Genómica funcional

Introducción

Valencia, Septiembre 2011
CEFIRE

Jorge Jiménez
jjimeneza@cipf.es
http://www.bioinformatico.es
Bioinformatics and Genomics Department
Centro de Investigacion Principe Felipe (CIPF)
(Valencia, Spain)

INB

PRINCIPE FELIPE
CENTRO DE INVESTIGACION

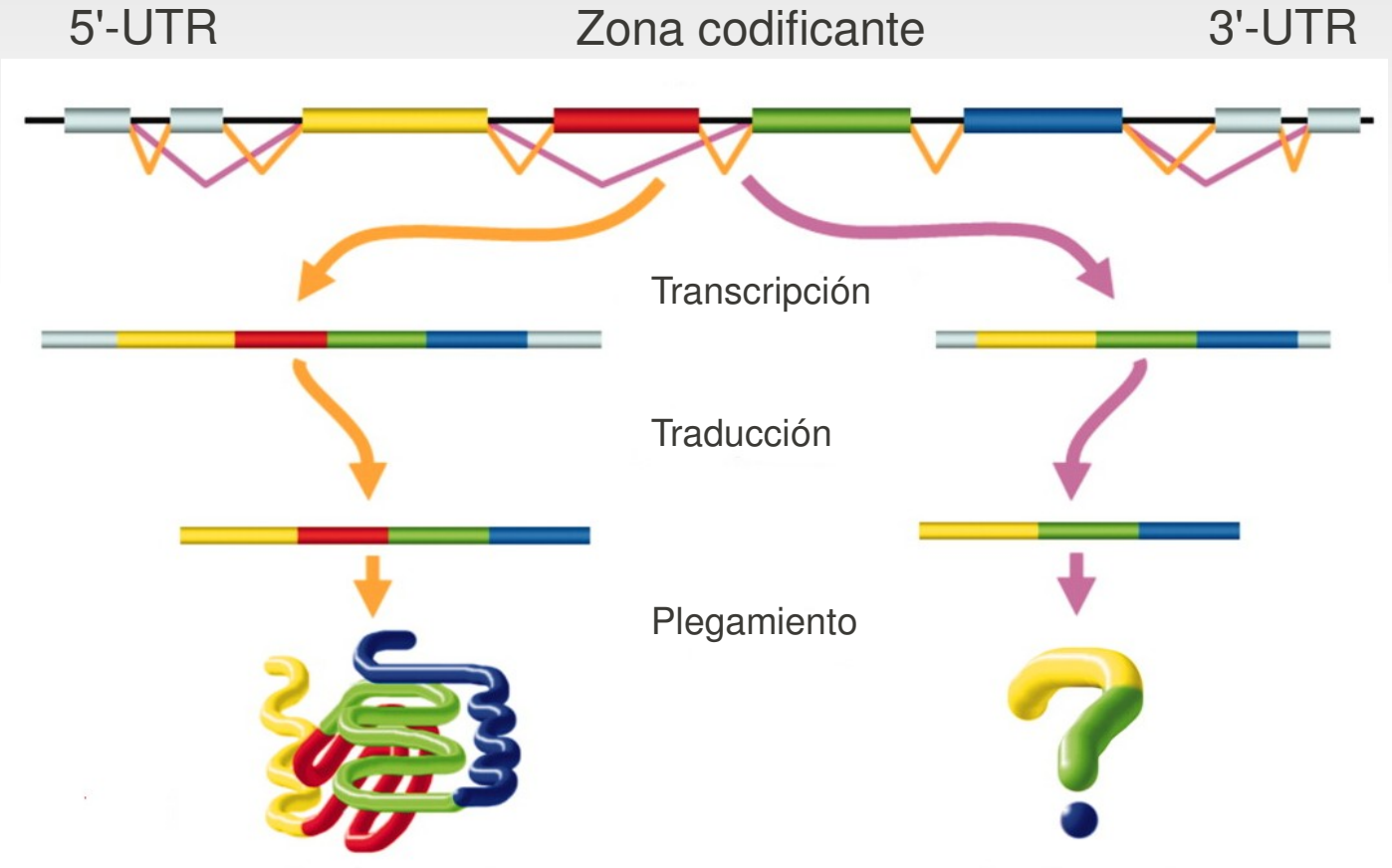
ciberer

Definición y objetivos

- Entender la relación entre el genotipo y el fenotipo
- Entender la función de los genes y la de los demás componentes del genoma
- Estudio a nivel de ADN, ARN y proteína
- Utiliza datos generados por máquinas de alto rendimiento

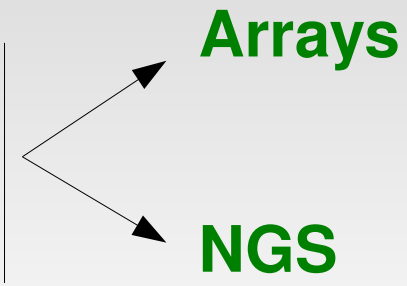
Estructura de un gen

- Exones, intrones, zonas UTR



Niveles

- ADN: SNPs, CNVs, etc.
- ARN: expresión, microARN, etc.

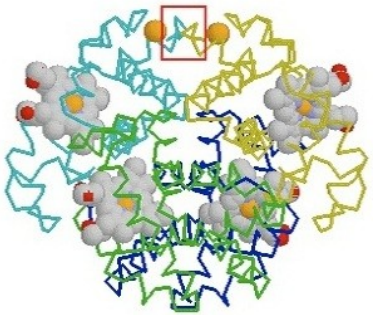


- Proteínas: interacciones proteína-proteína
- Anotación genómica
- Anotación funcional

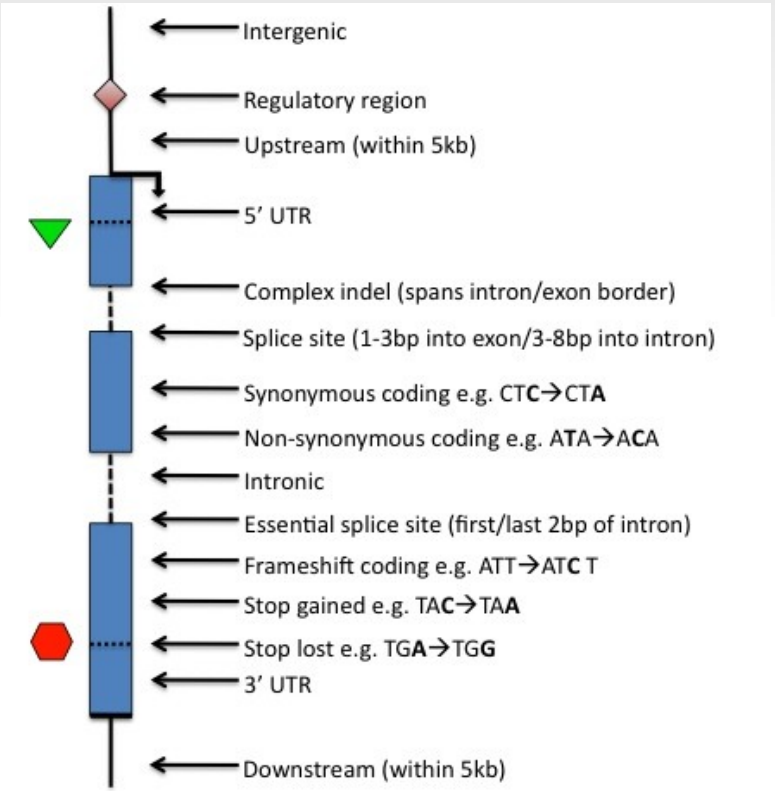
BIOINFORMÁTICA!!

ADN

SNP



Consecuencias



Others: Within non-coding gene, Within mature miRNA, NMD transcript

SNP

Técnicas de detección:

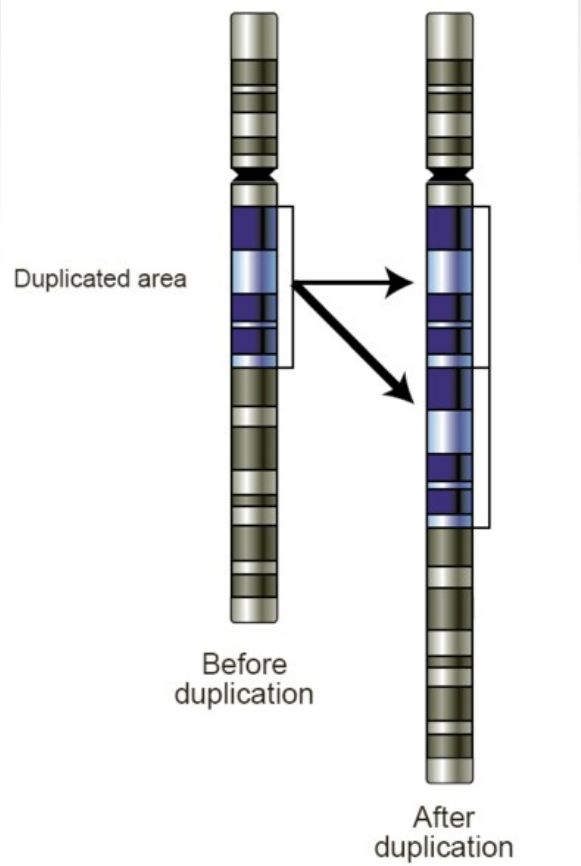
- Secuenciación: Sanger
- Espectrofotometría de masas
- single-strand conformation polymorphism (SSCP)
- Análisis electroquímico
- HPLC y electroforesis en gel
- restriction fragment length polymorphism
- Análisis por hibridación

Técnicas de alto rendimiento

- Chips de SNPs
- NGS: análisis de exoma, secuenciación genoma completo, resecuenciación dirigida

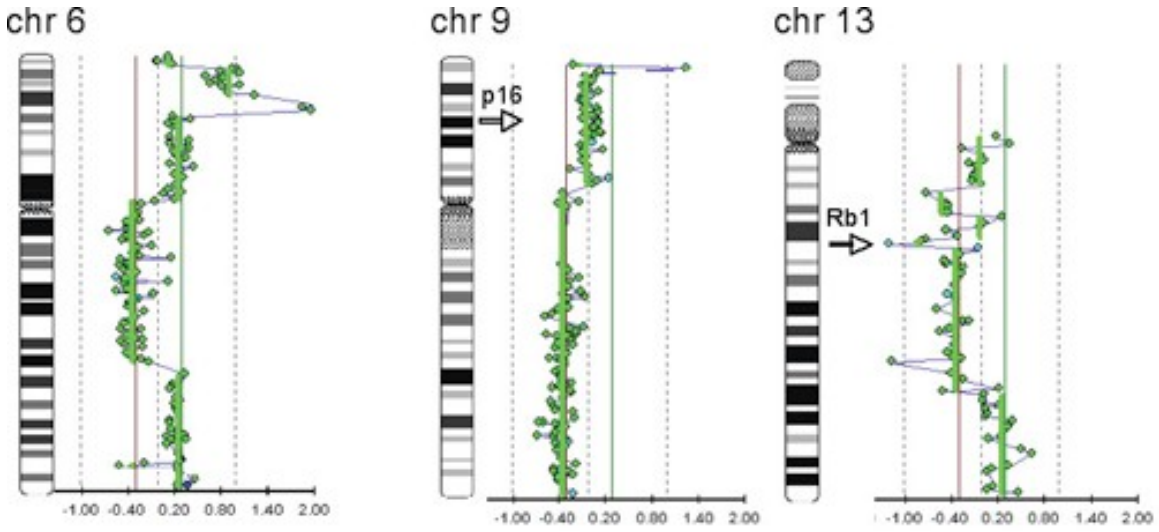
ADN - CNVs

CNVs (12%)

