Spring 5-26-2020

The crosstalk between COBRA1 and Wnt/Î²-catenin signaling in cervical cancer

Omnia Mahmoud Abdelraheem
The American University in Cairo

The Crosstalk between COBRA1 and Wnt/β-catenin signaling in Cervical Cancer

By

Omnia Mahmoud AbdelRaheem

A Thesis Submitted to the Biotechnology Master's Program

In partial fulfilment of the requirements for the Degree of Master of Science

Under the supervision of

Dr. Asma Amleh

Associate Professor, Department of Biology

School of Sciences and Engineering (SSE)

The American University in Cairo

May 2020

THE AMERICAN UNIVERSITY IN CAIRO
100 YEARS | SCHOOL OF SCIENCES AND ENGINEERING

The Crosstalk between COBRA1 and Wnt/β-catenin signaling in Cervical Cancer

A Thesis Submitted by

Omnia Mahmoud AbdelRaheem

To the Biotechnology Master's Program

May 2020

In partial fulfilment of the requirements for

The degree of Master of Science in Biotechnology

Has been approved by

Thesis Committee Supervisor/Chair _____

Affiliation _____

Thesis Committee Reader/Examiner _____

Affiliation _____

Thesis Committee Reader/Examiner _____

Affiliation _____

Thesis Committee Reader/External Examiner

_____

Affiliation _____

_____ _____ _____ _____

Dept. Chair/Director Date Dean Date

Dedication

To my ***Mum***, my back bone, for teaching me how to cope with changes, be resilient and pushing me out of my comfort zone. To my ***Father***, for his love and support. To my lovely and only brother, ***Ahmed***, you inspired me with your loyalty and sincerity. Without you, this thesis would have been completed two years later. To ***Yazan*** and ***Yunus***, my adorable nephews, for your existence and the joy you brought to my life. To ***Alaa***, there is no sister in law better than you.

Acknowledgments

حَسْبُنَا اللَّهُ سَيُؤْتِينَا اللَّهُ مِن فَضْلِهِ وَرَسُولُهُ إِنَّا إِلَى اللَّهِ رَاغِبُونَ

All praise due to Allah for giving me what I need not what I want

I would like to direct my appreciation and gratitude to my supervisor ***Dr. Asma Amleh***, I will always be indebted to you. Thank you for believing in me and for your advice to do what I am really passionate about. Your words made me aligned with my life path purpose, helped me to be energized and on point. I am thankful for the time and support you spent on me throughout the program. Your dedication and loyalty to your mission inspired me and showed me what I can do. The confidence I have now will never go away. I am honored to have been part of your research team.

I sincerely thank Dr. Amleh's research team members, who have been a great support during this journey. I am deeply grateful to **Myret Ghabriel,** for all the technical advices, fruitful debates and brain storming. To **Menna Ghouraba,** for helping me to have hands on lab experience and her guidance during my first experiments. To **Rowan Bahaa,** for being the best companion during the rough times in the lab. To **Alaa Youssef,** for managing our team's financial matters. I would like to acknowledge **Noha Saad,** for her precise data and valuable notes that helped me a lot. I would also like to thank Eric, Khaled, Sheri, Nancy and Nayla for all their support. I feel lucky to work with this incredible team.

I would like to Thank ***Dr. Ahmed Mustafa,*** for his endless backing and support to the Biotechnology program students. Also, I would like to express my sincere gratitude to all professors at the AUC who has taught me a lot during my time in the program. I genuinely thank *Amgad Ouf* for the doing whatever it takes to assure that our research work goes smoothly.

Finally, I would like to thank **AUC** for funding me with research grant and fellowships.

Abstract

The majority of cancer phenotypes, therapeutic resistance, and clinical prognosis are correlated to dysregulated transcriptional programs within cancer cells. Cofactor of BRCA1 (COBRA1), named also as NELF-B, is one of the principal components of the Negative Elongation Factor (NELF) complex. NELF complex negatively regulates the elongation of transcripts by halting RNA polymerase II at the proximal promoter. It also regulates the transcription of its target genes via interacting with other transcription factors like BRCA1 and AP-1.

A previous study by our team showed that the silencing of COBRA1 led to a noted decrease in the steady-state mRNA levels of *β-catenin* within HeLa cells. This decline indicates an effect for COBRA1 on the Wnt/β-catenin signaling pathway. In our study, quantitative real-time PCR showed that the knockdown of COBRA1 promoted *FOXF2* mRNA levels. *FOXF2* is a gene known to inhibit HeLa cells proliferation, migration, and invasion *in vitro* and growth *in vivo* by regulating the Wnt signaling pathway. Also, our study revealed that the upregulation of *FOXF2* in COBRA1 knocked down HeLa cells was accompanied by a suppressed expression of Wnt signaling pathway target genes *(β-catenin, C-MYC, and CCND1*). The constitutive firing of the Wnt signaling pathway is considered as a second trigger to cervical cancer development after chronic HPV infection.

So, COBRA1 silencing and *FOXF2* upregulation are accompanied by the downregulation of *β-catenin* and its downstream target genes. Based on these findings, we speculated that FOXF2 might be the modulator of COBRA1 effects on the β-catenin expression. By performing ChIP assay, the interaction between COBRA1 and the promoter of *FOXF2* was proven. Due to the lack of a DNA binding domain in COBRA1, this interaction should be mediated through complex formation. Data obtained from TRANSFAC® Database showed that COBRA1 regulates *FOXF2* expression via interacting with ERα, which has a predicted binding site in its promoter. ChIP assay results also indicated that COBRA1 within the NELF complex decreases *FOXF2* transcription by pausing the RNAPII at its promoter-proximal region. Altogether, this study could help as a start in identifying and describing the tumorigenic role of COBRA1 mediated by FOXF2 in cervical cancer. Understanding the role of COBRA1 as a transcription factor in cervical cancer opens new insights that have the potential to improve patient care.

Contents

List of Abbreviations

bps	base pairs
BRCA1	Breast cancer type 1 susceptibility protein
CC	Cervical cancer
CCND1	Cyclin D1
ChIP	Chromatin Immunoprecipitation
COBRA1	Cofactor of BRCA1
DMEM	Dulbecco's modified Eagle's medium
DRB	5,6-Dichloro-1-β-D-Dibofuranosyl Benz imidazole
DSIF	DRB sensitivity-inducing factor
EMT	epithelial mesenchyme transition
ER	Estrogen receptor
ER-α	Estrogen receptor-alpha
E-Value	Primer Efficiency
F	Forward
FBS	Fetal bovine serum
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
IARC	International Agency for Research on Cancer
ICO	Catalan Institute of Oncology
IP	Immunoprecipitation
Mins	Minutes
mL	Milliliters
NELF	Negative elongation factor
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
qPCR	Quantitative Real Time Polymerase Chain Reaction
R	Reverse
RNAi	RNA interference
RNAPII	RNA polymerase II
RT	Radiotherapy
RT-PCR	Reverse Transcription Polymerase Chain Reaction
RT-PCR	Reverse transcription polymerase chain reaction
SD	Standard deviation
siCOBRA1	COBRA1 siRNA
siNTC	Non Targeting siRNA (Negative Control siRNA)
siRNA	Short interfering RNA
TFBSs	Transcription Factors Binding Sites
TFs	Transcription Factors
TSS	Transcription Start Site
UGCs	Upper gastrointestinal cancers
WHO	World Health Organization
µls	microliters

List of Tables

List of Figures

Introduction

Cervical cancer

Cervical cancer is one of the most predominant cancers that influence the female population. Worldwide, it positions third in the frequency and half million cases are newly diagnosed each year. This number of cases is equivalent to 6.6% of all female cancers. Concerning mortality, it ranks fourth with 265,700 deaths each year. Also, it is rated as the second leading cause of cancer-related deaths amongst women in developing countries.[1], [2] According to the information available from The Information Centre on HPV and Cancer (ICO/IARC), cervical cancer is the 13th most abundant cancer between women in Egypt and ranks as the 10th most prevalent cancer amongst women between the ages of 15 and 44 years.[3] 13% of the cervical cancer diagnosis is made at stages that are unlikely to be cured. After metastasis, the 5-year survival decline to reach 16.5% compared to 91.5% for localized cervical cancer. The standard treatment for cervical cancer in its early stages includes surgery, chemotherapy, and radiotherapy (RT). Nonetheless, metastatic cervical cancer has no standard treatment due to the heterogeneous manifestation. The most common etiology for CC is the persistent Human Papillomavirus (HPV) infection, which is responsible for 99% of all cervical tumors.[4] HPV infection is usually eliminated by the immune system. However, smoking, estrogen, and immunocompromised responses are cofactors that halt the HPV elimination and development of chronic infection.[5] WHO classifies CC into three categories, squamous (70-80%), adenocarcinoma (10-15%), and other epithelial tumors.[6] For better management of CC, identifying its molecular mechanisms is necessary. The fact that not all women infected with HPV will develop CC has shed the light that other factors in addition to HPV infection are involved. For instance, HPV DNA integration into the host chromosomal genome is a crucial step that is present in all invasive CC. This integration leads to overexpression of viral

oncoproteins. This, in turn, promotes genomic instability, accumulation of secondary mutations, and malignant transformation of the host cells. Improved understanding and increased knowledge of the molecular biology events associated with CC progression will help to better understand and manage the disease.[5]

Transcription regulation

Despite having the same DNA in almost all the cells of the human body, the genes expression differs from cell type to another. Gene regulation allows each cell type to have specific sets of proteins to perform its specialized function. The regulation of gene expression can be achieved at different stages. Regulation can take place at the epigenetic level, transcriptional and post-transcriptional level or translational and post-translational level.[7]

For the transcriptional regulation, each gene has protein-binding regulatory DNA sequences that are found both close to and many kilo-bases away from Transcription Start Sites (TSS). Also, there are specific proteins that bind to a gene's transcription- control regions and determine where transcription will start and either activate or repress transcription. These specific proteins are called transcription factors (TFs) and they function as molecular switches. Approximately the humans have about 2,000 different proteins that act as these molecular switches, some as activators, some as repressors.[7] Transcription of a single gene can be controlled by the binding of numerous transcription factors to different alternative regulatory sites either upstream (opposite to the direction of transcription) or downstream (in the same direction as transcription) from the TSS. This binding directs the various expression patterns of the same gene in different cell types.[8]

The majority of TFs are expressed in any cell type. However, a small number of TFs are called lineage regulators or master TFs that are responsible for controlling the transcriptional program that regulates cell status in normal differentiated conditions. These master TFs are responsible for the management of gene expression programs specific to each cell type. Cell identity also depends

on tissue environment, DNA methylation, and histone modifiers. Altogether assure the chromatin is appropriate for positive and negative regulation. [9]

One of the main hallmarks of cancer is the dysregulation of gene expression. The majority of cancer phenotypes, therapeutic resistance, and clinical prognosis are correlated to dysregulated transcriptional programs within cancer cells. One of the recently identified concepts is transcriptional addiction. Transcriptional addiction is the behavior of the cancer cells showing dependence on oncogenic TF. This addiction is essential for the continuous proliferation of cancer cells. Understanding the role transcription plays in cancer pathogenesis is considered one of the approaches to recognize cancer and these new insights have the potential to improve patient care.[9]

Cofactor of BRCA1 (COBRA1)

Cofactor of BRCA1 is located on chromosome 9 and the protein consists of 580 amino acids. It has been first identified as a BRCA1 interacting protein. Results suggested that the recruitment of COBRA1 to the BRCT1 domain of BRCA1 is the mediator of *BRCA1*-dependent unfolding of higher levels of chromatin structure [10]. Later in 2003, COBRA1 and NELF-B, one of the principal components of the Negative Elongation Factor (NELF) complex, were found to be the same.[11]. NELF complex together with DRB- sensitivity inducing factor (DSIF) has been associated with a process known as promoter-proximal pausing. RNA- Pol II stalling is common at genes involved in development and response to stimulants, suggesting that Pol II halting during early elongation phase has critical roles in rapid and precise control of gene expression. Along with the lack of DNA binding domain, COBRA1 was found to regulate the transcription of its target genes via interacting with other transcription factors including BRCA1, nuclear receptors and AP-1 complex [12]–[14]. This network of interactions suggests COBRA1 being involved in the regulation of multiple cellular processes as proliferation, cell survival, and tumorigenesis.

The biological role of COBRA1 in cancer pathogenesis is not fully understood but there is a growing body of evidence which suggests that COBRA1 plays a role in the malignant transformation, proliferation, and invasion of cancer cells. Previous findings are suggesting a cancer-type dependent role for COBRA1, with different expression patterns in different cancers. It has been identified as a novel oncogene in UGCs with high expression levels of *COBRA1* mRNA and protein observed in tumor samples. Several studies have previously identified COBRA1 to act as a tumor suppressor in breast cancer. The silencing of COBRA1 expression in HepG2 cell line using RNA interference caused a significant decrease in the cellular proliferation and migratory ability of HepG2 cell. This was associated with a significant decrease in the mRNA expression levels of the proliferation marker *Ki-67* and the proto-oncogene *Survivin.*[15] The protein and mRNA expression of COBRA1 across four different cell lines representing different grades of HCC were investigated. Results revealed a gradual decrease in expression of COBRA1 with increased HCC aggressiveness. The highest expression was observed in the low-grade HepG2 cell line and the lowest expression in the high-grade SNU-387.[15]

A previous study by Dr. Amleh's research team investigated the role of COBRA1 in cervical cancer. To address this purpose, an *in-silico* analysis was done to identify the patterns of COBRA1 expression in cervical carcinoma relative to normal tissues from cervix. Publically available microarray databases were used to do this analysis. The four chosen datasets were all in cervical cancer and their data type was mRNA microarray. In two of these studies, Scotto Cervix and Pyeon multi-cancer, *COBRA1* mRNA was found to be significantly overexpressed in cervical cancer tissues relative to normal with a fold change of 1.245 (p= 0.003) and 1.76 (p=2.81E-4) respectively. As for Zhai cervix study, COBRA1 was found to be slightly upregulated with a fold change of 1.066 but this was found to be statistically insignificant (p=0.121). In contrast to the first

three studies, in the Biewenga Cervix study COBRA1 was found to be downregulated with -1.1441-fold change but again this was found to be insignificant with a very high p-value of 0.999.

Based on the finding of an upregulated expression pattern for COBRA1 in CC observed from the in silico studies. The effect of COBRA1 gene silencing via siRNA-mediated RNA interference (RNAi) in HeLa cells was analyzed. There were three controls for this transfection procedure. First, (untreated control) which was cells left un-transfected. Second, (Mock) was cells treated with the transfection agent only (lipofectamine). Third, the negative control (siNTC), which was cells transfected with scramble RNA that has no homology to any mammalian gene to control for any nonspecific effects that might happen because of the transfection procedure. Semi-quantitative RT-PCR showed a significant decrease in the mRNA levels of *β-catenin* in *COBRA1* siRNA treated cells relative to the negative control. This inicates an effect for COBRA1 depletion on the signaling pathway of Wnt/β-catenin.[5]

FOXF2

FOXF2 is an important member of the FOX family, which regulated the promoter's activity of its downstream genes, thereby regulating the biological processes of cells. FOXF2 was found to be associated with the development of multiple tumors. Kong et al demonstrated that FOXF2 was a new independent predictive factor of non-small cell lung cancer [16]. Its lower expression could lead to poor prognosis of patients, especially for patients with lung cancer. They also revealed in their article that the downregulation of *FOXF2* mRNA was a sign of early-onset metastasis and poor prognosis of patients with breast cancer [17]. Wang et al identified in their research that FOXF2 acted as a novel epithelial mesenchyme transition (EMT) suppressing transcription factor in basal-like breast cancer [18]. Their further research also showed that FOXF2 promoted basal-like breast cancer cells metastasis by upregulation of *TWIST1* as well as activating EMT. A recently published article by Zhang et al., 2018 was aimed to study the FOXF2 effects on cervical

cancer. It showed that Low FOXF2 expression was associated with poor outcomes of cervical cancer patients and that overexpression of FOXF2 inhibited HeLa cells proliferation, migration, and invasion *in vitro* and growth *in vivo*. Meanwhile, up-regulated FOXF2 stimulated E-cadherin expression and impaired Vimentin and Snail expression. Also, Zhang et al., 2018 observed that FOXF2 could suppress the expression level of *β-catenin* in the nucleus and its downstream target genes in the Wnt/β-catenin signaling pathway, such as c-Myc, CyclinDl, MMP9, and Lgr5. It is reported that c-Myc, CyclinDl, MMP9, and Lgr5 were involved in the development of tumors and their overexpression had significant promoting effects on the development of tumors. Taken together, FOXF2 inhibits the proliferation, migratory ability, and invasiveness of HeLa cells via modulating the Wnt signaling pathway.[19]

Hypothesis and objectives

As has been noted in the introduction section, both FOXF2 and COBRA1 genes have a role in cervical cancer via their effect on the Wnt/β-catenin signaling pathway.

We hypothesize that silencing COBRA1 subdues the Wnt/β-catenin signaling pathway by enhancing FOXF2 levels. This hypothesis is going to be tested by assessing the mRNA levels of *FOXF2* in silenced COBRA1 HeLa cells, examining the association between COBRA1 and FOXF2 promoter and assessing the down target genes of the Wnt/β-catenin signaling pathway. As far as we could possibly know, this is the first study that addresses the tumorigenic role of COBRA1 mediated by FOXF2 in cervical cancer. Understanding the role of COBRA1 as a transcription factor in cervical cancer opens new insights that will help for better management of the disease.

To address this hypothesis, our study had the following main specific aims:

- Assess the expression of *FOXF2* in HeLa cells with silenced COBRA1.
- q PCR to assess the expression of downstream target genes (at the mRNA level) of the Wnt/β-catenin signaling pathway in silenced COBRA1 HeLa cells
- Using the TRANSFAC® database to predict the Transcription Factors Binding sites (TFBSs) in the promoter of *FOXF2* gene.
- Chromatin Immunoprecipitation (ChIP) in HeLa cells. This assay is to capture a snapshot of specific interactions between COBRA1 within the NELF complex (protein) and *FOXF2* to determine whether the NELF complex is physically associated with *FOXF2* promoter (DNA).

Methods and Materials:

1. Cell lines and culture:

The human cervical adenocarcinoma cell line (HeLa) was a kind gift from ***Professor Marwan Emara,*** Director of Center for Aging and Associated Diseases, Zewail City of Science and Technology, Egypt. HeLa was grown in Dulbecco's modified Eagle's medium (DMEM) (GIBCO, USA) and supplied with 10% fetal bovine serum (GIBCO, USA), 100 U/ml penicillin and 100 mg/ml streptomycin (GIBCO, USA). Cells were maintained in a humidified incubator supplied with 5% CO2 at 37℃. In all the below-described experiments, cells in the logarithmic phase of growth from passage numbers 20-24 were used and propagated at 70-80% confluence according to the American type culture collection protocol. An inverted microscope (Olympus IX70, USA) was used to observe the cells. HeLa cells doubling time has been found to be approximately 25 hrs.

2. Viable Cell Count

Trypan Blue exclusion method was used to obtain the viable cell count prior to each experiment. Cells were harvested and re-suspended in fresh media by pipetting up and down until a homogenous cell suspension with no cell clumps was obtained. A mixture of Twenty microliters (µl) of the previous cell suspension and 20µl of 0.4% trypan blue in PBS was prepared. In each chamber of a hemocytometer (Hausser Scientific, USA) 10µl of the Trypan blue-cell suspension was loaded. The cells in each of the outer four squares in the two chambers were counted and the following equation was used to calculate the number of cells per 1 ml of cells suspension:

Number of cells/ml = (Total numbers of viable cells in all counted squares / total number of counted squares) x dilution factor x10000

3. Quantitative Real Time Polymerase Chain Reaction (qPCR):

Quantitative PCR was used to determine the differential gene expression at the messenger RNA level (mRNA) among tested conditions. Revert Aid First strand cDNA synthesis Kit (Thermo-Scientific, USA) was used to reverse transcribe (0.5µg) of total RNA in a final volume of 20µl following the instructions of the manufacturer's protocol.

Then PCR amplification reaction was performed using SYBR Green as a DNA-specific binding dye and continuous fluorescence monitoring. Each PCR reaction (10µl) consisted of 2x SYBR

Green I Master mix, 1µl of cDNA, and 0.5µl of each primer. All primers used for the amplification of the selected genes are listed in Table 3. The thermal cycler amplification program was set up with the following conditions: 50°C for 2 minutes, 95°C for 2 minutes, 95°C for 15 s, and 60°C for 1 minute. Forty cycles were included in this amplification reaction. The instrument was set to perform the default dissociation step under the following conditions, 95°C for 15 s, 60°C for 1 minute, and 95°C for 15 s. All experiments were performed three times using 7500 Real-Time PCR System (Applied Bio-systems, USA). Data analysis was done by the $2^{-\Delta\Delta Ct}$ method.

Table 1 qPCR primer sequences, PCR Conditions and amplicon sizes (F: forward primer, R: reverse primer, bp: base pair)

Gene		Primer Sequence	PCR Conditions	Amplicon size
β-catenin	**F**	GAGGAGCAGCTTCAGTCCCC	60°C annealing temperature, 40 cycles	139
	R	GCCATTGTCCACGCTGGATT		
FOXF2	**F**	5'-AAT GCC ACT CGC CCT ACA C-3'		199bp
	R	5'-GGC AGT CCC ACT GAG AGG TC-3'		
GAPDH	**F**	5'-AAG GTC ATC CCT GAG CTG AAC-3'		142bp
	R	5'-ACG CCT GCT TCA CCA CCT TCT-3'		
c-MYC	**F**	5'- AGC GAC TCT GAG GAG GAA C-3'		130bp
	R	5'- TGT GAG GAG GTT TGC TGT G – 3'		

CCND1	**F**	5’-TATTGCGCTGCTACCGTTGA-3′		90bp
	R	5’-CCAATAGCAGCAAACAATGTGAAA-3′		

4. FOXF2 promoter characterization:

TRANSFAC® in the geneXplain platform was used to better understand the specific characters of the *FOXF2* promoter. Thanks to Dr. Liling Huang, Sales and Marketing Manager at geneXplain GmbH, a free trial was offered to complete this study. TRANSFAC® release 2020.1 is a database containing published data on eukaryotic transcription factors and miRNAs, their experimentally-proven binding sites, and regulated genes. It can be considered as an encyclopedia of transcriptional regulation. Additionally, TRANSFAC® employs the Match algorithm which is able to predict identify potential transcription factor binding sites (TFBSs).

5. Chromatin Immunoprecipitation (ChIP) Assay:

Chromatin Immunoprecipitation (ChIP) is a powerful tool that has opened the secrets of chromatin and enhanced our understanding of the science behind it. The ChIP technique helps elucidate gene function and regulation in their native state. The scientific basis of this technique is selective enrichment of chromatin chunk containing a specific protein (antigen). Using an antibody that identifies this protein of interest, regions of the genome associated with this protein *in vivo* can be determined. So, the ChIP technique can provide insight into gene regulatory networks. ChIP methodology involves several steps mentioned below in detail.

I. *In vivo* Crosslinking and Lysis

37% formaldehyde-freshly prepared- was used to cross-link HeLa cells at ~ 80-90% confluence (10 million cells) in a 150mm culture dish containing DMEM (GIBCO, USA). The final concentration of formaldehyde was 1%. 10X Glycine was added to quench unreacted

formaldehyde. Cell Lysis Buffer and nuclear lysis buffer were added in the presence of protease inhibitors to ensure that the DNA is released.

II. Sonication to Shear DNA:

The optimum conditions to have cross-linked DNA within the length range of ~200-1000bp were 4 pulses for 15 seconds each at 50% power output of Branson Sonifier 150 (Emerson USA) and leave on ice for 60 seconds between each pulse. Agarose gel analysis of un-sheared DNA and sheared DNA was run to ensure the size of the DNA fragments. (1%) of the sheared chromatin were removed as the "Input" Sample and saved at 4°C until the Elution Process.

III. Immunoprecipitation (IP) of Cross-linked Protein/DNA

The immune-precipitating antibody and fully suspended protein G magnetic beads (Millipore USA, Cat. # CS200638) were mixed with the sheared chromatin. For the positive control, 1µg of Anti-RNA Polymerase II, clone CTD4H8 (Millipore USA, Cat. # 05-623B) was added per reaction. For the negative control, 1µg of Normal IgG (Millipore USA, Cat. # 12-371B) was added per reaction for assessing non-specific binding. For the sample, 10µg Anti-NELF-B Polyclonal antibody (Abcam, ab237027) was used. The appropriate amount of antibody needs was determined by adding different amounts from (1µg-10µg). 10µg was chosen as it gives an intense band in the agarose gel analysis. These samples were incubated overnight at 4°C with rotation. (BenchBlotter™-USA)

Protein G magnetic beads were pelleted with the magnetic separator (Invitrogen-USA) and the supernatant was removed completely. The Protein G beads-antibody/chromatin complex was re-suspended in each of the following cold buffers; Low Salt Immune Complex Wash Buffer (Millipore USA, Cat. # 20-154), High Salt Immune Complex Wash Buffer (Millipore USA, Cat. # 20-155), LiCl Immune Complex Wash Buffer (Millipore USA, Cat. # 20-156) and TE Buffer

(Millipore USA, Cat. # 20-157) and incubated for 3-5 minutes on a rotating platform followed by magnetic clearance and careful removal of the supernatant fraction.

IV. Elution of Protein/DNA Complexes and Reverse Crosslinking of Protein/DNA Complexes to Free DNA:

Proteinase K (Millipore USA, Cat. # 20-298) and the ChIP Elution Buffer (Millipore USA, Cat. # 200629) were added to all IP tubes as well as all Input tubes and incubate at 62°C for 2 hours with shaking. Proteinase K was deactivated by incubation at 95°C for 10 minutes and the samples were left to cool down at room temperature. Beads were separated using the magnetic separator. The supernatants were removed to new tubes. DNA Purification was done using Spin Columns (Millipore USA, Cat. # 20-290)

V. DNA concentration measurement

2µl of purified DNA samples including input, IP (NELF-B), and IP (RNA-Pol II) were used to measure the DNA concentration and purity using elution buffer as blank. Lvis Plate in SPECTROstar Nano (BMG LABTECH-USA) was used to measure the amounts of UV irradiation absorbed at 260 nm wavelength. The optical density (OD) readings at 260 nm and 280 nm were measured to assess (A260/A280) ratio to check the purity of the DNA preparation.

6. ChIP RT-PCR

After purification, DNA samples to undergo PCR amplification were positive and negative control antibody immunoprecipitations, input, samples, and a “no DNA” tube as a control for DNA contamination. MyTaq Red DNA Polymerase (Bioline, UK) was used to perform the PCR reactions using 1ul DNA per reaction in case of positive control and 4µL DNA in the case of NELF-B (COBRA1) immunoprecipitations samples, 0.5µl of forward and reverse primer were also included. The positive control Primers were specific for the human *GAPDH* gene. Primers

targeting FOXF2 promoter regions are listed in Table.2.PCR conditions used were the same among the regions. Amplifications conditions were programmed for 3 minutes at 94°C, then 35 cycles each cycle consists of (30 seconds denaturation set at 94ºC, followed by 30 seconds annealing adjusted at 60°C, and the cycle is ended with 45 seconds extension set at 72ºC). The process is then finalized with 10 minutes at 72ºC.The PCR products were then run on a 1.5 – 2 % agarose gel, separated depending on product size, and visualized by the help of the Gel Doc EZ System (Bio-Rad, USA). PCR amplicon sizes are listed in Table.4.

7. ChIP-qPCR

I. Primer design, efficiency assessment and specificity for ChIP-qPCR assay:

For the *FOXF2* gene, the primers (Eurofens Genomics, USA) were designed to target two specific regions in the promoter region. (Fig.1) Primer Blast [20] was used to design the primers and the *In-Silico* PCR function in the UCSC Genome Browser was used to test the specificity of them. The program outputs an amplicon table with only one single amplicon in the list. (Fig.2) This ensured the specificity of the designed primer. Primer Efficiencies (E-Values) were determined by diluting the clean input chromatin into a two-fold dilution series of four samples, then qPCR reactions were performed on these samples. The primer used for the targeting of the *FOXF2* gene promoter is listed in Table 2.

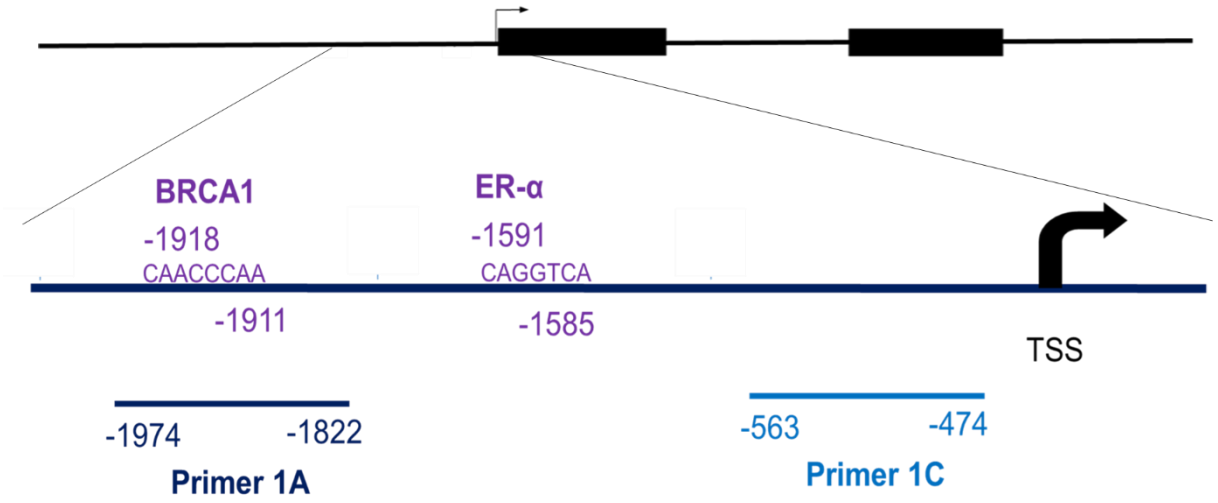

Figure 1 *FOXF2* genomic region illustration.

The gene is drawn to scale. The whole sequence is 12764bp. Black boxes represent exons. Exon 1 is 1543bp while exon 2 is 908bp. The bent arrow indicates the transcription start site (TSS). The predicted TFBS of BRCA1 and ER-α are illustrated with their sequences and relative positions. The horizontal lines represent the amplified regions using primer pairs 1A and 1C with their relative position to TSS.

Table 2 ChIP-qPCR primer sequences, amplicon sizes and primers efficiencies (E-values)

Primer Name	Sequence (5'->3')	Relative position	Amplicon size	E-value
Region 1A	F:CGCTCAGCAGTCAGTCAGAA	-1822	152	2.01
	R:AAAACACCGCCAAACCGAAC			
Region 1C	F:ACGTTACCTTTGGGCGTCTG	-474	89	2.029
	R:AGGGCCGATACTGACTCCAT			

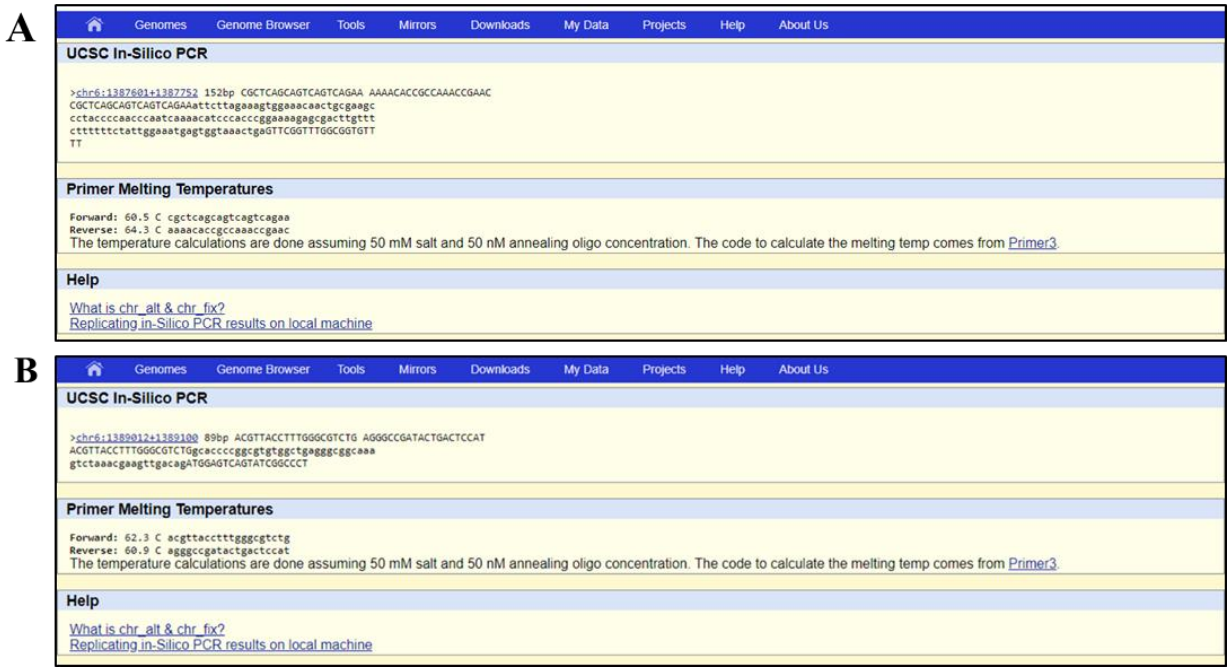

Figure 2 Primer Specificity:

The figure shows two screen shots from UCSC Genome Browse. Using the PCR in-silico tool gave only one hit on the list for each primer. (A) is the result for the primer 1A and (B) is the result for the primer (1C)

II. qPCR conditions and data analysis:

Then PCR amplification reaction was performed under the following conditions: 50°C for 2 minutes, 95°C for 2 minutes, 95°C for 15 s, and 60°C for 1 minute. Forty cycles were included in these amplification reactions, and 4µl of cDNA, 0.5µl of forward primer, 0.5µl of reverse primer were also added. The instrument was set to perform the default dissociation step under the following conditions, 95°C for 15 s, 60°C for 1 minute, and 95°C for 15 s. All experiments were performed three times using 7500 Real-Time PCR System (Applied Bio-systems, USA). All primers used for the amplification of the selected regions are listed in Table 2. Normalization of the cycle threshold (C_T) values from the qPCR assay was done by two methods, percent input method and fold enrichment method.

- **Percent Input Method:**

It normalizes (C_T) values according to the amount of chromatin input. In this method, the ChIP signals are divided by the input and the normalization is for both background levels and input.

