Jan 18, 2019

Transposon insertion sequencing (Tn-seq) library preparation protocol - includes UMI for PCR duplicate removal

DOI
https://dx.doi.org/10.17504/protocols.io.w9sfh6e

Nagendra palani[1]

[1]University of Minnesota Genomics Center

Nagendra Palani
University of Minnesota

OPEN ACCESS

DOI: https://dx.doi.org/10.17504/protocols.io.w9sfh6e

Protocol Citation: Nagendra palani 2019. Transposon insertion sequencing (Tn-seq) library preparation protocol - includes UMI for PCR duplicate removal. **protocols.io** https://dx.doi.org/10.17504/protocols.io.w9sfh6e

Protocol status: Working
We use this protocol and it's working

Created: January 18, 2019

Last Modified: January 18, 2019

Protocol Integer ID: 19474

Keywords: Tn-seq, tnseq, transposon insertion sequencing, transposons, mutagenesis, functional genomics, tradis, NGS, mariner, Tn5, sleeping beauty, masse map genomic locations of transposon insertion, synthetic mini transposons like tn5, transposon mutant, umi for pcr duplicate removal transposon insertion, transposon insertion, pcr duplicate removal transposon insertion, mariner transposon insertion, insertion mutant, genomic location, short genomic sequence available for alignment, using synthetic mini transposon, major approaches to tn, sequencing data, seq library preparation protocol, sequencing, seq mapping, genome junction, genome, genomic dna, novo tn, genetic interaction studies in microorganism, short genomic sequence, unique molecular identifier, transposon, seq library prep, 18bp genomic dna overhang, removing pcr duplicate, seq, pool of mutant, type iis restriction enzyme, pcr duplicate, gene, typhi gene, mutant, genetic interaction study, neb ultra ii fs kit for enzymatic fragmentation, tn5

Abstract

Transposon insertion sequencing (Tn-Seq) is an NGS method to *en masse* map genomic locations of transposon insertions in a pool of mutants. Insertion mutants are usually made using synthetic mini transposons like Tn5 and Himar (in bacteria), Sleeping Beauty & Piggybac (in mammalian cells), and others like Mos (in Drosophila / *C. elegans*).

There are several NGS methods to perform *de novo* Tn-Seq mapping of a library of mutants. Some of the key ones are listed below.

Van Opijnen, Tim, Kip L. Bodi, and Andrew Camilli. "Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms."*Nature methods* 6.10 (2009): 767.

- Uses MmeI (a type IIs restriction enzyme) digestion to generate ~ 18bp genomic DNA overhangs from Mariner transposon insertions. Adapters are ligated to the fragments, the transposon - genome junction is enriched by PCR before being sequenced. Because of the short genomic sequence available for alignment to reference, mapping accuracy can be lower compared to fragmentation based library prep methods.

Langridge, Gemma C., et al. "Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants."*Genome research* (2009).

- DNA extracted from a pool of mutants is sheared/fragmented to a desired fragment size (usually 300 - 400 bp), fragments are end-repaired, and the transposon - genome junction is enriched by PCR before being sequenced.

The list is only intended to highlight two major approaches to Tn-seq library prep - restriction digestion based or fragmentation based. There are several variations on each approach that are now available for the discerning scientist. There are also less known but interesting alternative methods like **HTML-PCR**.

After trying out several of the above methods, we have developed a Tn-Seq library preparation protocol that incorporates desirable features from existing protocols and new features that will improve data quality, while being easy to perform (an undergraduate student with some wet-lab experience can go through the protocol in a day).

The protocol is conceptually similar to the TraDIS protocol (see the Langridge *et al* paper above).

Key features are:

1. Uses the NEB Ultra II FS kit for enzymatic fragmentation of genomic DNA, end-repair of fragments, and ligation of adaptor. The kit allows performing all the reactions sequentially in a single tube (one-pot prep).