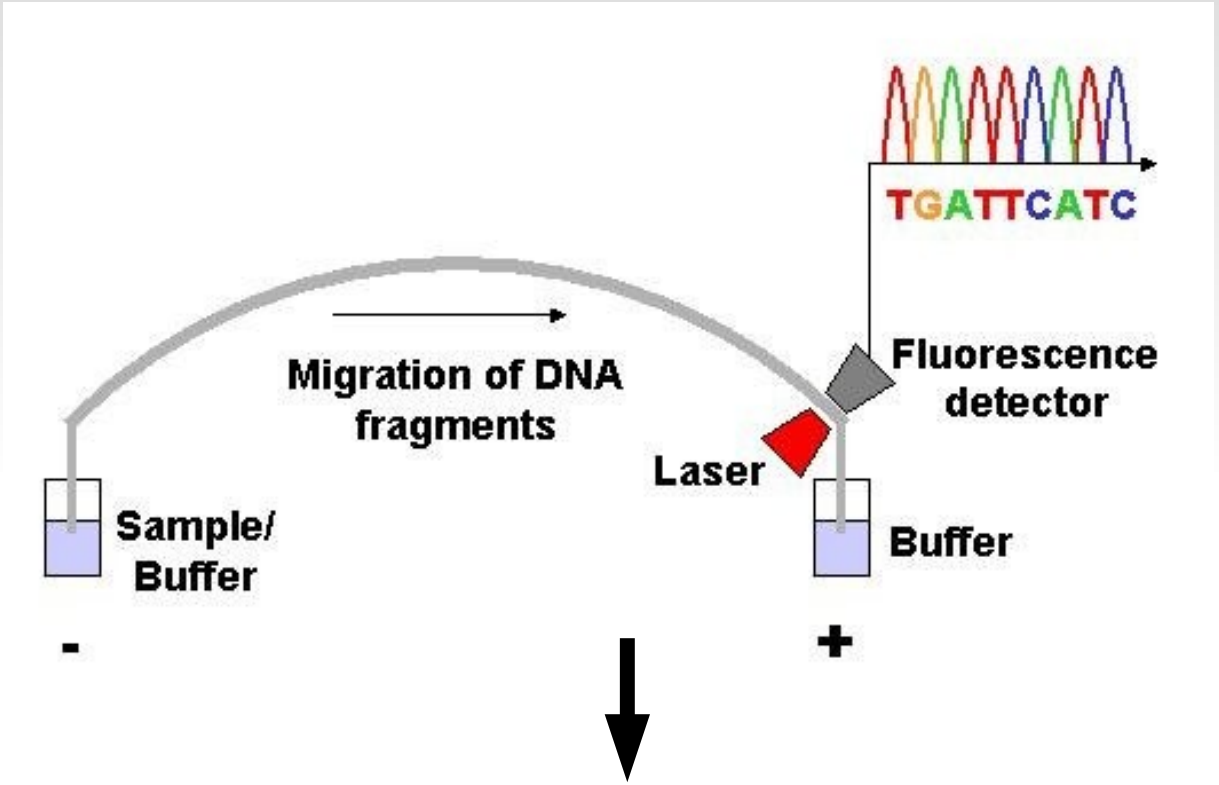


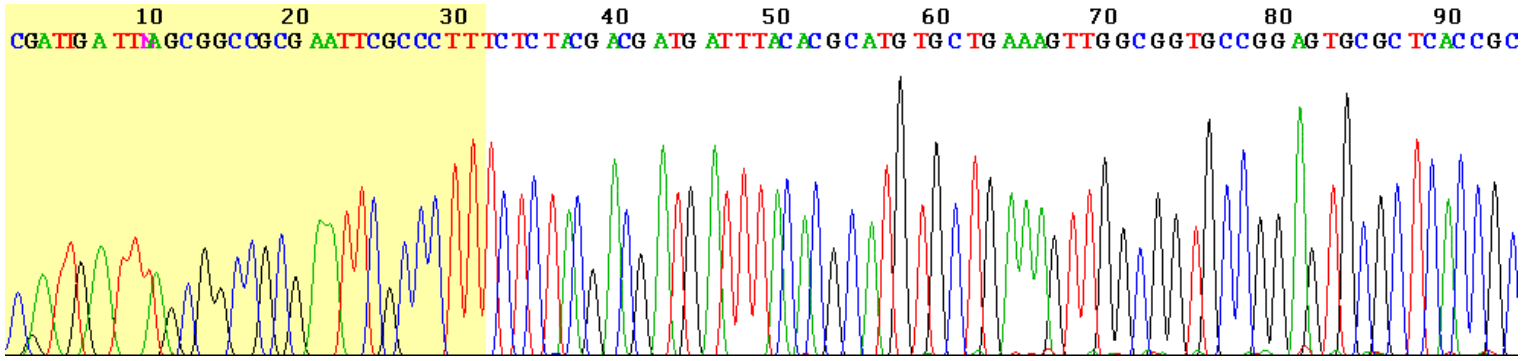
Detección:

