厚生労働科学研究費補助金

こころの健康科学研究事業

終板アセチルコリンエステラーゼ欠損症、及び、他の細胞外マトリックス分子欠損症におけるタンパク標的療法の開発研究

平成19年度　総括研究報告書

主任研究者　大野　欽司

平成20(2008)年　4月

目　　次

I. 終板アセチルコリンエステラーゼ欠損症、及び、
他の細胞外マトリックス分子欠損症におけるタンパク標的療法の開発研究

主任研究者　大野　欽司　名古屋大学医学系研究科・神経遺伝情報学・教授

研究要旨

先天性筋無力症候群は、神経筋接合部の先天的分子欠損症が原因であり、主任研究者らは欠損分子に応じた治療法を開発・臨床応用してきた(Engel, Ohno, Sine. *Nat Rev Neurosci* 4: 339, 2003)。しかし、collagen Q分子(*COLQ*遺伝子産物)欠損による終板アセチルコリンエステラーゼ(AChE)欠損症は全く治療法が存在しない(Ohno, et al. *Proc Natl Acad Sci USA* 95: 9654, 1998)。Collagen Q 3分子(*COLQ*遺伝子産物)は3重鎖を形成し、AChE catalytic subunit 12分子と結合し、非対称性A_{12}-AChEを形成する。本研究では、collagen Qがシナプス基底膜への係留シグナルを有する細胞外構造タンパクであることを利用し(Kimbell*, Ohno*, et al. *J Biol Chem* 279: 10997, 2004. *equal contribution)、*COLQ*欠損モデル動物のリンパ球に正常*COLQ*遺伝子と正常*ACHE*遺伝子を導入し、A_{12}-AChEを血中に発現させ、シナプス基底膜への係留を試みた。また、*COLQ*遺伝子導入骨格筋が産生をするA_{12}-AChEが*COLQ*遺伝子非導入骨格筋へ組織間液を介して移行することを期待し、骨格筋への正常*COLQ*遺伝子の導入を行った。

一般に、遺伝子治療においては導入遺伝子の細胞特異的・組織特異的なターゲッティングが障壁となり、培養細胞レベルで有効である手法の多くが臨床応用できない。本研究では、組み替え遺伝子を標的組織にターゲッティングさせる代わりに、欠損タンパクが細胞外分子であることと、タンパクに標的組織親和性があることを利用して標的組織へのターゲッティングを行う。本手法は、他の神経難治疾患であるperlecan欠損症、collagen I欠損症、α_2 laminin欠損症、α dystroglycanopathyを含む細胞外マトリックスタンパク欠損症一般への応用の可能性が期待される。

A. 研究目的

神経筋接合部の分子欠損症による先天性筋無力症候群は、世界中から数多くの症例が報告されているが日本からの報告は少ない。諸外国でも*de novo*遺伝子変異が数多く存在することから、本症候群は日本にも多数存在すると考えられる。胎生期からの神経筋接合部伝達障害によると思われる小奇形、筋萎縮、関節拘縮がみられ、諸外国で見られるように日本でも筋ジストロフィー症や先天性筋症と診断をされている例が少なからず存在すると思われる。また、重症筋無力症と診断をされ、不必要な胸腺摘出術や免疫抑制療法を受けている症例も、諸外国同様に日本にも数多く存在すると思われる。先天性筋無力症候群は欠損分子に応じた治療が可能であるタイプが多く、本症候群の診断、及び新規治療法開発研究は、患者利益につながると期待をされる。先天性筋無力症候群の中でも、*COLQ*遺伝子変異による終板AChE欠損症は、従来、全く治療法が存在

しない。わが国の終板AChE欠損症の一例も不幸な転帰を辿っている。本研究では、終板AChE欠損症に対して、臨床応用を近視野に入れたタンパク標的療法の開発研究を行う。本研究にて開発する手法は、神経難治疾患を含む他の細胞外マトリックスタンパク欠損症への応用の可能性がある。

B. 研究方法

Collagen Q（*COLQ*遺伝子産物）は、コラーゲンドメインを介して3分子が3重鎖構造を形成し、acetylcholinesterase (AChE) catalytic subunit（*ACHE*遺伝子産物）12分子と結合し、非対称性A_{12}-AChEを形成する。本研究では、collagen Qがシナプス基底膜への係留シグナルを有する細胞外構造タンパクであることを利用し、*COLQ*欠損モデル動物のTリンパ球に正常*COLQ*遺伝子と正常*ACHE*遺伝子を導入し、A_{12}-AChEを血中に発現させ、シナプス基底膜への係留を試みた。また、骨格筋に*COLQ*遺伝子を導入し、一部の骨格筋に発現したA_{12}-AChEが、血流・組織間液を介して*COLQ*遺伝子非導入骨格筋のシナプス基底膜への係留することを確認した。

I. *ColQ*欠損マウスの評価

ヘテロの*ColQ*欠損マウス（フランスINSERMのDr. Eric Krejci より譲渡）を交配し、そこから産まれた*ColQ*+/+, *ColQ*+/-, *ColQ*-/-マウスの表現型について評価を行う。運動機能を調べるため、ロタ・ロッド試験を行う。4分間で0～40 rpmまで加速回転する棒上で、マウスが乗り続けられる時間を計測する。また、*ColQ* +/+, *ColQ* +/-, *ColQ* -/-マウスの筋肉組織中のAChEの分子形状をショ糖濃度勾配遠心法を用いて調べる。

II. レトロウィルスベクターの構築

レトロウィルスベクターを用いて、モデルマウスにヒト*COLQ*遺伝子を導入、発現させる手法の確立を試みる。レトロウィルスベクターpSIREN-RetroQ (Clontech社)に、リンパ球系細胞で活性が高いEF1 αプロモーターを組み込み、その下流にヒト*ACHE* cDNA・IRES・ヒト*COLQ* cDNAを組み込む。*COLQ*欠損モデル動物のTリンパ球に遺伝子導入を行う。目的組織へのターゲッティングはColQがシナプス基底膜への係留シグナルを有する細胞外構造タンパクであることを利用する。正常ヒト*ACHE*遺伝子と正常ヒト*COLQ*遺伝子を導入する。このレトロウィルスをパッケージング細胞PLAT-Eに導入し、産出させた組み換えレトロウィルスをNIH 3T3細胞に感染させ、A_{12}-AChEの細胞内発現及び細胞外放出を確認する。

III. マウスリンパ球へのレトロウィルスの導入及びマウスへの注入

正常マウスの脾臓、リンパ節よりCD3及びCD28抗体を用いてTリンパ球を単離する。IL2の存在下に培養を行い、上記レトロウィルスを感染させる。次に、細胞培養液中の組み替えA_{12}-AChEをヘパリンアガロースカラムで精製を行い、ショ糖濃度勾配遠心法により分離し、形質転換Tリンパ球が、組み替えA_{12}-AChEを合成し培養液中に放出することを確認する。その後、形質転換Tリンパ

球を正常マウスに注入する。

IV. AAV (adeno-associated virus)を用いた*COLQ*遺伝子導入

AAVベクターは、神経・筋・肝細胞等の非分裂細胞に対し一回の導入で長期的な遺伝子発現が得られ、また安全性の点でも非病原性・低免疫原性など他のウイルスベクターに比べ利点が多い。ここでは筋肉組織への親和性が高いserotype 8を選び用いる。AAVベクターは、pAAV-CMV-MCS (Stratagene社)を用いて、骨格筋において高活性を示すCMV promoterの下流に*COLQ*を挿入したコンストラクトを作製する。AAVベクターをヘルパーベクターpDF6と共にリン酸カルシウム法によりHEK293T細胞にトランスフェクションし、得られたrAAVウィルス粒子をAAVHT1080に感染させ、GFP発現細胞をフローサイトメトリーを用いて測定し、細胞への遺伝子導入効率を調べる。

