An Introduction to Illumina Next-Generation Sequencing Technology for HLA Typing

Deciphering DNA sequences is essential for virtually all branches of biological research. Capillary electrophoresis (CE)-based sequencing has enabled scientists to elucidate genetic information from almost any organism or biological system. Although this technology has become widely adopted, inherent limitations in throughput, scalability, cost, speed, and resolution can hinder scientists from obtaining essential genomic information. Next-generation sequencing (NGS) is a fundamentally different approach to studying the genome that overcomes these barriers and paves the way for numerous groundbreaking discoveries. As a result of the introduction of NGS, scientists have experienced a major transformation in the way they extract genetic information from biological systems, revealing insight about the genome, transcriptome, and epigenome.

Illumina leads the NGS revolution, offering the most widely adopted, proven NGS technology. The MiSeq® System is the industry's most accurate, easiest-to-use desktop sequencer, making this technology accessible to all laboratories. With the ability to perform 2 × 300 bp paired-end reads and streamlined workflow, the MiSeq system makes high-resolution HLA typing accessible to any clinical laboratory. This introduction highlights the advantages of using NGS on the MiSeq system for HLA typing.

www.illumina.com/hlaseq

illumina®

Seek greater clarity in HLA typing.

High-resolution, unambiguous HLA typing is challenging due to the highly polymorphic nature of the HLA loci and the high levels of sequence homology between the loci. Using conventional HLA typing methods (e.g., serology, Sequence Specific Oligonucleotide [SSO], Sequence Specific Primers [SSP], and CE Sequence Based Typing [SBT]), resolving ambiguities can be laborious, requiring each sample to be processed through a different series of assays and technologies.

Illumina NGS offers a new paradigm in HLA typing: unambiguous, ultra-high–resolution typing for all loci on dozens of samples in a single assay on a single technology.

Sequence all important HLA genes simultaneously.

NGS has proven instrumental in advancing scientific fields from human disease research to environmental and evolutionary science. It lends itself particularly well to HLA typing. NGS enables sequencing of multiple HLA genes from many samples in a single run. The data generated is higher in resolution compared to conventional methods, yielding accurate results across the entire HLA region mapping to thousands of unique HLA alleles. The ability to interrogate more of the HLA region is becoming more important as new HLA disease associations are discovered. The comprehensiveness of this technique reduces the need for additional testing to resolve ambiguities, decreasing overall turnaround time.

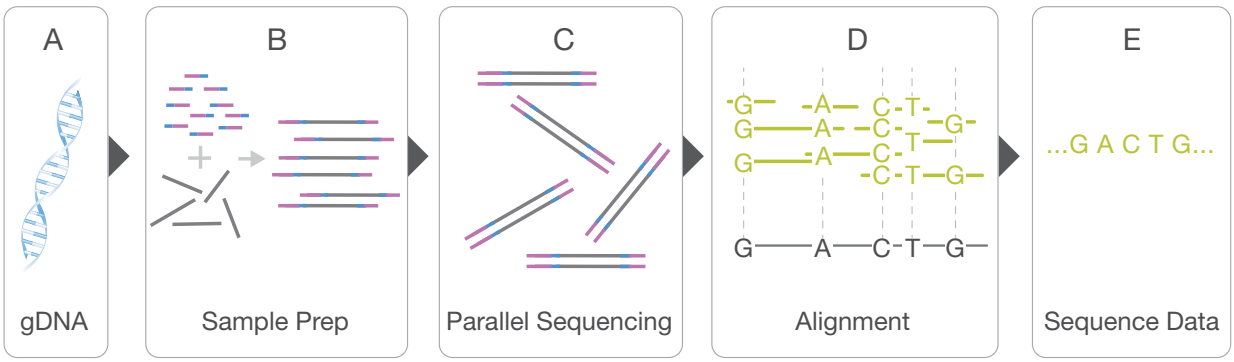
Basic concepts of NGS

In principle, NGS is similar to Sanger (CE-based) sequencing. The bases of a DNA fragment are sequentially identified from signals emitted as each fragment is resynthesized from a DNA 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 DNA fragments. This advance enables rapid sequencing of large stretches of DNA, with the latest instruments capable of producing hundreds of gigabases of data in a single sequencing run*.

To illustrate how this process works, consider a single genomic DNA (gDNA) sample from an individual. The gDNA is first fragmented into a library of smaller segments and sequenced. The newly identified strings of bases, called reads, are then reassembled using a known reference genome as a scaffold (resequencing), or compared to a database of known HLA types. The full set of aligned reads reveals the entire genomic sequence of the sample (Figure 1). After the sample library is prepared, all of the sequencing steps through primary data analysis (base calling) can be performed on a single instrument, facilitating rapid turnaround with minimal hands-on time.

* When using the HiSeq® 2500 next-generation sequencing system. The MiSeq system produces tens of gigabases of data in a single sequencing run.

Figure 1: Concepts of Next-Generation Sequencing



A. Extracted gDNA.
B. Library preparation fragments genomic DNA and adds adapters to the ends.
C. DNA fragments within the library are each sequenced in parallel.
D. Individual sequence reads are aligned to a reference sequence (e.g., a reference genome or known genes).
E. A consensus of aligned reads is generated allowing for subsequent variant calling.

Rapid, straightforward library preparation

The TruSight® HLA Sequencing Panel provides rapid, straightforward, two-step library preparation. It starts with highly specific long-range PCR amplification followed by a modified Nextera® protocol to take advantage of the dense variability and high sequence homology of the HLA region. The Nextera workflow involves simultaneous fragmentation and sequence adapter tagging using standard laboratory equipment. The prepared libraries are ready for sequencing on the MiSeq System.

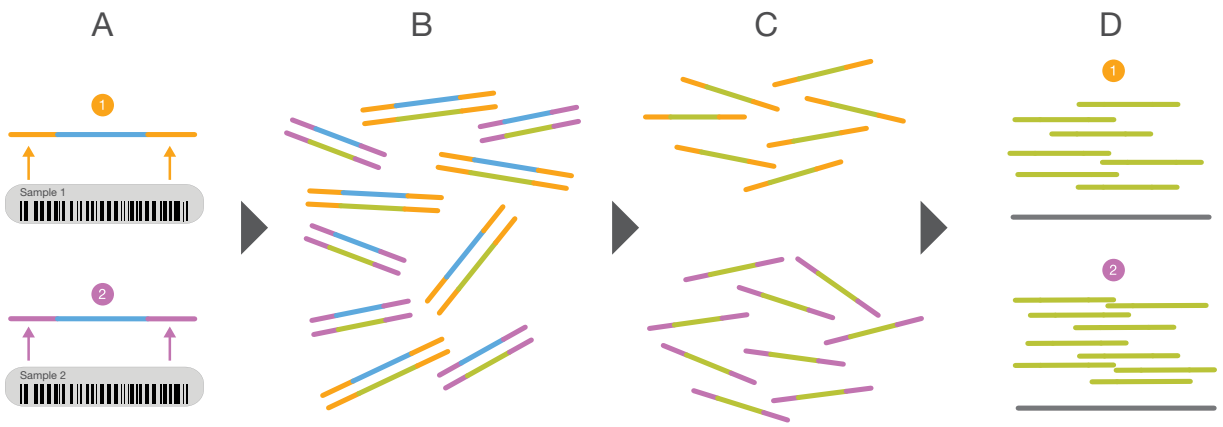
Multiplex for scalable analysis

The Illumina MiSeq NGS system offers the advantage of scalability, providing laboratories with the ability to meet their throughput needs. There's only one system to learn and maintain, enabling personnel to gain technological expertise quickly. The MiSeq can accommodate varying number of samples easily, reducing instrument downtime and increasing throughput in the lab.

When sequencing targeted regions of the genome, such as specific loci or genes of interest, researchers can use a lower output configuration and process a smaller number of samples per run. Alternatively, they can opt to process a larger number of samples through multiplexing. Multiplexing enables large numbers of samples to be simultaneously sequenced during a single experiment (Figure 2). To accomplish this, individual "barcode" sequences are added to each sample so they can be differentiated during the data analysis.

With multiplexing, NGS dramatically reduces the time to HLA typing data for large numbers of samples. HLA typing and analyzing hundreds of amplicons using CE-based sequencing may require several weeks whereas NGS sequencing and analysis can be completed in a few days. With highly automated, easy-to-use protocols, laboratories can go from sample to ultra-high–resolution data to typing faster and easier than ever before.

Figure 2: Sample Multiplexing Overview

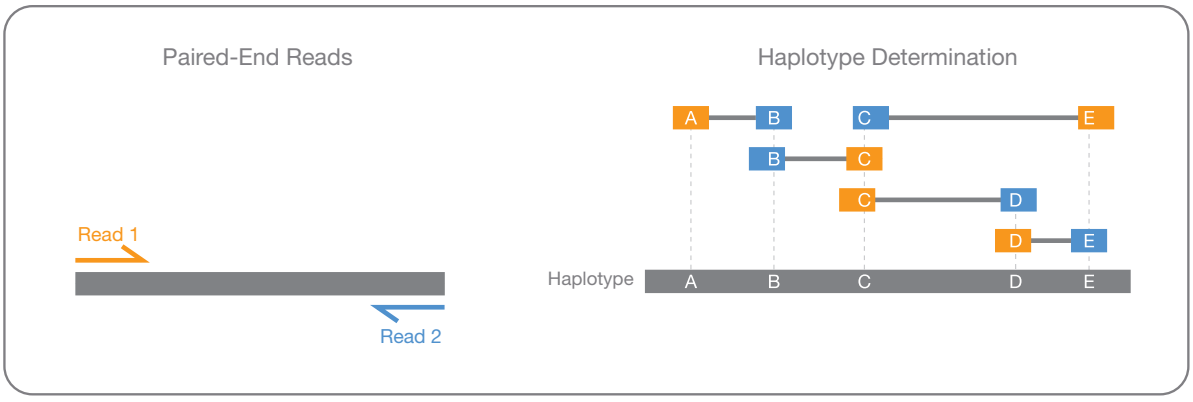


A. Two representative DNA fragments from two unique samples, each attached to a specific barcode sequence that identifies the sample from which it originated.
B. Libraries for each sample are pooled and sequenced in parallel. Each new read contains both the fragment sequence and its sample-identifying 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.

Phase-resolved HLA typing

Illumina NGS technology supports paired-end sequencing, a unique feature that is crucial for successful, unambiguous HLA typing. Sequencing the ends of the library DNA fragments (up to 300 bp on each end) generates high-quality base calls. The physical link between the two reads (originating from the same clonally amplified library DNA fragment) allows association of variants found in each read pair (haplotype). The distance between the paired reads varies as a result of the random library fragment generation process allowing the direct resolution of the phase of two variants that are located > 1 kb apart (Figure 3).

Figure 3: Paired-End Sequencing and HLA Haplotype Alignment



Paired-end sequencing enables both ends of a clonally amplified DNA fragment to be sequenced. If variants are found from the same read pair, then it can be determined that those variants are originally from the same chromosome. The distance between the paired reads can vary as a result of the random library preparation process. Given the highly polymorphic nature of HLA genes, the longer and shorter fragments both play an essential role in resolving phase ambiguity. Paired-end sequencing is automatically performed on the MiSeq System without any user intervention after the sequence run starts, a unique advantage currently unmatched by any competitor.

Analyze, store, and share data

Analysis of HLA sequencing data is challenging, requiring HLA-specific sequence alignment algorithms and comparison to the thousands of references in the IMGT database. Furthermore, analysis software must stay current with the rapid rate of discovery of new HLA alleles. Illumina is collaborating with leading HLA typing software experts to provide analysis software optimized for use with the TruSight HLA Sequencing Panel and Illumina MiSeq System.

MiSeq System: The highest accuracy in NGS-based HLA analysis.

