Molecular mechanism(s) of regulation of bidirectional promoters

A thesis

Submitted in partial fulfilment of the requirements

Of the degree of

Doctor of Philosophy

by

Rahul Kumar Jangid

(IISER ID No. 20103091)

Under the guidance of

Prof. Sanjeev Galande

at

INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH, PUNE

February 2016

Dedicated to my family…

Declaration

I declare that this written submission represents my ideas in my own words and where others' ideas have been included; I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Rahul Kumar Jangid

(Roll No. 20103091)

Date:

CERTIFICATE

Certified that the work incorporated in the thesis entitled "Molecular mechanism(s) of regulation of bidirectional promoters" Submitted by Rahul Kumar Jangid was carried out by the candidate, under my supervision. The work presented here or any part of it has not been included in any other thesis submitted previously for the award of any degree or diploma from any other University or institution.

Date: Prof. Sanjeev Galande

(Supervisor)

ACKNOWLEDGEMENTS

This thesis is a result of almost 6 years of work done during my Ph.D, which has been a wonderful and overwhelming experience. During these years, I have worked with a large number of people whose contributions in many different ways have made this thesis possible. I am grateful to many people for this journey of my Ph.D. It has been an unforgettable experience and it gives me pleasure to convey my gratitude towards them in form of this humble acknowledgment.

First of all, I am forever indebted to my supervisor, Prof. Sanjeev Galande, for his enthusiasm, guidance, and support throughout my Ph.D. He has gone beyond his duties to solve my worries, concerns and anxieties; and has worked to instill great confidence in me at both personal and professional levels. He is as cool and calm as anyone can be at the best level. I sincerely thank him for being extremely supportive during the tough times of my Ph.D. and for his encouragement to get past them.

I would like to express my gratitude to Prof. G. C. Mishra, ex-director, National Centre for Cell Science (NCCS) where I worked during my initial days of Ph.D. before moving to Indian Institute of Science Education and Research (IISER). I am deeply thankful to Prof. K. N. Ganesh; director, Prof. L. S. Shashidhara; head of biology division, and Dr. V. S. Rao, ex-registrar at IISER for providing an opportunity to register at IISER Pune for my Ph.D.

I would also like to thank members of my research advisory committee, Dr. Aurnab Ghose from IISER Pune and Dr. Vasudevan Seshadri from NCCS, for their support.

I would like to extend my sincere thanks to all the staff from academic section, computer section, accounts section, library, stores and purchase section of IISER Pune for their help and cooperation. It is my pleasure to acknowledge the funding sources that supported me during the tenure of my Ph.D. work. I wish to thank the Council of Scientific and Industrial Research (CSIR) for providing me the fellowship during my Ph.D.

I would like to express my sincere gratitude to IISER Pune for providing a healthy and exciting environment to do research. I would like to thank all the members of the biology department for their suggestions during my annual work presentations. I would like to thank Shabnam, Piyush and Kalpesh for maintaining the biology labs and instruments. It is their continuous hard work which enables all the research students to work smoothly. I would like to thank Vijay for maintaining the microscopy facility and for the initial training. I would like to express my sincere thanks to Miss Mrinalini Virkar for taking care of all the consumables orders. I would like to thank Mr. Prabhas, Miss Nayana, Mr Anil Jadhav and Mr Mahesh Rote for their help and support during my stay at IISER Pune. I am also grateful to all the security and housekeeping staff for their hard work round the clock to make our work does not suffer.

I am thankful to Dr. Krishanpal Karmodiya for timely discussions on my progress and also for sparing different antibodies and plasmid constructs which helped me to complete the second part of my thesis. My thanks are due to Dr. Vijay Kumar Muley for his help in ChIP-seq analysis; without his help I would not have been able to work on my hypothesis regarding involvement of specific epigenetic mechanisms in regulation of bidirectional promoter.

I consider myself very lucky to work with such wonderful people in SG lab throughout my PhD. I would like to thank each of those people who are directly or indirectly helped me during this period. I thank Pavan and Prabhat for spending some quality time with me while discussing about science during their visits to IISER. I wish to thank Amita and Sameet for their help in my initial days to PhD. My special thanks are due to Ranveer for being an exceptional person in many ways. His forever helping attitude sets him apart from other people. Praveena has been an amazing senior and a great human being during all these years in SG lab. I would like to sincerely thank her for all the help and guidance that she has given me during my Ph.D.

My special thanks are due to the pair of seniors (Kamal and Sunita) who joined SG lab on the same day. You both have played such an important part in my initial days at NCCS and IISER. The best time I had in my PhD was the time when you both were around. I thank you from the bottom of my heart for all your caring and invaluable support.

As I thank all my colleagues and seniors for their help, I find myself falling short of words while expressing my gratitude to a very important

person in my life, Sunita Singh. She is someone who has constantly encouraged and supported me through all my high and low times whether it was professionally or personally. Her contribution towards my journey has been indispensable, as a friend and a well-wisher. A formal written quote of thanks for her will never be able to define how much I owe to her.

I would like to thank Rafeeq for his help and support as well as for many useful discussions in lab. He is one of the most hardworking and dedicated person I have seen during my PhD. I thank Rini for her help and support in many aspects of my Ph.D. I am very thankful to Ankitha Shetty for helping me enormously at different times when I needed it most. I thank Indumathi and Divya Singh for taking care of my knockout cell line cultures when I was away or not keeping well. I appreciate Manu for his help in all computer and software related things. I thank Suyog for patiently listening to me, offering advice, and supporting me during this period. I would like to thank Saurabh and Amol for helping me with RNA seq experiment.

I wish to thank all the present and past SG lab members including Manu, Indumathi, Akhila, Ankita Sharma, Ankitha Shetty, Pallavi, Mouli, Ashwin, Satyjeet, Ameya, Suyog, Saurabh, Deepeshwari, Pavana, Neha Oli, Ishwar, Shubham, Geeta, Poornima, Bhagyashree Joshi, Rahul Sharma, Farhat, Amol, Nethra, Ram, Rachel, Soumitra, Vijay, Meghana, Georgina, Neha Sirwani, Priyanka, Shiny, Mithila, Madhujit, Amruth and Suman. I thank Divya Singh for her help in my project during her short stay at IISER Pune.

I would also like to all the present and past Krishanpal lab members namely, Mukul, Pratima, Madhavi, Abhishek. Swati and Bhagyashree Joshi.

My special thanks are due to Soumitra Athavale, a more Brilliant scientific brain I am yet to meet. 'Thank you' words are too small for all the help you provided me scientifically as well as non-scientifically.

I thank Manjunath for his help and suggestions in different aspects during my Ph.D. I wish to thank all the current post-doctoral fellows in SG lab including Mouli, Ameya, Satyajeet and Ashwin. I wish to thank Sonam Mehrotra for her suggestion and help. I thank Ashwin for giving his valuable time to discuss about my projects. I am thankful to Mouli for helping me with preparation of few figures for my presentations and thesis.

I thank Santosh for all the maintenance work he has been doing in the lab. Special thanks are due to our former lab manager Vipradas Sir who is always ready to help everyone in whatever capacity he can, both at professional and personal level. With his affection, care and love many times he made me feel home.

'Thank you' seems very small for my friends (Sandeep, Manish, Sanket, Ajay Labde, Mallu Chenna Reddy, Bijoy Mishra and Anupam Sawant, JP) who were always just a call away when I needed them at the time of difficulties.

I would like to thank my parents for their infinite love, encouragement and constant support throughout everything. I especially

wish to thank my father for keeping faith in me and allowing me to chase my dreams. Only because of his struggles and constant support I could reach to this stage.

Finally, I would like to thank everybody who helped me in one or the other way during some stage of my Ph.D. life as well as express my apology if I could not mention them personally here.

Last but not the least, I am deeply thankful to the one above all of us, the omnipresent God, for answering my prayers and for giving me the strength to carry on despite my constitution wanting to give up, thank you so much Dear Lord Shiva.

-Rahul Jangid

Table of Contents

List of Figures

List of Tables

List of Appendices

Abbreviations

ATP	Adenosine Triphosphate
Bp	Base pair
BRE	TFIIB recognition element
BDP	Bidirectional promoter
BDP S1	Bidirectional promoter clone 1 in sense orientation
BDP AS1	Bidirectional promoter clone 1 in antisense orientation
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeat
CAS9	CRISPR associated protein 9
CMV promoter	Cytomegalovirus (CMV)
ChIP	Chromatin Immunoprecipitation
ChIP-seq	ChIP-Sequencing
Cdk 9	Cyclin-Dependent Kinase 9
CTD	C terminal domain
CTD S2	C terminal domain serine residue 2
DPE	Downstream promoter element
DNA	Deoxyribonucleic Acid (DNA)
DTT	Dithiothreitol (DTT)
EGFP	Enhanced Green Fluorescent Protein
EDTA	Ethylenediaminetetraacetic acid
FACS	Flow cytometry
FBS	Fetal Bovine Serum
GTFs	General transcription factors
GABPA	(GA Binding Protein Transcription Factor, Alpha Subunit
H1	Histone H1
H2B	Histone H2A
H2A	Histone H2B
H3	Histone H3
H4	Histone H4
Hg19	Human genome assembly 19
H3K4me1	Histone H3 lysine 4 monomethyaltion
H3K4me2	Histone H3 lysine 4 dimethyaltion
H3K4me3	Histone H3 lysine 4 trimethyaltion
H2BK5me1	Histone H2B lysine 5 monomethyaltion
H3K27me1	Histone H3 lysine 27 monomethyaltion
H3K27me2	Histone H3 lysine 27 dimethyaltion
H3K27me3	Histone H3 lysine 27 trimethyaltion
H3K9ac	Histone H3 lysine 9 acetylation
H3K36me1	Histone H3 lysine 36 monomethyaltion
H3K36me3	Histone H3 lysine 36 trimethyaltion
H3K79me1	Histone H3 lysine 79 monomethyaltion

H3K79me2	Histone H3 lysine 79 dimethyaltion
H3K79me3	Histone H3 lysine 79 trimethyaltion
H3K9me1	Histone H3 lysine 9 monomethyaltion
H3K9me3	Histone H3 lysine 9 trimethyaltion
H4K20me1	Histone H4 lysine 20 monomethyaltion
H4K20me3	Histone H4 lysine 20 trimethyaltion
Inr	Initiator element
kDa	Kilo Dalton
Kb	Kilo basepair
KD	Knock down
KO	Knock out
µg	Microgram
µl	Microliter
NRF1	Nuclear respiratory factor-1
NaCl	Sodium Chloride
NaHCO3	Sodium bicarbonate
NFYA	Nuclear Transcription Factor Y, Alpha
ORF	Open reading frame
PDR	Plasmid with dual reporter
PCR	Polymerase chain reaction
qPCR	Quantitative polymerase chain reaction
RNA Pol ll	RNA polymerase II (RNAP II and Pol II)
RPMI	Roswell Park Memorial Institute medium
RA	Retinoic acid
RNA	Ribonucleic acid (RNA)
si-RNA	Small interfering RNA
sh-RNA	Small hairpin RNA or short hairpin RNA (shRNA) i
SP1	Specificity Protein 1
SAGA	Spt-Ada-Gcn5-Acetyl transferase
SDS	Sodium dodecyl sulfate
TBP	TATA binding protein
TAFs	TBP associated factors
TFIID	Transcription factor ll D
TFIIB	Transcription factor ll B
TFs	Transcription factors
TSS	Transcription start site
UNI P	Unidirectional promoter
WB	western blot
YY1	yin-yang 1 (Transcription Factors)
ZNF143	Zinc Finger Protein 143

Synopsis

Introduction

After the completion of the human genome sequence assembly, it became apparent that genes are arranged in three different modes on chromosomes; head-to-head, head-to-tail and tail-to-tail. In head-to-head gene arrangement, two genes are encoded on opposite strands of DNA and are regulated by a common intergenic region between the TSS (transcription start site) of both genes whereas, in head-to-tail gene arrangement, one gene is followed by the other gene and each gene has its own upstream regulatory region. In tail-to-tail gene arrangement, two genes are oriented towards each other. Upstream DNA elements of transcription start sites (TSSs) act as putative promoter element for head-to-tail and tail-to-tail arranged genes. These promoters act like conventional unidirectional promoters which regulate the transcription of downstream genes. In contrast, if two genes are arranged in head-to-head orientation, then the common DNA element between the two genes acts as promoter for both the genes; therefore is referred to as a bidirectional promoter. A "bidirectional gene pair" is defined as a pair of two adjacent genes located on opposite strands of DNA with their transcription start sites (TSSs) not more than 1 Kilo base pairs apart (Trinklein et al., 2004). The intergenic region between the two TSSs of a bidirectional gene pair is commonly designated as a putative "bidirectional promoter". Despite the large size of the human genome, a substantial proportion of genes (11%) is arranged in divergent, head-to- head fashion and is regulated by bidirectional promoters (Trinklein et al., 2004). This abundance of divergent arrangement of gene pairs has been observed across several mammalian genomes, suggesting an evolutionary pressure for conserving this type of gene pair arrangement, also known as bidirectional gene pair (Adachi and Lieber, 2002; Koyanagi et al., 2005a).

Bidirectional promoters orchestrate transcription of gene pairs whose expression levels are regulated in a coordinated manner for proper regulation of processes key to the survival of an organism.. Few bidirectional promoters serve to maintain a proper

stoichiometric relationship of gene expression; such as the histone genes (Ahn and Gruen, 1999; Albig et al., 1997; Maxson et al., 1983), whereas others regulate co-expression of the gene pairs which are part of a common biological pathway (Momota et al., 1998; Schmidt et al., 1993; Sugimoto et al., 1994). Analyses of occurrence of bidirectional promoters in human, mouse, rat, chicken, and fish genomes revealed that this kind of gene arrangement is conserved across different eukaryotic species (Li et al., 2006).

Several studies have analyzed bidirectional promoters to study if there is any correlation between the promoter sequence and transcription regulation from these promoters (Lin et al., 2007a). It was shown that most of the bidirectional promoters lack TATA boxes and are enriched in CpG islands (Adachi and Lieber, 2002; Takai and Jones, 2004). TATA motif enrichment study revealed that only 29% of non-bidirectional promoters harbor TATA box whereas only 9% of the bidirectional promoters show enrichment of TATA motif, suggesting that TATA motif is underrepresented on bidirectional promoters (Yang and Elnitski, 2008b). Downstream promoter element (DPE) shows enrichment on 46.5% and 50.6% of bidirectional and non-bidirectional promoters respectively. Initiator (INR) elements also show enrichment on 25.3% bidirectional promoters and 30.8% non-bidirectional promoters. Thus, DPE and INR promoter elements do not show any biased enrichment on bidirectional promoters. However, bidirectional promoters exhibit differential enrichment of CpG islands as compared to the non-bidirectional promoters. Nearly 90% of bidirectional promoters contain CpG islands as compared to 45% of non-bidirectional promoters (Yang and Elnitski, 2008a). Higher enrichment of CpG islands has been correlated with some important features of bidirectional promoters such as high basal level of transcription and higher RNA Pol II occupancy (Barski et al., 2007a). Presence of different promoter sequence elements on uni and bidirectional promoters is not sufficient enough to explain the functional difference between these two types of promoters. There are specific transcription factors which have been suggested to regulate the transcription from bidirectional promoters. Regarding this, several studies have focused on the regulatory mechanisms which govern transcription from these loci but detailed mechanistic insights are still lacking (Lin et al., 2007a; Trinklein et al., 2004). Studies have been performed to align the

bidirectional promoter regions from human genome to identify if these promoters are enriched for any particular motif as compared to the unidirectional promoters and categorized the enriched sites as underrepresented, shared and over-represented (Lin et al., 2007b). It was found that most of the common eukaryotic transcription factors lack binding sites on the bidirectional promotersand only a small set of motifs exhibit over represented enrichment on bidirectional promoters which include GABPA, MYC, E2F1, E2F4, NRF-1, CCAAT and YY1.

A recent study focused on differential distribution of histone modifications on unidirectional and bidirectional genes revealed that H3K4me3, H3K9ac, and H3K27ac are enriched more in antisense direction on bidirectional genes as compared to unidirectional genes. However no differential enrichment was observed in sense orientation (Bornelöv et al., 2015). These authors have also commented that antisense enrichment of these active transcription marks might be a consequence of antisense transcription rather than cause (Bornelöv et al., 2015). By using murine macrophage as model system another recent study showed that bidirectional transcription arises from two distinct hubs of transcription factor binding (Scruggs et al., 2015). A study focused on the nature of bidirectional promoter regulated genes revealed that most of the bidirectional promoter-driven genes are associated with promoters of transcriptional regulators (Lepoivre et al., 2013).

Our understanding of the process of transcription has changed drastically with the advancement of high throughput sequencing technology. Surprisingly, several studies using mammalian embryonic stem cells have shown that all active promoters possess the ability to drive transcription in both sense and antisense directions, however, mature transcription occurs only in the sense orientation whereas the antisense transcripts are targeted for immediate degradation (He et al., 2008; Ntini et al., 2013b; Preker et al., 2008a; Seila et al., 2008; Seila et al., 2009). In contrast, the gene pairs regulated by bidirectional promoters produce mature transcripts in both sense and antisense directions However, the mechanism underlying the same remains elusive. In light of the above findings, the aims and objectives of my thesis are as follows

Aims and Objectives

Recent evidences have shown that the majority of mammalian promoters exhibit transcription initiation on both sides in opposite orientations. Such divergent transcription does not produce mature transcripts in antisense orientation. However, a distinct class of promoters has been identified, which regulates transcription in a non-intuitive fashion, where mature transcription is observed in both the sense and antisense orientations. Since more than 10% of human genes are regulated in this manner, the understanding of this form of transcription regulation would yield valuable insights into the dynamic nature of chromatin and novel interplay between transcription and epigenetic modifications of chromatin. It is known that epigenetic modifications and transcription factors are linked functionally to the process of transcription (Donati et al., 2006; Sindhu et al., 2012).

Most of the studies focused on role of these overrepresented transcription factors in regulation of transcription from bidirectional promoter have revealed very narrow insights. Despite the advancement in modern genomics and proteomics techniques we still do not have answers for certain fundamental questions regarding the regulation of transcription from the bidirectional promoters. In this study, we attempt to elucidate the roles of epigenetic regulation in the form of dynamic histone modifications and transcription factor occupancies towards regulation of transcription from bidirectional promoters. Therefore, with the objective of deciphering the molecular mechanism of bidirectional promoter regulation, a study was undertaken with the following aims.

1. **To study the roles of histone modifications in regulation of bidirectional promoters**
2. **To study the roles of transcription factors in regulation of bidirectional promoters**

Summary of work done

1. To study the roles of histone modifications in regulation of bidirectional promoters

The first chapter of this thesis focuses on construction of a dual reporter vector tool and analyzing the mechanism of transcription regulation from bidirectional promoters. These promoters have the unique property of generating mature and functional transcripts in both the sense and antisense orientations in the cell. Regulation of bidirectional transcription is not clearly understood; this study aims at answering this question from an epigenetic perspective.

Majority of the studies aiming to characterize and understand the properties of bidirectional promoters have been using luciferase reporter system. However, a major limitation of this system is that the transcription from only single direction can be measured. To overcome this limitation, a vector system was constructed which allows us to score for the transcription happening from bidirectional promoters in both the directions. The vector was designed in a manner that any DNA fragment can be tested for its bidirectional promoter activity by scoring for the expression of two reporter genes – GFP and mCherry ; An essential control for the experiments is a unidirectional promoter that can transcribe either of the two reporters based on its directionality. This vector design allows us to monitor the expression of two fluorescent reporters (GFP and mCherry) in single cells by making use of microscopic as well as FACS based techniques. This vector was named the dual reporter (pDR) vector. This vector was validated in vivo by cloning the CMV promoter in sense and antisense orientation with respect to the reporters. The CMV promoter is a well characterized unidirectional promoter (Chambers et al., 2015); therefore, it drives the expression of GFP or mCherry depending on whether this promoter was cloned in sense or antisense orientation to the two reporters. To determine that the pDR vector system can indeed work as a potential tool to score for bidirectional transcription from any given DNA element which possesses bidirectional transcriptional activity, 8 human known bidirectional promoters were cloned in sense and antisense orientation to the two reporters in the pDR vector. Flow

cytometry (FACS) and microscopy analysis were used to assess bidirectional transcription from these promoters. The CMV promoter clones were used as a control for these experiments. Our FACS and microcopy analysis revealed that all the bidirectional promoters exhibit expression of both reporter proteins (EGFP and mCherry). However, expression of only one reporter protein from CMV promoter clone was observed suggesting that the pDR vector system can be used to identify bidirectional promoter activity from any given DNA element.

It has been known for years that post translation modifications of histone proteins alone or in combination play crucial role in transcription (Kouzarides, 2007). The Histone code paradigm suggests that histone modifications act as molecular code, read by regulatory proteins which either activate or repress the transcription process (Strahl and Allis, 2000). Using the knowledge of histone modifications and their effect on transcription, we asked if there is any link between epigenetic histone modifications and bidirectional transcription. The bidirectional promoter for the gene pair (NFYA/ OARD1) that exhibited bidirectional transcription in the pDR vector assay was chosen for chromatin immunoprecipitation (ChIP) analysis in Jurkat cells. ChIP was performed in Jurkat cells using highly specific antibodies to the following histone modifications- H3K4me3, H3K9ac, H3K79me3 and H4ac. Expression of both the genes was confirmed by quantitative PCR prior to the ChIP experiments. We analyzed the region 1 kb upstream and downstream of TSS of NFYA and OARD1 genes. This region includes the common bidirectional promoter and 1kb of the gene body for each gene. The ChIP analysis in Jurkat cells revealed that bidirectional promoters harbor unique distribution of active transcription-associated promoter and elongation marks in antisense orientation as compared to expression matched unidirectional promoter. We validated this observation by using one of our previously validated bidirectional promoter pDR clone. For this experiment, CMV promoter cloned in pDR vector was used as control unidirectional promoter. Interestingly, all the active elongation marks were enriched in both sense and antisense orientation only on the bidirectional promoter clone, whereas the unidirectional CMV promoter clone exhibited enrichment of active transcription elongation marks only in the sense direction to transcription. This observation prompted us to think that such unique distribution of active transcription marks on bidirectional promoters and their

respective gene bodies might be responsible for mature antisense transcription elongation from such loci.

To test our hypothesis comprehensively, we analyzed genome-wide occurrences of 39 histone modifications specifically at all bidirectional promoters and compared to that of all other promoters in the human CD4 T-cells. Only the promoters whose genes exhibited similar expression levels were compared between the two categories. We found that the histone modification marks H3K4me1, H3K4me3, H3K23ac, H2AK9ac and H3K36me1 exhibit enrichment in a bimodal fashion on the bidirectional promoters whereas they show enrichment only in the sense orientation on unidirectional promoters. H2BK5me1 which has been reported to occur downstream of actively transcribing promoters (Barski et al., 2007a) was observed to occur in a pattern concomitant with active transcription on bidirectional promoters. H3K27me1, H4K20me1 along with H3K36me3 and H3K79me3, which are associated with successful mature transcription occur in a bimodal distribution only on bidirectional promoters. This pattern overlays the distribution of RNA-seq tags on the genes regulated by these promoters. In cases where one of the genes of the bidirectional pair is not expressed, all the listed histone marks exhibit a profile similar to that of unidirectional promoters. These findings strongly suggest that occurrence of these marks correlates with the process of active transcription maturation from the bidirectional promoters.

Our analysis of histone modification profiles identifies an epigenetic signature of bidirectional promoters that sets them apart from all other transcribing loci in the genome. We propose that the bimodally distributed chromatin marks that occur flanking the transcription start site (TSS) could facilitate mature transcription from these sites in both sense and antisense orientations. We also addressed the functional importance of conservation of intergenic region in bidirectional promoters. We analyzed the distribution of epigenetic marks on bidirectional gene pairs with intergenic distance from 500 bp to 10 Kb. We observed a sharp decrease in the antisense peak of active promoter and transcription elongation associated marks associated with the increase in the intergenic region between two bidirectional genes.

Furthermore, to analyze if these epigenetic marks have an influence on bidirectional transcription in cellular context, we used RA mediated differentiation of NT2D1 as a model system and analyzed one of the bidirectional gene pair (NUP62CL-PIH1D3) for various transcription elongation marks (H3K79me3, H3K36me3 and H3K27me1). We analyzed the enrichment of these elongation marks in 2 Kb upstream and downstream region from TSS of NUP62CL-PIH1D3 gene pair including the intergenic region. Interestingly, all the 3 elongation marks revealed significantly high enrichment downstream to TSS on both gene bodies further strengthening the significance of bimodal distribution of elongation marks in mature bidirectional transcription from sense and antisense gene.

Thus, we demonstrate for the first time a strong link between the epigenetic marks on the bidirectional promoters and the transcriptional state of the bidirectional genes. Our data identifies an epigenetic signature of bidirectional promoters that sets them apart from all other transcribing loci in the genome. Observed bimodal distribution of epigenetic marks might be one of the unknown mechanisms which presumably plays an important role in mature sense and antisense transcription from bidirectional promoters. We propose that unique distribution of the active transcription marks on bidirectional promoters signals the transcription machinery to drive transcription in both sense and antisense orientations, however these active epigenetic marks exhibit enrichment only in one orientation on unidirectional promoters and therefore transcription maturation progresses only in one direction.

2. To study the role of transcription factors in regulation of bidirectional promoters

Most of the studies attempting to understand the role of transcription factors in regulating bidirectional promoters have drawn conclusions either based on the reporter assay systems or by determining their occupancies on bidirectional promoters. However, the mechanistic basis and significance of these observations remains poorly understood. It is an established fact that binding of a transcription factors might result in activation or

repression of gene expression depending on whether it recruits a co-activator or co-repressor. A typical transcription factor harbors multiple functional domains, not only for binding to specific DNA sequence, but also for interaction with other activator or repressor complexes. Since there is a lack of understanding in the field with regards to the role of transcription factors in regulating transcription from the bidirectional promoters, we have attempted to gain key insights into the same in the second chapter of the thesis.

Several studies have focused on delineating the common features in various bidirectional promoters. One of the common features that has been widely accepted is the overrepresentation of binding motifs for few specific transcription factors. These include motifs for GABPA, MYC, E2F1, E2F4, NRF-1, CCAAT and YY1 which are highly enriched on bidirectional promoters as compared to unidirectional promoters (Lin et al., 2007a) . In this study, we selected specific bidirectional promoters based on the presence of binding sites of these overrepresented transcription factors. To test the significance of this overrepresentation, knockdown of SP1, YY1, NRF1 and GABPA was performed followed by quantitative transcript profiling by quantitative RT-PCR analysis of 62 bidirectional genes. Surprisingly, we did not observe any dysregulation in transcription of bidirectional genes upon knockdown of these factors. Next, double knockdown of these factors was performed in combinations of SP1 and YY1, YY1 and NRF1 and NRF1 and SP1. Interestingly, double knockdown also did not result in any significant dysregulation in bidirectional gene expression barring few gene pairs which showed mild dysregulation. These results suggest two things about the function of these transcription factors in bidirectional promoter regulation; firstly, these factors might not be important for bidirectional promoter regulation and secondly, they might be playing redundant roles in terms of regulating the bidirectional promoters. .

In-depth literature survey of these factors in the recruitment of activator complexes led us to a study which showed that GABPA co-occupies with SAGA co-activator complex on many SAGA complex genomic targets (Krebs et al., 2011). Interestingly, GABPA also has been shown to bind multiple bidirectional promoters in various cell types (Collins et al., 2007). We used the co-occupancy data for GABPA and SAGA

complex to find out if these two factors co-occupy the bidirectional promoters. Our analysis showed that 54 bidirectional gene pairs exhibit co-occupancy of GABPA and SAGA complex. We randomly selected 8 gene pairs (16 genes) from this list for further analysis. To test if SAGA complex binds to the intergenic region of these 9 gene pairs, ChIP assay was performed using antibodies specific to SPT20 and PCAF, which are the structural and catalytic subunits of the SAGA complex. ChIP analysis revealed that SAGA complex binds to the intergenic region of all the 9 gene pairs selected. We also showed that GABPA direclty interact with the SAGA complex. These results encouraged us to analyze the expression of those 16 genes which showed co-occupancy of SAGA and GABPA upon GABPA knockdown. Surprisingly, we did not observe any dysregulation in any of these genes, suggesting the involvement of some other factors in regulation of bidrectional promoters. Further, co-immunoprecipitation assay revealed that SP1 and YY1 also interact with SAGA complex in addition to GABPA. To determine if YY1 and SP1 could also bind to SAGA complex bound sites on 9 selected gene pairs, we performed ChIP assay with YY1, SP1, and GABPA. We observed that SP1 and YY1 exhibited binding to all the 9 gene pairs, however GABPA was found to bind to only 2 intergenic regions. This finding explains the reason behind the lack of any significant effect on the regulation of chosen bidirectional gene pairs upon GABPA knockdown. Firstly, GABPA binds only to 2 intergenic regions, and secondly, in the absence of GABPA it is possible that SP1 and YY1 can recruit the SAGA complex to bidirectional promoters. To determine the redundancy in the function of SP1, YY1 and GABPA towards the regulation of bidirectional promoters, double knockdown of SP1 and YY1 was performed. However, the double knockdown also did not result in any dysregulation in the bidirectional gene expression.

Our data suggested that although SP1, YY1 and GABPA can interact with SAGA complex and bind to bidirectional promoters, the removal of these transcription factors does not affect bidirectional gene expression. Next, we tested involvement of another overrepresented transcription factor NRF1 in the regulation of bidirectional transcription from SAGA occupied genes. Co-immunoprecipitation analysis for NRF1 and SAGA complex revealed that NRF1 interacts directly with the SAGA complex. Furthermore, we showed that NRF1 also binds to SAGA occupied intergenic region of the bidirectional

gene pairs. Combined together, we showed that transcription factors SP1, YY1 and NRF1 interact with the SAGA complex and co-occupy the SAGA complex-bound bidirectional promoters.

To conclusively demonstrate the redundancy of these factors in bidirectional transcription we required to deplete all the three factors, however we could not achieve efficient triple knockdown for SP1, YY1 and NRF1 using siRNA approach. To circumvent this problem, we used the CRISPR/CAS9 genomic editing tool to generate knockouts for these factors. We generated NRF1 and SP1 knockout cell lines which were used as the background to knockdown other factors to address the redundancy between these transcription factors. siRNA mediated depletion of SP1 and YY1 in NRF1 knockout background exhibited significant downregulation of expression from bidirectional gene pairs. However, YY1 knockdown in SP1 knockout background did not lead to any dysregulation of bidirectional gene expression.

Collectively, our results show that the transcription factors SP1, YY1 and NRF1 interact with the SAGA complex and co-occupy SAGA complex-bound bidirectional promoters. The depletion of SP1, YY1 and NRF1 individually or in combinations of two, does not lead to any effect on bidirectional gene pairs, presumably pointing towards their redundancy. However, depletion of all the three factors (SP1, YY1 and NRF1) together results in significant downregulation of these bidirectional gene pairs. Collectively, we showed that SP1, YY1 and NRF1 function in a redundant manner to regulate gene expression from bidirectional promoters.

Conclusions

Substantial proportions of genes (11%) within the human genome are arranged in a head-to-head fashion and are controlled by bidirectional promoters. In contrast to unidirectional promoters, bidirectional promoters produce mature transcripts in both sense and antisense orientations. This interesting property of bidirectional promoters warrants analysis into the regulation of mature bidirectional transcription. Here we describe distinct epigenetic profiles of bidirectional promoters and the possible impact of

such epigenetic marks on the regulation of transcription elongation in both directions. We analyzed the distribution of 39 epigenetic marks using ChIP-Seq data from human CD4 T cells. Bidirectional promoters of gene pairs showing high, low and anti-correlated expression were compared with unidirectional promoters genome-wide. The active promoter histone modification marks H3K4 methylation (mono-, di- and tri-), H3K23ac, H3K36me1 and H2AK9ac were enriched surrounding the promoters in both directions only on active bidirectional promoters. Interestingly, H2BK5me1; a mark associated with active promoters downstream of TSS, was found to be distributed in bimodal fashion specifically on bidirectional promoters. Transcription elongation marks H3K36me3, H3K79me3 and H3K27me1, H4K20me1 exhibited a bimodal profile specific to bidirectional promoters. Additionally, motifs of only a few transcription factors have been shown to be overrepresented on bidirectional promoters but mechanistic significance of this overrepresentation is poorly understood. Individual and double knockdown of these transcription factors in combination; specifically, GABPA, SP1, NRF1 and YY1, followed by qRT-PCR analysis for bidirectional gene pairs revealed that these transcription factors function in redundant manner. Our data shows that all four (GABPA, SP1, NRF1 and YY1) transcription factors directly interact with the SAGA co-activator complex and recruit it to bidirectional promoters therefore depletion of one or two factors in combination did not yield any significant dysregulation in transcription from bidirectional promoters. We used the CRISPR/CAS9 genomic editing tool to generate knockouts for NRF1 and SP1. Interestingly, silencing of SP1 and YY1 in NRF1 knockout background resulted in downregulation in transcription of bidirectional promoters. Collectively, our data argues that a unique epigenetic signature at the bidirectional promoters is associated with mature transcription in both sense and antisense orientation. Furthermore, the transcription factors SP1, YY1 and NRF1 function in a redundant manner to regulate gene expression from bidirectional promoters.

References

Adachi, N., and Lieber, M.R. (2002). Bidirectional gene organization: a common architectural feature of the human genome. Cell *109*, 807-809.

Ahn, J., and Gruen, J.R. (1999). The genomic organization of the histone clusters on human 6p21. 3. Mammalian genome *10*, 768-770.

Albig, W., Kioschis, P., Poustka, A., Meergans, K., and Doenecke, D. (1997). Human histone gene organization: nonregular arrangement within a large cluster. Genomics *40*, 314-322.

Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Chambers, C.B., Halford, W.P., Geltz, J., Villamizar, O., Gross, J., Embalabala, A., Gershburg, E., and Wilber, A. (2015). A System for Creating Stable Cell Lines that Express a Gene of Interest from a Bidirectional and Regulatable Herpes Simplex Virus Type 1 Promoter. PloS one *10*, e0122253.

Collins, P.J., Kobayashi, Y., and Nguyen, L. (2007). The ets-related transcription factor GABP directs bidirectional transcription. PLoS genetics *3*, e208.

Donati, G., Imbriano, C., and Mantovani, R. (2006). Dynamic recruitment of transcription factors and epigenetic changes on the ER stress response gene promoters. Nucleic acids research *34*, 3116-3127.

He, Y., Vogelstein, B., Velculescu, V.E., Papadopoulos, N., and Kinzler, K.W. (2008). The antisense transcriptomes of human cells. Science *322*, 1855-1857.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693-705.

Koyanagi, K.O., Hagiwara, M., Itoh, T., Gojobori, T., and Imanishi, T. (2005). Comparative genomics of bidirectional gene pairs and its implications for the evolution of a transcriptional regulation system. Gene *353*, 169-176.

Krebs, A.R., Karmodiya, K., Lindahl-Allen, M., Struhl, K., and Tora, L. (2011). SAGA and ATAC histone acetyl transferase complexes regulate distinct sets of genes and ATAC defines a class of p300-independent enhancers. Molecular cell *44*, 410-423.

Li, Y.-Y., Yu, H., Guo, Z.-M., Guo, T.-Q., Tu, K., and Li, Y.-X. (2006). Systematic analysis of head-to-head gene organization: evolutionary conservation and potential biological relevance. PLoS Comput Biol *2*, e74.

Lin, J.M., Collins, P.J., Trinklein, N.D., Fu, Y., Xi, H., Myers, R.M., and Weng, Z. (2007). Transcription factor binding and modified histones in human bidirectional promoters. Genome research *17*, 818-827.

Maxson, R., Cohn, R., Kedes, L., and Mohun, T. (1983). Expression and organization of histone genes. Annual review of genetics *17*, 239-277.

Momota, R., Sugimoto, M., Oohashi, T., Kigasawa, K., Yoshioka, H., and Ninomiya, Y. (1998). Two genes, COL4A3 and COL4A4 coding for the human α3 (IV) and α4 (IV) collagen chains are arranged head-to-head on chromosome 2q36. FEBS letters *424*, 11-16.

Ntini, E., Jarvelin, A.I., Bornholdt, J., Chen, Y., Boyd, M., Jorgensen, M., Andersson, R., Hoof, I., Schein, A., Andersen, P.R.*, et al.* (2013). Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. Nature structural & molecular biology *20*, 923-928.

Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S., Mapendano, C.K., Schierup, M.H., and Jensen, T.H. (2008). RNA exosome depletion reveals transcription upstream of active human promoters. Science *322*, 1851-1854.

Schmidt, C., Fischer, G., Kadner, H., Genersch, E., Kühn, K., and Pöschl, E. (1993). Differential effects of DNA-binding proteins on bidirectional transcription from the common promoter region of human collagen type IV genes COL4A1 and COL4A2. Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression *1174*, 1-10.

Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., and Sharp, P.A. (2008). Divergent transcription from active promoters. Science *322*, 1849-1851.

Seila, A.C., Core, L.J., Lis, J.T., and Sharp, P.A. (2009). Divergent transcription: a new feature of active promoters. Cell cycle *8*, 2557-2564.

Sindhu, C., Samavarchi-Tehrani, P., and Meissner, A. (2012). Transcription factor-mediated epigenetic reprogramming. Journal of Biological Chemistry *287*, 30922-30931.

Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. Nature *403*, 41-45.

Sugimoto, M., Oohashi, T., and Ninomiya, Y. (1994). The genes COL4A5 and COL4A6, coding for basement membrane collagen chains alpha 5 (IV) and alpha 6 (IV), are located head-to-head in close proximity on human chromosome Xq22 and COL4A6 is transcribed from two alternative promoters. Proceedings of the National Academy of Sciences *91*, 11679-11683.

Takai, D., and Jones, P.A. (2004). Origins of bidirectional promoters: computational analyses of intergenic distance in the human genome. Molecular biology and evolution *21*, 463-467.

Trinklein, N.D., Aldred, S.F., Hartman, S.J., Schroeder, D.I., Otillar, R.P., and Myers, R.M. (2004). An abundance of bidirectional promoters in the human genome. Genome research *14*, 62-66.

Yang, M.Q., and Elnitski, L.L. (2008a). Diversity of core promoter elements comprising human bidirectional promoters. BMC genomics *9 Suppl 2*, S3.

Yang, M.Q., and Elnitski, L.L. (2008b). Diversity of core promoter elements comprising human bidirectional promoters. BMC genomics *9*, S3.

Chapter 1

1.1 Introduction

In eukaryotic cells, DNA is wrapped around histone core octamer along with other accessory proteins to form a complex structure known as chromatin. Chromatin provides a scaffold for various DNA related regulatory functions such as replication, transcription etc. The basic unit of chromatin is the nucleosome which is composed of two copies of each of the four core histone proteins H2A, H2B, H3 and H4. With the help of other proteins like histone H1, nucleosomes are further packed in 30nm fibers with 6 nucleosomes per turn in a spiral or solenoid arrangement (Hayes and Hansen, 2001; Kornberg and Lorch, 1999). The core histone octamer, which is an important components of ordered nucleosome structure, contains a 3 helix region called histone fold domain and two unstructured tails (Luger et al., 1997). Although, histone tail domains are dispensable for nucleosome formation, they play an important role in mediating nucleosome-nucleosome interaction, formation of transcription repressive chromatin referred to as heterochromatin, and transcriptionally active chromatin known as euchromatin (Grunstein et al., 1995). Inside the nucleus, DNA is read by tiny molecular machines to give rise to different form of RNA molecules. By this process genetic information stored in the DNA gets transferred into RNA which further gets translated into a functional protein. Flow of this genetic information from DNA to functional protein is described in the 'central dogma of molecular biology'. Central dogma of molecular biology states that genetic information which is wired in form of DNA molecules get transcribed into individual transportable RNA molecules which carry the information for the synthesis of functional proteins. Transcription is a universal process by which genes get transcribed to give rise to functional RNA molecules which in turn get translated into functional proteins. The process of transcription involves a plethora of ordered protein-DNA and protein-protein interactions. The mechanism by which cellular transcription machinery accesses DNA for transcription involves epigenetic remodeling of the locus

which further allows protein-protein and DNA-protein interactions, which subsequently have either positive or negative effect in transcription (Sandelin et al., 2007). Improved understanding of the process of transcription requires an understanding of DNA elements which regulate transcription such as promoters, enhancers, insulators etc. along with the proteins which bind to these DNA elements. Proteins which bind the upstream region of the transcription start site (TSS) and help in recruitment of RNA Polymerase II (Pol II) are called general transcription factors (TF) and the DNA element to which they bind in order to regulate transcription is called a cis regulatory element (Sperling, 2007). General transcription factors provide the core scaffold for organization of the pre-initiation complex which subsequently drives transcription downstream to the TSS (Lee and Young, 2000).The aforementioned mechanism holds true for basal transcription. Transcription is also regulated in a cell type specific manner where cell type specific proteins, known as regulatory transcription factors, bind to cis regulatory elements in a cell type specific manner and regulate the transcription (Sperling, 2007). In prokaryotes, genes having similar functions are clustered together in operons, and are regulated in a combinatorial fashion (Koonin, 2009; Rocha, 2008). Mammalian genomes also harbor gene clusters, most of these genes are functionally related or participate in the same physiological process. Particularly, divergent mode of gene arrangement is the closest evidence of conservation of operon concept in higher organisms. These divergently linked genes have gained considerable attention in the last decade, however, the mechanism of regulation of these closely linked gene clusters is still poorly understood. Various evolutionary studies present evidence that genes which show conservation across different species also harbor close proximity to each other. This proximity arrangement of genes and regulatory elements facilitates combinatorial regulation of genes in clusters. The variety of DNA elements such as promoter, enhancer and insulator add another level of complexity to the process of transcription.

1.1.1 Eukaryotic promoter

Promoter is a regulatory element located upstream of the 5’ region of the gene and plays an important role in the initiation of transcription (Fig. 1.1.1). In vitro studies of

eukaryotic promoter elements were carried out by using viral promoters or protein coding gene promoters which are actively transcribing at any given time in the cell. One of the first conserved motif identified in eukaryotic promoters is known as TATA box (Breathnach and Chambon, 1981), which is typically present at 25 to 30 bp upstream to the TSS. TATAAA sequence is identified as TATA box consensus sequence (Bucher and Trifonov, 1986). Mutagenesis studies have shown that even a single base pair mutation in TATA box results in a drastic decrease in the transcription suggesting the importance of TATA box in eukaryotic transcription (Wasylyk et al., 1980). Transcription from TATA box is regulated by TATA binding proteins (TBP) and TBP related factors (TRF). TFIIB recognition element (BRE) is another important motif which allows TFIIB binding to BRE element. Consensus sequence of BRE (G/C-G/C-G/A-C-G-C-C) was revealed with X-ray crystallographic study of DNA and TFIIB complex (Tsai and Sigler, 2000). Interestingly, 3' end of BRE consensus sequence is followed by 5' end of TATA box. Some eukaryotic promoters contain the Initiator element (Inr) instead of the TATA box. Subsequently, Inr elements were shown to be present in both TATA containing and TATA less promoters (Butler and Kadonaga, 2002). Inr element harbors cytosine residue at -1 and adenine residue at the +1 position to the TSS. These two nucleotides surrounding the TSS determine the strength of promoters. TFIID binds to Inr element in a sequence specific manner (Kaufmann and Smale, 1994; Martinez et al., 1994). Apart from TFIID, RNA Pol II along with TBP, TFIIB and TFIIF are also shown to recognize Inr element as well as regulate the transcription in Inr element dependent manner (Carcamo et al., 1991; Weis and Reinberg, 1997). In addition to the TATA box, BRE and Inr elements, there are other core promoter elements identified which play important roles in the transcription. One such element was identified as downstream promoter element (DPE) which has the ability to bind to TFIID. TFIID binds cooperatively with DPE and Inr element; and, mutation in any of the two element results in loss of TFIID binding (Burke and Kadonaga, 1996). DPE is precisely located at +28 to +32 from the Inr. All DPE containing promoters possess identical spacing between DPE and Inr element; any mutation or change in this spacing leads to loss in TFIID binding and several fold decrease in the transcription (Burke and Kadonaga, 1997). On the other hand, several eukaryotic promoters contain a stretch of CpG islands. CpG island containing promoters lack DPE and Inr elements and

harbor binding site for transcription factors like SP1 (Brandeis et al., 1994; Macleod et al., 1994). Transcription can initiate from multiple sites (ranging from 100 nucleotide) from CpG island containing promoters while it starts from a single unit from other core element promoters (Butler and Kadonaga, 2002).

