An Introduction to Next-Generation Sequencing for Pathologists
# Table of Contents

I. Genomics and Molecular Pathology 3

II. Why Next-Generation Sequencing? 3

III. What is Next-Generation Sequencing? 4
   - The Basic NGS Workflow 4
   - Multiplexing 4

IV. NGS Methods 5
   - Targeted Sequencing 5

V. NGS Applications in Molecular Pathology 5
   - Tumor Profiling 5
   - Liquid Biopsy Analysis 6
   - Inherited Condition Screening 6

VI. Summary 7

VII. Glossary 8

VIII. References 8
I. Genomics and Molecular Pathology

Scientific research has shown that genetic factors are implicated in many diseases. Multiple mutations have been associated with disease risk, hereditary disorders, and cancer onset and progression. Several well-characterized genes, such as \textit{HER2}, \textit{ALK}, \textit{EGFR}, \textit{BRAF}, and \textit{KRAS}, are used in molecular testing procedures today, while others are constantly being discovered and investigated. As research continues to uncover genetic variants associated with clinical outcomes, this information can lead to earlier diagnosis and more targeted therapies.

Next-generation sequencing (NGS) is the latest evolution in genomic technologies, enabling investigation of causative mutations and evaluation of risk level and prognosis.\textsuperscript{1} In fact, several of the leading hospital and treatment centers, such as the Dana Farber Cancer Institute and the Mayo Clinic, have announced that they are prepared to sequence “all patients for millions of tumor mutations.”\textsuperscript{2,3} Several organizations, including the Association for Molecular Pathology (AMP),\textsuperscript{4} the College of American Pathologists (CAP),\textsuperscript{4,5} the US Food and Drug Administration (FDA),\textsuperscript{6,7} and the American College of Medical Genetics (ACMG),\textsuperscript{8} recently published guidelines and recommendations for clinical sequencing. The advent of NGS is changing the molecular pathology paradigm, promising to transform diagnosis, treatment, and drug development in the future.

“With a panel of genes or whole-genome sequencing, we are able to look at more alterations that might be driving the patient’s cancer than with non-NGS methods, so the inquiry is significantly broader for a lower price and faster turnaround time.”

—Elaine Mardis, PhD, codirector of The McDonnell Genome Institute, Washington University School of Medicine\textsuperscript{9}

The growing adoption of NGS technology and the integration of sequencing standards into lab practices reflect the pathology community's commitment to making molecular testing an integral part of routine clinical practice. Technological advances have caused sequencing to become more accessible to even small academic and reference labs, with user-friendly desktop instrumentation that circumvents the need for highly trained sequencing experts. With recognition from leading researchers, medical professionals, and regulatory organizations, NGS is becoming a widely heralded technology for the future of human health.

“As [massively parallel sequencing] instruments continue to drive advances in translational research, it is inevitable that more laboratories will adopt the technology, and we will no doubt reach the day when next-generation sequencing becomes standard medical practice.”

—John ten Bosch, PhD and Wayne Grody, MD, PhD, FCAP, FACMG\textsuperscript{10}

II. Why Next-Generation Sequencing?

NGS offers several advantages over traditional approaches such as Sanger sequencing, PCR amplicon testing, and single-gene assays. Iterative single-gene tests can be labor-intensive, costly, and time-consuming processes, and sometimes do not produce an answer, requiring follow-on tests. Such sequential testing might be limited by tissue availability and might require additional biopsies. In contrast, by analyzing multiple genes and multiple samples in a single experiment, NGS drastically reduces the time to results. With a gene panel, the most significant genes are evaluated at the same time, largely eliminating the need for follow-on tests. In addition, NGS provides high sensitivity, enabling the detection of mutations present at as little as 5% of the DNA isolated from a tumor sample. Even degraded samples, such as formalin-fixed, paraffin-embedded (FFPE) tumor tissue, can be sequenced, starting from a small amount of DNA.
III. What is Next-Generation Sequencing?

In principle, the concept behind NGS technology is similar to Sanger sequencing. DNA polymerase catalyzes the incorporation of fluorescently labeled nucleotides into a DNA template strand during sequential cycles of DNA synthesis. During each cycle, the nucleotides are identified by fluorophore labels. The critical difference is that instead of sequencing a single DNA fragment, NGS extends this process across millions of fragments in a massively parallel fashion. This method is highly scalable—it can be applied to a subset of key genes or the entire genetic code.

