ANALYSIS OF REGULATED GENES OF BMP4 SIGNALING IN BREAST CANCER CELL LINES

Master's Thesis

Shanmugapriya Kalaichelvan

Institute of Biomedical Technology

University of Tampere

Finland

June 2013

DEDICATION

This work is dedicated to my dear brother Niranjanan Kalaichelvan and to my beloved husband Santhosh Kumar Chandrasekaran.

ACKNOWLEDGEMENT

It is my privilege to express my gratitude and respect to all those who have guided and inspired me during the course of this thesis work. With gratitude, respect and pride I take great pleasure in expressing my sincere thanks to Acting Professor Csaba Ortutay, PhD and Professor Matti Nykter for giving me the opportunity to carry out my project work and had been the constant resource for me throughout the project. I also thank Professor Anne Kallioniemi, MD, PhD for providing the data sets for this work.

On a personal note, I would like to thank my father Kalaichelvan and my mother Abiramasundari who encouraged me to pursue a Master's degree and who had also supported me financially. I also thank my in-laws Chandrasekaran and Mageswari for all their prayers and blessings. I extend my thanks to all my friends who had helped me in every possible way and making me to achieve this.

Finally, I thank Sri Shridi Sai and Satya Sai Baba for giving me strength and courage to complete this thesis.

MASTER'S THESIS:

Place: Institute of Biomedical Technology

School of Medicine

Department of Bioinformatics

University of Tampere

Author: Shanmugapriya Kalaichelvan

Title: Analysis of regulated genes of BMP4 signaling in breast cancer cell lines

Pages: 77

Supervisor: Acting Professor Csaba Ortutay, PhD

Reviewers: Acting Professor Csaba Ortutay, Professor Matti Nykter

Time: June 2013

ABSTRACT

Background and aim: The role of bone morphogenetic proteins (BMP) 4 and 7 in primary breast cancer is peculiar as they are known to induce cell proliferation in some tumor cells and also to reduce tumor growth in some cells. They are studied for their signaling on gene transcription. Characterizing the transcriptional response of primary breast cancer cells to BMP4 and BMP7 would serve better in understanding the role of these proteins in cancer biology. This work is aimed at finding out the genes that respond to these ligands. The data taken for analysis is an Agilent RNA microarray data. The data is produced from RNA of seven breast cancer cell lines that were treated with BMP4 and BMP7 ligands. Each cell line had shown different expression to these ligands and hence finding the genes that are most significant in their expression when treated with these ligands and also exploring the Gene Ontology, Transcription Factor and KEGG pathways of these genes would provide more information. These studies were performed also for the set of genes obtained from hierarchical clustering known as group C genes.

Methods: The Agilent microarray data was analyzed using Bioconductor's limma package in R programming environment. The cell lines that were treated with both BMP4 and BMP7 and that were treated only with BMP4 were analyzed separately. The significant genes were found out based on P-value and heat maps were generated using the log FC (Fold Change) values revealing their expression upon BMP4, BMP7 treatments. Furthermore, using WebGestalt the Gene Ontology, Transcription Factor and KEGG pathways were also analyzed. The group C genes were also subjected to these analyses using WebGestalt.

Results: In BMP4 and 7 treated cell lines, there were ninety three genes showing varying expression and for BMP4 treatment there were eighty one genes that showed expression. The genes showed over expression or up regulation and under expression or down regulation in each cell line indicating the influence of BMP4 and BMP7 signaling. The KEGG pathways like cancer and melanoma indicated the importance of genes and signaling of ligand molecules in cancer.

Conclusion: The aim of the thesis work was to identify the set of genes regulated by BMP4 and BMP7 signaling and it was obtained. It can be said that the BMP4 and BMP7 had a strong effect on the genes by varying their expression from each cell line to the other. The work had established basic information on signaling effects of BMP4, 7. It had also found out the biological and molecular functions, transcription factors and KEGG pathways involving these genes.

CONTENTS

ABBREVIATIONS

AnnotationDbi	Annotation Database Interface
BMP	Bone Morphogenetic Protein
BMP4	Bone Morphogenetic Protein 4
BMP7	Bone Morphogenetic Protein 7
BMPR	Bone Morphogenetic Protein Receptor
BRCA1	BReast CAncer gene one
BRCA2	BReast CAncer gene two
EST	Expressed Sequence Tag
ER	Estrogen Receptor
DEG	Differentially Expressed Genes
DEP	Differentially Expressed Probes
DNA	Deoxyribo Nucleic Acid
DCIS	Ductal Carcinoma In Situ
FISH	Fluorescence In Situ Hybridization
GEO	Gene Expression Omnibus
GEP	Gene Expression Profiling
GF	Growth Factor
GH	Growth Hormone
GO	Gene Ontology
h	hour
IGF	Insulin-like Growth Factor
KEGG	Kyoto Encyclopedia of Genes and Genomes

LCIS	Lobular Carcinoma In Situ
Limma	Linear Models for Microarray Data
MAPK	Mitogen Activated Protein Kinases
MPSS	Massively Parallel Signature Sequencing
mRNA	Messenger RNA
RNA	Ribo Nucleic Acid
SMAD	Homologue of Drosophila Mothers Against Decapentaplegic
SNP	Single Nucleotide Polymorphism
TGF - β	Transforming Growth Factor - β
WebGestalt	WEB-based GEne SeT AnaLysis Toolkit

1. INTRODUCTION

Bone morphogenetic proteins (BMP) are a family of ligands that belong to the transforming growth factor β (TGFβ) superfamily. BMPs have the ability to induce endochondral bone formation and regulation of transcription of target genes. BMPs interact with specific receptors on the cell surface, referred to as Bone morphogenetic protein receptors (BMPRs). Phosphorylation of downstream targets is mediated by signal transduction through BMPs. There are many BMPs like BMP 1, 2, 3, 4, 5, 6, 7, 8a, 8b, 10 and 15 with unique functions. In humans there are 21 members of BMP families. These regulate the transcription of target genes by signaling through type I and type II transmembrane serine-threonine receptors. BMP4 is essential for muscle development, bone mineralization and ureteric bud development. BMP7 or osteogenic protein-1 plays a vital role in transformation of mesenchymal cells in to bone and cartilage (Table 1.1).

The BMP signaling pathways regulates gene transcription. These pathways are initiated by the formation of heterotetramer. The BMP dimer binds to its type II receptor that recruits type I receptors resulting in the formation of heterotetramer with two receptors of each type. The type I receptor is phosphorylated by type II receptor. The SMAD cascade is one among the two pathways and the other pathway MAPK involves two mitogen activated protein kinase cascades.

TGF-β family receptors use the SMAD signaling pathway to transduce signals. The type I receptor is phosphorylated by type II receptor. Phosphorylation of R-SMAD 1, 5 and 8 takes place and the R-SMAD complex moves to nucleus. The downstream effect of R-SMADs is prevented by dorsomorphin.

BMP signaling is involved in tumor suppression, bone homeostasis, angiogenesis and metastasis. Among all the BMPs the BMP4 and BMP7 are known for their aberrant expression in primary breast cancer and bone metastases. The BMP4 and BMP7 treated cell lines reveal much information like cell proliferation and differentiation. Dealing with expression data determines differentially expressed genes and the transcriptional responses to the BMP4 and BMP7 signaling. BMP4 and BMP7 have been studied extensively in cancer biology to understand their role in metastasis and especially in breast cancer their role receives much attention from researchers.

Table 1.1 Information on BMP4 and BMP7

Protein	Name	Gene Name	Chromosome and location	Protein family	Functions
BMP4	Bone morphogenetic protein 4	BMP4	14q22-q23	TGF-beta family	Bone, cartilage development
BMP7	Bone morphogenetic protein 7	BMP7	20q13	TGF-beta family	Bone homeostasis

BMP4 has been identified as breast cancer metastasis suppressor gene. The 4T1 preclinical mammary mouse models had shown differences in the levels of BMP4 and coupled with highly metastatic or non-metastatic cells. Further studies indicated that the highly metastatic T1.2 mammary tumor cells express lower levels of BMP4 whereas the weakly or non-metastatic cells had higher levels of BMP4. Prevention of secondary tumor formation occurs as the 4T1.2 tumor cells become more susceptible to anoikis due the presence of BMP4 and prolongs the life of 4T1.2 tumor bearing mice. It is found that BMP4 act on both tumor cells and stromal components and suppresses metastasis. Further investigation was carried out on transcriptional alterations triggered by BMP4 using microarray gene expression profiling. From 4T1.2, primary tumor cells were isolated and subjected to gene expression profiling. Ontology analysis of differentially expressed genes (DEG) revealed many pathways and target factors that affect breast cancer metastasis (Cao, 2011).

The effect on global gene transcription in BMP4 and BMP7 treated breast cancer cell lines showed different expressions. As a response to these ligands the cellular functions, regulation of gene expression and signal transduction showed notable changes in the metabolism and cell proliferation. A set of genes expressed common molecular responses to BMP4 and BMP7 and were known as synexpression group of genes. This group of genes was obtained after several types of filtration (Rodriguez-Martinez *et al.*, 2011).

Microarrays help researchers to study thousands of gene expression simultaneously. The role of microarray in breast cancer studies in inevitable. Both DNA and RNA microarrays play a major role in gene expression studies. There are about 1.5 million

cases of breast cancer worldwide according to Breast Cancer Statistics. The occurrence, proliferation and cure can only be understood when large number of genes is studied and such a massive exploration can be easily handled with microarray technology. Gene expression profiling is a method of measuring the expression of thousands of genes at once. The gene expression values are later analyzed using one of the software called Bioconductor. This software can be used for pre-processing, quality assessment, differential expression, clustering, classification, gene set enrichment analysis and genetical genomics.

The aim of this thesis work was exploring the set of significant genes that respond to BMP4 and BMP7 by using the Bioconductor package and subject those genes to Gene Ontology, Transcription Factor and KEGG pathway analyses. This work was based on data provided by Rodriguez-Martinez from Laboratory of Cancer Genetics, Institute of Biomedical Technology, Finland. Finding out the significant genes would help in understanding their interactions with ligand molecules and also it helps in knowing the metabolic processes. The synexpression group of genes was also subjected to Gene Ontology, Transcription Factor and KEGG pathway analyses in order to find out the functions of genes and their pathways. The work has also been done with an approach that it opens an arena for investigating the genes that are expressed in primary breast cancer samples with that of identified set of genes from the breast cancer cell lines and there by drawing a comparison between *in vivo* and *in vitro* behavior of these genes.

2. REVIEW OF LITERATURE

2.1 Breast cancer – An overview

2.1.1 Breast cancer and its types

Breast cancer is a form of cancer which is caused in the tissues of breast that are the inner linings of the ducts and lobules. Cancer may occur from ducts namely ductal carcinoma or from lobule namely lobular carcinoma. Breast cancer is one of the major cause of death in females and also one of the most common invasive cancers (Alarmo *et al.,* 2010, Dumitrescu *et al.,* 2005, Mundy *et al.,* 1997, Naber *et al.,* 2012). There are about 1.15 million cases of breast cancer diagnosed all over the world every year. Thirty percent of all cancer in women occurs in breast (Park *et al.,* 2009). Abnormalities at gene level are the major cause of breast cancer. BRCA1, BRCA2, p53, PTEN, ATM, NBS1, LKB1, AR, ATM, BARD1, BRIP1, CHEK2, DIRAS3, ERBB2, NBN, PALB2, RAD50 and RAD51 when variation happen in these high penetration genes, breast cancer occurs (Table 2.1.3). Inheritance contributes to 5-10% of cancer cases (Rajnish *et al.,* 2012).

The forms of breast cancer are invasive and non-invasive. Invasive form spread from milk duct or lobules to other tissues in breast. Non-invasive does not spread to other tissues and it is also called as *"in situ"*. Ductal carcinoma *in situ* (DCIS) or intraductal carcinoma and lobular carcinoma *in situ* (LCIS) are two main types of cancer (Sotiriou *et al.,* 2003). Among estrogen receptor (ER) positive and estrogen receptor negative breast cancers, the estrogen receptor negative has the highest risk. ER negative breast cancer cells can be further classified into Erbb2/Her2/Neu positive, basal cell like and normal breast-like subtypes, with the basal cell like breast cancer having the worst prognosis and the normal breast-like the best prognosis (Otsuka *et al.,* 2009). ER negative breast cancer can be further classified into luminal 22 subtype A, B and C, with luminal C having the worst and luminal A having the best prognosis. Furthermore, the clinical outcome can be predicted based on a 70 gene signatures in the primary tumor. This led to the hypothesis that metastatic traits are already acquired during early tumor genesis (van't Veer *et al.,* 2002, Sorlie *et al.,* 2001, Weigelt *et al.,* 2005, Perou *et al.,* 2000, Honrado *et al.,* 2006).

2.1.2 Epidemiology of breast cancer

National Cancer Institute has estimated that 232,340 women will be diagnosed with and 39,620 women will die of breast cancer in 2013 in United States (Howlader *et al.,* 2013, Desai *et al.,* 2002). Breast cancer contributes to 16% of all female cancer, 22.9% of invasive cancer and 5% of breast cancer in women under age of 40 years, stated the World Health Organization in the year 2009. American Cancer Society provided the percentage of survival after diagnosis of breast cancer as 89% after five years, 82% after ten years of diagnosis and only 77% after fifteen years of diagnosis. The incidence of breast cancer is high in well developed countries and comparatively less in developing or under developed countries. United States has the highest incidence rates of 128.9 per 100,000 women. All these statistics indicate the growing risk of cancer and increasing rates of mortality. This is the very main reason for extensive research which is carried out in breast cancer in order to find the potential target.

Table 2.1.3 Genes associated with breast cancer (Source: OMIM)

Gene	**Gene/Locus MIM number**	**Location**	**Phenotype**
RAD54L	603615	1p34.1	Breast cancer, invasive ductal
CASP8	601763	2q33.1	Breast cancer, protection against
BARD1	601593	2q35	Breast cancer, susceptibility
PIK3CA	171834	3q26.32	Breast cancer, somatic
HMMR	600936	5q34	Breast cancer, susceptibility
NQO2	160998	6p25.2	Breast cancer susceptibility
RB1CC1	606837	8q11.23	Breast cancer, somatic
SLC22A1L	602631	11p15.4	Breast cancer, somatic
TSG101	601387	11p15.1	Breast cancer, somatic
ATM	607585	11q22.3	Breast cancer, susceptibility
KRAS	190070	12p12.1	Breast cancer, somatic
BRCA2	600185	13q13.1	Breast cancer, male, susceptibility
XRCC3	600675	14q32.33	Breast cancer, susceptibility
AKT1	164730	14q32.33	Breast cancer, somatic
RAD51A	179617	15q15.1	Breast cancer, susceptibility
PALB2	610355	16p12.2	Breast cancer, susceptibility

Gene	Gene/Locus MIM number	Location	Phenotype
CDH1	192090	16q22.1	Breast cancer, lobular
PHB	176705	17q21.33	Breast cancer, susceptibility
BRIP1	605882	17q23.2	Breast cancer, early-onset
PPM1D	605100	17q23.2	Breast cancer
CHEK2	604373	22q12.1	Breast cancer, susceptibility

2.1.3 Genes associated with breast cancer

The table (2.1.3) shows the different genes and their respective role in breast cancer. Studies indicate that the breast cancer gene one and breast cancer gene two (BRCA1&2) are associated with inherited cancer in most cases. The role of the BRCA genes is to repair cell damage and keep normal growing breast cells. But when these genes contain abnormalities or mutations that are passed from generation to generation, the genes don't function normally and breast cancer risk increases. 10% of all breast cancers or 1 out of every 10 cases has abnormal BRCA1 and BRCA2 genes. The abnormalities of BRCAs are not only the causative of breast cancer but there is single nucleotide polymorphism (SNP) that is also found to be linked to the cause of breast cancer. SNPs are mutations in small regions of chromosomes. It is also seen that women who are detected with this type of cancer have a family history of breast cancer or ovarian cancer or any other cancer there by showing the inheritance of abnormal BRCA genes (Bishop 1999, Ford *et al.,* 1998, Risch *et al.,* 2001).

Other than BCRA, there are several genes associated to breast cancer like ATM that helps to repair damaged DNA. Inheriting two abnormal copies of ATM gene causes the disease ataxia-telangiectasia, a rare disease that affects brain development. Inheriting one abnormal ATM gene has been linked to an increased rate of breast cancer in some families because the abnormal gene stops the cells from repairing damaged DNA. This gene and genes other than BRCA has less risk of causing cancer. p53 or TP53 gene provides instructions to the body for making a protein that stops tumor growth (Table 2.1.3). Inheriting an abnormal p53 gene causes Li-Fraumeni syndrome, a disorder that causes people to develop soft tissue cancers at a young age. People with this rare syndrome have a higher-than-average-risk of breast cancer and several other cancers

including leukemia, brain tumor and sarcomas (Annegien *et al.,* 2000, Georgia *et al.,* 2001).

The CHEK2 gene also provides instructions for making a protein that stops tumor growth. Li-Fraumeni syndrome can also be caused by an inherited abnormal CHEK2 gene. Even when an abnormal CHEK2 gene doesn't cause Li-Fraumeni syndrome, it can double breast cancer risk. The PTEN gene helps to regulate cell growth. An abnormal PTEN gene causes Cowden syndrome, a rare disorder in which people have a higher risk of both benign (not cancer) and cancerous breast tumor, as well as growths in the digestive tract, thyroid, uterus and ovaries. The CDH1 gene makes a protein that helps cells to bind together to form tissue. An abnormal CDH1 gene causes a rare type of stomach cancer at an early age. Women with an abnormal CDH1 gene also have an increased risk of invasive lobular breast cancer. Usually ATM, BRCA1, BRCA2, TP53 are the genes that are responsible for DNA damage recognition and repair pathways. Mutations in these genes increase the breast cancer rate from moderate to high. Another gene CHEK2 has also been reported which belongs to the same pathway causes cancer when deletion occurs in this gene but all these mutations do not contribute to a high incidence of cancer and only up to 2 – 5% which requires further studies to find out the real factors that cause mutation in these genes and there by leading to breast cancer (Caroline *et al.,* 2007).

Human Epidermal Growth Factor Receptor 2 (HER2) is another widely studied gene in breast cancer belonging to epidermal growth factor receptor family. Though, it does not belong to the transforming growth factor family it is studied for its essential part in breast cancer especially in pathogenesis and progression stages. This gene is also known as erbB2 or neu gene. It is found to be over expressed in about 30% breast cancers. The over expression or amplified presence of this gene in patients results in lower survival rates and make them more susceptible to other diseases (Tan *et al.,* 2007).

2.2 BMP and SMAD pathways

Bone morphogenetic proteins are a family of ligands that belong to the transforming growth factor β (TGFβ) superfamily. BMPs have the ability to induce endochondral bone formation and regulation of transcription of target genes. BMPs interact with specific receptors on the cell surface, referred to as bone morphogenetic protein receptors (BMPRs). Phosphorylation of downstream targets is mediated by signal transduction

through BMPs. There are several BMPs like BMP 1, 2, 3, 4, 5, 6, 7, 8a, 8b, 10 and 15 with unique functions (Table 2.2). In humans there are 21 members of BMP families that are involved in bone formation and developmental processes. These regulate the transcription of target genes by signaling through type I and type II transmembrane serine-threonine receptors. BMPs are extracellular signaling molecules that are able to regulate various cellular functions, proliferation, differentiation, apoptosis and migration. BMPs have been studied in several cancers and aberrant expression patterns of BMPs have been reported (Alarmo *et al.*, 2010, Herpin *et al.*, 2007).

