Dynamic transcriptional and epigenetic changes during tunicamycin-induced unfolded protein response in colorectal cancer cell

Su-Gyung Kim

Department of Medicine Science
The Graduate School, Yonsei University

Dynamic transcriptional and epigenetic changes during tunicamycin-induced unfolded protein response in colorectal cancer cell

Su-Gyung Kim

Department of Medicine Science
The Graduate School, Yonsei University

Dynamic transcriptional and epigenetic changes during tunicamycin-induced unfolded protein response in colorectal cancer cell

Directed by Professor Hyoung-Pyo Kim

The Master's Thesis
submitted to the Department of Medical Science,
the Graduate School of Yonsei University
in partial fulfillment of the requirements for the degree of
Master of Medical Science

Su-Gyung Kim

December 2021

This certifies that the Master's Thesis of

Su-Gyung Kim is approved.

Thesis Supervisor : Hyoung-Pyo Kim

Thesis Committee Member#1 : Myung-Shik Lee

Thesis Committee Member#2 : Hyun Seok Kim

The Graduate School

Yonsei University

December 2021

ACKNOWLEDGEMENTS

대학원에서의 2년의 석사과정을 정리하고, 어느새 졸업을 앞두고 있습니다. 이렇게 학위논문을 쓰기까지 도움을 주신 분들이 많습니다.

가장 먼저, 저의 지도 교수님이신 김형표 교수님께 감사하단 말씀 전하고 싶습니다. 단지 실험대에서의 일만 생각했던 저에게 생물정보 분석이라는 일을 제안해주시고, 충분히 고민하고 선택할 수 있게 해주셔서 이 길을 걸어갈 수 있었습니다. 개인적인 사정에 힘들었던 저에게 기둥이 되어 흔들리지 않게 붙잡아주신 점 마음 깊이 잊지 못할 것 같습니다. 항상 제가 하는 일에 예리한 조언과 아낌없는 지지를 해주셔서 자신감을 가지고 학업에 정진할 수 있었습니다. 또한, 저의 학위 논문을 관심있게 봐주시고, 피드백해주신 심사위원 이명식 교수님, 김현석 교수님 두 분께도 감사의 인사 드리고 싶습니다.

짧다면 짧고, 길다면 긴 2년의 학위 과정동안 많은 시간 함께 지냈던 연구원분들께도 감사드립니다. 시원 털털하시지만 누구보다 세심하신 미경 언니, 어떤 이야기라도 잘 경청해주시고 신경 써주시는 민지 언니, 조용하게 잘 챙겨주시는 웅재 오빠, 항상 긍정적으로 말씀해주시는 은총 오빠, 방장으로 항상 고생하시는 보배 언니, 저에게 부족한 지식을 서스럼없이 가르쳐주시는 정식 오빠, 대학원 생활에 있어서도 일에 대해서도 사수로서 많은 것을 알려주신 용진 오빠, 항상 편안한 분위기로 다가와주신 무구 오빠, 사소한 일에도 관심가져주신 경우 오빠, 입학 초에 당차게 이끌어주시던 예은 언니, 항상 멋지다고 응원해주시던 양철민 박사님 모두가

계셨기에 대학원 생활을 즐겁고 알차게 보낼 수 있었습니다.

마지막으로, 본인이 힘들지언정 지원을 아끼지 않으셨던 아버지, 항상 자랑스럽다고 말씀해주시고 저만을 생각하신 어머니, 인생의 동반자라고도 말할 수 있을 언니 그리고 나의 힐링 콩이. 우리 가족의 존재만으로 힘을 얻는 것 같습니다. 감사하고 또 감사합니다.

학위 과정동안 힘들었던 일들이 많았습니다. 그럼에도 꿋꿋이 앞으로 나아갔던 나 자신에게 수고했다는 말도 전하고 싶습니다.

도움주신 분들 덕분에 지난 학위 과정에서 소중한 추억과 많은 지식을 얻을 수 있었습니다. 이를 기반으로 앞으로도 더욱 발전해나가는 사람이 되도록 노력하겠습니다. 감사합니다.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Dynamic transcriptional and epigenetic changes during tunicamycin-induced unfolded protein response in colorectal cancer cell

Su-Gyung Kim

Department of Medical Science
The Graduate School, Yonsei University

(Directed by Professor Hyoung-Pyo Kim)

The abnormal accumulation of the mis/un-folding protein leads to cellular stress.[1] To maintain ER homeostasis, the unfolded protein response (UPR) signaling pathway is induced and is associated with the pathogenesis of various human diseases, especially cancer.[2,3] The molecular mechanism of the unfolded protein response (UPR) signaling pathway mediated by PERK, IRE1, ATF6 is well established.[1,2] However, whether and how chromatin structure, histone modifications, and distal regulatory elements control proteostasis under ER stress is less understood. In this study, i have employed the multi-omics sequencing data for the advance of understanding dynamic changes in chromatin state, especially enhancer activities, under ER stress. The

gene expression pattern of protein-coding genes and the genome-wide enhancer landscape were investigated by performing transcriptomic and epigenomic analyses, respectively. Taken together, our results unraveled dynamic changes in chromatin structure and enhancer activities which contribute to establishing an adequate gene expression program under ER stress.

Key words : ER-stress, tunicamycin, chromatin, enhancer

I. INTRODUCTION

Proteins form an appropriate tertiary structure through chaperone-mediated folding in the endoplasmic reticulum (ER) to function properly.[4] However, the accumulation of excessively secreted or mis/unfolded proteins induces ER stress. To relieve ER stress for proteostasis, eukaryotes activate the unfolded-protein response (UPR).[1] Three major pathways are processed by UPR, and they initiate respectively from IRE1a, PERK, and ATF6a, which are ER transmembrane protein sensors. They phosphorylate specific kinases or become cleaved forms to activate several transcription factors. These transcription factors move into the nucleus and regulate target genes involved in apoptosis, ERAD, and protein-folding.[1,2,5] It is well known that the UPR plays an important role in restoring proteostasis, but limited studies on epigenetic dynamics were reported when ER stress is induced.

Eukaryotes DNA organizes the chromatin structure by the basic unit, the nucleosome, together with histones. Histone modifications change the charge of nucleosomes or call-in multiple protein complexes to alter chromatin states.[6] The relationship between histone modifications and the transcriptional activity has been continuously revealed. For example, acetylation and phosphorylation generally accompany gene expression, while methylation and ubiquitination are involved in both activation and repression of transcription. That is, histone modifications play an important role in chromatin structure and gene expression.[7,8]

In addition, enhancers are key regulatory elements that control cell-specific spatiotemporal gene expression programs.[9] It regulates the transcription of target genes by physically contacting to long-distance genes that are several kilobases or hundreds of kilobases away or interacting with target genes via intermediate genes.[10] The fact that one enhancer can be linked to several genes and that one gene can be affected by multiple enhancers further complicates the relationship

between enhancers and genes.[11,12] In view of this, predicting the connection between an enhancer and a target gene only through location information on the genome map is highly likely to make an error.[13] Currently, the mechanism of cooperative regulation between enhancers and target genes in biology remains a major challenge. Many enhancer-promoter interactions have been discovered in the humans or mice genomes due to technological advances so far.

Colorectal cancer is a malignant tumor that progresses as a result of the continuous accumulation of variations in normal colonic epithelial cells, resulting in the development of colorectal adenoma or invasive adenocarcinoma.[14] It was predicted that 17,000 people would be diagnosed with colorectal cancer, and about 53,000 people with colorectal cancer would die.[15] The onset of colorectal cancer is caused by the complex action of genetic and environmental factors such as diet, overweight, activity, and smoking.[14] Surgery and chemotherapy have long been used for cancer patients, but the prognosis for colorectal cancer is not satisfactory, especially in cases with metastatic lesions. Currently, various therapies such as targeted therapy are being developed as a way to increase the survival rate of colorectal cancer patients.[16] In addition, the importance of relevant biomarkers for diagnosing and predicting response to treatment is emerging. It is important to study the biological phenomena of colorectal cancer in order to realize precision medicine and develop a treatment for colorectal cancer.

Disruption of ER proteostasis causes the promotion of unfolded protein response (UPR) along with ER stress.[1,2] In other words, it is a signal pathway that relieves cell stress through actions such as reducing the accumulation of un/mis-folded proteins by participating in protein translation, decomposing the produced protein, or correcting folding by increasing the ability of chaperones.[17] As a result, the amount of protein entering the ER decreases, restoring homeostasis again. However, if the ER stress is not relieved and the UPR is continuously activated, unstable cells have evolved to die to protect the organism.[2] This phenomenon is associated with several diseases such as metabolic diseases, neurodegenerative

diseases, cancer, and chronic inflammation.[1,3,5] In particular, from the viewpoint of cancer biology, UPR applies both functions of survival and apoptosis. In general, cancer develops in a stressful microenvironment, and the UPR response is used as a survival strategy. In fact, ER stress and UPR activity have been reported in various carcinomas, and it has been confirmed that they play an important role in cancer development step by step.[3,18,19] Therefore, the study of UPR signaling pathways will help to understand the process from the occurrence of tumors to the development of aggressive cancer cells.

Based on the research results that UPR plays an important role in cancer cells, UPR is induced by tunicamycin, which has already been identified as an ER stressor in cancer cell lines, especially HCT116, a colorectal cancer cell widely used in drug response studies. In this study, a variety of sequencing data were analyzed, which are commonly used in epigenetic studies such as ChIP-seq, ATAC-seq, RNA-seq and Hi-ChIP, to investigate how UPR affect the gene expression and chromatin states and whether the change of E-P interaction mediates the change of gene expression during UPR.

I explored ChIP-seq(Chromatin immunoprecipitation followed by sequencing) data to map transcription factor and histone modification throughout the genome.[20] Several ChIP-seq data using antibodies against various target proteins such as DNA binding protein and histone protein were processed with an in-house pipeline. In this study, XBP1, ATF4, ATF6 and CHOP, which are major protein for master regulator to induce UPR accumulating on transcriptionally active regulatory elements were sequenced and processed. Several histone modifications were also identified by widely used markers for chromatin transcriptional states such as H3K27ac, H3K4me1, H3K4me3, H3K27me3.[21,22] ATAC-seq(Assay for transposase-accessible chromatin) is a sequencing technique for profiling the genome-wide chromatin accessibility and the composition of accessible chromatin reflects a network of physical interaction between DNA regulatory elements or binding proteins and target genes.[23]

By integrating and analyzing these data, the expression pattern of protein-coding or non-coding genes in the UPR system is investigated. In addition, genes can be regulated by various mechanisms in addition to gene sequences. It is known that an enhancer, which is a regulatory element that contacts a target gene even at long distances or a modification of histone in contact with a gene sequence, is closely related to gene expression. The entire genome chromatin structure induced during UPR and the occupancy pattern of the related transcription elements are checked the interaction between the promoter of the target gene and the surrounding enhancer or histone changes, and functional aspects thereof are reviewed. In conclusion, by considering the results of changes in chromatin structure and enhancer regions according to the degree of UPR induction, it is intended to serve as a basis for subsequent studies to investigate the effect of epigenetic features on UPR-induced target genes.