The Basic NGS Workflow

The Illumina NGS workflow involves 3 general steps (Figure 1):

1. Sequencing library preparation begins by creating short DNA or cDNA fragments with 5’ and 3’ adapters ligated. For “cluster generation,” the library is attached to an oligonucleotide lawn on the surface of a flow cell. Through bridge amplification, each library fragment acts as a seed to generate a clonal cluster containing thousands of identical fragments. Across the entire flow cell, millions to billions of clusters are formed.

2. Next, the templates are ready for sequencing by synthesis (SBS). SBS technology utilizes a proprietary reversible terminator–based method that detects single bases as they are incorporated into DNA template strands. Because all 4 reversible, terminator-bound dNTPs are present during each sequencing cycle, natural competition minimizes incorporation bias and greatly reduces raw error rates compared to other technologies. The result is highly accurate base-by-base sequencing that virtually eliminates sequence-context-specific errors, even within repetitive regions and homopolymers.

3. The newly identified sequence reads are then exported to an output file and aligned to a reference genome by sequencing alignment software. Analysis and reporting software denotes the presence or absence of variants.

A detailed animation of SBS sequencing is available at www.illumina.com/SBSvideo.

![NGS Workflow](image)

**Figure 1: NGS Workflow**—The Illumina NGS workflow includes 3 general steps: library preparation, sequencing, and data analysis.

*Run in research mode.

Multiplexing

Multiplexing allows large numbers or batches of samples to be pooled and sequenced simultaneously during a single sequencing run. With multiplexed samples, unique index sequences are added to each DNA sample during library preparation so that each read can be identified and sorted before final data analysis. This dramatically reduces the time to data for multisample studies and enables researchers to go from experiment to answer faster and easier than ever before.
IV. NGS Methods

Compared to traditional methods, NGS offers advantages in accuracy, sensitivity, and speed that can make a significant impact on the field of molecular pathology in the future. Depending on the application, various NGS methods are available to suit diverse study designs and objectives, including whole-genome sequencing, exome sequencing, and targeted sequencing (Figure 2). Whole-genome sequencing determines the entire DNA sequence, while exome sequencing analyzes only on the coding portion of the genome. Targeted sequencing focuses on specific genes of interest and is the most commonly used NGS method in molecular pathology.

Targeted Sequencing

Targeted sequencing using NGS focuses on a preselected subset of genes with known involvement in the disease under study, enabling assessment of all potentially causative genes at the same time (Figure 3). Targeted gene sequencing offers several advantages:

- Analyzes multiple genes in a single assay
- Optimizes use of limited tissue samples by reducing need for sequential testing
- Enables the accurate identification of rare variants in heterogeneous tumor samples

Figure 2: NGS Methods—NGS methods include whole-genome sequencing, whole-exome sequencing, and targeted sequencing.

V. NGS Applications in Molecular Pathology

The most common application of targeted sequencing in molecular pathology is tumor profiling. NGS approaches are also becoming more common in analyzing circulating tumor DNA from “liquid biopsies” and screening for genetic disorders.

Tumor Profiling

Today, molecular profiling is a standard technique for classifying tumors, with established guidelines from the College of American Pathologists and the National Comprehensive Cancer Network. Tumor profiling using NGS focuses on a preselected subset of genes that have known involvement in cancer. This approach can be applied to both solid tumors and hematological malignancies, as well as germline risk assessment for determining cancer predisposition. Because the method focuses on a limited number of genes, it enables deep sequencing, resulting in high sensitivity and the ability to analyze rare variants and clonal tumors. Targeted sequencing using NGS follows a simple workflow that can be scaled easily, enabling labs to process hundreds of samples at a time and deliver answers sooner.
**Liquid Biopsy Analysis**

Cell-free, circulating tumor DNA (ctDNA) can act as a noninvasive cancer biomarker, offering a potential alternative to invasive tissue biopsies. Today, researchers are investigating the use of ctDNA as a biomarker for detecting the presence of tumors in liquid biopsies obtained through a simple blood draw. In the future, ctDNA could potentially serve as a noninvasive approach for real-time cancer detection, monitoring of therapeutic response, assessing remission or progression, and screening for disease. NGS offers the ability to reach high sensitivity and specificity needed to detect low levels of ctDNA in the bloodstream. Further refinement of this technology and the development of assays for ctDNA detection hold considerable potential to revolutionize the way cancer is identified and treated, leading to earlier diagnosis, improved survival rates, and better quality of life for cancer patients.

