

Hierarchical model of gene regulation by transforming growth factor β

Yaw-Ching Yang^{*†}, Ester Piek^{*§}, Jiri Zavadil^{*}, Dan Liang^{*¶}, Donglu Xie^{*}, Joerg Heyer^{||**}, Paul Pavlidis^{††}, Raju Kucherlapati^{||‡‡}, Anita B. Roberts[‡], and Erwin P. Böttinger^{*||§§}

^{*}Departments of Medicine and ^{||}Molecular Genetics, Albert Einstein College of Medicine, Bronx, NY 10461; [†]Laboratory of Cell Regulation and Carcinogenesis, National Cancer Institute, Bethesda, MD 20892; and ^{††}Columbia Genome Center, Columbia University, New York, NY 10032

Communicated by Patricia K. Donahoe, Massachusetts General Hospital, Boston, MA, July 1, 2003 (received for review March 5, 2003)

Transforming growth factor β s (TGF- β s) regulate key aspects of embryonic development and major human diseases. Although Smad2, Smad3, and extracellular signal-regulated kinase (ERK) mitogen-activated protein kinases (MAPKs) have been proposed as key mediators in TGF- β signaling, their functional specificities and interactivity in controlling transcriptional programs in different cell types and (patho)physiological contexts are not known. We investigated expression profiles of genes controlled by TGF- β in fibroblasts with ablations of Smad2, Smad3, and ERK MAPK. Our results suggest that Smad3 is the essential mediator of TGF- β signaling and directly activates genes encoding regulators of transcription and signal transducers through Smad3/Smad4 DNA-binding motif repeats that are characteristic for immediate-early target genes of TGF- β but absent in intermediate target genes. In contrast, Smad2 and ERK predominantly transmodulated regulation of both immediate-early and intermediate genes by TGF- β /Smad3. These results suggest a previously uncharacterized hierarchical model of gene regulation by TGF- β in which TGF- β causes direct activation by Smad3 of cascades of regulators of transcription and signaling that are transmodulated by Smad2 and/or ERK.

The transforming growth factor β (TGF- β) family of secreted signaling proteins is highly conserved among eukaryotic organisms and consists of regulators of cell fates in developmental and homeostatic processes including developmental tissue remodeling, histogenesis, and maintenance of epithelial homeostasis. Mutations and epigenetic dysregulation of TGF- β signaling mechanisms are common in major human diseases including cancer progression, immune, fibrotic, and vascular diseases (1–3). Consistent with their broad significance in biology, TGF- β signals control a wide range of cellular functions that depend on cell type and (patho)physiological context. In most epithelial cells, TGF- β may exert several functions including inhibition of cell growth and initiation of apoptosis or epithelial-to-mesenchymal transitions. In contrast, the effects of TGF- β on cell growth and/or apoptosis in stromal fibroblasts are minor compared with its potent stimulation of cell-matrix adhesion and matrix remodeling and promotion of cell motility. Thus, elucidation of cell type- and context-dependent molecular signaling mechanisms that control the variations in functional specificity of TGF- β signaling is of considerable importance to understanding key developmental and disease processes.

A large number of studies have established that TGF- β binding causes receptor serine/threonine kinases of the TGF- β receptor subfamily to phosphorylate and activate receptor-regulated Smads (R-Smads), Smad2 and Smad3, and/or initiate non-Smad signaling through activation of mitogen-activated protein kinases (MAPKs), phosphatidylinositol 3-kinase, and other mediators (2). Activated R-Smads heterooligomerize with the common partner (CO)-Smad4 before translocation to the nucleus, where they regulate gene expression. R-Smads and CO-Smads contain highly conserved Mad homology 1 (MH1) and MH2 domains, connected by a linker region. The MH1 domain of Smad3 mediates direct interaction of Smad3 with conserved DNA Smad-binding elements (SBEs), whereas an

extra exon encoding an additional 30 amino acids in the MH1 domain of Smad2 prevents its direct binding to DNA (4, 5). Studies of *JunB* and *Smad7* gene promoters indicate that Smad3 but not Smad2 may have a direct functional role in their inducibility by TGF- β (6–8). Using Smad2- and Smad3-deficient fibroblasts, we reported previously that activation of the SBE4-Lux reporter by TGF- β required Smad3 but not Smad2, whereas activation of the activin-response element Lux reporter required Smad2 and was enhanced in Smad3-deficient cells (9). Together these findings indicate specific roles for Smad2 and Smad3 in TGF- β signaling. On a genomic scale, however, the relative functional specificities and signaling hierarchies of Smad2 and Smad3 and their functional interactivity with non-Smad signaling via extracellular signal-regulated kinase (ERK) MAPK, remain largely undefined.

In this report we present a hierarchical model of TGF- β signaling in fibroblasts based on comprehensive transcriptional profiling and large-scale computational analysis of functions and regulatory motifs of TGF- β target genes. Our results suggest that Smad3 directly activates genes encoding regulators of transcription and signal transducers through Smad3/Smad4 DNA-binding motif repeats that are characteristic for immediate-early target genes of TGF- β but absent in intermediate target genes. In contrast, although Smad2 and ERK may be required for regulation of some genes, these mediators predominantly transmodulate immediate-early gene (IEG) and intermediate gene regulation by TGF- β /Smad3.

