Illumina Sequencing Technology

Highest data accuracy, simple workflow, and a broad range of applications.

Introduction

Illumina sequencing technology leverages clonal array formation and proprietary reversible terminator technology for rapid and accurate large-scale sequencing. The innovative and flexible sequencing system enables a broad array of applications in genomics, transcriptomics, and epigenomics.

Cluster Generation

Sequencing templates are immobilized on a proprietary flow cell surface (Figure 1) designed to present the DNA in a manner that facilitates access to enzymes while ensuring high stability of surface-bound template and low non-specific binding of fluorescently labeled nucleotides. Solid-phase amplification (Figures 2–7) creates up to 1,000 identical copies of each single template molecule in close proximity (diameter of one micron or less). Because this process does not involve photolithography, mechanical spotting, or positioning of beads into wells, densities on the order of ten million single-molecule clusters per square centimeter are achieved.

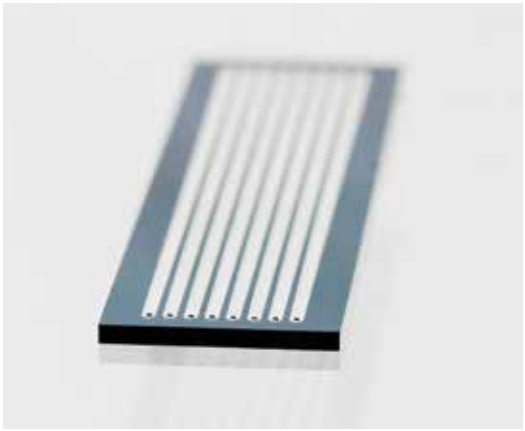
Sequencing by Synthesis

Sequencing by synthesis (SBS) technology uses four fluorescently-labeled nucleotides to sequence the tens of millions of clusters on the flow cell surface in parallel (Figure 8–12). During each sequencing cycle, a single labeled deoxynucleoside triphosphate (dNTP) is added to the nucleic acid chain. The nucleotide label serves as a terminator for polymerization, so after each dNTP incorporation, the fluorescent dye is imaged to identify the base and then enzymatically cleaved to allow incorporation of the next nucleotide. Since all four reversible terminator-bound dNTPs (A, C, T, G) are present as single, separate molecules, natural competition minimizes incorporation bias. Base calls are made directly from signal intensity measurements during each cycle, which greatly reduces raw error rates compared to other technologies. The end result is highly accurate base-by-base sequencing that eliminates sequence-context specific errors, enabling robust base calling across the genome, including repetitive sequence regions and within homopolymers.

Analysis Pipeline

The Illumina sequencing approach is built around a massive quantity of sequence reads in parallel. Deep sampling and uniform coverage is used to generate a consensus and ensure high confidence in determination of genetic differences. Deep sampling allows the use of weighted majority voting and statistical analysis, similar to conventional methods, to identify homozygotes and heterozygotes and to distinguish sequencing errors. Each raw read base has an assigned quality score so that the software can apply a weighting factor in calling differences and generating confidence scores.

Figure 1: Illumina Flow Cell



Several samples can be loaded onto the eight-lane flow cell for simultaneous analysis on an Illumina Sequencing System.

Data Collection, Processing, and Analysis

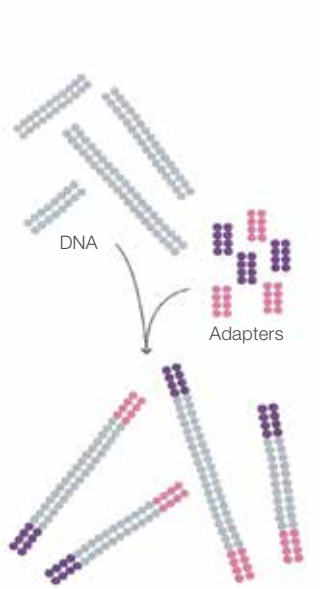
Illumina data collection software enables users to align sequences to a reference in resequencing applications (Figure 13). Developed in collaboration with leading researchers, this software suite includes the full range of data collection, processing, and analysis modules to streamline collection and analysis of data with minimal user intervention. The open format of the software allows easy access to data at various stages of processing and analysis using simple application program interfaces.

Data Accuracy and Workflow Simplicity

The TruSeq family of reagents represents the latest advancement of Illumina's sequencing by synthesis (SBS) technology. From sample prep through DNA sequencing, TruSeq reagent chemistry enables Illumina sequencing to provide the most accurate data across a broad range of applications. With highest yield of error-free reads and most base calls above Q30, researchers can have the highest confidence in their data integrity to draw sound biological conclusions.

