

Additional file 2

Supplementary Figures

Supplementary Figure 1. Distribution plot of all alternative splicing (FIRMA) scores in the test series of colorectal cancers. Numbers of log-2 transformed FIRMA scores for all 284,258 exons in 83 CRC samples in the test series are included. The FIRMA scores ranged from -24.3 to 15.0 with median value 0.0 and variance 0.6. The lower and upper 1st percentiles were used as threshold values to score exons with deviating skipping or inclusion (-2.2 and 1.9 respectively, marked in red). To investigate whether different thresholds for scoring exons with deviating skipping and inclusion would alter the distribution of sample-wise amounts of deviating exon usage, we tested the upper and lower 10th and 0.1st percentiles as well. The results from the 1st percentile thresholds had Pearson correlations ≥ 0.96 with the results from both alternative thresholds, for both deviating exon skipping and inclusion events ($P < 0.001$). Hence, the sample-wise distribution of deviating exon usage amounts remained largely unaltered.

Supplementary Figure 2. Comparison of deviating exon usage in paired colorectal cancer and normal colonic mucosa samples. a) There was significantly more deviating exon usage in the cancer samples compared to their normal mucosa counterparts. Eleven out of thirteen sample pairs had higher amounts in the cancer. The mean difference in total relative amounts of deviating exon usage was 1.2 (log2-scale; $P = 0.003$, by paired samples t-test for equality of means). This was true also for the individual events deviating exon skipping and inclusion (mean difference 1.2 with $P = 0.001$, and mean difference 1.2 with $P = 0.006$, respectively). b) The mean paired differences in alternative splicing (FIRMA) scores between thirteen paired cancer

and normal samples for thirteen exons known to be alternatively spliced between CRC and normal colonic mucosa (log2-scale). All six exons known to be primarily expressed in normal mucosa (blue bars), had higher mean paired alternative splicing scores in the normal samples, Of the seven exons known to be primarily expressed in CRC (red bars), five had higher mean alternative splicing scores in the CRC samples.

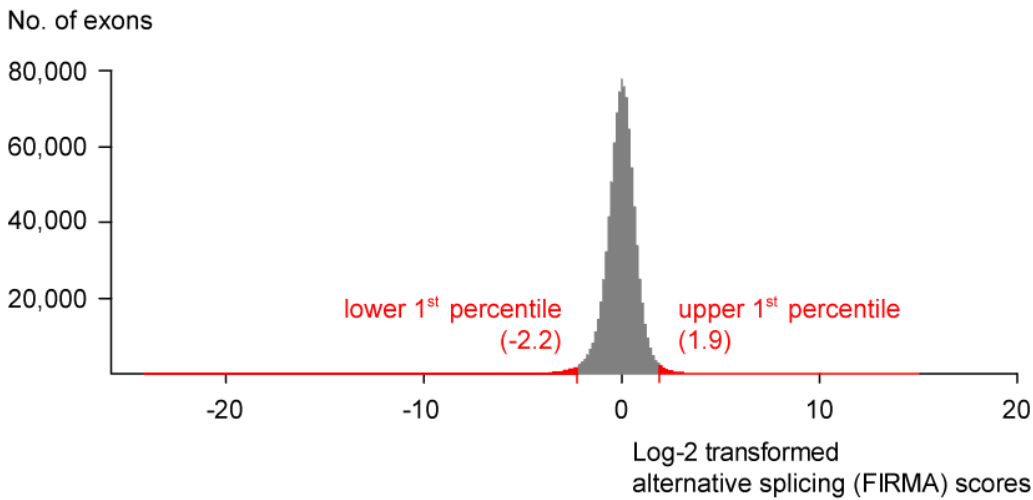
Supplementary Figure 3. Correlation to splicing factor expression levels for permutations of the total relative amounts of deviating exon usage across the test series. a) The total relative amounts of deviating exon usage across the 83 CRC samples were correlated to splicing factor ($n = 280$) expression levels (median Pearson correlation coefficient, $r = -0.23$). This value was stronger than all median Pearson correlations across 1000 permutations (hence, $P < 0.001$), which ranged from -0.17 to 0.18 (median = 0). b) For the observed amounts of deviating exon usage, there was 2.5 times more negatively than positively correlated splicing factor genes (log2-scale), and only 10 out of the 1,000 permuted values showed a stronger tendency towards negative correlation.

Supplementary Figure 4. Validation of the associations between splicing factor expression levels and the overall TIN subtype. a) Pearson correlation coefficients for the 133 splicing factors whose expression levels were significantly correlated to the sample-wise total relative amounts of deviating exon usage in the validation series ($P < 0.05$). There were 129 negatively and 4 positively correlated splicing factors. b) Unsupervised hierarchical clustering analysis of all the 77 samples based on the expression levels of all 280 splicing factor genes. As for the test sample series, the validation series was also separated into clusters of samples with

predominantly low (blue boxes) or high (red boxes) amounts of deviating exon usage. c)

Unsupervised hierarchical clustering analysis of the 30 samples designated with the oTIN subtype based on the expression levels of all 280 splicing factors. As for the test sample series, also the validation samples with low ($n = 21$) and high ($n = 9$) amounts of deviating exon usage were almost completely separated into two different clusters. Both clusters were created using Euclidean distance metrics and complete linkage.

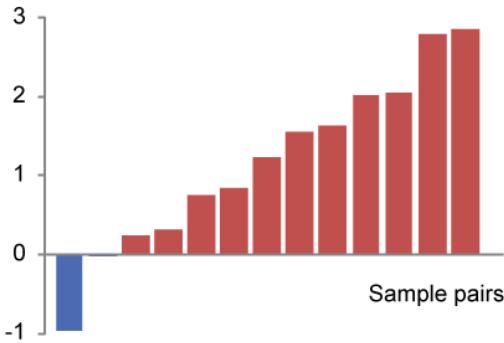
Supplementary Figure 1



Supplementary Figure 2

a

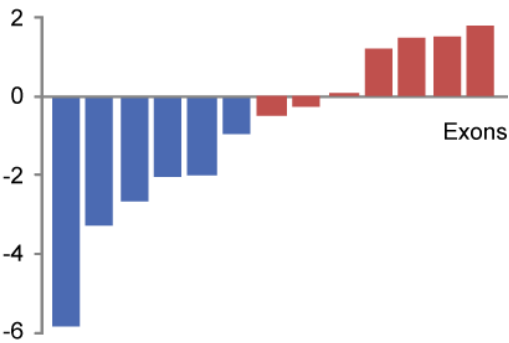
Total relative amounts of deviating exon usage
per sample pair (CRC vs. normal mucosa; log2)



Mean (CRC) = 0.35
Mean (normal mucosa) = -0.83
Mean difference = 1.2, P = 0.003

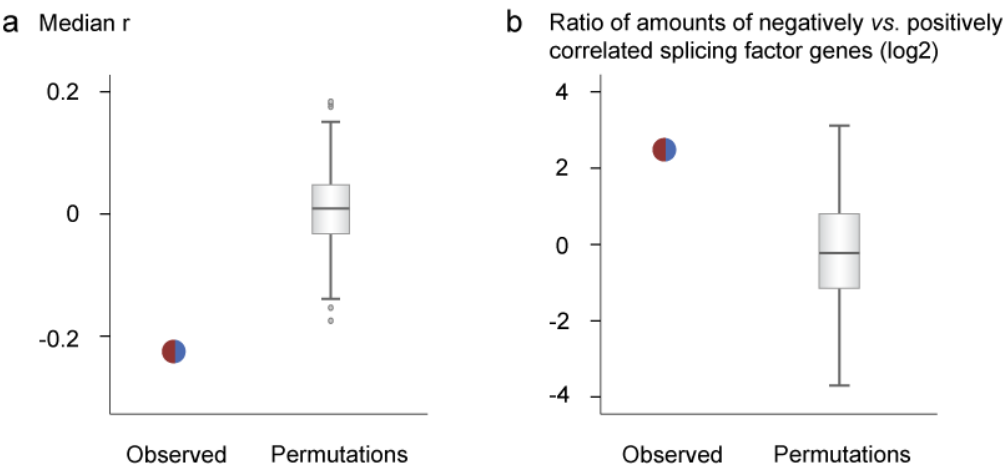
b

Mean paired difference in FIRMA scores
(CRC vs. normal mucosa: log2)

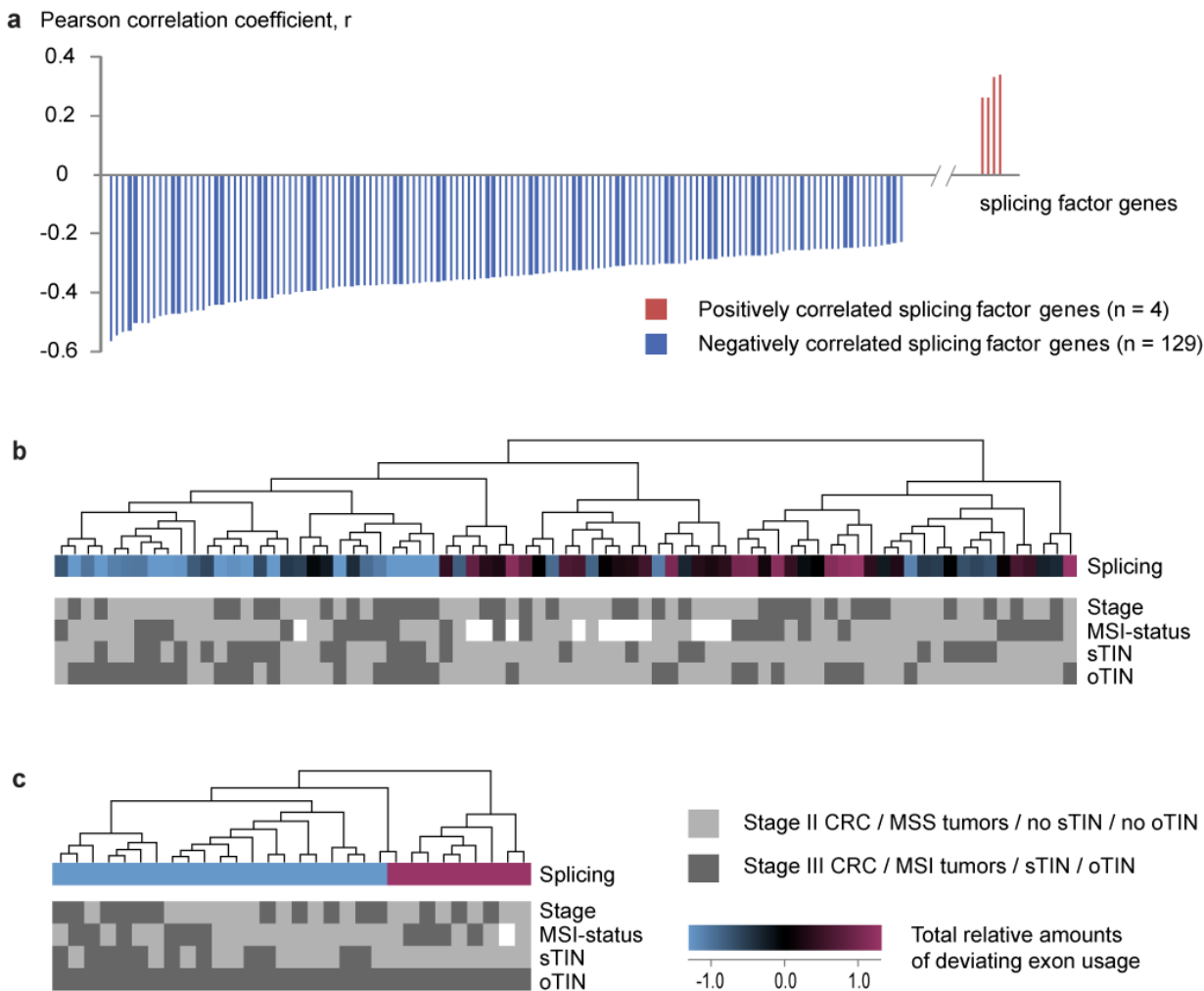


Alternatively spliced exons:
■ Expressed primarily in CRC
■ Expressed primarily in normal mucosa

