Epigenetic basis of transcriptional ripple in mammalian genome

A dissertation submitted for partial fulfillment of BS-MS dual degree in Science

Sayali Rangarao Chougale

MS10105

Indian Institute of Science Education and Research Mohali

Knowledge City, Sector 81, SAS Nagar, Manauli, PO 140306

April 2015

Certificate of Examination

This is to certify that the dissertation titled "Epigenetic basis of transcriptional ripple in mammalian genome" submitted by Sayali Rangarao Chougale (MS10105) for the partial fulfillment of BS-MS dual degree program of IISER Mohali has been examined by the thesis committee duly appointed by the institute. The committee finds the work done by the candidate satisfactory and recommends that the report be accepted.

Dr. Rajesh Ramachandran Dr. Shashi Bhushan Pandit Dr. Kuljeet Singh Sandhu
(Supervisor)

Dated: April 24th 2015

Declaration

The work presented in this dissertation has been carried out by me under the guidance of Dr. Kuljeet Singh Sandhu at the Indian Institute of Science Education and Research Mohali.

This work has not been submitted in part or in full for a degree, a diploma, or a fellowship to any other university or institute. Whenever contributions of others are involved, every effort is made to indicate this clearly, with due acknowledgement of collaborative research and discussions. This thesis is a bonafide record of original work done by me and all sources listed within have been detailed in the bibliography.

Sayali Rangarao Chougale

Dated: April 24th 2015

In my capacity as the supervisor of the candidate's project work, I certify that the above statements by the candidate are true to the best of my knowledge.

Dr. Kuljeet Singh Sandhu

(Supervisor)

Acknowledgment

I offer my warmest gratitude to my supervisor Dr. Kuljeet Singh Sandhu , who gave me the opportunity to do MS thesis in his research group. I would like to thank him for his guidance, support, valuable comments and encouragement from the start until the end of my project.

I would like to give respect to my thesis committee consisting of Dr. Kuljeet Singh Sandhu (Supervisor), Dr. Rajesh Ramachandran and Dr. Shashi Bhushan Pandit.

I am thankful to the Department of Biological Sciences, IISER Mohali for providing me this opportunity and Library, IISER Mohali for providing the facilities that allow us to do healthy research.

Special thanks to Arashdeep and Meenakshi for their constant support, help and advice throughout my project. I would never been able to finish my MS thesis without their guidance. I would like to thank Ananya, Aniket, Ashutosh, Preeti, Priya and Rivi for building up a positive environment and making my stay in a lab a wonderful experience.

I would also like to thank my friend Sukriti. She was always there cheering me up through the tough times. I would also like to thank Akshay, Sharmi, Mahima, Sukanya and Rahul who brought great joy to my life.

I would like to express my eternal gratitude to my parents and my brother for their everlasting love, encouragement and support at every step of my life.

Sayali Rangarao Chougale

Contents

List of Abbreviations

3C	**Chromosome Conformation Capture**
4C	**Circularized Chromosome Conformation Capture**
MCF-7	**Michigan Cancer Foundation-7**
IEG	**Immediate-early genes**
ER	**Estrogen Receptor**
ChIP-seq	**Chromosome Immunoprecipitation Sequencing**
SRA	**Sequence Read Archive**
ChIA-PET	**Chromatin Interaction Analysis by Paired-End Tag Sequencing**
Hg19	**Human Genome 19**
Hg18	**Human Genome 18**
TSS	**Transcription Start Site**
C	**Cytosine**
G	**Guanine**
CpG	**Cytosine-phosphate-Guanine**
CGI	**Cytosine Guanine Island**
LCP	**Low CpG Promoter**
HCP	**High CpG Promoter**
ICP	**Intermediate CpG Promoter**
NCBI	**National Center for Biotechnology Information**
H3K36me3	**Histone H3 trimethyl Lys36**
H3K4me2	**Histone H3 dimethyl Lys4**
H3K4me3	**Histone H3 trimethyl Lys4**
H3K9me3	**Histone H3 trimethyl Lys9**
H3K9ac	**Histone H3 acyl Lys36**
H3K27me3	**Histone H3 trimethyl Lys27**
CTCF	**CCCTC binding Factor**

List of Figures

Abstract

The phenomenon, where transcription of an induced gene propagates to its neighbouring genes is known as transcriptional ripple. A study by Ebisuya et al. demonstrated that addition of fibroblast growth factor in quiescent mouse NIH3T3 cells; rapidly induce immediate-early genes (IEGs) and this induction of target genes is accompanied by up -regulation of their neighboring genes. This suggests that transcriptional activation has a ripple-like effect. However, epigenetic basis and the functional significance of transcriptional ripple effect are not yet understood. Here, we have investigated the transcriptional ripple effect in Estrogen induced transcription in MCF-7 breast cancer cell line. We identified 151 estrogen-receptor(ER) target and their neighbouring genes; out of which 89 neighbouring genes showed significant transcriptional up/down regulation upon estrogen induction. Further, we comprehensively analyzed different genomic, epigenomic and functional attributes of estrogen-receptor target and their nearest neighboring genes. We had following key observations: 1) Affected neighbouring genes were more distant from the target genes as compared to unaffected genes, significance of which is not entirely clear. 2) Regions having affected and unaffected neighbouring gene did not differ in their gene density. 3) Chromatin state analyses through ChromHMM platform, suggested that the affected neighboring genes more likely to have 'active promoter' as compared to unaffected genes. 4) Most affected neighboring genes had 'High CpG' promoters as compared to ER-target genes and the unaffected neighboring genes. 5) Analyses of RNAPII ChIA-PET data revealed no significant association with the long range chromatin interaction between target and neighboring genes.

Overall, our observations suggest that the induction of a gene can interfere with the process of active transcription in the neighborhood and that the chromatin state and the CpG content of the neighboring gene-promoter seemingly determines whether or not the gene would be perturbed. Further exploration of other genome-wide datasets would, hopefully, unravel the underlying functional significance of such observations.

