TRANSCRIPTIONAL REGULATION OF BMP-2 ACTIVATED GENES IN OSTEOBLASTS USING GENE EXPRESSION MICROARRAY ANALYSIS: ROLE OF DLX2 AND DLX5 TRANSCRIPTION FACTORS

Stephen E. Harris [1], Dayong Guo [1], Marie A. Harris [1], Arvind Krishnaswamy [1], and Alex Lichtler [2]

[1] Department of Oral Biology, University of Missouri at Kansas City, School of Dentistry, Kansas City MO 64108, [2] University of Connecticut Medical Center, Farmington, Connecticut 06030

TABLE OF CONTENTS

1. ABSTRACT

This presentation will focus on using microarray data on a clonal osteoblast cell model to analyze the early BMP-2 responsive genes, as well as some of the later genes regulated by BMP2 during different phases of mineralization. We will focus on the early phases of gene expression that occur after BMP2 signaling from 30 min up to 1 day. The hypothesis is that understanding how these early genes are regulated during the initial multilayering and growth phase of osteoblasts will lead to models of how BMP activity stimulates cell growth, cell migration, multilayering, matrix deposition and remodeling phase that allows subsequent mineralization. The Dlx2 and Dlx5 homeobox genes have been shown to be critical for bone formation both *in vitro* and *in vivo*. Both Dlx 2 and Dlx5 are activated within 15-30 minutes after BMP2 addition to the mouse 2T3 osteoblast model and primary fetal rat calvarial osteoblasts.   The Dlx2 and Dlx5 genes stay elevated in the presence of BMP2 for up to 5 days, a time when overt mineralization is just beginning. To understand the genomic network that Dlx5 and Dlx2 regulate at the transcription level, we have taken an approach where we use a specific transcription repressor protein, Engrailed, ligated to the Dlx5 homeodomain. The idea is that this Eng-Dlx5 protein will interact with Dlx5 and possibly Dlx2 and related Dlx- regulated genes *in vivo* and down-regulate their transcriptional initiation. Using a microarray approach with over 5,000 known genes we can identify the genes that are directly and indirectly regulated by Dlx5 and Dlx2. This will allow us to build an initial genomic network of Dlx- regulated genes at the transcriptional level. We will present our model and preliminary efforts at understanding the genomic network regulated by this important BMP2-regulated transcription factor class in osteoblast biology.

2. INTRODUCTION

Experimental genomics in combination with the vast amount of new DNA sequence knowledge has drastically changed the approaches we can use to understand how cells grow and differentiate and in particular how an osteoblast differentiates and responds to such factors as bone morphogenetic protein 2 or BMP2, a member of the TGF-beta family with the potent capacity to stimulate bone *in vivo* and *in vitro*.

One of the major tools of experimental genomics is the use of DNA microarrays in which thousands to tens of thousands of cDNAs can be arrayed on filters or glass slides and used to monitor the expression of all of the genes on the gene "chip" or microarray in parallel (1,2). These gene-expression-profiling methods are revolutionizing such areas as basic yeast biology, C. elegans biology, cancer research, neural biology, aging research, and developmental biology (1,3,4,5,6,7,29). Genetic network architecture can and will be determined in the future (8,9). For example, in yeast, the entire coding genome can be placed on one array and now monitor over 300 diverse mutations and chemical treatments compared with wild-type expression patterns of all the yeast genes.   A "compendium" of expression profiles was developed that

became a powerful tool to predict which pathway is affected by various mutations. The authors of this study also demonstrated that the compendium could be used to characterize pharmacological perturbation by identifying new gene targets for commonly used drugs (Hughes *et al*, 2000). In this presentation we will present a small compendium of osteoblast genes that change during the first few hours after activating osteoblasts by BMP2. We will then show how one can begin to determine the genomic regulatory networks, using the early BMP2 responsive gene, Dlx5, as an example.

Various bioinformatic tools for analysis of these complex datasets are rapidly evolving (10,11). Many of the computational tools involve clustering algorithms for clustering patterns of expression and are available from the developers of these programs. These can be obtained for academic use without cost (GENECLUSTER www.wi.mit.edu/mpr/GeneCluster/GeneCluster.html;CLUSTER and TREE VIEW from http://rana.stanford.edu/clustering or www.dnachip.org/index.shtml or Michael Eisen's lab at UC Berkeley). Statistical programs are available to evaluate the reproducible and significance of measured changes in gene expression that do not necessarily depend on some arbitrary change in expression levels (12). An example of this type of program, SAM or statistical analysis of microarrays, will be demonstrated here in our osteoblast studies of early BMP2 responsive genes.

Gene expression microarray tools are now being applied to the field of bone biology. Using microarrays with 600 well characterized known mouse genes and arrays with 5,000 to 10,000 mouse known genes, we are developing a compendium of gene expression patterns during osteoblast differentiation using a clonal mouse cell model that responds to BMP2 and produces a well-organized mineral matrix. The use of a simple clonal cell system in which the cells can undergo the entire cascade of the stages of osteoblast differentiation will yield a useful database for developing experiments to test hypotheses on key pathways that must be integrated to produce a bone like mineralized matrix. The system will also allow one to explore the downstream genes that are regulated by almost any transcription factor. With new bioinformatics tools such as rVISTA (13), we can analyze the conserved nucleotide sequences in mouse and human genes and other species in any gene. This allows one to focus in on candidate cis-regulatory regions in a vast sea of non-conserved DNA sequences around a gene. The programs VISTA and rVISTA are available at http://www-gsd.lbl.gov/VISTA. Once you have run a alignment of a mouse and human DNA sequence, you can then process those files through rVISTA and systematically search up to 10 or 20 kb of the potential cis regulatory regions in and around the gene using over 276 transcription factor matrix binding sites. The program will now allow one to input your own DNA binding site or position weighted matrix (PWM). rVISTA will then show only the core binding sites at a parameter that you can set and with the extra requirement that 20 bp flanking the core transcription factor binding site has at least 80% homology. Important cis-regulatory regions or modules (CRM) often show clusters

of 2 or more of the same or similar sites within a 300-700 bp region of the gene (14). We will show an example of the read-out using the Dlx 5/6 locus. We have also analyzed over 50 other early BMP 2 responsive genes using rVISTA for development of a candidate Dlx transcription factor database for future use.

We have shown that BMP2 induces Cbfa1 and Dlx5 as well as Dlx2 within 30 min to 4 hours after the addition to 2T3 osteoblasts at the confluent single cell layer state (t = 0). The question is, can we find clusters of Cbfa1 binding sites or Dlx homeobox binding sites in the conserved non-coding sequence regions or CNSs of these genes? The answer is yes and all three genes could be autoregulated as well as cross-regulated. We then used a specific repressor of Dlx responsive genes, Eng-Dlx5, in which a strong Engrailed repressor domain is fused to the Dlx5 homeodomain DNA binding region of the protein. This construct lacks all Dlx5 activator domains and only contains the DNA binding homeodomain. The Eng-Dlx5 should block gene transcription from Dlx5 and possibly, other genes regulated by Dlx family members. We will test the hypothesis that genes suppressed by Eng-Dlx5 are directly regulated by Dlx5 or Dlx2 or possibly by other Dlx family members. This will be the beginning of defining the Dlx gene network in early osteoblast growth and multilayering phases. Since Dlx5 and Dlx2 remains elevated for up to 5 days after BMP2 addition, these Dlx family members likely play multiple roles at different stages of osteoblast differentiation.

In future experiments we will attempt to activate the Eng-Dlx5 specific repressor at later stages of osteoblast differentiation and then look at the downstream genes that are directly regulated by Dlx family members. These functional and bioinformatic approaches to define genomic networks for a given transcription factor will go hand in hand with other methods such as specific CHIP assays or chromatin immunoprecipitation. However, what is required in the CHIP assays is a good precipitating antibody to the given transcription factor and an efficient method for sorting through millions of DNA fragments of about 1,000 bp in length. One could use rVISTA or other programs such as Family Relations (15) to find candidate cis regulatory sites or modules (CRMs). At present we need to develop a good antibody to Dlx5 and Dlx2 that we could use in the CHIP assays.

3. MOUSE MODEL FOR STUDYING GROWTH AND DIFFERENTIATION OF OSTEOBLASTS

The clonal osteoblast model, 2T3, was developed from a transgenic mouse containing the weak BMP2 promoter driving T-antigen (16). The clonal cell line is immortal but not tumorogenic. The 2T3 model produces a well-organized mineralized matrix between 14 and 20 days and BMP-2 greatly accelerates this process. In BMP2 treated cultures, mineralized matrix first starts to form at 6-7 days and over 70% of the culture is mineralized by 20 days, with the formation osteocytes in a mineralized matrix. All stages of osteoblast differentiation can be studied in this model. In 2T3 model, osteoblasts differentiate into pre-

osteocytes and mature osteocytes *in vitro*. Osteocytes are the cells *in vivo* that differentiate from osteoblasts and become buried in the bone matrix and are thought to be the predominate bone cells that sense mechanical perturbation and sends signals to initiate new bone formation or repair., depending on the level of strain experienced by the bone.

Plasmids with various mutant and wild-type genes can be stably transfected into the 2T3 model, the cells recloned and then used to study the effect of the mutant gene on osteoblast differentiation. With some care the cells can be transiently transfected with over 80% efficiency (17; Harris, unpublished). This quality in fact allows us to study the short-term effects of various gene constructs, such as specific RNAsi (small interfering RNAs) and Engrailed-Dlx5HD (specific repressor of Dlx-regulated genes), on early aspects of BMP action in osteoblasts. We have also used the 2T3 model to study the effect of various mutant BMP receptors on BMP2 signaling pathways and on osteoblast differentiation. We discovered a differential role for BMP receptor IB and BMP receptor IA on both osteoblast differentiation and adipocyte differentiation using the 2T3 model (17). A truncated or dominant negative BMPR-IB (dn or trBMPR-IB) completely inhibited osteoblast differentiation to a mineralized matrix state without affecting growth, with or without BMP2 treatment. A constitutively active BMPR-IB (active without added BMP2 ligand) stimulated mineralized matrix production, in the absence of BMP2. By comparison of gene expression profiles in wild-type cells and mutant cells, insights into the pathways important for osteoblast differentiation can be obtained in a fairly straight-forward manner. , We have also used the 2T3 model to efficiently infect Eng-Dlx5HD with a retroviral vector with a GFP tag and then study the Dlx transcription factor gene regulatory network.

Recently, we have used the 2T3 model to analyze early and late BMP2 responsive genes using gene expression microarrays, first with arrays of 600 well annotated genes in defined functional categories, and later with arrays containing 5,000 well annotated genes. Some of these data sets will be used to demonstrate how we have approached the analysis of these complex data sets and methods for analysis of genomic networks in osteoblast biology.

4. EARLY BMP2 RESPONSIVE GENES USING GENE EXPRESSION ARRAYS AND STANDARD GENE CLUSTERING PROGRAMS

Analysis of global gene expression changes during early responses to the growth and differentiation factor, BMP2, is somewhat less complicated than analysis of complex tissues *in vivo*. But as we will show, as the osteoblast differentiate in culture, there are are states where cells are rapidly growing and form multilayers that will lead to focal mineralized structure and other areas where cells are not differentiating or differentiating at a slower rate. Even in this simple system, the complexity of the analysis of large numbers of gene expression patterns over time becomes a problem that will have to be

complimented by immunochemical and *in situ* hybridization procedures to localize the gene or protein expression changes. However the early responses to BMP2 within the first 8 hours to 1 day are less complicated and easier to interpret. Models of what BMP2 is initially doing to an immature and undifferentiated osteoblast and the role of the early BMP2 induced transcription factors, Dlx5 and Dlx2, will be the focus of this study.

