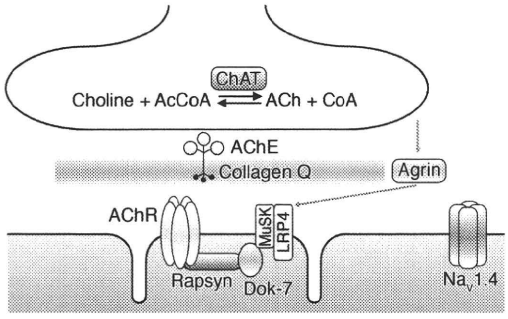
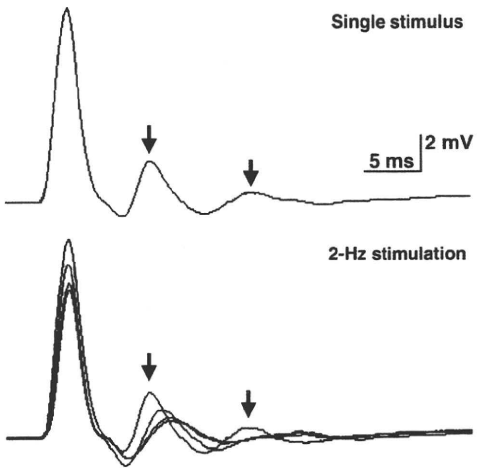
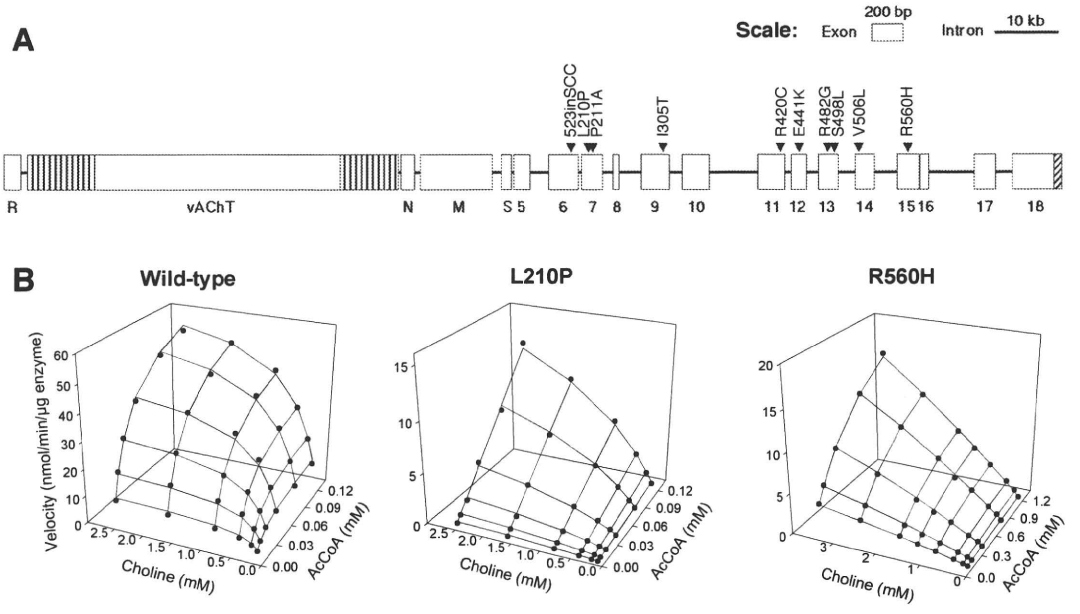
