Published in final edited form as:
Nat Protoc. 2009 ; 4(5): 789–798. doi:10.1038/nprot.2009.64.

A high-throughput splinkerette-PCR method for the isolation and sequencing of retroviral insertion sites

Anthony G. Uren[1,†], **Harald Mikkers**[2,†], **Jaap Kool**[1], **Louise van der Weyden**[3], **Anders H. Lund**[5], **Catherine H. Wilson**[3], **Richard Rance**[3], **Jos Jonkers**[4], **Maarten van Lohuizen**[1], **Anton Berns**[1,*], and **David J. Adams**[3,*]

[1]Division of Molecular Genetics, Cancer Genomics Centre, Netherlands Cancer Institute, Plesmanlaan. 1066CX, Amsterdam, The Netherlands. [2]Department of Molecular Cell Biology and Regenerative Medicine Program, Leiden University Medical Centre, 2300RC, Leiden, The Netherlands. [3]Experimental Cancer Genetics, The Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1SA, UK. [4]Division of Molecular Biology, Cancer Genomics Centre, Netherlands Cancer Institute, Plesmanlaan. 1066CX, Amsterdam, The Netherlands. [5]BRIC - Biotech Research & Innovation Centre, University of Copenhagen, Ole Maaløes Vej 5, 2200 Copenhagen N, Denmark.

Abstract

Insertional mutagens such as viruses and transposons are a useful tool for performing forward genetic screens in mice to discover cancer genes. These screens are most effective when performed using hundreds of mice, however until recently a major limitation to performing screens on this scale has been the cost effective isolation and sequencing of insertion sites. Here we present a method for the high-throughput isolation of insertion sites using a highly efficient splinkerette-PCR method coupled with capillary or 454 sequencing. This protocol includes a description of the procedure for DNA isolation, DNA digestion, linker or splinkerette ligation, primary and secondary PCR amplification, and sequencing. This method, which takes about 1 week to perform, has allowed us to isolate hundreds of thousands of insertion sites from mouse tumours and, unlike other methods, has been specifically optimised for the isolation of insertion sites generated with the murine leukaemia virus (MuLV), and can easily be performed in 96 well plate format for the efficient multiplex isolation of insertion sites.

Keywords

Splinkerette-PCR; Insertional Mutagenesis

INTRODUCTION

Insertional mutagenesis in mice

Human tumours arise as a result of the sequential acquisition of mutations, which may include point mutations, changes in DNA copy number, translocations and epigenetic changes in the genome. Not all of these alterations are causally implicated in oncogenesis,

[*]**Addresses for Correspondence** Dr. Anton Berns a.berns@nki.nl Telephone: +31-20-512-1991 Fax: +31-2-512-2011. Dr. David Adams da1@sanger.ac.uk Telephone: +44 (0) 1223-834-244 Fax: +44 (0) 1223-494-919.
[†]These authors contributed equally

CFI: The authors declare that they have no competing financial interests.

Europe PMC Funders Author Manuscripts

and separating the 'driver' mutations from the 'passenger' mutations[1] is not possible without functional validation *in vitro* or preferably *in vivo*. Insertional mutagenesis in the mouse using either viruses or transposons makes it possible to perform genome-wide screens for genes that co-operate in cancer formation. Once integrated in the genome the provirus or transposon acts as a 'tag' allowing its location to be determined and effects on gene function assessed. Therefore insertional mutagenesis studies in the mouse are a useful complement to the analysis of human cancer genomes and facilitate the identification of 'driver' cancer genes.

Historically, insertional mutagenesis screens in mice have been performed using slow transforming retroviruses. These viruses include murine leukaemia viruses (MuLV) and mouse mammary transforming viruses (MMTV), which generate tumours of the lymphoid system and mammary gland, respectively[2]. More recently transposon based methods, such as those that use the *Sleeping Beauty* transposon[3], have been developed to allow mutagenesis to be performed in other tissues including the liver[4], brain[5] and soft tissue compartments[6]. Although conceptually straightforward, isolating and sequencing insertion sites (either viral or transposon-mediated) from hundreds or thousands of tumours can be cumbersome and expensive. Hence, methods that allow insertion sites to be efficiently amplified and sequenced are critical for the success of these screens.

Strategies for isolating insertion sites

To isolate insertion sites it was once necessary to generate genomic libraries of each tumour and to screen these by colony lift for genomic fragments containing a virus-genome junction, usually using viral long terminal repeat (LTR) sequences as probes[7]. More recently a variety of approaches have been developed for insertion site isolation including inverse-PCR[8], vectorette[9] or splinkerette-PCR[10], panhandle-PCR, *Alu*-PCR, capture-PCR and boomerang DNA amplification[11]. Although inverse-PCR has been a widely used approach for isolating insertion sites, it is relatively inefficient compared to other PCR-based approaches[11]. Approaches such as panhandle-PCR, *Alu*-PCR, capture-PCR and boomerang DNA amplification have not been widely employed because they can be problematic to optimise and do not readily lend themselves to scaling-up for high-throughput applications. Vectorette-PCR can be highly sensitive but is prone to the amplification of contaminants by 'end-repair priming'[9]. End-repair priming involves the free cohesive ends of unligated vectorettes and inserts which are based on restriction sites that produce 5′ overhangs. During the first cycle of PCR these ends are 'filled in' and after the subsequent denaturing step, these ends are able to anneal to each other and to initiate priming. Extension across a vectorette sequence at either end of the insert molecule will result in the production of a sequence complementary to the vectorette primer. When this happens exponential PCR amplification may occur without involvement of the specific target DNA primer and the pool of PCR products in the reaction will become swamped by this contaminant, which is usually preferentially amplified. In contrast splinkerette-PCR[10,12], which is a variant of ligation-mediated PCR, was developed to overcome the problem of 'end-repair priming' by using a splinkerette 'hairpin loop' (Fig. 1a,b). Although not perfect, because like any PCR-based method Splinkerette-PCR is potentially prone to PCR amplification bias and contamination, it has become the most widely accepted technique for the amplification of viral and transposon insertion sites.