V. 治療マウスの評価

ヒトAChEに対するモノクローナル抗体(米国Mayo Clinic, Dr. Andrew G. Engelより供与)とヒトColQに対するポリクローナル交代を用いて、治療マウスのシナプス基底膜における組み換えA_{12}-AChEの発現を調べる。また、AChE活性染色を用いて同様の検討を行う。さらに、ショ糖濃度勾配遠心法によるAChE分画により、A_{12}-AChE分子の骨格筋における発現を調べ、微小終板電位の減衰時間を測定することにより神経接合部におけるAChE補充量を類推する。さらに、腎糸球体基底膜や中枢神経系への異所性のColQ発現を調べる。

C. 研究結果

I. レトロウィルスベクターを用いた *COLQ* 遺伝子と *ACHE* 遺伝子の導入

Clontech社のpSIREN-RetroQベクターを改変して、EF1αの下流にヒト*ACHE*遺伝子, IRES,*COLQ*遺伝子を導入し、p-SIREN-EF1 α -*ACHE*-IRES-*COLQ*, p-SIREN-EF1 α -*ACHE*, pSIREN-EF1 α -*COLQ*を構築した。作製したベクターを、PLAT-E細胞にトランスフェクションをし、組換えレトロウィルスを得た。組換えレトロウィルスをNIH3T3細胞に感染させ、タンパクの発現を確認した。*ACHE*-IRES-*COLQ*単独、及び、*ACHE*と*COLQ*の同時感染させたNIH3T3細胞において、その細胞内と培地中に放出されたAChEをショ糖濃度勾配遠心法で分画しAChEのformを調べた。どちらのレトロウィルス感染の場合も、G_1, G_2, G_4-formsに加えA_4, A_8, A_{12}-formsの活性が細胞内だけでなく、培地中でも検出された。感染により細胞内に導入をしたColQとAChEがA_4, A_8, A_{12}-formsの複合体を形成し、細胞外に放出されていることを示している。

II.マウスリンパ球へのレトロウィルスの導入及びマウスへの注入

正常マウスから脾臓・リンパ節を摘出し、血球をCD3及びCD28抗体存在下で培養してTリンパ球の単離を行ない、Tリンパ球の増殖の良好な培養系を確立した。T-cellに組換えレ

トロウィルスを感染させ、２日間培養後、細胞内と培地中のAChE formを調べた。いずれにおいて、A_4, A_8, A_{12}-formsが検出され、組換えTリンパ球により形成されたColQとAChEの複合体が細胞外に移行することを確認した。

現在、大量のTリンパ球の単離、及びレトロウィルスの生産を行い、マウス生体内に注入できる量の遺伝子導入Tリンパ球を作製している。今後、この組換えTリンパ球を正常マウスに静脈注射する。ヒトAChEに対するモノクローナル抗体を用いて、組み換えA_{12}-AChEの血中での発現を調べ、さらに、シナプス基底膜への係留の有無を調べる。良好な結果が得られた後、*ColQ-/-*マウスに適用し、非対象性A_{12}-AChE分子の発現、及び、神経接合部でのA_{12}-AChE分子の集積、生理学的・形態学的検査を行い、筋無力症状の機能回復・安全性について検討する。

III. AAV (adeno-associated virus)を用いた*COLQ*遺伝子導入

AAVベクターはpAAV-CMV-*COLQ*と、ColQと共にGFPを発現するpAAV-CMV-*COLQ*-IRES-*EGFP*のコンストラクトを作成した。AAV8型カプシドをコードするヘルパープラスミドと同時にHEK293細胞に導入し、rAAV8-ColQとrAAV8-ColQ-EGFPの２種類の組み替えAAVを作製した。強イオン膜を用いて精製した後、*ColQ-/-*マウスの尾静脈より全身投与（2×10^{11} genome copies）を行った。

投与2週間後から、ロタロッドでの運動テストに改善効果が見られ、４週間後には、運動テストと疲労テストで正常マウスと同成績まで回復した。治療マウスの骨格筋組織切片のAChE活性染色にて、αバンガロトキシンで染まるアセチルコリンレセプターと同一の部位に、AChE活性が見られた。また、ColQ抗体による免疫染色においても、同じ部位でシグナルが検出され、神経筋接合部においてColQが集積していることが確認できた。さらに、ショ糖勾配遠心法により、治療マウスの骨格筋中のAChE-formを検出したところ、未治療群では見られなかったA_4, A_8, A_{12}-formsのAChEが検出された。さらに、治療マウスの微小終板電位の減衰時間は正常マウスとノックアウトマウスの中間の値となり、A_{12}-AChEの発現は正常量までは回復をしていないと予想された。しかし、治療マウスの運動機能がほぼ正常化したことは、神経筋信号伝達の安全域を超える程度までA_{12}-AChEの発現量が回復をしたと考えられた。以上の結果から、AAV8を介して*COLQ*の遺伝子導入により、発現したColQがAChEと複合体を形成し神経筋接合部に係留すること、また治療マウスは、運動機能テストから筋無力症状が正常値まで回復したことが明らかになった。ColQは基底膜に係留シグナルを持つ細胞外タンパクであるため、遺伝子導入された骨格筋細胞の割合が低くても、自らの細胞外標的シグナルにより、十分量のA_{12}-AChEが神経筋接合部に移行をした考えられた。治療後最高３ヶ月間の観察では、改善した運動機能の再増悪もなく、また、神経筋接合部へのリンパ球浸潤もなく、良好な経過を辿っている。今後、筋無力症状の機能回復が可能なウィルス投与量の最低値の決定、組換えA_{12}-AChEの腎糸球体基底膜などへの異所性集積や抗collagen Q抗体の出現など、遺伝子導入の安全性を検討して行く予定である。

D. 考察

平成19年度は、レトロウィルスのTリンパ球への導入と、組換えA_{12}-AChEのTリンパ球からの放出を確認した。大量の遺伝子組換えTリンパ球の調整が完了していないために、モデル動物を用いた治療には至らなかった。

骨格筋への親和性を有するAAV8を用いた研究では、モデル動物の治療に成功した。5週齢のマウスに治療を行い、運動機能がほぼ正常化できたことは、本プロジェクトで提案をしてきた「タンパク自体の組織親和性を用いたタンパク標的療法」の有用性を示唆していると期待される。

E. 結論

タンパク標的療法は従来の神経筋疾患モデル動物の遺伝子治療ではみられない程度の良好な運動能力改善効果を示した。今後、本手法の臨床応用や他疾患への応用を含めて研究を行う。

F. 健康危険情報

特記事項なし。

G. 研究発表

1. 論文発表

1. Sahashi K, Masuda A, Matsuura T, Shinmi J, Zhang Z, Takeshima Y, Matsuo M, Sobue G, Ohno K. In vitro and in silico analysis reveals an efficient algorithm to predict the splicing consequences of mutations at the 5' splice sites. *Nucleic Acids Res* 2007, 35:5995-6003.
2. Ichihara M, Murakumo Y, Masuda A, Matsuura T, Asai N, Jijiwa M, Ishida M, Shinmi J, Yatsuya H, Qiao S, Takahashi M, Ohno K. Thermodynamic instability of siRNA duplex is a prerequisite for dependable prediction of siRNA activities. *Nucleic Acids Res* 2007, 35:e123.
3. Saito T, Amakusa,Y, Kimura T, Yahara O, Aizawa H, Ikeda Y, Day JW, Ranum LPW, Ohno K, Matsuura T. Myotonic dystrophy type 2 in Japan: ancestral origin distinct from Caucasian families. *Neurogenetics* 2008, 9:61-63.
4. Ito M, Masuda A, Jinno S, Katagiri T, Krejci E, Ohno K. Viral vector-medicated expression of human collagen Q in cultured cells. *Chemico-Biological Interactions* 2008, in press.
5. Kurosaki T, Matsuura T, Ohno K, Ueda S. Long-range PCR for the diagnosis of spinocerebellar ataxia type 10. *Neurogenetics* 2008, in press.
6. Gao K, Masuda A, Matsuura T, Ohno K. Human branch point consensus sequence is yUnAy. *Nucleic Acids Res* 2008, in press.
7. Shen X-M, Fukuda T, Ohno K, Sine SM, Engel AG. Novel AChR δ subunit mutation causing myasthenia hinders intersubunit link essential for channel gating. *J Clin Invest* 2008, in

