Published in final edited form as:

Methods Mol Biol. 2016 ; 1400: 79–93. doi:10.1007/978-1-4939-3372-3_5.

Long Interspersed Element Sequencing (L1-Seq): A Method to Identify Somatic LINE-1 Insertions in the Human Genome

Tara T. Doucet and Haig H. Kazazian Jr.

Abstract

L1-seq is a high-throughput sequencing technique which is utilized to identify novel L1 insertions in genomic DNA samples of interest. Using special diagnostic nucleotides unique to the youngest and most active L1 sequence, we can amplify new somatic insertions. This technique has helped to establish the number of L1 insertions present in the general population as well as the variation among individuals with regard to their complement of active L1 elements. More recently, this technique has been employed to assess the level of retrotransposition occurring in various diseases such as cancer. These efforts try to establish a connection between the process of retrotransposition and disease development and/or progression.

Keywords

Non-LTR retrotransposon; Retroelement; LINE-1; L1; Next-Generation DNA sequencing

1 Introduction

Retrotransposons are nearly ubiquitous in eukaryotes from slime molds [1] to humans [2] and have contributed greatly to genome composition of these organisms. Retrotransposons make up 45 % of the human genome [2]. In particular, the LINE-1 (L1) element has contributed to approximately 17 % of the human genome and continues to add to it via a copy and paste mechanism with an RNA intermediate [2]. L1 is the only autonomous retrotransposon in the human genome because it encodes two proteins necessary for mobilization and reinsertion into the genome; however, these two proteins, once expressed can mobilize other types of retrotransposons as well as processed pseudogenes [3–5]. Each individual has a different complement of potentially active L1 elements, although the majority of the L1s in each individual's genome are truncated and therefore inactive. L1-seq [6] was developed to help characterize L1 variation among individuals because L1s have contributed to a substantial fraction of the genome and are capable of inducing many types of mutations. L1-seq has since been used to evaluate several types of cancer to establish the level of retrotransposition occurring in colon cancer, lung cancer, breast cancer, and many other cancers [7, 8]. Additional sequencing techniques have confirmed the L1-seq data and demonstrated that L1 elements are active in many cancer types [8–13]. The results have demonstrated that L1s are active in a subset of patients with cancer; in addition, L1 elements are active in all epithelial cancers tested. The L1-seq technique consists of a DNA library

prep as well as the validation of the predicted new insertions detected in the samples used. Although few of the insertions may be directly responsible for the development of the disease, it should be possible to utilize known insertions present in a cancer sample for monitoring the cancer's progression to metastasis. To detect metastasis using a new L1 insertion, a PCR would be performed on serum DNA from a patient to determine whether or not the insertion was detectable in the blood and therefore potentially in a floating cancer cell. This technique is useful both for evaluating the overall complement of L1 elements in a genome as well as looking for new insertion events. L1-seq utilizes unique nucleotides, "ACA" 91–93 nucleotides from the 3' end of the element, to selectively amplify the young and active subset of elements in the human genome [6]. Following the initial five cycles of the PCR, wherein the linear amplification of L1 elements occurs, degenerate primers are added to the mixture to exponentially amplify both polymorphic and potentially somatic insertions present in the genome.

2 Materials

Store all reagents as specified by manufacturers. Diligently follow all waste disposal regulations when disposing of waste materials. All primers need to be diluted to 100 μ M upon receipt in diethylpurocarbonate (DEPC) water. Primers will be further diluted as specified later in the protocol.

2.1 DNA Isolation

1. DNeasy Blood and Tissue Kit (Qiagen).
2. 500 mL of absolute ethanol (200 proof).
3. Qubit™ dsDNA BR Assay kit (Life Technologies).

2.2 Library Preparation

1. Promega GoTaq Flexi.
2. 25 mM MgCl₂.
3. 10 mM dNTPs.
4. Diethylpurocarbonate (DEPC) water.
5. 100 % DMSO.
6. Pfu polymerase.
7. 1 μ g of good quality DNA per sample at a concentration of 100 ng/ μ L (*see Note 1*).
8. LE agarose.

