博士論文

Contribution of RNA degradation and post-transcriptional regulation to gene expression

(RNA 発現に対する RNA 分解および転写後制御の寄与)

前川　翔

Index

Introduction

Gene expression is determined by the eventual RNA abundance, that is the concentration of RNA molecule in a given cell at a given time. It is controlled by the rates of RNA generation, RNA processing and RNA degradation (Ben-Tabou de-Leon and Davidson, 2009; Kim et al., 2009; Komili and Silver, 2008; Wang et al., 2007). Majority of genome-wide studies have focused on the transcription activation to understand the mechanism of gene expression (Bujold et al., 2016; Roadmap Epigenomics Consortium et al., 2015; Stamatoyannopoulos et al., 2007; The ENCODE Project Consortium et al., 2012). There have been efforts to estimate the abundance of RNA through the epigenomics data (Dong et al., 2012; Karlic et al., 2010; Wang et al., 2012).

Transcription is regulated through the structure of chromatin and transcription factor binding to the DNA elements. Chromatin states are regulated through chromatin modifiers, which move the histones along the DNA to expose DNA elements to transcription factors (TF), that can either enhance or repress the target gene. Chromatin states have a 'histone code', which is a covalent mark on the histone tail that can infer the transcriptional state of a given loci. These histone codes include histone H3 lysine 4 tri-methylation (H3K4me3), which is an active mark of transcription and it is often observed around active transcription start sites (TSS) (Koch et al., 2007; Mikkelsen et al., 2007). Transcription factors can either be general transcription factors or sequence-specific transcription factors. General transcription factors include RNA polymerase II (Pol II) that is associated with actively transcribed regions (Roeder, 1996) and sequence-specific transcription factors include HIF-1 (hypoxia inducible factor 1), that bind to hypoxia response element (HRE) to activate genes involved in hypoxia response.

RNA degradation occurs through complex mechanisms, where more than 30 families of ribonucleases from more than 60 proteins are involved. They can be classed as 5'-3' and 3'-5' exonucleases and exonucleases as molecular basis of action and they can be either be nuclear

RNA degradation or cytoplasmic RNA degradation depending on the location of action (Arraiano et al., 2013; Garneau et al., 2007; Stoecklin and Muhlemann, 2013). In the nucleus, pre-mRNAs are capped and the pre-RNAs are spliced co-transcriptionally to remove the introns and followed by the addition of poly A tail to protect both the 5' and 3' ends of mRNAs. Errors in the mRNA processing lead to the degradation of mRNA by surveillance pathways, that ensures mRNAs are correctly processed. One of the key players in the surveillance is the nuclear exosome (Chlebowski et al., 2013), which is a multi-subunit complex that include RRP46 (EXOSC5). The complex quality controls of the 3' ends, where RNAs that have abnormal polyA tails are degraded (Porrua and Libri, 2013). In the cytoplasm, RNA decay occurs through multiple pathways, where the mRNA decay starts by the deadenylation of the 3' poly(A) tail (Godwin et al., 2013; Wahle and Winkler, 2013), where it is mediated by CCR4-NOT complex and followed by 3'-5' exonuclease degradation through the cytoplasmic exosome complex. In addition to the deadenylation of the 3' poly(A) tail, the 5'-cap is removed by DCP2/DCP1 decapping enzymes, followed by a 5'-3' exonuclease called XRN2 (Nagarajan et al., 2013). Another surveillance mechanisms include the nonsense-mediated decay (NMD), where the major player is UPF1 (Imamachi et al., 2012b; Schweingruber et al., 2013). It normally recognizes and selectively degrades RNA transcripts that have premature stop codons. In addition to the quality control mechanisms, there are reports that suggest the role of UPF1 in mediating approximately 3 to 20% of all transcripts, regardless of the presence or the absence of the PTC. STAU1 recognize double stranded RNA, and is a typical example of RNA binding protein. There is a report to suggest that STAU1 regulate approximately 1% of *bona fide* mRNA (Kim et al., 2007).

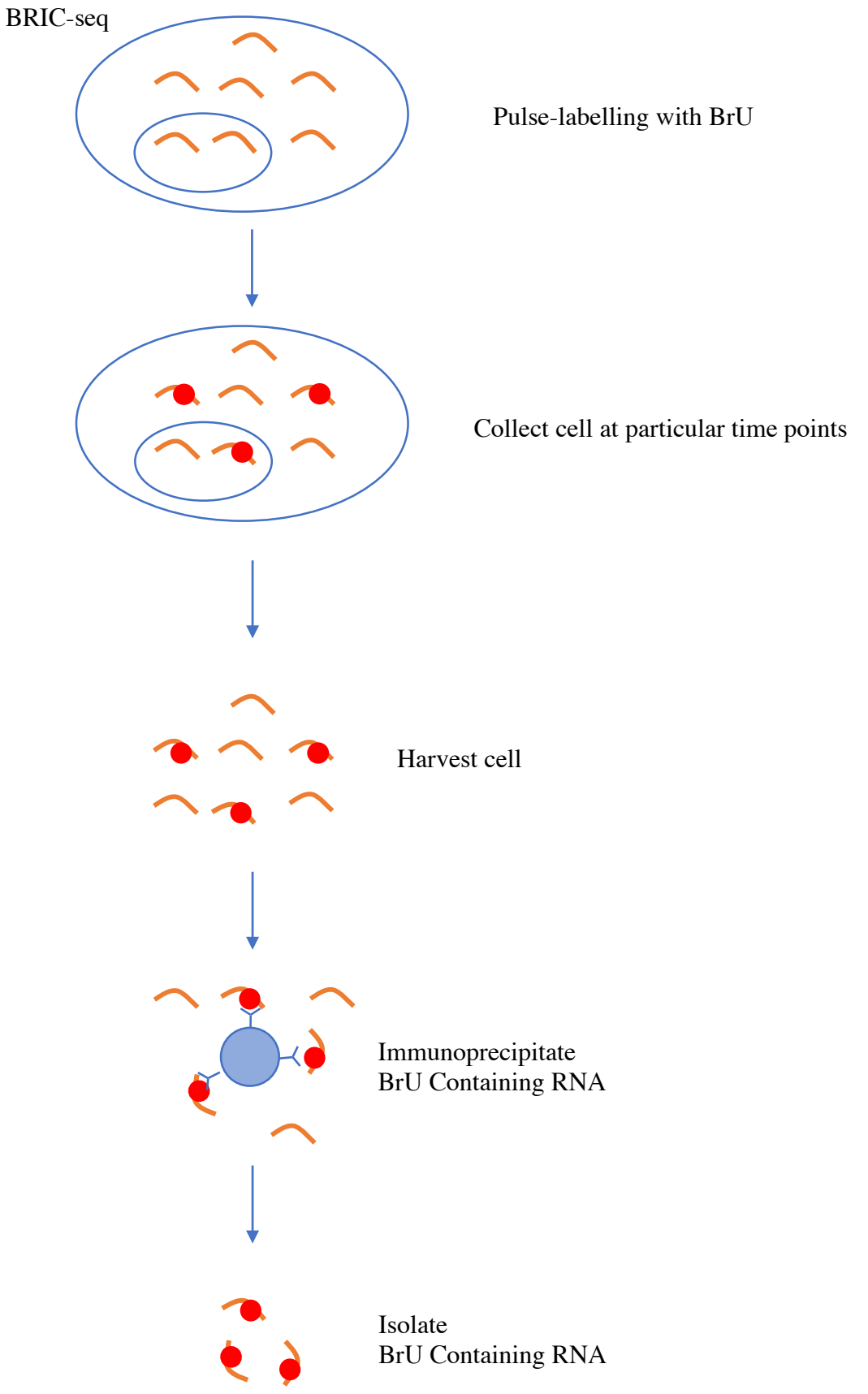
In theory, cells can regulate their RNA abundance through changes in the transcription and decay (Ben-Tabou de-Leon and Davidson, 2009). In previous reports, they have suggested that the majority of RNA abundance is regulated at the transcriptional level (Schwanhäusser et

al., 2011), and that RNA decay 'sharpens' the response (Rabani et al., 2011). However, in majority of reports the RNA decay is indirectly inferred using the data from RNA generation or they use transcriptional inhibition and assay the RNA decay, which limits the biological relevance of the assay, as transcriptional inhibitors are toxic to cells (Tani et al., 2012b) and indirect inference by definition cannot directly measure RNA decay and have various assumptions. In addition 4-thiouridine and 5-ethyl uridine are toxic when used as a metabolic label at required concentrations, thus there is an advantage in BrU based labelling (Tani et al., 2012b).

In many cancers, cells proliferate aggressively without adequate oxygen supply through the blood vessels, which lead to the low oxygen potential towards the centre of tumours. Numerous cancer cells still proliferate even in low oxygen potential, also known as the state of hypoxia (usually defined to be 0.02% to 3% of O_2), by gene expression changes in response to hypoxia (Keith and Simon, 2007; Pouysségur et al., 2006; Semenza, 2010). The molecular response to hypoxia is mediated through hypoxia-inducible factors (HIFs), that directly bind to DNA to activate downstream genes, where the protein levels increase dramatically with the decrease in the O_2 concentration (Semenza, 2014). HIFs are heterodimers with alpha and beta subunits, and the alpha subunit is regulated through the protein stability, where the subunit is unstable in normoxia and stable in hypoxia. This is achieved by hydroxylation of the proline residue by prolyl hydroxylase domain protein (PHD), which leads to von Hippel-Lindau protein (VHL) - dependent ubiquitination leading to proteosomal decay in normoxia (Semenza, 2014). However, HIFs does not regulate all of the transcription response in response to hypoxia with vascular endothelial growth factor (VEGF) only blocked partially in with HIF1 knockdown in hypoxia (Mizukami et al., 2005), suggesting an alternative mechanism of action in response to hypoxia.

There are studies to suggest that HIF-1 knockout is not sufficient to induce VEGFA up-regulation, and that other factors are necessary (Mizukami et al., 2005). There have been studies into the role of AU-rich element in VEGFA regulation through HuR; however, a systematic analysis of the RNA decay in this system was far from complete (Kurosu et al., 2011; Levy, 1998).

In this thesis, I will discuss on two topics. In the first chapter, I will discuss on the contributions of RNA decay in the RNA abundance in an unstressed state. In addition, I will discuss on the role of RNA decay factors in mediating the RNA stability to affect the RNA abundance (Maekawa et al., 2015). In the second chapter, I will discuss on the effects of hypoxia on RNA decay and its effects on the RNA abundance.



Graphic illustration of
BRIC-seq protocol

Chapter One: Regulation of the eventual RNA abundance in the steady-state

Rationale

The contribution of RNA decay to regulate the RNA abundance is far from elusive. To address this issue, in this chapter, I address the possible contribution of RNA decay on RNA abundance in the steady-state. Firstly, I analysed the correlation between active chromatin marks, and the eventual RNA abundance in HeLa cell, that is the concentration of RNA in a cell. This was done by comparing the histone intensities of histone H3K4me3 and RNA polymerase II ChIP-seq and RNA abundance estimated from RNA-seq. Secondly, I analysed the RNA decay profile by BRIC-seq for HeLa cell and compared against the ChIP-seq and RNA-seq data. Through the comparison with ChIP-seq, RNA-seq and BRIC-seq, I was able to characterise the types of genes that undergo regulation through RNA decay.

Material and Methods

Cell culture

HeLa cells were cultured in high glucose Dubelcco's Modified Eagle's Medium (D-MEM: Wako) supplemented with 10% Fetal bovine serum (FBS: Gibco) and 1% antimyotic-antibiotic (Gibco) in a humidified incubator at 37ºC in 5% CO_2.

Chromatin immunoprecipitation sequencing (ChIP-seq)

ChIP-seq was conducted as previously conducted in (Kanai et al., 2011; Tanimoto et al., 2011). In brief, cells were grown to 5×10^7 and crosslinked in 0.5% formaldehyde solution (50mM 4-(2-hydroxyethyl)-1-piperazineethanesulphonic acid (HEPES)-KOH pH7.5, 100mM NaCl, 1mM ethylendiaminetetraacetic acid (EDTA), 0.5 mM ethylene glycole tetraacetic acid (EGTA), 5.5% formaldehyde) by incubation for 10 minutes at room temperature. Crosslinking was quenched by the addition of glycine (150mM). Cross-linked cells were washed with phosphate buffered saline (PBS) and harvested using a scraper, and the pellet was frozen at -80ºC to be stored.

The frozen pellet was defrosted and resuspended in 5ml of Lysis buffer 1 (50mM HEPES-KOH pH7.5, 140mM NacCl, 1mM EDTA, 10% glycerol, 0.l5% NP-40 and 0.25% Triton X-100). The cell lysate was incubated on ice for 10 minutes and centrifuged at 1500 rpm for 5 minutes at 4ºC and the supernatant was discarded. The cell pellet was resuspended in 5ml Lysis buffer 2 (10 mM Tris-HCl, pH8.0, 200 mM NaCl, 1mM EDTA and 0.5 mM EGTA) and incubated on ice for 10 minutes. The lysate was centrifuged at 1500 rpm for 5 minutes at 4ºC and the supernatant was discarded. The cell pellet was resuspended in 1ml of lysis buffer 3 (10mM Tris-HCl, pH 8.0, 100mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na-deoxycholate and 0.5% N-lauroylsarcosine) and kept on ice. The supernatant was sonicated with a sonicator (Tomy Seiko) at maximum power with 16 cycles of 30 seconds

followed by 90 seconds of rest. 100μl of 10% Triton X-100 was added to the lysate, transferred to a new 1.5 ml tube and centrifuged for 10 minutes at 14,000 rpm at 4ºC. The supernatant was collected and transferred to a new 1.5 ml tube and 50μl of the lysate was aliquoted as a whole cell extract (WCE) DNA and stored at -20ºC.

To conduct the immunoprecipitation, 100μl of protein A or G magnetic beads were washed on ice with 1ml blocking buffer (0.5% Bovine serum albumin (BSA) in PBS) three times using the magnetic stands. 250μl of blocking buffer and 10μg of the antibodies were incubated at 4ºC overnight on rotation. The antibodies-beads were washed 3 times with 1ml of blocking buffer and resuspended in 100μl blocking buffer on ice. 100μl antibody-bead mixture was added to the lysate and incubated at 4ºC overnight with rotation. The immunoprecipitated lysate was placed on the magnetic stand to remove the lysate and washed 9 times with 1ml wash buffer (50mM HEPES-KOH pH 7.5, 500mM LiCl, 1mM EDTA, 1% NP-40, 0.7% Na-deoxycholate) on the magnetic stand. 1ml TE buffer with 10μl of 5M NaCl was added to the washed immunoprecipitated lysate and the supernatant was removed. The lysate is centrifuged for 3 minutes at 1000 rpm at 4ºC and the supernatant was removed. To elute, 200μl of the elution buffer (50mM Tris-HCl pH8.0, 10mM EDTA, 1% sodium dodecyl sulphate (SDS)) was added to the lysate and incubated at 65ºC for 15 minutes. The sample was centrifuged for 1 minute at 14000 rpm at room temperature and 200μl of the supernatant was transferred to a fresh 1.5ml tube. The samples were incubated overnight at 65ºC and to reverse cross-link. In addition, 150 μl of the elution buffer was added to the WCE-DNA samples and incubated overnight.

Both the immunoprecipitated samples and WCE samples are diluted with 200μl TE buffer. 8μl of RNase A is added and incubated at 2 hours at 37ªC, followed by the addition of 4μl of Proteinase K and 7μl $CaCl_2$ with 30-minute incubation at 55ºC. Following the RNase

and Proteinase K digestion, phenol/chloroform (1:1) purification was conducted by the addition of 400μl of phenol/chloroform followed by centrifugation for 10 mintues at 14000 rpm at room temperature. The sample in the liquid phase was transferred to a new tube and the DNA was isolated using the addition of 16μl of 5M NaCl, 1μl ethachinmate and 880μl of 100% ethanol. The samples were cooled at -80ºC and centrifuged for 15 minutes at 14000 rpm at 4ºC. The pellets were washed with 100μl 70% ethanol and resuspended in 35μl distilled water. The quality of the samples was verified using qPCR.

The antibodies used were the following: anti-H3K4me3 (Abcam, ab1012), anti-H3K27Ac (Abcam, ab4729), anti-H3K27me3 (Abcam, ab6002), anti-H3K36me3 (Abcam, ab9050), anti-RNA polymerase II (Abcam, ab817). The libraries were made using Tru-seq ChIP-seq (Illumina) and sequenced on Illumina Genome Analyzer IIx or HiSeq 2000 with 36bp single read.

With the analysis, the reads were base-called using CASAVA (Illumina) according to the manufacturer's instructions and mapped to the human genome (hg19) using Eland (Illumina) according to manufacturer's instructions. The peaks were called using MACS (v1.8.3) at default settings. Sequencing statistics are shown in supplementary table 5.

RNA sequencing (RNA-seq)

RNA was harvested from approximately 1×10^6 cells using 1ml TRIzol reagent (Invitrogen) and purified as described in the manufacturer's protocol. Approximately 1μg of RNA was used to create Illumina RNA-seq sequencing library using Tru-seq (Illumina) and sequenced on the Illumina HiSeq 2500 according to the standard protocol.

The sequenced tags were base-called using the standard manufacturer's protocol and aligned to the rRNA genes were removed by Bowtie 2 (v2.1.0) (Langmead and Salzberg,

2012). Reads unmapped to rRNA gene sequences were mapped to the human genome (hg19) using Tophat2 (Kim et al., 2013; Trapnell et al., 2009) using RefSeq (Pruitt et al., 2012) and lincRNA databases (Cabili et al., 2011) (downloaded on 2nd July 2013) as gene models, where reads were only mapped if compatible. For the enhancer RNA (eRNA) analysis, enhancer regions were defined using ChIP-seq peaks using H3K4me3 and H3K27Ac HeLa data from the ENCODE project, without any overlap within 1.5kb of the entire length of gene body. To map the eRNA reads, Tophat (2.0.8) was used to map the reads without a reference gene model. For RefSeq mRNA, lincRNA and eRNA, reads were quantified using Cufflinks (Trapnell et al., 2010) without generating novel isoforms. The transcript with the highest RNA expression was used as a reference isoform for which the transcription start site (TSS) was based. The sequencing statistics were shown in table XX.

Bromouridine-immunoprecipitation pulse chase sequencing (BRIC-seq)

Bromouridine immunoprecipitation pulse chase sequencing (BRIC-seq), was performed as conducted previously (Imamachi et al., 2014; Tani et al., 2012b). HeLa cells were incubated with 150 μM BrU containing high glucose D-MEM (Wako) supplemented with 10% FBS and 1% antimyotic-antibiotic in 5% CO_2 humidified incubator at 37ºC for 24 hours. Following the incubation with BrU, media containing BrU is removed, and the cells are washed with D-PBS (Wako) three times followed with the addition of BrU free media. The cells are incubated for 15 minutes to start the pulse-chase at time 0h and harvested at indicated time points. For the sample without siRNA, the time points collected were 0 min, 15 min, 30 min 45 min, 1 h, 1.5 h, 2 h, 3 h, 4 h, 6 h, 8 h, 10 h and 12 h. For EXOSC5 knock-down samples and its siCTRL, there were five time points at following times: 0 min, 4 h, 8h, 12 h and 24 h. For STAU1 knock-down samples and its siCTRL, there were 11 times point at the following times: 0 min, 15 min, 45 min 75 min, 105 min, 165 min, 225 min, 345 min, 465 min, 585 min and 705 min. siUPF1 and its control data were obtained from the previous study (Imamachi et al.,

2012a), where the BRIC-seq time points were: 0 min, 4 h, 8 h and 12 h. Total RNA was isolated using Takara RNAiso Plus (Takara), according to the manufacturer's protocol. Twelve micrograms of RNA of BrU-labelled total RNA were denatured by incubating at 80ºC for 1 minute and added to anti-BrdU mAB-conjugated beads with 2μg of anti-BrdU mAB (clone 2B1, MBL). The antibody-mRNA mixture was incubated at room temperature for 1 h by rotation. Beads were washed with 0.1 % BSA in PBS four times followed by RNA isolation using ISOGEN LS (Nippon Gene) to isolate BrU labelled RNA according to the manufacturer's protocol. The isolated BrU labelled RNA were checked for its quality and quantity by Bioanalyzer 2100 (Agilent Technologies). NGS library preparation was conducted on the BrU labelled RNA using Tru-seq RNA-seq library preparation kit (Illumina), without polyA selection and run on HiSeq 2500 platform (Illumina).

The bioinformatics pipeline follows previously reported methods (Imamachi et al., 2014; Tani et al., 2012a; Tani et al., 2012b). The sequenced reads are base-called in with CASAVA (Illumina) using the manufacturer's protocol. The reads are filtered against rRNA genes using Bowtie2 and filtered reads are mapped (to mRNA, lincRNA and eRNA) as described in the protocol for the RNA-seq. The quantified reads in RPKM are normalized to time = 0 h to calculate the relative abundance and then normalized to GAPDH at respective time points to remove the variation in the immunoprecipitation efficiency. The normalized abundance is then fitted to decay curves of three different models. Model 1 is $y_t = \frac{x_t}{x_0} = e^{-\lambda t}$, where t is the time following the pulse-chase, x_t is the normalised RPKM at time (t), y_t is the relative RNA abundance at time (t) and λ is the decay constant. The half-life for model 1 was calculated by $t_{1/2} = \frac{\log_e(2)}{\lambda}$. Model 2 is $y_t = ae^{-\lambda_1 t} + (1-a)e^{-\lambda_2 t}$, where λ_1 and λ_2 are the different half-lives of two populations, a is the asymptomatic value. The half-life for model 2

was calculated by the following equation $Z_t = (ae^{-\lambda_1 t} + (1-c)e^{-\lambda_2 t})^2$. Half-life is nearly equal to t , when Z_t is the minimal value. To assess the performance of the models, I used Akaike Information Criterion (AIC) to determine the fit relative to the increase in the parameter, where it is calculated by $AIC = 2k - n\log_e(L)$, where k is the number of parameters and L is the residual sum of squares and I chose the model with minimum AIC value.

