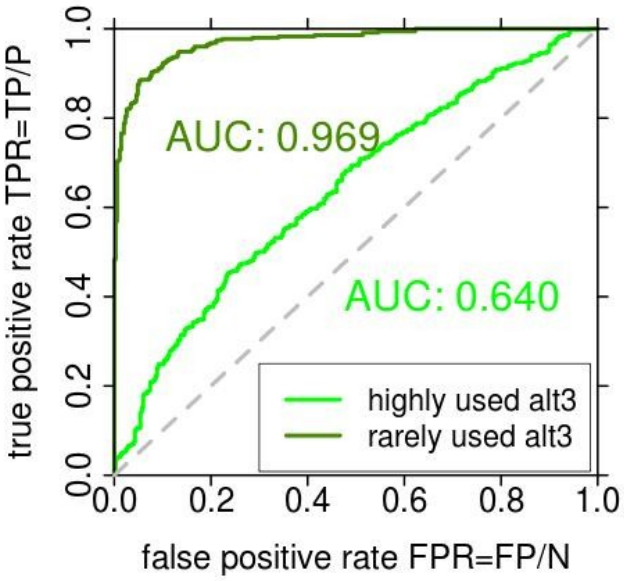
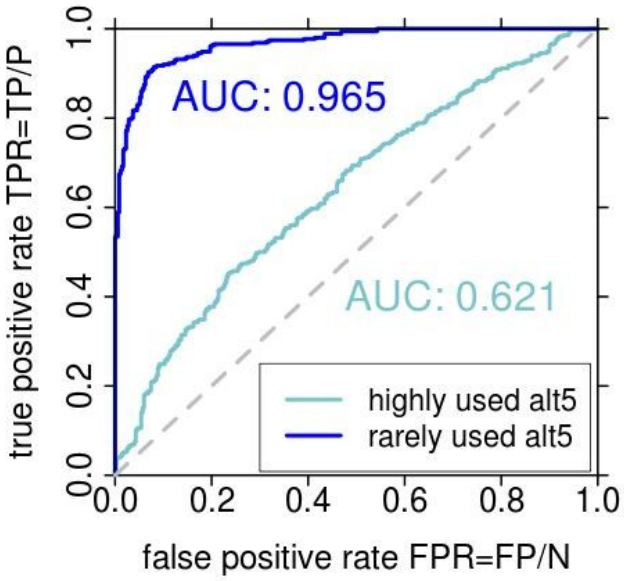
Supplemental Material

Supplemental Figure 1. Performance of SVMs distinguishing between cassette exons and exons with an alternative splice site. Highly included cassette exons were compared to exons with a highly used alternative splice site, while rarely included cassette exons were compared to exons with a rarely used alternative splice site.

(a) cassette exons vs. exons with an alternative 3' splice site

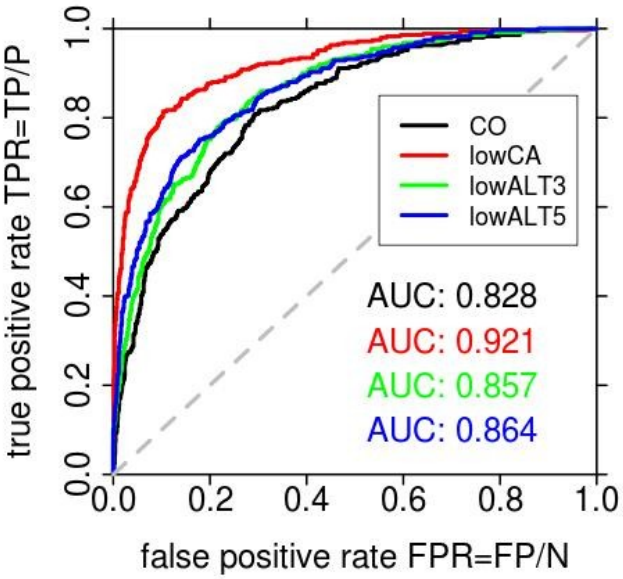


(b) cassette exons vs. exons with an alternative 5' splice site

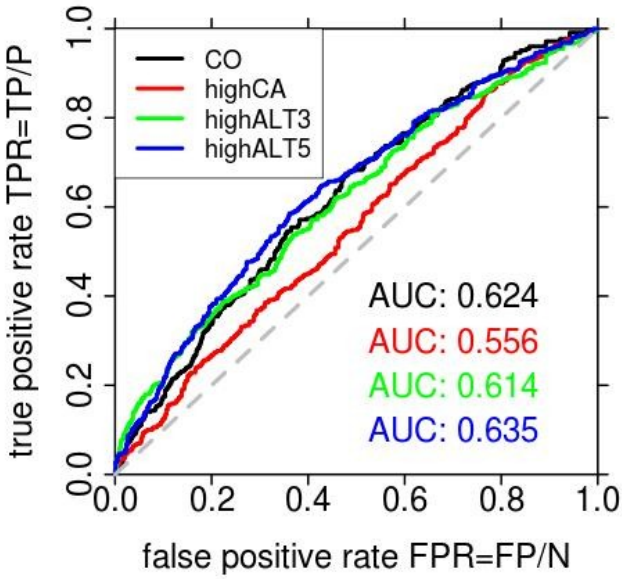


Supplemental Figure 2. Performance of the splicing code when increasing the compendium of RNA features to 826, including binding sites for splicing regulators as well as codon frequencies.

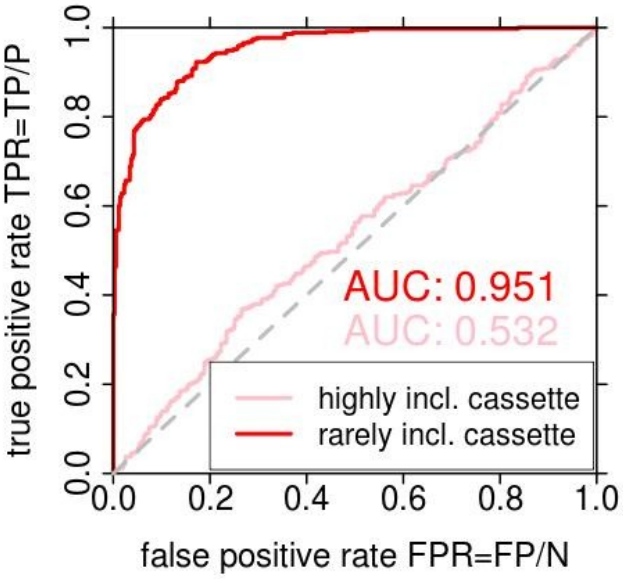
(a) all (LOW)



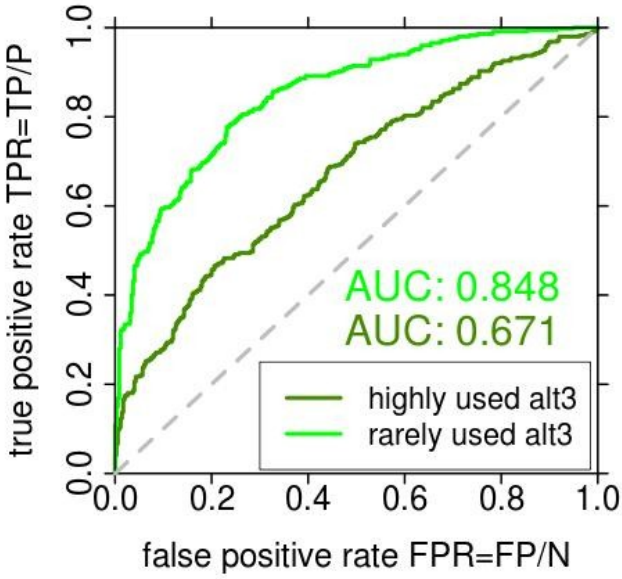
(b) all (HIGH)



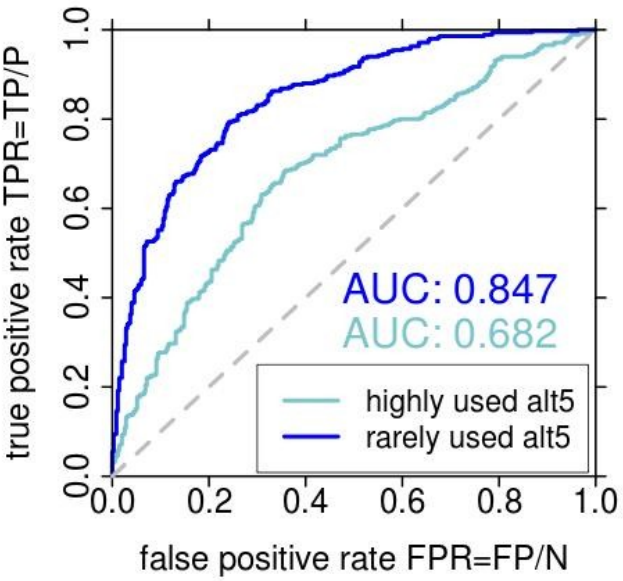
(c) constitutive vs. cassette exons



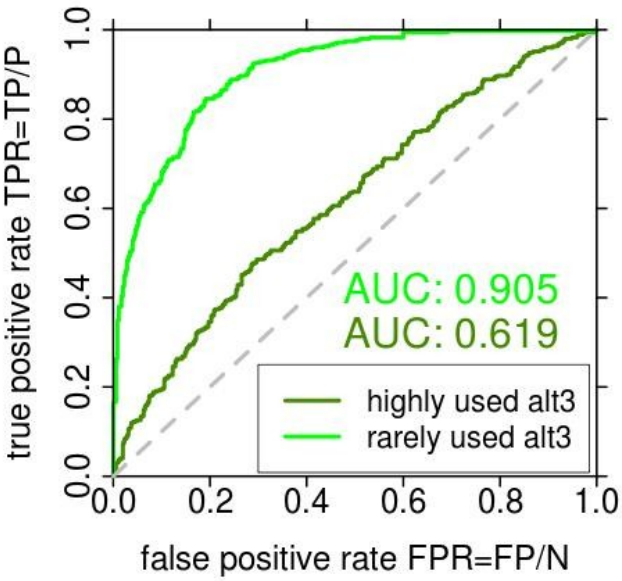
(d) constitutive exons vs. exons with an alternative 3'ss



(e) constitutive exons vs. exons with an alternative 5'ss

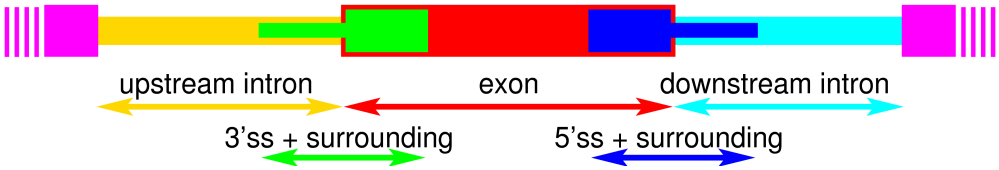


(f) exons with an alternative 3'ss vs. exons with an alternative 5'ss

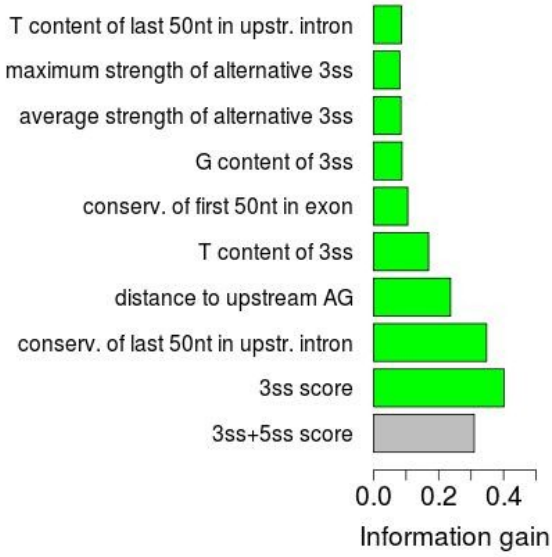


Supplemental Figure 3. Most influencing features when increasing the compendium of RNA features to 826, including binding sites for splicing regulators as well as codon frequencies.

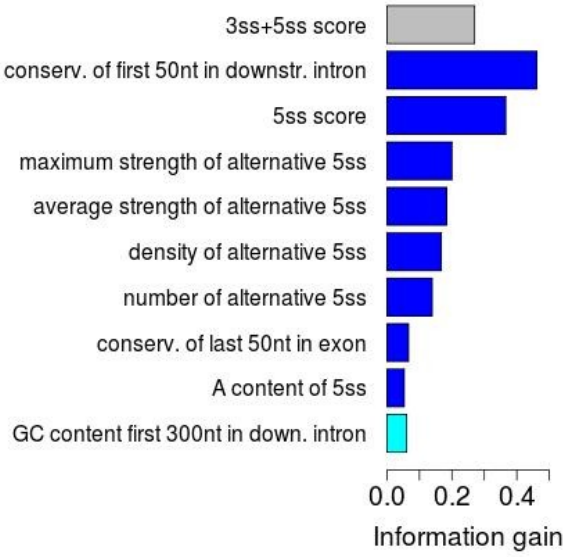
(a) color scheme



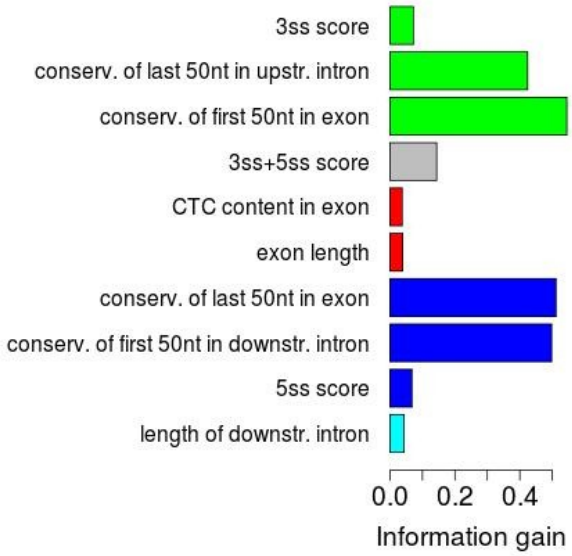
(b) constitutive vs. exons with a rarely used alt. 3' ss



(c) constitutive vs. exons with a rarely used alt. 5' ss

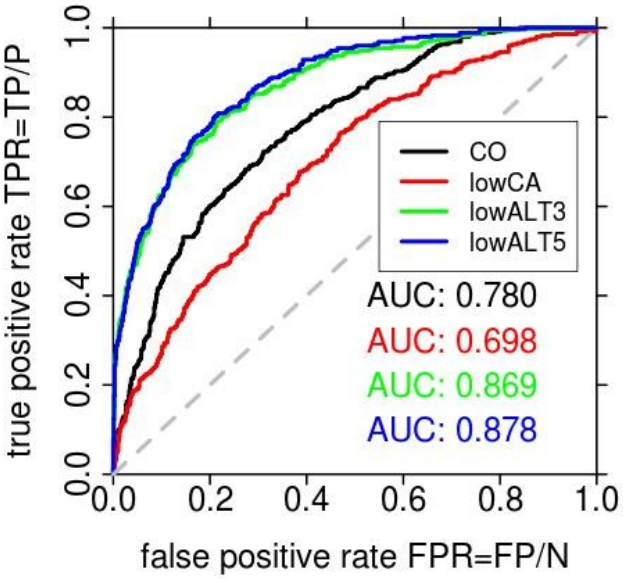


(d) constitutive vs. rarely included cassette exons

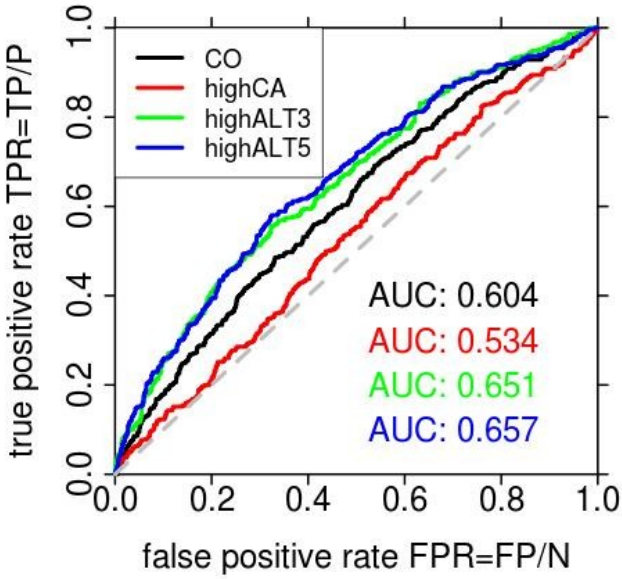


Supplemental Figure 4. Performance of the splicing code without using conservation features.