We first analyzed a set of early BMP2 responsive genes in 2T3 cells from 30 min to 8 hours, using arrays from Clontech that contained about 600 well-characterized mouse genes in a variety of functional categories. Based on an Error Model in which four independent control (0 hour) experiments were compared against each other, we found a 2.5 fold change in expression between control and BMP2 treated samples had a greater than 95% chance of being statistically significant (18). This estimate is based on the average of the whole data set in which most genes are not changing to any significant degree. With this associative t-test analysis we may have as many as 20-25% false positives. With small data sets of a few hundred genes, we can deal with these false positives by other validation methods, such as Northern analysis or Real-time RT-PCR. For larger arrays with 5K to 10K genes other statistical methods can be applied to reduce the false positives and not, at the same time, create many false negatives. What makes our data more robust is the fact that we have analysis of 5 different time points with replicates for most of the time points. When a gene changes at 2-3 of the time points in independent experiments we can be more confident of the observations.

We clustered the data on 126 genes that changed expression greater than 2.5 fold. The red colored squares represent genes that were up-regulated by BMP2 and green represents down-regulated genes in the Tree View graphs. We used both hierarchical clustering method (10) and self-organizing mapping cluster method (11). The results of the hierarchical cluster analysis of early BMP2 responsive genes using the 2T3 model are shown in Figure 2A-D.

All original data and others in this presentation can be found at http://www.umkc.edu/dentistry/oralbiology/body_faculty_steve_files.html. A dataset for early responsive BMP2 genes is in Microsoft Excel format and can be downloaded for further manipulations and analysis. A description of each file is given within the Excel file or a readme.txt or .doc file. A set of the Raw Data from 41 experiments is also available with GeneSymbol, MGI number, Gene Description, Accession No. and functional classification codes and can also be found at this site for building your own analysis criteria and sorting of the data.

Physiologically, what happens in the first 24 to 72 hours after treatment of primary fetal rat osteoblast cultures (19) or 2T3 osteoblast with BMP2 is a selective growth or multilayering of cells and an apparent reorganizing of the extracellular matrix where there is migration of cells and piling up along ridges that will be the first areas to mineralize into "nodules" or mineralized areas
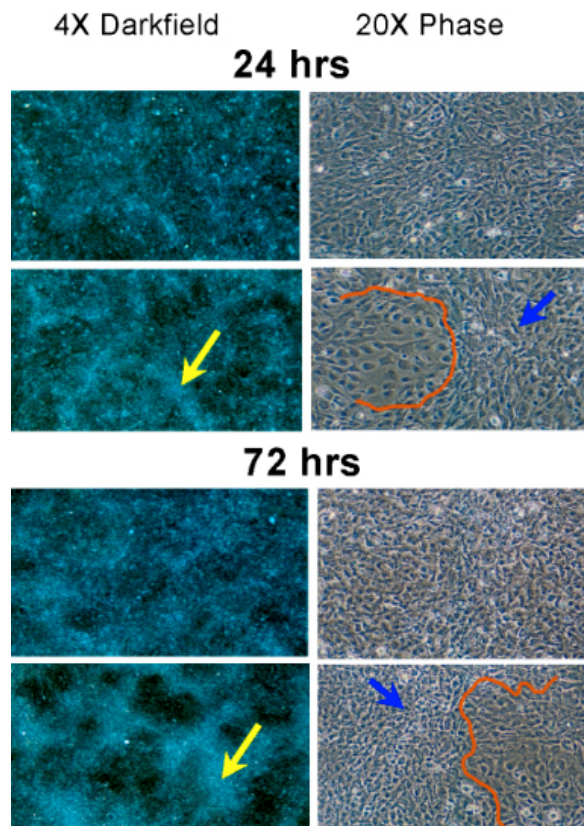
4X Darkfield   20X Phase
24 hrs

72 hrs

Figure 1. BMP2 stimulates differential growth and early responsive multilayering during the first 72 h. Arrows indicated differential growth and cell shape changes. Red line is the boundary of migration ridge. Upper panel: Control, Lower Panel: BMP-2, 100 ng/ml

(Figure 1). Even by 1 to 3 days it is apparent that this initial clonal system is becoming less homogenous or anisotropic in its responses. Major changes in the extracellular matrix are underway. This will require localization of gene expression changes by *in situ* hybridization or immunological methods in the future. Nevertheless useful information can be obtained and hypotheses developed that increase our understanding of how BMP2 is such an effective factor at producing new bone.

Associated with the selective growth or multilayer cell formation phenomena is the induction at 30 min to 8 hrs in a large number of genes associated with growth and the cell cycle, as well as genes associated with matrix reorganization. With this data we used the hierarchical clustering program from Eisen's Lab (http://www.rana.lbl.gov/EisenLab). In the program called Cluster there are a variety of clustering methods and ways to organize your input data. The TreeView, which is a separate graphical presentation program and is a part of this package which one can get from Eisen's lab at University of California at Berkley, allows one to graphically present the data, as shown in Figure 2. Other methods for data analysis and presentation will also be discussed. Figure 2A

shows the overall gene-clustering pattern for this set of 126 genes whose expression increase or decreases within the first 30min to 8h after adding BMP2 to 2T3 osteoblasts.

As noted in the expanded view or zoom of Cluster 2 and Cluster 6 in Figure 2B and Figure 2C, the figure shows a short description of the gene (eg laminin receptor 1) and Accession No., along with the expression pattern from 0 to 8 h after addition of 100 ng/ml of human recombinant BMP2.

We have found it is necessary that a unique identifier for each gene should be included in the initial gene list, such as the Gene Symbol, Locus Link, or for mouse, the MGI Number. After analysis and sorting of gene expression patterns, it is extremely important to have this unique identifier in your gene set for further analysis and sorting based on gene ontology, pathways, and functional classes, and various links on the internet. Most of this information can be found at locus link at the NCBI web site.

As noted in Figure 2B(and in Cluster 1, see http://www.umkc.edu/dentistry/oralbiology/body_faculty_s teve_files.html/ with all the original data and expanded diagrams), many of the immediate early response genes to BMP2 are associated with growth phenomena and signaling and are consistent with the observed physiological phenomena of selective BMP2 induced multilayer cell formation (eg. Egr1, PCNA, c-jun, cyclin D1, cyclin A, and even cyclin kinase inhibitors such a p19ink, etc.). It must be remembered that the osteoblast cultures are not synchronized with respect to the cell cycle and at t = 0, there are a mixture of cells at different cell cycle states.

In Figure 2C, a zoom of Cluster 6 of Figure 2A is shown for a collection of genes that on the average remain elevated up to 8 h after BMP2 addition. This simple set of data also demonstrates the value of having as many time points or conditions in the initial design of your experiment. Also, as noted earlier, most of these time points represent at least two independent experiments and we are presenting the average of each time point. Analysis of significance of each gene change at a given time point will be shown below, for at least the data in which the cells were treated with BMP2 for 30min. We used the SAM program for this analysis (12).

A large number of these genes in which expression remains elevated up to 8 h are also involved in growth pathways (9 out of 32), inhibition of cell death (4 of 32), matrix alterations (3 of 32), and transcription (7 or 32), including two homeobox genes, Hox A7(1.1) and the muscle specific Lbx1 gene. Of particular interest is the Akt gene, which, besides inhibiting apoptosis, plays a role in activating eNOS, a known stimulator of bone formation.

Another example of the diverse response of osteoblasts to a BMP2 signal is shown in Figure 2D, an expanded view or zoom of the Cluster 5 marked in Figure 2A.
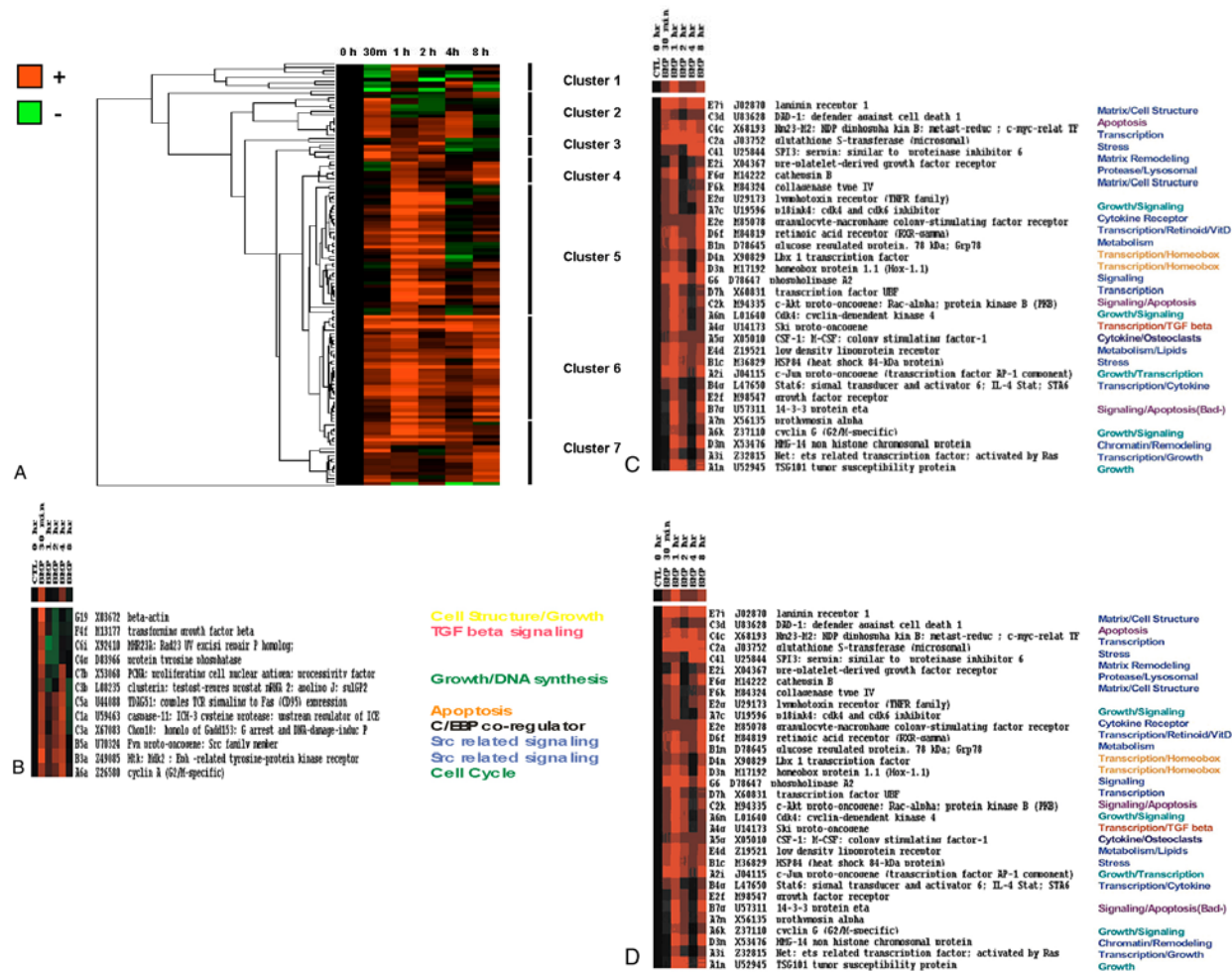
0 h 30m 1 h 2 h 4 h 8 h

