

Using DNA Microarray to Identify Sp1 as a Transcriptional Regulatory Element of Insulin-like Growth Factor-1 in Cardiac Muscle Cells

Tao Li¹, Yung-Hsiang Chen^{1,2}, Tsun-Jui Liu^{1,2}, Jia Jia¹, Steven Hampson³, Yue-Xin Shan¹,
Dennis Kibler³, Ping H. Wang^{1*}

¹Departments of Medicine, Biological Chemistry, Physiology and Biophysics, Center for Cardiovascular Hormone Research, University of California, Irvine, CA

²Institute of Clinical Medicine, National Yang-Ming University, Taiwan

³School of Information and Computer Science, University of California, Irvine, CA

T. Li and Y. Chen contributed equally in this study.

*To whom correspondence should be addressed: Department of Medicine, Med Sci 1, C240, Irvine, CA92697. Tel: 949-8246981. Fax: 949-8242200. E-mail: phwang@uci.edu.

Key words : Myocardium, DNA microarray, gene transcription, IGF, Sp1

Abstract

High throughput gene expression profiling with DNA microarray provides an opportunity to analyze transcriptional regulation of hundreds or thousands of similarly regulated genes. Transcriptional regulation of gene expression plays an important role in myocardial remodeling. We have studied cardiac muscle gene expression with DNA microarray and employed a computational strategy to identify common promoter motifs that respond to insulin-like growth factor-1 (IGF-1) stimulation in cardiac muscle cells. The analysis showed that the Sp1 binding site is a likely target of IGF-1 action. Further experiments with gel shift assay indicated that IGF-1 regulated the Sp1 site in cardiomyocytes, by increasing the abundance of Sp1 and Sp3 proteins. Using firefly luciferase as reporter gene, additional experiments showed that IGF-1 activated the promoter of Cyclin D3 and Glut1. Both promoters contain one Sp1 site. The effect of IGF-1 on these two promoters was abolished with siRNA for Sp1. Thus, the transcriptional activation of these two promoters by IGF-1 requires the induction of Sp1 protein. These experiments suggest that the global transcriptional regulatory actions of IGF-1 involve activation of the Sp1 site in cardiac muscle. The computational model we have developed is a prototypical method that may be further developed to identify unique cis- and trans-acting elements in response to hormonal stimulation during cardiac muscle growth, repair, and remodeling in normal and abnormal cardiac muscle.

Introduction

Clusters of microarray gene expression patterns indicate similar regulation of gene expression, which is likely due to common transcriptional regulatory mechanisms. The transcriptional regulatory mechanisms are mediated through transcription factors that typically interact with short DNA sequence motifs in the promoter region. Insulin-like growth factor 1 (IGF-1) modulates myocardial gene expression and regulates important cardiac muscle functions (1,2). In theory, a growth factor such as IGF-1 activates a specific set of signaling pathways that regulate specific transcription factors and hence modulate gene transcription. Thus, IGF-1 may modulate clusters of genes in microarray experiments through activation of common promoter sequence motifs. Recent developments in computational methods permit scientists to locate potential transcription factor binding sites based on microarray data (3,4,5). These computational models have been somewhat successful in identifying specific regulatory motifs in the yeast genome (3,4,5), but effective computational strategies to identify promoter motifs in mammalian genes have not yet been developed. Compared to the yeast, the structure of promoters in mammalian genes is more complex in size, location, and variability. Thus developing effective computational strategies to locate specific motifs from mammalian microarray experiments faces additional challenges. In this study, we have developed a computational approach and identified common promoter motifs that respond to IGF-1 stimulation in cardiac muscle cells.

Experimental Procedures

Primary Cardiomyocytes Culture and Microarray Study

Primary cultures of neonatal cardiomyocytes were prepared from Sprague-Dawley rats according to a protocol we previously described (6). Each cell preparation was prepared from approximately 40-60 neonates. To study the effects of IGF-1, the cells were serum-deprived overnight and then stimulated with IGF-1 (10^{-8} M) for the indicated time intervals. Total RNA was extracted from cardiomyocytes and the quality of each RNA sample was assessed with the Agilent Bioanalyzer 2100. Each sample was pooled from 4 to 5 P100 dishes. All samples used in this study passed quality assurance parameters according to the resulting electrophoresis. The total RNAs were used to synthesize single-stranded cDNA, and these single-stranded cDNA reactions are then used to generate double-stranded cDNA products. An *in vitro* transcription reaction was performed to produce biotin-labeled cRNA from the double-stranded cDNA. Prior to setup for hybridization, the biotin-labeled cRNA is fragmented to 50-200 bases in length. A small portion of each fragmented cRNA was analyzed on the Agilent instrument to verify that all the fragmented samples are of the appropriate size prior to preparation for hybridization onto the rat GeneChip U-34A (Affymetrix, Santa Clara, California). Hybridization was performed in a cocktail containing fragmented cRNA target, probe array controls, BSA, and herring sperm DNA for 16 hours. Then, the array underwent an automated washing and staining protocol using the Affymetrix Fluidics Station. All GeneChips were read twice by the Affymetrix (confocal laser) Scanner and the Affymetrix MicroArray Suite (MAS) software calculated an average of the two reads. Differential expression of genes was calculated with the Cyber T program (7), which employs a Bayesian statistical model based on the Student's t-test.

Calculation of Motif Over-representation

The approach we have taken is to write a discovery program, rather than a program that recognizing known binding sites. Consequently programs such as matInspector and rVista were not used. Most motif discovery approaches rely on the fact that the frequencies of motif instances in a set of co-regulated promoter regions are greater than their expected background occurrence rate. This can be explicitly computed as k-mer over-representation in the promoter of IGF-1 regulated genes. We represent motifs as specific words, called k-mers, which allows us to form a more accurate model of its background occurrence rate. Over-representation can be calculated in different ways depending on the choice of the background data and the statistical models. Our previously developed software suite of programs allows us to evaluate these various combinations (3). For this analysis of IGF-1 regulated promoters we used two different sets of background sequences. The first one is based on scrambled data which is equivalent to a zero-order Markov Model (8). The second set of background promoter sequences are derived from 50 genes that are not regulated by IGF-1 according to the microarray data. These 50 genes have the largest P value according to the Bayesian analysis as performed by Cyber T (7) and thus are least likely regulated by IGF-1 in cardiomyocytes. Given a choice of background models, over-representation can be computed using the standard binomial model for independent events (9, 10). For a particular k-mer m, the background data is used to estimate the probability (p) that m will occur. The most surprising motifs in some family are those whose occurrence is least likely. More precisely, the probability that a k-mer will occur x or more times is given by:

$$\text{Probability(k-mer m occurs x or more times)} = \sum_{i=x}^n p^i (1-p)^{n-i} \frac{n!}{i!(n-i)!}$$

Where n is the number of opportunities for the k-mer to occur in the data set of interest; x is the number of times the k-mer occurs in the data set, and p is the probability that the k-mer will

occur at in the background data set. Every k-mer is scored by the above formula and this value is used to sort the k-mers. The highest-ranking k-mers are those with least value and thus represents the k-mers with highest probability of over-representation. Data processing was carried out with a personal computer using programs we have developed.