press

2. 学会発表

1. Ito M, Masuda A, Jinno S, Katagiri T, Krejci E, Ohno K. Gene Therapy for Collagen Q defects in congenital myasthenic syndromes. The IXth International Meeting on Cholinesterases, Shuzou, China. May 10, 2007
2. Amakusa Y, Matsuura T, Saito T, Kimura T, Yahara O, Aizawa H, Ikeda Y, Day JW, Ranum LPW, Ohno K. Myotonic dystrophy type 2 in Japan: distinct ancestral origin from Caucasian families. 57th Annual Meeting of the American Society of Human Genetics, San Diego, California, USA. Oct 23-27, 2007
3. Ohno K, Ichihara M. Unstable siRNA duplex is a prerequisite for accurate prediction of siRNA efficiency – Proposal of a new parameter based on the linear regression model –. 57th Annual Meeting of the American Society of Human Genetics, San Diego, California, USA. Oct 23-27, 2007
4. Yamashita Y, Matsuura T, Shinmi J, Ibi T, Kinoshita M, Kimura T, Yahara O, Sahashi K, Ohno K. Comprehensive analysis of aberrantly spliced exons in myotonic dystrophy type 1 using Affymetrix Exon Array. 57th Annual Meeting of the American Society of Human Genetics, San Diego, California, USA. Oct 23-27, 2007

H. 知的財産権の出願・登録状況

特記事項なし。

II. わが国における先天性筋無力症候群未診断症例の病態・治療研究

分担研究者　祖父江　元　名古屋大学医学系研究科・神経内科学・教授

研究要旨

先天性筋無力症候群は、神経筋接合部の先天的分子欠損により、顔面・四肢・体幹の筋力低下・筋萎縮・奇形を主徴とする疾患群である。世界中からおよそ200例が報告されており、7種類の神経筋接合部分子において約180種類の遺伝子変異が同定されてきている。これらのうちファウンダー効果が明らかなものは*RAPSN*遺伝子における1変異のみであり、他の多くの遺伝子変異は個々の家系に特有の遺伝子変異か、de novo遺伝子変異であり、人種や地域を問わず本症候群が存在すると想定されるが、わが国からの報告は極めて稀である。本研究では、本症候群の診断に反復神経刺激が有用であることを利用し、筋力低下を主徴とする非定型的な神経筋疾患に対して積極的に反復神経刺激を行い、スローチャンネル症候群の一例を診断し、その治療効果について検討を行った。

A. 研究目的

先天性筋無力症候群は、神経筋接合部の先天的分子欠損症が原因であり、同定されてきた欠損分子には、(1) 神経終末からhigh affinity choline transporterを介して取り込まれたcholineからacetylcholineを再合成するcholine acetyltanferase、(2) acetylcholinesterase catalytic subunitとともにasymmetric acetylcholineを作り、synaptic basal laminaにacetylcholinesteraseをanchoringをさせるcollagen Q、(3) muscle nicotinic acetylcholine receptor (AChR)を形成するAChR subunits、(4) AChRを神経筋終板に集積させるrapsyn、(5) AChRによる脱分極を筋膜全般に伝播するmuscle voltage-gated sodium channel、(6) 神経終末より放出されるneural agrinのレセプターであり、AChRとrapsynの結合を促進し、endplateへのAChRの集積をドライブするMuSK (muscle specific kinase)、(7) MuSKシグナル伝達系の分子でMuSKやrapsynと同じく終板におけるAChR集積に関わるDok-7があげられる。スローチャンネル症候群は、AChRイオンチャンネルの開口時間が異常延長する病態である。先天性筋無力症候群は、スローチャンネル症候群のみが常染色体優性遺伝であり、他はいずれも常染色体劣性遺伝である。本症候群のなかで、ファウンダー効果が知られている遺伝子変異は、rapsynのN88K変異のみであり、大多数は、個々の家系に特有の変異か、de novo変異である。日本では先天性筋無力症候群と診断された症例は10例以下であり、多くは未診断であったり、誤診のもとに間違った治療を受けている可能性がある。本症候群は重症筋無力症と異なる臨床症状を呈することも診断を困難にしていると思われる。本症候群では、胎生期からの神経筋接合部信号伝達障害があることから顔面・頭蓋骨奇形や骨格筋低形成が認められことがある。また、日内変動が明らかではなく、複視を伴わない眼球運動障害も多く認められる。本症候群の診断には反復神経刺激が有用であることを利用し、筋力低下を主徴とする非定型的な神経筋疾患に対して積極的

に反復神経刺激を行い、スローチャンネル症候群の一例の診断を行なった。AChR β サブユニットに遺伝子変異を同定し、quinidineとfluoxetineによる治療を試みた。

B. 研究方法

筋力低下を主徴とする非定型的な神経筋疾患に対して、低頻度および高頻度反復神経刺激による複合筋活動電位(CMAP)減衰の計測を行い、同時に、単発神経刺激に対する反復CMAPの出現をモニターした。反復CMAPの出現および臨床症状よりスローチャンネル症候群を疑いAChRサブユニット遺伝子変異を同定した。発現実験を行いスローチャンネル症候群であることを確認した。さらに、quinidineとfluoxetineによる治療を試みた。

C. 研究結果

症例は、36歳発症の軽度四肢近位筋力低下と労作性呼吸困難を主訴とする37歳男性である。幼少時からの斜視があるが複視を訴えない。診察時に易疲労性を認めず、日内変動もなく、edrophoniumテストも陰性であり、重症筋無力症を含む神経筋接合部疾患を積極的に疑う臨床所見に欠けている。単発神経刺激にて2発の反復CMAPを認めた。反復CMAPは、(1) スローチャンネル症候群、(2) 終板acetylchoinesterase欠損症、(3) 抗cholinesterase剤や有機リン中毒によるacetylcholinesterase活性抑制で認められる。脛骨神経の3 Hzの反復神経刺激にて22%の異常減衰を認め、 single fiber EMGではMCDが98 ± 54μsと延長しており、神経筋接合部信号伝達の異常が示唆された。

AChRサブユニット遺伝子変異によるスローチャンネル症候群と、collagen Q遺伝子変異による終板acetylcholinesterase欠損症を疑い、網羅的な遺伝子変異の検索を行ったところ、AChR βサブユニットにV296A変異を認めた。患者は正常alleleと変異alleleを持つheterozygoteであり、βV296Aはドミナント変異であった。

βV296は第3膜貫通ドメインほぼ中央部に位置し、αサブユニットで対応するアミノ酸はαV285である。イオンチャンネルの開口時間が異常に短縮するファーストチャンネル症候群においてαV285I変異が報告されている(Wang H-L, Milone M, Ohno K, et al. *Nat Neurosci* 2: 226, 1999)。αV285I変異の研究において、αV285L、αV285I、αV285T、αV285Aの4種類の変異αサブユニットの解析が行われており、コドン285におけるアミノ酸側鎖のボリュームが、イオンチャンネルの動態を決定していることが判明している。つまり、valineよりもボリュームが大きいleucineやisoleucineではファーストチャンネルとなり、valineよりもボリュームが小さいthreonineやalanineではスローチャンネルとなる。この事実は、第3膜貫通ドメインほぼ中央部のコドン285におけるアミノ酸側鎖が、AChRのチャンネル孔を形成する第2膜貫通ドメインを背側から押していることを示している。AChR α, β, δ, εサブユニットは相同なサブユニットであり、第3膜貫通ドメインのβV296Aは第2膜貫通ドメインを背側から押す力を弱め、AChRのチャンネル孔を開きやすくさせるスローチャンネル変異と想定され、事実、単一チャンネル記録にてチャンネル開口が異常に遅延していることを実験的に証明した。