Materials and Methods

Cell Culture, RNA Extraction, and Immunoblotting. Generation and characterization of four fibroblast lines derived from Smad2-knockout (Smad2KO) and wild-type littermate (Smad2WT) mouse embryos (day 10.5) and Smad3-knockout (Smad3KO) and WT littermate (Smad3WT) mouse embryos (day 17) has been reported (9). Cell-culture methods, RNA extraction, and immunoblotting protocols are described in detail in *Supporting*

Abbreviations: TGF- β , transforming growth factor β ; MAPK, mitogen-activated protein kinase; SBE, Smad-binding element; ERK, extracellular signal-regulated kinase; KO, knockout; QC, quality control; S/N, signal-to-noise ratio; IEG, immediate-early gene; IIG, intermediate-induced gene; IRG, intermediate-repressed gene; MEK, MAPK kinase; 5'FS, 5'-flanking sequence.

[†]Present address: Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02453.

[§]Present address: Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX, Amsterdam, The Netherlands.

[¶]Present address: Department of Medicine, Vanderbilt University Medical Center, 529 Light Hall, Nashville, TN 37232.

^{**}Present address: GenPath Pharmaceuticals, Inc., 300 Technology Square, Cambridge, MA 02139.

^{††}Present address: Harvard Partners Center for Genetics and Genomics, 917 Thorn Building, 20 Shattuck Street, Boston, MA 02115.

^{§§}To whom correspondence should be addressed at: Departments of Medicine and Molecular Genetics, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461. E-mail: bottinger@aecom.yu.edu.

Materials and Methods, which is published as supporting information on the PNAS web site, www.pnas.org.

Microarray, Data Analysis, and Statistical Approaches. *Microarrays.* Mouse cDNA arrays (9M series) were obtained from the Albert Einstein College of Medicine cDNA Microarray Facility (www.aecom.yu.edu/home/molgen/facilities.html). Each slide contained an unbiased, random collection of 8,976 cDNA probe elements derived from the sequence-verified GEM1 clone set (Incyte Genomics, Palo Alto, CA). When analyzed by using the UniGene cluster database (July 2002), this probe set contains 7,183 unique mouse transcripts including 2,484 uniquely named mouse genes. Microarray procedures were performed as described (10) with minor modifications. For each fibroblast line, and U0126-treated WT cells, cDNA targets were prepared from total RNA samples (Cy5 fluorescent-labeled) obtained from TGF- β -stimulated cells. As reference RNA (Cy3 fluorescent-labeled), aliquots of time point zero (T_0) RNA samples obtained from Smad2WT and Smad2KO, Smad3WT and Smad3KO, and U0126-treated Smad2WT fibroblasts were used to provide separate baseline expression measurements for each genotype and inhibitor condition. Using a common baseline reference cDNA target allows comparisons of the relative expression of each gene across the entire series of timed TGF- β stimulations (0, 0.3, 1, 2, and 4 hr).

Quality control (QC) and data filtering (QC pass). A rigorous computational algorithm for spot quality and time profile QC was implemented based on (i) spot quality vote [signal-to-noise ratio (S/N): AVE CHI > AVE CHB + 2 SD = pass vote 0, AVE CHI \leq AVE CHB + 2 SD = fail vote 1]; (ii) reproducibility factor $R = \Sigma[N1(S/N \text{ vote}), N2(S/N \text{ vote}), N3(S/N \text{ vote})]$, $R \leq 1$ = pass vote 0, $R > 1$ = fail vote 1; and (iii) multimeasurement (time profile) QC final flag (FF) = $\Sigma[T_{0 \text{ hr}}(R), T_{0.3 \text{ hr}}(R), T_{1 \text{ hr}}(R), T_{2 \text{ hr}}(R), T_{4 \text{ hr}}(R)]$, $FF \leq 3$ = pass vote 0, $FF > 3$ = fail vote 1. Therefore, only Cy5/Cy3 intensity ratios for transcript profiles associated with FF vote 0 were accepted as quality-controlled (QC pass) and included in further analysis. Each experiment (microarray) was normalized by using a median-centering method based on \log_2 -transformed, QC-pass Cy5/Cy3 ratio values.

Criteria for identification of transcript profiles with "significant TGF- β effect." To filter TGF- β -responsive genes from all QC-pass gene expression profiles, statistical and threshold filters were applied to analyze significant changes in Cy5/Cy3 ratios between time points separately in each of four pathway mediator classes, represented by (i) WT fibroblasts [six replicates with combined Smad2WT ($n = 3$) and Smad3WT ($n = 3$) data]; (ii) U0126-pretreated Smad2WT fibroblasts (three replicates); (iii) Smad2KO (three replicates); and (iv) Smad3KO (three replicates). Groups of replicates of \log_2 -transformed ratios from each time point in a time series were compared for significant differences in mean ratios by using a t statistic with correction for false-discovery rate as implemented in "SIGNIFICANCE ANALYSIS OF MICROARRAY" (SAM) software (11). Statistical filtering identified transcript profiles that had significantly different ratio values at any time point. Next we inspected manually all replicates of each identified transcript profile to validate reproducibility. To enable direct comparisons of transcript profiles between different pathway conditions, median \log_2 -transformed ratios for each time point were normalized to baseline (T_0) median \log_2 ratios, and the normalized median \log_2 ratios were used as representative time-series patterns for all transcript profiles. Finally, we applied a threshold filter to all normalized, representative transcript profiles where median ratios at 1, 2, or 4 hr or any combination of these time points were required to be >1.4-fold different (up or down) compared with baseline (T_0). This threshold assured that significant transcript profiles included deviations of at least 2 SD (40%) of all averaged baseline

(T_0) median ratios. All data are deposited for public viewing, query, and download in the gene expression data repository Gene Expression Omnibus (GEO) at www.ncbi.nlm.nih.gov/geo under GEO accession GPL362.

Northern Blot Analysis. RNA was isolated from cells by column purification by using the RNeasy kit (Qiagen, Valencia, CA). For Northern blot analysis, RNA was electrophoresed on 1% agarose gels and transferred to a filter. Filters then were hybridized with ^{32}P -labeled cDNA probes for mouse Riken clone 1190017B18. The ribosomal 18S band was visualized by ethidium-bromide gel stain and used as an RNA loading control.

Promoter Analysis. Sequence information for 5'-flanking regions of mouse and human orthologs for genes with significant TGF- β effect and a random group of unregulated control genes was parsed to a local database from the Celera Genome Database by using Celera Discovery System (Celera, Rockville, MD). Genes selected for detailed promoter analysis had to pass the following filter criteria: (i) definitive mouse-human ortholog relationship, (ii) definitive transcription start site and exon 1 localization, and (iii) high-quality sequence data without genomic repeat sequences. Software was developed in-house to identify and parse information for transcription factor consensus binding motifs present in filtered 5'-flanking regions as defined in the TRANSFAC database and MATINSPECTOR PROFESSIONAL software (12, 13).

Gene Ontology. A custom program was developed to search AMIGO, LocusLink, and SwissProt databases for associations of Gene Ontology Consortium (14) terms in molecular function and biological process categories with named genes in lists of IEGs, intermediate-induced genes (IIG), and intermediate-repressed genes (IRGs).

Results and Discussion

Microarray Screens for TGF- β -Responsive Genes in Fibroblasts with Ablation of Smad2, Smad3, or MAPK ERK Reveal an Essential Role for Smad3 in TGF- β Signaling. We previously reported an experimental system consisting of stable fibroblast lines with genetic ablation of Smad2 (Smad2KO) or Smad3 (Smad3KO) (9). In addition, MEK/ERK activity was chemically ablated in WT control fibroblasts derived from WT littermate embryos by using the specific inhibitor U0126 (see *Supporting Materials and Methods*, and Fig. 4 which is published as supporting information on the PNAS web site). Experiments in WT cells pretreated with U0126 are referred to here as U0126. The impact of each ablation on TGF- β -dependent gene regulation in time-series experiments (0, 0.3, 1, 2, and 4 hr TGF- β 1) was assessed by using two-color cDNA microarray technology and significance analysis. Arrays contained an unbiased, random collection of 8,976 cDNA probe elements derived from EST clones that represent 7,183 unique mouse transcripts (UniGene clusters), indicating that a substantial portion of the mouse genome was assayed. Microarray procedures were performed on at least three independent repeats of each experimental condition as described (10) with minor modifications. This algorithm generates accurate, reproducible, and reliable gene expression data as previously verified by real-time PCR and tight-correlation coefficients of replicates (10). All data are deposited for public viewing, query, and download in the gene expression data repository Gene Expression Omnibus at www.ncbi.nlm.nih.gov/geo under GEO accession no. GPL362. A total of 360 gene expression profiles (referred to here as TGF- β target genes) with statistically significant and manually verified TGF- β effect were identified, representing 5% of the 7,183 assayed transcripts (the complete target-gene list with log-ratio data are available in Table 1, which is published as supporting information on the PNAS web site).

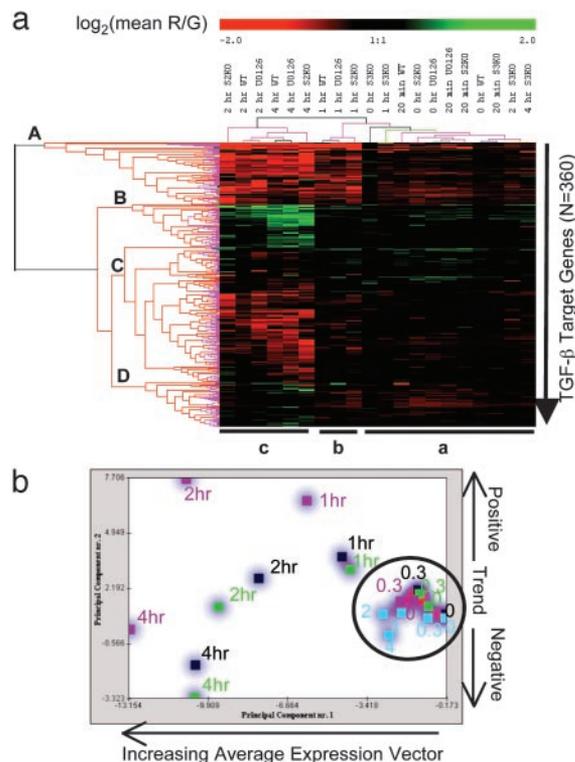


Fig. 1. Effects of ablation of Smad2, Smad3, and ERK function on gene regulation by TGF- β . (a) Resampled hierarchical clustering (bootstrapping) for 360 TGF- β target genes demonstrates four significant gene clusters (A–D) and three experimental variables clusters (a–c). Gene cluster A, immediate-early targets; gene cluster B, intermediate repressed targets; gene cluster C, intermediate-induced targets; gene cluster D, intermediate-repressed targets. Experimental variables cluster a, T_0 and $T_{0.3}$ of WT, Smad2KO (S2KO), Smad3KO (S3KO), and U0126 ablation together with T_1 , T_2 , and T_4 of S3KO; experimental variables cluster b, T_1 of WT, U0126, and S2KO; experimental variables cluster c, T_2 and T_4 of WT, U0126, and S2KO. (b) Principle component analysis (17) demonstrates the overall relatedness of expression profiles for 360 TGF- β target genes in each of the experimental variables in the four time series (WT, U0126, Smad2KO, and Smad3KO). Colored square dots and associated time-point values indicate WT (black), U0126 (green), Smad2KO (purple), and Smad3KO (blue) variables.

A recent study of extended TGF- β stimulation of human fetal lung fibroblasts up to 24 hr identified 146 genes with at least 2-fold induction (15). The number of genes with expression that was modulated significantly within 4 hr of TGF- β treatment was similar in WT ($n = 150$) and Smad2KO ($n = 161$; $P = 0.13$) but was increased in U0126 ($n = 238$; $P < 0.0001$) and dramatically decreased in Smad3KO ($n = 9$; $P < 0.0001$) compared with WT. Of 150 target genes in WT, 66 did not reach statistically significant levels of regulation in U0126 and/or Smad2KO. However, ERK inactivation and/or Smad2 deficiency were associated with partial reduction of TGF- β responsiveness of these genes compared with WT, whereas Smad3 deficiency completely blocked TGF- β response. Surprisingly, a considerable number of regulated transcripts were identified uniquely in U0126 ($n = 114$), Smad2KO ($n = 57$), or in both ($n = 39$) but not in WT.

Statistical variances and relatedness among the 20 experimental variables were determined by clustering the expression data for all 360 TGF- β target genes across the 20 experimental variables (four conditions \times five time points) by using two independent methods: hierarchical clustering with resampling (Fig. 1a) (16) and principal component analysis of time-series analysis (Fig. 1b) (17). The five Smad3KO experimental vari-

ables (T_0 , $T_{0.3}$, T_1 , T_2 , and T4S3KO) were clustered together with all baseline (T_0) and 20-min ($T_{0.3}$) variables for WT, U0126, and Smad2KO (Fig. 1a, cluster a, and 1b, circled), indicating that expression levels of 360 TGF- β target genes were statistically indistinguishable in TGF- β -treated Smad3KO relative to the general baseline gene expression levels of these genes. In contrast, the collective target-gene expression levels for T1WT, T2WT, T4WT, and corresponding U0126 and Smad2KO variables distinguished these variables from the baseline and Smad3KO cluster (Fig. 1). These results suggest that Smad2 and Smad3 exert fundamentally different functional roles in TGF- β signaling in fibroblasts.

Inactivation of Smad2 Is Primarily Associated with Hyperresponsiveness of Immediate-Early TGF- β Target Genes, Whereas Inactivation of MEK/ERK Primarily Increases the Overall Number of Intermediate TGF- β -Responsive Genes. We further refined our analysis by separating the 360 TGF- β target genes into kinetically distinct groups of IEGs ($n = 40$), IIGs ($n = 206$), and IRGs ($n = 112$) by using self-organizing maps (18) (Fig. 2a) and threshold criteria for single time points (see *Supporting Materials and Methods*). This analysis demonstrated that Smad2 ablation but not ERK inactivation was on average associated with hyperresponsiveness and/or extended activation of Smad3-dependent IEGs (Fig. 2a and b). In fact, some IEGs (*Smoc1* and *Tnfrsf1b*) were significantly activated only in Smad2KO cells (Table 1), and none of the IEGs (except for expressed sequence R75030) was significantly reduced in Smad2KO compared with WT fibroblasts. By Northern blot analysis we confirmed microarray-based expression profiles for select IEGs including Riken clone 1190017B18 (Fig. 2c), *Gadd45b*, and *Snail* (data not shown). As shown in our previous report for PAI-1 protein, and SBE- or 3TPlux reporter gene activities (9), reconstitution of Smad3 by adenoviral gene transfer in Smad3-deficient fibroblast lines (see Fig. 2c) and in primary dermal fibroblast cultures established from Smad3-deficient mice (data not shown) rescued immediate-early response profiles of these genes. ERK inactivation significantly inhibited TGF- β responsiveness of two IEGs (*Ier2* and *Bhlhb2*) and abolished induction of *Nr4a1*, indicating that ERK function is required for efficient activation of some IEGs by TGF- β . These findings suggest that Smad3 is required for transcriptional activation of IEGs by TGF- β in fibroblasts, whereas Smad2 may limit the magnitude and/or duration of their transient transcriptional activation. Smad2 is known to interact with the inducible transcriptional corepressors TG-interacting factor (TGIF) (19) and constitutive corepressors Ski and SnoN (20). Smad2 may also participate in Smad3-dependent transcriptional complexes (8, 21). Thus, inducible Smad2-dependent recruitment of corepressors to Smad3-dependent protein DNA-binding complexes may mediate a repressor signal that dynamically balances Smad3-dependent gene activation.

In contrast with IEGs, average expression profiles of IIGs and IRGs were similar in Smad2KO and U0126 compared with WT cells (Fig. 2a and Fig. 5, which is published as supporting information on the PNAS web site), although a limited fraction of IIGs (12%) that were regulated in WT did not respond to TGF- β in U0126. However, 66.5% (137) of IIGs and 41.9% (47) of IRGs were identified as target genes by our criteria only in U0126 and/or Smad2KO but not in WT. We conclude that baseline and/or inducible ERK signals enhance the magnitude of regulation of a limited group of IEGs and IIGs by TGF- β . In addition, an independent function of ERK may be to counteract the TGF- β responsiveness of a large group of genes to the extent that inactivation of ERK is required for these genes (IIGs and IRGs) to become responsive to TGF- β . These unexpected observations imply at least two distinct mechanisms of cross-talk between the TGF- β and Ras/Raf/MEK/ERK pathways. First, ERK signals may enhance Smad3-dependent activation of a

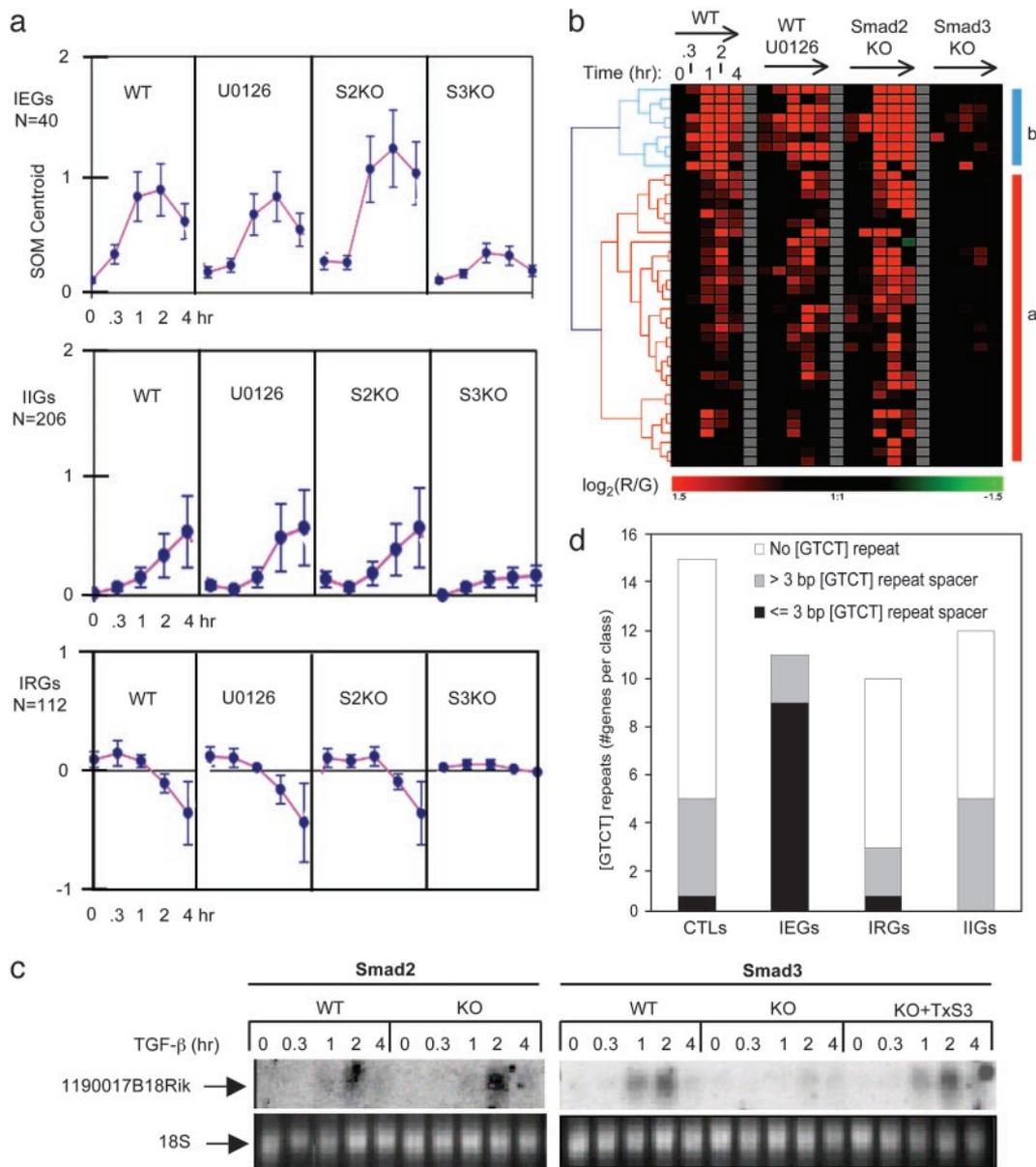


Fig. 2. Patterns of regulation and promoter elements of IEGs, IIGs, and IRGs. (a) The centroid (\pm SD) of expression profiles of IEGs, IIGs, and IRGs in WT, U0126-treated WT, Smad2KO (S2KO), and Smad3KO (S3KO) fibroblasts. A centroid of a cluster in self-organizing maps is a data point with coordinates that are the averages of the corresponding coordinates for median log ratios of gene expression for all gene expression profiles in a cluster. (b) Hierarchical clustering of IEG expression profiles in 20 experimental variables. (c) Northern blot demonstrates Riken clone 1190017B18 mRNA expression in Smad2KO and littermate-derived Smad2WT, Smad3KO and littermate-derived Smad3WT fibroblasts, and Smad3KO fibroblasts transduced with Smad3 adenovirus (KO+TxS3). Cells were treated with TGF- β as indicated. (d) Distribution of [GTCT] Smad3/Smad4 core binding-site repeats in unregulated control genes, IEGs, IRGs, and IIGs. Black box, number of genes containing [GTCT] core repeats with spacer lengths \leq 3 bp; gray box, number of genes harboring [GTCT] core repeats with spacer lengths >3 bp; white box, number of genes without [GTCT] core repeats in each group (the software and methods used are described in *Supporting Materials and Methods*).

limited number of IEGs by positive interactions with activated Smads (22, 23). Second, ERK signals may repress considerable subsets of intermediate TGF- β target genes through yet-unknown mechanisms.

Characteristic Smad3/4 Consensus Binding Motif Repeats Represent a Signature Regulatory Module of TGF- β /Smad3-Responsive IEGs. To understand the molecular basis for the kinetically defined and mediator-specific patterns of gene regulation that we observed, we searched for signature gene regulatory sequences in 5'-flanking sequences (5'FSs) of Smad3-dependent IEGs, IIGs, and IRGs and unregulated control genes. We analyzed 10 kb of 5'FS of all target genes for occurrences and locations of known

cis-acting elements defined in the TRANSFAC (13) database including putative 5'-GTCTG-3' Smad3/Smad4 consensus binding elements (SBEs) (24). We found no difference between groups in the occurrences of single SBE sites per gene. It has been proposed that Smad3 binds single 5'-GTCTG-3' sites with relatively low affinity, requiring direct interactions with other transcription factors for stable binding and effective transcriptional activation (25). We did not observe a statistically significant co-occurrence of SBEs and binding sites for unrelated known eukaryotic transcriptional regulators. However, inverted or direct repeats of the [GTCT] SBE core binding sequence may provide a higher-affinity binding matrix (26) that is sufficient for binding of inducible Smad3/Smad4 complexes independent of

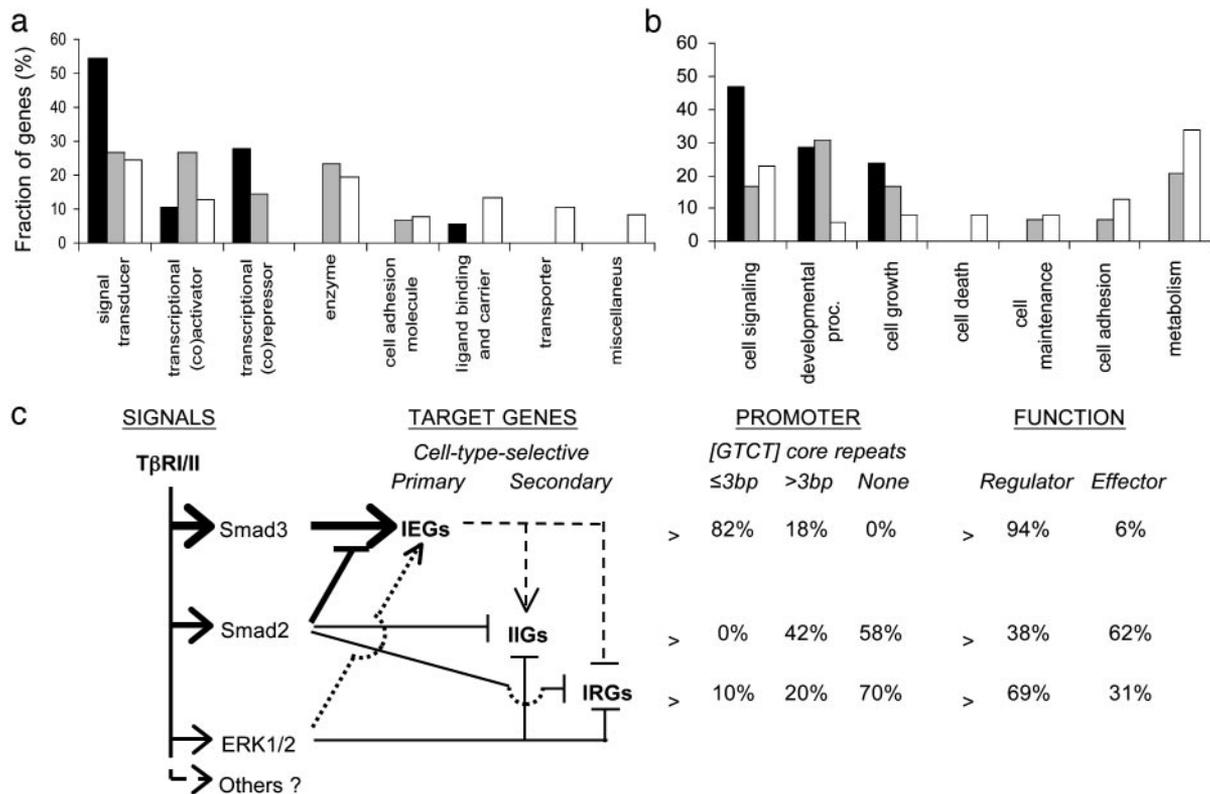


Fig. 3. Differential functional profiles of TGF- β target genes. Distribution (fraction of genes) of IEGs (black bars), IRGs (gray bars), and IIGs (white bars) in “molecular function” (a) and “biological process” (b) categories as defined by the Gene Ontology Consortium (14). (c) Hierarchical model of gene regulation after ligand-induced activation of the TGF- β receptor complex (T β RI/II) (see *Conclusions* for details).

unrelated factors (8) and for mediating Smad3/Smad4-dependent induction of the immediate-early *SMAD7* gene *in vivo* (7, 8, 27). When 5'FSs were searched for occurrences and locations of tandem or inverted SBE core repeats (see *Supporting Materials and Methods*), allowing for variable lengths of half-site spacing (0–20 bp), we observed that SBE core repeats with short spacer lengths (0–3 bp) were present specifically in proximal promoters (nucleotides –1,262 to +73) of Smad3-dependent IEGs (present in 82% of IEGs), compared with very rare occurrences in promoters of unregulated control genes, IRGs, and IIGs (Fig. 2c and Table 2, which is published as supporting information on the PNAS web site). Thus, our analysis of 5'FSs suggests a highly significant and specific “signature” SBE configuration for Smad3-dependent IEGs that consists of direct GTCT (≤ 3 bp) GTCT or inverted GTCT (≤ 3 bp) AGAC repeats of the general Smad3/Smad4 core binding sequence [GTCT] that are typically located in the proximal 1 kb of IEG 5'FSs (see Table 2). Because the IEG signature SBE core repeats occurred only at random frequency in IIGs and IRGs (Fig. 2c), these intermediate target genes may be regulated indirectly by Smad3-dependent activation of transcriptional regulators rather than directly by binding of Smad3-dependent protein complexes. However, it is still possible that Smad3/4 may bind intermediate target genes at single SBEs with considerably lower avidity as IEGs, therefore possibly requiring a longer duration of signal, or cooperation with other transcription factors.

Immediate-Early TGF- β -Responsive Genes Encode Signal Transducers and Transcriptional Regulators. If intermediate genes are not directly regulated by Smad3, one would expect an enrichment of transcriptional regulators among Smad3-dependent IEGs to control intermediate target genes. To explore this hypothesis, terms for “molecular function” and “biological process” from

the Gene Ontology Consortium database (14) were automatically linked with named target genes in each group (see data in Table 1 and *Supporting Materials and Methods*). Most IEGs (94%) encoded signal transducers (55%) and transcriptional regulators (39%) (Fig. 3a), referred to here as “regulators.” Surprisingly, none of the IEGs were repressed by TGF- β , and transcriptional (co)repressors outnumbered (co)activators three to one in this group, indicating that Smad3 itself may not have a significant role as a transcriptional repressor in fibroblasts. The majority of IRGs (69%) also encoded regulatory proteins, whereas the majority of IIGs (62%) encoded enzymes, cell adhesion molecules, ligand binding or carriers, transporters, and proteins of miscellaneous functions, referred to here as “effector” proteins (Fig. 3a). Consistent results were obtained when the target genes were linked with terms in the Gene Ontology Consortium category “biological process.” All named IEGs and 65% of IRGs encoded proteins mediating regulatory biological processes (cell signaling, developmental processes, and cell growth), whereas 63% of IIGs function in effector biological processes (cell maintenance, cell adhesion, cell death, and metabolism) (Fig. 3b). These results indicate that Smad3 may function primarily to activate directly a set of transcriptional regulators to initiate a cascade of secondary gene regulation and to induce expression of signal transducers that may mediate transduction of related signaling networks. Interestingly, in standardized comparisons (data not shown), we found little overlap between the IEGs in fibroblasts and a large set of TGF- β -regulated IEGs previously reported in human keratinocytes (HaCaT) (10). These preliminary observations may indicate that the complement of TGF- β -regulated IEGs may be highly cell-type-dependent (J.Z. and E.P.B., unpublished observations), consistent with well characterized differences in cellular responses controlled by TGF- β in fibroblasts and keratino-

cytes (28). Thus, specification of cellular and biological responses by TGF- β may be determined by cell-type-selective complements of directly activatable transcriptional regulators and signal transducers and their range of downstream mediators.

Conclusions

Our genome-level analysis has provided insights into the temporal regulation and integration of signaling pathways that could never have been seen at the level of single-gene analysis or conventional reporter gene-based approaches. Thus, we propose a hierarchical model of gene regulation by TGF- β (Fig. 3c), highlighting the complex interactions and distinctly different roles of Smad2 and Smad3 in direct target-gene regulation and a broad repressor effect of ERK signals predominantly on intermediate TGF- β targets. The data show that Smad3 is required for direct transcriptional activation of IEGs encoding regulator proteins including signal transducers and transcriptional regulators. In contrast, Smad2 may function as a negative modulator of Smad3 activity by yet-unknown mechanisms, limiting the extent and/or terminating the activation of most Smad3-dependent IEGs. Because secondary gene targets lack the high-affinity SBE promoter signature characteristic of IEGs, we propose that their induction or repression may be regulated either directly through lower-affinity Smad3/4-binding sites or indirectly by Smad3-dependent IEGs encoding transcriptional regulators. Although the majority of IEGs encode proteins with effector functions involved in metabolism, cell maintenance, and

cell adhesion, two thirds of IRGs encode other regulator proteins, indicating an additional layer of regulatory complexity of TGF- β -dependent transcriptional reprogramming. The unmasking of previously uncharacterized IEGs and IRGs by suppression of the ERK MAPK pathway and the demonstration of mediator-selective transcriptional targets of TGF- β provide paradigms for regulation of the transcriptome by signaling cross-talk. Thus, global cell-type- and context-selective mechanisms, for example MAPK activity in disease states (29–31), can permit or restrict regulation of distinct complements of TGF- β target genes. Such a hierarchical model of direct recruitment by TGF- β /Smad3 of multiple cell-type-selective and context-selective regulatory proteins that can regulate downstream effector proteins would provide a highly effective mechanism to enable context-dependent combinatorial signaling networks that may underlie the extraordinary multifunctionality characteristic of TGF- β s.

We are grateful to Dr. Michael Reiss (Robert Wood Johnson Medical School/University of Medicine and Dentistry of New Jersey, New Brunswick) for providing anti-phospho Smad3 antibody. We thank Mr. Aldo Massimi for assistance with microarrays and the Albert Einstein Biotechnology Center (through National Institute of Diabetes and Digestive and Kidney Diseases Grant U24 DK58768-A1) for technical support. We also thank Dr. Lalage Wakefield for critical reviews of this manuscript. J.Z. is Fellow of the American Association for Cancer Research–Sidney Kimmel Foundation for Cancer Research. This work was supported by National Institutes of Health Grants R01DK056077 and R01DK60043 (to E.P.B.).

1. Derynck, R., Akhurst, R. J. & Balmain, A. (2001) *Nat. Genet.* **29**, 117–129.
2. Roberts, A. B. (2002) *Cytokine Growth Factor Rev.* **13**, 3–5.
3. Massagué, J. (2000) *Nat. Rev. Mol. Cell Biol.* **1**, 169–178.
4. Dennler, S., Huet, S. & Gauthier, J. M. (1999) *Oncogene* **18**, 1643–1648.
5. Yagi, K., Goto, D., Hamamoto, T., Takenoshita, S., Kato, M. & Miyazono, K. (1999) *J. Biol. Chem.* **274**, 703–709.
6. Jonk, L. J., Itoh, S., Heldin, C. H., ten Dijke, P. & Kruijer, W. (1998) *J. Biol. Chem.* **273**, 21145–21152.
7. Nagarajan, R. P., Zhang, J., Li, W. & Chen, Y. (1999) *J. Biol. Chem.* **274**, 33412–33418.
8. von Gersdorff, G., Susztak, K., Rezvani, F., Bitzer, M., Liang, D. & Böttinger, E. P. (2000) *J. Biol. Chem.* **275**, 11320–11326.
9. Piek, E., Ju, W. J., Heyer, J., Escalante-Alcalde, D., Stewart, C. L., Weinstein, M., Deng, C., Kucherlapati, R., Böttinger, E. P. & Roberts, A. B. (2001) *J. Biol. Chem.* **276**, 19945–19953.
10. Zavadil, J., Bitzer, M., Liang, D., Yang, Y. C., Massimi, A., Kneitz, S., Piek, E. & Böttinger, E. P. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 6686–6691.
11. Tusher, V. G., Tibshirani, R. & Chu, G. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 5116–5121.
12. Quandt, K., Frech, K., Karas, H., Wingender, E. & Werner, T. (1995) *Nucleic Acids Res.* **23**, 4878–4884.
13. Wingender, E., Chen, X., Hehl, R., Karas, H., Liebich, I., Matys, V., Meinhardt, T., Pruss, M., Reuter, I. & Schacherer, F. (2000) *Nucleic Acids Res.* **28**, 316–319.
14. Gene Ontology Consortium (2001) *Genome Res.* **11**, 1425–1433.
15. Chambers, R. C., Leoni, P., Kaminski, N., Laurent, G. J. & Heller, R. A. (2003) *Am. J. Pathol.* **162**, 533–546.
16. Heyer, L. J., Kruglyak, S. & Yooshep, S. (1999) *Genome Res.* **9**, 1106–1115.
17. Raychaudhuri, S., Stuart, J. M. & Altman, R. B. (2000) *Pac. Symp. Biocomput.* **455–466**.
18. Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., Lander, E. S. & Golub, T. R. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 2907–2912.
19. Wotton, D., Lo, R. S., Swaby, L. A. & Massagué, J. (1999) *J. Biol. Chem.* **274**, 37105–37110.
20. Liu, X., Sun, Y., Weinberg, R. A. & Lodish, H. F. (2001) *Cytokine Growth Factor Rev.* **12**, 1–8.
21. Kawabata, M., Inoue, H., Hanyu, A., Imamura, T. & Miyazono, K. (1998) *EMBO J.* **17**, 4056–4065.
22. de Caestecker, M. P., Parks, T. W., Frank, C. J. T., Castagnino, P., Bottaro, D., Roberts, A. B. & Lechleider, R. J. (1998) *Genes Dev.* **12**, 1587–1592.
23. Zhang, Y., Feng, X. H. & Derynck, R. (1998) *Nature* **394**, 909–913.
24. Dennler, S., Itoh, S., Vivien, D., ten Dijke, P., Huet, S. & Gauthier, J. M. (1998) *EMBO J.* **17**, 3091–3100.
25. Moustakas, A., Souchelnytskyi, S. & Heldin, C. H. (2001) *J. Cell Sci.* **114**, 4359–4369.
26. Zawal, L., Dai, J. L., Buckhaults, P., Zhou, S., Kinzler, K. W., Vogelstein, B. & Kern, S. E. (1998) *Mol. Cell* **1**, 611–617.
27. Brodin, G., Ahgren, A., ten Dijke, P., Heldin, C. H. & Heuchel, R. (2000) *J. Biol. Chem.* **275**, 29023–29030.
28. Piek, E., Moustakas, A., Kurisaki, A., Heldin, C. H. & ten Dijke, P. (1999) *J. Cell Sci.* **112**, 4557–4568.
29. Lehmann, K., Janda, E., Pierreux, C. E., Rytomaa, M., Schulze, A., McMahon, M., Hill, C. S., Beug, H. & Downward, J. (2000) *Genes Dev.* **14**, 2610–2622.
30. Janda, E., Lehmann, K., Killisch, I., Jechlinger, M., Herzig, M., Downward, J., Beug, H. & Grunert, S. (2002) *J. Cell Biol.* **156**, 299–313.
31. Kretschmar, M., Doody, J., Timokhina, I. & Massagué, J. (1999) *Genes Dev.* **13**, 804–816.