AD_____

Award Number: DAMD17-02-1-0350

TITLE: Identification of the Downstream Promoter Targets of Smad Tumor Suppressors in Human Breast Cancer Cells

PRINCIPAL INVESTIGATOR: Xuedong Liu, Ph.D.

CONTRACTING ORGANIZATION: The University of Colorado
Boulder, Colorado 80291-0220

REPORT DATE: October 2004

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

20050603 180

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	October 2004	Annual (1 Oct 2003 - 30 Sep 2004)

4. TITLE AND SUBTITLE
Identification of the Downstream Promoter Targets of Smad Tumor Suppressors in Human Breast Cancer Cells

5. FUNDING NUMBERS
DAMD17-02-1-0350

6. AUTHOR(S)
Xuedong Liu, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
The University of Colorado
Boulder, Colorado 80291-0220

E-Mail: xliu@wi.mit.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 Words)

TGF-ß is an important negative growth regulator in mammalian epithelial cells and acts to prevent tumor formation in vivo. The molecular mechanisms of tumor suppression by TGF-ß remain poorly understood. The major research goal of this proposal is to identify the downstream promoter targets of Smad tumor suppressor proteins in breast cancer cells and characterize heritable changes in tumor cells due to the deletion of Smad4 using genomic and bioinformatic techniques. We have investigated TGF-beta signaling responses in human mammary epithelial cells using whole genome transcriptional profiling analysis coupling with a powerful and innovative bioinformatic tool developed ourselves. We utilized this new algorithm, which is based on frequency of occurrence and cross-species conservation, to search for sequence elements that may convey TGF-ß responsiveness to the target genes identified by our microarray analysis. Two regulatory elements predicted by computational analysis were further confirmed experimentally by reporter gene assays. Thus, a combination of genomic and bioinformatic approach has lead to successful identification and characterization of Smad/TGF-ß responsive elements.

14. SUBJECT TERMS
TGF-beta, Activin A, promoter analysis, DNA microarray, human mammary epithelial cells

15. NUMBER OF PAGES
60

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

Cover.....

SF 298.....

Table of Contents.....1

Introduction.....2

Body.....2

Key Research Accomplishments.....6

Reportable Outcomes.....7

Conclusions.....7

References.....

Appendices.....8

Progress report:

1. INTRODUCTION

Members of transforming growth factor ß (TGF-ß) superfamily play important role normal mammary gland development and serves as tumor suppressor function. TGF-ß signals through cell surface receptors to activate downstream signaling mediator Smad2, Smad3 and Smad4 which form oligomeric complexes and migrate into nucleus to function as transcription factors to modulate TGF-ß-responsive gene expression. The goal of our investigation is to understand the molecular mechanism of tumor suppression by TGF-ß by identifying the downstream promoter targets of Smads tumor suppressors in normal and breast cancer cells. We have systematically identified TGF-beta responsive genes in human mammary epithelial cells through whole genome DNA microarray transcriptional profiling. Using a new algorithm we developed, we revealed transcription factors binding sites that are enriched in TGF-beta responsive genes and conserved across human, mouse and rat. Some of these elements have been characterized in previous studies in the field and validate our analysis method. We have also experimental confirmed two novel TGF-beta responsive elements in two TGF-beta inducible genes using reporter gene assays.

2. BODY---Studies and Results

Three specific aims were proposed in the original application:

1. Development of a novel chromatin immunoprecipitation assay (CHIPS) using a TAP-TAG system to isolate *in vivo* binding targets of Smad3 and Smad4.
2. Identification of the downstream promoter targets of Smad3 or Smad4 in breast cancer cells.
3. Identify Smad4 regulated downstream target genes in tumor cells using DNA microarray technology

The approved Statement of Work (SOW) for the second reporting period is as follows:

Task 1. Development of a novel chromatin immunoprecipitation assay (CHIPS) using a TAP-TAG system to isolate *in vivo* binding targets of Smad3 and Smad4, **(months 1-24)**

- grow sufficient quantity of MDA-MB468 cell lines for CHIPS analysis **(months 1-2).**
- Optimize the experimental procedure for two step purification of TAP tagged Smad3 or Smad4 from cell lysates **(months 3-5)**
- Optimize the crosslinking and sonication conditions for Smad3 and Smad4 **(months 6-8)**
- Establish a efficient combination of crosslinking and TAP purification procedure for enrichment of PAI-1 promoter and the goal is to achieve 25,000 fold enrichment of the PAI-1 Smad3/4 binding site **(months 9-24)**
- Annual reports will be written

In the previous budget year, we have constructed and characterized human breast cancer cell lines expressing TAP-Tag Smad3 and Smad4. We have done preliminary DNA microarray analysis on these two cell lines. Alhtough we have optimized the experimental procedure for two-step purification of TAP tagged Smad3 or Smad4 from cell lysates and were able to isolate the Smad signaling complexes from these cell

lines, we encounter significant difficulties in purifying specific DNA fragments associated with the Smad signaling complex. A number of fragments we cloned from CHIP did not show significant responses to TGF-beta in the reporter gene assays. However, CHIP is effective to recover the binding site in the promoter regions of Smad-dependent TGF-beta regulated genes in a mixture of IPed fragment if we know regions that are important for TGF-beta responses.

The success of the above mentioned strategy requires identification of TGF-beta responsive and Smad4-dependent and build up extensive bioinformatic tools to carry out this study. In last reporting period, we reported our bioinformatics effort to allow us identify regulatory elements that are most likely involved in conveying TGF-beta responsiveness. We started constructing a whole-genome promoter analysis software called GeneACT to allow us conduct high throughput comparative genome analysis, in which a user can search for binding sites in a huge set of genes in a relatively short period of time. One of the tools looks for conserved sequences between genomes of different species, such as human, mouse, and rat. The tool can display the gene sequence alignments graphically as well as textually. The construction of the software architecture has been finished (http://enhancer.colorado.edu:6400/~hudakg/home.html). However, this has not been released to the general scientific community yet since the capacity of server can only handle limited amount of traffic right now. We are now trying to find additional resource outside to allow us to implement this software package in more powerful computational platform.

Since we have already generated a list of probable elements that are likely involved in TGF-beta signaling and associate with Smad proteins, in the next budget year, we will focus our effort to characterize some of the elements in our table to determine if they are indeed associated with Smads using CHIP assay described here. Task 1 has been initiated but delayed to be implemented in the next budget year due to the overall shift in the experimental strategy to accomplish the goal we set out.

Task 2. Identification of the downstream promoter targets of Smad3 or Smad4 in breast cancer cells **(months 20-48)**

- Workout ligation mediated PCR protocol for amplification of unknown targets of Smad3/4 binding sites **(months 20-24)**
- Cloning of the amplified Smad3/4 binding sites into a luciferase reporter construct **(months 25-28)**
- Make small pool library of the cloned putative Smad3/4 binding sites. Pool size=10. Initial plan is to make 100 pools **(months 29-32)**
- Transient transfection of HepG2 cells each small pool and screen for TGF-ß responsive pools **(months 33-36)**
- Subdivide each positive pool to identify individual clone that mediate TGF-ß transcriptional response **(months 36-38)**
- Sequence each positive clone and obtain the identity of the genes that are regulated by TGF-ß through the binding site **(months 39-42)**
- Confirm the binding of the identified DNA fragment to purified Smad3 or Smad4 in vitro by a gel shift assay **(months 43-45)**
- Mutational analysis to confirm the importance of the Smad binding site in mediating TGF-ß transcriptional response **(months 43-48)**
- Final report and initial manuscript will be drafted.

The difficulties we encountered in CHIP assay using TAP-Smad3 and TAP-Smad4 prompt us to think about alternative strategies to reliably identify TGF-ß responsive elements. We had obtained comprehensive DNA microarray data in human mammary epithelial cells. Time and dose dependent gene expression profiles of TGF-beta and Activin A responsive genes were identified. We hypothesized that the specific TGF-beta/Smad regulatory elements are embedded in the promoter regions of the responsive genes. It is our expectation that TGF-beta/Smad responsive elements should present at higher frequency in the promoter regions of the TGF-beta inducible genes than those of the non-responsive genes.

It has been shown previously that genomic response to TGF-beta and Activin A are highly conserved between human, mouse and rat. It is our expectation that authentic TGF-beta responsive elements are likely conserved across genome. We came up with an alternative strategy to identify Smad/TGF-beta responsive elements in TGF-beta target genes. To analyze the potential binding sites in promoter regions of a large set of gene data from our microarray studies we have developed a search algorithm (GeneACT) to search for all potential binding sites in a high throughput manner for the genes that we reported earlier. We used the Transcription Factor Database (TFD, www.ifti.org) as the source of our binding site database, which contains approximately 6000 experimentally defined transcription factor binding sites described in the literature. For the genomic sequence information, Homo sapiens, Mus musculus and Rattus norvegicus genomes (NCBI) were parsed into our database. For faster searching, sequence data was converted from string format into bitstring format. To minimize the false positives that resulted in using pattern matching, comparative genome analysis has been employed in which only binding sites that are conserved in more than one genome are reported. Binding site frequencies were reported in two ways. The first way is on an individual gene level, in which the location of the binding sites of each gene is reported along with the sequence and binding site name. The second way is that it reports the frequency of a particular binding site found in the whole set of input gene names.

We used a set of 108 genes that are differentially expressed upon TGF-beta stimulation (at least 1.8 fold induction or repression at the 2 hour time point) and a set of genes that are not regulated by TGF□ (fold changes on microarray in between -0.001 fold to 0.001 fold in all four replicates) to search for all binding sites of these genes in their promoter regions upstream from the transcription start site (TSS). We hypothesized that the frequency of the TGF-ß responsive binding sites present in the TGF-ß regulated genes is significantly higher or lower than that of the unregulated genes. To examine this we used a set of 644 unregulated genes as our control set to reflect a basal frequency of a particular binding site occurrence in the genome upon ligand treatment. 108 TGF-ß regulated genes were also chosen and the frequency of each of the transcription factor binding sites existing in the TFD was calculated. Comparing the frequency of transcription factor binding sites between these two datasets allows us to identify binding sites that exist only in the regulated gene set. In addition, those transcription factor binding sites that occur more frequently in the regulated gene set than in the control set (>= 2.9 fold) are also documented.

To visualize the global distribution pattern of the statistically significant binding sites identified in our analysis in relation to the transcriptional response, a two-dimensional heatmap was generated. A representative version of this heatmap with a few representative entries is shown in Figure 8a. The transcription factor binding sites that occur more frequently in the regulated genes were further ranked by their frequency of distribution in the up-regulated vs. down-regulated genes and plotted in descending order on the y-axis. The regulated genes were ranked according to their fold changes observed from DNA microarray analysis and were plotted on the x-axis. The colored dots indicate the presence of a specific binding site in the promoter region of the regulated genes. As shown in Figure 1b, the plot revealed that certain transcription factor binding sites are exclusively associated with up-regulated genes (red dots) and down-regulated genes (green dots). In addition, a group of transcription factor binding sites occurs more frequently in up-regulated genes and down-regulated genes (yellow dots). Therefore, transcription factor binding sites enriched in regulatory regions of the TGF-ß regulated genes exhibit a nonrandom distribution correlated with the levels of induction.

Only a limited number of transcription factor binding sites highly enriched in TGF-beta-responsive genes. The most abundant binding sites identified from this study are Sp1/AP2, Ap-1, CRE/ATF, NF-kappa B, CAC/EKLF, GATA, Oct-1 and Ets. Some of these sites, such as Sp1, Ap-1, CRE/ATF and NF-kappa B, have previously been shown experimentally to be present in TGF-beta responsive promoters (Figure 1a). These results suggest that our approach is able to pinpoint experimental defined the regulatory elements and thus provide strong support for validity of this type of analysis.

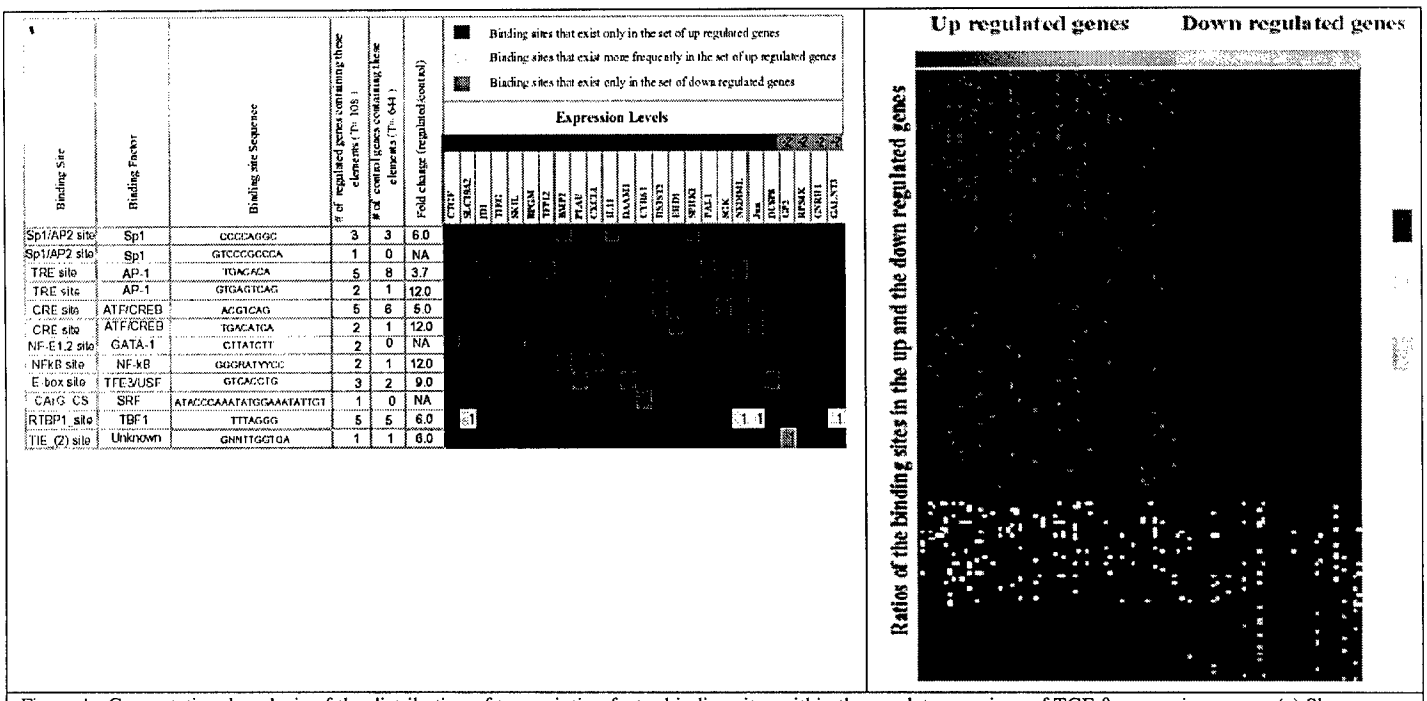


Figure 1. Computational analysis of the distribution of transcription factor binding sites within the regulatory regions of TGF-ß responsive genes. (a) Shown is a representative two-dimensional heatmap displaying the correlation between a few representative binding sites enriched in TGF-ß responsive genes and a number of representative differentially regulated genes sorted in descending order (from most induced to most repressed). The top row indicates approximate fold changes of these genes. Each row describes a specific transcription factor binding site that was found to exist exclusively (NA) or statistically more frequently in TGF-ß regulated genes. The presence of such a transcription factor binding site in TGF-ß responsive genes is designated as a colored square in the gene name column. The color code of the square is indicated in the figure. The columns on the left present all the detailed computational data associated with the transcription factor binding sites. (b) Two-dimensional heatmap displaying the correlation between all the transcription factor binding sites enriched in TGF-ß responsive genes and 108 differentially regulated genes sorted in descending order (from most induced to most repressed with changes at least 1.8 fold in either direction).

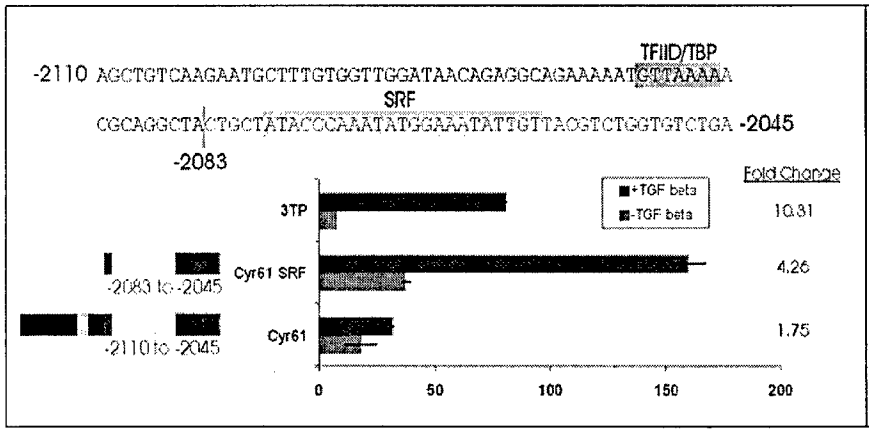


Figure 2a. Experimental validation of minimal TGF-ß responsive regulatory elements in CYR61 promoter region. Mink lung cells (CCL64) were transfected with reporter constructs indicated. p3TP-Lux reporter gene was used as the positive control. A schematic representation of the CYR61 promoter region identified by computational analysis was shown above the graph with known transcription factor binding sites highlighted. The fold induction by TGF-ß is indicated and error bars represent standard deviations from triplicate determinations.

