Human branch point consensus sequence is yUnAy

Kaiping Gao, Akio Masuda, Tohru Matsuura and Kinji Ohno*

Division of Neurogenetics, Center for Neurological Diseases and Cancer, Nagoya University Graduate School of Medicine, Nagoya 466-8550, Japan

Received December 6, 2007; Revised January 17, 2008; Accepted February 5, 2008

ABSTRACT

Yeast carries a strictly conserved branch point sequence (BPS) of UACUAAC, whereas the human BPS is degenerative and is less well characterized. The human consensus BPS has never been extensively explored in vitro to date. Here, we sequenced 367 clones of lariat RT-PCR products arising from 52 introns of 20 human housekeeping genes. Among the 367 clones, a misincorporated nucleotide at the branch point was observed in 181 clones, for which we can precisely pinpoint the branch point. The branch points were comprised of 92.3% A, 3.3% C, 1.7% G and 2.8% U. Our analysis revealed that the human consensus BPS is simply yUnAy, where the underlined is the branch point at position zero and the lowercase pyrimidines ('y') are not as well conserved as the uppercase U and A. We found that the branch points are located 21–34 nucleotides upstream of the 3' end of an intron in 83% clones. We also found that the polypyrimidine tract spans 4–24 nucleotides downstream of the branch point. Our analysis demonstrates that the human BPSs are more degenerative than we have expected and that the human BPSs are likely to be recognized in combination with the polypyrimidine tract and/or the other splicing cis-elements.

INTRODUCTION

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. Stepwise assembly of the spliceosome starts from recruitment of U1 snRNP, SF1, U2AF65 and U2AF35 to the 5' splice site, the branch site, the PPT and the 3' end of an intron, respectively (Complex E). SF1, a 75-kDa polypeptide, is a mammalian homolog of yeast BBP (branch point-binding protein). U2AF65 and U2AF35 bring U2snRNP to the BPS in place of SF1 (1,2). The BPS establishes base pairing interactions with a stretch of ‘GUAGUA’ of U2 snRNA (3,4), which then bulges out the branch site nucleotide, usually an adenosine (Complex A) (5). 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 nucleophile 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 leading to intron excision and ligation of the upstream and downstream exons (6). The branch site is thus involved in the first step of splicing, and potentially in the second step of splicing, although the detailed molecular mechanisms of contribution to the second step remain elusive (7).

The BPS is strictly conserved in yeast and has the sequence of UACUAAC, where the branch point adenosine is underlined. On the other hand, the human BPSs are degenerative. No extensive in vitro identification of human BPSs has been reported. Five communications address the mammalian consensus BPSs (Table 1). Three reports are based on 11–20 in vitro identified BPSs, and two are dependent on the in silico analysis of the human genome.

In an effort to establish the human consensus BPS based on in vitro experiments, we analyzed 367 clones of lariat RT-PCR products arising from 52 introns of 20 human housekeeping genes. We found that the human consensus BPS is yUnAy. Our analysis demonstrates that the human BPSs are more degenerative than we have expected and that the BPS is likely recognized in combination with the PPT and/or the other splicing cis-elements.

MATERIALS AND METHODS

Lariat RT-PCR primers for human housekeeping genes

Among the 575 human housekeeping genes registered at http://www.compugen.co.il/supp_info/Housekeeping_genes.html (8), we excluded 82 genes, for which we...
could not find entries in the EST profile viewer of the NCBI UniGene database (http://www.ncbi.nlm.nih.gov/UniGene/). Among 4188 introns of the remaining 493 human housekeeping genes, we excluded introns with a size of less than 300 nucleotides or with multiple repeated segments, because it was difficult to design appropriate PCR primers for such introns. We next sorted the 493 genes in the order of skin expression levels according to the EST profile viewer, and picked up the 20 best genes. We thus analyzed 52 introns of the 20 human housekeeping genes (Supplementary Table 1).

We placed the sense primers at least 100 nucleotides upstream of the 3' end of an intron, and the antisense primers at least 10 nucleotides downstream of the 5' end of an intron. The melting temperatures of the primers were designed to be 64–67°C according to the nearest neighbor method. Gene symbols, intron numbers and primer sequences are indicated in Supplementary Table 1.

Lariat RT-PCR to identify the branch point

We performed nested lariat RT-PCR to amplify a fragment spanning the 2'–5' phosphodiester bond at the branch point (9). We isolated total RNA from HEK293 cells grown to confluency in DMEM medium (Sigma-Aldrich) supplemented with 10% fetal bovine serum (Sigma-Aldrich) and penicillin–streptomycin (Invitrogen). First-strand cDNA was synthesized with SuperScript II reverse transcriptase (Invitrogen) using an intron-specific antisense primer C (Figure 1) located close to the 5' end of an intron. The first round of lariat RT-PCR was performed using primers C and D with Taq HS DNA polymerase (Takara) in 25 μl. The nested lariat RT-PCR was carried out with primers A and B using 0.2 μl of the first-round lariat RT-PCR product in 50 μl. The first-round PCR program was comprised of an initial denaturation step at 94°C for 3 min, followed by 30 cycles of 94°C for 30 s, 55°C for 30 s and 72°C for 1 min. For the nested PCR, we performed 35 cycles of amplification.

We purified the nested lariat RT-PCR products using the Wizard SV Gel and PCR Clean-up system (Promega), and cloned them into the pGEM-T-Easy vector (Promega). We sequenced 10 clones for each intron using the CEQ8000 genetic analyzer (Beckman Coulter).

Table 1. Previously reported consensus BPSs

| Consensus BPS | Note | References |
|---------------|------|------------|
| S. cerevisiae | UACUA| Invariant BPS (36,37) |
| Mammals | YNYURAY | 11 mammalian BPSs (25) |
| | YNCURAY | 20 mammalian BPSs (38) |
| | YNCURAY | 15 BPSs of human HBB (39) |
| | CURAY | In silico homology search (40) |
| | YUVAY | In silico homology search (41) |

Branch point ‘A’ is underlined.
Y, U or C; R, A or G; S, G or C; V, A, C or G.

Presentations of sequence motifs

Sequence motifs are presented using the Pictogram web server at http://genes.mit.edu/pictogram.html (10).

