VNIVERSITAT DE VALÈNCIA

Advanced clinical diagnostic techniques: omics and their application to the molecular study of diseases

BIOCHEMICAL INTEGRATION AND
CLINICAL BIOCHEMISTRY
2ND YEAR - DEGREE IN MEDICINE 2022-23

SEMINAR 4

PROFESSOR:
Amparo Galán Albiñana
amparo.galan@uv.es

Department of
Biochemistry and Molecular Biology

Analogue → Digital

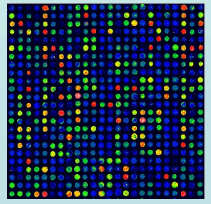
Paradigm shift in biomedical sciences

High performance analytical methods

Diagnosis based on individual parameters

Diagnosis based on global parameters

GLOBAL → omics



Bioinformatics

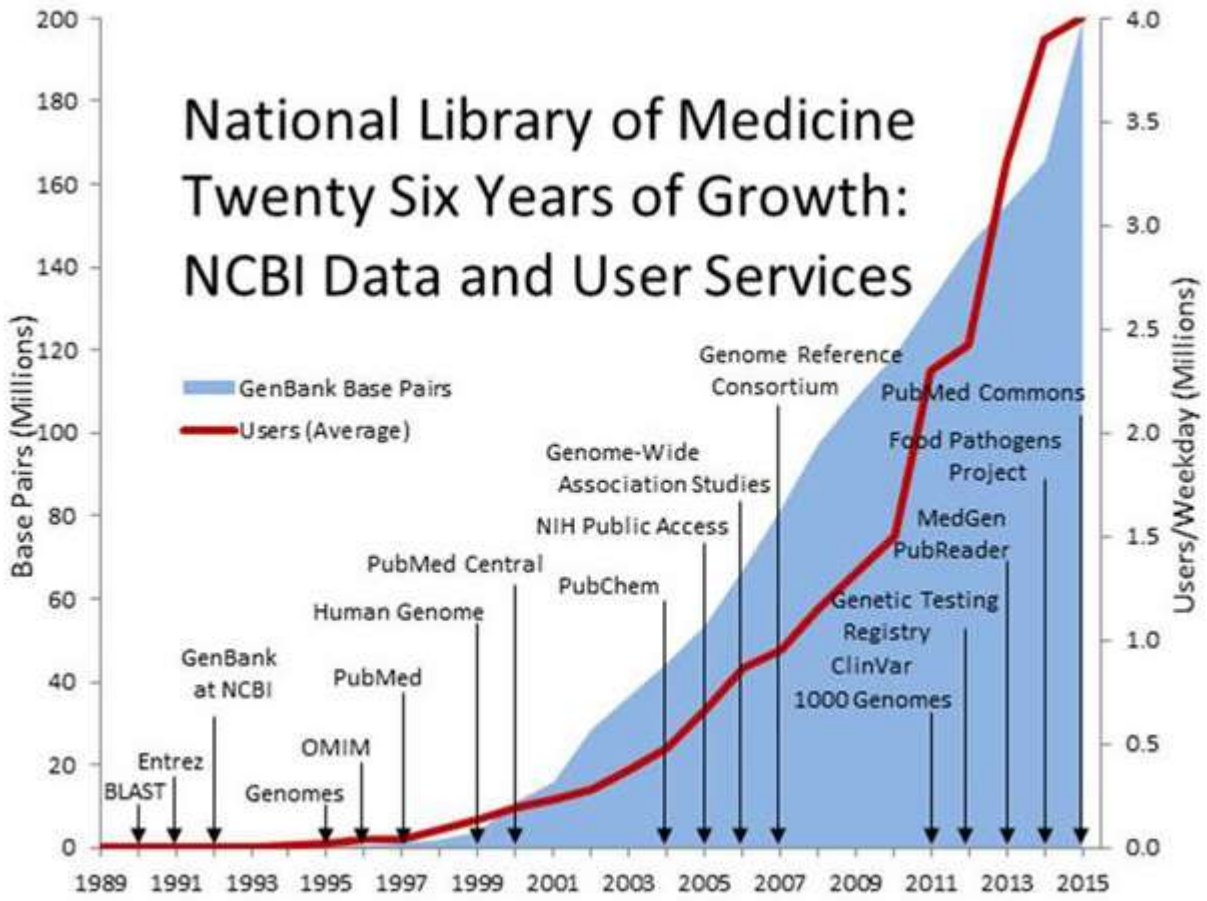
INTRODUCTION: OMICS

- Omics investigate organisms as a whole, from a holistic perspective.
- Omics are linked to the development of new high-performance technologies that arise from the collaboration across multiple disciplines (biology, computer science, chemistry, physics, engineering, etc.).
- Omics generate millions of biological data that require complex processing and analysis.
- The massive flow of data generated by the omics sciences has promoted the emergence and development of bioinformatics, which is defined as the discipline that applies computer science and statistics to the processing and analysis of biological data.

OMICS SCIENCES

Genes	→	Genomics
mRNAs	→	Transcriptomics
Proteins	→	Proteomics
Metabolites	→	Metabolomics
Cells	→	Cytomics

Explosion in the number of DNA sequences stored in public databases



NCBI

Update of the NCBI's Entrez databases repository

Nucleic Acids Research, 2021, Vol. 49, Database issue

Table 1. The Entrez Databases (as of 9 September 2020)

Database	Records	Description
Literature		
PubMed	31 471 600	scientific and medical abstracts/citations
PubMed Central	6 447 271	full-text journal articles
NLM catalog	1 619 856	index of NLM collections
Books	825 385	books and reports
MeSH	300 500	ontology used for PubMed indexing
Genomes		
Nucleotide	429 731 711	DNA and RNA sequences
BioSample	14 628 076	descriptions of biological source materials
SRA	11 807 161	high-throughput DNA and RNA sequence read archive
Taxonomy	2 401 136	taxonomic classification and nomenclature catalog
Assembly	837 406	genome assembly information
BioProject	458 893	biological projects providing data to NCBI
Genome	55 580	genome sequencing projects by organism
BioCollections	8 138	museum, herbaria and other biorepository collections
Genes		
GEO Profiles	128 414 055	gene expression and molecular abundance profiles
Gene	28 377 759	collected information about gene loci
GEO datasets	4 002 373	functional genomics studies
PopSet	350 627	sequence sets from phylogenetic and population studies
HomoloGene	141 268	homologous gene sets for selected organisms
Genetics		
SNP	720 643 623	short genetic variations
dbVar	6 030 887	genome structural variation studies
ClinVar	845 008	human variations of clinical significance
MedGen	335 277	medical genetics literature and links
GTR	76 814	genetic testing registry
dbGaP	1 397	genotype/phenotype interaction studies
Proteins		
Protein	874 272 642	protein sequences
Identical protein groups	329 946 078	protein sequences grouped by identity
Protein clusters	1 137 329	sequence similarity-based protein clusters
Structure	167 650	experimentally-determined biomolecular structures
Sparcle	149 462	conserved domain architectures
Conserved domains	59 951	conserved protein domains
Chemicals		
PubChem substance	285 048 146	deposited substance and chemical information
PubChem compound	111 325 418	chemical information with structures, information and links
PubChem BioAssay	1 229 071	bioactivity screening studies
BioSystems	983 968	molecular pathways with links to genes, proteins and chemicals

VNIVERSITAT DE VALÈNCIA

CYTOMICS

Department of
Biochemistry and Molecular Biology

FLOW CYTOMETRY

Analytical method for measuring the emission of multiple fluorescence and light scattering from cells or microscopic particles, aligned by a laminar liquid stream, when presented one at a time and at high speed in front of a light source of appropriate wavelength.

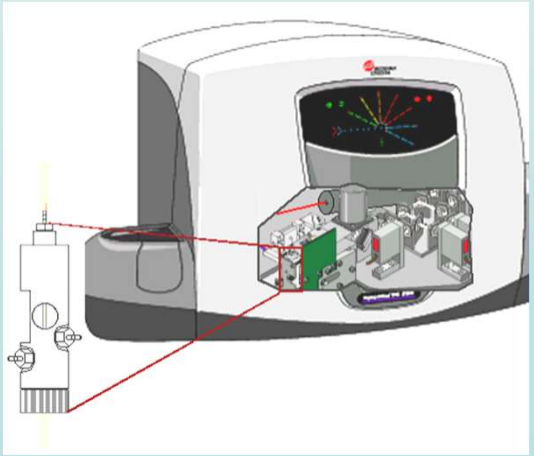
- **Fluorescence and fluorescent markers**
- **Parameters that can be analyzed by flow cytometry**
  - **What can be stained?**
  - **Which cells can be analyzed?**
- **Applications of flow cytometry**

FLOW CYTOMETRY

Individual biological cells or particles in suspension

Analytical method that measures the simultaneous emission of multiple fluorescence and light scattering