Nuclear Extracts Preparation and Electrophoretic Mobility Shift Assay (EMSA)

Nuclear protein extracts were prepared as previously described (11). In brief, cardiomyocytes were scraped off the plates in an ice-cold buffer (10 mM HEPES, pH 7.9, 10 mM KCl, 1 mM DTT, 1.0 mM PMSF, 1.5 mM MgCl₂, 2 µg/ml aprotinin, 2 µg/ml pepstatin, and 2 µg/ml leupeptin). After centrifugation at 300 g for 10 minutes at 4°C, the cells were solubilized with the above buffer+0.1% Triton X-100 and centrifuged at 12,000 g for 10 minutes at 4 °C. The nuclear pellets were resuspended in a buffer (10 mM HEPES, pH 7.9, 1.5 mM MgCl₂, 0.42 M NaCl, 1 mM DTT, 0.2 mM EDTA, 1.0 mM PMSF, 25% glycerol, 10.5 mM PMSF, 2 µg/ml aprotinin, 2 µg/ml pepstatin, and 2 µg/ml leupeptin), incubated for 30 minutes at 4 °C, and centrifugation at 15,000 g for 30 minutes at 4 °C. The probe for Sp1/Sp3 gel shift assays was a 29-mer synthetic double-stranded oligonucleotide (5'-CCC TTG GTG GGG GCG GGG CCT AAG CTG CG-3'; 3'-GGG AAC CAC CCC CGC CCC GGA TTC GAC GC-5') containing the consensus Sp1/Sp3 binding site. A mutant double-stranded oligonucleotide (5'-CCCTTGGTGGGTTGGGGGCCTAAGCTGCG-3', 3'-GGGAACCACCCAACCCCGGATTCGACGC-5') was used as a control. For EMSA, double-stranded DNA was end-labeled with dig-dUTP using terminal transferase (Roche, Basel, Switzerland). The DNA-binding reaction was carried out with 5 µg nuclear proteins, 0.8 ng digoxin-labeled oligonucleotide at room temperature for 15 minutes. Nuclear extract-

oligonucleotide mixtures were separated from the unbound DNA probe by electrophoresis through a native 6% polyacrylamide gel (acrylamide/bisacrylamide, 29:1) in 0.5X TBE (Tris-Borate-EDTA buffer, pH 8.0). The digoxin-labeled oligonucleotide was detected with anti-digoxin antibody conjugated with alkaline phosphatase. In competition analysis, nuclear extracts were incubated with 125X unlabeled oligonucleotide probe 15 minutes before the addition of labeled oligonucleotides. For supershift analyses, anti-Sp1 and anti-Sp3 monoclonal antibodies (Santa Cruz Biotechnology, Santa Cruz, California) were added to the reaction mixture 60 minutes before the addition of labeled oligonucleotides at 4°C.

Western Blot

The cells were lysed with lysis buffer (137 mM NaCl, 20 mM Tris-HCl, pH 7.5, 10% glycerol, 1% Triton X-100, 0.5% NP-40, 2 mM EDTA, pH 8.0, 3 µg/ml aprotinin, 3 µg/ml leupeptin, 2 mM phenylmethylsulfonyl fluoride, 20 mM NaF, 10 mM NaPP, and 2 mM Na₃VO₄) and equal amounts of proteins were separated by SDS-PAGE. The proteins were transferred to polyvinylidene difluoride membrane and incubated with a blocking buffer (5% nonfat milk in 20 mM Tris-HCl, pH 7.5, 137 mM NaCl, 0.1% Tween 20) for 1 h at room temperature. The membranes were incubated with anti-Sp1 or anti-Sp3 antibodies, washed three times (20 mM Tris-HCl, pH 7.5, 137 mM NaCl, and 0.1% Tween 20), incubated with HRP conjugated secondary antibodies (1:5000 to 1:10000 dilution) for 1h at room temperature, washed three times, and then detected with ECL.

Immunofluorescence Study

Cardiomyocytes were plated in 4-chamber glass slide (Lab-Tek, Naperville, Illinois) in DMEM+10% FBS. After overnight serum deprivation, the cardiomyocytes were incubated with IGF-1 ($2 \times 10^{-8} \text{M}$) in the presence or absence of cycloheximide (100 $\mu\text{g/ml}$). The cells were then rinsed and fixed with pre-chilled methanol at -20°C for 20 minutes. The slides were sequentially incubated with PBS+1%BSA for 2 hours, anti-Sp1 antibodies, and Texas-Red-tagged secondary antibodies (Molecular Probes, Eugene, Oregon). Finally, the cells were counterstained with 4,6-diamidino-2-phenylindole (DAPI) (0.1 $\mu\text{g/ml}$) for 10 min at room temperature. The slides were visualized with a Zeiss Axiophot epifluorescence microscope and the images were recorded with a Sensys digital camera and analyzed with PathVision imaging software (12).

Construction of Plasmids

A 1033 bp fragment of Cyclin D3 promoter (-1068 to -35 upstream from ATG) and a 976 bp fragment of Glut1 promoter (-1076 to -48 from ATG) was cloned by PCR amplification with genomic DNA extracted from rat cardiomyocytes (primer sets: for Cyclin D3 promoter, 5'-ATTACAGATTTGCACCACCACA-3' and 5'-GGCTAGCGAGTCCTAGGAGAA-3'; for GLUT1 promoter, 5'-GGGACCACAGAGGCTATTGA-3' and 5'-AACGGACGCGCTGTA ACTAT-3'). Both fragments of the promoter regions contain one Sp1 site. To create mutations at Sp1 site in these two promoter fragments, two sets of primers were used during PCR for site-directed mutagenesis (Cyclin D3: 5'-TCGTCGCGAGTTGGGG-3' and 5'-CCCCAACTCGCGACGA-3'; Glut1: 5'-AGGCCCCCAACCCTTC-3' and 5'-GAAGGGTTGGGGGCCT-3'). The Sp1 sites were mutated from GGCGGG/CCCGCC to GTTGGG/CAACCC. The wildtype (D3 and Glut1) and mutated (mD3 and mGlut1) promoter

fragments were cloned into the TOPO TA vector (Invitrogen, Carlsbad, CA), the cloned fragments were cut out with Asp718 and XhoI, and ligated to a firefly luciferase reporter vector pGL3-Basic (Promega, Madison, WI). These four plasmids, pGL3-D3-luc, pGL3-mD3-luc, pGL3-Glut1-luc, and pGL3-mGlut1-luc were used to analyze the effects of IGF-1 on the Sp1 motif. The fidelity of DNA sequences was verified by sequencing with ABI 3700 Capillary DNA Analyzer.

Gene Silencing with SiRNA

Sp1 siRNA targeted to AATGAGAACAGCAACAACCTCC and Sp3 siRNA targeted to AAGTTCTCAGACAATGACTGC were designed under the following principles: GC ratio \approx 50%, avoid GGG or CCC, sequence begin with AA, and dTdT overhang. These sequences had been digitally searched and no similar sequences were found in current database. The oligos were respectively annealed to create siRNA duplex and was transfected to the cells with TransMessenger Reagent according to the manufacturer's instructions (Qiagen). The control cells were transfected with a control siRNA duplex (sense UUCUCCGAACGUGUCACGUdTdT, anti-sense ACGUGACACGUUCGGAGAAAdTdT), this control siRNA has no known target in mammalian genomes. All siRNAs were purchased from Qiagen (Valencia, CA).

Transfection and Luciferase Assay

Neonatal cardiomyocytes were plated in 24 well plates using DMEM+10% FBS. H1299 cells were kindly provided by Dr. Rainer Brachmann (Irvine, CA). 24-48 hours after plating, the cells were co-transfected with Cyclin D3 or Glut1 plasmids with firefly luciferase reporter (1 ug/well)

and a Renilla luciferase control vector pRL-SV40 (0.1 ug/well) (Promega, Madison, WI) using Lipofectamine (Life Technologies) as a vehicle. pRL-SV40 constitutively express Renilla luciferase and thus served as an internal control to normalize specific activity of firefly luciferase. 24 hours after transfection, the cells were serum-deprived overnight and then stimulated with IGF-1 (1×10^{-8} M), recombinant growth hormone (GH) (1ug/ml), PMA (10ug/ml), or vehicles. Cells were lysed with a lysis buffer and the lysates were harvested and cleared with brief centrifugation. The activities of firefly and Renilla luciferase were measured using the Dual Luciferase Reporter Assay System (Promega) with a Monolight 2010 Luminometer (Analytical Luminescence Laboratory, San Diego, CA). Specific activities of firefly luciferase were normalized with the activities of Renilla luciferase in the same sample. pCMV-Sp1 was kindly provided by Dr. Eric Standbridge (Irvine, CA).

