

Buyer's Guide:

Next-Generation Sequencing Systems

Covering All the Bases: What to Consider Before Making Your NGS System Purchase.

This is an exciting time for biological research. Meaningful discovery is churning at a pace unmatched throughout history. Thanks to the pioneering world of next-generation sequencing (NGS), the data we can derive from organisms is deeper and more insightful than ever imagined.

NGS equipment, once reserved for the largest and busiest of research centers, is now accessible to enterprises of all sizes — unearthing new knowledge on cancer, microbiology, genetic disease, reproductive health, agriculture and forensics, and other emerging areas. With the advent of the first FDA-approved NGS instrument in late 2013, the technology is now gaining a stronger foothold in the clinical laboratory space for screening and diagnostic testing.

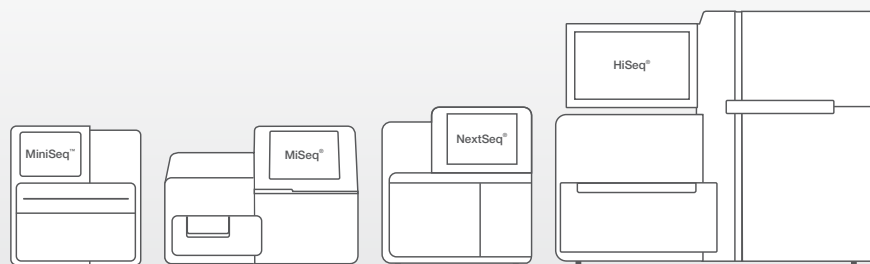
Whether you're seeking to conduct research worthy of publication in a respected journal, or looking for critical clues to the spread and evolution of disease, your work likely requires or would benefit from NGS technology.

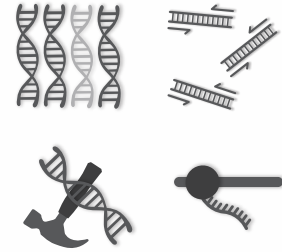
For this reason, labs that don't already own an NGS system are making a push to acquire equipment of their own. For many organizations, it's a matter of data privacy or a need for tight control over their projects and samples. For others, it's simply an issue of efficiency and timesaving. For all, it's seen as an investment in high-quality, reproducible data that leads to valuable biological insight.

If you're a principal investigator or manager for one of these labs, you have an important purchasing decision to make. The good news is that the market today offers plenty of choice. Next-generation systems range from ultra-flexible, high-output instruments that can read multiple samples in a single run, to desktop sequencers ideal for smaller-scale studies.

This guide will walk you through the purchasing process by helping you evaluate your research goals and laboratory needs. Among key questions to consider:

- What applications will I use most?
- What are the best instrumentation options for these applications?
- What are the ongoing costs beyond the initial purchase?
- How will an NGS system affect the lab's workflow and processes?
- How do I ensure quality in data, instrumentation and workflow?
- How will I analyze and store NGS data?
- What level of product support will I receive?





Step 1: Define Your Applications.

First, let's tackle some of the most important questions that will define your product search: What types of NGS experiments do you plan to perform? And which applications will you use most?

Think broadly here, considering your current needs as well as needs in the future. Purchasing NGS instrumentation is a commitment and you want to be sure that your chosen instrument will provide the versatility and the power to accommodate your lab for years to come.

For each application, consider your needs for these three areas:

Throughput per run. Your needs in this area will largely dictate the type of sequencer that's best suited for you. Assess your projected sample numbers per month and year, and let this number guide your selection process. For example, if you plan to run large sample numbers for applications such as whole-genome or whole-transcriptome sequencing, you're probably a good fit for a high-throughput system.

Read length. Longer read lengths are very important for many applications, with de novo sequencing standing out as a prime example. Why? The longer the read, the better the coverage across the genome. And because de novo sequencing does not use a reference genome to align data, it is essential to produce long overlapping reads to complete the assembly. Applications such as microbial sequencing need long reads, as many viruses and other microorganisms have not been previously sequenced. Large structural chromosomal rearrangements in cancer genomes can also be difficult to detect without long reads, as the reads must span the chromosomal breakpoints in order to detect them.

Paired-end sequencing. A broad range of applications can benefit from paired-end sequencing, which allows for maximum coverage across the entire genome. If you currently require or will require paired-end sequencing, be sure to investigate the solutions available for each of the instruments you're considering. It's important that the equipment supplier has established protocols and produces high-quality data with this technique. Easy sample preparation and data analysis tools are also valuable in this area, as a simple and integrated workflow will save you time and money.