Proven performance

The Illumina MiSeq system is the most accurate, easiest-to-use desktop sequencer available. Its small footprint—approximately four square feet—fits easily into virtually any laboratory environment (Figure 4). The MiSeq system employs Illumina sequencing by synthesis (SBS) technology, the most widely used, proven NGS chemistry with over 4,000 publications to date. For a list of publications citing the use of the MiSeq system for HLA analysis, visit www.illumina.com/hlaseq.

Exceptional data quality

The MiSeq system offers the ability to accurately sequence entire HLA genes with the highest resolution for accurate HLA analysis and phasing. Sequencing a larger region in one run provides data for more in-depth analyses, including discovery of new alleles, without additional cost and effort. The exceptional data quality is a product of long 2 × 250 bp read lengths, > 100× depth of coverage, and a consensus accuracy of 99.999%. Base calls are made directly from signal intensity measurements during each cycle, greatly reducing raw read error rates compared to other technologies[1–5]. The result is highly accurate base-by-base sequencing that virtually eliminates context-specific errors, even within repetitive sequence regions or homopolymers. Illumina sequencing delivers the highest yield of error-free data for the most sensitive sequencing samples.

Figure 4: MiSeq System

The Illumina MiSeq system is the ideal next-generation sequencing platform for high-resolution, cost-effective HLA typing.

AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACG
TCAACGTACCGTAACGAACGTATCATTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTA
CGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACC
AACGTACCATTAAGAGCTACCGTGCAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAA
AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACG
GATTACTTGATCCACTGATTCAACGTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAGCAACGACG
CGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGAAAAGAATGATAACAGTA

Accurate, Phase-Resolved HLA Typing

K Hosomichi et al.[6] demonstrated use of the MiSeq system and Nextera DNA Library Preparation Kit for accurate, phase-resolved sequencing of HLA genes. PCR amplicons encompassing six full-length HLA genes were used to generate libraries, which were then sequenced on the MiSeq system. The long paired-end reads enabled phase determination, an advantage over conventional HLA typing methods, as well as novel allele discovery.

Find your match.

TruSight HLA and the MiSeq System enable phase-resolved, sample-to-report HLA typing for eight loci in a single assay, with a single workflow, on a single instrument, for dozens of samples simultaneously. Gone are the days of resolving ambiguities for every sample at every locus; TruSight HLA delivers an ultra-high–resolution typing result the first time.

Learn more at www.illumina.com/hlaseq.

References

1. Junemann S, Sedlazeck FJ, Prior K, Albersmeier A, Uwe J, et al. (2013) Updating benchtop sequencing performance comparison. Nat Biotechnol. 31: 294–296.
2. Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, et al. (2013) Characterizing and measuring bias in sequence data. Gen Biol. 14: R51.
3. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, et al. (2012) Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol 30: 434–439.
4. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, et al. (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics 13: 341.
5. Liu L, Li Y, Li S, Hu N, He Y, et al. (2012) Comparison of next-generation sequencing systems. J Biomed Biotechnol. 2012: 251364.
6. Hosomichi K, Jinam TA, Mitsunaga S, Nakaoka H, Inoue I (2013) Phase-defined complete sequencing of the HLA genes by next-generation sequencing. BMC Genomics 14: 355.

Learn how you can achieve accurate, phase-resolved HLA typing with TruSight HLA and the MiSeq system at www.illumina.com/hlaseq.

FOR RESEARCH USE ONLY

© 2013–2014 Illumina, Inc. All rights reserved.
Illumina, Genetic Energy, HiSeq, MiSeq, Nextera, TruSight, the pumpkin orange color, and the Genetic Energy streaming bases design are trademarks or registered trademarks of Illumina, Inc. All other brands and names contained herein are the property of their respective owners.
Pub No. 0770-2013-038 Current as of 27 May 2014

illumina®

AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACG
TCAACGTACCGTAACGAACGTATCATTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTA
CGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACG
AACGTACCATTAAGAGCTACCGTGCAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAA
AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACG
GATTACTTGATCCACTGATTCAACGTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAGCAACGACG
CGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGAAAAGAATGATAACAGTA