Methods in Molecular Biology
DOI 10.1007/7651_2015_265

Library Construction for High-Throughput Mobile Element Identification and Genotyping

Hongseok Ha, Nan Wang, and Jinchuan Xing

Abstract

Mobile genetic elements are discrete DNA elements that can move around and copy themselves in a genome. As a ubiquitous component of the genome, mobile elements contribute to both genetic and epigenetic variation. Therefore, it is important to determine the genome-wide distribution of mobile elements. Here we present a targeted high-throughput sequencing protocol called Mobile Element Scanning (ME-Scan) for genome-wide mobile element detection. We will describe oligonucleotides design, sequencing library construction, and computational analysis for the ME-Scan protocol.

Keywords: Mobile element, ME-Scan, High-throughput sequencing, Population diversity, Polymorphism

1 Introduction

Mobile elements (MEs) are a major component of the human genome. As a consequence of their transposition and accumulation, roughly two-thirds of the human genome comprises MEs [1]. Based on the transposition mechanism, MEs can be divided into two classes. Class I elements, also known as retrotransposons, use a "copy-and-paste" mechanism. During a process called retrotransposition, class I elements create new copies of themselves at different genomic locations via RNA intermediates. Class II elements, also known as DNA transposons, use a "cut-and-paste" mechanism and mobilize a DNA element from one genomic location to another. DNA transposons have been inactive over the past 30 million years in the primate lineage, while retrotransposons remain active in all primate genomes studied to date [2]. Retrotransposons are further subdivided into long terminal repeat (LTR) and non-LTR classes. Long interspersed element-1 (LINE-1, or L1) is a representative of non-LTR retrotransposon and encodes proteins necessary for autonomous retrotransposition [3]. Alu and SVA (SINE/variable number of tandem repeat (VNTR)/Alu) are non-autonomous elements that do not encode functional mobilization

proteins by themselves. They rely on the enzymatic machinery of an L1 element to retrotranspose to other genomic locations [4–6].

MEs play a key role in genome evolution, creating structural variation both by generating new insertions and by promoting nonhomologous recombination [7, 8]. Mobile element insertions (MEIs) also shape gene regulatory networks by supplying and/or disrupting functional elements such as transcription factor binding sites, transcription enhancers, alternative splicing sites, nucleosome positioning signals, methylation signals, and chromatin boundaries [9, 10]. Some ME-derived or -targeted small RNAs, such as miRNAs and piRNAs, also affect transcriptional regulation in the host genome [11, 12]. Therefore, it is important to determine the genomic locations of MEIs.

Because of their ability to transpose in the genome, MEs have also been used extensively in genome engineering. For example, transposon systems *sleeping beauty* and *piggyBac* have been used for mutagenesis and nonviral gene delivery [13, 14]. Once new transposons are integrated in the genome, it is necessary to determine their genomic locations. An efficient, high-throughput method is crucial to identify the insertion sites.

Before the high-throughput sequencing technology became available, transposon display methods were used to identify polymorphic MEI loci [15]. Transposon display methods identify the junction of an ME and its upstream or downstream flanking genomic sequence. Usually a primer specific to the ME of interest and either a random primer or a primer specific to a generic linker sequence are used to amplify the ME/genomic junction site. Once candidate MEI loci are identified, locus-specific PCRs are used to determine the MEI genotypes in individual samples (e.g., [16]). Recently, a number of efforts have been made to identify polymorphic MEIs using high-throughput sequencing technology (Reviewed in refs. [17, 18]). Although high-coverage whole genome sequencing is suitable for studying MEIs in different species, the cost is still too high for large-scale population-level studies. On the other hand, low coverage strategy such as the one adopted by the 1000 Genomes Project [19] is not ideal because it is likely to under-sample polymorphic MEIs. Mobile element scanning (ME-Scan) protocol adapts the transposon-display concept to the high-throughput sequencing platform and provides both high sensitivity and high specificity for MEI detection [20, 21]. Because the resulting sequencing library contains only DNA fragments at the MEI-genomic junction sites, it is a cost-effective way to identify MEIs for both large-scale genomic studies and transposon-based mutagenesis studies. Here we describe the ME-Scan protocol in detail. Although we use AluYb and L1HS family of MEs in human as examples to illustrate the ME-Scan application, the protocol can be easily modified for other MEs in other species by changing the ME-specific primers to the ME of interest.

2 Materials

2.1 Reagents

2.1.1 Oligonucleotides (Adaptors, Primers)

The adaptor and primer sequences used for human AluYb and L1HS ME-Scan protocol are shown in Table 1. To capture ME-specific fragments, two PCR amplification steps are required. Table 1 show oligonucleotides used for both PCRs. The first round ME-specific primers include 5′ biotinylation modification for bead capture and all primers include a phosphorothioate bond at the 3′ end for stability. In addition, current Illumina sequencing technology requires near random representation of all four nucleotides in the first three sequencing cycles to establish baseline signals and positions for base calls. Therefore, we incorporated three random bases within the second amplification primers.

