Library Construction for NGS

3

Melanie Kappelmann-Fenzl

Contents

3.1 Introduction 40
3.2 Library Preparation Workflow 40
References 45

What You Will Learn in This Chapter

After finishing this chapter, you will have a basic overview of the individual experimental work steps regarding the NGS library preparation workflow, the connection between the scientific or clinical question and the choice of the corresponding library preparation. For this purpose, the technical and molecular biological relevance of the individual work steps is briefly described.

M. Kappelmann-Fenzl (✉)
Deggendorf Institute of Technology, Deggendorf, Germany

Institute of Biochemistry (Emil-Fischer Center), Friedrich–Alexander University Erlangen–Nürnberg, Erlangen, Germany
e-mail: melanie.kappelmann-fenzl@th-deg.de

M. Kappelmann-Fenzl (ed.), *Next Generation Sequencing and Data Analysis*, Learning Materials in Biosciences, https://doi.org/10.1007/978-3-030-62490-3_3

3.1 Introduction

Library preparation involves generating a collection of DNA/cDNA fragments for sequencing. NGS libraries are typically prepared by fragmenting a DNA or RNA sample and ligating specialized adapters to both fragments ends. In this textbook we will focus on the Illumina® Library Preparation workflow for short-read sequencing (https://emea.illumina.com/techniques/sequencing/ngs-library-prep.html). As already mentioned, a wide variety of NGS methods exists, and for almost every method a comprehensive sequencing library preparation solution. The principle is almost similar for all library preparation workflows [1, 2]. Further information on library preparation for long-read [3] or single-cell sequencing [4–6] can be found on the websites of the corresponding providers:

Single-cell sequencing

- Illumina (https://emea.illumina.com/techniques/sequencing/rna-sequencing/ultra-low-input-single-cell-rna-seq.html)
- 10xGenomics (https://support.10xgenomics.com/single-cell-gene-expression/automated-library-prep)
- Qiagen (https://www.qiagen.com/de/products/discovery-and-translational-research/next-generation-sequencing/library-preparation/qiaseq-fx-single-cell-dna-library-kit/#orderinginformation)
- And many others

Long-read sequencing:

- Illumina (https://emea.illumina.com/science/technology/next-generation-sequencing/long-read-sequencing.html)
- PacificBioscience (https://www.pacb.com/products-and-services/consumables/template-preparation-multiplexing-kits/)
- Oxford Nanopore (https://nanoporetech.com/products/kits)
- And many others

3.2 Library Preparation Workflow

The core steps in preparing RNA or DNA for NGS are:

- Fragmenting and/or sizing the target sequences to a desired length.
- Converting target to double-stranded DNA (in terms of RNA-Seq).
- Attaching oligonucleotide adapters to the ends of target fragments.
- Quantifying the final library product for sequencing.

The preparation of a high-quality sequencing library plays an important role in Next-Generation Sequencing (NGS). The first major step in preparing nucleic acids for NGS is fragmentation. The most common and effective *fragmentation methods* can be subdivided into three classes:

1. Physical fragmentation (are acoustic shearing, sonication, and hydrodynamic shear).
2. Enzymatic fragmentation (DNase I or other restriction endonuclease, non-specific nuclease, Transposase).
3. Chemical fragmentation (heat and divalent metal cation). This method is used to break up long RNA fragments, whereas the length of your RNA can be adjusted by modulating the incubation time.

But also, a PCR amplification of genetic loci of interest can be chosen. Each NGS approach has its own specific protocol. Available NGS sample preparation kits are: Illumina, New England BioLabs, KAPA Biosystems, Swift Bioscience, Enzymatics, BIOO, etc. The principle workflow after fragmentation can be briefly described by the following working steps, which are also illustrated in a simplified way in Fig. 3.1 (the numbering of each working step is analog to the numbering in Fig. 3.1):

