

bac
ZARAGOZA 2021
Online



Introducción al análisis de datos ómicos con Python

Introducción al análisis de datos ómicos con Python

Introducción al análisis de datos ómicos con Python

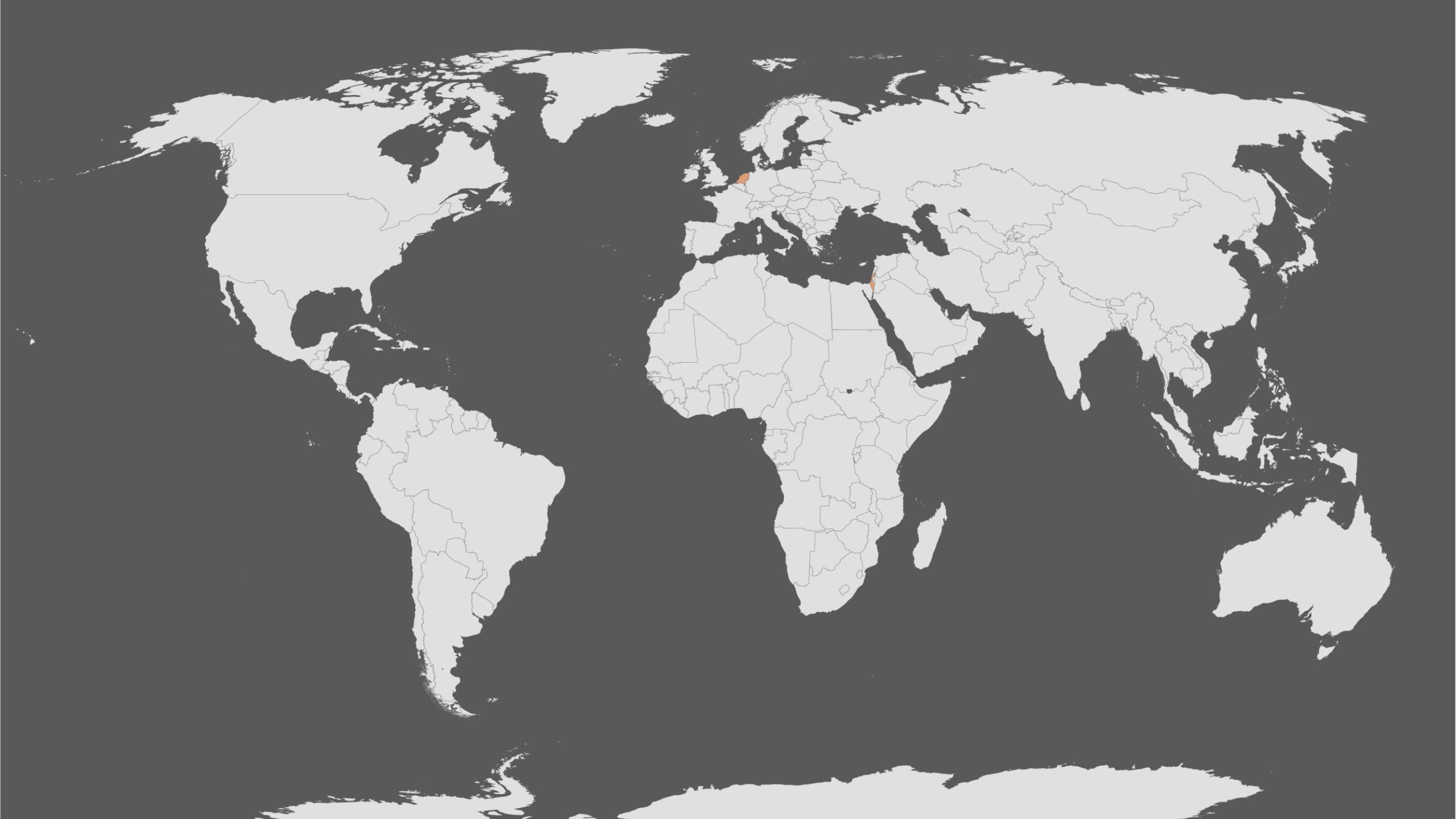


phyton

NO



python





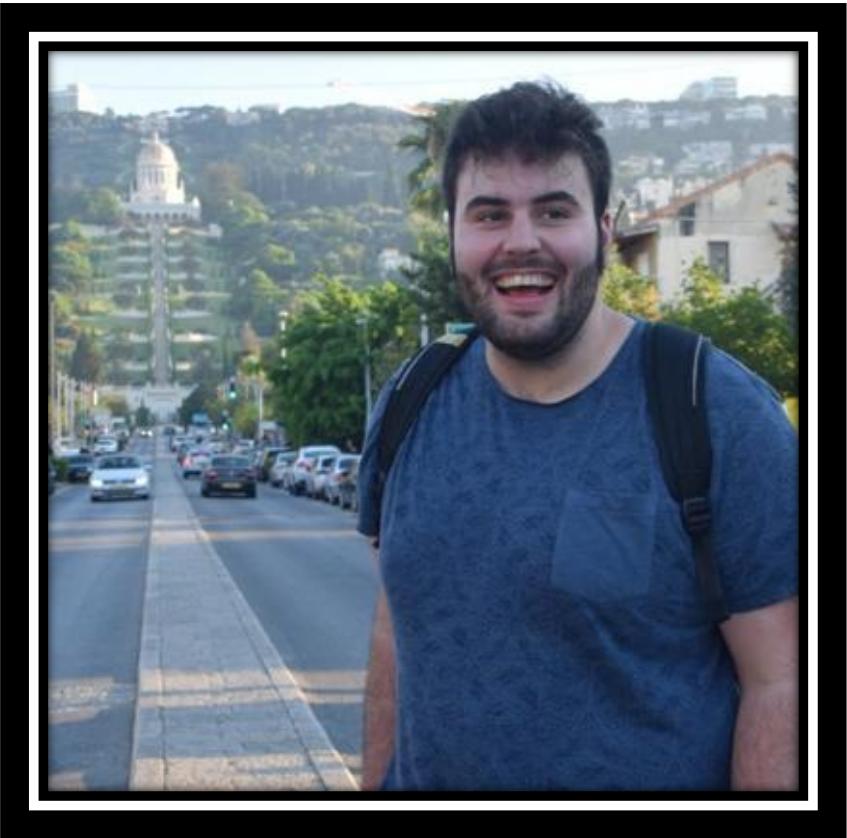


Jonatan Fernández
(Biostatisti ingeniusi)



Mango, MD. MSc.
Teaching assistant

Pulgui, PhD.
Teaching assistant



Jonatan Fernández
(Biostatisti ingeniusi)



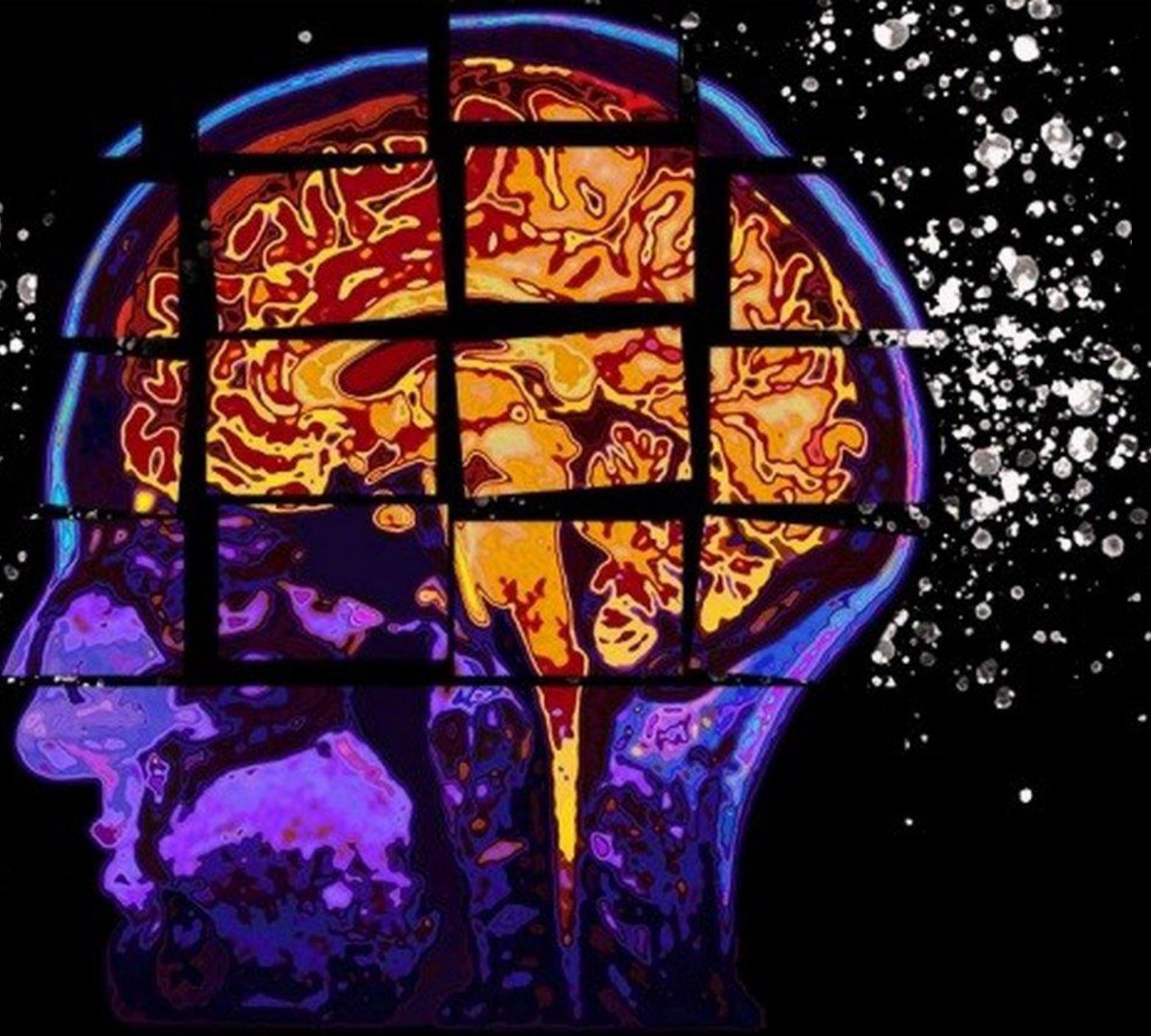
Mango, MD. MSc.
Teaching assistant

Pulgui, PhD.
Teaching assistant



METACAN
Metabolism Immunity Cancer







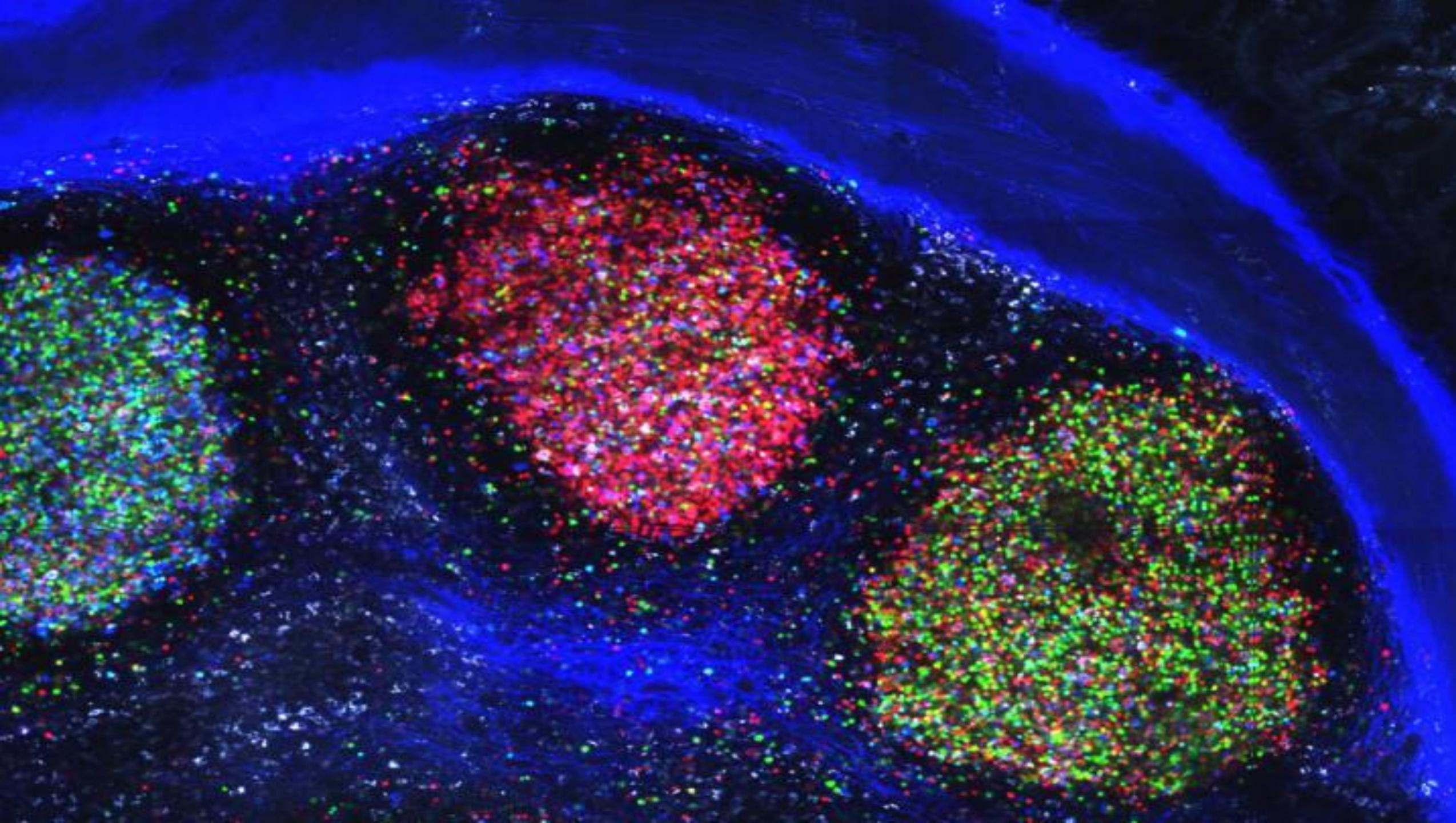
Rodrigo García

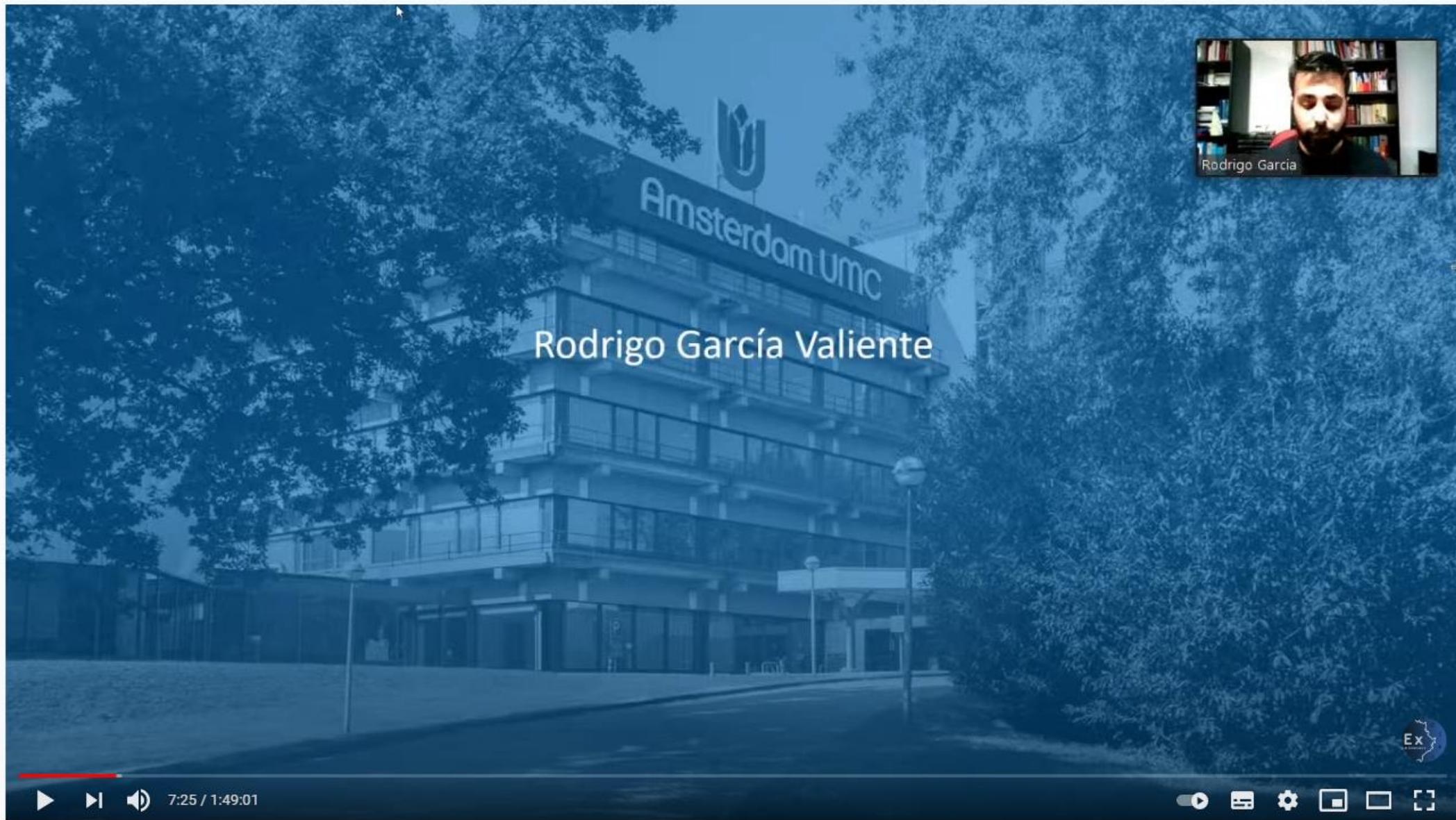
Socorro





Sombrita, PhD
Cavia porculis





ACHO, ¿Y DESPUÉS QUÉ? | Tercera sesión

523 visualizaciones • Emitido en directo el 14 dic 2020

19

0

COMPARTIR

GUARDAR

...



Objetivos

- Introducir a nivel teórico las distintas ómicas y sus características. Genómica (metagenómica), Epigenómica, Transcriptómica, Proteómica, Metabolómica. Análisis relacionados con cada ómica.
- Introducir las bases de la programación en Python.
- Presentar el esquema general de análisis de datos en RNA-Seq, y cómo se extrae a otras ómicas. Mención especial de casos en otras ómicas (genómica).
- Presentar las distintas opciones de análisis tras la generación de la tabla de datos: análisis exploratorios, análisis de expresión diferencial, análisis funcional, *machine learning*, Biología de Sistemas.
- *Wrap-up:* Presentar tendencias actuales en el sector y sus nuevas posibilidades. Recomendar recursos online para la formación en bioinformática, estadística y ML.
- Que os guste.

Sesiones

- **Sesión 1 (2 h): Introducción al curso.** Presentación. Ómicas, usos y peculiaridades. Instrumentación.
- **Sesión 2 (2 h): Introducción breve a la programación en Python.** Introducción a Python y a Jupyter Notebooks. Variables simples y complejas. Funciones. Clases y Métodos. Automatización. Bucles y Condicionales. Librerías. Estructura de una librería. Numpy y visualización de datos.
- **Sesión 3 (2 h): Procesado de datos RNA-Seq.** Resultados. Genes, proteínas. BBDD. Otros casos en ómicas.
- **Sesión 4 (2 h). Análisis fundamentales.** Repaso de estadística. Análisis exploratorio. Bootstrapping. Análisis de expresión diferencial. Análisis de coexpresión. Análisis funcional.
- **Sesión 5 (2h): Machine learning.** Una introducción a ML. Funciones de coste. Optimización matemática. Reflexión. TO DOs y DON'Ts, problemas en biología. Tamaño de muestras.
- **Sesión 6 (2h): Un paso más allá: Biología de Sistemas, tendencias y cierre.** Exprimiendo los datos: FBA. Biología de Sistemas. Presentación de “nuevas” tendencias en el área (scRNA-Seq, etc.). Wrap-up y despedida.

TOO MUCH
Time



TOO LITTLE





teachers
that use zoom

teachers
that use discord







Say it
Louder!

PYTHON



A close-up photograph of a person's hand holding a small, yellow sticky note. The word "PYTHON" is handwritten in blue ink on the note, with a blue horizontal line underneath it. The background is blurred, showing what appears to be a computer monitor and some colorful objects.



R vs Python

10 reason to learn **Python**



Ease of use

Extensive Support
Libraries

Machine learning

Extensible

Embeddable

Raspberry Pi

Speedy

Web development

Server-side scripting

A less limited
programming approach



Object Oriented Programming

CLASS

Human

Attributes

Name

Age

Email

Object

SomeName

Methods

Walk

sendEmail







The logo for the programming language Julia. It features the word "julia" in a bold, black, sans-serif font. Above the letter "j", there is a single blue circle. Above the letters "ullia", there is a cluster of four circles: green, red, purple, and another green, arranged in a small group.

Conda install — C x Python for ecology x ecology-bug-bbq x Comparing data x Home x Untitled x Carol

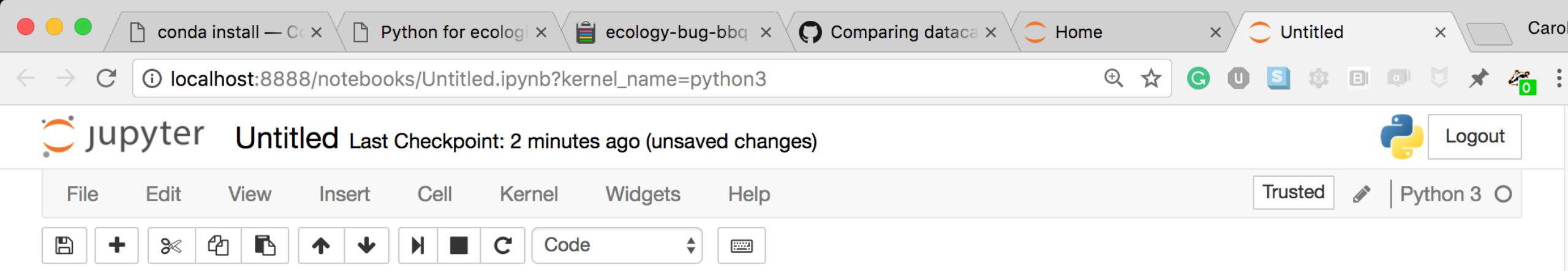
localhost:8888/notebooks/Untitled.ipynb?kernel_name=python3

jupyter Untitled Last Checkpoint: 2 minutes ago (unsaved changes) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

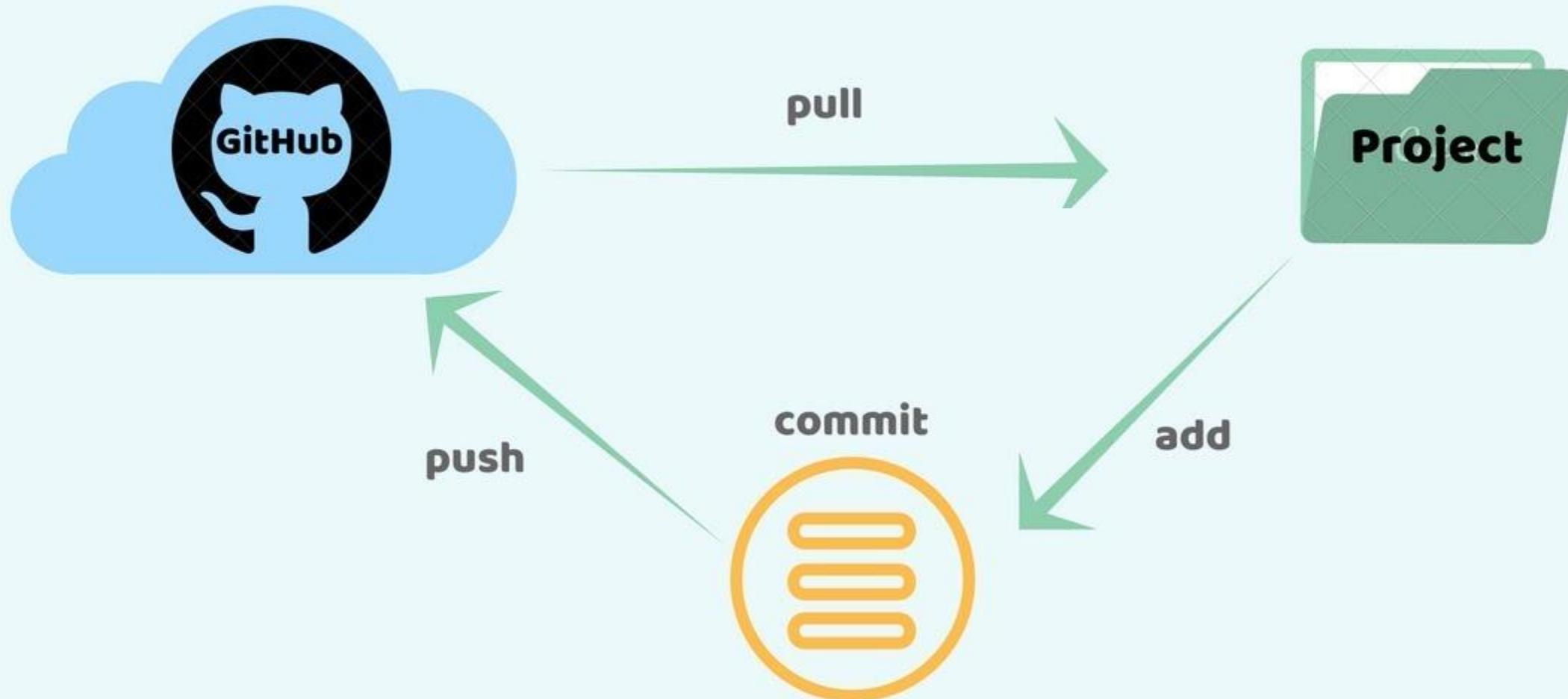
In []: `print('Hello world of ecology!')`



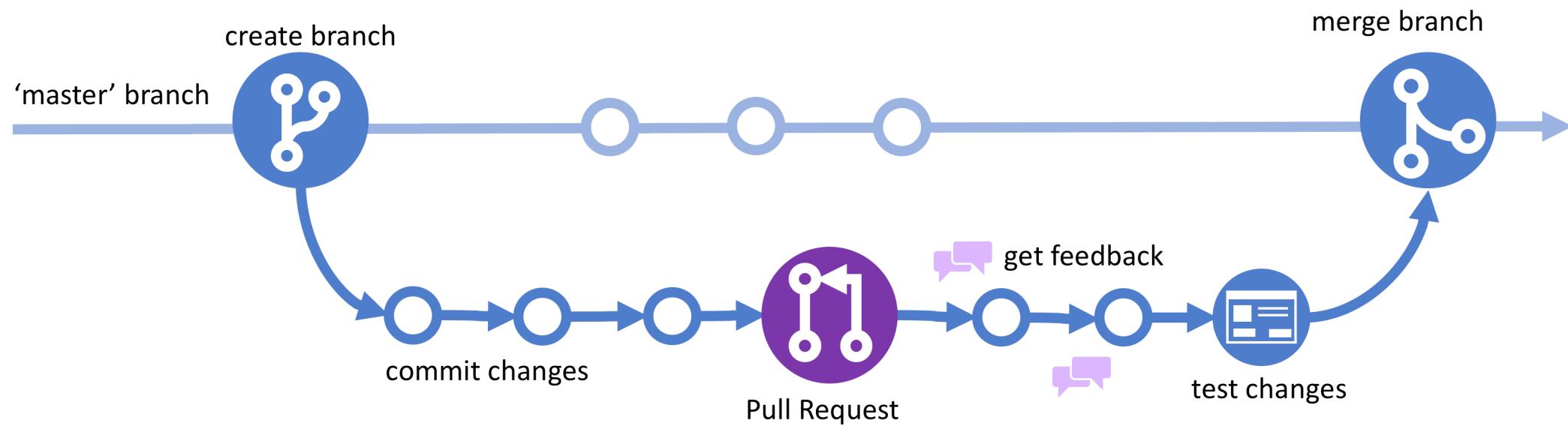


GitHub

Git Push Pull Commands



GitHub Flow



EVERYTHING



TOO MUCH
Time



TOO LITTLE

Inaccuracies



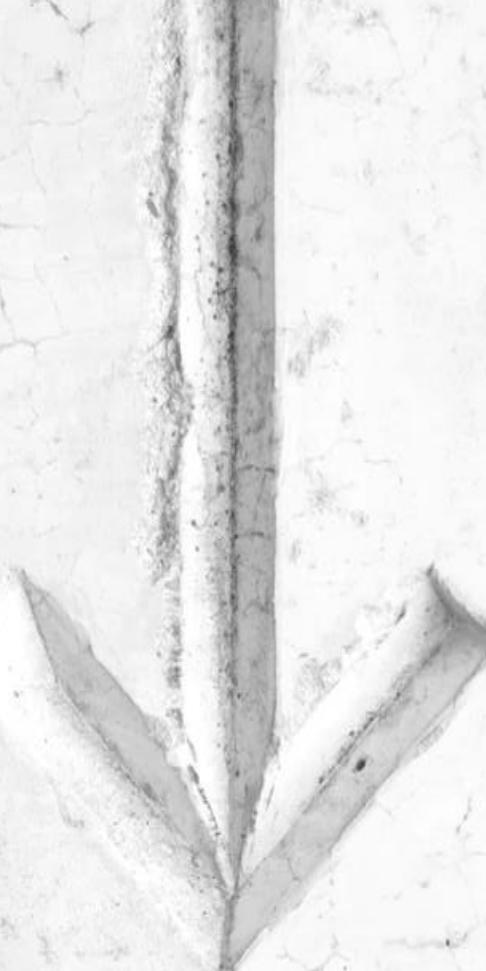


**CUANDO
EMPIECE A
HABLAR
OS VAIS A
CAGAR**

DON'T PANIC

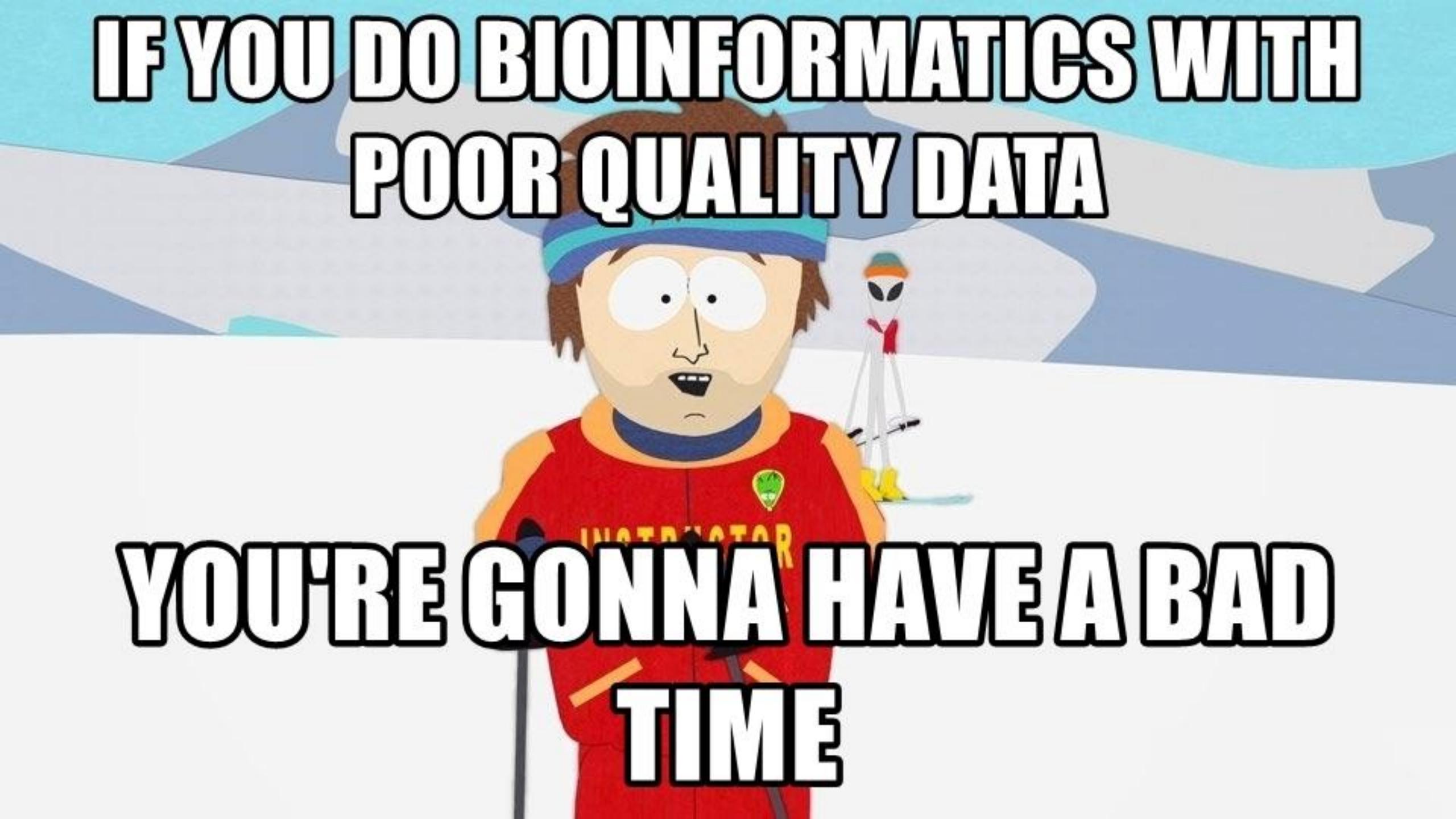
A tener siempre en cuenta

Orden de ejecución (y de errores)



1. Diseño experimental
2. Experimento
3. Datos crudos
4. Datos procesados
5. Análisis sobre datos procesados
6. Interpretación

**IF YOU DO BIOINFORMATICS WITH
POOR QUALITY DATA**



**YOU'RE GONNA HAVE A BAD
TIME**

Ómicas y su importancia.
Peculiaridades y usos.

ÓMICAS

ÓMICAS

ÓMICAS

ÓMICAS

ÓMICAS

ÓMICAS

ÓMICAS





HERE

THERE

EVERYWHERE

GENÓMICA

EPIGENÓMICA

TRANSCRIPTÓMICA

PROTEÓMICA

METABOLÓMICA

METAGENÓMICA

GENÓMICA *METAGENÓMICA*

EPIGENÓMICA

TRANSCRIPTÓMICA

PROTEÓMICA

METABOLÓMICA

GENÓMICA
EPIGENÓMICA
TRANSCRIPTÓMICA
PROTEÓMICA
METABOLÓMICA

METAGENÓMICA
LIPIDÓMICA
GLICÓMICA
FLUXÓMICA
FENÓMICA
METAPROTEÓMICA
ETCEÓMICA

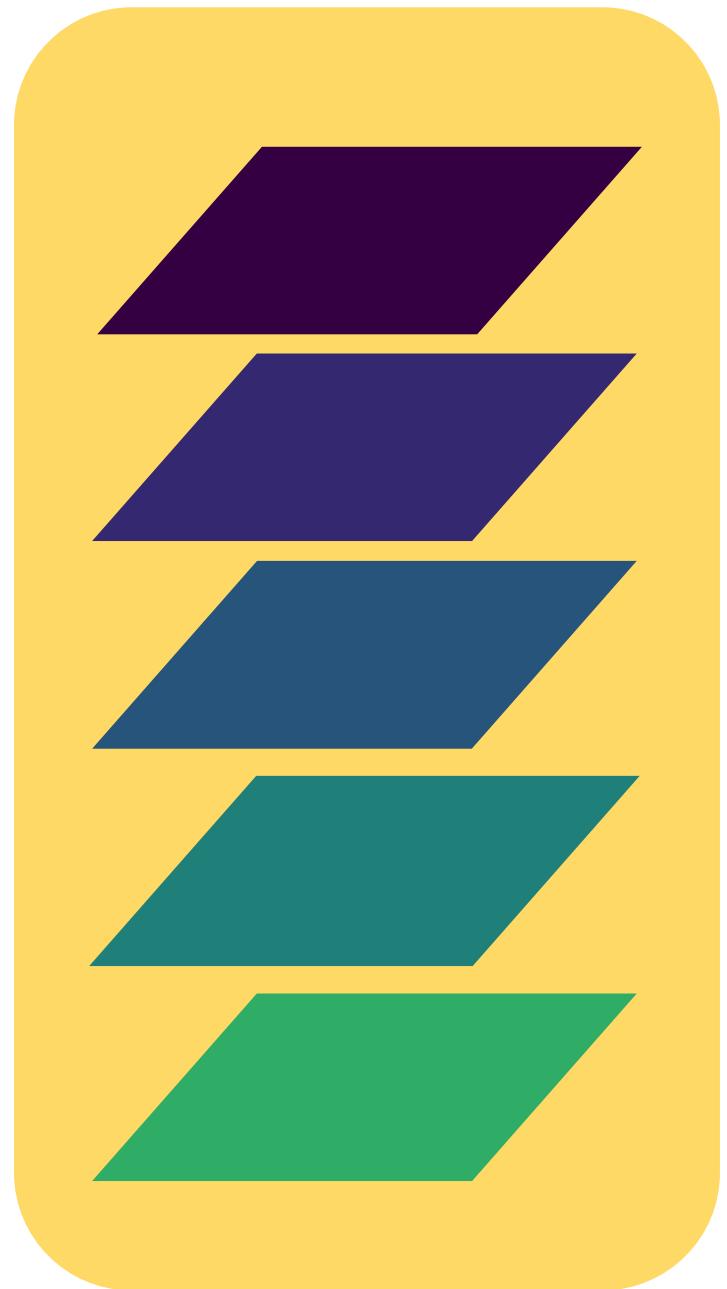
GENÓMICA *METAGENÓMICA*

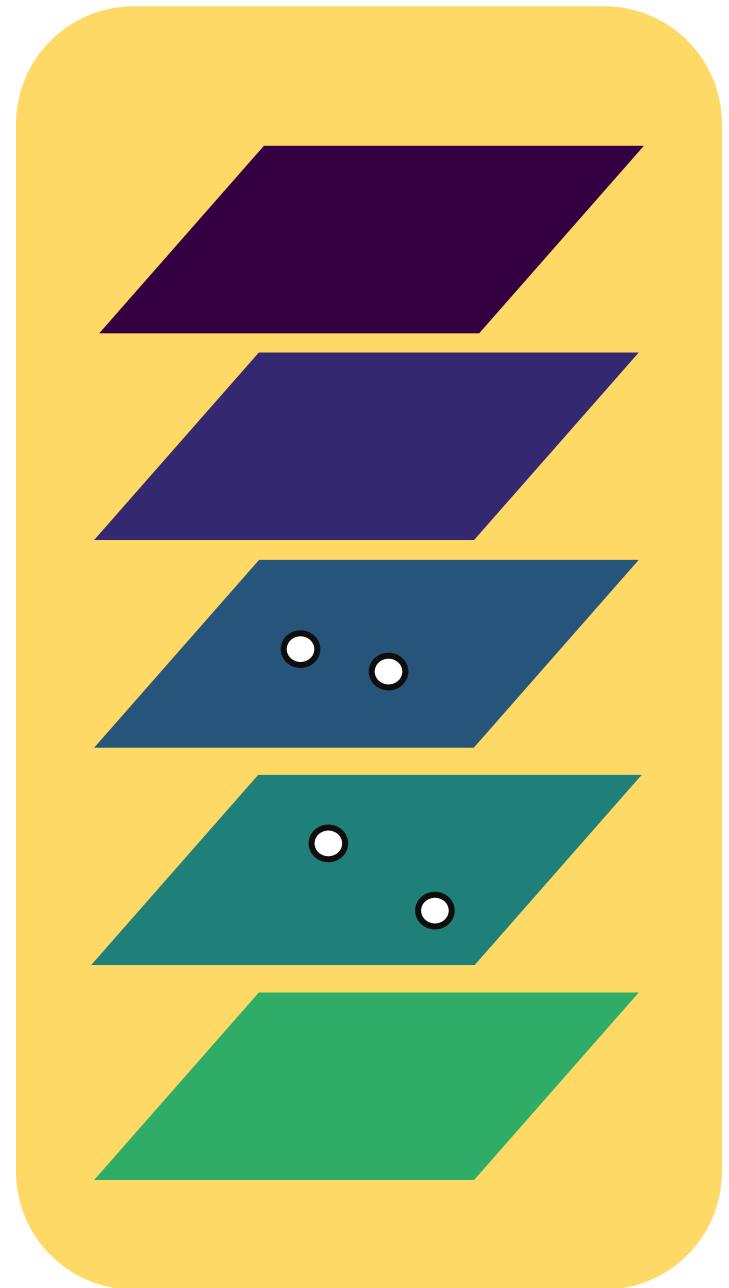
EPIGENÓMICA

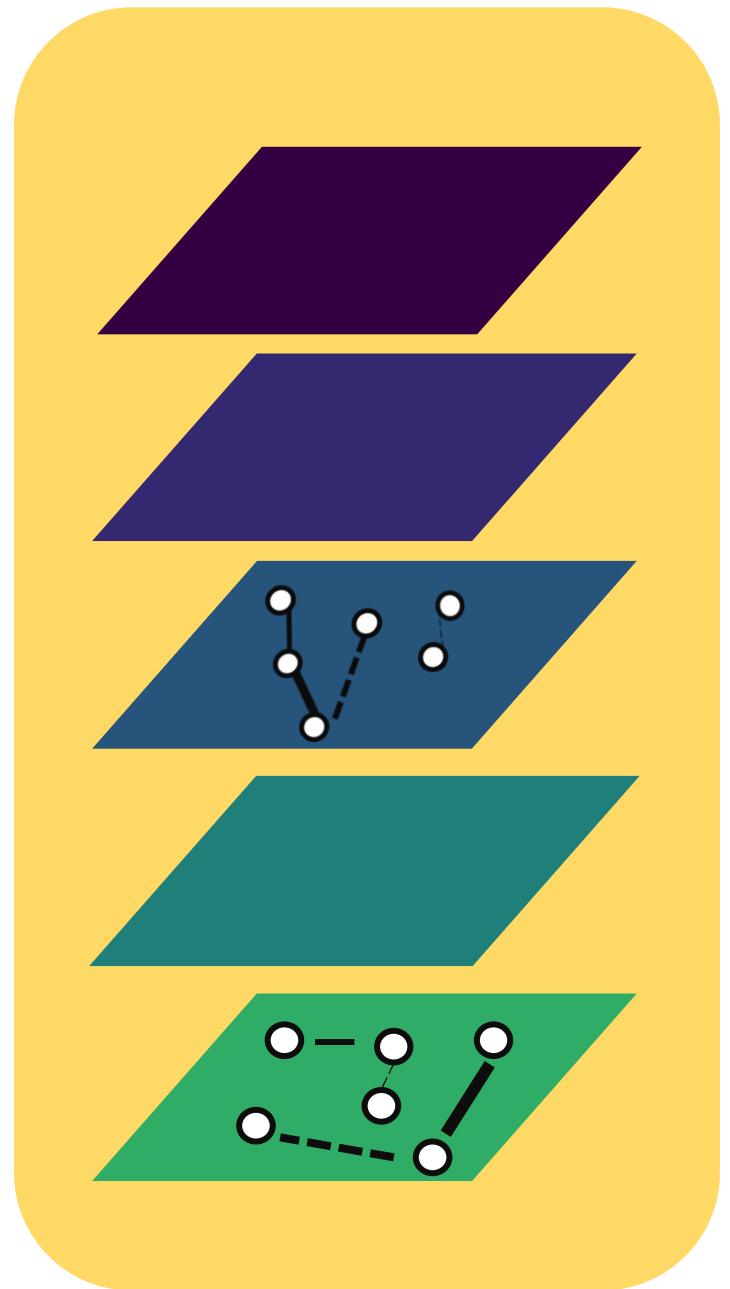
TRANSCRIPTÓMICA

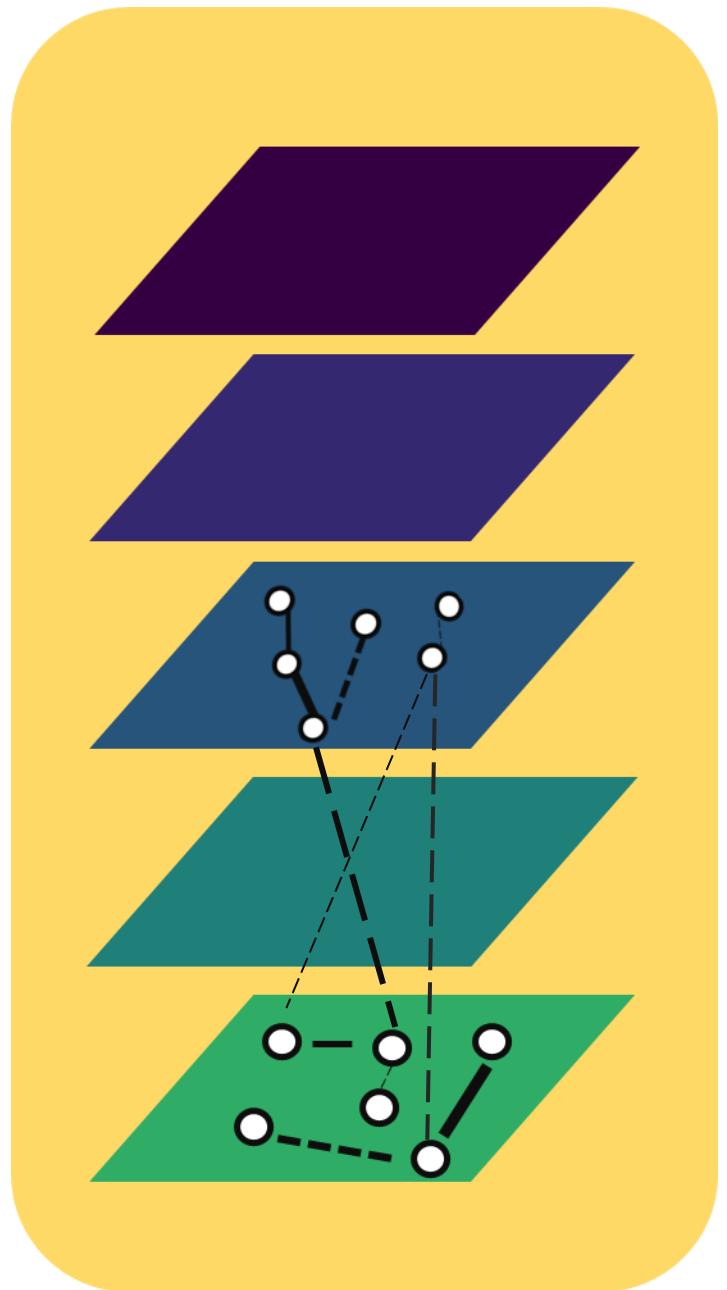
PROTEÓMICA

METABOLÓMICA

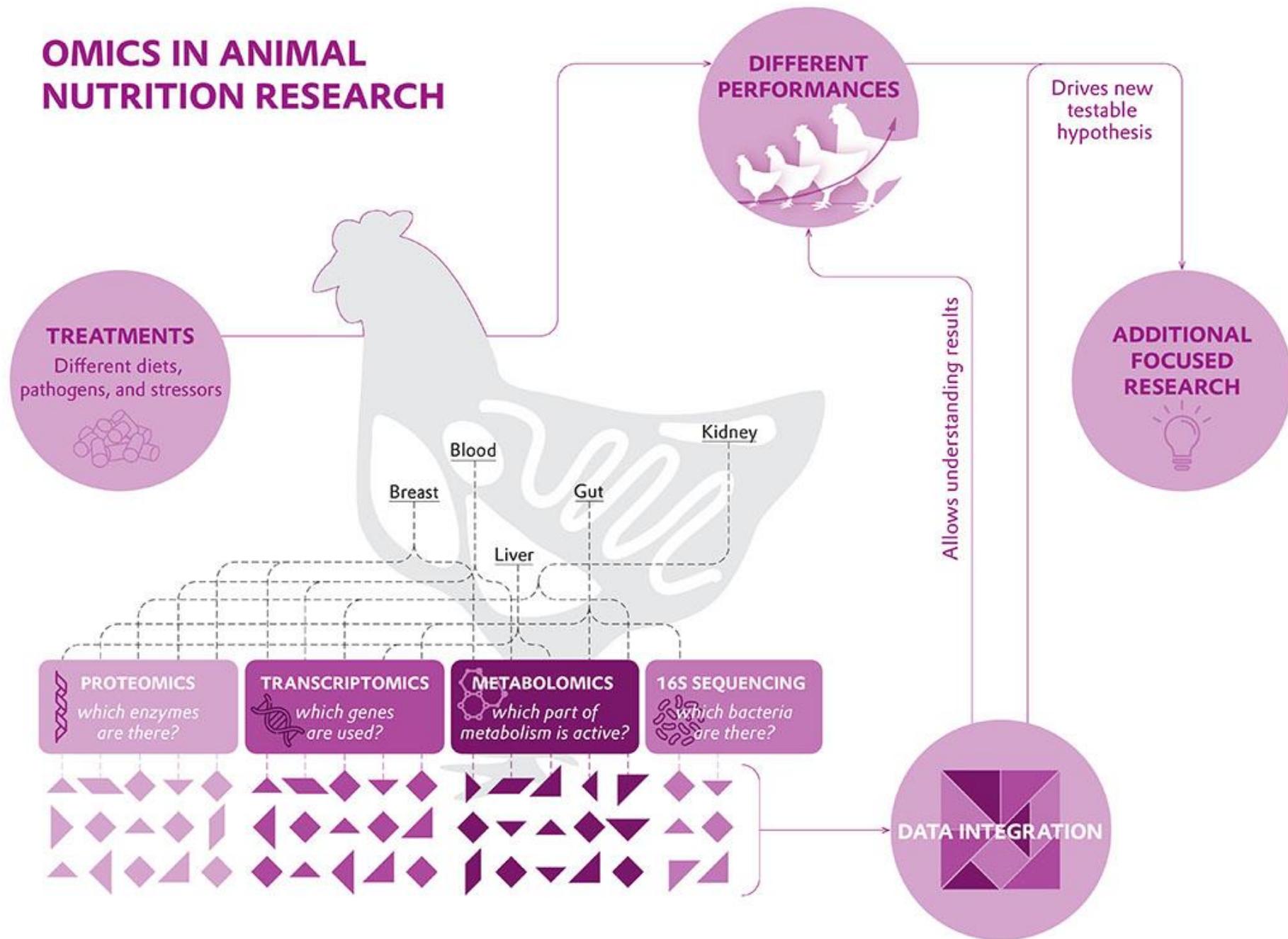






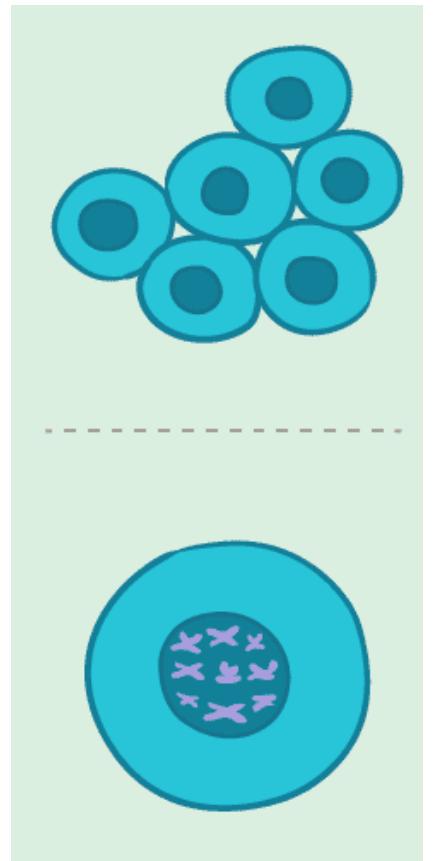
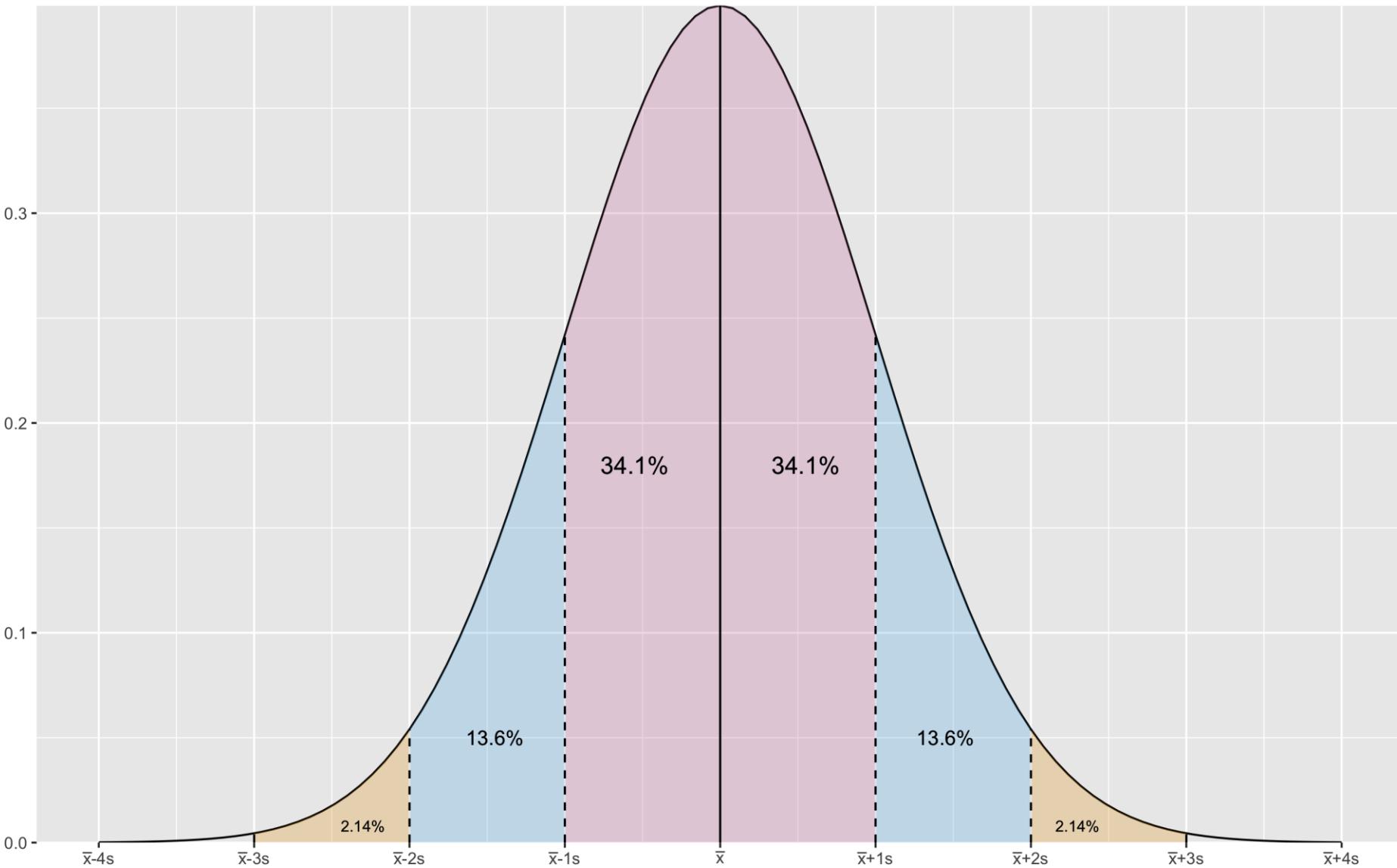


OMICS IN ANIMAL NUTRITION RESEARCH



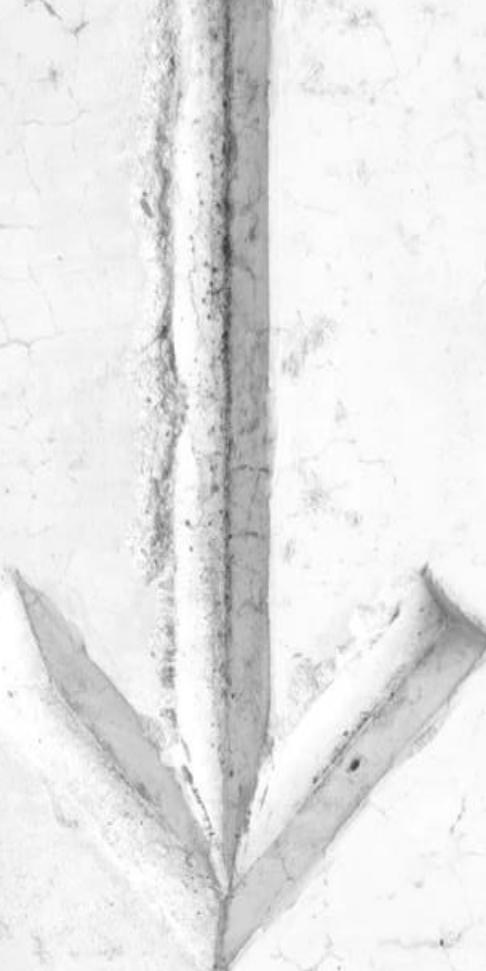


Normal Distribution





Orden de ejecución (y de errores)



1. Diseño experimental
2. Experimento
3. Datos crudos
4. Datos procesados
5. Análisis sobre datos procesados
6. Interpretación

4. Datos procesados

	Gen / Transcrito/ Proteína/ Metabolito A	Gen / Transcrito/ Proteína/ Metabolito B	Etc.
Muestra 1			
Muestra 2		0 vs NA	
Muestra 3			
Etc.			

Valor positivo



Cero



Missing value



GENÓMICA *METAGENÓMICA*

EPIGENÓMICA

TRANSCRIPTÓMICA

PROTEÓMICA

METABOLÓMICA

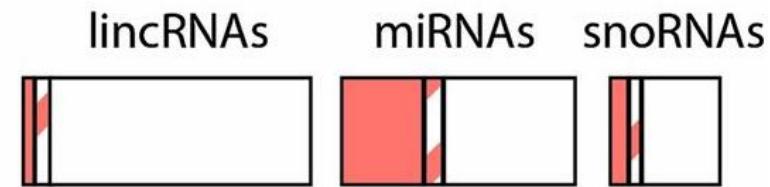
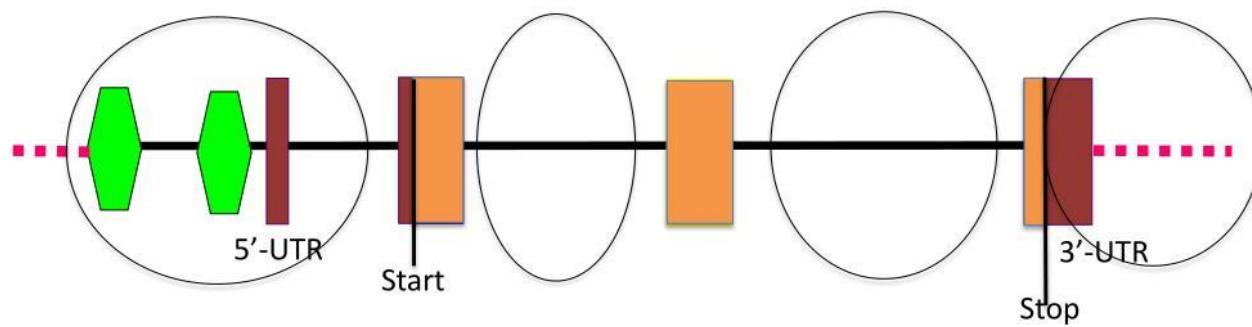


GENÓMICA



- Estudio de todo el genoma (WGS) o del exoma (EGS)

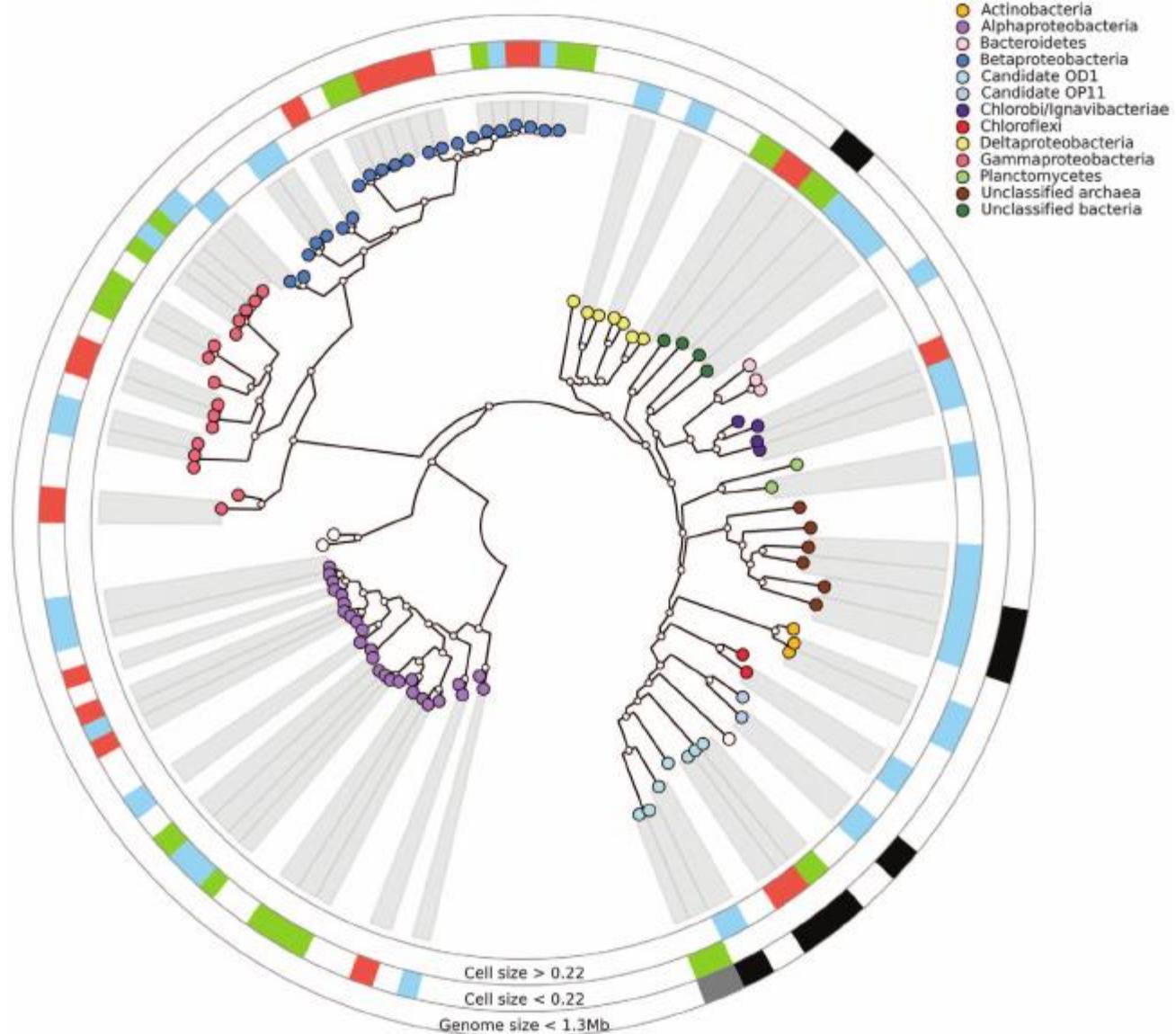
What Do You Miss With Whole Exome Sequencing?



