Sequencing Library Preparation

Sarah Boswell
Sarah_Boswell@hms.harvard.edu
Overview

- Starting material
- Standard RNASeq library prep (Tru-Seq)
- Multiplexing
- Genomic libraries
- Tagmentation-based approach (Nextera)
- Low input libraries (including single cell)
- Mate-pair sequencing (circularization)
Basic library preparation

Sheared Genomic DNA or cDNA

End Repair (blunt ends)

Add 3’ A Tail

Ligate Adapters

Enrich/Linearize with PCR

Sequencing

https://www.diagenode.com/applications/library-preparation
High Quality Starting Material

- Genomic DNA should be RNAsed
  - How will you shear your DNA?
- RNA needs to be DNAsed
  - RNA should be of a high RIN # (>7 / >8)
- Take care to avoid ethanol contamination in sample as it will inhibit downstream reactions
  - Ethanol precipitations are trickier than they seem
RNA Extraction

- Cultured cells – Easy!
  - Use your favorite column kit – Qiagen, Invitrogen, Zymo
  - For high throughput suggest bead based – MagMax

- Tissue samples
  - Dissect in cold room if possible
  - Use RNAlater solution to store tissue sample
  - Upon extraction proceed to homogenization in cold room

Practice your extraction before the real experiment
RNA enrichment

- mRNA enrichment – transcriptome
- rRNA removal – total RNASEq

Front Genet. 2015 Jan 26;6:2
RNA quality

- Degraded samples should only be used to make a “total” RNA seq library – rRNA removal
- FFPE optimally prepared as total RNA seq
RNASeq Stranded Library Prep (dUTP method)

1. rRNA depletion
2. RNA fragmentation
3. Random primer RT
4. dUTP incorporation
5. Adenosine tailing
6. Y-shaped adapter ligation
7. dUTP strand degradation or strand specific amplification
8. PCR

Index

www.support.illumina.com
RNASeq Stranded Library Prep

(dUTP method)

Read 1 sequencing reads are reverse compliment of mRNA

Read 2 sequencing reads match the starting mRNA

http://rnaseq.uoregon.edu/#library-prep
Library Strandedness

Read 1 sequencing reads are reverse complement of mRNA

Read 2 sequencing reads match the starting mRNA

Library Amplification
NOT UTP strand

First Strand Synthesis (RC)
Second Strand Synthesis (dUTP)

-dUTP is used in the Second Strand Marking Master Mix instead of dTTP. Second strand cDNA synthesis is required for adapter ligation.

Polymerase used in the assay will not incorporate past dUTP. Therefore, the second strand is effectively quenched during amplification.

Read 1 sequencing reads are reverse compliment of mRNA

Read 2 sequencing reads match the starting mRNA

RNASeq Stranded Library Prep (dUTP method)

1. rRNA depletion
2. RNA fragmentation
3. Random priming
4. dUTP incorporation
5. Adenosine tailing
6. Y-shaped adapter ligation
7. dUTP strand degradation or strand specific amplification

qPCR quantitate before sequencing
Multiplexing (barcodes and indices)

- Multiplexing allows optimal use of reads you will get
- Charges for sequencing are usually per lane of the flow cell
  - HiSeq generates ~150 million reads per lane
  - NextSeq generates ~ 450 million reads (one lane instrument)
- For RNA-Seq number of reads you need will depend on your experiment
  - 10 million standard for transcriptome
  - 20 million standard for total RNA (rRNA depleted)

Make sure multiplexing libraries of similar size
Consider Cluster Size in Multiplexing
Multiplexing (barcodes and indices)

- Sequencing Primer (SP)
- Barcode
  - Sample identifier read within the sequencing read
- Indexing Primer (IP)
  - Index read on separate read off instrument
Sequencing Read Order

1. Read 1
2. Index Read 1 (i7)
3. Index Read 2 (i5)
4. Read 2

HiSeq/MiSeq (4 color)
- A&C read on one camera
- G&T read on other

NextSeq (2 color)
Once you have sheared your DNA this is a quick process

- Acoustic shearing – Covaris
- Sonication
- Hydrodynamic shearing – nebulization

- Protocol same as for RNASEq once you have sheared dsDNA

https://www.diagenode.com/applications/library-preparation
Tagmentation Approach

- *Nextera* from Illumina
- Transposomes contain the transposase and an oligo (transposon+adapter)
- Very fast and efficient for DNA library preps
  - Works with small amounts of DNA
  - Important to RNAse treat your sample
  - Needs very precise DNA quantitation (Qubit)
- Often used as last stage of low input RNASeq library protocols
Tagmentation
(DNA fragmentation facilitated by transposon activity)
Low Input RNASeq

- Clontech – SMARTSeq
- Lower input = less chance to see transcript of interest
- Need to consider sampling error
- Differential expression observed is accurate
- Single cell methods will only capture 40-50% of transcriptome
Smart-Seq2 / Broad DGE (3') / Drop-Seq

DropSeq Single Cell Sequencing

- Cells lysed in the beads and hybridize to primer
- Droplets are broken and RT/template switching occurs on pool
- STAMP: single cell transcriptomes attached to microparticles

InDrop Single Cell Sequencing

- Lysis and reverse transcription occurs in the beads
- Samples are frozen after RT as RNA:DNA hybrid in gel.

