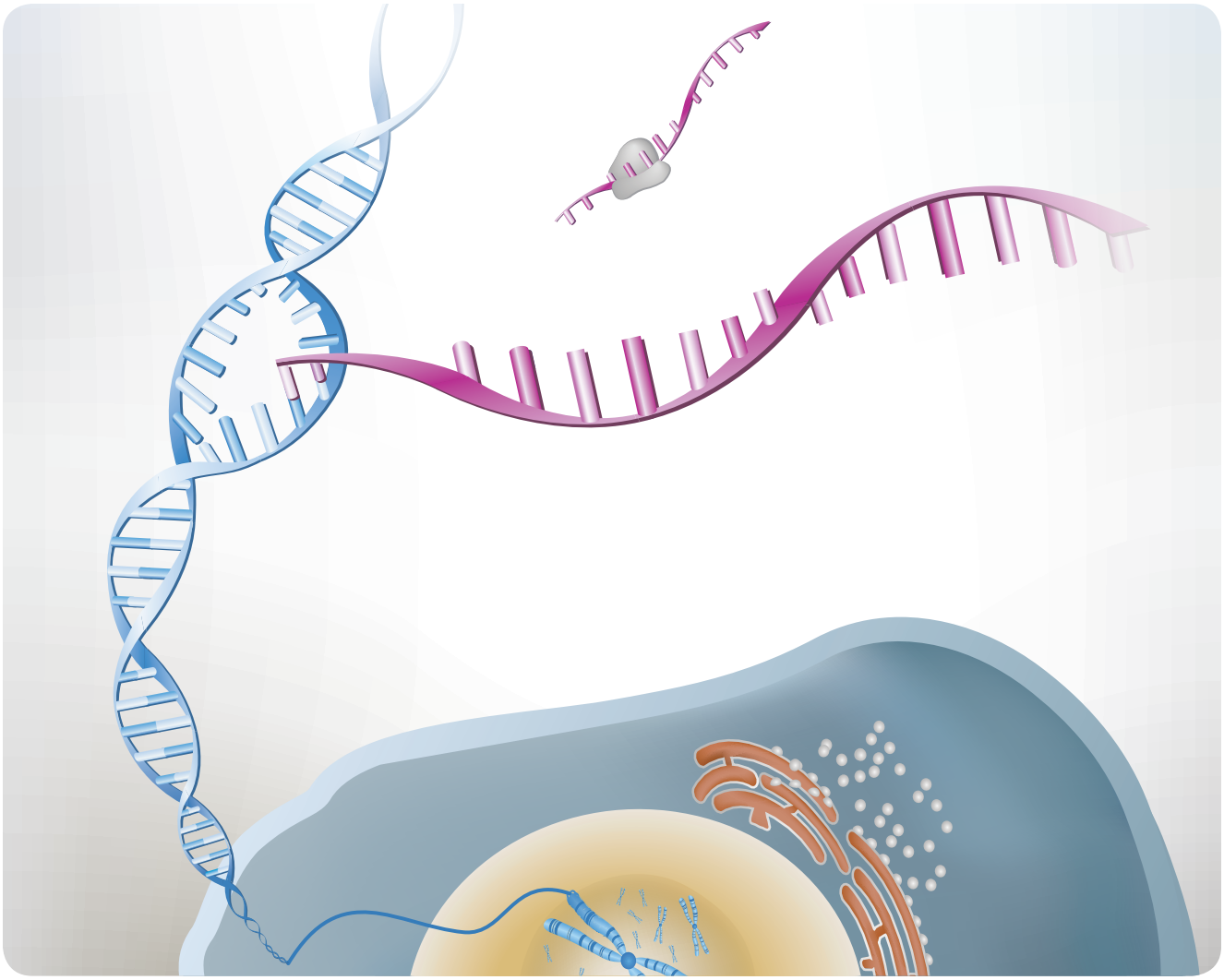

An Introduction to Next-Generation Sequencing for Cell Biologists

www.illumina.com/NGStech

illumina®

Table of Contents

Welcome to Next-Generation Sequencing

In 1665, Robert Hooke placed a thin slice of cork under his microscope and saw that it divided into “a great many little Boxes,” which he referred to as “cells.” His findings and the invention of the microscope led to the cell theory, the foundation of biology. As the basic unit of life, cells play a role in all aspects of human physiology, including development and disease. At the time, the microscope revolutionized how scientists understand cellular function, and it has become an invaluable tool for studying the appearance and behavior of cells. However, new discoveries have revealed that cell activity involves more than physical characteristics alone.

Today, biologists studying disease etiology have many methods at their disposal to analyze cells—microscopy, flow cytometry, and staining comprise only a few. Many conventional approaches for studying cell biology rely on antibody staining or other signal-based methods to visualize protein interactions. Using staining alone to study protein function can be challenging, because antibodies are not available for all proteins and only a limited number of targets can be investigated simultaneously. These methods reveal physical attributes of cells and protein function, but do not capture an important component of biology—namely, the genetic code that dictates protein behavior. Increasingly, scientists are finding strong links between genetics and disease phenotypes, indicating that a true understanding of proteins and their behavior requires knowing the underlying genetics and regulation.

Understanding genetic factors is essential for virtually all branches of biological research, and the introduction of next-generation sequencing (NGS) technology has transformed the study of biology.[1] Recent collaborations such as the Encyclopedia of DNA Elements Consortium (ENCODE)[2] used NGS to provide functional information about human health at the molecular level by exploring transcription, regulation, DNA–protein binding, and epigenetics. By digitally counting sequence reads rather than measuring continuous signal intensities, as many current methods do, NGS provides quantitative information about protein translation, gene expression, and regulation at higher resolution than traditional microarray—or antibody-based—methods. Knowing how these processes are altered in disease states can help researchers predict changes to cell behavior and design additional studies to assess the proteins in specific functional pathways. NGS opens new avenues to exploring and understanding the cellular activity of disease, so scientists can ultimately develop therapeutic approaches to target the biological pathways contributing to disorders.

Why Next-Generation Sequencing?

As an unbiased technology, NGS can be used to investigate gene expression and protein translation in disease biology and generate hypotheses for further functional studies. It offers a high-throughput alternative to current single-analyte assays, enabling screening of more samples in less time. Using NGS, biologists can analyze cell states and fates, study signal transduction pathways, determine causal variants, and examine tissue-specific gene expression.

High-Throughput Science

Often, biological research involves single-gene knockout models that assess how the absence or modification of a certain protein affects a specific pathway. This approach is time-consuming and does not always result in conclusive findings, but often leads to additional experiments designed to test alternate hypotheses. Today, many scientists are using other approaches to find targets of interest. NGS enables researchers to survey the entire genome, transcriptome, or epigenome in a single sequencing run and observe multiple changes at one time. Sequencing results can inform experimental design by identifying proteins of interest for subsequent studies, so that researchers can investigate targets efficiently, save time, and publish sooner.

AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACG
TCAACGTACCGTAACGAACGTATCATTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTA
CGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACC
AACGTACCATTAAGAGCTACCGTGCAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAA
AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACG
GATTACTTGATCCACTGATTCAACGTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAGCAACGACG
CGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGAAAAGAATGATAACAGTA

Unlimited Dynamic Range

Dynamic range is the difference between the highest and lowest signals that can be detected. The digital nature of NGS supports an unlimited dynamic range, providing high sensitivity for quantification-based applications, such as gene expression analysis. With NGS, researchers can quantify RNA activity at much higher resolution than traditional microarray-based methods, which is important for capturing subtle gene expression changes associated with biological processes. While microarrays measure continuous signal intensities, with a detection range limited by noise at the low end and signal saturation at the high end, NGS quantifies discrete, digital sequencing read counts. By increasing or decreasing the number of sequencing reads, researchers can tune the sensitivity of an experiment to accommodate different study objectives.