Supplementary Figure 3



Supplementary Figure 4



Supplementary Table 1. Sources for the gene list of 280 splicing factors

Affymetrix Transcript Cluster ID	HGNC symbol*	Gene Ontology [†]		Swiss-Prot		Gene Cards
		GO:0000398	GO:0005681	Splicing	Spliceosome	
2320048	<i>TARDBP</i>			X		
2325526	<i>SRRM1</i>	X		X	X	
2327259	<i>PPP1R8</i>			X	X	
2331602	<i>PPIE</i>			X	X	
2332711	<i>PPIH</i>		X		X	
2336383	<i>PRPF38A</i>			X	X	
2341565	<i>SFRS11</i>	X		X	X	
2348060	<i>PTBP2</i>			X	X	
2350287	<i>PRPF38B</i>			X		
2351940	<i>DDX20</i>	X		X	X	
2353773	<i>TTF2</i>			X	X	
2356181	<i>RBM8A</i>	X		X	X	
2358171	<i>PRPF3</i>	X	X	X	X	
2358743	<i>SCNM1</i>			X		
2370991	<i>DHX9</i>	X		X	X	
2401275	<i>HNRNPR</i>	X		X		
2402068	<i>SYF2</i>			X	X	
2403470	<i>DNAJC8</i>	X		X	X	
2403740	<i>SFRS4</i>	X		X	X	
2404377	<i>SNRNP40</i>	X	X	X	X	
2406064	<i>SFPQ</i>			X		

2406722	<i>LSM10</i>			X	X	
2407439	<i>SF3A3</i>	X	X	X	X	
2413008	<i>RPS13</i>			X		
2413180	<i>MAGOH</i>	X		X		
2418000	<i>ZRANB2</i>			X		
2421782	<i>RBMX</i>	X		X	X	
2422227	<i>GEMIN8</i>	X		X		
2427930	<i>WDR77</i>	X				
2434159	<i>SF3B4</i>	X	X	X	X	
2435149	<i>TNRC4</i>	X		X		
2448073	<i>IVNS1ABP</i>		X	X	X	
2464499	<i>HNRNPU</i>	X		X		
2468920	<i>CPSF3</i>	X		X	X	
2470654	<i>DDX1</i>	X			X	
2477980	<i>GEMIN6</i>	X		X	X	
2487527	<i>SNRNP27</i>			X		
2487639	<i>PCBP1</i>	X		X	X	
2491702	<i>USP39</i>	X			X	
2514441	<i>PPIG</i>				X	
2544179	<i>SF3B14</i>	X		X	X	
2545811	<i>PPM1G</i>					X
2548970	<i>SFRS7</i>	X		X	X	
2562115	<i>LSM3</i>			X		
2565262	<i>ASCC3L1</i>	X	X		X	
2575134	<i>POLR2D</i>	X		X	X	
2581548	<i>PRPF40A</i>			X	X	
2587747	<i>CIR</i>			X	X	

2593670	<i>SF3B1</i>	x	x	x	x	
2594497	<i>CLK1</i>					x
2594535	<i>PPIL3</i>				x	
2598099	<i>BARD1</i>					x
2611122	<i>TSEN2</i>			x		
2622469	<i>RBM5</i>	x	x	x	x	
2623154	<i>RBM15B</i>	x		x	x	
2629693	<i>PPP4R2</i>			x	x	
2649532	<i>RSRC1</i>					
2655773	<i>POLR2H</i>	x		x	x	
2694617	<i>ISY1</i>			x	x	
2697331	<i>DBR1</i>			x	x	
2709062	<i>SFRS10</i>	x		x	x	
2713074	<i>NCBP2</i>	x		x		
2728448	<i>POLR2B</i>	x		x		
2746269	<i>LSM6</i>			x	x	
2756029	<i>FRG1</i>			x	x	
2763805	<i>DHX15</i>		x	x	x	
2772017	<i>YTHDC1</i>			x		
2775463	<i>HNRNPD</i>	x		x	x	
2790570	<i>PLRG1</i>				x	
2809885	<i>SKIV2L2</i>			x	x	
2812273	<i>PPWD1</i>			x	x	
2812539	<i>SFRS12</i>			x	x	
2824315	<i>ZRSR2</i>		x	x	x	
2829488	<i>DDX46</i>			x	x	
2834093	<i>TCERG1</i>					x

2837499	<i>LSM11</i>			X	X	
2859494	<i>SFRS12IP1</i>			X		
2877141	<i>HNRNPA0</i>	X		X	X	
2881521	<i>RBM22</i>			X		
2882897	<i>GEMIN5</i>	X		X		
2884658	<i>SLU7</i>	X	X	X	X	
2890148	<i>HNRNPH1</i>	X		X	X	
2892738	<i>PRPF4B</i>			X	X	
2893847	<i>SNRNP48</i>		X	X	X	
2904248	<i>SNRPC</i>	X	X	X	X	
2905118	<i>SFRS3</i>	X		X	X	
2908572	<i>CDC5L</i>			X		
2921086	<i>CDC40</i>	X	X		X	
2934089	<i>WTAP</i>			X		
2935475	<i>QKI</i>			X	X	
2944068	<i>DEK</i>					X
2948485	<i>DHX16</i>			X	X	
2949038	<i>BAT1</i>	X		X	X	
2949431	<i>LSM2</i>	X		X	X	
2951674	<i>SRPK1</i>			X		
2952065	<i>PPIL1</i>				X	
2963407	<i>SYNCRIP</i>			X	X	
2964052	<i>SRRP35</i>	X		X	X	
2966253	<i>SFRS18</i>					X
2977690	<i>SF3B5</i>	X		X	X	
3020804	<i>LSM8</i>	X		X		
3041519	<i>TRA2A</i>	X		X	X	

3041550	<i>TRA2A</i>					X
3042421	<i>HNRNPA2B1</i>	x	x	x	x	
3044753	<i>LSM5</i>			x		
3065015	<i>POLR2J</i>	x		x	x	
3066297	<i>SRPK2</i>	x			x	
3071636	<i>RBM28</i>			x	x	
3072014	<i>TNPO3</i>					X
3098935	<i>TGS1</i>	x				
3109191	<i>POLR2K</i>	x		x	x	
3117384	<i>KHDRBS3</i>					X
3131844	<i>LSM1</i>			x	x	
3145020	<i>KIAA1429</i>			x		
3157817	<i>PUF60</i>			x		
3158516	<i>CPSF1</i>	x			x	
3168508	<i>MELK</i>					X
3181302	<i>NCBP1</i>	x		x		
3185558	<i>PRPF4</i>	x	x	x		
3199790	<i>PSIP1</i>					X
3205033	<i>YBX1</i>	x		x	x	
3212294	<i>HNRNPK</i>	x		x	x	
3233547	<i>RBM17</i>			x	x	
3234760	<i>CUGBP2</i>					X
3235932	<i>PRPF18</i>		x	x	x	
3240340	<i>WAC</i>				x	
3249738	<i>HNRNPH3</i>	x		x		
3249788	<i>CCAR1</i>	x		x	x	
3277662	<i>UPF2</i>					X

3286286	<i>HNRNPF</i>	x		x	x	
3306516	<i>SMNDC1</i>			x	x	
3311775	<i>DHX32</i>					x
3331392	<i>CLP1</i>	x		x	x	
3333622	<i>POLR2G</i>	x		x	x	
3335774	<i>SART1</i>			x	x	
3335907	<i>SF3B2</i>	x	x	x	x	
3336422	<i>RBM4</i>			x		
3342525	<i>PCF11</i>			x	x	
3358492	<i>POLR2L</i>	x		x	x	
3368304	<i>WT1</i>					x
3368520	<i>CSTF3</i>	x		x	x	
3372253	<i>CUGBP1</i>	x		x		
3375049	<i>PRPF19</i>			x		
3375340	<i>CPSF7</i>	x		x	x	
3376779	<i>TRPT1</i>			x		
3377044	<i>SF1</i>	x	x		x	
3378411	<i>RBM4B</i>			x		
3382948	<i>CLNS1A</i>	x		x	x	
3387171	<i>CWC15</i>	x		x		
3406421	<i>STRAP</i>			x	x	
3414104	<i>PRPF40B</i>			x	x	
3416036	<i>PCBP2</i>	x		x	x	
3416483	<i>HNRNPA1</i>	x	x	x	x	
3427014	<i>SNRPF</i>	x	x		x	
3435681	<i>ARL6IP4</i>			x		
3438417	<i>SFRS8</i>	x				

3439268	<i>NHP2L1</i>	x		x		
3444195	<i>MAGOHB</i>			x		
3445670	<i>WBP11</i>			x		
3451318	<i>ZCRB1</i>		x	x	x	
3452145	<i>SFRS2IP</i>	x			x	
3453319	<i>DDX23</i>	x	x		x	
3470253	<i>SART3</i>					
3474502	<i>SFRS9</i>	x				x
3475679	<i>ZCCHC8</i>			x	x	
3482112	<i>PABPC1</i>			x	x	
3486807	<i>WBP4</i>			x	x	
3494502	<i>DHX9</i>					
3503224	<i>UPF3A</i>					x
3529082	<i>PABPN1</i>	x		x		x
3533397	<i>SIP1</i>	x	x	x	x	
3533435	<i>PNN</i>			x	x	
3534201	<i>PRPF39</i>			x		
3536706	<i>LGALS3</i>					
3542207	<i>SFRS5</i>	x				x
3543411	<i>RBM25</i>	x		x	x	
3548788	<i>CPSF2</i>	x		x	x	
3550392	<i>PAPOLA</i>	x			x	
3551935	<i>WDR25</i>					
3556888	<i>RBM23</i>					x
3558745	<i>NOVA1</i>			x		x
3573261	<i>SNW1</i>	x	x	x	x	
3584443	<i>SNRPN</i>	x	x	x	x	

3617757	<i>AQR</i>			X	X	
3632107	<i>BRUNOL6</i>			X	X	
3633522	<i>SNUPN</i>					
3642162	<i>SNRPA1</i>	X	X	X	X	X
3642572	<i>SNRNP25</i>		X	X	X	
3645253	<i>SRRM2</i>			X	X	
3646613	<i>A2BP1</i>			X		
3656904	<i>FUS</i>	X		X		
3662750	<i>POLR2C</i>	X		X	X	
3665262	<i>NOL3</i>			X	X	
3665949	<i>PSKH1</i>					
3666189	<i>PRMT7</i>	X		X		X
3667281	<i>SF3B3</i>	X	X	X	X	
3667902	<i>DHX38</i>	X		X	X	
3674960	<i>LUC7L</i>					
3676669	<i>RNPS1</i>	X		X	X	X
3687698	<i>CD2BP2</i>	X		X	X	
3692895	<i>NUDT21</i>	X		X		
3698055	<i>TXNL4B</i>			X		
3708704	<i>POLR2A</i>	X		X		
3722554	<i>DHX8</i>		X	X	X	
3726772	<i>CROP</i>			X	X	
3730899	<i>DDX42</i>					
3734865	<i>TSEN54</i>			X		X
3739812	<i>GEMIN4</i>	X		X	X	
3740479	<i>PRPF8</i>	X	X	X	X	
3742708	<i>C1QBP</i>					

3759356	<i>EFTUD2</i>	x	x	x	x	x
3764103	<i>SFRS1</i>	x			x	
3766893	<i>DDX5</i>			x	x	
3771800	<i>SFRS2</i>	x		x	x	
3773312	<i>EIF4A3</i>			x	x	
3774331	<i>THOC4</i>	x		x	x	
3781082	<i>SNRPD1</i>	x			x	
3795680	<i>THOC1</i>			x		
3804452	<i>BRUNOL4</i>	x		x		
3814734	<i>TXNL4A</i>	x	x	x	x	
3815165	<i>PTBP1</i>	x		x	x	
3816333	<i>SF3A2</i>	x	x		x	
3819543	<i>HNRNPM</i>	x	x	x	x	
3820161	<i>UBL5</i>					
3820342	<i>PPAN</i>			x	x	x
3821805	<i>MORG1</i>	x	x	x		
3825383	<i>UPF1</i>					
3833757	<i>SNRPA</i>	x	x	x	x	x
3834089	<i>HNRNPUL1</i>	x		x	x	
3835983	<i>SFRS16</i>			x		
3836044	<i>GEMIN7</i>	x		x	x	
3838185	<i>SNRNP70</i>	x	x	x	x	
3838757	<i>SCAF1</i>			x		
3841231	<i>PRPF31</i>	x			x	
3841310	<i>TSEN34</i>			x		
3842345	<i>U2AF2</i>	x	x	x	x	
3844952	<i>POLR2E</i>	x		x	x	