The principle of splinkerette-PCR

Splinkerette-PCR was originally developed to enable PCR amplification of sequences that lie between a known DNA sequence[10], such as a viral LTR or transposon IR/DR, and a nearby restriction site. Unlike vectorettes, which contain a central DNA mismatch within them, splinkerette adaptors incorporate a hairpin structure (Fig. 1b). Annealing of oligos to

create the splinkerette adaptor creates an overhang compatible with overhangs generated by the restriction of the (genomic) DNA with the restriction enzyme of choice. However, since the oligonucleotides used to create the splinkerette adaptor are not phosphorylated at their 5′ terminus they should only undergo single stranded ligation to the 5′ end of the template sequences. To amplify DNA a primer that can only bind to the newly synthesized strand of the splinkerette adaptor is used in the PCR reaction. This ensures that DNA cannot be amplified until synthesis from a primer that binds to the target sequence, for example the viral LTR, has been achieved. This feature improves specific amplification of splinkerette-adapted sequences. Furthermore, during the elongation step of the PCR reaction the free 3′ end of the splinkerette will 'flip back' on itself to form a stable double stranded DNA hairpin which is unable to participate in end-repair priming. An additional factor that increases the specificity of the splinkerette reaction is that only one-strand is available for non-specific priming, unlike vectorettes where mis-priming can occur from both strands. Some splinkerette protocols have also been adapted to incorporate a C3 spacer 'blocking group' at the 3′ end of the splinkerette oligos which can also help to prevent mis-priming[6,10]. Collectively, these features make splinkerette-PCR efficient, accurate and highly sensitive.

Sequence analysis of splinkerette insertion sites

Tumours generated using either transposons or retroviruses are mostly oligoclonal. In addition, cells within each tumour contain multiple insertion events making the amplification of insertion sites from a tumour a multiplex reaction. For sequence analysis it is necessary to de-multiplex each splinkerette-PCR so that the sequence of individual insertion sites can be retrieved.

There are several approaches that can be used for de-multiplexing. The first is shotgun cloning of individual splinkerette-PCR products followed by capillary read sequencing, and the second is barcode splinkerette-PCR followed by 454 parallel sequencing. The major advantage of shotgun cloning followed by capillary read sequencing is that it is extremely robust. We have developed an approach called TOPO-shotgun (Fig. 2) which we have used to generate the DNA sequence of over 100,000 insertion sites from over 1,000 splinkerette-PCR reactions generated from both retroviral[13] and transposon induced tumours. The other advantage of TOPO-shotgun is that the sequences generated from capillary reads are long, up to 1 Kb, so it is possible to obtain the sequence from the viral LTR or transposon IR/DR, the sequence of the junction between the insertional mutagen and the genome, a length of genomic sequence, and to also read sequence into the splinkerette adaptor. This facilitates the removal of contaminated splinkerette-PCR products and the accurate alignment and orientation of insertion sites on the genome. The major disadvantage of TOPO-shotgun cloning is that it is expensive and requires considerable infrastructure for colony picking, prepping and sequencing. It is however still the method of choice for the analysis of small numbers of tumour samples (<20) where the use of parallel sequencing is unnecessary and uneconomical.

The alternative approach for obtaining the sequence of insertion sites is to employ 454 parallel sequencing of pools of barcoded splinkerette-PCR reactions from multiple independent tumours (Fig. 2). Because 454 employs single molecule sequencing technology, pools of PCR products can be effectively de-multiplexed on the machine. By incorporating 'barcode' sequences (short DNA sequence identifiers) in the splinkerette-PCR primers, it is possible to pool together splinkerette products from multiple tumours and to use the barcode as a tag to determine which sequence belongs to each tumour within the pool. 454 sequencing is a very efficient way of analyzing insertion sites. A single 454 run can be performed in one day, yielding over 400 Mb of sequence[14]. Importantly, it is possible to run up to eight independent samples, or pools of samples, on each flow cell. DNA sequence

reads on the Roche 454 Gs20 machine are short, ~100 bp whereas the sequences generated by the 454 FLX are on average 250 bp in length. For most insertion sites the length of DNA sequence generated on the FLX is sufficient to identify the IR/DR or LTR, the junction between the IR/DR or LTR and the genome and in many cases to generate sequence into the splinkerette adaptor. Although this is the first working protocol describing the use of splinkerette-PCR and 454 for insertion site isolation from MuLV induced tumours[13] similar methods exist for HIV and other viruses[15].

EXPERIMENTAL DESIGN

In this protocol we have focused on the isolation of insertion sites from MuLV-induced mouse lymphomas but with adaptation of the primer sequences this protocol is applicable to the isolation of insertion sites generated by any insertional mutagen. Here we present an optimised protocol for the amplification of splinkerette-PCR products representing the junction between the MuLV viral U3LTR and a genomic insertion point. This protocol uses the restriction enzyme *Sau*3AI, and a compatible splinkerette adaptor, but can be adapted for any restriction enzyme and splinkerette adaptor combination. The technique we describe is applicable to the isolation of insertion sites from transposons or other insertional mutagens with adaptation of the oligonucleotides used at each PCR step and the restriction enzymes used for cutting the genomic DNA. For simplicity this protocol is in single tube format but can be scaled up to 96-well plate format. We have provided information on the equipment and reagents needed for performing this splinkerette method in 96-well plates throughout this protocol.

DNA extraction method

The quality of DNA used for splinkerette-PCR is critical. The DNA must be free of salt or other contaminants that may affect the PCR amplification reactions. We have obtained good results using DNA extracted with Gentra Puregene kits (Qiagen). Although it is possible to perform splinkerette-PCR using DNA extracted by other methods, such as phenol-chloroform, we have found this less successful.

Choice of restriction enzyme

Since splinkerette-PCR is a multiplex-PCR reaction it will generally favour the amplification of shorter PCR products from within the pool of insertion sites found within a sample. The sequence of the genomic DNA between the viral LTR or transposon IR/DR is also likely to influence the recovery of an insertion since GC rich sequences may be harder to amplify. To obtain maximal coverage of insertion sites it is preferable to use multiple restriction enzymes on each tumour DNA and several independent splinkerette reactions. For the isolation of retroviral insertion sites (from MuLV-induced tumours), we have had success using *Sau*3AI, *Tsp509*I, and *Msp*I although best results have been obtained with *Sau*3AI and *Msp*I. For isolation of transposon insertion sites (from *Sleeping Beauty*-induced tumours), we have used *Nla*III and *Bfa*I[6]. These enzymes all have a 4 bp recognition sequence and the average distance between each restriction site in the mouse genome is 414 bp (SD 8,014 bp) for *Sau*3AI, 170 bp (SD 5,142 bp) for *Tsp509*I, 1,665 bp (SD16,258 bp) for *Msp*I, 213 bp (SD 5,748 bp) for *Nla*III and 316 bp (SD 340 bp) for *Bfa*I. It is unlikely that saturation isolation of insertion sites is possible by using splinkerette-PCR, or any other PCR based method. However, it is clear that the absolute recovery of insertion sites increases when additional enzymes are used to analyse each tumour in separate splinkerette-PCR reactions.

Design of the splinkerette adaptor sequences

Although several splinkerette adaptor sequences have been developed the protocols described below use an adaptor sequence first reported by Mikkers et al[16].