Sequencing Applications for Cell Biology

NGS platforms enable a wide variety of applications, allowing researchers to investigate any cell type or organism. In disease cases, various factors can influence the affected phenotype, including gene mutations, changes to gene expression at different times and under varying environmental conditions, and differences in epigenetic regulation. Cell biologists must consider these potential alterations when looking for the molecular pathways that contribute to complex traits or diseases. Researchers can leverage the flexibility of NGS to quantify gene expression in a signaling pathway using a single technology (Figure 1).

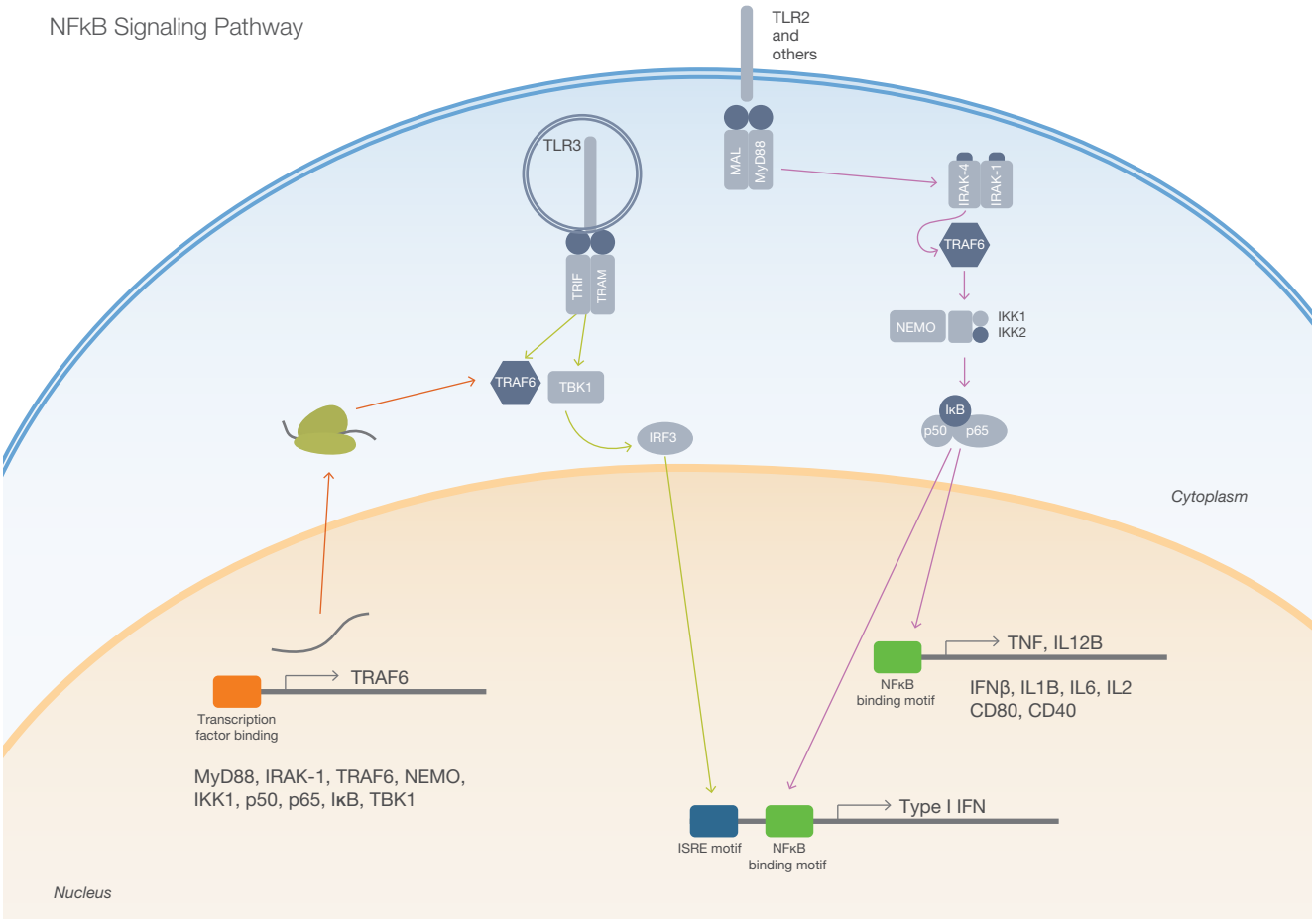

Figure 1: Molecular Analysis with NGS. NGS enables precise and sensitive detection of signaling pathway dynamics. Targeted RNA sequencing provides dynamic expression data of key signaling molecules and transcription targets. Chromatin immunoprecipitation sequencing allows researchers to identify binding locations across the genome, at a precise moment, or in response to a cellular stimulus, while ribosome profiling enables quantification of actively translated proteins.

Gene Expression Profiling

Traditional approaches for evaluating gene expression include microarrays, which rely on continuous signal intensity and require prior knowledge of probe sequences. With arrays, noise and signal saturation limit detection, so low and high signal levels are often overlooked. RNA sequencing (RNA-Seq) is a more sensitive approach; by providing information about the RNA template, it can help add significant details to studies of protein structure and function. It quantifies RNA transcripts by measuring discrete, digital sequencing reads, capturing both low and high levels of expression. RNA-Seq does not require prior knowledge about the target sequences, enabling discovery of novel transcripts and splice variants. In contrast to single-analyte methods, such as quantitative real-time polymerase chain reaction (qPCR), RNA-Seq enables researchers to profile and quantify all RNA transcripts in a cell at one time, saving time and resources. Targeted RNA-Seq offers the ability to multiplex thousands of targets and investigate all genes of interest in a pathway, processing more RNA transcripts in less time compared to qPCR. RNA information can then be used to predict changes to protein structure and function caused by missense, nonsense, frameshift, and other mutations. By characterizing transcriptional changes, researchers can better understand protein activity in disease, leading to a better understanding of signaling pathways and how signaling disruption affects downstream cellular response.

Without requiring probe or primer design, mRNA sequencing offers an unbiased method for investigating the transcriptome and gaining a deeper understanding of gene regulation and transcriptional states at the individual cell level. Sequencing mRNA using NGS can identify strand information, indicating from which of the two complementary DNA strands a particular mRNA transcript was derived. Strand orientation allows identification of antisense expression, an important mediator of gene regulation.[3] The ability to capture the relative abundance of sense and antisense expression provides visibility into regulatory interactions that might otherwise be missed. Initial RNA-Seq experiments can identify several targets of interest at one time, generating hypotheses about their functional effects. These hypotheses can be explored through focused studies using protein-based approaches in combination with targeted RNA-Seq and ribosome profiling to investigate the role of gene expression and protein translation in cell function.

Real-Time Analysis of Protein Translation

The speed at which a cell can activate or silence gene expression is relatively rapid, suggesting high turnover rates at both the RNA and protein level. Current methods used to measure RNA or protein levels tend to overlook the regulatory potential of translation as a means to control protein abundance. Ribosome profiling is a novel technique for investigating the RNA directly associated with the translation machinery. This method relies on deep RNA sequencing of ribosome-protected mRNA fragments to provide a complete view of the actively translating RNA in a cell. This information can then be used to determine which proteins are actively translated in a cell at a specific point in time, so that biologists can better understand time-dependent changes to cellular signaling pathways.

Ribosome profiling offers a bridge to link gene expression data with protein abundance measurements as compared to single-analyte, antibody-based methods. These approaches require antibodies with specificity for particular antigens and can only detect one protein per assay. In contrast, ribosome profiling surveys all proteins in a cell at one time rather than relying on antibody availability and tedious sequential analyses. Because ribosome profiling uses RNA sequencing as a means by which to estimate protein levels, it has a much higher dynamic range of detection as compared to other methods designed for direct protein detection. Ribosome profiling using NGS does not require prior sequence knowledge and captures all actively translated mRNA transcripts, avoiding the potential for missed signals. Ribosome profiling provides a more accurate representation of protein translation and generates less signal noise compared to other high-throughput methods, such as mass spectrometry. With ribosome profiling, researchers can measure gene expression, identify translation start sites, predict protein abundance, and investigate translational and co-translational processes *in vivo*.

Characterizing DNA Methylation

Recent studies have shown that lifestyle and environmental factors can lead to epigenetic changes to DNA, which can cause or exacerbate disease. Aberrant DNA methylation and its impact on gene expression have been implicated in many disease processes, including Alzheimer's disease.[4] Methylation of DNA at cytosine nucleotides impacts various cellular activities involving gene expression, RNA processing, and protein function. These processes can influence cell biology by activating or suppressing certain genes at specific times, with implications in development, disease, aging, and immune defense. Sequencing-based methylation analysis applies the coverage density of NGS to provide a detailed view of methylation in individual cell types or states. It relies on bisulfite sequencing, which uses a sodium bisulfite conversion reaction to distinguish between methylated and non-methylated cytosine bases. More stable than short-lived RNA species, methylation profiles provide an additional route to understanding gene expression patterns over time. Examining methylation and other epigenetic signatures allows researchers to characterize the causative DNA modifications that alter protein production.

Understanding DNA–Protein Interactions

Chromatin immunoprecipitation sequencing, or ChIP-Seq, can be used to survey interactions between proteins, DNA, and RNA. ChIP-Seq using NGS enables researchers to identify the binding sites of multiple protein targets, including transcription factors and histones, across the entire genome. Analysis of DNA–protein interactions can provide insight into the regulation events that are essential for many biological processes and disease states. Current methods for transcription factor analysis, such as arrays and qPCR, are limited in scope and provide information about only a subset of genes. NGS allows a broader scope of analysis, enabling identification and investigation of the many genes potentially activated by transcription factors in disease states. With ChIP-Seq, researchers can better understand how chromatin modifications and local structural changes impact transcription factor activity in a cell and within signaling pathways.

Single-Cell Analysis

Single-cell sequencing is an emerging method that examines the genomes or transcriptomes of individual cells, providing a high-resolution view of cell-to-cell variation in a tissue. Single-cell sequencing identifies the signals and behavior of a cell in the context of its surrounding environment. It enables researchers to assess cells individually rather than relying on the average signal from an aggregate of cells, which can result in missed signals. Strong signal identification plays a role in various disciplines; for example, it is important in neurological research for assessing connectivity in and between different regions of the brain. Single-cell sequencing is an advantageous method for biologists studying cell function and heterogeneity in time-dependent processes such as differentiation, proliferation, and tumorigenesis.

What is Next-Generation Sequencing?

NGS is an advanced technology that uses fluorescence to provide a base-by-base view of the genome, transcriptome, or epigenome. Illumina sequencing chemistry, sequencing by synthesis (SBS), is the most widely adopted and published NGS technology worldwide. SBS detects single bases as they are incorporated into growing DNA strands.

The Basic Next-Generation Sequencing Workflow

In principle, NGS is similar to Sanger (capillary electrophoresis–based) sequencing. The bases of a DNA or RNA fragment are sequentially identified from signals emitted as each fragment is resynthesized from a template strand. NGS scales up this process; millions of reactions occur in a massively parallel fashion, rather than being limited to a single or a few fragments. This advance enables rapid sequencing of large stretches of genomic information.

To illustrate how this process works, consider a single DNA or RNA sample. The nucleic acid is first fragmented into a library of smaller segments that can be sequenced efficiently. The newly identified strings of bases, called reads, are then reassembled using a known reference genome as a scaffold (resequencing). The full set of aligned reads reveals the entire sequence of the sample (Figure 2).

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

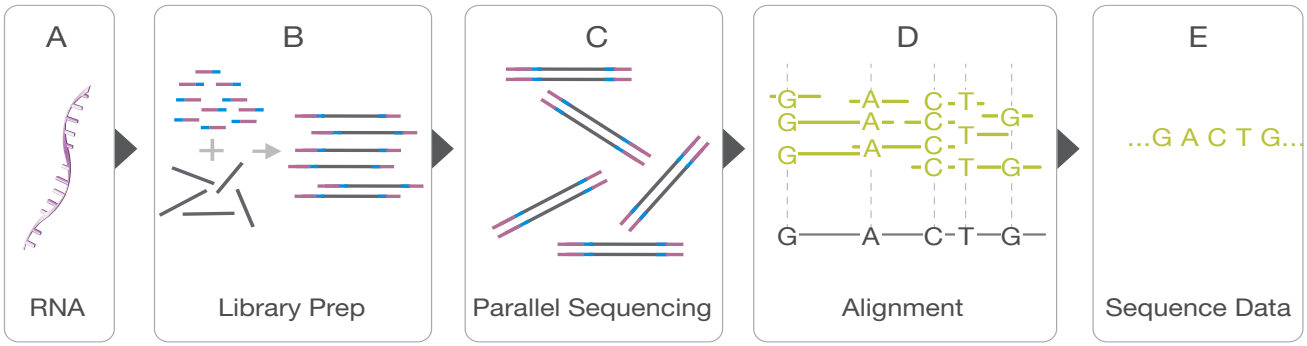

Figure 2: Concepts of NGS.
A. Nucleic acid (DNA or RNA) is extracted. If sequencing RNA, cDNA is reverse transcribed from the RNA template.
B. Library preparation fragments the DNA and adds adapters to the ends.
C. Fragments within the library are each sequenced in parallel.
D. Individual sequence reads are aligned to a reference sequence, such as a reference genome or known genes.
E. A consensus of aligned reads is generated allowing for subsequent variant calling.

Multiplexing

In addition to the increased amount of data generated on NGS instruments, the sample throughput per run has also increased over time. Multiplexing allows large numbers or batches of samples to be pooled and sequenced simultaneously during a single sequencing run (Figure 3). With multiplexed samples, unique index sequences (or "barcodes") are added to each DNA fragment during library preparation so that each read can be identified and sorted before final data analysis. Multiplexing with NGS allows investigators to process large sample numbers quickly, generating the statistical power to detect expression patterns confidently. Multiplexing dramatically reduces the time to data for multi-sample studies and enables researchers to go from experiment to answer faster and easier than ever before.

