

**Supplemental information**

**Nonconditioned ADA-SCID gene therapy reveals ADA requirement in the hematopoietic system and clonal dominance of vector-marked clones**

**Toru Uchiyama, Sirirat Takahashi, Kazuhiko Nakabayashi, Kohji Okamura, Kaori Edasawa, Masafumi Yamada, Nobuyuki Watanabe, Emi Mochizuki, Toru Yasuda, Akane Miura, Motohiro Kato, Daisuke Tomizawa, Makoto Otsu, Tadashi Ariga, and Masafumi Onodera**









5	69938092	69938284	5_69938138	0	1	1	intergenic	SMA5(dist=56590),SERF1A(dist=258332)	5.89275E-05	0.005892752
5	70117564	70117756	5_70117610	0	1	1	intergenic	SMA5(dist=236062),SERF1A(dist=78860)	5.89275E-05	0.005892752
5	72248598	72248790	5_72248746	1	0	1	intergenic	TNPO1(dist=38532),FCHO2(dist=3042)	5.89275E-05	0.005892752
5	99026574	99026766	5_99026722	1	0	1	intergenic	100289230(dist=760010),LOC100133050(dist=688)	5.89275E-05	0.005892752
5	118742871	118743063	5_118742917	0	1	1	intergenic	TNFaip8(dist=12624),HSD17B4(dist=45201)	5.89275E-05	0.005892752
6	32937137	32937329	6_32937285	1	0	1	intronic	BRD2	5.89275E-05	0.005892752
6	47890139	47890328	6_47890284	1	0	1	intronic	PTCHD4	5.89275E-05	0.005892752
6	81948198	81948388	6_81948244	0	1	1	intergenic	BCKDHB(dist=892258),FAM46A(dist=507183)	5.89275E-05	0.005892752
6	102249984	102250176	6_102250132	1	0	1	intronic	GRIK2	5.89275E-05	0.005892752
6	113537263	113537455	6_113537411	1	0	1	intergenic	RFPL4B(dist=864914),MARCKS(dist=641096)	5.89275E-05	0.005892752
6	151541947	151542139	6_151542095	1	0	1	intergenic	MTHFD1L(dist=119073),AKAP12(dist=19019)	5.89275E-05	0.005892752
7	2545985	2546177	7_2546133	1	0	1	intergenic	CHST12(dist=71918),LFNG(dist=6010)	5.89275E-05	0.005892752
7	36507941	36508133	7_36508089	1	0	1	intergenic	ANLN(dist=14690),AOAH(dist=44440)	5.89275E-05	0.005892752
7	101122703	101122892	7_101122749	0	1	1	intronic	COL26A1	5.89275E-05	0.005892752
8	121781	121973	8_121827	0	1	1	intergenic	OR4F21(dist=4804),RPL23AP53(dist=36498)	5.89275E-05	0.005892752
8	92563902	92564094	8_92563948	0	1	1	intergenic	SLC26A7(dist=153571),RUNX1T1(dist=403227)	5.89275E-05	0.005892752
9	27695643	27695835	9_27695689	0	1	1	intergenic	C9orf72(dist=121826),LINGO2(dist=252375)	5.89275E-05	0.005892752
9	42413976	42414168	9_42414124	1	0	1	intergenic	ANKRD20A3(dist=2262),FAM95B1(dist=54445)	5.89275E-05	0.005892752
10	10736746	10736938	10_10736792	0	1	1	intergenic	NONE(dist=NONE),SFTA1P(dist=89590)	5.89275E-05	0.005892752
10	12136491	12136680	10_12136537	0	1	1	intronic	DHTKD1	5.89275E-05	0.005892752
10	13473850	13474042	10_13473896	0	1	1	intergenic	SEPHS1(dist=83599),BEND7(dist=6568)	5.89275E-05	0.005892752
10	104359065	104359257	10_104359111	0	1	1	intronic	SUFU	5.89275E-05	0.005892752
10	106330497	106330689	10_106330645	1	0	1	intergenic	CCDC147(dist=115798),SORCS3(dist=70194)	5.89275E-05	0.005892752
11	49748902	49749094	11_49749050	1	0	1	ncRNA_intronic	LOC440040	5.89275E-05	0.005892752
11	129983153	129983345	11_129983199	0	1	1	intronic	APLP2	5.89275E-05	0.005892752
12	16905572	16905764	12_16905720	1	0	1	intergenic	LMO3(dist=144573),SKP1P2(dist=235941)	5.89275E-05	0.005892752
12	44502720	44502912	12_44502868	1	0	1	intronic	TMEM117	5.89275E-05	0.005892752
12	67134253	67134445	12_67134299	0	1	1	intergenic	GRIPI(dist=61375),CAND1(dist=528742)	5.89275E-05	0.005892752
12	120337763	120337913	12_120337869	1	0	1	intergenic	CIT(dist=22775),CCDC64(dist=89759)	5.89275E-05	0.005892752
13	19855775	19855911	13_19855867	1	0	1	ncRNA_intronic	ANKRD26P3	5.89275E-05	0.005892752
14	41655205	41655397	14_41655353	1	0	1	intergenic	FBXO33(dist=1753650),LRFN5(dist=421391)	5.89275E-05	0.005892752
14	55553266	55553458	14_55553414	1	0	1	intergenic	MAPK1IP1L(dist=16503),LGALS3(dist=42501)	5.89275E-05	0.005892752
14	103306548	103306740	14_103306696	1	0	1	intronic	TRAF3	5.89275E-05	0.005892752
15	70798937	70799129	15_70798983	0	1	1	intergenic	TLE3(dist=408728),UACA(dist=147890)	5.89275E-05	0.005892752
18	25822469	25822661	18_25822617	1	0	1	intergenic	CDH2(dist=65173),MIR302F(dist=2056239)	5.89275E-05	0.005892752
18	28393063	28393255	18_28393109	0	1	1	intergenic	MIR302F(dist=514184),DSC3(dist=176923)	5.89275E-05	0.005892752
18	30979192	30979384	18_30979238	0	1	1	intronic	CCDC178	5.89275E-05	0.005892752
19	5124198	5124390	19_5124244	0	1	1	intronic	KDM4B	5.89275E-05	0.005892752
21	21397016	21397208	21_21397062	0	1	1	intergenic	MPRSS15(dist=1621093),LINC00320(dist=717831)	5.89275E-05	0.005892752
22	50228568	50228708	22_50228614	0	1	1	intergenic	BRD1(dist=10163),ZBED4(dist=18863)	5.89275E-05	0.005892752
X	19565380	19565572	X_19565528	1	0	1	intronic	SH3KBP1	5.89275E-05	0.005892752
X	94547832	94548024	X_94547878	0	1	1	intergenic	AM133A(dist=1580606),LOC643486(dist=104418)	5.89275E-05	0.005892752
X	132959058	132959222	X_132959178	1	0	1	intronic	GPC3	5.89275E-05	0.005892752