First, the difference in C_T (ΔC_T) which equals (C_T input – C_T test samples) was calculated. Second, the fold difference was calculated using the formula: $f = E^{\Delta Ct}$, where E is the *FOXF2* primer efficiency value (E= 2.01 for primer 1A and E=2.03 for primer 1C). Finally, the % input was obtained by this equation (f * 1%) (1% is the percentage of the chromatin used in the input sample.)

- **Fold enrichment method:**

This method normalizes (C_T) values according to the background level but are not normalized to the input. This method of normalization does not require an input sample. First, the ΔC_T value between the test samples and the negative antibody sample (Normal IgG) was calculated. Then the fold enrichment in a linear range was obtained by the application of the $2^{-\Delta Ct}$ method.

8. Statistical Analysis of Data

Statistical analyses and graphical representations were performed using Graph Pad Prism 5.0 (Graph Pad, San Diego California USA, http://www.graphpad.com/). For comparisons made between two different groups, statistical significance was assessed using an unpaired student's t-test (two-tailed). P–value less than 0.05 was considered to be statistically significant (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

Results

1. The Effect of COBRA1 Silencing on the expression of *FOXF2* gene:

The mRNA steady state levels of *FOXF2* was examined post COBRA1 siRNA transfection using qPCR. As shown in (Fig.3), there was a significant increase by an average of 53.4% in the mRNA levels of *FOXF2* in knockdown COBRA1 cells relative to the negative control (siNTC) with p-value < 0.01. As mentioned before, the negative control (siNTC) in the transfection procedure was HeLa cells transfected with scramble RNA that has no homology to any mammalian gene.

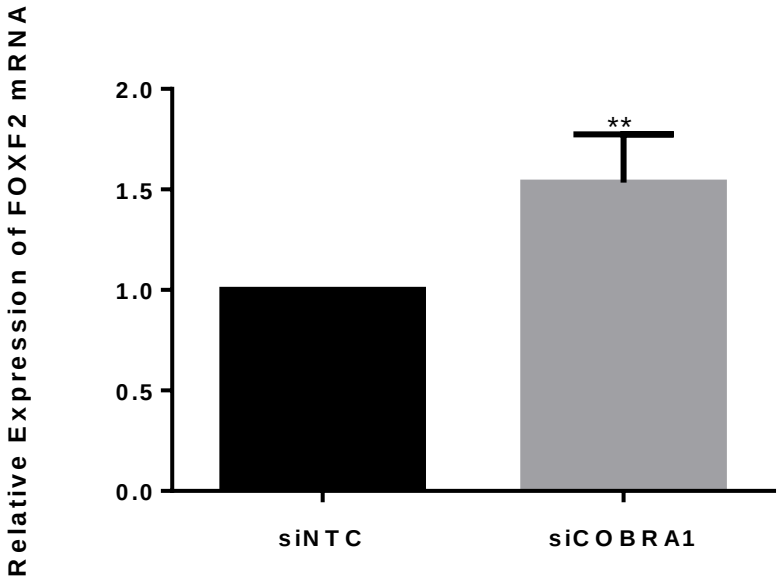

Figure 3 Upregulation of *FOXF2* in COBRA1 Knockdown HeLa cells

mRNA expression was analyzed by qPCR. C_T were obtained and *FOXF2* expressions were normalized to the internal control *GAPDH*. The relative expression is indicated as fold change to the negative control (siNTC). Data represent the mean ± SD of 2 independent experiments (n = 6). The statistical significance was calculated using two-tailed unpaired Student t-test and **p < 0.01. siNTC: Negative siRNA and siCOBRA1: COBRA1 Knockdown.

2. COBRA1 Silencing effect on some of the target genes in the Wnt/β-catenin signaling pathway:

I. *β-catenin*:

In our team's previous study, the mRNA steady state levels of *β-catenin* were examined post COBRA1 silencing using semi quantitative RT-PCR. In this study, qPCR was carried out to confirm the previous results. The mRNA level of the *β-catenin* gene following COBRA1 siRNA transfection, as shown in (Fig.4), was significantly decreased relative to the negative control with p-value <0.001.

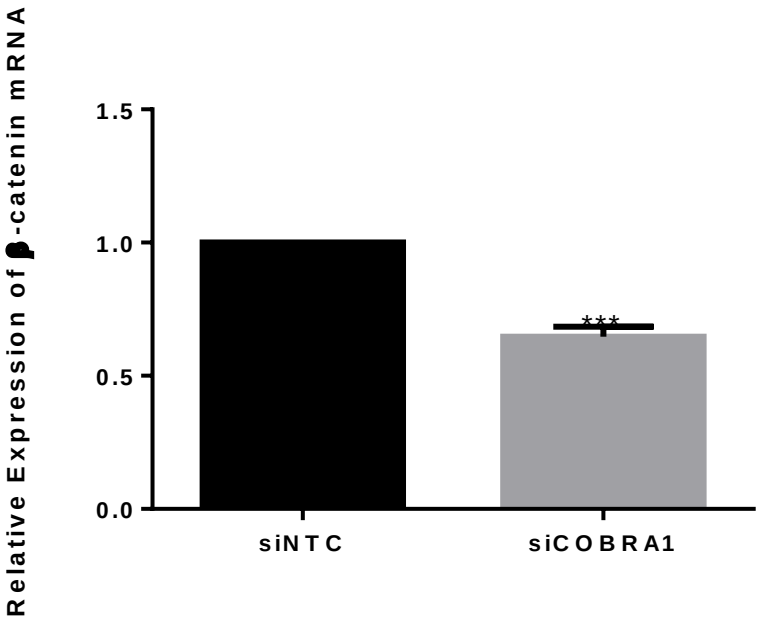

Figure 4 Downregulation of *β-catenin* in COBRA1 Knockdown HeLa cells

mRNA expression was analyzed by qPCR. C_T were obtained and *β-catenin* expressions were normalized to the internal control *GAPDH*. The relative expression is indicated as fold change to the negative control (siNTC). Data represent the mean ± SD of 2 independent experiments (n = 6). Statistically significant at ***p < 0.001 (unpaired Student t-test, two-tailed). siNTC: Negative siRNA and siCOBRA1: COBRA1 Knockdown.

II. CCND1:

The mRNA steady state levels of *CCND1, which is a downstream target of b-catenin,* was examined post COBRA1 knockdown using qPCR. As shown in (Fig.5), upon COBRA1 knockdown, there was a significant decrease in the mRNA levels of *CCND1* by 31.51% in COBRA1 siRNA treated cells relative to the negative control with p-value < 0.0001.

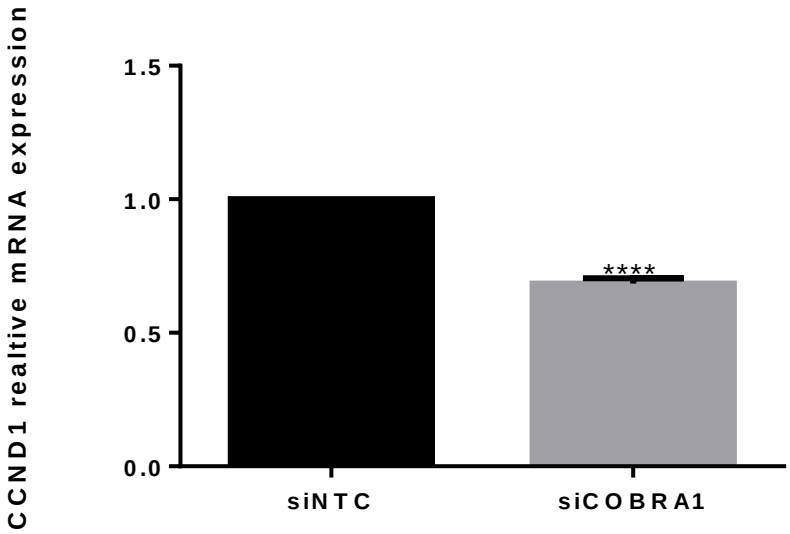

Figure 5 Downregulation of *CCND1* in COBRA1 Knockdown HeLa cells

mRNA expression was analyzed by qPCR. Ct were obtained and *CCND1* expressions were normalized to the internal control *GAPDH*. The relative expression is indicated as fold change to the negative control (siNTC). Data represent the mean ± SD of 2 independent experiments (n = 6). The statistical significance was calculated using two-tailed unpaired Student t-test and ****p < 0.0001 (siCOBRA1: COBRA1 siRNA)

C-MYC*:*

The mRNA steady state levels of *C-MYC*, which is a target gene in Wnt/β-catenin signaling pathway, was assessed using qPCR. A significant decrease in *c-MYC*, mRNA relative expression was found in COBRA-1 knockdown group when compared with control group and (P<0.0001)

(Fig.6), demonstrating that silencing of COBRA1 inhibited *c-MYC*. The decrease in expression level was estimated to be 31.72%.

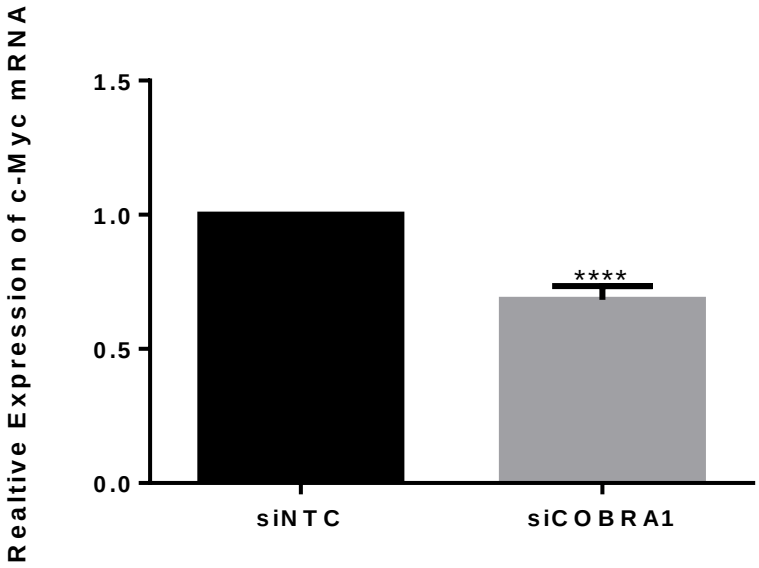

Figure 6 Downregulation of *c-MYC* in COBRA1 Knockdown HeLa cells

mRNA expression was analyzed by qPCR. C_T were obtained and *c-MYC* expressions were normalized to the internal control *GAPDH*. The relative expression is indicated as fold change to the negative control (siNTC). Data represent the mean ± SD of 2 independent experiments (n = 6). The statistical significance was calculated using two-tailed unpaired Student t-test and ****$p < 0.0001$ (siCOBRA1: COBRA1 siRNA)

3. Analysis of *FOXF2* promoter features:

Promoter report provided by the TRANSFAC® database showed that (SP-1) is the only transcription factor that is proven experimentally to bind to *FOXF2* promoter. (Fig.7) SP-1 is known to be the "Specificity Protein" that binds to TATA-less promoters, which contain a clear CpG island, as the case with the FOXF2 promoter shown in (Fig.7) it's an almost ubiquitous eukaryotic protein. [21] The binding site start position is at nucleotide 1389492 and the end position is at nucleotide 1389501. (Table.6) Additionally, other promoter features were included in the report. The features include CpG islands, repeats, histone modifications, and single nucleotide polymorphism (SNPs). The features are listed in (Table.4). Reviewing the predicted

transcription binding sites, a consensus DNA-binding sequence for BRCA1 protein complexes was observed at -1918 position relative to TSS. Also Estrogen Receptor alpha (Erα) has a binding site at position -1591 with higher core and matrix similarity score than BRCA1. (Table.3) However, COBRA1, which lacks a DNA binding site, was not included in the list. The full tables are provided in the supplementary data section.

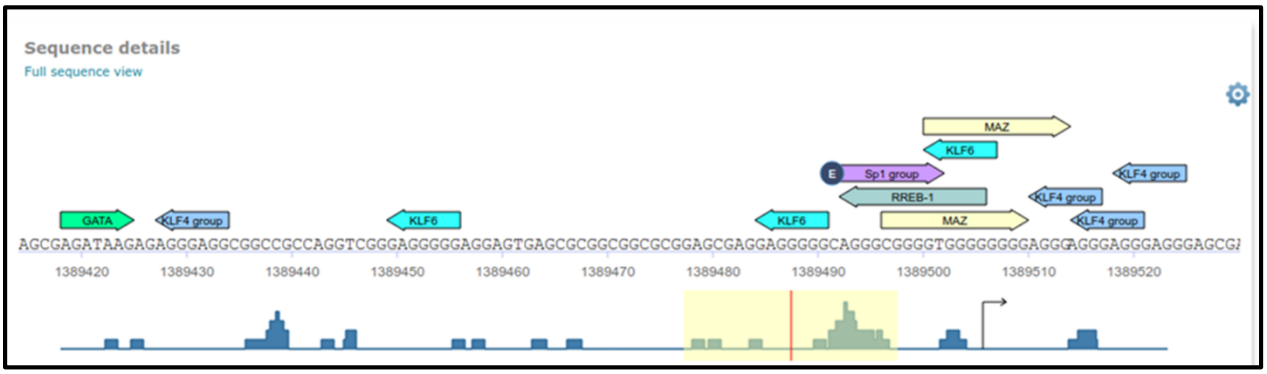

Figure 7 *FOXF2* promoter sequence view

The figure views the sequence of FOXF2 gene from nucleotide 1389414 to nucleotide 1389530.The purple arrow represents the SP-1 transcription factor and its relative binding site. The letter E in the circle represents the experimental validation.

Table 3 part of the predicted transcription factors binding sites using Match tool embedded in the TRANSFAC database.

Factor name	Core Similarity Score	Matrix Similarity Score	Start Position	Relative start position	Sequence
BRCA1:USF2	0.997	0.996	1387658	-1918	CAACCCAA
ER-alpha	1	1	1387985	-1591	CAGGTCA

Table 4 The promoter information provided by the TRANSFAC database and the Match tool predictions:

Features	Results
Histone modifications	41 entries
Experimental binding sites	2 entries
TSS associated with the promoter	1 entry
Repeats	37 entries
CpG islands	4 entries
Match predicted transcription factor binding sites	508 entries
SNPs	2751 entries

4. ChIP optimization:

Having established that COBRA1 knockdown RNA behaved differently than RNA from Negative siRNA, chromatin immunoprecipitation (ChIP) was performed in HeLa cells. The ChIP assays were carried out according to the protocol of the kit. However, several rounds of troubleshooting were performed to optimize the technique to our laboratory and experiment.

I. Shearing Optimization

First, it was important to optimize the shearing conditions to have cross-linked DNA in the range of ~200-1000 base pairs in length. 3,4,5and 6 pulses were tried and each time 5µl aliquot was withdrawn and mixed with RNAse (10 mg/mL) and incubated for 30 minutes at 37°C. Then 1µL of Proteinase K was added and incubated at 62°C for 2 hours. Then all samples were loaded 2% agarose gel with a 100bp and 1kb DNA marker. As seen in (Fig.8A), the bands have comparable

sizes. So, 4 pulses were used in the biological replicates to have proper shearing and also avoid over shearing and exposing the DNA to unnecessary heat. (Fig.8B)

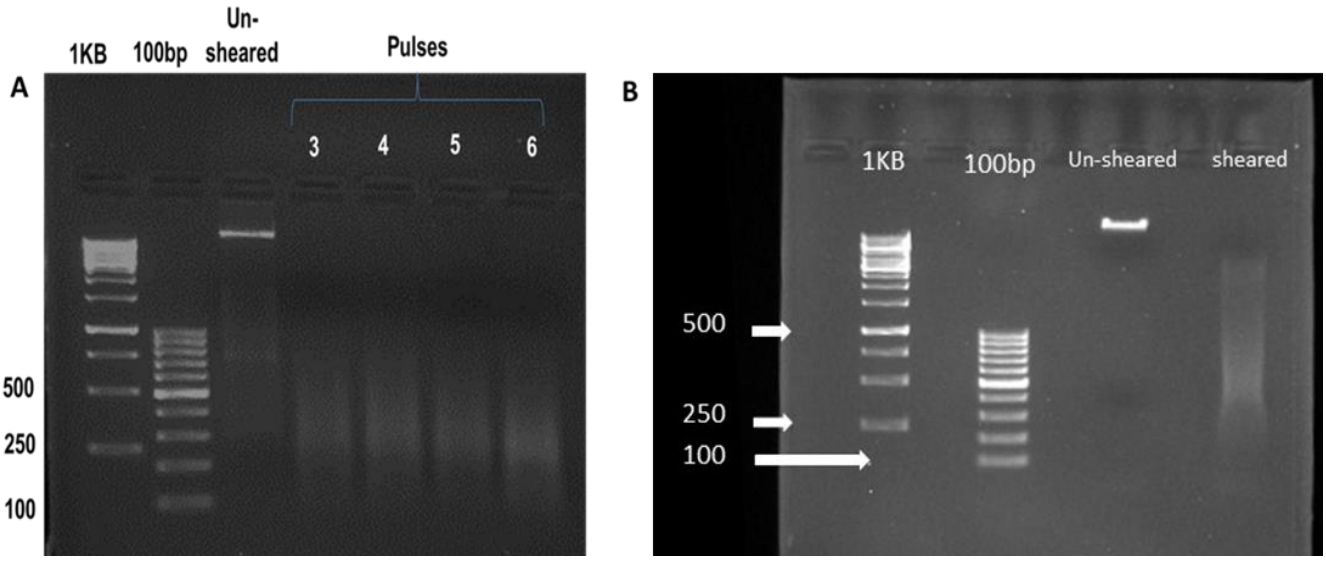

Figure 8 Shearing optimization

In (A), sheared chromatin from HeLa after sonication for 3,4,5 and 6 pulses. Two molecular weight markers (1kb and 100bp) were used. In (B), sheared chromatin using 4pulses. The smear indicates the proper shearing of chromatin between 250 and 750bp.

II. Positive control ChIP assay:

Chromatin immunoprecipitation was performed according to the optimized protocol using chromatin from HeLa cells and either Anti-Pol II or Normal IgG as the immunoprecipitating antibody. Purified DNA was then analyzed by RT- PCR using control primers specific for the *GAPDH* promoter. *GAPDH* promoter-specific DNA was observed in the 1% Input. PCR product was also observed in the IP from Anti RNA-Pol II and no PCR product was detected in the Normal IgG ChIP. (Fig.9)

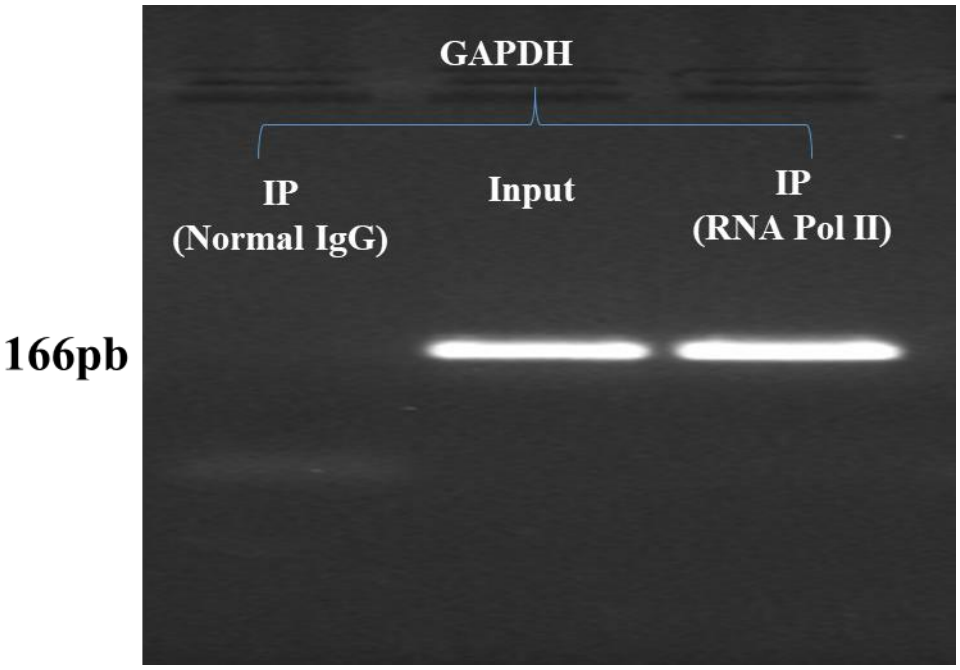

Figure 8 PCR Analysis of Chromatin Immunoprecipitation for Positive control :

GAPDH promoter specific DNA was observed in the 1% Input. PCR product was also observed in the IP from Anti RNA-Pol II and no PCR product was observed in the Normal IgG ChIP. This ensures the efficiency of the optimized protocol.

5. COBRA1 is required for transcriptional repression of *FOXF2* in HeLa cells:

I. ChIP RT-PCR:

Chromatin immunoprecipitation was performed according to the optimized protocol using chromatin from HeLa cells and Normal IgG or Anti-NELF-B (Anti-COBRA1) as the immunoprecipitating antibody. Purified DNA was then analyzed by RT- PCR using primers specific for the *FOXF2* promoter regions (Region 1A with relative position -1822 and Region 1C with relative position -474). PCR products were observed in the input samples and the IP from Anti-NELF-B (Anti-COBRA1). However, the bands for the region 1A were more intense than 1C. Also, the band from IP (NELF-B) using 1C primers was less intense than the input sample. No bands appear in the Normal IgG ChIP for both primers. (Fig.10)

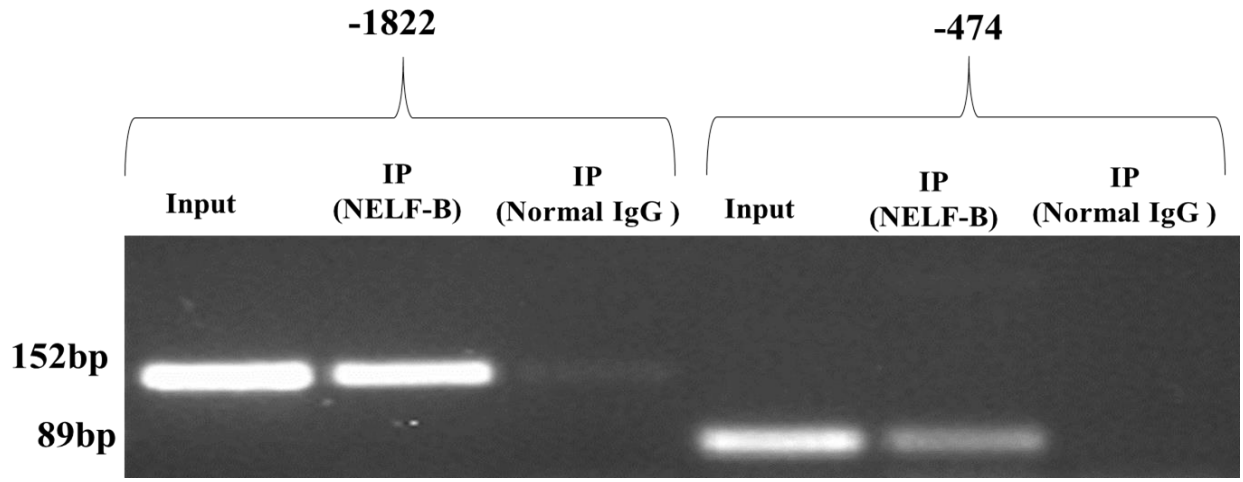

Figure 9 PCR Analysis of Chromatin Immunoprecipitation:

Anti NELF-B (COBRA1) showed increased precipitation of the FOXF2 promoter region fragment (1A) and (1C) in comparison to the IgG negative control fragment after agarose gel electrophoresis.

II. ChIP-qPCR:

To further confirm the previous results, and add a quantitative dimension to the ChIP assay, qPCR was used to analyze ChIP samples. Normalization of the cycle threshold (C_T) values from the qPCR assay was done by two methods, percent input method and fold enrichment method.

- Percent Input results:

This method represents the amount of DNA pulled down by using the Anti-NELF-B (Anti-COBRA1) and Normal IgG in the ChIP reaction, relative to the amount of starting material (input sample). For primers targeting region 1C in *FOXF2* promoter, as shown in (Fig.11), the % input of IP from Anti-NELF-B (Anti-COBRA1) is 0.0164 vs. %input of 0.00036 of IP from Normal IgG. For the other region in the *FOXF2* promoter (1A), as shown in (Fig.12), the % input of IP from Anti-NELF-B (Anti-COBRA1) is 0.0189 vs. %input of 0.0008 of IP from Normal IgG. Data in (Fig.11B) and (Fig.12B) represent the mean ± SD of 3 independent experiments (n = 6). Statistically significant at ***$p < 0.001$ for (1C) region and **$p<0.01$ for (1A) region. (unpaired Student t-test, two tailed).

- Fold enrichment results:

Anti NELF-B (Anti-COBRA1) antibody showed enhanced precipitation of the *FOXF2* promoter region fragment (1C) and (1A) in comparison to the IgG negative control fragment upon agarose gel electrophoresis. This was confirmed using qPCR, which showed ~ 33-fold enrichment and ~ 23-fold enrichment in comparison to the IgG negative control for (1C) and (1A) respectively. (Fig.11C) and (Fig.12C). An antibody for the RNA-Pol II was used as a positive control and showed an enrichment of approximately 45-fold. (Data not shown)

A

Sample Name	Promoter region	C_T	ΔC_T to input	Fold difference ($E^{\Delta CT}$)	%Input	Fold Enrichment
Input	1C	23.983006	0	1	1	-
IP (Anti-NELF-B)	1C	29.819245	-5.836239	0.016433	**0.0164331**	**33.08**
IP (IgG)	1C	35.161264	-11.178258	0.00037	**0.0003698**	**1**

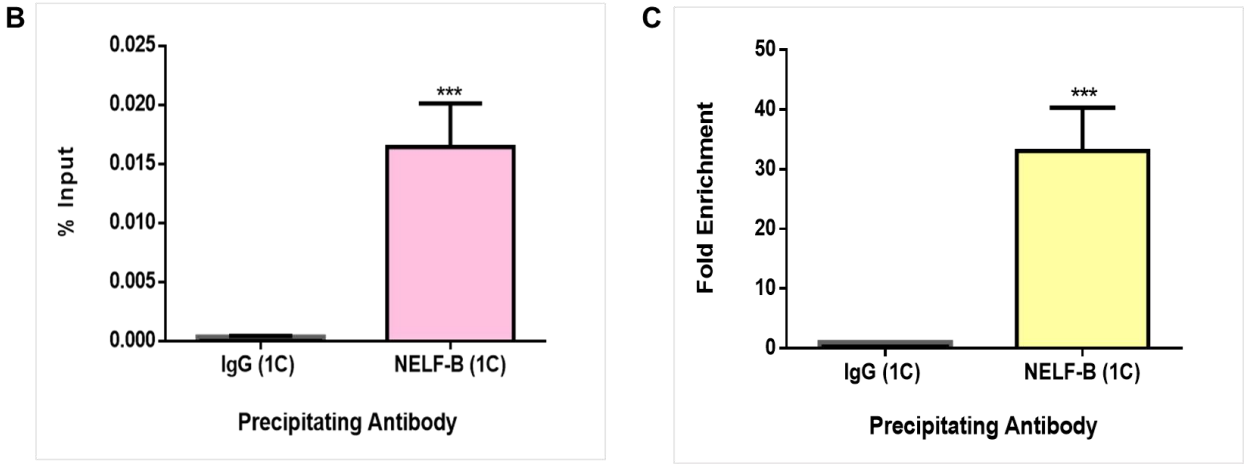

Figure 10: ChIP-qPCR calculations and data presentation for primers targeting 1C (-474) region.

(A) ChIP-qPCR data testing the NELF-B (COBRA1) occupation on the *FOXF2* promoter region (1C) (B) The %input calculated in (A) was plotted. The data demonstrates that the % input of IP from Anti-NELF-B (Anti-COBRA1) is 0.0164 vs. %input of 0.0003698 of IP from Normal IgG. (C) ChIP-qPCR results showing enhanced binding of NELF-B (COBRA1) to the *FOXF2* promoter of approximately 33 folds. The chart represents the mean ± SD of 3 independent experiments (n = 6). The statistical significance was calculated using two tailed unpaired Student t-test and ***$p < 0.001$.

A

Sample Name	Promoter region	C_T	ΔC_T to input	Fold difference ($E^{\Delta CT}$)	%Input	Fold Enrichment
Input	1A	24.63571701	0	1	1	-
IP (Anti-NELF-B)	1A	30.31818581	-5.6824688	0.018927701	**0.018927701**	**23.01**
IP (IgG)	1A	34.8379364	-10.2022194	0.000806728	**0.000806728**	**1**

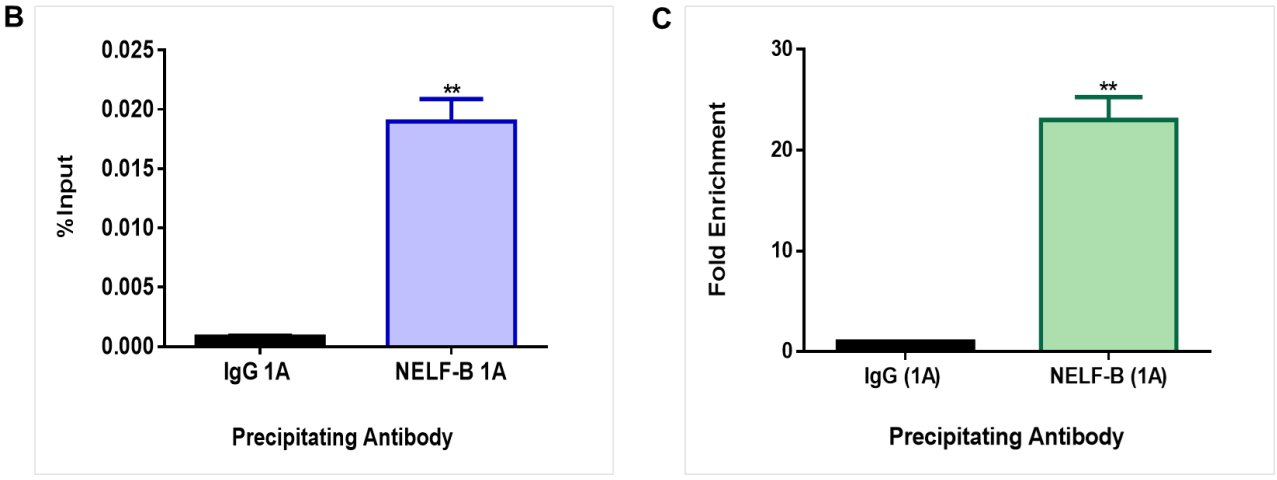

Figure 11 ChIP-qPCR calculations and data presentation for primers targeting 1A (-1822) region.

(A)ChIP-qPCR data analyzing NELF-B (COBRA1) occupancy on the FOXF2 promoter region (1C) (B) The result from the calculations in (A) plotted as % input. The data demonstrates that the % input of IP from Anti-NELF-B (Anti-COBRA1) is 0.0189 vs. % input of 0.0008 of IP from Normal IgG. (C) ChIP-qPCR results showing enhanced binding of NELF-B (COBRA1) to the *FOXF2* promoter of approximately 23 folds. The chart represents the mean ± SD of 3 independent experiments (n = 6). The statistical significance was calculated using two tailed unpaired Student t-test and **$p < 0.01$.

Discussion

Forkhead box F2 (FOXF2) is a gene member of the FOXF subfamily in the FOX gene family with biased expression in the lung and prostate.[22][23] Several studies confirmed the association between *FOXF2* and various tumors' development. *FOXF2* was reported to be downregulated in hepatocellular carcinoma, and a predictive factor of lung cancer. It was revealed that *FOXF2* downregulation leads to poor prognosis of patients with breast and lung cancer. [24], [25]

Cervical cancer ranks 13th as the most frequent cancer amongst women in Egypt and ranks as the 10th most frequent cancer among women between the ages of 15 and 44 years. It accounts for a total number of 866 cancer new cases and 373 cancer-related deaths annually. Worldwide, cervical cancer ranks fourth as the most common cancer in women. Cervical cancer is a significant public health issue influencing moderately aged women, especially in less-resourced nations. [26]

A recent study showed that low FOXF2 expression was associated with poor outcomes of cervical cancer patients and that overexpression of FOXF2 inhibited HeLa cells proliferation, migration, and invasion *in vitro* and growth *in vivo* through regulating the Wnt signaling pathway.[19]

A previous study was done by our group on the role of COBRA1 in cervical cancer. Using the HeLa cell line, the knockdown of COBRA1 expression was achieved by performing the RNA interference technique. The efficiency of silencing was assessed at the RNA and protein levels. COBRA1 depletion resulted in a significant decrease in the *β-catenin* at the level of mRNA. This finding indicates an effect for COBRA1 on the Wnt/β-catenin signaling pathway.[5]

The fundamental purpose of this study is to identify if COBRA1 effect on the Wnt/β-catenin signaling pathway is by enhancing *FOXF2* levels.

Effect on the mRNA expression level of *FOXF2*:

Although FOXF2 was associated to numerous tumor development including cervical cancer[19], the lack of reported results concerning its transcriptional regulation is considered a hindrance to fully understand its role in cancer pathogenesis. In fact, the only reported transcription factor that regulate the expression of FOXF2 was SP-1. Tian et al (2015) described a new regulatory mechanism for *FOXF2* expression in breast cancer cells and reported that the effects of DNA methylation and SP-1 transcriptional regulation together affect the subtype-specific expression and function of *FOXF2* in breast cancer cell lines. [21]

In this study, examining the FOXF2 mRNA levels in HeLa cell line upon COBRA1 silencing revealed an inverse correlation between COBRA1 and FOXF2. There was a significant increase in *FOXF2* mRNA expression in COBRA1 siRNA treated cells compared to the negative siRNA treated cells ($p<0.01$). This finding suggests that COBRA1 negatively regulates FOXF2 in HeLa cells and could be considered as a potential transcriptional regulator of FOXF2. However, the TRANSFAC® data did not show COBRA1 as a predicted transcription factor. This is due to the lack of DNA binding domain in COBRA1 protein.[27] Taking the previous findings into consideration suggested that COBRA1 regulates the transcription of *FOXF2* by indirect association with DNA via binding to other site-specific transcription factors.