~3-5% of Exons, Promoters, Untranslated Regions, and the Bulk of Intron Sequences are not Included in Exome Sequencing

- Estudio de todo el genoma (WGS) o del exoma (EGS)
- Instrumento: Secuenciador
- Archivo de salida del secuenciador: Fastq

- Estudio de todo el genoma (WGS) o del exoma (EGS)
- Instrumento: Secuenciador
- Archivo de salida del secuenciador: Fastq
- ¿Para qué?
 - Ensamblar genomas nuevos
 - Encontrar polimorfismos
 - Encontrar duplicaciones, delecciones, etc.
 - CNVs
- ¿Y después?



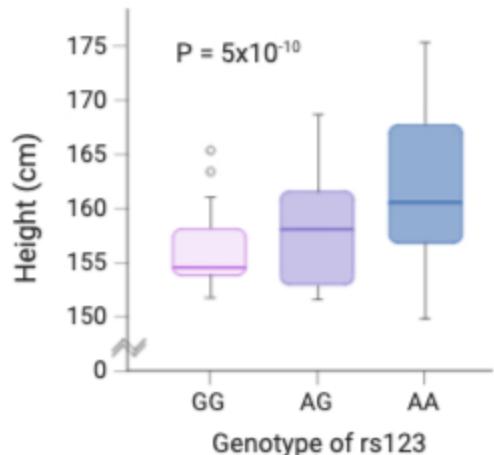
The Principle of a Genome-wide Association Study (GWAS)

- 1 Height and genetic data for individuals in study



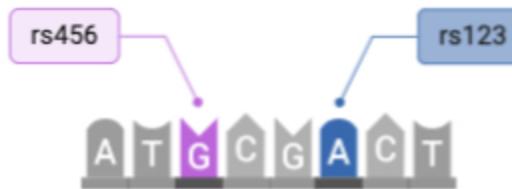
- 2 Single-variant association test with a candidate variant, rs123

A alleles increase height on average*

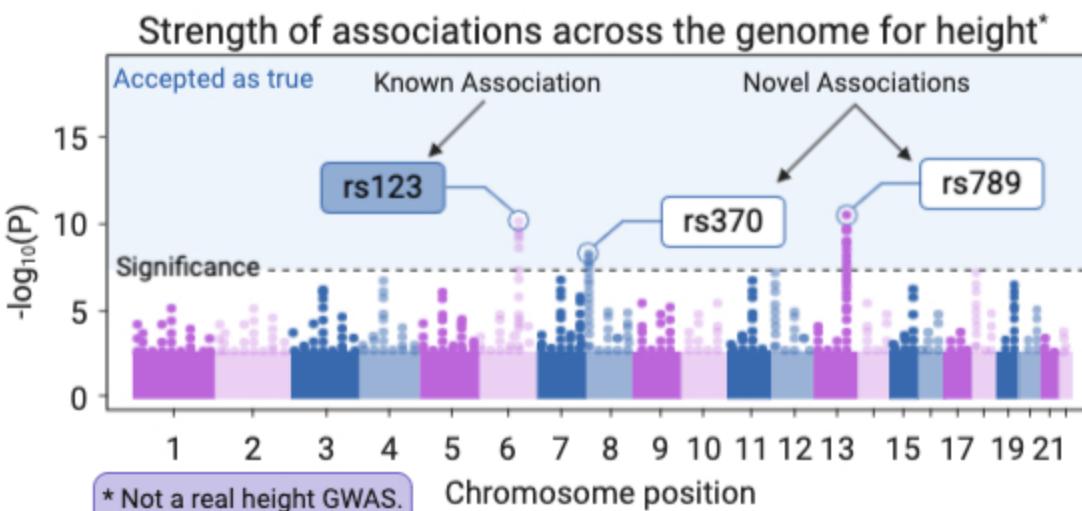


* Liberties were taken with the size of the allelic effect shown in this example.

- 3 rs123 may not be causal, but a measured proxy for the causal variant, rs456



- 4 GWAS use evenly-spaced proxies for association tests across the genome; plots show if $-\log_{10}(P)$ passes the genome-wide significance threshold

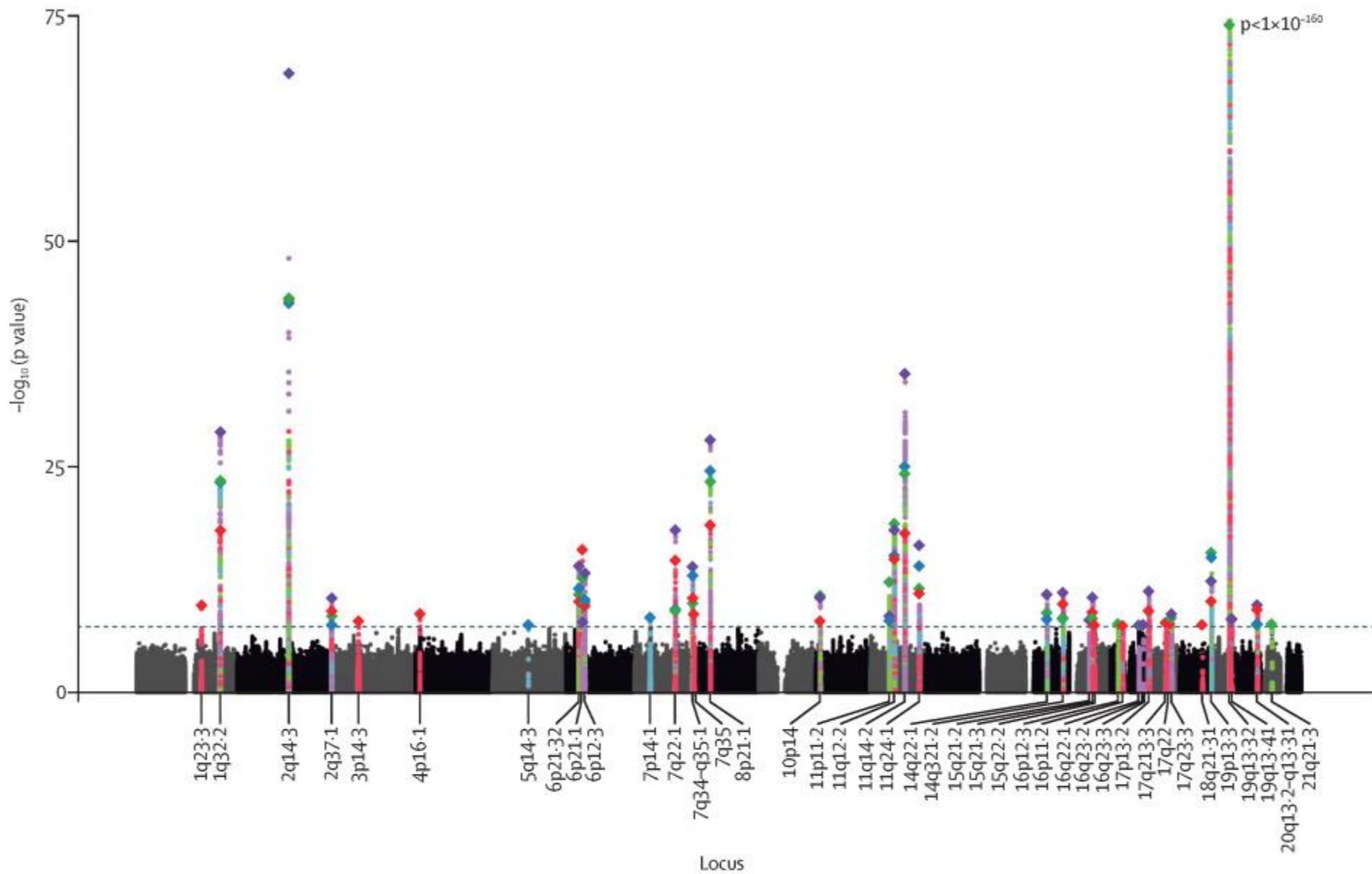


- 5 Further analyses of the regions can identify causal variants and their function

Fine-mapping

Functional work

Meta-analysis

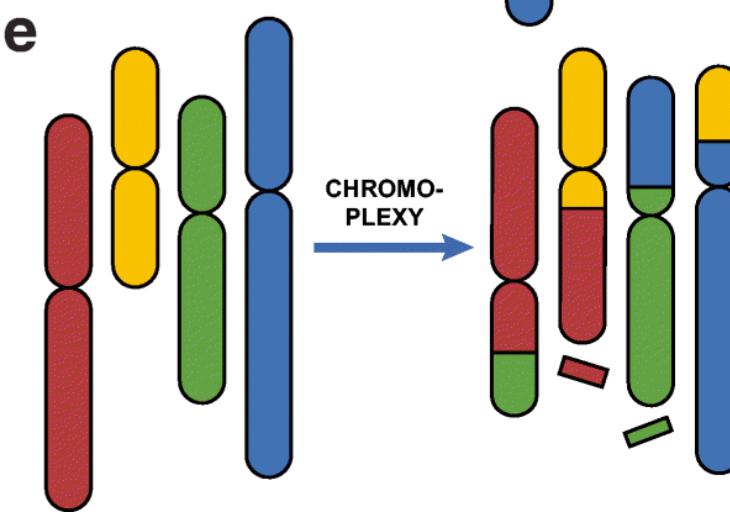
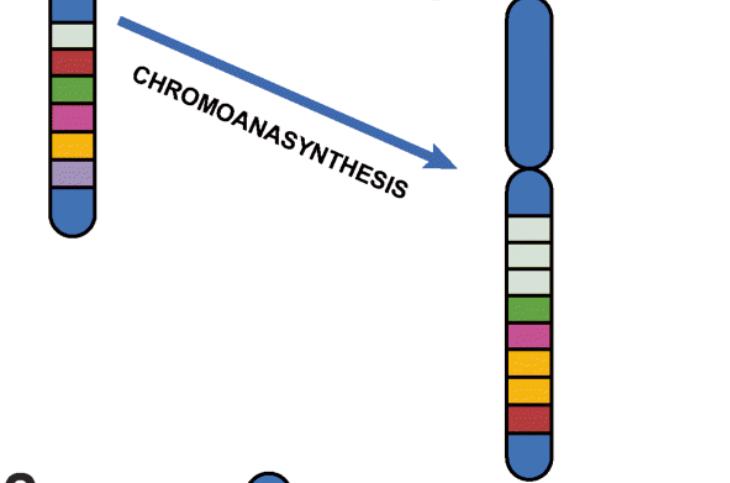
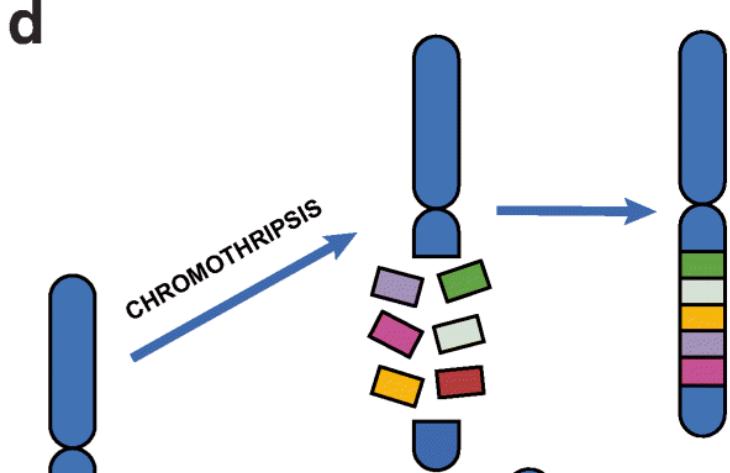
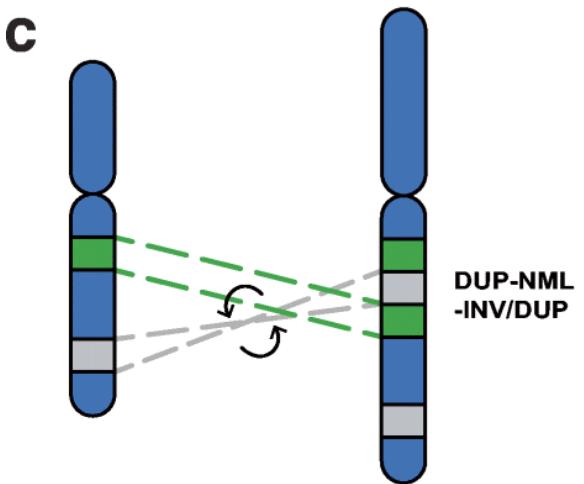
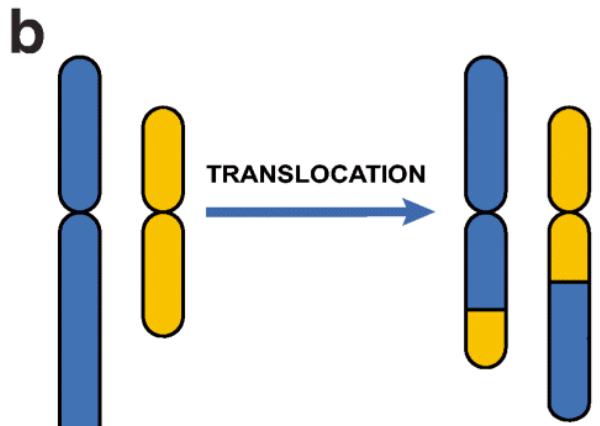
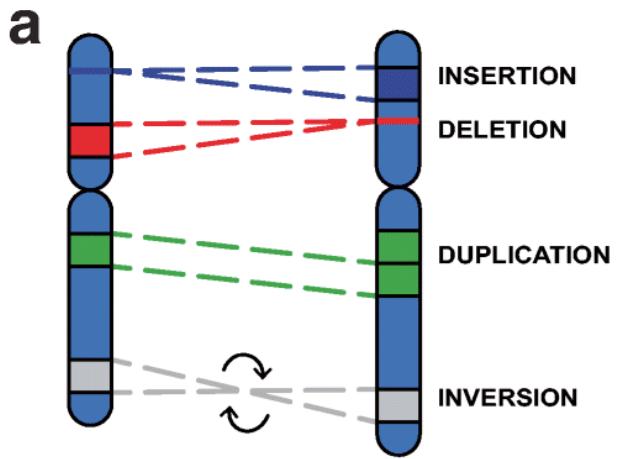


GWAS 2013⁵

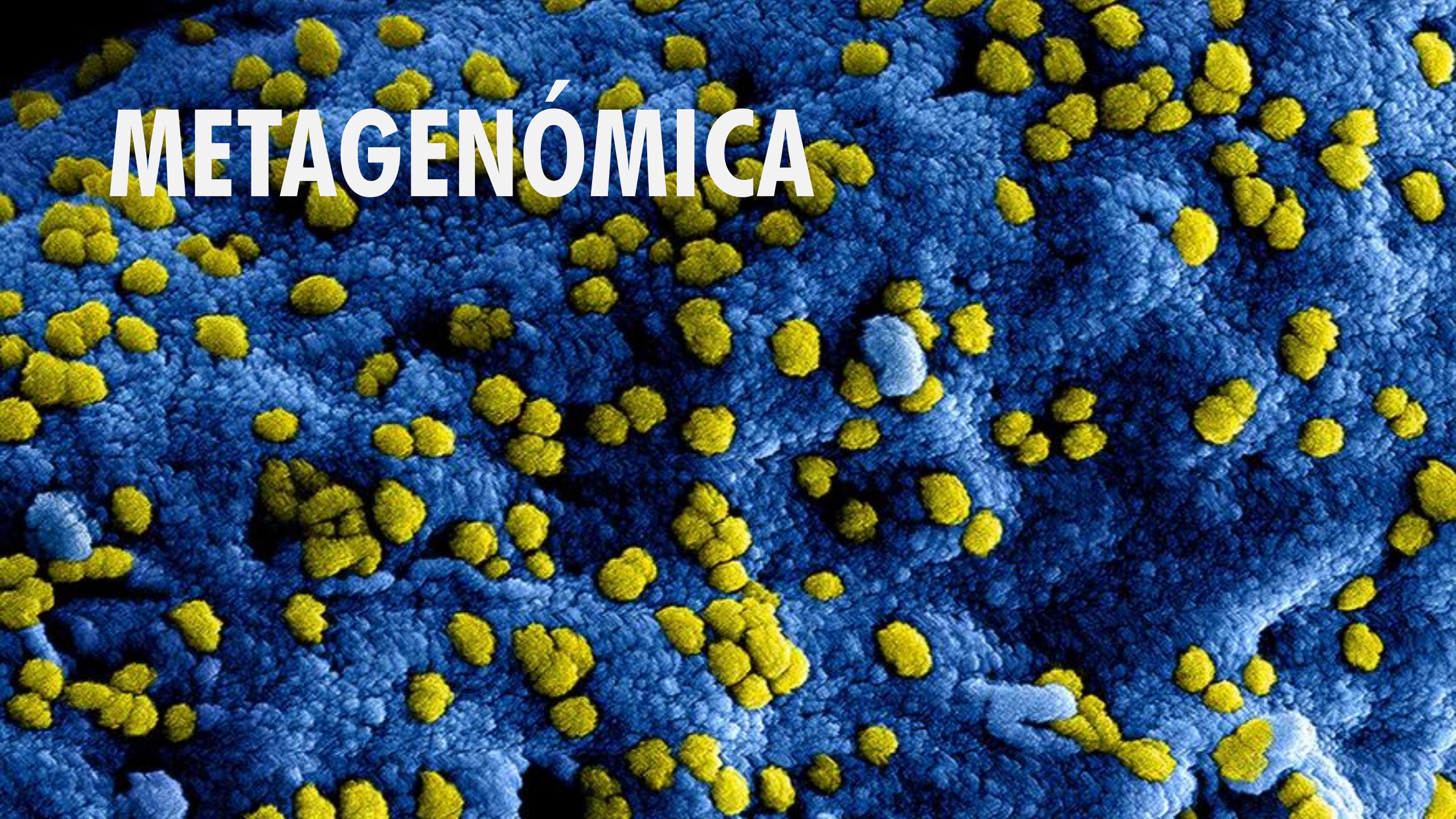
GWAS 2019⁹

GWAX 2018^B

GWAX 2019⁷



METAGENÓMICA





BACTERIA IN NATURE

eating literal dirt

defying the physical
limits of life

this is my third
eukaryotic
extinction event
in a row



BACTERIA IN THE LAB

not my favourite sugar ☹ ☹ ☹

the pH is off by 0.001

is this tap water? I'm allergic



- Estudio del microbioma, 16S vs WGS

16S vs. Shotgun Metagenomic Sequencing

	16S Sequencing	Shotgun Sequencing
Bacterial Coverage	High	Limited
Cross-Domain Coverage	No	Yes
False Positives	Low Risk	High Risk
Taxonomy Resolution	Genus-Species	Species-Strains
Host DNA Interference	No	Yes
Functional Profiling	No	Yes
Minimum DNA Input	10 copies of 16S	1 ng
Recommended Sample Type	All	Human Microbiome
Cost per Sample	~\$80	~\$200

- Estudio del microbioma, 16S vs WGS
- 16S: OTUs vs ASVs

16S Sequencing Challenges

Targeted sequence with a few bases differentiating species

- Sequencing is imperfect
 - Illumina usually makes some base call errors
 - Nanopore makes more
 - Errors are not necessarily evenly-distributed
- We do not want errors to be confused with real diversity/new species

FOR OUR PURPOSES:

 = 1 sequence* and one species
*May contain errors

REAL BACTERIA

 One species can have multiple, different copies of a gene
 Two similar species may share an identical sequence

- Estudio del microbioma, 16S vs WGS
- 16S: OTUs vs ASVs

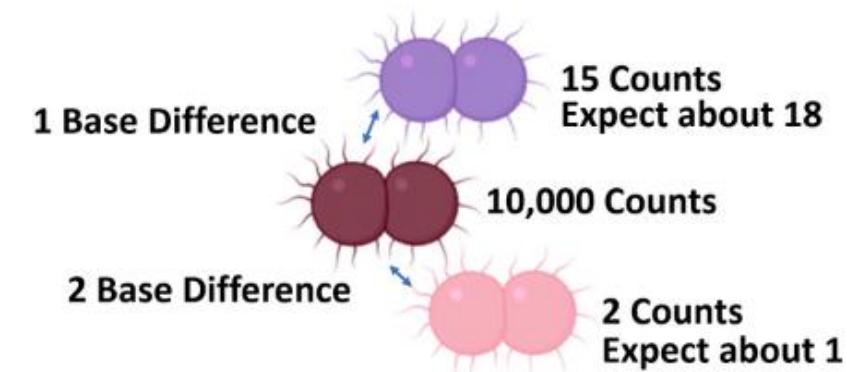
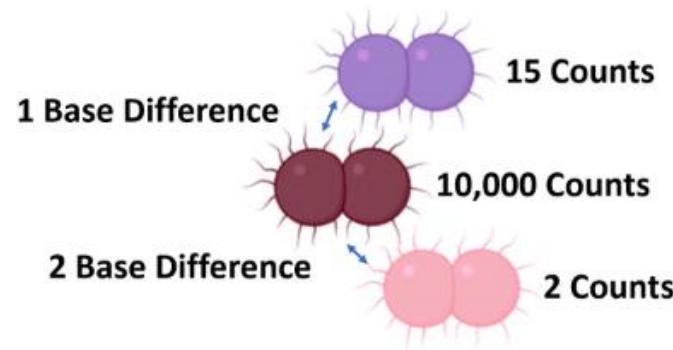
Operational Taxonomic Unit (OTU) Approach

We know some of these sequences arose from error/artifact

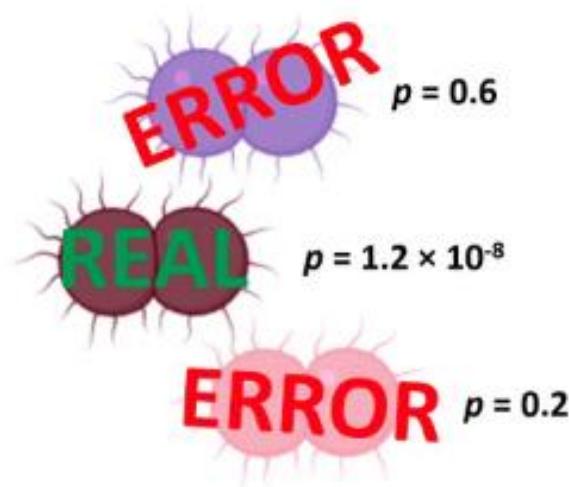
Combine extremely similar sequences (usually 97% or more identity) to minimize the effect of observed errors. Then treat each OTU as a representative sequence.

- OTUs can be represented by a representative consensus sequence
 - Closed reference OTUs are fast to create, but are subject to reference bias
 - *De novo* OTUs are free of reference bias, but computationally expensive and can change with changed samples
 - Open reference OTUs are in between, with sequence similar to reference behaving like closed reference and more novel sequences behaving like *de novo*
-

- Estudio del microbioma, 16S vs WGS
- 16S: OTUs vs ASVs

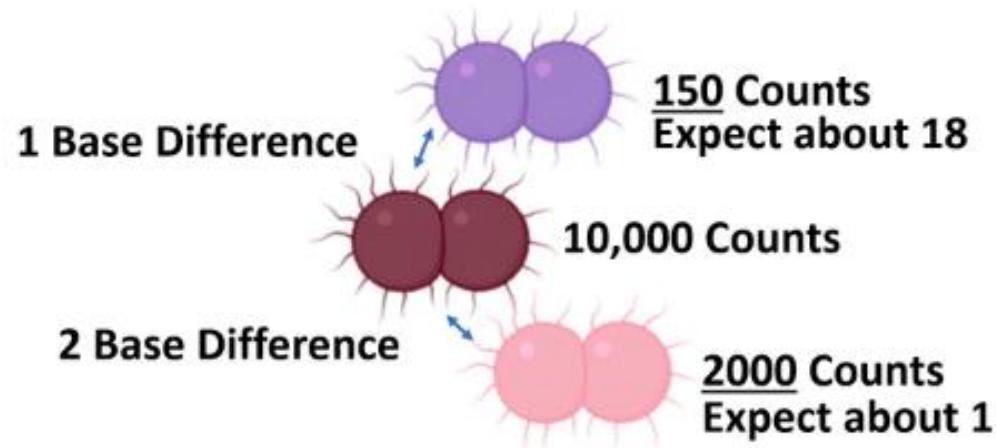


- Estudio del microbioma, 16S vs WGS
- 16S: OTUs vs ASVs



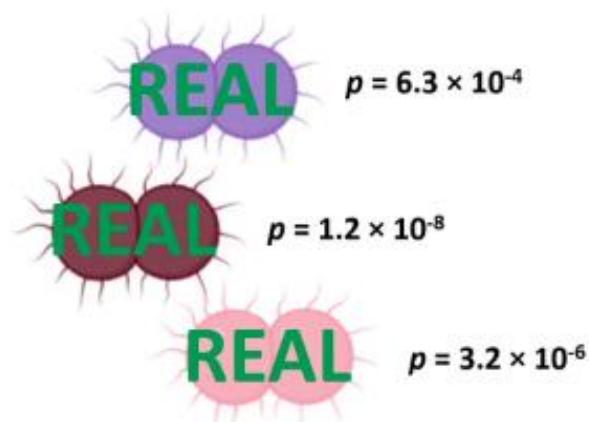
- Estudio del microbioma, 16S vs WGS
- 16S: OTUs vs ASVs

What If The Counts Were Far Higher?



- Estudio del microbioma, 16S vs WGS
- 16S: OTUs vs ASVs

What If The Counts Were Far Higher?



- Estudio del microbioma, 16S vs WGS
- 16S: OTUs vs ASVs

Amplicon Sequence Variant (ASV) Approach

What is the statistical support for each sequence's existence?

Throw out amplicon sequences that lack strong statistical support for not being artifacts of sequencing. Cost: potential loss of real sequence that was present at very low levels.

