



BIA

Plataforma
Bioinformática
Argentina

Adrian Turjanski
Director de la Plataforma de Bioinformática
Investigador CONICET, Profesor FCEyN, UBA.

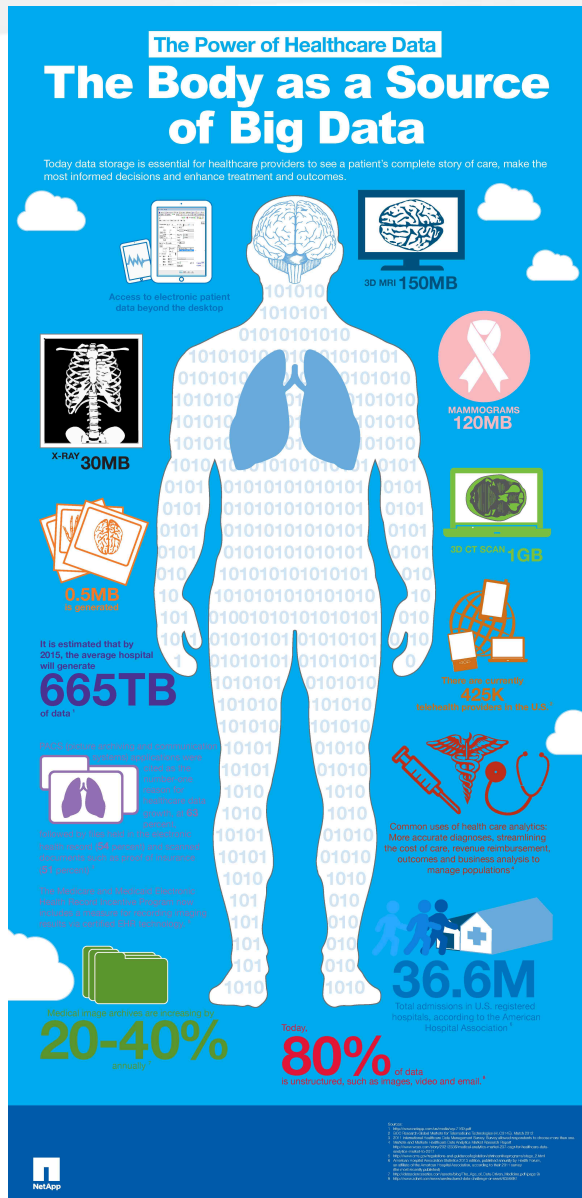
HPC DAY 2013, CORDOBA, ARGENTINA.

Sobre Genómica, Bioinformática y Salud...

Por que cuando hablamos de Genómica
Hablamos de Big-Data?

Por que es TAN importante el “CALCULO
INTENSIVO” en esta área?

Sobre Genómica, Bioinformática, Salud y BIG DATA...



HUMAN GENOME, IN NUMBERS



6 billion DNA LETTERS

A T C G T A

22,000 GENES

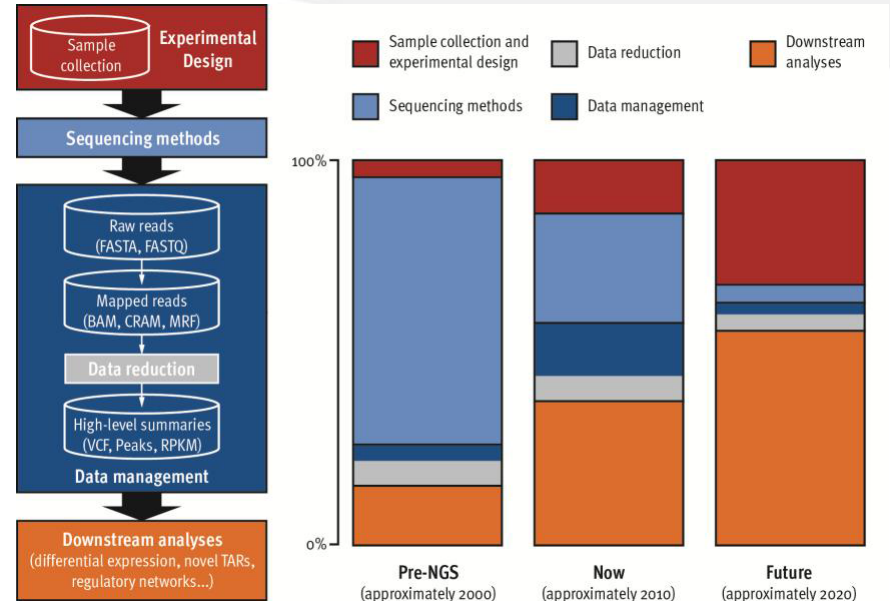
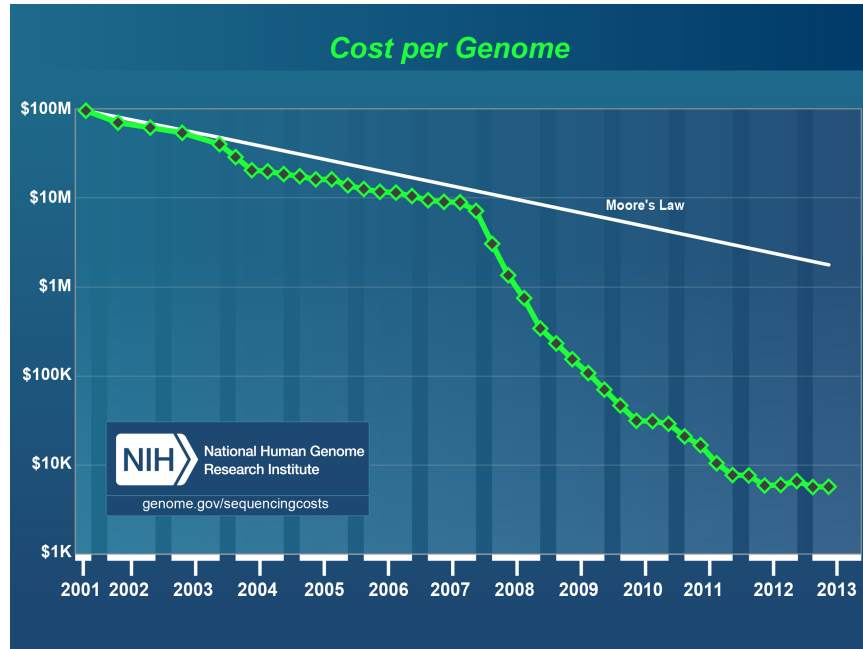
46 CHROMOSOMES

\$9,500 COST TO SEQUENCE

500GB SIZE ON DISK

Sources: NIH, Illumina

Revolución en costos!



Secuenciar es un commodity pero el análisis no!

Proyecto Genoma Humano (2003)