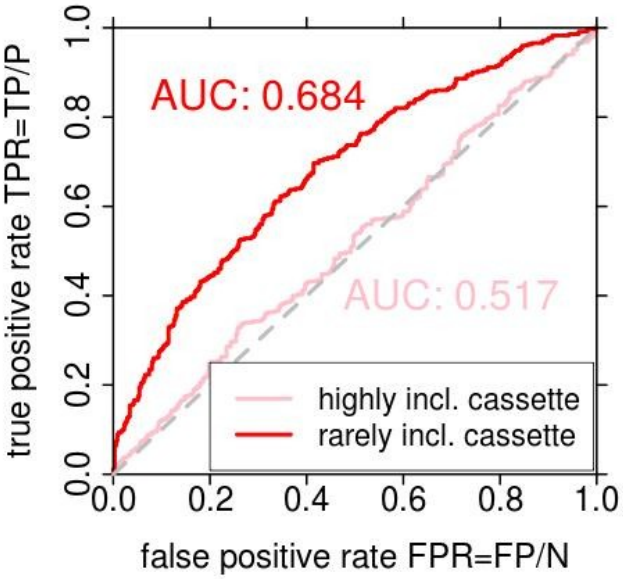
(a) all (LOW)



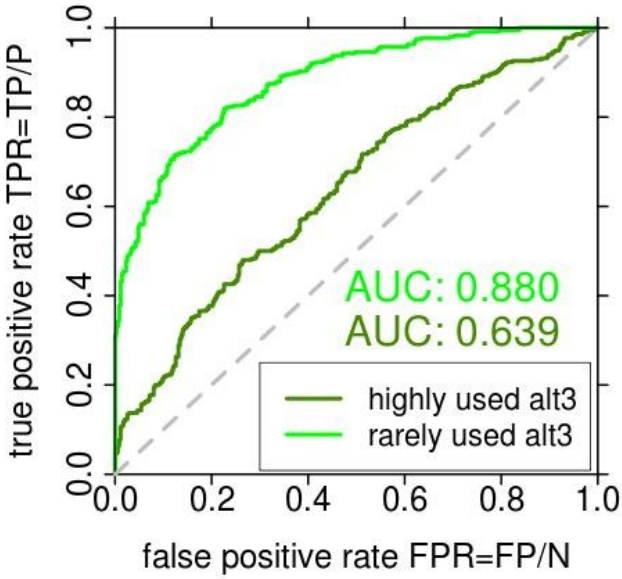
(b) all (HIGH)



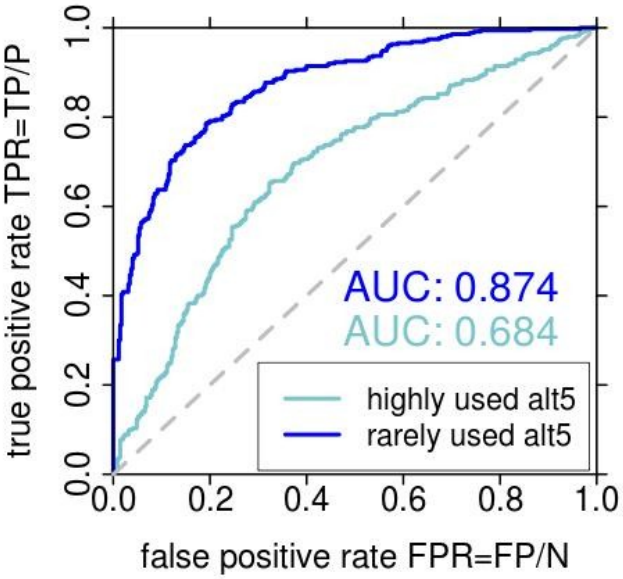
(c) constitutive vs. cassette exons



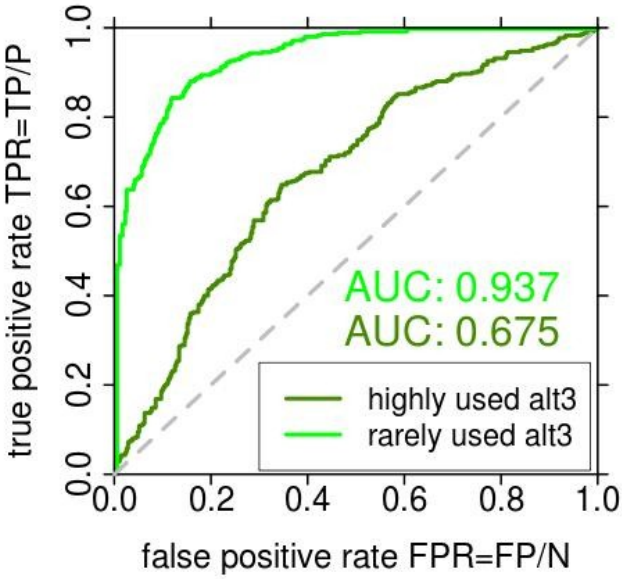
(d) constitutive exons vs. exons with an alternative 3'ss



(e) constitutive exons vs. exons with an alternative 5'ss

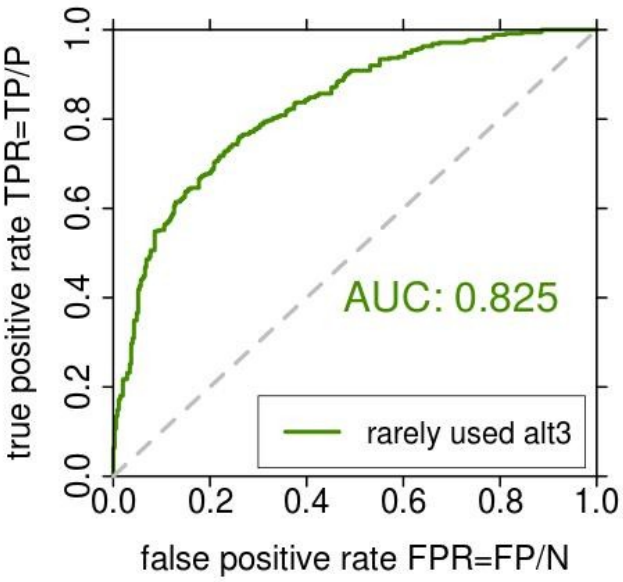


(f) exons with an alternative 3'ss vs. exons with an alternative 5'ss

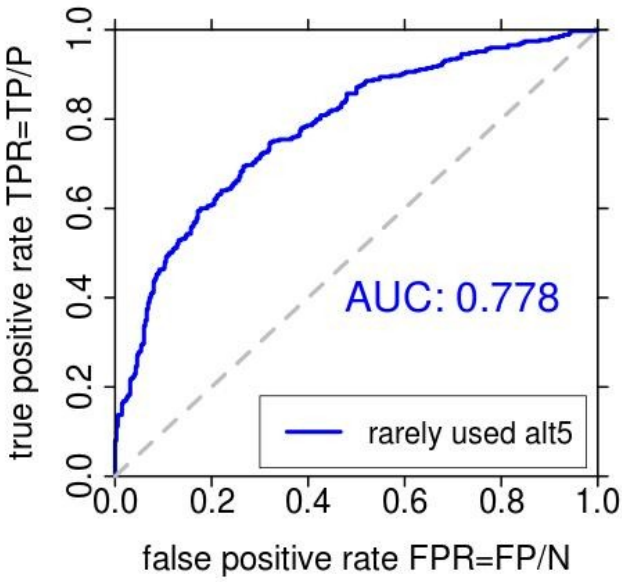


Supplemental Figure 5. Performance of SVMs distinguishing between highly included/used exons/splice sites and rarely included/used exons/splice sites.

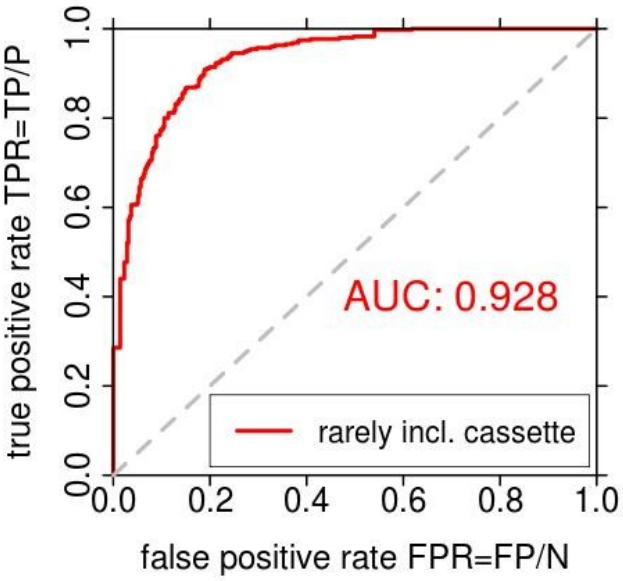
(a) exons with an alternative 3' splice site



(b) exons with an alternative 5' splice site

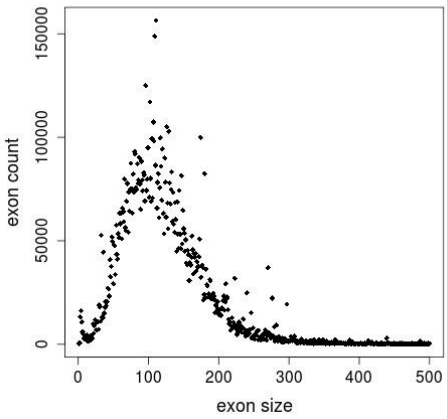


(c) cassette exons

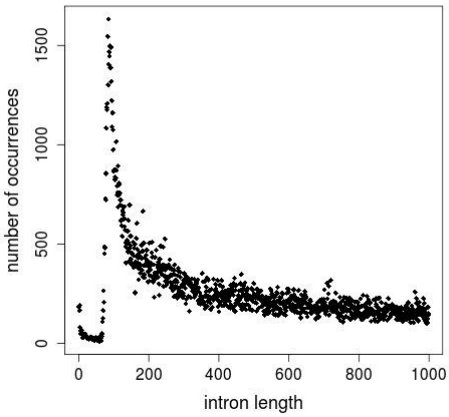


Supplemental Figure 6. The distribution of (a) all internal human exons (up to a size of 500nt) and (b) all human introns (up to a size of 1000nt).

(a) frequencies of internal human exons depending on their size

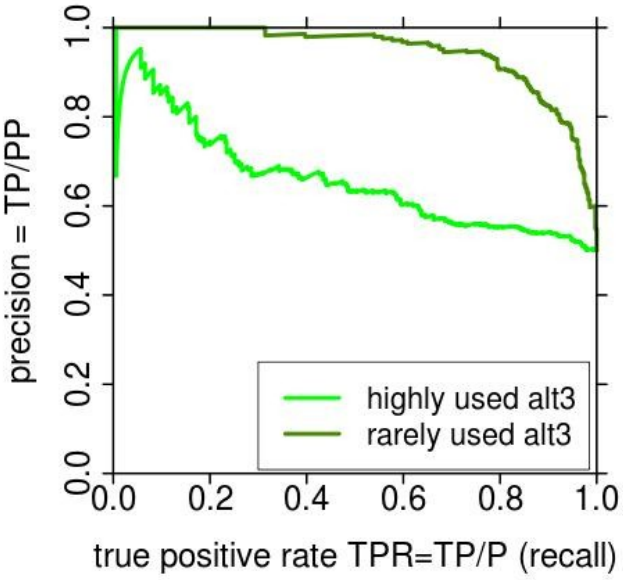


(b) frequencies of human introns depending on their size

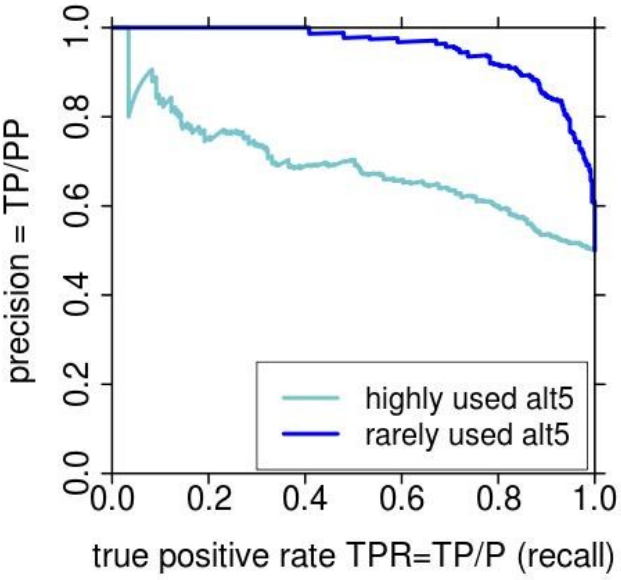


Supplemental Figure 7. Precision-recall plots of the splicing code as described in the main manuscript. The recall is defined as the true positive rate TPR = TP/P (the number of true positives TP divided by the number of all real positives P=TP+FN). The precision is defined as the positive predictive value PPV = TP/PP (the number of true positives TP divided by the number of all that are predicted to be positive PP=TP+FP).

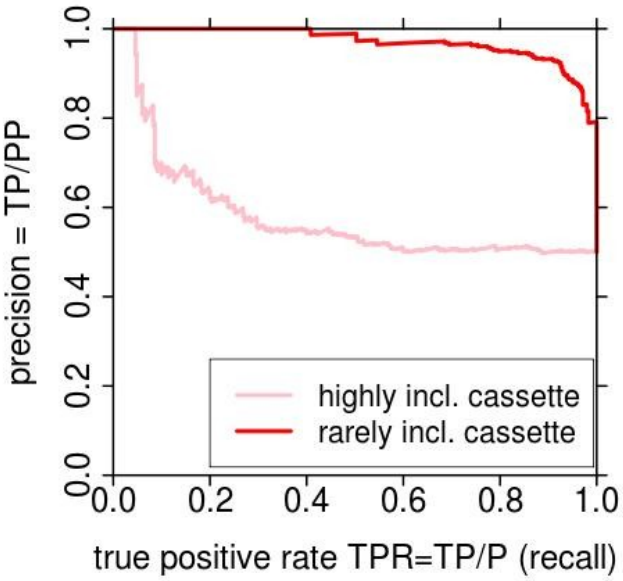
(a) constitutive vs. exons with an alternative 3' splice site



(b) constitutive vs. exons with an alternative 5' splice site



(c) constitutive vs. cassette exons



Supplemental Table 1. A complete list of RNA features.

feature name	data type	region
length	integer	E
length-upstream-intron	integer	I1
length-downstream-intron	integer	I2
length-upstream-exon	integer	E1
length-downstream-exon	integer	E2
3ss-score	real	3ss
5ss-score	real	5ss
3ss5ss-score	real	3ss+5ss
averagePU-4mers-3ss	real	3ss
averagePU-4mers-5ss	real	5ss
ESEdensity-all	real	E
ESEaver-rel-strength-all	real	E
ESEmax-rel-score-all	real	E
dist-start-first-ESE-all	integer	E
dist-end-last-ESE-all	integer	E
ESEdensity-all-PU	real	E
ESEaver-rel-strength-all-PU	real	E
ESEmax-rel-score-all-PU	real	E
dist-start-first-ESE-all-PU	integer	E
dist-end-last-ESE-all-PU	integer	E
ESEdensity-ASF	real	E
ESEaver-strength-ASF	real	E
ESEaver-rel-strength-ASF	real	E
ESEmax-score-ASF	real	E
dist-start-first-ESE-ASF	integer	E
dist-end-last-ESE-ASF	integer	E
ESEdensity-ASF-PU	real	E
ESEaver-strength-ASF-PU	real	E
ESEaver-rel-strength-ASF-PU	real	E
ESEmax-score-ASF-PU	real	E
dist-start-first-ESE-ASF-PU	integer	E
dist-end-last-ESE-ASF-PU	integer	E
ESEdensity-IgM-BRCA1	real	E
ESEaver-strength-IgM-BRCA1	real	E
ESEaver-rel-strength-IgM-BRCA1	real	E
ESEmax-score-IgM-BRCA1	real	E
dist-start-first-ESE-IgM-BRCA1	integer	E
dist-end-last-ESE-IgM-BRCA1	integer	E
ESEdensity-IgM-BRCA1-PU	real	E
ESEaver-strength-IgM-BRCA1-PU	real	E
ESEaver-rel-strength-IgM-BRCA1-PU	real	E
ESEmax-score-IgM-BRCA1-PU	real	E
dist-start-first-ESE-IgM-BRCA1-PU	integer	E
dist-end-last-ESE-IgM-BRCA1-PU	integer	E
ESEdensity-SC35	real	E
ESEaver-strength-SC35	real	E
ESEaver-rel-strength-SC35	real	E
ESEmax-score-SC35	real	E
dist-start-first-ESE-SC35	integer	E
dist-end-last-ESE-SC35	integer	E

ESEdensity-SC35-PU	real	E
ESEaver-strength-SC35-PU	real	E
ESEaver-rel-strength-SC35-PU	real	E
ESEmax-score-SC35-PU	real	E
dist-start-first-ESE-SC35-PU	integer	E
dist-end-last-ESE-SC35-PU	integer	E
ESEdensity-SRp40	real	E
ESEaver-strength-SRp40	real	E
ESEaver-rel-strength-SRp40	real	E
ESEmax-score-SRp40	real	E
dist-start-first-ESE-SRp40	integer	E
dist-end-last-ESE-SRp40	integer	E
ESEdensity-SRp40-PU	real	E
ESEaver-strength-SRp40-PU	real	E
ESEaver-rel-strength-SRp40-PU	real	E
ESEmax-score-SRp40-PU	real	E
dist-start-first-ESE-SRp40-PU	integer	E
dist-end-last-ESE-SRp40-PU	integer	E
ESEdensity-SRp55	real	E
ESEaver-strength-SRp55	real	E
ESEaver-rel-strength-SRp55	real	E
ESEmax-score-SRp55	real	E
dist-start-first-ESE-SRp55	integer	E
dist-end-last-ESE-SRp55	integer	E
ESEdensity-SRp55-PU	real	E
ESEaver-strength-SRp55-PU	real	E
ESEaver-rel-strength-SRp55-PU	real	E
ESEmax-score-SRp55-PU	real	E
dist-start-first-ESE-SRp55-PU	integer	E
dist-end-last-ESE-SRp55-PU	integer	E
fraction-A-3ss	real	3ss
fraction-C-3ss	real	3ss
fraction-G-3ss	real	3ss
fraction-T-3ss	real	3ss
fraction-A-5ss	real	5ss
fraction-C-5ss	real	5ss
fraction-G-5ss	real	5ss
fraction-T-5ss	real	5ss
fraction-A-exon	real	E
fraction-C-exon	real	E
fraction-G-exon	real	E
fraction-T-exon	real	E
fraction-CG-exon	real	E
fraction-A-50nt-upstream-exon	real	E1
fraction-C-50nt-upstream-exon	real	E1
fraction-G-50nt-upstream-exon	real	E1
fraction-T-50nt-upstream-exon	real	E1
fraction-CG-50nt-upstream-exon	real	E1
fraction-A-50nt-downstream-exon	real	E2
fraction-C-50nt-downstream-exon	real	E2
fraction-G-50nt-downstream-exon	real	E2
fraction-T-50nt-downstream-exon	real	E2
fraction-CG-50nt-downstream-exon	real	E2

ISSdensity-up50nt-all	real	I1
ISSaver-rel-strength-up50nt-all	real	I1
ISSmax-rel-score-up50nt-all	real	I1
dist-start-firstISS-up50nt-all	integer	I1
dist-end-lastISS-up50nt-all	integer	I1
ISSdensity-down50nt-all	real	I2
ISSaver-rel-strength-down50nt-all	real	I2
ISSmax-rel-score-down50nt-all	real	I2
dist-start-firstISS-down50nt-all	integer	I2
dist-end-lastISS-down50nt-all	integer	I2
number-altSS-around-3ss-score>=0	integer	3ss+/-200nt
number-altSS-around-3ss-score>=1	integer	3ss+/-200nt
number-altSS-around-3ss-score>=2	integer	3ss+/-200nt
number-altSS-around-3ss-score>=3	integer	3ss+/-200nt
number-altSS-around-3ss-score>=4	integer	3ss+/-200nt
number-altSS-around-3ss-score>=5	integer	3ss+/-200nt
number-altSS-around-3ss-score>=6	integer	3ss+/-200nt
number-altSS-around-3ss-score>=7	integer	3ss+/-200nt
number-altSS-around-3ss-score>=7.5	integer	3ss+/-200nt
number-altSS-around-3ss-score>=8	integer	3ss+/-200nt
number-altSS-around-3ss-score>=8.5	integer	3ss+/-200nt
number-altSS-around-3ss-score>=9	integer	3ss+/-200nt
number-altSS-around-3ss-score>=9.5	integer	3ss+/-200nt
number-altSS-around-3ss-score>=10	integer	3ss+/-200nt
number-altSS-around-3ss-score>=10.5	integer	3ss+/-200nt
number-altSS-around-3ss-score>=11	integer	3ss+/-200nt
density-altSS-around-3ss-score>=0-norm-by-length	real	3ss+/-200nt
density-altSS-around-3ss-score>=1-norm-by-length	real	3ss+/-200nt
density-altSS-around-3ss-score>=2-norm-by-length	real	3ss+/-200nt
density-altSS-around-3ss-score>=3-norm-by-length	real	3ss+/-200nt
density-altSS-around-3ss-score>=4-norm-by-length	real	3ss+/-200nt
density-altSS-around-3ss-score>=5-norm-by-length	real	3ss+/-200nt
density-altSS-around-3ss-score>=6-norm-by-length	real	3ss+/-200nt
density-altSS-around-3ss-score>=7-norm-by-length	real	3ss+/-200nt
density-altSS-around-3ss-score>=7.5-norm-by-length	real	3ss+/-200nt
density-altSS-around-3ss-score>=8-norm-by-length	real	3ss+/-200nt
density-altSS-around-3ss-score>=8.5-norm-by-length	real	3ss+/-200nt
density-altSS-around-3ss-score>=9-norm-by-length	real	3ss+/-200nt
density-altSS-around-3ss-score>=9.5-norm-by-length	real	3ss+/-200nt
density-altSS-around-3ss-score>=10-norm-by-length	real	3ss+/-200nt
density-altSS-around-3ss-score>=10.5-norm-by-length	real	3ss+/-200nt
density-altSS-around-3ss-score>=11-norm-by-length	real	3ss+/-200nt
averStrength-altSS-around-3ss-score>=0	real	3ss+/-200nt
averStrength-altSS-around-3ss-score>=1	real	3ss+/-200nt
averStrength-altSS-around-3ss-score>=2	real	3ss+/-200nt
averStrength-altSS-around-3ss-score>=3	real	3ss+/-200nt
averStrength-altSS-around-3ss-score>=4	real	3ss+/-200nt
averStrength-altSS-around-3ss-score>=5	real	3ss+/-200nt
averStrength-altSS-around-3ss-score>=6	real	3ss+/-200nt
averStrength-altSS-around-3ss-score>=7	real	3ss+/-200nt
averStrength-altSS-around-3ss-score>=7.5	real	3ss+/-200nt
averStrength-altSS-around-3ss-score>=8	real	3ss+/-200nt
averStrength-altSS-around-3ss-score>=8.5	real	3ss+/-200nt

averStrength-altSS-around-3ss-score>=9	real	3ss+/-200nt
averStrength-altSS-around-3ss-score>=9.5	real	3ss+/-200nt
averStrength-altSS-around-3ss-score>=10	real	3ss+/-200nt
averStrength-altSS-around-3ss-score>=10.5	real	3ss+/-200nt
averStrength-altSS-around-3ss-score>=11	real	3ss+/-200nt
maxStrength-altSS-around-3ss	real	3ss+/-200nt
number-altSS-around-5ss-score>=0	integer	5ss+/-200nt
number-altSS-around-5ss-score>=1	integer	5ss+/-200nt
number-altSS-around-5ss-score>=2	integer	5ss+/-200nt
number-altSS-around-5ss-score>=3	integer	5ss+/-200nt
number-altSS-around-5ss-score>=4	integer	5ss+/-200nt
number-altSS-around-5ss-score>=5	integer	5ss+/-200nt
number-altSS-around-5ss-score>=6	integer	5ss+/-200nt
number-altSS-around-5ss-score>=7	integer	5ss+/-200nt
number-altSS-around-5ss-score>=7.5	integer	5ss+/-200nt
number-altSS-around-5ss-score>=8	integer	5ss+/-200nt
number-altSS-around-5ss-score>=8.5	integer	5ss+/-200nt
number-altSS-around-5ss-score>=9	integer	5ss+/-200nt
number-altSS-around-5ss-score>=9.5	integer	5ss+/-200nt
number-altSS-around-5ss-score>=10	integer	5ss+/-200nt
number-altSS-around-5ss-score>=10.5	integer	5ss+/-200nt
number-altSS-around-5ss-score>=11	integer	5ss+/-200nt
density-altSS-around-5ss-score>=0-norm-by-length	real	5ss+/-200nt
density-altSS-around-5ss-score>=1-norm-by-length	real	5ss+/-200nt
density-altSS-around-5ss-score>=2-norm-by-length	real	5ss+/-200nt
density-altSS-around-5ss-score>=3-norm-by-length	real	5ss+/-200nt
density-altSS-around-5ss-score>=4-norm-by-length	real	5ss+/-200nt
density-altSS-around-5ss-score>=5-norm-by-length	real	5ss+/-200nt
density-altSS-around-5ss-score>=6-norm-by-length	real	5ss+/-200nt
density-altSS-around-5ss-score>=7-norm-by-length	real	5ss+/-200nt
density-altSS-around-5ss-score>=7.5-norm-by-length	real	5ss+/-200nt
density-altSS-around-5ss-score>=8-norm-by-length	real	5ss+/-200nt
density-altSS-around-5ss-score>=8.5-norm-by-length	real	5ss+/-200nt
density-altSS-around-5ss-score>=9-norm-by-length	real	5ss+/-200nt
density-altSS-around-5ss-score>=9.5-norm-by-length	real	5ss+/-200nt
density-altSS-around-5ss-score>=10-norm-by-length	real	5ss+/-200nt
density-altSS-around-5ss-score>=10.5-norm-by-length	real	5ss+/-200nt
density-altSS-around-5ss-score>=11-norm-by-length	real	5ss+/-200nt
averStrength-altSS-around-5ss-score>=0	real	5ss+/-200nt
averStrength-altSS-around-5ss-score>=1	real	5ss+/-200nt
averStrength-altSS-around-5ss-score>=2	real	5ss+/-200nt
averStrength-altSS-around-5ss-score>=3	real	5ss+/-200nt
averStrength-altSS-around-5ss-score>=4	real	5ss+/-200nt
averStrength-altSS-around-5ss-score>=5	real	5ss+/-200nt
averStrength-altSS-around-5ss-score>=6	real	5ss+/-200nt
averStrength-altSS-around-5ss-score>=7	real	5ss+/-200nt
averStrength-altSS-around-5ss-score>=7.5	real	5ss+/-200nt
averStrength-altSS-around-5ss-score>=8	real	5ss+/-200nt
averStrength-altSS-around-5ss-score>=8.5	real	5ss+/-200nt
averStrength-altSS-around-5ss-score>=9	real	5ss+/-200nt
averStrength-altSS-around-5ss-score>=9.5	real	5ss+/-200nt
averStrength-altSS-around-5ss-score>=10	real	5ss+/-200nt
averStrength-altSS-around-5ss-score>=10.5	real	5ss+/-200nt

averStrength-altSS-around-5ss-score>=11	real	5ss+/-200nt
maxStrength-altSS-around-5ss	real	5ss+/-200nt
averConservation-up50nt-3ss	real	I1
averConservation-down50nt-3ss	real	E
averConservation-up50nt-5ss	real	E
averConservation-down50nt-5ss	real	I2
distance-upstreamAG	integer	I1
maxPU-3ss	real	3ss
minPU-3ss	real	3ss
maxPU-5ss	real	5ss
minPU-5ss	real	5ss
averPU-70nt-upstream-exon	real	I1
maxPU-70nt-upstream-exon	real	I1
minPU-70nt-upstream-exon	real	I1
averPU-first70nt-exon	real	E
maxPU-first70nt-exon	real	E
minPU-first70nt-exon	real	E
averPU-last70nt-exon	real	E
maxPU-last70nt-exon	real	E
minPU-last70nt-exon	real	E
averPU-70nt-downstream-exon	real	I2
maxPU-70nt-downstream-exon	real	I2
minPU-70nt-downstream-exon	real	I2
upstream-intron-Urich-count	integer	I1
upstream-intron-Urich-density	real	I1
downstream-intron-Urich-count	integer	I2
downstream-intron-Urich-density	real	I2
upstream-intron-MBNL1-count	integer	I1
upstream-intron-MBNL1-density	real	I1
downstream-intron-MBNL1-count	integer	I2
downstream-intron-MBNL1-density	real	I2
upstream-intron-CUrich-count	integer	I1
upstream-intron-CUrich-density	real	I1
downstream-intron-CUrich-count	integer	I2
downstream-intron-CUrich-density	real	I2
upstream-intron-Fox-count	integer	I1
upstream-intron-Fox-density	real	I1
downstream-intron-Fox-count	integer	I2
downstream-intron-Fox-density	real	I2
upstream-intron-CUGrich-count	integer	I1
upstream-intron-CUGrich-density	real	I1
downstream-intron-CUGrich-count	integer	I2
downstream-intron-CUGrich-density	real	I2
upstream-intron-Qkl-count	integer	I1
upstream-intron-Qkl-density	real	I1
downstream-intron-Qkl-count	integer	I2
downstream-intron-Qkl-density	real	I2
upstream-intron-Nova-count	integer	I1
upstream-intron-Nova-density	real	I1
downstream-intron-Nova-count	integer	I2
downstream-intron-Nova-density	real	I2
exon-Nova-count	integer	E
exon-Nova-density	real	E