

Multiplexed Sequencing with the Illumina Genome Analyzer System

Introducing index sequences onto DNA fragments enables sequencing of 96 different samples on a single flow cell. This greatly increases experimental scalability, while maintaining extremely low error rates and conserving read length.

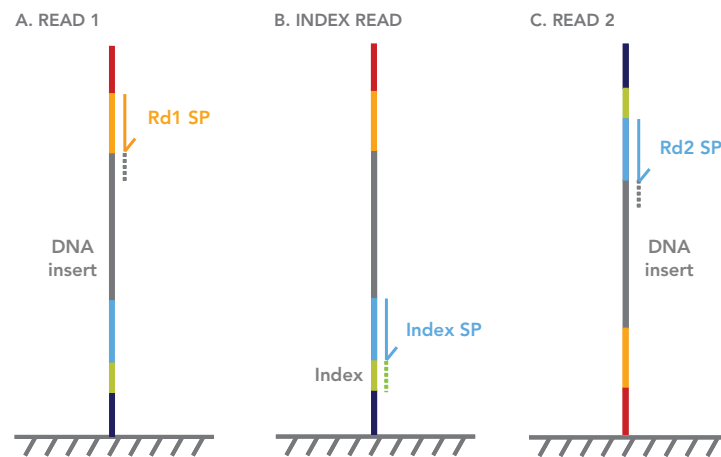
HIGH-THROUGHPUT SEQUENCING

Using the industry's leading next-generation sequencing technology, the Genome Analyzer system offers proven, exceptionally high data yields and the largest number of error-free reads. Harnessing this sequencing power in a multiplex fashion increases experimental throughput while reducing time and cost. This is especially useful when targeting genomic sub-regions or studying small genomes. To make multiplexed sequencing on the Genome Analyzer available to any laboratory, Illumina offers the Multiplexing Sample Preparation Oligonucleotide Kit and the Multiplexing Sequencing Primers and PhiX Control Kit.

HIGHLIGHTS OF ILLUMINA MULTIPLEXED SEQUENCING

- **Fast, High-Throughput Strategy:** Automated sequencing of 96 samples per flow cell
- **Cost-Effective Method:** Multi-sample pooling improves productivity by reducing time and reagent use
- **High-Quality Data:** Accurate maintenance of read length for unknown sequences
- **Simplified Analysis:** Automated sample association with index using Pipeline Analysis software

FIGURE 1: MULTIPLEXED SEQUENCING PROCESS



Sample multiplexing involves a total of three sequencing reads, including a separate index read, which is generated automatically on the Genome Analyzer equipped with the Paired-End Module. A: Application read 1 (dotted line) is generated using the Read 1 Sequencing Primer (Rd1 SP). B: The read 1 product is removed and the Index Sequencing Primer (Index SP) is annealed to the same strand to produce the 6-bp index read (dotted line). C: If a paired-end read is required, the original template strand is used to regenerate the complementary strand. Then, the original strand is removed and the complementary strand acts as a template for application read 2 (dotted line), primed by the Read 2 Sequencing Primer (Rd2 SP). Pipeline Analysis software identifies the index sequence from each cluster so that the application reads can be assigned to a single sample. Hatch marks represent the flow cell surface.

In the multiplexed sequencing method, DNA libraries are “tagged” with a unique identifier, or index, during sample preparation. Multiple samples are then pooled into a single lane on a flow cell and sequenced together in one Genome Analyzer run. An automated three-read sequencing strategy (Figure 1) identifies each uniquely tagged sample

for individual downstream analysis. Using this approach, sample identification is highly accurate.

APPLICATIONS

Multiplexed sequencing on the Genome Analyzer can be used in a wide range of applications. For example, following genome-wide association studies of human

disease, multiplexed sequencing can be performed for time- and cost-effective resequencing of targeted regions in many individuals. Multiplexed sequencing can also be used to characterize small, non-human genomes, such as when determining genetic variations between bacterial strains responsible for separate disease outbreaks. In studies of gene structure and regulation, multiplexed sequencing can be applied to whole-transcriptome sequencing and to DNA recovered by chromatin immunoprecipitation (ChIP) experiments.