Digestion of the adaptor-ligated splinkerette products

When the primary amplification primers are in a sequence that is repeated internally (for example the U3 LTR of MuLV) an appropriate restriction site must be found for digestion of the ligated DNA to prevent amplification of internal fragments. This site must be closer to the internal primer than the enzyme used to ligate the splinkerette to the DNA, i.e. digestion of the genomic DNA. In the case of isolating insertion sites from MuLV-induced tumour genomic DNA, it is necessary to digest the adaptor-ligated splinkerette genomic DNA to prevent amplification of internal MuLV fragments from the internal U3 LTR sequence during the primary amplification PCR step. If *Sau*3AI, *Tsp*509I or *Msp*I are used, then *Eco*RV is recommended for digestion of the adaptor-ligated splinkerette products since the *Eco*RV site is closer to the internal LTR than these sites (Fig. 1a).

Experimental Controls

As with any PCR-based protocol it is important to always run a ‘water only’ negative control at each PCR step to ensure that there is no systematic contamination of reagents. We also recommend including control of DNA that does not contain insertion sites. The reason for this control is to ensure that non-specific amplification of sequence from endogenous ‘psuedo-LTR’ sequencing does not occur. An optimal result from splinkerette-PCR is a ladder of bands, as shown in Fig. 1c. Novice users of this protocol may wish to run a positive control that produces a known banding pattern on the agarose gel has described in the expected results.

MATERIALS

Reagents

! CAUTION - When handling all the reagents wear gloves and suitable protective clothing, avoid contact with skin and eyes and do not eat, drink or smoke. For further information for individual products, search at http://www.sigmaaldrich.com/technical-service-home/product-catalog.html.

Gentra Purgene Cell Kit [Qiagen, 158767]

! CAUTION - Some of the reagents used in this kit are classed as irritants. Gloves, lab coats, and protective eyewear should worn when using this kit.

Propan-2-ol (NORMAPUR AR analytical reagent, iso-propyl alcohol) [VWR, 20842.323]

! CAUTION - Propan-2-ol is flammable and is an irritant. Always store in an appropriate flammables cabinet and handle away from spark sources or flames. Gloves, lab coats, and protective eyewear should be worn when handling Propan-2-ol.

Ethanol (99.7% v/v, AnalaR) [VWR, 101077Y]

*Sau3A*1 (4 U μl^{-1}; supplied with 10× buffer & 100× BSA) [NEB, R0169]

HPLC-purified oligonucleotides (Table 1) [Sigma-Genosys]

*Eco*RV (20 U μl^{-1}; supplied with 10× buffer & 100× BSA) [NEB, R0195S]

NEB buffer 2 [NEB, B7002S]

Bacto-tryptone [Difco, 0123-01-1]

Bacto-yeast extract [Difco, 0127-05-3]

Sodium Chloride, NaCl [Sigma, S9625]

Potassium Chloride, KCl [Sigma, P3911]

Glucose [Sigma, G8270]

Sodium Hydroxide, NaOH [Sigma, S8045]

! CAUTION - NaOH is an irritant. Gloves, lab coats, and protective eyewear should worn when handling NaOH or solutions containing it.

Bacto-agar [Difco, 214530]

Ampicillin [Sigma, A6140]

T4 DNA ligase (20 U μl^{-1}; supplied with 10× buffer) [NEB, M0202S]

QIAquick Gel Extraction Kit [Qiagen, 28704]

! CAUTION - Some of the reagents used in this kit are classed as irritants. Gloves, lab coats, and protective eyewear should worn when using this kit.

SigmaSpin™ Sequencing Reaction Clean-Up [Sigma, S4309]

CRITICAL Required if splinkerette-PCR is to be performed in 96-well plates.

dNTP Set (100 mM Solution, 25 μMol) [Abgene, AB-0315/A]

Pfu Turbo Hotstart DNA Polymerase (supplied with buffer) [Stratagene, 600320]

Qiagen Multiplex PCR kit [Qiagen, 206143]

1 Kb DNA Ladder [Invitrogen, 15615024]

Agarose (electrophoresis grade) [Invitrogen, 15510027]

Ethidium Bromide Solution (10 mg/mL) [Sigma-Aldrich, E1510]

! CAUTION - Ethidium Bromide is a suspected mutagen and is a possible carcinogen. Gloves, lab coats, and protective eyewear should worn when handling this reagent and any gels or solutions carrying it. Ethidium Bromide containing material should be disposed of in accordance with local institutional safety procedures.

TOPO TA Cloning® Kit for Sequencing (containing pCR®4-TOPO®) [Invitrogen, K4575-40]

ElectroMAX DH10B™ Cells [Invitrogen, 18290-015]

†GS emPCR kit III (Amplicon B, paired end) [Roche, 04891392001]

CRITICAL Required if splinkerette-PCR sequencing is to be performed on the 454 platform.

†Exonuclease 1 (20,000 U/mL) [M0293L]

CRITICAL Required if splinkerette-PCR sequencing is to be performed on the 454 platform.

†Shrimp alkaline phosphatase (2,000 U/mL). Supplied with 10×SAP buffer [M0289B]

CRITICAL Required if splinkerette-PCR sequencing is to be performed on the 454 platform.

Equipment

1.5 mL microtubes [Axygen, 311-04-051]

PCR tubes and lids (strips of 8) [Abgene, AB-0771 and -0784]

Thin-walled 96-well PCR plates [Abgene, SP-1043]

CRITICAL These are required if splinkerette-PCR is to be performed in 96-well plates.

Electroporation Cuvettes (0.1 cm, Brown Top) [Biorad, 1652089]

Deep-well 96-well plates [PLPL0082]

Millipore filter plates [LSKMPCR50]

CRITICAL Required if splinkerette-PCR sequencing is to be performed on the 454 platform. 96 well plate thermal cycler

37°C incubator

Heating block

Gel running equipment including gel tank and power supply.

Micro-centrifuge

Electroporator

Benchtop 96 well plate centrifuge

CRITICAL Necessary if splinkerette-PCR is to be performed in 96 well plates.

454 FLX DNA sequencer

CRITICAL If 454 FLX instruments are not available, commercial access to 454 technology can be obtained through the following pay for service organizations including:

www.454.com

http://www.mwg-biotech.com

http://www.cogenics.com

www.gatc-biotech.com

http://www.biotech.ufl.edu/services.html

www.macrogen.com/

Reagent Set-up

Preparation of the splinkerette adapter mix—'Long strand adapter' and 'short strand adapter' oligonucleotides (shown in Table 1) need to be HPLC-purified. The oligos should be resuspended in 5× NEB buffer 2 (250 mM NaCl, 50 mM Tris-HCl, 50 mM $MgCl_2$, 1 mM Dithiothreitol; pH 7.9) to a concentration of 50 μM. Stocks can be frozen at −20°C. Add 50 μL of 50 μM long strand adapter and 50 μL of 50 μM short strand adapter to a 1.5 mL tube and vortex to mix. Each adapter primer will be present at a final concentration of 25 μM. This adapter mix can be stored indefinitely at −20°C and will provide enough splinkerette-adaptor for 100 splinkerette-PCR reactions.

SOC medium—To 950 mL double-distilled water, add 20 g Bacto-tryptone, 5 g Bacto-yeast extract, 0.5 g NaCl and 2.5 mL of 1 M KCl. Adjust the pH to 7.0 with 10 N NaOH and then make up to final volume of 1 L using double-distilled water. Autoclave to sterilise, then store at room temperature (21°C–25°C). Add 20 mL sterile 1 M glucose immediately before use. SOC medium is stable for 1 week after the addition of the glucose. SOC medium with glucose provides a very rich environment for bacterial growth and even with good sterile technique can easily become contaminated. It is advisable to inspect SOC medium for contamination, which will make the medium 'cloudy', prior to use.

LB+amp plates—Add 10 g Bacto-tryptone, 5 g Bacto-yeast extract, 10 g NaCl and 15 g Bacto-agar to double-distilled water (final volume is 1 L). Autoclave to sterilize, cool to 55°C, add 1 mL 100 mg/mL ampicillin (amp), then pour into sterile 10 cm^2 petri dishes. Store at 4°C and dispose of unused plates after 3 weeks.

LB+amp medium—Add 10 g Bacto-tryptone, 5 g Bacto-yeast extract and 10 g NaCl to double-distilled water (final volume is 1 L). Autoclave to sterilize, then store at room temperature. To 1 L add 1 mL 100 mg/mL ampicillin. After the addition of ampicillin store medium for a maximum of 2 weeks at 4°C.

PROCEDURE

Tumour Collection and Storage

1: Collect mouse tumour material into screw-topped cryo-vials and snap frozen in liquid nitrogen. Tissue can be stored at −80°C (or in liquid nitrogen for long term storage).

Extraction of tumour genomic DNA – TIMING: 4 hr.

2: Remove cryo-vial from the liquid nitrogen and place on dry ice. With a sharp scalpel blade shave 5-10 mg of tissue from the tumour sample and extract the genomic DNA using the Gentra Purgene Cell Kit (Qiagen) according to the manufacturer's instructions. Unused tissue can be returned to the cryovial and stored in liquid nitrogen.

Caution – The solutions used with in the Gentra Puregene Cell Kit contain irritants such as proteinase K and lyticase. Gloves, lab coats and eye protection are recommended. Liquid nitrogen should be handled with care.

Pause Point – The DNA can be stored at 4°C or at −20°C indefinitely.

Digestion of tumour genomic DNA with restriction enzyme – TIMING: 12-16 hr.

3: Digest the genomic DNA by setting up the reaction listed below in a 1.5 mL tube. DNA quantities in the range of 0.5-5 μg are acceptable. 2 μg is optimal. We suggest quantifying DNA by spectrophotometry using a nanodrop.

Component	Amt. per tube (μL)	Final
Genomic DNA	1 *	2 μg genomic DNA
*Sau3*A1 Enzyme (20 U/μL)	2	40 U
10× NEB buffer 1	3	1×
Acetylated BSA (10 mg/mL)	1	0.33 μg/μL
Double distilled Water	23 *	-
Final volume	30	

* Varies depending on DNA concentration.

4: Incubate the digestions at 37°C overnight (12-16 hr), either in a heating block or incubator.

Pause Point – The digested genomic DNA can be stored at 4°C or at −20°C indefinitely.

Splinkerette-adaptor ligation– TIMING: 12 hr.

5: Thaw the adaptor mix (prepared as described in 'reagent set up') on ice. Pipette the desired amount into a PCR tube (1 μL of adaptor mix per digested genomic DNA sample from step 4). The adapter mix is then denatured and annealed by heating it to 95°C for 5 min and then cooling to room temperature at the rate of 1°C every 15 sec. This is best performed in a thermal cycler. Once completed the PCR tube should be kept on ice.

6: Immediately before use, heat the *Sau3*AI-digested genomic DNA to 65°C for 20 min then place on ice. Set-up the following splinkerette adaptor ligation reaction:

Component	Amt. per tube (μL)	Final
*Sau3*AI digested genomic DNA (2 μg in 30 μL)	4.5	300 ng (~2 pmol) *
Adaptor mix (25 μM)	1	25 pmol (10-fold molar excess)
T4 DNA-ligase (20 U μl)	1	20 U
10× T4 DNA-ligase buffer	4	1×
Double distilled Water	29.5	-
Final volume	40	

* Using a 4 bp restriction enzyme, 300 ng genomic DNA equates to ~1-2 pmol genomic DNA depending on the enzyme used.

7: Incubate the ligation reaction at 4°C overnight (12-16 hr).

8: Heat inactivate the T4 DNA ligase at 65°C for 20 min (in a heating block).

Pause Point – The spinkerette ligation reaction can be stored at −20°C indefinitely.

Digestion to remove genomic DNA fragments with splinkerette-adaptor ligated on both ends. – TIMING: 6 hr.

9: *EcoRV* digestion of the ligation reaction (from step 8) removes any adaptor that has ligated onto the LTR-end of the *Sau*3AI digested genomic DNA (Fig. 1a) and prevents amplication of internal proviral fragments. Set-up the following digestion reaction in a 1.5 mL tube:

Component	Amt. per tube (μL)	Final
Splinkerette adaptor ligation	40	-
10× NEB buffer 3	10	1×
EcoRV(20 U/μL)	1	20 U
Acetylated BSA (10 mg/mL)	1	0.1 μg/μL
Double distilled Water	48	-
Final volume	100	

10: Incubate the tube at 37°C for at least 6 hr (overnight is permitted).

11: Heat inactivate the enzyme by heating the tube at 65°C for 20 min (in a heating block).

Pause Point – The *Eco*RV-digested spinkerette ligation reaction can be stored at −20°C indefinitely.

Clean up the *Eco*RV digestion reaction – TIMING: 10 mins.

12: To clean up the splinkerette adaptor ligated DNA use the QIAquick Gel Extraction Kit according to the manufacturer's instructions, with the exception that the DNA does not need to be run on an agarose gel. Simply add 600 μL of Buffer QG (supplied in the kit) to the reaction from step 11 and complete the protocol as per the manufacturer's instructions. Elute the DNA in 40 μL of 10 mM Tris (pH 7.4).

Caution – Some of the solutions used with the QIAquick Gel Extraction Kit contain chaotropic salts which are irritants. Gloves, lab coats and eye protection are recommended.

13: Although not always necessary, the samples can be quantified using a nanodrop (and are typically 5-15 ng/μL).

Pause Point – The purified spinkerette ligation reaction can be stored at −20°C indefinitely.

Primary PCR Amplification – TIMING: 2 hr.