+
-

Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5
Cluster 6
Cluster 7

A

C

E7i J02870   laminin receptor 1 — Matrix/Cell Structure
C3d U83628   DAD-1: defender against cell death 1 — Apoptosis
C4c X68193   Nm23-M2: NDP dinkosnha kin B: metast-reduc : c-myc-relat TF — Transcription
C2a J03752   glutathione S-transferase (microsomal) — Stress
C4l U25844   SPI3: serpin; similar to proteinase inhibitor 6 — Matrix Remodeling
E2i X04367   pre-platelet-derived growth factor receptor — Protease/Lysosomal
F6d M14222   cathepsin B — Matrix/Cell Structure
F6k M84324   collagenase type IV
E2u U29173   lymphotoxin receptor (TNFR family) — Growth/Signaling
A7c U19596   p18ink4: cdk4 and cdk6 inhibitor — Cytokine Receptor
E2e M85078   granulocyte-macrophage colony-stimulating factor receptor — Transcription/Retinoid/VitD
D6f M84819   retinoic acid receptor (RXR-gamma) — Metabolism
B1n D78645   glucose regulated protein. 78 kDa; Grp78 — Transcription/Homeobox
D4a X90829   Lbx 1 transcription factor — Transcription/Homeobox
D3n M17192   homeobox protein 1.1 (Hox-1.1) — Signaling
G6   D78647   phospholipase A2 — Transcription
D7k X60831   transcription factor UBF — Signaling/Apoptosis
C2k M94335   c-Akt proto-oncogene: Rac-alpha: protein kinase B (PKB) — Growth/Signaling
A6n L01640   Cdk4: cyclin-dependent kinase 4 — Transcription/TGF beta
A4g U14173   Ski proto-oncogene — Cytokine/Osteoclasts
A5u X05010   CSF-1: M-CSF: colony stimulating factor-1 — Metabolism/Lipids
E4d Z19521   low density lipoprotein receptor — Stress
B1c M16029   HSP84 (heat shock 84-kDa protein) — Growth/Transcription
A2i J04115   c-Jun proto-oncogene (transcription factor AP-1 component) — Transcription/Cytokine
B4r L47650   Stat6: signal transducer and activator 6: IL-4 Stat: STA6
E2f M98547   growth factor receptor — Signaling/Apoptosis(Bad-)
B7u U57311   14-3-3 protein eta
A7m X56135   prothymosin alpha — Growth/Signaling
A6k Z37110   cyclin G (G2/M-specific) — Chromatin/Remodeling
D3n X53676   HMG-14 non histone chromosomal protein — Transcription/Growth
A3i Z32015   Net: ets related transcription factor: activated by Ras — Growth
A1n U52945   TSG101 tumor susceptibility protein

B

G19 X83672   beta-actin
F4f M13177   transforming growth factor beta
C6i X92410   MHR23A: Rad23 UV excisi repair P homolog;
C4e D43966   protein tyrosine phosphatase
C7b X53068   PCNA: proliferating cell nuclear antigen: processivity factor
C3h L08235   clusterin: testost-repres prostat mRG 2: apolipo J: sulGP2
C5a U44088   TDAG51: complex TCR stimulatio to Fas (CD95) expression
C1a U59463   caspase-11: ICH-3 cysteine protease: upstream regulator of ICE
C3a X67083   Chop10: homolo of Gadd153: G arrest and DNA-damage-induc P
B5a U70324   Fyn proto-oncogene: Src family member
B3a Z49005   Htk: Mdk2 : Enh -related tyrosine-protein kinase receptor
A6a X26500   cyclin A (G2/M-specific)

Cell Structure/Growth
TGF beta signaling

Growth/DNA synthesis

Apoptosis
C/EBP co-regulator
Src related signaling
Src related signaling
Cell Cycle

D

E7i J02870   laminin receptor 1 — Matrix/Cell Structure
C3d U83628   DAD-1: defender against cell death 1 — Apoptosis
C4c X68193   Nm23-M2: NDP dinkosnha kin B: metast-reduc : c-myc-relat TF — Transcription
C2a J03752   glutathione S-transferase (microsomal) — Stress
C4l U25844   SPI3: serpin; similar to proteinase inhibitor 6 — Matrix Remodeling
E2i X04367   pre-platelet-derived growth factor receptor — Protease/Lysosomal
F6d M14222   cathepsin B — Matrix/Cell Structure
F6k M84324   collagenase type IV
E2u U29173   lymphotoxin receptor (TNFR family) — Growth/Signaling
A7c U19596   p18ink4: cdk4 and cdk6 inhibitor — Cytokine Receptor
E2e M85078   granulocyte-macrophage colony-stimulating factor receptor — Transcription/Retinoid/VitD
D6f M84819   retinoic acid receptor (RXR-gamma) — Metabolism
B1n D78645   glucose regulated protein. 78 kDa; Grp78 — Transcription/Homeobox
D4a X90829   Lbx 1 transcription factor — Transcription/Homeobox
D3n M17192   homeobox protein 1.1 (Hox-1.1) — Signaling
G6   D78647   phospholipase A2 — Transcription
D7k X60831   transcription factor UBF — Signaling/Apoptosis
C2k M94335   c-Akt proto-oncogene: Rac-alpha: protein kinase B (PKB) — Growth/Signaling
A6n L01640   Cdk4: cyclin-dependent kinase 4 — Transcription/TGF beta
A4g U14173   Ski proto-oncogene — Cytokine/Osteoclasts
A5u X05010   CSF-1: M-CSF: colony stimulating factor-1 — Metabolism/Lipids
E4d Z19521   low density lipoprotein receptor — Stress
B1c M16029   HSP84 (heat shock 84-kDa protein) — Growth/Transcription
A2i J04115   c-Jun proto-oncogene (transcription factor AP-1 component) — Transcription/Cytokine
B4r L47650   Stat6: signal transducer and activator 6: IL-4 Stat: STA6
E2f M98547   growth factor receptor — Signaling/Apoptosis(Bad-)
B7u U57311   14-3-3 protein eta
A7m X56135   prothymosin alpha — Growth/Signaling
A6k Z37110   cyclin G (G2/M-specific) — Chromatin/Remodeling
D3n X53676   HMG-14 non histone chromosomal protein — Transcription/Growth
A3i Z32015   Net: ets related transcription factor: activated by Ras — Growth
A1n U52945   TSG101 tumor susceptibility protein

Figure 2. A. Hierarchical Cluster analysis or 126 genes whose expression changes > 2.5-fold by BMP2, from 30 min to 8 h, in the 2T3 Osteoblast model — Overall view using TreeView program. B. Hierarchical cluster analysis –immediate early responsive BMP2 genes. C. Hierarchical cluster analysis-early response and remain activated up to 8 h after BMP2 treatment. D. Hierarchical cluster analysis of genes increased at 1 h after BMP and decrease at later time points.

Of the 39 genes in Cluster 5 and 9 are associated with growth or growth associated pathways, 19 of the 39 genes in this Cluster 5 are associated with transcription, and 3 of these 19 are belonged to the Kruppel family of transcription factors. This coordinate response suggests important roles for the Kruppel family of transcription factors in osteoblast differentiation and is worth further exploration. Many of the Kruppel family of Zn-finger genes have been shown to be involved in regulation of various homeobox genes. This type of coordinate response is useful for designing focused experiments on determining which genes may be downstream of Kruppel expression. For example, we found that BMP2 induced HoxA7 with a similar time frame.

We have begun attempts at functional classification of these data sets using the approximately 160 functional codes that Clontech has developed for these 588 genes. This will allow one to cluster genes based on functional categories and will assist in developing pathway maps. The functional classification codes and lists can be found at the Clontech.com and in our raw data files (http://www.umkc.edu/dentistry/oralbiology/body_faculty_steve_files.html).

5. SELF-ORGANIZING MAPS (SOMS) OF EARLY BMP2 RESPONSIVE GENES IN THE 2T3 MODEL

Another complimentary method to organize and visualize complex expression datasets is the use of self-organizing maps or SOMs. SOM program is another clustering method based on a different algorithm than used in hierarchical method or K-mean method of clustering (11). It is a more supervised clustering method and gives complementary data to the other gene clustering programs. A link to the SOM program developed at the Whitehead Institute can also be found at http://www.dnachip.org/restech.html. The program is called GeneCluster. We have taken a similar data set used in the hierarchical method and applied the 2 dimensional SOM analysis. As shown in Figure 3A, a cluster, labeled c7 or cluster7, and outlined in the yellow box, contains 11

G...	Descriptions
1	A1a M88127 APC (adenomatous polyposis coli protein)
...	B4b U33626 Pml; murine homolog of the leukemia-associated PML gene
...	C2d D30687 GST Pi 1; glutathione S-transferase Pi 1; preadipoce GF
...	C4a D17571 NADPH-cytochrome P450 reductase
...	C4d J04113 Nur77 early response protein; thyroid hormone (TR3) receptor
...	D2i M20157 Egr-1 Zn-finger regulatory protein
...	D4l J03168 interferon regulatory factor 2 (IRF 2)
...	F4f M13177 transforming growth factor beta
...	F6d X51438 vimentin
...	F6h X53337 cathepsin D
...	G19 X03672 beta-actin

A

G...	Descriptions
20	A2i J04115 c-Jun proto-oncogene (transcription factor AP-1 component)
53	A5g X05010 CSF-1; M-CSF; colony stimulating factor-1
71	A6m L01640 Cdk4; cyclin-dependent kinase 4
89	B1c M36829 HSP84 (heat shock 84-kDa protein)
...	B3c M83336 interleukin-6 receptor beta chain; membrane glycoprotein gp130
...	B4g L47650 Stat6; signal transducer and activator 6; IL-4 Stat; STA6
...	C2a J03752 glutathione S-transferase (microsomal)
...	C2k M94335 c-Akt proto-oncogene; Rac-alpha; protein kinase B (PKB)
...	C4c X68193 Nm23-M2; NDP diphospha kin B; metast-reduc ; c-myc-relat TF
...	D3n M17192 homeobox protein 1.1 (Hox-1.1)
...	D7h X60831 transcription factor UBF
...	E4d Z19521 low density lipoprotein receptor
...	F6k M84324 collagenase type IV

B

Figure 3. A. SOM analysis of early BMP2 responsive genes in 2T3-Cluster 7 or C7. Note the yellow box with 11 genes that increase by BMP2 at 30 min and then decrease over time. The red line represents the standard deviation of the expression levels of the 11 genes in Cluster 7. B. SOM analysis of Cluster 4 or C4, a set of 13 genes whose expression peaks at 1 h and remain activated after BMP2 up to 8 h.

genes that are immediate BMP2 responsive genes (ie up at 30min and decrease after time). 7 out of 11 of this set were also found in Cluster 1 and 2 using the hierarchical method and also indicates acceleration of the genetic program for growth or multilayer cell formation.

Another example of how SOMs can be utilized is shown in Figure 3B. We could identify a set of genes whose expression peaks at 1 h after BMP2 treatment. We can compare the genes modulated at 30 min with the genes modulated at 1 h and other clusters such as C2 or Cluster 2 with 11 genes whose expression peaks at 2 h after BMP2 treatment.

