



UNIVERSIDAD
DE CHILE

Introducción a ChIP-seq

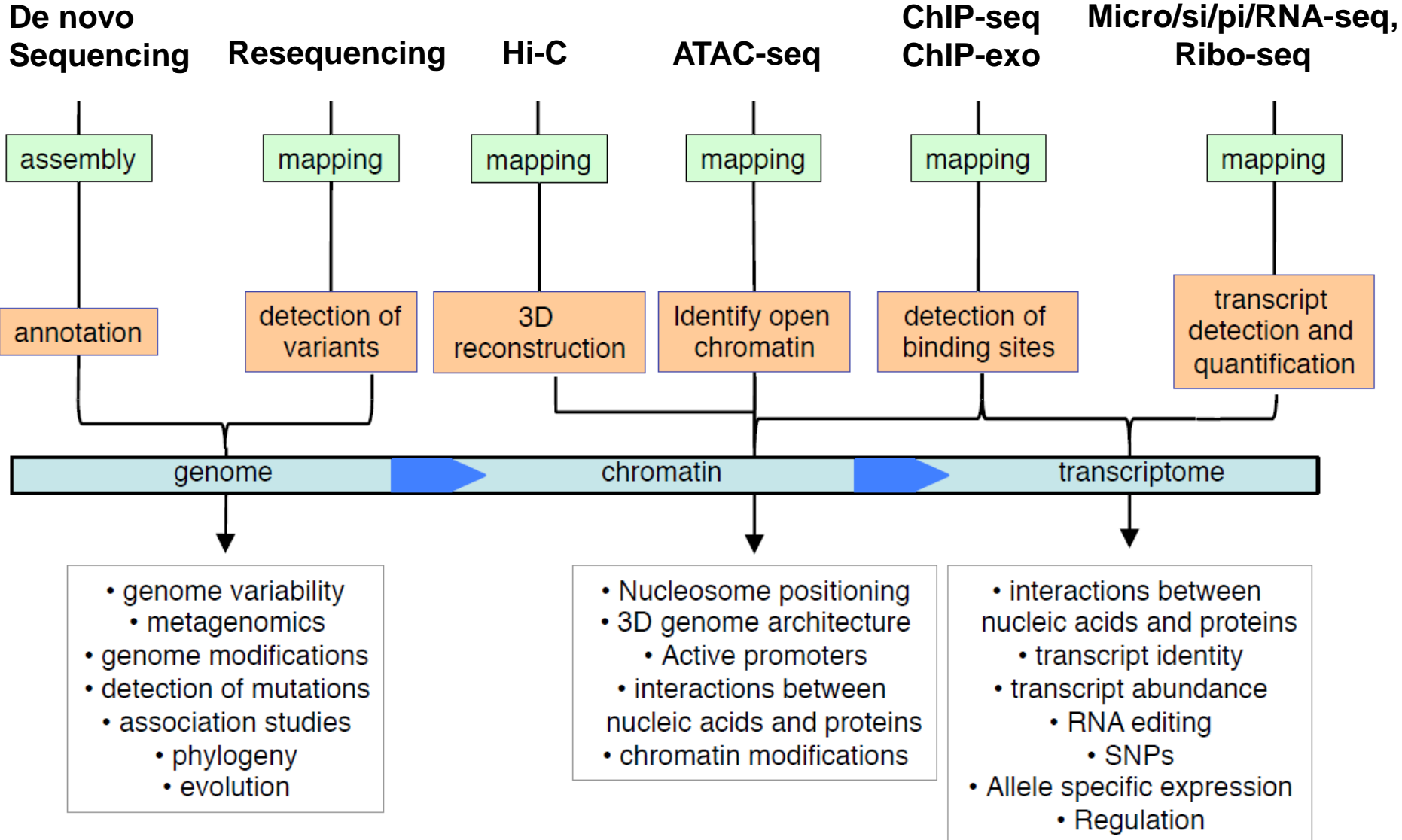
Luis Valenzuela Villa,
luis.valenz.v@gmail.com

Laboratorio de Genética, Facultad de Ciencias,
Laboratorio de BioMatemática y Ómica Integrativa, Facultad de Medicina,
Universidad de Chile.

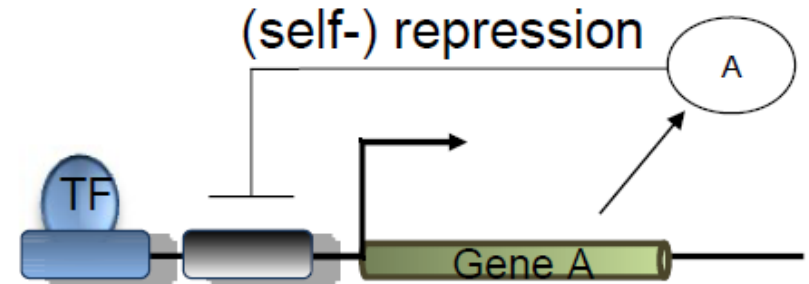
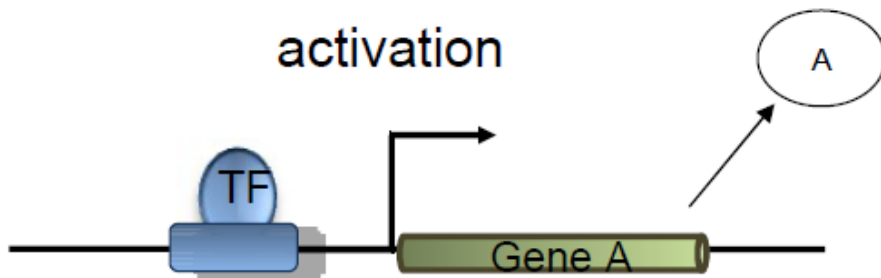
18 de Octubre, 2017



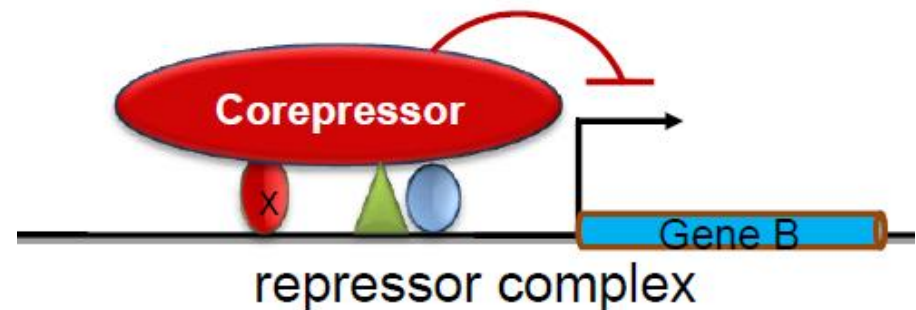
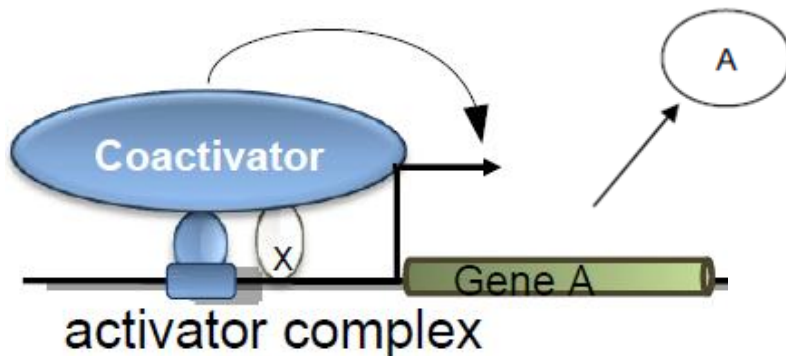
Introducción a ChIP-seq Aplicaciones de NGS



- Regulación Directa



- Regulación Indirecta





- **In silico**

- Alineamiento de secuencias
- herramientas de búsqueda: MathInspector, TFBSearch, JASPAR, Perl::TFBS, TRANSFAC.

- **In vivo**

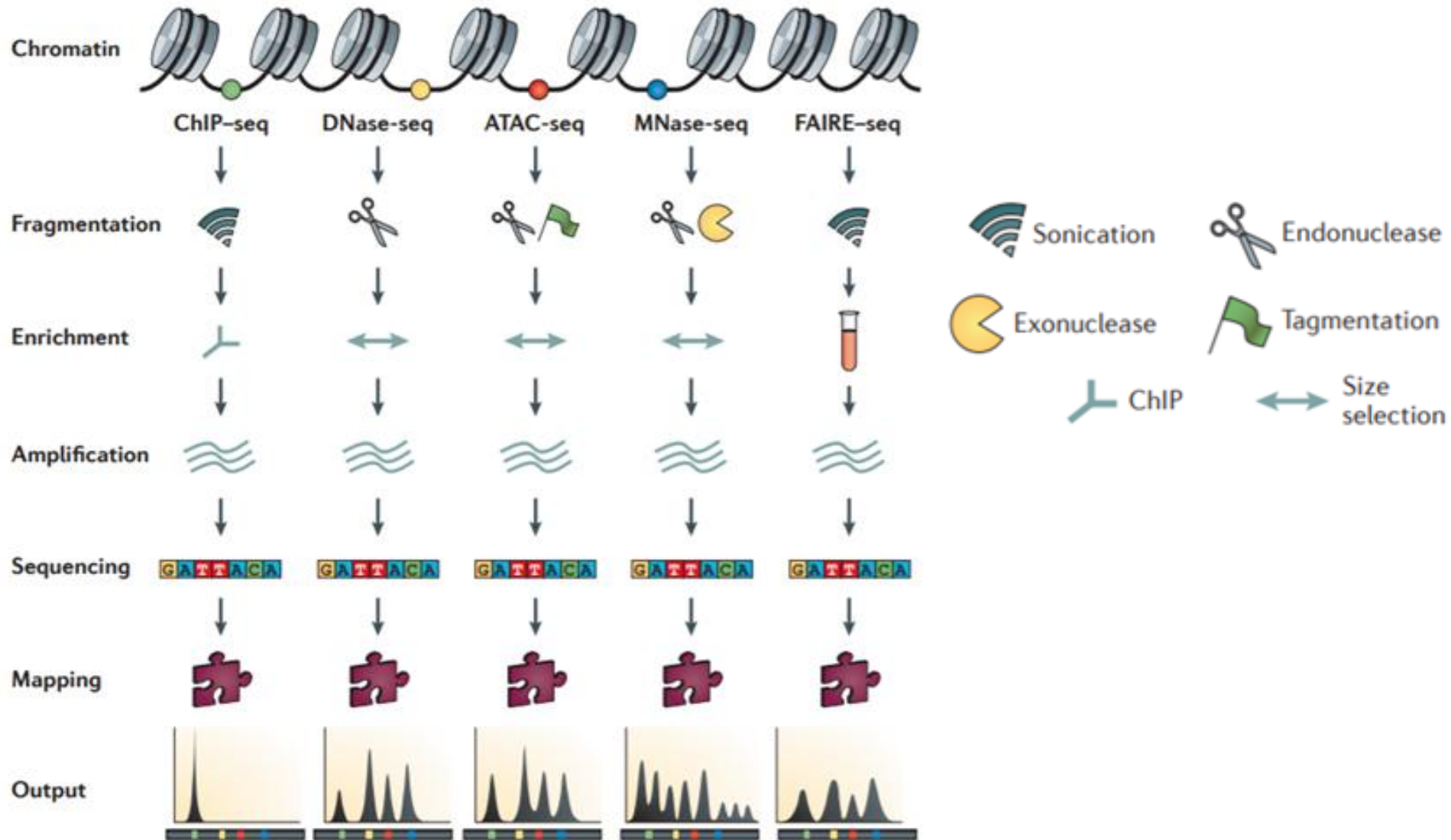
- DNase footprinting assay
- Electrophoretic Mobility Shift Assay (EMSA) (co-binding)
- ChIP-chip
- ChIP-seq



UNIVERSIDAD DE CHILE

Introducción a ChIP-seq

Identificación de sitios de unión de TFs: in vivo





UNIVERSIDAD DE CHILE

Introducción a ChIP-seq

Position frequency matrix y Position Weight Matrix



PFM

A	7	1	0	0	0	0	0	14	3	4	7
C	0	2	7	0	0	0	15	0	3	1	0
G	0	6	1	0	15	0	0	0	4	3	2
T	1	0	2	15	0	15	0	1	3	4	3

$$w = \log_2 \left(\frac{(f + \sqrt{(N)} \times p)}{(N + \sqrt{(N)})} \right) = \log_2 \left(\frac{(f + \sqrt{(N)} \times p)}{(N + \sqrt{(N)}) / p} \right) \longrightarrow \log_2 \frac{(7 + \sqrt{8} \times 0,25)}{(8 + \sqrt{8})} = 0,96265427$$

PWM

A	0,963	-1,100	-2,187	-2,187	-2,187	-2,187	-2,187	1,879	-0,058	0,272	0,963
C	-2,187	-0,487	0,963	-2,187	-2,187	-2,187	1,973	-2,187	-0,058	-1,100	-2,187
G	-2,187	0,767	-1,100	-2,187	1,973	-2,187	-2,187	-2,187	0,272	-0,058	-0,487
T	-1,100	-2,187	-0,487	1,973	-2,187	1,973	-2,187	-1,100	-0,058	0,272	-0,058



TGIF_01
consensus
sequence



JASPAR²⁰¹⁸

Cart ⁰ JASPAR Blog

Examples: SPI1, P17676, ChIP-seq, Homo sapiens

Advanced Options

Home

About

MIG1

Search

Brows

Examples: SPI1, P17676, ChIP-seq, Homo sapiens

Advanced Options

Brows

Tools

1 profile(s) found

RESTf

Display 10 profiles

Filter:

Down



ID

Name

Species

Class

Family

Logo

Matrix



MA0337.1

MIG1

Saccharomyces
cerevisiae

C2H2 zinc finger
factors

Other factors with up to three adjacent zinc
fingers

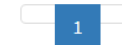


Genor

Copy

CSV

Showing 1 profiles of page 1 from 1 pages



Analyze selected profiles

Please select matrix profiles on the left side to add to your cart or perform the following analysis.

Add to cart

Scan

Input a (FASTA-formatted) sequence to scan with selected matrix models.

Load example sequence

Enter FASTA sequence here: (3000 nucleotides left)

Enter fasta-formatted sequence ...

Relative profile score threshold 80 %

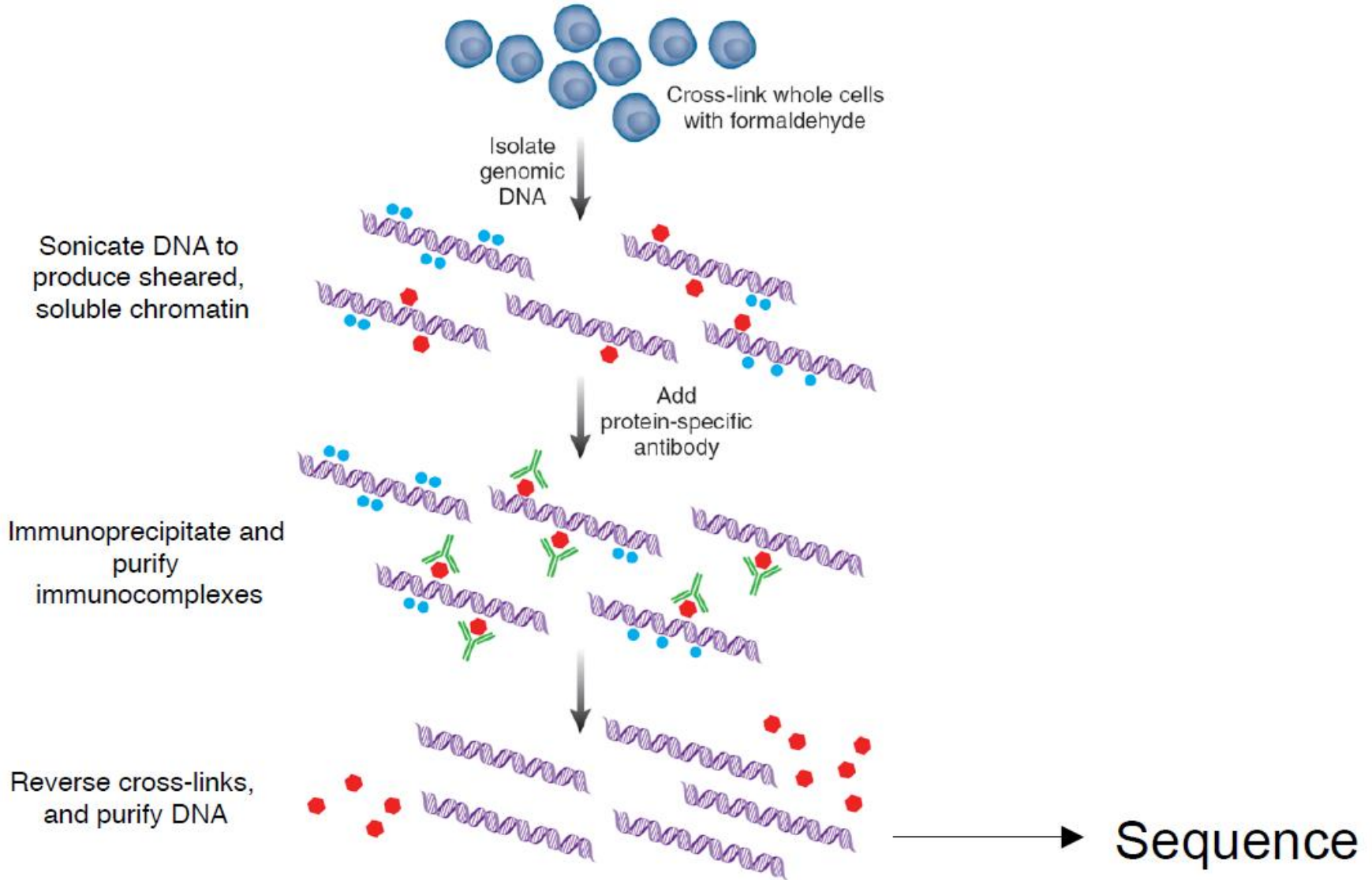
Scan



UNIVERSIDAD
DE CHILE

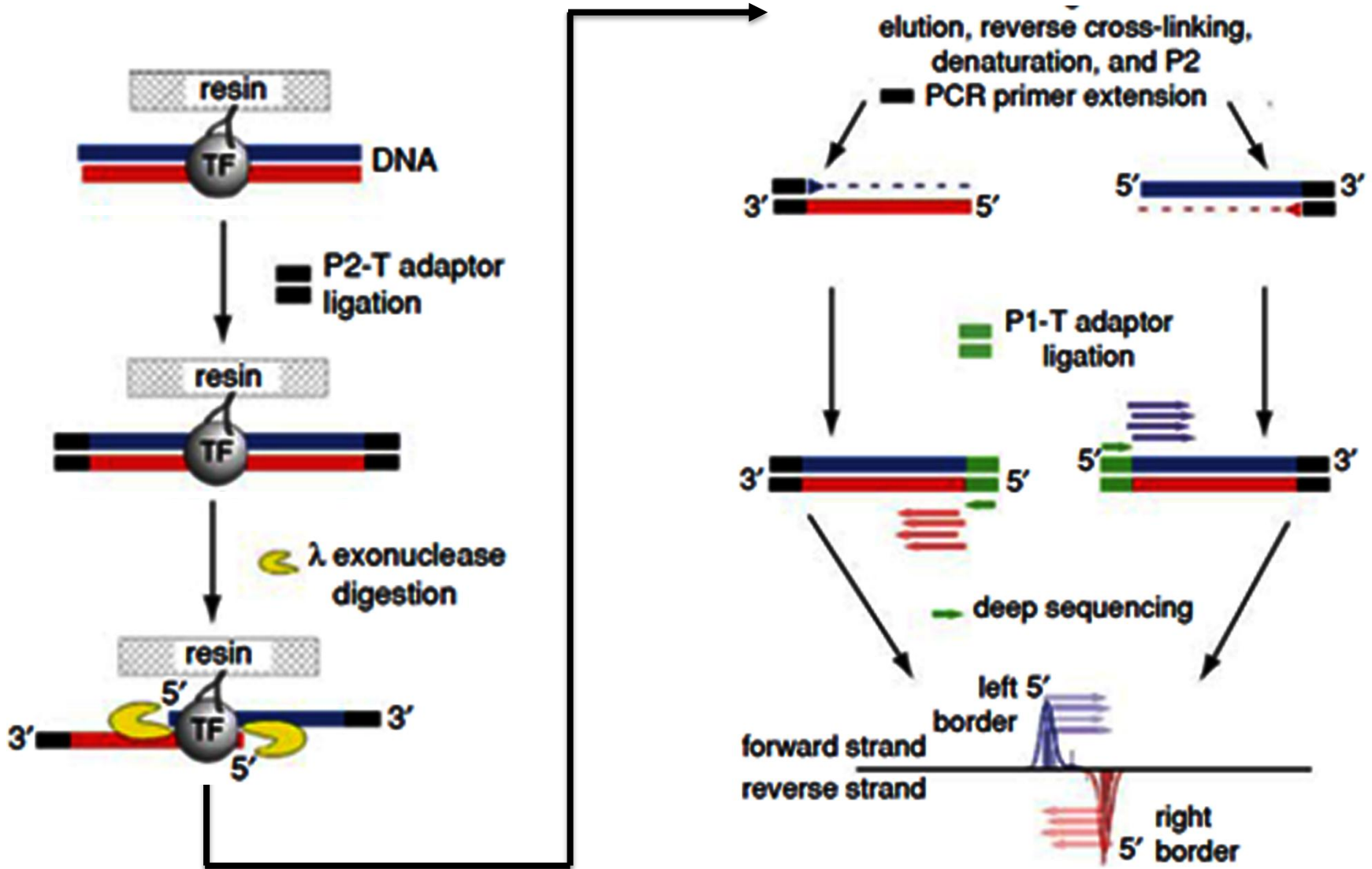
Introducción a ChIP-seq

ChIP-seq





Introducción a ChIP-seq ChIP-exo





UNIVERSIDAD
DE CHILE

Introducción a ChIP-seq

Consideraciones previas



¿Cuán bueno es el anticuerpo?

¿Se necesita controles?

Input: IP sin anticuerpo

IgG: IP con anticuerpo inespecifico

Profundidad de secuenciación: tamaño genoma, número y tamaño de los sitios de unión de la proteína..

¿Cuántas replicas?

Con más de 2 no aumentamos el número de targets.

¿Cuántos reads se necesitan?

20M de reads para mamíferos, 4M para moscas y c. elegans.

