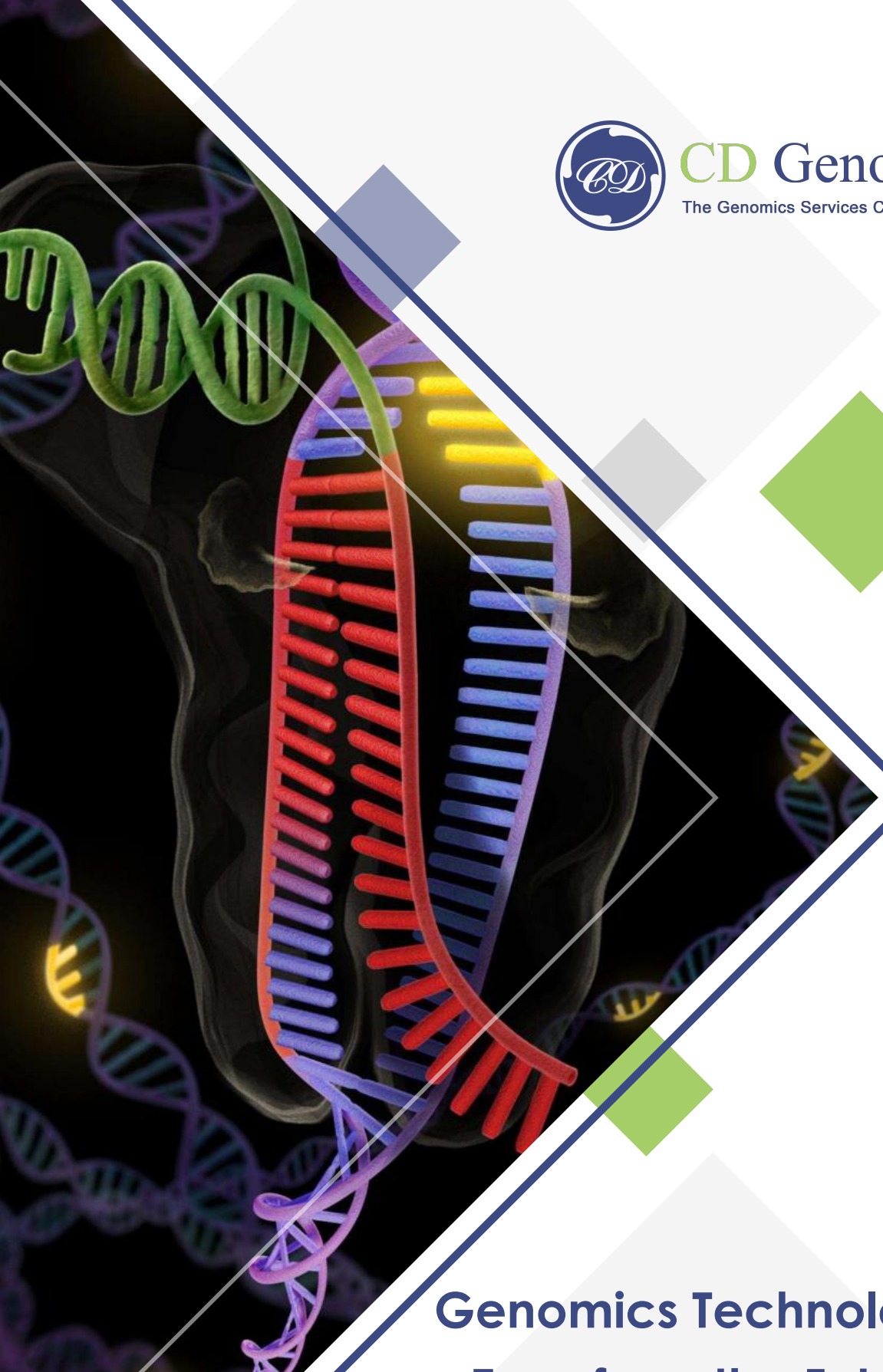

CD Genomics
The Genomics Services Company

Genomics Technologies Transform the Future of Biomarker Discovery

Clinical Implications of Biomarkers

Biomarkers can be classified into two major groups, disease-related and drug-related biomarkers. Disease-related biomarkers are very helpful for clinical stratification, identification of patients within the general population, disease staging, and evaluation of likely outcome of disease. In contrast, drug-related biomarkers are biological or clinical characteristics that provide valuable information on prognosis and survival after therapeutic intervention. Such predictive factors can be employed to identify subgroups of patients who are most likely to respond to a given treatment, to create a treatment plan, and to assess the efficacy, adverse reactions, and complications of therapy. An ideal biomarker should be easily, specifically, stably, and inexpensively measured in a noninvasive or minimally invasive way with high analytical specificity and sensitivity.

The discovery of biomarkers has become increasingly vital in clinical science and practice in the era of targeted therapy. Discovery of biological parameters that may contribute to the differential response of disease to a given treatment allows developing an individualized therapy. Individualized therapy, also known as personalized medicine, refers to the prediction of the response to a given therapy and followed by a tailored treatment intervention based on the individual biological and genetic information.

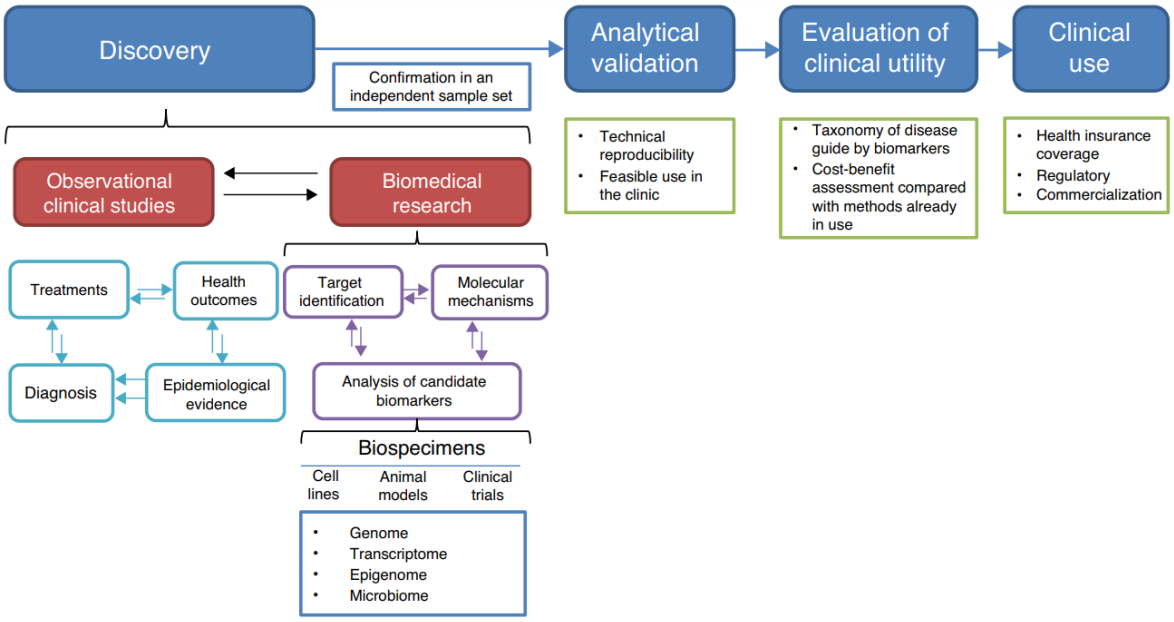

Figure 1. Genomics technologies accelerate biomarker discovery and development (adapted from Quezada *et al*. 2017).

CD Genomics
The Genomics Services Company

Genomic Technologies for Biomarker Discovery

The human genome sequencing was completed in 2001, revealing a complex genome with many variations. With the rapid development of genomic technologies, comprehensive analyses of biologic interaction and clinical significance of these genetic aberrations become feasible, and these data enable researchers to discover biomarkers. Next-generation sequencing (NGS) provides exquisite sensitivity and resolution in the discovery of DNA, RNA, and epigenomic biomarkers. Long-read sequencing, including PacBio SMRT sequencing and Oxford nanopore sequencing, is remarked by low cost, long reads, no PCR bias, high resolution in repetitive sequence, direct detection of base modifications, direct RNA sequencing and so on. Microarray-based technologies provide a cost-effective and rapid solution for biomarker discovery.

Table 1. Summary of Genomics technologies for biomarker discovery.

Genomic Measurement	Technology	Platform
DNA sequence	Sequencing	NGS, PacBio, Oxford nanopore
Gene expression	RNA-seq	NGS, PacBio, Oxford nanopore
	Gene expression microarray	Affymetrix GeneChip Agilent gene expression microarray Illumina BeadChip
DNA binding	ChIP-chip	Roche NibleGen Agilent ChIP-on-chip microarray
	ChIP-seq	NGS, PacBio, Oxford nanopore
Methylation	Methylation array	Illumina BeadChip Agilent methylation microarray
	Bisulfite sequencing	NGS
	Binding assays (MBDSeq, MeDIP)	NGS or microarray
	Direct sequencing	PacBio, Oxford nanopore
SNP	Sequencing	NGS, PacBio, Oxford nanopore
	SNP-chip	Affymetrix SNP array Illumina SNP array Agilent SNP Microarray
miRNAs and other sncRNAs	Sequencing	NGS, PacBio, Oxford nanopore
	Microarray	Affymetrix GeneChip Agilent miRNA microarray Illumina BeadChip

Workflow of Sequencing Data Analyses

Assessment of Quality

First of all, the quality of NGS reads is evaluated and poor-quality reads are removed or trimmed. Base-calling errors are also assessed in this step with tools like FASTQC.

Sequence Alignment

Qualified reads are then aligned to the reference genome or *de novo*. For NGS, paired-end reads are a better option. The combination of data from NGS and long-read sequencing generates an alignment with higher accuracy. Mutations originated from reads with many mismatches should be discarded.

Variant Calling

There are many types of variants, germline variants, somatic variants, copy number variations (CNVs), and structural variants (SVs). In cancer, somatic mutations are commonly targeted for detection while in rare diseases, germline mutations are usually focused. There are two major types of SVs: balanced and unbalanced SVs. Balanced SVs (like inversion, same chromosomal translocation, different chromosomal translocation) do not change the content of DNA while unbalanced SVs (like duplication, deletion) change the content of DNA.

Table 2. Tools for variant identification (Wadapurkar & Vyas 2018).

Tools	Input files	Output files	Identifies
Germline caller tools			
Galaxy platform	BAM/SAM	VCF, VarScan CSV	SNP, INDEL
SanGeniX platform*	BAM/SAM	VCF, VarScan CSV	SNP, INDEL
VarScan2	pileup/mpileup	VCF, VarScan CSV	SNP, INDEL
SNVer	BAM/SAM	VCF	SNP, INDEL
CRISP	BAM/SAM	VCF	SNP, INDEL
GATK(Unified Genotyper)	BAM/SAM	VCF	SNP, INDEL
SAMtools	BAM/SAM, FASTA	VCF	SNP, INDEL
Somatic callers tools			
Galaxy platform	BAM/SAM	VCF, VarScan CSV	SNP, INDEL
SanGeniX platform*	BAM/SAM	VCF, VarScan CSV	SNP, INDEL
VarScan2	pileup/mpileup	VCF, VarScan CSV	INDEL, SNP,CNV
GATK (Somatic Indel Detector)	BAM/SAM	VCF	INDEL
SAM tools	BAM/SAM, FASTA	BCF	SNP, INDEL
CNV identification tools			
ExomeCNV	BAM/SAM, pileup + bed + FASTA	CSV	CNV, LOH
CNVnator	BAM/SAM, FASTA	CSV	CNV
CONTRA	BAM/SAM, FASTA	VCF, CSV	CNV
RDXplorer	BAM/SAM, FASTA	CSV	CNV
SV identification tools			
GASVPro (GASVPro-HQ)	BAM/SAM	clusters file	INDEL, INV, TRANS
CLEVER	BAM/SAM, FASTA	CLEVER format	INDEL
BreakDancer	BAM/SAM, config file	BED, CSV	INDEL, CNV, INV, TRANS
Breakpointer	BAM/SAM	GFF	INDEL
SVMerge	BAM/SAM, FASTA	BED	INDEL, INV, CNV

* Proprietary software.

Variant Annotation

Annotation of variants can be achieved via comparison with databases such as dbSNP etc. Many computational annotation tools provide links to databases, such as Galaxy platform, VARIANT, snpEff, VEP, ANNOVA-R, AnnTools, SeattleSeq, SVA, NGS-SPN.

Gene Expression Regulation

In addition to genetic mutations, noncoding RNAs (ncRNAs), methylation, and modified histones are also closely associated with diseases via the regulation of gene expression to serve as potential biomarkers. Therefore, it is very important to identify differentially expressed microRNAs (miRNAs) and differential modification patterns.

Pathway Analysis

Pathway analysis has recently been highlighted in the discovery of cancer biomarkers. A number of databases are available for signaling and metabolic pathways including Kyoto Encyclopedia of Genes and Genomes (KEGG), GeneGO by MetaCore, Molecular Signatures Database (MSigDB), IngenuityPathway Analysis (IPA) and Gene Set Enrichment Analysis (GSEA).

Our solutions

Genomic technologies have now been extensively applied to the biomarker development process. It is a versatile tool to identify genomic variants, differentially expressed genes, and differential methylation patterns, as well as investigate pathway and gene regulation network. CD Genomics is dedicated to offering a wide range of genomic solutions and bioinformatics analysis for biomarker discovery. We are equipped with advanced Illumina, PacBio, and Oxford nanopore sequencing instruments.

- Genomics
- Transcriptomics
- Metagenomics
- Epigenomics
- Microarray

Contact CD Genomics for more inspiration and service content.