14: To amplify the junction between the genome and the viral insertion a nested-PCR reaction is used. Set-up the primary PCR as follows using the proofreading enzyme, *Pfu*. Primer sequences are shown in Table 1.

Component	Amt. per tube (μL)	Final
Purified splinkerette adaptor ligated genomic DNA	5	-
Splink1 primer (0.1 μg/μL)	1	0.1 μg

Component	Amt. per tube (μL)	Final
U3LTR#5 primer (0.1 μg/μL)	1	0.1 μg
dNTPs (25 mM)	0.5	0.25 mM
10× *Pfu* buffer	5	1×
Pfu Turbo Hotstart DNA polymerase (2.5 U/μL)	0.5	1.25 U
Double distilled Water	37	-
Final volume	50	

15: The PCR protocol is as follows: 3 min at 94°C then 29 cycles of 15 sec at 94°C, 30 sec at 68°C, 3 min at 72°C, with a final extension of 5 min at 72°C. A water only negative control reaction should be run to test for possible reagent contamination.

Pause Point – The primary PCR reaction can be stored at −20°C indefinitely.

Secondary PCR Amplification – TIMING: 2 hr.

16: The secondary PCR primers are nested relative to the primers used in the primary PCR and therefore increase the specificity of isolating virus/genome junctions. These primers may also be barcoded and tagged with the 454 ‘primer A’ and ‘primer B’ sequences (Table 1). Set-up the secondary PCR as follows.

Component	Amt. per tube (μL)	Final
Primary PCR product	1	-
Splink2 primer (0.1 μg/μL)*	1	0.1 μg
U3LTR#1 primer (0.1 μg/μL)*	1	0.1 μg
Qiagen Multiplex Mastermix	25	-
Double distilled Water	22	-
Final volume	50	

*Splink2-454 (A) and U3LTR#1-454 (B) must be used (see Table 1) if splinkerette PCR is to be sequenced on a 454 sequencer (see Step 19B).

17: The PCR protocol is as follows: 15 min at 94°C then 25 cycles of 15 sec at 94°C, 30 sec at 60°C, 5 min at 72°C, with a final extension of 5 min at 72°C. A water only negative control reaction should be run to test for possible reagent contamination.

Pause Point – The secondary PCR reaction can be stored at −20°C indefinitely.

Agarose Gel Analysis of splinkerette-PCR reactions – TIMING: 2 hr.

18: Visualize the results of the secondary splinkerette-PCR reaction by loading 15-20 μL on a 2-4% agarose gel (in 1× TAE) with ethidium bromide (0.5 μg/mL). The PCR products should appear as a ladder (Fig. 1c). The remaining PCR products should be purified using the QIAquick Gel Extraction Kit (as described in step 12).

Caution – Ethidium bromide is a known mutagen and a suspected carcinogen.

De-multiplexing and sequencing of splinkerette-PCR products

19: The splinkerette-PCR products now represent a pool that must be de-multiplexed prior to sequencing. This can be done either by shotgun cloning and capillary read sequencing (option A) or 454 sequence analysis (option B). 454 sequencing is only an option if the secondary PCR has been performed using the Splink2-454 (A) primer and the U3LTR#1-454 (B) primer (Table 1) since the 'primer A' and 'primer B' sequences in these primers are essential for 454 sequencing.

A: Shotgun cloning prior to capillary read sequencing – TIMING: ~1 week.

i. To shotgun clone the PCR products set-up the following TOPO cloning reaction in a 1.5 mL tube (note: Using the water supplied with the "TOPO TA Cloning® Kit", make a 1 in 4 dilution of the "salt solution" (1.2 M NaCl, 0.06 M $MgCl_2$) before using it in the cloning reaction listed below).

Component	Amt. per tube (μL)	Final
Purifed splinkerette DNA	1	-
pCR®4-TOPO® plasmid (10 ng/μL)	0.14	1.4 ng
Diluted (1/4) salt solution	1	0.05 M NaCl, 2.5 mM $MgCl_2$
Double distilled water (supplied with kit)	3.9	-
Final volume	6.04	

ii. Incubate the ligation at room temperature for 20 min before stopping the reaction by transferring the tube to −80°C.

Pause Point – The Spinkerette PCR ligation reactions can be stored at −80°C indefinitely.

iii: To generate transformants of the ligation reactions, add 1 μL of the reaction and 50 μL of ElectroMax electrocompetent DH10B™ bacterial cells to a pre-chilled 0.1 cm gap electroporation cuvette (sitting on ice). Transform the bacterial cells (electroporate using the following conditions: 1.8 kV, 25 μF and 200 Ω).

iv: Recover transformants by addition of 1 mL SOC medium, transfer the transformants to a 15 mL culture tube, and incubate in a shaking incubator at 37°C for 1 hr.

v: Plate transformants onto LB+amp plates (prepared as described in 'reagent set up') and grow for 12-16 hr until colonies appear. Normally it is only necessary to plate 20 μL of the transformation to yield approximately 100-200 colonies per plate. Colony numbers are usually reproducible because the ligation reaction is driven to completion. Be sure to include a negative 'no insert' control to test reagents for contamination. The remaining (unplated) transformation can stored at 4°C for up to 4 days or alternatively glycerol stocks can be prepared by adding sterile glycerol to a final concentration of 15%.

vi: Pick colonies into deep-well 96-well plates containing 1 mL LB+amp medium (prepared as described in 'reagent set up') per well and grow in a shaking incubator at 37°C for 12-16 hr. Purify plasmids from the colonies and sequence using standard plasmid-based sequencing protocols[17] on a high-throughput sequencing platform described previously[18].

B: 454 'barcode' sequencing of splinkerette-PCR reactions – TIMING: ~1 day.

i. An alternative approach of de-multiplexing and sequencing insertion sites is to use 454 parallel sequencing (Fig. 2)[14].,For optimal results when using the 454 parallel sequencing platform splinkerette-PCR products should be cleaned of any unwanted dNTPs and primers – set up an Exo-Sap reaction usually in 96 well plates as follows:

Component	Amt. per tube (μL)	Final
Splinkerette-PCR DNA	15	-
10× SAP buffer	2	20 mM Tris-HCl,10 mM $MgCl_2$ (pH 8)
Exonuclease I (20,000 U/mL)	0.25	5 U
Shrimp alkaline phosphatase (2,000 U/mL)	0.25	0.5 U
Double distilled water	2.50	-
Final volume	20	

ii. Incubate the reaction at 37°C for 60 min, followed by 80°C for 20 min

iii. Purify the PCR products using a Millipore filter plate. Add 20 μl of PCR product and 80 μL of HPLC water to the filter plate. Centrifuge the plate at 3,220 *g* for 10 min. Add 20 μL of HPLC water to the filter plate and centrifuge at 242 *g* for 10 min. Collect the re-suspended splinkerette-PCR products and place in a clean 96 well plate.