Table 2.2 Types of BMPs, their functions and Gene locus:

(Source: http://en.wikipedia.org/wiki/Bone_morphogenetic_protein)

BMP	**Functions**	**Gene Locus**
BMP1	BMP1 does not belong to the TGF-β family of proteins. It is a metalloprotease that acts on procollagen I, II and III. It is involved in cartilage development	Chromosome: 8; Location: 8p21
BMP2	Acts as a disulphide-linked homodimer and induces bone and cartilage formation. It is a candidate as a retinoid mediator. Plays a key role in osteoblast differentiation	Chromosome: 20; Location: 20p12
BMP3	Induces bone formation.	Chromosome: 14; Location: 14p22
BMP4	Regulates the formation of teeth, limbs and bone from mesoderm. It also plays a role in fracture repair, epidermis formation, dorsal-ventral axis formation and ovarian follicle development.	Chromosome: 14; Location:14q22-q23
BMP5	Performs functions in cartilage development.	Chromosome: 6; Location: 6p12.1
BMP6	Plays a role in joint integrity in adults. Controls iron homeostasis via regulation of hepcidin.	Chromosome: 6; Location: 6p12.1
BMP8a	Involved in bone and cartilage development.	Chromosome: 1; Location: 1p35–p32

BMP	Functions	Gene Locus
BMP8b	Expressed in the hippocampus.	Chromosome: 1; Location: 1p35–p32
BMP10	May play a role in the trabeculation of the embryonic heart.	Chromosome: 2; Location: 2p14
BMP15	May play a role in oocyte and follicular development.	Chromosome: X; Location: Xp11.2

2.2.1 BMP Signaling pathway

The BMP signaling pathways regulate gene transcription. These pathways are initiated by the formation of heterotetramer. The BMP dimer binds to its type II receptor that recruits type I receptors resulting in the formation of heterotetramer with two receptors of each type. The type I receptor is phosphorylated by type II receptor. The SMAD cascade is one among the two pathways and the other pathway MAPK involves two mitogen activated protein kinase cascades. TGF-β family receptors use the SMAD signaling pathway to transduce signals (Figure 2.2.1). Phosphorylation of the cytoplasmic signaling molecules SMAD2 and SMAD3 takes place for the TGF-β pathway. The type I receptor is phosphorylated by type II receptor. Phosphorylation of R-SMAD 1, 5 and 8 takes place and the R-SMAD complex moves to nucleus in case of BMP pathway (Beck *et al.,* 2006, Chen *et al.,* 2004, Ikushima *et al.,* 2010, Kitisin *et al.,* 2007, Horbelt *et al.,* 2012, Meulmeester *et al.,* 2011).

The downstream effects of R-SMADs is prevented by dorsomorphin. The TGF – β and BMP signaling is highly regulated because of their important role in bone development and tissue homeostasis. Diverse biological effects are regulated by activated SMAD. They couple with transcription factors resulting in cell-state specific modulation of transcription. DLX2, ID1, ID2, JUNB, SOX4, STAT1 are the BMP responsive genes. The activin and BMP pathways are themselves attenuated by MAPK signaling at a number of levels (Candia *et al.,* 1997, de Gorter *et al.,* 2010, Kowanetz *et al.,* 2004, Montero *et al.,* 2008). In certain contexts, TGF-β signaling can also affect SMAD-independent pathways, including Erk, SAPK/JNK and p38 MAPK pathways. Activation

of SMAD-independent pathways through TGF-β signaling is also common (Verheyen *et al.,* 2007).

The intracellular BMP signaling has been many times found ambiguous whether it is a simple pathway or network. BMPs bind to specific serine/threonine kinase receptors, which transduce the signal to the nucleus through SMAD proteins. The BMP signaling is regulated by many factors at cytoplasmic, nucleic and cellular levels. The factors when identified revealed the principle behind diverse effects of BMP signaling. These factors are BMP inducible and they form negative feedback loop thus resulting in inhibition of BMP pathway. An interesting fact that was found out is that the members of BMP-SMAD pathway interact with the members of other pathways and resulting in crosstalk (Von Bubnoff *et al.,* 2001).

The present approach of research on BMPs is finding the clinical interventions of these proteins. Besides, playing a role in embryonic development and cellular functions, recombinant BMPs find its application in several clinical interventions such as non-union features and spinal fusions. Few knock out mouse models were constructed and the role BMPs was studied. These mouse models having mutated BMPs showed differences in heart, bone and cartilage developments. To know how BMP affects these developments, a tissue specific knock out must be studied (Chen *et al.,* 2004, Xiao *et al.,* 2007).

The BMP signaling pathway is initiated by the BMP receptors at the plasma membrane forming the heterotetrameric complex of type II and type I receptors and these are signaled by the SMAD to the nucleus. RGM proteins serve as co-receptors for BMPs. By interaction with membrane anchoring proteins R-SMADs exist in the cytoplasm. When phosphorylation by type I receptors takes place R-SMADs form complexes with SMAD - 4 (Figure 2.2.1). Translocation takes place into the nucleus and regulates transcription of target genes through interaction with transcription factor DNA-binding proteins and transcriptional co-activators. The growth of sophisticated technologies help in a better way in exploring signaling mechanisms of BMPs. Chromatin immunoprecipitation on microarray (ChIP-chip) analysis is a powerful method for identification of binding sites of transcription factors. The studies about BMP receptor inhibitors serve as a good source in understanding the regulation of many diseases such as anemia (Kohai Miyazono *et al.,* 2010, Schmierer *et al.,* 2007).

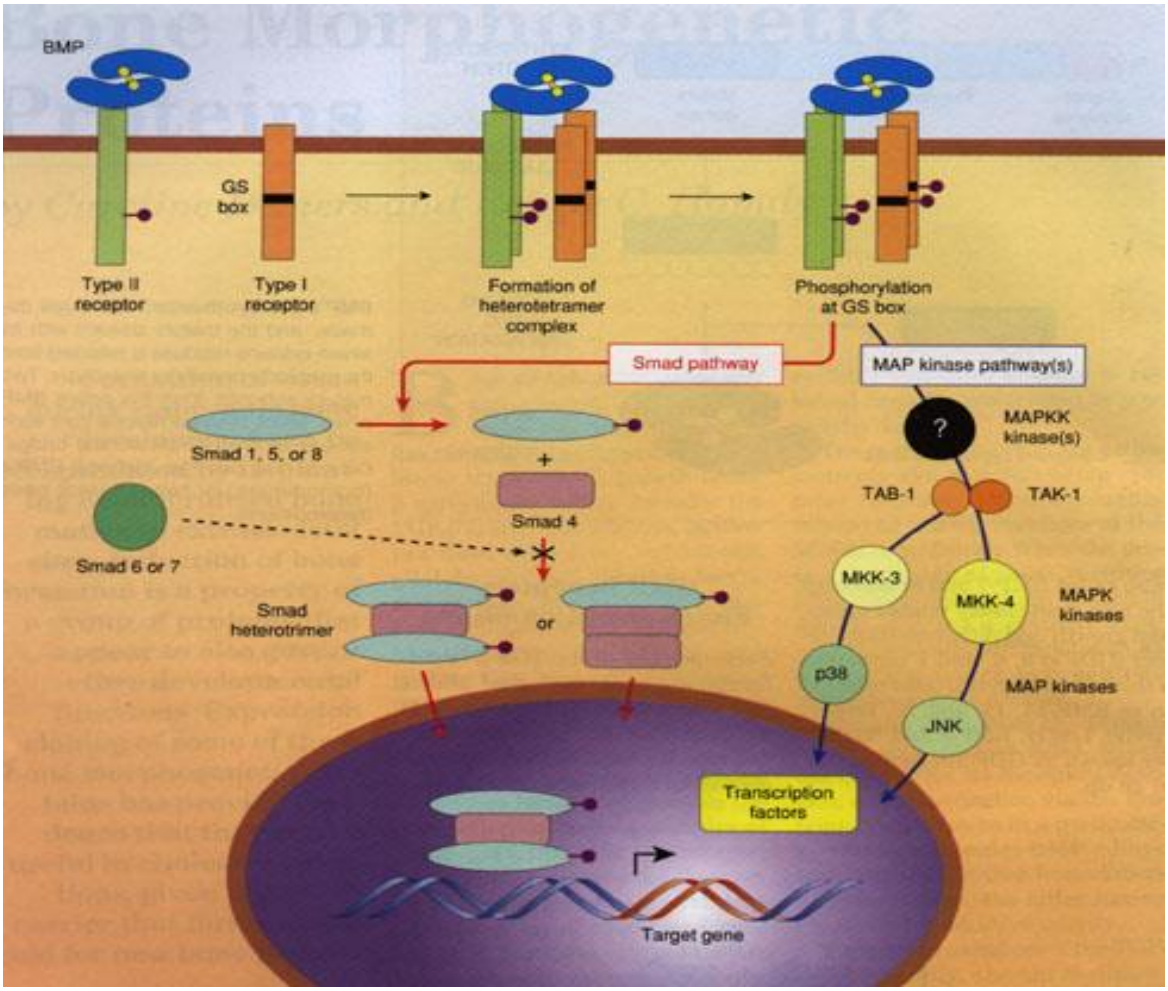

Figure 2.2.1 BMP Signaling Pathway (source**:** http://www.humpath.com)

2.3 Role of BMPs in breast cancer

Bone morphogenetic proteins exhibit diversified roles in breast cancer and they are dependent on time, concentration and cell type. The BMPs can serve as potential therapeutic agents because it shows differential patterns of expression in breast cancer. Immunohistochemical methods determine the location and distribution of BMPs. Few BMPs showed significant levels of expression namely BMP2 and BMP7 whereas few others showed inconsistent variation in transcript levels and they are BMP1, 3, 4 and 5. The BMPs and their role in breast cancer is still a mystery to many researchers for their unpredictable behavior in cancerous cells (Davies *et al.,* 2008).

Studies suggest that BMPs are involved in both cancer promotion and inhibition. This makes it vivid that further exploration of BMPs is essential to conclude their role in cancer. The BMPs are the ligand molecules that increase and decrease cancerous cell growth and migration (Alarmo *et al.,* 2010).

The inhibitory effect of BMPs in breast cancer is of much interest. This effect was proven when tumorous cells were induced by estrogen. Human breast cancer MCF-7 cells were found to express estrogen receptors α and β, BMP receptors and SMAD signaling molecules. MCF-7 cell proliferation was stimulated by estradiol and membrane-impermeable estradiol. Estradiol induced cell mitosis was suppressed by BMP 2, 4, 6 and 7. BMP2 and 4 decreased the expression more than the other BMPs. Estradiol induced differential modulation of expression in BMP receptor, thereby causing different BMP responses. BMP6 and 7 preferentially inhibited estradiol-induced p38 phosphorylation. The inhibitory effects of BMP6 and 7 on p38 signaling expression were functionally involved in the suppression of estrogen-induced mitosis of breast cancer cells (Otsuka *et al.,* 2009).

Comparison of two subsets of estrogen receptor positive breast cancers was carried out, in order to identify proteins involved in the progression of estrogen receptor positive breast cancers. These subsets differed in tumor grade, cytogenetic instability and tumor proliferation for their differential gene expression. The studies suggested that the bone morphogenetic protein receptor IB (BMPR-IB) plays a major role in the progression and dedifferentiation of breast cancer cells. In tumor, estrogen receptor expression indicates better tumor differentiation and clinical outcome in invasive breast cancer but there is also existence of estrogen receptor positive, poorly differentiated carcinomas showing poor clinical outcome (Mike Helms *et al.,* 2005).

Western blot analysis revealed that downstream signaling of this receptor is mainly mediated via phosphorylation of SMAD 1 in estrogen receptor positive breast cancer. Estrogen receptor positive and negative breast cancers showed BMPR-IB expression. But impact on tumor grade, proliferation and cytogenetic instability as parameters of tumor progression were demonstrated in estrogen receptor positive carcinomas. Significant anti-apoptotic activity complemented the pro-proliferative effect. This was indicated by XIAP and IAP-2 expression. The same effect was found in many of the BMPR related cancer. Activated BMP/SMAD pathway in breast cancer leads to progression and dedifferentiation in estrogen receptor positive breast cancer (Mike Helms *et al.,* 2005).

The evidence that BMP act as tumor suppressor can be found out from that it plays its role in tumorigenesis. BMPs are also known as growth stimulators and anti-proliferative

or anti-growth molecules. BMPs act as growth and proliferation inhibitor but they are also found to enhance the growth and proliferation of the tumorous cells. They behave differently in different cancer environments and their mechanism of action is also different at molecular level. Proliferation is inhibited by BMP2 and BMP5. They also modulate steroidogenesis. This was found in human adrenocortical tumor cells carried out as an *in vitro* study. BMP2 act as growth inhibitor by creating cell cycle arrest in the G1-phase mediated by p21/WAF1/CIP1. Pepsinogen II - a gene serves as differentiation marker of glandular cells of stomach, is activated by this activity of BMP2 and there by indicating its role in gastric cancer. BMP2 is found to have proliferative and apoptotic activity on cancerous cells (Wen *et al.,* 2004). Furthermore, BMP2 was also reported for its repressing effect on inhibitory activity of TGF-β pathway and there by establishing less proliferative cells of breast cancer. Similar effects were reported on BMP7 and BMP6 (Du *et al.,* 2008).

Migration and invasion are found to be characteristic of BMPs. BMP7 was reported to have enhanced cell migration and invasion. This differs from cell line to cell line and from an environment to another. Increased suppression of growth rate, invasion and migration of cells were found to be the regulatory effects of BMP7. BMP6 also exhibit similar action during metastasis and angiogenesis processes (Masudha *et al.,* 2003). BMPs effect was profoundly seen in prostate cancer. In prostate cancer cells BMP2, 4 and 7 have seen to be having an effect. BMP2 have been reported for its role in many types of cancer such as human pancreatic cancer, ovarian cancer, mesothelioma, mucinous adenocarcinoma and many more (Hatakeryama *et al.,* 1993, Le Page *et al.,* 2006).

2.4 Role of BMP4 and BMP7 in breast cancer

The signal transduction pathway of BMP4 is closely associated with the transcriptional response of the target genes. BMP4 like every other TGF-β superfamily member binds to BMPR 1 and BMPR 2 (Figure 2.4.1). This binding is followed by signal transduction. This signal transduction can take place by SMAD pathway or MAPK pathway but both resulting in the transcription of the target genes of BMP4 represented in figure 2.4.1. At certain circumstances, transcription is affected by downstream signaling with in the cell and this is due to transphosphorylation of BMPR1 by BMPR2 (Miyazono *et al.,* 2010).

BMP4 belongs to the TGF-β superfamily. BMP4 plays its role in various types of cancer like breast cancer, brain cancer, bladder cancer and colorectal cancer. Breast cancer and effects of BMP4 is widely studied and such studies indicate that its expression is more in primary breast cancer tumor and cell lines.

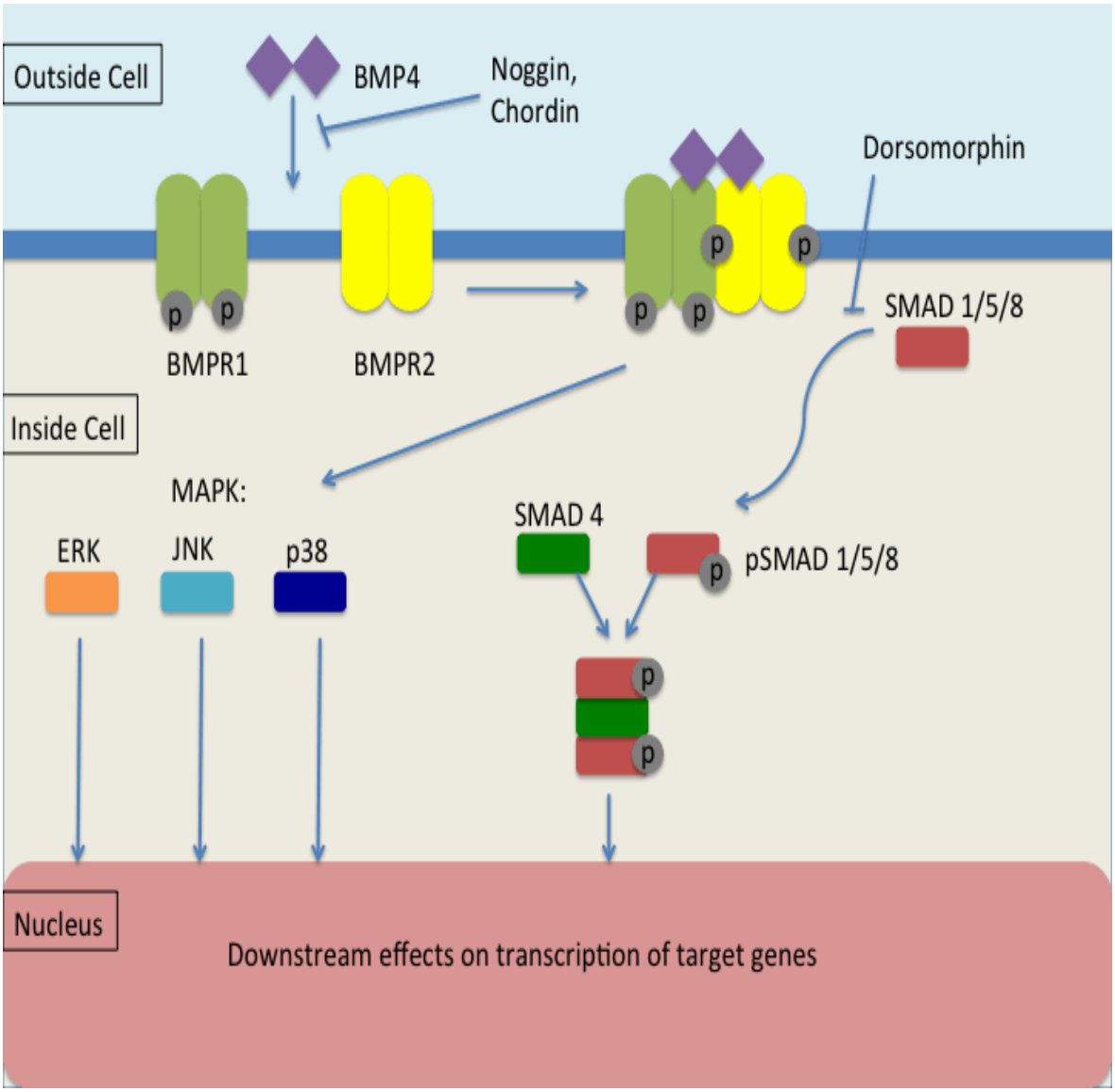

Figure 2.4.1 Signal Transduction Pathway of BMP4

(source: http://en.wikipedia.org/wiki/File:BMP4_Signal_Transduction_Pathways.gif)

The BMP4 effect was found in many types of breast cancers like node negative and steroid receptor positive. A series of findings had been reported on breast cancer and BMP4. In the year 2007 Montesano *et al.*, reported that in mammary epithelial cells invasive growth is promoted by exogenous treatment of BMP4. BMP4 was found capable of inducing growth factor that results in proliferation as the cells get induced.