¹If very little DNA is available for both library prep and validation PCRs, L1-seq can still be successfully performed. L1-seq has successfully been executed with as little as 25 ng of input per sample for the library prep. For the steps following next-generation sequencing, whole genome amplification can be used (e.g. the Qiagen Repli-G kit) to provide more DNA to use for the validation PCRs. If adjusting the amount of DNA used, be sure to account for volume changes and the concentrations of the other reagents to ensure all final concentrations are the same as described in the original technique.

9. 1× TAE; Prepare 50× solution by dissolving 242 g Tris base in 750 mL of deionized water. Carefully add 57.1 mL glacial acetic acid, and 100 mL of 0.5 M EDTA, pH 8.0 and adjust solution to final volume of 1 L. Dilute the 50× solution to 1× in deionized water.
10. Ethidium bromide (10 mg/mL).
11. QIAquick Gel Extraction Kit (Qiagen).
12. MinElute PCR Purification Kit (Qiagen).
13. Isopropanol (200 proof).
14. 500 mL of Ethanol (200 proof).
15. Agilent DNA 1000 kit or high sensitivity DNA kit (choose as needed, *see* Note 1).
16. L1-seq primers (order HPLC grade for library preparation). *See* Table 1.

2.3 Next-Generation Sequencing Data Analysis

1. Server with at least 4GB of RAM and L1-seq scripts properly formatted (<https://github.com/adamewing/l1seq>).
2. Bowtie2 (<http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>) and all relevant human genome files and indices as per Bowtie2 instructions.

2.4 Data Validation

1. GoTaq Green Master Mix (2×).
2. Diethylpurocarbonate (DEPC) water.
3. DNA at concentration of 12.5 ng/μL.
4. LE agarose.
5. 1× TAE; Prepare 50× solution by dissolving 242 g Tris base in 750 mL of deionized water. Carefully add 57.1 mL glacial acetic acid, and 100 mL of 0.5 M EDTA (pH 8.0) and adjust solution to final volume of 1 L. Dilute the 50× solution to 1× in deionized water.
6. Ethidium bromide (10 mg/mL).
7. QIAquick Gel Extraction Kit (Qiagen).
8. Isopropanol (200 proof).
9. 500 mL of Ethanol (200 proof).
10. L1SP1A2 primer, L1nt112out, L1 “G” primer (*see* Table 1).

3 Methods

3.1 Embedding and Cryosectioning Tissue

1. To begin, embed each piece of tissue to be assayed in OCT freezing medium. You can simply put a thin layer of the media onto a pre-chilled (- 20 °C) chuck.
2. Quickly placing the thawed tissue section onto the OCT.
3. Immediately cover the tissue in more OCT until it is barely visible through the medium. The OCT medium will change from clear to white, when the entire block of tissue/OCT is completely frozen, you can begin to cryosection the tissue for DNA extraction. It is best for the freezing to occur as rapidly as possible, to this end, a heat extractor can be used to enhance and shorten the freezing process and adherence to the chuck on which the tissue is being embedded in the OCT freezing medium.
4. Set the cryostat to slice sections of tissue between 10 and 30 µm.
5. During sectioning, carefully remove each roll of tissue and place 10–20 slices into a pre-chilled (- 20 °C) 1.5 mL microtube. *See Note 2.*

3.2 Extracting DNA from Sectioned Tissue

1. Remove microtubes with tissue slices from -80° freezer and place them on ice.
2. Add 360 µL of Buffer ATL (DNeasy Blood and Tissue kit). There is no need to further homogenize the tissue. Add 40 µL of proteinase K and mix thoroughly by vortexing (*see Note 1*).
3. Incubate at 55 °C overnight until the tissue is completely lysed. Vortex occasionally during incubation to help disperse samples.
4. Vortex for 15 s.
5. Add 400 µL of buffer AL (DNeasy Blood and Tissue Kit) to the sample and mix thoroughly by vortexing.
6. Immediately add 400 µL of ethanol (200 proof) and mix again thoroughly by vortexing (notes).
7. Pipette 750 µL of the mixture (including any precipitate) into the DNeasy Mini spin column placed in a 2 mL collection tube (provided in the kit).
8. Centrifuge at 6000 *g* for 1 min. Discard flow through.
9. Repeat **steps 7 and 8** until all of mixture has been run through the same column for each sample.
10. After the final spin with the aforementioned mixture, discard the collection tube and replace it with a new 2 mL collection tube.

