

2	OCT, VIE	● 2 – 5pm	IB - Unidad 3.1
9	OCT, VIE	● 2 – 5pm	IB - Unidad 3.2
16	OCT, VIE	● 2 – 5pm	IB - 1° Parcial
23	OCT, VIE	● 2 – 5pm	IB - Unidad 4.1
30	OCT, VIE	● 2 – 5pm	IB - Unidad 4.2; 4.3; 4.4
6	NOV, VIE	● 2 – 5pm	IB - Unidad 5.1
		● 2 – 5pm	IB - Unidad 5.2
13	NOV, VIE	● 2 – 5pm	IB - Unidad 5.3
20	NOV, VIE	● 2 – 5pm	IB - 2° Parcial
27	NOV, VIE	● 2 – 5pm	IB - Recuperatorio

### **Unidad 3:** Base de datos y Fuentes de Datos

3.1. Bioinformática, biología y biología computacional. Organización de la información, almacenamiento de la información y tipos de datos (*Big Data*).

3.2. Generación de datos: Metodología experimental para el estudio de las diferentes moléculas: experimentos clásicos y a gran escala.

## \*. Definición para el “público”:

La bioinformática es la aplicación de la computación y las matemáticas a la gestión, el análisis y la comprensión de los datos para resolver preguntas/cuestiones biológicas relacionadas con la medicina, quemo-, neuro-, etc, informática.

- ***La bioinformática es la aplicación de la computación y las matemáticas...***

ver la definición canónica

- ***...a la gestión, el análisis y la comprensión de datos...***

La bioinformática se ocupa de todo el proceso desde la captura inicial de los datos, su administración en bases de datos, el análisis de estos datos y la formulación de los resultados en un contexto que da como resultado una nueva comprensión genuina. El objetivo está impulsado por la cantidad y calidad de los datos.

- ***...para resolver preguntas/cuestiones biológicas...***

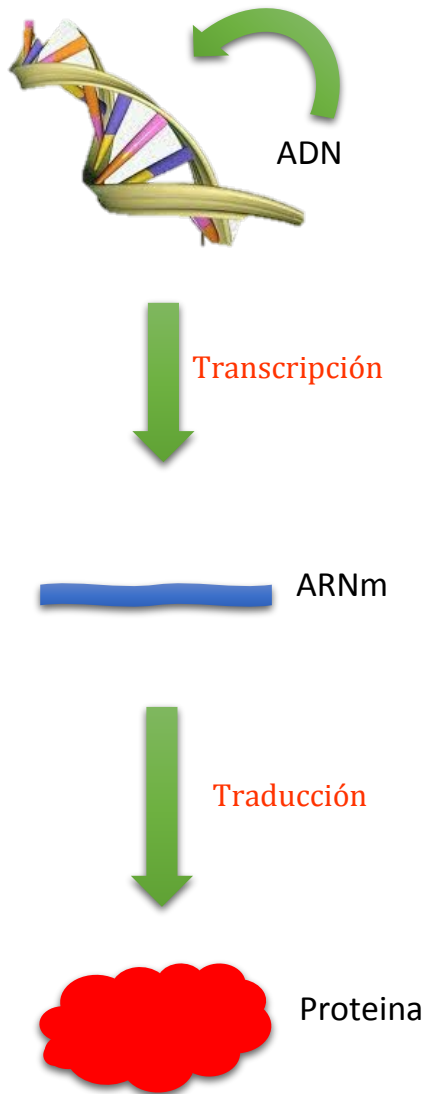
La bioinformática se aplica a problemas biológicos y no puramente cuestiones de informática

- ***...relacionadas con la medicina, quemo-, neuro-, etc, informática.***

Los desarrollos matemáticos e informáticos que se han aplicado a datos en otras áreas del conocimiento también pueden ser útiles en bioinformática.

# Biología, Bioinformática y Biología Computacional

La biología computacional es a veces definida como sinónimo de Bioinformática y a veces como una disciplina emparentada, pero distinta, de esta. El [NIH](#) define a ambas disciplinas como distintas aunque con cierto grado de solapamiento, según esta definición la bioinformática está más relacionada con el desarrollo de herramientas computacionales con el fin de analizar y procesar datos y la biología computacional con el estudio por medios computacionales de sistemas biológicos .