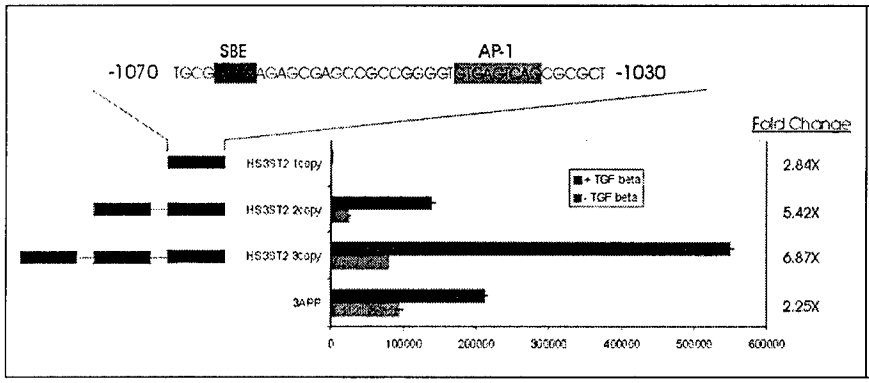


Figure 2b. Experimental validation of minimal TGF-ß responsive regulatory elements in HS3ST2 promoter region. Hep3B cells was transfected with reporter constructs containing one, two or three copies of the putative minimal TGF-ß responsive element from the promoter region of the HS3ST2 gene. p3APP-Lux was used as the positive control in this experiment.

Identification and Characterization of Two TGF-beta Responsive Elements

Our computational analysis suggested a collection of potential TGF-ß responsive elements in the genome. Whether any of these elements other than the ones that are well-characterized in the literature make biological senses remains to be determined experimentally. To begin with, we chose two TGF-ß targets genes CYR61 and HS3ST2 from our microarray list. The regulatory elements that are responsible for TGF-ß responsiveness in the promoter regions of these two genes have never been characterized. We demonstrated that a 40 bp sequence consisting of the TRE element next to AGAC is a TGF-ß responsive element for HS3ST2 and 38-bp region consisting of the SRF element is responsible for CYR61 transcriptional response to TGF-beta. In summary, a majority of the goals in Task 2 was accomplished even though the route to get there is not exactly as planned.

In the next budget year, we will further characterize these two elements through mutation analysis. In addition, we will test whether these two elements bind Smads, AP-1 and SRF in vitro by DNA affinity assay or in vivo by the CHIP assay.

Task 3. Identify Smad4 regulated downstream target genes in tumor cells by DNA microarray **(months 12-30)**

- prepare high quality of mRNA for DNA microarray analysis **(months 12-14)**
- run test chip experiment to familiar with the procedure and calibrate the reagent **(months 15-16)**
- Prepare high quality cRNA for hybridization to the U95 CHIP **(months 17-18)**
- Hybridization, scan and data collection **(months 19-20)**
- Analysis the DNA microarray data using gene spring or cluster software **(months 20-24)**
- Annual report will be written **(months 20-24)**
- Repeat the DNA microarray experiment to ensure the high reproducibility of the data **(months 25-30)**
- Summary of DNA microarray data will be written and initial manuscript will be drafted **(months 25-30)**

The task 3 is ahead of the schedule. We have collected DNA microarray data from three pairs of cell lines differ by the Smad4 expression. Human 1A Oligo microarrays instead of U95CHIP (Agilent Technologies, Palo Alto, CA) were used to perform all the DNA microarry analysis. Some of the informative genes were further confirmed by real-time PCR analysis. The data obtained were also analyzed by the GeneACT software package developed in our lab with the support of this award. We are in the process of preparing manuscripts to describe our findings.

3. KEY RESEARCH ACCOMPLISHMENTS

- Obtained gene expression profiling data in human mammary epithelial cells in response to TGF-beta and Activin A.
- Obtained gene expression profiling data in human mammary epithelial cells in response to various concentrations of Activin A.
- Construct a human, mouse and rat promoter database for bioinformatic analysis of TGF-ß responsive promoters
- Obtained a complete dataset for the regulatory elements in the promoter regions of the TGF-beta responsive genes that conserved across human, mouse and rat genome.

- Identified two novel TGF-beta responsive elements that are responsible for TGF-beta induced transcriptional activation of Cyr61 and HS3ST2
- Further characterized the Smad4-dependent and Smad4-independent TGF-ß responsive genes in MDA-MB468, SW480 and CFPAC-1 tumor cell lines. Secondary confirmation of the microarray data were acquired through real-time PCR analysis
- Improved the functionality and usability of GeneACT bioinformatics analysis tool.

4. REPORTABLE OUTCOMES

Cheung, H.T., Collins, P.J., Riquelme, C., Kwan, P., Doan, T.B and X.Liu Specificity of TGF-beta and Activin Signaling Responses Revealed by the Analysis of Their Transcriptional Programs. *Submitted to Molecular Cellular Biology and in revision.*

Web-based GeneACT Promoter Analysis Algorithm

Identified two novel TGF-beta responsive elements

Publications not supported by the grant

Macdonald, M, Wan,Y., Wang, W., Erickson,R.E., Cheung,T and X.Liu Control of cell cycle dependent degradation of c-Ski proto-oncoprotein by Cdc34. *Oncogene* 23(33):5643-53, 2004

Royer, Y, Menu, C, Liu, X, and S.N. Constantinescu. High-Throughput Gateway Bicistronic Retroviral Vectors for Stable Expression in Mammalian Cells: Exploring the Biologic Effects of STAT5 Overexpression. *DNA Cell Biol.* 6: 355-65, 2004.

Liang, M, Liang, Y, Wrighton, K, Ungermannova, D Wang, X, Brunicardi, F, Liu, X, Feng, X and X. Lin Ubiquitination and Proteolysis of Cancer-derived Smad4 Mutants by SCF^{Skp2}. *Molecular Cellular Biology.* Sep;24(17):7524-37, 2004

Wang, W, Ungermannova,D, Chen, L and X.Liu Molecular and Biochemical characterization of Skp2-Cks1 interface. *J Biol Chem. 2004 Sep 27 in press.*

5. CONCLUSIONS

The goal of the proposed studies is to identify the downstream promoter targets of Smad tumor suppressors in normal and breast cancer cells. We have performed comprehensive transcriptional profiling of the normal mammary and breast cancer cells. We have developed a new algorithm, called GeneACT which is based on frequency of occurrence and cross-species conservation, to search for sequence elements that may convey TGF-beta responsiveness to the target genes identified by our microarray analysis. A list of transcription factor binding sites that are over-represented in TGF-beta responsive genes was identified. Some of the binding identified in our analysis match exactly the binding sites characterized experimentally by numerous investigators in this field. Two novel TGF-beta responsive elements predicted by our analysis were characterized experimental and confirmed that they are able to convey TGF-ß signaling. We are working to improve the chromatin immunoprecipitation (CHIP) assay to determine if Smad3 and Smad4 associate directly with some of the promoter elements identified in our bioinformatics and experimental analysis.

Specificity of TGF-ß and Activin A Signaling Responses Revealed by the Analysis of Their Transcriptional Programs

Tom H Cheung,[1] Patrick J. Collins,[2] Cecilia Riquelme,[1] Phoenix Kwan,[1] Tri B. Doan[2] and Xuedong Liu[1,3]

[1]Department of Chemistry and Biochemistry, University of Colorado-Boulder, Boulder, Colorado 80309

[2]Agilent Technologies, 3500 Deer Creek Road, 25U, Palo Alto, CA 94304

[3]Corresponding author: Tel: (303) 735-6161

Fax: (303) 735-6161

Email: Xuedong.Liu@Colorado.edu

Running title: Activin A and TGF-ß Transcription Responses

Keywords: Activin A, TGF-ß, DNA microarray, transcription, Smads, bioinformatics

Abstract

Activins and TGF-ß can signal through distinct cell surface receptors and activate the same downstream intermediates Smad 2, Smad3 and Smad4. However, the biological activities of these two ligands only partially overlap in diverse biological systems. The molecular basis for signaling similarity and specificity between these two ligands is not yet understood. We investigated Activin A and TGF-ß transcriptional responses in immortalized normal human mammary epithelial cells by gene expression profiling. We demonstrated that Activin A and TGF-ß elicit overlapping but distinct transcriptional programs. Activin A signaling is relatively transient compared to TGF-ß and correlated with quantitative levels of Smad2 phosphorylation and nuclear translocation in response to variable concentrations of ligands. In addition, we analyzed and compared sequence compositions of the regulatory regions of TGF-ß-responsive genes in human, mouse and rats genomes using a unique computational method. Our analysis revealed that a distinct set of sequence elements conserved across species is either unique or occurs at a much higher frequency in TGF-ß-regulated genes. These regulatory elements include some of the previous well-characterized TGF-ß-responsive elements as well as a number of transcription factor binding sites that have not been implicated TGF-ß signaling. Two regulatory elements in two separate TGF-ß target genes predicted by our computational analysis were further confirmed experimentally. Thus, TGF-ß-regulated transcription appears to be conducted by a limited set of regulatory elements alone or in combinations.

Introduction

TGF-ß, Activin, Bone morphogenic proteins (BMPs), Mullerian inhibiting substances (MIS), and GDFs are members of the transforming growth factor-ß superfamily and play important roles in cell growth, differentiation and development (20). These structurally related growth factors function as ligands to trigger signal transduction programs that control gene expression (37). Members of the TGF-ß superfamily interact with two different types of cell surface serine/threonine kinase receptors known as type I and type II receptors. Ligand binding results in assembly of a type I and type II receptor complex, phosphorylation of the type I receptor by the type II receptor and activation of the kinase activity of the type I receptor. The type I receptors then recognize their intracellular substrates, receptor -regulated Smads (R-Smads), and phosphorylate them at the carboxyl terminal SSXS motif (1, 28). R-Smads are pathway-specific signaling traducers, which include Smad1, 2, 3, 5 and 8. Once phosphorylated upon ligand stimulation, selective pathway-specific Smads form complexes with Smad4, the common-partner Smad (13). The resulting Smad complexes translocate into the nucleus, bind DNA directly and recruit other transcription factors and other cofactors to positively or negatively regulate gene expression (7).

Activin A and TGF-ß are two distinct but structurally related members of the TGF-ß superfamily. Despite that there is only 30% homology between these two ligands they appear to share much of the same signaling machinery downstream of their respective receptors (6, 31). Both ligands activate Smad2, Smad3 and Smad4 and the constitutively active Activin and TGF-ß type I receptors modulate a similar

transcription output upon overexpression in cultured cells (34, 35). However, in many developmental systems, Activins and TGF-ßs trigger distinct and sometimes opposite biological effects on cell proliferation in different tissues (3, 4). The qualitative and quantitative aspects of biological responses elicited by Activins and TGF-ßs under physiologically relevant settings remains to be determined.

The TGF-ß superfamily triggers a myriad of transcriptional responses. Identification of the genes that are regulated by the signaling cascade of each family member and elucidation of the mechanisms underlying specificity of gene induction are crucial for understanding the biological activity of these ligands in physiologically relevant processes. Induction of TGF-ß target genes could be mediated by Smad-dependent and Smad-independent signaling cascades (7). R-Smads and Co-Smad, Smad4, share highly conserved MH1 and MH2 domains separated by a variable linker region. The MH1 domain exhibits sequence-specific DNA binding activity whereas the MH2 domain is involved in transactivation and homo- or hetero-oligomerization (28). An 8-bp DNA sequence element (5'-GTCTAGAC-3') was identified as the high-affinity binding sites for the DNA binding domain of Smad3 and Smad4 using an *in vitro* selection approach (46). Characterization of the sequence elements in known TGF-ß-inducible genes revealed few DNA elements identical to the 8-bp high affinity-binding site; instead, most of the TGF-ß responsive elements contain only a 4-bp DNA sequence (5'-GTCT-3' or 5'-AGAC-3') (7). In addition, binding sites for other transcription factors are frequently located adjacent to these 4-bp elements suggesting Smads activate transcription through functional cooperation with other sequence-specific transcription factors (7).

TGF-ß-inducible gene expression can also occur through Smad-independent pathways. Activation of the MAP kinase pathway can regulate the activity of downstream transcription factors and turn on transcription of the target genes even in the absence of Smads (7). Thus, TGF-ß signaling converges at the promoter regions of the targets genes and elicits transcriptional responses by assembling transcription complexes recruited by Smads or activated by Smad-independent signaling pathways such as the MAP kinase pathway.

Hundreds of genes have been shown to be regulated by TGF-ß experimentally. A diverse group of transcription factor binding sites were found to mediate the TGF-ß transcriptional response. However, there has been no systematic analysis at the genome scale to enumerate the predominant binding elements associated with TGF-ß responsive genes. Here we investigated genomic responses upon TGF-ß and Activin A stimulation in the TGF-ß superfamily in telomerase immortalized human mammary epithelial cells. Comparative analysis of the DNA binding elements in the regulatory regions of the responsive genes identified in our study suggests that TGF-ß- responsiveness could be conferred by a distinct set of regulatory elements in the promoter regions of TGF-ß inducible genes.

Materials and Methods

Cell lines

HME (Human Mammary Epithelial) cells were purchased from Clontech and cultured in MEBM (Mammary Epithelium Basal Medium) supplemented with 52 μg/ml BPE, 0.5 μg/ml hydrocortisone, 10 ng/ml hEGF, 5 μg/ml insulin, 50 μg/ml gentamicin

and 50 μg/ml amphotericin-B (Clonetics). HME cells are immortalized by overexpression of the catalytic subunit of telomerase (TERT). Mink lung epithelial cells and Human Hep3B cells were purchased from ATCC and maintained in DME medium.

Northern blot analysis

HME cells were treated with the indicated concentrations of TGF-ß and Activin A (R&D system) at various times. Total RNA was isolated from HME cells using a RNeasy kit (Qiagen) following the manufacturer's instructions. RNA samples (10ug of each) were electrophoresed on a 1% agarose-formaldehyde gel and were blotted to a Nytran membrane (Amersham). The cDNA probes for various genes were randomly labeled with ^{32}P-dCTP using a RediprimeTM II kit (Amersham Bioscience) and hybridized overnight at 42°C with the membrane in Ultrahyb Hybridization buffer (Ambion).

Antibodies and Immunoblotting analysis

Protein extracts were prepared from HME cells by lysing equal numbers of cells directly in passive lysis buffer in the presence of protease and phosphatase inhibitors (Promega). The protein concentrations were measured by Bradford assay (Biorad). Samples were resolved on 12% SDS-PAGE gels and electrphoretically transferred to nitrocellulose membranes. Western blot analysis was performed using phospho-Smad2 antibody (kindly provided by Peter Ten Dijke, Aris Moustakis and Carl Heldin). The proteins were detected using HRP conjugated rabbit secondary antibody (Amersham) with a WestDura detection kit (Pierce).

DNA Microarray Experiments

Human 1A Oligo microarrays (Agilent Technologies, Palo Alto, CA) were used in this study. Cyanine 3-labeled or cyanine 5-labeled amplified RNA targets were generated from 100 ng of total RNA using Agilent's Low Input RNA Fluorescent Linear Amplification kit (Agilent Technologies). In each experiment, cyanine 3-labeled RNA amplified from "control" cells was mixed with cyanine 5-labeled RNA from "test" cells and hybridized to Human 1A microarrays. Each experiment consisted of four labeling and hybridization replicates. Microarrays were scanned using the Agilent dual laser scanner and data were extracted using Agilent's Feature Extraction software(Agilent Technologies).

Microarray Data Analysis

Data were analyzed using a combination of Rosetta Resolver Gene Expression Analysis System (Rosetta Biosoftware, Seattle, WA) and customized algorithms. Customized scripts were implemented in Structured Query Language (SQL). Log ratio error values derived from Agilent's Feature Extraction software were used in error-weighted averaging of replicate log ratio values. Spotfire's Functional Genomics (Spotfire, Somerville, MA) was used for Hierarchical clustering of selected gene sets.