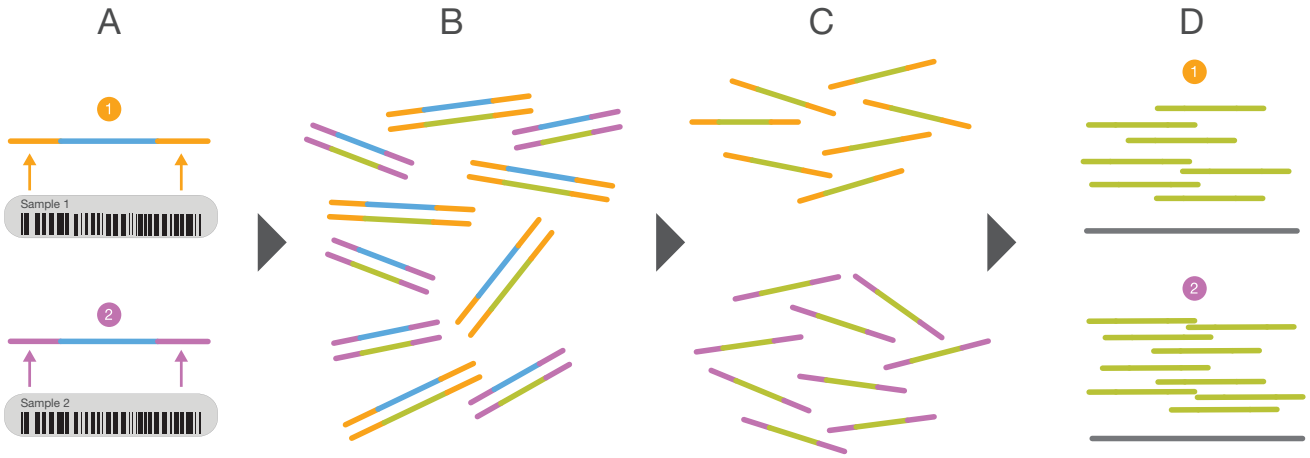

Figure 3: Sample Multiplexing Overview.
A. Two representative DNA fragments from two unique samples are each attached to a specific barcode sequence that identifies the sample from which it originated.
B. Libraries for each sample are pooled and sequenced in parallel. Each new read contains both the fragment sequence and its sample-identifying barcode.
C. Barcode sequences are used to de-multiplex, or differentiate, reads from each sample.
D. Each set of reads is aligned to the reference sequence.

Sequencing Services and Systems

Service laboratories are a cost-effective option for researchers getting started with NGS. Illumina certified service providers (CSPro®) are located worldwide and provide sequencing services from various sample types. Specialized services are available for biologists using formalin-fixed, paraffin-embedded (FFPE) tissue or limited samples. CSPro labs offer many different services to accommodate diverse study designs, including RNA-Seq, bisulfite sequencing, ChIP-Seq, small RNA analysis, and others. Each service delivers analyzed data, so that researchers spend less time analyzing data and can focus on the next discovery.

To locate a nearby service provider, visit www.illumina.com/cspro.

Because of the advantages offered by NGS, many researchers choose to acquire their own NGS systems. Available systems range from high-output instruments that can process many samples in a single run to desktop sequencers ideal for smaller-scale studies.

To find the sequencer that best fits your lab, visit www.illumina.com/sequencer.

AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACG
TCAACGTACCGTAACGAACGTATCATTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTA
CGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACC
AACGTACCATTAAGAGCTACCGTGCAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAA
AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACG
GATTACTTGATCCACTGATTCAACGTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAGCAACGACG
CGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGAAAAGAATGATAACAGTA

Summary

Over the last decade, advances in NGS technology have led to an improved understanding of genomics, which, in turn, has led to new approaches to understanding cellular function and variation. Investigating DNA and RNA sequences can provide insight into protein function and regulation, which has significant implications for disease research.

With NGS, scientists can perform multiplexed molecular analyses instead of sequential single-molecule analyses, establishing unbiased starting points for studies, progressing through research faster, and ultimately publishing sooner. The result is high-resolution, quantitative analyses to find signals from more protein precursors. As NGS is adopted into more laboratories and studies continue to map specific genetic profiles to phenotypes, more biologists than ever before have access to powerful genetic tools that can guide their experimental designs.

Illumina is committed to providing the highest data quality in the industry, exemplified by implementation of the largest instrument install base of any NGS technology company[5] and relationships with leaders in many research fields. Together, we are bringing the promise of NGS towards a deeper understanding of human biology and disease.

Glossary

Adapters: Specialized oligos bound to the 5' and 3' ends of each DNA fragment in a sequencing library. The adapter sequences are complementary to the oligos bound to the surface of Illumina sequencing flow cells.

Bridge amplification: An amplification reaction that occurs on the surface of an Illumina flow cell—also known as cluster generation. The flow cell surface is coated with a lawn of two distinct oligonucleotides. Repeated denaturation and extension cycles (similar to PCR) result in localized amplification of a single fragment into thousands of identical fragments. Millions to billions of unique, clonal clusters cover the flow cell. For Illumina NGS, cluster generation occurs on the sequencing instrument or in a separate fluidics instrument called a cBot.

Chromatin immunoprecipitation sequencing (ChIP-Seq): By combining chromatin immunoprecipitation (ChIP) and massively parallel sequencing, ChIP-Seq can be used to accurately survey interactions between protein, DNA, and RNA, enabling the interpretation of regulation events central to many biological processes and disease states.

Clusters: A clonal grouping of template DNA bound to the surface of a flow cell. Seeded by a single, template DNA strand, each cluster is clonally amplified through bridge amplification until the cluster has roughly 1,000 copies. Each cluster on the flow cell produces a single sequencing read. For example, 1 million clusters on a flow cell would produce 1 million reads.