To indicate the amount of information conferred by each nucleotide at each position, we employed the WebLogo program at http://weblogo.berkeley.edu/ (11,12). We also calculated the total amount of information content at each position using the following formula:

\[
R_{\text{sequence}} = 2 + \sum P_i \log_2 P_i \quad (i = \text{A,C,G and U})
\]

where \(P_i\) represents the probability of nucleotide \(i\) at each position. \(R_{\text{sequence}}\) represents the degree of conservation.
respectively. The deduced consensus BPS thus becomes (72.4%), 21 (72.4%), 23 (79.3%) and 25 sites (86.2%),

### RESULTS

**Collation of previously identified mammalian BPSs**

As shown in Table 1, five communications address the mammalian consensus BPSs. In order to understand the *in vitro* determined mammalian consensus BPS, we collated 29 previously reported BPSs comprised of 25 mammalian and four viral introns (Table 2). Viral introns should be spliced in the same way as the mammalian genes. The branch points are located between positions −38 to −4 (mean and SD, −26.9 ± 7.8). Nucleotides ‘C’, ‘U’, ‘A’ and ‘Y’ at positions −3, −2, 0 and +1 are observed at 21 (72.4%), 21 (72.4%), 23 (79.3%) and 25 sites (86.2%), respectively. The deduced consensus BPS thus becomes CUNAy at positions −3 to +1 (Figure 2A and B), when we arbitrarily assume that positions with the information contents above 0.45 are significant.

**Lariat RT-PCR with or without a misincorporated nucleotide at the branch point**

To further explore the human consensus BPS, we chose 52 introns of 20 human housekeeping genes. We performed nested lariat RT-PCR and cloned the amplified products. We sequenced ten clones from each intron, and 367 clones carried available inserts, which represented 117 possible branch sites (Table 3). The remaining 153 clones carried either no inserts or PCR artifacts.

The 367 clones were divided into two classes: 181 clones carrying misincorporated nucleotides at the branch points, and 186 clones without misincorporated nucleotides. For those carrying misincorporated nucleotides, we could pinpoint the exact branch points (Figure 1A). On the other hand, for those carrying no misincorporated nucleotides, the reverse transcriptase might have skipped one or two nucleotides at the 2′–3′ phosphodiester bond at the branch points (Figure 1B).

Among the 367 clones, we observed two or more possible branch sites in 36 of 52 introns. The 36 introns carried a total of 101 possible branch sites. Among the 101 sites, 25 were followed by an immediate downstream branch site, making 25 possible branch-site pairs. Among the 25 upstream branch sites, 19 carried no misincorporated nucleotides. In addition, 13 of the 19 upstream sites were followed by an ‘A’ nucleotide. Furthermore, when we simply deduced the consensus BPS from all 367 clones, the consensus BPS became more degenerative and less informative (data not shown). These findings suggest that the observed upstream branch points are likely due to skipping of a nucleotide in lariat RT-PCR. We thus employed the 181 clones carrying misincorporated nucleotides at the branch points in the following analyses unless otherwise stated.

### Table 2. Previously identified mammalian and viral BPSs

| Species | Gene | Intron | BPS | Position | Predicted BP | BPS Score | Reference |
|---------|------|--------|-----|----------|-------------|-----------|-----------|
| H. sapiens | CALCA | 4 | CACTAC | −36b | −36A | 3.85 | (42) |
| H. sapiens | CALCA | 3 | TACTGTC | −23b | −42A | 2.60 | (21) |
| H. sapiens | CALCA | 3 | GTACTGT | −34b | −42A | 2.60 | (21) |
| H. sapiens | CALCA | 3 | GTTGCTAT | −32b | −42A | 2.60 | (21) |
| H. sapiens | CS1H | 1 | CCTCCAT | −23b | −23A | 2.75 | (22) |
| H. sapiens | DQ6J | 3 | CACAGAC | −21b | −21A | 3.25 | (17) |
| H. sapiens | G1H | 1 | CTCTGTTC | −22b | na | na | (22) |
| H. sapiens | G1H | 1 | GGCTCC | −28b | na | na | (22) |
| H. sapiens | G1H | 1 | TGTCCTC | −36b | na | na | (22) |
| H. sapiens | G1H | 4 | GCTCTCTC | −24b | −37A | 2.80 | (22) |
| H. sapiens | G1H | 4 | ACCCAAG | −37b | −37A | 2.80 | (22) |
| H. sapiens | G1H | 4 | TACCATAA | −38b | −37A | 2.80 | (22) |
| H. sapiens | HBA | 1 | CCCTCACC | −10b | −37A | 3.25 | (25) |
| H. sapiens | HBA | 2 | CACTGAC | −18b | −18A | 3.95 | (25) |
| H. sapiens | HBB | 1 | CACTGAC | −37b | −37A | 3.95 | (25) |
| H. sapiens | HBB | 1 | CTCTAATT | −31b | −31A | 3.45 | (25) |
| H. sapiens | HBE1 | 1 | CTCTTGAC | −30b | −30A | 3.85 | (25) |
| H. sapiens | HBG1 | 1 | CTCTTGAC | −30b | −30A | 3.85 | (25) |
| H. sapiens | MYH10 | 5 | TGCTTAC | −31b | na | na | (15) |
| H. sapiens | XPC | 3 | TGTTGAT | −4b | na | na | (16) |
| H. sapiens | XPC | 3 | TACTGAT | −24b | na | na | (16) |
| M. musculus | Hbb-b2 | 1 | CACTAA | −36b | −36A | 3.85 | (25) |
| M. musculus | Igf | 5 | AATTAC | −22b | −22A | 3.30 | (14) |
| R. norvegicus | Ins1 | 1 | CCTACA | −18b | −18A | 3.15 | (25) |
| O. cuniculus | Hbb | 1 | TGCTGAC | −34b | −34A | 3.85 | (25) |
| O. cuniculus | Hbb | 2 | TGCTGAC | −32b | −32A | 3.75 | (29) |
| Adenovirus 5 | E1A | 1 | GTTATAA | −30b | −30A | 2.70 | (25) |
| Adenovirus 2 | E2a-2 | 1 | GACTGAC | −26b | −26A | 3.70 | (42) |
| Adenovirus 5 | Major Late | 1 | TACCTAAT | −24b | −24A | 3.05 | (42) |
| SV40 | T antigen | 1 | ATTCTGAC | −19b | −19A | 2.00 | (42) |

*aPredicted BPs and BPS scores are according to the Branch-Site Analyzer at http://ast.bioinfo.tau.ac.il/BranchSite.htm (14). na, not available.

*bIdentified by the primer extension method.