Computational Genome-wide Transcription Factor Location Analysis

Genomic data (*Homo sapiens, Mus musculus* and *Rattus norvegicus*) was downloaded from NCBI and the TFD (Transcription Factor Database) were downloaded from IFTI (Institute for Transcriptional Informatics). TFD is a transcription factor database that contains binding site sequences reported in the

literature. A custom designed genome database was built using these data sets. In brief, 108 genes regulated by TGF-ß and 644 not regulated by TGF-ß (control) were fed into the database to search for all potential binding sites from –2500 to +100 upstream from the transcription site for all three species. Only those binding sites that go across more than 2 species were selected for further analysis. The frequencies of bindings sites found from the 108 regulated genes were then compared to the 644 control genes. Frequency ratios of binding sites were calculated, in which binding sites with a two-fold difference or more and those only exist in the regulated but not the control set, were then mapped onto the gene features that were differentially expressed by more than 1.8 fold (at the 2 hour time point of TGF-ß treatment) on the microarray analysis. A two-dimensional heatmap diagram was generated using the expression data on one dimension and binding site sequences on the other dimension. The expression data were sorted from low to high (most repressed to most induced) and the binding sites were sorted using the ratios of the binding sites in the up and down regulated genes.

Gene Ontology Analysis

Gene Ontology Analysis was performed by using the GO-Getter mapping program (http://bayes.colorado.edu/go-getter)(12). In brief, probe IDs from the Agilent Human 1A array were used directly to link to different GO ontology IDs using GO-Getter (12). TGF-ß up regulated genes were compared to the Activin A up regulated genes and vice versa. Percentages of genes in each GO category for each treatment were calculated from each main GO category (Molecular Function, Cellular Component and Biological Process). Bar charts comparing numbers of genes in each GO category for each treatment were then plotted.

Luciferase Reporter Gene Assay

The TGF-ß responsive reporter constructs p3TP-Lux and p3APP-Lux have been described previously (14, 42). To test whether sequence elements identified by computational analysis are able to mediate TGF-ß induced transcriptional activation, we cloned some of the representative promoter elements into the KpnI-PstI site of p3APP-lux by substituting the existing TGF-ß responsive elements cloned previously between the two restriction sites. Genomic sequences corresponding -2110 to -2045 of CYR61 (LocusID: 3491) or -1070 to -1030 of HS3ST2(heparan sulfate (glucosamine) 3-O-sulfotransferase 2) (LocusID: 9956) relative to their transcription initiation sites were PCR amplified using two pairs of primers with KpnI and PstI attached at the end of primers and subsequently cloned to the compatible site in p3APP-Lux. The SRF binding site in CYR61 and TRE-like sequence elements were cloned in p3APP-lux by annealing two pairs of oligonucleotides with KpnI and PstI overhangs. To test the TGF-ß responses of various reporter constructs, we transient transfected the indicated constructs into mink lung cells or Hep3B cells using Fugene 6(Roche). Twenty-four hours after transfection, cells were switch into low serum medium (0.1% FBS) either in the presence or absence of TGF-ß and incubated for another twenty four hours prior to harvesting. Luciferase activity was determined as described previously using a Dynex luminometer (26).

Results

Analysis of TGF-ß and Activin A regulated gene expression patterns in human mammary epithelial cells using an oligonucleotide microarray

TGF-ß and Activin A play important roles in mammary gland differentiation in mammals (4). Mammary epithelial cells are responsive to treatment by both ligands (25). Human mammary epithelial cells (HME), immortalized by telomerase overexpression maintain the properties of normal mammary epithelial cells and identifying the gene expression patterns in this cell line is likely to be biologically relevant. To compare TGF-ß and Activin A-regulated gene expression patterns, total RNA was isolated from human mammary epithelial cells was treated with 100 pM TGF-ß or Activin A for 2, 4 and 8 hours. To assess the changes in relative abundance of transcripts in response to TGF-ß and Activin A treatment, total RNA from non-treated control cells (T0 for control cells that were not treated with TGF-ß and A0 for Activin A non-treated cells) or treated cells were amplified and labeled with either Cy3 or Cy5 fluorescent dyes. In each experiment, Cy3 labeled amplified RNA (aRNA) from non-treated cells was mixed with Cy5 labeled amplified RNA derived from TGF-ß or Activin A treated cells at the indicated time points and hybridized to Human 1A 60-mer oligonucleotide arrays representing more than 17,000 human genes. Each experiment consisted of four replicates of hybridization.

The overall patterns of time-dependent differential gene expression induced by Activin A and TGF-ß are shown in Figure 1. It becomes evident that the number of genes differentially regulated by TGF-ß differs significantly from Activin A. Table 1 shows the summary of the transcriptional responses elicited by the treatment of TGF-ß

or Activin A. Only genes that are differentially expressed by more than 1.8 fold with a p value <0.01 were selected for further analysis. We have identified 129 genes that are differentially expressed in response to TGF-ß and only 64 genes for Activin A (Tables of the complete dataset are presented in the supplementary data 1). There is a significant overlap between the TGF-ß and Activin A responsive genes (Figure 2). About half of the Activin A regulated genes are also responsive to TGF-ß. At the same ligand concentration (100 pM), TGF-ß responsiveness is more robust than that of Activin A treated cells (Figure 1, Suppl. 1). Therefore, Actvin A appears to induce similar but not identical transcriptional responses. The spectrum of TGF-ß responsive genes is much broader than Activin A regulated genes in human mammary epithelial cells.

A list of representative genes that are regulated by TGF-ß and Activin A treatment is shown in Figure 3. Genes that have not been previously reported as TGF-ß regulated genes are indicated by "*". To validate the microarray data, four genes that are differentially expressed by both TGF-ß and Activin A treatment were selected for northern blot analysis. As shown in Figure 4a, northern blot results are highly consistent with the microarray data and provide secondary confirmation of the DNA microarray data.

Expression patterns of TGF-ß and Activin A regulated genes

Examination of the kinetics of responsive genes that are commonly regulated by TGF-ß and Activin A revealed that there are significant differences in the duration of signaling responses between these two ligands in HME cells. Activin A triggered transcriptional responses are short-lived while TGF-ß responses are relatively more

persistent (Figure 4b). For example, activation of Angiopoietin-4 occurs within 2 hr of treatment with Activin A and its induction levels drop off by the 4 hr time point (Figure 3 & 4a). In contrast, TGF-ß induced Angiopoeitin-4 expression persists more than 4 hr and even at the 8 hr time point there is still significant expression. Plotting the number of differentially expressed genes in response to Activin A and TGF-ß treatment revealed that transcriptional responses to Activin A peaked at 4 hr and declined afterwards (Figure 4b). In contrast, TGF-ß responses were persistent and increased with the time of treatment (Figure 4b). The reasons behind these apparent differences between TGF-ß and Activin A in gene induction are likely to be complex. One of the obvious hypotheses is that these two ligands have different capacities to activate downstream signal transducer-Smads. To test this hypothesis, we performed immunoblotting analysis to investigate R-Smad activation in response to TGF-ß and Activin A in HME cells. Both Smad2 and Smad3 have been reported to be activated by type I Activin A or type I TGF-ß receptor kinases upon ligand binding (32). Phosphorylation of the carboxyl-terminal SSXS motif can be analyzed by a specific antibody raised against the phosphorylated SSXS peptide (32). As shown in Figure 5, phosphorylated Smad2 is readily detectable in both TGF-ß and Activin A treated HME cells. While TGF-ß induces rapid and persistent Smad2 phosphorylation, Activin A only induces transient Smad2 phosphorylation and its magnitude of activation is significantly less dramatic than TGF-ß. Therefore, TGF-ß and Activin A have different capacities to activate Smad2, an effect that could be a result of differences in their type I receptor Ser/Thr kinase activities, different rates of receptor endocytosis or

dephosphorylation of Smad2 by an unknown phosphatase that is differentially regulated by these two ligands (8, 15).

Genes that are similarly regulated can be clustered together based on their overall gene response kinetics. Genes that fall into the same cluster are likely to be coordinately regulated. Using K-means clustering analysis, we can readily assign ten different clusters of gene induction patterns (Suppl. 2 and Suppl. 3). For example, Angiopoeitin-4, HEF1 and CTGF show similar ON/OFF patterns even though the induction magnitudes are quite different. The fact that a significant number of TGF-ß and Activin A responsive genes shared similar cluster patterns implies that these two pathways must share a similar information flow paths but have different kinetic responses.

Dose-dependent response of Activin A target genes

Another potential mechanism that could account for the difference between TGF-ß and Activin A transcriptional programs may be the difference in the effective concentrations of respective ligands used in our experiments. Activin A is a classical example of a gradient morphogen that triggers concentration-dependent cell fate determination in early embryo development (9, 11). To determine whether Activin A displays dose-dependent transcription regulation in HME cells, we first examined Smad2 phosphorylation in response to increasing concentrations of Activin A. As shown in Figure 5, Smad2 phosphorylation increased significantly in cells treated with higher concentrations of Activin A; however, higher concentrations of Activin A do not appear to affect the kinetics of Smad2 phosphorylation. Phosphorylated Smad2 is still diminished after 8 hrs of Activin A treatment even at 800 pM (Figure 5). The

efficiency of translocation of phosphorylated Smad2 from the cytosol to the nucleus in response to ligand treatment was also investigated using immunofluorescence (data not shown). As observed by immunoblotting experiments, the levels of intracellular phospho-Smad2 accumulation increase with higher concentrations of Activin A; however, there is little difference in phospo-Smad2 accumulation when concentrations of TGF-ß were varied from 50 pM to 800 pM (data not shown).

To identify genes whose transcription varies in response to different concentrations of Activin A, HME cells were treated with 50 pM, 200 pM and 800 pM Activin A for 4 hr and total RNA was isolated for each treatment. Figure 6a shows a list of Activin A dose-responsive genes. To validate the identification of dose-dependent genes, northern blot analysis was performed with HEF1 (Figure 6b), a typical Activin A concentration dependent gene. Again, we found good agreement between northern blot results and DNA microarray experiments. Interestingly, the magnitude and spectrum of Activin A transcriptional response are still far more subdued even though the concentration of Activin A is eight times higher than the concentration of TGF-ß used in the result shown in Figure 3. Taken together, these experiments suggest Activin A signaling appears to be tunable and transcription of some of the Activin A target genes is ligand dose-dependent.

TGF-ß and Activin A Signaling Program

Treatment of HME cells with TGF-ß and Activin A effectively changed the gene expression programs and reprogrammed the cellular output. The readjustment of cellular content results in resetting the response network to enable cells to adopt a different identity. Detailed classification of the Activin A and TGF-ß regulated genes

could offer unique insights into the biological processes they may influence. Each of the differentially regulated genes by Activin A and TGF-ß was assigned to a designated category, namely, "molecular function", "cellular component" and "biological process" as defined by the Gene Ontology Consortium database using custom designed software (GO-Getter). The spectrum of molecular functions of target genes up-regulated by Activin A and TGF-ß showed a similar distribution except that a higher percentage of hydrolase activity was observed in Activin A up-regulated genes (Figure 7). Another notable difference is that TGF-ß appears to regulate genes associated with motor activity, structural molecule activity and oxidoreductase activity while Activin A does not. A similar pattern is observed in the cellular component classification in that only TGF-ß regulates gene products related to cytoskeleton, membrane and ribonucleoprotein complexes. When the up-regulated target genes by either ligand are annotated by the biological processes involved, it becomes evident that TGF-ß activates a number of genes that are involved in cell cycle control, cell death, cell proliferation and differentiation while Activin A fails to do so. These results suggest that TGF-ß rather than Activin A has a more pronounced effect on cellular proliferation programs.

Whereas a similar gene ontology is displayed in the genes that are up-regulated by both ligands, there is significant divergence among genes that are down-regulated by Activin A and TGF-ß. For example, more than 25% of genes down-regulated in response to Activin A belong to nucleic acid binding proteins compared to only 7% of genes suppressed by TGF-ß with regard to molecular function. Activin A appears to selectively repress genes in the category of structural molecule activity (22% vs. 2%).

Oxidoreductase activity is negatively regulated by Activin A but positively regulated by TGF-ß. Along the same line, genes associated with ribonucleoprotein complex, which accounts for close to 20% of genes down-regulated by Activin A, were not the targets of TGF-ß down-regulation at all (Figure 7). Taken together, these data suggest that there is a significant similarity among genes that are up-regulated by Activin A and TGF-ß with regard to their functional category but significant differences among genes that are down-regulated by these two ligands, at least in human mammary epithelial cells. Such a difference may contribute to their distinct functions during mammary gland cell differentiation and development.

Computational Analysis of the Promoter Regions of TGF-ß Target Genes

To understand how TGF-ß selectively turns on or off transcription of its targets at a global level, it is crucial to identify the specific regulatory DNA elements embedded in the promoter regions of the responsive genes. TGF-ß induced transcriptional responses could occur through regulatory modes of a hierarchal or parallel nature or a combination of both. A hierarchal model would predict that TGF-ß transcriptional activation involves a stepwise activation scheme. Early response genes are activated to set up the expression of the delayed response gene. It would also predict that regulatory regions of early response genes must have some unique features to allow them to be sensitive to TGF-ß. A parallel regulatory model would suggest that when Smads are activated, they directly participate in regulation of TGF-ß responsive genes alone or associate with other transcription factors to effect direct activation or repression of target genes. By computational analysis, Bottinger and coworkers

examined TGF-ß transcriptional responses in mouse fibroblasts. When the putative Smad binding site (SBE, 5'-GTCTG-3') was used to search promoter regions of TGF-ß responsive genes, they found no statistical significance in the concurrence of the SBE and binding sites for unrelated eukaryotic transcription factors. Instead, they found that the sequence GTCT in a direct repeat with variable spacing between units occurs significantly higher in their dataset of early-responsive genes but not the delayed responsive genes and these data support a hierarchical model of transcriptional response (44). However, there is a plethora of experimental evidence supporting the notion that Smads activate transcription through physical interactions and functional collaborations with other sequence-specific transcription factors.

It has been well established that specific transcription factor binding elements in the promoter region are largely responsible for differential gene expression. It is our expectation that there is a differential distribution of regulatory sequence elements between TGF-β responsive and nonresponsive genes. Furthermore, there is a considerable conservation in the TGF-β regulated gene expression pattern between human and mouse genomes (5, 34, 43, 44). It is therefore reasonable to assume that at least some of the TGF-β responsive sequence elements would be conserved across species. We aimed to determine whether there are unique or high occurrence regulatory elements in the control regions of the TGF-β responsive genes that are conserved across at least two species.

To analyze the potential binding sites in promoter regions of a large set of gene data from our microarray studies we have developed a search algorithm (GeneACT) to search for all potential binding sites in a high throughput manner for the genes that we

reported earlier. We used the Transcription Factor Database (TFD, www.ifti.org) as the source of our binding site database, which contains approximately 6000 experimentally defined transcription factor binding sites described in the literature. For the genomic sequence information, *Homo sapiens, Mus musculus* and *Rattus norvegicus* genomes (NCBI) were parsed into our database. For faster searching, sequence data was converted from string format into bitstring format. To minimize the false positives that resulted in using pattern matching, comparative genome analysis has been employed in which only binding sites that are conserved in more than one genome are reported. Binding site frequencies were reported in two ways. The first way is on an individual gene level, in which the location of the binding sites of each gene is reported along with the sequence and binding site name. The second way is that it reports the frequency of a particular binding site found in the whole set of input gene names.

We used a set of 108 genes that are differentially expressed upon TGF-β stimulation (at least 1.8 fold induction or repression at the 2 hour time point) and a set of genes that are not regulated by TGFβ (fold changes on microarray in between -0.001 fold to 0.001 fold in all four replicates) to search for all binding sites of these genes in their promoter regions upstream from the transcription start site (TSS). We hypothesized that the frequency of the TGF-ß responsive binding sites present in the TGF-ß regulated genes is significantly higher or lower than that of the unregulated genes. To examine this we used a set of 644 unregulated genes as our control set to reflect a basal frequency of a particular binding site occurrence in the genome upon ligand treatment. 108 TGF-ß regulated genes were also chosen and the frequency of each of the transcription factor binding sites existing in the TFD was calculated. The

results of the calculations of both the control and the regulated genes are summarized in Suppl. 4. Comparing the frequency of transcription factor binding sites between these two datasets allows us to identify binding sites that exist only in the regulated gene set. In addition, those transcription factor binding sites that occur more frequently in the regulated gene set than in the control set (>= 2.9 fold) are also documented.