For studies involving multiple samples, Illumina provides 6 bp index sequences for pooling multiple samples in one sequencing library. We tested 48 indexes and these index sequences have good uniformity and show no systematic biases. Therefore, we designed our customized linker sequences using the Illumina index sequences (Table 1).

2.1.2 Enzymes and Buffer Solutions

Several commercial kits were used in the protocol. For example, for sequencing library construction, we used KAPA Library Preparation Kit with SPRI solution for Illumina (KAPA Biosystems, Wilmington, MA, USA, cat. no KK8232). Other comparable reagents can be used as substitutes.

1. 1× TE buffer: 10 mM Tris (pH 8.0), 1 mM EDTA
2. KAPA Library Preparation Kit with SPRI solution for Illumina (KAPA Biosystems, cat. no KK8232)
3. Streptavidin-coupled Dynabeads magnetic beads (Life Technologies, Grand Island, NY, USA, cat. no 65305)
4. Agencourt AMPure XP beads (Beckman coulter, Indianapolis, IN, USA, cat. no A63880)
5. 2× B&W Buffer: 10 mM Tris–HCl (pH 7.5), 1 mM EDTA, 2 M NaCl
6. Agarose Gel: NuSieve GTG (Lonza, Cologne, Germany, cat. no 50084) and GeneMate LE (BioExpress, Kaysville, UT, USA, cat. no E-3120-500) (3:1)
7. 1× TBE buffer
8. 100 bp DNA ladder (New England Biolabs, Ipswich, MA, USA, cat. no N3231S)
9. Wizard SV Gel Clean-Up System (Promega, Madison, WI, USA, cat. no A9281)

Table 1
Oligonucleotides for ME-Scan protocol for human AluYb and L1HS MEs

Library	Description	Sequences (5′ → 3′)
L1HS L1HS L1HS	Biotinylated L1-specific primer cocktail for first amplification	/5Biosg/GGGAGATATACCTAATGCTAGATGACAC*A /5Biosg/GGGAGATATACCTAATGCTAGATGACAC*G /5Biosg/GGGAGATATACCTAATGCTAGATGACAA*G
L1HS	L1HS-specific primer for second amplification	AATGATACGGCGACCACCGlAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNN GGGAGATATACCTAATGCTAGATGAC*A
AluYb	Biotinylated AluYb-specific primer for first amplification	/5Biosg/CAGGCCGGACTGCGGA*C
AluYb	AluYb-specific primer cocktail for second amplification	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTNNN AGTGCTGGGATTACAGGCGTG*A
Common	Typical Illumina adaptor pair including P7 region and individual index	CAAGCAGAAGACGGCATACGAGAT**CGTGAT**GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T AGATCGGAAGAGCGTCGTG
Common	P7 adaptor-specific primer	CAAGCAGAAGACGGCATACGAGA*T

/5Biosg/: 5′ Biotin; *: 3′ Phosphorothioate bond
Underlined sequences indicate random sequences; *bold letters* indicate one example of Illumina index sequence

10. KAPA Library Quantification Kit for Illumina (KAPA Biosystems, cat. no KK4824)
11. Zero Blunt TOPO PCR Cloning Kit (Life Technologies, Grand Island, NY, USA, cat. no K270020).

2.2 Equipment

1. Heat block (Corning, Corning, NY, USA)
2. Covaris system with Crimp-Cap Micro-Tube (Covaris, Woburn, MA, USA)
3. NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA)
4. Magnetic stand (Promega, Madison, WI, USA, cat. no Z5342) or 96 well micro plate magnetic separation rack (New England Biolabs, cat. no S1511S)
5. Vortex mixer (Scientific Industries, Bohemia, NY, USA)
6. Thermal cycler PCR machine (Bio-Rad Laboratories, Hercules, CA, USA)
7. Gel electrophoresis system (Bio-Rad Laboratories)
8. Real-time PCR machine (Bio-Rad Laboratories)
9. High-throughput sequencer (Hiseq 2500, Miseq (Illumina, San Diego, CA, USA) and PACBIO RS (Pacific Biosciences, Menlo Park, CA, USA) were tested)
10. Water bath (Precision/Thermo Fisher Scientific, Waltham, MA, USA)

3 Methods

Procedures of the ME-Scan protocol are illustrated in Fig. 1. First, genomic DNA is randomly fragmented to ~1 kb in size (Fig. 1a). The DNA fragments are then end-repaired (Fig. 1b), A-tailed (Fig. 1c), and ligated to adaptors on both ends (Fig. 1d). DNA fragments containing ME-genomic junction are then amplified from the whole-genome library using ME-specific PCR (Fig. 1e). The amplified, biotinylated DNA fragments are enriched by streptavidin beads (Fig. 1f) and further amplified (Fig. 1g) into the final sequencing library. After the quality assessment (Fig. 1h), the library is sequenced (Fig. 1i). Below we describe each step in detail.