イオンチャンネルブロッカであり、スローチャンネル症候群に対する有効性が確立しているquinidine常用量(400 mg/day)を投与したところ、四肢筋力が改善し、反復神経刺激に対するCMAPの異常減衰も軽減し、単発神経刺激に対する反復CMAPも軽減した。しかし、寒冷誘発性のアセチルコリン放出効率の低下とAChR開口効率の低下によると思われる寒冷時の四肢筋力低下と呼吸不全を認めたため、fluoxetine常用量(40 mg/day)を追加し、さらなる四肢筋力の改善を認めた。

D. 考察

今回解析を行った症例に見られるように先天性筋無力症候群の中には、成人発症で、日内変動を伴わず、易疲労性も明らかでない症例が多く存在すると思われる。本症例も、呼吸筋が優位に侵され、明らかな易疲労性を認めず、鑑別診断として重症筋無力症や先天性筋無力症候群を考えにくい臨床像であった。本症例でも、単発神経刺激による反復CMAPの出現と、反復神経刺激におけるCMAP電位の異常減衰が診断に有用であり、筋力低下を主徴とする神経疾患に対して電気生理学手法を用いた神経筋接合部欠損の検索を行う重要性が示唆された。

E. 結論

先天性筋無力症候群は、スローチャンネル症候群をはじめとして病態に応じた治療方法が可能な疾患群が多く、本症候群に対して積極的に電気生理学診断を行う重要性が示された。

F. 健康危険情報

特記事項なし。

G. 研究発表

1. 論文発表

1. Adachi H, Waza M, Tokui K, Katsuno M, Minamiyama M, Tanaka F, Doyu M, Sobue G. CHIP overexpression reduces mutant androgen receptor protein and ameliorates phenotypes of the spinal and bulbar muscular atrophy transgenic mouse model. *J Neurosci* 27: 5115-5126, 2007.
2. Niwa J, Yamada S, Ishigaki S, Sone J, Takahashi M, Katsuno M, Tanaka F, Doyu M, Sobue G. Disulfide bond mediates aggregation, toxicity, and ubiquitylation of familial amyotrophic lateral sclerosis-linked mutant SOD1. *J Biol Chem* 282: 28087-28095, 2007.
3. Oki Y, Koike H, Iijima M, Mori K, Hattori N, Katsuno M, Nakamura T, Hirayama M, Tanaka F, Shiraishi M, Yazaki S, Nokura K, Yamamoto H, Sobue G. Ataxic vs painful form of paraneoplastic neuropathy. *Neurology* 69: 564-572, 2007.
4. Suzuki K, Katsuno M, Banno H, Takeuchi Y, Atsuta N, Ito M, Watanabe H, Yamashita F, Hori N, Nakamura T, Hirayama M, Tanaka F, Sobue G. CAG repeat size correlates to electrophysiological motor and sensory phenotypes in SBMA. *Brain* 131: 229-239, 2008.

H. 知的財産権の出願・登録状況

特記事項なし。

III. 研究成果の刊行に関する一覧表

発表者名	論文タイトル名	発表誌名	巻号	ページ	出版年
Sahashi K, Masuda A, Matsuura T, Shinmi J, Zhang Z, Takeshima Y, Matsuo M, Sobue G, Ohno K	In vitro and in silico analysis reveals an efficient algorithm to predict the splicing consequences of mutations at the 5' splice sites	*Nucleic Acids Res*	35	5995-6003	2007
Ichihara M, Murakumo Y, Masuda A, Matsuura T, Asai N, Jijiwa M, Ishida M, Shinmi J, Yatsuya H, Qiao S, Takahashi M, Ohno K	Thermodynamic instability of siRNA duplex is a prerequisite for dependable prediction of siRNA activities	*Nucleic Acids Res*	35	e123	2007
Adachi H, Waza M, Tokui K, Katsuno M, Minamiyama M, Tanaka F, Doyu M, Sobue G	CHIP overexpression reduces mutant androgen receptor protein and ameliorates phenotypes of the spinal and bulbar muscular atrophy transgenic mouse model	*J Neurosci*	27	5115-5126	2007
Niwa J, Yamada S, Ishigaki S, Sone J, Takahashi M, Katsuno M, Tanaka F, Doyu M, Sobue G	Disulfide bond mediates aggregation, toxicity, and ubiquitylation of familial amyotrophic lateral sclerosis-linked mutant SOD1	*J Biol Chem*	282	28087-28095	2007
Oki Y, Koike H, Iijima M, Mori K, Hattori N, Katsuno M, Nakamura T, Hirayama M, Tanaka F, Shiraishi M, Yazaki S, Nokura K, Yamamoto H, Sobue G	Ataxic vs painful form of paraneoplastic neuropathy	*Neurology*	69	564-572	2007
Saito T, Amakusa,Y, Kimura T, Yahara O, Aizawa H, Ikeda Y, Day JW, Ranum LPW, Ohno K, Matsuura T	Myotonic dystrophy type 2 in Japan: ancestral origin distinct from Caucasian families	*Neurogenetics*	9	61-63	2008
Suzuki K, Katsuno M, Banno	CAG repeat size correlates to electrophysio-	*Brain*	131	229-239	2008

H, Takeuchi Y, Atsuta N, Ito M, Watanabe H, Yamashita F, Hori N, Nakamura T, Hirayama M, Tanaka F, Sobue G	logical motor and sensory phenotypes in SBMA				
Ito M, Masuda A, Jinno S, Katagiri T, Krejci E, Ohno K	Viral vector-medicated expression of human collagen Q in cultured cells	*Chemico-Biological Interactions*		in press	2008
Kurosaki T, Matsuura T, Ohno K, Ueda S	Long-range PCR for the diagnosis of spinocerebellar ataxia type 10	*Neurogenetics*		in press	2008
Gao K, Masuda A, Matsuura T, Ohno K	Human branch point consensus sequence is yUnAy	*Nucleic Acids Res*		in press	2008
Shen X-M, Fukuda T, Ohno K, Sine SM, Engel AG	Novel AChR δ subunit mutation causing myasthenia hinders intersubunit link essential for channel gating	*J Clin Invest*		in press	2008

Published online 28 August 2007

Nucleic Acids Research, 2007, Vol. 35, No. 18 ***5995–6003***
doi:10.1093/nar/gkm647

In vitro and *in silico* analysis reveals an efficient algorithm to predict the splicing consequences of mutations at the 5′ splice sites

Kentaro Sahashi[1,2], Akio Masuda[1], Tohru Matsuura[1], Jun Shinmi[1], Zhujun Zhang[3], Yasuhiro Takeshima[3], Masafumi Matsuo[3], Gen Sobue[2] and Kinji Ohno[1,*]

[1]Division of Neurogenetics and Bioinformatics, Center for Neurological Diseases and Cancer, Nagoya University Graduate School of Medicine, [2]Department of Neurology, Nagoya University Graduate School of Medicine, Nagoya and [3]Department of Pediatrics, Kobe University Graduate School of Medicine, Kobe, Japan

Received April 5, 2007; Revised August 6, 2007; Accepted August 6, 2007

ABSTRACT

We have found that two previously reported exonic mutations in the *PINK1* and *PARK7* genes affect pre-mRNA splicing. To develop an algorithm to predict underestimated splicing consequences of exonic mutations at the 5′ splice site, we constructed and analyzed 31 minigenes carrying exonic splicing mutations and their derivatives. We also examined 189 249 U2-dependent 5′ splice sites of the entire human genome and found that a new variable, the SD-Score, which represents a common logarithm of the frequency of a specific 5′ splice site, efficiently predicts the splicing consequences of these minigenes. We also employed the information contents (R_i) to improve the prediction accuracy. We validated our algorithm by analyzing 32 additional minigenes as well as 179 previously reported splicing mutations. The SD-Score algorithm predicted aberrant splicings in 198 of 204 sites (sensitivity = 97.1%) and normal splicings in 36 of 38 sites (specificity = 94.7%). Simulation of all possible exonic mutations at positions −3, −2 and −1 of the 189 249 sites predicts that 37.8, 88.8 and 96.8% of these mutations would affect pre-mRNA splicing, respectively. We propose that the SD-Score algorithm is a practical tool to predict splicing consequences of mutations affecting the 5′ splice site.

INTRODUCTION

In eukaryotes, splicing of the nuclear mRNA precursor (pre-mRNA) takes place mostly within the U2-dependent spliceosome, a complex of five uridine-rich small nuclear (sn) ribonucleoproteins (RNPs): U1, U2, U4, U5 and U6 snRNPs and numerous non-snRNP proteins. In the first step of spliceosome formation, U1 snRNP recognizes the 5′ splice site and regulates initiation of pre-mRNA splicing (1).

The 5′ splice site is composed of the last three nucleotides of an exon (positions −3, −2 and −1) and the first six nucleotides of an intron (positions +1 to +6). The consensus sequence of the U2-dependent 5′ splice sites is (C/A)AG|GT(A/G)AGT (2), where the vertical line (|) represents the exon–intron boundary, and the 'GT' dinucleotide at the 5′ end of an intron is invariable (Figure 1A) (3). In the latter stage of pre-mRNA splicing, U1 snRNA dissociates from the 5′ splice site, and U6 snRNA subsequently binds to nucleotides at positions +2, +5 and +6 (Figure 2A) (4–6).

In the course of identification of exonic splicing mutations in genetic forms of Parkinson's disease, we found two splicing mutations at the 5′ splice site that compromise binding to U1 snRNA. To clarify how exonic mutations at the 5′ splice site cause aberrant splicing, we analyzed 31 minigenes *in vitro* and examined 189 249 putative U2-dependent 5′ GT splice sites of the entire human genome *in silico*. We found that a new variable, the SD-Score, in combination with the information contents (R_i), which represents the amount of information in bits (7,8), can efficiently predict the splicing consequences of exonic mutations at the 5′ splice site. We validated our algorithm with 32 additional minigenes and with 179 previously reported splicing mutations, and found that the SD-Score algorithm has a sensitivity of 97.1% and a specificity of 94.7%. We believe that the SD-Score algorithm is a practical tool for predicting the splicing consequences at the 5′ splice site of mutations causing human disease.

*To whom correspondence should be addressed. Tel: +81-52-744-2446; Fax: +81-52-744-2449; Email: ohnok@med.nagoya-u.ac.jp

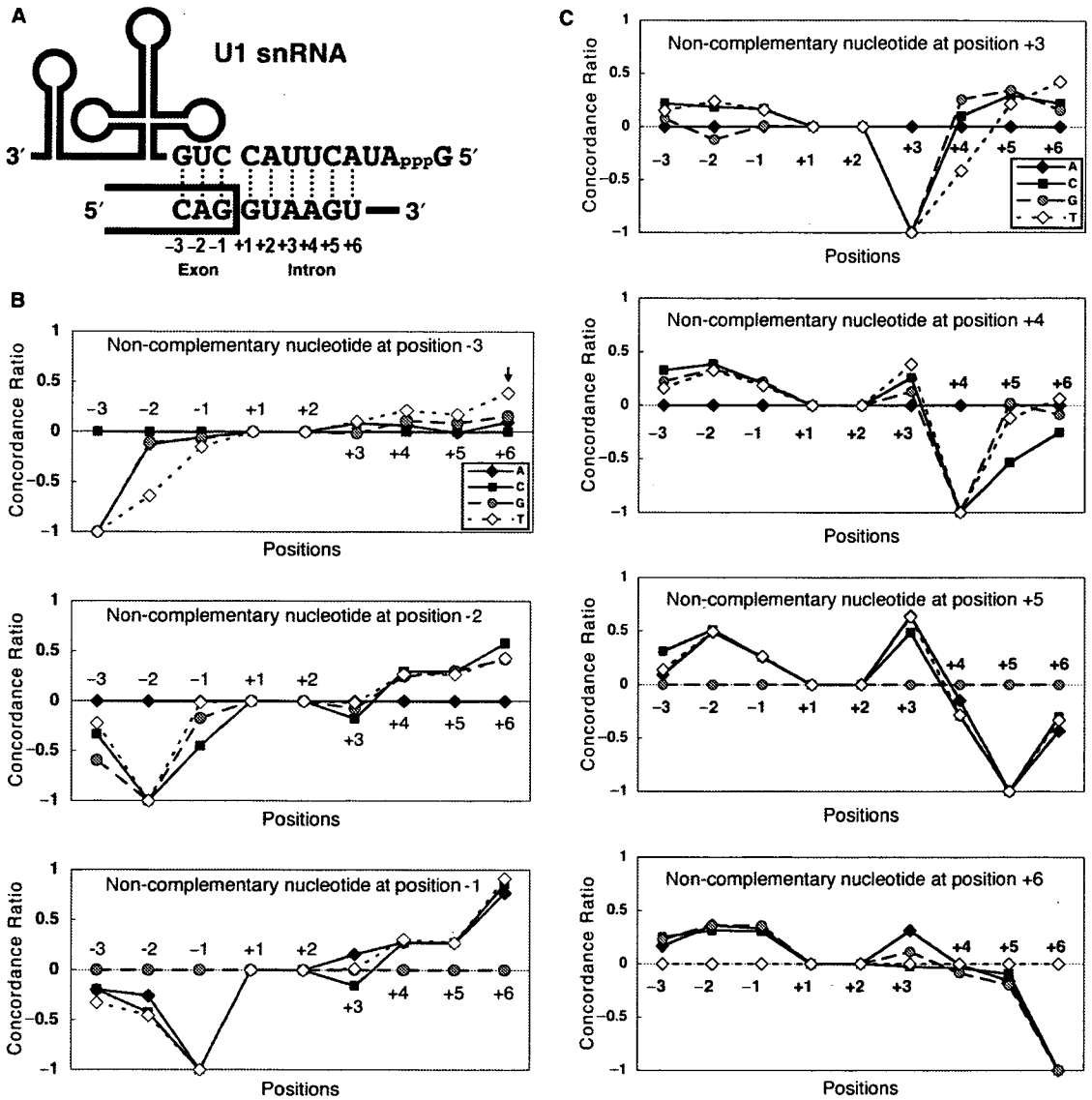

Figure 1. (**A**) The 5′ end of U1 snRNA recognizes three nucleotides at exonic positions −3, −2 and −1 and six nucleotides at intronic positions +1 to +6. Note that the consensus sequence, 5′-(C/A)AG|GT(A/G)AGT-3′, is complementary to U1 snRNA at most positions. (**B** and **C**) The 'two-point' analysis reveals that NCp-nucs at exonic positions −3, −2 and −1 are compensated for by Cp-nucs at positions +4, +5 and +6, and that NCp-nucs at intronic positions +4, +5 and +6 are compensated for by Cp-nucs at positions −3, −2, −1 and +3. For example, when a complementary C is used at position −3 (68 353 sites), the frequency of a complementary T at position +6 is 42.9% (29 307 of 68 353). In contrast, when a noncomplementary T is used at position −3 (22 667 sites), the frequency of a complementary T at position +6 is increased to 60.2% (13 636 of 22 667). The concordance ratio is calculated as (60.2–42.9)/42.9 = 0.403 (arrow). This means that when position −3 is a noncomplementary T, we observe a complementary T at position +6 40.3% more frequently than when position −3 is a complementary C. A positive concordance ratio at a specific position indicates that a Cp-nuc to U1 snRNA is preferentially used to compensate for an NCp-nuc at another position.

MATERIALS AND METHODS

Exon trapping vector

To examine splicing consequences of exonic mutations, we constructed minigenes in an exon-trapping vector, **pSPL3** (a discontinued product of Invitrogen, Carlsbad, CA, USA), which was kindly provided by Dr Kazunori Imaizumi, Department of Anatomy, University of Miyazaki, Miyazaki, Japan. We introduced a cytomegalovirus promoter in place of the simian virus 40 promoter by means of a megaprimer-based, site-directed mutagenesis method (9) and also introduced a PacI recognition sequence at the multiple cloning site. Because the nonsense-mediated mRNA decay can degrade a specific splicing product with a premature termination codon and cause misinterpretation of splicing assays (10), we eliminated a premature stop codon and constructed **pSPL3** vectors with three reading frames to insert any chimeric exons in-frame (Supplementary Figure 1).

Minigene constructs and mutagenesis

We used the polymerase chain reaction (PCR) to amplify an exon and the flanking introns of 200 bp of the genes of our interest using normal human genomic DNA extracted

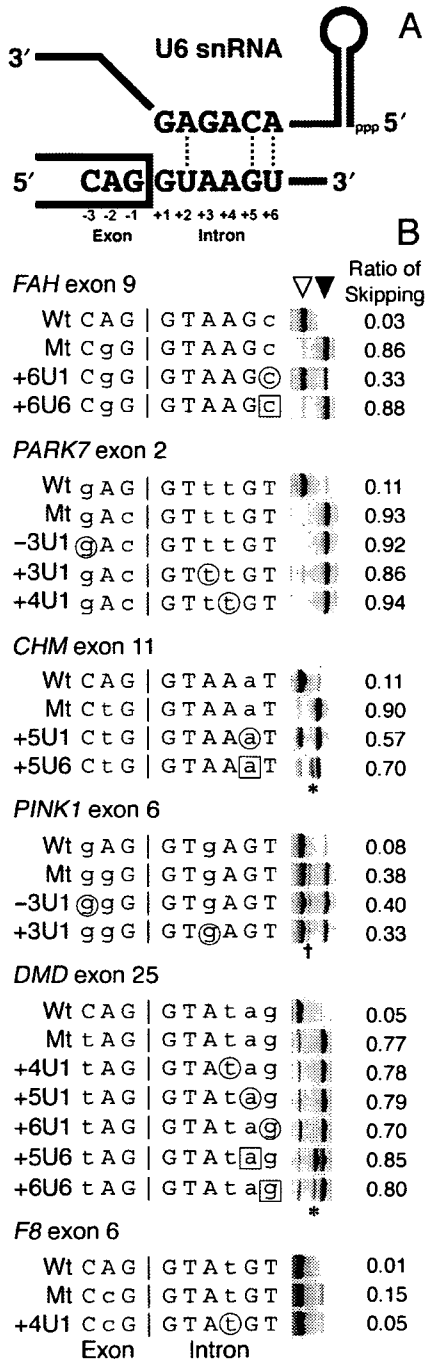

Figure 2. (**A**) The U6 snRNA base pairs with nucleotides at positions +2, +5 and +6. (**B**) RT–PCR analysis of minigene constructs transfected into HEK293 cells with artificial U1 or U6 snRNA. A single Cp-nuc is introduced into U1 or U6 snRNA while retaining the mismatch at the mutation. Wt, wild-type construct; Mt, mutation observed in a patient. For example, +6U1 indicates that a nucleotide on U1 snRNA corresponding to position +6 is substituted to match the 5′ splice site. Circles and squares represent nucleotides that become complementary to the artificial U1 and U6 snRNAs, respectively. The open arrowhead indicates a normally spliced fragment, whereas the closed arrowhead indicates an exon-skipped fragment. The rightmost column shows the densitometric ratio of the exon-skipped fragment. The asterisk indicates a mixture of fragments due to activated 5′ splice sites four and 13 nucleotides downstream of the native 5′ splice site at the 5′ exon of pSPL3. The dagger indicates a heteroduplex formed by normally spliced and exon-skipped products. Uppercase nucleotides are complementary to U1 snRNA, whereas lowercase nucleotides are not.

from HEK293 cells. NotI and PacI recognition sites were introduced to the 5′ and 3′ ends, respectively, of the PCR product. Each amplicon was inserted into one of the three **pSPL3** vectors so that the reading frame of the chimeric exon was retained.

For the *PARK7* and *DYSF* genes, wild-type pSPL3 constructs yielded a large proportion of exon-skipped products. We thus amplified a genomic segment spanning the mutation-harboring exon, the flanking introns and the neighboring exons, and then inserted the amplicon into the **pcDNA3.1(+)** mammalian expression vector (Invitrogen). Different splicing consequences between **pSPL3** and **pcDNA3.1(+)** constructs likely represent the complexity of splicing analysis.

The U1 snRNA gene with its own promoter was kindly provided by Dr Alan M. Weiner, Department of Biochemistry, University of Washington, Seattle, WA, USA. The U6 snRNA gene with the 5′ promoter region of 367 bp and a 3′ end region of 149 bp was amplified using normal human genomic DNA extracted from HEK293 cells and was inserted into the **pGEM-T Easy** Vector (Promega, Madison, WI, USA).

Naturally occurring and artificial mutations were introduced into the inserts with the QuikChange Site-Directed Mutagenesis Kit (Stratagene, La Jolla, CA, USA). We confirmed by sequencing that there were no artifacts in any insert.

Transfection and RNA analysis

HEK293 cells were maintained in the Dulbecco's minimum essential medium (DMEM; Sigma–Aldrich, St Louis, MO, USA) with 10% fetal bovine serum (FBS; Sigma–Aldrich). At ~50% confluency (~5×10^5 cells) in a 6-well plate, 1 ml of fresh Opti-MEM I (Invitrogen) was substituted for DMEM, and 1 μg of a minigene with 3 μl of the FuGENE6 Transfection Reagent (Roche Diagnostics, Indianapolis, IN, USA) were then added. After 4 h, 2 ml of DMEM with 10% FBS was overlaid, and the cells were incubated overnight. The transfection medium was replaced with 2 ml of fresh DMEM with 10% FBS, and the transfected cells were incubated for 48 h before RNA extraction. When artificial U1 or U6 snRNA vector was used, 50 ng of a minigene and 950 ng of each snRNA vector were introduced. Total RNA was extracted using the GenElute Mammalian Total RNA Kit (Sigma–Aldrich). DNA was degraded on-column with the DNase I (Qiagen, Valencia, CA, USA). Twenty percent of the isolated RNA was used as a template for cDNA synthesis with the Oligo(dT)$^{12-18}$ primer (Invitrogen) and the SuperScript II Reverse Transcriptase (Invitrogen). Ten percent of the synthesized cDNA was used as a template for reverse transcriptase (RT)–PCR amplification with primers SD6 (5′-TCTGAGTCACCTGGACAACC-3′) and SA2 (5′-GTGAACTGCACTGTGACAAGCTGC-3′), both of which were on the **pSPL3** vector. For minigenes in **pcDNA3.1(+)**, gene-specific primers were employed. Amplification was performed for 30 to 35 cycles of denaturation at 94°C for 20 s, annealing at 52°C for 20 s and extension at 72°C for 45 s. We measured the signal

intensities of the normal and aberrant fragments with the NIH Image 1.63 program. When the ratio of the aberrant product of the mutant construct was increased by 2.5-fold compared with that of the wild-type construct, we considered the mutant construct to have resulted in aberrant splicing. We tried several different thresholds and found that the threshold of 2.5-fold best represents the results of our visual inspections (data not shown).

For the *PINK1*, *CHM*, *BRCA1*, *DYSF*, *F8* and *DMD* genes, we cloned and sequenced all RT–PCR fragments to confirm that the expected normal and aberrant splicings indeed had taken place in these minigenes.