Keywords Transcriptional ripple effect, chromatin states, Normalized CpG content, chromatin interaction.

Chapter 1

Introduction, Objective, Materials and methods

1 Introduction

1.1 Transcription and its regulation

Transcription is the process by which a template DNA strand is transcribed into RNA. This process is regulated at multiple levels by the different mechanisms ranging from differential binding of transcription factors, Enhancer promoter interactions and promoter-promoter interactions. In particular, regulation of transcription in eukaryotes is a result of the combined effects of structural properties such as packaging of DNA and the interaction with proteins called transcription factors and chromatin remodeling complexes.

This control allows the cell or organism to respond a variety of intra- and extra cellular environment. Any perturbation in the normal physiology of the cell starts a cascade of signals which can ultimately modify the gene expression. One example of this kind of transcriptional induction or repression is Estrogen signaling [1]. This signaling results in the activation and translocation of estrogen receptor inside the nucleus, where it binds its target genes and modulates their expression profiles. It has been thought that such kind of transcriptional regulation affects only target genes of effector molecule (Transcription Factor). But recently, it has been demonstrated that this kind of transcriptional induction spills beyond the boundaries of target genes [2], which may lead to activation/suppression of neighboring genes, the phenomena which has been termed Transcription Ripple.

1.2 The Transcription Ripple Effect

As described above, Induced transcription at one locus frequently spills over into its physical neighboring loci (2). This phenomenon is described in figure 1. Considering the phenomena of transcriptional ripples and given the fact that genome is pervasively transcribed, where most of the DNA gives rise to RNA, it is possible that co-regulation of neighbouring genes is achieved by ripple

effect, where the transcriptional activation at one loci reaches the other nearby genes via intergenic regions.

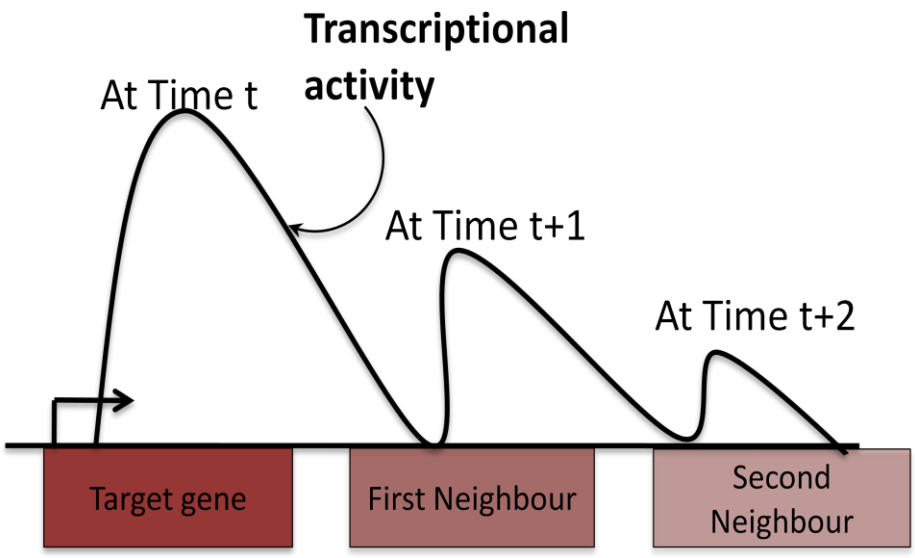

Fig1: Transcription Ripple- Transcription propagates outside the boundaries of the initial target gene. Transcription of a target gene spills over into its neighboring genes over the time.

In bacteria also, it is well known that functionally similar genes are present in close neighborhood and regulated as a single unit known as operon systems, so in Eukaryotes also, it is possible the neighbouring genes are co-regulated by the phenomena of transcription ripples.

However not much is known about the molecular basis or functional significance of transcription ripple effect. It is possible that genes which are associated with ripple effect have unique epigenetic signatures. Further, it is also possible that ripple effect is not limited to linear neighbors, but is also exhibited by the physically close neighbors in the 3-D space of the nucleus.

1.3 **Epigenetics and Chromatin modification**

The DNA of all eukaryotic cells is tightly bound to histones, forming chromatin. The basic structural unit of chromatin is the nucleosome, which consists of 146 base pairs of DNA wrapped around two molecules each of histones H2A, H2B, H3 and H4, with one molecule of histone H1 bound to the DNA as it enters the nucleosome core

particle. The chromatin is then further condensed by being coiled into higher-order structures organized into loops of DNA. This packaging of eukaryotic DNA in chromatin clearly has important consequences in terms of its availability as a template for transcription, so chromatin structure is a critical aspect of gene expression in eukaryotic cells.

This condensed chromatin structure is dynamical nature, with various covalent modifications (acetylation, methylation) being involved to change the extent of condensation of chromatin fiber.

The combination of different kinds of histone modifications that are present at various loci define the epigenetic state of a given chromatin loci. Epigenetics is heritable changes in gene activity and expression that occur without alteration in DNA sequence. These non-genetic alterations are tightly regulated by two major epigenetic modifications:

- Chemical modifications to the cytosine residues of DNA (DNA methylation)
- Histone modifications

The patterns of epigenetic modifications can serve as epigenetic markers to represent gene expression and chromatin state.

1.4 **Long range chromatin interactions**

Nucleus of higher eukaryotes contains large molecules of DNA and those molecules are not linear. The packaging of DNA of huge length inside the nucleus brings various distant elements in close proximity to each other. This give rise to chromatin interactions, which has been proved to be crucial in affecting various aspects of gene expression, DNA replication and DNA repair. With the advent of 3C based (3C, 4C, Hi-C) technologies, it has become possible to measure such chromatin interactions at genome wide scale. This enables us to check the ripple effect in context of long range interactions.