3.1 Preparation of Double-Strand DNA Adaptor

1. Mix equal volumes of paired oligonucleotides (100 μM). A pair of typical Illumina adaptors is shown in Table 1.
2. Incubate in a heat block for 5 min at 95 °C.
3. With tubes still in the heat block, turn off the heat block and allow tubes to cool to room temperature.
4. Store at 4 °C.

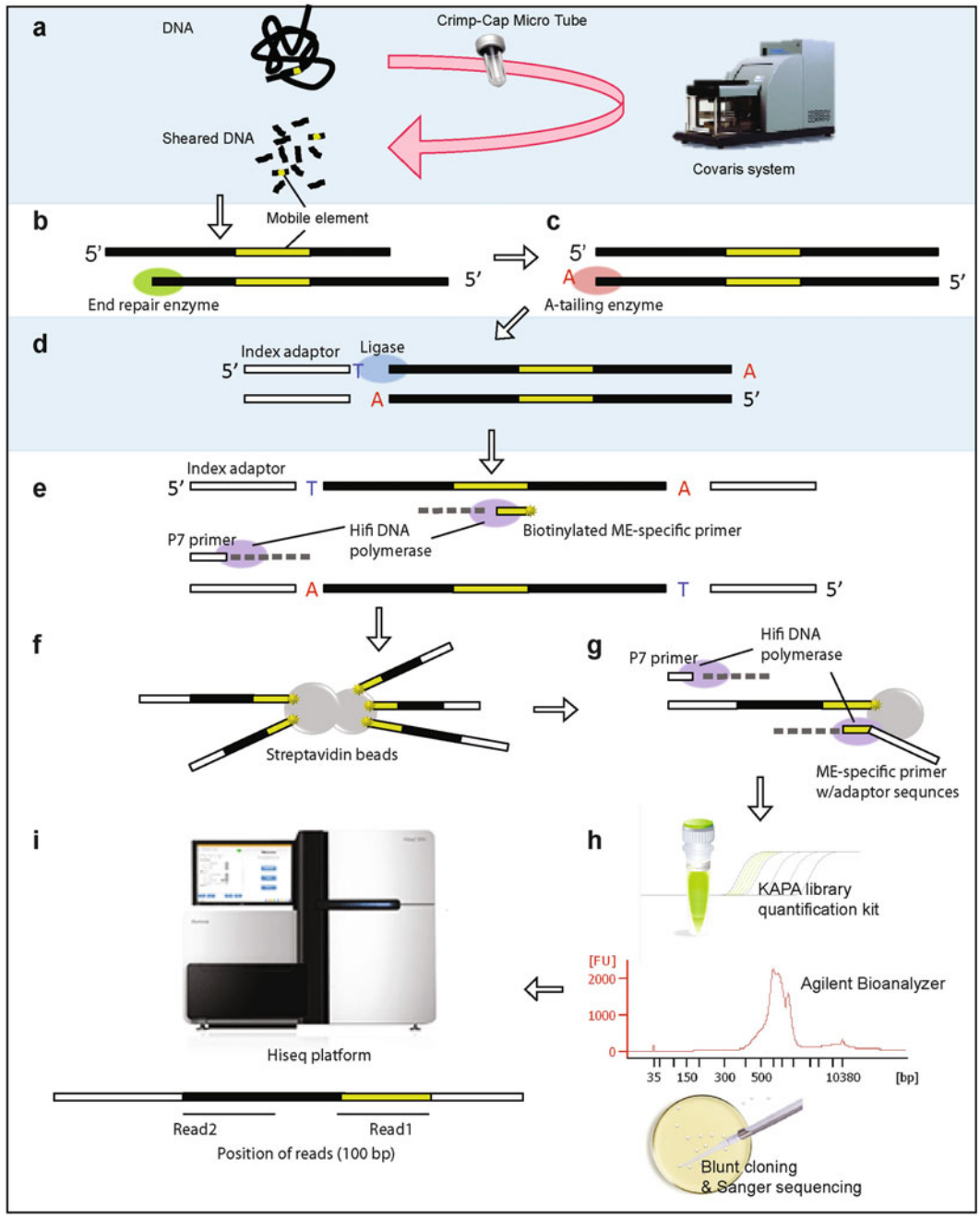

Fig. 1 ME-Scan library construction procedure. (**a**) DNA fragmentation; (**b**) end repair; (**c**) A-tailing; (**d**) adaptor ligation; (**e**) first PCR amplification; (**f**) beads capture; (**g**) second PCR amplification; (**h**) library validation; (**i**) high-throughput sequencing

3.2 Genomic DNA Fragmentation

1. Prepare 1–10 μg genomic DNA in 120 μL TE buffer.
2. Targeted fragment length is around 1,000 bp, and the operating conditions for the Covaris system are: Duty Cycle—5 %, Intensity—3, Cycle per Burst—200, Time—15 s.