From many findings it was found that the real role of BMP4 remains critical yet unknown.

Ketolainen *et al.,* 2010, had shown G1 cell cycle arrest leading to suppression of growth upon BMP4 treatment in nine different cell lines. The decrease in the metastatic activity in a cell line was reported by Shon *et al.,* 2009. Another exploratory work was carried out by Rodriguez-Martinez *et al.,* 2011 was an approach that revealed BMP4 and BMP7's transcriptional responses in breast cancer cell lines. The work involved seven breast cancer cell lines that were treated both with BMP4 and BMP7 and also three of the cell lines were treated with both of the ligands and the treatment was carried out on different hours of incubation from 30 minutes to 24 hours with an increase of hours as a time series experiment. The experiment was performed in a RNA microarray using Agilent whole human genome oligo array. The results were subjected to several types of filtration such as cell line specific, time point specific and general filtration. These filtrations were done in order to find out the differentially expressed genes.

The transcriptional response to BMP4 and BMP7 were stated as more number of differentially expressed genes was found in BMP4 treated than in BMP7 treated cell lines. But there may be a conclusion that it is true in all cases because the cell line that were treated with both BMP4 and BMP7 did not show any such results and further another analyses was made in order to find out the time point specific effect on the response of the genes. Temporal variation was noticed when these time point specific observations were made. Temporal variation is the pattern shown by gene clusters in such a way that they remain up regulated in some cell lines and down regulated in some cell lines. Later upon general filtering of the expression data, more pronounced effect of BMP4 was observed than BMP7 on gene transcription. Another unique outcome of this work was synexpression group of genes that is these genes show similar expression either up regulated or down regulated simultaneously. They also contain common cis and trans acting elements. There were 210 such probes resulting in 154 annotated genes. The genes were found to be involved in development and morphogenesis when they were subjected to GO analysis (Rodriguez-Martinez *et al.,* 2011).

Though BMP7 has been reported for its potential role in inhibiting cell proliferation in many forms of cancer like prostate, thyroid and colorectal cancer but its role in breast cancer is still under experimentation. Studies indicate that elevated levels of BMP7 in

breast tumor tissues when compared with normal mammary gland tissue. The expression of BMP7 in primary tumor has been extensively reported (Alarmo *et al.,* 2009). BMP7 is referred to as pleiotropic signaling molecule. Unlike other TGF-β family members, BMP7 reduces the bone metastasis formation and growth in tumorous cells by making their association in the early stages of bone metastasis. This work made a different approach to study the breast cancer cell phenotype regulated by BMP7. Among the cell lines chosen for the study, few cell lines were BMP7 silenced and few were added with BMP7.

The phenotype of each cell line was observed. The changes in growth of the BMP7 silenced cells were due to the mechanisms of BMP7 like G1 cell cycle arrest and resulted in inhibition of growth. On the other hand, the cells treated with BMP7 showed no apoptosis. There were significant changes in the cell migration, cell invasion and growth all being influenced by BMP7.

So far in prostate cancer, myeloma and anaplastic thyroid carcinoma addition of BMP7 was observed to reduce proliferation further enunciated by decreased tumor growth in mouse xenograft. The BMP7 phosphorylates the SMAD1/5/8 and there by BMP signaling pathway takes place in both types of cell lines that exhibited cell growth or cell inhibition. This proves that the SMAD 4 mutations are unlikely to happen in breast cancer cells and it was also expressed in different levels. It can be said that there are multiple factors that influence the BMP7 functioning (Alarmo *et al.,* 2009).

The study carried out with breast tumor cell lines and breast tumor tissue samples reveal that the expression was not uniform for all cell lines but were cell line specific. Few cell lines showed BMP7 expression at mRNA level. BMP7 expression was not enhanced by the epidermal growth factor (EGF). Usually EGF enhances the expression of BMPs as they are dependent on EGF for regulation and this kindles further interest in the behaviour of BMP7. BMP6 another family member of BMP was enhanced by EGF makes it a very intriguing mechanism. All these studies showed that BMP7 undergoes a unique expression level in different samples and under different conditions. There were no incidences of similar expression among uniformly treated cell lines. Either cell proliferation or cell cycle arrest was observed (Schwalbe *et al.,* 2003).

Overexpressed BMP7 in primary breast cancer was reported in a study carried out by (Alarmo *et al.,* 2006). In twenty two cell lines and in about hundred and forty six primary

breast tumor samples the gene copy number and expression of BMP7 was observed. In 16% of primary tumor cases, the gene copy number was increased. With the help of fluorescence *in situ* hybridization technique (FISH) variations in gene copy numbers of BMP7 in each cell line was determined. The increase in gene copy number does not greatly influence the mRNA levels in either cell lines or tissue samples. The alterations in the levels of BMP7 remain one of the reasons for the overexpression of BMP7 protein and also for the development of breast cancer apart from being playing an important role in vertebrate bone development.

2.5 DNA and RNA Microarrays and its applications in Gene Expression Profiling

Microarray technology employs a two dimensional array on a solid surface such as glass or silicon thin film which is used to assay higher quantity of biological samples. It was first invented by Patrick Brown in 1995. DNA and RNA microarrays are those that replaced the traditional biological assays. DNA microarrays have been used for many experiments that involve gene expression studies that measures DNA and RNA microarrays for studying thousands of messenger RNA (mRNA) transcripts. These serve as an invincible tool to study thousands of genes simultaneously and often referred to as a high-capacity system that takes few microliters of sample to determine several number of genes (Adi *et al.,* 2006, Stuart *et al.,* 2001, Ziv Bar 2004). Microarrays can be classified based on length of probes, manufacturing method and number of samples that can be profiled on to the array simultaneously.

Affymetrix, Agilent, Illumina and Nimblegen are the manufacturers of different kinds of arrays. The largest number of probes is found in complementary DNA array (cDNA) and lowest number of probes is found in oligonucleotide arrays. Here, probe stands for the nucleotide surface that is attached to the surface of the slide. The manufacturing methods also differ from array to array. Usually, cDNA arrays are manufactured using deposition, while oligonucleotide arrays are manufactured using in-situ technologies. There are three different types of in-situ technologies photolithography, ink-jet printing and electrochemical synthesis. The type of array used determines the type of experiment carried out. There are many types of array experiments such as Gene Expression Profiling, Single Nucleotide Polymorphisms detection, Tiling Array, Comparative Genomic Hybridization and many more (Maynard *et al.,* 2003, Romero *et al.,* 2002, Weinheim 2003) .

All these experiment types are useful in research and find its application in many ways such as the diagnosis and prognosis of hereditary diseases by detecting deletions and other mutations in genes. Single-channel arrays analyze a single sample at a time whereas multiple-channel arrays can analyze two or more samples simultaneously. Time series experiments using microarrays are challenging experiments where samples under different treatment and different conditions are studied and hence multiple channel arrays are used (Houts 2000). There are also growing need for algorithms in order to analyze these time series experiments (Schena 2000, Spellman *et al*, 1998, Friedman *et al.*, 2000).

In medicine exploring gene transcriptional responses and conditions to specific target, environment and drug is very essential to understand the mechanism of the disease and to develop therapies. Diseases that are not well characterized because of the lack of information on cellular, molecular and genetic mechanisms are answered easily using microarrays (Knudsen 2004, Leppert *et al,* 2006, Muhle *et al.,* 2001). The role of microarray is found very useful in complete cycle of disease mechanism such as identifying diagnostic or prognostic biomarkers, classifying diseases like tumors with different prognosis that are in differentiable by microscopic examination, monitoring the response to therapy and understanding the mechanisms involved in the genesis of disease processes.

The applications of microarray in various fields of biological science is fast growing and limitless. The following are few applications of it in scientific research:

Cancer research finds its easiest way out to study cancerous samples using microarray technology. In a research study conducted by (Grigoriadis *et al.,* 2006), a comparative study was done by isolating malignant neoplastic epithelial cells from primary breast cancers, luminal and myoepithelial cells isolated from normal human breast tissue by immunomagnetic separation methods. Massively parallel signature sequencing (MPSS) and four different genome wide microarray platforms were used for this expression profiling.

6,553 differentially expressed genes were identified by MPSS method. Microarray profiling of primary tumor epithelial cells and 90% of normal luminal cells was carried out. Using microarray technology significant expression level changes between these two samples were detected and 4,149 transcripts were found out. There were 8,051 genes that

resulted in a combined differential tumor epithelial transcriptome. A list of 907 and 955 transcripts whose expression differed between luminal epithelial cells and myoepithelial cells were identified by microarray gene signatures. As a result of these findings, genes that encode for periostin were found out. In this way the molecular changes in malignant epithelial cells in breast cancers were studied using microarray.

Whole genome sequencing is another fascinating field of research made easy with the application of microarray. The whole genome of several organisms are sequenced by sequencing methods and differentially expressed genes are found out using arrays. Studies carried out by (Kristen *et al.,* 2006) involved whole genome sequencing. Growth Hormone (GH) is important in the development and maintenance of bone. The molecular pathways such as insulin-like growth factors (IGF) still remain to be studied in detail to understand its association with GH in a wider way. This experiment showed the GH signaling pathway using mice which was deficient of GH and later was injected with it to study the differences in the pathways.

The microarray analysis of the RNA samples showed important aspects such as differential expression of six thousand one hundred and sixty genes followed by higher number genes that were up regulated after injection of growth hormone. The study was also able to characterize the genes as expressed sequence tags (EST) and many genes were categorized under Gene Ontology. On the microarray, nineteen thousand and eighty one genes were found to an intensity of greater than hundred and there were three hundred and ninety five genes that were categorized under forty nine GO categories (Kristen *et al.,* 2006).

Though the IGF independent and dependent pathways by which growth hormone acts in bone still remains unrevealed, with the help of microarray several genes that are associated with common signaling pathways were revealed on whole genome had been a notable contribution. These types of experiments greatly reveal information on associated genes and sequencing of those genes may help in finding out different forms of genetic abnormalities, percentage of inheritance of those abnormalities to the progeny and also causatives of such abnormalities as in most cases the carcinogens.

Breast cancer research and its importance had been already discussed in the previous parts and here the role of microarray in experiments to understand the multiple molecular interactions associated with breast cancer is being discussed. An excellent, integrated,

approach towards understanding the multiple molecular events and mechanisms by which cancer may develop can be studied with the application of gene expression profiling using microarray. Cellular pathways such as cellular pathways including cell cycle, growth, survival and apoptosis are disrupted by the process of oncogenesis. A snapshot of the complete cellular transcriptome on a single microarray chip provided by the microarray hybridization helps the researchers to find out the complex interactions of genes (Plamena *et al.,* 2010).

Changes in gene expression of epithelial and stromal cells occur during progression of cancer. The stromal cells exhibit extensive gene expression and these were studied using expression profiling. Fourteen patients with primary ductal breast cancer were chosen and 78.6% were ER positive and 78.6% were lymph node positive. They all belonged to premenopausal group and their stromal and epithelial tumor cells were studied for expression under different stages of tumor progression (Xiao *et al.,* 2009, de visser *et al.,* 2006, Clement *et al.,* 2008, Strausberg *et al.,* 2005).

This exploratory analysis of microenvironment in breast cancer is very limited till date. These findings reveal that the tumor associated stromal cells expression is influenced by the components of the extracellular matrix and the extracellular matrix remodeling matrix metalloproteases. The interactions between tumor cells and the various immune cells are complex however, ranging from tumor growth-suppressing effects to tumor growth-promoting effects (Xiao *et al.,* 2009, de visser *et al.,* 2006, Clement *et al.,* 2008, Strausberg *et al.,* 2005).

Several specific gene alterations have been implicated in breast cancer progression. Evaluation of global gene expression patterns using microarray technology helps in understanding the molecular basis of the disease. Gene expression in four different breast cancer cell lines T47D, MDA231, SKBR3 and BT474 were studied using cDNA microarrays containing 1,700 and 19,000 sequence-verified human cDNAs (Nalan *et al.,* 2001).

Each hybridization was compared between Cy5-labeled complementary DNA from one of the cell lines with Cy3-labeled cDNA from a reference sample. This work aimed at finding out the consistency and reproducibility of the technology and for that purpose subsequent hybridization with reciprocal labeling was carried out. With this method, a system was developed for analysis of breast cancer samples and also to determine genes

that have common pathways that are also involved in breast carcinogenesis (Nalan *et al.*, 2001).

3. OBJECTIVES

1. Detection of target genes of BMP4 and BMP7 ligands from top most regulated genes upon BMP4 and BMP7 treatment of the breast cancer cell lines.
2. Detection of BMP4 target genes from list of top most regulated genes upon BMP4 treatment of the breast cancer cell lines.
3. Analysis of Gene Ontology, Transcription Factor Target (Enrichment Analysis) and KEGG pathway of genes from cell line samples treated with BMP4 and BMP7 and those treated only with BMP4 using WEB-based GEne SeT AnaLysis Toolkit (WebGestalt) tool.
4. Analysis of Gene Ontology, Transcription Factor Target (Enrichment Analysis) and KEGG pathway of group C genes that resulted from hierarchical clustering.

4. MATERIALS AND METHODS

4.1 MATERIALS

4.1.1 Microarray data

The microarray data was downloaded from gene expression omnibus (GEO) database using the accession number GSE31605. This microarray data was submitted by Rodriguez-Martinez, Laboratory of Cancer Genetics, Institute of Biomedical Technology, University of Tampere and Centre of Laboratory of Medicine, Tampere University Hospital, Finland. (Analysis of BMP4 and BMP7 signaling in breast cancer cells unveils time-dependent transcription patterns and highlights a common synexpression group of genes, Rodriguez-Martinez *et al.*, BMC Medical Genomics 2011, 4:80). This microarray data was used for the entire analysis of this work.

The experiment carried out by Rodriguez-Martinez was expression profiling by array and the platform used was Agilent -014850 Whole Human Genome Oligo Microarray 4x44K G4112F. This experiment involved treatment of seven breast cancer cell lines namely HCC1954, MDA-MB-361, ZR-75-30, HCC1419, SK-BR-3, MDA-MB-231 and T-47D with recombinant human BMP4 and BMP7 proteins (ligands). The cell lines HCC1954, MDA-MB-361, ZR-75-30 were treated with both BMP4 and BMP7 separately while HCC1419, SK-BR-3 cell lines were treated only with BMP4. MDA-MB-231 and T-47D cell lines were treated with only BMP7 (Table 4.1.1). Total RNA from these cell lines were extracted at six different time points like 30min, 1h, 3h, 6h, 12h and 24h after the treatment with BMP4 and BMP7. Later, these RNA were labeled with Cy-5 and vehicle treated cells with Cy-3 for generating fluorescence and were hybridized to the microarray and resulting in a two-color data. The scanned microarray data had been submitted to GEO. For this work, the cell lines treated with both BMP4 and BMP7 and also the cell lines treated with only BMP4 were chosen for analysis. The genes that respond to the signaling of the ligands were found out.

Table 4.1.1: List of treatment of cell lines with BMP4 and BMP7

Name of the cell line	**Name of the ligands used for treatment**
HCC1954	BMP4 and BMP7
MDA-MB-361	BMP4 and BMP7

Name of the Cell line	Name of the ligands used for treatment
ZR-75-30	BMP4 and BMP7
HCC1419	BMP4
SK-BR-3	BMP4
MDA-MB-231	BMP7
T-47D	BMP7

4.1.2 Hierarchically clustered group C genes

A list containing expression values represented as log2 ratios was provided as an additional file from the Laboratory of Cancer Genetics, Institute of Biomedical Technology, University of Tampere and Centre of Laboratory of Medicine, Tampere University Hospital, Finland. The group C genes showed a common synexpression in response to BMP4 and BMP7 signaling and it had 210 probes. These set of genes revealed elements of similarity in transcriptional response to the BMP4 and BMP7 treatment indicating that they share common signaling pathways of BMP4 and BMP7. The functions of these genes are regulation of gene expression and regulation of development and morphogenesis. Hence, was found interesting to explore further in this work.

4.1.3 Tools used for analysis of Microarray data

For the analysis of microarray data, Bioconductor version 2.12 package in R programming environment version 2.15.2 was used. Bioconductor is an open source platform that serves to carry out pre-processing, quality assessment, differential expression, clustering and classification, gene set enrichment analysis and genetical genomics of data sets by developing various packages. It can be used on various platforms like Affymetrix, Illumina, Nimblegen and Agilent. Limma, a library of Bioconductor was used to analyze, construct linear models and to study differential expression of genes and also for producing heat maps. This library served as an inevitable tool for finding significant genes. Annotation Database Interface (AnnotationDbi) and Genomic Features are packages curated by Bioconductor were also used. AnnotationDbi provides user interface and database connection code for annotation data packages using SQLite data storage and required by annotations of all types. Genomic Feature exposes an annotation database generated from UCSC by exposing

these as FeatureDb objects. plyr is a package of tools for splitting, applying and combining data which was also used. gplots is a package for plotting data which was used to produce heat maps from the resulting significant genes after each analysis.

4.1.4 WEB-based GEne SeT AnaLysis Toolkit

This is an online tool with so many enrichment analysis options like Phenotype Analysis, Disease Association Analysis, Drug Association Analysis and Cytogenetic Band Analysis. To study the Gene Ontology, Transcription Factor Target (Enrichment Analysis) and KEGG pathway of genes WebGestalt was used and the above mentioned enrichment analyses were carried out in this work using gene lists. There are three different GO slim classification analysis provided by WebGestalt namely Biological process, Molecular function and Cellular component.

4.2 METHODS

In order to carry out various analyses the R programming environment was set up by downloading the R version 2.15.2 from http://cran.r-project.org/. The cran is the repository for the packages in R and the packages in Bioconductor can be downloaded from its own website. The Bioconductor packages were installed by typing the following commands in R window:

source (http://bioconductor.org/biocLite.R)

biocLite()

Further, the packages GenomicFeatures and AnnotationDbi were installed by typing the following command in R window:

biocLite(c("GenomicFeatures", "AnnotationDbi"))

The package plyr was installed from the option ‘Install Package’ in R window and plyr was selected from the list. gplots was downloaded as zip file from (http://cran.r-project.org/web/packages/gplots/index.html) and unzipped in R window.

The microarray data series was obtained from GEO by browsing under the option ‘Series’ using the accession number GSE31605 (http://www.ncbi.nlm.nih.gov/geo/). From the webpage, GSE31605_RAW.tar file was downloaded and the link is as follows (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE31605). A tab delimited text file named as *‘Targets.txt’* was created which served as the target frame. This file contains

different columns such as FileName which gives the name of the image analysis output file, Name, CellType - name of the cell lines, Cy5 Cy3 - RNA type labelled with the dye, Time - 30mins, 1h, 3h, 6h, 12h and 24h. The target file needs to have the columns FileName, Cy3, Cy5 but other columns like time, date, replicates are optional. The purpose of this file is that it lists the RNA target hybridized to each channel of each array and has a row of information on each microarray in general.

