Simplifying your next-generation sequencing workflow

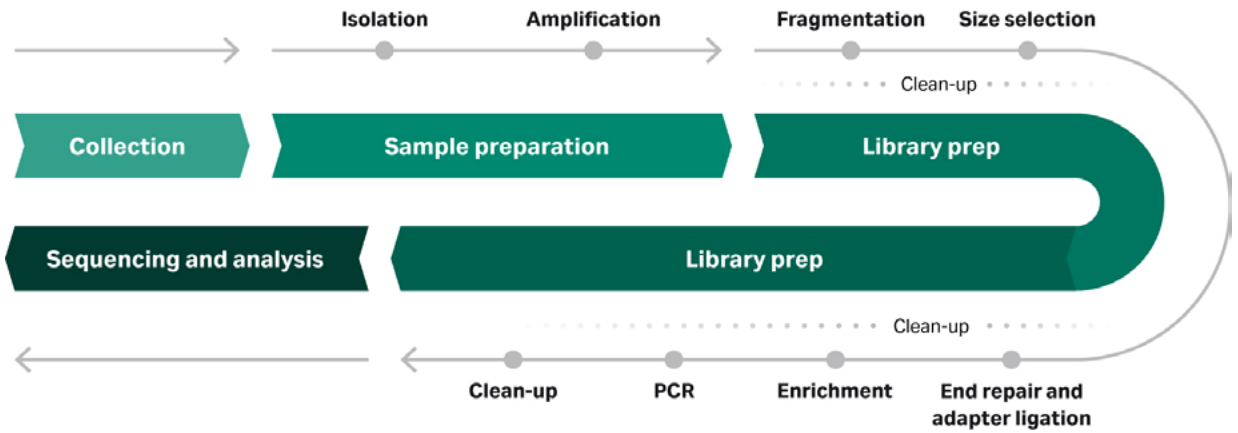
cytiva

Contents

Introduction

Nucleic acid sequencing technologies have come a long way since the first Sanger sequencing of the human genome. As the capabilities of sequencing have increased, so too has the range of applications and the questions answered. Next-generation sequencing (NGS) provides the ability to parallel sequence on a large scale which has had a massive impact and has enabled biological questions at the level of the genome to be addressed. NGS technology now underpins genomics research enabling researchers to sequence an entire genome, transcriptome or exome carrying out experiments and analyzing data which until recently would have been considered impossible. Sequencing supports a diverse range of applications including clinical diagnostics and other aspects of medical care ranging from disease risk to therapeutic interventions.

The next steps are to streamline this process and lower the overall cost to make it more accessible to all markets. We hope to be at the forefront of the simplification and accessibility of the NGS technologies for day-to-day research, clinical diagnostics and much more.



Clinical NGS initiatives: making an impact with sequencing sample preparation

Donald Green, Field Application Scientist, Genomics and Diagnostic Solutions, Cytiva

As next-generation sequencing (NGS) makes inroads into clinical applications such as cancer diagnostics, we look at some of the NGS initiatives and collaborations aiming to change the landscape of clinical sequencing and improve patient's lives.

In December 2018, Genomics England announced that they had reached the main goal of their ambitious "100 000 Genomes Project"—to sequence 100 000 genomes. The results are already making an impact on the lives of people with cancer and a range of other diseases.

The aims of Genomics England aren't unique. All over the world, huge NGS initiatives and collaborations are launching or ongoing—all aiming to advance our understanding of genomics and improve patients' lives through the power of DNA sequencing.

We look at the status of next-generation sequencing (NGS) in the clinic in our white paper. Here, we've picked out some of the DNA sequencing initiatives, where today's scientists work together to change outcomes for patients tomorrow. Who are they, and what are trying to achieve?

100 000 ways to improve molecular diagnostics

One of the key aims of the 100 000 Genomes Project is to provide diagnostic tools

for patients with rare diseases: those that have been difficult to diagnose in the past. These patients have often gone through 'diagnostic odysseys' as doctors struggled to uncover a cause for their condition.

Launched in 2013, the 100 000 genomes project didn't just look at DNA containing harmful mutations, it also created healthy reference genomes. Cancer patients, for example, had both their healthy and tumor DNA sequenced. For inherited diseases, the project used parental genomes for comparison.

Genomics England expects to present the final results to the UK's National Health Service (NHS) in 2019. But early results from this NGS initiative have already identified causative mutations in patients with previously undiagnosed conditions, enabling more targeted treatments and often ending years of uncertainty.

Setting the standard for cancer diagnostics

As the use of clinical NGS expands and the number of available genetic tests increases, there's an emerging need for improved standardization and regulatory oversight.

One collaboration that's working on addressing this issue in the field of cancer sequencing is the Actionable Genome Consortium. Its central goal is to work towards a clearer definition of an 'actionable cancer genome'. It aims to set out clear standards that define tumors (and their treatments) by genetic makeup.

The collaboration began in 2014 and involves sequencing giant Illumina and four major US cancer centers: the Dana-Farber Cancer Institute, the Fred Hutchinson Cancer Research Center, the MD Anderson Cancer Center, and the Memorial Sloan Kettering Cancer Center.

The added value of this collaboration is that these major cancer centers have large, multidisciplinary cancer boards. Their know-how can help other clinicians working in cancer diagnostics assess the clinical significance of complex NGS data.

This collaboration is also already having an impact on the way organizations develop new sequencing panels.

Assessing the broader impact of clinical NGS

When it comes to gathering clinical sequencing data, the central paradigm is often "more is better." But this quest for retrieving more and more information about our genomes overlooks a concern that many people have in our society: does more information always improve well-being?

This concern is one of the aspects of sequencing that the BabySeq project is investigating. BabySeq is a randomized clinical trial, in which scientists are sequencing the genomes of around 150 babies in the treatment group and comparing to a control group where no sequencing takes place.

The aim of this approach is to investigate the broader impact of whole genome sequencing on the well-being of both babies and their parents. Alongside data on the babies' health and the care they receive, the research also considers answers received from questionnaires given to parents about how access to their child's genetic information affects their family life.

Researchers hope to use this information to gain an insight into the effects of DNA sequencing that might otherwise be overlooked with a purely clinical approach.

These are a few examples of NGS initiatives where clinical research using sequencing is telling us about more than just the mutations we have in our genes. Read our white paper for more information on current trends and applications of clinical NGS.

Donald has worked in a variety of disciplines related to the molecular field including Biomedical Engineering research at Mississippi State and Autism Spectrum Disorder studies at the University of Mississippi Medical Center. More recently, he has worked for well established life sciences organizations, gaining experience in clinical diagnostics, lab developed tests, and helping develop start-ups while focusing on Next Generation Sequencing (NGS) and targeted sequencing in Oncology.

Fundamentals of NGS sample preparation

Andrew Gane, Strategy & Technology Manager, Genomics and Diagnostic Solutions, Cytiva

Next-generation sequencing (NGS) has enabled us to extract genetic information from samples faster, more reliably, and at lower cost than ever before. Getting your DNA ready for sequencing requires the preparation of a sequencing library as well as a few other steps that depend on the type of sample and the NGS platform.

In this blog we cover the fundamentals of preparing your samples for NGS, as well as considerations for each step: DNA extraction, amplification, library preparation, selection or purification, and quality control.

Next-generation sequencing: DNA extraction protocol

The first step in every sample prep protocol is extracting the genetic material– DNA or RNA– from cells and tissues. Other molecules, such as RNA and proteins, interfere with the sequencing process and must be removed before doing anything else. The specific tissue type and storage conditions determine the details of this extraction process.

Extraction entails breaking down the extracellular matrix and opening the cell membranes using enzymes, solvents, or surfactants. The DNA in the resulting mixture must then be isolated.

The traditional gold standard in DNA isolation is phenol-based extraction. Phenol is a hydrophobic solvent that denatures and dissolves proteins, removing them from the DNA-containing aqueous phase. However, it can be tricky to work with, and users need to be careful not to contaminate the aqueous phase with phenol.

Spin columns that specifically bind DNA provide an alternative and are an easy-to-use, but more expensive, method to wash away the debris. Chloroform-based extraction, another alternative, enables you to isolate high-quality DNA without phenol, and commercial kits can include a resin that minimizes the risk of contamination.

Next-generation sequencing: amplification methods

Amplification after extraction is optional, depending on your application and sample size. For example, whole genome sequencing (WGA) with 2 µg of starting material does not necessarily require further amplification. But, with nanograms—or even picograms—of starting material, amplification becomes essential to obtain sufficient coverage for reliable sequence calls.

Isothermal amplification and polymerase chain reaction (PCR) are two common methods to increase the amount of input DNA. PCR uses generic primers to amplify the starting material in a highly uniform manner, but tends to be more error-prone than multiple displacement amplification (MDA).

MDA is an isothermal method, often based on Phi29 polymerase, and excels in accuracy with low rates of false-positives and false-negatives. MDA's main drawback is overrepresentation of some regions of the genome.

More recently developed hybrid methods, such as MALBAC, aim to correct this issue with MDA, but these methods also rely on PCR, and have some of the same associated drawbacks.

The different advantages and disadvantages of these methods mean that each is better suited to detect some features over others. For example, MDA outperforms the other two methods in detecting single-nucleotide variants (SNVs), whereas PCR and MALBAC are better for studying copy number variation (CNV), as described in this Nature review article.

DNA library preparation for next-generation sequencing

Most NGS platforms analyze DNA in uniform, bite-size pieces, created by DNA fragmentation. This process generates a 'library' of fragments with a narrow length distribution that is optimal for the sequencing platform.

DNA fragmentation

Both mechanical fragmentation (shearing) and enzymatic methods are suitable for NGS. Mechanical methods enable random shearing to produce a variety of overlapping fragments for any given region of the genome. This is ideal for de novo assembly.

Enzymatic methods are relatively fast and require less investment upfront but have some 'bias', cleaving some sites preferentially, making de novo assembly more challenging without the variety of overlapping fragments.

DNA end-repair

The fragments generated have single-stranded, 'sticky' ends. The next step, end-repair, fills in these sticky ends to create blunt ends, ready for adaptor ligation.

Adaptors

Adaptors are then bound to both the 5' and the 3' ends of the library fragments. They are specific to the sequencing platform, but ultimately all serve to enable in-platform clonal amplification, i.e. Illumina's bridge amplification or BGI's rolling circle amplification.The adaptors are designed to bind to the sequencer-specific substrate, such as a patterned flow cell, contain sequences to enable amplification, and can have barcodes for fragment identification.

Targeted sequencing

These library preparation steps are generally applicable to whole genome sequencing. If you're looking to perform targeted sequencing, library preparation differs.In amplicon-based target enrichment, the fragmentation and end-repair

steps tend to be unnecessary. Pulling the targeted regions out as amplicon fragments with blunt ends enables you to go directly to adaptor ligation. Hybridization-based enrichment does require fragmentation. The hybridization probes pull out the regions of interest from the library of overlapping fragments, ready for end-repair.

DNA sequencing: size selection and purification

To speed up your workflow, it might be necessary to 'clean up' your library before sequencing by removing fragments that won't produce relevant data. For NGS workflows that have narrow size requirements, discarding fragments that are either too large or too small to produce useful results can improve sequencing efficiency.There are different protocols for size selection, which might involve gel electrophoresis or magnetic bead-based selection. Magnetic beads also provide a quick and easy method for final clean-up.

DNA quality control

A final step before proceeding to sequencing is to confirm the quality and quantity of your DNA. Both parameters contribute to the confidence in your sequencing data. You can measure the quantity of your DNA using fluorescence- or qPCR-based methods.

For qualitative validation, many protocols use the Agilent TapeStation™ or Bioanalyzer™. Have a look at our blog on the challenges in NGS sample preparation for possible solutions for quality or quantity issues.

These are the basic steps that researchers use to prepare DNA for sequencing. You can find more information about specific NGS workflows and applications in our other NGS blogs.

Andrew is the the Product Strategy and Technology Manager within the Genomics and Diagnostic Solutions business responsible for building the innovation pipeline in collaboration with the R&D and commercial teams. His knowledge and understanding of emerging trends and new applications has been fundamental to developing the product portfolio into workflow based solutions, with a particular focus on next generation sequencing. Andrew has more than 30 years' experience in immunodiagnostics and molecular diagnostics in both lab-based and product development roles.

Unraveling the challenges of nucleic acid isolation

Chike Ejiofer, Field Application Scientist, Genomics and Diagnostic Solutions, Cytiva

Explore the challenges of DNA extraction (and RNA extraction) from a variety of common sample types, and the methods and technologies that facilitate this, including magnetic beads, which provide a versatile, high capacity solution.

Analyzing nucleic acids is enormously powerful, providing us with insight into a variety of biological processes for basic research and clinical applications. DNA isolation (and RNA isolation) is the first step for many modern genomics techniques and applications, which require high-quality starting material free of contaminants.

For lab managers complexity remains at the heart of nucleic acid extraction. You could say there are both too many and too few choices out there. What is the 'right' isolation protocol for your sample or application?

In this blog, I'll explore the challenges of isolating nucleic acids from a variety of common sample types and pick out several approaches, highlighting magnetic bead DNA extraction, which has become one of the most versatile methods for nucleic acid isolation.

How to extract DNA (or RNA)

Genomic DNA extraction is the first step in many molecular biology studies, and all recombinant DNA techniques. Protocols involve breaking open the cells and separating the DNA you need from other nucleic acids and cellular components in the sample, while also keeping it in good condition for downstream analysis.

There are several approaches that you might take, varying from gentle to aggressive. The choice depends on several factors, including the target DNA, source organism,

the type and quality of your starting material, and the application. They generally all share three common steps: lysis, contaminant removal and DNA recovery.

Step 1: Lysis

Cell lysis involves chemical, mechanical, or enzymatic disruption of cell membranes and denaturation of proteins. The exact method depends on your starting material. Bacteria, mammalian cells, plant cells, and human tissues all might require a slightly different approach.

'Gentle' lysis might involve using a detergent, such as sodium dodecyl sulfate (SDS), or enzymes to break up cell membranes; aggressive lysis might take the form of homogenization to physically break open cell walls.