- CGH array: comparación entre dos muestras.
- Arrays de SNPs
- NGS



Secuenciación de primera generación

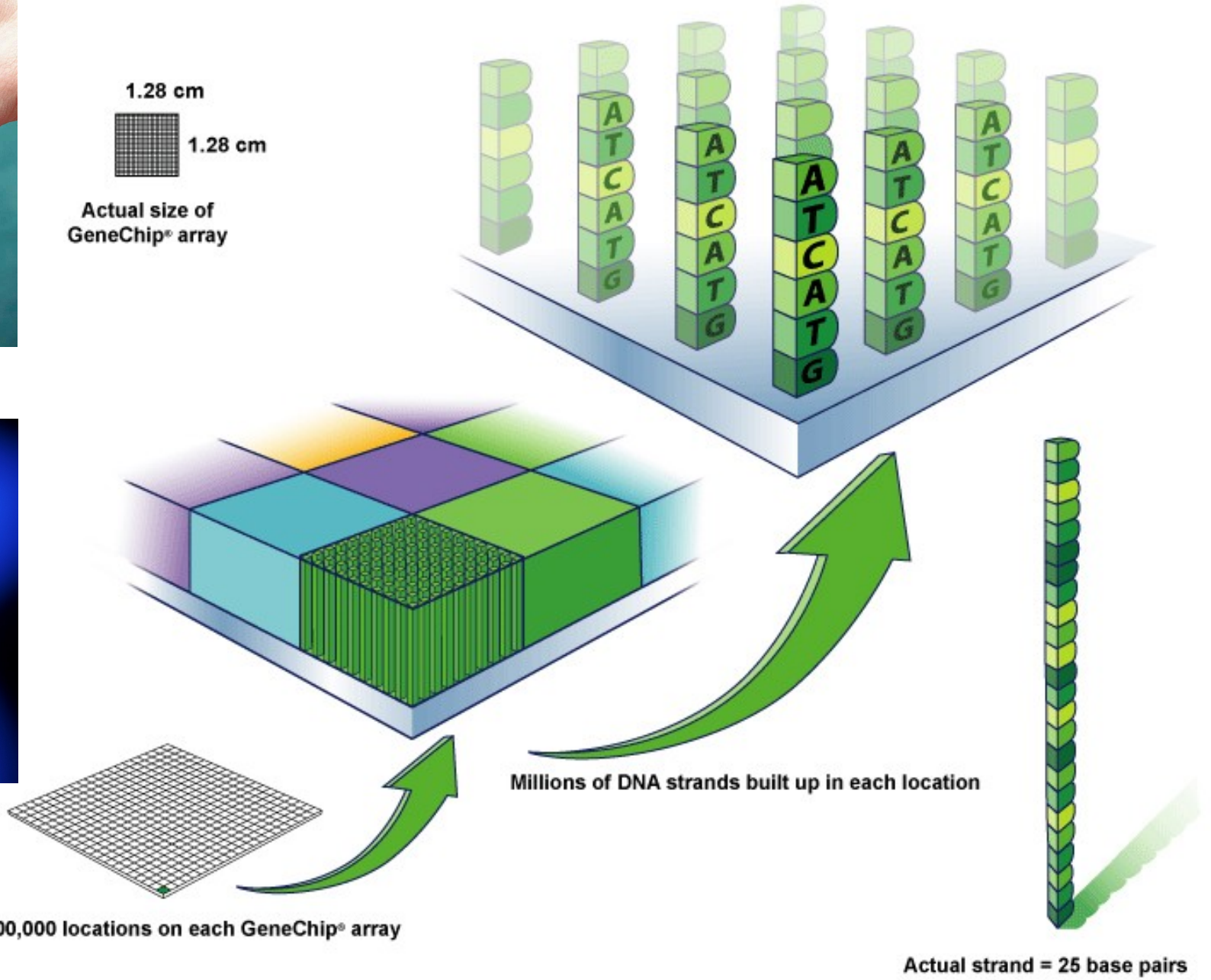




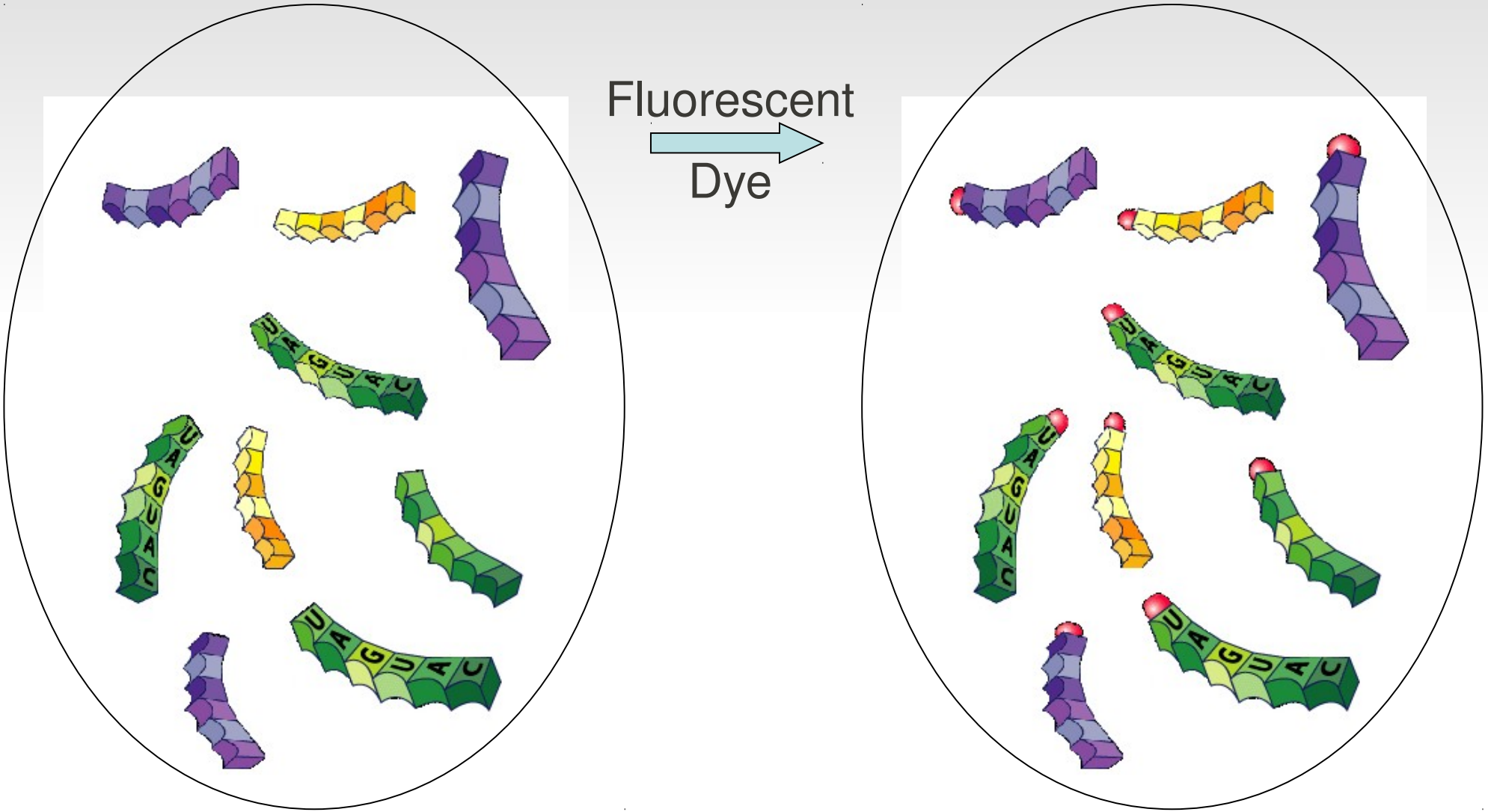
Chips de ADN

- Primer método de alto rendimiento
- Capaces de determinar genes, SNP, mRNA, CNVs,...
- Medida en diferentes condiciones
- Análisis a escala genómica

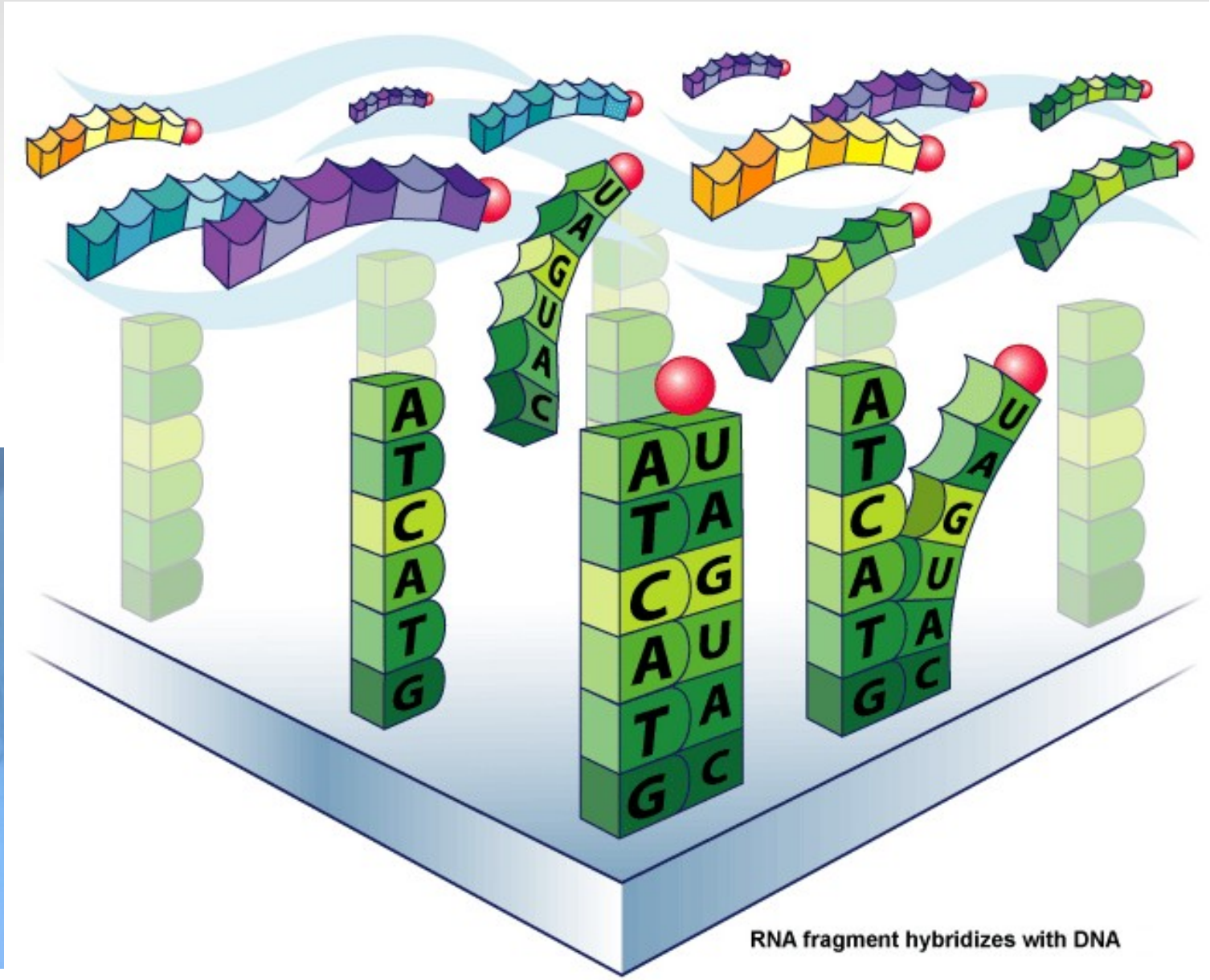
Microarrays de ADN



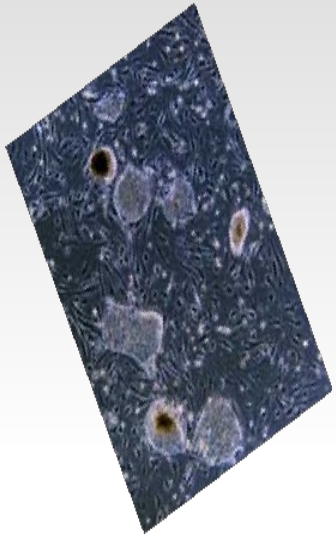
Labelling the Sample



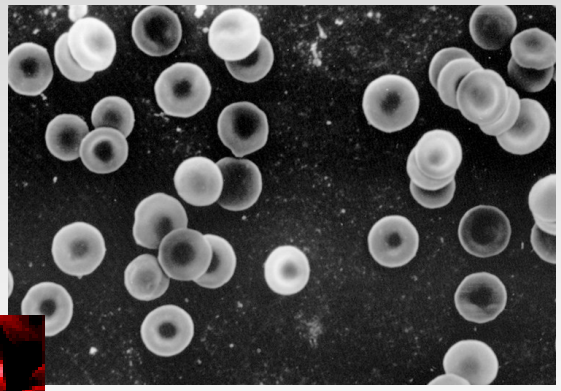
Hibridación

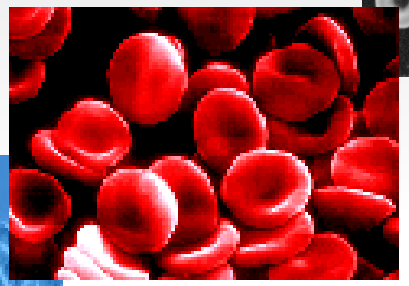


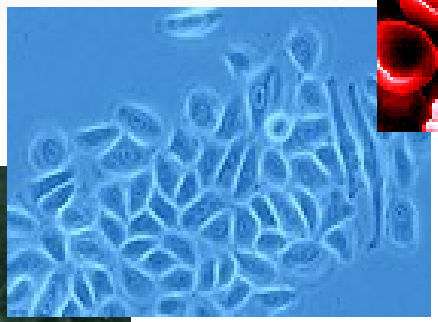
ARN

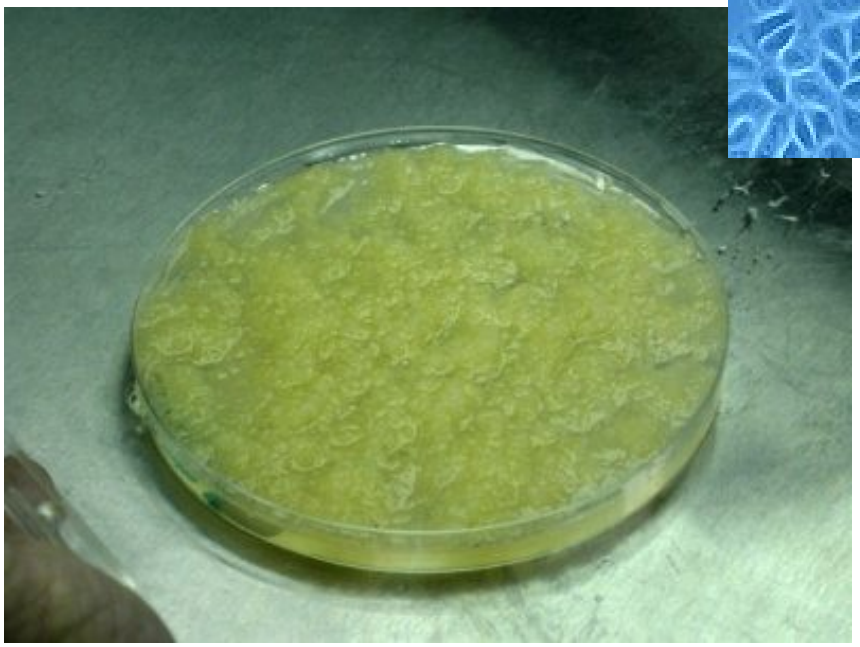


Queremos saber qué genes están expresados en condiciones celulares muy concretas





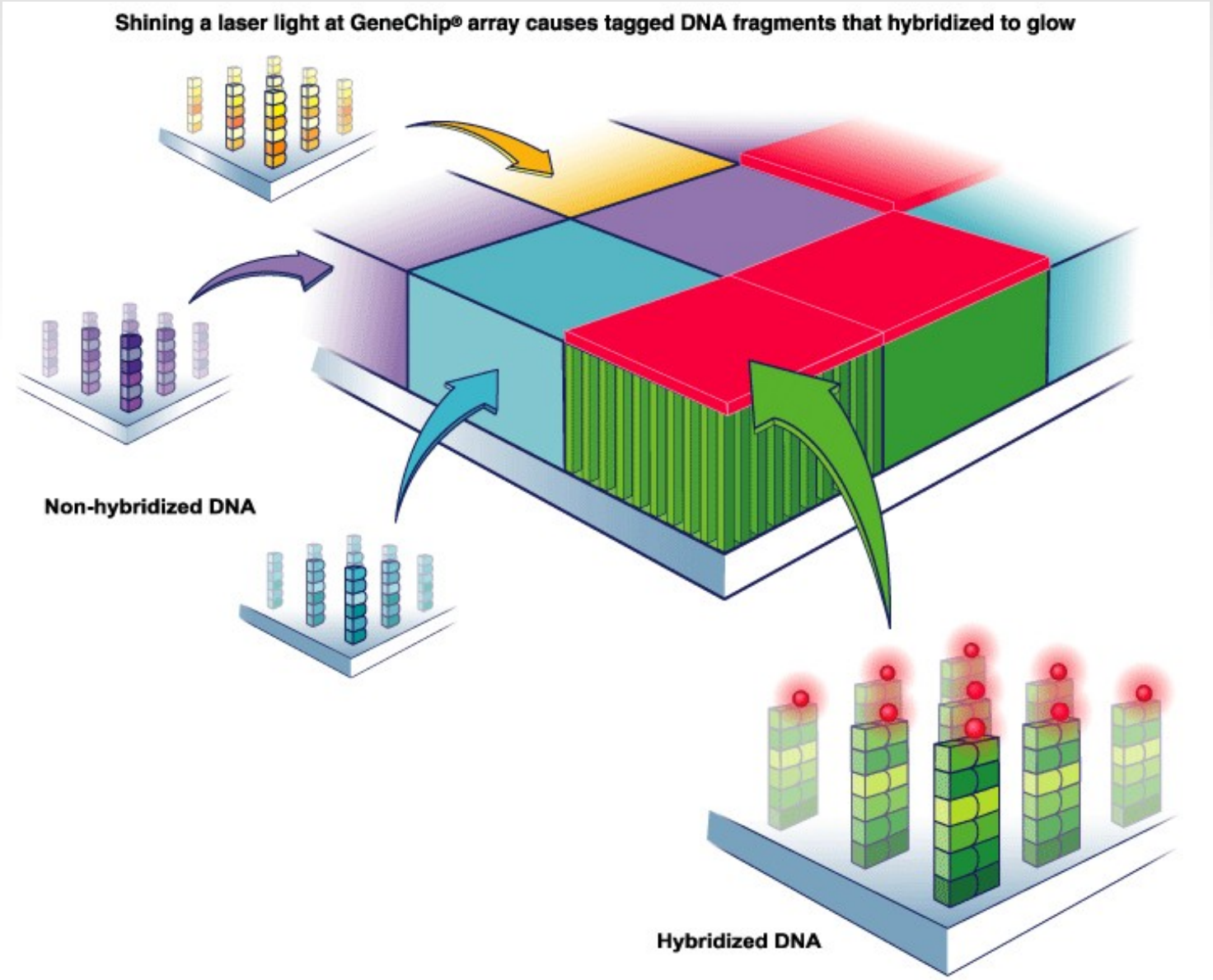




Somos capaces de extraer todas las moléculas de mARN presentes en las células

y conocer el nivel de expresión como el indicador de la concentración en la muestra biológica

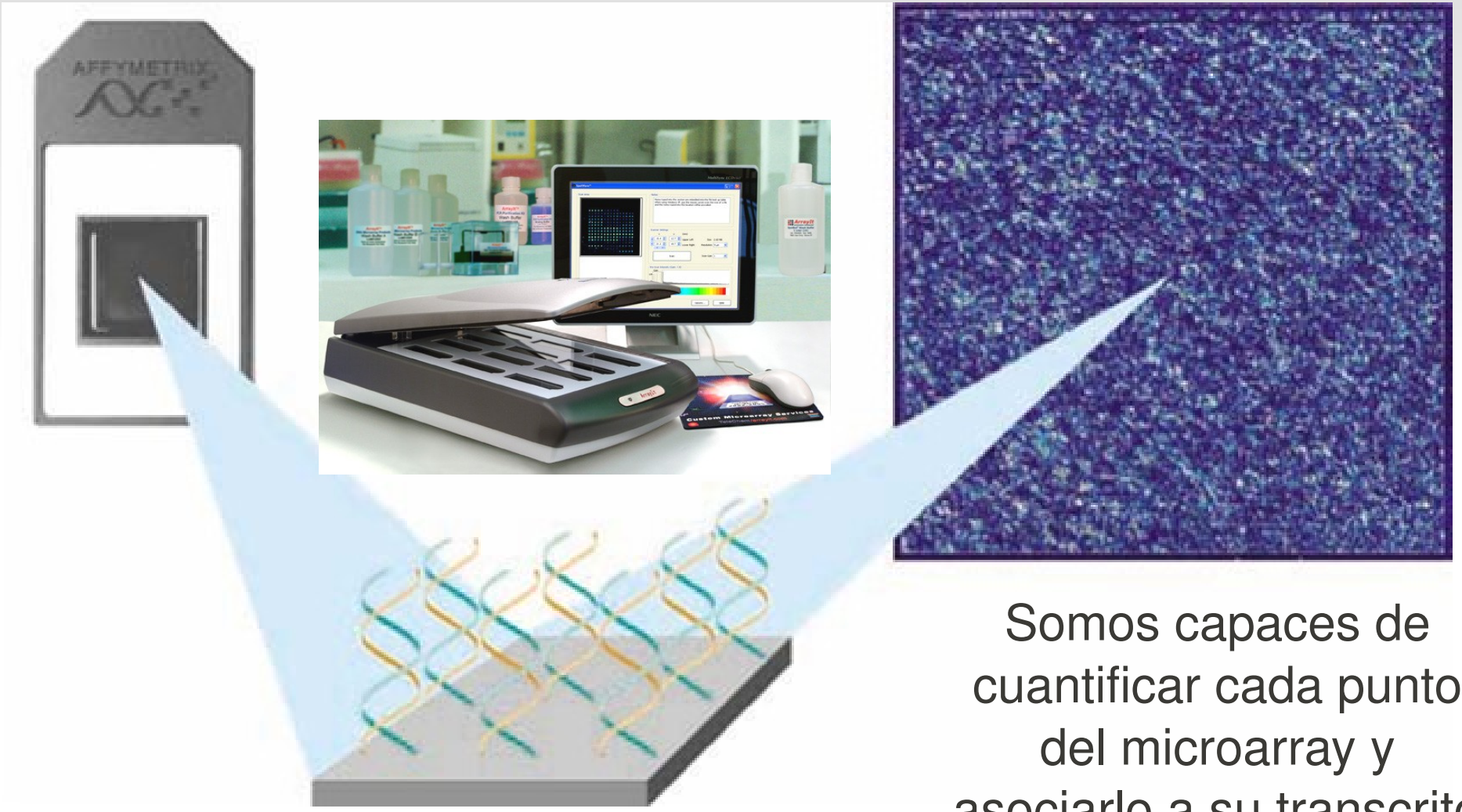
Expression Measurement



Detectaremos aquellos genes que se estén expresando cuando se unan a las sondas marcadas

Cuánto mayor sea la florescencia, mayor será la concentración del transcrito

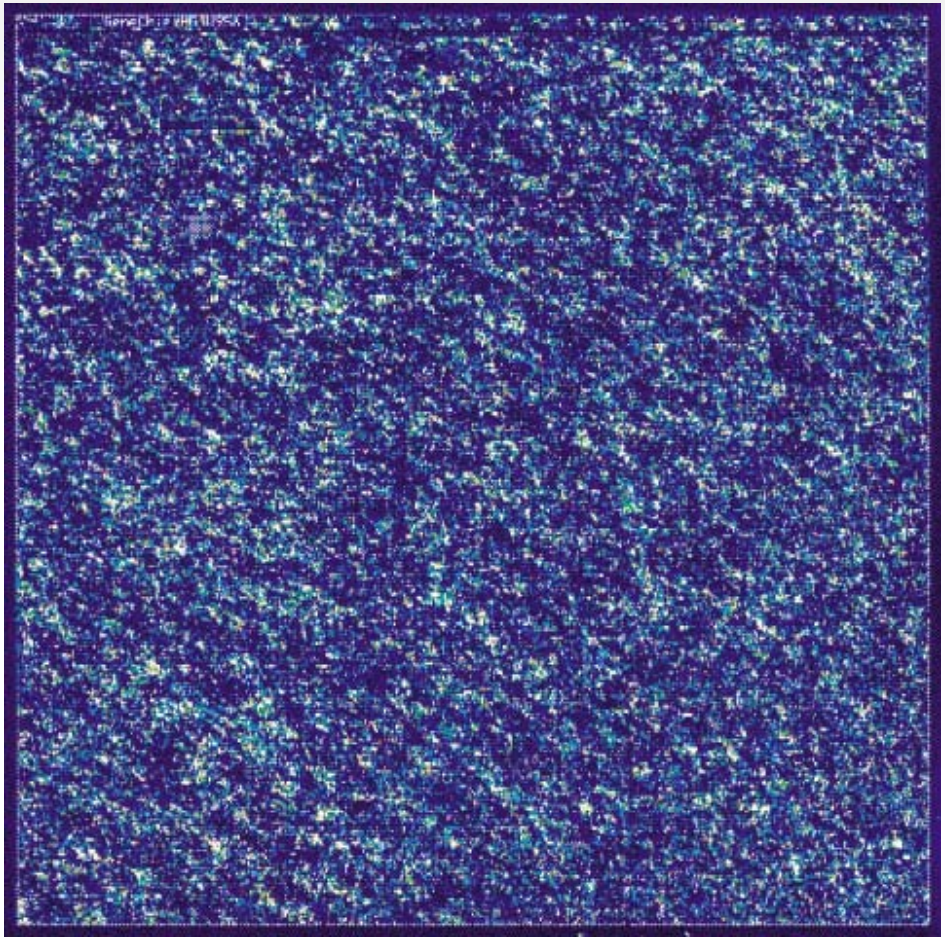
Escáner del microarray



Somos capaces de cuantificar cada punto del microarray y asociarlo a su transcrito