CSF-1 was identified in this cluster and the CSF-1 promoter has recently been shown to have a direct BMP response region as well as AP-1 sites that respond to fos-jun complexes (20). Hox A7 (Hox 1.1) is in this class, as well as c-jun and Akt, both involved in growth and survival pathways. We hypothesize that Hox A7 is involved in early osteoblast specification and upstream of Cbfa1 gene. This is based on the fact that many of the Hox A class are expressed very early in mesenchymal tissue from a large number of tissue sites such as somites, kidney, stomach, and a variety of vertebrae and rib precursors, as well as in neuronal precursors and neural crest cells (21.22). Also many of the Hox A group are most likely regulated by
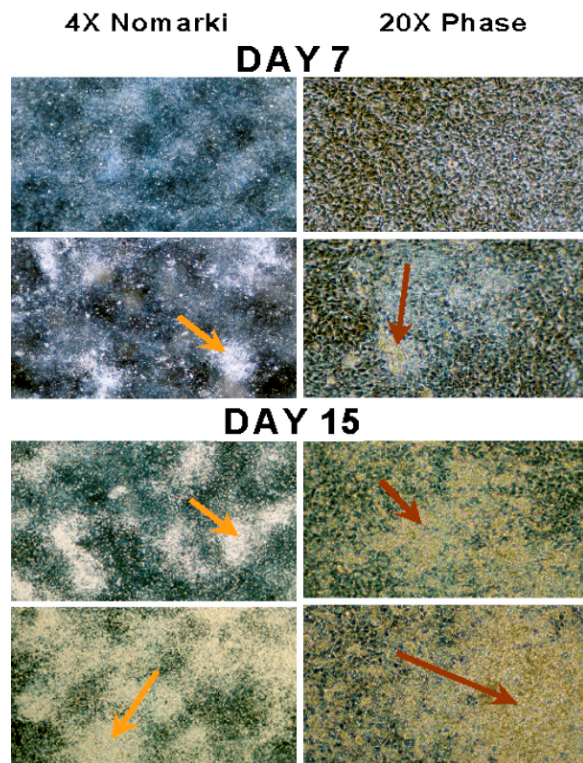
Figure 4. BMP2 accelerates the formation of a well organized Fractal – like mineral matrix with the mineral associated with collagen fibrils. Arrows point to mineral. Upper panel: Control, Lower panel: BMP-2 100 ng/ml.

members of the Kruppel family of transcription factors that we observed earlier and were induced by BMP2 at 100 ng/ml.

6. GENE EXPRESSION PROFILES OF 137 GENES WHOSE EXPRESSION IS ALTERED BY TIME OR BMP2(> 2.5 FOLD) DURING OSTEOBLAST DIFFERENTIATION - 1 TO 15 DAYS AND COMPARISON TO MUTANT 2T3 WITH A DOMINANT NEGATIVE BMPR-IB CONSTRUCT

Another example of how we can gain some insight into osteoblast differentiation is shown below where we analyzed BMP2 regulated genes from 1 day to 15 days. In these long-term experiments it was critical to have a control at each time point. It is not that the control culture is static. The control culture is also multilayering and differentiating into mineralized structures except that it is occurring at a slower rate. This slow versus fast rate of differentiation will eventually allow us to use various phase analysis to find sets of genes that are out of phase for some few days and those subsets should be very informative. An example of what phenomena of osteoblasts differentiation are occurring in these cultures is shown below. This is highly reproducible and sufficient mineralized structures are formed to allow a global analysis of gene expression pattern to be derived in the near future.

Figure 4 shows that by Day 7, the BMP2 treated cultures are producing detectable mineralized matrix, while the control untreated cultures show no mineral. This can be seen in the phase contrast where the mineral is slightly yellow and sponge-like or fractal in nature, and in the Nomarki/Darkfield photographs where the mineral structure is strongly light refractal.

By Day 15, the BMP 2 treated cultures are over 50% mineralized (at least in 2 dimensions), while the control culture is beginning to form mineralized matrix or nodules in the absence of BMP2. This "control" mineralization can however be blocked by 0.5 ng/ml of noggin, indicating endogenous BMP activity is required for the control cultures to mineralize. The mineralized structures are formed between the cell layers on a collagen matrix (16,17).

We would like to demonstrate and present the current data in a form where one can visualize the BMP2 regulated set. This will allow us to define some of the components of BMP2 stimulated bone cell differentiation program. We first normalized all the control values at the various time points to 1.0 for both the wild-type and mutant datasets. Then we looked at this same set of 137 genes out of about 527 that were on the array and gave a valid hybridization signal. This data set can be found at http://www.umkc.edu/dentistry/oralbiology/body_faculty_s teve_files.html (LateBMP2Responsive2T3Osteoblast.zip), and can be downloaded and used in various Clustering programs. Excel file for the total 527-gene set, before sorting for regulation is in this zip file. The main point in this example is the clustering programs, which allow the user to look and explore the data in a variety of ways. Other simple Excel equations can be written to explore subsets of this file.

The TreeView presentation program of this normalized data is shown in Figure 5A, 5B, and 5C. A zoom of Cluster 1 is shown in Figure 5B, which shows a set of genes that are in general negatively regulated at Day 1 by BMP2 but increase in expression dramatically at Day 5 to 7 in the BMP2 treated cultures. However in the Mutant dnBMPR-IB 2T3 cells, this set of genes shows little change or decreases with BMP2 treatment. For example, the transcription factors, Sox3, Kruppel-like factor GKLF, and split hand/foot gene show this pattern. BMP1 increases at 5 to 7 days in the wild-type cells but is decreased by BMP2 in the Mutant cells. BMP1 is a protease and important in regulating BMP activity by degrading negative regulators of BMP action, such as chordin and to a lesser degree, gremlin and noggin.

Cluster 2 from Figure 5A is shown in Figure 5C and contains a subset of BMP2 regulated genes that increase (red) at Day 1 to 3 and at Day 7. Many of these genes are negatively regulated (green) by BMP2 in the Mutant 2T3 cells. Many pathways suggested by these experiments are involved in BMP stimulated bone formation, including the Wnt pathway (EB 1 APC binding protein), TGFβ/Activin pathways (Smad2), stress responses (HSP27, GST Mu1), matrix interaction (integrin b, N-cadherin), cell structure(β-actin), growth pathways (PI3-Kinase p85 regulatory subunit—linked to PDGF and Akt,
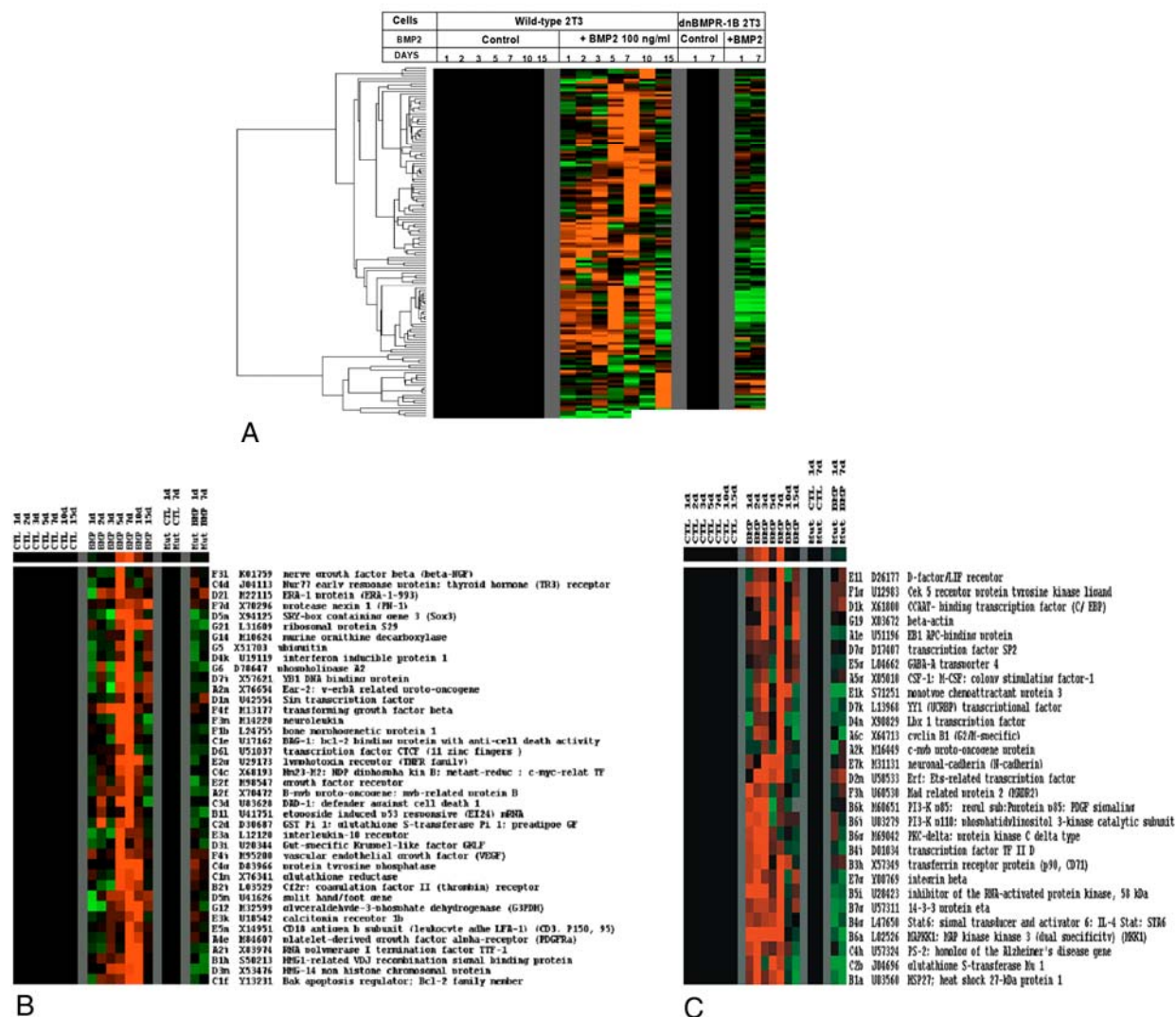
Figure 5. A. Hierarchical cluster analysis of BMP2 regulated Genes (137). All control values are normalized to 1.0 and appear as a uniform black region in Cluster diagram. B. Cluster 1 from Fig. 6A with a subset of genes that are increased by BMP2 at 7 to 10 days in the wild-type 2T3 but not in the mutant dnBMPR-IB 2T3. C. Cluster 2 of Fig.10 is enriched in a set of genes whose expression is increased at 1 to 7 days in wild-type cells but is negatively regulated at 1 and 7 days in the mutant 2T3 cells with the dn-BMPR-IB.

cyclin B1, MAPKK1), transcription (general TF IID, homeobox gene Lbx1, YY1, Stat6), and chemoattractant (MCP-3). This brief list demonstrates the complex responses that underlie BMP action in this simple clonal osteoblast model with a simple array system.

7. EVALUATION OF STATISTICAL SIGNIFICANCE OF INDIVIDUAL GENE EXPRESSION CHANGES USING SAM

As with any biological measure, it is important to have some measure of the variance in individual gene expression changes in microarray data. Ideally, one would like to have completely independent experiments, but at a minimum have the same RNA analyzed on independent arrays. A program called significance of analysis of microarrays or SAM is a useful statistical approach that does not depend on some arbitrary change in fold expression but looks at the relationship of the change in expression between two states or conditions and the variance of the replicate experiments (12). SAM also has an algorithm to permutate the data 100s of ways and look at the same number of genes that might change on a random basis. From this calculation an estimate of the false discover rate or FDR can be derived. In the program there is a sliding scale so one can adjust the level of FDR which one is willing to tolerate. The program is a plug-in and can be placed directly into the Excel program containing ones data.