Step 2: Removing contaminants

You can use both solution-based and solid-phase methods to separate DNA from unwanted lysis debris and potential contaminants. Phenol chloroform DNA extraction, for example, separates water-soluble DNA and denatured proteins into different phases. This is cheap, but slow, and risks carryover of phenol that can affect downstream applications. Solid-phase extraction binds DNA to a column or bead surface. Silica resins or silica-coated magnetic beads, for example, use chaotropic salts to disrupt hydrogen bonds and bind nucleic acids, enabling contaminants to be washed away. Oligonucleotide-coated resins can also add a level of specificity, but column kits can quickly add up in cost.

Step 3: Recovering the target nucleic acid

Downstream applications require your DNA in a suitable format (solvent and concentration). Often, this will be just a matter of precipitating your DNA with ethanol, washing, and resuspending in an appropriate buffer. For solid-phase methods, it will first require adjusting the pH or salt concentration of the buffer to release the nucleic acids.

Sample-specific DNA isolation challenges

Cultured mammalian cells and tissues

Cultured cells are relatively easy to lyse with osmotic shock or detergent treatments, while isolating DNA from tissue requires breaking down the extracellular matrix, not just cell membranes. This often requires homogenization followed by silica column (e.g. GenomicPrep kits) or mag bead-based (e.g. SeraSil-Mag™) purification, or less favorable phenol-chloroform extraction.

Using formalin-fixed, paraffin-embedded (FFPE) tissue is common in clinical applications and some research studies. It's excellent for preserving tissue structures, but can introduce all sorts of DNA damage with profound effects. That is, as the quality of the DNA isolated directly affects the assay results, positive samples might be overlooked simply because of poor extraction.

Blood

A challenge of DNA extraction from blood is the variability in DNA quantity depending on blood fraction. Red blood cells don't contain DNA, so there's much less per cell in whole blood compared to buffy coat or bone marrow-derived fractions.

Blood coagulation also presents challenges: clotting can prevent effective sample digestion, and some anticoagulants can interfere with PCR amplification.

Bacteria

There are differences between gram-positive and gram-negative samples in DNA extraction from bacteria. Gram-positive samples usually require lysozyme treatment to digest the higher levels of peptidoglycan in the cell wall, whereas for gram-negative samples, a simple osmotic shock might be enough.

DNA is unlikely to be scarce with either type, and it's common to use fast methods, like alkaline extraction and diatomaceous earth, to extract the DNA. Both methods are reliable, but alkaline extraction might not provide the highest purity by itself, and diatomaceous earth can be high cost.

Plant material

Plant cells can be embedded in a tough matrix and have cell walls consisting of glycans and cellulose that are difficult to break. The solvent-based cetyltrimethylammonium bromide (CTAB) extraction method is common for plant material, but it is an aggressive approach. It uses harsh chemicals, is laborious, and often requires further clean-up and optimization for different samples and applications.

To bead or not to bead for DNA extraction

Magnetic beads provide an excellent alternative to traditional isolation and clean-up methods due to their versatility and ease of use. They don't require additional centrifugation of a potentially already agitated sample, improving the likelihood of recovering larger fragments, and can be scaled up to have a higher binding capacity than columns.

Using magnetic beads is straightforward, needing no hazardous solvents, and releasing the DNA or RNA is just a matter of adjusting the buffer properties (Figure 1). This simplicity also makes magnetic beads well suited to automation in high-throughput applications.

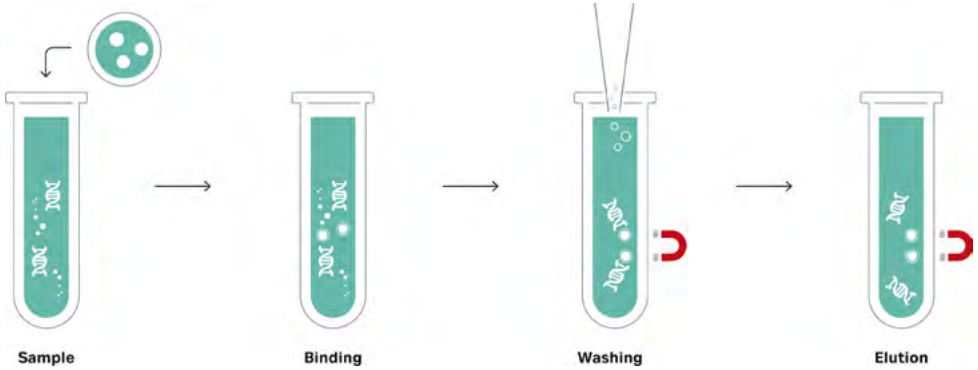


Figure 1. *The principle of magnetic beads for nucleic acid isolation.*

SeraSil-Mag silica coated magnetic beads are an appropriate example. They help address several challenges in DNA extraction and clean-up I've mentioned here, and suit a range of applications, including all the sample types I've described, when used with appropriate buffers. Their binding capacity and tight size distribution deliver highly reliable results while being easy to use without centrifugation.

Chike is a Senior Field Applications Specialist for Genomics and Diagnostics Solutions, focusing on supporting the commercial teams in their understanding of the NGS workflow, DNA sequencing, RNAi and Real-Time PCR. Chike has worked in the healthcare industry for over 15 years, after spending a year as a Research Scientist at the University of Westminster, where he focused on validating ATG-like genes in T. Vaginalis.

The enzymes making the cut in NGS library preparation

Angeliki Achimastou, Modality Specialist, Genomics and Diagnostic Solutions, Cytiva

Enzymes play key roles in many applications, including NGS. See how a range of enzymes from your molecular biology toolbox make the NGS library preparation workflow easier.

Back to basics: what are enzymes and what do they do?

Enzymes are macromolecular biological catalysts (usually proteins) that accelerate virtually all chemical reactions in cells. They work by reducing the activation energy needed for reactions that might otherwise take a long time, or not take place at all.

An enzyme's active site binds one or more substrates in such a way that it minimizes the necessary energy input. Molecules enter the active site, and the enzyme facilitates the chemical reaction. The speed of these reactions is determined by the action of the enzyme.

Some enzymes bind a range of substrates while others are quite selective, either through the shape of the active site or associated targeting molecules, like oligonucleotides.

In next-generation sequencing (NGS) sample and library preparation, you probably use enzymes at almost every critical step. They can digest tissue and other unwanted cellular material, degrade or reverse transcribe RNA, amplify starting material, cut DNA to optimum fragment lengths. The list goes on.

The enzymes in NGS library prep

A typical NGS library preparation workflow has multiple steps, several of which involve enzymes:

- Extracting and purifying DNA (or RNA) from a sample
- Fragmenting DNA (if necessary)
- Size selecting fragments of optimum lengths for sequencing
- Fragment end-repair
- Ligating adaptors
- Amplifying the library (if necessary)
- Quantitating and pooling for sequencing

Using enzymes for DNA extraction

Getting a purified sample can be a messy business. Depending on the starting material, you might have challenges with yield, integrity, and purity.

Tissue samples, for example, might be a plentiful source of nucleic acids, but preservation methods can damage DNA. They also often need homogenization to break down the extracellular matrix for DNA or RNA extraction, risking further unwanted damage and fragmentation.

Single-cells, on the other hand, provide a limited yield. The nucleic acids require amplification to be usable for NGS or any other genomics application.