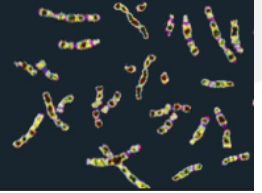


National Human
Genome Research
Institute

1000 Genomas (2008)

1000 Genomes

A Deep Catalog of Human Genetic Variation



10.000 Genomas (2013)



UK
10K

RARE GENETIC VARIANTS IN HEALTH AND DISEASE

100.000
Genomas
(2018-2020)



Genomics
england



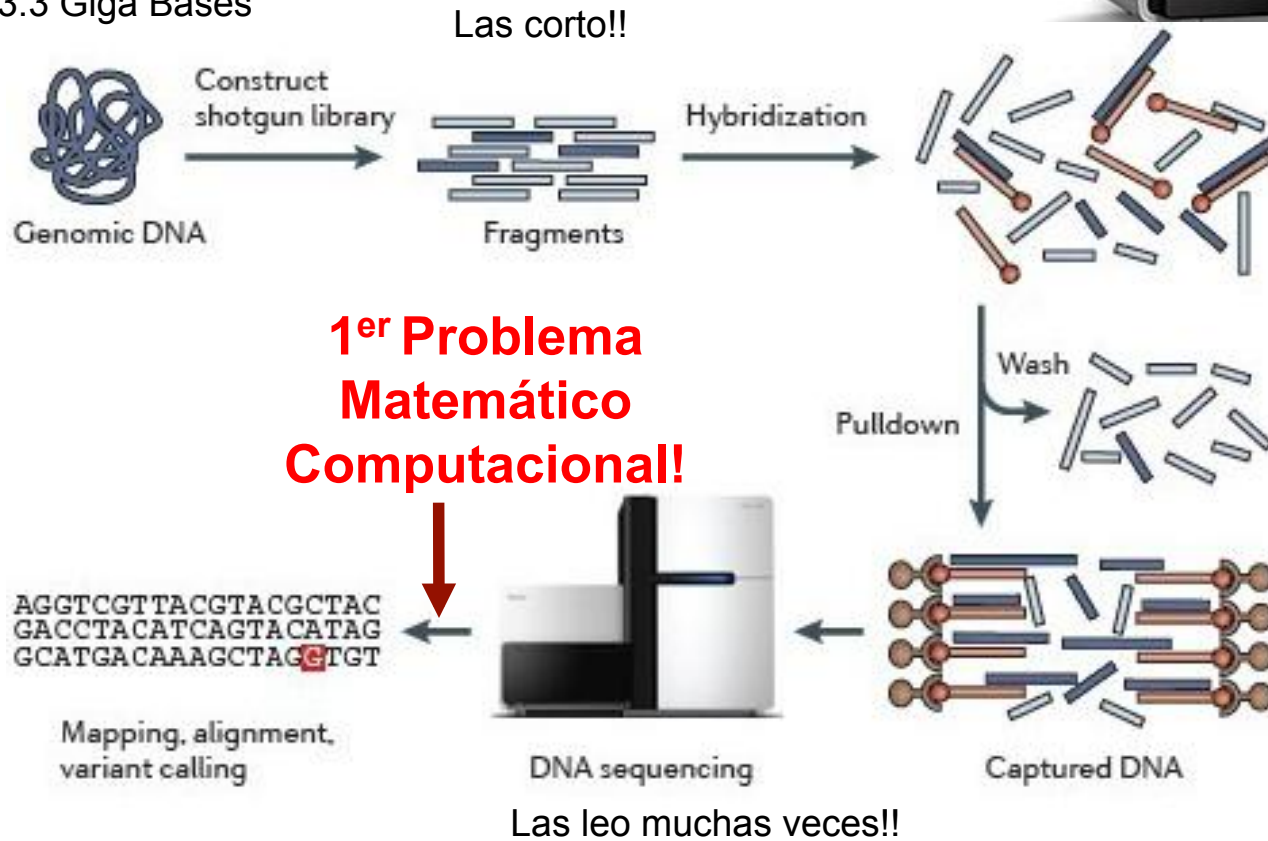
3.000.000 Genomas



华大基因
BGI

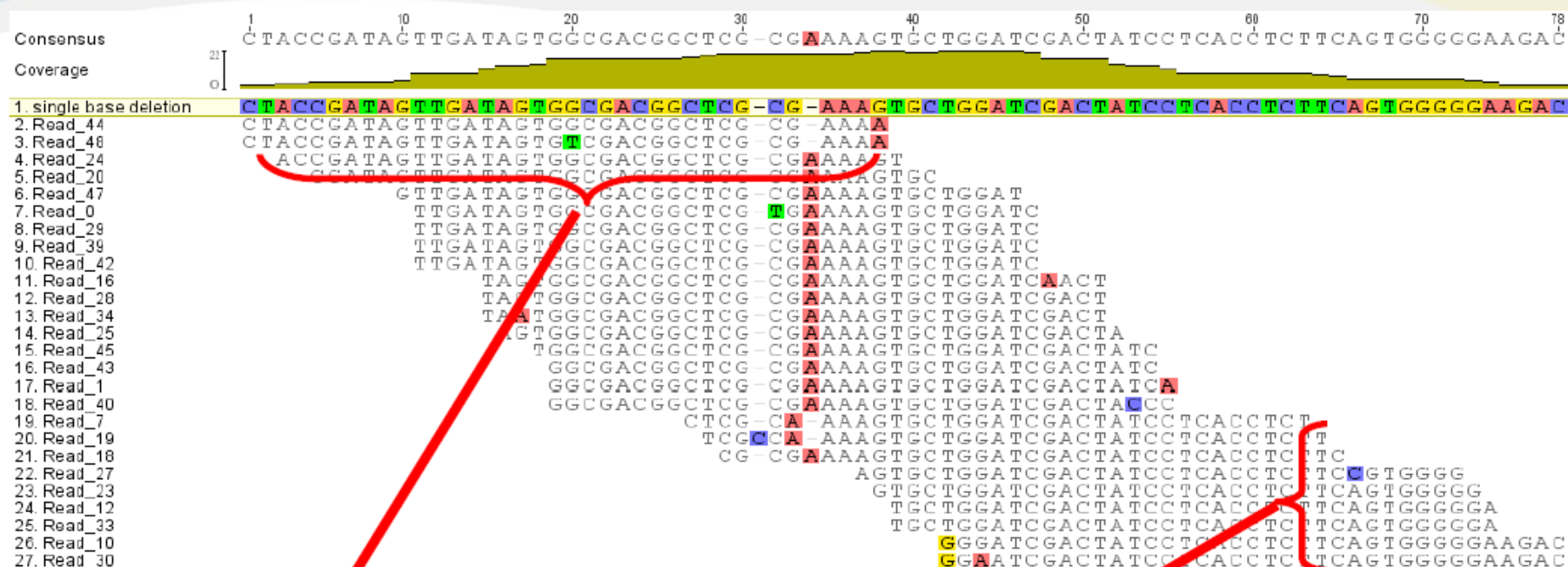
Secuenciamiento ADN

3.3 Giga Bases



Genera un máximo de 600GB por corrida!!

El problema es el ensamblado (Real):



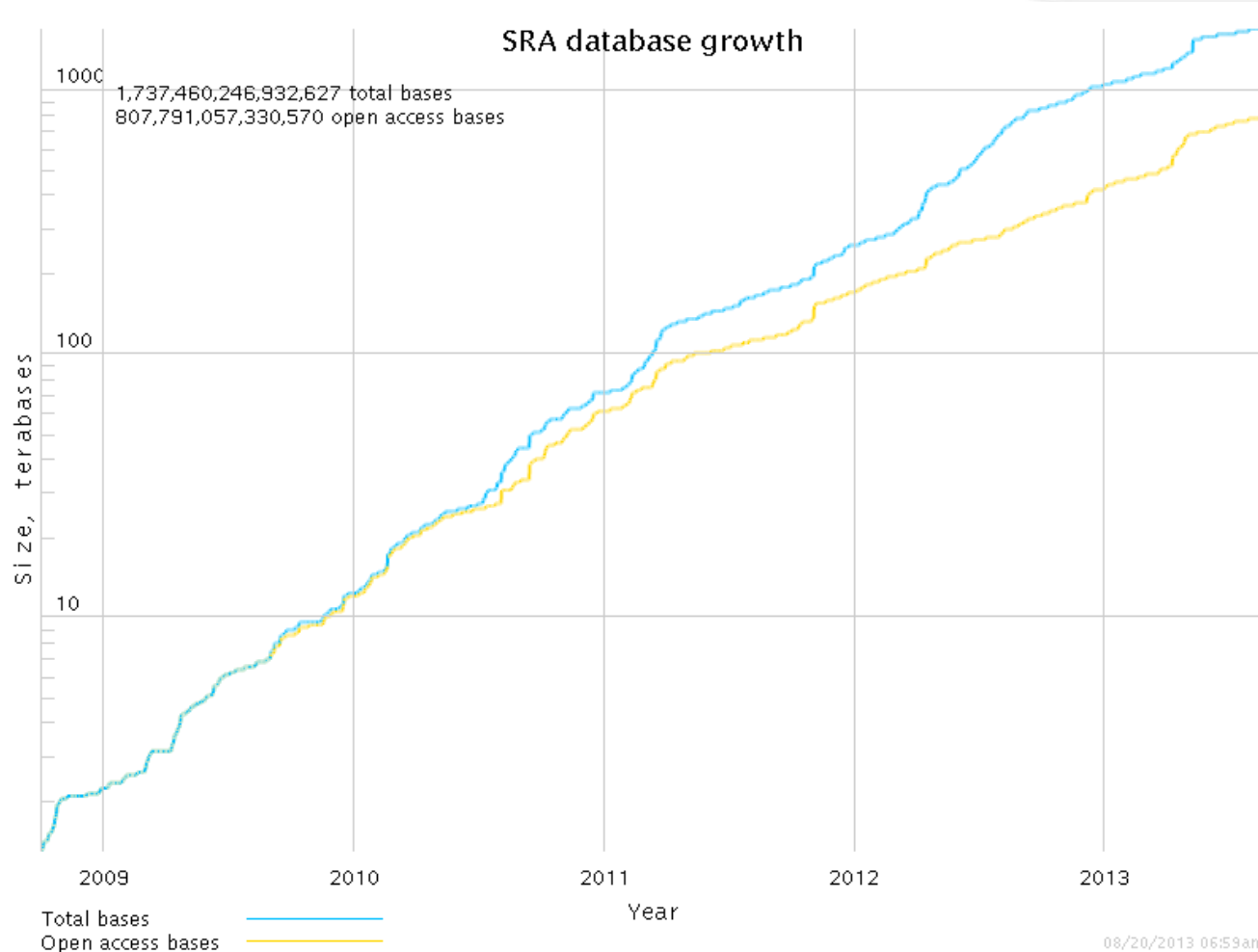
Longitud del
"Read"

Cuanto mas largos mas
fácil ensamblar!

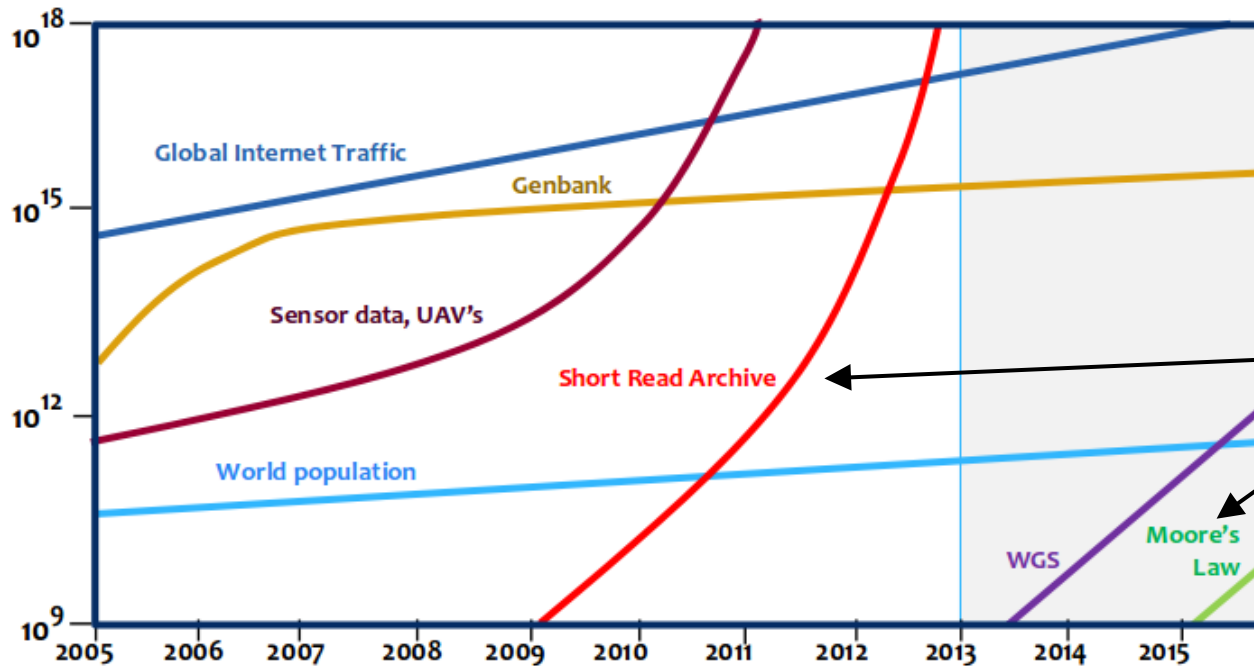
Cobertura

Mayor cobertura, mayor precisión
(menos errores)
Minimo 10X, Usual Genoma
Humano 30X

- **Sequence Read Archive** (Short Read Archive; SRA): repositorio de datos crudos obtenidos mediante secuenciación de 2da generación (NGS; 454, Illumina, IonProton).



Crecimiento de los datos de secuenciación (SRA) en comparación con otras poblaciones de datos



Los datos de secuenciación crecen en mayor medida que la velocidad de cómputo

Figure 1. Approximate growth of different data populations. Note that the 'short read archive' (SRA) has grown faster than what would be predicted by Moore's law. NGS began with a 'bang' in mid-2005 when Jonathan Rothberg's company, 454, assembled a group of cutting-edge technologies and commercialized the first next-generation sequencer. Curves as indicated by colored text above the graph and as annotated, forecasts in growth indicated by the shaded region of the graph.

- **World human population growth in numbers [2].**
- **The 'short read' archive (SRA) generated from 2nd generation NGS [3].**
- **Moore's Law - number of transistors per chip [4].**
- **Number of whole human genomes sequenced (WGS) [5].**
- **The number of entries in Genbank [6].**
- **Global internet traffic [7].**
- **Sensor data, U.S. unmanned aerial vehicle (UAV)'s, including Global Hawk and Predator**

¿Qué es Bioinformática?

La investigación, desarrollo y aplicación de herramientas y enfoques computacionales para la manipulación de datos biológicos, médicos, de comportamiento y de salud, incluyendo aquellos para obtener, almacenar, organizar, analizar y visualizar dicha información.

Fuente: NIH Working Definition of Bioinformatics and Computational Biology

Las agujas en los pajares: encontrar la variante génica relevante en todo un genoma (1 o 2 bytes en 3.2 Gb)



Mellizos con Enfermedad Genética. Cuál Gen?

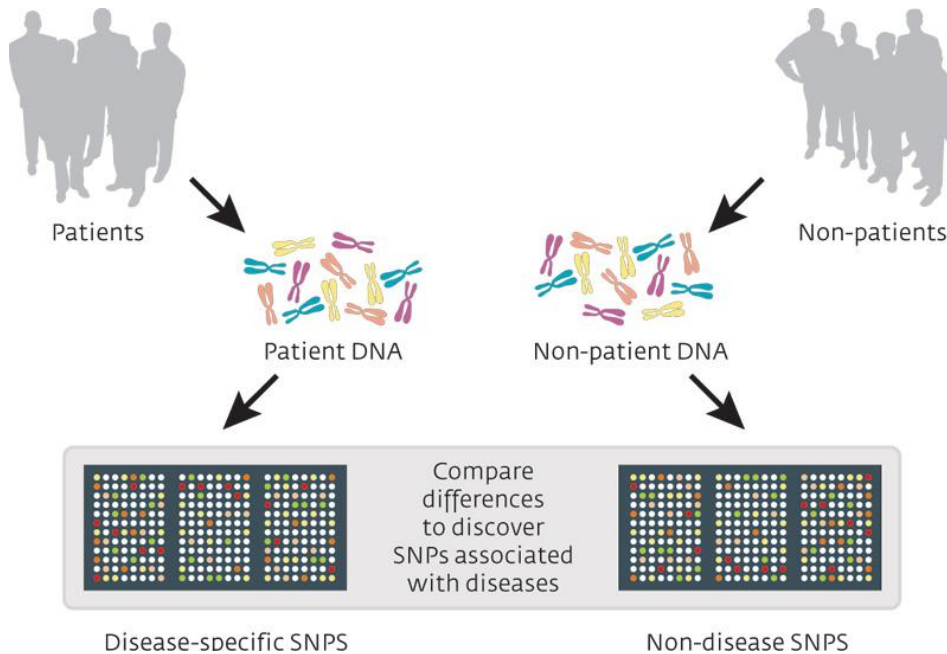
Nucleotide variants	IV-2 (male subject)	IV-3 (female subject)	Shared
All variants	2.427.038	2.504.162	1.631.770
% dbSNP129	88,7	88,1	
Variant density (bp ⁻¹)	1/1112	1/1078	
Coding	13.352	14.961	9531
Nonsynonymous	6432	7141	4605
Rare nonsynonymous	174	175	77
Candidate genes	6	9	3

dbSNP: 73.345.505 variantes registradas.