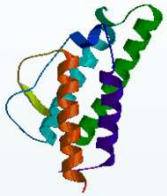
FLOW CYTOMETRY

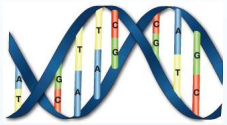


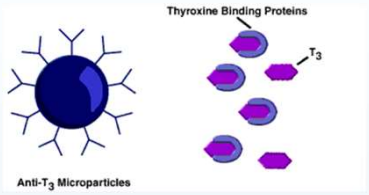
Sequentially aligned by a laminar flow and presented one at a time at high speed at an optimal point of illumination

FLOW CYTOMETRY

Molecular level

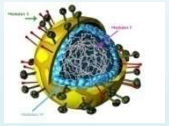


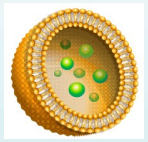


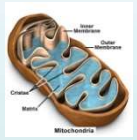


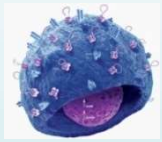
- Extracellular proteins
- Free DNA/RNA sequences
- Circulating immune complexes

Subcellular level



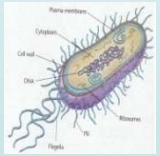


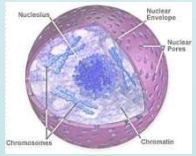


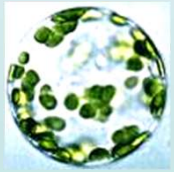


- Individual virions
- Liposomes
- Isolated chromosomes
- Isolated organelles
- Isolated nuclei

Cell level

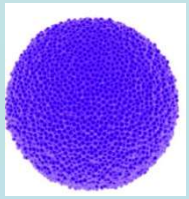


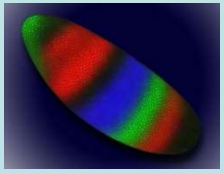




- Bacteria
- Unicellular fungi
- Human and animal cells

Supracellular level





- Hybridomas and cell fusions
- Spheroids
- Embryoid bodies
- Larvae and embryos
- Pluricellular organisms

DIAGNOSTIC APPLICATION

Cells

↓

Cytometer

↙ ↘

Numerical information (analyzer)

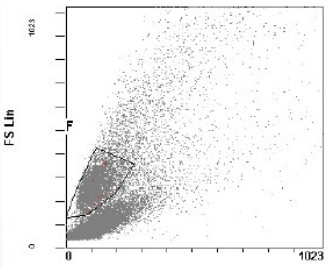
Purified cells (separator)

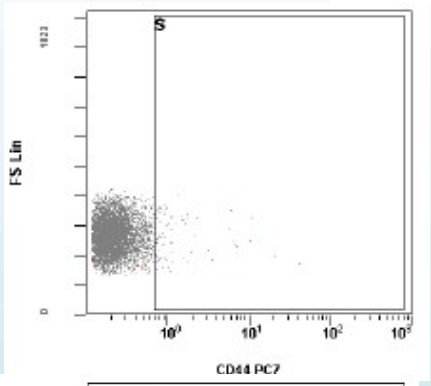
↓

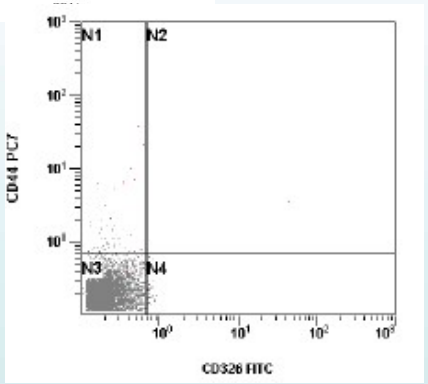
↓

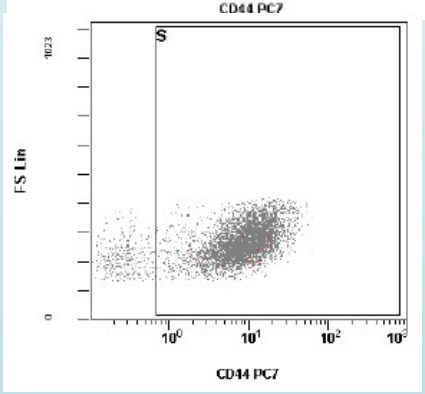
Therapeutic/ diagnostic decision-making

Cell transfusion

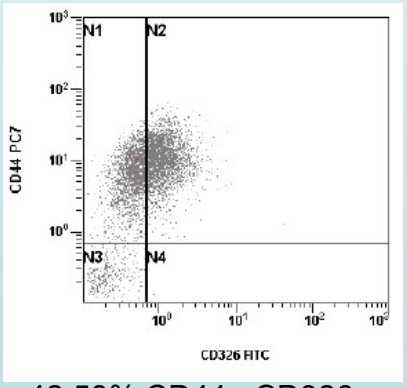








93.73% CD44+

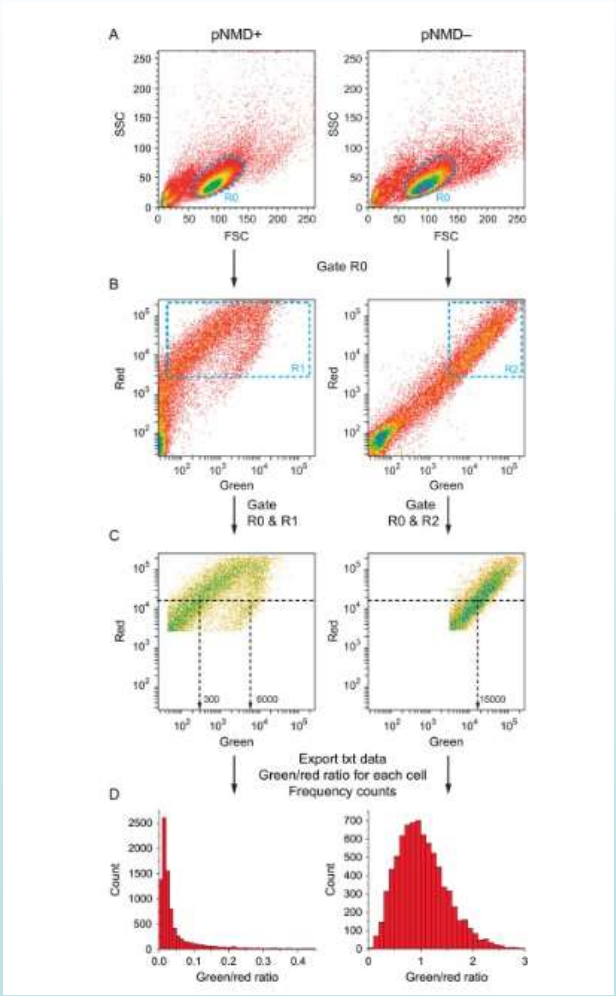


42.53% CD44+ CD326-
51.89% CD44+ CD326+
6.22% CD44- CD326-
0.06% CD44- CD326+

DIAGNOSTIC APPLICATION

Main applications: oncology and hematology

- **Diagnosis and prognosis of leukemias, lymphomas, lymphoproliferative syndromes and myelodysplastic syndrome**
- **Absolute CD34⁺ cell count for auto- or allotransplantation of hematopoietic stem cells**
- **Identification of proliferating cells and tumor cells in blood, biological fluids and complex samples**
- **Study of cell cycle regulation**



METABOLOMICS

Department of
Biochemistry and Molecular Biology

METABOLOMICS

Metabolomics is the study of metabolites.

Metabolites are the intermediates and products of metabolism. They are any molecule of less than 1 KDa (except some macromolecules: lipoproteins, albumin, etc.).

Unique footprints left by specific cellular processes.

Endogenous metabolites: produced by the organism.

Exogenous metabolites: from external substances (drugs, poisoning, etc.).

Also known as xenometabolites.

The metabolome represents the collection of all the metabolites in a cell, tissue, organ, or organism that are the product of cellular processes.

Biological fluids: mostly urine and plasma (non-/minimally invasive).

But they also include: tissue biopsies, cells, exudates, saliva, bile salts, intestinal aspirates, and fluids (cerebrospinal, seminal, amniotic, and synovial).

Metabolites have been roughly categorized as:

2,500 endogenous metabolites

1,200 drugs

3,500 food components

Human metabolome data: www.hmdb.ca

ROUTINE DIAGNOSTIC APPLICATIONS

BIOCHEMICAL AUTOANALYZERS

Samples: urine and plasma

- State-of-the-art analyzers for analyzing hundreds of metabolites in a few minutes
- Based on spectrophotometry
- Non-complex techniques used
- Quick and precise

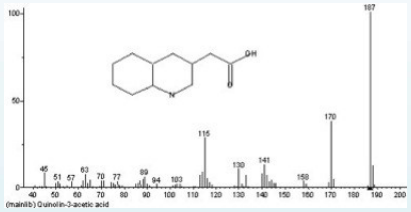
Particle detection of:

- Enzymes
- Ions:
  - Na^+/K^+
- Biochemical markers:
  - Glucose, cholesterol, triglycerides, uric acid, proteins, serum albumin, creatinine
- Immune assays by antibodies
- Hematologic assays
  - Erythrocyte sedimentation/coagulation

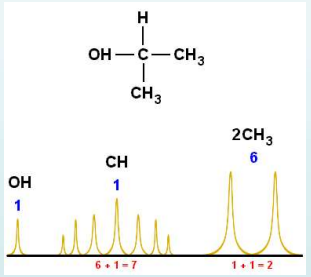
Routinely detect common metabolites and big molecules / cells

METHODS OF BIOMARKER DETECTION IN METABOLOMICS

Mass spectrometry (MS)



Nuclear magnetic resonance (NMR) spectrometry



Samples

Urine (by NMR): high metabolite concentration and low protein concentration

Blood/plasma (by MS and NMR): metabolic status of the organism

Whole tissue → **Lipid phase**
Whole tissue → **Aqueous phase**

In pathophysiological processes, metabolites increase *in situ.*

Low sample amounts (10 µg in tissue and 10 µl in fluids)

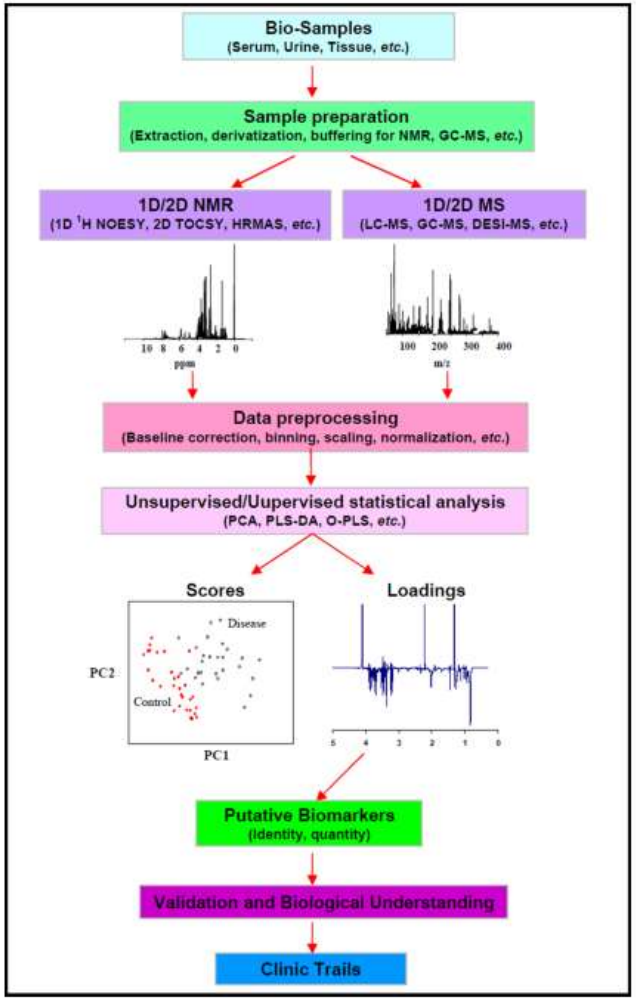
DIAGNOSTIC APPLICATION

Metabolite detection endogenous/exogenous origin

- Vitamin C (AA/DHAA)
- Vitamin B6 and derivatives
- Taurine
- Amino acids
- Hydroxymethylated DNA cytosines
- Purine and pyrimidine bases
- Lactulose/mannitol
- Prostaglandines
- Nucleosides/nucleotides
- Toxicology (alkaloids, benzodiazepines, amphetamines and derivatives)
- Eicosanoids (polyunsaturated fatty acids)
- Drugs of abuse
- Emerging drugs

DIAGNOSTIC APPLICATION

Biomarkers: value for early diagnosis of diseases (i.e. preclinical phase of trials with therapeutic possibilities).
Prognosis/treatment implication



Gowda *et al.*, Expert Rev Mol Diagn. 2008

Analysis

Targeted — Non-targeted

Diagnosis/prognosis:

- Congenital metabolic diseases
- Congenital non-metabolic diseases
- Diabetes
- Cancer
- Autism
- Cardiovascular diseases
- Food and drug poisoning
- Biomarker screening (novel or not)

VNIVERSITAT DE VALÈNCIA

MOLECULAR BIOLOGY TECHNIQUES

Department of
Biochemistry and Molecular Biology

MOLECULAR DIAGNOSTICS

VNIVERSITAT DE VALÈNCIA

Old and new molecular diagnostic techniques

One molecule: one disease

- Southern blot
- PCR
- DNA sequencing (Sanger)
- Cytogenetics

Set of markers that characterize a disease

- Gene expression microarrays
- Comparative genomics and epigenetic microarrays
- Next generation sequencing: NGS

PCR/qPCR

Polymerase chain reaction

It detects presence/expression of individual genes

PCR: endpoint PCR

Qualitative/semi-quantitative

qPCR: real-time PCR

- Quantitative, sensitive and specific
- Quick

PCR/qPCR

DIAGNOSTIC APPLICATION

Diagnosis of infectious diseases

Detection of DNA and/or cDNA from infectious organisms

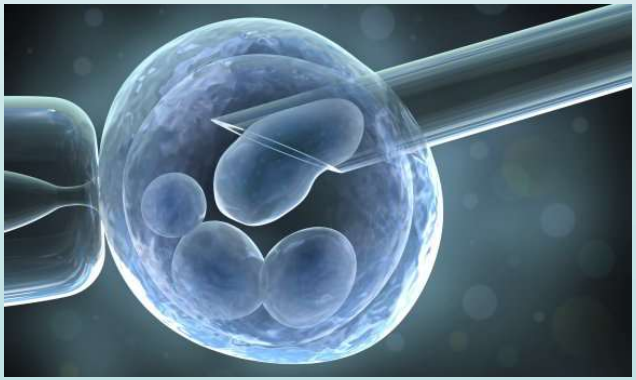
Fungi, bacteria, virus

Diagnosis of monogenic diseases

Monogenic diseases: failures or mutations in the information of a single gene that can be inherited or *de novo*.
Mutations are detected in patients by DNA analysis and confirmed by screening of family members.

Preimplantation genetic diagnosis

Performed on the embryo prior to its transfer to the uterus

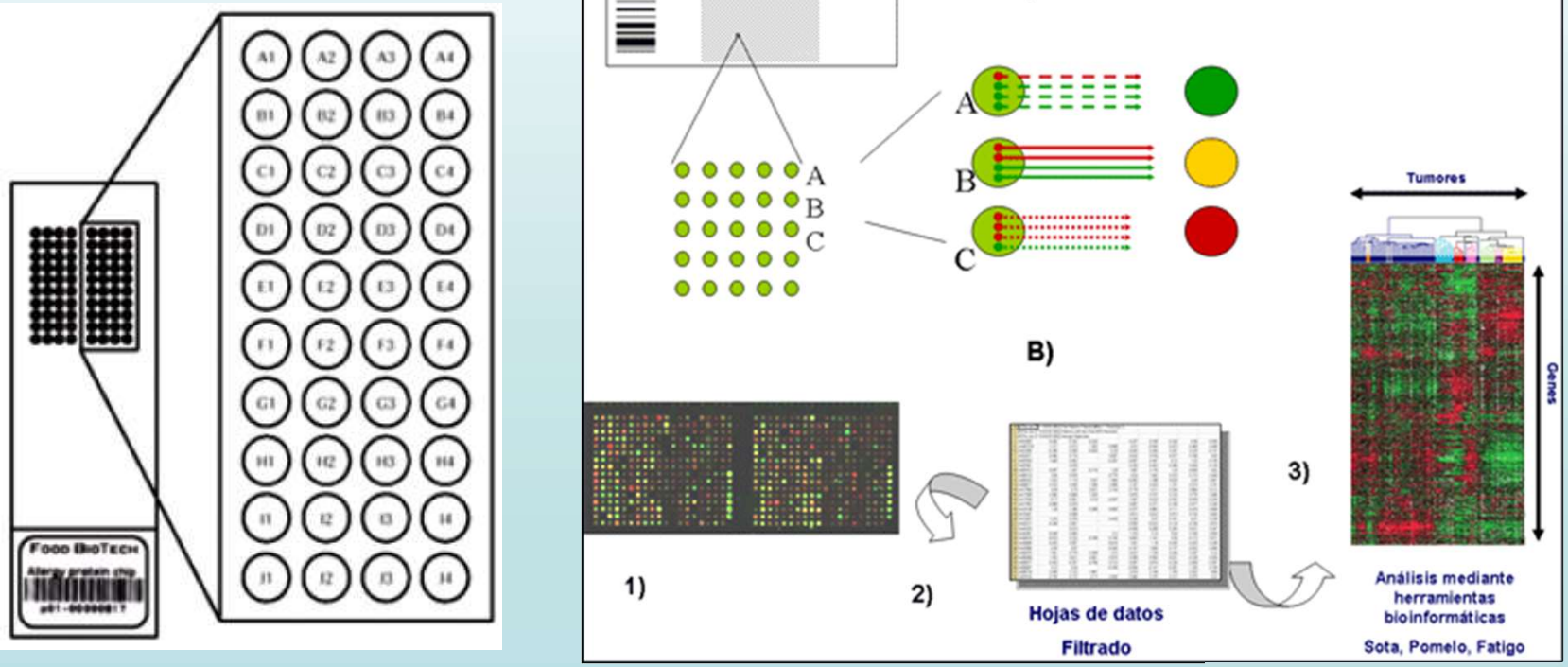


<u>Diseases:</u>

- **Autosomal dominant inheritance: single parent**
- **Autosomal recessive inheritance: two unaffected carrier parents**
- **Sex-linked inheritance: an affected father or a carrier mother**