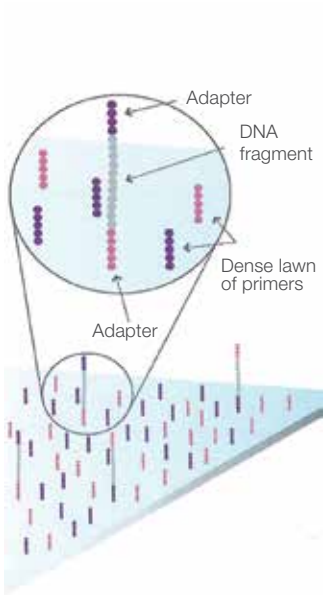
A highly automated, streamlined workflow requires minimal instrument hands-on time. With the ability to generate several gigabases of DNA sequence per run, even large mammalian genomes can be sequenced in weeks rather than years. The capacity to accommodate many samples per flow cell means that runs can be tailored to the demands of diverse applications.

Figure 2: Prepare Genomic DNA Sample



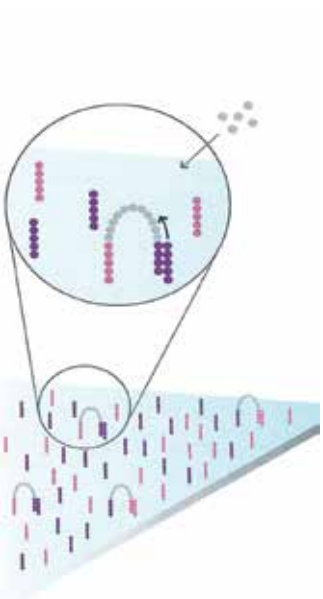
Randomly fragment genomic DNA and ligate adapters to both ends of the fragments.

Figure 3: Attach DNA to Surface



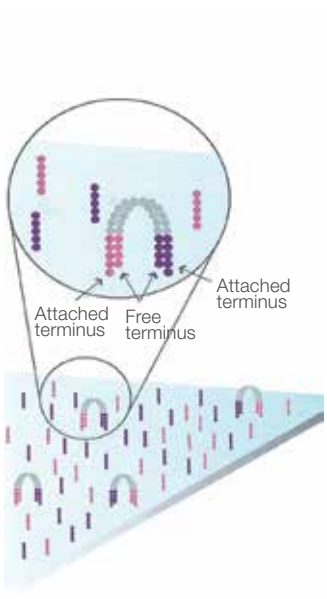
Bind single-stranded fragments randomly to the inside surface of the flow cell channels.

Figure 4: Bridge Amplification



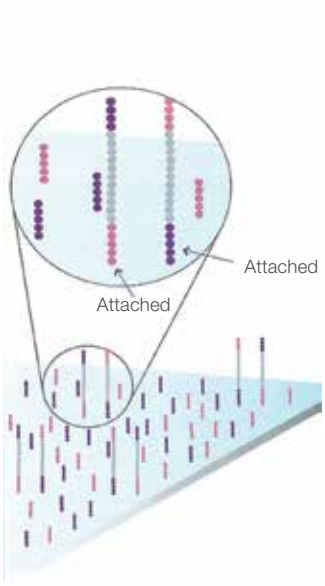
Add unlabeled nucleotides and enzyme to initiate solid-phase bridge amplification.

Figure 5: Fragments Become Double Stranded



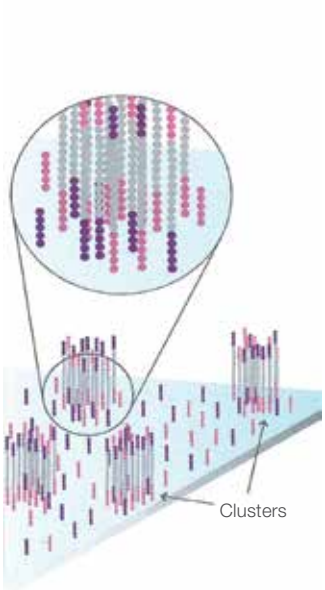
The enzyme incorporates nucleotides to build double-stranded bridges on the solid-phase substrate.