Input Parameters

Imputation Engine	10-Nearest Neighbor Imputer
Data Type	Two Class, unpaired data
Number of Permutations	200
RNG Seed	1234567
(Delta, Threshold)	(0.40671, 0.00000)
(Upper Cutoff, Lower Cutoff)	(1.34842, -2.21946)

Computed Quantities

Computed Exchangeability Factor S0	0.338660432
S0 percentile	0.3
False Significant Number (Median, 90 percentile)	(13.00000, 70.00000)
False Discovery Rate (Median, 90 percentile)	(18.05556, 97.22222)

65 Positive Significant Genes

Row	Gene Name	Gene ID	Score(d)	Numerator(r)	Denominator(s+s0)
555	G19 X03672 beta-actin	Actb	3.84498887	4.34806459	1.130839907
478	F2l X81582 insulin-like growth factor binding protein-4 (IGFBP-4)	Igfbp4	3.23030885	1.71995061	0.532441538
326	D4n X90829 Lbx 1 transcription factor	Lbx1h	2.97695569	1.1667704	0.391934082
359	D7h X60831 transcription factor UBF	Tcfubf	2.64025283	1.30444979	0.494062453
76	A6m L01640 Cdk4; cyclin-dependent kinase 4	Cdk4	2.59995423	1.43483143	0.551867957
80	A7c U19596 p18ink4; cdk4 and cdk6 inhibitor	Cdkn2c	2.5118562	1.06707234	0.424814261
65	A6a Z26580 cyclin A (G2/M-specific)	Ccna2	2.44837737	1.58351175	0.646759673
340	D5n X94125 SRY-box containing gene 3 (Sox3)	Sox3	2.44569566	2.12145534	0.867424091
448	E7j J02870 lamimin receptor 1	Lamr1	2.43219537	2.2752138	0.93545684
194	C1m X76341 glutathione reductase	Gr1	2.43132634	1.2976968	0.533740282
226	C4g D83966 protein tyrosine phosphatase	Ptpn13	2.3420489	2.64805142	1.130655906
223	C4c X68193 Nm23-M2; NDP diphospha kin B; metast-reduc ; c-myc-relat TF	Nme2	2.25353525	1.9100871	0.847595838
491	F3l K01759 nerve growth factor beta (beta-NGF)	Ngfb	2.22563804	1.17049593	0.525914774
549	G5 X51703 ubiquitin	Ubb	2.14075996	2.2302612	1.041808163
263	C7e U02098 pur-alpha transcrip activa; seq.-specific ssDNA-binding protein	Pura	2.10147433	0.86533731	0.411776295
515	F5i M11686 cytoskeletal epidermal keratin (18 human)	Krt1-18	1.90351421	1.98937272	1.045105263
454	F1b L24755 bone morphogenetic protein 1	Bmp1	1.8990971	1.11664936	0.587989608
93	B1b X53584 HSP60 (heat shock 60-kDa protein1); chaperonin, GroEL homol;	Hsp60	1.88411479	0.80888404	0.429317815
452	E7n D13867 VLA-3 alpha subunit	Itga3	1.86032081	0.97189908	0.522436281
260	C7b X53068 PCNA; proliferating cell nuclear antigen; processivity factor	Pcna	1.85477232	1.075221	0.579705114
157	B6a L02526 MAPKK1; MAP kinase kinase 3 (dual specificity) (MKK1)	Map2k1	1.84985602	1.6005228	0.865214793
248	C6a X96618 PA6 stromal protein; RAG1 gene activator	Rga	1.82469202	0.73200919	0.401168622
550	G6 D78647 phospholipase A2	Ywhaz	1.779834	1.51570328	0.851598117
339	D5m U41626 split hand/foot gene	Shfdg1	1.773655	2.98080541	1.680600464
286	D1n U42554 Sim transcription factor	Sim2	1.75110414	1.67440803	0.956201283
319	D4g X59252 homeobox protein 8 (Hox-8)	Msx1	1.73886745	0.89357217	0.513881705
275	D1c L12721 adipocyte differentiation-associated protein	Dlk1	1.67776254	1.18066085	0.703711537
490	F3k M11434 nerve growth factor alpha (alpha-NGF)	Ngfa	1.66491167	1.41222894	0.848230551
420	E5g L04662 GABA-A transporter 4	Gabt4	1.6649035	1.65288382	0.992780554
459	F1g U12983 Cek 5 receptor protein tyrosine kinase ligand	Efnb1	1.63492979	1.12090601	0.685598862
298	D2l M22115 ERA-1 protein (ERA-1-993)	Hoxa1	1.61897784	2.30654696	1.424693351
210	C3d U83628 DAD-1; defender against cell death 1	Dad1	1.60525132	2.59631217	1.61738672
199	C2d D30687 GST Pi 1; glutathione S-transferase Pi 1; preadipoc GF	Gstp2	1.60183621	2.67998205	1.67306871
11	A1k U36789 p130; retinoblastoma gene product-related protein Rb2/p130	Rbl2	1.58881569	1.11660555	0.702791116
200	C2e U19463 A20 zinc finger protein; apoptosis inhibitor	Tnfaip3	1.57945479	1.00378148	0.635524038

Figure 6. Part of a SAM output table showing the d parameter which measures significance and the r parameter that is the log2 of the fold change and the s+s0 parameter that is a measure of the variance. D then represents ratio of the fold change to the variability in that measure. See reference 12 for details.

An example of the use of SAM with the 2T3 data, using two duplicate control experiments and two 30 min BMP2 treated experiments with RNA from independent cultures is shown below. After log transforming the data so that increases and decreases are of equal magnitude, SAM is evoked in the Excel file and the data run through the program. The output is an Excel workbook with the original data, a graph of expected genes versus observed genes (based on the permutation and randomization algorithm) and a table of the significant gene expression changes, in order from highest to lowest. The program also allows one to explore less significant changes with a corresponding increase in the FDR. The output of part of this table (Figure 6) is shown below, where we set the FDR at 18% (i.e. 1 in 5 gene expression changes could be false) and found 65 positive significant genes and 5 negative significant genes. The score (d) or what is referred to as the T-statistic, represents the ratio of the Numerator(s) to the Denominator(s +s0). The Numerator is a log2 value of the fold change in expression (ie for beta Actin—4.34 =log2 of 23 fold). The Denominator represents the log 2 of the variance or approximately, the log 2 of the standard deviation in the duplicate experiments.

For details on use and an extensive description of how the program works, see the SAM Users guide and technical documents that can be obtained as described in reference 12. An Excel file in the appropriate format for SAM after log2 transformed and the excel workbook with the original data and the SAM analysis, using the 2T3 model and the 30min BMP2 treatment data can be found at http://www.umkc.edu/dentistry/oralbiology/body_faculty_steve_files.html (SAM30minBMPResponsive_2T3Osteoblast.zip)

8. DLX 5 AND DLX 2 ARE EARLY BMP2 RESPONSIVE GENES IN OSTEOBLASTS

A variety of experiments indicate that Dlx5 and Dlx 2 are important homeobox transcription factors for osteoblast growth and differentiation. Overexpression of Dlx 5 in chicken osteoblasts accelerates the entire cascade of bone formation (23). Overexpression in MC3T3-E1 osteoblast can lead to increased osteocalcin expression and mineralization, and they demonstrated that Dlx5 is BMP4 inducible in cultured osteoblast and in embryo cultures (24). Knockout experiments of Dlx5 and Dlx6 genes have
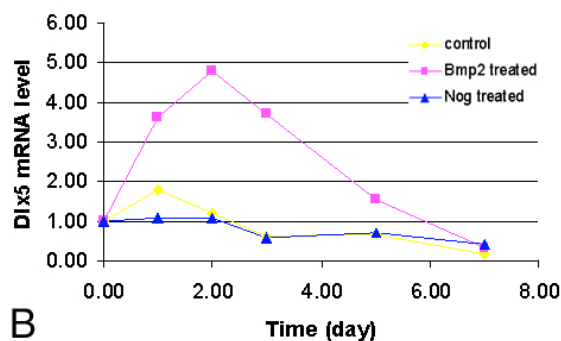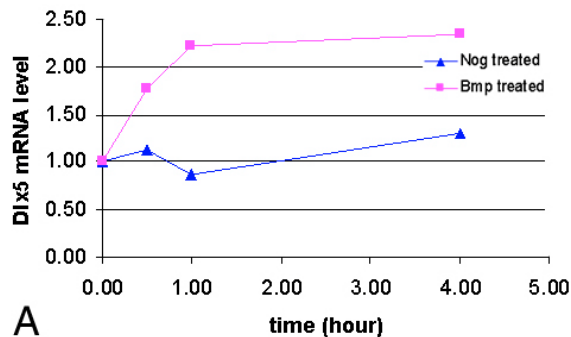
Figure 7. Expression of Dlx5 mRNA in 2T3 Osteoblasts. Dlx5 is an early BMP2 response genes, increased almost 2-fold by 30 min and remains elevated up to 5 days in BMP-2 treated cultures. The concenration of BMP-2 is 100 ng/ml.
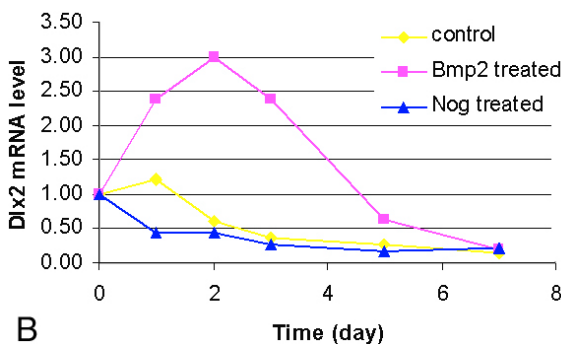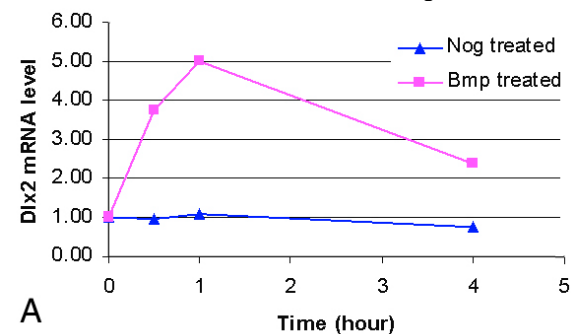

Figure 8. Expression of Dlx2 mRNA in 2T3 Osteoblasts. Dlx2 is an early BMP2 response genes, increased over 4 fold by 30 min and remains elevated up to 5 days in BMP2 treated cultures. The concentration of BMP2 is 100 ng/ml. Noggin at 0.5 μg/ml inhibits Dlx2 basal expression at Day 1 after confluency and addition of ascorbic acid at 100 μg /ml and 4 mM Beta –glycerol phosphate.

demonstrated that these genes are required for craniofacial, axial, and appendicular skeletal development (25). Mice in which just the Dlx5 is deleted have regional branchial arch defects and over 20% of the mice fail to form a calvariae (26).

We have recently demonstrated that both Dlx5 and Dlx2 are rapidly induced in 2T3 osteoblasts by 30 minutes after the addition of BMP2 to 2T3 osteoblasts. This data is shown in Figure 7 and Figure 8.

Both genes remain elevated for up to 5 days after BMP2 addition and support the hypothesis that Dlx5 and Dlx2 are important transcription factors for BMP2 induced bone formation and may function at multiple steps during both the growth and differentiation stages. Noggin treatment at 0.5 μg/ml to 2T3 cells blocks the cell growth, the multiplayer cell formation and mineral formation in this model, but does not affect the basal levels of Dlx5, as shown in Figure 7. However the basal level of Dlx2 decreases in the presence of noggin at 1 day suggesting that the Dlx 2 gene is responding to the endogenous BMP signaling that is occurring in the absence of added BMP2 (Figure 8). The effect of Noggin alone at 0.5 μg/ml is shown in Figure 9. Noggin inhibits the multilayer cell formation and cell growth that is required for later stages of differentiation and mineralization.