To visualize the global distribution pattern of the statistically significant binding sites identified in our analysis in relation to the transcriptional response, a two-dimensional heatmap was generated. A representative version of this heatmap with a few representative entries is shown in Figure 8a. The transcription factor binding sites that occur more frequently in the regulated genes were further ranked by their frequency of distribution in the up-regulated vs. down-regulated genes and plotted in descending order on the y-axis. The regulated genes were ranked according to their fold changes observed from DNA microarray analysis and were plotted on the x-axis. The colored dots indicate the presence of a specific binding site in the promoter region of the regulated genes. As shown in Figure 8b, the plot revealed that certain transcription factor binding sites are exclusively associated with up-regulated genes (red dots) and down-regulated genes (green dots). In addition, a group of transcription factor binding sites occurs more frequently in up-regulated genes and down-regulated genes (yellow dots). Therefore, transcription factor binding sites enriched in regulatory regions of the TGF-ß regulated genes exhibit a nonrandom distribution correlated with the levels of inductions. Some of the most frequent binding sites in TGF-ß regulated genes are Sp1, AP-1, NF-κB and ATF/CRE. It has been very well documented that these binding sites often mediate TGF-ß transcriptional responses in a

number of well-characterized genes (7, 16-19, 22-24, 47). For example, our analysis indicate that the AP-1 site (ATGTGTCAG) in IL11, the Sp1/AP-2 site (CCCCACCCCC) in TIEG and the ATF/CRE site (GTGACGTMR) in ID1 are enriched in TGF-ß regulated genes (Suppl. 4). All three binding sites are exactly the ones reported in the literature shown to be experimentally responsible for TGF-ß induction of these genes(10, 17, 38). Thus, there is a very good agreement between our computational analysis and experimental data, suggesting that our approach is valid. The fact that these binding sites exist in a number of other TGF-ß genes in our dataset suggests that these elements could contribute to their responsiveness to TGF-ß.

Experimental Validation of TGF-ß Responsive Elements in CYR61 and HS3ST2 Promoters Identified from Computational Analysis

Our computational analysis suggested a collection of potential TGF-ß responsive elements in the genome. Whether any of these elements other than the ones that are well-characterized in the literature make biological senses remains to be determined experimentally. To begin with, we chose two TGF-ß targets genes CYR61 and HS3ST2 from our microarray list. The regulatory elements that are responsible for TGF-ß responsiveness in the promoter regions of these two genes have never been characterized. Data presented in Figure 8 implicated that the region surrounding -2110 and -2045 in CYR61 and -1070 and -1030 in HS3ST2 are likely to be involved in mediating TGF-ß responses. Another reason for selecting these two regions is because the nucleotide sequences of these regions are conserved between human, mouse and rat. The indicated regions (Figure 9) were cloned into a luciferase reporter construct (pGL3). To test whether the promoter fragment containing -2110 to -2045 of CYR61

can confer a TGF-ß response, the reporter construct was transfected into mink lung epithelial cells and Hep3B cells. These two cell lines were selected because they are highly transfectable and have been used as model cell lines for analyzing TGF-ß signaling. HME cells are less transfectable and TGF-ß transcriptional responses in HME are transient (Figure 4 and 5) thus made it difficult to perform reporter gene assays. As positive controls, p3TP-Lux and p3APP-Lux, two standard TGF-ß signaling reporters, were also transfected. As shown in Figure 9a, the region spanning -2110 to -2045 is able to confer a modest TGF-ß response (1.75 fold increase). A consensus SRF binding site is located between -2083 and -2045. To test whether this SRF site is responsible for TGF-ß induction, a pair of oligos containing the SRF sequence (underlined) was inserted into pGL3 (Figure 9). In the presence of TGF-ß, this reporter gene showed 4.26 fold activation indicating the SRF sequence element is able to mediate TGF-ß induction and most likely the CYR61 gene itself.

The gene encoding for heparan sulfate (glucosamine) 3-O-sulfotransferase 2 (HS3ST2) was found to be TGF-ß regulated gene in this study. The region spanning -1070 and -1030 was found to be theone containing the candidate regulatory elements by the computational analysis. Within this region there is a TRE element (GTGAGTCAG) and a potential Smad binding element (SBE) (Figure 9b). To test effectiveness of this relative small region to enable TGF-ß induction, reporter constructs consisting of one, two or three copies of this elements were made and transfected into Hep3B and mink lung cells. The results shown in Figure 9b is data obtained with Hep3B, similar results were obtained with mink lung cells. A single copy of this element is able to elicit a 2.84 fold of activation in the presence of TGF-ß. As the copy of number of this

responsive element increases, so is TGF-ß induction as well as the basal levels of transcription. This result indicates that this 40 bp sequence consisting of the TRE element next to AGAC is a TGF-ß responsive element for HS3ST2. Taken together our experimental studies in the two cases we investigated support our computational predictions. Further experiments will be necessary to validate other candidate elements in an effort to fully categorize the regulatory elements responsible for TGF-ß induction.

Discussion

Activin A and TGF-ß signal transduction pathways appear to overlap significantly. Both of them can activate Smad2, 3 and 4 and inhibit cell proliferation in certain cell types. The qualitative and quantitative differences between Activin A and TGF-ß pathways are poorly understood. In this report, we investigated Activin A and TGF-ß signaling in an immortalized non-tumorigenic human mammary epithelial cell line. Our data clearly revealed the qualitative and quantitative differences between these signaling pathways at the genomic level. Whereas Activin A signaling is transient and quickly terminated, TGF-ß signaling is more persistent and robust. Activin A regulates only a subset of genes controlled by TGF-ß. Transcriptional responses to Activin A in HME cells are concentration dependent which correlate with the levels of Smad2 phosphorylation. Therefore, ligand concentrations and signaling durations could contribute to the specificity of biological effects of the TGF-ß family of growth factors.

Previous studies of transcriptional responses to Activin A and TGF-ß in human pancreatic tumor cell lines infected with constitutively active Activin (ALK4m) and TGF-ß receptors (ALK5) indicate that overexpression of these two receptors by Adenoviral-mediated gene transfer results in remarkably similar transcriptional responses, suggesting an essential redundancy of these two related ligands (34). Data presented here reveal qualitative and quantitative differences between Activin A and TGF-ß signaling programs when cells were exposed to various concentrations of ligands. It is quite possible that expression of constitutively active receptors elicits persistent signaling and negates the differences observed with ligand treatment. The inability to trigger robust TGF-ß signaling has been observed with other cell types as well. For example, significant transcriptional induction of Smad7 only occurred when more than 700 pM Activin A was used to treat cells (2). What is likely to be responsible for the transient nature of Activin A signaling? The receptors for Activin A appear to be functional in HME cells judging by the robust early induction of genes like Angiopoetin-4 and CTGF. The transient transcriptional responses to Activin A signaling correlates with the intensity of Smad2 phosphorylation and nuclear translocation. A number of mechanisms could account for transient Smad2 phosphorylation. The Activin A receptors could be quickly down regulated by internalization or feedback regulation through association with intracellular or extracellular inhibitors (41). Follistatin is a competitive inhibitor of Activin A and is induced upon Activin A treatment in several cell types (41). In HME cells, Activin A treatment does not affect Follistatin transcription based on our DNA microarray analysis but we cannot rule out the possibility Activin A may enhance Follistatin

expression post-transcriptionally. Alternatively, transient Smad2 phospohrylation could be due to the action of phosphatases that either inactivate the Activin A receptor or Smad2/3. The potential involvement of phosphatases in terminating TGF-ß signaling through dephosphorylation of Smad2 has recently been demonstrated by Hill and coworkers (15). Others have demonstrated that association of specific phosphatases with activated receptors could be involved in down-regulation of the activity of the receptors (15, 36). Regardless of the exact mechanisms, there appear to be fundamental differences in regulating Smad2 phosphorylation between TGF-ß and Activin A.

There have been a number of investigations into the TGF-ß gene induction profile by DNA microarray analysis using a variety of tumor derived cell lines in the literature (5, 34, 43, 45). TGF-ß appears to be able to induce transcription of a number of genes regardless of cell lines employed. For example, PAI-1, Smad7, CTGF, TIEG, BHLHB2/DEC-1 and JunB are among them. All of these genes are also strongly regulated by Activin A but exhibit different induction kinetics. Our study also identified Angiopoietin-4 (Ang-4) as a TGF-ß and Activin A early inducible gene. Angiopoietins have been recently recognized as important growth factors for vascular endothelial cells through interaction with Tie2 receptors (39, 40). Strong transcriptional induction of Ang-4 by TGF-ß and Activin A in HME cells suggests that these two ligands could influence epithelial-endothelial cell interactions during mammary gland development through modulation of the levels of Ang-4. It is also tempting to speculate that the ability of TGF-ß or Activin A to stimulate tumor

metastasis could be attributed in part to transcriptional induction of angiogenic factors like Ang-4.

A comprehensive understanding of TGF-ß signaling specificity requires a genomic scale examination of TGF-ß responsive profiles and systematic identification and evaluation of potential regulatory elements in the promoter regions of responsive genes that mediate TGF-ß induction. Genome-wide transcriptional profiling analysis yielded unprecedented insights into the TGF-ß and Activin A signaling pathways. Identification of the target genes of these pathways will help our understanding of how the TGF-ß family of ligands influences various biological processes. The overriding question as to what determines whether a gene will be subject to TGF-ß regulation still remains. Based on decades of research on transcriptional regulation, it is reasonable to assume that recruitment of specific transcription factors to their cognate binding sites in the regulatory region of the responsive genes is likely to be responsible for the specificity of gene induction. This would predict that transcription factor binding sites that are involved in mediating TGF-ß responsiveness should occur exclusively or at least more frequently in the TGF-ß target genes vs. control genes. TGF-ß responsive sequence elements are often identified using "promoter bashing" experiments. Such an analysis often yields a few informative elements in a gene of interest. It is not clear whether there exists a unique set of transcription factor binding sites shared by many TGF-ß responsiveness genes. Our computational analysis of the transcription factor binding sites in the promoter regions of TGF-ß responsive genes indicates that out of more than 6000 experimentally characterized transcription factor binding sites in the TFD, there is only a limited number of transcription factor binding sites highly enriched

in the TGF-ß responsive genes. These bindings sites are highly conserved and exist in at least two of the three genomes we investigated. The most abundant binding sites are Sp1/AP2, AP-1, CRE, NF-κB, CAC/EKLF, GATA-1, Oct-1 and Ets. The involvement of Sp1, AP-1, CRE and NF-κB in TGF-ß responsiveness has been extensively documented in the literature (reviewed in (7)). Interestingly, the corresponding transcription factors associated with these sites have all been shown to be coactivators of Smads through direct physical interactions (16, 21-24, 27, 29, 30, 33, 47). It will be interesting to test whether EKLF, Oct-1, Ets and other informative transcription factors revealed from our analysis bind Smad2/3 directly to activate transcription of a given TGF-ß responsive gene.

Smads play a central role in transcriptional regulation of TGF-ß responsive genes. Core Smad-binding elements (GTCT or AGAC) have been shown to be necessary but often not sufficient to enable TGF-ß responsiveness. We searched for the occurrence of the Smad-binding elements (SBE including, 5'-GTCT-3, 5'-AGAC-3', 5'-CAGA-3', 5'-GTCTG-3' 5'-GTCTGGAC-3' and 5'-GTCTAGAC-3') in the promoter regions of the TGF-ß responsive and nonresponsive genes. Consistent with the previous report, we found no significant difference between these two groups of genes in SBE occurrence except that the occurrence of 5'-GTCTGGAC-3' is 1.68 fold more in the responsive genes. It has been suggested that tandem or inverted GTCT repeats with 0-3 spacer lengths were present specifically in proximal promoter regions of Smad3-dependent immediate early genes (44). We searched our data set using tandem or inverted GTCT repeats with variable spacers between them. The occurrence of these repeats was found not to be statistically significant between these two groups in our

data set when either the proximal or the 10 kb regulatory region was searched. This discrepancy may stem from the different data set used in the search. It is possible that only Smad3-dependent immediate early genes harbor significant GTCT repeats. The paucity of these repeat elements in the promoter region of TGF-ß responsive and control genes makes it difficult to assess the statistical significance of their occurrence.

Our experimental data and computational analysis of the regulatory regions of TGF-ß responsive genes favor the model that multiple transcription factor binding sites are responsible for TGF-ß induction. This observation is consistent with the notion that transcription factors associated with these sequence elements are more likely to partner with Smads to trigger transcription of these target genes. Thus, delineation of transcription factor binding sites enriched in TGF-ß responsive genes could be informative for identifying additional Smad partners in TGF-ß signal transduction pathways.

Acknowledgments

We thank Drs. Ahn, Massague, Moustakis, ten Dijke, Heldin and Loeb for the generous supply of reagents used in this study. In addition, we thank Rob Knight, Natalie Ahn and members of Liu laboratory for helpful discussions. We also want to thank Kristen Barthel and Richard Erickson for critically reading the manuscripts. Tom H Cheung was supported by a predoctoral training grant from NHLBI (5T32HL07851). This work is supported by a grant from US Army Breast Cancer Research Program (DAMD17-02-1-0350) to Xuedong Liu.

Figure Legends

Figure 1. Comparison of time-dependent expression profiles on human mammary epithelial cells (HME) treated with either TGF-ß (A) or Activin A (B) for indicated times. All data points represent combined values across four replicate arrays. Log ratios colored blue are unchanged (not significantly different than 0), those shown in red are up-regulated (significantly greater than 0, p value <0.01) and those in green are down-regulated (significantly less than 0, p value <0.01). The data were processed by the Resolver software (Rosetta Biosoft) and plotted - Log(10) Ratio vs. Log(10) Intensity.

Figure 2. Comparison of Activin A and TGF-ß responses in HME cells. Shown are three plots comparing differentially expressed genes in response to Activin A and TGF-ß after 2(A), 4(B) or 8 hours (C). Cyan datapoints represent genes that are common signatures (p<0.01) (i.e., upregulated in response to both ligands or downregulated in response to both). The magenta colored datapoints represent anti-correlated signatures.

Figure 3. A list of representative Activin A and TGF-ß responsive genes. Activin A and TGF-ß responsive genes correlate with the time of the treatment in HME cell line. 35 representative genes were selected from a list of 217 genes in which they are differentially expressed by > 2 fold in at least one time point. A Heat map of the selected genes is shown. Data are presented in two sets of three time points each (2, 4

and 8 hours compared to the untreated), with the leftmost set representing Activin A treatment and the rightmost set representing TGF-ß treatment. Genes that have not been previously identified as Activin A and TGF-ß responsive genes are indicated by "*".

Figure 4. (a) Northern blot analysis of selected Activin A and TGF-ß differentially regulated genes from microarray analysis. HME cells were treated with Activin A or TGF-ß for indicated times. Total RNA was harvested and blotted onto a Nytran membrane. The blot was hybridized with indicated radiolabled probes. (b) The duration of Activin A and TGF-ß transcriptional response. Shown is a plot comparing the numbers of signature genes responding to Activin A and TGF-ß after 2, 4 and 8 hours as determined by Resolver.

Figure 5. Ligand-induced Smad2 phosphorylation in HME cells treated with various concentrations of Activin A and TGF-ß. Western blot analyses were performed on HME cell lysates exposed to the indicated concentrations of ligands for the indicated times using a phospho-Smad2 monoclonal antibody.

Figure 6. (a) Dose-dependent transcriptional response to Activin A in HME cells. A list of representative Activin A dose-responsive genes is shown. HME cells were treated with increasing concentrations of Activin A for 4 hr. Total RNA was isolated and profiled as described in Figure 1. (b) Northern blot analysis of a representative Activin A dose responsive gene HEF1.

Figure 7. Comparison of Gene Ontology in HME cells upon Activin A and TGF-ß treatment. GO categories were assigned to each of the genes found to be differentially

expressed in response to Activin A and TGF-ß treatment. These categories are grouped and plotted on a bar graph. The percentages of genes in each GO category are shown on the y-axis while different GO categories are shown on the x-axis. In (a), genes that are up-regulated upon Activin A treatment (2 hour time point) are compared to TGF-ß treatment (2 hour time point). In (b), genes that are down-regulated upon Activin A treatment (2 hour time point) are compared to TGF-ß treatment (2 hour time point).

Figure 8. Computational analysis of the distribution of transcription factor binding sites within the regulatory regions of TGF-ß responsive genes. (a) Shown is a representative two-dimensional heatmap displaying the correlation between a few representative binding sites enriched in TGF-ß responsive genes and a number of representative differentially regulated genes sorted in descending order (from most induced to most repressed). The top row indicates approximate fold changes of these genes. Each row describes a specific transcription factor binding site that was found to exist exclusively (NA) or statistically more frequently in TGF-ß regulated genes. The presence of such a transcription factor binding site in TGF-ß responsive genes is designated as a colored square in the gene name column. The color code of the square is indicated in the figure. The columns on the left present all the detailed computational data associated with the transcription factor binding sites. (b) Two-dimensional heatmap displaying the correlation between all the transcription factor binding sites enriched in TGF-ß responsive genes and 108 differentially regulated genes sorted in descending order (from most induced to most repressed with changes at least 1.8 fold in either direction). The complete dataset for this graph is presented in Suppl. 4.