Table S2. Vector type of each integration in Pt2

Gene	Vector type
<i>SMARCC1</i>	GCsapM-ADA
<i>NAGA</i>	LASN
<i>LOC100129316</i>	LASN
<i>RGS9</i>	LASN
<i>DPP4</i>	LASN

Table S3

GO Term

Immune system

GO:0051414, response to cortisol(Pathway)  
GO:0002924, negative regulation of humoral immune response mediated by circulating immunoglobulin(Pathway)  
GO:2000318, positive regulation of T-helper 17 type immune response(Pathway)  
GO:0002829, negative regulation of type 2 immune response(Pathway)  
GO:0002827, positive regulation of T-helper 1 type immune response(Pathway)  
GO:0002890, negative regulation of immunoglobulin mediated immune response(Pathway)  
GO:0002921, negative regulation of humoral immune response(Pathway) G  
GO:0090713, immunological memory process(Pathway)  
GO:0002923, regulation of humoral immune response mediated by circulating immunoglobulin(Pathway)  
GO:0006957, complement activation, alternative pathway(Pathway)  
GO:0051412, response to corticosterone(Pathway)  
GO:2000316, regulation of T-helper 17 type immune response(Pathway)  
GO:0002433, immune response-regulating cell surface receptor signaling pathway involved in phagocytosis(Pathway)  
GO:0002825, regulation of T-helper 1 type immune response(Pathway)  
GO:0002313, mature B cell differentiation involved in immune response(Pathway)  
GO:0002507, tolerance induction(Pathway)  
GO:0002230, positive regulation of defense response to virus by host(Pathway)  
GO:0002828, regulation of type 2 immune response(Pathway)  
GO:0002431, Fc receptor mediated stimulatory signaling pathway(Pathway)  
GO:0048536, spleen development(Pathway)  
GO:0042092, type 2 immune response(Pathway)  
GO:0072538, T-helper 17 type immune response(Pathway)  
GO:0042088, T-helper 1 type immune response(Pathway)  
GO:0002920, regulation of humoral immune response(Pathway)  
GO:0097720, calcineurin-mediated signaling(Pathway)  
GO:0002823, negative regulation of adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains(Pathway)  
GO:0002820, negative regulation of adaptive immune response(Pathway)  
GO:0002889, regulation of immunoglobulin mediated immune response(Pathway)  
GO:0071384, cellular response to corticosteroid stimulus(Pathway)  
GO:0002294, CD4-positive, alpha-beta T cell differentiation involved in immune response(Pathway)  
GO:0002287, alpha-beta T cell activation involved in immune response(Pathway)  
GO:0002293, alpha-beta T cell differentiation involved in immune response(Pathway)  
GO:0002292, T cell differentiation involved in immune response(Pathway)  
GO:0002312, B cell activation involved in immune response(Pathway)  
GO:0001776, leukocyte homeostasis(Pathway)  
GO:0032602, chemokine production(Pathway)  
GO:0002824, positive regulation of adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains(Pathway)  
GO:0006958, complement activation, classical pathway(Pathway)  
GO:0002698, negative regulation of immune effector process(Pathway)  
GO:0002456, T cell mediated immunity(Pathway)  
GO:0002821, positive regulation of adaptive immune response(Pathway)  
GO:0002286, T cell activation involved in immune response(Pathway)  
GO:0002455, humoral immune response mediated by circulating immunoglobulin(Pathway)  
GO:0002224, toll-like receptor signaling pathway(Pathway)  
GO:0006911, phagocytosis, engulfment(Pathway)  
GO:0045089, positive regulation of innate immune response(Pathway)  
GO:0006956, complement activation(Pathway)  
GO:0031960, response to corticosteroid(Pathway)  
GO:0002822, regulation of adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains(Pathway)  
GO:0002819, regulation of adaptive immune response(Pathway)  
GO:0002285, lymphocyte activation involved in immune response(Pathway)  
GO:0050777, negative regulation of immune response(Pathway)  
GO:0016064, immunoglobulin mediated immune response(Pathway)  
GO:0019724, B cell mediated immunity(Pathway)  
GO:0045088, regulation of innate immune response(Pathway)  
GO:0002699, positive regulation of immune effector process(Pathway)  
GO:0002366, leukocyte activation involved in immune response(Pathway)  
GO:0002263, cell activation involved in immune response(Pathway)  
GO:0002429, immune response-activating cell surface receptor signaling pathway(Pathway)  
GO:0002757, immune response-activating signal transduction(Pathway)  
GO:0002768, immune response-regulating cell surface receptor signaling pathway(Pathway)  
GO:0006959, humoral immune response(Pathway)  
GO:0002697, regulation of immune effector process(Pathway)  
GO:0002449, lymphocyte mediated immunity(Pathway)  
GO:0002460, adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains(Pathway)  
GO:0002253, activation of immune response(Pathway)  
GO:0002683, negative regulation of immune system process(Pathway)  
GO:0002443, leukocyte mediated immunity(Pathway)  
GO:0002764, immune response-regulating signaling pathway(Pathway)  
GO:0050778, positive regulation of immune response(Pathway)  
GO:0002252, immune effector process(Pathway)  
GO:0002250, adaptive immune response(Pathway)  
GO:0045087, innate immune response(Pathway)  
GO:0050776, regulation of immune response(Pathway)  
GO:0002684, positive regulation of immune system process(Pathway)  
GO:0002520, immune system development(Pathway)  
GO:0002682, regulation of immune system process(Pathway)  
GO:0006955, immune response(Pathway)  
GO:0002376, immune system process(Pathway)