(si es Pol II, 60M)

Generalidad: >10M reads para peaks finos, >20M para peaks anchos

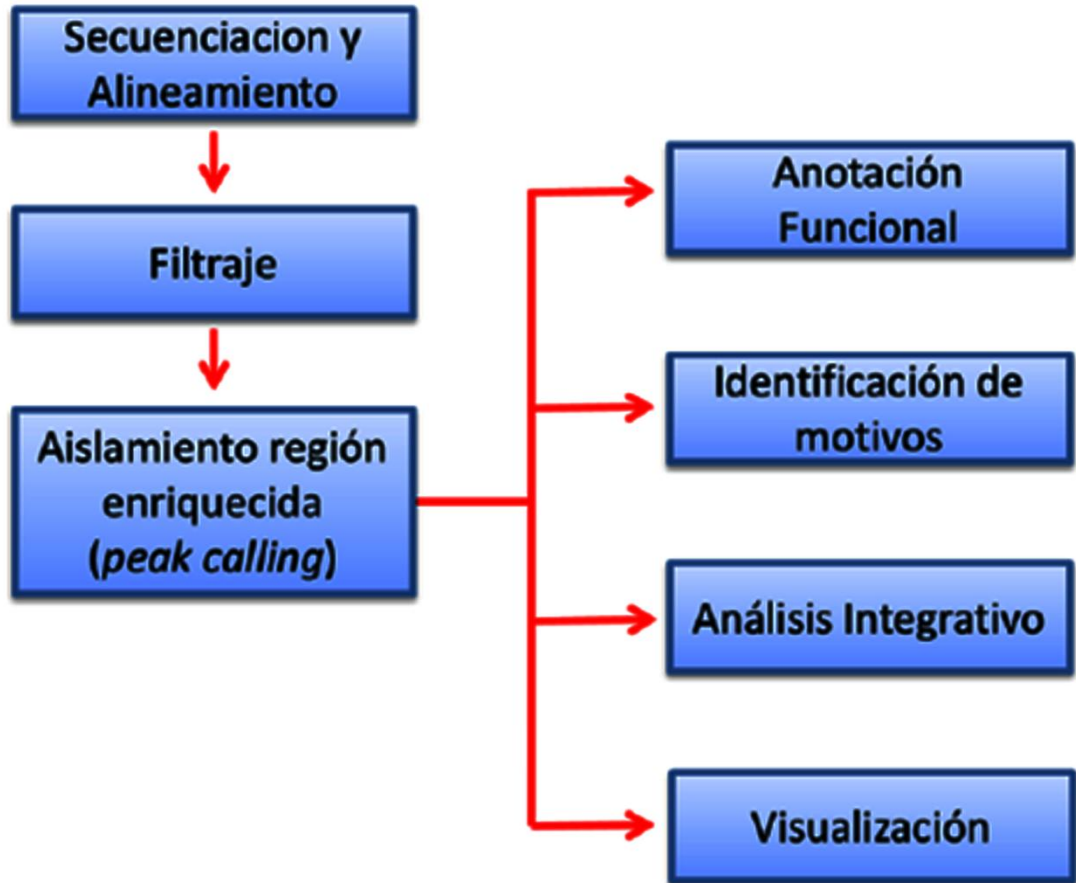


UNIVERSIDAD
DE CHILE

Introducción a ChIP-seq Flujo de trabajo



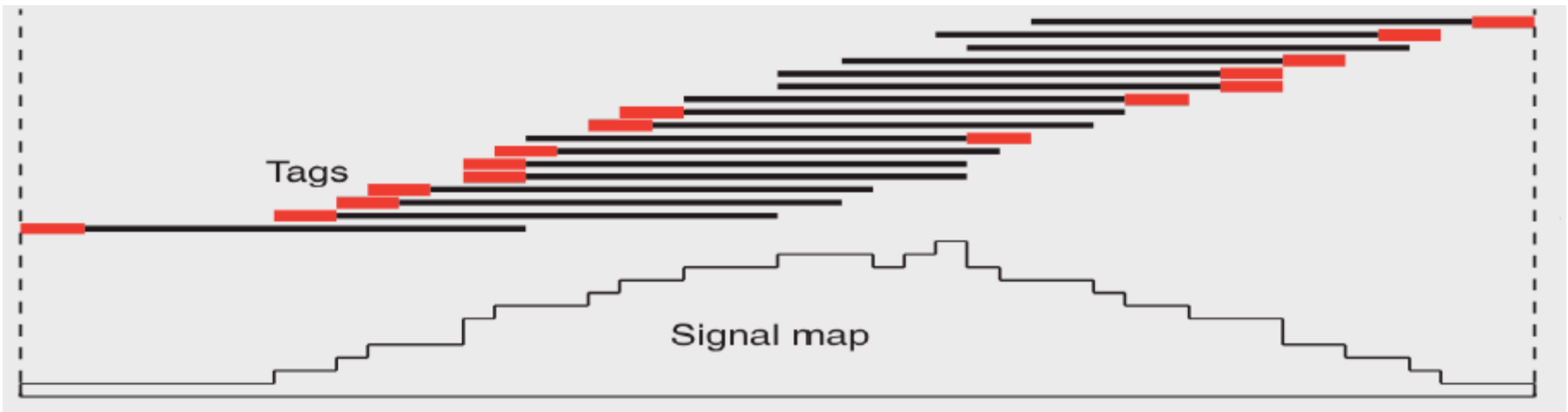
Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS





UNIVERSIDAD
DE CHILE

Introducción a ChIP-seq

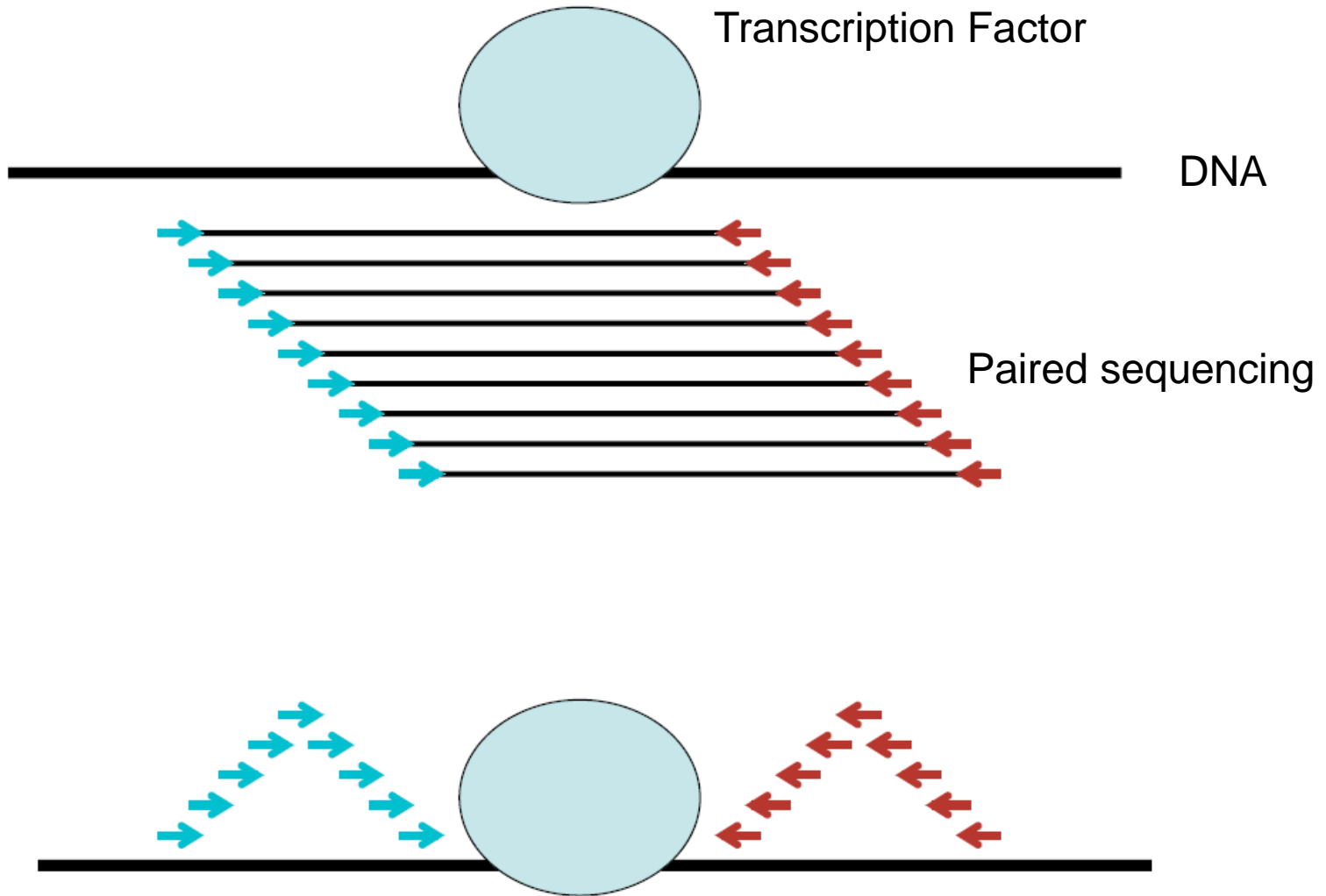




UNIVERSIDAD DE CHILE

Introducción a ChIP-seq

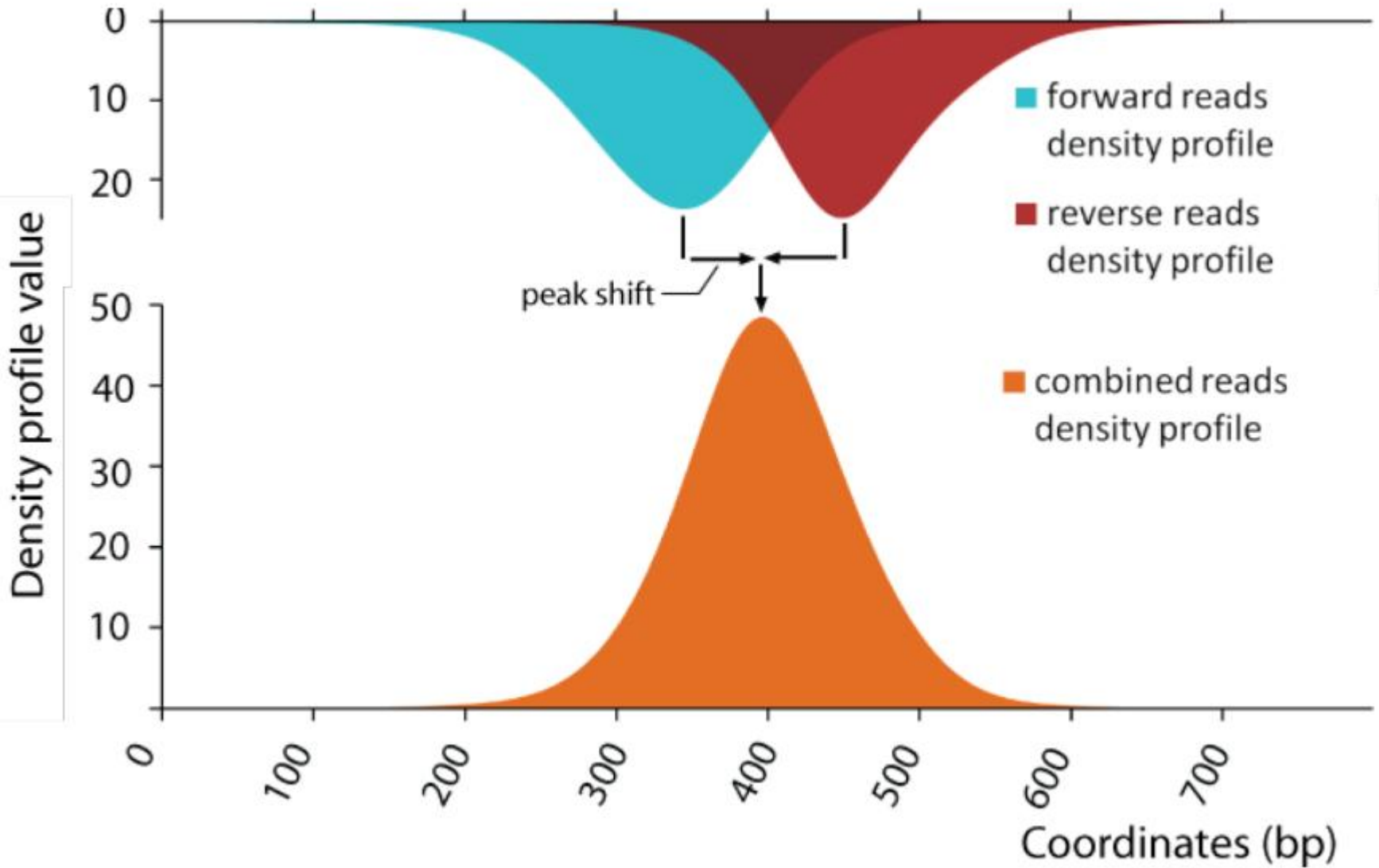
ChIP-seq o ChIP-exo: Peak Calling





Introducción a ChIP-seq

ChIP-seq o ChIP-exo: Peak Calling

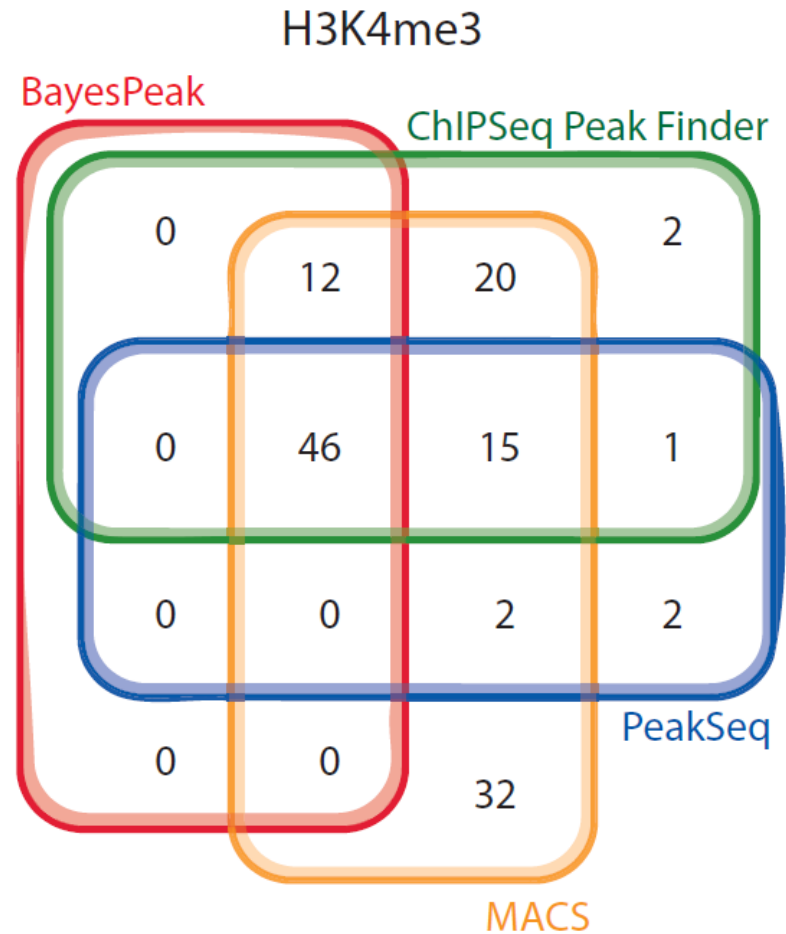
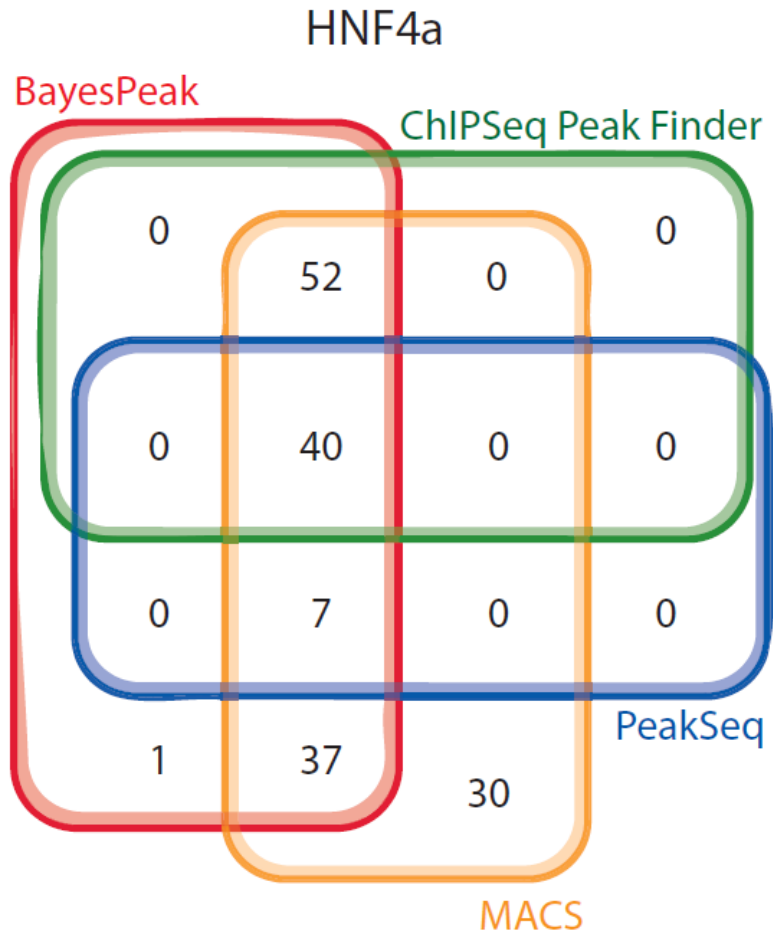