II. MATERIALS AND METHODS

1. Dataset

All sequencing libraries were prepared in our laboratory and sequenced.

2. Annotation

A. Human Genome Sequence

Reference Human genome sequence file(v19, GRCh37.p13) and corresponding gene annotation GTF file were obtained from gencode website (http://www.gencodegenes.org/).[24]

B. Human transcription factor

The information for transcription factors in the human genome were downloaded from the combination of various databases (http://humantfs.ccbr.utoronto.ca/).[25]

3. Definition of Regulatory Element

A. Promoter

Promoter is defined as TSS±2.5kb flanking region.

B. Enhancer

H3K27ac ChIP-seq peaks called by MACS2 software were determined as enhancers. (see method below)

C. Super-enhancer

Super-enhancers were defined by ROSE algorithm based on enhancer regions (see method below).

4. Sequencing data processing

A. RNA-seq

Adapters of paired-end reads were trimmed using trim_galore. The trimmed

reads were aligned to the human genome assembly hg19 using STAR[26](v2.6.0a) with the default parameters but with the twopassMode set at Basic. Post-alignment quality was assessed using an in-house customized R script. BedGraph files were generated using deeptools bamCoverage with parameters --normalizeUsing CPM --binSize 1 and bigWigToBedGraph. Read count tables are created using RSEM[27](v1.3.1). Expression dynamics are evaluated using DESeq2[28](v1.26.0) with normalization by the number of uniquely mapped reads. Significant differentially expressed genes were determined by adjusted p value (FDR) < 0.05 and |log2FoldChange| > 1.

For hierarchical clustering of gene expression, Euclidean's distance measure and complete linkage clustering as implemented in the R stats(v3.6.1) package were performed. I selected the value of K at 6 to identify meaningful clusters by plotting dendrogram. The normalized counts of significant gene filtered by adjusted p value < 0.05 for all contrasts were used as input. The heatmap to visualize normalized signals of each cluster was generated by R pheatmap(v1.0.12). Functional analysis was performed using Metascape[29](v3.5) to identify GO terms enriched in differentially regulated genes and the results were plotted by R ggplot2.

B. ATAC-seq

ATAC-seq raw sequence data was trimmed using trim_galore to remove adapters. Mapping the trimmed reads to the reference genome using bowtie2[30](v2.3.5.1) with parameters --end-to-end --very-sensitive -X 2000 was performed. Following marking duplicated reads with Picard, mitochondrial reads and duplicates were eliminated using samtools. Fragment size distributions of each sample were checked using Picard CollectInsertSizeMetrics. For minimizing Tn5 transposase binding biases, uniquely mapped reads were shifted by +4bp on the positive strand and -5bp on the negative strand using deeptools function alignmentSieve. Only reads with a length of less than 100bp were extracted using deeptools alignmentSieve to

filter NFR(nucleosome free region). Peak calling was performed by MACS2 with parameters -f BAMPE --nomodel --min-length 100. Peaks located in the blacklisted regions were excluded. Bigwig files to visualize and use in other analyses were generated using deeptools bamCoverage with parameters --normalizeUsing CPM --binSize 1. Representative for downstream analysis is replicate 1.

C. ChIP-seq

Paired-end sequencing libraries were constructed for all types of sequencing methods and sequenced on the Illumina platform. Sequencing reads were quality-checked using fastqc and adapter sequences were trimmed using trim_galore(v0.6.4) (Table 1). The trimmed reads were aligned to the human reference genome (hg19) using bwa mem[31](v0.7.12) with the default parameters. Reads mapped on mitochondrial chromosomes, low-quality reads and duplicates marked with picard(v2.18.23) were eliminated using samtools[32](v1.9). Enriched regions defined as ChIP-seq peaks were called from bam file using MACS[33](v2.2.7.1) with options -g hs -f BAMPE --nomodel and scored against matched input libraries to remove artificial bias. Significant peaks with a q-value (false discovery rate) threshold of 0.01 are remained and peaks on genomic blacklisted regions were filtered. For visualization and downstream analysis, deeptools[34](v3.3.0) bamCoverage is used to convert uniquely mapped reads into bigwig files with parameters --normalizeUsing CPM --binSize 1. Representative for downstream analysis is replicate 1 and merged data of all sequencing data except H3K27ac ChIP-seq and H3K27ac ChIP-seq, respectively.

Regions at the peak center ±2 kb were sorted in descending order based on the mean value of the signal and visualized using deepTools plotHeatmap.

Peak distribution on genomic loci was calculated by HOMER[35] annotatePeaks.pl with a gtf file of GENCODE hg19 reference genome.

D. PRO-seq

After trimming the adapter sequences and checking quality, the paired reads were mapped to the human reference genome (hg19) using bwa mem[31] with the default parameters. Reads at mitochondrial chromosome, those with low-quality and duplicates marked with picard were removed using samtools[32]. PRO-seq peaks were called from bam file using HOMER[33] with options –style groseq. For visualization and downstream analysis, deeptools[34](v3.3.0) bamCoverage is used to convert uniquely mapped reads into bigwig files with parameters --normalizeUsing CPM --binSize 1. Representative for visualization is replicate 2. Regions at the peak center ±2 kb were sorted in descending order based on the mean value of the signal and visualized using deepTools plotHeatmap.

5. Motif enrichment analysis

Motif analysis was performed using findMotifsGenome.pl from HOMER to investigate enriched motifs within differential accessible regions and each cluster of enhancers. The enrichment score was defined as -log10(p value) meaning the significance of each motif site in a specific region compared to control regions randomly selected from the genome. For illustrating the motif search of differential accessible regions and enhancer clusters as heatmap, i filtered the results by P value < 1e-10 and P value < 1e-2, respectively.

6. Super-enhancer calling

H3K27ac peaks from pooled data were used as input of ROSE algorithm[36,37](v0.1) for calling super-enhancers after excluding peaks within promoter, ±2.5 kb of the transcription start site (TSS). Individual enhancer element was stitched together within a distance of 12.5 Kb, forming a long enhancer domain. The H3K27ac ChIP-seq signal of each stitched region was calculated and ranked. Based on a threshold at the tangent slope of 1, super-enhancers and typical-enhancer were classified, for example,

super-enhancers were defined as H3K27ac domains with the threshold above 1.

Table 1. Resources used in this study

Software and Algorithms	Source	Identifier
fastqc v0.11.9	Babraham Bioinformatics	https://github.com/s-andrews/FastQC
trim_galore v0.6.4	Babraham Bioinformatics	https://github.com/FelixKrueger/TrimGalore
bowtie2 v2.3.5.1	Ben Langmead et al., 2012	https://github.com/BenLangmead/bowtie2
bwa v0.7.12	Li an Durbin, 2010	https://github.com/lh3/bwa
SAMtools v1.9	Li et al., 2009	http://samtools.sourceforge.net
Picard v2.18.23	Broad Institute	https://github.com/broadinstitute/picard
deeptools v3.3.0	Ramírez et al., 2016	https://github.com/deeptools/deepTools/
MACS2 v2.2.7.1	Zhang et al., 2008	https://github.com/taoliu/MACS
bedtools v2.29.2	Aaron R Quinlan et al., 2010	https://github.com/arq5x/bedtools2
ROSE v0.1	Warren A. Whyte et al., 2013	http://younglab.wi.mit.edu/young.html
STAR v2.6.0a	Dobin et al., 2013	https://github.com/alexdobin/STAR
RSEM v.1.3.1	Li and Dewey, 2011	https://deweylab.github.io/RSEM/
DESeq2 v1.26.0	Love et al., 2014	https://doi.org/10.18129/B9.bioc.DESeq2
Metascape v3.5	Yingyao Zhou et al.,2019	https://github.com/data2code/msbio/

III. RESULTS

1. Expression pattern of protein-coding genes during the UPR

All sequencing data were generated from tunicamycin-treated cells for 0h, 6hr, 16hr, which are labeled as Con, 6hr, 16hr, respectively (Fig. 1A). To identify how transcriptional activity is changed during UPR and what characteristics it has, RNA-seq analysis was performed (Table 2). RNA-seq is one of widely used sequencing methods to quantify mRNA transcripts. At first, it was confirmed whether UPR was well-induced by tunicamycin and the data can reflect transcription profiles through the genomic viewer at the target genes regions (Fig. 1B). High correlation among replicates was verified for the differential analysis(Fig. 1C). Differentially expressed genes of con vs 6hr, 6hr vs 16hr and con vs 16hr were represented (Fig. 1D-F). 683 genes were up-regulated and 746 of genes were down-regulated significantly after the treatment of tunicamycin (Fig. 1D). The number of genes up-regulated and down-regulated in 6hr compared to 16hr were 316 and 1134, each (Fig. 1E). The expression of 524 and 96 genes were respectively increased and decreased in 16hr versus Con (Fig. 1F). To classify the transcriptional dynamics in more detail, i clustered significant genes by normalized counts (Fig. 2A). The transcription-decreased genes in 6h compared to Con were involved in cluster 1, 5, 6 and they mainly serve morphogenesis or ribosome biogenesis (Fig. 2B,F,G). On the contrary, focused on genes with increased expression during UPR, i investigated GO terms of cluster number 2, 3, and 4 among 6 clusters (Fig. 2C-E). Not surprisingly, a majority of up-regulated genes when ER stress was induced belong to the GO terms related to the response to endoplasmic reticulum stress. Cluster 2 was defined as genes with largely increased expression at the 6hr time point and a decrease to Con level at 16hr. While, up-regulated genes in 6hr compared to Con

whose transcription levels in 16hr were below the level of Con were called as cluster 3. The genes that were up-regulated continuously to at the 16hr time point belongs to cluster 4. Via GO analysis by cluster, It was found that ‘peptide biosynthetic process’ and ‘translation’ was for cluster 2 (Fig. 2C). GO terms related to ‘mRNA processing and splicing’ were shown as the result of cluster 3 (Fig. 2D). It was represented that genes with a long-lasting increase in expression mainly have ‘response to endoplasmic reticulum stress’, ‘proteasomal protein catabolic process’ and ‘ERAD pathway’ (Fig. 2E). The results show that the characteristics of genes induced by ER stress were distinguished according to the change of transcription activity from the basal state to the adapted cells through activation. Moreover, these suggest that UPR proceeds in the order of protein synthesis, accompanied by changes in associated transcriptional activity.