These are both situations where enzymes can be useful. Proteases can help degrade nucleases and other proteins, though it's also common to have DNases or RNases in lysis buffers to degrade the unwanted nucleic acid type.

Commercial kits containing cocktails of repair enzymes can restore damaged DNA, such as that from formalin-fixed paraffin-embedded (FFPE) samples, to a more useful state.

When you're dealing with limited or insufficient yields, where extraction alone is impractical, whole genome amplification (WGA) can help enable otherwise impossible analyses. Using Phi29 DNA polymerase and random primers for multiple displacement amplification (MDA) generates micrograms of high molecular weight DNA from picograms of starting material, providing a simpler alternative to PCR-based methods.

Library fragmentation

Part of generating a high-quality library for NGS involves making sure fragment sizes are within a range centered around an optimal average length, typically a few hundred base pairs for Illumina™ systems.

There are several options for DNA fragmentation, including mechanical, enzymatic, and chemical approaches. Mechanical methods might involve hydrodynamic shearing by sonication, focused acoustic shearing, or nebulization. But each comes with caveats:

- Sonication is easy and effective, but requires careful calibration and can be slow compared to other methods.
- Focused acoustic shearing provides tight size distribution, but requires specialized equipment with potentially high upfront costs.
- Nebulization is fast, but produces a rather wide size distribution and degrades much of the sample, and so requires high input.

All mechanical methods also have the potential to introduce unwanted DNA damage as they won't necessarily make clean breaks.

Enzymes provide a scalable and cost-efficient alternative. They don't require any complex equipment, meaning upfront costs are low. Enzyme reactions are generally quite gentle reactions compared to mechanical shearing, minimizing unwanted sample degradation or DNA damage.

Reaction volumes can also be as small as a few microliters, and there's little risk of losing a precious or limited sample.

The enzymes used for DNA fragmentation fall into three types: restriction enzymes, nicking enzymes, and transposases:

1. DNA Fragmentation: Restriction Enzymes

Restriction enzymes (or restriction endonucleases) are essential tools for all labs working with recombinant DNA. There is a large variety available, typically targeting 4 to 8 base pair sequences, and they serve a useful function in NGS library prep.

On binding their recognition site, restriction endonucleases create either blunt-ended or overhanging double-stranded breaks, the latter requiring end-repair by fill-in (e.g. by the Klenow fragment of DNA Polymerase I) before adaptor ligation (Figure 1A).

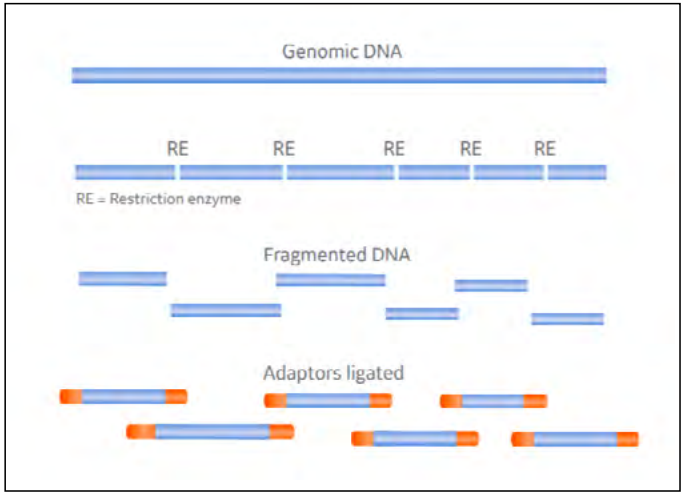
By selecting an enzyme with a recognition site that appears roughly as often as the target average fragment length, you stand the best chance of creating an even distribution of fragments for NGS.

But these recognition sites are also the main drawback of restriction digests. They will introduce some fragmentation bias (i.e. fragmentation is not random). This isn't an issue in all DNA sequencing applications but means some genomic regions might have lower coverage than others, a challenge for applications reliant on deep sequencing.

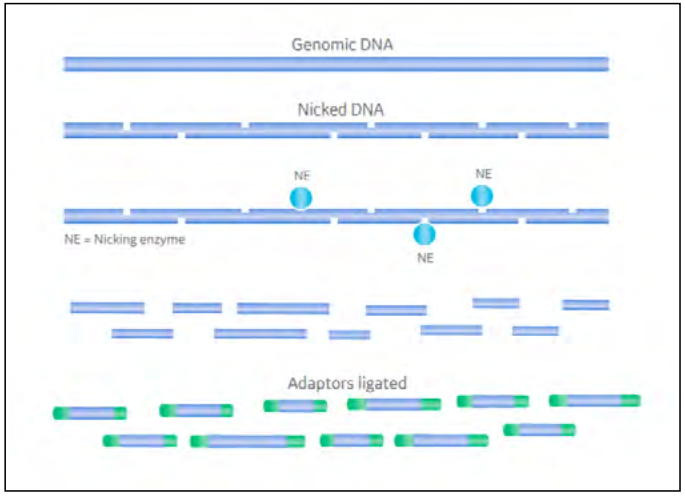
2. DNA Fragmentation: Nicking Enzymes

An alternative to the potentially biased fragmentation introduced by restriction enzymes is the use of nicking enzymes. DNase I, for example, can make random single-stranded cuts in the DNA. A second, single-strand-specific enzyme that recognizes nicked sites then cleaves the second strand.

The result is a distribution of fragments with short overhangs that need fill-in before adaptor ligation, but are otherwise unbiased (Figure 1B). There's also potential here for modulating the average fragment length by varying reaction conditions and time.



A



B

Figure 1. *Comparing restriction endonuclease- and nicking enzyme-based DNA fragmentation in NGS. Genomic DNA fragmented at specific recognition sites by restriction enzymes in one step (A) and random sites by nicking enzyme and single-strand-specific endonuclease in two-step process (B).*

3. DNA Fragmentation: Transposomes

Whereas using restriction and nicking enzymes depends on cutting specific or random sites in the genome and performing end-repair, transposon-based fragmentation can both cleave DNA at random sites and insert a short double-stranded oligonucleotide on both ends. These are then ready for index and adaptor ligation without further processing.

Illumina's Nextera™ kits use this 'tagmentation' approach to produce libraries compatible with Illumina technology in one step.

But one challenge with the use of transposases is that the reaction (and so the quality of the library) is sensitive to the amount of starting material. Each transposase only works once, so the average fragment length is critically dependent on the DNA:transposome ratio, though this does provide a way of modulating the target fragment length too.

Enzymes in DNA fragment end-repair and adaptor ligation

Adding adaptors to library fragments first requires clean blunt ends with a single-nucleotide 3' A-tail amenable to ligation. As mechanical shearing and enzymatic methods tend to create damaged ends or overhangs, most tend to need repair before ligating adaptors. This doesn't apply to Illumina's 'tagmentation' method though, which fragments and adds short blunt-ended oligonucleotides as part of the same step.

Enzymes are key to end-repair. A typical blunting enzyme mix might, for example, contain T4 DNA polymerase and T4 polynucleotide kinase (PNK). T4 DNA polymerase (in the presence of dNTPs) can fill-in 5' overhangs and trim 3' overhangs down to the dsDNA interface to generate the blunt ends (Figure 2A-B). The T4 PNK can then phosphorylate the 5' terminal nucleotide (Figure 2C).