Bainbridge, Matthew N., Wojciech Wiszniewski, David R. Murdock, Jennifer Friedman, Claudia Gonzaga-Jauregui, Irene Newsham, Jeffrey G. Reid, et al. "Whole-Genome Sequencing for Optimized Patient Management." *Science Translational Medicine* 3, no. 87 (June 15, 2011): 87re3–87re3. doi:10.1126/scitranslmed.3002243.

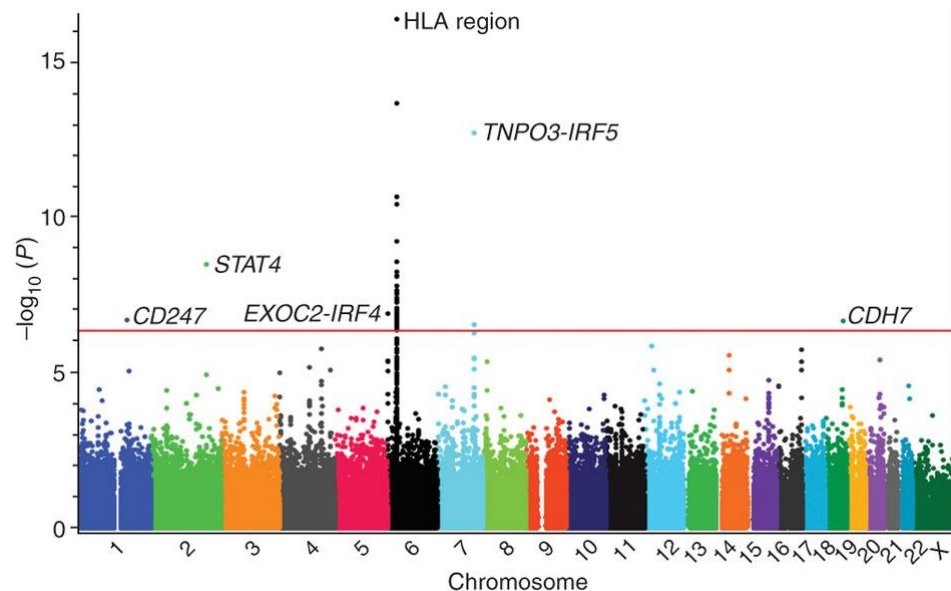
GWAS (Genome-Wide Association Study)

Enfermedades poligénicas(Cancer)



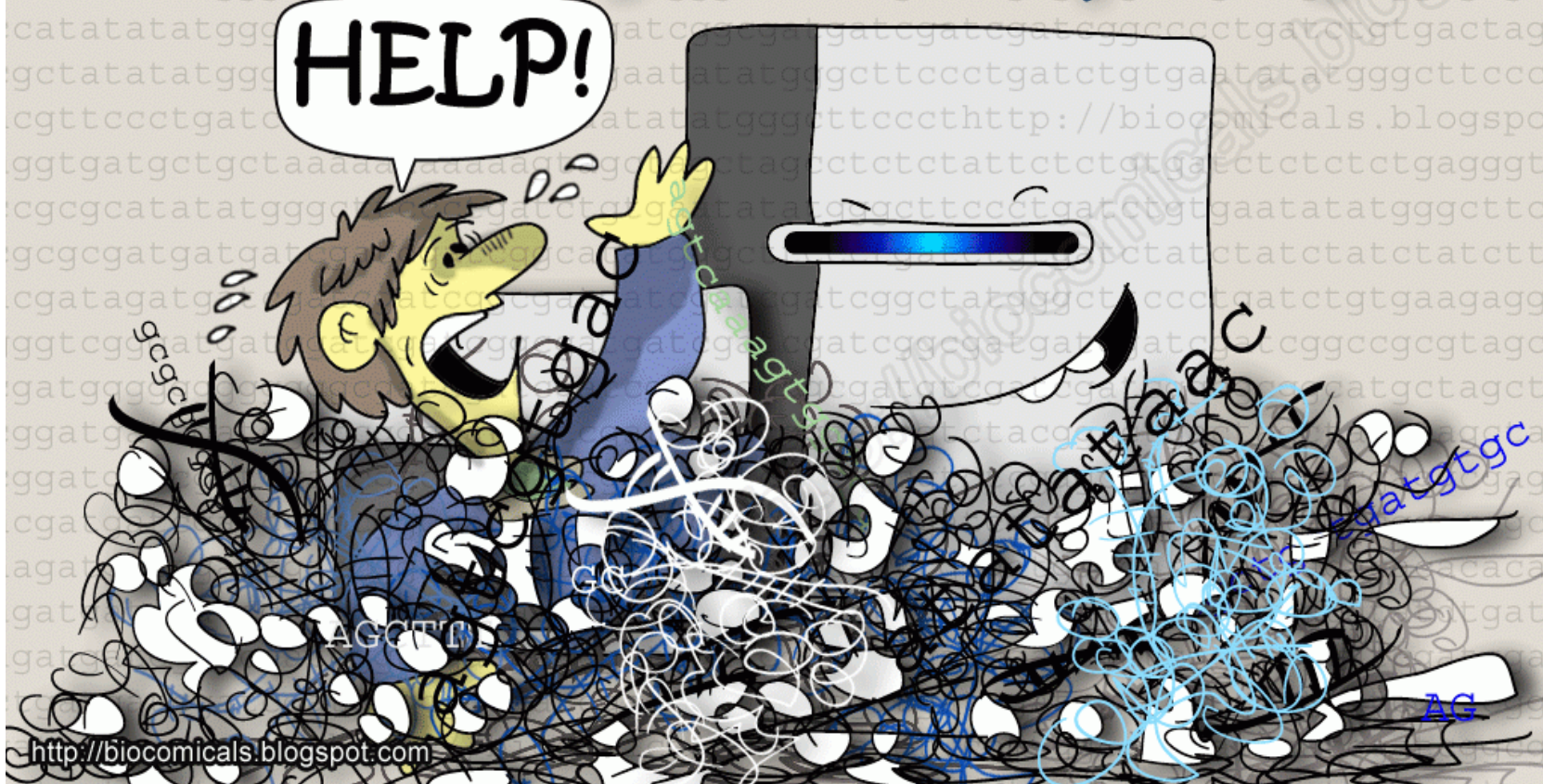
up to 14,900 European ancestry cases, up to 29,485 European ancestry controls

Conjunto de Genes: NO Genoma completo



Drowned in next generation sequencing data

HELP!



<http://biocomicals.blogspot.com>

the **BIG BANG** THEORY™





BIA

Plataforma
Bioinformática
Argentina

BIA - Quiénes somos



Universidad de Buenos Aires (FCEyN – CONICET)

Universidad Católica de Córdoba (CONICET)

Instituto de Agrobiotecnología Rosario (CONICET)

Universidad Nacional de San Martín (IIB-InTeCh, CONICET)

OBJETIVO GENERAL DEL PROYECTO

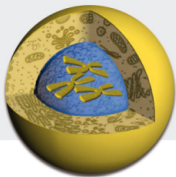
“Desarrollar una plataforma que
brinde
servicios bioinformáticos a las
instituciones públicas y privadas”

OBJETIVO GENERAL DEL PROYECTO

“Desarrollar una plataforma que
brinde
servicios bioinformáticos a las
instituciones públicas y privadas”

DE METAGENOMAS AL DESARROLLO DE DROGAS

CÉLULA

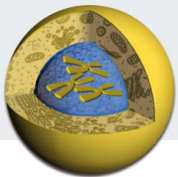


OBJETIVO GENERAL DEL PROYECTO

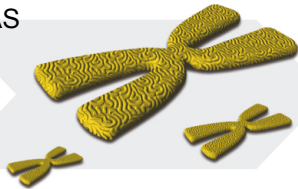
“Desarrollar una plataforma que
brinde
servicios bioinformáticos a las
instituciones públicas y privadas”

DE METAGENOMAS AL DESARROLLO DE DROGAS

CÉLULA



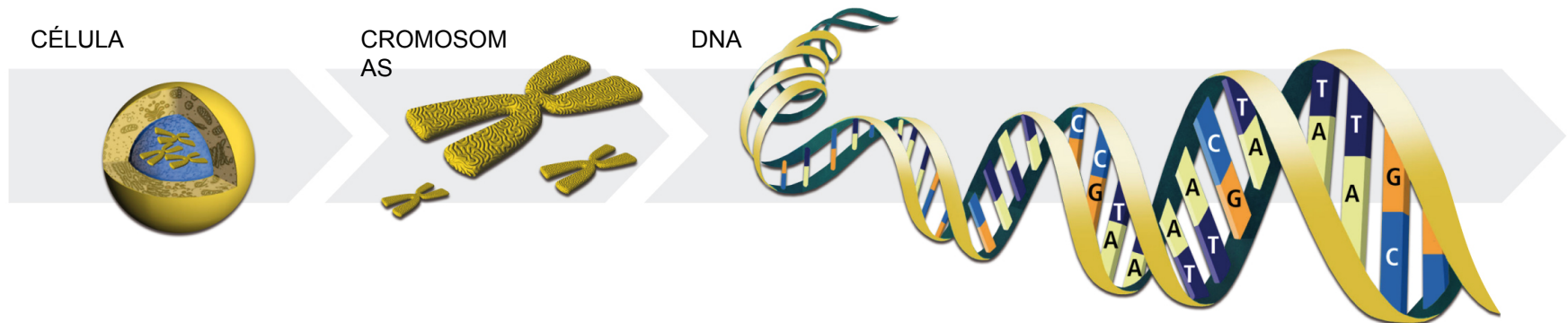
CROMOSOMAS



OBJETIVO GENERAL DEL PROYECTO

“Desarrollar una plataforma que
brinde
servicios bioinformáticos a las
instituciones públicas y privadas”

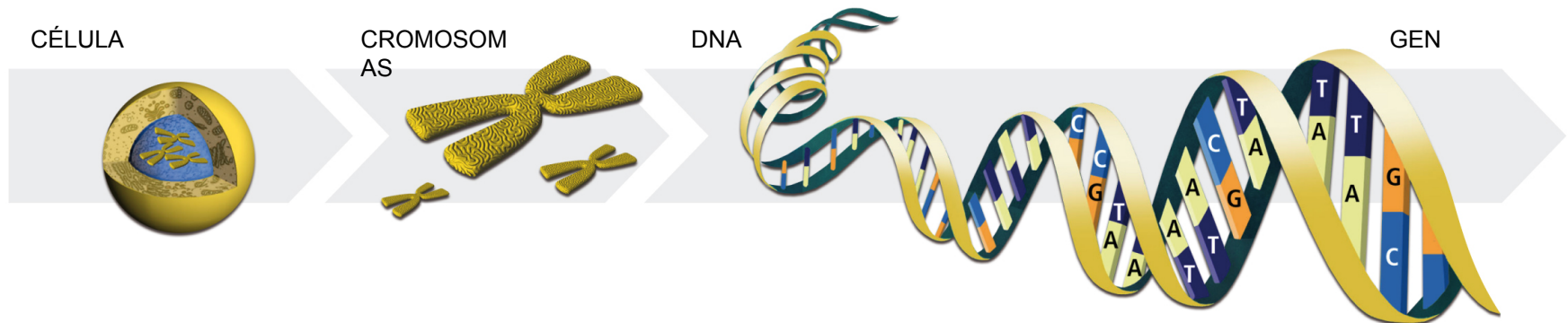
DE METAGENOMAS AL DESARROLLO DE DROGAS



OBJETIVO GENERAL DEL PROYECTO

“Desarrollar una plataforma que
brinde
servicios bioinformáticos a las
instituciones públicas y privadas”

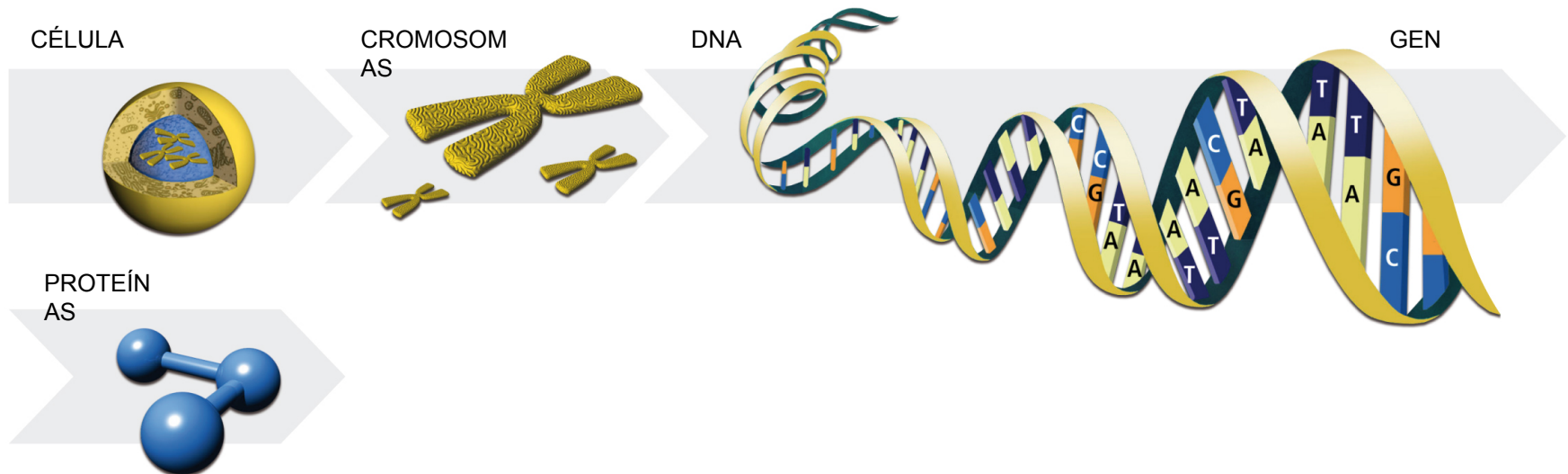
DE METAGENOMAS AL DESARROLLO DE DROGAS



OBJETIVO GENERAL DEL PROYECTO

“Desarrollar una plataforma que
brinde
servicios bioinformáticos a las
instituciones públicas y privadas”

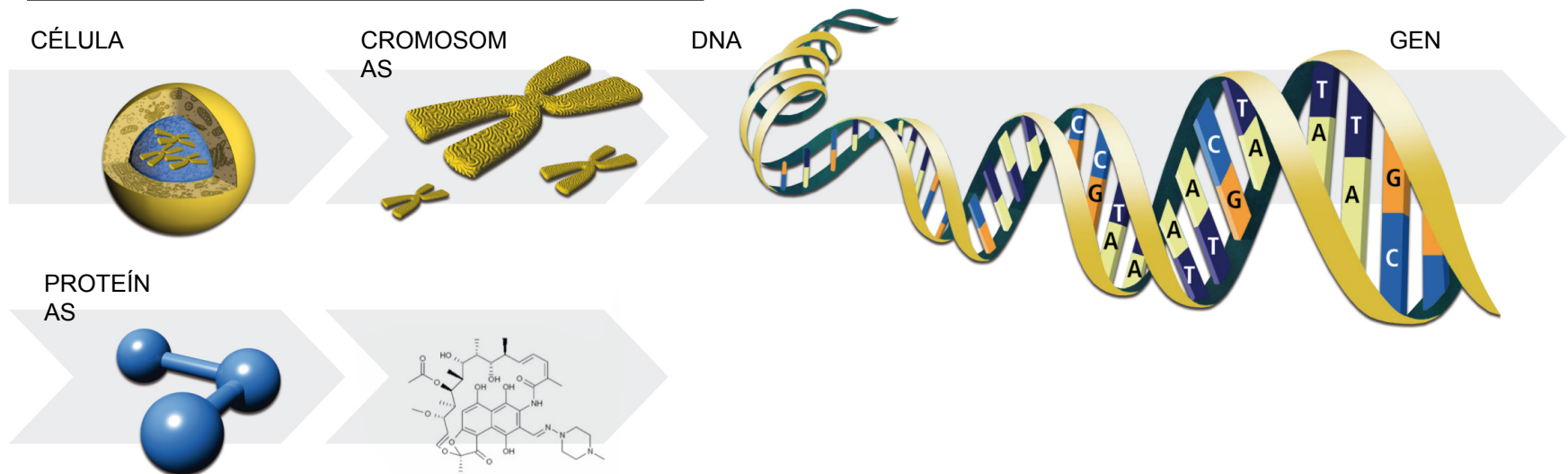
DE METAGENOMAS AL DESARROLLO DE DROGAS



OBJETIVO GENERAL DEL PROYECTO

“Desarrollar una plataforma que
brinde
servicios bioinformáticos a las
instituciones públicas y privadas”

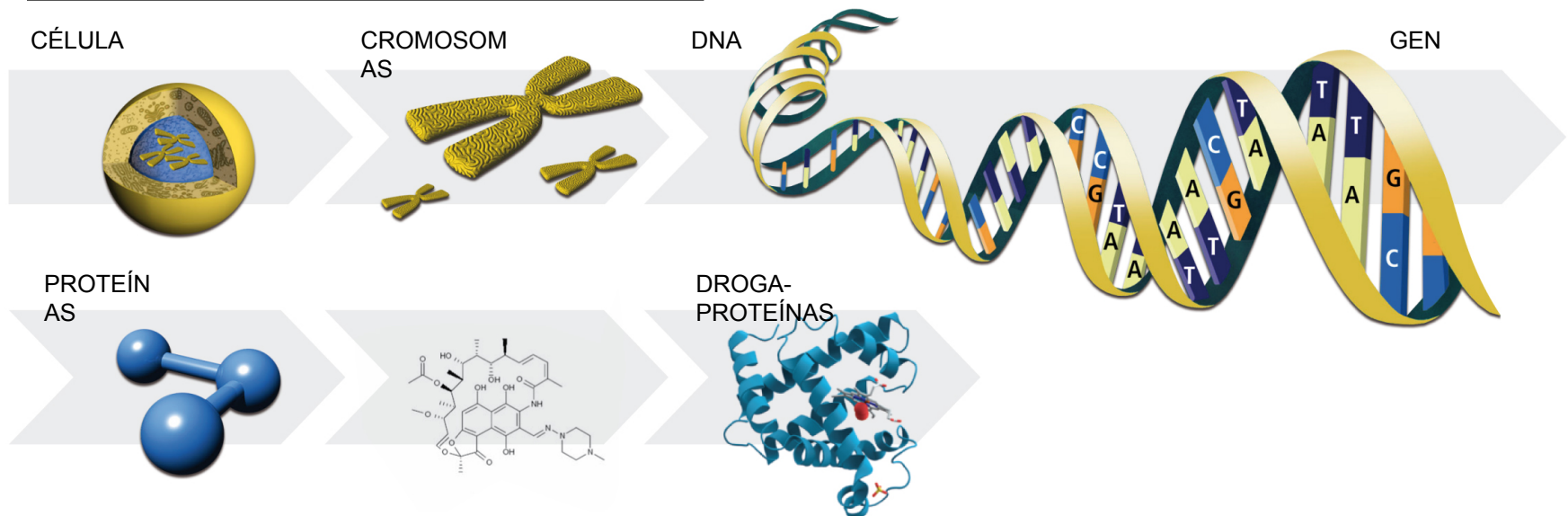
DE METAGENOMAS AL DESARROLLO DE DROGAS

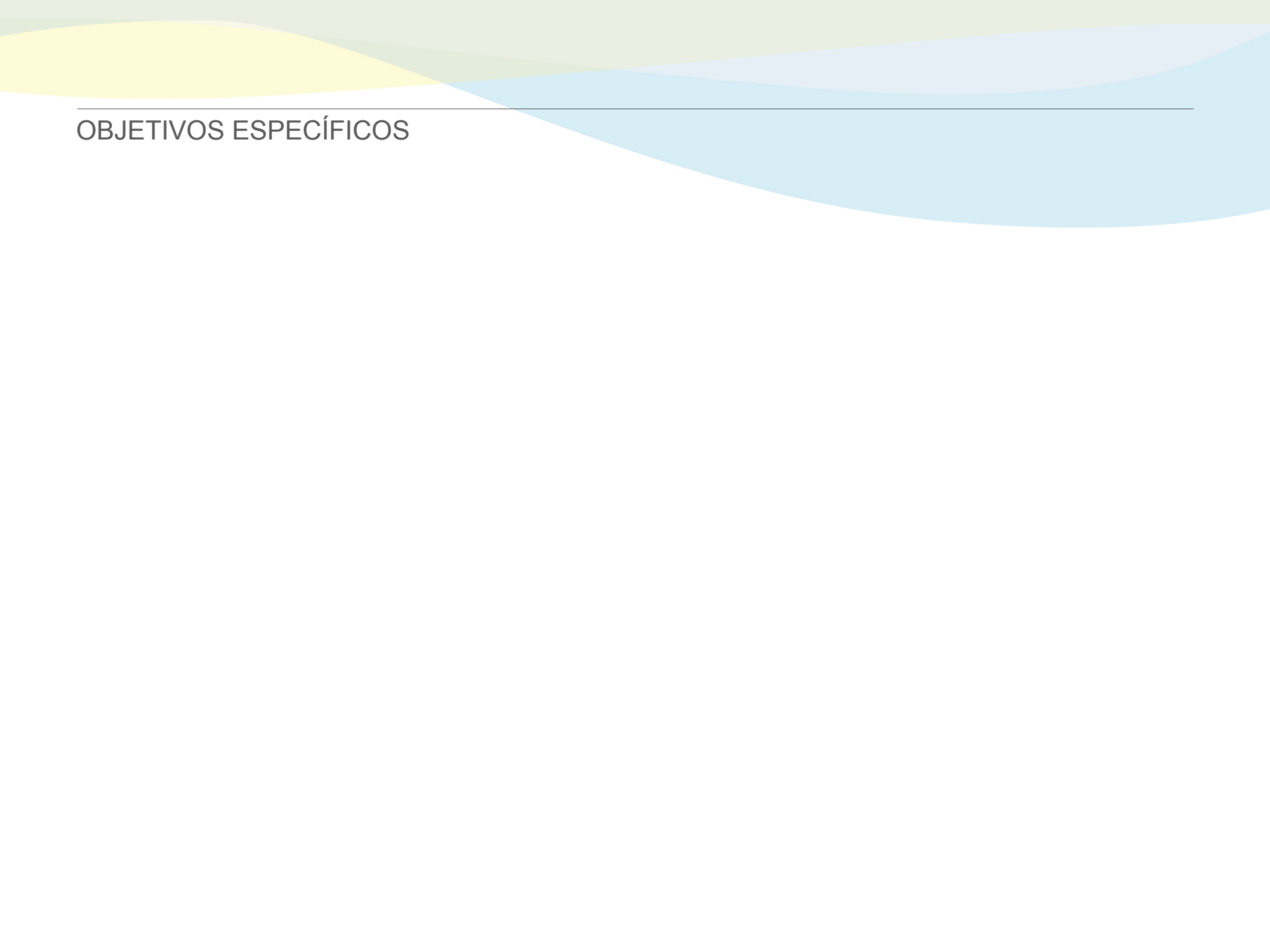


OBJETIVO GENERAL DEL PROYECTO

“Desarrollar una plataforma que
brinde
servicios bioinformáticos a las
instituciones públicas y privadas”

DE METAGENOMAS AL DESARROLLO DE DROGAS





OBJETIVOS ESPECÍFICOS



OBJETIVOS ESPECÍFICOS

Desarrollo de métodos,
algoritmos y herramientas en
diversas ramas bioinformáticas
para prestación de servicios.