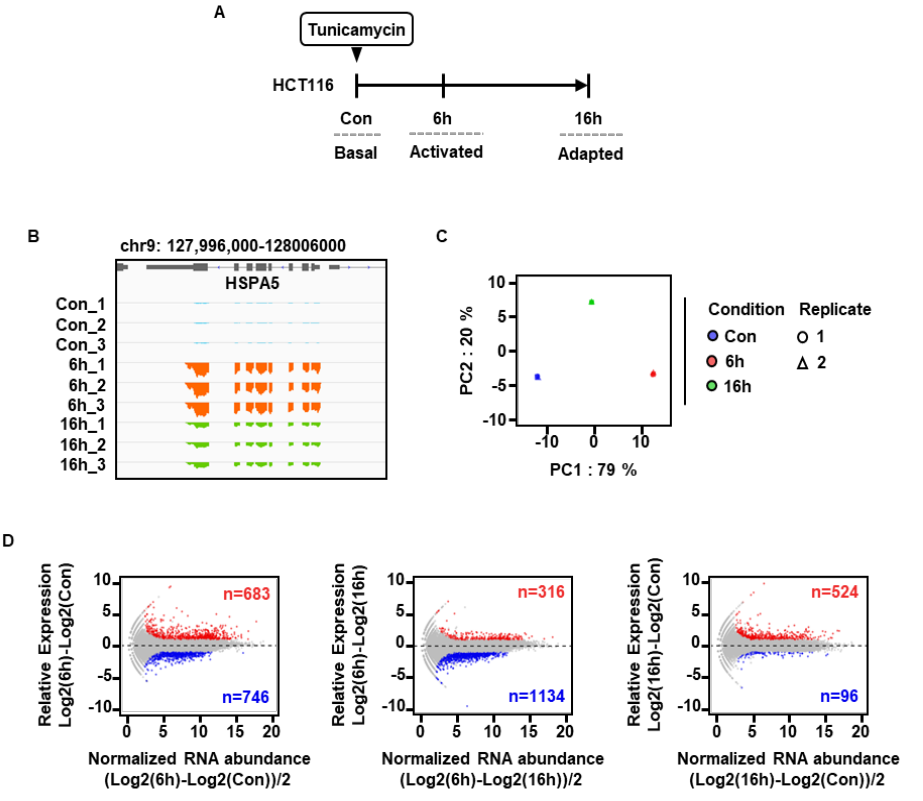

Figure 1. Global transcriptome profiling under ER stress

(A) The scheme of our experiment; Con, 6hr, 16hr were 0hr, 6hr, 16hr after treatment of tunicamycin in HCT116 cell lines, respectively. (B) Snapshot showing RNA transcription. (C) PCA analysis using normalized counts of RNA-seq to confirm correlation between replicates. (D) The identification of significant differentially expressed gene; con vs 6hr (left), 6hr vs 16hr (middle), and con vs 16hr (right). Red dots are up-regulated and blue ones are down-regulated genes. The cutoff applied on all contrasts is FDR < 0.05 & |log2FoldChange| > 1.

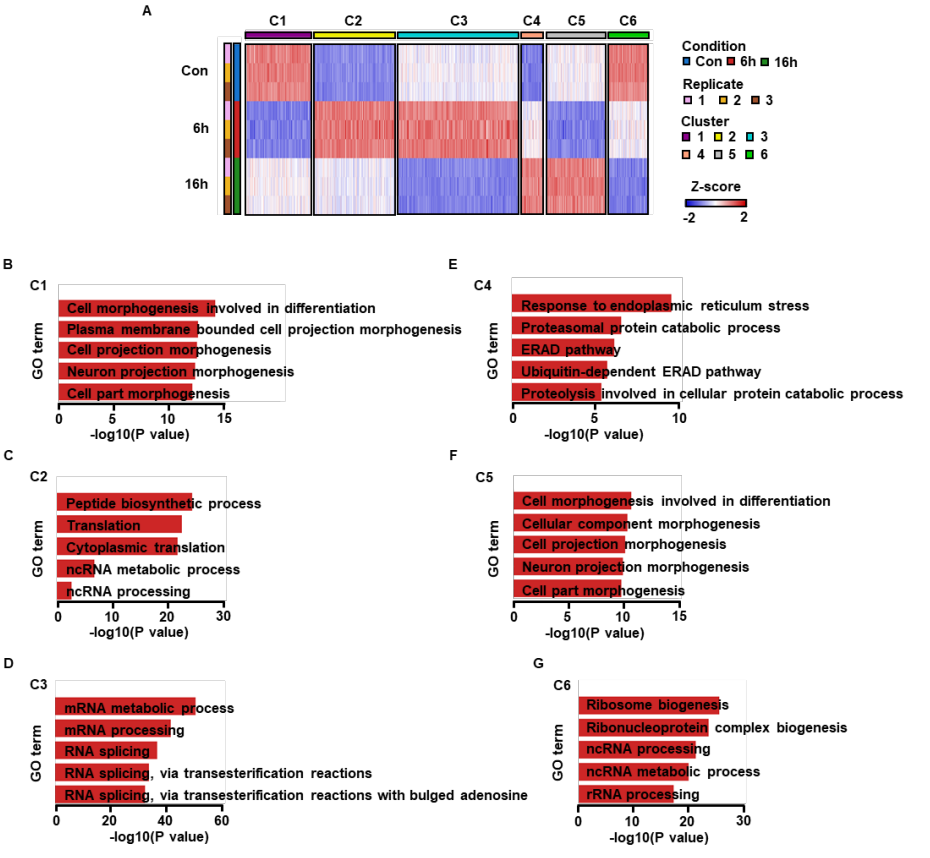

Figure 2. The change and characteristics of gene expression when ER stress induced (A) Hierarchical clustering of gene expression pattern. Z-score was calculated by the normalized gene counts. (B-G) GO terms analyzed using cluster 1 (B), 2 (C), 3 (D), 4 (E), 5 (F), 6 (G).

Table 2. Summarized mapping results of RNA-seq, ATAC-seq, and ChIP-seq

Assay	Antibody	Treatment	Replicate	Aligned Reads	Duplication Rate	Used reads in Analysis
ChIP-seq	XBP1	Tunicamycin 0h	1	19,820,599	15 %	15,026,621
ChIP-seq	XBP1	Tunicamycin 0h	2	21,123,553	16 %	15,995,208
ChIP-seq	XBP1	Tunicamycin 6h	1	20,992,467	15 %	16,119,904
ChIP-seq	XBP1	Tunicamycin 6h	2	22,489,952	16 %	17,043,073
ChIP-seq	XBP1	Tunicamycin 16h	1	34,213,609	19 %	25,065,817
ChIP-seq	XBP1	Tunicamycin 16h	2	33,424,797	18 %	24,646,599
ChIP-seq	ATF4	Tunicamycin 0h	1	35,720,984	23 %	24,859,317
ChIP-seq	ATF4	Tunicamycin 0h	2	35,903,020	22 %	25,244,512
ChIP-seq	ATF4	Tunicamycin 6h	1	35,852,391	25 %	24,372,424
ChIP-seq	ATF4	Tunicamycin 6h	2	36,182,748	21 %	25,838,813
ChIP-seq	ATF4	Tunicamycin 16h	1	35,562,042	19 %	25,826,487
ChIP-seq	ATF4	Tunicamycin 16h	2	36,457,311	21 %	26,041,506
ChIP-seq	ATF6	Tunicamycin 0h	1	23,241,552	17 %	17,272,515
ChIP-seq	ATF6	Tunicamycin 0h	2	24,652,538	17 %	18,283,714
ChIP-seq	ATF6	Tunicamycin 6h	1	21,526,872	19 %	15,739,453
ChIP-seq	ATF6	Tunicamycin 6h	2	27,854,843	21 %	19,837,027
ChIP-seq	ATF6	Tunicamycin 16h	1	21,876,725	17 %	16,267,252
ChIP-seq	ATF6	Tunicamycin 16h	2	21,590,622	16 %	16,220,984
ChIP-seq	CHOP	Tunicamycin 0h	1	23,653,602	28 %	15,351,670
ChIP-seq	CHOP	Tunicamycin 0h	2	36,450,282	35 %	21,177,447
ChIP-seq	CHOP	Tunicamycin 6h	1	27,295,849	17 %	20,456,139
ChIP-seq	CHOP	Tunicamycin 6h	2	23,819,541	15 %	18,316,180
ChIP-seq	CHOP	Tunicamycin 16h	1	22,853,331	20 %	16,540,215
ChIP-seq	CHOP	Tunicamycin 16h	2	23,538,799	22 %	16,628,407
ChIP-seq	H3K27ac	Tunicamycin 0h	1	101,097,774	14 %	80,535,607
ChIP-seq	H3K27ac	Tunicamycin 0h	2	116,125,866	15 %	91,208,604
ChIP-seq	H3K27ac	Tunicamycin 6h	1	104,312,368	15 %	82,253,690
ChIP-seq	H3K27ac	Tunicamycin 6h	2	125,199,444	15 %	98,583,061
ChIP-seq	H3K27ac	Tunicamycin 16h	1	109,578,058	15 %	86,038,567
ChIP-seq	H3K27ac	Tunicamycin 16h	2	104,368,723	15 %	82,208,529
ChIP-seq	H3K27ac	Tunicamycin 0h	Merged	217,223,562	15 %	171,184,353
ChIP-seq	H3K27ac	Tunicamycin 6h	Merged	229,511,830	16 %	180,168,013
ChIP-seq	H3K27ac	Tunicamycin 16h	Merged	213,946,782	15 %	167,812,918
ChIP-seq	H3K27me3	Tunicamycin 0h	1	27,477,786	10 %	22,648,383
ChIP-seq	H3K27me3	Tunicamycin 0h	2	36,519,769	9 %	30,733,274
ChIP-seq	H3K4me1	Tunicamycin 0h	1	36,146,159	5 %	31,738,947
ChIP-seq	H3K4me1	Tunicamycin 0h	2	33,176,502	12 %	27,218,512
ChIP-seq	H3K4me3	Tunicamycin 0h	1	36,294,604	11 %	29,720,683
ChIP-seq	H3K4me3	Tunicamycin 0h	2	35,994,092	10 %	29,632,624
ChIP-seq	H3K27me3	Tunicamycin 6h	1	31,873,290	12 %	25,863,819
ChIP-seq	H3K27me3	Tunicamycin 6h	2	36,371,614	11 %	30,025,446
ChIP-seq	H3K4me1	Tunicamycin 6h	1	36,138,203	7 %	31,224,295
ChIP-seq	H3K4me1	Tunicamycin 6h	2	36,215,801	7 %	31,268,067
ChIP-seq	H3K4me3	Tunicamycin 6h	1	32,618,673	15 %	25,548,231
ChIP-seq	H3K4me3	Tunicamycin 6h	2	36,344,846	11 %	29,731,093
ChIP-seq	H3K27me3	Tunicamycin 16h	1	31,388,329	10 %	26,131,505
ChIP-seq	H3K27me3	Tunicamycin 16h	2	31,832,246	13 %	25,735,901
ChIP-seq	H3K4me1	Tunicamycin 16h	1	31,281,954	12 %	25,732,065
ChIP-seq	H3K4me1	Tunicamycin 16h	2	30,352,669	12 %	24,922,166
ChIP-seq	H3K4me3	Tunicamycin 16h	1	33,445,823	13 %	26,719,978
ChIP-seq	H3K4me3	Tunicamycin 16h	2	29,295,787	15 %	23,064,833
RNA-seq	-	Tunicamycin 0h	1	-	-	81,129,072
RNA-seq	-	Tunicamycin 0h	2	-	-	80,037,976
RNA-seq	-	Tunicamycin 0h	3	-	-	81,548,544
RNA-seq	-	Tunicamycin 6h	1	-	-	79,690,194
RNA-seq	-	Tunicamycin 6h	2	-	-	75,497,464
RNA-seq	-	Tunicamycin 6h	3	-	-	69,365,122
RNA-seq	-	Tunicamycin 16h	1	-	-	70,594,306
RNA-seq	-	Tunicamycin 16h	2	-	-	73,615,496
RNA-seq	-	Tunicamycin 16h	3	-	-	81,230,142
ATAC-seq	-	Tunicamycin 0h	1	83,826,695	21 %	55,447,672
ATAC-seq	-	Tunicamycin 0h	2	84,916,422	24 %	54,799,610
ATAC-seq	-	Tunicamycin 6h	1	79,923,541	20 %	54,687,986
ATAC-seq	-	Tunicamycin 6h	2	87,348,995	21 %	58,695,526
ATAC-seq	-	Tunicamycin 16h	1	83,287,929	23 %	54,562,460
ATAC-seq	-	Tunicamycin 16h	2	82,826,070	17 %	58,715,314
PRO-seq	-	Tunicamycin 0h	1	61,084,859	27 %	33,724,063
PRO-seq	-	Tunicamycin 0h	2	60,614,132	24 %	35,148,661
PRO-seq	-	Tunicamycin 6h	1	61,908,436	28 %	31,917,037
PRO-seq	-	Tunicamycin 6h	2	56,844,589	27 %	29,090,642
PRO-seq	-	Tunicamycin 16h	1	54,411,174	20 %	31,694,315
PRO-seq	-	Tunicamycin 16h	2	54,879,992	24 %	29,258,249