A-tailing also requires a polymerase. Taq DNA polymerase the most common as it has terminal transferase activity and naturally leaves a 3' terminal adenine (Figure 2D). DNA polymerase I Llarge (Klenow) fragment is another common option, which can also double as a blunting enzyme. Using either of these polymerases leaves A-tailed ends that complement standard Illumina sequencing adaptors.

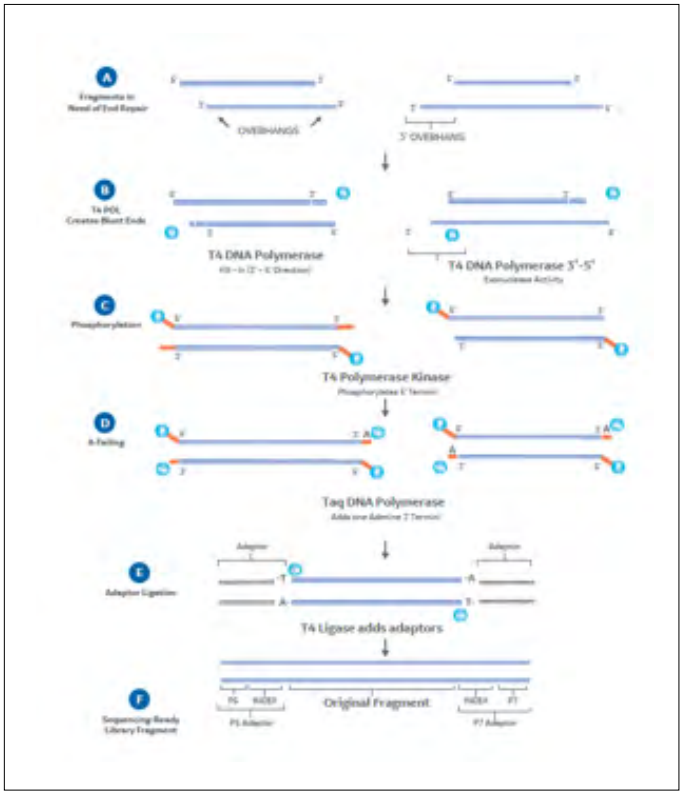


Figure 2. *End-repair and adaptor ligation in NGS library preparation. Fragments with 5' and 3' overhangs (A) are filled-in by T4 DNA polymerase to create blunt ends (B). T4 PNK then phosphorylates 5' termini (C) and Taq DNA polymerase A-tails 3' termini (D), leaving ends amenable to adaptor ligation. T4 ligase adds sequencing adaptors (E) to leave complete sequencing-ready library fragments (F).*

Adding an adaptor at this stage just requires an incubation with T4 DNA ligase. This enzyme will join both blunt and so-called 'sticky' ends, in this case catalyzing the formation of a phosphodiester bond between the 5' and 3' termini of the end-repaired fragments and sequencing adaptors (Figure 2E-F).

Each of these steps must be accurate and efficient for the library to produce reliable NGS data, and so relies on using high-quality enzymes under optimum reaction conditions. The result is a library of fragments that might need quantitation and pooling, but is otherwise ready for sequencing.

Enzymes in DNA amplification

The last step in NGS library prep that might involve enzymes is amplification by PCR. This step isn't always necessary. High-quality samples producing high yields or single-cell samples already amplified by MDA are unlikely to need a further amplification step.

PCR does serve a dual purpose though. As well as amplifying, in some protocols it's needed for adding the functional elements. For example, Illumina's 'tagmentation' approach has a reduced-cycle PCR step for adding sequencing adaptors. The adaptors bind and extend from the short oligonucleotides at the ends of fragments.

Whether you're using PCR for adding indexes, or also for amplification, it's essential that it be error-free. That requires using a particularly high-fidelity thermostable DNA polymerase with excellent proof-reading capabilities and appropriate reaction conditions. This approach maximizes amplification efficiency, and minimizes both amplification bias and the risk of introducing sequencing artifacts.

These factors combined ultimately help generate uniform coverage, even across high G-C and other difficult regions.

So, as you can see, enzymes are crucial to multiple steps in NGS sample and library preparation. They provide shortcuts for slow or difficult reactions, and enable us to modify, repair, and amplify nucleic acids for a variety of applications, particularly NGS.

Angeliki is a European Modality Specialist for the Genomics and Diagnostic Solutions business. Her current role includes account management, working with R&D on new product development and collaborating with the customized solutions team on bespoke customer projects. Prior to Cytiva, Angeliki worked in other product specialist roles in life sciences and was a post-doc at the National Institute for Medical Research (NIMR) now the Francis Crick Institute.

Size selection brings better data to NGS workflows

Andrew Gane, Strategy & Technology Manager, Genomics and Diagnostic Solutions, Cytiva

High quality libraries are key to keeping next-generation sequencing costs down and maximizing usable data. Take a closer look at how size selection can improve your data quality, and the methods you can use in your library construction workflow.

NGS sample preparation, cost, and data quality

The introduction of next-generation sequencing (NGS) technology resulted in a fundamental shift in our approach to genomics. Even now, more than a decade after second generation sequencers arrived, the market continues to grow.

This is partly because of the constant drive to reduce the cost of sequencing and open up the technology to more researchers and applications. Despite these year-on-year cost reductions, individual sequencing runs remain expensive. Maximizing the usable data from any given run, which can be achieved by optimizing upstream library construction and sample preparation steps, can lead to additional savings.

These processes are relatively inexpensive and have substantial influence on final data quality. Here's why library fragment size selection is a key step towards data quality, and my recommendations on the main methods for carrying out size selection, their advantages and disadvantages.

What does a typical NGS sample prep look like?

Although there are multiple approaches to sequencing, Illumina's sequencing-by-synthesis approach continues to be the most widespread. We've previously discussed the fundamentals of NGS sample prep, which has several common steps

for library construction, including:

- Fragmentation through enzymatic or mechanical means.
- End-repair and processing to homogenize the heterogeneous fragment ends.
- Adapter ligation for cluster generation and in-cell clonal amplification.
- Size selection to remove suboptimal fragment sizes and any adaptor dimers.

The significance of size selection

Genomic sequencing relies on having high quality libraries. Part of this is making sure library fragment sizes are within the optimum range for a given instrument, typically 200-500 bp for Illumina™ systems. This range is a consequence of the effect of fragment length on cluster generation and the efficiency of the sequencing process itself.

Small fragments tend to cluster more efficiently on the flow cell than larger fragments. A bias towards smaller fragments leaves much of the sequencing capacity unused. Selecting fragment sizes below 150 bp can risk carryover of unwanted adaptor and primer dimers, the sequencing of which leads to a lot of unusable data and further wasting of capacity.

Fragments larger than optimum pose the opposite challenge. Although it's possible to sequence fragments >1 kb in length, this is inefficient and prone to errors—an issue that third generation sequences attempt to solve.

Individual samples might also have different shearing profiles, with narrow to wide distributions. Setting an instrument up for 600 bp fragments when there is a 200-1,000 bp distribution, for example, means that many of the sequencing templates won't be viable or read to sufficient depth. This produces little useful data and low uniformity of coverage.