-

Desarrollo de métodos, algoritmos y herramientas en diversas ramas bioinformáticas para prestación de servicios.

-

Cubrir necesidades bioinformáticas identificadas en el sector público - privado:

OBJETIVOS ESPECÍFICOS

Desarrollo de métodos, algoritmos y herramientas en diversas ramas bioinformáticas para prestación de servicios.

-

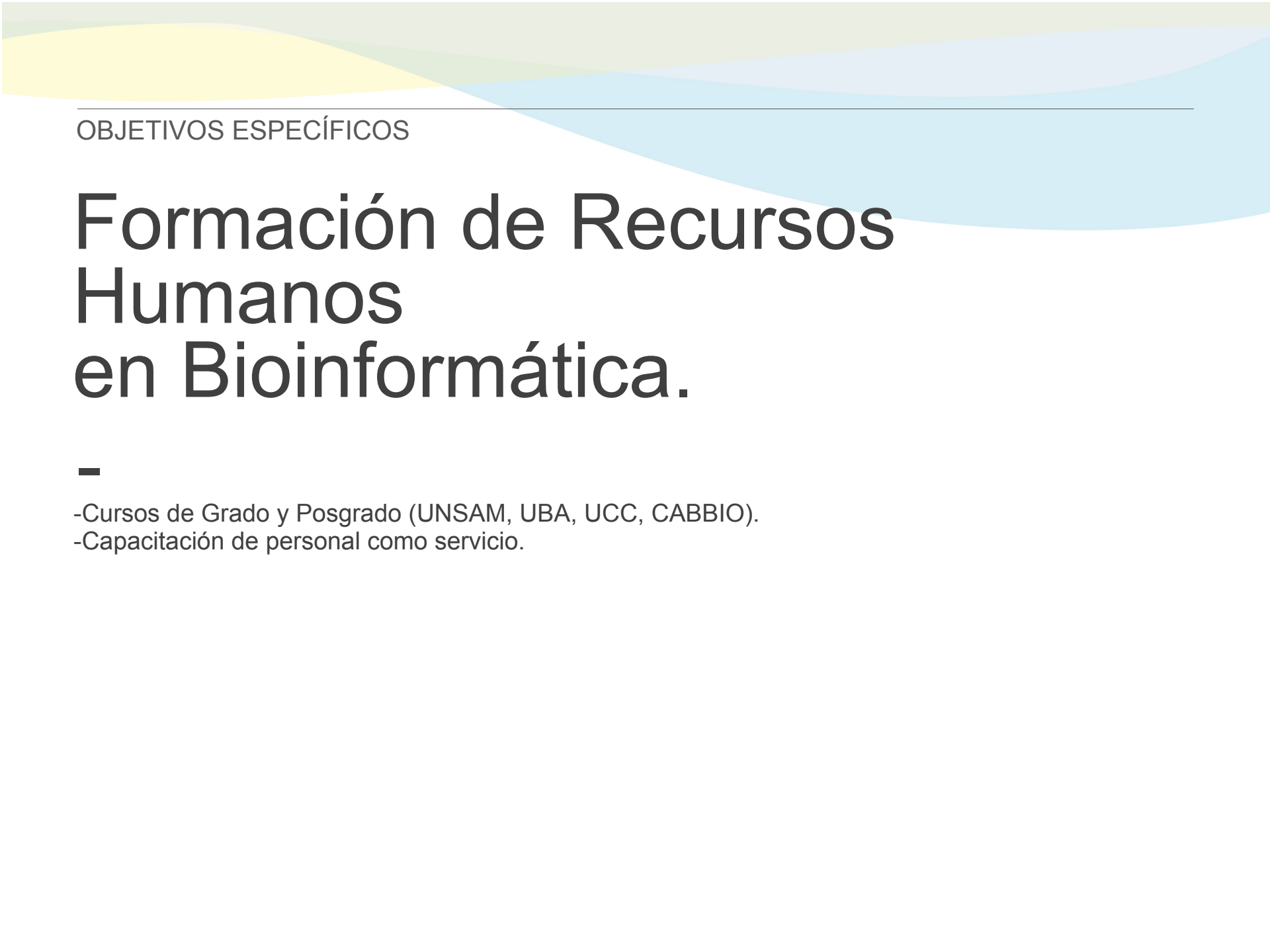
Cubrir necesidades bioinformáticas identificadas en el sector público - privado:

- Grupos de Investigación y Entidades Públicas (Hospitales, ONGs, etc..).
- Empresas: Agroindustria, Salud y Biotecnología.



OBJETIVOS ESPECÍFICOS

Formación de Recursos Humanos en Bioinformática.

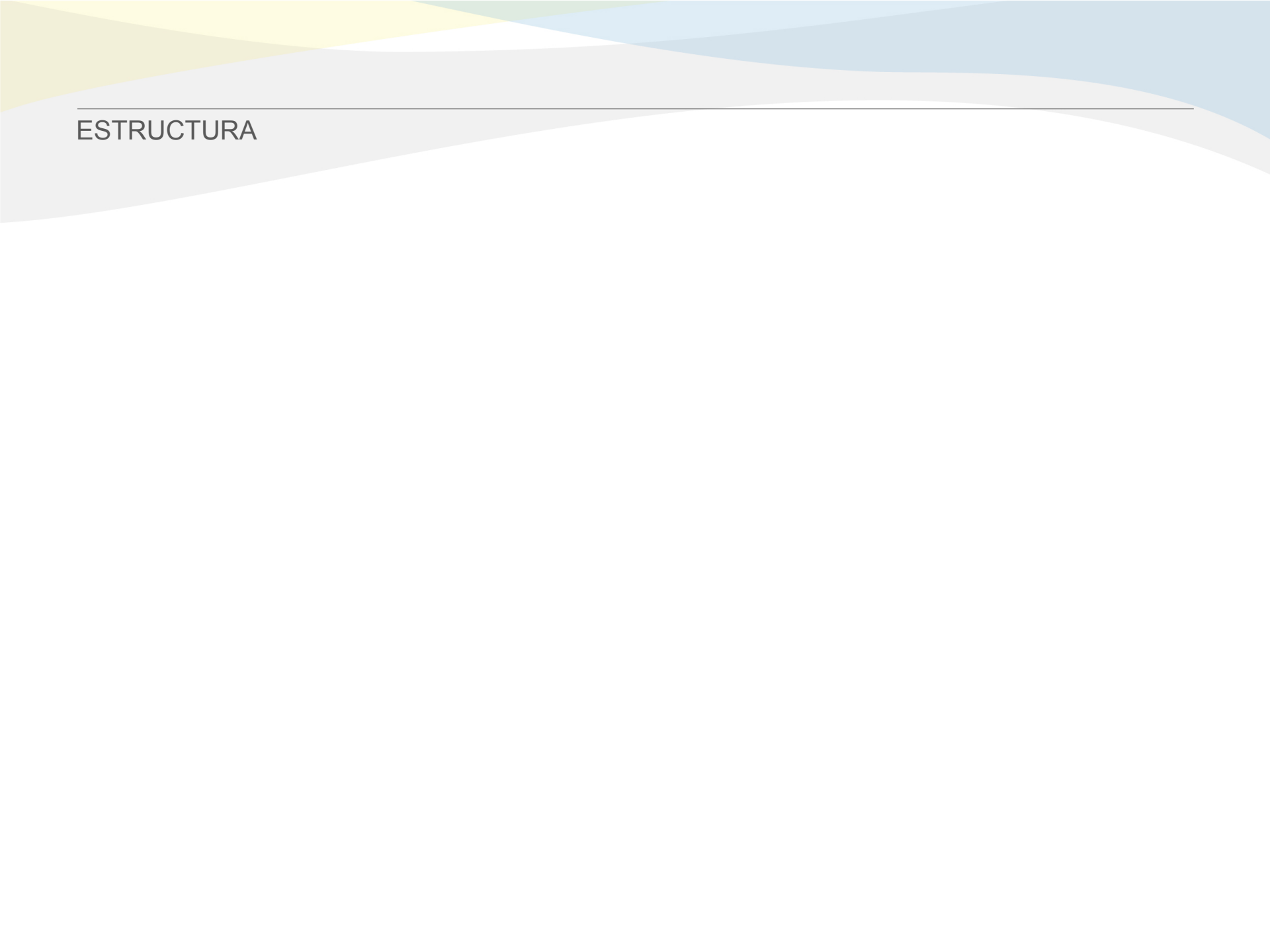


OBJETIVOS ESPECÍFICOS

Formación de Recursos Humanos en Bioinformática.

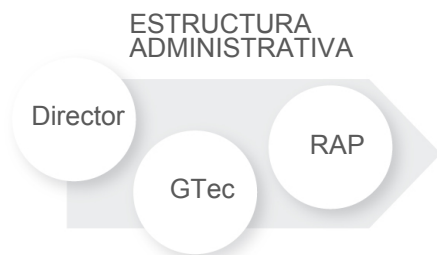
—

- Cursos de Grado y Posgrado (UNSAM, UBA, UCC, CABBIO).
- Capacitación de personal como servicio.