Los datos

Para cada muestra biológica

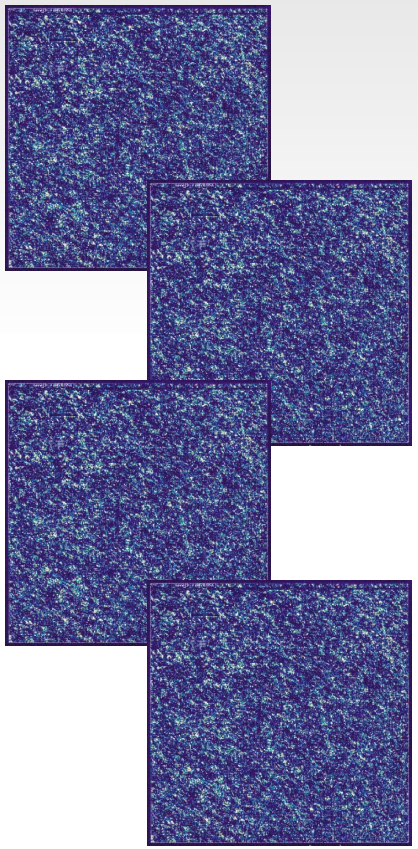


Obtenemos la intensidad de fluorescencia de miles de transcritos

La medida de intensidad es un indicador de la expresión génica

200000_s_at	134.4
200001_at	586.5
200002_at	1868.4
200003_s_at	1232.7
200004_at	1071.6
200005_at	312.8
200006_at	1712.6
200007_at	606.5
200008_s_at	421.9
200009_at	395.6
200010_at	1228.6
200011_s_at	132.5
200012_x_at	2606.3
200013_at	1572.9
200014_s_at	138.7
200015_s_at	124.1
200016_x_at	1058.7
200017_at	889.4
200018_at	3964.2
200019_s_at	1069.9
200020_at	212.1
200021_at	1018.1
200022_at	1254.8
200023_s_at	1202.8
200024_at	2460.6

Varios Microarrays



31307_at	5.53
31308_at	7.07
31309_r_at	6.05
31310_at	7.42
31311_at	7.77
31312_at	9.47
31313_at	8.58
31314_at	7.14
31315_at	9.62
31316_at	5.45
31317_r_at	10.27
31318_at	5.7
31319_at	9.25
31320_at	11.5
31321_at	7.79
31322_at	6.98
31323_r_at	11.18
31324_at	7.97
31325_at	9.53
31326_at	9.67
31327_at	6.48
31328_at	8.92
31329_at	6.11
31330_at	14.44
31331_at	6.3

31307_at	5.66
31308_at	7.14
31309_r_at	5.33
31310_at	7.02
31311_at	7.83
31312_at	9.43
31313_at	8.67
31314_at	7.3
31315_at	9.62
31316_at	5.53
31317_r_at	10.75
31318_at	5.53
31319_at	9.19
31320_at	11.51
31321_at	7.91
31322_at	6.93
31323_r_at	10.27
31324_at	8.12
31325_at	9.37
31326_at	10.16
31327_at	6.2
31328_at	9.11
31329_at	5.86
31330_at	14.32
31331_at	6.4

31307_at	5.52
31308_at	7.05
31309_r_at	5.35
31310_at	7.02
31311_at	7.79
31312_at	9.34
31313_at	8.52
31314_at	7.19
31315_at	9.22
31316_at	5.3
31317_r_at	10.41
31318_at	5.59
31319_at	9.24
31320_at	11.35
31321_at	7.76
31322_at	6.91
31323_r_at	10.32
31324_at	8.06
31325_at	9.33
31326_at	9.92
31327_at	6.2
31328_at	8.81
31329_at	5.81
31330_at	14.3
31331_at	6.42

Chips de SNPs

- Análisis de miles de SNPs a la vez
- Aplicaciones
  - Genética de poblaciones
  - GWAS
  - Análisis de enfermedades
  - Análisis del número de copias del SNP

NGS

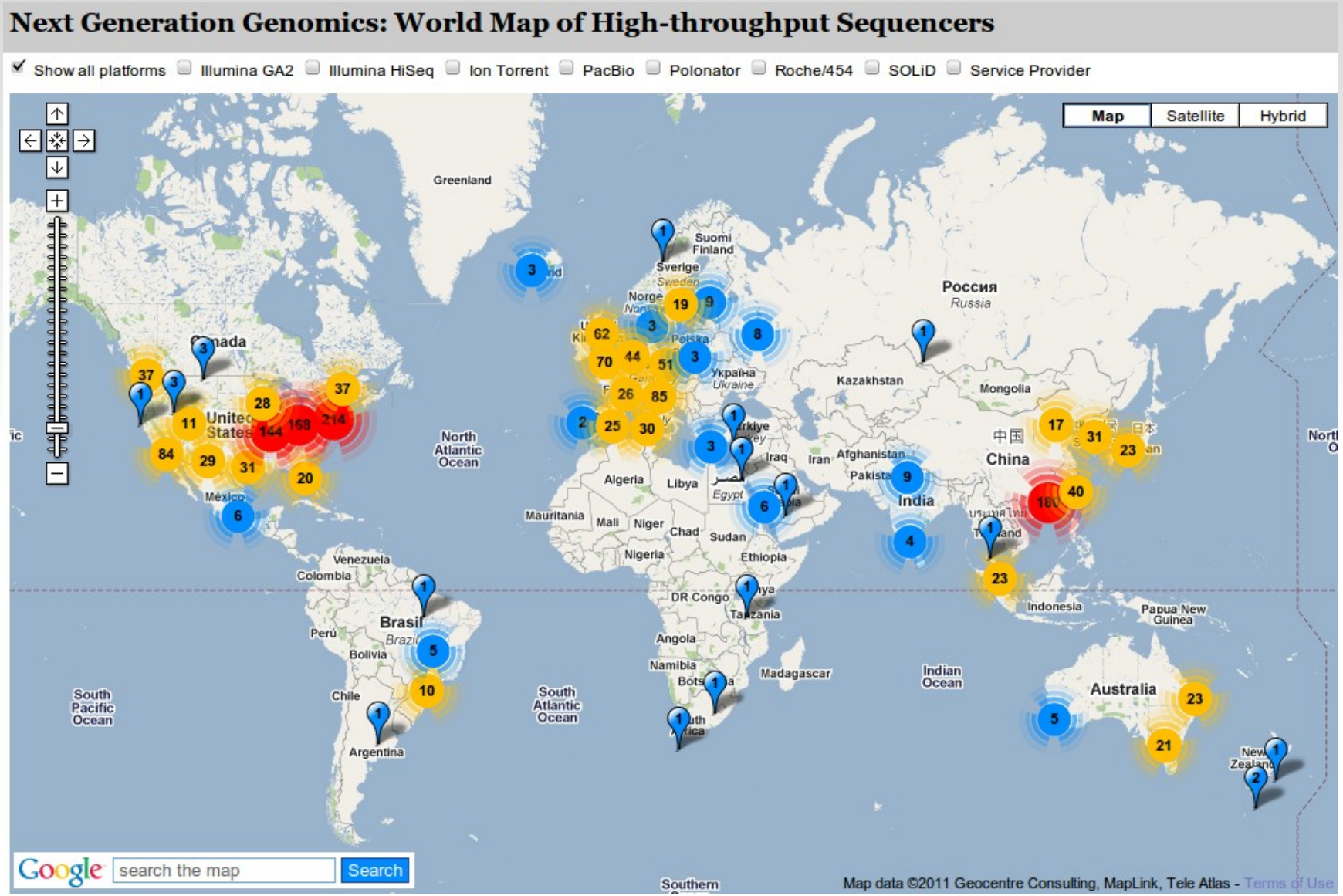
- Características más importantes:
  - Paralelización del proceso
  - Generación de millones de secuencias por experimento