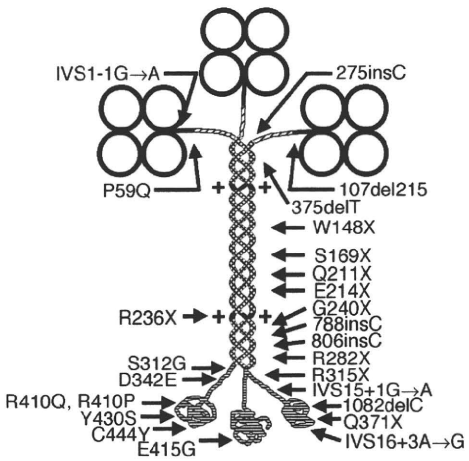
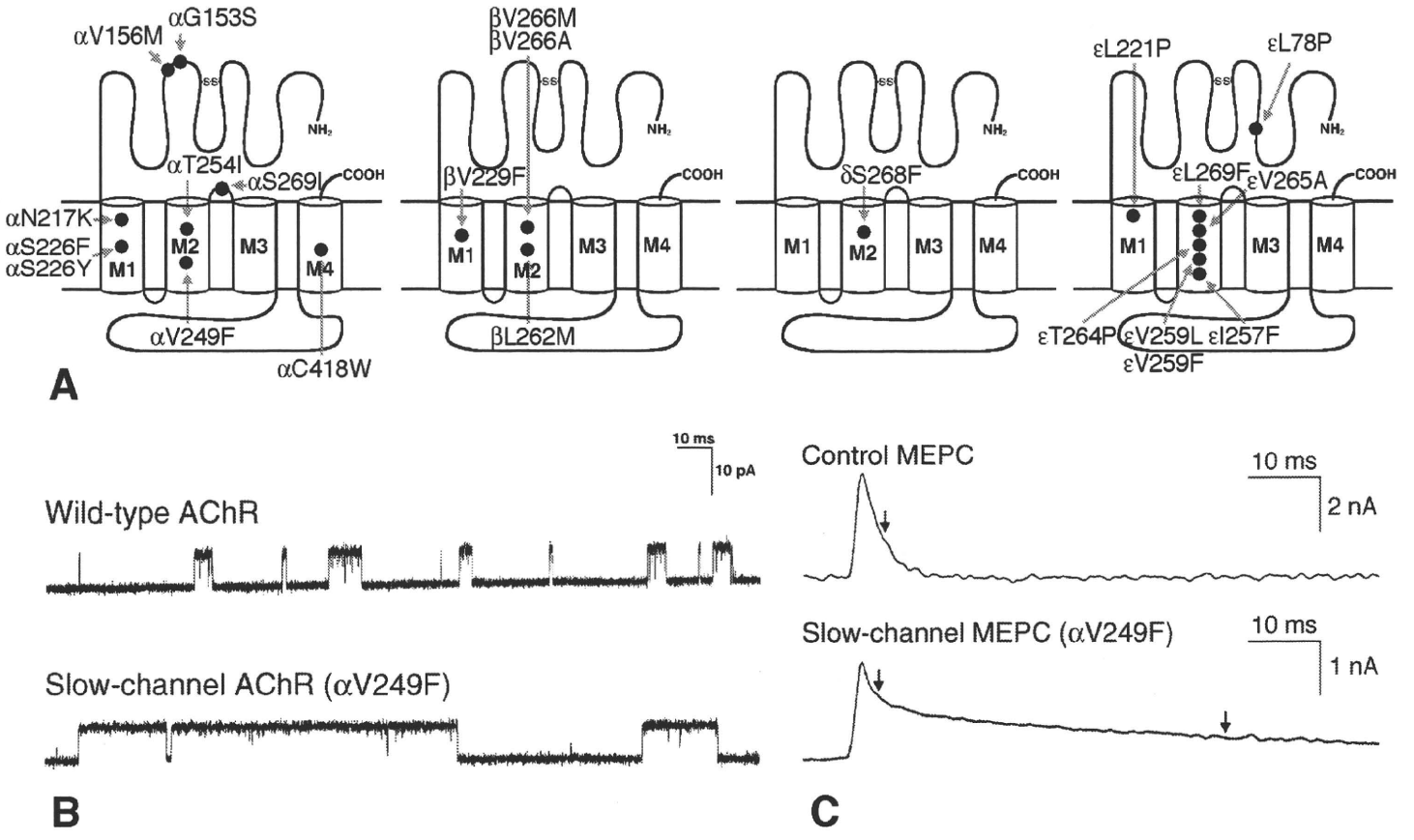
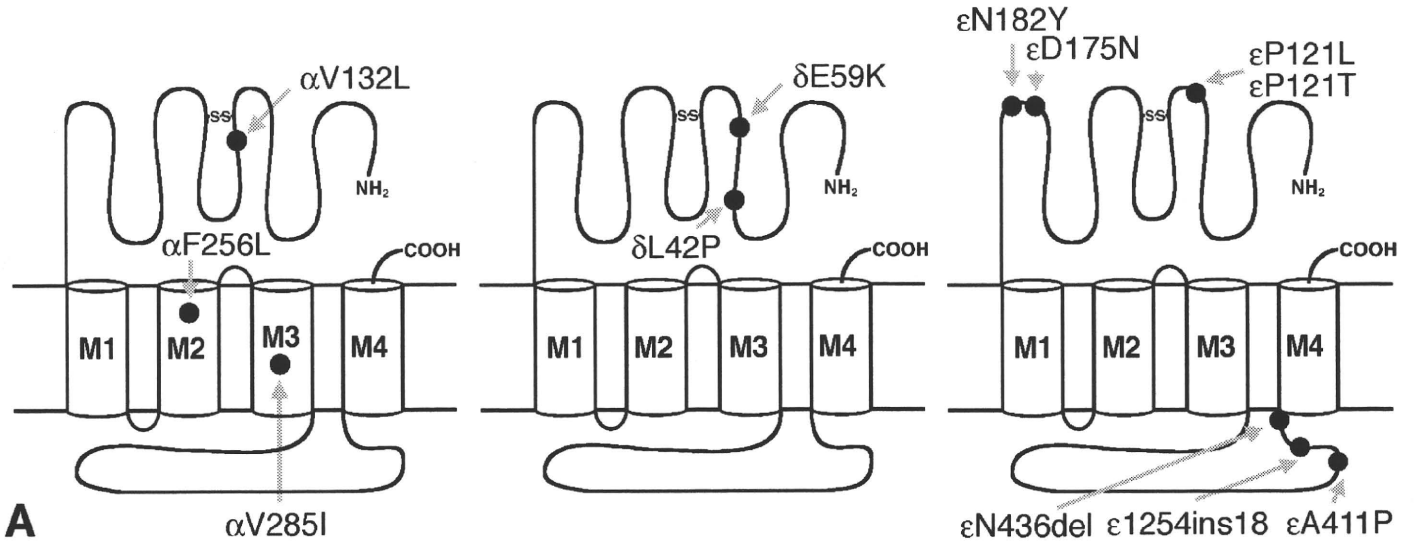