3.3 ME-Scan Library Construction

1. Ensure that the AMPure XP Beads are equilibrated to room temperature, and thoroughly resuspended.

3.3.1 Concentrate DNA Samples

2. Mix 120 μl DNA fragments in TE buffer and 120 μl AMPure XP Beads per tube/well. For small sample size, mix in tubes; for large sample size, mix in 96-well plates. Because the total volume is more than 200 μl, use a microtiter plate (250 μl working volume) instead of a standard PCR plate for this step.
3. Mix thoroughly on a vortex mixer or by pipetting up and down at least ten times.
4. Incubate at room temperature for 5 min to allow DNA to bind to the beads.
5. Capture the beads by placing the tube/microtiter plate on an appropriate magnetic stand at room temperature for 10 min or until the liquid is completely clear.
6. If working with the microtiter plate, carefully remove and discard 120 μl supernatant (half of the total volume) per well. Do not disturb or discard any of the beads. If working with the tube, go directly to step 9.
7. Remove the microtiter plate from the magnetic stand, mix well and transfer the samples from the microtiter plate to a PCR plate (multichannel pipette can be used when processing multiple samples).
8. Capture the beads by placing the PCR plate on an appropriate magnetic stand at room temperature for 10 min or until the liquid is completely clear.
9. Carefully remove and discard the supernatant. Do not disturb or discard the beads. Some liquid may remain visible in the tube/well.
10. Remove the PCR plate from the magnetic stand, add 50 μl double-distilled water, and incubate at room temperature for 5–10 min to recover the DNA fragments.

3.3.2 End Repair Reaction

1. Assemble the end repair reaction in the PCR plate containing DNA fragments and AMPure XP Beads. For each well, add 20 μl End Repair Mix (8 μl water, 7 μl 10× KAPA End Repair Buffer, 5 μl KAPA End Repair Enzyme). For multiple library construction, master mix can be made for the End Repair Mix based on the number of libraries to improve the consistency. When making a master mix, add 1 or 2 more reaction volumes

to ensure sufficient volume. The same principle applies for making other master mixes in this protocol.

2. Mix each reaction thoroughly on a vortex mixer or by pipetting up and down, and incubate the plate at 20 °C for 30 min.

3.3.3 End Repair Cleanup

1. To each 70 μl end repair reaction, add 120 μl PEG/NaCl SPRI Solution.
2. Mix thoroughly by pipetting up and down multiple times and/or by vortexing.
3. Incubate the plate at room temperature for 15 min, allowing the DNA to bind to the beads.
4. Place the plate on a magnetic stand at room temperature to capture the beads for 10 min or until the liquid is completely clear.
5. Remove and discard the supernatant.
6. While keeping the plate on the magnetic stand, add 200 μl of 80 % ethanol.
7. Incubate the plate at room temperature for 30 s to 1 min.
8. Remove and discard the ethanol.
9. Repeat the wash (steps 6–8).
10. Allow the beads to dry sufficiently for 5 min at room temperature and ensure that all the ethanol has evaporated.

3.3.4 A-Tailing Reaction

1. To each well containing the dried beads and end repaired DNA, add: 50 μl A-Tailing Master Mix (42 μl water, 5 μl 10× KAPA A-Tailing Master Buffer, 3 μl KAPA A-Tailing Enzyme).
2. Mix thoroughly by pipetting up and down multiple times, or by vortexing, to resuspend the beads.
3. Incubate the plate at 30 °C for 30 min.

3.3.5 A-Tailing Cleanup

1. To each well containing the 50 μl A-tailing reaction with beads, add 90 μl PEG/NaCl SPRI Solution.
2. Capture beads and perform cleanup as described in Section 3.3.3.
3. Remove the PCR plate from the magnetic stand, add 32 μl double-distilled water and incubate at room temperature for 5–10 min to recover the DNA fragments.

3.3.6 Calculate the Amount of Pre-annealed Adaptor Needed for Each Sample

1. Quantify the DNA concentration with 2 μl of each sample using the NanoDrop (As a quality control, the 260/280 ratio should be >2).
2. In ligation reactions, the molarity of sample (M_s) can be calculated using the following equation:

$$M_s = \frac{\text{Sample concentration(ng/μl)} \times 1{,}000{,}000 \times 10\ \mu\text{l}}{1000\ \text{bp} \times 650\ \text{Da} \times 50\ \mu\text{l}}$$

Then, the volume (in μl) of adaptor (10 μM) used in ligation should be:

$$\text{Volum of adaptor (μl)} = \frac{M_s \times 10 \times 50\ \mu\text{l}}{10\ \mu\text{M} \times 1000}$$

3.3.7 Adaptor Ligation Reaction

1. To each well containing 30 μl A-tailed product, add 15 μl Ligation Master Mix (10 μl 5× KAPA Ligation Buffer, 5 μl KAPA T4 DNA Ligase, supplied by the library preparation kit) and 5 μl adaptor (use the volume of adaptor determined in Section 3.3.6 and water for the remaining volume).
2. Mix thoroughly to resuspend the beads.
3. Incubate the plate at 20 °C for 15 min.