²If the tissue sample is large enough, more than one tube of tissue slices can be made. Following the sectioning, tissue slices should be stored at -80 °C until it is time to isolate DNA. Embedding tissue in OCT freezing medium is only one way of extracting DNA from frozen tissue.

11. Add 500 μ L of Buffer AW1. (Ensure that ethanol has been added to Buffer AW1 before use.)
12. Centrifuge for 1 min at 6000 *g*.
13. Discard flow-through and collection tube and place the spin column in a new 2 mL collection tube.
14. Add 500 μ L of Buffer AW2. (Ensure ethanol has been added to Buffer AW2 before use.)
15. Centrifuge for 3 min at 20,000 *g* to dry the DNeasy membrane and then discard flow-through and collection tube.
16. Place the DNeasy Mini spin column in a clean 1.5 mL micro-centrifuge tube and pipet 100 μ L of pre-warmed (55 °C) Buffer AE onto the DNeasy membrane.
17. Incubate at room temperature for 10 min.
18. Centrifuge for 1 min at 6000 *g* to elute.
19. Pipette an additional 50 μ L of pre-warmed AE buffer directly onto DNeasy membrane. (There is no need to replace the micro-centrifuge tube. The additional DNA elution can be collected into the same tube up to a volume of no more than 150 μ L.)
20. Incubate for 10 min at room temperature.
21. Centrifuge for 1 min at 20,000 *g*.

This protocol is adapted from the Qiagen handbook for the DNeasy Blood and Tissue Kit.

3.3 Measuring DNA Concentration (Qubit™)

Use the Qubit™ fluorometer to measure DNA concentration because it is one of the most accurate methods. Follow manufacturer protocols exactly and see Note 1. (http://www.ebc.uu.se/digitalAssets/176/176882_3qubitquickrefcard.pdf).

3.4 L1-seq Library Preparation (see Note 3)

1. Before beginning the relevant library prep PCRs, it is necessary to determine which samples will be pooled together in the library preparation. As many as ten samples can be pooled together without using barcoding (notes). Equal amounts of each sample must be put into the DNA pool to be used in the library. A total of 24 μ L of pooled DNA at a concentration of 100 ng/ μ L is needed. If there is not enough DNA available from one or more samples, a modified protocol can be used (notes).

³When first performing L1-seq it is prudent to execute a TA cloning step after completing the libraries and mixing them in equimolar ratios, but before completing the end-polishing step. To do this, simply take 1 μ L from the mixed libraries and use it in a Topo TA cloning reaction. Follow kit instructions and after growing colonies overnight, select 12 or more from each plate for colony PCR. Following colony PCR, run the product on a gel to be sure the cloning worked effectively, select some or all of the successful clones for Sanger sequencing. When analyzing the Sanger sequencing, look for different L1Ta elements from many different areas of the genome. Essentially, this is a step to check that the library does not consist of amplicons of only a handful of LINE-1 elements in the genome and that elements in the genome are equally represented in the library. This step does not need to be performed for every

2. Round 1 PCR: Linear amplification of L1 flanks followed by a hemi-specific PCR incorporating the Illumina sequencing primer (Fig. 1). To prevent running out of master mix, make enough for nine reactions even though only eight reactions will be assembled. Master mix (per 1 reaction): 10 μ L of Promega Go-Taq flexi buffer (5 $\times$), 6 μ L of 25 mM $MgCl_2$, 2 μ L of 20 μ M L1SP1A2 primer (the second primer for the reaction, the degenerate primers previously mentioned, will be added following the completion of five cycles of the PCR which consists of the linear amplification step), 0.5 μ L of DMSO 100 %, 0.5 μ L of FlexiTaq, 1 μ L of 10 mM dNTPs, 2 μ L of pooled DNA (at 100 ng/ μ L), 24 μ L of DEPC water. The reaction should total 46 μ L before the addition of 4 μ L of the degenerate primers to be added after linear amplification is finished. When the linear amplification is finished, add 1 of each of the degenerate primers at 5 μ M (e.g. DEGSeq1N5TCTGT) to each of the eight reactions. Use the following cycling program:

