

SUPPLEMENTARY TABLE AND FIGURES LEGENDS

Figure S1-9. U2/U12-like Non-canonical splice junctions present at ETV1, IDH3B, ITPR1, NRK, C1orf63, POLR2E, RBFOX2 and TSPYL2 were validated by RT-PCR and Sanger sequencing. However, the U2/U12-like non-canonical splice junction present at CTNNB1 was not detected in the sequenced subclones. A) UCSC genome browser images show the identified non-canonical U2/U12-like splice junctions and the Sanger sequences that confirm their existence. B) RT-PCR of the non-canonical U2/U12-like splice junctions using 3 different RTs. The correspondence of RT-PCR products and expected isoforms are show to the right of each gel image, where § indicate that the isoform is not annotated in GENCODE v.17 and * indicate that the presence of the isoform was corroborated by Sanger sequencing.

Table S1. The primers used for RT-PCRs of the U2/U12-like and non-U2/U12 non-canonical splice junctions.

Table S2. Odd ratios and Pearson Chi-square P-values calculated of the differences of IE densities between the different regions of U2/U12-like non-canonical and canonical splice sites.

Table S3. Odd ratios and Pearson Chi-square P-values calculated of the differences of IE densities between the different regions of U2/U12-like non-canonical and alternative canonical splice sites.

Figure S10. A picture of UCSC genome browser shows the detection of the non-U2/U12 splice junction of XBP1 gene. Red letters highlight the 6-nt direct repeat associated to this splice junction. As non-canonical splice sites present in GENCODE and in our annotation are located in direct repeats, the two alignments are equivalent.

Figure S11-12. Non-U2/U12 splice junctions present at PSENEN and NAPA were tried to validate through RT-PCR and Sanger sequencing. A) Pictures of UCSC genome browser show the non-U2/U12 splice sites and the alignment of the Sanger sequences. B) RT-PCR of the non-U2/U12 splice junctions using 2 different RTs. C) Secondary RNA structures associated to the non-U2/12 splice junctions. Direct repeats are highlighted in red and orange.

Figure S1

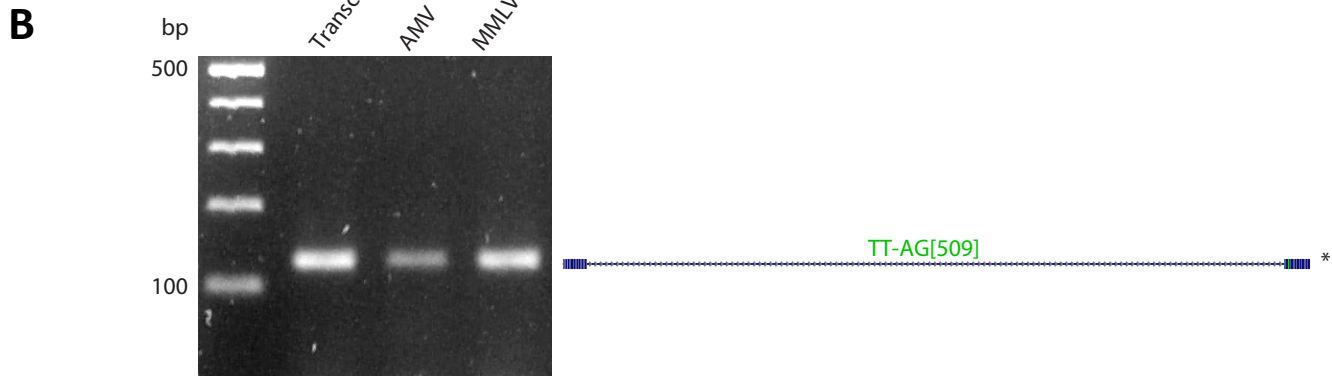
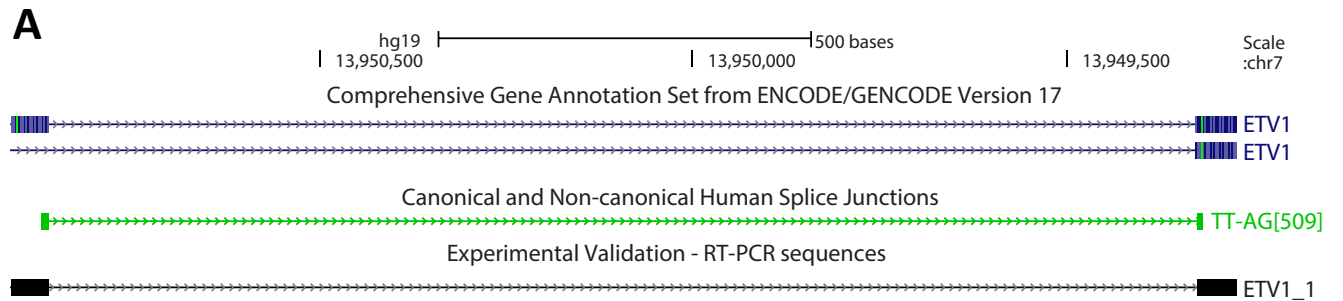