1. *Quantification and profile your isolated DNA or RNA samples*. This is one of the most important steps after sample preparation. For sequencing the samples have to be of a very good quality, the concentration must be determined, and the fragmentation efficiency must be checked (target size for short-read sequencing is commonly 200bp–800bp) before you can go any further.
2. *Perform End Repair and size selection via AMPure XP Beads* (Beckman Coulter Genomics [7]). This process converts the overhangs resulting from fragmentation into blunt ends and serves for size selection. End Repair is not performed during RNA-Seq library preparation.
3. *In Terms of RNA-Seq Library Preparation* [8].
   Depletion of rRNA and fragmentation or polyA-capture or another procedure for isolating the RNA type of interest (depending on your research question and library preparation kit you use: https://emea.illumina.com/products/by-type/sequencing-kits/library-prep-kits.html) is necessary.
4. The RNA fragments obtained must then be transcribed into cDNA for sequencing by synthesis of the first cDNA strand followed by synthesis of the second cDNA strand.
5. *Adenylate 3'-Ends of DNA/cDNA*.
   A single "A" nucleotide is added to the 3' ends of the blunt fragments to prevent them from ligating to one another during the adapter ligation reaction. A corresponding single "T" nucleotide on the 3' end of the adapter provides a complementary overhang for ligating the adapter to the fragment.
6. *Adapter Ligation and Size Selection via AMPure XP Beads of DNA/cDNA*
   Adapter ligation is a crucial step within the NGS library preparation. Adapters are design with a single "T" nucleotide on the 3' end to recognize the "A" nucleotide