Results

To profile gene expression pattern in response to IGF-1 stimulation, we have previously used DNA microarray spotted on nylon membrane, which revealed that the majority of genes modulated by IGF-1 can be identified after two hours of IGF-1 stimulation in primary cardiomyocytes (6). We also found that we could reduce false positive results by increasing sample size and excluding minimally expressed genes (6). Using the same strategy, instead of nylon membrane microarray, we used the Affymetrix rat genome chip U-34A to identify those genes regulated in the cardiomyocytes after two hours of IGF-1 incubation. Five sets of independent samples were included in the control group and the IGF-1-stimulated group; differential regulation of gene expression were analyzed with Cyber T as we previously reported (6, 7). Only those genes achieved statistical significance ($P < 0.05$) were included in subsequent analysis. The results showed that IGF-1 up-regulated 107 genes and down-regulated 52 genes. After excluding those genes that were minimally expressed, we searched existing databases, including GenBank, and retrieved 63 up-regulated genes with more than 500 base pairs of 5'-end promoter sequences immediately upstream from the translation start codon. Our computational analysis was restricted to a maximum of 1000 bases immediately upstream from the start codon. There are several reasons for limiting the search to 1000 base pair, although we recognize that binding sites could well occur much further away. However, it is likely that the immediate upstream region contains the most important regulatory elements. In order to establish statistical significance via over-representation it is necessary to demonstrate that a particular k-mer occurs much more frequently than one expects. A binding site is typically only over-represented in some local region and the only local region that we can readily identify for X gene is the immediate upstream region.

Motif search assumes that the frequency of the motifs of interest is over-represented in the IGF-1-upregulated genes. Such “over-representation” is defined by comparing the IGF-1-upregulated promoters against a set of background promoter sequences (3). In our analysis, we have used two different types of background sequences. The first is simply scrambled sequences from the promoter included in the analysis (8); the second is the promoter sequences from 50 genes that are not regulated by IGF-1 according to our microarray data. Up-regulated genes were compared respectively to these two sets of background sequences. The over-represented 7-mers in the promoter of the up-regulated genes were correspondingly ranked according to their statistical probability and the top twenty 7-mers were investigated as potential regulatory motifs. To enhance the specificity of analysis, only those 7-mers that were over-represented when compared against both sets of background sequences were selected as potential response elements of IGF-1 activation (Table 1). All these 7-mers matched known transcription factors in the TRANSFAC database. The Sp1 consensus motif, adding two different motifs together, is the most common motif in the promoters of those up-regulated genes. To further verify that these sites are potential targets of IGF-1 regulation, we performed similar analysis on the genes that were upregulated by IGF-1 as determined by our previous experiments using microarray spotted on nylon membrane (6). The results showed that again the Sp1 site ranked as the most likely promoter motif regulated by IGF-1 in cardiomyocytes. These results suggest that the Sp1 binding site is a target of IGF-1 receptor signaling in cardiac muscle cells.

To determine whether the Sp1 binding site is a target of IGF-1 action, the nuclear proteins were extracted from the control and IGF-1 stimulated cardiomyocytes and analyzed with gel shift

assay. As shown in Fig. 1A, in the IGF-1-stimulated cells there was an increased nuclear protein binding to the labeled oligos containing a consensus Sp1 binding site. Both Sp1 and Sp3 protein bind to the same consensus sequence. To investigate IGF-1 effects on Sp1 and Sp3 binding, specific antibodies were added to the binding mixture for supershift assay. The results showed that Sp1 protein was mainly in complex 1, 3, and 4, whereas Sp3 were present in complexes 1-4. To explore whether IGF-1 regulation of Sp1 is modulated through the synthesis of Sp1, cycloheximide was added to the culture medium to inhibit protein synthesis during IGF-1 stimulation. The results showed that IGF-1 activation of nuclear protein binding to the consensus Sp1 site was completely blocked by cycloheximide, suggesting IGF-1 induction of Sp1 and Sp3 binding complexes requires protein synthesis (Fig. 1B). Time-course experiments showed that IGF-1 increased accumulation of Sp1 and two forms of Sp3 in the nuclear protein extracts within 30 to 60 minutes of stimulation (Fig. 2A). To further confirm the effect of IGF-1 on Sp1 protein, immunofluorescence study was carried out using specific Sp1 antibody. The results showed that IGF-1 increased the abundance of Sp1 proteins in cardiomyocytes and cycloheximide inhibited IGF-1 induction of Sp1 proteins (Fig. 2B). These experiments proved that IGF-1 signaling can activate Sp1/Sp3 binding motifs through increasing the abundance of Sp1 and Sp3 in the cardiomyocytes nucleus, and thus provides direct evidence that the computational algorithm we used to locate IGF-1 response elements in those IGF-1-regulated genes is valid.

IGF-1 activation of Sp1 motif likely leads to modulation of promoter activities. The following experiments are designed to determine the effect of IGF-1 on the promoter of the genes that have Sp1 site in their 5' upstream region. To this end, we have respectively cloned two promoter

fragments from the 5' regions of Cyclin D3 and glucose transporter 1 (Glut1) gene upstream from ATG and inserted these two promoter fragments into a firefly luciferase reporter gene vector (pGL3-D3-luc and pGL3-Glut1-luc). Cyclin D3 and Glut1 promoters were chosen because 1) IGF-1 increased the expression of Cyclin D3 (158%) and Glut1 (153%), 2) these two genes were involved in the regulation of myocardial biology, and 3) promoter sequences allowed convenient site-directed mutagenesis. In this series experiments, a constitutive Renilla luciferase reporter gene vector (pRL-SV40) was co-transfected to cardiomyocytes with pGL3-D3-luc or pGL3-Glut1-luc to standardize firefly luciferase activities. As shown in Fig. 3A and 3B, IGF-1 activated pGL3-D3-luc and pGL3-Glut1-luc. Phorbol-12-myristate-13-acetate (PMA) served as positive control and growth hormone as negative control in these experiments. We next mutated the Sp1 site in these two promoter fragments and then inserted into firefly luciferase reporter gene vector (pGL3-mD3-luc and pGL3-mGlut1-luc). IGF-1 failed to increase firefly luciferase activities in the cardiomyocytes transfected with pGL3-mD3-luc or pGL3-mGlut1-luc (Fig. 3C and 3D), This series of study suggest that IGF-1 induction of Cyclin D3 and Glut1 promoter requires the presence of the Sp1 site. However, because the promoter activities were generally low in the constructs with mutated Sp1 site, we could not exclude the possibility that mutation on Sp1 site abolish the promoter activity and thus became unresponsive to IGF-1 stimulation. To confirm the role of Sp1 and Sp3 protein, we used an alternative approach with siRNA to reduce endogenous expression of Sp1 and Sp3. As shown in Fig 4A, transfecting cardiomyocytes with siRNAs respectively targeting Sp1 and Sp3 reduced the abundance of Sp1 and Sp3 proteins. In the cells transfected with Sp1 siRNA, the stimulatory effects of IGF-1 on Cyclin D3 and Glut1 promoter were completely abolished (Fig 4B). In comparison, the effects of PMA were not abolished, but attenuated by the Sp1 siRNA. In the cells transfected with Sp3 SiRNA, the

effects of IGF-1 and PMA on Cyclin D3 and Glut1 promoter were not affected. Therefore, the stimulatory effects of IGF-1 on these two promoter fragments were not dependent on Sp3. To verify that increased expression of Sp1 lead to activation of Cyclin D3 and Glut1 promoter, we co-transfected pCMV-Sp1 and promoter constructs to H1299 cells as shown in Fig 4C.

Overexpression of Sp1 significantly increased luciferase reporter gene activities in the wildtype promoter construct (pGL3-D3-luc and pGL3-Glut1-luc), but Sp1 overexpression did not activate the mutated promoter constructs (pGL3-mD3-luc and pGL3-mGlut1-luc). These data indicate that the transcriptional stimulatory effect of IGF-1 involves increased abundance of Sp1 protein in cardiomyocytes.

Discussion

Understanding how the expression of thousands of genes is regulated in the cell remains a significant and difficult challenge in biology. Traditionally, gene regulation has been studied one gene at a time. It has been anticipated that high-throughput gene profiling may provide an opportunity to compute and identify consensus transcription binding sites in the similarly regulated genes. But, to our knowledge, this is the first report that validates a computational strategy for the identification of common cis- and trans-acting elements in the similarly regulated genes in mammalian cells. This computational strategy may become a useful tool to identify unique cis- and trans-acting elements in response to hormonal stimulation during cardiac muscle growth, repair, and remodeling in the normal and diseased hearts.

The GC-rich region is a known target of IGF-1 regulation (13), and potential Sp1 sites can be found in the GC-rich regions (14). Previous literatures rarely explored the relationship between IGF-1 receptor signaling and the Sp1 domain, but indirect evidence suggests that the Sp1 sites within the GC-rich region may be targets of IGF-1 receptor signaling. Over-expressing a dominant negative Sp1 attenuated IGF-1 activation of adenosine deaminase promoter in MCF-7 cells (15). Using a promoter fragment of elastin gene, another study showed that IGF-1 might preferentially modulate the abundance of a DNA-protein complex containing Sp3 in smooth muscle cells (16). In cardiac fibroblast, van Eickels et al. had showed that IGF-1 increased SP1 mRNA levels (17). These studies render additional support to our finding that the Sp1 motif is a common response element for IGF-1 signaling. Interestingly, another study by Kaytor et al. showed that insulin increased Sp1 binding to the promoter of IGF-1 gene (18), which suggest Sp1 is also a target of insulin action (18).

Sp1 and Sp3 belong to the Sp zinc-finger superfamily (14). Both Sp1 and Sp3 are essential for survival as Sp1 knockout and Sp3 knockout mice died before or at birth (19, 20). Sp1 is a transactivator/enhancer of gene transcription, but the role of Sp3 in gene regulation is less clear (14). Sp3 is homologous to Sp1 with similar affinities to the Sp1 binding sites. While Sp3 may enhance Sp1-mediated transcription in a small number of genes, Sp3 may suppress Sp1-mediated transcription in most genes that have been studied (14). It has been proposed that Sp3 competes with Sp1 binding to the Sp1 motif and thus serves as negative regulator of Sp1 sites. Our data indicated that IGF-1 activation of Cyclin D3 and Glut1 promoters did not involve Sp3. However, the role of Sp3 in the global transcriptional regulatory effects of IGF-1 will require further investigation. Our result nonetheless proved that up-regulation of Sp1 play a key role when IGF-1 activated the promoter of Cyclin D3 and Glut1.

Among the genes regulated by IGF-1 in our microarray study, approximately 58% exhibit potential Sp1 sites in their 1 kb promoter regions; the majority of the Sp1 sites are located within the first 500 bp next to the ATG. In comparison, only 6% of the 50 non-regulated genes have Sp1 sites in the 5' region. The presence of Sp1 sites in a small number of the non-regulated genes suggests that the Sp1 motif alone may not be sufficient for the transcriptional regulatory action of IGF-1 in some genes. Full activation of transcription by IGF-1 may involve motifs other than Sp1 in those genes. Alternatively, other transcription regulatory steps, such as chromatin remodeling and scaffolding, may be necessary during IGF-1 activation of Sp1.

The Sp family of transcription factors is involved in the regulation of many important cardiac genes (21). For example, Sp1 is required for transcription of sarcoplasmic reticulum Ca^{+2} -ATPase gene (22), and our microarray results indicate IGF-1 increased the expression of Ca^{+2} -ATPase. Evidence from experimental models of cardiomyopathy provides additional links between the Sp family and cardiac genes of functional significance. A study by Sack et al. indicates that induction of the Sp family is involved in the regulation of myocardial fetal gene expression pattern during cardiac hypertrophic response (23). Hypoxia is another stimuli that increases Sp1 expression and also is associated with increased expression of fetal gene pattern in myocardium (24). These observations indicate that the Sp transcription factor family may play an important role in the regulation of cardiac muscle biology.

Different statistical approaches have different biases and give different results. If possible, it is useful to try several and look for areas of agreement. Our program includes a range of methods and we have found that k-mer over-representation generally gives the best results, although most methods generally work to some extent. Calculation of k-mer patterns longer than 10 can sometimes be discovered by examining overlapping shorter patterns. On the other hand, short patterns, say of length 4 or 5, are difficult to find. Many patterns of this length will repeat at random, so establishing a statistically significant difference is difficult. We had analyzed our data using 5-mer and 6-mer approach, and the results showed that Sp1 is one of the highest-ranking motif in the promoters of those genes regulated by IGF-1. Our program permits a wide range of background models and statistical measures which we regularly use, although for brevity, only two were reported in this particular study. Nevertheless, among the different

background models we have tried, Sp1 sites consistently ranked among the most likely over-representation (i.e. most likely target of IGF 1 action).

In general, our program provides a much wider range of interactive analysis tools than other currently available programs. In the current case, relatively straightforward analysis using our previously published methods was immediately successful, so a more extensive discussion of possible or alternative analysis methods seemed unnecessary, nor was there room for it within the scope of this paper. For example, looking for conserved DNA regions across species is a powerful technique for analyzing a single gene in the absence of expression data, but is comparatively complex and is not necessary if a large pool of co-regulated genes is available. It might well provide additional information, but like a wide array of other DNA analysis techniques, it does not address the issue at hand as simply or directly as k-mer analysis. Other algorithms were tried and did not yield useful results. In particular, Gibb's sampling was used without success. The closest algorithm to ours, since it based on the same model, is the method developed by van Helden et al (10). However the van Helden model does not allow for interactive examination of the result nor does it allow for a comparison set of sequences.

The strategy we have developed for the analysis of microarray data represents a prototypical approach; this is likely a start toward more refined computational strategies. The development of such computational methods will provide a new way to study transcriptional regulation of thousands of co-regulated genes in cardiac and other tissues in response to hormonal stimulation, injuries, and pathophysiological adaptations. Although this approach may offer a new window of opportunity for future research on gene transcription, many new questions arise from the

results of this study. For examples, what are the specificity and sensitivity of our computational strategy? How can we tweak our computational approach to achieve better sensitivity and specificity? Can we define a statistical threshold to differentiate true regulatory motifs from false positive ones, such as p value in t-test? Can we develop similar strategies to analyze chromatin remodeling that is essential to IGF-1 actions on transcription? How are we going to categorize and warehouse the complex mammalian promoter sequences and structure in a database for future motif analysis? These and many other questions are beyond the scope of the present study and can't be answered by the data presented here. There are a lot to be learned.

Acknowledgement

This work is supported by grants from NHLBI and NIDDK (to PHW).

Figure Legends

Fig. 1. Gel shift assays demonstrated IGF-1 activation of the Sp1 motif. EMSA was performed using nuclear proteins extracted from the control and IGF-1-stimulated (60 minutes) neonatal cardiomyocytes. A: IGF-1 activation of the Sp1 motif involves Sp1 and Sp3 protein. 125X unlabeled oligos (Sp1 competitor, lane 3) or labeled mutant oligos (lane 4) were added to the binding mixture (in the presence of labeled oligos containing the consensus Sp1 site) to verify the specificity of protein-DNA interaction. The binding mixtures were pre-incubated with Sp1 antibodies (lane 5) or Sp3 antibodies (lane 6) to generate supershift of protein-DNA complexes. B: IGF-1 activation of Sp1 motif is blocked by cycloheximide. The cardiomyocytes were pre-incubated with Cycloheximide prior to IGF-1 stimulation when indicated (lane 5).

Fig 2. IGF-1 increased accumulation of Sp1 protein. A: Time-course effect of IGF-1 on the abundance of Sp1 and Sp3 protein in cardiomyocyte nuclear protein extracts. The abundance of Sp1 and Sp3 proteins in nuclear extracts was determined with Western blot. B: Immunofluorescence staining of Sp1 proteins in cardiomyocytes. The cardiomyocytes were stimulated with vehicles or IGF-1 (30 minutes) and then stained with Sp1 antibodies. When indicated, the cells were pre-incubated with Cycloheximide (Cx) to block protein synthesis. The nuclei were counterstained with DAPI. The left panels represent composite images of Sp1 and DAPI staining.

Fig 3. IGF-1 activated Cyclin D3 and Glut1 promoters. Wildtype promoter constructs (pGL3-D3-luc and pGL3-Glut1-luc) or mutated promoter constructs (pGL3-mD3-luc and pGL3-

mGlut1-luc) were co-transfected with a Renilla luciferase reporter construct to neonatal cardiomyocytes. The Sp1 sites were mutated in the pGL3-mD3-luc and pGL3-mGlut1-luc to prevent activation of Sp1 motif. The cardiomyocytes were overnight serum-fasted before stimulation with IGF-1, growth hormone (GH), or PMA. The activities of firefly luciferase were standardized with Renilla luciferase activities. A: Activation of wildtype Cyclin D3 promoter. B: Activation of wildtype Glut1 promoter. C: Activation of mutated Cyclin D3 promoter. D: Activation of mutated Glut1 promoter. * p value < 0.03, vs. basal, Student's t-test. # p value, insignificant vs. basal, Student's t-test.

Fig 4. IGF-1 activation of Cyclin D3 and Glut1 promoter requires Sp1 protein. Sp1 siRNA and Sp3 siRNA were transfected to cardiomyocytes to suppress the expression of Sp1 and Sp3 protein. A non-silencing siRNA was transfected to the control cells. A: Western blots showed decreased level of Sp1 and Sp3 protein in the cells transfected with specific siRNA. B: Firefly luciferase activities upon IGF-1 stimulation. Wildtype promoter constructs (pGL3-D3-luc and pGL3-Glut1-luc) were co-transfected with a Renilla luciferase reporter construct to neonatal cardiomyocytes. The cardiomyocytes were overnight serum-fasted before stimulation with IGF-1 or PMA. The activities of firefly luciferase were standardized with Renilla luciferase activities. * p value > 0.6, vs. basal, Student's t-test. # p value < 0.05 vs. basal, Student's t-test. C: Overexpression of Sp1 activated Cyclin D3 and Glut1 promoter. These experiments were carried out in H1299 cells. Sp1 protein was overexpressed by transfecting H1299 cells with pCMV-Sp1. The upper panel showed increased expression of Sp1 protein after pCMV-Sp1 transfection to the cells plated on 24-well plates. Overexpression of Sp1 (30 ng pCMV-Sp1/well) lead to increased activation of pGL3-D3-luc and pGL3-Glut1-luc. * p < 0.01, vs.

without pCMV-Sp1.

Table 1. Over-represented Promoter Motifs in the IGF-1 Up-Regulated Genes

Motifs	Matching Factors	Occurrence of 7-mer Motif in the Regulated Promoters	Binominal Probability of Over-representation
GCGGCGG	T-Ag	35	2.18×10^{-18}
TATATAT	TATA box	18	6.17×10^{-12}
ATATATA	TATA box	16	4.79×10^{-10}
CCCGCCC	SP1	23	3.90×10^{-9}
GGGCGGG	SP1	22	1.81×10^{-8}

Reference

1. Fazio S, Palmieri EA, Biondi B, Cittadini A, Sacca L. The role of the GH-IGF-I axis in the regulation of myocardial growth: from experimental models to human evidence. *Eur J Endocrinol.* 2000;142(3): 211-6.
2. Ren J, Samson WK, Sowers, JR. Insulin-like growth factor I as a cardiac hormone: physiological and pathophysiological implications in heart disease. *J Mol Cell Cardiol.* 1999;31(11): 2049-61.
3. Hampson S, Kibler D, Baldi P. Distribution patterns of over-represented k-mers in non-coding yeast DNA. *Bioinformatics.* 2002;18(4): 513-28.
4. Zhu Z, Pilpel Y, Church GM. Computational identification of transcription factor binding sites via a transcription-factor-centric clustering (TFCC) algorithm. *J Mol Biol.* 2002;318(1): 71-81.
5. Wolfsberg TG, Gabrielian AE, Campbell MJ, Cho RJ, Spouqe JL, Landsman D. Candidate regulatory sequence elements for cell cycle-dependent transcription in *Saccharomyces cerevisiae*. *Genome Res.* 1999 ;(8): 775-92.
6. Liu T, Lai H, Wu W, Chinn S, Wang PH. Developing a strategy to define the effects of insulin-like growth factor-1 on gene expression profile in cardiomyocytes. *Circ Res.* 2001 ;88:1231-8.
7. Long AD, Mangalam HJ, Chan BY, Tollerli L, Hatfield GW, Baldi P. Improved statistical inference from DNA microarray data using analysis of variance and a Bayesian statistical framework. Analysis of global gene expression in *Escherichia coli* K12. *J Biol Chem.* 2001;276(23):19937-44.
8. Baldi P, Brunak S. *Bioinformatics : the Machine Learning Approach.* 2nd Ed Cambridge, Mass: MIT Press; 2001.

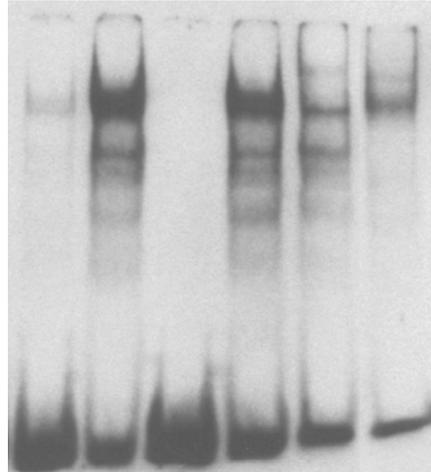
9. Ross SM. Introduction to probability, 5th Ed., San Diego, CA, Academic; 2003.
10. van Helden JV, Andre B, Collado-Vides J. Extracting Regulatory Sites from the Upstream Region of Yeast Genes by Computational Analysis of Oligonucleotide Frequencies. *J Mol Biol* 1998;281: 827-42.
11. Ji YS, Xu Q, Schmedtje JF. Hypoxia induces high-mobility-group protein I (Y) and transcription of the cyclooxygenase-2 gene in human vascular endothelium. *Circ Res.* 1998 ;83: 295-304.
12. Wu W, Lee W, Wu Y, Chen D, Liu T, Jang A, Sharma PM, Wang PH. Expression of constitutively active phosphatidylinositol 3 kinase inhibits activation of caspase 3 and apoptosis of cardiac muscle cells. *J Biol Chem.* 2000;275: 40113-40119.
13. Urban RJ, Shupnik MA, Bodenbun YH. Insulin-like growth factor-I increases expression of the porcine P-450 cholesterol side chain cleavage gene through a GC-rich domain. *J Biol Chem.* 1994;269(41): 25761-9.
14. Suske G. The Sp-family of transcription factors. *Gene* 1999;238: 291-300.
15. Xie W, Duan R, Safe S. Activation of adenosine deaminase in MCF-7 cells through IGF-estrogen receptor α crosstalk. *J Mol Endocrinol.* 2001;26: 217-228.
16. Conn KJ, Rich CB, Jensen DE, Fontanilla M, Bashir M, Rosenbloom J, Foster JA. Insulin-like growth factor-I regulates transcription of the elastin gene through a putative retinoblastoma control element. *J Biol Chem.* 1996;272: 28853-60.
17. van Eickels M, Vetter H, Grohe C. Angiotensin-converting enzyme (ACE) inhibition attenuates insulin-like growth factor-I (IGF-I) induced cardiac fibroblast proliferation. *Br J Pharmacol.* 2000;131(8):1592-6.

18. Kaytor EN, Zhu JL, Pao CI, Phillips LS. Insulin-responsive nuclear proteins facilitate Sp1 interactions with the insulin-like growth factor-I gene. *J Biol Chem.* 2001;276(40):36896-901.
19. Marin M, Karis A, Visser P, Grosveld F, Philipsen S. Transcription factor Sp1 is essential for early embryonic development but dispensable for cell growth and differentiation. *Cell.* 1997;89(4): 619-28.
20. Bouwman P, Gollner H, Elsasser HP, Eckhoff G, Karis A, Grosveld F, Philipsen S, Suske G. Transcription factor Sp3 is essential for post-natal survival and late tooth development. *EMBO J.* 2000;19(4): 655-61.
21. Flesch M. On the trail of cardiac specific transcription factors. *Cardiovasc Res.* 2001;50(1): 3-6.
22. Baker D, Dave V, Reed T, Periasmy M. Multiple Sp1 binding sites in the cardiac/slow twitch muscle sarcoplasmic reticulum Ca⁺²-ATPase gene promoter are required for expression of in So10 muscle cells. *J Biol Chem.* 1996;271: 5921-28.
23. Sack MN, Disch DL, Rockman HA, Kelly DP. A role for Sp and nuclear receptor transcription factors in a cardiac hypertrophic growth program. *Proc Nat Acad Sci. USA* 1997;94: 6438-43.
24. Xu Q, Ji Y, Schmedtje JF. Sp1 increases expression of cyclooxygenase-2 in hypoxic vascular endothelium. *J Biol Chem.* 2000;275: 24583-9.

A.

Sp1 competitor	-	-	+	-	-	-
Sp1 (m) oligo	-	-	-	+	-	-
Sp1 Ab	-	-	-	-	+	-
Sp3 Ab	-	-	-	-	-	+

Complex 1 →
 Complex 2 →
 Complex 3 →
 Complex 4 →



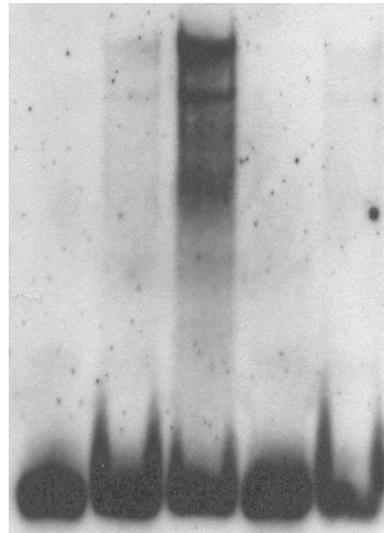
← Super shift

← Free probe

IGF-1 - + + + + +

B.

Sp1 competitor	-	-	-	+	-
Nuclear Extract	-	+	+	+	+
IGF-1	-	-	+	+	+
Cycloheximide	-	-	-	-	+