**Hematopoietic system**

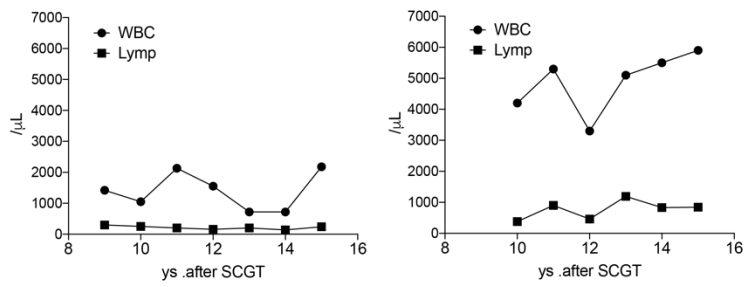
GO:1902035, positive regulation of hematopoietic stem cell proliferation(Pathway)  
GO:1902033, regulation of hematopoietic stem cell proliferation(Pathway)  
GO:0071425, hematopoietic stem cell proliferation(Pathway)  
GO:0060218, hematopoietic stem cell differentiation(Pathway)  
GO:1901532, regulation of hematopoietic progenitor cell differentiation(Pathway)  
GO:0002244, hematopoietic progenitor cell differentiation(Pathway)  
GO:0048534, hematopoietic or lymphoid organ development(Pathway)  
GO:0002901, mature B cell apoptotic process(Pathway)  
GO:0003159, morphogenesis of an endothelium(Pathway)  
GO:0061154, endothelial tube morphogenesis(Pathway)  
GO:0045446, endothelial cell differentiation(Pathway)  
GO:0003158, endothelium development(Pathway)  
GO:0030097, hemopoiesis(Pathway)

## Supplemental figure

Figure S1

A

Pt1



B

Pt2

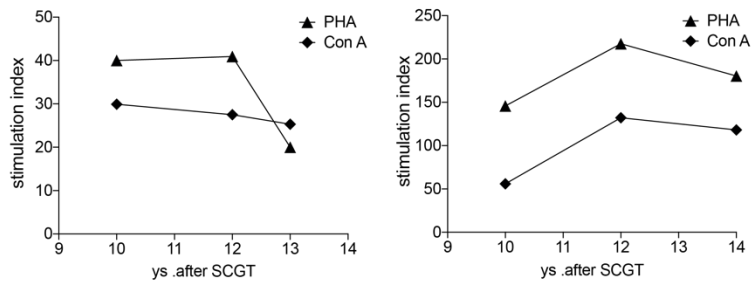


Figure S1. Levels of immune reconstitutions more than 10 years after SCGT in both patients. (A and B) Lymphocyte counts and the responses to mitogens were shown for both patients.