A size selection step enables you to take a randomly fragmented library and pull out only the fragments fitting the optimal/target range for the instrument and application (Figure 1). This saves time and cost by maximizing the efficiency of sequencing runs.

A note on DNA fragmentation methods

There are various options for fragmentation, some of which attempt to bypass the need for size selection altogether. The choice of method may depend on your application, starting material, and equipment available.

Enzymatic methods tend not to be completely random, but provide some control over fragment sizes through varied incubation times. However, these are less well suited for de novo assembly due to the likelihood of making fewer overlapping fragments.

There are various options for mechanical shearing, which use sonication or focused acoustic technologies. These are random, and can be tuned to produce predictable shearing profiles.

Size selection methods

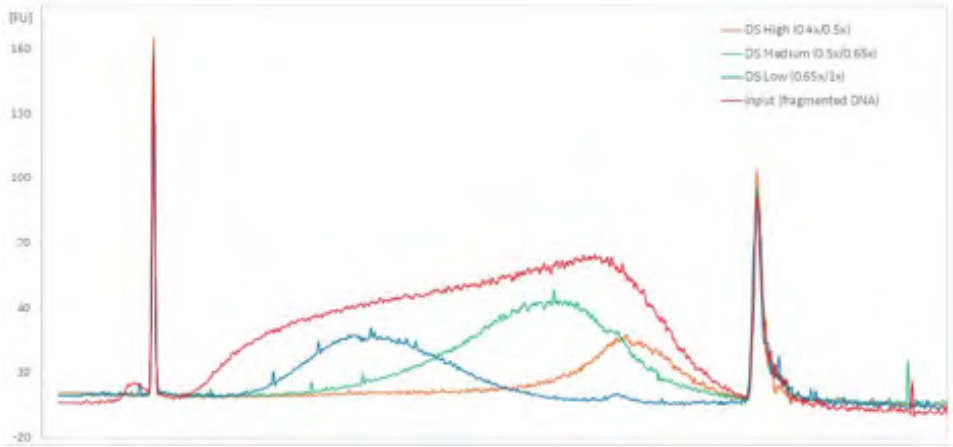


Figure 1. *Enzyme fragmented DNA with dual size selection*

The approaches to size selection include enzymatic, gel-based, and magnetic bead-based methods, the suitability of each depending on the needs of the experiment. These also provide an opportunity to clean up adaptor dimers and any other leftover reagents.

Enzymatic approach

Illumina's Nextera™ kits produce libraries for various applications compatible with Illumina technology in one step.

When launched, they attempted to get around the need for size selection by using transposon-based fragmentation and tagging, known as 'tagmentation', saving several workflow steps. However, library profiles tended to be broad, leaving users often reverting to a separate size selection step.

Nextera kits now include magnetic bead-based size selection reagents.

Gel-based approach

Gels have long been used for nucleic acid purification, enabling you to physically remove the chosen fragment size. Gel-based systems, such as Sage's Pippin Prep™, help automate this process, but have inherently limited throughput. A typical 96-sample batch requires close to 10 hours to process.

Magnetic bead-based approach

The introduction of magnetic beads for convenient and high throughput size selection and clean-up has transformed NGS workflows, with Sera-Mag beads integral to this success.

Originally developed for the isolation of PCR products, these beads have polystyrene cores covered in magnetite and a layer of carboxyl molecules. Nucleic acids bind to them reversibly in the presence of polyethene glycol (PEG) and salt; a process known as solid phase reversible immobilization.

The beads are otherwise inert and have high binding capacities, due to large surface areas. The size of fragment bound can be adjusted by simply altering the volumetric ratio of PEG/salt/beads to DNA. From a practical point of view, this bead chemistry makes it straightforward to size select a very specific range of fragments consistently and reproducibly.

The magnetic bead-based approach is well suited for high throughput applications with automation, and the cost of reagents is also low compared to other approaches. These properties make magnetic beads a simple solution for optimizing NGS sample prep.

Andrew is the the Product Strategy and Technology Manager within the Genomics and Diagnostic Solutions business responsible for building the innovation pipeline in collaboration with the R&D and commercial teams. His knowledge and understanding of emerging trends and new applications has been fundamental to developing the product portfolio into workflow based solutions, with a particular focus on next generation sequencing. Andrew has more than 30 years' experience in immunodiagnostics and molecular diagnostics in both lab-based and product development roles.

Tackling the challenges of multiplexing in NGS

Andrew Gane, Strategy & Technology Manager,
Genomics and Diagnostic Solutions, Cytiva

Library multiplexing helps drive down the cost of NGS, but doing so creates a new set of challenges: index misassignment and parallel sample preparation. Solve these and improve your workflow and data quality, reducing your cost per sample.

The need for multiplexing

Next generation sequencing (NGS) is fast. Runs capable of sequencing an entire genome are measured in just hours, rather than days or even the years needed for the human genome project. The capacity and availability of NGS technology is also greater than ever, with hundreds of millions of reads producing hundreds of gigabases of data in a single run.

No single genome sequencing experiment requires such vast capacity, however, with most read depth needs comfortably covered by a fraction of that capacity. As the cost of an individual run is still substantial, making use of this full capacity by multiplexing hundreds or thousands of libraries enables us to continue driving down the cost of sequencing a genome.Multiplexing is not without its challenges, though. Let's look at the challenges of index misassignment and parallel sample preparation, and how we can address them to improve both workflow and data quality, ultimately reducing your cost per sample.

The challenges associated with multiplexed sequencing

Despite the goal of making each sequencing run as productive as possible, any approach that involves scaling up a workflow will introduce some unique challenges. In the case of NGS multiplexing, these can be found in both the library preparation and the sequencing data analysis stages.

Index misassignment

Multiplexing requires tagging each library with indexing barcodes before pooling together and running on a single patterned flow cell. This allows the reads for each library to be identified and separated out by software from the bulk sequencing data. This might be only a handful of libraries, or thousands, each with unique indexes.

Sometimes, however, indexes are misassigned, meaning that the read for a molecule in one library is mistakenly assigned to another, complicating the analysis. These events are fairly uncommon (perhaps 1-2%, though can be as high as 10%). Given the sheer numbers involved, this can add up to many misassigned reads.

This is a particular issue in low frequency allele detection, where it may be impossible to distinguish true and false positives, or where sample might be scarce and you need to maximize yield.

Tackling index misassignment

Index misassignment is largely attributed to 'index hopping,' a process where contaminating free adapters and index primers are thought to bind clustered molecules in flow cells. These can extend during clonal amplification and produce reads with another library's index. Index hopping appears to be more common with Illumina's ExAmp clustering chemistry, compared to the older bridge amplification method, according to tests conducted by Illumina and the NGS community.

Misassignment is a well-known problem, and now largely addressed by dual indexing and the use of unique molecular identifiers (UMIs). This approach enables the software to discard the data for any reads that do not have the correct combination of two indexes or UMIs.

Parallel sample preparation

Parallel sample preparation of multiple individual libraries for multiplexing requires extra time and resources, and adds potential sources of error.