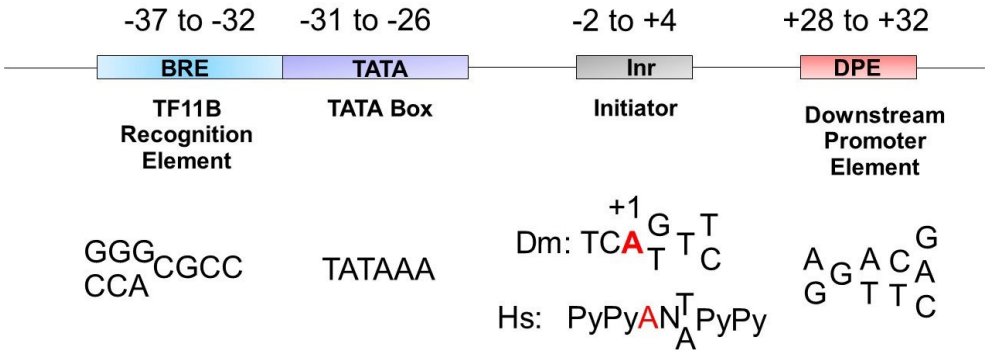


Figure 1.1.1 Core promoter elements of eukaryotic promoter.
Eukaryotic core promoter elements play important roles in RNA Pol II driven transcription. TFIIB recognition element (BRE) is the binding site for TFIIB and is present immediately upstream to the TATA box (TATA). Downstream to the TATA Box there is an initiator element (Inr). Inr sequence for both Dm (*Drosophila melanogaster*) and Hs (*Homo sapiens*) is shown. The most downstream core promoter element is known as downstream promoter element (DPE).

1.1.2 Unidirectional and bidirectional promoters

With the completion of the human genome sequence assembly, it became apparent that there are three different modes of gene arrangement on chromosomes; head to head, head to tail and tail to tail. In head to head gene arrangement, two genes are present on opposite strand of DNA and regulated by a common intergenic region between the TSS of both genes whereas in head to tail gene arrangement, one gene is followed by the other gene and each gene has its own upstream regulatory region. In tail to tail gene arrangement, two genes are oriented towards each other. Upstream DNA elements of transcription start sites (TSSs) act as putative promoter element for head to tail and tail to tail arranged genes (Fig 1.1.2). These promoters act like conventional unidirectional promoters which regulate the transcription of downstream genes. In contrast, if two genes are arranged in head to head orientation, then common DNA element between the two genes acts as a promoter for both; and hence is known as a

bidirectional promoter. Despite having a large genome size , a substantial proportion of genes (11%) are arranged in head to head fashion and are regulated by bidirectional promoters (Trinklein et al., 2004). This abundance of divergent arrangement of gene pair has been observed across several mammalian genomes, suggesting that there is some evolutionary pressure for conserving this type of gene pair structure, also known as bidirectional gene pair (Adachi and Lieber, 2002; Koyanagi et al., 2005a).

A "bidirectional gene pair" is defined as a pair of two adjacent genes located on opposite strands of DNA with transcription start sites (TSSs) not more than 1 Kilo base (Kb) apart. The intergenic region between the two TSSs of a bidirectional gene pair is commonly designated as a putative "bidirectional promoter". It has been shown that a bidirectional promoter regulates transcription of gene pairs whose expression levels are required to be regulated in a coordinated manner for proper biological functioning of the organism. Some bidirectional promoters serve to maintain a proper stoichiometric relationship of genes' expressions; such as histone genes (Ahn and Gruen, 1999; Albig et al., 1997; Maxson et al., 1983), whereas others regulate co-expression of the gene pairs which are part of a common biological pathway (Momota et al., 1998; Schmidt et al., 1993; Sugimoto et al., 1994).

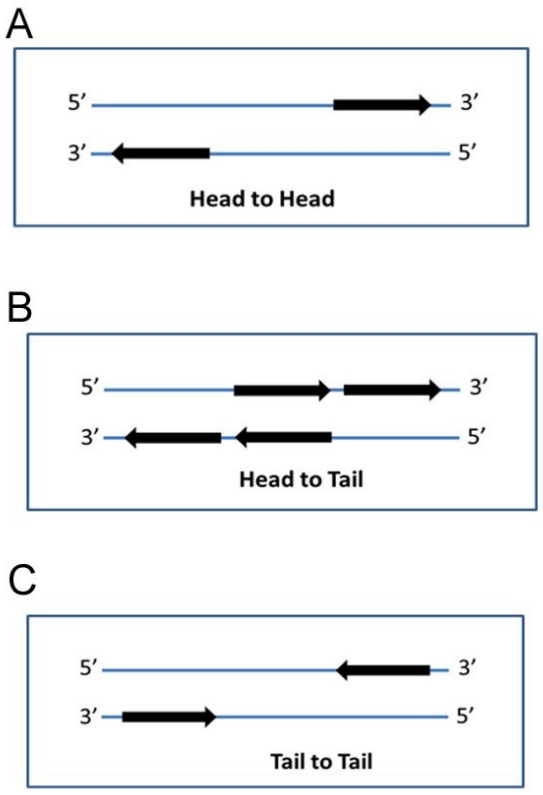


Figure 1.1.2 Gene arrangement in mammalian genomes.
Genes are arranged primarily in three distinct ways (A) In head to head gene arrangement, two genes are present on opposite strand of DNA and regulated by a common intergenic region between the TSS of both genes. (B) In head to tail gene arrangement one gene is followed by the other gene and each gene has its own upstream regulatory region. (C) In tail to tail gene arrangement two genes are oriented towards each other and each gene is regulated by its own upstream regulatory region.

1.1.3 Occurrence and Conservation of bidirectional promoters in different species

Occurrence of bidirectional promoters has been observed across different species. One of the first study by Trinklein et al. revealed that more than 10% of human genes are arranged in head to head orientation (Trinklein et al., 2004). Study on occurrence of bidirectional promoter in human, mouse, rat, chicken and fish have revealed that this kind of gene arrangement is conserved across different eukaryotic species (Li et al., 2006). Genome-wide comparative analysis of bidirectional promoters in rice (*Oryza sativa*), *Arabidopsis thaliana* and black cottonwood (*Populus trichocarpa*) has shown that occurrence of bidirectional promoters in these plants species is equal to that

of the human genome (Dhadi et al., 2009). Furthermore, occurrence of bidirectional promoters is also observed in insect genomes and analysis of 23 different insect genomes has revealed that occurrence of bidirectional promoters depends on the compactness of genome and density of genes (Behura and Severson, 2015). Enrichment of bidirectional promoters in the insect genome is correlated with presence of a transcriptional hotspot as compared to unidirectional promoters (Behura and Severson, 2015). Comparative study of bidirectional promoters in *Drosophila melanogaster* and vertebrates revealed that 31.6% of genes are arranged in divergent manner and show remarkable conservation with respect to relative gene density as compared to other eukaryotic species (Yang and Yu, 2009).

The presence of diverse bidirectional promoters in different plants and animal species suggests the prevalence of some kind of evolutionary pressure to preserve this kind of gene arrangement. Bidirectional gene pairs show evolutionary conservation in plants, though it is less as compared to mammalian genomes. Interestingly, conservation was observed across the genome of rice (*Oryza sativa*), *Arabidopsis thaliana* and black cottonwood (*Populus trichocarpa*) for only those bidirectional gene pairs which showed high co-expression or shared same gene ontology classification (Krom and Ramakrishna, 2008). Bidirectional gene pairs show less conservation in *Saccharomyces cerevisiae* than those in mammals (Tsai et al., 2007). Studies on conservation of head to head, tail to tail and tail to head gene pairs revealed that there is specific selective evolutionary pressure to conserve head to head gene architecture (Franck et al., 2008; Yang and Elnitski, 2008a). Many studies have been carried out to study origin and evolution of bidirectional gene pair arrangement and showed that bidirectional gene arrangement plays an important role in regulating mammalian genomes. Several lines of evidences show that bidirectional gene pairs are co-regulated. Bidirectional gene pairs from *Arabidopsis thaliana* genome showed more correlation in expression as compared to randomly selected gene pairs (Wang et al., 2009). It has also been established that bidirectional promoters with correlated functions are highly co-expressed. In insects, it has been shown that around 85% of the bidirectional gene pairs exhibit either positive or negative correlation with respect to gene expression (Yang and Yu, 2009). In humans most of the bidirectional gene pairs are co-expressed (Trinklein et al., 2004) and have

diverse functions. Many machine learning methods have been developed to identify bidirectional promoters from other genomic regulatory elements in human and mouse (Yang and Elnitski, 2008c).

1.1.4 Different promoter elements and their impact on bidirectional promoters

Several studies have analyzed the sequence features of bidirectional promoters to determine if there is any correlation between the promoter sequence and transcription regulation from these promoters (Lin et al., 2007a). It was shown that most of the bidirectional promoters lack TATA boxes and are enriched in CpG islands (Adachi and Lieber, 2002; Takai and Jones, 2004). They also exhibit a mirror DNA sequence composition, such that Gs and Ts dominate on one side of midpoint, while 'C's and 'A's on the other side (Engström et al., 2006). Earliest documentation of functional promoter elements is the TATA box element which has been shown to be a major DNA element for RNA Pol II recruitment but now we know TATA box regulates only a small proportion of eukaryotic promoters (Burke and Kadonaga, 1997). TATA motif enrichment study have shown that only 29% of non-bidirectional promoters harbor TATA box while only 9% of the bidirectional promoters show enrichment of TATA motif which suggests that TATA motif is under represented on bidirectional promoters (Yang and Elnitski, 2008b). DPE shows enrichment on 46.5% and 50.6% on bidirectional and non-bidirectional promoters respectively. INR elements also show enrichment at its functional position on 25.3% bidirectional promoters and 30.8% non-bidirectional promoters. Thus, DPE and INR promoter elements do not show any biased enrichment on bidirectional promoters. However, bidirectional promoters exhibit differential enrichment of CpG islands as compared to the non-bidirectional promoters. Nearly 90% of bidirectional promoters contain CpG island as compared to 45% of non-bidirectional promoters (Yang and Elnitski, 2008a). Higher enrichment of CpG islands has been correlated with some important features of bidirectional promoters such as high basal level of transcription and higher RNA Pol II occupancy (Barski et al., 2007a). Ontology analysis of gene pairs which are regulated by bidirectional promoters show that these genes are

overrepresented in some of the pathways including DNA repair, chaperone function and mitochondrial genes (Wakano et al., 2012).

1.1.5 Bidirectional promoters function as a single unit

Since a bidirectional promoter regulates two genes simultaneously, it is an interesting question to ask whether they function as a single unit or independent regulatory elements for those two genes. By using luciferase reporter assays with different truncation in intergenic region it has been shown that most of the bidirectional promoters function as a single unit and share a cis regulatory element which is essential and sufficient for bidirectional transcription (Trinklein et al., 2004). Any perturbation to common cis regulatory elements disrupts the transcriptional potential of bidirectional promoters in both the orientation (Lin et al., 2007b; Trinklein et al., 2004). These shared regulatory elements regulate transcription in both directions; however, mechanism of regulation of bidirectional gene pairs by these regulatory elements is poorly understood.

1.1.6 Crosstalk between epigenetics and transcription

The process of transcription requires unfolding of the compacted chromatin, which is achieved by post-translation modifications of histone tails in the nucleosomes. The histone tails as well as the globular domains of histones are susceptible to various types of covalent chemical modifications including acetylation, phosphorylation, methylation, and ubiquitination. These covalent modifications play important roles in demarcating the active and repressive domains of chromatin. Modifications involved in transcriptional activation are referred to as euchromatic modifications, while transcription repressive modifications are known as heterochromatic modifications (Table 1). Two major mechanisms have been proposed to address the function of these histone modifications; the first mechanism involves nucleosome unraveling by disruption of the interaction between them and second involves recruitment of non-histone proteins. However, the second mechanism is better understood in terms of transcriptional regulation. Recruitment of these non-histone proteins depends on the combination of histone

modifications present on the recruitment site. These non-histone proteins possess various enzymatic activities, which allow them to modify the chromatin as per the cellular requirement. The physiological processes regulated by these modifications (transcription, replication and DNA repair) are multistep in nature and require distinct set of modifications at different stages. Therefore, an ordered series of enzymatic activities are required at different stages of these physiological processes. In this study, emphasis has been placed on the process of transcription.

Epigenetic Mark	Enriched Site	Effect on transcription	Reference
H3K4me1	Upstream of TSS (Transcriptionally active genes)	Positive	(Liang et al., 2004; Noma et al., 2001)
H3K4me2	Upstream of TSS (Transcriptionally active genes)	Positive	(Liang et al., 2004; Noma et al., 2001)
H3K4me3	Upstream of TSS (Transcriptionally active genes)	Positive	(Liang et al., 2004; Noma et al., 2001)
H2BK5me1	Downstream of TSS (Transcriptionally active genes)	Positive	(Barski et al., 2007b)
H3K27me1	Gene body of actively transcribing genes	Positive	(Ferrari et al., 2014)
H3K27me2	Gene body of actively transcribing genes	Positive	(Ferrari et al., 2014)
H3K27me3	Promoter and gene body of repressed gene	Negative	(Cao et al., 2002)
H3K9ac	Near the TSS of active genes	Positive	(Nishida et al., 2006)
H3K36me1	Near TSS of actively transcribing genes	Positive	(Tanaka et al., 2007)
H3K36me3	Gene body of actively transcribing genes	Positive	(Wagner and Carpenter, 2012)
H3K79me1	Gene body of actively transcribing genes	Positive	(Nguyen and Zhang, 2011)
H3K79me2	Gene body of actively transcribing genes	Positive	(Schübeler et al., 2004)
H3K79me3	Gene body of actively transcribing genes	Positive	(Nguyen and Zhang, 2011; Steger et al., 2008a)
H3K9me1	Gene body of actively transcribing genes	Positive	(Barski et al., 2007a)
H3K9me3	Promoter of repressed genes	Negative	(Barski et al., 2007a; Sati et al., 2012)
H4K20me1	Gene body of actively transcribing genes	Positive	(Barski et al., 2007a)
H4K20me3	Promoter	Negative	(Schotta et al., 2004)

Table 1.1.1 Histone modifications and their effect on transcription.
Various histone modifications analyzed in this study are summarized with respect to their enrichment and function.

The phenomenon of transcription can be divided into four distinct steps which include initiation, pausing, elongation, and termination. Epigenetic mechanisms have been shown to govern all the above mentioned steps of transcription. Several studies have shown that based on the epigenetic status of the promoter element, genes can be either switched ‘ON’ or ‘OFF’ (Fig 1.1.3). All the epigenetic marks which play an

important role in transcription are summarized in (Table 1.1.1). Transcription initiation is regulated by promoter associated epigenetic marks which include H3K4me2, H3K4me3, acetylation of histone H3 and H4, H3K27me3, H3K9me3 etc.. Enrichment of active transcription marks (H3K4me3, H3K9ac) at promoters is conducive for transcription initiation downstream to promoter, while enrichment of repressive marks (H3K27me3 and H3K9me3) at the promoter results in repression of transcription initiation. Even after initiation of transcription, RNA Pol II remains paused downstream to the TSS until its release which is induced by CDK9 mediated phosphorylation of serine 2 residue in CTD tail of RNA Pol II; this constitutes the second checkpoint of transcription. After the pause release, the transcript enters elongation phase, which is also regulated by epigenetic marks. H3K36me3, H3K79me3, H3K27me1/2, and H3K9me1 have been shown to be enriched on gene bodies of actively transcribing genes. Transcription factors read these epigenetic marks and regulate the transcription process accordingly.

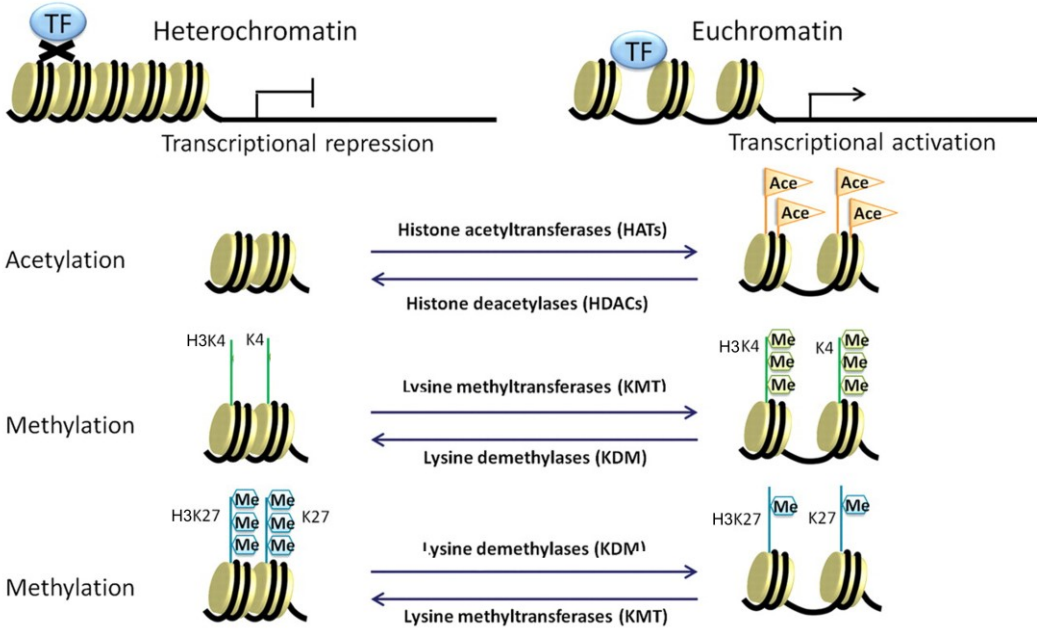


Figure 1.1.3 Schematic illustration of the crosstalk between epigenetics and transcription. Chromatin structure is regulated by epigenetic marks imposed by chromatin remodelling complexes. Transcriptionally inactive chromatin is called as heterochromatin while active form is known as euchromatin. Heterochromatin prevents the access of transcription factors (TF), whereas euchromatin is accessible for transcriptional activation. Transcriptional output from any genomic loci depends on the epigenetic status of the loci. Enrichment of histone acetylation and histone H3 lysine 4 tri methylation positively regulates transcription while enrichment of H3K9me3 and H3K27me3 result in repression of transcription. Model adapted from (Ohtani and Dimmeler, 2011).

1.1.7 Rationale of present study

One interesting aspect of bidirectional promoter is that it drives transcription in both directions; however, it's not clear as to how directionality is maintained from these promoters. Our understanding of process of transcription has changed drastically with the advancement of high throughput sequencing technology. ChIP-on-ChIP assays on divergent gene pairs and other genomic loci have revealed that on bidirectional promoters the RNA Pol II occupancy is two-fold enriched as compared to unidirectional promoters, which suggests effective RNA Pol II recruitment on bidirectional promoters (Lin et al., 2007b). Surprisingly, some studies in mammals have shown that all active promoters possess the ability to drive transcription in both, sense and antisense, direction but mature transcription occurs only in the sense orientation whereas the antisense transcripts are targeted for immediate degradation (He et al., 2008; Ntini et al., 2013b; Preker et al., 2008a; Seila et al., 2008; Seila et al., 2009). Contrary to this, bidirectional promoters transcribe mature transcripts in both sense and antisense directions; however, the reason underlying the same remains elusive. Lin et al. showed that chromatin structure of bidirectional promoters is more open as compared to any other promoter locus in the genome, indicating towards a default transcriptionally active state of bidirectional promoters (Lin et al., 2007b). Difference in promoter sequence elements on uni and bidirectional promoters is insufficient to explain the functional difference between these two types of promoters. Moreover, very little is known about the gene regulation from bidirectional promoters.

Under this chapter, we attempt to understand the mechanism underlying bidirectional promoter mediated gene regulation by analyzing the patterns of histone marks which might regulate the transcription from bidirectional gene pairs. We propose that epigenetic modifications at such promoters could play important roles in regulation of bidirectional transcription. Our proposal is largely based on the fact that epigenetic modifications are known to dictate the transcriptional status of unidirectional genes. In this study we have focused on identifying the epigenetic profile of bidirectional promoters and their possible impact in regulation of transcription from these promoters.

1.2 Material and Methods

1.2.1 Construction of pDR vector

To study bidirectional transcription, a dual reporter vector (pDR) was designed by modifying pmCherry-N1 vector (Clontech) where, EGFP and mCherry were cloned in sense and anti-sense orientation with respect to each other. Briefly, pmCherry-N1 vector was digested with AseI and NheI to remove the CMV promoter. Next, EGFP sequence along with SV40 polyA sequence was PCR amplified from pEGFP-N1 vector. SV40 polyA sequence was included to ensure that there is no difference in RNA stability of mCherry and EGFP. Amplified PCR product was purified by phenol-chloroform extraction and subsequently digested with AseI and NheI. Restriction digestion product was purified by gel extraction (Qiagen). Linearized vector (pmCherry-N1) and insert (EGFP-SV40 pA) were ligated using T4 DNA Ligase (NEB) and transformed into *E. coli* (DH5α) cells. The resulting colonies were screened by restriction digestion for presence of EGFP insert. A number of positive clones were selected and confirmed by DNA sequencing. Primers used for EGFP amplification were designed such that EGFP insertion occurred in an orientation opposite to that of the mCherry .

1.2.2 Cloning of CMV promoter in sense and antisense orientation in pDR vector

For validation of the pDR vector construct, CMV promoter was used as a unidirectional promoter control. CMV promoter was cloned in sense and anti-sense orientation to EGFP and mCherry respectively. The specific directional property of CMV promoter was used to validate our in-house pDR vector system. For sense cloning, CMV promoter was PCR amplified using AgeI (forward primer) and NheI (reverse primer) linker primers. For antisense cloning, linkers were swapped so that now NheI linker was in forward primer and AgeI in reverse primer. Both, sense and antisense PCR amplified and digested CMV fragments were ligated with AgeI and NheI digested pDR vector and transformed into *E.coli*. (DH5α) cells. The resulting colonies were screened *via* restriction digestion for presence of CMV insert. A number of positive clones were selected and confirmed by DNA sequencing.

1.2.3 Cloning of putative bidirectional promoters in pDR vector

To study bidirectional transcriptional activity from human bidirectional promoters, 8 intergenic regions from head to head gene pairs were selected. All the 8 intergenic regions were cloned in sense and antisense orientations in pDR vector. Briefly, specific intergenic regions were amplified using genomic DNA from Jurkat cell line as template. PCR products were purified by phenol-chloroform-isoamylalcohol followed by digestion with NheI and AgeI. These intergenic regions were then ligated with AgeI and NheI digested pDR vector and transformed into *E. coli.* (DH5α) cells. The resulting colonies were screened by restriction digestion for presence of CMV insert. A number of positive clones were selected and confirmed by DNA sequencing.

1.2.4 Cell culture, microscopy and FACS analysis

HEK-293T cells were grown in Dulbecco's Modified Eagle's Medium (DMEM, Invitrogen) supplemented with 10% fetal bovine serum and Penicillin/Streptomycin, at 37^{o}C under 5% CO_2 atmosphere. For transfections, HEK-293T cells were grown up to 60% confluency in 6 well culture plates at 37^{o}C in DMEM (Gibco) supplemented with 10% FBS (Fetal Bovine Serum, Invitrogen) and penicillin/streptomycin, under 5% CO_2 atmosphere. Cells were transfected using Lipofectamine 2000 as per manufacturer's instructions (Invitrogen). The medium was supplemented with 10% fetal bovine serum 6 h post-transfection. The cells were allowed to grow for 48 hrs post transfection before being imaged on confocal microscopy and flow cytometry analysis. For flow cytometry, cells were washed twice with ice cold PBS, followed by fixing with 3.7% paraformaldehyde. Flow cytometry was performed using FACS CANTO II (Becton Dickinson).

1.2.5 Antibodies and reagents

Normal rabbit IgG (12-370) and normal mouse IgG (12-371) were purchased from Millipore/Upstate. H3K79me3 (17-10130), H3K36me3 (17-10032), H3K27me1 (17-643), H3K4me3 (07-473), Acetyl-Histone H4 (17-630) and H3K9ac (17-658) antibodies were

procured from Millipore. Oct-3/4 (sc-8628) and Sox2 (sc-17319) antibodies were procured from Santa Cruz Biotechnology. Anti-Nanog (AF1997) was purchased from R&D System. Anti-H3 (ab1791) was procured from Abcam.

1.2.6 Chromatin immunoprecipitation (ChIP) assay

HEK-293T and Jurkat cells were crosslinked for 10 min at 37^{o}C by adding with formaldehyde (to a final concentration of 1%) directly to the culture medium in the flask. Crosslinking was quenched by the addition of glycine to a final concentration of 125mM. Cells were washed twice with ice-cold PBS and harvested using cell scraper and pelleted down. Subsequently, the cell pellet was resuspended in 6 volume of swelling buffer (25 mM Tris pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCL,0.1% NP40, 1 mM DTT, 0.5 mM PMSF, 1x Protease inhibitor cocktail) followed by 10 min incubation on ice. Nuclei were prepared by using Dounce homogenizer 25 times using loose piston. Centrifugation was performed for isolating the nuclei from homogenized mixture. Cells were resuspended in 8 volumes of sonication buffer (50 mM Tris-Cl pH 7.9, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Sodium deoxycholate, 1.0% SDS, 0.5mM PMSF, 1X Protease inhibitor cocktail), and sonicated using Bioruptor sonicator (Diagenode, Belgium) (10 min, 30 sec “ON”, 30 sec “OFF”; for 3 times). The sonicated sample was centrifuged at 13000 X g for 10 min at 4^{o}C and the supernatant was collected as soluble crosslinked chromatin. To analyze if chromatin preparation is of correct size (200 to 300 bp); 20 µl of the supernatant was taken (rest frozen and kept at -80^{o}C) and added with 20 µl of 5 M NaCl (0.3M final conc.) and 1 µl of RNase A (10 mg/ml), total volume was made up to 300 µl with sonication buffer. De-crosslinking was performed at 65 ^{o}C with overnight incubation in thermomixer. Twenty ug proteinase K, 20 µl of Tris pH 7.9 (1 M) and 10 µl of EDTA (0.5 M) were added followed by incubation at 42^{o}C for 1 h. DNA was purified by phenol:chloroform:isoamylalcohol extraction. Chromatin preparation was analyzed on 1% agarose gel. The fragment size was always between 200 to 300 bp. The chromatin was quantified and equal amount of chromatin (50 µg) was used for each ChIP. Final IP volume was made up to 1 ml with ChIP dilution buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-Cl pH 8.0, 167 mM NaCl). The chromatin solution was then precleared by addition of 10 µl of protein G magnetic beads cocktail (50% protein G

beads, 100 µg of salmon sperm DNA/ml, 500 µg of bovine serum albumin/ml) and kept on rocker at 4°C for 2 h. Precleared chromatin was collected by transferring tube on a magnetic rack, followed by immunoprecipitation by normal rabbit IgG (as control) at 4°C on an end-to-end rocker for overnight. Twenty µl of protein G-plus bead cocktail was added and rocking was continued for another 4 h. The beads were then harvested by magnet by placing the tubes on magnetic rack. Beads were washed thrice (10 minutes each) with low salt wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-Cl pH 8.0,150 mM NaCl), followed by three washes (10 minutes each) of high salt wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-Cl pH 8.0, 500 mM NaCl), followed by three washes (10 minutes each) of LiCl (0.25 M LiCl, 1% NP40, 1% Sodium deoxycholate, 1 mM EDTA, 10 mM Tris-Cl pH 8.0) , and final three washes (10 minutes each) of TE (50 ml of TE contain 250 ul of 2.0 M Tris-Cl pH 7.5 and 100 µl of EDTA (0.5M)). Chromatin-antibody complexes were eluted from the protein G beads by adding elution buffer (1 %SDS, 0.1 M $NaHCO_3$ (freshly prepared) and 10 mM dithiothreitol) to the beads. Twenty µl of 5M NaCl (0.3M final conc.) and 1 µl of RNase A (10 mg/ml) was added to the eluate and de-crosslinking was performed at 65°C with overnight incubation on thermomixer. Next day, proteinase treatment was performed by adding 20 ug proteinase K, 20 µl of Tris pH7.9 (1M) and 10 µl of EDTA (0.5 M) and incubating the chromatin at 42°C for 1 h. DNA was recovered by phenol:chloroform:isoamylalcohol extraction followed by a chloroform-isoamylalcohol extraction and precipitated by addition of 0.1 volume of 3 M sodium acetate (pH 5.2), 20 µg of glycogen, and 2.5 volumes of ethanol. Precipitated DNA was dissolved in water, and was analyzed by quantitative PCR (q-PCR).

1.2.7 Quantitative ChIP-PCR

ChIP products were diluted to 5 times with nuclease free water. Diluted ChIP products were used as template for the PCR with specific set of primers. For quantification of enrichment, the efficiency of chromatin immunoprecipitation of particular

genomic locus can be calculated from qPCR data and reported as a percentage of starting material: % (ChIP/ Total input) which is calculated according to following formula:

% (ChIP/ Total input)= 2^[(Ct(x%input) – log(x%)/log2) - Ct(ChIP)]x 100%

Here, 2 is the amplification efficiency (AE); Ct (ChIP) and Ct (x%input) are threshold values obtained from exponential phase of qPCR for the IPed DNA sample and input sample respectively; the compensatory factor (logx%/log2) is used to take into account the dilution 1:x of the input. The recovery is the % (ChIP/ Total input). Relative occupancy can be calculated as a ratio of specific signal over background:

Occupancy= % input (specific loci) / % input (background loci)

Relative occupancy is then used as a measure of the protein association with a specific locus.

1.2.8 ChIP-Seq of histone modifications

The human reference genome assembly hg19 was used to obtain coordinates for the genomic locations flanking bidirectional and unidirectional promoters in both directions. These locations were then used as targets and ChIP-Seq data for 39 epigenetic modifications obtained from GEO ID: SRA000287 was analyzed with respect to ChIP enrichment signal in these targets. Analysis of ChIP enrichment signal within/near genes was performed using CEAS "Cis-regulatory Element Annotation System" (Shin et al., 2009) . RNA-seq data (Abraham et al., 2013) (Accession id GSE39537) for CD4 cells was obtained from Gene Expression Omnibus (GEO). Using distribution of RPKM values bidirectional gene pairs were divided into low, high and anti-correlated expression sets. Genes from each set were analyzed for enrichment of different epigenetic marks.

1.2.9 Differentiation of NT2D1 cells

All-trans-retinoic acid (RA)-induced differentiation of N2D1 cells was performed in 100mm culture dish. RA was reconstituted at a concentration of 5mg/ml in DMSO (Sigma) and stored in dark at -80^{o}C. For differentiation experiments, NT2D1 cells were

harvested using 0.05% Trypsin and resuspended in fresh media. Cells were counted and 2 x 10^6 cells were seeded in each 100mm dish. Cells were allowed to grow for 24 h before adding RA. Next day, RA was added to a final concentration of 13.7 μM and cells were maintained in RA upto 7 days, with media replacement (containing freshly thawed RA) every day. After 7 days of RA treatment, cells were harvested for ChIP, RNA and protein.

1.2.10 Real-time quantitative PCR

RNA was isolated from control and 7 day differentiated NT2D1 cells using TRI reagent (Sigma). One μg of RNA was used for cDNA preparation per 20 μl reaction. The cDNA was used as template for the PCR amplification with gene-specific primers. The changes in threshold cycle (Ct) values were calculated as follows: ΔCt = (Ct target gene - Ct of actin). These ΔCt values were used to calculate fold change using equation as relative fold change = $2^{-(\Delta(\Delta Ct))}$ with respect to endogenous control and plotted graph for the average fold change.

1.3 Results

1.3.1 Design and construction of a vector system to study bidirectional promoter activity

To study bidirectional transcription *in vivo,* a vector system with two reporters (EGFP and mCherry) were designed (Fig. 1.3.1 A). These reporter genes were cloned in a head to head orientation and the expression of these reporters can be driven by a common promoter placed in between the two transcription start sites. Any genomic region of interest can be cloned into this vector as the intergenic region between these two reporters. Mutually exclusive or simultaneous expressions of these two reporters can be scored from this vector depending on directionality of the promoter placed in between the two reporters. Depending on the expression of one or both the reporters, uni or bidirectional transcription can be scored. If both reporters are simultaneously expressed, then the intergenic region is exhibiting bidirectional promoter activity. In contrast, mutually exclusive expression of the reporters, depending on the orientation of the

cloned sequence would score for transcriptional activity of a unidirectional promoter. Collectively, a unidirectional promoter would drive the expression of only one of the reporter genes depending on the orientation of cloning while a bidirectional promoter would drive the expression of both the reporters independent of what orientation it was cloned into the vector. This vector was generated by modifying pmCherry N1 vector (Clontech, cat. no. 632523). The constitutive CMV promoter in the mCherryN1 vector was replaced by EGFP (Fig. 1.3.1 B). EGFP was cloned in the (AseI and NheI sites) such that the ORF of EGFP would be in opposite (anti-sense) orientation to mCherry.

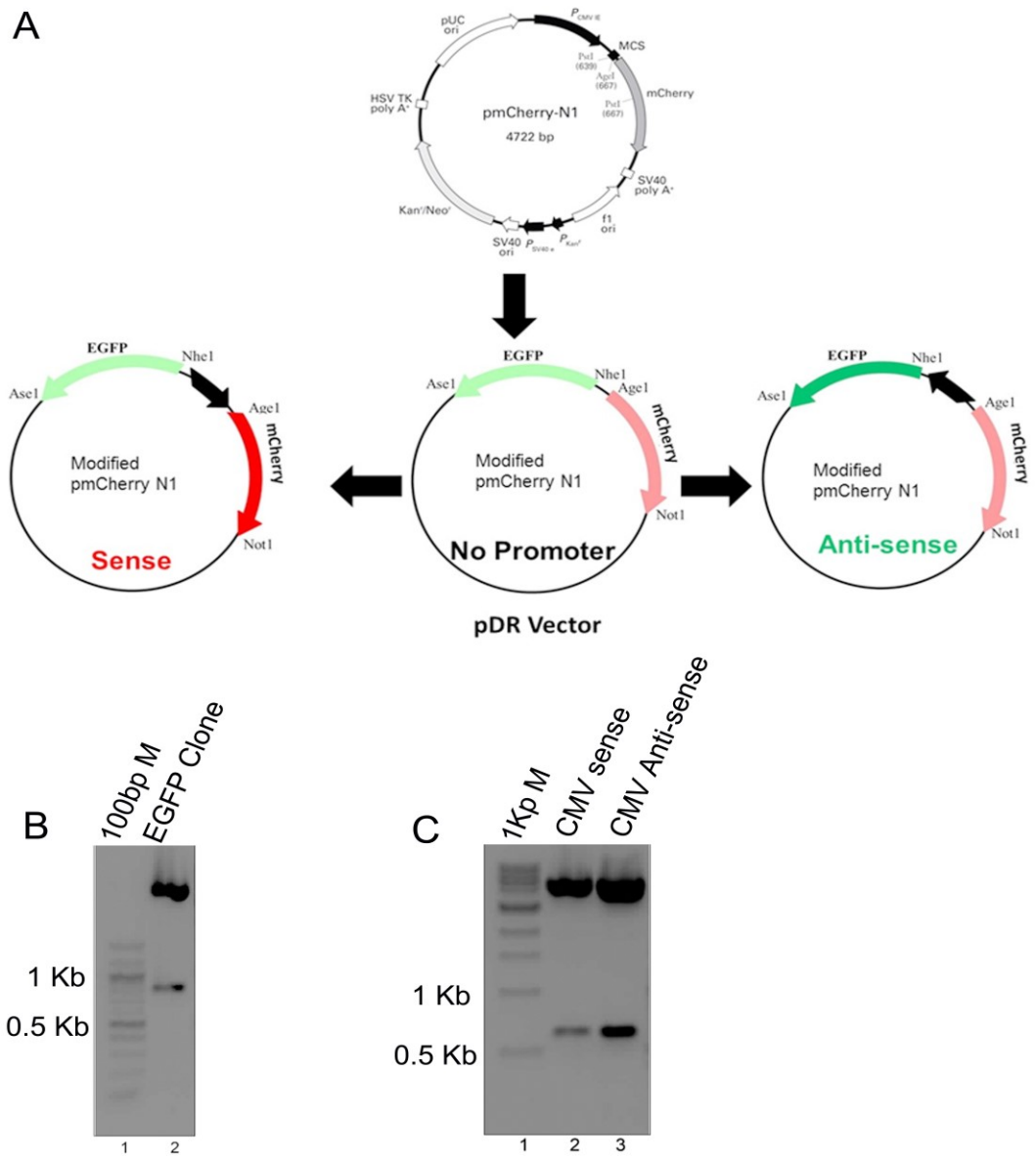


Figure 1.3.1 Design and construction of pDR vector.
(A) Strategy for introducing GFP and mCherry under a common regulatory DNA element is shown. This system was designed to provide a quantitative readout in live cells based on strand-specific promoter activity. Intense red and green colors indicate direction of promoter activity. There are three constructs shown in the scheme- *Middle construct* is pDR vector where no promoter element is present; *Right construct* is pDR vector showing CMV sense clone (to mCherry) and *Left construct* is pDR vector showing CMV antisense clone (to mCherry). (B) Agarose gel picture showing insert (EGFP) release from pDR vector after digestion with AseI and NheI. (C) Agarose gel picture showing insert (CMV) release from CMV-pDR vector clones after digestion with AgeI and NheI.

1.3.2 Validation of pDR vector

Since the CMV promoter is unidirectional promoter, we exploited its directional properties to validate our pDR construct. The dual reporter vector, named as the pDR vector (plasmid with dual reporters), as such does not contain any promoter element. Therefore, the pDR vector, on its own, does not express any of the reporter genes. If a CMV promoter is cloned in between the mCherry and EGFP, it should be able to express either mCherry or GFP depending on the orientation of the cloned promoter. To test this, we cloned CMV promoter in between the mCherry and EGFP in both sense and antisense orientation (Fig. 1.3.1 C). To validate pDR vector, we transfected CMV clones (CMV in pDR vector in sense and antisense orientations) in HEK-293T cells. Cells were analyzed 48 hrs post-transfection by microscopy and FACS (Fluorescence-activated cell sorting). We observed that cells transfected with CMV sense and antisense clones specifically express either mCherry or GFP depending on its orientation with respect to the reporter (Fig 1.3.2 A and B). CMV clone in sense to mCherry showed expression of only mCherry protein while antisense clone showed the expression of EGFP (Fig. 1.3.2 A and B). Since pDR vector does not contain any promoter element in between of mCherry and EGFP therefore, we used pDR vector as the negative control. As expected, we did not detect expression of any of reporter protein from pDR vector alone (Fig. 1.3.2 A and B).

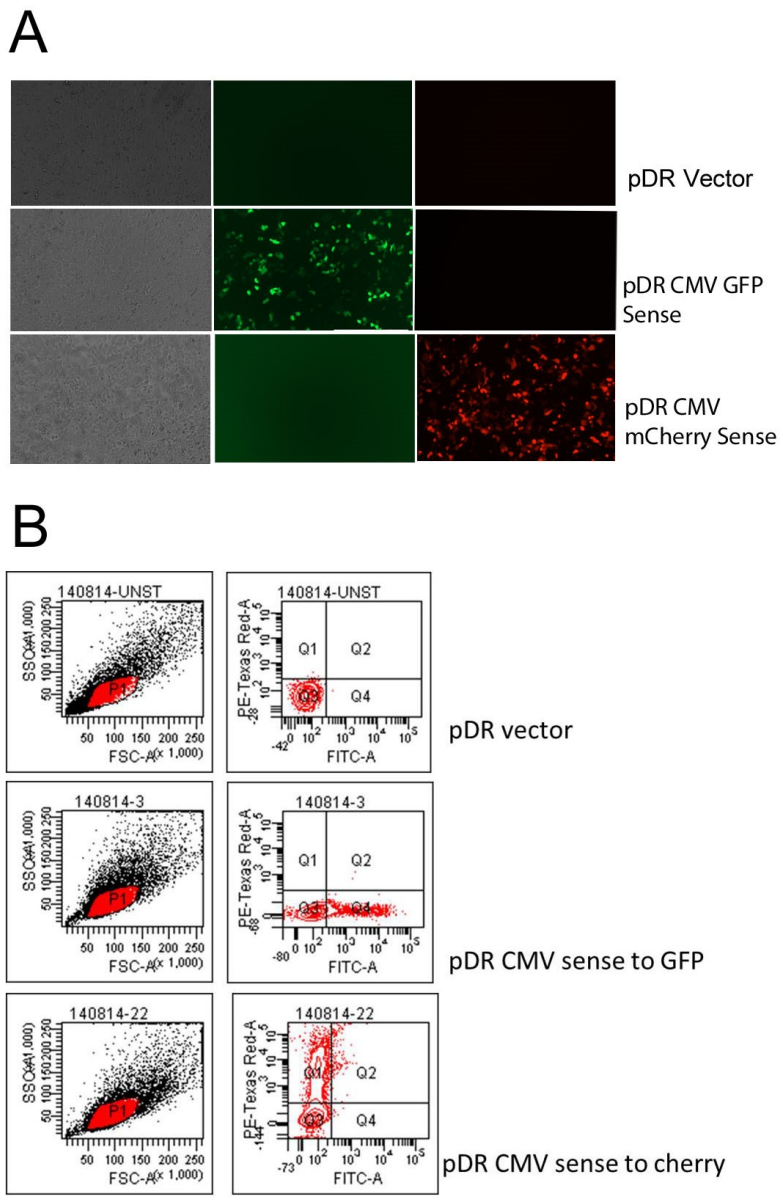


Figure 1.3.2 Validation of pDR vector.
pDR vector system with CMV promoter in sense and antisense orientation was transfected in HEK-293T cells and analyzed by microscopy and FACS. (A) Microscopic analysis show that expression of EGFP and Cherry depend on the orientation of CMV promoter in pDR clone. pDR vector with CMV either expresses EGFP or mCherry depending on the CMV promoter with respect the two reporters (B) FACS analysis shows that pDR vector itself does not express any protein due to lack of any regulatory element between both reporter protein. CMV promoter clones show mutually exclusive expression of either EGFP or mCherry depending upon the orientation of CMV promoter in pDR vector.

1.3.3 Cloning of human putative bidirectional promoters and bidirectional transcription

After validating the pDR vector, we cloned 8 putative bidirectional promoter regions from previously reported bidirectional gene pairs (Wang et al., 2013) in the pDR vector. The pDR constructs with cloned putative bidirectional promoter regions were named based on their gene pair name and subsequently sense or antisense depending on their orientation. These regions were cloned in both sense and antisense orientation and validated by restriction digestion (Fig. 1.3.3) and PCR sequencing.

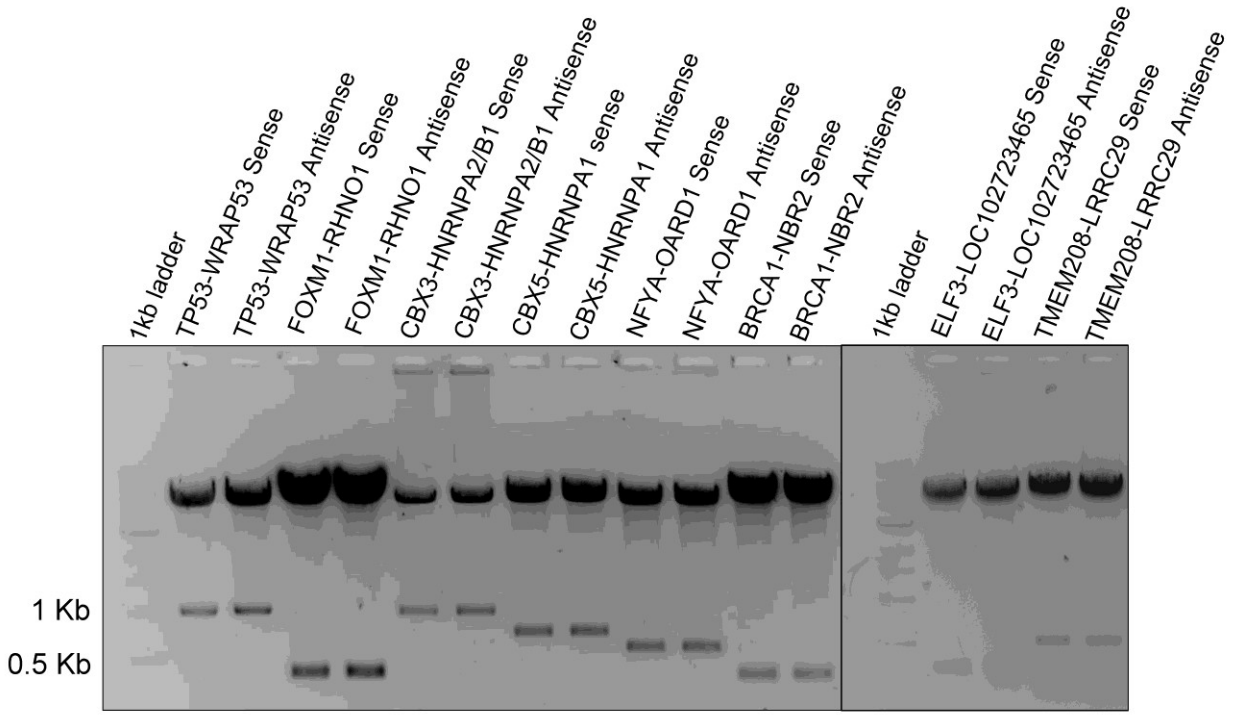


Figure 1.3.3 Cloning of human putative bidirectional promoter regions in pDR vector.
The putative bidirectional promoter regions were cloned in pDR vector in sense and anti-sense orientation. Clones were screened by restriction digestion with NheI and AgeI. Positive clones showed insert release at appropriate size.