Effect on the mRNA expression level of *β-catenin*:

Although HPV infection considered to be the crucial factor in the development of cervical cancer, numerous studies reported the activation of Wnt/β-catenin signaling as a second trigger to CC development.[28] Furthermore, targeting positive regulators in Wnt/β-catenin signaling impairs cervical cancer cell growth.[29] First, Wnt binds to its receptors and activates the canonical pathway, one of the multiple activated pathways, and then induces the entry of β-catenin to the nucleus. β-catenin is a multifunctional protein that has the dual activity of mediating cell adhesion

and signal transduction. After accumulation in the nucleus, β-catenin would form a transcription factor complex with the transcription factor TCF/LEF through its C-terminal transcriptional activator binding site, thereby facilitating the transcription of downstream target genes such as *CCNDl, C-JUN, and C-MYC*.[19] Thus, Wnt/β-catenin signaling pathway control multiple biological processes such as cellular proliferation, migration, and organogenesis. A recent study revealed that in cervical cancer, FOXF2 depletion induces cellular proliferation via β-catenin signaling. This was reflected in the expression levels of target genes (C-MYC and CCND1) [7]. Additionally, a study by our laboratory observed decreased expression of β-catenin at the mRNA and protein level in COBRA1 siRNA treated cells compared to negative siRNA treated cells. This decrease was found to be statistically significant at a p-value <0.05.[5] In line with these results, we observed decreased expression of *β-catenin* in COBRA1 siRNA treated cells compared to negative siRNA treated cells. This decrease was found to be statistically significant at a p-value <0.0001. So, the increased mRNA levels of *FOXF2* upon COBRA1 silencing was associated with decreased *β-catenin* mRNA steady-state levels.

Since the effect of COBRA1 knockdown on β-catenin was assessed, the following relevant step is to evaluate the expression of the Wnt signaling pathway down target genes that are affected by the increase in *FOXF2*. CCND1, LGR5, C-MYC, and MMP9 were reported to be inhibited by FOXF2 upregulation.[19] Yet, *CCND1* and *C-MYC* were the two genes that we validated.

Effect on the mRNA expression level of *CCND1*:

Cyclin D1 (CCND1)-is located at human chromosome 11q13. The expressed protein controls the transition of the cell cycle from G1 to S, which is a key regulator in the cell cycle. CCND1 expression is upregulated in tumor tissues, and its dysregulation has been proved to be tied to different malignant cancers.[30] Concerning CC, evidence supported the correlation between CCND1 upregulation and radio-resistance of cervical cancer via influencing proliferation, cell

cycle arrest, and apoptosis. Besides, the knockdown of CCND1 also impaired the proliferative ability of HeLa and SiHa cells.[31] Several transcription regulators of *CCND1* have been identified including Epidermal Growth Factor Receptor (EGFR), Phosphatidylinositol 3-kinase (PI3K), and the NF-κB transcription factor family. Added to the above-mentioned transcription regulators, the β-catenin/LEF-1 complexes bind at position -75 and -15 within the *CCND1* promoter. [32] As previously mentioned, FOXF2 depletion induces cellular proliferation via β-catenin signaling. This was reflected in the expression levels of the Wnt signaling pathway target genes c-MYC and Cyclin D1.[19] In agreement with these results, we observed a decreased expression of *CCND1* in COBRA1 siRNA treated cells compared to negative siRNA treated cells. This decrease was found to be statistically significant at a p-value <0.0001. Hence, the enhancement of the *FOXF2* mRNA level upon COBRA1 silencing was associated with decreased *CCND1* expression via β-catenin signaling.

Effect on the mRNA expression level of *C-MYC*:

C-MYC is a proto-oncogene that is usually tightly controlled by the extracellular signals [33] when altered by mutation, C-MYC becomes an oncogene that plays a role in proliferation regulation, cell cycle progression, and cellular transformation. Higher levels of this gene is usually noticed in various types of human cancers.[34] Regarding CC, the high expression of C-MYC promotes its invasion and metastasis.[35] As mentioned before, *FOXF2* reduction was echoed on the expression levels of the Wnt signaling pathway target genes (c-Myc, CyclinDl, MMP9, and Lgr5).[19] Conforming to these results, we observed declined expression of *C-MYC* in COBRA1 siRNA treated cells compared to negative siRNA treated cells. This decline was found to be statistically significant at a p-value <0.0001. Thus, the increased levels of *FOXF2* mRNA after COBRA1 knockdown was associated with decreased *C-MYC* transcription via β-catenin signaling.

COBRA1 is required for transcriptional repression of *FOXF2* in HeLa cells:

As shown above, the knockdown of COBRA1 was associated with an increase in the mRNA levels of *FOXF2*. Besides, overexpression of FOXF2 and silencing of COBRA1 have comparable effects on β-catenin and its downstream target genes. Based on the aforementioned results, carrying out the ChIP assay will be of value in examining the relation between COBRA1 and the promoter of the *FOXF2* gene. Hence COBRA1 has no DNA binding domain, the transcription regulation of the target gene is achieved through complex formation with other transcription factors including BRCA1 and NELF complex.[27]

In addition to its roles with BRCA1 and as a functional component of the NELF complex, a novel role for COBRA1 in the regulation of hormone-responsive transcription was identified by Aiyar et al., (2004). In breast cancer cells, one of the mechanisms for attenuating the ERα mediated gene activation is through direct binding with COBRA1. Moreover, COBRA1 and the rest of the NELF complex are recruited to a number of endogenous ERα-responsive promoters in an estrogen-stimulated fashion. The promoter-bound NELF complex acts to stall RNA polymerase II (RNAPII) movement at the promoter-proximal region and attenuate ERα-dependent transcription.[36]

In our study, analysis of the data from the TRANSFAC® database indicated that Estrogen receptor alpha (ERα) and BRCA1 were of the predicted TFs that bind to the promoter of *FOXF2*. After designing primer pairs to target this region, data from RT-PCR and qPCR showed enhanced precipitation of the FOXF2 promoter region fragment (1A) in comparison to the IgG negative control fragment. As the scores that demonstrate the quality of match were higher in the case of ER-α, the data obtained from TRANSFAC® Database combined with PCR data suggests that COBRA1 might regulate the transcription of *FOXF2* by interacting directly with ERα rather than BRCA1.

Another potential complex to regulate and bind to the *FOXF2* promoter is the NELF complex. To design primers targeting the RNA Pol II binding site, it should be spanning around -70 to 200bp relative to the transcription site. However, by specifying this region in Primer Blast to create primers for it, all the primers rendered were not specific to *FOXF2* and 100 other hits were obtained. The nearest primer that could be designed to this region and at the same time specific to *FOXF2* was at position -474 relative to the TSS. After designing primer pairs to target this region, data from RT-PCR and qPCR showed enhanced precipitation of the FOXF2 promoter region fragment (1C) in comparison to the IgG negative control fragment. Consequently, this enhanced precipitation indicates that COBRA1 within the NELF complex might decrease FOXF2 transcription by halting the RNAPII at its promoter-proximal region.

Although the ChIP-qPCR data shown in Fig.12 and Fig.13 looks convincing, it is important to understand that it is only credible if the enrichment observed in the IP from Anti-NELF-B (Anti-COBRA1) is significant. Therefore, the threshold of significant enrichment should be experimentally determined for our specific antibody/chromatin complex.[37] To achieve this purpose, a comparison of enrichment signals from wild-type cells versus ChIP target negative control cells has to be performed. The best control is a knockout cell line deficient for *FOXF2* while the second-best choice is to use the siRNA transfection procedure to knock down NELF-B.

Based on ChIP results the following model for regulation of *FOXF2* via COBRA1 is suggested. COBRA1 regulates *FOXF2* transcription via interacting directly with ERα, which has a predicted binding site in its promoter. Also, COBRA1 decreases *FOXF2* transcription by pausing the RNAPII at its promoter-proximal region. In (Fig.13), a layout of the results is combined to show that knockdown of COBRA1 promoted the expression of *FOXF2*. Amplified *FOXF2* expression suppressed the expression of the Wnt signaling pathway downstream target genes, including (*β-Catenin-C-MYC*-and *CCND1*).

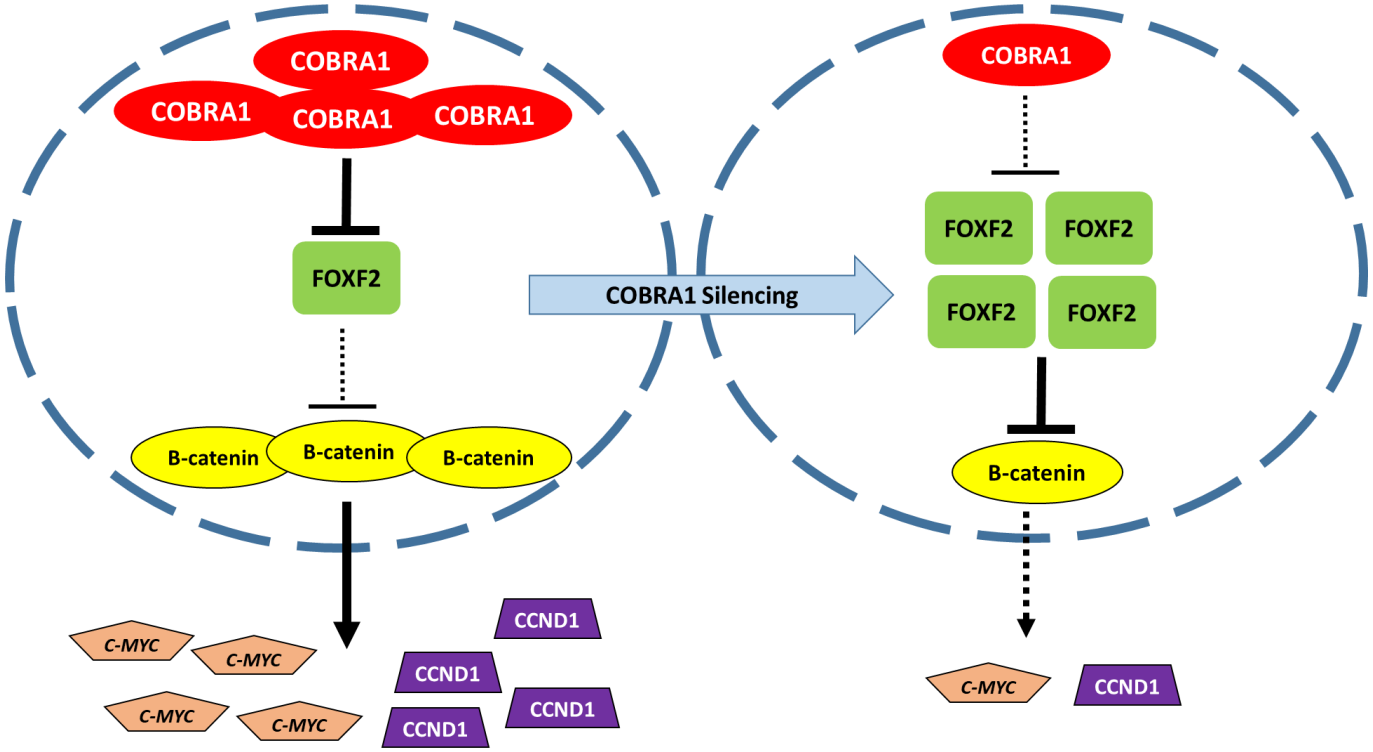

Figure 12 Schematic diagram illustrating the consequences of COBRA1 silencing on mRNA expression levels of examined genes.

Successful COBRA1 depletion in HeLa cells caused a significant increase in the expression of *FOXF2* leading to decreased expression levels of the *β-catenin* resulting in decreased levels of *β-catenin* target genes, *CCND1* and *C-MYC*.

Conclusion

In short, this research explored that knockdown of COBRA1 promoted the mRNA expression of *FOXF2*. Upregulation of *FOXF2* mRNA expression appeared to be associated with repression in the mRNA expression levels of target genes in the Wnt signaling pathway, including (*β-Catenin*, *C-MYC*, and *CNDD1*). Based on these findings, we suspected that FOXF2 might be the modulator of the effects observed on the β-catenin expression. By performing ChIP assay, the interaction between COBRA1 and *FOXF2* was proven. Due to the absence of a DNA binding domain in COBRA1, this interaction should be mediated through complex formation. Data obtained from TRANSFAC® Database showed that COBRA1 might regulate the transcription of *FOXF2* by interacting directly with ERα, which has a predicted binding site in the *FOXF2* promoter. In addition, COBRA1 within the NELF complex might decrease *FOXF2* transcription by halting the RNAPII at its promoter-proximal region. Altogether considered, this study could help as an initial step in identifying the role of COBRA1 in cervical cancer tumorigenesis. Nonetheless, future studies need to be conducted to further validate our work.

Future Recommendations

Future studies are needed to further understand the role of COBRA1 as a transcription regulator for *FOXF2* gene and more characterization for its role in CC pathogenesis. Performing the ChIP assay in COBRA1-knocked HeLa cells to assess the significance of enrichment and confirm the previous results is essential. Also, combining ChIP assay with sequencing would be beneficial to map DNA-binding proteins in a genome-wide manner at base-pair resolution. Another beneficial test is luciferase assay, to establish a functional connection between the presence of the COBRA1 and the amount of FOXF2 gene product that is produced. Finally, adding exogenous estrogen together with the silencing process could help in proving the model suggested based on the TRANSFAC® data and ChIP results.

References

[1] A. M. da Costa, D. Hashim, J. H. T. G. Fregnani, and E. Weiderpass, "Overall survival and time trends in breast and cervical cancer incidence and mortality in the Regional Health District (RHD) of Barretos, São Paulo, Brazil," *BMC Cancer*, vol. 18, no. 1, p. 1079, Nov. 2018, doi: 10.1186/s12885-018-4956-7.

[2] "WHO Cervical cancer," *WHO*. http://www.who.int/cancer/prevention/diagnosis-screening/cervical-cancer/en/ (accessed Apr. 29, 2020).

[3] Bruni L, Barrionuevo-Rosas L, Albero G, Serrano B, Mena M, Gómez D, Muñoz J, Bosch FX, de S. S. 27. Human Papillomavirus and Related Diseases in the World. Summary Report. ICO/IARC Information Centre on HPV and Cancer (HPV Information Centre). (2017).

[4] Colombo, N., Carinelli, S., Colombo, A., Marini, C., Rollo, D., & Sessa, C. Cervical cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. Annals of Oncology, 23(SUPPL. 7). https://doi.org/10.1093/annonc/mds268 (2012).

[5] Saad, N. HeLa cell line, a model to study the role of cofactor of BRCA1 in cervical cancer. American University in Cairo. (2018)

[6] C. Marth *et al.*, "Cervical cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up," *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol.*, vol. 28, no. suppl_4, pp. iv72–iv83, Jul. 2017, doi: 10.1093/annonc/mdx220.

[7] "Harvey Lodish, Arnold Berk, Chris A. Kaiser, Monty Krieger, Anthony Bretscher, Hidde Ploegh, Angelika Amon, Kelsey C. Martin-Molecular Cell Biology-W. H. Freeman (2016).pdf.".

[8] S. Ramírez-Clavijo and G. Montoya-Ortíz, *Gene expression and regulation*. El Rosario University Press, 2013.

[9] J. E. Bradner, D. Hnisz, and R. A. Young, “Transcriptional Addiction in Cancer,” *Cell*, vol. 168, no. 4, pp. 629–643, Feb. 2017, doi: 10.1016/j.cell.2016.12.013.

[10] Y. Q. al et, “BRCA1-induced large-scale chromatin unfolding and allele-specific effects of cancer-predisposing mutations. - PubMed - NCBI.” https://www.ncbi.nlm.nih.gov/pubmed/11739404 (accessed Oct. 29, 2018).

[11] T. Narita *et al.*, “Human transcription elongation factor NELF: identification of novel subunits and reconstitution of the functionally active complex,” *Mol. Cell. Biol.*, vol. 23, no. 6, pp. 1863–1873, Mar. 2003.

[12] S. E. Aiyar, A. L. Blair, D. A. Hopkinson, S. Bekiranov, and R. Li, “Regulation of clustered gene expression by cofactor of BRCA1 (COBRA1) in breast cancer cells,” *Oncogene*, vol. 26, no. 18, pp. 2543–2553, Apr. 2007, doi: 10.1038/sj.onc.1210047.

[13] J. Sun, A. L. Blair, S. E. Aiyar, and R. Li, “Cofactor of BRCA1 modulates androgen-dependent transcription and alternative splicing,” *J. Steroid Biochem. Mol. Biol.*, vol. 107, no. 3–5, pp. 131–139, Dec. 2007, doi: 10.1016/j.jsbmb.2007.05.031.

[14] H. Zhong *et al.*, “COBRA1 inhibits AP-1 transcriptional activity in transfected cells,” *Biochem. Biophys. Res. Commun.*, vol. 325, no. 2, pp. 568–573, Dec. 2004, doi: 10.1016/j.bbrc.2004.10.079.

[15] Saad, N. HeLa cell line, a model to study the role of cofactor of BRCA1 in cervical cancer. American University in Cairo. (2018)

[16] Kong, P.Z. et al. (2016) Decreased expression of FOXF2 as new predictor of poor prognosis in stage I non-small cell lung cancer. Oncotarget 7, 55601–55610, https://doi.org/10.18632/oncotarget.10876

[17] P.-Z. Kong, F. Yang, L. Li, X.-Q. Li, and Y.-M. Feng, “Decreased FOXF2 mRNA Expression Indicates Early-Onset Metastasis and Poor Prognosis for Breast Cancer Patients with

Histological Grade II Tumor," *PLOS ONE*, vol. 8, no. 4, p. e61591, Apr. 2013, doi: 10.1371/journal.pone.0061591.

[18] Wang, Q.S. et al. (2015) FOXF2 deficiency promotes epithelial-mesenchymal transition and metastasis of basal-like breast cancer. Breast Cancer Res.17, 30, https://doi.org/10.1186/s13058-015-0531-1

[19] J. Zhang, C. Zhang, L. Sang, L. Huang, J. Du, and X. Zhao, "FOXF2 inhibits proliferation, migration, and invasion of Hela cells by regulating Wnt signaling pathway," *Biosci. Rep.*, vol. 38, no. 5, p. BSR20180747, Oct. 2018, doi: 10.1042/BSR20180747.