**BIG**

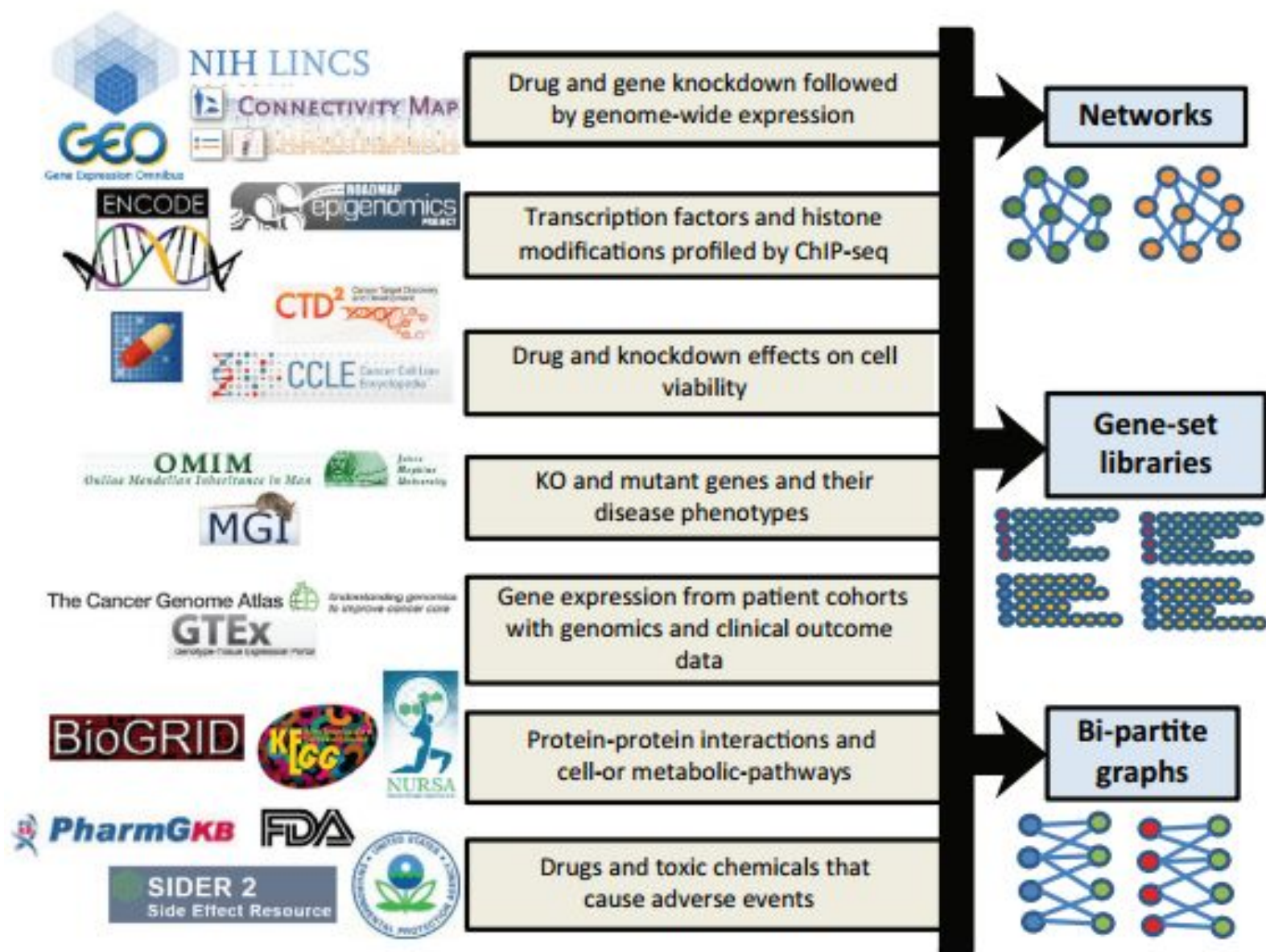
**DATA**

## Contents [hide]

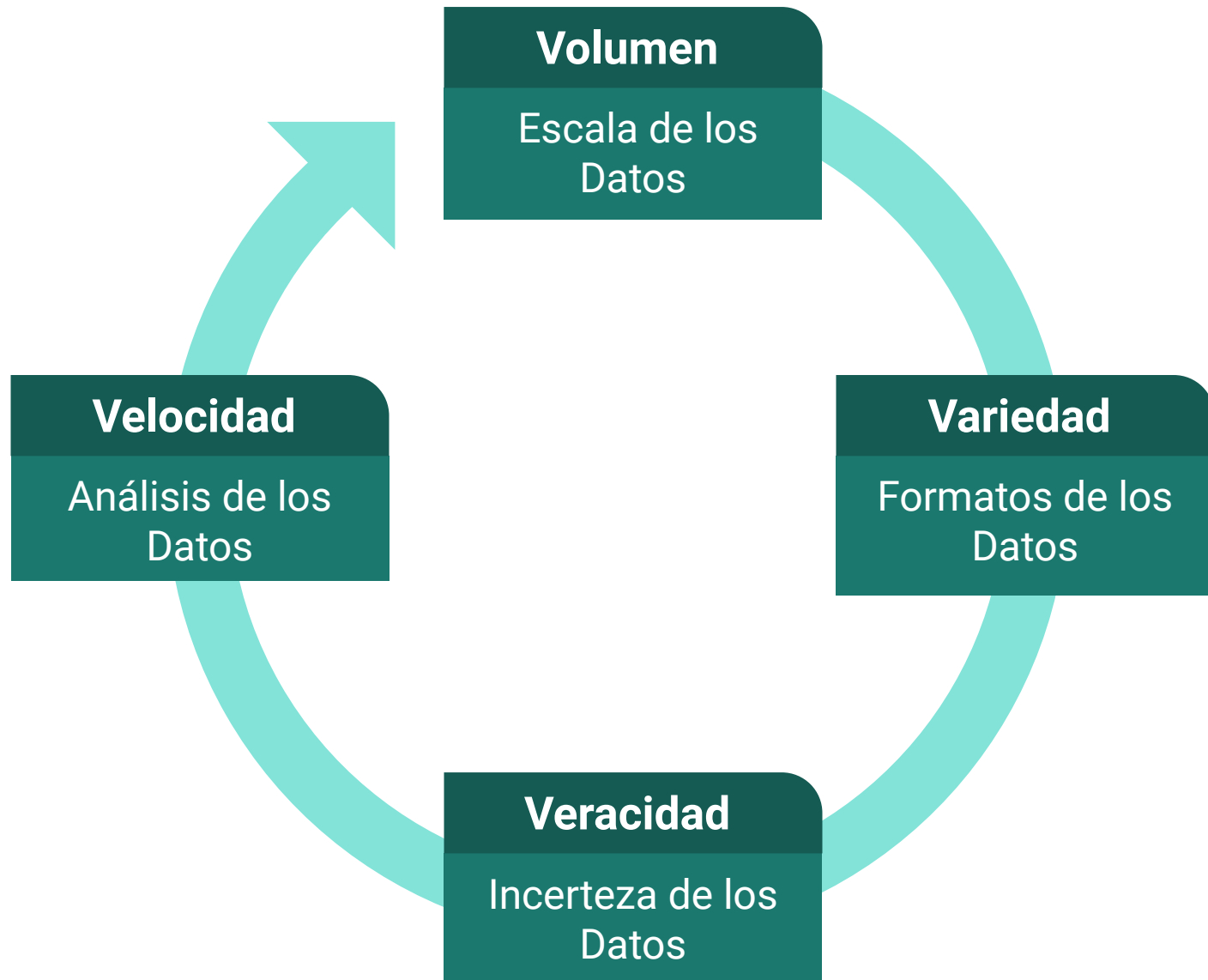
- 1 Meta databases
- 2 Model organism databases
- 3 Nucleic acid databases
  - 3.1 DNA databases
  - 3.2 Gene expression databases (mostly microarray data)
  - 3.3 Phenotype databases
  - 3.4 RNA databases
- 4 Amino acid / protein databases
  - 4.1 Protein sequence databases
  - 4.2 Protein structure databases
  - 4.3 Protein model databases
  - 4.4 Protein-protein and other molecular interactions
  - 4.5 Protein expression databases
- 5 Signal transduction pathway databases
- 6 Metabolic pathway and protein function databases
- 7 Additional databases
  - 7.1 Exosomal databases
  - 7.2 Mathematical model databases
  - 7.3 Taxonomic databases
  - 7.4 Radiologic databases
  - 7.5 Antimicrobial resistance databases
- 8 Wiki-style databases
- 9 Specialized databases
- 10 References
- 11 External links







# BIG DATA $\Rightarrow$ 4 V





### Tema 3.1

#### Biología, Bioinformática y Biología Computacional

1. Existen controversias y similitudes entre las definiciones de Bioinformática y de Biología Computacional que puede ser evidenciada en la siguiente definición:

Lea cuidadosamente los textos de las siguientes páginas webs y discuta con sus compañeros si hay o no diferencia entre Bioinformática y Biología Computacional.