L1-Seq PCR 1

- a. 95 $^{\circ}$ C for 2 min 30 s.
 - b. 95 $^{\circ}$ C for 30 s.
 - c. 58 $^{\circ}$ C for 1 min.
 - d. 72 $^{\circ}$ C for 2 min.
 - e. Go to **step b** (5 $\times$).
 - f. 60 $^{\circ}$ C (pause and add 4 μ L of degenerate primer into each of the eight reactions, one primer per reaction for each of the eight different degenerate primers).
 - g. 95 $^{\circ}$ C for 30 s.
 - h. 55 $^{\circ}$ C for 30 s.
 - i. 72 $^{\circ}$ C for 1 min and 30 s.
 - j. Go to **step g** (14 $\times$).
 - k. 72 $^{\circ}$ C for 10 min.
 - l. 4 $^{\circ}$ C hold.
3. Purify all eight reactions on eight separate Qiagen PCR Clean-up columns following the Qiagen protocol, eluting in 50 μ L of pre-warmed (55 $^{\circ}$ C) EB (do a 10 min final incubation before elution to optimize DNA eluted from column).
 4. L1-Seq PCR 2: Amplification of library and addition of the Illumina sequencing adapters (Fig. 1). Master mix (for 1 reaction): 12.5 μ L of master mix (Promega GoTaq Green 2 $\times$), 1.5 μ L of 20 mM primer Adap1L1HsG, 1.5 μ L of 20 μ M Adap2Seq1, 2.5 μ L of purified round 1 product (1 degenerate primer per

library prep; however, if a problem occurs with next-generation sequencing, this step could consequently be taken to determine whether or not overrepresentation of a few elements precluded successful sequencing.

reaction), 7 μ L of DEPC ddH₂O. Again, make enough for nine reactions to prevent running out of master mix for the samples to be amplified. The reactions will each have a total of 25 μ L. Use the following cycling program:

L1-Seq PCR 2

- a. 95 °C for 2 min.
 - b. 95 °C for 30 s.
 - c. 62 °C for 30 s.
 - d. 72 °C for 1 min.
 - e. Go to **step b** (19 $\times$).
 - f. 72 °C for 5 min.
 - g. 4 °C Hold.
5. Resolve products on a 1 % TAE gel.
 6. Excise the constellation of bands between 200 and 500 nucleotides with a sterile scalpel (using a different scalpel for each reaction) and purify the DNA using the Qiagen Gel Purification protocol.
 7. Elute the library in 50 μ L of pre-warmed EB buffer (55 °C with a 10 min incubation before elution).
 8. Run each DNA sample on the Agilent Bioanalyzer with the DNA 1000 kit to get an accurate measure of concentration and the average size of the DNA amplified. Using the concentration and average size of the molecules, calculate how to add the DNA from all eight reactions in equimolar ratios to one tube. *See Note 4.*
 9. After mixing the DNAs together, purify the entire mixture with the Qiagen MinElute PCR purification kit eluting in 50 μ L of pre-warmed EB (55 °C and 10 min incubation at room temperature before elution).
 10. End-polishing must be performed on the library because Taq leaves adenine overhangs which could cause problems when the library is annealed to the Illumina flow cell. To accomplish the end-polishing: mix 6 μ L of 10 $\times$ Pfu buffer, 2.5 μ L of Pfu polymerase, 2.5 μ L of 10 mM dNTPs, and 49 μ L of library. Incubate for 1 h at 72 °C.
 11. Purify reaction on a Qiagen MinElute column and elute in 10 μ L of pre-warmed (55 °C) EB following a 10 min room-temperature incubation before elution.

⁴Occasionally, one of the degenerate primer reactions will not be as robust as the other reactions and when the libraries are run on a gel, the amount of DNA present is variable between reactions. This may not be an issue if there is enough DNA present after the gel purification for the samples to easily be mixed in equal amounts. However, if the concentration of the DNA isolated from the gel purification step is too little to continue without grossly diminishing the amount of total DNA in the combined library, simply repeat the second reaction of L1-seq and combine the isolated DNA from both gel purifications and concentrate the DNA. If the DNAs from the respective degenerate primer reactions are run on the Bioanalyzer and produce very different size distributions of products, it may be necessary to repeat the second L1 PCR again on that DNA sample as well. Ideally, the average product size for each degenerate primer reaction should be within one standard deviation of 350 nucleotides. If the size varies more than one standard deviation from 350, the reaction should be repeated and rerun on a gel. If the size is wrong, it is likely that the excision was initially imprecise.