2. Objective:

Our investigation to understand the transcription ripple effect is using Estrogen regulated MCF-7 cell line.

Firstly, we decided to see if estrogen regulated genes are showing transcriptional ripple effect (domain wide transcriptional response) using time-course microarray data of the same cell line. As the rapid induction of immediate-early genes (IEGs) in response to growth factor stimulation is accompanied by co-upregulation of their neighboring genes [10] thus we examined the expression status of estrogen regulated genes and their neighboring genes after estrogen induction in mammalian cells.

Secondly, we decided to see if chromatin states and long-range interactions play an important role in the domain wide transcriptional response. Further, we comprehensively analyzed different genomic, epigenomic and functional attributes of estrogen-receptor target and their nearest neighboring genes.

2.1 Our System: Estrogen regulated MCF-7 cell line

Estrogen hormones are key regulators of diverse cellular processes in a wide array of target tissues. Effect of estrogen is mediated through its interaction with its nuclear receptors, estrogen receptors. Estrogen binds with its receptor and the dimerization of the estrogen receptor results into the interaction with chromatin. This interaction led to the regulation of transcriptional activity of estrogen responsive genes. Figure 2 is a cartoon representation of the system we have used for our investigation.

MCF-7, breast cancer cell line undergoes a robust proliferative response to estrogens.

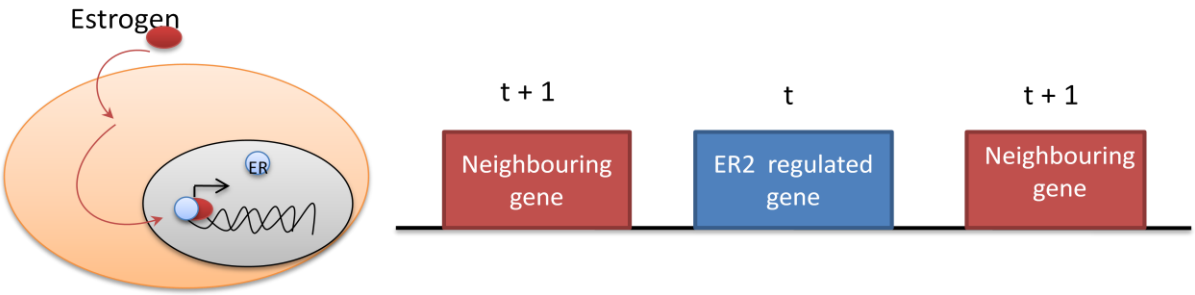

Fig 2: Our System- Estrogen regulated transcriptional activity. Estrogen enters the cell and dimerizes with estrogen receptor. This complex binds to the DNA and regulates the transcriptional activity of genes.

3. Materials and methods

Perl was used as a programming language for file handling. R and Microsoft Excel were used for the statistical analysis. Datasets that were used in this study were

obtained from publically available sources. Wherever necessary we have generated the data using different softwares for detailed analysis.

3.1 ER-α binding data:

Chromatin Immunoprecipitation sequencing (ChIP – seq) have identified the chromosomal locations of ER-α binding sites [13].

We have taken the mid-point of those genomic region and calculated 10 KB region around the point. Further, we have found out the genes present in that region. These genes are the ER-α bound genes or target genes.

Next, using the genomic coordinates we have found out the neighbors of all the target genes.

3.2 Genome Information:

Genomic coordinates in this study are based on UCSC Genome Browser [14] hg18. We have considered the TSS of all the genes.

3.3 Gene Expression Data:

We have used Time-course microarray data for estrogen treated human breast adenocarcinoma cells [15] (MCF-7 cell line). Gene expression data was already normalized.

3.4 Clustering of Genes:

Cluster 3.0[16] was used to cluster the genes. We have clustered the 212 estrogen bound by k-means clustering where k=10.

In Short, given a set of observations ($x_1, x_2, ..., x_n$), where each observation is a d-dimensional real vector. The k-means algorithm is used to cluster n objects based on attributes into k partitions where $k < n$.

Cluster 3.0 gave 2 output files CDT and kgg. The kgg file gave the 10 clustered genes, CDT file was used to visualize the clustered data in R. we have used heatmaps to visualize the clustered data.

$$J = \sum_{j=1}^{k} \sum_{n \in s_j} |x_n - \mu_j|^2$$

x_n = Vector representing nth data point
μ_j = Geometric mean of data points in a subset s_j

3.5 Detection of transcriptionly affected neighboring genes

We have used Wilcoxon test in R to get the transcriptionly affected neighbouring genes. Out of 6 times course columns (3hr, 6hr, 9hr, 12hr, 24hr, and 48hr), each time we have kept a column fixed and checked whether the value of the column is significantly different from the mean of all other columns. At least one column should have p value <= 0.05, then we have considered the gene as transcriptionly perturbed gene.

3.6 Histone modification Data

We have used H3K4me2, H3K9ac, H3K4me3, CTCF, H3K27ac, and H3K27me3, H3K9me3, and H3K36me3 histone modification data to get the chromatin modification state data of the MCF-7 cell line.

We did not get bed files for H3K9me3, H3K4me2, H3K9ac and H3K4me1 however sequence read archive (SRA) files were available in NCBI sequence read archive.

As aligned data was not present in the sra files, we have aligned the fastq files to the reference genome using Bowtie 2 and the bed files were used as input for the ChromHMM.

3.7 Generation of chromatin states

We have used ChromHMM [17] to integrate histone modification datasets into 8 different chromatin states. ChromHMM works based on a multivariate Hidden Markov Model that models the observed combination of chromatin marks using a product of independent Bernoulli random variables, which enables robust learning of complex patterns of many chromatin modifications. Chromatin states were analyzed at 200-base pair intervals.

TSS of target and neighboring genes were mapped to this data and chromatin state for each gene was determined.

3.8 Chromatin Interaction Data

Using genome-wide Chromatin Interaction Analysis with Paired-End-Tag sequencing (ChIA-PET), long-range chromatin interactions associated with RNA polymerase II in MCF-7 cell line were discovered [13].

We have used this spatial proximity data to get the interacting gene pairs of target and its neighbor.

3.9 Gene Density Data

The gene density of an organism's genome is the ratio of the number of genes per number of base pairs.

We were having gene density data for hg19 genome assembly; by using 'liftover' tool of UCSC we have converted it to hg18 genome assembly.

Target and neighboring gene TSS were mapped to this data and their gene density was obtained.

3.10 Promoter Type

The CpG sites are regions of DNA where a cytosine nucleotide occurs next to guanine nucleotide in the linear sequence of bases. CpG islands (CG islands or CGI) are regions with a high frequency of CpG sites. CGIs are generally associated with promoters. Genes with High CpG Promoter tend to be expressed in most tissues whereas genes with Low CpG Promoter tend to be tissue-specific. [18, 19]

Normalized CpG content can be calculated by the ratio of observed to expected CpG.

Observed CpG = (number of CpGs * length of sequence)

Expected CpG = $((\text{number of Cs} + \text{number of Gs})/2)^2$

We have taken 3kb region around the TSS of all target and their neighboring genes from UCSC's Galaxy genome browser and calculated Normalized CpG content for them. As a reference, we have taken Normalized CpG content for whole human

genome. The distribution of promoters' normalized CpG content is bimodal and shows three different regions 0 to 0.45 is Low CpG promoters (LCP), 0.45 to 0.75 is Intermediate CpG promoters (ICP) and 0.75 to 1 is High CpG promoters (HCP). Figure 3 shows the bimodal distribution of normalized CpG content of whole human genome.

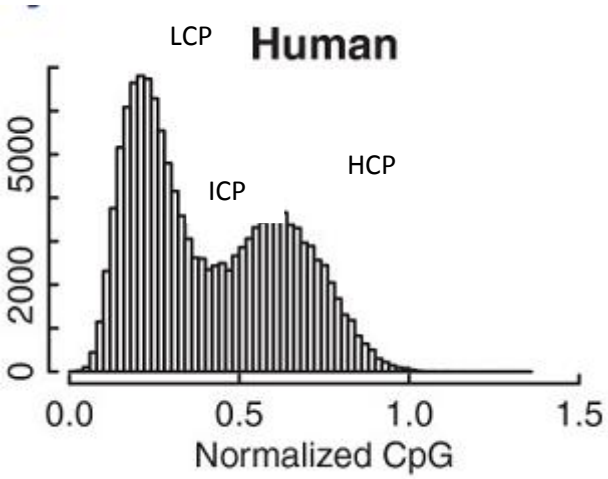

Fig 3: Bimodal Distribution of normalized CpG content at the promoters of whole human genome. It shows three different regions 0 to 0.45 is Low CpG promoters (LCP), 0.45 to 0.75 is Intermediate CpG promoters (ICP) and 0.75 to 1 is High CpG promoters (HCP).

We have compared the normalized CpG content of target and their neighboring genes. In addition, we have checked the same for transcriptionly affected and unaffected genes.

3.11 Functional similarity

As the co-regulated neighbouring genes show functional similarity in operon systems we wanted to check whether co-regulated neighbours of estrogen target genes show functional similarity with their target genes. We thought, Functional similarity help us to understand the biological role and function of the transcriptional ripple effect.

We have used GOSemSim[20], an R[21] package to get functional similarity of the genes. It uses Gene Ontology terms as the basis for measuring gene functional similarity.

3.12 Approach

- We have used estrogen binding data to get the target genes and using genomic coordinates we were able to find out neighboring genes.
- Once we got the chromosomal locations of ER-α binding sites, we have mapped those chromosomal locations to genes. We have taken 5kb genomic region on both the sides of ER-α binding site and we got 263 target genes.
- Using genomic coordinate information, neighbouring genes of those target genes were found out.
- We have mapped time-course microarray data of estrogen induced MCF-7 cell line to target genes and its neighboring gene (151). We were able to map 212 genes to their expression data. Further, we have used heatmaps to visualize the clustering these genes.
- Using Wilcoxon test, we have differentiated neighbouring genes into two types: transcriptionly affected (89) and transcriptionly unaffected (62).
- Then, we did genomic attributes analysis, chromatin states analysis, promoter type analysis, chromatin interaction analysis, and gene functionality analysis.

Chapter 2

Results, Discussion and Conclusion

To test the hypothesis, first we did expression analysis of target and neighboring genes. We have clustered target and neighboring genes into 10 clusters and visualized those results using heatmap. Figure 4 shows the heatmap of clustered target and neighbouring genes.

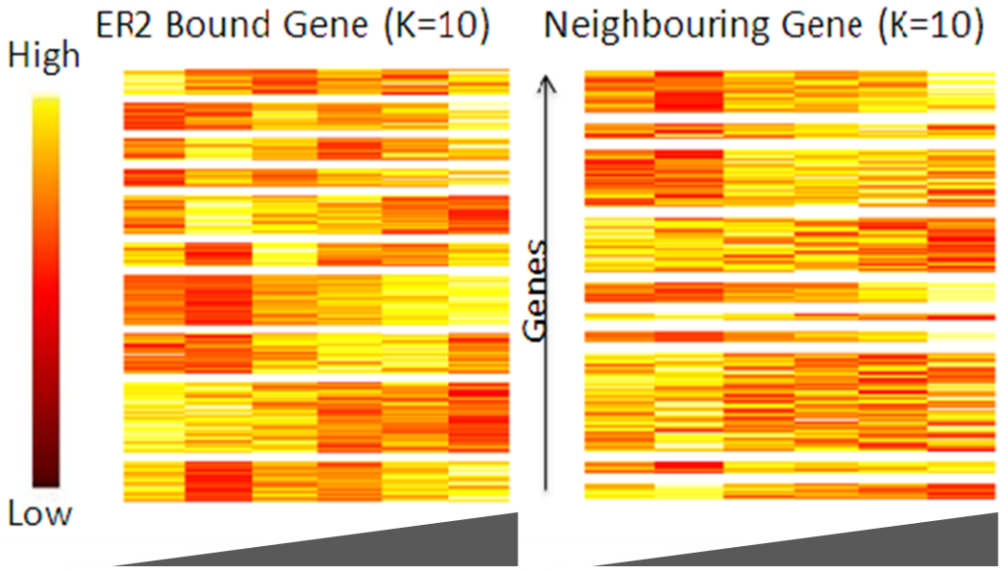

Fig 4: Heatmaps visualizing clustered target and neighboring genes. Cluster 3.0 was used to cluster target and neighbouring genes. To find out the transcriptionly affected neighboring genes Wilcocxon one-tailed test was used.

Out of 151 neighbouring genes, 89 genes showed significant transcriptional activity in which 58 were upregulated and 31 were downregulated.

4.1 Genomic Attributes

We have analyzed two genomic attributes: 1) Distance between the target and its neighboring gene. 2) Gene density of the genomic region in which target and its neighboring genes are present.

Distance

We wanted to check whether the distance between target and its neighbouring gene

show any effect on the transcriptional activity of the neighbouring gene. Bar plot in figure 5 shows that distance weakly affect the transcriptional activity of the neighbouring gene. Transcriptionly affected neighbouring genes are more distant from the estrogen target gene.

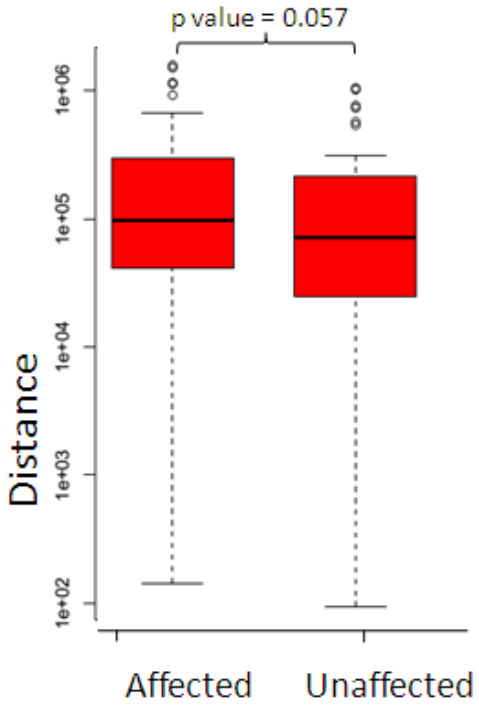

Fig 5: Boxplot for Distance between target and neighbouring genes of transcriptionly affected and unaffected. P-value was calculated using one tailed t test.

Gene Density

Gene density of the genomic region in which it is present influences its transcriptional activity [26, 26]. We have checked the effect of gene density on the transcriptional activity of neighbors of target genes. We did not did not find any significant effect of gene density on the transcriptional activity of the neighbouring gene. Also, there was no significant difference in gene densities of target and neighbouring genes. Boxplot in the figure 6 shows non-significant effect of gene density of the target and neighbouring genes on their transcriptional activity.

4.3 Chromatin Modification states

ChromHMM outputs the chromatin state modification data for each genomic position. The Figure 7 shows the heatmap of chromatin states generated as an output of

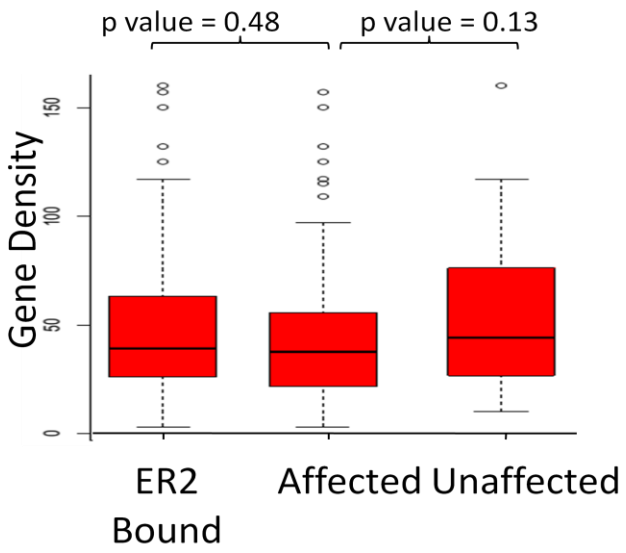

Fig 6. Boxplot for gene density between target and neighbouring genes of transcriptionly affected and unaffected. P-value was calculated using one tailed t test.

ChromHMM. Using the genomic coordinates of the target and neighbouring genes, we have found out the chromatin states of these genes. Then we checked the chromatins states of target and its neighboring genes.

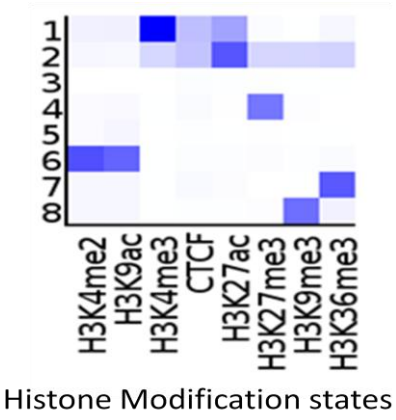

Fig 7. ChromHMM output: Eight different chromatin states were generated. Heatmap3 of chromatin state shows the enrichment for histone modifications. State 1 – Active promoter, State 2- Repeat region, State 3- Heterochromatin, State 4- Polycomb repressed, State 5- Heterochromatin, State 6- Enhancer, State 7- Transcribed, State 8- Heterochromatin. Same chromatin states were merged thus, six chromatin states were analyzed.

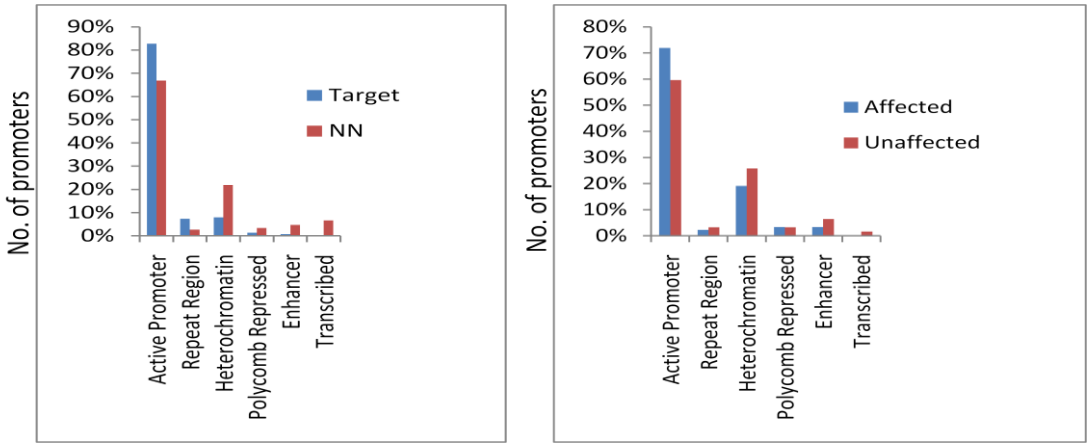

Fig 8: Relative percentage of Chromatin states of target and neighbouring genes. Where target is the estrogen bound gene and NN is nearest neighbor of the target gene. Active promoter state is dominant in both target and neighbouring gene.

Fig 9: Relative percentage of Chromatin states of transcriptionly affected and unaffected genes neighboring genes. Affected genes are transcriptionly affected and unaffected ones are transcriptionly unaffected. Active promoter state is dominant in transcriptionly affected neighbouring gene.

Chromatin sate data showed that Active promoter state is predominantly present in target and neighboring genes. Also, heterochromatin state was relatively dominant in the neighbouring genes. Bar plot in Figure 8 gives the relative proportion of all the chromatin states. Further, we have checked the distribution of active promoter state in transcriptionly affected and unaffected neighbouring genes. Active promoter state was more dominant in affected genes than unaffected genes. Bar plot in figure 9 shows the relative proportion of chromatin states in transcriptionly affected and unaffected neighbouring genes.

Next, we did a chi square test to check is the association between chromatin state modifications of neighbouring genes and their transcriptional activity. In particular, we have checked the association of active promoter state of neighbouring genes with their transcriptional activity. Figure 10 shows the proportional bar of the association.

We got weak association between chromatin state (Active promoter) and the transcriptional activity of the neighboring genes with the marginal significant p-value.

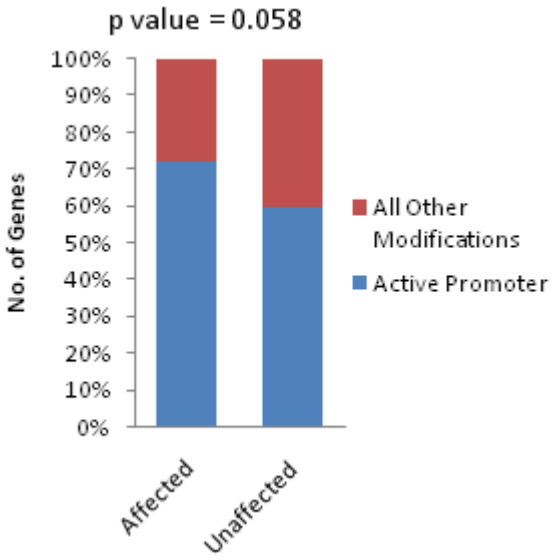

Fig 10: Relative percentage Active promoter chromatin state and rest of the chromatin modifications in transcriptionly affected and unaffected genes. P-value was calculated using Chi-square test.

4.4 Promoter Type (LCP and HCP)

After the chromatin state analysis, we wanted to check the promoter type of target and transcriptionly affected and unaffected neighboring genes.

Normalized CpG content at the promoter of a gene tells about the methylation pattern at the promoter and thus gives information about the transcriptional activity of the gene. So, we have checked the normalized CpG content of all target and neighboring genes. Firstly, we have checked the normalized CpG content of whole human genome as a reference. Figure 11 shows the density plot of normalized CpG content of all the promoters of the human genome. Density plot in Figure 12 shows that all the estrogen target genes are having low normalized CpG content at their promoters and tends to be tissue-specific.

Futher, we have checked the association of transcriptional activity of the neighboring genes with their promoter type. Significant numbers of transcriptionly affected genes