3.3.8 Adapter Ligation Cleanup

1. To each 50 μl ligation reaction with beads, add 50 μl PEG/NaCl SPRI Solution.
2. Capture beads and perform cleanup as described in Section 3.3.3.
3. Remove the PCR plate from the magnetic stand, add 50 μl double-distilled water and incubate at room temperature for 5–10 min to recover the DNA fragments.
4. Place the plate on a magnetic stand to capture the beads until the liquid is clear. Transfer the supernatant containing ligation product to a new plate. Discard the beads.

3.3.9 First PCR Amplification

Measure DNA concentration of each individual sample using NanoDrop. Normalize the sample concentration based on the NanoDrop quantification result and pool up to 48 individual samples with different index sequences together in one single tube with equal amount.

1. Set up PCR reactions according to Table 2.
2. Perform PCR reactions using the following conditions: initial denaturation for 45 s at 98 °C followed by 5–10 cycles of 98 °C for 15 s, anneal at 65 °C for 30 s, extension at 72 °C for 30 s followed by a final extension at 72 °C for 1 min.

3.3.10 ME-Containing Fragments Pull Down by Streptavidin Beads

Preparation

1. Dilute 2× B&W Buffer to 1× B&W Buffer with distilled water.
2. Calculate the amount of beads required based on their binding capacity [1 mg (100 μl) Dynabeads magnetic beads binds 10 μg double-stranded DNA].
3. Prepare appropriate amount of Dynabeads magnetic beads following the manufacturer's instructions.

Table 2
Pre-mix for PCR reaction

	For first amplification		For second amplification	
Component	**Working concentration**	**Volume**	**Working concentration**	**Volume**
PCR grade water		As needed		17 µl
2× KAPA HiFi HS RM	1×	25 µl	1×	25 µl
Adapter primer (P7)	10 µM	2.5 µl	10 µM	2.5 µl
(Biotinylated-) ME-specific primer (refer Table 1)	10 µM	2.5 µl	10 µM	2.5 µl
DNA		As needed		3 µl[a]
Total		50 µl		50 µl

[a]The template DNA solution contains the DNA fragments captured on the streptavidin beads

Immobilization of Nucleic Acids

1. Resuspend beads in 30 µl 2× B&W Buffer.
2. To immobilize DNA fragments, add an equal volume of the biotinylated DNA in H_2O to dilute the NaCl concentration in the 2× B&W Buffer from 2 M to 1 M for optimal binding.
3. Incubate for 15 min at room temperature using gentle rotation. Incubation time depends on the nucleic acid length: DNA fragments up to 1 kb require 15 min.
4. Separate the biotinylated DNA coated beads with a magnetic stand for 2–3 min or until the liquid is clear. Remove supernatant using a pipette while the tube is on the magnetic stand.
5. While keeping the tube on the magnetic stand, add 30 µl 1× B&W Buffer.
6. Incubate the tube at room temperature for 30 s to 1 min.
7. Remove and discard the B&W Buffer.
8. Repeat steps 5–7 twice, for a total of three washes.
9. Remove the tube from the magnetic stand and resuspend beads in 24 µl double-distilled water.

3.3.11 Second PCR Amplification

1. Set up PCR reactions according to Table 2.
2. Perform PCR reactions using the following conditions: initial denaturation for 45 s at 98 °C followed by at most 25 cycles of 98 °C for 15 s, anneal at 65 °C for 30 s, extension at 72 °C for 30 s followed by a final extension at 72 °C for 1 min.

3.3.12 Size Selection and Gel Extraction

1. Prepare a 2 % agarose gel using 3 quarters of NuSieve GTG and 1 quarter of GeneMate LE agarose.
2. Run the gel at 100 V for 55 min.

Table 3
The size of different components of the DNA fragments in a completed ME-Scan library

Parts	Size	Remarks
Read2 indexed adaptor	65 bp	The size of an index is 6 bp.
Read1 adaptor	58 bp	
Random sequences	3 ~ 5 bp	At least 3 bp random sequences at the beginning of Read 1 are required by current Illumina sequencing technology.
ME fragment	e.g., 123 bp for L1HS	The region from the ME-specific primer to the boundary of an ME.
Variable region	Variable length	The experimenter should consider variable sized regions such as a poly(A) tail at the 3′ end of an ME.
Genomic Flanking region	>20 bp	The genomic region should be large enough (e.g., >20 bp) to ensure the resulted sequencing reads can be mapped to the reference genome with high confidence.