The pDR constructs with cloned putative bidirectional promoter regions were transfected in HEK-293T cells. Sense orientation clones were named as bidirectional promoter sense clone (BDP S) while antisense clones were named as bidirectional promoter antisense clones (BDP AS) (Fig. 1.3.4 A). 48 hrs post-transfection, cells were subjected to microscopy and FACS analysis. Microscopy analysis showed that all 8

clones express both the reporter proteins in sense orientation (Fig. 1.3.4 B). Additionally, the bidirectional transcription from these bidirectional promoter pDR clones was confirmed by FACS analysis for sense clones (Fig. 1.3.4 C). We analyzed the transfected cells with antisense clones as well and observed that all the 8 antisense clones also show transcription of both reporter proteins (Fig. 1.3.4 D). These results show that pDR vector can be used to score for bidirectional transcriptional ability from any given DNA elements. A significant number of gene pairs are known to be transcribed from bidirectional promoters; however the transcriptional regulation imposed on them is poorly understood. Our dual reporter vector can be used to identify regulatory elements as well as to decipher the transcriptional regulation at such elements.

A

Clone Number	Gene Pair	Genomic assemby	Cloned Region	Length	Cloning Site
BDP S1	FOXM1-RHNO1	hg19	chr12:2986293-2986670	380bp	Sense to cherry
BDP S2	LOC105369155-TMEM208	hg19	chr16:67260620+67261122	503Bp	Sense to cherry
BDP S3	TP53-WRAP53	hg19	chr17:7590787-7591798	1000bp	Sense to cherry
BDP S4	TMEM234-EIF3I	hg19	chr1:32687713-32688023	310bp	Sense to cherry
BDP S5	CIB1-NGRN	hg19	chr15:90777156-90777587	431bp	Sense to cherry
BDP S6	HNRNPA2B1-CBX3	hg19	chr7:26240385-26241390	1005bp	Sense to cherry
BDP S7	CBX5-HNRNPA1	hg19	chr12:54653296-54654073	778bp	Sense to cherry
BDP S8	NFYA-OARD	hg19	chr6:41040191+41040774	584bp	Sense to cherry
BDP AS1	FOXM1-RHNO1	hg19	chr12:2986293-2986670	380bp	Sense to EGFP
BDP AS2	LOC105369155-TMEM208	hg19	chr16:67260620+67261122	503Bp	Sense to EGFP
BDP AS3	TP53-WRAP53	hg19	chr17:7590787-7591798	1000bp	Sense to EGFP
BDP AS4	TMEM234-EIF3I	hg19	chr1:32687713-32688023	310bp	Sense to EGFP
BDP AS5	CIB1-NGRN	hg19	chr15:90777156-90777587	431bp	Sense to EGFP
BDP AS6	HNRNPA2B1-CBX3	hg19	chr7:26240385-26241390	1005bp	Sense to EGFP
BDP AS7	CBX5-HNRNPA1	hg19	chr12:54653296-54654073	778bp	Sense to EGFP
BDP AS8	NFYA-OARD	hg19	chr6:41040191+41040774	584bp	Sense to EGFP

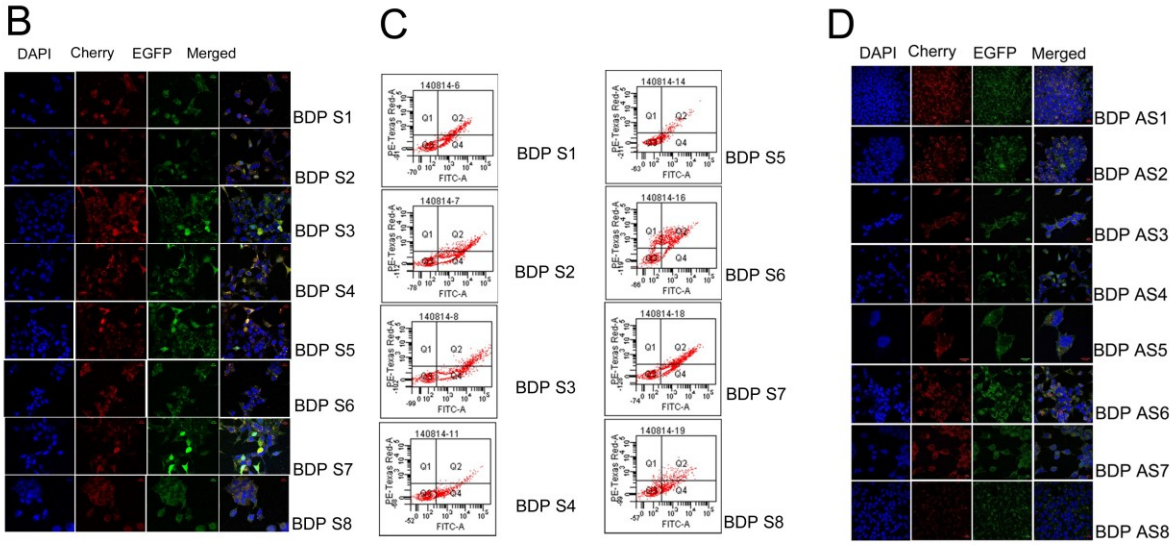


Figure 1.3.4 Bidirectional transcription from human bidirectional promoters. (A) Table showing the details of bidirectional promoter gene pairs which were cloned in pDR vector for the analysis. The bidirectional promoter pDR clones were transfected in HEK-293T cells and the transfected cells were analysed by microscopy and FACS. (B) Microscopic analysis show that bidirectional promoter sense orientation clones drive expression of both EGFP and mCherry. (C) FACS analysis also confirmed that the bidirectional promoter sense orientation clones show expression of both reporter proteins. (D) Same bidirectional promoters cloned in antisense orientation also drive the expression of both reporter proteins suggesting the mature bidirectional transcription from all the clones.

1.3.4 Bidirectional promoters possess distinct epigenetic histone marks than unidirectional promoters

A couple of recent studies have focused on the cause of abortive antisense transcription and revealed that the polyadenylation sites, in antisense direction to the unidirectional promoter, cause decay of transcripts upstream to the promoter (Almada et al., 2013; Ntini et al., 2013a). Since only bidirectional promoters exhibit mature transcription in both orientations, we were interested in understanding the mechanistic differences between bidirectional and unidirectional promoters. Epigenetic mechanisms have been shown as one of the important factors which regulate transcription and are a prime candidate to provide clues regarding regulation of bidirectional promoters. To understand, if the epigenetic mechanisms play any role in the functional difference between uni and bidirectional promoters we set out to analyze the epigenetic landscape of these two types of promoters. The endogenous bidirectional promoter for the gene pair (NFYA/OARD1) that exhibited bidirectional transcription in pDR vector assay (Fig. 1.3.4 A) was chosen to analyze the epigenetic status of bidirectional promoter. For this, we analyzed the endogenous bidirectional promoter for the NFYA/OARD1 gene pair for various histone marks including H3K4me3, H3K9ac, H3K79me3, and H4ac in Jurkat cells by ChIP. Expression of both the genes was confirmed by q-PCR (Fig. 1.3.5 A) prior to the ChIP experiments. We see more expression from NFYA gene as compared to its antisense partner OARD (Fig. 1.3.5 A). We analyzed the region 1 kb upstream and downstream of the TSS of NFYA and OARD1 genes. This region includes the common bidirectional promoter and 1kb of the gene body for each gene. ChIP qPCR was performed and data was plotted as a relative occupancy with respect to input. Interestingly, active promoter marks H3K4me3, H3K9ac and H4ac showed enrichment in the both sense and antisense orientation for the bidirectional promoter (Fig. 1.3.5 B). We also observed the spread of H3K79me3 in both sense and antisense direction (Fig. 1.3.5 B). Enrichment of all transcription associated marks show more enrichment towards NFYA gene direction which might be because NFYA shows more expression as compare to ORAD (Fig. 1.3.5 A and B).

To examine if these epigenetic marks are exhibited in a similar manner on all the bidirectional promoters, we decided to perform similar experiments using one of our

bidirectional promoter pDR clone which displayed bidirectional transcription of both reporter proteins. Our reasoning was, if the epigenetic marks play an important role in bidirectional transcription, then one would expect the bimodal distribution of transcription elongation marks on both reporter gene bodies even in this artificially generated system. Observed bimodal distribution of all histone modifications is considered with respect to center of transcription start site. To test this, we repeated similar ChIP experiments with one of our previously verified bidirectional promoter pDR clone which harbors the putative bidirectional promoter region for TP53-WRAP53 gene pair (Fig. 1.3.5 D). CMV promoter cloned into the 'sense' orientation with respect to mCherry was used as a control (Fig. 1.3.5 C). Even in this artificially generated system, a bimodal distribution of transcription elongation marks such as H3K79me3, H3K36me3 and H3K27me1 was observed only from bidirectional promoter clone (Fig. 1.3.5 D) but not with the CMV clone (Fig. 1.3.5 C). Thus, this chromatin structure is not a result of two genes lying in a head to head orientation with each other but is correlated to mature bidirectional transcription. It is to be noted that in case of the CMV clone in pDR vector, the CMV promoter lies between two ORFs that are only 600bp apart. The histone marks in this case do not spread into the antisense direction; however, in case of bidirectional promoter clone, histones marks are spread on either direction of the intergenic region. Therefore, active transcription occurs in both sense and antisense orientation from this bidirectional promoter unlike the unidirectional promoter which shows transcription only in sense. This suggests that the active bidirectional transcription could be responsible for this unique chromatin architecture of the bidirectional promoters

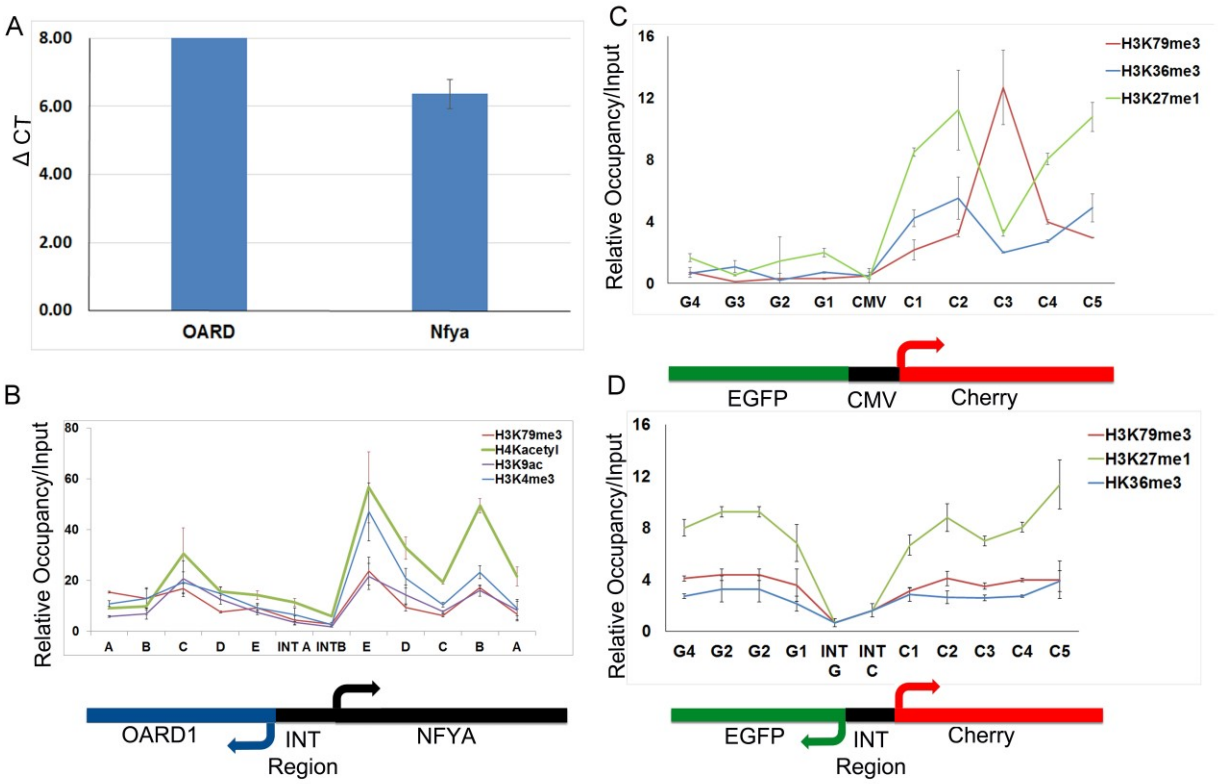


Figure 1.3.5 Epigenetic modifications on bidirectional and unidirectional promoters. (A) qPCR showing the expression of NFYA/ORAD genes. Data is plotted in form of Δ CT value, which is inversely proportional to the expression. (B) ChIP analysis at NFYA/ORAD genomic loci which include 1Kb upstream and downstream region from the TSS of both genes. Active promoter marks H3K4me3, H3K9ac and Pan H4ac show peculiar enrichment of these marks on promoter region. However in this case, both marks show enrichment in both sense and antisense orientation. Moreover, H3K79me3, a mark for active transcription elongation, shows bimodal enrichment on sense and antisense gene body of this bidirectional gene pair. (C) ChIP analysis of bidirectional promoter pDR clone withTP53-WRAP intergenic region revealed that all three active transcription elongation marks (H3K79me3, H3K36me3 and H3K27me1) are enriched on both GFP and mCherry gene body in a similar way. (C) ChIP analysis of CMV promoter pDR clone which is in sense orientation to mCherry showed enrichment of active transcription elongation marks only on mCherry gene body; no enrichment of any of these marks was seen on the GFP gene body.

1.3.5 Active transcription associated promoter marks are enriched in both sense and antisense to TSS on bidirectional promoters in the whole genome

Epigenetic histone modifications such as methylation and acetylation have been extensively documented in regulating transcription by modulating chromatin structure of the region they reside in (Goldberg et al., 2007; Koch et al., 2007; Kouzarides, 2007). Unlike unidirectional promoters, bidirectional promoters drive transcription in two opposite directions; but the chromatin structure of such promoters that permits unique

transcription is poorly understood. From our ChIP experiment, it was evident that a bidirectional promoter has distinct distribution of epigenetic marks as compared to a unidirectional promoter. We asked if these epigenetic marks occur in a similar manner on all bidirectional promoters in the genome. To address this, we analyzed previously published ChIP-seq and RNA-seq data from human CD4 T cells (Wang et al., 2008). Only those bidirectional and unidirectional genes with comparable gene expression were selected for further analysis. We divided the bidirectional gene pairs in 4 categories based on their gene expression profile. The bidirectional gene pairs where both the genes are actively transcribed were termed as “UP”. The bidirectional gene pairs where only the gene in the sense orientation was expressed were termed as “OneUP”. Contrary to this, the bidirectional gene pairs where only the gene in antisense orientation was expressed were classified as “OneDOWN”. Finally, the bidirectional gene pairs where neither sense nor antisense gene expression was observed were categorized as “DOWN”. Unidirectional genes that were expressed at a similar level as bidirectional gene pairs were classified as “UniUP”, while unidirectional genes which were not expressed were categorized as “UniDOWN”. “UniUP” genes were comparable to the “UP” category of bidirectional gene pairs while “UniDOWN” were comparable to “DOWN” category of bidirectional gene pairs in terms of their expression. Next, we analyzed the active transcription associated epigenetic marks across these categories of bidirectional and unidirectional genes with a window of +/- 3kb from TSS.

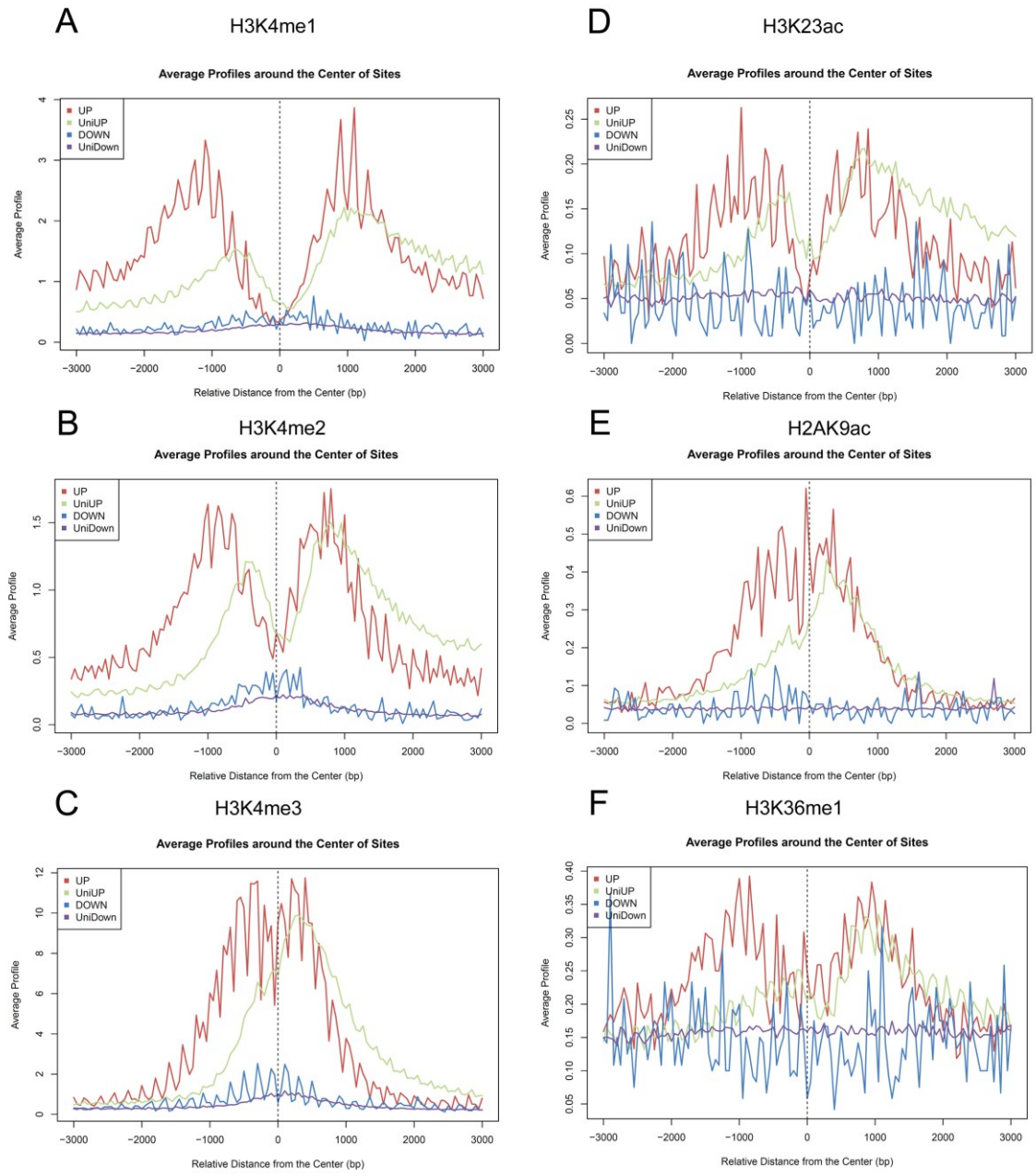


Figure 1.3.6 Genome-wide analysis of active transcription associated marks on bidirectional and unidirectional promoters.

Genome-wide analysis was performed to assess the distribution of epigenetic marks on bidirectional and unidirectional promoters. Genomic regions 3Kb upstream and downstream from TSS were taken for analysis. (A-F) Bidirectional promoters transcribing in both sense and antisense orientation are enriched for active promoter marks (H3K4me1, H3K4me2, H3K4me3, H3K23ac, H2AK9ac and H3K36me1) in both sense and antisense orientation; however unidirectional promoter show enrichment only in sense orientation.

Our analysis revealed that the bidirectional gene pairs expressing both in sense and antisense orientation ("UP" category), exhibit an enrichment of H3K4 methylation (mono, di and tri), H3K23ac, H3K36me1 and H2AK9ac in both sense and antisense direction (Fig. 1.3.6 A-F). Whereas for a unidirectional gene with comparable expression ("UniUP" category), these marks were present only in the direction of matured transcription i.e only in sense direction (Fig. 1.3.6 A-F). H3K4me1, H3K4me2, H2AK9ac and H3K23ac are enriched upto 2 kb in the antisense orientation on bidirectional promoters whereas on unidirectional promoters it was only upto 1 kb (Fig. 1.3.6 A-E). When transcription is off from bidirectional ("DOWN" category) and unidirectional promoters ("UniDOWN" category) there was no enrichment of these marks on both types of the promoters (Fig. 1.3.6 A-F). Differential enrichment of active epigenetic marks in the antisense orientation on bidirectional promoters implies that bidirectional promoters have active transcription associated chromatin state in both sense and antisense orientation. This suggests that chromatin state of bidirectional promoters is permissive for transcription and allows the transcription machinery to initiate the transcription in both directions.

1.3.6 Bidirectional promoters exhibit enriched transcription elongation marks on the antisense gene body

Recently a new concept in transcription biology was introduced where it was shown that RNA polymerase II (RNA Pol II) initiates transcription divergently from most active gene promoters, but productive elongation occurs primarily in the sense-coding direction only (Preker et al., 2008a; Seila et al., 2008). A couple of studies have provided evidence for emergence of abortive antisense transcription by demonstrating polyadenylation site enrichment and loss of U1 snRNP binding to transcripts originating in the antisense direction to promoter. These two mechanisms are known to induce the decay of antisense transcripts (Almada et al., 2013; Ntini et al., 2013a). This is the only mechanism known so far which explains the abortive antisense transcription and decay. However, mature bidirectional transcription occurs from bidirectional promoters and the molecular mechanisms for the same are unknown. Additionally, active human gene

promoters show localization of H3K4me3, the transcription initiation hallmarks, both at sense and antisense of TSS.; however, H3K79me2, indicative of elongating RNA Pol II, are only present downstream of TSSs in sense orientation (Seila et al., 2008). This suggests that though the transcription is initiated in both sense and antisense orientation, transcriptional maturation occurs only in the sense orientation. Contrary to unidirectional promoters, genes from bidirectional promoters exhibit both initiation and maturation of transcription in both sense and antisense orientation. Therefore, it becomes an obvious question as to whether the difference between these two types of promoters has any correlation with the transcriptional initiation and/ or elongation associated histone marks. To answer this question, we analyzed the enrichment of H2BK5me1, H3K36me3, H3K79me1/2 on unidirectional and bidirectional gene promoters. H2BK5me1 has been reported to be associated exclusively downstream to TSS of actively transcribing genes (Barski et al., 2007b; Li et al., 2011). We observed enrichment of H2BK5me1 on gene body of both sense and antisense from bidirectional promoters unlike unidirectional promoters. This indicated that the active transcription occurs in both orientations only in case of bidirectional promoters (Fig. 1.3.7 A).The next immediate question we addressed was what would be the fate of transcriptional maturation from bidirectional promoters. H3K36me3 and H3K79me3 have been linked to transcriptional elongation and these marks have been shown to be enriched specifically on the gene body of actively transcribing genes (Krogan et al., 2003; Li et al., 2002; Schaft et al., 2003; Steger et al., 2008) . We analyzed the distribution of transcriptional elongation marks on the gene body of bidirectional and unidirectional promoter regulated genes. We observed that H3K79me2, H3K79me1 and H3K36me3 showed bimodal distribution from the TSS of bidirectional promoter genes whereas genes from unidirectional promoters show enrichment only in the sense direction of gene body (Fig. 1.3.7 B, C and D).

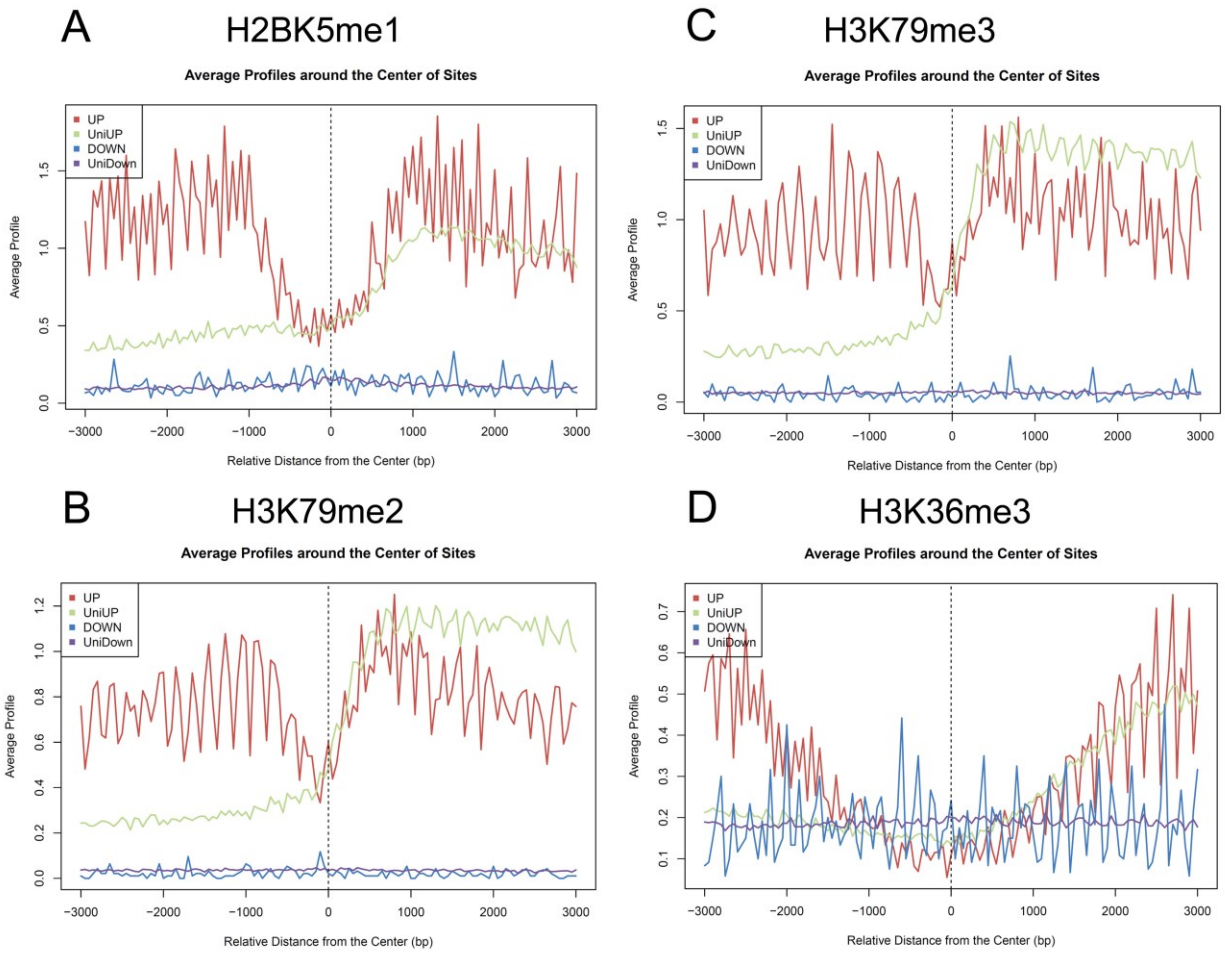


Figure 1.3.7 Genome-wide analysis of transcription elongation marks on bidirectional and unidirectional promoters.
Genome-wide analysis was conducted to score for the enrichment of transcription elongation marks up to 3Kb upstream and downstream from TSS. (A-D) Analysis was performed with expression matched unidirectional and bidirectional genes. Transcriptional elongation marks are enriched on both sense and antisense gene body only on bidirectional promoters while unidirectional promoters show enrichment only in sense orientation.

The profiles of histone modifications are distinct on bidirectional promoters and they mirror the transcriptional status of the genes regulated by the bidirectional promoters. Histone marks associated with transcriptional initiation and transcriptional elongation display a bimodal distribution on active bidirectional promoters. In contrast, active unidirectional promoters show enrichment of these marks only in the direction of matured transcription and not in a bimodal orientation. These data suggest that

transcription initiates in both the directions from bidirectional promoters and matures with elongation in sense and anti-sense.

1.3.7 Bidirectional promoters are depleted of repressive epigenetic marks and exhibit a bimodal distribution for mono-methylation marks for gene pairs

Epigenetic marks regulate transcription via activating and repressing transcription by means of active or repressive histone modifications. H3K27me3, H3K9me3 and H4K20me3 have been shown to be involved in silencing of transcription by different mechanisms. H3K9me3 induces transcriptional silencing by recruitment of HP1 whereas H3K27me3 recruits PolyComb (PcG) complexes and dictates transcriptional silencing (Munshi et al., 2009; van Kruijsbergen et al., 2015). Analysis of bidirectional promoters revealed that H3K27me3, H3K9me3 and H4K20me3 are depleted on gene bodies of bidirectional gene pairs as compared to unidirectional genes (Fig. 1.3.8 A-C). This implies that default chromatin state of bidirectional gene pairs is active. A recent study has shown that the H3K27me1 mark overlaps with H3K36me3 which is enriched in the actively transcribing gene bodies (Ferrari et al., 2014) . H4K20me1 and H3K9me1 have been shown to be enriched on the genes with higher expression and is correlated with transcriptional activation in human CD4 T cells (Wang et al., 2008). As we have observed earlier that bidirectional gene pairs show enrichment of active promoter marks and transcriptional elongation marks in bimodal fashion, we assessed distribution of transcription associated mono methylation marks on the gene bodies of bidirectional genes as compared to unidirectional genes. Our analysis indicates that H3K9me1, H4K20me1 and H3K27me1 are enriched in both sense and antisense gene body. This suggests that chromatin state of bidirectional promoters and their respective gene body is poised for transcriptional activation (Fig. 1.3.8 D-F).

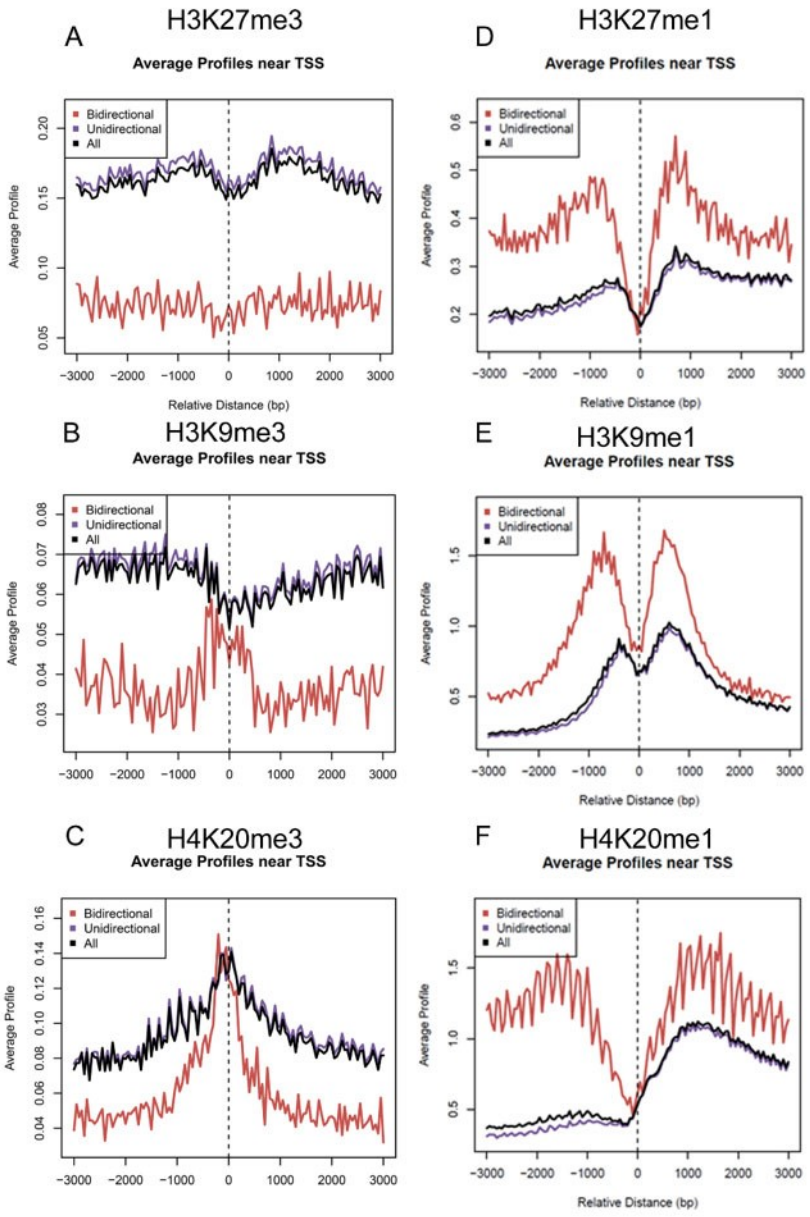


Figure 1.3.8. Genome-wide distribution of repressive marks and mono-methylation marks on bidirectional and unidirectional promoters.
ChIP seq reads for repressive marks and mono-methylation marks on bidirectional and unidirectional promoters were aligned. 3kb upstream and downstream region was selected for the analysis. (A-C) Repressive marks which include H3K27me3, H3K9me3, H4K20me3 shows underrepresentation on bidirectional promoters as compared to unidirectional promoters (B-D). H3K27me1, H3K9me1 and H4K20me1, all three mono-methylation marks which have been shown as marks of transcriptional activation, show higher enrichment on bidirectional promoters as compared to unidirectional promoters.

1.3.8 Transcription from bidirectional promoters is correlated with distinct epigenetic marks

To test the significance of the observed bimodal pattern of active transcription associated histone marks, we decided to study the distribution of epigenetic marks on bidirectional gene pairs where only one gene out of the pair is transcribed. In this situation, the bidirectional promoter would behave similar to a unidirectional promoter with respect to transcription. Here, a bidirectional gene pair which shows expression for only one of the two genes was compared with the unidirectional transcribing genes. Our analysis showed that when both the genes of a bidirectional gene pair are transcribed, active promoter and elongation marks are enriched in both sense and antisense orientations to the TSS of bidirectional gene pair (Fig. 1.3.9). In contrast, active promoter and elongation marks are enriched only in the sense direction on the unidirectional gene. When only one of the gene expresses from bidirectional gene pairs ("OneUP" or "OneDOWN" category), interestingly the enrichment of promoter and transcription elongation marks show a pattern similar to unidirectional promoters. We then analyzed other transcriptional activation and elongation marks for all four categories of bidirectional gene pairs (refer section 1.3.5 in Results). We observed enrichment of H2BK5me1, H3K79me3, H3K36me3 and H4K20me1 in both the gene bodies of bidirectional gene pairs only when both genes are transcribed (Fig. 1.3.8 C, D, E and F). It was found that they show enrichment only in the sense orientation to unidirectional gene. When only one gene is expressed from the bidirectional gene pair, all marks showed enrichment pattern equivalent to unidirectional gene (Fig. 1.3.8 C, E and F). H3K4me1, H3K4me2 also shows the antisense peak only if both genes are expressed, however, if only one of the genes for bidirectional gene pair is expressed then both the promoter marks show enrichment equivalent to a unidirectional gene.

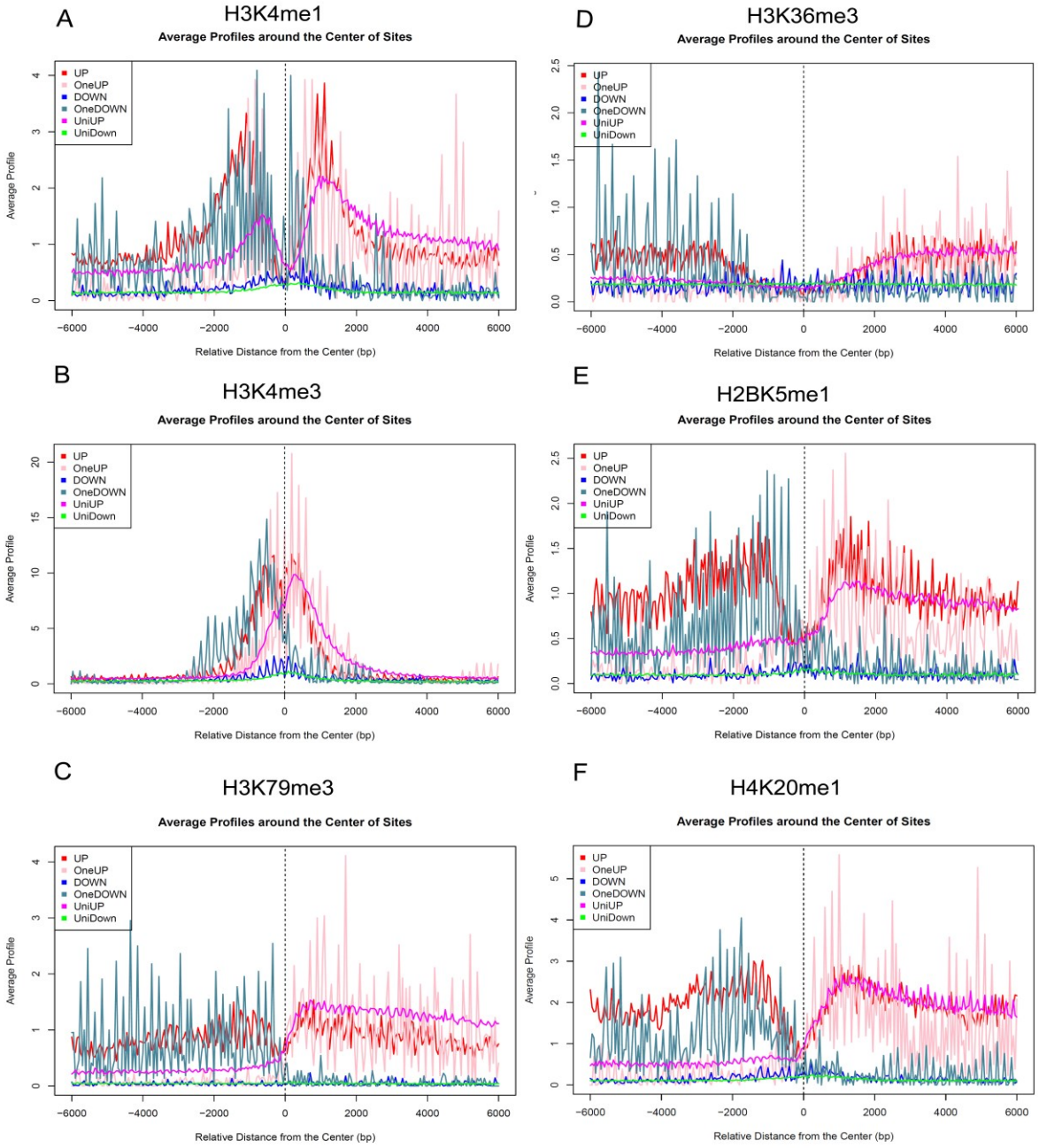


Figure 1.3.9 Enrichment of epigenetic marks correlate with bidirectional transcription. To test if epigenetic marks correlate with transcriptional status of bidirectional gene pairs, we aligned the ChIP-seq reads on bidirectional gene pairs where both gene express, only one gene expresses and when none of the genes express. This was then compared with the unidirectional promoter regulated gene expression. (A-B) Active promoter marks H3K4me1 and H3K4me3 shows enrichment in both sense and antisense direction only to gene pairs where both genes show expression. However if either sense or antisense gene is switched off then the enrichment profile is similar to unidirectional promoter. (C-D) H3K79me3 and H3K36me3 marks of active transcription elongation, show enrichment on the sense and antisense gene body only to gene pairs where both gene express. Enrichment profiles of bidirectional promoters are similar to unidirectional promoters when only one gene out of the bidirectional pair is expressed. (E) H2BK5me1 is known to be enriched downstream to TSS of actively transcribing genes.

Interestingly, H2BK5me1 shows bimodal enrichment only in bidirectional promoters where both genes show expression. In case when only one of gene is expressed, enrichment of H2BK5me1 follows the pattern observed in case of unidirectional promoters. (F) All mono-methylation marks show the same pattern which were observed in case of elongation marks (H3K79me3, H3K36me3).

From above observations, it is evident that the histone marks are associated with active transcription both at the promoters and the gene body that mirrors the transcriptional status of genes. These marks are exclusive to transcriptional activity and even in bidirectional gene pairs where two genes are in a head to head orientation, the transcriptional status of the genes dictate the presence of the corresponding histone marks. Thus, histone modifications are important in driving bidirectional transcriptional initiation and maturation at bidirectional promoters.

1.3.9. Intergenic distance between a bidirectional gene pair is important for bidirectional transcription

A number of studies which have analyzed the intergenic distance between head to head arranged gene pairs in different eukaryotes. These studies revealed that intergenic distance between bidirectional gene pairs are conserved across the species (Davila Lopez et al., 2010; Koyanagi et al., 2005b). It has been reported that nearly 67% of head to head gene pairs have intergenic distance less than 300bp (Trinklein et al., 2004).

Conservation of intergenic regions is a well-known fact but functional importance of maintaining this distance between the two TSS is yet to be deciphered. The distance of the intergenic region between the bidirectionally regulated genes could play an important role in facilitating bidirectional transcription from these promoters. To address this question, we analyzed the gene pairs which are arranged in head to head orientation with differing intergenic distances ranging from 500 bp to 10 Kb. We hypothesized that if intergenic distance conservation is an important factor for bidirectional promoter function then the observed bimodal distribution of different epigenetic marks would be affected as a function of increasing intergenic distance. We analyzed ChIP-seq reads for various histone modifications upto 10kb upstream and downstream of the TSS for these gene

pairs (Fig. 1.3.10). Interestingly, genes with 2 Kb intergenic distances showed above observed bimodal distribution of H3K4me1, H3K4me3, H3K79me3, H3K27me1 and H4K20me1. However, as the intergenic distance increased between head to head gene pairs, we observed a sharp decrease in the active promoter and transcription elongation marks; specifically, in the antisense orientation. This implies that if the intergenic distance is increased between the bidirectional gene pairs, the peak antisense to the orientation of the promoter decreased significantly.

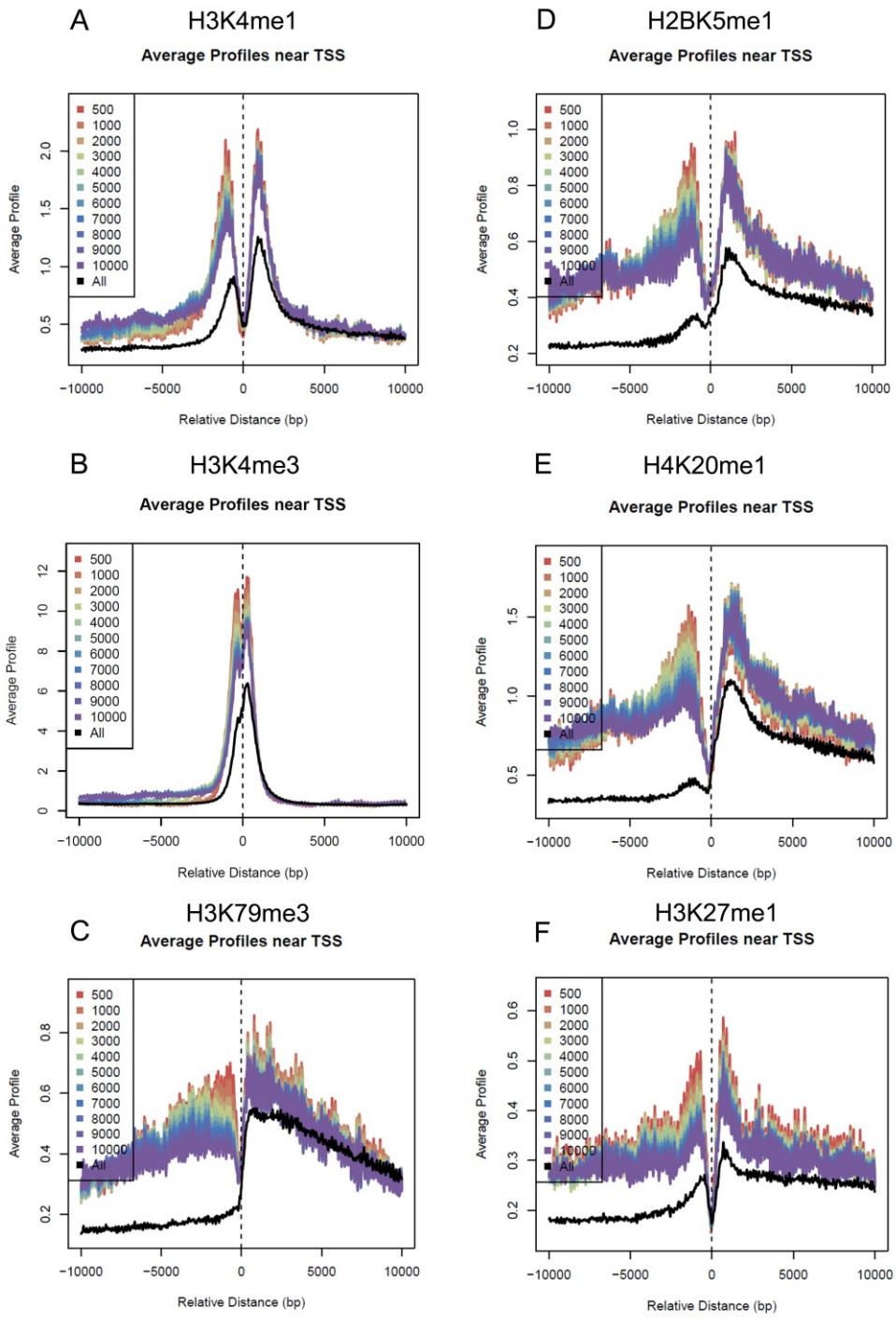


Figure 1.3.10 Intergenic distance between a bidirectional gene pair is important for bidirectional transcription.

To study the significance of intergenic region we analyzed bidirectional gene pairs with varying intergenic region and aligned the ChIP seq reads to 10Kb upstream and downstream to TSS. (A-B) Enrichment of active promoter marks in antisense orientation sharply decreases as the intergenic region between bidirectional gene pairs is increased; suggesting that enrichment of active promoter marks in antisense orientation is a function of intergenic region distance. (C-F) All active transcription elongation marks show sharp decrease only in the antisense orientation as a function of an increase in intergenic distance.

1.3.10. Epigenetic modifications are functionally associated with bidirectional transcription

To address whether the epigenetic modifications discussed earlier would be functionally associated with bidirectional transcription in a biological context, we analyzed the presence of these in retinoic acid (RA) mediated differentiation of NT2D1 cells. We differentiated NT2D1 cells for 7 days and analyzed gene expression from a few bidirectional promoters. We specifically selected X-linked genes for this analysis as NT2D1 cells contain a single copy of the X—chromosome which ensures expression of the bidirectional genes from the same locus and in single copy. By q-PCR analysis we confirmed that gene expression for the gene pair NUP62CL-PIH1D3, which comprises of highly differential genes increases 10 fold upon differentiation (Fig. 1.3.11E). RA mediated differentiation of NT2D1 cells was confirmed by scoring for decrease in pluripotency markers after RA treatment of NT2D1 cells (Fig. 1.3.11A). ChIP for H3K79me3, H3K36me3 and H3K27me1 was performed after 7 days of differentiation. A region spanning 2 Kb on either side of the TSS for each gene was analyzed which included the intergenic bidirectional promoter (INT). A concomitant increase in these marks on the gene bodies of both the genes was observed (Fig. 1.3.11 B-D) which correlated perfectly with the increase in transcription from these genes. As expected the intergenic region (INT) did not have any enrichment of the marks associated with transcriptional elongation.

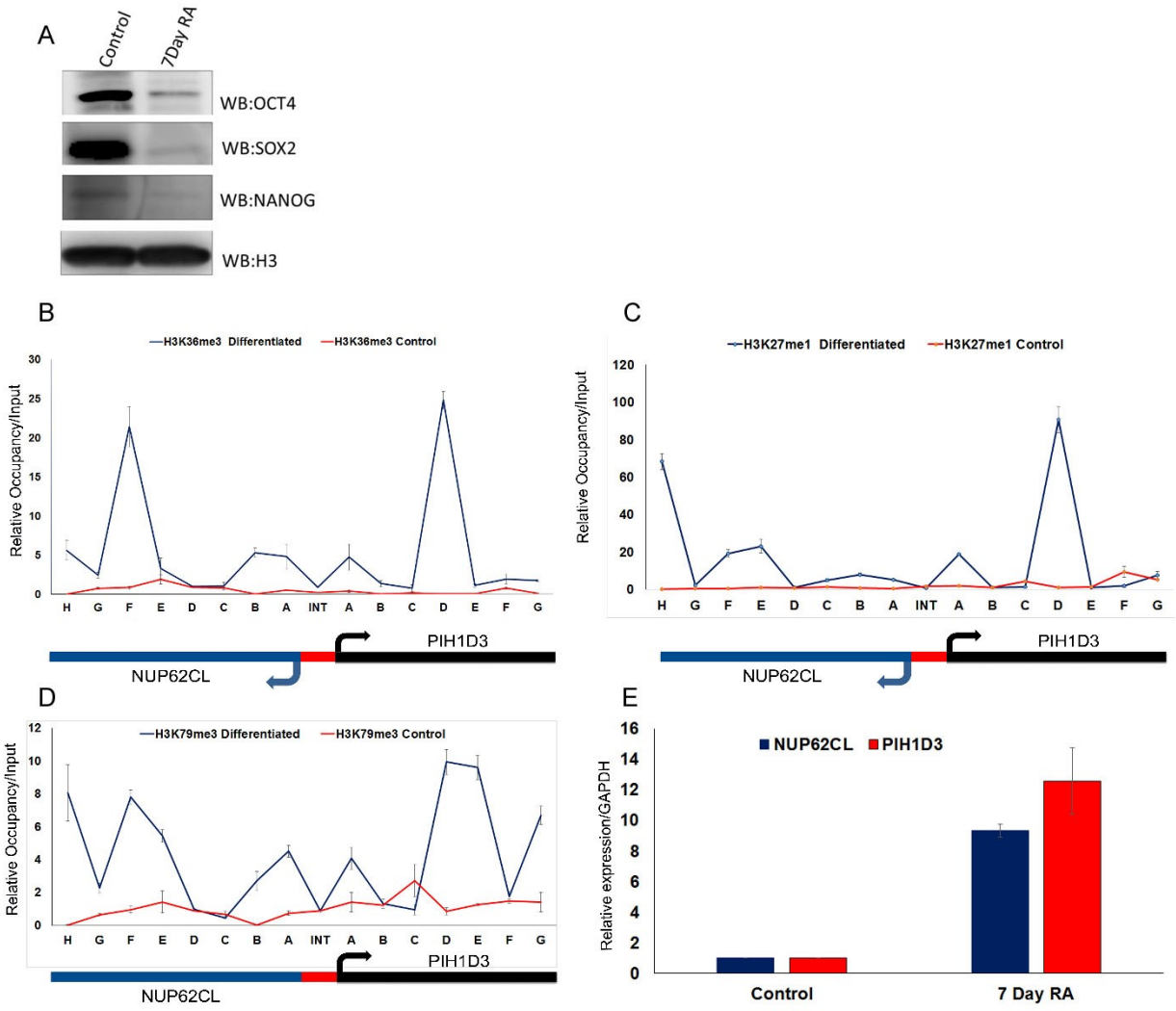


Figure 1.3.11 Epigenetic marks regulate bidirectional transcription in cellular context. To study the role of epigenetics in regulation of bidirectional promoters in cellular context, NT2D1 cells were differentiated with RA, followed by western blot and qPCR. (A) We confirmed the differentiation of NT2D1 cells by western blot; differentiated cells show a decrease in expression of the pluripotency factors OCT4, SOX2 and Nanog. (B-D) ChIP analysis show that all three elongation marks analyzed in this experiment (H3K36me3, H3K27me1 and H3K79me3) show increased enrichment in both sense and antisense gene body after differentiation. (E) qPCR analysis showing the differential expression of NUP62CL-PIH1D3 gene pair in control and differentiated cells.

1.4. Discussion

This study focuses on analyzing the mechanism of transcriptional elongation from bidirectional promoters. These promoters have the unique property to generate transcripts in both sense and antisense orientations, which are mature and functional in the cell. The regulation of bidirectional transcription is not clearly understood and this study aims at answering this question from an epigenetic perspective.

Until now, luciferase system was used to characterize bidirectional promoter. However, there is a one major limitation in that system, one can score for only single directional firing of promoter at one time. To overcome this limitation, we have constructed a vector system that can be used for studying bidirectional promoters. In this vector system, any DNA fragment acting as a bonafide bidirectional promoter can fire two reporters (GFP and mCherry) simultaneously; and a unidirectional promoter can fire either of the two reporters based on its directionality. We have named this vector as the pDR (plasmid with dual reporter) vector. We validated this vector *in vivo* by cloning CMV promoter in sense and antisense orientation with respect to the reporter genes. CMV is a unidirectional promoter; therefore, it drives the expression of GFP or mCherry depending on whether this promoter was cloned in sense or antisense orientation between the two reporters. Further, we cloned 8 bidirectional promoters from human genome in pDR vector and characterized them *in vivo* by microscopy and FACS.

A study by Bornelöv et al., 2015 has focused on differential distribution of histone modifications on unidirectional and bidirectional protein coding genes and report that H3K4me3, H3K9ac, and H3K27ac marks are enriched more on antisense direction on bidirectional genes as compared to unidirectional genes. However, they did not observe any differential enrichment in the sense orientation (Bornelöv et al., 2015). The authors have also commented that antisense enrichment of these active transcription marks might be a consequence of antisense transcription rather than the cause (Bornelöv et al., 2015). By using murine macrophages as a model system, a study by Scruggs et al., 2015 shows that bidirectional transcription arises from two distinct hubs of transcription factor binding. Lepoivre et al., 2013 has also focused on the nature of bidirectional

promoter regulated genes and has shown that most of the bidirectional promoter driven genes are associated with promoters of transcriptional regulators. We tried to decipher if there is any link between epigenetic modifications and bidirectional transcription. Our ChIP experiments in Jurkat cells shows that bidirectional promoters harbor unique distribution of active transcription associated promoter and elongation marks in antisense orientation as compared to expression matched unidirectional promoters. We validated this observation *in vitro* as well by using one of our previously validated bidirectional promoter pDR clone. For this experiment CMV promoter clone in pDR vector was used as a control unidirectional promoter. Interestingly, all the active elongation marks were enriched in both sense and antisense orientation only on bidirectional promoter clones while unidirectional CMV promoter clones showed enrichment of active transcription elongation marks only in sense direction to transcription. This prompted us to think that this unique distribution of active transcription marks on bidirectional promoters and their respective gene body might be responsible for mature antisense transcription elongation from such loci.

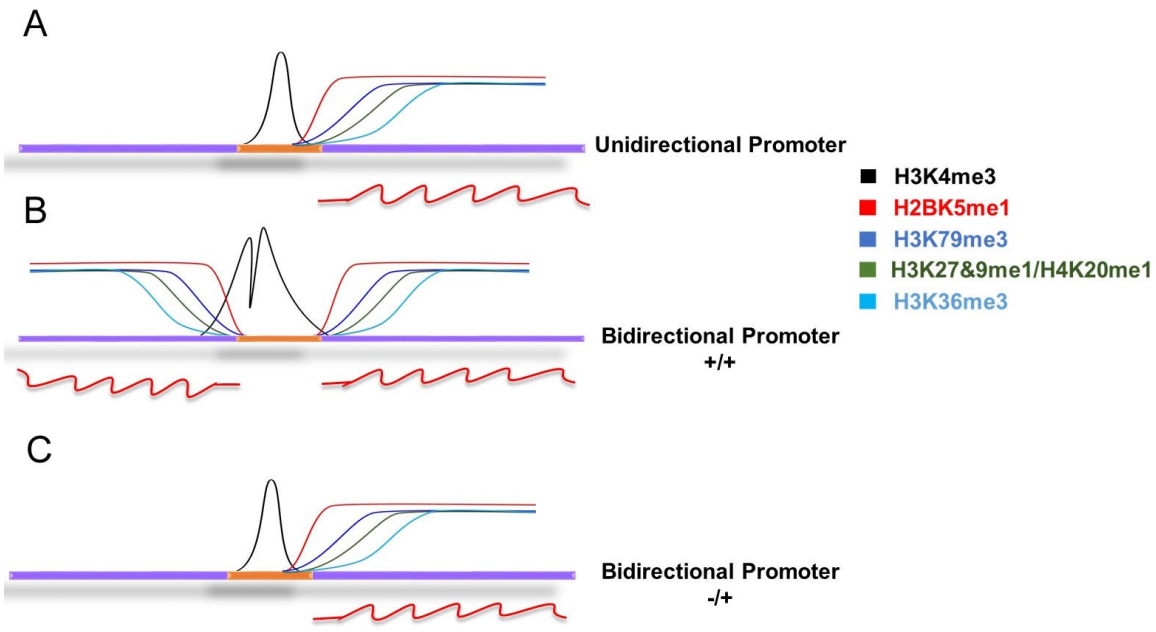


Figure 1.4 Model depicting the role of epigenetics in bidirectional promoter regulation. (A) Unidirectional promoters show mature transcription only in sense direction and active promoter marks are enriched on the promoter region as unimodal peak. Transcriptional elongation marks show (H3K79me3, H3K36me3) enrichment only in the sense orientation and no enrichment of elongation marks

in antisense orientation is observed. Mono-methylation marks which are marks of transcriptional activation are enriched only in the sense direction of gene body. **(B)** Bidirectional promoters which show expression of both sense and antisense genes, display bimodal distribution of active promoter marks. Elongation marks (H3K79me3, H3K36me3), and active transcription associated mono-methylation marks show enrichment in both sense and antisense gene body, suggesting that differential enrichment of epigenetic marks allows the transcription machinery to transcribe in both sense and antisense orientation. **(C)** Bidirectional promoters where only one gene is expressed behave same as the unidirectional promoters in context of epigenetic modifications. All the epigenetic marks studied here show similar kind of profile as observed in unidirectional promoters.

To test our hypothesis comprehensively, we analyzed 39 epigenetic modifications on all bidirectional promoters and compared it to that of all other promoters in the human CD4 T cells. Only the promoters whose genes had similar expression levels were compared between the two categories. We found that the histone marks H3K4me1, H3K4me3, H3K23ac, H2AK9ac and H3K36me1 show enrichment in a bimodal fashion on the bidirectional promoters whereas they show enrichment only in the sense orientation on unidirectional promoters. H2BK5me1 which has been reported to occur downstream of actively transcribing promoters (Barski et al., 2007a) was observed to occur in a pattern concomitant with active transcription on bidirectional promoters. H3K27me1, H4K20me1, H3K36me3 and H3K79me3, which are marks associated with successful mature transcription, occur in a bimodal distribution only on bidirectional promoters. This pattern overlays with the distribution of RNA-seq tags on the genes regulated by these promoters. In cases where one of the genes of the bidirectional pair is not expressed, all the listed histone marks exhibit a profile similar to that of unidirectional promoters (Fig. 4.1). These findings strongly imply that the occurrence of these marks is correlated with the process of active transcription maturation from the bidirectional promoters. The data regarding the epigenetic marks on the bidirectional promoters is conclusive with respect to the epigenetic state of the bidirectional promoters.

Our data identifies an epigenetic signature of bidirectional promoters that sets them apart from all other transcribing loci in the genome. We propose that the bimodally distributed chromatin marks that occur flanking the transcription start site (TSS) could facilitate mature transcription from these sites in both sense and antisense orientations. We also addressed the functional importance of conservation of intergenic region in bidirectional promoters. We analyzed the distribution of epigenetic marks on

bidirectionally arranged gene pairs with intergenic distance from 500 bp to 10 Kb. We observed a sharp decrease in the antisense peak of active promoter and transcription elongation associated marks with an increase in intergenic region between two genes of a bidirectional gene pair.

Furthermore, to see if these epigenetic marks have an influence on bidirectional transcription in cellular context, we used RA mediated differentiation of NT2D1 as the model system and analyzed one of the bidirectional gene pair (NUP62CL-PIH1D3) for various transcription elongation marks (H3K79me3, H3K36me3 and H3K27me1). We analyzed the enrichment of these elongation marks in 2 Kb upstream and downstream regions from TSS of NUP62CL-PIH1D3 gene pair including the intergenic region. Interestingly, all the 3 elongation marks revealed significantly high enrichment downstream to TSS on both gene bodies further strengthening the significance of bimodal distribution of elongation marks in mature bidirectional transcription from sense and antisense gene.

This is for the first time that we demonstrate a strong link between the epigenetic marks on the bidirectional promoters and the transcriptional state of the bidirectional genes. Our data identifies an epigenetic signature of bidirectional promoters that sets them apart from all other transcribing loci in the genome. Observed bimodal distribution of epigenetic marks might be one of unknown mechanisms in the field which play an important role in mature sense and antisense transcription from bidirectional promoters. We propose that unique distribution of active transcription marks on bidirectional promoters signal the transcription machinery to drive the transcription in both sense and antisense orientation. However, these active epigenetic marks show enrichment only in one orientation on unidirectional promoters and therefore transcription maturation progresses only in one direction.

1.5 References

Abraham, B.J., Cui, K., Tang, Q., and Zhao, K. (2013). Dynamic regulation of epigenomic landscapes during hematopoiesis. BMC genomics *14*, 193.

Adachi, N., and Lieber, M.R. (2002). Bidirectional gene organization: a common architectural feature of the human genome. Cell *109*, 807-809.

Ahn, J., and Gruen, J.R. (1999). The genomic organization of the histone clusters on human 6p21. 3. Mammalian genome *10*, 768-770.

Albig, W., Kioschis, P., Poustka, A., Meergans, K., and Doenecke, D. (1997). Human histone gene organization: nonregular arrangement within a large cluster. Genomics *40*, 314-322.

Almada, A.E., Wu, X., Kriz, A.J., Burge, C.B., and Sharp, P.A. (2013). Promoter directionality is controlled by U1 snRNP and polyadenylation signals. Nature *499*, 360-363.

Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007a). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007b). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Behura, S.K., and Severson, D.W. (2015). Bidirectional Promoters of Insects: Genome-Wide Comparison, Evolutionary Implication and Influence on Gene Expression. Journal of molecular biology *427*, 521-536.

Bornelöv, S., Komorowski, J., and Wadelius, C. (2015). Different distribution of histone modifications in genes with unidirectional and bidirectional transcription and a role of CTCF and cohesin in directing transcription. BMC genomics *16*, 300.

Brandeis, M., Frank, D., Keshet, I., Siegfried, Z., Mendelsohn, M., Temper, V., Razin, A., and Cedar, H. (1994). Spl elements protect a CpG island from de novo methylation.

Breathnach, R., and Chambon, P. (1981). Organization and expression of eucaryotic split genes coding for proteins. Annual review of biochemistry *50*, 349-383.

Bucher, P., and Trifonov, E.N. (1986). Compilation and analysis of eukaryotic POL II promoter sequences. Nucleic acids research *14*, 10009-10026.

Burke, T.W., and Kadonaga, J.T. (1996). Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. Genes & development *10*, 711-724.

Burke, T.W., and Kadonaga, J.T. (1997). The downstream core promoter element, DPE, is conserved fromDrosophila to humans and is recognized by TAFII60 of Drosophila. Genes & development *11*, 3020-3031.

Butler, J.E., and Kadonaga, J.T. (2002). The RNA polymerase II core promoter: a key component in the regulation of gene expression. Genes & development *16*, 2583-2592.

Carcamo, J., Buckbinder, L., and Reinberg, D. (1991). The initiator directs the assembly of a transcription factor IID-dependent transcription complex. Proceedings of the National Academy of Sciences *88*, 8052-8056.

Davila Lopez, M., Martinez Guerra, J.J., and Samuelsson, T. (2010). Analysis of gene order conservation in eukaryotes identifies transcriptionally and functionally linked genes. PloS one *5*, e10654.

Dhadi, S.R., Krom, N., and Ramakrishna, W. (2009). Genome-wide comparative analysis of putative bidirectional promoters from rice, Arabidopsis and Populus. Gene *429*, 65-73.

Engström, P.G., Suzuki, H., Ninomiya, N., Akalin, A., Sessa, L., Lavorgna, G., Brozzi, A., Luzi, L., Tan, S.L., and Yang, L. (2006). Complex loci in human and mouse genomes. PLoS genetics *2*, e47.

Ferrari, K.J., Scelfo, A., Jammula, S., Cuomo, A., Barozzi, I., Stützer, A., Fischle, W., Bonaldi, T., and Pasini, D. (2014). Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity. Molecular cell *53*, 49-62.

Franck, E., Hulsen, T., Huynen, M.A., de Jong, W.W., Lubsen, N.H., and Madsen, O. (2008). Evolution of closely linked gene pairs in vertebrate genomes. Molecular biology and evolution *25*, 1909-1921.

Goldberg, A.D., Allis, C.D., and Bernstein, E. (2007). Epigenetics: a landscape takes shape. Cell *128*, 635-638.

Grunstein, M., Hecht, A., Fisher-Adams, G., Wan, J., Mann, R., Strahl-Bolsinger, S., Laroche, T., and Gasser, S. (1995). The regulation of euchromatin and heterochromatin by histones in yeast. Journal of Cell Science *1995*, 29-36.

Hayes, J.J., and Hansen, J.C. (2001). Nucleosomes and the chromatin fiber. Current opinion in genetics & development *11*, 124-129.

He, Y., Vogelstein, B., Velculescu, V.E., Papadopoulos, N., and Kinzler, K.W. (2008). The antisense transcriptomes of human cells. Science *322*, 1855-1857.

Kaufmann, J., and Smale, S.T. (1994). Direct recognition of initiator elements by a component of the transcription factor IID complex. Genes & development *8*, 821-829.

Koch, C.M., Andrews, R.M., Flicek, P., Dillon, S.C., Karaoz, U., Clelland, G.K., Wilcox, S., Beare, D.M., Fowler, J.C., Couttet, P.*, et al.* (2007). The landscape of histone modifications across 1% of the human genome in five human cell lines. Genome research *17*, 691-707.

Koonin, E.V. (2009). Evolution of genome architecture. The international journal of biochemistry & cell biology *41*, 298-306.

Kornberg, R.D., and Lorch, Y. (1999). Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. Cell *98*, 285-294.
Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693-705.

Koyanagi, K.O., Hagiwara, M., Itoh, T., Gojobori, T., and Imanishi, T. (2005a). Comparative genomics of bidirectional gene pairs and its implications for the evolution of a transcriptional regulation system. Gene *353*, 169-176.

Koyanagi, K.O., Hagiwara, M., Itoh, T., Gojobori, T., and Imanishi, T. (2005b). Comparative genomics of bidirectional gene pairs and its implications for the evolution of a transcriptional regulation system. Gene *353*, 169-176.

Krogan, N.J., Kim, M., Tong, A., Golshani, A., Cagney, G., Canadien, V., Richards, D.P., Beattie, B.K., Emili, A., Boone, C*., et al.* (2003). Methylation of Histone H3 by Set2 in Saccharomyces cerevisiae Is Linked to Transcriptional Elongation by RNA Polymerase II. Molecular and cellular biology *23*, 4207-4218.

Krom, N., and Ramakrishna, W. (2008). Comparative analysis of divergent and convergent gene pairs and their expression patterns in rice, Arabidopsis, and Populus. Plant physiology *147*, 1763-1773.

Lee, T.I., and Young, R.A. (2000). Transcription of eukaryotic protein-coding genes. Annual review of genetics *34*, 77-137.

Li, J., Moazed, D., and Gygi, S.P. (2002). Association of the histone methyltransferase Set2 with RNA polymerase II plays a role in transcription elongation. The Journal of biological chemistry *277*, 49383-49388.

Li, Y.-Y., Yu, H., Guo, Z.-M., Guo, T.-Q., Tu, K., and Li, Y.-X. (2006). Systematic analysis of head-to-head gene organization: evolutionary conservation and potential biological relevance. PLoS Comput Biol *2*, e74.

Li, Y., Sun, L., Zhang, Y., Wang, D., Wang, F., Liang, J., Gui, B., and Shang, Y. (2011). The histone modifications governing TFF1 transcription mediated by estrogen receptor. The Journal of biological chemistry *286*, 13925-13936.

Lin, J.M., Collins, P.J., Trinklein, N.D., Fu, Y., Xi, H., Myers, R.M., and Weng, Z. (2007a). Transcription factor binding and modified histones in human bidirectional promoters. Genome research *17*, 818-827.

Lin, J.M., Collins, P.J., Trinklein, N.D., Fu, Y., Xi, H., Myers, R.M., and Weng, Z. (2007b). Transcription factor binding and modified histones in human bidirectional promoters. Genome research *17*, 818-827.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature *389*, 251-260.

Macleod, D., Charlton, J., Mullins, J., and Bird, A.P. (1994). Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. Genes & development *8*, 2282-2292.

Martinez, E., Chiang, C.-M., Ge, H., and Roeder, R. (1994). TATA-binding protein-associated factor (s) in TFIID function through the initiator to direct basal transcription from a TATA-less class II promoter. The EMBO journal *13*, 3115.

Maxson, R., Cohn, R., Kedes, L., and Mohun, T. (1983). Expression and organization of histone genes. Annual review of genetics *17*, 239-277.

Momota, R., Sugimoto, M., Oohashi, T., Kigasawa, K., Yoshioka, H., and Ninomiya, Y. (1998). Two genes, COL4A3 and COL4A4 coding for the human α3 (IV) and α4 (IV) collagen chains are arranged head-to-head on chromosome 2q36. FEBS letters *424*, 11-16.

Munshi, A., Shafi, G., Aliya, N., and Jyothy, A. (2009). Histone modifications dictate specific biological readouts. Journal of Genetics and Genomics *36*, 75-88.

Ntini, E., Järvelin, A.I., Bornholdt, J., Chen, Y., Boyd, M., Jørgensen, M., Andersson, R., Hoof, I., Schein, A., and Andersen, P.R. (2013a). Polyadenylation site–induced decay of upstream transcripts enforces promoter directionality. Nature structural & molecular biology *20*, 923-928.

Ntini, E., Jarvelin, A.I., Bornholdt, J., Chen, Y., Boyd, M., Jorgensen, M., Andersson, R., Hoof, I., Schein, A., Andersen, P.R*., et al.* (2013b). Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. Nature structural & molecular biology *20*, 923-928.

Ohtani, K., and Dimmeler, S. (2011). Epigenetic regulation of cardiovascular differentiation. Cardiovascular research, cvr019.

Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S., Mapendano, C.K., Schierup, M.H., and Jensen, T.H. (2008). RNA exosome depletion reveals transcription upstream of active human promoters. Science *322*, 1851-1854.

Rocha, E.P. (2008). The organization of the bacterial genome. Annual review of genetics *42*, 211-233.

Sandelin, A., Carninci, P., Lenhard, B., Ponjavic, J., Hayashizaki, Y., and Hume, D.A. (2007). Mammalian RNA polymerase II core promoters: insights from genome-wide studies. Nature Reviews Genetics *8*, 424-436.

Schaft, D., Roguev, A., Kotovic, K.M., Shevchenko, A., Sarov, M., Shevchenko, A., Neugebauer, K.M., and Stewart, A.F. (2003). The histone 3 lysine 36 methyltransferase, SET2, is involved in transcriptional elongation. Nucleic acids research *31*, 2475-2482.

Schmidt, C., Fischer, G., Kadner, H., Genersch, E., Kühn, K., and Pöschl, E. (1993). Differential effects of DNA-binding proteins on bidirectional transcription from the common promoter region of human collagen type IV genes COL4A1 and COL4A2. Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression *1174*, 1-10.

Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., and Sharp, P.A. (2008). Divergent transcription from active promoters. Science *322*, 1849-1851.

Seila, A.C., Core, L.J., Lis, J.T., and Sharp, P.A. (2009). Divergent transcription: a new feature of active promoters. Cell cycle *8*, 2557-2564.

Shin, H., Liu, T., Manrai, A.K., and Liu, X.S. (2009). CEAS: cis-regulatory element annotation system. Bioinformatics *25*, 2605-2606.

Sperling, S. (2007). Transcriptional regulation at a glance. BMC bioinformatics *8*, S2.

Steger, D.J., Lefterova, M.I., Ying, L., Stonestrom, A.J., Schupp, M., Zhuo, D., Vakoc, A.L., Kim, J.E., Chen, J., Lazar, M.A.*, et al.* (2008). DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells. Molecular and cellular biology *28*, 2825-2839.

Sugimoto, M., Oohashi, T., and Ninomiya, Y. (1994). The genes COL4A5 and COL4A6, coding for basement membrane collagen chains alpha 5 (IV) and alpha 6 (IV), are located head-to-head in close proximity on human chromosome Xq22 and COL4A6 is transcribed from two alternative promoters. Proceedings of the National Academy of Sciences *91*, 11679-11683.

Takai, D., and Jones, P.A. (2004). Origins of bidirectional promoters: computational analyses of intergenic distance in the human genome. Molecular biology and evolution *21*, 463-467.

Trinklein, N.D., Aldred, S.F., Hartman, S.J., Schroeder, D.I., Otillar, R.P., and Myers, R.M. (2004). An abundance of bidirectional promoters in the human genome. Genome research *14*, 62-66.

Tsai, F.T., and Sigler, P.B. (2000). Structural basis of preinitiation complex assembly on human pol II promoters. The EMBO journal *19*, 25-36.

Tsai, H.-K., Su, C.P., Lu, M.-Y.J., Shih, C.-H., and Wang, D. (2007). Co-expression of adjacent genes in yeast cannot be simply attributed to shared regulatory system. BMC genomics *8*, 352.

van Kruijsbergen, I., Hontelez, S., and Veenstra, G.J. (2015). Recruiting polycomb to chromatin. The international journal of biochemistry & cell biology.

Wakano, C., Byun, J.S., Di, L.J., and Gardner, K. (2012). The dual lives of bidirectional promoters. Biochimica et biophysica acta *1819*, 688-693.

Wang, G., Qi, K., Zhao, Y., Li, Y., Juan, L., Teng, M., Li, L., Liu, Y., and Wang, Y. (2013). Identification of regulatory regions of bidirectional genes in cervical cancer. BMC medical genomics *6 Suppl 1*, S5.

Wang, Q., Wan, L., Li, D., Zhu, L., Qian, M., and Deng, M. (2009). Searching for bidirectional promoters in Arabidopsis thaliana. BMC bioinformatics *10*, S29.

Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Peng, W., Zhang, M.Q.*, et al.* (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet *40*, 897-903.

Wasylyk, B., Derbyshire, R., Guy, A., Molko, D., Roget, A., Teoule, R., and Chambon, P. (1980). Specific in vitro transcription of conalbumin gene is drastically decreased by single-point mutation in TATA box homology sequence. Proceedings of the National Academy of Sciences *77*, 7024-7028.

Weis, L., and Reinberg, D. (1997). Accurate positioning of RNA polymerase II on a natural TATA-less promoter is independent of TATA-binding-protein-associated factors and initiator-binding proteins. Molecular and cellular biology *17*, 2973-2984.

Yang, L., and Yu, J. (2009). A comparative analysis of divergently-paired genes (DPGs) among Drosophila and vertebrate genomes. BMC evolutionary biology *9*, 55.

Yang, M.Q., and Elnitski, L.L. (2008a). Diversity of core promoter elements comprising human bidirectional promoters. BMC genomics *9*, S3.

Yang, M.Q., and Elnitski, L.L. (2008b). Diversity of core promoter elements comprising human bidirectional promoters. BMC genomics *9 Suppl 2*, S3.

Yang, M.Q., and Elnitski, L.L. (2008c). Prediction-based approaches to characterize bidirectional promoters in the mammalian genome. BMC genomics *9*, S2.

Chapter 2

2.1 Introduction

Bidirectional promoters have received considerable interest in the recent past due to their property of regulating two genes from common intergenic region. In the first chapter, I have discussed the properties of a bidirectional promoter and associated epigenetic marks that contribute to the bidirectional nature of these promoters (Chapter 1). In addition to these specific DNA associated features, specific transcription factors also regulate the transcription from bidirectional promoters. Multiple studies have focused on the regulatory mechanisms governing transcription from such loci; however, mechanistic insights are still lacking (Lin et al., 2007b; Trinklein et al., 2004). Alignments of the sequences of bidirectional promoters from human genome to identify if these promoters are enriched for any particular motif as compared to the unidirectional promoters yielded few enriched motifs which were then categorized as underrepresented, shared and over-represented (Lin et al., 2007a). It was found that most of the common eukaryotic transcription factors lack binding sites on the bidirectional promoters and only a small set of motifs show over represented enrichment on bidirectional promoters. These motifs are for the transcription factors namely GABPA, MYC, E2F1, E2F4, NRF-1, CCAAT and YY1. Interestingly, SP1 binding motifs show similar degree of enrichment on both unidirectional as well as bidirectional promoters (Lin et al., 2007a). Further studies were performed to analyze the role of individual transcription factors in regulation of bidirectional promoters. The roles of some of these factors are discussed more elaborately in the following section.

2.1.1 GA-binding protein transcription factor (GABPA)

GA-binding protein transcription factor (GABPA) has been shown to bind nearly 80% of bidirectional promoters in the human genome from the studies carried out in

HeLa, Jurkat, and K562 cell lines (Collins et al., 2007). GABPA belongs to ETS family of transcription factors which are present in species ranging from sponges to human (Sementchenko and Watson, 2000). All ETS family transcription factors harbor a conserved 85 amino acid long DNA binding motif which is known as ETS domain (Fig 2.1.1) (Sementchenko and Watson, 2000). Occurrence of GABPA binding site is correlated with bidirectional transcription and the same has been validated by luciferase reporter assay. It has been clearly demonstrated that any unidirectional promoter when cloned in dual luciferase reporter vector system exhibits reporter activity only in one direction; however, addition of GABPA consensus binding sites to same promoter results in bidirectional reporter activity. This finding suggests that the presence of GABPA consensus binding site promotes bidirectional transcription. (Collins et al., 2007). Moreover, binding of GABPA was verified with 121 and 291 randomly selected bidirectional and unidirectional promoters respectively in three different human cell lines (HeLa, Jurkat, and K562) by performing chromatin immunoprecipitation analysis. Although the referred study provides evidence showing positive correlation between occurrence of GABPA site and the bidirectionality of the promoter, the molecular mechanism underlying this regulation is not well understood.

Interestingly, expression of GABPA itself is driven by a bidirectional promoter (John Patton et al 2005). GABPA has been shown to cooperate with YY1 and bind the intergenic region to regulate transcription from PREPL-C2ORF34 bidirectional gene pair. (Huang and Chang, 2009). The authors also show that the disruption of binding site of any of these transcription factors by mutation results in dysregulation in transcription from this locus indicating the importance of these factors in regulating the stated bidirectional gene pair (Huang and Chang, 2009).

2.1.2 Nuclear respiratory factor (NRF1)

Nuclear respiratory factor (NRF1) has been shown to regulate a number of biological pathways such as antioxidant response, purine biosynthesis and mitochondrial biogenesis (Biswas and Chan, 2010; Chen et al., 1997b; Kelly and Scarpulla, 2004).

Though bidirectional promoters are enriched for NRF1 motif (Lin et al., 2007b), the role of NRF1 in bidirectional promoter regulation is poorly understood. It is noteworthy that there has been only one study till date demonstrating the binding to NRF1 to GPAT-AIRC bidirectional promoter (Chen et al., 1997b). This study shows that the binding of NRF1 to intergenic region of these two genes stabilize SP1 binding, thereby, regulating gene expression from this locus (Chen et al., 1997b).

2.1.3 Nuclear Transcription Factor Y, Alpha (NFYA)

NF-YA is a member of a heterotrimeric transcription factors (NFY A & B & C). It binds to CCAAT boxes in the promoter region of several genes (Serra et al., 1998) and hence is also known as CCAAT-box binding factor (CBF) (Fig 2.1.1) and CCAAT binding protein-1 (CP1). Nuclear Transcription Factor Y, Alpha (NFYA) has been shown to influence the directionality of transcription from Mrps12-Sarsm gene pair in both human and mouse cells (Ernesto Zanotto et al., 2009). It has been shown to regulate PRR11-SKA2 bidirectional gene pair (Wang Y et al., 2015). Genome wide binding analysis of NFYA to unidirectional and bidirectional promoter revealed that bidirectional promoters have specific allocation of CCAAT boxes to which NF-YA binds (Häkkinen et al., 2011). This unique allocation of CCAAT boxes on bidirectional promoters suggests for a prominent role of NFYA in bidirectional promoter regulation.

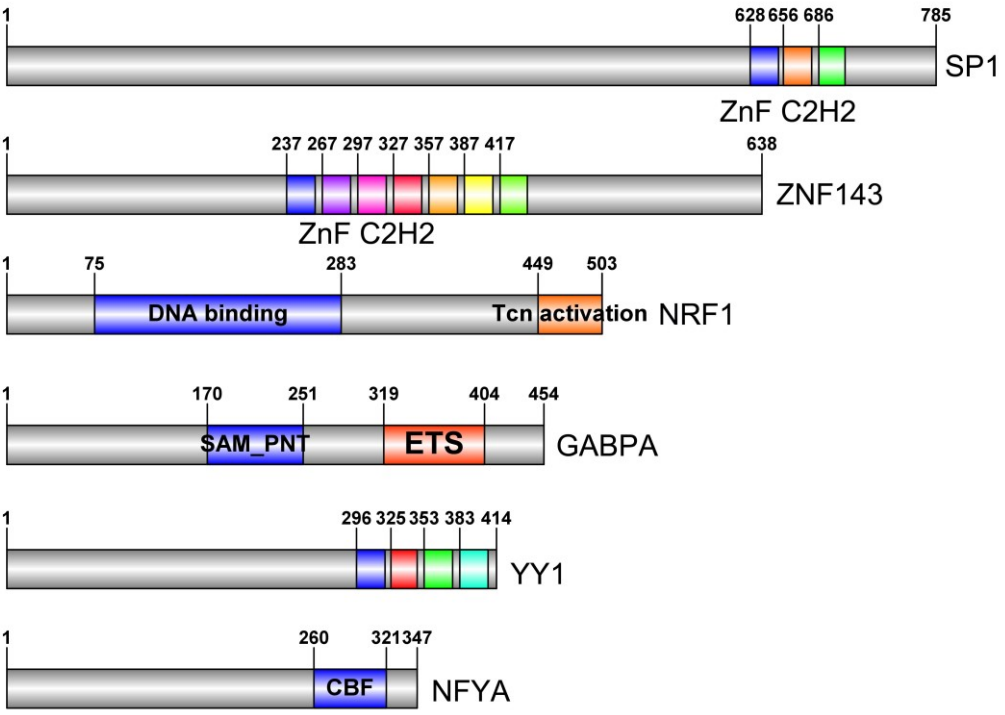


Figure 2.1.1 Domain organization of SP1, ZNF143, NRF1, GABPA, YY1 and NFYA. Cartoon depicting the domain organization of the transcription factors SP1, ZNF143, NRF1, GABPA, YY1 and NFYA. All the characterized domains and their respective positions are marked by corresponding amino acid residues.

2.1.4 Zinc Finger Protein 143 (ZNF143)

Zinc Finger Protein 143 (ZNF143) is a Zinc finger related transcription family member (Fig 2.1.1) and has been shown to mediate long-range interactions between gene promoters with distal regulatory elements. It directly binds to the promoters and allows lineage specific chromatin interactions and gene expression (Bailey et al., 2015). ZNF143 consensus sequence is one of the most widely occurring transcription factor motif present in mammalian genome which is represented at 2500 sites on 2000 promoters (Myslinski et al., 2006). A recent study has discovered that ZNF143 binds and regulates a subset of bidirectional promoters (Anno et al., 2011b). ZNF143 binding motifs are over represented on bidirectional promoters as compared to the unidirectional promoters (Anno et al., 2011b). Knockdown of ZNF143 results in dysregulation of

ZNF143 bound bidirectional gene pairs suggesting the importance of this factor in bidirectional promoter regulation (Anno et al., 2011a).

2.1.5 Yin Yang 1 (YY1)

Yin Yang 1 **(**YY1) is a ubiquitous transcription factor and belongs to the GLI-Kruppel class of zinc finger proteins (Fig 2.1.1). YY1 is involved in a variety of biological processes such as embryogenesis, cell differentiation, replication, and cellular proliferation (Donohoe et al., 1999; Palko et al., 2004; Petkova et al., 2001; Shi et al., 1997). YY1 has been shown to bind and regulate the expression of human Surf-1-Surf-2 bidirectional gene pair in response to serum growth factors (Cole and Gaston, 1997; Gaston and Fried, 1994). Furthermore, CpG methylation has differential effects on binding of YY1 and ETS related proteins to human Surf-1-Surf-2 bidirectional promoter (Cole and Gaston, 1997). Another study showed that YY1 cooperates with MYC to positively regulate the expression of Surf-1-Surf-2 gene pair (Vernon and Gaston, 2000). GABPA and ATP synthase coupling factor 6 bidirectional gene pair harbors binding sites for YY1, NRF1, SP1 and GABPA (Chinenov et al., 2000), suggesting a possible cross talk between these factors in the regulation of bidirectional promoters. Apart from binding to bidirectional promoters, nothing is known about the mechanistic role played by YY1 in regulating transcription from bidirectional promoters.

2.1.6 Specificity Protein (SP1)

Specificity Protein (SP1) is also an ubiquitously expressed transcription factor which possesses 3 C2H2-type zinc finger motifs (Fig 2.1.1) as DNA-binding domain (Kadonaga et al., 1987; Oka et al., 2004). SP1 binds to GC- and CT-boxes both, however affinity for CT-boxes is significantly lower as compared to GC-boxes (Briggs et al., 1986; Kadonaga et al., 1986). In addition to the DNA binding domain, SP1 also harbors two transactivation domains (TAD) known as domain A and domain B (Courey and Tjian, 1988; Pascal and Tjian, 1991) which directly interact with TATA-binding protein (TBP) and TATA associated factors (TAFs) (Chen et al., 1994; Emili et al., 1994).

SP1 has been shown to regulate the transcription of the bidirectional gene pair SIRT3-PSMD13 by binding to their shared intergenic promoter region (Bellizzi et al., 2007). In another study, bidirectional promoter of human monoamine oxidase A (MAO A) has been shown to be regulated by SP1 (Zhu et al., 1994). Interestingly, Human Ly49 class I receptors have been shown to be controlled by bidirectional promoters and harbor binding sites for SP1, YY1 and ETS family for transcription factors (Davies et al., 2007).

2.1.7 SAGA complex

Transcription factors, upon induction, bind to target gene loci, where they lead to recruitment of chromatin remodeling complexes like co-activators and co-repressors. These remodeling complexes modulate the chromatin structure locally to bring about transcriptional regulation at promoter elements. Transcriptional co-activators work as a mediator between external signals and transcription machinery. Several co-activator complexes have been shown to acetylate histones, suggesting that activators may be involved in targeting histone acetylation to promoters. These activator complexes include Spt-Ada-Gcn5 acetyltransferase (SAGA) complex (Grant et al., 1997), p300/CBP (Bannister and Kouzarides, 1996; Ogryzko et al., 1996), PCAF (Yang et al., 1996), Src-1 (Spencer et al., 1997) and ACTR (Chen et al., 1997a). SAGA complex is a multi-subunit histone acetyltransferase complex which is composed of Ada, Spt and TATA-binding protein-associated factors (TAFs) (Grant et al., 1997; Grant et al., 1998). The histone acetyltransferase module in the SAGA complex is composed of either GCN5 or PCAF. SPT20, one of subunits of SAGA complex provides structural integrity to complex (Fig 2.1.2) (Nagy et al., 2009). SAGA complex plays important roles in transcription by generating and interacting with various histone modifications including acetylation, methylation, ubiquitylation and phosphorylation (Baker and Grant, 2007). SAGA complex has been shown to acetylates histones and helps in recruitment RNA polymerase II, which suggest a prominent role of SAGA complex in transcription (Bonnet et al., 2014). Role of the SAGA complex in transcription process has been studied to quite a good extend, however importance of SAGA complex in transcription regulation from bidirectional promoter has not been studied.

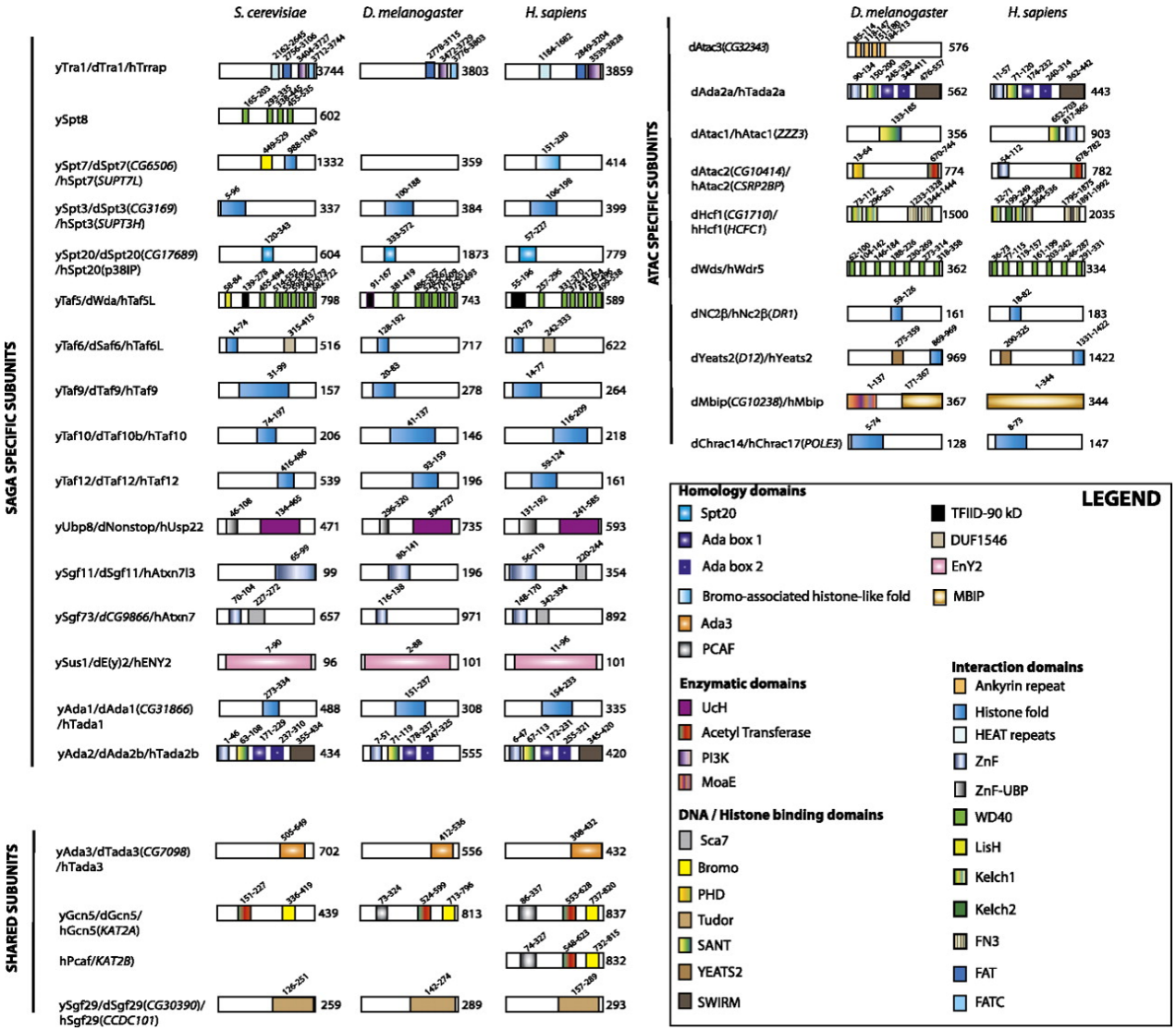


Figure 2.1.2. The SAGA (Spt-Ada-Gcn5 Acetyltransferase) complex.
List and domain architecture of SAGA subunit from *S.cerevisiae, D. melanogaster* and human. SAGA is a 1.8- MDa histone acetyl transferases complex composed of more than 20 polypeptides. Spt subunits provide the structural scaffold to the complex while TAFs required for structural integrity and interaction with basal transcriptional machinery. Gcn5 and Ada subunits play an important role in nucleosome acetylation. Table taken from (Spedale et al., 2012).

2.1.8 Rationale of the present study

Most of the studies attempting to understand the role of transcription factors in regulating transcription from the bidirectional promoters are either based on luciferase

reporter assays or binding of these factors on bidirectional promoters. However, the mechanism by which binding of these transcription factors regulate transcription from these promoters is poorly understood. Binding of a transcription factor might result into upregulation or downregulation of gene expression depending on whether it further recruits a co-activator or a co-repressor. A typical transcription factor harbors multiple functional domains, not only for binding to the specific DNA sequence, but also for interaction with other activator or repressor complexes. Despite the vast amount of literature present, it is still unknown as to how these transcription factors regulate transcription from bidirectional promoters. In this chapter, I have focused on the role(s) of these transcription factors in regulation of bidirectional promoters. In order to determine whether any of these transcription factors can potentially regulate transcription from bidirectional promoters, RNA interference mediated knockdown strategy was used. We hypothesized that if any of these factors is indeed involved in regulating bidirectional promoters, then knockdown of these factors would affect the transcription from bidirectional promoter-driven genes which can be scored by quantitative RT-PCR (q-PCR) for bidirectional gene pairs.

2.1.9 Summary of the work

In the present study, we selected the candidate transcription factors that have been shown to be overrepresented on bidirectional promoters namely GABPA, YY1, SP1 and NRF1 (Lin et al., 2007b). To test the role of these factors in bidirectional transcription, the levels of these factors were perturbed using siRNA mediated knockdown in NTera-2D1 (NT2D1) followed by quantitative RT-PCR (qRT PCR) analysis for 62 bidirectional genes. Bidirectional genes were selected based on the presence of consensus motifs for these transcription factors. Our results show that individual knockdown of GABPA, YY1, SP1 and NRF1 do not cause any alteration in gene expression from the selected bidirectional gene pairs, indicating redundancy in their function(s). In order to carefully address this, we performed double knockdown of these transcription factors and scored for the same gene pairs. However, no appreciable effect on gene expression was observed. These results prompted us to analyze the reason behind such strong redundancy in the function of these transcription factors. Therefore,

we set out to study if these transcription factors recruit any co-activator or co-repressor complexes to their target loci.

A detailed literature survey was performed to find out if any of the bidirectional promoter-associated transcription factors have been shown to recruit or co-occupy the co-activator complexes on genomic targets. We came across an important study demonstrating GABPA co-occupancy with SAGA complex on majority of SAGA-bound sites. In this chapter we report results of experiments designed to unravel the molecular mechanism(s) underlying the regulation of bidirectional promoters by these transcription factors. We show that many of these transcription factors physically interact with SAGA complex and have the ability to recruit the SAGA complex onto the bidirectional promoters. We provide evidence that all the carefully chosen transcription factors have the ability of recruiting SAGA complex to the selected bidirectional gene pairs. Consequently, we do not observe any appreciable dysregulation in transcription from these genes. For unequivocally proving this point, we generated knockout lines for NRF1 and SP1 and performed knockdown of other transcription factors under investigation in this background. We provide compelling evidence that perturbing the levels of three factors by the described strategy leads to downregulation of transcription from bidirectional gene pairs. Collectively, we demonstrate the coordinated role of multiple transcription factors in regulation of bidirectional promoters.

2.2 Materials and Methods

2.2.1 Design and Cloning of shRNA and overexpression constructs for transcription factors

To perform knockdown studies, we designed shRNA against YY1, SP1, GABPA and NRF1. For each gene two shRNA were designed. shRNA were designed using the Dharmacon shRNA design tool (http://dharmacon.gelifesciences.com/design-center). All shRNA was cloned in pSUPER puro vector and shGFP was used as the control shRNA. We also cloned FLAG tagged overexpression construct for SP1, GABPA, NFYA and

YY1. All these genes were cloned in p3xFLAG-CMV-10 expression vector (Sigma E7658). The sequences and other details of these primers are mentioned in appendix.

2.2.2 Antibodies and Reagents

Anti-YY1 (ab12131) and anti-NRF1 (ab175932 for western blot; ab34682 for ChIP) were purchased from Abcam. Anti-GABPA (sc-22810) was procured from Santa Cruz Biotechnology anti-SP1 (#9389) was procured from Cell Signaling Technology. SPT20 and GCN5 antibodies were generously gifted by Dr. Krishanpal Karmodiya (IISER, Pune).

2.2.3 Cell Culture and Transfections

HEK-293T and NT2D1 cells were grown in Dulbecco's Modified Eagle's Medium (DMEM, Invitrogen, Carlsbad, CA, USA) supplemented with 10% fetal bovine serum (FBS) and penicillin/streptomycin, at 37^{o}C under 5% CO_2 atmosphere. DLD1 cells were grown in RPMI1640 medium supplemented with 10% fetal bovine serum (FBS) and penicillin/streptomycin, at 37^{o}C under 5% CO_2 atmosphere (Gibco 11875-085). Knockdown experiments were performed using Lipofectamine RNAiMAX transfection reagent (Invitrogen, Carlsbad, CA, USA) according to manufacturer's protocol. Briefly, cells were grown up to 60% confluency in 60 mm culture dishes in DMEM supplemented with 10% FBS and penicillin/streptomycin, at 37^{o}C under 5% CO_2 atmosphere. 20 µL of 20 µM siRNAs (for knockdown) and 8 µg of DNA (for overexpression) were transfected using RNAiMAX and Lipofectamine 2000 respectively as per manufacturer's instructions, in serum-free medium. The medium was supplemented with 10% fetal bovine serum 6 h post-transfection. The cells were allowed to grow for 48 h and harvested by scraping and used for RNA preparation and quantitative RT-PCR.

2.2.4 Quantitative RT-PCR

RNA was isolated from control and siRNA transfected cells using TRI reagent (Sigma-Aldrich Co., St. Louis, MO, USA). One µg of RNA was used for cDNA preparation per 20 µl of reaction. The cDNA was used as template for the PCR with specific set of primers. Changes in threshold cycle (Ct) values were calculated as

follows: $\Delta Ct = (Ct_{\text{target genes}} - Ct_{\beta\text{-actin}})$ for transcript analysis. These ΔCt values were used to calculate fold change using the following formula: relative fold change = $2^{(-\Delta(\Delta Ct))}$. The graphs were plotted for the average fold values along with standard deviation from three independent experimental samples.

2.2.5 Western blotting

Cells were lysed in lysis buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.5% Triton X-100, 5% glycerol, 1% SDS, 1 mM Na_3VO_4, 10 mM NaF, 1 mM PMSF). Protein concentrations were determined using BCA protein estimation kit (Pierce™ BCA Protein Assay Kit 23225). Lysate was boiled with 6X SDS sample buffer (0.5 M Tris-HCl pH 6.8, 28% glycerol, 9% SDS, 5% 2-mercaptoethanol, 0.01% bromophenol blue) and electrophoresed on a 12.5% SDS-Polyacrylamide gel and transferred onto a PVDF membrane. Membranes were blocked in 5% milk and incubated overnight at 4°C with primary antibodies, washed thrice with TST (20 mM Tris buffer pH 7.4, , 500 mM NaCl and 0.05% tween 20) followed with incubation with horseradish peroxidase (HRP) conjugated secondary antibodies and washed thrice with TST. The signals were developed using Immobilon Western Chemiluminescent HRP Substrate (Millipore, Billerica, MA, USA) and detected using ImageQuant LAS 4000 (GE Healthcare, Piscataway, NJ, USA) according to the manufacturer's instructions.

2.2.6 Co-immunoprecipitation

Co-immunoprecipitation was performed to detect interaction of transcription factors and SAGA Complex component *in vivo*. For this, HEK-293T cells were grown upto 60% confluency and then transfected with FLAG tagged constructs of YY1, GABPA, NRF1 and NFYA. Transfections were performed with Lipofectamine 2000 (Life Technologies) as per manufacturer's protocol. Cell lysate was made in extraction buffer (25 mM HEPES pH 8.0, 150 mM NaCl, 0.2 mM EDTA pH 8.0, 5% glycerol, 0.5% NP 40 and 1X EDTA free complete protease inhibitor cocktail) without dithiothreitol (DTT). All the downstream processes were carried out under non-reducing conditions (i.e. in absence of dithiothreitol or β–mercaptoethanol in the lysis buffer and sample buffer). The lysate was diluted to a final concentration of 1 µg/µl with 1X chilled extraction buffer,

containing 1X EDTA-free complete protease inhibitor cocktail (Grunstein et al.). For each immunoprecipitation reaction, 500 µg of the lysate was precleared for 2 h at 4°C on a test tube rocker with 10 µl Protein G Dynabeads (Invitrogen). Beads were recovered by using a magnetic tube stand. Precleared extract was then incubated with 1 µg each of IgG, anti-FLAG and anti-SP1 antibodies for overnight at 4°C on a test tube rocker. To this, 20 µl of Protein G Dynabeads (Invitrogen) were added and incubation was further continued for 4 h. The protein-antibody complexes bound to Protein G Dynabeads were washed four times with IP Buffer (25 mM HEPES, pH 8.0, 150 mM NaCl, 0.5 % NP40 and 1X Protease inhibitor cocktail) followed by one wash with PBS (150 mM NaCl in 10 mM phosphate buffer). The complexes were eluted by incubating the beads in 10 mM Glycine buffer (pH 2.5) at 37°C for 5 min with intermittent mixing and the eluate was resolved on a 10 % SDS-PAGE gel and transferred to PVDF membrane. Immunoblotting was performed using anti-CDK9 antibody.

2.2.7 Chromatin immunoprecipitation (ChIP) assay

ChIP assay was performed as mentioned in Chapter 1 (Karmodiya et al., 2012) . For ChIP assay Anti-YY1 (ab 12132), Anti-SP1 antibody - ChIP Grade (ab13370) were procured from Abcam. Anti-GABPA (sc-28312X) and anti-PCAF (sc-13124) were procured from Santa Cruz Biotech. Anti-GCN5 and anti-SPT20 antibodies were kindly gifted by Dr Krishanpal Karmodiya (IISER Pune).

2.2.8 Designing and cloning of guide RNAs into pSpCas9(BB)-2A-Puro (PX459) vector for generating CRISPR mediated knockout cell lines

We designed guide RNA for cloning into pSpCas9(BB)-2A-Puro (PX459) vector which already has humanized CAS9 protein and puromycin as selection marker. For each gene two guide RNA's were designed. We designed guide RNA with Deskgen tool (https://horizon.deskgen.com/advanced.html). All the details of guide RNAs are mentioned in the guide RNA oligo list. Guide RNAs were designed with an overhang of BbsI restriction sites. For cloning of guide RNA, we used a published protocol (Ran et al., 2013). Briefly, guide RNA oligos were resuspended to the final concentration 100 µM.

Phosphorylation and annealing of the guide RNA oligos (top and bottom strands) was performed in thermocycler by using T4 PNK (polynucleotide Kinases) (NEB). px459 vector was digested with BbsI (NEB) followed by gel purification. Ligation reaction was performed using T4 DNA quick ligase (NEB) followed by transformation in the Stbl3 competent cells. A number of positive clones were selected and confirmed by sequencing and subjected to large-scale DNA purification via CsCl density gradient method.

2.2.9 Screening of positive knockout clones

DLD1 cells were used for creating CRISPR mediated knockout clones for transcription factors. Transfections were performed with 2 guide RNA clones for each transcription factor in 24 well culture plates. Forty-eight hours post-transfection, cells were trypsinized and seeded in 6 cm culture dish. Post 24 hrs of seeding, cells were supplemented with selection media containing puromycin (final concentration 2 µg/µl) and kept on selection for 2 weeks. Cells were supplemented with fresh selection media containing puromycin on each alternate day. After 2 weeks, when small puromycin resistant colonies become visible, puromycin selection was removed and cells were allowed to grow in RPMI media supplemented with 10% FCS till they become big enough to pick after which they were picked and individual colonies were seeded in 96 well plates. Cells were allowed to grow in 96 well plates till they reach adequate confluency. Once cells reached adequate confluency they were trypsinized and reseeded into two 96 well plates in 1:9 ratio, so that way that one plates get 10% of cells and other gets 90%. 90% cells containing plate were allowed to grow for 24 h followed by genomic DNA isolation. Genomic DNA was isolated as described previously (Ramírez-Solis et al., 1992).

2.3 Results

2.3.1 Knockdown of bidirectional promoter associated transcription factors

To study the role of transcription factors in regulating the transcription from the bidirectional promoters, we performed shRNA and siRNA mediated knockdown of SP1, YY1, GABPA and NRF1 factors in NT2D1 cells. Briefly, we transfected NT2D1 cells with shRNA and siRNA followed by qRT PCR and western blot to confirm the knock down. We observed better knockdown efficiency using siRNA as compared to shRNA clones (data not shown) therefore we used siRNAs for our further knockdown experiments. siRNA mediated knockdown of SP1, YY1, GABPA and NRF1 showed significant decrease in the expression of these genes, both at RNA as well as protein level (Figure 2.3.1). cDNAs from these knockdown experiments were further used to analyze the expression of bidirectional gene pairs.

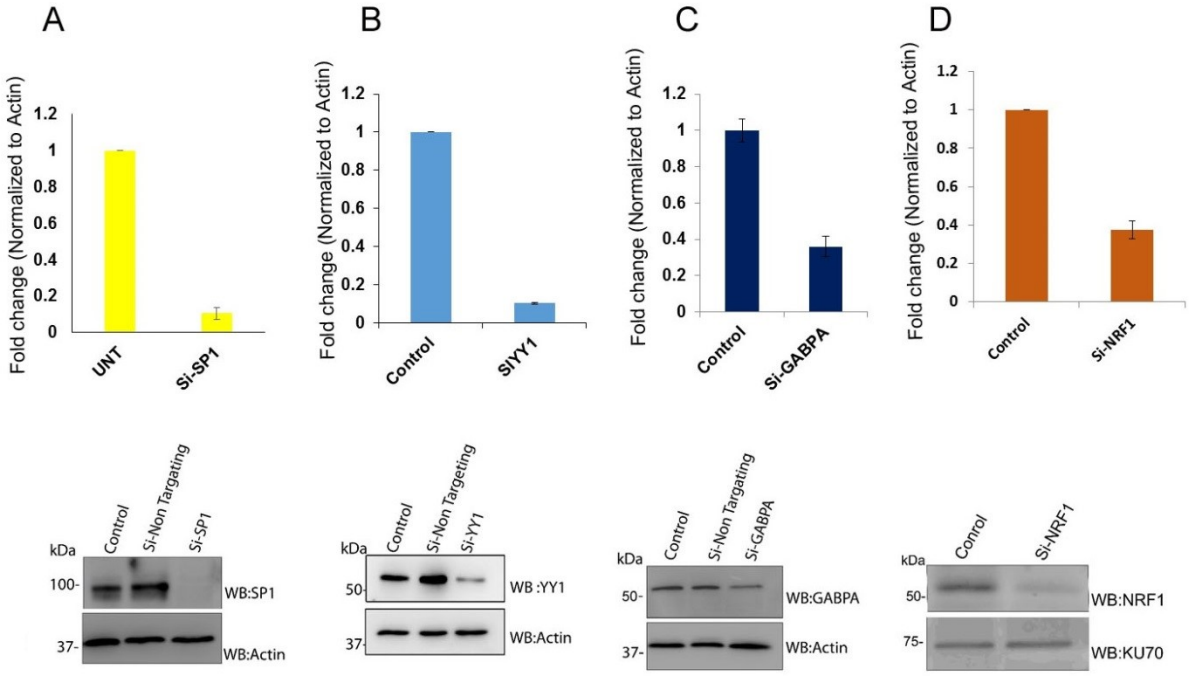


Figure 2.3.1 Specific siRNA transfections in NT2D1 cells result in efficient knockdown of the candidate transcription factors.
NT2D1 cells were transfected with non-targeting RNA (control) or with si-SP1, si-YY1, si-NRF1 and si-GABPA as described in “Materials and Methods”. Forty-eight hours post transfection, cells were harvested and RNA was extracted followed by cDNA synthesis.

Real time RT-PCR was performed with cDNA from these samples. β-actin was used as internal control to normalize the values. Error bars indicate standard deviation calculated from triplicates. Fold change as compared to control is depicted on Y axis. Knock down efficiency was analyzed by qRT-PCR and Western blotting. (A-D) Q-PCR analysis demonstrated significant knockdown of SP1, YY1, GABPA and NRF1. (D-G) Western blot showing knockdown of YY1, SP1, GABPA and NRF1 at protein level.

2.3.2 Knockdown of individual transcription factors does not alter gene expression profile from bidirectional promoters

To test whether knockdown of above mentioned transcription factors affects transcription from bidirectional promoters, we analyzed several bidirectional gene pairs upon knockdown of GABPA, SP1, NRF1 and YY1. These bidirectional gene pairs were selected based on the presence of binding sites for one or more of these transcription factors in their intergenic region (Odrowaz and Sharrocks, 2012; Reed et al., 2008; Tong et al., 2013) (Table 2.3.1). To rule out any allele specific variation in gene expression, we also included few bidirectional gene pairs from the X chromosome and used NT2D1 cell line for this experiment which is a male cell line. This proves to be a good system to score for the expression of monoallelic bidirectional gene pairs. We performed qRT PCR for the selected gene pairs in control versus knockdown condition and plotted the data as fold change normalized to beta-actin levels which acted as endogenous control. We observed that, none of the knockdowns (YY1, SP1, NRF1 and GABPA) led to any significant change in the expression of bidirectional gene pairs (Figure 2.3.2). Results from this experiment suggest two possibilities – (i) Either these transcription factors do not play any role in regulating bidirectional promoters or (ii) There is functional redundancy in their roles.

Number	Bidirectional Gene Pairs
1	TP 53 -WRAP53
2	CBX3-HNRNPA2B1
3	BRCA1-NBR2
4	SERPINI1- PDCD10
5	CIB1 -GDPGP1
6	SMC1A -RIBC1
7	GLA -HNRNPH2
8	ZCCHC16- NUP62CL
9	ZNF182-SPACA5
10	COL4A6 - COL4A5
11	BCAP31 - ABCD1
12	BRCC3-MTCP1
13	ZNF75-ZNF449
14	APOO-CXorf58
15	PSMD10 - ATG4A
16	PDZD11 -KIF4A
17	FOXM1-Anti FOXM1
18	MB4-WDR10
19	METTL23 -JMJD6
20	MCM7 -AP4M1
21	TMEM208-LRRC29
22	NSA2 -GFM2
23	OARD-NFYA
24	CEBPA antisense RNA -CEBPA
25	CBX5 -HNRNPA1
26	NUP88-RAPIN
27	ZNF81 -RPL7P57
28	TRAPPC2 -OFD1
29	NDUFB11 -RBM10
30	CETN2 -NSDHL

Table 2.3.1 List of bidirectional gene pairs used for the analyzing effect of TF KD on regulation of bidirectional promoters.

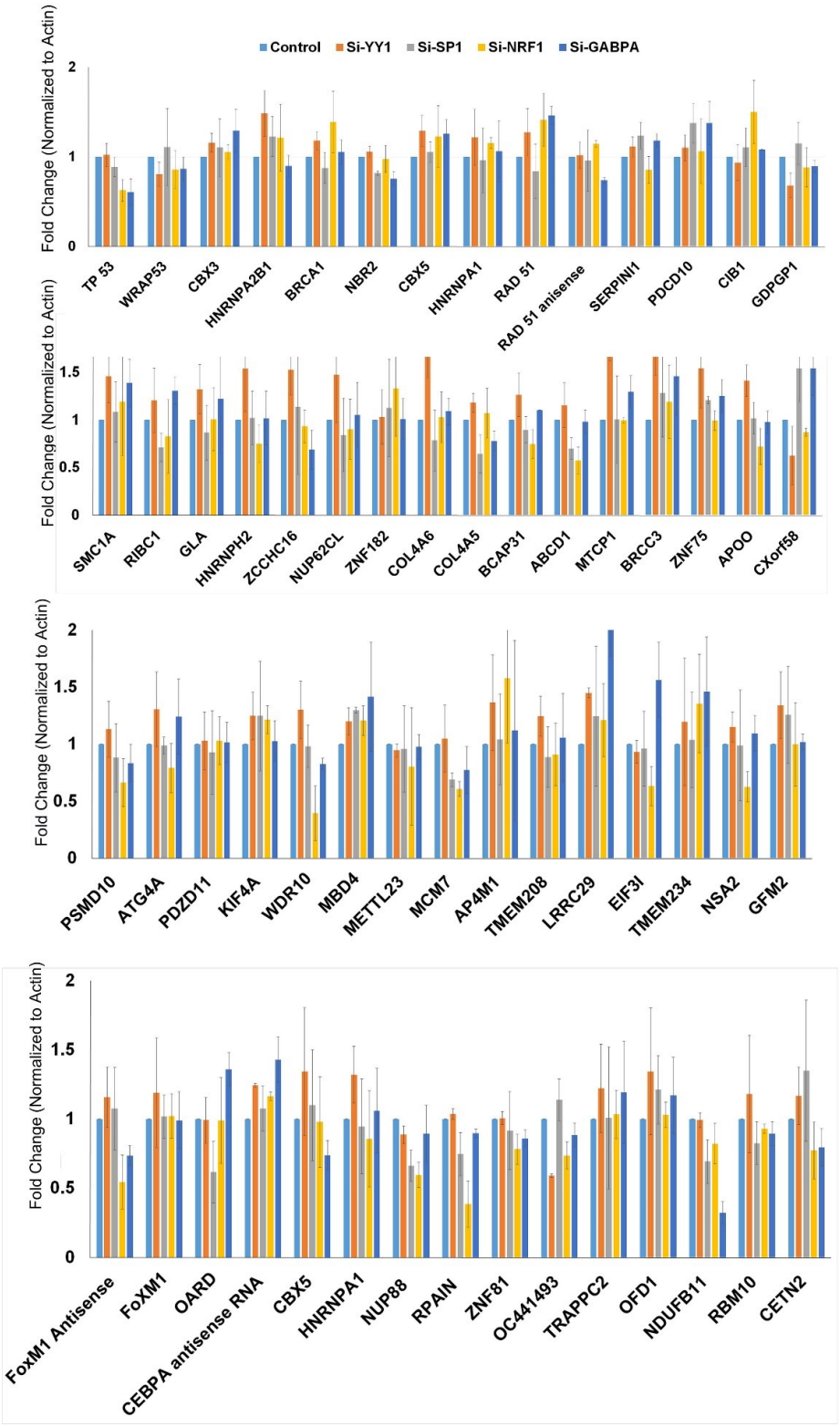


Figure 2.3.2 Knockdown of individual transcription factors does not alter gene expression profile from bidirectional promoters.
NT2D1 cells were transfected with non-targeting RNA (control) or with si-SP1, si-YY1, si-

NRF1 and si-GABPA as described in “Materials and Methods”. Forty-eight hours post transfection, cells were harvested and RNA was extracted followed by cDNA synthesis. Real time RT-PCR was performed with cDNA synthesized from these samples. β-actin was used as internal control to normalize the values. Error bars indicate standard deviation calculated from triplicates. Fold change as compared to control is depicted on Y axis. After confirming the knockdown of individual transcription factors (shown in Figure 2.3.1), these samples were screened for expression of several bidirectional gene pairs to study the effects of the knockdown on bidirectional transcription. qRT-PCR was performed from same cDNA samples which were used to score the knockdown efficiency of SP1, YY1, NRF1 and GABPA. Q-PCR analysis revealed that there is no significant dysregulation upon SP1, YY1, NRF1 and GABPA knockdown.

2.3.3 Knockdown of two transcription factors in combination results in mild dysregulation of bidirectional transcription

As showed above, individual knockdown of YY1, SP1, NRF1 and GABPA did not show significant dysregulation of bidirectional gene expression which hinted towards a possible redundancy in the function of the assessed transcription factors. To delineate the mechanism further, we performed double knockdowns of these transcription factors (YY1 and SP1, YY1 and NRF1, NRF1 and SP1) in NT2D1. The knockdowns were validated by qRT-PCR (Figure 2.3.3).

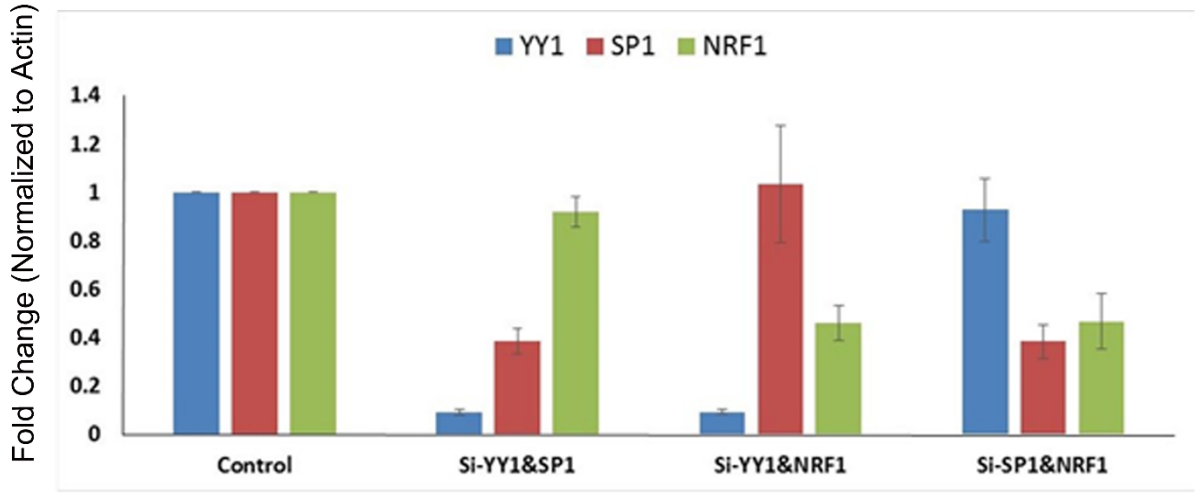


Figure 2.3.3 Validation of double knockdown of the transcription factors. NT2D1 cells were transfected with non-targeting RNA (control) or with si-YY1&SP1, si-YY1&NRF1 and si-SP1&NRF1 as described in “Materials and Methods”. Forty-eight hours post transfection, cells were harvested and RNA was extracted followed by cDNA

synthesis. Real time RT-PCR was performed with cDNA synthesized from these samples. β-actin was used as internal control to normalize the values. Error bars indicate standard deviation calculated from triplicates. Fold change as compared to control is depicted on Y axis. Significant knockdown was observed for each of these factors in all three combinations as revealed by qRT PCR analysis.

cDNA synthesized from these samples were used to analyze the expression of the bidirectional gene pairs. Interestingly, double knockdown of SP1 and YY1 as well as SP1 and NRF1 resulted in moderate dysregulation of bidirectional gene pair transcription (as shown in Figure 2.3.4). We did not observe any significant dysregulation in case of YY1 and NRF1 double knockdown (Figure 2.3.4). Altogether, our results point towards the possibility of functional redundancies.

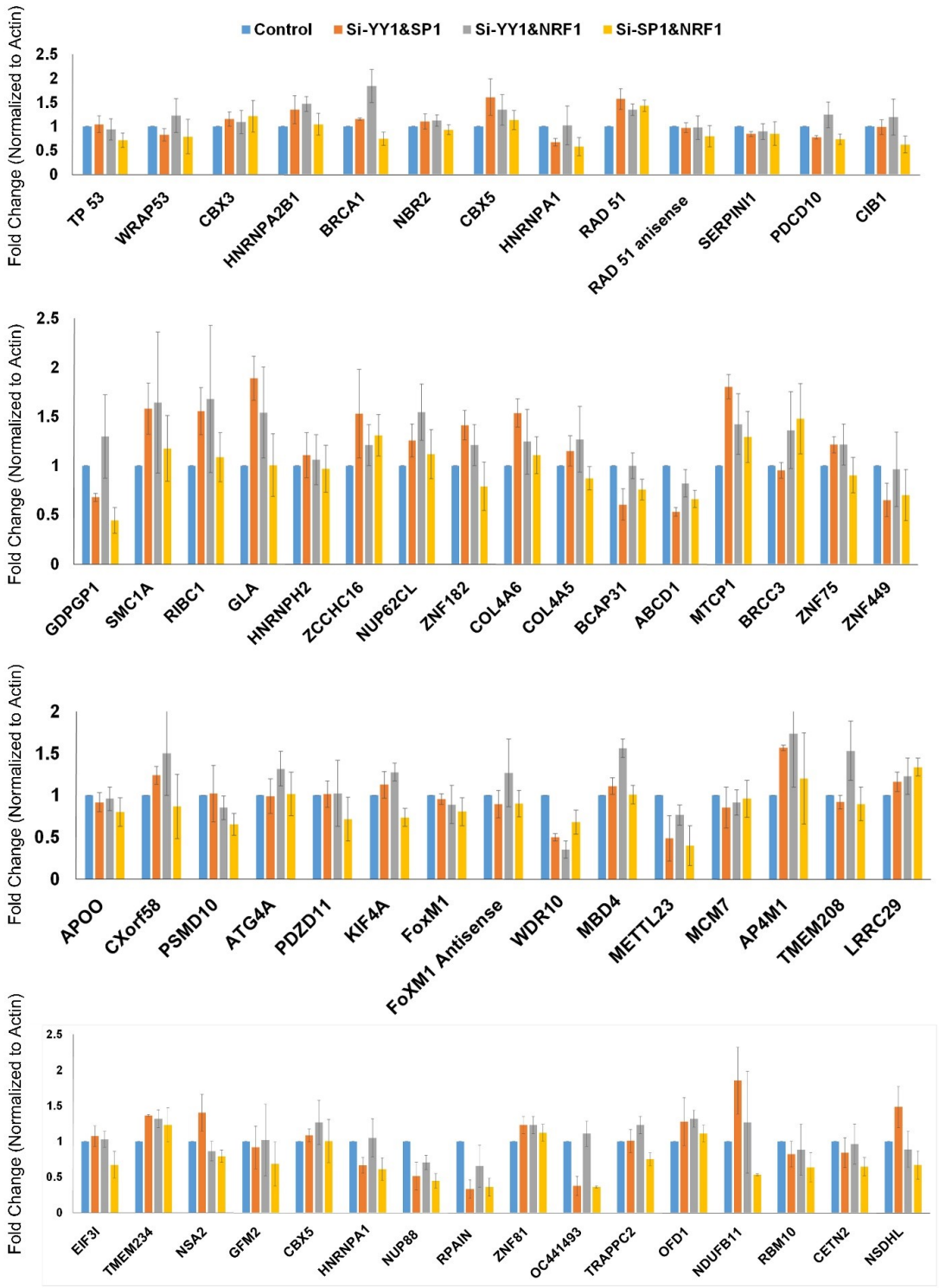


Figure 2.3.4 Double knockdown of transcription factors shows mild dysregulation in bidirectional transcription.

Double knockdown of transcription factor was performed in NT2D1 cells. NT2D1 cells were transfected with non-targeting RNA (control) or with Si-SP1, Si-YY1, Si-NRF1 and Si-GABPA as

described in “Materials and Methods”. Forty-eight hours post transfection, cells were harvested and RNA was extracted followed by cDNA synthesis. Real time RT-PCR was performed with cDNA synthesized from these samples. β-actin was used as internal control to normalize the values. Error bars indicate standard deviation calculated from triplicates. Fold change as compared to control is depicted on Y axis. To score the effect of double knockdown of transcription factors in combination. qRT-PCR was performed for bidirectional gene pairs. We did not observe any significant dysregulation in bidirectional gene expression.

2.3.4 SAGA complex binds to bidirectional promoters and interacts with GABPA

Above results revealed that though individual and double knockdown of YY1, SP1, NRF1 and GABPA does not affect transcription from bidirectional gene pairs. These experiments suggested that there is functional redundancy among these transcription factors in terms of regulating the bidirectional promoters. It is a well established fact that the transcription factors help in recruiting the chromatin modifying complexes to the promoters, thereby regulating the transcription. In order to attribute the functional significance to the aforementioned transcription factors we sought to determine if any of these factors are known to recruit co-activator complexes. It has been shown that binding by SAGA complex correlates with GABPA occupancy on many genomic targets (Krebs et al., 2011) (Figure 2.3.5 A). It is known that GABPA binding sites are present in nearly 80% of the bidirectional promoters (Collins et al., 2007). In the light of these findings, it seemed plausible that GABPA and SAGA might be regulating transcription from bidirectional promoters. To carefully address this, data from the referred study was used to analyze co-occupancies by GABPA and SAGA on the bidirectional promoters. From our analysis we found out that 54 bidirectional promoters showed co-enrichment of GABPA-SAGA on their intergenic region (Table 2.3.2). In order to test the biological significance of this finding, we randomly selected 18 (9 pairs) bidirectional genes from this gene list for further analysis. For all further experiments we have used these gene pairs. Since NT2D1 cells harbor high degree of aneuploidy (ATCC), we decided to use DLD1 cell line instead which maintained nearly diploid condition during the passages.

To see if SAGA complex can bind to bidirectional promoter in DLD1 cells, we performed ChIP experiments with two SAGA components, SPT20 and PCAF. We observed binding of SPT20 and PCAF on all 9 bidirectional promoters (Figure 2.3.5 B). Next, to see if GABPA can directly interact with SAGA complex and recruit it on

bidirectional promoters, we performed co-immunoprecipitation for GABPA and SAGA complex. Our co-immunoprecipitation data suggests that GABPA directly interacts with SAGA complex. Also, analysis of previously published data by Krebs et al. (2011) and validatory ChIP-PCR experiments revealed that GABPA and SAGA complex co-occupy the tested bidirectional promoters (Figure 2.3.5 C). For all further experiments we used the same 9 bidirectional gene pairs.

No	Gene ID	Gene Name	Gene Name
1	NM_145206	vti1a	vesicle transport through interaction with t-SNAREs homolog 1A (yeast)
2	NM_012460	TIMM9	translocase of inner mitochondrial membrane 9 homolog (yeast)
3	NM_016067	MRPS18C	mitochondrial ribosomal protein S18C
4	NM_033506	FBXO24	F-box protein 24
5	NM_014187	tmem208	transmembrane protein 208
6	NM_032437	EFCAB7	EF-hand calcium binding domain 7
7	M_0010246	LIN52	lin-52 homolog (C. elegans)
8	NM_006466	Polr3f	polymerase (RNA) III (DNA directed) polypeptide F, 39 kDa
9	NM_197956	Naif1	nuclear apoptosis inducing factor 1
10	NM_170691	GFM2	G elongation factor, mitochondrial 2
11	NM_032869	NUDCD1	NudC domain containing 1
12	NM_032741	AGPAT1	1-acylglycerol-3-phosphate O-acyltransferase 1
13	NM_020850	RANBP10	RAN binding protein 10
14	NM_052857	znf830	zinc finger protein 830
15	NM_015425	POLR1A	polymerase (RNA) I polypeptide A, 194kDa
16	NM_018663	PXMP2	hypothetical LOC100129532; peroxisomal membrane protein 2, 22kDa
17	NM_018663	LOC100129532	hypothetical LOC100129532; peroxisomal membrane protein 2, 22kDa
18	NM_002491	NDUFB3	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 3, 12kDa
19	NM_015072	TTLL5	tubulin tyrosine ligase-like family, member 5
20	NM_017748	Cwc25	coiled-coil domain containing 49
21	NM_020401	NUP107	nucleoporin 107kDa
22	M_0010985	USP5	ubiquitin specific peptidase 5 (isopeptidase T)
23	NM_014078	mrpl13	mitochondrial ribosomal protein L13
24	NM_005469	ACOT8	acyl-CoA thioesterase 8
25	NM_014177	C18orf55	chromosome 18 open reading frame 55
26	NM_015342	PPWD1	peptidylprolyl isomerase domain and WD repeat containing 1
27	NM_003442	ZNF143	zinc finger protein 143
28	NM_005791	mphosph10	M-phase phosphoprotein 10 (U3 small nucleolar ribonucleoprotein)
29	NM_004175	snrpd3	small nuclear ribonucleoprotein D3 polypeptide 18kDa
30	NM_006331	EMG1	EMG1 nucleolar protein homolog (S. cerevisiae)
31	NM_005869	Cwc27	serologically defined colon cancer antigen 10
32	NM_024818	uba5	ubiquitin-like modifier activating enzyme 5
33	M_0011666	PFKM	phosphofructokinase, muscle
34	NM_138358	C19orf52	chromosome 19 open reading frame 52
35	NM_032120	c7orf64	chromosome 7 open reading frame 64
36	NM_015528	Rnf167	ring finger protein 167
37	NM_004450	ERH	enhancer of rudimentary homolog (Drosophila)
38	NM_006083	Ik	similar to CG18005; IK cytokine, down-regulator of HLA II
39	NM_006083	LOC644456	similar to CG18005; IK cytokine, down-regulator of HLA II
40	NM_016399	TRIAP1	TP53 regulated inhibitor of apoptosis 1
41	NM_033415	Armc6	armadillo repeat containing 6
42	NM_016930	STX18	syntaxin 18
43	NM_020817	KIAA1407	KIAA1407
44	NR_003138	Snhg10	small Cajal body-specific RNA 13; small nucleolar RNA host gene 10 (non-protein coding)
45	NR_003138	SCARNA13	small Cajal body-specific RNA 13; small nucleolar RNA host gene 10 (non-protein coding)
46	NR_026826	Ints9	integrator complex subunit 9
47	NR_003545	c5orf44	chromosome 5 open reading frame 44
48	NR_024148	RNF121	ring finger protein 121
49	NR_027249	LOC253724	Grp94 neighboring nucleotidase pseudogene
50	NR_002944	LOC645691	heterogeneous nuclear ribonucleoprotein A1 pseudogene 35
51	NR_002944	HNRPA1L-2	heterogeneous nuclear ribonucleoprotein A1 pseudogene 10
52	NR_002944	LOC728643	heterogeneous nuclear ribonucleoprotein A1 pseudogene 33
53	NR_002944	HNRNPA1P2	heterogeneous nuclear ribonucleoprotein A1 pseudogene 2
54	NR_002944	HNRNPA1	heterogeneous nuclear ribonucleoprotein A1

Table 2.3.2: SAGA-GABPA co-occupied bidirectional gene pairs.
GABPA and SAGA complex exhibit co-occupancy on many genomic targets. Analysis of GABPA- and SAGA-bound bidirectional gene pairs revealed that 54 bidirectional gene promoters are co-occupied by GABPA and SAGA.

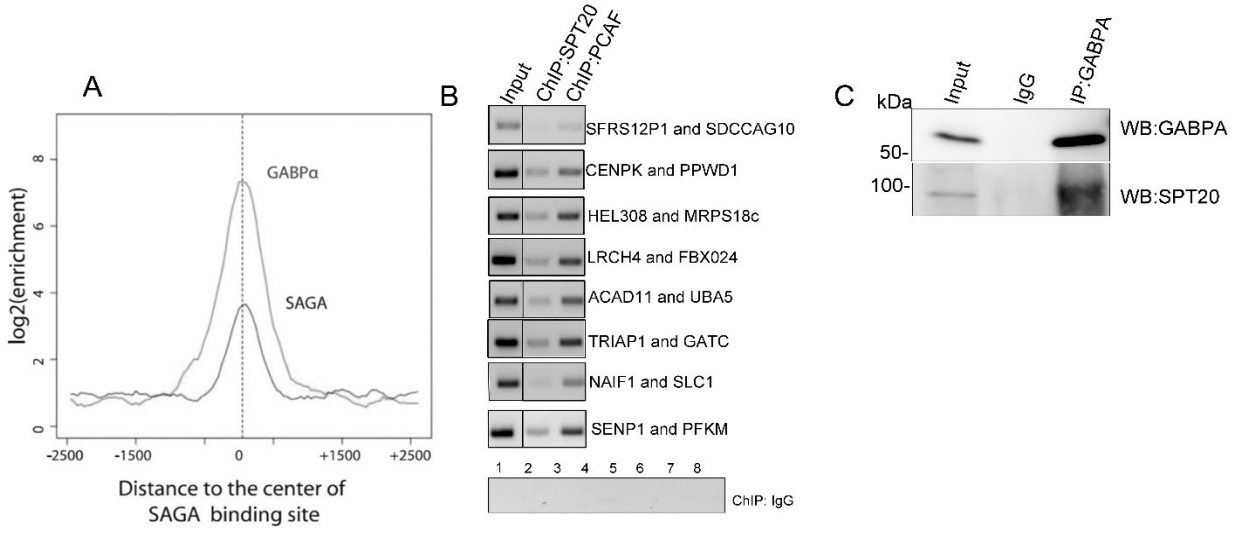


Figure 2.3.5 SAGA complex binds to the bidirectional promoters and interacts with GABPA. (A) GABPA and SAGA co-occupancy on genomic targets. (adapted from (Krebs et al., 2011). GABPA and SAGA co-occupy on 54 bidirectional promoters (Table 2.3.2). (B) Eight gene pairs were selected to analyze the binding of SAGA complex to these bidirectional promoters. ChIP analysis was done on the same locus where SAGA and GABPA showed co-occupancy (Krebs et al., 2011). ChIP experiment was performed in DLD1 using SPT20 and PCAF which are structural scaffold and catalytic subunit of SAGA complex. IgG was used as negative control. ChIP-PCR was performed using specific genomic primers designed for amplifying the intergenic region which showed SAGA-GABPA co-occupancy (Krebs et al., 2011). Input was used as positive control for ChIP PCR. (C) To test if GABPA can directly interact with SAGA complex, co-immunoprecipitation was performed. Immunoprecipitation and western blot was done with anti-GABPA and anti-SPT20 respectively. Presence of SPT20 in GABPA pull down fraction suggested the interaction between GABPA and SAGA complex.