- Ventajas
  - Barato
  - Gran versatilidad de análisis

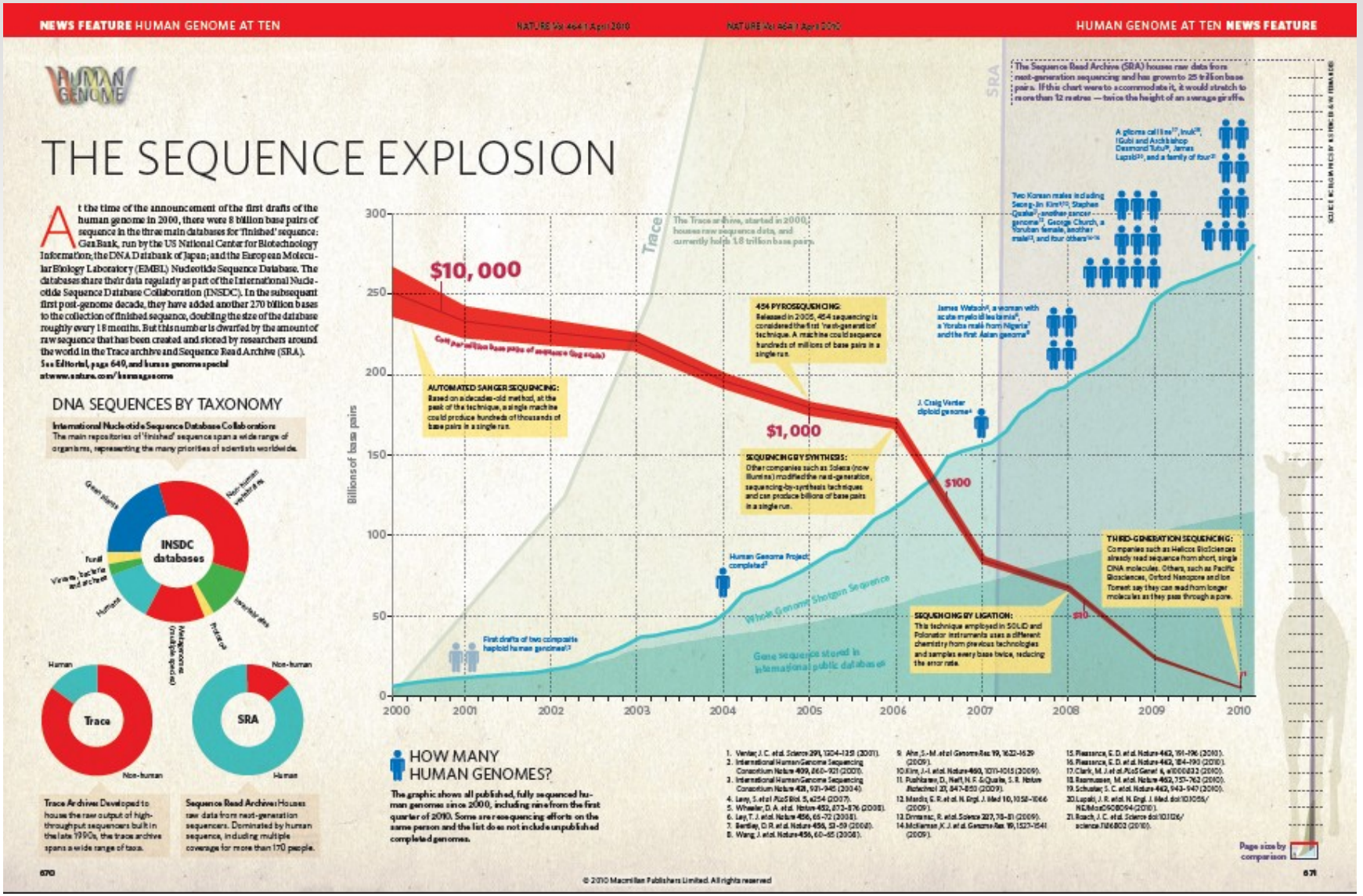
NGS

- Se utitilizan para secuenciación de novo, resecuenciación dirigida de ADN (ej: exoma) y estudio de expresión (RNA-Seq)
- Secuenciadores de segunda generación
- Plataformas más importantes:
  - 454 – pirosecuenciación
  - Illumina
  - SOLiD
  - etc.

NGS


Next Generation Genomics: World Map of High-throughput Sequencers
Show all platforms
Illumina GA2
Illumina HiSeq
Ion Torrent
PacBio
Polonator
Roche/454
SOLiD
Service Provider
Map
Satellite
Hybrid
Greenland
Suomi Finland
Sverige
Sweden
Norge
Polska
Україна
Ukraine
Россия
Russia
Kazakhstan
Mongolia
China
India
Afghanistan
Pakistan
Iran
Iraq
Algeria
Libya
Egypt
Mauritania
Mali
Niger
Chad
Sudan
Nigeria
Ethiopia
DR Congo
Tanzania
Angola
Namibia
Madagascar
Indonesia
Papua New Guinea
Australia
New Zealand
Venezuela
Colombia
Brasil
Perú
Bolivia
Chile
Argentina
North Atlantic Ocean
South Pacific Ocean
South Atlantic Ocean
Indian Ocean
Southern
Google
search the map
Search
Map data ©2011 Geocentre Consulting, MapLink, Tele Atlas - Terms of Use

Revolución tecnológica



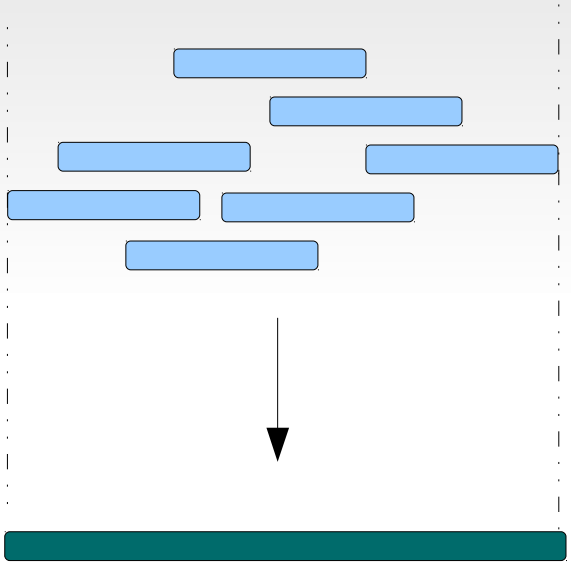
Comparativa de plataformas

Platform	Library/ template preparation	NGS chemistry	Read length (bases)	Run time (days)	Gb per run	Machine cost (US$)	Pros	Cons	Biological applications	Refs
Roche/454's GS FLX Titanium	Frag, MP/ emPCR	PS	330*	0.35	0.45	500,000	Longer reads improve mapping in repetitive regions; fast run times	High reagent cost; high error rates in homo-polymer repeats	Bacterial and insect genome *de novo* assemblies; medium scale (<3 Mb) exome capture; 16S in metagenomics	D. Muzny, pers. comm.
Illumina/ Solexa's GA$_{II}$	Frag, MP/ solid-phase	RTs	75 or 100	4‡, 9§	18‡, 35§	540,000	Currently the most widely used platform in the field	Low multiplexing capability of samples	Variant discovery by whole-genome resequencing or whole-exome capture; gene discovery in metagenomics	D. Muzny, pers. comm.
Life/APG's SOLiD 3	Frag, MP/ emPCR	Cleavable probe SBL	50	7‡, 14§	30‡, 50§	595,000	Two-base encoding provides inherent error correction	Long run times	Variant discovery by whole-genome resequencing or whole-exome capture; gene discovery in metagenomics	D. Muzny, pers. comm.
Polonator G.007	MP only/ emPCR	Non-cleavable probe SBL	26	5§	12§	170,000	Least expensive platform; open source to adapt alternative NGS chemistries	Users are required to maintain and quality control reagents; shortest NGS read lengths	Bacterial genome resequencing for variant discovery	J. Edwards, pers. comm.
Helicos BioSciences HeliScope	Frag, MP/ single molecule	RTs	32*	8‡	37‡	999,000	Non-bias representation of templates for genome and seq-based applications	High error rates compared with other reversible terminator chemistries	Seq-based methods	91
Pacific Biosciences (target release: 2010)	Frag only/ single molecule	Real-time	964*	N/A	N/A	N/A	Has the greatest potential for reads exceeding 1 kb	Highest error rates compared with other NGS chemistries	Full-length transcriptome sequencing; complements other resequencing efforts in discovering large structural variants and haplotype blocks	S. Turner, pers. comm.

*Average read-lengths. ‡Fragment run. §Mate-pair run. Frag, fragment; GA, Genome Analyzer; GS, Genome Sequencer; MP, mate-pair; N/A, not available; NGS, next-generation sequencing; PS, pyrosequencing; RT, reversible terminator; SBL, sequencing by ligation; SOLiD, support oligonucleotide ligation detection.

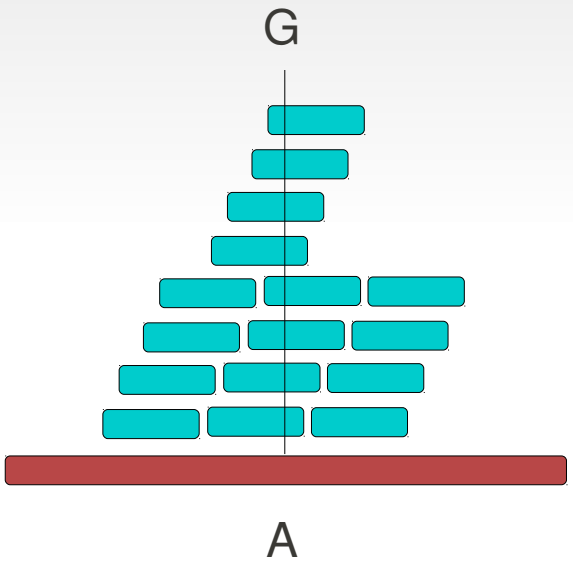
NGS

Ensamblaje



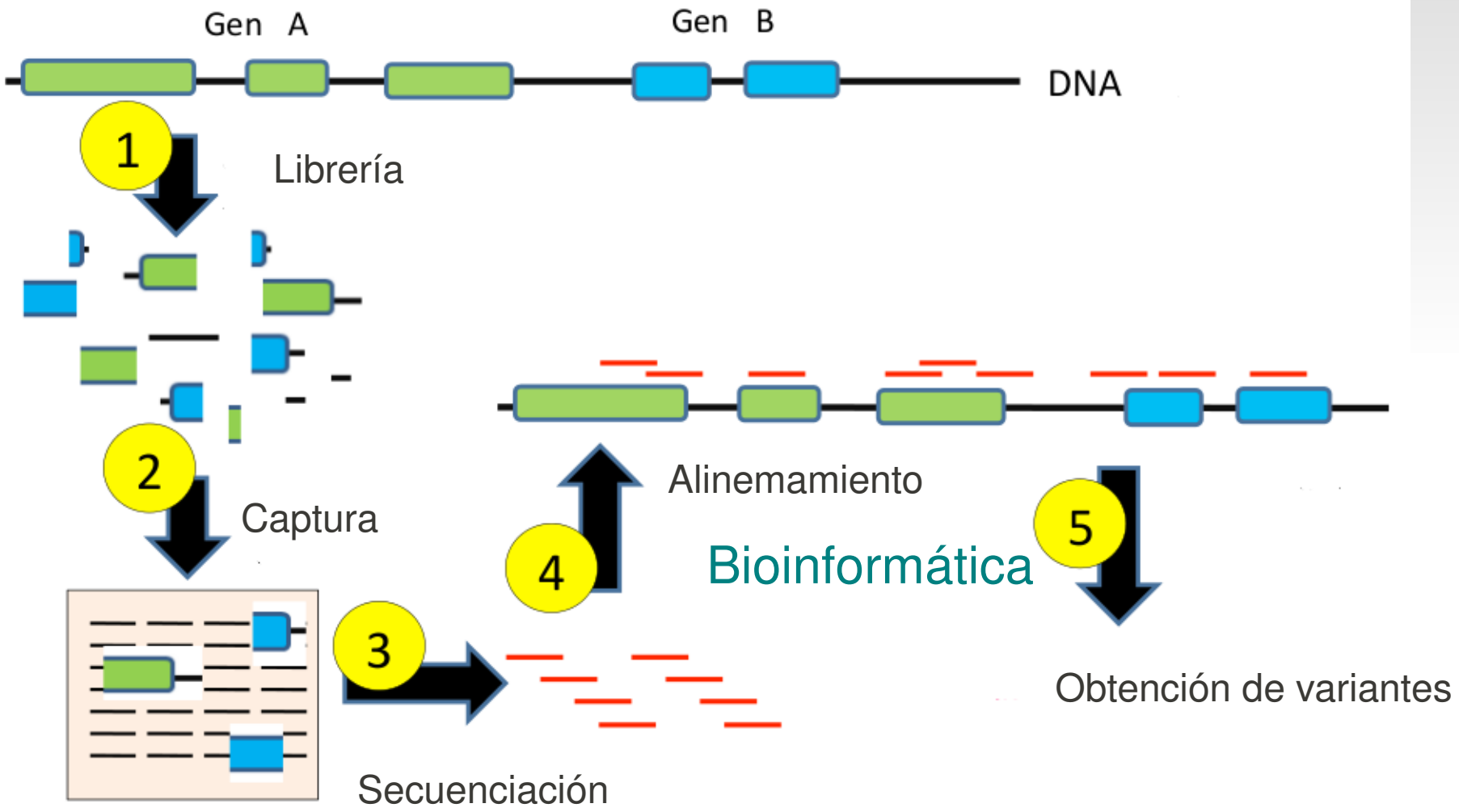
454-Pirosecuenciación
Lecturas largas
Secuenciación de genomas