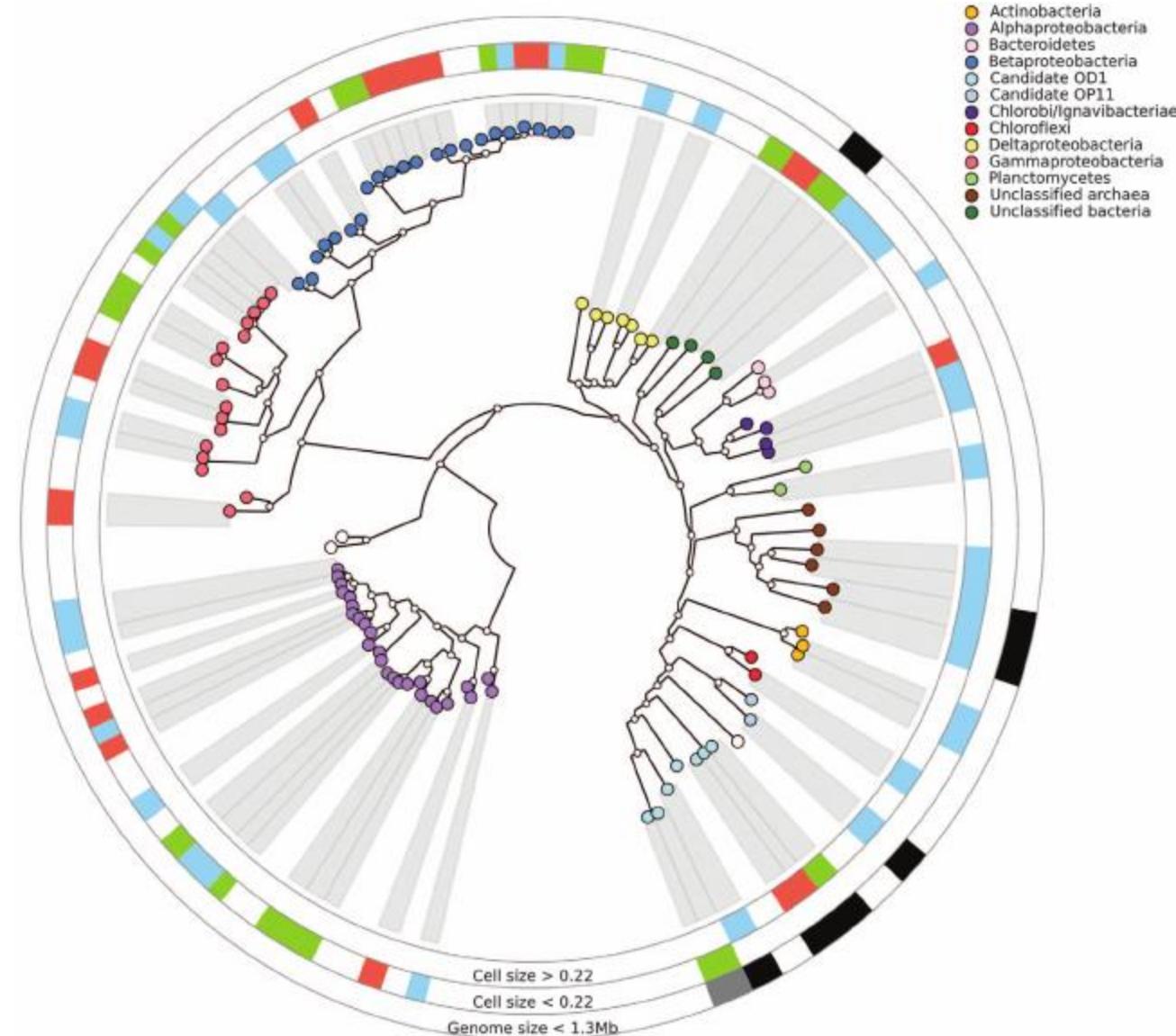
- No representative consensus sequence
 - Each sequence is supported as being present in the sample
- Potentially higher resolution, no potential to combine multiple “real” sequences into an abstract
- Generated without the use of a reference, no risk of reference bias.
- May also be called an ESV (exact sequence variant) or zOTU (zero-radius OTU)

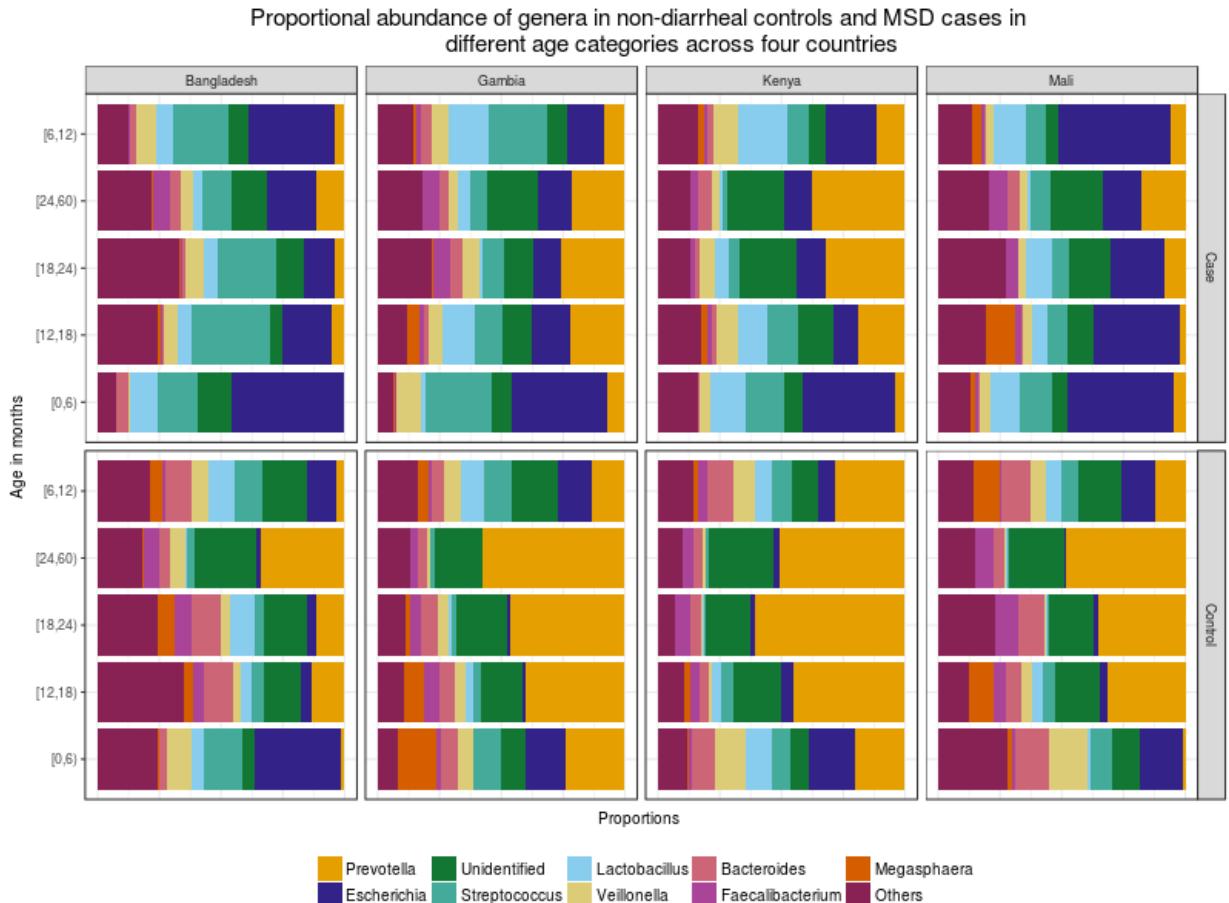
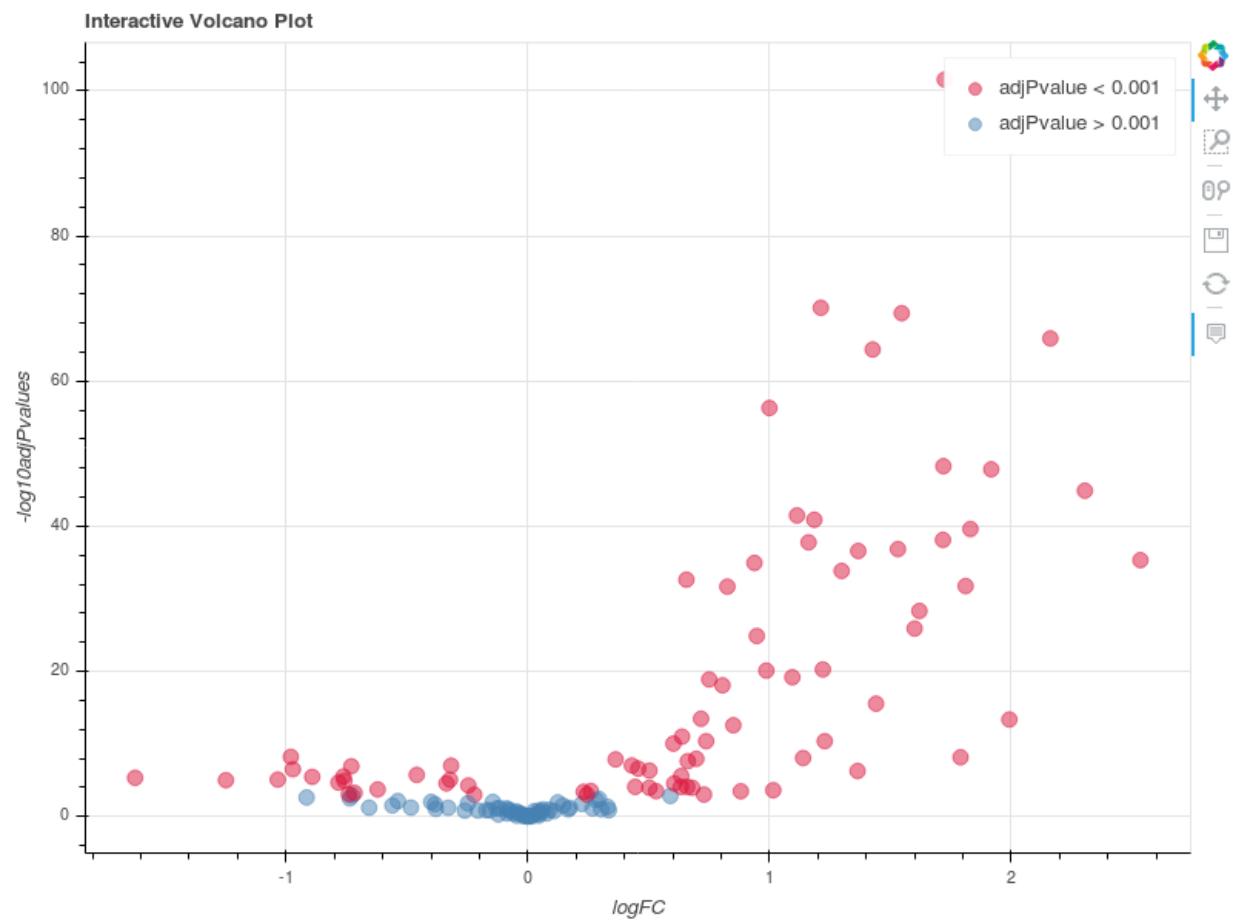
- Estudio del microbioma, 16S vs WGS
- 16S: OTUs vs ASVs

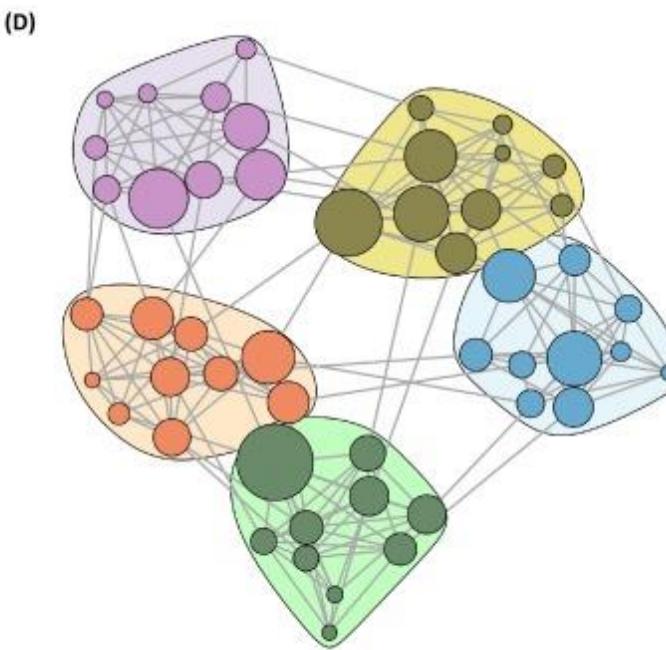
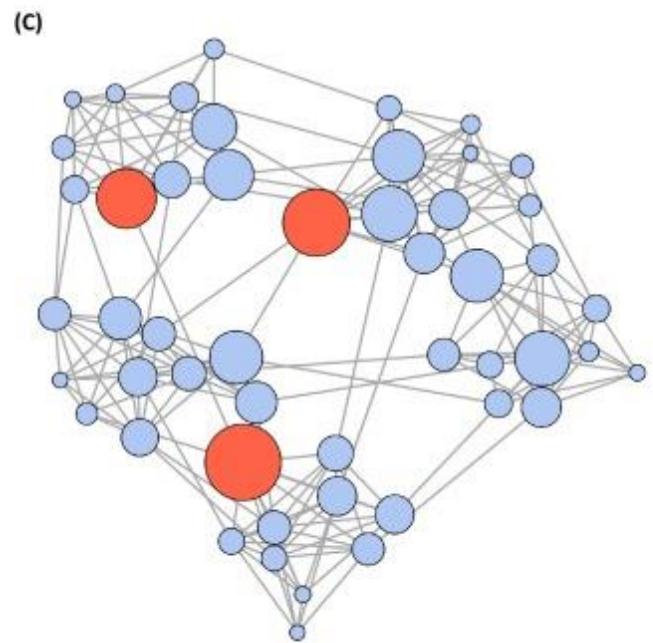
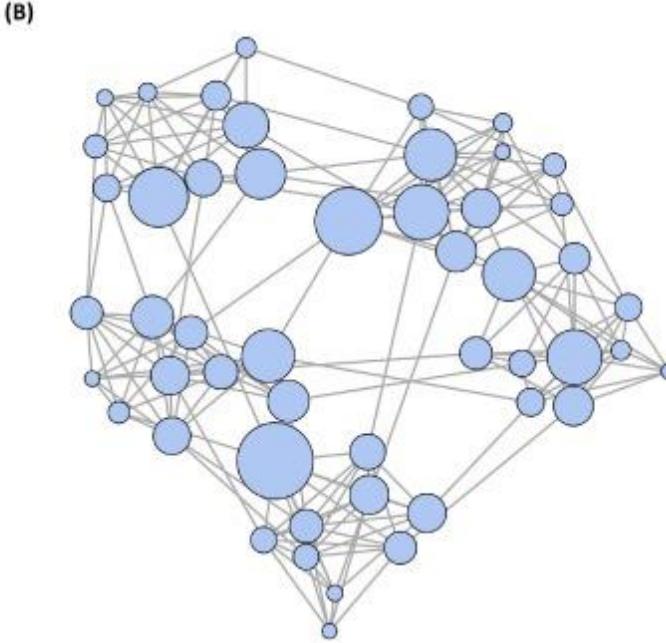
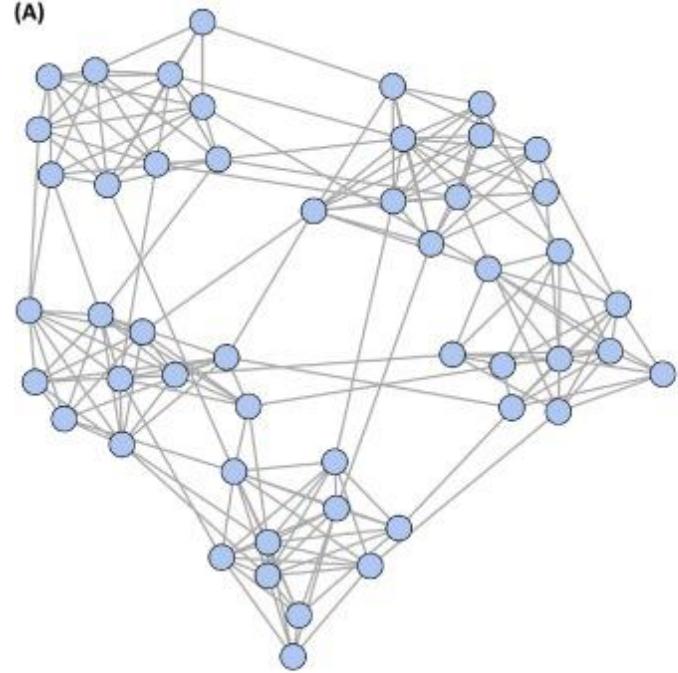
OTU	ASV
Can be subject to reference bias	Reference is not used until taxonomy assignment
OTU tables cannot be combined between studies	ASV tables can be compared across studies
Represented by a consensus sequence	Represented by an exact sequence
Can represent multiple species with different sequences	If it represents multiple species, it is because they share the sequence
Subject to chimeric sequences	Subject to chimeric sequences
Chimera detection can be complex and may require reference bias	Chimera detection is simple and reference-free

- Estudio del microbioma, 16S vs WGS
- 16S: OTUs vs ASVs
- Instrumento: Secuenciador
- Archivo de salida del secuenciador: Fastq
- ¿Pa qué?
 - Determinar composición microbiana de la muestra

- Estudio del microbioma, 16S vs WGS
- 16S: OTUs vs ASVs
- Instrumento: Secuenciador
- Archivo de salida del secuenciador: Fastq
- ¿Pa qué?
 - Determinar composición microbiana de la muestra
- ¿Y después?







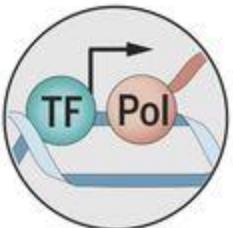
EPIGENÓMICA

ON



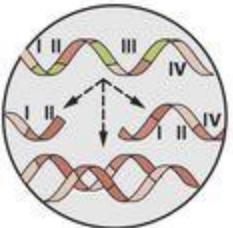
OFF

- Estudio del epigenoma, modificaciones reversibles del ADN o de las histonas que afectan a la expresión génica sin alterar la secuencia de ADN.
- Instrumento: Secuenciador
- Archivo de salida del secuenciador: Fastq
- ¿Para qué?
 - Metilación del ADN
 - Modificaciones de las histonas
 - ¿Dónde se une X factor de transcripción?
 - Modificaciones/estructura de la cromatina
 - Modificaciones/estructura de los cromosomas
- ¿Y después?

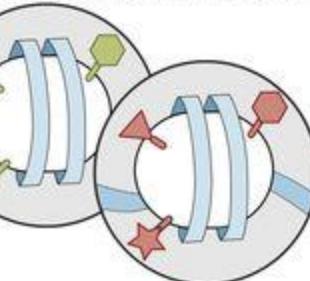


Transcription factor binding

TF binding interacts with DNA methylation and chromatin accessibility



Transcription and RNA maturation



Histone modifications

Modifications can be active marks (e.g., H3K4me3 in green) or repressive marks (e.g., H2K27m3 in red)



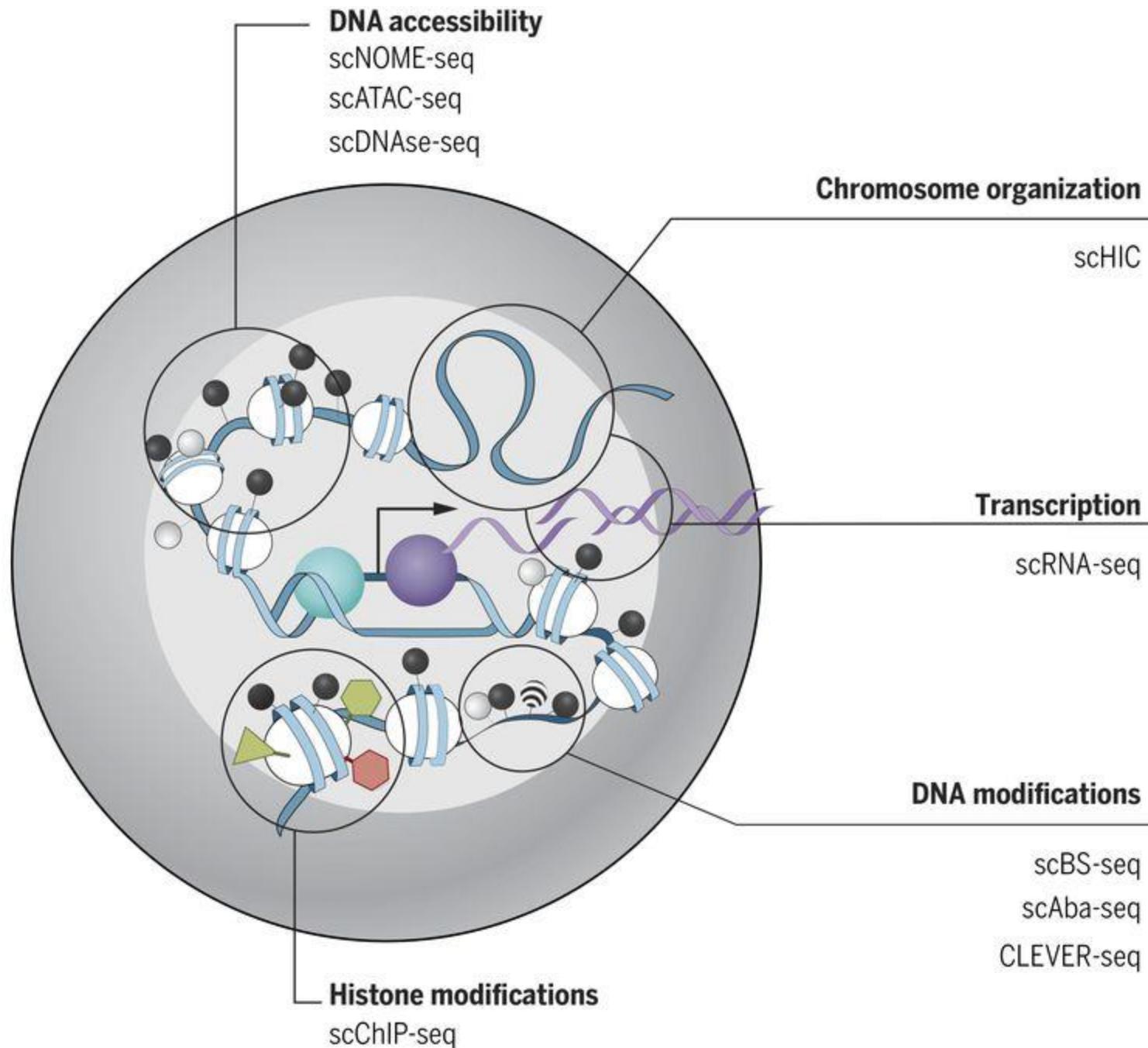
DNA modifications

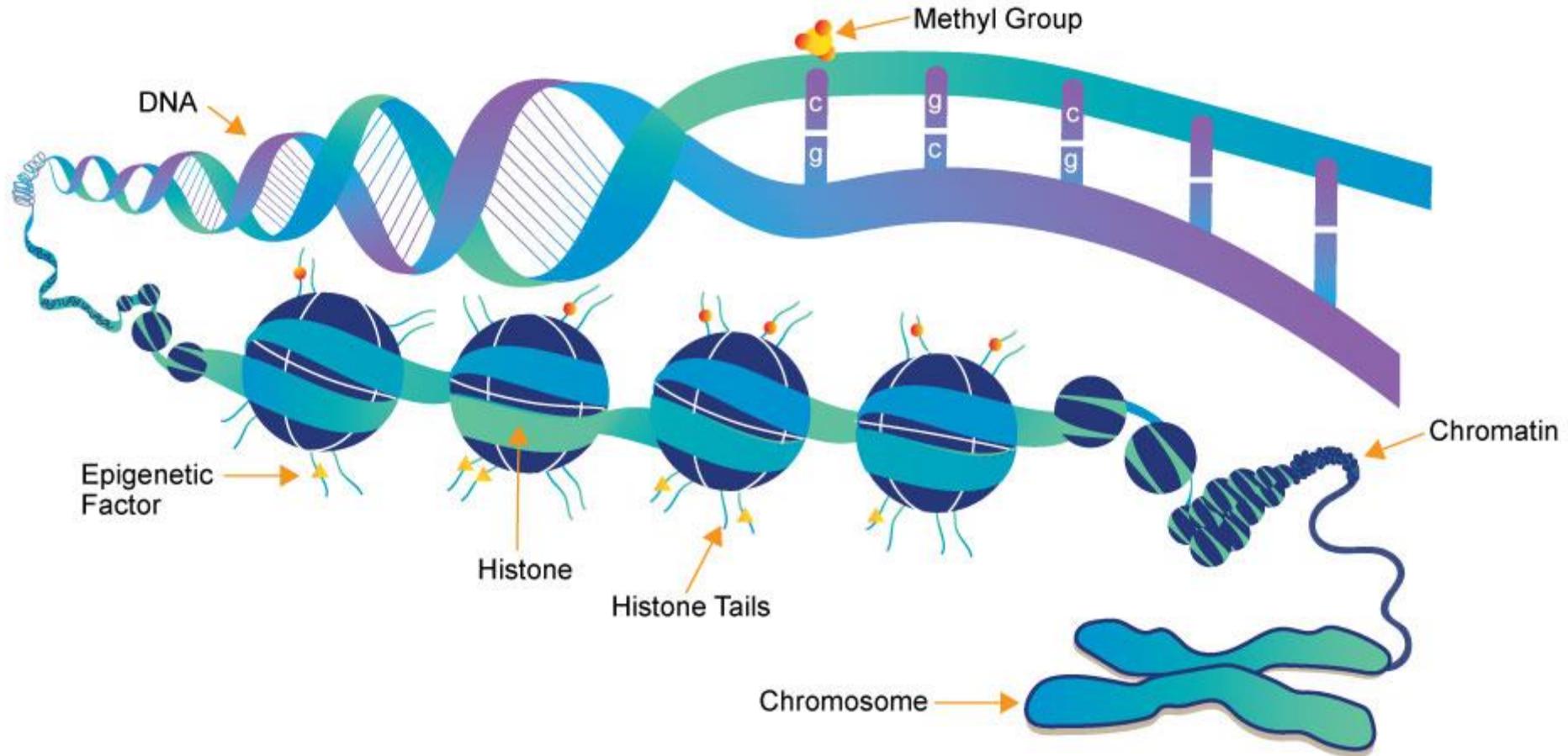
- C
- 5mC
- 5hmC / 5fC / 5caC



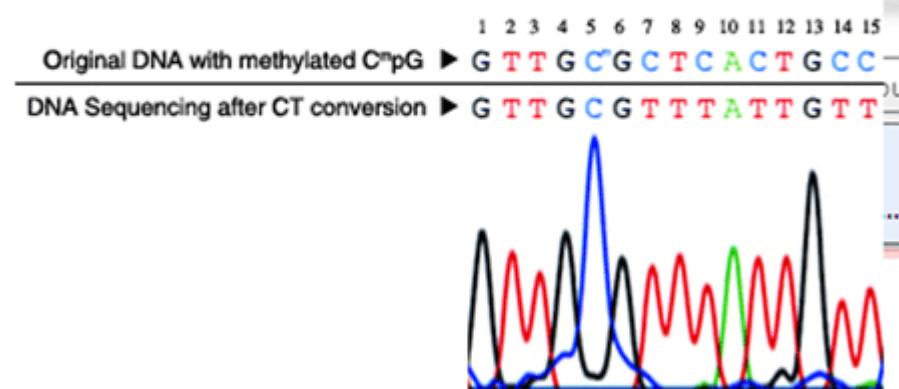
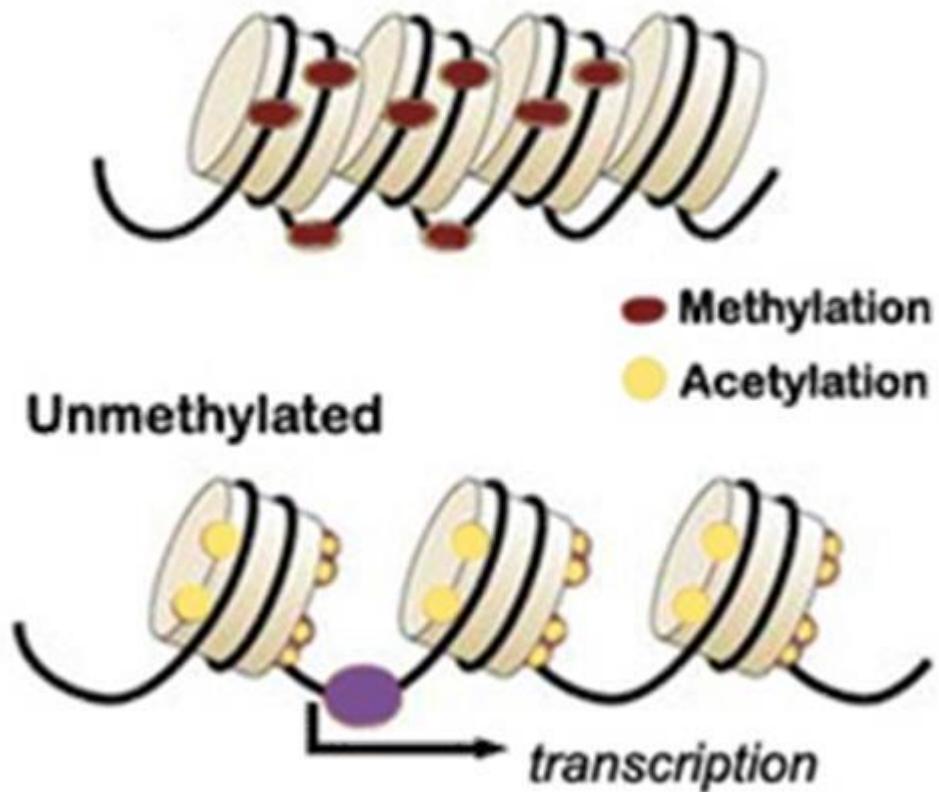
Chromosome organization

Higher-order chromatin organization into LADs and TADs

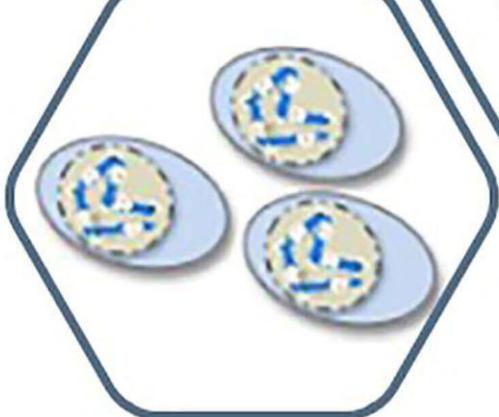




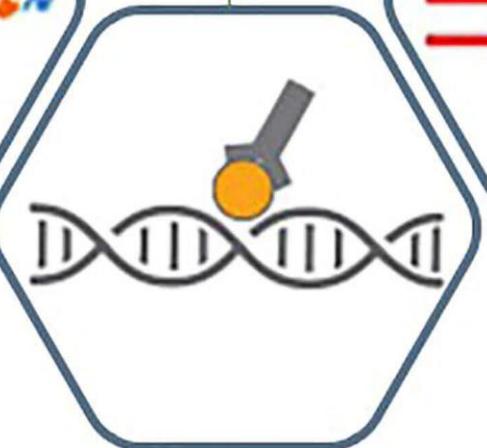
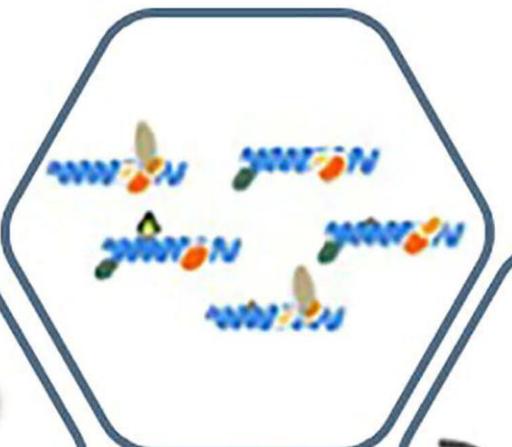
Methylated DNA