In silico analysis

We extracted all the nonredundant 5′ GT splice sites in the entire human genome using the CDS tags in the NCBI RefSeq Database Build 36.2. Each 5′ splice site on the genome is counted once, even if it is used multiple times in alternatively spliced transcripts. The analysis was performed with the PrimePower HPC2500/Solaris 9 supercomputer (Fujitsu Ltd., Tokyo, Japan). Using the JMP-IN Ver. 5.1.2 software (SAS Institute, Cary, NC, USA), we statistically determined a threshold for each variable using the default settings.

In humans, ~0.1–0.3% of introns are spliced by the minor U12-dependent spliceosome (2,11,12), and ~70% of the U12-dependent introns have GT-AG terminal dinucleotides (13). Previous *in silico* analyses of the human genome identified 275 (12), 469 (14) and 487 (13) GT-AG U12-dependent introns. We thus eliminated 487 U12-dependent 5′ GT splice sites from our analysis, according to the U12 Intron Database (http://genome.imim.es/cgi-bin/u12db/u12db.cgi). Our training and validation data sets (see 'Results' section) did not include any of the known U12-dependent splice sites.

RESULTS

Screening of exonic splicing mutations in genes causing Parkinson's disease

To identify exonic splicing mutations in genetic forms of Parkinson's disease, we analyzed 57 missense, nonsense and synonymous mutations deposited in the Human Gene Mutation Database (http://www.hgmd.cf.ac.uk/) (15) in the *SNCA*, *PARK2*, *PINK1* and *PARK7* genes using minigenes (Supplementary Table 1). We found that no mutation affected an exonic splicing enhancer (ESE) or silencer (ESS). Two mutations at the 5′ splice site, however, resulted in skipping of the mutation-harboring exon likely by compromising binding to U1 snRNA. One mutation was E417G in *PINK1* (16) and the other mutation was E64D in *PARK7* (17). To understand how complementary nucleotides (Cp-nucs) to U1 snRNA at the 5′ splice site compensate for noncomplementary nucleotides (NCp-nucs) at other positions, we introduced a single Cp-nuc to the mutant *PINK1* and *PARK7* minigenes while retaining the mutation (Supplementary Figure 2A and Supplementary Table 2). We employed these results to develop an algorithm to predict splicing consequences of mutations at the 5′ splice site.

Recapitulation of aberrant splicings due to previously reported exonic splicing mutations

We next recapitulated aberrant splicings of six previously reported exonic splicing mutations at position −2 in the *CHM*, *FAH*, *HMBS*, *UROS*, *BRCA1* and *CYP27A1* genes (Supplementary Figure 2B and Supplementary Table 2). All mutations except for *CYP27A1* caused the mutation-harboring exon to be skipped. The *CYP27A1* mutation is exceptional because it introduces a Cp-nuc rather than disrupting complementarity (18). We similarly introduced a Cp-nuc to the mutant constructs while retaining the mutation. These results were also used for developing a prediction algorithm of splicing mutations.

Site-directed mutagenesis of a single nucleotide of U1 snRNA and U6 snRNA

Because the 5′ splice site is recognized by both U1 and U6 snRNAs at different stages of pre-mRNA splicing (Figures 1A and 2A), we wondered which nucleotide of the 5′ splice site is most important for binding each snRNA. To this end, we introduced a single Cp-nuc to U1 or U6 snRNA while retaining the mismatch between U1 or U6 snRNA and the mutation.

Among 11 experiments with artificial U1 snRNAs, corrections of U1 snRNA corresponding to position +6 in *FAH* and to +5 in *CHM* ameliorated aberrant splicings, while the others were inefficient (Figure 2B). Among four experiments with artificial U6 snRNAs, only a correction corresponding to position +5 in *CHM* partially normalized aberrant splicing (Figure 2B). In contrast to manipulation of the splicing *cis*-elements, introduced artificial snRNAs are competed by endogenous snRNAs, and hence their effects tend to be compromised. In addition, substitution of these nucleotides might have modified the core secondary structure of U1 or U6 snRNA and made it nonfunctional. Nevertheless, it is interesting to note that corrections of U1 and U6 snRNAs ameliorate aberrant splicings in some mutants even in the presence of mismatch at the mutation. To our knowledge, no similar study has been performed in this scale (19), but the study size was still too small to draw a definite conclusion.

In silico analysis of human 5′ splice sites: the consensus sequence

To examine how the human 5′ splice sites are organized and why the identified exonic mutations resulted in aberrant splicings, we analyzed the 5′ splice sites of the entire human genome. According to the NCBI RefSeq Database, the human genome comprises 28 714 annotated genes with 192 643 5′ splice sites. Of these sites, 189 718 (98.5%) sites carry an invariant GT dinucleotide at positions +1 and +2, whereas 1859 (1.0%) and 311 (0.2%) sites have GC and AT dinucleotides, respectively. The remaining 755 (0.4%) sites carry other dinucleotides and likely include erroneous annotations. We excluded 487 U12-dependent 5′ GT splice sites (see 'Materials and Methods' section) and extracted nine nucleotide segments

Table 1. Nucleotide frequencies (%) at U2-dependent 189 249 human 5′ GT splice sites

Position	−3	−2	−1	+1	+2	+3	+4	+5	+6
A	33.4	63.5	10.0			59.5	69.4	8.9	17.9
C	36.1	10.9	2.8			2.9	7.7	5.7	15.0
G	18.5	11.6	80.3	100.0		34.6	11.8	77.6	19.4
T	12.0	14.0	6.8		100.0	3.0	11.1	7.8	47.7
Consensus sequence	C/A	A	G	G	T	A/G	A	G	T

Nucleotides that are complementary to U1 snRNA are underlined. In this study, we calculated the CV (21) with the equation $CV = \sum_{l=-3}^{6} (F(n,l) - 0.570)/5.772$, where $F(n, l)$ is a ratio of a nucleotide 'n' at position 'l'. Similarly, we calculated the R_i (7,8) with the equation $R_i = \sum_{l=-3}^{6} (2 + \log_2(F(n,l)))$. We ignored the error function of R_i, because $F(n, l)$ values are calculated using a large number of observations (189 249 sites), and hence the contribution of the error function should be negligible.

spanning positions −3 and +6 from the remaining 189 249 5′ GT splice sites.

Analysis of nucleotide frequencies at each position showed that the 'winner sequence' comprising the most frequently used nucleotides was CAG|GTAAGT (Table 1), which is entirely complementary to U1 snRNA (Figure 1A). The frequency of Cp-nuc was highest at position −1, followed in descending order by positions +5, +4, −2, +3, +6 and −3 (excluding positions +1 and +2, which are invariant in our analysis).

In silico analysis of human 5′ splice sites: the 'two-point' analysis

We next analyzed how an NCp-nuc to U1 snRNA at a specific position is compensated for by a Cp-nuc at the other positions. The 'two-point' analysis revealed that NCp-nucs at positions +4, +5 and +6 are compensated for by Cp-nucs at positions −3, −2, −1 and +3 (Figure 1C). Conversely, NCp-nucs at positions −3, −2 and −1 are associated with high concordance ratios at positions +4, +5 and +6 (Figure 1B). These results suggest that a stretch of Cp-nucs either in an exonic or an intronic region is essential for proper splicing, which also conforms to the notion that consecutive base pairings with U1 snRNA contribute to recognition of the 5′ splice site (20). It is also interesting to note that a Cp-nuc at position +6 most frequently compensates for an NCp-nucs at position −3, −2 and −1, although the frequency of a Cp-nuc at position +6 is only 47.7% in the human genome.

In our analysis, we assumed that only A at position +3 is a Cp-nuc. When we regarded both A and G as Cp-nucs at position +3, the concordance ratio at position +3 became always low (Supplementary Figure 3), likely because A or G is observed at position +3 in 94.1% of the human 5′ splice sites. This implies that the concordance ratio is less informative, when the frequency of a Cp-nuc is high.