UNIQUE INDEX TAGS

In a multiplexed run on the Genome Analyzer, multiple samples are sequenced in a single lane of a flow cell. To identify samples after pooling, each sample is uniquely tagged with a sequence index during the sample preparation protocol.

The Multiplexing Sample Preparation Oligonucleotide Kit provides 12 index oligos for pooling up to 12 samples per lane, or 96 samples per flow cell. Each index is six bases in length. This permits accurate differentiation between samples, even if an index read contains an error.

SIMPLE SAMPLE PREPARATION

Sample preparation for multiplexing on the Genome Analyzer is highly robust and familiar. The straightforward workflow requires as little as one microgram of input DNA. It is based on the simplified Paired-End sample preparation procedure with minimal changes (Figure 2). Index sequences are added to adapter-modified DNA fragments during the PCR enrichment step (Figure 3). The protocol does not require use of restriction enzymes to prepare

fragments, avoiding constraints on read length or fragment size and maximizing yield and data utility. Prepared samples can be used on both single-read and paired-end flow cells.

FULLY AUTOMATED SEQUENCING

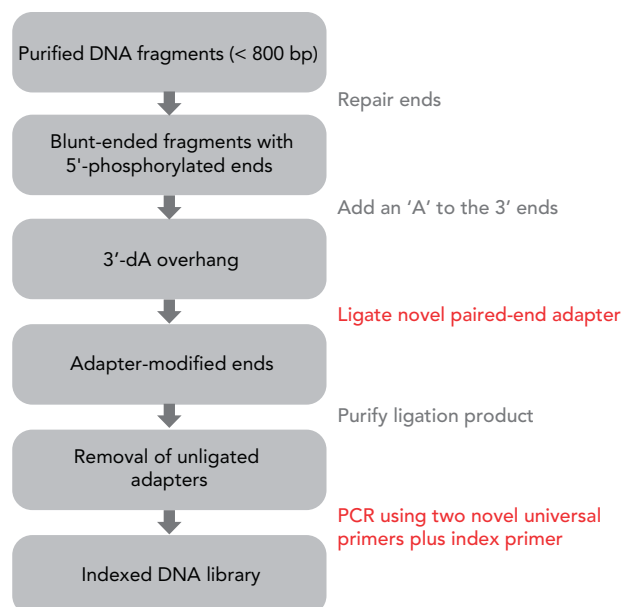
The multiplexed sequencing process is fully automated using the Genome Analyzer, Cluster Station, and Paired-End Module.

The Cluster Station amplifies DNA from up to 96 samples on the flow cell surface to create clusters containing 500–1,000 clonal copies of each molecule. The resulting high-density array of templates is sequenced using the Genome Analyzer and the Paired-End Module.

Sequencing by synthesis is

performed in parallel with novel fluorescently labeled reversible terminator nucleotides. A total of three sequencing reads are performed (Figure 1). The first read is identical to that of a paired-end experiment and uses the standard Read 1 Sequencing Primer provided in the Paired-End Cluster Generation Kit. At the end of the first read, the extended sequencing primer is removed and the Index Sequencing Primer, provided in the Multiplexing Sequencing Primers and PhiX Control Kit, is annealed to the same strand. This approach leverages the Paired-End Module to avoid the loss of high-quality sequencing data from the unknown sample that would occur if the index sequence had been included at the start of an application read.

FIGURE 2: THE MULTIPLEXED SEQUENCING SAMPLE PREPARATION METHOD FOLLOWS THE FAMILIAR PAIRED-END PROTOCOL



The multiplexed sequencing sample preparation method follows standard sample preparation protocols for paired-end sequencing, with the exception of using a novel adapter and PCR primers (shown in red).

Prior to application read 2, the index sequencing product is removed and the clusters are modified *in situ* to regenerate the complementary strand using the Paired-End Module. The Multiplexing Read 2 Sequencing Primer is annealed to the complementary strand and extended to complete the final read.