ESTRUCTURA

ESTRUCTURA



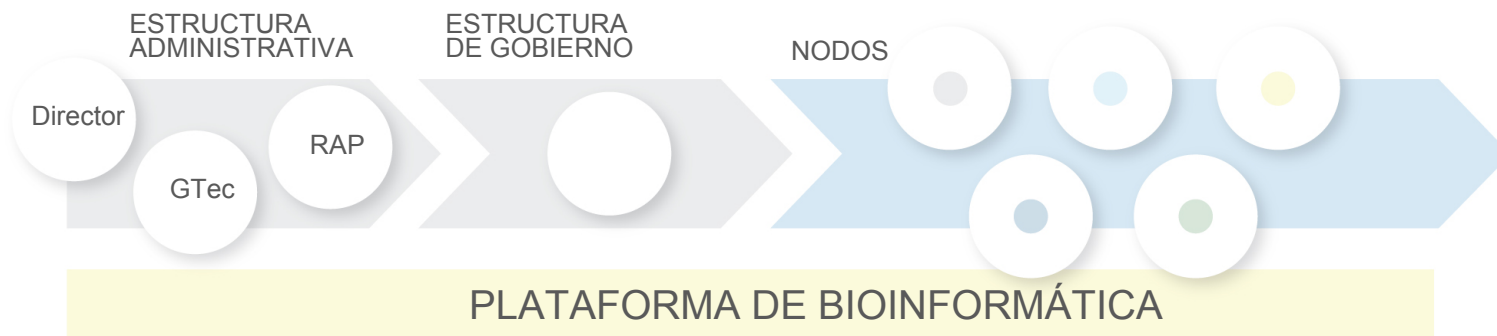
PLATAFORMA DE BIOINFORMÁTICA

ESTRUCTURA

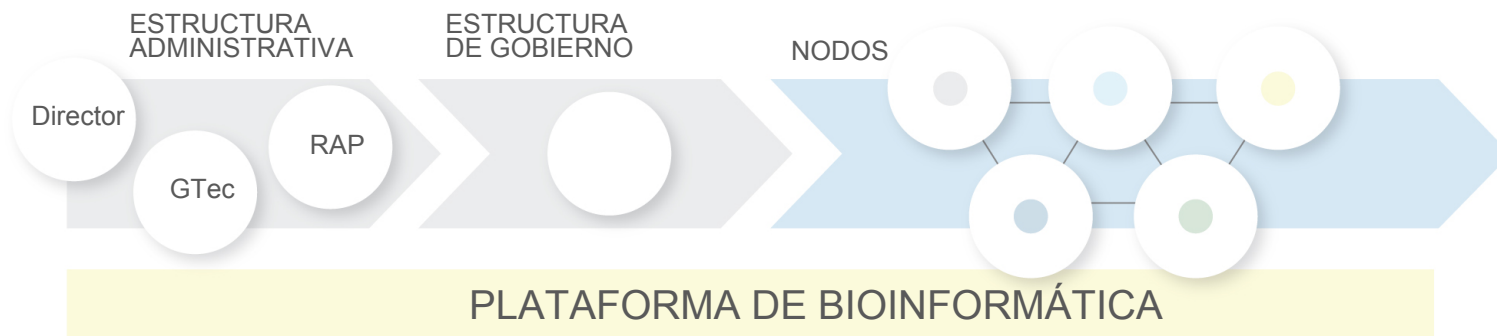


PLATAFORMA DE BIOINFORMÁTICA

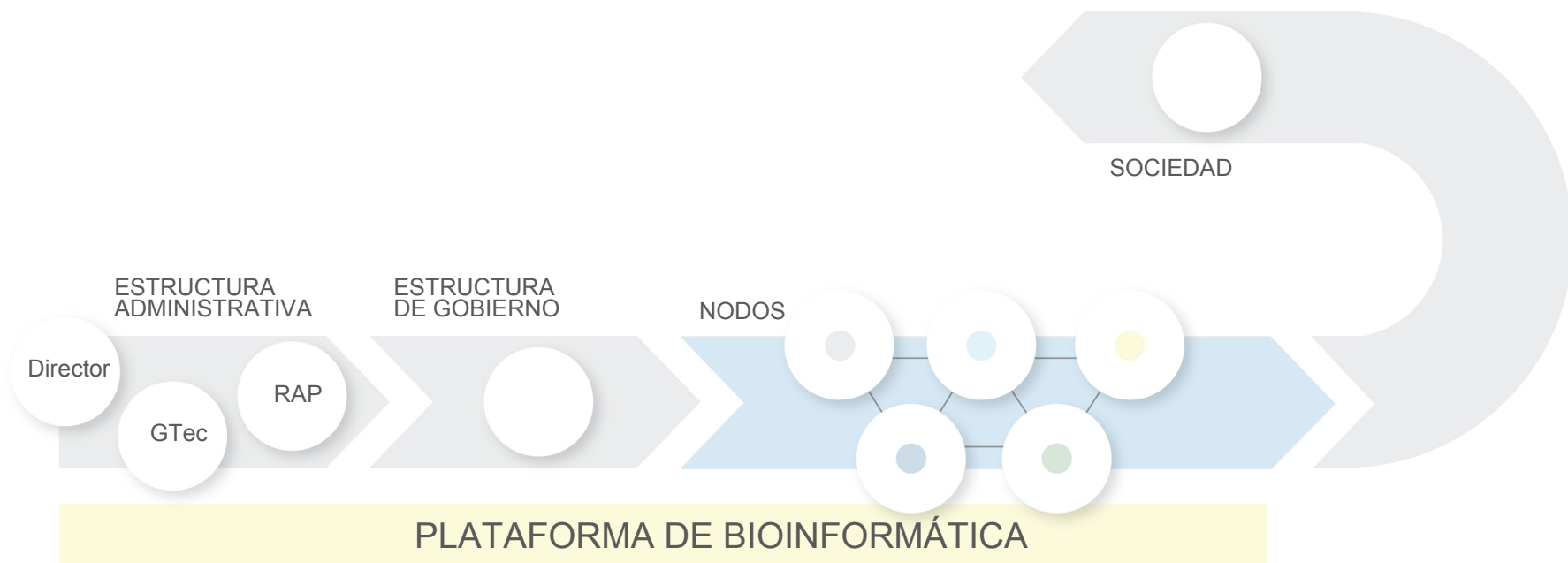
ESTRUCTURA



ESTRUCTURA



ESTRUCTURA



ESTRUCTURA

USUARIOS

SOCIEDAD

ESTRUCTURA
ADMINISTRATIVA

ESTRUCTURA
DE GOBIERNO

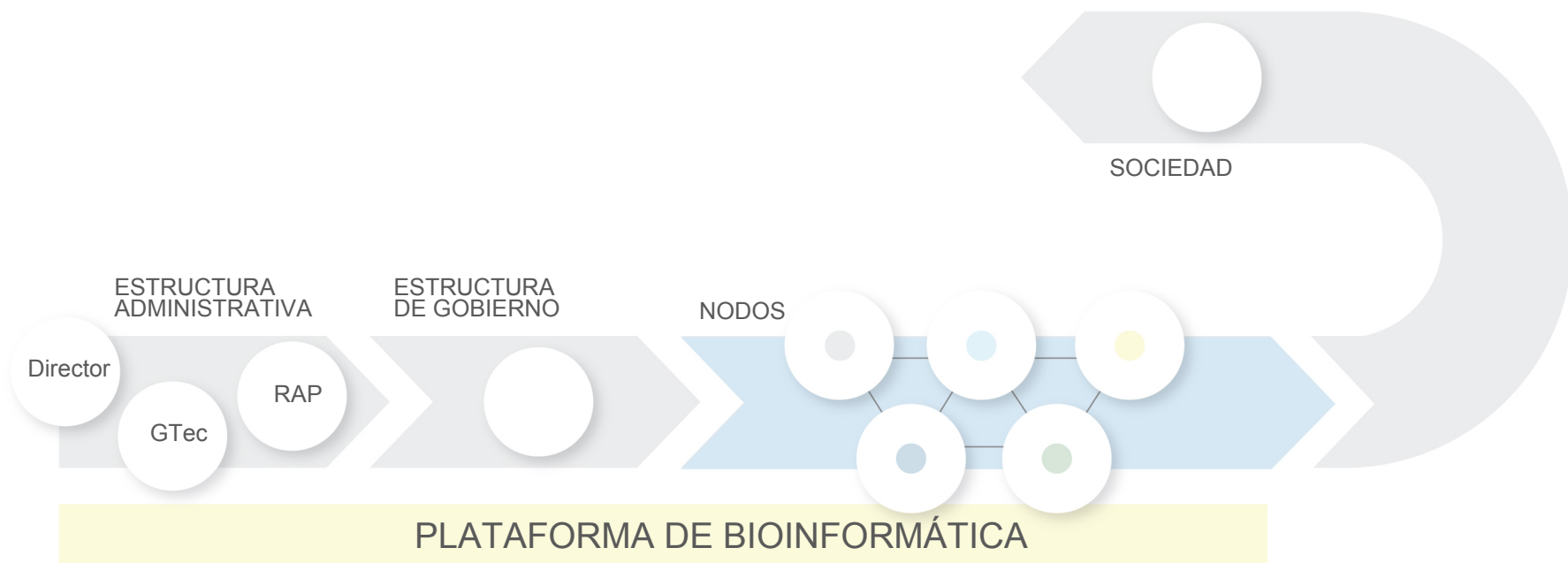
NODOS

Director

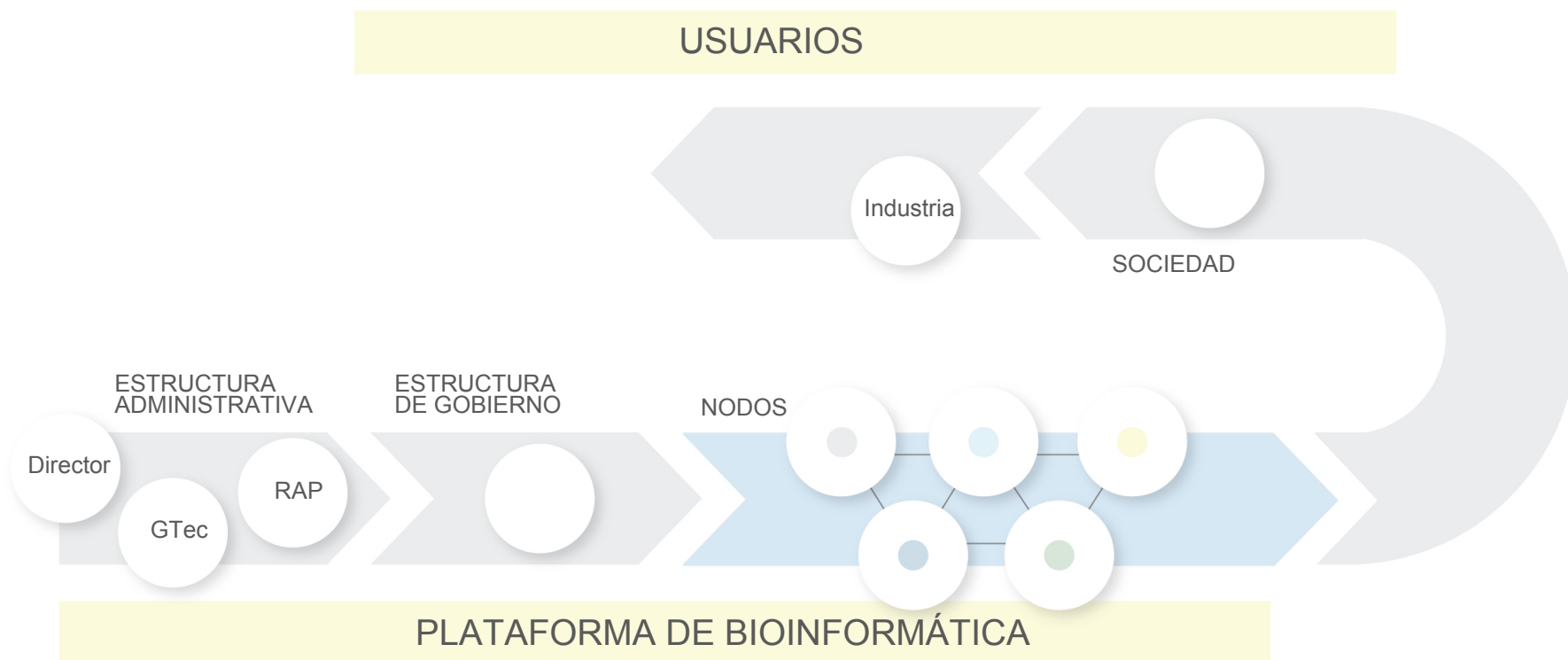
GTec

RAP

PLATAFORMA DE BIOINFORMÁTICA

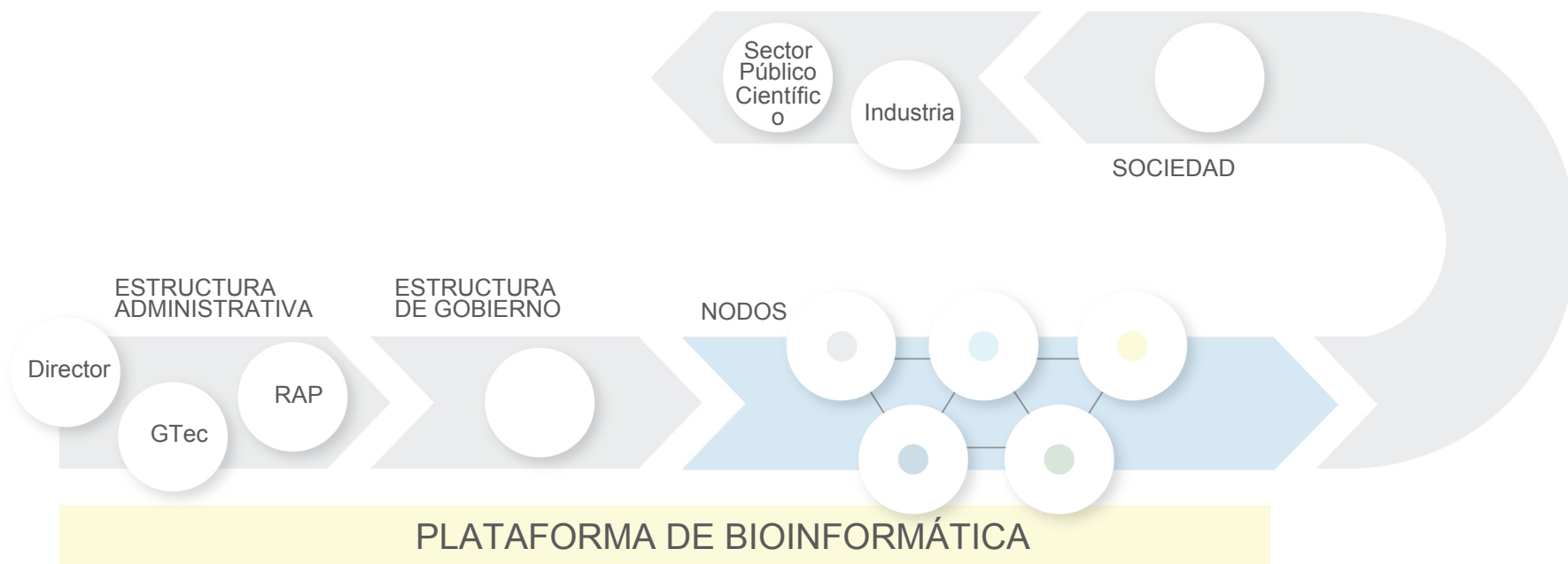


ESTRUCTURA

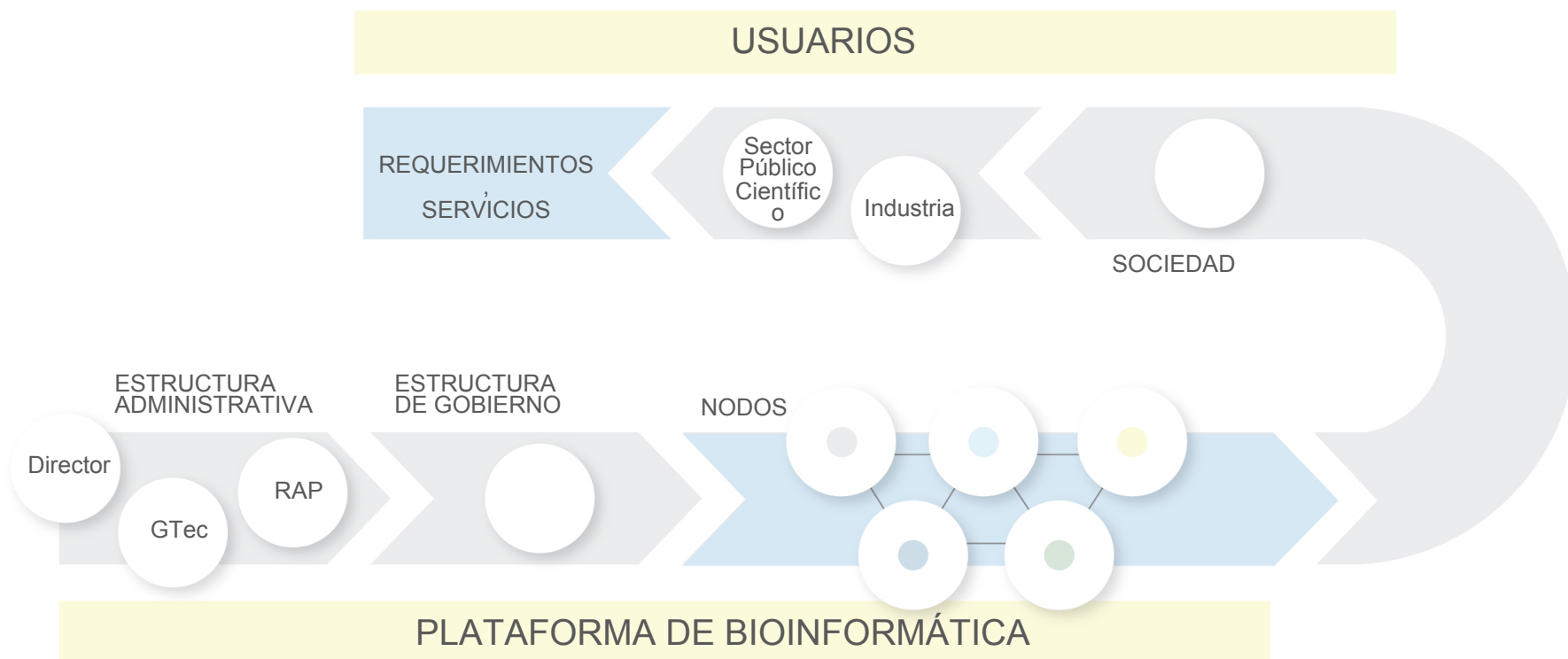


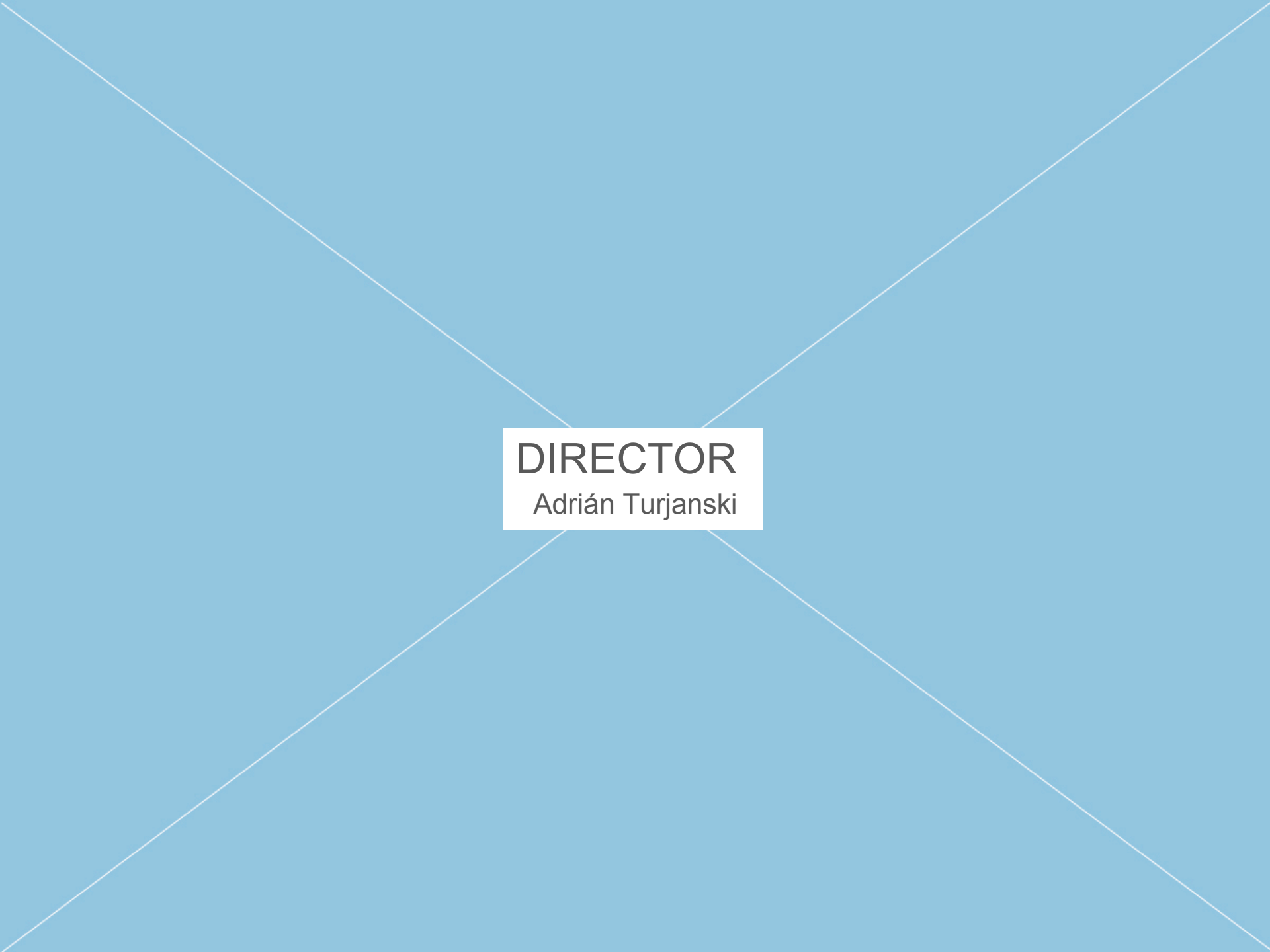
ESTRUCTURA

USUARIOS



ESTRUCTURA





DIRECTOR

Adrián Turjanski

DESARROLLO DE SOLUCIONES INTEGRADAS

DIRECTOR

Adrián Turjanski

```
graph TD; Director["DIRECTOR  
Adrián Turjanski"] --- A["DESARROLLO DE SOLUCIONES INTEGRADAS"]; Director --- B["NODOS DISTRIBUIDOS POR ÁREAS TEMÁTICAS  
COORDINADOS POR UNA GERENCIA TÉCNICO  
ADMINISTRATIVA"]
```

DESARROLLO DE SOLUCIONES INTEGRADAS

DIRECTOR

Adrián Turjanski

NODOS DISTRIBUIDOS POR ÁREAS TEMÁTICAS
COORDINADOS POR UNA GERENCIA TÉCNICO
ADMINISTRATIVA

BASE DE DATOS BIOLÓGICAS

1

IIB-INTECH /
UNSAM

DESARROLLO DE BASES DE DATOS
BIOLÓGICAS
Y DE APLICACIONES WEB PARA VISUALIZACIÓN
Y CONSULTA.

DIRECTOR

Dr. Fernán Agüero



UNSAM

UNIVERSIDAD
NACIONAL DE
SAN MARTÍN

SERVICIOS/ HERRAMIENTAS

Bases de datos de variación genética (SNPs).

-

Aplicaciones quimiogenómicas para búsqueda y validación de blancos (drug discovery).