3. Based on comparison to a DNA ladder, cut out the gel slice of the required size and place the gel slice in a 1.5 ml micro-centrifuge tube. The required library size depends on the ME of interest and the sequencing platform. Refer to Table 3 for a size calculation example.
4. Extract DNA fragments from the gel slice using Wizard SV Gel Clean-Up System (Promega) following the manufacturer instruction. Elute DNA in 30 μl of elution buffer.

3.4 Library Validation and Sequencing

3.4.1 Validation of ME-Scan Library

1. Using Agilent Bioanalyzer, or similar technology, confirm the size distribution of the completed library. An example of the library size calculation is shown in Table 3.
2. Quantify the concentration of DNA fragments that can be sequenced by quantitative PCR using sequencing-specific primers (e.g., KAPA Library Quantification Kit). In general, the library should have a concentration of 10 nM or higher.
3. To validate the sequencing library, clone the library using a blunt-end cloning kit (e.g., Zero Blunt TOPO PCR Cloning Kits). Sequence a number of colonies to validate the DNA fragments within the library. Examine the DNA fragments in the library to ensure the presence of the proper library structure (e.g., sequencing primer binding sites, index) and the targeted ME sequences. We suggest that at least 24 colonies should be sequenced when a new ME-specific primer is used.

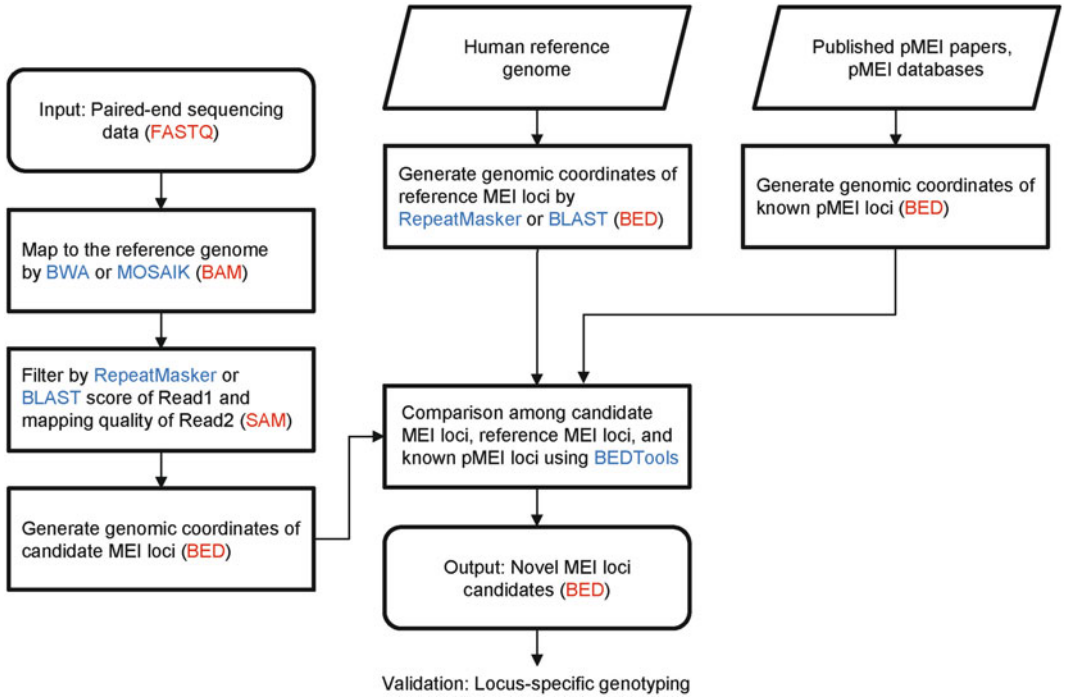

Fig. 2 Computational workflow for ME-Scan analysis. File format is shown in *red*, program name is shown in *blue*

3.4.2 High-Throughput Sequencing

Sequence the library on an Illumina HiSeq 2000/2500 platform using pair-end 100 base-pair format.

3.4.3 Analysis Pipeline

Figure 2 shows a flowchart of the analysis pipeline. First, raw sequencing reads were aligned to the reference genome using aligner such as BWA [22] or MOSAIK [23]. Pair-end reads that can be mapped to the genome were then filtered by two criteria: Read1 (containing targeted MEI sequence) is filtered using RepeatMasker [24] or BLAST [25] programs to ensure the presence of the expected MEI sequence; Read2 (genomic flanking sequences of MEIs) in each pair is filtered based on its mapping quality to ensure the unique mapping of the read-pair. Read pairs that failed either of the filters will be excluded from further analyses. After the filtering steps, the candidate loci are compared with known MEIs in the reference genome and known polymorphic MEI loci in previous studies and databases (e.g., [8, 19, 20, 26–31]) to identify novel polymorphic MEI loci.