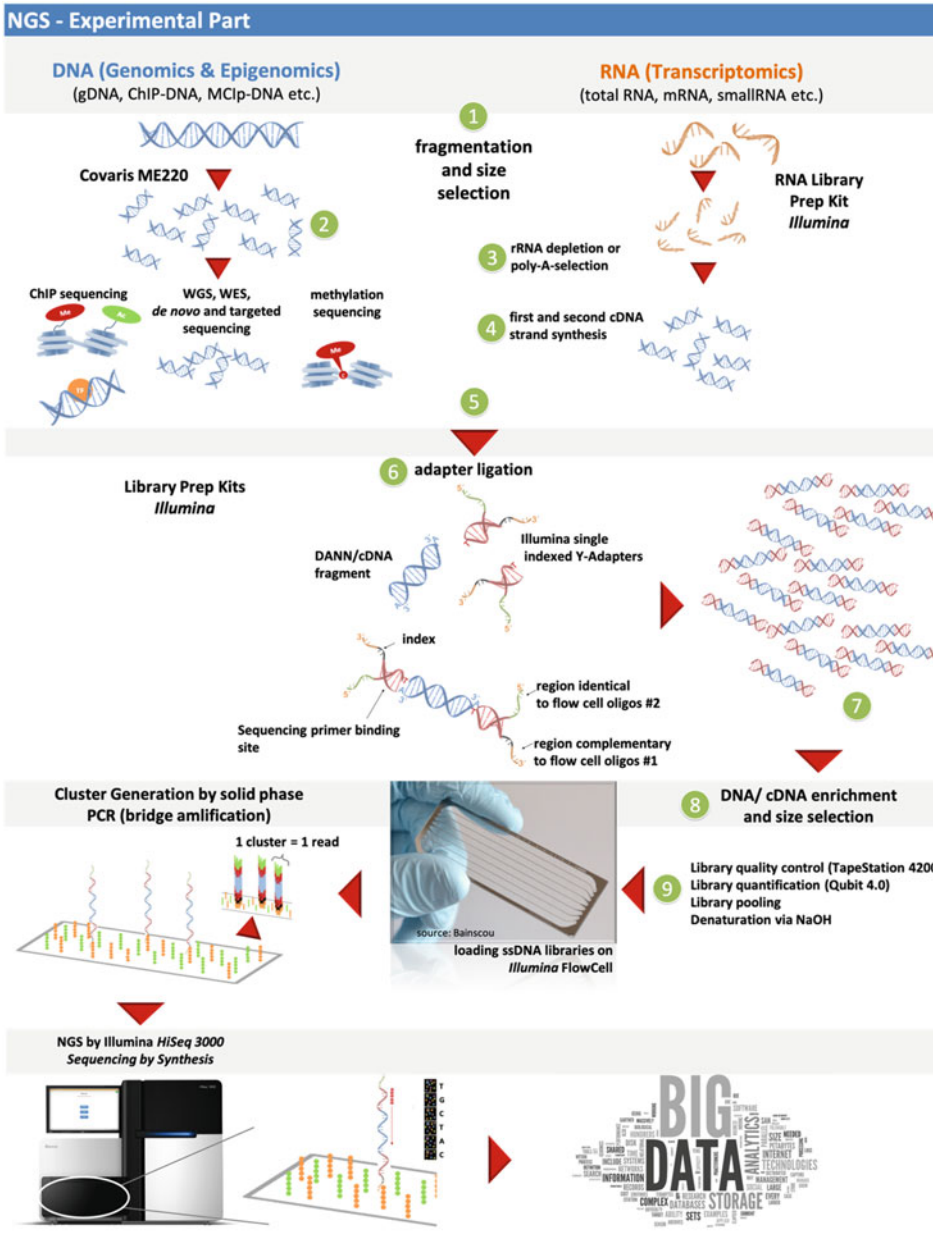


Fig. 3.1 NGS Library preparation workflow for DNA or RNA samples, respectively (source: © Melanie Kappelmann-Fenzl)

overhang (see Step 5) of each DNA/cDNA fragment. One part of the adapter sequence is complementary to the oligos covering the sequencing flow-cell and thus guarantees binding and another part is complementary to the later added sequencing primer and

thus guarantees sequencing of the fragments. You can also ligate multiple indexing adapters allowing to load more than one sample onto a flow-cell. This short index enables you to distinguish between all the loaded samples carrying different indexing adapters. Adapters have a defined length of ~60 bp, hence ~120 bp long fragments can easily be identified as adapter dimers without any DNA/cDNA insert.

7. *Purify Ligation Products* (e.g., Pippin™ size selection; https://sagescience.com/products/pippin-prep/).

   This process purifies the products of the ligation reaction on a gel and removes unligated adapters, as well as any adapters that might have ligated to one another.

8. *Enrich DNA/cDNA Fragments and size selection via AMPure XP Beads*.

   This process uses PCR to selectively enrich those DNA/cDNA fragments that have adapter molecules on both ends and to amplify the amount of DNA/cDNA in the library. Additionally, it serves for size selection.

9. *Validate Library and normalize and pool libraries*.

   This procedure is performed for quality control analysis on your sample library and for quantification. Therefore, a Agilent Technologies Bioanalyzer or TapeStation (https://www.agilent.com/en/product/automated-electrophoresis) and a Qubit 4 Fluorometer (https://www.thermofisher.com/de/de/home/industrial/spectroscopy-elemental-isotope-analysis/molecular-spectroscopy/fluorometers/qubit/qubit-fluorometer.html) are used.

   This process describes how to prepare DNA/cDNA templates for cluster generation. Indexed DNA/cDNA libraries are normalized to 10nM, and then pooled in equal volumes.

Example library preparation protocols for RNA-Seq or ChIP-Seq, respectively, can be found in the Appendix section (Sect. 13.1).

Take Home Message

- Different research or clinical questions require different library preparation workflows.
- DNA sequencing approaches require proper fragmentation before library preparation.
- RNA molecules are not directly sequenced due to their chemical instability and the difficulty of processing and amplifying single-stranded nucleic acids. Thus, RNA has to be converted into cDNA (reverse transcription) for sequencing purpose.
- Quality and quantity determinations are essential within all library preparation workflows.

Review Questions

Review Question 1

Why should a final library have a median insert size of ~250–300 bp to support long paired end 2 × 150 read lengths?

Review Question 2

Which different read-outs can be obtained by different RNA-Seq Library Preparation methods?

Review Question 3

A crucial factor leading to misrepresentation of data is the bias prevalent in almost all steps of NGS sample preparation. Discuss a few possible solutions to this kind of bias!

