An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers

Philip J. Smith¹, Chaolin Zhang¹,2, Jinhua Wang³, Shern L. Chew⁴, Michael Q. Zhang¹ and Adrian R. Krainer¹,*

¹Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA, ²Department of Biochemical Engineering, State University of New York at Stony Brook, NY 11794, USA, ³St Jude Children’s Research Hospital, Memphis, TN 38105, USA and ⁴Department of Endocrinology, St Bartholomew’s Hospital, Queen Mary, University of London, London EC1M 6BQ, UK

Received May 16, 2006; Revised and Accepted July 4, 2006

Numerous disease-associated point mutations exert their effects by disrupting the activity of exonic splicing enhancers (ESEs). We previously derived position weight matrices to predict putative ESEs specific for four human SR proteins. The score matrices are part of ESEfinder, an online resource to identify ESEs in query sequences. We have now carried out a refined functional SELEX screen for motifs that can act as ESEs in response to the human SR protein SF2/ASF. The test BRCA1 exon under selection was internal, rather than the 3'-terminal IGHM exon used in our earlier studies. A naturally occurring heptameric ESE in BRCA1 exon 18 was replaced with two libraries of random sequences, one seven nucleotides in length, the other 14. Following three rounds of selection for in vitro splicing via internal exon inclusion, new consensus motifs and score matrices were derived. Many winner sequences were demonstrated to be functional ESEs in S100-extract-complementation assays with recombinant SF2/ASF. Motif-score threshold values were derived from both experimental and statistical analyses. Motif scores were shown to correlate with levels of exon inclusion, both in vitro and in vivo. Our results confirm and extend our earlier data, as many of the same motifs are recognized as ESEs by both the original and our new score matrix, despite the different context used for selection. Finally, we have derived an increased specificity score matrix that incorporates information from both of our SF2/ASF-specific matrices and that accurately predicts the exon-skipping phenotypes of deleterious point mutations.

INTRODUCTION

Most eukaryotic transcripts comprise multiple short exons interrupted by much longer introns. Mature mRNA is generated through splicing of the pre-mRNA in a well-characterized process that involves exon recognition via intronic sequences at the 5' and 3' splice sites and branch site, removal of the intron and ligation of the exons (1). The information content provided by the splice sites is insufficient for accurate exon definition (2). Additional information is provided by cis-acting regulatory sequences that serve to enhance or repress splicing, and that may be exonic or intronic in nature (3).

Exonic splicing enhancers (ESEs) were first identified as regulators of alternative splicing (4) and later demonstrated to participate in the splicing of constitutive exons (5,6). Members of the SR protein family function both as general splicing factors and as regulators of alternative splicing, and they act in part through recognition of, and binding to, ESEs (4,7). SR proteins share a conserved domain structure consisting of one or two copies of an RNA-recognition motif (RRM) followed by a C-terminal domain highly enriched in arginine/serine dipeptides (RS domain) (8). The RRMs provide substrate specificity via sequence-specific RNA binding, whereas the RS domain is involved mainly in protein—protein interactions, with recent data suggesting an additional role in protein—RNA interactions (9,10). ESE-bound SR proteins function in exon definition by recruitment of the splicing machinery through their RS domains (11,12), and/or by

*cTo whom correspondence should be addressed. Tel: +1 5163678417; Fax: +1 5163678453; Email: krainer@cshl.edu

© The Author 2006. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org
antagonizing the action of nearby splicing silencer elements (13–15). In addition to their role in splicing, recent work points to a role for SR proteins in numerous steps of RNA metabolism. These include nuclear export, mRNA quality control and translation (reviewed in 16). Furthermore, a recent study provided evidence for the involvement of ESE-bound SR proteins in ensuring the correct linear order of exons in mature mRNA (17).

Many studies have demonstrated that a diverse array of sequences can function as ESEs. An important tool for ESE identification has been the employment of functional systematic evolution of ligands by exponential enrichment (SELEX) both in vivo (18) and in vitro (19–21). A refinement of functional SELEX experiments utilized the ability of recombinant SR proteins to complement SR protein-deficient S100 extracts, and thus derived SR-protein-specific sequence motifs (22,23). These consensus motifs are short (6–8 nt), degenerate, and were found to be frequently inactivated by disease-associated mutations that cause exon-skipping in vivo and in vitro (24,25). Nucleotide-frequency scoring matrices derived from the SELEX consensus motifs are available in a web-based program called ESEfinder (26) (http://rulai.cshl.edu/tools/ESE/). Use of ESEfinder has demonstrated that numerous disease-associated mutations that cause exon-skipping correlate with a reduction in high-score ESE motifs to below-threshold values (24,25,27–32).

The Burge and Chasin labs have employed ab initio computational approaches to identify ESE motifs (33,34). RESCUE-ESE identified hexamer motifs that were over-represented in exons with weak splice sites, and also in exons versus introns (33). Zhang and Chasin utilized a similar approach to identify octamers over-represented in internal non-versus introns (33). Zhang and Chasin also identifying a number of exonic-splicing silencers. Definitive classification of any given genomic variation as deleterious remains problematic (reviewed in 35) and it is likely that the current versions of ESE-prediction methodologies described earlier require further refinement. Our original SR-protein-specific functional SELEX was carried out in the context of an \textit{IGHM} (IgM) two-exon minigene pre-mRNA, and involved the replacement of a naturally occurring segment with ESE activity (36) in the 3 ‘-terminal M2 exon with a random 20 nt library (22,23). The natural enhancer, and by extension the sequences under selection, are adjacent to an exonic-splicing silencer, recently demonstrated to function through the binding of the splicing repressor polypyrimidine-tract-binding protein (PTB) (15). The original functional SELEX experiments were limited by the inability to accommodate all possible 20mers within an \textit{in vitro} splicing reaction. In this study, we have performed the SF2/ASF functional SELEX experiments in a different exonic context, namely a \textit{BRCA1} three-exon minigene pre-mRNA (24), to examine the effects of sequence context on motif derivation and functionality. The position of the enhancer within the middle exon makes this test system typical of the majority of naturally occurring ESEs. Furthermore, the shorter lengths of random sequence inserted made it possible to achieve total representation of all possible sequences available for selection. We demonstrate that the new SF2/ASF-specific motifs we identified are functional ESEs, both \textit{in vitro} and \textit{in vivo}, with motif scores correlating with levels of exon inclusion. Moreover, a reduction in SF2/ASF motif score correlates with the exon-skipping phenotypes associated with a number of disease-causing mutations.

**RESULTS**

**Identification of ESE motifs recognized by SF2/ASF under splicing conditions**

Our previous functional SELEX experiments utilized libraries of random sequence 20 nt in length, in the context of a two-exon IgM minigene construct (22,23). The derived SR-protein-specific matrices have proven to be very useful tools for ESE prediction. We wanted to refine the original SF2/ASF score matrix for a number of reasons: (1) to investigate the effects of sequence context on motif selection; (2) to ascertain whether there are differences in the motifs selected from internal rather than terminal exons; (3) to derive a more quantitative measurement of the threshold value, i.e. the score above which we consider a given motif to be significant and therefore a probable ESE; (4) to improve the robustness of the ESE predictions. We chose to introduce random libraries of 7 and 14 nt into the three-exon \textit{BRCA1} minigene, because the motifs recognized by SF2/ASF are 7mers (22). When we used the original 20mer library we did not know \textit{a priori} the length of the recognition motif; the new libraries make it possible to achieve complete representation in the context of an \textit{in vitro} splicing reaction, and they also avoid or reduce the complication of having multiple ESEs within each insert. The n7 library provides a fixed context, whereas the n14 library allows the influence of potential position effects and flanking-sequence upon motif functionality to be directly assessed.

\textit{In vitro} splicing conditions were optimized using previously described minigenes containing either the wild-type (WT) \textit{BRCA1} SF2/ASF-specific exon 18 ESE or the naturally occurring E1694X nonsense mutation (MT) (24), in both nuclear extract and S100 extract complemented by recombinant SF2/ASF. In addition, \textit{in vitro} splicing experiments were performed in which the WT SF2/ASF-specific ESE was replaced with individual 14-nt sequences, such as two tandem WT or MT heptamers, or a single WT or MT heptamer at a central position within the 14 nt sequence. The expected pattern of exon inclusion was observed in experiments carried out both in nuclear extract and in S100 extract complemented by recombinant SF2/ASF, namely predominant exon inclusion with constructs containing the WT ESE, and exon-skipping with MT-containing constructs (data not shown).