-

Presidente de la Asociación Argentina de Bioinformática.

-

GENÓMICA DE ALTO RENDIMIENTO

INDEAR

2

SECUENCIACIÓN ADN DE ALTO RENDIMIENTO
Y ANÁLISIS DE GRANDES VOLUMENES DE
DATOS.

DIRECTOR

Dr. Martín Vázquez

CONICET



INDEAR
Instituto de Agrobiotecnología Rosario

SERVICIOS/ HERRAMIENTAS

INDEAR webservices 1.0, primera plataforma Argentina de análisis e interpretación de grandes volúmenes de datos de secuencia (DNA y RNA).

-

BIO DATA MINING

3

MÉTODOS ESTADÍSTICOS/
COMPUTACIONALES PARA
ANÁLISIS DE DATOS MASIVOS
EN BIOLOGÍA.

CONICET /
UCC

DIRECTOR
Dr. Elmer Fernández



SERVICIOS/ HERRAMIENTAS

Hosting de Bases de Datos Microarrays (caARR AY), clinical trials (OpenClinica®), Biobancos(BSI®).

-

Diseño de experimentos, Análisis de datos en tecnologías de alto rendimiento (Microarrays, RNAseq, Proteomica, 2D-DIGE, ITRAQ). InfoGen.

-

BIOINFORMÁTICA ESTRUCTURAL

FCEYN /
UBA

4

ANÁLISIS DE ESTRUCTURA FUNCIÓN DE
BIOMOLÉCULAS
Y DISEÑO RACIONAL DE DROGAS Y PROTEÍNAS.

DIRECTOR

Dr. Darío Estrin

CONICET



SERVICIOS/ HERRAMIENTAS

Búsqueda Virtual de Fármacos, Diseño de Proteínas p/uso Biotecnológico, Quimioinformática, modelos farmacocinéticos, Análisis estructural de proteínas.

-

Co-developers de Amber (Dinámica Molecular).
Desarrollo de métodos QM, QM/MM y High Througput Docking.

-

HPC IT

5

COMPUTACIÓN DE ALTO
RENDIMIENTO.

Depto. De
Computación
FCEYN /
UBA

DIRECTOR
Dr. Esteban Mocskos



SERVICIOS/ HERRAMIENTAS

Asistencia tecnológica en Supercómputo.

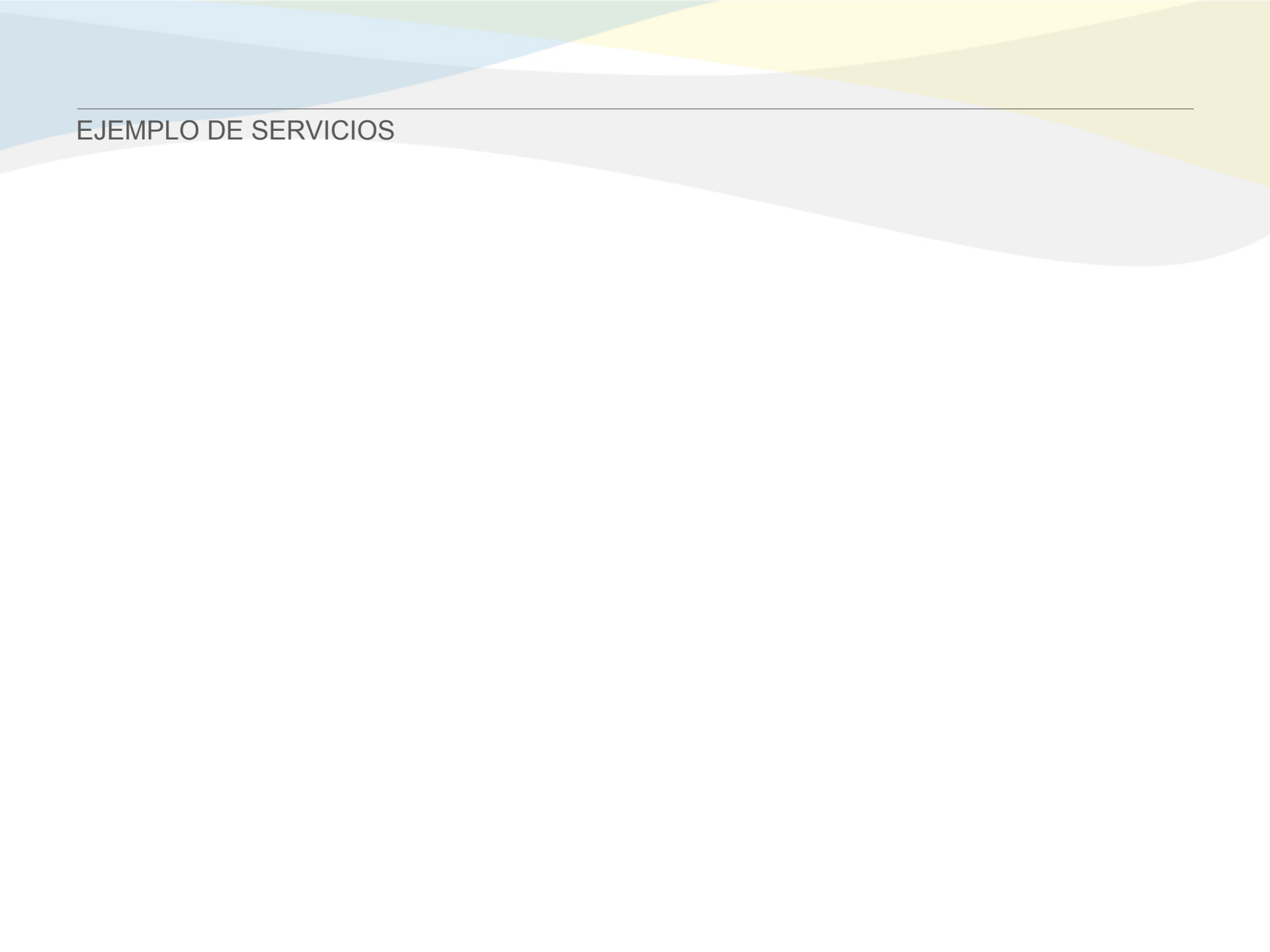
Implementación de nuevas tecnologías de cómputo.

Coordinación de desarrollo distribuido de aplicaciones.

-

Coordinador Centro de Supercómputo de Exactas (SNCAD).

Responsable Argentino de Red Iberoamericana de Supercómputo.



EJEMPLO DE SERVICIOS

Servicios Automáticos de requerimiento alto

- Análisis de secuencias.
- Modelado Estructural de Biomoléculas.

Servicios Semi-Automáticos de requerimiento alto

- Anotación de un Genoma.
- RNA-Seq: Cuantificación de Transcriptomas.
- Mejoramiento de Software para HPC.

Servicios Personalizados de requerimiento bajo

- Desarrollo de Base Datos customizada.
- Análisis de Genomas. GWAS.
- High Throughput Docking de Moléculas.



BIA

Plataforma
Bioinformática
Argentina

Un grupo consolidado de excelencia científica
y de referencia a nivel nacional con
experiencia
en servicios.



Dr. Elmer
Fernández



Dr. Esteban
Mocskos



Dr. Adrián Turjanski



Dr. Fernán Agüero



Dr. Darío Estrin

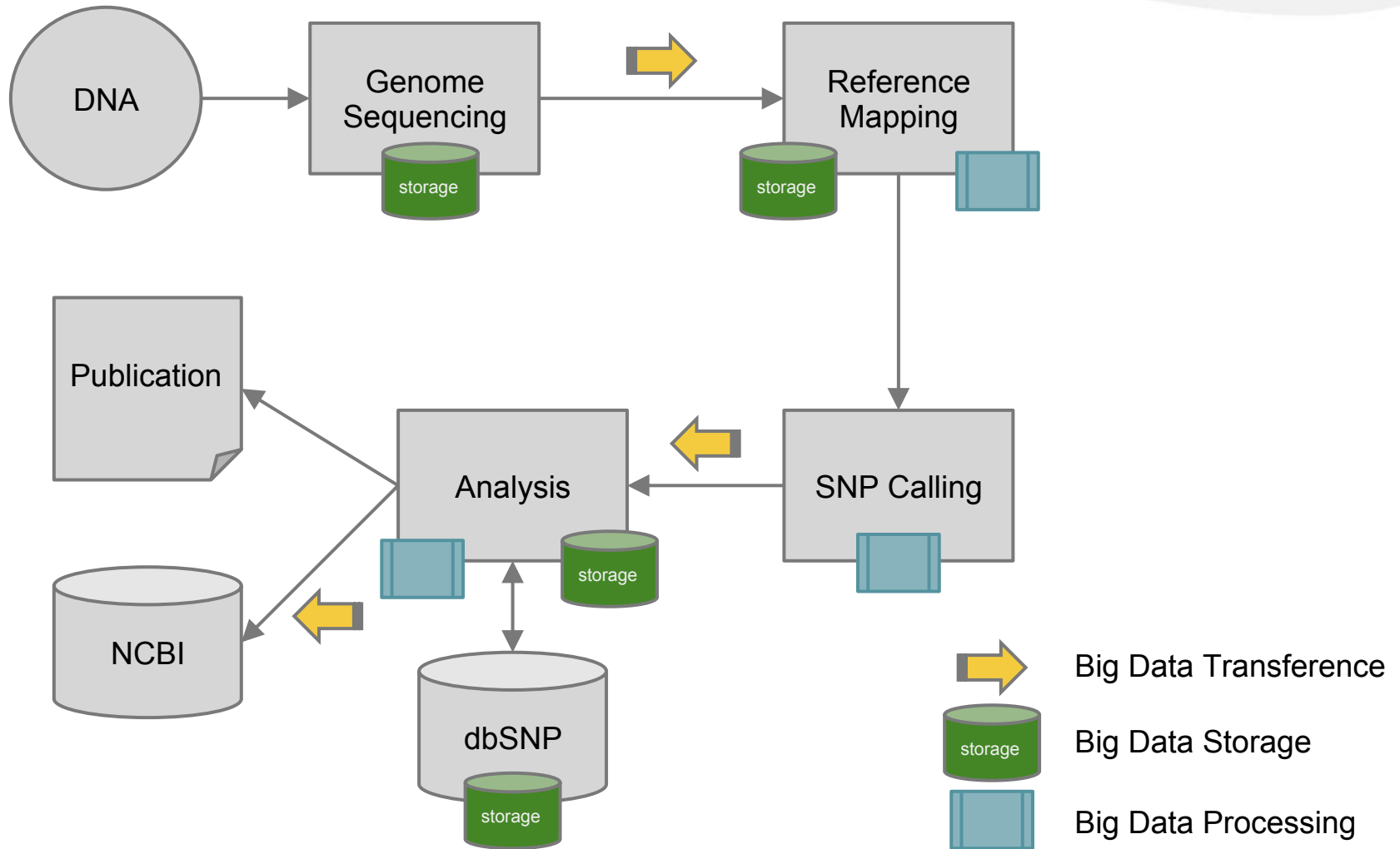


Dr. Martín Vázquez

BIA - Qué hacemos

- Desarrollo de métodos, algoritmos y herramientas en diversas ramas bioinformáticas para prestación de servicios.
- Cubrimos necesidades bioinformáticas identificadas en el sector público - privado:
 - Grupos de Investigación y Entidades Públicas (Hospitales, ONGs, etc...).
 - Empresas: Agroindustria, Salud y Biotecnología.
- Formación de Recursos Humanos en Bioinformática.
 - Cursos de Grado y Posgrado (UNSAM, UBA, UCC, CABBIO).
 - Capacitación de personal como servicio.

Big Data en Genómica



Necesitamos soluciones BIG DATA

Muchas Gracias!