Review Question 4

Graphically illustrate the structure of adapter-lying DNA/cDNA fragments and label the individual sequence sections of the adapters!

Answers to Review Questions

Answer to Question 1: Otherwise the percentage of adapter contaminated reads increases;

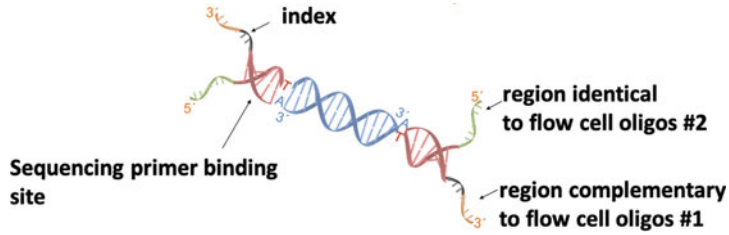
Answer to Question 2:

Objective	Principles of approach
Gene expression	Target poly(A) mRNAs (enrich or selectively amplify)
Alternative splicing	Target exon/intron boundaries by either long-read sequencing (>300 bp) or paired end sequencing (≥2 × 75) and rRNA depletion
miRNA (or small RNAs)	Target short reads (miRNAs: 18–23 bp) using size selection purification. piRNAs, snoRNAs, tRNAs are all <100 bps

Answer to Question 3: Bias during amplification of AT- and GC-rich regions: PCR-free amplification could yield better read distribution and coverage compared to PCR methods, but would require large quantities of starting DNA material.

PCR bias during library preparation for RNA-Seq can be introduced by the additional steps to convert RNA to cDNA. KAPA HiFi DNA polymerase can be used for the amplification step to reduce this kind of bias.

Answer to Question 4:



Acknowledgements We are grateful to Dr. Ines Böhme (Institute of Biochemistry (Emil-Fischer Center), Friedrich–Alexander University Erlangen–Nürnberg, Erlangen, Germany) for reviewing this chapter and Alexander Oliver Matthies (Institute of Biochemistry (Emil-Fischer Center), Friedrich–Alexander University Erlangen–Nürnberg, Erlangen, Germany) for critically reading this text.

References

1. van Dijk EL, Jaszczyszyn Y, Thermes C. Library preparation methods for next-generation sequencing: tone down the bias. Exp Cell Res. 2014;322(1):12–20.
2. Head SR, Komori HK, LaMere SA, Whisenant T, Van Nieuwerburgh F, Salomon DR, et al. Library construction for next-generation sequencing: overviews and challenges. Biotechniques. 2014;56(2):61–4, 6, 8, passim.
3. Sakamoto Y, Sereewattanawoot S, Suzuki A. A new era of long-read sequencing for cancer genomics. J Hum Genet. 2020;65(1):3–10.
4. Trombetta JJ, Gennert D, Lu D, Satija R, Shalek AK, Regev A. Preparation of Single-Cell RNA-Seq Libraries for Next Generation Sequencing. Curr Protoc Mol Biol. 2014;107:4–22. 1-17.
5. Jemt A, Salmen F, Lundmark A, Mollbrink A, Fernandez Navarro J, Stahl PL, et al. An automated approach to prepare tissue-derived spatially barcoded RNA-sequencing libraries. Sci Rep. 2016;6:37137.
6. Jemt A, Salmen F, Lundmark A, Mollbrink A, Navarro JF, Stahl PL, et al. Corrigendum: an automated approach to prepare tissue-derived spatially barcoded RNA-sequencing libraries. Sci Rep. 2017;7:41109.
7. Greenwald WW, Li H, Benaglio P, Jakubosky D, Matsui H, Schmitt A, et al. Subtle changes in chromatin loop contact propensity are associated with differential gene regulation and expression. Nat Commun. 2019;10(1):1054.
8. Podnar J, Deiderick H, Huerta G, Hunicke-Smith S. Next-Generation Sequencing RNA-Seq Library Construction. Curr Protoc Mol Biol. 2014;106:4–21. 1-19.