Two SELEX libraries, one with n7, and one with n14, randomized regions in exon 18 of \textit{BRCA1} in place of the SF2/ASF ESE heptamer, were constructed by overlap-extension PCR. Randomized regions of 7 and 14 nt allow complete sequence representation in standard \textit{in vitro} splicing reactions. The position of the ESE (beginning at nucleotide +4 in exon 18) required the engineering of a \textit{BaeI} restriction site within exon 17 for rebuilding the full-length minigene construct after each round of SELEX. The engineered construct was
tested by in vitro splicing and behaved in an identical manner
to the parental construct (data not shown). An additional con-
struct, utilized solely for cloning purposes (BRCA1 C), con-
taining the reciprocal BaeI site in exon 18 was generated to
allow rebuilding of the full-length construct. The SELEX
libraries were cloned into pCR-Blunt (Invitrogen), and
random clones sequenced. Sequence analysis revealed no
strong bias towards any of the four nucleotides, and therefore
the starting pools were optimal for the SELEX experiments.
The nucleotide composition for the n14 library was 25.8% 
A, 19.9% C, 27.2% G, 27.1% T (73 clones sequenced),
and 28.1% A, 18.9% C, 28.1% G, 24.9% T (55 clones sequenced)
for the n7 library. About $6 \times 10^4$ molecules of pre-mRNA
were used as the input for the SELEX experiments. A conser-
vative estimate of coverage of all possible 7mers and 14mers
in the n7 and n14 libraries, respectively, can be made based
upon the limiting nucleotide in the random pools. For both
libraries, the limiting nucleotide is C (18.9% in the
n7 library, 19.9% in the n14 library). For the n14 library all
possible 14mers should be represented at least once, with
the exception of C14, which is expected to occur
$6 \times 10^4 \times (0.199)^{14} = 0.92$ times. Coverage of all possible
heptamers is complete. For example, C7, the most infrequent
heptamer based upon the limiting nucleotide, is expected to
occur $6 \times 10^4 \times (0.189)^7 = 5 \times 10^4$ times.

Figure 1 illustrates the functional SELEX procedure. The
SELEX libraries were spliced in S100 extract complemented
by recombinant SF2/ASF. As controls, equivalent samples
from the SELEX libraries were spliced in nuclear extract,
and BRCA1 WT and MT minigenes were spliced under the
same conditions. Spliced exon-18-containing mRNAs were
recovered and rebuilt into full-length splicing constructs.
Three rounds of selection were performed with both libraries.
The winner pools from each round of selection were sub-
cloned into pCR-Blunt (Invitrogen), and random clones
sequenced. The results of splicing of the round-three winner
pools are shown in Figure 2A. As previously reported,
BRCA1 WT predominantly includes exon 18 in both nuclear
extract and in S100 extract complemented by SF2/ASF
(Fig. 2A, lanes 1 and 3), whereas exon 18 is predominantly
skipped when BRCA1 MT is spliced, especially in the
S100 + SF2/ASF sample (Fig. 2A, lane 6). Both the n7 and
n14 round-three winner pools undergo splicing with almost
complete inclusion of exon 18 (Fig. 2A, lanes 7–12). The
experiment was performed in triplicate and the reaction pro-
ducts quantified by phosphorimager analysis. The data were
normalized to the levels of splicing obtained with the
BRCA1 WT construct, and expressed as both normalized
exon inclusion [included mRNA/(included mRNA + skipped
mRNA)] (Fig. 2B), and normalized inclusive splicing
[included mRNA/(included mRNA + skipped mRNA +
pre-mRNA)] (Fig. 2C). Expression of the data in this way
allows both the level of exon inclusion and the overall splicing
efficiency afforded by the selected RNA pools to be compared.
The SELEX winner pools included exon 18 to a higher degree
than BRCA1 WT, and showed a greatly enhanced level of
splicing activity in response to SF2/ASF.

Sequencing of random clones from both winner pools
demonstrated a striking degree of both positive and negative
selection, reflected in the overall nucleotide composition
(Table 1). The round-three winner pools had very similar
GC contents of 70.96 and 69.64% for the n7 and n14 libraries,
respectively, increasing from 47% GC in the unselected
library pools. The increase in GC content is accounted for
by a decrease in the A content and a dramatic decrease in
T content. The IgM SF2/ASF SELEX-winner pool had a
GC content of 62%; however, the GC content of the IgM
starting pool was 58% (22), and therefore the increase
observed for the BRCA1 SELEX winners is more significant.

Consensus motifs and score matrices were derived from
the n7 and n14 round-three winner sequences. An alignment
step was not required for the n7 winners, due to the ESE position
being fixed by the length of the random sequence insert (see
below). The n14 round-three winners were aligned using
three different motif-finding algorithms: Gibbs sampler (37),
MEME (38) and DME (39). A fourth alignment was per-
formed by scoring the n14 winners with the n7-derived
matrix. The highest scoring 7mer from each winner was
then used to generate the consensus motif. The matrix
derived from alignment of the n14 winners scored with the
n7-derived matrix was subsequently found to be the most
accurate predictor of both in vitro and in vivo splicing when
compared with the n14 matrices derived using the alignments
generated by the motif-finding algorithms. Therefore, we have
limited our discussion to the results obtained with this
matrix, designated n14(n7). It should be noted that the motifs derived
using MEME and Gibbs were somewhat similar, whereas the
DME-derived motif resembled the motif derived from the
n14(n7) matrix. When used to predict in vitro splicing activity,
only the DME and n14(n7)-derived matrices resulted in
statistically significant correlations.

The aligned winner sequences and consensus motifs are
shown in Figure 3. The consensus motifs are relatively degene-
rate, as reported for the previously derived SR protein-specific
motifs (22,23). The winner sequences were then used to derive
score matrices according to the frequency of each nucleotide
at each position of the consensus motif, with an adjustment to take
into account the compositional bias of the initial random pools
(22,23). A second n7 matrix was derived, in which the initial
matrix was used to score the round-three winner sequences
plus three exonic nucleotides upstream of the ESE and six
nucleotides downstream (constant flanking regions). If this
resulted in a higher score for a given winner, the flanking
nucleotide(s) were included in a new alignment of that
winner, instead of the original 7mer motif. This second
matrix proved to be slightly less accurate at predicting experimen-
tal splicing, and therefore the original n7 matrix was
retained. Clones with the BRCA1 WT ESE sequence were
found in both the n7 unselected pool and rounds two and
three winner pools, and were not included in the sequence
analysis as they probably represent PCR contamination.

The scores of the n7 round-three winners ranged from 0.344
to 4.337, with a mean score of 2.649. The 66 clones sequenced
from the unselected (round-zero) pool had scores that range
from −4.637 to 3.529, with a mean score of −0.952. Only
three sequences in the round-zero pool had scores higher
than the mean of the winner pool, whereas 14 sequences in
the winner pool had scores higher than this mean, and all 27
winner clones had scores higher than the mean of the round-
zero pool. The scores of the round-zero and winner pools
were compared by plotting the receiver operating characteristic (ROC) curve (Supplementary Material, Fig. S1). This analysis is a measure of the sensitivity (number of high-scores in the winner pool) and specificity (number of high-scores in the random pool) of the score matrix at all possible threshold values, from the minimum to the maximum score from both pools. Discrimination of the pools was observed for all threshold values (Supplementary Material, Fig. S1).

The n14 round-three winner scores ranged from 1.259 to 6.372, with a mean score of 3.919. The 73 clones from the round-zero pool had a score range of 2.3467 to 4.410, with a mean score of 0.876. Only two sequences in the round-zero pool had scores higher than the mean of the winner pool, whereas 18 sequences in the winner pool had scores higher than this mean, and all 33 winner clones had scores higher than the mean of the unselected pool. Plotting the ROC curve for the n14 library demonstrated discrimination of the winner pool from the round-zero pool for all threshold values (Supplementary Material, Fig. S1). There is a significant bias for the highest-score motif being present at the beginning of the selected 14mer sequences. 13 out of the 33 winners have the highest score motif beginning at position 1.
indicating that there is selective pressure for functional ESEs being at this position, which corresponds to the position of the WT BRCA1 ESE (Supplementary Material, Fig. S2). The SELEX winners comprise functional ESEs

To investigate the functionality of the individual SELEX winner sequences, we tested the corresponding pre-mRNAs carrying individual sequences from the n7 and n14 round-three pools for their ability to include exon 18. The selected putative ESE motifs are in the same context as in the libraries used for functional SELEX, i.e. in place of the SF2/ASF-specific ESE of BRCA1 exon 18. A representative experiment showing splicing of 10 n7 and 10 n14 winner sequences is shown in Figure 4A and B, respectively. The pre-mRNAs were incubated in both nuclear extract and S100 extract complemented by recombinant SF2/ASF. Under these splicing-reaction conditions, the WT BRCA1 ESE is a weakly responsive SF2/ASF-specific ESE. As a control, we chose the level of enhancer function afforded by the BRCA1 WT ESE as the lower limit for ESE function, and normalized all of the data to this level. Significantly, the winner sequences demonstrated a clear enhancement of splicing in response to SF2/ASF in comparison with the BRCA1 WT construct (Fig. 4A and B), with many of the clones tested resulting in almost complete, and in a number of examples complete, exon inclusion.

Sixty-seven n7 and n14 winner sequences were spliced, in triplicate experiments, in S100 extract complemented by recombinant SF2/ASF, and the products of the splicing reactions quantified by phosphorimage analysis. The results were normalized to the BRCA1 WT control run in each experiment, and expressed as normalized inclusive splicing [included mRNA/(included mRNA + skipped mRNA + pre-mRNA)] (Supplementary Material, Fig. S3). Sixty-six of the sequences supported levels of exon inclusion and inclusive splicing levels greater than that of the BRCA1 WT control. The mean inclusive splicing level for the pooled data was 5.19 ± 3.17 for the n7 and n14 winners. Significantly, the single winner sequence that spliced poorly was found to contain a below-threshold (see below) motif when scored with the n7-derived matrix and the n14-derived matrix: 0.466 and 0.345, respectively.

Table 1. Nucleotide composition of the SELEX pools from the n7 and n14 libraries

<table>
<thead>
<tr>
<th>Round</th>
<th>A</th>
<th>C</th>
<th>G</th>
<th>T</th>
<th>Number of clones</th>
</tr>
</thead>
<tbody>
<tr>
<td>n7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>28.10</td>
<td>18.90</td>
<td>28.10</td>
<td>24.90</td>
<td>66</td>
</tr>
<tr>
<td>1</td>
<td>26.33</td>
<td>23.81</td>
<td>29.13</td>
<td>20.73</td>
<td>51</td>
</tr>
<tr>
<td>2</td>
<td>21.85</td>
<td>32.77</td>
<td>33.19</td>
<td>12.19</td>
<td>34</td>
</tr>
<tr>
<td>3</td>
<td>18.57</td>
<td>36.19</td>
<td>34.76</td>
<td>10.48</td>
<td>27</td>
</tr>
<tr>
<td>n14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>25.80</td>
<td>19.90</td>
<td>27.20</td>
<td>27.10</td>
<td>73</td>
</tr>
<tr>
<td>1</td>
<td>27.65</td>
<td>26.73</td>
<td>29.49</td>
<td>16.13</td>
<td>31</td>
</tr>
<tr>
<td>2</td>
<td>21.85</td>
<td>25.21</td>
<td>35.50</td>
<td>17.44</td>
<td>34</td>
</tr>
<tr>
<td>3</td>
<td>20.76</td>
<td>31.92</td>
<td>37.72</td>
<td>9.60</td>
<td>33</td>
</tr>
</tbody>
</table>
In contrast, when a total of 45 random pre-mRNAs from the n7 and n14 unselected round-zero pools were tested, the levels of splicing observed were much lower. A representative experiment showing splicing of 10 n14 round-zero sequences spliced in S100 extract complemented by recombinant SF2/ASF is shown in Figure 5A. A number of sequences resulted in inclusion of exon 18 at levels above the BRCA1 WT control (e.g. Fig. 5A, clones 9 and 10). Significantly, many of the sequences that spliced well were subsequently found to contain high-score motifs. The experiments were performed in triplicate, and the mean inclusive splicing levels for each sequence are shown in Figure 5B. The data in Figures 4C and 5B are plotted with the same ordinate scale to allow easier comparison. The mean inclusive splicing level for the n7 and n14 round-zero sequences was 1.54 ± 1.21, significantly lower than the splicing observed with the round-three winner sequences (P < 10⁻¹⁰, two-sample t-test).

Figure 3. Analysis of the SF2/ASF-selected sequences. Sequence alignment and identification of consensus motifs from the n7 (A) and n14 round-three winner sequences (B). The n14 sequences were aligned on the basis of the highest score motif for each sequence. Nucleotides matching the consensus are shown as white on a black background, mismatched nucleotides are not shaded. The consensus shown is only an approximation that indicates the most frequent nucleotide(s) at each position. The lower case c at position 4 of the n7 consensus denotes a slight preference for this nucleotide over A and G, which occur at similar frequencies. R, purine; S, C or G; M, A or C. Pictogram representations (62) of the n7-derived consensus motif (C) and n14-derived consensus motif (D). The diagrams show the frequency of each nucleotide at each position of the heptamer consensus, adjusted for the compositional bias of the initial pool. The height of each letter is proportional to its frequency; black and grey letters indicate higher and lower than background frequencies, respectively. T is used instead of U for convenience.
Figure 4. The winner sequences are functional SF2/ASF-dependent ESEs. In vitro splicing of 10 n7 (A) and 10 n14 (B) round-three winner sequences in HeLa nuclear extract, S100 extract alone and S100 complemented by recombinant SF2/ASF. The structures of the precursor, intermediates and products are indicated next to the autoradiograms. (C) Phosphorimage analysis of in vitro splicing, in S100 extract complemented by recombinant SF2/ASF, of 67 n7 and n14 round-three winner sequences. The experiment was performed three times and the data expressed as normalized inclusive splicing [included mRNA/(included mRNA + skipped mRNA + pre-mRNA)]. The horizontal white line represents the BRCA1 WT level of inclusive splicing (1), error bars the standard error.
protein as the predominant skipping of exon 7 results in an unstable protein (SMN\(\Delta 7\)) (43). Three round-three winner sequences from the n7 library and two round-zero sequences from the starting pool were inserted in place of the SMN1 exon 7 ESE by overlap-extension PCR. The SMN clones were incubated in nuclear extract (Fig. 6), with SMN1 and SMN2 serving as positive and negative controls, respectively. All of the winner sequences promoted levels of exon inclusion that were greater than those observed for SMN1 (Fig. 6A, lanes 5−7 compared with lane 1). In contrast, the two round-zero clones were similar to SMN2, resulting in predominant skipping of exon 7 (Fig. 6A, lanes 3 and 4). The experiment was performed in triplicate and the levels of exon inclusion for each construct normalized to those of SMN1 (Fig. 6B).

Derivation of motif-score threshold values and matrix analysis

Motif scores are considered potentially significant if they are above a defined threshold value, in which case they are classified as high-score motifs. The default thresholds in ESEfinder were defined as the median of the highest scores for each sequence in a set of 30 randomly chosen 20-nt sequences from the starting pool used for the previous functional SELEX experiments (26). This statistical value gives a good differentiation between the starting pool and the SELEX winners in terms of scores, but does not take into consideration the splicing activity associated with different motif scores. We chose to derive a threshold value based on the experimentally determined extent of splicing of individual BRCA1 SELEX clones. The n7- and n14-derived matrices were used to score the 112 clones that had been spliced in vitro in S100 extracts complemented with SF2/ASF. The amount of exon inclusion afforded by the WT BRCA1 ESE was selected as the lower limit for positive enhancer activity, and all data were normalized to this level. The experimental threshold was defined as the minimum motif score that results in no false positives. In other words, this threshold means that all of the tested high-score clones give levels of exon inclusion at least equal to the WT ESE.

The n14 clones contain eight 7mers within each winner sequence. The highest scoring 7mer from each clone was counted as the putative ESE motif. Plotting in vitro splicing data against motif score (Fig. 7) allowed experimental threshold values of 1.121 and 1.748 to be derived for the n7 and n14 matrices, respectively. It should be noted that scores generated from different matrices are not numerically comparable, because of the different degree of matrix degeneracy. Using this stringent criterion for threshold derivation results in highly accurate segregation of low motif-score clones that do not enhance splicing (lower left quadrants, Fig. 7A and B), from high motif-score clones that result in above WT levels of exon inclusion (upper right quadrants). Using these threshold values, only nine (of the 112 tested) clones behaved as enhancers that were not predicted by the n7 and n14 matrices. This indicates that scores generated from different matrices are not numerically comparable, because of the different degree of matrix degeneracy. Using this stringent criterion for threshold derivation results in highly accurate segregation of low motif-score clones that do not enhance splicing (lower left quadrants, Fig. 7A and B), from high motif-score clones that result in above WT levels of exon inclusion (upper right quadrants). Using these threshold values, only nine (of the 112 tested) clones behaved as enhancers that were not predicted by the n7 and n14 matrices. For both the n7 and n14 matrices there is a highly significant correlation between motif score and observed enhancer activity (Table 2). Significantly, the n7 matrix scored the WT BRCA1 ESE as above threshold (1.225) and the E1694X nonsense mutation as below threshold (−0.128). The n14 matrix correctly
ascertained a below-threshold score to E1694X, but did not score the WT ESE as above-threshold. These results suggest that the n7-derived matrix is the more robust predictor of ESE potential.

The matrices and experimentally derived thresholds were then tested against an independent panel of SMN constructs that were spliced in vivo (40). In that study, 30 SMN minigene constructs containing either the WT SMN1 exon 7 SF2/ASF-dependent ESE, or mutations of the ESE motif, were transiently transfected into HEK293 cells and exon inclusion/skipping measured by semi-quantitative RT–PCR. We compiled the results into a dataset presented as scatter plots to demonstrate the relationship between exon 7-inclusion and matrix score for the n7 (Fig. 8A) and n14 (Fig. 8B) matrices, respectively. As was observed for the in vitro BRCA1 splicing data, an accurate segregation of low-score clones that do not include exon 7 (lower left quadrants) from high-score clones that demonstrate high levels of exon inclusion (upper right quadrants) was achieved. The n7 matrix was again slightly more accurate in terms of predicting enhancer activity than the n14 matrix (23/30 correct predictions versus 21/30), and correctly predicted the splicing activity of both SMN1 and SMN2 (Fig. 8A). In addition, as mentioned earlier, the WT BRCA1 ESE and the version with the E1694X nonsense mutation have high and low scores, respectively, with the n7 matrix. In this heterologous assay, in which these elements are tested in the context of an SMN minigene, the BRCA1 WT ESE construct includes exon 7 at levels greater than SMN2 (although below our 80% exon-inclusion threshold), whereas the BRCA1 MT ESE results in complete exon-skipping. There is a significant correlation between motif scores and exon 7 inclusion (Table 2). The results are in accordance with those obtained with the SF2/ASF matrix in ESEfinder (40), and provide further evidence for exon 7 skipping in SMN2 being a consequence of loss of a functional SF2/ASF ESE.

We repeated the matrix analysis utilizing other threshold values based on statistical discrimination of the winner pools from the unselected round-zero sequences. The n14 round-zero sequences were used because they represent a large number of random sequences and, as described earlier, exhibit no nucleotide-composition bias. We calculated a threshold value corresponding to the 60th percentile value of the highest scores from the 73 round-zero clones. This
threshold, when applied to the \textit{BRCA1 in vitro} splicing data (Table 2) and \textit{in vivo} \textit{SMN} splicing experiments (Table 2) gave results that were in good accordance with the results obtained with the experimentally derived thresholds. It would appear that the ability of the matrices to predict the enhancer activity of a given motif is due less to the precise threshold value applied than to the discriminatory power of the matrices.

To investigate the degree of overlap between the n7 and n14 matrices, we scored all possible heptamers (16 384 sequences) with both matrices using the thresholds described earlier. The number of heptamers recognized as high scores by both matrices had a highly significant degree of overlap. For example, when the statistically derived thresholds were used, 81% of the heptamers recognized as high score by the n14 matrix were also high score with the n7 matrix, and 79% of the heptamers recognized as high score by the n7 matrix were also high score with the n14 matrix. Heptamer scores with the n14 matrix (\(P < 10^{-10}\), Fisher’s exact test, for the comparison of the expected number of overlapping high scores with the observed number of overlapping high scores).

**An increased specificity score matrix for SF2/ASF**

An important objective of this study was the derivation of an SF2/ASF score matrix with increased specificity. The current SF2/ASF matrix in ESEfinder generates a number of false positive high-scores, which may be at least in part a consequence of the context from which the motifs were derived. Increasing the specificity of the score matrix will enable more robust ESE prediction. A score matrix that incorporates information from both the original IgM SF2/ASF experiments and the \textit{BRCA1} SELEX experiments should allow recognition of ESE motifs capable of functioning in multiple exonic contexts.

A combined IgM-\textit{BRCA1} SF2/ASF-specific score matrix was created. The \textit{BRCA1} n7 matrix was chosen, as it gave the most accurate predictions of both \textit{in vitro} and \textit{in vivo} splicing activity when compared with the n14 matrix. We scored all possible heptamers (16 384) with both the n7 matrix and the ESEfinder SF2/ASF matrix. For increased specificity we used the original statistically derived IgM SF2/ASF matrix threshold (26), and the threshold corresponding to the 60th percentile of the \textit{BRCA1} n14 R0 highest scores. Heptamer motifs that were recognized as high-scores by both matrices were used to generate the combined matrix. This analysis resulted in 395 overlapping high-score sequences, compared with the 678 IgM ESEfinder high-scores. Given the percentage of all possible heptamers that are high-score motifs with the n7 matrix, by chance one would expect only 91 overlapping sequences, so the observed degree of overlap is highly significant (\(P < 10^{-10}\), Fisher’s exact test). A consensus motif (Fig. 9A) and score matrix were derived from the overlapping high-scores, adjusted for the genomic exonic nucleotide...
frequency as the background (25.9691% A, 25.0262% C, 25.4594% G, 23.5453% T, derived from a large set of human internal protein-coding exons) (44). It is highly probable that both the IgM and the BRCA1-derived matrices predict a number of motifs that function only in the context from which they were derived. Therefore the threshold for the combined matrix was derived statistically, rather than using the experimental threshold based on BRCA1 in vitro splicing (1.867, equal to the 70th percentile of the highest score from the BRCA1 R0 clones). About 57.5% of the combined matrix high-score motifs are recognized by the ESEfinder SF2/ASF matrix and 81.4% are recognized by the n7 matrix.

As an independent assay of the robustness of the new matrix, we scored the panel of SMN mutants analyzed in Figure 8, and correlated motif scores with the observed in vivo splicing activity. The results are shown as a scatter plot to demonstrate the relationship between SMN exon 7 inclusion and motif score (Fig. 9B). The combined matrix correctly predicts the ESE functionality of 25/30 of the motifs, including both SMN1 and SMN2. In addition, BRCA1 WT and the E1694X nonsense mutation have high and low scores, respectively. The correlation between motif score and exon 7 inclusion is highly significant (Pearson correlation coefficient of 0.72; 95%CI 0.48–0.86).

ESE inactivation by point mutations is a documented mechanism of disease (reviewed in reference 3). For example, the BRCA1 nonsense mutation E1694X and the translationally silent C → T difference between SMN1 and SMN2 disrupt SF2/ASF-specific ESEs, causing skipping of the exons harboring these mutations and progression to breast cancer and SMA, respectively (24,25,40). We also analyzed a set of 73 exonic single-base substitutions in human genes that cause exon-skipping in vivo (3,45). This set was created from a list of 50 exonic mutations documented to cause
in vivo (45), and 42 exonic mutations associated with altered splicing in vivo (3). The list was filtered to remove mutations that cause increased exon inclusion, and mutations that fall in the last two nucleotides of an exon, overlapping the 5′ splice-site. We scored the WT and mutant sequences for the presence of above-threshold (>1.867) motifs using the IgM-BRCA1 combined matrix. Table 3 summarizes the results obtained. For completeness we have included the results obtained with the score matrices previously derived for SRp40, SRp55 and SC35 (22,23,26). The motifs recognized by SF2/ASF and SRp40 are heptamers, SRp55 recognizes hexamers and SC35 octamers. Therefore, we scored 13mers for SF2/ASF and SRp40, 11mers for SRp55 and 15mers for SC35, with the mutant nucleotide at the central position. Twenty of the 73 WT sequences contained above-threshold SF2/ASF motifs that were reduced in the mutant sequences, and of these, 15 were reduced below the threshold. In contrast, only seven sequences had higher scores in the mutant than the WT sequence. To eliminate the background from sequences that lack putative SF2/ASF motifs in both WT and mutant versions, we compared only the sequences with a high-score motif in either the WT version, the mutant version, or both. The highest score motif in each of these sequences was compared with the highest score motif in the corresponding WT or mutant sequence. This analysis revealed a statistically significant difference in SF2/ASF motif scores between the WT and mutant sequences (P < 0.01, Fisher’s exact test). The combined output of the four SR protein-specific matrices resulted in the prediction of 23 mutant sequences having loss of ESE function, compared with the creation of 12 putative ESEs in the mutants. The effects of mutations in putative SF2/ASF ESEs appear to be more detrimental than mutations occurring in putative ESEs for the other three SR proteins.

Two ab initio computational ESE prediction methods have recently been developed by the Burge and Chasin labs (33,34). We reported that the motifs recognized by ESEfinder do not overlap above the chance expectation with those predicted by RESCUE-ESE or the PESEs of Zhang and Chasin, with the exception of a significant overlap in the number of SF2/ASF high-score motifs in the set of PESEs (44). We hypothesized that the exon-skipping mutants predicted to be deleterious by our functionally derived matrices may be different from those recognized by the computational prediction methodologies, which presumably include ESEs recognized by a wide variety of proteins. We scored the 73 WT and mutant sequences for the presence of RESCUE-ESE motifs and PESEs or PESSs (PESX) (33,34). We scored 11mer sequences for the presence of RESCUE-ESE hexamers (http://genes.mit.edu/burgelab/escape/index.html), and 15mer sequences for the presence of PESX octamers (http://cubweb.biology.columbia.edu/pesx/). The Burge lab recently reported the results of an in vivo screen for motifs that function as ESSs (46). We therefore also scored the WT and mutant sequences (11mers) for the presence of motifs that match the putative ESS motifs (FAS-hex2 hexamers, http://genes.mit.edu/fas-ess/). Supplementary Material, Tables S1 and S2 summarize the results of this analysis. RESCUE-ESE, FAS-hex2 and PESX were able to predict the exon-skipping behavior of 25, 21 and 32 out of the 73 mutations, respectively. As for the ESEfinder analysis, predictions were taken to correlate with exon-skipping if there was an overall loss of putative ESEs, or gain of ESSs. Without experimental validation, it is not possible to conclude which of multiple motifs within any given sequence is functional. Overall, RESCUE-ESE predicted the loss of 25 ESEs and the creation of 15 new ESEs. FAS-hex2 predicted the creation of 21 new ESSs and the loss of only four ESS motifs in the mutant sequences. The predictions of exon-skipping with PESX were robust; there was a net loss of 20 PESE-containing sequences in the mutants, with the creation of six PESEs. The mutant sequences contain 15 new PESSs and lose only one. Significantly, most of the mutants whose ESE functionality was correctly predicted by RESCUE-ESE and PESE were different from those predicted by ESEfinder.

The combined output of scoring the set of exon-skipping mutants with all four methods is shown in Table 4. In total, the exon-skipping phenotype of the 73 point mutations is predicted correctly 61 times, a significant increase over the ability of any one of the methods alone. None of the mutants were predicted to skip by all three ESE predictors. It should be noted that a direct comparison of the three methods is not possible, as ESEfinder searches for specific motifs for only four of the SR proteins, whereas RESCUE-ESE and PESE are not protein-specific. Including more SR protein score matrices should increase the number of mutations amenable to this type of analysis.

SF2/ASF motif frequencies in human protein-coding exons

Functional ESEs should be present at a higher frequency in exons compared with their flanking introns, as we observed for the original SR-protein-specific matrices in ESEfinder (44) and as was found for the RESCUE-ESE hexamers and PESE octamers (33,34). We created a dataset of 16 635 constitutively spliced internal protein-coding exons ≥106 nt in length (100 consecutive heptamers), plus 100 nt each of flanking upstream and downstream intronic sequence. We created a second dataset of 5041 exons that undergo alternative splicing. To standardize for differences in exon length, we created composite 100-nt exon sequences consisting of 25 nt from each end plus 50 nt from the center. We scored the datasets with the IgM-BRCA1 matrix for the presence of high-score motifs (above the threshold of 1.867).

SF2/ASF scores were calculated at each position and high-score motif frequencies plotted at the first position of the motif (Fig. 10). As expected, given the SF2/ASF consensus motif, there is a sharp decrease in the frequency of SF2/ASF motifs found in the areas corresponding to the polypyrimidine tract and 5′ splice-site. ESE motif frequencies were approximately constant within the exons, as we previously observed with the ESEfinder high-score motif frequency distributions within exons (44). Comparison of the ESE-motif frequencies revealed that the higher density of ESE motifs in exons than in introns was statistically significant for both the constitutive and alternatively spliced exons (paired t-test, P < 10−10). Analysis of the datasets with ESEfinder SF2/ASF (26) also revealed a significant difference in ESE-motif frequency between exons and introns (data not shown), in accordance with our previous data (44).
Table 3. Disease-associated mutations that cause exon-skipping correlate with a reduction in SF2/ASF motif scores

<table>
<thead>
<tr>
<th>Gene*</th>
<th>Mutationb</th>
<th>sub.</th>
<th>SF2/ASF</th>
<th>SRp40</th>
<th>SRp55</th>
<th>SC35</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Missense mutations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ADAT</strong></td>
<td>A215T</td>
<td>G→A</td>
<td>↓↓</td>
<td>40 ↓</td>
<td><strong>ADAT</strong></td>
<td>R142X</td>
</tr>
<tr>
<td><strong>ATM</strong></td>
<td>E2032K</td>
<td>G→A</td>
<td>↓↑</td>
<td>40 ↑</td>
<td><strong>ATM</strong></td>
<td>R645X</td>
</tr>
<tr>
<td><strong>ATP7A</strong></td>
<td>G1302R</td>
<td>G→A</td>
<td>↓↑</td>
<td>40 ↑</td>
<td><strong>ATP7A</strong></td>
<td>E1694X</td>
</tr>
<tr>
<td><strong>BRCAl</strong></td>
<td>E1649K</td>
<td>G→A</td>
<td>↓</td>
<td>40 ↓</td>
<td><strong>BRCAl</strong></td>
<td>E60X</td>
</tr>
<tr>
<td><strong>CFTR</strong></td>
<td>G85E</td>
<td>G→A</td>
<td>↓↓</td>
<td>40 ↓</td>
<td><strong>CFTR</strong></td>
<td>R75X</td>
</tr>
<tr>
<td><strong>CFTR</strong></td>
<td>D565G</td>
<td>A→G</td>
<td>↑↑</td>
<td>40 ↑</td>
<td><strong>CFTR</strong></td>
<td>R553X</td>
</tr>
<tr>
<td><strong>F8</strong></td>
<td>R1997W</td>
<td>C→T</td>
<td>↓↓</td>
<td>40 ↓</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>HEXB</strong></td>
<td>P404L</td>
<td>C→T</td>
<td>↓</td>
<td>55 ↓</td>
<td><strong>HEXB</strong></td>
<td>E1211X</td>
</tr>
<tr>
<td><strong>HPRT</strong></td>
<td>G40V</td>
<td>G→T</td>
<td>↑↑</td>
<td>55 ↑</td>
<td><strong>HPRT</strong></td>
<td>E1987X</td>
</tr>
<tr>
<td><strong>HPRT</strong></td>
<td>R48H</td>
<td>G→A</td>
<td>↓</td>
<td>55 ↓</td>
<td><strong>HPRT</strong></td>
<td>F8</td>
</tr>
<tr>
<td><strong>HPRT</strong></td>
<td>A161E</td>
<td>C→A</td>
<td>↓</td>
<td>40 ↓</td>
<td><strong>HPRT</strong></td>
<td>FANCC</td>
</tr>
<tr>
<td><strong>HPRT</strong></td>
<td>G180E</td>
<td>G→A</td>
<td>↓</td>
<td>40 ↓</td>
<td><strong>HPRT</strong></td>
<td>FBNI</td>
</tr>
<tr>
<td><strong>HPRT</strong></td>
<td>G180V</td>
<td>G→T</td>
<td>↓</td>
<td>55 ↓</td>
<td><strong>HPRT</strong></td>
<td>HMGCL</td>
</tr>
<tr>
<td><strong>HPRT</strong></td>
<td>G182K</td>
<td>G→A</td>
<td>↓</td>
<td>55 ↓</td>
<td><strong>HPRT</strong></td>
<td>E30X</td>
</tr>
<tr>
<td><strong>HPRT</strong></td>
<td>P184L</td>
<td>C→T</td>
<td>↓↓</td>
<td>40 ↓</td>
<td><strong>HPRT</strong></td>
<td>E47X</td>
</tr>
<tr>
<td><strong>HPRT</strong></td>
<td>D194Y</td>
<td>G→T</td>
<td>↑↓</td>
<td>40 ↑</td>
<td><strong>HPRT</strong></td>
<td>R51X</td>
</tr>
<tr>
<td><strong>HPRT</strong></td>
<td>E197K</td>
<td>G→A</td>
<td>↓</td>
<td>55 ↓</td>
<td><strong>HPRT</strong></td>
<td>K55X</td>
</tr>
<tr>
<td><strong>HPRT</strong></td>
<td>E197V</td>
<td>A→T</td>
<td>↑↑</td>
<td>55 ↑</td>
<td><strong>HPRT</strong></td>
<td>C66X</td>
</tr>
<tr>
<td><strong>HPRT</strong></td>
<td>D201V</td>
<td>A→T</td>
<td>↓</td>
<td>55 ↓</td>
<td><strong>HPRT</strong></td>
<td>K103X</td>
</tr>
<tr>
<td><strong>IVD</strong></td>
<td>R21C</td>
<td>C→T</td>
<td>↓</td>
<td>55 ↓</td>
<td><strong>IVD</strong></td>
<td>G190X</td>
</tr>
<tr>
<td><strong>IVD</strong></td>
<td>R1P</td>
<td>G→T</td>
<td>↓</td>
<td>55 ↓</td>
<td><strong>IVD</strong></td>
<td>L125X</td>
</tr>
<tr>
<td><strong>IVD</strong></td>
<td>D40N</td>
<td>G→A</td>
<td>↓</td>
<td>40 ↓</td>
<td><strong>IVD</strong></td>
<td>G180X</td>
</tr>
<tr>
<td><strong>MLHI</strong></td>
<td>R659P</td>
<td>G→C</td>
<td>↓</td>
<td>40 ↓</td>
<td><strong>MLHI</strong></td>
<td>E182X</td>
</tr>
<tr>
<td><strong>MLHI</strong></td>
<td>R659L</td>
<td>G→T</td>
<td>↓</td>
<td>40 ↓</td>
<td><strong>MLHI</strong></td>
<td>E197X</td>
</tr>
<tr>
<td><strong>PDHA1</strong></td>
<td>A175T</td>
<td>G→A</td>
<td>↓↓</td>
<td>40 ↓</td>
<td><strong>PDHA1</strong></td>
<td>Y198X</td>
</tr>
<tr>
<td><strong>PMM2</strong></td>
<td>E139K</td>
<td>G→A</td>
<td>↓</td>
<td>40 ↓</td>
<td><strong>PMM2</strong></td>
<td>MLHI</td>
</tr>
<tr>
<td><strong>RHAG</strong></td>
<td>G380V</td>
<td>G→T</td>
<td>↓</td>
<td>40 ↓</td>
<td><strong>RHAG</strong></td>
<td>CFTR</td>
</tr>
<tr>
<td><strong>Silent mutations</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>APC</strong></td>
<td>R623R</td>
<td>G→T</td>
<td>↓</td>
<td>40 ↓</td>
<td><strong>APC</strong></td>
<td>OAT</td>
</tr>
<tr>
<td><strong>AR</strong></td>
<td>S888S</td>
<td>C→T</td>
<td>↓</td>
<td>55 ↓</td>
<td><strong>AR</strong></td>
<td>OAT</td>
</tr>
<tr>
<td><strong>CP27A1</strong></td>
<td>G112G</td>
<td>G→T</td>
<td>↓</td>
<td>35 ↓</td>
<td><strong>CP27A1</strong></td>
<td>PROSI</td>
</tr>
<tr>
<td><strong>FBNI</strong></td>
<td>I21181I</td>
<td>C→T</td>
<td>↓</td>
<td>35 ↓</td>
<td><strong>FBNI</strong></td>
<td>TG</td>
</tr>
<tr>
<td><strong>HPRT</strong></td>
<td>F199F</td>
<td>C→T</td>
<td>↓</td>
<td>40 ↓</td>
<td><strong>HPRT</strong></td>
<td>WAS</td>
</tr>
<tr>
<td><strong>ITGB3</strong></td>
<td>T420T</td>
<td>G→A</td>
<td>↑↑</td>
<td>40 ↓</td>
<td><strong>ITGB3</strong></td>
<td>MLHI</td>
</tr>
<tr>
<td><strong>NF1</strong></td>
<td>K354K</td>
<td>G→A</td>
<td>↑↑</td>
<td>40 ↓</td>
<td><strong>NF1</strong></td>
<td>OAT</td>
</tr>
<tr>
<td><strong>PAH</strong></td>
<td>V399Y</td>
<td>A→T</td>
<td>↑↑</td>
<td>40 ↑</td>
<td><strong>PAH</strong></td>
<td>OAT</td>
</tr>
<tr>
<td><strong>PDHA1</strong></td>
<td>G185Q</td>
<td>G→A</td>
<td>↓</td>
<td>55 ↓</td>
<td><strong>PDHA1</strong></td>
<td>OAT</td>
</tr>
<tr>
<td><strong>RET</strong></td>
<td>I6471</td>
<td>C→T</td>
<td>↓</td>
<td>55 ↓</td>
<td><strong>RET</strong></td>
<td>OAT</td>
</tr>
<tr>
<td><strong>SMN1</strong></td>
<td>F280F</td>
<td>C→T</td>
<td>↓</td>
<td>55 ↓</td>
<td><strong>SMN1</strong></td>
<td>SC35</td>
</tr>
<tr>
<td><strong>TNFRSF5</strong></td>
<td>T136T</td>
<td>A→T</td>
<td>↓↓</td>
<td>40 ↓</td>
<td><strong>TNFRSF5</strong></td>
<td>Total</td>
</tr>
</tbody>
</table>

Seventy-three point mutations that cause exon-skipping and disease were scored with the combined IgM-BRCAl SF2/ASF motif score matrix and the score matrices for SRp40, SRp55 and SC35 from ESEfinder. Sequence motifs for the same or for a different SR protein can overlap. Only the WT or MT sequence motifs with scores greater than or equal to the threshold for the corresponding SR protein were considered. Downward arrows denote a reduction, or elimination (bold) of the motif score as a result of the mutation. Upward arrows denote a higher score in the MT than in the WT, with bold denoting an increase from below to above the threshold (creation of new putative ESE motif).

**Genes and their encoded proteins are as follows: ADAT, adenosine deaminase; APC, adenomatous polyposis coli; AR, androgen receptor; ATM, ataxia telangiectasia mutated; ATP7A, ATPase, Cu2+ transporting, α-polypeptide; BRCAl, breast cancer 1, early onset; CFTR, cystic fibrosis transmembrane conductance regulator; CP27A1, sterol-27-hydroxylase; DMD, dystrophin; F8, coagulation factor VIII; FANCC, Fanconi anemia, complementation group C; FBNI, fibrillin 1; HEXB, hexosaminidase B, β-polypeptide; HMGCL, 3-hydroxy-3-methylglutaryl Coenzyme A lyase; HPRT, hypoxanthine phosphoribosyltransferase 1; IDUA, α-L-iduronidase; ITGB3, integrin-β3; IVD, isovaleryl coenzyme A dehydrogenase; MLHI, mutl homologue; NF1, neurofibromin 1; OAT, ornithine amino-transferase; PAH, phenylalanine hydrolase; PDHA1, pyruvate dehydrogenase (lipoamide) α1; PMM2, phosphomannomutase 2; PROSI, protein S-α; RHAG, Rhesus blood group-associated glycoprotein; SMN1, survival of motor neuron 1; TG, thyroglobulin; TNFRSF5, tumor-necrosis factor receptor superfamily, member 5 (CD40); WAS, Wiskott–Aldrich syndrome.

1The specific proteins are identified by the WT amino acid in the one-letter code, followed by the residue number in the protein sequence and the MT amino acid (X denotes one of the three nonsense codons) as it would be in the absence of exon-skipping.

2Nucleotide substitution.

3The number of high-score motifs that are reduced in the WT sequences compared with the WT.

4The number of high-score motifs that are increased or created in the MT sequences compared with the WT.

5Comparison of the highest score in the WT sequence with the highest score in the MT when a high score is present in either the WT or MT sequence (Fisher’s exact test).
DISCUSSION

The prototypical SR protein SF2/ASF (47,48) functions in both the regulation of alternative splicing and in constitutive splicing (6,49), and participates in other cellular processes, including nonsense-mediated mRNA decay, maintenance of genomic stability, RNA export and translation (50–53). We have performed a functional SELEX screen for sequence motifs that are recognized as ESEs by SF2/ASF. Our data support the validity of the SR protein-specific ESE matrix derived in the context of the IgM two-exon minigene (22,23), and provide further information regarding the motifs functionally recognized by SF2/ASF.

We carried out parallel SELEX experiments utilizing one random library of seven nucleotides, and one of 14 nucleotides, in place of the naturally occurring SF2/ASF-specific ESE in BRCA1 exon 18 (24). In vitro splicing conditions in S100 extracts complemented by recombinant SF2/ASF were optimized to minimize the possibility of non-functional motifs passing through the ESE screen. Three rounds of functional SELEX were performed, resulting in the derivation of consensus motifs and score matrices from both libraries. There was a highly significant overlap in the motifs derived from the two libraries, with both winner pools being greatly enriched for C and G nucleotides. The consensus motifs are

Table 4. Prediction of the effects of exon-skipping disease-associated point mutations by scoring for the presence of ESEfinder, RESCUE-ESE, FAS-hex2 and PESX motifs

<table>
<thead>
<tr>
<th>Gene</th>
<th>Mutation</th>
<th>Exon-skipping predicted</th>
<th>Gene</th>
<th>Mutation</th>
<th>Exon-skipping predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADA</td>
<td>A215T</td>
<td>Yes</td>
<td>ADA</td>
<td>R142X</td>
<td>Yes</td>
</tr>
<tr>
<td>ATM</td>
<td>E2022K</td>
<td>Yes</td>
<td>ATM</td>
<td>R645X</td>
<td>Yes</td>
</tr>
<tr>
<td>ATPTA</td>
<td>G1302R</td>
<td>Yes</td>
<td>ATPTA</td>
<td>E1694X</td>
<td>Yes</td>
</tr>
<tr>
<td>BRCA1</td>
<td>E1694K</td>
<td>Yes</td>
<td>CFTR</td>
<td>E060X</td>
<td>Yes</td>
</tr>
<tr>
<td>CFTR</td>
<td>G85E</td>
<td>Yes</td>
<td>CFTR</td>
<td>R75X</td>
<td>Yes</td>
</tr>
<tr>
<td>CFTR</td>
<td>D565G</td>
<td>Yes</td>
<td>CFTR</td>
<td>R553X</td>
<td>Yes</td>
</tr>
<tr>
<td>H3X</td>
<td>P404L</td>
<td>Yes</td>
<td>CFTR</td>
<td>W128X</td>
<td>Yes</td>
</tr>
<tr>
<td>HPRT</td>
<td>G40V</td>
<td>Yes</td>
<td>DM</td>
<td>E1211X</td>
<td>Yes</td>
</tr>
<tr>
<td>HPRT</td>
<td>R48H</td>
<td>Yes</td>
<td>F8</td>
<td>E1987X</td>
<td>Yes</td>
</tr>
<tr>
<td>HPRT</td>
<td>A161E</td>
<td>Yes</td>
<td>FANCC</td>
<td>R2116X</td>
<td>Yes</td>
</tr>
<tr>
<td>HPRT</td>
<td>G180E</td>
<td>Yes</td>
<td>F8</td>
<td>R185X</td>
<td>Yes</td>
</tr>
<tr>
<td>HPRT</td>
<td>G180V</td>
<td>Yes</td>
<td>HMGCL</td>
<td>Y2113X</td>
<td>Yes</td>
</tr>
<tr>
<td>HPRT</td>
<td>G182K</td>
<td>Yes</td>
<td>E37X</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>HPRT</td>
<td>P184L</td>
<td>Yes</td>
<td>E30X</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>HPRT</td>
<td>D194Y</td>
<td>Yes</td>
<td>E47X</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>HPRT</td>
<td>E197K</td>
<td>Yes</td>
<td>K55X</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>HPRT</td>
<td>E197V</td>
<td>Yes</td>
<td>C66X</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>HPRT</td>
<td>D201V</td>
<td>Yes</td>
<td>K103X</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>IVD</td>
<td>R21C</td>
<td>Yes</td>
<td>G119X</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>IVD</td>
<td>R21P</td>
<td>Yes</td>
<td>L125X</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>IVD</td>
<td>D40N</td>
<td>Yes</td>
<td>G180X</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>MLH1</td>
<td>R659P</td>
<td>Yes</td>
<td>E182X</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>MLH1</td>
<td>R659L</td>
<td>Yes</td>
<td>E197X</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>PDHA1</td>
<td>A175T</td>
<td>Yes</td>
<td>HPRRT</td>
<td>Y198X</td>
<td></td>
</tr>
<tr>
<td>PHM2</td>
<td>E139K</td>
<td>Yes</td>
<td>IDUA</td>
<td>Y64X</td>
<td></td>
</tr>
<tr>
<td>RHAG</td>
<td>G380V</td>
<td>Yes</td>
<td>MLHI</td>
<td>R659X</td>
<td></td>
</tr>
<tr>
<td>APC</td>
<td>R623R</td>
<td>Yes</td>
<td>NF1</td>
<td>Y2264X</td>
<td></td>
</tr>
<tr>
<td>AR</td>
<td>S888S</td>
<td>Yes</td>
<td>NF1</td>
<td>Y2264X</td>
<td></td>
</tr>
<tr>
<td>CYP27A1</td>
<td>G112G</td>
<td>Yes</td>
<td>OAT</td>
<td>W178X</td>
<td></td>
</tr>
<tr>
<td>FBN1</td>
<td>I2118I</td>
<td>Yes</td>
<td>OAT</td>
<td>W275X</td>
<td></td>
</tr>
<tr>
<td>HPRT</td>
<td>F199F</td>
<td>Yes</td>
<td>PROSI</td>
<td>S62X</td>
<td></td>
</tr>
<tr>
<td>ITGB3</td>
<td>T420T</td>
<td>Yes</td>
<td>TG</td>
<td>R717X</td>
<td></td>
</tr>
<tr>
<td>NF1</td>
<td>K354K</td>
<td>Yes</td>
<td>WAS</td>
<td>Q99X</td>
<td></td>
</tr>
<tr>
<td>PAH</td>
<td>V399V</td>
<td>Yes</td>
<td>Total</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>PDHA1</td>
<td>G185G</td>
<td>Yes</td>
<td></td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>RET</td>
<td>I647I</td>
<td>Yes</td>
<td></td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>SMN1</td>
<td>F280F</td>
<td>Yes</td>
<td></td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>TNFRFS5</td>
<td>T136T</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

aSeventy-three point mutations that cause exon-skipping and disease were scored with the combined IgM-BRCA1 SF2/ASF score matrix and the score matrices for SRp40, SRp55 and SC35 from ESEfinder.

bScored for the presence of RESCUE-ESE (RESE) hexamers.

cScored for the presence of FAS-hex2 hexamers.

dScored for the presence of PESX octamers.

eTotal number of correct predictions when all four of the motif analysis methodologies are used.
constitutive and alternative exons, respectively. The mean intronic motif densities of 0.0354 ESEs/nt and 0.0407 ESEs/nt for the constitutive and alternative exons, respectively.

degenerate, as was observed for the IgM-derived motifs (22,23,26), and this probably reflects the need for SF2/ASF-specific ESEs to function in multiple exonic contexts, with differential protein-coding specificities. The selected winner pools demonstrated enhanced splicing activity in response to SF2/ASF, compared with the WT BRCA1 ESE control, indicating that functional selection had been successful. In addition, the enhancer activity of individual winner clones was higher than both the WT ESE and random clones from the unselected library pool for both the n7 and n14 sequences. In total 112 clones were tested, with all the winner clones splitting well, with the exception of one clone, later found to be a low-score motif. Our results support the conclusion that the position of functional ESEs within a given exon is constrained to a certain degree. The observation that many of the highest score motifs selected from the n14 library were in the same location, with respect to the 3’ splice-site, as the WT enhancer, implies that this position is optimal for enhancer function in BRCA1 exon 18.

The context for ESE selection provided by the three-exon BRCA1 minigene makes this splicing substrate more typical of the majority of exons, in that the exon under selection is internal, rather than the terminal M2 exon used in the IgM SELEX experiments. There is evidence that the mechanism of exon definition for 3’-terminal exons is different to the definition of internal exons, in that it involves coupling to 3’-end cleavage and polyadenylation of the pre-mRNA (54,55). Hypothetically, the function of ESEs in terminal exons may be subtly different to their function in internal exons, and this in turn may influence the ESEs selected from within a terminal exon. In addition, the local sequence environment may influence ESE selection. Our data support the conclusion that many of the motifs derived from the IgM-based SELEX are able to function in a different exonic context, in that we found significant overlap between the high-score motifs derived from this screen and our experiments in the context of BRCA1. In addition, we recently reported that ESEfinder high-score motifs are present at significantly higher frequencies in internal exons than in their flanking intronic regions (44). The overlap between the IgM- and BRCA1-derived motifs implies that, by analogy, many of the BRCA1-derived motifs will function as ESEs in other exons.

The score matrices in ESEfinder utilize threshold values, above which a given motif is considered significant, which were based on a statistical discrimination of the winner sequences from the unselected pool (26). We derived thresholds based on experimental data and compared them with statistical thresholds derived in an analogous manner to the ESEfinder threshold values. Our data suggest that the absolute threshold value is less important for the ability to predict splicing outcomes, both in vitro and in vivo, than the discriminatory power of the matrix. Our results, with the n7-derived matrix in particular, demonstrated that this score matrix was highly accurate in predicting the in vitro splicing behavior of individual BRCA1 clones, and of an independently tested set of 30 SMN mutants spliced in vivo (40).

A number of winner sequences were demonstrated to enhance exon inclusion in a heterologous exonic context, indicating that the SELEX procedure does not result in motifs that are able to function only in the context of BRCA1. Furthermore, our new matrices were able to accurately predict the splicing outcome of the majority of 30 SMN mutants spliced in vivo (40). Our data give further support to the notion that the splicing defect in SMN2 that contributes to spinal muscular atrophy is the consequence of the disruption of an SF2/ASF-specific ESE (25,40).

The large number of winner clones demonstrated to be functional ESEs validates our ESE screen. However, it remains possible that a number of the winner sequences are able to function only in the context of BRCA1 exon 18. This limitation also applies to the functional IgM SELEX. For example, the IgM-derived matrix is not as accurate at predicting the splicing behavior of the BRCA1 clones tested, although the results were still highly statistically significant ($P < 10^{-5}$, correlation of splicing with motif score). We have addressed this concern by creating a combined IgM-BRCA1 score matrix. A combined matrix improves specificity, an important consideration for ESEfinder users, as the probability of any given motif being a real ESE is increased if the score is derived from information generated from independent experiments. The combined SF2/ASF matrix was derived from the significant number of overlapping high-score motifs from the IgM and BRCA1 SELEX matrices. The combined matrix filters out many ESEs that function only in the context of BRCA1 exon 18 or IgM exon M2, and includes ESEs able to function in both internal and terminal exons. A statistical, rather than the BRCA1-biased experimental threshold value was calculated for the combined matrix, based on scoring of the n14 round-zero clones. A value equal to the 70th percentile of the highest score from each
of the round-zero clones was chosen for increased stringency. A number of different threshold values were calculated and applied to the experimental data. The 70th percentile was selected as it gave the best correlation with observed ESE function and largest statistical discrimination between the putative exonic versus intronic ESE frequencies. Scoring 30 $\text{SMN}$ mutants with the combined SF2/ASF matrix gave significant correlation with the observed in vivo levels of exon inclusion, including accurately predicting the behavior of both $\text{SMN}_1$ and $\text{SMN}_2$.

An important outcome of our experiments was the derivation of a score matrix that could be applied to identify ESEs in other contexts. ESEfinder (26) has been widely utilized by the research community, and a number of groups have identified disease-associated mutations that correlate with reduced ESEfinder motif scores (24,25,27–32,40). In addition, a recent report described a significant correlation between antisense oligonucleotides able to induce specific exon-skipping with high-score ESEfinder motifs in the target exon sequence (56). We analyzed the ability of the combined SF2/ASF plus the ESEfinder score matrices specific for SRp40, SRp55 and SC35 to predict the effects of 73 single-base substitutions in human disease-associated genes that cause exon-skipping in vivo (3,45). The majority of WT sequences containing SF2/ASF motifs had reduced scores in the mutated sequences, and many were reduced below the threshold. The results obtained with the combined matrix were better than using ESEfinder SF2/ASF (data not shown). The effects of point mutations in putative SRp40, SRp55 and SC35 were less conclusive, with a relatively high number of the mutant sequences containing higher score motifs than the WT sequences. One possibility is that a mutated SF2/ASF motif has more pronounced effects upon exon inclusion than mutations in the other SR protein-specific matrices, or that the motifs recognized are partially redundant. Alternatively, further refinement of these other score matrices may be necessary, as the data may be a consequence of the matrices incorrectly predicting some non-existent ESEs.

As a comparison to the predictive power of our new SF2/ASF matrix and ESEfinder, we scored the exon-skipping mutations with the $ab$ $initio$ computational methods of RESCUE-ESE and PESX (33,34). Both correctly predicted the exon-skipping phenotype associated with a number of the mutations. There was some overlap between the motifs recognized by the three methods, and this was more significant for motifs recognized by both RESCUE-ESE and PESE. Our data is in accord with a recent report analyzing mutations in 22 predicted PESEs (57). Both ESEfinder and RESCUE-ESE motifs were affected by 11/18 of the PESE mutations demonstrated to have an effect on exon inclusion, and the majority of these were in different sequences (57). We analyzed the overlap between the motifs recognized by our new SF2/ASF matrix and RESCUE-ESE and PESX, as we have previously calculated for the ESEfinder matrices (44). The motifs recognized as ESEs by the SF2/ASF matrix do not overlap with the RESCUE-ESE motifs. However, we found a significant overlap between our high-score SF2/ASF motifs and PESEs (data not shown), as we previously described for the IgM SELEX-derived SF2/ASF motifs (44). All three methods are useful tools for ESE prediction, and all have been experimentally validated. The constraints, experimental or computational, imposed by the three methods result in significant differences in the motifs predicted to be ESEs. Our data support the notion that all three methods should be used in parallel when searching for putative ESEs.