2.3.5 GABPA knockdown does not affect the transcription from GABPA-SAGA co-occupied genes.

We demonstrated that SAGA and GABPA co-occupy 54 bidirectional promoters and GABPA directly interacts with SAGA complex. These results prompted us to hypothesize that GABPA might be a candidate transcription factor for recruitment of the SAGA complex on the bidirectional promoters. This can be tested by perturbing the levels of GABPA in DLD1 cells followed by qRT-PCR for bidirectional gene pairs. Surprisingly, we did not observe any dysregulation of bidirectional gene pairs (Figure

2.3.6). This data suggests that there might be other factors involved in recruiting SAGA complex.

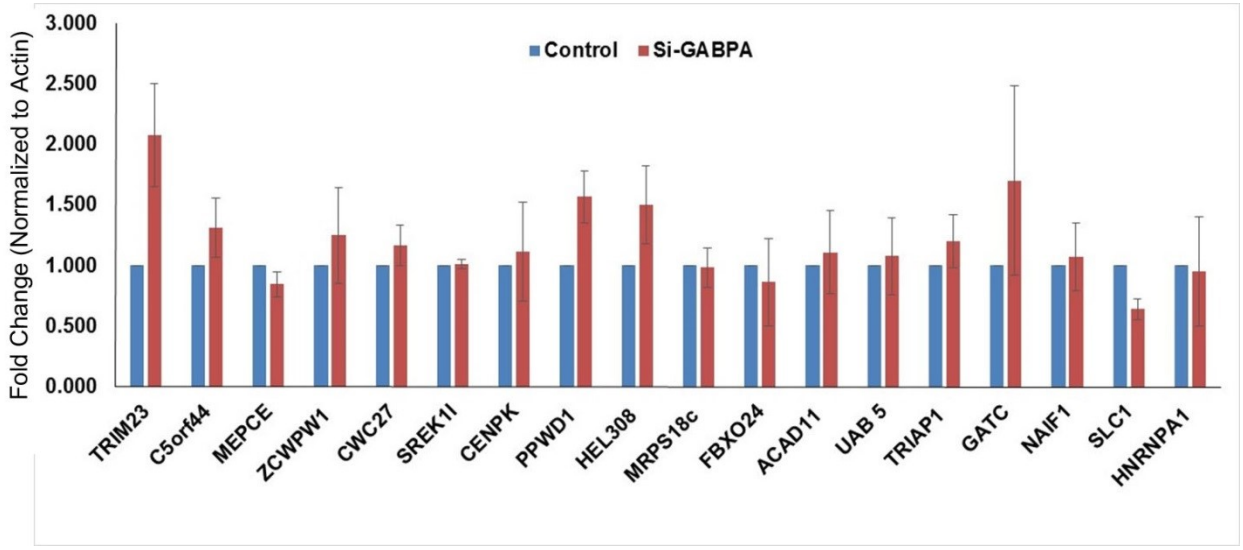


Figure 2.3.6 GABPA knockdown does not affect the transcription from GABPA-SAGA co-occupied bidirectional gene pairs.
DLD1 cells were transfected with non-targeting RNA (control) or si-GABPA as described in "Materials and Methods". Real time RT-PCR was performed with cDNA synthesized from these samples. β-actin was used as internal control to normalize the values. Error bars indicate standard deviation calculated from triplicates. Fold change as compared to control is depicted on Y axis. To determine the effect of GABPA knock down, qRT PCR analysis was done for the same set of bidirectional gene pairs which showed SAGA occupancy (Figure 2.3.5 B). However, the qRT PCR analysis shows that GABPA knockdown does not alter the expression from these genes.

2.3.6 YY1 and SP1 interact with SAGA

GABPA knockdown does not alter gene expression from GABPA-SAGA co-occupied bidirectional promoters which raised an interesting question of redundancy in the role of these factors in the SAGA complex recruitment on bidirectional promoters. We hypothesized that other proteins which are overrepresented on bidirectional promoters might be involved in recruitment of SAGA complex on these loci. Towards this end, we asked if any of other proteins including YY1, SP1 and GABPA can interact with SAGA complex. To test this, we generated overexpression constructs for YY1, SP1 and GABPA

in 3xFLAG-CMV-10 vector and transfected them into DLD1 cells. The overexpression was validated by western blotting (Figure 2.3.7 A). Further we performed immunoprecipitation using FLAG antibody and performed immunoblotting with two components of SAGA complex, GCN5 and PCAF. Immunoprecipitation results confirmed that significant amount of overexpressed protein is pulled down by the anti-FLAG antibody (Figure 2.3.7 B). We used interaction of GABPA and SAGA complex (GCN5 and PCAF) as positive control for co-immunoprecipitation experiment. Interestingly, we observed interaction of both SP1 and YY1 with GCN5 (Figure 2.3.7 C), however YY1 showed interaction with PCAF containing SAGA complex as well (Figure 2.3.7 D). These results suggests that in addition to GABPA, both SP1 and YY1 also can interact with SAGA complex.

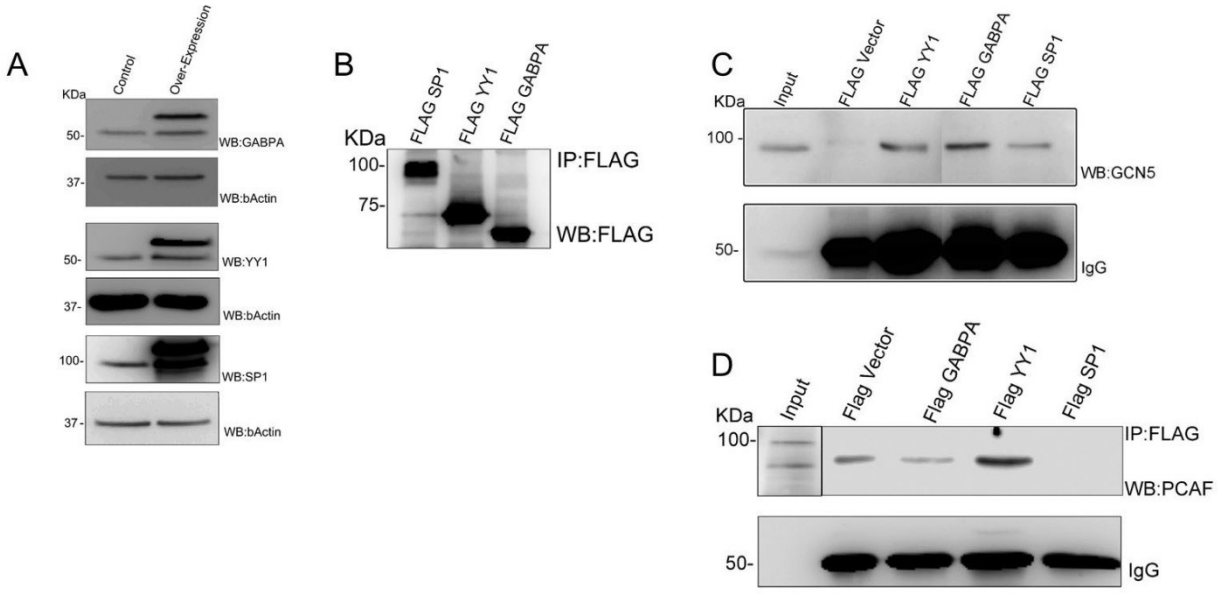


Figure 2.3.7 YY1 and SP1 interact with SAGA complex.
(A) Western blots showing the validation of YY1, SP1 and GABPA overexpression constructs. (B) Western blot showing that the positive immunoprecipitation for YY1, SP1 and GABPA. Immunoprecipitation and western blot was performed with anti-FLAG (C) Co-immunoprecipitation showing that YY1, SP1 and GABPA interact with SAGA complex catalytic subunit GCN5. (D) Co-immunoprecipitation showing that GABPA and YY1 interact with PCAF, other catalytic subunit of SAGA complex. SP1 does not interact with PCAF.

2.3.7 YY1 and SP1 co-occupy the SAGA complex bound sites on bidirectional promoters

The co-immunoprecipitation experiments revealed that not only GABPA, but SP1 and YY1 also interact with SAGA complex. Further we asked if SP1 and YY1 also occupy SAGA bound sites on bidirectional promoters. To test this, we performed ChIP experiment for the SP1, YY1 and GABPA and analyzed the occupancy of these transcription factors on the SAGA complex occupied promoter regions. Our semi-quantitative ChIP PCR results revealed that SP1 and YY1 occupy all of the candidate bidirectional promoters however we observed binding of GABPA only on two of the bidirectional gene pairs (Figure 2.3.8). This might be a reason for observed redundancy in our qRT PCR results after knockdown of these transcription factors. Since all of these transcription factors have the ability to bind and recruit the SAGA complex on bidirectional promoter, knocking down single factor did not yield any significant change in gene expression. Since GABPA displayed occupancy only 2 of the tested bidirectional gene pairs, we omitted GABPA from the future experiments for the current study and focused on YY1 and SP1.

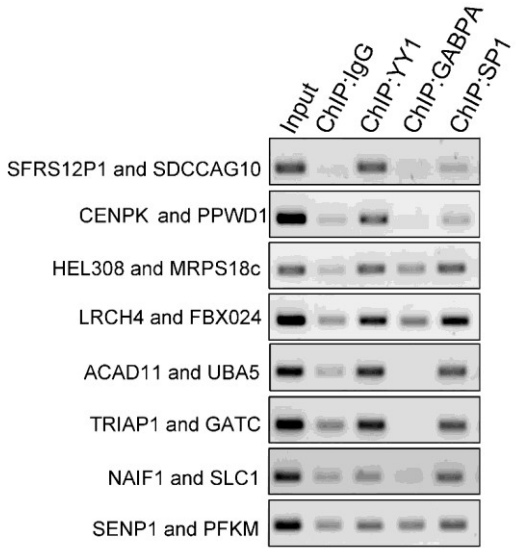


Figure 2.3.8 YY1 and SP1 co-occupy SAGA bound sites on bidirectional promoters. ChIP analysis was done on the same locus where SAGA and GABPA showed co-occupancy (Fig 2.3.5). .ChIP was performed in DLD1 cells using anti-SP1, anti-YY1, anti-GABPA. IgG was used as negative control. ChIP-PCR was performed using specific genomic primers designed for amplifying the intergenic region which showed SAGA-GABPA co-occupancy (Krebs et al., 2011). Input was used as positive control for ChIP-PCR.

2.3.8 Individual as well as double knockdown of YY1 and SP1 do not affect the transcription from SAGA-GABPA co-occupied bidirectional genes.

Our previous experiments demonstrated fair evidence for functional redundancies between various transcription factors with respect to bidirectional promoter- driven gene regulation. Our ChIP-PCR data showed that SP1 and YY1 bind to all candidate bidirectional promoters while GABPA showed occupancy only on two of the bidirectional gene pair promoters. We analyzed the role of SP1 and YY1 in the regulation of these bidirectional promoters by performing individual and double knockdown for SP1 and YY1, followed by qRT-PCR for these bidirectional gene pairs. The knockdowns were significant as shown by qRT-PCR (Figure 2.3.9 A). Next, qRT-PCR was performed to monitor if the individual and double knockdown of SP1 and YY1 have any effect on the expression of the bidirectional gene pairs. Surprisingly, we did not observe any dysregulation from any of the gene pairs, both upon individual as well as double knockdown (Figure 2.3.9 B and C). This data again suggested an evidence for involvement of some other factor in the bidirectional promoter regulation.

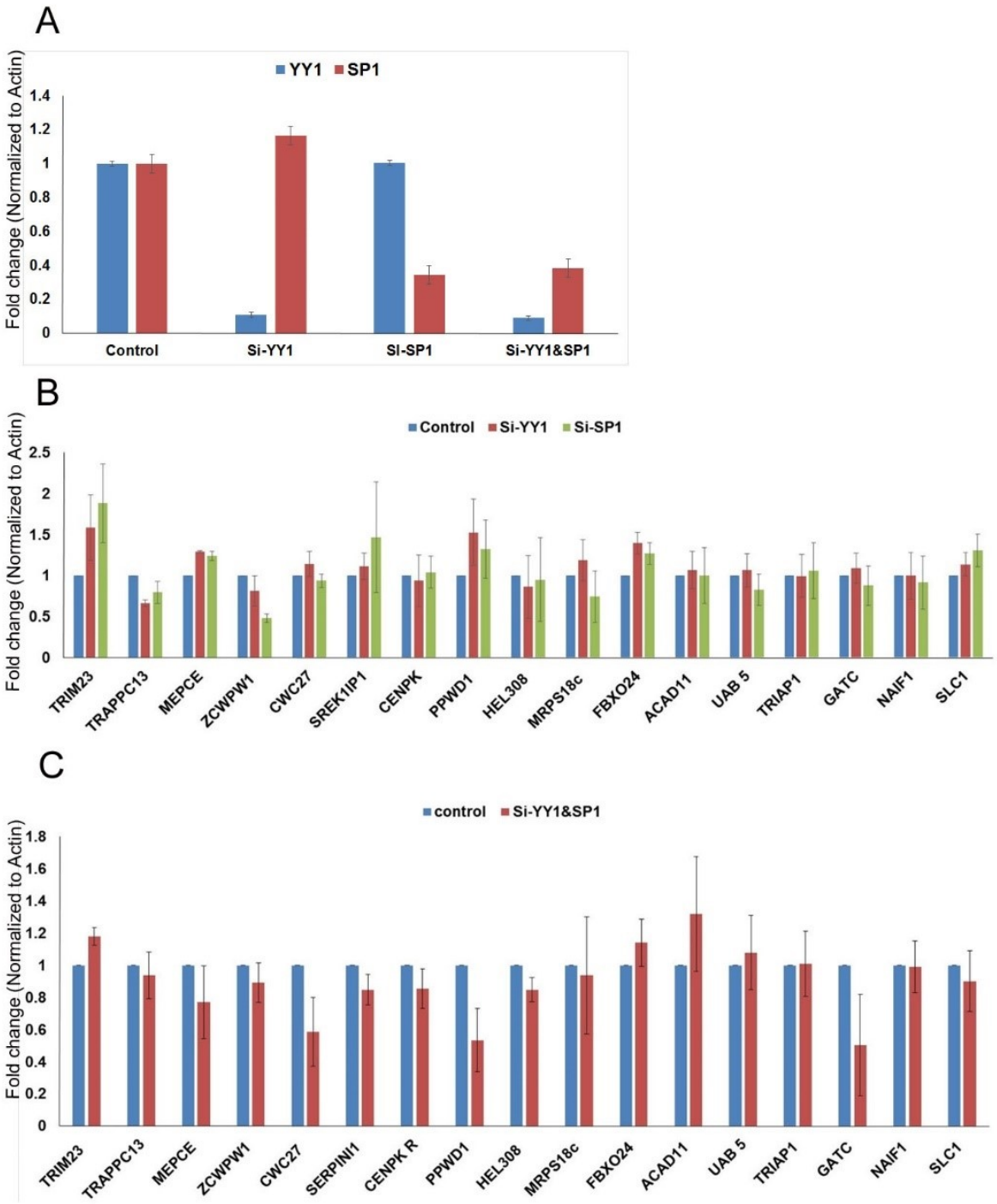


Figure 2.3.9 Individual as well as double knockdown of YY1 and SP1 does not affect the transcription from bidirectional gene pairs.
Q-PCR was performed to determine to effect of YY1 and SP1 knock down on SAGA-GABPA co-occupied bidirectional gene pairs. (A) Q-PCR analysis showing a significant decrease of both SP1 and YY1 upon knockdown. cDNAs from same knockdown samples was used to score the transcription from the bidirectional gene pairs. Q-PCR for bidirectional gene pairs revealed that (B) individual as well as (C) double knockdown of SP1 and YY1 do not affect the transcription from the bidirectional gene pairs.

2.3.9 NRF1 interacts with SAGA complex and binds to bidirectional promoters

Above results showed ample evidence for involvement of some other protein which might be involved in the SAGA recruitment and transcription from bidirectional gene pairs. Since we do not observe any dysregulation upon double knockdown of YY1 and SP1, therefore, we hypothesized that there might be sites for other transcription factors on these bidirectional promoters which would compensate for the loss of both SP1 and YY1. To understand this complexity of redundancy in bidirectional promoter gene regulation we analyzed the role of NRF1 on these bidirectional promoters as NRF1 has been shown among the most overrepresented motif on bidirectional promoters in addition to GABPA and YY1 (Lin et al., 2007b). We performed co-immunoprecipitation to check if NRF1 can interact with SAGA complex component. Our co-immunoprecipitation data revealed that NRF1 interact with GCN5 and SPT20 (Figure 2.3.10 A) which forms the catalytic and structural scaffold of the SAGA complex respectively. Next we asked if NRF1 also binds to the same genomic loci to which SP1 and YY1 bind along with the SAGA complex. We performed ChIP with NRF1 and analyzed the same genomic regions that were used before for ChIP analysis (Figure 2.3.5 B and 2.3.8). Our ChIP data revealed that NRF1 also binds to the same genomic loci which are occupied by SP1, YY1 and SAGA complex (Figure 2.3.10 B). Since all of the three factors have capability to interact and recruit the SAGA complex on these bidirectional promoters, therefore knockdown of single or even double factor did not yield any significant dysregulation in gene expression.

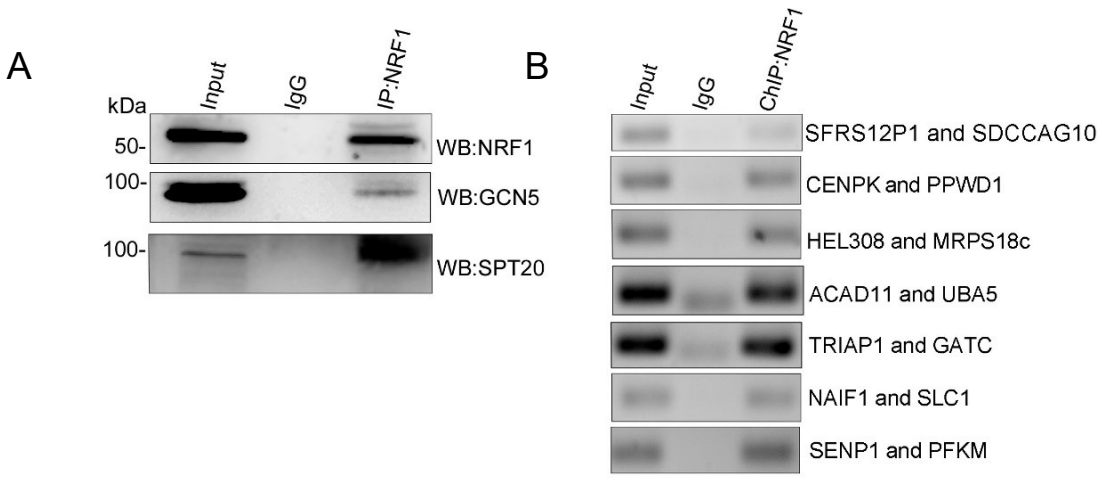


Figure 2.3.10 NRF1 interacts with SAGA complex and binds to SAGA bound sites on bidirectional promoters.
ChIP analysis was done in DLD1 cells on the same locus where SAGA and GABPA showed co-occupancy (Figure 2.3.5). ChIP was performed in DLD1 cells using anti-NRF1. IgG was used as negative control. ChIP-PCR was performed using specific genomic primers designed for amplifying the intergenic region which showed SAGA-GABPA co-occupancy (Krebs et al., 2011). Input was used as positive control for ChIP PCR (A) Co-immunoprecipitation showing that NRF1 interacts with GCN5 and SPT20 subunit of SAGA complex. (B) ChIP experiment using NRF1 was performed to analyze the occupancy of NRF1 on SAGA complex co-occupied bidirectional promoter regions. ChIP results showed that similar to YY1 and SP1, NRF1 also co-occupies on the SAGA complex bound sites on bidirectional promoters. IP: immunoprecipitation, IB: immunoblot.

2.3.10 Generation and validation of NRF1 and SP1 knockout cell lines

To address the functional redundancy between SP1, YY1 and NRF1, we performed triple knockdown for SP1, YY1 and NRF1 however we could not achieve significant knockdown of the three factors together (data not shown). To circumvent this problem, we utilized the recent and very efficient CRISPR/Cas9 system to generate NRF1 and SP1 knockout cell lines.

CRISPR (Clustered regularly interspaced short palindromic repeats) was discovered as a part of microbial adaptive immune system that uses RNA guided nucleases to cleave foreign genetic material (Bhaya et al., 2011; Deveau et al., 2010; Horvath and Barrangou, 2010). CRISPR-Cas9 system is based on a RNA-guided Cas9 nuclease where a Cas9 nuclease is targeted to the specific genomic target by a sgRNA (single-guide RNA) which consists of a 20 nucleotide long guide sequence and a scaffold (Figure 2.3.11). The guide sequence pairs with the DNA target, directly upstream of a

requisite 5′-NGG adjacent motif known as protospacer adjacent motif (PAM) which mediate the Cas9 cleavage. After recognizing the guide RNA and PAM motif, CAS9 binds to its target and induces a double strand break (DSB) at 3 base pairs upstream of PAM sequence (Ran et al., 2013). The resultant double strand break in the DNA can be repaired by two DNA repair pathways, homologous DNA recombination (HDR) or non-homologous end joining (NHEJ). HDR is very specific and highly error proof method of DNA recombination which is based on homology based DNA recombination and repair. On other hand, NHEJ is an error prone DNA repair pathway and it causes random insertions or deletions of nucleotides in the DNA leading to frameshift mutations in the gene, which also might lead to introduction of premature stop codon. NHEJ mediated repair of CRISPR-Cas9 generated DSBs is exploited efficiently to generate gene knockout (Ran et al., 2013).

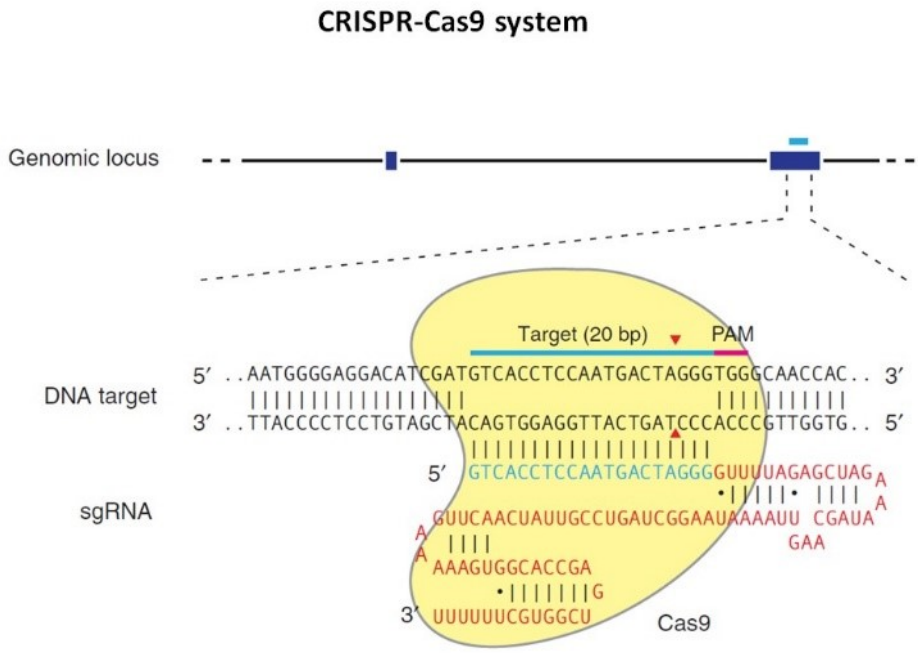
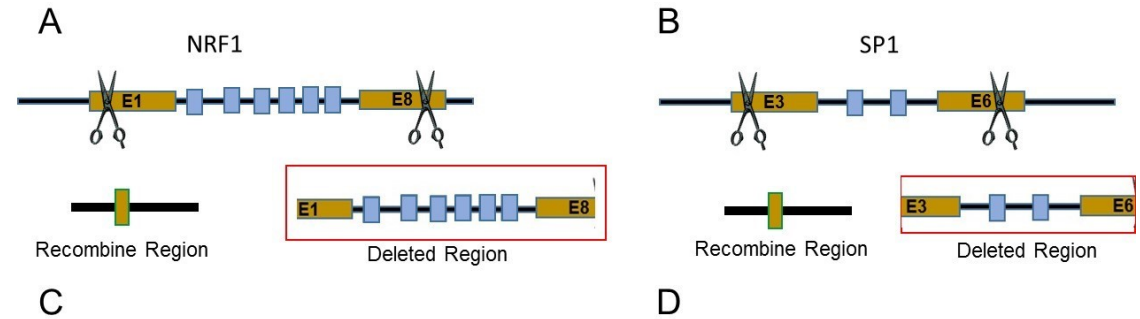


Figure 2.3.11 CRISPR-CAS9 mediated genomic editing.
CRISPR-Cas9 system is based on a RNA-guided Cas9 nuclease where a Cas9 nuclease (yellow) is targeted to the specific genomic target (blue) by an sgRNA (blue) which consists of a 20 nucleotide long guide sequence and a scaffold (red). Guide RNA pairs to the complementary DNA sequence on the target site directly upstream to a PAM motif (pink) which is followed by CAS9 recruitment which induces a double strand break, 3 base pair upstream (red cursor) to the PAM motif. Figure adapted from (Ran et al., 2013)

We used two-guide RNA strategy to generate knockout cell lines. This was found to be a more efficient strategy as opposed to one-guide RNA approach. Two guide-RNAs were designed against two exons for each gene to delete out the intervening DNA sequences, resulting into truncated, nonfunctional proteins. To generate SP1 and NRF1 knockout cell lines, we designed a pair of guide RNAs which targeted these genes at two different sites in the coding region (Figure 2.3.12). Guide RNAs for NRF1 was designed against exon 1 and 8 while for SP1 the guides were designed against exon 3 and 6. We cloned all the guide RNAs in pSpCas9 (BB)-2A-Puro (PX459) vector as described before (Ran et al., 2013). We used pSpCas9 (BB)-2A-Puro (PX459) vector as it co-expresses Cas9 and guide RNA both. The guide RNA clones for SP1 and NRF1 were then transfected in DLD1 cells. Targeted region is deleted from the genome followed by non-homologous end joining which will result in formation of newly recombined region (Figure 2.3.12). Due to deletion of exonic regions there will be loss of open reading frame which results in generation of knockout cell line for SP1 and NRF1.



Primer Pair	Expected PCR fragment Size
P1	220bp
P2	230bp
P3	311bp

Primer Pair	Expected PCR fragment Size
P1	200bp
P2	230bp
P3	270bp

Figure 2.3.12 Schematic diagram of steps involved in screening knockout clones. Two guide RNAs were designed against two exons (E) for each gene which will target Cas9 to these sites to induce DSBs (shown as scissors) and delete out the intervening DNA sequences, thus resulting into nonfunctional proteins. Presence of recombinant fragment (P3) using specific screening primers validated

the successful deletion. (A) For SP1, two guides were designed against exon1 and exon 8. Cas9 cut sites are shown with scissors. After CAS9 mediated cleavage, the DNA region to be deleted from the genome is shown in red box. (B) For NRF1, two guides were designed against exon 3 and 6 (Refer to the text for more details on screening strategy). (C&D) Strategy for screening the knockouts. Three sets of primers were designed to screen the knockouts. P3 set of primer is designed to screen the presence of indel, while other two sets (P1 and P2) were designed to score for homozygous and heterozygous.

To screen for the positive clones, two pairs of screening primers were designed around the two guide RNA target sites which flanked the Cas9 cut sites for those genes (Figure 2.3.12). We designed the screening primers in such a manner that both first pair (P1) and second pair (P2) of screening primers would yield amplification in wild type cells but would fail to produce any amplicon in the repaired allele after Cas9 mediated DSB repair. The reason behind this being that forward primer of first pair (P1) binds just upstream to the cut site and reverse primer binds within the deleted region (and opposite is true for the second pair (P2) of primers). Hence upon Cas9 mediated cleavage and NHEJ repair, these sets of primers will not yield any product because reverse primer of P1 (and forward primer of P2) cannot bind owing to the deletion. However, now the forward primer of first pair and reverse primer of the second pair (together termed as the third pair, P3) would amplify the deleted region in the knockout cell line (which would be too long to amplify in case of wild type cells having intact intervening DNA sequence). Furthermore, if the knockout clone is heterozygous then one should get amplification of P1 and/or P2 in addition to P3, however only P3 should give amplification in the homozygous knockout clones (Figure 2.3.12). We designed these two sets of screening primers for both SP1 and NRF1. For SP1 and NRF1 knockout clones, PCR product (P3) indicating the recombinant allele would be about 300 and 270 bp respectively (Figure 2.3.12). Since deleted region size is too long to amplify by PCR in given elongation time therefore negative clones (wild type cells) will not show any product. Homozygous and heterozygous clones for SP1 and NRF1 were selected by PCR using specific sets of primers (P1, P2 and P3) for both genes. As described in the schematic (Figure 2.2.12), we performed first round of screening (using P3 pair of primers). We got 22 and 17 colonies for SP1 and NRF1 respectively which were positive for presence of indel (Figure 2.3.13 A). Next round of screening was performed for isolating homozygous knockout colonies. We performed PCR with specific set of primers as described in (Figure 2.3.12).

Two colonies out of 22 knockout colonies for SP1 showed homozygous deletion. For NRF1, we found that 1 out of 17 colonies was homozygous for the deletion (Figure 2.3.13 B). These homozygous colonies were expanded more for further validation by western blot.

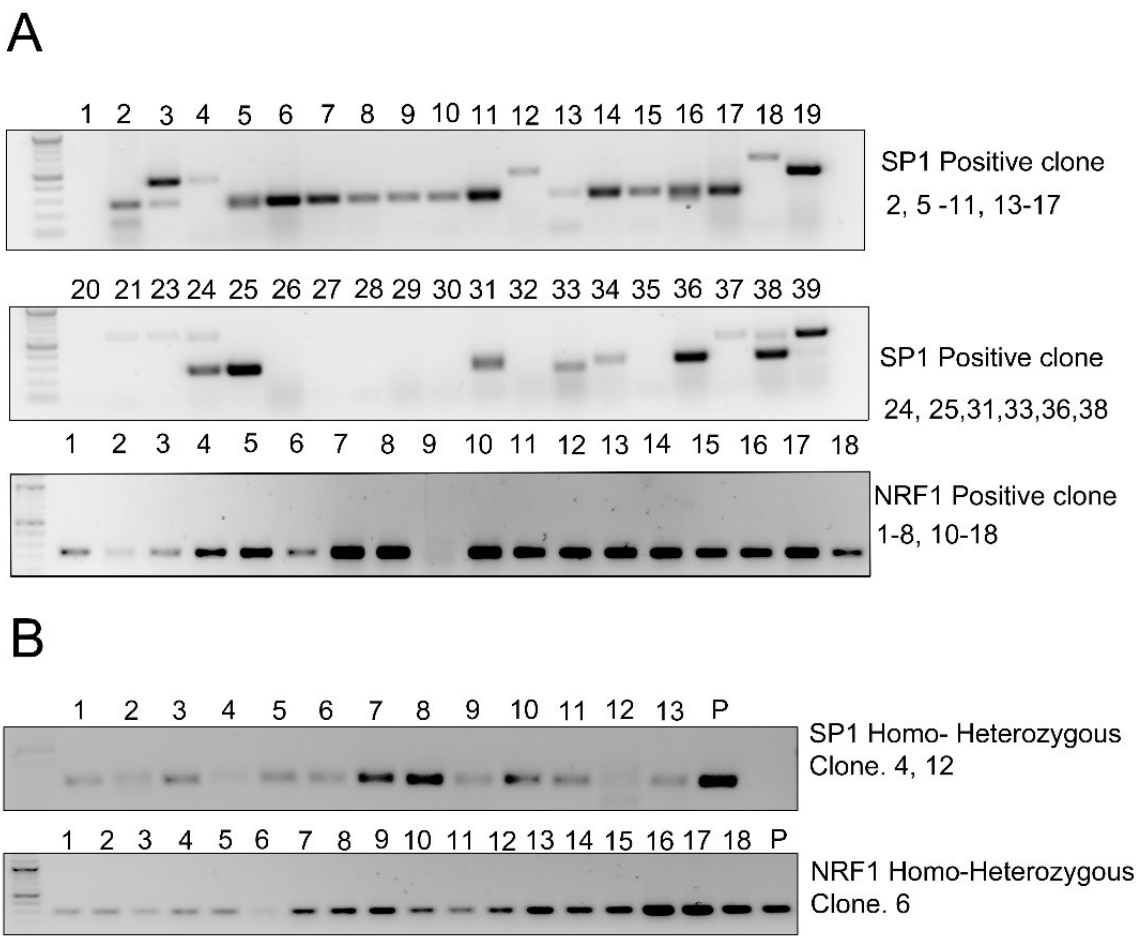


Figure 2.2.13 Screening of knockout clones for SP1 and NRF1. PCR was carried out for checking the indel from the individual colonies genomic DNA using specific screening primers. (A) Colonies showing deletion by performing PCR using the primer pair flanking both of the CAS9 cut sites (P3). Colonies which show expected size band 311 bp for SP1 and 270 bp for NRF1 were selected for next round of screening. (B) Next round of screening was carried out to screen for the colonies which are homozygous for the deletion. We are here showing the PCR results generated by P1 primer pair. Colonies which showed expected size of band were further grown for validation by immunoblot.

Homozygous clones for SP1 and YY1 were selected and grown further to validate these lines for gene knockout at protein level by western blot analysis. SP1 and NRF1 knockout lines showed complete loss of these proteins in the knockout cells as compared to the wild type cells. We repeated the western blot after passaging the cells multiple times to rule out the presence of any contaminating wild type cells. None of these replicated western blots showed presence of SP1 and NRF1 at the protein level, thus confirming stable knockouts for SP1 and NRF1 (Figure 2.3.14 A and B).

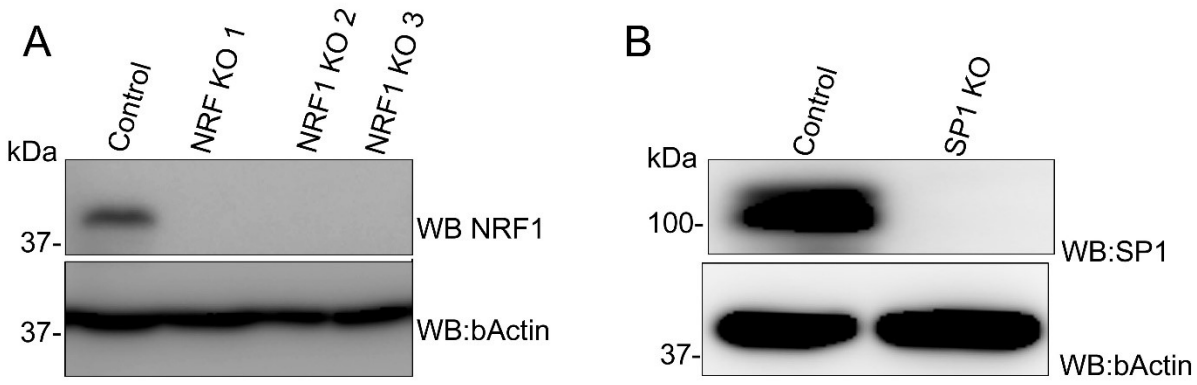


Figure 2.3.14 Validation of SP1 and NRF1 knockout DLD1 cell lines. Homozygous knockout colonies selected based on the screening PCRs were grown further for multiple passages followed by western blot analysis for SP1 and NRF1 expression. (A) Western blot for NRF1 showing the complete loss of NRF1 in the NRF1 knockout clones as compared to the wild type cells. (B) Western blot for SP1 showing the complete loss of SP1 in SP1 knockout clones as compared to the wild type cells.

2.3.11 Knockdown of SP1 and YY1 in NRF1 knockout background affect transcription from bidirectional gene pairs

We showed previously that single as well as double knockdown of YY1, SP1 and NRF1 in combination did not show any significant dysregulation in the bidirectional gene expression due to functional redundancy between these factors. All of these factors have shown the capability to recruit activator SAGA complex and occupy the SAGA bound loci on bidirectional promoters. To test for the hypothesized functional redundancies in the roles played by these transcription factors, we performed SP1 and YY1 knockdown in the NRF1 knockout background. We also performed YY1 knockdown in SP1 knockout background. Knockout and knockdown of all these factors were validated by western blot

(Figure 2.3.15). After validating the knockout and knockdown of these factors, we isolated RNA from these cells and analyzed the effect of these factors on bidirectional transcription in these knockout and knockdown cells.

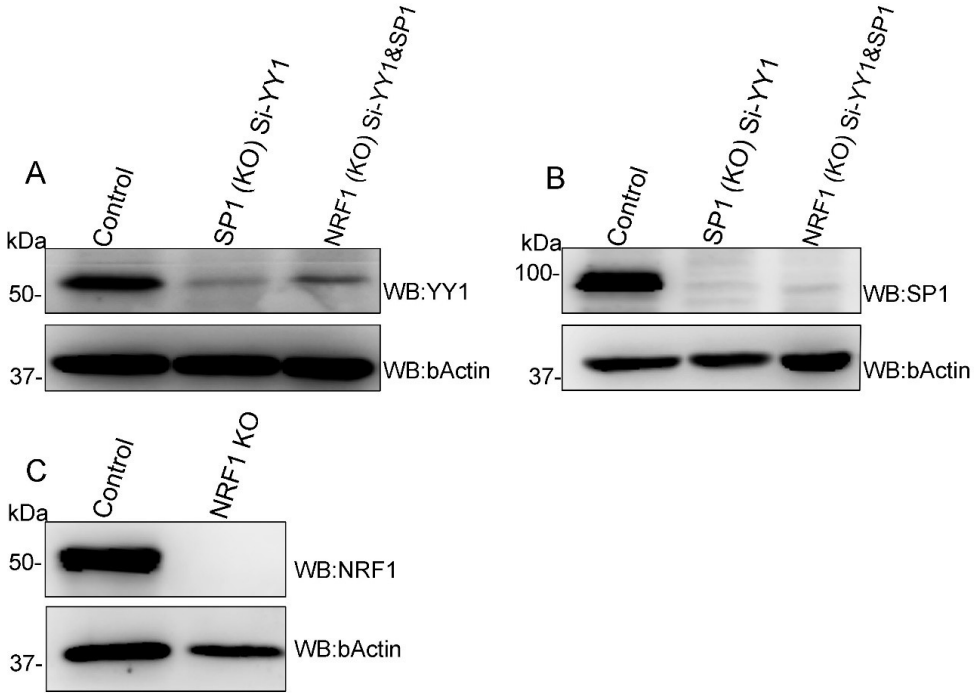


Figure 2.3.15 Validation of SP1 and YY1 knockdown in NRF1 knockout background.
(A) Western blot for YY1 showing efficient knockdown of YY1 in SP1 KO background and NRF1 KO background. (B) Western blot for SP1 confirming knock down of SP1 in NRF1 KO background (lane 3). (C) Western blot showing the complete loss of NRF1 protein in NRF1 knockout cells.

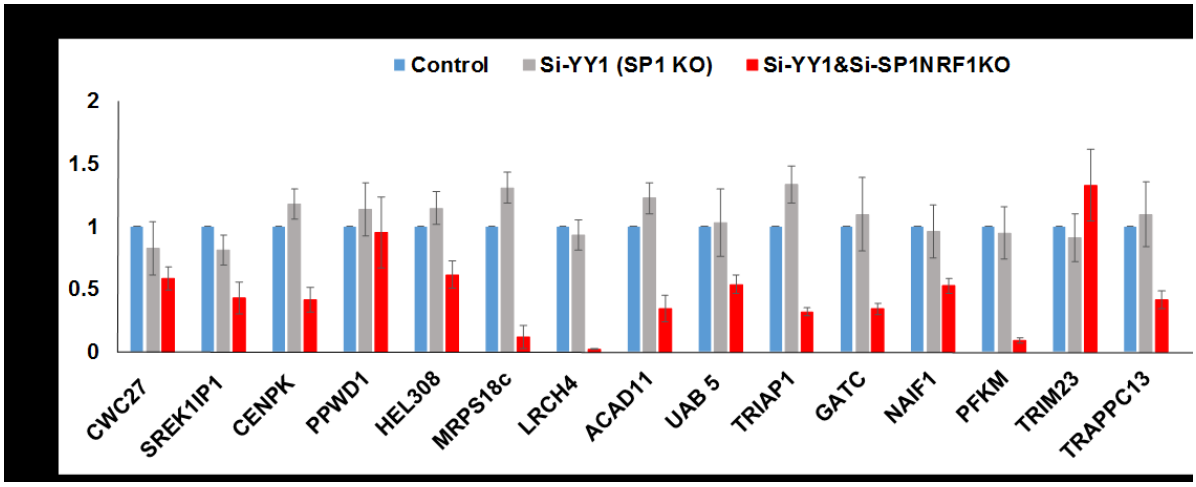


Figure 2.3.16 Knockdown of SP1 and YY1 in NRF1 knockout background affect the gene expression from bidirectional promoters.
siRNA mediated knockdown of SP1 and YY1 was performed in NRF1 knockout (KO) background and YY1 was knocked down in SP1 KO background. qRT-PCR was performed to analyze expression of bidirectional genes. As shown earlier in our double knockdown for YY1 and SP1, we did not observe any dysregulation in bidirectional gene expression after YY1 knockdown in SP1 knockout background. However, significant downregulation in gene expression from these bidirectional promoters was observed upon knockdown of SP1 and YY1 in NRF1 KO background. The qRT PCR data was normalized with endogenous control, actin.

Quantitative PCR results revealed that 13 genes exhibited significant downregulation after knockdown of SP1 and YY1 in NRF1 knockout background (Figure 2.3.16). Knockdown of YY1 in SP1 knockout background did not reveal any effect as seen earlier in the double knockdown experiment (Figure 2.3.9 and Figure 2.3.16). Collectively, these results demonstrate that all three factors SP1, YY1 and NRF1 complement each other's function in bidirectional promoter regulation. Single or double factor loss did not show any dysregulation in gene expression from bidirectional loci however removal of all the three factors results in significant dysregulation.

2.4 Discussion

Multiple factors including DNA methylation, histone modifications, histone variants and the binding of non-histone architectural proteins regulate the structure of chromatin. The multiple parameters that can affect transcription inside the nucleus can be arranged hierarchically. At the most basic level, regulation of the gene expression occurs through the interplay of multiple regulatory elements, such as promoters and more distal regulatory sequences which are usually classified under the operational definition of 'enhancer' or 'silencers'. Promoter is a DNA element to which transcription factors bind and recruit chromatin remodeling complexes, which in turn modify the chromatin, making it accessible to the transcriptional machinery. With the advancement in genome sequencing and better annotations, our understanding of eukaryotic promoters has changed drastically. In the mammalian genome, a special class of promoters has been enriched during the course of evolution and they are known as bidirectional promoters. Conceptually, bidirectional promoters are considered as an extension of the operon concept which is a hallmark of prokaryotes. However, among all the known mechanisms for transcriptional regulation, role of transcription factors and epigenetics in regulation of bidirectional transcription is not clear. We have already discussed about the role of epigenetic regulation towards gene expression from bidirectional promoters in the previous chapter. In this chapter, we have aimed to determine the role of various transcription factors in regulation of transcription from bidirectional promoters.

Several studies have focused on delineating the common feature in different bidirectional promoters. One common feature which has been discussed is the overrepresentation of some specific transcription factor binding motifs. Motifs for binding of several transcription factors including GABPA, MYC, E2F1, E2F4, NRF-1, CCAAT and YY1 have been shown to be overrepresented on bidirectional promoters as compared to unidirectional promoters. In this study, we selected specific bidirectional promoters based on the presence of binding sites of these overrepresented transcription factors (Odrowaz and Sharrocks, 2012; Reed et al., 2008; Tong et al., 2013). To test the significance of this overrepresentation, knockdown of SP1, YY1, NRF1 and GABPA was performed followed by qRT PCR for 62 bidirectional genes. Surprisingly, we did not observe any

dysregulation in transcription of bidirectional genes upon knockdown of these factors. Next, double knockdowns for SP1 and YY1, YY1 and NRF1 and NRF1 and SP1 were performed. Interestingly, double knockdown also did not result in any significant dysregulation in bidirectional gene expression barring few pairs which showed mild dysregulation. These results suggest two aspects about the function of these transcription factors in bidirectional promoter regulation; firstly, these factors might not be important for bidirectional promoter regulation and, secondly, there might be a redundancy in the function of these factors regarding the bidirectional promoter regulation.

