Characterization of Alternative Splicing of LSD-1 in Neuroblastoma Cell Lines

Weichun Li
Department of Biological Science
Fordham University

Abstract

LSD1 is a gene that encodes a chromatin modifying protein which functions as a histone demethylase. Although numerous researches have been done to investigate how this protein regulates gene transcription and interacts with DNA methylation, only a few is known about the alternative splicing of this gene. In this project, I detect three possible alternative spliced forms in neuroblastoma cell lines, the skipping of exon 2, exon 7 and a novel insertion between exon 2 and 3. Also, RT-PCR is conducted to see if there is any differential expression between transcript variants among six cell lines. According to the RT-PCR and sequencing results, exon 2 and 7 are preserved in the RNA of LSD1 in six neuroblastoma cell lines. There is a novel insertion of 60 bases between exon 2 and 3 which encodes 20 amino acids in one transcript variant observed. The sequences before and after this insertion are in great similarity to the consensus sequence of donor and acceptor site with only one nucleotide difference. This may explain why in some cases, this insertion is cut out while in other cases, it is preserved. Meanwhile, the location of this insertion is right in front of the RNA that encodes SWIRM domain which is a highly conserved domain present in many histone modifying and remolding complexes. According to recently published papers, this domain functions by improving the stability of LSD1 protein and interaction with amine oxidase domain to form catalytic cavity.

Therefore, this 20 amino-acid peptide encoded by the 60-base insertion may exert structural and functional influence on this domain and the protein but details remain to be identified.

Introduction

DNA in the nucleus of eukaryotic cell is wrapped around octamers of histone protein, called nucleosomes that allow for compaction of DNA (Luger, K *et al.* 1997). Besides the structural role in nucleosome, histone proteins are also key players in the regulation of gene expression. The known regulations depend partially on their post-translational modification including phosphorylation, acetylation, methylation, etc primarily on the N-terminal tail of histones. These epigenetic markers endowed by various modifications are under tight regulation of specific chromatin-modifying enzymes, which usually function as part of large multiprotein complexes (Kouzaride, T 2007).

LSD-1 is the gene that encodes a histone demethylase, catalyzing demethylation of mono- and di-methylated histone at H3K4 or H3K9 (Qingjun Zhu *et al*. 2008). This protein contains a SWIRM domain, which functions as a putative protein-protein interaction motif, an amine oxidase domain which harbors the demethylase activity and requires FAD as the cofactor (Eric Metzger, *et al* 2005). Besides, LSD1 usually functions as a subunit in protein complexes such as Co-REST, NRD, etc. (Lee, M.G. *et al.* 2005). The crystal structure and biochemical studies of LSD1 reveal that this enzyme requires H3 peptide at least 16 residues in length for substrate recognition (Ronen Marmorstein, Raymond C. Trievel 2009) and the distinction between methylated H3K4, H3K9, both of

which are substrates of this enzyme and other methylated histone sites depends on chemical selection instead of steric control (Pete Stavropoulos, *et al.* 2006).

As a histone demethylase, LSD1 can either activate or repress certain gene transcription. There is report that it silences genes by targeting at H3K4me1/2, a methylation site usually associated with transcriptional active genes (Federico Forneris, *et al.* 2008). However, there are reports that LSD1 could demethylate repressive histone markers to promote transcription. Like in normal human prostate and prostate tumor, people find LSD1 co-localizes with the androgen receptor and stimulates androgen-receptor-dependent transcription by demethylating H3K9 (Eric Metzger, *et al*. 2005). Such regulatory effect on gene expression predicts a potential link between cancer and LSD1 (Wang, G.G *et al.* 2007). Moreover, people have found in prostate cancer, the over-expression of LSD1 is associated with increasing grade and may be used as a prognostic predictor (Kahl *et al.* 2006). Although major substrate of LSD1 is histone, it also demethylates and stabilizes Dnmt1 (DNA methyltrasferase 1), thereby influencing DNA methylation system (Jing Wang, *et al* 2009).

Although the structure and function of the protein encoded by LSD 1 is under intensive investigation, which is proportional to its importance, the alternative splicing of this gene is not fully characterized. According to Entrez Gene, this gene maps on chromosome 1, at 1p36.12. It covers 64.25 kb (NCBI 36, March 2006). There are 19 exons, 27 different gt-ag introns. Transcription produces 12 alternatively spliced variants. The mRNAs appear to differ by truncation of the 5' end, truncation of the 3' end, presence or absence of 7 cassette exons, overlapping exons with different boundaries (Genebank). However, up to now, little is known about the expression of this gene in

neuroblastoma cells. Therefore, the goal of this project is to characterize alternative splicing of LSD1 in neuroblastoma cells and see if there is any differential expression among the variants observed.

Material and Methods

Cell lines

Six neuroblastoma cell lines are analyzed in this study. They are IMR-32, SH-SY5Y, BE (2)-C, CB-JMN, LA1-5S, SH-EP1. All of them are kindly provided by Dr. Ross's lab.

RNA extraction

RNA was extracted using RNasy Plus Mini Kit (QIAGEN) according to the manufacture's instruction.

RT-PCR

Reverse transcriptase PCR was run using a total reaction volume of 20 µl which included 0.8 µl dNTP mix, 0.8 µl enzyme mix, 8 pmol each of forward and reverse primers, 4 µl reverse transcriptase 5X buffer solution and 20 ng template. The volume of RNA added was determined by measuring the optical density of each RNA extraction on a spectrophotometer. Reaction volumes were brought up to 20 µl by adding appropriate volume of distilled water for each reaction. Pre-PCR holds were 30 minutes at 50 °C for cDNA synthesis and 15 minutes at 95 °C for inactivation of reverse transcriptase and activation of polymerase. The reaction was run as follows: denaturation at 94 °C (30 seconds), annealing at 57 °C (30 seconds), and extension at 72 °C (30 seconds). 50 cycles were run for primer set A and B. 35 cycles were run for primer set C. Primers were created to amplify specific regions of different exons on gene LSD 1 (Table 1). They

were designed to span over at least one intron to distinguish PCR product from RNA or genomic DNA. The RT-PCR products were analyzed on a 1.5% agarose gel and visualized by ethidium bromide in a UV trans-illuminator (BioRad).

PCR product purification

The 313 bp RT-PCR product amplified from primer C was purified using QIAquick PCR Purification Kit according to the manufacturer's protocol.

Gel purification

The 355 bp, 295 bp RT-PCR products from primer A and 266 bp, 206 bp RT-PCR products from primer B were cut from 1.5% agarose gel and purified using QIAquick Gel Extraction Kit following the manufacturer's protocol.

Primer	Position	Sequence	Product size from RNA	Product size from DNA	Designed to detect
A	Exon 1 Exon 3	F:5'-CCATGGAAACTGGAATAGCAG-3' R:5'GGCTGCTTCTTGAGAAGTCATCC3'	129 bp 295 bp 355 bp	30637 bp	Skipping of exon 2 Insertion between exon 2 and 3
B	Exon 2 Exon 3	F:5'-GAGATGGATGAAAGCTTGG-3' R:5'-GAAGTCATCCGGTCATGAGG-3'	0 bp 206 bp 266 bp	19958 bp	Skipping of exon 2 Insertion between exon 2 and 3
C	Exon 5 Exon 8	F: 5'-GATACTGTGCTTGTCCACC-3' R:5'-CTTGTTTGCTGACCACAGC-3'	206 bp 313 bp	3585 bp	Skipping of exon 7

Table 1. Primers designed for specific amplification of possible alternative spliced transcripts.

Results

Primer A specifically amplifies the region between exon 1 and exon 3. The products show two transcription variants. The size difference between them is 60 base pairs.

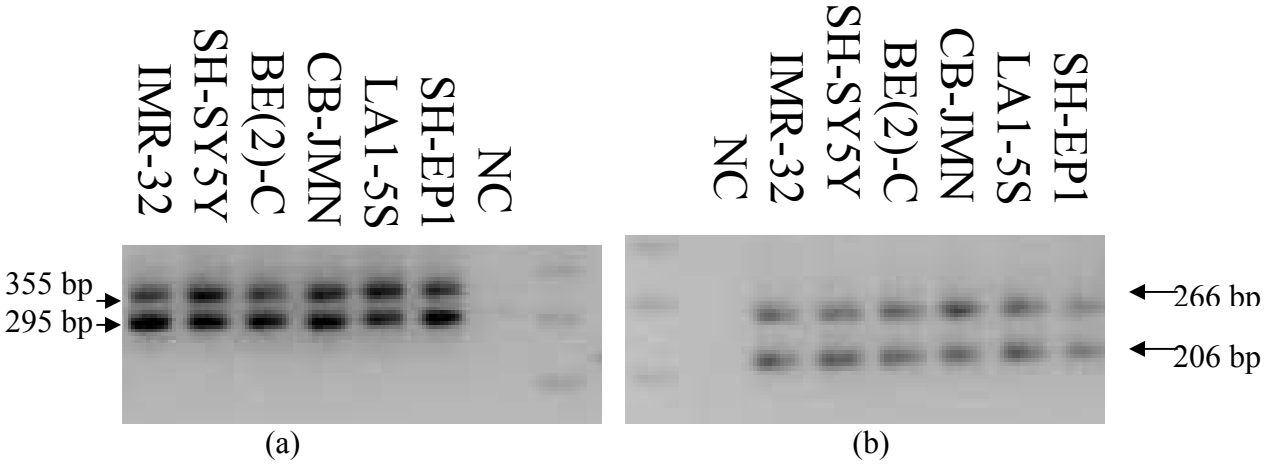

Figure 1: (a) RT-PCR (50 cycles) products amplified from exon 1 to exon 3 of LSD1 from six neuroblastoma cell lines show the same two variants. The difference in size indicates there is a 60 bp piece existing in the large band but not the small one. Meanwhile, there is no apparent differential expression between two variants in six cell lines.
(b) Two bands are produced by amplification from primer B (50 cycles RT-PCR) of LSD 1 in six neuroblastoma cell lines. The difference in size is the same as that of products amplified from primer A. Still, there is no apparent differential expression of the two variants in six cell lines

Both bands from two gels are cut and purified as described in Material and Method. The sequencing results confirm that the products are from exon 1 to exon 3 and exon 2 to exon 3 (BLAST data not shown), respectively, of LSD1. Exon 2 is not skipped.

```
Query  260      TACCNTCGNNCGCTTGCGCCGGCTGGTCCGACGCCCCTCCGGAGTCTCTGCTATTCCAGT  319
                ||||  |||  ||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  6145011  TACCTTCGCCCGCTTGCGCCGGCTGGTCCGACGCCCCTCCGGAGTCTCTGCTATTCCAGT  6144952

Query  320      T  320
                |
Sbjct  6144951  T  6144951
```

Figure 2: BLAST result of products from primer A confirms that the product is produced from exon 1 to exon 3 and exon 2 is preserved in the RNA of six neuroblastoma cell lines.

To identify the 60-base insertion which makes the difference between the two variants, we compare the sequence results of both large and small bands from primer B. The result shows the novel insertion comes from intron 2.

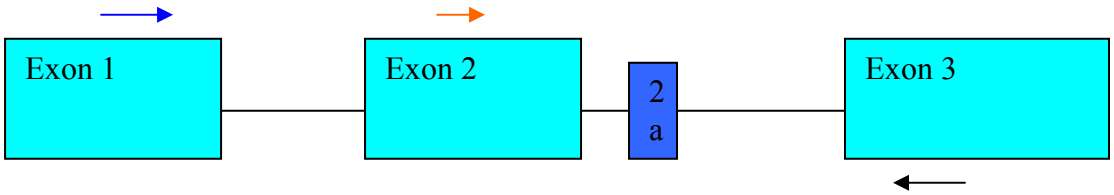

Figure 3: Primer A amplifies from exon 1 (blue arrow) to exon 3 (black arrow). Primer B amplifies from exon 2 (orange arrow) to exon 3 (black arrow). In both cases, exon 2 is preserved in mRNA. The novel insertion of 60 bases (dark blue square) comes from intron 2. The sequence of 2a is 5'-GGCAAGCAGGAGGACTTCAAGACGACAGTTCTGGAGGGTATGGAGACGGCCAAGCATCAG-3'

Primer C, which intends to identify if exon 7 of LSD 1 is skipped in neuroblastoma cell lines produces only one band in six neuroblastoma cell lines.

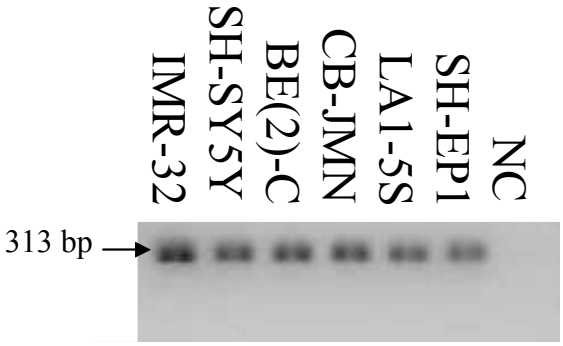

Figure 4: Primer C of LSD1 produces only one band whose size is 313 bp which is in accordance with the prediction when exon 7 is not skipped. All six cell lines produce the same product.

The sequence result of the DNA purified from RT-PCR product confirms the existence of exon 7 in mature RNA in six neuroblastoma cell lines.

```
Query 82                                                      141
CCCTGGCTTCCAAAAGTGTGACATCCATTCCAAAACTTTGTAACTGTCGAGCTGCTGCCA
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CCCTGGCTTCCAAAAGTGTGACATCCATTCCAAAACTTTGTAACTGTCGAGCTGCTGCCA
Subject 6206874                                             6206815
```

```
Query 142                                          189
AGCCTGAGACCCCAGAGCCTATAATAATTACCTTTCCTGTCTTTTTAG
|||||||||||||||||||||||||||||||||||||||||||||||
AGCCTGAGACCCCAGAGCCTATAATAATTACCTTTCCTGTCTTTTTAG
Subject 6206814                                    6206767
```

Figure 5: BLAST of sequencing result of product of primer C confirms that exon 7 of LSD1 is not skipped in splicing process in six neuroblastoma cell lines.

Discussion

Since there have been some reports on alternative splicing of LSD1 (Table 2) in other cell lines (Ohara O., *et al.* 1997), I intend to see if they apply to neuroblastoma cell lines and if there are other alternative splicing events in the six neuroblastoma cell lines studied in this project.

Variant Rep	Event Location	Event Detected
CN 341844.1	2<>2	Exon skipped
NM 001009999.1	2><3	Novel insertion
BE 538437.1	7<>7	Exon skipped

Table 2: Some of the reported alternative splicing events of LSD1

Primers are designed to see if exon 2 is skipped in RNA (primer A), if there is novel insertion between exon 2 and exon 3 (primer B) and if exon 7 is skipped (primer C). Meanwhile, RT-PCR is conducted for six cell lines under respective condition (details described in Material and Method) to see if there is differential expression among variants.

According to the size of RT-PCR products and their sequencing results, exon 2 is preserved in RNA of six neuroblastoma cell lines studied and so does exon 7. Meanwhile,

there is a novel insertion of 60 bases in RNA coming from intron 2. The sequence before this insertion is AG which is the same as the consensus sequence of acceptor site. The sequence after this insertion is GTGAGG while the consensus sequence of alternative splicing donor site is GTA/GAGT. Except for the last nucleotide, the six-base sequence after the insertion is the same as the consensus donor site sequence. This may explain why in some cases, the 60-base piece is preserved in mature RNA while in other cases, it is spliced (Figure 6). Besides, no apparent differential expression is observed between these two variants in the six cell lines.

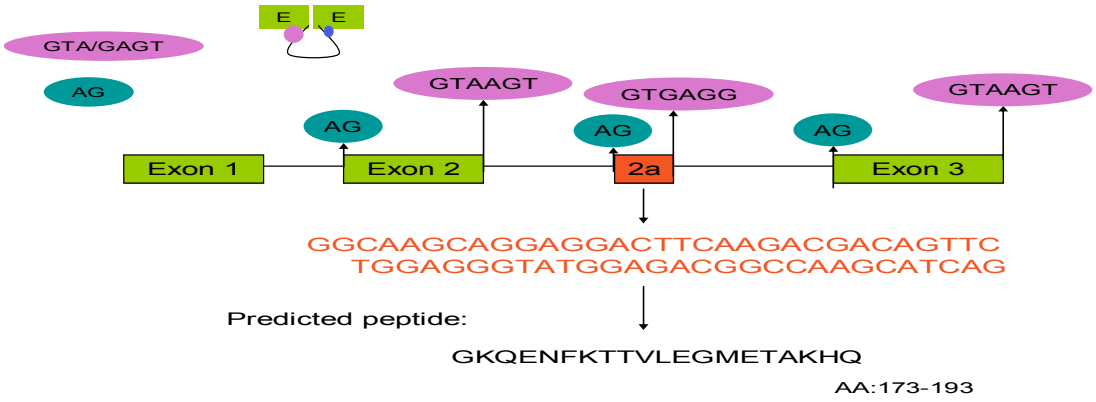

Figure 6: The pink sequence is the consent sequence of donor site. The dark green one is that of acceptor site. Comparison shows that the acceptor site of the insertion is the same as that of normal exon and consensus sequence but there are five nucleotides similarity and one nucleotide difference between the sequence of donor site and that of normal exon or consent sequence. This may be the reason why this insertion is occasionally included in final transcription product. The red sequence is the identified sequence of insertion which encodes the 20-amino acids locating from AA 173 to 193.

In order to see the possible influence of the 60 bases insertion, the protein encoded by this gene needs to be considered. There are two important domains in the protein encoded by LSD1, the SWIRM domain (pfam04433) which locates from amino acid 198

to 284 and the amine oxidase domain which locates at amino acid 544 to amino acid 845 (pfam01593). Since the coding region of LSD1 starts from nucleotide 145 of exon 1, the peptide encoded by the novel insertion locates from amino acid 173 to 193 which is right in front of the SWIRM domain. The SWIRM domain is a highly conserved domain present in multiple histone remodeling and modifying complexes (Guoping Da, Jeffery Lenkart, *et al.* 2005). It consists of a helical bundle containing a long central helix and two smaller helix-turn-helix motifs. It is important for the stability of LSD1 since a mutant lacking this domain produces an insoluble LSD1 protein in bacterial expression system. Meanwhile, the interaction between SWIRM domain and amine oxidase domain forms a spacious catalytic cavity for binding to both the lysine and neighboring residues of the targeted histone tail (Yong Chen *et al.* 2006). Any mutation that weakens the hydrophobic interaction in the interface between SWIRM domain and oxidase domain would greatly reduce the catalytic activity (Pete Stavropoulos, *et al.* 2006). According to the sequence of insertion, the coded peptide should be GKQENFKTTVLEGMETAKHQ. Although this sequence is not specific to any known domain or function, this 20-amino acid insertion may still have structural or functional influence on SWIRM domain as well as LSD1 in this variant. In order to see the influence, western blotting and other functional assays may be required.

Reference

Luger, K et al. (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* 389, 251-260

Lee, M.G., Wynder, C., Cooch, N., Shiekhattar, R. (2005). An essential role fro CoREST in nucleosomal histone 3 lysine 4 demethylation. *Nature* 437, 432-435

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

Eric Metzger, Melanie Wissmann, Na Yin *et al*. (2005) LSD1 demethylates repressive histone marks to promote androgen-receptor-dependent transcription. *Nature* 437, 436-439

Ronen Marmorstein, Raymond C. Trievel (2009) Histone modifying enzymes: Structures, mechnisms, and specificities. *Biochimica et Biophysica Acta (BBA)* 1789(1) 58-68

Jing Wang, Sarah Hevi, Julia K Kurash, *et al.* (2009) The lysine demethylase LSD1 (KDM1) is required for maintenance of alobal DNA methylation. *Nature Genetics* 41 (1), 125-129

Wang, G.G. *et al.* (2007) Chromatin remodeling and cancer, part I: covalent histone modifications. *Trends Mol. Med.* 13, 363-372

Kahl P, Gullotti L, Heukamp LC, *et al.* (2006) Androgen receptor coactivators lysine-specific histone demethylase 1 and four and a half LIM domain protein 2 predict risk of prostate cancer recurrence. *Cancer Res* 66: 11341-11347

Qingjun Zhu, Cheng Liu, Zheng Ge, *et al.* (2008) Lysine-specific demethylase 1 (LSD 1) is required for the transcriptional repression of the *telomerase reverse transcriptase (hTERT)* gene. PLoS ONE. 3(1), e1446-1447

Yong Chen, Yuting Yang, Feng Wang, Ke Wan, *et al.* (2006) Crystal structure of human histone lysing-specific demethylase 1 (LSD1). *PNAS* 103.38 13956-13961

Federico Forneris, Claudia Binda, Elena Battaglioli, Andrea Mattevi (2008) LSD1: oxidative chemistry for multifaceted functions in chromatin regulation. *Trends in Biochemical Sciences.* 33 (4) 181-189

Ohara, O., Nagase, T., Ishikawa, K. *et al.* (1997) Construction and characterization of human brain cDNA libraries suitable for analysis of cDNA clones encoding relatively large proteins. *DNA research* 4 (1), 53-59

Guoping Da, Jeffrey Lenkart *et al.* 2005 Structure and function of the SWIRM domain, a conserved protein module found in chromatin regulatory complex. PNAS 103 (7) 2057-2062

Pete Stavropoulos, Gunter Blobel & Andre Hoelz (2006) Crystal structure and mechanism of human lysine-specific demethylase-1.*Nature Structural & Molecular Biology*. 13 (7) 626-632