Step 2: Understand the Equipment Options.

With your applications defined, it's time to begin exploring the market for next-generation sequencing systems. You'll find there's a full spectrum of options available — depending on your budget, the needs of your laboratory and the nature of your research. Considering the breadth of price points, sizes, and capabilities, it's safe to say that one of the options will meet your needs.

Here's a look at the main categories of NGS instrumentation, along with factors that make each one the right fit for a given lab.

Instrumentation Platforms

Desktop Low-Throughput Sequencer

- Most affordable option in terms of initial purchase price.
- Ideal for smaller- to medium-scale studies.
- Accessible and convenient, allowing for privacy and the ability to work on your own schedule.
- Fast workflow for targeted samples.
- Applications include: Targeted DNA sequencing, targeted gene expression, 16S metagenomics, and HLA typing. Also, small-genome and microbial sequencing. Some larger labs choose to have desktop sequencers in addition to their high-throughput equipment for small-scale follow-up studies.
- Clinical version of this equipment was FDA approved in 2013 for medical screenings and diagnostics.

Desktop High-Throughput Sequencer

- Mid-range price point.
- Accessible and convenient, allowing for privacy and the ability to work on your own schedule.
- Higher throughput sequencer that maintains a small-footprint desktop format.
- Applications include: Exome, mRNA, and whole-genome sequencing, shotgun metagenomics, and HLA typing. Ability to sequence a whole human genome in about a day.

High-Throughput / High-Volume Sequencer

- Most expensive option in terms of initial capital outlay. But for busy labs, it delivers the lowest price per sample.
- Highest-throughput available.
- Applications include: Large-scale whole-genome sequencing, targeted resequencing, whole-transcriptome sequencing, *de novo* sequencing, epigenetic regulation. Also: Exome and mRNA sequencing, shotgun metagenomics, HLA typing.
- Can complete projects in least number of runs for highest operational efficiency.

Using today's highly automated NGS equipment, you can go from experiment to data to publication faster and easier than ever before. The scientific community has used this technology platform to develop a broad range of applications that have transformed study designs, surpassing boundaries, and unlocking information never imaginable.

But not all sequencers are built the same or offer the same features. To help you make smart buying comparisons, let's first look at the best approach to sizing up costs.

Step 3: Think About Budget Holistically, from DNA to Data.

Your budget can help you narrow your decision on the type of sequencer to purchase, but the issue of investment isn't as clear-cut as you might think. There are many important factors to consider beyond the initial capital expenditure of the instrument itself, including the cost of ongoing operations, the quality of the data (ensuring the process won't have to be repeated at an additional cost) and the hands-on time required to get the job done right.

This is where it's critical to consider your research needs, as well as the labor required to prepare samples for sequencing and to analyze your results.

What reagents will you need on an ongoing basis? How long will it take to interpret the data? How and where will you store the data? These are some of the most important issues to consider. Examine your budget per sample and make sure everything is included in that cost, including DNA extraction and informatics.

Operational Expenses: What's the Cost Per Sample?

Consumables are one of the biggest drivers of operational expenses, as they're required for both library preparation and sequencing.

Library prep costs will vary depending on the study size, application, and vendor. Many third-party vendors sell library prep kits, but not all kits will work on every machine. That's why it's smart to select a sequencing technology for which various kits are sold. This ensures a larger range of options and, oftentimes, more competitive prices.

When comparing operational costs, look at the cost per base or cost per sample. This figure is dictated by the amount of DNA that can be sequenced per run. For higher-output applications, economies of scale can help reduce cost.

Tip: Sample indexing allows you to pack more samples into each sequencing run, lowering the cost per base. For labs running high sample numbers, the higher expense of a high-throughput sequencer is offset over time by the lower operating costs associated with each sample.

Hands-On Labor: Time Is Money

Another factor to consider with operational costs is the hand-on time required by lab technicians. The more time a lab tech must spend to carry out a given sequencing experiment, the less time that person has available for other important projects. It becomes an issue of both money and time. Consider the element of efficiency as you compare platforms.

Tip: In addition to hands-on time, pay attention to the amount of user intervention needed per experiment. Even if hands-on time is minimal, you may have to return to the equipment often to intervene. This cuts into a user's ability to do other things in the lab, and as a result increases the operational cost. Frequent interventions also increase the likelihood of human error, which can come at very high price.