4.2.1 Analysis of cell lines treated with both BMP4 and BMP7

The cell lines treated with both BMP4 and BMP7 were MBA-MB-361, HCC-1954, ZR-75-30. Limma, plyr and gplots libraries were loaded in R for this analysis. At first, the *'Target.txt'* file was read in and followed by the data function of studying the foreground and background intensities of the images in to an RG list. RG is Red-Green list which is a data object and a class used to store raw two color intensities as they are read in from an image analysis output file and then it is read. Normalization is a preprocessing step used to correct the systematic difference between genes or arrays. This step is also known as background correction where subtraction of background intensity from the foreground intensity is carried out for each spot. There are two types of normalization namely within array and between array. For this step, between array normalization with Aquantile method was carried out and this method normalizes intensities or log-ratios to be comparable across arrays. The quantile normalization when applied to the A-values makes the distributions uniform across each array and each channel. Normalization is also a quality control measure for Agilent two color data. The density plots were also constructed. These were the preliminary steps involved in the analysis.

As a next step, a model matrix was designed which describes the hybridization of RNA targets to the arrays in the experiment. Following the designing of a model matrix, a contrast matrix was built. The contrast matrix allows the coefficients defined by the design matrix to be combined into the contrasts of interest and specifies which comparison to be made between the RNA targets. In this step, it was analysis of both BMP4 and BMP7 RNA targets. In design matrix, each row corresponds to an array in the experiment and each column corresponds to a coefficient. The matrix which was constructed was used to fit a linear model. Emprical Bayes was the method used and it is one of the widely used methods in limma than other statistical methods like t - test, etc., it is the statistical approach with parameter inferred from the data.

The results after applying the empirical Bayes were stored in an object and top Table was saved. The top Table consisted of columns such as row numbers in the original data, logFC - log2 based fold change values, t- moderated t-statistics from the empirical Bayes method, P values, adj.P.value - p-values corrected for multiple comparisons using Benjamini & Hochberg's false discovery rate, B - log odds of DEG. Based on the log FC values the regulation of genes were estimated. The negative log FC values indicated down regulation and positive log FC values indicated up regulation of genes. The obtained top Table was subjected to various filtrations like filtering out the probes with p - values less than 0.00, positive and negative control probes were removed. For construction of heat map another table was created from the top Table. This table had only those probes that occurred in common in all the three cell lines and their log FC values were added to the table and were used to produce a heat map in R. In this way only significant probes that expressed in all the three cell lines were identified and their regulation was studied.

4.2.2 Analysis of cell lines treated only with BMP4

The cell lines treated only with BMP4 were alone chosen for this step. They were MBA-MB-361, HCC-1954, ZR-75-30, HCC-1419 and SK-BR-30. The analysis steps were so close to the analysis of both BMP4 and BMP7 treated cell lines but when selecting the targets only BMP4 was selected for analysis. As a first step, the *'Target.txt'* file was read and followed by the preliminary steps such as making RG list and normalization.

The linear model was set after constructing the design matrix and contrast matrix. Empirical Bayes method was applied and a top Table was obtained with columns such as row numbers, log FC, t values, P - values, adj.P.values and B values. The top Table was then filtered out from positive and negative controls. Common genes present in all the five cell lines were chosen. These genes and their log FC values were made as a separate table and were used for construction of heat map and their regulation was studied.

4.2.3 Gene Ontology, Transcription Factor, KEGG pathway analysis of cell lines treated with both BMP4 and BMP7 and only with BMP4

The list of probes that were generated from above analysis 4.2.1 and 4.2.2 were made in to separate text files that contain only the probe Id's column, as this web tool can process files in *".txt and .tsv"* format. At first, the text file consisting of probe Ids of both BMP4 and BMP7 was analyzed. The text file for BMP4 and BMP7 treated cell lines contained 92 probes and the text file for BMP4 treated cell lines had 81 probes. The text files

served as target files for all the three analyses namely Gene Ontology, Transcription Factor and KEGG pathway. The following steps were performed for Gene Ontology analysis:

- In the webpage of WebGestalt (http://bioinfo.vanderbilt.edu/webgestalt/), the target text file was uploaded and Homo sapiens were selected as the organism of interest.
- As a next step, Gene ID type was chosen. The list of genes that corresponds to the type of gene list that was uploaded had been selected and it was Agilent whole genome 4*44K_v1 of Homo sapiens for this analysis.
- From the menu under the "Enrichment Analysis" - GO analysis was chosen and the analysis was carried out with following parameters such as reference set for enrichment analysis - hsapiens_Agilent_ Wholegenome_4*44k_v1. This parameter decides the reference set of genes to be compared with for statistical analysis in order to find out enriched gene sets. Hypergeometric statistical method was chosen as a default parameter for enrichment evaluation analysis. Multiple test adjustment was an essential parameter which adjusts the p value using Benjamini and Hochberg method.
- Top 10 significant pathways were chosen by selecting the default parameter top 10 for significance level. The final parameter was selecting a minimum number of genes from the uploaded file that serves as most interesting genes in a pathway and so that the pathway can be considered. It was set as two in order to get those pathways that had minimum of two significant genes involved in it from the uploaded file.

The result was obtained as separate file and was downloaded. The same procedure was followed for Transcription Factor analysis and KEGG pathway analysis but with a change in the step where under 'Enrichment Analysis' - Transcription Factor and KEGG pathway were chosen for each analysis respectively among many other analysis that were listed. For all the analyses the parameters remained the same as above. For GO, Transcription Factor and KEGG pathway analyses of only BMP4 treated cell lines, the probe Ids were extracted and made into a target text file and the other steps were carried out in the same way as it was carried out for both BMP4 and BMP7 treated cell lines using WebGestalt. The result files were downloaded and saved.

4.2.4 Gene Ontology, Transcription Factor, KEGG pathway analysis of group C genes

List of probe Ids of group C genes from excel file were extracted and made in to a text file that contains only the probe Id's column, as this web tool can process files in *".txt* and *.tsv"* format. It contained 210 probes in total. This text file served as target file for all the three analyses namely Gene Ontology, Transcription Factor and KEGG pathway. The following steps were performed for Gene Ontology analysis.

- In the webpage of WebGestalt (http://bioinfo.vanderbilt.edu/webgestalt/), the target text file was uploaded and Homo sapiens were selected as the organism of interest.
- As a next step, Gene ID type was chosen. The list of genes that corresponds to the type of gene list that was uploaded had been selected and it was Agilent whole genome 4*44K_v1 of Homo sapiens for this analysis.
- From the menu under the "Enrichment Analysis" – GO analysis was chosen and the analysis was carried out with following parameters such as reference set for enrichment analysis - hsapiens_Agilent_ Wholegenome_4*44k_v1. This parameter decides the reference set of genes to be compared with for statistical analysis in order to find out enriched gene sets. Hypergeometric statistical method was chosen as a default parameter for enrichment evaluation analysis. Multiple test adjustment was an essential parameter which adjusts the p value using Benjamini and Hochberg method.
- Top 10 significant pathways were chosen by selecting the default parameter top 10 for significance level. The final parameter was selecting a minimum number of genes from the uploaded file that serves as most interesting genes in a pathway and so that the pathway can be considered. It was set as two in order to get those pathways that had minimum of two significant genes involved in it from the uploaded file.

The result was obtained as separate file and was downloaded. The same procedure was followed for Transcription Factor Analysis and KEGG pathway analysis but with a change in the step where under 'Enrichment Analysis' – Transcription Factor and KEGG pathway were chosen for each analysis respectively. For all the analyses the parameters remained the same and result files were saved.

5. RESULTS

5.1 Results of analysis of cell lines treated with both BMP4 and BMP7

5.1.1 Heat map

Heat map was constructed in order to find out the up and down regulated genes using the list of genes from the top Table. Figure 5.1.1 shows about ninety three probes that codes for different genes and their log FC values plotted against each other for each line. From the heat map it was seen that the genes followed different patterns of expression in each cell line.

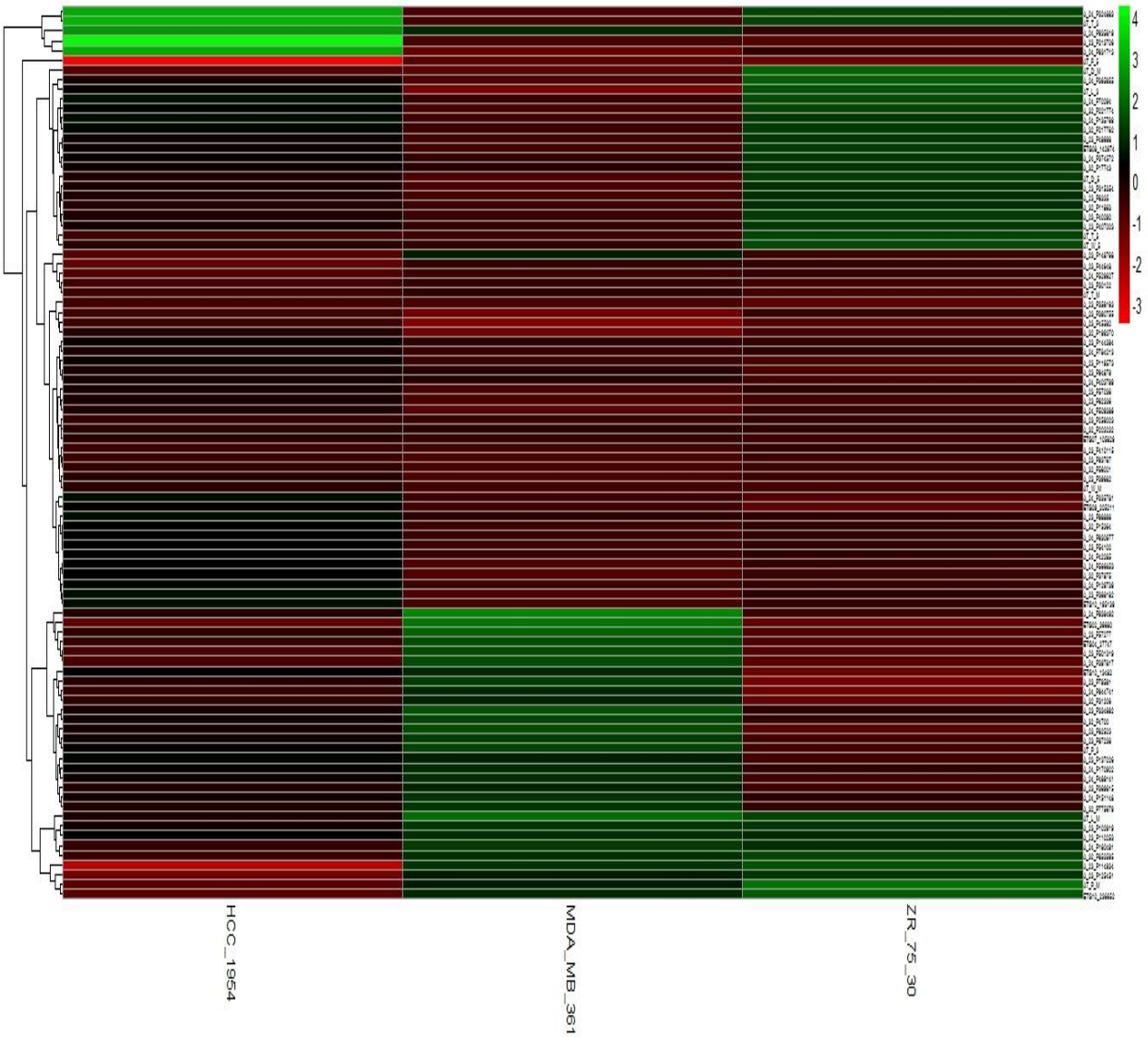

Figure 5.1.1 Heat map showing over expressed and under expressed genes

Heat map showing over expressed and under expressed genes of cell lines HCC_1954, MDA_MB_361 and ZR_75_30 treated with BMP4 and BMP7. In the heat map, each row represents a gene, represented by its probe ID and the columns represent respective cell lines. The log FC values of each gene in each cell line was plotted in order to read the expression of genes. The heat map showing, up regulated genes in red and down expressed genes in green and was identified as indicating different levels of expression of same gene but differing from cell line to cell line (Figure 5.1.1).

The red color indicates over expressed or up regulated genes while green color indicates under expressed or down regulated genes. There were twenty genes that under expressed in cell line ZR_75_30 but the same set of genes were over expressed in cell lines HCC_1954 and MDA_MB_361. Further, it was also seen that the heat map showing another set of twenty one genes which were under expressed in MDA_MB_361 cell line but over expressed in HCC_1954 and ZR_75_30. In ZR_75_30 and MDA_MB_361 nine genes were found to under express and in HCC_1954 they had remained showing over expression. The set of genes were also found to have biological functions when they were subjected to GO analysis.

5.1.2 GO analysis

Among 92 probe Ids that were uploaded, 42 probes were mapped to Entrez Ids. The GO analysis provided enriched GO categories under Biological processes, Molecular function and Cellular component. The following table (5.1.2 a & b) shows the set of genes and their biological functions. The genes were grouped under one biological process based on the values such as number of reference genes in the category (C), number of genes in the gene set and also in the category (O), expected number in the category (E), Ratio of enrichment (R), p value from hypergeometric test (rawP) and p value adjusted by multiple test adjustment (adjP). Similarly the tables were obtained for set of genes for molecular functions and cellular components. Here, genes involved in positive regulation of isotype switching to IgG isotypes and peptidyl-tyrosine phosphorylation are shown. There are four genes playing a role in positive regulation of isotype switching to IgG isotypes and five genes taking part in peptidyl-tyrosine phosphorylation. The genes were grouped based on the p value (Table 5.1.2 a & b).

Table 5.1.2 (a) Biological processes of GO analysis

GO	Biological Process	C	O	E	R	rawP	adjP
0048304	positive regulation of isotype switching to IgG isotypes	7	4	0.01	269.68	5.75e-10	3.97e-0

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P125451	NA	5788	ENSG00000081237	PTPRC	protein tyrosine phosphatase, receptor type, C
A_23_P57036	NA	958	ENSG00000101017	CD40	CD40 molecule, TNF receptor superfamily member 5
A_23_P30122	NA	3558	ENSG00000109471	IL2	interleukin 2
A_23_P213706	NA	3565	ENSG00000113520	IL4	interleukin 4

Table 5.1.2 (b)

GO	Biological Process	C	O	E	R	rawP	adjP
0018108	peptidyl-tyrosine phosphorylation	201	8	0.43	18.78	6.64e-09	9.29e-07

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P125451	NA	5788	ENSG00000081237	PTPRC	protein tyrosine phosphatase, receptor type, C
A_23_P93787	NA	3082	ENSG00000019991	HGF	hepatocyte growth factor (hepapoietin A; scatter factor)
A_24_P397817	NA	3952	ENSG00000174697	LEP	leptin
A_23_P57036	NA	958	ENSG00000101017	CD40	CD40 molecule, TNF receptor superfamily member 5
A_23_P110253	NA	3815	ENSG00000157404	KIT	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog

5.1.3 Transcription Factor analysis

Among 92 probes that were uploaded, 42 probes were mapped to Entrez Ids. The table (5.1.3 a & b) shows some of the significantly enriched gene sets. The Entrez genes were represented by Entrez Ids and the description defines about the gene. The following defines each of the category by which the set of genes were categorized - number of reference genes in the category (C), number of genes in the gene set and also in the category (O), expected number in the category (E), Ratio of enrichment (R), p value from hypergeometric test (rawP) and p value adjusted by multiple test adjustment (Table 5.1.3 a & b).

Table 5.1.3 (a) Transcription Target: hsa_TGCTGAY_UNKNOWN DB_ID:2367
C=530; O=6; E=0.52; R=11.62; rawP=1.21e-05; adjP=0.0012

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_24_P335781	NA	116	ENSG00000141433	ADCYAP1	adenylate cyclase activating polypeptide 1 (pituitary)
A_23_P148768	NA	2153	ENSG00000198734	F5	coagulation factor V (proaccelerin, labile factor)
A_23_P6335	NA	3053	ENSG00000099937	SERPIND1	serpin peptidase inhibitor, clade D (heparin cofactor), member 1
A_23_P407203	NA	9839	ENSG00000169554	ZEB2	zinc finger E-box binding homeobox 2
A_23_P501319	NA	51666	ENSG00000005981	ASB4	ankyrin repeat and SOCS box containing 4
A_23_P213706	NA	3565	ENSG00000113520	IL4	interleukin 4

Table 5.1.3 (b) Transcription Target: hsa_V$NKX3A_01 DB_ID:2109
C=232; O=4; E=0.23; R=17.70; rawP=7.78e-05; adjP=0.0019

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P110253	NA	3815	ENSG00000157404	KIT	v-kit Hardy-Zuckerman 4
A_23_P410115	NA	150350	ENSG00000176177	ENTHD1	ENTH domain containing 1
A_23_P407203	NA	9839	ENSG00000169554	ZEB2	zinc finger E-box binding homeobox 2
A_23_P501319	NA	51666	ENSG00000005981	ASB4	ankyrin repeat and SOCS box containing 4

5.1.4 KEGG pathway analysis

42 probes were mapped to Entrez Ids out of 92 probes that were uploaded and their KEGG pathways are listed below (Table 5.1.4 a & b). Each gene was categorized based on its associated pathway and also the description of the function of the gene in that pathway was provided. The table shows the information on Entrez Id, Ensembl Id, Gene symbol and description for each of the probe Id that was uploaded.

Table 5.1.4 (a) KEGG pathway: Cytokine-cytokine receptor interaction 04060 C=265; O=7; E=0.26; R=27.12; rawP=6.86e-09; adjP=1.17e-07

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P369815	NA	356	ENSG00000117560	FASLG	Fas ligand (TNF superfamily, member 6)
A_23_P93787	NA	3082	ENSG00000019991	HGF	hepatocyte growth factor (hepapoietin A; scatter factor)
A_24_P397817	NA	3952	ENSG00000174697	LEP	leptin
A_23_P57036	NA	958	ENSG00000101017	CD40	CD40 molecule, TNF receptor superfamily member 5

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P110253	NA	3815	ENSG00000157404	KIT	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog
A_23_P30122	NA	3558	ENSG00000109471	IL2	interleukin 2
A_23_P213706	NA	3565	ENSG00000113520	IL4	interleukin 4

Table 5.1.4 (b) KEGG pathway: Allograft rejection 05330 C=37; O=4; E=0.04; R=111.01; rawP=5.01e-08; adjP=4.26e-07

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P369815	NA	356	ENSG00000117560	FASLG	Fas ligand (TNF superfamily, member 6)
A_23_P57036	NA	958	ENSG00000101017	CD40	CD40 molecule, TNF receptor superfamily member 5
A_23_P30122	NA	3558	ENSG00000109471	IL2	interleukin 2
A_23_P213706	NA	3565	ENSG00000113520	IL4	interleukin 4

The tables 5.1.2, 5.1.3, 5.1.4 are showing few genes and their GO, Transcription factor and KEGG pathways. The rest of the genes and their information are provided in the appendices. Each table is divided in to parts a & b as they differ in their functions.

5.2 Results of analysis of cell lines treated only with BMP4

5.2.1 Heat map

Heat map was constructed in order to find out the up and down regulated genes using the list of genes from the top Table. Figure 5.2.1 shows about eighty one genes and their log FC values plotted against each other for each line. From the heat map it was seen that the genes followed different patterns of expression in each cell line. The red color indicates over expressed or up regulated genes while green color indicates under expressed or down regulated genes. There were nine genes that under expressed in all the cell lines.

