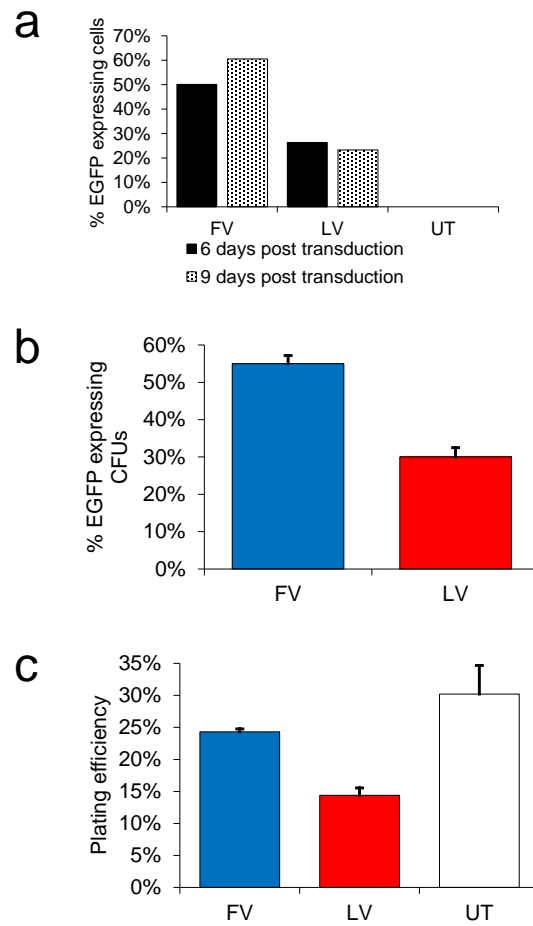
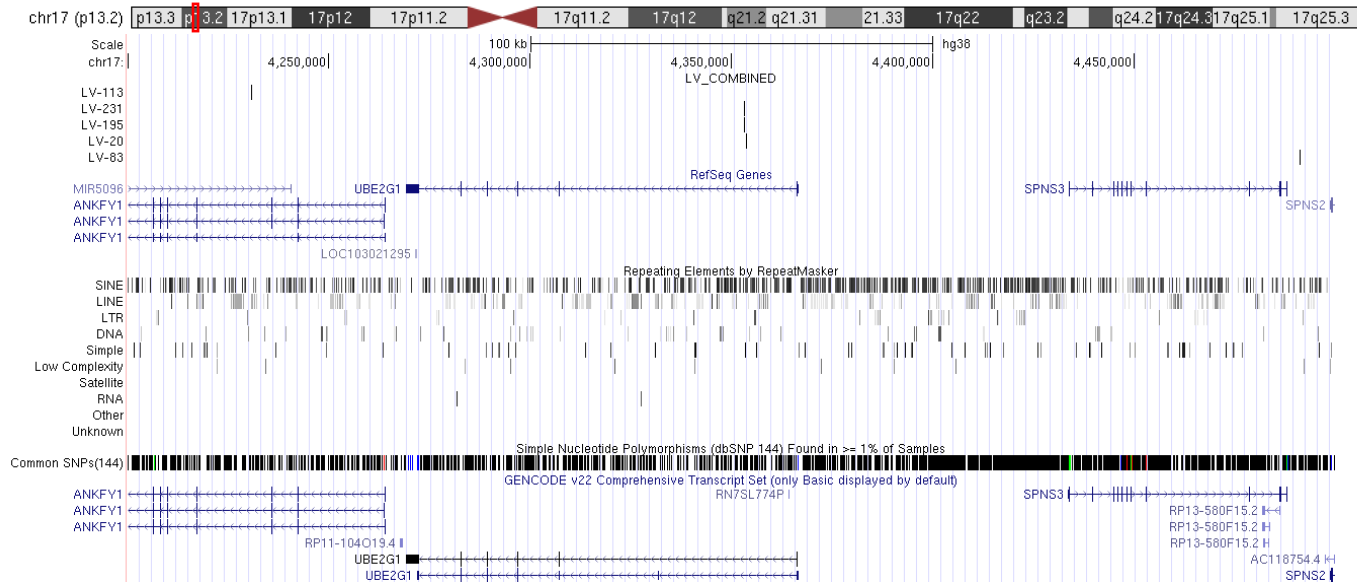