12. Measure final DNA concentration with Qubit™ fluorometer to get an accurate concentration for sending samples for next-generation sequencing on the Illumina HiSeq 2500. Opt for single end sequencing and at least 100 bp reads. See Notes 3 and 5.

3.5 Analyzing the Next-Generation Sequencing Data (see Note 6)

1. Obtain the sequencing reads from the core or center where the samples were sequenced and transfer them into the L1-seq directory on the server in which all the correctly formatted L1-seq scripts reside. These scripts can be acquired from <https://github.com/adamewing/l1seq>.
2. Obtain the contents of a database with all reference L1 insertions and polymorphic L1 insertions which have been previously published to use for filtering sequencing data. A database of L1 insertions may be obtained at: <http://nar.oxfordjournals.org/content/43/D1/D43> [14].
3. Once the documents are downloaded through the terminal, they must be unzipped. To unzip the fastq.gz files type “gunzip -d FILE_NAME.fastq.gz &” (the & symbol allows the unzipping to run in the background so that you can set all the files and can unzip simultaneously by typing this command for each file in turn.)
4. After the files are unzipped, use the script to run bowtie and create indices for your data. You can execute this process with the command “./run_bowtie.py/ whatever_fastq/directions/to/bowtie directions/to/hg19.fa &”. For any process which takes more than 10 min, it is helpful to use the screen function by typing “screen -rAad” which will allow for the monitoring of all the processes simultaneously running. It also enables the user to monitor the total memory being used for analysis and the length of time each process has been running.
5. Once the run_bowtie script finishes, run the l1seq.py script as follows: “./l1seq.py -bam whatever.ba, > whatever.l1seq.txt &”.
6. Once all of the L1-seq.txt files are made, all the files need to be compressed for sorting. To compress the files, type “bgzip whatever_l1seq.txt &”.
7. To sort the files, type “tabix -s 1 -b 3 -e 4 whatever.l1seq.txt.gz &”.
8. All the files must be compared to one another (e.g. normal compared with tumor, etc.) to do this analysis, type “./compare.py group1_L1seq.txt.gz group2_l1seq.txt.gz group3_l1seq.txt.gz > filename_for_comparisons.tsv &”.
9. Finally, after the comparison file has been made, primers must be designed for validation of the data. It is best to run the makeprimers.pl script on the entire

⁵If the DNA being measured at any point in the library prep is at a low concentration and undetectable with the standard Qubit™ broad range kit or the Agilent 1000 DNA chip, there are low concentration versions of these reagents available.

⁶Barcoding may also be utilized with this technique; however, results may vary. In 2012, Evrony et al. performed L1-seq using barcoding and were able to validate some new LINE-1 insertions following sequencing analysis. However, other groups have had more difficulty getting the technique to work well and seem to have more success with pooling samples without barcodes. Pooling samples without barcodes does create more work for the validation steps of the technique; however, it seems to have more reproducible results.

comparison file before looking at the data because the script does not take long to run and having the primer sequences ready to order is very useful. To run this final script, type “./makeprimers.pl filename_for_comparisons.tsv > filename_for_comparisons_with_primers.tsv &”. Use sftp to transfer the files back to your local computer if desired.

3.6 Validating the Predicted Insertions from L1-Seq with Site-Specific PCR

1. The presence of nonreference insertions is validated with site-specific PCR (Fig. 2). If the samples are not barcoded (*see* Note 6), all samples in a pool must be evaluated for the presence or absence of a predicted insertion. The DNA from each input sample from a pool needs to be at 12.5 ng/μL and 2 μL of DNA used per reaction. The primers will be named by the makeprimers.pl script as “filled site” or FS and “empty site” ES refer to Fig. 2 for orientation of primers with regard to the potential insertion. For each validation to be complete, the FS and L1SP1A2 primer reaction needs to be performed on all samples in the pool from which the prediction came. If comparing two states of the same tissue such as tumor and normal and the insertion is predicted only in one, the reaction must also be performed on both DNA samples. A control reaction can also be performed with the “empty site” or ES primer and the FS primer. Both the FS and ES primers are genomic and will produce a product of predetermined size in any DNA sample regardless of presence or absence of an insertion.
2. For the FS/L1SP1A2 (filled site PCR) use the following master mix (1×): 12.5 μL of Promega GoTaq green (2×) master mix, 0.8 μL of 20 μM FS primer, 1.6 μL of 20 μM L1SP1A2, 2 μL of genomic DNA (12.5 ng/μL), 8.2 μL of DNA-free H₂O. For the FS/ES primers (empty site PCR) use the following master mix (1×): 12.5 μL of Promega GoTaq green (2×) master mix, 1 μL of 20 μM FS primer, 1 μL of 20 μM ES primer, 2 μL of genomic DNA (12.5 ng/μL), 9.5 μL of DNA-free H₂O. Use the following parameters for the PCR:
 - 3' L1 Validation PCR*
 - a. 95 °C for 2 min.
 - b. 95 °C for 30 s.
 - c. 57 °C for 30 s.
 - d. 72 °C for 1 min and 30 s.
 - e. Go to **step b** (29×).
 - f. 72 °C for 5 min.
 - g. 4 °C Hold.
3. Run the PCR products on a 1.5 % TAE gel to resolve the products (*see* Note 7). Take images of the gel while it is exposed to UV to visualize the products.

⁷If validation PCRs are unsuccessful after many attempts, be sure to check the specificity of the primers being used in the amplification. Oftentimes, it is helpful to perform a nested PCR following the first conventional PCR to amplify difficult or low-copy

Excise fragments which are unique to only one of the samples upon which the PCR was run (Fig. 2b). Isolate the DNA from the band and send for sequencing. If no clear filled site band is uniquely present in one of the samples tested, a nested PCR may be necessary (Fig. 2a). Alternatively, the PCR conditions can be further optimized to attempt to amplify the insertion (*see* Notes 7 and 8).

4. Finally, send the DNA for Sanger sequencing to ensure it is the correct product. Sequence the product with both the FS primer as well as the L1SP1A2 primer. When the sequence from the FS is aligned to the genome with BLAT or another alignment algorithm, part of the sequence should align to the genome and a poly T tract should also be visible adjacent to the aligning sequence. For the sequence from the reaction performed with the L1-specific primer, the 3' end of the L1 should be visible in addition to the poly-A tail (Fig. 2c) (*see* Notes 7 and 8).
5. To find the 5' end of the insertion, several different methods can be utilized. Because many new L1 insertions are truncated on the 5' end of the element, it is frequently possible to detect the 5' end of the element by using the reverse complement of the L1SP1A2 primer (L1 GTG primer) with the ES primer. To do this, make the master mix as follows: mix (1×): 12.5 µL of Promega GoTaq green (2×) master mix, 0.8 µL of 20 µM ES primer, 1.6 µL of 20 µM L1 GTG primer, 2 µL of genomic DNA (12.5 ng/µL), 8.2 µL of DNA-free H₂O. For insertions with a longer 5' end present, this PCR will likely fail; however, it is possible to tile across the L1 element with various primers at different locations (e.g. L1nt112out) in the element accompanied by the empty site primer to find the 5' end. For this PCR, use the following master mix (1×): 12.5 µL of Promega GoTaq green (2×) master mix, 0.8 µL of 20 µM ES primer, 1.6 µL of 20 µM L1 internal primer, 2 µL of genomic DNA (12.5 ng/µL), 8.2 µL of DNA-free H₂O.

5' L1 GTG PCR Parameters

- a. 95 °C for 2 min.
- b. 95 °C for 30 s.
- c. 57.5 °C for 1 min and 30 s.
- d. 72 °C for 3 min.
- e. Go to **step b** (29×).
- f. 72 °C for 5 min.
- g. 4 °C Hold.

insertions which may have been easily detectable with next-generation sequencing and not with Sanger sequencing. You can nest both the filled site primers as well as the L1Ta-specific primers to increase the specificity of the reaction greatly. Nested PCR along with an increase in cycle numbers and/or altering the melting temperature of the PCR often alleviates validation PCR issues.

⁸With regard to choosing predicted insertions for validation, one of two main methods may be employed. A random number generator can be used to select putative somatic insertions for validation which will potentially give a good estimate of the number of true somatic insertions in the data set. Alternatively, putative somatic insertions with unique read counts above 5, map scores of 1, and alignment windows of at least 100 base pairs can be selected for validation. Depending on the validation rate with the primary insertions selected, the level of stringency can be altered until the ideal validation rate is achieved. A validation rate above 60 % is generally acceptable for this technique; however, PCR optimization, good primer design, and good DNA are key to successful validations.