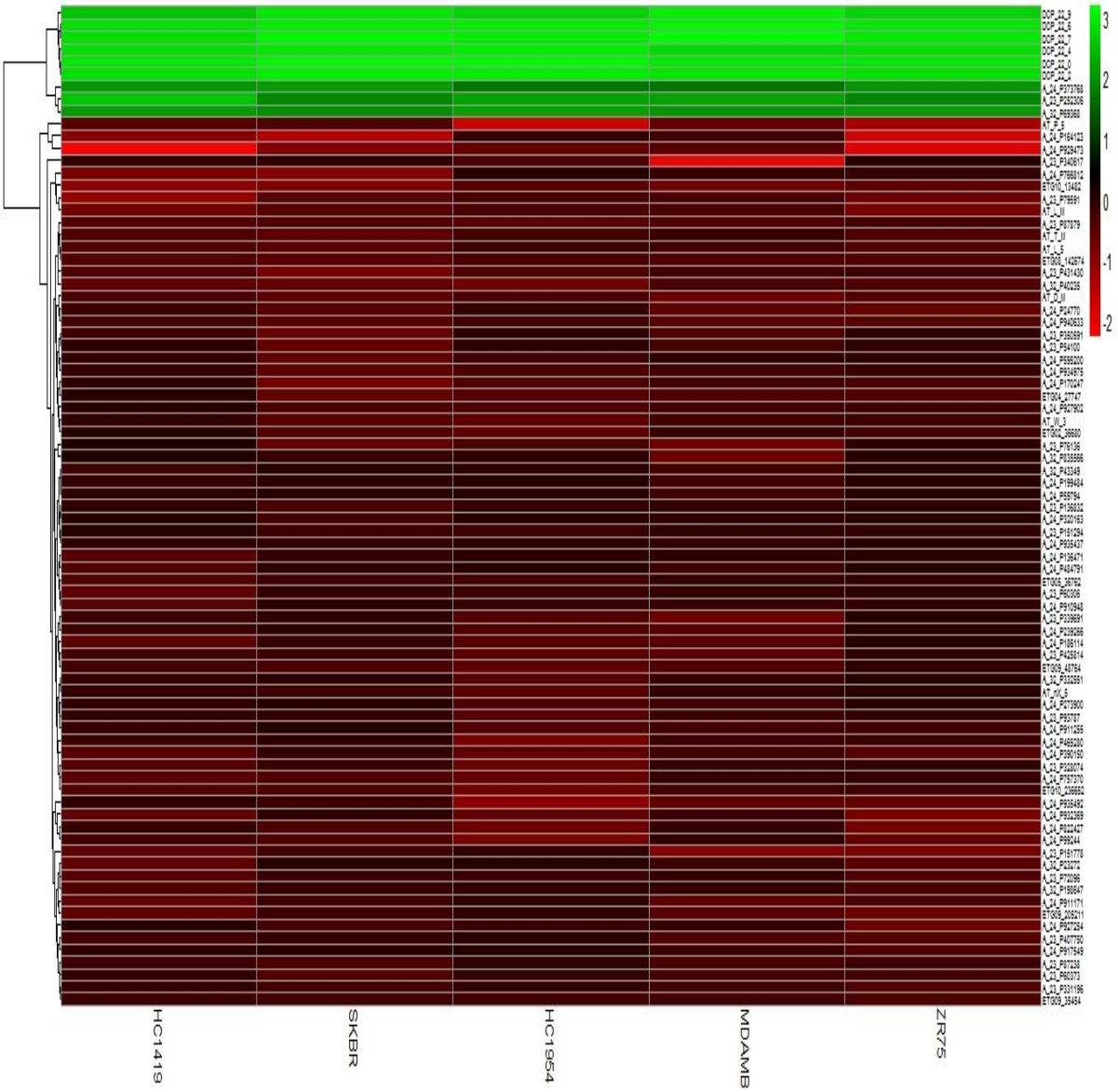

Figure 5.2.1 Heat map showing up regulated and down regulated genes of BMP4 signaling in cell lines

Heat map showing up regulated and down regulated genes of cell lines HCC_1419, SK_BR_30, HCC_1954, MDA_MB_361, ZR_75_30 treated with BMP4. A very few set of genes were found to under express and several genes were found to over express in all the five cell lines. Each row represents each gene, represented by its probe ID and the columns represent respective cell lines. The log FC values of each gene in each cell line was plotted in order to read the expression of genes. The heat map showing up regulated genes in red and down expressed genes in green and was identified as indicating different levels of expression of same gene but differing from cell line to cell line (Figure 5.2.1).

There were seventy two genes in cell lines HCC_1954, HCC_1419, SK_BR_30, ZR_75_30 and MDA_MB_361 that showed over expression. The set of genes were also found to have biological functions when they were subjected to GO analysis.

5.2.2 GO analysis

Among 81 probe Ids that were uploaded, 33 probes were mapped to Entrez Ids. The GO analysis provided enriched GO categories under Biological processes, Molecular function and Cellular component. The following table (5.2.2 a & b) shows the set of genes and their biological functions. The genes were grouped under one biological process based on the values such as number of reference genes in the category (C), number of genes in the gene set and also in the category (O), expected number in the category (E), ratio of enrichment (R), p value from hypergeometric test (rawP) and p value adjusted by multiple test adjustment. Similarly the tables were obtained for Molecular functions and Cellular components using the same set of genes.

Table 5.2.2 (a) Biological processes of GO analysis

GO	**Biological Process**	**C**	**O**	**E**	**R**	**rawP**	**adjP**
0030855	epithelial cell differentiation	307	7	0.55	12.83	7.81e-07	0.0003

User Id	**Value**	**Entrez ID**	**Ensembl Gene Stable ID**	**Gene Symbol**	**Description**
A_23_P339691	NA	353135	ENSG00000186226	LCE1E	late cornified envelope 1E
A_32_P69368	NA	3398	ENSG00000115738	ID2	inhibitor of DNA binding 2, dominant negative helix-loop-helix protein
A_23_P151294	NA	3458	ENSG00000111537	IFNG	interferon, gamma
A_23_P252306	NA	3397	ENSG00000125968	ID1	inhibitor of DNA binding 1, dominant negative helix-loop-helix protein
A_23_P54100	NA	2100	ENSG00000140009	ESR2	estrogen receptor 2 (ER beta)
A_23_P72096	NA	3552	ENSG00000115008	IL1A	interleukin 1, alpha

Table 5.2.2 (b)

GO	Biological Process	C	O	E	R	rawP	adjP
0060429	epithelium development	604	9	1.07	8.38	5.46e-07	0.0003

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P151294	NA	3458	ENSG00000111537	IFNG	interferon, gamma
A_23_P252306	NA	3397	ENSG00000125968	ID1	inhibitor of DNA binding 1, dominant negative helix-loop-helix protein
A_23_P328074	NA	6299	ENSG00000103449	SALL1	sal-like 1 (Drosophila)

5.2.3 Transcription Factor analysis

There were 33 probes mapped with Entrez Ids out of 81 probes that were loaded. The table (5.2.3 a & b) shows significantly enriched gene sets. The Entrez genes were represented by Entrez Ids and the description defines about the gene. The following table defines each of the category by which the set of genes were categorized - number of reference genes in the category (C), number of genes in the gene set and also in the category (O), expected number in the category (E), ratio of enrichment (R), p value from hypergeometric test (rawP) and p value adjusted by multiple test adjustment.

Table 5.2.3 (a) Transcription Target: hsa_V$FOXJ2_01 DB_ID:2088 C=181; O=3; E=0.14; R=21.66; rawP=0.0004; adjP=0.0107

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_32_P69368	NA	3398	ENSG00000115738	ID2	inhibitor of DNA binding 2, dominant negative helix-loop-helix protein
A_23_P252306	NA	3397	ENSG00000125968	ID1	inhibitor of DNA binding 1, dominant negative helix-

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
					loop-helix protein
A_24_P199484	NA	9068	ENSG00000116194	ANGPTL1	angiopoietin-like 1

Table 5.2.3 (b) Transcription Target: hsa_TATAAA_V$TATA_01 DB_ID:2456
C=1276; O=6; E=0.98; R=6.15; rawP=0.0004; adjP=0.0107

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_32_P69368	NA	3398	ENSG00000115738	ID2	inhibitor of DNA binding 2, dominant negative helix-loop-helix protein
A_23_P151294	NA	3458	ENSG00000111537	IFNG	interferon, gamma
A_23_P252306	NA	3397	ENSG00000125968	ID1	inhibitor of DNA binding 1, dominant negative helix-loop-helix protein
A_23_P54100	NA	2100	ENSG00000140009	ESR2	estrogen receptor 2 (ER beta)
A_23_P87879	NA	969	ENSG00000110848	CD69	CD69 molecule
A_24_P199484	NA	9068	ENSG00000116194	ANGPTL 1	angiopoietin-like 1

5.2.4 KEGG pathway analysis

There were 33 probes that were mapped to Entrez Ids out of 81 probes and their KEGG pathways are listed below (Table 5.2.4 a, b & c). Each gene was categorized based on its associated pathway and also the description of the gene in that pathway was provided. The table shows the information on Entrez Id, Ensembl Id, Gene symbol and description for each of the probe Id that was uploaded.

Table 5.2.4 (a) KEGG pathway: Malaria 05144 C=51; O=3;E=0.04; R=76.88; rawP=8.29e-06; adjP=0.0001

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P151294	NA	3458	ENSG00000111537	IFNG	interferon, gamma
A_23_P93787	NA	3082	ENSG00000019991	HGF	hepatocyte growth factor (hepapoietin A; scatter factor)
A_23_P60306	NA	7099	ENSG00000136869	TLR4	toll-like receptor 4

Table 5.2.4 (b) KEGG pathway: Cytokine-cytokine receptor interaction 04060 C=265; O=4; E=0.20; R=19.73; rawP=4.96e-05; adjP=0.0002

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P151294	NA	3458	ENSG00000111537	IFNG	interferon, gamma
A_23_P60373	NA	3467	ENSG00000177047	IFNW1	interferon, omega 1
A_23_P93787	NA	3082	ENSG00000019991	HGF	hepatocyte growth factor (hepapoietin A; scatter factor)
A_23_P72096	NA	3552	ENSG00000115008	IL1A	interleukin 1, alpha

Table 5.2.4 (c) KEGG pathway: Rheumatoid arthritis 05323 C=91; O=3; E=0.07; R=43.08; rawP=4.74e-05; adjP=0.0002

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P151294	NA	3458	ENSG00000111537	IFNG	interferon, gamma
A_23_P72096	NA	3552	ENSG00000115008	IL1A	interleukin 1, alpha

5.3 Group C genes analysis

5.3.1 GO analysis

From the 210 probes loaded for the analysis, 119 were mapped to be Entrez user Ids. The GO analysis provided enriched GO categories under Biological processes, Molecular function and Cellular component. The following table (5.3.1) shows the set of genes and their biological functions. The genes were grouped under one biological process based on the values - number of reference genes in the category (C), number of genes in the gene set and also in the category (O), expected number in the category (E), ratio of enrichment (R), p value from hypergeometric test (rawP) and p value adjusted by multiple test adjustment. Similarly the tables were obtained for set of genes for Molecular functions and Cellular components.

Table 5.3.1 Biological processes of GO analysis

GO	**Biological Process**	**C**	**O**	**E**	**R**	**rawP**	**adjP**
0050789	Regulation of biological process	8132	70	50.58	1.38	1.82 e-05	0.0075

User Id	**Value**	**Entrez ID**	**Ensembl Gene Stable ID**	**Gene Symbol**	**Description**
A_24_P14731	NA	27344	ENSG00000102109	PCSK1N	proprotein convertase subtilisin/kexin type 1 inhibitor
A_24_P416595	NA	284346	ENSG00000176472	ZNF575	zinc finger protein 575
A_24_P113725	NA	6658	ENSG00000134595	SOX3	SRY (sex determining region Y)-box 3
A_24_P213884	NA	7148	ENSG00000168477	TNXB	tenascin XB
A_24_P216165	NA	1050	ENSG00000245848	CEBPA	CCAAT/enhancer binding protein (C/EBP), alpha
A_23_P54573	NA	161882	ENSG00000179588	ZFPM1	zinc finger protein, multitype 1
A_24_P44916	NA	148170	ENSG00000167617	CDC42EP5	CDC42 effector protein (Rho GTPase binding) 5
A_24_P279060	NA	10021	ENSG00000138622	HCN4	hyperpolarization activated cyclic nucleotide-gated potassium channel 4
A_24_P10674	NA	115703	ENSG00000004777	ARHGAP33	Rho GTPase activating protein

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
					33
A_24_P878992	NA	10297	ENSG00000115266	APC2	adenomatosis polyposis coli 2
A_24_P203315	NA	254263	ENSG00000174871	CNIH2	cornichon homolog 2 (Drosophila)
A_24_P251040	NA	9048	ENSG00000117407	ARTN	artemin

5.3.2 Transcription Factor analysis

There were 119 probes mapped with Entrez Ids out of 210 probes that were loaded. The table shows significantly enriched gene sets. The Entrez genes were represented by Entrez Ids and the description defines about the gene. The following table (5.3.2) defines each of the category by which the set of genes were categorized - number of reference genes in the category (C), number of genes in the gene set and also in the category (O), expected number in the category (E), ratio of enrichment (R), p value from hypergeometric test (rawP) and p value adjusted by multiple test adjustment.

Table 5.3.2 Transcription Target: hsa_CAGGTG_V$E12_Q6 DB_ID:2409
C=2450; O=30; E=6.70; R=4.48; rawP=2.55e-12; adjP=8.64e-10

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_24_P416595	NA	284346	ENSG00000176472	ZNF575	zinc finger protein 575
A_24_P43959	NA	55691	ENSG00000151474	FRMD4A	FERM domain containing 4A
A_24_P227326	NA	283248	ENSG00000167771	RCOR2	REST corepressor 2
A_23_P34554	NA	777	ENSG00000198216	CACNA1E	calcium channel, voltage-dependent, R type, alpha 1E subunit
A_24_P117782	NA	85508	ENSG00000215397	SCRT2	scratch homolog 2, zinc finger protein (Drosophila)
A_24_P328446	NA	3196	ENSG00000115297	TLX2	T-cell leukemia homeobox 2
A_24_P113725	NA	6658	ENSG00000134595	SOX3	SRY (sex determining region Y)-box

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
					3
A_23_P96072	NA	2902	ENSG00000176884	GRIN1	glutamate receptor, ionotropic, N-methyl D-aspartate 1

5.3.3 KEGG pathway analysis

There were 119 probes that were mapped to Entrez Ids out of 210 probes and their KEGG pathways are listed below (Table 5.3.3). Each gene was categorized based on its associated pathway and also the description of the gene was provided. The table shows the information on Entrez Id, Ensembl Id, Gene symbol and description for each of the probe Id that was uploaded.

Table 5.3.3 KEGG pathway: Neuroactive ligand-receptor interaction 04080 C=272; O=6; E=0.74; R=8.06; rawP=0.0001; adjP=0.0014

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P153549	NA	2906	ENSG00000105464	GRIN2D	glutamate receptor, ionotropic, N-methyl D-aspartate 2D
A_23_P146885	NA	2837	ENSG00000181408	UTS2R	urotensin 2 receptor
A_23_P150162	NA	1815	ENSG00000069696	DRD4	dopamine receptor D4
A_23_P20458	NA	2831	ENSG00000183729	NPBWR1	neuropeptides B/W receptor 1
A_23_P101761	NA	84634	ENSG00000116014	KISS1R	KISS1 receptor
A_23_P96072	NA	2902	ENSG00000176884	GRIN1	glutamate receptor, ionotropic, N-methyl D-aspartate 1

6. DISCUSSION

6.1 Analysis of cell lines treated with both BMP4 and BMP7

This analysis showed number of genes that were regulated by BMP4 and BMP7 treatment. The signaling effect of both ligands together on these genes was found to vary between cell line to cell line. A set of genes showed over expression in two cell lines and under expression in one cell line. Another set of genes remained under expressed in two cell lines and over expressed in one cell line (Figure 5.1.1). This indicates the possibility of influence of one ligand over the other, thereby making dissimilar expression of genes in all the three cell lines. The GO analysis of these genes showed their biological process such as positive regulation of isotype switching, peptidyl-tyrosine modification, peptidyl-tyrosine phosphorylation, positive regulation of DNA regulation, peptide hormone receptor binding. The molecular functions were peptidase activity, growth factor activity, heparin binding, carbohydrate derivative binding, SMAD binding and enzyme binding. The pathways such as T cell receptor signaling pathway, pathways in cancer, primary immunodeficiency and autoimmune thyroid disease were found to be involving these genes.

The variable expression of BMPs in primary breast cancer samples and cell lines were demonstrated by Alarmo *et al.,* 2010 had shown that the BMP4 and BMP7 were expressed in majority in both cell lines and primary breast cancer samples and the levels of expression in normal cells were relatively less. The expression had widely targeted the genes in both cell line and samples. These genes when studied further may help to find out highly regulated gene and its network thereby in finding out the mechanism of action of BMP4 and BMP7. It can be said that the genes showed a contrast in expression in each cell line but shows essential information like pathways involved.

6.2 Analysis of cell lines treated only with BMP4

As BMP4 is widely studied for its role in cancer, the main focus of this thesis was to find out the genes responding to BMP4 treatment. It was seen that there were eighty one genes totally responding to the ligand. Nine genes were found to be under expressed in all the five cell lines showing uniform expression and seventy two genes were found to over express in all the five cell lines (Figure 5.2.1). Further, GO analysis of these genes showed their role in biological functions like Interleukin-1 production, Cytokine metabolic process, positive regulation of MHC class II biosynthetic process and epithelium development. The molecular functions were receptor binding, cytokine

activity, Serine-type endopeptidase activity and many more. KEGG pathways like TGF-beta signaling pathway, melanoma and Graft-versus-host disease were noticed involving these genes. These results indicate that the role of these genes in TGF-signaling pathway may be altered upon BMP4 treatment. If the network of pathways when studied, can bring out the interactions of these genes in the pathway and also the risk of diseases associated when these genes are not regulated.

The validation with primary breast cancer samples from database would bring about the comparison of these genes and their expression both at cell line and tissue samples there by establishing a comparison between *in vivo* and *in vitro* conditions playing a role in their expression. Earlier a comparative work was carried out by Alarmo *et al.,* 2010 had shown the effect of BMP4. It was reported that BMP4 had shown elevated levels of expression in primary tumors and cell lines when compared to all the other BMPs. Rodriguez-Martinez *et al.,* 2011 had reported the transcriptional regulation that leads to metabolic changes and cell proliferation in breast cancer cell lines. Hence, a comparative study with the genes that have been found out from this work can bring about some highly regulated genes of BMP4. Having used a microarray platform, the experiment has wider scope for comparison with other data sets that are also from same Agilent platform. Bioconductor would serve as a potential tool in comparison steps and can bring about new findings. On the whole, it can be said that this work had focused on throwing light on the genes that were regulated by BMP4 and BMP7 treatments and also provides a scenario for works in future.