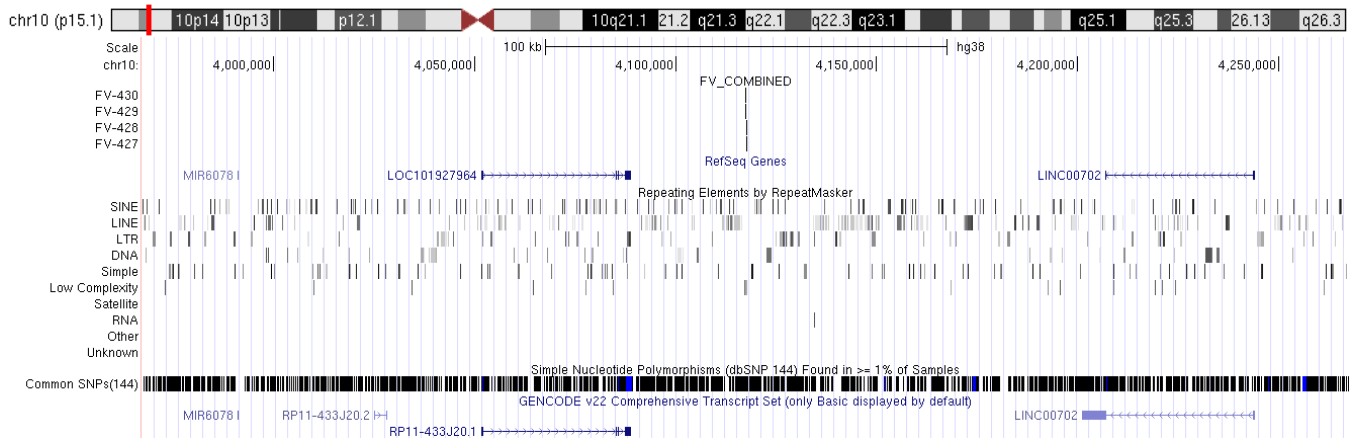
Supplementary Figure S1 In vitro marking analysis of FV and LV vector transduced CD34⁺ cells. **(a)** Marking in CD34⁺ cells maintained in liquid culture six and nine days post-transduction. **(b)** Progenitors expressing EGFP in the CFU assay. **(c)** Plating efficiency from CFU assay.



C.

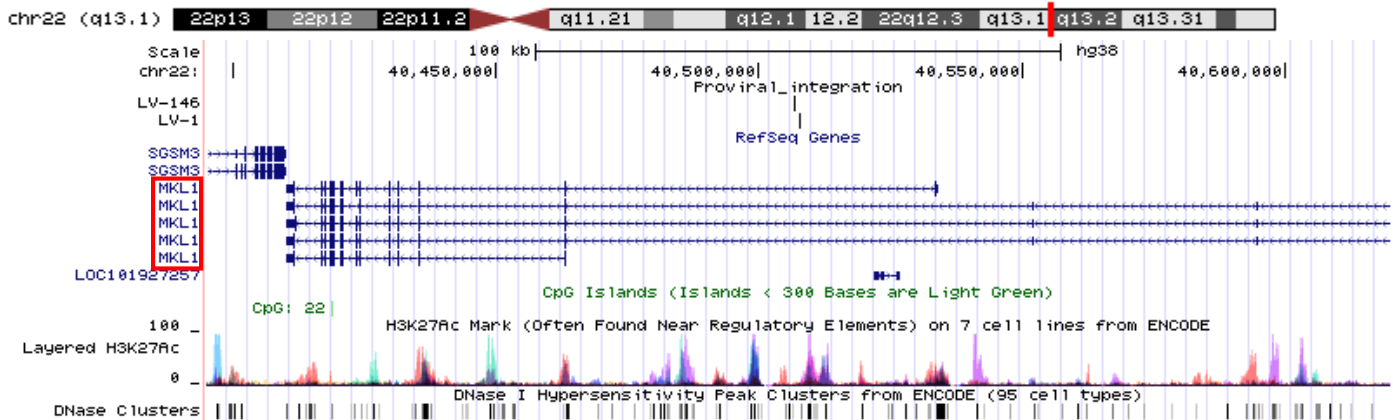


C.