Introducción a ChIP-seq Peak Calling



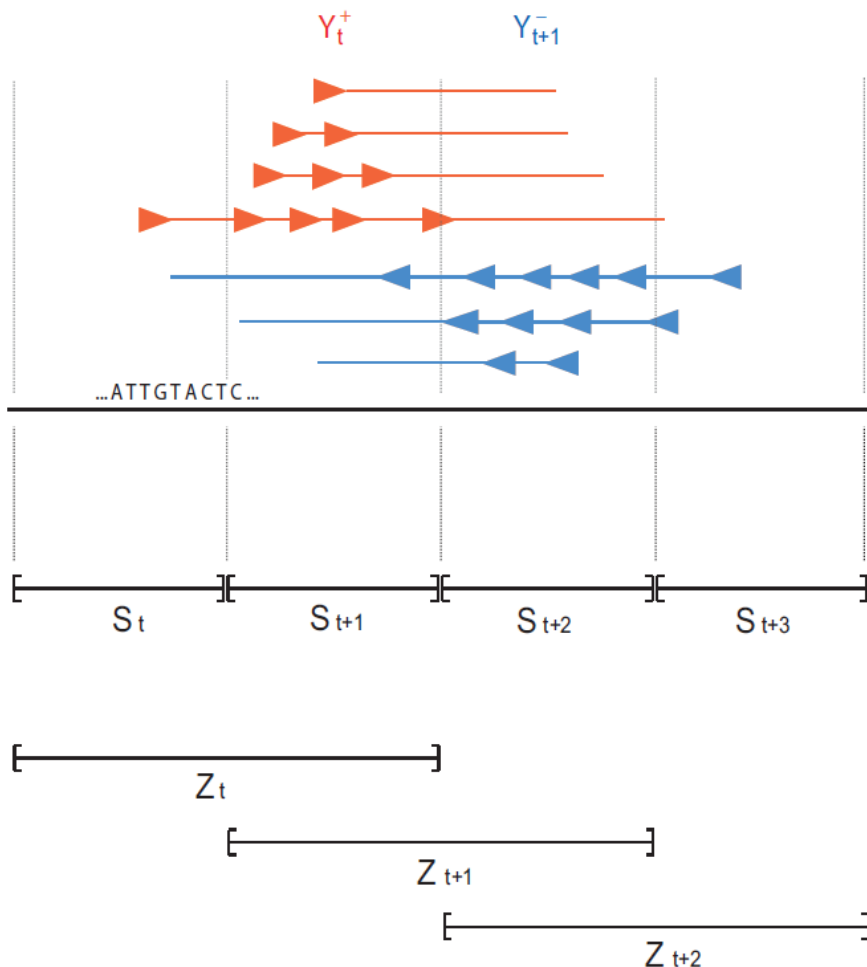
Method	A	B	C	D	E	F	G
CSPF	control or IP only	read length no orientation	merge strands no shift	N	simple height criteria	ROC curve (empirically)	both
XSET	IP only	fragment length orientation	merge strands no shift	Y	simple height criteria	FDR estimate using Poisson distribution	both
Mikkelsen et al.	IP only	no orientation	no merge no shift	Y	p -values from permutations	no official FDR	both
MACS	control or IP only	fragment length orientation no duplicated reads	shift reads merge strands	N	Poisson p -values	FDR estimate by peaks in control:IP	both
QuEST	control	orientation	shift reads merge strands	N	kernel density estimation	FDR estimate by permutations of the control	better for TF
FindPeaks	IP only	fragment length orientation	no merge no shift	N	simple height criteria	FDR estimate by permutations of the IP	both
SISSR	control or IP only	fragment length orientation	no merge no shift	N	compares reads on different strands	FDR estimate by peaks in background:IP	better for TF
Kharchenko et al.	control	orientation	no merge no shift	N	Poisson distribution	FDR estimate by permutations of the control	better for TF
PeakSeq	control	fragment length orientation	merge strands	Y	sample normalisation Binomial distribution	FDR estimate, q - values (BH correction)	both
BayesPeak	control or IP only	fragment length orientation	no merge no shift	N	negative binomial distribution, Bayesian posterior probabilities	posterior enrichment probabilities	both





UNIVERSIDAD DE CHILE

Introducción a RNA-seq Peak Calling: BayesPeak



$$Z_t = \begin{cases} 0 & \text{if } (S_t, S_{t+1}) = (0, 0) \\ 1 & \text{if } (S_t, S_{t+1}) = (0, 1) \\ 2 & \text{if } (S_t, S_{t+1}) = (1, 0) \\ 3 & \text{if } (S_t, S_{t+1}) = (1, 1) \end{cases}$$

$$Y_t^+, Y_{t+1}^- \mid Z_t = 0 \sim \text{Poisson}(\lambda_0 \gamma^{w_t})$$

$$Y_t^+, Y_{t+1}^- \mid Z_t = 1, 2, 3 \sim \text{Poisson}((\lambda_0 + \lambda_1) \gamma^{w_t})$$

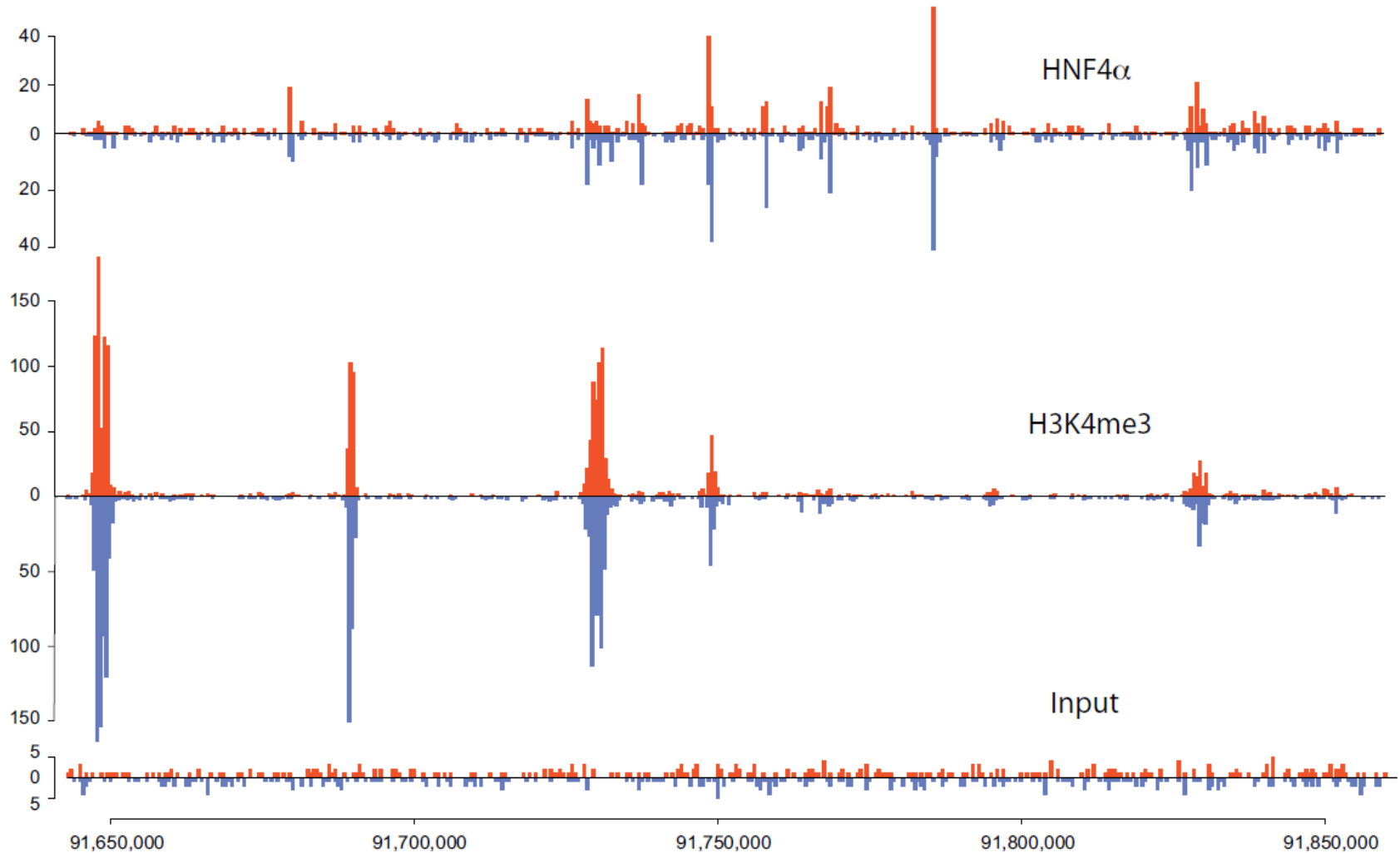
$$\lambda_0 \sim \Gamma(\alpha_0, \beta_0)$$