*Identified by lariat RT-PCR.
We counted each clone as a single occurrence of a branch point in order to weigh the preferred branch points. For example, \( PGK1 \) intron 6 had eight clones mapping to position \( +23 \), whereas \( EEF1A1 \) intron 1 had ten clones all mapping to position \( -23 \). A single clone of \( PGK1 \) might have arisen from one of many branch points, and we might have sequenced it by chance. On the other hand, it is likely that \( EEF1A1 \) intron 1 indeed had a single branch point. We analyzed our data using both methods and obtained similar results (data not shown), except that the frequency of \( C \) at position \( -1 \) was slightly lower with the alternative method (44.8% versus 36.3%). In the current communication, we employed the former method, in which each clone was counted as a single occurrence of a branch site.

**Positions and sequence motif of the branch points**

Analysis of the 181 clones revealed that the positions of the branch points were from \( -50 \) to \( -5 \), where position \( -1 \) represents the 3\(^{0}\) end of an intron (Figure 3A). Among the 181 sites, 150 (83%) were at positions \( +34 \) to \( +21 \). We observed \( U \) at position \( +2 \) in 74.6% branch sites, and \( A \) at position \( 0 \) in 92.3% branch sites (Table 4). In addition, pyrimidines were observed at positions \( +3 \) and \( +1 \) in 79.0% and 75.1% branch sites, respectively (Table 4). We can thus conclude that the human consensus BPS is \( \text{yUnA} \text{y} \) at positions \( +3 \) to \( +1 \) (Figure 2C), where the branch site is underlined and the less conserved nucleotides are indicated in lowercase letters. The information contents of 0.27 and 0.23 at positions \( +3 \) and \( +1 \), however, were not as high as those of 0.85 and 1.48 at positions \( -2 \) and \( 0 \) (Figure 2D), or \( 0.39 \pm 0.12 \) (mean \( \pm \) SD) of the polypyrimidine tract at positions \( +4 \) to \( +24 \) (Figure 3C). Therefore, the consensus sequence alternatively becomes \( \text{UnA} \) according to the information contents (Figure 2D).

Among the 41 introns yielding the 181 clones, 14 introns carried multiple branch sites. In eight of the 14 introns (57%), the most downstream branch sites were most frequently used (Table 3). Although the ratio of 57% was not high, the downstream branch sites were four to eight times more frequently used than the upstream sites in four of the eight introns. We could not observe this magnitude of differential branch site usage in the remaining six introns, in which the downstream branch points were not overrepresented. Accordingly, when there are multiple branch points, downstream branch points are more likely to be employed than their upstream counterparts.

We also predicted BPSs of our housekeeping genes with the Branch Site Analyzer (14), and found that the actual branch sites matched to the predicted positions in 80 of the 181 sites (44.2%) (Table 3).

**Alignment of polypyrimidine tract in respect to the branch point**

We next aligned the PPT’s in respect to the 181 branch points (Figure 3B and C). We observed a polypyrimidine stretch from position \( +4 \) down to position \( +24 \). The ‘\( U \)’ nucleotide was preferred over ‘\( C \)’ especially at positions \( +4 \) to \( +12 \) in the PPT. Alignment of the PPT in respect to the 3\(^{0}\) end of an intron also demonstrated a stretch of pyrimidines from positions \( -20 \) to \( -4 \) (Figure 3D and E).
Figure 3. (A) Positions and nucleotides of 181 branch points with misincorporated nucleotides in our studies, where position −1 represents the 3' end of an intron. The median value of the branch points is −26, and the mean and SD is −27.7 ± 7.6. Among the 181 sites with misincorporated nucleotides at the branch points, 150 sites (83%) are at positions −34 to −21 (horizontal bar on top). Native nucleotides, not the misincorporated nucleotides, are indicated. Nucleotide preferences (B and D) and information contents (C and E) are deduced from 181 branch points. (B and C) Plots are aligned in respect to the branch point (closed arrows), which is designated as position 0. Open arrows point to peaks of information contents at positions +7 and +8. A polypyrimidine stretch starts from position +4 down to position +24 (bars). The plots are truncated at position +25, because the numbers of observations fall below 40 after position +25, and the plots become less informative and uneven. The last three nucleotides of introns are excluded from the plots. (D and E) Plots are aligned in respect to the 3' end of each intron, which is designated as position −1. A polypyrimidine stretch spans positions −19 to −5 (bars).

Figure 4. Representative composition of the branch point sequence (arrow) and the PPT deduced from our studies.
Table 3. Analyzed introns and observed branch points

| Gene      | Intron size (bp) | Predicted BPS Score | Observed BPS | Number of clones | Misincorporated nucleotide | Intronic sequence from BPS position -5 to the 3' end of an intron |
|-----------|------------------|---------------------|--------------|------------------|-----------------------------|---------------------------------------------------------------|
| ACTB      | 3                | −24A                | 3.10         | −30A             | 10                          | TCCCCAAGTGACATGTTGACCATCCTGCTCTTACAG                           |
| ALOD      | 8                | −24A                | 3.05         | −26T             | 2                           | TGTCCTAAGTGACCTTACCTGCTCTGCCCCAACAG                           |
|           |                  |                     |              |                  |                             | GTCTAAGTGACCTTACCTGCTCTGCCCCAACAG                              |
|           |                  | −25A                | 2            | T                | −5A                         | ACCCCCAAGTGACCTTACCTGCTCTGCCCCAACAG                           |
| ACTB      | 4                | −21A                | 3.30         | −32A             | 3                           | TGAATAGTGACATGTTGACCATCCTGCTCTTACAG                           |
| CCT3      | 11               | −44A                | 2.95         | −24A             | 6                           | TGTGACATGTTGACCATCCTGCTCTTACAG                               |
|           |                  | −10C                | 1            | −10C             |                             | AATTGTCCTGACCTTACCTGCTCTTACAG                                |
| EEFL1A    | 2                | −24A                | 2.80         | −28A             | 2                           | TAGTACCAAGTAGACTTACAAATCCCTTACAG                             |
|           |                  | −19C                | 1            | −19C             |                             | AATTGTCCTGACCTTACCTGCTCTTACAG                                |
| EEFL1A    | 1                | −33A                | 3.25         | −23A             | 10                          | TGGTCAAGTTTTCCTTCCATTACAG                                    |
| ENO1      | 2                | −33A                | 3.25         | −27T             | 6                           | ATGTCCTACATGCATTTTCTCTCCTCATACAG                             |
|           |                  | −25C                | 2            | T                | −25C                        | TCACTCCTACATGCATTTTCTCTCCTCATACAG                             |
| ENO1      | 4                | −21A                | 3.45         | −21A             | 10                          | TCCCTACATCCCTCCCTCCTCCATTAG                                   |
|           |                  | −24C                | 1            | −24C             |                             | TCCCTACATCCCTCCCTCCTCCATTAG                                   |
| ENO1      | 6                | −36A                | 2.85         | −30T             | 2                           | CCCCAGTGCACATGCTCCTCTGCCTCTCCTCATACAG                         |
|           |                  | −27C                | 3            | A                | −27C                        | AGTGCCTACATGCATTTTCTCTCCTCATACAG                               |
| ENO1      | 7                | −96                | 2.50         | −25A             | 3                           | CGGCAGCTGACTTACATGCTCCTCTCCTCATACAG                            |
|           |                  | −26A                | 3            | −26A             |                             | CGGCAGCTGACTTACATGCTCCTCTCCTCATACAG                            |
| ENO1      | 9                | −54                | 2.70         | −5T              | 2                           | TGGTCCTACATGCTCCTCTCCTCATACAG                                 |
| ENO1      | 11               | −26A                | 2.95         | −48T             | 4                           | AAGGACGACTTACATGCTCCTCTCCTCATACAG                              |
|           |                  | −46A                | 3            | T                | −46A                        | AAGGACGACTTACATGCTCCTCTCCTCATACAG                              |
| G22P1     | 1                | −28A                | 2.90         | −30A             | 2                           | AGGCAAAAGACATTTTCTCTCATTATTTTCTCTCTCATACAG                    |
| G22P1     | 2                | −26A                | 2.50         | −31A             | 2                           | AACAAAGACATTTTCTCTCATTATTTTCTCTCTCATACAG                      |
|           |                  | −26A                | 1            | T                | −26A                        | AACAAAGACATTTTCTCTCATTATTTTCTCTCTCATACAG                      |
|           |                  | −25A                | 1            | T                | −25A                        | AACAAAGACATTTTCTCTCATTATTTTCTCTCTCATACAG                      |
| G22P1     | 10               | −33A                | 3.60         | −33A             | 10                          | GACTCAGGAGGACTTACATGCTCCTCTCTCATACAG                            |
| GAPDH     | 2                | −26A                | 3.35         | −6T              | 10                          | TGTGCCTACATGCTCCTCTCCTCATACAG                                 |
| HSPA8     | 1                | −39A                | 3.15         | −19A             | 1                           | TTTTAAAGACATTTTCTCTCATTATTTTCTCTCTCATACAG                    |
|           |                  | −17A                | 1            | C                | −17A                        | TTTTAAAGACATTTTCTCTCATTATTTTCTCTCTCATACAG                    |
| HSPCB     | 6                | −18A                | 3.15         | −22A             | 6                           | GTACCCCTATTTTTTGTGACCATCCTACAG                                |
|           |                  | −23C                | 1            | −23C             |                             | GTACCCCTATTTTTTGTGACCATCCTACAG                                |
| HSPCB     | 10               | −22A                | 3.20         | −24T             | 1                           | CAACTGAAAGCAGTTTGGGATGCTACCATACAG                             |
|           |                  | −22A                | 2            | T                | −22A                        | AATTGAAAGCAGTTTGGGATGCTACCATACAG                              |
| HSPCB     | 1                | −22A                | 2.75         | −18A             | 1                           | AATTGAAAGCAGTTTGGGATGCTACCATACAG                              |
| LDHB      | 2                | −22A                | 3.35         | −24T             | 4                           | GGGTCCTATGCGTTGACCATCCTTACAG                                  |
|           |                  | −23A                | 1            | T                | −23A                        | GGGTCCTATGCGTTGACCATCCTTACAG                                  |
|           |                  | −22A                | 6            | T                | −22A                        | GGGTCCTATGCGTTGACCATCCTTACAG                                  |
| PGK1      | 1                | −28A                | 3.00         | −29G             | 2                           | AGTGTGAAAGCAGTTTGGGATGCTACCATACAG                             |
|           |                  | −28A                | 4            | T                | −28A                        | AGTGTGAAAGCAGTTTGGGATGCTACCATACAG                             |
| PGK1      | 2                | −35A                | 3.00         | −26T             | 4                           | AGTGTGAAAGCAGTTTGGGATGCTACCATACAG                             |
|           |                  | −26T                | 4            | −26T             |                             | AGTGTGAAAGCAGTTTGGGATGCTACCATACAG                             |
| PGK1      | 4                | −29A                | 3.35         | −33C             | 1                           | TGACCTAAGGACATTTTCTCTCCTCCTTACTACAG                            |
|           |                  | −29A                | 1            | T                | −29A                        | TGACCTAAGGACATTTTCTCTCCTCCTTACTACAG                            |

(Continued)
Table 3. Continued

| Gene  | Intron size (bp) | Predicted BP | BPS Score | Observed BP | Number of clones | Misincorporated nucleotide | Intronic sequence from BPS position −5 to the 3′ end of an intron |
|-------|------------------|--------------|-----------|-------------|-------------------|----------------------------|---------------------------------------------------------------|
| PKM2  | 2 1443           | −39A 3.05    | −31A 2    | −29A 1 T    | 1 T               | −23A 8 T                 | −23G 1 C                                                      |
|       |                  |              |           | −29A 1 T    |                   | −25G 1 C                 | −23G 1 C                                                      |
| PKM2  | 3 6930           | −21A 2.75    | −25T 3 −  | −21A 4 T    | 1 T               | −21A 4 T                 | −21A 4 T                                                      |
|       |                  |              |           | −16A 1 T    |                   | −16A 1 T                 | −16A 1 T                                                      |
| PKM2  | 4 487            | −26A 2.85    | −38T 1 −  | −33A 1 T    |                  | −33A 1 T                 | −33A 1 T                                                      |
|       |                  |              |           | −23C 1 T    |                   | −23C 1 T                 | −23C 1 T                                                      |
| PKM2  | 5 781 na         | na na        | −32T 1 −  | −31C 1 −    |                  | −31C 1 −                 | −31C 1 −                                                       |
|       |                  |              |           | −28C 3 −    |                   | −28C 3 −                 | −28C 3 −                                                       |
| PKM2  | 6 1343           | −29A 3.10    | −39G 1 C  | −31C 1 −    |                  | −31C 1 −                 | −31C 1 −                                                       |
|       |                  |              |           | −65T 2 −    |                   | −65T 2 −                 | −65T 2 −                                                       |
| PKM2  | 8 4107 na        | na na        | −52T 1 −  | −50A 5 T    |                  | −50A 5 T                 | −50A 5 T                                                       |
|       |                  |              |           | −52T 1 −    |                   | −52T 1 −                 | −52T 1 −                                                       |
| PKM2  | 10 717           | −25A 3.75    | −27T 1 −  | −25A 6 T    |                  | −25A 6 T                 | −25A 6 T                                                       |
|       |                  |              |           | −26C 2 −    |                   | −26C 2 −                 | −26C 2 −                                                       |
| PSMB4 | 4 393            | −40A 3.05    | −44A 4 T  | −40A 5 T    |                  | −40A 5 T                 | −40A 5 T                                                       |
|       |                  |              |           | −40A 5 T    |                   | −40A 5 T                 | −40A 5 T                                                       |
| RPL13 | 4 583            | −21A 3.30    | −26A 1 C  | −26A 1 C    |                  | −26A 1 C                 | −26A 1 C                                                       |
|       |                  |              |           | −26A 1 C    |                   | −26A 1 C                 | −26A 1 C                                                       |
| RPL13 | 5 492            | −22A 3.30    | −22A 9 −  | −20A 1 T    |                  | −20A 1 T                 | −20A 1 T                                                       |
|       |                  |              |           | −20A 1 T    |                   | −20A 1 T                 | −20A 1 T                                                       |
| RPL13A| 1 2205           | −26A 3.55    | −22C 7 −  | −8T 1 −     |                  | −8T 1 −                  | −8T 1 −                                                        |
|       |                  |              |           | −8T 1 −     |                   | −8T 1 −                  | −8T 1 −                                                        |
| RPL3  | 2 770            | −21A 3.70    | −21A 4 T  | −19T 3 −    |                  | −19T 3 −                 | −19T 3 −                                                       |
|       |                  |              |           | −19T 3 −    |                   | −19T 3 −                 | −19T 3 −                                                       |
| RPL3  | 4 1150           | −22A 3.30    | −24T 10  | −24T 10     |                  | −24T 10                  | −24T 10                                                        |
|       |                  |              |           | −24T 10     |                   | −24T 10                  | −24T 10                                                        |
| RPL3  | 5 522            | −22A 2.50    | −29T 1 −  | −29T 1 −    |                  | −29T 1 −                 | −29T 1 −                                                       |
|       |                  |              |           | −29T 1 −    |                   | −29T 1 −                 | −29T 1 −                                                       |
| RPL8  | 4 806            | −23C 5 −     | −21A 3 T  | −21A 3 T    |                  | −21A 3 T                 | −21A 3 T                                                       |
|       |                  |              |           | −21A 3 T    |                   | −21A 3 T                 | −21A 3 T                                                       |
| SLC25A3| 2 1591           | −30A 2.85    | −31A 1 − | −31A 1 −    |                  | −31A 1 −                 | −31A 1 −                                                       |

(Continued)
The information contents at the PPT’s were similar between the two alignments. We observed peaks of information contents seven and eight nucleotides downstream of the branch point. The functional significance of these peaks, however, remains elusive.

Information obtained from lariat RT-PCR clones without misincorporated nucleotides

We next asked if we could exploit the 186 clones without misincorporated nucleotides at the branch point. If there was an ‘A’ nucleotide one or two nucleotides downstream of the sequenced branch point and the sequenced branch point was not ‘A’, we assumed that one or two nucleotides were skipped by the reverse transcriptase and that the particular downstream ‘A’ was the actual branch point. A similar assumption has also been applied to three other genes in previous reports (15–17). We aligned the branch points under this assumption, and plotted the nucleotide preferences and the information contents (Figure 2E and F). Compared to those of misincorporated nucleotides, the information contents were generally lower, but the Pictogram and WebLogo presentations resulted in similar patterns. These analyses suggest that one or two nucleotides were skipped when there were no misincorporated nucleotides, but definite experimental evidence is lacking to employ these clones to deduce the human consensus BPS.

**DISCUSSION**

**Highly degenerative human BPS**

We determined splicing branch points in 52 introns of 20 human housekeeping genes by lariat RT-PCR. Our analysis disclosed the following features (Figure 4). First, 83% of the branch points are located 21–34 nucleotides upstream of the 3′ end of an intron (Figure 3A). Second, a polypyrimidine stretch spans 4–24 nucleotides downstream of the branch point (Figure 3B and C). Third, the human branch point consensus sequence is yUnAy (Figure 2C and D). The first and the second features underscore the previous *in silico* observations (6,14), whereas the degeneracy of the human BPS is more than we have expected.

It is interesting to note that among the six consensus BPSs proposed for the mammalian branch points (Table 1), the shared nucleotides are yUnAy, which is identical to that determined by our analysis.

**SF1 binds to BPS using its KH domain**

SF1 binds to BPS using its KH domain (18). NMR analysis of SF1 bound to the BPS revealed that a hydrophobic motif of Gly-Pro-Arg-Gly within the KH domain builds hydrogen bonds with ‘UAyA’ at positions −2 to 0 of the yeast BPS, ‘UACUAAC’ (19). Our analysis suggests that the binding of the KH domain to position −1 may enhance, but may be dispensable for, the recognition of the BPS. Berglund and colleagues (20) also demonstrate that, in ‘UACUAAC’ at positions −5 to +1, nucleotide substitutions only at position −2 or 0, but not at the other positions, compromise the binding of SF1.

**Non-‘A’ nucleotides at position 0**

We observed an ‘A’ nucleotide at 92.3% of the branch points. Non-‘A’ nucleotides at the branch point have been reported in *CALCA1* (21) and *GH1* (22) (Table 2). The two reports demonstrate six such examples in...
four introns. As these unusual branch points constitute 21% (6/29) of the previously reported in vitro determined branch points, the ratio of ‘A’ at the branch point is reduced to 79% (Table 2). Additionally, the potential observation bias posed by these unusual BPSs may account for the differences in the Pictogram and WebLogo patterns between the previously identified BPSs (Figure 2A and B) and our BPSs (Figure 2C and D).