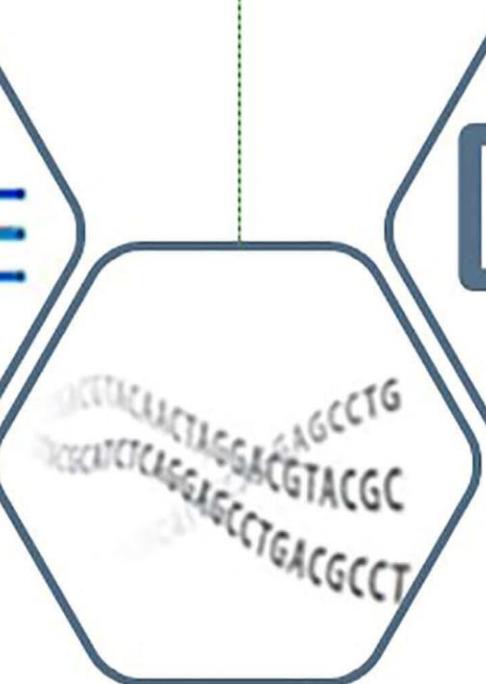
➤ Cell Preparation



➤ Immunoprecipitation



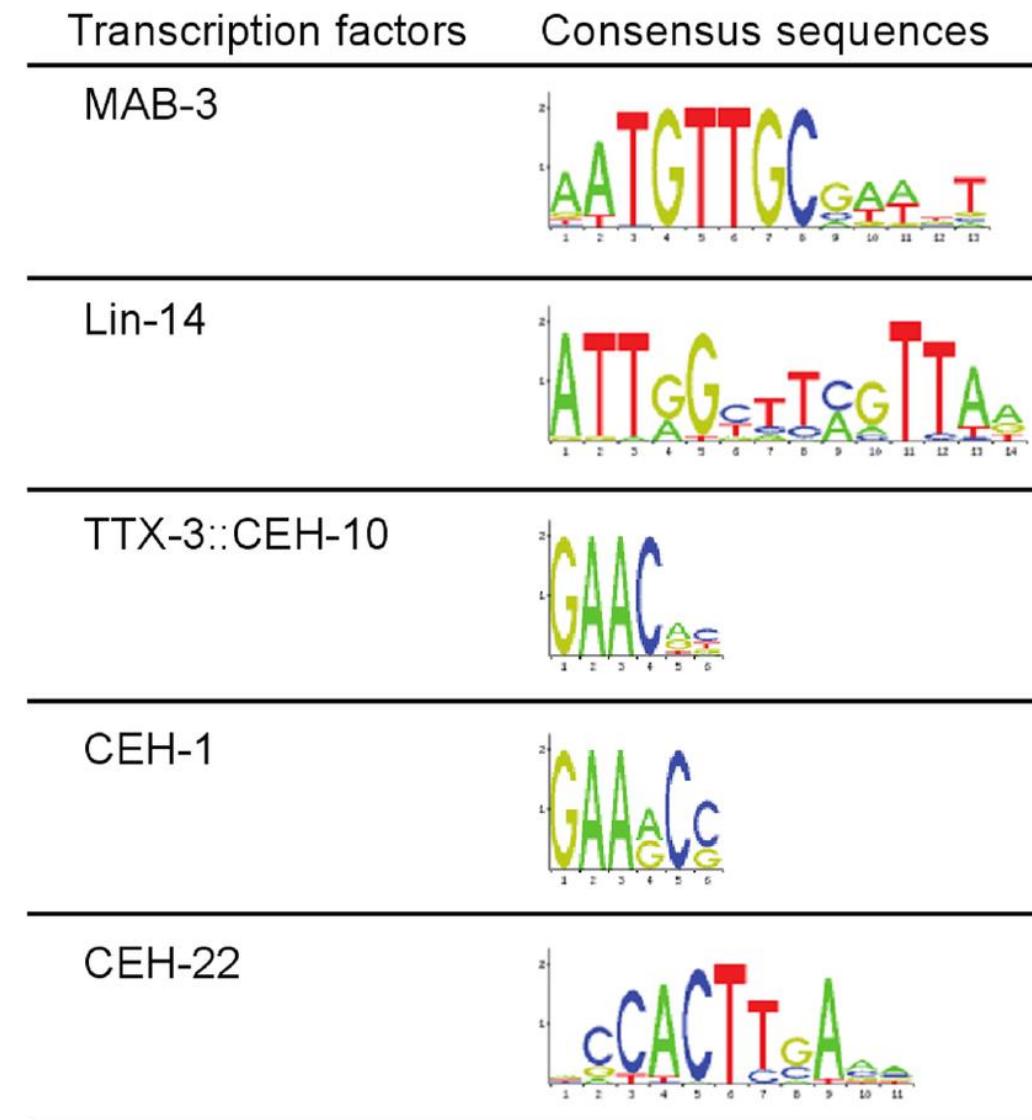
➤ Sequencing



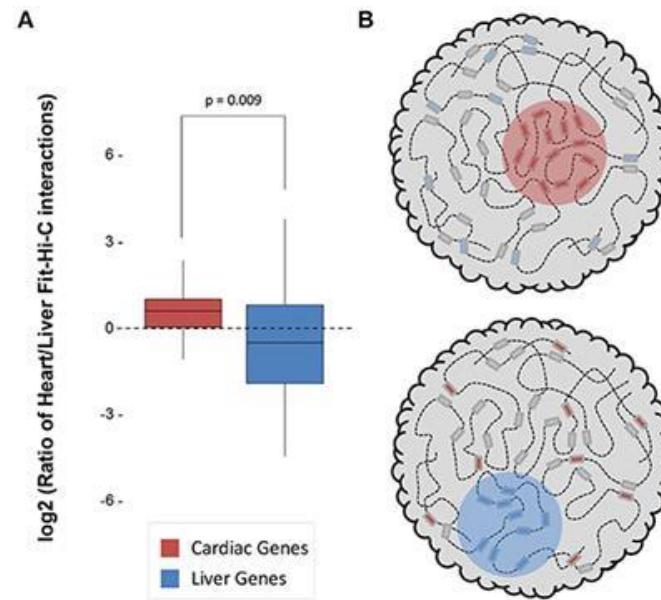
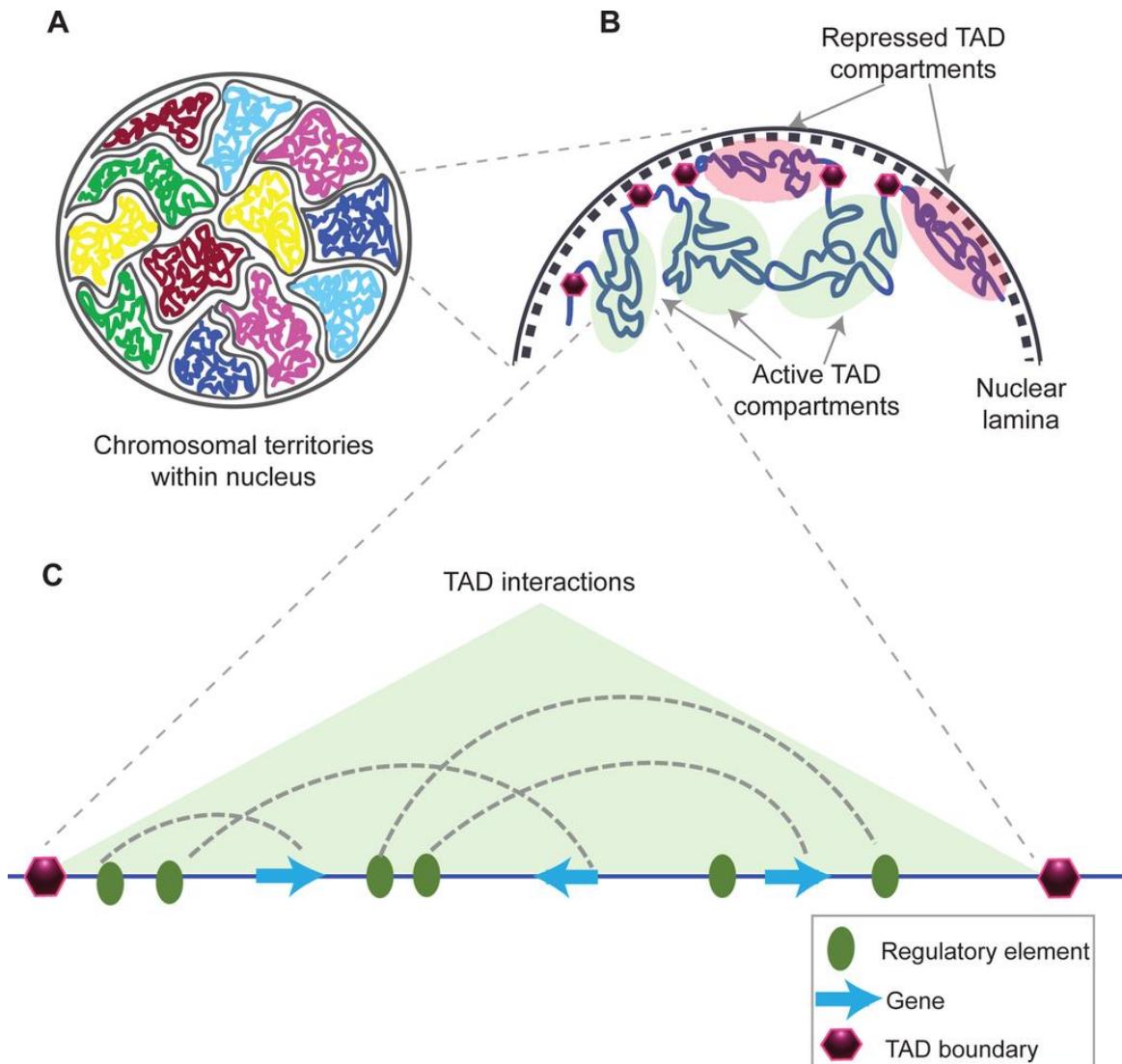
➤ Isolate and
Sonicate Chromatin

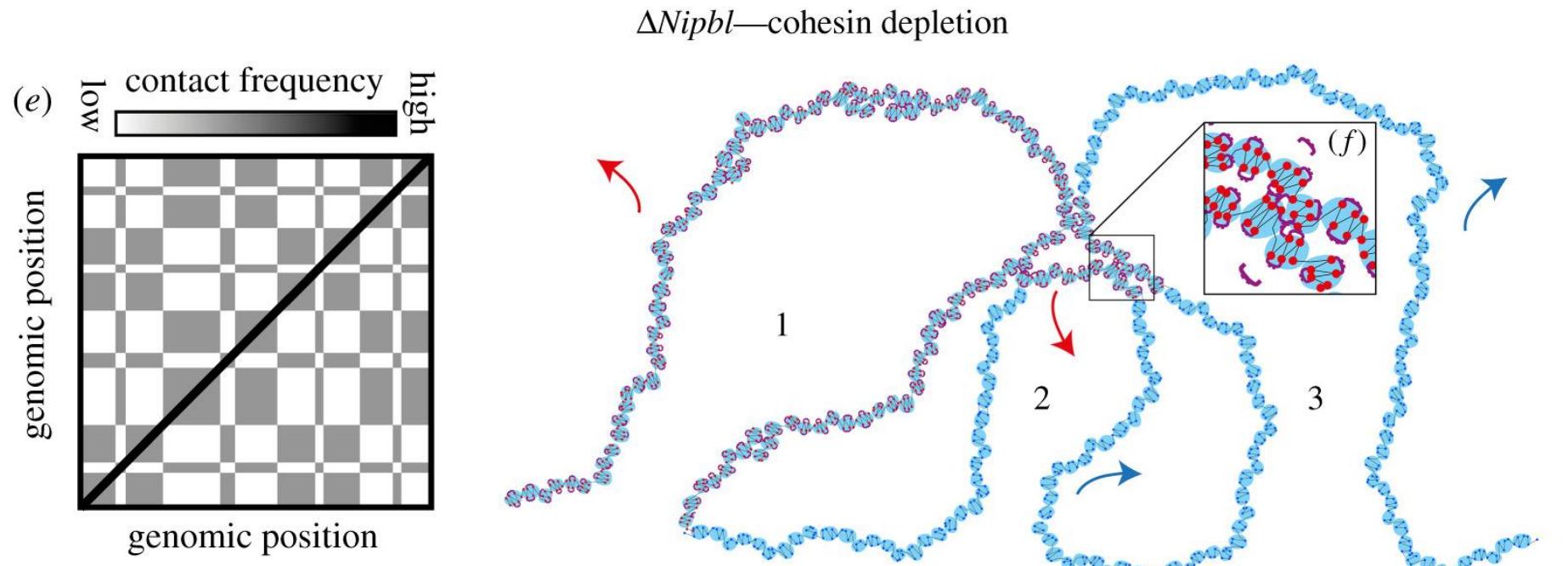
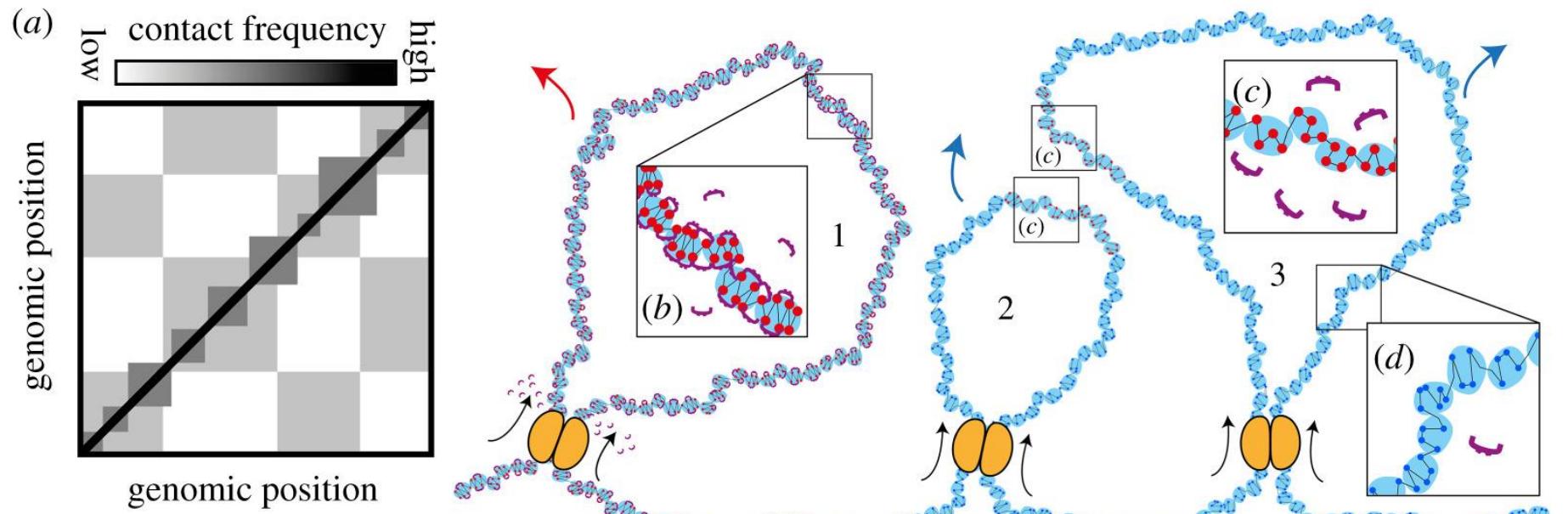
➤ Library Preparation

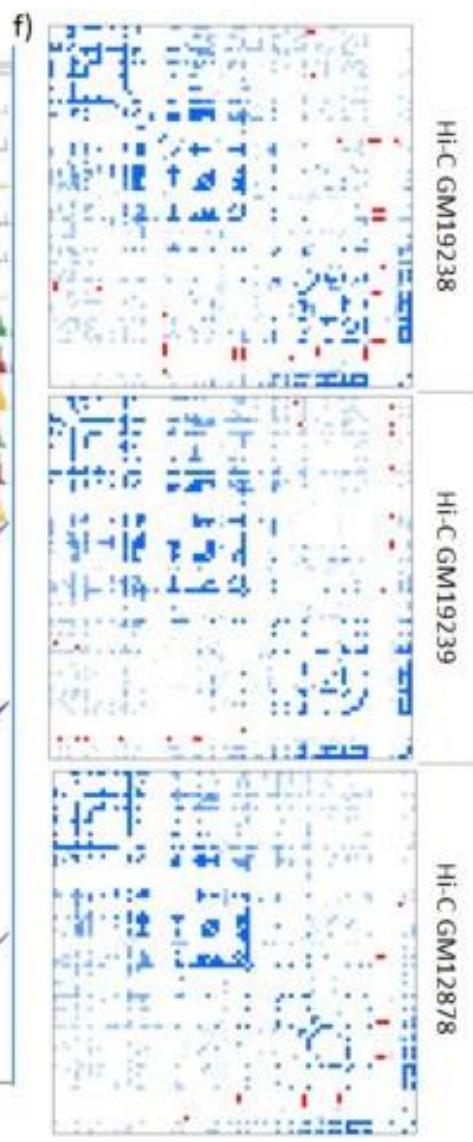
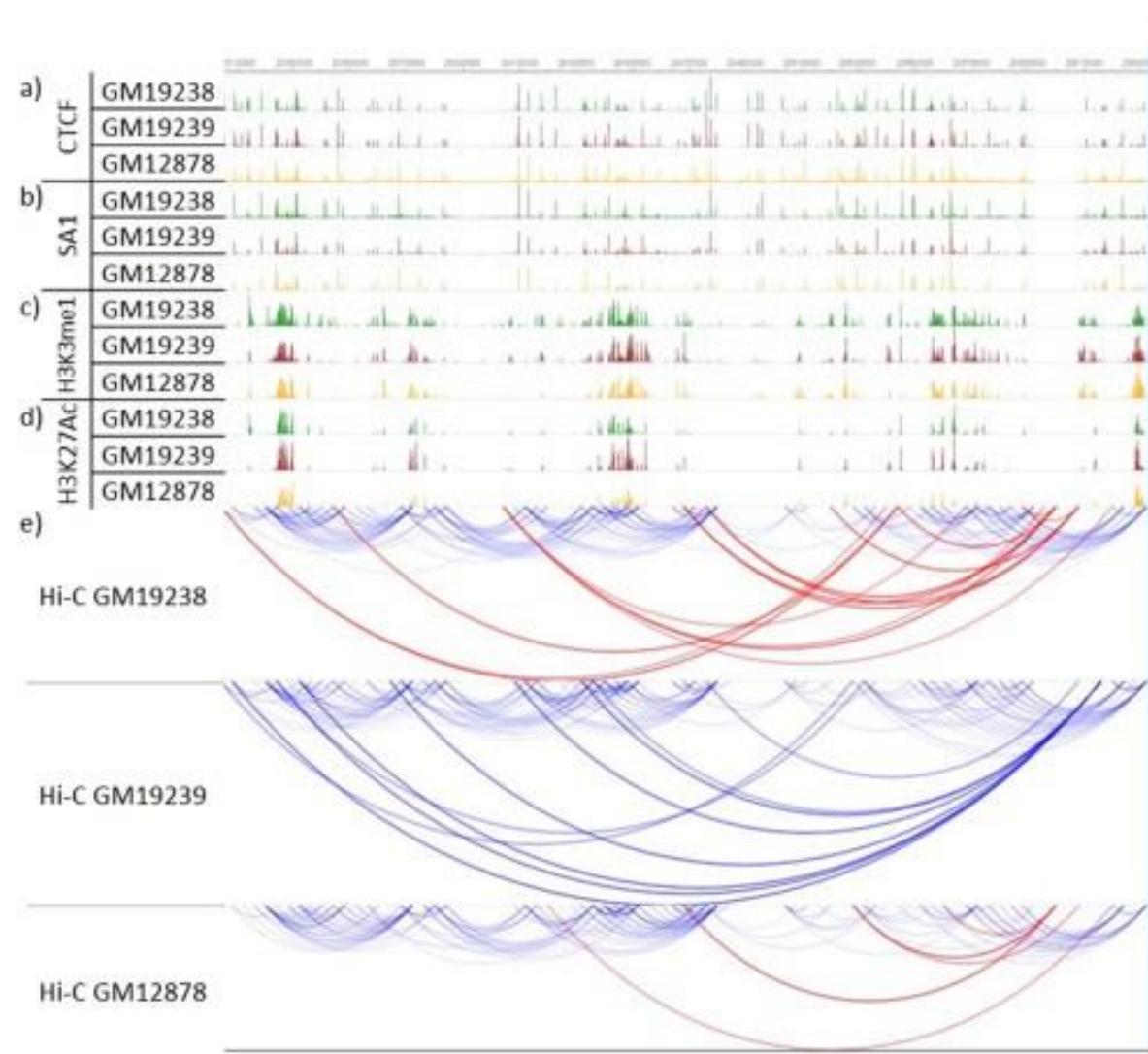
➤ Bioinformatics
Analysis



Un dominio asociado topológicamente (TAD) es una región genómica que interactúa automáticamente, lo que significa que las secuencias de ADN dentro de un TAD interactúan físicamente entre sí con más frecuencia que con secuencias fuera del TAD





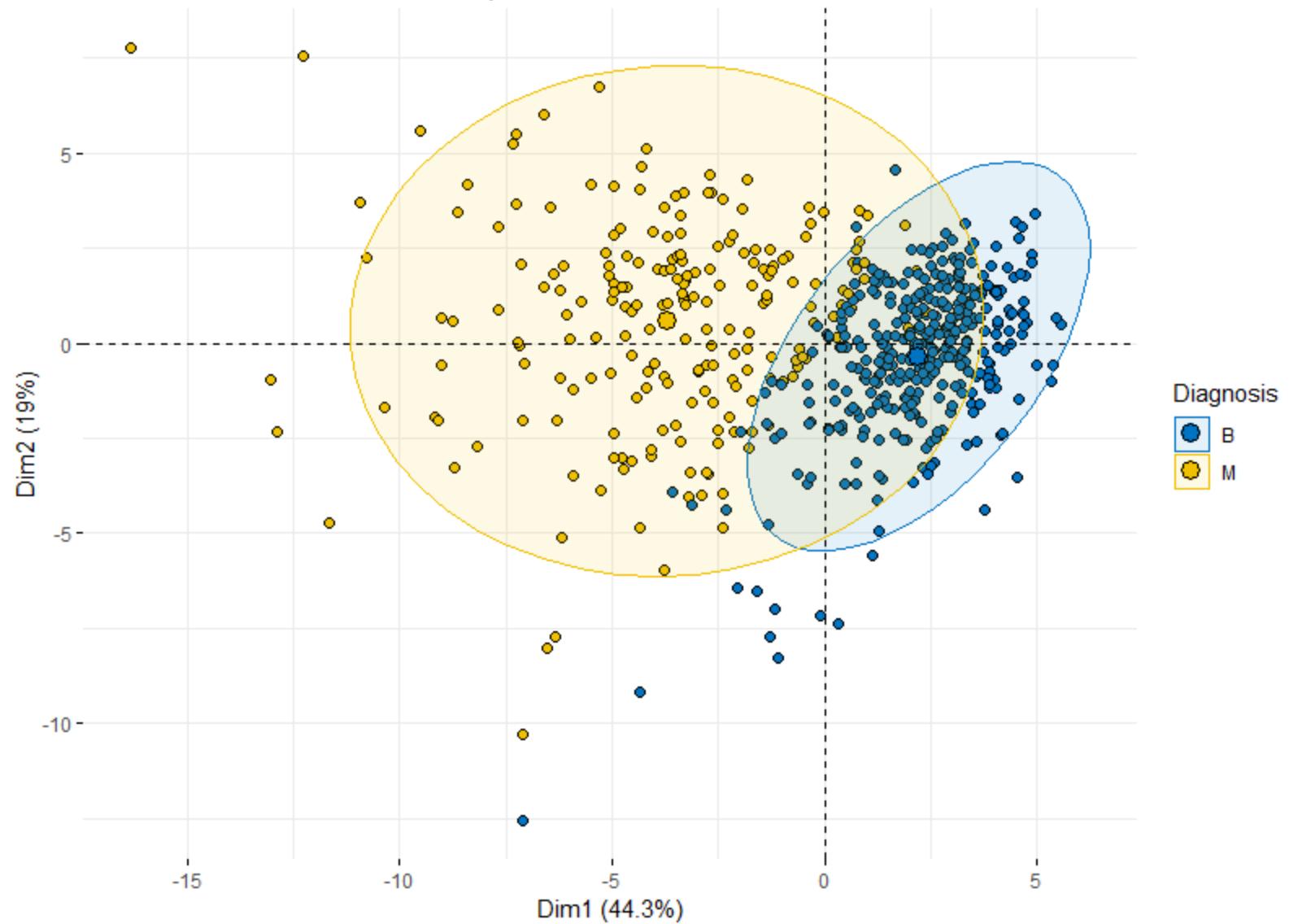


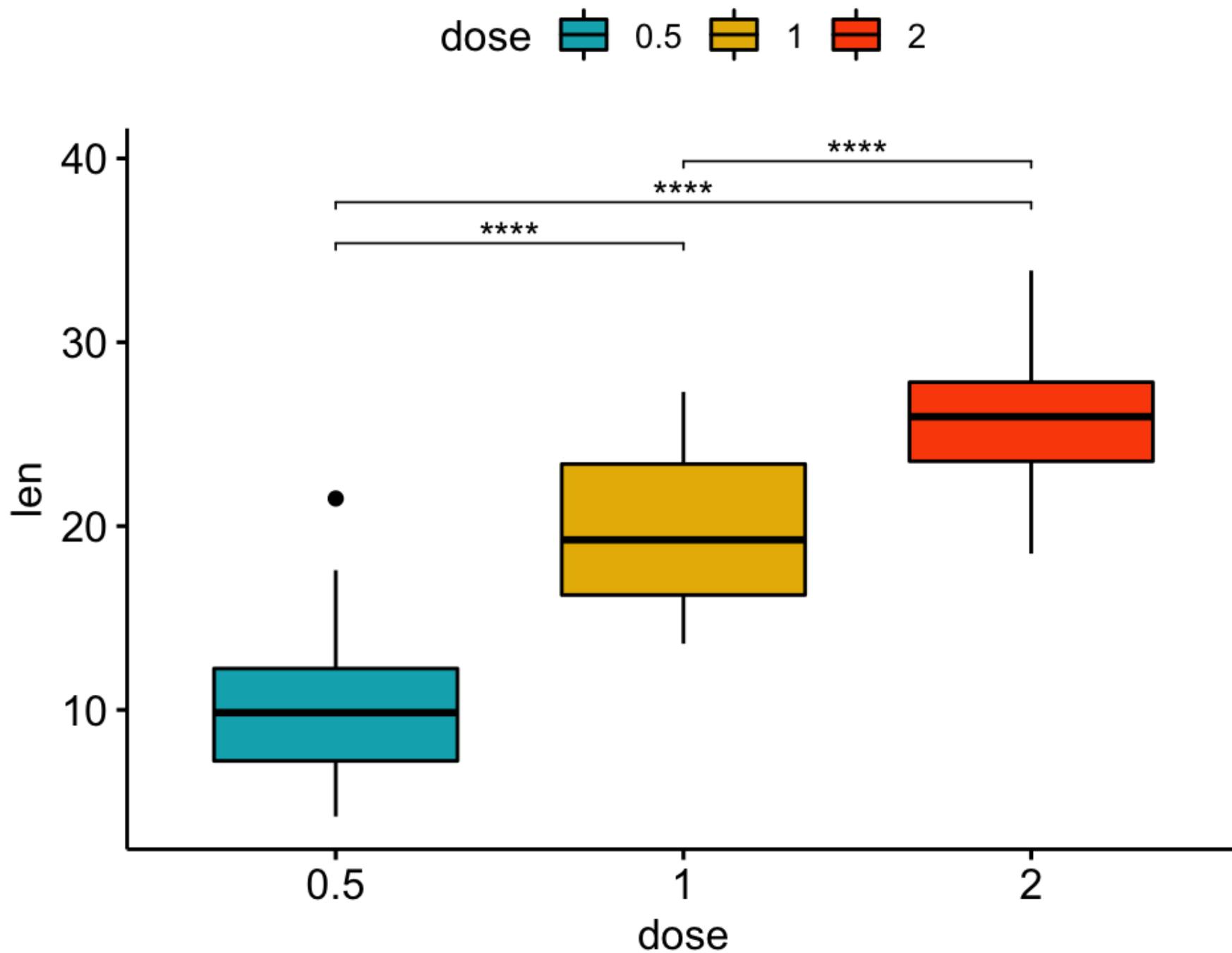
TRANSCRIPTÓMICA

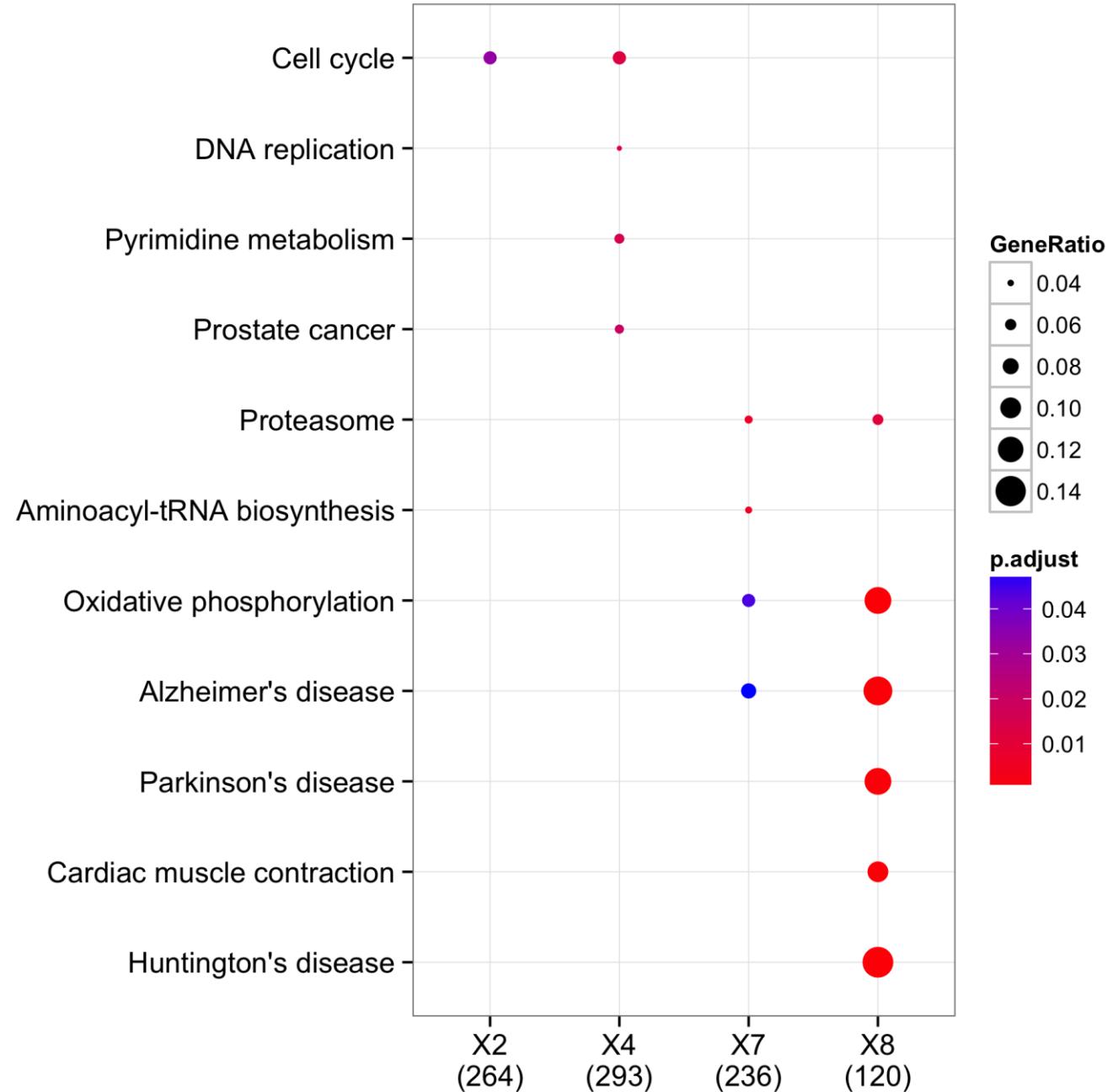


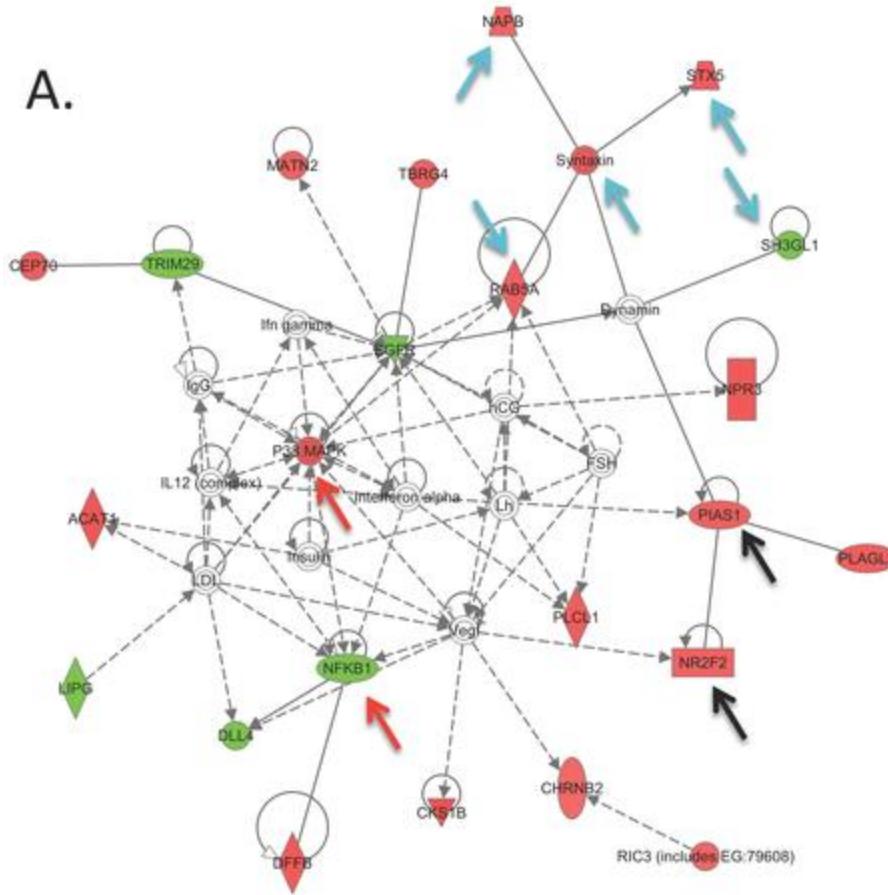
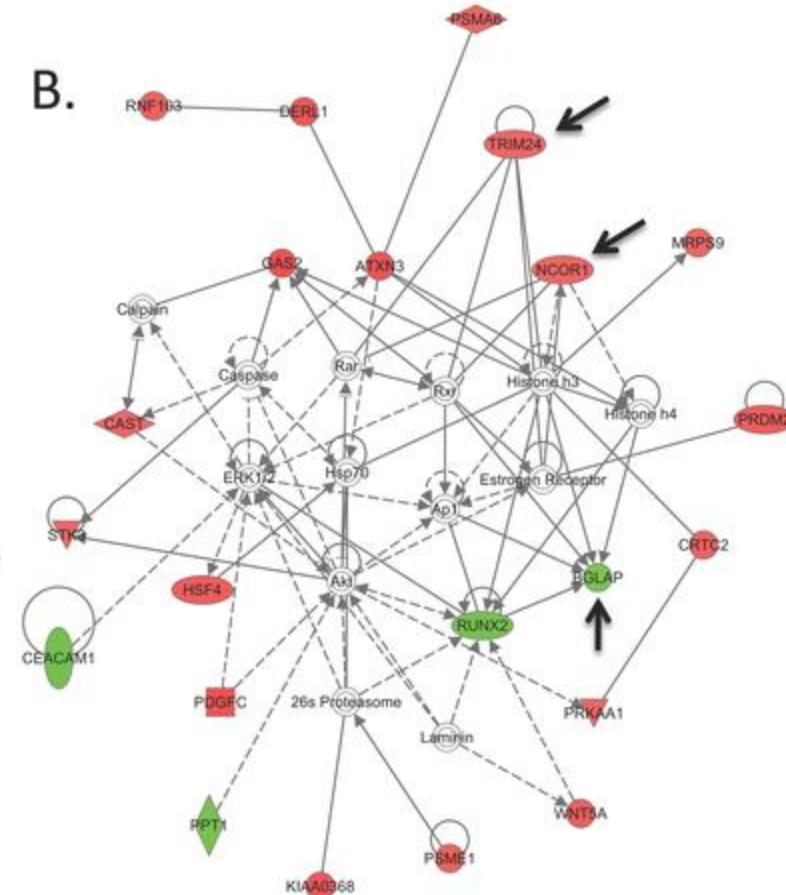
- Estudio de todo el transcriptoma
- Instrumento: Secuenciador
- Archivo de salida del secuenciador: Fastq
- ¿Para qué?
 - Ver los niveles de transcritos (codificantes y/o no codificantes)
 - Ver isoformas
 - Ver velocidad de transcripción
- ¿Y después?