In-depth literature survey of these factors in the recruitment of activator complexes led us to one study which showed that GABPA co-occupies with SAGA co-activator complex on many genomic targets of the SAGA complex (Krebs et al., 2011). Interestingly, GABPA also has been shown to bind to multiple bidirectional promoters in various cell types (Collins et al., 2007). We used the co-occupancy data for GABPA and SAGA complex to find out if these two factors co-occupy the bidirectional promoters. Our analysis revealed that 54 bidirectional gene pairs exhibit co-occupancy of GABPA and SAGA complex. We randomly selected 9 gene pairs (18 genes) from this list for further analysis. To test if SAGA complex binds to these 9 gene pairs' intergenic region, ChIP assay was performed using SPT20 and PCAF antibody, which are the structural and catalytic subunits of SAGA complex. ChIP analysis revealed that SAGA complex binds to the intergenic region of all the 9 gene pair selected. We further found that GABPA direclty interacts with the SAGA complex. These results encouraged us to analyze the expression of those 18 genes which showed co-occupancy by SAGA and GABPA upon GABPA knockdown. Surprisingly, we did not observe any dysregultion in any of these genes which suggested the involvment of some other factors in bidrectional promoter regultion. Co-immunoprecipitation assay revealed that SP1 and YY1 also interact with the SAGA complex along with GABPA. To determine if YY1 and SP1 could also bind to SAGA complex-bound sites on 9 selected gene pairs, we performed ChIP assay with YY1, SP1, and GABPA. We observed that SP1 and YY1 exhibited binding to all the 9 gene pairs, however, GABPA exhibited binding to only 2 intergenic regions. This result explains the reason behind the lack of any significant bidirectional gene regulation upon

GABPA knock down. It is evident that GABPA binds only to 2 intergenic regions and in absence of GABPA, SP1 and YY1 can recruit the SAGA complex to bidirectional promoters. To determine the redundancy in the function of SP1, YY1 and GABPA in bidirectional promoter regulation double knockdown of SP1 and YY1 was performed, however, this also did not lead to any dysregulation in bidirectional gene expression.

Our data suggested that though SP1, YY1 and GABPA all can interact with SAGA complex and bind to bidirectional promoters, removal of these transcription factors does not affect bidirectional gene expression. Next, we tested involvement of another overrepresented transcription factor NRF1 in the regulation of bidirectional transcription from SAGA occupied genes. Co-immunoprecipitation analysis of NRF1 and SAGA complex revealed that NRF1 directly interacts with the SAGA complex. Furthermore, we showed that NRF1 also binds to SAGA occupied intergenic regions of bidirectional gene pairs. Combined together, we showed that SP1, YY1 and NRF1 interact with the SAGA complex and co-occupy the SAGA complex-bound bidirectional promoters.

To demonstrate the redundancy of these factors in bidirectional transcription we required to deplete all the three factors, however we could not achieve efficient triple knockdown for SP1, YY1 and NRF1. To circumvent this problem, we used

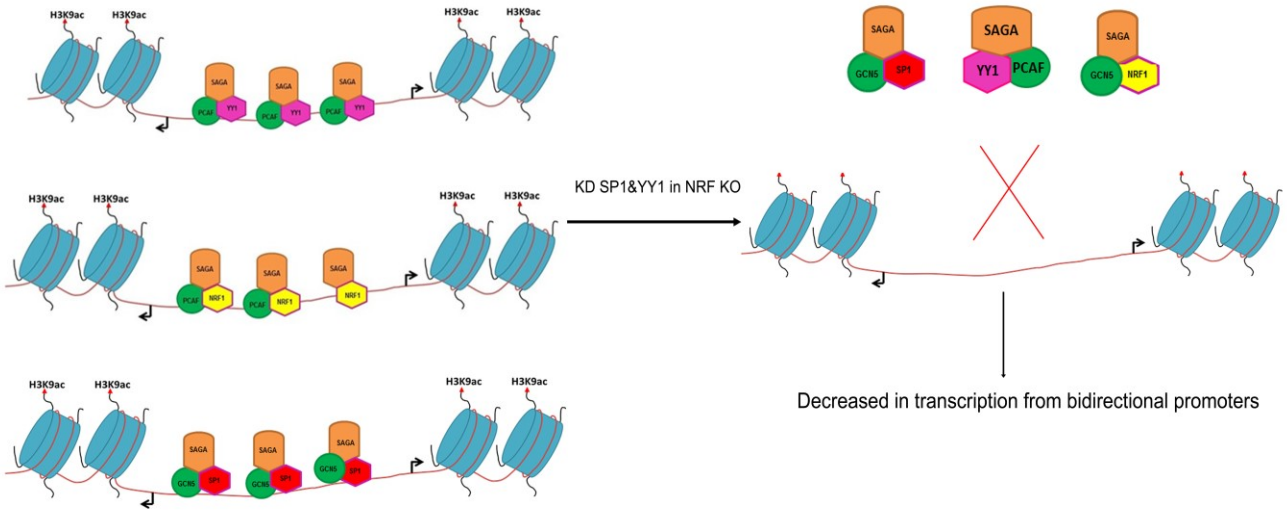


Figure 2.3.17 Proposed model for transcription regulation of bidirectional promoters. Our data suggests that all three factors (SP1, YY1 and NRF1) can interact with SAGA complex and co-occupy the bidirectional promoter region with SAGA complex. Depletion of one or two factors fails to alter the transcription from these promoters due to redundancy in the function of these factors. However depletion of all three together results in reduced transcription from these promoters which might be due to impaired recruitment of SAGA complex on these loci.

CRISPR/CAS9 genomic editing tool to generate knockout for these factors. We generated NRF1 and SP1 knockout cell lines which were used as the background to knockdown other factors to address the redundancy between these transcription factors. siRNA mediated depletion of SP1 and YY1 in NRF1 knockout background showed significant downregulation in expression from bidirectional gene pairs. However, YY1 knockdown in SP1 knockout background did not result in any dysregulation of bidirectional gene expression.

Our results show that SP1, YY1 and NRF1 interact with SAGA complex and co-occupy SAGA complex-bound bidirectional promoters. Though depletion of SP1, YY1 and NRF1, individually or in combination of two, do not show any effect on bidirectional gene pairs, however depletion of all the three factors (SP1, YY1 and NRF1) together results in significant downregulation of these bidirectional gene pairs. Depletion of all

three factors would lead to impaired recruitment of the SAGA complex on these promoters which would result in decrease in active transcription associated histone acetylation marks (Figure 2.1.17).

Ontology analysis of gene pairs regulated by bidirectional promoters show that these genes are involved in nucleosome assembly, regulation of cell proliferation, transcription, metabolic process, histone modification, cell cycle, chromatin modification, DNA repair and chromatin assembly or disassembly (Wakano et al., 2012). Since bidirectional promoter-driven genes are involved in functions related to housekeeping functions of cell therefore one would expect stable expression from these genes against different physiological stress conditions. Since large number of genes are regulated by bidirectional promoters, therefore regulating these genes by a single transcription factor would be very risky for cell survival because any mutation in that one regulatory protein would lead to failure of multiple cellular functions and ultimately cell death. To avoid such catastrophic outcome, redundant mechanism would have been selected to regulated these large number of genes. For any cellular process redundant mechanisms show more resistance in change against any external or internal damage. Our data also supports this notion and revealed that bidirectional promoter driven genes are regulated by redundant mechanism. We show that depletion of one or two transcription factors does not lead to any significant change in gene expression from bidirectional promoters. However simultaneous depletion of three factors (NRF1, SP1 and YY1) which would be rare event in natural conditions, results in dysregulation of gene expression from bidirectional promoters. Collectively, our results demonstrate that SP1, YY1 and NRF1 function in a redundant manner to regulate gene expression from bidirectional promoters

2.5. References

Abraham, B.J., Cui, K., Tang, Q., and Zhao, K. (2013). Dynamic regulation of epigenomic landscapes during hematopoiesis. BMC genomics *14*, 193.

Adachi, N., and Lieber, M.R. (2002). Bidirectional gene organization: a common architectural feature of the human genome. Cell *109*, 807-809.

Ahn, J., and Gruen, J.R. (1999). The genomic organization of the histone clusters on human 6p21. 3. Mammalian genome *10*, 768-770.

Albig, W., Kioschis, P., Poustka, A., Meergans, K., and Doenecke, D. (1997). Human histone gene organization: nonregular arrangement within a large cluster. Genomics *40*, 314-322.

Almada, A.E., Wu, X., Kriz, A.J., Burge, C.B., and Sharp, P.A. (2013). Promoter directionality is controlled by U1 snRNP and polyadenylation signals. Nature *499*, 360-363.

Anno, Y.-N., Myslinski, E., Ngondo-Mbongo, R.P., Krol, A., Poch, O., Lecompte, O., and Carbon, P. (2011a). Genome-wide evidence for an essential role of the human Staf/ZNF143 transcription factor in bidirectional transcription. Nucleic acids research *39*, 3116-3127.

Anno, Y.N., Myslinski, E., Ngondo-Mbongo, R.P., Krol, A., Poch, O., Lecompte, O., and Carbon, P. (2011b). Genome-wide evidence for an essential role of the human Staf/ZNF143 transcription factor in bidirectional transcription. Nucleic acids research *39*, 3116-3127.

Bailey, S.D., Zhang, X., Desai, K., Aid, M., Corradin, O., Cowper-Sal, R., Akhtar-Zaidi, B., Scacheri, P.C., Haibe-Kains, B., and Lupien, M. (2015). ZNF143 provides sequence specificity to secure chromatin interactions at gene promoters. Nature communications *2*.

Baker, S., and Grant, P. (2007). The SAGA continues: expanding the cellular role of a transcriptional co-activator complex. Oncogene *26*, 5329-5340.

Bannister, A.J., and Kouzarides, T. (1996). The CBP co-activator is a histone acetyltransferase.

Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007a). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007b). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Behura, S.K., and Severson, D.W. (2015). Bidirectional Promoters of Insects: Genome-Wide Comparison, Evolutionary Implication and Influence on Gene Expression. Journal of molecular biology *427*, 521-536.

Bellizzi, D., Dato, S., Cavalcante, P., Covello, G., Di Cianni, F., Passarino, G., Rose, G., and De Benedictis, G. (2007). Characterization of a bidirectional promoter shared between two human genes related to aging: SIRT3 and PSMD13. Genomics *89*, 143-150.

Bhaya, D., Davison, M., and Barrangou, R. (2011). CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. Annual review of genetics *45*, 273-297.

Biswas, M., and Chan, J.Y. (2010). Role of Nrf1 in antioxidant response element-mediated gene expression and beyond. Toxicology and applied pharmacology *244*, 16-20.

Bonnet, J., Wang, C.-Y., Baptista, T., Vincent, S.D., Hsiao, W.-C., Stierle, M., Kao, C.-F., Tora, L., and Devys, D. (2014). The SAGA coactivator complex acts on the whole transcribed genome and is required for RNA polymerase II transcription. Genes & development *28*, 1999-2012.

Bornelöv, S., Komorowski, J., and Wadelius, C. (2015). Different distribution of histone modifications in genes with unidirectional and bidirectional transcription and a role of CTCF and cohesin in directing transcription. BMC genomics *16*, 300.

Brandeis, M., Frank, D., Keshet, I., Siegfried, Z., Mendelsohn, M., Temper, V., Razin, A., and Cedar, H. (1994). Spl elements protect a CpG island from de novo methylation.

Breathnach, R., and Chambon, P. (1981). Organization and expression of eucaryotic split genes coding for proteins. Annual review of biochemistry *50*, 349-383.

Briggs, M.R., Kadonaga, J.T., Bell, S.P., and Tjian, R. (1986). Purification and biochemical characterization of the promoter-specific transcription factor, Sp1. Science *234*, 47-52.
Bucher, P., and Trifonov, E.N. (1986). Compilation and analysis of eukaryotic POL II promoter sequences. Nucleic acids research *14*, 10009-10026.
Burke, T.W., and Kadonaga, J.T. (1996). Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. Genes & development *10*, 711-724.
Burke, T.W., and Kadonaga, J.T. (1997). The downstream core promoter element, DPE, is conserved fromDrosophila to humans and is recognized by TAFII60 of Drosophila. Genes & development *11*, 3020-3031.
Butler, J.E., and Kadonaga, J.T. (2002). The RNA polymerase II core promoter: a key component in the regulation of gene expression. Genes & development *16*, 2583-2592.
Carcamo, J., Buckbinder, L., and Reinberg, D. (1991). The initiator directs the assembly of a transcription factor IID-dependent transcription complex. Proceedings of the National Academy of Sciences *88*, 8052-8056.
Chambers, C.B., Halford, W.P., Geltz, J., Villamizar, O., Gross, J., Embalabala, A., Gershburg, E., and Wilber, A. (2015). A System for Creating Stable Cell Lines that Express a Gene of Interest from a Bidirectional and Regulatable Herpes Simplex Virus Type 1 Promoter. PloS one *10*, e0122253.
Chen, H., Lin, R.J., Schiltz, R.L., Chakravarti, D., Nash, A., Nagy, L., Privalsky, M.L., Nakatani, Y., and Evans, R.M. (1997a). Nuclear receptor coactivator ACTR is a novel histone acetyltransferase and forms a multimeric activation complex with P/CAF and CBP/p300. Cell *90*, 569-580.
Chen, J.-L., Attardi, L.D., Verrijzer, C.P., Yokomori, K., and Tjian, R. (1994). Assembly of recombinant TFIID reveals differential coactivator requirements for distinct transcriptional activators. Cell *79*, 93-105.
Chen, S., Nagy, P.L., and Zalkin, H. (1997b). Role of NRF-1 in bidirectional transcription of the human GPAT-AIRC purine biosynthesis locus. Nucleic acids research *25*, 1809-1816.
Chinenov, Y., Coombs, C., and Martin, M.E. (2000). Isolation of a bi-directional promoter directing expression of the mouse GABPα and ATP synthase coupling factor 6 genes. Gene *261*, 311-320.
Cole, E.G., and Gaston, K. (1997). A functional YY1 binding site is necessary and sufficient to activate Surf-1 promoter activity in response to serum growth factors. Nucleic acids research *25*, 3705-3711.
Collins, P.J., Kobayashi, Y., and Nguyen, L. (2007). The ets-related transcription factor GABP directs bidirectional transcription. PLoS genetics *3*, e208.
Courey, A.J., and Tjian, R. (1988). Analysis of Sp1 in vivo reveals mutiple transcriptional domains, including a novel glutamine-rich activation motif. Cell *55*, 887-898.
Davies, G., Locke, S., Wright, P., Li, H., Hanson, R., Miller, J., and Anderson, S. (2007). Identification of bidirectional promoters in the human KIR genes. Genes and immunity *8*, 245-253.
Davila Lopez, M., Martinez Guerra, J.J., and Samuelsson, T. (2010). Analysis of gene order conservation in eukaryotes identifies transcriptionally and functionally linked genes. PloS one *5*, e10654.
Deveau, H., Garneau, J.E., and Moineau, S. (2010). CRISPR/Cas system and its role in phage-bacteria interactions. Annual review of microbiology *64*, 475-493.
Dhadi, S.R., Krom, N., and Ramakrishna, W. (2009). Genome-wide comparative analysis of putative bidirectional promoters from rice, Arabidopsis and Populus. Gene *429*, 65-73.
Donati, G., Imbriano, C., and Mantovani, R. (2006). Dynamic recruitment of transcription factors and epigenetic changes on the ER stress response gene promoters. Nucleic acids research *34*, 3116-3127.
Donohoe, M.E., Zhang, X., McGinnis, L., Biggers, J., Li, E., and Shi, Y. (1999). Targeted disruption of mouse Yin Yang 1 transcription factor results in peri-implantation lethality. Molecular and cellular biology *19*, 7237-7244.
Emili, A., Greenblatt, J., and Ingles, C.J. (1994). Species-specific interaction of the glutamine-rich activation domains of Sp1 with the TATA box-binding protein. Molecular and cellular biology *14*, 1582-1593.

Engström, P.G., Suzuki, H., Ninomiya, N., Akalin, A., Sessa, L., Lavorgna, G., Brozzi, A., Luzi, L., Tan, S.L., and Yang, L. (2006). Complex loci in human and mouse genomes. PLoS genetics *2*, e47.
Ferrari, K.J., Scelfo, A., Jammula, S., Cuomo, A., Barozzi, I., Stützer, A., Fischle, W., Bonaldi, T., and Pasini, D. (2014). Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity. Molecular cell *53*, 49-62.
Franck, E., Hulsen, T., Huynen, M.A., de Jong, W.W., Lubsen, N.H., and Madsen, O. (2008). Evolution of closely linked gene pairs in vertebrate genomes. Molecular biology and evolution *25*, 1909-1921.
Gaston, K., and Fried, M. (1994). YY1 is involved in the regulation of the bi-directional promoter of the Surf-1 and Surf-2 genes. FEBS letters *347*, 289-294.
Goldberg, A.D., Allis, C.D., and Bernstein, E. (2007). Epigenetics: a landscape takes shape. Cell *128*, 635-638.
Grant, P.A., Duggan, L., Côté, J., Roberts, S.M., Brownell, J.E., Candau, R., Ohba, R., Owen-Hughes, T., Allis, C.D., and Winston, F. (1997). Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. Genes & development *11*, 1640-1650.
Grant, P.A., Schieltz, D., Pray-Grant, M.G., Steger, D.J., Reese, J.C., Yates, J.R., and Workman, J.L. (1998). A subset of TAF II s are integral components of the SAGA complex required for nucleosome acetylation and transcriptional stimulation. Cell *94*, 45-53.
Grunstein, M., Hecht, A., Fisher-Adams, G., Wan, J., Mann, R., Strahl-Bolsinger, S., Laroche, T., and Gasser, S. (1995). The regulation of euchromatin and heterochromatin by histones in yeast. Journal of Cell Science *1995*, 29-36.
Häkkinen, A., Healy, S., Jacobs, H.T., and Ribeiro, A.S. (2011). Genome wide study of NF-Y type CCAAT boxes in unidirectional and bidirectional promoters in human and mouse. Journal of theoretical biology *281*, 74-83.
Hayes, J.J., and Hansen, J.C. (2001). Nucleosomes and the chromatin fiber. Current opinion in genetics & development *11*, 124-129.
He, Y., Vogelstein, B., Velculescu, V.E., Papadopoulos, N., and Kinzler, K.W. (2008). The antisense transcriptomes of human cells. Science *322*, 1855-1857.
Horvath, P., and Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. Science *327*, 167-170.
Huang, C.-C., and Chang, W.-S.W. (2009). Cooperation between NRF-2 and YY-1 transcription factors is essential for triggering the expression of the PREPL-C2ORF34 bidirectional gene pair. BMC molecular biology *10*, 67.
Kadonaga, J.T., Carner, K.R., Masiarz, F.R., and Tjian, R. (1987). Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. Cell *51*, 1079-1090.
Kadonaga, J.T., Jones, K.A., and Tjian, R. (1986). Promoter-specific activation of RNA polymerase II transcription by Sp1. Trends in Biochemical Sciences *11*, 20-23.
Karmodiya, K., Krebs, A.R., Oulad-Abdelghani, M., Kimura, H., and Tora, L. (2012). H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. BMC genomics *13*, 424.
Kaufmann, J., and Smale, S.T. (1994). Direct recognition of initiator elements by a component of the transcription factor IID complex. Genes & development *8*, 821-829.
Kelly, D.P., and Scarpulla, R.C. (2004). Transcriptional regulatory circuits controlling mitochondrial biogenesis and function. Genes & development *18*, 357-368.
Koch, C.M., Andrews, R.M., Flicek, P., Dillon, S.C., Karaoz, U., Clelland, G.K., Wilcox, S., Beare, D.M., Fowler, J.C., Couttet, P.*, et al.* (2007). The landscape of histone modifications across 1% of the human genome in five human cell lines. Genome research *17*, 691-707.

Koonin, E.V. (2009). Evolution of genome architecture. The international journal of biochemistry & cell biology *41*, 298-306.
Kornberg, R.D., and Lorch, Y. (1999). Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. Cell *98*, 285-294.
Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693-705.
Koyanagi, K.O., Hagiwara, M., Itoh, T., Gojobori, T., and Imanishi, T. (2005a). Comparative genomics of bidirectional gene pairs and its implications for the evolution of a transcriptional regulation system. Gene *353*, 169-176.
Koyanagi, K.O., Hagiwara, M., Itoh, T., Gojobori, T., and Imanishi, T. (2005b). Comparative genomics of bidirectional gene pairs and its implications for the evolution of a transcriptional regulation system. Gene *353*, 169-176.
Krebs, A.R., Karmodiya, K., Lindahl-Allen, M., Struhl, K., and Tora, L. (2011). SAGA and ATAC histone acetyl transferase complexes regulate distinct sets of genes and ATAC defines a class of p300-independent enhancers. Molecular cell *44*, 410-423.
Krogan, N.J., Kim, M., Tong, A., Golshani, A., Cagney, G., Canadien, V., Richards, D.P., Beattie, B.K., Emili, A., Boone, C*., et al.* (2003). Methylation of Histone H3 by Set2 in Saccharomyces cerevisiae Is Linked to Transcriptional Elongation by RNA Polymerase II. Molecular and cellular biology *23*, 4207-4218.
Krom, N., and Ramakrishna, W. (2008). Comparative analysis of divergent and convergent gene pairs and their expression patterns in rice, Arabidopsis, and Populus. Plant physiology *147*, 1763-1773.
Lee, T.I., and Young, R.A. (2000). Transcription of eukaryotic protein-coding genes. Annual review of genetics *34*, 77-137.
Li, J., Moazed, D., and Gygi, S.P. (2002). Association of the histone methyltransferase Set2 with RNA polymerase II plays a role in transcription elongation. The Journal of biological chemistry *277*, 49383-49388.
Li, Y.-Y., Yu, H., Guo, Z.-M., Guo, T.-Q., Tu, K., and Li, Y.-X. (2006). Systematic analysis of head-to-head gene organization: evolutionary conservation and potential biological relevance. PLoS Comput Biol *2*, e74.
Li, Y., Sun, L., Zhang, Y., Wang, D., Wang, F., Liang, J., Gui, B., and Shang, Y. (2011). The histone modifications governing TFF1 transcription mediated by estrogen receptor. The Journal of biological chemistry *286*, 13925-13936.
Lin, J.M., Collins, P.J., Trinklein, N.D., Fu, Y., Xi, H., Myers, R.M., and Weng, Z. (2007a). Transcription factor binding and modified histones in human bidirectional promoters. Genome research *17*, 818-827.
Lin, J.M., Collins, P.J., Trinklein, N.D., Fu, Y., Xi, H., Myers, R.M., and Weng, Z. (2007b). Transcription factor binding and modified histones in human bidirectional promoters. Genome research *17*, 818-827.
Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature *389*, 251-260.
Macleod, D., Charlton, J., Mullins, J., and Bird, A.P. (1994). Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. Genes & development *8*, 2282-2292.
Martinez, E., Chiang, C.-M., Ge, H., and Roeder, R. (1994). TATA-binding protein-associated factor (s) in TFIID function through the initiator to direct basal transcription from a TATA-less class II promoter. The EMBO journal *13*, 3115.
Maxson, R., Cohn, R., Kedes, L., and Mohun, T. (1983). Expression and organization of histone genes. Annual review of genetics *17*, 239-277.
Momota, R., Sugimoto, M., Oohashi, T., Kigasawa, K., Yoshioka, H., and Ninomiya, Y. (1998). Two genes, COL4A3 and COL4A4 coding for the human α3 (IV) and α4 (IV) collagen chains are arranged head-to-head on chromosome 2q36. FEBS letters *424*, 11-16.
Munshi, A., Shafi, G., Aliya, N., and Jyothy, A. (2009). Histone modifications dictate specific biological readouts. Journal of Genetics and Genomics *36*, 75-88.

Myslinski, E., Gérard, M.-A., Krol, A., and Carbon, P. (2006). A genome scale location analysis of human Staf/ZNF143-binding sites suggests a widespread role for human Staf/ZNF143 in mammalian promoters. Journal of Biological Chemistry *281*, 39953-39962.
Nagy, Z., Riss, A., Romier, C., Le Guezennec, X., Dongre, A.R., Orpinell, M., Han, J., Stunnenberg, H., and Tora, L. (2009). The human SPT20-containing SAGA complex plays a direct role in the regulation of endoplasmic reticulum stress-induced genes. Molecular and cellular biology *29*, 1649-1660.
Ntini, E., Järvelin, A.I., Bornholdt, J., Chen, Y., Boyd, M., Jørgensen, M., Andersson, R., Hoof, I., Schein, A., and Andersen, P.R. (2013a). Polyadenylation site–induced decay of upstream transcripts enforces promoter directionality. Nature structural & molecular biology *20*, 923-928.
Ntini, E., Jarvelin, A.I., Bornholdt, J., Chen, Y., Boyd, M., Jorgensen, M., Andersson, R., Hoof, I., Schein, A., Andersen, P.R*., et al.* (2013b). Polyadenylation site-induced decay of upstream transcripts enforces promoter directionality. Nature structural & molecular biology *20*, 923-928.
Odrowaz, Z., and Sharrocks, A.D. (2012). The ETS transcription factors ELK1 and GABPA regulate different gene networks to control MCF10A breast epithelial cell migration. PloS one *7*, e49892.
Ogryzko, V.V., Schiltz, R.L., Russanova, V., Howard, B.H., and Nakatani, Y. (1996). The transcriptional coactivators p300 and CBP are histone acetyltransferases. Cell *87*, 953-959.
Ohtani, K., and Dimmeler, S. (2011). Epigenetic regulation of cardiovascular differentiation. Cardiovascular research, cvr019.
Oka, S., Shiraishi, Y., Yoshida, T., Ohkubo, T., Sugiura, Y., and Kobayashi, Y. (2004). NMR structure of transcription factor Sp1 DNA binding domain. Biochemistry *43*, 16027-16035.
Palko, L., Bass, H.W., Beyrouthy, M.J., and Hurt, M.M. (2004). The Yin Yang-1 (YY1) protein undergoes a DNA-replication-associated switch in localization from the cytoplasm to the nucleus at the onset of S phase. Journal of cell science *117*, 465-476.
Pascal, E., and Tjian, R. (1991). Different activation domains of Sp1 govern formation of multimers and mediate transcriptional synergism. Genes & development *5*, 1646-1656.
Petkova, V., Romanowski, M.J., Sulijoadikusumo, I., Rohne, D., Kang, P., Shenk, T., and Usheva, A. (2001). Interaction between YY1 and the Retinoblastoma Protein REGULATION OF CELL CYCLE PROGRESSION IN DIFFERENTIATED CELLS. Journal of Biological Chemistry *276*, 7932-7936.
Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S., Mapendano, C.K., Schierup, M.H., and Jensen, T.H. (2008). RNA exosome depletion reveals transcription upstream of active human promoters. Science *322*, 1851-1854.
Ramírez-Solis, R., Rivera-Pérez, J., Wallace, J.D., Wims, M., Zheng, H., and Bradley, A. (1992). Genomic DNA microextraction: a method to screen numerous samples. Analytical biochemistry *201*, 331-335.
Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nature protocols *8*, 2281-2308.
Reed, B.D., Charos, A.E., Szekely, A.M., Weissman, S.M., and Snyder, M. (2008). Genome-wide occupancy of SREBP1 and its partners NFY and SP1 reveals novel functional roles and combinatorial regulation of distinct classes of genes. PLoS genetics *4*, e1000133-e1000133.
Rocha, E.P. (2008). The organization of the bacterial genome. Annual review of genetics *42*, 211-233.
Sandelin, A., Carninci, P., Lenhard, B., Ponjavic, J., Hayashizaki, Y., and Hume, D.A. (2007). Mammalian RNA polymerase II core promoters: insights from genome-wide studies. Nature Reviews Genetics *8*, 424-436.
Schaft, D., Roguev, A., Kotovic, K.M., Shevchenko, A., Sarov, M., Shevchenko, A., Neugebauer, K.M., and Stewart, A.F. (2003). The histone 3 lysine 36 methyltransferase, SET2, is involved in transcriptional elongation. Nucleic acids research *31*, 2475-2482.
Schmidt, C., Fischer, G., Kadner, H., Genersch, E., Kühn, K., and Pöschl, E. (1993). Differential effects of DNA-binding proteins on bidirectional transcription from the common promoter region of human

collagen type IV genes COL4A1 and COL4A2. Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression *1174*, 1-10.
Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., and Sharp, P.A. (2008). Divergent transcription from active promoters. Science *322*, 1849-1851.
Seila, A.C., Core, L.J., Lis, J.T., and Sharp, P.A. (2009). Divergent transcription: a new feature of active promoters. Cell cycle *8*, 2557-2564.
Sementchenko, V.I., and Watson, D.K. (2000). Ets target genes: past, present and future. Oncogene *19*, 6533-6548.
Serra, E., Zemzoumi, K., Lardans, V., Dissous, C., di Silvio, A., and Mantovani, R. (1998). Conservation and divergence of NF-Y transcriptional activation function. Nucleic acids research *26*, 3800-3805.
Shi, Y., Lee, J.-S., and Galvin, K.M. (1997). Everything you have ever wanted to know about Yin Yang 1. Biochimica et Biophysica Acta (BBA)-Reviews on Cancer *1332*, F49-F66.
Shin, H., Liu, T., Manrai, A.K., and Liu, X.S. (2009). CEAS: cis-regulatory element annotation system. Bioinformatics *25*, 2605-2606.
Sindhu, C., Samavarchi-Tehrani, P., and Meissner, A. (2012). Transcription factor-mediated epigenetic reprogramming. Journal of Biological Chemistry *287*, 30922-30931.
Spedale, G., Timmers, H.T.M., and Pijnappel, W.P. (2012). ATAC-king the complexity of SAGA during evolution. Genes & development *26*, 527-541.
Spencer, T.E., Jenster, G., Burcin, M.M., Allis, C.D., Zhou, J., Mizzen, C.A., McKenna, N.J., Onate, S.A., Tsai, S.Y., and Tsai, M.-J. (1997). Steroid receptor coactivator-1 is a histone acetyltransferase. Nature *389*, 194-198.
Sperling, S. (2007). Transcriptional regulation at a glance. BMC bioinformatics *8*, S2.
Steger, D.J., Lefterova, M.I., Ying, L., Stonestrom, A.J., Schupp, M., Zhuo, D., Vakoc, A.L., Kim, J.E., Chen, J., Lazar, M.A*., et al.* (2008). DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells. Molecular and cellular biology *28*, 2825-2839.
Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. Nature *403*, 41-45.
Sugimoto, M., Oohashi, T., and Ninomiya, Y. (1994). The genes COL4A5 and COL4A6, coding for basement membrane collagen chains alpha 5 (IV) and alpha 6 (IV), are located head-to-head in close proximity on human chromosome Xq22 and COL4A6 is transcribed from two alternative promoters. Proceedings of the National Academy of Sciences *91*, 11679-11683.
Takai, D., and Jones, P.A. (2004). Origins of bidirectional promoters: computational analyses of intergenic distance in the human genome. Molecular biology and evolution *21*, 463-467.
Tong, C.-W., Wang, J.-L., Jiang, M.-S., Hsu, C.-H., Chang, W.-T., and Huang, A.-M. (2013). Novel genes that mediate nuclear respiratory factor 1-regualted neurite outgrowth in neuroblastoma IMR-32 cells. Gene *515*, 62-70.
Trinklein, N.D., Aldred, S.F., Hartman, S.J., Schroeder, D.I., Otillar, R.P., and Myers, R.M. (2004). An abundance of bidirectional promoters in the human genome. Genome research *14*, 62-66.
Tsai, F.T., and Sigler, P.B. (2000). Structural basis of preinitiation complex assembly on human pol II promoters. The EMBO journal *19*, 25-36.
Tsai, H.-K., Su, C.P., Lu, M.-Y.J., Shih, C.-H., and Wang, D. (2007). Co-expression of adjacent genes in yeast cannot be simply attributed to shared regulatory system. BMC genomics *8*, 352.
van Kruijsbergen, I., Hontelez, S., and Veenstra, G.J. (2015). Recruiting polycomb to chromatin. The international journal of biochemistry & cell biology.
Vernon, E.G., and Gaston, K. (2000). Myc and YY1 mediate activation of the Surf-1 promoter in response to serum growth factors. Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression *1492*, 172-179.
Wakano, C., Byun, J.S., Di, L.J., and Gardner, K. (2012). The dual lives of bidirectional promoters. Biochimica et biophysica acta *1819*, 688-693.

Wang, G., Qi, K., Zhao, Y., Li, Y., Juan, L., Teng, M., Li, L., Liu, Y., and Wang, Y. (2013). Identification of regulatory regions of bidirectional genes in cervical cancer. BMC medical genomics *6 Suppl 1*, S5.
Wang, Q., Wan, L., Li, D., Zhu, L., Qian, M., and Deng, M. (2009). Searching for bidirectional promoters in Arabidopsis thaliana. BMC bioinformatics *10*, S29.
Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Peng, W., Zhang, M.Q.*, et al.* (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet *40*, 897-903.
Wasylyk, B., Derbyshire, R., Guy, A., Molko, D., Roget, A., Teoule, R., and Chambon, P. (1980). Specific in vitro transcription of conalbumin gene is drastically decreased by single-point mutation in TATA box homology sequence. Proceedings of the National Academy of Sciences *77*, 7024-7028.
Weis, L., and Reinberg, D. (1997). Accurate positioning of RNA polymerase II on a natural TATA-less promoter is independent of TATA-binding-protein-associated factors and initiator-binding proteins. Molecular and cellular biology *17*, 2973-2984.
Yang, L., and Yu, J. (2009). A comparative analysis of divergently-paired genes (DPGs) among Drosophila and vertebrate genomes. BMC evolutionary biology *9*, 55.
Yang, M.Q., and Elnitski, L.L. (2008a). Diversity of core promoter elements comprising human bidirectional promoters. BMC genomics *9*, S3.
Yang, M.Q., and Elnitski, L.L. (2008b). Diversity of core promoter elements comprising human bidirectional promoters. BMC genomics *9 Suppl 2*, S3.
Yang, M.Q., and Elnitski, L.L. (2008c). Prediction-based approaches to characterize bidirectional promoters in the mammalian genome. BMC genomics *9*, S2.
Yang, X.-J., Ogryzko, V.V., Nishikawa, J.-i., Howard, B.H., and Nakatani, Y. (1996). A p300/CBP-associated factor that competes with the adenoviral oncoprotein E1A.
Zhu, Q.S., Chen, K., and Shih, J.C. (1994). Bidirectional promoter of human monoamine oxidase A (MAO A) controlled by transcription factor Sp1. The Journal of neuroscience *14*, 7393-7403.

Chapter 3: General Discussion of thesis work

3.1 Discussion

This study focuses on analyzing the mechanism of transcription elongation from bidirectional promoters. In this thesis I have studied the role of histone modifications and transcription factors in regulation of transcription from bidirectional promoters. Towards understanding the mechanism governing bidirectional transcription, we searched for a reporter system that can be used to score for bidirectional transcription from any given DNA element. Until now, a luciferase reporter based system was used to characterize the activity of bidirectional promoters (Trinklein et al., 2004). Previous studies have scored the bidirectional transcription by cloning the intergenic regions of bidirectional gene pairs in both sense and antisense orientation in pGL3 basic reporter vector (Lin et al., 2007a; Trinklein et al., 2004). Rationale behind these assays was to score the luciferase reporter activity from both sense and antisense clones. Any intergenic region acting as a true bidirectional promoter would exhibit reporter activity from both sense and antisense clones. However, one major limitation in this system is that one can score for firing of the promoter in only one direction at a time. To overcome this limitation, we have designed and constructed a vector system that can be used for studying bidirectional promoters wherein any DNA fragment acting as bonafide bidirectional promoter will express two reporter proteins independent of its orientation. We have named this construct as the pDR (plasmid with Dual Reporter) vector. To validate this vector in vivo we cloned the CMV promoter which is a well characterized unidirectional promoter (Chambers et al., 2015) and demonstrated that it drives the expression of GFP or mCherry depending on whether this promoter was cloned in sense or antisense orientation between the two reporters. Further, we cloned 8 bidirectional promoters from human genome in the pDR vector and characterized them in vivo qualitatively and quantitatively by fluorescence microscopy and flow cytometry respectively. Thus this reporter system can be used to monitor bidirectional transcription from any given DNA element.

Our understanding of transcription has improved significantly with the advancement of next generation sequencing technologies. Recent studies have documented that all mammalian active promoters intrinsically harbor the capability to transcribe in both sense and antisense directions, however, antisense transcripts get aborted after polymerization of 50-250 nucleotides followed by immediate degradation by the RNA exosome degradation machinery (Core et al., 2008; Preker et al., 2008; Seila et al., 2008). In contrast, bidirectional promoters exhibit unique property of generating mature and functional transcripts in both the sense and antisense orientations. This interesting property of bidirectional promoters sets them apart from other transcribing loci. However, regulation of bidirectional transcription is not clearly understood. Transcription factors, histone modifications, various promoter elements and higher-order complexes have been shown to regulate bidirectional transcription in various model systems (Kouzarides, 2007; Li et al., 2007; Wakano et al., 2012). Previous studies have revealed that bidirectional promoters harbor more active histone modifications as compared to expression matched unidirectional promoters, however, the role played by these histone modification in driving antisense transcription from bidirectional promoter is not well understood. (Lin et al., 2007b). Recently it has been documented that active promoter marks exhibit enrichment in both sense and antisense directions only on bidirectional promoters as compared to other promoters (Bornelöv et al., 2015). The presence of active promoter mark H3K4me3 in antisense direction is an established fact and might be responsible for initiation of antisense transcription from human promoters (Seila et al., 2008). However, the most striking difference between any unidirectional and bidirectional promoter is mature transcription elongation in antisense direction which is specific only to bidirectional promoters. In the case of unidirectional promoters antisense transcription occurs but is inherently abortive in nature.

It has been known for years that post-translational modifications of histone proteins alone or in combination play crucial roles in transcription (Kouzarides, 2007). The Histone code paradigm suggests that histone modifications act as a molecular code that is read by regulatory proteins which either activate or repress the transcription process (Strahl and Allis, 2000). Using the knowledge of histone modifications and their effect on transcription, we asked if there is any link between epigenetic histone modifications and

bidirectional transcription. To test if histone modifications play any role in regulation of bidirectional promoter, the bidirectional promoter for the gene pair (NFYA/ OARD1) that exhibited bidirectional transcription in the pDR vector assay was chosen for chromatin immunoprecipitation (ChIP) analysis in Jurkat cells. ChIP analysis in Jurkat cells revealed that active promoter marks H3K4me3, H3K9ac and H4ac are enriched in both sense and antisense orientations. Enrichment of active promoter marks in antisense orientation was supported by a recent study which also revealed that active promoter marks are enriched in antisense direction only on bidirectional promoters as compared to unidirectional promoters (Bornelöv et al., 2015). Although unidirectional promoters have also been shown to harbor H3K4me3 mark in antisense orientation, nevertheless the antisense transcription does not mature from these promoters. These studies suggest that antisense transcription initiation is a property of eukaryotic promoters and transcription machinery, however, hitherto uncharacterized mechanisms do operate to regulate mature transcription from bidirectional promoters. These interesting results prompted us to analyze the epigenetic landscape of bidirectional promoters as compared to all other promoter regions. To test our hypothesis comprehensively, we analyzed genome-wide occurrences of 39 histone modifications specifically at all bidirectional promoters and compared to that of all other promoters in the human CD4 T-cells. To rule out the bias of variable transcription, only the promoters whose genes exhibited similar expression levels were compared between the two categories. We found that multiple active promoter associated histone modification marks H3K4me1, H3K4me3, H3K23ac, H2AK9ac and H3K36me1 exhibit enrichment in bimodal fashion on the bidirectional promoters whereas they are enriched only in the sense orientation on unidirectional promoters. Presence of active promoter marks in the antisense direction on bidirectional promoters might change the local chromatin environment which favors the proper initiation of antisense transcription. Next we analyzed the enrichment of H2BK5me1, which has been reported to occur downstream of actively transcribing promoters (Barski et al., 2007). Interestingly, H2BK5me1 enrichment was observed to occur in a pattern concomitant with active transcription on bidirectional promoters. Most strikingly, H3K27me1, H4K20me1 along with H3K36me3 and H3K79me3, which are associated with successful mature transcription, occur in a bimodal distribution only on bidirectional

promoters. Furthermore, this pattern overlays the distribution of RNA-seq tags on the genes regulated by these promoters, corroborating their mature transcription.

To test the significance of the observed bimodal pattern of active transcription associated histone marks, we studied the distribution of epigenetic marks on bidirectional gene pairs where only one gene out of the pair is transcribed. In this situation, the bidirectional promoter would behave similar to any unidirectional promoter with respect to epigenetic marks. Histone modification profiles of a bidirectional gene pair which exhibits expression of only one of the two genes were compared with that of the unidirectionally transcribing genes. Our analysis revealed that when both the genes of a bidirectional gene pair are transcribed, active promoter and elongation marks are enriched in both sense and antisense orientations with respect to TSS of the bidirectional gene pair, whereas these marks are enriched only in the sense direction on the unidirectional gene. When only one of the gene expresses from bidirectional gene pairs ("OneUP" or "OneDOWN" category), interestingly the enrichment of promoter and transcription elongation marks show a pattern similar to unidirectional promoters. We went on to analyze other transcription activation and elongation marks for all four categories of bidirectional gene pairs. We observed enrichment of H2BK5me1, H3K79me3, H3K36me3 and H4K20me1 in both the gene bodies of bidirectional gene pairs only when both genes are transcribed and they show enrichment only in the sense orientation to unidirectional gene. When only one gene is expressed from the bidirectional gene pair, all marks exhibited an enrichment pattern equivalent to unidirectional gene. Peaks of H3K4me1 and H3K4me2 were observed in the antisense orientation only if both genes are expressed, however, if only one of the genes among a bidirectional gene pair is expressed then both the promoter marks show enrichment equivalent to a unidirectional gene. From the above observations, it is evident that the histone marks are associated with active transcription both at the promoters and the gene body mirroring the transcriptional status of genes. These marks are exclusive to transcriptional activity and even in the bidirectional gene pairs wherein two genes are in a head to head orientation, the transcriptional status of the genes dictates the presence of the corresponding histone marks. One might argue that observed bimodal distribution of active transcription associated histone modifications

on bidirectional promoter may be a result of two genes organized in head to head orientation.

To rule out this possibility we performed ChIP analysis on plasmid DNA (pDR vector). ChIP analysis revealed that the transcription elongation marks H3K79me3, H3K36me3 and H3K27me1 exhibited bimodal enrichment only on bidirectional promoter clones, however, we observed enrichment of these marks on CMV promoter only in the sense direction. Thus, it is clear that the intergenic region is almost solely responsible for recruiting transcription factors and RNA PolII to both the sense and the antisense genes in a bidirectional pair. If the mere act of mature transcription in both orientations was enough to deposit histone marks in the gene bodies of GFP and mCherry (genes not normally found in any mammalian system), then the role of ascribing bidirectional transcription status lies solely with the intergenic region and the bimodal distribution of histone marks follows as a consequence of this. This is supported by our observation that in a situation where only one out of the two genes in a pair is transcribed, the histone marks seem to mirror transcriptional status as opposed to just the presence of a bidirectionally responsive promoter element.

Recently, it has been suggested that human promoters are intrinsically unidirectional and antisense transcription arises due to antisense cognate promoter elements (Duttke et al., 2015). However, bidirectional promoters have been shown to share common regulatory region which acts as a promoter for both sense and antisense genes (Trinklein et al., 2004). Mutation or deletion in these common regulatory elements has been shown to affect both sense and antisense transcription, suggesting an important role for the common intergenic region in the regulation of bidirectional transcription (Lin et al., 2007a; Trinklein et al., 2004). These intergenic regions are fairly conserved across all mammalian species (Piontkivska et al., 2009b) but The functional significance of conservation of intergenic region distance in the bidirectional promoters is not known.

There are many studies which have analyzed the intergenic distance between head to head arranged gene pairs in different eukaryotes. These studies have shown that intragenic distance between bidirectional gene pairs is conserved across the species (Davila Lopez et al., 2010; Koyanagi et al., 2005). It has been reported that nearly 67% of

head to head gene pairs have intergenic distance less than 300 bp (Trinklein et al., 2004). Conservation of the intergenic region has been documented, however, the functional importance of maintaining this distance between the two TSS is not understood (Piontkivska et al., 2009a). This constraint of distance between genes vis a vis the length of a bidirectional promoter could be involved in facilitating bidirectional transcription from these promoters. To address this question, we analyzed the gene pairs which are arranged in head to head orientation with the intergenic distance varying from 500 bp to 10 Kb. We hypothesized that if intergenic distance conservation is an important factor for bidirectional promoter function then the observed bimodal distribution of different epigenetic marks would get affected as a function of increasing intergenic distance. We analyzed ChIP-seq reads for various histone modifications for the head to head gene pairs which harbor intergenic distance from 500 bp to 10 Kb. Analysis were performed up to 10 Kb upstream and downstream of the TSS for these gene pairs. Interestingly, as the intergenic distance increased between the head to head gene pairs, it correlated with a sharp decrease in the active promoter and transcription elongation marks, specifically in the antisense orientation. Our data identify unique epigenetic landscape of bidirectional promoters which can be used to predict if two genes are present in close proximity would be co-regulated by bidirectional promoter.