**Disease-causing mutations disrupting BPSs**

According to the Human Gene Mutation Database (23), splicing mutations account for 13.7% (1768 of 12 879) of single nucleotide substitutions. Most splicing mutations, however, are at the splice donor or acceptor sites. To our knowledge, sixteen disease-causing mutations and a single polymorphism disrupt BPSs and give rise to aberrant splicings (Table 5). Nine variants are at position 0, and the other eight are at position −2. Among the nine variants affecting position 0, seven are A-to-G mutations, which supports the notion reported by Kralovicova and colleagues (24) that A-to-G transitions at position 0 are more deleterious than A-to-T or A-to-C transversions. For all the variants, aberrant splicings have been determined either in patients or minigenes. The actual branch points, however, have been identified only in two variants by lariat RT-PCR, whereas the remaining fourteen variants have been mapped to putative BPSs. Exclusive confinement of BPS-disrupting nucleotide changes at positions −2 and 0 also underscores our observation that the BPS consensus is yUn/AC.

Conversely, mutations disrupting yUn/Ay are not always deleterious. When the branch point ‘A’ is mutated or deleted, a neighboring cryptic ‘A’ residue is employed as a branch point (25–27), or the mutant ‘C’, ‘G’ or ‘U’ residue is used as a surrogate branch point (28). Additionally, we observed two or more branch sites in 15 of 41 introns (Table 3), which also implies that a mutation-harboring BPS can be readily substituted for by another BPS.

**How is the highly degenerative BPS recognized?**

It is hard to believe that SF1 simply recognizes yUn/Ay. We expect that SF1 recognizes the BPS along with the other cis-element(s) and their interacting trans-factor(s).

The SELEX screening of the yeast BBP binding motifs revealed a stem and loop structure immediately upstream of the BPS of ‘UACUAAC’ in 9 out of 48 selected motifs (29). A gel shift assay also showed preferential binding of human SF1 to ‘UACUAAC’ carrying an upstream stem and loop. Our BPSs, however, had no upstream stem and loop structures (data not shown). An upstream stem and loop may help recognize highly degenerative mammalian BPSs for a subset of introns.

In the early step of the spliceosome assembly, SF1, U2AF65 and U2AF35 bind to the BPS, the PPT and AG at the 3′ end of an intron, respectively, to form complex E (1,2). In *S. pombe*, SF1/BBP is tightly associated with U2AF59, a yeast homolog of mammalian U2AF65 recognizing the PPT, as well as with U2AF23, a yeast homolog of mammalian U2AF35 recognizing the 3′ AG (30).

**Table 5. Sixteen mutations and a single polymorphism disrupting BPSs**

| Gene and intron | Sequence | Consequence | Reference |
|-----------------|----------|-------------|-----------|
| *LCAT* intron4  | CCCTGAC  | Intron retention<sup>a</sup> | (43,44)   |
| Wild-type       | CCCGGAC  | Intron retention<sup>b</sup> | (45)      |
| Mutant          | CCCAGAC  | Intron retention<sup>b</sup> | (45)      |
| *FBX2* intron30 | TAATAG   | Exon skipping<sup>a</sup> | (46)      |
| Wild-type       | TAATAG   | Exon skipping<sup>a</sup> | (46)      |
| Mutant          | TAATAG   | Exon skipping<sup>a</sup> | (46)      |
| *COLS1* intron32| GACTGAC  | Intron retention<sup>a</sup> | (48)      |
| Wild-type       | GACTGAC  | Intron retention<sup>a</sup> | (48)      |
| Mutant          | GACTGAC  | Intron retention<sup>a</sup> | (48)      |
| *ITGB4* intron31| GCCCAC   | Intron retention<sup>a</sup> | (48)      |
| Wild-type       | GCCCAC   | Intron retention<sup>a</sup> | (48)      |
| Mutant          | GCCGAC   | Intron retention<sup>a</sup> | (48)      |
| *TH* intron10   | GCTGAT   | Exon skipping<sup>a</sup> | (49)      |
| Wild-type       | GCTGAT   | Exon skipping<sup>a</sup> | (49)      |
| Mutant          | GCTGAT   | Exon skipping<sup>a</sup> | (49)      |
| *LICAM* intron18| ATCCAAG  | Exon skipping<sup>a</sup> | (50)      |
| Wild-type       | ATCCAAG  | Exon skipping<sup>a</sup> | (50)      |
| Mutant          | ATCCAAG  | Exon skipping<sup>a</sup> | (50)      |
| *LIPC* intron1  | CCAAT    | Cryptic 3′ splice site<sup>a</sup> | (51)     |
| Wild-type       | CCAAT    | Cryptic 3′ splice site<sup>a</sup> | (51)     |
| Mutant          | CCAAT    | Cryptic 3′ splice site<sup>a</sup> | (51)     |
| *FBX2* intron28 | TTGCAAT  | Exon skipping<sup>a</sup> | (52)      |
| Wild-type       | TTGCAAT  | Exon skipping<sup>a</sup> | (52)      |
| Mutant          | TTGCAAT  | Exon skipping<sup>a</sup> | (52)      |
| *HEXB* intron10 | TTGCAAT  | Cryptic 3′ splice site<sup>a</sup> | (53)     |
| Wild-type       | TTGCAAT  | Cryptic 3′ splice site<sup>a</sup> | (53)     |
| Mutant          | TTGCAAT  | Cryptic 3′ splice site<sup>a</sup> | (53)     |
| *NF2* intron5   | TTCTAGC  | Intron retention<sup>a</sup> | (54)      |
| Wild-type       | TTCTAGC  | Intron retention<sup>a</sup> | (54)      |
| Mutant          | TTCTAGC  | Intron retention<sup>a</sup> | (54)      |
| *TSC2* intron38 | GCGTAC   | Cryptic 3′ splice site<sup>a</sup> | (55)     |
| Wild-type       | GCGTAC   | Cryptic 3′ splice site<sup>a</sup> | (55)     |
| Mutant          | GCGTAC   | Cryptic 3′ splice site<sup>a</sup> | (55)     |
| *XPC* intron3<sup>o</sup> | TACTGAT  | Exon skipping<sup>a</sup> | (56)     |
| Wild-type       | TACTGAT  | Exon skipping<sup>a</sup> | (56)     |
| Mutant          | TACTGAT  | Exon skipping<sup>a</sup> | (56)     |
| *NPC1* intron6<sup>o</sup> | CACTAT  | Exon skipping<sup>a</sup> | (57)     |
| Wild-type       | CACTAT  | Exon skipping<sup>a</sup> | (57)     |
| Mutant          | CACTAT  | Exon skipping<sup>a</sup> | (57)     |
| *F9* intron2    | CTTTAAT  | Exon skipping<sup>a</sup> | (58)     |
| Wild-type       | CTTTAAT  | Exon skipping<sup>a</sup> | (58)     |
| Mutant          | CTTTAAT  | Exon skipping<sup>a</sup> | (58)     |
| *DQB1* intron3<sup>cd</sup> | CAGAGC  | Exon skipping<sup>a</sup> | (17)     |
| Genotype A      | CAGAGC  | Exon skipping<sup>a</sup> | (17)     |
| Genotype U      | CAGAGC  | Exon skipping<sup>a</sup> | (17)     |

Mutations or a polymorphism are underlined. Aberrant splicings have been determined in patients<sup>a</sup> or minigenes<sup>b</sup>.<br>Branch points have been identified by lariat RT-PCR. Others are putative BPSs lacking in vitro evidence.<br>Polymorphism.

