An Introduction to Next-Generation Sequencing for *in vitro* Fertilization

www.illumina.com/IVFprimer

illumina®

Table of Contents

Part I. Welcome to Next-Generation Sequencing

NGS for *in vitro* Fertilization

Deciphering DNA sequences is essential for virtually all branches of biological research. With its unprecedented throughput, scalability, and speed, next-generation sequencing (NGS) enables more in-depth analysis of the genome than ever before, revealing new biological insights. In addition, NGS offers a faster, more comprehensive, streamlined method with more precise performance and higher throughput capabilities for obtaining DNA data than embryo morphology-based reviews. This new level of access to DNA information has catalyzed several important breakthroughs, making significant impacts in reproductive and genetic health studies, in particular in the field of *in vitro* fertilization (IVF).

As infertility rates rise and women wait until later in life to have children, so does the demand for IVF and assisted pregnancy. Unfortunately, IVF success rates remain low with only 30–35% of cycles resulting in a live birth.[1] Chromosome aneuploidy (abnormal number of chromosomes) is a major cause of IVF failure. Most embryos with aneuploidy do not implant, and those embryos that do often miscarry during the first trimester of pregnancy.[2,3] Using preimplantation genetic screening (PGS) to identify euploid embryos, those embryos with a normal number of chromosomes, increases the chance that a viable embryo is selected for transfer during IVF, resulting in:

- Improved implantation rates[4,5]
- Reduced spontaneous abortion[4,6]
- Increased ongoing pregnancy rates[4,6]

Using NGS to perform PGS delivers highly detailed, accurate, and scalable results for improved selection of euploid embryos over visual assessment alone.[7]

According to Dagan Wells, Associate Professor at the University of Oxford and Director at Reprogenetics-UK, "With next-generation sequencing technologies and PGS from Illumina, we can select embryos that are more likely to result in successful pregnancies".

Simon Fishel, Managing Director of CARE Fertility, adds "PGS is a game changer for IVF clinics and labs. The ability to assess all 24 chromosomes in an embryo and the rapid delivery of the VeriSeq™ PGS Kit* empowers clinicians with accurate chromosomal insights into the health of the embryo to optimize IVF success."

With NGS, researchers and clinicians have unprecedented access to genetic information. By enabling simultaneous analysis of multiple samples in fully automated workflows, NGS opens opportunities to increase efficiency, reduce costs, and shorten workflows. By generating highly specific and sensitive results, Illumina NGS solutions are setting new standards of success in the IVF landscape, enabling informed reproductive choices that can change patients' lives.

Part II. What is Next-Generation Sequencing?

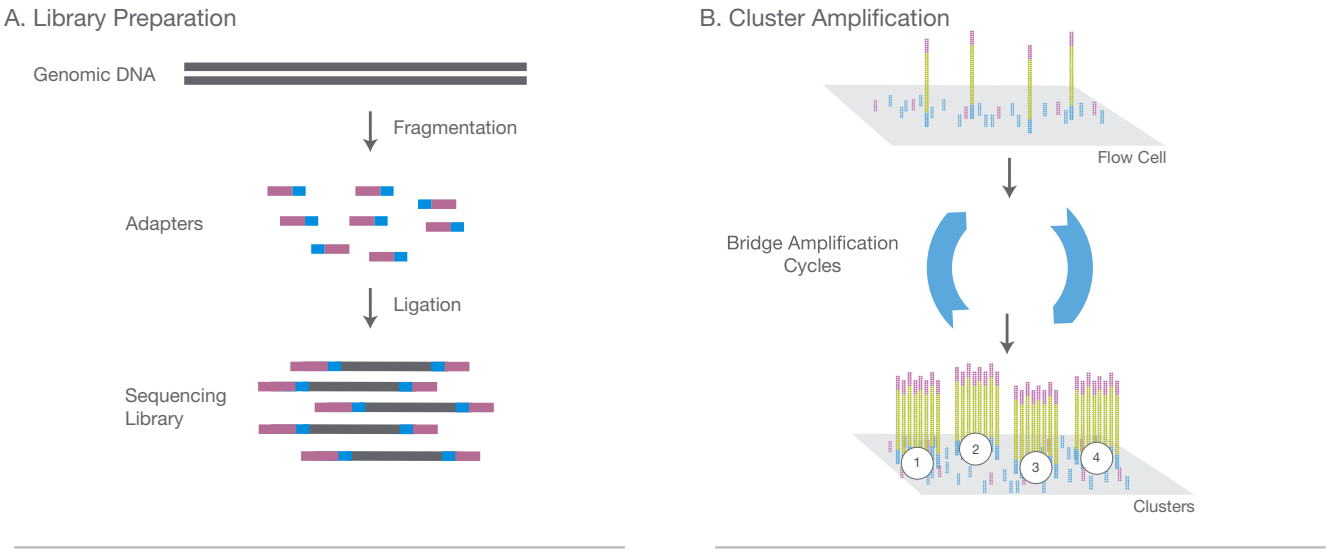
The Basic NGS Workflow

All Illumina NGS workflows include 4 basic steps.

1. Sequencing library preparation begins by creating short DNA or cDNA fragments with 5' and 3' adapters ligated (Figure 1A).
2. During 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 (Figure 1B). Next, the templates are ready for sequencing by synthesis (SBS). SBS technology uses a proprietary reversible terminator–based method that detects single bases as they are incorporated into DNA template strands.[8] 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.[9,10]
3. The result is highly accurate base-by-base sequencing that virtually eliminates sequence-context-specific errors, even within repetitive regions and homopolymers (Figure 1C).
4. The newly identified sequence reads are then exported to an output file and aligned to a reference genome by sequencing alignment software (Figure 1D).

A detailed animation of SBS sequencing is available at
www.youtube.com/watch?v=HMyCqWhwB8E&list=UUxWMU29FF4klG8YmQf6Zv0g

* The Illumina VeriSeq PGS Kit uses NGS to provide comprehensive, accurate screening of all 24 chromosomes simultaneously with an improved assay workflow and higher throughput compared to other methods.7



NGS library is prepared by fragmenting a gDNA sample and ligating specialized adapters to both fragment ends.

Library is loaded into a flow cell and the fragments hybridize to the flow cell surface. Each bound fragment is amplified into a clonal cluster through bridge amplification.

Sequencing reagents, including fluorescently labeled nucleotides, are added to the flow cell and the first base is incorporated. The flow cell is imaged and the emission from each cluster is recorded. The emission wavelength and intensity are used to identify the base. This cycle is repeated “n” times to create a read length of “n” bases.

Reads are aligned to a reference sequence with bioinformatics software. After alignment, differences between the reference genome and the newly sequenced reads can be identified.

Figure 1: Next-Generation Sequencing (NGS) Workflow—The Illumina NGS workflow follows 4 basic steps: library preparation, cluster generation, sequencing, and data alignment.

Multiplexing

In addition to the rise of data output per run, the sample throughput per run in NGS 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 2). With multiplexed samples, unique index sequences are added to each DNA fragment during library preparation so that each read can be identified and sorted before final data analysis. Use of these indexes 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.

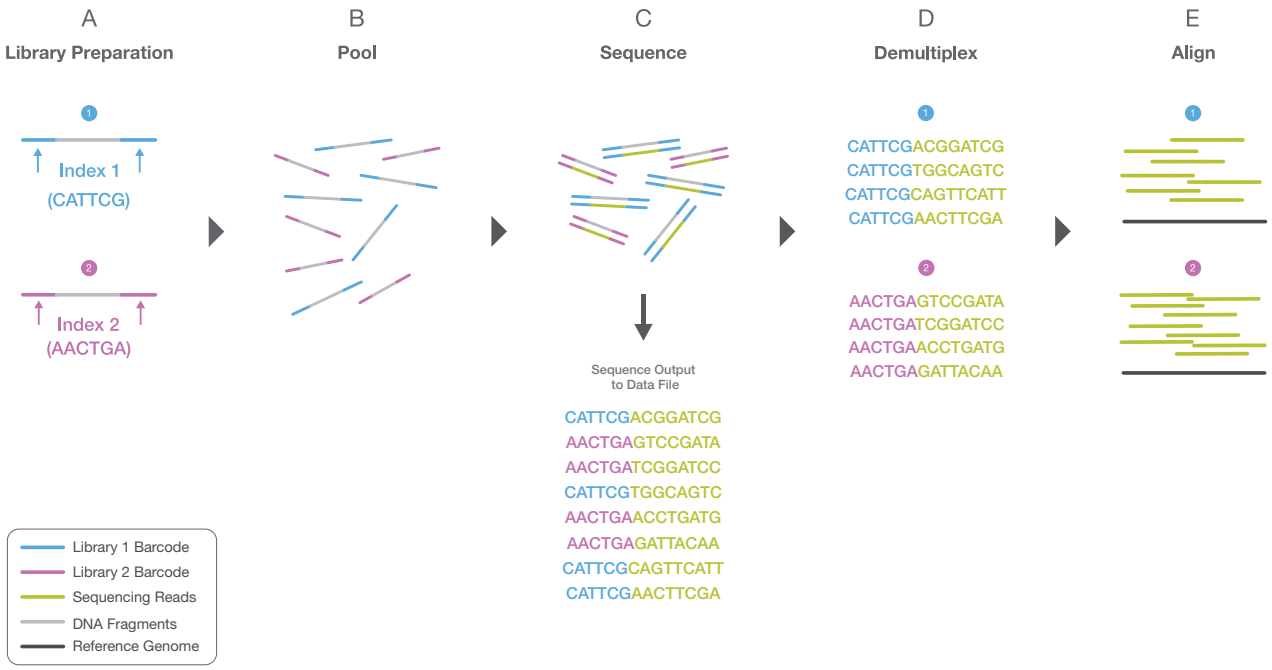


Figure 2: Library Multiplexing Overview—A. Two distinct samples are attached to unique index sequences. Index sequences are attached during library preparation. B. Libraries for each sample are pooled together and loaded into the same flow cell lane. C. Samples are sequenced together during a single instrument run. All sequences are exported to a single output file. D. A demultiplexing algorithm sorts the reads into different files according to their indexes. E. Each set of reads is aligned to the appropriate reference sequence.

Part III. NGS Methods

NGS platforms enable a wide variety of applications, allowing researchers to investigate the genome, transcriptome, or epigenome of any organism. After sequencing libraries are prepared, the actual sequencing process remains fundamentally the same regardless of the application. Streamlined workflows, higher throughput capabilities, and customizable assays support easy portfolio expansion for any laboratory.

Sequencing for PGS

PGS performed using NGS follows a method known as targeted sequencing. With targeted sequencing, a preselected subset of genes is isolated and sequenced. These customized content sets are referred to as sequencing panels. Due to the tightly defined content, sequencing panels allow the genes within each sample to be sequenced to a greater depth than large-scale approaches like whole-genome or exome sequencing. This multigene approach confers several advantages:

- Decreases time-to-answer
- Minimizes issues with limited material for sequential testing
- More accurate and reliable results

Additional Sequencing Methods

The same platform used for sequencing in PGS can be used for various additional methods, including exome and whole-genome sequencing. This expands the utility of the system and enables users to expand their laboratory offerings.

Exome Sequencing

The exome, the protein-coding portion of the human genome, represents less than 2% of the genetic code, but contains ~85% of known disease-related variants.[11] With exome sequencing, the exonic regions of the genome are isolated and sequenced allowing researchers to focus on specific genetic regions of interest.

Whole-Genome Sequencing

Whole-genome sequencing (WGS) involves sequencing the full genome of a given organism. WGS can provide insight into the intronic (noncoding) regions of the genome that exome sequencing does not capture.

Part IV. 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 genetic screening. NGS has enabled significant strides in reproductive and genetic health, notably in IVF, setting a new standard in PGS, improving IVF success rates, and changing patients' lives.

Illumina is committed to providing the highest-quality data in the industry, setting a new standard of IVF success.

- Advancing trusted PGS and PGD genomic solutions
- Delivering accurate, timely, and reliable answers
- Offering educational resources and support

Learn more about Illumina NGS solution for reproductive and genetic health at www.illumina.com/NGSforIVF.

Glossary

Adapters: Specialized oligos bound to the 5'and 3'end 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.

Aneuploidy: Chromosomal abnormality in which there is an abnormal number of chromosomes (too many or too little) in the nucleus of a cell, giving rise to genetic disorders

Aneuploidy Screening: The procedure of screening cells to assess the presence of a correct complement of chromosomes

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 2 distinct oligonucleotides. Repeated denaturation and extension cycles (similar to PCR) results in localized amplification of a single fragment into thousands of identical fragments. Millions to billions of unique, clonal clusters across the flow cell. For Illumina NGS, cluster generation occurs on the sequencing instrument or in a separate fluidics instrument called a "cBot."

Clusters: A clonal grouping of template DNA bound to the surface of a flow cell. Each cluster is seeded by a single template DNA strand 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, 1 million clusters on a flow cell would produce 1 million reads.

Euploidy: Having the normal number of chromosomes in the nucleus of a cell

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.

***In vitro* Fertilization (IVF):** Fertilization of the oocyte by sperm in a petri dish, outside of the human body

Indexes: Also known as "barcodes" or "tags," indexes 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.

Preimplantation Genetic Screening (PGS): Screening embryos or oocytes for the correct number and type of chromosomes; PGS does not look for a specific genetic disorder

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. One 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, the genome must be constructed by contiguous assembly (*de novo* sequencing)

Sequencing by Synthesis (SBS): SBS technology 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.[8]

A global genomics leader, Illumina delivers complete next-generation sequencing workflow solutions to the basic and translational research communities. Illumina technology is responsible for generating more than 90% of the world's sequencing data. Through collaborative innovation, Illumina is fueling groundbreaking advancements in the fields of oncology, reproductive health, genetic disease, microbiology, agriculture, and forensic science.

References

1. Clinic Summary Report. sart cors. https://www.sartcorsonline.com/rptCSR_PublicMultYear.aspx?ClinicPKID=0. Accessed May 11, 2015.
2. Scott RT Jr, Ferry K, Su J, et al. Comprehensive chromosome screening is highly predictive of the reproductive potential of human embryos: a prospective, blinded, nonselection study. *Fertil Steril*. 2012;97(4):870-875. doi: 10.1016/j.fertnstert.2012.01.104.
3. Tobias E, Connor JM, Ferguson-Smith M *Essential Medical Genetics.* 6th ed. Chichester, West Sussex, UK: Wiley-Blackwell; 2011.
4. Grifo JA, Hodes-Wertz B, Lee HL, et al. Single thawed euploid embryo transfer improves IVF pregnancy, miscarriage, and multiple gestation outcomes and has similar implantation rates as egg donation. *J Assist Reprod Genet*. 2013;30(2):259-264. doi: 10.1007/s10815-012-9929-1.
5. Scott RT Jr, Upham KM, Forman EJ, et al. Blastocyst biopsy with comprehensive chromosome screening and fresh embryo transfer significantly increases *in vitro* fertilization implantation and delivery rates: a randomized controlled trial. *Fertil Steril*. 2013;100(3):697-703. doi: 10.1016/j.fertnstert.2013.04.035.
6. Forman EJ, Hong KH, Ferry KM, et al. *In vitro* fertilization with single euploid blastocyst transfer: a randomized controlled trial. *Fertil Steril.* 2013;100(1):100-107 e101. doi: 10.1016/j.fertnstert.2013.02.056.
7. Fiorentino F, Bono S, Biricik A, et al. Application of next-generation sequencing technology for comprehensive aneuploidy screening of blastocysts in clinical preimplantation genetic screening cycles. *Hum Reprod*. 2014;29(12):2802-2813. doi: 10.1093/humrep/deu277.
8. Bentley DR, Balasubramanian S, Swerdlow HP, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 2008;456(7218):53-59. doi: 10.1038/nature07517.
9. 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(10):e77910. doi: 10.1371/journal.pone.0077910.
10. Ross MG, Russ C, Costello M, et al. Characterizing and measuring bias in sequence data. *Genome Biol*. 2013;14(5):R51. doi: 10.1186/gb-2013-14-5-r51.
11. van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. Ten years of next-generation sequencing technology. Trends Genet. 2014;30(9):418-426. doi: 10.1016/j.tig.2014.07.001.

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

For Research Use Only. Not for use in diagnostic procedures.

illumina®

AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACG
TCAACGTACCGTAACGAACGTATCATTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTA
CGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACC
AACGTACCATTAAGAGCTACCGTGCAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAA
AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACG
GATTACTTGATCCACTGATTCAACGTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAGCAACGACG
CGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGAAAAGAATGATAACAGTA