厚生労働科学研究費補助金（障害者対策総合研究事業）
分担研究報告書

下部神経管閉鎖障害症例に対するエキソームシークエンス解析研究

研究分担者　鬼頭浩史（名古屋大学医学部附属病院・整形外科講師）

研究要旨　下部神経管閉鎖障害に対する分子レベルでの病態解明のため、網羅的なエキソームシークエンス解析手法を用いて、原因遺伝子を同定することを試みる。症例のリクルートは着実に進んでおり、DNA サンプルによるエキソームシークエンス解析を実施している。

A．研究目的

下部神経管閉鎖障害は下肢の運動・知覚障害、膀胱直腸障害などを引き起こし、患者の移動能力および日常生活動作を著しく低下させる。また、神経障害に伴う下肢関節の変形は難治性である。本症における神経障害の発症機序はいまだ解明されていない。家族内発症例の報告もあり、遺伝的因子の関連も示唆されるが、ヒトにおける遺伝子変異の報告はVANGL1およびVANGL2のみである。本研究では、下部神経管閉鎖障害症例に対し、網羅的な エキソームシークエンス解析を行い、原因となる候補遺伝子を同定することを目的とする。

B．研究方法

名古屋大学医学部倫理委員会、および各分担研究者の施設における倫理委員会の承認を得た上で研究を遂行する。患者および両親の血液を採取してgenomic DNAを調整し、Agilent社SureSelect Human All Exon V4を用いてエキソームライブラリーを作成する。名古屋大学神経遺伝情報学において、SOLiD 5500xlにて塩基配列決定を行い、BioScope 1.3を用いてヒトゲノムへのmappingを行う。SNP解析ツールを用いて、de novo遺伝子変異解析を行い、候補遺伝子を同定する。

（倫理面への配慮）

下部神経管閉鎖障害症例に対する臨床解析および遺伝子解析は、「臨床研究に関する倫理指針」および「ヒトゲノム・遺伝子解析研究に関する倫理指針」に則り、各施設の生命倫委員会ならびに各施設長の承認を得て遂行する。

C．研究結果

東京大学附属病院および福岡こども病院から 8 家系の下部神経管閉鎖障害症例の血液サンプルを採取し、genomic DNA を調整した。現在、それらサンプルに対し、エクソームシークエンス解析を実施中である。

D．考察

近年、次世代シークエンサーによる網羅的な遺伝子配列解析が可能になり、本症を含む希少疾患に対して、ゲノムレベルからの新たな病態解明を進めることが期待されている。本症に対する根本的で有効な治療法は確立されておらず、分子レベルでの病態解明により、新たな治療戦略が導かれる可能性がある。

E．結論

8家系の下部神経管閉鎖障害症例に対し、エクソームシークエンス解析を実施している。

F．研究発表

1. 論文発表

Matsushita M, **Kitoh H**, Kaneko H, Mishima K, Itoh Y, Hattori T, Ishiguro N. Novel compound heterozygous mutations in the cathepsin K gene in Japanese female siblings with pyknodysostosis. Mol Syndromol 2:254-258, 2012

Kaneko H, **Kitoh H**, Wasa J, Nishida Y, Ishiguro N. Chondroblastoma of the femoral neck as a cause of hip synovitis. J Pediatr Orthop B 21:179-182, 2012

2. 学会発表

進行性骨化性線維異形成症に対するマレイン酸ペルヘキシリンの治療経験（第 23 回日本整形外科学会骨系統疾患研究会、2012.11、鬼頭浩史、金子浩史、他）

軟骨無形成症における低身長に対する治療薬の検討（第 23 回日本整形外科学会骨系統疾患研究会、2012.11、松下雅樹、鬼頭浩史、他）

G．知的財産権の出願・登録状況

なし

研究成果の刊行に関する一覧表

III. 研究成果の刊行に関する一覧表

書籍

著者氏名	論文タイトル名	書籍全体の編集者名	書　籍　名	出版社名	出版地	出版年	ページ
Engel AG, Shen X-M, Ohno K, Sine SM.	Congenital myasthenic syndromes	Engel AG	*Myasthenia gravis and myasthenic disorders 2nd ed*	Oxford University Press	New York	2012	173-230
Ohno K, Ito M, and Engel AG.	Congenital Myasthenic Syndromes – Molecular Bases of Congenital Defects of Proteins at the Neuromuscular Junction–		*Myopathy*	InTech	Rijeka	2012	175-200 (査読有)
Ohno K, Ito M, Ichihara M, and Ito M.	Molecular Hydrogen as an Emerging Therapeutic Medical Gas for Neurodegenerative and Other Diseases	Pereira MD	*Oxid Med Cell Longev*	Hindawi Publishing Corp.	New York	2012	vol.2012 ID 353152, 11 pages (査読有)
Ohno K, Ito M, Kawakami Y, Krejci E, Engel AG.	Specific binding of collagen q to the neuromuscular junction is exploited to cure congenital myasthenia and to explore bases of myasthenia gravis	*Chem Biol Interact*	*Elsevier*	Amsterdam	2012	2012	203: pp 335-340 (査読有)
大野欽司	RNA 異常と神経疾患		*Annual Review 神経2012*			2012	97-103
大野欽司	先天性筋無力症候群		*Clinical Neuroscience*			2012	30 (6): 685-687
Ohe K, Masuda A, Ohno K.	Intronic and exonic nucleotide variations that affect rna splicing in humans		*Introduction to Sequence and Genome Analysis*	iConcept Press	Hong Kong	2012	in press

雑誌

発表者氏名	論文タイトル名	発表誌名	巻号	ページ	出版年
Masuda A, Andersen HS, Doktor TK, Okamoto T, Ito M, Andresen BS, Ohno K.	CUGBP1 and MBNL1 preferentially bind to 3' UTRs and facilitate mRNA decay	*Scientific Reports*	2	209	2012
Yamashita Y*, Matsuura T*, Shinmi J, Amakusa Y, Masuda A, Ito M, Kinoshita M, Furuya H, Abe K, Ibi T, Sahashi K, Ohno K. *Equal contribution	Four parameters increase the sensitivity and specificity of the exon array analysis and disclose twenty-five novel aberrantly spliced exons in myotonic dystrophy	*J Hum Genet*	57	368-374	2012
Ito M, Suzuki Y, Okada T, Fukudome T, Yoshimura T, Masuda A, Takeda S, Krejci E, Ohno K.	Protein-anchoring strategy for delivering acetylcholinesterase to the neuromuscular junction	*Mol Ther*	20	1384-1392	2012
Matsuura T, Minami N, Arahata H, Ohno K, Abe K, Hayashi YK, Nishino I.	Myotonic dystrophy type 2 (DM2) is rare in the Japanese population	*J Hum Genet*	57	219-220	2012
Ishigaki S, Masuda A, Fujioka Y, Iguchi Y, Katsuno M, Shibata A, Urano F, Sobue G, Ohno K.	Position-dependent fus-rna interactions regulate alternative splicing events and transcriptions	*Sci Rep*	2	529	2012
Kurosaki T, Ueda S, Ishida T, Abe K, Ohno K, Matsuura T.	The unstable CCTG repeat responsible for myotonic dystrophy type 2 originates from an *Alu*Sx element insertion into an early primate genome	*PLoS ONE*	7	e38379	2012
Ito M, Hirayama M, Yamai K, Goto S, Ichihara M, Ohno K.	Drinking hydrogen water and intermittent hydrogen gas exposure, but not lactulose or continuous hydrogen gas exposure, prevent 6-hydorxydopamine-induced Parkinson's disease in rats	*Med Gas Res*	2	15	2012

Yoshinaga H, Sakoda S, Good J M, Takahashi M P, Kubota T, Arikawa-Hirasawa E, Nakata T, Ohno K, Kitamura T, Kobayashi K, Ohtsuka Y.	A novel mutation in SCN4A causes severe myotonia and school-age-onset paralytic episodes	*J Neurol Sci*	315	15-19	2012
Weng, Q., Chen, Y., Wang, H., Xu, X., Yang, B., He, Q., Shou, W., Chen, Y., Higashi, Y., van den Berghe, V., Seuntjens, E., Kernie, S.G., Bukshpun, P., Sherr, E.H., Huylebroeck, D. and Lu, Q.L.	Dual-Mode Modulation of Smad Signaling by Smad-Interacting Protein Sip1 Is Required for Myelination in the Central Nervous System	*Neuron*	73	713-728	2012
Suzuki K, Yokoyama C, Higashi Y, Daikoku T, Mizoguchi S, Saika S andYamada G.	Wakayama symposium: epithelial-mesenchymal interaction regulates tissue formation and characteristics: insights for corneal development	*Ocul Surf*	10	217-220	2012
芳賀信彦	小児運動器疾患のリハビリテーション：現状と展望	*東北医学雑誌*	124（1）	38-40	2012
Sayeed S, Asano E, Ito S, Ohno K, Hamaguchi M, Senga T.	S100a10 is required for the organization of actin stress fibers and promotion of cell spreading	*Mol Cel Biochem*	374	105-111	2013
Yamamoto R, Matsushita M, Kitoh H, Masuda A, Ito M, Katagiri T, Kawai T, Ishiguro N, Ohno K.	Clinically applicable antianginal agents suppress osteoblastic transformation of myogenic cells and heterotopic ossifications in mice	*J Bone Miner Metab*	31	26-33	2013
Iio A, Ito M, Itoh T, Terazawa R, Fujita Y, Nozawa Y, Ohsawa I, Ohno K, Ito M.	Molecular hydrogen attenuates fatty acid uptake and lipid accumulation through downregulating cd36 expression in hepg2 cells	*Med Gas Res*	3	6	2013

Tanisawa K, Mikami E, Fuku N, Honda Y, Honda S, Ohsawa I, Ito M, Endo S, Ihara K, Ohno K, Kishimoto Y, Ishigami A, Maruyama N, Sawabe M, Iseki H, Okazaki Y, Hasegawa-Ishii S, Takei S, Shimada A, Hosokawa M, Mori M, Higuchi K, Takeda T, Higuchi M, Tanaka M.	Exome sequencing of senescence-accelerated mice (sam) reveals deleterious mutations in degenerative disease-causing genes	*BMC Genomics*	14	248	2013
McKinsey GL, Lindtner S, Trzcinski B, Visel A, Pennacchio LA, Huylebroeck D, Higashi Y. and Rubenstein JL.	Dlx1&2-dependent expression of Zfhx1b (Sip1, Zeb2) regulates the fate switch between cortical and striatal interneurons	*Neuron*	77	83-98	2013
芳賀信彦、田中信幸、田中弘志	先天性無痛無汗症の治療戦略	*日整会誌*	87	57-60	2013
Nakata T, Ito M, Azuma Y, Otsuka K, Noguchi Y, Komaki H, Okumura A, Shiraishi K, Masuda A, Natsume J, Kojima S, Ohno K.	Mutations in the c-terminal domain of colq in endplate acetylcholinesterase deficiency compromise colq-musk interaction	*Hum Mutat*			in press
Zhang Y, Ogata N, Yozu A, Haga N	Two-dimensional video gait analyses in patients with congenital insensitivity to pain	*Dev Neurorehabil*			in press

研究成果の刊行物・別刷

Intronic and exonic nucleotide variations that affect RNA splicing in humans

Kenji Ohe, Akio Masuda, Kinji Ohno*
Division of Neurogenetics, Center for Neurological Diseases and Cancer, Nagoya University Graduate School of Medicine, Nagoya, Japan

1 Introduction

Identification of the 3.2 billion nucleotides of the human genome, the human genome project, and the gene annotation projects thereafter have revealed that human has merely 22,000-25,000 protein coding genes. Comparison with other genomes has highlighted the significance of a higher degree of alternative splicing in human to produce the vast proteome compared to other species. During the millions of years of evolution, our genome has been extensively mutated and exposed to natural selection. Ten percent of mutations, which cause hereditary disease, disrupt the splice site consensus sequence at the 5' splice site and 3' splice site (Krawczak et al., 2007). Indeed, 11,525 splicing mutations out of 123,656 total mutations are registered in the Human Genome Mutation Database (HGMD) professional release 2012.1 (http://www.hgmd.org/). Splicing, however, does not only rely on the splice site consensus sequences. Considering auxiliary splicing *cis*-elements along with the recent advances in resequencing technologies of the human genome, the rate of disease-causing mutations affecting splicing has been estimated to be up to 60% (Hammond&Wood, 2011). Studying the mechanisms of splicing disruption will elucidate the molecular bases of development of disease and also give us essential information on how and why the human genome has evolved.

Mechanisms of splicing disruption can be categorized into three groups: a) disruption of splicing *cis*-element; b) aberrant expression or abnormal function of splicing *trans*-factors; and c) RNA gain-of-function disorders, in which a splicing *trans*-factor is sequestered to abnormally expanded RNA repeats and forms an abnormal aggregate. Abnormal expression of splicing *trans*-factors is observed in various types of cancer, in which SRSF1 (ASF/SF2) is an oncogenic product (Karni et al., 2007). Another example of an abnormally regulated *trans*-factor is the oncogenic product HMGA1a (HMG I/Y), which binds upstream of the 5' splice site and traps U1 snRNP at the 5' splice site (Ohe&Mayeda, 2010) and results in aberrant alternative splicing of *PSEN2* encoding presenilin-2 in sporadic Alzheimer's disease (Manabe et al., 2003). RNA gain-of-function mechanisms are observed in neurological disorders including myotonic dystrophy, spinocerebellar ataxia, and paraneoplastic syndromes (Licatalosi&Darnell, 2006; O'Rourke&Swanson, 2009). This chapter is dedicated to disruption of intronic and exonic splicing *cis*-elements. Please refer to other review articles for other abnormal splicing mechanisms (Licatalosi&Darnell, 2006; O'Rourke&Swanson, 2009; Ohno K, 2011).

*Corresponding Author, E-mail: ohnok@med.nagoya-u.ac.jp

Splicing occurs as a two-step *trans*-esterification reaction after the step-wise assembly of uridine-rich small nuclear ribonucleoproteins (U snRNPs) of the spliceosome (Smith et al., 2008; Wahl et al., 2009) (Figure 1A). Though the average length of human introns is more than 3,300 nucleotides (Lander et al., 2001), variation in length is immense; from a 43-nucleotide intron in *ESRP2* (epithelial splicing regulatory protein 2) (Sasaki-Haraguchi et al., 2012) to a 740,920-nucleotide intron found in the heparan sulfate 6-O-sulfotransferase 3 gene (Fedorova&Fedorov, 2005). Thus, in higher eukaryotes, splicing is mostly defined by exon rather than intron (Figure 1B) (Berget, 1995). Using this exon definition model, a number of algorithms have been developed to give a probable answer on the consequence of a single nucleotide variation (SNV). No algorithm, however, is 100% accurate. Splicing consequences should be tested with a combination of different algorithms and should be verified by splicing substrates *in vitro* and/or *in vivo*.

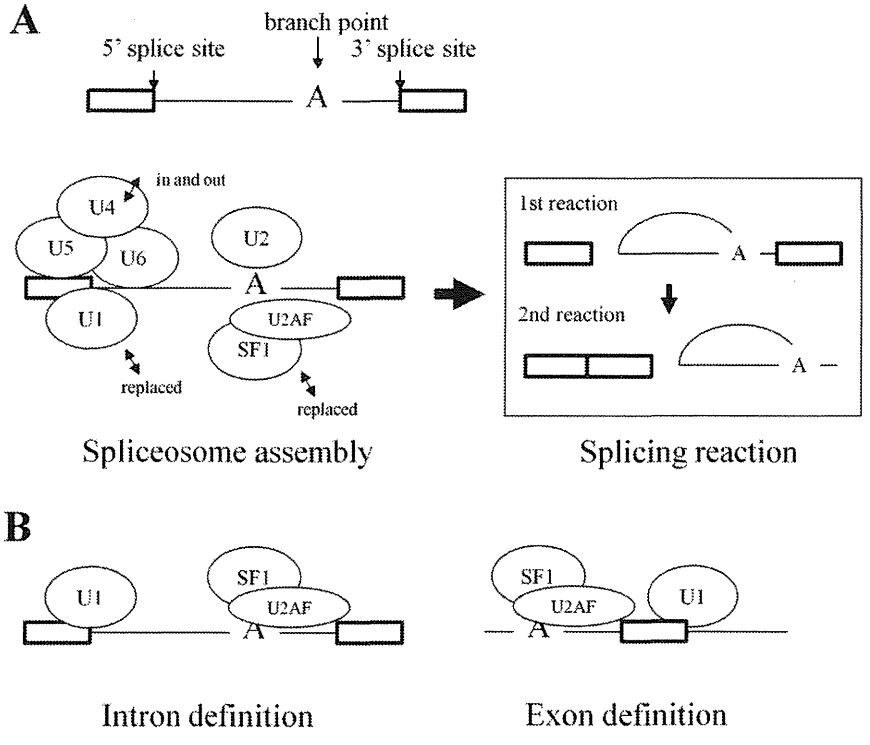

Figure 1: A, Assembly of the spliceosome by U snRNPs and associated factors (left) and the two-step *trans*-esterification reaction after the assembly of the spliceosome (right). **B,** The intron definition model for lower eukaryotes and exon definition model for higher eukaryotes. U1-6, U1-6 snRNP; SF1, splicing factor 1; U2AF, U2 snRNP auxiliary factor.

2 Intronic nucleotide variations

2.1 Intronic splicing *cis*-elements

In eukaryotes, pre-mRNA splicing is conducted by essential nucleotide sequences on the splicing substrate (*cis*-elements) and *trans*-factors that bind to these sequences. Core *cis*-elements involve the 5' splice site,

*Corresponding Author, E-mail: ohnok@med.nagoya-u.ac.jp

the branch point sequence (BPS), polypyrimidine tract (PPT) and the 3' splice site, to which the *trans*-factors, U1 small nuclear ribonucleoprotein (U1 snRNP), SF1/mBBP and later U2 small nuclear ribonucleoprotein (U2 snRNP), U2 snRNP auxiliary factor 65 (U2AF65), and U2 snRNP auxiliary factor 35 (U2AF35) associate, respectively (Figure 1A). Besides these classical *cis*-elements, a growing number of intronic splicing enhancer/silencer (ISE, ISS) are known to function in alternative splicing, which, however, are difficult to predict from the nucleotide sequence alone. The tissue- and development-specific expression of these *trans*-factors that recognize these *cis*-elements makes the fine-tuned alternative splicing possible.

2.1.1 5' splice site

The consensus sequence of the 5' splice site is "CAG/GUAAGUAU", where "/" indicates an exon-intron boundary (Lerner et al., 1980; Roca&Krainer, 2009). Several computational methods have been developed to evaluate the 5' splice site including the Shapiro and Senapathy matrix (Shapiro&Senapathy, 1987), the weight matrix model (Staden, 1984), the first-order Markov model (Salzberg, 1997), the maximum dependence decomposition model (Burge&Karlin, 1997), and the maximum entropy model (Yeo&Burge, 2004). Many reports have been made on aberrant splicing due to disease-associated SNVs at the 5' splice site. The computational methods above are optimized for evaluating the strength of U1 snRNP-binding to the 5' splice site, but are not designed for evaluating the effect of an SNV on aberrant splicing. Algorithms to predict an activated cryptic 5' splice site due to an SNV at the 5' splice site have been created: an EST-based method called Cryptic Splice Finder (Kapustin et al., 2011), which uses the information content of nucleotide frequencies at each 5' splice site (Rogan et al., 1998; Nalla&Rogan, 2005). Calculation of free energy for stable binding of U1 snRNA to the 5' splice site has also been proposed (ΔG) (Roca et al., 2005). More recently, a structure-based method, the structure profiles and odds measure (SPO algorithm), was developed for identifying cryptic 5' splice sites (Tsai&Wang, 2012). In order to predict a splicing consequence of an SNV at the 5' splice site, we have also analyzed splicing patterns of 31 mini-genes carrying naturally occurring or artificially introduced mutations at various nucleotides of the 5' splice site using a cell culture system (Sahashi et al., 2007). We found that a new parameter, the SD-score, efficiently predicts the splicing consequences of SNVs at the 5' splice site. SD-score is a common logarithm of the frequency in the human genome of a specific 9-nucleotide sequence spanning 3 nucleotides at the 3' end of an exon and 6 nucleotides at the 5' end of an intron. The SD-score predicted the splicing consequences of the 31 mini-genes and 179 reported SNVs at the 5' splice site with the sensitivity of 97.1 % and specificity of 94.7 %. We simulated all possible SNVs at 189,249 5' splice sites in the human genome using this SD-score, and found that 37.8%, 88.8%, 96.8%, 82.8%, 95.0%, 96.5%, and 63.0% of SNVs at exonic positions -3, -2, -1, and intronic position +3, +4, +5, +6, respectively, were predicted to cause aberrant splicing, which were all higher than we expected. Using the SD-score we indeed detected four mutations at the 5' splice site causing aberrant splicing, which have been overlooked. The SD-score web service program is available at http://www.med.nagoya-u.ac.jp/neurogenetics/SD_Score/sd_score.html.

A problem shared by these computational methods is that they rely on prediction of binding of U1 snRNP to the 5' splice site. The problem emerges at the non-canonical 5' splice sites: 0.9% of 5' splice sites carry GC dinucleotide and 0.36% harbor U12-type 5' splice sites with AT dinucleotide (Sheth et al., 2006). Also it has been reported that base pairing is shifted by one nucleotide between the 5' splice site and U1 snRNA at 59 5' splice sites (Roca&Krainer, 2009). A subsequent report shows that as much as ~5% of 5' splice sites are recognized by U1 snRNA by bulged nucleotides on either the 5' splice site or U1 snRNA (Roca et al., 2012).

*Corresponding Author, E-mail: ohnok@med.nagoya-u.ac.jp

2.1.2 Human branch point sequence (BPS)

Yeast carries a strictly conserved BPS of "UACUAAC" (Langford&Gallwitz, 1983). In eukaryotes, however, the BPS had not been evaluated through global experimental verification. Previously reported consensus sequences of eukaryote BPS were either based on analyses of only a few experimentally identified branch points or *in silico* search for sequences which resemble the yeast consensus BPS.

As depletion of SF1 does not abrogate splicing in HeLa nuclear extract, SF1 may be dispensable in most pre-mRNAs (Tanackovic&Kramer, 2005). An SF-1 affinity model, in which validated SF1 affinity data is employed, efficiently predicts BPS in putative SF1-dependent introns (Pastuszak et al., 2011). A computational algorithm for branch point prediction, the SVM-BP finder (http://regulatorygenomics.upf.edu/Software/SVM_BP/), has also been developed (Corvelo et al., 2010). Recently, large-scale sequencing of lariat RT-PCR products determined individual branchpoints in human pre-mRNAs and also identified 66 disrupted branchpoints associated with human diseases (Taggart et al., 2012).

Development of lariat RT-PCR has markedly increased efficiency of identifying BPSs (Vogel et al., 1997). We conducted lariat RT-PCR of 52 introns of 20 housekeeping genes. Analysis of 367 lariat RT-PCR clones disclosed that the human consensus BPS is "yUnAy" (y, pyrimidine; n, any nucleotide) (Gao et al., 2008). The fourth nucleotide "A" is the branchpoint (position +0) and is conserved in 92.3% of the clones. The "U" at position -2 is conserved in 74.6%. Collation of 46 experimentally verified BPSs that have been previously published also resulted in a "yUnAy" sequence.

2.1.3 3' splice site

The 3' splice site consists of 3' end of an intron and 5' end of an exon, having a consensus sequence of "NYAG/R" (N, any nucleotide; Y, pyrimidine; R, purine; "/", intron-exon boundary) (Zhang, 1998). U2AF35 binds to this sequence at early steps of the splicing reaction. At later steps, the dinucleotide "AG" is scanned downstream of the branchpoint and the first "AG" is recognized as the 3' end of the intron, "AG scanning model" (Smith et al., 1989). SNVs of this kind have been reported; an 18-nucleotide duplication comprising 16 intronic and 2 exonic residues of the *HEXB* gene encoding the β subunit of β-hexosaminidase results in an active upstream copy of the 3' splice sites (Dlott et al., 1990). A 69-nucleotide duplication comprising 7 intronic and 62 exonic residues of *SLC4A1* encoding anion exchanger member 1 also results in an active upstream copy of the 3' splice sites (Bianchi et al., 1997). We have characterized a mutation of this kind in a patient with congenital myasthenic syndrome. The mutation was a duplicated sequence of 16 nucleotides at the junction of intron 10 and exon 11 of the acetylcholine receptor ε subunit gene (*CHRNE*) (Ohno et al., 2005). The upstream of the duplicated "AG" was always recognized and generated aberrant protein.

On the contrary, a mutation of the *BTD* gene encoding biotinidase creates a downstream 3' splice site with a stronger polypyrimidine tract which results in usage of the downstream AG, not following the AG scanning model (Pomponio et al., 1997). Analyses of the β-globin gene in β-thalassemia patients (Spritz et al., 1981; Westaway&Williamson, 1981; Metherall et al., 1986) and deletion mutagenesis analyses of the rabbit β-globin intron 2 have disclosed the following rule. When the upstream cryptic 3' splice site is more than 24 nucleotides away from the authentic 3' splice site, the authentic 3' splice site is used; less than 21 nucleotides, the cryptic 3' splice site is used; between 21 and 24 nucleotides, both cryptic and authentic 3' splice sites are used (Chua&Reed, 2001; Wieringa et al., 1984). A chromatin- and RNA-associated protein, DEK, facilitates the U2AF35-AG interaction and prevents binding of U2AF65 to pyrimidine tracts not followed by AG (Soares et al., 2006). As described later, hnRNPA1 also

*Corresponding Author, E-mail: ohnok@med.nagoya-u.ac.jp

discriminates authentic 3' splice sites from pyrimidine-rich RNA segments by recognizing AG at the 3' end of an intron (Tavanez et al., 2012).

Extensive mutational analyses of nucleotides surrounding the 3' splice site of *FAS* revealed that, though U2AF35-binding sequences are highly degenerative, the auxiliary nucleotides are essential for binding of U2AF35 (Corrionero et al., 2011).

Recent analysis of large-scale mapping of the human branchpoints has revealed 80% of the branchpoints follow the AG scanning model. A decision tree optimized by the ID3 algorithm was developed for AG selection to an accuracy of 95.5%. This decision tree includes four steps: 1) whether the AG of interest is the first one after the branchpoint or not, 2) whether the first AG is more than 7 nucleotides downstream the branchpoint or not, 3) whether the distance between the first and next AG is more than 14 nucleotides or not, 4) if the AG decided is within an optimal range from the branchpoint (16-25 nucleotides) (Taggart et al., 2012).

2.1.4 Intronic splicing enhancers (ISE) / intronic splicing silencers (ISS)

In human, prediction of disruption of intronic *cis*-elements would be more difficult than exonic *cis*-elements due to the fact that introns are much longer than exons. Many examples of aberrant activation of pseudo-exons due to mutations have been documented (Dhir&Buratti, 2010). Most of these pseudo-exons originate from transposable elements that make up half of the human genome (Vorechovsky, 2010). They are likely to be suppressed by nucleotide substitutions acquired in the course of evolution (Schmitz&Brosius, 2011), and are reactivated by a disease-causing mutation. Currently available exon scanning tools cannot fully predict aberrant activation due to these mutations. We thus have listed the position in gene and resulting disease in Table 1, with the hope of assisting development of an accurate algorithm which predicts exonization due to an intronic SNV and for aid of the development of the recent antisense therapies which correct expression of these pseudo-exons (Wood et al., 2007) especially in neuromuscular disorders (Wood et al., 2010).

The average length of these 124 pseudo-exons identified in 61 diseases is 109 nucleotides, which is slightly shorter than the average size of human exons of 126 nucleotides (Schwartz et al., 2009). Recently disclosed link of histone modifications (Kolasinska-Zwierz et al., 2009) and nucleosome (Beckmann&Trifonov, 1991; Tilgner et al., 2009) to alternative splicing has led to speculation that 120 bp of DNA wrapping a subnucleosome structure, the $(H3\text{-}H4)_2$ histone tetramer, is protected from nucleotide substitutions and marks the exon for the splicing machinery in higher eukaryotes (Keren et al., 2010). The growing list of aberrantly activated pseudo-exons predicts that intronic mutations causing abnormal splicing are more involved in disease mechanisms than have been previously expected. Indeed, the frequency of these types of mutations has been calculated up to 7% in some diseases (Pros et al., 2009; Gurvich et al., 2008).

We have also identified a G-to-A mutation at 8 nucleotides upstream the 3' end of intron 3 (IVS3-8G>A) of the *CHRNA1* gene. In this patient, a missense mutation that decreases the expression of acetylcholine receptor was also identified on one allele, and the IVS3-8G>A mutation was found on the other. We thus checked splicing products in the patient's biopsied muscle and found that an alternative exon P3A downstream of exon 3 is exclusively activated by IVS3-8G>A. As the mutation does not disrupt any known splicing *cis*-element, we sought for a responsible splicing *trans*-factor and identified that hnRNP H binds to this *cis*-element. Biacore surface plasmon resonance (SPR) analysis revealed that the mutation decreases the binding affinity for hnRNP H to 1/100 (Masuda et al., 2008). We also determined polypyrimidine tract binding protein (PTB) as a silencing *trans*-factor for P3A and that tannic acid increases expression of PTB to correct abnormal activation of P3A (Bian et al., 2009).

*Corresponding Author, E-mail: ohnok@med.nagoya-u.ac.jp

Disease	Gene name	Size of pseudo-exon and localization
Afibrinogenemia	*FGB*	50 bp in intron 1
	FGG	75 bp in intron 6
Alport syndrome	*COL4A3*	74 bp in intron 5
	COL4A5	30 bp in intron 29
		147 bp in intron 6
Amyotrophic lateral sclerosis	*SOD1*	43 bp in intron 4
Ataxia telangiectasia	*ATM*	58 bp in intron 19
		65 bp in intron 20
		112bp in intron 28
		137 bp in intron 40
Autosomal dominant hearing loss	MYO6	108 bp in intron 23
Autosomal dominant retinitis pigmentosa	PRPF31	175 bp in intron 13
Autosomal recessive polycystic kidney disease	PKHD1	116 bp in intron 46
β+ thalassemia	*HBB*	165bp in intron 2
		125 bp in intron 2
		73 bp in intron 2
Breast cancer	*BRCA1*	66 bp in intron 13
	BRCA2	65 bp in intron 16
	ESR1	69 bp in intron 5
Brooke-Spiegler syndrome	*CYLD*	65 bp in intron 9
Choroideremia	*CHM*	98 bp in intron 5
Chronic granulomatous disease	*CYBB*	56 bp in intron 6
		61 bp and 110 bp in intron 5
Congenital cataracts facial dysmorphism neuropathy syndrome	*CTDP1*	95 bp in intron 6
Congenital disorders of glycosylation type Ia	*PMM2*	66 bp in intron 7
		123 bp in intron 7
Chronic myeloid leukemia	*ABL*	35 bp in intron8
	BCR-ABL	42 bp of *ABL* intron 1b
Coagulation factor V deficiency	*F5*	295 bp in intron 8
Cystic fibrosis	*CFTR*	183 bp in intron 3
		101 bp in intron 6
		104 bp in intron 10
		49 bp in intron 11
		214 bp in intron 18
		84 bp in intron 19
DHPR deficiency	*DHPR*	152 bp in intron 3
Duchenne muscular dystrophy	*DMD*	149 bp in intron 1M (Becker type)
		132 and 46 bp in intron 2
		70 bp in intron 4

*Corresponding Author, E-mail: ohnok@med.nagoya-u.ac.jp

		90 bp in intron 9
		(Becker type)
		in intron 11
		79 bp in intron 11
		(Becker type)
		95 bp in intron 25
		172 and 202 bp in intron
		25
		119 bp in intron 27
		137 bp in intron45
		72 bp in intron 47
		125 bp in intron 48
		108 bp in intron 48
		98 bp in intron 49
		180 bp in intron 49
		160 bp in intron 49
		149 bp in intron 49
		134 bp in intron 56
		89 bp in intron 60
		(intermediate type)
		67 bp in intron 62
		58 bp in intron 62
		67 bp in intron 63
		147 bp in intron 65
		53 bp in intron 65
		121 bp in intron 67
Fabry disease	*GLA*	57 bp in intron 4
Familial hemophagocytic lymphohistiocytosis	*UNC13D*	130 bp in intron 1
Fibrochondrogenesis	*COL11A1*	50 bp in intron 48
Fukutin related congenital muscular dystrophy	*FKTN*	64 bp in intron 5
Gitelman's syndrome	*SLC12A3*	238 bp in intron 13
		90 bp in intron 21
Globoid-cell leukodystrophy	*GALC*	34 bp in intron 6
Glutathione synthetase deficiency	*GSS*	In exon 2
Growth-hormone insensitivity (Laron syndrome)	*GHR*	108 bp in intron 6
Haemophilia A	*FⅧ*	191 bp in intron 1
Holoprosencephaly	*SHH*	129 bp in intron 1
Homocystinuria	*MTRR*	140 bp in intron 6
17α-Hydroxylase Deficiency	*CYP17A1*	95 bp in intron 1
Hyper immunoglobulin M (IgM) syndrome	*CD40L*	59 bp in intron 3
Leber congenital amaurosis	*CEP290*	128 bp in intron 26
Leigh syndrome	*NDUFS7*	122 bp in intron 1
Lynch syndrome	*MSH2*	75 bp in intron 1
Malignant rhabdoid tumor	*SNF5/INI1*	72 bp in intron 1

*Corresponding Author, E-mail: ohnok@med.nagoya-u.ac.jp

Maple syrup urine disease	*E2*	126 bp in intron 8
Marfan syndrome	*FBN1*	93 bp in intron 63
Megalencephalic leukoencephalopathy with subcortical cysts	*MLC1*	246 bp in intron 10
3-methylcrotonyl-CoA carboxylase deficiency	*MCBB*	64 bp in intron 10 also deletes exon 11
Methylmalonic acidemia	*MUT*	76 bp in intron 11
Mitochondrial trifunctional protein deficiency	*HADHB*	106 and 56 bp in intron7
Mucopolysaccharidosis Type II (Hunter syndrome)	*IDS*	78 bp in intron 7 103 bp in intron 3
Mucopolysaccharidosis type VII (Sly syndrome)	*GUSB*	68 bp in intron 8
Multiminicore Disease	*RYR1*	119 bp in intron 101
Myopathy with lactic acidosis	*ISCU*	53 bp in intron 4 86 and 100 bp in intron 5
Neurofibromatosis type 1	*NF-1*	107 bp in intron 3 58 bp in intron 6 76 and 54 bp in intron 10 177 bp in intron 30 172 bp in intron 30 70 bp in intron 45
Neurofibromatosis type 2	*NF-2*	106 bp in intron 5
Niemann–Pick type C disease.	*NPC1*	374 bp in intron 1 108 bp in intron 6 194 bp in intron 9
Ocular albinism type 1	*OA1*	165 bp in intron7
ornithine δ-aminotransferase deficiency	*OAT*	142 bp in intron 3
Propionic acidemia	*PCCA*	84 bp in intron 14
	PCCB	72 bp in intron 6
Primary ciliary dyskinesia (Kartagener syndrome)	*CCDC39*	116 bp in intron 9
Retinoblastoma	*RB1*	103 bp in intron 23
Schwartz-Jampel syndrome	*HSPG2*	130 bp in intron 6
Tetrahydrobiopterin deficiency	*PTS*	45 bp in intron 2 79 bp in intron 1
Tuberous sclerosis complex	*TSC2*	89 bp in intron 8
Urea transport deficiency	*JK*	136 bp in intron3 also deletes exon4,5
Usher syndrome type 2	*USH2A*	152 bp in intron 40
X-linked hypophosphatemia	*PHEX*	50, 100, 170 bp in intron 7 59 bp in intron 19
Werner syndrome	*WRN*	69 bp in intron 25

Table 1: Disease-associated mutations causing aberrant activation of pseudo-exon

*Corresponding Author, E-mail: ohnok@med.nagoya-u.ac.jp

3 Exonic nucleotide variations

3.1 Exonic splicing *cis*-elements

3.1.1 First nucleotide of exon

As mentioned above, the consensus 3' splice site is "YAG/G" (Y, pyrimidine). Many SNVs disrupting the AG are known, but few have been reported that disrupt the first nucleotide G in the exon.

We have analyzed the effect of mutations at the first nucleotide in the exon on splicing regulation (Fu et al., 2011). When the length of the pyrimidine stretch of the polypyrimidine tract is longer than 10-15 nucleotides, U2AF65 binds so tight that U2AF35 binding to the YAG/G is dispensable, which is known as the AG-independent splice site. When the pyrimidine stretch is short, U2AF35 needs to bind to the YAG/G for efficient splicing, which is known as the AG-dependent splice site. We have found that when a mutation of the first nucleotide of exon is at an AG-dependent splice site, aberrant splicing occurs, but when a mutation is at an AG-independent splice site, normal splicing takes place. Analyses of 1,796,809 3' splice sites in human revealed that when the first nucleotide of the exon is other than G, the average size of pyrimidine stretches is significantly shorter than that when it is a G. This means that when the 3' splice site is AG-dependent, the probabilities of a short pyrimidine stretch and the first nucleotide of the exon to be a G are higher. We have reported 5 mutations at the first nucleotide of an exon that cause aberrant splicing. The importance of the first nucleotide of an exon is also reported in *FAS* encoding *CD95* gene, in which U2AF35 strictly binds to RCAG/G (R, purine), and is sensitive to a single mutation even at -4, -3, and +1 (Corrionero et al., 2011). Mutations at the first nucleotide of an exon may have been underestimated and attention must be evoked in the field of human genetics.

A total of 20 mutations at the first nucleotide of an exon have been reported: 10 mutations cause aberrant splicing (AS), whereas 10 result in normal splicing (NS) (Table 2). According to our studies (Fu et al., 2011), we can assume that the AS and NS mutations are at the AG-dependent and AG-independent sites, respectively. As expected, all the AS mutations affect a G at the first nucleotide of an exon. We constructed sequence logos using a tool available on-line (Crooks et al., 2004) (http://weblogo.berkeley.edu/logo.cgi) using the 10 AS (Figure 2A) and 10 NS (Figure 2B) mutants. Figure 2A shows that in the 10 AS sites a C at intron -3 is rarely observed. As the consensus sequence for U2AF35 is YAG/G, lack of C at intron -3 makes binding of U2AF35 solely dependent on G at exon +1 in the AS sites. A mutation affecting G at exon +1 thus compromises binding of U2AF35, leading to aberrant splicing. The lengths of uninterrupted stretch of PPT (underlined in Table 2) of the AS and NS sites are 7.4 ± 2.9 and 9.3 ± 4.2 (mean and SD), respectively. Although the difference is not statistically significant, the difference in PPT also suggests that the AS sites are AG-dependent.

Valcarcel and colleagues recently report that U2AF heterodimer requires hnRNPA1 to discriminate *bona fide* 3' splice sites from pyrimidine-rich RNA segments (Tavanez et al., 2012). hnRNP A1 recognizes AG at the 3' end of an intron and facilitates binding of U2AF heterodimer to *bona fide* PPT. As the consensus sequence of hnRNP A1 is "UAGGGW" (W, A or U) (Burd&Dreyfuss, 1994), a G at the first nucleotide of an exon may also be essential for binding to hnRNP A1.

In the course of analysis of mutations at the first nucleotide of an exon, we also found that T is preferentially used at exonic position +3, +4, and +5 in the human genome (Fu et al., 2011). SELEX analysis of U2AF35 also shows that U2AF35 binds up to 12 nucleotides downstream the intron-exon boundary into the exon (Wu et al., 1999). Indeed, preferential usage of T at exonic position +3, +4, +5 is

*Corresponding Author, E-mail: ohnok@med.nagoya-u.ac.jp

observed in the SELEX dataset. Splicing mutations affecting these thymine nucleotides, however, have not been reported to date.

Mutation at first nucleotide of exon	Gene name /exon	Consequence
Mutations affecting splicing (AG-dependent 3' splice sites to which U2AF35 binds)		
atttaAtttctctgtttttccccttttgtag (G>T)AAGTCACCAAA	*CFTR* / exon 4	Exon skipping, activation of cryptic exon
cacttAcgcattcgtctatctttcgcatag (G>T)TTGGTCCAATG	*FECH* / exon 9	Exon skipping
gactgAtcttgttttctgttccccacag (G>T)GGGTTCCAGCC	*UROS* / exon 9	Exon skipping
gtagAccttggtgggcggtccttctcctag (G>T)AAGAAGCCTAT	*GH1* / exon 3	Exon skipping
tcttcAcctgtcatattcttattttag (G>T)ATCCACCCACT	*EYA1* / exon 12 (previously exon10)	Exon skipping
ccatgAcagaatttaccagaaatgtaaccatctcag (G>T)TCTCTGATGAA	*PKHD1* / exon 25	Exon skipping, activation of cryptic 3' splice site
ctgtaActgtttatttccaacag (G>T) GTGCTGCTGGT	*COL1A2* / exon37	Exon skipping
agatgAcattgccccctggggccttatgtttgggaa cacag (G>A)GAGCGCAGAGT	*CLCN2* / exon 19	Exon skipping
ctgtgAccccaaattggtcttcatcctctctctaag (G>T)CTCCATGGTTC	*CAPN3* / exon 5	Exon skipping
atctgAagcatcttcctttctgtttcttctcaag (G>T)TTCCCAAAGAG	*CAPN3* / exon 12	Exon skipping
Mutations not affecting splicing (AG-independent 3' splice sites to which U2AF35 does not bind)		
ttctaAgcagtttacgtgccaattcAatttcttaacc tatctcaaag (A>G)TGGAGATCAGT	F9 / exon 4	No effect on splicing
ggcttAtatctgtttattattcAgtattcctgtgtaca ttttctgtttttatttttatacag (G>del)CTATGTAGAAC	*MSH2* / exon 12	No effect on splicing
gagtgAgtccagggtgcctagacaagaggtagc agcctgtggatgtccagcactttgcagggaatac agggcccaatctggcacatgccccttttcctccag (G>C)CCCAGAGCAGG	*HEXA* / exon 13	No effect on splicing
aatttAcaaatctgtgttcctgctttttttccctttaag (G>T)CCTCGATCCAG	*LPL* / exon 5	No effect on splicing
ttgtgAtctcttgatttatttcag (G>A)CAAATCCTAAG	*APC* / exon 16	No effect on splicing
ttttaAgtgtttattttcattgactttgcag (A>del)TATTAATGAAT	*FBN1* / exon 20	Predicted to be only amino acid substitution

*Corresponding Author, E-mail: ohnok@med.nagoya-u.ac.jp

ctgtaActgtttatttccaacag (G>A)GCCCTCCTGGT	*COL1A2* / exon23	No effect on splicing
ctctgAgggtctcttgtctttcctcag (G>A)TGAATGTGGAA	*LAMA2* / exon 24	No effect on splicing
tgttgAcccttcctcctcccatgacag (G>A)TGCAGCCGCTG	*NEU1* / exon 2	No effect on splicing
cactaAtgcccctctctcctcctgcccccag (G>T)GCGTTCCTGGC	*COL6A2* / exon 8	No effect on splicing

Table 2: Mutations at the first nucleotide of exon

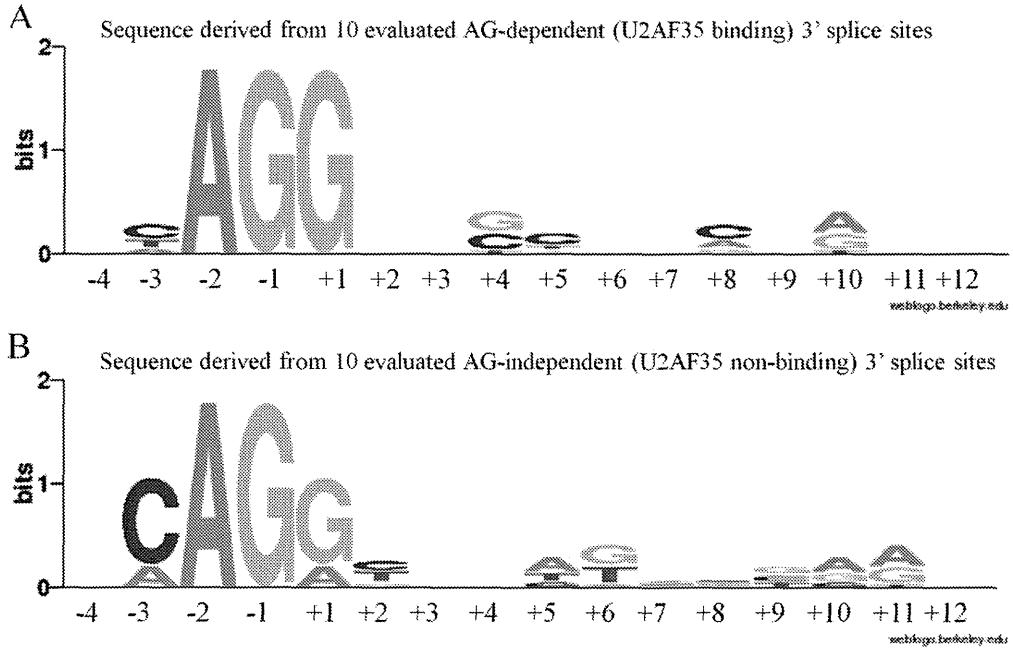

Figure 2: Sequence logos at the 3' splice sites. **A.** Ten mutations in Table 2 causing aberrant splicing. **B.** Ten mutations in Table 2 immune to aberrant splicing.

3.1.3 Exonic splicing enhancers (ESE) / Exonic splicing silencers (ESS)

It has been predicted that 16-20% of missense mutations disrupts ESEs and change the splicing pattern of the gene (Gorlov et al., 2003). Comparison of effects of mutations in HGMD and neutral SNPs in dbSNP on ESE/ESS hexamers and subsequent validation of an example of *de novo* generated ESSs revealed that 7,154 out of 27,681 missense and nonsense mutations (26%) are likely to cause exon skipping (Sterne-Weiler et al., 2011). Indeed, we reported that E154X and EF157V in the *CHRNE* gene (Ohno et al., 2003) and E415G in the *COLQ* gene (Kimbell et al., 2004) in congenital myasthenic syndromes were not nonsense or missense mutations, but were splicing mutations. Although the predicted ratios of 16-26% may be too high, many ESE/ESS-disrupting mutations are likely to be underestimated.

*Corresponding Author, E-mail: ohnok@med.nagoya-u.ac.jp

Although ESEs/ESSs are poorly conserved and difficult to predict, several motif search tools are available online: ESE finder 3.0 (Cartegni et al., 2003), ESRsearch (Goren et al., 2006), FAS-ESS (Wang et al., 2004), PESXs (Zhang&Chasin, 2004) (Zhang et al., 2005), RESCUE-ESE (Fairbrother et al., 2002), Human Splicing Finder (Desmet et al., 2009), SpliceAid (Piva et al., 2009), and SpliceAid2 (Piva et al., 2009). Two additional prediction algorithms employ different parameters to predict auxiliary splicing elements including ESE/ESS/ISE/ISS. CRYP-SKIP is based on a dataset of 250 skipped exons and 204 activated cryptic splice sites due to disease-causing mutations (Divina et al., 2009). Spliceman is based on positional distributions of hexamers that comprise auxiliary splicing elements. Spliceman calculates the "L1 distance", which is a measure how much a mutation changes the position of a cluster of hexamers (Lim et al., 2011). As these predictive algorithms tend to show false positives, *in vivo* and/or *in vitro* splicing analyses are essential to prove the splicing effect of a mutation. But it must be kept in mind that these web sites are easily discontinued and difficult to maintain due to accelerating development of new tools.

4 NAS and NASRE

4.1 Nonsense-associated altered splicing (NAS) and NMD-associated skipping of a remote exon (NASRE)

Originally found in the fibrillin gene (Dietz et al., 1993; Dietz&Kendzior, 1994), a premature terminating codon (PTC) sometimes leads to aberrant splicing, which is known as nonsense-associated altered splicing (NAS) (Hentze&Kulozik, 1999). Involvement of NMD as a cause of NAS has been suggested (Mendell et al., 2002; Wang et al., 2002). Disruption of an ESE is also indicated as a cause of NAS (Caputi et al., 2002; Liu et al., 2001). Another mechanism, suppression of splicing (SOS), in which neither NMD nor ESE is involved, is also proposed (Wachtel et al., 2004). As PTC itself has no effect on efficiency of removal of an adjacent intron (Lytle&Steitz, 2004), NAS is likely to be mediated by disruption of an ESE. In addition to an effect of ESEs on splicing, disruption of ESEs also stabilize pre-mRNA by unknown mechanisms (Imam et al., 2010; Muhlemann et al., 2001) .

A mutation can also induce skipping of a remote exon on which the mutation is not located. In congenital myasthenia syndrome, we found that a 7-bp deletion in *CHRNE* exon 7 leads to skipping of the preceding exon 6 of 101 bp, and we named it NMD-associated skipping of a remote exon (NASRE). Exon 6 has weak splicing signals and is skipped even in normal conditions. The exon 6-skipped product, however, is completely degraded by NMD due to a PTC generated by a frameshift. When exon 7 has a 7-bp deletion, the open reading frame is resumed and the exon 6-skipped product becomes immune to NMD (Ohno et al., 2003). We reported other examples of NASRE, but NASRE is likely to be underestimated because remote exons are rarely scrutinized in the analysis of human diseases.

Conclusions

We have summarized canonical and auxiliary splicing *cis*-elements in physiological and pathological conditions. We hope to emphasize that splicing *cis*-elements can be located anywhere on the exon and intron, and aberrant splicing may be observed in any disease in any tissue. As correction of aberrant splicing by antisense oligonucleotides and small molecules are intensively sought for by many investigators, the identification of splicing-disrupting mutations is essential for understanding disease mechanisms and also for ameliorating pathomechanisms.

*Corresponding Author, E-mail: ohnok@med.nagoya-u.ac.jp

References

Beckmann, J. S.&Trifonov, E. N. (1991). "Splice junctions follow a 205-base ladder." *Proceedings of the National Academy of Sciences of the United States of America* 88,2380-2383.

Berget, S. M. (1995). "Exon recognition in vertebrate splicing." *The Journal of biological chemistry* 270,2411-2414.

Bian, Y., et al. (2009). "Tannic acid facilitates expression of the polypyrimidine tract binding protein and alleviates deleterious inclusion of CHRNA1 exon P3A due to an hnRNP H-disrupting mutation in congenital myasthenic syndrome." *Human molecular genetics* 18,1229-1237.

Bianchi, P., et al. (1997). "A variant of the EPB3 gene of the anti-Lepore type in hereditary spherocytosis." *British journal of haematology* 98,283-288.

Burd, C. G.&Dreyfuss, G. (1994). "RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing." *EMBO J* 13,1197-1204.

Burge, C.&Karlin, S. (1997). "Prediction of complete gene structures in human genomic DNA." *J Mol Biol* 268,78-94.

Caputi, M., et al. (2002). "A nonsense mutation in the fibrillin-1 gene of a Marfan syndrome patient induces NMD and disrupts an exonic splicing enhancer." *Genes & development* 16,1754-1759.

Cartegni, L., et al. (2003). "ESEfinder: A web resource to identify exonic splicing enhancers." *Nucleic acids research* 31,3568-3571.

Chua, K.&Reed, R. (2001). "An upstream AG determines whether a downstream AG is selected during catalytic step II of splicing." *Molecular and cellular biology* 21,1509-1514.

Corrionero, A., et al. (2011). "Strict 3' splice site sequence requirements for U2 snRNP recruitment after U2AF binding underlie a genetic defect leading to autoimmune disease." *RNA* 17,401-411.

Corvelo, A., et al. (2010). "Genome-wide association between branch point properties and alternative splicing." *PLoS computational biology* 6,e1001016.

Crooks, G. E., et al. (2004). "WebLogo: a sequence logo generator." *Genome research* 14,1188-1190.

Desmet, F. O., et al. (2009). "Human Splicing Finder: an online bioinformatics tool to predict splicing signals." *Nucleic acids research* 37,e67.

Dhir, A.&Buratti, E. (2010). "Alternative splicing: role of pseudoexons in human disease and potential therapeutic strategies." *FEBS J* 277,841-855.

Dietz, H. C.&Kendzior, R. J., Jr. (1994). "Maintenance of an open reading frame as an additional level of scrutiny during splice site selection." *Nat Genet* 8,183-188.

Dietz, H. C., et al. (1993). "The skipping of constitutive exons in vivo induced by nonsense mutations." *Science* 259,680-683.

Divina, P., et al. (2009). "Ab initio prediction of mutation-induced cryptic splice-site activation and exon skipping." *European journal of human genetics : EJHG* 17,759-765.

Dlott, B., et al. (1990). "Two mutations produce intron insertion in mRNA and elongated beta-subunit of human beta-hexosaminidase." *The Journal of biological chemistry* 265,17921-17927.

Fairbrother, W. G., et al. (2002). "Predictive identification of exonic splicing enhancers in human genes." *Science* 297,1007-1013.

Fedorova, L.&Fedorov, A. (2005). "Puzzles of the human genome: Why do we need our introns?" *Current Genomics* 6,589-595.

*Corresponding Author, E-mail: ohnok@med.nagoya-u.ac.jp