Figure 6: Denature the Double-Standed Molecules



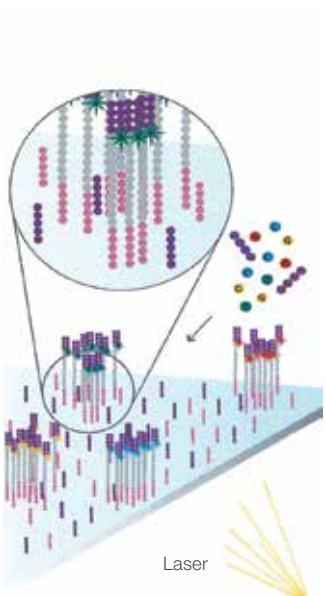
Denaturation leaves single-stranded templates anchored to the substrate.

Figure 7: Complete Amplification



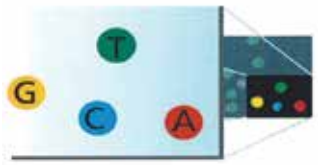
Several million dense clusters of double-stranded DNA are generated in each channel of the flow cell.

Figure 8: Determine First Base



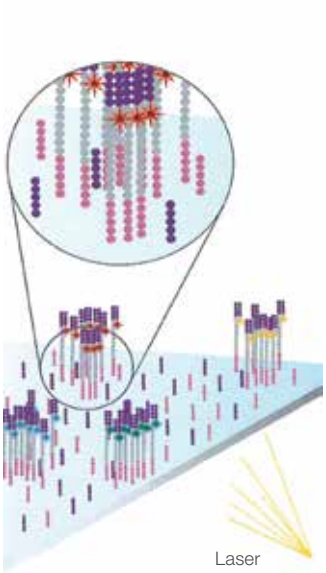
The first sequencing cycle begins by adding four labeled reversible terminators, primers, and DNA polymerase.

Figure 9: Image First Base



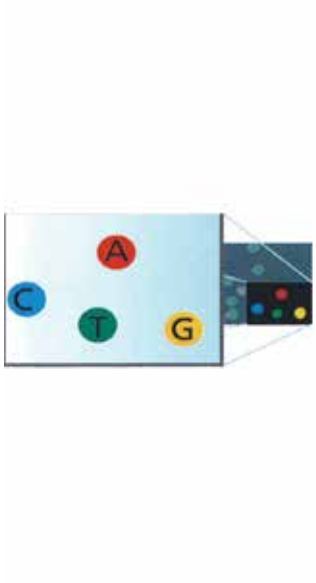
After laser excitation, the emitted fluorescence from each cluster is captured and the first base is identified.

Figure 10: Determine Second Base



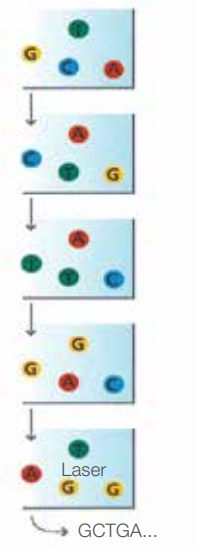
The next cycle repeats the incorporation of four labeled reversible terminators, primers, and DNA polymerase.

Figure 11: Image Second Chemistry Cycle



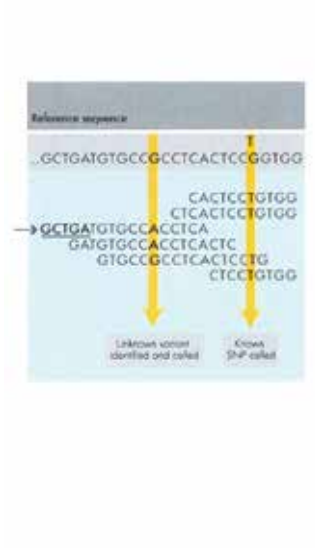
After laser excitation, the image is captured as before, and the identity of the second base is recorded.

Figure 12: Sequencing Over Multiple Chemistry Cycles



The sequencing cycles are repeated to determine the sequence of bases in a fragment, one base at a time.

Figure 13: Align Data



The data are aligned and compared to a reference, and sequencing differences are identified.

Additional Information

Visit our website or contact us at the address below to learn more about Illumina sequencing technology and applications.

Illumina, Inc. • 9885 Towne Centre Drive, San Diego, CA 92121 USA • 1.800.809.4566 toll-free • 1.858.202.4566 tel • techsupport@illumina.com • illumina.com

FOR RESEARCH USE ONLY

Pub. No. 770-2007-002 Current as of 11 October 2010

illumina®