4 Notes

1. When testing the protocol on a new type of ME, PCR-based locus-by-locus validation is strongly recommended to assess the sensitivity/specificity of the ME-specific primer.

2. Because PCR amplification is initiated from randomly sheared DNA fragments, a smear will be generated during the size selection step. Cutting a thin slice of gel (e.g., ~ 1 mm) can help to control the size distribution of the DNA fragments for downstream analysis. Also, the amount of DNA loaded for size selection should be carefully controlled. Overloading the gel could interfere with size separation of the DNA pool. Alternatively, if the size distribution of the final library is in a wide range, an additional size-selection step can be added after the first round PCR amplification (Section 3.3.9) to further improve specificity.
3. There are two types of bead-captures in the protocol for different purposes. Among the sections, different components (e.g., beads or the supernatant) were kept. The experimenter should pay close attention to these sections to make sure the correct component is kept.
4. We use the in-solution protocol from KAPA to improve the yield and reduce the cost for library construction [32]. In this protocol, AMPure XP Beads are kept in every step without replacement until the adaptor ligation step.
5. ME-specific primers should be reverse-complementary to a target region that is highly conserved in the ME consensus and close to the ME-genomic junction. If both ME's junctions (5′- or 3′-) are available, select the less variable junction is preferred (e.g., not attempting to capture the junction associated with the poly(A) tail at the 3′ end of many MEs). Degenerate primers can be used if there are subtype mutations in targeted ME (refer to L1HS primers in Table 1 for an example). The ME-specific primer (non-biotinylated) for the second amplification can be designed in the internal region of the first amplicon (i.e., nested PCR) to improve the specificity of the protocol.

Acknowledgement

The authors declare no competing financial interests. We thank Drs. David Ray and Roy Platt for their valuable comments. This study was supported by grants from the National Institutes of Health (R00HG005846).

References

1. de Koning AP, Gu W, Castoe TA, Batzer MA, Pollock DD (2011) Repetitive elements may comprise over two-thirds of the human genome. PLoS Genet 7(12), e1002384. doi:10.1371/journal.pgen.1002384
2. Pace JK II, Feschotte C (2007) The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. Genome Res 17(4):422–432. doi:10.1101/gr.5826307

3. Ostertag EM, Kazazian HH Jr (2001) Biology of mammalian L1 retrotransposons. Annu Rev Genet 35:501–538
4. Dewannieux M, Esnault C, Heidmann T (2003) LINE-mediated retrotransposition of marked Alu sequences. Nat Genet 35(1):41–48
5. Hancks DC, Goodier JL, Mandal PK, Cheung LE, Kazazian HH Jr (2011) Retrotransposition of marked SVA elements by human L1s in cultured cells. Hum Mol Genet 20 (17):3386–3400. doi:10.1093/hmg/ddr245
6. Raiz J, Damert A, Chira S, Held U, Klawitter S, Hamdorf M, Lower J, Stratling WH, Lower R, Schumann GG (2012) The non-autonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery. Nucleic Acids Res 40(4):1666–1683. doi:10.1093/nar/gkr863
7. Burns KH, Boeke JD (2012) Human transposon tectonics. Cell 149(4):740–752. doi:10.1016/j.cell.2012.04.019
8. Xing J, Zhang Y, Han K, Salem AH, Sen SK, Huff CD, Zhou Q, Kirkness EF, Levy S, Batzer MA, Jorde LB (2009) Mobile elements create structural variation: analysis of a complete human genome. Genome Res 19 (9):1516–1526. doi:10.1101/gr.091827.109
9. Ichiyanagi K (2013) Epigenetic regulation of transcription and possible functions of mammalian short interspersed elements, SINEs. Genes Genet Syst 88(1):19–29
10. Cowley M, Oakey RJ (2013) Transposable elements re-wire and fine-tune the transcriptome. PLoS Genet 9(1), e1003234. doi:10.1371/journal.pgen.1003234
11. Piriyapongsa J, Marino-Ramirez L, Jordan IK (2007) Origin and evolution of human microRNAs from transposable elements. Genetics 176(2):1323–1337. doi:10.1534/genetics.107.072553
12. Rouget C, Papin C, Boureux A, Meunier AC, Franco B, Robine N, Lai EC, Pelisson A, Simonelig M (2010) Maternal mRNA deadenylation and decay by the piRNA pathway in the early Drosophila embryo. Nature 467(7319):1128–1132. doi:10.1038/nature09465
13. Wilson MH, Coates CJ, George AL Jr (2007) PiggyBac transposon-mediated gene transfer in human cells. Mol Ther 15(1):139–145. doi:10.1038/sj.mt.6300028
14. Mann MB, Jenkins NA, Copeland NG, Mann KM (2013) Sleeping Beauty mutagenesis: exploiting forward genetic screens for cancer gene discovery. Curr Opin Genet Dev 24:16–22. doi:10.1016/j.gde.2013.11.004
15. Van den Broeck D, Maes T, Sauer M, Zethof J, De Keukeleire P, D'Hauw M, Van Montagu M, Gerats T (1998) Transposon display identifies individual transposable elements in high copy number lines. Plant J 13(1):121–129. doi:10.1046/j.1365-313X.1998.00004.x
16. Xing J, Wang H, Han K, Ray DA, Huang CH, Chemnick LG, Stewart CB, Disotell TR, Ryder OA, Batzer MA (2005) A mobile element based phylogeny of Old World monkeys. Mol Phylogenet Evol 37(3):872–880. doi:10.1016/j.ympev.2005.04.015
17. Xing J, Witherspoon DJ, Jorde LB (2013) Mobile element biology: new possibilities with high-throughput sequencing. Trends Genet 29(5):280–289. doi:10.1016/j.tig.2012.12.002
18. Ray DA, Batzer MA (2011) Reading TE leaves: new approaches to the identification of transposable element insertions. Genome Res 21 (6):813–820. doi:10.1101/gr.110528.110
19. Stewart C, Kural D, Stromberg MP, Walker JA, Konkel MK, Stutz AM, Urban AE, Grubert F, Lam HY, Lee WP, Busby M, Indap AR, Garrison E, Huff C, Xing J, Snyder MP, Jorde LB, Batzer MA, Korbel JO, Marth GT, Genomes P (2011) A comprehensive map of mobile element insertion polymorphisms in humans. PLoS Genet 7(8), e1002236. doi:10.1371/journal.pgen.1002236
20. Witherspoon DJ, Xing J, Zhang Y, Watkins WS, Batzer MA, Jorde LB (2010) Mobile element scanning (ME-Scan) by targeted high-throughput sequencing. BMC Genomics 11:410. doi:10.1186/1471-2164-11-410
21. Witherspoon DJ, Zhang Y, Xing J, Watkins WS, Ha H, Batzer MA, Jorde LB (2013) Mobile element scanning (ME-Scan) identifies thousands of novel Alu insertions in diverse human populations. Genome Res 23 (7):1170–1181. doi:10.1101/gr.148973.112
22. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14):1754–1760. doi:10.1093/bioinformatics/btp324
23. Lee WP, Stromberg MP, Ward A, Stewart C, Garrison EP, Marth GT (2014) MOSAIK: a hash-based algorithm for accurate next-generation sequencing short-read mapping. PLoS One 9(3), e90581. doi:10.1371/journal.pone.0090581
24. Smit AF, Hubley R, Green P (1996-2010) RepeatMasker Open-3.0. http://www.repeatmasker.org.
25. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215(3):403–410