Free Probes →

Fig. 1

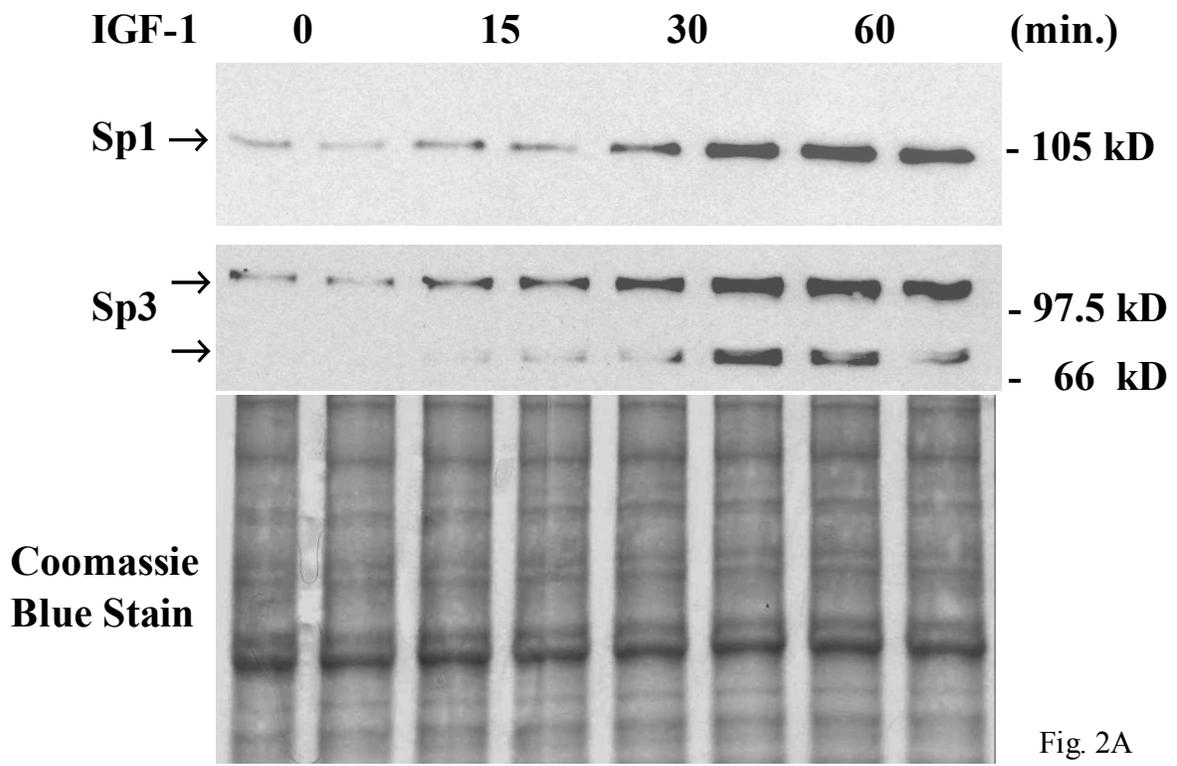


Fig. 2A

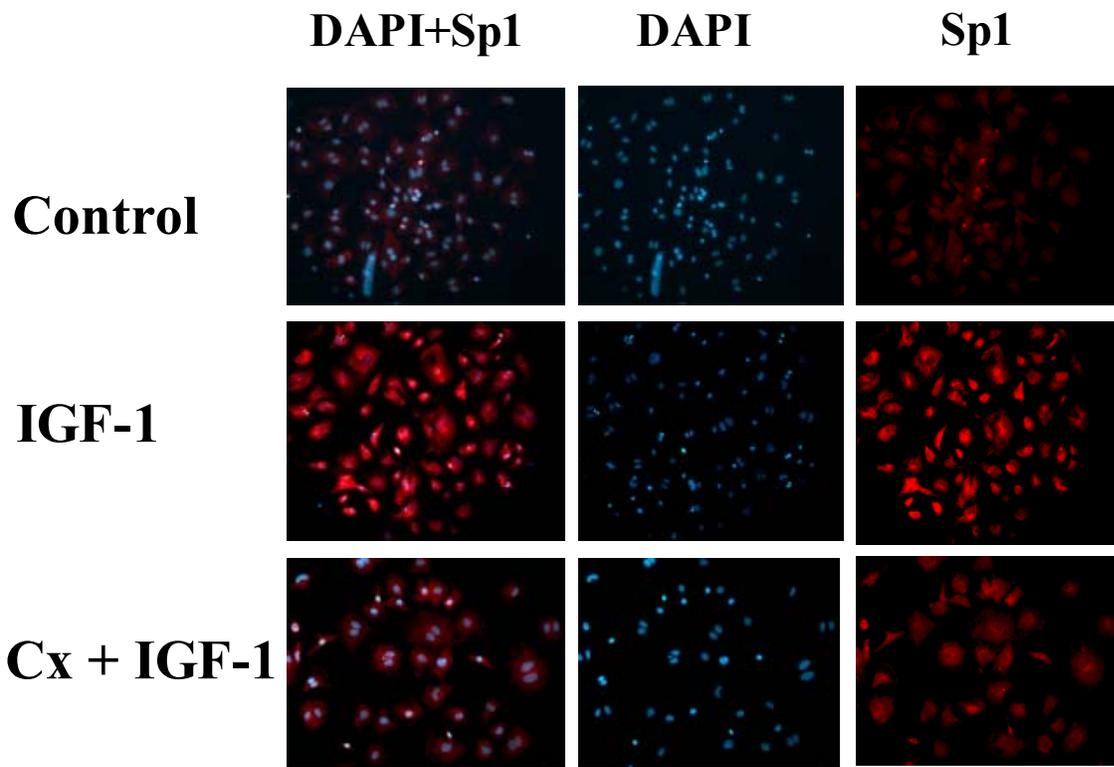


Fig. 2B

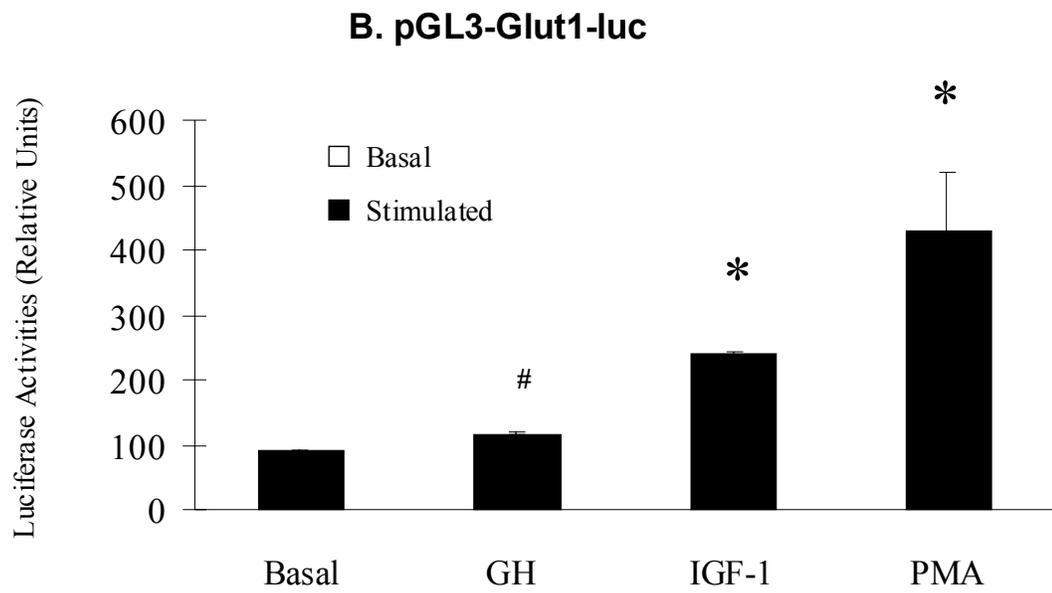
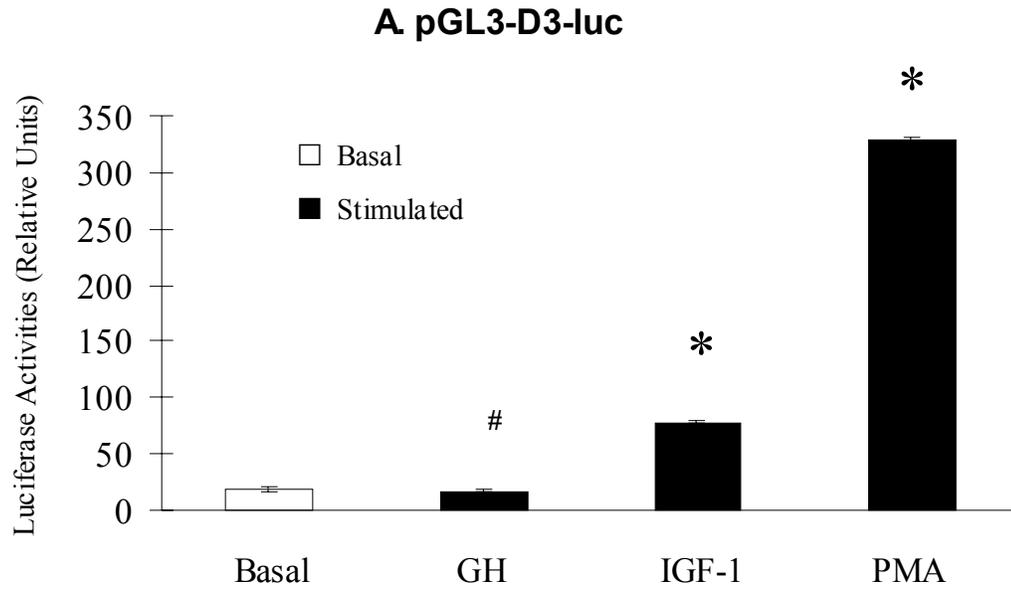
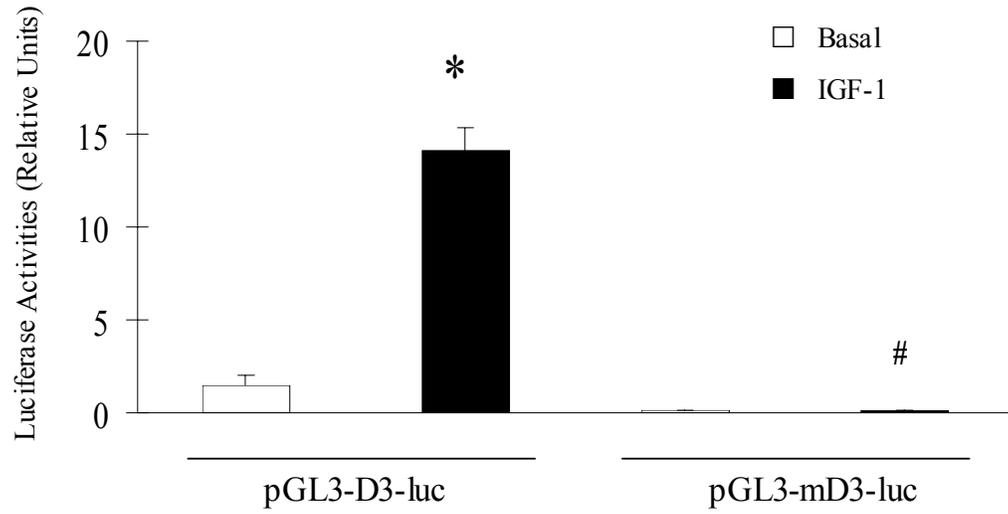