¿Deberían fusionarse los términos y por consiguiente su definición?

Webs:

\*. <http://www.bioinformaticos.com.ar/biologia-computacional-en-argentina/>

\*.

<https://biology.meta.stackexchange.com/questions/168/merging-bioinformatics-and-computational-biology-tags>

\*.

<https://respuestas.me/q/La-biolog-a-computacional-es-diferente-de-la-bioinform-tica-34039920924>

\*.

<https://rubenyciencia.wordpress.com/tag/diferencias-entre-bioinformatica-y-biologia-computacional/>

\*. <http://www.euskonews.com/0334zbnk/gaia33402es.html>

**2. Explique brevemente cómo se aplican las 4V del “Big Data” a las siguientes Bases de Datos Biológicas. Para ello, indague sobre el tipo de datos, estructura de datos, curado de los datos, etc.**

Desarrolle las 4V para:

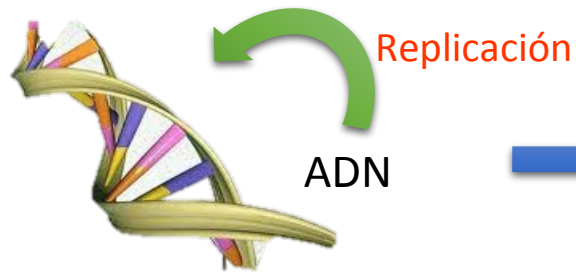
- \*. GeneBank: <https://www.ncbi.nlm.nih.gov/genbank/>
- \*. RNACentral: <https://rnacentral.org/>
- \*. Uniprot: <https://www.uniprot.org/>
- \*. KEGG: <https://www.genome.jp/kegg/>
- \*. PubMed: <https://www.ncbi.nlm.nih.gov/pubmed>

### **Unidad 3:** Base de datos y Fuentes de Datos

3.1. Bioinformática, biología y biología computacional. Organización de la información, almacenamiento de la información y tipos de datos (*Big Data*).

3.2. Generación de datos: Metodología experimental para el estudio de las diferentes moléculas: experimentos clásicos y a gran escala.

# DOGMA CENTRAL



ARNm



Proteína

Southern blot  
y  
Secuenciamiento  
de ADN (método de Sanger)

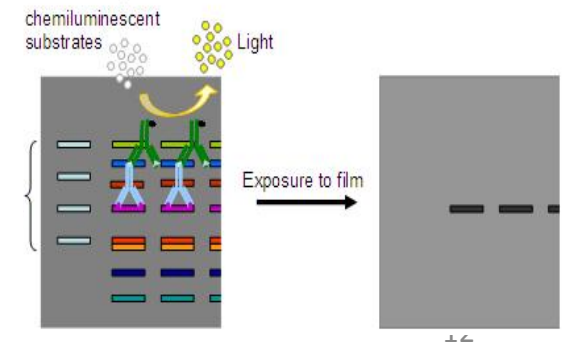
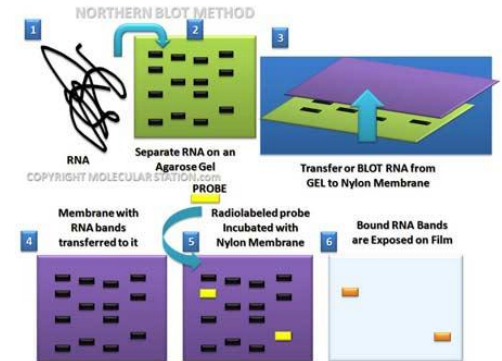
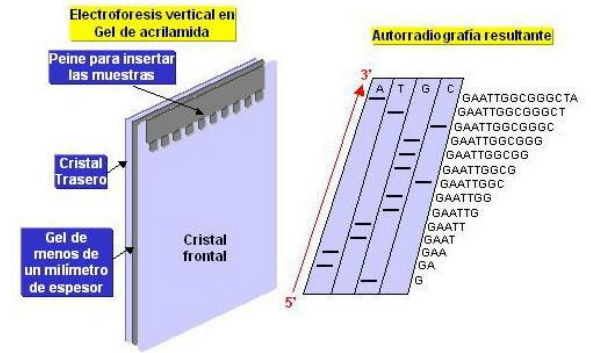
UN locus

