Fundamentals of NGS Library Preparation

How poor quality starting material can affect sequencing data

Agilent
Trusted Answers

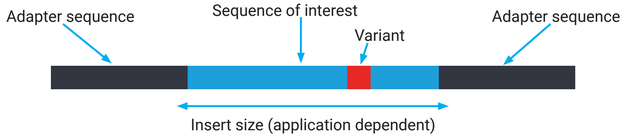
What is library preparation?

Library preparation is the first step in next-generation sequencing (NGS) analysis. Genomic DNA or RNA are often used as starting material for NGS applications.

Library preparation begins with fragmentation, either by mechanical or enzymatic shearing. Adapter sequences are then added to the 3' and 5' ends of fragments. Common sequencing applications include whole genome (WGS), exome (WES), transcriptome (RNA-Seq), and targeted sequencing. Target enrichment methods are shown here as examples:

Amplicon-based NGS methods

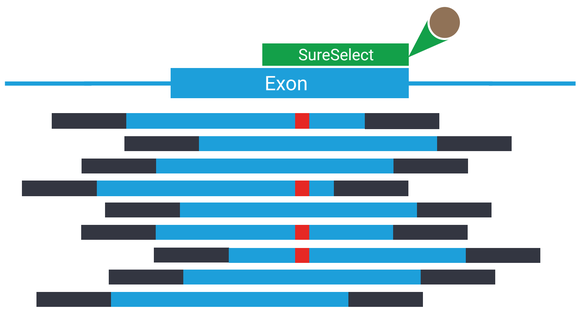
The most common methods for target enrichment provide a pool of primer pairs to amplify a specific region (creating an amplicon) with PCR, for example by using the Agilent HaloPlex custom kits. The primer distance defines the insert size for sequencing.



Example: Agilent HaloPlex NGS assays

Capture-based NGS methods

Random shearing generates fragments that are captured by biotinylated baits complementary to the target region. In turn, these baits with attached DNA fragments are captured by magnetic beads.



Example: Agilent SureSelect NGS assays

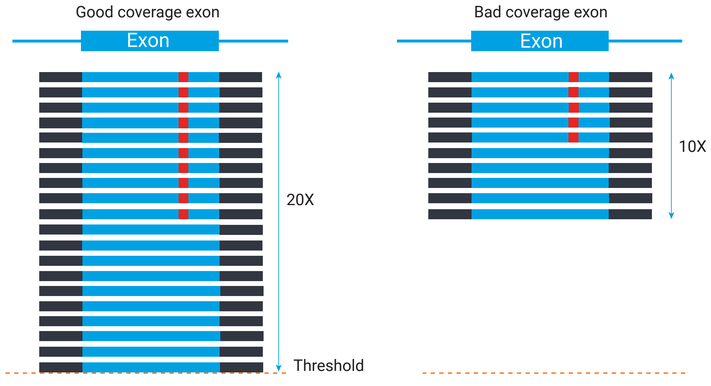
After library preparation, the final library is quantified. Libraries are normalized to a specific molarity, and pooled. The pooled libraries are then sequenced, with the sequencing quality typically characterized by depth of coverage, library complexity, and the on-target rate. High quality NGS libraries should provide reads with a high coverage that is evenly distributed accross the entire sequence of interest.

Published in the USA, September 1, 2019
5994-0865EN

Measuring NGS data quality

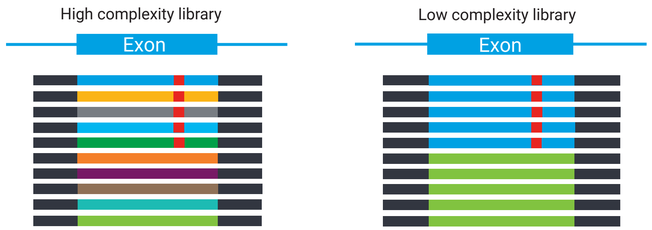
Depth of coverage

The depth of coverage is defined as the average number of times a base is read during sequencing. A higher number of reads will increase the confidence in sequencing data. Usually a minimum depth is set as a threshold for variant calls.



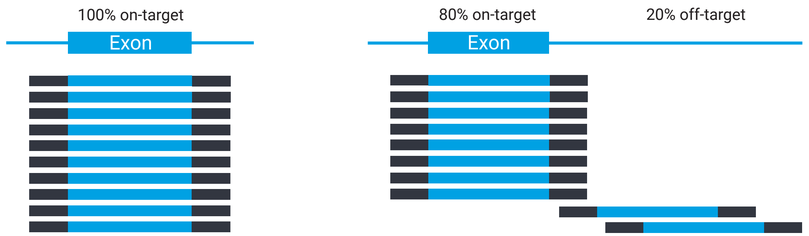
Library complexity

Ideally, sequencing libraries should be as complex as possible, reflecting the true nature of the starting material. Complexity can be measured by the number of duplicate reads; highly complex libraries have lower numbers of these. Greater library complexity typically provides greater confidence that a variant detected in the sample is real.

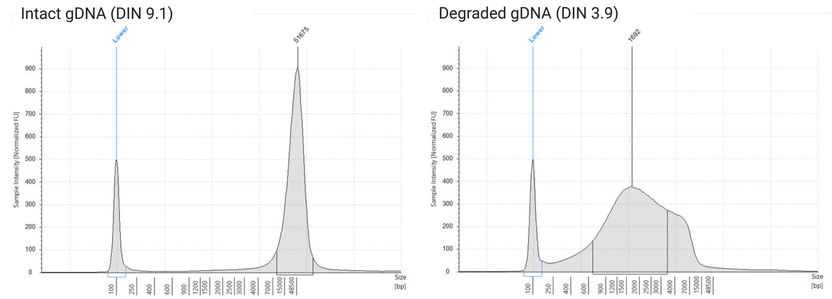


On-target rate

The on-target rate is the percentage of sequencing data covering a region of interest. Reads outside of the region of interest are referred to as off-target: these reduce the depth of coverage. To compensate, increased sequencing is needed for each library.

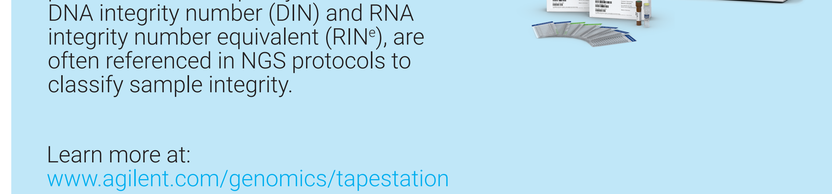


Effects of poor quality DNA or RNA



gDNA analyzed with the Genomic DNA ScreenTape assay.

Degradation of DNA can interfere with amplicon-based NGS methods. If the initial fragments are smaller in size than the expected amplicon, the depth of coverage can be impacted. PCR cycles can be increased to compensate for this, but can result in decreased library complexity.



RNA analyzed with the RNA ScreenTape assay.

The degradation of RNA can directly influence the depth of coverage. PolyA enrichment is often performed at the beginning of RNA-Seq protocols to remove ribosomal RNA. In degraded RNA samples, sequences further from the polyA tail are less represented and the depth of coverage is not uniform.

Agilent TapeStation systems

Library quality and sequencing data can be directly affected by low integrity or low concentration of starting material. Agilent offers multiple solutions to provide reliable results for NGS QC. One option is a TapeStation system. They are useful QC tools to easily assess both parameters. Their quality metrics, the DNA integrity number (DIN) and RNA integrity number equivalent (RIN^e), are often referenced in NGS protocols to classify sample integrity.

Learn more at:
www.agilent.com/genomics/tapestation