Challenges in sample preparation are nothing new, of course, and the work is compounded by the multiplex rate. Preparing a handful of libraries manually over several days was acceptable just a few years ago. Now, you might need to prepare hundreds or thousands of samples in that time to maintain a competitive cost per sample. This is all while performing painstakingly accurate quantitation, fragment size identification, and normalization for every library to generate high quality, reproducible data.

Multiplexing libraries with different sized fragments already creates inconsistencies with read depth, as there's a natural bias towards sequencing smaller fragments more efficiently than larger ones. So, pooling needs to be as accurate as possible.

If there are differences in quantitation method, because of time, resource, or equipment limitations, the same experiment using the same sample sources might produce data of different quality. Simple, fast methods, like spectrophotometry, aren't accurate enough and quantitate all nucleic acids, including primers and nucleotides. Electrophoretic methods are good for size determination, but not reliable for quantitation either. Ideally, you would use qPCR, but this is time-consuming.

Normalization is also a source of error, requiring careful and precise handling of volumes often less than 10 µl. Small errors or user-to-user variation here affects data quality and confidence, and repeating preparation and pooling isn't always easy, especially with scarce samples.

So, there is a need for a more practical workflow that doesn't compromise on data quality.

Solving the sample preparation

A certain amount of automation does help alleviate some of the issues with parallel sample preparation. But this is much like putting a more powerful engine in a vehicle that's not aerodynamically efficient. It would make more sense to improve the workflow. Rather than processing everything in parallel, pooling libraries at an early stage and processing them all together would be more practical. This would be an ideal solution, reducing the overall workload and minimizing any sample-to-sample variation introduced by the user.

Key to making this approach successful would be easy or automatic barcoding and normalization of libraries, and a high tolerance for variation in input amounts to avoid compromising data quality. For example, a simple molecular tagging step for each library could pull out equimolar quantities of fragments. This would remove the need for separate quantitation and normalization steps, simplifying and easily integrating into existing workflows.

We are working to resolve the challenges of multiplexed library preparation. Our aim is to develop a practical workflow that maintains high data quality, reduces cost per sample, and doesn't require any specialist knowledge or resources.

Andrew is the the Product Strategy and Technology Manager within the Genomics and Diagnostic Solutions business responsible for building the innovation pipeline in collaboration with the R&D and commercial teams. His knowledge and understanding of emerging trends and new applications has been fundamental to developing the product portfolio into workflow based solutions, with a particular focus on next generation sequencing. Andrew has more than 30 years' experience in immunodiagnostics and molecular diagnostics in both lab-based and product development roles.

DNA library normalization for NGS: why and how?

David Tesin, Modality Specialist, Genomics and Diagnostics Solutions, Cytiva

How and why do you normalize your NGS libraries? Read up on why we need to normalize, best practices for quantitating your libraries (hint: use qPCR), and how magnetic beads enable a new approach to normalization.

What is DNA normalization and why is it important in NGS?

Normalization in next-generation sequencing (NGS) is the process of equalizing the concentration of DNA libraries for multiplexing. Multiplexing helps maximize the use of the ever-increasing capacity of NGS technology, enabling you to run multiple—often thousands—of libraries on a single flow cell, and drive down costs.

These ever-reducing DNA sequencing costs have led to its adoption in an array of molecular diagnostics applications, including reproductive health and oncology, as well as enabling a range of clinical research initiatives around the world.

Both basic research and clinical NGS rely on obtaining reliable data. But uneven library concentrations from different typesand qualities of sample can lead to inconsistencies in data quality (Figure 1).

Those libraries with a high concentration are likely to be overrepresented on the flow cell while those with low concentration are underrepresented. Overrepresentation isn't necessarily a problem, likely increasing read depth, though it does waste capacity. Underrepresentation might result in poor read depth and unreliable data, wasting capacity and potentially your precious sample.

This highlights the importance of normalization in making sure every library is represented equally and sequenced to sufficient depth (Figure 1).

What are the implications of normalizing and not normalizing libraries?

From a cost standpoint, wasting capacity means you end up spending additional work time re-preparing libraries, assuming there is sample available. This time could be better spent on downstream analysis or preparing the next batch of libraries.

From an application and outcome standpoint, analyses and decisions based on potentially inaccurate or incomplete data will at best confuse research results or lead to repeating experiments. At worst, clinicians might, for example, miss key information like a rare allele or single nucleotide variation (SNV) that could have led them down a more appropriate treatment avenue.Normalization helps address these challenges.

How DNA normalization works

Each library prep used in a multiplexed DNA sequencing run is unique in terms of both content and concentration. The final concentration depends on the efficiency of your DNA extraction protocol, and quality and quantity of starting material. Evening out these libraries through normalization helps produce consistent and reliable NGS data.

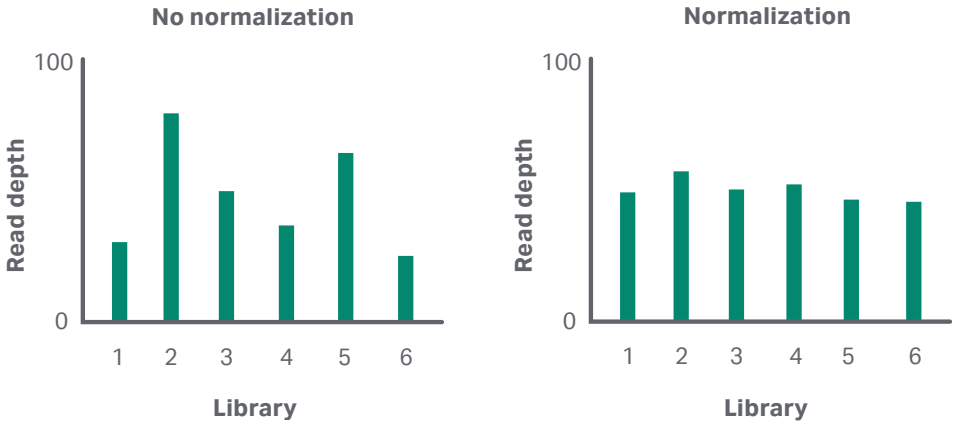


Figure 1. *DNA library normalization addresses the challenge of inconsistent read depth. Variation in read depth without normalization (A), and consistency in read depth with normalization (B).*

There are opportunities for normalization at several stages of a multiplexed sequencing workflow. You might normalize the concentration of input DNA, size distribution of library fragments, and concentration of library prep before pooling.

Checking the concentration of library preps can have a direct effect on clustering efficiency, clonal amplification, and read uniformity across the pooled libraries. So, standard protocols will often involve quantitatively checking individual library preps and adjusting them to equimolar ratios before pooling. This helps to make sure that all libraries are represented equally on the flow cell.

Methods of quantitating NGS libraries

There are several options for quantitating library preps, varying in ease and accuracy.

The quickest and most convenient methods (i.e. spectrophotometry-based) tend not to be that accurate.

The most accurate methods, like quantitative PCR (qPCR), take time and precision, and rely on knowing the average fragment size in each library for dilution calculations (adding more steps to the workflow).

One crucial factor influencing the accuracy of quantitation and subsequent normalization is whether the quantitation method can specifically count adaptor-ligated (i.e. amplifiable) double-stranded DNA (dsDNA) molecules. These are the only molecules that will cluster on the flow cell and contribute to sequencing output.