Northern blot  
y  
PCR, qPCR

UN gen

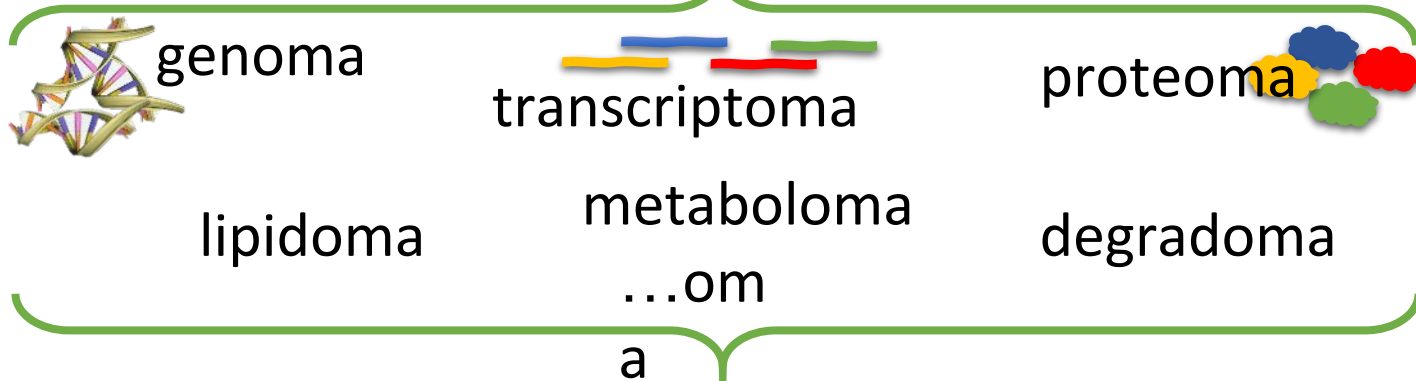
Western blot y  
sec. de Edman

UNA proteína



# La era “-ómica”/“-oma”

Es el estudio a **gran escala** de los procesos biológicos

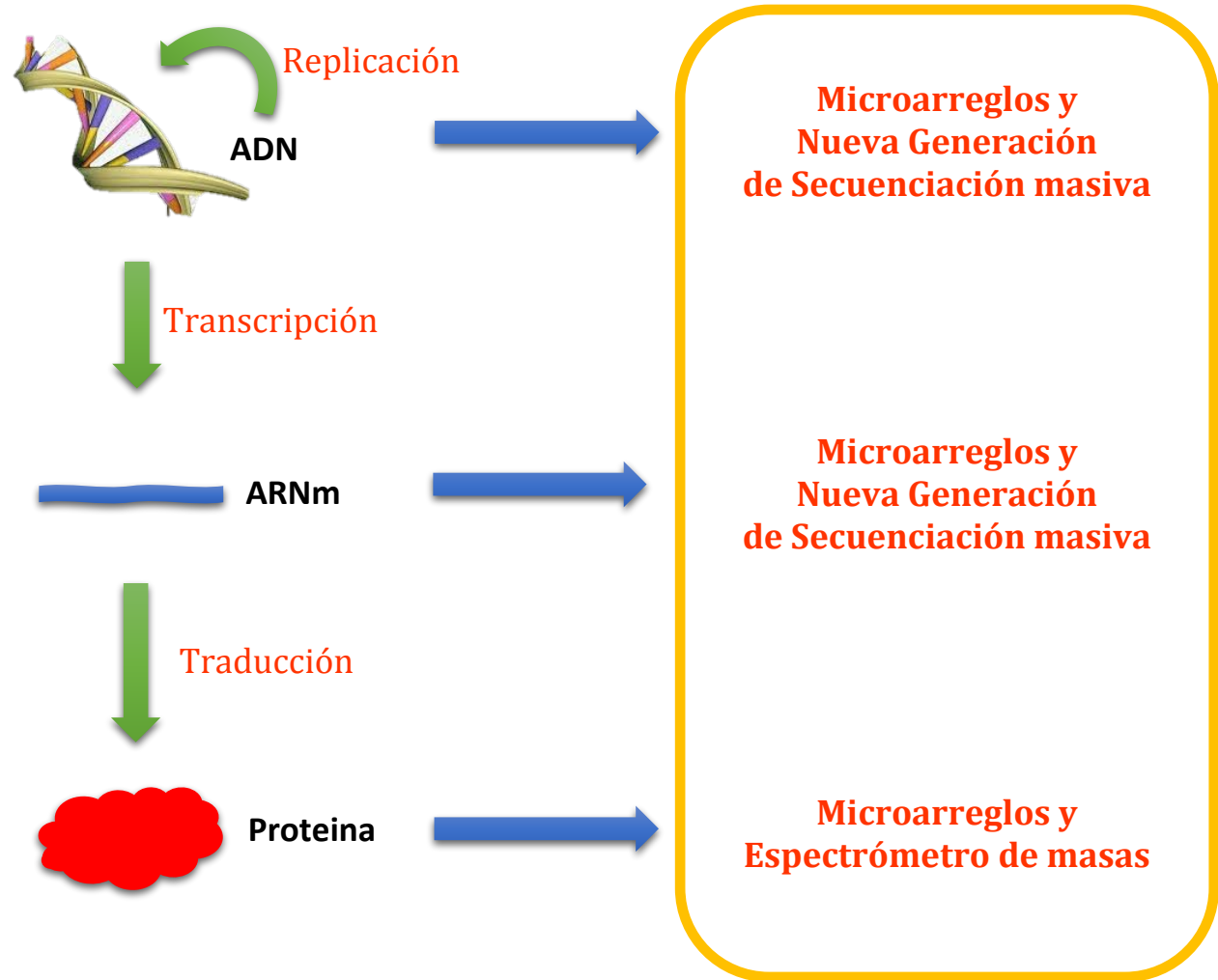


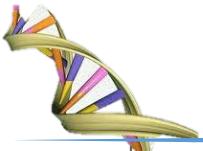


## **SYSTEM BIOLOGY**

El **análisis de datos** obtenidos por técnicas moleculares a **gran escala** comprende un **enfoque interdisciplinario** que involucra ciencias como:



**BIOINFORMÁTICA  
Y BIOLOGÍA  
COMPUTACIONAL**



	Metodología	Organismo modelo	Organismo NO modelo
 ADN	<i>Southernblot</i>	✓	✗✗✗✗✓
	Secuenciación de Sanger	✓	✓
	<i>Microarrays</i>	✓	✗✗✗✗✓
	NGS	✓	✓
 ARNm	<i>Northernblot</i>	✓	✗✗✗✗✓
	PCR, qPCR	✓	✗✗✗✗✓
	<i>Microarrays</i>	✓	✗✗✗✗✓
	NGS	✓	✓
 Proteína	<i>Western blot</i>	✓	✗✗✗✗✓
	Sec. Edman	✓	✓
	<i>Microarrays</i>	✓	✗✗✗✗✓
	2D-Espect. De Masas	✓	✓



## **Tema 3.1**

**1. Busque y desarrolle brevemente metodologías y técnicas utilizadas en los laboratorios, tanto a escala simple como a gran escala, para el estudio de:**

- \*. Ácidos Nucléicos.

- \*. Proteínas

- \*. Metabolitos

# Historia de la Secuenciación de ADN

Efficiency  
(bp/person/year)

1

15

150

1,500

15,000

25,000

50,000

200,000

50,000,000

100,000,000,000

500,000,000,000

1870

Miescher: Discovers DNA

1940

Avery: Proposes DNA as 'Genetic Material'

1953

Watson & Crick: Double Helix Structure of DNA

1965

Holley: Sequences Yeast tRNA<sup>Ala</sup>

1968

10 consecutives bases in a DNA strand

1970

Wu: Sequences  $\lambda$  Cohesive End DNA

1977

Sanger: Dideoxy Chain Termination  
Maxam-Gilbert: Chemical Degradation

1980

Messing: M13 Cloning

1986

Hood et al.: Partial Automation

1990

- Cycle Sequencing
- Improved Sequencing Enzymes
- Improved Fluorescent Detection Schemes
- US Human Genome Project and Celera