2. Incorporates Unique Molecular Identifiers (UMI) to enable removing PCR duplicates from sequencing data. Improves read count accuracy when the data is to be used for read-frequency based analyses (fitness or log fold

change).

3. Includes 'best practices' library features like color diversity phasing and separate indexing step to save cost on adaptor oligos.

The method is also automation friendly.

Guidelines

- Time from genomic DNA sample to sequencing-ready library takes ~ 1 to 1.5 days.
- Protocol can be performed with 1.5 ml tubes or 96-well plates.
- Because Tn-seq libraries are amplicon-type libraries with lower base diversity than typical DNA-seq or RNA-seq libraries, the libraries need to be phased (See article https://www.nature.com/articles/nmeth.2634 suppl. fig 2 & 3 for explanation) or a PhiX library spiked into the Tn-seq library before cluster generation. At UMGC, we typically spike in PhiX at 15% of the total reads to add base color diversity to Tn-seq libraries. You can design phased oligos to add base color diversity during library prep and reduce the amount of PhiX added before clustering. We still recommend adding at least 2% PhiX to phased libraries because PhiX spike-in provides valuable sequencing run-time metrics in Basespace.
- To use the UMI for PCR duplicate removal, **sequencing needs to be performed in paired-end mode**. If you just need the transposon - genome junction, single-end read mode is sufficient. **Minimum read length in either mode should be 100 bases.**

Materials

MATERIALS

TE buffer

NEBNext Ultra II Q5 Master Mix **New England Biolabs Catalog #**E7649

NEBNext Ultra II FS DNA Library Prep with Sample Purification Beads - 24 rxns **New England Biolabs Catalog #**E6177S

Oligonucleotides required:

TA_Adaptor_Top - /5Phos/CTCACCGCTCTTGTAGS NNNNNNNN CTGTCTCTTATACACATCTCCGAG*C
TA_Adaptor_Bottom - CTACAAGAGCGGTGAGT

Transposon_Enrich_Rev - GTCTCGTGGGCTCGGAGATGTGTATAAGAGACA*G

Transposon_Enrich_Fwd - The 3' portion of this primer (in italics) is transposon specific. The primer can be designed to amplify from both ends of the transposon or amplify specifically one side of the transposon. Designs below amplify from both ends of a transposon. Re-design the transposon-specific region according to your

Mariner transposon:
TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG NNN *CCGGGGACTTATCAGCCAAC*C*

Sleeping Beauty transposon:
TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG NNN *GTGTATGTAAACTTCCGACTTCAACT*G*

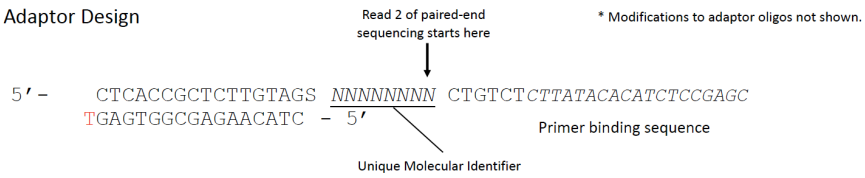
Nextera Indexing Oligo design
Nextera_R1 : AATGATACGGCGACCACCGAGATCTACAC [i5] TCGTCGGCAGCGT*C
Nextera_R2 : CAAGCAGAAGACGGCATACGAGAT [i7] GTCTCGTGGGCTCG*G

You will need to replace the [i5] & [i7] with unique index barcodes. You can use the Hamming barcodes from https://doi.org/10.1371/journal.pone.0036852. Order the appropriate number of indexing oligos for your samples based on if you want to do combinatorial indexing (cheaper) or unique dual indexing (better with the patterned flowcells).

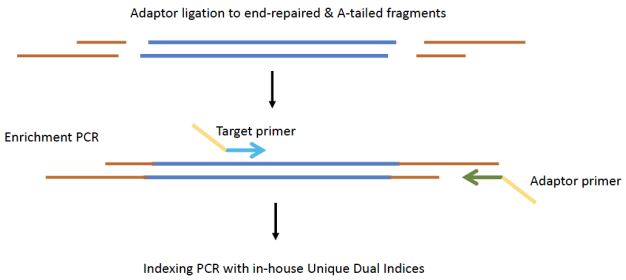
Note: When selecting index barcodes, make sure

1) there is no collision with other barcodes that will be run on the same flowcell as your samples
2) don't start with or have consecutive G nucleotides if the samples will be run on instruments that use two-color (NextSeq, NovaSeq) or one-color (iSeq) chemistry.

A



B



Safety warnings

WARNING: This method is designed to be used for transposons EXCEPT Tn5. The Nextera indexing system we use has a sequence conflict with the Tn5 transposon mosaic ends (Nextera is based on the Tn5 system). However, once you understand the protocol, you should be able to modify the indexing oligo sequences accordingly to use with Tn5 without changing any of the steps. Or you can contact me - nagendra at umn.edu and I can send you the TruSeq indexing oligo design that is compatible with Tn5.

Before start

Make sure that the genomic DNA is clean (260/230 & 260/280 > 1.8) and is in a volume less than 26 ul. Measure DNA concentration by fluorometry (Picogreen or Qubit).

Input DNA mass can vary from 10 ng to 500 ng. Make sure that you are using the appropriate mass of DNA as input to avoid sampling bottlenecks in library prep that can exclude low abundance mutants.

My rule of thumb to determine the amount of input genomic DNA for a transposon library is to have at least 100 copies of the genome for **each** mutant in the pool (assuming the input pool is uniformly distributed). You can calculate the number of genome copies in a given amount of DNA at http://cels.uri.edu/gsc/cndna.html . Plug in DNA mass and genome size in bp to get number of genome copies.

Genome copies per mutant (average) =

total number of genome copies per unit mass of DNA / number of mutants in the pool.

For example, 100 ng of a 5 Mb bacterial genome contains ~ 1.8×10^7 genome copies which works out to ~ 180,000 mutants that can be present in a non-saturated pool.

In order to ensure that you are able to pick low abundance mutants in the sequencing data, aim to have at least 1000 copies of the genome per mutant. This is particularly important if the mutant pool has been subjected to a long growth period (like animal infections) since the dynamic range of mutant frequencies will have increased.

The protocol here uses 100 ng of input DNA.

Adaptor formation

1

- Use 0.1x TE to prepare 100 µM stock of all the oligos.
- In a 0.2 ml PCR tube, add the following reagents, vortex well, and spin down. Place on the thermal cycler at 95 °C for 00:03:00 (heated lid).

15 µL Oligo TA_Adaptor_Top (100 uM)
15 µL Oligo TA_Adaptor_Bottom (100 uM)
70 µL dH2O
---
100 µL Total

- After incubation, terminate the incubation and let the heat block cool down to room temperature. Remove the tube and place it on the bench for 1 hr.
- The oligos will have annealed and formed the adaptor.

DNA Fragmentation, end-repair, ligation

2

This section of the protocol is from the manual for NEB Ultra II FS kit - Catalog # E7805.

(Refer to https://www.neb.com/-/media/catalog/datacards-or-manuals/manuale7805.pdf : Section for Input > 100 ng for detailed protocol)

- Aliquot 100 ng of genomic DNA into a well of a 96-well plate and use 1x TE to bring the volume to 26 µL.
- Add the NEBNext Ultra II FS buffer + enzyme mix to the DNA (follow protocol from the linked manual)
- Set up the following program on the thermal cycler 75 °C heated lid

37 °C for 00:15:00
65 °C for 00:30:00

4 °C hold.

Note

This should fragment the library to have a distribution between 200 - 450 bp.