Finding out the most responsive genes to BMP4 and its transcriptional responses would bring in some potential targets that may be developed as therapeutic agents for breast cancer and it may also serve in understanding the unrevealed mechanism of BMP4 in breast cancer. Webgestalt can be used in order to investigate disease association, drug association and also microRNA target analysis. It is essential in cancer studies to relate diseases associated to gene and its function. The malfunctioning of a gene may cause disease and by understanding this relationship, developing drugs would become easier.

6.3 Analysis of group C genes

The analysis of group C genes was carried out with an intention to find out the biological, molecular and cellular functions of the genes. The genes were found interesting from the data set as they were the synexpression group of genes. Synexpression group of genes are known to be involved in the same process and are all coordinately expressed and are found to have common cis and trans acting elements. There were 210 probes of which 119 were mapped to Entrez Ids that is found to be Entrez annotated genes.

The Gene Ontology analysis of the probes showed their role in essential biological processes like regulation of biological processes, epithelium development of kidney, positive regulation of transcription from RNA polymerase II promoter, regulation of multicellular organismal process, cellular nitrogen compound biosynthetic process and nephron epithelium morphogenesis. The molecular functions in which these genes involved are nucleic acid binding transcription factor activity, sequence-specific DNA binding transcription factor activity, DNA binding, sequence-specific DNA binding, RNA polymerase II distal enhancer sequence-specific DNA binding transcription factor activity, sequence-specific DNA binding RNA polymerase II transcription factor activity, nucleic acid binding and double-stranded DNA binding.

It is seen that all these processes are closely associated to transcription and it can said that the genes that had showed response to BMP4 and BMP7 treatment are all regulatory genes in nature and these two ligands had very strong effect on transcription process of these genes. Few of the pathways that are related to these genes are basal cell carcinoma pathway, cancer pathways and Wnt signaling pathway. These information help to find out the genes that play a major role in these pathways and by altering its expression defective pathways can be made functional and thereby preventing a non-functional pathway leading to diseases. These genes were found to be expressed in cell lines but their expression in primary breast cancer samples still remains unexplored. However, in future studies can be focused on these genes to find out their activities in tumor cells.

7. CONCLUSION

This work was an attempt towards finding out the target genes from the most regulated genes of BMP4 and BMP7 signaling. From the heat map results it was seen that there were a set of genes that formed clusters and showed expression to BMP4 and BMP7 signaling. Similarly there were clusters of genes that showed expression upon treatment only with BMP4. There were ninety three genes that showed expression upon BMP4 and BMP7 treatment and in case of BMP4 treatment it was eighty one genes. The results established the effect of influence of these ligands on the target genes. Further, validation of these genes would result in acquiring exactly or highly regulated genes by BMP4 and BMP7 signaling. From this work, the GO analysis, Transcription Factor analysis and KEGG pathways of the genes provide information on each gene and its role in transcription. As BMP4 is much studied for its role in transcriptional responses in these cell lines, the transcription factor analysis would bring in new insights on BMP4's transcriptional effect. These analytical studies indicated the role of genes in major pathways and also their role in many molecular processes. On the whole, the work attempted at bringing out the genes and their role when treated with ligands like BMP4 and BMP7.

The analysis of group C genes was made to find out further information as they were showing strong response to both BMP4 and BMP7. They contributed to a large set of genes about hundred and nineteen genes and they were all found to be involved in pathways like Wnt signaling pathway, basal cell carcinoma and many pathways in cancer. All these results from the work had created a basic incite to further explore on genes regulated by signaling of BMP4 and BMP7.

8. REFERENCES

1. Adi L.Tarca, Roberto Romero, Sorin Draghici (2006), Analysis of microarray experiments of gene expression profiling, *American Journal of Obstetrics and Gynecology*, 195 373-88.
2. Alarmo EL, Kallioniemi A (2010). Bone morphogenetic proteins in breast cancer: dual role in tumourigenesis? *Endocr Relat Cancer* Apr 21;17(2):R123-39.
3. Alarmo EL, T. Korhonen, T. Kuukasjarvi, H. Huhtala, K. Holli, A. Kallioniemi (2008), Bone morphogenetic protein 7 expression associates with bone metastasis in breast carcinomas, *Ann. Oncol* 19, pp. 308–314
4. Alarmo EL, Jenita Pärssinen, Johanna M. Ketolainen, Kimmo Savinainen, Ritva Karhu, Anne Kallioniemi (2009), BMP7 influences proliferation, migration and invasion of breast cancer cells, *Cancer Letters volume 275, Issue 1, 8 Pages 35–43*
5. Alarmo EL, T. Kuukasjarvi, R. Karhu, A. Kallioniemi (2007), A comprehensive expression survey of bone morphogenetic proteins in breast cancer highlights the importance of BMP4 and BMP7, *Breast Cancer Res. Treat* 103, pp. 239–246
6. Alarmo EL, J. Rauta, P. Kauraniemi, R. Karhu, T. Kuukasjarvi, A. Kallioniemi (2006), Bone morphogenetic protein 7 is widely overexpressed in primary breast cancer, *Genes Chromosomes Cancer*, 45, pp. 411–419
7. Annegien Broeks, Jos H. M. Urbanus, Arno N. Floore, Ellen C. Dahler, Jan G. M. Klijn, Emiel J. Th. Rutgers, Peter Devilee, Nicola S. Russell, Flora E. van Leeuwen and Laura J. van't Veer (2000), ATM-Heterozygous Germ line Mutations Contribute to Breast Cancer–Susceptibility, *Am J Hum Genet*, Feb 66(2): 494–500.
8. Beck SE, Jung BH, Fiorino A, Gomez J, Rosario ED, Cabrera BL, Huang SC, Chow JY, Carethers JM (2006), Bone morphogenetic protein signaling and growth suppression in colon cancer, *Am J Physiol Gastrointest Liver Physiol*, Jul;291(1):G135-45.
9. Bishop DT, (1990). BRCA1 and BRCA2 and breast cancer incidence: a review. *Ann Oncol*. 10 Suppl 6:113-9.
10. Cao Y, (2011). BMP4: a metastasis suppressor gene in breast cancer (http://repository.unimelb.edu.au/10187/11433). Source - PhD thesis report.

11. Candia A.F, Watabe T, Hawley S.H, Onichtchouk D, Zhang Y, Derynck R, Niehrs C and Cho K.W.(1997), Cellular interpretation of multiple TGF- signals: intracellular antagonism between activin/BVg1 and BMP-2/4 signaling mediated by Smads, *Development*, 124, 4467-4480.
12. Caroline Baynes, Catherine S Healey, Karen A Pooley, Serena Scollen, Robert N Luben, Deborah J Thompson, Paul DP Pharoah, Douglas F Easton, Bruce AJ Ponder, Alison M Dunning and the SEARCH breast cancer study, (2007), Common variants in the ATM, BRCA1, BRCA2, CHEK2 and TP53 cancer susceptibility genes are unlikely to increase breast cancer risk, *Breast Cancer Research*, 9:R27.
13. Chen D, Zhao M, Mundy GR (2004), Bone Morphogenetic Proteins, *Growth Factors* Dec;22(4):233-41.
14. Chen D, Zhao M, Harris SE, Mi Z (2004), Signal transduction and biological functions of bone morphogenetic proteins, *Front Biosci* Jan 1;9:349-58.
15. Clement JF, Meloche S, Servant MJ (2008), The IKK-related kinases: from innate immunity to oncogenesis, *Cell Res* 18:889-899.
16. Davies SR, Watkins G, Douglas-Jones A, Mansel RE, Jiang WG (2008). Bone morphogenetic proteins 1 to 7 in human breast cancer, expression pattern and clinical/prognostic relevance. *J Exp Ther Oncol.* 7(4):327-38.
17. de Gorter D.J, van D.M, Korchynskyi O and ten Dijke P. (2010), Biphasic effects of trans-forming growth factor - on bone morphogenetic protein-induced osteoblast differentiation, *J.Bone Miner.Res* Jun;26(6):1178-87
18. Desai K.V, C J Kavanaugh, A Calvo and J E Green (2002), Chipping away at breast cancer: insights from microarray studies of human and mouse mammary cancer, *Endocrine-Related Cancer* 9 207–220
19. de Visser KE, Eichten A, Coussens LM (2006), Paradoxical roles of the immune system during cancer development, *Nat Rev Cancer* 6:24-37.
20. Dumitrescu RG, Cotarla I (2005), Review Understanding breast cancer risk - where do we stand in 2005? *J Cell Mol Med.* Jan-Mar; 9(1):208-21.
21. Du J, Yang S, Wang Z, Zhai C, Yuan W, Lei R (2008) Bone morphogenetic protein 6 inhibit stress-induced breast cancer cells apoptosis via both Smad and p38 pathways *J Cell Biochem*. 103(5):1584–97

22. F Otsuka, M Takahashi, T Miyoshi, M Yamashita, H Otani, J Goto, T Ogura, H Doihara & H Makino (2009). Inhibitory effects of bone morphogenetic proteins on estrogen-induced proliferation of breast cancer cells. *Endocrine Abstracts* 19 P164
23. Ford D, D F Easton, M Stratton, S Narod, D Goldgar, P Devilee, D T Bishop, B Weber, G Lenoir, J Chang-Claude, H Sobol, M D Teare, J Struewing, A Arason, S Scherneck, J Peto, T R Rebbeck, P Tonin, S Neuhausen, R Barkardottir, J Eyfjord, H Lynch, B A Ponder, S A Gayther, M Zelada-Hedman (1998, Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. *Am J Hum Genet*. March; 62(3): 676–689.
24. Friedman N, Linial M, Nachman I. and Pe'er D. (2000) Using Bayesian network to analyze expression data, *J. Comput. Biol* 7, 601–620.
25. Georgia Chenevix-Trench, Amanda B. Spurdle, Magtouf Gatei, Helena Kelly, Anna Marsh, Xiaoqing Chen, Karen Donn, Margaret Cummings, Dale Nyholt, Mark A. Jenkins, Clare Scott, Gulietta M. Pupo, Thilo Dörk, Regina Bendix, Judy Kirk, Katherine Tucker, Margaret R. E. McCredie, John L. Hopper, Joseph Sambrook, Graham J. Mann and Kum Kum Khanna (2001), Dominant Negative ATM Mutations in Breast Cancer Families, *JNCI J Natl Cancer Inst* 94 (3): 205-215.
26. Grigoriadis A, Mackay A, Reis-Filho JS, Steele D, Iseli C, Stevenson BJ, Jongeneel CV, Valgeirsson H, Fenwick K, Iravani M, Leao M, Simpson AJ, Strausberg RL, Jat PS, Ashworth A, Neville AM, O'Hare MJ (2006), Establishment of the epithelial-specific transcriptome of normal and malignant human breast cells based on MPSS and array expression data, *Breast Cancer Res* 8(5):R56.
27. Herpin A, Cunningham C (2007), Cross-talk between the bone morphogenetic protein pathway and other major signaling pathways results in tightly regulated cell-specific outcome, *The FEBS J*, Jun;274(12):2977-85
28. Honrado E, Osorio A, Palacios J, Benitez J (2006), Review Pathology and gene expression of hereditary breast tumors associated with BRCA1, BRCA2 and CHEK2 gene mutations. *Oncogene*. Sep 25; 25(43):5837-45.
29. Horbelt D, Denkis A, Knaus P (2012), A portrait of transforming growth factor β superfamily signalling: Background matters, *Int. J. Biochem Cell Biol*, 44(3), 469–74.

30. Houts TM. Improved 2-color exponential normalization for microarray analyses employing cyanine dyes. Lin, S. 2000. Durham, NC, Duke University Medical Center. Proceedings of CAMDA 2000, ‘‘Critical assessment of techniques for microarray data mining’’.
31. Howlader N, Noone AM, Krapcho M, Garshell J, Neyman N, Altekruse SF, Kosary CL, Yu M, Ruhl J, Tatalovich Z, Cho H, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA (eds) (*1975-2010)*. *SEER Cancer Statistics Review*, National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/csr/1975_2010/, based on November 2012 SEER data submission, posted to the SEER web site, 2013.
32. Ikushima H, Miyazono K (2010), TGFbeta signalling: a complex web in cancer progression, *Nat. Rev. Cancer* 10(6), 415–24.
33. Ketolainen JM, Alarmo EL, Tuominen VJ, Kallioniemi A (2010), Parallel inhibition of cell growth and induction of cell migration and invasion in breast cancer cells by bone morphogenetic protein 4, *Breast Cancer Res Treat* Nov ;124(2):377-86
34. Kitisin K, Saha T, Blake T, Golestaneh N, Deng M, Kim C, Tang Y, Shetty K, Mishra B, Mishra L (2007), TGF-Beta signaling in development, *Sci. STKE* (399).
35. Kowanetz,M., Valcourt,U., Bergstrom,R., Heldin,C.H and Moustakas,A (2004), Id2 and Id3 define the potency of cell proliferation and differentiation responses to transforming growth factor- and bonemorphogenetic protein, *Mol.Cell Biol.*, 24, 4241-4254.
36. Knudsen S (2004), Guide to analysis of DNA microarray data. Hoboken, NJ: John Wiley & Sons, Inc.
37. Kristen E. Govoni, Seong Keun Lee, Robert B. Chadwick, Hongrun Yu, Yuji Kasukawa, David J. Baylink and Subburaman Mohan (2006), Whole genome microarray analysis of growth hormone-induced gene expression in bone: T-box3, a novel transcription factor, regulates osteoblast proliferation, *Am J Physiol Endocrinol Metab* 291: E128–E136.
38. Le Page C, Ouellet V, Madore J, Ren F, Hudson TJ, Tonin PN (2006) Gene expression profiling of primary cultures of ovarian epithelial cells identifies novel molecular classifiers of ovarian cancer, *Br J Cancer* 94(3):436–45

39. Leppert PC, Catherino WH, Segars J (2006), A new hypothesis about the origin of uterine fibroids based on gene expression profiling with microarrays, *Am J Obstet Gynecol* 195:415-20.
40. Masuda H, Fukabori Y, Nakano K, Takezawa Y, T CS, Yamanaka H (2003), Increased expression of bone morphogenetic protein-7 in bone metastatic prostate cancer. *Prostate* 54(4):268–74.
41. Maynard SE, Min JY, Merchan J, Lim KH, Li J, Mondal S, et al (2003), Excess placental soluble fms-like tyrosine kinase 1 (sFlt1) may contribute to endothelial dysfunction, hypertension and proteinuria in preeclampsia. *J Clin Inves*, 111:649-58.
42. Meulmeester E, Ten Dijke P (2011), The dynamic roles of TGF-β in cancer, *J. Pathol*, 223(2), 205–18.
43. Mike W Helms, Jens Packeisen, Christian August, Birgit Schittek, Werner Boecker, Burkhard H Brandt and Horst Buerger (2005), First evidence supporting a potential role for the BMP/SMAD pathway in the progression of oestrogen receptor-positive breast cancer, *Journal of Pathology*, 206: 366–376
44. Miyazono K, Kamiya Y, Morikawa M (January 2010), Bone morphogenetic protein receptors and signal transduction, *Journal of Biochem* 147 (1): 35–51
45. Miyazono K, Yuto Kamiya and Masato Morikawa (2010), Bone morphogenetic protein receptors and signal transduction, *J. Biochem* 147(1):35–51
46. Montero J.A., Lorda-Diez,C.I., Ganan,Y., Macias,D and Hurle,J.M. (2008), Activin/TGF-and BMP crosstalk determines digit chondrogenesis, *Dev.Biol.*, 321, 343-356.
47. Muhle RA, Pavlidis P, Grundy WN, Hirsch E(2001), A high-throughput study of gene expression in preterm labor with a subtractive microarray approach. *Am J Obstet Gynecol*, 185:716-24.
48. Mundy G.R (1997) Mechanisms of bone metastasis. *Cancer*, 80, 1546-1556
49. Naber, Hildegonda Petronella Henriëtte (2012), TGF-beta and BMP in breast cancer cell invasion, Signal Transduction and Ageing Section, Department of Molecular Cell Biology, Faculty of Medicine, Leiden University Medical Center (LUMC), Leiden University.

50. Nalan Gokgoz, Xiang Sun, Shelly Bull, Jim Woodgett & Irene Andrulis (2001), Analysis of gene expression patterns in breast cancer by microarray technology, *Nature* http://omim.org/entry/114480
51. Park Y, Shackney S, Schwartz R *(2009),* Network-based inference of cancer progression from microarray, *IEEE/ACM Trans Comput Biol Bioinform*, Apr-Jun; 6(2):200-12
52. Perou C.M, Sorlie T, Eisen M.B, van de Rijn M, Jeffrey S.S, Rees C.A, Pollack J.R, Ross D.T, Johnsen H, Akslen L.A, Fluge O, Pergamenschikov A, Williams C, Zhu S.X, Lonning P.E, Borresen-Dale A.L, Brown P.O and Botstein D. (2000), Molecular portraits of human breast tumors. *Nature*, 406, 747-752.
53. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lønning PE, Børresen-Dale AL, Brown PO, Botstein D (2000), Molecular portraits of human breast tumours. *Nature*. Aug 17; 406(6797):747-52
54. Perou CM., Therese Sørlie, Michael B. Eisen, Matt van de Rijn, Stefanie S. Jeffrey, Christian A. Rees, Jonathan R. Pollack, Douglas T. Ross, Hilde Johnsen, Lars A. Akslen, Oystein Fluge, Alexander Pergamenschikov, Cheryl Williams, Shirley X. Zhu, Per E. Lønning, Anne-Lise Børresen-Dale, Patrick O. Brown & David Botstein (200), Molecular portraits of human breast tumours, *Nature* 406, 747-752.
55. Plamena N. Gabrovska, Robert A. Smith, Larisa M. Haupt and Lyn R. Griffiths (2010), Gene Expression Profiling in Human Breast Cancer - Towards Personalised Therapeutics? *The Open Breast Cancer Journal 2,* 46-59 46
56. Rajnish Kumar, Anju Sharma and Rajesh Kumar Tiwari (2012), Application of microarray in breast cancer: An overview, *J Pharm Bio allied Science*, Jan-Mar; 4(1): 21–26.
57. Risch HA, McLaughlin JR, Cole DE, Rosen B, Bradley L, Kwan E, Jack E, Vesprini DJ, Kuperstein G, Abrahamson JL, Fan I, Wong B, Narod SA (2001), Prevalence and penetrance of germ line BRCA1 and BRCA2 mutations in a population series of 649 women with ovarian cancer, *Am J Hum Genet*, Mar;68(3):700-10.
58. Rodriguez-Martinez Alejandra, Emma-Leena Alarmo, Lilli Saarinen, Johanna Ketolainen, Kari Nousiainen, Sampsa Hautaniemi and Anne Kallioniemi (2011).

Analysis of BMP4 and BMP7 signaling in breast cancer cells unveils time-dependent transcription patterns and highlights a common synexpression group of genes. *BMC Medical Genomics* 4:80.