ChAT
Choline + AcCoA
ACh + CoA
AChE
Collagen Q
Agrin
AChR
MuSK
LRP4
Rapsyn
Dok-7
$Na_v1.4$

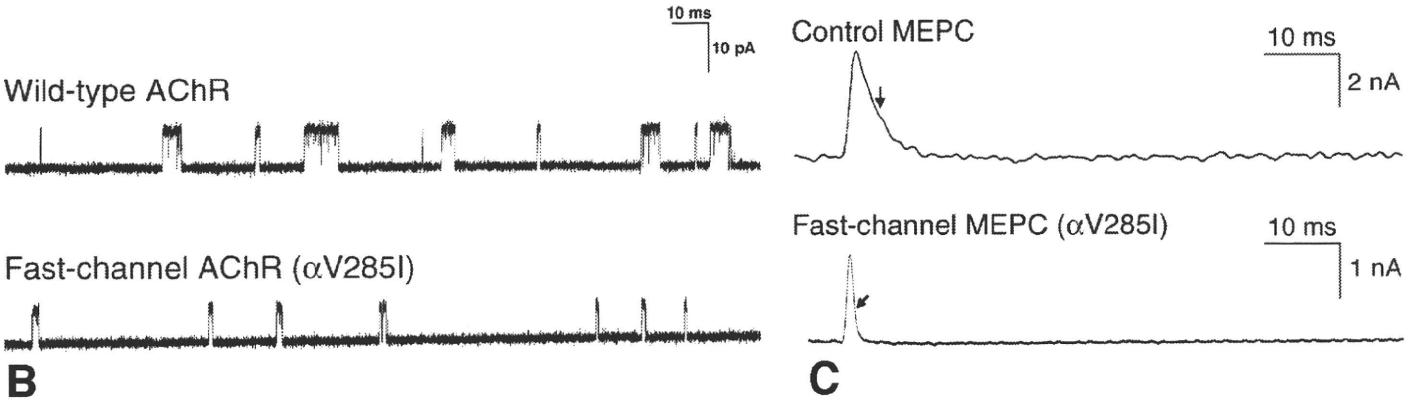

Single stimulus
2 mV
5 ms
2-Hz stimulation

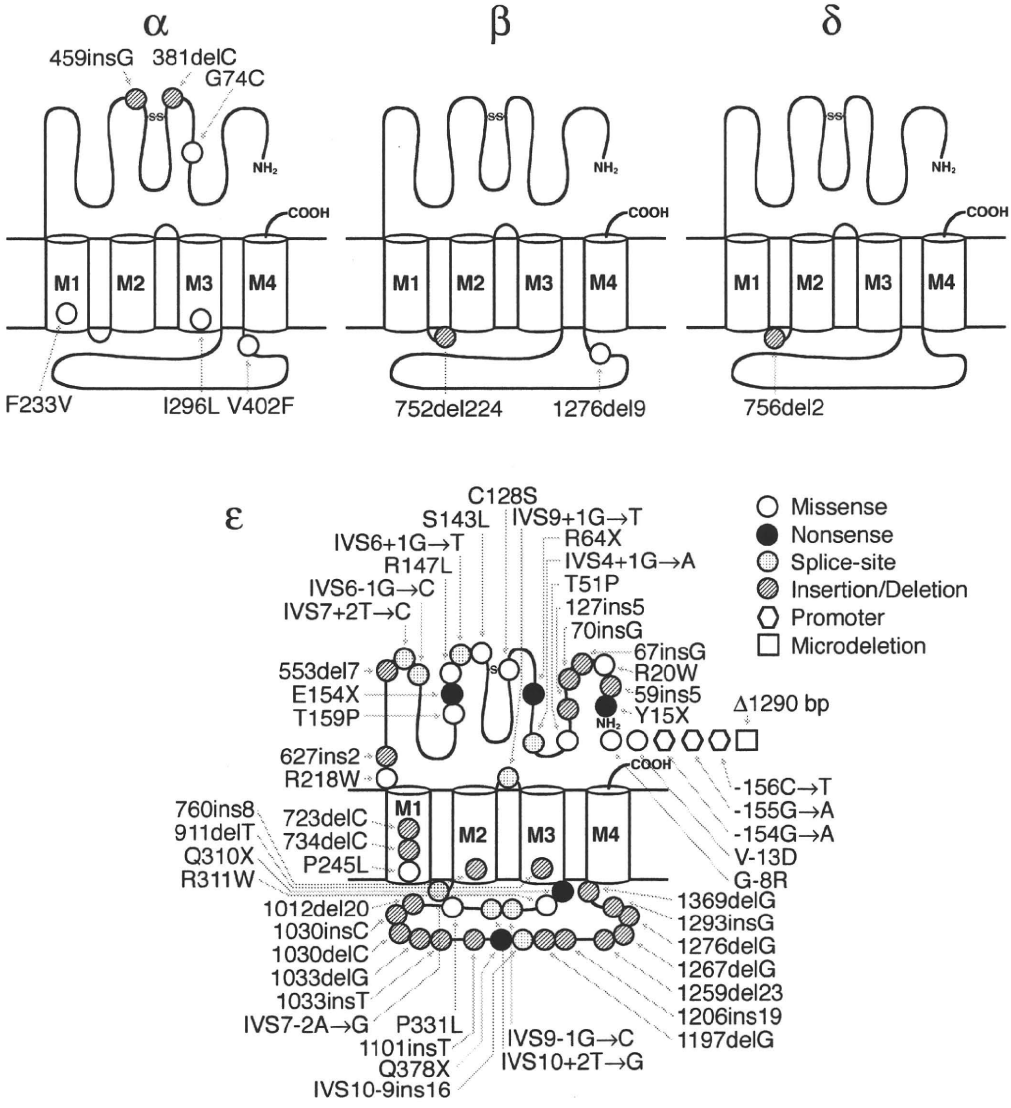

A
Scale:
Exon
200 bp
Intron
10 kb
523insSCC
L210P
P211A
I305T
R420C
E441K
R482G
S498L
V506L
R560H
R
vAChT
N
M
S 5
6
7
8
9
10
11
12
13
14
15
16
17
18
B
Wild-type
L210P
R560H
Velocity (nmol/min/µg enzyme)
Choline (mM)
AcCoA (mM)

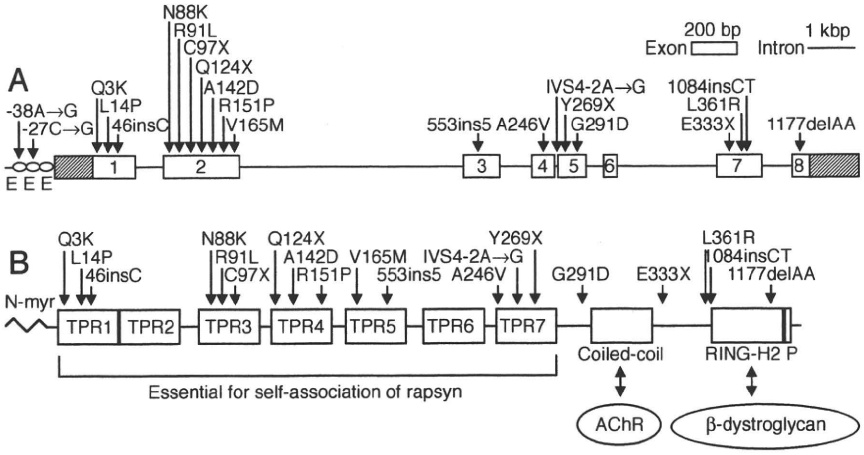

IVS1-1G→A
275insC
P59Q
107del215
375delT
W148X
S169X
Q211X
E214X
R236X
G240X
788insC
806insC
S312G
R282X
D342E
R315X
IVS15+1G→A
R410Q, R410P