26. Ewing AD, Kazazian HH Jr (2010) High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes. Genome Res 20 (9):1262–1270. doi:10.1101/gr.106419.110
27. Iskow RC, McCabe MT, Mills RE, Torene S, Pittard WS, Neuwald AF, Van Meir EG, Vertino PM, Devine SE (2010) Natural mutagenesis of human genomes by endogenous retrotransposons. Cell 141(7):1253–1261. doi:10.1016/j.cell.2010.05.020
28. Beck CR, Collier P, Macfarlane C, Malig M, Kidd JM, Eichler EE, Badge RM, Moran JV (2010) LINE-1 retrotransposition activity in human genomes. Cell 141(7):1159–1170. doi:10.1016/j.cell.2010.05.021
29. Huang CR, Schneider AM, Lu Y, Niranjan T, Shen P, Robinson MA, Steranka JP, Valle D, Civin CI, Wang T, Wheelan SJ, Ji H, Boeke JD, Burns KH (2010) Mobile interspersed repeats are major structural variants in the human genome. Cell 141(7):1171–1182. doi:10.1016/j.cell.2010.05.026
30. Hormozdiari F, Hajirasouliha I, Dao P, Hach F, Yorukoglu D, Alkan C, Eichler EE, Sahinalp SC (2010) Next-generation VariationHunter: combinatorial algorithms for transposon insertion discovery. Bioinformatics 26(12):i350–i357. doi:10.1093/bioinformatics/btq216
31. Wang J, Song L, Grover D, Azrak S, Batzer MA, Liang P (2006) dbRIP: a highly integrated database of retrotransposon insertion polymorphisms in humans. Hum Mutat 27(4):323–329. doi:10.1002/humu.20307
32. Fisher S, Barry A, Abreu J, Minie B, Nolan J, Delorey TM, Young G, Fennell TJ, Allen A, Ambrogio L, Berlin AM, Blumenstiel B, Cibulskis K, Friedrich D, Johnson R, Juhn F, Reilly B, Shammas R, Stalker J, Sykes SM, Thompson J, Walsh J, Zimmer A, Zwirko Z, Gabriel S, Nicol R, Nusbaum C (2011) A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. Genome Biol 12(1):R1. doi:10.1186/gb-2011-12-1-r1