2002

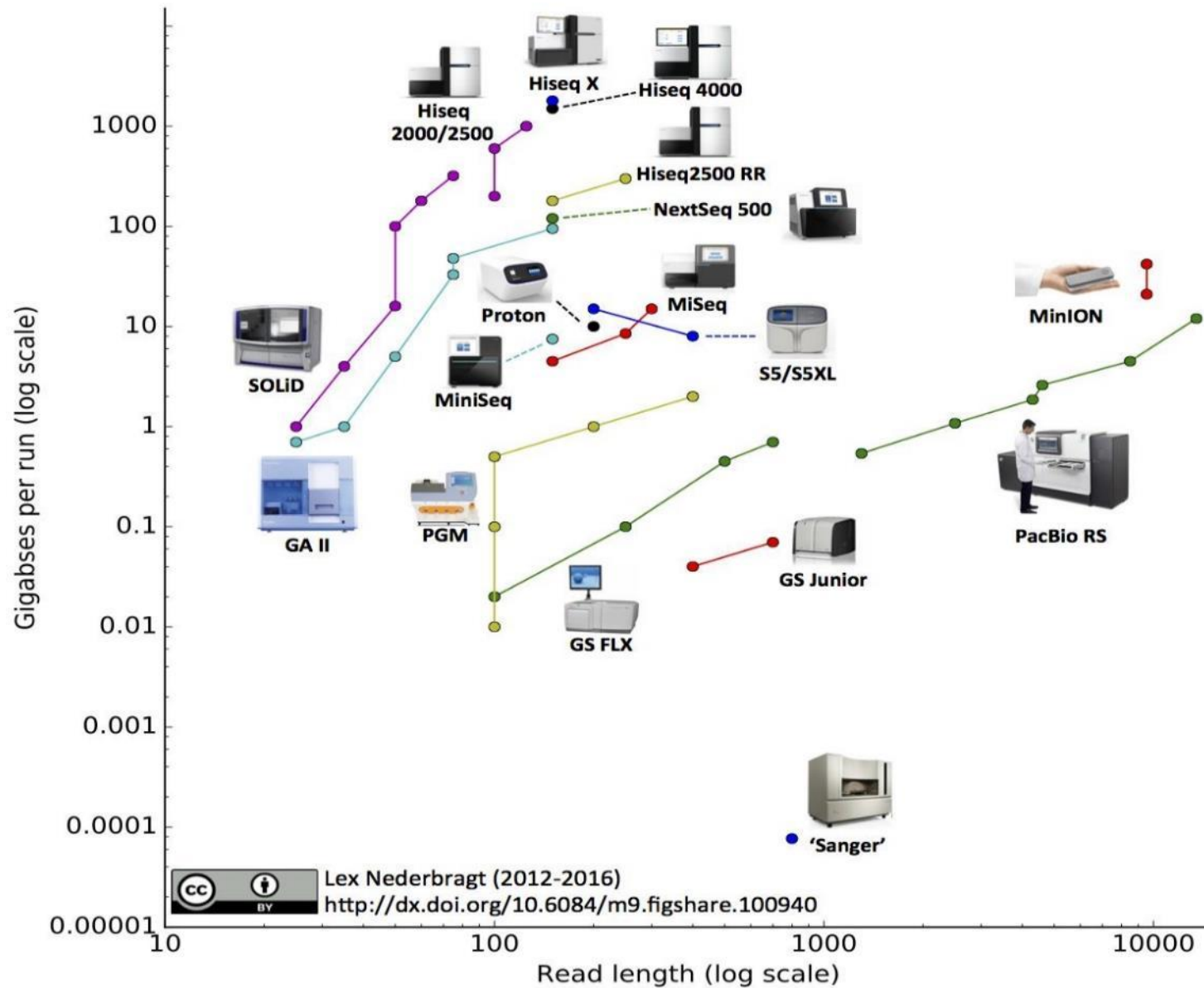
2007

- 2<sup>nd</sup> Next Generation Sequencing
- Improved enzymes and chemistry
- Improved image processing

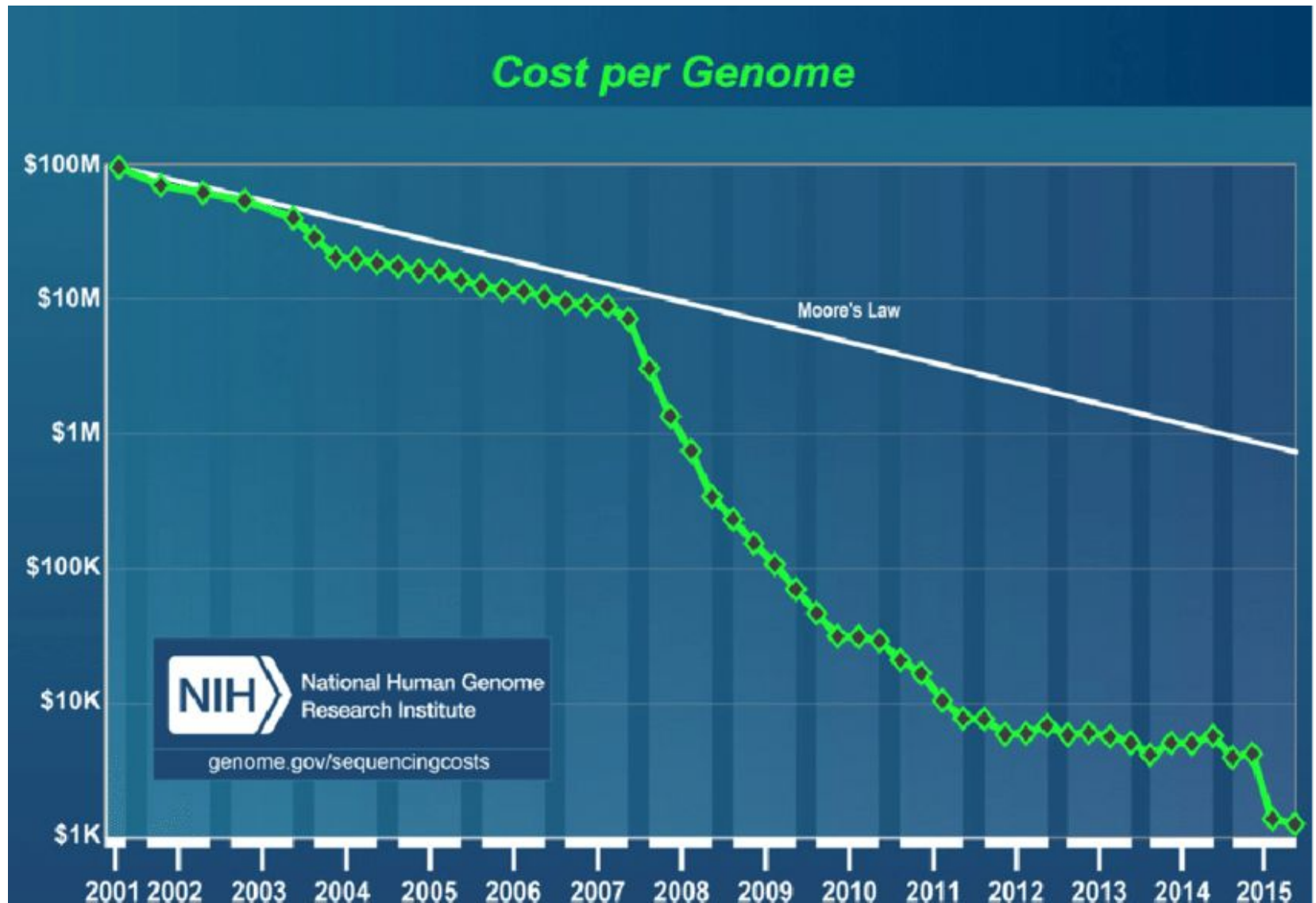
2013

- 3<sup>rd</sup> Next Generation Sequencing (SMS)

# Developments in high throughput sequencing



# ¿Cuánto cuesta secuenciar un genoma?



**\$3,000,000,000**

**2003** Human Genome Project



**\$20,000,000**

**2006** 1<sup>st</sup> individual genome



**\$2,000,000**

**2007** 1<sup>st</sup> NGS Genome



**\$200,000**

**2008** 1<sup>st</sup> 30x genome



**\$10,000**

**2010** 1<sup>st</sup> sub-10K genome



**\$1,000**

**2014** 1<sup>st</sup> \$1,000 genome



**\$100**

**2017** 1<sup>st</sup> \$100 genome