1082delC
Y430S
Q371X
C444Y
IVS16+3A→G
E415G

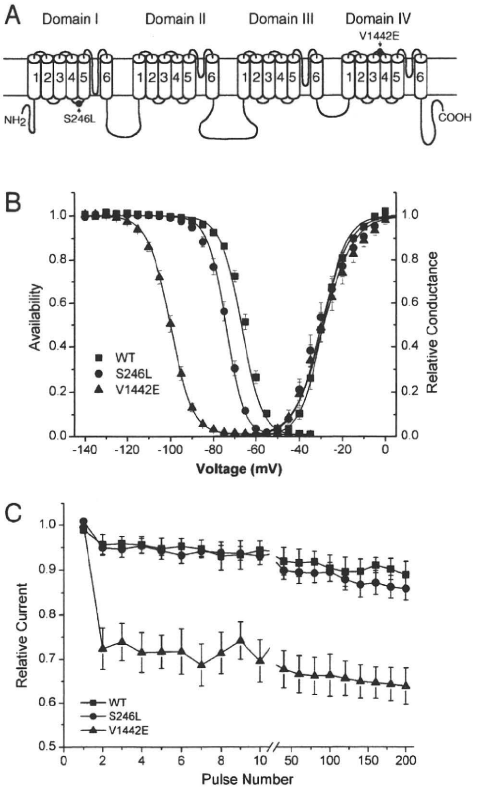

αV156M
αG153S
αT254I
αS269I
αN217K
αS226F
αS226Y
αV249F
αC418W
βV266M
βV266A
βV229F
βL262M
δS268F
εL221P
εL78P
εL269F
εV265A
εT264P
εV259L
εI257F
εV259F
M1
M2
M3
M4
COOH
NH2
A
10 ms
10 pA
Wild-type AChR
Slow-channel AChR (αV249F)
B
Control MEPC
10 ms
2 nA
Slow-channel MEPC (αV249F)
10 ms
1 nA
C


αV132L
δE59K
εN182Y
εD175N
εP121L
εP121T
αF256L
δL42P
αV285I
εN436del ε1254ins18 εA411P
M1
M2
M3
M4
COOH
NH2
SS
A


10 ms
10 pA
Wild-type AChR
Fast-channel AChR (αV285I)
B
Control MEPC
10 ms
2 nA
Fast-channel MEPC (αV285I)
10 ms
1 nA
C


α
β
δ
459insG
381delC
G74C
SS
NH2
COOH
M1
M2
M3
M4
F233V
I296L V402F
752del224
1276del9
756del2
ε
C128S
S143L
IVS9+1G→T
IVS6+1G→T
R64X
R147L
IVS4+1G→A
IVS6-1G→C
T51P
IVS7+2T→C
127ins5
70insG
67insG
553del7
R20W
E154X
59ins5
T159P
Y15X
Δ1290 bp
627ins2
R218W
-156C→T
-155G→A
760ins8
723delC
-154G→A
911delT
734delC
Q310X
P245L
V-13D
R311W
G-8R
1369delG
1012del20
1293insG
1030insC
1276delG
1030delC
1267delG
1033delG
1259del23
1033insT
1206ins19
IVS7-2A→G
P331L
1197delG
1101insT
IVS9-1G→C
Q378X
IVS10+2T→G
IVS10-9ins16
Missense
Nonsense
Splice-site
Insertion/Deletion
Promoter
Microdeletion


A
N88K
R91L
C97X
Q124X
A142D
R151P
V165M
Q3K
L14P
46insC
-38A→G
-27C→G
553ins5 A246V
IVS4-2A→G
Y269X
G291D
1084insCT
L361R
E333X
1177delAA
200 bp
Exon
1 kbp
Intron
E E E
1
2
3
4
5
6
7
8
B
Q3K
L14P
46insC
N88K
R91L
C97X
Q124X
A142D
R151P
V165M
553ins5
IVS4-2A→G
A246V
Y269X
G291D
E333X
L361R
1084insCT
1177delAA
N-myr
TPR1
TPR2
TPR3
TPR4
TPR5
TPR6
TPR7
Coiled-coil
RING-H2 P
Essential for self-association of rapsyn
AChR
β-dystroglycan


A
Domain I
Domain II
Domain III
Domain IV
V1442E
S246L
NH2
COOH
B
Availability
Relative Conductance
WT
S246L
V1442E
Voltage (mV)
C
Relative Current
WT
S246L
V1442E
Pulse Number

Handbook of Neurochemistry and Molecular Neurobiology 3rd Edition

Volume No. 25 "Neurochemical Mechanisms in Disease"

"RNA Pathologies in Neurological Disorders"

Kinji Ohno, Akio Masuda

Division of Neurogenetics, Center for Neurological Diseases and Cancer, Nagoya University Graduate School of Medicine, Nagoya, Japan

E-mail: ohnok@med.nagoya-u.ac.jp

Abstract

RNA is not a simple intermediate linking DNA and protein. RNA is widely transcribed from a variety of genomic regions, and extensive studies on the functional roles and regulations of non-coding RNAs including antisense RNAs and small RNAs are in progress. In addition, the human genome project revealed that we humans carry as few as ~22,000 genes. Humans exploit tissue-specific and developmental stage-specific alternative splicing to generate a large variety of molecules in specific cells at specific developmental stages. Neurological disorders are also subject to aberrations of the splicing mechanisms. This review focuses mostly on splicing abnormalities due to pathological alterations of splicing *cis*-elements and *trans*-factors. Pathomechanisms associated with disrupted splicing *cis*-elements can be applied to any human diseases, and we did not restrict the descriptions to neurological diseases. On the other hand, we limited the descriptions of dysregulated splicing *trans*-factors to neurological disorders. Neurological diseases covered in this review include congenital myasthenic syndromes, spinal muscular atrophy, myotonic dystrophy, Alzheimer's disease, frontotemporal dementia with parkinsonism linked to

chromosome 17, facioscapulohumeral muscular dystrophy, fragile X-associated tremor/ataxia syndrome, Prader-Willi syndrome, Rett syndrome, spinocerebellar atrophy type 8, and paraneoplastic neurological disorders.

Key words: The RNA world, pre-mRNA splicing, splicing *cis*-elements, splicing *trans*-factors, branch point sequence (BPS), exonic splicing enhancer (ESE), exonic splicing silencer (ESS), intronic splicing enhancer (ISE), intronic splicing silencer (ISS), nonsense-mediated mRNA decay (NMD), nonsense-associated skipping of a remote exon (NASRE), congenital myasthenic syndromes, spinal muscular atrophy (SMA), myotonic dystrophy (DM1, DM2), Alzheimer's disease, frontotemporal dementia with parkinsonism linked to chromosome 17 (FTDP-17), facioscapulohumeral muscular dystrophy (FSHD), fragile X-associated tremor/ataxia syndrome (FXTAS), Prader-Willi syndrome, Rett syndrome, spinocerebellar atrophy type 8 (SCA8), and paraneoplastic neurological disorders (PND).

1. Introduction

The central dogma first enunciated by Francis Crick depicts that RNA is an intermediate that links DNA and protein (Crick, 1970). The beginning of life, however, was the RNA world where there was no DNA or proteins (Gilbert, 1986). In the RNA world, RNA was the only carrier of genetic information that DNA currently serves, and the only functional molecule that proteins currently serve. Although the RNA transmits no genetic information to progeny and constitutes a limited number of functional molecules in our human body, the RNA world is still in effect in our body. Humans transcribe more than half of our entire genome including noncoding regions. The transcripts work as antisense RNAs, microRNAs, and snoRNAs. Researchers are now working to disclose the functional significance of these noncoding RNAs.

The human genome project and the subsequent annotation efforts revealed that we humans carry as few as 22,000 genes. Tissue-specific and developmental stage-specific splicing enables to us to generate more than 100,000 molecules from a limited number of genes (Black, 2003; Licatalosi and Darnell, 2006). Small RNA molecules and RNA splicing mechanisms potentially become targets of neurological diseases (Ranum and Cooper, 2006). This review focuses mostly on splicing aberrations associated with neurological disorders.

2. Physiology of splicing mechanisms

In higher eukaryotes, pre-mRNA splicing is mediated by degenerative splicing *cis*-elements comprised of the branch point sequence (BPS), the polypyrimidine tract (PPT), the 5' and 3' splice sites, and exonic/intronic splicing enhancers/ silencers **(Fig. 1)**. Stepwise assembly of the spliceosome starts from recruitment of U1 snRNP to the 5' splice site, SF1 to the BPS, U2AF65 to the PPT, and U2AF35 to the 3' end of an

intron to form a spliceosome complex E (Sperling et al., 2008). SF1, a 75-kDa protein, is a mammalian homolog of yeast BBP (branch point-binding protein). U2AF65 and U2AF35 bring U2 snRNP to the BPS in place of SF1 (Wu et al., 1999; Zorio and Blumenthal, 1999). The BPS establishes base pairing interactions with a stretch of 'GUAGUA' of U2 snRNA (Arning et al., 1996; Abovich and Rosbash, 1997), which then bulges out the branch site nucleotide, usually an adenosine to form a spliceosome complex A (Query et al., 1994). Thereafter, pre-mRNAs are spliced in two sequential transesterification reactions mediated by the spliceosome. In the first step, the 2'-OH moiety of the branch site nucleotide carries out a nucleophilic attack against a phosphate at the 5' splice site, generating a free upstream exon, as well as a lariat carrying the intron and the downstream exon. In the second step, the 3'-OH moiety of the upstream exon attacks the 3' splice site of the lariat leading to intron excision and ligation of the upstream and downstream exons (Query et al., 1996).

In addition to the "classical" spliceosomal mechanisms, splicing is modulated by exonic/intronic splicing enhancers/silencers (ESE, ISE, ESS, ISS). The *trans*-factors for the splicing enhancers/silencers carry repeats of arginine and serine, are accordingly called SR proteins. Tissue-specific and developmental stage-specific expressions of the splicing *trans*-factors enable precise spatial and temporal regulations of the gene expressions. In addition, the splicing *trans*-factors also work on constitutively spliced exons to compensate for highly degenerative "classical" splicing *cis*-elements.

3. Disorders associated with disruption of splicing *cis*-elements

3.1 Aberrations of the 5' splice sites

Mutations disrupting the 5' splice sites have been most frequently reported. U1 snRNA recognizes three nucleotides at the end of an exon and six nucleotides at the beginning of an intron **(Fig. 2)**. The completely matched nucleotides to U1 snRNA are CAG|GTAAGT, where the vertical line represent the exon/intron boundary. The completely matched sequence is observed at 1,597 sites out of the entire 189,249 5' splice sites in the human genome (Sahashi et al., 2007), which is the 10th most common sequence. The completely matched 5' splice site is rather avoided because, in the second stage of splicing, U1 snRNA is substituted for with U5 snRNA. If U1 snRNA is tightly bound to the 5' splice site, it hinders binding of U5 snRNA.

Degeneracy of the 5' splice site and its vulnerability to disease-causing mutations have been extensively studied. Three algorithms have been proposed. First, Shapiro and Senapathy collated nucleotide frequencies at each position of the 5' splice site. They assumed that nucleotide frequencies at each position of the 5' splice site represent the splicing signal intensity. They thus constructed a linear regression model so that the most preferred 5' splice site becomes 1.0 and the most unfavorable 5' splice site becomes 0.0 (Shapiro and Senapathy, 1987). Second, Rogan and Schneider invented the information contents, R*i*. For example, at a specific position, if a single nucleotide is exclusively used, the information content at this position becomes $-\log_2(1/4) = 2$ bits. Similarly, if two nucleotides are equally used, the information content becomes $-\log_2(2/4) = 1$ bit. In R*i*, the similarity to the consensus sequence is represented by the sum of information bits (Rogan and Schneider, 1995; O'Neill et al., 1998). Third, we found that a new parameter, the SD-Score, which represents a common logarithm of the frequency of a specific 5' splice site in the human genome, efficiently predicts the splicing signal intensity (Sahashi et al., 2007). Our algorithm predicts the splicing consequences of mutations with the sensitivity of

97.1% and the specificity of 94.7%. Simulation of all the possible mutations in the human genome using the SD-score algorithm predicts high frequencies of splicing mutations from exon -3 to intron +6 **(Table 1)**. Especially at exon position -3, about one third of mutations are predicted to cause aberrant splicing. Using our algorithm, we predicted and proved that *DYSF* G1842D in Miyoshi myopathy, *ABCD1* R545W in adrenoleucodystrophy, *GLA* Q333X in Fabry disease, and *DMD* Q119X and Q1144X in Duchenne muscular dystrophy are not missense or nonsense mutations but are splicing mutations. Algorithms by us and by others all point to the notion that aberrant splicing caused by mutations at the 5' splice sites are likely to be underestimated.

3.2 Human branch point consensus sequence

In an effort to seek for an algorithm to predict the position of the branch point sequence (BPS) in human, we sequenced 367 clones of lariat RT-PCR products arising from 52 introns of 20 human housekeeping genes and identified that the human consensus BPS is simply yUnAy, where 'y' represents U or C (Gao et al., 2008) **(Fig. 3)**. The consensus BPS was more degenerative than we had expected and we failed to construct a dependable algorithm that predicts the position of the BPS. Sixteen disease-causing mutations and a polymorphism, however, have been reported to date that disrupt a BPS and cause aberrant splicing (Gao et al., 2008). Among these, eight mutates U at position -2, whereas nine affects A at position 0, which also supports the notion that U at -2 and A at 0 are essential nucleotides.

3.3 Ectopic AG dinucleotide abrogates the AG-scanning mechanism

The 3' end of an intron and the 5' end of an exon carry a consensus sequence of CAG|G, where the vertical line represents the intron/exon boundary. The AG dinucleotide is scanned from the branch point and the first AG is recognized as the 3' end of the intron (Chen et al., 2000). In a patient with congenital myasthenic syndrome, we identified duplication of a 16-nt segment comprised of 8 intronic and 8 exonic nucleotides at the intron 10/exon 10 boundary of *CHRNE* encoding the acetylcholine receptor epsilon subunit (Ohno et al., 2005). We found that the upstream AG of the duplicated segment is exclusively used for splicing and that one or two mutations in the upstream BPS had no effect whereas complete deletion of the upstream BPS partially activated the downstream AG. Similar exclusive activation of the upstream AG is reported in *HEXB* (Dlott et al., 1990) and *SLC4A1* (Bianchi et al., 1997). Creation of a cryptic AG dinucleotide close to the 3' end of an intron should be carefully scrutinized in mutation analysis.

3.4 Mutations that disrupt ESE and ESS

Gorlov and colleagues predicted that more than 16-20% of missense mutations are splicing mutations that disrupt an ESE (Gorlov et al., 2003). According to our own experience, their estimates are likely to be too high. Most ESE/ESS-disrupting mutations, however, are likely to be underestimated, because the positions and sequences of ESE/ESS are highly degenerative.

Four web services provide valuable information to locate ESE and ESS. First, the ESE Finder (http://rulai.cshl.org/ESE/) calculates the similarity of a given nucleotide sequence to the consensus sequences of four splicing *trans*-factors, SF2/ASF, SC35, SRp40, and SRp55 (Cartegni et al., 2003; Smith et al., 2006). Second, the RESUCE-ESE web server (http://genes.mit.edu/burgelab/rescue-ese/)

shows the similarity of a given sequence to ESE elements of unidentified splicing *trans*-factors (Fairbrother et al., 2002). The same group also provides the FAS-ESS web service to screen for ESS elements (http://genes.mit.edu/fas-ess/) (Wang et al., 2004). Third, the PESX web server (http://cubweb.biology.columbia.edu/pesx/) indicates an RNA octamer with putative exonic splicing enhancing or silencing activities (Zhang and Chasin, 2004; Zhang et al., 2005). Fourth, the ESRsearch web server (http://ast.bioinfo.tau.ac.il/) shows 285 candidate ESE/ESS sequences (Goren et al., 2006), as well as ESE/ESS elements indicated by the RESCUE-ESE, FAS-ESS, and PESX services.

In patients with congenital myasthenic syndromes, we identified that *CHRNE* E154X and EF157V (Ohno et al., 2003), as well as *COLQ* E415G (Kimbell et al., 2004), disrupt an ESE and cause aberrant splicing. The ESE/ESS servers above indicate disruption of candidate splicing *cis*-elements for all the three mutations, but we frequently obtain false positives and we cannot simply rely on the servers. Analysis of patient mRNA or analysis using a minigene is generally expected.

3.5 Mutations that disrupt ISE and ISS

Identification of mutations disrupting intronic splicing *cis*-elements is more challenging than that of exonic mutations, because introns are longer than exons and splicing mutations can be anywhere in the introns, and because we do not have a dependable algorithm to predict ISE/ISS. The ESRsearch web server described above is able to indicate consensus sequences recognized by a variety of splicing *trans*-factors including intronic ones.

In a patient with congenital myasthenic syndrome, we identified that *CHRNA1* IVS3-8G>A attenuates binding of hnRNP H ~100-fold and causes exclusive inclusion

of the downstream exon P3A (Masuda et al., 2008) **(Fig. 4)**. We also identified that polypyrimidine tract binding protein (PTB) silences recognition of exon P3A and tannic acid facilitates the expression of PTB by activating its promoter region (submitted).

3.6 Spinal muscular atrophy (SMA)

SMA is an autosomal recessive disorder characterized by degeneration of the anterior horn cells of the spinal cord, which causes muscular weakness and atrophy. SMA is caused by loss-of-function mutations including deletion of the *SMN1* gene that encodes the survival of motor neuron 1. Humans carry almost identical *SMN1* and *SMN2* genes both on chromosome 5q13. *SMN2* carries a C-to-T transition at position 6 of exon 7 compared to *SMN1*, which results in loss of an SF2/ASF-dependent ESE activity (Cartegni et al., 2006). In addition, *SMN2* carries an A-to-G transition at position +100 of intron 7, which creates a high-affinity hnRNP A1-binding site and promotes skipping of exon 7 (Kashima et al., 2007). Skipping of exon 7 in *SMN2* can be ameliorated by therapeutic doses of valproic acid (Brichta et al., 2003; Brichta et al., 2006) and of salbutamol (Angelozzi et al., 2008).

4. Skipping of multiple exons caused by a single splicing mutation

4.1 Skipping of multiple contiguous exons

A mutation disrupting a splicing *cis*-element generally affects splicing of a single exon or intron, but sometimes generates aberrant transcripts affecting multiple neighboring exons. Skipping of multiple contiguous exons is accounted for by

ordered removal of introns and consequent clustering of neighboring exons (Schwarze et al., 1999; Takahara et al., 2002).

4.2 Nonsense-associated skipping of a remote exon (NASRE)

A single mutation infrequently causes skipping of a remote exon. In a patient with congenital myasthenic syndrome, we found that a 7-nt deletion in exon 7 of *CHRNE* causes complete skipping of the preceding exon 6. *CHRNE* exon 6 is composed of 101 nucleotides. It carries weak splicing signals and is partially skipped even in normal subjects. The exon 6-skipped transcript, however, is removed by the nonsense-mediated mRNA decay (NMD) mechanism. The 7-nt deletion in exon 7 restores the open reading frame of the exon 6-skipped transcript and renders it immune to NMD. On the other hand, the normally spliced transcript carries a premature stop codon (PTC) after the 7-nt deletion, and is degraded by NMD **(Fig. 5)**. We dubbed this mechanism NASRE, and found that it is in effect in *SLC25A20* (Hsu et al., 2001), *DBT* (Fisher et al., 1993), *BTK* (Haire et al., 1997), and *MLH1* (Clarke et al., 2000).

5. Disorders associated with dysregulation of splicing *trans*-factors

5.1 Myotonic dystrophy

Myotonic dystrophy is an autosomal dominant multisystem disorder affecting skeletal muscles, eye, heart, endocrine system, and central nervous system. The clinical symptoms include variable degrees of muscle weakness and wasting, myotonia, cataract, insulin resistance, hypogonadism, cardiac conduction defects, frontal balding, and intellectual disabilities (Harper and Monckton, 2004). Myotonic

dystrophy is caused by abnormally expanded CTG repeats in the 3' untranslated region of the *DMPK* gene encoding the dystrophia myotonica protein kinase on chromosome 19q13 (myotonic dystrophy type 1, DM1) (Brook et al., 1992) or by abnormally expanded CCTG repeats in intron 1 of the *ZNF9* gene encoding the zinc finger protein 9 on chromosome 3q21 (myotonic dystrophy type 2, DM2) (Liquori et al., 2001). In DM1, normal individuals have 5 to 30 repeats, mildly affected patients have 50 to 80 repeats, and severely affected individuals have 2,000 or more copies of CTG (Gharehbaghi-Schnell et al., 1998). In DM2, the size of expanded repeats is extremely variable, ranging from 75 to 11,000 repeats, with a mean of 5,000 CCTG repeats (Liquori et al., 2001).

In both DM1 and DM2, expanded CTG or CCTG repeats in the noncoding regions sequestrate a splicing *trans*-factor muscleblind encoded by *MBNL1* to intranuclear RNA foci harboring the mutant RNA, and somehow upregulate another splicing *trans*-factor CUG-binding protein encoded by *CUGBP1* (Ranum and Cooper, 2006) **(Fig. 6)**. Dysregulation of the two splicing *trans*-factors then causes aberrant splicing of their target genes. The aberrantly spliced genes identified to date in skeletal and cardiac muscles include *ATP2A1* (*SERCA1*) exon 22, *ATP2A2* (*SERCA2*) intron 19, *CAPN3* exon 16, *CLCN1* intron 2 and exons 6b/7a, *DMD* exons 71 and 78, *DTNA* exons 11A and 12, *FHOD1* (*FHOS*) exon 11a, *GFPT1* (*GFAT1*) exon 10, *INSR* exon 11, *KCNAB1* exons 2b/2c, *LDB3* (*ZASP*) exon 11 (189-nt exon 7 according to RefSeq Build 36.3), *MBNL1* exon 7 (54-nt exon 6 according to RefSeq), *MBNL2* exon 7 (54 nt, no exonic annotation in RefSeq), *MTMR1* exons 2.1 and 2.2, *NRAP* exon 12, *PDLIM3* (*ALP*) exons 5a/5b, *RYR1* exon 70, *TNNT2* exon 5, *TNNT3* fetal exon, *TTN* exons Zr4 and Zr5 (138-nt exon 11 and 138-nt exon 12 according to RefSeq), and *TTN* exon Mex5 (303-nt exon 315 according to RefSeq) (Philips et al.,