Supplementary Table S4: Vector insertion sites (VIS) distribution according to gene classes (+/- 150 kb), DBISA.

Sets	gamma-retroviral LSK #11-15 and STHSC #31-35	gamma-retroviral LK #16-20 and MPP #36-40	gamma-retroviral IDDb neutral	gamma-retroviral IDDb all	Lentiviral LTHSC #41-45 and STHSC #46-50
N° of VIS	107	63	161	278	43
N° of genes	745	342	918	1607	91
Classes					
RTCGD CIS (%)	8.0	9.9	7.9	8.7	1.1*
RTCGD hits (%)	17.9	16.4	17.6	18.6	5.5*
Tumor suppressor (%)	0.4	0	0.3	0.3	0
Oncogenesis (%)	1.6	2.3	2.3	2.8	0
Signal transduction (%)	13.0	13.4	13.5	14.3	20.9
Unclassified (%)	26.4	25.4	25.6	26.6	19.8

Analysis performed for window +/-150 kb from VIS; Sets, indicate datasets analysed: LSK, STHSC, LTHSC, LK, MPP VIS obtained from primary recipients of LSK, STHSC, LTHSC, LK and MPP EGFP progeny respectively; IDDb neutral: insertional dominance database[24] re-analysis focusing on VIS from primary recipients with fluorescent or surface markers; IDDb all: re-analysis of all entries of the IDDb; Classes, different categories analysed, corresponding to RTCGD genes from http://rtcgd.abcc.ncifcrf.gov or to biological process retrieved from PantherDB www.pantherdb.org. Data files downloaded June 21, 2008. For each dataset we analyzed the indicated amount of sequences (N° of VIS). This yielded the indicated number of genes within a +/-150 kb window (N° of genes). For each class, the % of genes overlapping with the class is given. *, Fisher Exact 2-sided $p<0.05$, compared to the gamma-retroviral LSK and STHSC datasets.