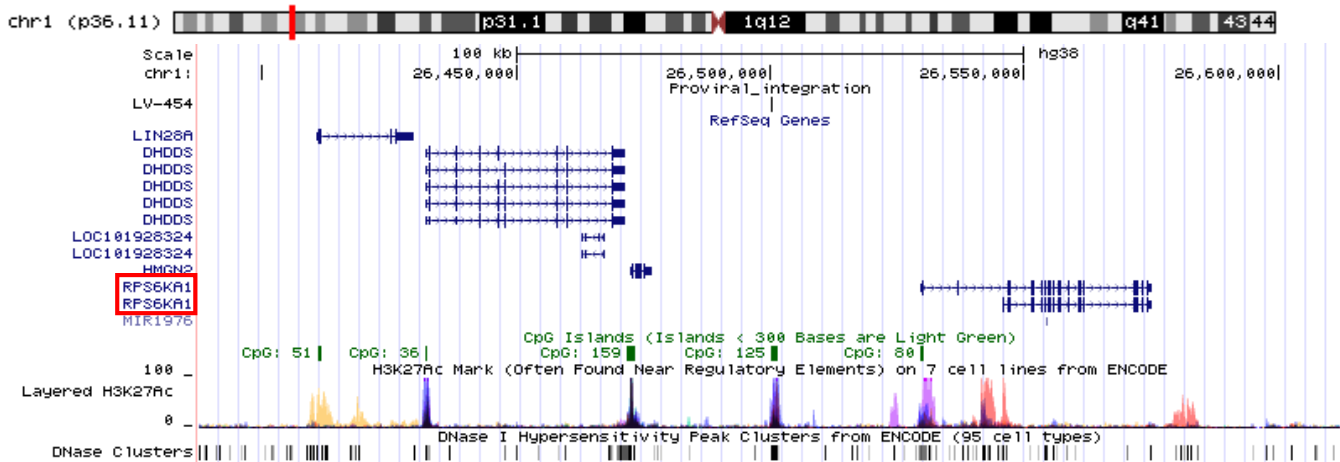


Supplementary Figure S4. Proviral integrations of dominant clones within 100 kb of proto-oncogenes mapped on the human genome using the UCSC genome browser (hg38). **(a)** Proviral integration sites mapped on the human genome. **(b)** The proviral LTR-chromosome junctions and query mapped on the human genome.

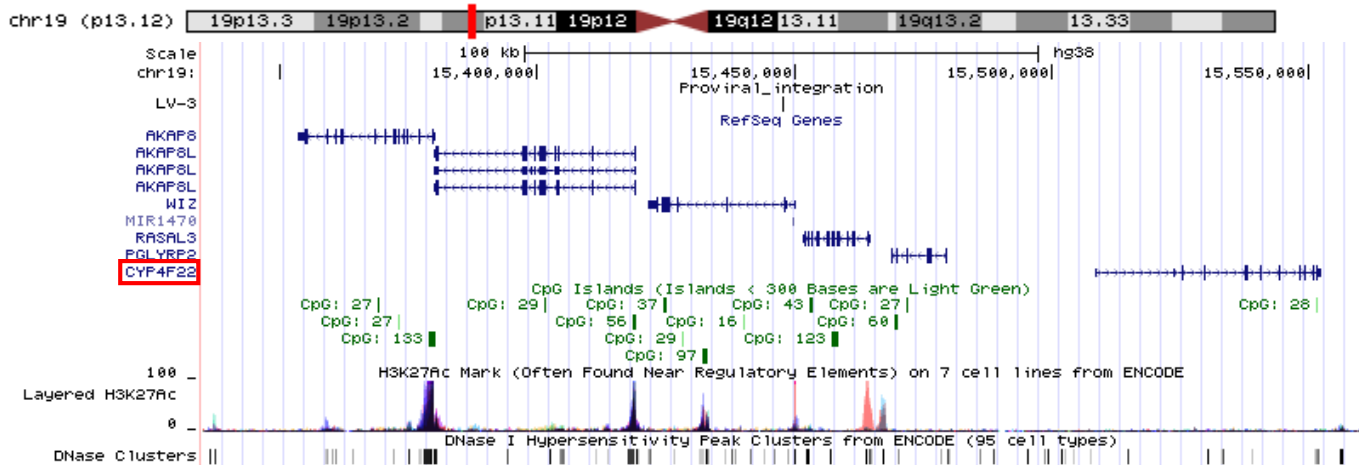
4a1. Proviral integration site in *MKL1* on chromosome 22.