9. DETERMINATION OF THE EARLY DOWNSTREAM GENES IN OSTEOBLASTS THAT ARE REGULATED BY DLX FAMILY MEMBERS USING ENGRAILED-DLX5 HOMEODOMAIN REPRESSOR (ENG-DLX5HD) TOOL

To develop a deeper genomic understanding of how any transcription factor works and what genes that transcription factor contributes to their regulation, we began to build networks of genes that are directly regulated by that particular transcription factor. One can do this by selectively knocking-down or inhibiting at the genome level the activation of the specific transcription factor or class of transcription factors. In our case we would like to selectively inhibit the immediate downstream genes that are directly regulated at the transcription level by Dlx5 and Dlx2. Other approaches include chromatin immunoprecipitation techniques combined with various bioinformatics tools (27).

Recent work has shown that attaching the DNA binding domain of a given transcription factor to the strong repressor domain of the Engrailed transcription factor (Eng) creates a tool to selectively inhibit only those genes that are regulated by that given transcription factor (15,28). We have used this technique to begin to explore the downstream genes in osteoblasts that are directly regulated by Dlx5 and other members of the Dlx family. The overall strategy is shown in Figure 10A. The infection protocol is shown in Figure 10B.

We initially attempted to produce stable clones with the control, Eng, and the Eng-Dlx5 expression vectors. We produced stable clones that initially expressed Eng-
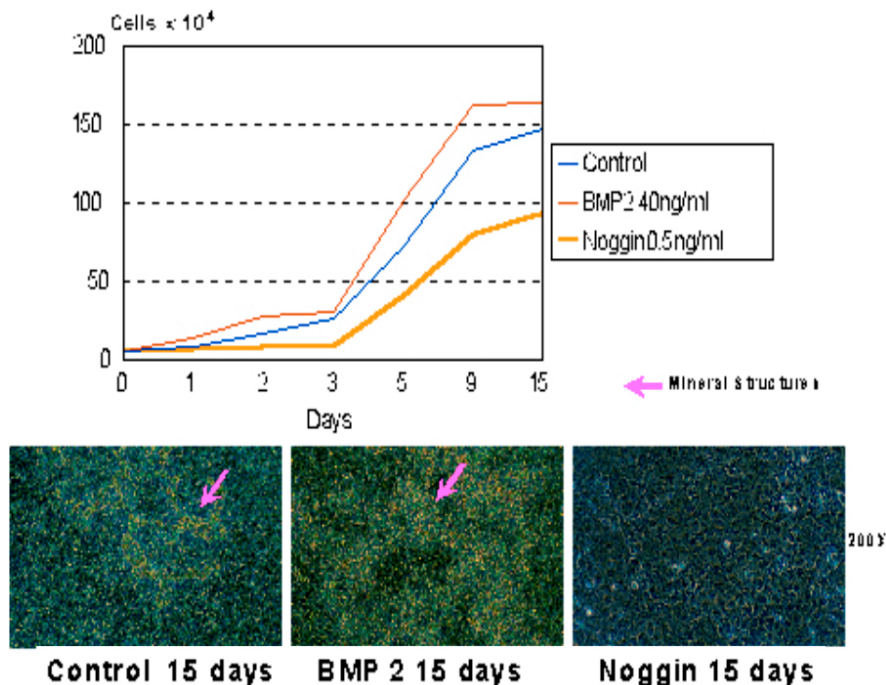
Figure 9. Noggin at 0.5 μg /ml inhibits the early multilayering and growth that is required for differentiation to a well-organized mineral matrix. BMP2 accelerates both the growth and multilayering, and futher differentiation to a mineralized matrix. This indicates that endogenous BMP signaling is most likely responsible for the slow multilayering and eventual mineralized matrix that is formed at 15-22 days in the control cultures without added BMP2.

RES GFP but the expression of Eng–Dlx5 and the GFP was lost upon expansion of the stable clones. This suggests that Eng-Dlx5 inhibits genes involved in some aspects of cell growth or survival requirements of the cells at this stage. . Recently we have used the Eng-Dlx5 retrovirus and have now gotten over 95% of the cells positive for GFP by 2-3 days after infection. Figure 11 shows an example of this analysis with the Eng-Dlx5 retrovirus and Control Retrovirus.

After the cells were confluent at 4 days after infection, all cells were still GFP positive.. BMP2 at 100 ug/ml was added to the cultures and after 1h, RNA was collected, poly A –RNA prepared, and then analyzed on denaturing agarose gels. The Northern blots were first probed with a GFP probe and the level of IRES-GFP mRNA and Eng-Dlx5HD-IRES-GFP was quantitated and normalized to the housekeeping gene, GAPDH or GAP. This data indicated that the expression levels of the control retroviral vector and the Eng-Dlx5 HD vector are similar, but the Eng-Dlx5 vector appears higher..

10. GENE EXPRESSION ARRAY ANALYSIS OF ENG/2T3 OSTEOBLASTS AND ENG-DLX5HD/2T3 OSTEOBLASTS AFTER BMP2 FOR 1HOUR

Poly A RNA was used to prepare P33 labeled cDNA probes to RNA samples. The cDNA probes were hybridized to plastic arrays that contain 5,000 well annotated genes, represented as 80mer long oligonucleotides(Clontech).   After collection of the data

and analysis using Atlas Image 2.7 beta and Microsoft Access, we derived a set of 136   genes that changed expression at least 1.6-12 fold , either up or down. We then identified the same set of genes in the array data from the Eng-Dlx5HD/2T3 cells and prepared the appropriate spread-sheets.   Using Biomine software(gnsbiotech.com) we then analyzed and present these data as outlined above. Figure 12A shows the overview of the 136 BMP2 regulated genes, from 1.6 fold to 12 fold change after 1 hr BMP2 treatment in the control cells and in the Eng-Dlx5 infected cells. The data have been normalized such that the control 1hr-Control Vector is set at 1.0 and the control 1h-EngDlx5 Vector is set at 1.0.   As indicated we can identify classes or clusters of genes that areBMP-Dlx dependent by their failure to be induced or inhibited in the presence of the Eng-Dlx5HD vector. Also there is a large class of genes that are partially or completely independent of Dlx, as this class of genes show a nice BMP2 response in the presence of the Eng-Dlx5 vector.

Figure 13 is our current model of a partial genomic regulatory network of BMP2 responsive genes that are dependent or independent of Dlx family of transcription factors.   This model is based on our first iteration of this complex process. The gene symbol is used to identify the gene and simple insertion of this symbol into NCBI GenBank can take you to the appropriate sequence file. This can then take you to a variety of Bioinformatic Web sites, such as MGI, Locus Link, AmiGo for gene ontology. The MGI data base can quickly give you a list of references and is linked to PubMe. The MGI data base can

Control

Eng-Dlx5HD

Eng

Activate

DlxAD
DlxHD

Msx

BTA

CNGTAANTG

Dlx Binding Site

DlxAD
DlxHD

Inhibit

Msx

Eng
DlxHD

BTA

CNGTAANTG

Dlx Binding Site

HD =Homeo Domain   AD=Activation Domain   BTA=Basal Transcription Apparatus
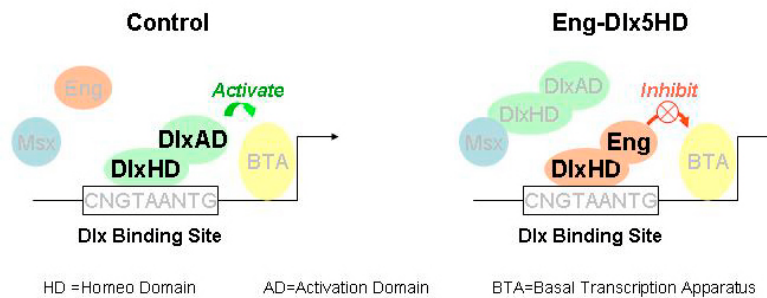
•Engrailed (Eng) is a strong repressor when attached to a DNA binding domain of a transcription factor
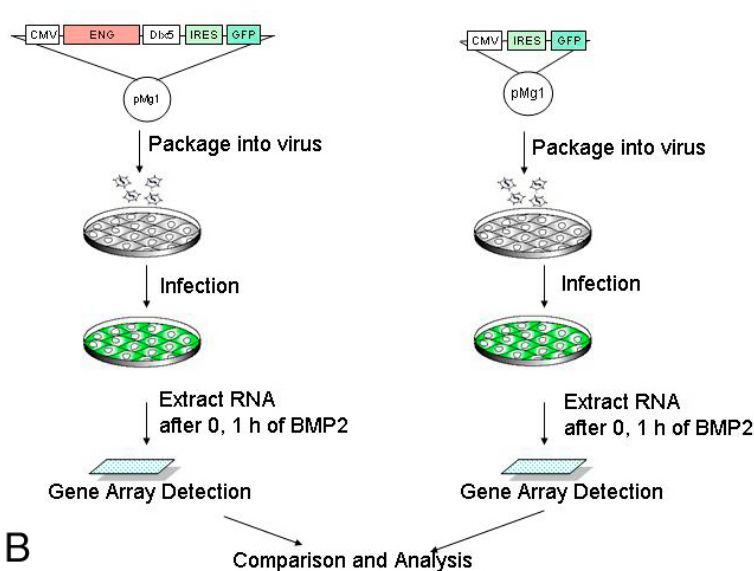
•Eng-Dlx5HD inhibits Dlx5 responsive genes

A

CMV | ENG | Dlx5 | IRES | GFP

pMg1

↓ Package into virus

↓ Infection

↓ Extract RNA after 0, 1 h of BMP2

Gene Array Detection

CMV | IRES | GFP

pMg1

↓ Package into virus

↓ Infection

↓ Extract RNA after 0, 1 h of BMP2

Gene Array Detection

B

Comparison and Analysis

Figure 10. Outline of Experiments to Selectively inhibit Dlx 5 responsive genes at the transcription level. A. Mechanism of Eng-DlxHD. The Eng-Dlx5HD will bind Dlx binding sites in chromatin and inhibit a gene that is normally Dlx regulated. B. Infection Protocol. After packaging the Empty Control vector(pMg1-IRES-GFP) and the Eng-Dlx5HD-IRES-GFP into the retrovirus and determining the appropriate titer, 2T3 cultures were infected such that by 2-3 days, more than 90% of the cells were GFP positive. After the cells reach confluency, 3-4 days, they were treated with BMP2 at 100 μg/ml for 1hr.. Poly A- RNA was collected, and analysed by Northern analysis for various expressed genes. Part of the PolyA RNA is then converted to cDNA and used to analyze a 5,000 gene microarray with well-annotated genes. Right panels: pMg1 empty vector. Left panels: Eng D1x5HD.

also take you to SwissPro and ExPasy databases for protein modeling/information.   One conlusion from these studies is however, that Dlx family of transcription factors in osteoblasts is initiall playing key roles in autoregulation of Dlx5 itself, a large number of genes involved with growth, and a large collection of genes that are in fact involved in suppressing BMP and TGFbeta function, such as Smad7, GADD45b, Noggin, Notch 1. We hypothesize that when we look at Dlx regulated genes at 3-6 days just before mineralization has been initiated, we will find a whole new set of genes regulated by this set of transcription factors.

Further analysis of the BMP-Dlx gene network will now involve identifying the appropriate cis-regulatory modules (CRM) in the Dlx dependent genes using a bioinformatics approach coupled to testing the candidate CRM both *in vitro* using cell models and, in some cases testing the model *in vivo* in transgenic animal models.   We are now developing ways in which we can turn on the Eng-Dlx5 HD in a more controlled fashion using inducible gene switch systems and using stable 2T3 clones containing continually expressing Dlx5 small interfering RNAs Finally we would like to show an example of how we can use bioinformatics
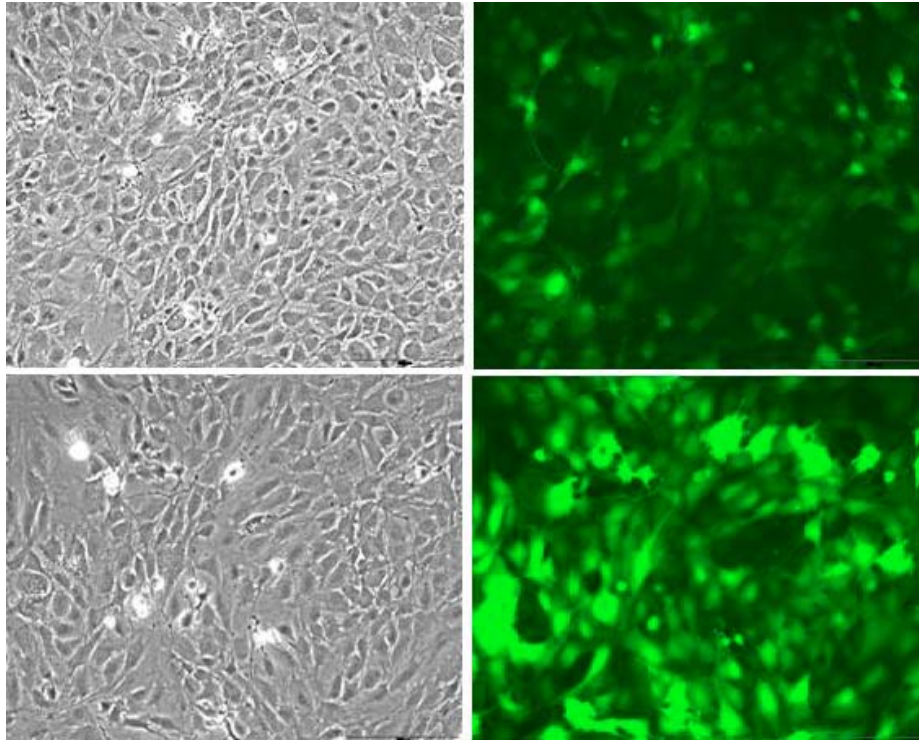
Figure 11. Retroviral Infection of 2T3 cells with   CMV-IRES–GFP or CMV- ENG-DLX5HD-IRES GFP. Photographs were taken 3 day after infection. . By determinining the number of cells and the number of GFP positive cells we can determine the infection efficiency. We calculate that >90%% of the cells are GFP positive.

approaches to greatly narrow down the cis-regulatory modules that are directly regulated by Dlx family members.

11. ANALYSIS OF HUMAN/MOUSE GENE HOMOLOGIES IN CONSERVED NON-CODING REGIONS OF GENES IMPORTANT FOR OSTEOBLAST BIOLOGY

We can now search computationally for clusters of Dlx binding sites in a gene, in which we suspect, are regulated directly by Dlx family members. We have used the rVISTA program (http://pga.lbl.gov/rvista.html). This program has over 360 matrices of transcription factor binding sites.   The new program allows user specific binding sites and an update in the future will allow one to input a give position weighted matrix table (PWM). For our analysis we used the Msx binding matrix (PWM). As best we know this is a good surrogate for Dlx binding sites, since Msx and Dlx family members bind the same DNA sequence in *in vitro* selection assays.   The sequence we used for a Dlx binding site is CNGTAANTG. There is the probability that this sequence will occur by chance 1 in 16,000 bp. Another criteria that rVISTA allows the user to set is the % matrix identity for the core binding sequence. In general we use from 80 to 90%, depending on the gene. At 85-90%, we accept a potential interesting region, if there are 2 or more clusters of the Dlx binding site within a 500bp region or window. Clustering of several weak or less tight binding sites of a given transcription factor is a

common strategy for transcriptional regulation in higher organism (2,14). rVISTA also has criteria we set such that if a Dlx binding sequence is considered conserved, then 20 bp flanking the core binding site must also 80% identity.

We have begun the rVISTA analysis (human –mouse at this stage) of a set of 40 early BMP2 response genes that were identified earlier and have confirmed BMP2 induced expression in many of this set of genes by Northern analysis. The idea is to begin to catalogue the position and extent of clustering of Dlx/Msx binding sites in these early BMP2 responsive genes.(see Figure 2B,C, and D). As we identify other Dlx regulated genes, we can apply this rVISTA approach or Family Relations approach (15) to identify the regions likely involved in   direct transcriptional activity of Dlx regulated genes. What we were surprised to find in our initial analysis of these 40 genes was the vast distances between CNS and clusters of regulatory sites as the clusters spread out over the gene of interest. For higher eukaryotic gene regulation studies, a map of candidate regulatory regions will be necessary to guide our experiments in the lab and in animal models. On average, 10-20kb of 5'-flanking region, the intron regions, and 10-20 kb of the 3'-flanking region should be explored when analyzing large numbers of genes.   Many genes will have control regions even at greater distances from the transcription unit that are hard to predict at the present time. However, programs such as rVISTA and Family Relations have a chance at identifying these distant
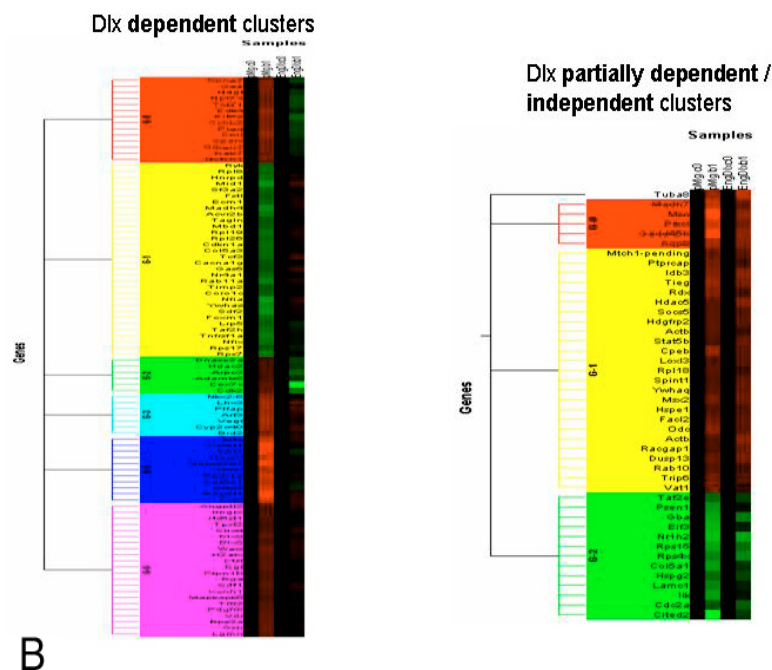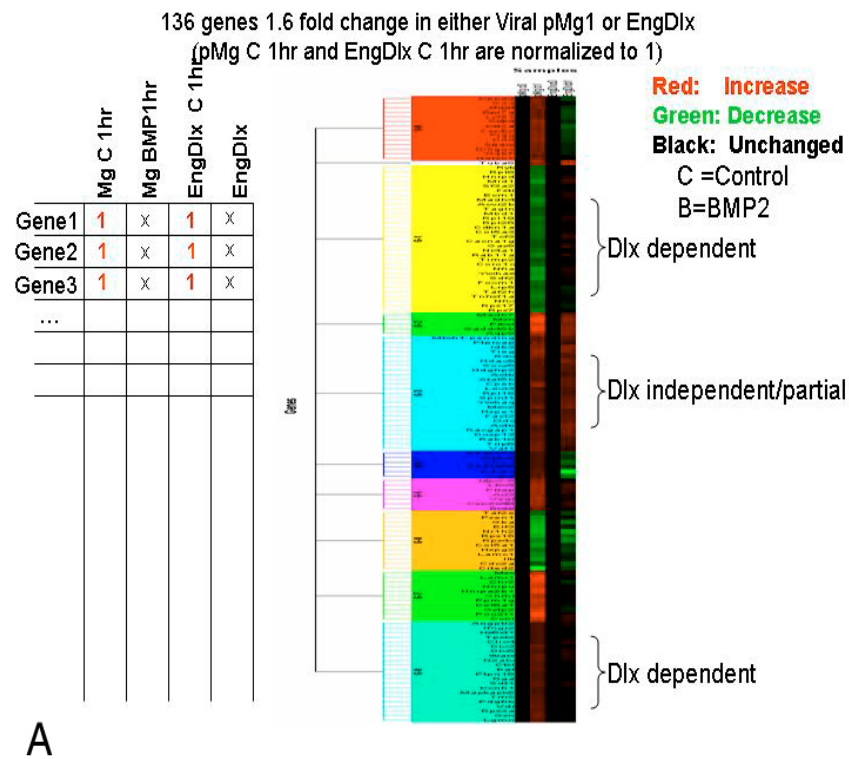
Figure 12. Percolation Clustering and Kmedian clustering of 136 genes whose expression change at least 1.6 fold by BMP2 treatment after 1hr. . A. Overview of 10 clusters of BMP2 regulated genes. Both the control at 1hr for the empty vector (Mg) and the EngDlx were set at 1.0 for comparison. It is now easy to visualize candidate genes that are Dlx dependent or independent, as indicated. B. Blow of of Clusters of Dlx dependent genes and Dlx independent or partially dependent. From the gene symbol on the cluster, access to GenBenk and other databases is fairly straight forward as outlined.
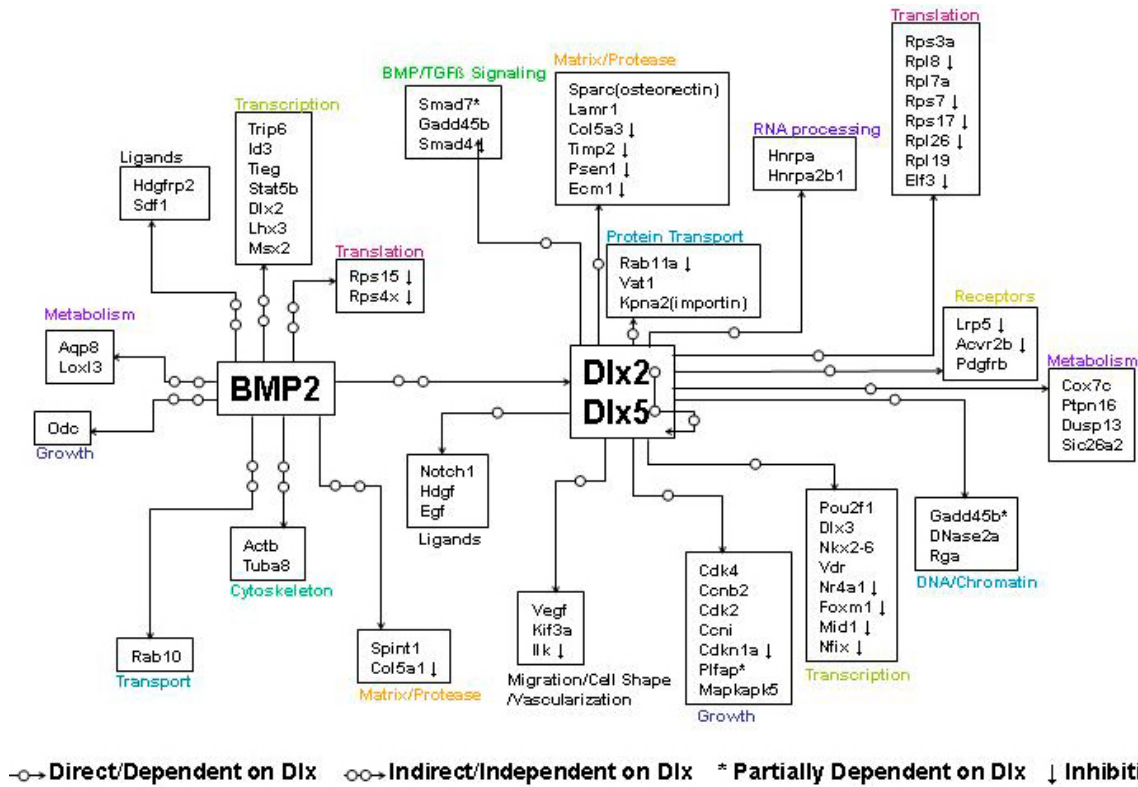
Figure 13. Overall view of part of the set of genes that are dependent or independent of Dlx family of transcription factors. Clearly Dlx is directly involved in the initial growth program induced by BMPs in osteoblast.
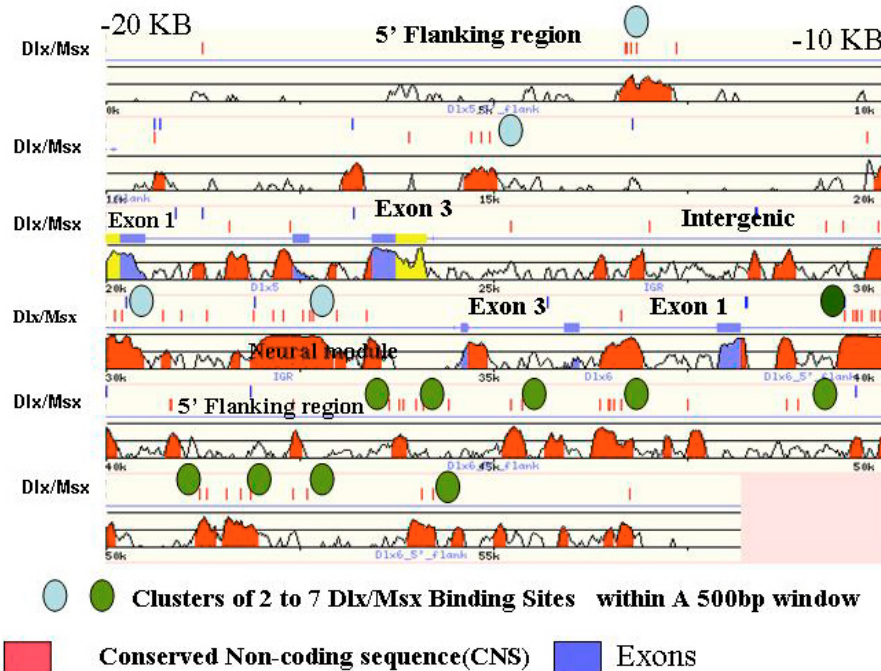
Figure 14. The Dlx5 and Dlx6 gene locus. rVISTA plot of conserved DNA regions in the 20kb of 5'flanking region, the transcription units, intergenic region (IGR), and about 20kb of 5'flanking region of the Dlx6 gene. We set the parameters to look for Dlx/Msx type binding sites (CNGTAANTG) with 85% matrix identity with the standard hula-hoop step that requires 20 bp flanking the core-binding site to have at least 80% homology. We consider that clusters of 2 or more sites within a 500 bp window to be highly significant and worthy of further investigation.