Ancillary Equipment: What Add-Ons Are Required?

We've learned that the list price, or initial capital outlay, only tells one piece of the story. After factoring in the cost of consumables and hands-on labor, the next element to consider is ancillary equipment required to run your applications. Ancillary equipment can add significant unforeseen expense, so it's important to address this topic head on.

Tip: When speaking with sales representatives, ask questions that are highly specific to your application to determine whether you need additional equipment to get the job done. For example, if you often run enrichment experiments, you should find out whether you can use a single piece of instrumentation or will be required to purchase add-ons.

Data Storage: On-Site or in the Cloud?

Since next-generation sequencing has emerged, most labs have traditionally stored their large quantities of data on site, which necessitates the purchase of servers for their facility.

However, managing data on-site comes with added administrative burdens, such as maintaining an infrastructure to back up data. Also, in these instances, users may encounter challenges in sharing data with others outside of their network. The storage capacity must be continually scaled as larger volumes of sequencing data are generated.

A more recent option to consider is cloud-based storage platforms, which do not require on-site computer and storage infrastructure. Sequencing data are loaded from the instrument directly into the "cloud." Users can analyze data in the cloud and share data around the globe. Cloud-based systems easily scale to meet growing sequencing needs.

In addition to comparing the costs of your various data storage options, you may want to explore a cloud-based storage solution to determine if this would work for your lab.

Tip: For security reasons, some institutions are not allowed to store data in the cloud. If you're interested in the integrated computing and storage of a cloud platform, ask the provider if an on-site version is available.

Step 4: Envision Your New Workflow and Informatics.

You want your NGS system to foster the easiest workflow possible. Considering the scope of many DNA experiments, you should seek out any and all opportunities to save time and ensure accuracy.

The quantity of data NGS produces can seem overwhelming at first. Next-generation sequencing data output has increased at a rate that surpasses Moore's law, more than doubling each year since it was invented. In 2007, a single sequencing run could produce a maximum of around one gigabase of data. Today, that rate exceeds a terabase of data in a single sequencing run — more than a 1,000x increase.

What does this mean for you? For starters, it means that you'll have to find a way to manage a very large amount of data. So as you size up your NGS options, think about what each system will mean for your workflow, from sample preparation to informatics. How exactly will you turn that raw data into actionable information? Make sure you're comfortable with the workflow, not intimidated by it.

Sample and Library Preparation

Consider the types of experiments you do and how many samples they usually involve. When evaluating your NGS options, look specifically at how many days it will take to create the libraries. How much of that time is "hands-on" time? Does the equipment vendor offer a library solution?

In general, sample preparation protocols for NGS are more rapid and straightforward than those for Sanger sequencing. With NGS, you can start directly from a gDNA or cDNA library. The DNA fragments are then ligated to platform-specific oligonucleotide adapters to perform the sequencing biochemistry, requiring as little as 90 minutes to complete.

When considering sample preparation, ask vendors about the versatility of their products: Do the sample prep solutions support a broad range of applications? Is it easy to purchase all of your needed solutions in one place and access support when questions arise?

Data Analysis

Along with sample preparation, data analysis is an area that simply cannot be overlooked when comparing NGS equipment.

A wide range of data analysis algorithms are available that perform specific tasks related to a given application. Some applications require specialized assembly of sequencing reads. Others require quantification of read counts to provide information about gene expression levels.

While some of these data analysis applications are commercially available from software vendors, many are freely available open-source algorithms from academic institutions.

Explore what types of analysis tools and protocols are available to users of a given platform, and ask whether these tools are specialized in working with data from the equipment you're looking to purchase. Ease of workflow and data integrity are at stake. It's even better if some tools are open source because users and developers continually make improvements.

Step 5: Address the Big Issue of Quality.

If you're trying to put together a jigsaw puzzle and you can't make out the design on some of the pieces, you're going to have a difficult time achieving your goal. The same is true of DNA-driven research.

Data: Q-Scores Are Important

Exceptional data quality is at the core of why NGS has become such a huge driver of discovery in recent years. However, not all systems can offer the same reassurance that bases will be called correctly. The consequences of poor quality are immense: Not only can errors give you the wrong answer to your research questions, but they can also cause you to miss the right answer. False positives and false negatives can lead to higher costs and clinical scares or oversights.