3845868	<i>LSM7</i>	x		x	x	
3847814	<i>KHSRP</i>			x		
3848437	<i>XAB2</i>			x	x	
3852691	<i>DDX39</i>	x		x	x	
3854892	<i>LSM4</i>		x	x	x	
3855410	<i>SFRS14</i>			x		
3855660	<i>SF4</i>	x		x	x	
3859915	<i>U2AF1L4</i>			x	x	
3860277	<i>POLR2I</i>	x		x	x	
3861617	<i>HNRNPL</i>	x		x	x	
3865568	<i>SNRPD2</i>	x	x		x	
3877776	<i>SNRPB2</i>	x	x	x	x	
3882720	<i>RALY</i>			x	x	
3884922	<i>DHX35</i>			x	x	
3886050	<i>SFRS6</i>	x				
3890154	<i>CSTF1</i>	x		x	x	
3893849	<i>PRPF6</i>	x	x	x	x	
3894995	<i>SNRPB</i>	x	x	x	x	
3899954	<i>CRNKL1</i>	x	x	x	x	
3904226	<i>RBM39</i>			x		
3918696	<i>SON</i>					
3928866	<i>SFRS15</i>					x
3933999	<i>U2AF1</i>	x	x	x	x	x
3940185	<i>SNRPD3</i>	x	x		x	
3945133	<i>POLR2F</i>	x		x	x	
3952508	<i>DGCR14</i>			x	x	
3955875	<i>TFIP11</i>		x	x	x	

3956984	<i>ZMAT5</i>		x	x	x	
3957260	<i>SF3A1</i>	x	x		x	
3959203	<i>RBM9</i>			x		
3961955	<i>PHF5A</i>	x	x	x		
3962054	<i>NHP2L1</i>					
3969946	<i>ZRSR2</i>		x	x	x	x
3975987	<i>RBM10</i>			x		
3980887	<i>NONO</i>			x		
3984536	<i>CSTF2</i>	x		x	x	
3984779	<i>HNRNPH2</i>	x		x	x	
4000269	<i>GEMIN8</i>	x		x		
4009238	<i>SMC1A</i>	x		x	x	
4019570	<i>UPF3B</i>	x		x	x	
4020444	<i>THOC2</i>			x		
4045780	<i>FUSIP1</i>	x		x		
4052378	<i>SNRNP35</i>			x	x	

*HGNC: Human genome nomenclature committee (<http://www.genenames.org>). †GO:0000398

denotes “nuclear mRNA splicing, via spliceosome”, GO:00005681 denotes “spliceosomal complex”.

Supplementary Table 2. Alternative splicing (FIRMA) scores for known alternative splicing events between CRC and normal colonic mucosa

Alternatively spliced gene	Affymetrix probe set targeting alternatively spliced exon(s)		Mean difference in paired FIRMA scores [*]	Source
	Expressed in CRC	Expressed in normal mucosa		
<i>SLC39A14</i> (Ex4A and 4B)	- 3089381 and 3089382	3089375 -	-5.8 1.5	Thorsen <i>et al.</i>
<i>ACTN1</i> (Ex19a and 19b [†])	3569830	-	1.2	Gardina <i>et al.</i> [‡]
<i>ATP2B4</i> (Ex21)	-	2375764, 2375765 and 2375766	-3.3	
<i>VCL</i> (Ex19)	-	3252128 and 3252129	-2.7	
<i>CALD1</i> (extended Ex5 and Ex6)	-	3025632 and 3025638	-2.0	
<i>SLC3A2</i> (Ex2, 3 and 4)	3333716 and 3333717	-	0.1	
<i>COL6A3</i> (Ex3 and 4)	2605390, 2605391, 2605395 and 2605396	-	-0.5	
<i>COL6A3</i> (Ex6)	2605386	-	1.5	
<i>CTTN</i> (Ex11)	3338589	-	1.8	
<i>FN1</i> (Ex25)	-	2598321	-0.9	
<i>TPM1</i> (Ex7 and 8)	3597384 -	- 3597388	-0.2 -2.0	

^{*}CRC vs. normal mucosa; n = 13 pairs; log2

[†]Probe set targeting exon 19b (3569827) was not assigned a FIRMA score using the custom made annotation file