2. The change of genome-wide chromatin structure during UPR

ATAC-seq method was used to identify one of various chromatin features, chromatin accessibility. Chromatin accessibility means the extent to which chromatin opens for DNA-binding proteins. Given the heatmap, it is observed that average signals of ATAC-seq were similar among all samples (Fig. 3A). It is implied that chromatin accessibility and its characteristics were generally unchanged upon the treatment of tunicamycin. However, regions with significantly differential accessibility were detected at a few specific regions. Samples tunicamycin-treated for 6hr have 924 and 416 regions that are more and less opened compared to Con, respectively (Fig. 3B). Motif enrichment analysis was performed by the differential accessible regions of Con vs 6hr to focus on the direct effect of ER stress. Mostly, enriched motifs were also similar. But, motifs with higher enrichment of gained than lost accessible regions in 6hr were AARE, ATF4 and CHOP, etc. They are renowned transcription factors associated with UPR, verifying that ATAC-seq data reflects the chromatin open status well (Fig. 3C). Among enriched motifs at gained accessible regions, ATF3, ATF4, CHOP have up-regulated expression in 6hr versus Con. They are likely to be the putative transcription factors that affect the expression of target genes when ER stress induced (Fig. 3D-E).

Furthermore, i explored ChIP-seq data for a variety of histone markers such as H3K4me1, H3K4me3, H3K27me3 and H3K27ac to find out what chromatin status are at each time point under ER stress. The regions that are enriched with H3K4me1 and H3K4me3 mean the poised enhancer and the active promoter, respectively. H3K27me3 is a repressive mark and H3K27ac is defined as an active enhancer marker. By identifying profiles of those histone markers, i examined whether ER stress influences histone modifications genome-wide. All histone ChIP-seq signals except for H3K27ac were partially altered at only low signals during UPR, suggesting that tunicamycin treatment had little effect on histone modifications (Fig. 4A-D).

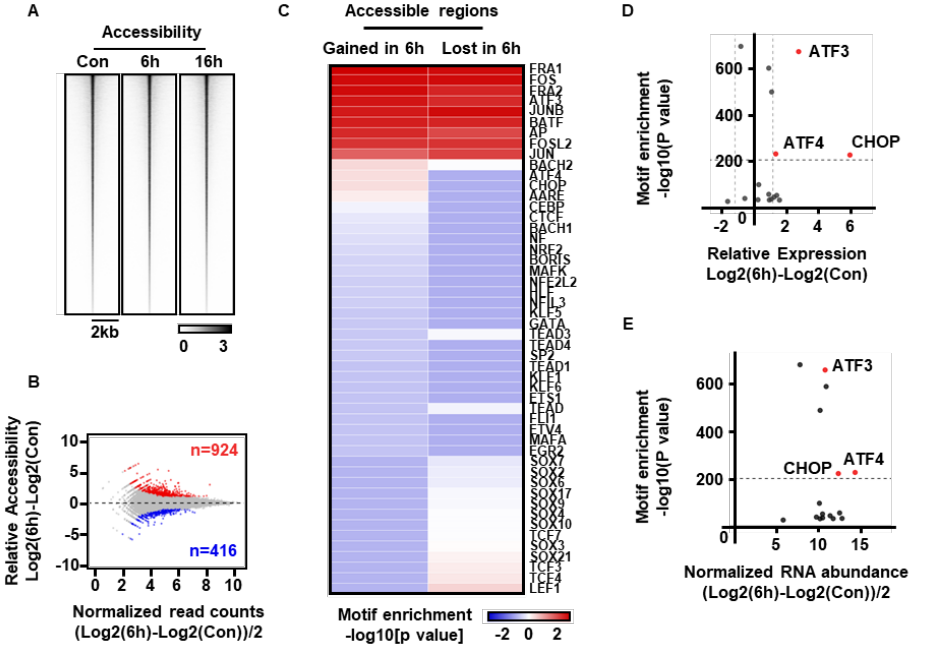

Figure 3. The pattern of chromatin accessibility under ER stress (A) Heatmap of ATAC-seq signals for genomic accessibility. (B) Significant differentially accessible regions; con vs 6hr. Red and blue dots represents gained and lost accessible regions, each. The cutoff applied on all contrasts is p value < 0.05 & |log2FoldChange| > 1. (C) Motif enrichment analysis performed using gained and lost accessible regions in 6hr sample versus Con sample. (D-E) The scatter plots showing the transcription level of the motifs enriched at the regions with increased accessibility at 6hr compared to Con. X axis shows the relative expression in 6hr compared to Con (D) and the normalized RNA abundance (E).

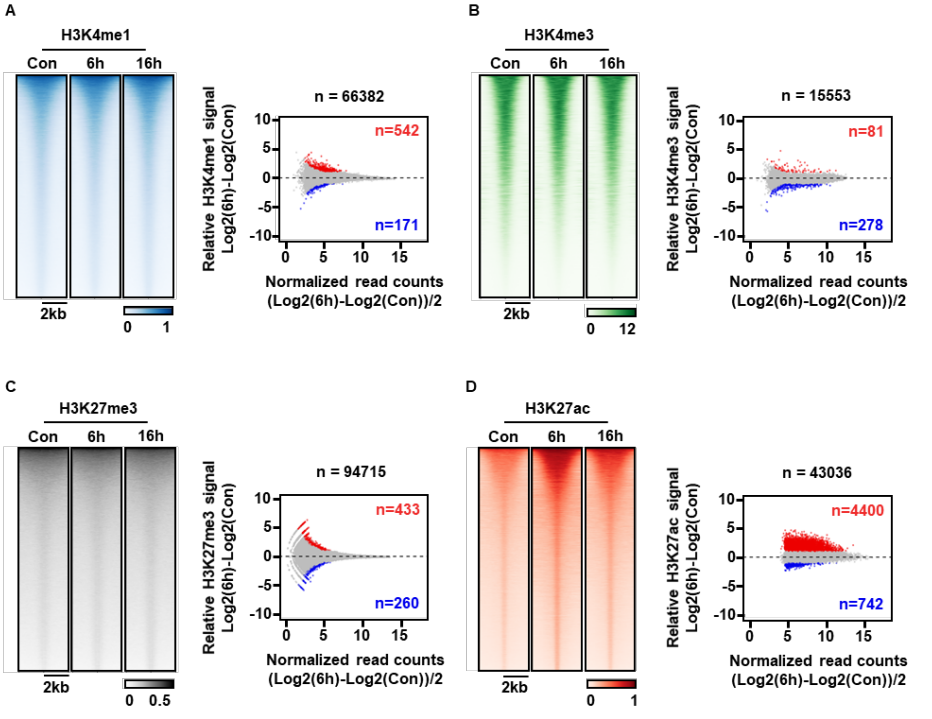

Figure 4. The profiling of histone modification (A-D) Heatmaps and MA plots of H3K4me1 (A), H3K4me3 (B), H3K27me3 (C), H3K27ac (D) ChIP-seq signals with or without tunicamycin treatment. Left : Heatmap of ATAC-seq signals for genomic accessibility. Right : Significant differentially accessible regions; con vs 6hr. Red and blue dots represents gained and lost accessible regions, each. The cutoff applied on all contrasts is p value < 0.05 & |log2FoldChange| > 1.

3. The dynamics in enhancer landscape during UPR

To identify the effects of ER stress on enhancer landscape, H3K27ac ChIP-seq data was analyzed after pooling replicates. Given that average enhancer signal was strengthen during UPR and reduced below the level of Con at the 16hr time point (Fig. 5A-B), the number of enhancers with an increased intensity were 4400 and 3588 in 6hr samples versus Con and 16hr, respectively. Also, 1709 of enhancer signals were gained in 16hr compared to Con (Fig. 5C). For detail classification of enhancers changes, clustering by normalized counts was performed. Though i got 6 clusters total, cluster 4, 5 and 6 were ignored because of few contents (Fig. 5D). Each cluster was applied by motif enrichment analysis. To concentrate on the increasing trends, it is observed by heatmap that motifs of ATF1, ATF4, CHOP, CEBP, NFIL3 and HLF were more enriched in cluster 1 than cluster 2 (Fig. 5E). Among them, ATF4 and CHOP were more enriched at H3K27ac peaks of C1 than C2 and their expressions also were up-regulated at the 6hr time point versus Con (Fig. 5F-G). It indicates that they can be putative transcription factors to influence transcriptional signals via the change of enhancer activity, particularly enhancers that do not have less signal than control in ER stress-adapted cells due to rebound.

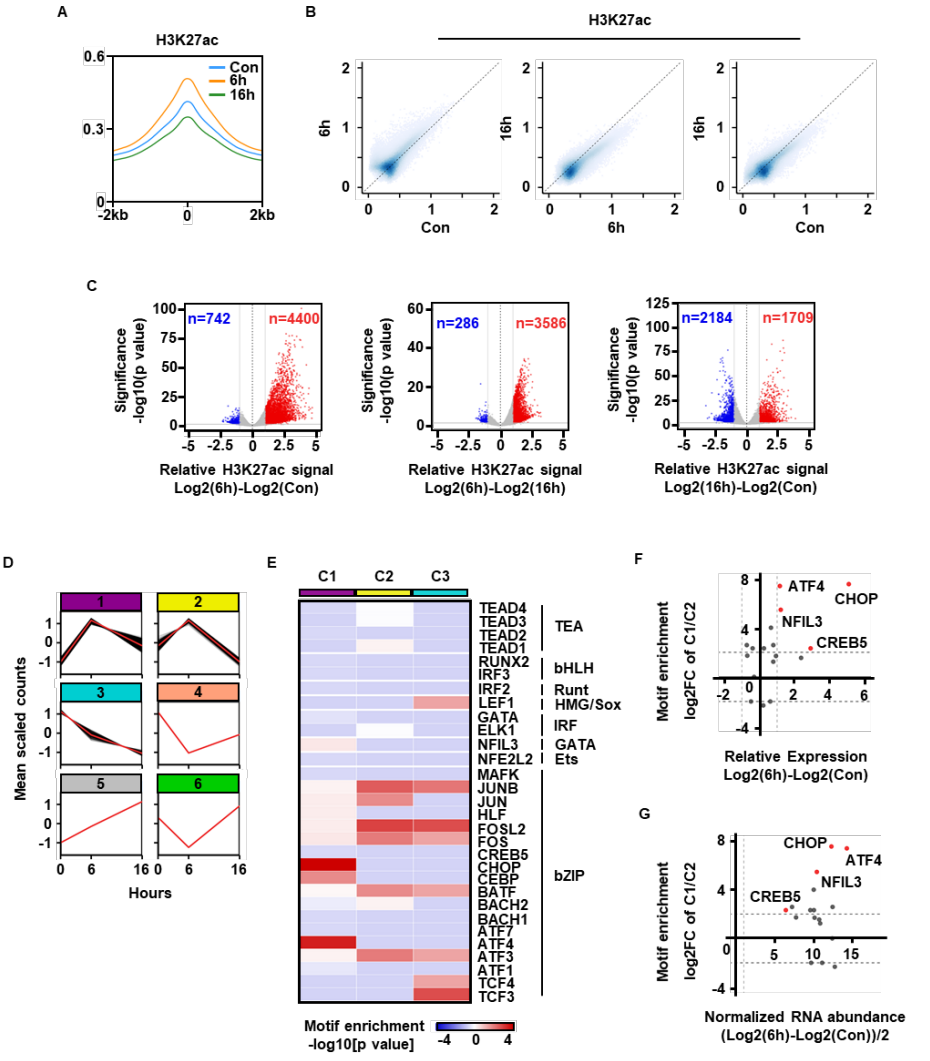

Figure 5. ER stress influences the enhancer landscape by diverse patterns (A) Average enrichment of H3K27ac ChIP-seq signal as enhancer activity in 3 conditions; Blue : cells without the tunicamycin treatment, Red : tunicamycin-treated HCT116 for 6hr, Green : tunicamycin-treated HCT116 for 16hr. (B-C) The scatter plots (B) and the volcano plots (C) showing the increase upon the treatment of tunicamycin and reductions upon ER stress adaptation in global occupancies. (D) The clusters of enhancer pattern by the normalized counts of ChIP-Seq signals called for H3K27ac. (E) The heatmap generated for showing the motif enrichment at each cluster of H3K27ac ChIP-seq signals. Only motifs which are in human transcription factors database were represented. (F-G) The change of expression (F) and the average of normalized expression (G) of genes with motifs that were enriched at cluster 1 of enhancer signal than cluster 2.

4. Genome-wide binding pattern of UPR-related transcription factors

The ChIP-seq data for specific target proteins were applied to examine how binding profile of transcription factors related to UPR were changed and what features each TF has. There are a diversity of transcription factors associated with UPR. Among them, i selected 4 transcription factors, XBP1, CHOP, ATF4, ATF6, which are activated through main signaling pathway when ER stress occurs. As the treatment of tunicamycin for 6hr, the number of their binding sites and their intensity all were shown to be increased. After UPR-induced for 16hr, TF binding signals were declined to or below the level of Con (Fig. 6A-B). Here, i also discovered that ATF4 and ATF6 have basal level even at the Con sample in contrasts to XBP1 and CHOP. It means that unlike XBP1 and CHOP, ATF4 and ATF6 are likely to affect expression of target genes even through the decrease in binding strength. I investigated the genomic distributions of each TF binding site, showing that XBP1 and ATF6 have tendency to binds at promoter. On the other hands, binding sites of CHOP and ATF4 were mostly on intron or intergenic regions (Fig. 6C). It indicates that CHOP and ATF6 work in an enhancer-centric manner, whereas XBP1 and ATF4 have a promoter-centric binding pattern. From these results, i suggest the possibility that XBP1 and ATF6 affect the transcriptional activity mainly at its promoter, and ATF4 and CHOP indirectly regulates the expression of target genes by locating on distal region.

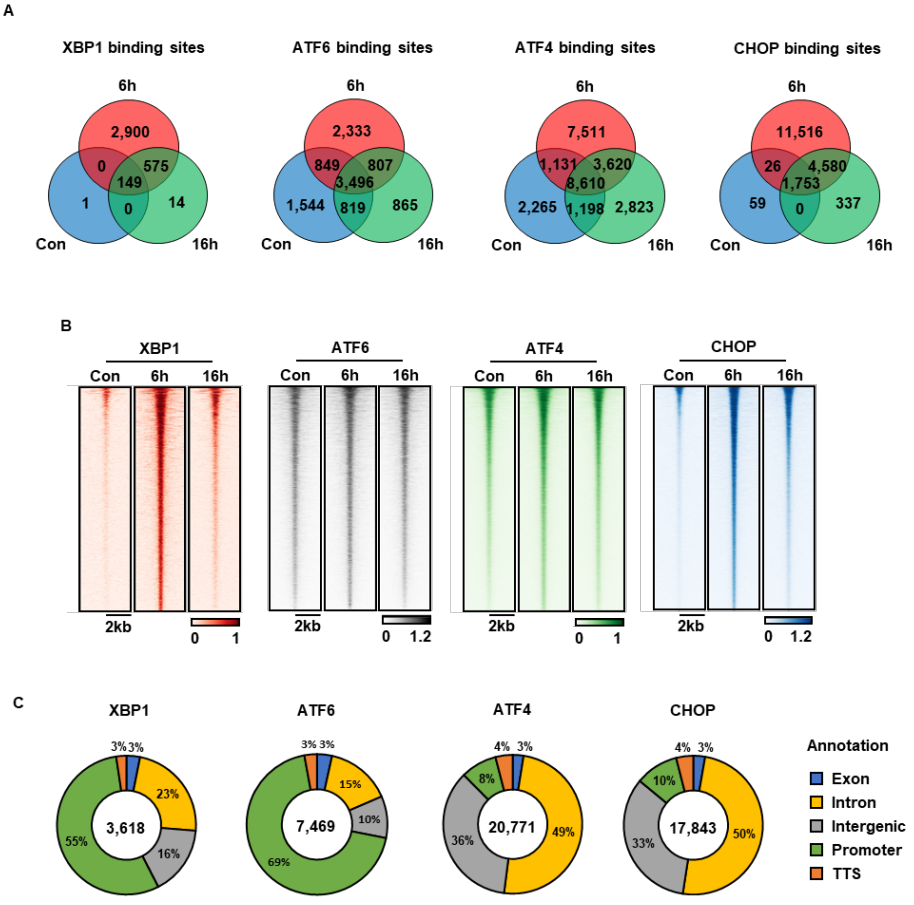

Figure 6. The occupancy and intensity of well-known transcription factors activated by ER stress (A) The ChIP-seq peaks of all conditions, Con, 6hr and 16hr for XBP1 (first left), ATF6 (second left), ATF4 (second right) and ATF6 (first right). Con (blue): untreated wild-type HCT116, 6hr (red): wild-type cell treated with tunicamycin for 6 hours, 16hr (green): HCT116 treated with tunicamycin for 3 hours. (B) The heatmaps showing average signals of each target proteins at the flanking regions ±2kb of their binding sites. Reds are for XBP1 and greys are for ATF6. ATF4 and CHOP signals represent as green and blue, respectively. (C) The genomic distributions of XBP1, ATF6, ATF4, CHOP from left to right side. The numbers in the center of the doughnut plot are the number of peaks.

5. The relationship among enhancer landscape, UPR-related transcription factors and gene expression during UPR

Given that enhancer activity was increased during UPR, i explored the TF ChIP-seq and H3K27ac ChIP-seq together to examine whether the transcription factors induced by ER stress are related to the change, especially the increase, of enhancer landscape. The occupancy of transcription factors per each H3K27ac cluster was identified by generating heatmaps. Here too, clusters from number 4 to 6 were also not considered due to a small number of subjects. All of the transcription factors are correlated to the gained enhancer regions in 6hr versus Con, and transcription factors did not bind on the H3K27ac signals with a decrease. The result of ATF4 and CHOP with signals spreading in most of the H3K27ac region is due to their enhancer-centric nature, as opposed to XBP1 and ATF6 (Fig. 7A). Viewing from the opposite perspective, i distinguished the H3K27ac peaks as TF bound and unbound, and compared H3K27ac signals statistically (Fig. 7B). It is examined that transcription factors related to UPR have a connection with enhancer activity. Moreover, XBP1 and ATF6, ATF4 and CHOP are grouped by two to show a difference in pattern. Significantly gained H3K27ac peaks are much bound by ATF4 and CHOP contrary to XBP1 and ATF6, suggesting that ATF4 and CHOP are likely to affect the enhancement of enhancer signal (Fig. 7C). Then, i processed the integrative analysis with RNA-seq, H3K27ac ChIP-seq and TF ChIP-seq to identify the effect of TF-bound enhancer on gene expression. The results show that adjacent genes of TF-bound enhancer have more increased expression in 6hr compared to Con than those of TF-unbound enhancer (Fig. 7D). It is suggested that TFs regulate the expression of target genes through the modulation of enhancer signal.

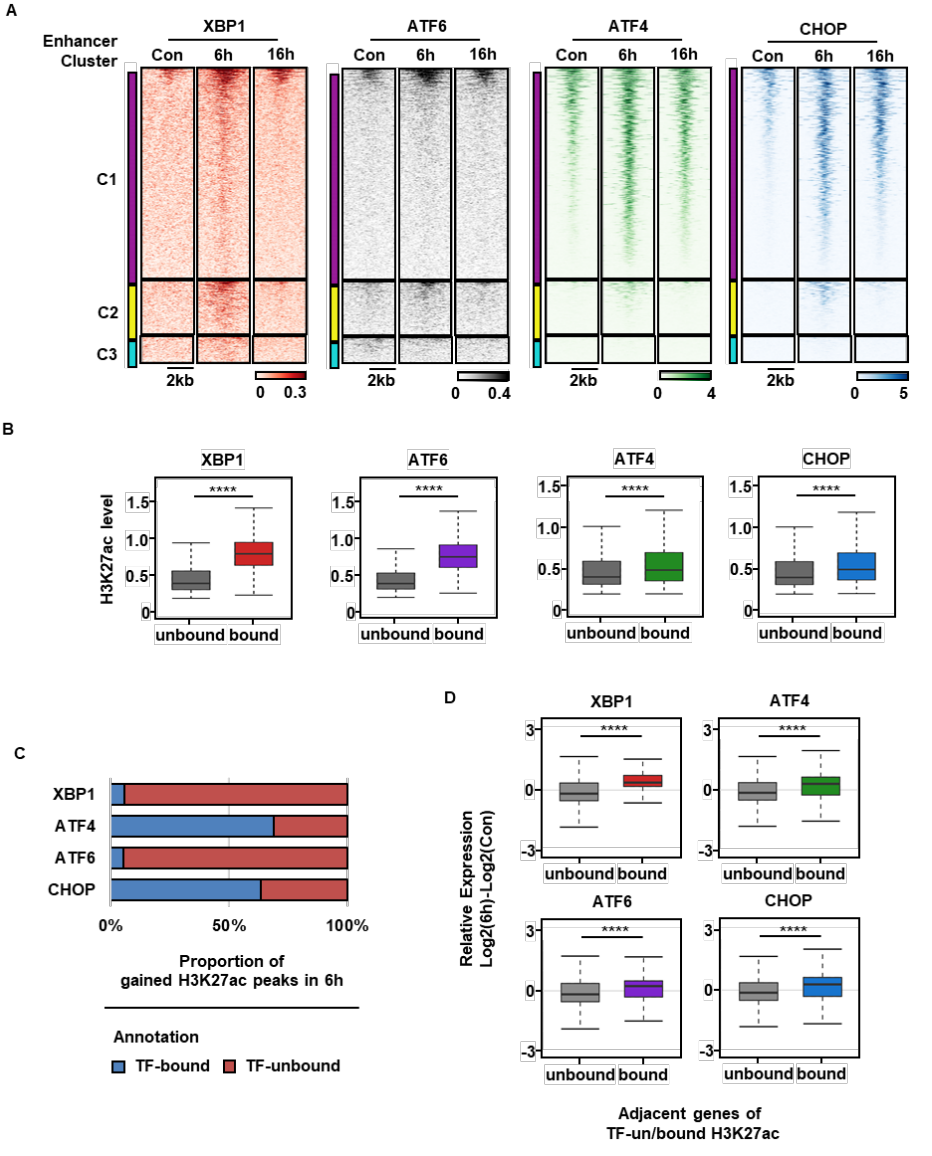

Figure 7. The dynamics of enhancer landscape was associated with the binding of transcription factors related to UPR (A) Average binding intensity of the target binding proteins at each H3K27ac cluster. Red heatmaps were generated for XBP1 signals and grey ones were for ATF6. ATF4 and CHOP were illustrated as green and blue heatmap, each. The cluster 1 of H3K27ac signal was marked by purple. The cluster 2 and 3 were denoted as yellow and turquoise blue, respectively. (B) The difference of H3K27ac level according to the transcription factor binding in all conditions; The grey box plot represents the H3K27ac signals at transcription factor un-bound enhancers and the box plot with each color shows the H3K27ac signals at the enhancers that are bound by each transcription factor. (**** : $p < 0.0001$, wilcox test) (C) The proportion of the significantly gained H3K27ac peaks after 6hr compared to Con with (red) and without (blue) transcription factors signals. (D) The box plot showing the expression of adjacent genes of TF-bound or unbound H3K27ac. (**** : $p < 0.0001$, wilcox test)

6. The profiling of super-enhancer when ER stress-induced

To determine the super-enhancer region, i utilized ROSE algorithm and plotted the line plot showing the total enhancers ordered by magnitude of signal. Con, 6hr, and 16hr samples have 330, 453, and 375 super-enhancers, respectively (Fig. 8A-B). The positions of most of the super-enhancer regions were maintained with a few exceptions (Fig. 8C). However, since this comparison is approved only with the presence or absence of a called super-enhancer, there may be bias from ranking the H3K27ac signal. Thus, i analyzed the average signal intensity at super-enhancers and carried out the differential anlysis. It shows that super-enhancers in Con have higher H3K27ac signal than those in 6hr and the signal was declined in 16hr below Con. This parallels the dynamics of all enhancer activity (Fig. 8D). At this time, the comparison of Con and 6hr showed the most striking difference, and the super-enhancer showed a significant increase in all (Fig. 8E). It illustrates that super-enhancers that have key roles in regulation of gene expression were increased when ER stress-induced.

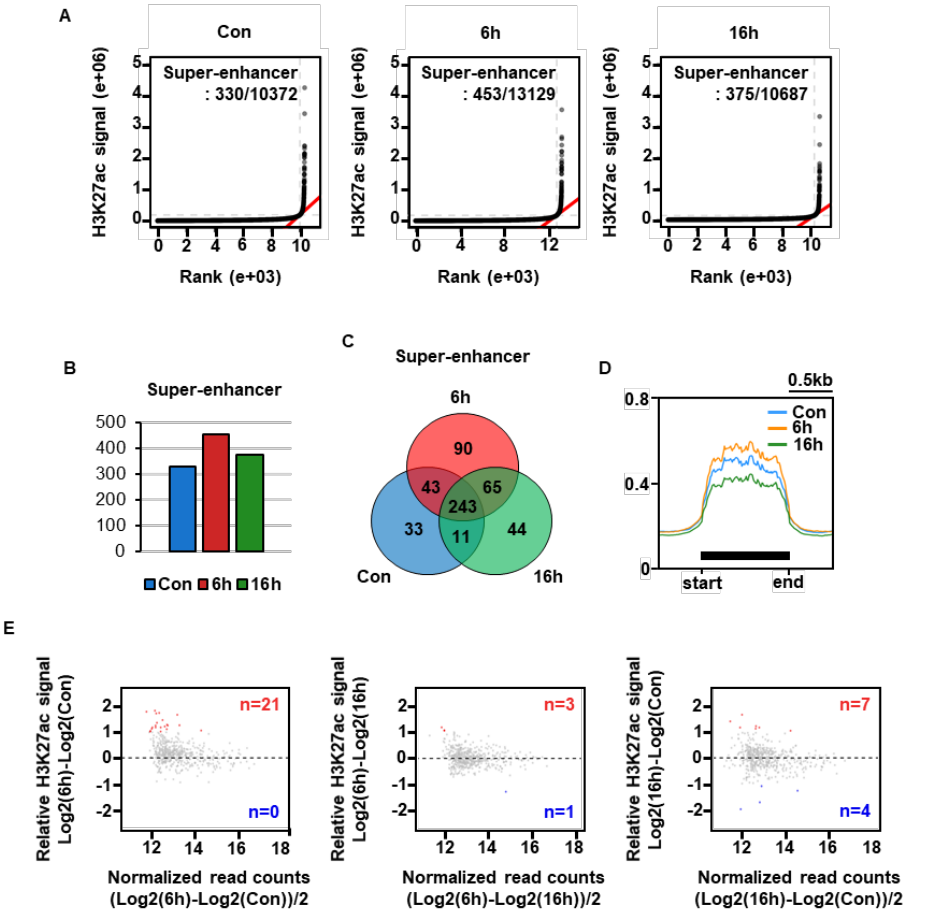

Figure 8. The identification of super-enhancer and the change of its activity
(A) Distribution of stitched enhancer that are ranked by increasing H3K27ac ChIP-seq signal. (B) The bar graph showing the number of super-enhancer (C) The venndiagram of the super-enhancers of all conditions, con, 6hr, and 16h. (D) Average signal of super-enhancer in each sample; Blue : con, Red : 6hr, Green : 16hr. (E) Differential super-enhancer activity; con vs 6hr (left), 6hr vs 16hr (middle), con vs 16hr (right). Red dots means super-enhancers with p value < 0.05 & log2FoldChange > 1 and blue ones are with p value < 0.05 & log2FoldChange < -1.

7. Super-enhancer associated with UPR-related transcription factors affects gene expression during UPR

In consideration of the fact that increased enhancer activity during UPR is related to UPR-related transcription factors, i wanted to determine whether super-enhancers also have a relationship with transcription factors. It was figured out that transcription factors, especially ATF4 and CHOP, were bound to most of the increased super-enhancers (Fig. 9A). At this time, i focused on the comparison with Con and 6hr for the direct effect of ER stress. Then, i pointed out the relative expression of genes that are adjacent to differential super-enhancers. The nearest genes to gained super-enhancers were far more increased than those to constant super-enhancers in 6hr (Fig. 9B). It means that the ER stress affect the transcriptional activity via the change of super-enhancer activity following UPR-related transcription factors binding. For example, NFE2L1 was adjacent genes of gained super-enhancers in 6hr and the increased enhancer region bound ATF4 and CHOP. And, expression of NFE2L1 was increased. Similarly, ARHGEF2 with increased transcription was located on increased super-enhancer region that binds ATF4 and CHOP (Fig. 9C).

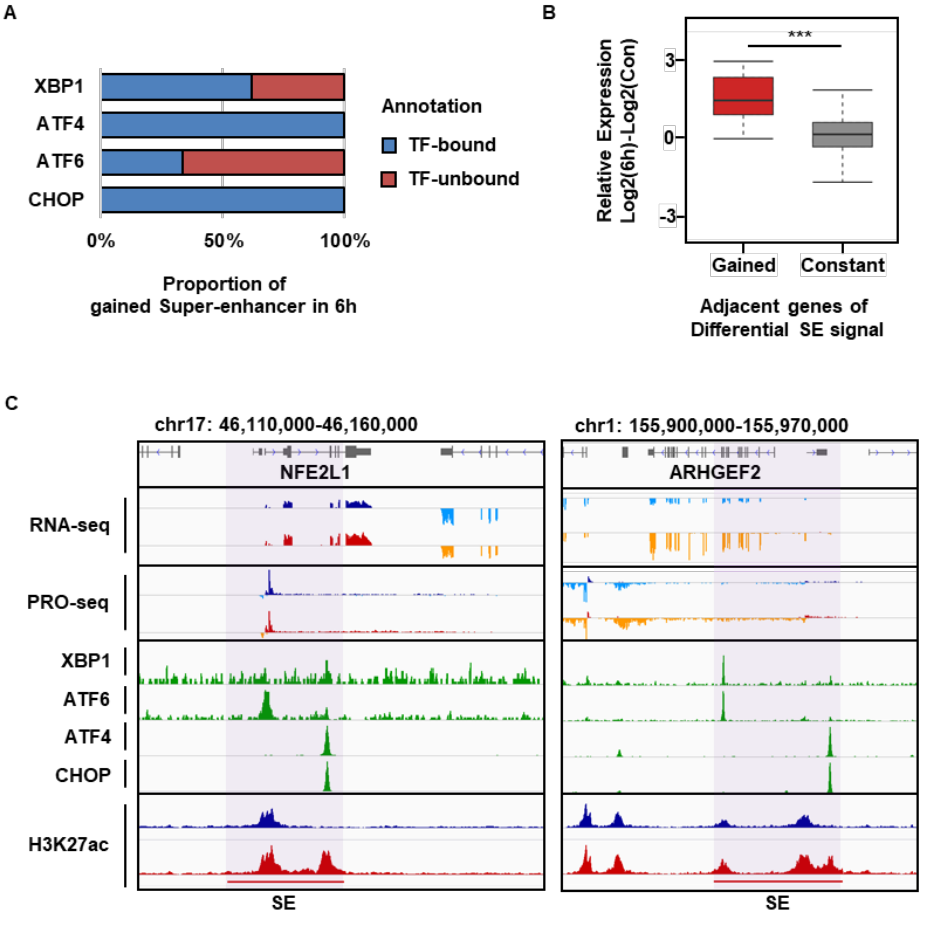

Figure 9. Super-enhancer is related to UPR activator TFs and affect gene expression (A) The proportion of the significantly gained super-enhancer regions in 6hr compared to Con with (red) and without (blue) transcription factors binding. (B) The Box plot illustrates the expression of adjacent genes of gained or constant super-enhancer in 6hr versus con (***: $p < 0.0001$, wilcox test). (C) The representatives of genes with increased expression among adjacent genes of gained super-enhancers at 6 hr compared to con; Blue : Con, Red : 6hr. Purple vertical bars highlight the location of the signal increased super-enhancer region whose adjacent genes are up-regulated in 6hr.

8. The identification of enhancer RNA and its change under ER stress

PRO-seq is a technique that provides a profile of nascent RNA transcripts, unlike RNA-seq. I explored PRO-seq data for identification of eRNA and to come out the changes of enhancer RNA activity during UPR. Firstly, the PRO-seq data has good reproducibility, and captures both coding and non-coding transcripts well (Fig. 10A-B). While steady state RNA levels and immediate transcriptional activity have slightly different meanings, it is expected that both tend to change is simliar under stimulus induction. The box plot shows that the relative expression from RNA-seq and that of PRO-seq per gene. The trend line has a positive slope and there were only few genes with change values of opposite signs between PRO-seq and RNA-seq (Fig. 10C). Moreover, the PRO-seq signal of each gene divided into four categories according to the RNA-seq signal strength was calculated. The genes with most strong RNA-seq signal, 0~25% category, had many transcripts in the PRO-seq data. As the RNA-seq signal decreases, the PRO-seq signal also decreases, showing that the two parallel similarly (Fig. 10D). Based on the observation that ER stress greatly affect enhancer landscape, the identification of transcripts on enhancer, eRNA, was needed as one of key regulatory elements. Here, eRNA was briefly defined as a transcript produced by the enhancer that located on the non-protein-coding region. It has been previously revealed that UPR-related transcription factors affect the increase in enhancer activity. I verified that its transcriptional activity also inclines according to increased enhancer signal via transcription factor binding (Fig. 10E-F). There are representatives of the nearest genes to gained enhancers with increased transcripts in 6hr compared to Con through UPR-related transcription factors that are located in intergenic regions when ER stress induction (Fig. 10G). It indicates that UPR-related transcription factors, especially ATF4 and CHOP, increase enhancer signal and eRNA activity, and regulate the expression of target genes.

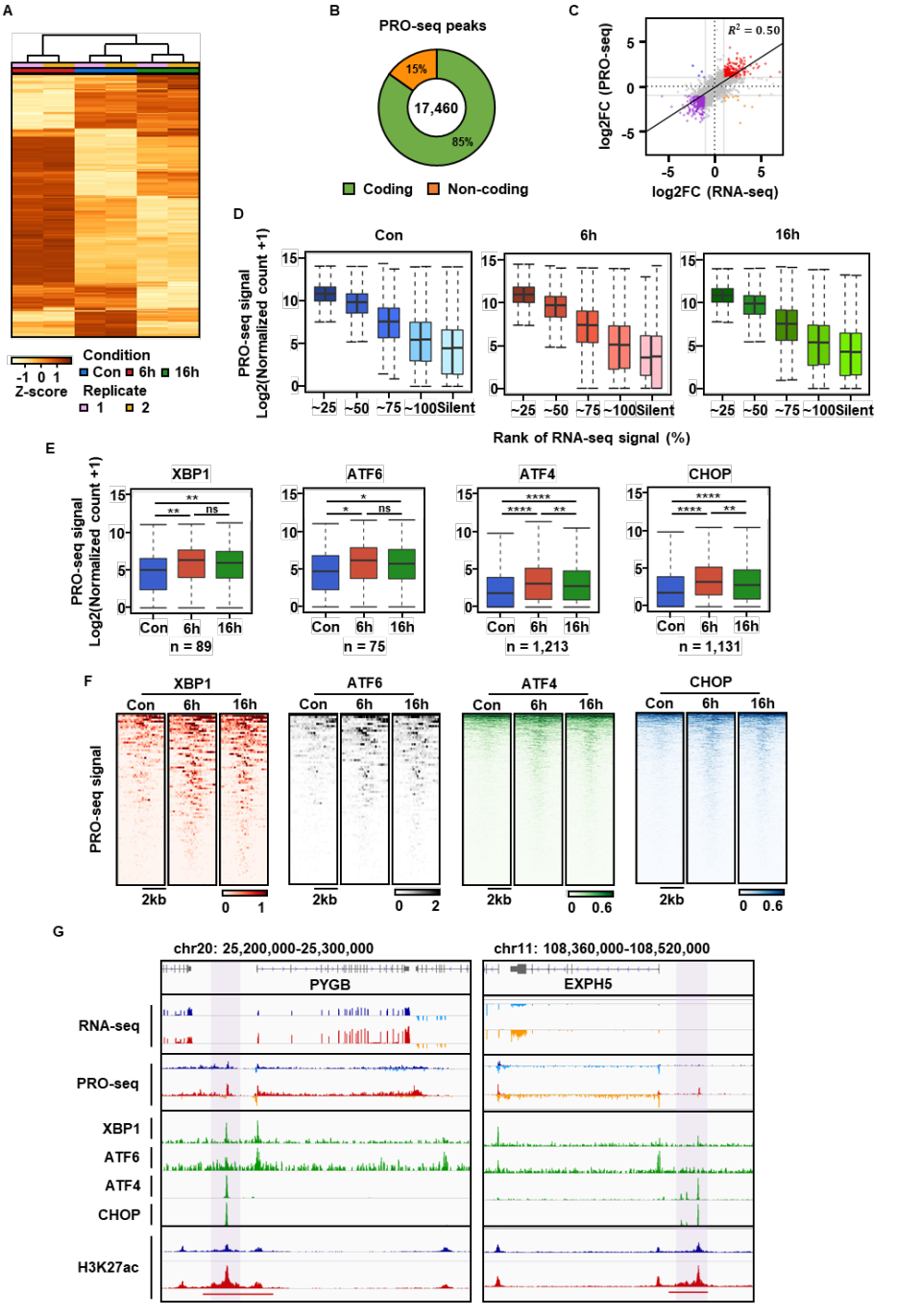

Figure 10. The expression of non-coding genes with enhancer activity (A) The heatmaps showing the PRO-seq signal in Con, 6hr, 16hr. (B) The genomic distribution of PRO-seq peaks; green : peaks in coding regions, orange : peaks in non-coding regions. (C) The scatter plot showing that log2FoldChange of RNA-seq and that of PRO-seq per gene; log2FoldChange is a result from comparison with Con and 6hr. (D) The box plot representing the PRO-seq signal of genes in each quarter of RNA-seq signal; ~25 : genes with the highest 0~25% intensity, ~50 : genes with the highest 25~50% intensity, ~75 : genes with the highest 50~75% intensity, ~100 : genes with the highest 75~100% intensity when ordered by RNA-seq signal, Silent : genes with zero RNA-seq signal. (E-F) The box plots(E) and heatmaps(F) showing that PRO-seq signals at each TF- bound gained H3K27ac peaks in 6h versus Con that not located in coding genes. (ns : not significant, * : $p < 0.1$, ** : $p < 0.01$, **** : $p < 0.0001$, wilcox test) (G) The IGV snapshot of examples that TF-bound intergenic gained enhancers with increased transcripts regulate the adjacent gene expression; Blue : Con, Red : 6hr. Purple vertical bars highlight the location of the putative eRNA with increased enhancer signal whose nearest genes are up-regulated in 6hr.

IV. DISCUSSION

A stressful environment in the endoplasmic reticulum makes the cells activate the unfolded-protein response (UPR) for proteostasis. Three main signaling pathways which start from IRE1a, PERK and ATF6a were well-known, and target genes associated with apoptosis, protein-folding and translation are regulated. Although the molecular mechanism of UPR has been researched in detail, the epigenetic dynamics when ER stress-induced are not still well-defined. The epigenetic studies on chromatin state, 3D structure, and various regulatory elements are needed to understand the regulation of target genes. I analyzed several omics data to examine the epigenetic features that may affect the change of transcription during UPR.

All sequencing data except RNA-seq were performed using two biological replicates of HCT116 cell lines. The transcriptional change when tunicamycin was treated is shown using three replicates of RNA-seq data. The up-regulated genes at the 6hr time point versus Con which are likely to involve in UPR have several patterns that were associated with different GO terms.

The chromatin states such as histone modifications and chromatin accessibility have mostly continued the same whether or not ER stress was applied. However, significantly differential accessible regions were detected at specific regions. And, i discovered that ATF3, ATF4, CHOP have an increase of transcriptional activity with higher motif enrichment in gained accessible regions than lost.

Unlike other histone modifications, the enhancer landscape represented by H3K27ac ChIP-seq was changed significantly. Average enhancer signals are up-regulated at the 6hr time point after tunicamycin treatment and have a reduction in 16hr sample below the level of Con. I clustered enhancer activity into six clusters by the normalized counts of H3K27ac signals. Although most of the H3K27ac peaks have increased signals, disparate patterns emerged. For example,

there was a difference in whether the reduction from 6hr to 16hr was greater than the increase from Con to 6hr despite the same trend to signals being increased in 6hr. For identification of the putative causes which affect enhancers, motif enrichment analysis was performed using the main three clusters of enhancer signals. ATF4 and CHOP are enriched at cluster 1 compared to cluster 2 as well as have up-regulated expression when ER stress-induced. The result demonstrated that ATF4 and CHOP can be a putative transcription factors to influence the enhancer activity.

The well-known transcription factors related to UPR are discovered to have an effect on the dynamics of enhancer landscape and partial accessibility change. Thus, ChIP-seq was processed with the transcription factor antibodies to identify their binding profiles. They are confirmed to have increased occupancy and binding intensity by the treatment of tunicamycin. Among them, XBP1 and ATF6 work in promoter-centric manners while ATF4 and CHOP have enhancer-centric features for binding the genome. It means that XBP1 and ATF6 regulate gene expression more directly in promoters, and ATF4 and CHOP affect transcription indirectly via specific methods such as looping.

The contribution of well-studied transcription factors in the UPR pathway on the enhancer landscape was revealed by various methods. The results show that all transcription factors, for example, XBP1, ATF6, ATF4, CHOP were related to the enhancer signals. Especially, ATF4 and CHOP are highly bound on the gained H3K27ac signals in 6hr, indicating that they may affect the strengthen of enhancer landscape when ER stress-induced. Furthermore, i suggest the relationship among UPR-related transcription factors, enhancer activity, and target gene expression provided that the adjacent genes to TF-bound enhancers have more increased transcription than those to TF-unbound enhancers.

The super-enhancer which plays an important role in regulating target genes was identified using ROSE algorithm. Likewise total enhancers, super-enhancers have increased in 6hr and decreased in 16hr compared to Con. Gained super-enhancers

by ER stress were associated with UPR-related transcription factors, especially ATF4 and CHOP. Also, it is shown that they affect gene expression in an increasing direction.

The non-coding RNA in the enhancer region, called eRNA, is one of the key regulatory elements. Identification of putative eRNA was achieved and its signal in gained enhancer by ER stress was analyzed. The results indicate that the enhancer activity and its transcriptional activity were increased via transcription factor binding, modulating the expression of target genes during UPR.

In summary, it has been reported that UPR plays an important role in the maintenance of ER homeostasis and protein function. However, how epigenetic profiles appear and what effect they are on gene expression are still unknown. Here, these results demonstrate that chromatin states except for H3K27ac signal are maintained throughout the genome. ER stress drives the dynamics of the enhancer landscape towards increasing. The strengthened enhancer activity, also super-enhancer activity are involved with UPR-related transcription factors such as XBP1, ATF6, ATF4 and CHOP, regulating the expression of target genes. In addition, i showed the potential contribution of increased putative eRNA transcription in intensified enhancer regions in the coordination of the gene expression.

In this study, i defined the adjacent genes as the target genes. Some enhancers present in promoters are more likely to affect adjacent genes, whereas the intergenic or intronic enhancers can modulate the distal target genes as well as nearest genes. Considering enhancers characteristics, it deserves further investigation using high-throughput sequencing for chromatin interactions to verify the certain target genes from enhancer-promoter interaction.

In conclusion, our results provide evidence of epigenetic mechanisms in response to ER stress. For example, transcriptional activities of target genes are collaboratively regulated via interaction with enhancers and their eRNA which are associated with specific proteins.

V. CONCLUSION

The disrupted proteostasis by ER stress is the one pathogenesis of various diseases, indicating that the study of UPR is necessary to develop therapeutic avenues. The Gene expression process in eukaryotes is precisely regulated by highly complex and coordinated systems. There is a need to explore the relationship between transcriptional activity and chromatin features during UPR. Here, i used multiple sequencing data like ChIP-seq, ATAC-seq, RNA-seq and PRO-seq to examine genome-wide profiles about epigenetics which may contribute to UPR-related gene regulation in HCT116 colorectal cancer cells. For the classification of profiles in detail, clustering was processed, which was integrated with other omics data. Genomic accessibility and histone modification states represent the same level under ER stress. However, it is revealed that there is an alteration of the enhancer landscape, which is associated with well-known transcription factors related to UPR. The results show that specific transcription factors that strengthen the enhancer activity contribute to the transcriptional regulation via E-P interactions, not genome-wide chromatin states in response to ER stress. Further, It is plausible that the putative eRNA may play a role in the interaction between the target gene and the enhancer. Our study suggests the de novo mechanism of UPR that the enhancer dynamics, including the change of super-enhancer acitivity, are associated with transcription factors and regulate gene expression with eRNA transcription.

REFERENCES

1. Hetz C, Zhang K, Kaufman RJ. Mechanisms, regulation and functions of the unfolded protein response. Nat Rev Mol Cell Biol 2020;21:421-38.
2. Hetz C. The unfolded protein response: controlling cell fate decisions under ER stress and beyond. Nat Rev Mol Cell Biol 2012;13:89-102.
3. Wang M, Kaufman RJ. The impact of the endoplasmic reticulum protein-folding environment on cancer development. Nat Rev Cancer 2014;14:581-97.
4. Hartl FU, Bracher A, Hayer-Hartl M. Molecular chaperones in protein folding and proteostasis. Nature 2011;475:324-32.
5. Todd DJ, Lee AH, Glimcher LH. The endoplasmic reticulum stress response in immunity and autoimmunity. Nat Rev Immunol 2008;8:663-74.
6. Stillman B. Histone Modifications: Insights into Their Influence on Gene Expression. Cell 2018;175:6-9.
7. Rice JC, Allis CD. Histone methylation versus histone acetylation: new insights into epigenetic regulation. Curr Opin Cell Biol 2001;13:263-73.
8. Karlic R, Chung HR, Lasserre J, Vlahovicek K, Vingron M. Histone modification levels are predictive for gene expression. Proc Natl Acad Sci U S A 2010;107:2926-31.
9. Plank JL, Dean A. Enhancer function: mechanistic and genome-wide insights come together. Mol Cell 2014;55:5-14.
10. Furlong EEM, Levine M. Developmental enhancers and chromosome topology. Science 2018;361:1341-5.
11. Schoenfelder S, Furlan-Magaril M, Mifsud B, Tavares-Cadete F, Sugar R, Javierre BM, et al. The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. Genome Res 2015;25:582-97.
12. Misteli T. The Self-Organizing Genome: Principles of Genome Architecture and Function. Cell 2020;183:28-45.
13. Schoenfelder S, Fraser P. Long-range enhancer-promoter contacts in gene expression control. Nat Rev Genet 2019;20:437-55.
14. Siegel RL, Miller KD, Goding Sauer A, Fedewa SA, Butterly LF, Anderson JC, et al. Colorectal cancer statistics, 2020. CA Cancer J Clin 2020;70:145-64.

15. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. CA: A Cancer Journal for Clinicians 2021;71:7-33.
16. Xie YH, Chen YX, Fang JY. Comprehensive review of targeted therapy for colorectal cancer. Signal Transduct Target Ther 2020;5:22.
17. Walter P, Ron D. The unfolded protein response: from stress pathway to homeostatic regulation. Science 2011;334:1081-6.
18. Avril T, Vauleon E, Chevet E. Endoplasmic reticulum stress signaling and chemotherapy resistance in solid cancers. Oncogenesis 2017;6:e373.
19. Spaan CN, Smit WL, van Lidth de Jeude JF, Meijer BJ, Muncan V, van den Brink GR, et al. Expression of UPR effector proteins ATF6 and XBP1 reduce colorectal cancer cell proliferation and stemness by activating PERK signaling. Cell Death Dis 2019;10:490.
20. Park PJ. ChIP-seq: advantages and challenges of a maturing technology. Nat Rev Genet 2009;10:669-80.
21. Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ, et al. Genomic maps and comparative analysis of histone modifications in human and mouse. Cell 2005;120:169-81.
22. Zhang T, Zhang Z, Dong Q, Xiong J, Zhu B. Histone H3K27 acetylation is dispensable for enhancer activity in mouse embryonic stem cells. Genome Biol 2020;21:45.
23. Klemm SL, Shipony Z, Greenleaf WJ. Chromatin accessibility and the regulatory epigenome. Nat Rev Genet 2019;20:207-20.
24. Frankish A, Diekhans M, Ferreira AM, Johnson R, Jungreis I, Loveland J, et al. GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Res 2019;47:D766-D73.
25. Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, et al. The Human Transcription Factors. Cell 2018;172:650-65.
26. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013;29:15-21.
27. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 2011;12:323.
28. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550.
29. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun 2019;10:1523.
30. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie

2. Nat Methods 2012;9:357-9.
31. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 2010;26:589-95.
32. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25:2078-9.
33. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol 2008;9:R137.
34. Ramirez F, Dundar F, Diehl S, Gruning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res 2014;42:W187-91.
35. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 2010;38:576-89.
36. Loven J, Hoke HA, Lin CY, Lau A, Orlando DA, Vakoc CR, et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell 2013;153:320-34.
37. Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell 2013;153:307-19.

ABSTRACT(IN KOREAN)

대장암세포에서 튜니카마이신에 의해 유도되는 미접힘 단백질 반응 동안 일어나는 역동적 전사 및 후성유전학적 변화

< 지도교수 김 형 표 >

연세대학교 대학원 의과학과

김 수 경

mis/un-folding 단백질의 비정상적인 축적은 세포 스트레스를 유발한다. ER 항상성을 유지하기 위해 UPR(Unfolded protein response) 신호 전달 경로가 유도되고 이는 인간의 다양한 질병, 특히 암의 발병기전과 관련이 있다. 현재, PERK, IRE1, ATF6에 의해 매개되는 비접힘 단백질 반응(UPR) 신호 전달 경로의 분자적 메커니즘은 잘 확립되어 있다. 그러나, 염색질 구조, 히스톤 변형 및 말단 조절 요소가 ER 스트레스 하에서 단백질 항상성을 조절하는지 여부와 그 방법에 대해서는 덜 이해되고 있다. 이 연구에서 저는 ER 스트레스 하에서 인핸서 활동의

동적 변화를 이해하기 위해 다중 오믹스 시퀀싱 데이터를 사용했다. 전사체 및 후성유전체 분석을 수행하여 단백질 코딩 유전자의 발현 패턴과 게놈 전체의 인핸서 랜드스케이프를 조사했다. 이러한 데이터 통합 분석의 결과, ER 스트레스 하에서 적절한 유전자 발현 프로그램을 확립하는 데 기여하는 염색질 구조 및 인핸서 활동의 동적 변화를 밝혀냈다.

핵심되는 말 : ER 스트레스, 튜니카마이신, 크로마틴, 인핸서