5' L1 Internal Primer (e.g. L1nt112out) PCR Parameters

- a. 95 °C for 2 min.
- b. 95 °C for 30 s.
- c. 57 °C for 30 s.
- d. 72 °C for 45 s.
- e. Go to **step b** (29×).
- f. 72 °C for 5 min.
- g. 4 °C Hold.

Acknowledgments

This work was funded by a P-50 grant awarded to H.H.K. Jr.

References

1. Voytas DF, Cummings MP, Konieczny A, Ausubel FM, Roderick SR. *copia*-like retrotransposons are ubiquitous among plants. *Proc Natl Acad Sci*. 1992; 89(15):7124–7128. [PubMed: 1379734]
2. Hancks DC, Kazazian HH Jr. Active human retrotransposons: variation and disease. *Curr Opin Genet Dev*. 2012; 22(3):191–203. [PubMed: 22406018]
3. Dewannieux M, Esnault C, Heidmann T. LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet*. 2003; 35(1):41–48. [PubMed: 12897783]
4. Ostertag EM, Goodier JL, Zhang Y, Kazazian HH Jr. SVA elements are nonautonomous retrotransposons that cause disease in humans. *Am J Hum Genet*. 2003; 73(6):1444–1451. [PubMed: 14628287]
5. Esnault C, Maestre J, Heidmann T. Human LINE retrotransposons generate processed pseudogenes. *Nat Genet*. 2000; 24(4):363–367. [PubMed: 10742098]
6. Ewing AD, Kazazian HH Jr. High-throughput sequencing reveals extensive variation in human specific L1 content in individual human genomes. *Genome Res*. 2010; 20(9):1262–1270. [PubMed: 20488934]
7. Solyom S, Ewing AD, Rahrmann EP, Doucet T, Nelson HH, Burns MB, Harris RS, Sigmon DF, Casella A, Erlanger B, Wheelan S, Upton KR, Shukla R, Faulkner GJ, Largaespada DA, Kazazian HH Jr. Extensive somatic L1 retrotransposition in colorectal tumors. *Genome Res*. 2012; 22(12):2328–2338. [PubMed: 22968929]
8. Lee E, Iskow R, Yang L, Gokcumen O, Haseley P, Luquette LJ, Lohr JG, Harris CC, Ding L, Wilson RK, Wheeler DA, Gibbs RA, Kucherlapati R, Lee C, Kharchenko PV, Park PJ. Landscape of somatic retrotransposition in human cancers. *Science*. 2012; 337(6097):967–971. [PubMed: 22745252]
9. Shukla R, Upton KR, Muñoz-Lopez M, Gerhardt DJ, Fisher ME, Nguyen T, Brennan PM, Baillie JK, Collino A, Ghisletti S, Sinha S, Iannelli F, Radaelli E, Dos Santos A, Rapoud D, Guettier C, Samuel D, Natoli G, Carninci P, Ciccarelli FD, Garcia-Perez JL, Faivre J, Faulkner GJ. Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. *Cell*. 2013; 153(1):101–111. [PubMed: 23540693]
10. Helman E, Lawrence MS, Stewart C, Sougnez C, Getz G, Meyerson M. Somatic retrotransposition in human cancer revealed by whole-genome and exome sequencing. *Genome Res*. 2014; 24(7):1053–1063. [PubMed: 24823667]
11. Baillie JK, Barnett MW, Upton KR, Gerhardt DJ, Richmond TA, De Sapio F, Brennan PM, Rizzu P, Smith S, Fell M, Talbot RT, Gustincich S, Freeman TC, Mattick JS, Hume DA, Heutink P, Carninci P, Jeddeloh JA, Faulkner GJ. Somatic retrotransposition alters the genetic landscape of the human brain. *Nature*. 2011; 479(7374):534–537. [PubMed: 22037309]

12. Evrony GD, Cai X, Lee E, Hills LB, Elhosary PC, Lehmann HS, Parker JJ, Atabay KD, Gilmore EC, Poduri A, Park PJ, Walsh CA. Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. *Cell*. 2012; 151(3):483–496. [PubMed: 23101622]
13. Tubio JM, Li Y, Ju YS, Martincorena I, et al. Mobile DNA in cancer. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science*. 2014; 345(6196):1251343. [PubMed: 25082706]
14. Mir AA, Philippe C, Cristofari G. euL1db: the European database of L1HS retrotransposon insertions in humans. *Nucleic Acids Res*. 2014; 43(Database issue):D43–D47. [PubMed: 25352549]
15. Doucet-O'Hare T, Rodic N, Sharma R, Darbari I, Abril G, Choi JA, Young Ahn J, Cheng Y, Anders RA, Burns KH, Meltzer SJ, Kazazian HH Jr. LINE-1 expression and retrotransposition in Barrett's esophagus and esophageal carcinoma. *Proc. Natl. Acad Sci*. 2015; 112(35):4894–4900.

Fig. 1.

Scheme of L1-seq. The first PCR consists of five cycles which enrich for sequences containing human-specific L1 sequences via primer extension with the L1Sp1A2 primer using diagnostic nucleotides for the human specific subfamily of L1. After enrichment for human-specific L1 flanks in the first five cycles, each reaction has one degenerate primer added. There are eight degenerate primers, each with a specified 5-mer at the 3' end and preceded by five degenerate bases (NNNNN) and a sequencing primer used for the Illumina platform. Eight different reactions are performed, each with a unique 5mer. The second PCR enriches for human-specific L1 3' flanks by utilizing another diagnostic nucleotide in the 3' UTR of L1 and adds the necessary adapter sequences via primer overhangs. The products of this reaction are mixed in equimolar ratios before sequencing on the Illumina 2500 platform. Following sequencing and initial processing, the reads are aligned to the human genome (hg19) and L1 insertions are identified for validation

Fig. 2.

Validation of results. **(a)** Diagram of the PCR validation scheme for putative insertions, the 3' end of the LINE-1 insertion is pictured adjacent to a poly A tail. The nested empty site and filled site primers are flanking the empty and filled site primers. In a nested PCR, the nested primers are used in the first of two reactions. One and a half μ L of product from the first reaction (with ES and FS primers) are used as template in a second PCR with the nested primers to amplify difficult or rare products. **(b)** Two examples of validations for insertions present in only tumor and absent from normal DNA. On the left, a PCR result depicting both the empty site (ES) and filled site (FS) products for both the normal and tumor DNA samples from patients. Only in the tumor of patient 11 is a filled site band present confirming the insertion is present. In the image on the right side on **b**, a PCR depicting another validation of a somatic insertion present in BE and absent from normal esophageal and white blood cell DNA. There is only a band present in the BE sample for the FS PCR; however, the ES PCR has bands for all three DNA samples as a positive control. **(c)** An insertion sequence with the unique genomic DNA (*blue*), target site duplications (*purple*), LINE-1 sequence (*red*), and the poly-A tail sequence (*orange*) [15]

Table 1

Primers used for L1-seq

Sequences 5' to 3'	
<i>Primers with Illumina adapters:</i>	
Adap 1 L 1HsG	CAAGCAGAAGACGGCATAOGAGCTCTTCCGATC TTGCACATGAOCCTAAAACTTAG
Adap2Seq1	AATGATACGGCGACCACCGAGATCTACACTTTOCC TACACGACGACGCTCTTCCGATCT
<i>L1 specific primers:</i>	
L1HsSP1A2	GGGAGATATACCTAATGCTAGATGACAC (specific for L 1Hs subset)
L1 "G" primer	TGCACATGTACOCTAAAACTTAG (specific for L 1Hs subset)
L1nt112out	GATGAACCCGTACCTCAGA
<i>Degenerate primers:</i>	
DEG Seq 1N5TCTGT	ACACTCTTTCCCTACACGACGACGCTCTTCCGA TCTNNNNNTCTGT
DEGSeq1N5CTTCT	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNN NNNCTTCT
DEGSeq1N5TGCCT	ACACTCTTTCCCTACACGACGACGCTCTTCCGAT CTNNNNNTGCCT
DEGSeq1NTCTCA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTN NNNTCTCA
DEGSeq1N5CAGAG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNN NNNCAGAG
DEGSeq1N5TTGAA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNN NNNTTGAA
DEGSeq1N5CTTTG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNN NNNCTTTG