**Table 4. Nucleotide frequencies at the 181 branch sites**

| Position | –5 | –4 | –3 | –2 | –1 | 0 | 1 | 2 | 3 |
|----------|----|----|----|----|----|---|---|---|---|
| A        | 0.254 | 0.232 | 0.083 | 0.066 | 0.166 | 0.923 | 0.182 | 0.302 | 0.201 |
| C        | 0.210 | 0.227 | 0.470 | 0.160 | 0.448 | 0.033 | 0.331 | 0.274 | 0.391 |
| G        | 0.254 | 0.193 | 0.127 | 0.028 | 0.177 | 0.017 | 0.066 | 0.112 | 0.112 |
| U        | 0.282 | 0.348 | 0.320 | 0.746 | 0.210 | 0.028 | 0.420 | 0.313 | 0.296 |
In mammals, the association between SF1 and U2AF65 is mediated by the 28 N-terminal amino acids of the KH domain of SF1 (31) and by the third RBD of U2AF65 (32). Wang and colleagues determined that Ser20 in the N-terminal region of the KH domain is essential for binding to U2AF65 and that phosphorylation of Ser20 inhibits its binding and formation of complex A (33). Berglund and colleagues also report that the SF1-U2AF65 interaction promotes cooperative binding to the BPS and the PPT (32). Our analysis also demonstrates positional association of the BPS and the PPT (Figure 3B and C). On the other hand, Kent and colleagues demonstrate that U2AF65 and U2AF35 are dispensable for the binding of SF1 to the BPS (34). Sharma and colleagues similarly show that complex H includes SF1 in the absence of U2AF65 and U2AF35 (35). Although the exact order of the SF1, U2AF65 and U2AF35 assembly remains elusive, the BPS is possibly recognized along with the PPT and the 3' AG. Alternatively, SF1 is bound to any yUnAy sequences in complex H, and a particular SF1 that successfully associates with the U2AF heterodimer exclusively survives to form complex E.

**SUPPLEMENTARY DATA**

Supplementary Data are available at NAR Online.

**ACKNOWLEDGEMENTS**

We appreciate Jun Shinmi and Keiko Itano for technical assistance. This work was supported by Grants-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Culture, Sports, Science, and Technology of Japan, as well as from the Ministry of Health, Labor, and Welfare of Japan. Funding to pay the Open Access publication charges for this article was provided by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

Conflict of interest statement. None declared.

**REFERENCES**

1. Wu, S., Romfo, C.M., Nilsen, T.W. and Green, M.R. (1999) Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature*, 402, 832–835.

2. Zorio, D.A. and Blumenthal, T. (1999) Both subunits of U2AF recognize the 3' splice site in Caenorhabditis elegans. *Nature*, 402, 835–838.

3. Arning, S., Gruter, P., Bilbe, G. and Kramer, A. (1996) Mammalian splicing factor SF1 is encoded by variant cDNAs and binds to RNA. *RNA*, 2, 794–810.

4. Abovich, N. and Rosbash, M. (1997) Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell*, 89, 403–412.

5. Queru, C.C., Moore, M.J. and Sharp, P.A. (1994) Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. *Genes Dev.*, 8, 587–597.

6. Will, C.L. and Lührmann, R. (2006) Spliceosome structure and function. In Gesteland, R.F., Cech, T.R. and Atkins, J.F. (eds), *The RNA World*, 3rd edn. Cold Spring Harbor Laboratory Press, Plainview, NY, pp. 369–400.

7. Queru, C.C., Strelbel, S.A. and Sharp, P.A. (1996) Three recognition events at the branch-site adenine. *EMBO J.*, 15, 1392–1402.

8. Eisenberg, E. and Levanon, E.Y. (2003) Human housekeeping genes are compact. *Trends Genet.*, 19, 362–365.

9. Vogel, J., Hess, W.R. and Borner, T. (1997) Precise branch point mapping and quantification of splicing intermediates. *Nucleic Acids Res.*, 25, 2030–2031.

10. Garge, C.B., Tuschl, T. and Sharp, P.A. (1999) Splicing of precursors to mRNAs by the spliceosomes. In Gesteland, R.F., Cech, T.R. and Atkins, J.F. (eds), *The RNA World*, Cold Spring Harbor Laboratory Press, Plainview, NY, pp. 525–560.

11. Schneider, T.D. and Stephens, R.M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.*, 18, 6197–6199.

12. Crooks, G.E., Hon, G., Chandonia, J.M. and Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.*, 14, 1188–1190.

13. Rogan, P.K. and Schneider, T.D. (1995) Using information content and base frequencies to distinguish mutations from genetic polymorphisms in splice junction recognition sites. *Hum. Mutat.*, 6, 74–76.

14. Kol, G., Lev-Maor, G. and Ast, G. (2005) Human-mouse comparative analysis reveals that branch-site plasticity contributes to splicing regulation. *Hum. Mol. Genet.*, 14, 1559–1568.

15. Guo, N. and Kawamoto, S. (2000) An intronic downstream enhancer promotes 3' splice site usage of a neural cell-specific exon. *J. Biol. Chem.*, 275, 33641–33649.

16. Khan, S.G., Metin, A., Czurkoka, E., Ini, H., Shahlavi, T., Muniz-Medina, V., Baker, C.C., Ueda, T., Aiken, J.R., Schneider, T.D. et al. (2004) Two essential splice lariat branchpoint sequences in one intron in a xeroderma pigmentosum DNA repair gene: mutations result in reduced XPC mRNA levels that correlate with cancer risk. *Hum. Mol. Genet.*, 13, 343–352.

17. Kralovicova, J., Houngninou-Molango, S., Kramer, A. and Vorechovsky, I. (2004) Branch site haplotypes that control alternative splicing. *Hum. Mol. Genet.*, 13, 3189–3202.