conserved sites with clusters of particular transcription factors as well as clusters of combinations of classes of transcription factors.

Figure 14 presents the rVISTA plot of CNS in the Dlx5 and Dlx6 gene locus. As you may recall, both genes are within a 60 kb region and are coordinately regulated in neural and branchial arch tissue during development. The red peaks in the rVISTA plot represent the areas in the mouse and human genes that share over 75 % identity. The red tick marks are clusters of Dlx/Msx candidate binding sites within islands of CNS. There are 4 main Dlx/Msx clusters with 2 or more sites around the Dlx5 gene and 10 clusters in the intergenic region between Dlx5 and Dlx6 and the flanking region of Dlx6. We hypothesize that one of the 4 clusters around the Dlx5 gene controls BMP2 responsiveness through Dlx5 action.

In summary, we have begun to collect a compendium of gene expression profiles of osteoblasts during their growth and differentiation in a simple clonal cell model. The focus has been on the earliest response genes to BMP2 or BMP activity. Both the Dlx 2 and Dlx 5 genes were coordinately induced by BMP2 as early as 30 minutes. Using a sequence specific repressor approach we are beginning to define the downstream genes that are directly and indirectly regulated by important transcription factors for osteoblasts using a microarray approaches.

12. ACKNOWLEDGEMENT

Partially presented at the 2002 International Osteoporosis Conference, Shanghai, China, October 22, 2002.

14. REFERENCES

1. Alizadeh AA, MB Eisen, RE, Davis Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J, Jr, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Levy R, Wilson W, Grever MR, Byrd JC, Botstein D, Brown PO, Staudt LM. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 403:503-511, (2000)

2. Arnosti DN, S Barolo, M Levine, and S Small. The eve stripe 2 enhancer employs multiple modes of transcriptional synergy. *Development* 122: 205-214, (1996)

3. Basset DE, Jr, Eisen MB, Boguski MS. Gene expression informatics - it's all in your mine. *Nature Genetics* Suppl 21:51-55, (1999)

4. Brown PO, Botstein D. Exploring the new world of the genome with DNA microarrays. *Nature Genetics* Suppl 21:33-37, (1999)

5. Golub, TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES. Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring. *Science* 286:531-537, (1999)

6. Hughes, T. Jones, A. Stoughton, R. Bennett, H. Yudong, D. Meyer, M. *et al.* Functional Discovery via a Compendium of Expression Profiles. *Cell* 102: 109-126, (2000)

7. Ly, D. Lockhart, D. Lerner, R. Schultz, P. Mitotic Misregulation and Human Aging. *Science* 287,2486-2492, (2000)

8. Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM. Systematic determination of genetic network architecture. Nature Genetics 22,281-285 (1999)

9. Ihmels I, Frielander G, Bergmann S, Sarig O, Ziv Y, and Karkai N. Revealing modular organization in the yeast transcriptional network. *Nature Genetics* 31:370-377, (2002)

10. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci* USA 95:14863-14868, (1998)

11. Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E, Lander ES, Golub TR. Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation. *Proc Natl Acad Sci USA* 96,2907-2912 (1999)

12. Tuscher, Tibshirani and Chui. Significance of analysis of microarrays applied to the ionizing radiation response. Proc. Natl. Acad Sci USA 98,5116-5121(2000)

13. Loots GG, Ovcharenko I, Pachter , Dubchak I, and Rubin EM. rVISTA for comparative sequence-based discovery of functional transcription factor binding sites. *Genome Research* 12,832-839, (2002)

14. Berman BP, Nibu Y, Pfeiffer Bd, Tomancak P, Celniker SE, Levine M, Rubin GM, and Eisen MB. Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the Drosophila genome. *Proc. Natl. Acad. Sci USA* 99:757-762.

15. Davidson EH, *et al. Science* 295: 1669-1678,(2002)

16. Ghosh-Choudhury, N, Windle J,Koop B, Harris M, Guerrero D, Wozney J, Mundy G, Harris S. Immortialized murine osteoblasts derived from BMP2-T-antigen expressing transgenic mice. *Endocrinology*, 137:331-339, (1996)

17. Chen , D, Ji X, M, Harris, M, J, Feng J, G, Karsenty G, A, Celeste A, V, Rosen V, G, Mundy G, Harris , S. Differential Roles for Bone Morphogenetic Protein (BMP) Receptor Type IB and IA in Differentiation and Specification of Mesenchymal Precursor Cells to osteoblast and Adipocyte Lineages. *J. of Cell Biology*, 142: 295-306, (1998)

18. Harris, S.E. and Harris, M.A. Gene expression profiling in osteoblast biology:bioinformatics tools. *Molecular Biology Reports* 28: 139-156.

19. Harris, S.E., Feng, J.Q., Harris, M.A., Ghosh-Choudhury, N., Dallas, M.R., Wozney, J., Mundy, G.R. Recombinant Bone Morphogenetic Protein 2 Accelerates Bone Cell Differentiation and Stimulates BMP-2 MRNA Expression and BMP-2 Promoter Activity in Primary Fetal Rat Calvarial Osteoblast Cultures. *Molecular and Cellular Differentiation*. 3(2) 137-155, (1995)

20. Ghosh-Choudhury, N, Ghosh-Choudhury, G, Woodruff, K, Bsoul, S, Nishimura, R, Oi, W, Celeste, A, Abboud, S. Identification of a Smad Binding Element (SBE) in BMP-2 Responsive CSF-1Gene Promotre: A Mechanism of BMP-2 mediated Transregulation of Osteoclastogenesis During Bone Remodeling. *22nd Meeting of the American Society for Bone and Mineral Research, Toronto, Canada September 22-26, 2000.* Vol. 15, Supplemental 1, p 5203. (2000)

21. Krumlauf, R. Hox Genes in Vertebrate Development. Cell 78:191-201, (1994)

22. Stein, S, Fritsch, R, Lemaire, L, Kessel, M. Checklist: Vertebrate homeobox genes. *Mechanism of Development* 55, 91-108, (1996)

23. Tadic T, Dodig M, Erceg I, Marijanovic I, Mina M, Kalajzic Z, Velonis D, Kronenberg MS, Kosher RA, Ferrari D, and Lichtler AC. Overexpression of Dlx5 in Chick Calvarial Cells Accelerates Osteoblastic Differentiation. *J. Bone and Mineral Res*. 17,1008-1014(2002)

24. Miyanma K, Yamado G, Yamamoto TS, Takagi C, Miyado, K, Sakai M, Ueno N, and Shibuya H. A BMP-Inducible Gene, Dlx5, Regulates Osteoblast Differentiation and Mesoderm Induction. *Developmental Biology* 208, 123-133, (1999)

25. Robledo RF, Rajan L, Li X, and Lufkin T. The Dlx 5 and Dlx6 homeobox genes are essential for craniofacial, axial, and appendicular skeletal development. *Genes and Dev*. 16,1089-1101 (2002)

26. Depew MJ, Liu JK, Long JE, Presley JE, Meneses JJ, Pedersen RA, and Rubenstein JL. Dlx 5 regulates regional development of the branchial arches and sensory capsules. *Development*, 126 : 3831-3846, (1999)

27. Lui XS, Brutlag DL, and Liu JS. An algorithm for finding protein-DNA binding sites with applications to chromati-immunoprecipation microarray experiments. *Nature Biotechnology* 20, 835-839.

28. Kessler DS. Siamois is required for formation of Spemann's organizer. *Proc. Natl. Acad. Sci USA* 94:13017-13022, (1997)

29. Lockhart, D. Winzeler, E. Genomics, gene expression and DNA arrays. *Nature* 405:827-836, (2000)