Pause point – The purified splinkerette-PCR can be stored at −80°C indefinitely.

iv: Quantify the splinkerette-PCR products on a nanodrop and then pool the samples. We routinely pool 10-20 μL of PCR product from up to 100 splinkerette-PCR reactions (the DNA can be concentrated by ethanol precipitation to reduce the total final volume or using a Qiaquick column). The degree of multiplexing depends greatly on the application and the expected complexity of the pool and it is probably best to optimize this factor for each experiment. We have designed primers so that in the second splinkerette-PCR primer B is incorporated into the splinkerette product at the LTR end and primer A is incorporated at the linker end.

v: Emulsion PCR should then be performed on these products using "GS emPCR kit III" from Roche. A short nucleotide 'barcode' in the LTR end primer is used as a tag to computationally de-multiplex the 454 splinkerette-PCR sequences. A list of suitable barcodes, designed to lack homopolymer stretches, is provided in Supplementary Table 1.

CRITICAL STEP The choice of barcodes should be carefully considered depending on the application. 5 bp barcodes represent a minimum length that should be used. 10 bp is optimal. To increase the unambiguous assignment of sequences to tumours within the pool of sequenced products it may be desirable to use longer barcode sequences. Barcode length largely depends of the degree of multiplexing within each pool[19].

DNA sequence trace analysis

20: After the DNA sequences have been generated they need to be analyzed. The first phase of this process is to remove contaminating sequence by identifying and discarding those sequence traces that do not contain the end of the LTR and where the barcode cannot be read

if sequencing was performed on the 454 platform. Software for mapping and analysis of insertion sites within the remaining sequences has been described elsewhere[13,20-23].

Timing

The entire splinkerette-PCR procedure usually takes 4-5 days depending on the duration of the digestion and ligation steps.

Step 2: Extraction of tumour genomic DNA: 4 hr

Steps 3-4: Digestion of tumour genomic DNA with *Sau*3AI: 12-16 hr

Steps 5-8: Splinkerette adaptor ligation: 12 hr

Steps 9-11: Digestion to remove genomic DNA fragments with splinkerette-adaptor ligated on both ends: 6 hr

Step 12: Clean-up of *Eco*RV digestion reaction: 10 min

Steps 14-15: Primary PCR amplification: 2 hr

Steps 16-17: Secondary PCR amplification: 2 hr

Step 18: Agarose gel analysis: 2hr

Step 19A: Shotgun cloning prior to capillary read sequencing ~1 week

Step 19B: 454 'barcode sequencing of splinkerette-PCR reactions ~1 day

Step 20: DNA sequence trace analysis ~ 1 week

Troubleshooting

Troubleshooting advice can be found in Table 2.

Anticipated Results

We have successfully used this approach to generate the DNA sequence of over 1,000 splinkerette-PCR reactions generated using both viruses and transposons. Successful splinkerette-PCR will yield a ladder of bands (Fig. 1c). The number and the size of the bands greatly depends on the number of insertion sites in the tumour. When analysing large numbers of tumours it is important to perform several splinkerette-PCR reactions on control genomic DNA that does not contain any viral or transposon insertions. This is because there are elements such as pseudo-LTR or pseudo-transposon repeat-like sequences in every genome and these may be non-specifically amplified. Although rare, mispriming of splinkerette primers can also yield PCR artefacts. Sequencing of a splinkerette-PCR product is the only sure way to ensure that it faithfully represents a true insertion site.

Supplementary Table 1: 10 bp sequences suitable for 'barcoding' splinkerette products for 454 sequencing. These sequences facilitate the computational 'demultiplexing' of pooled splinkerette-PCR products following 454 sequencing.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Europe PMC Funders Author Manuscripts

Acknowledgments

A.G.U, J.K, M.v.L and A.B are funded by the NWO Genomics program and the Netherlands Genomics Initiative/ Netherlands Organization for Scientific Research (NWO). D.J.A and his laboratory are funded by Cancer Research-UK and the Wellcome Trust. L.v.d.W is funded by the Kay Kendall Leukaemia Fund. Thanks to Lara Collier for critically reading this manuscript.

References

1. Greenman C, et al. Patterns of somatic mutation in human cancer genomes. Nature. 2007; 446(7132):153. [PubMed: 17344846]
2. Mikkers H, Berns A. Retroviral insertional mutagenesis: tagging cancer pathways. Adv Cancer Res. 2003; 88:53. [PubMed: 12665053]
3. Dupuy AJ, et al. Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system. Nature. 2005; 436(7048):221. [PubMed: 16015321]
4. Keng, VW., et al. Forward genetic screen for hepatocellulear carcinoma associated genes. Proceedings of the 99th Annual Meeting of the American Association for Cancer Research; 2008.
5. Wu, X., et al. Murine metastatic Medulloblastoma driven by Sleeping Beauty transposition identifies genes and pathways important for medulloblastoma initiation and progression. Proceedings of the 99th Annual Meeting of the American Association for Cancer Research; 2008. ; Bender, AM., et al. Sleeping Beauty mediated gliomagenesis. Proceedings of the 99th Annual Meeting of the American Association for Cancer Research; 2008.
6. Collier LS, et al. Cancer gene discovery in solid tumours using transposon-based somatic mutagenesis in the mouse. Nature. 2005; 436(7048):272. [PubMed: 16015333]
7. Cuypers HT, et al. Murine leukemia virus-induced T-cell lymphomagenesis: integration of proviruses in a distinct chromosomal region. Cell. 1984; 37(1):141. [PubMed: 6327049]
8. Ochman H, Gerber AS, Hartl DL. Genetic applications of an inverse polymerase chain reaction. Genetics. 1988; 120(3):621. [PubMed: 2852134]
9. McAleer MA, Coffey AJ, Dunham I. DNA rescue by the vectorette method. Methods Mol Biol. 2003; 226:393. [PubMed: 12958526]
10. Devon RS, Porteous DJ, Brookes AJ. Splinkerettes-improved vectorettes for greater efficiency in PCR walking. Nucleic Acids Res. 1995; 23(9):1644. [PubMed: 7784225]
11. Hui EK, Wang PC, Lo SJ. Strategies for cloning unknown cellular flanking DNA sequences from foreign integrants. Cell Mol Life Sci. 1998; 54(12):1403. [PubMed: 9893713]
12. Hengen PN. Vectorette, splinkerette and boomerang DNA amplification. Trends Biochem Sci. 1995; 20(9):372. [PubMed: 7482706]
13. Uren AG, et al. Large-scale mutagenesis in p19(ARF)- and p53-deficient mice identifies cancer genes and their collaborative networks. Cell. 2008; 133(4):727. [PubMed: 18485879]
14. Margulies M, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005; 437(7057):376. [PubMed: 16056220]
15. Wang GP, et al. HIV integration site selection: analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. Genome Res. 2007; 17(8):1186. [PubMed: 17545577]
16. Mikkers H, et al. High-throughput retroviral tagging to identify components of specific signaling pathways in cancer. Nat Genet. 2002; 32(1):153. [PubMed: 12185366]
17. Marra MA, Kucaba TA, Hillier LW, Waterston RH. High-throughput plasmid DNA purification for 3 cents per sample. Nucleic Acids Res. 1999; 27(24):e37. [PubMed: 10572189]
18. Mullikin JC, McMurragy AA. Techview: DNA sequencing. Sequencing the genome, fast. Science. 1999; 283(5409):1867. [PubMed: 10206892]
19. Parameswaran P, et al. A pyrosequencing-tailored nucleotide barcode design unveils opportunities for large-scale sample multiplexing. Nucleic Acids Res. 2007; 35(19):e130. [PubMed: 17932070]
20. Giordano FA, et al. New bioinformatic strategies to rapidly characterize retroviral integration sites of gene therapy vectors. Methods Inf Med. 2007; 46(5):542. [PubMed: 17938776]