18. Berglund, J.A., Fleming, M.L. and Rosbash, M. (1998) The KH domain of the branchpoint sequence binding protein determines specificity for the pre-mRNA branchpoint sequence. *RNA*, 4, 998–1006.

19. Liu, Z., Luyten, J., Bottomley, M.J., Messias, A.C., Houngninou-Molango, S., Sprangers, R., Zanier, K., Kramer, A. and Sattler, M. (2001) Structural basis for recognition of the intron branch site RNA by splicing factor 1. *Science*, 294, 1098–1102.

20. Berglund, J.A., Chua, K., Abovich, N., Reed, R. and Rosbash, M. (1997) The splicing factor BBF interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell*, 89, 781–787.

21. Adema, G.J., Bovenberg, R.A., Janss, H.S. and Baas, P.D. (1988) Unusual branch point selection involved in splicing of the alternatively processed Calcitonin/CGRP-I pre-mRNA. *Nucleic Acids Res.*, 16, 9513–9526.

22. Hartmuth, K. and Barta, A. (1988) Unusual branch point selection in processing of human growth hormone pre-mRNA. *Mol. Cell. Biol.*, 8, 2011–2020.

23. Krawczak, M., Ball, E.V., Fenton, J., Stenson, P.D., Abeysinghe, S., Thomas, N. and Cooper, D.N. (2000) Human gene mutation database-a biomedical information and research resource. *Hum. Mutat.*, 15, 45–51.

24. Kralovicova, J., Lei, H. and Vorechovsky, I. (2006) Phenotypic consequences of branch point substitutions. *Hum. Mutat.*, 27, 803–813.

25. Reed, R. and Maniatis, T. (1985) Intron sequences involved in lariat formation during pre-mRNA splicing. *Cell*, 41, 95–105.

26. Ruskin, B., Greene, J.M. and Green, M.R. (1985) Cryptic branch point activation allows accurate in vitro splicing of human beta-globin intron mutants. *Cell*, 41, 833–844.

27. Padgett, R.A., Konarska, M.M., Aeblu, M., Hornig, H., Weissmann, C. and Sharp, P.A. (1985) Nonconsensus branch-site sequences in the in vitro splicing of transcripts of mutant rabbit beta-globin genes. *Proc. Natl Acad. Sci. USA*, 82, 8349–8353.

28. Hornig, H., Aeblu, M. and Weissmann, C. (1986) Effect of mutations at the lariat branch acceptor site on beta-globin pre-mRNA splicing in vitro. *Nature*, 324, 589–591.

29. Garrey, S.M., Voelker, R. and Berglund, J.A. (2006) An extended RNA binding site for the yeast branch point-binding protein and
39. Zhang, Y., Goldstein, A.M. and Weiner, A.M. (1989) UACUAAC is a conserved functional domain involved in RNA binding and protein-protein interactions in human and Saccharomyces cerevisiae pre-mRNA splicing factor SF1. RNA, 4, 551–565.

40. Huang, T., Vilardebell, J. and Query, C.C. (2002) Pre-spliceosome formation in S. pombe requires a stable complex of SF1-U2AF(59)-U2AF(65). EMBO J., 21, 5516–5526.

41. Rain, J.C., Rafi, Z., Rhani, Z., Legrain, P. and Kramer, A. (1998) Conservation of functional domains involved in RNA binding and protein-protein interactions in human and Saccharomyces cerevisiae pre-mRNA splicing factor SF1. RNA, 4, 551–565.

42. Langford, C.J. and Gallwitz, D. (1983) Evidence for an intron–exon boundary signal between the third and fourth exons of the pre-mRNA of the human lamin A/C gene. EMBO J., 18, 4549–4559.

43. Kent, O.A., Ritchie, D.B. and Macmillan, A.M. (2005) Characterization of a U2AF-independent commitment complex (E') in the mammalian spliceosome assembly pathway. Mol. Cell. Biol., 25, 233–240.

44. Wang, X., Bruderer, S., Rafi, Z., Xue, J., Milburn, P.J., Kramer, A. and Rhani, Z. (1998) Polypyrimidine tract binding protein blocks the 5' splice site-dependent assembly of U2AF and the prespliceosomal E complex. Mol. Cell, 34, 395–403.

45. Nelson, K.K. and Green, M.R. (1989) Mammalian U2 snRNP has a T–G splice site-dependent assembly of U2AF and the prespliceosomal E complex. Mol. Cell. Biol., 9, 485–496.

46. Green, M.R. (1986) Pre-mRNA splicing. Annu. Rev. Genet., 20, 671–708.

47. Zhang, Y., Goldstein, A.M. and Weiner, A.M. (1989) UACUAAC is the preferred branch site for mammalian mRNA splicing. Proc. Natl Acad. Sci. USA, 86, 2752–2756.

48. Harris, N.L. and Senapathy, P. (1990) Distribution and consensus of branch point signals in eukaryotic genes: a computerized statistical analysis. Nucleic Acids Res., 18, 3015–3019.

49. Zhang, M.Q. (1998) Statistical features of human exons and their flanking regions. Hum. Mol. Genet., 7, 919–932.

50. Nelson, K.K. and Green, M.R. (1989) Mammalian U2 snRNP has a sequence-specific RNA-binding activity. Genes Dev., 3, 1562–1571.

51. Kuivenhoven, J.A., Weibusch, H., Pritchard, P.H., Funke, H., Benne, R., Assmann, G. and Kastelein, J.J. (1996) An intronic mutation in a lariat branchpoint sequence is a direct cause of an inherited human disorder (fish-eye disease). J. Clin. Invest., 98, 358–364.

52. Li, M. and Pritchard, P.H. (2000) Characterization of the effects of mutations in the putative branchpoint sequence of intron 4 on splicing within the human lecithin:cholesterol acyltransferase gene. J. Biol. Chem., 275, 18079–18084.

53. Di Leo, E., Panico, F., Tarugi, P., Battisti, C., Federico, A. and Calandra, S. (2004) A point mutation in the lariat branch point of intron 6 of NPC1 as the cause of abnormal pre-mRNA splicing in Niemann-Pick type C disease. Hum. Mutat., 24, 440.

54. Ketterling, R.P., Drost, J.B., Searinge, W.A., Liao, D.Z., Liu, J.Z., Kasper, C.K. and Sommer, S.S. (1999) Reported in vivo splice-site mutations in the factor IX gene: severity of splicing defects and a hypothesis for predicting deleterious splice donor mutations. Hum. Mutat., 13, 221–231.