To test if the bimodal distribution of active transcription-associated histone modifications correlate with transcriptional output from bidirectional promoter in cellular context, we used differentiation of the human embryonic carcinoma NT2D1 cells as a model system. NT2D1 cells were differentiated to neuronal phenotype using retinoic acid treatment followed by screening of differentially expressed bidirectional gene pairs. We found out that the NUP62CL-PIH1D3 gene pair exhibited 10-fold increased expression in differentiated state of cells as compared to control NT2D1 cells. ChIP analysis was performed for various transcription elongation marks (H3K79me3, H3K36me3 and H3K27me1). We analyzed the enrichment of these elongation marks in 2 Kb upstream and downstream region from TSS of NUP62CL-PIH1D3 gene pair including the intergenic region. Interestingly, all the 3 elongation marks revealed significantly high enrichment downstream to TSS on both gene bodies, further strengthening the

significance of bimodal distribution of elongation marks in mature bidirectional transcription from sense and antisense genes.

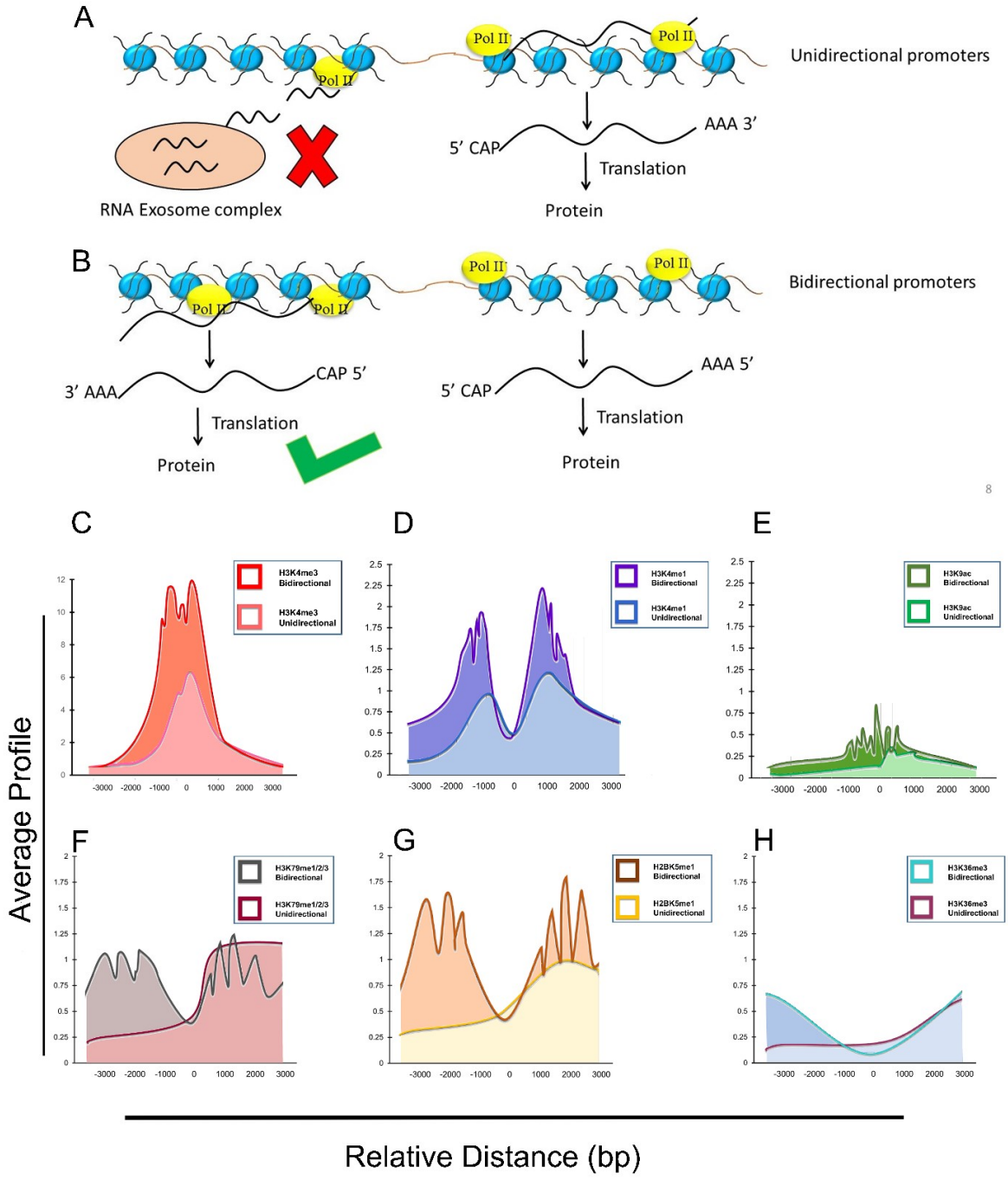


Figure 3.1 Model depicting the role of epigenetic modifications in regulation of bidirectional promoter. (A) Unidirectional promoters exhibit mature transcription only in sense direction. Antisense transcripts get aborted after 50-250 nucleotides which are subsequently degraded by the RNA exosome

machinery. (B) Bidirectional promoters display mature transcription in sense and antisense direction. (C-F) Distribution of active promoter marks on unidirectional and bidirectional promoter display contrasting differences. All active promoter marks display bimodal enrichment profile on bidirectional promoters while enrichment was observed only in sense orientation on the unidirectional promoters. (D-F) Transcription elongation marks (H3K79me3, H3K36me3) are enriched only in the sense orientation and no enrichment of elongation marks in antisense orientation is observed on unidirectional promoters, which correlates with their transcriptional output. Bidirectional promoters that exhibit expression of both sense and antisense genes, display bimodal distribution of elongation marks (H3K79me3, H3K36me3), suggesting that differential enrichment of epigenetic marks enables the transcription machinery to transcribe in both sense and antisense orientations. For details see text.

Thus, we demonstrate for the first instance a strong correlation between the epigenetic marks on the bidirectional promoters and the transcriptional status of the bidirectional genes. Observed bimodal distribution of epigenetic marks might be one of unknown mechanisms which presumably plays an important role in mature sense and antisense transcription from bidirectional promoters. We propose that unique distribution of the active transcription marks on bidirectional promoters signals the transcription machinery to drive transcription in both sense and antisense orientations, however these active epigenetic marks exhibit enrichment only in one orientation on unidirectional promoters and therefore transcription maturation progresses only in one direction. These findings strongly suggest that occurrence of these marks correlates with the process of active transcription maturation from the bidirectional promoters. Our study opens up an interesting possibility that mature transcription in the silenced gene is oppressed by a separate mechanism that does not allow the deposition of histone marks that enhance transcription. Thus, it is evident that histone deposition can act as memory of active transcription status but just how downstream deposition of histone marks is regulated remains an open question. Our data identifies candidate bidirectional gene pairs that can be further used for functional analysis of transcriptional suppression of bidirectional transcription.

In the second part of thesis we aimed to study the role of transcription factors in regulation of bidirectional promoters. Most of the studies regarding the role of transcription factors in regulation of bidirectional promoters are based on reporter assays

or occupancy of these factors on bidirectional promoters. However, how binding of these transcription factors regulates transcription from these promoters is poorly understood. Binding of a transcription factor might result in activation or repression of gene expression depending on whether it recruits a coactivator or corepressor. A typical transcription factor harbors multiple functional domains, not only for binding to specific DNA sequence, but also for interaction with other activator or repressor complexes (Hahn and Young, 2011; Näär et al., 2001). Despite large number of investigations and reports in literature, how these transcription factors regulate transcription from bidirectional promoters is poorly understood. In this chapter, we have focused on the role of key transcription factors in regulation of bidirectional promoters.

Several studies have focused on delineating common motif(s) in bidirectional promoters. One common feature that has been discussed is the overrepresentation of few specific transcription factor-binding motifs. Motifs for binding of several transcription factors including GABPA, MYC, E2F1, E2F4, NRF-1, CCAAT and YY1 have been shown to be overrepresented on bidirectional promoters as compared to unidirectional promoters (Lin et al., 2007a). Occurrence of GABPA-binding site is correlated with bidirectional transcription and the same has been validated by luciferase reporter assay. It has been clearly demonstrated that any unidirectional promoter when cloned in dual luciferase reporter vector system exhibits reporter activity only in one direction; however, addition of GABPA consensus binding sites to same promoter results in bidirectional reporter activity. This finding suggests that the presence of GABPA consensus binding site promotes bidirectional transcription (Collins et al., 2007). Moreover, binding of GABPA was verified with 121 and 291 randomly selected bidirectional and unidirectional promoters respectively in three different human cell lines viz. HeLa, Jurkat, and K562 by performing chromatin immunoprecipitation analysis. Although the referred study provides evidence showing positive correlation between occurrence of GABPA site and the bidirectionality of the promoter, the molecular mechanism underlying this regulation is not well understood. Though bidirectional promoters are enriched for NRF1 motif (Lin et al., 2007b), the role of NRF1 in bidirectional promoter regulation is poorly understood. It is noteworthy that there has been only one study till date demonstrating the binding to NRF1 to GPAT-AIRC bidirectional promoter (Chen et al., 1997). Nuclear Transcription

Factor Y, Alpha (NFYA) has been shown to influence the directionality of transcription from Mrps12-Sarsm gene pair in both human and mouse cells (Zanotto et al., 2009).) Genome-wide binding analysis of NFYA to unidirectional and bidirectional promoter revealed that bidirectional promoters have specific allocation of CCAAT boxes to which NF-YA binds (Häkkinen et al., 2011). This unique allocation of CCAAT boxes on bidirectional promoters suggests for a prominent role of NFYA in bidirectional promoter regulation.

YY1 has been shown to bind and regulate the expression of human Surf-1-Surf-2 bidirectional gene pair in response to serum growth factors (Cole and Gaston, 1997; Gaston and Fried, 1994). Furthermore, CpG methylation has differential effects on binding of YY1 and ETS related proteins to human Surf-1-Surf-2 bidirectional promoter (Cole and Gaston, 1997). Another study showed that YY1 cooperates with MYC to positively regulate the expression of Surf-1-Surf-2 gene pair (Vernon and Gaston, 2000). The GABPA and ATP synthase coupling factor 6 bidirectional gene pair harbors binding sites for YY1, NRF1, SP1 and GABPA (Chinenov et al., 2000), suggesting a possible cross-talk between these factors in the regulation of bidirectional promoters. Apart from binding to bidirectional promoters, virtually nothing is known about the mechanistic role played by YY1 in regulating transcription from bidirectional promoters. Despite the advancement in modern genomics and proteomics techniques we still do not have answer for certain fundamental questions regarding the transcriptional process from the bidirectional promoters. In this study, we attempt to elucidate the roles of transcription factors in bidirectional promoter regulation. First to delineate the role of specific transcription factors we chose the approach of siRNA mediated knockdown of transcription factors followed by quantitative RT-PCR (qRT-PCR) analysis of bidirectional gene pairs. We selected specific bidirectional promoters based on the presence of binding sites of these overrepresented transcription factors. To test the significance of this overrepresentation, knockdown of SP1, YY1, NRF1 and GABPA was performed followed by quantitative transcript profiling by real-time RT-PCR analysis of 62 bidirectional genes. Surprisingly, we did not observe any dysregulation in transcription of bidirectional genes upon knockdown of these individual factors. Next, double knockdown of these factors was performed in combinations of SP1 and YY1, YY1 and NRF1 and

NRF1 and SP1. Interestingly, double knockdown also did not yield any significant dysregulation in bidirectional gene expression barring few gene pairs which exhibited mild dysregulation. These results suggest two things about the function of these transcription factors in bidirectional promoter regulation; first, these factors might not be important for bidirectional promoter regulation and secondly, there might be a redundancy in the function of these factors towards regulation of bidirectional promoters.

In-depth literature survey of these factors in the recruitment of activator complexes led us to a study which showed that GABPA co-occupies with SAGA coactivator complex on many SAGA complex genomic targets (Krebs et al., 2011). The SAGA complex has been shown to acetylate histones and helps in recruitment RNA polymerase II, which suggest a prominent role of SAGA complex in transcription (Bonnet et al., 2014). Role of the SAGA complex in transcription process has been studied extensively, however, its role in the regulation of transcription from the bidirectional promoters has not been studied.

Interestingly, GABPA also has been shown to bind multiple bidirectional promoters in various cell types (Collins et al., 2007). We used the co-occupancy data for GABPA and SAGA complex to find out if these two factors co-occupy the bidirectional promoters. Our analysis revealed that 54 bidirectional gene pairs exhibit co-occupancy of GABPA and SAGA complex. We randomly selected 9 gene pairs (18 genes) from this list for further analysis. To test if SAGA complex binds to the intergenic regions of these 9 gene pairs, ChIP assay was performed using antibodies specific to SPT20 and PCAF, which are the structural and catalytic subunits of the SAGA complex. ChIP analysis revealed that SAGA complex binds to the intergenic region of all the 9 gene pairs selected. We also showed that GABPA direclty interacts with the SAGA complex. These results encouraged us to analyze the expression of those 18 genes which showed co-occupancy of SAGA and GABPA upon GABPA knockdown. Surprisingly, we did not observe any dysregulation in any of these genes, suggesting the involvement of additional factors in regulation of bidrectional promoters. Further, co-immunoprecipitation assay revealed that SP1 and YY1 also interact with SAGA complex along with GABPA. To determine if YY1 and SP1 could also bind to the SAGA complex-bound sites on the 9 selected gene pairs, we performed ChIP assay using antibodies for YY1, SP1, and GABPA. We observed that

SP1 and YY1 exhibited binding on the intergenic regions of all 9 gene pairs, however GABPA was found to bind to only 2 intergenic regions. This finding presumably provides explanation for the lack of any significant effect on transcription of bidirectional gene pairs upon GABPA knockdown. First, GABPA binds only to 2 intergenic regions, and second, in absence of GABPA, SP1 and YY1 can recruit the SAGA complex to bidirectional promoters. To determine the redundancy in the function of SP1, YY1 and GABPA towards the regulation of bidirectional promoters, double knockdown of SP1 and YY1 was performed. However, this double knockdown also did not result in any dysregulation in the bidirectional gene expression.

Our data suggested that although SP1, YY1 and GABPA can interact with the SAGA complex and bind to bidirectional promoters, the removal of these transcription factors does not affect bidirectional gene expression. Next, we tested involvement of another overrepresented transcription factor NRF1 in the regulation of bidirectional transcription from SAGA occupied genes. Co-immunoprecipitation analysis for NRF1 and SAGA complex revealed that NRF1 interacts directly with the SAGA complex. Furthermore, we showed that NRF1 also binds to the SAGA occupied intergenic regions of the bidirectional gene pairs. Combined together, our results demonstrate that transcription factors SP1, YY1 and NRF1 interact with the SAGA complex and co-occupy the SAGA complex-bound bidirectional promoters.

To conclusively demonstrate the redundancy of these factors in bidirectional transcription we depleted all the three factors. However we could not achieve efficient triple knockdown for SP1, YY1 and NRF1 using the siRNA transfection mediated silencing approach. To circumvent this problem, we used the CRISPAR/CAS9 genomic editing tool to generate knockouts for these factors. We generated NRF1 and SP1 knockout cell lines which were used as the background to knockdown other factors to address the redundancy between these transcription factors. siRNA mediated depletion of SP1 and YY1 in NRF1 knockout background exhibited significant downregulation of expression from bidirectional gene pairs. However, YY1 knockdown in SP1 knockout background did not lead to any dysregulation of bidirectional gene expression.

Collectively, our results show that the transcription factors SP1, YY1 and NRF1 interact with the SAGA complex and co-occupy SAGA complex-bound bidirectional promoters. SAGA complex plays an important role in transcription by histone acetylation, deubiquitination and TBP delivery to regulate transcription (Baker and Grant, 2007; Bonnet et al., 2014). The SAGA complex is known to deposit and interact with a number of histone modifications, including acetylation, methylation, ubiquitylation and phosphorylation (Baker and Grant, 2007). Deubiquitination of H2B by the SAGA complex has been shown to affect histone H3 methylation status which in turn affect the gene transcriptional activity (Weake and Workman, 2008). The deubiquitination activity residing in the SAGA complex also crosstalks with H3K36me3 and S2 phosphorylation of RNA PolII, both of which are mark of active transcription elongation. The SAGA complex is a transcription coactivator complex that regulates transcription by coordinated of multiple post-translational modifications of histone proteins. The SAGA complex performs the acetylation of histone H3 at various positions including H3K9ac, H3K14ac, H3K18ac and H3K23ac. In Chapter 1 I have discussed the enrichment of H2K23ac in the antisense orientation on bidirectional promoters. Along with the acetylation activity, SAGA is known to generate and interact with number of histone modifications, including acetylation, methylation, ubiquitylation and phosphorylation (Baker and Grant, 2007). Ubiquitination activity of the SAGA complex has been shown to play an important role in phosphorylating the S2 of RNA PolII which is a mark of elongating RNA PolII. H2B ubiquitination and deubiquitination by SAGA complex regulates multiple methylation events on histone H3. Histone modifications, and even modification of RNA polymerase II, are involved in the crosstalk at multiple steps during transcription and may serve as checkpoints for the correct assembly of the machinery required to accurately load and launch RNA polymerase during gene expression. We are therefore more interested in understanding the crosstalk between the SAGA complex with other histone medications. It will be interesting to study which of the activities of the SAGA complex helps in regulation of the bidirectional promoters.

The depletion of SP1, YY1 and NRF1 individually or in combinations of two does not lead to any effect on bidirectional gene pairs, presumably pointing towards their redundancy. However, depletion of all the three factors (SP1, YY1 and NRF1) together

results in significant downregulation of most of these bidirectional gene pairs. Depletion of all three factors might result in impaired recruitment of the SAGA complex on the studied bidirectional promoters, resulting in loss of SAGA transcriptional activity on these loci. Collectively, we have demonstrated that the transcription factors SP1, YY1 and NRF1 function in a redundant manner to regulate gene expression from bidirectional promoters.

3.2 Future experiments

In summary, results presented in this thesis reveal comprehensive epigenomic landscape of bidirectional promoters, which correlates with the transcriptional output from these loci. We also demonstrate redundancy in the function of transcription factors in the regulation of bidirectional promoters. Based on the results presented in the thesis we propose the following experiments to further prove or disprove our model unequivocally in future.

1. To test whether the bimodal presence of histone marks enables transcription in both directions, or they are present as a consequence of bidirectional transcription occurring at these sites, we propose ChIP analysis using pDR bidirectional promoter clones for histone active marks in presence of α-Amanitin or actinomycin D. In brief, we would transfect the pDR bidirectional promoter and CMV promoter clones in 293T cells which had been treated with α-Amanitin or actinomycin D prior to transfection. We believe that bidirectional promoters harbor some unknown signature which are read by cellular chromatin modifying machinery resulting in the deposition of bimodal active marks which further helps in bidirectional transcription. If histone marks are consequence of active transcription, then we would not observe the enrichment of any active histone modification marks on plasmid DNA. To get genome-wide comprehensive picture, similar kind of experiment can be

performed in cell lines wherein we proposed to perform ChIP-seq analysis for various histone marks in control and α-Amanitin or Actinomycin D treated cells.

2. In continuation with experiment suggested above, we proposed to perform RNA-seq experiment after perturbing the activity of specific histone methyl- and acetyl-transferases by using specific inhibitors or siRNA mediated knockdown. Results of this experiment will also help us to understand the cause and consequence relationship between transcription and histone modifications.

3. We propose RNA-seq experiment after depletion of SP1, YY1 and NRF1 individually as well as in combination (siRNA mediated depletion of SP1 and YY1 in NRF1 knockout background). Results from this experiment will reveal the level of redundancy in function of these factors in bidirectional promoter regulation.

4. ChIP analysis of the SAGA complex after depletion of SP1, YY1 and NRF1 performed individually or in various combinations and all three knocked down simultaneously.

5. ChIP-seq analysis of the SAGA complex after depletion of SP1, YY1 and NRF1 performed individually or in various combinations and all three knocked down simultaneously. We have already proposed RNA-seq analysis in similar conditions. Analysis of data from these two experiments will uncover the complex interplay of these transcription factors and the SAGA complex.

6. We would like to further investigate the molecular mechanism of bidirectional promoter regulation by these transcription factors. Above experiment will provide the list of bidirectional gene pairs which exhibit dysregulation after depletion of SP1, YY1 and NRF1. To investigate the possible crosstalk between transcription factors and histone modifications, ChIP analysis will be performed on the bidirectional gene pairs after depletion of SP1, YY1 and NRF1. If there is any crosstalk between these factors and histone

modifications on bidirectional promoters, then observed bimodal pattern of histone modifications would get affected.

7. Above experiment will provide more insight about the role of these transcription factors in bidirectional promoter regulation. We would like to perform the rescue experiment where we can rescue the dysregulated gene expression by overexpressing these transcription factors.

3.4 References

Baker, S., and Grant, P. (2007). The SAGA continues: expanding the cellular role of a transcriptional co-activator complex. Oncogene *26*, 5329-5340.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Bonnet, J., Wang, C.-Y., Baptista, T., Vincent, S.D., Hsiao, W.-C., Stierle, M., Kao, C.-F., Tora, L., and Devys, D. (2014). The SAGA coactivator complex acts on the whole transcribed genome and is required for RNA polymerase II transcription. Genes & development *28*, 1999-2012.
Bornelöv, S., Komorowski, J., and Wadelius, C. (2015). Different distribution of histone modifications in genes with unidirectional and bidirectional transcription and a role of CTCF and cohesin in directing transcription. BMC genomics *16*, 300.
Chambers, C.B., Halford, W.P., Geltz, J., Villamizar, O., Gross, J., Embalabala, A., Gershburg, E., and Wilber, A. (2015). A System for Creating Stable Cell Lines that Express a Gene of Interest from a Bidirectional and Regulatable Herpes Simplex Virus Type 1 Promoter. PloS one *10*, e0122253.
Chen, S., Nagy, P.L., and Zalkin, H. (1997). Role of NRF-1 in bidirectional transcription of the human GPAT-AIRC purine biosynthesis locus. Nucleic acids research *25*, 1809-1816.
Chinenov, Y., Coombs, C., and Martin, M.E. (2000). Isolation of a bi-directional promoter directing expression of the mouse GABPα and ATP synthase coupling factor 6 genes. Gene *261*, 311-320.
Cole, E.G., and Gaston, K. (1997). A functional YY1 binding site is necessary and sufficient to activate Surf-1 promoter activity in response to serum growth factors. Nucleic acids research *25*, 3705-3711.
Collins, P.J., Kobayashi, Y., and Nguyen, L. (2007). The ets-related transcription factor GABP directs bidirectional transcription. PLoS genetics *3*, e208.
Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845-1848.
Davila Lopez, M., Martinez Guerra, J.J., and Samuelsson, T. (2010). Analysis of gene order conservation in eukaryotes identifies transcriptionally and functionally linked genes. PloS one *5*, e10654.
Duttke, S.H., Lacadie, S.A., Ibrahim, M.M., Glass, C.K., Corcoran, D.L., Benner, C., Heinz, S., Kadonaga, J.T., and Ohler, U. (2015). Human promoters are intrinsically directional. Molecular cell *57*, 674-684.
Gaston, K., and Fried, M. (1994). YY1 is involved in the regulation of the bi-directional promoter of the Surf-1 and Surf-2 genes. FEBS letters *347*, 289-294.
Hahn, S., and Young, E.T. (2011). Transcriptional regulation in Saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. Genetics *189*, 705-736.
Häkkinen, A., Healy, S., Jacobs, H.T., and Ribeiro, A.S. (2011). Genome wide study of NF-Y type CCAAT boxes in unidirectional and bidirectional promoters in human and mouse. Journal of theoretical biology *281*, 74-83.
Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693-705.
Koyanagi, K.O., Hagiwara, M., Itoh, T., Gojobori, T., and Imanishi, T. (2005). Comparative genomics of bidirectional gene pairs and its implications for the evolution of a transcriptional regulation system. Gene *353*, 169-176.
Krebs, A.R., Karmodiya, K., Lindahl-Allen, M., Struhl, K., and Tora, L. (2011). SAGA and ATAC histone acetyl transferase complexes regulate distinct sets of genes and ATAC defines a class of p300-independent enhancers. Molecular cell *44*, 410-423.
Li, B., Carey, M., and Workman, J.L. (2007). The role of chromatin during transcription. Cell *128*, 707-719.
Lin, J.M., Collins, P.J., Trinklein, N.D., Fu, Y., Xi, H., Myers, R.M., and Weng, Z. (2007a). Transcription factor binding and modified histones in human bidirectional promoters. Genome research *17*, 818-827.
Lin, J.M., Collins, P.J., Trinklein, N.D., Fu, Y., Xi, H., Myers, R.M., and Weng, Z. (2007b). Transcription factor binding and modified histones in human bidirectional promoters. Genome research *17*, 818-827.
Näär, A.M., Lemon, B.D., and Tjian, R. (2001). Transcriptional coactivator complexes. Annual review of biochemistry *70*, 475-501.
Piontkivska, H., Yang, M.Q., Larkin, D.M., Lewin, H.A., Reecy, J., and Elnitski, L. (2009a). Cross-species mapping of bidirectional promoters enables prediction of unannotated 5' UTRs and identification of species-specific transcripts. BMC genomics *10*, 189.

Piontkivska, H., Yang, M.Q., Larkin, D.M., Lewin, H.A., Reecy, J., and Elnitski, L. (2009b). Cross-species mapping of bidirectional promoters enables prediction of unannotated 5'UTRs and identification of species-specific transcripts. BMC genomics *10*, 189.
Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S., Mapendano, C.K., Schierup, M.H., and Jensen, T.H. (2008). RNA exosome depletion reveals transcription upstream of active human promoters. Science *322*, 1851-1854.
Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., and Sharp, P.A. (2008). Divergent transcription from active promoters. Science *322*, 1849-1851.
Strahl, B.D., and Allis, C.D. (2000). The language of covalent histone modifications. Nature *403*, 41-45.
Trinklein, N.D., Aldred, S.F., Hartman, S.J., Schroeder, D.I., Otillar, R.P., and Myers, R.M. (2004). An abundance of bidirectional promoters in the human genome. Genome research *14*, 62-66.
Vernon, E.G., and Gaston, K. (2000). Myc and YY1 mediate activation of the Surf-1 promoter in response to serum growth factors. Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression *1492*, 172-179.
Wakano, C., Byun, J.S., Di, L.J., and Gardner, K. (2012). The dual lives of bidirectional promoters. Biochimica et biophysica acta *1819*, 688-693.
Weake, V.M., and Workman, J.L. (2008). Histone ubiquitination: triggering gene activity. Molecular cell *29*, 653-663.
Zanotto, E., Hakkinen, A., Teku, G., Shen, B., Ribeiro, A.S., and Jacobs, H.T. (2009). NF-Y influences directionality of transcription from the bidirectional Mrps12/Sarsm promoter in both mouse and human cells. Biochimica et biophysica acta *1789*, 432-442.

Appendix

A. Details of primer used for qPCR

Primer list for qPCR (Gene name and orientation)	Sequence (5’ – 3’)
TP 53 F	GCCATCTACAAGCAGTCACAG
TP 53 R	TCATCCAAATACTCCACACGC
WRAP53 F	CTTGCGAATTTATAACCTGCCC
WRAP53 R	GAGGACATCAGAGAATACCAGC
CBX3 F	GGCCTCCAACAAAACTACATTG
CBX3 R	TCCACTTTCCCATTCACTACAC
HNRNPA2B1 F	ACCAGCAACCTTCTAACTACG

HNRNPA2B1 R	CACCATAACCCCCACTTCC
BRCA1 F	AGGAGATGTGGTCAATGGAAG
BRCA1 R	GTTGATCTGTGGGCATGTTG
NBR2 F	GCCTGTTTTCTCATCCATACAATG
NBR2 R	CTCCTTTCCACATTCCAAATTCC
CBX5 F	CCAATTTCTCAAACAGTGCCG
CBX5 R	GTTGCCCCAATGATCTTTTCTG
HNRNPA1 F	ACCTATTGTCCAAAGCAGTCG
HNRNPA1 R	CTCTACTCTTCATCTTCCTCATCC
RAD 51 F	GTGGTAGCTCAAGTGGATGG
RAD 51 R	GGGAGAGTCGTAGATTTTGCAG
RAD 51 anisense F	CGAGTTTACAGACTGCCCTC
RAD 51 antisense R	CTAATAGTCCAGCTGCGATGG
FoXM1 Antisense F	ACCACAGCACCATCACTTC
FoXM1 Antisense F	AGATGTGGAAACTTGGAGGTG
SERPINI1 F	TCAGTCAAAATGTAGCCGTGG
SERPINI1 R	AGTTCCCCTTGAAATAGACAGC
PDCD10 F	TGAGCTAGAACGAGTAAATCTGTC
PDCD10 R	AAGGGACTCCGTGAAGTTAAC
CIB1 F	CTTCTCCACATCCCCAGC
CIB1 R	TCTGTTCAAGGTTCCGTCATC
GDPGP1 F	GCTATTTACTGAGGACCCACTG
GDPGP1 R	AGTTTGGGTTCCTTCAGCTC
JMJD6 F	TTGGAAGACTACAAGGTGCC
JMJD6 R	GGTCGATGTGAATCCCAGTTC

METTL23 F	CCTTGCATCTGATGTGTTCTTTG
METTL23 R	CCAGTCAGCACTCCTAACTTG
MCM7 F	GATGCCACCTATACTTCTGCC
MCM7 R	TCCTTTGACATCTCCATTAGCC
AP4M1 F	CTTCTGGGCGATTACTGTGG
AP4M1 R	AATTCCTCAGCATCTCCGTG
h NRF1 F	TGCAGGTCCTGTGGGAATG
h NRF1 R	TGAGGCCGTTTCCGTTTC
h GABP F	TGCACTGGAAGGCTATAGGAAAG
h GABP R	GGACCACTGTATGGGATCATAGG
h YY1 F	TGGAGAGAACTCACCTCCTGA
h YY1 R	TCTTTAATTTTTCTTGGCTTCATTC
Hu CWC27 F	CATTTGGAGAGGAAGCTGAGG
Hu CWC27 R	TGGAACAGAACTGAGATGTGG
Hu SREK1IP1 F	AACAAGGACAGTGTCAGAGC
Hu SREK1IP1 R	TCTACTCGGAGAAAATTGCGG
CENPK F	GTTCCAAAAGCTGAGACAAGATC
CENPK R	CCATTATCTGTTGCTGTTCATCC
PPWD1 F	AGTTGGAGAAGGTTGATGCTG
PPWD1 R	TTCTTGTTTGCCTAAAATCCGC
HEL308 F	CATTTATCAAGACGCCAAGCC
HEL308 R	AGCACCAGGGAAATCAGAAG
MRPS18c F	CGCTGTGGTTGCTGTTTG
MRPS18c R	GTCCTCATTGCTGGATACCTG
LRCH4 F	CAACGAGCTCCAATCCCTG

LRCH4 R	ACGCGGTTACAGGAGAAATC
FBXO24 F	TGACCTTCAAGCAGATCGTG
FBXO24 R	GAACCGTGTAGGAGCGTG
ACAD11 F	AGGATATGGTATAGGTGCTGGG
ACAD11 R	TCTCTTCATTGTCATTATCGGGC
UAB 5 F	GTGTGGACAATTTTGAAGCTCG
UAB 5 R	CGCAAAACAAGCAGATTCTCC
TRIAP1 F	CTGGTTCGCCGAGAAATTTC
TRIAP1 R	CCCATGAACTCCAGTCCTTC
GATC F	CTGAGATCCGACAATGTGGTAG
GATC R	GCTCTTGTTCATCCAGCTTTG
NAIF1 F	ACCACAGAGATCCACCCT
NAIF1 R	CTTGACCGACGTGTCTGCATG
SLC1 F	CAACGTCCTCAAAATTGCCC
SLC1 R	CAAGCCTCTCGTGAATCCTC
PFKM F	AAGTCTTTACCTCGAAACCCG
PFKM R	ACGACATGAACCACTCCAAG
PFKM F	TGACCAAAGATGTGACCAAGG
PFKM R	GCGAACCACTCTTAGATACCG
Hu TRIM23 F	TGTCAAACTAGCCCACTCATG
Hu TRIM23 R	GATTTCCTCTGTGAAGGTCCG
C5orf44 F	GACCAGCCAACTTCAAAGAATG
C5orf44 R	AACCAGATCCATAGTCCTTTCAC
MEPCE F	ACACATCAGTCTTCCCCAAC
MEPCE R	TCGTCTCCCCAGTTCAGAT

ZCWPW1 F	TCAACGGATCTAACAGTAATGGG
ZCWPW1 R	CGAATGGGCAAAATTGGGTC
h SP1 F	GCCCCAGGTGATCATGGA
h SP1 R	CTGGGCTGTTTTCTCCTTCCT

B. Details of primers used for cloning.

Cloned Region	**RE Sites**	**Sequence (5' – 3')**
TP53-WRAP53 INT region sense F	Age 1	GCGACCGGTCACTGTGTTCCTTAG CACCG
TP53-WRAP53 INT region sense R	Nhe 1	GCGGCTAGCCGGTGGCTCTAGAC TTTTGAG
TP53-WRAP53 INT region Antisense F	Nhe1	GCGGCTAGCCACTGTGTTCCTTAG CACCG
TP53-WRAP53 INT region Antisense R	Age1	GCGACCGGTCGGTGGCTCTAGAC TTTTGAG
FoXM1- RHNO1 int region sense F	Age 1	GCGACCGGTTGAAAAGGGGAGCA GAGGAGC
FoXM1- RHNO1 int region sense R	Nhe 1	GCGGCTAGCGAGCTTTCAGTTTGT

		TCCGCTGTTTG
FoXM1- RHNO1 int region sense F	Nhe1	GCGGCTAGCTGAAAAGGGGAGCAGAGGAGC
FoXM1- RHNO1 int region sense R	Age1	GCGACCGGTGAGCTTTCAGTTTGTTCCGCTGTTTG
CBX 3 Int region sense F	Age 1	GCGACCGGTCCGCTTTTCTAGAACCTTCC
CBX 3 Int region sense R	Nhe 1	GCGGCTAGCCCTACAGCTCAAGCCACATCC
CBX 3 Int region Anti-sense F	Nhe1	GCGGCTAGCCCGCTTTTCTAGAACCTTCC
CBX 3 Int region Anti-sense R	Age1	GCGACCGGTCCTACAGCTCAAGCCACATCC
CBX 5 Int region sense F	Age 1	GCGACCGGTGAGCACGTGACCTCAAATGAT
CBX 5 Int region sense R	Nhe 1	GCGGCTAGCATCGCCCCAGTTCTTTCTTTC
CBX 5 Int region Anti-sense F	Nhe1	GCGGCTAGCGAGCACGTGACCTCAAATGAT
CBX 5 Int region Anti-sense R	Age 1	GCGACCGGTATCGCCCCAGTTCTTTCTTTC
NFYA-ORAD sense F	Age 1	GCGACCGGTCTTAATTGCACGCATCTAAGATG
NFYA OARD- sense R	BamH1	GCGGGATCCCGGACTCCGAAACCCAATC
NFYA-OARD Antisense F	BamH1	GCGGGATCCCTTAATTGCACGCATCTAAGATG
NFYA-OARD antisense R	Age 1	GCGACCGGTCGGACTCCGAAACCCAATC
BRCA1-NBR2 Int sense F	Age1	GCGACCGGTGATTGGGACCTCTTC

		TTACGACT
BRCA1-NBR2 Int sense R	Nhe1	GCGGCTAGCCGCAGTTTTAATTTATCTGTAATTCC
BRCA1-NBR2Int Antisense F	Age1	GCGGCTAGCTGAAAAGGGGAGCAGAGGAGC
BRCA1-NBR2Int Antisense R	Nhe1	GCGACCGGTCGCAGTTTTAATTTATCTGTAATTCC
EIF3I-LOC102723465 int sense F	Age1	GCGACCGGTCTTCCCTCCTTCGCTCTCTTC
EIF3I-LOC102723465 int sense R	Nhe1	GCGGCTAGCGGCCGCAACGTGAGTAAGAC
EIF3-ILOC102723465 int Antisense F	Nhe1	GCGGCTAGCCTTCCCTCCTTCGCTCTCTTC
EIF3I-LOC102723465 int Antisense R	Age1	GCGACCGGTGGCCGCAACGTGAGTAAGAC
TMEM2080-LRRC29 int Sense F	Age1	GCGACCGGTCTGTTCTCGCGGAGATGACAG
TMEM208-LRRC29 int Sense R	BamH1	GCGGGATCCTGTGCGAGGCAAATACCCAG
TMEM208-LRRC29 int Antisense F	BamH1	GCGGGATCCCTGTTCTCGCGGAGATGACAG
TMEM208-LRRC29 int Antisense R	Age1	GCGACCGGTTGTGCGAGGCAAATACCCAG
CMV promotor Nhe1 F	Nhe1	GAGAGCTAGCGCCGTATTACCGCCATGC
CMV promotor Age1 R	Age1	TATAACCGGTGGTGGCGACCGGTAG
CMV promotor Age1 F	Age1	GAGAACCGGTGCCGTATTACCGCCATGC
CMV promotor Nhe1 R	Nhe1	GCATGCTAGCGGTGGCGACCGGT

		A
GFP FL Cloning in antisense Nhe1 F	Nhe1	GAAGCTAGCGATGGTGAGCAAGG G
GFP FL cloning in antisense Ase1 R	Ase1	GCGCATTAATTTATCTAGATCCGGT GG
hSP1 BamH1 F	BamH1	CGGAATTCGATGGATGAAATGACA GCTGTGGTG
hSP1 Kpn1 R	Kpn1	GGTACCTCAGAAGCCATTGCCACT GATATT
NFYa FL EcorI F	EcorI	GCGGAATTCGATGGAGCAGTATAC AGCAAACAG
NFYa FL XbaI F	XbaI	CGCTCTAGATTAGGACACTCGGAT GATCTGT

C. Details of primers used for ChIP-PCR

ChIP Primer description	**Sequence (5' – 3')**
hNYFA ChIP F1	CAAGGGTATCAATCTTCTAAGAGTG
hNYFA ChIP R1	GGCTCGTCATTTCCTCTCTTC
hNYFA ChIP F2	GAAGAGAGGAAATGACGAGCC
hNYFA ChIP R2	GAGCCGGGTCTGTCTTATCTG
hNYFA ChIP F3	CAGATAAGACAGACCCGGCTC
hNYFA ChIP R3	GACTCAGGCCCAGGTTCTCAG

hNYFA ChIP F4	CTGAGAACCTGGGCCTGAGTC
hNFYA ChIP R4	GTGAGTGTGAGGAGCCAATATCC
hNFYA ChIP F5	GGATATTGGCTCCTCACACTCAC
hNYFA ChIP R5	CGATCCCTGAACTGGAGTTAGTG
hNYFA ChIP F6	CACTAACTCCAGTTCAGGGATCG
hNYFA ChIP R6	GTCCTAGTGGCCACTTGGAAAG
hNYFA ChIP F7	CTTTCCAAGTGGCCACTAGGAC
hNYFA ChIP R8	CTTAATTGCACGCATCTAAGATGG
hNYFA ChIP F9	CCATCTTAGATGCGTGCAATTAAG
hNYFA ChIP R9	GCCTCCCATTCTCTGTCTCCTAC
hNYFA ChIP F10	GTAGGAGACAGAGAATGGGAGGC
hNYFA ChIP R10	CGATTTAGGACGGTCTCCTTTTC
hNYFA ChIP F11	GAAAAGGAGACCGTCCTAAATCG
hNYFA ChIP R11	CTGATTGCCCTGTACAACCAC
hNYFA ChIP F12	GTGGTTGTACAGGGCAATCAG
hNYFA ChIP R12	CGAATTTCAGTTCACCAATAAGTC
hNYFA ChIP F13	GACTTATTGGTGAACTGAAATTCG
hNYFA ChIP R13	CTCTCTTCCCAAACCAATTACAG
Mcherry F1	ATGGTGAGCAAGGGCGAG
Mcherry R1	ACCCTTGGTCACCTTCAGCT
Mcherry F2	AGCTGAAGGTGACCAAGGGT
Mcherry R2	GTCCTCGAAGTTCATCACGC
Mcherry F3	GCGTGATGAACTTCGAGGAC
Mcherry R3	CATGGTCTTCTTCTGCATTACG

Mcherry F4	CGTAATGCAGAAGAAGACCATG
Mcherry R4	AGCTGCACGGGCTTCTTG
Mcherry F5	CAAGAAGCCCGTGCAGCT
Mcherry R5	TACTTGTACAGCTCGTCCATGC
Mcherry Reverse Common	GAGCGGGAACGAGTGGTA
CMV primer for mCHeery	AGTCCCATAAGGTCATGTACTGG
mCherry N1 first reverse	CTCCTTGATGATGGCCATGTTAT
CMV primer for mCHeery near to cherry	GGAAATCCCCGTGAGTCAAAC
CMV primer for mCHeery near to GFP	GACGTATGTTCCCATAGTAACG
EGFP F1	ATGGTGAGCAAGGGCGAG
EGFP R1	CAGGGTCAGCTTGCCGTAG
EGFP F2	CTACGGCAAGCTGACCCTG
EGFP R2	GTCGTCCTTGAAGAAGATGGTG
EGFP F3	CACCATCTTCTTCAAGGACGAC
EGFP R3	GCTTGTCGGCCATGATATAGA
EGFP F4	TCTATATCATGGCCGACAAGC
EGFP R4	GTACAGCTCGTCCATGCCG
TP53 F for EGFP ChIP SENSE	TGCTTTAAGAATTACCGCG
SDCCAG10-SFRS12IP1 Intergenic	GTGTAAACACCGCCAACACTTA
SDCCAG10-SFRS12IP1 Intergenic	AATGTACTCCAGGGAACGAGAA
CENPK-PPWD1 intergenic	GAGAAGAACGGGACTCAAAGC
CENPK-PPWD1 intergenic	GGGAGGTTGATATAACGGTTGA
HEL308-MRPS118c intergenic F	CGCATAAACTTCTACCCTGTCC
HEL308-MRPS118c intergenic R	GCGAGAGTAGGTGAGGTTTTTC

LRCH4-FBX024 intergenic F	AAAGACCAATCGTAAGCCAGAT
LRCH4-FBX024 intergenic R	GCCCATGCTATTGGTAGGTAGA
ACAD11-UBA5 Intergenic F	AGGGTCCCGTTCAGAAAAAG
ACAD11-UBA5 Intergenic R	GATAAGTGTCCAGGGCAGGAG
TRIAP1-GATC intergenic F	CAAGGAAGGAAGAAATGTGGTC
TRIAP1-GATC intergenic R	CCTGGCCTTTACCTGAGGAT
NAIF1-SLC1 intergenic F	TGGTAGCTGCTCAAATATCACG
NAIF1-SLC1 intergenic R	CGGACACAATTTTAGACCAATG
SENP1-PFKM Intergenic F	GCAGGAGGTCAAAGAACAGAGT
SENP1-PFKM Intergenic F	CCCAGACTGAAAAGGGTACTGA

D. Primers used for guide RNA clones and screening of Knockout cells

SP1 guide1(Exon3) sense	CACCGTCAAGGCCAGACACCCCAGA
SP1 guide1(Exon3) antisense	AAACTCTGGGGTGTCTGGCCTTGAC
SP1 guide2(Exon6) sense	CACCGCAGTGGCATCAACGTCATGC
SP1 guide2(Exon6) antisense	AAACGCATGACGTTGATGCCACTGC
NRF1 guide1(Exon1) sense	CACCGAGTGACCCAAACCGAACATA
NRF1 guide1(Exon1) antisense	AAACTATGTTCGGTTTGGGTCACTC
NRF1 guide2(Exon8) sense	CACCGTCTTACCTCTCCATCAGCCA

NRF1 guide2(Exon8) antisense	AAACTGGCTGATGGAGAGGTAAGAC
SP1 Ko check RT F1	CAGCTAGTTCAAGGGGGACA
SP1 KO check RT R1	CAATGGGTGTGAGAGTGGTG
NRF1 Ko check RT F1	GGGCATTTATCCCAGAGATG
NRF1 KO check RT R1	CTGCTTTTGCTCTTCTGTGC
SP1 F1 for screening KO	ACCATCAGTTCTGCCAGCTT
SP1 R1 screening KO	AGATCTGCCACCTGCATGAC
SP1 R2 for screening KO	CTGGTTTTGCTGGATGTTCA
SP1 F2 for screening KO	CTGCCCTGAGTGTCCTAAGC
NRF1 F1 for screening KO	TCAAGGTTCCTCTGCTCTTGA
NRF1 R1 for screening KO	AATGAAAGAGTGCCGCAGAC
NRF1 F2 for screening KO	TCTGTGCTGAATTTGGGATT
NRF1 R2 for screening KO	TCAAGGTTCCTCTGCTCTTGA
SP1 screening KO	CTGGTTTTGCTGGATGTTCA