InDrop Library Prep

- Second Strand Synthesis to make full length dsDNA
- In vitro transcription (IVT) back to RNA off T7 promoter from primer
- Fragmentation RNA
- RT with random hexamer primer containing adaptor
- PCR off adaptors to add index and illumina adaptors
Low Input or Single Cell

- Only use if needed for experiment
- Commercially available kits all rely on PolyA tail
- If you can get more starting material then you will get better results.
- Gold standard is TruSeq with >500ng input RNA
- Plan a small scale starter experiment to see if protocol will give useful results
RNA Capture Sequencing

- Capture targeted sequence using biotinylated RNA bait
- Sequencing library applied to beads
- Retain only library covering genes of interest
- Saves money on sequencing
- IDT lockdown probes expensive but good for small number genes
Whole Genome Sequencing (WGS) vs Whole Exome Sequencing (WES)

https://www.my46.org/intro/whole-genome-and-exome-sequencing
Capture Probes

- Tiling is the number of times a base is covered by a different probe.
- Difficult to design probes if looking at a single gene family or pseudogenes.
Circularization Library Prep Methods

For assembly of non-standard libraries

- NET-Seq: Native Elongating Transcript-Sequencing
- Ribo-Seq: Ribosomal profiling
- Mate-Pair Sequencing: genomic library preparation

- All involve ligation of a linker and circularization of insert
Mate-Pair Genomic Library

- Insert pieces of DNA can be long (2-5kb)
  - RNA based applications all inserts around 100bp

- Allows for better de novo genome assembly and finishing of genome assemblies.

- Helps in determining presence and location of genomic rearrangements and amplifications.
Mate-pair sequencing

Genomic DNA (blue) is tagmented by the tagment enzyme, which attaches a biotinylated junction adapter (green) to both ends of the tagmented molecule.

The tagmented DNA molecules are then circularized and the ends of the genomic fragment are linked by two copies of the biotin junction adapter.

http://www.illumina.com/
Mate-pair sequencing

Genomic DNA (blue) is tagmented by the tagment enzyme, which attaches a biotinylated junction adapter (green) to both ends of the tagmented molecule.

The tagmented DNA molecules are then circularized and the ends of the genomic fragment are linked by two copies of the biotin junction adapter.

Circularized molecules are fragmented again, yielding smaller fragments. Fragments containing the original junction are enriched via the biotin tag (B) in the junction adapter.

After end repair and A-tailing, adapters (gray and purple) are added, enabling amplification and sequencing.

http://www.illumina.com/
Mate-pair sequencing

After end repair and A-tailing, adapters (gray and purple) are added, enabling amplification and sequencing.
Mate-pair sequencing

http://www.illumina.com/
Mate-pair sequencing

Going backwards
Mate-pair sequencing
Mate-Pair Genomic Library

- Insert pieces of DNA can be long (2-5kb)
  - RNA based applications all inserts around 100bp

- Allows for better de novo genome assembly and finishing of genome assemblies.

- Helps in determining presence and location of genomic rearrangements and amplifications.
NET-seq: Native Elongating Transcript-sequencing

Final Thoughts

Practice your library prep on a control sample be sure you understand what is happening at each step.

qPCR

Precise quantitation is key to good sequencing!
Sequencing Library Preparation

Sarah Boswell
Sarah_Boswell@hms.harvard.edu
Useful Websites

- support.illumina.com/
- seqanswers.com/
- core-genomics.blogspot.com/2012/04/how-do-sprite-beads-work.html
Modifying / Custom Sequencing

- What is the length of the species of interest?
- How can you purify this species?
- Are there end modifications you can exploit?
- Use high quality enzymes / kits.
- READ the manual in its entirety. Talk to someone who has used the kit.
- Understand the primers and protocol completely.
Micro RNA Library Prep

- Ligate 3’ and 5’ adaptors
- RT PCR – 1\textsuperscript{st} strand synthesis
- Amplify
- Gel purify
- Validate library
Paired End Reads

- Sequence one end and then sequence other end
- Is useful for aligning repetitive sequences & de novo assembly
- If make library small fragment size the when do PE will read same sequence 2x. This helps identify sequencing error.