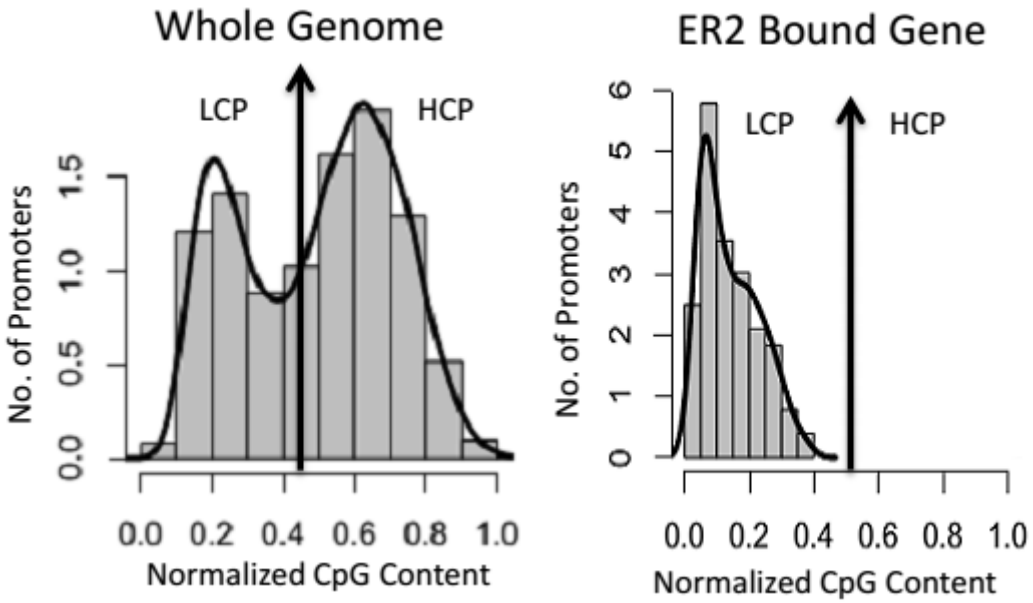

Fig 11: Normalized CpG content of all the promoters of human genome. Vertical line is at 0.45 CpG content. 0 to 0.45 is Low CpG promoters (LCP) and 0.46 to 1 is High CpG promoters (HCP).

Fig 12: Normalized CpG content of target gene promoters. Vertical line is at 0.45 CpG content. 0 to 0.45 is Low CpG promoters (LCP) and 0.46 to 1 is High CpG promoters (HCP)

are having high CpG promoters and thus tends to be housekeeping. Figure 13 shows the density plot comparison between transcriptionaly affected and unaffeced genes.

4.5 Chromatin interaction

To check whether the target and neighboring genes are in spatial proximity in the nucleus, we did chromatin interaction analysis. Out of 151 gene pairs, 20 show interaction however there is no significant association between chromatin interaction and the transcriptional activity of the neighbouring gene. Proportional bar plot in the figure 14 shows that there is no significant enrichment of the chromatin interaction in transcriptinaly affected neighboring genes.

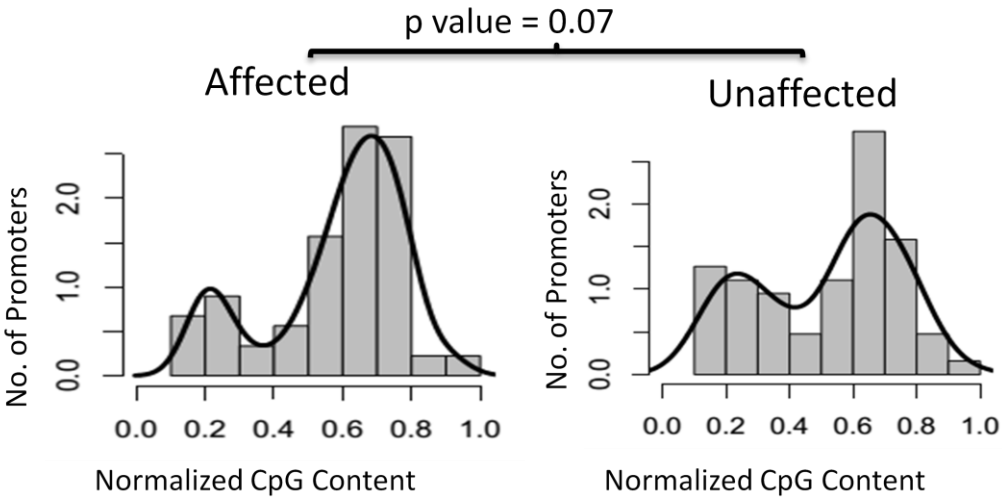

Fig 13: Normalized CpG content of transcriptionaly affected and unaffected neighboring genes. Number of high CpG promoters are significantly more in transcriptionly affected genes. P-value was calculated using Wilcoxon test.

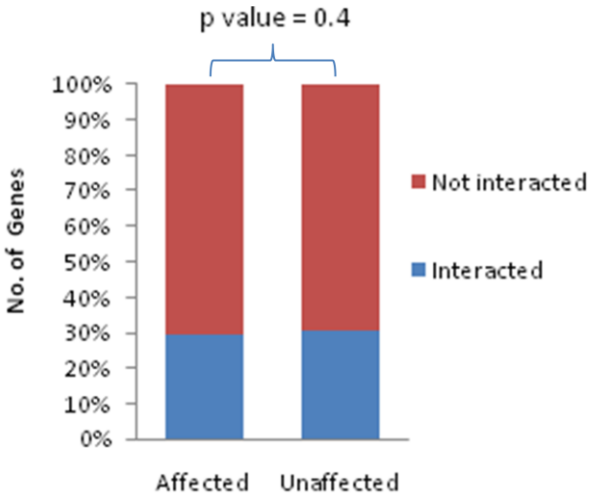

Fig 14: Spatial proximity data was used to get interacting target and neighboring gene pairs. Proportional bar plot between interacting and non-interacting gene pairs in transcriptionly affected and unaffected genes. P-value was calculated using chi-square test.

4.6 Functional Similarity

Further, we wanted to check functional similarity between the target and its neighbouring genes. We did not get any significant functional similarity between target and its neighbouring gene.

Chapter 3 Discussion

A previous study by Ebisuya et.al has demonstrated that transcriptional induction of a gene is not an independent process and is accompanied by simultaneous modulation of the expression profiles of neighboring genes. Here in our study, we confirm the ripple effect using a different induction system and found that out of 151 estrogen targets, 89 neighbouring genes showed significant alteration in their expression profiles.