Figure 9. Experimental validation of minimal TGF-ß responsive regulatory elements in CYR61 and HS3ST2 promoter region. (A) Mink lung cells (CCL64) were transfected with reporter constructs indicated. p3TP-Lux reporter gene was used as the positive control. A schematic representation of the CYR61 promoter region identified by computational analysis was shown above the graph with known transcription factor binding sites highlighted. The fold induction by TGF-ß is indicated and error bars represent standard deviations from triplicate determinations. (B) Hep3B cells was transfected with reporter constructs containing one, two or three copies of the putative minimal TGF-ß responsive element from the promoter region of the HS3ST2 gene. p3APP-Lux was used as the positive control in this experiment.

Table1. Summary of the changes of gene expression profiles determined by DNA microarray array analysis.

Supplementary data

Suppl. 1. The dataset of time-dependent DNA microarray experiments in HME cells treated with 100 pM Activin A or TGF-ß (1.8 fold up or down-regulated).

Suppl. 2. Clustering genes exhibiting similar expression kinetics in response to ligand stimulation using the Spotfire software.

Suppl. 3. Detailed summary of gene clusters.

Suppl. 4. Dataset used for construction of Figure 8b.

References

1. **Attisano, L., and J. L. Wrana.** 2002. Signal transduction by the TGF-beta superfamily. Science **296:**1646-7.
2. **Bilezikjian, L. M., A. Z. Corrigan, A. L. Blount, Y. Chen, and W. W. Vale.** 2001. Regulation and actions of Smad7 in the modulation of activin, inhibin, and transforming growth factor-beta signaling in anterior pituitary cells. Endocrinology **142:**1065-72.
3. **Brummel, T., S. Abdollah, T. E. Haerry, M. J. Shimell, J. Merriam, L. Raftery, J. L. Wrana, and M. B. O'Connor.** 1999. The Drosophila activin receptor baboon signals through dSmad2 and controls cell proliferation but not patterning during larval development. Genes Dev **13:**98-111.
4. **Chang, H., C. W. Brown, and M. M. Matzuk.** 2002. Genetic analysis of the mammalian transforming growth factor-beta superfamily. Endocr Rev **23:**787-823.
5. **Chen, C. R., Y. Kang, and J. Massague.** 2001. Defective repression of c-myc in breast cancer cells: A loss at the core of the transforming growth factor beta growth arrest program. Proc Natl Acad Sci U S A **98:**992-9.
6. **Chen, Y. G., H. M. Lui, S. L. Lin, J. M. Lee, and S. Y. Ying.** 2002. Regulation of cell proliferation, apoptosis, and carcinogenesis by activin. Exp Biol Med (Maywood) **227:**75-87.
7. **Derynck, R., and Y. E. Zhang.** 2003. Smad-dependent and Smad-independent pathways in TGF-beta family signalling. Nature **425:**577-84.
8. **Di Guglielmo, G. M., C. Le Roy, A. F. Goodfellow, and J. L. Wrana.** 2003. Distinct endocytic pathways regulate TGF-beta receptor signalling and turnover. Nat Cell Biol **5:**410-21.
9. **Dyson, S., and J. B. Gurdon.** 1998. The interpretation of position in a morphogen gradient as revealed by occupancy of activin receptors. Cell **93:**557-68.
10. **Fautsch, M. P., A. Vrabel, M. Subramaniam, T. E. Hefferen, T. C. Spelsberg, and E. D. Wieben.** 1998. TGFbeta-inducible early gene (TIEG) also codes for early growth response alpha (EGRalpha): evidence of multiple transcripts from alternate promoters. Genomics **51:**408-16.
11. **Gurdon, J. B., P. Harger, A. Mitchell, and P. Lemaire.** 1994. Activin signalling and response to a morphogen gradient. Nature **371:**487-92.

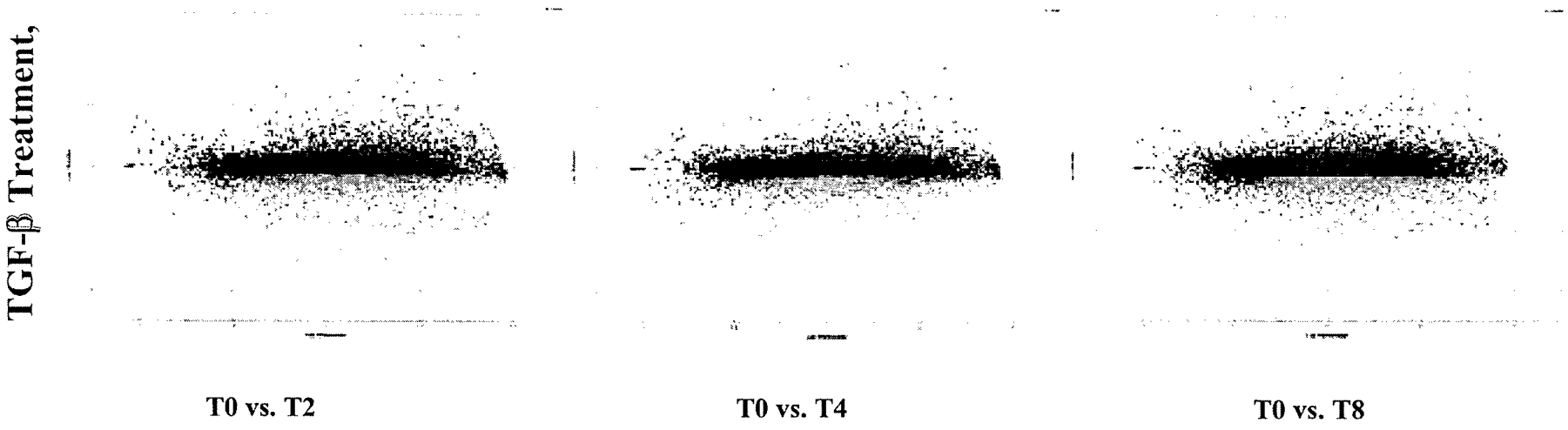
12. **Hamady, M., Cheung, T. H. and R. Knight.** Unpublished results. GO-Getter: Using the Gene Ontology to Detect Differences Between Expression Data Sets.
13. **Heldin, C. H., K. Miyazono, and P. ten Dijke.** 1997. TGF-beta signalling from cell membrane to nucleus through SMAD proteins. Nature **390:**465-71.
14. **Hua, X., X. Liu, D. O. Ansari, and H. F. Lodish.** 1998. Synergistic cooperation of TFE3 and smad proteins in TGF-beta-induced transcription of the plasminogen activator inhibitor-1 gene. Genes Dev **12:**3084-95.
15. **Inman, G. J., F. J. Nicolas, and C. S. Hill.** 2002. Nucleocytoplasmic shuttling of Smads 2, 3, and 4 permits sensing of TGF-beta receptor activity. Mol Cell **10:**283-94.
16. **Jono, H., T. Shuto, H. Xu, H. Kai, D. J. Lim, J. R. Gum, Jr., Y. S. Kim, S. Yamaoka, X. H. Feng, and J. D. Li.** 2002. Transforming growth factor-beta - Smad signaling pathway cooperates with NF-kappa B to mediate nontypeable Haemophilus influenzae-induced MUC2 mucin transcription. J Biol Chem **277:**45547-57.
17. **Kang, Y., C. R. Chen, and J. Massague.** 2003. A self-enabling TGFbeta response coupled to stress signaling: Smad engages stress response factor ATF3 for Id1 repression in epithelial cells. Mol Cell **11:**915-26.
18. **Kim, S. J., P. Angel, R. Lafyatis, K. Hattori, K. Y. Kim, M. B. Sporn, M. Karin, and A. B. Roberts.** 1990. Autoinduction of transforming growth factor beta 1 is mediated by the AP-1 complex. Mol Cell Biol **10:**1492-7.
19. **Kim, S. J., F. Denhez, K. Y. Kim, J. T. Holt, M. B. Sporn, and A. B. Roberts.** 1989. Activation of the second promoter of the transforming growth factor-beta 1 gene by transforming growth factor-beta 1 and phorbol ester occurs through the same target sequences. J Biol Chem **264:**19373-8.
20. **Kingsley, D. M.** 1994. The TGF-beta superfamily: new members, new receptors, and new genetic tests of function in different organisms. Genes Dev **8:**133-46.
21. **Kon, A., L. Vindevoghel, D. J. Kouba, Y. Fujimura, J. Uitto, and A. Mauviel.** 1999. Cooperation between SMAD and NF-kappaB in growth factor regulated type VII collagen gene expression. Oncogene **18:**1837-44.
22. **Li, J. M., M. B. Datto, X. Shen, P. P. Hu, Y. Yu, and X. F. Wang.** 1998. Sp1, but not Sp3, functions to mediate promoter activation by TGF-beta through canonical Sp1 binding sites. Nucleic Acids Res **26:**2449-56.
23. **Li, J. M., M. A. Nichols, S. Chandrasekharan, Y. Xiong, and X. F. Wang.** 1995. Transforming growth factor beta activates the promoter of cyclin-dependent kinase inhibitor p15INK4B through an Sp1 consensus site. J Biol Chem **270:**26750-3.
24. **Li, J. M., X. Shen, P. P. Hu, and X. F. Wang.** 1998. Transforming growth factor beta stimulates the human immunodeficiency virus 1 enhancer and requires NF-kappaB activity. Mol Cell Biol **18:**110-21.
25. **Liu, Q. Y., B. Niranjan, P. Gomes, J. J. Gomm, D. Davies, R. C. Coombes, and L. Buluwela.** 1996. Inhibitory effects of activin on the growth and morpholgenesis of primary and transformed mammary epithelial cells. Cancer Res **56:**1155-63.
26. **Liu, X., Y. Sun, S. N. Constantinescu, E. Karam, R. A. Weinberg, and H. F. Lodish.** 1997. Transforming growth factor beta-induced phosphorylation of

Smad3 is required for growth inhibition and transcriptional induction in epithelial cells. Proc Natl Acad Sci U S A **94:**10669-74.

27. **Lopez-Rovira, T., E. Chalaux, J. L. Rosa, R. Bartrons, and F. Ventura.** 2000. Interaction and functional cooperation of NF-kappa B with Smads. Transcriptional regulation of the junB promoter. J Biol Chem **275:**28937-46.
28. **Massague, J.** 1998. TGF-beta signal transduction. Annu Rev Biochem **67:**753-91.
29. **Moustakas, A., and D. Kardassis.** 1998. Regulation of the human p21/WAF1/Cip1 promoter in hepatic cells by functional interactions between Sp1 and Smad family members. Proc Natl Acad Sci U S A **95:**6733-8.
30. **Pardali, K., A. Kurisaki, A. Moren, P. ten Dijke, D. Kardassis, and A. Moustakas.** 2000. Role of Smad proteins and transcription factor Sp1 in p21(Waf1/Cip1) regulation by transforming growth factor-beta. J Biol Chem **275:**29244-56.
31. **Piek, E., C. H. Heldin, and P. Ten Dijke.** 1999. Specificity, diversity, and regulation in TGF-beta superfamily signaling. Faseb J **13:**2105-24.
32. **Piek, E., A. Moustakas, A. Kurisaki, C. H. Heldin, and P. ten Dijke.** 1999. TGF-(beta) type I receptor/ALK-5 and Smad proteins mediate epithelial to mesenchymal transdifferentiation in NMuMG breast epithelial cells. J Cell Sci **112 (Pt 24):**4557-68.
33. **Rasty, S., P. Thatikunta, J. Gordon, K. Khalili, S. Amini, and J. C. Glorioso.** 1996. Human immunodeficiency virus tat gene transfer to the murine central nervous system using a replication-defective herpes simplex virus vector stimulates transforming growth factor beta 1 gene expression. Proc Natl Acad Sci U S A **93:**6073-8.
34. **Ryu, B., and S. E. Kern.** 2003. The essential similarity of TGFbeta and activin receptor transcriptional responses in cancer cells. Cancer Biol Ther **2:**164-70.
35. **Schnepp, R., and X. Hua.** 2003. A tumor-suppressing duo: TGFbeta and activin modulate a similar transcriptome. Cancer Biol Ther **2:**171-2.
36. **Shi, W., C. Sun, B. He, W. Xiong, X. Shi, D. Yao, and X. Cao.** 2004. GADD34-PP1c recruited by Smad7 dephosphorylates TGF{beta} type I receptor. J Cell Biol **164:**291-300.
37. **Shi, Y., and J. Massague.** 2003. Mechanisms of TGF-beta signaling from cell membrane to the nucleus. Cell **113:**685-700.
38. **Tang, W., L. Yang, Y. C. Yang, S. X. Leng, and J. A. Elias.** 1998. Transforming growth factor-beta stimulates interleukin-11 transcription via complex activating protein-1-dependent pathways. J Biol Chem **273:**5506-13.
39. **Tsigkos, S., M. Koutsilieris, and A. Papapetropoulos.** 2003. Angiopoietins in angiogenesis and beyond. Expert Opin Investig Drugs **12:**933-41.
40. **Valenzuela, D. M., J. A. Griffiths, J. Rojas, T. H. Aldrich, P. F. Jones, H. Zhou, J. McClain, N. G. Copeland, D. J. Gilbert, N. A. Jenkins, T. Huang, N. Papadopoulos, P. C. Maisonpierre, S. Davis, and G. D. Yancopoulos.** 1999. Angiopoietins 3 and 4: diverging gene counterparts in mice and humans. Proc Natl Acad Sci U S A **96:**1904-9.

41. **Welt, C., Y. Sidis, H. Keutmann, and A. Schneyer.** 2002. Activins, inhibins, and follistatins: from endocrinology to signaling. A paradigm for the new millennium. Exp Biol Med (Maywood) **227**:724-52.
42. **Wrana, J. L., J. Carcamo, L. Attisano, S. Cheifetz, A. Zentella, F. Lopez-Casillas, and J. Massague.** 1992. The type II TGF-beta receptor signals diverse responses in cooperation with the type I receptor. Cold Spring Harb Symp Quant Biol **57**:81-6.
43. **Xie, L., B. K. Law, M. E. Aakre, M. Edgerton, Y. Shyr, N. A. Bhowmick, and H. L. Moses.** 2003. Transforming growth factor beta-regulated gene expression in a mouse mammary gland epithelial cell line. Breast Cancer Res **5**:R187-98.
44. **Yang, Y. C., E. Piek, J. Zavadil, D. Liang, D. Xie, J. Heyer, P. Pavlidis, R. Kucherlapati, A. B. Roberts, and E. P. Bottinger.** 2003. Hierarchical model of gene regulation by transforming growth factor beta. Proc Natl Acad Sci U S A **100**:10269-74.
45. **Zavadil, J., M. Bitzer, D. Liang, Y. C. Yang, A. Massimi, S. Kneitz, E. Piek, and E. P. Bottinger.** 2001. Genetic programs of epithelial cell plasticity directed by transforming growth factor-beta. Proc Natl Acad Sci U S A **98**:6686-91.
46. **Zawel, L., J. L. Dai, P. Buckhaults, S. Zhou, K. W. Kinzler, B. Vogelstein, and S. E. Kern.** 1998. Human Smad3 and Smad4 are sequence-specific transcription activators. Mol Cell **1**:611-7.
47. **Zhang, Y., X. H. Feng, and R. Derynck.** 1998. Smad3 and Smad4 cooperate with c-Jun/c-Fos to mediate TGF-beta-induced transcription. Nature **394**:909-13.

A.


TGF-β Treatment,
T0 vs. T2
T0 vs. T4
T0 vs. T8

B.