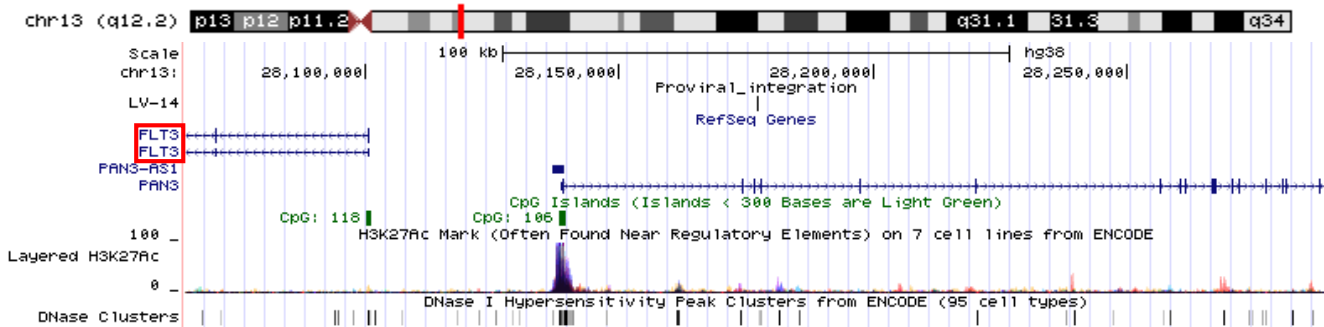
4a2. Proviral integration site near *RPS6KA1* on chromosome 1.



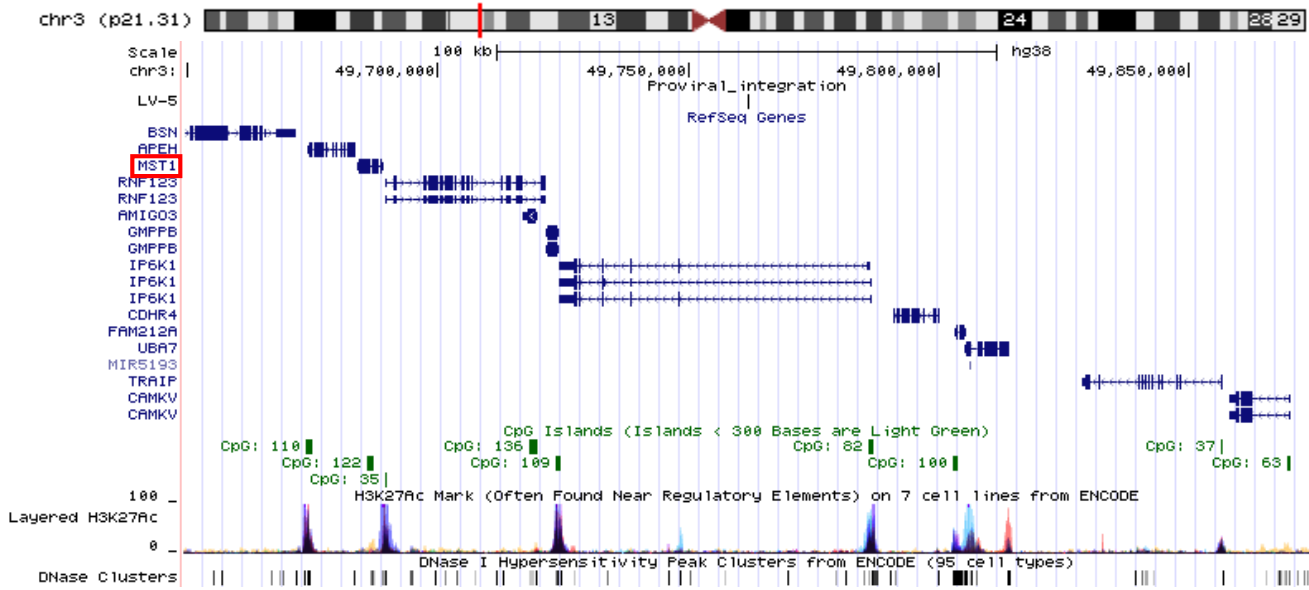
4a3. Proviral integration site near *CYP4F22* on chromosome 19.



4a4. Proviral integration site near *FLT3* on chromosome 13.



4a5. Proviral integration site near *MST1* on chromosome 3.



4b1. MKL1

LTR-Chromosome junction sequence read

AGTGTGGAAAATCTCTAGCAAGGGGGGATCACTTGAGCCTGTCAAGGCTGCAGTGAGCTATGACGGTGCCACTGCACT
CCCACCTGTGTGACAGAGCAAGGCCCTCTCTGAAAGAAAAAAAAAATTGTCCTAGATTAAGAACCCTGGGTTATT
TCAAGGACTAAAGAAGTTGATTTAGGCCAGATGCAGTTGCNNACGCTTGTAAATCCCAGCACCTTAGGCCAAGGTGG
TGGATCACGAGGTCAGGAATTAGAGACCAGCCTGACCAACATGGCGAAACCCCGTCTC

Query sequence localized to chromosome 22 using BLAT

ttcgagacca	gcctgggtaa	catggcaaaa	tctgtctct	aaaaagaata	40507477
caaaaaatta	gaaggggtgcc	tcagtccta	gctattagag	aggctgaggt	40507527
GGGGGGATCA	CTTGAGCCTG	TCAAGGCTGC	AGTGAGCTAT	GACGGTGCCA	40507577
CTGCACTCCC	ACCTGTGTGA	CAGAGCAAGG	CCCTCTCTGA	AAGAAAAAAAA	40507627
AAAATTGTCC	TAGATTAAGA	ACCACTGGGT	TATTTCAAGG	ACTAAAGAAG	40507677
TTGATTTAGG	CCAGATGCAG	TTGTcACGC	TTGTAATCCC	AGCACCTTAG	40507727
GCCAAGGTGG	GTGGATCACG	AGGTCAGGAA	TTAGAGACCA	GCCTGACCAA	40507777
CATGGCGAAA	CCCCGTCTCt	actaaaaata	caaaaattat	ccaggcatgg	40507827
tgacacacac	ctgtagtccc	agctaccagg	aaggctgagg	aatgagaatt	40507877
gcttgaacc	gggaggcag				

4b2. RPS6KA1

LTR-Chromosome junction sequence read

AGTGTGGAAAATCTCTAGCAATAACAAAACAAAAAAGCTGGTTGGTTGTTGGGAGGCCTGGCTGCTTTTGGGAAGG
ATAATCCTCGTCCACACCAAGGTGATTTCAGTGAGCAGGGTGAGGGCAGATGCTGGGTGTCCACGCCCGGCCCCACA
GCAGCCAGAGGAAGATACCAGCCCCGCGCACCCGTCGCCATGGCAACCTCAACCCCACTCCTGGACCCTAGCTCGGAA
GGGGGCAGGGTATCGAAAGC

Query sequence localized to chromosome 1 using BLAT

gctggggctg	cgcgcggggg	cttatgcctg	taaccccagc	actttgggag	26499978
gacagcctgg	ccaacagagc	gagaccctat	ctctatTTTT	TTTTAAAAAA	26500028
ATAACAAAAC	AAAAAAGCT	GGTTGGTTGT	TGGGAGGCCT	GGCTGCTTTT	26500078
GGGAAGGATA	ATCCTCGTCC	ACACCAAGGT	GATTCAGTGA	GCAGGGTGAG	26500128
GGCAGATGCT	GGGTGTCCAC	GCCCCGCCCC	CACAGCAGCC	AGAGGAAGAT	26500178
ACCAGCCCCG	GCACCCGTGC	CCATGGCAAC	CTCAACCCCA	CTCCTGGACC	26500228
CTAGCTCGGA	AGGGGGCAGG	GTATCGAAAG	Ccgatcgcca	gaggctcctc	26500278
ggtgcctcgt	gcgagtcccc	gccgtcagcc	ccgaggagcc	tcttggcacc	26500328
aagcagcgcg	tgcccccttc	ggccacacgg	t		

4b3. CYP4F22

LTR-Chromosome junction sequence read

AGTGTGGAAAATCTCTAGCA**TTGTCCATTTCCCCTACTAGAGTGT**CAGCCCCTTGAGGGCAGGAATTCTTGCCTCTT
TGGTTTACAGCTGTCTTCCCAGTGCCTAGCACATACCCTGGCACACCGAAGGTGCTCAGTATTTAATGAATGAATAA
CACGTCTGGAGGTGGGGAGAGTGTGAGATACCTTTTTGCTGGGTAAACCAAGGGGTAGCCAGCATCGCTGACTACCC
AGCCTCNCCCTAGCCACTACAGAAGCTGTTTTGGCCCGTGGTT**CAGGAACAGGATACACCCAGACAGCCAGAGGGAG**

Query sequence localized to chromosome 19 using BLAT

gagggctctcc	caggtctccc	aggcatgatc	ctggcactct	gtttaatgcc	15447398
ttcacagcat	ctgtcactac	ctgtccttct	gatatttggt	tagttgcttg	15447448
TTGTCCATTT	CCCCTACTAG	AGTGT CAGCC	CCTTGAGGGC	AGGAATTCTT	15447498
GCCTCTTTGG	TTTACAGCTG	TCTTCCCAGT	GCCTAGCACA	TACCCTGGCA	15447548
CACCGAAGGT	GCTCAGTATT	TAATGAATGA	ATAACACGTC	TGGAGGTGGG	15447598
GAGAGTGTGA	GATACCTTTT	TGCTGGGTAA	ACCAAGGGGT	AGCCAGCATC	15447648
GCTGACTACC	CAGCCTctCC	CTAGCCACTA	CAGAAGCTGT	TTTGGCCCGT	15447698
GGTTCAGGAA	CAGGATACAC	CCAGACAGCC	AGAGGGAGct	gggatagact	15447748
gatgggagtg	gggaatacct	gtggaataag	tgggctgggg	gcagggagtc	15447798
tggggtgctg	agatgggcct	ctggacccaa	ttagtatg		

4b4. FLT3

LTR-Chromosome junction sequence read

AGTGTGGAAAATCTCTAGCA**CTTACCTTCATTA**AAATTAACATGAATTTAATATTAATAACTAATAGTGCAAAGAG
CAAGATGTTAATAAAATAGCATGTTGAAATTTTTGTATATAACTTATACACTTTTCTTTTTTTTTTTAATAATAGAG
ACATGGTCTTGTTATATTGNCNAAGTTGGTTTTCAA**ACTCCTGGGCTCAATCATTCTCCACCTGAGCCTCCCAA**

Query sequence localized to chromosome 13 using BLAT

attaaatgcc	cttacgtaaa	ggaaccttaa	atctcttcac	aatgtagtac	28177022
atatagactg	ctgcttcatt	tctctgcctt	cttcttctg	gtttttcaga	28176972
CTTACCTTCA	TTAAAATTAA	CATGAATTTA	ATATTAAATA	ACTAATAGTG	28176922
CAAAGAGCAA	GATGTTAATA	AAATAGCATG	TTGAAATTTT	TGTATATAAC	28176872
TTATACACTT	TTCTTTTTTT	TTTTAATAAT	AGAGACATGG	TCTTGTTATA	28176822
TTGcCcAAGT	TGGTTTCAAA	CTCCTGGGCT	CAATCATTCC	TCCCACCTGA	28176772
GCCTCCCAA	gtgctgggat	tacagggatg	agccaccaca	cctggcctgg	28176722
aatctaaatc	tcaatagtta	gatctgagtg	ggaagataat	aatttttatt	28176672
ttcttcttta					

4b5. MST1

LTR-Chromosome junction sequence read

AGTGTGGAAAATCTCTAGCA**GTTCTCATTAAACA**ACTTAGAAAAAAGGCCAGGTGTGATGGCTTGCACCTGTAAACCC
AACATTTTCAGGAGACCAAGGCAGGAGGATGACTTGAGGCCAGGAGTTCAAGAGCACCCCTGGTCAACAGAGAAAGACT
CTGTCTCTACAGAAAAAATAAAAATTAGGCAGGCATGGTGGTACATGCAGTCCCAGTTTGGAGGCTGAGGCACCGAG
ATTACTTAAGACCAGGAGTTT**GAGGTTACAGTAAGCCATGATTGCACCATTGC**ACTCCAGCCTGGGT

Query sequence localized to chromosome 3 using BLAT

gagcttgcag	tgagccaaga	tcgtgccact	gcactccagc	ctgggcaaca	49761596
gcgcgagact	ccgtctcaaa	aaaaaaaaaa	aaaaaaatta	ataacctaag	49761646
GTTCTCATT	ACA ACTTAGA	AAAA AGGCCA	GGTGTGATGG	CTTGCACCTG	49761696
TAAACCCAAC	ATTT CAGGAG	ACCA AGGCAG	GAGGATGACT	TGAGGCCAGG	49761746
AGTTCAAGAG	CACCCTGGTC	AACAGAGAAA	GACTCTGTCT	CTACAGAAAA	49761796
AATgAAAATT	AGGCAGGCAT	GGTGGTACAT	GCAGTCCCAG	TTTGGAGGCT	49761846
GAGGCACCGA	GATTACTTAA	GACCAGGAGT	TTGAGGTTAC	AGTAAGCCAT	49761896
GATTGCACCA	TTGCACTCCA	GCCTGGGTga	cagagtgaaa	ccctgtctct	49761946
ctctctttca	tttatttttt	ctggtgagat	ggggtctcac	tgttgcccat	49761996
gctgaagtgg	cctggcacga	tcttggt			

Supplementary Table S1. Enriched gene classes near RIS in human SCID repopulating cells (P<0.01)

Vector	Gene class	No. of integrants	% of integrants	Benjamini value
FV	phosphoprotein	169	49.9	1.7×10^{-8}
	alternative splicing	167	49.3	2.1×10^{-3}
LV	phosphoprotein	156	60.7	2.4×10^{-12}
	alternative splicing	152	59.1	9.2×10^{-10}
	splice variant	152	59.1	8.0×10^{-9}
	coiled coil	57	22.2	3.3×10^{-6}
	cytoplasm	74	28.8	1.8×10^{-4}
	acetylation	58	22.6	3.1×10^{-3}
	repressor	18	7.0	3.5×10^{-3}
	chromatin regulator	12	4.7	4.7×10^{-3}

Supplementary Table S2. FV and LV vector clones contributing to more than 20% of all RIS

Vector	Mouse	Chr.	RIS	Contribution to all RIS ^a	Closest proto-oncogene (within 100 kb)	Closest proto-oncogene (within 500 kb)
LV	6	22	40507529	99%	Within <i>MKL1</i>	Within <i>MKL1</i>
LV	8	1	26500029	66%	<i>RPS6KA1</i> (30 kb)	<i>RPS6KA1</i> (30 kb)
LV	1	12	123718848	63%	n/a	n/a
		19	15447449	28%	<i>CYP4F22</i> (61 kb)	<i>CYP4F22</i> (61 kb)
LV	3	X	78694471	63%	n/a	<i>GPR174</i> (465,500 kb)
		13	28176971	33%	<i>FLT3</i> (76 kb)	<i>FLT3</i> (76 kb)
FV	5	X	36964413	39%	n/a	n/a
		22	30608050	27%	n/a	<i>SEC14L3</i> (136,003 kb)
LV	4	9	137093265	34%	n/a	<i>MAMDL4</i> (240,899 kb)
		4	39714912	22%	n/a	<i>RHOH</i> (476,098 kb)
LV	5	3	49761647	30%	<i>MST1</i> (73 kb)	<i>MST1</i> (73 kb)
		9	122215765	25%	n/a	<i>ORIJ2</i> (295,036 kb)
LV	2	2	238004311	23%	n/a	<i>LRRFIP1</i> (376,731 kb)
		5	122826778	22%	n/a	n/a
LV	7	1	207939040	22%	n/a	<i>MIR29B2</i> (136,517 kb)
FV	6	4	106716013	20%	n/a	<i>DKK2</i> (320,283 kb)

Abbreviations: Chr., chromosome, RIS, retroviral integration site, n/a, not applicable

^acapture frequency per mouse