MICROARRAYS

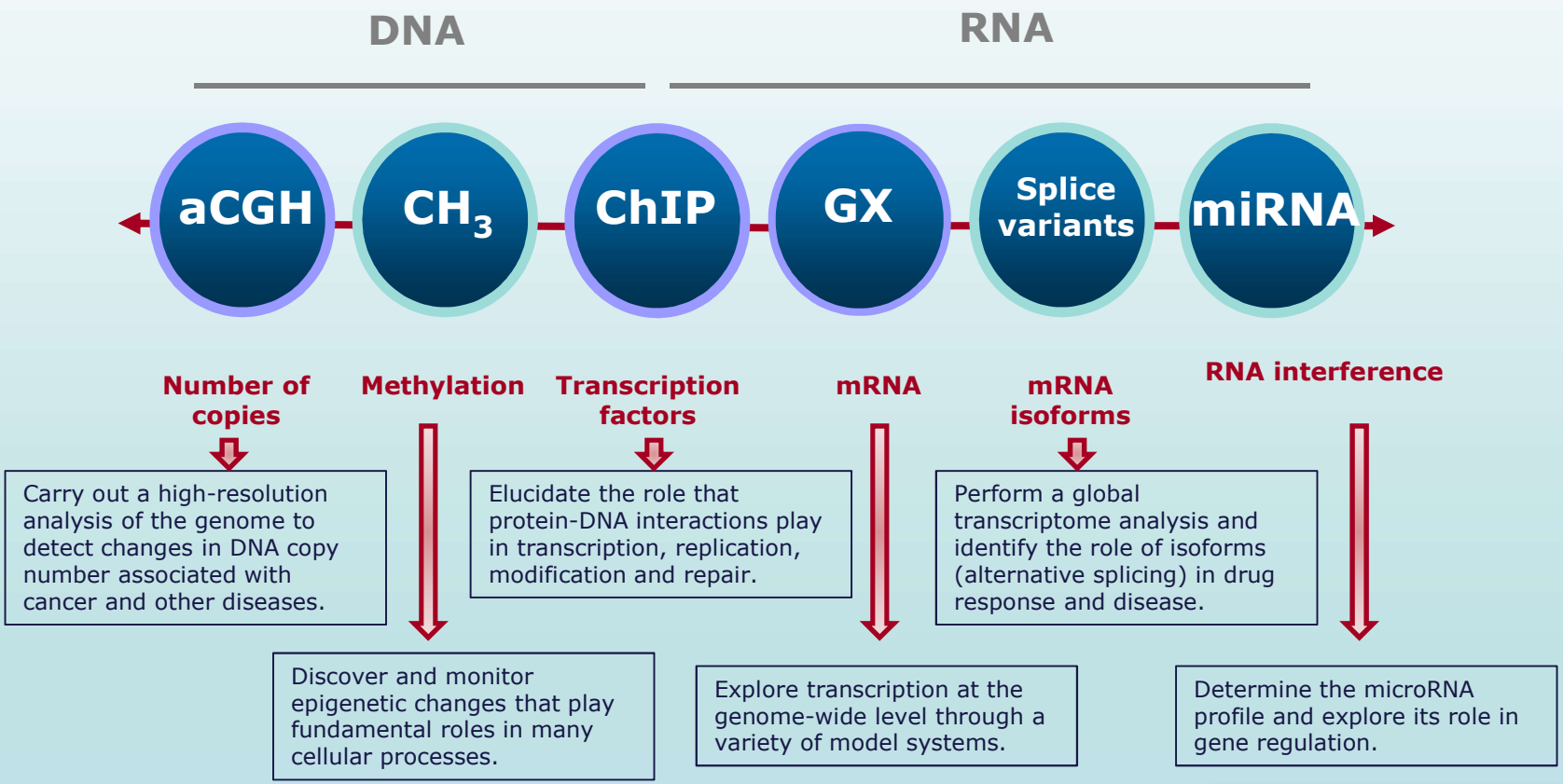
Array is the equivalent to matrix.
It is a solid surface containing a huge number of fixed probes (nucleic acids, proteins, metabolites, tissues…) that will be exposed to target molecules (sample).



MICROARRAYS

Massive analysis of gene sequences

Classification

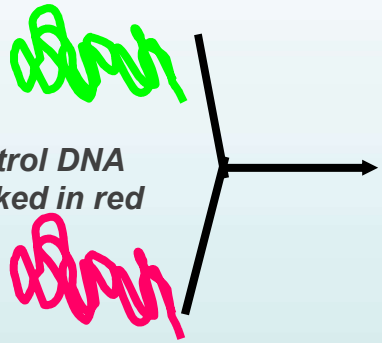


MICROARRAYS

CGH (comparative genomic hybridization)

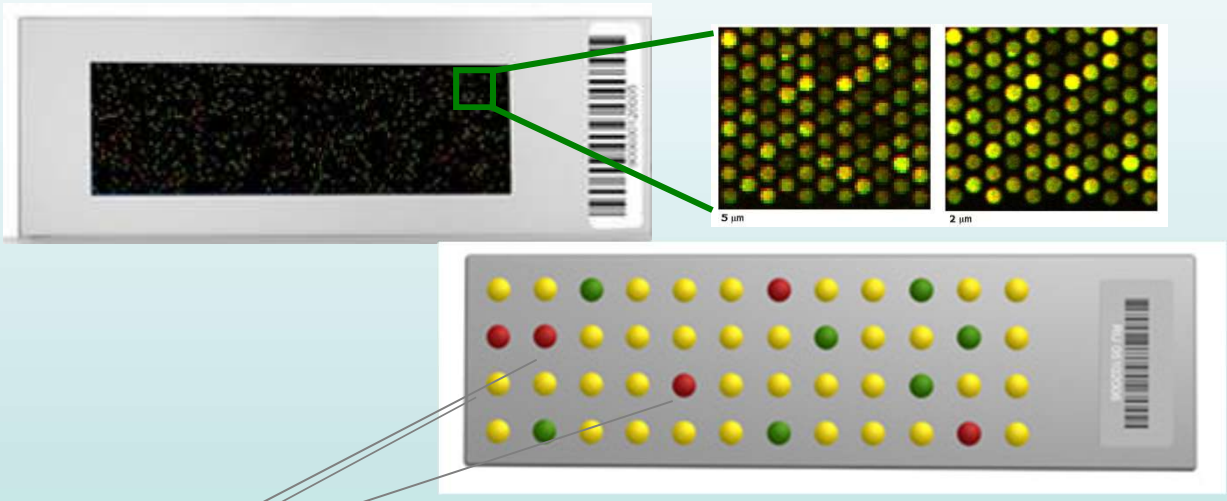
Patient DNA marked in green

Control DNA marked in red

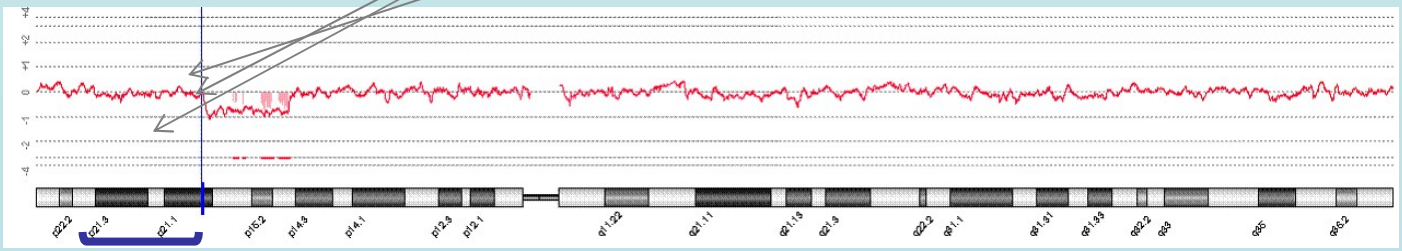


MICROARRAY

Each point corresponds to a known sequence location in the genome



Diagnosis of diseases associated with chromosomal abnormalities



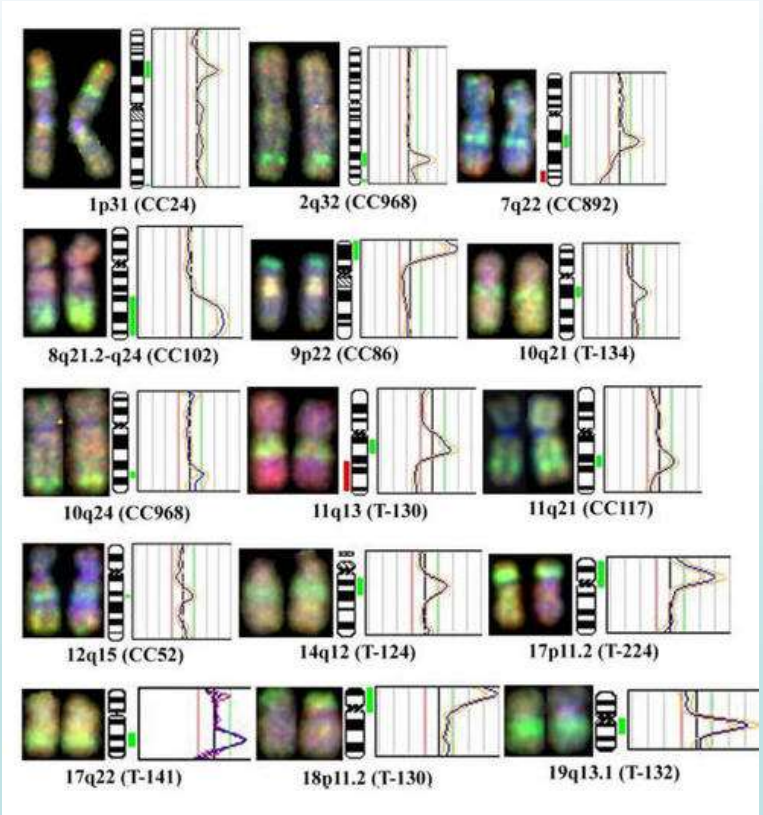
MICROARRAYS

CGH (comparative genomic hybridization)

DIAGNOSTIC APPLICATION

Cancer

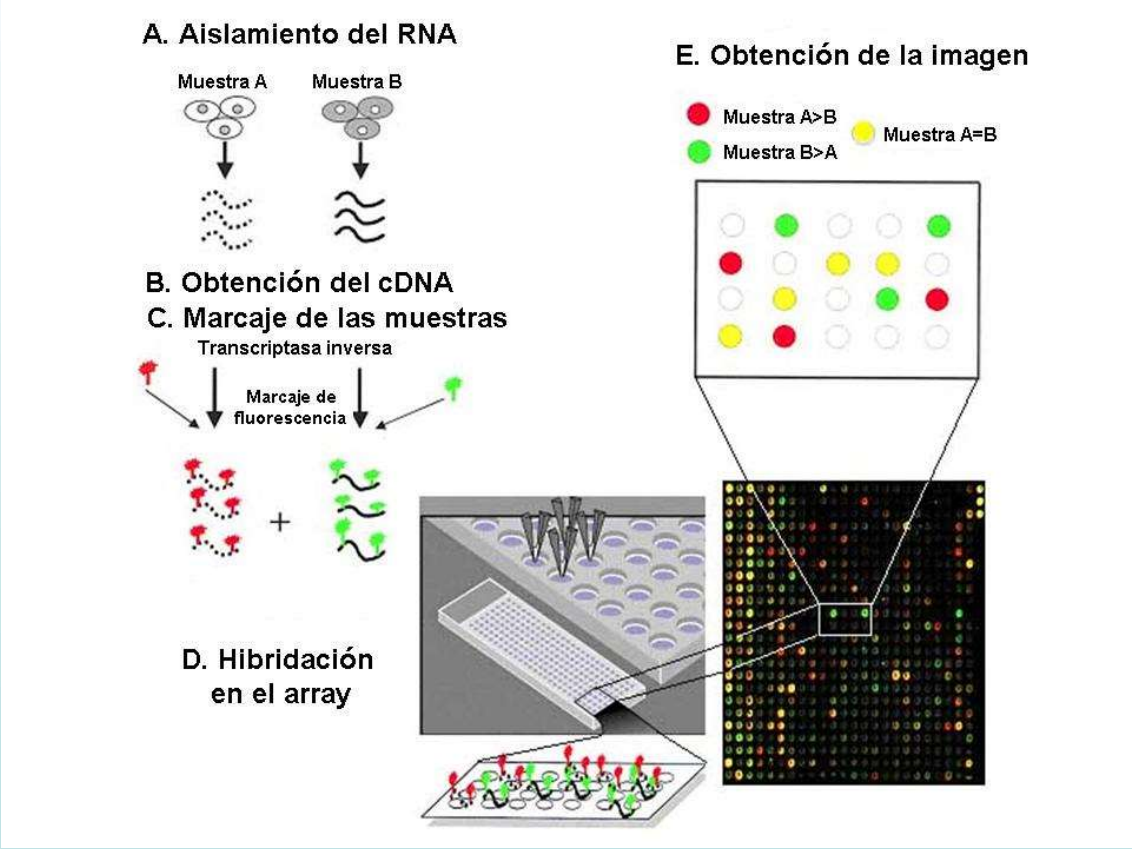
Preimplantation diagnosis



On chromosomes

MICROARRAYS

EXPRESSION MICROARRAYS



Comparative results

MICROARRAYS

EXPRESSION MICROARRAYS

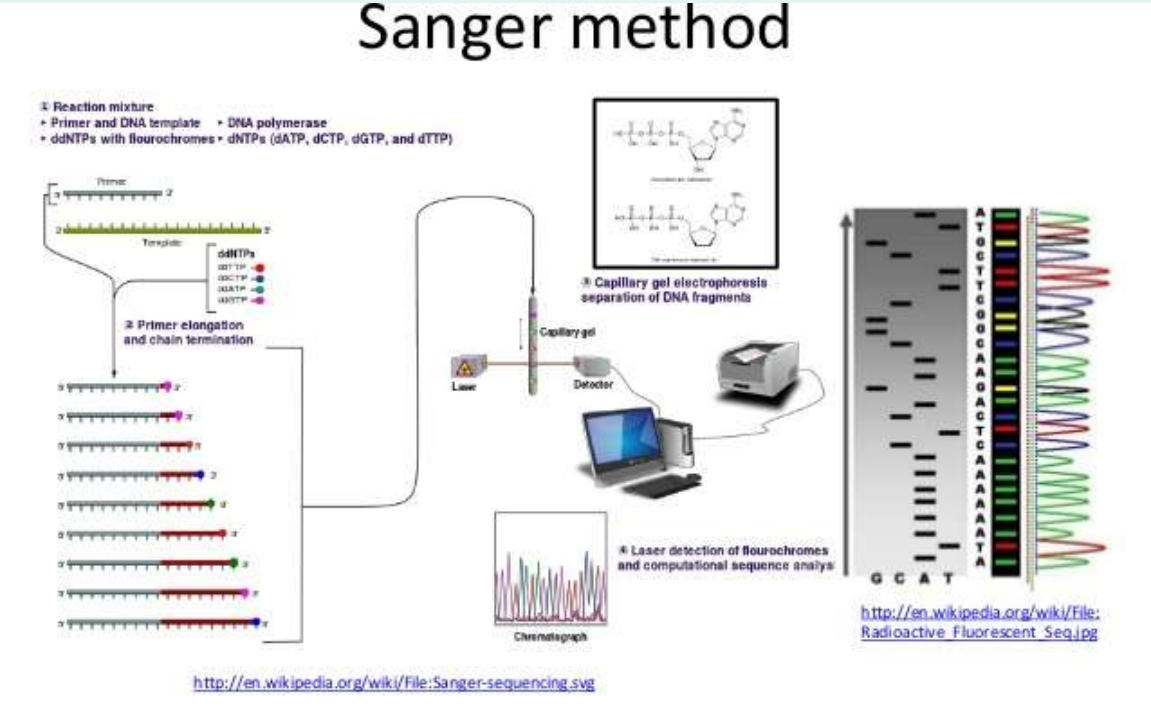
DIAGNOSTIC APPLICATION

- Detailed molecular phenotype
- Applications in oncology:
  - Tumor stage: molecular classification
  - Determination of treatment modalities
  - Identification of prognostic markers
- Variation of expression profile with nutrition → **Nutrigenomics**
- Variation of expression profile with exposure to toxic substances → **Toxicogenomics**
- **Personalized medicine:** patients can be treated with specific drugs based on their own expression signatures (patterns) → **Pharmacogenomics**

NUCLEIC ACID SEQUENCING

We have 3 billion ($3x10^9$) bases arranged in a unique order, with ~ 20,000 genes regulating gene transcription to protein synthesis.

Sanger sequencing is a well-established method.