$$\lambda_1 \sim \Gamma(\alpha_1, \beta_1)$$



UNIVERSIDAD DE CHILE

Introducción a ChIP-seq Peak Calling



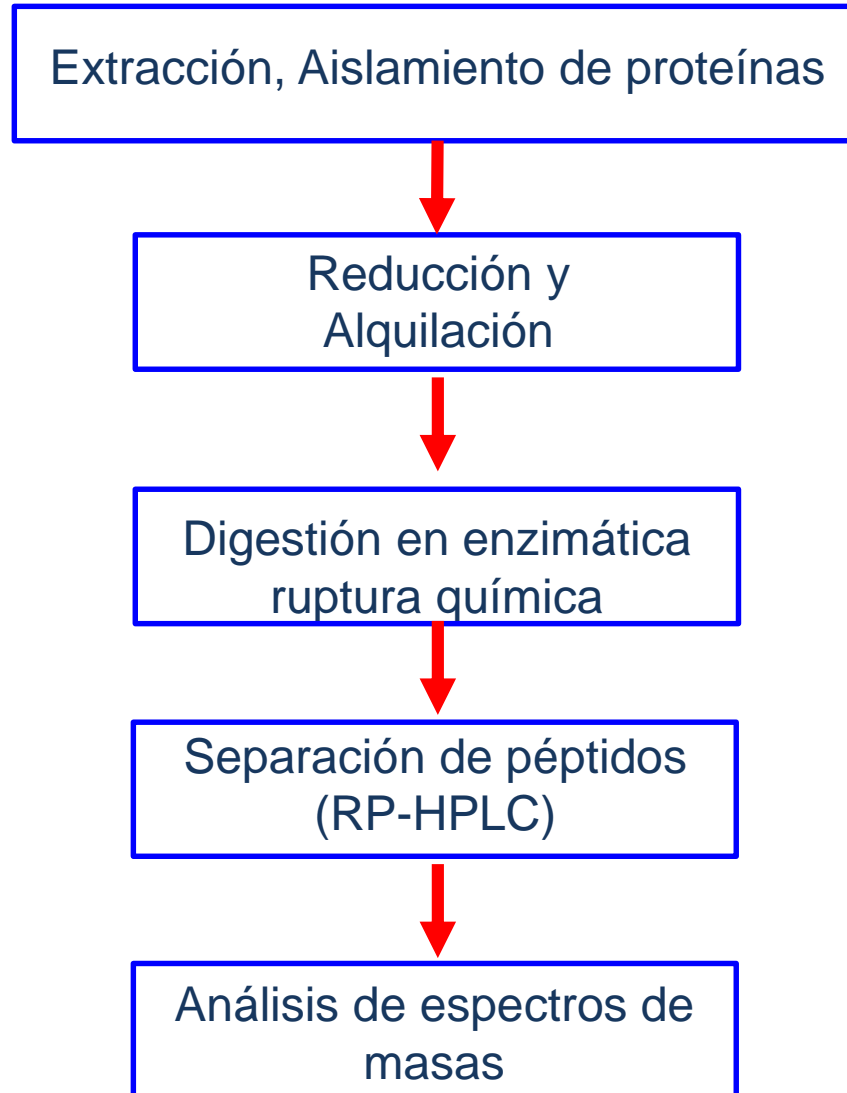
BMC Bioinformatics 2009, 10:299

- 1970: Geles en 2D y separación de cientos de proteínas.
- 1990: Expansión masiva
MALDI, ESI.
Genomas, bases de datos.
- 2000: Consolidación
Cuantificación relativa y absoluta
Modificaciones postraduccionales
aplicaciones biológicas y análisis dirigido



