

Table S4

read#	vector/ sample	#	vector specific sequence length	splice site used	cellular exon/ gene	Blat /UCSC(GRCh37/hg19)	comment	result
1	LV	A01	-	-	-	-	TOPO religated	-
1	LV	A02	108	B	STAT5B; 3rd Exon	chr17:40,379,546-40,379,643	SD matches but only 3rd Exon (60 nt 5'missing); (100% match to STAT5B; 97% to STAT5A)	fusion transcript
1	LV	A03	80	SA4	POLA2; Exon 1-7	chr11:65,029,809-65,047,096	lacks part of 1st Exon corresponding to 5'UTR	fusion transcript
1	LV	A04	78	SA4	homolog of Pongo FAM126b; 1st Exon	chr2:201,936,177-201,936,349	SD matches; 15 nt at 5' missing	fusion transcript
1	LV	A05	-	-	-	-	both primers with 14 nt inbetween	unspecific
1	LV	A06	-	-	-	-	TOPO religated	-
1	LV	B01	80	SA4	EEF1G; Exon 1-5	chr11:62,338,439-62,341,332	analogue 1_F01	fusion transcript
1	LV	B02	78	SA4	RPL22; 1st+2nd Exon	chr1:6,257,712-6,259,653	1st Exon: 26 nt at 5' missing; 2nd Exon 4 nt 5' missing; SD to LV matches	fusion transcript
1	LV	B03	76	SA3	COBRA1; Exon 1-4	chr9:140,149,981-140,151,505	Exon 1 incomplete at 5'	fusion transcript
1	LV	B04	306	-	CRNDE; last intron, antisense	chr16:54,953,234-54,953,354	complete provirus including dU3	read through
1	LV	B05	142	-	-	-	~analogue 1_D05	LTR transcript
1	LV	B06	-	-	-	-	poor sequencing results	-
1	LV	C01	-	-	-	-	DWcrypSA2 sequence present, but no other LV leader sequence	unspecific
1	LV	C02	80	SA4	FAM126b	chr2:201,936,178-201,936,351	analogue 1_A04	fusion transcript
1	LV	C03	79	SA4	SIRPB1; 1st Exon + novel alternative(?) Exon in first annotated intron	chr20:1,599,788-1,600,655	novel alternative(?) Exon in first annotated intron; compare 2_H06: similar but SA3	putative fusion transcript
1	LV	C04	80	SA4	EEF1G; Exon 1-5	chr11:62,338,439-62,341,332	analogue 1_F01	fusion transcript
1	LV	C05	75	SA3	CHKA; Exon 6-11	chr11:67,829,420-67,837,727	premature termination of RT?	fusion transcript
1	LV	C06	80	SA4	ZNF426; Exon 1-5	chr19:9,643,521-9,649,290		fusion transcript
1	LV	D01	149	-	-	-	~analogue 1_D05	LTR transcript
1	LV	D02	150	-	-	-	~analogue 1_D05	LTR transcript
1	LV	D03	80	SA4	EEF1G; Exon 1-5	chr11:62,338,439-62,341,332	analogue 1_F01	fusion transcript
1	LV	D04	81	SA4	CLCN7; internal exon	chr16:1,509,110-1,509,146	premature 5'RACE termination or novel transcription initiation? SD-SA4 junction fits	fusion transcript
1	LV	D05	138	-	-	-	incomplete 5' RACE (TAR deletion?)	LTR transcript
1	LV	D06	79	SA4	FBX038; 1st Exon	chr5:147,763,546-147,763,602		fusion transcript
1	LV	E01	80	SA4	EEF1G; Exon 1-5	chr11:62,338,439-62,341,332	analogue 1_F01	fusion transcript
1	LV	E02	-	B	STAT5B;3rd Exon (partial)	chr17:40,379,550-40,379,643	analogue 1_A02	fusion transcript
1	LV	E03	77	SA4	MECP2; 1st+2nd Exon	chrX:153,357,644-153,363,130		fusion transcript
1	LV	E04	77	SA3	COX6B1; 1st+2nd Exon	chr19:36,139,241-36,142,250		fusion transcript
1	LV	E05	81	(SA4)	6 nt		LV leader seq. Until SA4 + 6 nt	-
1	LV	E06	74	SA4	22nt		LV leader sequence until SA4 + 22 nt (not mappable)	-
1	LV	F01	80	SA4	EEF1G; Exon 1-5	chr11:62,338,439-62,341,332	=1_H05, 1_B01, 1_E01, 1_C04, 1_D03	fusion transcript
1	LV	F02	76	SA3	GATAD2B; 1st Exon	chr1:153,895,212-153,895,338	lacks part of 1st Exon	fusion transcript
1	LV	F03	109	B	STAT5B;3rd Exon (partial)	chr17:40,379,547-40,379,643	analogue 1_A02	fusion transcript
1	LV	F04	-	-	ZNF426		~analogue 1_H04; read length insufficient	-
1	LV	F05	59	novel SA in PBS	MAP4K2; Exon 10-12	chr11:64,567,583-64,568,363	Exon 10 extended into intron+X	fusion transcript
1	LV	F06	79	SA4	SAPS2/BC000976; 1st Exon	chr22:50,810,505-50,810,577		fusion transcript
1	LV	G01	-	-	-	-	no sequence retrieved	-
1	LV	G02	45	-	-	-	unspecific binding of SMARTer oligo in leader sequence (SMARTer oligo residues present/not mediated by NUP)	unspecific
1	LV	G03	80	SA4	TRABD; Exon 2-5	chr22:50,631,489-50,633,468	lacks 1st Exon and part of 2nd exon corresponding to 5'UTR/ 5'End homolog to bos TRABD (->Exon 1-4)	fusion transcript
2	LV	G04	79	SA4	ZBTB43; 1st+2nd Exon	chr9:129,567,320-129,571,194	Exon 1 incomplete at 5'	fusion transcript
1	LV	G05	72	SA3	HMHA1; Exon 4-11	chr19:1,073,525-1,078,036	Exon 4 partial	fusion transcript
1	LV	G06	73	SA3	EEF1A1; 1st Exon	chr6:74,230,723-74,230,754	73 bp between SA2 and Exon1 not mappable; analogue to 2_F06	-
1	LV	H01	80	SA4	PRMT1; Exon 1-6	chr19:50,180,528-50,185,515	lacks part of 1st Exon corresponding to 5'UTR	fusion transcript
1	LV	H02	151	-	-	-	~analogue 1_D05	LTR transcript
1	LV	H03	144	-	-	-	~analogue 1_D05	LTR transcript
1	LV	H04	-	-	ZNF426; 1st+ 2nd Exon(partial)	chr19:9,646,926-9,649,290	read length insufficient; aligns with 1_C06	-
1	LV	H05	80	SA4	EEF1G; Exon 1-5	chr11:62,338,439-62,341,332	analogue 1_F01	fusion transcript
1	LV	H06	77	SA4	CPSF1; 1st+2nd Exon	chr8:145,634,399-145,634,764	extended 5'RACE	fusion transcript
2	LV	A05	-	-	-	-	poor sequencing results	-
2	LV	A06	-	-	SEC16A; 1st Exon	chr9:139,377,390-139,377,497	LV to R but inbetween no sensible sequence; no proper primer binding site	-
2	LV	B05	-	-	-	-	poor sequencing results	-
2	LV	B06	77	SA3	COBRA1; Exon1-4	chr9:140,149,980-140,151,504	analogue 1_B03	fusion transcript
2	LV	C05	109	B	STAT5B;3rd Exon (partial)	chr17:40,379,547-40,379,646	analogue 2_D06	fusion transcript
2	LV	C06	76	SA3	SIRPB1; 1st Exon + novel alternative(?) Exon in first annotated intron	chr20:1,599,788-1,600,655	analogue 2_H06	fusion transcript
2	LV	D05	80	SA4	EEF1G; Exon 1-5	chr11:62,338,438-62,341,333	analogue 1_F01	fusion transcript
2	LV	D06	109	B	STAT5B;3rd Exon (partial)	chr17:40,379,547-40,379,646	analogue 1_A02	fusion transcript
2	LV	E05	-	-	EEF1G; partial Exon 2	-	compare 1_F01	-
2	LV	E06	18	-	RBM6; Exon 1+2	chr3:49,977,605-50,000,119	Exon2+DWcrypSA2(3' 8 nt 87% complementary)	unspecific