21. Kong J, Zhu F, Stalker J, Adams DJ. iMapper: a web application for the automated analysis and mapping of insertional mutagenesis sequence data against Ensembl genomes. Bioinformatics. 2008; 24(24):2923. [PubMed: 18974167]
22. Kustikova OS, Modlich U, Fehse B. Retroviral insertion site analysis in dominant haematopoietic clones. Methods Mol Biol. 2009; 506:373. [PubMed: 19110639]
23. Ambrosi A, Cattoglio C, Di Serio C. Retroviral integration process in the human genome: is it really non-random? A new statistical approach. PLoS Comput Biol. 2008; 4(8):e1000144. [PubMed: 18688267]

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

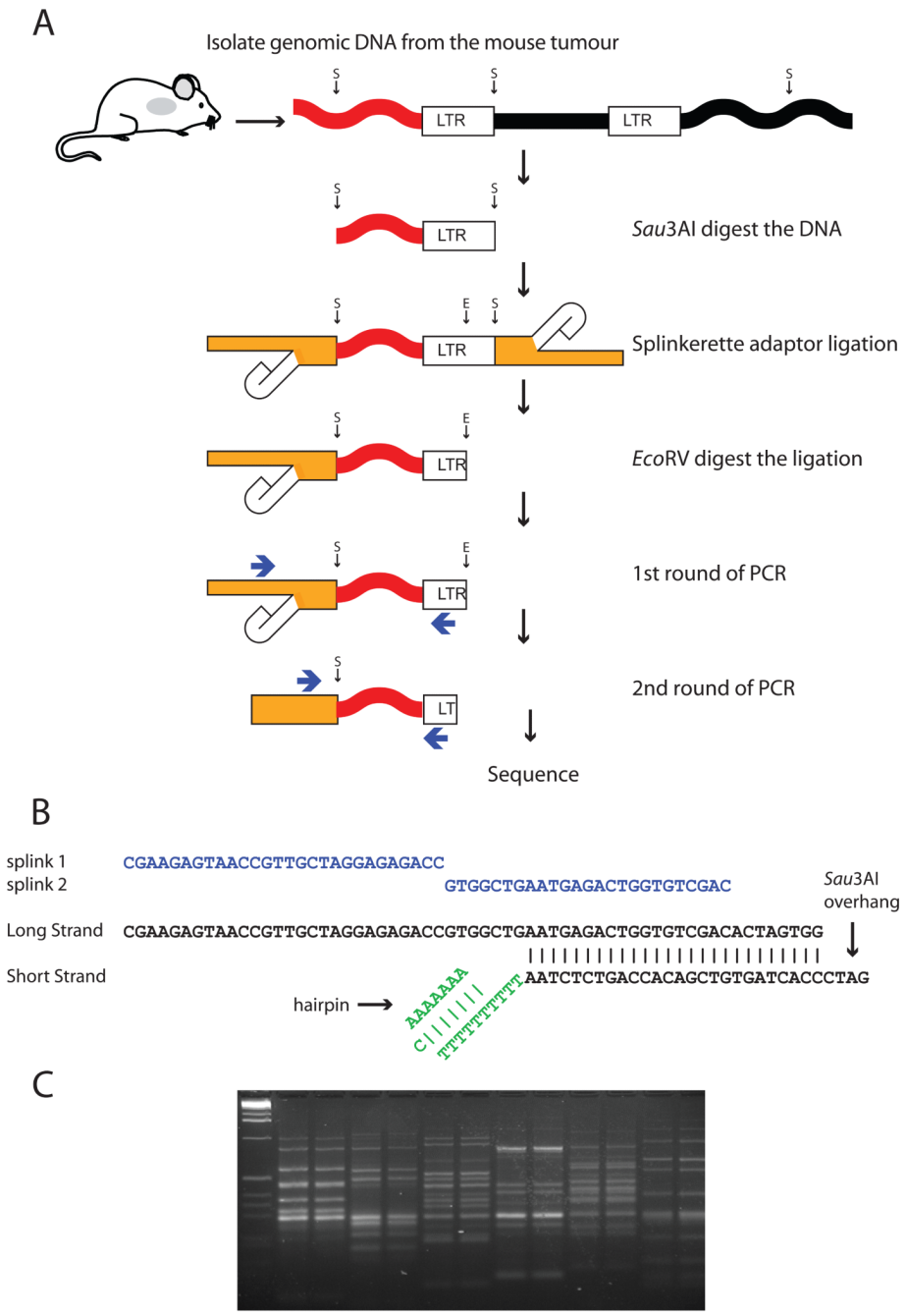

Figure 1.

Overview of the splinkerette-PCR protocol. **A.** The steps performed to generate splinkerette-PCR products for sequencing. E and S refer to *Eco*RV and *Sau3*AI, respectively. **B.** The structure of the splinkerette-adaptor. **C.** The expected result for a successful splinkerette-PCR reaction. The first lane of this 4% gel is a 100 bp marker followed by duplicate splinkerette-PCR reactions for 6 MuLV tumour samples.

Europe PMC Funders Author Manuscripts

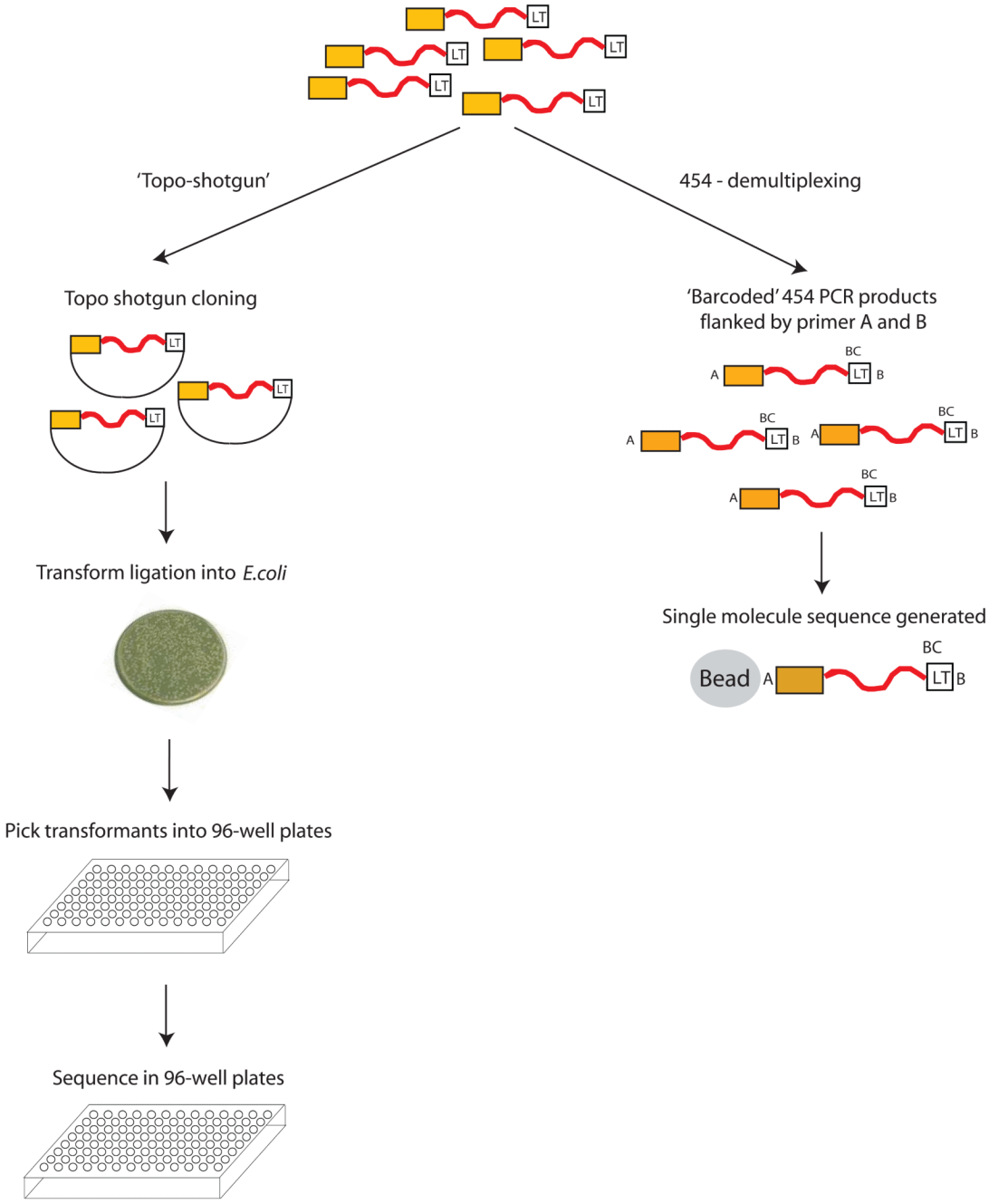

Figure 2.

There are two methods for high-throughput sequencing of splinkerette-PCR products. These include the 'TOPO-shotgun' approach where pools of splinkerette-PCR products from each tumour are individually shotgun cloned and sequenced on a capillary sequencing platform, and the 454 sequencing approach where spinkerette-PCR products from multiple tumours are pooled together and subjected to single molecule sequencing on the 454 platform. When 454 sequencing is used, products from individual splinkerette-PCR reactions are barcoded (BC) using a short nucleotide 'tag' sequence and the second round PCR primers contain sequences for primer A and primer B which are necessary for the emulsion PCR and sequencing steps on the 454 platform.

Table 1
Oligonucleotides required for this protocol

The underlined sequences are the 454 Primer A (normal text) and 454 Primer B (Italics) sequences respectively. The bold **XXXXXXXXXX** represents the 10bp barcode sequence used to identify each tumour. A list of suitable barcodes is provided in the supplementary Table 1.

Name	Sequence (5′ to 3′)	Comment
For standard splinkerette-PCR:		
Long strand adapter	CGAAGAGTAACCGTTGCTAGGAGAGACCGTGGCTGAATGAGACTGGTGTCGACACTAGTGG	Add bases to 3′ end for 3′ overhang enzymes
Short strand adapter	GATCCCACTAGTGTCGACACCAGTCTCTAATTTTTTTTTTCAAAAAAA	Contains hairpin and GATC 5′ overhang
Splink1	CGAAGAGTAACCGTTGCTAGGAGAGACC	1st round PCR primer on splinkerette adaptor
U3 LTR#5	GCGTTACTTAAGCTAGCTTGCCAAACCTAC	1st round PCR primer on MuLV viral LTR
Splink2	GTGGCTGAATGAGACTGGTGTCGAC	2nd round PCR primer on splinkerette adaptor
U3 LTR#1	CCAAACCTACAGGTGGGGTCTTTC	2nd round PCR primer on MuLV viral LTR
For 454 splinkerette-PCR:		
Splink2-454A	GCCTCCCTCGCGCCATCAGGTGGCTGAATGAGACTGGTGTCGAC	2nd round PCR primer adapted with 454 primer A
U3 LTR#1-454B	*GCCTTGCCAGCCCGCTCAG***XXXXXXXXXX** CCAAACCTACAGGTGGGGTCTTTC	2nd round PCR primer adapted with 454 primer B

Europe PMC Funders Author Manuscripts

Table 2

Troubleshooting

Step	Problem	Solution
18	No splinkerette-PCR products on gel	The splinkerette-PCR reaction has many steps and as such are there are numerous places where the protocol may fail. The most frequent source of failure is an incorrect ratio of splinkerette-adaptor to genomic DNA in the splinkerette ligation. Be sure to carefully prepare the splinkerette-adaptor and to quantify the genomic DNA prior to digestion.
20	Chimeric splinkerette products	One of the major contaminants in splinkerette-PCR is chimeric genomic sequences ligated between the LTR and the adaptor. These are splinkerettes carrying multiple genomic fragments between the end of the insertional mutagen and the linker. It is very hard to remove these by modifying the splinkerette-PCR reaction but these should be carefully masked out when the sequence data is analyzed.
20	Linker – genomic DNA – linker products	Occasionally linker – genomic DNA – linker fragments are generated. The generation of these products is suppressed by cutting with an internal enzyme or they can be masked out during the analysis of the sequence.