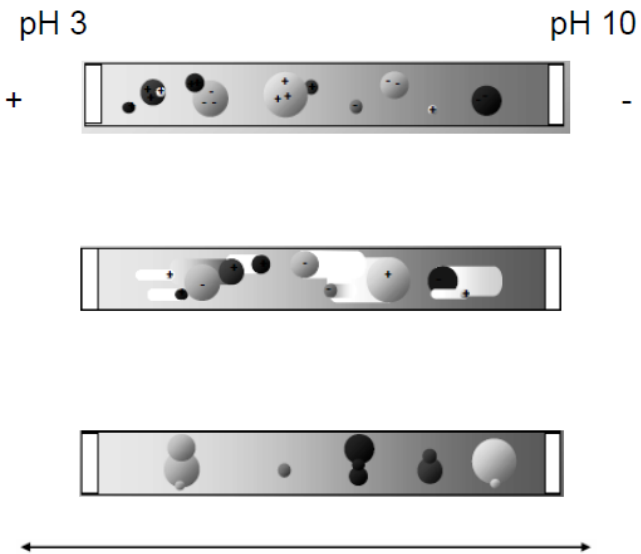
Introducción a Proteómica

Flujo de trabajo con Proteómica





Isoelectroenfoque



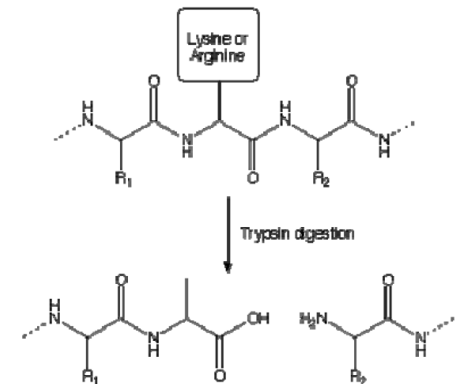
SDS-PAGE



CONTROL



EXPERIMENTAL





UNIVERSIDAD
DE CHILE

Introducción a Proteómica

2. Flujo de trabajo con Proteómica



Ionización de los fragmentos

-MALDI: Matrix Assisted Laser Desorption Ionization

-ESI: Electrospray Ionization

Acelerar a través de un espectrometro de masa

Producir los “peptide mass fingerprint”s



UNIVERSIDAD
DE CHILE

Introducción a Proteomica Peptide Mass Fingerprints

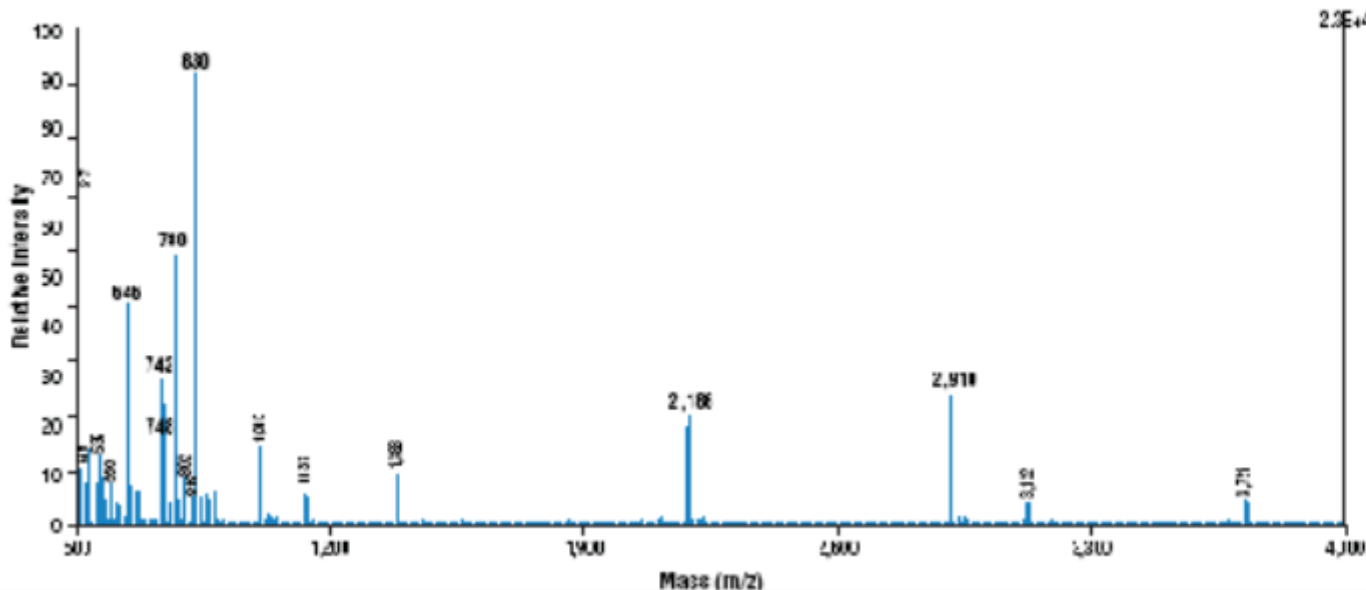


¿Qué es?

Método para identificar proteínas basado en las masas de los péptidos generados a partir de ella.

Cada proteína genera un patrón específico de péptidos luego de la digestión por una proteasa dada.

Los patrones de las masas son los que permiten la identificación.



2978.4567
2646.2992
2186.1678
1981.8621
1131.4384
830.4519
780.4978
748.3698
742.4497
646.3228
....



Introducción a Proteomica Peptide Mass Fingerprints



```
>sp|P02666|CASB_BOVIN Beta-casein OS=Bos taurus
MKVLILACLVALALARELEELNVPGEIVESLSSEESITRINKKIEKFQSEEQQQTEDEL
QDKIHFAQTQSLVYFPFGPIPNLPLTQTPVVVPPFLQPEVMGVSKVKEAMAPK
HKEMPFKYPVEPFTESQSLTLTDVENLHLPLLLQSWMHQPHQPLPPTVMFPPQSVLSL
SQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLPVVRGPFPIIV
```

```
MK VLILACIVALALAR ELEELNVPGEIVESLSSEESITR INKKIEK FQSEEQQQTEDEL
QDK IHFAQTQSLVYFPFGPIPNLPLTQTPVVVPPFLQPEVMGVSKVK EAMAPK
HK EMPEPK YPVEPFTESQSLTLTDVENLHLPLLLQSWMHQPHQPLPPTVMFPPQSVLSL
SQSK VLPVPQK AVPYPQR DMPIQAFLLYQEPVLPVVR GPFPIIV
```

2646.2992
2186.1678
1981.8621
830.4519
780.4978
748.3698
742.4497
646.3228

Match

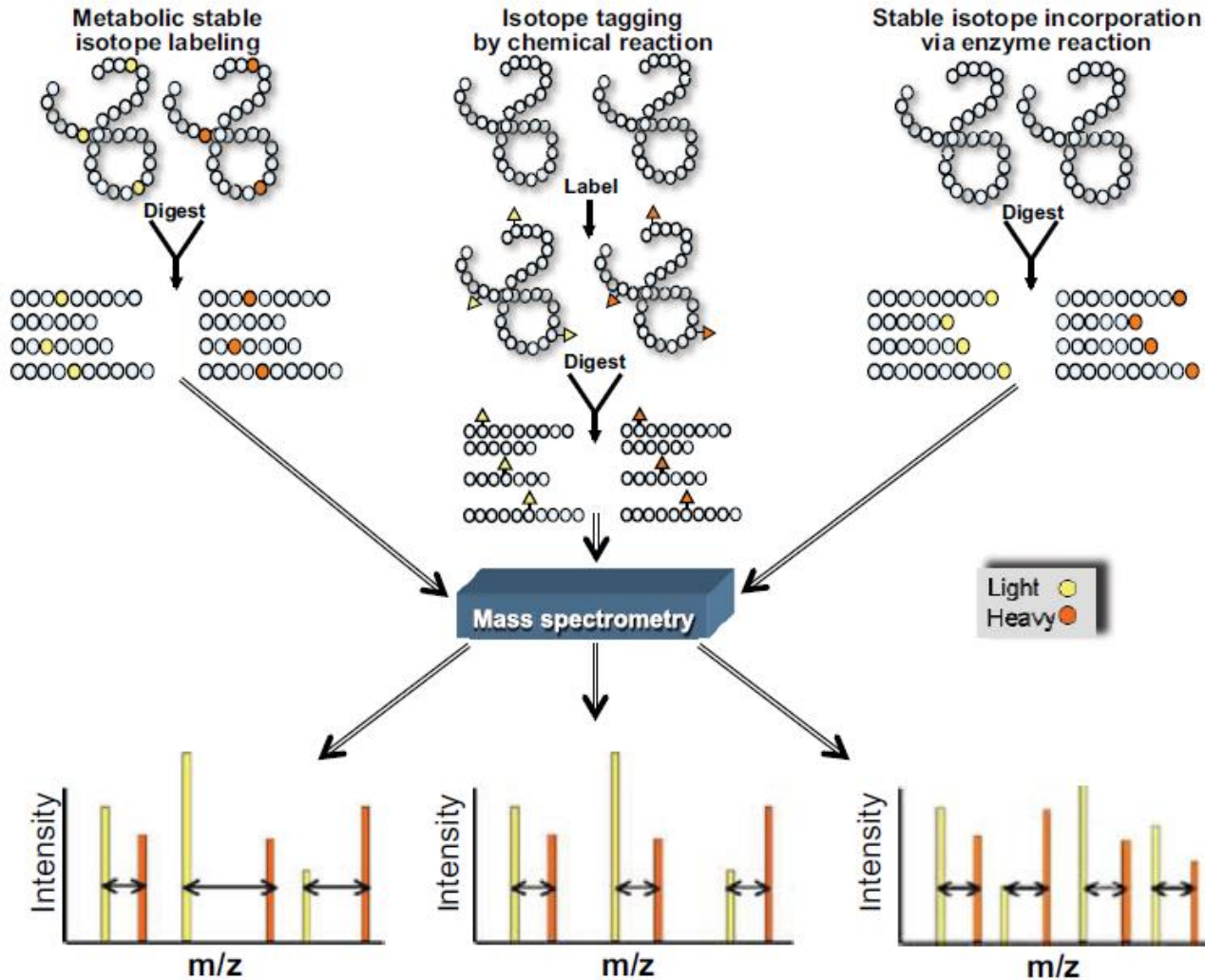
2978.4567
2646.2992
2186.1678
1981.8621
1131.4384
830.4519
780.4978
748.3698
742.4497
646.3228
....



Introducción a Proteómica Proteómica Cuantitativa



Estrategias para marcar con isotopos estables





UNIVERSIDAD
DE CHILE

Introducción a ChIP-seq y Proteómica



- Práctico.