could therefore reflect SRF binding to either site.

prediction of a conserved SRE in *ARHE*, we were unable to confirm this site experimentally.

The less stringent V\$SRF_Q6 matrix detected all of the sites predicted by V\$SRF_C, as well as additional sites in *ETR101*, *CCL8*, *RGS2*, *SLC21A3*, and *TIEG*. In contrast to the sites predicted by V\$SRF_C, none of the additional sites predicted by V\$SRF_Q6 demonstrated enrichment in chromatin immunoprecipitations (Fig. 5). Although *ETR101* was clearly enriched in anti-SRF chromatin immunoprecipitates, these experiments cannot distinguish between SRF occupancy at the position computationally predicted by V\$SRF_Q6 (–884) and the previously demonstrated site in the mouse ortholog outside of the 1 kb window (–1188), which is recognized by V\$SRF_C. Because of the proximity of these sites, we think it is more likely that the positive chromatin immunoprecipitations reflect binding to the –1188 site, rather than to the –884 site predicted by V\$SRF_Q6. It thus appears that the V\$SRF_Q6 matrix predicted a higher number of false positive binding sites than V\$SRF_C, consistent with the higher frequency of V\$SRF_Q6 sites in the background set of promoters.

Networks of Regulated Gene Expression—We next sought to integrate the experimental data and our computational predictions into a transcriptional regulatory network. To generate this network, we combined the computational results from TRANSFAC matrices that were redundant or represented sites for families of related transcription factors. Thus, the 40 significant binding sites matrices identified in Fig. 3 corresponded to 18 unique transcription factors or families (Table II). For

TABLE II
Transcription factors with over-represented binding sites in the upstream sequences of PDGF-induced genes

Summary of transcription factors with statistically over-represented ($p \leq 0.01$) binding sites upstream of each group of co-expressed genes as assessed by the Student's *t*-test and the corresponding *p* values from the permutation test. Related transcription factors with similar binding sites are presented as a single family (for example, ATF/CREB). The component matrices represented by each factor can be found in Supplementary Table III. Transcription factors with binding sites limited to one group of co-expressed genes are indicated in bold.

| Transcription factor | Student's <i>t</i> test <i>p</i> value | Permutation <i>p</i> value |
|-------------------------------|---|-------------------------------|
| PI3K- and MEK/ERK-independent | | |
| ATF/CREB | <0.001 | 0.006 |
| CDP/Cut | 0.010 | 0.070 |
| NF-1/Myogenin | 0.001 | 0.013 |
| OCT7 | 0.001 | 0.094 |
| STAT1/5 | 0.001 | 0.010 |
| PI3K-dependent | | |
| ATF/CREB | <0.001 | 0.002 |
| MEF2 | 0.001 | 0.006 |
| NFκB | 0.009 | 0.048 |
| SRF | <0.001 | 0.002 |
| C/EBPα | 0.003 | 0.008 |
| NFIL3 | 0.001 | 0.005 |
| EVI | <0.001 | 0.024 |
| Forkhead | 0.001 | 0.010 |
| NGFI-C | <0.001 | 0.005 |
| NKX2-5 | 0.004 | 0.013 |
| OCT1/2 | <0.001 | 0.006 |
| ROAZ | 0.002 | 0.106 |
| MEK/ERK-dependent | | |
| ATF/CREB | <0.001 | 0.003 |
| NFκB | 0.002 | 0.009 |
| SRF | <0.001 | 0.005 |
| PI3K- and MEK/ERK-dependent | | |
| ATF/CREB | <0.001 | 0.001 |
| MEF2 | <0.001 | 0.004 |
| SRF | 0.005 | 0.009 |
| PBX1 | <0.001 | 0.019 |
| RORα2 | 0.001 | 0.056 |

each of these factors, Table II indicates the *p* value of the most significant matrix as determined by both the Student's *t* test and the permutation test. 14 of 18 factors identified as highly significant by the *t* test were also scored as significant ($p < 0.05$) by the permutation test. However, 4 factors (CDP/Cut, OCT7, ROAZ, and RORα2) identified as significant by the Student's *t* test were not statistically significant by the permutation test. As discussed further below, it is noteworthy that the binding sites predicted for these factors were identified in only 1 or 2 target genes and were not supported by experimental evidence, suggesting that they may represent false positives in the Student's *t* test.

The network of genes regulated by all 18 factors is presented in Fig. 6. All genes identified as having binding sites predicted by any of the TRANSFAC matrices for these factors are included, although (as discussed above for V\$SRF_Q6) some are expected to represent false positives corresponding to the frequency of sites predicted by each matrix in the background set of promoter sequences (see Supplementary Table II). In addition to SRF, predicted binding sites for STAT5, NF-κB, and ATF/CREB have been demonstrated experimentally (orange lines). At an additional level of confirmation, orthologous mouse sequences were obtained and aligned with 45 of the 64 human promoter regions (Supplementary Table IV). Within these regions, 50% of the predicted human binding sites were conserved in the mouse (green lines). For example, 36 ATF/CREB sites were detected in 23 human sequences for which a mouse ortholog was available. Twenty-three of these sites were conserved, 6 of which have been experimentally verified, supporting the role of ATF/CREB as a regulator of these genes.

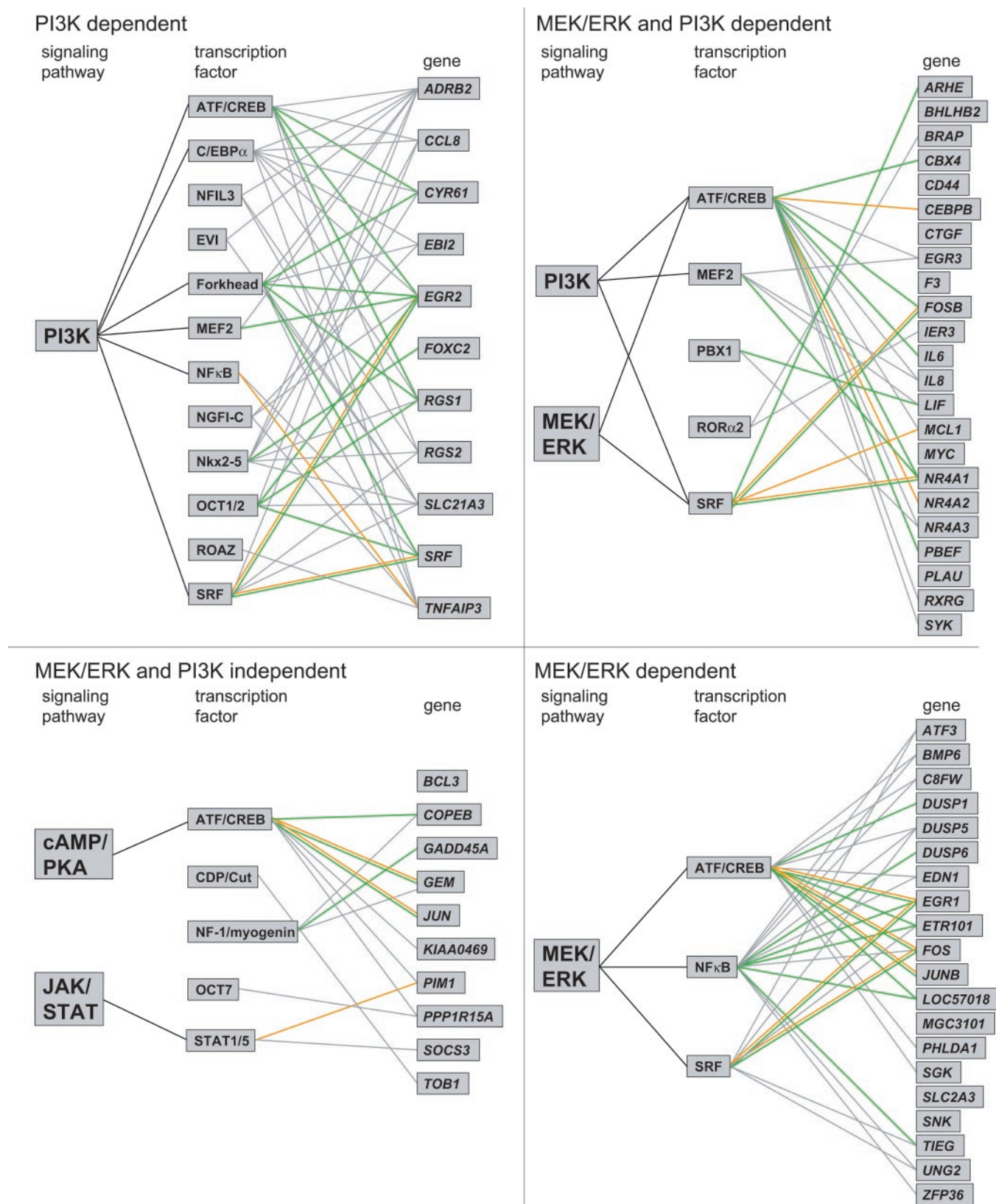


FIG. 6. Network diagram illustrating relationships between transcription factors and co-expressed genes. Genes connected to transcription factors by gray, orange, or green lines have computationally identified binding sites. Orange lines indicate genes for which the identified binding site has been previously reported. Green lines indicate genes with a least one conserved site detected in the aligned orthologous mouse sequence (54% of these sites were identical, and 20 of the remaining 28 differed in only a single position); see Supplementary Table IV for complete alignment information). Black lines indicate previously reported regulation of a transcription factor as a result of post-translational modification by the corresponding signaling pathway. References for PI3K pathway-transcription factor connections: (36, 40, 41, 43, 44, 46). References for MEK/ERK pathway-transcription factor connections: (36, 42, 46). Reference for JAK/STAT pathway-transcription factor connection: (45). Reference for cAMP/PKA pathway-transcription factor connections: (46). Reference for NF κ B-*TNFAIP3* connection: (64). Reference for STAT-*PIM1* connection: (65). ATF/CREB gene connections: (66–72). References for SRF gene connections are listed in the legend of Fig. 4.

Several of the predicted transcription factors are known targets of relevant signaling pathways (*black lines*). Binding sites for Forkhead (FOX) family members were over-represented among PI3K-dependent genes, consistent with regulation of Forkhead family members by PI3K/Akt signaling (40). NF- κ B binding sites were over-represented in PI3K- and MEK/ERK-dependent clusters, consistent with its known regulation (41, 42). Mef-2 had predicted binding sites in the PI3K-dependent as well as the MEK/ERK and PI3K-dependent clusters. This result is consistent with its known regulation by PI3K signaling (43). Binding sites for C/EBP α were also over-represented within the PI3K-dependent group of genes, consistent with regulation of C/EBP α by GSK-3 β downstream of PI3K/Akt (44). Likewise, binding sites for STATs, which are directly activated by receptor-associated kinases (45), were over-represented in the PI3K- and MEK/ERK-independent genes. Other factors, including SRF, were over-represented in multiple groups. For example, binding sites for ATF/CREB were over-represented in all 4 groups of genes, consistent with activation of CREB by cAMP/PKA signaling, as well as by PI3K/Akt and MEK/ERK/Rsk-2 (46). Overall, the regulation of 7 of the 18 predicted transcription factors was consistent with previous experimental data.

In combination, the conservation of predicted human regulatory elements in orthologous mouse genes and previous experimental verification of either predicted transcription factor binding sites or their cognate transcription factors provided validation for 11 of the 18 transcription factors that were predicted by our analysis (ATF/CREB, NF-1/myogenin, STAT1/5, MEF2, NF κ B, SRF, C/EBP α , Forkhead, Nkx2-5, OCT1/2, and PBX1). Predicted binding sites for most of these factors were identified in upstream sequences of multiple genes in each co-expressed group (Fig. 6), consistent with the hypothesis that common transcription factor binding sites would be shared among co-expressed immediate early genes. Of the 18 unique predictions, 14 were confirmed by the permutation test (Table II). It is noteworthy that the 4 factors not confirmed by the permutation test (CDP/Cut, OCT7, ROAZ, and ROR α 2) were also not validated by either experimental data or conservation in the mouse. Moreover, binding sites for 3 of these factors (OCT7, CDP/Cut, and ROAZ) were predicted in only a single gene and binding sites for ROR α 2 in only two genes. These factors may thus represent false positives, in contrast to the physiologically significant factors that have predicted binding sites in a number of co-expressed genes.

The agreement of many of our predictions with previous experimental data, the conservation of predicted sites in the mouse, and the direct validation of SRF binding sites by chromatin immunoprecipitation demonstrates the presence of common *cis*-regulatory elements in groups of co-expressed human genes. A critical element of this analysis was the experimental grouping of genes based on their regulation by specific signaling pathways that directly target transcription factors. By focusing on the specific induction of immediate early genes, we were able to establish a direct relationship between groups of genes and their transcriptional regulators. This allowed statistical analysis of the frequencies of regulatory elements in groups of co-expressed genes, addressing the problem of frequently occurring sequences that resemble transcription factor binding sites in genomic DNA. The accuracy of the identification of transcription factor binding sites in groups of co-expressed genes is coupled to both the stringency of the statistical analysis and the results of phylogenetic footprinting. Although we expect false positives in the *cis*-elements identified in individual genes, corresponding to the background associated with each matrix, the high frequencies of particular transcription factor binding sites in the co-expressed gene

groups substantiates these factors as likely targets of the relevant signaling pathways. Additional computational improvements would be expected to further enhance the power of this approach. Such improvements might include the development of better-defined matrices for identification of transcription factor binding sites, as indicated by the false positives revealed by the experimental validations of the V\$SRF_C and V\$SRF_Q6 predictions, as well as analysis of clustered transcription factor binding sites (5–7, 47–49) and phylogenetic footprinting with multiple organisms (50–52).

Acknowledgments—We thank Ulla Hansen, Zhiping Weng, and Stan Letovsky for helpful discussions and critical review of the article.