[‡]Splicing events marked with confidence level "good"

Supplementary Table 3. Genes correlated to relative amounts of aberrant alternative exon usage within the splicing factor gene set and one hundred random gene sets of equal size in the test sample series

	Median of 100		
	Splicing factor gene set	random gene sets	<i>P</i> -value
n (significantly correlated genes)	151	102	< 0.0001 [*]
n (negative significant correlation; positive significant correlation)	144; 7	78; 24	< 0.0001 [†]
Mean $r^{\ddagger}$, all 282 genes	-0.21	-0.07	< 0.0001 [§]
Mean $r^{\ddagger}$, significantly correlated genes	-0.33	-0.19	< 0.0001 [§]
Mean $ r ^{**}$, all 282 genes	0.24	0.18	<0.0001 [§]
Mean $ r ^{**}$, significantly correlated genes	0.36	0.32	0.002 [§]

^{*} *P*-value by Fisher's exact test comparing the amounts of significantly correlated genes to non-significantly correlated genes in the two gene sets

[†] *P*-value by Fisher's exact test comparing the amounts of significant negatively correlated genes to significant positively correlated genes in the two gene sets

[‡] r = Pearson correlation coefficient

[§] *P*-value by independent samples t-test for equality of means in the two gene sets

^{**} $|r|$ = absolute value of the Pearson correlation coefficient

Supplementary Table 4. Ten years disease specific survival rates for groups of patients in the test sample series stratified by different thresholds of the skewed TIN (sTIN) phenotype

Threshold*	Disease specific survival		<i>P</i> -value [‡]	n		n (events)	
	(%) [†]						
	Within	Outside		Within	Outside	Within	Outside
	threshold	threshold		threshold	threshold	threshold	threshold
+/-0.9	54	14	0.008	76	7	35	6
+0.9	51	0	0.12	82	1	40	1
-0.9	53	17	0.03	77	6	36	5
+/-0.8	55	20	0.02	73	10	33	8
+0.8	51	50	0.93	81	2	40	1
-0.8	55	13	.008	75	8	34	7
+/-0.7	56	17	0.001	71	12	31	10
+0.7	51	33	0.31	80	3	39	2
-0.7	55	11	.001	74	9	33	8
+/-0.6	55	33	0.04	65	18	29	12
+0.6	52	33	0.18	77	6	37	4
-0.6	54	33	0.13	71	12	33	8

*threshold = (relative amounts of aberrant exon inclusion) - (relative amounts of aberrant exon exclusion) [†] Survival rates from Kaplan Meier survival analysis. Events: dead from disease (n = 43), censored: survival throughout follow-up period (n = 44), ignored: recurrences [‡]Significance level from log rank test for equality of survival distributions

Supplementary Table 5. Disease specific survival analyses for patients in the two independent sample series divided into groups by the TIN phenotypes

A. Test series

Phenotype	Ten years disease		<i>P</i> -value [*]	n		n (events)	
	specific survival (%)						
	No			No		No	
	Phenotype	phenotype		Phenotype	phenotype	Phenotype	phenotype
sTIN [†]	17	56	0.001	12	71	10	31
oTIN [‡]	43	52	0.43	14	69	8	33
Either sTIN or							
oTIN	33	58	0.02	24	59	16	25
Both sTIN and							
oTIN	0	52	0.02	2	81	2	39

B. Validation series

Phenotype	Five years disease		<i>P</i> -value [*]	n		n (events)	
	specific survival (%)						
	No			No		No	
	Phenotype	phenotype		Phenotype	phenotype	Phenotype	phenotype
sTIN [†]	80	88	0.57	24	53	4	6
oTIN [‡]	79	91	0.20	30	47	6	4
Either sTIN or oTIN	84	88	0.83	43	34	6	4

Both sTIN and

oTIN	64	91	0.03	11	66	4	6
------	----	----	------	----	----	---	---

*Significance level from log rank test for equality of survival distributions. †sTIN: Preferential exon inclusion or skipping (difference in relative amounts of aberrant exon skipping and inclusion $> \pm 0.7$) ‡oTIN: Total relative amounts of aberrant splicing $> \pm 1.0$.