Identification of mutated genes containing premature termination codon and their translational suppression after nonsense-mediated mRNA decay block

Nara Shin

Department of Medical Science

The Graduate School, Yonsei University

Identification of mutated genes containing premature termination codon and their translational suppression after nonsense-mediated mRNA decay block

Nara Shin

Department of Medical Science

The Graduate School, Yonsei University

Identification of mutated genes containing premature termination codon and their translational suppression after nonsense-mediated mRNA decay block

Directed by Professor Hoguen Kim

The Doctoral Dissertation
submitted to the Department of Medical Science,
the Graduate School of Yonsei University
in partial fulfillment of the requirements for the degree
of Doctor of Philosophy

Nara Shin
June 2012

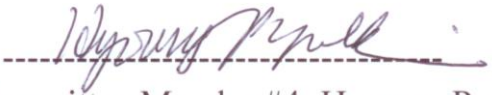
This certifies that the Doctoral Dissertation of Nara Shin is approved.

Thesis Supervisor: Hoguen Kim

Thesis Committee Member#1: Jeon Han Park

Thesis Committee Member#2: Kyung-Sup Kim

Thesis Committee Member#3: Ho-Geun Yoon



Thesis Committee Member#4: Hyoung-Pyo Kim

The Graduate School
Yonsei University

June 2012

ACKNOWLEDGEMENTS

학문에 조금 더 접근해 보고자 학위를 시작해야겠다 생각했고 지금까지 왔지만, 박사를 마무리하는 이 시점에서도 갈 길은 멀며 저는 아직 많이 부족하다는 생각이 듭니다. 그렇지만 이러한 저를 이 자리로 이끌어 주시고 힘들 때 버팀목이 되시고 기대를 가지고 나아가게 하시는 하나님께 정말 깊은 감사를 드립니다.

이 곳에서 연구할 수 있는 기회를 주시고 또한 지도하여 주신 김호근 교수님께 감사드립니다. 부족한 저를 인내하여 주시고 따뜻한 격려와 조언으로 지도해주신 박전한 교수님, 김경섭 교수님, 윤호근 교수님, 김형표 교수님께도 진심으로 감사드립니다. 서로 다른 주제로 연구하고 있지만, 디스커션에 힘써 준 Y3 디스커션 모임인 한나, 원석, 창훈, 종주, 욱진 에게도 감사의 말을 전하며, 멀리 메릴랜드에서 부족한 질문에도 정성껏 답해주신 길호 외삼촌께도 이 자리를 빌어 진심으로 감사드립니다. 그리고 4 년 동안 함께 해준 실험실 친구들과, 따뜻한 격려와 조언으로 함께 해준 지희, 민지, 김상겸 선생님, 민경언니와 현진이에게도 감사의 말을 전합니다.

마지막으로, 학위를 진행하며 어렵다 생각들 때, 한결 같은 응원으로 큰 힘이 되어주신 양가 부모님들과 가족, 그리고 사랑하는 남편 김윤태에게 진심으로 감사하다는 말을 전하며 모든 도움에 감사하는 마음으로 앞으로 더욱 힘써 한동에서의 가르침 '사랑, 겸손, 봉사' 하는 삶을 살 수 있도록 노력하겠습니다.

신 나 라 올림

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Identification of mutated genes containing premature termination codon and their translational suppression after nonsense-mediated mRNA decay block

Nara Shin

Department of Medical Science
The Graduate School, Yonsei University

(Directed by Professor Hoguen Kim)

Frameshift mutations at coding mononucleotide repeats (cMNR) are frequent in cancers of high-microsatellite instability (MSI-H). Frameshift mutations in cMNR result in the formation of a premature termination codon (PTC) in the transcribed mRNA, and these abnormal mRNAs are generally degraded by nonsense mediated mRNA decay (NMD). In this study, 12 novel genes were identified that are frequently mutated at their cMNR by blocking NMD in two MSI-H cancer cell lines. After blocking NMD, differentially-expressed genes were screened using DNA microarrays, and then database analysis was used to select 28 candidate genes containing cMNR with more than 9 nucleotide repeats. Among them, mutations at cMNR of 15 genes have not been previously reported in cancers. Mutations at cMNR of each of the 15 genes in 10 MSI-H cell lines and 21 MSI-H cancers were analyzed, and frequent mutations of 12 genes in MSI-H cell lines and cancers were found, but not in microsatellite stable (MSS) cancers. In addition, these mutated

genes are degraded by NMD and protein expressions are down-regulated in MSI-H cancers. Although NMD is an efficient mechanism that down-regulate the PTC-containing mRNA at post-transcription level, there are several stressors to inhibit NMD *in vivo*. Thus, down-regulated transcripts by NMD can be recovered and might produce truncated proteins. To clarify whether truncated proteins are produced or not, β-globin modified vector system was generated which has been used as a conventional model construct to study NMD mechanism. When NMD was blocked using siRNA of h*UPF1*, three different drugs, and hypoxic condition, mutant proteins were rarely detected, although PTC-containing transcripts were sufficiently recovered. To verify this discrepancy, several possibilities including translation repression, rapid degradation, and insufficiency of recovered transcript, were checked by using polysome analysis, proteasomal inhibition by MG132, and quantitation of their mRNA expressions, respectively. These experiments indicated that the translations are repressed from PTC-containing transcripts which are recovered by NMD inhibition. In addition, eIF4A3 was involved in this translation repression of PTC-containing mRNAs. All of these results demonstrated that PTC-containing mRNA derived from frameshift mutation is inhibited in the generation of mutant protein production via NMD. Although mutant mRNA expression is recovered by NMD inhibition through cellular stressors, translation repression is developed, and therefore harmful cellular changes by truncated proteins are prevented.

Key words: coding mononucleotide repeat, frameshift mutation, premature termination codon, nonsense-mediated mRNA decay, translation repression, h*eIF4A3*

Identification of mutated genes containing premature termination codon and their translational suppression after nonsense-mediated mRNA decay block

Nara Shin

Department of Medical Science
The Graduate School, Yonsei University

(Directed by Professor Hoguen Kim)

I. INTRODUCTION

In mammalian cells, mRNAs containing premature termination codons (PTCs) due to a nonsense or frameshift mutation are typically degraded by nonsense-mediated mRNA decay (NMD). Generally, NMD is mediated through the recognition of the PTC-containing mRNA, which is recognized by its position relative to the last exon-exon junction. In brief, mammalian transcripts that contain a PTC more than 50 nucleotides upstream of the last exon-exon junction are subjected to NMD[1-3].

Central to the NMD pathway, exon-junction complexes (EJC) and UPF complex play key roles. EJC is deposited just upstream of exon-exon junctions during splicing[4]. EJC is composed with heterogenous protein composition, among these proteins eIF4A3, Y14, MAGOH, and Barentsz (Btz) are core complexes[5,6]. When translation termination occurs at PTC,

eukaryotic release factors eRF1 and eRF3 interact with UPF1 and SMG1 at the A-site of the terminating ribosome[7]. This SMG1-UPF1-eRF (SURF) complex associates with UPF2 which is recruited to EJC, and sequential phosphorylation of UPF1 by PI3-kinase-like SMG1 kinase recruits decay molecules[7].

Microsatellites, short repetitive sequences of various types and lengths[8], are distributed non-randomly throughout the genome[9]. A subset of carcinomas exhibits length alterations in several coding as well as non-coding microsatellites, a molecular phenotype commonly referred to as high-microsatellite instability (MSI-H)[10]. The MSI-H pathway begins with the inactivation of one of a group of genes responsible for DNA nucleotide mismatch repair, which leads to extensive mutations in both repetitive and non-repetitive DNA sequences, low frequencies of allelic losses, and rare alterations in tumor DNA content[11,12]. The mechanism of tumorigenesis in MSI-H tumors is length alterations of microsatellite repeats (frameshift mutations) within the coding regions of affected genes, and the inactivation of these genes is believed to contribute directly to tumor development and progression[13,14]. The frameshift mutations at coding microsatellites of target genes invariably result in the formation of premature termination codon (PTC), and these PTC-containing mRNAs are subject to degradation at the mRNA level. The identification of frequently mutated genes and measurement of their expression is essential for the diagnostic and therapeutic application of these genes in MSI-H cancers[15-18].

According to the recent reports, there are various physiological and environmental stressors to inhibit NMD machinery, such as hypoxia[19,20]. These cellular changes could be occurred frequently, especially in the tumor. The growth and a faulty vascular system of tumor cause significant cellular stresses, hypoxia, and amino acid deprivation. Expected following consequence is recovering NMD target transcripts, and these PTC-containing

transcripts are able to generate truncated proteins which have the potential to be deleterious. A fundamental question is whether truncated proteins are produced or down-regulated by certain mechanisms from elevated PTC-containing mRNAs, in NMD inhibited condition.

In this study, the goal was identification of frameshift mutated genes in MSI-H cancer cell lines by NMD block and validation of potential to produce truncated proteins from these endogenous targets under certain circumstances. Newly identified 12 genes are frequently mutated in MSI-H cell lines and cancers, and their expressions are down-regulated by NMD. I have also demonstrated that recovered PTC-containing NMD target transcript due to NMD inhibition is regulated at translation level and eIF4A3 is involved in this translation repression.

II. MATERIALS AND METHODS

1. Cell lines and tumor tissues

To prepare cells for microarray experiments, MSI-H colorectal carcinoma cell lines, HCT116 and RKO, and MSS colorectal carcinoma cell line, SW480, were maintained in RPMI (HyClone, Logan, UT, USA) containing 10% FBS (Invitrogen, Carlsbad, CA, USA), according to ATCC guidelines. For mutation analysis, MSI-H colorectal carcinoma cell lines, DLD1, HCT116, HCT8, LoVo, LS174T, RKO, SNU C2A, SNU C2B, SNU C4, and SNU407, MSS colorectal carcinoma cell lines WiDr and non-colorectal cancer cell line, HeLa were maintained according to ATCC guidelines. For real-time PCR experiments, HCT116 and RKO cells were cultured as above.

To measure mutation frequencies in tissues, 21 MSI-H primary colon carcinomas were included in this study. Authorization for use of the tissues for research was obtained from the Institutional Review Board of Yonsei Medical Center. Some of the fresh specimens were obtained from the Liver Cancer Specimen Bank of the National Research Resource Bank Program of the Korea Science and Engineering Foundation of the Ministry of Science and Technology. Tumor specimens were micro-dissected on a cryostat and fractionated to enrich the tumor cell population. Briefly, before cutting sections for DNA isolation, a slide was prepared for hematoxylin-eosin stain to allow the selection of samples with 60% or more tumor cell content; non tumor cells of samples with less than 60% tumor cell content were cut away to enrich the tumor cell population.

2. DNA microarray

The quality of the total RNA samples was checked using an Agilent Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) to determine if it was suitable for DNA arrays. Sentrix® BeadChip Array HumanWG-6-v3

(Illumina, San Diego, CA, USA) was used for human whole gene expression analysis. To select candidate genes containing MNR, a total of 4071 coding sequences containing MNR with more than seven nucleotide repeats previously reported database were used[21]. The Hs numbers (UniGene accession number for human) of genes identified from array analysis were to merge with database. Genes containing cMNR with more than nine nucleotide repeats were selected for further analysis. All of these processes were manually performed in the Microsoft Excel program.

3. PCR based mutation analysis and sequencing

Frameshift mutations in the cMNRs were analysed using a PCR-based assay using the primers described in Table 2. Primers were designed to amplify approximately 100 bp encompassing the cMNR of each gene. PCR reactions were carried out in a 20 μl mixture containing 1.5 mM $MgCl_2$, 20 pmol primer, 0.2 mM of each dATP, dGTP, dTTP, 5 μM dCTP, 1 μCi of [αP^{32}]-dCTP, 50 ng of sample DNA, 1 x PCR buffer and 1.25 U Taq DNA polymerase (Applied Biosystems, Foster City, CA, USA) using the following conditions; an initial denaturation at 95 ℃ for 5 min, followed by DNA amplification for 25-30 cycles of denaturation at 95 ℃ for 30 s, primer annealing at 55-60 ℃ for 30 s, and primer extension at 72 ℃ for 30 s. PCR products were separated in 6 % polyacrylamide gels containing 5.6 M urea, followed by autoradiography.

In order to confirm that a shifted band represented a frameshift mutation, PCR products displaying a mobility shift were cut from the gel for sequencing analysis. DNA eluted from the gel was subjected to PCR without isotope and sequencing (Genotech, Daejeon, Korea).

4. siRNA treatment

siRNA against h*UPF1* was used to block NMD in HCT116, RKO, SW480,

and HeLa. Endogenous *eIF4A3* or *Y14* was down-regulated by siRNAs in HeLa to check the inhibition effect of translation repression. Cells were cultured in 60 mm dishes and were transiently transfected with 100 nM of siRNAs (siGENOME SMARTpool) (Dharmacon, Lafayette, CO, USA) using Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol. For transfection of β-globin construct, cells were transfected with 1 μg of each specific plasmid as indicated in the text, 2 days after transfection with the siRNA. Cells were harvested 24 hours later and were used for protein purification and RNA extraction. The inhibition of UPF1, eIF4A3, and Y14 expressions by siRNA targeting were evaluated by western blotting.

5. Drug treatment

Cells were treated with 100 μg/mL puromycin (Sigma, St. Louis, MO, USA) for 4 hours before harvesting for NMD inhibition. Actinomycin D was added to reduce the stress effects of emetine treatment[16]. Briefly, each cell line was treated with emetine (100 μg/ml) and actinomycin D (4 μg/ml) for 4 hours prior to harvest. For NMD inhibition using another drug, cells were treated with 100 μg/mL cycloheximide (Sigma) for 4 hours before harvesting. To inhibit proteolysis, 25 μM MG132 was used for 2 hours before harvesting. The inhibition of proteasome was evaluated by western blotting by antibody against HIF1-α.

6. Plasmid constructs

To generate the 3 x FLAG GI construct, β-globin gene of pmCMV-GI Norm[22] was digested with Not1 and ligated into 3 x FLAG CMV10 (Invitrogen). Using site directed mutagenesis, PTCs of 3 x FLAG GI P39, P66, P101, P127 were generated respectively. To generate the 3 x FLAG hGI construct, human β-globin was amplified from the 5' - HindIII and 3' – Xba1 conjugated primers. To generate the 3 x FLAG EGFP control plasmid, the

EGFP from rN-EGFP plasmid[23] was digested with HindIII and Not1 and ligated into 3 x FLAG CMV10 (Invitrogen).

7. Northern blot analysis

Total RNA was extracted from HeLa cells with TRIZOL Reagent according to the manufacturer's protocol (Invitrogen). Total cellular RNA (10 μg) was denatured, electrophoretically separated on 6 % Novex TBE-Urea gel (Invitrogen), and blotted to nylon membrane (Amersham, Boston, MA, USA). After UV cross-linking, the membrane was hybridized to P^{32}-labeled specific probe which was amplified by RT-PCR. Hybridized membrane was washed with 2 x SSC (1 x SSC is 0.15 M NaCl plus 0.015 M sodium citrate, pH 7) containing 0.2 % SDS, two or three times, and exposed to X-ray film.

8. RT-PCR and quantitative real-time PCR

Total RNA was isolated from cells using Trizol (Invitrogen). Reverse transcription was performed with 2 μg of total RNA using M-MLV reverse transcriptase, and random hexamer primers (Invitrogen). RT-PCR was performed with the following conditions; an initial denaturation at 95 ℃ for 5 min, followed by DNA amplification for 25-30 cycles of denaturation at 95 ℃ for 30 s, primer annealing at 55-60 ℃ for 30 s, and primer extension at 72 ℃ for 30 s. PCR products were separated in 1 - 1.5 % agarose gels. Real-time PCR was performed using the ABI PRISM 7500 Sequence Detector (Applied Biosystems) and SYBR Premix Ex TaqII (TaKaRa, Shiga, Japan), according to instructions provided by the manufacturers. The relative expression levels of target genes were determined by generating a 5-point serial standard curve. The amount of target mRNA was normalized to GAPDH mRNA or transfected control plasmid mRNA.

9. Western blot analysis

Cell lysates used in the microarray analysis were prepared by using passive lysis buffer (Promega, Madison, WI, USA). Protein samples were separated by SDS-PAGE and transferred to PVDF membranes. Blots were blocked with TBST containing 5 % skim milk, and were incubated for 1 hour at room temperature with primary antibodies against GAPDH (Trevigen, Gaithersburg, MD, USA), CDC7 (Cell Signaling, Danvers, MA, USA), TBC1D23 (ProteinTech, Chicago, IL, USA), UPF1, FLAG (Sigma), GFP (BD Biosciences, San Jose, CA, USA), HIF1-α (BD Biosciences), eIF4A3 (ProteinTech), and Y14 (Bethyl Laboratories, Montgomery, TX, USA). After washing, the membranes were incubated with HRP-conjugated secondary antibody (Santa Cruz, CA, USA) for 1 hour at room temperature, washed, and developed with ECL-Plus (Santa Cruz).

10. Polysome fractionation analysis

HeLa cells were incubated with 100 μg/mL cycloheximide for 5 minutes at room temperature, 24 hours after transfection with plasmids as described in the text, and were washed three times with ice-cold PBS. Cells were collected by scraping into PBS, and were then incubated in lysis buffer [15 mM Tris-Cl, pH 7.4, 3 mM $MgCl_2$, 10 mM NaCl, 0.5 % Triton X-100, 100 μg/mL cycloheximide, 1 mg/mL heparin, and 200 U RNasin (Intron, Gyeonggi, Korea)]. Where indicated, puromycin (100 μg/mL) was added to the cultures 2 hours prior to harvest. Nuclei and debri were removed by centrifugation at 12,000 × *g* for 2 minutes. One milliliter of each sample was layered onto an 11 mL 10 % – 50 % sucrose gradient, and centrifuged for 2 hours at 4^{o}C using an SW40 rotor at 39,000 rpm. Twelve fractions were collected from the top of each gradient with concomitant measurement of absorbance at 254 nm, using a fraction collection system. RNA was extracted from each fraction using TRIZOL and was analyzed by RT-PCR.

III. RESULTS

1. Identification of differentially expressed genes following NMD inhibition

Genes containing coding mononucleotide repeats (cMNR) are frequently mutated in MSI-H cancers. The mRNAs transcribed from cMNR mutated genes invariably harbor PTC and are usually degraded by the NMD system[24]. It was hypothesized that blocking NMD in an MSI-H cell line would cause the PTC-containing mRNAs to accumulate in the cell, thereby enables detection of them. To test this hypothesis, NMD was blocked in two MSI-H cell lines, HCT116 and RKO, using two independent methods: emetine and actinomycin D treatment to chemically inhibit NMD, and siRNA of hUPF1 to inhibit NMD genetically. Emetine inhibits translation by binding to the 40S subunit of the ribosome, and this inhibition has been shown to abrogate degradation of mutant transcript mediated NMD[25]. Actinomycin D, a transcription inhibitor, was added in the chemical block to reduce stress response to emetine[16,26]. In these two different experiments, the efficiency of the NMD block was checked by examining endogenous NMD substrates, ASNA, CARS, and SC35, which were reported as physiologic NMD substrates[27] and found the increased expression of these three genes. In addition, gene expression changes in a microsatellite stable cell line, SW480 were examined to control for non-specific gene expression changes after NMD block (Fig. 2).

Total 12 RNA samples were prepared from four different conditions (mRNAs from 3 cell lines with control siRNA, 3 cell lines with siRNA of h*UPF1*, 3 cell lines treated actinomycin D alone, and 3 cell lines treated with emetine and actinomycin D) for Illumina whole-genome expression bead array. All lysates derived from cells treated with siRNA against h*UPF1* were checked for UPF1 down-regulation before the DNA array analysis (Fig. 1B).

Average-linkage hierarchical clustering analysis was conducted for 12 arrays. Each cell line clustered into 1 of 2 subgroups, chemical block and genetic block, as expected (Fig. 1A). All array data for the samples described in this study can be accessed on our web page (http://www.molpathol.org).

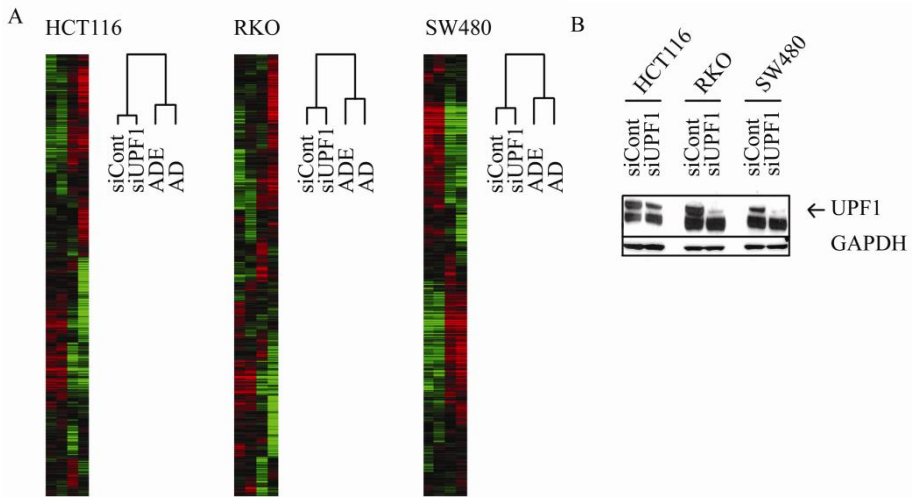


Figure 1. Hierarchical clustering analyses of mRNA expression in HCT116, RKO, and SW480 cell lines. A. Each cell line clustered into two subgroups, cells with genetic NMD block and cells with chemical NMD block. B. Western blotting of human UPF1 for validation of h*UPF1* knockdown in each cell line. siCont denotes negative control of siRNA; siUPF1 denotes siRNA of h*UPF1*; AD denotes actinomycin D; ADE denotes actinomycin D and emetine.

From a total of 48,804 genes, 7,870 were up-regulated more than 1.2 fold in HCT116 or RKO when NMD was blocked. Among these, 3,684 genes that were up-regulated in the SW480 (MSS cell line) by NMD block were excluded to identify specific NMD targets. With the 4,186 genes selected for further analysis, only 2,837 genes were able to merge with previously reported mononuclear repeat database[21]. Using the data search program, 39 genes containing MNRs with more than 9 nucleotide repeats, and 154 genes containing MNR with 8 nucleotide repeats were found. After manual analysis,

28 out of 39 genes were confirmed more than 9 nucleotide repeats in their coding MNR (Table 1). Mutations of these 28 genes were searched because the mutations rated of cMNR are expected to increase with MNR length[28].

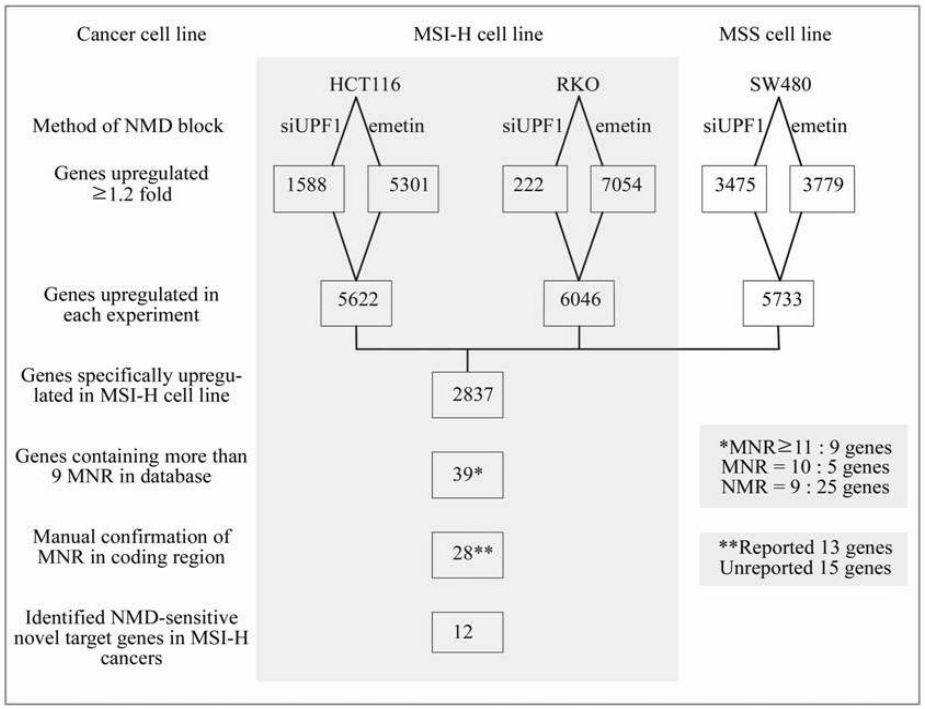


Figure 2. Schematic diagram of flow chart to identify NMD-sensitive target genes in MSI-H cancers. Both h*UPF1* knockdown and drug treatment were used to block NMD. Genes up-regulated in the SW480 cells (MSS cell line) after NMD block, were regarded as non-specifically up-regulated gene and were eliminated from the pool of candidate genes. Finally, 39 genes that showed elevated expression in the MSI-H cell lines after NMD block and contained MNR with more than 9 nucleotide repeats were selected. Among these genes, only 28 genes were confirmed to have MNR in their coding regions. Among the 28 genes, 13 genes had been reported to have mutations in MSI-H colon cancers; therefore, 15 genes were finally selected for mutation analysis in MSI-H tumors.

Table 1. List of 28 MSI specific target genes containing coding more than 9 mononucleotide repeats

Gene Name	Type of Repeat	Function	Reference
CEP164	A, 11	Involved in cell cycle checkpoint, genomic stability	21
MARCKS	A,11	Involved in cell motility, phagocytosis, membrane trafficking	29
ATR	A, 10	Serine/threonine protein kinase, DNA damage sensor	30
MNS1	A, 10	Involved in meiotic division and germ cell differentiation	21
PHACTR4	A, 10	A member of the phosphatase and actin regulator	
TFAM	A, 10	Key activator of mitochondrial transcription	31
CCDC43	A, 9	Unknown	
CDC7	A, 9	Positive regulator of cell proliferation	
CHEK1	A, 9	Negative regulator of cell proliferation	32
DIAPH3	A, 9	Unknown	
ELAVL3	G, 9	Neural specific RNA-binding protein/ involved in neurogenesis	31
ERCC5	A, 9	Involved in excision repair of UV-inducted DNA damage	33
HDAC2	A, 9	component of transcriptional repressor complexes	34
HMMR	A, 9	Involved in cell motility, component of complex with BRCA1 and BRCA2	26
KIAA0391	T, 9	Involved in coronary artery disease	
MLL3	A, 9	Involved in transcriptional co-activation	
OFD1	A, 9	Involved in oral-facial-digital type1 syndrome	35
RAPGEF6	A, 9	GTP-dependent protein binding	
RBBP8	A, 9	Regulator of cell proliferation	36
RUFY2	A, 9	Involved in late-onset Alzheimer disease	
SGOL2	T, 9	Essential for centromere cohesion	
SLC23A2	C, 9	Sodium-dependent vitamin C transporter	37
SMG7	A, 9	Involved in Nonsense Mediated mRNA Decay	
SPAG9	A, 9	Positive regulator of cell migration	
TBC1D23	A, 9	Unknown	
UGDH	T, 9	Involved in biosynthesis of glycosaminoglycans	
VEZT	T, 9	Involved in listeria entry into epithelial cells	
ZMYM4	A, 9	Unknown	

Note; Among 39 genes containing MNRs with more than 9 nucleotide repeats, 28 genes were

confirmed as they have NMR in coding region. Genes represented by bold type are 15 unknown target genes.

2. Identification of 12 novel cMNR-containing genes frequently mutated in MSI-H cancer cell lines

Of the 28 cMNR-containing genes with more than 9 nucleotide repeats, 13 had already been reported as having mutations in cMNR[21,29-37]. Thus the mutation frequency of 15 novel genes and 1 previously reported gene were analyzed in 10 MSI-H cancer cell lines by isotope PCR using αP^{32} dCTP (Fig. 3A). The primers used for mutation analysis are listed in Table 2. Genomic DNA samples were isolated from 10 MSI-H colorectal cancer cell lines, DLD1, HCT116, HCT8, LoVo, LS174T, RKO, SNU C2A, SNU C2B, SNU C4, and SNU407. A MSS colorectal cancer cell line, WiDr, and HeLa cells were used as controls. As expected, MSI-H colon cancer cell lines have frameshift mutations in the cMNR of 12 of these genes; however, 3 of the genes were not mutated in these cell lines. Of the 12 genes that were mutated in MSI-H cancer cell lines, 5 were not mutated in HCT116 and RKO. These findings suggest that 8 out of 15 genes underwent a non-specific up-regulation following a NMD block in these latter two cell lines. In the 12 genes with cMNR mutations, 6 had frequencies of more than 40 %. The mutation frequencies were 70 % for *MLL3* and *PHACTR4*, 50 % for *RUFY2* and *TBC1D23*, 40 % for *CDC7* and *SPAG9*, 20 % for *SMG7* and *ZMYM4*, and 10 % for *KIAA0391*, *RAPGEF6*, *SGOL2*, and *VEZT*. The mutation frequency of the previously reported gene, *HMMR*, was 10 % (Table 3)

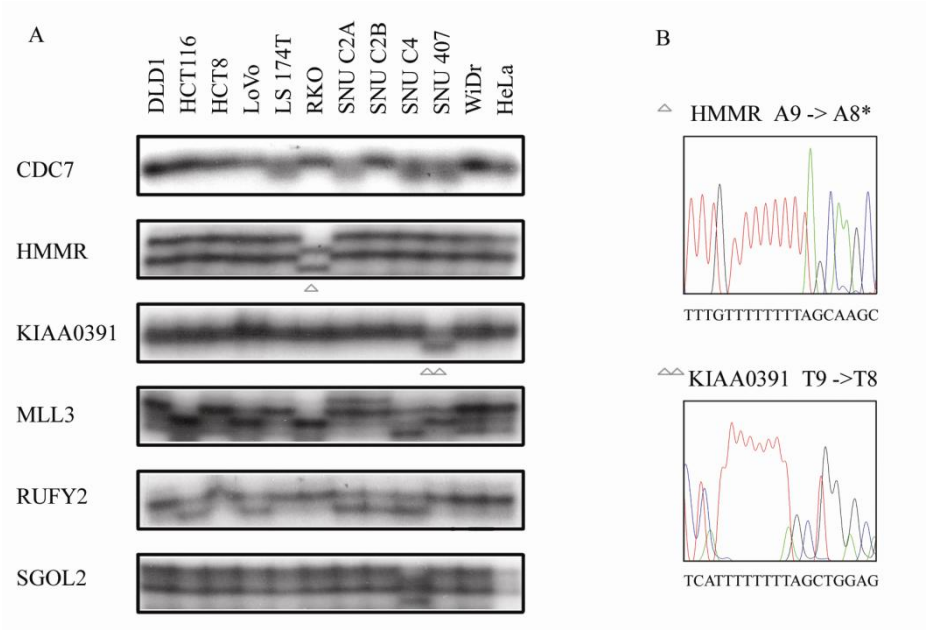


Figure 3. Analysis of frameshift mutations in MSI-H tumors. A. Among the NMD sensitive target genes, 15 genes containing MNR with more than 9 nucleotide repeats were selected and analyzed for mutations in cMNR in MSI-H colorectal cancer cell lines. Frameshift mutations of *CDC7*, *KIAA0391*, *MLL3*, *RUFY2*, *SGOL2*, and previously reported gene, *HMMR* were represented. B. PCR products containing αP^{32} dCTP of *HMMR* and *KIAA0391* were separated on acrylamide gels. PCR products of abnormal sizes were cut from the gel and DNA was eluted and amplified for further sequencing. *Reverse sequence using reverse primer

Table 2. Primers used for RT-PCR analysis

Gene	Sequence of primer
CCDC43	F: 5’-agctgcagagggagagagac-3’ /R: 5’-gtaggccaaggttatcgcttt-3’
CDC7	F: 5’-tctttggggattcagatgga-3’ /R: 5’-tgcaagtttaaaattctgctcgt-3’
DIAPH3	F: 5’-agttttagaagctttaacttcagctggt-3’ /R: 5’-cggaggccttccacaatac-3’
HMMR	F: 5’-caaaactccgctgtcagctt-3’ /R: 5’-tgaaaagcctttgaaggatca-3’
KIAA0391	F: 5’-tcaggaaagatgagggcagt-3’ /R: 5’-ttcagtttgagaattcatctgtga-3’
MLL3	F: 5’-ggaactagaccttccaattgatga-3’ /R: 5’-tgaatgtttatcagagagaaccaga-3’
PHACTR4	F: 5’-cctcctaccaaaaggaagagc-3’ /R: 5’-cttctgaagtctctttaaatttatcactac-3’
RAPGEF6	F: 5’-ggactgaacaagagaaatctgg-3’ /R: 5’-tgttcaatatctcctggcaca-3’

RUFY2	F: 5’-tctgggtcgagcaagagc-3’ /R: 5’-gagatccctctgaataattaagcaac-3’
SGOL2	F: 5’-ggataatgataaagatgtggtgcat-3’ /R: 5’-ttggaaacttctatgttttcagagat-3’
SMG7	F: 5’-cgtcattgcaacctcctgta-3’ /R: 5’-tgggaccttgacttctgagg-3’
SPAG9	F: 5’-ggcatcacgagaaaatccag-3’ /R: 5’-ttttcttgtctcatgtgctgtct-3’
TBC1D23	F: 5’-acagtctcgacaagcgctga-3’ /R: 5’-tcctgaagcactgctatttcc-3’
UGDH	F: 5’-ccaggactaaaagaagtggtagaa-3’ /R: 5’-caagatcagcttctttgatgg-3’
VEZT	F: 5’-aggtatcctgttaaaagtggctga-3’ /R: 5’-ttgtgaagtaagtcatctttcttattg-3’
ZMYM4	F: 5’-aaagatttttgcagtcagtcatgt-3’ /R: 5’-acatgctgcatttggttgaa-3’

Table 3. Mutation profile of cMNR containing more than 9 nucleotides in MSI-H colorectal cancer cell lines

Cell line	PHACTR4	CDC7	HMMR	KIAA0391	MLL3	RAPGEF6	RUFY2	SGOL2	SMG7	SPAG9	TBC1D23	VEZT	ZMYM4	CCDC43	DIAPH3	UGDH
DLD1	.	.	.	.	.	.	.	.	.	.	-1/w	.	.	.	.	.
HCT 116	-1/w	.	.	.	-1/-1	.	-1/w	.	.	-1/w	.	-1/w	.	.	.	.
HCT 8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
LoVo	-1/w	.	.	.	-1/w	.	-1/w	.	.	.	.	.	.	.	.	.
LS174T	-1/w	-1/w	.	.	.	.	.	.	-1/w	.	-1/w	.	.	.	.	.
RKO	.	.	-1/-1	.	-1/-1	+1/w	.	.	.	-1/w	-1/w	.	.	.	.	.
SNUC2A	-1/w	-1/w	.	.	+1/w	.	-1/w	.	.	-1/w	-1/w	.	-1/w	.	.	.
SNU C2B	-1/w	.	.	.	+1/w	.	-1/w	.	.	-1/w	-1/w	.	-1/w	.	.	.
SNU C4	-1/w	-1/w	.	.	-2/w	.	-1/w	-1/-1	.	.	.	.	.	.	.	.
SNU 407	-1/w	-1/w	.	-1/w	-1/w	.	.	.	-1/w	.	.	.	.	.	.	.
Incidence	**70%**	**40%**	**10%**	**10%**	**70%**	**10%**	**50%**	**10%**	**20%**	**40%**	**50%**	**10%**	**20%**	**0%**	**0%**	**0%**

Note; Among 28 genes with cMNRs containing more than 9 nucleotide repeats, unreported 15 genes were analyzed their frameshift mutations. *HMMR* is used as a mutated control gene.

3. The identified novel genes are frequently mutated in MSI-H colorectal carcinoma tissues

In order to check the mutations in cancer tissues, 12 genes of interest were analyzed in 21 MSI-H cancer tissues. The cases used were consecutively identified from the Gastrointestinal Tumor Working Group Tissue Bank at Yonsei University Medical Center (Seoul, Korea) between 2003 and 2005. The MSI status of each of the 21 MSI-H colorectal carcinomas were evaluated using five MSI markers, *BAT25*, *BAT26*, *D17S250*, *D2S123*, and *D5S346*. The percentage of tumor cells, determined on histological slides after cryostat fractionation to enrich the tumor cell population, amounted to more than 60%. All but one of the 12 novel genes had mutations in MSI-H cancer tissues. The mutation frequency of each gene in MSI-H cancer cell lines and tissues is depicted in Figure 4. The status of mutation profile for the 21 MSI-H colorectal cancer tissues is presented in Table 4.

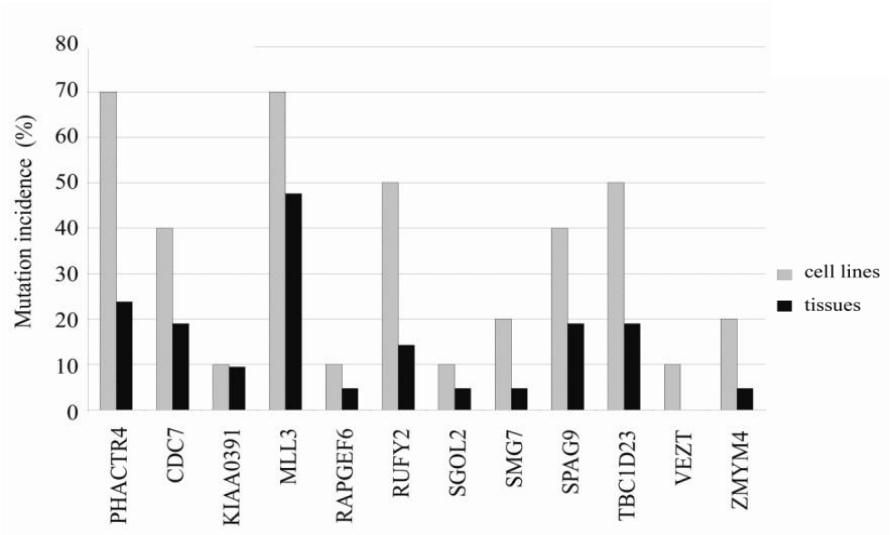


Figure 4. Frequency of frameshift mutations of 12 genes in 10 MSI-H cell lines (dashed bars) and 21 tissues (dark bars).

Table 4. Mutation profile of cMNR containing more than 9 nucleotides in MSI-H colorectal cancer tissues.

Tissue number	PHACTR4	CDC7	KIAA0391	MLL3	RAPGEF6	RUFY2	SGOL2	SMG7	SPAG9	TBC1D23	VEZT	ZMYM4
3345	-1/w	.	-1/w	-1/w	.	.	.	.	.	.	.	.
3368	-1/w	.	.	.	.	.	.	.	.	-1/w	.	.
4010	-1/w	.	.	-1/w	.	-1/w	.	.	-1/w	.	.	.
4118	.	.	.	-1/w	.	.	.	.	.	.	.	.
4124	.	.	.	.	.	.	.	.	.	.	.	.
4130	.	+1/w	.	-1/-1	.	.	.	.	-1/w	+1/w	.	.
4191	.	.	.	.	.	.	.	.	.	-1/w	.	.
4289	.	.	.	.	.	.	.	.	.	.	.	.
5205	.	.	.	-1/w	.	.	.	.	.	-1/w	.	-1/w
3597	.	.	.	.	.	.	.	.	.	.	.	.
3622	.	.	.	.	.	.	.	.	.	.	.	.
3806	.	.	.	.	.	.	.	.	.	.	.	.
4131	-1/w	.	.	-1/w	.	-1/w	.	.	.	.	.	.
4259	-1/w	.	.	-1/w	.	.	.	.	.	.	.	.
4296	.	.	.	-1/w	.	.	.	.	-1/w	.	.	.
4367	.	-1/w	-1/w	-1/w	-1/w	.	.	-1/w	.	.	.	.
4382	.	.	.	.	.	.	.	.	.	.	.	.
4565	.	.	.	.	.	.	.	.	.	.	.	.
4572	.	-1/w	.	-1/w	.	.	.	.	.	.	.	.
4791	.	-1/w	.	.	.	-1/w	-1/w	.	.	.	.	.
4797	.	.	.	.	.	.	.	.	+1/w	.	.	.
Incidence	**24%**	**19%**	**10%**	**48%**	**5%**	**14%**	**5%**	**5%**	**19%**	**19%**	**0%**	**5%**

Note; Twelve genes were analyzed in 21 MSI-H cancer tissues. (.) denotes that both alleles show wild type; (w) denotes no mutation in the cMNR; (-1) denotes a 1-bp deletion in the cMNR; (-2) denotes a 2-bp deletion in the cMNR; (+1) denotes a 1-bp insertion.

4. mRNA expression of the mutated genes is restored after emetine and puromycin treatment

In order to confirm the decreased mRNA expression level of the mutated genes was due to NMD activity, NMD block experiments were conducted by using drugs. NMD is blocked when translation is inhibited, and resulted in the accumulation of PTC-containing mRNA transcripts[25]. For the NMD blocking, previously used emetine and actinomycin D for the screen of genes in the microarray experiment were used. In order to assure the experimental validity another chemical drug, puromycin was used which is an amino-acyl tRNA analog that prematurely terminates peptide bond elongation on the ribosome, and generally used for translation block[38]. HCT116 and RKO cells were treated with puromycin, and the quantity of transcript was measured by real-time PCR. *HMMR*, *MLL3*, and *RUFY2* were selected for this study and *GAPDH* was used as a control. RKO cells had homozygous mutations in *HMMR* and *MLL3* and HCT116 cells carried homozygous mutations in *MLL3* (Table 3). By real-time PCR analysis, increased mRNA levels for *HMMR* and *MLL3* in RKO cells and *MLL3* and *RUFY2* in HCT116 cells were found (Fig. 5). Wild type *HMMR* mRNA expressions in emetine treated HCT116 cells were increased, however this up-regulated expressions might be due to nonspecific effect of emetine, considering that mRNA expressions of wild type *HMMR* in puromycin treated HCT116 cells were maintained (Fig. 5). All of these findings indicate that transcripts with cMNR mutations in *HMMR*, *MLL3*, and *RUFY2* are normally degraded by NMD in these two cell lines.

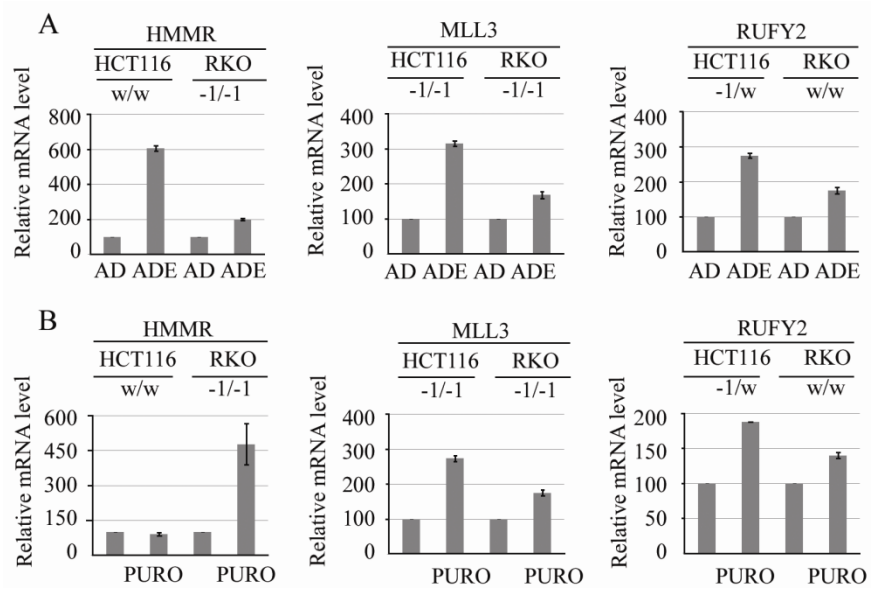
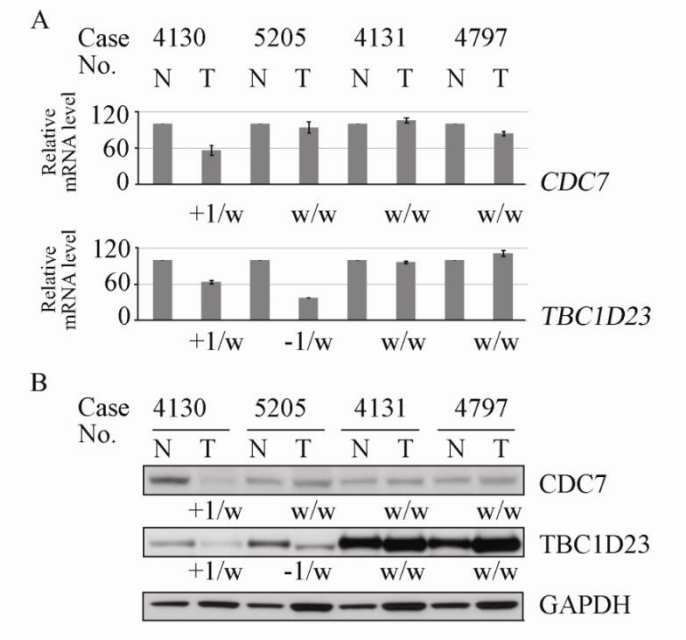


Figure 5. mRNA expressions of the mutated genes of cMNR are restored by NMD blocking. A. Real-time PCR analysis of *HMMR*, *MLL3*, and *RUFY2* genes in HCT116 and RKO cell lines treated with 100 μg/mL emetine and 4 μg/mL actinomycin D. Treatment of actinomycin D alone was used as a control. Each target mRNA was normalized to the amount of *GAPDH* mRNA. The normalized level of mRNA from cells treated actinomycin D was defined as 100 %. AD denotes actinomycin D; ADE denotes actinomycin D and emetine; (w) denotes no mutation in the cMNR; (-1) denotes a 1-bp deletion in the cMNR. B. Real-time PCR analysis of *HMMR*, *MLL3*, and *RUFY2* genes in HCT116 and RKO cell lines treated with 100 μg/mL puromycin in order to block NMD. Each target mRNA was normalized to the amount of *GAPDH* mRNA. The normalized level of mRNA from non treated cells was defined as 100 %. PURO denotes puromycin; (w) denotes no mutation in the cMNR; (-1) denotes a 1-bp deletion in the cMNR.

5. Protein expressions of colon cancer cells with mutant *CDC7* and *TBC1D23* genes are down-regulated

Among the 21 colorectal cancer tissues, 16 colorectal cancer tissues were evaluated for protein expressions of the mutated genes according to the availability for western blot analysis. *CDC7* and *TBC1D2*3 were selected to this study according to the availability of commercial antibodies. Three out of

16 tumors showed one base pair insertion or deletion in cMNR of *CDC7* gene. *CDC7* mRNA expressions in tumors with cMNR mutation were lower than the matched normal mucosa, and its protein expressions were also lower (Fig. 6A and B). Two out of 16 tumors had mutations in cMNR of *TBC1D23*, and mRNA and protein expressions of *TBC1D23* in these two tumors were markedly low (Fig. 6A and B). Protein expressions of *CDC7* and *TBC1D23* in case 4130, 5205, 4131, and 4797 were depicted in Figure 6. The down-regulation of mutated *TBC1D23* was also confirmed by immunohistochemical analysis. Tumors with wild type *TBC1D23* showed strong expression of TBC1D23. In contrast, two colorectal tumors with frameshift mutation in cMNR of *TBC1D2* showed low TBC1D23 expression (Fig. 6C).



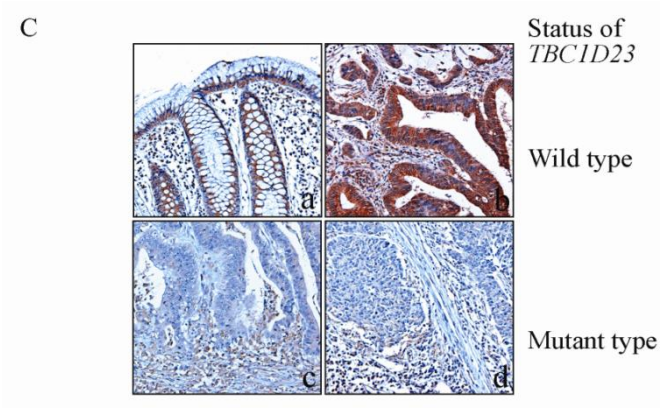


Figure 6. mRNA and protein expressions of colon cancer cells with mutant *CDC7* and *TBC1D23* genes are down-regulated. A. Real-time PCR analyses of *CDC7* and *TBC1D23* in tumor tissues. Each target mRNA was normalized to the amount of *GAPDH* mRNA. The normalized level of mRNA from matched normal mucosa was defined as 100 %. mRNA expression of *CDC7* was decreased in the tumor of case 4130, which containing heterozygous 1-bp insertion in cMNR. The expressions of TBC1D23 were decreased in the tumors containing 1-bp insertion (case 4130) and 1-bp deletion (5205). N denotes normal mucosa; T denotes colonic tumor; (w) denotes no mutation in the cMNR; (+1) denotes a 1-bp insertion in the cMNR; (-1) denotes a 1-bp deletion in the cMNR. B. Western blotting analysis of CDC7 and TBC1D23 in tumor tissues. GAPDH used as a control. CDC7 expression was decreased in tumor of 4130 (case with 1-bp insertion). The expressions of TBC1D23 were decreased in tumors of 4130 (case with 1-bp insertion) and 5205 (1-bp deletion). N denotes normal mucosa; T denotes colonic tumor; (w) denotes no mutation in the cMNR; (+1) denotes a 1-bp insertion in the cMNR; (-1) denotes a 1-bp deletion in the cMNR. C. Immunohistochemical analysis of TBC1D23 in colon cancer tissues. a. Non tumorous normal mucosa shows strong expression of TBC1D23 in colonic epithelial cells. b. The expression of TBC1D23 in tumor cells of case 4791 with wild type TBC1D23 shows strong expression of TBC1D23. c. The expression of TBC1D23 in tumor cells of case 4130 with *TBC1D23* frameshift mutation was markedly decreased. d. Case 4191 with *TBC1D23* frameshift mutation shows decreased expression of TBC1D23 in tumor cells.

6. Endogenous expression of NMD sensitive PTC-containing mRNAs in hypoxic condition

It was recently reported that there are various physiological and

environmental stressors to global down-regulation of pioneer round translation, for example, viral infection, amino acid starvation, hypoxia, or heat shock[19]. These environmental changes are able to inhibit NMD, because NMD is a translation dependent mRNA surveillance pathway. Thus it could be hypothesized that if the down-regulated transcripts by NMD are able to be recovered, these recovered PTC-containing transcripts are able to produce truncated proteins which are potentially deleterious. To investigate whether endogenous NMD targets are regulated in cellular condition practically, cells were cultivated under hypoxic condition, one of regulatory environments to NMD, to several novel targets. *RAPGEF6*, *TBC1D23*, and *SPAG9* have heterogenous mutation in RKO cell line, and these genes are degraded by NMD[39]. RKO cells were incubated at 1 % O_2 prior to harvest, and total RNAs were isolated for quantitation of target mRNAs. The mRNA expressions of *RAPGEF6*, *TBC1D23*, and *SPAG9* were elevated in hypoxia 1.3 fold, 3.2 fold, and 1.5 fold, respectively. Through this experiment, it might be speculated that endogenous PTC-containing mRNAs degraded by NMD have chance to increase their stability in certain cellular condition and also these genes have chance to generate truncated proteins.

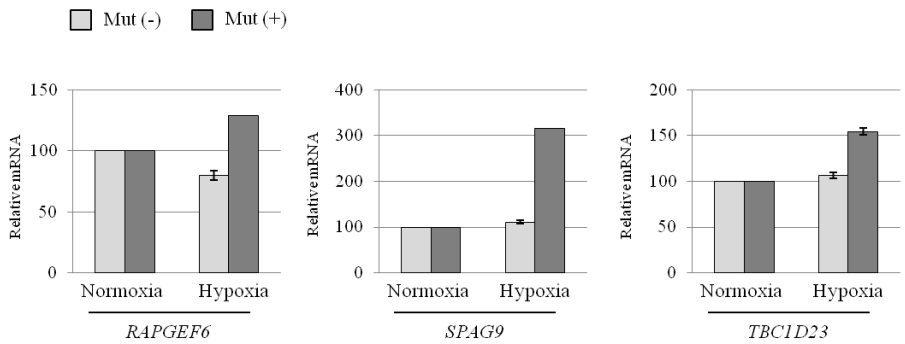
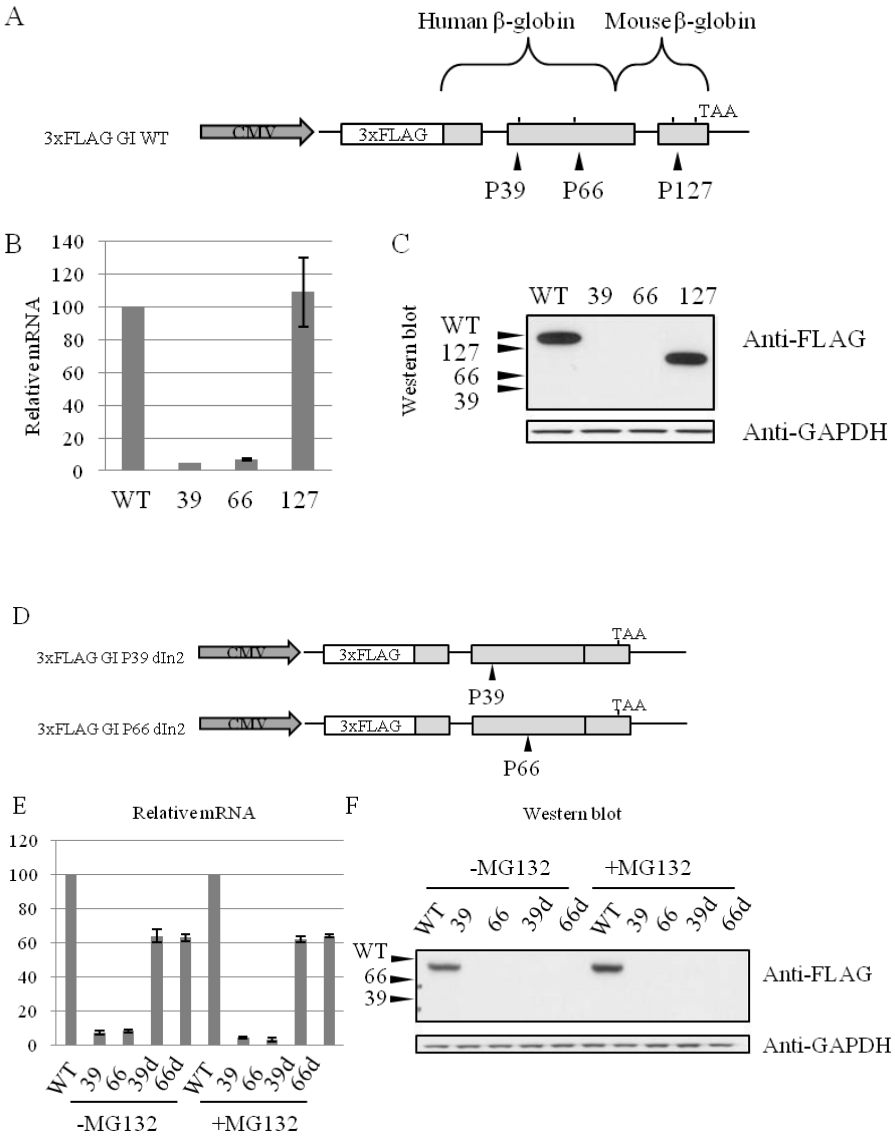


Figure 7. mRNA expressions of endogenous NMD targets are increased by hypoxia. A. Real-time PCR analysis of RKO cells in hypoxic condition. Cells were exposed in 1 % oxygen for 3 hours prior to harvest. Each target mRNA was normalized to the amount of 18S rRNA.

The normalized level of mRNA from cells in normoxia was defined as 100 %.

7. β-globin constructs containing 39th, 66th PTC either could not be proper vector model to examine translation level

In order to identify whether the PTC-containing transcripts produce truncated proteins in NMD inhibited condition, protein production by using a model construct, β-globin (GI) was checked, which has been used as a conventional model construct to study NMD mechanism[3,40]. 3 x FLAG GI plasmid is based on a human mouse hybrid β-globin mRNA[3]. These vectors contain either normal stop or a PTC at each amino acid position of the globin gene at 39th, 66th, and 127th, respectively (Fig. 8A). One day after transfection, total RNAs and proteins were extracted, and quantitative real-time PCR, and western blot were conducted. As expected, the levels of transcripts with PTC located in NMD sensitive position at 39th or 66th were down-regulated by NMD (Fig. 8B). Protein expressions of these two plasmids were not detected at expected size. 3 x FLAG GI with 127th PTC, which represents NMD irrelevant position because the PTC is located in the last exon, showed comparable mRNA expression to the 3 x FLAG GI WT and protein expression at expected size (Fig. 8C). To evaluate truncated protein production by using these conventional NMD sensitive constructs, 3 x FLAG GI P39, P66, proper control plasmids should be required which generate the same size of protein. To produce the same size of β-globin with 39th or 66th termination either, two different ways of control plasmids were designed (Fig. 8D), by second intron deletion, 39th PTC and 66th PTC were located in the last exon, and cDNA construct (Fig. 8G).


A
Human β-globin
Mouse β-globin
3xFLAG GI WT
CMV
3xFLAG
TAA
P39
P66
P127
B
Relative mRNA
140
120
100
80
60
40
20
0
WT
39
66
127
C
WT 39 66 127
Western blot
WT
127
66
39
Anti-FLAG
Anti-GAPDH
D
3xFLAG GI P39 dIn2
3xFLAG GI P66 dIn2
CMV
3xFLAG
TAA
P39
P66
E
Relative mRNA
120
100
80
60
40
20
0
WT
39
66
39d
66d
-MG132
+MG132
F
Western blot
-MG132
+MG132
WT
39
66
39d
66d
Anti-FLAG
Anti-GAPDH

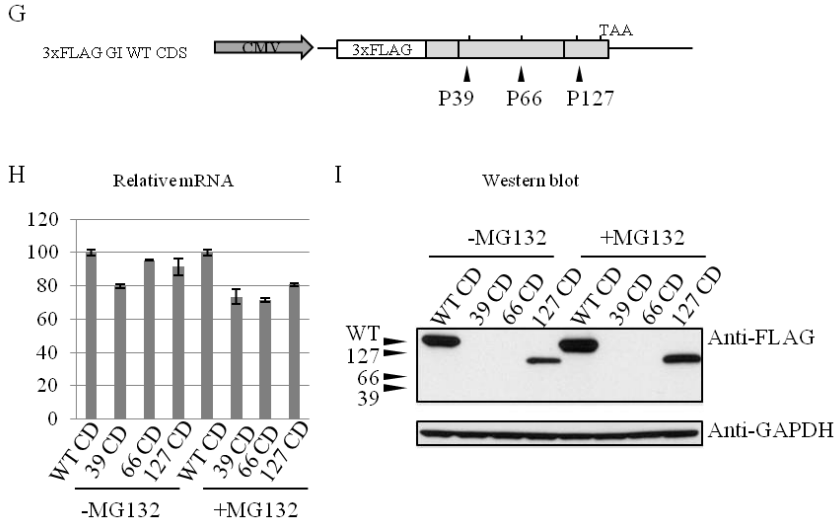


Figure 8. Truncated proteins from 39th or 66th PTC-containing transcripts in the last exon were not detected. A. Construct of conventional β-globin plasmid with several PTC positions. 3 x FLAG GI WT is consisted with human and mouse hybrid β-globin. The position of the PTC is shown below the diagram at 39th, 66th, and 127th amino acid position. B. Relative mRNA levels of each plasmid through real-time PCR. Each target mRNA was normalized to the amount of *GAPDH* mRNA. The normalized level of mRNA from cells with 3 x FLAG GI WT plasmid was defined as 100 %. C. Western blot analyses were carried out with the identical cellular extract of B. D. Schematic representation of the second intron deletion of 3 x FLAG GI P39, and 3 x FLAG GI P66, respectively. E. Real-time PCR analyses of HeLa cells with second intron deletion control plasmids. Each target mRNA was normalized to the amount of *GAPDH* mRNA. The normalized level of mRNA from cells with 3 x FLAG GI WT plasmid was defined as 100 %. F. Western blot analyses were carried out with the identical cellular extract of E. For proteolysis block, 25 μM MG132 was treated to transfected cells for 2 hours prior to harvest. G. Schematic representation of the cDNA construct of β-globin with several PTC positions. The position of the PTC is shown below the diagram at 39th, 66th, and 127th amino acid position. H. Relative mRNA levels of each plasmid through real-time PCR. Each target mRNA was normalized to the amount of *GAPDH* mRNA. The normalized level of mRNA from cells with 3 x FLAG GI WT CDS plasmid was defined as 100 %. I. Western blot analyses were carried out with the identical cellular extract of H. For proteolysis block, 25 μM MG132 was treated to transfected cells for 2 hours prior to harvest.

(WT) denotes 3 x FLAG GI WT, (39) denotes 3 x FLAG GI P39, (66) denotes 3 x FLAG GI

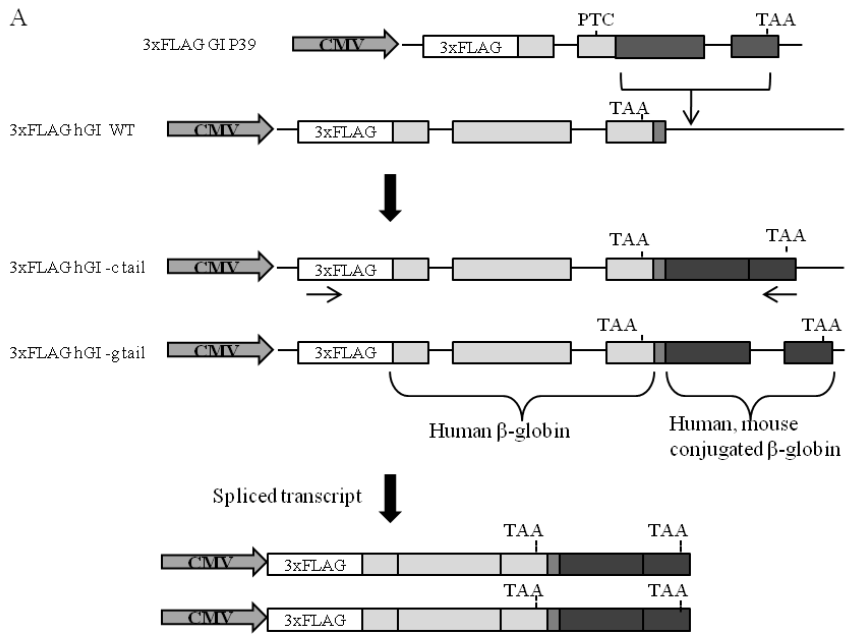
P66, and (127) denotes 3 x FLAG GI P127. Arrow heads represent expected size of each protein.

Modified plasmid with deletion of second intron (3 x FLAG GI P39 dIn2 and 3 x FLAG GI P66 dIn2) showed elevated mRNA expression contrary to plasmid with PTC at NMD sensitive position (3 x FLAG GI P39 and 3 x FLAG GI P66) (Fig. 8E). However, protein expression of these two control plasmids were not detected in common experimental condition, even proteasome inhibition by MG132 treatment (Fig. 8F). In case of non-splicing control plasmids (3 x FLAG GI P39 CDS and 3 x FLAG GI P66 CDS) showed comparable mRNA expressions to 3 x FLAG GI WT CDS (Fig. 8H). However, protein expressions of these two control plasmids were not detected, although two types of β-globin protein were detected well which have normal stop or PTC at irrelevant position, respectively (Fig. 8I). In conclusion, conventional NMD sensitive models, β-globin with 39^{th} PTC or 66^{th} PTC cannot be applicable to this study, because their control proteins are not detected.

8. Generation of new β-globin construct producing stable proteins

To make proper control plasmid which can generate identical stable protein after recovery by NMD inhibition, and more stable protein production, human complete β-globin, not hybrid β-globin was selected. By using 3 x FLAG hGI WT as a backbone, two vectors were constructed following with extra sequences in downstream of normal termination. The extra sequences were obtained from the segment which is 104^{th} amino acid to stop codon of 3 x FLAG GI P39 that is well defined and expected to mRNA decay by NMD[3,41] (Fig. 8B). Through this process, control plasmid, 3 x FLAG hGI-c tail has two termination codons, and the first one is considered as a PTC. There is no intron downstream of the 1^{st} termination codon, therefore the produced PTC-containing mRNA is irrelevant of NMD, and expected to

produce wild type β-globin of same size (Fig. 9D), 3 x FLAG hGI-g tail is a NMD sensitive vector, because an intron is present after first termination codon (Fig. 9C). Transcripts from these two plasmids are identical after splicing. To confirm this, the accurate size of both transcripts was verified by northern blot analysis through total cell RNAs. The different quantity of transcripts is due to NMD. In alternative confirmation, mRNAs of two constructs were determined by RT-PCR using a set of primers that can amplify the full-length cDNA covering the whole ORF in both transcripts (Fig. 9B). Through these experiments, it is expected that same proteins can be generated from both constructs.



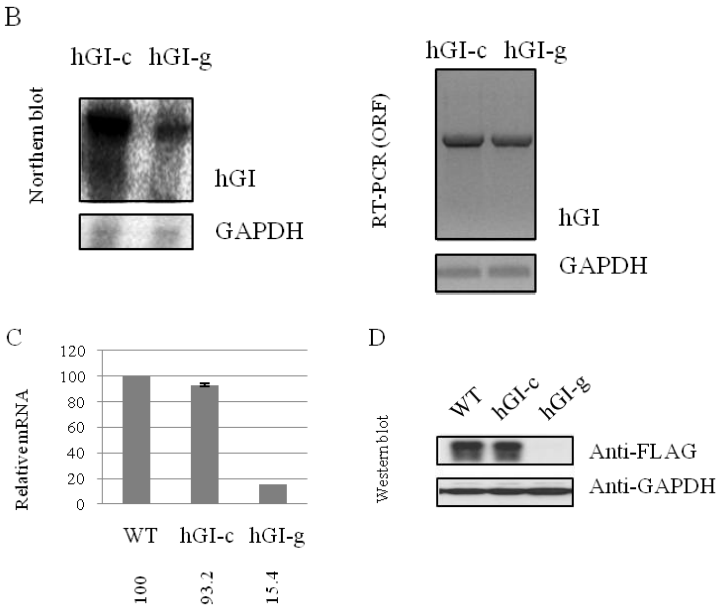


Figure 9. Schematic representations and expressions of 3 x FLAG hGI-c tail and 3 x FLAG hGI-g tail constructs. A. Sequences from downstream of PTC in hybrid GI construct were added after 3 x FLAG hGI WT stop codon. Control construct, 3 x FLAG hGI-c tail contained PTC in irrelevant position, and 3 x FLAG hGI-g tail contained PTC in sensitive position. B. Northern blot analysis and RT-PCR for full-length transcript. HeLa cells were transfected with 3 x FLAG hGI-c tail or 3 x FLAG hGI-g tail, and total RNAs were used for northern blot analysis. *GAPDH* was used as a loading control. RT-PCR experiments using cDNA of these extracts were conducted for validation of proper splicing in each transcript. Two experiments showed that transcripts of 3 x FLAG hGI-c tail and 3 x FLAG hGI-g tail are identical. C. Real-time PCR analyses of the HeLa cells with 3 x FLAG hGI WT, 3 x FLAG hGI-c tail, and 3 x FLAG hGI-g tail, respectively. Each target mRNA was normalized to the amount of *GAPDH* mRNA. The normalized level of mRNA from cells with 3 x FLAG hGI WT plasmid was defined as 100 %. D. Detectable protein producing new construct expressed identical protein of human β-globin wild type. The protein expression of hGI-g tail was hardly detected. GAPDH was used as a loading control.

For validation of their mRNA and protein expressions, quantitative real-time PCR and western blotting were performed. 3 x FLAG hGI-c tail plasmid showed comparable mRNA expressions to 3 x FLAG hGI WT plasmid

(93.2 %), while 3 x FLAG hGI-g tail plasmid showed down-regulated to 15.4 % of the wild type mRNA expressions (Fig. 9C). Properly expressed proteins of hGI-c tail were detected at the same size and comparable quantity of wild type β-globin proteins. However hGI-g tail proteins were hardly detected (Fig. 9D). These data indicate that mRNAs generated from 3 x FLAG hGI-g tail showed NMD sensitive tendency and these transcripts are identical with the control construct, 3 x FLAG hGI-c tail. Table 5 showed simplified expressions of mRNA and protein from transfected cells with each construct.

Table 5. Total mRNA and protein expressions of β-globin constructs

Vector construct	mRNA -qRT	Protein	Protein -MG132
A. 3xFLAG GI WT	100	100	100
B. 3xFLAG GI P39	4.8	0.22	2.83
C. 3xFLAG GI P66	6.68	0.00	1.58
D. 3xFLAG GI P39 dIn2	65.62	0.01	1.29
E. 3xFLAG GI P66 dIn2	64.07	0.00	0.51
F. 3xFLAG GI P127	110.2	87.45	92.91
G. 3xFLAG hGI-c tail	93.2	98.94	-
H. 3xFLAG hGI-g tail	15.4	0.32	-

Note; List of mRNA and protein expressions of 8 conststructs. The normalized level of expression from cells with 3 x FLAG GI WT plasmid was defined as 100 %.

9. Proteins are rarely detected from the PTC-containing β-globin mRNAs recovered by NMD inhibition

Whether modified NMD sensitive construct shows authentic NMD substrate and recovered transcripts can generate proteins in NMD inhibited condition, several ways of NMD inhibition were conducted to the transfected cells with each construct, respectively. 3 x FLAG EGFP was used as a control for the transfection efficiency. Two days treatments of si-h*UPF1*, which is one of NMD key molecule, second transfection was conducted with 3 x FLAG EGFP and 3 x FLAG hGI-c tail or 3 x FLAG hGI-g tail either. The inhibition of UPF1 was evaluated by western blotting. The transcripts of hGI-g tail were recovered to almost three fold (66.9 % expression of control) by si-h*UPF1*. This result showed that NMD worked to hGI-g tail transcript efficiently. However protein expression from this transcript was not increased in comparison to control siRNA treated cells with hGI-g tail, despite substantial increase in mRNA level (Fig. 10A). For the second way of NMD blocking, three chemicals, puromycin, emetine, and cycloheximide were used which are general translation inhibitors. Two days after transfection, cells were treated with 100 μg/mL of puromycin, emetine, or cycloheximide for 4 hours prior to harvest, respectively. hGI-g tail transcripts of each chemical treated cells were restored at 64.8 %, 62.5 %, and 71.2 % of control transcript. However, proteins from these increased transcripts were not detected. For the third way of NMD blocking, hypoxic condition was applied to the transfected cells. According to a previous report, it has been demonstrated that hypoxia inhibited NMD activity[42], and hypoxic condition are known to be frequently produced in cancer environment. In order to mimic this cellular condition, transfected cells were cultivated at 1 % of oxygen for 3 hours prior to harvest. As similar recovery efficiency to previous results (Fig. 10A), NMD affected transcripts were increased about three fold (Fig. 9C). In this case, protein expressions were not also increased, although control proteins were expressed constantly. Table 6 showed simplified mRNA and protein expressions from the transfected cells with each construct. In addition, the translation efficiency

(normalized protein expressions per normalized relative mRNA expressions) was calculated to assess each protein proportion from relative mRNA in NMD inhibited conditions (Table 6). The translational efficiencies of 3 x FLAG hGI-g tail mRNA are under 9 % in several NMD inhibited conditions due to rarely detected proteins from recovered mRNAs. Based on all of the results, it can be concluded that proteins are hardly detected from the PTC-containing transcripts recovered by NMD inhibition.

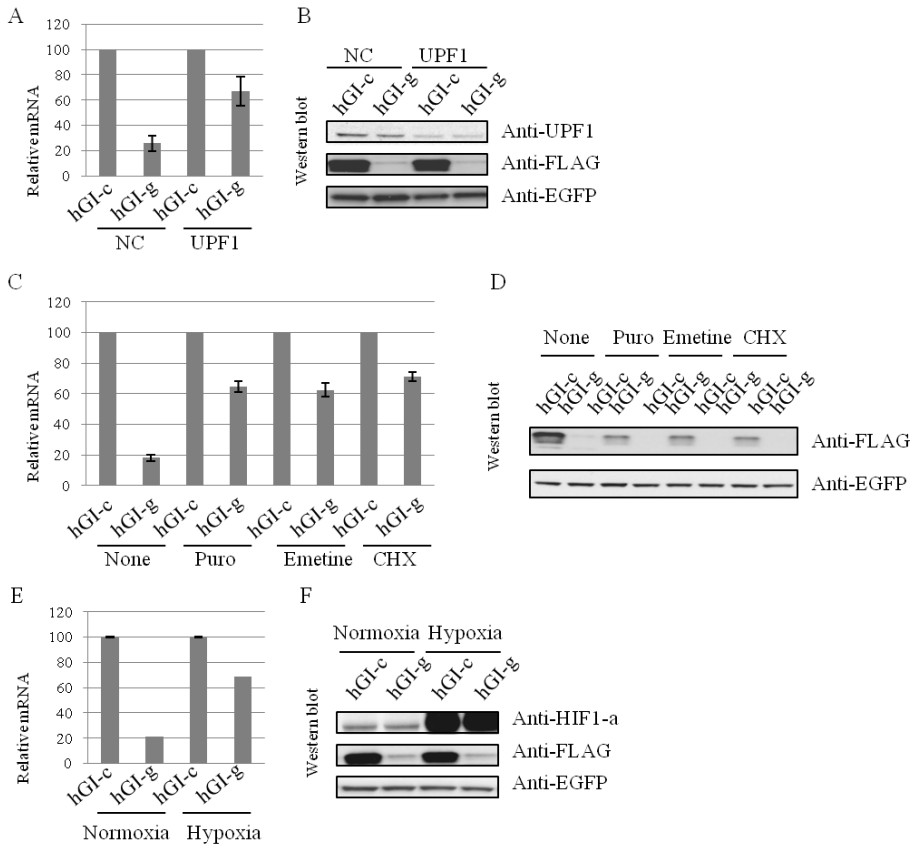


Figure 10. Proteins from recovered 3 x FLAG hGI-g tail transcripts by NMD reverse treatment are not increased. A. Real-time PCR analysis of HeLa cells transfected with 3 x FLAG EGFP and either 3 x FLAG hGI-c tail or 3 x FLAG hGI-g tail, two days after siRNA transfection. Transfected β-globin mRNA was normalized to the amount of co-transfected

EGFP mRNA. The normalized level of mRNA from cells with 3 x FLAG hGI-c tail was defined as 100 %. B. Western blotting analysis of the same cellular extracts of A. Anti-EGFP was used as a control. C. Real-time PCR analysis of NMD inhibited HeLa cells transfected with 3 x FLAG EGFP and either 3 x FLAG hGI-c tail or 3 x FLAG hGI-g tail. Cells were treated each drug at 100 μg/mL for 4 hours prior to harvest. Transfected β-globin mRNA was normalized to the amount of co-transfected EGFP mRNA. The normalized level of mRNA from cells with 3 x FLAG hGI-c tail was defined as 100 %. D. Western blotting analysis of the same cellular extracts of C. Anti-EGFP was used as a control. E. Real-time PCR analysis of HeLa cells transfected with 3 x FLAG EGFP and either 3 x FLAG hGI-c tail or 3 x FLAG hGI-g tail. To inhibit NMD, cells were in 1 % oxygen condition for 3 hours prior to harvest. Transfected β-globin mRNA was normalized to the amount of co-transfected EGFP mRNA. The normalized level of mRNA from cells with 3 x FLAG hGI-c tail was defined as 100 %. F. Western blotting analysis was used to quantitate the level of HIF1-α, transfected β-globin, and EGFP. Anti-HIF1-α was used for check of hypoxic effect. EGFP was used as a control.

Table 6. mRNA and protein expressions of 3 x FLAG hGI-g tail by NMD inhibition and translation efficiency

Vector	NMD inhibition method	mRNA	Protein	Translation efficiency
3xFLAG hGI-c tail	-	100	100	100
3xFLAG hGI-g tail	-	16.52	2.61	15.8
	si-h*UPF1*	66.9	3.5	4.7
	Cycloheximide	71.2	6.34	8.9
	Emetine	62.5	4.62	7.4
	Puromycin	64.8	5.4	8.3
	Hypoxia	69	2.67	3.9

Note; Quantitation of mRNA, protein expressions and translation efficiency of each NMD inhibition methods was represented. The normalized level of expression from cells with 3 x FLAG hGI-c tail plasmid was defined as 100 %. Translation efficiency was calculated by normalized protein expressions per normalized mRNA expressions. Translation efficiency of 3 x FLAG hGI-c tail mRNA was defined as 100 %.

10. Proteasome inhibition has no effect on the detection of proteins from PTC-containing β-globin mRNAs by inhibition of NMD

For the constantly rare level of protein expressions from the restored transcripts, there are two possibilities, rapid degradation and translation regulation. Firstly using proteolysis inhibitor, MG132, protein degradation issues was verified. For the induction of mutant transcript recovery, cells were transfected with si-h*UPF1*. One day after re-transfection with 3 x FLAG EGFP and 3 x FLAG hGI-c tail, or 3 x FLAG hGI-g tail, respectively, cells were treated 25 nM MG132 for 2 hours prior to harvest. Similar to the previous results (Fig. 10A), mRNAs of hGI-g tail were elevated three fold by h*UPF1* knockdown (Fig. 11A). Although the transcripts are increased by NMD block, proteins are hardly detected from the cells with 3 x FLAG hGI-g tail transfection before and after proteasome inhibition by MG132 (Fig. 11B). These results indicate that differences between protein expressions of hGI-c tail and hGI-g tail were not resulted from proteolysis.

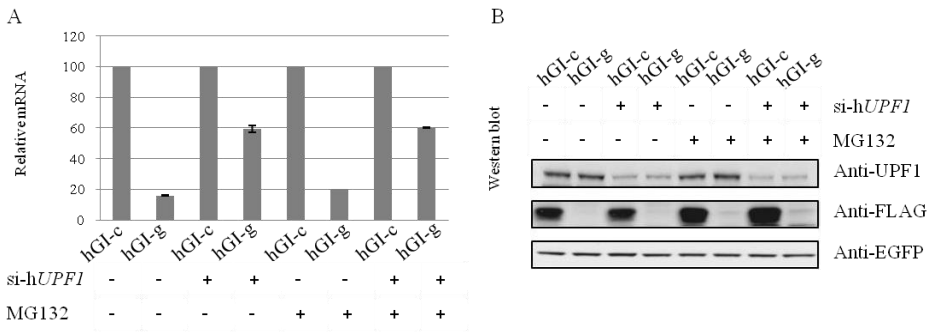


Figure 11. Low level of protein expression of hGI-g tail was not resulted from proteasomal degradation. A. Real-time PCR analysis of hGI in HeLa cells transfected with control siRNA or si-h*UPF1* was performed, two days after co-transfection with 3 x FLAG EGFP, either 3 x FLAG hGI-c tail or 3 x FLAG hGI-g tail. Cells were treated with 25 μM MG132 for 2 hours prior to harvest. Transfected β-globin mRNA was normalized to the amount of co-transfected EGFP mRNA. The normalized level of mRNA from cells with 3 x FLAG hGI-c tail was defined as 100 %. B. Western blotting analysis of the same cellular extracts of A. EGFP was used as a control.

11. Translation of PTC-containing β-globin mRNAs are repressed

To check the second possibility of rare level of protein expressions even from sufficient transcripts, polysome analysis was used and hGI-g tail mRNAs in their distribution on polysomes were assessed. HeLa cells were transfected with h*UPF1* or nonspecific control siRNA, for 1 day, after which they were re-transfected with 3 x FLAG hGI-c tail or 3 x FLAG hGI-g tail. Cells were lysed and cytosol part was prepared and subjected to sedimentation on 10 to 50 % sucrose gradient. In cells transfected with the 3 x FLAG hGI-c tail, this mRNA was detected in the polysome-containing fractions (similar to endogenous GAPDH mRNA) (Fig. 12A), indicating that this mRNA is being actively translated. Cells were then treated with puromycin to inhibit translation at the same time as transfection with 3 x FLAG hGI-c tail. A large percentage of the 3 x FLAG hGI-c tail mRNA and *GAPDH* mRNA shifted to monosome-containing fractions (Fig. 12B), indicating a decrease in its translation rate. In cells transfected with the 3 x FLAG hGI-g tail plasmid which is up-regulated by si-h*UPF1*, this mRNA positioned to monosome-containing and light polysome fractions similar to 3 x FLAG hGI-c tail trated with puromycin (Fig. 12B and 12C). These results indicate that this mRNA is not translated efficiently, and rare protein expressions of PTC-containing mRNAs were due to translational repression even though PTC-containing mRNAs were recovered from NMD inhibition.

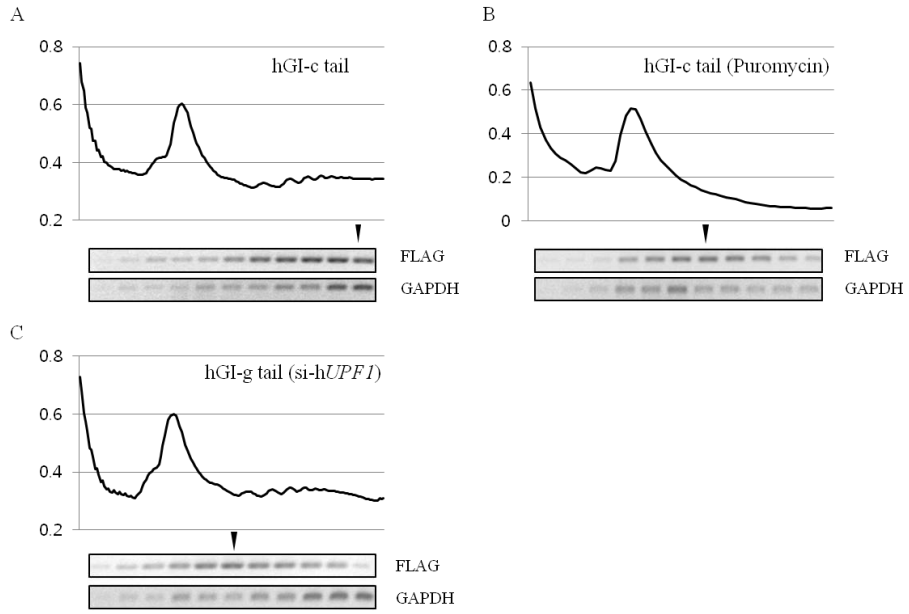


Figure 12. Recovered 3 x FLAG hGI-g tail mRNAs are released from polysomes by NMD inhibition. HeLa cells transfected with siRNA of negative control or h*UPF1*, and 1 day after, were re-transfected with 3 x FLAG hGI-c tail (A), 3 x FLAG hGI-c tail and subsequently treated with 100 μg/mL puromycin 2 hours prior to harvest (B), or with 3 x FLAG hGI-g tail (C). Cytoplasmic lysates were prepared and layered onto 10 – 50 % sucrose gradient and centrifuged for 2 hours at 4 ℃ at 39,000 rpm. Fractions were collected and absorbance at 254 nm was read for each fraction (top panels). RNAs extracted from each fraction were subjected to RT-PCR for transfected β-globin mRNA (middle panels) or GAPDH (bottom panels). Data are representative of three independent experiments.

12. eIF4A3 is involved in the translational repression of PTC-containing mRNAs recovered from NMD inhibition

One of the differences between hGI-c tail and hGI-g tail plasmid is splicing event which occurs at next to the first stop codon, thus EJC complex and surrounding molecules might have roles in this translation repression. To prove this hypothesis, among EJC components, eIF4A3 and Y14 were individually down-regulated by siRNAs. HeLa cells were transfected with

each siRNA, two days after re-transfected with 3 x FLAG EGFP, either 3 x FLAG hGI-c tail or 3 x FLAG hGI-g tail. To inhibit NMD, the 100 μM emetine was treated to transfected cells for 4 hours prior to harvest. RT-PCR analyses were used for validation of knockdown of siRNAs, and real-time PCR analyses were used for validation of up-regulation of hGI-g tail mRNAs by emetine. Without a significant difference between hGI-g tail mRNAs treated with siRNA of control or target, inhibition of eIF4A3 results in the protein expressions from the PTC-containing transcripts recovered from NMD inhibition (Fig. 13).

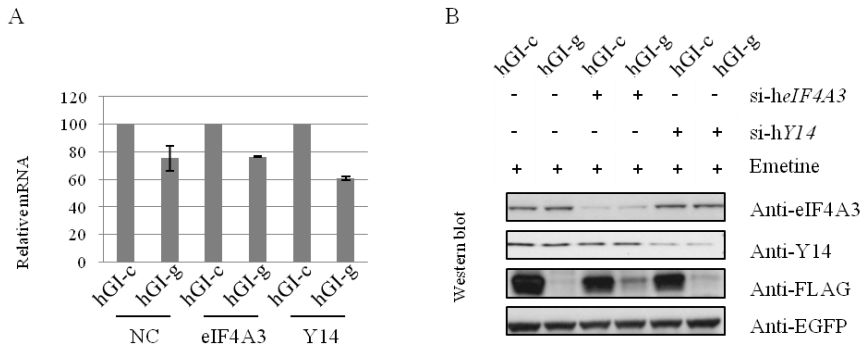


Figure 13. eIF4A3 is involved in translation repression of PTC-containing mRNAs, when NMD is blocked. A. Real-time PCR analysis of hGI in HeLa cells transfected with control siRNA or si-h*eIF4A3*, two days after co-transfected with 3 x FLAG EGFP, either 3 x FLAG hGI-c tail or 3 x FLAG hGI-g tail. Cells were treated emetine at 100 μM for 4 hours prior to harvest. Transfected β-globin mRNA was normalized to the amount of co-transfected EGFP mRNA. The normalized level of mRNA from cells with 3 x FLAG hGI-c tail was defined as 100 %. B. Western blotting analysis of the same cellular extracts of A. EGFP was used as a control.

IV. DISCUSSION

NMD is a post-transcriptional mechanism that detects PTC-containing mRNAs and subsequently inhibits translation at the initiation step and actively degrades abnormal mRNA. Here, 12 novel cMNR-containing genes frequently mutated in MSI-H tumors were identified using gene expression and mutation analysis in NMD-blocked cells. The systematic identification of frameshift mutations in cMNR is an effective way to identify target genes in MSI-H cancers. Previous studies showed mononucleotide repeats in the gene database and identified several target genes[21,43-46]. However, a second gene database search is needed to determine whether the MNR is located in the coding region and mutation analysis is needed to confirm that the genes are related to MSI-H tumorigenesis. The identification of MNR using gene data base searches results in more than 4,300 genes (in the genes containing more than 7 nucleotide repeat numbers), and a large portion of these MNRs are not located in coding regions. Owing to the large number of genes containing cMNR, random selection of genes to confirm the mutations has been used in the past. The inclusion of an NMD block to facilitate the identification of genes in MSI-H tumors has two advantages. First, the selection of MNR-containing genes with increased expression after NMD block can identify target genes of MSI-H tumorigenesis more specifically. Second, this method can identify the homogenous group of inactivated genes through gene dosage effects in MSI-H tumors[47]. The other genes that produce mutant transcripts might play different roles in MSI-H tumorigenesis and are not detected by this method.

There are two reports for the identification of target genes involved in MSI-H carcinogenesis by blocking NMD. Ionov *et al*. identified two novel genes, *UVRAG* and *p300* that are up-regulated following emetine and actinomycin D treatment in two cell lines (LS180, RKO) using cancer specific

microarrays[16]. In their study, drug treatment was not a specific block method, as they mentioned, and they could analyze only 2,500 cancer-related genes and found only two novel genes. El-Bchiri *et al*. blocked NMD using siRNA of h*UPF1* in HCT116 cells and identified four cMNR-containing genes with more than nine nucleotide repeats that are mutated in MSI-H tumors. In their study, among the 19 cMNR-containing genes up-regulated following a NMD block, only six were mutated in HCT116 cells[17]. These two previous reports suggest NMD block induces many non-specific alterations in gene expression, and therefore can identify faulty target genes. To overcome the previous limitations of NMD block, I tried to exclude non-specifically altered genes by blocking NMD in an MSS cell line and subtracting the elevated genes which shared with the two MSI-H cancer cell lines and analyzed with database. Using this approach, I was able to find specific mutations in 12 novel cMNR-containing genes with more than nine nucleotide repeats. These genes had altered expression following NMD block, and mutations in these genes were found. The combination of NMD block in many frequently frameshift mutated cell lines, microarray analysis, subtraction of non-specific genes, gene data base search for cMNR, and mutation analysis in MSI-H tumors will make it possible to accurately identify MSI-H target genes in the future.

Among newly identified 12 novel genes here, most frequently mutated gene, mixed-lineage leukemia 3 (*MLL3*), was reported as a cancer candidate (CAN) gene for colon cancer[48]. In addition, MLL3 is known to act as a p53 coactivator[49] therefore decreased MLL3 is expected to inhibit p53 associated tumor suppressor activity. Functional study of mutation in cMNR of *MLL3* is needed to elucidate the direct relationship between mutations in cMNRs and tumorigenesis. Other novel genes have been reported to affect a wide range of cellular functions, including coronary artery disease and Alzheimer disease (*KIAA0391 and RUFY2*), cell cycle regulation (*CDC7 and SGOL2*), development (*PHACTR4*), cell signaling (*RAPGEF6*), mRNA surveillance

component (*SMG7*), and cell migration (*SPAG9*)[50-54]. Although there have been no reports directly relating colon cancer to the other target genes, the frameshift mutations of these genes might contribute to the tumorigenesis of MSI-H tumors[13,14]. Further functional studies of these mutations will provide clues to understanding their roles in MSI-H tumor development and progression.

This study indicates that PTC-containing mRNAs recovered by NMD inhibition are translationally repressed. The importance of this finding is that there are several physiological conditions of NMD inhibition. NMD is inhibited under certain circumstances, thus transcripts down-regulated by NMD have chance to translate proteins which have potential to be deleterious. Considering one of three alternative spliced transcripts are degraded by NMD and numerous endogenous NMD targets, NMD inhibition and its expected results are crucial issues[55]. In fact, recent studies have demonstrated that NMD is inhibited in hypoxic cells[42]. Hypoxia is occurred in tumor inner mass and effects on tumor micro-environment including angiogenesis.

Because truncated protein from PTC-containing mRNA has potential to be deleterious, many studies tried to determine truncated protein detection. Several studies reported that truncated proteins have not been detected through western blot or immunoprecipitation experiments: truncated XPC[56], BubR1[57], PRPF31[58], BRCA1[59], CHK2[59], LMAN[46]. They explained that difficulty of detection for these endogenous truncated proteins is due to uncertain stability, inefficiency of antibody, and experimental lack of sensitivity. Thus it is difficult to conclude that these truncated proteins can be generated. In this study, it is firstly examined that truncated protein production issues through stable human β-globin protein which is regarded as truncated protein. I investigated that truncated protein from PTC-containing mRNA is almost repressed at translation level.

There are possible explanations of discrepancy between constantly rare

protein expressions from the sufficiently restored PTC-containing transcripts under the conditions of NMD block. One explanation is that truncated protein is intrinsically unstable or quickly degraded because this protein might fail to fold properly. However this possibility can be ruled out in this study, because the proteins are stably expressed from control vector (Fig. 9), and truncated proteins are not increased after MG132 treatment (Fig. 11). In this study, I designed that the same transcripts should be generated from both constructs 3 x FLAG hGI-c tail and 3 x FLAG hGI-g tail, to exclude translation differences from different proteins. 3 x FLAG hGI-c tail, new constructed NMD irrelevant control construct is expressed stable β-globin protein, considering its protein was showed at the same size of wild type human β-globin protein (Fig. 9). Therefore the low level of relative protein expressions from 3 x FLAG hGI-g tail is not related to protein stability. In addition, when HeLa cells were treated with MG132, a proteasome inhibitor, truncated protein was still hardly detected (Fig. 11). In fact, Rio Frio *et al.* blocked lysosomal and proteasomal degradation with NMD inhibition in primary cultured mutant cell line for the detection of mutant PRPF31 protein. Despite this inhibition, they could not detect the truncated PRPF31 protein[58]. Second possible explanation is truncated protein translate improperly. You *et al.* reported that mutant alleles in cancer cell line are repressed at translation level, in which mutations encode PTC, however its transcript escape NMD[29]. Similar to this observation, NMD sensitive PTC-containing mRNAs are also translationally repressed (Fig. 12). The last possible explanation is that the rare truncated protein is due to insufficient mRNA recovery. The PTC-containing transcripts from the cells transfected with 3 x FLAG hGI-g tail were restored up to about 70 % of control transcripts. However, unlike the cells tranfected with control vectors which produce large amount of proteins, the cells transfected with 3 x FLAG hGI-g tail did not produce any detectable proteins. Considering all of these findings, the discrepancy between rare

protein expressions and sufficiently restored transcripts is not due to protein stability, rapid degradation, and insufficient mRNA recovery. The translational repression of the PTC-containing mRNAs is mostly suspected.

In this study, I suggested that eIF4A3 is involved in this repression (Fig. 13). eIF4A3 is a member of the DEAD-box helicase proteins, and has about 70 % amino acid homology with eIF4A[60]. eIF4A which has a crucial function in cap-dependent translation initiation by part of the heterotrimeric complex eIF4F (eIF4A, eIF4E, and eIF4G)[61]. eIF4A3 is well known as a central component of EJC. EJC is deposited on the mRNA during splicing upstream of the exon junction[4,62], and eIF4A3 serves as a molecular placeholder for the subsequent assembly by binding mRNA directly [5,63-65]. Thus eIF4A3 is essential for NMD[66,67]. Except the role of eIF4A3 in NMD, specific mechanism of eIF4A3 in the translation repression of recovered PTC-containing mRNA is unclear yet, further functional studies will provide clues to understanding this mechanism.

In conclusion, I have demonstrated that PTC-containing mRNAs derived from frameshift mutation are degraded by NMD and consequently their protein expressions are low. I verified that this rare protein expression is maintained although their relative mRNA is recovered by NMD inhibition, and resulted from translation repression. I also identified eIF4A3 protein is involved in this translation repression. All of these findings give clues of MSI-H tumor carcinogenesis and gene regulation in specific cellular condition relevant to NMD.

V. CONCLUSION

In order to identify the target genes of MSI-H tumor carcinogenesis and the functional relevance of mutant genes, I have systematically searched the mutant genes by NMD inhibition and validated the proteins expression from the transcripts of mutated genes.

1. Twelve novel genes frequently mutated in MSI-H colon cancers.
2. mRNAs of these mutated genes contain PTC and are decayed by NMD.
3. Protein expressions of these mutated genes are rarely detected.
4. mRNAs of NMD sensitive construct are recovered by NMD inhibition, however their proteins are rarely detected.
5. Low level of truncated proteins from recovered mRNAs of NMD target are due to translation repression.
6. eIF4A3 protein is involved in this translation repression.

In this study, I demonstrated clues for the MSI-H tumor carcinogenesis with identification of novel twelve genes through functional inactivation, and suggested their translational repression in a certain tumor micro-environment in which NMD is inhibited.

REFFERENCES

1. Hillman RT, Green RE, Brenner SE. An unappreciated role for RNA surveillance. Genome Biol. 2004;5(2):R8.
2. Zhang J, Sun X, Qian Y, LaDuca JP, Maquat LE. At least one intron is required for the nonsense-mediated decay of triosephosphate isomerase mRNA: a possible link between nuclear splicing and cytoplasmic translation. Mol Cell Biol. 1998;18(9):5272-83.
3. Zhang J, Sun X, Qian Y, Maquat LE. Intron function in the nonsense-mediated decay of beta-globin mRNA: indications that pre-mRNA splicing in the nucleus can influence mRNA translation in the cytoplasm. RNA. 1998;4(7):801-15.
4. Le Hir H, Izaurralde E, Maquat LE, Moore MJ. The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. EMBO J. 2000;19(24):6860-9.
5. Andersen CB, Ballut L, Johansen JS, Chamieh H, Nielsen KH, Oliveira CL, et al. Structure of the exon junction core complex with a trapped DEAD-box ATPase bound to RNA. Science. 2006;313(5795):1968-72.
6. Bono F, Ebert J, Lorentzen E, Conti E. The crystal structure of the exon junction complex reveals how it maintains a stable grip on mRNA. Cell. 2006;126(4):713-25.
7. Kashima I, Yamashita A, Izumi N, Kataoka N, Morishita R, Hoshino S, et al. Binding of a novel SMG-1-Upf1-eRF1-eRF3 complex (SURF) to the exon junction complex triggers Upf1 phosphorylation and nonsense-mediated mRNA decay. Genes Dev. 2006;20(3):355-67.
8. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860-921.

9. Ellegren H. Microsatellites: simple sequences with complex evolution. Nat Rev Genet. 2004;5(6):435-45.
10. Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. Nature. 1998;396(6712):643-9.
11. Ionov Y, Peinado MA, Malkhosyan S, Shibata D, Perucho M. Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. Nature. 1993;363(6429):558-61.
12. Thibodeau SN, Bren G, Schaid D. Microsatellite instability in cancer of the proximal colon. Science. 1993;260(5109):816-9.
13. Perucho M. Tumors with microsatellite instability: many mutations, targets and paradoxes. Oncogene. 2003;22(15):2223-5.
14. Markowitz S, Wang J, Myeroff L, Parsons R, Sun L, Lutterbaugh J, et al. Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. Science. 1995;268(5215):1336-8.
15. Ionov Y, Yamamoto H, Krajewski S, Reed JC, Perucho M. Mutational inactivation of the proapoptotic gene BAX confers selective advantage during tumor clonal evolution. Proc Natl Acad Sci U S A. 2000;97(20):10872-7.
16. Ionov Y, Nowak N, Perucho M, Markowitz S, Cowell JK. Manipulation of nonsense mediated decay identifies gene mutations in colon cancer Cells with microsatellite instability. Oncogene. 2004;23(3):639-45.
17. El-Bchiri J, Guilloux A, Dartigues P, Loire E, Mercier D, Buhard O, et al. Nonsense-mediated mRNA decay impacts MSI-driven carcinogenesis and anti-tumor immunity in colorectal cancers. PLoS One. 2008;3(7):e2583.
18. Grady WM, Myeroff LL, Swinler SE, Rajput A, Thiagalingam S, Lutterbaugh JD, et al. Mutational inactivation of transforming growth

factor beta receptor type II in microsatellite stable colon cancers. Cancer Res. 1999;59(2):320-4.

19. Maquat LE, Tarn WY, Isken O. The pioneer round of translation: features and functions. Cell. 142(3):368-74.

20. Gardner LB. Nonsense-mediated RNA decay regulation by cellular stress: implications for tumorigenesis. Mol Cancer Res. 8(3):295-308.

21. Kim NG, Rhee H, Li LS, Kim H, Lee JS, Kim JH, et al. Identification of MARCKS, FLJ11383 and TAF1B as putative novel target genes in colorectal carcinomas with microsatellite instability. Oncogene. 2002;21(33):5081-7.

22. Matsuda D, Hosoda N, Kim YK, Maquat LE. Failsafe nonsense-mediated mRNA decay does not detectably target eIF4E-bound mRNA. Nat Struct Mol Biol. 2007;14(10):974-9.

23. Daigle N, Ellenberg J. LambdaN-GFP: an RNA reporter system for live-cell imaging. Nat Methods. 2007;4(8):633-6.

24. El-Bchiri J, Buhard O, Penard-Lacronique V, Thomas G, Hamelin R, Duval A. Differential nonsense mediated decay of mutated mRNAs in mismatch repair deficient colorectal cancers. Hum Mol Genet. 2005;14(16):2435-42.

25. Ishigaki Y, Li X, Serin G, Maquat LE. Evidence for a pioneer round of mRNA translation: mRNAs subject to nonsense-mediated decay in mammalian cells are bound by CBP80 and CBP20. Cell. 2001;106(5):607-17.

26. Ivanov I, Lo KC, Hawthorn L, Cowell JK, Ionov Y. Identifying candidate colon cancer tumor suppressor genes using inhibition of nonsense-mediated mRNA decay in colon cancer cells. Oncogene. 2007;26(20):2873-84.

27. Linde L, Boelz S, Nissim-Rafinia M, Oren YS, Wilschanski M, Yaacov Y, et al. Nonsense-mediated mRNA decay affects nonsense

transcript levels and governs response of cystic fibrosis patients to gentamicin. J Clin Invest. 2007;117(3):683-92.

28. Chung KY, Kim NG, Li LS, Kim H, Nam CM, Shin DH. Clinicopathologic characteristics related to the high variability of coding mononucleotide repeat sequences in tumors with high-microsatellite instability. Oncol Rep. 2003;10(2):439-44.

29. You KT, Li LS, Kim NG, Kang HJ, Koh KH, Chwae YJ, et al. Selective translational repression of truncated proteins from frameshift mutation-derived mRNAs in tumors. PLoS Biol. 2007;5(5):e109.

30. Woerner SM, Gebert J, Yuan YP, Sutter C, Ridder R, Bork P, et al. Systematic identification of genes with coding microsatellites mutated in DNA mismatch repair-deficient cancer cells. Int J Cancer. 2001;93(1):12-9.

31. Woerner SM, Benner A, Sutter C, Schiller M, Yuan YP, Keller G, et al. Pathogenesis of DNA repair-deficient cancers: a statistical meta-analysis of putative Real Common Target genes. Oncogene. 2003;22(15):2226-35.

32. Park J, Betel D, Gryfe R, Michalickova K, Di Nicola N, Gallinger S, et al. Mutation profiling of mismatch repair-deficient colorectal cncers using an in silico genome scan to identify coding microsatellites. Cancer Res. 2002;62(5):1284-8.

33. Ikenoue T, Togo G, Nagai K, Ijichi H, Kato J, Yamaji Y, et al. Frameshift mutations at mononucleotide repeats in RAD50 recombinational DNA repair gene in colorectal cancers with microsatellite instability. Jpn J Cancer Res. 2001;92(6):587-91.

34. Ropero S, Fraga MF, Ballestar E, Hamelin R, Yamamoto H, Boix-Chornet M, et al. A truncating mutation of HDAC2 in human cancers confers resistance to histone deacetylase inhibition. Nat Genet.

2006;38(5):566-9.

35. Paun BC, Cheng Y, Leggett BA, Young J, Meltzer SJ, Mori Y. Screening for microsatellite instability identifies frequent 3'-untranslated region mutation of the RB1-inducible coiled-coil 1 gene in colon tumors. PLoS One. 2009;4(11):e7715.

36. Duval A, Rolland S, Compoint A, Tubacher E, Iacopetta B, Thomas G, et al. Evolution of instability at coding and non-coding repeat sequences in human MSI-H colorectal cancers. Hum Mol Genet. 2001;10(5):513-8.

37. Li LS, Kim NG, Kim SH, Park C, Kim H, Kang HJ, et al. Chromosomal imbalances in the colorectal carcinomas with microsatellite instability. Am J Pathol. 2003;163(4):1429-36.

38. Blobel G, Sabatini D. Dissociation of mammalian polyribosomes into subunits by puromycin. Proc Natl Acad Sci U S A. 1971;68(2):390-4.

39. Shin N, You KT, Lee H, Kim WK, Song M, Choi HJ, et al. Identification of frequently mutated genes with relevance to nonsense mediated mRNA decay in the high microsatellite instability cancers. Int J Cancer. 128(12):2872-80.

40. Baserga SJ, Benz EJ, Jr. Nonsense mutations in the human beta-globin gene affect mRNA metabolism. Proc Natl Acad Sci U S A. 1988;85(7):2056-60.

41. Neu-Yilik G, Gehring NH, Thermann R, Frede U, Hentze MW, Kulozik AE. Splicing and 3' end formation in the definition of nonsense-mediated decay-competent human beta-globin mRNPs. EMBO J. 2001;20(3):532-40.

42. Gardner LB. Hypoxic inhibition of nonsense-mediated RNA decay regulates gene expression and the integrated stress response. Mol Cell Biol. 2008;28(11):3729-41.

43. Woerner SM, Kloor M, Mueller A, Rueschoff J, Friedrichs N,

Buettner R, et al. Microsatellite instability of selective target genes in HNPCC-associated colon adenomas. Oncogene. 2005;24(15):2525-35.

44. Woerner SM, Kloor M, Schwitalle Y, Youmans H, Doeberitz MK, Gebert J, et al. The putative tumor suppressor AIM2 is frequently affected by different genetic alterations in microsatellite unstable colon cancers. Genes Chromosomes Cancer. 2007;46(12):1080-9.

45. Korff S, Woerner SM, Yuan YP, Bork P, von Knebel Doeberitz M, Gebert J. Frameshift mutations in coding repeats of protein tyrosine phosphatase genes in colorectal tumors with microsatellite instability. BMC Cancer. 2008;8:329.

46. Roeckel N, Woerner SM, Kloor M, Yuan YP, Patsos G, Gromes R, et al. High frequency of LMAN1 abnormalities in colorectal tumors with microsatellite instability. Cancer Res. 2009;69(1):292-9.

47. Noensie EN, Dietz HC. A strategy for disease gene identification through nonsense-mediated mRNA decay inhibition. Nat Biotechnol. 2001;19(5):434-9.

48. Sjoblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, et al. The consensus coding sequences of human breast and colorectal cancers. Science. 2006;314(5797):268-74.

49. Lee J, Kim DH, Lee S, Yang QH, Lee DK, Lee SK, et al. A tumor suppressive coactivator complex of p53 containing ASC-2 and histone H3-lysine-4 methyltransferase MLL3 or its paralogue MLL4. Proc Natl Acad Sci U S A. 2009;106(21):8513-8.

50. Alsmadi O, Muiya P, Khalak H, Al-Saud H, Meyer BF, Al-Mohanna F, et al. Haplotypes encompassing the KIAA0391 and PSMA6 gene cluster confer a genetic link for myocardial infarction and coronary artery disease. Ann Hum Genet. 2009;73(Pt 5):475-83.

51. Xu Z, Cetin B, Anger M, Cho US, Helmhart W, Nasmyth K, et al. Structure and function of the PP2A-shugoshin interaction. Mol Cell.

2009;35(4):426-41.

52. Kim TH, Goodman J, Anderson KV, Niswander L. Phactr4 regulates neural tube and optic fissure closure by controlling PP1-, Rb-, and E2F1-regulated cell-cycle progression. Dev Cell. 2007;13(1):87-102.

53. Majercak J, Ray WJ, Espeseth A, Simon A, Shi XP, Wolffe C, et al. LRRTM3 promotes processing of amyloid-precursor protein by BACE1 and is a positional candidate gene for late-onset Alzheimer's disease. Proc Natl Acad Sci U S A. 2006;103(47):17967-72.

54. Dube N, Kooistra MR, Pannekoek WJ, Vliem MJ, Oorschot V, Klumperman J, et al. The RapGEF PDZ-GEF2 is required for maturation of cell-cell junctions. Cell Signal. 2008;20(9):1608-15.

55. Lewis BP, Green RE, Brenner SE. Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. Proc Natl Acad Sci U S A. 2003;100(1):189-92.

56. Khan SG, Oh KS, Shahlavi T, Ueda T, Busch DB, Inui H, et al. Reduced XPC DNA repair gene mRNA levels in clinically normal parents of xeroderma pigmentosum patients. Carcinogenesis. 2006;27(1):84-94.

57. Matsuura S, Matsumoto Y, Morishima K, Izumi H, Matsumoto H, Ito E, et al. Monoallelic BUB1B mutations and defective mitotic-spindle checkpoint in seven families with premature chromatid separation (PCS) syndrome. Am J Med Genet A. 2006;140(4):358-67.

58. Rio Frio T, Wade NM, Ransijn A, Berson EL, Beckmann JS, Rivolta C. Premature termination codons in PRPF31 cause retinitis pigmentosa via haploinsufficiency due to nonsense-mediated mRNA decay. J Clin Invest. 2008;118(4):1519-31.

59. Anczukow O, Ware MD, Buisson M, Zetoune AB, Stoppa-Lyonnet D, Sinilnikova OM, et al. Does the nonsense-mediated mRNA decay mechanism prevent the synthesis of truncated BRCA1, CHK2, and

p53 proteins? Hum Mutat. 2008;29(1):65-73.

60. Li Q, Imataka H, Morino S, Rogers GW, Jr., Richter-Cook NJ, Merrick WC, et al. Eukaryotic translation initiation factor 4AIII (eIF4AIII) is functionally distinct from eIF4AI and eIF4AII. Mol Cell Biol. 1999;19(11):7336-46.

61. Rozen F, Edery I, Meerovitch K, Dever TE, Merrick WC, Sonenberg N. Bidirectional RNA helicase activity of eucaryotic translation initiation factors 4A and 4F. Mol Cell Biol. 1990;10(3):1134-44.

62. Le Hir H, Moore MJ, Maquat LE. Pre-mRNA splicing alters mRNP composition: evidence for stable association of proteins at exon-exon junctions. Genes Dev. 2000;14(9):1098-108.

63. Ballut L, Marchadier B, Baguet A, Tomasetto C, Seraphin B, Le Hir H. The exon junction core complex is locked onto RNA by inhibition of eIF4AIII ATPase activity. Nat Struct Mol Biol. 2005;12(10):861-9.

64. Le Hir H, Andersen GR. Structural insights into the exon junction complex. Curr Opin Struct Biol. 2008;18(1):112-9.

65. Shibuya T, Tange TO, Stroupe ME, Moore MJ. Mutational analysis of human eIF4AIII identifies regions necessary for exon junction complex formation and nonsense-mediated mRNA decay. RNA. 2006;12(3):360-74.

66. Ferraiuolo MA, Lee CS, Ler LW, Hsu JL, Costa-Mattioli M, Luo MJ, et al. A nuclear translation-like factor eIF4AIII is recruited to the mRNA during splicing and functions in nonsense-mediated decay. Proc Natl Acad Sci U S A. 2004;101(12):4118-23.

67. Shibuya T, Tange TO, Sonenberg N, Moore MJ. eIF4AIII binds spliced mRNA in the exon junction complex and is essential for nonsense-mediated decay. Nat Struct Mol Biol. 2004;11(4):346-51.

ABSTRACT (IN KOREAN)

조기 종결 코돈을 가진 돌연변이 유전자들과 nonsense-mediated mRNA decay 억제 후의 번역 억제 규명

<지도교수 김 호 근>

연세대학교 대학원 의과학과

신 나 라

유전자에 발생한 돌연변이 중 체이동 돌연변이 (Frameshift mutation) 는 각종 암 특히 현미부수체 불안정형 (Microsatellite instability; MSI-H) 암세포에서 흔히 존재하는 원인적 유전 변화이다. 특정 유전자에 체이동 돌연변이가 일어날 경우, 전사된 mRNA는 예외 없이 조기 종결 코돈 (premature termination codon; PTC)이 생기게 되는데 이는 mRNA 수준에서 nonsense mediated mRNA decay (NMD) 라는 기전에 의해 전사체의 분해가 일어나 최종적인 단백질 생성은 발생하지 않는다. 본 연구에서는 두 개의 현미부수체 불안정형 암세포주에서 NMD 를 저해하여 체이동 돌연변이가 빈번한 새로운 유전자들을 찾았다. NMD 저해 후, DNA microarray를 통하여 발현에 변화가 있는 유전자들을 선정한 후, database와의 비교를 통해 coding region에 9개 이상의 단염기 서열을 가진 28개의 유전자를 골라내었다. 이 중에서 13개의 유전자는 이전에 해당 단염기 서열에서 돌연변이가 보고되어 있었다. 나머

지 15개의 유전자 중 12개의 유전자가 10개의 현미부수체 불안정형 암세포주와 암환자의 조직에서 실제로 체이동 돌연변이가 있는 것을 확인하였으며 또한 이러한 체이동 돌연변이에 의해 조기 종결 코돈이 생성되어 NMD 에 의해 전사체가 분해되고 궁극적으로는 단백질이 생성되지 않음을 증명하였다. 그러나 NMD는 전사체 수준에서 변형된 단백질의 생성을 억제하는 효율적인 기전이기는 하나 생체 내의 여러 자극들에 의해 NMD 가 저해되는 경우들이 있는데, 이 때 조기 종결 코돈을 포함한 전사체들의 발현이 존재하게 될 것으로 예측된다. 따라서 생체 내 NMD 가 저해되는 상황에서, 분해되지 않은 전사체에 의해 절단형 돌연변이 단백질이 생성될 수 있는지 확인하고자, β-globin 유전자를 변형한 vector system을 사용하여 NMD를 저해하여 그 단백질 발현을 조사하였다. NMD를 저해하는 여러 약제들과, NMD에 관여하는 upf1 유전자의 siRNA 처리, hypoxia 유도 등으로 NMD를 저해하였을 때, 조기 종결 코돈을 포함한 vector의 전사체가 충분히 발현함을 확인하였고, 그럼에도 그 단백질은 거의 발현되지 않는 것을 확인하였다. 이상과 같이 돌연변이형 전사체는 존재하나 해당 단백질이 발현되지 않는 원인이 단백질 합성 저해, 생성된 단백질의 조기 분해, 발현된 전사체 양의 부족 중 어느 것인가를 규명하기 위해 발현 전사체의 정량적 측정, polysome assay를 이용한 효과적 단백질 합성 탐색, MG132 처리를 통한 프로테아좀 분해 억제 등의 방법으로 조사한 결과 조기 종결 코돈을 포함한 전사체에서 단백질 합성이 억제되어 궁극적으로 단백질 발현이 억제됨을 밝혔다. 돌연변이 전사체의 단백질 발현 억제에 어떠한 단백질이 관여하는지를 규명하기 위해 exon junction complex에 속하는 eIF4A3 단백질의 억제시 단백질 합성 저해의 효과가 줄어드는 것을 발견하였다. 이상의 연구를 통해 체이동 돌연

변이에 의해 생성된 돌연변이 중 조기 종결 코돈을 포함한 전사체는 NMD에 의해 단백질 발현이 저해되며, 설령 생체 내 여러 자극에 의해 전사체 분해가 이루어지지 않더라도 단백질 합성 수준에서 억제가 일어나 궁극적으로 돌연변이형 단백질 생성에 의한 세포 및 개체 손상이 예방된다고 사료된다.

핵심되는 말: coding region에 존재하는 단염기 서열, 체이동 돌연변이, 조기 종결 코돈, nonsense mediated mRNA decay, 전사 저해, h*eIF4A3*

PUBLICATION LIST

1. Lee H, Song M, **Shin N**, Shin CH, Min BS, Kim HS, et al. (2012) Diagnostic significance of serum HMGB1 in colorectal carcinomas. PLoS One. 7(4):e34318

2. **Shin N**, You KT, Lee H, Kim WK, Song M, Choi HJ, et al. (2011) Identification of frequently mutated genes with relevance to nonsense mediated mRNA decay in the high microsatellite instability cancers. Int J Cancer. 128(12):2872-80

3. Lee H*, **Shin N***, Song M, Kang UB, Yeom J, Lee C, et al (2010) Analysis of nuclear high mobility group box 1 (HMGB1)-binding proteins in colon cancer cells: clustering with proteins involved in secretion and extranuclear function. J Proteome Res. 9(9):4661-70 *equal contribution

4. Lee J, Lee SH, **Shin N**, Jeong M, Kim MS, Kim MJ, et al (2009) Tumor necrosis factor-alpha enhances IL-15-induced natural killer cell differentiation. Biochem Biophys Res Commun. 386(4):718-23

5. Chung JW, Kim MS, Piao ZH, Jeong M, Yoon SR, **Shin N**, et al (2008) Osteopontin promotes the development of natural killer cells from hematopoietic stem cells. Stem Cells. 26(8):2114-23

6. **Shin N**, Lee J, Lee J, Jeong M, Kim MS, Lee SH, et al (2008) Expression of Gpnmb in NK cell development from hematopoietic stem cells. Immune Network. 8(2):53-58