Using Illumina's Pipeline Analysis software, each index is associated with a particular read-pair, identifying samples for downstream analysis.

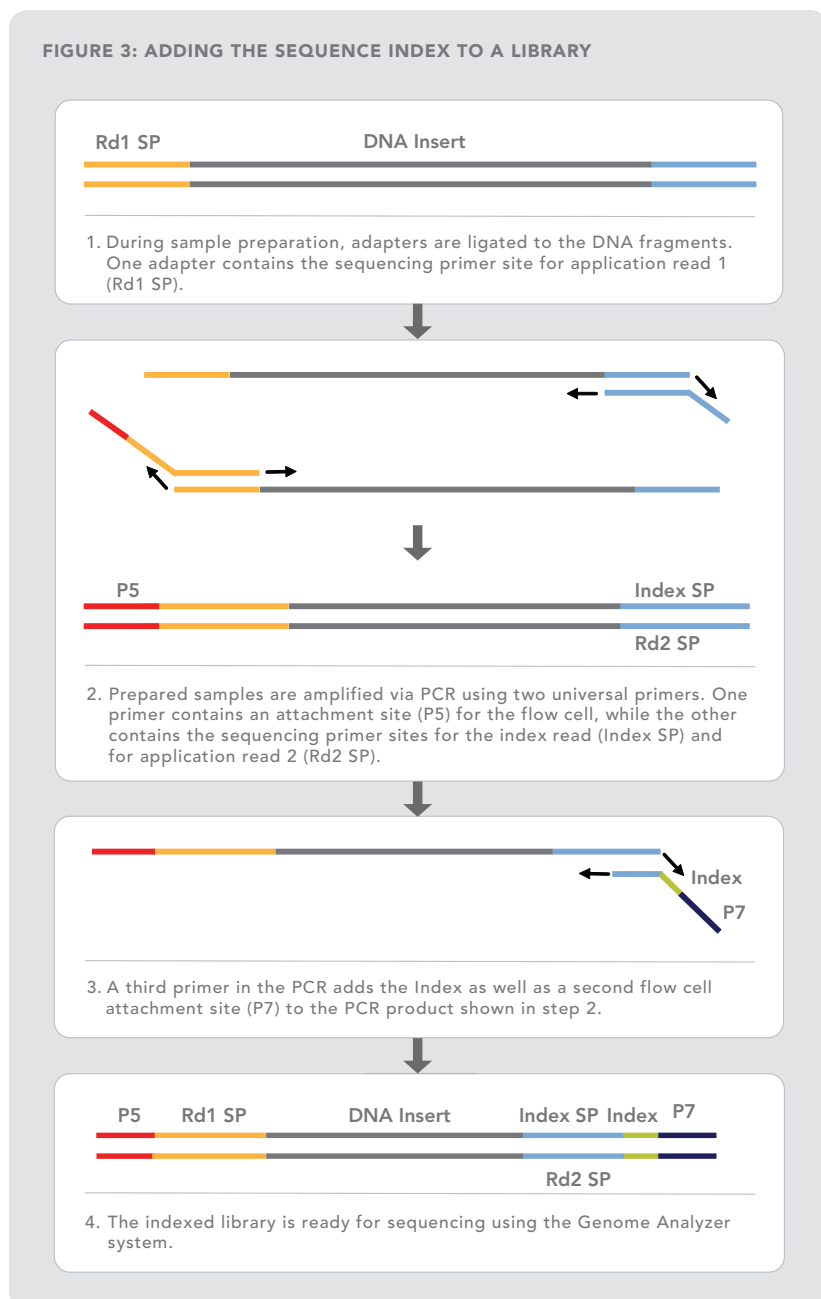
HIGH-QUALITY DATA

Sample multiplexing on the Genome Analyzer system produces high-throughput sequence information with industry-leading accuracy. Data quality is equivalent to that routinely achieved in a single-read or paired-end run. Due to the inherent redundancy in the index design, both perfect index reads and those that differ by one base can be used as sample identifiers (Figure 4).