Figure S2

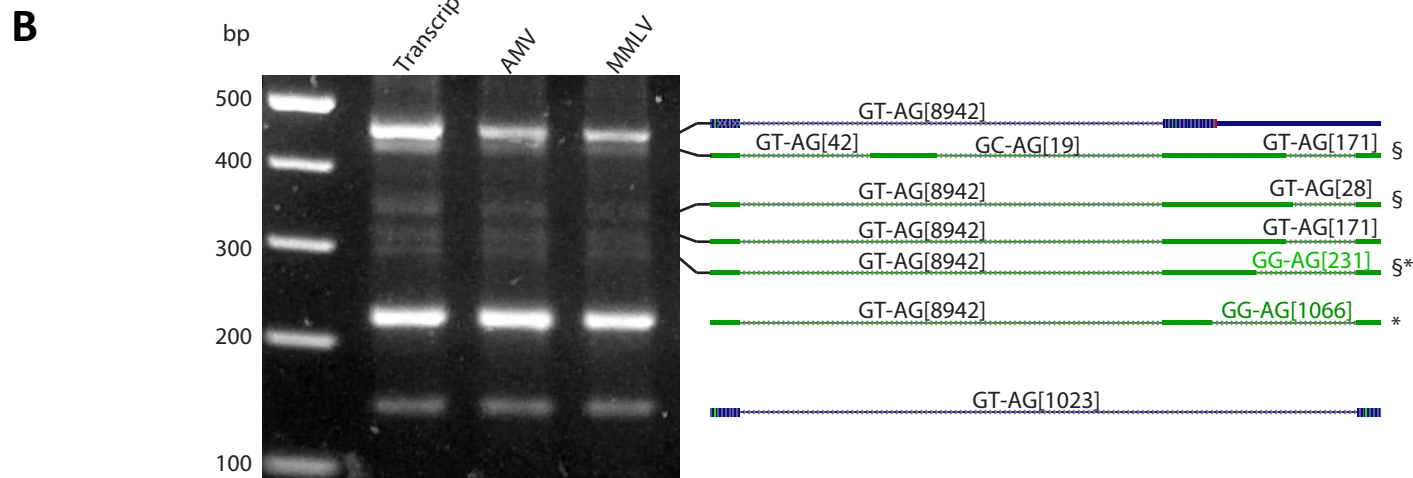
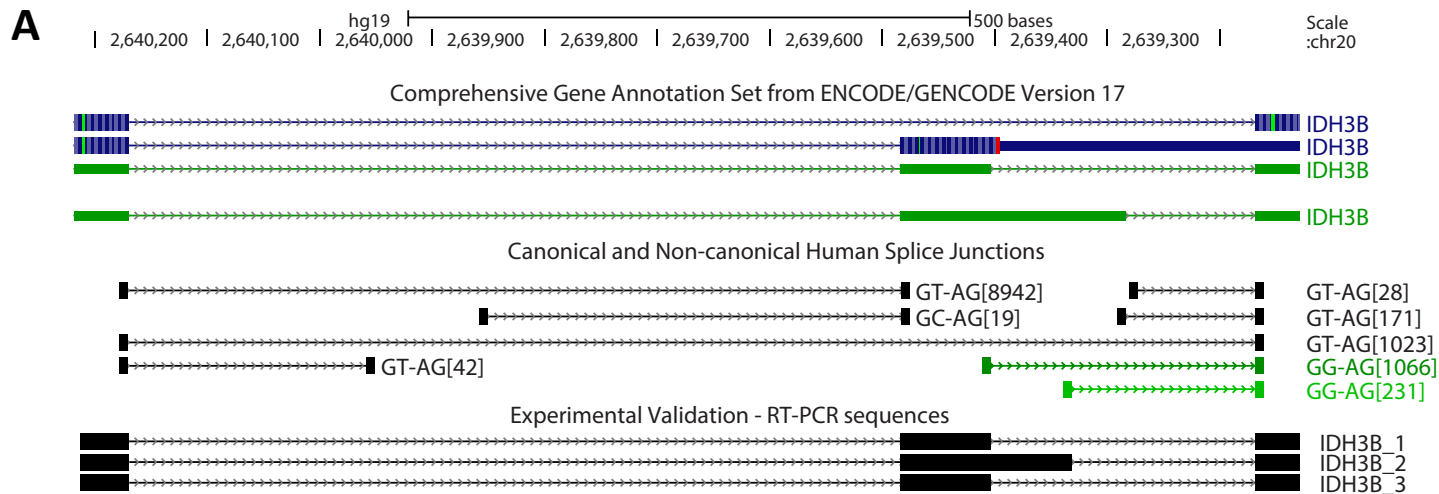
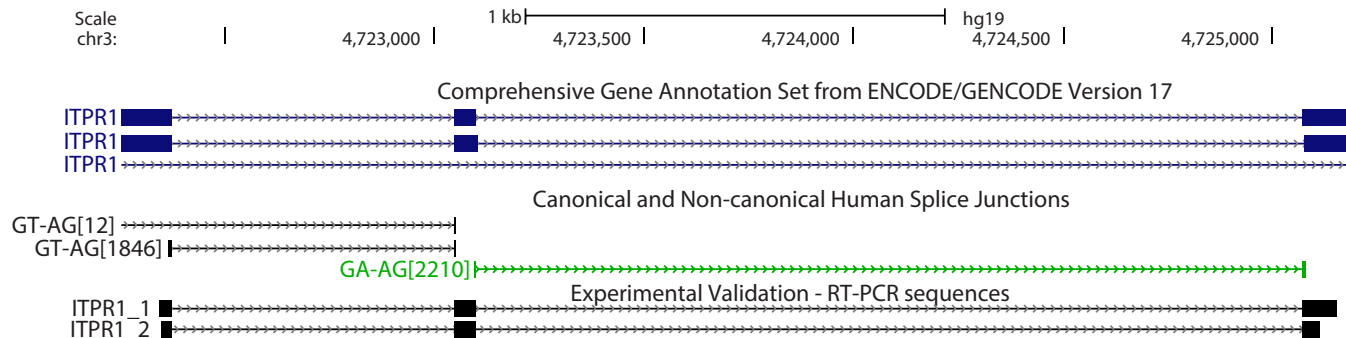


Figure S3

A



B

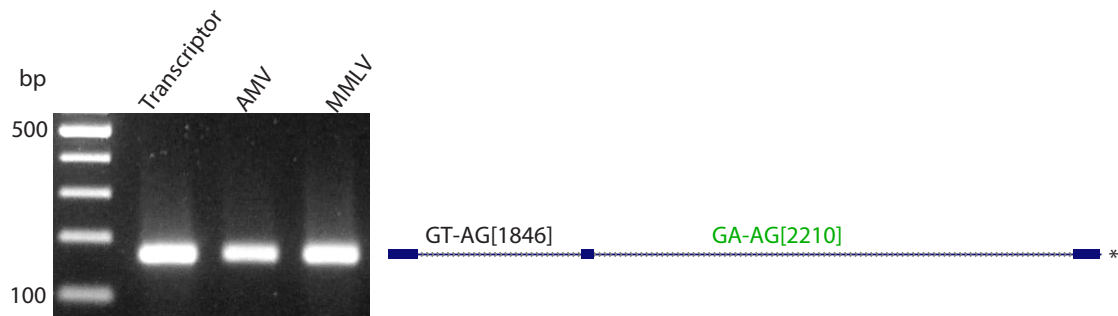
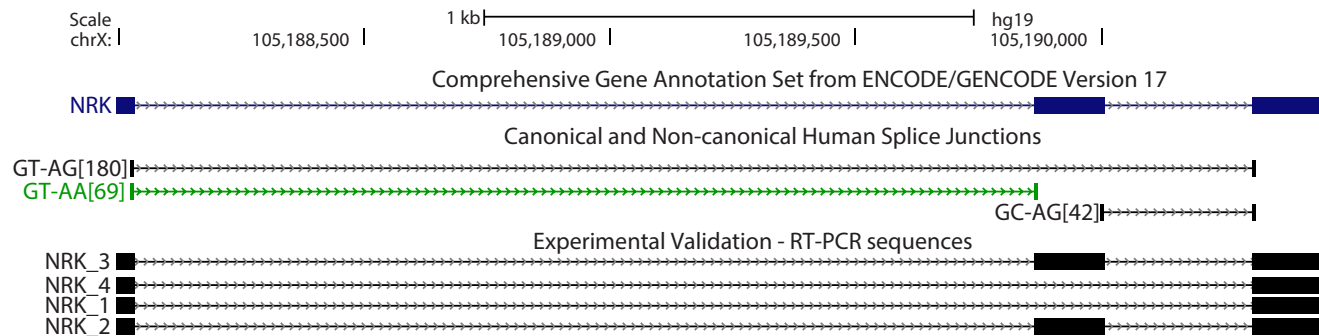


Figure S4

A



B

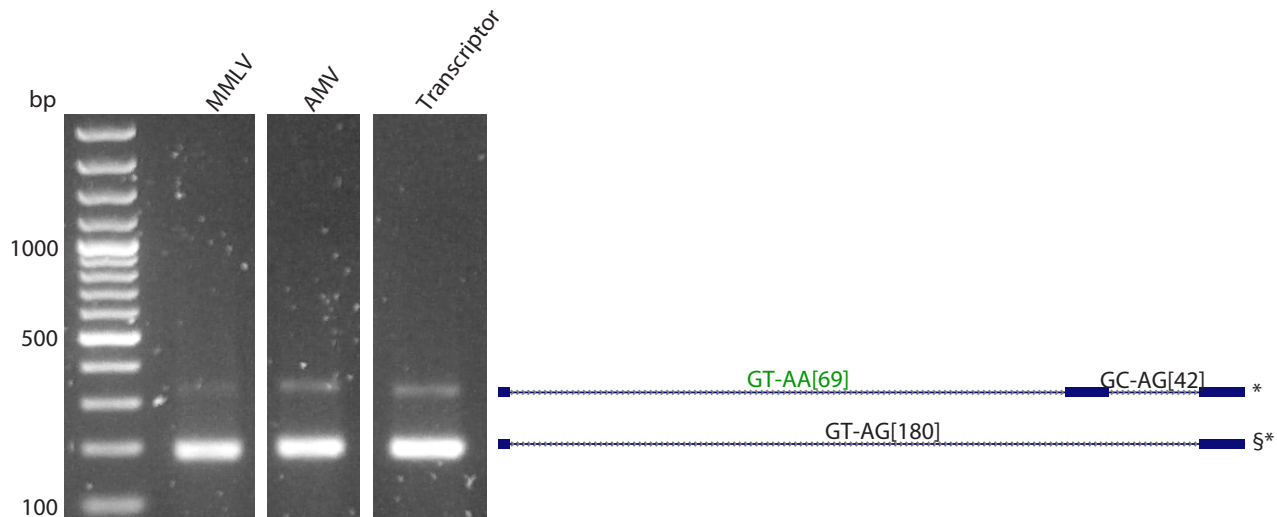
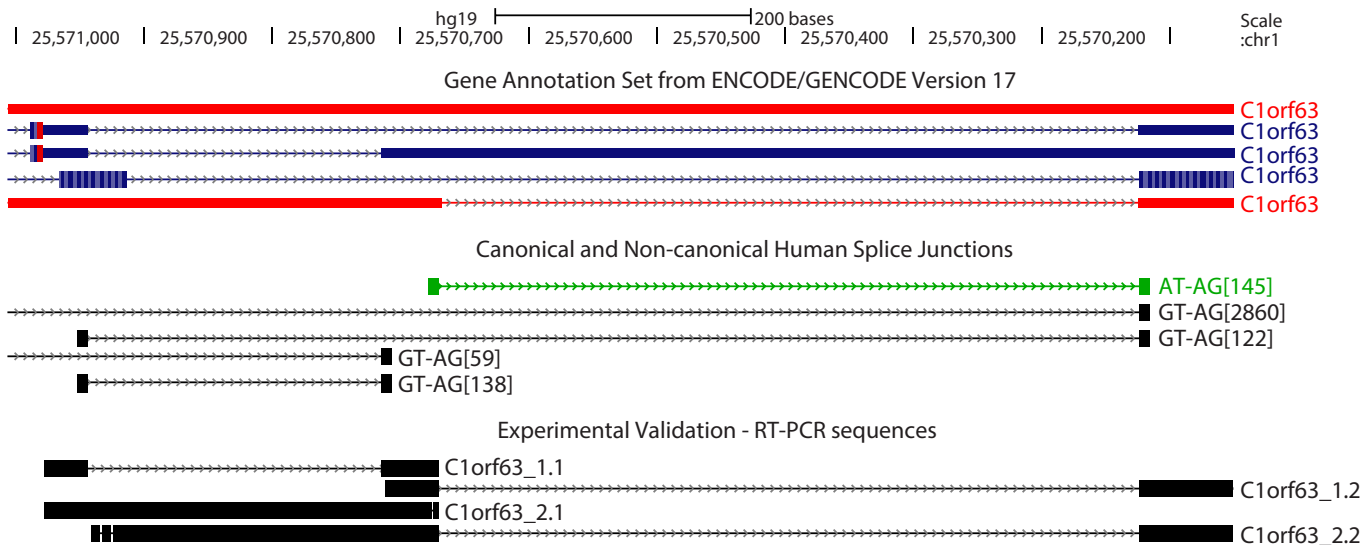


Figure S5

A



B

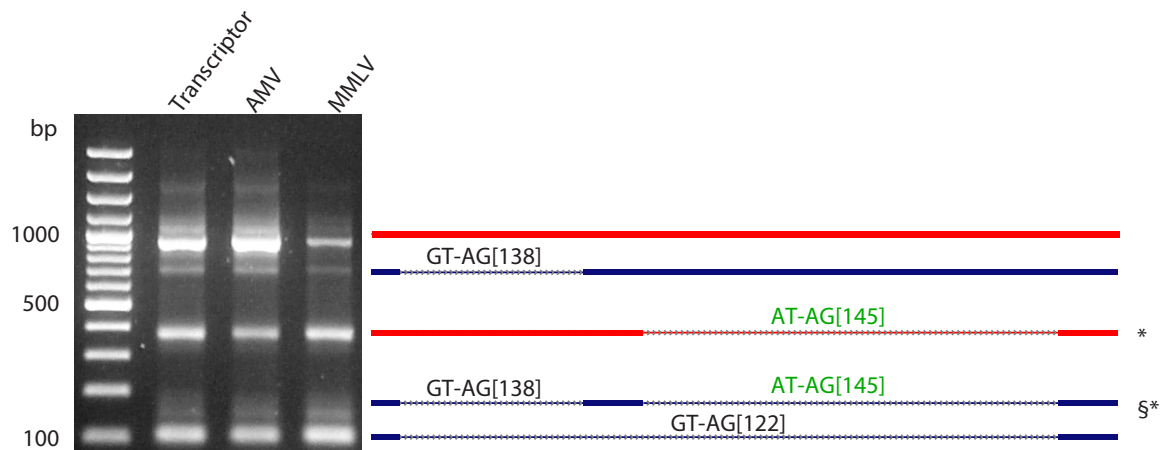
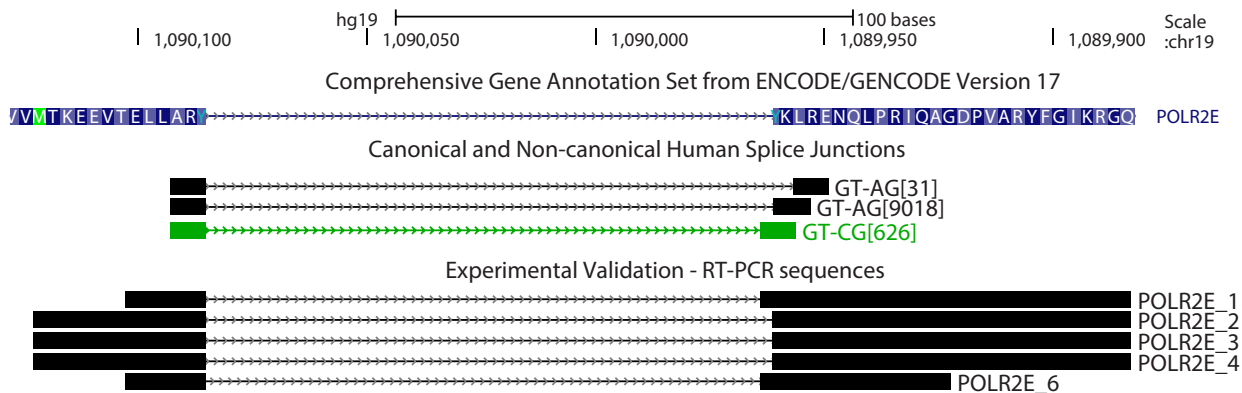


Figure S6

A



B

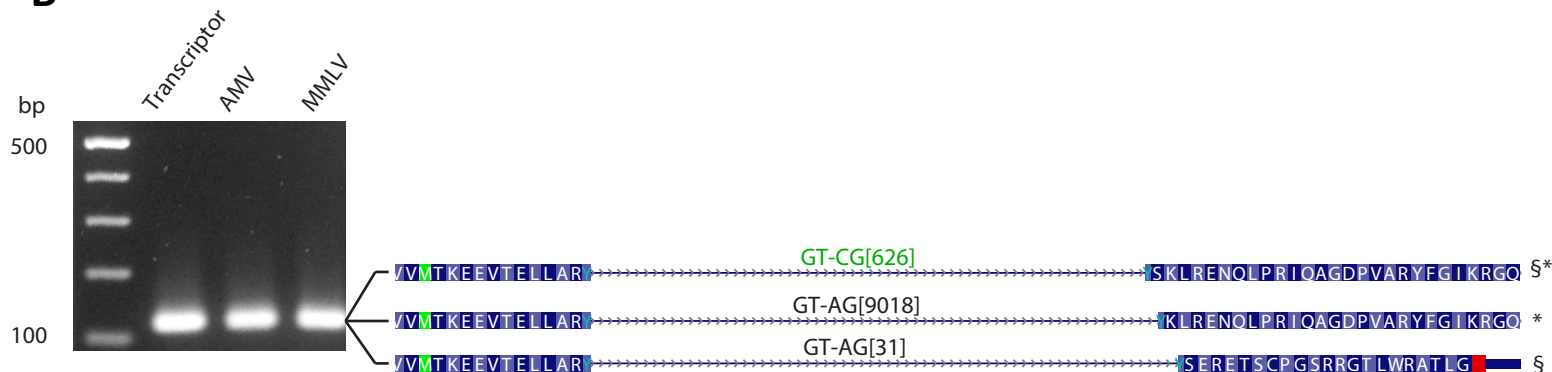
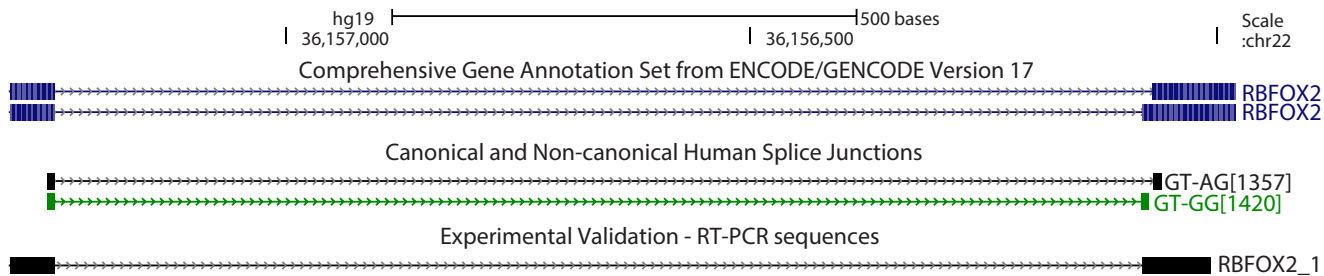


Figure S7

A



B

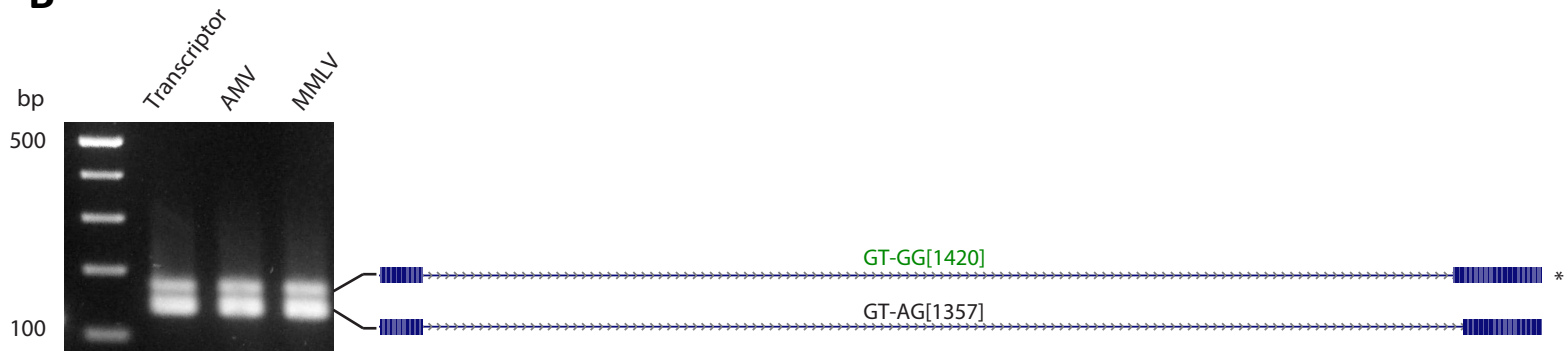


Figure S8

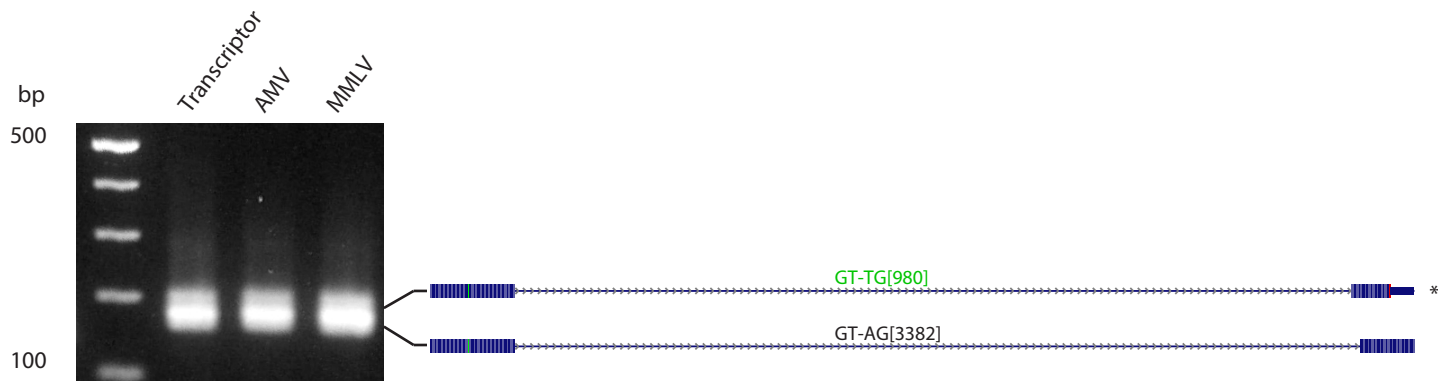
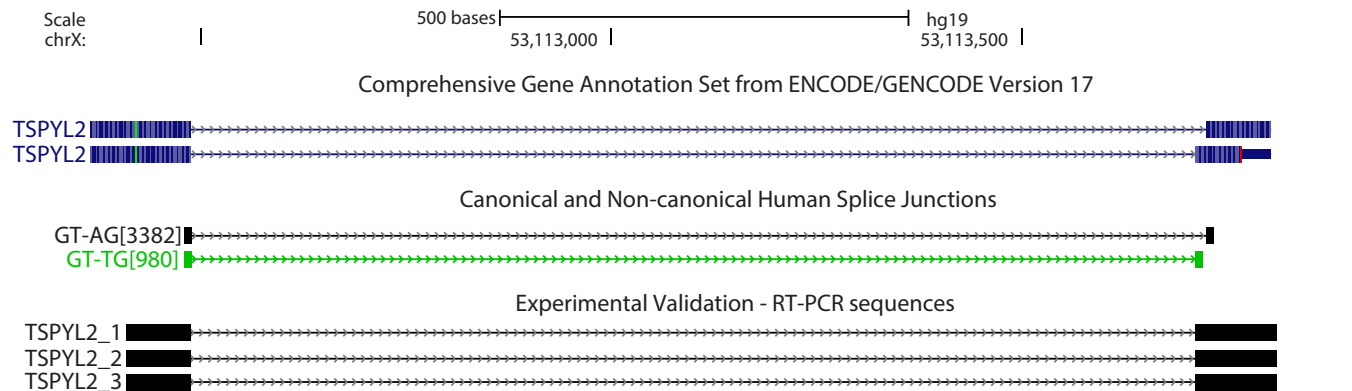
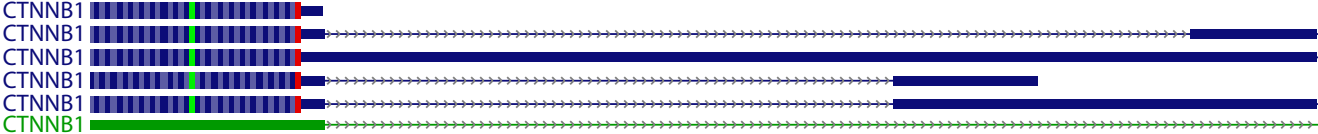


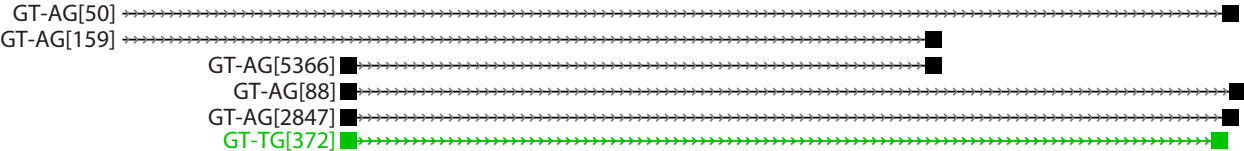
Figure S9

Scale chr3: 41,280,800 | 41,280,900 | 200 bases | 41,281,000 | 41,281,100 | hg19 | 41,281,200 | 41,281,300 |

Comprehensive Gene Annotation Set from ENCODE/GENCODE Version 17



Canonical and Non-canonical Human Splice Junctions



Experimental Validation - RT-PCR sequences

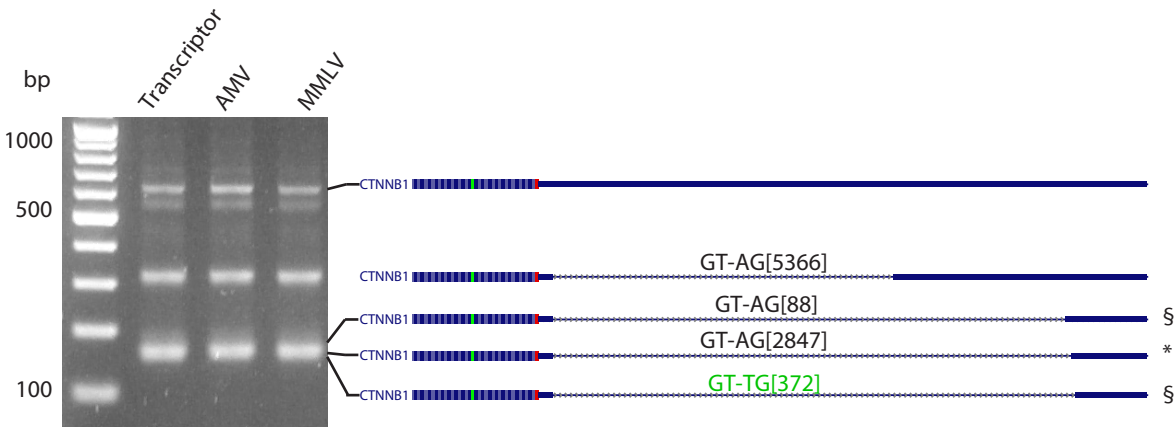
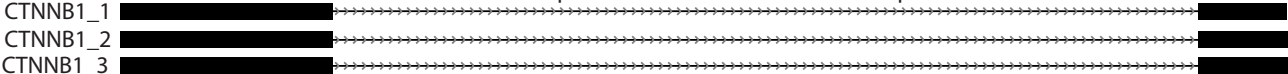


Table S1

Gene	Primers (5'→3')
ETVI	AGTGCCTAGCTGCCACTCCATTTA CTGGGACAACACAGGTGTCATCAT
IDH3B	CATGATCGCAGATGCGGTGAAGAA TTACATGGGTATGGACAGGGCCTA
ITPR1	CAGCCAAGAAGGGCCAAGTAATGT AGTCATGCATCGTCAAGTGGAGCA
NKR	CTTTATGCATGGGCACCAAAGTCC TCTTTGGCAGGGTCACATCAGACA
C1orf63	GTGGTCTGCAGACATTGATGAAGAA AGCTATGCTTCTTTGCTGGGTAGG
POLR2E	ATGACCAAGGAGGAGGTGACAGAG TGCCACGCTTTATCCCAAAGTAG
RBFOX2	ATGTGTCCCTAGGCAATGATGCAG TCAAATGGGCTCCTCTGAAAGCG
CTNNB1	ATATTCCTCGTTAGGGCAGGCG GTCAGTTGCCAATGGGACATTCT
CCNG1	ATATTCCTCGTTAGGGCAGGCG GTCAGTTGCCAATGGGACATTCT
PSENE1	GGGTTTGCTTTCCTGCCTTT ATGTCAGCAGAGAACGTGGGACAT
NAPA	ATCAAGAAGACCATCCAGGGCGAT AAGCATTAGGAAGGAGCAGGTGT

Table S2

U2/U12-like splice sites	Region	Identity elements	Odd ratio	P-value Pearson chi-square	
5' non-canonical	5' exon	IIE	1.3855	3.4466E-09	
		EIE	1.1474	1.2380E-05	
	5' intron	IIE	1.1375	1.6969E-03	
		EIE	1.1493	1.4575E-04	
	3' intron	IIE	1.0039	9.3516E-01	
		EIE	1.1166	3.5408E-03	
	3' exon	IIE	1.3856	3.4217E-09	
		EIE	1.1102	7.4518E-04	
	3' non-canonical	5' exon	IIE	0.9406	2.4444E-01
			EIE	1.1047	9.7384E-05
5' intron		IIE	1.1069	1.5311E-03	
		EIE	1.0146	6.2872E-01	
3' intron		IIE	1.1167	1.8233E-03	
		EIE	1.1833	3.1760E-09	
3' exon		IIE	0.9406	2.4495E-01	
		EIE	1.1191	5.0888E-06	

Table S3

U2/U12-like splice sites	Region	Identity elements	Odd ratio	P-value Pearson chi-square
5' non-canonical	5' exon	IIE	1.3187	6.5445E-07
		EIE	1.1781	2.0693E-07
	5' intron	IIE	1.1229	4.7619E-03
		EIE	1.1761	9.5263E-06
	3' intron	IIE	1.0014	9.9628E-01
		EIE	1.1148	4.3347E-03
3' exon	IIE	1.3189	6.4759E-07	
	EIE	1.1400	2.5601E-05	
3' non-canonical	5' exon	IIE	0.8952	3.7673E-02
		EIE	1.1342	8.8481E-07
	5' intron	IIE	1.0927	6.0074E-03
		EIE	1.0382	2.1624E-01
	3' intron	IIE	1.1139	2.4714E-03
		EIE	1.1814	4.9915E-09
3' exon	IIE	0.8953	3.7870E-02	
	EIE	1.1491	1.9077E-08	