REFERENCES

- Pennacchio, L. A., and Rubin, E. M. (2001) *Nat. Rev. Genet.* **2**, 100–109
- Wyrick, J. J., and Young, R. A. (2002) *Curr. Opin. Genet. Dev.* **12**, 130–136
- Fickett, J. W., and Wasserman, W. W. (2000) *Curr. Opin. Biotechnol.* **11**, 19–24
- Ohler, U., and Niemann, H. (2001) *Trends. Genet.* **17**, 56–60
- Halfon, M. S., Grad, Y., Church, G. M., and Michelson, A. M. (2002) *Genome Res.* **12**, 1019–1028
- Markstein, M., Markstein, P., Markstein, V., and Levine, M. S. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99**, 763–768
- Berman, B. P., Nibu, Y., Pfeiffer, B. D., Tomancak, P., Celniker, S. E., Levine, M., Rubin, G. M., and Eisen, M. B. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99**, 757–762
- Hughes, J. D., Estep, P. W., Tavazoie, S., and Church, G. M. (2000) *J. Mol. Biol.* **296**, 1205–1214
- Chu, S., DeRisi, J., Eisen, M., Mulholland, J., Botstein, D., Brown, P. O., and Herskowitz, I. (1998) *Science* **282**, 699–705
- DeRisi, J. L., Iyer, V. R., and Brown, P. O. (1997) *Science* **278**, 680–686
- Roth, F. P., Hughes, J. D., Estep, P. W., and Church, G. M. (1998) *Nat. Biotechnol.* **16**, 939–945
- Tavazoie, S., Hughes, J. D., Campbell, M. J., Cho, R. J., and Church, G. M. (1999) *Nat. Genet.* **22**, 281–285
- Wolfsberg, T. G., Gabrielian, A. E., Campbell, M. J., Cho, R. J., Spouge, J. L., and Landsman, D. (1999) *Genome Res.* **9**, 775–792
- Elkon, R., Linhart, C., Sharan, R., Shamir, R., and Shiloh, Y. (2003) *Genome Res.* **13**, 773–780
- Cho, R. J., Campbell, M. J., Winzler, E. A., Steinmetz, L., Conway, A., Wodicka, L., Wolfsberg, T. G., Gabrielian, A. E., Landsman, D., Lockhart, D. J., and Davis, R. W. (1998) *Mol. Cell* **2**, 65–73
- Spellman, P. T., Sherlock, G., Zhang, M. Q., Iyer, V. R., Anders, K., Eisen, M. B., Brown, P. O., Botstein, D., and Futcher, B. (1998) *Mol. Biol. Cell* **9**, 3273–3297
- Ueda, H. R., Chen, W., Adachi, A., Wakamatsu, H., Hayashi, S., Takasugi, T., Nagano, M., Nakahama, K., Suzuki, Y., Sugano, S., Iino, M., Shigeyoshi, Y., and Hashimoto, S. (2002) *Nature* **418**, 534–539
- Brivanlou, A. H., and Darnell, J. E., Jr. (2002) *Science* **295**, 813–818
- Herschman, H. R. (1991) *Annu. Rev. Biochem.* **60**, 281–319
- Chang, L., and Karin, M. (2001) *Nature* **410**, 37–40
- Datta, S. R., Brunet, A., and Greenberg, M. E. (1999) *Genes Dev.* **13**, 2905–2927
- Tseng, G. C., Oh, M. K., Rohlin, L., Liao, J. C., and Wong, W. H. (2001) *Nucleic Acids Res.* **29**, 2549–2557
- Suzuki, Y., Yamashita, R., Nakai, K., and Sugano, S. (2002) *Nucleic Acids Res.* **30**, 328–331
- Wingender, E., Chen, X., Fricke, E., Geffers, R., Hehl, R., Liebich, I., Krull, M., Matys, V., Michael, H., Ohnhauser, R., Pruss, M., Schacherer, F., Thiele, S., and Urbach, S. (2001) *Nucleic Acids Res.* **29**, 281–283
- Ewens, W. J., and Grant, G. R. (2001) *Statistical Methods in Bioinformatics: An Introduction*, pp. 119–121, Springer-Verlag, New York, NY
- Rice, P., Longden, I., and Bleasby, A. (2000) *Trends Genet.* **16**, 276–277
- Luo, R. X., Postigo, A. A., and Dean, D. C. (1998) *Cell* **92**, 463–473
- Miralles, F., Posern, G., Zaromytidou, A. I., and Treisman, R. (2003) *Cell* **113**, 329–342
- Takahashi, Y., Rayman, J. B., and Dynlacht, B. D. (2000) *Genes Dev.* **14**, 804–816
- Stein, G. H. (1979) *J. Cell. Physiol.* **99**, 43–54
- Fambrough, D., McClure, K., Kazlauskas, A., and Lander, E. S. (1999) *Cell* **97**, 727–741
- Iyer, V. R., Eisen, M. B., Ross, D. T., Schuler, G., Moore, T., Lee, J. C., Trent, J. M., Staudt, L. M., Hudson, J., Jr., Boguski, M. S., Lashkari, D., Shalon, D., Botstein, D., and Brown, P. O. (1999) *Science* **283**, 83–87
- Vlahos, C. J., Matter, W. F., Hui, K. Y., and Brown, R. F. (1994) *J. Biol. Chem.* **269**, 5241–5248
- Favata, M. F., Horiuchi, K. Y., Manos, E. J., Daulerio, A. J., Stradley, D. A., Feeser, W. S., Van Dyk, D. E., Pitts, W. J., Earl, R. A., Hobbs, F., Copeland, R. A., Magolda, R. L., Scherle, P. A., and Trzaskos, J. M. (1998) *J. Biol. Chem.* **273**, 18623–18632
- Wasserman, W. W., Palumbo, M., Thompson, W., Fickett, J. W., and Lawrence, C. E. (2000) *Nat. Genet.* **26**, 225–228
- Gineitis, D., and Treisman, R. (2001) *J. Biol. Chem.* **276**, 24531–24539
- Latinkic, B. V., Mo, F. E., Greenspan, J. A., Copeland, N. G., Gilbert, D. J., Jenkins, N. A., Ross, S. R., and Lau, L. F. (2001) *Endocrinology* **142**, 2549–2557
- Perez-Albuerna, E. D., Schattelman, G., Sanders, L. K., and Nathans, D. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90**, 11960–11964

39. Latinkic, B. V., and Lau, L. F. (1994) *J. Biol. Chem.* **269**, 23163–23170
40. Tran, H., Brunet, A., Griffith, E. C., and Greenberg, M. E. (2003) *Sci. STKE* **2003**, RE5
41. Sizemore, N., Leung, S., and Stark, G. R. (1999) *Mol. Cell. Biol.* **19**, 4798–4805
42. Zhao, Q., and Lee, F. S. (1999) *J. Biol. Chem.* **274**, 8355–8358
43. Tamir, Y., and Bengal, E. (2000) *J. Biol. Chem.* **275**, 34424–34432
44. Ross, S. E., Erickson, R. L., Hemati, N., and MacDougald, O. A. (1999) *Mol. Cell. Biol.* **19**, 8433–8441
45. Aaronson, D. S., and Horvath, C. M. (2002) *Science* **296**, 1653–1655
46. Mayr, B., and Montminy, M. (2001) *Nat. Rev. Mol. Cell. Biol.* **2**, 599–609
47. Pilpel, Y., Sudarsanam, P., and Church, G. M. (2001) *Nat. Genet.* **29**, 153–159
48. Rebeiz, M., Reeves, N. L., and Posakony, J. W. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99**, 9888–9893
49. Frith, M. C., Spouge, J. L., Hansen, U., and Weng, Z. (2002) *Nucleic Acids Res.* **30**, 3214–3224
50. Boffelli, D., McAuliffe, J., Ovcharenko, D., Lewis, K. D., Ovcharenko, I., Pachter, L., and Rubin, E. M. (2003) *Science* **299**, 1391–1394
51. Dubchak, I., Brudno, M., Loots, G. G., Pachter, L., Mayor, C., Rubin, E. M., and Frazer, K. A. (2000) *Genome Res.* **10**, 1304–1306
52. Blanchette, M., Schwikowski, B., and Tompa, M. (2002) *J. Comput. Biol.* **9**, 211–223
53. Hipkind, R. A., and Nordheim, A. (1991) *J. Biol. Chem.* **266**, 19583–19592
54. Liu, X., Chen, X., Zachar, V., Chang, C., and Ebbesen, P. (1999) *J. Gen. Virol.* **80**, 3073–3081
55. Rangnekar, V. M., Aplin, A. C., and Sukhatme, V. P. (1990) *Nucleic Acids Res.* **18**, 2749–2757
56. Thiel, G., and Cibelli, G. (2002) *J. Cell. Physiol.* **193**, 287–292
57. Wu, S. Q., Minami, T., Donovan, D. J., and Aird, W. C. (2002) *Blood* **100**, 4454–4461
58. Chavrier, P., Janssen-Timmen, U., Mattei, M. G., Zerial, M., Bravo, R., and Charnay, P. (1989) *Mol. Cell. Biol.* **9**, 787–797
59. Lucibello, F. C., Lowag, C., Neuberg, M., and Muller, R. (1989) *Cell* **59**, 999–1007
60. Williams, G. T., and Lau, L. F. (1993) *Mol. Cell. Biol.* **13**, 6124–6136
61. Xi, H., and Kersh, G. J. (2003) *J. Immunol.* **170**, 315–324
62. Lazo, P. S., Dorfman, K., Noguchi, T., Mattei, M. G., and Bravo, R. (1992) *Nucleic Acids Res.* **20**, 343–350
63. Spencer, J. A., and Misra, R. P. (1996) *J. Biol. Chem.* **271**, 16535–16543
64. Krikos, A., Laherty, C. D., and Dixit, V. M. (1992) *J. Biol. Chem.* **267**, 17971–17976
65. Morcinek, J. C., Weisser, C., Geissinger, E., Scharl, M., and Wellbrock, C. (2002) *Oncogene* **21**, 1668–1678
66. Nakajima, K., Kusafuka, T., Takeda, T., Fujitani, Y., Nakae, K., and Hirano, T. (1993) *Mol. Cell. Biol.* **13**, 3027–3041
67. Niehof, M., Manns, M. P., and Trautwein, C. (1997) *Mol. Cell. Biol.* **17**, 3600–3613
68. Rolli, M., Kotlyarov, A., Sakamoto, K. M., Gaestel, M., and Neining, A. (1999) *J. Biol. Chem.* **274**, 19559–19564
69. Vanhoutte, P., Barnier, J. V., Guibert, B., Pages, C., Besson, M. J., Hipkind, R. A., and Caboche, J. (1999) *Mol. Cell. Biol.* **19**, 136–146
70. Santoro, T., Maguire, J., McBride, O. W., Avraham, K. B., Copeland, N. G., Jenkins, N. A., and Kelly, K. (1995) *Genomics* **30**, 558–564
71. Herdegen, T., and Leah, J. D. (1998) *Brain Res. Brain Res. Rev.* **28**, 370–490
72. McEvoy, A. N., Murphy, E. A., Ponnio, T., Conneely, O. M., Bresnahan, B., Fitzgerald, O., and Murphy, E. P. (2002) *J. Immunol.* **168**, 2979–2987