(Table S4 continued)

2	LV	F05	-	-	CHKA; Exon 6-11	chr11:67,829,474-67,837,722	read length insufficient; aligns with 1_C05	-
2	LV	F06	73	SA3	EEF1A1; 1st Exon	chr6:74,230,723-74,230,754	73 bp between SA2 and Exon1 not mappable; analogue to 1_G06	-
2	LV	G05	68	novel SA in PBS	ZNF251; Exon 2 + 3 + alternative exon (?)	chr8:145,975,509-145,979,714	putative novel Exon intron 3	putative fusion transcript
2	LV	G06	80	SA4	FAM126b	chr2:201,936,178-201,936,349	analogue 1_A04	fusion transcript
2	LV	H05	172	-	-	-	incomplete 5' RACE/majority of U5	LTR transcript
2	LV	H06	76	SA3	SIRPB1; 1st Exon + novel alternative(?) Exon	chr20:1,599,788-1,600,655	novel alternative(?) Exon in first annotated intron; compare 1_C03: similar but SA4	putative fusion transcript
2	alpha	A01	-	-	-	-	TOPO religated	-
2	alpha	A02	-	-	-	-	NUP primer only	unspecific
2	alpha	B01	-	-	-	-	no sequence retrieved	-
2	alpha	B02	-	-	-	-	aDWSAF2 primer only	unspecific
2	alpha	C01	-	-	-	-	Universal Primer long + SMARTer residues only	unspecific
2	alpha	C02	-	-	-	-	no sequence retrieved	-
2	alpha	D01	-	-	-	-	no sequence retrieved	-
2	alpha	D02	-	-	-	-	no sequence retrieved	-
2	alpha	E01	-	-	-	-	no sequence retrieved	-
2	alpha	E02	-	-	-	-	NUP primer; no aDWSAF2 sequence	unspecific
2	alpha	F01	-	-	-	-	no sequence retrieved	-
2	alpha	F02	-	-	-	-	no sequence retrieved	-
2	alpha	G01	-	-	-	-	no sequence retrieved	-
2	alpha	H01	-	-	-	-	Universal Primer long + SMARTer residues only	unspecific