Figure S10



Gene Annotation from ENCODE/GENCODE Version 17

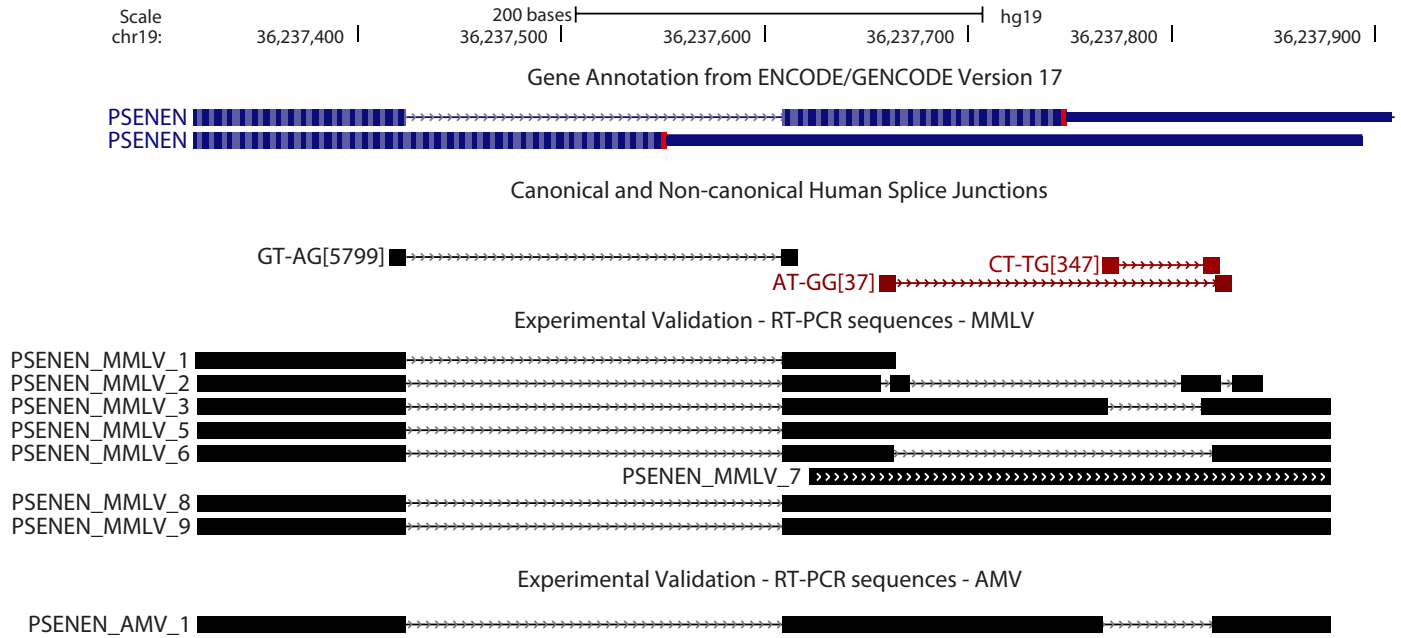


Canonical and Non-canonical Human Splice Junctions

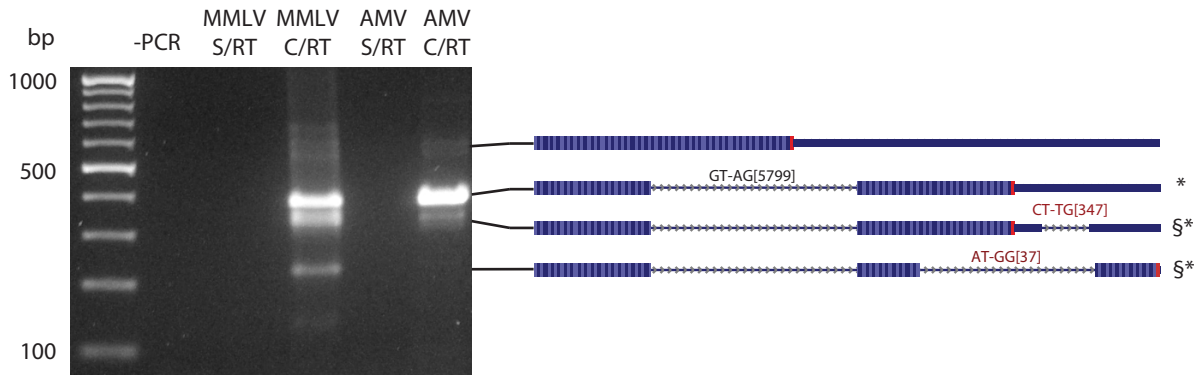


Figure S11

A



B



C

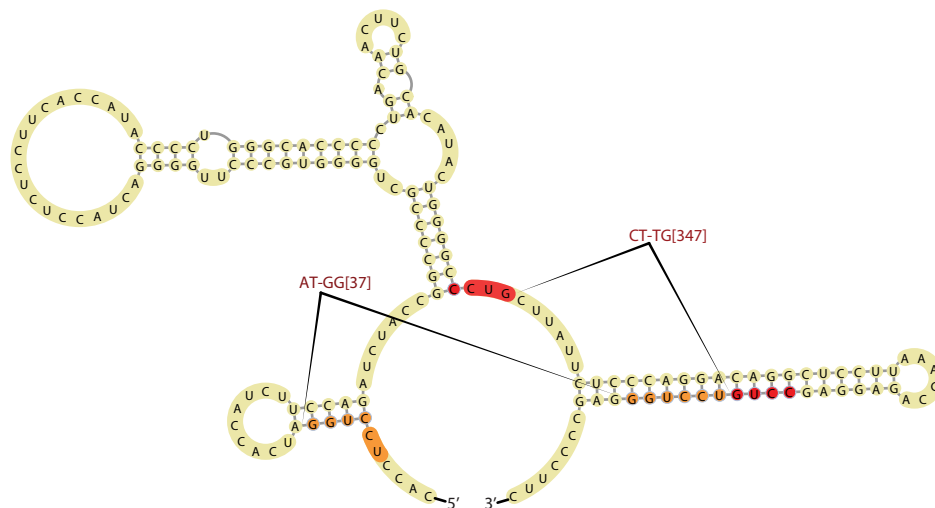


Figure S12