Fragments: 500 pb – 1 Kb
Maximum: 96 reads

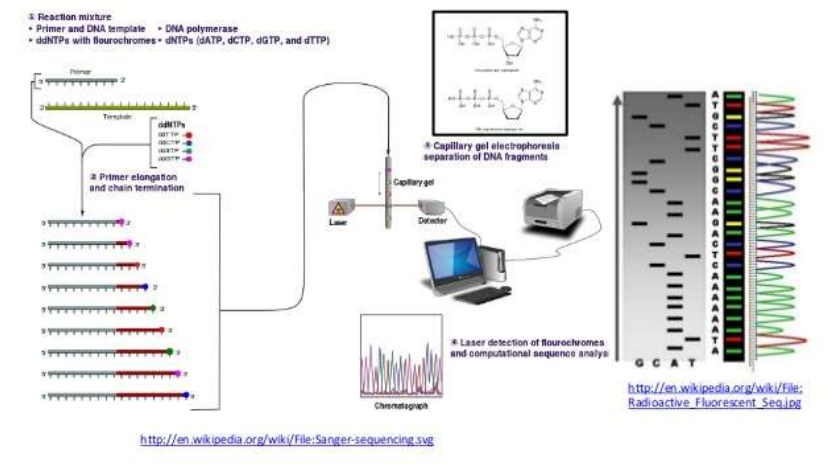
Time-consuming

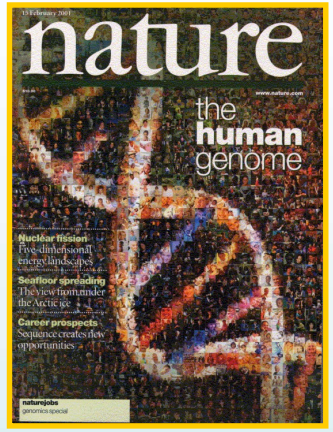
Expensive for long reads

MOLECULAR DIAGNOSTICS

The Human Genome Sequencing Project

Sanger sequencing method

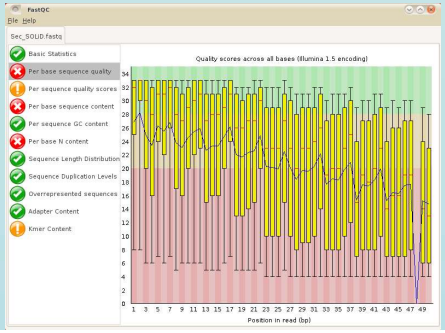




Sequencing of 25,000 genes

Next generation sequencing (NGS)

High throughput

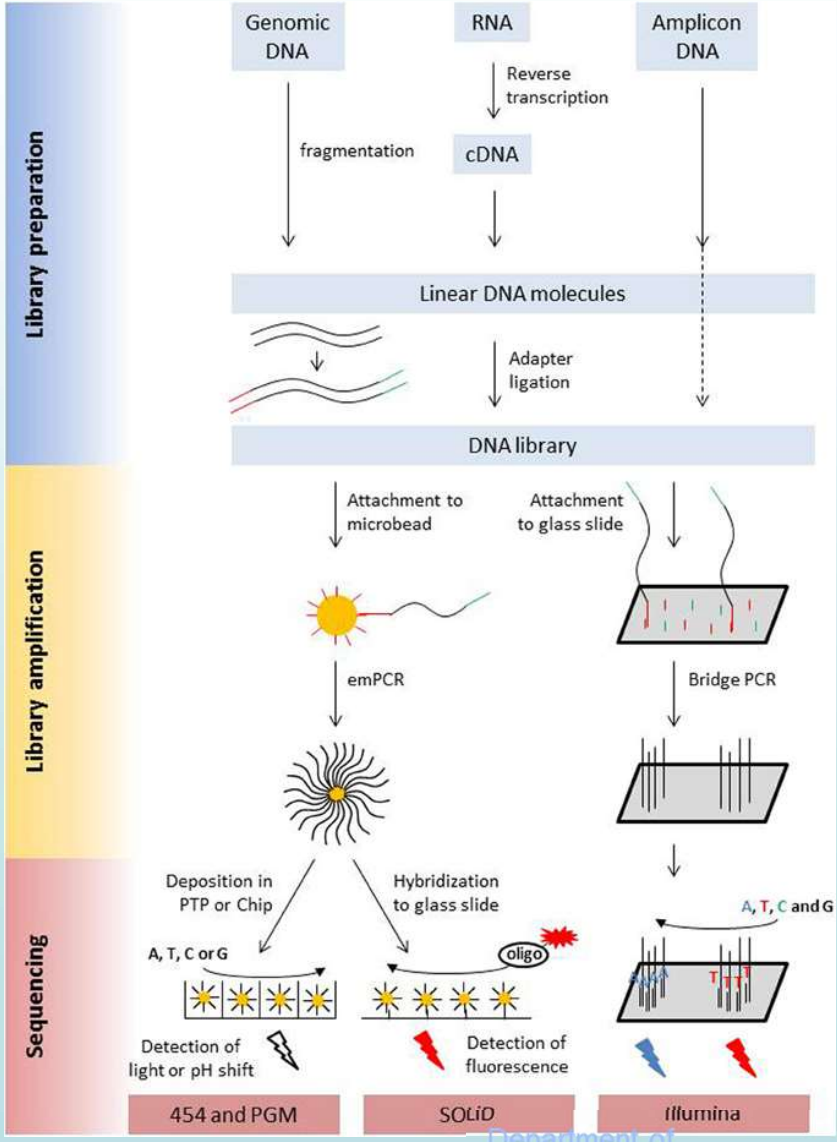


Bioinformatic interpretation

NGS

High-throughput sequencing:
"Next generation"
Able to sequence a large number of copies in a single reaction

Same procedure
Different technology depending on the platform



MOLECULAR DIAGNOSTICS

NGS

Step 1.
Sample preparation

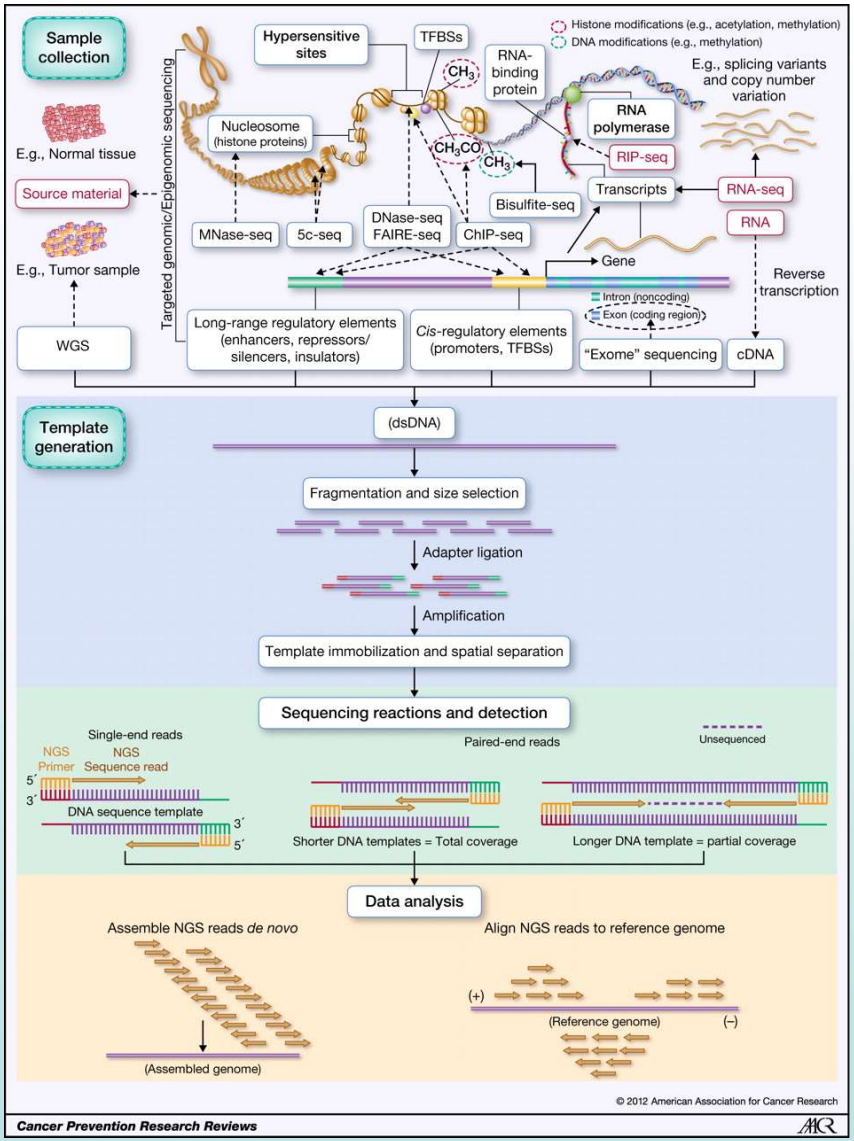
Step 2.
Template generation/ amplification/sequencing

Reads: 30-400 bps

Step 3.
Data analysis

Data analysis is essential.
Millions or trillions of data are generated.
Platform software assembles the results.
Sequence assembling:
- Known genomic sequences
- *De novo* sequences