Although under given set of conditions, the transcription profile of a cell is optimized to maximize the survival of a cell, yet it is prone to random fluctuations in its expression profile, which is termed as transcriptional noise. Although such random fluctuations may have an underlying biological significance, it is a homeostatic challenge for a cell to maintain its steady state transcription and a cell has to come up with strategies to minimize expression noise. Given all this, it is highly unlikely that modulation of gene expression that we observed in case of transcriptional ripples is a random effect associated with transcriptional induction and there must be epigenetic, molecular and functional basis of the transcriptional ripples. Since affected genes were associated with active chromatin state and were not associated with repressed chromatin states, one obvious fact that we have established is that ripples can affect the genes which show marks for active promoters, but cannot activate any nearby inactive gene. But this needs to be cross-validated using other systems as our P-value was only marginally significant (0.058). Further we found that affected genes are associated with High CpG promoters which means they should be housekeeping genes. Thus in case of Estrogen induced systems; transcriptional ripples can be a mechanism to modulate the expression of housekeeping genes to maintain the optimal relative gene expression levels.

Quite unexpectedly, we didn’t get any significant functional assocation between target and neighboring genes, which indicates that ripples are not mediated via direct relationship between target and neighbors (as observed in Operons of bacteria, where linear neighbors are functionally related and are regulated as a single unit).

Interestingly, the transcriptional ripples are associated only with its linear neighbours, and do not affect physical neighbors, which are brought there by long-range chromatin interactions. But as the genome of MCF-7 is multiploid, so the interaction data for the MCF-7 may not capture the actual chromatin interactions hence it would be useful to revisit the effect of spatial proximity on transcription ripples using cell lines other than MCF-7.

Chapter 4 Conclusion

- Affected gene is more distant from ER2 target gene than the unaffected and its significance is not yet known.
- No significant difference in the gene density of affected and unaffected gene regions was observed.
- Affected genes are more often actively transcribing and have high CpG promoters as compared to unaffected ones, suggesting that induction of a target gene can interfere with the promoter of actively transcribing gene in the neighbour for which the mechanism is not known.
- No significant enrichment of long range chromatin interactions between target-neighbouring was observed in affected and unaffected genes.

Chapter 5 Future Perspectives

- It will be insightful to look into other hormone induced or Transcription factor induced systems to check whether our observations are a generic feature of transcriptional regulation.
- Further, we would like to revisit the effect of spatial proximity on transcription ripples using some cell lines other than MCF-7 as its genome is multiploid. Recently the interaction data in ES cell lines have become available at unprecedented resolution, which can be a good system to study ripples in context of spatial proximity.
- Functional associations between target - neighbouring genes.
- Whether the alteration of neighbouring is not within the range of stochastic fluctuations in gene expression.

Long term objective:

By understanding the transcriptional ripple effect, we can explain the unexplained disease phenotype problem, where dysregulated or mutated gene has no direct link with the observed phenotype.

References

1) Soltysik K, Czekaj P. Membrane estrogen receptors - is it an alternative way of estrogen action? Journal of Physiology and pharmacology 2013.
2) Ebisuya et.al. Ripples from neighboring transcription. Nature 2008
3) Jacob, F.; Monod, J. (1961). "Genetic regulatory mechanisms in the synthesis of proteins". J. Mol. Biology 1961
4) Busby, S.; Ebright, R. Promoter Structure, Promoter Recognition, and Transcription Activation in Prokaryotes. Cell 1994
5) Payankaulam, S., Arnosti, D. Transcriptional repression: conserved and evolved features. Current Biology 2010.
6) Gruber, T., Gross, C. Multiple sigma subunits and the partitioning of bacterial transcription space. Annual Review Microbiology 2003.
7) Struhl K. Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. Cell 1999.
8) Wysocka, J. Modification of enhancer chromatin: what, how, and why? Molecular Cell 2013
9) Van Arensbergen, J; van Steensel, B; Bussemaker, HJ. In search of the determinants of enhancer-promoter interaction specificity. Trends in cell biology 2014
10) Mercer, TR; Mattick, JS Understanding the regulatory and transcriptional complexity of the genome through structure. Genome Research 2013.
11) Dekker, J; Marti-Renom, MA; Mirny, LA. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. Nature reviews. 2013
12) Ebisuya et.al. Ripples from neighboring transcription. Nature 2008
13) Guoliang Li et.al. Extensive Promoter-centered Chromatin interaction provide a topological basis for transcription regulation. Cell 2012.
14) Soule, HD; Vazquez J; Long A; Albert S; Brennan M. A human cell line from a pleural effusion derived from a breast carcinoma. Journal of the National Cancer Institute 1973.
15) Melissa J. Fullwood. An estrogen-receptor-α-bound human chromatin interactome. Nature 2009.

16) Kent WJ et al. The human genome browser at UCSC. Genome 2002.
17) Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. Nature Methods 2012.
18) Amir Ben-Dor, Ron Shamir, and Zohar Yakhini. Clustering Gene Expression Patterns. Journal of Computational Biology 1999.
19) M.J.L. de Hoon, S. Imoto, J. Nolan, and S. Miyano. Open source Clustering Software. Bioinformatics 2004.
20) Chunhui Hou, Li Li, Zhaohui S. Qin, Victore G. Corces. Gene Density, Transcription, and Insulators contribute to the partition of the drosophila genome into physical domains. Cell 2012
21) Nicola L. Mahy, Paul E. Perrey, Wendy A. Bickmore. Gene density and transcription influence the localization of chromatin outside of chromosome territories detectable by FISH. The journal of Cell Biology 2002.
22)
23) Jabbari K, Bernardi G. Cytosine methylation and CpG, TpG (CpA) and TpA frequencies. Gene 2004.
24) Serge Saxonov, Paul Berg, Douglas L. Brutlag. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. PNAS 2005.
25) Yu G, Li F, Qin Y, Bo X, Wu Y and Wang S. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. Bioinformatics 2010.
26) R development core team. R: a language and environment for statistical computing
27) Langmead B, Salzberg S. Fast gapped-read alignment with Bowtie2. Nature Methods 2012.
28) Li H et.al. The Sequence Alignment/Map format and SAM tools. Bioinformatics 2009
29) Derek W. Barnett et. al. Bam Tools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics 2011.