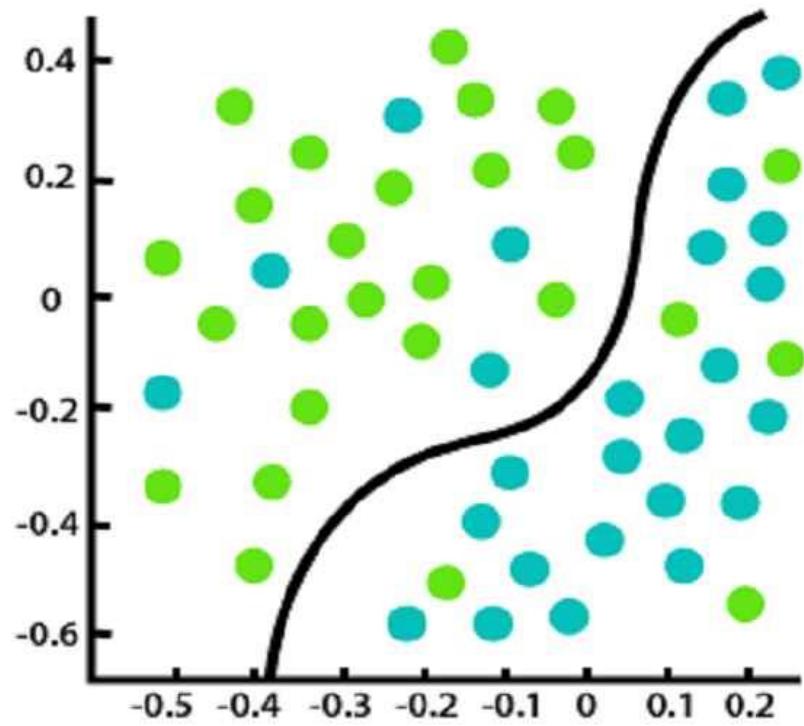
2D PCA-plot from 30 feature dataset



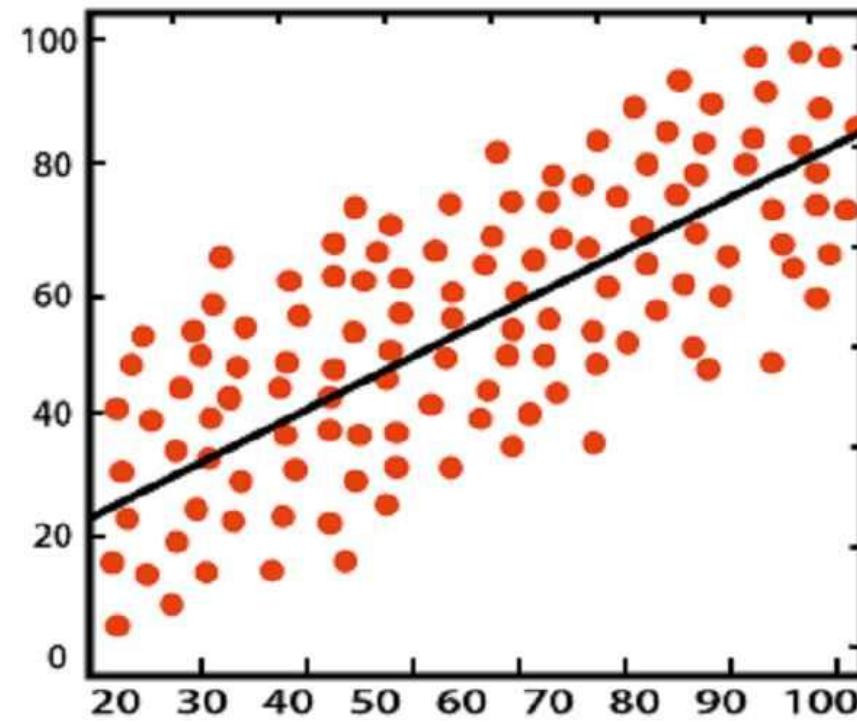




A.**B.**



Classification



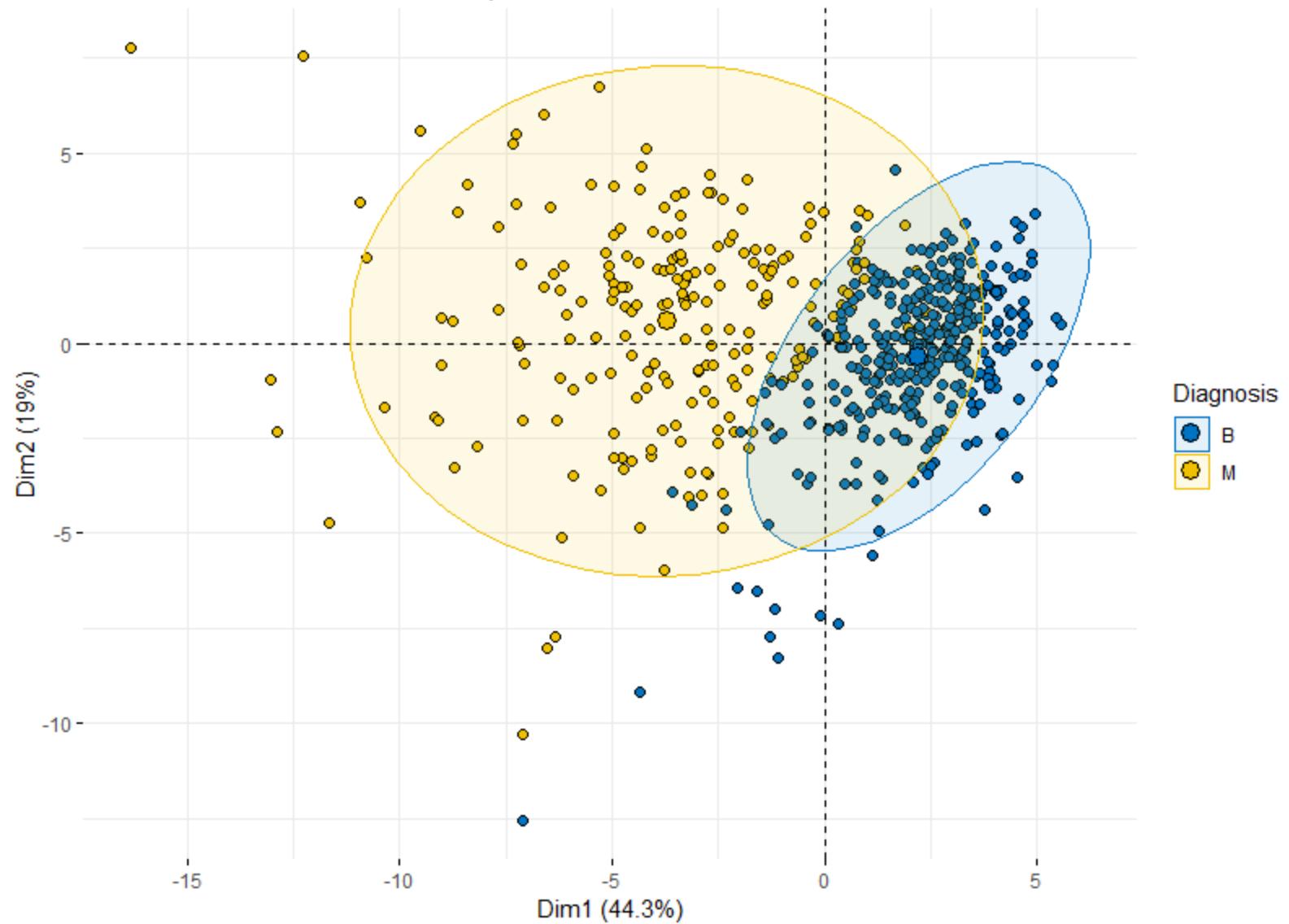
Regression

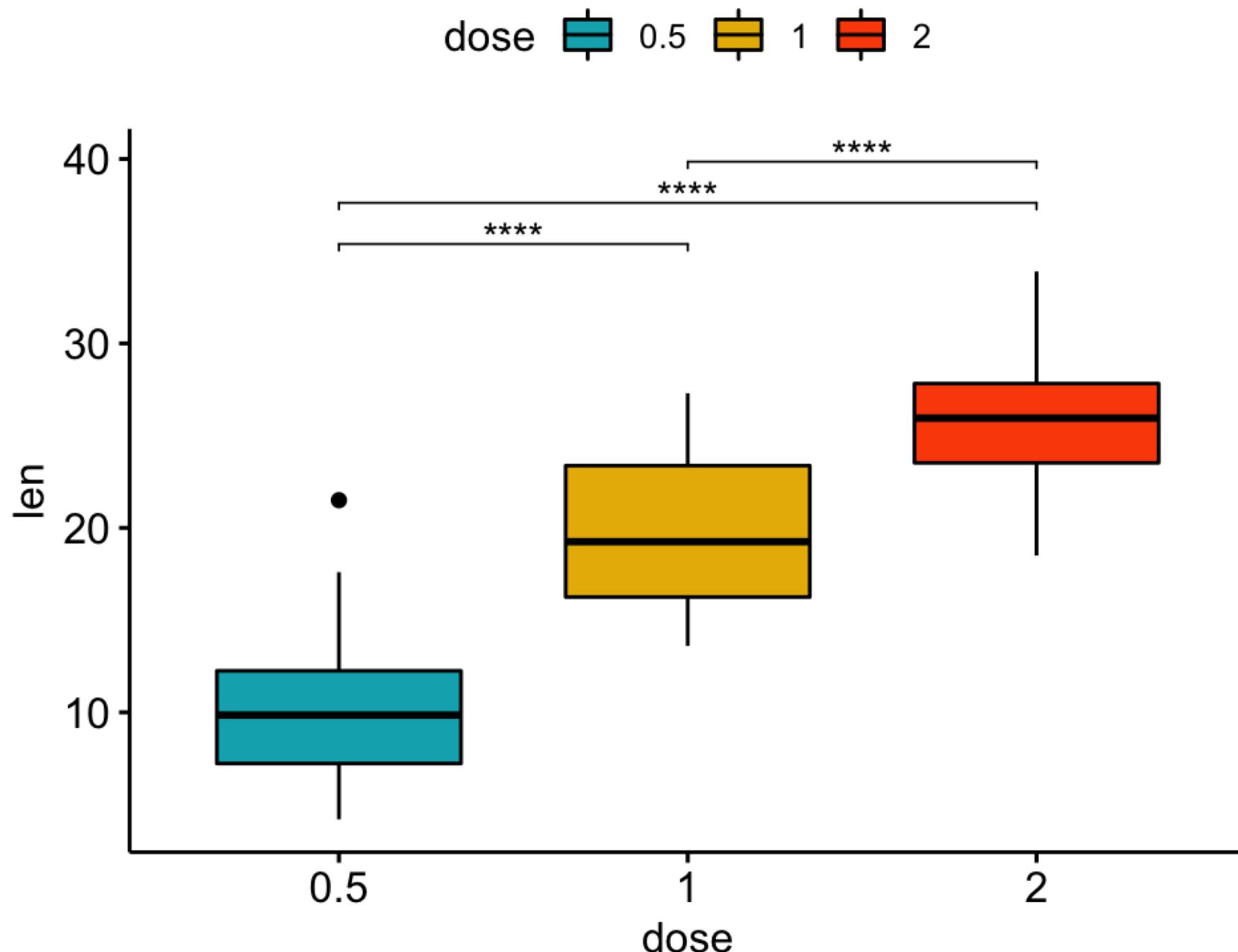
PROTEÓMICA

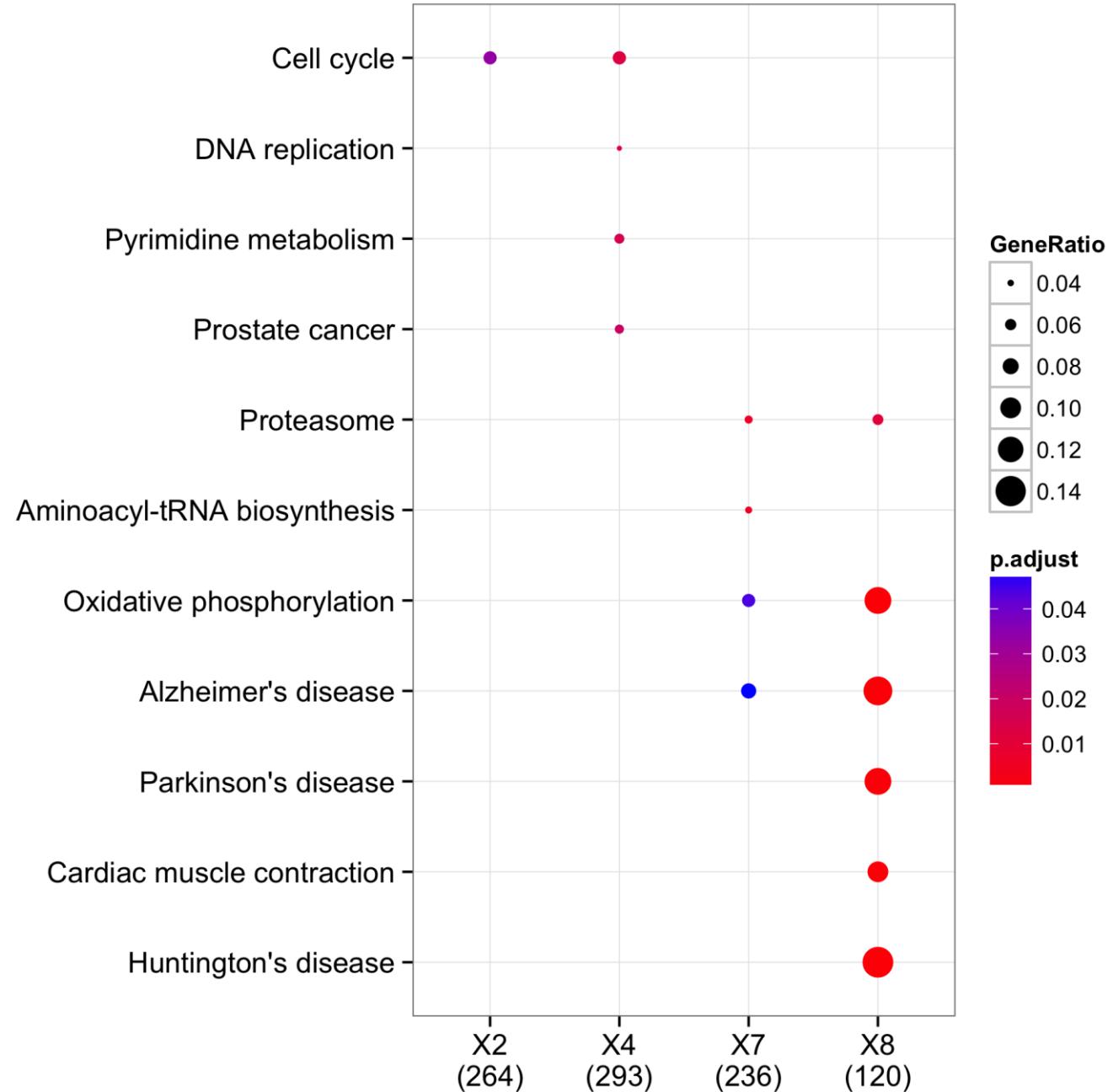


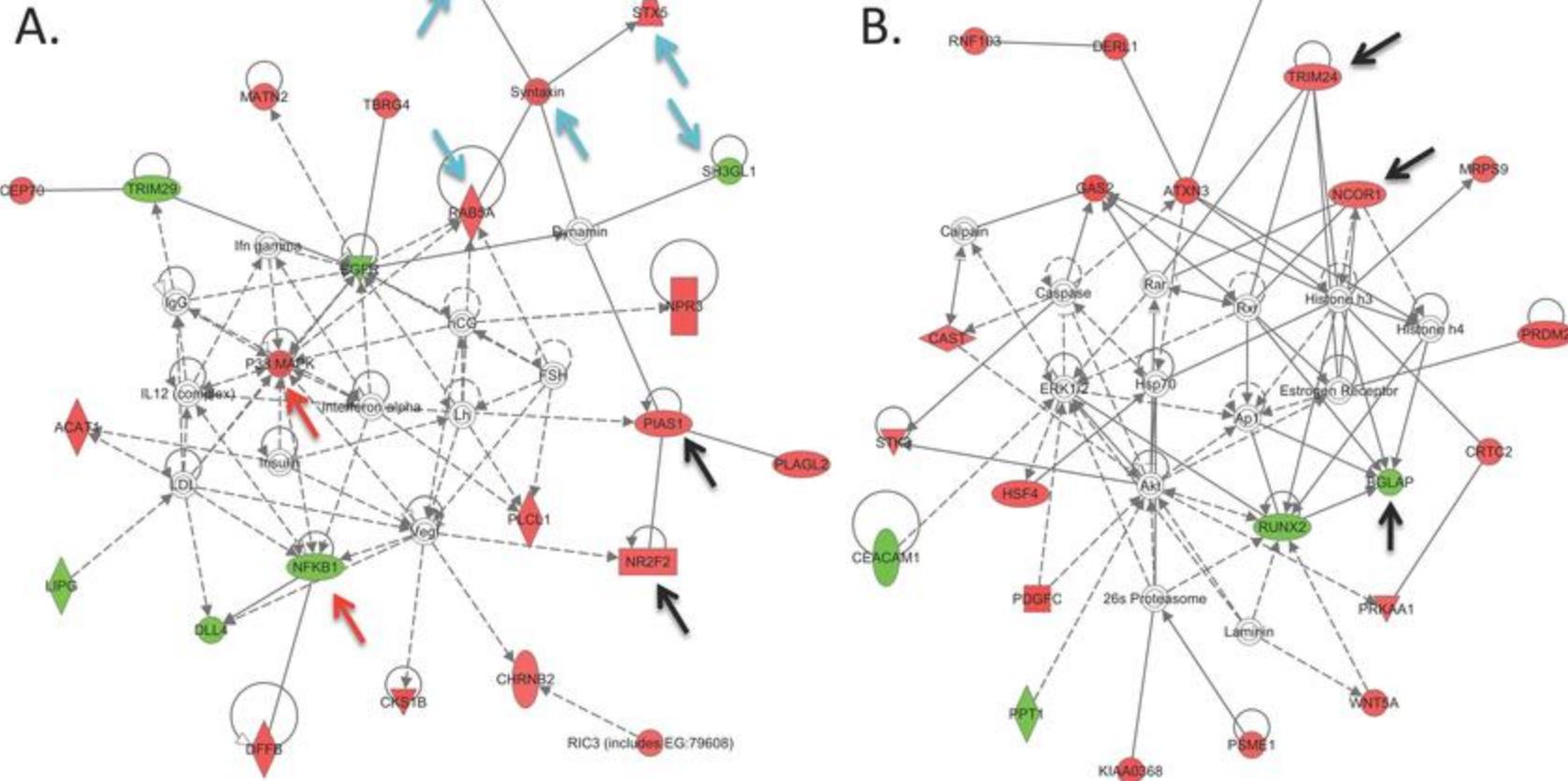
- Estudio de todo el proteoma
- Instrumento: LC-MS/MS
- Archivo de salida del secuenciador: mzXML/.raw
- ¿Para qué?
 - Para ver los niveles de proteínas en las muestras
 - Para ver los niveles de modificaciones proteicas en las muestras
 - Para ver interacciones: proteína-proteína
- ¿Y después?

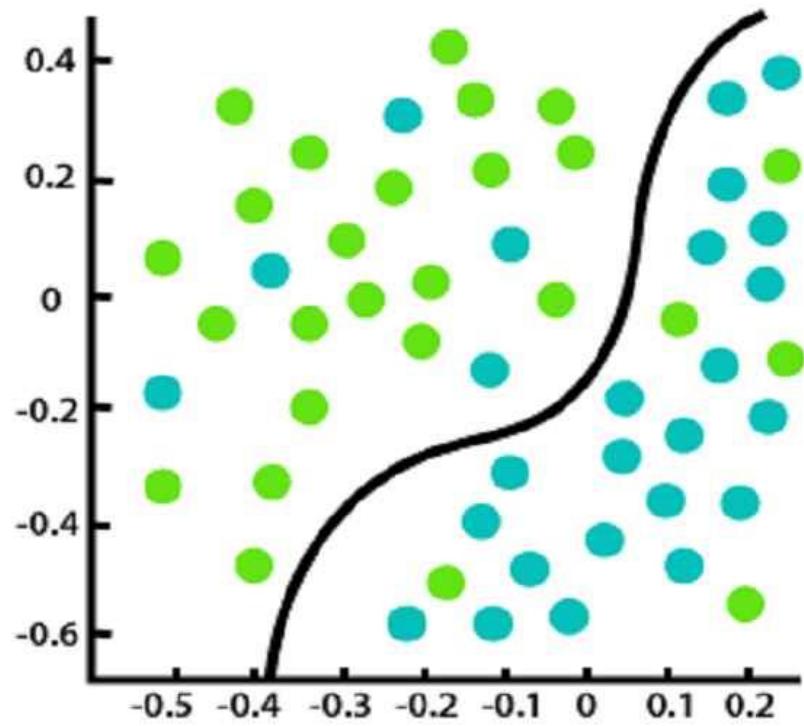
2D PCA-plot from 30 feature dataset



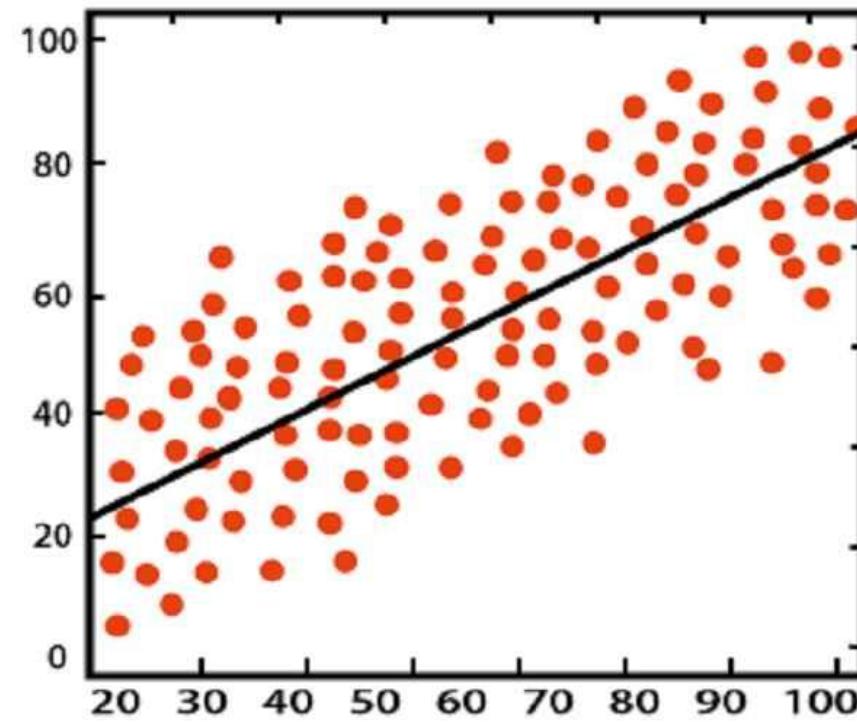








Classification



Regression

Valor positivo



Cero



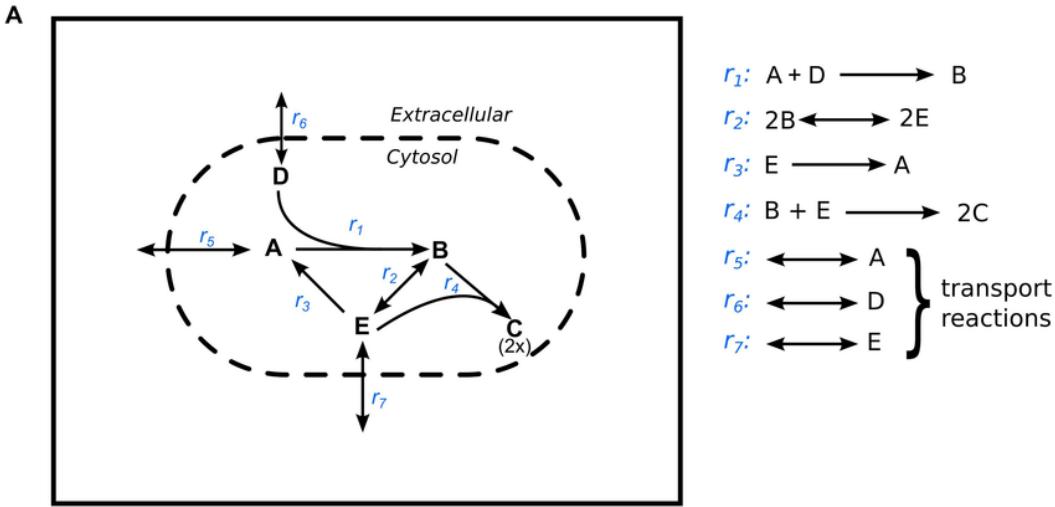
Missing value



METABOLÓMICA



- Estudio de todo el metaboloma (muy cerca del fenoma)
- Instrumento: Espectrómetro de masas (LC – MS, LC –MS/MS)
- Archivo de salida del secuenciador: mzXML / raw
- ¿Para qué?
 - Ver el producto último, los niveles de metabolitos
 - FBA
- ¿Y después?



B

$$S = \begin{matrix} & r_1 & r_2 & r_3 & r_4 & r_5 & r_6 & r_7 \\ \text{A} & -1 & 0 & 1 & 0 & 1 & 0 & 0 \\ \text{B} & 1 & -2 & 0 & -1 & 0 & 0 & 0 \\ \text{C} & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ \text{D} & -1 & 0 & 0 & 0 & 0 & 1 & 0 \\ \text{E} & 0 & 2 & -1 & -1 & 0 & 0 & 1 \end{matrix}$$

(Stoichiometric values)

$$\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \end{bmatrix}$$

(Metabolic flux values)

C

Objective function
 $\max Z = v_4$
 (Production of compound C)

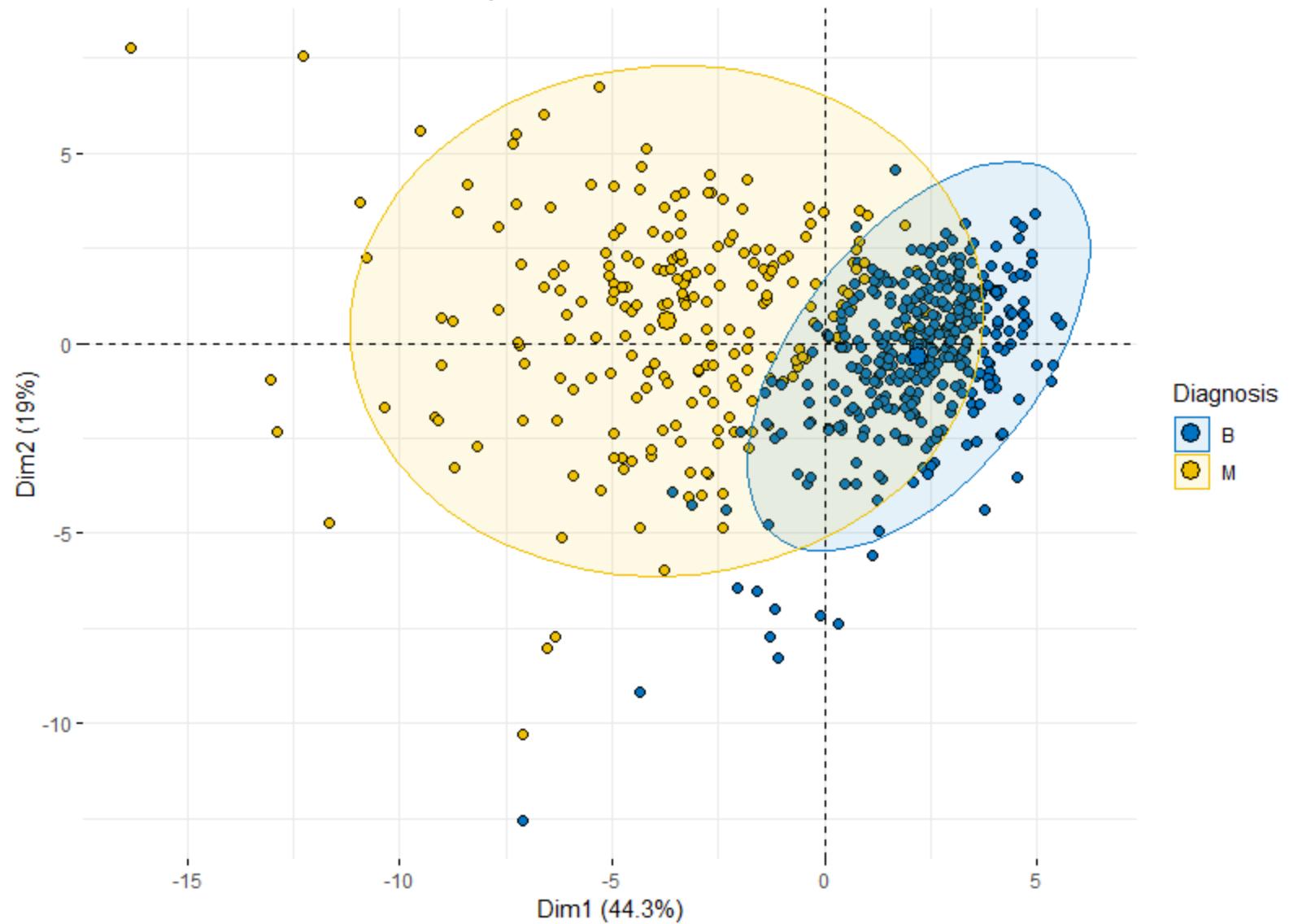
Subject to

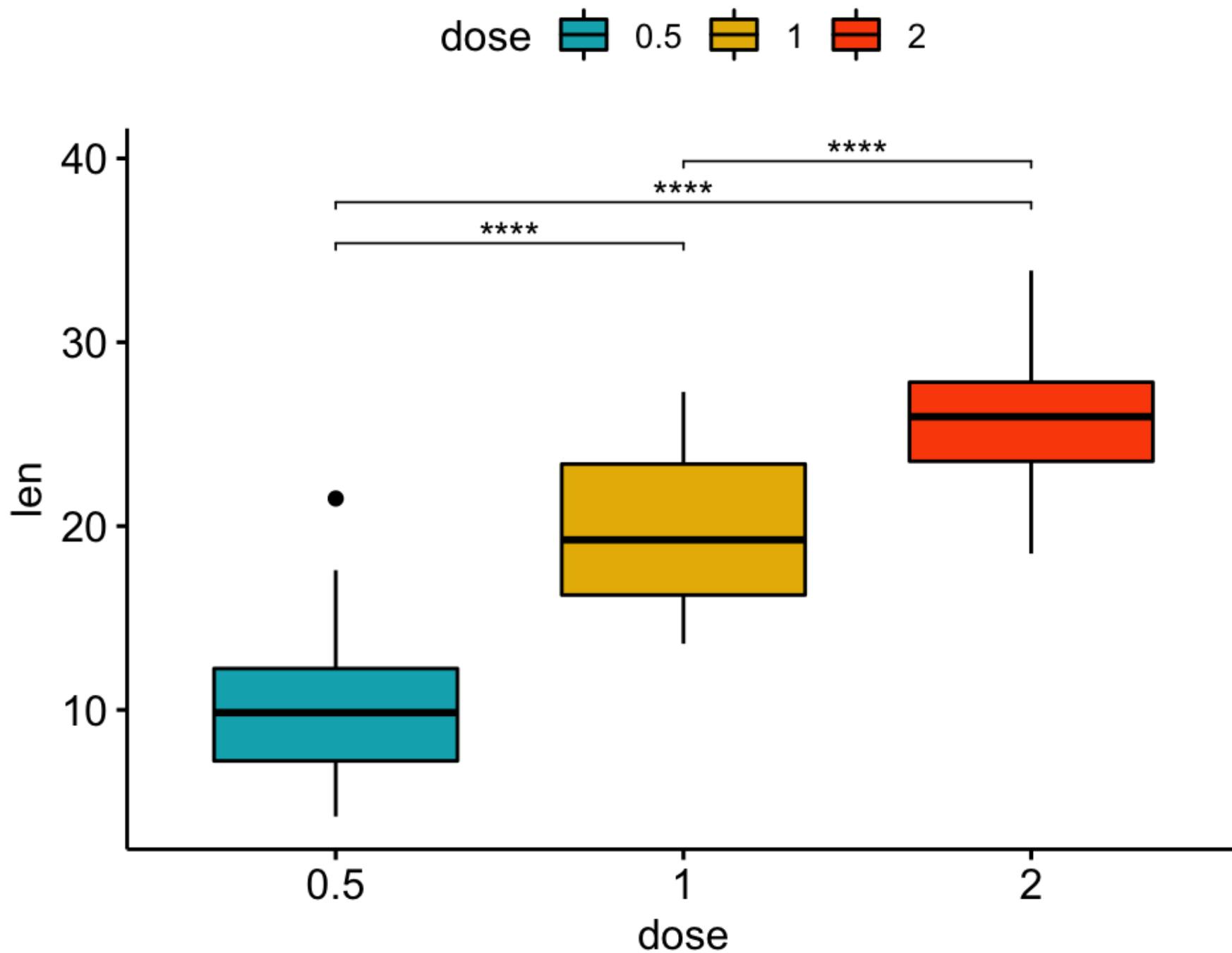
$$S\vec{v} = \vec{0} = \begin{cases} \frac{dA}{dt} = -v_1 + v_3 + v_5 & 0 \leq v_1 < \infty \\ \frac{dB}{dt} = v_1 - 2v_2 - v_4 & -\infty < v_2 < \infty \\ \frac{dC}{dt} = 2v_4 & 0 \leq v_3 < \infty \\ \frac{dD}{dt} = -v_1 + v_6 & 0 \leq v_5 \leq \infty \\ \frac{dE}{dt} = 2v_2 - v_3 - v_4 + v_7 & -\infty < v_6 < \infty \\ & 0 \leq v_7 \leq \infty \end{cases}$$

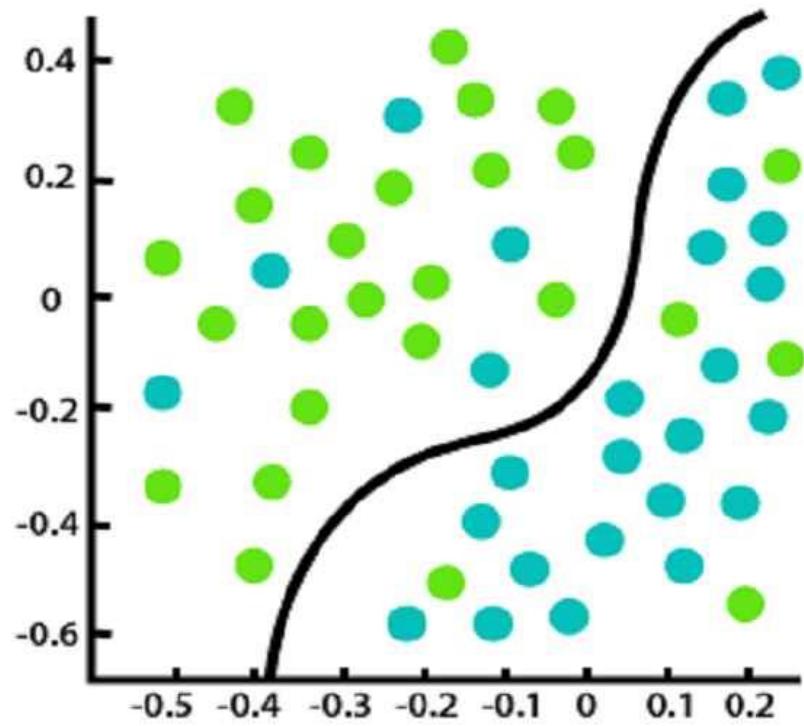
(Steady state system)

(Reaction bounds)

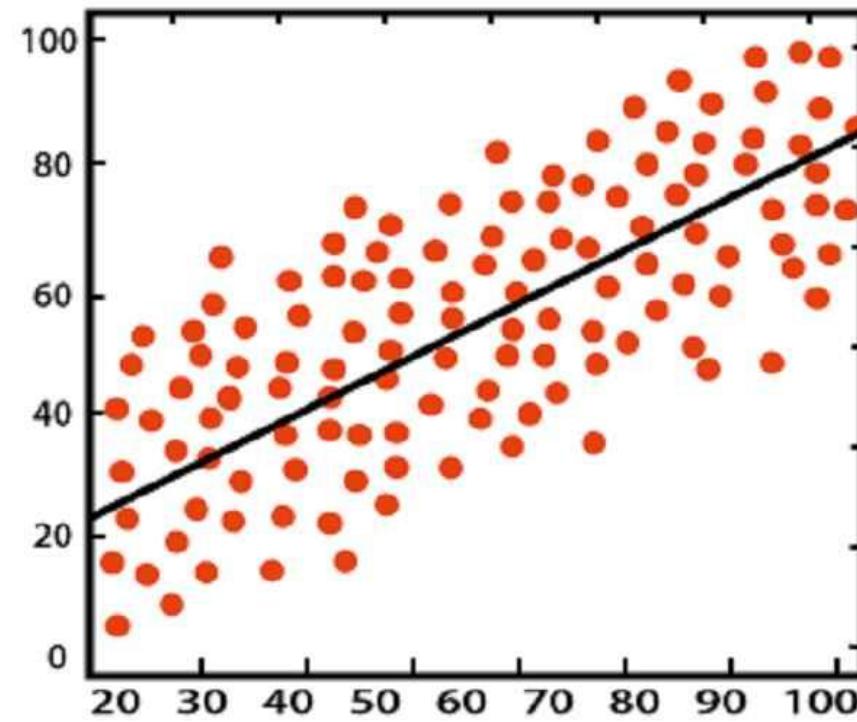
2D PCA-plot from 30 feature dataset







Classification



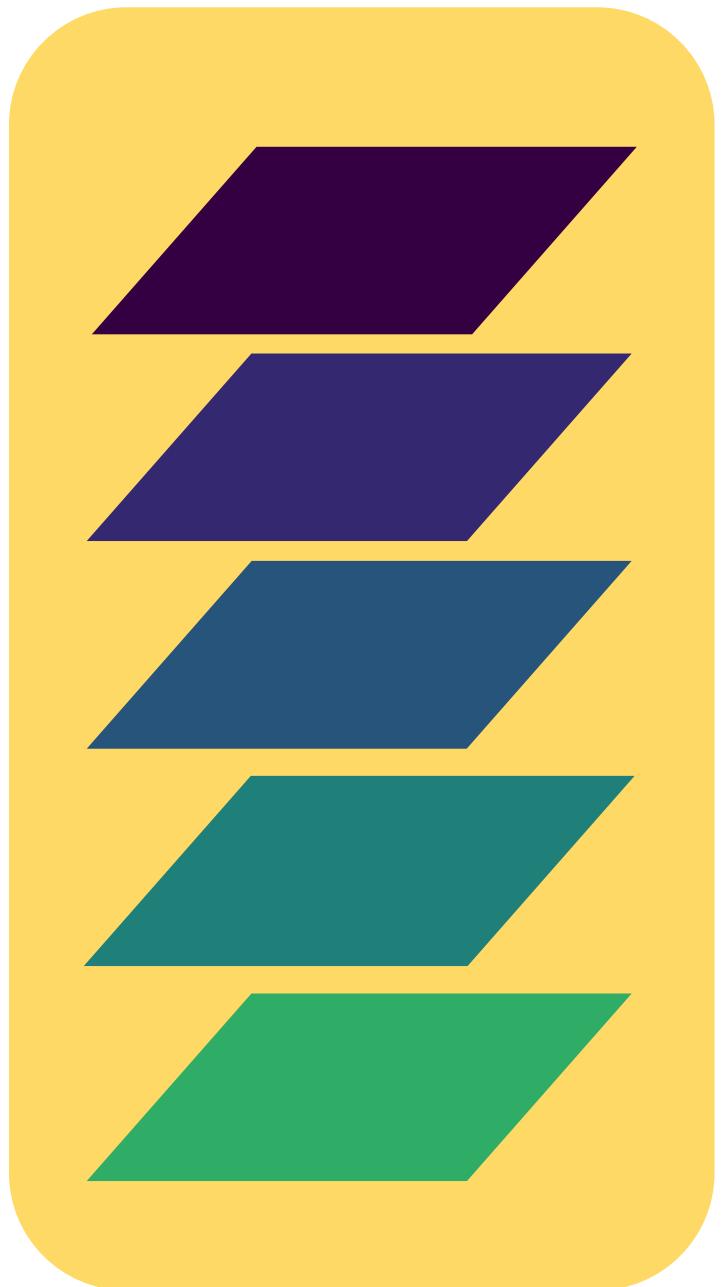
Regression

GENÓMICA *METAGENÓMICA*
EPIGENÓMICA

TRANSCRIPTÓMICA

PROTEÓMICA

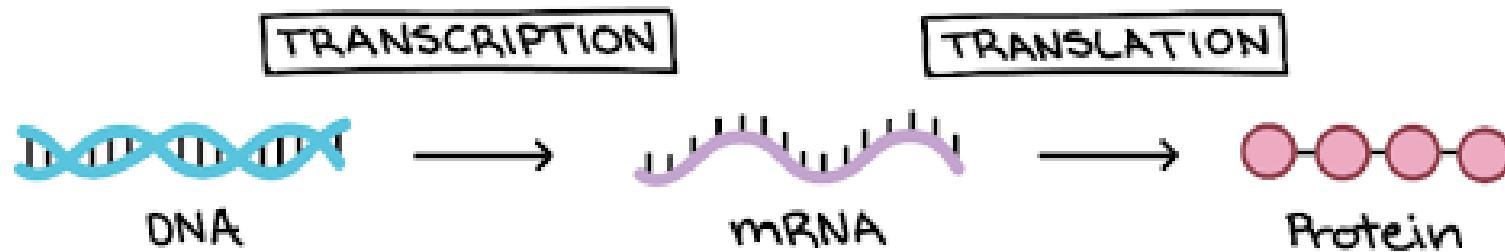
METABOLÓMICA





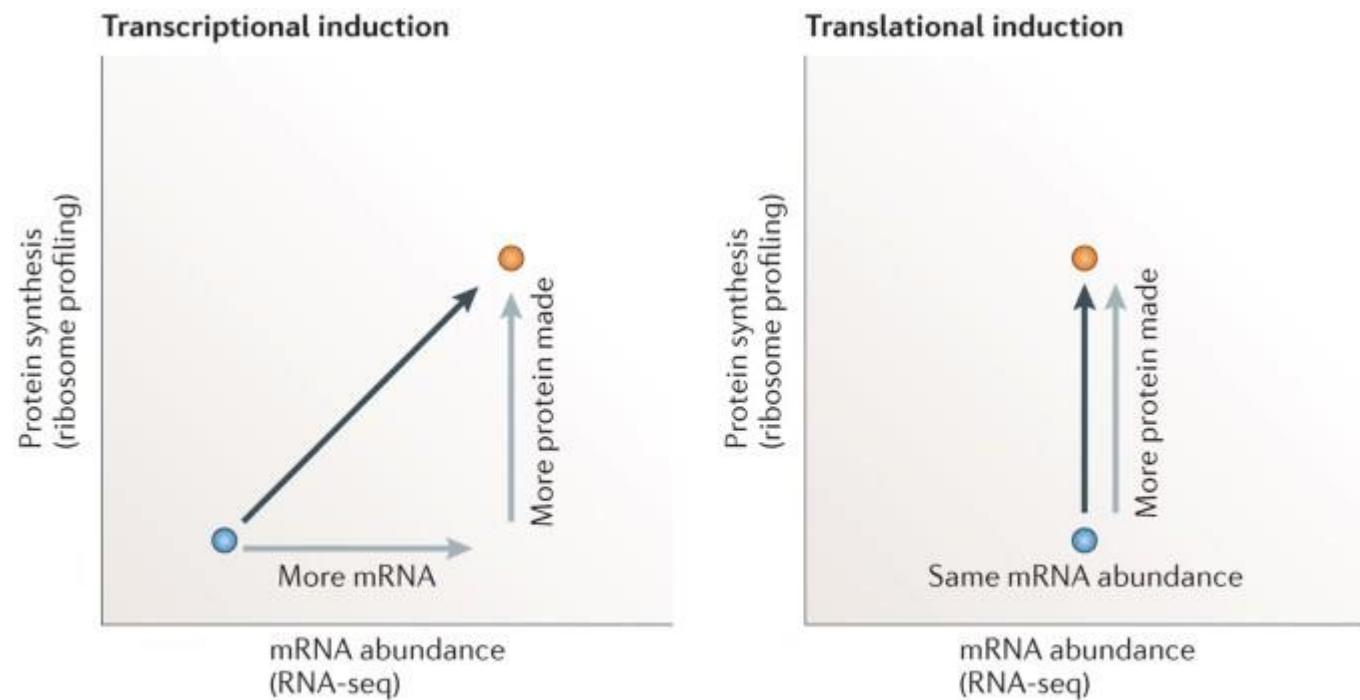
TRANSCRIPTÓMICA

TRANSCRIPTÓMICA



TRANSCRIPTÓMICA

¿Cuánto más ARNm, más proteínas?



RNA-Seq != Transcriptómica



15 minutos

Instrumentos para generar datos ómicos



GENÓMICA *METAGENÓMICA*
EPIGENÓMICA

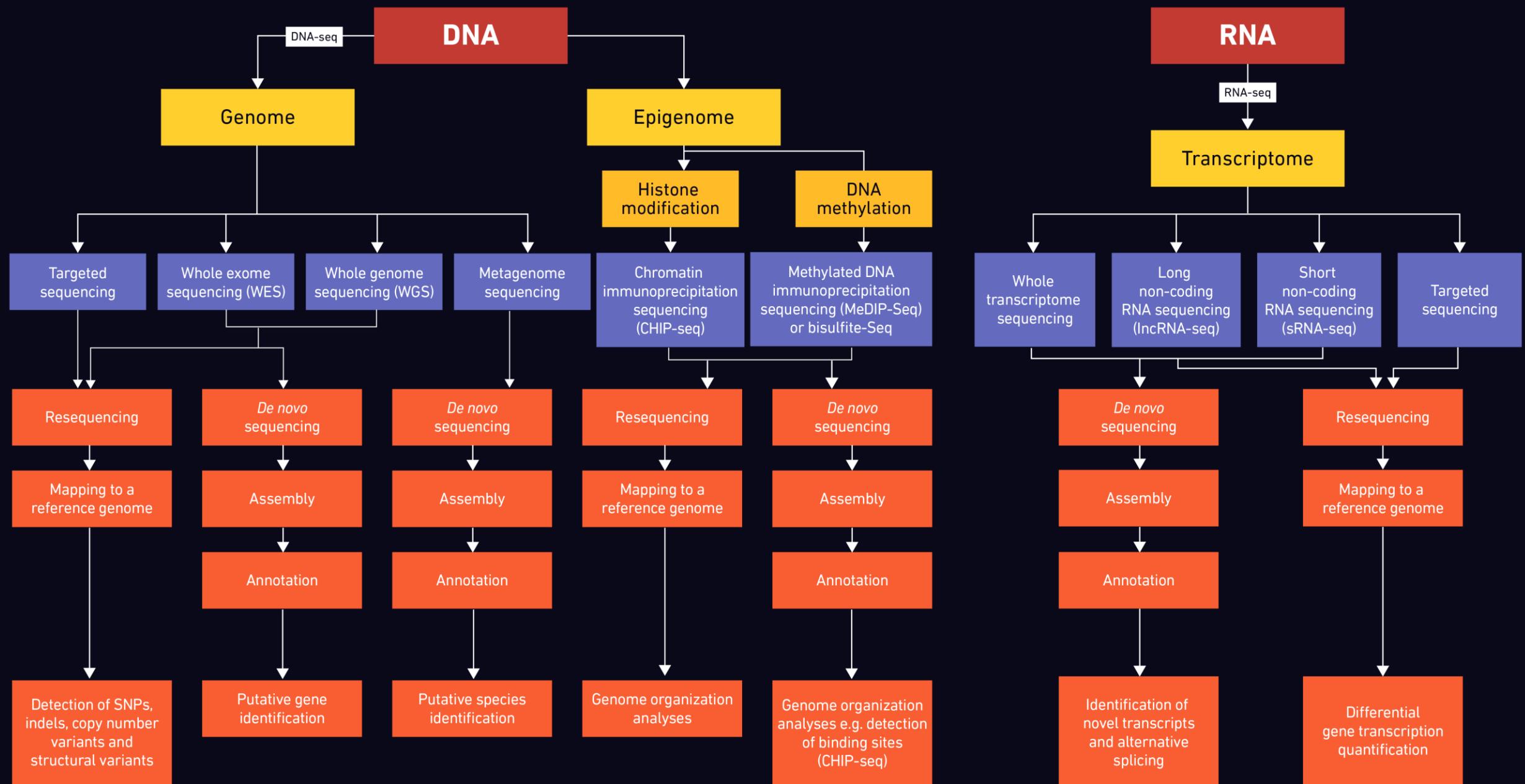
TRANSCRIPTÓMICA

PROTEÓMICA

METABOLÓMICA

GENÓMICA *METAGENÓMICA*
EPIGENÓMICA
TRANSCRIPTÓMICA

Secuenciadores



DNA Sequencing Methods

• Sequence Rearrangements

- [2b-RAD](#)
- [CPT-seq](#)
- [ddRADseq](#)
- [Digenome-seq](#)
- [EC-seq](#)
- [hyRAD](#)
- [RAD-Seq](#)
- [Rapture](#)
- [RC-Seq](#)
- [Repli-Seq](#)
- [SLAF-seq](#)
- [TC-Seq](#)
- [Tn-Seq/INSeq](#)
- [Bubble-Seq](#)
- [NSCR](#)
- [NS-Seq](#)
- [Rep-Seq/Ig-Seq/MAF](#)

• DNA Break Mapping

- [BLESS](#)
- [DSB-Seq](#)
- [GUIDE-seq](#)
- [HTGTS](#)
- [LAM-HTGTS](#)
- [Break-seq](#)
- [SSB-Seq](#)

• DNA Protein Interactions

- [DNaseI Seq or DNase-Seq](#)
- [Pu-seq](#)
- [3-C/Capture-C/Hi-C](#)
- [4C-seq](#)
- [5C](#)
- [ATAC-Seq/Fast-ATAC](#)
- [CATCH IT](#)
- [Chem-seq](#)
- [ChIA-PET](#)
- [CHIPmentation](#)
- [ChIP-Seq/HT-ChIP/ChIP-exo/Mint-ChIP](#)
- [DamID](#)
- [DNase I SIM](#)
- [FAIRE-seq/Sono-Seq](#)
- [FiT-Seq](#)
- [HiTS-FLIP](#)
- [MINCE-seq](#)
- [MNase-Seq/MAINE-Sequcledo-Sequc-seq](#)
- [MPE-seq](#)
- [NG Capture-C](#)
- [NOMe-Seq](#)
- [ORGANIC](#)
- [PAT-ChIP](#)
- [PB seq](#)
- [SELEX or SELEX-seq / HT-SELEX](#)
- [THS-seq](#)
- [UMI-4C](#)
- [X-ChIP-seq](#)

• Epigenetics

- [Aba-seq](#)
- [BisChIP-Seq/ChIP-BS-Seq/ChIP-BMS](#)
- [BSAS](#)
- [BSPP](#)
- [BS-Seq/Bisulfite-Seq/WGBS](#)
- [CAB-Seq](#)
- [EpiRADseq](#)
- [fCAB-seq](#)
- [fC-CET](#)
- [fC-Seal](#)
- [hMeDIP-seq](#)
- [JBP1-seq](#)
- [MAB-seq](#)
- [MBDCap-seq/MethylCap-Seq/MiGS](#)
- [MeDIP-Seq/DIP-seq](#)
- [MIRA](#)
- [MRE-Seq and Methyl-Seq](#)
- [oxBS-Seq](#)
- [PBAT](#)
- [redBS-Seq/caMAB-seq](#)
- [RRBS-Seq](#)
- [RRMAB-seq](#)
- [TAB-Seq](#)
- [TAmC-Seq](#)
- [T-WGBS](#)

• Low-Level DNA Detection

- [Safe-SeqS](#)
- [scAba-seq](#)
- [scATAC-Seq \(Cell index variation\)](#)
- [scATAC-Seq \(Microfluidics variation\)](#)
- [scBS-Seq](#)
- [scM&T-Seq](#)
- [scRC-Seq](#)
- [SMDB](#)

RNA Sequencing Methods

• Low-Level RNA Detection

- [CEL-Seq](#)
- [CirSeq](#)
- [CLaP](#)
- [CytoSeq](#)
- [Digital RNA Sequencing](#)
- [DP-Seq](#)
- [Drop-Seq](#)
- [Hi-SCL](#)
- [InDrop](#)
- [MARS-Seq](#)
- [Nuc-Seq](#)
- [PAIR](#)
- [Quartz-Seq](#)
- [scM&T-Seq](#)
- [SCRB-Seq](#)
- [scRNA-Seq](#)
- [scTrio-seq](#)
- [Smart-Seq](#)
- [Smart-Seq2](#)
- [snRNA-Seq](#)
- [STRT-Seq](#)
- [SUPeR-Seq](#)
- [TCR-LA-MC PCR](#)
- [TIVA](#)
- [UMI](#)
- [5C](#)
- [5C](#)
- [Div-Seq](#)
- [FRISCR](#)
- [TCR Chain Pairing](#)
- [AbPair](#)

• RNA Modifications

- [ICE](#)
- [MeRIP-Seq](#)
- [miCLIP-m6A](#)
- [Pseudo-Seq](#)
- [PSI-Seq](#)

• RNA Structure

- [CAP-seq](#)
- [Cap-Seq](#)
- [CIP-TAP](#)
- [PARS-Seq](#)
- [SPARE](#)
- [Structure-Seq/DMS-Seq](#)
- [CIRS-Seq](#)
- [icSHAPE](#)
- [SHAPE-MaP](#)
- [SHAPE-Seq](#)

• RNA Transcription

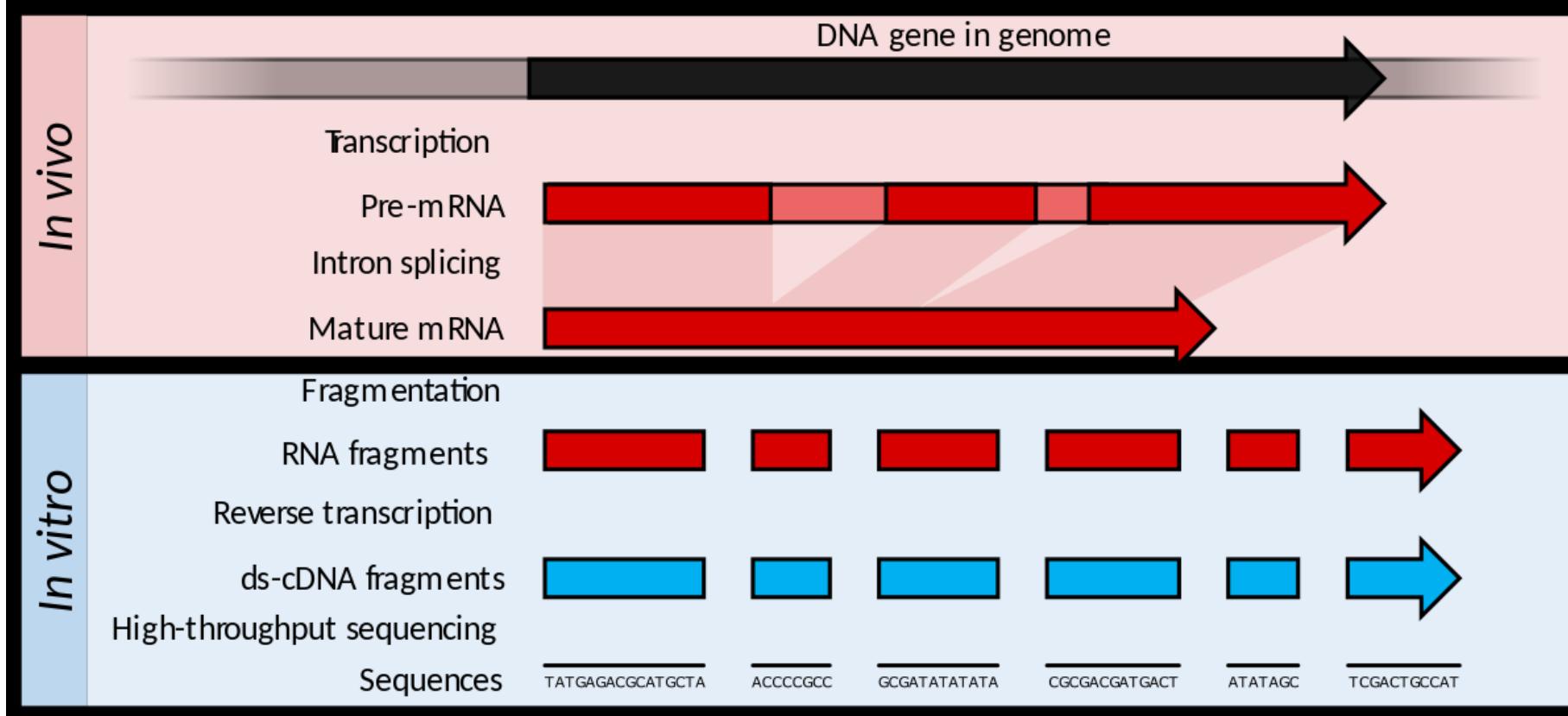
- [2P-Seq](#)
- [3'NT Method](#)
- [3P-Seq](#)
- [3Seq](#)
- [3'-Seq](#)
- [5'-GRO-Seq](#)
- [BruChase-Seq](#)
- [BruDRB-Seq](#)
- [Bru-Seq](#)
- [CAGE](#)
- [CHART](#)
- [ChIRP](#)
- [ClickSeq](#)
- [GRO-seq](#)
- [NET-Seq](#)
- [PAL-Seq](#)
- [PARE-Seq](#)
- [PEAT](#)
- [PRO-Cap](#)
- [PRO-Seq](#)
- [RAP](#)
- [RARseq](#)
- [RASL-Seq](#)