59. Romero R, Kuivaniemi H, Tromp G(2002). Functional genomics and proteomics in term and preterm parturition. *J Clin Endocrinol Metab*, 87:2431-4.
60. Schena M, Shalon D, Davis RW, Brown PO (1995), Quantitative monitoring of gene expression patterns with a complementary DNA microarray, *Science* 270:467-70.
61. Schena M 2000, Microarray biochip technology, Sunnyvale, CA: Eaton Publishing.
62. Schmierer B, Hill CS (2007), TGFbeta-SMAD signal transduction: molecular specificity and functional flexibility, *Nat. Rev. Mol. Cell Biol*, 8(12), 970–82.
63. Schwalbe M, J. Sanger, R. Eggers, A. Naumann, A. Schmidt, K. Hoffken, J.H. Clement (2003), Differential expression and regulation of bone morphogenetic protein 7 in breast cancer, *Int. J. Oncol*. 23, pp. 89–95
64. Shon SK, Kim A, Kim JY, Kim KI, Yang Y, Lim JS (2009), Bone morphogenetic protein-4 induced by NDRG2 expression inhibits MMP-9 activity in breast cancer cells, *Biochem Biophys Res Commun* Jul 24;385(2):198-203.
65. Sotiriou C, Neo S.Y, McShane L.M, Korn E.L, Long P.M, Jazaeri A, Martiat P, Fox S.B, Harris A. and Liu E.T. (2003), Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc.Natl.Acad.Sci.U.S.A*, 100, 10393-10398.
66. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron J.S, Nobel A., Deng S., Johnsen H, Pesich R, Geisler S, Demeter J, Perou C.M, Lonning P.E, Brown P.O, Borresen- Dale A.L and Botstein D. (2003, Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc.Natl.Acad.Sci.U.S.A*, 100, 8418-8423.
67. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Lonning PE, Børresen-Dale AL (2001), Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A*. Sep 11; 98(19):10869-74.

68. Spellman P.T, Sherlock G, Zhang M, Iyer V, Anders K, Eisen M, Brown P, Botstein D. and Futcher B. (1998), Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. *Mol. Biol.Cell*, 9, 3273–3297.
69. Strausberg RL (2005), Tumor microenvironments, the immune system and cancer survival. *Genome Biol*, 6:211.
70. Stuart C.Sealfon and Tearina T. Chu (2011), RNA and DNA microarrays, *Methods in Molecular Biology*, 671:3-34
71. Tan Ming and Yu Dihua (2007), Molecular mechanisms of erbB2-mediated breast cancer chemo resistance, *Advances in experimental medicine and biology*, 608:119-29.
72. van 't Veer L.J, Dai H, van de Vijver M.J, He Y.D, Hart A.A, Mao M, Peterse H.L, vander Kooy K, Marton M.J, Wieen A.T, Schreiber G.J, Kerkhoven R.M, Roberts C, Linsley P.S, Bernards R and Friend S.H. (2002), Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415, 530-536.
73. Verheyen EM (2007), Opposing effects of Wnt and MAPK on BMP/Smad signal duration, *Dev. Cell*, 13(6), 755–6
74. von Bubnoff A, Cho KW (2001), Intracellular BMP signaling regulation in vertebrates: pathway or network? *Developmental Biology*, 239(1):1-14
75. Weigelt B, Peterse J.L and van't Veer L.J. (2005) Breast cancer metastasis: markers and models. *Nat.Rev.Cancer*, 5,591-602.
76. Weinheim (2003), Analysing gene expression: a handbook of methods, possibilities and pitfalls, Germany: Wiley-VCH.
77. Wen XZ, Miyake S, Akiyama Y, Yuasa Y (2004), BMP-2 modulates the proliferation and differentiation of normal and cancerous gastric cells. *Biochem Biophys Res Commun* 316(1):100–6.
78. Xiao YT, Xiang LX, Shao JZ (2007), Bone morphogenetic protein, *Biochemistry, Biophysics. Research Community*, 362(3), 550 - 3.
79. Xiao-Jun Ma, Sonika Dahiya, Elizabeth Richardson, Mark Erlander and Dennis C Sgroi (2000), Gene expression profiling of the tumor microenvironment during breast cancer progression, *Breast Cancer Research* 11:R7
80. Ziv Bar-Joseph (2004), Analyzing time series gene expression data, *Bioinformatics*, vol: 20, no:16204, 2493-2503

9. APPENDICES

9.1 Analysis of cell lines treated with both BMP4 and BMP7

9.1.1 Tables showing results of GO analysis

Biological process: positive regulation of isotype switching to IgG isotypes
GO:0048304 C=7;O=4;E=0.01;R=269.68;rawP=5.75e-10;adjP=3.97e-07

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P125451	NA	5788	ENSG00000081237	PTPRC	protein tyrosine phosphatase, receptor type, C
A_23_P57036	NA	958	ENSG00000101017	CD40	CD40 molecule, TNF receptor superfamily member 5
A_23_P30122	NA	3558	ENSG00000109471	IL2	interleukin 2
A_23_P213706	NA	3565	ENSG00000113520	IL4	interleukin 4

Biological process: peptidyl-tyrosine phosphorylation
GO:0018108 C=201;O=8;E=0.43;R=18.78;rawP=6.64e-09;adjP=9.29e-07

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P125451	NA	5788	ENSG00000081237	PTPRC	protein tyrosine phosphatase, receptor type, C
A_23_P93787	NA	3082	ENSG00000019991	HGF	hepatocyte growth factor (hepapoietin A; scatter factor)
A_24_P397817	NA	3952	ENSG00000174697	LEP	leptin
A_23_P57036	NA	958	ENSG00000101017	CD40	CD40 molecule, TNF receptor superfamily member 5
A_23_P110253	NA	3815	ENSG00000157404	KIT	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog
A_23_P39682	NA	7535	ENSG00000115085	ZAP70	zeta-chain (TCR)

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
					associated protein kinase 70kDa
A_23_P30122	NA	3558	ENSG00000109471	IL2	interleukin 2
A_23_P213706	NA	3565	ENSG00000113520	IL4	interleukin 4

Biological process: peptidyl-tyrosine modification
GO:0018212 C=203;O=8;E=0.43;R=18.60;rawP=7.18e-09;adjP=9.29e-07

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P125451	NA	5788	ENSG00000081237	PTPRC	protein tyrosine phosphatase, receptor type, C
A_23_P93787	NA	3082	ENSG00000019991	HGF	hepatocyte growth factor (hepapoietin A; scatter factor)
A_24_P397817	NA	3952	ENSG00000174697	LEP	leptin
A_23_P57036	NA	958	ENSG00000101017	CD40	CD40 molecule, TNF receptor superfamily member 5
A_23_P110253	NA	3815	ENSG00000157404	KIT	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog
A_23_P39682	NA	7535	ENSG00000115085	ZAP70	zeta-chain (TCR) associated protein kinase 70kDa
A_23_P30122	NA	3558	ENSG00000109471	IL2	interleukin 2
A_23_P213706	NA	3565	ENSG00000113520	IL4	interleukin 4

Biological process: positive regulation of isotype switching
GO:0045830 C=12;O=4;E=0.03;R=157.31;rawP=8.07e-09;adjP=9.29e-07

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P125451	NA	5788	ENSG00000081237	PTPRC	protein tyrosine phosphatase, receptor type, C
A_23_P57036	NA	958	ENSG00000101017	CD40	CD40

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
					molecule, TNF receptor superfamily member 5
A_23_P30122	NA	3558	ENSG00000109471	IL2	interleukin 2
A_23_P213706	NA	3565	ENSG00000113520	IL4	interleukin 4
A_23_P30122	NA	3558	ENSG00000109471	IL2	interleukin 2
A_23_P213706	NA	3565	ENSG00000113520	IL4	interleukin 4

Biological process: positive regulation of tyrosine phosphorylation of STAT protein GO:0042531 C=42;O=5;E=0.09;R=56.18;rawP=2.45e-08;adjP=2.12e-06

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_24_P397817	NA	3952	ENSG00000174697	LEP	leptin
A_23_P57036	NA	958	ENSG00000101017	CD40	CD40 molecule, TNF receptor superfamily member 5
A_23_P110253	NA	3815	ENSG00000157404	KIT	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog
A_23_P30122	NA	3558	ENSG00000109471	IL2	interleukin 2
A_23_P213706	NA	3565	ENSG00000113520	IL4	interleukin 4

9.1.2 Tables showing results of Transcription Factor analysis

Transcription Target: hsa_TGCTGAY_UNKNOWN DB_ID:2367 C=530;O=6;E=0.52;R=11.62;rawP=1.21e-05;adjP=0.0012

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_24_P335781	NA	116	ENSG00000141433	ADCYAP1	adenylate cyclase activating polypeptide 1 (pituitary)
A_23_P148768	NA	2153	ENSG00000198734	F5	coagulation factor V (proaccelerin, labile factor)
A_23_P407203	NA	9839	ENSG00000169554	ZEB2	zinc finger E-box binding homeobox 2
A_23_P501319	NA	51666	ENSG00000005981	ASB4	ankyrin repeat and SOCS box containing 4

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P213706	NA	3565	ENSG00000113520	IL4	interleukin 4

Transcription Target: hsa_V$NKX3A_01 DB_ID:2109
C=232;O=4;E=0.23;R=17.70;rawP=7.78e-05;adjP=0.0019

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P110253	NA	3815	ENSG00000157404	KIT	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog
A_23_P410115	NA	150350	ENSG00000176177	ENTHD1	ENTH domain containing 1
A_23_P407203	NA	9839	ENSG00000169554	ZEB2	zinc finger E-box binding homeobox 2
A_23_P501319	NA	51666	ENSG00000005981	ASB4	ankyrin repeat and SOCS box containing 4

Transcription Target: hsa_CAGGTA_V$AREB6_01 DB_ID:2404
C=780;O=6;E=0.76;R=7.90;rawP=0.0001;adjP=0.0019

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P93787	NA	3082	ENSG00000019991	HGF	hepatocyte growth factor (hepapoietin A; scatter factor)
A_23_P57036	NA	958	ENSG00000101017	CD40	CD40 molecule, TNF receptor superfamily member 5
A_23_P54100	NA	2100	ENSG00000140009	ESR2	estrogen receptor 2 (ER beta)
A_24_P944741	NA	55805	ENSG00000109771	LRP2BP	LRP2 binding protein
A_23_P30122	NA	3558	ENSG00000109471	IL2	interleukin 2
A_23_P407203	NA	9839	ENSG00000169554	ZEB2	zinc finger E-box binding homeobox 2
A_24_P135769	NA	4013	ENSG00000110002	VWA5A	von Willebrand

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
					factor A domain containing 5A
A_23_P30122	NA	3558	ENSG00000109471	IL2	interleukin 2
A_23_P407203	NA	9839	ENSG00000169554	ZEB2	zinc finger E-box binding homeobox 2

Transcription Target: hsa_TGCCAAR_V$NF1_Q6 DB_ID:2436
C=709;O=5;E=0.69;R=7.24;rawP=0.0006;adjP=0.0065

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P93787	NA	3082	ENSG00000019991	HGF	hepatocyte growth factor (hepapoietin A; scatter factor)
A_23_P259003	NA	51555	ENSG00000114757	PEX5L	peroxisomal biogenesis factor 5-like
A_24_P335781	NA	116	ENSG00000141433	ADCYAP1	adenylate cyclase activating polypeptide 1 (pituitary)
A_23_P110253	NA	3815	ENSG00000157404	KIT	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog
A_23_P407203	NA	9839	ENSG00000169554	ZEB2	zinc finger E-box binding homeobox 2

Transcription Target: hsa_GATTGGY_V$NFY_Q6_01 DB_ID:2440
C=1141;O=6;E=1.11;R=5.40;rawP=0.0008;adjP=0.0078

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P57036	NA	958	ENSG00000101017	CD40	CD40 molecule, TNF receptor superfamily member 5
A_23_P114934	NA	84944	ENSG00000143194	MAEL	maelstrom homolog (Drosophila)

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_24_P944741	NA	55805	ENSG00000109771	LRP2BP	LRP2 binding protein
A_23_P368182	NA	254428	ENSG00000133065	SLC41A1	solute carrier family 41, member 1
A_23_P407203	NA	9839	ENSG00000169554	ZEB2	zinc finger E-box binding homeobox 2
A_23_P213706	NA	3565	ENSG00000113520	IL4	interleukin 4

9.1.3 Tables showing results of KEGG pathway analysis

KEGG pathway: Cytokine-cytokine receptor interaction 04060
C=265;O=7;E=0.26;R=27.12;rawP=6.86e-09;adjP=1.17e-07

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P369815	NA	356	ENSG00000117560	FASLG	Fas ligand (TNF superfamily, member 6)
A_23_P93787	NA	3082	ENSG00000019991	HGF	hepatocyte growth factor (hepapoietin A; scatter factor)
A_24_P397817	NA	3952	ENSG00000174697	LEP	leptin
A_23_P57036	NA	958	ENSG00000101017	CD40	CD40 molecule, TNF receptor superfamily member 5
A_23_P110253	NA	3815	ENSG00000157404	KIT	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog
A_23_P30122	NA	3558	ENSG00000109471	IL2	interleukin 2
A_23_P213706	NA	3565	ENSG00000113520	IL4	interleukin 4

KEGG pathway: Autoimmune thyroid disease 05320
C=52;O=4;E=0.05;R=78.99;rawP=2.03e-07;adjP=1.15e-06

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P369815	NA	356	ENSG00000117560	FASLG	Fas ligand (TNF superfamily, member 6)

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P57036	NA	958	ENSG00000101017	CD40	CD40 molecule, TNF receptor superfamily member 5
A_23_P30122	NA	3558	ENSG00000109471	IL2	interleukin 2
A_23_P213706	NA	3565	ENSG00000113520	IL4	interleukin 4

KEGG pathway: T cell receptor signaling pathway 04660
C=108;O=4;E=0.11;R=38.03;rawP=3.87e-06;adjP=1.64e-05

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P125451	NA	5788	ENSG00000081237	PTPRC	protein tyrosine phosphatase, receptor type, C
A_23_P39682	NA	7535	ENSG00000115085	ZAP70	zeta-chain (TCR) associated protein kinase 70kDa
A_23_P30122	NA	3558	ENSG00000109471	IL2	interleukin 2
A_23_P213706	NA	3565	ENSG00000113520	IL4	interleukin 4

KEGG pathway: Primary immunodeficiency 05340
C=35;O=3;E=0.03;R=88.01;rawP=5.50e-06;adjP=1.87e-05

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P125451	NA	5788	ENSG00000081237	PTPRC	protein tyrosine phosphatase, receptor type, C
A_23_P57036	NA	958	ENSG00000101017	CD40	CD40 molecule, TNF receptor superfamily member 5
A_23_P39682	NA	7535	ENSG00000115085	ZAP70	zeta-chain (TCR) associated protein kinase 70kDa

KEGG pathway: Intestinal immune network for IgA production 04672
C=48;O=3;E=0.05;R=64.18;rawP=1.44e-05;adjP=4.08e-05

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P57036	NA	958	ENSG00000101017	CD40	CD40 molecule,
A_23_P30122	NA	3558	ENSG00000109471	IL2	interleukin 2
A_23_P213706	NA	3565	ENSG00000113520	IL4	interleukin 4

KEGG pathway: Pathways in cancer 05200
C=326;O=4;E=0.32;R=12.60;rawP=0.0003;adjP=0.0007

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P369815	NA	356	ENSG00000117560	FASLG	Fas ligand (TNF superfamily, member 6)
A_23_P93787	NA	3082	ENSG00000019991	HGF	hepatocyte growth factor (hepapoietin A; scatter factor)
A_23_P110253	NA	3815	ENSG00000157404	KIT	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog
A_23_P98898	NA	1017	ENSG00000123374	CDK2	cyclin-dependent kinase 2

KEGG pathway: Asthma 05310
C=30;O=2;E=0.03;R=68.46;rawP=0.0004;adjP=0.0009

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P57036	NA	958	ENSG00000101017	CD40	CD40 molecule, TNF receptor superfamily member 5

9.2 Analysis of cell lines treated only with BMP4

9.2.1 Tables showing results of GO analysis

Biological process: epithelial cell differentiation
GO:0030855 C=307;O=7;E=0.55;R=12.83;rawP=7.81e-07;adjP=0.0003

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P339691	NA	353135	ENSG00000186226	LCE1E	late cornified envelope 1E
A_32_P69368	NA	3398	ENSG00000115738	ID2	inhibitor of DNA binding 2, dominant negative helix-loop-helix protein
A_23_P151294	NA	3458	ENSG00000111537	IFNG	interferon, gamma
A_23_P252306	NA	3397	ENSG00000125968	ID1	inhibitor of DNA binding 1, dominant negative helix-loop-helix protein
A_23_P54100	NA	2100	ENSG00000140009	ESR2	estrogen receptor 2 (ER beta)
A_23_P72096	NA	3552	ENSG00000115008	IL1A	interleukin 1, alpha
A_23_P328074	NA	6299	ENSG00000103449	SALL1	sal-like 1 (Drosophila)

Biological process: epithelium development
GO:0060429 C=604;O=9;E=1.07;R=8.38;rawP=5.46e-07;adjP=0.0003

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P151294	NA	3458	ENSG00000111537	IFNG	interferon, gamma
A_23_P252306	NA	3397	ENSG00000125968	ID1	inhibitor of DNA binding 1, dominant negative

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
					helix-loop-helix protein
A_23_P328074	NA	6299	ENSG00000103449	SALL1	sal-like 1 (Drosophila)
A_32_P69368	NA	3398	ENSG00000115738	ID2	inhibitor of DNA binding 2, dominant negative helix-loop-helix protein
A_24_P99244	NA	2252	ENSG00000140285	FGF7	fibroblast growth factor 7
A_23_P339691	NA	353135	ENSG00000186226	LCE1E	late cornified envelope 1E
A_23_P93787	NA	3082	ENSG00000019991	HGF	"hepatocyte growth factor (hepapoietin A; scatter factor)"
A_23_P72096	NA	3552	ENSG00000115008	IL1A	interleukin 1, alpha
A_23_P54100	NA	2100	ENSG00000140009	ESR2	estrogen receptor 2 (ER beta)

Biological process: Regulation of steroid metabolic process
GO:0019218 C=67;O=4;E=0.12;R=33.59;rawP=5.57e-06;adjP=0.0013

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P151294	NA	3458	ENSG00000111537	IFNG	interferon, gamma
A_23_P72096	NA	3552	ENSG00000115008	IL1A	interleukin 1, alpha
A_23_P79591	NA	338	ENSG00000084674	APOB	apolipoprotein B (including Ag(x) antigen)
A_24_P239266	NA	55004	ENSG00000149357	LAMTOR1	late endosomal/lysosomal adaptor, MAPK and MTOR

					activator 1

Biological process: Epithelial cell development
GO:0002064 C=82;O=4;E=0.15;R=27.45;rawP=1.25e-05;adjP=0.0021

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P151294	NA	3458	ENSG00000111537	IFNG	interferon, gamma
A_23_P252306	NA	3397	ENSG00000125968	ID1	inhibitor of DNA binding 1, dominant negative helix-loop-helix protein
A_23_P54100	NA	2100	ENSG00000140009	ESR2	estrogen receptor 2 (ER beta)
A_23_P328074	NA	6299	ENSG00000103449	SALL1	sal-like 1 (Drosophila)

Biological process: Positive regulation of MHC class II biosynthetic process
GO:0045348 C=8;O=2;E=0.01;R=140.67;rawP=8.45e-05;adjP=0.0027

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P151294	NA	3458	ENSG00000111537	IFNG	interferon, gamma
A_23_P60306	NA	7099	ENSG00000136869	TLR4	toll-like receptor 4

9.2.2 Tables showing results of Transcription Factor analysis

Transcription Target: hsa_V$FOXJ2_01 DB_ID:2088
C=181;O=3;E=0.14;R=21.66;rawP=0.0004;adjP=0.0107

