Cytogenetics to Cytogenomics: An Introduction to Genomic Technologies

www.illumina.com/cytogenetics

illumina®

Table of Contents

I. From Microscopes to Genomes

To understand the role that chromosomal variations play in both constitutional disorders and cancer, cytogenetic analysis is an integral part of current genomic medicine. Chromosomal abnormalities, including aneuploidies, deletions, duplications, and rearrangements, may result in misregulation of gene expression or generation of novel proteins. Such abnormalities are a common cause of cancer, infertility, and various congenital disorders such as Down syndrome, DiGeorge syndrome, and cri du chat syndrome.[1]

Traditional cytogenetic studies use karyotyping and/or fluorescence *in situ* hybridization (FISH) to analyze chromosomes. These approaches have been instrumental in identifying major chromosomal abnormalities such as monosomies, trisomies, chromosomal rearrangements, and large deletions or duplications. However, these methods are limited by low resolution or narrow target range. For example, karyotyping is capable of detecting only large chromosomal changes (typically > 5 Mb).[2] It is a subjective technique, and detection of abnormalities often depends on the expertise of the analyst. FISH and real-time quantitative PCR (qPCR) are highly targeted approaches, but require prior knowledge of the abnormality.

Due to recent advancements in high-throughput technologies, cytogeneticists are using genomics tools to overcome these limitations when studying chromosomal abnormalities. Compared to karyotyping, arrays offer a more reliable method for identifying copy number variations and unbalanced chromosomal rearrangements. Because it offers wider genome coverage with higher resolution, array-based research is often used as a first line of testing to identify both small and large alterations associated with congenital disorders and various cancer subtypes.[3–6]

The most recent evolution of genomic technology is next-generation sequencing (NGS), a high-throughput process for determining DNA sequences. Sequencing can be used to explore variants detected by arrays or to identify small variants missed by other methods. By providing a base-by-base view of the genome, NGS can identify single nucleotide variants (SNV), small structural changes, and balanced translocations, increasing information while decreasing costs with a genome-wide view of variation.

> "Now that we have whole-genome tools such as microarrays and NGS, we can analyze the entire genome from a few malignant cells at a resolution that allows us to identify abnormalities and look for rearrangements with a very high degree of precision."
>
> —Trilochan Sahoo, MD, Director of Cytogenetics, CombiMatrix[7]

Cytogeneticists can use complementary array and NGS technologies for thorough analysis of chromosomal changes to obtain a comprehensive view of genomic alterations, both large and small (Table 1). Genomic approaches can detect:

- Amplifications and deletions
- Loss of heterozygosity (LOH)
- Uniparental disomy (UPD)
- Balanced translocations or inversions
- Unbalanced translocations
- Mosaicism
- Copy number variations (CNVs)
- Single Nucleotide Variations (SNVs)
- Small insertions and deletions
- Structural variants

Table 1: Genomic Technologies for Cytogenetics

	Parameter	FISH	Karyotyping	Arrays	NGS (large gene panel)	NGS (whole genome)
	Whole-Genome View	No	Yes	Yes	No	Yes
	Resolution	> 50 kb	> 5 Mb	< 1 kb	1 base pair	1 base pair
	Aneuploidy	Yes	Yes	Yes	Yes	Yes
	Unbalanced Translocation	Yes, if known	Yes, if large	Yes	Yes[a]	Yes
	Balanced Translocation or Inversion	Yes, if known	Yes	No	No	Yes
	Mosaicism	Yes	Yes	Yes, if 20% of cells present	Yes	Yes
	Polyploidy	Yes (indirect)	Yes	Yes (SNP arrays only)	Yes	Yes
	UPD	No	No	Yes (SNP arrays only)	Yes[b]	Yes
	Copy-Neutral LOH	No	No	Yes (SNP arrays only)	No	Yes
	SNVs	No	No	No	Yes	Yes
	Gene Fusions	Yes	No	Yes, if unbalanced No, if balanced	Yes	Yes

Using arrays and NGS to complement traditional methods, cytogeneticists can obtain a comprehensive view of genetic abnormalities, both large and small.[2]
a. Yes, if branch points are targeted.
b. Yes, if both parents are analyzed.

"We are moving from cytogenetics to cytogenomics. It seems in the near future, array CGH and NGS will replace conventional cytogenetics in most clinical and research applications."[2]

II. Genomic Technologies for Cytogenetics

Genomic technologies improve upon previous methods by allowing a wider view of the genome in a single assay. Array and NGS applications offer advantages in resolution and scale for the detection of chromosomal abnormalities.

Arrays

Both FISH and array methods rely upon nucleic acid hybridization, with the use of designed probes to detect specific DNA targets. However, arrays can probe thousands of genetic loci simultaneously, providing wider coverage of the genome and higher throughput in initial stages of testing than FISH. Arrays can validate known abnormalities, such

as deletions or amplifications, previously identified using standard cytogenetics, and discover numerous abnormalities that were below the resolution of other cytogenetic tests.[8]

Chromosomal abnormalities that don't involve copy number variations, such as inversions and balanced translocations, might not be detected by arrays measuring signal intensity only.[8] Single nucleotide polymorphism (SNP) arrays combine intensity information with genotyping information to enable identification of abnormalities that would be missed by either parameter alone, such as uniparental disomy (UPD). Combining genotyping with intensity information can also extract copy number information in a background of normal cells in a heterogenous population. Intercellular heterogeneity, also referred to as mosaicism, is often found in tumor samples mixed with normal cells, as well as some genetic disorders. Mosaicism can be detected on Infinium® SNP genotyping arrays (Figure 1), but might not be correctly identified on array CGH platforms that provide only intensity information.[9]

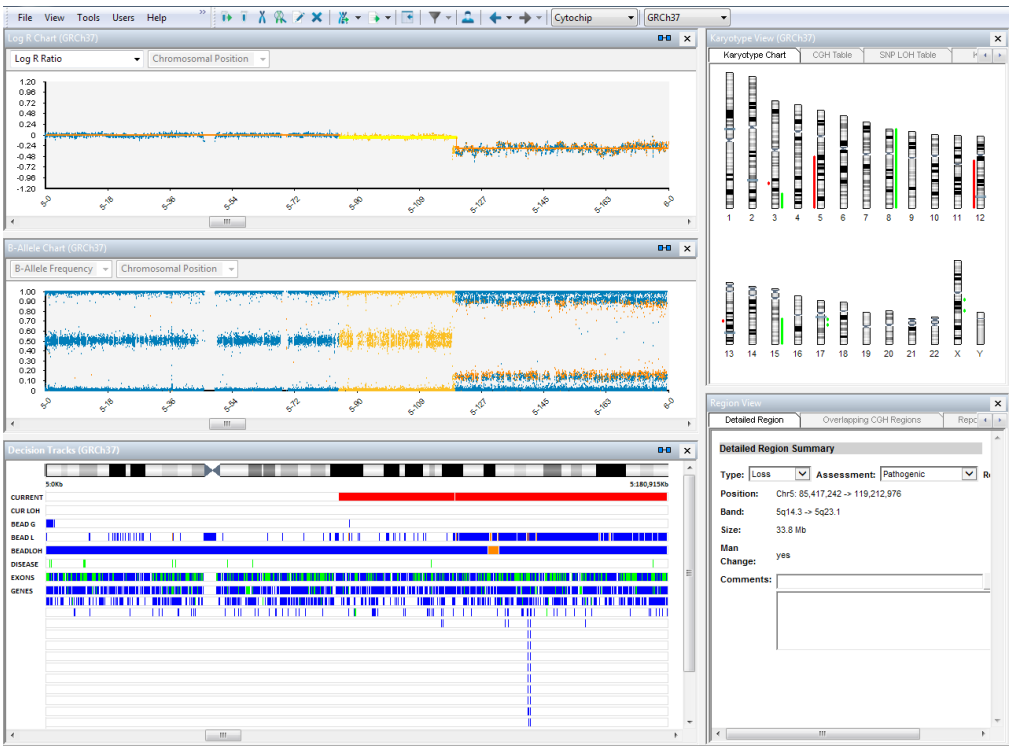


Figure 1: Genome-Wide View of Chromosomal Alterations—Horizontal view of the entire genome from BlueFuse® Multi Software showing mosaicism detection. By correlating the intensity information from the log R ratio (top plot) and the genotyping information from the B Allele Frequency Chart (middle plot), a mosaic deletion is evidenced on the long arm of chromosome 5. A region demonstrating low-level mosaic loss (highlighted yellow) lies adjacent to a region of high-level mosaic loss, which is indicated by the position of the heterozygous track on the B Allele Frequency Chart. The mosaic is unlikely to have been correctly characterized on an array CGH platform, which provides intensity information only. Data shown is from the Infinium CytoSNP-850K BeadChip.

"NGS enables the detection of unexpected aberrations, such as SNVs, that aren't well characterized. Pairing these 2 methods—dose analysis for CNVs and sequence analysis for SNVs—will determine the genomic basis for an increasing number of inherited disorders."

—Udo Koehler, PhD, Lab Director, MGZ Medical Genetics Center, Munich, Germany[10]

Next-Generation Sequencing

Arrays can be designed for wide coverage of the genome with specific applications. However, there are a few things that they cannot detect, such as balanced rearrangements.[7] Although array technologies can detect variations at high resolution, new sequence variations cannot be detected, as prior knowledge is required for assay design. NGS data can complement array data by providing additional information on single nucleotide level sequence variations, and chromosomal abnormalities such as balanced translocations.[11]

NGS technology is fundamentally similar to traditional Sanger sequencing, resulting in single base resolution of DNA samples. The critical difference is that, instead of sequencing a single DNA fragment, NGS extends this process across millions of fragments in a massively parallel fashion, providing enhanced coverage of the genome and higher throughput. As with arrays, NGS offers flexibility in its application. Cytogeneticists have the option of rapidly sequencing the entire genome or deeply sequencing targeted regions.

III. What is Next-Generation Sequencing?

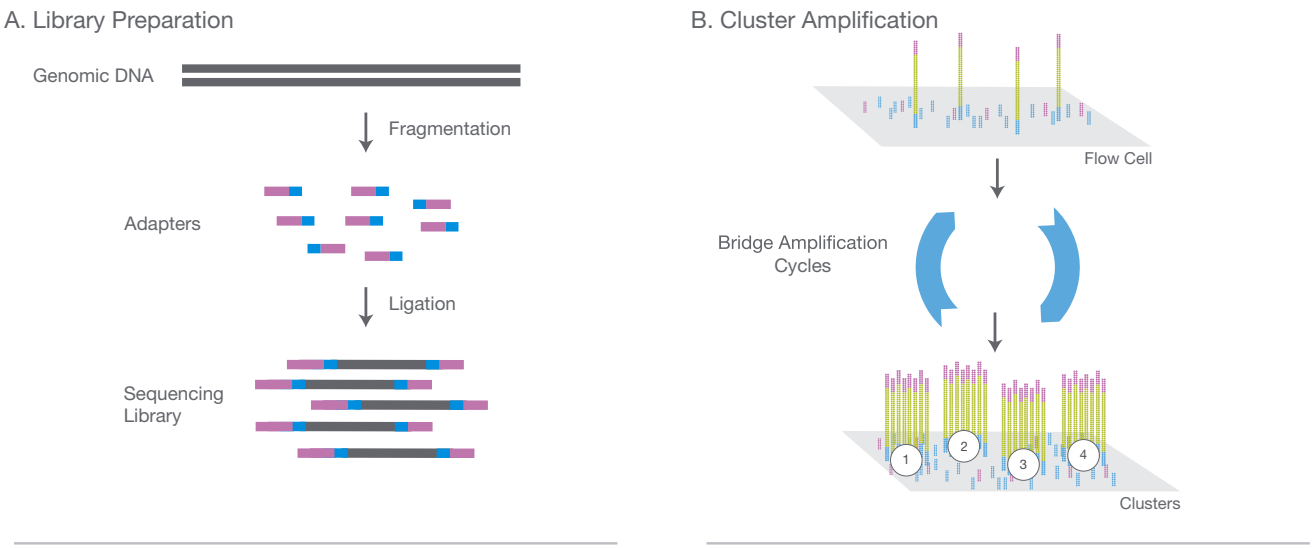
NGS is a high-throughput process for determining DNA sequence. Illumina sequencing by synthesis (SBS) chemistry uses a DNA polymerase to catalyze the incorporation of fluorescently labeled nucleotides into a DNA template strand during sequential cycles of DNA synthesis. During each cycle, the nucleotides are identified by fluorophore labels. This method is highly scalable and can be applied to a subset of key genes or the entire genome.

The Basic NGS Workflow

All Illumina, NGS workflows include 4 basic steps.

1. **Library Preparation**. Sequencing library preparation begins by creating short DNA or cDNA fragments with 5' and 3' adapters ligated (Figure 2A).
2. **Cluster Amplification**. For 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 2B).
3. **Sequencing**. Next, the templates are ready for sequencing. SBS technology uses a proprietary reversible terminator-based method that detects single bases as they are incorporated into DNA template strands.[12] 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.[13,14] The result is highly accurate base-by-base sequencing that virtually eliminates sequence context-specific errors, even within repetitive regions and homopolymers (Figure 2C).
4. **Alignment and Data Analysis**. Newly identified sequence reads are exported to an output file and aligned to a reference genome by sequencing alignment software (Figure 2D).

For a detailed animation of SBS technology, visit www.illumina.com/SBSvideo.



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

Multiplexing

Over time, NGS has increased data output per run, while also increasing sample throughput. Multiplexing allows large numbers or batches of libraries to be pooled and sequenced simultaneously during a single sequencing run (Figure 3). With multiplexing, 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. This dramatically reduces the time-to-data for multisample studies and enables researchers to go from experiment to answer faster and easier than ever before.

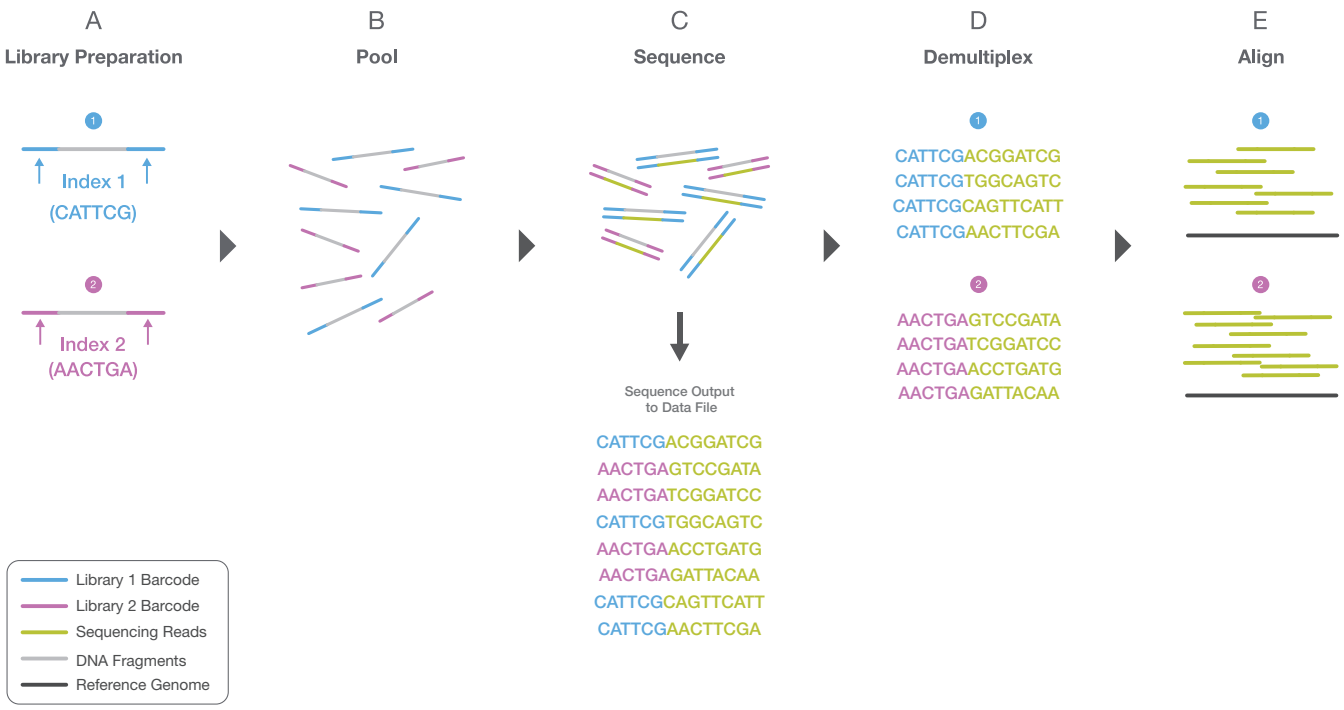


Figure 3: Library Multiplexing Overview—A. Distinct samples are attached to unique index sequences. Index sequences are attached during library preparation. B. Libraries for each sample are pooled together and loaded into the same flow cell lane. C. Samples are sequenced together during a single instrument 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.

IV. Summary

Over the last decade, advances in genomic technologies have led to improved approaches to characterizing genetic abnormalities. In turn, improved analyses have increased understanding of genetic influences on many diseases. With more thorough, accurate methods for detection of chromosomal aberrations, genomic technology offers new possibilities for identifying and monitoring diseases in the future.

Illumina is committed to advancing molecular testing tools and collaborating with industry leaders to transform health care. Together, we aim to bring the promise of genomics toward widespread clinical adoption and improvements in patient diagnosis, treatment, and outcomes.

V. Glossary

Adapters: Oligos bound to the 5' and 3' ends of each DNA fragment in a sequencing library. The adapters are complementary to the lawn of oligos present on the surface of Illumina sequencing flow cells.

Bridge amplification: An amplification reaction that occurs on the surface of an Illumina flow cell. During flow cell manufacturing, the surface is coated with a lawn of 2 distinct oligonucleotides referred to as “p5” and “p7.” In the first step of bridge amplification, a single-stranded sequencing library is injected into the flow cell. Individual molecules in the library bind to complementary oligos as they “flow” across the oligo lawn. Priming occurs as the free end of a ligated fragment bends over and “bridges” to another complementary oligo on the surface. Repeated denaturation and extension cycles (similar to PCR) result in localized amplification of single molecules into millions of unique, clonal clusters across the flow cell.

Clusters: A clonal grouping of template DNA bound to the surface of a flow cell. Each DNA template strand that binds to the flow cell acts as a seed 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, 10,000 clusters on the flow cell would produce 10,000 single reads.

Flow cell: A glass slide coated with a lawn of surface-bound, adapter-complimentary oligos. A pool of 8–384 multiplexed libraries can be sequenced simultaneously, depending on application parameters.

Indexes/Barcodes/Tags: A unique DNA sequence ligated to fragments within a sequencing library for downstream *in silico* sorting and identification. Libraries with unique indexes can be pooled together, loaded into a lane of a sequencing flow cell, and sequenced in the same run. Reads are later identified and sorted via software.

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.

Read: In general terms, a sequence “read” refers to the data string of “A, T, C, and G” bases corresponding to the sample DNA. With Illumina technology, millions of reads are generated in a single sequencing run. In specific terms, each cluster on the flow cell produces a single sequencing read.

VI. References

1. Morozova O, Marra MA. From cytogenetics to next-generation sequencing technologies: advances in the detection of genome rearrangements in tumors. *Biochem Cell Biol.* 2008;86:81-91.
2. Halder A, Jain M. Cytogenetics to cytogenomics: transition from chromosome to DNA sequence. *Global Journal of Human Genetics & Gene Therapy.* 2014;2:53-67.
3. Petrone J (2010) ACMG Recommends Replacing Karyotyping with Chromosomal Microarrays as 'First-Line' Postnatal Test. Genomeweb (www.genomeweb.com/arrays/acmg-recommends-replacing-karyotyping-chromosomal-microarrays-first-line-postnat) Accessed 10 October 2015.
4. Jackson EM, Sievert AJ, Gai X, et al. Genomic analysis using high-density single nucleotide polymorphisms-based oligonucleotide arrays and multiplex ligation-dependent probe amplification provides a comprehensive analysis of *INI1/SMARCB1* in malignant rhabdoid tumors. *Clin Cancer Res.* 2009;15:1923-1930.
5. Ho CY, Mosier S, Safneck J, et al. Genetic profiling by single-nucleotide polymorphism-based array analysis defines three distinct subtypes of orbital meningioma. *Brain Pathol.* 2015;25:193-201.
6. Harada S, Henderson LB, Eshleman JR, et al. Genomic changes in gliomas detected using single nucleotide polymorphism array in formalin-fixed, paraffin-embedded tissue: superior results compared with microsatellite analysis. *J Mol Diagn.* 2011;13:541-548.
7. Illumina (2015) Microarray-based cytogenetic testing illuminates congenital abnormalities and cancer. (www.illumina.com/content/dam/illumina-marketing/documents/icommunity/article_2015_01_combimatrix_cytosnp850.pdf)
8. Illumina (2010) Cytogenetics into Cytogenomics: SNP Arrays Expand the Screening Capabilities of Genetics Laboratories. (www.illumina.com/documents/products/appnotes/appnote_cytogenetics.pdf)
9. Illumina (2012) Molecular Cytogenetics. (www.illumina.com/documents/products/datasheets/datasheet_cytogenetics.pdf)
10. Illumina (2015) From arrays to NGS: how chromosomal genetics evolved from structure to disease. (www.illumina.com/content/dam/illumina-marketing/documents/icommunity/koehler-mgz-cyto-article-1570-2015-008.pdf)
11. Liang D, Peng Y, Lv W, et al. Copy Number Variation Sequencing for Comprehensive Diagnosis of Chromosome Disease Syndromes. *J Mol Diagn*. 2014;16(5): 519-526.
12. Bentley DR, Balasubramanian S, Swerdlow HP, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature.* 2008;456(7218):53-59.
13. Ross MG, Russ C, Costello M, et al. Characterizing and measuring bias in sequence data. *Gen Biol.* 2013;14:R51.
14. CARDIoGRAMplusC4D Consortium. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet.* 2012;45(1):25-33.

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.

 Pub. No. 1170-2015-007 Current as of 13 March 2016

illumina®