In silico analysis of human 5′ splice sites: the SD-Score

The 'two-point' analysis disclosed interdependence between two nucleotides at the 5′ splice sites. However, we could not develop a scoring system using the 'two-point' analysis. We thus sought another quantitative measure to predict splicing consequences. We expected that the frequency of a specific 5′ splice site sequence in the human genome would represent the splicing signal intensity that we hoped to score. We analyzed the entire human genome and determined the frequency of each U2-dependent GT splice site sequence. The common logarithm of the frequency was calculated to give a new variable, the SD-Score (Supplementary Table 5). For example, the SD-Score for CAG|GTGAGG, which was observed at 2562 sites, is log (2562/189 249) = −1.868. The SD-Score of a splice site sequence that never appears in the human genome should be log (0/189 249) = −∞ but is defined as log (0.25/189 249) = −5.879 to simplify calculations. The correlation coefficients of the SD-Score with the R_i (7,8) and the consensus value (CV), which represents the similarity of a splice site sequence to the consensus splice site sequence (21), are 0.678 and 0.694, respectively, indicating that the SD-Score is similar to, but distinct from, the other variables.

Prediction of splicing consequences using the SD-Score algorithm

We next examined whether the SD-Score is indeed an effective scoring variable. To this end, we plotted the SD-Scores and the ΔSD-Scores of the 31 constructs of the *PARK7*, *PINK1*, *CHM*, *FAH*, *HMBS*, *UROS*, *BRCA1* and *CYP27A1* genes (Supplementary Table 2). The ΔSD-Score is calculated by subtracting the wild-type SD-Score from the mutant SD-Score. We found that a 5′ splice site with ΔSD-Score > −0.34 does not affect pre-mRNA splicing in 13 out of 13 sites, whereas a mutant site with ΔSD-Score < −0.34 and SD-Score < −2.9 causes aberrant splicing in 9 out of 10 sites. The 5′ splice sites with ΔSD-Score < −0.34 and SD-Score > −2.9 include a mixed population of three normal and five aberrant splicings. We thus employed the ΔR_i value, which is calculated by subtracting the wild-type R_i from the mutant R_i, and found that three out of three sites with $\Delta R_i > -1.45$ are normally spliced, whereas five out of five sites with $\Delta R_i < -1.45$ are aberrantly spliced. Therefore, these thresholds efficiently predict splicing consequences of our 31 minigene constructs (Figure 3).

Previously unrecognized exonic splicing mutations at positions −2 and −1

To validate the SD-Score algorithm, we employed previously unrecognized splicing mutations at positions

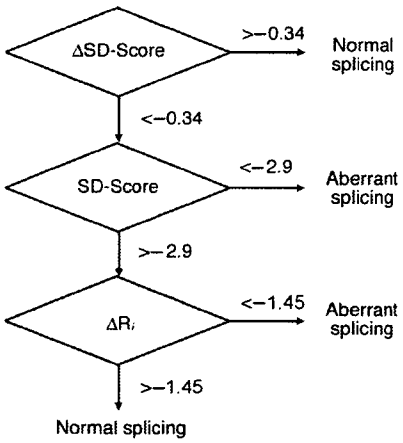

Figure 3. The diagram demonstrates the SD-Score algorithm to predict aberrant splicings due to mutations at the 5′ splice site. The algorithm is based on a training dataset of 31 minigenes and was validated with testing data sets of 32 additional minigenes and 179 naturally occurring splicing mutations (Supplementary Tables 2–4).

−2 and −1. To this end, we scrutinized 2477 exonic mutations from the Human Gene Mutation Database, and searched for mutations at positions −2 and −1. We then randomly selected three mutations in the *DYSF*, *F8* and *ABCD1* genes (Supplementary Table 2) whose splicing consequences have not been characterized. The SD-Score algorithm predicted that all the mutations would affect pre-mRNA splicings, and minigene analyses confirmed it (Supplementary Figure 2C and Supplementary Table 2).

We further introduced a single Cp-nuc to each mutant minigene while retaining the mutation. We thus constructed seven artificial minigenes, and six of these were spliced as predicted (Supplementary Figure 2C and Supplementary Table 2).

Previously unrecognized splicing mutations at position −3

We next sought exonic splicing mutations at position −3, which has been reported only in two mutations (22,23) according to the Human 5′ Splice Site Database (http://www.uni-duesseldorf.de/rna/). In the 2477 exonic mutations described above, we identified six mutations that disrupt a complementary C nucleotide at position −3 (Supplementary Table 2). The SD-Score algorithm predicted that four mutations in the *GLA*, *DMD* and *PARK2* genes would affect pre-mRNA splicing, whereas two mutations in the *ABCD1* and *NPC1* genes would not. We analyzed six mutant minigenes and found that all, except the *PARK2* mutant, were spliced as predicted (Supplementary Figures 2D and 5, and Supplementary Table 2). We confirmed in patient's lymphocytes that a C-to-T mutation at position −3 in *DMD* exon 5 indeed caused the same aberrant splicing as we observed with the minigene (data not shown). The aberrant splicing due to a mutation in *DMD* exon 5, however, was likely successfully predicted by the SD-Score algorithm, because additional mutagenesis at position −4, which did not create a novel cryptic site, failed to show exon skipping (Supplementary Figure 5).

We also constructed seven artificial minigenes, and the SD-Score algorithm successfully predicted the splicing consequences of all the minigenes (Supplementary Figure 2D and Supplementary Table 2).

Splicing mutations in the literature database

To further validate the SD-Score algorithm, we employed other exonic and intronic splicing mutations in the literature database (Supplementary Tables 3 and 4). We randomly examined 2, 9, 26, 45, 3, 83 and 11 splicing mutations at positions −3, −2, −1, +3, +4, +5 and +6, respectively. Our algorithm correctly predicted aberrant splicings in 174 of the 179 reported mutations and falsely predicted normal splicings in five mutations.

DISCUSSION

Clinical implications of exonic splicing mutations

Although our analysis failed to detect ESE- and ESS-affecting mutations in genetic forms of Parkinson's disease, we identified two exonic splicing mutations at the 5′ splice site: E417G in *PINK1* and E64D in *PARK7*. These mutations, as well as six other previously unrecognized exonic splicing mutations in the *DYSF*, *F8*, *ABCD1*, *GLA* and *DMD* genes (Supplementary Table 2), have been reported as synonymous, missense or nonsense mutations. Discrimination of splicing mutations from other types of mutations is essential for understanding human diseases, because different phenotypes and different therapeutic options should be considered for different disease mechanisms. For example, splicing abnormalities in the *IKBKAP* and *SMN2* genes can be normalized with kinetin (24) and sodium valproate (25), respectively.

Prediction of aberrant splicings using the SD-Score algorithm

To predict aberrant splicings due to mutations at the 5′ splice site, we developed the SD-Score algorithm using a training dataset and tested it using a validation dataset. Except for the *PINK1* and *PARK7* genes, we selected mutations without any bias in both minigenes and previously reported splicing mutations in the literature database. Of the 63 minigenes examined in the present study, six normally spliced and seven aberrantly spliced minigenes required the use of ΔR_i values for analysis. In contrast, of the 179 splicing mutations in the literature database, only four mutations required the ΔR_i values for analysis. Artificial minigenes that we constructed to understand interdependence between Cp-nucs and NCp-nucs carry two nonnative nucleotides, whereas naturally occurring mutations carry a single nonnative nucleotide. The SD-Score alone may not be powerful enough to predict the splicing consequences of mutants carrying two or more nonnative nucleotides at the 5′ splice site.

Recognition of an exon, however, is dependent not only on the 5′ splice site sequence, but also on other splicing *cis*-elements, including the branch point, the polypyrimidine tract, the 3′ splice site and ESEs/ESSs and intronic enhancers/silencers. Lack of information about the other