Alineamiento

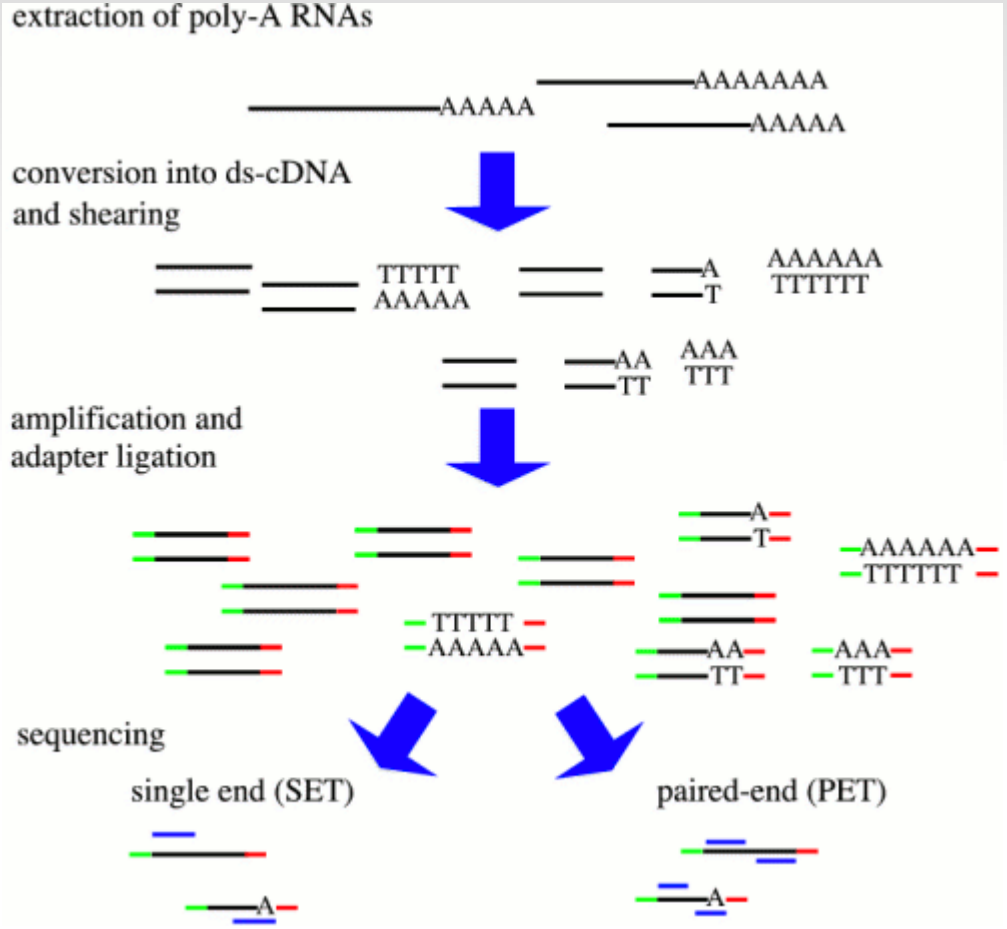


SOLiD - Illumina
Lecturas cortas
Resecuenciación ADN

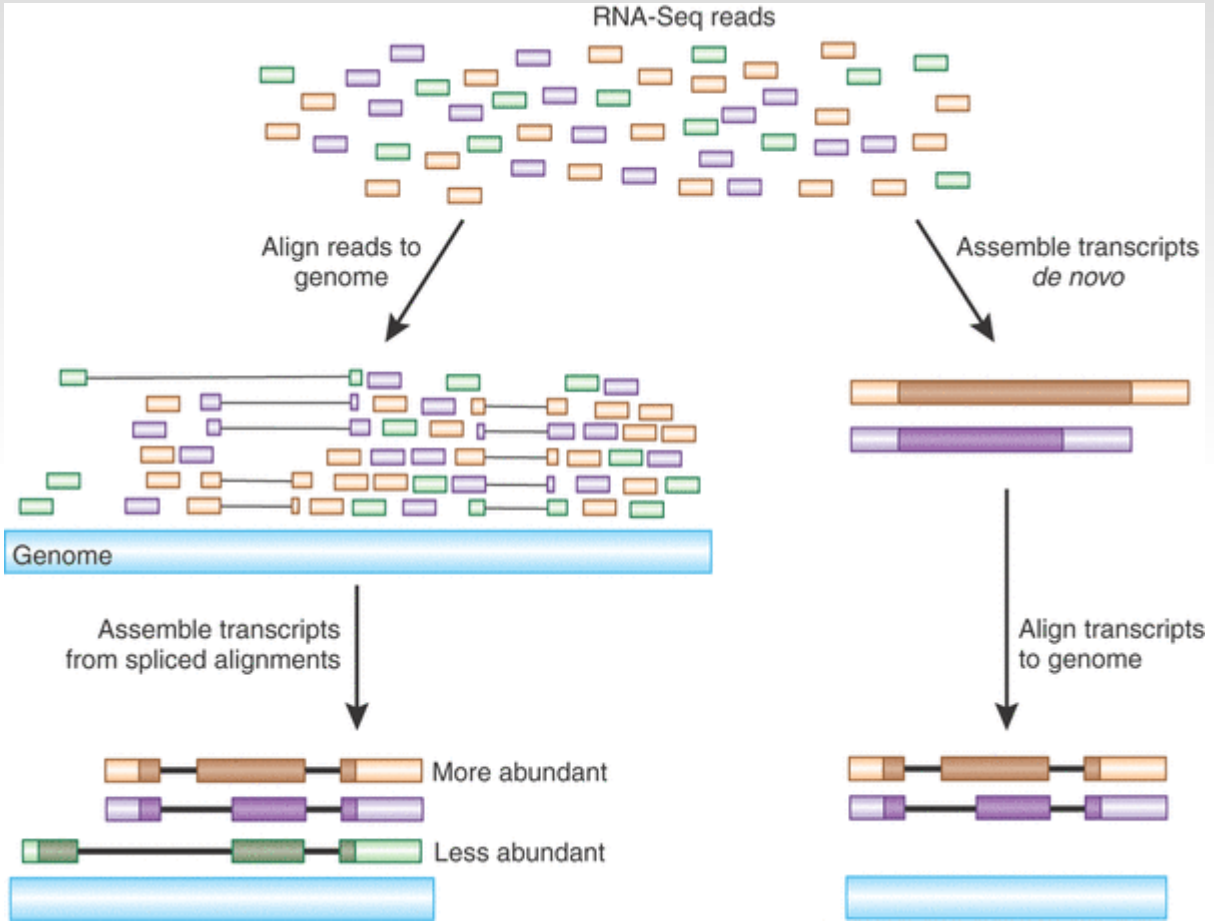
Exoma


Gen A
Gen B
DNA
1
Librería
2
Captura
3
Secuenciación
4
Alinemamiento
Bioinformática
5
Obtención de variantes

RNA-Seq


extraction of poly-A RNAs
AAAAAAA
AAAAA
AAAAA
conversion into ds-cDNA
and shearing
TTTTT
AAAAA
A
T
AAAAAA
TTTTTT
AA
TT
AAA
TTT
amplification and
adapter ligation
A
T
AAAAAA
TTTTTT
TTTTT
AAAAA
AA
TT
AAA
TTT
sequencing
single end (SET)
paired-end (PET)
A
A

RNA-Seq


RNA-Seq reads
Align reads to genome
Assemble transcripts de novo
Genome
Assemble transcripts from spliced alignments
Align transcripts to genome
More abundant
Less abundant

RNA-Seq

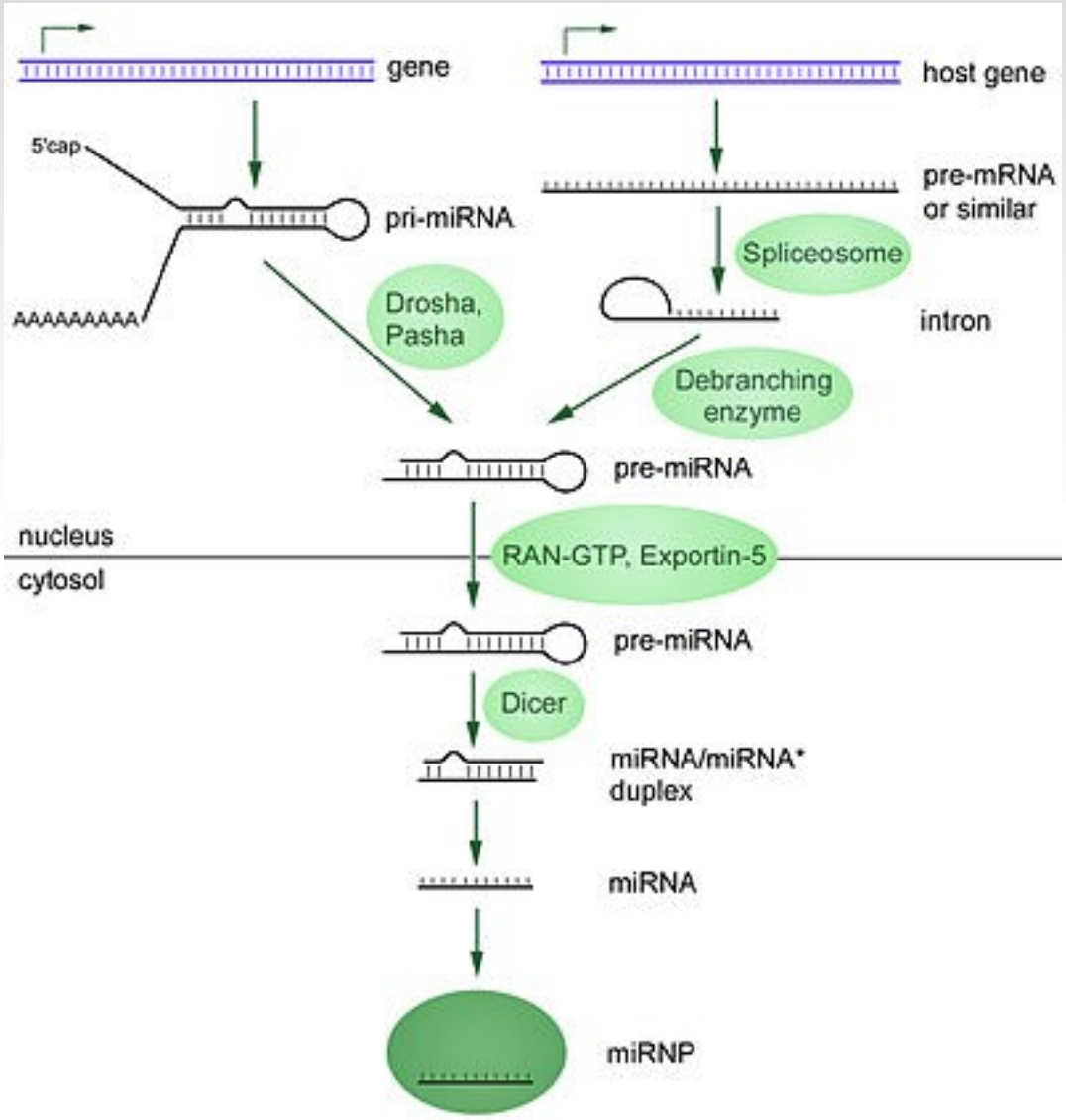
RNA-Seq

- detección de transcrito no dirigida
- no necesita genoma de referencia
- específico de cadena
- descubrimiento de splicing
- detecta variaciones en el ADN
- detecta transcritos raros

Microarrays de ARN

- restringido a las sondas en el diseño
- necesita conocer el genoma
- no es específico de cadena
- difícil trabajar con exones
- no aporta información sobre la secuencia de los transcritos
- difícil de detectar transcritos poco raros
- técnica y análisis muy optimizado

microRNAs



Mecanismos de actuación:

- degradación de la proteína durante la traducción
- inhibición de la elongación de la traducción
- terminación prematura de la traducción (disgregación de los ribosomas)
- inhibición de la iniciación de la traducción

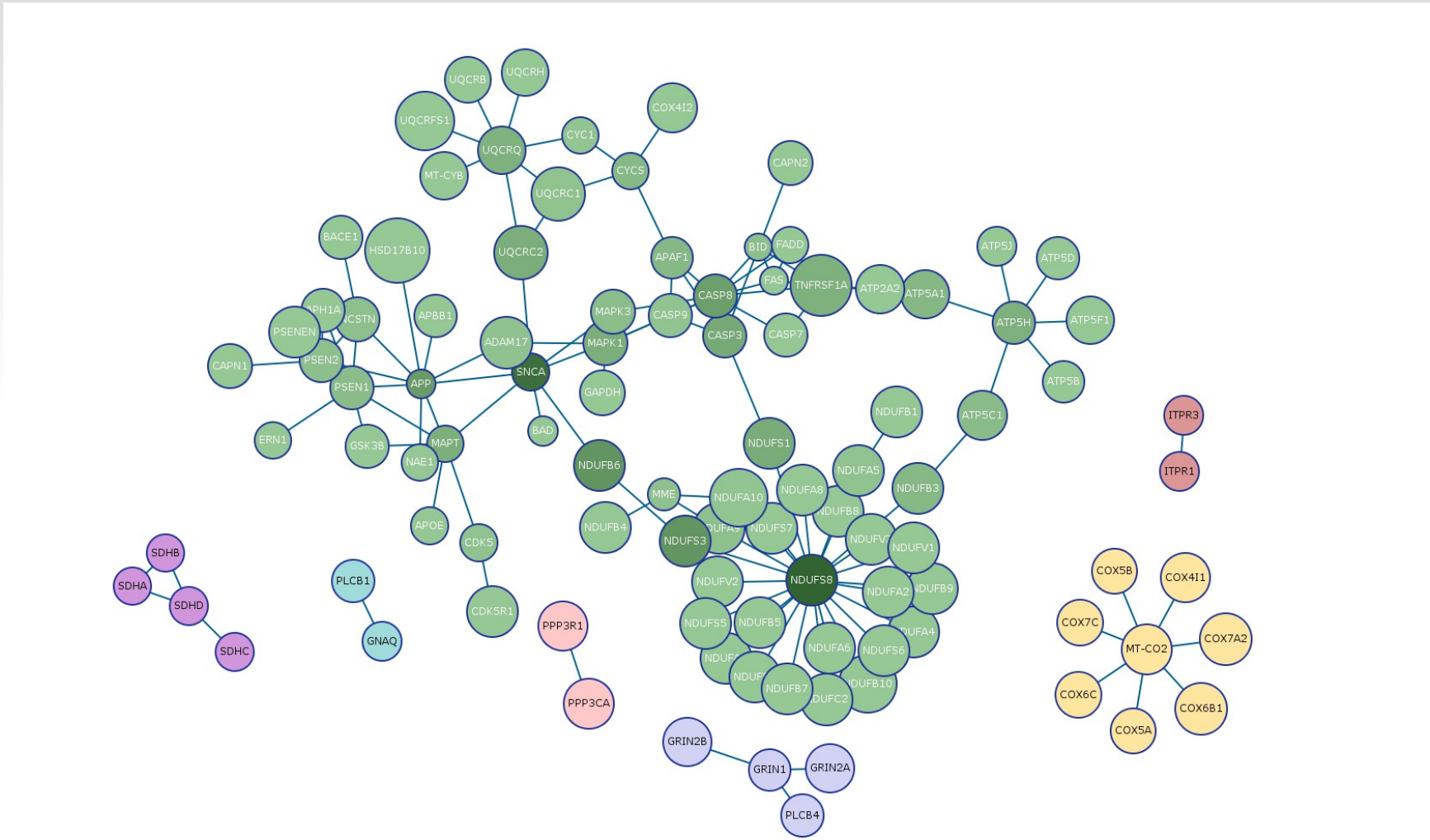
Técnicas de detección:

- RT-PCR + qRT-PCR
- Microarrays de expresión
- NGS

Interacciones proteína-proteína

- Técnicas de detección moleculares
  - Doble híbrido
  - Espectrofotometría de masas
- Técnicas de alto rendimiento
  - Microarrays de expresión

Interacciones proteína-proteína



Anotación genómica

- Identificación de elementos del genoma
- Anotación estructural: búsqueda de genes y otros elementos en el genoma
    - Búsqueda de secuencias de traducción
    - Búsqueda de dominios
    - Comparación con bases de datos: BLAST

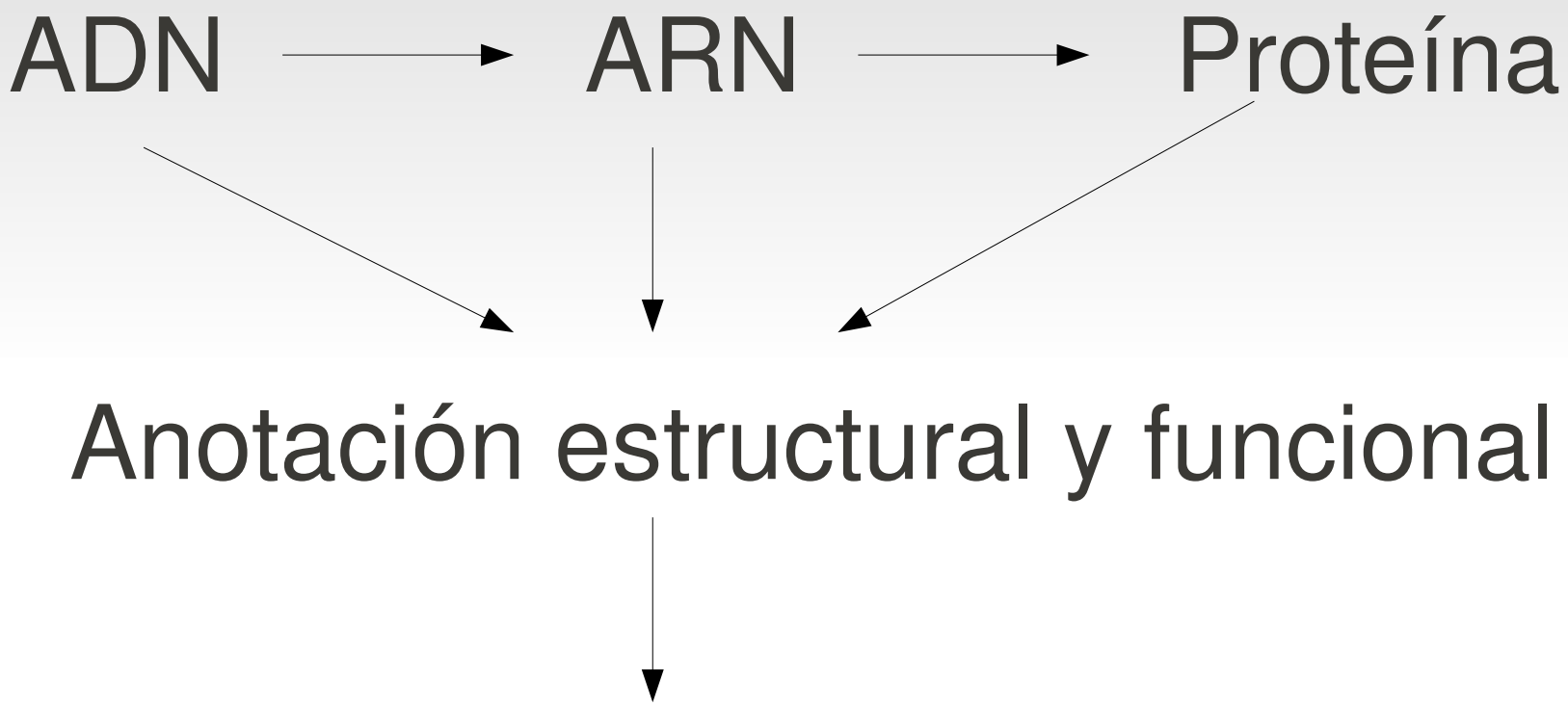
Anotación genómica

- Una vez se han encontrado los posibles elementos, se procede a la anotación funcional:
  - Definir posibles funciones
  - Comparar con bases de datos
  - Anotar con bases de datos (Gene Ontology)
- Ejemplo de grandes proyectos:
  - ENCODE (http://www.genome.gov/10005107)
  - ENSEMBL (http://www.ensembl.org)

Bases de datos

- Toda esta información se recopila en bases de datos:
  - GenBank (http://www.ncbi.nlm.nih.gov/genbank/)
  - dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/)
  - Reactome (http://www.reactome.org)
  - PDB (http://www.rcsb.org)
  - Uniprot (http://www.uniprot.org)
  - Etc.

Resumen


ADN
ARN
Proteína
Anotación estructural y funcional
Bases de datos

Gracias por vuestra atención!!

¿Preguntas?