Figure S2

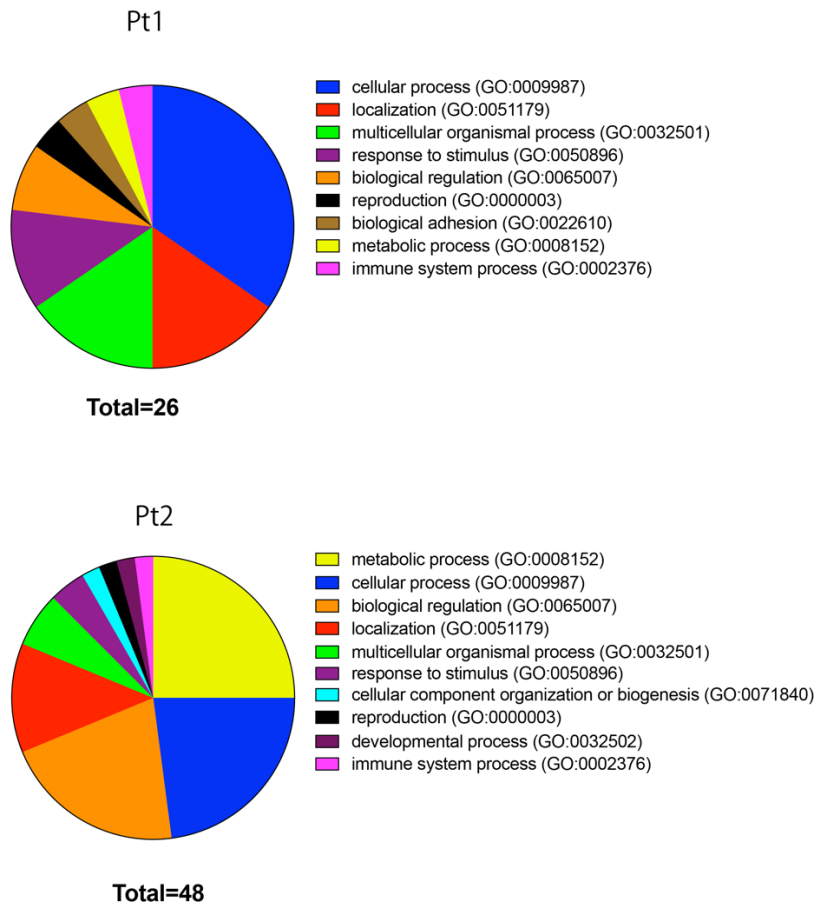


Figure S2. Biological categories of the genes proximal to the vector integration sites.

In gene ontology enrichment analysis, genes near ISs with total read numbers of more than 1% of all integrations were not statistically enriched in specific categories in both patients.

Figure S3

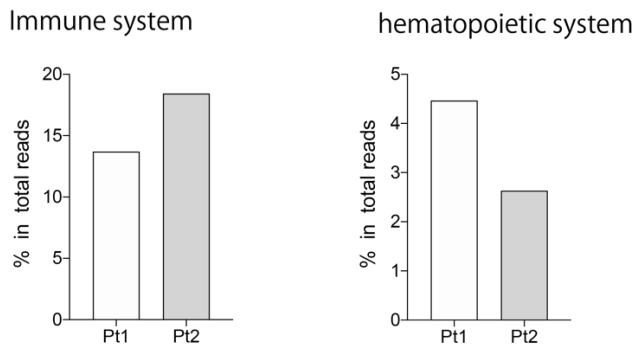


Figure S3. Function of genes hit by retroviral vector in the patients.

The functions of the genes with vector integration were analyzed. Frequencies of the genes that were involved in immune and hematopoietic system were shown.

Figure S4

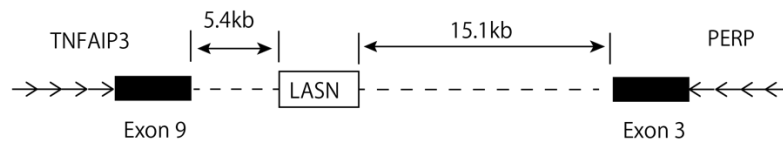


Figure S4. Structure of LASN integration between *TNFAIP3* and *PERP* in Pt2. LASN provirus located 5.4 kb downstream of *TNFAIP3* and 15.1 kb downstream of *PERP*, both of which have been reported as cancer genes.