Fig. 3A, 3B

C. Cyclin D3



D. Glut1

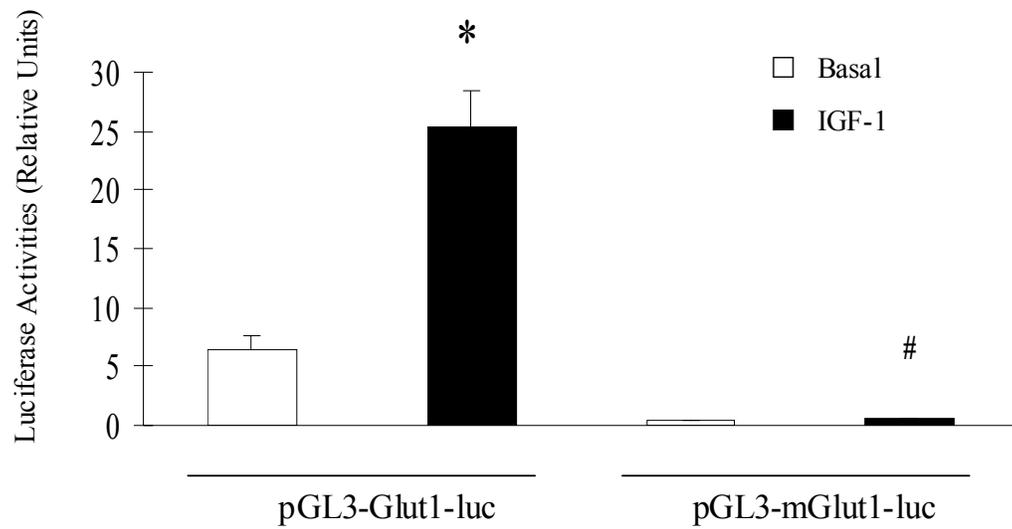


Fig. 3C, 3D

A.

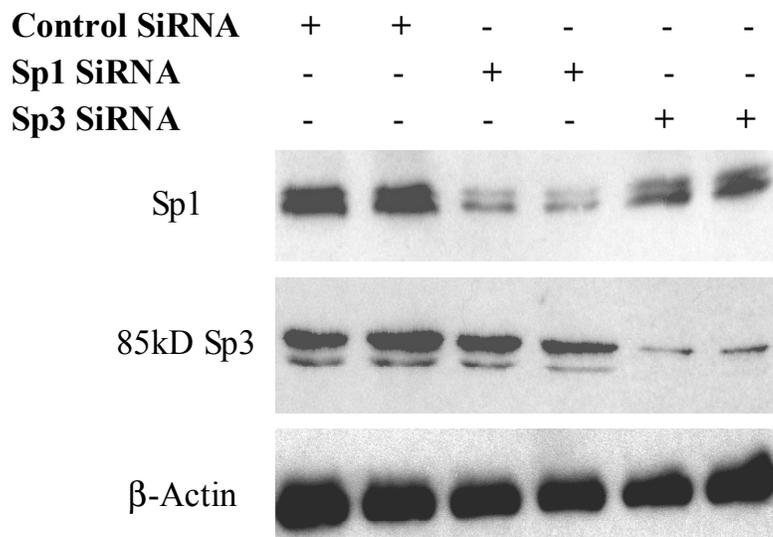


Fig. 4A

B.

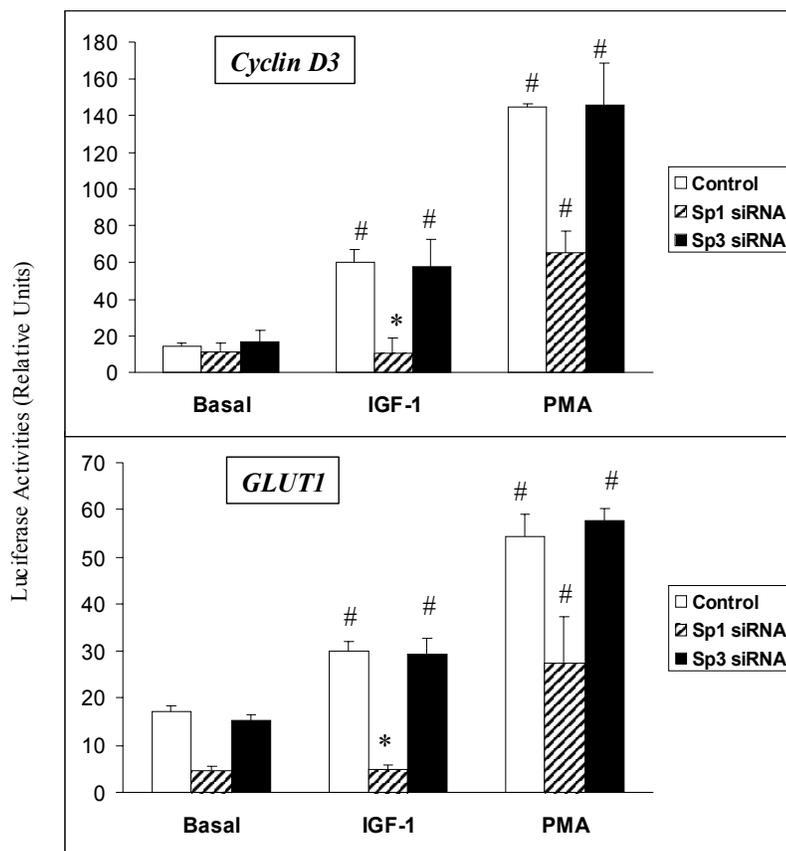


Fig. 4B

C.

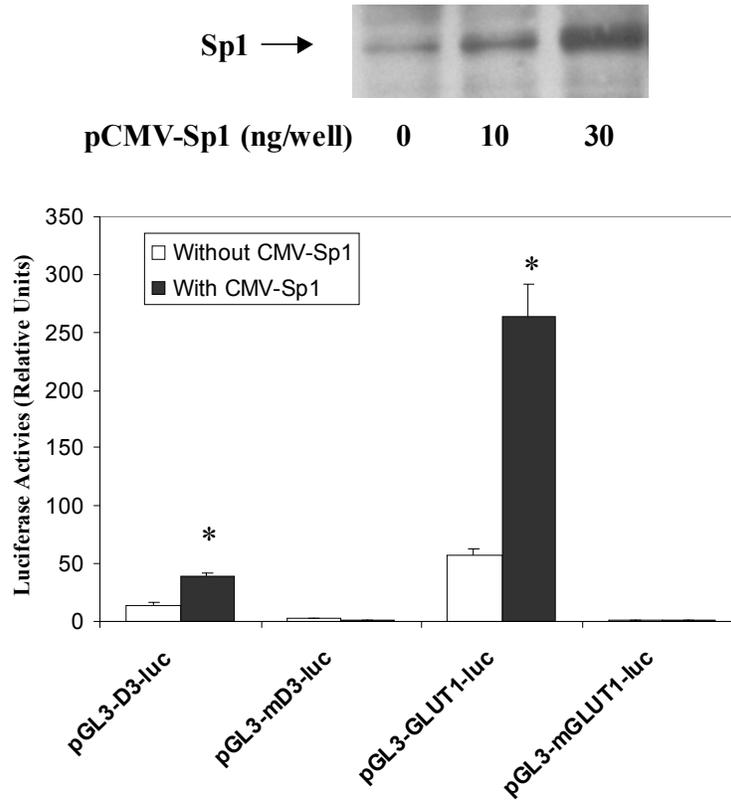


Fig. 4C