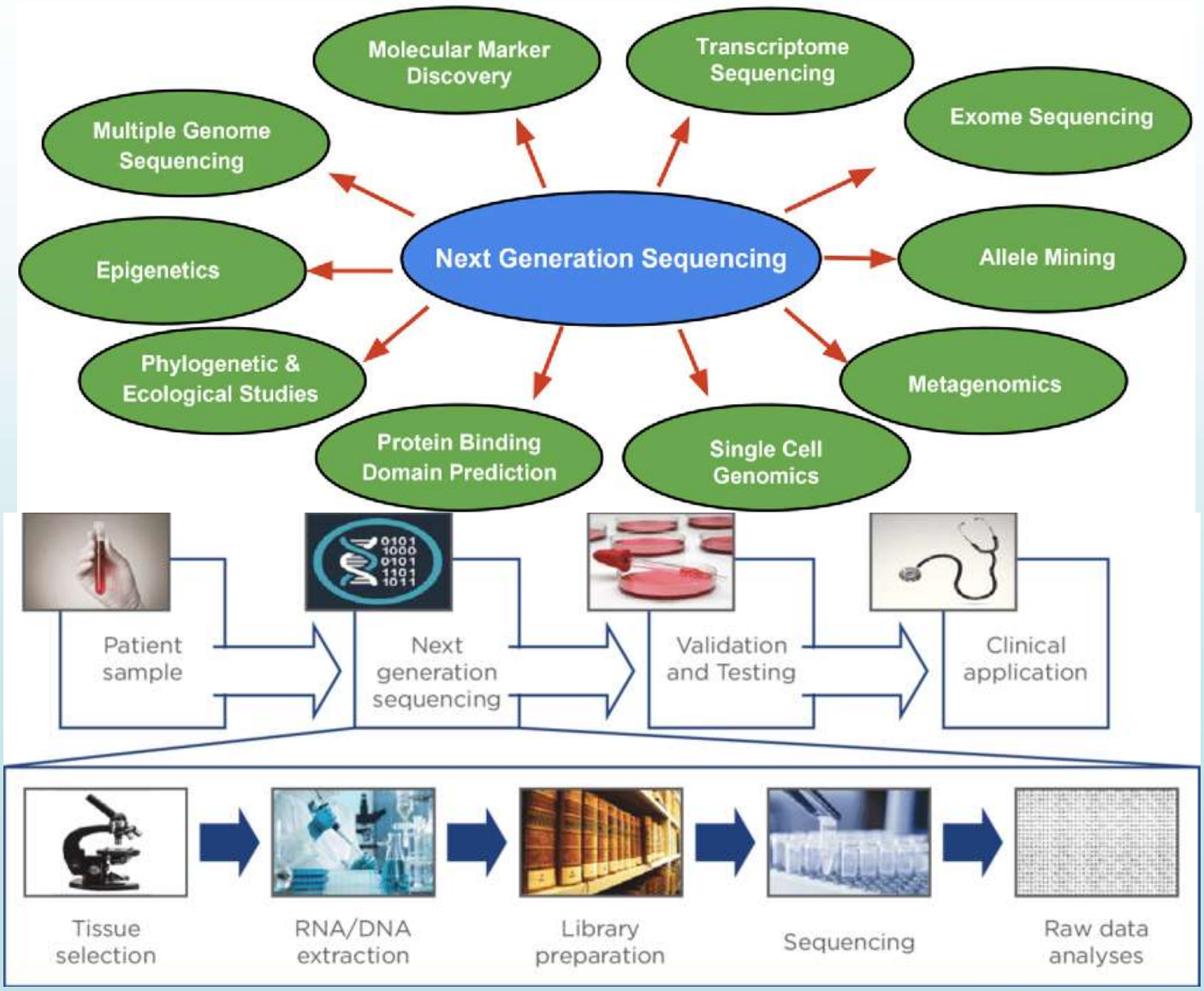
Proper bioinformatics analysis is essential.



Jason M. Rizzo, and Michael J. Buck Cancer Prev Res 2012;5:887-900

MOLECULAR DIAGNOSTICS

NGS



Landolt *et al.*, EMJ. 2016;1[2]:50-57

NGS

DIAGNOSTIC APPLICATION

- Exome and targeted sequencing
- Whole genome sequencing (WGS)
- Circulating DNA
- Transcriptome sequencing (RNA-seq)
- NGS biobanking

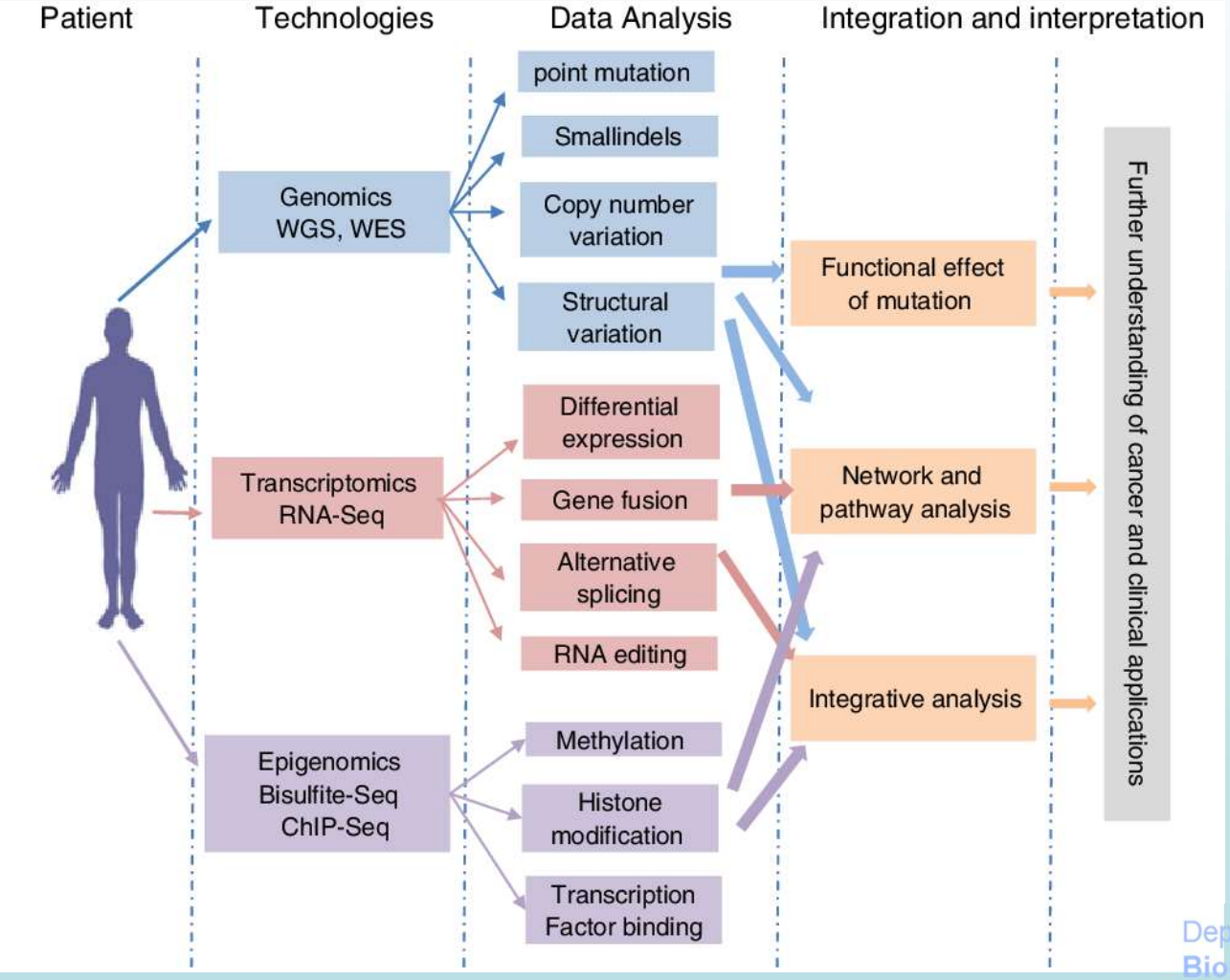
Experiment	Source DNA (input)	Description
WGS	gDNA	Identifies an individual's complete genome sequence (coding and noncoding regions); including copy number variation (e.g., repeats, indels) and structural rearragements (e.g., translocations)
Targeted "exome" sequencing	Protein-encoding gDNA (i.e., exons)	Identifies the sequence for all coding regions (exons), including copy number variation (e.g., repeats, indels) and structural rearrangements (e.g. translocations)
RNA-seq	cDNA made from various sources of RNA	Can identify all transcribed sequences (transcriptome) or just coding RNA sequences; can also provide information on sequence content (e.g., splicing variants) and copy number/abundance (e.g., gene expression profiling)
Bisulfite-seq	Bisulfite-treated DNA	Identifies sites of DNA methylation (e.g., genetic imprinting)
ChIP-seq	Immunoprecipitated DNA	Identifies sites of protein–DNA interactions such as transcription factor–binding sites
RIP-seq	cDNA made from immunoprecipitated RNA	Identifies sites of protein–RNA interactions; a ChIP-seq for RNA-binding proteins
DNase-seq	DNase-digested chromatin DNA	Identifies genomic regions susceptible to enzymatic cleavage by DNase, i.e., hypersensitive sites and potential regulatory regions
FAIRE-seq	Open/accessible chromatin DNA	Identifies open/accessible chromatin regions, i.e., hypersensitive sites and potential regulatory regions
MNase-seq	Nucleosome-associated DNA	Identifies nucleosome positions on genomic DNA (i.e., primary chromatin structure); also provides information on histone/ nucleosome density at each location
Hi-C/5C-seq	Captured chromosome conformations	Identifies intra- and interchromosomal interactions; determines the spatial organization of chromosomes at high resolution
Metagenomics	Microbial DNA populations	Genomic analysis of microbial communities; identifies bacterial/viral populations present in specific environments (e.g., human gut and tumor samples)

NOTE: Immunoprecipitated (IP) DNA and RNA can be collected for any protein that has an antibody or using an epi
IP DNA sources can include histone proteins (e.g., histone H3 or H4), as a paired or alternative approach to MNa
(77). IP DNA sources can also include covalently modified histone proteins (i.e., specific histone acetylation
H3K36Me3)–to map "histone code."
Abbreviations: cDNA, reverse-transcribed RNA or "complementary DNA" (i.e., introns removed during RNA splic
collection of DNA and proteins in the nucleus; openness/accessibility of chromatin, regions of looser DNA packa
enzymatic cleavage and protein binding/gene regulation (see Fig. 1: "Hypersensitive Sites"); indels, insertions/
tome, all transcribed DNA sequences, includes small noncoding RNAs, miRNAs, and coding RNAs (i.e. genes)

NGS

DIAGNOSTIC APPLICATION

Workflow for integrating omics data into cancer research (disease) and clinical applications



NGS

DIAGNOSTIC APPLICATION

- ❑ **Exome and targeted sequencing**

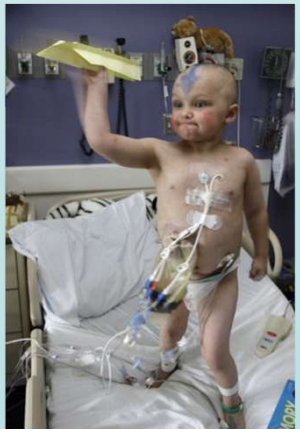
histopat LABORATORIS

Cancer panels study

- **500 variants analyzed, including deletions, inversions, insertions and substitutions in 22 genes (*EGFR, ALK, ERBB2, ERBB4, FGFR1, FGFR2, TP53*…)**
- **Related with different types of solid neoplasia**

- **Presence or absence of mutations**
- **Diagnostic factors / prognosis / biomarkers that indicate/advise against therapies**

Diagnosis of unknown genetic diseases



- **Exome sequencing to understand recessive hereditary diseases**
- **Child developing inflammatory bowel disease**
  - **Wounds in the gut / unknown cause**
  - **100 visits to hospital from age 4**
- **Exome sequencing reveals mutation in *XIAP* gene**
  - **Bone marrow transplant as a treatment**

NGS

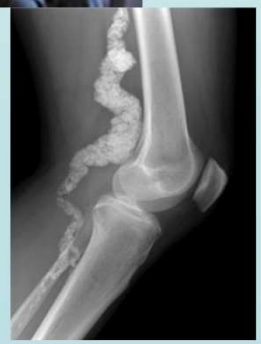
DIAGNOSTIC APPLICATION

- **Whole genome sequencing (WGS)**

Oncological application

Studies of mutations, methylation, sequence fusion, SNPs, etc.

Diagnosis of unknown genetic diseases



- **Study of all regions of the genome (SNP array) to understand recessive hereditary disorders**
- **Patients who suffer from joint pain and show calcium deposits in their arteries on x-ray**
  - **WGS shows mutation in *NT5E* gene, involved in breaking down calcifications in arteries**

MOLECULAR DIAGNOSTICS

VNIVERSITAT DE VALÈNCIA

NGS

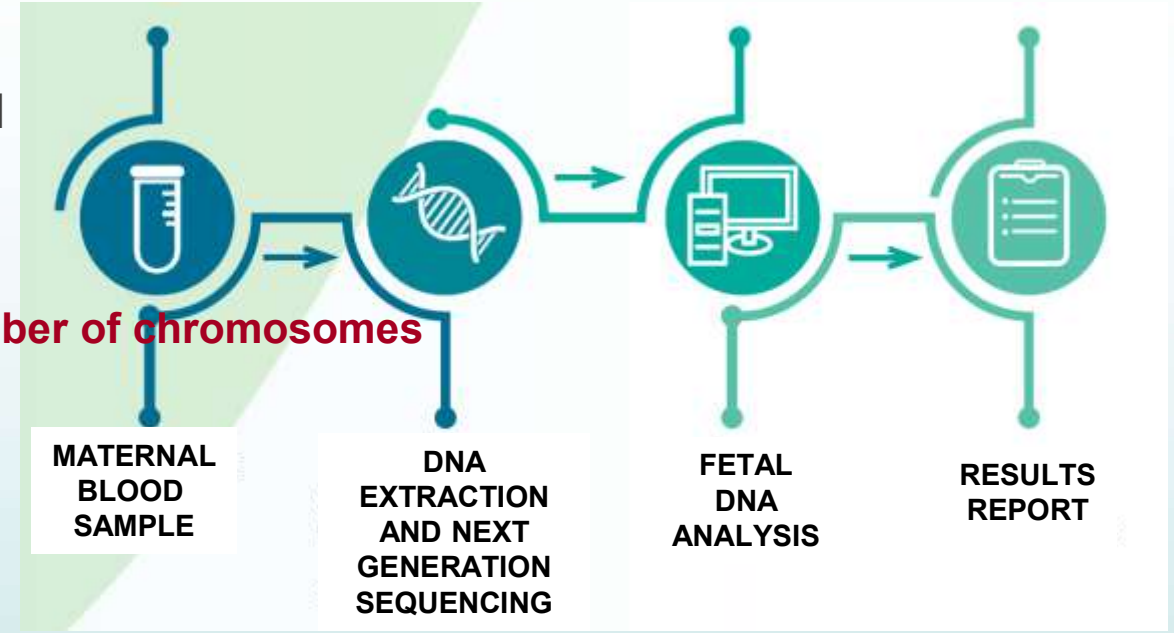
DIAGNOSTIC APPLICATION

- **Circulating DNA**

Prenatal diagnosis

Detection of anomalies in the number of chromosomes
99% precision

- Down's syndrome
- Edwards' syndrome
- Patau's syndrome



SENSITIVITY >99.9%

Identifies the presence of anomalies

With reduced possibilities of false negatives

SPECIFICITY >99.9%

Excludes the presence of anomalies

With reduced possibilities of false positives

Results in 5-7 days

Diagnosis of a great number of disorders with a genetic cause

Department of
Biochemistry and Molecular Biology

NGS

DIAGNOSTIC APPLICATION

- **Circulating DNA**

Circulating tumor DNA

Liquid biopsy

New UltraSEEK™ Lung Panel

Gene	Coverage (Missense mutations)
BRAF	Codon 469 (exon 11) and codons 594, 600 (exon 15)
EGFR	E709A, E709G, E709K, E709V, G719A, G719D, G719S, G719C, S768I, T790M, L858R, L861Q, L861R, C797S, Exon 19 indels, Exon 20 insertions
KRAS	G12A, G12C, G12D, G12R, G12S, G12V, G13C, G13D, Q61H, Q61K, Q61E, Q61P, Q61R, Q61L
ERBB2	A775_G776insYVMA, G776>VC
PIK3CA	Codons 542, 545 (exon 9), codon 1047 of (exon 20)
TOTAL	5 Genes

MOLECULAR DIAGNOSTICS

NGS

Emergence of biotechnology companies (ie. in Valencia)

Non-invasive sample collection
Basic analysis techniques (from the NGS spectrum)

Great number of services offered

Low cost/price

The techniques applied are not optimal

Results based on database availability

Metadata analysis / bioinformatics interpretation

Incomplete / partially wrong results

NGS

DIAGNOSTIC APPLICATION

Future perspectives

Is NGS the ultimate "personalized medicine"?

- It has a predictive value of disease risk and potential for delaying or preventing the development of the disease.
- It can predict adequate drug dosage, drug response or adverse effects.
- It can help determine the treatment of diseases based on the patient's genetic profile.

Mol Biosyst. 2016 May 24;12(6):1818-30. doi: 10.1039/c6mb00115g.

Next generation sequencing: implications in personalized medicine and pharmacogenomics.

Rabbani B[1], Nakaoka H[2], Akhondzadeh S[3], Tekin M[4], Mahdieh N[1].

Author information

http://www.labgenetics.com.es/catalogo_enfermedades_hereditarias.htm

- Legal/political adaptations
- Work adaptations
- Social adaptations

Post-genomic era of medicine

CONCLUSIONS

The application of omics for the diagnosis, prognosis and study of diseases has great value and potential.

Cytomics is used for the diagnosis and prognosis of hematological and oncological diseases.

Metabolomics develops biomarkers for the current and potential diagnosis of multiple pathologies.

Molecular diagnostics has the greatest application and potential at present.

- **CGH arrays**
- **Expression microarrays**
- **Sequencing/NGS**

VNIVERSITAT DE VALÈNCIA

Advanced clinical diagnostic techniques: omics and their application to the molecular study of diseases

BIOCHEMICAL INTEGRATION AND
CLINICAL BIOCHEMISTRY
2ND YEAR - DEGREE IN MEDICINE 2022-23

SEMINAR 4

PROFESSOR:
Amparo Galán Albiñana
amparo.galan@uv.es

Department of
Biochemistry and Molecular Biology