Figure S5

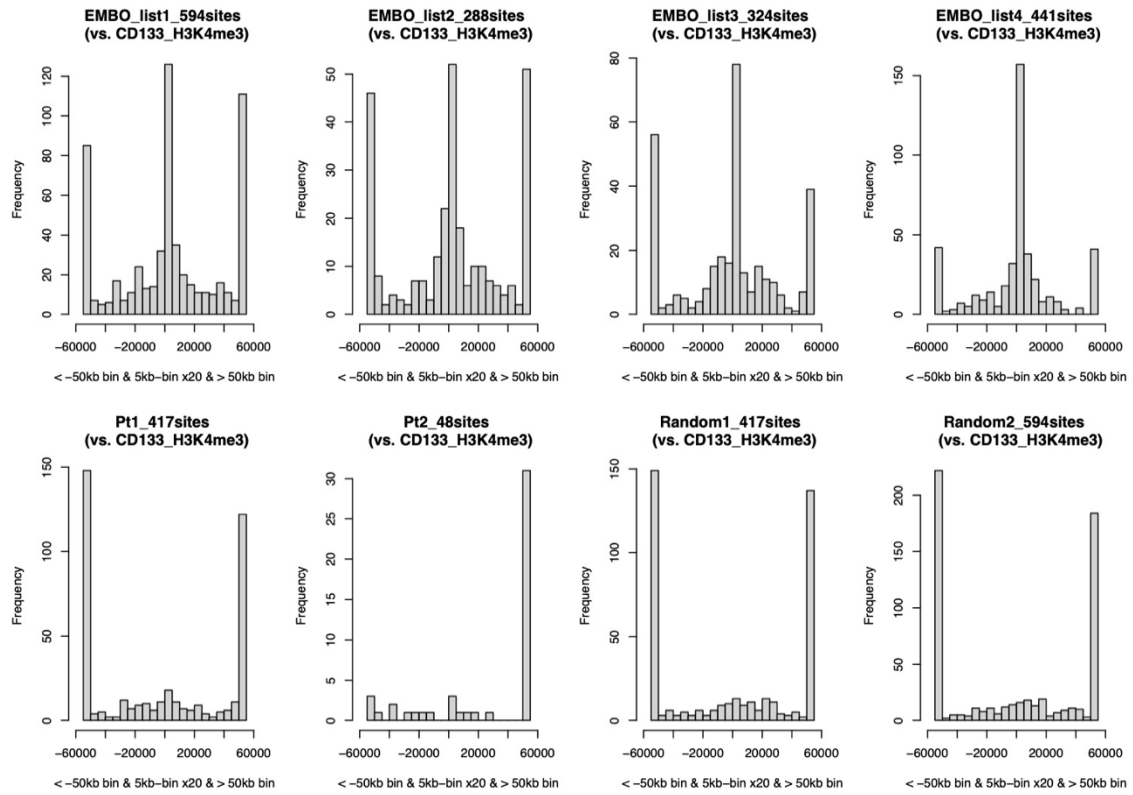


Figure S5. Distribution of the distance from the active transcriptional start site (TSS) in each integration. Each integration is analyzed for the distance from the sites of H3K4me3 modification corresponding to TSS of the genes. The data of the integration in the previous report (Biasco, et al<sup>14</sup>) and randomly selected region as controls were also shown.

Figure S6

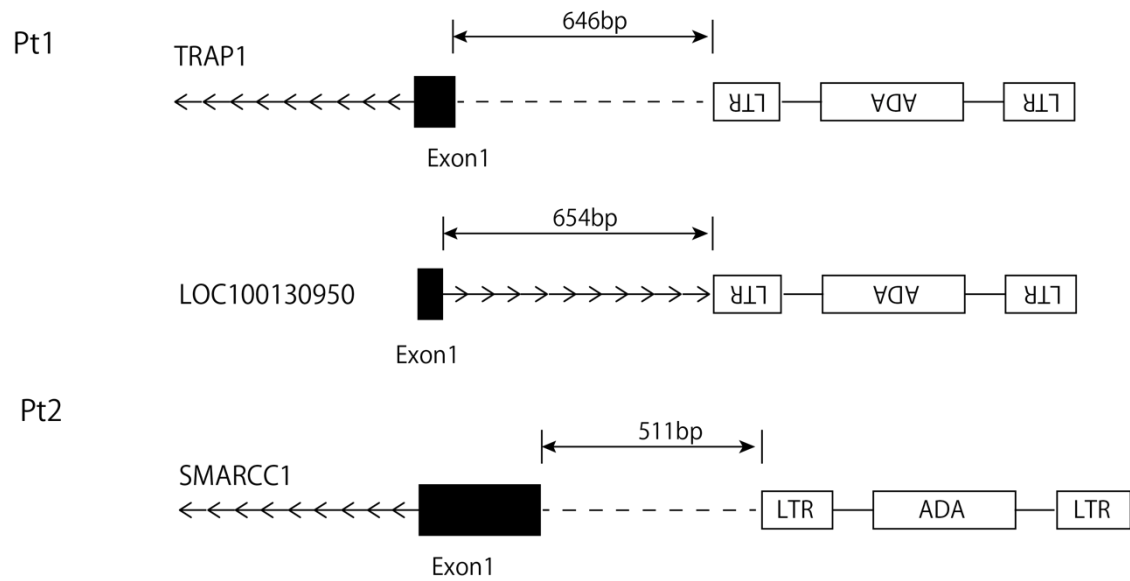


Figure S6. GCsapM-ADA provirus in the dominant clones in the patients. Integration of GCsapM-ADA was detected in active transcriptional start site (TSS) of *TRAP1* and *LOC10012950* in Pt1, and *SMARCC1* in Pt2.

# Supplemental file 1

## LTR sequences of GCsapM-ADA and LASN

MSPV-LTR  
sequence  
as a part of GCsapADA sequence

```
>MSPV-LTR_444bp
AATGAAAGACCCACCTGTAGGTTTGGCAAGCTAGCTTAAGTAACGCCATTTGCAAGGCATGGAAAAATACATAACTGAGAATAGAGAAGTTCAGATCAAGGTCAGG
AACAGAGAAACAGGAGAATATGGGCCAAACAGGATATCTGTGGTAAGCAGTTCCTGCCCGCTCAGGGCCAAGAACAGTTGGAACAGGAGAATATGGGCCAAACAG
GATATCTGTGGTAAGCAGTTCCTGCCCGCTCAGGGCCAAGAACAGATGGTCCCAGATCGGTCCCGCCCTCAGCAGTTTCTAGAGAACCATCAGATGTTTCCAGG
GTGCCCAAGGACCTGAAATGACCCTGTGCCTTATTTGAACTAACCAATCAGTTCGCTTCTCGCTTCTGTTTCGCGCGCTTCTGCTCCCGAGCTCAATAAAAGAGCC
CACAAACCCTCACTCGGC
```