AUTOMATIC SAMPLE IDENTIFICATION AND PROCESSING

Pipeline Analysis software (version 1.0 and higher) includes the ability to discriminate between the three reads. Once the sequencing chemistry is complete, the alignment software identifies the index sequence and annotates each read with its respective index. From this point on, reads derived from a given multiplexed sample can be positively identified.

Just as in a non-multiplexed sequencing run, Pipeline software provides automated base-calling, calculation of quality values for every base, and alignment of read pairs to a reference sequence. The ability to combine multiplexing with paired-end reads is crucial for optimizing sequence alignment,



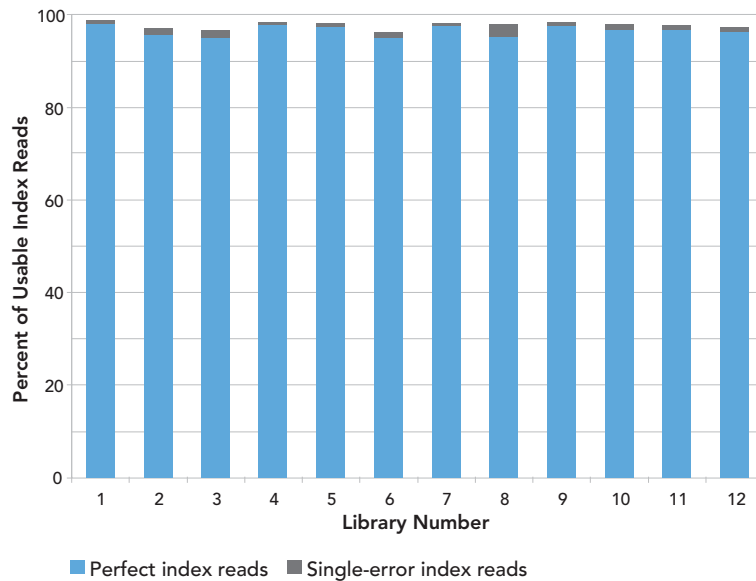
detecting structural variation, and assembling sequences *de novo*.

Like all of Illumina's software solutions, Pipeline software offers an open architecture that allows for easy customization of downstream analysis and integration with a variety of innovative analysis tools.

UNLIMITED ACCESSIBILITY

The Multiplexing Sample Preparation Oligonucleotide Kit and Multiplexing Sequencing Primers and PhiX Control Kit simplify high-throughput multiplexed sequencing with the Genome Analyzer system. In addition to multiplexing, the Genome

FIGURE 4: PERCENTAGE OF USABLE INDEX READS



Most index sequences are perfect, but error correction in the index design means that even the small number of reads with a single error can be used. Twelve libraries were prepared, each tagged with a different index, and sequenced in individual flow cell lanes. The percentage of index sequences that can be used (bar height) is a combination of perfect index reads (blue) and those with a single error (grey).

Analyzer system can be used for single-read sequencing, mRNA-Seq, ChIP-Seq studies, and more. As an open platform for genetic analysis, the Genome Analyzer enables the broadest range of applications. View the latest accomplishments using Illumina sequencing technology at www.illumina.com/publications.

ORDERING INFORMATION

PRODUCT	DESCRIPTION	CATALOG NO.
Multiplexing Sample Preparation Oligonucleotide Kit	Kitted oligonucleotides used to prepare up to 96 samples for multiplexed sequencing*.	PE-400-1001
Multiplexing Sequencing Primers and PhiX Control Kit	Kitted multiplexing sequencing primers, multiplexing control DNA, and buffer set. Sufficient for up to 10 Genome Analyzer runs.	PE-400-1002

*Requires Genomic DNA or Paired-End Sample Prep Kits, available separately.

ADDITIONAL INFORMATION

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

Illumina, Inc.
Customer Solutions
 9885 Towne Centre Drive
 San Diego, CA 92121-1975
 1.800.809.4566 (toll free)
 1.858.202.4566 (outside the U.S.)
techsupport@illumina.com
www.illumina.com

FOR RESEARCH USE ONLY