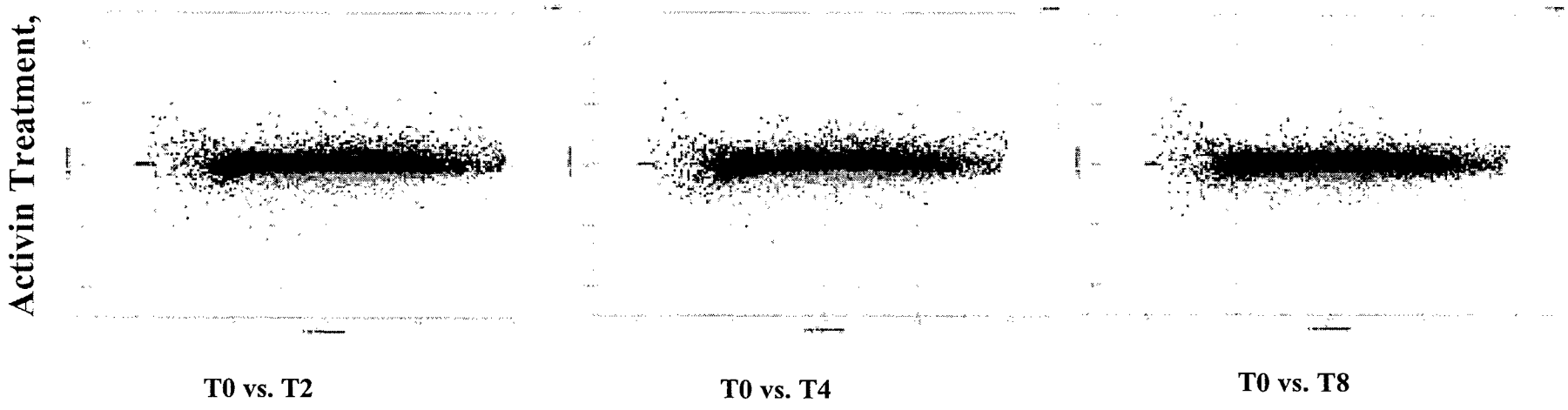

Activin Treatment,
T0 vs. T2
T0 vs. T4
T0 vs. T8

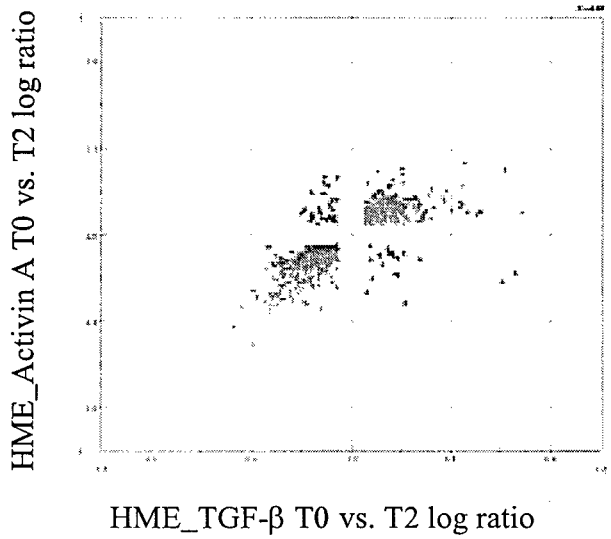
Figure 1

Table 1

	Activin A		TGF-beta		Common	
	Up	**Down**	**Up**	**Down**	**Up**	**Down**
2 hour	22	61	126	114	12	10
4 hour	24	40	87	69	6	20
8 hour	14	24	92	151	5	13

A.

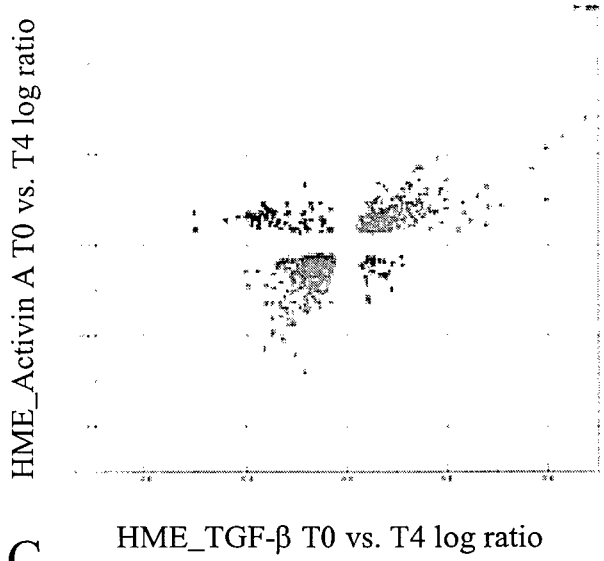
Actinvin A vs. TGF-β



Correlated genes

Anti-correlated genes

B.



C.

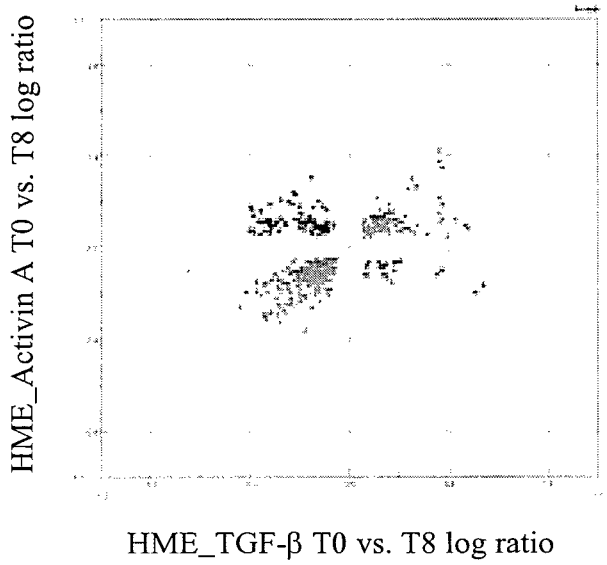
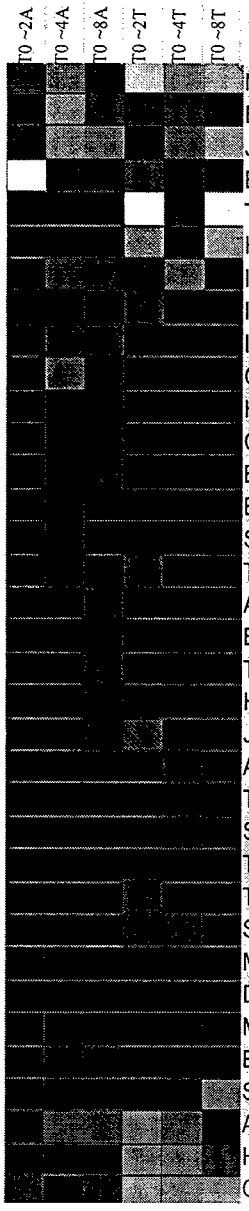


Figure 2



	Gene Name	Gene Description	Activin A			TGF-β		
			T0vsT2 A	T0vsT4 A	T0vsT8 A	T0vsT2 B	T0vsT4 B	T0vsT8 B
	MYC	V-myc avian myelocytomatosis viral oncogene homolog	-1.3	-1.9	-1.1	-2.5	-1.9	-2.3
	DUSP6	Dual specificity phosphatase 6	1.5	-2.2	-1.3	-1.1	-1.3	-1.0
	JUN	V-jun homolog	-1.1	-2.1	-1.8	2.0	-1.6	-2.5
*	EGR1	Early growth response 1	-3.7	-1.2	-1.2	-1.5	-1.1	1.0
	TIMP3	Tissue inhibitor of metalloproteinases 3	1.4	1.5	1.1	-4.1	-1.1	-3.5
	HDAC10	Member of the histone deacetylase family	1.5	1.7	1.5	-2.2	1.1	-2.3
	ID1	Inhibitor of DNA binding 1	1.1	-1.5	-1.3	3.6	-1.9	1.2
	ID3	Inhibitor of DNA binding 3	-1.0	-1.1	1.0	4.1	1.2	2.0
	IER3	Immediate early response 3	1.1	-1.5	-1.4	2.4	1.5	1.5
	GRO1	Growth related oncogene (melanoma growth stimulating activity)	1.5	-2.1	-1.1	2.5	1.6	1.3
	HRY	Hairy Drosophila homolog	1.2	-1.1	-1.1	2.1	1.0	1.4
	CYR61	Cysteine-rich angiogenic inducer 61	1.3	-1.0	-1.0	2.3	1.7	2.0
	BMP2	Bone morphogenetic protein 2	1.5	-1.1	-1.2	2.9	2.5	1.4
*	EMP1	Epithelial membrane protein-1 (tumor-associated membrane protein)	1.7	-1.1	1.1	2.1	1.3	1.2
	SNAI2	Snail homolog 2,	1.6	-1.1	1.1	2.5	1.7	1.9
*	TBX3	T-box 3 (ulnar mammary syndrome)	1.1	-1.0	1.3	3.9	2.7	2.7
	ATF3	Activating transcription factor 3	1.2	1.0	-1.3	1.7	2.1	1.4
	BMP6	Bone morphogenetic protein 6	1.1	1.2	-1.0	1.4	2.2	2.4
	TMEPAI	Transmembrane (prostate androgen induced RNA)	1.2	1.2	-1.2	2.2	2.9	2.6
*	HSPC056	Protein containing eight armadillo or beta-catenin repeats	2.1	1.9	-1.1	1.9	1.6	1.3
	JUNB	Jun B proto-oncogene	1.4	1.1	-1.1	4.1	2.0	2.5
*	APP	Amyloid beta (A4) precursor protein	1.9	2.2	2.5	1.8	-1.1	2.2
	TGFBI	Transforming growth factor beta induced	1.0	1.2	1.2	1.4	2.0	3.0
	STK38L	Protein with strong similarity to human STK38	1.2	1.1	1.0	1.7	3.0	2.7
	TFPI2	Tissue factor pathway inhibitor 2	1.4	1.3	1.0	2.9	1.9	2.0
	TIEG	TGFB inducible early growth response	1.8	1.0	1.0	3.6	1.6	1.8
	SKIL	SKI-like oncoprotein	1.2	1.3	1.2	3.2	3.4	2.6
*	MAGOH	Mago-nashi homolog	1.7	2.2	1.5	1.1	1.1	1.7
*	DUSP8	Dual specificity phosphatase 8	1.6	1.5	2.0	2.0	1.9	1.4
*	MAGEB4	Melanoma antigen family B 4	2.0	2.9	2.5	1.7	1.6	1.5
*	ENPP2	Ectonucleotide pyrophosphatase/phosphodiesterase 2 (autotaxin)	2.3	3.0	2.8	1.6	1.9	1.8
	SERPINE1	Plasminogen activator inhibitor 1	1.5	2.1	1.1	2.2	3.0	4.7
*	ANGPTL4	Angiopoietin-like 4	2.9	-1.6	-1.4	6.8	4.4	3.2
	HEF1	Enhancer of filamentation 1	1.6	1.3	1.1	6.4	4.9	3.9
	CTGF	Connective tissue growth factor	3.6	1.1	-1.1	7.8	4.8	5.1

Figure 3

4 Fold Down or more

5 Fold Up or more

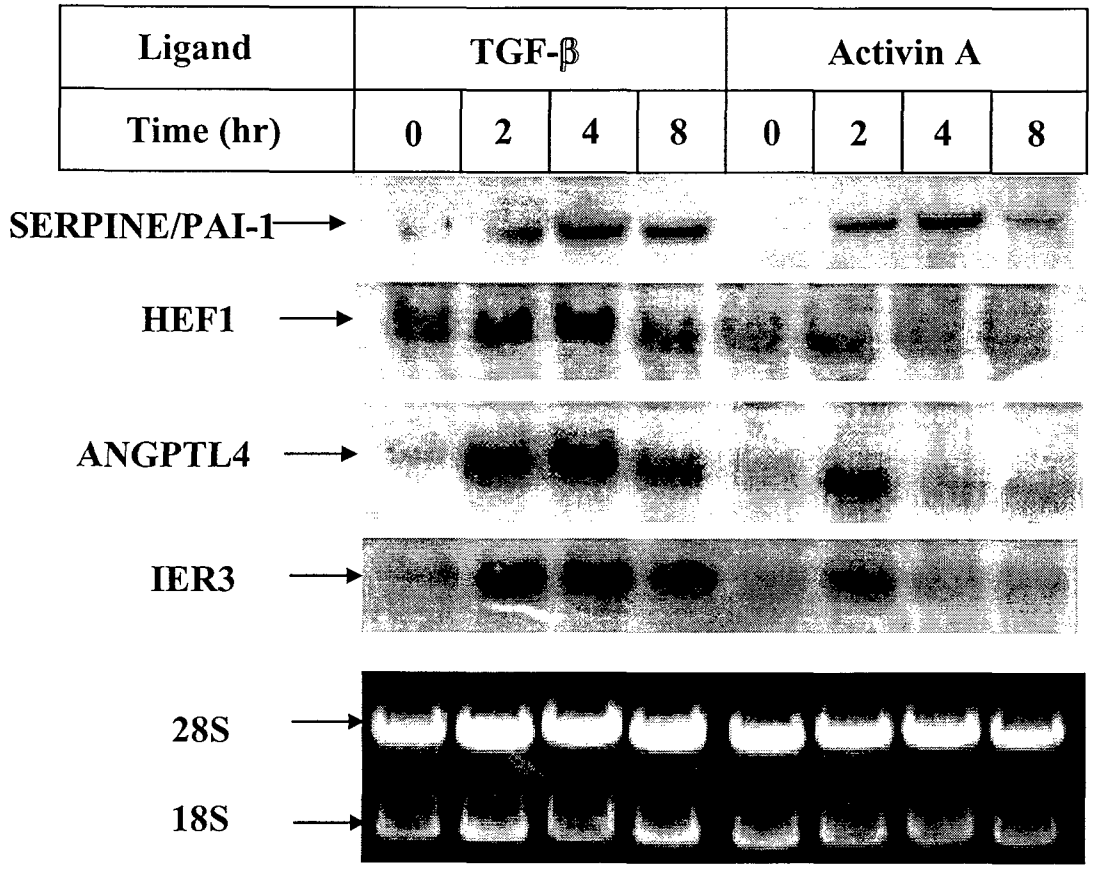

Ligand
TGF-β
Activin A
Time (hr)
0
2
4
8
0
2
4
8
SERPINE/PAI-1
HEF1
ANGPTL4
IER3
28S
18S

Figure 4a

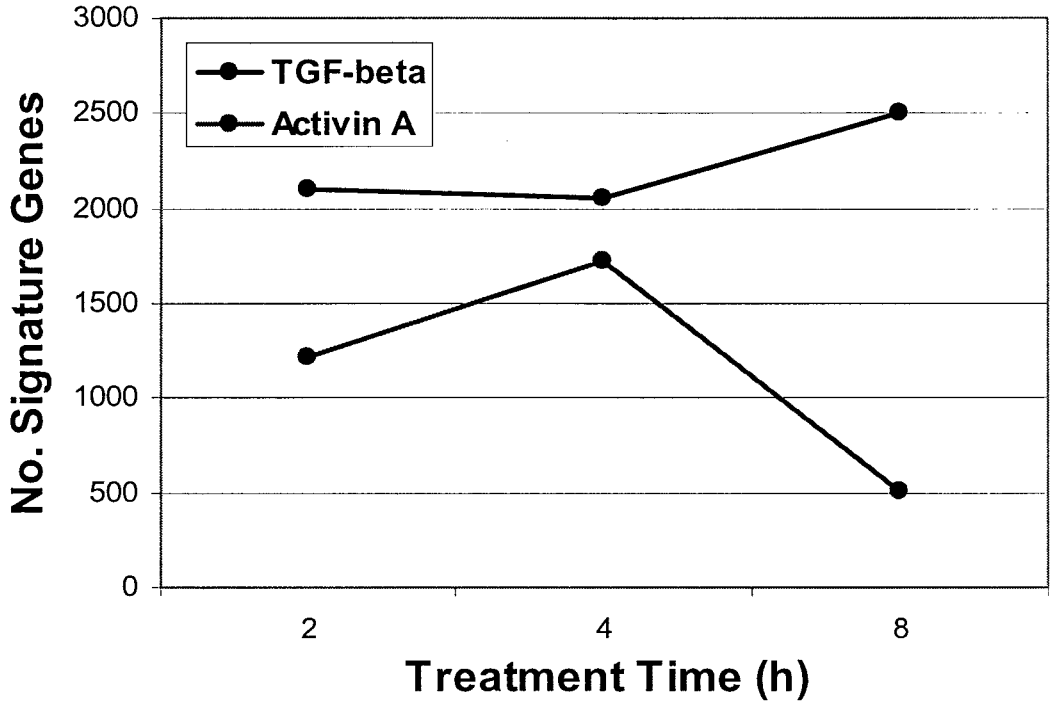

3000
2500
2000
1500
1000
500
0
No. Signature Genes
TGF-beta
Activin A
2
4
8
Treatment Time (h)

Figure 4b

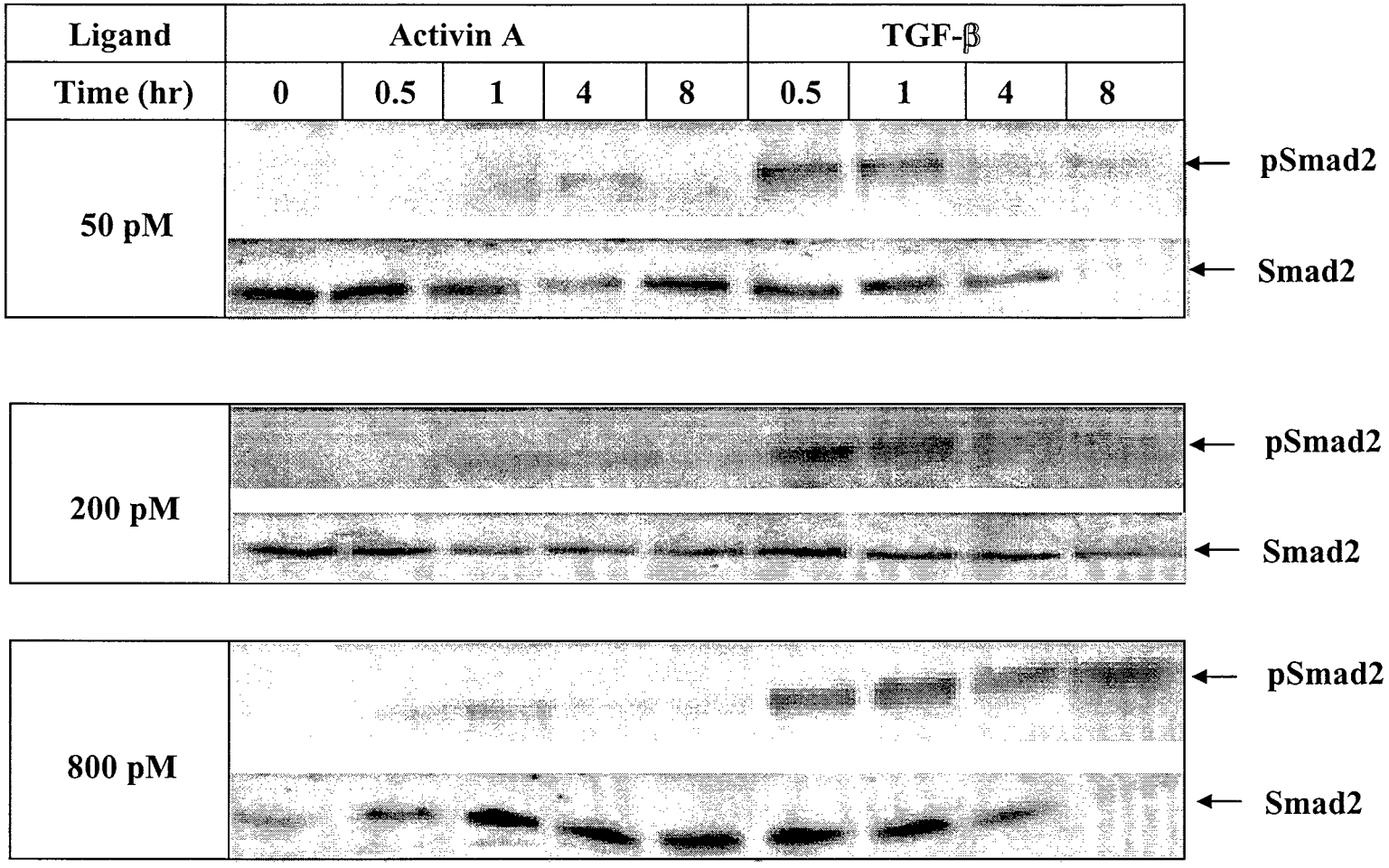

Ligand
Activin A
TGF-β
Time (hr)
0
0.5
1
4
8
0.5
1
4
8
50 pM
200 pM
800 pM
pSmad2
Smad2
pSmad2
Smad2
pSmad2
Smad2

Figure 5

Gene Name	Gene Description	50pM Activin A	200pM Activin A	800pM Activin A
EGR1	Early growth response 1	-1.39	-2.51	-4.64
I_958789	Protein containing two leucine rich repeats	-2.29	-3.11	-3.06
C1orf24	Protein of unknown function	-1.70	-2.76	-2.81
DSPG3	Dermatan sulfate proteoglycan 3	-1.76	-2.46	-2.76
PLAB	Macrophage inhibitory cytokine 1	-1.47	-2.35	-2.49
FLJ13189	Protein of unknown function	-2.29	-2.34	-2.48
POU4F3	POU domain class 4 transcription factor 3	-2.01	-2.26	-2.32
I_1151867	Protein containing two uncharacterized domains DUF6	-2.18	-1.98	-2.34
KLRF1	Killer cell lectin-like receptor subfamily F member 1	-2.19	-1.88	-2.23
SEP2	Septin 6, a protein with high similarity to septin (C. elegans UNC-61)	-2.05	-2.30	-1.93
CD74	CD74 antigen (invariant polypeptide of major histocompatibility complex	-1.93	-1.96	-2.16
MAGEA3	Melanoma antigen family A 3, member of the MAGE family	-1.82	-1.95	-2.09
PRAME	Preferentially expressed antigen of melanoma	-1.62	-2.03	-2.27
RPL10A	Ribosomal protein L10a	-1.73	-1.97	-2.25
KCNA4	Voltage-gated potassium channel shaker-related subfamily member 4	-1.68	-2.01	-2.18
EEF1B2	Eukaryotic translation elongation factor 1 beta 2	-1.75	-1.83	-2.24
ARHGAP6	Rho GTPase activating protein 6	-1.69	-2.15	-1.95
I_932377	Protein with high similarity to mouse Unc5h3	-1.45	-1.97	-2.02
LZ16	Protein containing four ankyrin (Ank) repeats	-1.28	-1.80	-2.09
MGC16332	Protein of unknown function,	-1.71	-1.72	-2.13
RPLP0	Ribosomal protein P0	-1.68	-1.66	-2.12
RPS25	Ribosomal protein S25	-1.67	-1.60	-2.07
RPS17	Ribosomal protein S17	-1.66	-1.58	-2.02
RPL38	Ribosomal protein L38	-1.57	-1.63	-2.19
RPL34	Ribosomal protein L34	-1.50	-1.54	-2.11
I_1000646	Protein with strong similarity to LINE retrotransposable element 1	-1.58	-1.40	-2.03
LRE1	LINE retrotransposable element 1	-1.56	-1.41	-2.03
FKBP10	Protein with strong similarity to FK506 binding protein (mouse Fkbp10)	1.20	-2.04	-1.89
GNAT2	Guanine nucleotide binding protein alpha transducing 2	1.07	1.27	2.10
ID3	Inhibitor of DNA binding 3	1.30	1.32	2.26
HEF1	Enhancer of filamentation 1	1.35	1.38	2.16
HMOX1	Heme oxygenase 1	1.40	1.52	2.12
ANGPTL4	Angiopoietin-like 4	1.34	1.51	2.30
ANGPTL3	Angiopoietin-like 3	1.28	1.98	2.09
CTGF	Connective tissue growth factor	1.98	2.03	2.06
GRO1	Growth related oncogene (melanoma growth stimulating activity)	2.36	2.15	1.89
I_1201840	Member of the ephrin family	1.61	1.78	2.96
TIEG	TGFB inducible early growth response	1.96	2.04	2.78
BHLHB2	Differentiated embryo chondrocyte 1	2.19	2.22	2.89

Activin A 4hr

4 Fold Down or more **3 Fold Up or more**

Figure 6a

Ligand	Activin A			
Conc. (pM)	0	50	200	800

HEF1 →

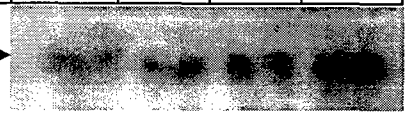


Figure 6b

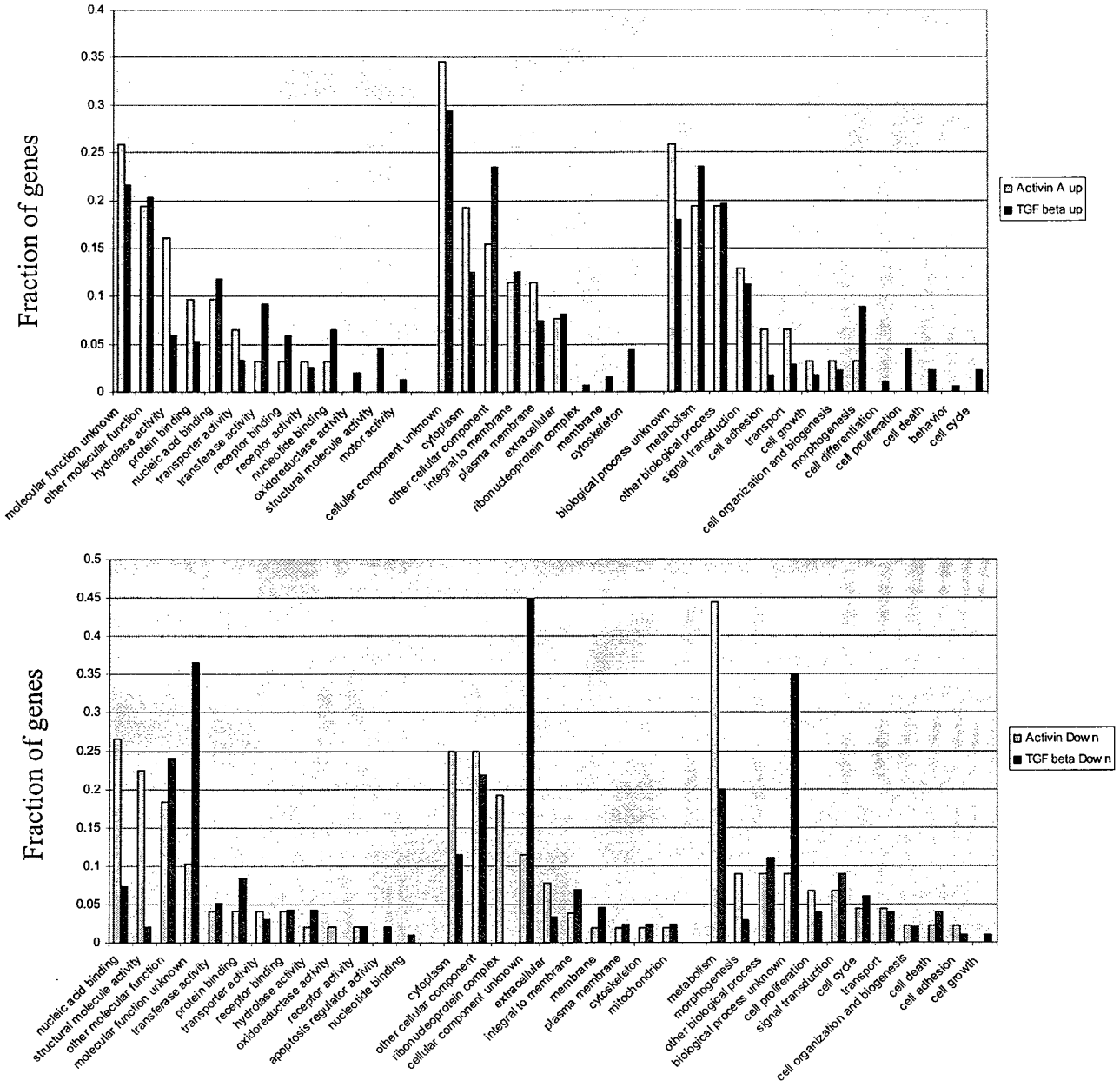

Fraction of genes
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Activin A up
TGF beta up
molecular function unknown
other molecular function
hydrolase activity
protein binding
nucleic acid binding
transporter activity
transferase activity
receptor binding
receptor activity
nucleotide binding
oxidoreductase activity
structural molecule activity
motor activity
cellular component unknown
cytoplasm
other cellular component
integral to membrane
plasma membrane
extracellular
ribonucleoprotein complex
membrane
cytoskeleton
biological process unknown
metabolism
other biological process
signal transduction
cell adhesion
transport
cell growth
cell organization and biogenesis
morphogenesis
cell differentiation
cell proliferation
cell death
behavior
cell cycle
Fraction of genes
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Activin Down
TGF beta Down
nucleic acid binding
structural molecule activity
other molecular function
molecular function unknown
transferase activity
protein binding
transporter activity
receptor binding
hydrolase activity
oxidoreductase activity
receptor activity
apoptosis regulator activity
nucleotide binding
cytoplasm
other cellular component
ribonucleoprotein complex
cellular component unknown
extracellular
integral to membrane
membrane
plasma membrane
cytoskeleton
mitochondrion
metabolism
morphogenesis
other biological process
biological process unknown
cell proliferation
signal transduction
cell cycle
transport
cell organization and biogenesis
cell death
cell adhesion
cell growth

Figure 7

Binding Site	Binding Factor	Binding site Sequence	# of regulated genes containing these elements (T= 108)	# of control genes containing these elements (T= 644)	Fold change (regulated/control)
Sp1/AP2 site	Sp1	CCCCAGGC	3	3	6.0
Sp1/AP2 site	Sp1	GTCCCGCCCA	1	0	NA
TRE site	AP-1	TGACACA	5	8	3.7
TRE site	AP-1	GTGAGTCAG	2	1	12.0
CRE site	ATF/CREB	ACGTCAG	5	6	5.0
CRE site	ATF/CREB	TGACATCA	2	1	12.0
NF-E1.2 site	GATA-1	CTTATCTT	2	0	NA
NFkB site	NF-kB	GGGRATYYCC	2	1	12.0
E-box site	TFE3/USF	GTCACCTG	3	2	9.0
CArG_CS	SRF	ATACCCAAATATGGAAATATTGT	1	0	NA
RTBP1_site	TBF1	TTTAGGG	5	5	6.0
TIE_(2) site	Unknown	GNNTTGGTGA	1	1	6.0

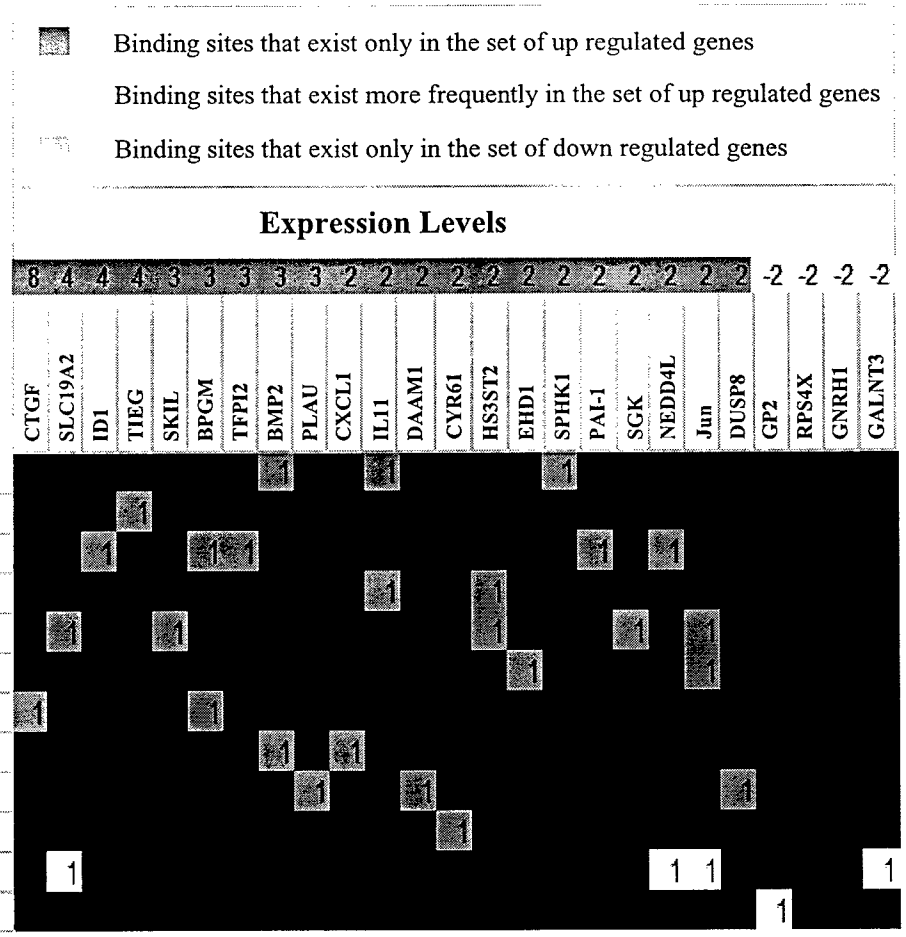


Figure 8a

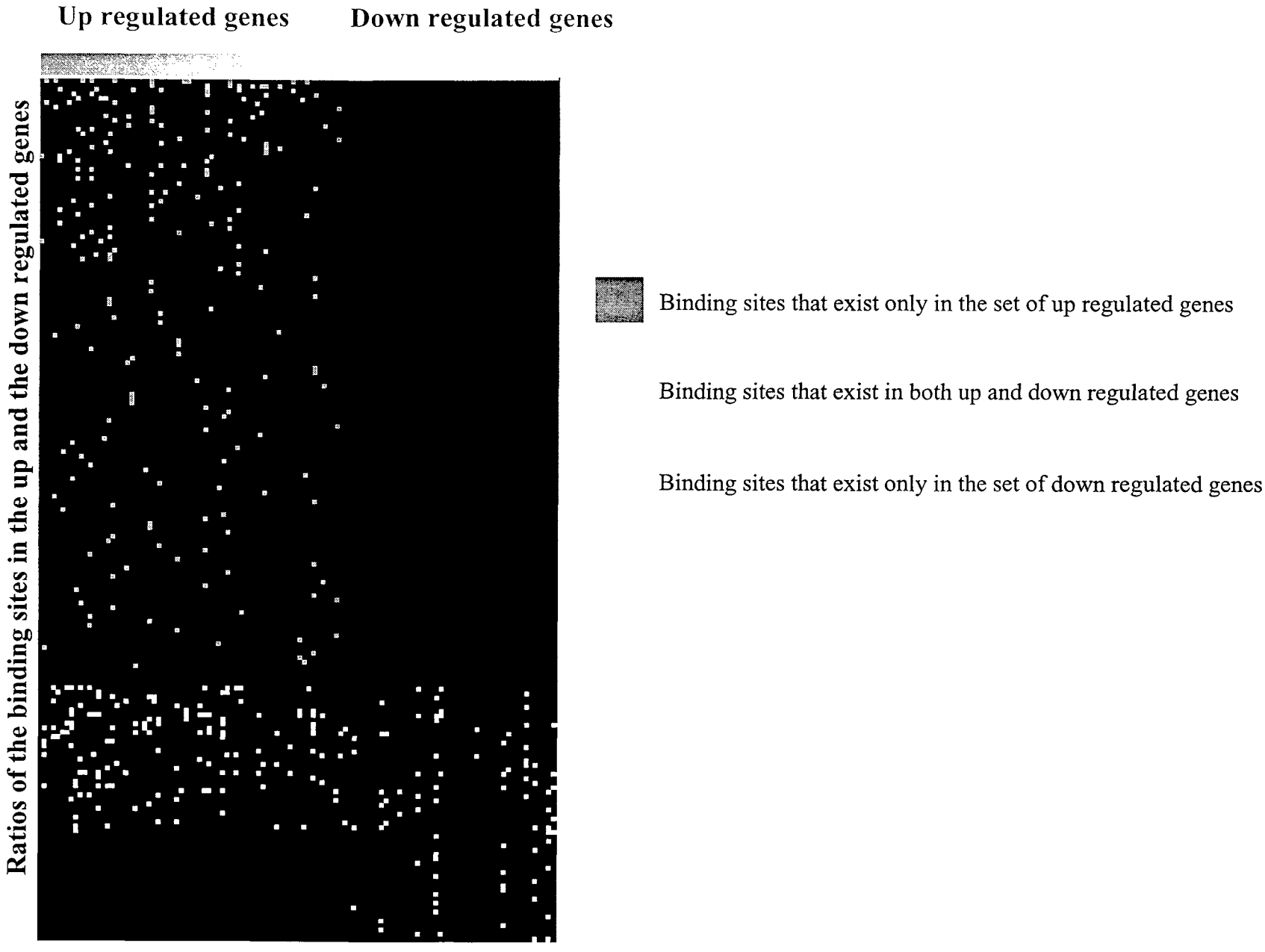

Up regulated genes
Down regulated genes
Ratios of the binding sites in the up and the down regulated genes
Binding sites that exist only in the set of up regulated genes
Binding sites that exist in both up and down regulated genes
Binding sites that exist only in the set of down regulated genes

Figure 8b

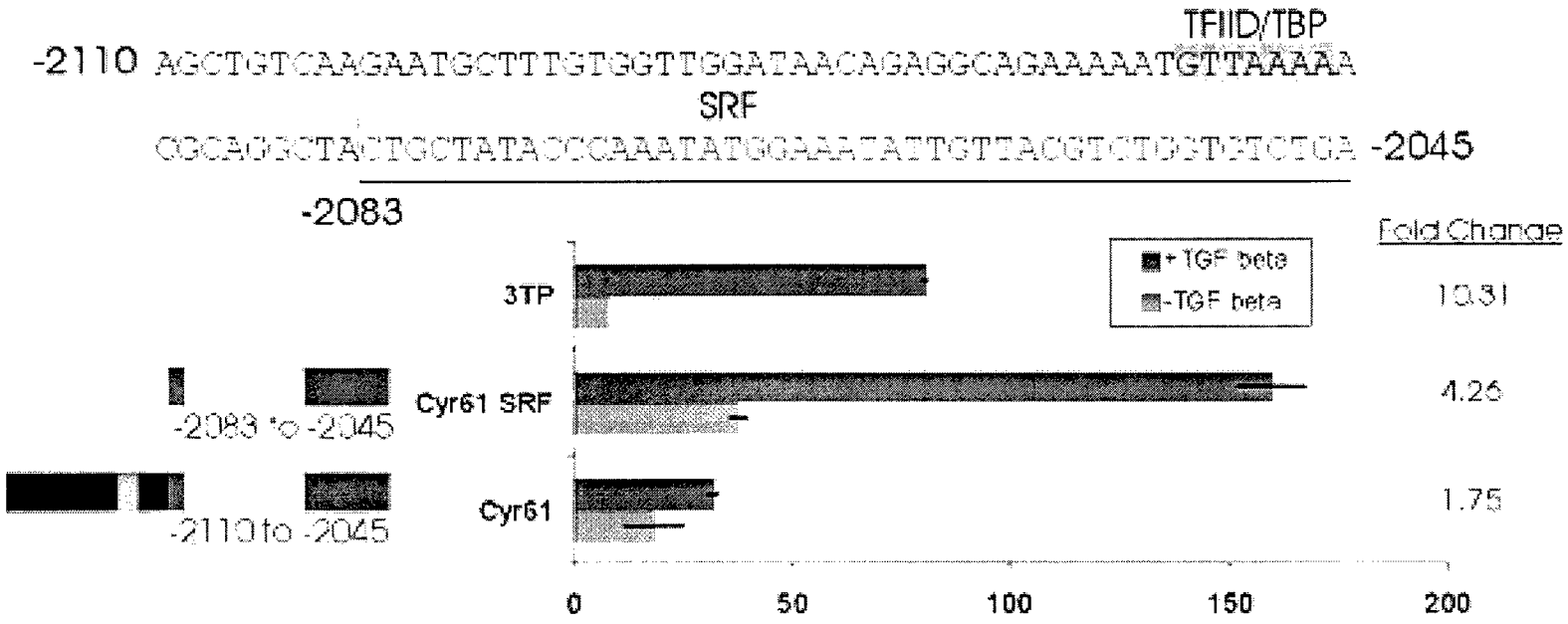

TFIID/TBP
-2110 AGCTGTCAAGAATGCTTTGTGGTTGGATAACAGAGGCAGAAAAATGTTAAAAA
SRF
CGCAGGCTACTGCTATACCCAAATATGGAAATATTGTTACGTCTGGTCTCTGA -2045
-2083
Fold Change
+TGF beta
-TGF beta
3TP
10.31
Cyr61 SRF
4.26
-2083 to -2045
Cyr61
1.75
-2110 to -2045
0
50
100
150
200

Figure 9a

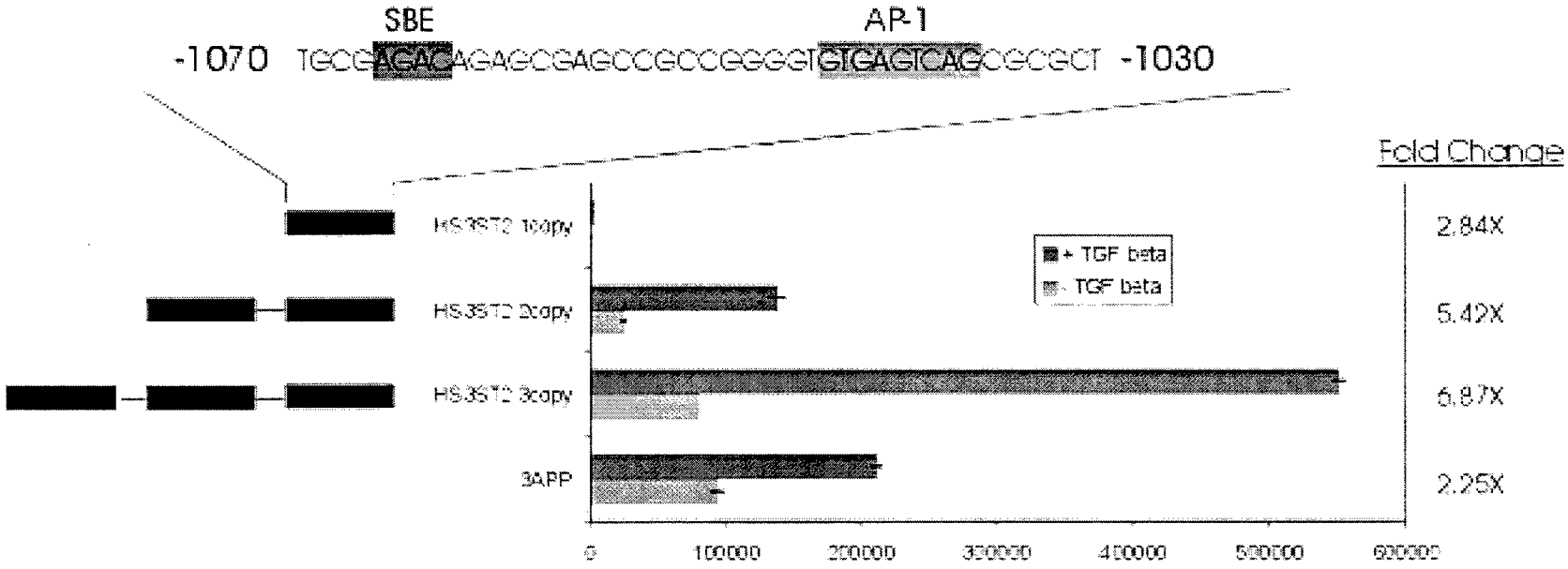
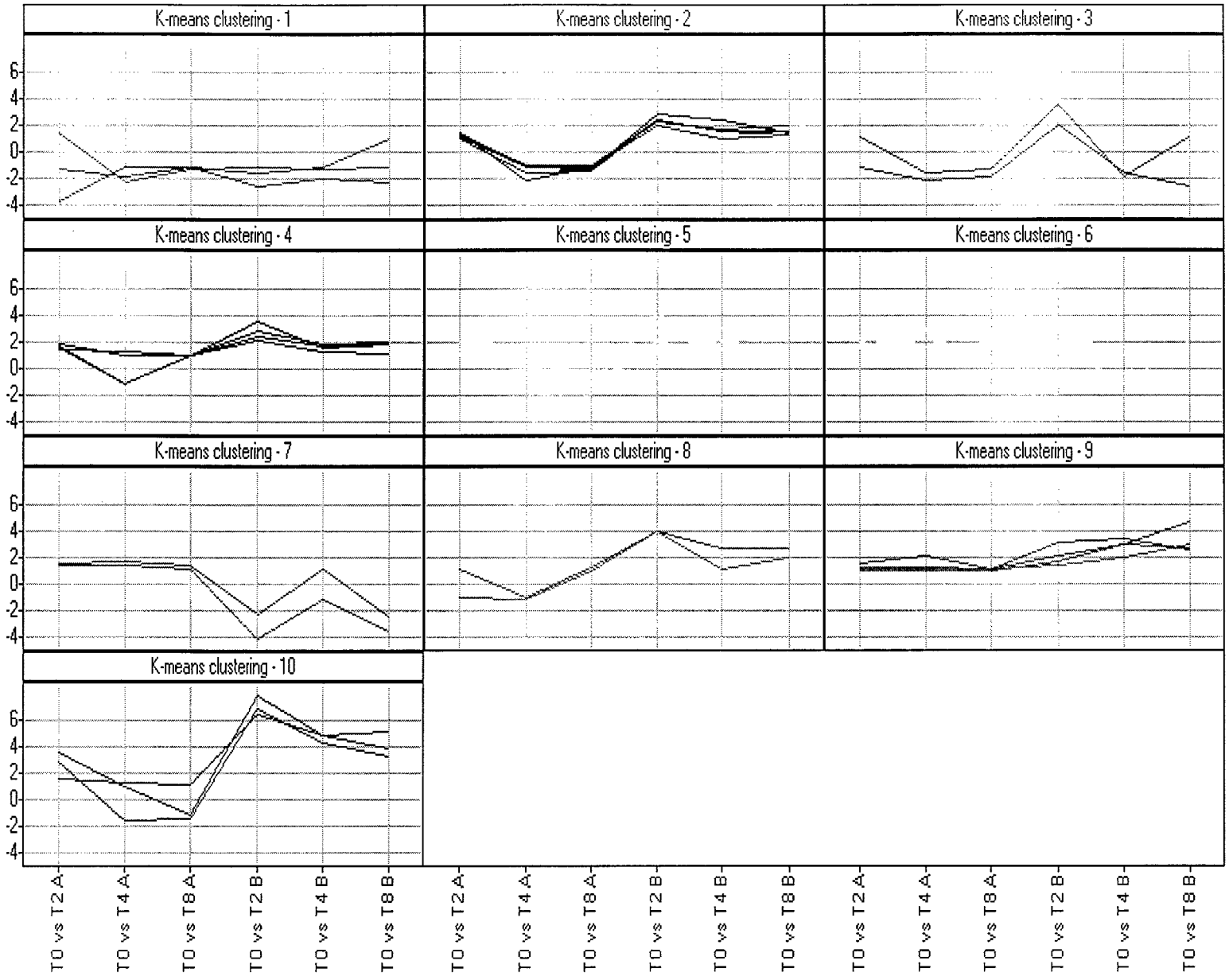

SBE
AP-1
-1070 TGCGAGACAGAGCGAGCCGCCCGGGGTGTGAGTCAGCGCGCT -1030
Fold Change
HS3ST2 1copy
HS3ST2 2copy
HS3ST2 3copy
3APP
+ TGF beta
- TGF beta
2.84X
5.42X
6.87X
2.25X
0
100000
200000
300000
400000
500000
600000

Figure 9b


K-means clustering - 1
K-means clustering - 2
K-means clustering - 3
K-means clustering - 4
K-means clustering - 5
K-means clustering - 6
K-means clustering - 7
K-means clustering - 8
K-means clustering - 9
K-means clustering - 10
6
4
2
0
-2
-4
T0 vs T2 A
T0 vs T4 A
T0 vs T8 A
T0 vs T2 B
T0 vs T4 B
T0 vs T8 B

Suppl. 2

Gene Description	Gene Nam	T0 vs T2 A	T0 vs T4 A	T0 vs T8 A	T0 vs T2 B	T0 vs T4 B	T0 vs T8 B
Dual specificity phosphatase 6	DUSP6	1.50	-2.21	-1.30	-1.07	-1.29	-1.05
Early growth response 1	EGR1	-3.67	-1.16	-1.17	-1.48	-1.13	1.02
V-myc avian myelocytomatosis viral oncogene	MYC	-1.27	-1.88	-1.09	-2.53	-1.91	-2.29
Bone morphogenetic protein 2	BMP2	1.50	-1.08	-1.22	2.86	2.45	1.42
Cysteine-rich angiogenic inducer 61	CYR61	1.31	-1.03	-1.02	2.32	1.70	1.97
Growth related oncogene (melanoma growth st	GRO1	1.46	-2.05	-1.14	2.47	1.60	1.31
Hairy Drosophila homolog	HRY	1.18	-1.08	-1.13	2.08	1.05	1.37
Immediate early response 3	IER3	1.08	-1.54	-1.43	2.38	1.53	1.55
Inhibitor of DNA binding 1	ID1	1.14	-1.49	-1.29	3.63	-1.89	1.19
V-jun homolog	JUN	-1.07	-2.10	-1.78	1.99	-1.59	-2.46
Epithelial membrane protein-1 (tumor-associat	EMP1	1.73	-1.13	1.07	2.11	1.25	1.20
Snail homolog 2	SNAI2	1.59	-1.06	1.08	2.49	1.71	1.90
Tissue factor pathway inhibitor 2	TFPI2	1.40	1.31	1.00	2.93	1.87	1.98
TGFB inducible early growth response	TIEG	1.81	1.02	1.02	3.62	1.59	1.84
Member of the histone deacetylase family	HDAC10	1.53	1.75	1.46	-2.19	1.09	-2.32
Tissue inhibitor of metalloproteinases 3	TIMP3	1.38	1.51	1.10	-4.09	-1.14	-3.48
Inhibitor of DNA binding 3	ID3	-1.02	-1.09	1.00	4.05	1.22	1.96
T-box 3 (ulnar mammary syndrome)	TBX3	1.12	-1.00	1.25	3.94	2.66	2.72
Plasminogen activator inhibitor 1	SERPINE1	1.53	2.15	1.13	2.17	3.02	4.67
SKI-like oncoprotein	SKIL	1.23	1.30	1.22	3.21	3.38	2.56
Protein with strong similarity to human STK38	STK38L	1.20	1.05	1.02	1.67	3.01	2.70
Transforming growth factor beta induced	TGFBI	1.01	1.19	1.20	1.42	2.05	3.02
Angiopoietin-like 4	ANGPTL4	2.86	-1.56	-1.43	6.83	4.35	3.22
Connective tissue growth factor	CTGF	3.59	1.08	-1.06	7.83	4.78	5.09
Enhancer of filamentation 1	HEF1	1.59	1.29	1.11	6.41	4.91	3.92

Suppl. 3