Flow cell: A glass slide with 1, 2, or 8 (depending on instrument platform) physically separated lanes. Each lane is coated with a lawn of surface-bound, adapter-complimentary oligos. A single sample or pool of up to 96 multiplexed samples can be run per lane depending on application parameters.

Indexes: Also known as barcodes or tags, these are unique sequences, usually 8–12 base pairs long that are ligated to fragments in a sequencing library for identification in subsequent data analysis steps. The index sequences (typically part of the adapter) are added during the library preparation stage.

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.

Quantitative real-time polymerase chain reaction (qPCR): Developed in 1992, qPCR is an advanced PCR method that allows accurate quantification of the number of DNA molecules in a sample. It simultaneously amplifies DNA and detects the products of amplification.[6]

Read: A unique sequence resulting from a single cluster on the flow cell. The length of the sequence read depends on the number of programmed sequencing cycles during the instrument run. For example, a 150-cycle sequencing run would produce a 150 base pair read, and 1 million clusters on the flow cell would result in 1 million unique reads. All sequence reads are exported to a data file following the completion of a sequencing run.

Reference genome: A known or previously sequenced genome. The reference genome acts as a scaffold against which new sequence reads are aligned (resequencing). In the absence of a reference genome, contig assembly (*de novo* sequencing) must be used to construct the genome.

Ribosome profiling: Ribosome profiling is a method based on deep sequencing of ribosome-protected mRNA fragments. Purification and sequencing of these fragments provides a snapshot of all the ribosomes active in a cell at a specific time point. This information can determine what proteins are being actively translated in a cell.

RNA sequencing (RNA-Seq): RNA sequencing reveals the products of gene expression, also known as the transcriptome. It has led to breakthrough discoveries about the complexity of RNA species.[7] Unlike arrays, RNA-Seq technology does not require species- or transcript-specific probes. It enables unbiased detection of novel transcripts, gene fusions, single nucleotide variants, indels (small insertions and deletions), and other previously unknown changes that arrays cannot detect.

Sequencing by synthesis (SBS): SBS technology uses four 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.

References

1. Schuster SC. Next-generation sequencing transforms today's biology. *Nat Methods.* 2008;5:16-18.
2. ENCODE Project Consortium. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.* 2011;9:e1001046.
3. Nagai K, Kohno K, Chiba M, et al. Differential expression profiles of sense and antisense transcripts between HCV-associated hepatocellular carcinoma and corresponding non-cancerous liver tissue. *Int J Oncol.* 2012;40:1813-1820.
4. Lunnon K, Smith R, Hannon E, et al. Methylomic profiling implicates cortical deregulation of *ANK1* in Alzheimer's disease. *Nat Neurosci.* 2014;17:1164-1170.
5. Nakazato T, Ohta T, Bono H. Experimental design-based functional mining and characterization of high-throughput sequencing data in the sequence read archive. *PLoS One.* 2013;8:e77910.
6. Higuchi R1, Dollinger G, Walsh PS, Griffith R. Simultaneous amplification and detection of specific DNA sequences. *Biotechnology (NY)*. 1992 Apr;10(4):413-7.
7. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*. 2009 Jan;10(1):57-63.

Illumina • 1.800.809.4566 toll-free (U.S.) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

FOR RESEARCH USE ONLY

© 2015 Illumina, Inc. All rights reserved.
Illumina, CSPro, and the pumpkin orange color are trademarks of Illumina, Inc. and/or its affiliate(s) in the U.S. and/or other countries.
Pub. No. 1070-2014-006 Current as of 09 January 2015

illumina®

AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACG
TCAACGTACCGTAACGAACGTATCATTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTA
CGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACG
AACGTACCATTAAGAGCTACCGTGCAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAA
AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACG
GATTACTTGATCCACTGATTCAACGTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAGCAACGACG
CGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGAAAAGAATGATAACAGTA