Supplemental Protocol S1 - PCR amplification and sequencing of integration sites.

Integration sites were PCR-amplified using a splinkerette-based PCR approach (Devon et al. 1995; Mikkers et al. 2002). Chromosomal DNA was digested and ligated to annealed adapter oligos (splinkerettes) and the fragment containing the lentivirus was amplified by nested PCR using primers complimentary to the splinkerette and the LTRs and subsequently sequenced.

To achieve this, DNA was isolated from the HEK293 clones (see Methods) and 1 μg of DNA was digested overnight with 4 units of *Nla*III (New England Biolabs) or *Taq*I (Roche) in the presence of BSA (200 μg/μl) and a total volume of 40 μl. Both enzymes cut the lentiviral construct in between the LTRs. *Nla*III digestions were heat-inactivated at 65 °C for 20 min. To create splinkerettes, adapter oligos were annealed in equal molar amounts (see Supplemental Table S4 for DNA oligos used for sequencing of integration sites). For *Taq*I digested DNA, TaqSpAA and TaqSpBB adapter oligos were used (creating a GC overhang), for *Nla*III digested DNA, NlaSpAA and NlaSpBB adapter oligos were used (creating a CATG overhang). For annealing, 10 μl of each oligo (250 pmol/μl in Milli-Q water) plus 2.2 μl of SuRE/Cut buffer M (Roche) was heated to 95 °C for 3 min and gradually cooled to room temperature in a thermal cycler (PTC-100, MJ research Inc.) by 1 °C per 15 s. Annealed oligos were then diluted to 4 pmol/μl by adding 603 μl of Milli-Q water. Digested DNA was heated to 45 °C and directly put on ice. Ligation of splinkerette and digested chromosomal DNA was performed overnight at 15 °C in a total volume of 133.3 μl, using 13.5 units of T4 DNA ligase (Roche), 40 μl of digested DNA (1 μg) and 32 μl of splinkerettes (4 pmol/μl). This is approximately a 20-fold molar excess of splinkerettes. After ligation, ligase was inactivated at 65 °C for 20 min and 466.6 μl of Milli-Q water was added to a total volume 500 μl. Samples were purified on a Microcon YM-100 column (Millipore) and spun down at 14,000g for 12 min. The sample (template) was eluted in with 100 μl of Milli-Q water at 1,000g for one min.

Sequencing and PCR was performed in a thermal cycler (PTC-100, MJ research Inc). PCR reactions with a total volume of 50 μl were performed with 2.5 units of Super Taq (HT Biotechnology), 100 μM dNTP's and 10 pmoles of each primer. 10x PCR buffer consisted of 500 mM KCl, 100 mM Tris-HCl pH 8.0, 0.1 % gelatin and 15, 25 or 30 mM of $MgCl_2$, depending on the primers used. The first PCR contained 10 μl of template (10 ng/μl) and the primers Ta and 5a (for picking up the 5' LTR fragment; 25 mM $MgCl_2$) or Ta and 3a (for picking up the 3' LTR

fragment; 30 mM $MgCl_2$). PCR conditions were 3 min at 94 °C followed by two cycles of 15 s at 94 °C, 30 s at 68 °C, 5 min at 72 °C and 27 cycles of 15 s at 94 °C, 30 s at 66 °C, 5 min at 72 °C followed by 5 min at 72 °C. The second PCR contained 2 μl of product from the first PCR reaction, 15 mM of $MgCl_2$ and the primers Tb and 5b (for picking up the 5' LTR fragment) or Tb and 3b (for picking up the 3' LTR fragment). PCR conditions were 3 min at 94 °C followed by 25 cycles of 15 s at 94 °C, 30 s at 60 °C, 3 min at 72 °C followed by 5 min at 72 °C. For sequencing, the product of the second PCR was purified using a QIAquick PCR Purification Kit (Qiagen). Samples were sequenced with 5 ng of template in 10 μl reactions with BigDye Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems). As a primer, 1.6 pmol of 5s, 3s or Ts oligo was used for sequencing from the 5' LTR, the 3' LTR or splinkerette respectively. Sequencing conditions were 2 min at 96 °C followed by 29 cycles of 10 s at 96 °C, 5 s at 50 °C and 4 min at 60 °C. 10 μl of Milli-Q water was added to reactions, which were analyzed on an ABI 3730 Automated Sequencer (Applied Biosystems). Sequences were base-called with Sequencing Analysis Software (Applied Biosystems).

References

Devon, R.S., Porteous, D.J., and Brookes, A.J. 1995. Splinkerettes Improved Vectorettes for Greater Efficiency in Pcr Walking. *Nucleic Acids Res.* **23**:1644-1645

Mikkers, H., Allen, J., Knipscheer, P., Romeyn, L., Hart, A., Vink, E., and Berns, A. 2002. High-throughput retroviral tagging to identify components of specific signaling pathways in cancer. *Nature Genetics* **32**:153-159