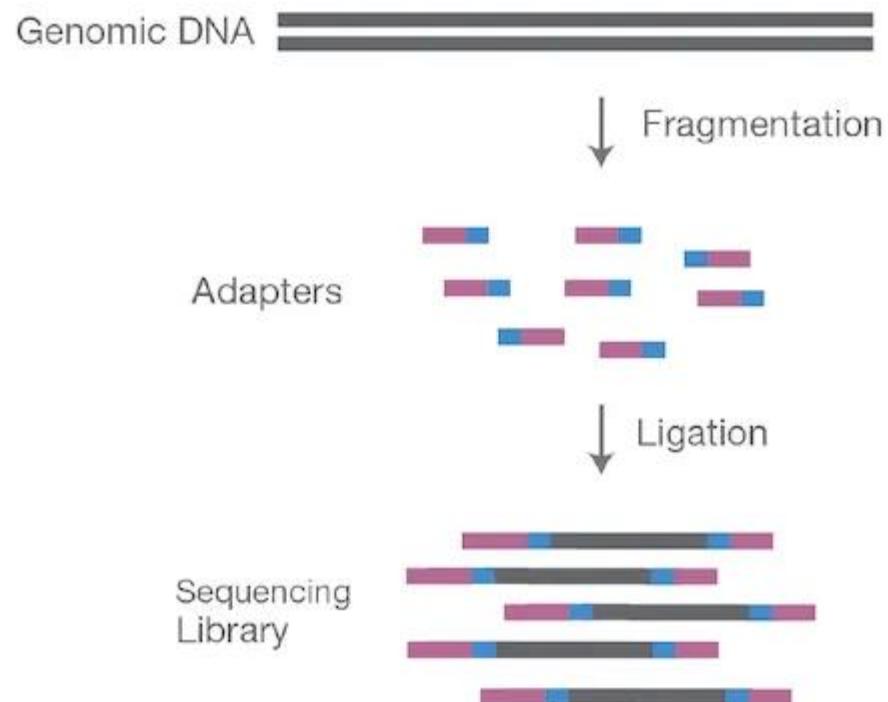
- [RNA-Seq](#)
- [SMORE-Seq](#)
- [TAIL-Seq](#)
- [TATL-Seq](#)
- [TIF-Seq](#)
- [TL-Seq](#)
- [4sUDRB-Seq](#)
- [CaptureSeq](#)
- [cP-RNA-Seq](#)
- [FRT-Seq](#)
- [GMUCT](#)
- [mNET-Seq](#)

• RNA-Protein Interactions

- [AGO-CLIP](#)
- [CLASH](#)
- [CLIP-Seq or HITS-CLIP](#)
- [DLAF](#)
- [eCLIP](#)
- [hiCLIP](#)
- [iCLIP](#)
- [miR-CLIP](#)
- [miTRAP](#)
- [PAR-CLIP](#)
- [PIP-Seq](#)
- [Pol II CLIP](#)
- [RBNS](#)
- [Ribo-Seq or ARTSeq](#)
- [RIP-Seq](#)
- [TRAP-Seq](#)
- [TRIBE](#)
- [BrdU-CLIP](#)
- [HiTS-RAP](#)
- [irCLIP](#)

**A C
G T**







First Generation



Sanger Sequencing
Maxam and Gilbert
Sanger Chain-termination

- Infer nucleotide identity using dNTPs then visualize with electrophoresis
- 500-1000 bp fragments

Second Generation (Next Generation Sequencing)



454, Solexa,
Ion Torrent
Illumina

- High throughput from the parallelization of sequencing reactions
- ~50-500 bp fragments

Third Generation



PacBio
Oxford Nanopore

- Sequence native DNA in real time with single-molecule resolution
- Tens of kb fragments, on average

Short-read sequencing

Long-read sequencing

Next Generation Sequencing Technology based on three things

(A) Chemistry Involved

Sequencing by Ligation

- >ABI SOLiD: Base encoded Probes
- >Complete Genomics: Combinatorial Probe Anchor Ligation

Sequencing by Synthesis

- >Illumina Platforms: Cyclic Reversible Terminator technology,
- >454 (Pyrosequencing), Ion Torrent: Single Nucleotide Amplification

Nanopore

- >MinIONs: Oxford Nanopore, Genia Nanopore

(B) Detection

Optical (CCD, ZMW)

- >Roche 454, ABI SOLiD, Illumina and Pacific BioSciences

Solid State Detection

- >Ion Torrent

Electrical Detection

- >Oxford Nanopore, Genia Nanopore (Roche)

(C) Amplification

Required for all second generation sequencers

By Emulsion PCR

- >454 Roche (Pyrosequencing)
- >Ion Torrent

By Bridge Amplification

- >Solexa Genome Analyzer
- >Illumina Platforms

Not required for third and fourth generation sequencers

SMRT

- >Pacific Biosciences

Nanopore

- >Oxford Nanopore



iSeq 100

MiniSeq

MiSeq Series+

NextSeq 550 Series+

NextSeq 1000 & 2000

Popular Applications & Methods	Key Application	Key Application	Key Application	Key Application	Key Application
Large Whole-Genome Sequencing (human, plant, animal)					
Small Whole-Genome Sequencing (microbe, virus)	●	●	●	●	●
Exome & Large Panel Sequencing (enrichment-based)				●	●
Targeted Gene Sequencing (amplicon-based, gene panel)	●	●	●	●	●
Single-Cell Profiling (scRNA-Seq, scDNA-Seq, oligo tagging assays)				●	●
Transcriptome Sequencing (total RNA-Seq, mRNA-Seq, gene expression profiling)				●	●
Targeted Gene Expression Profiling	●	●	●	●	●
miRNA & Small RNA Analysis	●	●	●	●	●
DNA-Protein Interaction Analysis (ChIP-Seq)			●	●	●
Methylation Sequencing				●	●
16S Metagenomic Sequencing		●	●	●	●
Metagenomic Profiling (shotgun metagenomics, metatranscriptomics)				●	●
Cell-Free Sequencing & Liquid Biopsy Analysis				●	●
Run Time	9.5–19 hrs	4–24 hours	4–55 hours	12–30 hours	11–48 hours
Maximum Output	1.2 Gb	7.5 Gb	15 Gb	120 Gb	330 Gb*
Maximum Reads Per Run	4 million	25 million	25 million [†]	400 million	1.1 billion [*]
Maximum Read Length	2 × 150 bp	2 × 150 bp	2 × 300 bp	2 × 150 bp	2 × 150 bp

Benchtop Sequencers

Production-Scale Sequencers



NextSeq 550 Series +



NextSeq 1000 & 2000



NovaSeq 6000

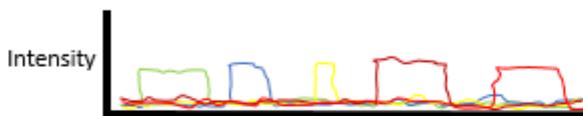
Popular Applications & Methods	Key Application	Key Application	Key Application
Large Whole-Genome Sequencing (human, plant, animal)			●
Small Whole-Genome Sequencing (microbe, virus)	●	●	●
Exome & Large Panel Sequencing (enrichment-based)	●	●	●
Targeted Gene Sequencing (amplicon-based, gene panel)	●	●	●
Single-Cell Profiling (scRNA-Seq, scDNA-Seq, oligo tagging assays)	●	●	●
Transcriptome Sequencing (total RNA-Seq, mRNA-Seq, gene expression profiling)	●	●	●
Chromatin Analysis (ATAC-Seq, ChIP-Seq)	●	●	●
Methylation Sequencing	●	●	●
Metagenomic Profiling (shotgun metagenomics, metatranscriptomics)	●	●	●
Cell-Free Sequencing & Liquid Biopsy Analysis	●	●	●
Run Time	12–30 hours	11–48 hours	~13 - 38 hours (dual SP flow cells) ~13–25 hours (dual S1 flow cells) ~16–36 hours (dual S2 flow cells) ~44 hours (dual S4 flow cells)
Maximum Output	120 Gb	330 Gb*	6000 Gb
Maximum Reads Per Run	400 million	1.1 billion*	20 billion
Maximum Read Length	2 × 150 bp	2 × 150 bp	2 × 250**

PacBio SMRT seq

DNA passes thru polymerase in an illuminated volume



Raw output is fluorescent signal of the nucleotide incorporation, specific to each nucleotide

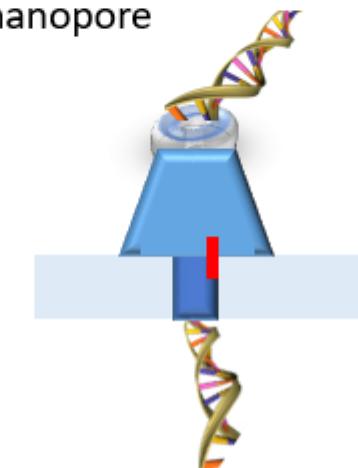


A,C,T,G have known pulse durations, which are used to infer methylated nucleotides

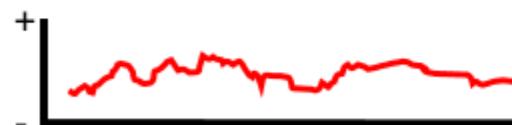


Oxford Nanopore

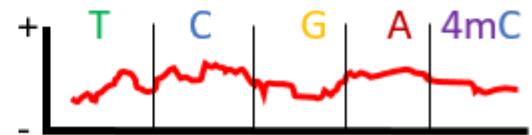
DNA passes thru nanopore



Raw output is electrical signal caused by nucleotide blocking ion flow in nanopore



Each nucleotide has a specific electric "signature"





- Coverage
- Calidad de las lecturas
- Longitud de la lectura (short vs long)
- Single-end, paired-end o mate-pair

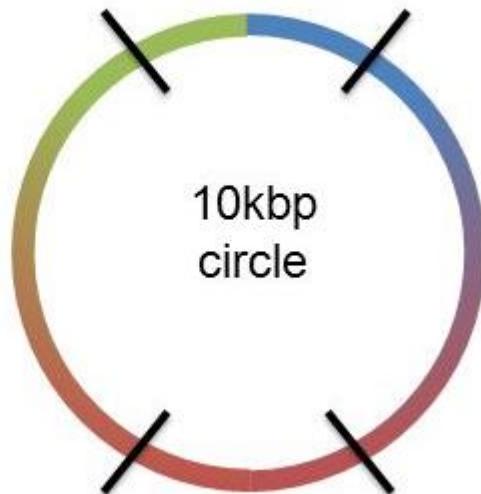
Paired-end sequencing

- Read one end of the molecule, flip, and read the other end
- Generate pair of reads separated by up to 500bp with inward orientation



Mate-pair sequencing

- Circularize long molecules (1-10kbp), shear into fragments, & sequence
- Mate failures create short paired-end reads



2x100 @ ~10kbp (outies)

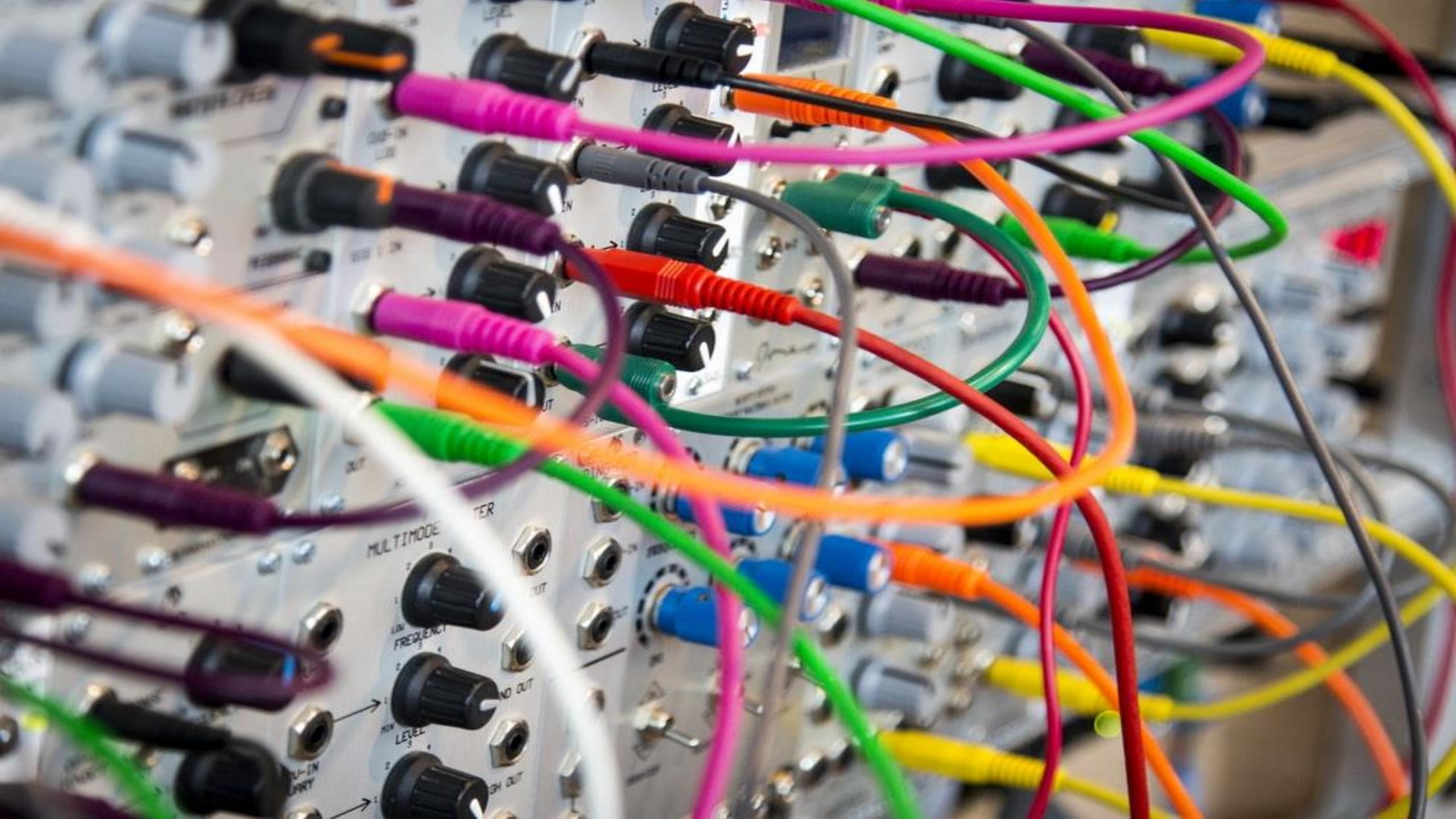


2x100 @ 300bp (innies)



Espectrometría de masas

PROTEÓMICA METABOLÓMICA





PROT

METAB

SEPARAR

METAB

ROMPER

ANALIZAR

(abundancia, tiempo
de retención y m/z)

LC - MS

SEPARAR

ROMPER

PROT

ROMPER

ANALIZAR

(abundancia, tiempo
de retención y m/z)

SEPARAR

ROMPER

RETENER

ROMPER

PROT

ANALIZAR

SEPARAR

ROMPER

RETENER

ROMPER

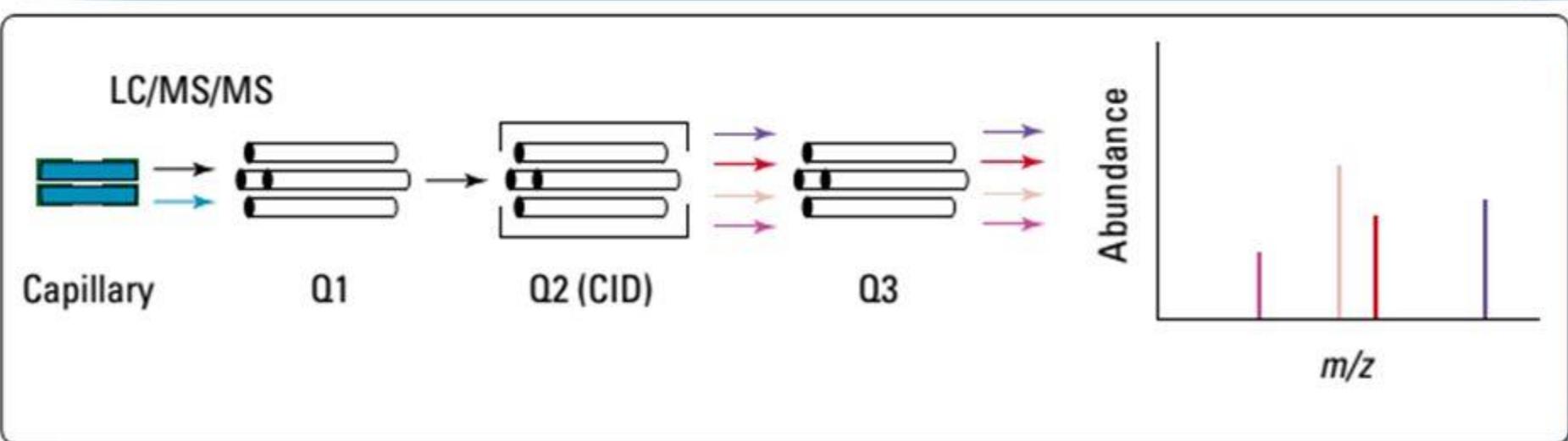
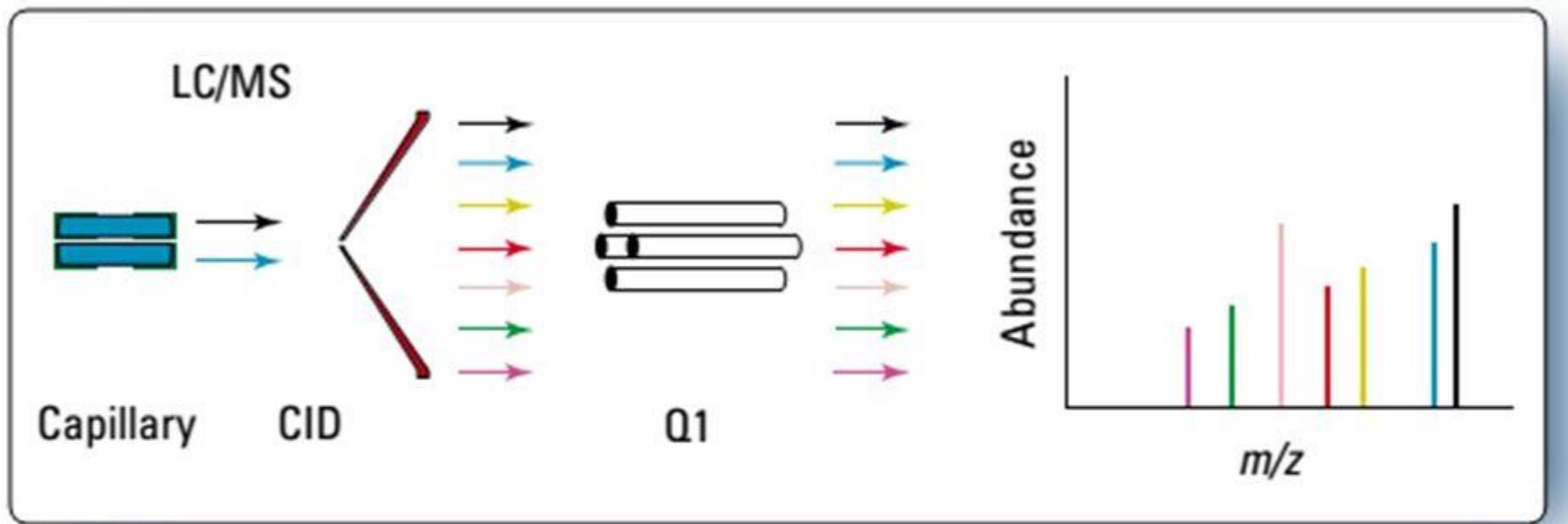
METAB

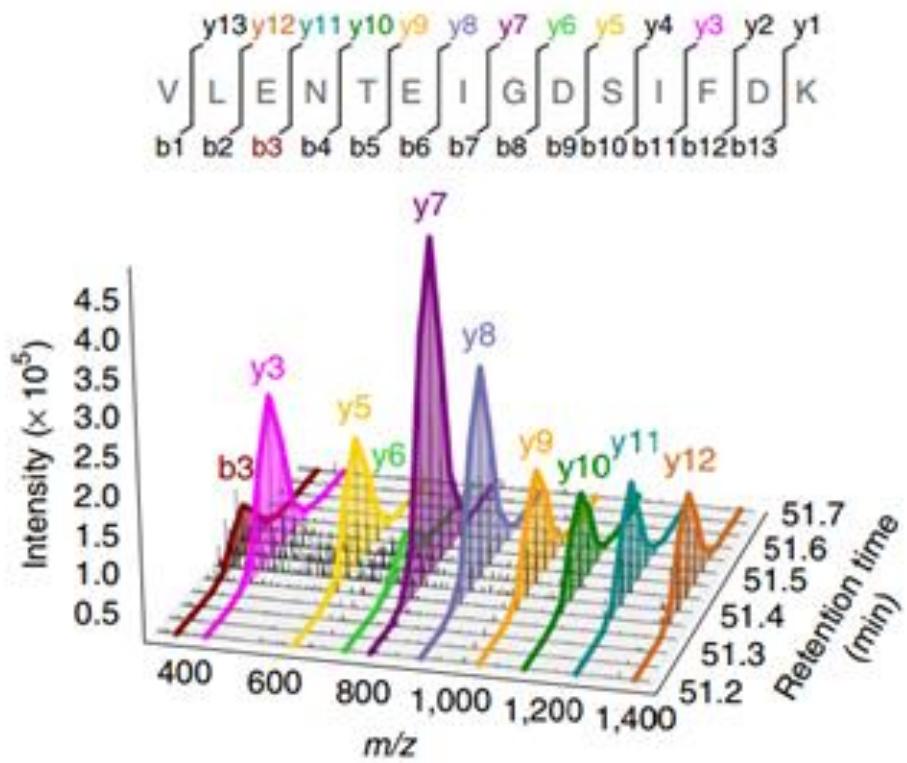
ANALIZAR

(abundancia, tiempo
de retención y m/z)

LC – MS/MS

LC-MS vs LC-MS/MS





Valor positivo

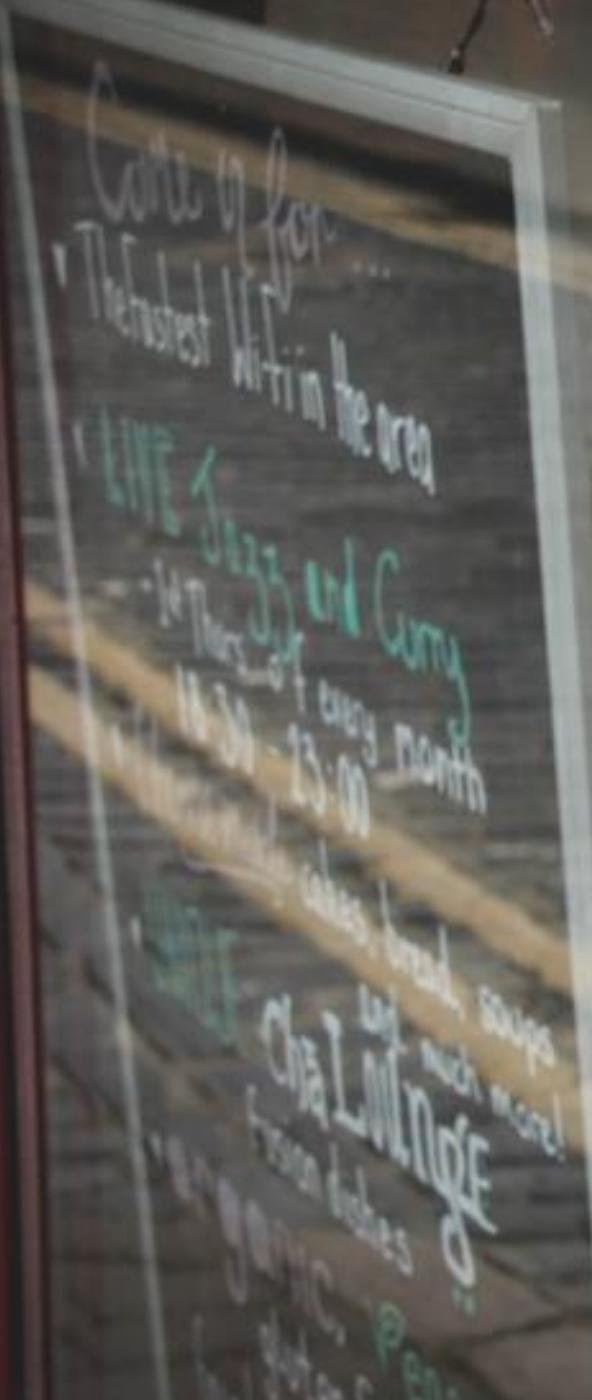


Cero



Missing value





#becurious

WED - Saturday 11am - 10pm (Friday & Sat until 11pm)
A typical day sees us re-stocking shelves, preparing food through pre-cooking
We believe in supporting local suppliers, please ask for details.
We believe in reducing food waste by using up-to-date cooking.

Item	Price
Chilli Con Carne	£7.50
Spicy Fish	£7.50
Spicy Tuna	£7.50
Spicy Chicken	£7.50
Spicy Beef	£7.50
Spicy Lamb	£7.50
Spicy Pork	£7.50
Spicy Duck	£7.50
Spicy Beef	£7.50
Spicy Lamb	£7.50
Spicy Pork	£7.50
Spicy Duck	£7.50
Spicy Beef	£7.50
Spicy Lamb	£7.50
Spicy Pork	£7.50
Spicy Duck	£7.50

We source high quality, local produce whenever possible.
We have an extensive list of produce which changes daily. Please ask us for whatever you
need and we will do our best to get a price prior to collection. Please also ask us about
our Pig Roast.

TO ORDER: PLEASE EMAIL
WE WILL CONTACT
YOU WHEN YOUR ORDER ARRIVES AND YOU CAN COME AND COLLECT FROM
CHI LOUNGE. YOU CAN PAY VIA BANKS.
DELIVERY DAYS ARE FRIDAY AND TUESDAY.
THANK YOU, THE CHI LOUNGE TEAM.

En resumen



En resumen



GENÓMICA *METAGENÓMICA*
EPIGENÓMICA

TRANSCRIPTÓMICA

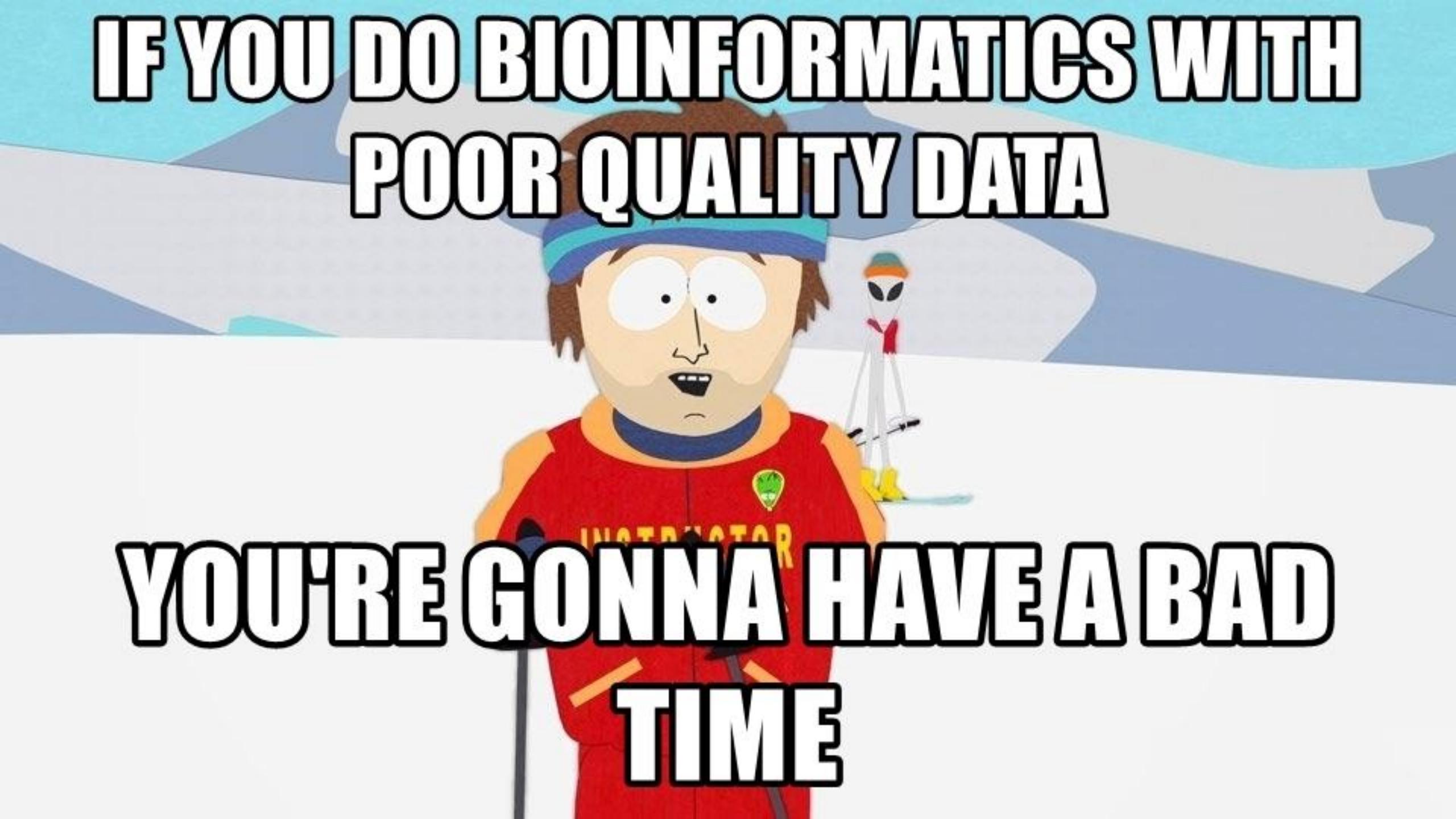
PROTEÓMICA

METABOLÓMICA



1. Diseño experimental
2. Experimento
3. Datos crudos
4. Datos procesados
5. Análisis sobre datos procesados
6. Interpretación

**IF YOU DO BIOINFORMATICS WITH
POOR QUALITY DATA**



**YOU'RE GONNA HAVE A BAD
TIME**



bac
ZARAGOZA 2021
Online