**Inherited Condition Screening**

Targeted sequencing of genes that contribute to inherited diseases, such as cystic fibrosis or cardiac conditions, provides the opportunity to analyze multiple potentially causative genes at the same time. Researchers can analyze not only known causal genes, but also expand analysis to include emerging genes cited in literature or risk-associated genes that contribute to disease onset. This approach can shorten the time to answer and uncover the causative variant or variants for a particular disorder.

![Figure 3: Targeted Sequencing](image)

*Figure 3: Targeted Sequencing* — Targeted sequencing focuses on a subset of genes related to the disease or phenotype under study. This approach results in deeper sequencing for higher sensitivity of detection, enabling accurate investigation of potentially causative mutations.

**VI. Summary**

Over the last decade, advances in genomics have led to an improved understanding of disease biology, which, in turn, has led to new approaches to diagnosing and treating disorders. The adoption of sequencing-based methods continues to grow, reducing time and costs for molecular testing procedures and offering the potential for more specific, individualized patient assessment. This revolution in health care will lead to a shorter time to diagnosis and more effective therapies, ultimately saving lives.

Illumina is committed to advancing molecular analysis tools and collaborating with industry leaders to transform health care. Together, we will bring the promise of NGS toward widespread clinical adoption and improvements in patient diagnosis, treatment, and outcomes.
VII. Glossary

Adapters: Oligos bound to the 5’ and 3’ ends of each DNA fragment in a sequencing library. The adapters are complementary to the lawn of oligos present on the surface of Illumina sequencing flow cells.

Bridge amplification: An amplification reaction that occurs on the surface of an Illumina flow cell. During flow cell manufacturing, the surface is coated with a lawn of 2 distinct oligonucleotides referred to as “p5” and “p7.” In the first step of bridge amplification, a single-stranded sequencing library is injected into the flow cell. Individual molecules in the library bind to complementary oligos as they “flow” across the oligo lawn. Priming occurs as the free end of a ligated fragment bends over and “bridges” to another complementary oligo on the surface. Repeated denaturation and extension cycles (similar to PCR) result in localized amplification of single molecules into millions of unique, clonal clusters across the flow cell.

Clusters: A clonal grouping of template DNA bound to the surface of a flow cell. Each DNA template strand that binds to the flow cell acts as a seed and is clonally amplified through bridge amplification until the cluster has roughly 1000 copies. Each cluster on the flow cell produces a single sequencing read. For example, 10,000 clusters on the flow cell would produce 10,000 single reads.

Flow cell: A glass slide coated with a lawn of surface-bound, adapter-complimentary oligos. A pool of 8–384 multiplexed libraries can be sequenced simultaneously, depending on application parameters.

Indexes/Barcodes/Tags: A unique DNA sequence ligated to fragments within a sequencing library for downstream in silico sorting and identification. Libraries with unique indexes can be pooled together, loaded into a lane of a sequencing flow cell, and sequenced in the same run. Reads are later identified and sorted via software.

Multiplexing: Multiple samples, each with a unique index, can be pooled together, loaded into the same flow cell, and sequenced simultaneously during a single sequencing run. Depending on the application and the sequencing instrument used, 10–384 samples can be pooled together.

Read: In general terms, a sequence “read” refers to the data string of “A, T, C, and G” bases corresponding to the sample DNA. With Illumina technology, millions of reads are generated in a single sequencing run. In specific terms, each cluster on the flow cell produces a single sequencing read.

Sequencing by synthesis (SBS): SBS chemistry uses 4 fluorescently labeled nucleotides to sequence the millions to billions of clusters on a flow cell surface in parallel. During each sequencing cycle, a single labeled dNTP is added to the nucleic acid chain. The nucleotide label serves as a reversible terminator for polymerization. After dNTP incorporation, the fluorescent dye is identified through laser excitation and imaging, then enzymatically cleaved to allow the next round of incorporation. Base calls are made directly from signal intensity measurements during each cycle.\(^1\)

VIII. References