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_32_P69368	NA	3398	ENSG00000115738	ID2	inhibitor of DNA binding 2, dominant negative helix-loop-helix protein
A_23_P252306	NA	3397	ENSG00000125968	ID1	inhibitor of DNA binding 1, dominant negative helix-loop-helix protein
A_24_P199484	NA	9068	ENSG00000116194	ANGPTL1	angiopoietin-

					like 1

Transcription Target: hsa_TATAAA_V$TATA_01 DB_ID:2456
C=1276;O=6;E=0.98;R=6.15;rawP=0.0004;adjP=0.0107

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_32_P69368	NA	3398	ENSG00000115738	ID2	inhibitor of DNA binding 2, dominant negative helix-loop-helix protein
A_23_P151294	NA	3458	ENSG00000111537	IFNG	interferon, gamma
A_23_P252306	NA	3397	ENSG00000125968	ID1	inhibitor of DNA binding 1, dominant negative helix-loop-helix protein
A_23_P54100	NA	2100	ENSG00000140009	ESR2	estrogen receptor 2 (ER beta)
A_23_P87879	NA	969	ENSG00000110848	CD69	CD69 molecule
A_24_P199484	NA	9068	ENSG00000116194	ANGPTL1	angiopoietin-like 1

Transcription Target: hsa_V$HFH4_01 DB_ID:2237
C=195;O=3;E=0.15;R=20.11;rawP=0.0004;adjP=0.0107

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_32_P69368	NA	3398	ENSG00000115738	ID2	inhibitor of DNA binding 2, dominant negative helix-loop-helix protein
A_23_P252306	NA	3397	ENSG00000125968	ID1	inhibitor of DNA binding 1, dominant negative helix-loop-helix protein
A_24_P199484	NA	9068	ENSG00000116194	ANGPTL1	angiopoietin-like 1

Transcription Target: hsa_GTCNYYATGR_UNKNOWN DB_ID:2335
C=108;O=2;E=0.08;R=24.20;rawP=0.0031;adjP=0.0248

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P252306	NA	3397	ENSG00000125968	ID1	inhibitor of DNA binding 1, dominant negative helix-loop-helix protein
A_23_P328074	NA	6299	ENSG00000103449	SALL1	sal-like 1 (Drosophila)

Transcription Target: hsa_CAGGTA_V$AREB6_01 DB_ID:2404
C=780;O=4;E=0.60;R=6.70;rawP=0.0029;adjP=0.0248

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P93787	NA	3082	ENSG00000019991	HGF	hepatocyte growth factor (hepapoietin A; scatter factor)
A_23_P54100	NA	2100	ENSG00000140009	ESR2	estrogen receptor 2 (ER beta)
A_32_P835566	NA	401024	ENSG00000188738	FSIP2	fibrous sheath interacting protein 2
A_24_P239266	NA	55004	ENSG00000149357	LAMTOR1	late endosomal/lysosomal adaptor, MAPK and MTOR activator 1

Transcription Target: hsa_AAAYRNCTG_UNKNOWN DB_ID:1846
C=368;O=3;E=0.28;R=10.65;rawP=0.0028;adjP=0.0248

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P93787	NA	3082	ENSG00000019991	HGF	hepatocyte growth factor (hepapoietin A; scatter factor)
A_23_P252306	NA	3397	ENSG00000125968	ID1	inhibitor of DNA binding 1, dominant negative helix-loop-helix protein
A_32_P835566	NA	401024	ENSG00000188738	FSIP2	fibrous sheath interacting protein 2

9.2.3 Tables showing results of KEGG pathway analysis

KEGG pathway: Rheumatoid arthritis
05323 C=91;O=3;E=0.07;R=43.08;rawP=4.74e-05;adjP=0.0002

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P151294	NA	3458	ENSG00000111537	IFNG	interferon, gamma
A_23_P72096	NA	3552	ENSG00000115008	IL1A	interleukin 1, alpha
A_23_P60306	NA	7099	ENSG00000136869	TLR4	toll-like receptor 4

KEGG pathway: Leishmaniasis
05140 C=72;O=3;E=0.06;R=54.45;rawP=2.35e-05;adjP=0.0002

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P151294	NA	3458	ENSG00000111537	IFNG	interferon, gamma
A_23_P72096	NA	3552	ENSG00000115008	IL1A	interleukin 1, alpha
A_23_P60306	NA	7099	ENSG00000136869	TLR4	toll-like receptor 4

KEGG pathway: TGF-beta signaling pathway
04350 C=84;O=3;E=0.06;R=46.67;rawP=3.73e-05;adjP=0.0002

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_32_P69368	NA	3398	ENSG00000115738	ID2	inhibitor of DNA binding 2, dominant negative helix-loop-helix protein
A_23_P151294	NA	3458	ENSG00000111537	IFNG	interferon, gamma
A_23_P252306	NA	3397	ENSG00000125968	ID1	inhibitor of DNA binding 1, dominant negative helix-loop-helix protein

KEGG pathway: Type I diabetes mellitus
04940 C=43;O=2;E=0.03;R=60.79;rawP=0.0005;adjP=0.0011

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P151294	NA	3458	ENSG00000111537	IFNG	interferon, gamma
A_23_P72096	NA	3552	ENSG00000115008	IL1A	interleukin 1, alpha

KEGG pathway: Graft-versus-host disease
05332 C=41;O=2;E=0.03;R=63.75;rawP=0.0005;adjP=0.0011

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P151294	NA	3458	ENSG00000111537	IFNG	interferon, gamma
A_23_P72096	NA	3552	ENSG00000115008	IL1A	interleukin 1, alpha

KEGG pathway: Melanoma
05218 C=71;O=2;E=0.05;R=36.81;rawP=0.0014;adjP=0.0028

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_24_P99244	NA	2252	ENSG00000140285	FGF7	fibroblast growth factor 7
A_23_P93787	NA	3082	ENSG00000019991	HGF	hepatocyte growth factor (hepapoietin A; scatter factor)

KEGG pathway: Chagas disease (American trypanosomiasis) 05142
C=104;O=2;E=0.08;R=25.13;rawP=0.0029;adjP=0.0048

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P151294	NA	3458	ENSG00000111537	IFNG	interferon, gamma
A_23_P60306	NA	7099	ENSG00000136869	TLR4	toll-like receptor 4

KEGG pathway: Amoebiasis 05146
C=106;O=2;E=0.08;R=24.66;rawP=0.0030;adjP=0.0048

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P151294	NA	3458	ENSG00000111537	IFNG	interferon, gamma
A_23_P60306	NA	7099	ENSG00000136869	TLR4	toll-like receptor 4

9.3 Analysis of group C genes

9.3.1 Tables showing results of GO analysis

Biological process: regulation of biological process
GO:0050789 C=8132;O=70;E=50.58;R=1.38;rawP=1.82e-05;adjP=0.0075

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_24_P14731	NA	27344	ENSG00000102109	PCSK1N	proprotein convertase
					subtilisin/kexin type 1 inhibitor
A_24_P113725	NA	6658	ENSG00000134595	SOX3	SRY (sex determining region Y)-box 3
A_24_P213884	NA	7148	ENSG00000168477	TNXB	tenascin XB
A_24_P216165	NA	1050	ENSG00000245848	CEBPA	CCAAT/enhancer binding protein (C/EBP), alpha
A_23_P54573	NA	161882	ENSG00000179588	ZFPM1	zinc finger protein, multitype 1
A_24_P44916	NA	148170	ENSG00000167617	CDC42EP5	CDC42 effector protein (Rho GTPase binding) 5
A_24_P279060	NA	10021	ENSG00000138622	HCN4	hyperpolarization activated cyclic nucleotide-gated potassium channel 4
A_24_P10674	NA	115703	ENSG00000004777	ARHGAP33	Rho GTPase activating protein 33
A_24_P878992	NA	10297	ENSG00000115266	APC2	adenomatosis polyposis coli 2
A_24_P203315	NA	254263	ENSG00000174871	CNIH2	cornichon homolog 2 (Drosophila)
A_24_P251040	NA	9048	ENSG00000117407	ARTN	artemin
A_24_P52882	NA	8739	ENSG00000135116	HRK	harakiri, BCL2 interacting protein (contains only BH3 domain)
A_23_P109133	NA	551	ENSG00000101200	AVP	arginine vasopressin
A_24_P269101	NA	4762	ENSG00000181965	NEUROG1	neurogenin 1
A_32_P74615	NA	389058	ENSG00000204335	SP5	Sp5 transcription factor
A_24_P34575	NA	5455	ENSG00000198914	POU3F3	Receptor 3
A_32_P35375	NA	10522	ENSG00000177030	DEAF1	deformed epidermal autoregulatory factor 1 (Drosophila)

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_24_P293089	NA	163033	ENSG00000218891	ZNF579	zinc finger protein 579
A_23_P342612	NA	610	ENSG00000099822	HCN2	hyperpolarization activated cyclic nucleotide-gated potassium channel 2
A_23_P140527	NA	27023	ENSG00000171956	FOXB1	forkhead box B1
A_24_P144465	NA	63940	ENSG00000213654	GPSM3	G-protein signaling
A_23_P107394	NA	84667	ENSG00000179111	HES7	hairy and enhancer of split 7 (Drosophila)
A_24_P207195	NA	79191	ENSG00000177508	IRX3	iroquois homeobox 3
A_24_P112803	NA	6656	ENSG00000182968	SOX1	SRY (sex determining region Y)-box 1
A_24_P208513	NA	7475	ENSG00000115596	WNT6	wingless-type MMTV integration site family, member 6
A_23_P20458	NA	2831	ENSG00000183729	NPBWR1	neuropeptides B/W receptor 1
A_23_P90189	NA	27113	ENSG00000105327	BBC3	BCL2 binding component 3
A_24_P57898	NA	128408	ENSG00000125533	BHLHE23	basic helix-loop-helix family, member e23
A_24_P923765	NA	399664	ENSG00000181588	MEX3D	mex-3 homolog D (C. elegans)
A_24_P405992	NA	11346	ENSG00000171992	SYNPO	synaptopodin
A_23_P50773	NA	23373	ENSG00000105662	CRTC1	CREB regulated transcription coactivator 1
A_23_P206466	NA	57524	ENSG00000167971	CASKIN1	CASK interacting protein 1
A_24_P227326	NA	283248	ENSG00000167771	RCOR2	REST corepressor 2
A_24_P328446	NA	3196	ENSG00000115297	TLX2	T-cell leukemia homeobox 2
A_24_P117782	NA	85508	ENSG00000215397	SCRT2	scratch homolog 2, zinc finger protein (Drosophila)
A_24_P129107	NA	85409	ENSG00000145506	NKD2	naked cuticle homolog 2 (Drosophila)
A_24_P215445	NA	401	ENSG00000165462	PHOX2A	paired-like homeobox 2a
A_23_P134433	NA	2020	ENSG00000164778	EN2	engrailed homeobox 2
					motif) receptor 3

A_24_P56689	NA	7755	ENSG00000122386	ZNF205	zinc finger protein 205
A_23_P150162	NA	1815	ENSG00000069696	DRD4	dopamine receptor D4
A_32_P226918	NA	5586	ENSG00000065243	PKN2	protein kinase N2
A_23_P150664	NA	387733	ENSG00000206013	IFITM5	interferon induced
					transmembrane protein 5
A_24_P173294	NA	83855	ENSG00000129911	KLF16	Kruppel-like factor
A_24_P266466	NA	2843	ENSG00000204882	GPR20	G protein-coupled receptor 20
A_23_P15135	NA	84891	ENSG00000130182	ZSCAN10	zinc finger and SCAN domain containing 10

9.3.2 Tables showing results of Transcription Factor analysis

Transcription Target: hsa_CAGGTG_V$E12_Q6 DB_ID:2409
C=2450;O=30;E=6.70;R=4.48;rawP=2.55e-12;adjP=8.64e-10

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_24_P416595	NA	284346	ENSG00000176472	ZNF575	zinc finger protein 575
A_24_P43959	NA	55691	ENSG00000151474	FRMD4A	FERM domain containing 4A
A_24_P227326	NA	283248	ENSG00000167771	RCOR2	REST corepressor 2
A_23_P34554	NA	777	ENSG00000198216	CACNA1E	calcium channel, voltage-dependent, R type, alpha 1E subunit
A_24_P117782	NA	85508	ENSG00000215397	SCRT2	scratch homolog 2, zinc finger protein (Drosophila)
A_24_P328446	NA	3196	ENSG00000115297	TLX2	T-cell leukemia homeobox 2
A_24_P113725	NA	6658	ENSG00000134595	SOX3	SRY (sex determining region Y)-box 3
A_23_P96072	NA	2902	ENSG00000176884	GRIN1	glutamate receptor, ionotropic, N-methyl D-aspartate 1
A_24_P215445	NA	401	ENSG00000165462	PHOX2A	paired-like homeobox 2a
A_24_P218074	NA	168544	ENSG00000181444	ZNF467	zinc finger protein 467
A_23_P140527	NA	27023	ENSG00000171956	FOXB1	forkhead box B1
A_24_P20807	NA	80758	ENSG00000131188	PRR7	proline rich 7 (synaptic)
A_24_P56689	NA	7755	ENSG00000122386	ZNF205	zinc finger protein 205

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P107394	NA	84667	ENSG00000179111	HES7	hairy and enhancer of split 7 (Drosophila)
A_24_P166434	NA	170680	ENSG00000204538	PSORS1C2	psoriasis susceptibility 1 candidate 2
A_32_P226918	NA	5586	ENSG00000065243	PKN2	protein kinase N2
A_23_P25790	NA	64403	ENSG00000139880	CDH24	cadherin 24, type 2
A_23_P18406	NA	94032	ENSG00000163888	CAMK2N2	calcium/calmodulin-dependent protein r
A_24_P251040	NA	9048	ENSG00000117407	ARTN	artemin
A_24_P208513	NA	7475	ENSG00000115596	WNT6	wingless-type MMTV integration site family, member 6
A_32_P11425	NA	55752	ENSG00000138758	SEPT11	septin 11
A_24_P376451	NA	2668	ENSG00000168621	GDNF	glial cell derived neurotrophic factor
A_24_P61537	NA	1152	ENSG00000166165	CKB	creatine kinase, brain
A_23_P304530	NA	254910	ENSG00000186207	LCE5A	late cornified envelope 5A
A_23_P86493	NA	10660	ENSG00000138136	LBX1	ladybird homeobox 1
A_24_P46946	NA	64321	ENSG00000164736	SOX17	SRY (sex determining region Y)-box 17
A_24_P34575	NA	5455	ENSG00000198914	POU3F3	POU class 3 homeobox 3
A_23_P206466	NA	57524	ENSG00000167971	CASKIN1	CASK interacting protein 1
A_23_P50773	NA	23373	ENSG00000105662	CRTC1	CREB regulated transcription coactivator 1
A_24_P333364	NA	116988	ENSG00000133612	AGAP3	ArfGAP with GTPase domain, ankyrin repeat and PH domain 3

Transcription Target: hsa_GGGCGGR_V$SP1_Q6 DB_ID:2452
C=2891;O=27;E=7.91;R=3.41;rawP=1.43e-08;adjP=9.70e-07

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_24_P14731	NA	27344	ENSG00000102109	PCSK1N	proprotein convertase subtilisin/kexin type 1 inhibitor
A_24_P416595	NA	284346	ENSG00000176472	ZNF575	zinc finger protein 575
A_24_P117782	NA	85508	ENSG00000215397	SCRT2	scratch homolog 2, zinc finger protein (Drosophila)

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_24_P328446	NA	3196	ENSG00000115297	TLX2	T-cell leukemia homeobox 2
A_24_P215445	NA	401	ENSG00000165462	PHOX2A	paired-like homeobox 2a
A_24_P216165	NA	1050	ENSG00000245848	CEBPA	CCAAT/enhancer binding protein (C/EBP), alpha
A_24_P218074	NA	168544	ENSG00000181444	ZNF467	zinc finger protein 467
A_23_P140527	NA	27023	ENSG00000171956	FOXB1	forkhead box B1
A_24_P20807	NA	80758	ENSG00000131188	PRR7	proline rich 7 (synaptic)

9.3.3 Tables showing results of KEGG pathway analysis

KEGG pathway: Neuroactive ligand-receptor interaction 04080
C=272;O=6;E=0.74;R=8.06;rawP=0.0001;adjP=0.0014

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P153549	NA	2906	ENSG00000105464	GRIN2D	glutamate receptor, ionotropic, N-methyl D-aspartate 2D
A_23_P146885	NA	2837	ENSG00000181408	UTS2R	urotensin 2 receptor
A_23_P150162	NA	1815	ENSG00000069696	DRD4	dopamine receptor D4
A_23_P20458	NA	2831	ENSG00000183729	NPBWR1	neuropeptides B/W receptor 1
A_23_P101761	NA	84634	ENSG00000116014	KISS1R	KISS1 receptor
A_23_P96072	NA	2902	ENSG00000176884	GRIN1	glutamate receptor, ionotropic, N-methyl D-aspartate 1

KEGG pathway: Wnt signaling pathway 04310
C=150;O=4;E=0.41;R=9.75;rawP=0.0008;adjP=0.0056

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_24_P208513	NA	7475	ENSG00000115596	WNT6	wingless-type MMTV integration site family, member 6
A_24_P46946	NA	64321	ENSG00000164736	SOX17	SRY (sex

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
					determining region Y)-box 17
A_24_P878992	NA	10297	ENSG00000115266	APC2	adenomatosis polyposis coli 2
A_24_P129107	NA	85409	ENSG00000145506	NKD2	naked cuticle homolog 2 (Drosophila)

KEGG pathway: Calcium signaling pathway 04020
C=177;O=4;E=0.48;R=8.26;rawP=0.0015;adjP=0.0070

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P153549	NA	2906	ENSG00000105464	GRIN2D	glutamate receptor, ionotropic, N-methyl D-aspartate 2D
A_24_P399871	NA	8911	ENSG00000100346	CACNA1I	calcium channel, voltage-dependent, T type, alpha 1I subunit
A_23_P34554	NA	777	ENSG00000198216	CACNA1E	calcium channel, voltage-dependent, R type, alpha 1E subunit
A_23_P96072	NA	2902	ENSG00000176884	GRIN1	glutamate receptor, ionotropic, N-methyl D-aspartate 1

KEGG pathway: Basal cell carcinoma 05217
C=55;O=2;E=0.15;R=13.29;rawP=0.0100;adjP=0.0280

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_24_P208513	NA	7475	ENSG00000115596	WNT6	wingless-type MMTV integration site family, member 6
A_24_P878992	NA	10297	ENSG00000115266	APC2	adenomatosis polyposis coli 2

KEGG pathway: Amyotrophic lateral sclerosis (ALS) 05014
C=53;O=2;E=0.15;R=13.79;rawP=0.0093;adjP=0.0280

User Id	Value	Entrez ID	Ensembl Gene Stable ID	Gene Symbol	Description
A_23_P153549	NA	2906	ENSG00000105464	GRIN2D	glutamate receptor,
					ionotropic, N-methyl D-aspartate 2D
A_23_P96072	NA	2902	ENSG00000176884	GRIN1	glutamate 1 receptor, ionotropic, N-methyl D-aspartate