We recently reported that the frequency of putative SR-protein-ESE motifs is higher in constitutive exons than in their flanking intronic regions (44). We repeated this analysis with our new SF2/ASF matrix on a second set of constitutively spliced exons and a set of alternatively spliced exons, and observed a similar distribution of ESE motifs. Our data reinforce the hypothesis that ESEs are required for splicing of most, if not all, exons, and are not limited to a function in the splicing of alternative exons. The statistical difference in ESE frequency between exonic and intronic regions was greater with our second generation SF2/ASF matrix than the distribution of high-score SF2/ASF motifs generated by scoring the same datasets with the original SF2/ASF matrix. This large-scale analysis provides further evidence that the new matrix is more specific. ESEfinder has now been updated to include the new SF2/ASF matrix, as well as the original four SR protein-specific matrices, increasing the robustness of ESE prediction.

**MATERIALS AND METHODS**

**Library construction and functional SELEX**

Two SELEX libraries were constructed in which the naturally occurring SF2/ASF-specific 7-nt ESE in exon 18 of $\text{BRCA1}$ (24) was replaced with random sequences of either 7 or 14 nt—n7 and n14, respectively—by sequential PCR using high-fidelity Deep Vent polymerase (New England Biolabs). The $\text{BRCA1}$ template is a three-exon minigene with a T7 RNA polymerase promoter, allowing RNA transcripts to be derived. Several oligonucleotide pools containing the randomized regions were obtained in which the manufacturer (Sigma) had optimized the phosphoramidite concentrations to balance the nucleotide composition within the randomized region. Libraries were constructed, random clones sequenced and the n7 and n14 libraries that contained randomized regions closest to 25% representation of each nucleotide were chosen for the SELEX experiments. The location of the $\text{BRCA1}$ exon 18 natural ESE at position +4 made it impossible to regenerate full-length constructs by PCR following the SELEX procedure. Therefore, for cloning purposes, a BaeI restriction site was created by site-directed mutagenesis within exon 17. A second construct ($\text{BRCA1}$ C), in which the reciprocal BaeI site was engineered in $\text{BRCA1}$ exon 18, was made solely for cloning purposes. The SELEX procedure was carried out as previously described, with some modifications (22,23). An overview of the procedure is illustrated in Figure 1. The $\text{BRCA1}$ SELEX libraries were in vitro transcribed with T7 polymerase (Promega), generating radiolabeled pre-mRNA substrate pools. About 10 fmol of the pre-mRNA pools was incubated under in vitro splicing conditions in S100 extract plus 10 pmol of recombinant SF2/ASF in 25 µl reaction mixtures. The RNAs were separated by denaturing polyacrylamide gel electrophoresis, and
spliced mRNAs excised and eluted from the gel in 0.5 M sodium acetate, 1 mM ethylenediaminetetraacetic acid and 0.2% sodium dodecyl sulfate. RNAs were recovered by ethanol precipitation and reverse-transcribed using Superscript II, as described by the manufacturer (Invitrogen), using an exon-specific primer. Full-length cDNAs were amplified by PCR using Deep Vent polymerase, and then digested with BaeI (New England Biolabs) followed by agarose-gel purification (Qiagen gel-extraction kit). Exon 17 plus intron 17, produced from BaeI digestion of the cloning construct (C) was then ligated to the BaeI-digested SELEX products using T4 DNA ligase (Gibco-BRL). Ligation products were purified by agarose-gel electrophoresis, and full-length splicing constructs rebuilt by overlap-extension PCR. Oligonucleotide sequences are available upon request. Following each round of selection, the winner pools were sub-cloned into the vector pCR-Blunt (Invitrogen) and sequenced by use of a Dye Terminator Cycle Sequencing kit (Perkin-Elmer) and an automated ABI 377 sequencer.

Preparation of HeLa cell extracts and recombinant SF2/ASF

HeLa nuclear and S100 extracts, and recombinant SF2/ASF were prepared as described (58,59). The integrity and purity of SF2/ASF was checked by SDS–PAGE and the specific activity determined by in vitro splicing of β-globin pre-mRNA in S100 extract. SF2/ASF and extract concentrations for use in the SELEX experiments were optimized using the splicing minigenes BRCA1-WT and nonsense mutant BRCA1 E1694X (BRCA1-MT), the latter of which is known to cause skipping of exon 18 (24). The same S100 and SF2/ASF preparations were used for all of the in vitro experiments.

In vitro splicing

Uniformly [α-32P] UTP-labeled, 5′-capped, T7 runoff transcripts were produced from the SELEX libraries, individual SELEX clones, BRCA1 control substrates and SMN constructs. Transcripts were purified by denaturing PAGE and spliced in HeLa cell nuclear extract or S100 post-nuclear extract under standard conditions with 1.6 mM MgCl2 (60). SELEX experiments were optimized using the splicing minigenes BRCA1-WT and nonsense mutant BRCA1 E1694X (BRCA1-MT), the latter of which is known to cause skipping of exon 18 (24). The same S100 and SF2/ASF preparations were used for all of the in vitro experiments.

Sequence analysis and construction of score matrices

Consensus motifs were identified from the n14 winner sequences by alignment of the sequences using three motif-finding algorithms: (1) Gibbs sampler (37); (2) MEME (38); (3) DME (39). An additional alignment of the n14 winners was performed by scoring the sequences with the n7-derived matrix. Alignment of the n7 winners was not required because of the ESE position being fixed. The aligned sequences were used to derive consensus motifs from which nucleotide-frequency score matrices were constructed using established methods (22,23). The compositional bias of the initial RNA pools was taken into account. A detailed description of score-matrix construction is given in Liu et al. (22).

Analysis of genomic ESE motif frequencies

Datasets of constitutively spliced and alternatively spliced human internal protein-coding exons were retrieved from the Alternative Splicing Database (61). Genomic sequences and transcript annotations were downloaded from http://www.ebi.ac.uk.asd/. Constitutively spliced exons and their flanking intronic sequences were extracted by a Perl script from 6291 intron-containing genes without any EST or cDNA evidence of alternative splicing. The dataset was filtered to remove exons flanked by short introns (<250 nt), and short exons (<106 nt). Composite exons were created from this set of 16 635 exons, comprising 25 nt from each end and 50 nt from the center. The composite exons, and 100 nt of upstream and downstream flanking intronic sequence, were scored for the presence of high-score SF2/ASF motifs. A total of 5041 alternatively spliced cassette exons and their flanking intronic regions were extracted from 3644 genes, filtered and composite exons created as earlier.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG Online.

ACKNOWLEDGEMENTS

This work was supported by NIH grants GM42699 to A.R.K. and HG01696/CA88351 to M.Q.Z. and by a postdoctoral fellowship from the U.S. Army Medical Research and Matériel Command to P.J.S.

Conflict of Interest statement. None declared.

REFERENCES

27. Colapietro, P., Gervasio, C., Natafeci, F., Rossi, L., Riva, P. and
28. Ferrari, S., Giliani, S., Insalaco, A., Al-Ghoniaim, A., Soresina, A.R.,
30. Mas, C., Taske, N., Deutsch, S., Guipponi, M., Thomas, P., Covasin, A.,
31. Aretz, S., Uhlihaas, S., Sun, Y., Pagenstecher, C., Mangold, E., Caspari, R.,
Science, 297, 1007–1013.
37. Lawrence, C.E., Altschul, S.F., Boguski, M.S., Liu, J.S., Neuwald, A.F. and
tissue-selective transcription factor binding sites in vertebrate promoters.
40. Cartegni, L., Hastings, M.L., Calarco, J.A., de Stanchina, E. and
nucleotide in the SMN gene regulates splicing and is responsible for
42. Lefebvre, S., Burglen, L., Reboullet, S., Clermont, O., Burlet, P.,
Identification and characterization of a spinal muscular atrophy-determining gene. Cell, 80, 155–165.
43. Lorson, C.L. and Androphy, E.J. (2000) An exonic enhancer is required
for inclusion of an essential exon in the SMA-determining gene SMN.
protein exonic splicing enhancer motifs in human protein-coding genes.
46. Wang, Z., Rolish, M.E., Yeo, G., Tung, V., Mawson, M. and Burge, B.
Cell, 119, 831–845.
characterization of co-factor proteins that enhance pre-mRNA splicing by
Genes Dev., 17, 705–733.
Cell, 114, 973–982.
4, 571–582.
U2AF65 is required for splicing. Genes Dev., 14, 403–413.
56. Aartsma-Rus, A., De Winter, C.L., Janson, A.A., Kaman, W.E.,
57. Aartsma-Rus, A., De Winter, C.L., Janson, A.A., Kaman, W.E.,
58. Aartsma-Rus, A., De Winter, C.L., Janson, A.A., Kaman, W.E.,
59. Aartsma-Rus, A., De Winter, C.L., Janson, A.A., Kaman, W.E.,