Acquired data

The siUPF1 knockdown data for RNA-seq, BRIC-seq was used from a previous study (Tani et al., 2012a), with the sequence accession numbers DRA000591 and DRA001215. ENCODE (The ENCODE Project Consortium, 2011) and DBTSS (Yamashita et al., 2012) data were obtained from the URL shown in tables 1-2.

siRNA transfection

siRNA transfection was conducted using Lipofectamine RNAiMAX reagent (Invitrogen) with 10nM as a final concentration according to the manufacturer's protocol. The siRNA sequences are shown in table 3. Cells were harvested at 72 hours following the transfection and the knockdown efficiencies were determined using qRT-PCR and western blot.

qRT-PCR

RNA were reverse-transcribed into cDNA using PrimeScript RT Master Mix (Takara) and the target cDNA were amplified by SYBR Premix Ex Taq II (Takara) according to the manufacturer's protocol. GAPDH was used for normalization. Quantitative real-time reverse transcription PCR analysis was performed using a Thermal Cycler Dice Real Time System (Takara). The oligonucleotides used as primers are listed in table 4.

Western blot

The cells were harvested using RIPA buffer (50mM Tris-HCl, pH7.4, 150 mM NaCl,

5 mM EDTA, 1% NP-40, 1% Na-deoxycholate, 0.1% SDS, 1% proteinase inhibitor cocktail (Sigma-Aldrich)). Proteins from lysates were resolved using 10% SDS-PAGE gel and transferred to a PVDF membrane (GE Healthcare). Membranes were incubated for 1 hour with primary antibodies indicated, followed by incubation with anti-mouse or anti-rabbit secondary antibodies conjugated to horseradish peroxidase (HRP). The blotted protein was imaged by using the ECL Prime plus (GE Healthcare), which uses HRP substrate to allow the chemiluminsecnce to be detected with Luminescent Image Analyzer LAS-4000 (Fujifilm). The following antibodies were used for as primary antibodies: rabbit anti-UPF1 (Abcam), rabbit anti-STAU1 (kindly provided by Dr. Ortín), rabbit anti-EXOSC5 (Sigma-Aldrich), rabbit anti-actin (Sigma-Aldrich, 1978) and rabbit anti-tubulin (MBL).

Integrated Analysis

To assign histone intensities to genes, for H3K4me3 and RNA pol II, peaks that were called by MACS were assigned to genes if any of the peak was present within +/- 1.5kbp of the TSS of the most expressed isoform (determined by Cufflinks). To count the intensity, wig files generated by MACS were used to calculate the number of tags in the given loci. Pearson product-moment coefficient was calculated by using the log-transformed H3K4me3 tags + 1, where 1 was used as a pseudocount to avoid errors generated by log(0) and log-transformed gene expression in RPKM + 1×10^{-6}. Gene ontology data (Ashburner et al., 2000) were used by downloading from the National Center for Biotechnology Information (NCBI) on 8 May 2014. To calculate the enrichment of a particular gene ontology (GO) term, the number of genes that possessed the annotation of a particular GO term was calculated in the sample gene-list and to the total mRNA followed by hypergeometric test to calculate the probably of occurrence through p-value. The *p-values* were normalised for the multiple testing by Benjamini-Hochberg false-discovery rate (Benjamini and Hochberg, 1995) using custom script in R (R Core Team, 2012).

Computational Simulation and Modelling

To conduct the simulation, linear regression was used to predict the eventual RNA abundance. Linear regression was conducted by using number of sequenced tags regardless of the presence of peaks. To calculate the tags for H3K4me3 and H3K27Ac, the frequency of sequenced tags was calculated +/- 1kb of the TSS. For H3K27me3 and H3K36me3, the gene body was used to calculate the ChIP-seq intensities through .wig files generated by MACS.

The tag counts were then log-transformed and z-standardised (mean = 0 and standard deviation = 1) to build a linear model using the following:

$$\text{Model A: mRNA level } \sim b_0 N_{\text{H3K4me3}} + b_2 N_{\text{H3K27Ac}} + b_3 N_{\text{H3K27me3}} + b_4 N_{\text{H3K36me3}} + e$$
$$\text{Model B: mRNA level } \sim b_0 N_{\text{H3K4me3}} + b_2 N_{\text{H3K27Ac}} + b_3 N_{\text{H3K27me3}} + b_4 N_{\text{H3K36me3}} + b_5 \text{half-life} + e$$

Where N is studentised read coverage. mRNA level is log-transformed RPKM+1×10^{-6}, half-life is log-transformed decay constant $\left(\lambda = \frac{\log(2)}{\text{half-life}}\right)$ and e is the residual error.

To simulate the contribution of RNA stability for individual genes, there was an assumption that RNA stability does not play a role to changes in the transcription initiation and the following equation was used for the simulation:

$$\frac{dY}{dt} = b - aY$$
$$\text{when } \frac{dY}{dt} = 0$$
$$Y = \frac{b}{a} \text{RNAseq}_{\text{siCTRL}} = \frac{b}{a_{\text{siCTRL}}}$$
$$b = a_{\text{siCTRL}} \cdot \text{RNAseq}_{\text{siCTRL}}$$
$$\text{RNA half-life: } T_{1/2} = \log(2)/a$$
$$a_{\text{siCTRL}} \cdot \text{RNAseq}_{\text{siCTRL}} = a_{\text{siX}} \cdot \text{RNAseq}_{\text{siX}}$$
$$\therefore \text{RNAseq}_{\text{siX}} = a_{\text{siCTRL}} \cdot \frac{\text{RNAseq}_{\text{siCTRL}}}{a_{\text{siX}}}$$

a and b are constants;

Y : total RNAabundance

RNAseq_y : RNA expression in RPKM

where X : UPF1, EXOSC5 or STAU1

ENCODE Data Analysis

To analyse the ENCODE dataset for ChIP-seq and RNA-seq, cells that possess ChIP-seq data for H3K4em3, H3K36me3 and RNA pol II with RNA-seq data was used for seven cell types (The ENCODE Project Consortium et al., 2012). In addition, the same analysis was conducted for the DLD-1 from DBTSS (Yamashita et al., 2012) (URLs in supplementary tables 1-2). Average enrichment for the H3K4me3 ChIP-seq data were used to compare with their gene expression values.

Results

Correlation between histone epigenome and the transcriptome data

To investigate the contribution of transcription initiation in regulating the eventual RNA, I conducted ChIP-seq for histone H3K4me3 and RNA polymerase II to estimate the transcription initiation and I also conducted RNA-seq to estimate the RNA abundance, which were conducted on the Illumina HiSeq2500. For the analysis of ChIP-seq, I used MACS software (Zhang et al., 2008) to identify the peaks that were statistically significantly enriched with *p*-value of less than 1×10^{-10} in the sample compared to the negative control. In ChIP-seq of H3K4me3, I identified in total of 11116 genes with H3K4me3 peaks (2732, low peak genes 8,384 high peak genes), with 6319 genes that possess both H3K4me3 and pol II, respectively, within 1.5kb of TSS, as shown in the Figure 1. When the ChIP-seq peaks were quantitatively compared against the RNA abundance of the target genes, I observed a positive correlation of R=0.71 (p-value < 2.2x10-16), as shown in the Figure 2a and Figure 2b. Despite the correlation, there are genes that vary from the correlation. In particular, there were 2861 genes that show high H3K4me3 ChIP-seq intensity with less than 10 RPKM in RNA abundance (ChIP+/RNA- in upper-left hand corner of Figure 2b, example in Figure 3b), estimated through RNA-seq and 2897 genes that show low H3K4me3 ChIP-seq intensity, no pol II ChIP-seq peak, with more than 10 RPKM in RNA abundance (ChIP-/RNA+ in lower-right hand corner of the Figure 2b, Example in Figure 3c). Through these comparisons, I identified discrepancies between the ChIP-seq and RNA-seq data.

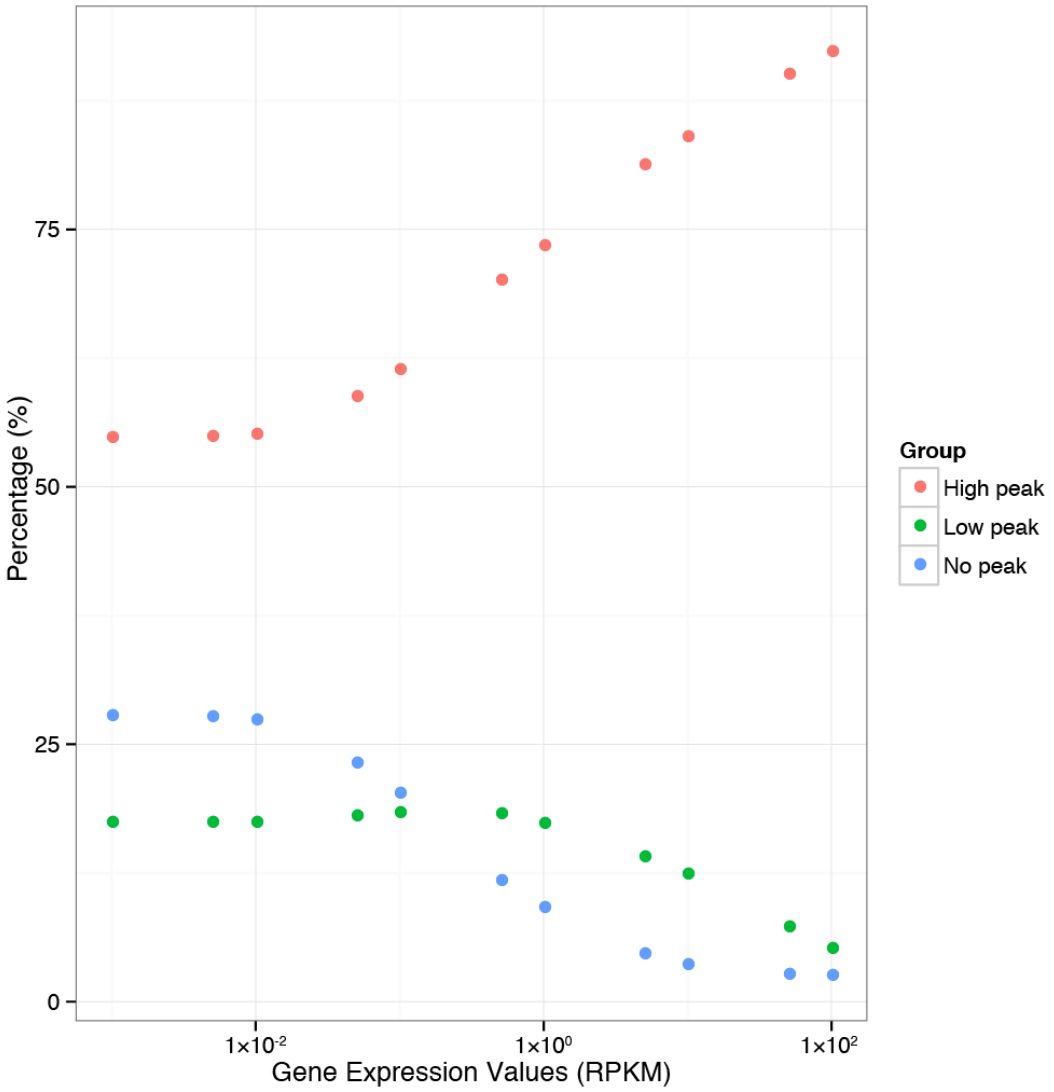


Figure 1
The correlation between the H3K4me3 ChIP-seq peaks and gene expression
The positive correlation is present between the presence of the H3K4me3 peaks near the TSS and the gene expression values
(Maekawa et al., 2015)

a b

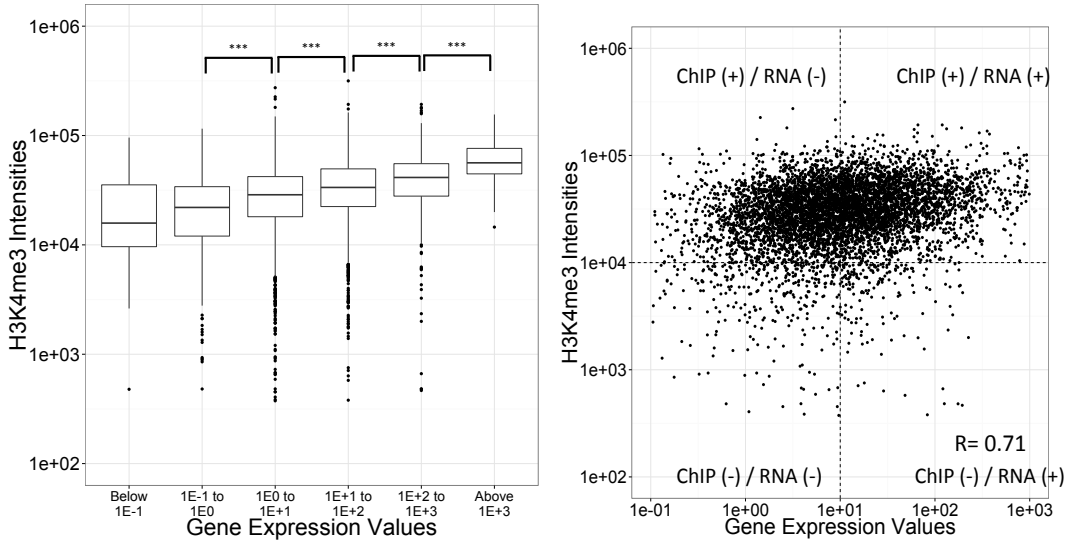


Figure 2

The relationship between the ChIP-seq intensities for H3K4me3 and gene expression

A) Binned boxplot shows that there is a positive correlation between H3K4me3 intensities and gene expression (RPKM).

Scatterplot also shows the positive correlations between H3K4me3 intensities and gene expression values (RPKM). Genes can be classed as ChIP+/RNA+, ChIP+/RNA-, ChIP-/RNA+, ChIP-/RNA- depending on the H3K4me3 intensities and gene expression values

(Maekawa et al., 2015)

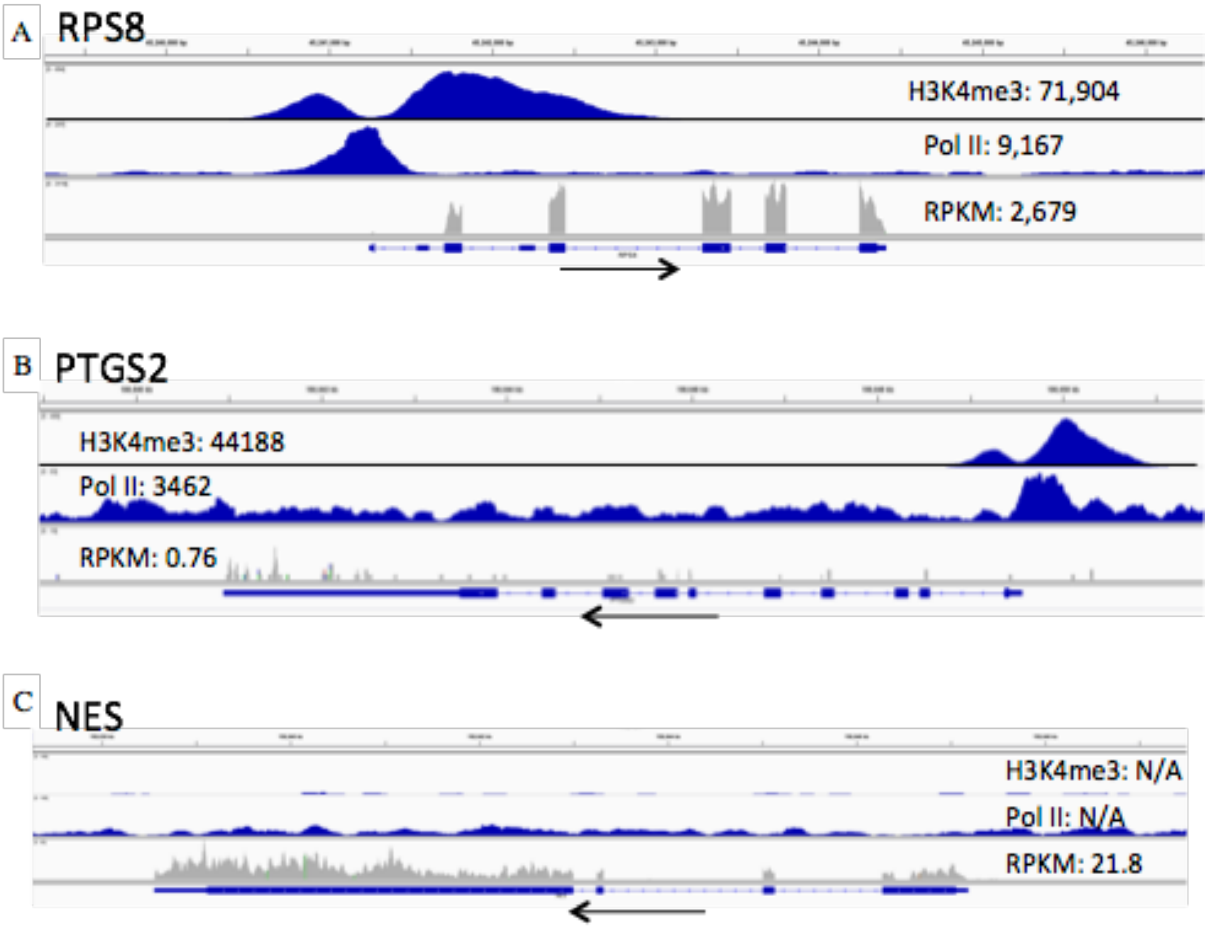


Figure 3
Examples of genes in different regions
B) RPS8 as an example of ChIP+/RNA+ gene with ChIP-seq peak of H3K4me3 and RNA polymerase II with RNA accumulation
C) PTGS2 as an example of ChIP+/RNA- gene with ChIP-seq peak of H3K4me3 and RNA polymerase II without RNA accumulation
NES as an example of ChIP-/RNA+ gene without ChIP-seq peak but with RNA accumulation (Maekawa et al., 2015)

Correlation amongst RNA half-life, RNA abundance and chromatin marks

To investigate the hypothesis that RNA decay is a factor that explain the discrepancy, I used BRIC-seq to determine the RNA stability. BRIC-seq uses 5'-bromouridine (BrU) as a nucleoside analogue to pulse-chase the RNA from BrU saturation at time = 0h, and it can be used to measure the RNA half-life at a genome-wide level by using next-generation sequencing (Imamachi et al., 2014; Tani et al., 2012a; Tani et al., 2012b). In this study, I sequenced 11 time points from time=0 hours to 12 hours to determine the RNA stability, with the average of 10 million sequenced mapped reads per time point. When I compared the RNA abundance and RNA half-life, there was an overall positive correlation where the increase in RNA half-life correlated with the increase in RNA abundance (Figure 4a). However, I detected no significant correlations between the intensity of active chromatin marks and RNA stability (Figure 4b). These observations suggest the potential of RNA stability in contributing to the mRNA abundance.

Through these observations, I speculated the possibility of the RNA stability to explain those that show discrepancies between chromatin marks and RNA abundance. To conduct the analysis, I compared the RNA half-lives between different regions in from the Figure 2b, I observed that ChIP+/RNA- region have significantly shorter RNA half-life in comparison to genes that show correlation between ChIP-seq and RNA-seq (Figure 5a), with ChIP+/RNA- genes having the median RNA half-life of 6.0h in comparison to the ChIP+/RNA+ genes having 11.6h. Through the correlation of ChIP-seq and RNA-seq, I was able to define genes that fit the least-squares regression line, which may potentially act as genes with default half-lives, where genes in the 2x and 1.1x of the least squared regression line had 11.0h and 10.9 hours respectively.

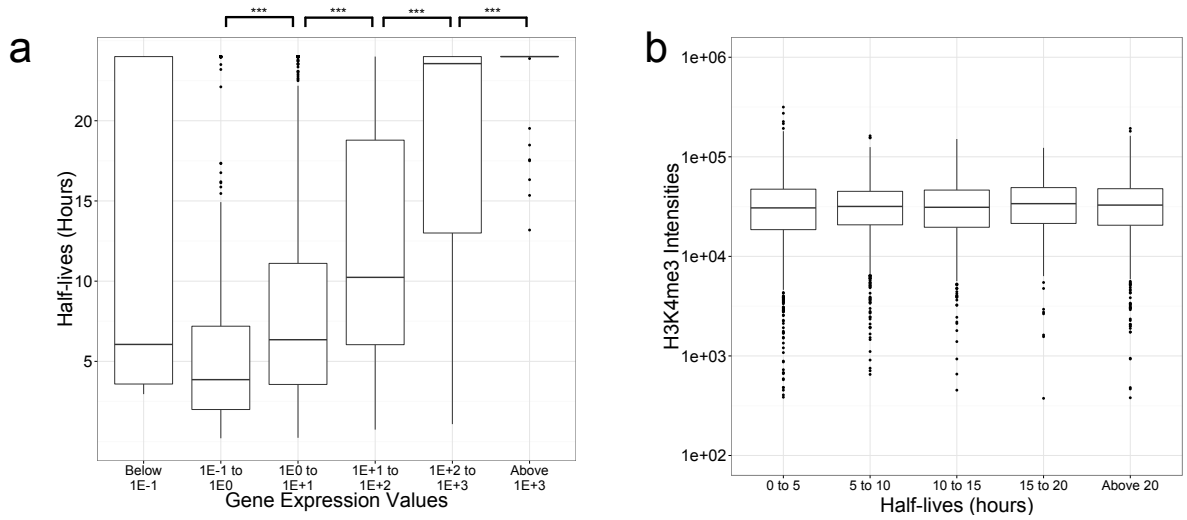


Figure 4
Relationships between RNA half-life, chromatin marks and gene expression
Correlation between RNA half-life and gene expression shows that genes with longer RNA half-lives correlate with genes with higher gene expresssion values
Correlation between RNA half-life and H3K4me3 intensities and it shows that there are no statistical significance in the correlation between these two variables
(Maekawa et al., 2015)

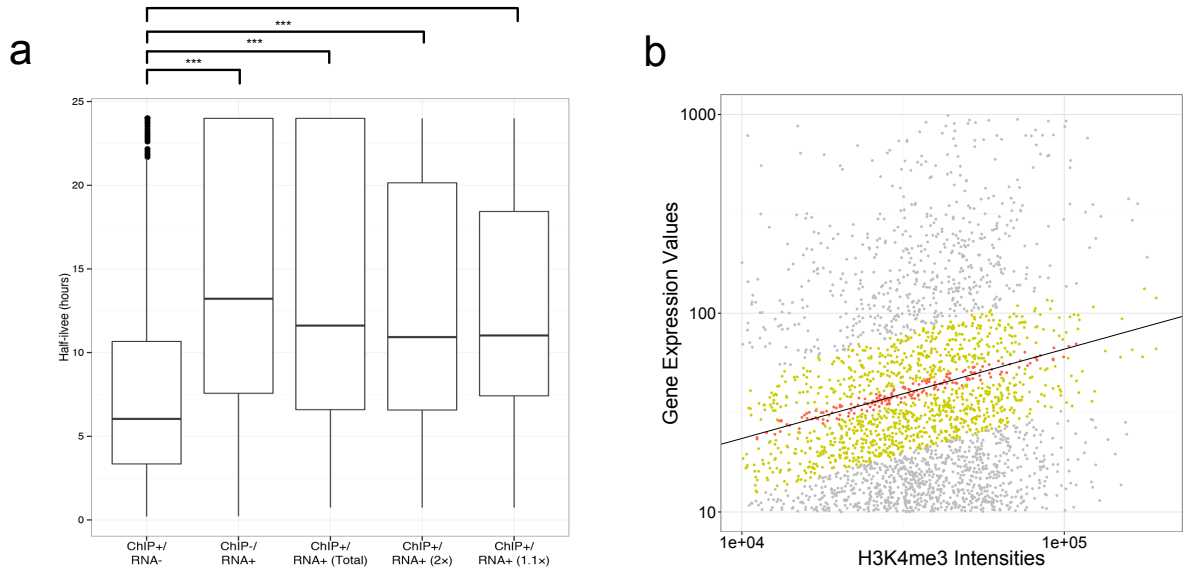


Figure 5
A) Genes with ChIP+/RNA- half-lives have shorter RNA half-life
Genes that are ChIP+/RNA- have the median half-life of 6.0h in contrast 10.9h of ChIP+/RNA+ genes with correlation between ChIP-seq and RNA-seq
B) Genes that correlate between ChIP-seq and RNA-seq. Genes in orange are within 1.1x of the least squares regression line and genes in yellow are within 2x of the least squares regression line. (Maekawa et al., 2015)

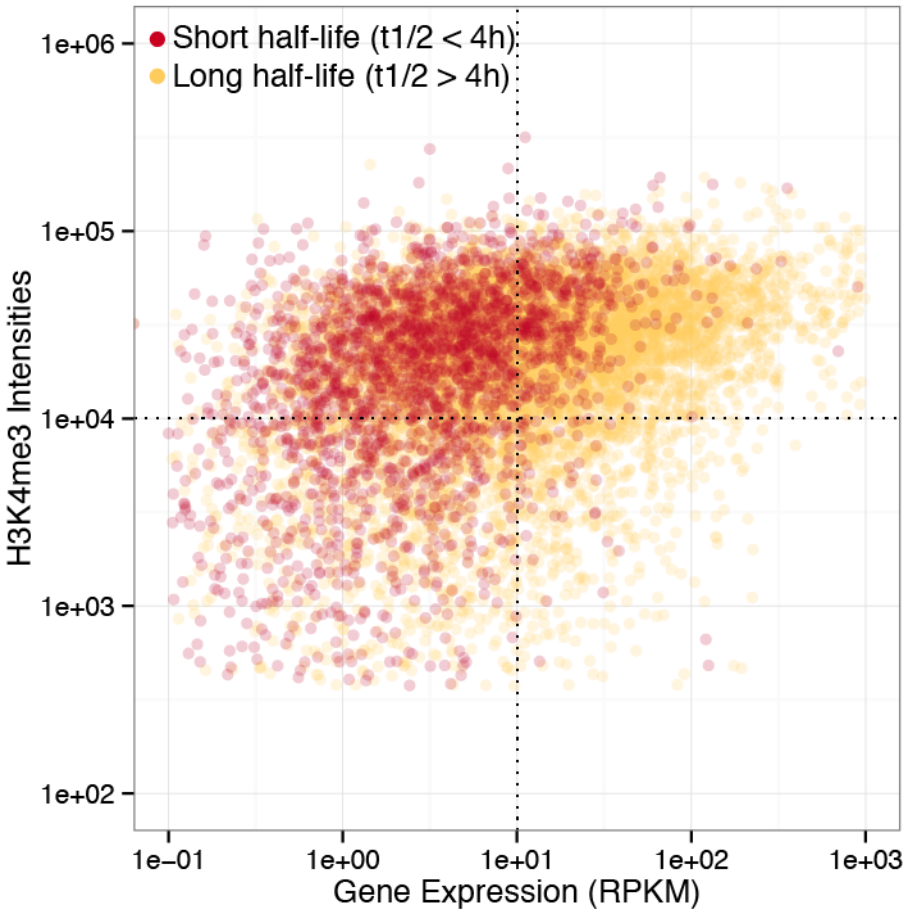


Figure 6

Relationship between ChIP-seq, RNA-seq and RNA half-life
Genes that have shorter RNA half-life (t1/2 < 4h) have tendencies to be ChIP+/RNA- genes (Maekawa et al., 2015)

I defined genes with 'short' RNA half-life to be less than RNA half-life of 4 hours, as 4 hours was 2 standard deviations away from the median ChIP+/RNA+ genes. I overlaid the RNA half-life to the scatterplot with RNA abundance and chromatin marks and it showed that genes with shorter RNA half-life tend to be in the ChIP+/RNA- region with 865 genes in ChIP+/RNA-/short RNA half-life, further suggesting the potential role of RNA decay in regulating the RNA expression in HeLa cells (Figure 6).

In order to identify the types of genes that was in different regions, I conducted gene ontology analysis to determine GO terms that statistically enrich in each region. In the ChIP+/RNA+ region, I found that there are terms associated with RNA processing; for ChIP-/RNA+ genes, GO terms enriched in cytoplasm as a location. Interestingly, I found that genes in ChIP+/RNA- region are strongly enriched in terms associated with transcription factors and even more so for ChIP+/RNA- genes with short RNA half-life ($t1/2 < 4$ h).

In addition, I conducted a simple computational simulation to predict the RNA abundance when the RNA stability was simulated to be 10.9h (median RNA half-life for ChIP+/RNA+ 1.1x) and fitted against the original ChIP-seq and RNA-seq. I found that out of total of 9407 genes, 1540 genes and 229 genes were simulated to have the correlation between ChIP-seq and RNA-seq data, between 2x and 1.1x of the least-squared regression line; thus, suggesting the role of RNA stability in mediating changes to the RNA abundance (Figure 7). Furthermore, I conducted a linear regression using the normalised Z-scores of H3K4me3, H3K27me3, H3K36me3, Pol2 and RNA decay constant and I tested the fit with and without RNA decay constants. I found that the fitting increased from R=0.41 to R=0.58 upon the addition of RNA decay component (Figure 8).

Table 1

List of Gene ontology enrichments

ChIP(+) / RNA(+) genes

GO ID	**GO Term**	**GO Genes**	**FDR**
GO:0044822	poly(A) RNA binding	618	6.00E-216
GO:0006412	translation	293	3.86E-196
GO:0010467	gene expression	387	5.61E-162

ChIP (-) / RNA (+) genes

GO ID	**GO Term**	**GO Genes**	**FDR**
GO:0005737	cytoplasm	138	2.26E-02
GO:0070062	extracellular vesicular exosome	74	2.61E-02
GO:0005635	nuclear envelope	15	2.81E-02

ChIP (+) / RNA (-) genes

GO ID	**GO Term**	**GO Genes**	**FDR**
GO:0003677	DNA binding	378	2.96E-28
GO:0006355	regulation of transcription, DNA-templated	266	4.44E-17
GO:0006351	transcription, DNA-templated	361	8.60E-17

ChIP (+) / RNA (-) / short RNA half-life genes

GO ID	**GO Term**	**GO Genes**	**FDR**
GO:003677	DNA binding	216	5.19E-58
GO:0006351	transcription, DNA-templated	192	1.37E-37
GO:0006355	regulation of transcription, DNA-templated	147	5.13E-34

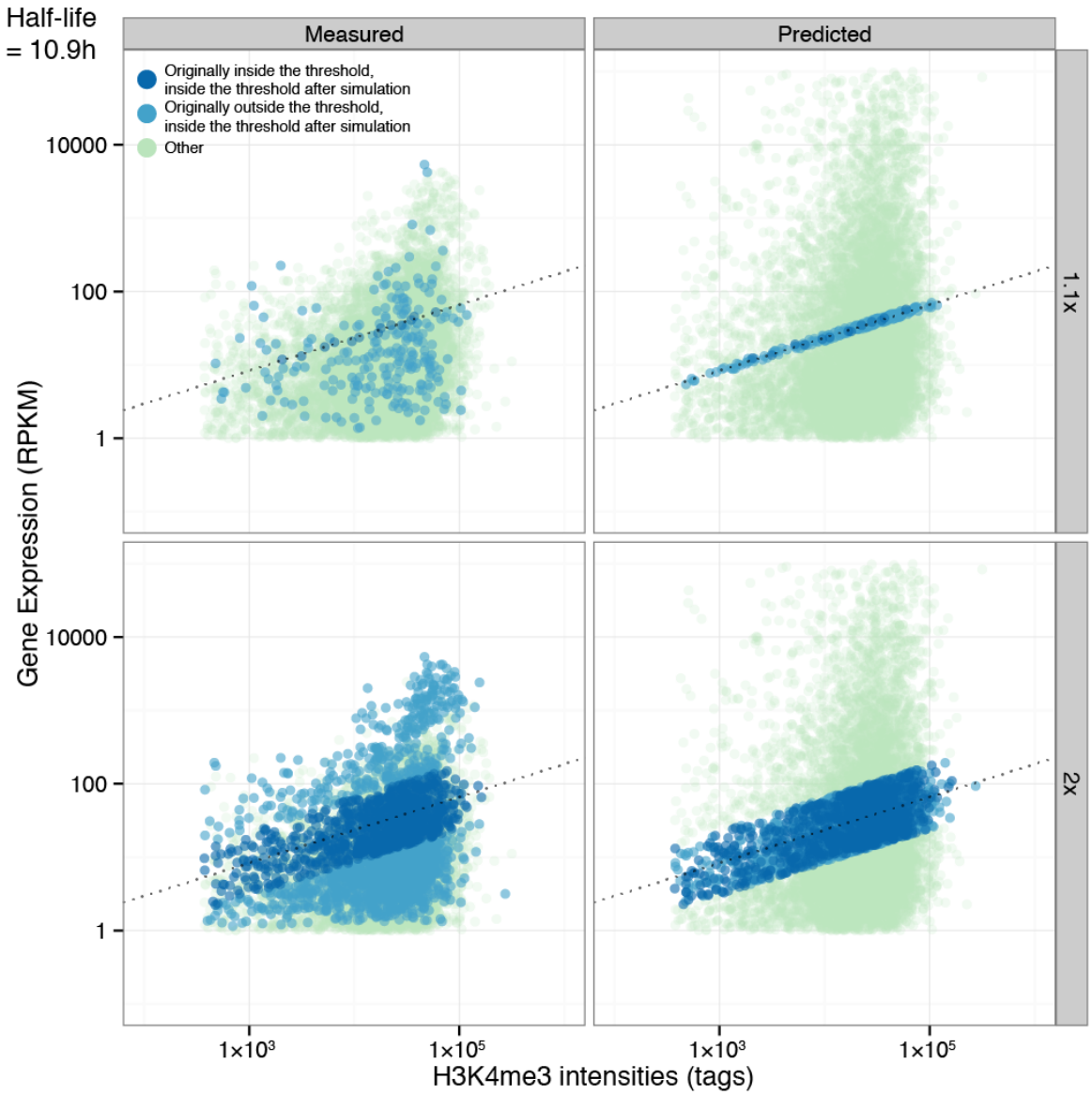


Figure 7

Computational simulation on the role of RNA stability on RNA abundance

This was done through normalising the RNA decay of each gene to 10.9 hours and predict RNA decay Left and right panels show the measured and predicted RNA abundance, respectively. Top and bottom panels show genes when the 'correct' measurement is defined as 1.1x and 2x, respectively.
(Maekawa et al., 2015)

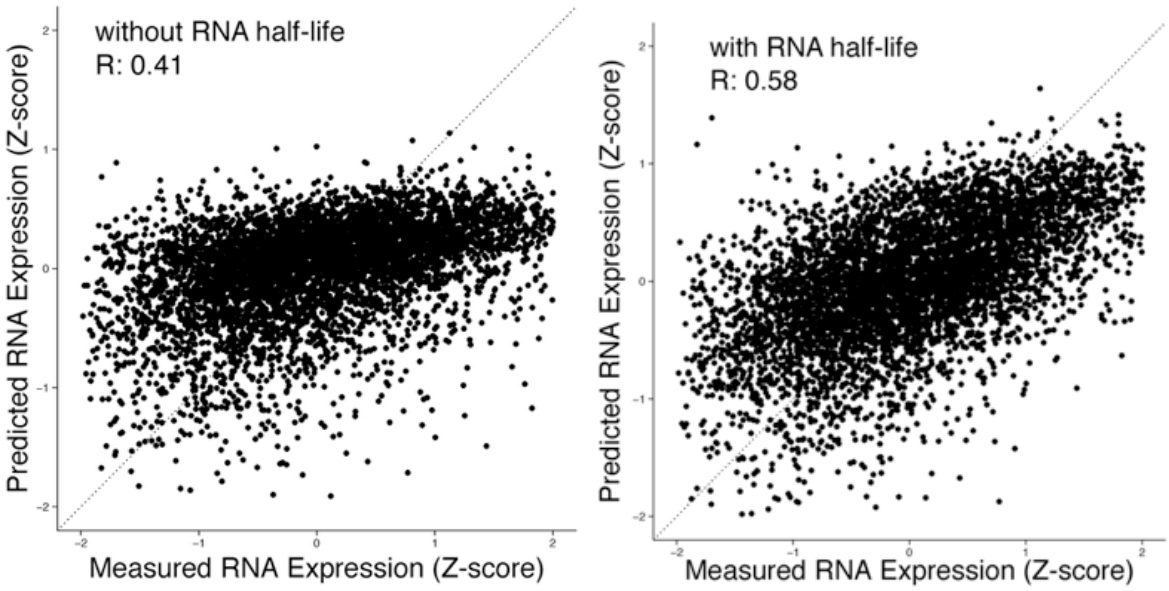


Figure 8
Linear regression to predict RNA abundance

A) shows linear regression of H3K4me3, RNA Pol2, H3K27me3 and H3K36me3
B) shows linear regression of a with a linear RNA decay constant component.
(Maekawa et al., 2015)

Role of RNA decay factors in mediating changes to the RNA abundance

To identify the potential role of RNA decay factors in mediating changes to the RNA abundance through the RNA stability, I investigated the role of UPF1, EXOSC5 and STAU1. For EXOSC5 and STAU1, I conducted siRNA knockdown to these factors and conducted RNA-seq and BRIC-seq to identify the targets of these RNA decay factors. With UPF1, I used the previous siRNA followed by RNA-seq and BRIC-seq in the same cell line from the Akimitsu laboratory. By looking at the overlap of genes that were up-regulated in RNA abundance by more than two-fold (UPF1 and EXOSC5) or 1.5-fold (STAU1) and elongated in RNA half-life by more than two-fold (UPF1 and EXOSC5) or 1.5-fold (STAU1), I identified 266, 219 and 39 genes upon depletion in UPF1, EXOSC5 and STAU1, respectively. Of those target genes 23, 40 and 4 were in the ChIP+/RNA-/short RNA half-life region of the Figure 2, which is 3, 5 and 0.5% out of 895 genes in this region, for UPF1, EXOSC5 and STAU1 respectively (Figures 9-11). I also examined the overlap between the target genes and found no evidence of overlap. Through these analysis, I was able to start to observe the extent of RNA decay factor contribution for limited RNA decay factors.

Interestingly, when I analysed genes that were elongated in RNA half-life upon siRNA knock-down to UPF1 and EXOSC5, I observed that 975 and 6309 genes showed elongation of RNA half-life without up-regulation of the RNA abundance. This phenomenon could be explained by the feedback of RNA decay to transcription factor, where the elongation of RNA decay to a negative regulator of transcription leads to the increase in the transcription repressor, which increase in the protein expression and lead to the repression of the downstream genes that oppose the effect of RNA stability increase. To test the hypothesis, I conducted a motif search of the promoter regions for those 975 and 6309 genes (Figure 12). In the case of UPF1, I was able to detect statistically significant binding motif for HIC1, a zinc finger transcription factor

and validate the changes in expression and RNA half-life for HIC1; however, I could not conduct a ChIP as no working antibodies were available (Figure 13).

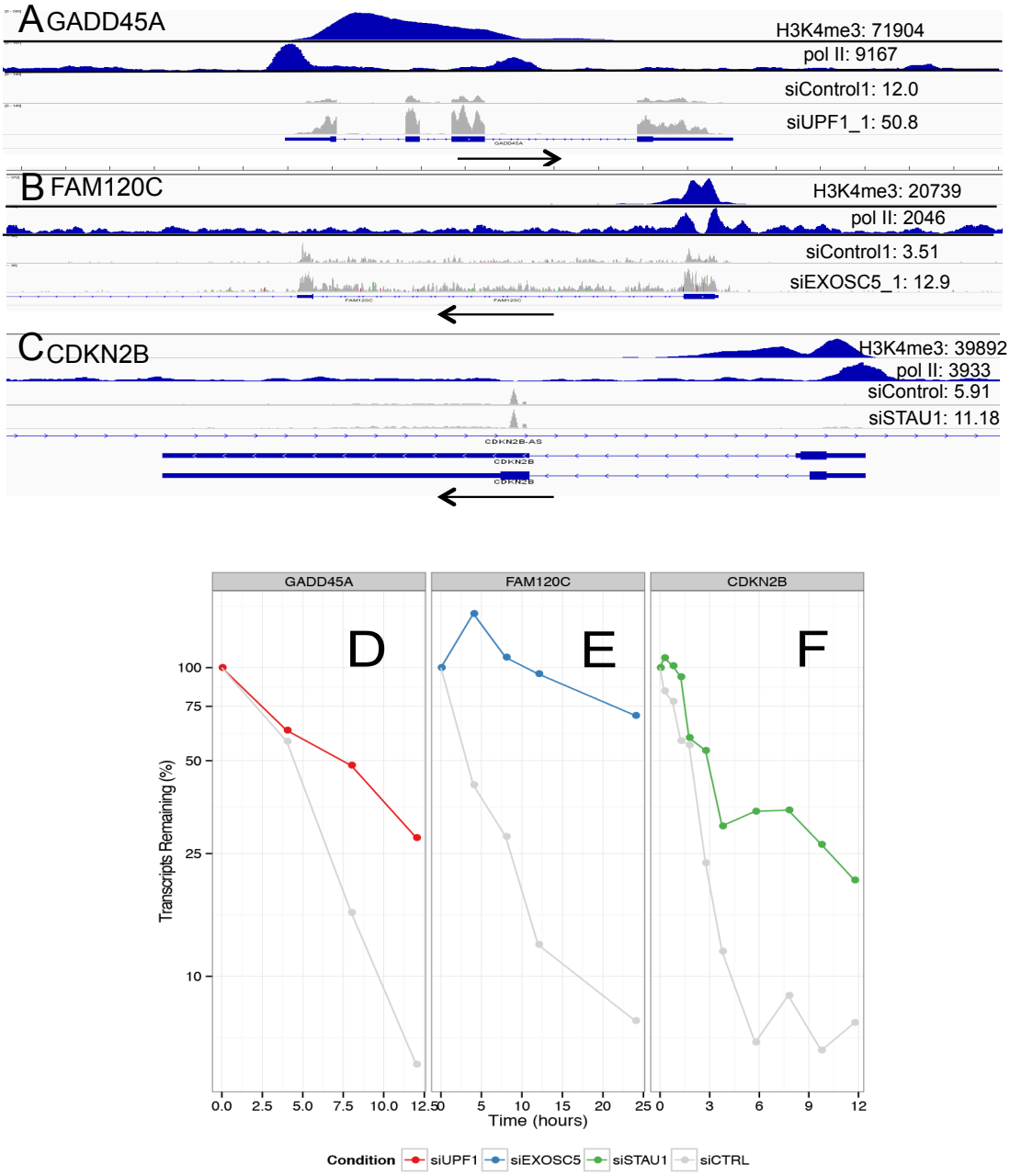


Figure 9
Examples of genes that are regulated through RNA decay factors
D) and D) Example of GADD45A that is regulated by UPF1
E) and E) Example of FAM120C that is regulated by EXOSC5
and F) Example of CDKN2B that is regulated by STAU1
(Maekawa et al., 2015)

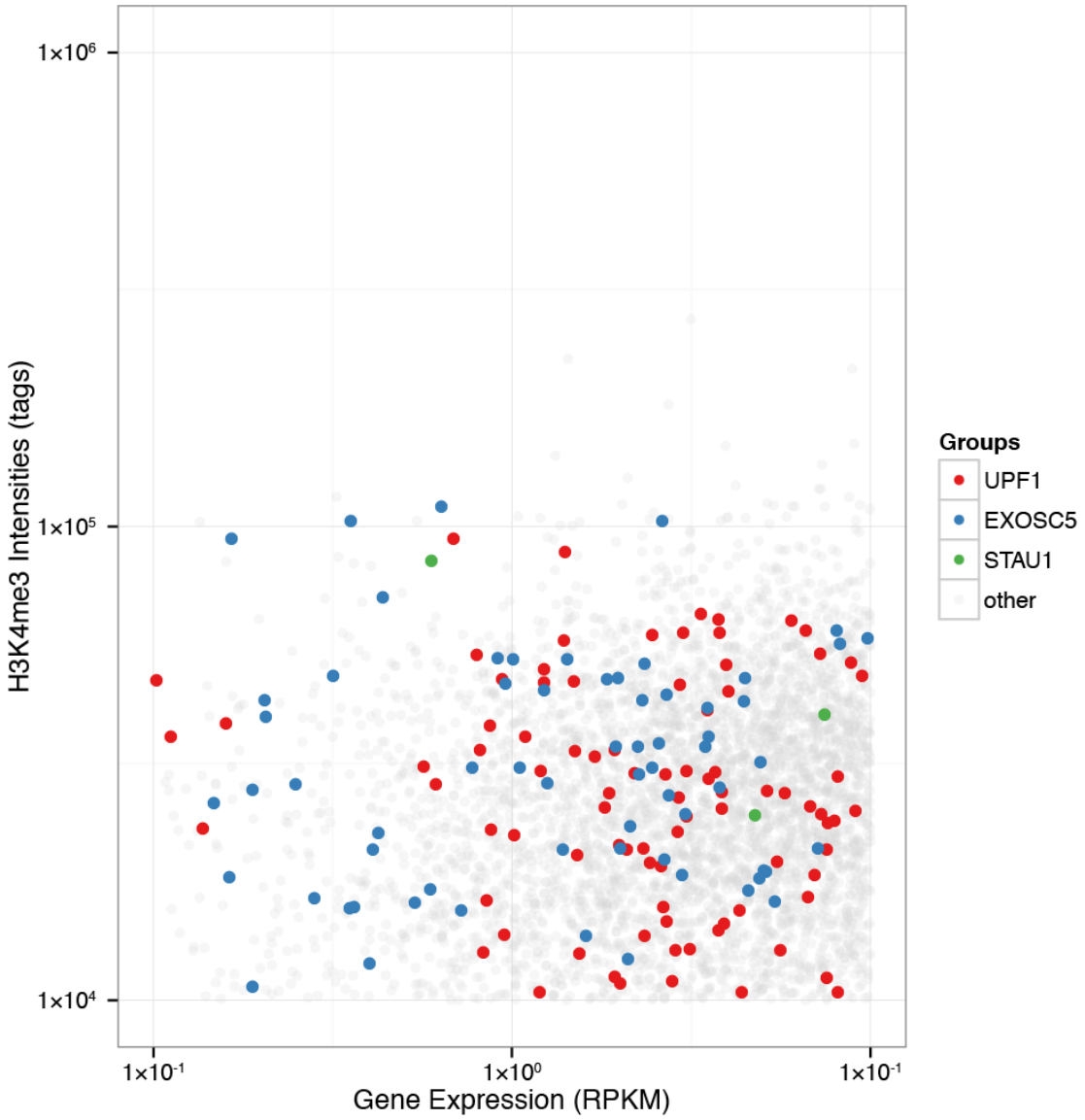


Figure 10
The targets of RNA decay factors for ChIP+/RNA- genes
Genes in colours show genes that are targeted by UPF1 (red), EXOSC5 (blue) and STAU1 (green).
Genes in grey are genes that are not regulated by RNA decay factors
(Maekawa et al., 2015)

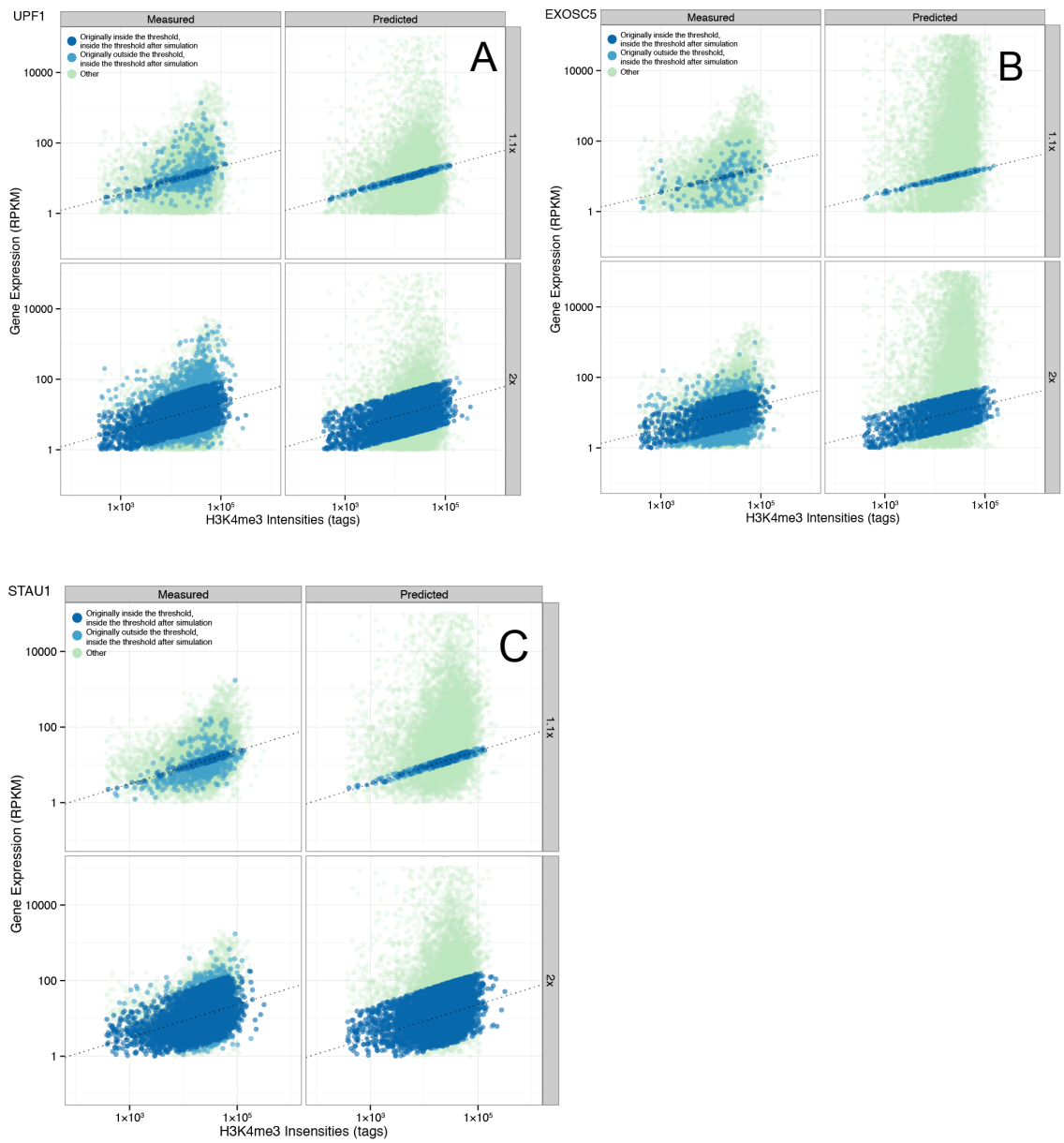


Figure 11
Computational simulation to show the effects of changes to the RNA half-life upon siRNA knockdown to RNA decay factors A) UPF1, B) EXOSC5 and C) STAU1
(Maekawa et al., 2015)

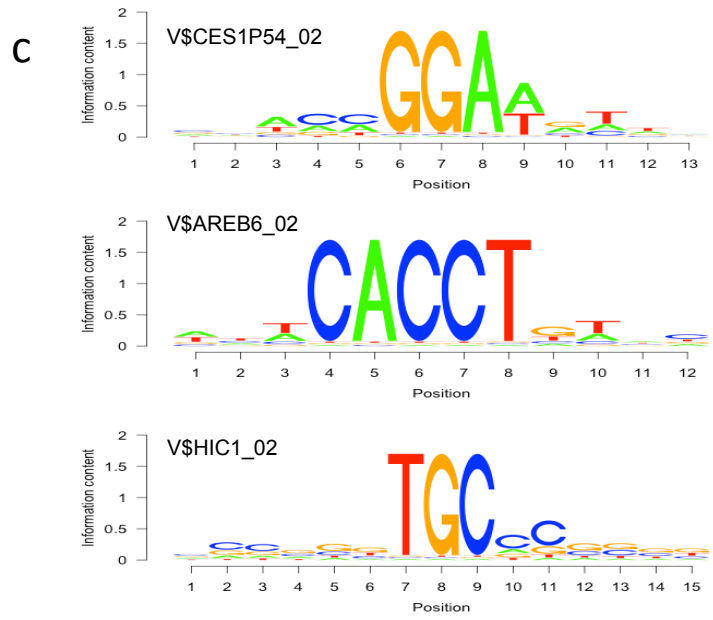


Figure 12
Transcription factor search for ENCODE and DBTSS cell types
These transcription factor motifs were enriched in genes suspected for negative feedback
(Maekawa et al., 2015)

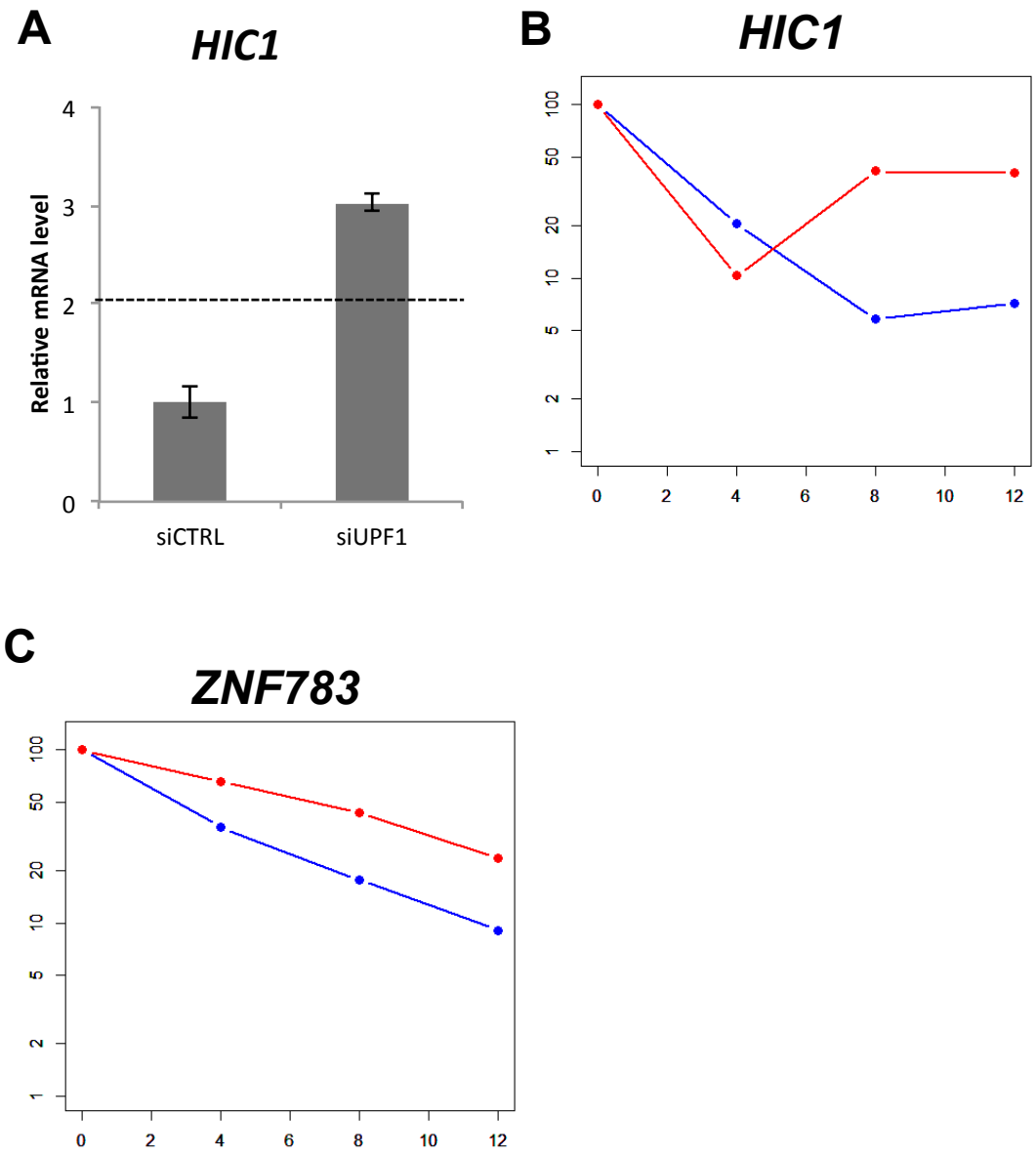


Figure 13
Transcription factor that may be regulated by RNA decay factors

A) and B) HIC1 is induced upon UPF1 knockdown with changes to the RNA decay
B) C) siUPF1 in read and siCTRL in blue

(Maekawa et al., 2015)

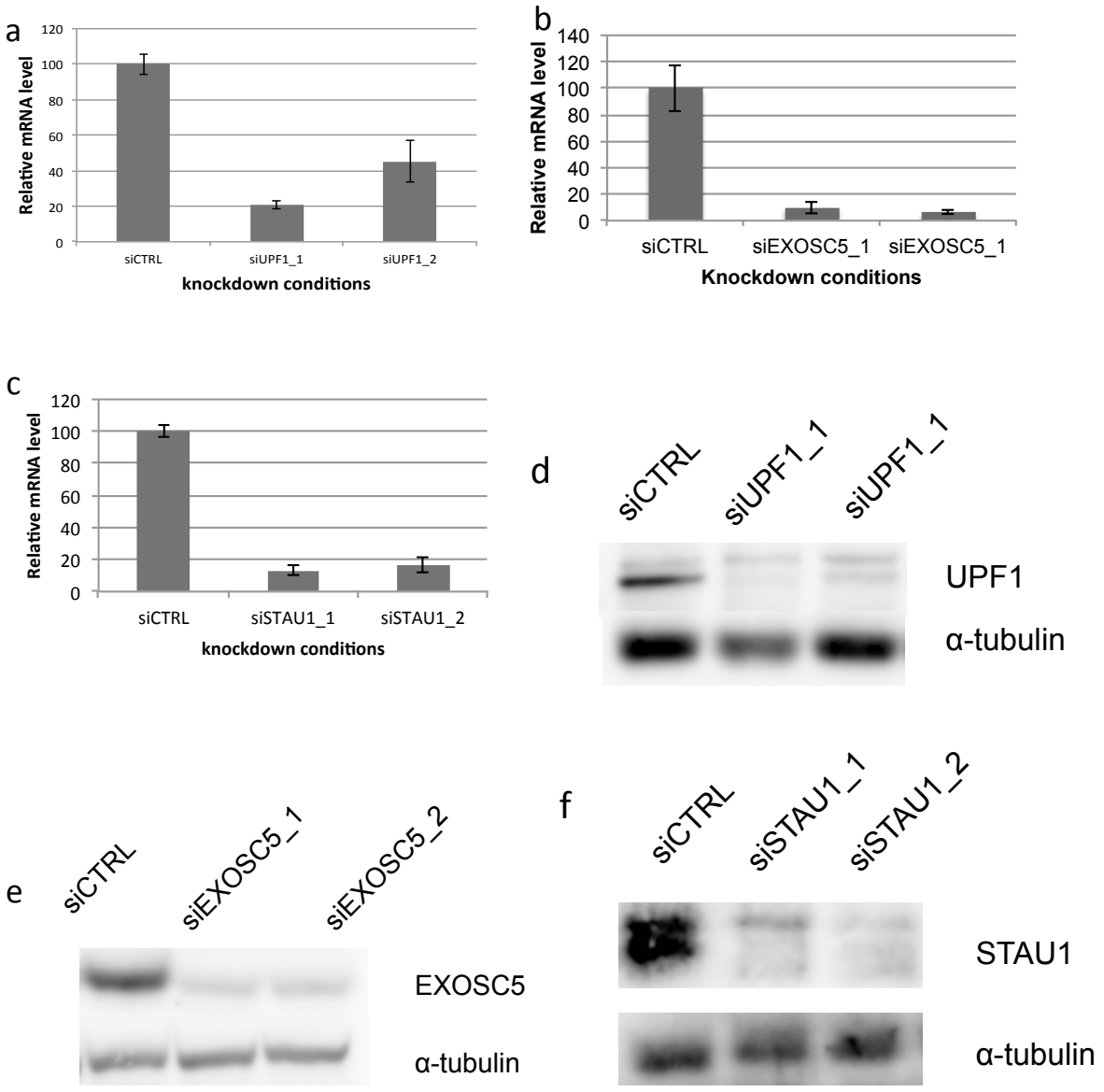


Figure 14
Validation of RNA decay factor knockdown

A) And D) Validation of UPF1 knockdown by qRT-PCR (A) and Western blot (D)
B) And E) Validation of EXOSC5 knockdown by qRT-PCR (B) and Western blot (E)
C) And F) Validation of STAU1 knockdown by qRT-pCR (c) and Western blot (F)

(Maekawa et al., 2015)

Analysis of ENCODE cell lines

To evaluate the role of RNA decay contributions on mRNA abundance for genes encoding transcription factors, I used ENCODE data and DBTSS data to analyse on the types and distribution of genes that have the chromatin marks but without the accumulation of RNA. As ENCODE and DBTSS datasets do not possess RNA stability data, I analysed ChIP-seq and RNA-seq data and I ran the dataset through similar criteria to the previous analysis. I found that there were on average of 338 candidate genes that were ChIP+/RNA- with 2705 genes that are specific to the particular cell type (Figures 15 and 16). When I ran the GO analysis, I found that 7 out of 8 cell types available had shown statistically significant enrichment of gene ontology terms associated with DNA binding. This suggests a more general role of transcription factors having shorter RNA half-life.

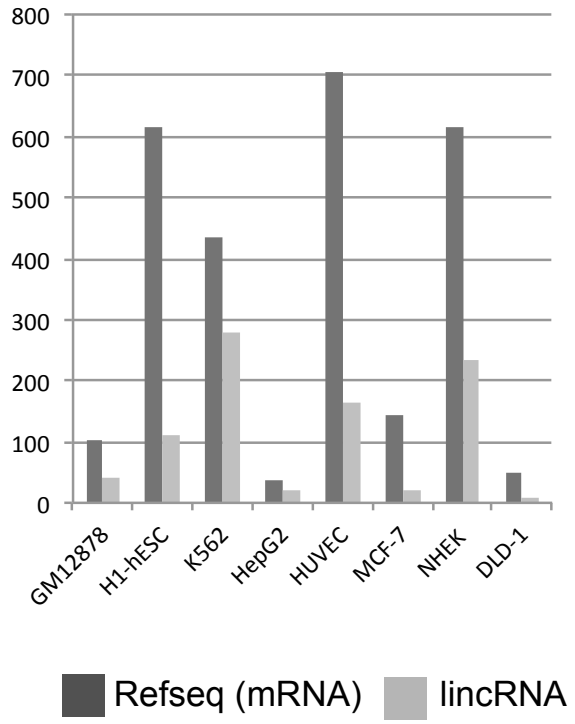


Figure 15
Number of genes that were ChIP(+)/RNA(-)
Barplot showing the number of genes that were ChIP+/RNA-
(Maekawa et al., 2015)

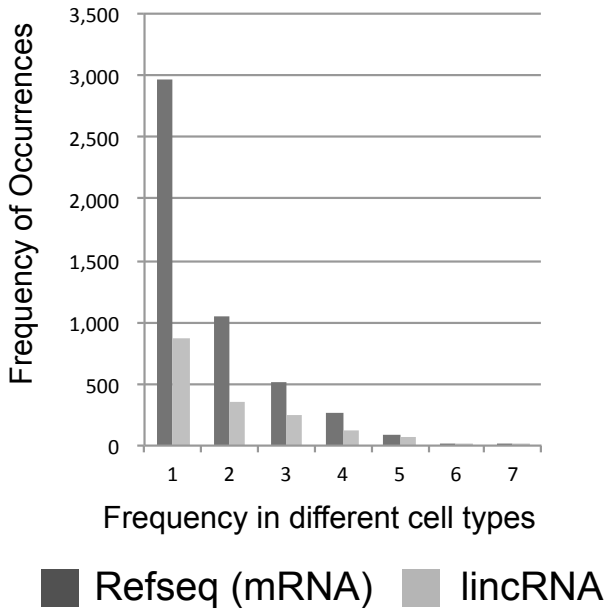


Figure 16
Frequency of overlap of genes in the ChIP+/RNA- region
2705 genes are unique in one particular cell type
(Maekawa et al., 2015)

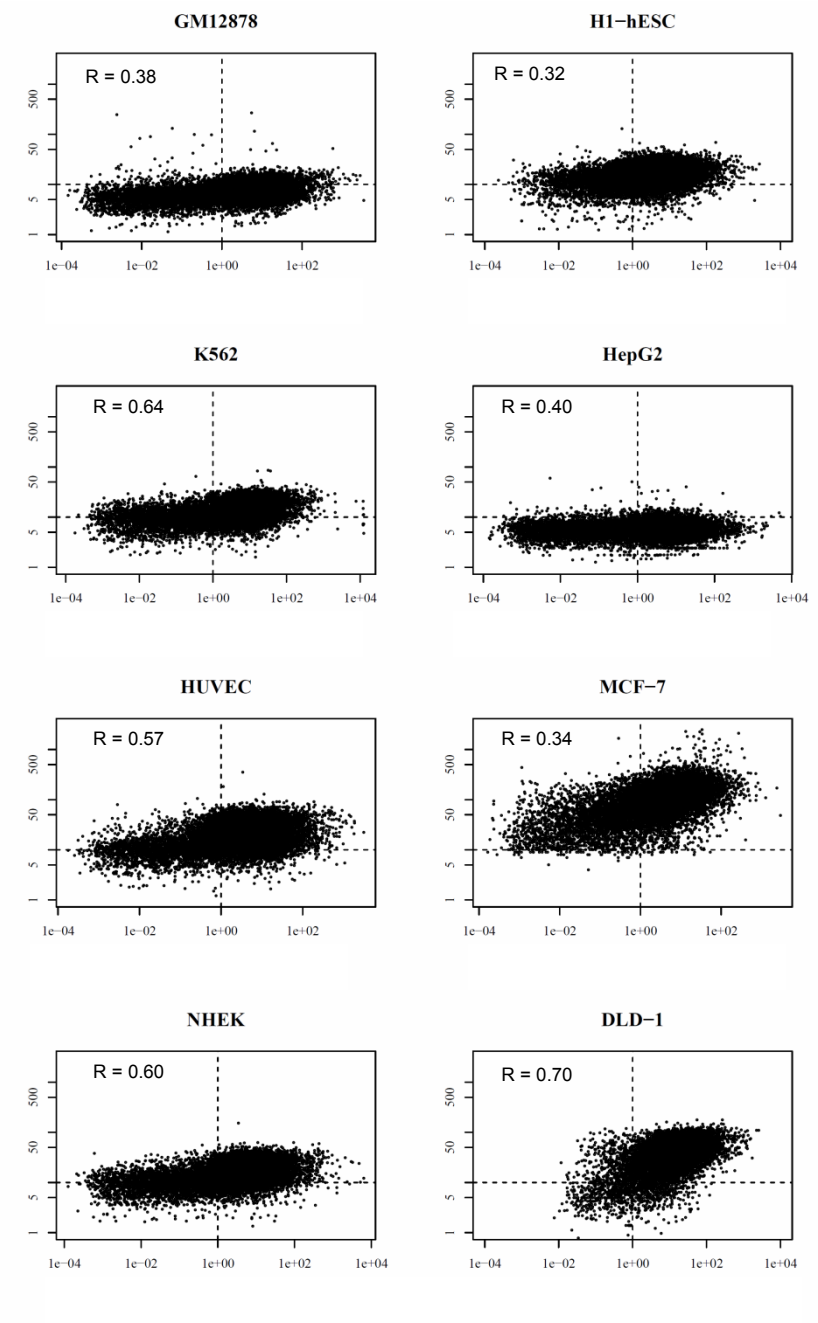


Figure 17
Correlation between ChIP-seq and RNA-seq data in ENCODE and DBTSS cell-line

y-axes are the ChIP-seq intensities and x-axes are the RNA expression data (RPKM)
R is calculated with log transformed Pearson's correlation coefficient

(Maekawa et al., 2015)

Distinct regulation of non-coding RNAs in comparison with mRNAs

In addition to mRNAs, I also investigated the role of RNA decay in determining the level of RNA abundance, as previous studies have identified that ncRNA have low expression; which some can be attributed to RNA decay. To analyse the regulation of ncRNA, here I focused on the long intergenic non-coding RNA (lincRNA) and used ChIP-seq and RNA-seq to determine the number of lincRNA that are potentially regulated. Interestingly, out of 141 lincRNA in HeLa cells, 103 lincRNA had short RNA half-life, with 84 lincRNA being in the ChIP+/RNA- region of the figure 18. I analysed the contribution of UPF1 and EXOSC5 to regulate these lincRNA, as there are recent reports on the possible role of UPF1 in regulating ncRNA and EXOSC5 is also thought to be regulating ncRNA. Interestingly, no lincRNA was regulted by UPF1, whereas 26 were regulated by EXOSC5; suggesting a more extensive role of exosome in regulating the lincRNA but not the nonsense mediated decay in this primitive analysis (examples in figure 19). In addition to the lincRNA, I investigated the role of enhancer RNA (eRNA), that is transcribed from the enhancer regions of the genome. To identify enhancers, I used H3K4me1 and H3K27ac ChIP-seq, from ENCODE dataset of HeLa cells, which are representative chromatin marks for active enhancers. I defined enhancers as loci that possess both H3K4me1 and H3K27ac peaks that are 1.5kb away from any RefSeq gene. Out of 49903 genomic locations identified as potential enhancers, 77 and 358 eRNA loci that are potentially regulated by UPF1 and EXOSC5, respectively, with an example in figure 20. Although the study is still primitive, it highlights the different regulation that ncRNA receives in comparison with mRNA and that they may be controlled by these RNA decay factors.

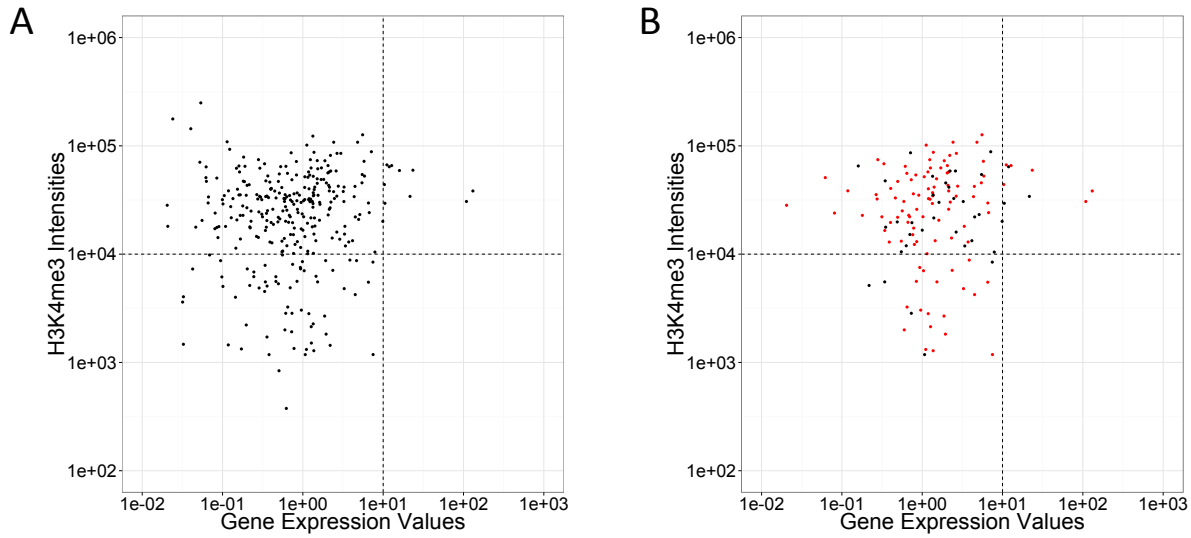


Figure 18
Correlation between ChIP-seq and RNA-seq (A) and with RNA half-life (B)

A) Scatterplot for ChIP-seq and RNA-seq for HeLa cells for lincRNA

Majority of genes have low gene expression values as expected but many still have ChIP-seq signal

B) With RNA decay data, it shows that majority of lincRNA genes have shorter RNA half-life (red dots t1/2 < 4h)

(Maekawa et al., 2015)

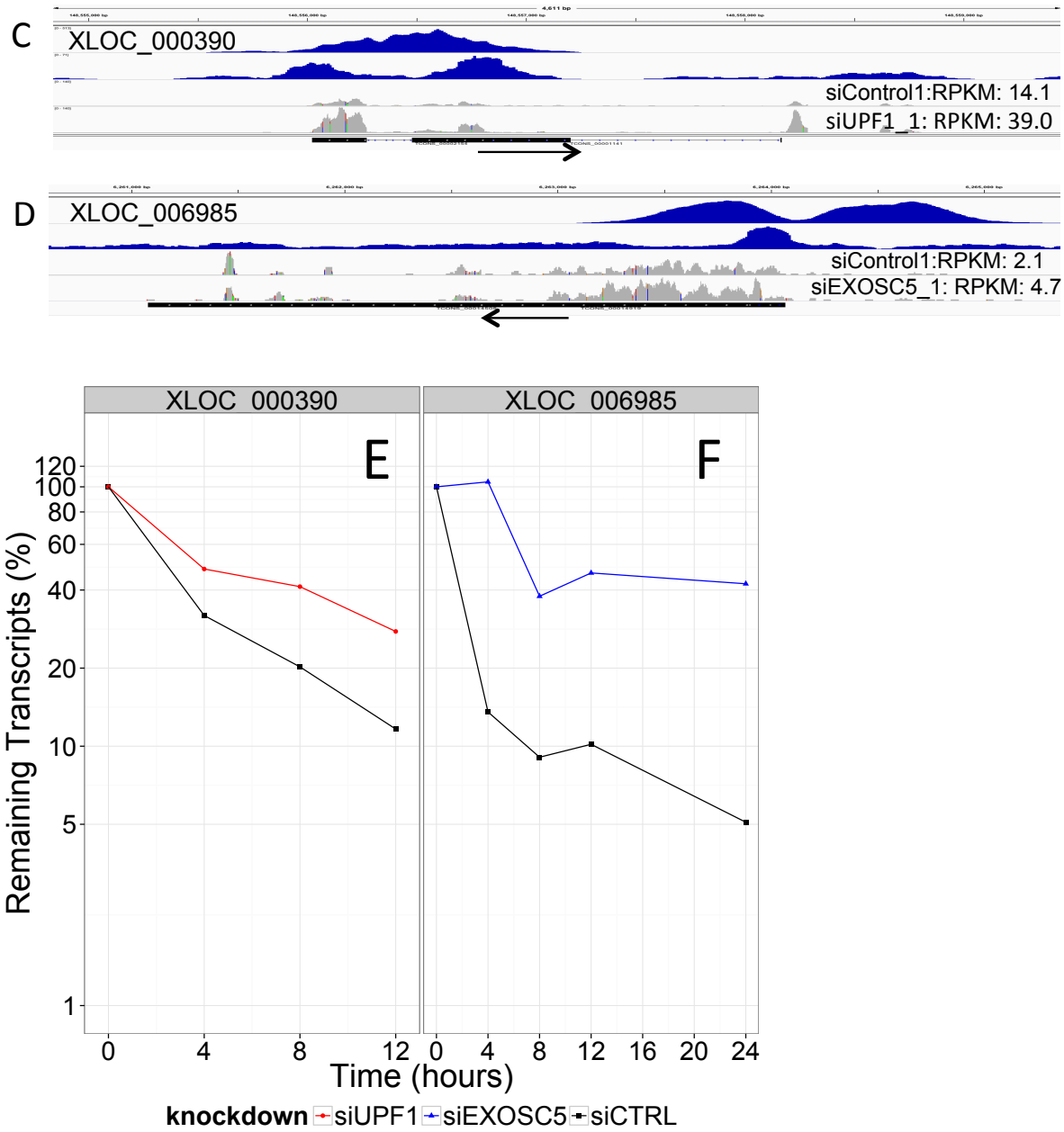


Figure 19
Examples of genes that could be regulated by UPF1 and EXOSC5 for lincRNA

A) And C) example of a lincRNA regulated by UPF1
B) And D) example of a lincRNA regulated by EXOSC5

(Maekawa et al., 2015)

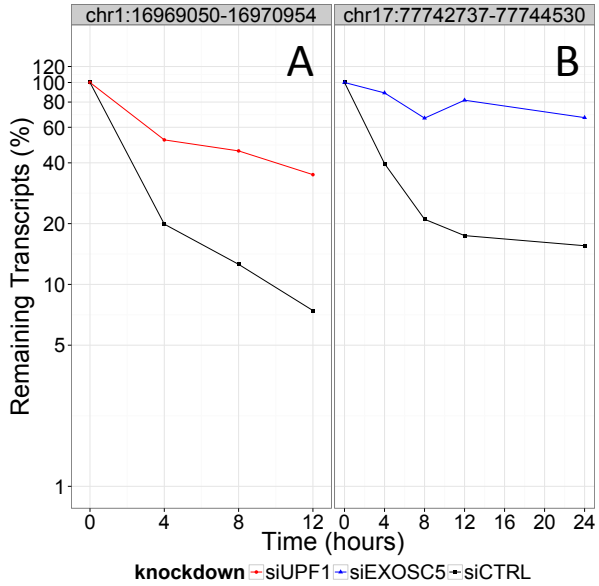


Figure 20
Example of an eRNA possibly regulated by RNA decay
a) By UPF1 b) by EXOSC5

(Maekawa et al., 2015)

Discussion

Through this study, I was able to determine the contribution of RNA decay in mediating the RNA abundance levels. I identified that RNA decay play an important role in subset of genes for those that have inconsistencies between the ChIP-seq and RNA-seq. I was able to identify 865 genes with inconsistency between ChIP-seq and RNA-seq and also showed shorter RNA half-life. Through the gene ontology analysis, I was able to identify that these 865 genes were statistically enriched in GO terms associated with transcription factors. The analysis of different cell lines in the ENCODE dataset has come to the similar conclusions regarding the inconsistencies between ChIP-seq and RNA-seq and that genes that are ChIP+/RNA- are enriched in transcription factors for majority of cell lines. In addition, I was able to evaluate the contributions of three RNA decay factors and I found that out of 865 genes, 60 genes (8.5%) were targeted by UPF1, EXOSC5 and STAU1 to control their RNA abundance through their RNA decay. I also conducted analysis on long ncRNA and eRNAs, as they are known to have lower RNA abundance. I was able to identify that the lincRNA and eRNA seemed to have different modes of regulation in contrast to the mRNA as majority of RNA were ChIP+/RNA- with short RNA half-life. To my knowledge, this is one of the first attempts to directly compare ChIP-seq, RNA-seq and RNA half-life information to understand the regulatory mechanisms behind human cells.

Chapter Two: Contribution of effects on RNA decay to hypoxia stress

Rationale

In the previous chapter, I have predominantly analyzed the absolute level of RNA decay. In this chapter I will discuss on the role of RNA decay under stress to analyze the contributions of the RNA decay in mediating changes to the transcriptome, in contrast to transcription. To investigate on the effect of changes to the RNA decay, I used hypoxia as a model in DLD-1 human colorectal cancer cell-line. Hypoxia was used as a model as the role of transcription is relatively well established in hypoxia response, as hypoxia inducible factors (HIFs) have been identified and extensively studied (Keith et al., 2011; Keith and Simon, 2007; Kim et al., 2006; Semenza, 2010; Semenza, 2014; Tanimoto et al., 2011; Tsuchihara et al., 2009).

In this chapter of the thesis, I highlight the extent of the RNA decay contributions to the RNA abundance upon cellular stress to uncover the extent of the contribution of RNA stability changes in the changes to the RNA expression. Firstly, I profiled the RNA stability by using BRIC-seq for hypoxia and normoxia in DLD-1 cancer cells. Secondly, I investigated the role of RNA stability on the change in RNA abundance by comparing the BRIC-seq to the RNA-seq data. Following the comparison, I conducted some computational simulation to predict RNA abundance in hypoxia with the contribution of RNA decay upon hypoxia stimulation. Finally, I analysed the potential *trans*-factors that could play a role in the feedback loop of gene regulation in hypoxia.

Materials and Methods

Cell culture

DLD-1 cells were grown in high glucose DMEM supplemented with anti-anti and Fetal bovine serum in a humidied 5% CO_2 incubator. Hypoxia condition was induced with 1% O_2 condition for 24 hours.

Western blot

The cell culture media was removed from the cell then washed with PBS. Cells were then treated with 200 μl cell lysis buffer (50mM Tris-HCl, 1mM EDTA, 1% Triton X-100, 160mM NaCl) directly on to the cells. Cells were scraped then boiled at 100ºC for 5 minutes. Proteins were quantified by BCA and run on 7.5% SDS-PAGE. Proteins were transferred to the PVDF membrane with a semi-dry blot and the membrane was blocked with 5% BSA at room temperature for 1 hour and primary antibodies (anti-HIF1: Novus, anti-ACTB: Abcam) were diluted by 1:1000 (HIF1) and 1:2500 (ACTB) with 5% BSA at 1hour. Membrane was incubated with diluted secondary antibodies (1:25000) at room temperature for 1 hour. The imagine was conducted using ECL Prime Western blot (GE Healthcare) and imaged on LAS-4000 system (GE Healthcare).

BRIC-seq

BRIC-seq was conducted as above and previously reported (Tani et al., 2012b) with the following time-points: 0h, 0.25h, 0.5h, 1h, 1.5h, 2h, 3h, 4h, 6h, 8h, 10, 12h. The hypoxia condition was conducted in a 1% O_2 over 24 hours and the control normoxia at 20% O_2. The immunopurified RNA were sequenced using Illumina Tru-seq RNA-seq library preparation kit without polyA selection on HiSeq 2500. The sequencing statistics are on the supplementary

table 11. The reads were mapped against rRNA genes to remove rRNA in the library by Bowtie2 (Langmead and Salzberg, 2012) then mapped to the human genome reference of hg19 with GENCODE gene annotation (v19) using Tophat2 (Kim et al., 2013; Trapnell et al., 2009) and the mapped sequences were quantified using Cufflinks (Trapnell et al., 2010). The RNA abundances were normalized to ACTB, to account for the immunoprecipitation efficiency variances. The normalized RNA abundance was normalized to time 0h to calculate the relative RNA remaining for the particular gene. The relative RNA abundances were fitted against the one-phase exponential decay model and to calculate the decay constant and the residuals. The decay constant was then converted to RNA half-life using t1/2 = ln(2)/decay constant. The genes were then filtered for the fitting. The half-life was used if the gene had $r2 > 0.7$ and with a positive RNA half-life. For genes that had RNA half-life of more than 24 hours, the half-life was set to 24 hours as a maximum possible RNA half-life to avoid over-extrapolation.

ChIP-seq and RNA-seq

RNA-seq, and ChIP-seq raw sequence reads were obtained from the previous study (Tanimoto et al., 2011). RNA-seq was mapped to the human genome (hg19) using Tophat2 with GENCODE annotation (V19) (Kim et al., 2013; Trapnell et al., 2009) and quantified with Cufflinks (Trapnell et al., 2010).

ChIP-seq was mapped to the human genome (hg19) using Bowtie2 (Langmead and Salzberg, 2012) and the peaks were called using MACS v1.8.2 (Zhang et al., 2008). The peaks were assigned to genes by assessing whether the peak summit of ChIP-seq was within +/- 1.5kb of the TSS of the longest isoform in GENCODE v19 database.

Gene ontology analysis

Gene ontology database was obtained from gene ontology database (Ashburner et al., 2000). The presence of each gene ontologies was counted for the background (all mRNA genes) and the subset of genes of interest. To find the statistical significance, hypergeometric test was used to calculate the *p*-value with False-discovery rate (FDR) to correct for the multiple testing.

Transcription factor and RNA binding factor enrichments

To find the enrichment of transcription factor binding sites, Regulome DB (Boyle et al., 2012) was used. It has over 400 transcription factor ChIP-seq from the ENCODE dataset. To conduct the analysis, I used any loci that is marked as level 5, where the binding is inferred from ChIP-seq data. I analysed the presence of the ChIP-seq binding sites for -500 bp to +100 bp of the TSS and compared between the background and the gene list of interest. The statistical enrichment analysis was conducted using hypergeometric distribution.

HuR binding sites were investigated by using the CLIP-seq data from (Lebedeva et al., 2011), where they had the conservative estimates for the genes that are bound by HuR in HeLa cells. These HuR binding was compared between the background and the sample to find the enrichment and the statistical analysis was conducted using hypergeometric distribution.

Computational simulation to predict the RNA abundance

The computation simulation was conducted by normalizing the RNA decay constant in hypoxia by RNA decay constant in normoxia and multiplying with RNA abundance.

$$\mathrm{RNA}_{\mathrm{hypo}} = \frac{\lambda_{\mathrm{norm}} \cdot \mathrm{RNA}_{\mathrm{norm}}}{\lambda_{\mathrm{hypo}}}$$

Results

Comparison of RNA half-lives in hypoxia and normoxia

I used the previous method of BRIC-seq (Imamachi et al., 2014; Tani et al., 2012b) to determine the RNA half-lives of DLD-1 colorectal adenocarcinoma cell-lines, where the nascent RNAs are labelled with 5'-bromouridine (BrU) and pulse-chased in a time-course over 12 hours, with 11 time points. The labelled RNAs were immunoprecipitated through anti-BrU antibodies and then sequenced through an RNA-seq on Illumina HiSeq2500, where we sequenced on average of 20 million reads per time point (Detailed sequencing statistics on the Table 13). The reads were mapped to the human genome (hg19) using Tophat2 mapping software aided by the Gencode v19 gene model and the gene expression was quantified by normalisation using reads per kilobase exon per million mapped reads (RPKM). By using RPKM normalised RNA expression, the fraction of the RNA remaining were calculated from the start of the pulse-chase, and also normalised by the stable gene (ACTB) in order to minimise the variation between time-points. The obtained relative abundances were used to calculate the RNA half-life where they were modelled to the one-phase exponential decay, the simplest model of decay function. I was able to obtain the RNA half-life measurements of 4902 and 4546 genes, with r^2 value of more than 0.7 and initial RNA expression of more than 5 RPKM, in hypoxia and normoxia respectively (Figure 21). When I compared the distributions of RNA decay, I observed elongation of the overall RNA-life in the hypoxia condition, with the RNA half-lives being 5.62 and 4.59 hours in hypoxia and normoxia respectively (Figure 22). When I compared the RNA stability at the gene level, we found 361, 50 and 3552 genes that were elongated, shortened and unchanged in RNA stability when subjected to hypoxia (Figure 22), further suggesting that RNA becomes more stable on average; however, the changes in the RNA stability is not uniform, with gene specific changes to RNA decay in hypoxia.

Interestingly for 361 genes that increase in RNA half-life by more than two-fold, gene ontology terms for genes involved in regulation of DNA-templated transcription is significantly enriched; whereas I fail to see the enrichment of any gene ontologies for 50 genes that shorten in RNA half-life. This suggest a potential role of RNA decay induced transcriptional network change that mediate changes in the eventual RNA abundance. Another point is that the changes in that there is an inverse correlation between the original (normoxia) RNA half-life to the changes in RNA abundance (Figure 23) One example of gene that is involved in the regulation of transcription that is elongated in RNA half-life that is KDM3A (also known as JMJD1A) shown in Figure 22, which activate target genes through histone demethylation of the histone H3K9me1 and H3K9me2. KDM3A is a factor that is known to be regulated by HIF-1 (Lendahl et al., 2009) and KDM3A regulates GLUT3 (SLC2A3) aided by cooperative DNA binding of HIF-1 (Mimura et al., 2012). Du *et. al.* identifies that endogenous miRNA MiR-155 directly targets KDM3A and downregulate KDM3A nasopharyngeal carcinoma (Du et al., 2011). This suggest the potential role of RNA decay in mediating genes that are biologically relevant for hypoxia, with KDM3A as an example.

One of the regulators that are known to regulate genes in the hypoxia response include HuR, which is a known regulator to stabilise VEGF mRNA, by direct binding to the target mRNA(Levy et al., 1998). To investigate on the effect of HuR, I used CLIP-seq data from previous publication(Lebedeva et al., 2011) to obtain the positions of HuR binding in the transcriptome and observe the overlap with those that are elongated in RNA half-life. Out of 361 genes that have their mRNA stabilised in hypoxia, 40 genes had HuR binding sites, which was statistically significant (*p*-value: 5.26×10^{-5}). In the example of KDM3A, I was able to detect a CLIP-seq peak at the 3'UTR of the KDM3A gene, suggesting that HuR is involved in the elongation of RNA half-lives.

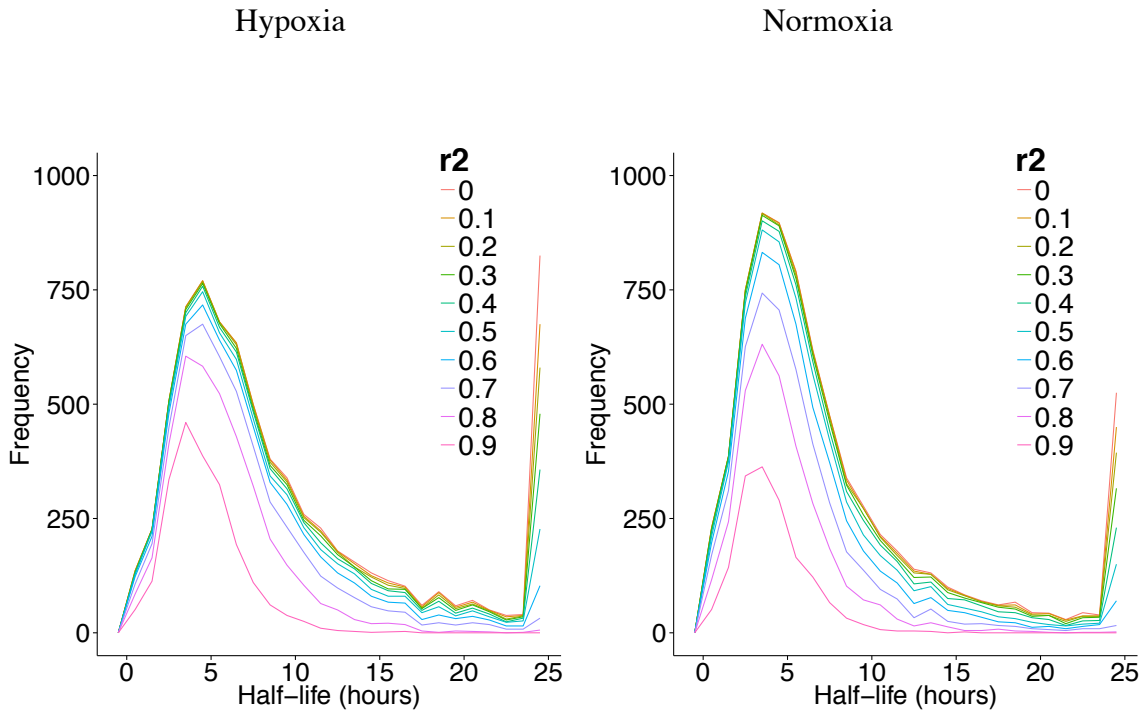


Figure 21
Distribution of RNA stabilities for DLD-1 in hypoxia and normoxia
RNA decay was modelled to the exponential decay model in order to compare between two conditions.
R2 value is the fit to the decay model.

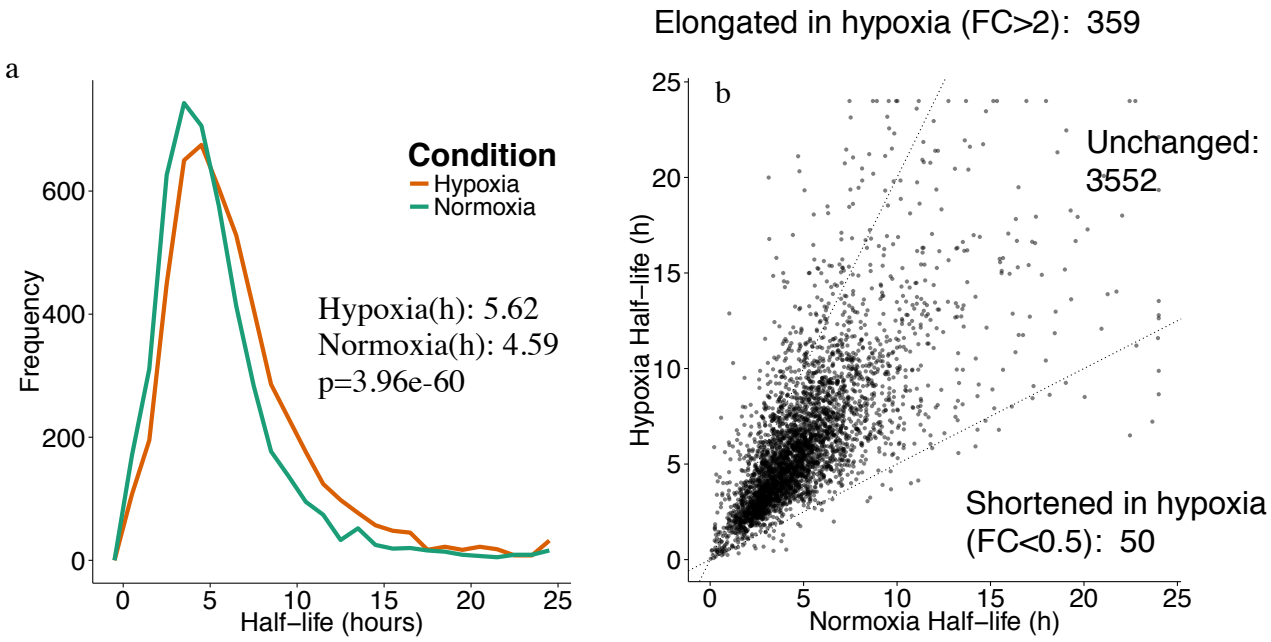


Figure 22
Comparison of RNA half-lives in hypoxia and normoxia

a) Histogram of RNA half-life for both hypoxia and normoxia
b) Comparison of RNA half-life for each gene in hypoxia and normoxia

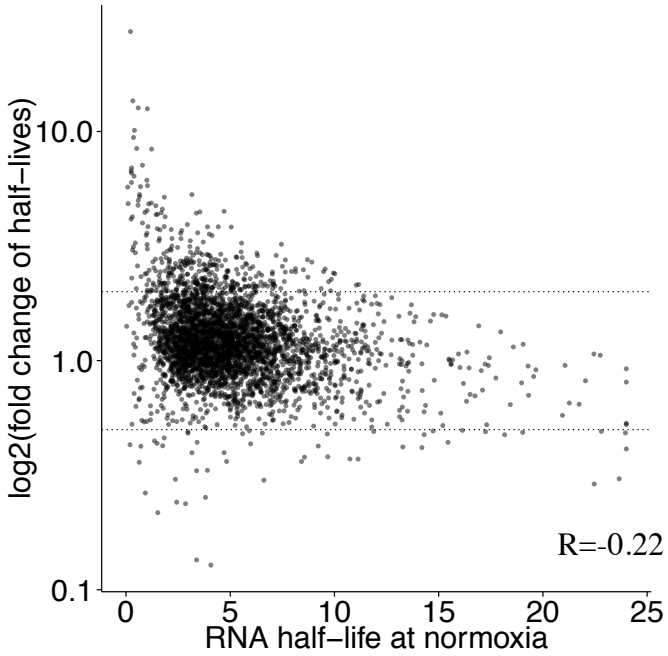


Figure 23
Comparison of original RNA half-life to the changes of RNA half-life

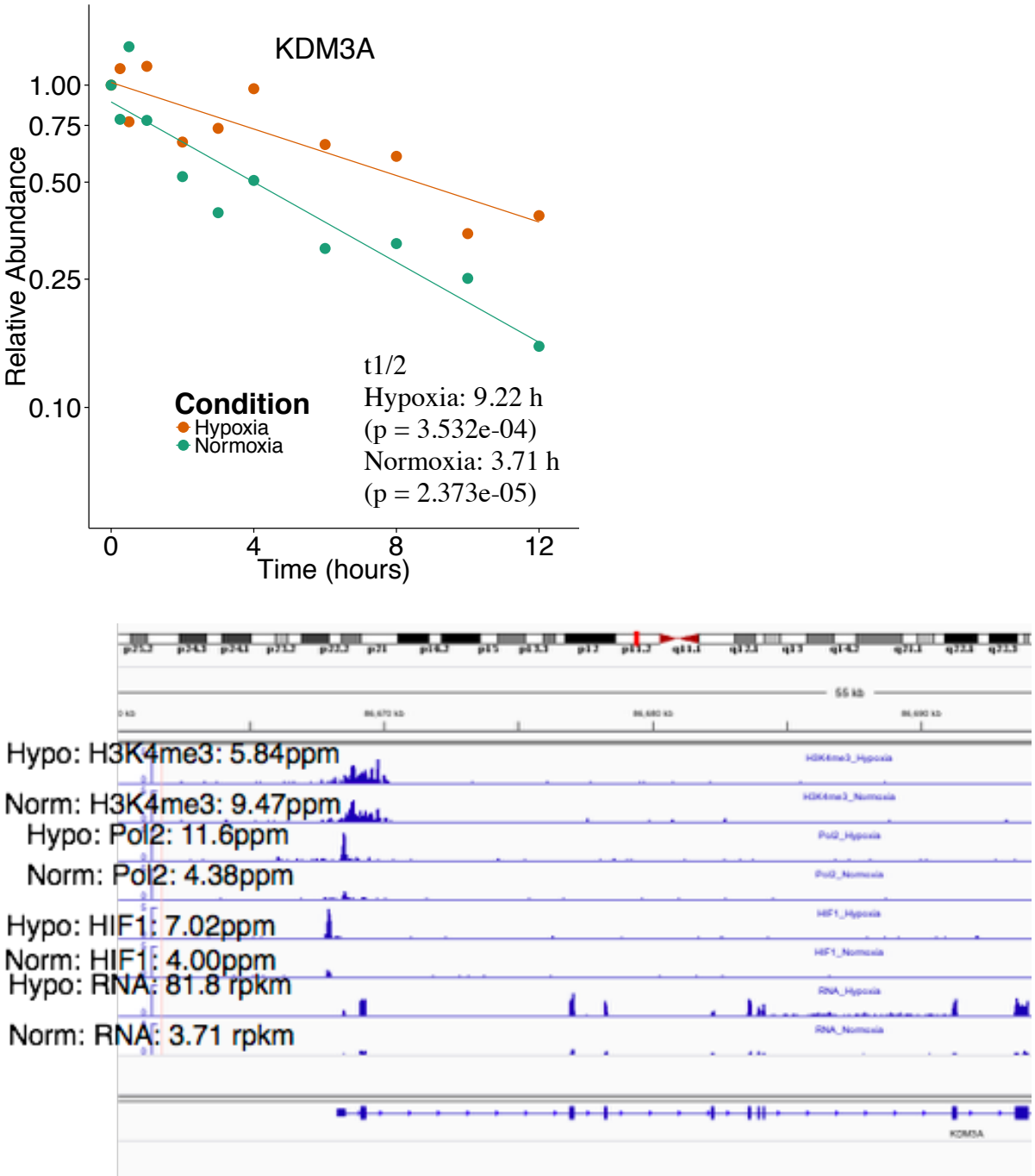


Figure 24
Example of a gene that is regulated by changes in RNA half-life (KDM3A)

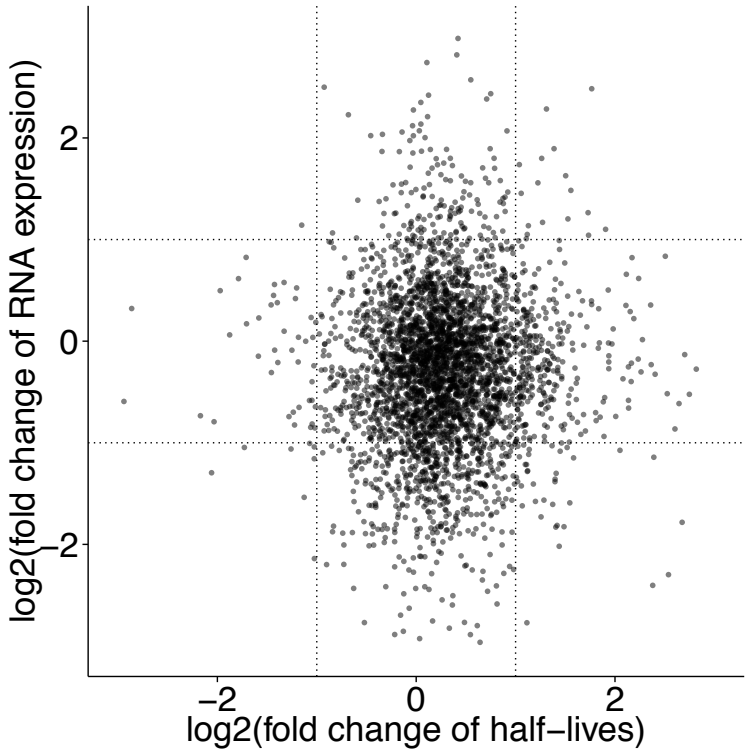
Comparison between changes in RNA expression and changes in RNA decay

In order to identify the contribution of changes in RNA decay that determine the eventual level of RNA expression, I compared the BRIC-seq based RNA decay data to the RNA-seq based eventual RNA abundance data from the previous study (Tanimoto et al., 2011). I directly compared the fold-change of RNA half-lives and RNA expression as shown in the scatterplot (Figure 25). For 179 genes that showed more than two-fold increase in the eventual RNA abundance upon hypoxia, 20 genes showed elongation of RNA half-lives by more than two-fold in hypoxia and 158 genes showed no changes in RNA half-lives, consistent with the idea that transcription is the major player in regulating the changes of gene expression in hypoxia. This includes the example of KDM3A from the previous figure, which is up-regulated by more than two-fold in the eventual RNA abundance as well as the increase in its mRNA stability.

To identify the contributions of hypoxia inducible factors, I used HIF1A ChIP-seq data in hypoxia and normoxia conditions from the previous study to identify genes that are targeted by HIF1A. HIF1A is a sequence specific transcription factor that is expected to bind to hypoxia element in *cis*-regulatory regions of genes that they transcriptionally activate in hypoxia. I used MACS software (v1.8.9) to call the peaks that significantly enrich in a particular locus in comparison with the background control. When I set the *p*-value threshold for 1×10^{-10} to call the ChIP-seq peaks, I identified 618 peaks in the whole-genome corresponding to 172 genes in hypoxia. To identify the target genes of HIF1A, I assigned HIF1A peaks to genes, if the loci of the peak was between 1kb either side of the transcription start sites (TSS). I assigned 172 peaks to genes that could be compared against RNA decay and RNA abundance, and I found that 36 genes that showed HIF1A binding showed no changes in RNA stability, in comparison

to 9 genes that showed HIF1A binding and showed elongation in RNA stability in hypoxia, suggesting that genes that are regulated through RNA decay do not tend to be regulated by HIF1A, and that they have separate mechanisms (Figure 26). In the example of KDM3A, it is regulated by both HIF-1 and RNA half-life. When I additionally analysed the epigenome, I was able to observe the increase in the levels of H3K4me3, pol2, HIF1 and RNA abundance, suggesting that both the transcription initiation and RNA decay are involved in regulating KDM3A.

In addition, I was able to identify 301 genes that show increase in RNA stability in hypoxia without changes in the eventual RNA abundance. If I assume the simplest model of regulation, the increase in the RNA half-life by two-fold will increase the RNA abundance by two-fold, if the rates of RNA generation were constant. It means that these 301 genes are under an alternative methods regulation other than RNA decay to balance the increase in stability of RNA. Interestingly, when these 301 genes were subjected to gene ontology analysis, GO terms for transcription factor activities were enriched suggesting that transcription factor undergoes tight regulatory process, which is consistent with the fact that mRNA encoding transcription factors tend to have shorter RNA half-lives, therefore they have more potential in being able to use RNA half-life. One of the example of this is ELK4, which is ETS domain containing transcription factor (Figure 27), where there are little changes to the epigenome, lack of HIF-1 binding and lack of changes to the RNA abundance, but with changes to the RNA decay (Figure 25).



Fold change of RNA expression

1	158	20
41	2881	301
8	513	38

Fold change of RNA half-life

Figure 25
Comparison of the changes to the RNA half-life and RNA abundance
Top figure shows the scatterplot between the fold change (hypoxia/normoxia) of RNA expression and RNA half-life.
Bottom figure shows the frequency of occurrence

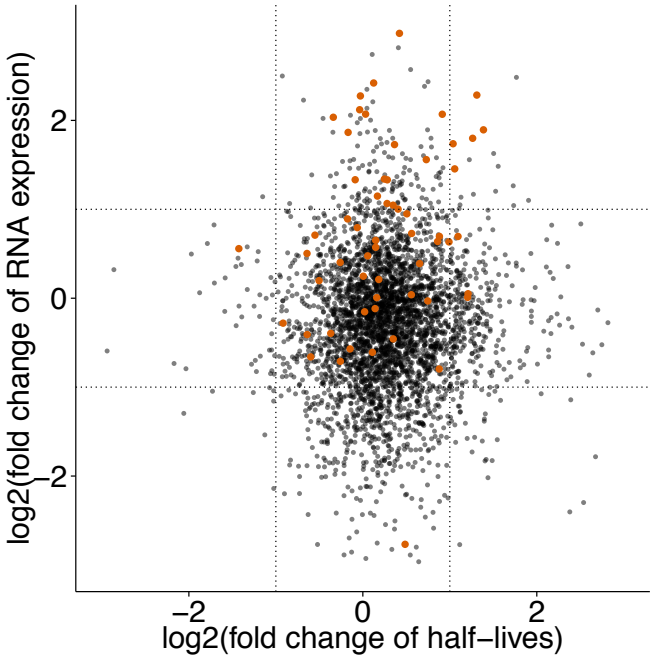


Figure 26
Comparison of the changes to the RNA half-life and RNA abundance for HIF-1 peaks
The same graph as figure 24 with genes targeted by HIF1 shown in orange.

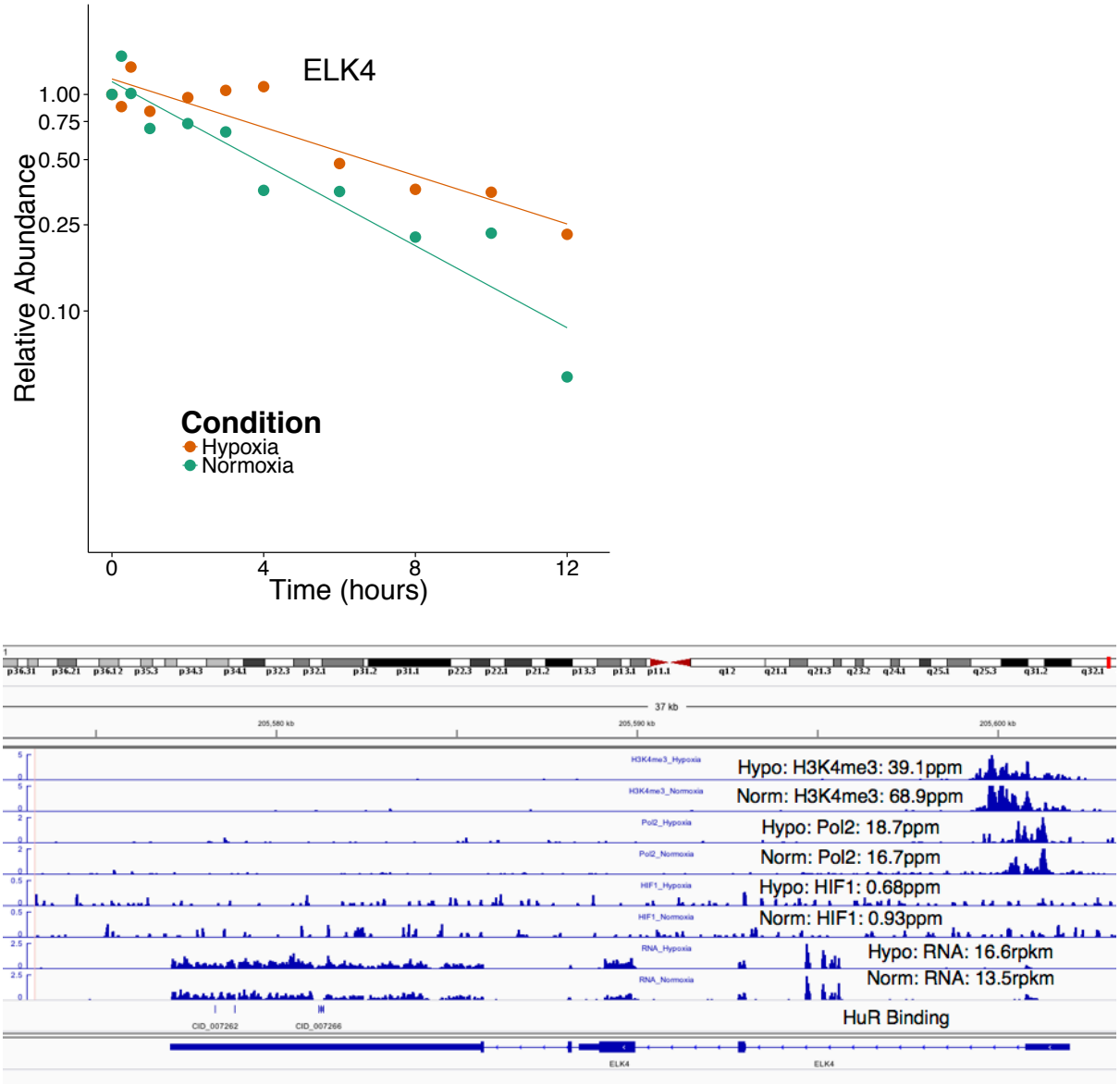


Figure 27
An example of ELK4, a gene that change in RNA stability without changes to the epigenome or the RNA abundance

Computational simulation to find the contribution of RNA decay in mediating changes in the transcriptome

I evaluated the contribution of RNA decay through simple computational simulation to predict the eventual RNA abundance, so that the contribution of the RNA decay can be inferred. In order to achieve the simulation, I simply used changes in the RNA decay constants and multiplied by the RNA abundance in normoxia (in RPKM) to estimate the RNA abundance in hypoxia. When the simulation was conducted, I observed that out of 179 genes that are up-regulated in the eventual RNA abundance by more than two-fold in hypoxia, there were 10 genes (5.6%) that could predict more than 80% of the changes in the eventual RNA abundance through changes in the RNA decay (Figure 28). In contrast, for 559 genes that are down-regulated in the eventual RNA abundance by less than 0.5-fold in hypoxia, the simulation was able to explain 133 genes (23.8%) that could explain the 80% of the changes in the eventual RNA abundance (Figure 29). This suggests that the on average the contribution of RNA decay is modest in regulating the RNA abundance, especially for genes that are up-regulated in hypoxia.

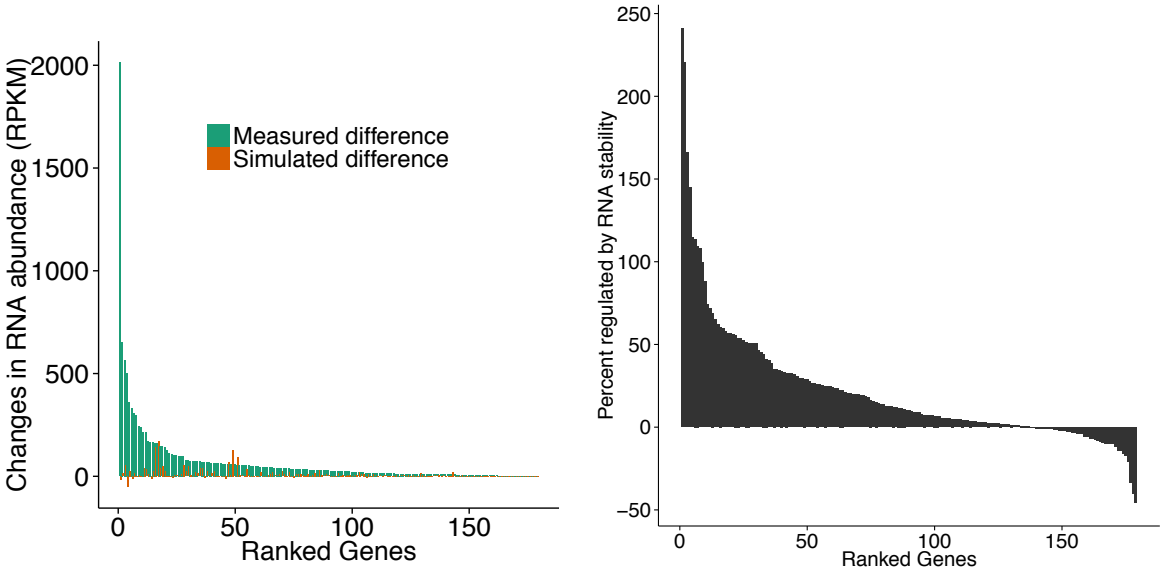


Figure 28
Simulation of effects of changes in RNA decay on RNA abundance for up-regulated genes Left panel shows the absolute change of RNA abundance and right panel shows the percent of the measured RNA difference between hypoxia and normoxia.

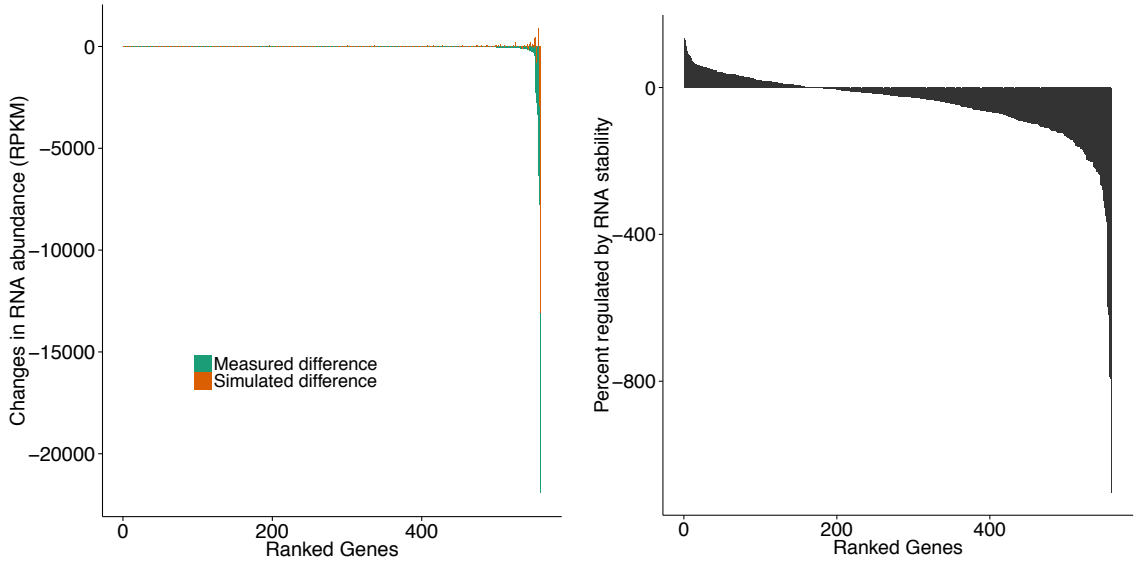


Figure 29
Simulation of effects of changes in RNA decay on RNA abundance for dowm-regulated genes
Left panel shows the absolute change of RNA abundance and right panel shows the percent of the measured RNA difference between hypoxia and normoxia.

Potential feedback mechanisms

I investigated on the regulatory mechanisms of genes that show elongation in RNA half-life without changes in the eventual RNA abundance, by investigating the enrichment of ChIP-seq peaks around the TSS of those genes. This was done by using the RegulomeDB (Boyle et al., 2012), which collates ChIP-seq of numerous transcription factors as well as other functional genomics datasets from ENCODE ChIP-seq data. When compared to the background, there were transcription factors that enrich in between -500 to +100 bp of the TSS, including E2F1 (Table 2). E2F1 belongs to the E2F family of transcription factors that activate their target gene, and E2F1, E2F2 and E2F3 are all down-regulated in hypoxia (Figure 30). This suggest that these E2F factors could play a role in the negative feedback of these target genes.

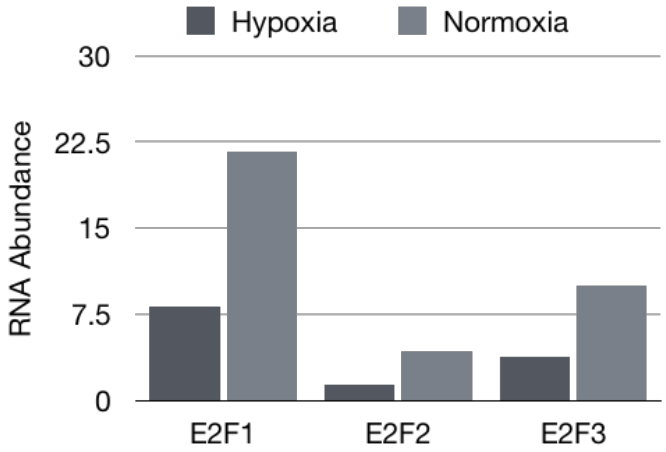


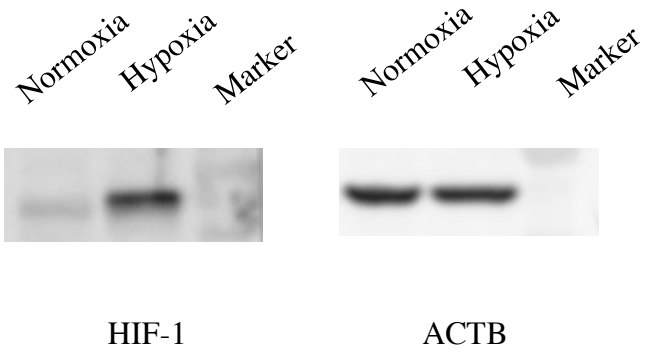
Figure 30
Expression of E2F family genes from RNA-seq data
These E2F activators are down-regulated in hypoxia and potentially targeting genes that are enriched in E2F1 binding sites.

Gene ID	Description	List	FDR
TBP	TATA-box binding protein	229	1.27E-21
ELF1	E74 like ETS transcription factor 1	222	3.98E-21
POLR2A	RNA polymerase II subunit A	273	1.01E-20
E2F1	E2F transcription factor 1	210	1.59E-20
HEY1	hrs-related family bHLH transcription factor with YRPW motif 1	217	1.04E-19

Table 2
Transcription factor enrichment for promoters of genes that show changes in RNA decay without changes in the RNA abundance
Enrichment was conducted by using the RegulomeDB (Boyle et. al. 2012) with elements that are within -500bp and +100bp of the TSS.

Gene ID	Description	List	P-value
HuR	Hu antigen binding R protein	39	1.51E-06

Table 3
RNA binding factor that may explain the elongation of RNA decay
HuR binding sites were obtained from Lebedeva et. al. 2011 and enrichment conducted against the background



Supplementary figure 1
Confirmation of the hypoxia condition through western blot

Conclusion

In this chapter I have characterized the RNA decay profiles for cancer cells that are subjected to hypoxia. Through this analysis, I was able to observe that the RNA in general is more stable in hypoxia; however, it is gene specific rather than a general form. I was able to indicate the possibility that some key genes that are up-regulated at the transcriptional initiation level is also regulated at the RNA decay level including KDM3A. I identified 361 genes that showed stabilization in its RNA upon hypoxia; however, only 20 genes showed increase in the RNA abundance as well, suggesting a potential feedback through other mechanisms. I was able to infer that HuR binding occurs for a subset of genes that are elongated in RNA half-life, suggesting the possibility of HuR mediated RNA stabilization and it included KDM3A. I was able to identify that those genes that increase in RNA stability without changes in the RNA abundance have enrichment for transcription factors including E2F1, which as an activator it is downregulated in hypoxia. This is further evidence to suggest that feedbacks do occur.

Through these analyses, I was able to start to uncover the extent of the contributions of RNA decay, which could play an important role in mediating the gene expression. As previously reported, RNA decay does not seem to be the major player in determining the absolute levels of RNA abundance nor the changes in the RNA abundance upon a stimulus. However, it is still biologically relevant as it is involved in selected subset of genes, especially transcription factor that I have highlighted here. The improvement in the resolution of the understanding of different layers of gene expression is only going to increase the current knowledge of biology and to the disease.

Supplementary tables

Cell Types	ChIP-seq Conditions	URL	Project
GM12878	H3K4me3	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeBroadHistone/wgEncodeBroadHistoneGm12878H3k04me3StdPkV2.broadPeak.gz	ENCODE
GM12878	Pol II	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeOpenChromChip/wgEncodeOpenChromChipGm12878Pol2Pk.narrowPeak.gz	ENCODE
GM12878	H3K36me3	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeBroadHistone/wgEncodeBroadHistoneGm12878H3k36me3StdPk.broadPeak.gz	ENCODE
H1-hESC	H3K4me3	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeBroadHistone/wgEncodeBroadHistoneH1hescH3k4me3StdPk.broadPeak.gz	ENCODE
H1-hESC	Pol II	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeOpenChromChip/wgEncodeOpenChromChipH1hescPol2Pk.narrowPeak.gz	ENCODE
H1-hESC	H3K36me3	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeBroadHistone/wgEncodeBroadHistoneH1hescH3k36me3StdPk.broadPeak.gz	ENCODE
K562	H3K4me3	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDC	ENCOD

		C/wgEncodeBroadHistone/wgEncodeBroadHistoneK562H3k4me3StdPk.broadPeak.gz	E
K562	Pol II	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDC/wgEncodeBroadHistone/wgEncodeBroadHistoneK562Pol2bStdPk.broadPeak.gz	ENCODE
K562	H3K36me3	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDC/wgEncodeBroadHistone/wgEncodeBroadHistoneK562H3k36me3StdPk.broadPeak.gz	ENCODE
HepG2	H3K4me3	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDC/wgEncodeBroadHistone/wgEncodeBroadHistoneHepg2H3k04me1StdPk.broadPeak.gz	ENCODE
HepG2	Pol II	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDC/wgEncodeOpenChromChip/wgEncodeOpenChromChipHepg2Pol2Pk.narrowPeak.gz	ENCODE
HepG2	H3K36me3	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDC/wgEncodeBroadHistone/wgEncodeBroadHistoneHepg2H3k36me3StdPk.broadPeak.gz	ENCODE
HUVEC	H3K4me3	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDC/wgEncodeBroadHistone/wgEncodeBroadHistoneHuvecH3k4me3StdPk.broadPeak.gz	ENCODE
HUVEC	Pol II	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDC/wgEncodeBroadHistone/wgEncodeBroadHistoneHuvecPol2bStdPk.broadPeak.gz	ENCODE
HUVEC	H3K36me	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDC	ENCOD

	3	C/wgEncodeBroadHistone/wgEncodeBroadHistoneHuvecH3k36me3StdPk.broadPeak.gz	E
MCF-7	H3K4me3	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDC/wgEncodeUwHistone/wgEncodeUwHistoneMcf7H3k4me3StdPkRep1.narrowPeak.gz	ENCODE
MCF-7	Pol II	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDC/wgEncodeOpenChromChip/wgEncodeOpenChromChipMcf7Pol2PkRep1.narrowPeak.gz	ENCODE
MCF-7	H3K36me3	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDC/wgEncodeSydhHistone/wgEncodeSydhHistoneMcf7H3k36me3bUcdPk.narrowPeak.gz	ENCODE
NHEK	H3K4me3	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDC/wgEncodeBroadHistone/wgEncodeBroadHistoneNhekH3k4me3StdPk.broadPeak.gz	ENCODE
NHEK	Pol II	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDC/wgEncodeBroadHistone/wgEncodeBroadHistoneNhekPol2bStdPk.broadPeak.gz	ENCODE
NHEK	H3K36me3	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDC/wgEncodeBroadHistone/wgEncodeBroadHistoneNhekH3k36me3StdPk.broadPeak.gz	ENCODE
DLD-1	H3K4me3	http://dbtss.hgc.jp/cgi-bin/downloader2.cgi/DLD1_H3k4me3_peaks.xls.gz	DBTSS
DLD-1	Pol II	http://dbtss.hgc.jp/cgi-bin/downloader2.cgi/DLD1_Pol2_peaks.xls.gz	DBTSS

DLD-1	H3K36me3	http://dbtss.hgc.jp/cgi-bin/downloader2.cgi/DLD1_H3k36me3_peaks.xls.gz	DBTSS
HeLaS3	H3K27Ac	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeBroadHistone/wgEncodeBroadHistoneHelas3H3k27acStdPk.broadPeak.gz	ENCODE
HeLaS3	H3K4me1	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeBroadHistone/wgEncodeBroadHistoneHelas3H3k04me1StdPk.broadPeak.gz	ENCODE

Supplementary Table 1

List of URLs of ChIP-seq used

Cell Type	RNA sub-fractionation	URL	Project
GM12878	Cell	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeCshlLongRnaSeq/wgEncodeCshlLongRnaSeqGm12878CellPapGeneGencV7.gtf.gz	ENCODE
GM12878	Nucleus	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeCshlLongRnaSeq/wgEncodeCshlLongRnaSeqGm12878NucleusPapGeneGencV7.gtf.gz	ENCODE
GM12878	Cytosol	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeCshlLongRnaSeq/wgEncodeCshlLongRnaSeqGm12878CytosolPapGeneGencV7.gtf.gz	ENCODE
H1-hESC	Cell	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeCshlLongRnaSeq/wgEncodeCshlLongRnaSeqH1hescCellPapGeneGencV7.gtf.gz	ENCODE
H1-hESC	Nucleus	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeCshlLongRnaSeq/wgEncodeCshlLongRnaSeqH1hescNucleusPapGeneGencV7.gtf.gz	ENCODE
H1-hESC	Cytosol	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeCshlLongRnaSeq/wgEncodeCshlLongRnaSeqH1hescCytosolPapGeneGencV7.gtf.gz	ENCODE
K562	Cell	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeCshlLongRnaSeq/wgEncodeCshlLongRnaSeqK562CellPapGeneGencV7.gtf.gz	ENCODE

K562	Nucleus	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeCshlLongRnaSeq/wgEncodeCshlLongRnaSeqK562NucleolusTotalGeneGencV7.gtf.gz	ENCODE
K562	Cytosol	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeCshlLongRnaSeq/wgEncodeCshlLongRnaSeqK562CytosolPapGeneGencV7.gtf.gz	ENCODE
HepG2	Cell	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeCshlLongRnaSeq/wgEncodeCshlLongRnaSeqHepg2CellPapGeneGencV7.gtf.gz	ENCODE
HepG2	Nucleus	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeCshlLongRnaSeq/wgEncodeCshlLongRnaSeqHepg2NucleusPapGeneGencV7.gtf.gz	ENCODE
HepG2	Cytosol	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeCshlLongRnaSeq/wgEncodeCshlLongRnaSeqHepg2CytosolPapGeneGencV7.gtf.gz	ENCODE
HUVEC	Cell	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeCshlLongRnaSeq/wgEncodeCshlLongRnaSeqHuvecCellPapGeneGencV7.gtf.gz	ENCODE
HUVEC	Nucleus	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeCshlLongRnaSeq/wgEncodeCshlLongRnaSeqHuvecNucleusPapGeneGencV7.gtf.gz	ENCODE
HUVEC	Cytosol	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeCshlLongRnaSeq/wgEncodeCshlLongRnaSeqHuvecCytosolPapGeneGencV7.gtf.gz	ENCODE

MCF-7	Cell	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeCshlLongRnaSeq/wgEncodeCshlLongRnaSeqMcf7CellPapGeneGencV7.gtf.gz	ENCODE
MCF-7	Nucleus	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeCshlLongRnaSeq/wgEncodeCshlLongRnaSeqMcf7NucleusPapGeneGencV10.gtf.gz	ENCODE
MCF-7	Cytosol	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeCshlLongRnaSeq/wgEncodeCshlLongRnaSeqMcf7CytosolPapGeneGencV10.gtf.gz	ENCODE
NHEK	Cell	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeCshlLongRnaSeq/wgEncodeCshlLongRnaSeqNhekCellPapGeneGencV7.gtf.gz	ENCODE
NHEK	Nucleus	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeCshlLongRnaSeq/wgEncodeCshlLongRnaSeqNhekNucleusPapGeneGencV7.gtf.gz	ENCODE
NHEK	Cytosol	ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeCshlLongRnaSeq/wgEncodeCshlLongRnaSeqNhekCytosolPapGeneGencV7.gtf.gz	ENCODE
DLD-1	Cell	http://dbtss.hgc.jp/cgi-bin/downloader2.cgi/DLD1_RNAseq.wig.gz	DBTSS

Supplementary table2

List of URLs of RNA-seq used

Name	Sense sequece (5'-3')	Antisense sequece (5'-3')
Control siRNA-1	GTACCTGACTAGTCGCAGAA G	TCTGCGACTAGTCAGGTACG G
Control siRNA-2	UUCUCCGAACGUGUCACGUT T	ACGUGACACGUUCGGAGAAT T
UPF1 siRNA-1	GAUGCAGUUCCGCUCCAUUd TdT	AAUGGAGCGGAACUGCAUCd TdT
UPF1 siRNA-2	AAUUUCUGUAACUUGUUUCC U	GAAACAAGUUACAGAAAUUA C
EXOSC5 siRNA-1	CAACACGUCUUCCGUUUCUd TdT	AGAAACGGAAGACGuGuuGdT dT
EXOSC5 siRNA-2	GCAAAGAGAUUUUCAACAAd TdT	UUGUUGAAAAUCUCUUUGCd TdT
STAU1 siRNA-1	CUCUGCGUGUGGUCCGUAUG G	AUACGGACCACACGCAGAGC C
STAU1 siRNA-2	CAGGGGAUCAAUCCGAUUAG C	UAAUCGGAUUGAUCCCCUGG C

Supplementary table 3

Sequences of siRNA used

Name	Sense sequece (5'-3')	Antisense sequece (5'-3')
GAPDH	GCACCGTCAAGGCTGAGAAC	TGGTGAAGACGCCAGTGGA
UPF1	AGATCACGGCACAGCAGAT	TGGCAGAAGGGTTTTCCTT
EXOSC5	CCACACTCGAAGTGATCCTG	CCGGCTCTTCTCTGCAAC
STAU1	TATCGGCAAGGATGTGGAGT	TGGTCCAACTCAGACAGCAA
HIC1	GATGCTGGACACGATGGA	CTTGGTGCGCTGGTTGTT
PVT1	CTCTTCCTGGTGAAGCATCTG	ATGGCTGTATGTGCCAAGGT
XLOC_003734	TTTCAGAGGTTCTTTAGGGAAAAG	ACGTGCAAGAGGTCAAAGAAC
XLOC_007604	GGACTGAGGCAACCCATCTA	TTCAAGATGAACCTTATGAGTGGT
XLOC_013194	GGGGGCTCTGTGATATGCTA	TTCCTTTTCCTCTTGGGTTTAGT
XLOC_000048	TGGAATATGGCTAATGTAAAGTTCA	GCCCGTTCTTGTGGTAGAAG

Supplementary table 4

Sequences of primer pairs used for qRT-PCR

Sample Name	Sequencing Depth	Uniquely Mapped Reads	Number of peaks
H3K4me3	29607438	22520220	13829
PolII	186897451	139395594	11395
Input	71328884	54200494	

Supplementary table 5

Sequence statistics for ChIP-seq data

Sample Name	Sequencing Depth	Uniquely Mapped Reads
Basal	38260387	16146827
siCTRL_1	34070704	15353237
siCTRL_2	35590030	15965080
siUPF1_1	30991212	15029160
siUPF1_2	41238364	17146466
siCTRL_1	46202054	23153491
siCTRL_2	52056079	24794738
siEXOSC5_1	35065084	17853409
siEXOSC5_2	40062193	20404942
siCTRL	36876205	29287082
siSTAU1_1	30256722	23896759
siSTAU1_2	36353208	29049849

Supplementary table 6

Sequence statistics for RNA-seq data

Sample Name	Sequencing Depth	Uniquely Mapped Reads
siCTRL_1 0h	32113983	11795163
siCTRL_1 4h	27332602	8410082
siCTRL_1 8h	29910113	8223544
siCTRL_1 12h	29209782	4683418
siUPF1 0h	35319906	11151822
siUPF1 4h	40875119	7125974
siUPF1 8h	32442201	6047454
siUPF1 12h	30350033	4353768
siCTRL_1 0h	16889847	10067525
siCTRL_1 4h	14850192	8384827
siCTRL_1 8h	16570098	8854992
siCTRL_1 12h	11755236	6132423
siCTRL_1 24h	14667835	5230211
siEXOSC5 0h	18576472	9683224
siEXOSC5 4h	17433615	8799438
siEXOSC5 8h	15412791	7458473
siEXOSC5 12h	17570157	9208784
siEXOSC5 24h	31969879	13096031
siCTRL 0min	45023135	10594823

siCTRL 15min	59225958	13936994
siCTRL 45min	50683722	12952619
siCTRL 75min	28327195	6211525
siCTRL 105min	48647819	11909083
siCTRL 165min	40151717	8021772
siCTRL 225min	33619287	7437648
siCTRL 345min	38757947	6494652
siCTRL 465min	60184115	8497967
siCTRL 585min	59306310	7014734
siCTRL 705min	54748461	6036560
siSTAU1 0min	40475633	6053372
siSTAU1 15min	38958412	5826463
siSTAU1 45min	63492284	9828708
siSTAU1 75min	53641757	8666838
siSTAU1 105min	38332782	6163057
siSTAU1 165min	48164040	6399917
siSTAU1 225min	26029128	3633723
siSTAU1 345min	41558057	4375736
siSTAU1 465min	55302129	4631584
siSTAU1 585min	31968358	1731351
siSTAU1 705min	67427989	3747637

Supplementary table 7

Sequence statistics for BRIC-seq

RPKM	Total	No peak	%	Low peak	%	High peak	%
Total	18853	7737	41%	2732	14.5%	8384	44.5%
> 0.001	15251	4245	27.8%	2652	17.4%	8354	54.8%
> 0.005	15222	4221	27.7%	2650	17.4%	8351	54.9%
> 0.01	15146	4153	27.4%	2645	17.5%	8348	55.1%
> 0.05	14159	3282	23.2%	2558	18.1%	8319	58.8%
> 0.1	13493	2730	20.2%	2478	18.4%	8285	61.4%
> 0.5	11491	1348	11.7%	2095	18.2%	8048	70%
> 1	10421	957	9.2%	1808	17.3%	7656	73.5%
> 5	6736	315	4.7%	947	14.1%	5474	81.3%
> 10	4848	175	3.6%	600	12.4%	4073	84%
> 50	1488	39	2.6%	108	7.3%	1341	90.1%
> 100	775	20	2.6%	40	5.2%	715	92.3%

Supplementary table 8

Distribution of ChIP-seq and RNA-seq data

	Total	**ChIP(-)/RNA(-)**	**ChIP(-)/RNA(+)**	**ChIP(+)/RNA(-)**	**ChIP(+)/RNA(+) (Total)**	**ChIP(+)/RNA(+) (×2)**	**ChIP(+)/RNA(+) (×1.1)**
Number of genes	12,479	6,235	603	2,745	2,896	1,617	187
Median half-life (hours)	nd	nd	13.2	6	11.6	11	10.9

Supplementary table 9

RNA half-life statistics

Cell Type	GO Term	GO Name	Number	P-value
GM12878	GO:0003677	DNA binding	21	1.5E-02
H1-hESC	GO:0003677	DNA binding	211	2.6E-32
HUVEC	GO:0003677	DNA binding	222	6.7E-13
K562	GO:0008270	zinc ion binding	116	3.9E-05
HepG2	GO:0005886	plasma membrane	15	3.9E-02
MCF7	GO:0008270	zinc ion binding	59	1.2E-08
NHEK	GO:0003677	DNA binding	177	2.3E-04
DLD-1	GO:0006355	regulation of transcription, DNA-templated	19	4.8E-02

Supplementary table 10

List of GOs enriched for ChIP+/RNA- genes in ENCODE cell line

Mapped reads	Hypoxia	Normoxia
0h	8,973,103	4,840,754
0.25h	12,356,596	3,234,528
0.5h	11,020,498	3,708,792
1h	9,788,648	3,818,593
2h	13,408,793	3,463,636
3h	14,603,126	5,137,848
4h	18,782,748	6,064,246
6h	14,055,035	3,379,641
8h	10,818,309	4,856,825
10h	12,223,632	4,711,979
12h	14,859,722	3,478,927
Average	12,808,201	4,159,841

Supplementary table 11

Sequencing statistics for BRIC-seq in hypoxia

Bibliography

Arraiano, C.M., Mauxion, F., Viegas, S.C., Matos, R.G., and Seraphin, B. (2013). Intracellular ribonucleases involved in transcript processing and decay: precision tools for RNA. Biochim Biophys Acta *1829*, 491-513.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., *et al*. (2000). Gene Ontology: tool for the unification of biology. Nature Genetics *25*, 25-29.

Ben-Tabou de-Leon, S., and Davidson, E.H. (2009). Modeling the dynamics of transcriptional gene regulatory networks for animal development. Dev Biol *325*, 317-328.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B-Methodological *57*, 289-300.

Boyle, A.P., Hong, E.L., Hariharan, M., Cheng, Y., Schaub, M.A., Kasowski, M., Karczewski, K.J., Park, J., Hitz, B.C., Weng, S., *et al*. (2012). Annotation of functional variation in personal genomes using RegulomeDB. Genome Res *22*, 1790-1797.

Bujold, D., Morais, D.A., Gauthier, C., Cote, C., Caron, M., Kwan, T., Chen, K.C., Laperle, J., Markovits, A.N., Pastinen, T., *et al*. (2016). The International Human Epigenome Consortium Data Portal. Cell Syst *3*, 496-499 e492.

Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., and Rinn, J.L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes & Development *25*, 1915-1927.

Chlebowski, A., Lubas, M., Jensen, T.H., and Dziembowski, A. (2013). RNA decay machines: The exosome. *1829*, 552-560.

Dong, X., Greven, M.C., Kundaje, A., Djebali, S., Brown, J.B., Cheng, C., Gingeras, T.R., Gerstein, M., Guigó, R., Birney, E., *et al*. (2012). Modeling gene expression using chromatin features in various cellular contexts. Genome Biology *13*, R53.

Du, Z.M., Hu, L.F., Wang, H.Y., Yan, L.X., Zeng, Y.X., Shao, J.Y., and Ernberg, I. (2011). Upregulation of MiR-155 in nasopharyngeal carcinoma is partly driven by LMP1 and LMP2A and downregulates a negative prognostic marker JMJD1A. PLoS One *6*, e19137.

Garneau, N.L., Wilusz, J., and Wilusz, C.J. (2007). The highways and byways of mRNA decay. Nature Reviews Molecular Cell Biology *8*, 113-126.

Godwin, A.R., Kojima, S., Green, C.B., and Wilusz, J. (2013). Kiss your tail goodbye: the role of PARN, Nocturnin, and Angel deadenylases in mRNA biology. Biochim Biophys Acta *1829*, 571-579.

Imamachi, N., Tani, H., and Akimitsu, N. (2012a). Up-frameshift protein 1 (UPF1): multitalented entertainer in RNA decay. Drug Discov Ther *6*, 55-61.

Imamachi, N., Tani, H., and Akimitsu, N. (2012b). Up-frameshift protein 1 (UPF1): multitalented entertainer in RNA decay. Drug Discoveries & Therapeutics *6*, 55-61.

Imamachi, N., Tani, H., Mizutani, R., Imamura, K., Irie, T., Suzuki, Y., and Akimitsu, N. (2014). BRIC-seq: A genome-wide approach for determining RNA stability in mammalian cells. Methods *67*, 55-63.

Kanai, A., Suzuki, K., Tanimoto, K., Mizushima-Sugano, J., Suzuki, Y., and Sugano, S. (2011). Characterization of STAT6 target genes in human B cells and lung epithelial cells. DNA research : an international journal for rapid publication of reports on genes and genomes *18*, 379-392.

Karlic, R., Chung, H.R., Lasserre, J., Vlahovicek, K., and Vingron, M. (2010). Histone modification levels are predictive for gene expression. Proceedings of the National Academy of Sciences of the United States of America *107*, 2926-2931.

Keith, B., Johnson, R.S., and Simon, M.C. (2011). HIF1α and HIF2α: sibling rivalry in hypoxic tumour growth and progression. *12*, 9-22.

Keith, B., and Simon, M.C. (2007). Hypoxia-inducible factors, stem cells, and cancer. Cell *129*, 465-472.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology *14*, R36.

Kim, H.D., Shay, T., O'Shea, E.K., and Regev, A. (2009). Transcriptional regulatory circuits: predicting numbers from alphabets. Science *325*, 429-432.

Kim, J.-w., Tchernyshyov, I., Semenza, G.L., and Dang, C.V. (2006). HIF-1-mediated expression of pyruvate dehydrogenase kinase: A metabolic switch required for cellular adaptation to hypoxia. Cell Metabolism *3*, 177-185.

Kim, Y.K., Furic, L., Parisien, M., Major, F., DesGroseillers, L., and Maquat, L.E. (2007). Staufen1 regulates diverse classes of mammalian transcripts. *26*, 2670-2681.

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

Komili, S., and Silver, P.A. (2008). Coupling and coordination in gene expression processes: a systems biology view. Nature Reviews Genetics *9*, 38-48.

Kurosu, T., Ohga, N., Hida, Y., Maishi, N., Akiyama, K., Kakuguchi, W., Kuroshima, T., Kondo, M., Akino, T., Totsuka, Y., *et al*. (2011). HuR keeps an angiogenic switch on by stabilising mRNA of VEGF and COX-2 in tumour endothelium. Br J Cancer *104*, 819-829.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods *9*, 357-359.

Lebedeva, S., Jens, M., Theil, K., Schwanhäusser, B., Selbach, M., Landthaler, M., and Rajewsky, N. (2011). Transcriptome-wide Analysis of Regulatory Interactions of the RNA-Binding Protein HuR. *43*, 340-352.

Lendahl, U., Lee, K.L., Yang, H., and Poellinger, L. (2009). Generating specificity and diversity in the transcriptional response to hypoxia. Nature Reviews Genetics *10*, 821-832.

Levy, N.S. (1998). Hypoxic Stabilization of Vascular Endothelial Growth Factor mRNA by the RNA-binding Protein HuR. *273*, 6417-6423.

Levy, N.S., Chung, S., Furneaux, H., and Levy, A.P. (1998). Hypoxic Stabilization of Vascular Endothelial Growth Factor mRNA by the RNA-binding Protein HuR. Journal of Biological Chemistry *273*, 6417-6423.

Maekawa, S., Imamachi, N., Irie, T., Tani, H., Matsumoto, K., Mizutani, R., Imamura, K., Kakeda, M., Yada, T., Sugano, S., *et al*. (2015). Analysis of RNA decay factor mediated RNA stability contributions on RNA abundance. BMC Genomics *16*, 154.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., Koche, R.P., *et al*. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *448*, 553-560.

Mimura, I., Nangaku, M., Kanki, Y., Tsutsumi, S., Inoue, T., Kohro, T., Yamamoto, S., Fujita, T., Shimamura, T., Suehiro, J., *et al*. (2012). Dynamic change of chromatin conformation in response to hypoxia enhances the expression of GLUT3 (SLC2A3) by cooperative interaction of hypoxia-inducible factor 1 and KDM3A. Mol Cell Biol *32*, 3018-3032.

Mizukami, Y., Jo, W.-S., Duerr, E.-M., Gala, M., Li, J., Zhang, X., Zimmer, M.A., Iliopoulos, O., Zukerberg, L.R., Kohgo, Y., *et al*. (2005). Induction of interleukin-8 preserves the angiogenic response in HIF-1α–deficient colon cancer cells. Nature Medicine, 1-7.

Nagarajan, V.K., Jones, C.I., Newbury, S.F., and Green, P.J. (2013). XRN 5'-->3' exoribonucleases: structure, mechanisms and functions. Biochim Biophys Acta *1829*, 590-603.

Porrua, O., and Libri, D. (2013). RNA quality control in the nucleus: The Angels' share of RNA. 1-8.

Pouysségur, J., Dayan, F., and Mazure, N.M. (2006). Hypoxia signalling in cancer and approaches to enforce tumour regression. *441*, 437-443.

Pruitt, K.D., Tatusova, T., Brown, G.R., and Maglott, D.R. (2012). NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. Nucleic Acids Research *40*, D130-D135.

R Core Team (2012). R: A language and environment for statistical computing (R Foundation for Statistical Computing, Vienna, Austria. ISBN).

Rabani, M., Levin, J.Z., Fan, L., Adiconis, X., Raychowdhury, R., Garber, M., Gnirke, A., Nusbaum, C., Hacohen, N., Friedman, N., *et al*. (2011). Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. Nature Biotechnology *29*, 436-442.

Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., *et al*. (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317-330.

Roeder, R. (1996). The role of general initiation factors in transcription by RNA polymerase II. Trends in Biochemical Sciences *21*, 327-335.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature *473*, 337-342.

Schweingruber, C., Rufener, S.C., Zünd, D., Yamashita, A., and Mühlemann, O. (2013). Nonsense-mediated mRNA decay - mechanisms of substrate mRNA recognition and degradation in mammalian cells. Biochimica et biophysica acta *1829*, 612-623.

Semenza, G.L. (2010). HIF-1: upstream and downstream of cancer metabolism. Current Opinion in Genetics & Development *20*, 51-56.

Semenza, G.L. (2014). Oxygen sensing, hypoxia-inducible factors, and disease pathophysiology. Annu Rev Pathol *9*, 47-71.

Stamatoyannopoulos, J.A., Dutta, A., Guigó, R., Thurman, R.E., Kuehn, M.S., Neph, S., Asthana, S., Malhotra, A., Adzhubei, I., Greenbaum, J.A., *et al*. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *447*, 799-816.

Stoecklin, G., and Muhlemann, O. (2013). RNA decay mechanisms: specificity through diversity. Biochim Biophys Acta *1829*, 487-490.

Tani, H., Imamachi, N., Salam, K.A., Mizutani, R., Ijiri, K., Irie, T., Yada, T., Suzuki, Y., and Akimitsu, N. (2012a). Identification of hundreds of novel UPF1 target transcripts by direct determination of whole transcriptome stability. RNA Biol *9*, 1370-1379.

Tani, H., Mizutani, R., Salam, K.A., Tano, K., Ijiri, K., Wakamatsu, A., Isogai, T., Suzuki, Y., and Akimitsu, N. (2012b). Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. *22*, 947-956.

Tanimoto, K., Tsuchihara, K., Kanai, A., Arauchi, T., Esumi, H., Suzuki, Y., and Sugano, S. (2011). Genome-wide identification and annotation of HIF-1α binding sites in two cell lines using massively parallel sequencing. The HUGO Journal *4*, 35-48.

The ENCODE Project Consortium, Bernstein, B.E., Birney, E., Dunham, I., Green, E.D., Gunter, C., and Snyder, M. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57-74.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature Biotechnology *28*, 516-520.

Tsuchihara, K., Suzuki, Y., Wakaguri, H., Irie, T., Tanimoto, K., Hashimoto, S.i., Matsushima, K., Mizushima-Sugano, J., Yamashita, R., Nakai, K., *et al*. (2009). Massive transcriptional start site analysis of human genes in hypoxia cells. Nucleic Acids Research *37*, 2249-2263.

Wahle, E., and Winkler, G.S. (2013). RNA decay machines: deadenylation by the Ccr4-not and Pan2-Pan3 complexes. Biochim Biophys Acta *1829*, 561-570.

Wang, C., Tian, R., Zhao, Q., Xu, H., Meyer, C.A., Li, C., Zhang, Y., and Liu, X.S. (2012). Computational inference of mRNA stability from histone modification and transcriptome profiles. Nucleic Acids Research *40*, 6414-6423.

Wang, R.S., Zhang, X.S., and Chen, L. (2007). Inferring transcriptional interactions and regulator activities from experimental data. Mol Cells *24*, 307-315.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., *et al*. (2008). Model-based Analysis of ChIP-Seq (MACS). Genome Biology *9*.

Acknowledgements

I would like to thank my supervisors Prof. Sumio Sugano and Prof. Yutaka Suzuki for their supervision during the years that I have spent in their laboratory. I am very grateful for their support as without their supervision, this thesis and research will not have been possible.

I would like to thank Prof. Nobuyoshi Akimitsu as a collaborator on this thesis and also as a thesis committee member. I am very grateful for his large contribution to the work especially with his help in the experimental side of the thesis and his large contribution to the discussion.

I would like to thank Dr. Takuma Irie, Dr. Kyoko Matsumoto (Graduate School of Frontier Sciences), Dr. Naoto Imamachi, Dr. Rena Mizutani, Dr. Katsutoshi Imamura, Ms. Miho Kakeda (Isotope Science Center), Dr. Hidenori Tani (National Institute of Advanced Institute of Science and Technology), Prof. Tetsushi Yada (Kyushu Institute of Technology) for their valuable contribution to the thesis and the publication through experimental work and analysis as well as valuable discussion.

I would like to thank Prof. Kenta Nakai (Institute of Medical Sciences), Prof. Koichi Matusda and Prof. Kozo Tomita for their invaluable comments and discussion on the thesis as a thesis committee member.

I would like to thank Dr. Katsuya Tsuchihara (National Cancer Center East) and Dr. Hideaki Makinoshima (National Cancer Center East) for their help and discussion in cancer and hypoxia biology.

I would like to thank all of the current and past lab members of the Sugano lab and Suzuki lab, who guided me through the time as a student.

I would also like to thank my funding sources. Japan Society of the Promotion of Sciences (JSPS) has provided my living support through Fellowship for Young Scientists and my

research funding through KAKENHI, as well as support from the Global Leaders Program for Social Design and Management (GSDM).

I would like to thank my family and friends who have supported me through the good and the bad times through my education. Finally, I would like to thank God for all his grace and guidance.