[20] J. Ye, G. Coulouris, I. Zaretskaya, I. Cutcutache, S. Rozen, and T. L. Madden, "Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction," *BMC Bioinformatics*, vol. 13, p. 134, Jun. 2012, doi: 10.1186/1471-2105-13-134.

[21] H.-P. Tian *et al.*, "DNA Methylation Affects the SP1-regulated Transcription of FOXF2 in Breast Cancer Cells," *J. Biol. Chem.*, vol. 290, no. 31, pp. 19173–19183, Jul. 2015, doi: 10.1074/jbc.M114.636126.

[22] P. Carlsson and M. Mahlapuu, "Forkhead Transcription Factors: Key Players in Development and Metabolism," *Dev. Biol.*, vol. 250, no. 1, pp. 1–23, Oct. 2002, doi: 10.1006/dbio.2002.0780.

[23] "FOXF2 forkhead box F2 [Homo sapiens (human)] - Gene - NCBI." https://www.ncbi.nlm.nih.gov/gene/2295 (accessed Apr. 16, 2020).

[24] Kong, P.Z. et al. (2013) Decreased FOXF2 mRNA expression indicates early-onset metastasis and poor prognosis for breast cancer patients with histological grade II tumor. PLoS ONE 8, e61591, https://doi.org/10.1371/journal.pone.0061591

[25] P.-Z. Kong, G.-M. Li, Y. Tian, B. Song, and R. Shi, "Decreased expression of FOXF2 as new predictor of poor prognosis in stage I non-small cell lung cancer," *Oncotarget*, vol. 7, no. 34, pp. 55601–55610, Jul. 2016, doi: 10.18632/oncotarget.10876.

[26] M. Arbyn *et al.*, “Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis,” *Lancet Glob. Health*, vol. 8, no. 2, pp. e191–e203, Feb. 2020, doi: 10.1016/S2214-109X(19)30482-6.

[27] A. Amleh, S. J. Nair, J. Sun, A. Sutherland, P. Hasty, and R. Li, “Mouse Cofactor of BRCA1 (Cobra1) Is Required for Early Embryogenesis,” *PLoS ONE*, vol. 4, no. 4, p. e5034, Apr. 2009, doi: 10.1371/journal.pone.0005034.

[28] C. Perez-Plasencia, A. Duenas-Gonzalez, and B. Alatorre-Tavera, “Second hit in cervical carcinogenesis process: involvement of wnt/beta catenin pathway,” *Int. Arch. Med.*, vol. 1, p. 10, Jul. 2008, doi: 10.1186/1755-7682-1-10.

[29] H. T. Kwan *et al.*, “AMPK Activators Suppress Cervical Cancer Cell Growth through Inhibition of DVL3 Mediated Wnt/β-Catenin Signaling Activity,” *PLoS ONE*, vol. 8, no. 1, Jan. 2013, doi: 10.1371/journal.pone.0053597.

[30] M. Yang, H. Zhu, T. Hu, S. Liu, and H. Wang, “Association of CCND1 gene polymorphism with cervical cancer susceptibility in Caucasian population: a meta-analysis,” *Int. J. Clin. Exp. Med.*, vol. 8, no. 8, pp. 12983–12988, Aug. 2015.

[31] D. Han, J. Wang, and G. Cheng, “LncRNA NEAT1 enhances the radio-resistance of cervical cancer via miR-193b-3p/CCND1 axis,” *Oncotarget*, vol. 9, no. 2, pp. 2395–2409, Dec. 2017, doi: 10.18632/oncotarget.23416.

[32] S. Qie and J. A. Diehl, “Cyclin D1, Cancer Progression and Opportunities in Cancer Treatment,” *J. Mol. Med. Berl. Ger.*, vol. 94, no. 12, pp. 1313–1326, Dec. 2016, doi: 10.1007/s00109-016-1475-3.

[33] “Robert A. Weinberg, Robert A Weinberg - The biology of cancer (2014, Garland Science).pdf.”.

[34] ""MYC MYC proto-oncogene, bHLH transcription factor [Homo sapiens (human)] - Gene - NCBI," National Center for Biotechnology Information. [Online]. Available: https://www.ncbi.nlm.nih.gov/gene [Accessed: 23-May-2020].

[35] W. Sh, Z. Xf, W. P, Z. Y, and L. W, "[The Expression and Significance of C-Myc and bcat1 in Cervical Cancer]," *Sichuan da xue xue bao. Yi xue ban = Journal of Sichuan University. Medical science edition*, Sep. 2018. https://pubmed.ncbi.nlm.nih.gov/(accessed Apr. 28, 2020).

[36] S. E. Aiyar *et al.*, "Attenuation of estrogen receptor α-mediated transcription through estrogen-stimulated recruitment of a negative elongation factor," *Genes Dev.*, vol. 18, no. 17, pp. 2134–2146, Sep. 2004, doi: 10.1101/gad.1214104.

[37] N. Visa and A. Jordán-Pla, Eds., *Chromatin Immunoprecipitation: Methods and Protocols*, vol. 1689. New York, NY: Springer New York, 2018.

Supplementary Data

Match predicted Transcription factors binding sites in FOXF2 gene

Table 5 Predicted TFBSs

Factor name	Core Similarity Score	Matrix Similarity Score	Start Position	Relative start position	End Position	Sequence
C/EBPalpha	1	0.996	1379575	-10001	1379587	ATCTTTTGCAAAG
IRF-4	1	0.982	1379609	-9967	1379615	GAAAGTC
44107	0.799	0.82	1379624	-9952	1379636	AATAAGCAAACGC
DBP	1	1	1379628	-9948	1379634	AGCAAAC
AML1	0.904	0.907	1379647	-9929	1379655	TTGCGGTCG
c-Myc	1	0.799	1379655	-9921	1379666	GGCCACGTCCGC
Churchill	1	1	1379749	-9827	1379754	CGGGGG
BCL-6	0.963	0.889	1379762	-9814	1379771	CTCCTCGAAT
REST	1	0.92	1379797	-9779	1379809	CCCCAGTGCTGCA
HSF1	0.657	0.657	1379907	-9669	1379919	GAGCCCTAGAGAA
AML2	1	0.925	1379925	-9651	1379935	CCACAGCCCGA
CSX	1	0.992	1379978	-9598	1379988	CACACTTGTAG
Nanog	1	0.96	1379988	-9588	1379999	GACAATGGACCC
Helios A	1	0.987	1380002	-9574	1380012	ATTTTCCTCAT
POU6F1	1	0.927	1380003	-9573	1380019	TTTTCCTCATTAAAGAC
RREB-1	1	0.835	1380022	-9554	1380035	ACCCAAATCCGCCT
c-Myb	1	0.997	1380168	-9408	1380184	TTTATAACCGTTATTTC
Helios A	1	0.988	1380231	-9345	1380241	GTTTTCCTGAT
AIRE	0.987	0.933	1380237	-9339	1380262	CTGATCCCCATTTTACCATAAAACAG
YY1	1	1	1380244	-9332	1380250	CCATTTT
P53	0.993	0.99	1380263	-9313	1380273	AAACAAGCCCA
NKX3-1	1	0.935	1380283	-9293	1380291	TTAAGTGAC
RelA-p65	0.927	0.835	1380390	-9186	1380401	GGGAAATGCTAA

Factor name	Core Similarity Score	Matrix Similarity Score	Start Position	Relative start position	End Position	Sequence
FAC1	1	0.964	1380402	-9174	1380415	GAGCACAACATGGA
CDP CR1	0.91	0.905	1380410	-9166	1380419	CATGGATTGT
HNF-6	1	1	1380442	-9134	1380449	AATCAATA
Xvent-1	1	0.935	1380452	-9124	1380464	CAACTATTTGTGA
GATA	1	1	1380513	-9063	1380519	CTTATCT
MZF-1	1	1	1380538	-9038	1380544	CTCCCCA
AP-2gamma	0.993	0.963	1380580	-8996	1380589	TCCCTCAGGC
TEAD4	0.913	0.865	1380607	-8969	1380616	AAAATTCTTC
LEF-1	1	1	1380698	-8878	1380704	TCAAAGG
c-Myc	1	0.91	1380703	-8873	1380714	GGACACGTGCTC
GATA-2	0.98	0.895	1380717	-8859	1380723	GTGATAC
TFIIB	0.941	0.88	1380739	-8837	1380749	ATATATGGAAG
Pax	0.9	0.891	1380755	-8821	1380765	CTGGGATTAAC
AHR	1	1	1380813	-8763	1380818	TGCGTG
Nanog	0.945	0.94	1380846	-8730	1380857	GGGAACATTTGC
Pax	0.818	0.86	1380907	-8669	1380917	CTGTAAGTCAC
FAC1	1	0.941	1380953	-8623	1380966	ATCCACAACAATGT
AHR	1	1	1381002	-8574	1381007	TGCGTG
c-Myb	0.862	0.89	1381165	-8411	1381181	TTGGTAACTGTTACATG
Irx2	1	0.987	1381171	-8405	1381187	ACTGTTACATGTAGGTG
ARNT	1	0.964	1381214	-8362	1381221	TCCACGTT
ATF-2	1	0.849	1381240	-8336	1381247	TTTCGTAA
TATA	1	0.985	1381254	-8322	1381268	ATATAAAAAGATCCT
RREB-1	0.901	0.834	1381326	-8250	1381339	CCCCACAACGACCC
BEN	1	0.96	1381443	-8133	1381450	AATCGCTG
Ets	1	1	1381485	-8091	1381492	ACTTCCTC

Factor name	Core Similarity Score	Matrix Similarity Score	Start Position	Relative start position	End Position	Sequence
ZF5	1	0.938	1381510	-8066	1381523	GCCGCGCAGAGTCG
ZBTB26	0.776	0.785	1381573	-8003	1381589	TGCTTGTTTTCCACCAA
AML3	0.885	0.897	1381588	-7988	1381597	AAACCGCAGG
ER-alpha	0.616	0.625	1381632	-7944	1381646	AAGACACAGAGCCCC
Churchill	1	1	1381643	-7933	1381648	CCCCCG
CTCFL	0.793	0.831	1381651	-7925	1381662	CCTTGCTTCAGA
Xvent-1	1	0.916	1381664	-7912	1381676	AACCAAATAACGG
ERG	1	0.866	1381672	-7904	1381681	AACGGAACTG
ETS1	0.874	0.863	1381672	-7904	1381681	AACGGAACTG
AR	0.726	0.799	1381675	-7901	1381689	GGAACTGAATGATCG
Bach1	0.8	0.692	1381676	-7900	1381690	GAACTGAATGATCGT
ZNF561	0.657	0.681	1381677	-7899	1381688	AACTGAATGATC
44107	0.859	0.774	1381679	-7897	1381691	CTGAATGATCGTT
IRF-4	0.971	0.853	1381687	-7889	1381693	TCGTTTC
ZNF690	0.798	0.719	1381687	-7889	1381698	TCGTTTCTTCAG
LEF-1	1	0.972	1381693	-7883	1381705	CTTCAGCAAAGCC
FOXP1	0.873	0.755	1381700	-7876	1381714	AAAGCCTAAAGACGC
ZASC1	0.632	0.656	1381744	-7832	1381759	GACTCGGCTGCCCCCT
YY1	0.798	0.81	1381756	-7820	1381762	CCCTCTT
FOXO1A	0.73	0.729	1381758	-7818	1381765	CTCTTCTT
PEBP2beta	0.919	0.859	1381765	-7811	1381771	TGCGGGT
Osx	1	0.887	1381783	-7793	1381792	AGGGCGGGAA
Sp3	1	0.92	1381783	-7793	1381794	AGGGCGGGAACG
KLF14	0.815	0.792	1381793	-7783	1381806	CGCGGGACGTGCCC
HIF1	1	0.946	1381795	-7781	1381808	CGGGACGTGCCCTA
NF-KAPPAB1	0.93	0.915	1381795	-7781	1381807	CGGGACGTGCCCT

Factor name	Core Similarity Score	Matrix Similarity Score	Start Position	Relative start position	End Position	Sequence
44166	1	0.933	1381796	-7780	1381806	GGGACGTGCCC
p53 decamer	0.907	0.913	1381797	-7779	1381806	GGACGTGCCC
ARNT	1	0.977	1381798	-7778	1381805	GACGTGCC
CTCF	0.908	0.82	1381800	-7776	1381819	CGTGCCCTAAGCGGGCCAGC
ZNF2	0.775	0.709	1381811	-7765	1381822	CGGGCCAGCCGA
Smad3	0.926	0.923	1381824	-7752	1381836	CGCGAGACAGGCG
Smad4	0.889	0.892	1381826	-7750	1381833	CGAGACAG
CTCFL	0.81	0.83	1381835	-7741	1381846	CGATGCTGCAGA
ZNF553	0.636	0.77	1381835	-7741	1381844	CGATGCTGCA
AML1	0.904	0.794	1381858	-7718	1381866	CTGCGGGTC
COUP-TF1	1	0.917	1381861	-7715	1381867	CGGGTCA
KLF9	1	0.834	1381863	-7713	1381875	GGTCACACACCGG
HNF-3alpha	0.757	0.746	1381865	-7711	1381872	TCACACAC
Sox-9	0.943	0.86	1381868	-7708	1381884	CACACCGGCCGGCCTCG
ZSCAN30	1	0.939	1381879	-7697	1381893	GCCTCGGCTGCAGGG
RFX1	1	0.96	1381892	-7684	1381908	GGGTTGCGGGGCCACGG
c-Myc	0.8	0.795	1381901	-7675	1381912	GGCCACGGGTCG
MYCN	0.824	0.82	1381901	-7675	1381912	GGCCACGGGTCG
Pax-5	0.896	0.767	1381901	-7675	1381918	GGCCACGGGTCGCGGGGC
Zbtb44	0.97	0.899	1381903	-7673	1381908	CCACGG
MAX	0.915	0.871	1381929	-7647	1381940	GCGAGCAGGTGA
GATA-2	0.761	0.734	1381930	-7646	1381936	CGAGCAG
SIP1	1	0.966	1381930	-7646	1381944	CGAGCAGGTGACGGC
MyoD	0.996	0.97	1381931	-7645	1381939	GAGCAGGTG
slug	1	0.997	1381931	-7645	1381939	GAGCAGGTG
HTF4	1	1	1381933	-7643	1381939	GCAGGTG

Factor name	Core Similarity Score	Matrix Similarity Score	Start Position	Relative start position	End Position	Sequence
CREB1	1	0.943	1381934	-7642	1381948	CAGGTGACGGCCGCG
TGIF2	0.81	0.698	1381934	-7642	1381949	CAGGTGACGGCCGCGC
ZF5	1	0.926	1381943	-7633	1381956	GCCGCGCCGGGCAG
RelA-p65	0.857	0.833	1381950	-7626	1381961	CGGGCAGCCCCG
ZFHX2	0.973	0.91	1381950	-7626	1381960	CGGGCAGCCCC
ZFX	1	0.867	1381950	-7626	1381961	CGGGCAGCCCCG
sin3A	0.764	0.746	1381954	-7622	1381967	CAGCCCCGCGGGCC
AP-2gamma	1	0.99	1381957	-7619	1381966	CCCCGCGGGC
YY2	0.772	0.806	1381964	-7612	1381974	GGCCGCGATTC
ZNF610	0.822	0.672	1381973	-7603	1381984	TCCCGAGAGCCT
Muscle initiator	1	0.934	1381981	-7595	1382001	GCCTGGCGCCACCCCGCGGAA
ctcf	0.793	0.855	1381983	-7593	1382002	CTGGCGCCACCCCGCGGAAG
Rad21	0.793	0.813	1381983	-7593	1382001	CTGGCGCCACCCCGCGGAA
CTCF	0.959	0.822	1381984	-7592	1382003	TGGCGCCACCCCGCGGAAGC
SC-1	0.446	0.669	1381984	-7592	1381995	TGGCGCCACCCC
Kox1	0.911	0.72	1381986	-7590	1381997	GCGCCACCCCGC
KLF8	1	0.965	1381987	-7589	1381996	CGCCACCCCG
SMC-3	0.777	0.769	1381987	-7589	1382000	CGCCACCCCGCGGA
ERG	1	0.809	1381995	-7581	1382004	CGCGGAAGCC
GABP-beta	1	0.89	1381995	-7581	1382005	CGCGGAAGCCG
ZBTB20	0.517	0.773	1381995	-7581	1382006	CGCGGAAGCCGG
Elf-1	1	0.884	1382002	-7574	1382013	GCCGGAGGAACT
GABPA	1	0.811	1382002	-7574	1382011	GCCGGAGGAA
ESE-1	1	0.952	1382006	-7570	1382014	GAGGAACTG
ZNF785	0.581	0.579	1382008	-7568	1382028	GGAACTGCGGCTTCTTCCCGC
SPI1	0.596	0.716	1382010	-7566	1382023	AACTGCGGCTTCTT

Factor name	Core Similarity Score	Matrix Similarity Score	Start Position	Relative start position	End Position	Sequence
GLI4	0.759	0.768	1382016	-7560	1382027	GGCTTCTTCCCG
Aiolos	1	0.972	1382022	-7554	1382030	TTCCCGCTC
MAZ	1	0.951	1382034	-7542	1382047	CGCGGAGGGGAGAG
MAZR	0.935	0.9	1382034	-7542	1382046	CGCGGAGGGGAGA
MZF-1	1	0.98	1382038	-7538	1382047	GAGGGGAGAG
WT1	0.996	0.968	1382041	-7535	1382052	GGGAGAGGGAGG
ZNF37A	0.976	0.965	1382042	-7534	1382050	GGAGAGGGA
NRF-1	0.999	0.797	1382045	-7531	1382055	GAGGGAGGCGC
ZF5	1	0.903	1382050	-7526	1382063	AGGCGCAGCGCGGA
ZNF143	0.696	0.617	1382123	-7453	1382138	GGCGCAGCCCGCACCG
ZFP2	0.757	0.771	1382127	-7449	1382138	CAGCCCGCACCG
p54NRB	0.953	0.824	1382134	-7442	1382152	CACCGCGCAATCCCGGGGC
ZER6	0.799	0.752	1382135	-7441	1382146	ACCGCGCAATCC
Nanog	0.741	0.777	1382136	-7440	1382147	CCGCGCAATCCC
PRDM14	0.725	0.686	1382147	-7429	1382160	CGGGGCTCTGGCGG
CtBP1	0.964	0.883	1382150	-7426	1382159	GGCTCTGGCG
KLF17	1	0.853	1382154	-7422	1382160	CTGGCGG
ZNF341	0.8	0.845	1382154	-7422	1382165	CTGGCGGGAGGG
Egr-2	0.928	0.939	1382156	-7420	1382165	GGCGGGAGGG
ZBTB11	0.773	0.784	1382158	-7418	1382169	CGGGAGGGAAGG
BCL-11A	0.965	0.688	1382159	-7417	1382166	GGGAGGGA
FOXM1	0.73	0.736	1382165	-7411	1382175	GAAGGTGACTT
Nur77	0.796	0.832	1382165	-7411	1382175	GAAGGTGACTT
AHR	1	1	1382175	-7401	1382180	TGCGTG
ARNT	0.997	0.992	1382175	-7401	1382182	TGCGTGGG
Bach1	0.8	0.681	1382180	-7396	1382194	GGGACGACTCCAAGA

Factor name	Core Similarity Score	Matrix Similarity Score	Start Position	Relative start position	End Position	Sequence
ctcf	0.753	0.748	1382180	-7396	1382199	GGGACGACTCCAAGAGGCCA
XBP-1	0.816	0.738	1382182	-7394	1382192	GACGACTCCAA
GLIS1	0.597	0.779	1382214	-7362	1382229	AGTTCACCCACGAGGA
hdac2	0.833	0.82	1382215	-7361	1382223	GTTCACCCA
AML3	0.899	0.823	1382218	-7358	1382227	CACCCACGAG
ZNF143	1	0.643	1382218	-7358	1382233	CACCCACGAGGATACC
GATA-4	0.946	0.911	1382227	-7349	1382233	GGATACC
ZNF189	0.878	0.79	1382228	-7348	1382239	GATACCGCGACC
AML1	0.768	0.798	1382243	-7333	1382251	CCACCAAAG
E2F-6	0.876	0.873	1382245	-7331	1382252	ACCAAAGC
ZXDB	1	1	1382271	-7305	1382277	GACCCCT
ZNF600	0.595	0.768	1382273	-7303	1382284	CCCCTTTTCCCC
ZNF770	1	0.797	1382276	-7300	1382286	CTTTTCCCCCG
RBP-Jkappa	1	0.847	1382279	-7297	1382286	TTCCCCCG
Churchill	1	1	1382281	-7295	1382286	CCCCCG
Blimp-1	0.971	0.947	1382290	-7286	1382301	CCCAAGAGAAAA
Zic2	0.712	0.804	1382323	-7253	1382331	GAACTCCCC
c-Myc	0.8	0.804	1382353	-7223	1382364	TGGCACGGGGCA
AR	0.984	0.891	1382359	-7217	1382373	GGGGCAGGGTGTGCG
PRDM10	0.79	0.782	1382363	-7213	1382374	CAGGGTGTGCGG
EKLF	1	0.96	1382364	-7212	1382373	AGGGTGTGCG
Egr-1	0.942	0.913	1382380	-7196	1382389	GTGCGGGCGC
KLF16	0.908	0.873	1382382	-7194	1382392	GCGGGCGCGGG
ZNF449	1	0.703	1382382	-7194	1382393	GCGGGCGCGGGC
USF	1	0.982	1382396	-7180	1382409	GAGTCACGTGGAGG
c-Myc	1	0.902	1382397	-7179	1382408	AGTCACGTGGAG

Factor name	Core Similarity Score	Matrix Similarity Score	Start Position	Relative start position	End Position	Sequence
Ebox	1	0.998	1382397	-7179	1382406	AGTCACGTGG
MAX	1	0.95	1382399	-7177	1382410	TCACGTGGAGGC
SP2	1	0.921	1382415	-7161	1382429	GGAGAGGGCGAGGGG
GLIS2	1	0.795	1382428	-7148	1382444	GGGCCCCCGTAAATCAT
Nanog	0.728	0.791	1382436	-7140	1382447	GTAAATCATCCT
NF-1	1	0.983	1382455	-7121	1382472	TTTTGGCCCAAGTCCCTG
XBP-1	1	0.836	1382499	-7077	1382509	GACGTTGATTA
44107	0.923	0.817	1382505	-7071	1382517	GATTATGTAGCTG
CSX	1	0.993	1382524	-7052	1382534	TTTCAAGTGCG
MAZ	0.984	0.973	1382538	-7038	1382551	CCCTCCCCACCGCA
MZF-1	1	1	1382540	-7036	1382546	CTCCCCA
Smad1	0.968	0.933	1382547	-7029	1382555	CCGCAGCCC
AML3	1	0.9	1382600	-6976	1382609	GAAGTGGTGG
ZNF449	0.799	0.712	1382606	-6970	1382617	GTGGGTGATGGA
Prep-1	0.862	0.895	1382608	-6968	1382619	GGGTGATGGATT
CDP CR1	0.91	0.905	1382612	-6964	1382621	GATGGATTGT
Pbx	0.942	0.939	1382612	-6964	1382623	GATGGATTGTCT
HSF1	0.67	0.869	1382630	-6946	1382642	TTCTGGGACGTTC
GLI	1	0.982	1382642	-6934	1382651	CTGGGTGGGG
p53 decamer	0.88	0.912	1382705	-6871	1382714	AGACCAGCCC
Churchill	1	1	1382712	-6864	1382717	CCCCCG
ZF5	1	0.91	1382717	-6859	1382730	GCCGCGCAGCCCCG
AP-2alphaA	0.954	0.882	1382730	-6846	1382744	GAGGCCCCAGGCCTG
44107	0.903	0.871	1382801	-6775	1382813	GCAGATGGAGATT
FPM315 (ZNF263)	0.929	0.939	1382936	-6640	1382947	AGGGGAGCAGGA
GABP-beta	1	0.925	1382942	-6634	1382952	GCAGGAAGGGG

Factor name	Core Similarity Score	Matrix Similarity Score	Start Position	Relative start position	End Position	Sequence
XBP-1	0.8	0.803	1382962	-6614	1382972	GCGGACACGCC
AP-2gamma	0.976	0.978	1382970	-6606	1382979	GCCTCCAGCC
CTCF	0.917	0.834	1382984	-6592	1383003	TGCTCCCCTAGGAGGATCTC
NF-1A	1	1	1383058	-6518	1383064	CTTGGCA
Ikaros	1	1	1383123	-6453	1383129	TGGGAGG
REST	1	0.922	1383130	-6446	1383142	CAGGGGTGCTGGA
COUP-TF1	1	0.929	1383144	-6432	1383150	AGGGTCA
SP1	1	0.98	1383197	-6379	1383207	GCCACGCCCCA
AP-2alphaA	0.954	0.906	1383199	-6377	1383213	CACGCCCCAGGCCAG
AR	0.909	0.857	1383299	-6277	1383313	GGGGCCCTCCGTGCT
ZFP202	1	0.77	1383330	-6246	1383341	CTGGGTCCCCAC
CTCF	0.917	0.884	1383334	-6242	1383353	GTCCCCACGAGGAGGAGGCT
Nanog	0.966	0.866	1383377	-6199	1383388	GGAAATGAAGGG
AP-2gamma	0.982	0.99	1383432	-6144	1383441	GCCTGGGGGC
AP-2alphaA	0.954	0.903	1383521	-6055	1383535	TCGGCCTCGGGCGCC
ZF5	0.847	0.873	1383561	-6015	1383574	TCCGCGGAGGGCTT
P53	0.993	0.986	1383568	-6008	1383578	AGGGCTTGTGT
p53 decamer	1	0.851	1383569	-6007	1383578	GGGCTTGTGT
RNF96	1	0.989	1383646	-5930	1383655	GGCTGCGGGG
MAZ	1	0.978	1383651	-5925	1383664	CGGGGAGGGGTGGG
RREB-1	1	0.912	1383658	-5918	1383671	GGGTGGGGTTGGGG
MZF-1	1	1	1383674	-5902	1383680	CTCCCCA
CTCFL	0.813	0.62	1383677	-5899	1383688	CCCAAAGCAGCT
STAT3	0.683	0.673	1383682	-5894	1383702	AGCAGCTTCTCGGAGGCGGCG
PRDM14	0.967	0.848	1383715	-5861	1383728	GAGGTCCCTAGTCG
CREB1	1	0.987	1383793	-5783	1383804	TTGGACGTCAGA

Factor name	Core Similarity Score	Matrix Similarity Score	Start Position	Relative start position	End Position	Sequence
GABP-beta	1	0.918	1383910	-5666	1383920	CCAGGAAGTTC
44107	0.907	0.811	1383914	-5662	1383926	GAAGTTCATCTAT
FOXP1	0.872	0.853	1383979	-5597	1383993	ATTTATTTATTTGGG
STAT3	0.55	0.733	1383983	-5593	1384003	ATTTATTTGGGGGAAGGGAGG
BEN	1	0.982	1384005	-5571	1384012	CAGCGGGG
Tal-1 (Scl)	0.996	0.905	1384028	-5548	1384040	CGGGAGAAGAAAG
SOX	0.781	0.701	1384030	-5546	1384042	GGAGAAGAAAGCC
p53	0.944	0.902	1384039	-5537	1384058	AGCCCTGCCCGGGCTGGGTT
Zfx	1	0.967	1384127	-5449	1384142	CAGGCGCGGCCTGGCG
ERG	0.848	0.829	1384142	-5434	1384151	GACCTCCGGG
AP-2alphaA	1	0.906	1384150	-5426	1384164	GGCGCCTTTGGCTGG
MTF-1	1	0.993	1384195	-5381	1384204	CCGCGTGCAG
Nanog	0.741	0.789	1384197	-5379	1384208	GCGTGCAGTGCC
CTCF	0.745	0.824	1384211	-5365	1384230	GAGTGACCTCTTCAGGCCCC
COUP-TF1	1	0.994	1384214	-5362	1384220	TGACCTC
Muscle initiator	1	0.93	1384237	-5339	1384257	GCCTCCTGGGTGGGCCCCGCG
GLI	1	0.987	1384242	-5334	1384251	CTGGGTGGGC
RNF96	1	0.995	1384251	-5325	1384260	CCCCGCGGCC
MZF-1	1	1	1384304	-5272	1384310	CTCCCCA
FPM315 (ZNF263)	1	0.991	1384325	-5251	1384336	GCGGGAGGAGGG
Churchill	1	1	1384347	-5229	1384352	CGGGGG
RNF96	0.958	0.97	1384351	-5225	1384360	GGCCGCGGAG
PRDM14	0.967	0.817	1384372	-5204	1384385	CGGACCCCTAAACG
RFX	1	0.981	1384394	-5182	1384402	GCGTTGCCA
CTCFL	1	0.863	1384395	-5181	1384406	CGTTGCCACAGC
DEAF1	0.961	0.857	1384430	-5146	1384454	CAGAGCTCGGGGGTCTCCCAGGACG

Factor name	Core Similarity Score	Matrix Similarity Score	Start Position	Relative start position	End Position	Sequence
HSF1	0.655	0.753	1384526	-5050	1384538	TTCCAGCAGCATC
44107	0.952	0.834	1384529	-5047	1384541	CAGCAGCATCAGT
BBX	1	0.98	1384546	-5030	1384562	GAAAAGTTAACATGCTG
AML1	0.904	0.903	1384599	-4977	1384607	TTGCGGTAC
BEN	1	0.957	1384631	-4945	1384638	AGCCGCTG
Muscle initiator	1	0.908	1384631	-4945	1384651	AGCCGCTGGGTGCCCCCCCTT
Tal-1 (Scl)	0.996	0.947	1384695	-4881	1384707	GCCCCTTCTCTCT
RFX1	1	0.926	1384751	-4825	1384767	CCGCAACCCCGCAACTC
BEN	1	0.995	1384782	-4794	1384789	CAGCGGGC
Zfx	1	0.971	1384821	-4755	1384836	GTACCGCGGCCTGCGC
BEN	1	0.996	1384856	-4720	1384863	GGCCGCTG
ZF5	1	0.892	1384913	-4663	1384926	CCCGCGCCCGGCAA
Xvent-1	1	0.903	1384921	-4655	1384933	CGGCAATTTGGGA
SPI1	0.815	0.745	1384932	-4644	1384945	GAAGTGCGGAAATG
SREBP	1	0.981	1384941	-4635	1384955	AAATGGGGTGAGGCC
p53 decamer	0.81	0.853	1384951	-4625	1384960	AGGCCGGTCT
HSF1	0.627	0.749	1385000	-4576	1385012	TTCAAGCACCCTC
44107	0.907	0.812	1385083	-4493	1385095	TCTCATGAAGGTG
TATA	1	0.994	1385097	-4479	1385111	CTATAAAAAGAGCCC
PRDM14	1	0.825	1385140	-4436	1385153	GAAGACTCTATCCA
Tbx5	1	0.999	1385238	-4338	1385249	AGAGGTGTGAAA
RelA-p65	1	0.947	1385243	-4333	1385254	TGTGAAATTCCC
CTCF	0.908	0.837	1385247	-4329	1385266	AAATTCCCACTCCCGGCCAC
SMC-3	0.915	0.754	1385321	-4255	1385334	CATCTTCTGCTGTT
BEN	1	0.947	1385361	-4215	1385368	CAGCGTGC
GEN_INI	0.995	0.995	1385392	-4184	1385399	CCTCAATC

Factor name	Core Similarity Score	Matrix Similarity Score	Start Position	Relative start position	End Position	Sequence
Muscle initiator	1	0.911	1385398	-4178	1385418	TCACGCTGACACCCCCAGGCC
Kaiso	1	0.996	1385456	-4120	1385465	CTCGCAGGAG
POU6F1	1	0.964	1385717	-3859	1385733	CGCAATAATGAGGAGAT
Bcl-6	0.984	0.982	1385733	-3843	1385742	TTTCCTAGAA
STAT1	0.991	0.993	1385733	-3843	1385742	TTTCCTAGAA
STAT5B	0.964	0.977	1385733	-3843	1385743	TTTCCTAGAAT
Smad4	1	1	1386142	-3434	1386148	TGTCTGC
Ikaros	1	1	1386169	-3407	1386175	CCTCCCA
TTF-1	1	1	1386218	-3358	1386224	ACTCAAG
Pax	1	0.868	1386300	-3276	1386310	TGGAGTTCCAG
p53	0.969	0.874	1386334	-3242	1386353	TCACTTGCCCCAGCGTGACT
CSX	1	0.979	1386420	-3156	1386430	GCCACTTGGTG
USF	1	0.797	1386486	-3090	1386499	TGTAGACGTGGTGG
ARNT	1	0.967	1386490	-3086	1386497	GACGTGGT
MAFA	1	1	1386574	-3002	1386580	CTGCTGA
GATA	1	1	1386607	-2969	1386613	CTTATCT
CP2	1	0.997	1386670	-2906	1386679	TTCCAGTCAG
RBP-Jkappa	1	0.974	1386722	-2854	1386732	CTTCCCACGAC
FAC1	0.956	0.937	1386742	-2834	1386755	CCTCATAACACTCA
PLZF	0.978	0.8	1386785	-2791	1386813	TTAGAGTTTAGGTAAAATGGATACTGCTG
ATF-2	0.831	0.891	1386792	-2784	1386799	TTAGGTAA
YY1	1	1	1386798	-2778	1386804	AAAATGG
HMGIY	0.96	0.934	1386839	-2737	1386853	TTAAAAAAATCCCTC
POU6F1	1	0.909	1386919	-2657	1386935	TGTTTTAATGAGTACCT
ARNT	0.783	0.773	1387061	-2515	1387068	ATCACTGG
Irx2	0.791	0.841	1387173	-2403	1387189	CAAATACATATAATTTT

Factor name	Core Similarity Score	Matrix Similarity Score	Start Position	Relative start position	End Position	Sequence
RFX	1	0.988	1387254	-2322	1387262	CCGTTGCCC
Egr-1	1	0.977	1387264	-2312	1387273	CCGCCCCCAC
DBP	0.985	0.987	1387336	-2240	1387342	AGCACAC
ZSCAN4	1	0.924	1387337	-2239	1387349	GCACACACACACA
MZF-1	1	1	1387415	-2161	1387421	CTCCCCA
POU2F1	0.959	0.973	1387436	-2140	1387447	TTATGAAAATAA
GEN_INI	0.996	0.997	1387544	-2032	1387551	AAATGAGG
HMGIY	0.96	0.939	1387557	-2019	1387571	AGAAAAAAATACCAT
GLIS1	1	0.59	1387587	-1989	1387602	GCACCCCCCAAAAACG
MAFA	1	1	1387604	-1972	1387610	TCAGCAG
RREB-1	1	0.845	1387655	-1921	1387668	CCCCAACCCAATCA
BRCA1:USF2	0.997	0.996	1387658	-1918	1387665	CAACCCAA
E2F	1	0.924	1387738	-1838	1387749	GTTTGGCGGTGT
YY1	1	1	1387780	-1796	1387786	CCATTTT
Dlx-3	1	1	1387868	-1708	1387875	CTAATTAC
ZNF596	0.999	0.918	1387921	-1655	1387935	ACTCGGTCTCTGTGT
ER-alpha	1	1	1387985	-1591	1387991	CAGGTCA
Zic2	0.903	0.875	1388014	-1562	1388022	CACCCCCCT
WT1	0.996	0.996	1388016	-1560	1388027	CCCCCCTCCCAC
Ikaros	1	1	1388020	-1556	1388026	CCTCCCA
MAZ	1	0.994	1388032	-1544	1388045	CCCTCCCCTCCCCA
MAZR	0.935	0.935	1388033	-1543	1388045	CCTCCCCTCCCCA
MZF-1	1	1	1388039	-1537	1388045	CTCCCCA
Osx	1	0.996	1388044	-1532	1388053	CACCCGCCCT
OSR2	0.797	0.714	1388046	-1530	1388061	CCCGCCCTCTCTGCAT
KLF17	1	0.998	1388047	-1529	1388053	CCGCCCT

Factor name	Core Similarity Score	Matrix Similarity Score	Start Position	Relative start position	End Position	Sequence
DBP	0.985	0.987	1388075	-1501	1388081	AGCACAC
ZSCAN4	1	0.938	1388076	-1500	1388088	GCACACTCAGCCA
Muscle initiator	0.985	0.933	1388095	-1481	1388115	GTCCCCCGGCACTCCAGGGGC
Churchill	1	1	1388097	-1479	1388102	CCCCCG
Kox1	1	0.86	1388100	-1476	1388111	CCGGCACTCCAG
znf580	1	0.837	1388126	-1450	1388136	CCTGCCGCCCC
GLIS1	0.801	0.735	1388128	-1448	1388143	TGCCGCCCCAGGGGGC
ctcf	1	0.807	1388129	-1447	1388148	GCCGCCCCAGGGGGCCTGCC
Roaz	0.8	0.764	1388130	-1446	1388143	CCGCCCCAGGGGGC
GLIS2	1	0.672	1388140	-1436	1388156	GGGCCTGCCTAGGCTGT
p53 decamer	0.907	0.932	1388140	-1436	1388149	GGGCCTGCCT
ZNF121	1	0.759	1388157	-1419	1388168	GCGGCTGTGAAG
ZNF768	0.623	0.678	1388169	-1407	1388180	GCTCTTTGCCTG
sin3A	0.693	0.74	1388215	-1361	1388228	TCTCCTGCCTCCTG
ZNF76	1	0.816	1388216	-1360	1388234	CTCCTGCCTCCTGCAACTC
Zipro1	1	0.913	1388221	-1355	1388231	GCCTCCTGCAA
44107	0.858	0.811	1388222	-1354	1388234	CCTCCTGCAACTC
Pax	0.818	0.848	1388226	-1350	1388236	CTGCAACTCCC
MZF-1	1	1	1388232	-1344	1388238	CTCCCCA
ZNF143	0.932	0.753	1388242	-1334	1388257	TCCCCAGCAGGCACAC
44166	0.845	0.847	1388253	-1323	1388263	CACACCCGTGC
PRDM10	0.62	0.697	1388291	-1285	1388302	CGGGGTGCCCAG
ZXDB	1	0.953	1388291	-1285	1388297	CGGGGTG
PRDM14	1	0.729	1388298	-1278	1388311	CCCAGCTCTACCTC
ZBTB20	0.745	0.863	1388299	-1277	1388310	CCAGCTCTACCT
ZID	0.896	0.846	1388300	-1276	1388312	CAGCTCTACCTCC

Factor name	Core Similarity Score	Matrix Similarity Score	Start Position	Relative start position	End Position	Sequence
BEN	1	0.996	1388319	-1257	1388326	GTCCGCTG
ZFP2	0.506	0.706	1388328	-1248	1388339	TCTGCAGCACCT
REST	1	0.867	1388330	-1246	1388342	TGCAGCACCTTCT
ZNF143	0.933	0.652	1388336	-1240	1388351	ACCTTCTTTCTGGGCC
ZNF366	0.6	0.702	1388337	-1239	1388350	CCTTCTTTCTGGGC
CTCF	1	0.819	1388348	-1228	1388367	GGCCTCCCCCTTCAGCCCTT
ZNF341	1	0.748	1388351	-1225	1388362	CTCCCCCTTCAG
Ikaros	1	1	1388367	-1209	1388373	TGGGAGG
KLF7	0.8	0.856	1388373	-1203	1388382	GCCCCGCCAG
Osx	0.8	0.785	1388373	-1203	1388382	GCCCCGCCAG
KLF17	1	0.853	1388376	-1200	1388382	CCGCCAG
MAZ	1	0.899	1388394	-1182	1388407	TGCTCTCCTCCTCC
FPM315 (ZNF263)	1	0.997	1388399	-1177	1388410	TCCTCCTCCCTC
SP2	0.94	0.901	1388399	-1177	1388413	TCCTCCTCCCTCTCC
WT1	0.996	0.968	1388403	-1173	1388414	CCTCCCTCTCCC
Aiolos	1	0.973	1388441	-1135	1388449	TTCCCTGCA
c-Myc	1	0.831	1388445	-1131	1388456	CTGCACGTGTAA
ZNF2	0.877	0.719	1388450	-1126	1388461	CGTGTAATCCAA
Gfi1b	0.977	0.937	1388452	-1124	1388461	TGTAATCCAA
ZNF643	0.873	0.786	1388455	-1121	1388466	AATCCAATAACT
ZNF155	0.472	0.68	1388461	-1115	1388472	ATAACTAACTGT
c-Myb	1	0.992	1388464	-1112	1388475	ACTAACTGTCGG
ZNF524	0.754	0.733	1388491	-1085	1388504	ATTGCTCCCGAGGG
ZNF837	0.772	0.793	1388527	-1049	1388538	TCTTACTTAGGG
ATF-2	0.831	0.765	1388529	-1047	1388536	TTACTTAG
Insm2	0.971	0.792	1388540	-1036	1388550	GGAATTAGCCA

Factor name	Core Similarity Score	Matrix Similarity Score	Start Position	Relative start position	End Position	Sequence
MZF-1	1	0.965	1388554	-1022	1388563	CTGGGGAGAA
MZF-1	1	1	1388555	-1021	1388561	TGGGGAG
AhR	0.821	0.82	1388600	-976	1388616	GTTTTGGGTGGGCAATG
ZER6	0.799	0.863	1388600	-976	1388611	GTTTTGGGTGGG
GLI	1	0.984	1388603	-973	1388612	TTGGGTGGGC
EKLF	1	0.955	1388604	-972	1388613	TGGGTGGGCA
ZNF770	0.79	0.862	1388606	-970	1388616	GGTGGGCAATG
MEIS1	1	0.998	1388638	-938	1388649	TAGTGACAGCTG
Prep-1	1	0.922	1388638	-938	1388649	TAGTGACAGCTG
ZNF600	0.61	0.707	1388639	-937	1388650	AGTGACAGCTGC
MyoD	1	1	1388641	-935	1388649	TGACAGCTG
KLF8	1	0.978	1388649	-927	1388658	GCACACCCCT
ZNF629	1	0.791	1388686	-890	1388697	TCGCTTATCTCA
GATA	1	1	1388689	-887	1388695	CTTATCT
Blimp-1	1	0.992	1388712	-864	1388723	CTTTCTCTTTCT
Churchill	1	1	1388787	-789	1388792	CGGGGG
COE1	0.983	0.976	1388824	-752	1388837	TAAGTCCCATGAGA
HSF1	0.914	0.756	1388845	-731	1388857	TTCCAGAGTCTTT
STAT1	0.986	0.959	1388901	-675	1388910	GTTCCGAAAA
XBP-1	1	0.81	1388904	-672	1388914	CCGAAAACGTC
BEN	1	0.996	1388924	-652	1388931	CAGCGGCC
GLIS2	1	0.699	1388934	-642	1388950	CAACCCTGCGGCGGCCC
PRDM14	0.967	0.765	1388953	-623	1388966	AGTCGCCCTAACAA
ZNF366	0.789	0.706	1388961	-615	1388974	TAACAATTGTCGGG
NF-1	1	0.987	1389005	-571	1389022	GATTGGCACGTTACCTTT
Nanog	0.704	0.778	1389008	-568	1389019	TGGCACGTTACC

Factor name	Core Similarity Score	Matrix Similarity Score	Start Position	Relative start position	End Position	Sequence
Muscle initiator	1	0.918	1389024	-552	1389044	GGCGTCTGGCACCCCGGCGTG
Roaz	1	0.838	1389032	-544	1389045	GCACCCCGGCGTGT
KLF9	1	0.864	1389037	-539	1389049	CCGGCGTGTGGCT
RNF96	0.835	0.896	1389046	-530	1389055	GGCTGAGGGC
MZF-1	1	1	1389100	-476	1389106	TGGGGAG
REST	1	0.858	1389141	-435	1389153	AGCAGCACCACTC
ZNF547	0.862	0.763	1389157	-419	1389168	CCGGCGGGCCAG
CTCF	0.917	0.89	1389161	-415	1389180	CGGGCCAGGAGGAGGAAGGG
KLF16	1	0.923	1389178	-398	1389188	GGGGGCGCGGA
SP2	1	0.982	1389185	-391	1389199	CGGAGGGGCGGGGCT
MAZR	0.918	0.933	1389188	-388	1389200	AGGGGCGGGGCTG
SP1	1	0.981	1389188	-388	1389198	AGGGGCGGGGC
KLF7	1	0.977	1389189	-387	1389198	GGGGCGGGGC
Osx	1	0.913	1389189	-387	1389198	GGGGCGGGGC
Sp1	1	1	1389189	-387	1389198	GGGGCGGGGC
Sp3	1	0.997	1389189	-387	1389200	GGGGCGGGGCTG
Churchill	1	1	1389230	-346	1389235	CGGGGG
ZBTB7A	1	0.96	1389265	-311	1389276	GAGGTGGTCGAG
Zic2	1	0.909	1389265	-311	1389273	GAGGTGGTC
ZER6	0.599	0.753	1389276	-300	1389287	GTTTTGCGAGGG
KLF8	1	0.973	1389284	-292	1389293	AGGGGTGGGG
EKLF	1	0.988	1389285	-291	1389294	GGGGTGGGGG
ZXDB	1	0.998	1389290	-286	1389296	GGGGGTC
SIP1	0.8	0.806	1389295	-281	1389309	TCGTCCCCCTGGCAG
ZBTB20	0.522	0.731	1389304	-272	1389315	TGGCAGAACCGG
ZNF189	1	0.809	1389327	-249	1389338	GATCCCTCAATT

Factor name	Core Similarity Score	Matrix Similarity Score	Start Position	Relative start position	End Position	Sequence
ZBTB26	0.776	0.703	1389333	-243	1389349	TCAATTGGAAAGAGGGT
GEN_INI	0.993	0.994	1389350	-226	1389357	GAGTGAGG
ZFP2	0.826	0.823	1389350	-226	1389361	GAGTGAGGGAAA
GLIS1	0.401	0.65	1389367	-209	1389382	GGAGGCCGCGAGAGGA
ZNF643	0.843	0.789	1389411	-165	1389422	GGGAGCGAGATA
GATA-1	1	0.995	1389416	-160	1389425	CGAGATAAGA
GATA	1	1	1389418	-158	1389424	AGATAAG
GATA-6	1	1	1389418	-158	1389424	AGATAAG
MAX	0.915	0.749	1389436	-140	1389447	GCCGCCAGGTCG
Egr-1	1	0.924	1389449	-127	1389458	GAGGGGGAGG
CTCF	1	0.827	1389476	-100	1389495	GGAGCGAGGAGGGGGCAGGG
RREB-1	0.901	0.848	1389492	-84	1389505	AGGGCGGGGTGGGG
Sp1	1	0.973	1389492	-84	1389501	AGGGCGGGGT
WT1	1	0.994	1389496	-80	1389507	CGGGGTGGGGGG
MAZ	0.981	0.976	1389500	-76	1389513	GTGGGGGGGGAGGG
ZNF341	1	0.902	1389500	-76	1389511	GTGGGGGGGGAG
PRDM10	0.839	0.697	1389532	-44	1389543	GAGGGCGGCCGG
Ets	1	1	1389559	-17	1389566	GAGGAAGT
TFIIB	0.659	0.767	1389563	-13	1389573	AAGTCCGAGAG
FOXP1	0.782	0.683	1389565	-11	1389579	GTCCGAGAGACAGAG
ZNF596	0.999	0.956	1389566	-10	1389580	TCCGAGAGACAGAGC
HSF1	0.816	0.728	1389584	8	1389596	GAGCGCTCCTGAA
Zfx	1	0.984	1389622	46	1389637	GTCCAGGCCGCGGTCC
AML3	0.899	0.805	1389634	58	1389643	GTCCCACTCG
RNF96	1	0.995	1389675	99	1389684	CCCCGCGGCC
Ikaros	1	1	1389685	109	1389691	TGGGAGG

Factor name	Core Similarity Score	Matrix Similarity Score	Start Position	Relative start position	End Position	Sequence
CTCFL	1	0.703	1389693	117	1389704	CGTTGCGCAAGG
C/EBPbeta	0.995	0.991	1389694	118	1389704	GTTGCGCAAGG
C/EBPbeta	0.995	0.99	1389694	118	1389704	GTTGCGCAAGG
C/EBPdelta	0.986	0.988	1389695	119	1389704	TTGCGCAAGG
p53 decamer	0.907	0.926	1389702	126	1389711	AGGCAGGGCC
ELK1	0.788	0.734	1389768	192	1389777	GCCAGAAGCC
BCL-11A	0.965	0.737	1389787	211	1389794	CTGAGGGA
SP100	0.942	0.911	1389795	219	1389809	GGCGACGCCGAAGCG
NRF-1	0.999	0.874	1389888	312	1389898	GCGCCTCCGCT
ERG	0.822	0.777	1389895	319	1389904	CGCTTCCCGC
AML1	0.822	0.845	1389949	373	1389957	TGACCACCG
c-Myc	0.8	0.726	1390076	500	1390087	GACCACCTCCTC
CTCF	0.908	0.855	1390191	615	1390210	GGCGCCGGGAGCGGGGGCGC
Egr-1	1	0.977	1390201	625	1390210	GCGGGGGCGC
Churchill	1	1	1390202	626	1390207	CGGGGG
E2F	1	0.926	1390204	628	1390215	GGGGCGCCAAGA
E2F-1	0.957	0.881	1390204	628	1390219	GGGGCGCCAAGAAGGC
ZFX	1	0.933	1390225	649	1390236	CGGGGCTGCGGC
44107	0.81	0.804	1390271	695	1390283	CATCGTCATGGCC
BEN	1	0.989	1390316	740	1390323	CAGCGAGA
GRHL2	1	0.85	1390322	746	1390336	GATCTACCAGTTCCT
ESE-1	1	0.971	1390329	753	1390337	CAGTTCCTG
AP-2gamma	0.993	0.987	1390333	757	1390342	TCCTGCAGGC
ZFHX2	0.958	0.906	1390342	766	1390352	CGCGCTTCCCC
ZNF600	1	0.792	1390342	766	1390353	CGCGCTTCCCCT
SPI1	1	0.799	1390344	768	1390357	CGCTTCCCCTTCTT

Factor name	Core Similarity Score	Matrix Similarity Score	Start Position	Relative start position	End Position	Sequence
Elf-1	0.882	0.879	1390345	769	1390356	GCTTCCCCTTCT
SP100	0.933	0.904	1390353	777	1390367	TTCTTCCGCGGCGCC
GLIS2	0.62	0.685	1390388	812	1390404	GGTGCGCCACAATCTCT
BCL-11A	0.778	0.796	1390429	853	1390436	TGCCTAAG
AP-2alphaA	0.954	0.92	1390435	859	1390449	AGGGCCTCGGGCGGC
p53 decamer	0.977	0.889	1390451	875	1390460	CGGCAAGGGC
Nanog	0.762	0.81	1390452	876	1390463	GGCAAGGGCCAC
PRDM14	0.926	0.764	1390456	880	1390469	AGGGCCACTACTGG
BCL-6	0.963	0.897	1390496	920	1390505	GTTCGAGGAG
Elf-1	0.863	0.832	1390534	958	1390545	TCAGGCGGAAGT
ERG	1	0.86	1390537	961	1390546	GGCGGAAGTG
FLI-1	1	0.939	1390537	961	1390546	GGCGGAAGTG
GABP-beta	1	0.947	1390537	961	1390547	GGCGGAAGTGC
Pax	0.883	0.862	1390540	964	1390550	GGAAGTGCCAG
ZNF561	0.432	0.529	1390559	983	1390570	GCCCATGTACCA
CTCF	0.959	0.806	1390563	987	1390582	ATGTACCACCGCGTGGTGAG
c-Myc	0.8	0.841	1390569	993	1390580	CACCGCGTGGTG

TRANSFAC experimental binding sites

Table 6 Experimentally proven TFs of FOXF2 gene

Binding Factor(s)	Binding Site	Start Position	End Position
Sp1(h)	R68293	1389179	1389193
Sp1(h)	R68294	1389492	1389501