That's why, when comparing systems, you should pay close attention to the quality score, also known as a Q-score.

Quality scores measure the probability that a base is called incorrectly. Thus, a higher quality score indicates a smaller probability of error. The below chart will help you to understand how Q-scores are calculated, and what it means for your research.

The Relationship Between Quality and Base Call Accuracy

Quality Score	Probability of Incorrect Base Call	Inferred Base Call Accuracy
10 (Q10)	1 in 10	90%
20 (Q20)	1 in 100	99%
30 (Q30)	1 in 1000	99.9%

Systematic Errors: Address the Homopolymer Issue

While the Q-Score is a standard measurement of data quality, other factors relate to the overall quality of the instrument design and performance.

For example, learn how the instrument you're considering performs with homopolymers, or regions of identical bases. Some short-read NGS technologies have encountered problems in correctly counting the number of bases in homopolymer runs. Look for user groups and see what experience people report having, and how that compares with competing instruments.

Workflow Design: Can You Plug and Play?

Integration is the buzz word here. As you move from raw data to meaningful results, you want the process to be as simple and coordinated as possible. This includes sample preparation, sequencing, and data analysis.

Consider how each supplier handles integration and find out whether it offers complete workflow solutions for the applications that matter most to you. When gauging quality in this area, look for customer reviews and community forums. Ask colleagues for advice and speak in depth with sales representatives about how easy it will be for you to extract answers to your research questions in an integrated, meaningful way.

Underlying Technology: Gauge Your Comfort

Each of the major NGS suppliers relies on a different underlying sequencing platform. Sequencing by synthesis, semiconductor technology, and nanopore technology are a few of the terms you'll come across when researching the different platforms.

As you explore your options, learn about how each system's technology works. Is it proven? Is it scalable to your future needs? As with any technology, earlier-stage platforms come with the risk of the unknown, while more mature platforms offer larger user communities and more reliable performance.

Whether you're considering a new technology or an established one, be sure that all core accuracy issues have been resolved and that your machine will be able to produce enough sequencing volume per run and data quality to meet your application needs.



Step 6: Know the Terms.

If you've never purchased a next-generation sequencing system before, it helps to first gain familiarity with the terms. This will help you better understand comparisons from system to system, and it will expand your vocabulary so you can ask the right questions of your field application specialist.

Glossary of NGS Terms

Coverage. The average number of sequencing reads that align to each base within the sample DNA. For example, a whole genome sequenced at 30x coverage means that, on average, each base in the genome was covered by 30 sequencing reads.

Coverage uniformity. A measurement of the evenness of the coverage depth of target.

False negative. Variant base that is called as non-variant in a sequencing experiment.

False positive. Non-variant base that is called as variant in a sequencing experiment.

FASTA/FASTQ. FASTA format is used to store nucleic acid sequences or protein sequences; FASTQ is used for storing both a nucleotide sequence and its corresponding quality scores.

Gb. Gigabase, or one billion nucleotides. The Gb required for a given application can vary by the size of the genome (for example, human genome versus a microbe).

Library. A collection of DNA fragments with adapters ligated to each end.

Long reads. Sequence reads that generally are at least 400 base pairs long in a single direction. Long reads make nucleic acid scaffold construction from subsequences an easier bioinformatic task.

Log scale. A logarithmic scale is a scale of measurement that displays the value of a physical quantity using intervals corresponding to orders of magnitude, rather than a standard linear scale.

Mapped read depth. The total number of bases sequenced and aligned at a given reference base position.

Mean mapped read depth. The sum of the (mapped) read depths at each reference base position, divided by the number of known bases in the reference genome.

Percentage of genome mapped. Percentage of the reference genome that is successfully mapped after alignment.

Quality score (Q-score). Quality scores measure the probability that a base is called incorrectly. Thus, a higher quality score indicates a smaller probability of error.

Raw read depth. Total amount of sequence data output by the instrument (pre-alignment), divided by the reference genome size.

Run speed. The speed at which an NGS instrument can produce data; must be weighed against data quality and ease of workflow for true indication of overall time of experiment.

SAM. Sequence Alignment/Map format is a generic alignment format for storing read alignments against reference sequences, supporting short and long reads (up to 128 Mb) produced by different sequencing platforms.

Short reads. Sequence reads that are normally 150 base pairs long or shorter in one direction. They normally have higher accuracy than long reads.

SNP detection. The ability to detect genetic variations of single nucleotide polymorphisms (SNPs) between members of a species.

Targeted sequencing. With targeted sequencing, only a subset of genes or defined regions in a genome are sequenced, allowing researchers to focus time, expenses, and data storage on the regions of the genome in which they are most interested.

Variant calling. Variant detection remains a major research focus of sequencing experiments. Errors in base calling may lead to the identification of non-existent variants, or to missing true variants in the sample. Such erroneous calls are termed false positive and false negative identifications, respectively.

Step 7: Consider Reputation, Community, and Support.

When making any new purchase, there's often a fear of the unknown. It can be daunting making a switch to a new technology. But it doesn't have to be. This is where a company's strong reputation, established community of users, and responsive service support will help put you at ease.

Reputation. We'll say this again: data quality is important. Investigate published research that relied on the equipment to generate data. Seek out other users and ask if the workflow is easy. The more you can learn, the more comfortable you'll feel with your purchasing decision.

Community. Look for an established community of users who specialize in your key applications. Engage with this community to learn about others' experiences using the systems you're considering. Scientists are always inventing new ways of doing things; as such, an active community offers the benefit of deep knowledge and innovative ideas.

Support. Find out what happens after the purchase. You want to do business with a company that will help you get up and running, not just drop off an instrument at your door. Look for a well-defined and personalized onboarding process — a company that will help you set up your machine, train your lab workers, and follow up to make sure that everything is working as planned. Ask the sales representative whether the company offers any extra training or webinars that will help you keep your knowledge fresh.



NGS Buyer's Checklist: Covering All the Bases.

Thanks to the pioneering world of next-generation sequencing (NGS), meaningful discovery is churning at a pace unmatched throughout history. For this reason, labs that don't already own an NGS system are making a push to acquire equipment of their own.

If you're a principal investigator or manager for one of these labs, you have an important purchasing decision to make. The good news is that the market today offers plenty of choice. Use this checklist to guide your buying process.

Define Your Applications.

What types of NGS experiments do you plan to perform and which applications will you use most? Consider your application's needs for these three areas:

- Throughput per run.
- Read length.
- Paired-end sequencing.

Understand the Equipment Options.

Buying equipment is a commitment; consider your needs now and for the future.

- Desktop Low-Throughput Sequencer.
- Desktop High-Throughput Sequencer.
- High-Throughput / High-Volume Sequencer.

Think About Budget Holistically.

There are many factors to consider beyond the initial capital expenditure.

- Operational Expenses: What's the cost per sample? Factor in consumables and library prep kits.
- Hands-On Labor: Consider the element of efficiency. The more time a lab tech must spend to carry out a given sequencing experiment, the less time that person has available for other important projects.
- Ancillary Equipment: What add-ons are required? Determine whether additional equipment will be needed to get the job done.
- Data Storage: On-site or in the cloud? Inquire about the costs and learn about the benefits and drawbacks of each solution.

Envision Your New Workflow and Informatics.

- Seek out any and all opportunities to save time and ensure accuracy.
- Sample and Library Preparation: How many days will it take to create the libraries? How much of that time is "hands-on"? Does the equipment vendor offer a library solution? Do the sample prep solutions support a broad range of applications?
- Data Analysis: Explore what types of analysis tools and protocols are available to users of a given platform.

Address the Big Issue of Quality.

If you're trying to put together a jigsaw puzzle and you can't make out the design on some of the pieces, you're going to have a difficult time achieving your goal. The same is true of DNA-driven research.

- Q-Score: Quality scores measure the probability that a base is called incorrectly. Thus, a higher quality score indicates a smaller probability of error.
- Instrument Operation: Address the homopolymer issue.
- Workflow Design: Can you plug and play?
- Underlying Technology: Different instruments use different underlying sequencing platforms. Make sure the technology is well proven and scalable to your future needs.

Know the Terms.

Download the full Buyer's Guide for a glossary of the most frequently used terms. If, during the buying process, you run into a word or phrase you don't know, ask a field application specialist or company representative for clarification.

Consider Reputation, Community and Support.

Ensure there's an established base of users and responsive service support. Consider the following items.

- Reputation: Investigate published research and ask existing users about ease of workflow.
- Community: Look for an established group of users who specialize in your key applications.
- Support: Seek out a well-defined and personalized onboarding process.

Notes