Illumina's best practice suggests using fluorometric or qPCR-based quantitation with genomic DNA samples in most cases. Table 1 summarizes the common methods for quantitation.

Table 1. *Common approaches to NGS library prep quantitation for normalization.*

Method	Spectrophotometry	Electrophoresis	Fluorometry	Quantitative PCR (qPCR)
Description	Detects the absorption of UV light by molecules in the sample, with concentration calculated against a standard curve. Estimated purity is based on the ratio of measured absorbance at 260 and 280 nm.	Estimates fragment sizes through capillary electrophoresis, and concentration through intercalating dyes.	Uses dsDNA-specific intercalating fluorescent dyes for assessing the concentration of nucleic acids against a standard curve.	Probe-based chemistries use adaptor-specific primers with fluorescent dyes and quenchers to quantitate library preps against standard curves. Digital droplet qPCR is a variation that can provide absolute quantitation without reference samples.
Advantages and disadvantages	Advantages: • Quick and low cost Disadvantages: • Inaccurate, measuring all nucleic acids, not just adaptor-ligated molecules • Not very sensitive • Affected by contaminating RNA and proteins	Advantages: • Accurate for estimating fragment size and distribution Disadvantages: • Quantitation cannot discern between adaptor-ligated and other molecules • Potentially expensive equipment requirements for a single purpose	Advantages: • Sensitive and accurate estimation of concentration of dsDNA • Can also be used to specifically quantitate single-stranded DNA, RNA, and protein • Reasonably fast and low cost Disadvantages: • Cannot discern between adaptor-ligated and other molecules. • Not able to estimate fragment size	Advantages: • Accurate quantitation of adaptor-ligated molecules (viable sequencing templates) • High sensitivity (suitable for quantitation of dilute libraries) • Amenable to automation Disadvantages: • Higher cost and requires more hands-on time than other methods • Not able to estimatefragment size

qPCR provides the ultimate accuracy in quantitation

It's interesting that no single method provides all the data you need with enough accuracy for normalization. Though fluorometry and qPCR enable the most accurate quantitation, neither can estimate average fragment size. So, it's often still necessary to check this by electrophoresis.

Of these two most accurate methods, only qPCR can specifically target the adaptor-ligated molecules. It uses primers complementary to the adaptor sequences. Quantitating only these viable sequencing templates gives you the best chance at normalizing your libraries accurately.

Adaptor ligation efficiency can vary between individual samples and batches. It's reliant on enzymatic reactions that could be affected by impurities and differences in the quality of starting material. So, quantitating with no specificity for adaptor-ligated molecules (fluorometry) means you're more likely to overestimate the sequencing-competent library concentration and over-dilute.

Having said that, if your starting material is of high and consistent quality, and the end repair/adaptor ligation step of your library prep workflow is efficient, fluorometry can be a cheaper, faster, and nearly as accurate an option.

If you're looking for the ultimate accuracy in quantitation though, qPCR is the way to go.

Magnetic beads-based normalization as an alternative

What if you didn't need to go through the trouble of quantitating your libraries at all?

It's increasingly common to find magnetic beads popping up in NGS sample prep workflows. For example, they are already being used for size selection—another challenge in NGS sample prep—providing a reliable and established way to safely handle nucleic acids.

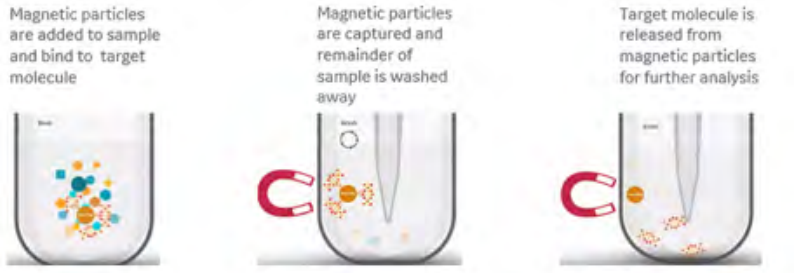


Figure 2a. *Principle of magnetic bead-based normalization of DNA libraries.*

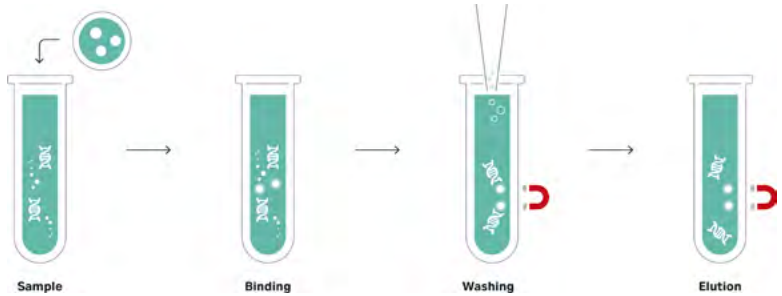


Figure 2b. *Principle of magnetic bead-based normalization of DNA libraries with silica core.*

The idea behind magnetic bead-based normalization is that a given volume of beads can bind a consistent quantity of nucleic acid molecules. That is, if there are enough molecules in each library to saturate the beads, an essentially equimolar quantity of library fragments will bind and be retained from each sample (Refer to figure 2a and 2b). All unbound molecules are then washed away so that each library is represented by just the bead-bound molecules.

There are several coating options available to suit any given application: carboxyl- and silica-coated magnetic beads for generic, non-specific binding based on buffer conditions; oligo(dT)-coated beads for binding mRNA; and streptavidin-coated beads for binding biotinylated samples.

This approach is reasonably straightforward and studies in recent years have indicated that bead-based normalization produces more consistent read depth than several existing quantitation-based methods. Illumina has exploited this approach for normalization, modifying its transposon-based 'tagmentation' system for NGS library prep to use magnetic beads.

The bead-based approach, however, can be wasteful: the number of molecules in each library needs to equal or exceed the binding capacity of the beads, with the excess discarded. If your sample is precious or in short supply, it might be worth taking the extra time for qPCR-based quantitation.

Best practice for selecting a normalization method

- Use fluorometry for library normalization when:
- Your samples are of good quality
- Your sample prep workflow has a history of producing consistent concentrations
- Some variation in quantitation, and so read depth, is acceptable
- Use qPCR for library normalization when:
- Your samples are from varied sources, are precious, or in limited supply
- You need the ultimate accuracy for normalization
- It's essential that you achieve a minimum target read depth
- Use magnetic beads for library normalization when:
- Your samples are in plentiful supply but might vary in quality
- Your library prep yields are usually high (at least 10–15 nM, according to Illumina best practice)
- You have many samples and need to minimize time spent quantitating

David was initially involved with Sera-Mag™ Magnetic beads sales and technical support before transitioning to a Modality Specialist for the Genomics and Diagnostic Solutions business, driving commercial excellence on the East and South coasts of America. Utilizing his extensive knowledge of magnetic bead technology, David was instrumental to the launch of Sera-Mag Select size selection and PCR clean-up reagent. Prior to Cytiva, David spent several years on the bench in molecular biology and protein purification and in business development roles responsible for initiating and developing assay development opportunities with IVD, biotechnology and pharmaceutical companies.

For local office contact information, visit cytiva.com/contact

CY10090-21Oct20-EB

cytiva