MoMLV-LTR  
sequence  
as a part of the LASN sequence

```
>LASN-LTR (=MoMLV-_589bp)
TTTGAAGACCCACCCGTAGGTGGCAAGCTAGCTTAAGTAACGCCACTTTGCAAGGCATGGAAAAATACATAACTGAGAATAGGAAAGTTCAGATCAAGGTCAGGA
ACAAAGAAACAGCTGAATACCAAACAGGATATCTGTGGTAAGCAGTTCCTGCCCGCTCAGGGCCAAGAACAGATGAGACAGCTGAGTATGGGCCAAACAGGAT
ATCGTGTGTAAGCAGTTCCTGCCCGCTCGGGCCAAGAACAGATGGTCCCAGATCGGTCCAGCCCTCAGCAGTTTCTAGTGAATCATCAGATGTTTCCAGGG
TGCCCAAGGACCTGAAATGACCCTGTACCTTATTTGAACTAACCAATCAGTTCGCTTCTCGCTTCTGTTTCGCGCGCTTCCGCTCTCCGAGCTCAATAAAAGAGCCC
ACAACCCTCACTCGGCGCGCCAGTCTCCGATAGACTGCGTCGCCCGGTACCCGATTCCAATAAAGCCTCTTGCTGTTTGCATCCGAATCGTGGTCTCGCTG
TTCCTTGGGAGGGTCTCCTCTGAGTGATTGACTACCCACGACGGGGTCTTTCATT
```

Custom SureSelect LTR baits for three LTR sequences (MoMLV-LTR, MPSV-LTR)

The baits also contains baits for SHH exons.

Design ID: 0812211

TargetID	ProbelD	Sequence	Replication
MoMLV-LTR	MoMLV-LTR_1-120	TTTGAAGACCCACCCGTAGGTGGCAAGCTAGCTTAAGTAACGCCACTTTGCAAGGCATGGAAAAATACATAACTGAGAATAGGAAAGTTGAGATCAAGGTGAGAACAAAGAAACAGC	1
MoMLV-LTR	MoMLV-LTR_31-150	TAGCTTAAGTAACGCCACTTTGCAAGGCATGGAAAAATACATAACTGAGAATAGGAAAGTTGAGATCAAGGTGAGAACAAAGAAACAGCTGAATACCAAACAGGATATCTGTGGTAAGC	1
MoMLV-LTR	MoMLV-LTR_61-180	GGAAAAATACATAACTGAGAATAGGAAAGTTGAGATCAAGGTGAGAACAAAGAAACAGCTGAATACCAAACAGGATATCTGTGGTAAGCGGTTCTGCCCGGCTCAGGGCCAAGAAC	1
MoMLV-LTR	MoMLV-LTR_91-210	TCAGATCAAGGTGAGAACAAAGAAACAGCTGAATACCAAACAGGATATCTGTGGTAAGCGGTTCTGCCCGGCTCAGGGCCAAGAACAGATGAGACAGCTGAGTGATGGGCCAAACAG	1
MoMLV-LTR	MoMLV-LTR_121-240	TGAATACCAAACAGGATATCTGTGGTAAGCGGTTCTGCCCGGCTCAGGGCCAAGAACAGATGAGACAGCTGAGTGATGGGCCAAACAGGATATCTGTGTAAGCAGTTCTGCCCGG	1
MoMLV-LTR	MoMLV-LTR_151-270	GGTTCTGCCCGGCTCAGGGCCAAGAACAGATGAGACAGCTGAGTGATGGGCCAAACAGGATATCTGTGTAAGCAGTTCTGCCCGGCTCGGGCCAAGAACAGATGGTCCCCAGAT	1
MoMLV-LTR	MoMLV-LTR_181-300	GATGAGACAGCTGAGTGATGGGCCAAACAGGATATCTGTGTAAGCAGTTCTGCCCGGCTCGGGCCAAGAACAGATGGTCCCCAGATGCGGTCCAGCCCTCAGCAGTTTCTAGTGAA	1
MoMLV-LTR	MoMLV-LTR_211-330	GATATCTGTGTAAGCAGTTCTGCCCGGCTCGGGCCAAGAACAGATGGTCCCCAGATGCGGTCCAGCCCTCAGCAGTTTCTAGTGAATCATCAGATGTTTCCAGGGTGCCCAAGGA	1
MoMLV-LTR	MoMLV-LTR_241-360	CTCGGGGCCAAGAACAGATGTTCCAGATGCGGTCCAGCCCTCAGCAGTTTCTAGTGAATCATCAGATGTTTCCAGGGTGCCCAAGGACCTGAAATGACCCCTGTACCTATTTGAAC	1
MoMLV-LTR	MoMLV-LTR_271-390	GCGGTCCAGCCCTCAGCAGTTTCTAGTGAATCATCAGATGTTTCCAGGGTGCCCAAGGACCTGAAATGACCCCTGTACCTATTTGAACAAACCAATCAGTTCGCTTCTCGCTTCTGTT	1
MoMLV-LTR	MoMLV-LTR_301-420	TCATCAGATGTTTCCAGGGTGCCCAAGGACCTGAAATGACCCCTGTACCTATTTGAACAAACCAATCAGTTCGCTTCTCGCTTCTCGCTTCTCGCTTCTCGCTTCTCGCTTCTGTT	1
MoMLV-LTR	MoMLV-LTR_331-450	CCTGAAATGACCCCTGTACCTATTTGAACAAACCAATCAGTTCGCTTCTCGCTTCTCGCTTCTCGCTTCTCGCTTCTCGCTTCTCGCTTCTCGCTTCTCGCTTCTCGCTTCTCGCTTCTG	1
MoMLV-LTR	MoMLV-LTR_361-480	TAACCAATCAGTTCGCTTCTCGCTTCTCGCTTCTCGCTTCTCGCTTCTCGCTTCTCGCTTCTCGCTTCTCGCTTCTCGCTTCTCGCTTCTCGCTTCTCGCTTCTCGCTTCTCGCTTCTG	1
MoMLV-LTR	MoMLV-LTR_391-510	CGGCGCTTCCGCTCCTCGAGCTCAATAAAAGAGCCCAACCCCTCACTCGGCGGCCAGTCTTCCGATAGACTGCGTCCGCCGGGTACCCGATTTCCCAATAAAGCCTCTTGCTGTTT	1
MoMLV-LTR	MoMLV-LTR_421-540	AGAGCCCAACCCCTCACTCGGCGGCCAGTCTTCCGATAGACTGCGTCCGCCGGGTACCCGATTTCCCAATAAAGCCTCTTGCTGTTTGCATCCGAATCGTGGTCTCGCTGTTCCCTG	1
MoMLV-LTR	MoMLV-LTR_451-570	GTCTTCCGATAGACTGCGTCCGCCGGGTACCCGATTTCCCAATAAAGCCTCTTGCTGTTTGCATCCGAATCGTGGTCTCGCTGTTCCCTTGGAGGGTCTCCTCTGAGTGATTGACTACCC	1
MoMLV-LTR	MoMLV-LTR_470-589	CGCCGGGTACCCGATTTCCCAATAAAGCCTCTTGCTGTTTGCATCCGAATCGTGGTCTCGCTGTTCCCTTGGAGGGTCTCCTCTGAGTGATTGACTACCCACGACGGGGTCTTTTCATT	1
MPSV-LTR	MPSV-LTR_1-120	AATGAAAGACCCACCTGTAGGTTGGCAAGCTAGCTTAAGTAACGCCACTTTGCAAGGCATGGAAAAATACATAACTGAGAATAGAGAAGTTGAGATCAAGGTGAGAACAGAGAAACAG	1
MPSV-LTR	MPSV-LTR_31-150	GCTAGCTTAAGTAACGCCACTTTGCAAGGCATGGAAAAATACATAACTGAGAATAGAGAAGTTGAGATCAAGGTGAGAACAGAGAAACAGGAGAATATGGGCCAAACAGGATATCTGTGG	1
MPSV-LTR	MPSV-LTR_61-180	TGGAAAAATACATAACTGAGAATAGAGAAGTTGAGATCAAGGTGAGAACAGAGAAACAGGAGAATATGGGCCAAACAGGATATCTGTGGTAAGCAGTTCTGCCCGCTCAGGGCCAAG	1
MPSV-LTR	MPSV-LTR_91-210	TTGAGATCAAGGTGAGAACAGAGAAACAGGAGAATATGGGCCAAACAGGATATCTGTGGTAAGCAGTTCTGCCCGCTCAGGGCCAAGAACAGTTGGAACAGGAGAATATGGGCCAAA	1
MPSV-LTR	MPSV-LTR_121-240	GAGAATATGGGCCAAACAGGATATCTGTGGTAAGCAGTTCTGCCCGCTCAGGGCCAAGAACAGTTGGAACAGGAGAATATGGGCCAAACAGGATATCTGTGGTAAGCAGTTCTGCC	1
MPSV-LTR	MPSV-LTR_151-270	TAAGCAGTTCTGCCCGCTCAGGGCCAAGAACAGTTGGAACAGGAGAATATGGGCCAAACAGGATATCTGTGGTAAGCAGTTCTGCCCGCTCAGGGCCAAGAACAGATGGTCCCCAG	1
MPSV-LTR	MPSV-LTR_181-300	AACAGTTGGAACAGGAGAATATGGGCCAAACAGGATATCTGTGGTAAGCAGTTCTGCCCGCTCAGGGCCAAGAACAGATGGTCCCCAGATCGGTCCCGCCCTCAGCAGTTTCTAGAGA	1
MPSV-LTR	MPSV-LTR_211-330	CAGGATATCTGTGGTAAGCAGTTCTGCCCGCTCAGGGCCAAGAACAGATGGTCCCCAGATCGGTCCCGCCCTCAGCAGTTTCTAGAGAACCATCAGATGTTTCCAGGGTGCCCAAGG	1
MPSV-LTR	MPSV-LTR_241-360	CGCTCAGGGCCAAGAACAGATGGTCCCCAGATCGGTCCCGCCCTCAGCAGTTTCTAGAGAACCATCAGATGTTTCCAGGGTGCCCAAGGACCTGAAATGACCCCTGTGCCTATTTGAAC	1
MPSV-LTR	MPSV-LTR_271-390	ATCGGTCCCGCCCTCAGCAGTTTCTAGAGAACCATCAGATGTTTCCAGGGTGCCCAAGGACCTGAAATGACCCCTGTGCCTATTTGAACAAACCAATCAGTTCGCTTCTCGCTTCTGTT	1
MPSV-LTR	MPSV-LTR_301-420	ACCATCAGATGTTTCCAGGGTGCCCAAGGACCTGAAATGACCCCTGTGCCTATTTGAACAAACCAATCAGTTCGCTTCTCGCTTCTGTTCCGCGCTTCTGCTCCCGAGCTCAATAAA	1
MPSV-LTR	MPSV-LTR_325-444	CCAAGGACCTGAAATGACCCCTGTGCCTATTTGAACAAACCAATCAGTTCGCTTCTCGCTTCTGTTCCGCGCTTCTGCTCCCGAGCTCAATAAAAGAGCCCAACCCCTCACTCGGC	1



## Baits for SHH

browser position chr7:155595548-155596430

track name="Target Regions" description="Agilent SureSelect DNA - SHH\_151219 - Target regions of interest given as input to probe selection" color=0,128,0

chr7 155595547 155596430 SHH

chr7 155598979 155599261 SHH

chr7 155604506 155604977 SHH

track name="Missed" description="Agilent SureSelect DNA - SHH\_151219 - Target regions of interest that will not be sequenced" color=128,0,0

chr7 155595568 155595608 SHH

chr7 155595728 155595781 SHH

chr7 155604923 155604950 SHH

track name="Covered" description="Agilent SureSelect DNA - SHH\_151219 - Genomic regions expected to be sequenced" color=0,128,0

chr7 155595388 155595568 SHH

chr7 155595608 155595728 SHH

chr7 155595781 155596441 SHH

chr7 155598970 155599270 SHH

chr7 155604503 155604923 SHH

chr7 155604950 155605130 SHH

## Supplemental file 2

### Primers and probes for VCN analysis

Target	primer/probe	primer/probe ID	sequence
packaging signal $\psi$	Fw	Packaging Fw	GCAACCTATCTGTGTCTGTCC
	Rv	Packaging Rv	GGTCCGCCAGATACAGAG
	Probe	Packaging Probe	/56-FAM/TGCGGCCTGC/ZEN/GTCGTACTAGTTAG/3IABkFQ/
GCsapM-ADA	Fw	GCsapM-ADA Fw	TAGACGGCATCGCAGCTTGG
	Rv	GCsapM-ADA Rv	TCCGTCTAGGTGGACATGCAGT
	Probe	GCsapM-ADA Probe	/56-FAM/CGCCCGCCT/ZEN/TCGACAAGCCCAAAG/3YABkFQ/
RPP30	Fw	RPP30 Fw	TCCAGGAGGGAGAATTGATG
	Rv	RPP30 Rv	ATGGTCCGTCTCAGGAAATG
	Probe	RPP30 Probe	/HEX/TCCCTAGGT/ZEN/GGCCTGAGCAG/3IABkFQ/

## Primers and probes for IS-specific ddPCR

### Pt1

Gene name	primer/probe	primer/probe ID	sequence
TRAP1	Fw	ddTP Up Fw1	GAGGCACAGTCTCAAAGGTC
	Rv	ddTP Up Rv1	TGAGTGATTGACTACCCACGA
	Probe	ddTP Up Probe 1	/56-FAM/GCTTGGGTTACAGCTTGCTT/3BHQ
LOC100130950	Fw	ddLOC Down Fw2	CAAACCTACAGGTGGGGTCT
	Rv	ddLOC Down Rv2	GCACTGACAGTTTTGCTTCG
	Probe	ddLOC Down Probe 2	/56-FAM/TAAGATGTCCAACCCCAAGC/3BHQ
ZFAND3	Fw	ddZFAND3 Fw	TGCATCCGAATCGTGGTCTC
	Rv	ddZFAND3 Rv	CCTGGCCAACAGTTGCTTTC
	Probe	ddZFAND3	/56-FAM/CCACGACGG/ZEN/GGGTCTTTCACCTTG/3IABkFQ/
DIP2A	Fw	ddDIP2A/USP4 Fw	TCCATGCCTTGCAAATGGC
	Rv	ddDIP2A Rv	TCATTGAGGCTGGTCCAACC
	Probe	ddDIP2A/USP4	/56-FAM/GCTTGCCAA/ZEN/ACCTACAGGTGGGGT/3IABkFQ/
USP4	Fw	ddDIP2A/USP4 Fw	TCCATGCCTTGCAAATGGC
	Rv	ddUSP4 Rv	TCTGGTCCCTTGAGTCTCCC
	Probe	ddDIP2A/USP4	/56-FAM/GCTTGCCAA/ZEN/ACCTACAGGTGGGGT/3IABkFQ/
LMO2	Fw	ddPCRLMO2Sense4	CTTGCAAATGGCGTTA
	Rv	ddPCRLMO2Anti-sense4	GCTGGAATCGAGACAA
	Probe	Lower_Junc 2 LMO	/56-FAM/TTCAATTCCTTGCTAGAACTCAACA/3BHQ

### Pt2

Gene name	primer/probe	primer/probe ID	sequence
SMARCC1	Fw	ddSMARCFW2	AATCGTGGTCTCGCTGTTCC
	Rv	ddSMARCRV2	TCGAAAGAGCCAGTGCAAGG
	Probe	SMARCC probe-new	/56-FAM/TTCAATTCCTGCCCCAGGG/3BHQ

## Primers/probes for transcription levels of integrated genes

Patient	Gene name	IDT Assay ID
Pt1	TRAP1	Hs.PT.58.19693862
	GPX1	Hs.PT.58.39247474.g
	RAP1B	Hs.PT.58.26098638
Pt2	DPP4	Hs.PT.58.39108231.g
	TNFAIP3	Hs.PT.58.1824217
	PERP	Hs.PT.58.40332203
	MLLT10	Hs.PT.58.1482781
	SMARCC1	Hs.PT.58.3919159
Reference	TBP	Hs.PT.58v.39858774