- Add the NEBNext Ultra II ligation mix + enhancer to the well. Add 2.5 µL of the adaptor to the mixture. Mix well and incubate at 20 °C for 00:15:00 heated lid off. Post incubation, add 28.5 µL of 0.1x TE and bring volume to 100 µL.

- Perform bead clean-up of the ligation mix using the sample purification beads that came with the kit.

Use 30 µL of beads for 1st bead addition

15 µL of beads for 2nd bead addition (to select for an insert size distribution of 200 - 350 bp).

- Elute in 10 µL of dH_2O.

Enrichment PCR

3

Make 10 uM dilutions of the following primers.

Transposon_Enrich_Fwd
Transposon_Enrich_Rev

For each sample set up the following PCR in a 96-well plate.

Transposon - genome junction enrichment

7.5 µL Adaptor ligated DNA
12.5 µL Ultra II Q5 master mix
2.5 µL Transposon_Enrich_Fwd

2.5 µL Transposon_Enrich_Rev

Run the following program on a thermal cycler.

Note

- Use the NEB Tm Calculator to calculate the primer Tm values. If recommended Tm is greater than 65 ºC, use 65 ºC in the annealing & extension step below.

For the Q5 master mix that comes with the NEB Ultra II FS kit (or the separately bought version), NEB recommends a combined annealing + extension step at 65 ºC. If your enrichment primers require a different annealing temperature, you can use the standard Q5 protocol (anneal at primer Tm, extend at 72 ºC). See **here** for more information).

- The number of PCR cycles needs to be decided by the user to get enough enriched DNA for indexing. Lower number of cycles is better to maintain sequence diversity.

98 °C 00:00:30 Initial denaturation - use heated lid

98 °C 00:00:10 Denaturation
65 °C 00:01:15 Annealing & Extension
10 - 15 cycles (to be decided by user)

4 °C Hold

After end of PCR, use 1 µL of the PCR end-product for Picogreen / Qubit.
Use 2 µL to run the sample on a Tapestation (D5000 high-sensitivity) to check for library size.

Each sample library should be a unimodal distribution between ~ 300 bp - 600 bp. All libraries should have an average size distribution within 15% across all samples.

Indexing PCR

4

Make 10 uM dilutions of the **Nextera_R1** & **Nextera_R2** primers.

For each sample, use 10 ng **of Enrichment PCR product** as template for the indexing reaction.

Set up the following PCR:

12.5 µL Ultra II Q5 master mix
1.25 µL Nextera_R1
1.25 µL Nextera_R2
10 ng Template
0 µL dH2O As required
---
25 µL Total

PCR Program:

98 °C 00:00:30 Initial Denaturation

98 °C 00:00:10 Denaturation
65 °C 00:01:30 Annealing & Extension
10 cycles

4 °C Hold

After end of PCR, use 1 µL of the PCR end-product for Picogreen / Qubit.

Library pooling, QC, and sequencing

5

- Pool samples by equal mass (assuming the library size distribution for the samples are similar). Otherwise, pool by equal molar mass.
- Do a 1.2x SPRI bead clean-up. Elute in 30 µL 0.1x TE
- Check concentration of the pool by Picogreen/Qubit and quantify library using Kapa Illumina Library quantification kit. Dilute the pool to appropriate molarity required for the sequencing instrument.

- Load the pool on the sequencer and run with 15% PhiX spike-in for base diversity (works on MiSeq, HiSeq 2500, iSeq, and Novaseq. Might need more PhiX for NextSeq.)

Sequencing parameters : Paired-end sequencing with at least 100 base read length. Set Index read length appropriately.

Data analysis

6

Scripts for pre-processing the data (fastq files to readcounts per insertion) will be made available at https://github.com/nppalani/TnSeq/ . The readcounts can then be consolidated and used with software like TRANSIT, TnSeqDiff etc., for gene essentiality analysis.

Note that the scripts undergo frequent revision, so ensure that you are working with the latest version. Feel free to contact me if you have questions.