TP 6: Transcriptómica

Biología Celular y Molecular
Lic. en Ciencias Biológicas
Universidad Nacional del Comahue

Transcriptómica

Contexto

Información Genética

- Composición genómica de la especie
- Varicación intraespecífica (alelos)

→

Expresión Génica

- Transcripción
- Traducción

→

Procesos Biológicos

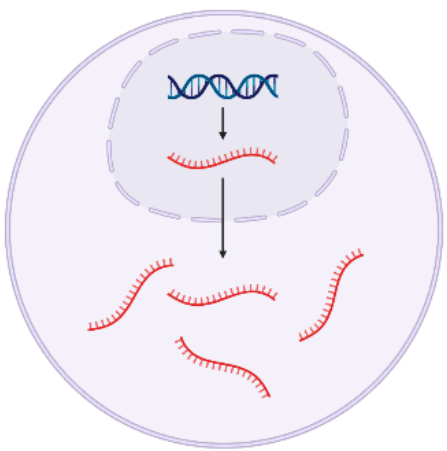
- Fisiología
- Desarrollo
- Comportamiento

Transcriptómica

Northern blot

Expresión Génica

- Transcripción

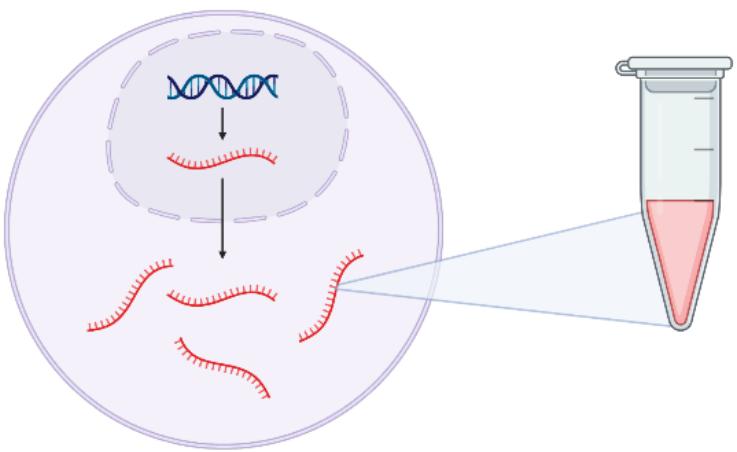


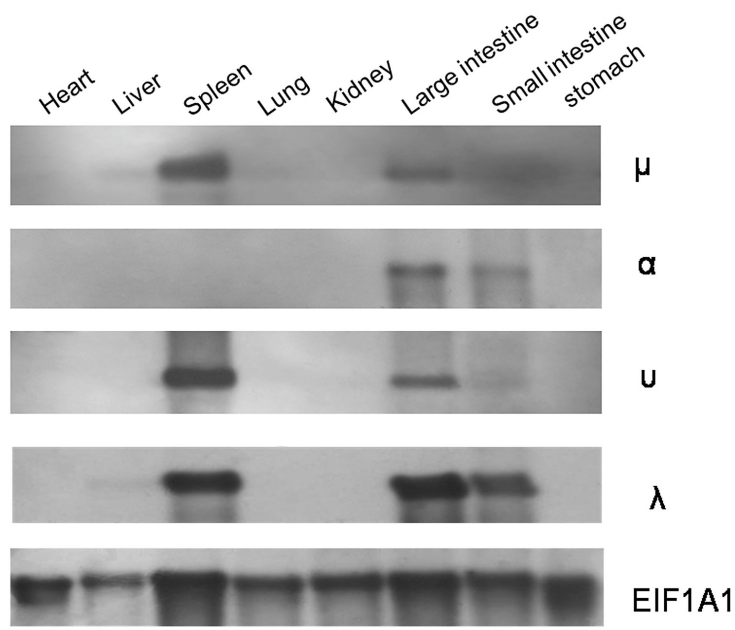
Transcriptómica

Northern blot

Expresión Génica

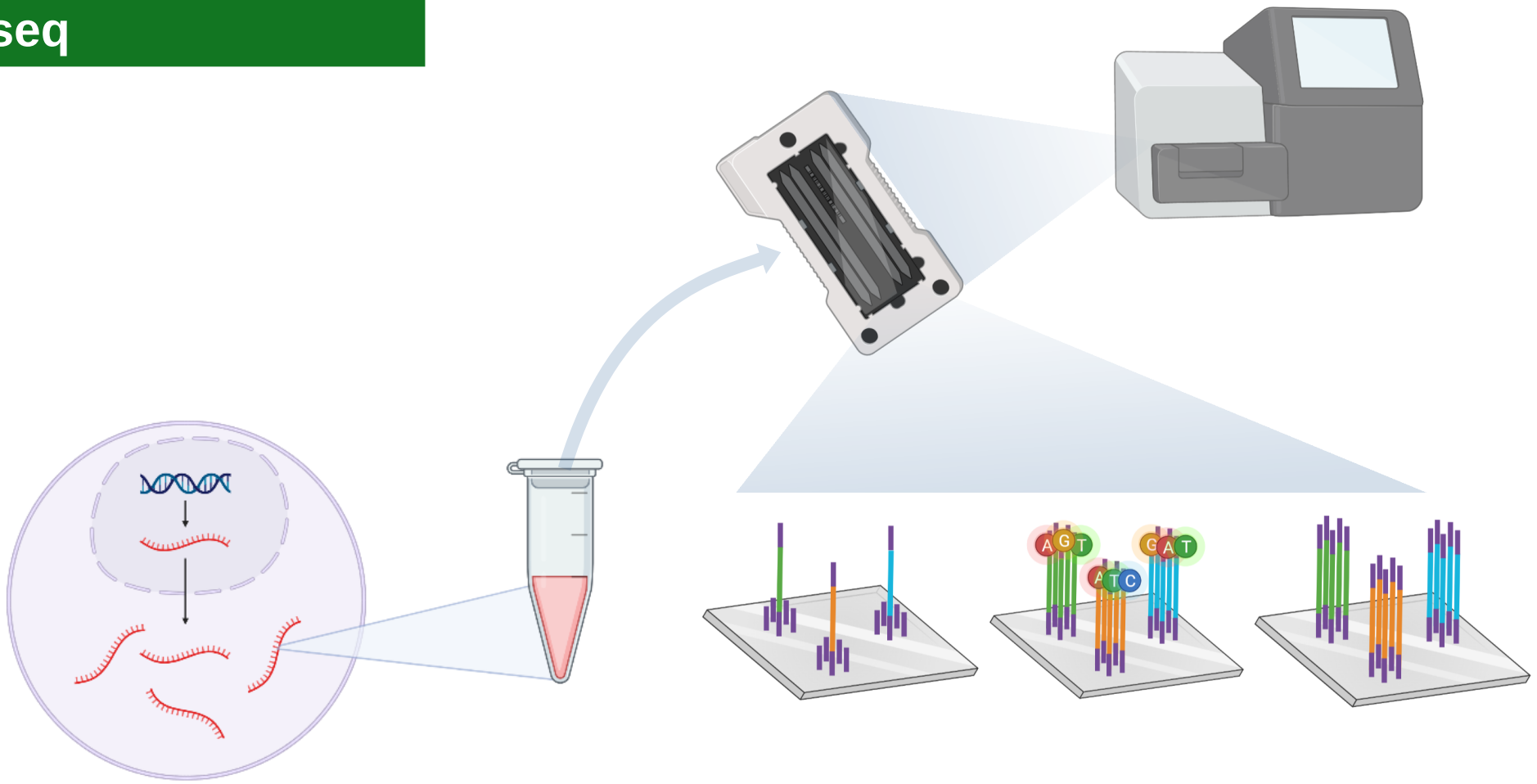
- Transcripción





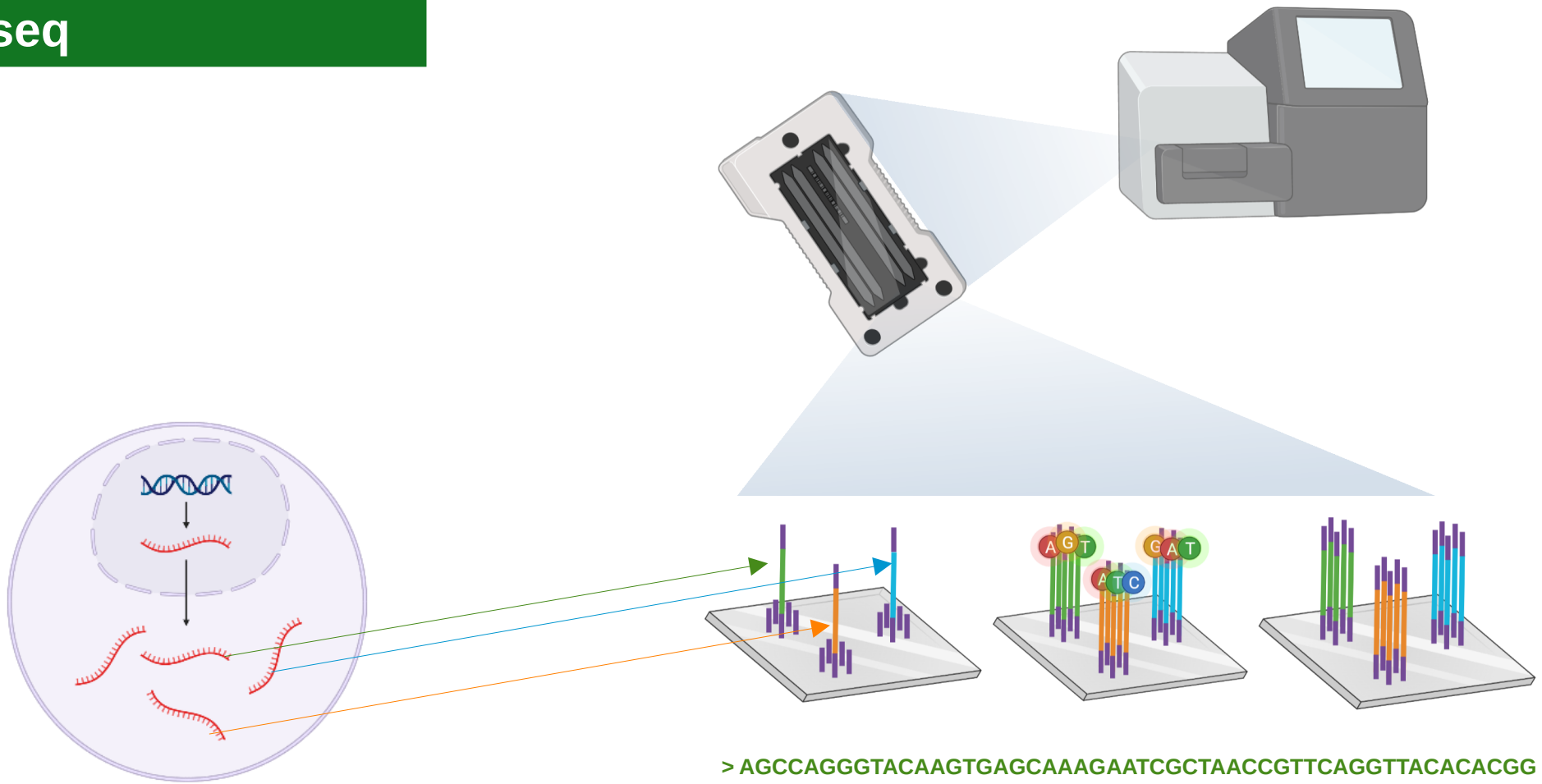
Transcriptómica

RNAseq



Transcriptómica

RNAseq



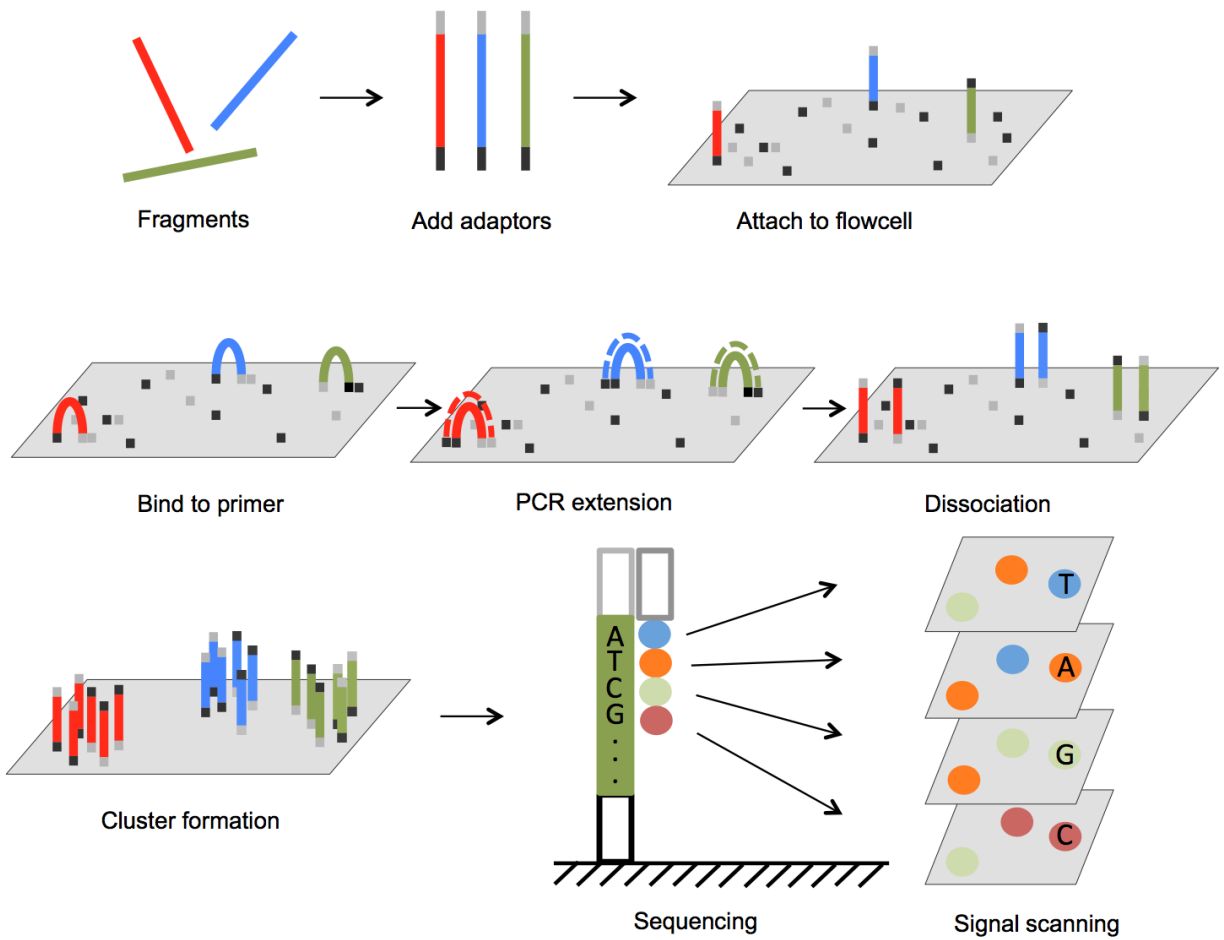
FASTA

```
> AGCCAGGGTACAAGTGAGCAAAGAATCGCTAACCGTTCAGGTTACACACGG
> TAGCCAGGGTACAAGTGAGCAAAGAATCGCTAACCGTTCAGGTTACA
> CTAGCCAGGGTACAAGTGAGCAAAGAATCGCTAACCGTTCAGGTTACACACGG
```

Transcriptómica

RNAseq

https://www.youtube.com/watch?v=fCd6B5HRaZ8&ab_channel=Illumina



Archivo FASTA

```
> Read_1
AGCCAGGGTACAAGTGAGCAAAGAATCGCTAGAGCAA

> Read_2
TAGCCAGGGTACAAAAAGAGTGAGCGTGAGCACA

> Read_3
CTAGCCAGGGTACAAGTGAGCAAAGAATCAGCAAAGCAA
```

Datos de RNAseq

Alineamiento/Mapeo

TAGCCAGGGTACAAGTGAGCAAAGAATCGCTAACCGTTCAGGTTACACACGGGGATATCAGTTCGTACAA

ACTAGCCAGGGTACAAGTGAGCAAAGAATCGCTAACCGTTCAGGTTACACACGGGGATATCAGTTCGTA

AAGGAAGAAGACTAGCCAGGGTACAAGTGAGCAAAGAATCGCTAACCGTTCAGGTTACACACGG

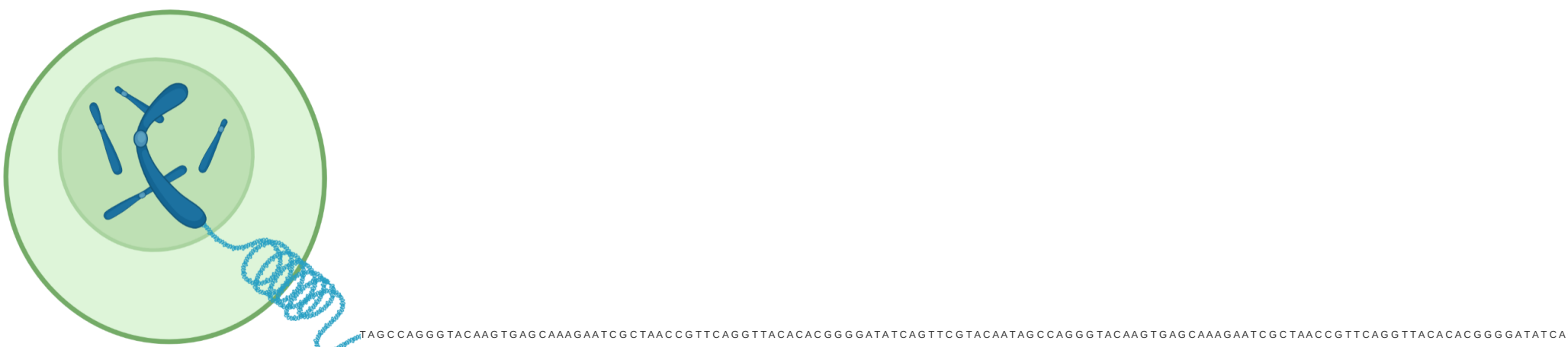
Datos de RNAseq

Alineamiento/Mapeo

TAGCCAGGGTACAAGTGAGCAAAGAATCGCTAACCGTTCAGGTTACACACGGGGATATCAGTTCGTACAA

ACTAGCCAGGGTACAAGTGAGCAAAGAATCGCTAACCGTTCAGGTTACACACGGGGATATCAGTTCGTA

AAGGAAGAAGACTAGCCAGGGTACAAGTGAGCAAAGAATCGCTAACCGTTCAGGTTACACACGG



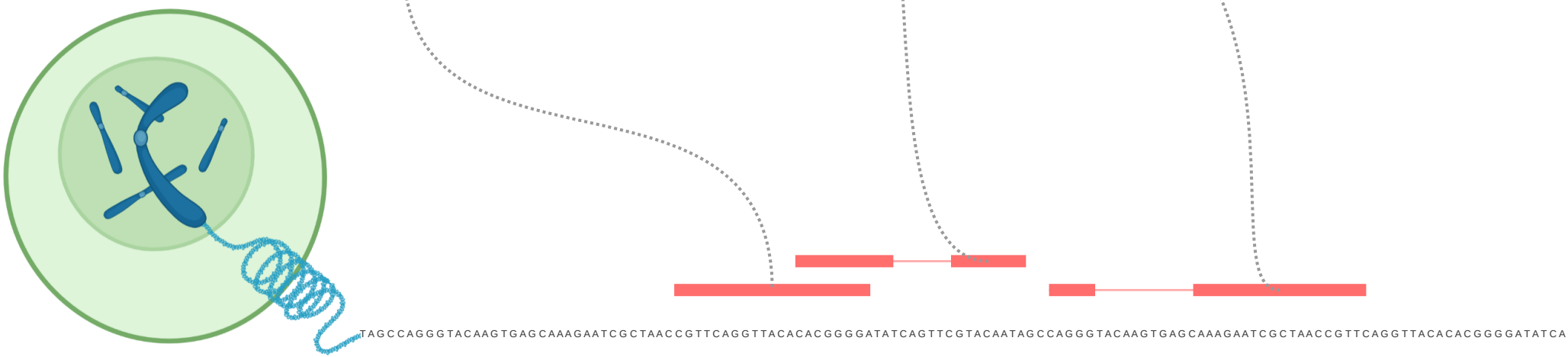
Datos de RNAseq

Alineamiento/Mapeo

TAGCCAGGGTACAAGTGAGCAAAGAATCGCTAACCGTTCAGGTTACACACGGGGATATCAGTTCGTACAA

ACTAGCCAGGGTACAAGTGAGCAAAGAATCGCTAACCGTTCAGGTTACACACGGGGATATCAGTTCGTA

AAGGAAGAAGACTAGCCAGGGTACAAGTGAGCAAAGAATCGCTAACCGTTCAGGTTACACACGG



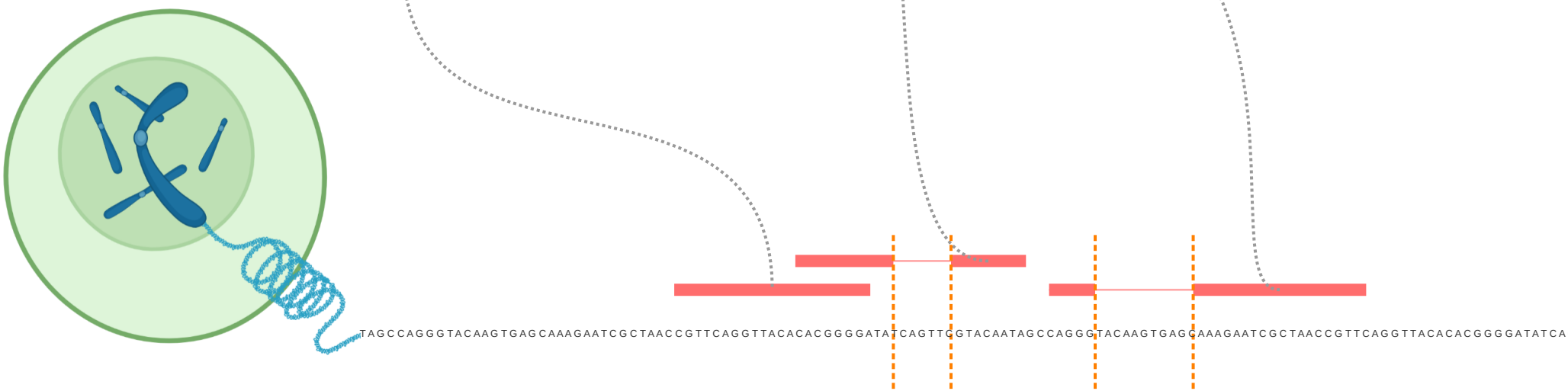
Datos de RNAseq

Alineamiento/Mapeo

TAGCCAGGGTACAAGTGAGCAAAGAATCGCTAACCGTTCAGGTTACACACGGGGATATCAGTTCGTACAA

ACTAGCCAGGGTACAAGTGAGCAAAGAATCGCTAACCGTTCAGGTTACACACGGGGATATCAGTTCGTA

AAGGAAGAAGACTAGCCAGGGTACAAGTGAGCAAAGAATCGCTAACCGTTCAGGTTACACACGG



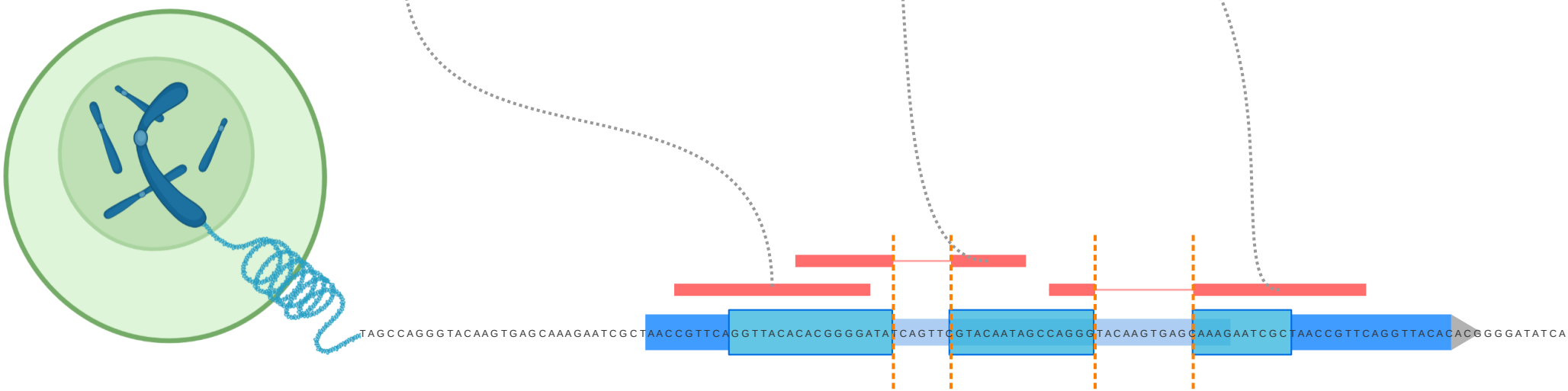
Datos de RNAseq

Alineamiento/Mapeo

TAGCCAGGGTACAAGTGAGCAAAGAATCGCTAACCGTTCAGGTTACACACGGGGATATCAGTTCGTACAA

ACTAGCCAGGGTACAAGTGAGCAAAGAATCGCTAACCGTTCAGGTTACACACGGGGATATCAGTTCGTA

AAGGAAGAAGACTAGCCAGGGTACAAGTGAGCAAAGAATCGCTAACCGTTCAGGTTACACACGG



Datos de RNAseq

Modelos de genes

- Identificar regiones codificantes en un genoma de referencia

GCCAGGGTACAAGTGAGCAAAGAATCGCTAACCGTTCAGGTTACACACGGGGATATCAGTTCGTACAATAGCCAGGGTACAAGTGAGCAAAGAATCGCTAACCGTTCAGGTTACACACGGGGATATCAGTTCGAAGTGAGCAAAGAATCGCTAACCGTTCAGGTTACACACGGG

Datos de RNAseq

Modelos de genes

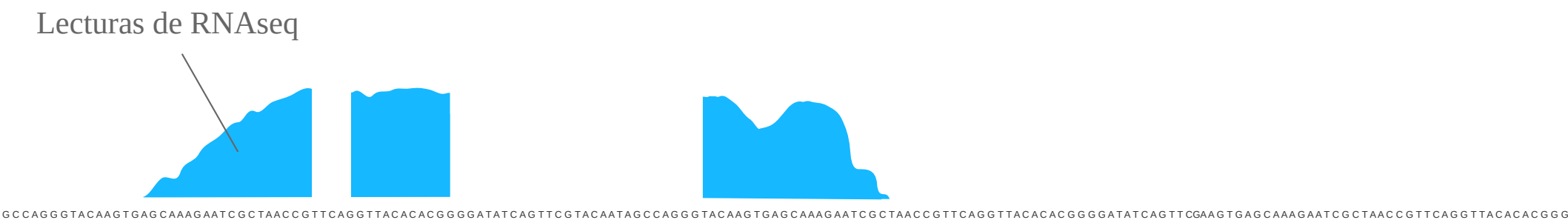
- Identificar regiones codificantes en un genoma de referencia (generar anotaciones)

GCCAGGGTACAAGTGAGCAAAGAATCGCTAACCGTTCAGGTTACACACGGGGATATCAGTTCGTACAATAGCCAGGGTACAAGTGAGCAAAGAATCGCTAACCGTTCAGGTTACACACGGGGATATCAGTTCGAAGTGAGCAAAGAATCGCTAACCGTTCAGGTTACACACGGG

Datos de RNAseq

Modelos de genes

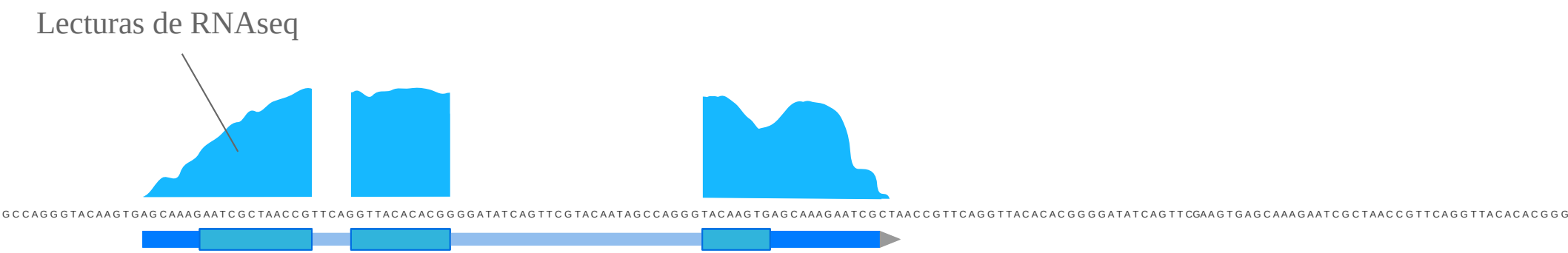
- Identificar regiones codificantes en un genoma de referencia (generar anotaciones)



Datos de RNAseq

Modelos de genes

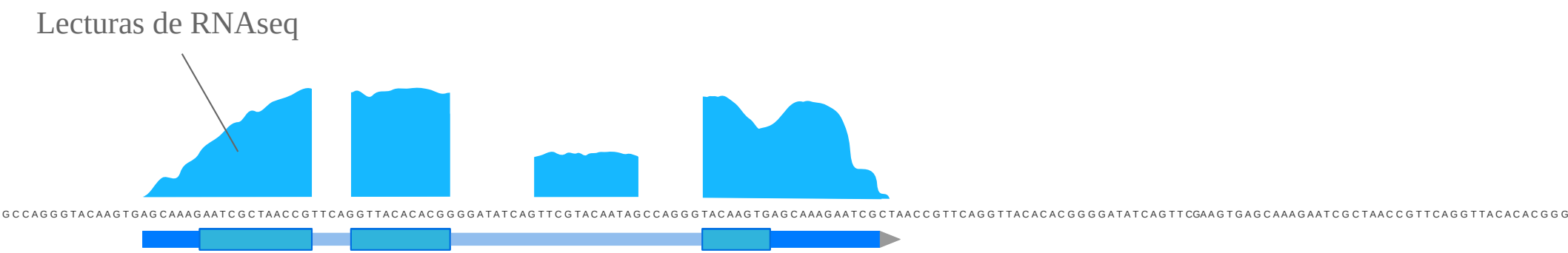
- Identificar regiones codificantes en un genoma de referencia (generar anotaciones)



Datos de RNAseq

Modelos de genes

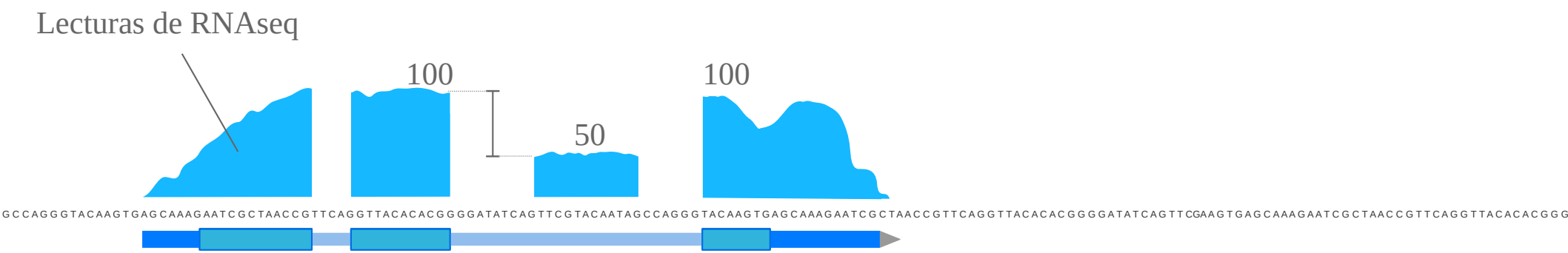
- Identificar regiones codificantes en un genoma de referencia (generar anotaciones)



Datos de RNAseq

Modelos de genes

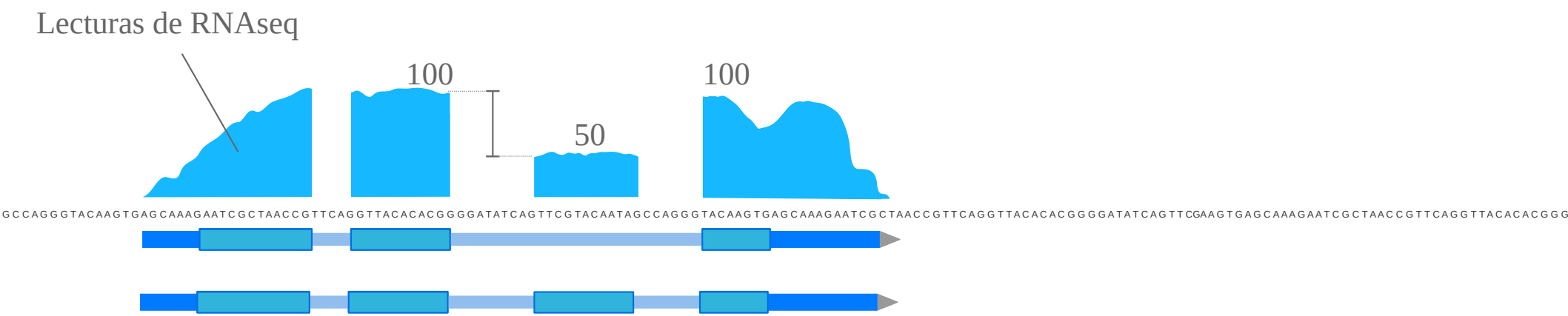
- Identificar regiones codificantes en un genoma de referencia (generar anotaciones)



Datos de RNAseq

Modelos de genes

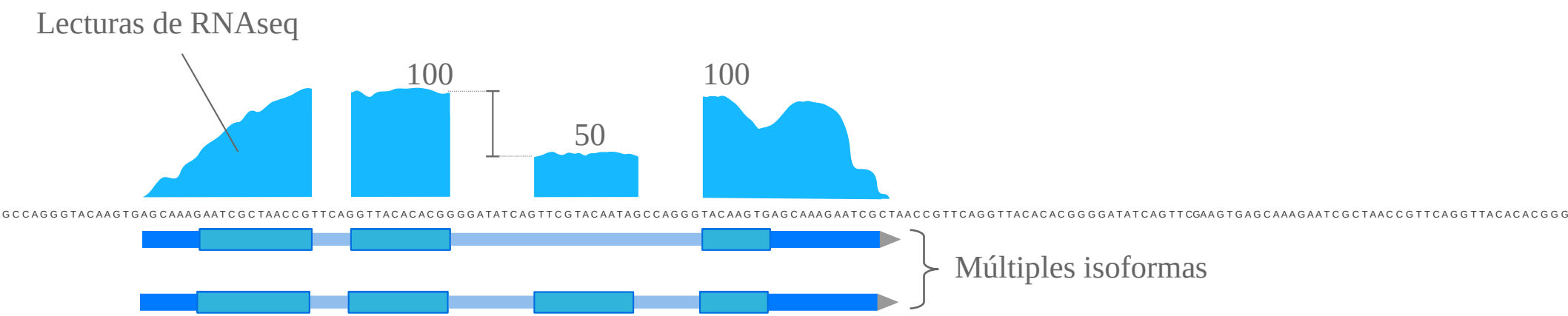
- Identificar regiones codificantes en un genoma de referencia (generar anotaciones)



Datos de RNAseq

Modelos de genes

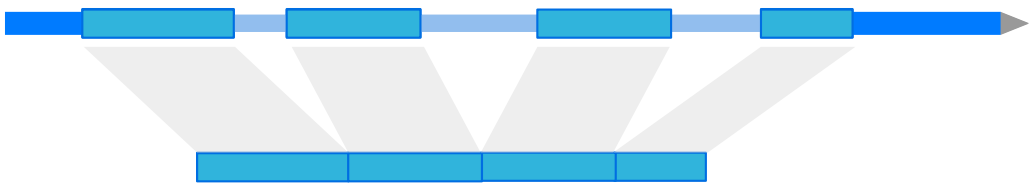
- Identificar regiones codificantes en un genoma de referencia (generar anotaciones)



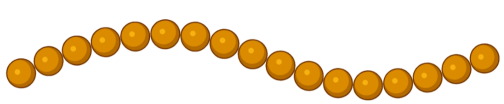
Datos de RNAseq

Modelos de genes

- Obtener secuencias proteicas



	T	C	A	G	
T	TTT Phe	TCT Ser	TAT Tyr	TGT Cys	T
	TTC Phe	TCC Ser	TAC Tyr	TGC Cys	C
	TTA Leu	TCA Ser	TAA stop	TGA stop	A
	TTG Leu	TCG Ser	TAG stop	TGG Trp	G
C	CTT Leu	CCT Pro	CAT His	CGT Arg	T
	CTC Leu	CCC Pro	CAC His	CGC Arg	C
	CTA Leu	CCA Pro	CAA Gln	CGA Arg	A
	CTG Leu	CCG Pro	CAG Gln	CGG Arg	G
A	ATT Ile	ACT Thr	AAT Asn	AGT Ser	T
	ATC Ile	ACC Thr	AAC Asn	AGC Ser	C
	ATA Ile	ACA Thr	AAA Lys	AGA Arg	A
	ATG Met	ACG Thr	AAG Lys	AGG Arg	G
G	GTT Val	GCT Ala	GAT Asp	GGT Gly	T
	GTC Val	GCC Ala	GAC Asp	GGC Gly	C
	GTA Val	GCA Ala	GAA Glu	GGA Gly	A
	GTG Val	GCG Ala	GAG Glu	GGG Gly	G



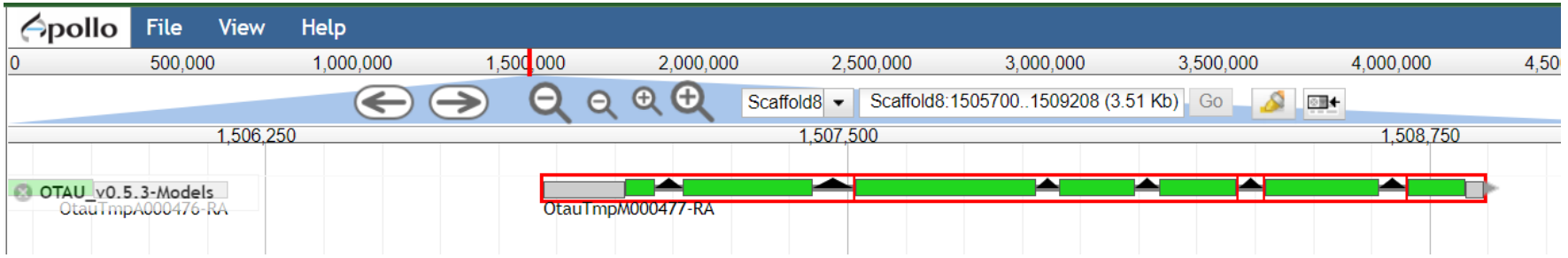
Secuencia de aminoácidos

Datos de RNAseq

Browsers / navegadores

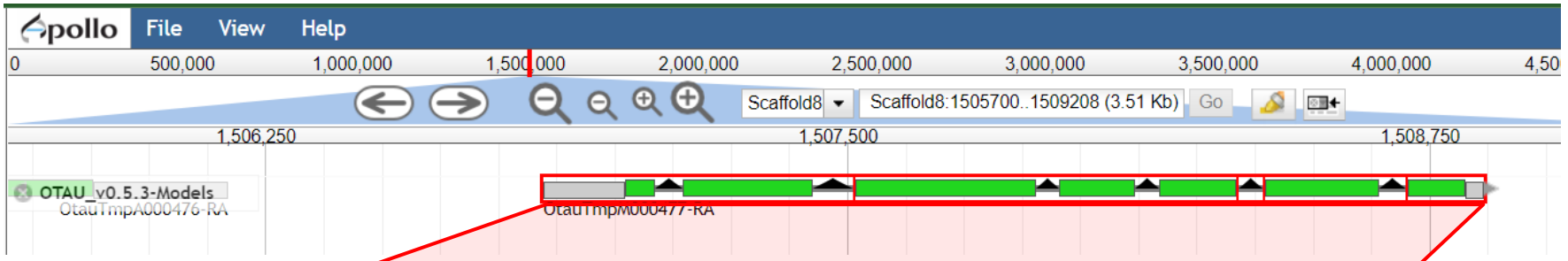
Datos de RNAseq

Browsers / navegadores



Datos de RNAseq

Browsers / navegadores



Primary Data

Name	OtauTmpM000477-RA
Type	mRNA
Position	Scaffold8:1506847..1508866 (+ strand)
Length	2,020 bp

Attributes

Dbxref (7)			
SMART:SM00382	SUPERFAMILY:SSF52540	Gene3D:G3DSA:1.10.8.60	Gene3D:G3DSA:3.40.50.300
ProSiteProfiles:PS50896	ProSitePatterns:PS00674	Pfam:PF00004	

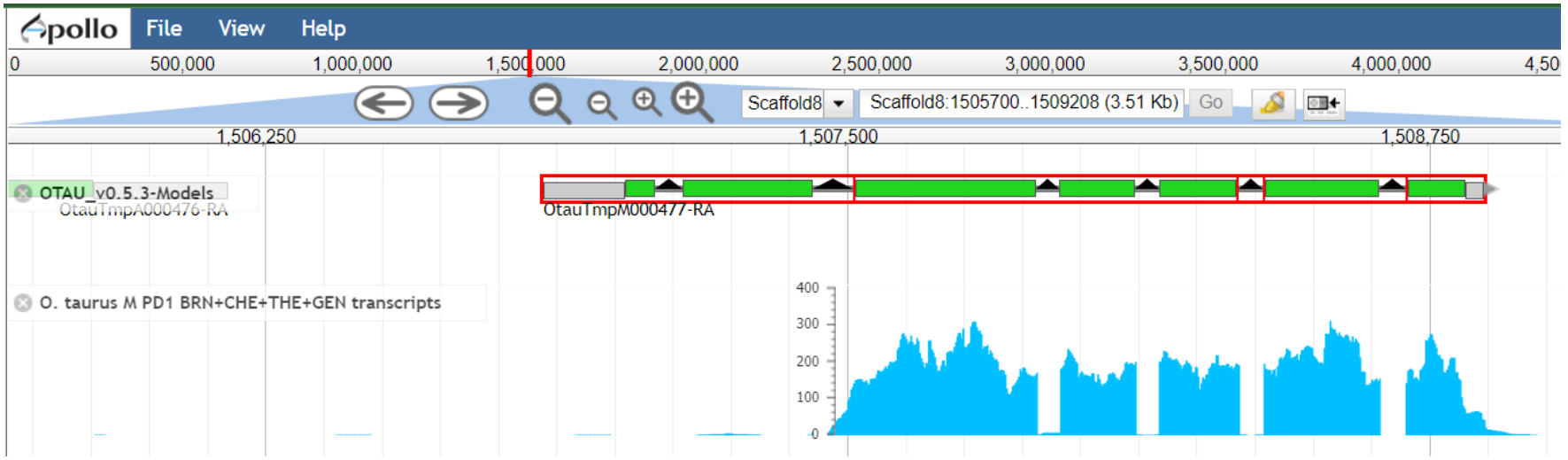
Id	OTAU000477-RA
Seq_id	Scaffold8
Source	I5K

Region sequence

```
>Scaffold8 Scaffold8:1506847..1508866 (+ strand) class=mRNA
length=2020
TCTGTAATTCTTGCAAATTATCTTTAATAATATGAATCATGAATTTTGTGTATTATTGAATA
TAATTACAGCTGCTTAGCAACTTGCTTGTTTTAAAGGGGGATGAAAAGTTCAACCGTCAACA
TAAAACCTCAATTTTACCTTAGTCATTATTTAATAATCGCCACGTTAAACCATGTCGGAGTT
TCTTATTAATAAAAATAGTTCTTCTAATTTCAAAATGAAAACTGTAAAAGAATACGTAAGTT
TCTCAAGTGTATTATTTGTATATCATTATTTTTTTTTTTAAATAATGTTTAGGAAAAAAGTG
AATCAAATGAAAGACGTCGCAATATATTACATTTAATAGCATCGTATTTAAAAGATAGAAAC
TTAATGGATTCATATCAGTCGCTATCAAACGAAGCTCAGTTAACAAATCAATACGAAGTATG
TGATAACATCGATCTTGATATTATATTACAAGAATACCAAAGTTATTACTTTACGAAATTTA
ATAAATATCCAAAGTTGGTTAAAAAGTTAGATGTTGACGTAATAAAACAAGATGAACATCAT
```

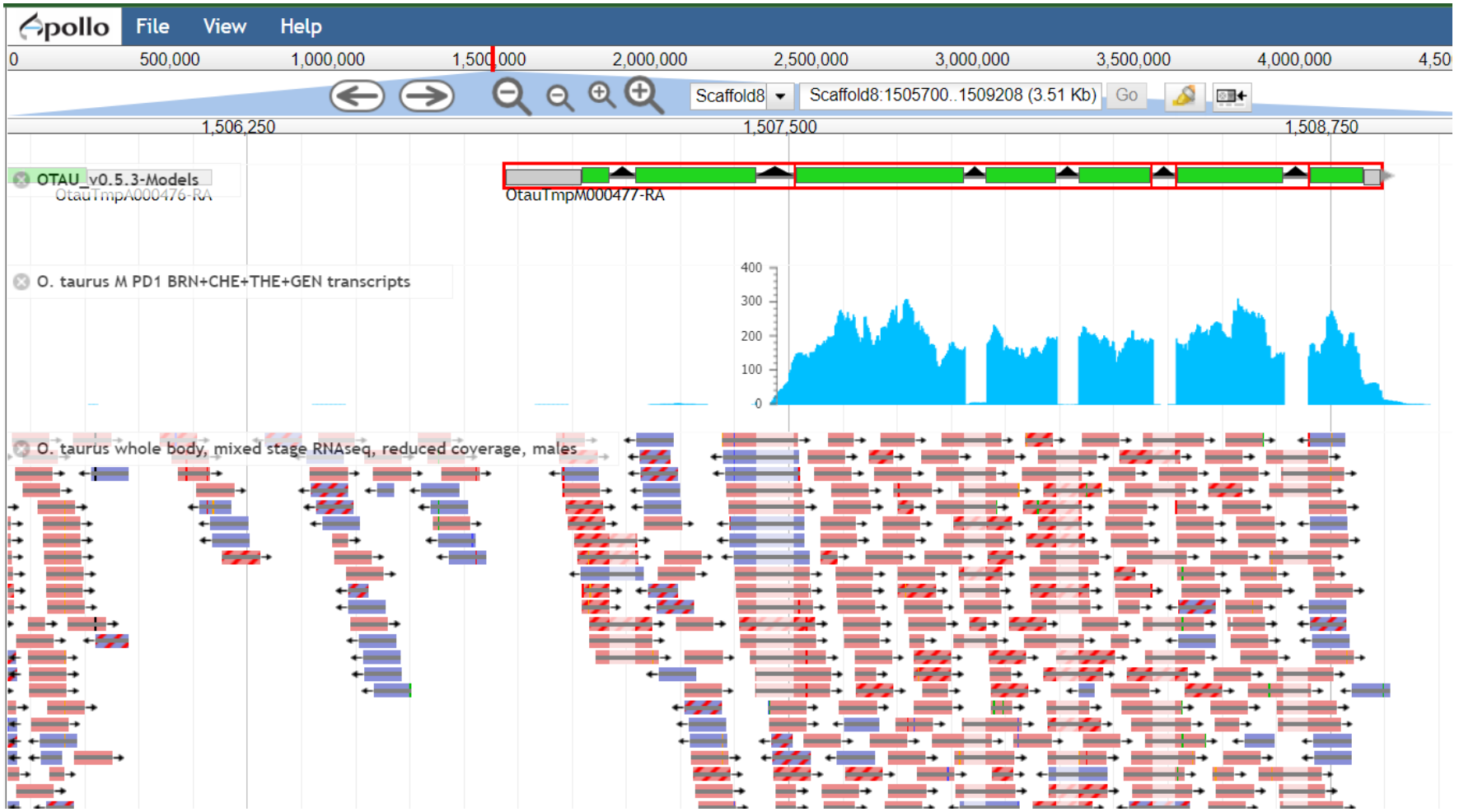
Datos de RNAseq

Browsers / navegadores



Datos de RNAseq

Browsers / navegadores



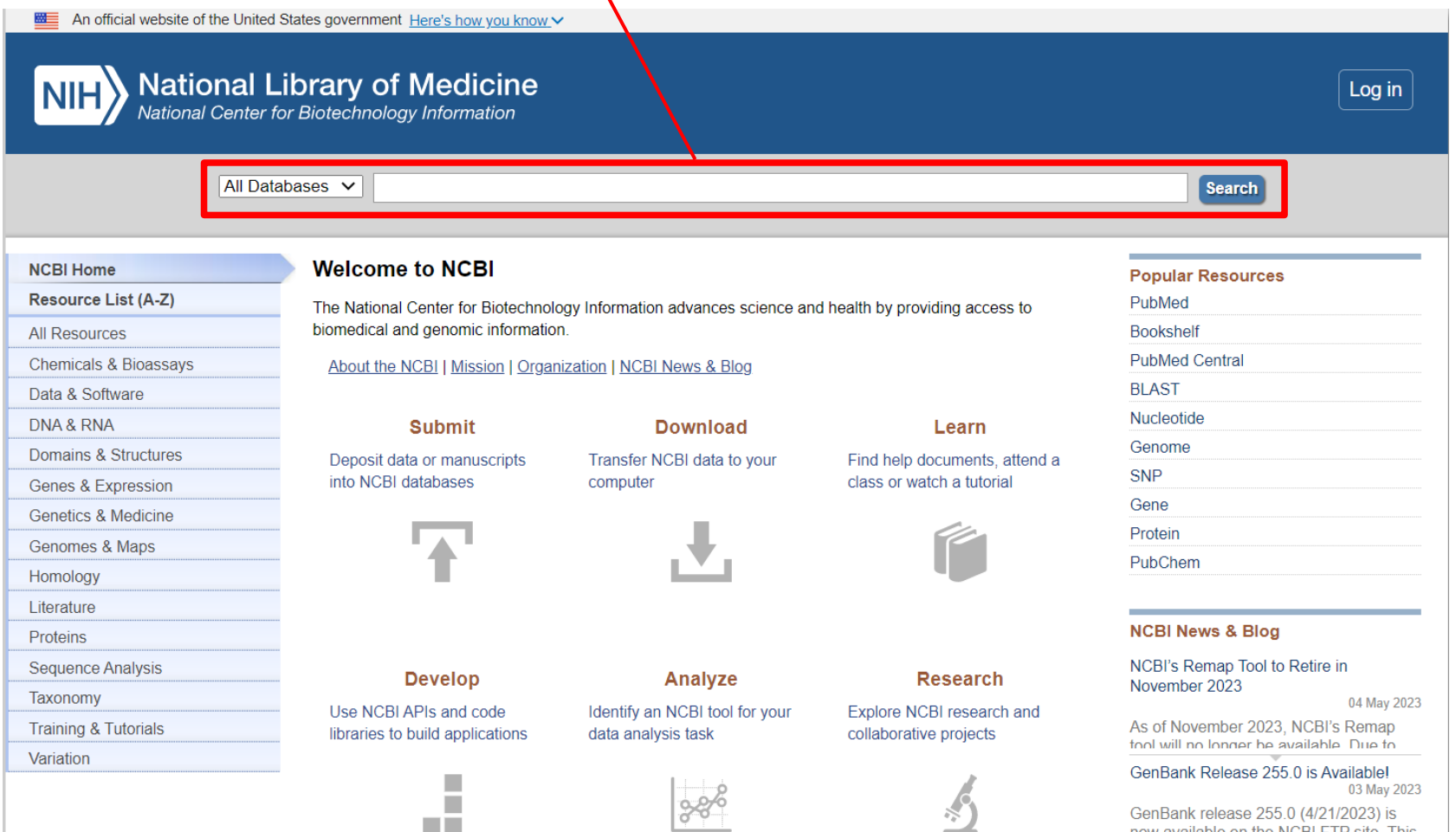
Datos de RNAseq

NCBI National Center for Biotecnology Information

Datos de RNAseq

NCBI

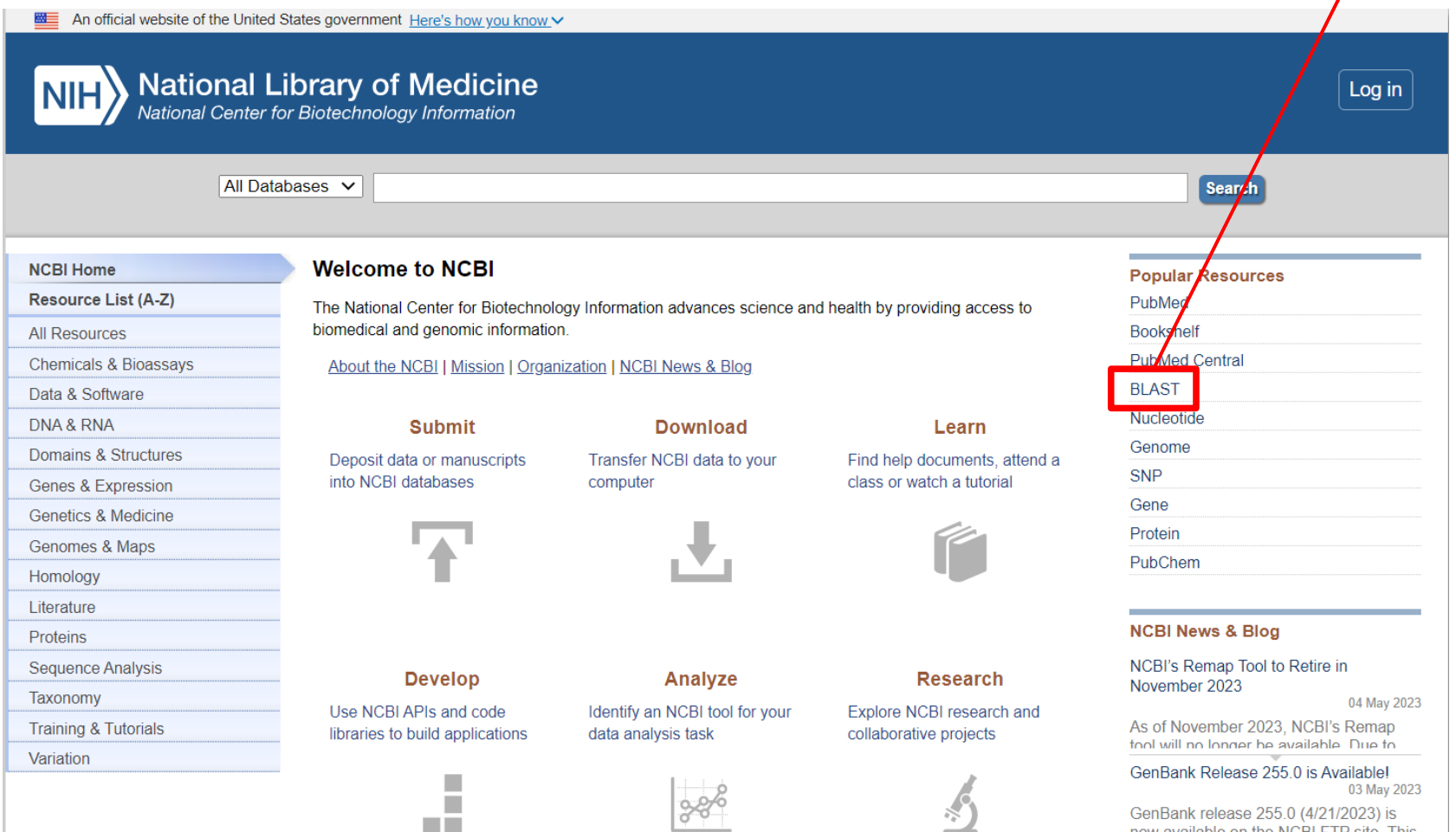
Disponible: genes, genomas, transcriptomas, parpers, etc...



Datos de RNAseq

NCBI

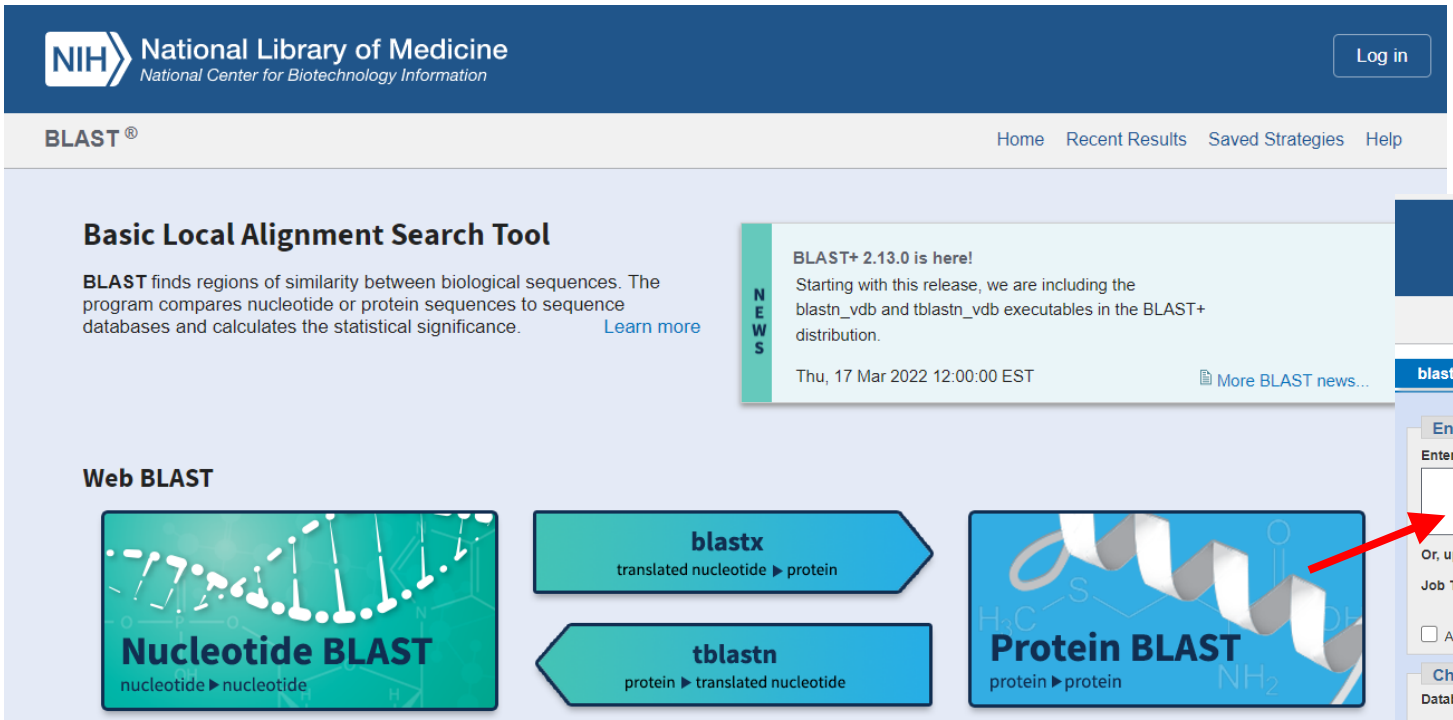
Buscar por secuencia

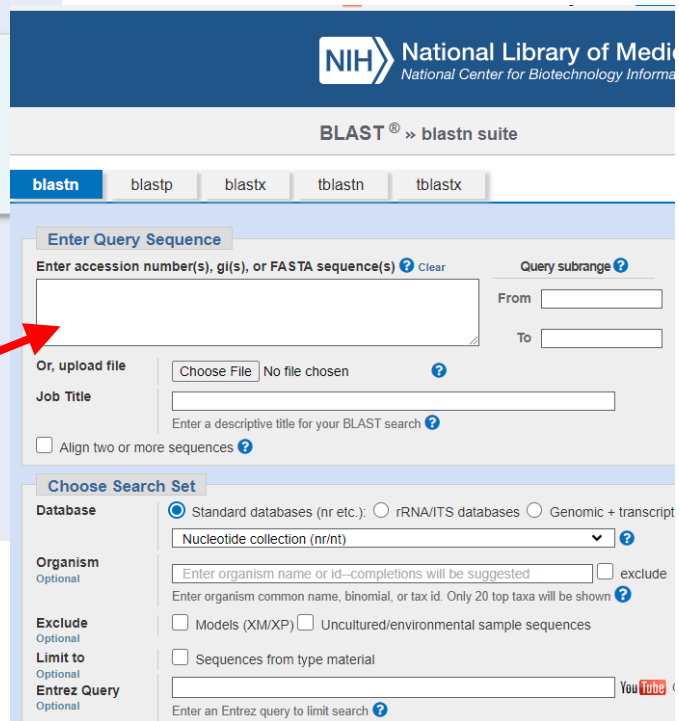


Datos de RNAseq

Blast

Basic Local Alignment Search Tool





Datos de RNAseq

Expresión génica

Datos de RNAseq

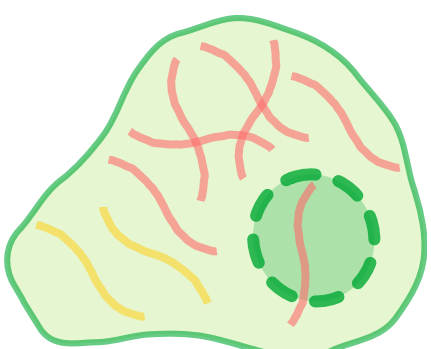
Expresión génica

- Cuantificar la tasa de transcripción

Datos de RNAseq

Expresión génica

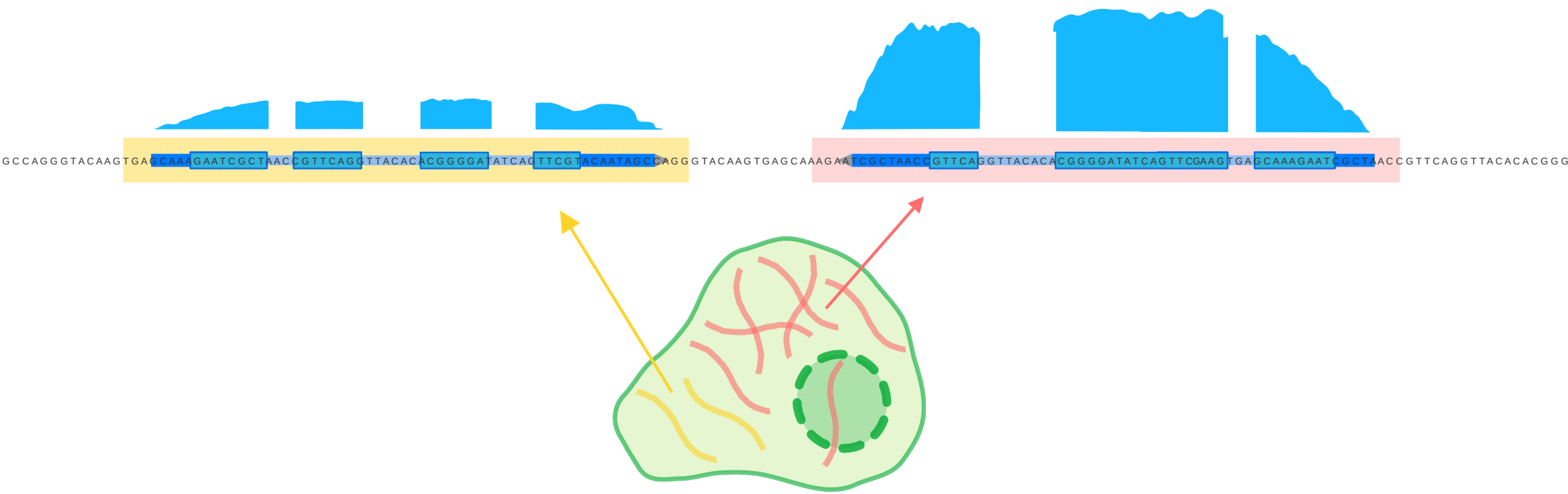
- Cuantificar la tasa de transcripción



Datos de RNAseq

Expresión génica

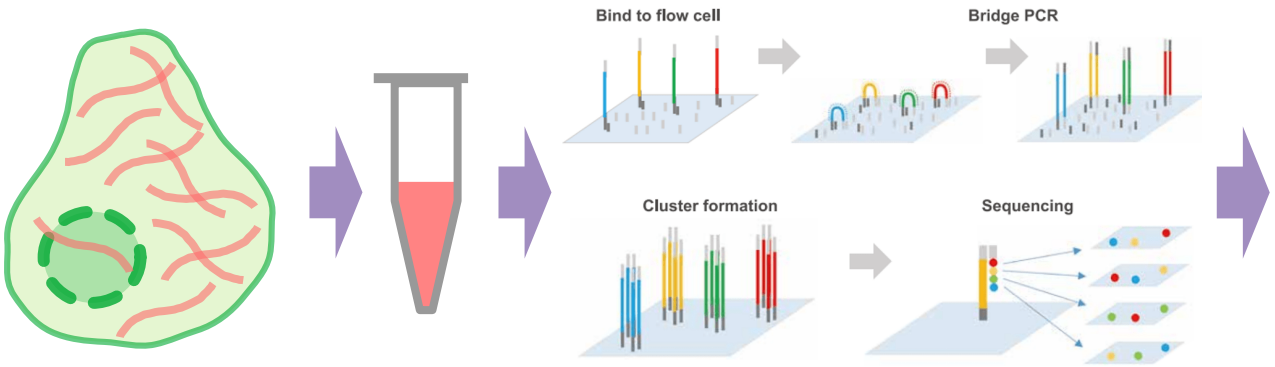
- Cuantificar la tasa de transcripción



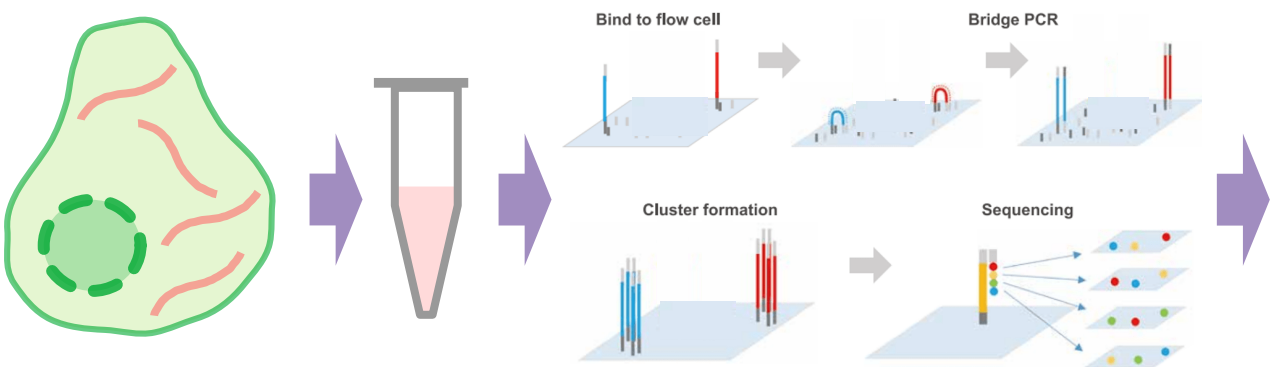
Datos de RNAseq

Expresión génica

- Comparación entre muestras



AGCCAGGGTACAAGTGAGCAAAGAATCGCTAACCGTTCAGGTTACACACGGGGATATCAGTTC
TAGCCAGGGTACAAGTGAGCAAAGAATCGCTAACCGTTCAGGTTACACACGGGGATATCAGTTCGTACAA
CTAGCCAGGGTACAAGTGAGCAAAGAATCGCTAACCGTTCAGGTTACACACGGGGATATCAGTT
ACTAGCCAGGGTACAAGTGAGCAAAGAATCGCTAACCGTTCAGGTTACACACGGGGATATCAGTTCGTA
CTAGCCAGGGTACAAGTGAGCAAAGAATCGCTAACCGTTCAGGTTACACACGGGGATATCAGTTCG
TAGCCAGGGTACAAGTGAGCAAAGAATCGCTAACCGTTCAGGTTACACACGGGGATATCAGT
CTAGCCAGGGTACAAGTGAGCAAAGAATCGCTAACCGTTCAGGTTACACACGGGGATATCAGTTCGTACAA
CAGGGTACAAGTGAGCAAAGAATCGCTAACCGTTCAGGTTACACACGGGGATATCAGTTCG
AAGGAAGAAGACTAGCCAGGGTACAAGTGAGCAAAGAATCGCTAACCGTTCAGGTTACACACGG



TAGCCAGGGTACAAGTGAGCAAAGAATCGCTAACCGTTCAGGTTACACACGGGGATATCAGTTCGTACAA

ACTAGCCAGGGTACAAGTGAGCAAAGAATCGCTAACCGTTCAGGTTACACACGGGGATATCAGTTCGTA

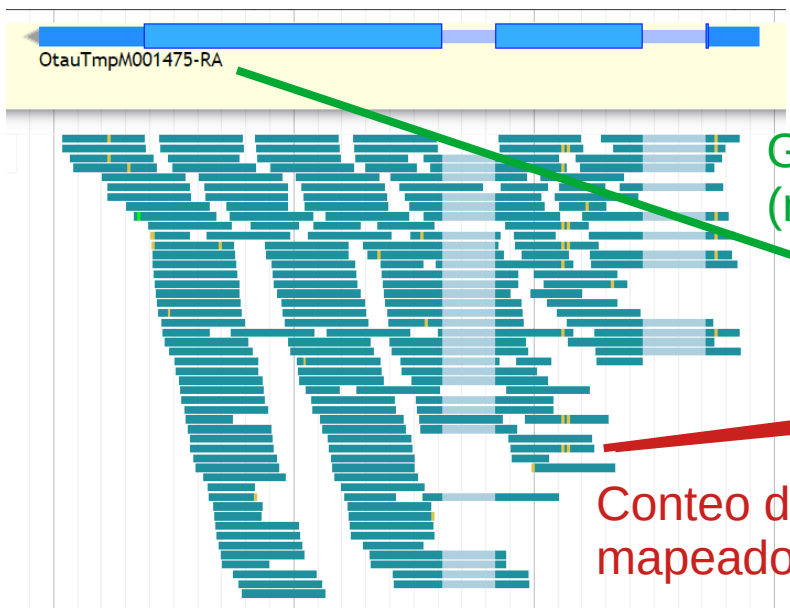
AAGGAAGAAGACTAGCCAGGGTACAAGTGAGCAAAGAATCGCTAACCGTTCAGGTTACACACGG

Datos de RNAseq

Expresión génica

Muestras

...



Gen (modelo)

Conteo de reads mapeados

	gene	ctrlRNAi_M.BRN_S19	ctrlRNAi_M.BRN_S20	ctrlRNAi_M
1	OTAU000001	166	155	
2	OTAU000002	955	975	
3	OTAU000003	2318	2228	
4	OTAU000004	2	1	
5	OTAU000005	29	26	
6	OTAU000006	97	78	
7	OTAU000007	0	0	
8	OTAU000008	1841	1902	
9	OTAU000009	730	549	
10	OTAU000010	490	535	
11	OTAU000011	1	4	
12	OTAU000012	21	14	
13	OTAU000013	25	29	
14	OTAU000014	39	35	
15	OTAU000015	3625	3178	
16	OTAU000016	1130	1115	
17	OTAU000017	638	607	
18	OTAU000018	304	242	
19	OTAU000019	15	10	
20	OTAU000020	0	0	

Datos de RNAseq

Matriz de diseño

Factores

Reads totales

IDs de las muestras

	sampleID	Tissue	Treatment	Sex	genome_unique	X_unique
1	ctrlRNAi_M.BRN_S19	BRN	ctr	M	10748437	8861358
2	ctrlRNAi_M.BRN_S20	BRN	ctr	M	10129825	8339020
3	ctrlRNAi_M.BRN_S21	BRN	ctr	M	10571624	8676801
4	ctrlRNAi_M.BRN_S22	BRN	ctr	M	10899420	9015026
5	ctrlRNAi_M.BRN_S23	BRN	ctr	M	10275664	8460658
6	ctrlRNAi_M.BRN_S24	BRN	ctr	M	10281103	8359537
7	ctrlRNAi_M.THE_S43	THE	ctr	M	10175292	9123613
8	ctrlRNAi_M.THE_S44	THE	ctr	M	4640668	4110648
9	ctrlRNAi_M.THE_S45	THE	ctr	M	10431983	9411583
10	ctrlRNAi_M.THE_S46	THE	ctr	M	10239974	9337273
11	ctrlRNAi_M.THE_S47	THE	ctr	M	10954425	9844328
12	ctrlRNAi_M.THE_S48	THE	ctr	M	9431346	8379494
13	ctrlRNAi_M.CHE_S67	CHE	ctr	M	10063065	8960920
14	ctrlRNAi_M.CHE_S68	CHE	ctr	M	10213914	9003514
15	ctrlRNAi_M.CHE_S69	CHE	ctr	M	14606729	13274821
16	ctrlRNAi_M.CHE_S70	CHE	ctr	M	10461943	9431847
17	ctrlRNAi_M.CHE_S71	CHE	ctr	M	15317489	13661185
18	ctrlRNAi_M.CHE_S72	CHE	ctr	M	13292199	11776005
19	ctrlRNAi_M.GEN_S91	GEN	ctr	M	17698786	15493944
20	ctrlRNAi_M.GEN_S92	GEN	ctr	M	9542977	8330501

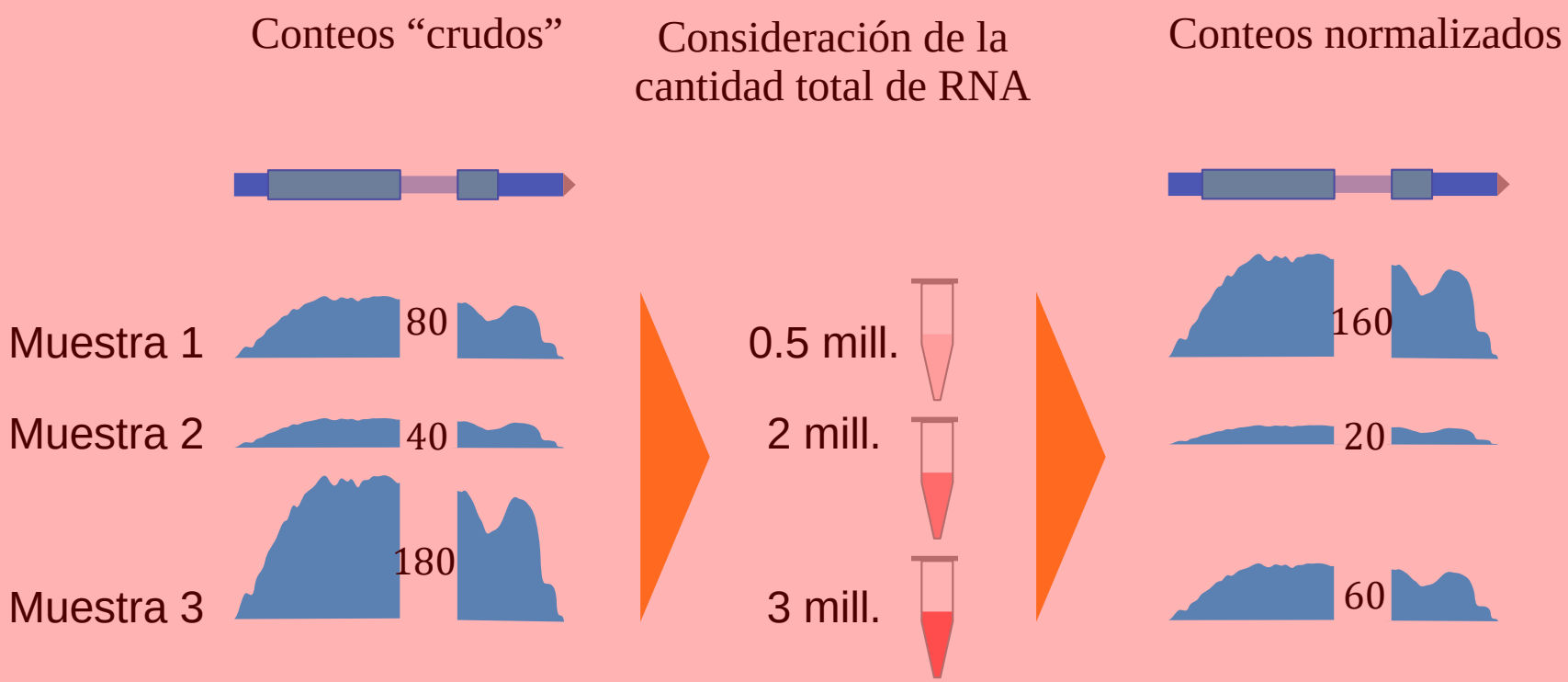
...

	gene	ctrlRNAi_M.BRN_S19	ctrlRNAi_M.BRN_S20	ctrlRNAi_M
1	OTAU000001	166	155	
2	OTAU000002	955	975	
3	OTAU000003	2318	2228	
4	OTAU000004	2	1	
5	OTAU000005	29	26	
6	OTAU000006	97	78	
7	OTAU000007	0	0	
8	OTAU000008	1841	1902	
9	OTAU000009	730	549	
10	OTAU000010	490	535	
11	OTAU000011	1	4	
12	OTAU000012	21	14	
13	OTAU000013	25	29	
14	OTAU000014	39	35	
15	OTAU000015	3625	3178	
16	OTAU000016	1130	1115	
17	OTAU000017	638	607	
18	OTAU000018	304	242	
19	OTAU000019	15	10	
20	OTAU000020	0	0	

Tabla de conteos

Datos de RNAseq

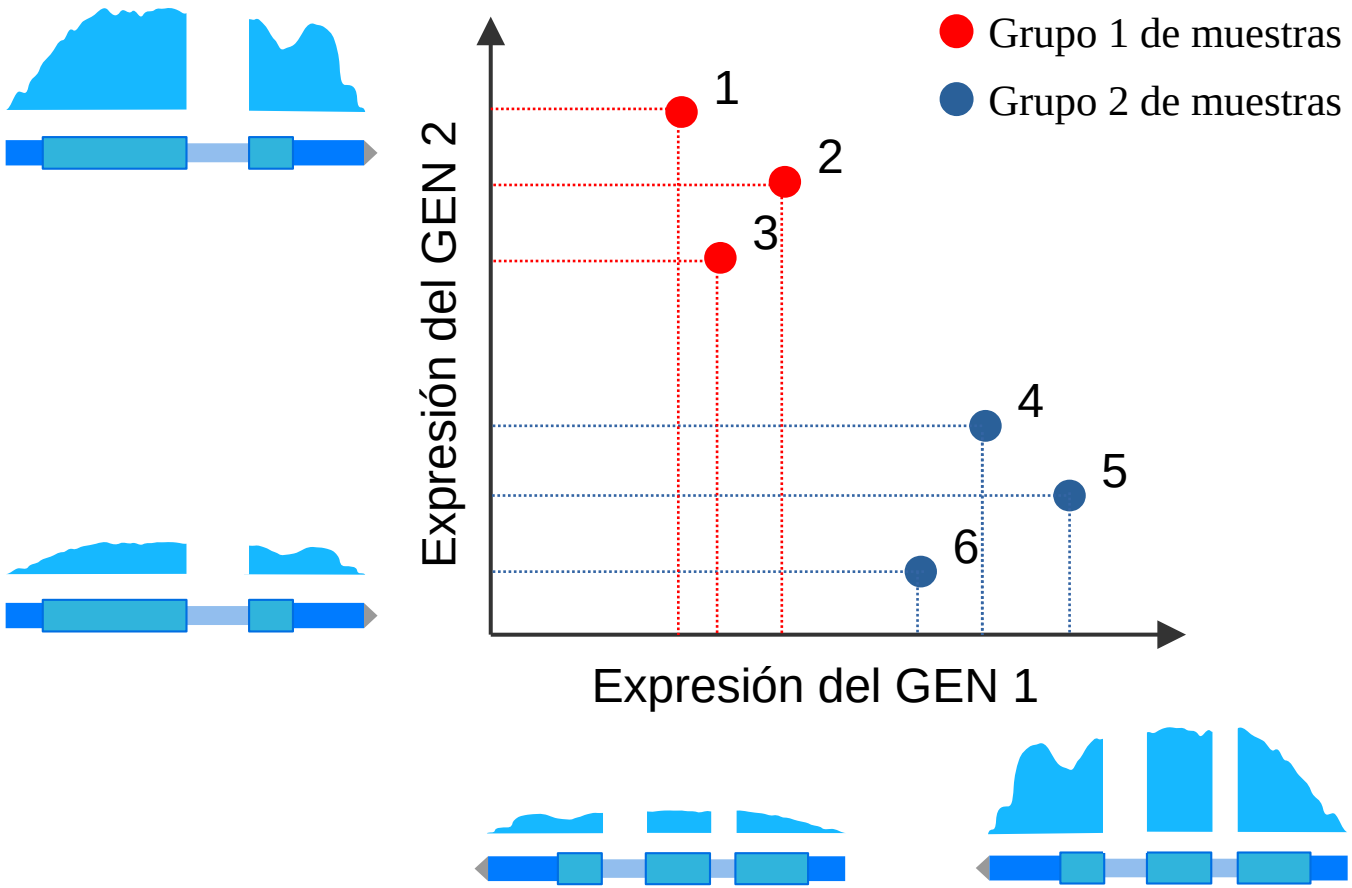
Normalización



Análisis de datos de RNAseq

Expresión génica

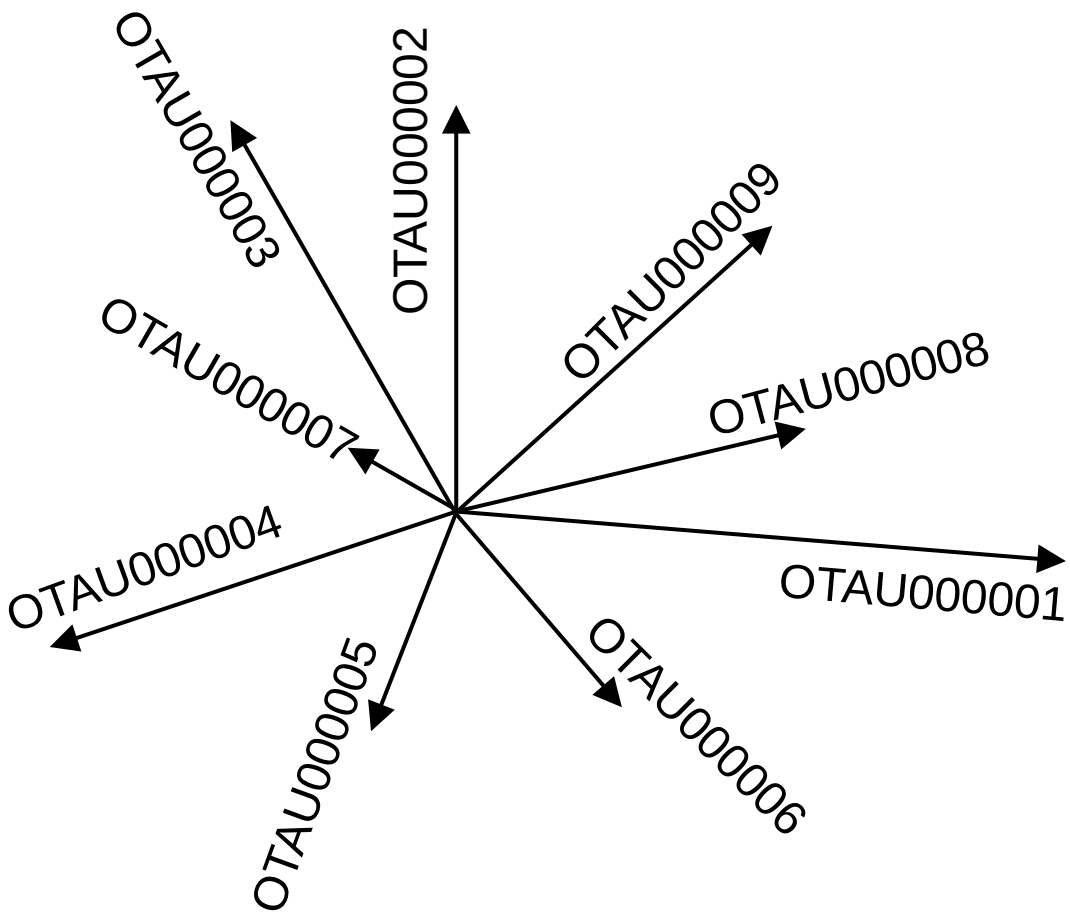
Posición/coordenadas = Patrón de expresión génica



Análisis de datos de RNAseq

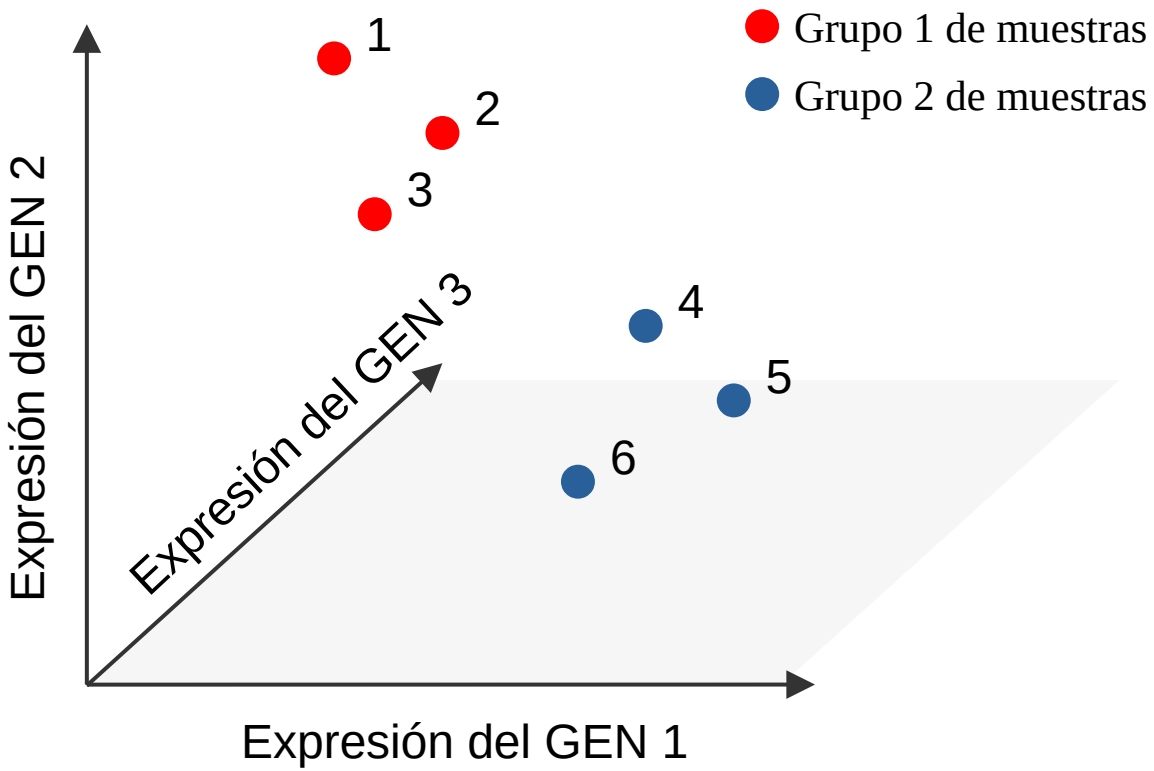
Expresión génica

	gene	ctrlRNAi_M.BRN_S19	ctrlRNAi_M.BRN_S20	ctrlRNAi_M
1	OTAU000001	166	155	
2	OTAU000002	955	975	
3	OTAU000003	2318	2228	
4	OTAU000004	2	1	
5	OTAU000005	29	26	
6	OTAU000006	97	78	
7	OTAU000007	0	0	
8	OTAU000008	1841	1902	
9	OTAU000009	730	549	
10	OTAU000010	490	535	
11	OTAU000011	1	4	
12	OTAU000012	21	14	
13	OTAU000013	25	29	
14	OTAU000014	39	35	
15	OTAU000015	3625	3178	
16	OTAU000016	1130	1115	
17	OTAU000017	638	607	
18	OTAU000018	304	242	
19	OTAU000019	15	10	
20	OTAU000020	0	0	



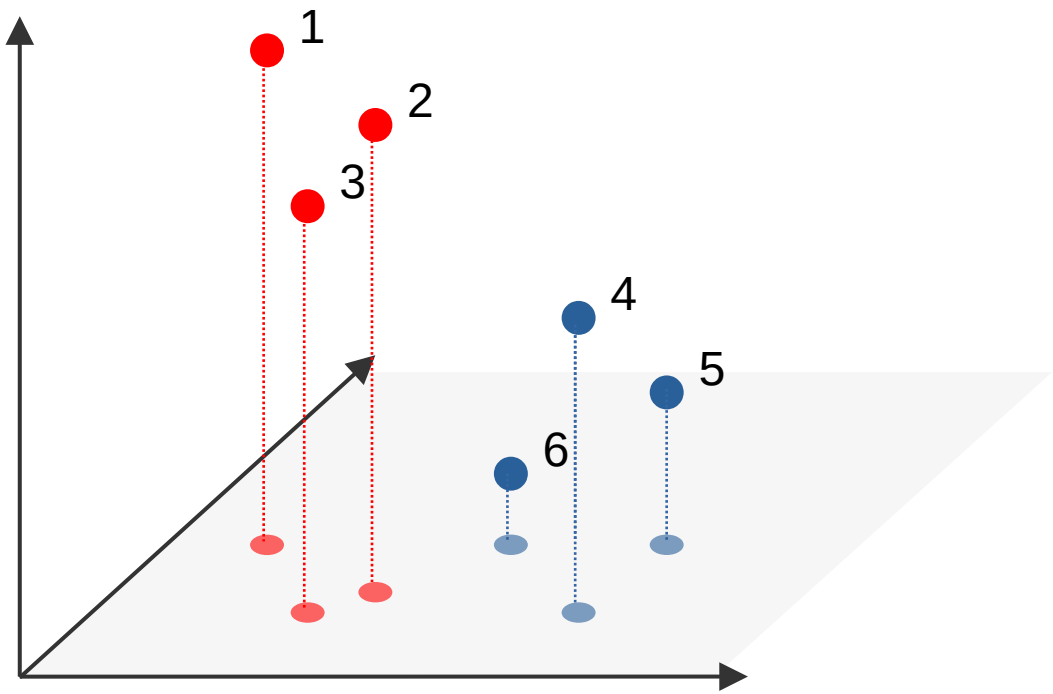
Análisis de datos de RNAseq

Expresión génica



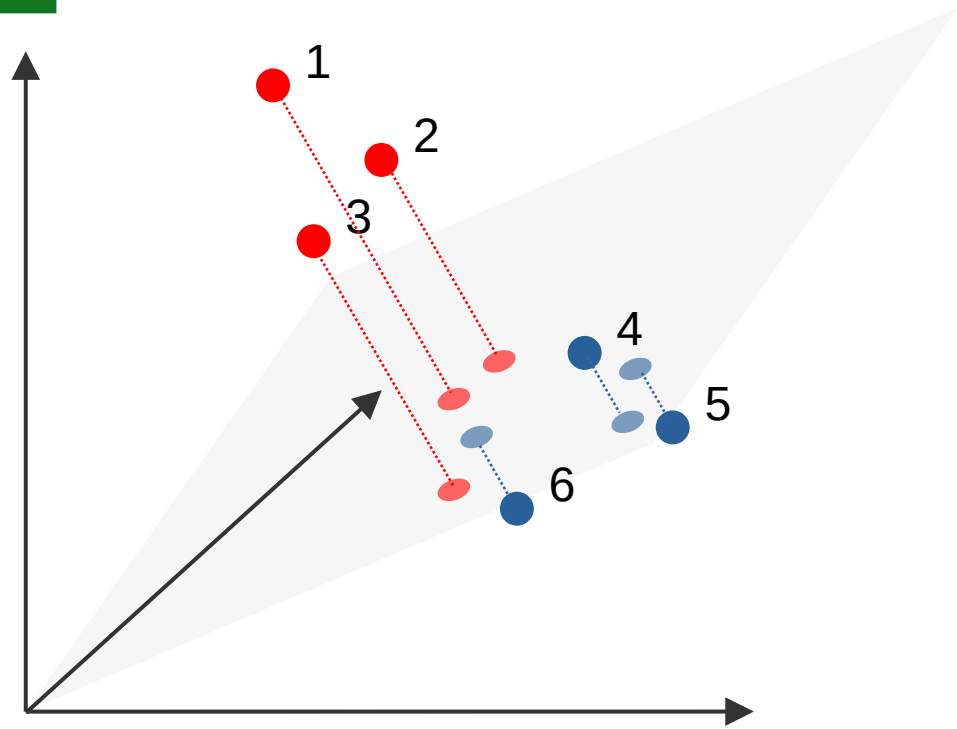
Análisis de datos de RNAseq

Expresión génica



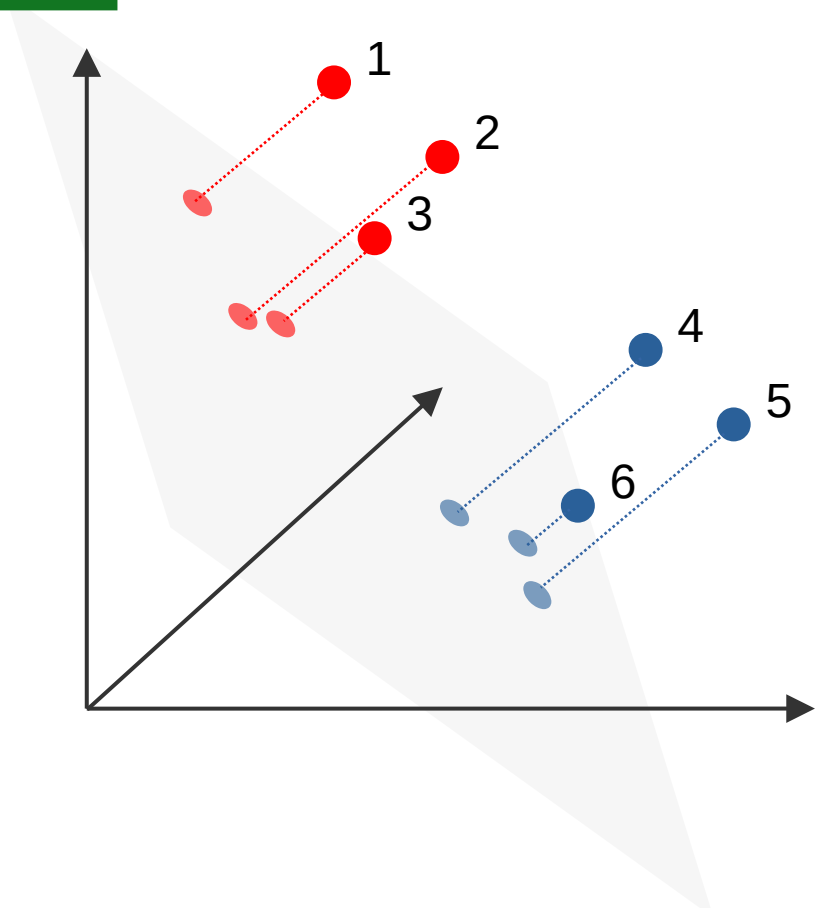
Análisis de datos de RNAseq

Expresión génica



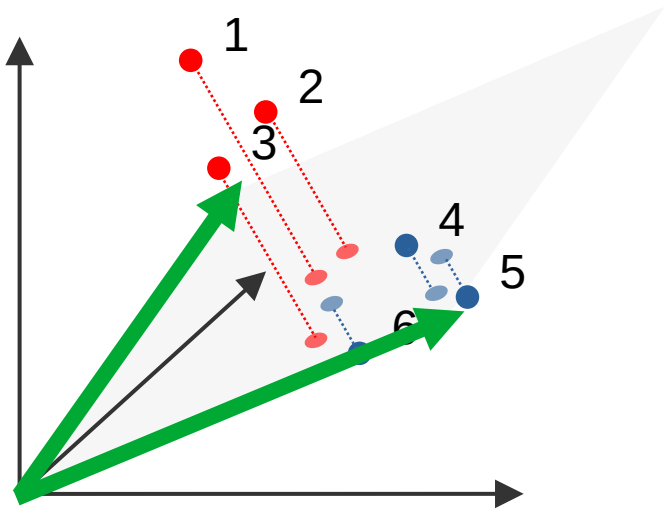
Análisis de datos de RNAseq

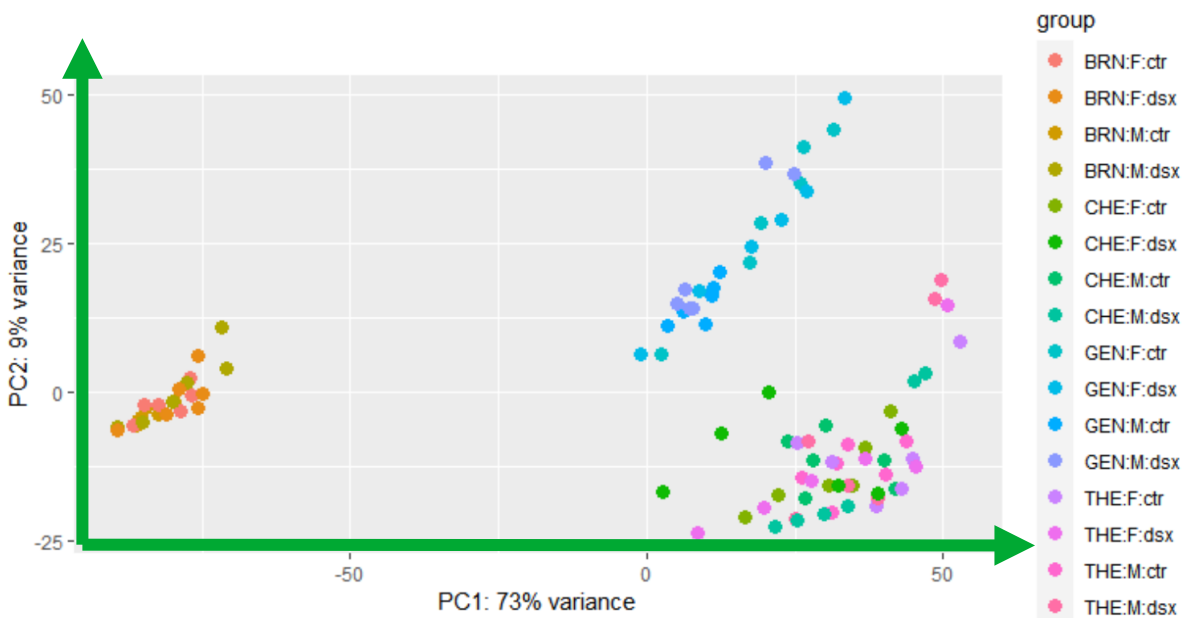
Expresión génica



Análisis de datos de RNAseq

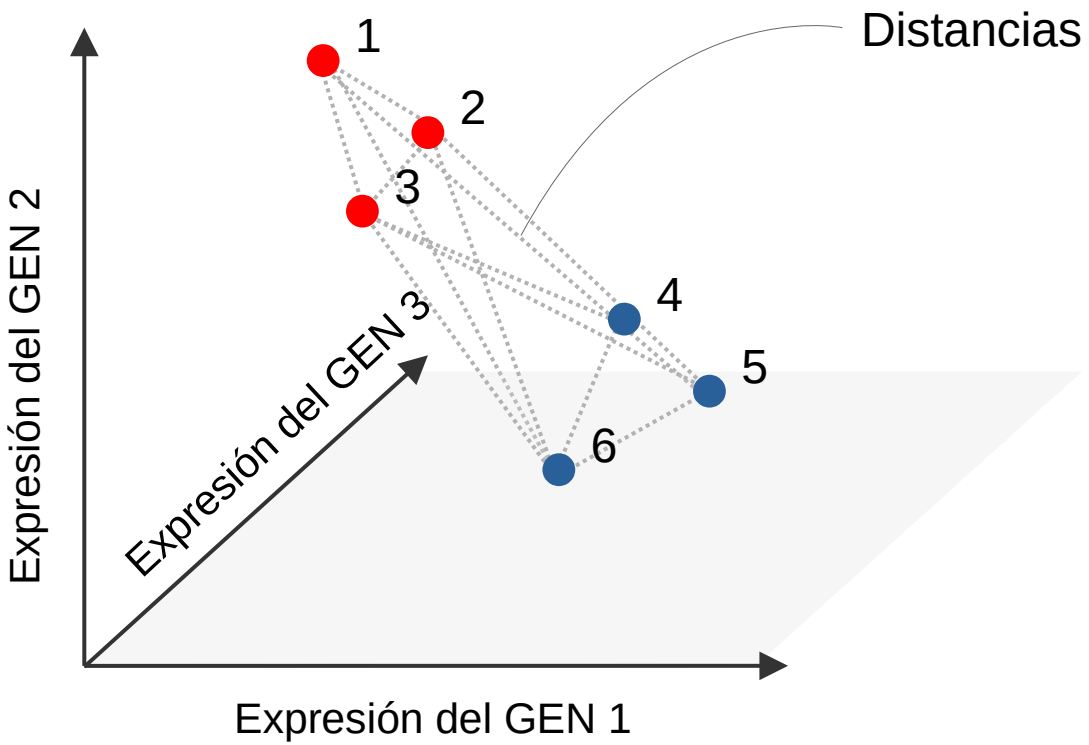
Comp. principales (PCA)

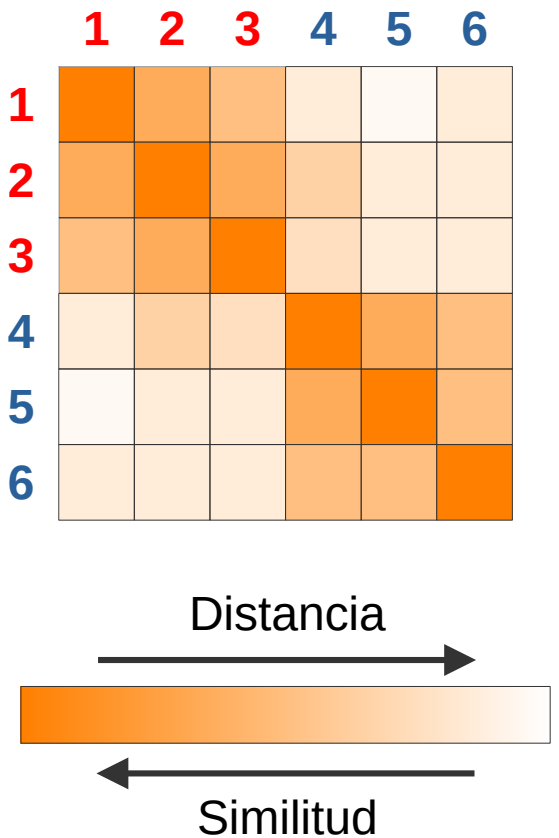




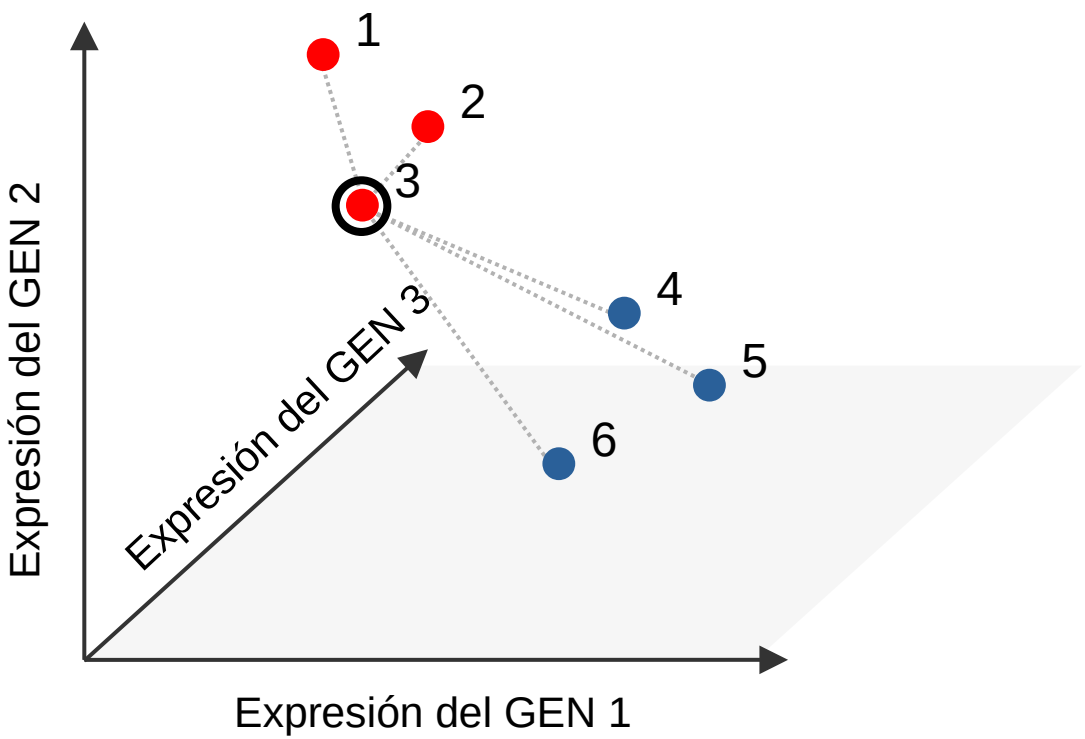
Análisis e interpretación de datos de RNAseq

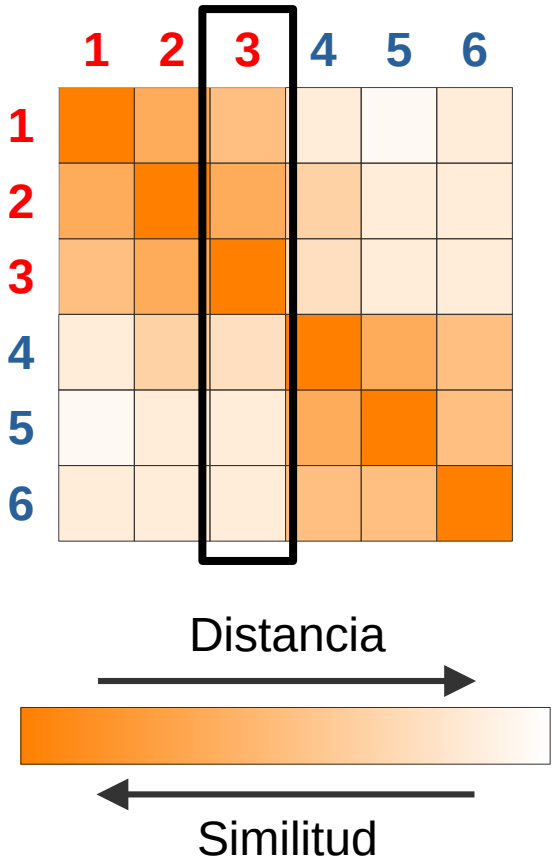
Heatmaps



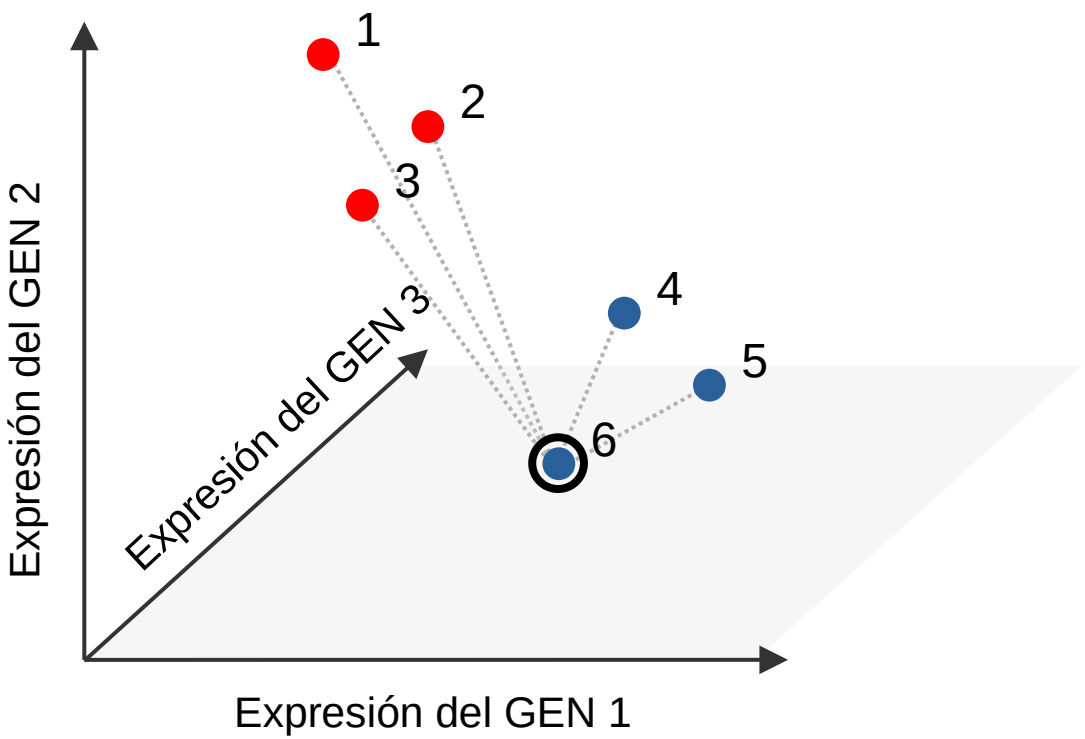


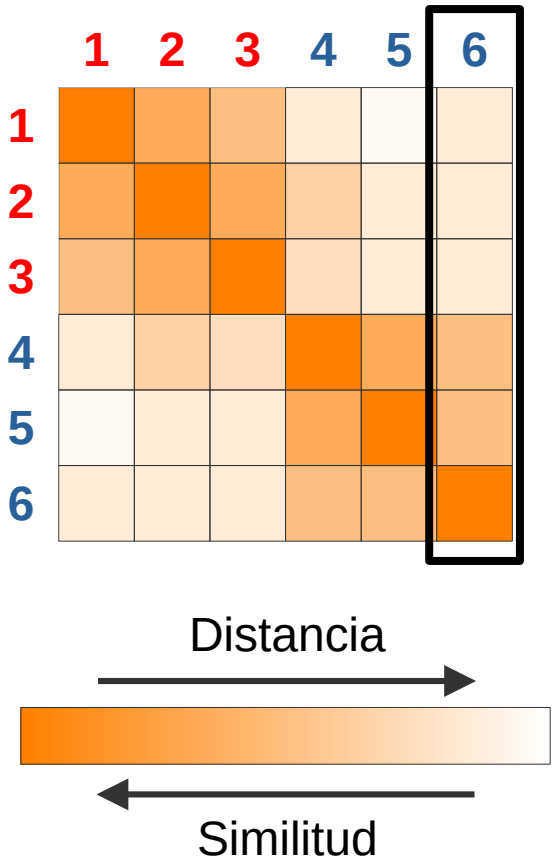
Heatmaps


1
2
3
4
5
6
Expresión del GEN 2
Expresión del GEN 3
Expresión del GEN 1

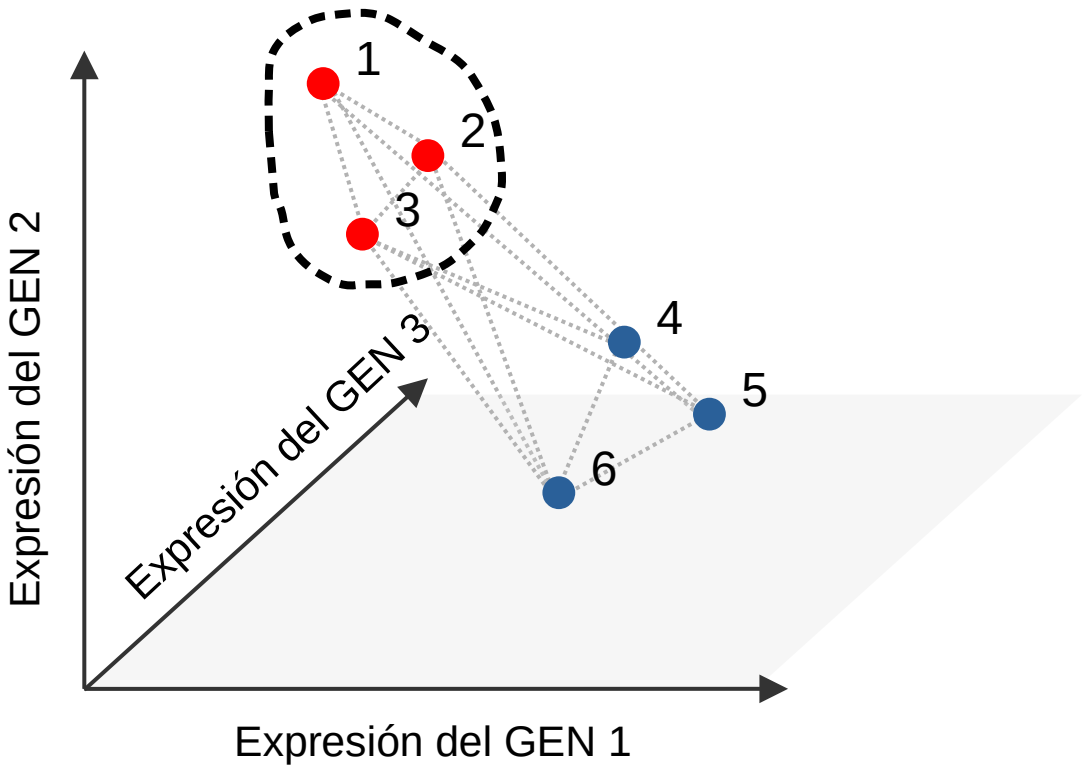

1 2 3 4 5 6
1
2
3
4
5
6
Distancia
Similitud

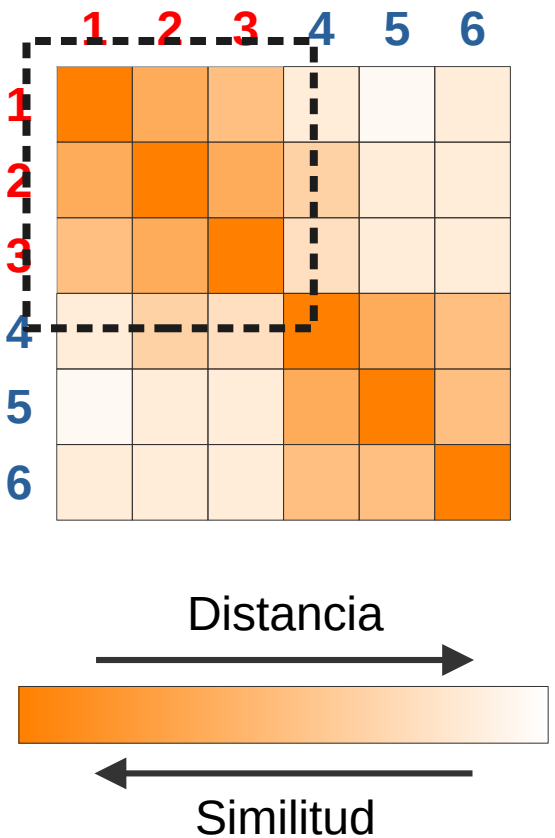
Heatmaps


1
2
3
4
5
6
Expresión del GEN 2
Expresión del GEN 3
Expresión del GEN 1


1 2 3 4 5 6
1
2
3
4
5
6
Distancia
Similitud

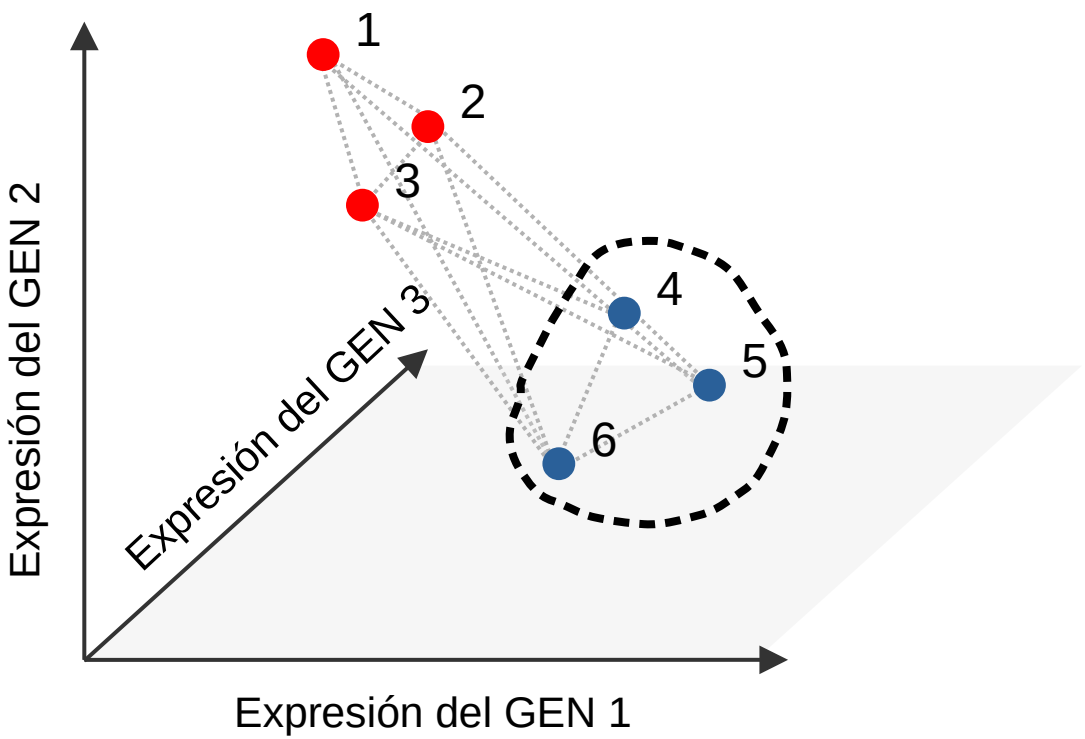
Heatmaps

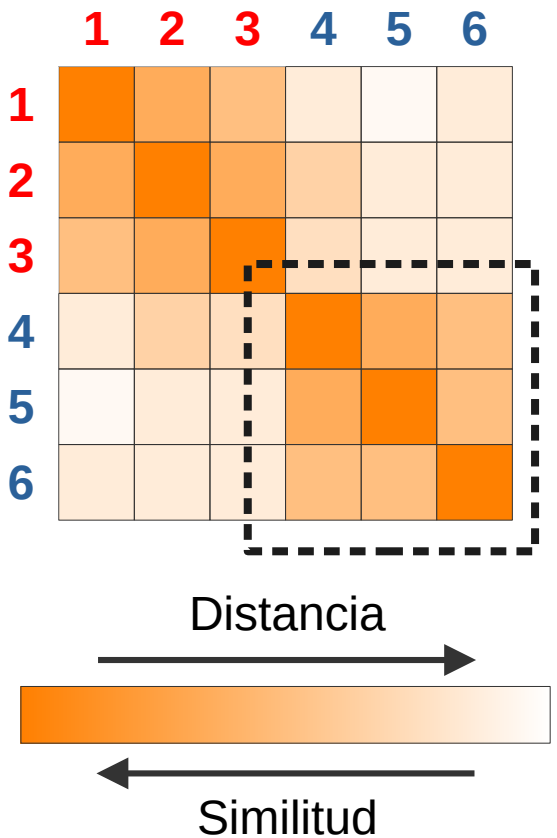

1
2
3
4
5
6
Expresión del GEN 2
Expresión del GEN 3
Expresión del GEN 1


1 2 3 4 5 6
1
2
3
4
5
6
Distancia
Similitud

Análisis e interpretación de datos de RNAseq

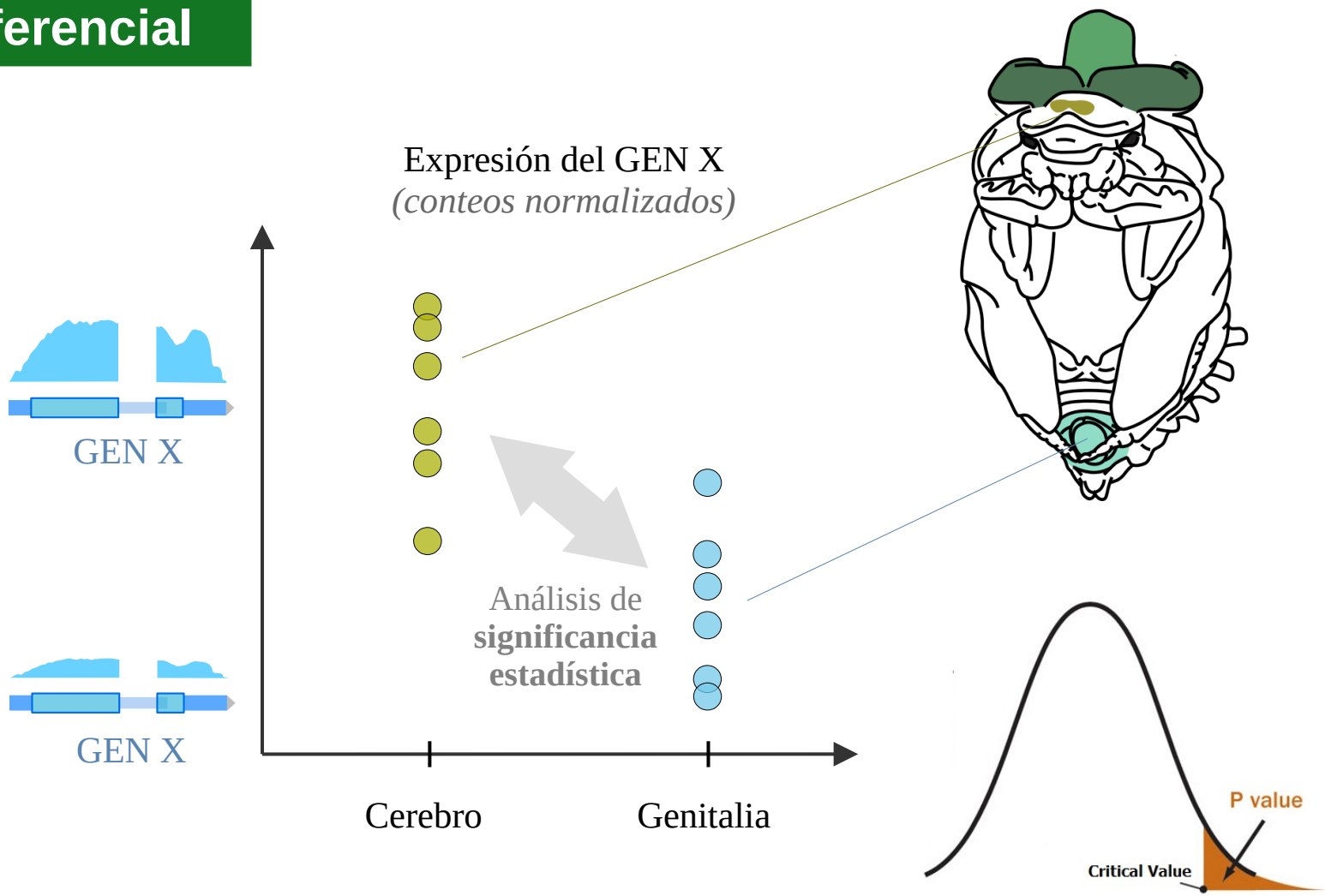
Heatmaps





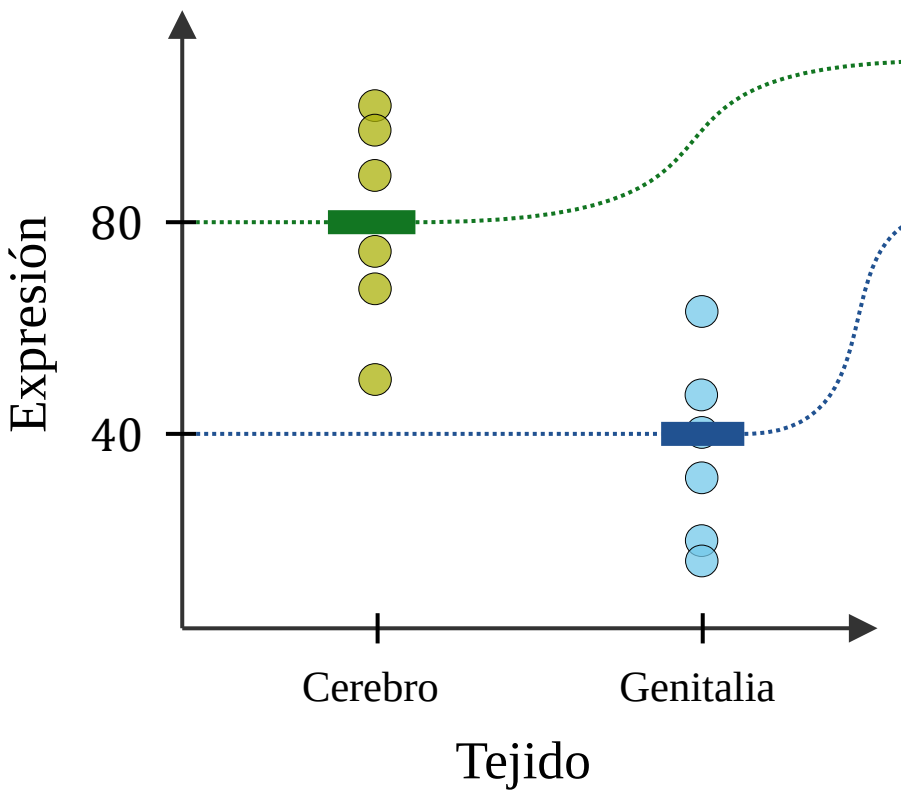
Análisis e interpretación de datos de RNAseq

Expresión diferencial



Análisis e interpretación de datos de RNAseq

Expresión diferencial



Relación de cambio
(Fold Change)

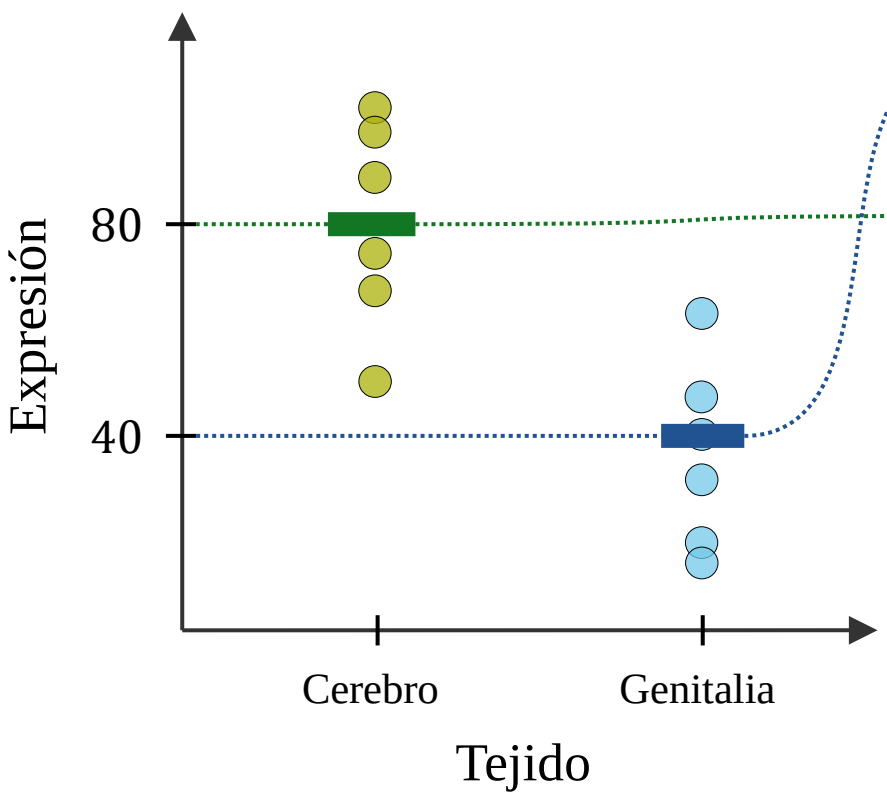
$$\frac{E_1}{E_2} = C_{12} = 2$$

$$Log_2(C_{12}) = 1$$

Log (2) Fold Change
(LFC)

Análisis e interpretación de datos de RNAseq

Expresión diferencial



Relación de cambio
(Fold Change)

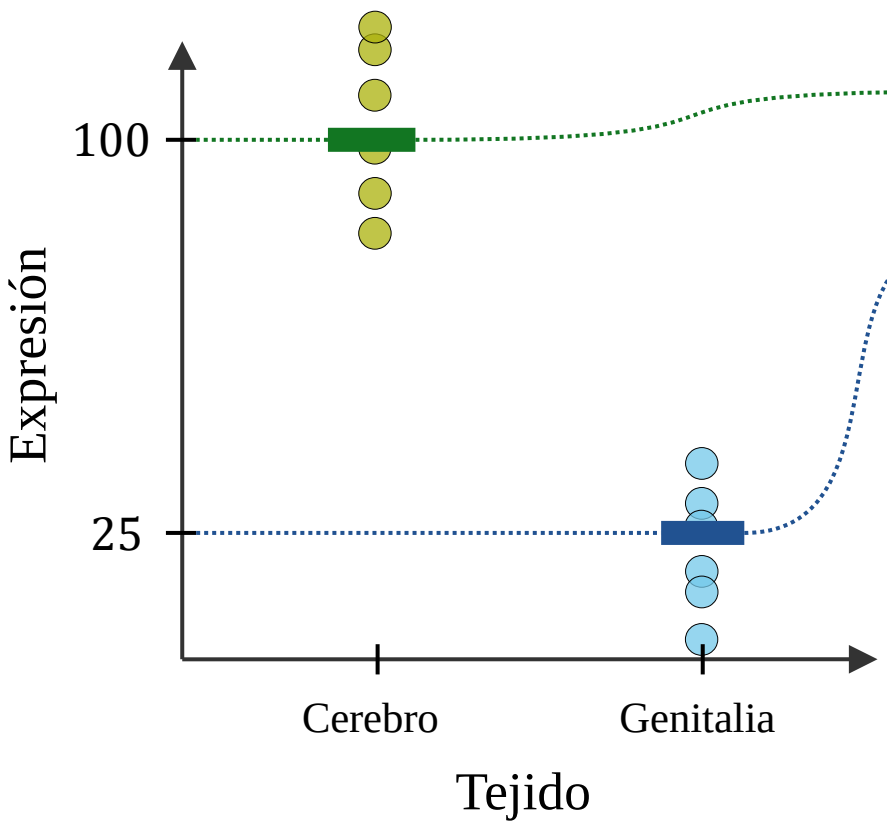
$$\frac{E_2}{E_1} = C_{21} = \boxed{0.5}$$

$$Log_2(C_{21}) = -1$$

Log (2) Fold Change
(LFC)

Análisis e interpretación de datos de RNAseq

Expresión diferencial



Relación de cambio
(Fold Change)

$$\frac{E_1}{E_2} = C_{12} = 4$$

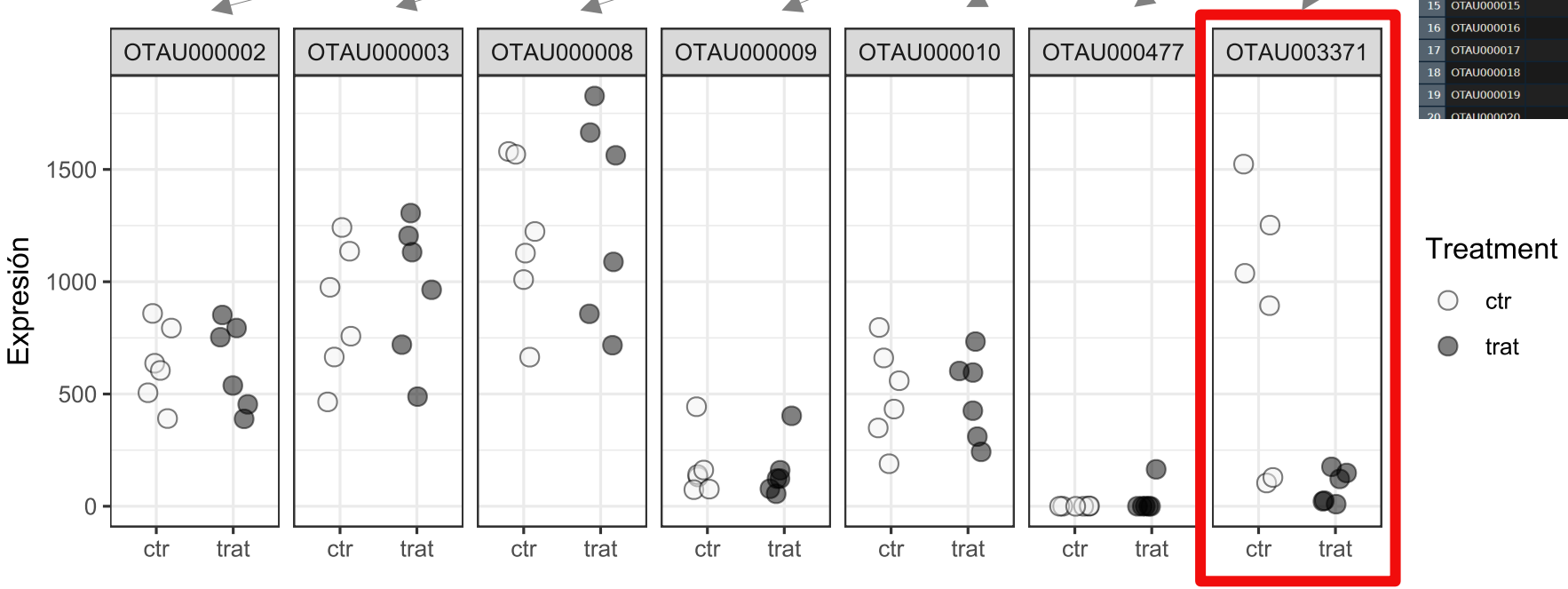
$$Log_2(C_{12}) = 2$$

Log (2) Fold Change
(LFC)

Análisis e interpretación de datos de RNAseq

Expresión diferencial

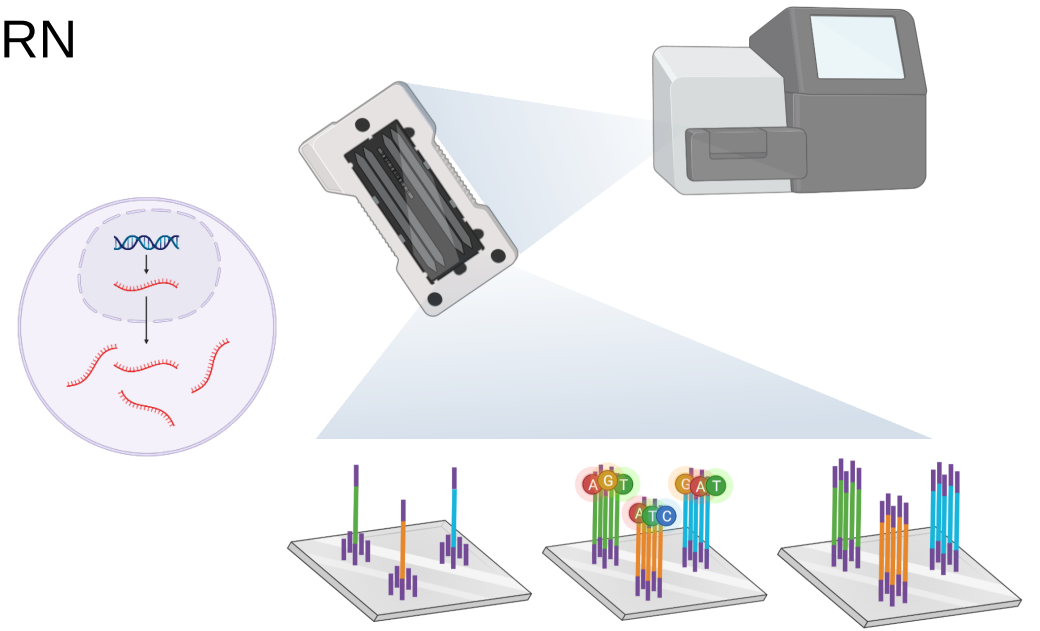
	gene	ctrlRNAi_M.BRN_S19	ctrlRNAi_M.BRN_S20	ctrlRNAi_M
1	OTAU000001	166	155	
2	OTAU000002	955	975	
3	OTAU000003	2318	2228	
4	OTAU000004	2	1	
5	OTAU000005	29	26	
6	OTAU000006	97	78	
7	OTAU000007	0	0	
8	OTAU000008	1841	1902	
9	OTAU000009	730	549	
10	OTAU000010	490	535	
11	OTAU000011	1	4	
12	OTAU000012	21	14	
13	OTAU000013	25	29	
14	OTAU000014	39	35	
15	OTAU000015	3625	3178	
16	OTAU000016	1130	1115	
17	OTAU000017	638	607	
18	OTAU000018	304	242	
19	OTAU000019	15	10	
20	OTAU000020	0	0	



Resumen hasta acá

Resumen hasta acá

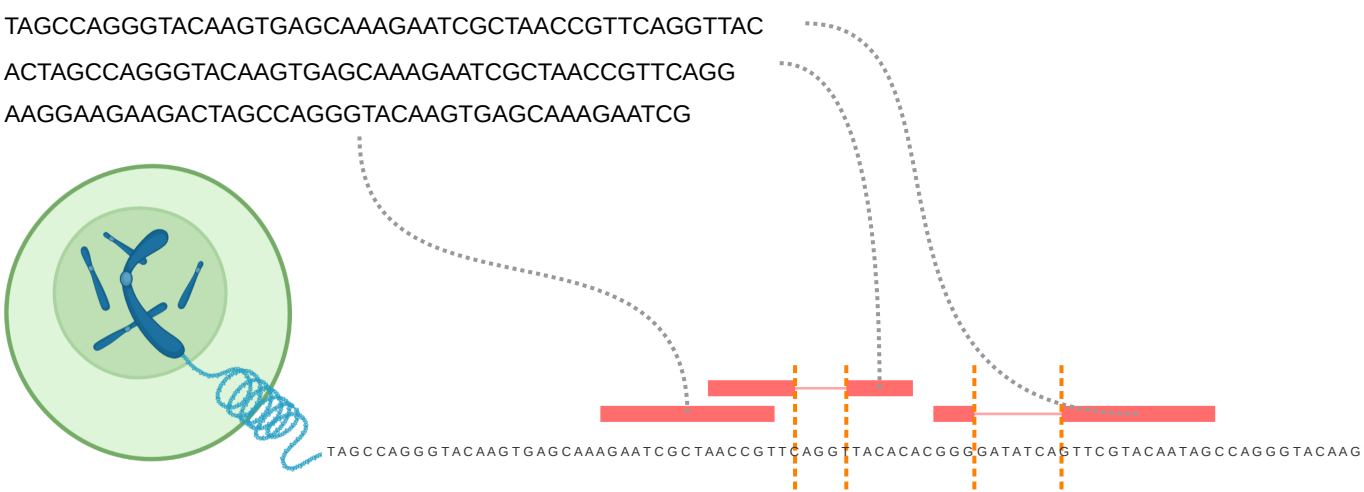
- Secuenciación paralela masiva de ARN



> AGCCAGGGTACAAGTGAGCAAAGAATCGCTAACCGTTCAGGTTACACACGG
> TAGCCAGGGTACAAGTGAGCAAAGAATCGCTAACCGTTCAGGTTACA
> CTAGCCAGGGTACAAGTGAGCAAAGAATCGCTAACCGTTCAGGTTACACACGG

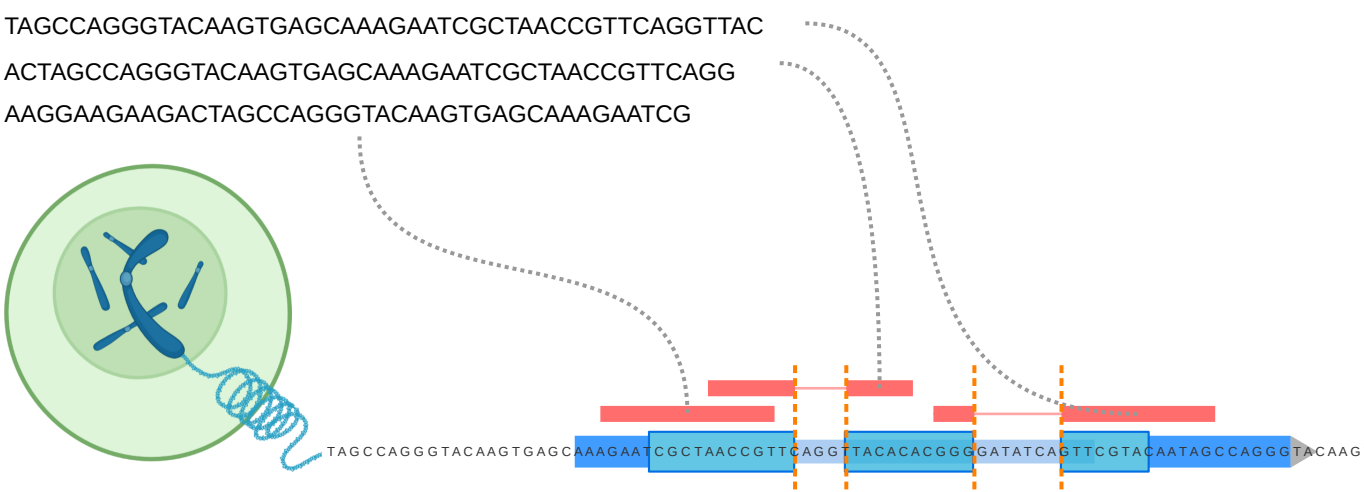
Resumen hasta acá

- Secuenciación paralela masiva de ARN
- Mapeo contra genoma de referencia



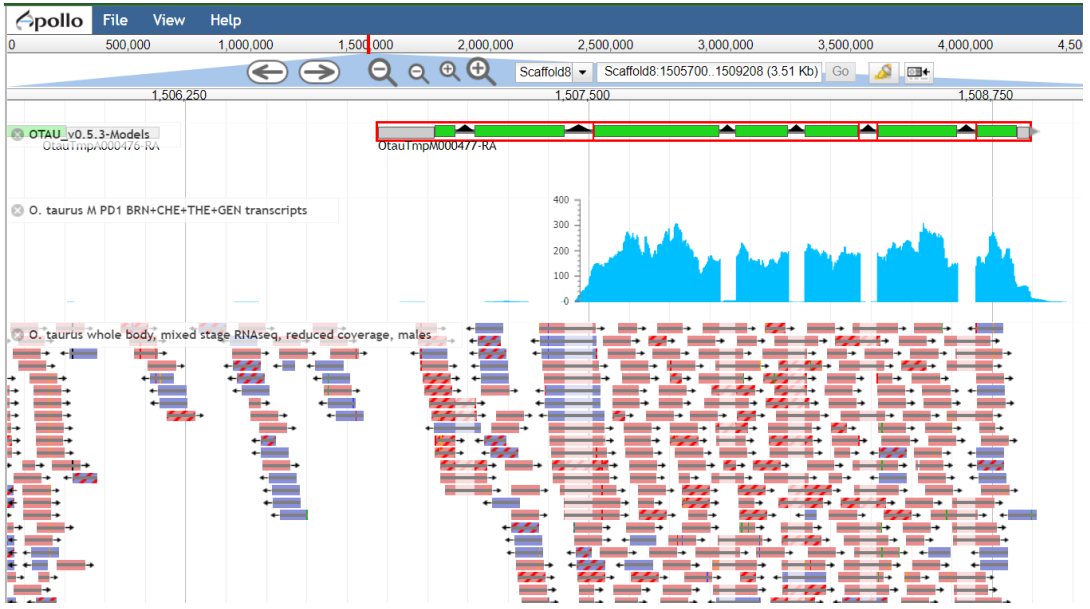
Resumen hasta acá

- Secuenciación paralela masiva de ARN
- Mapeo contra genoma de referencia
- Generación de anotaciones / Modelos de genes



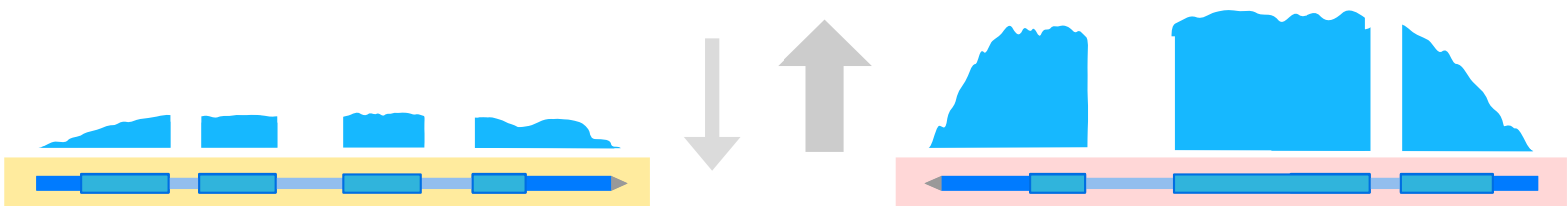
Resumen hasta acá

- Secuenciación paralela masiva de ARN
- Mapeo contra genoma de referencia
- Generación de anotaciones / Modelos de genes
  - Browsers
  - BLAST



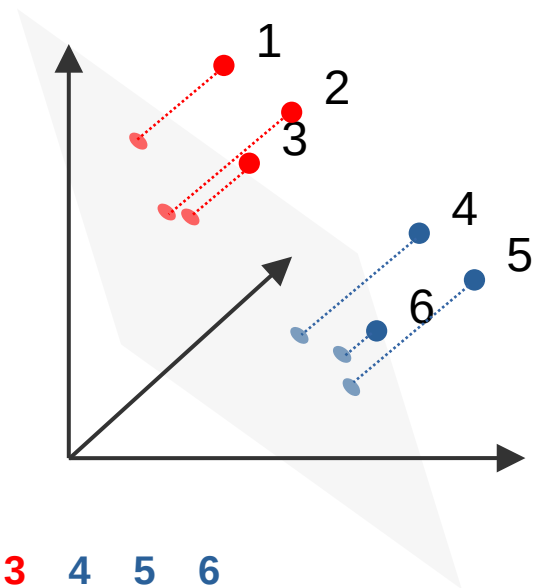
Resumen hasta acá

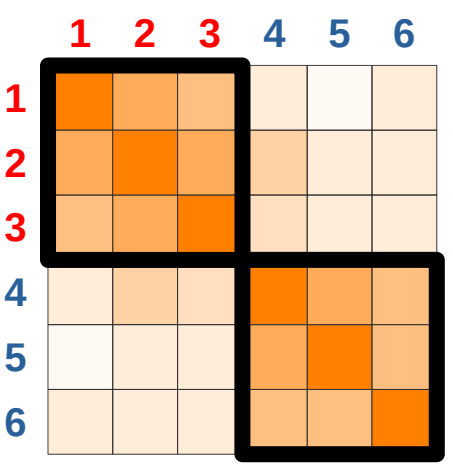
- Secuenciación paralela masiva de ARN
- Mapeo contra genoma de referencia
- Generación de anotaciones / Modelos de genes
  - Browsers
  - BLAST
- Cuantificación de la expresión génica



Resumen hasta acá

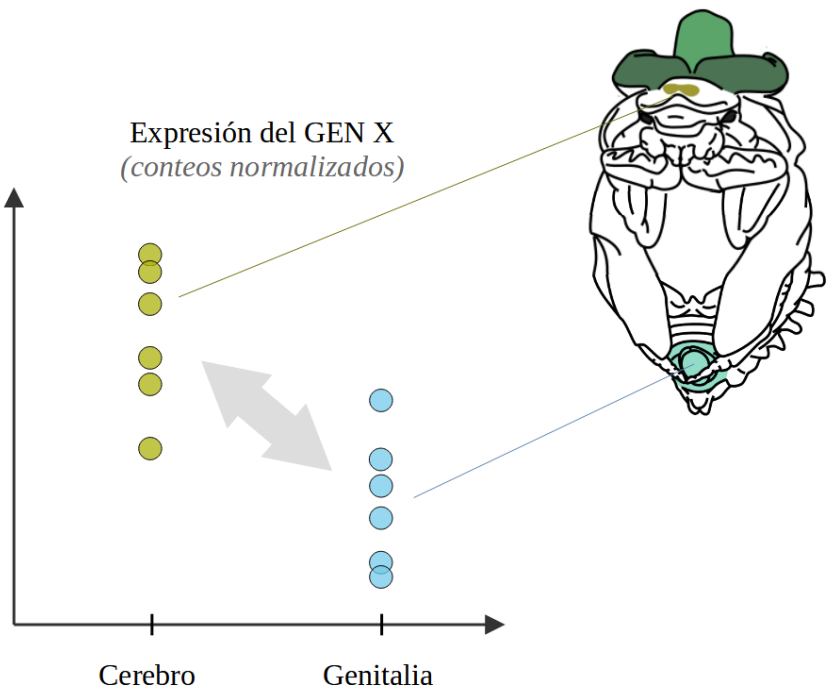
- Secuenciación paralela masiva de ARN
- Mapeo contra genoma de referencia
- Generación de anotaciones / Modelos de genes
  - Browsers
  - BLAST
- Cuantificación de la expresión génica
  - Comparación de muestras de distinto origen
    - PCA
    - Heatmaps





Resumen hasta acá

- Secuenciación paralela masiva de ARN
- Mapeo contra genoma de referencia
- Generación de anotaciones / Modelos de genes
  - Browsers
  - BLAST
- Cuantificación de la expresión génica
  - Comparación de muestras de distinto origen
    - PCA
    - Heatmaps
  - Comparación de genes particulares



Sistema de estudio del TP

Dimorfismo sexual *Onthophagus taurus*

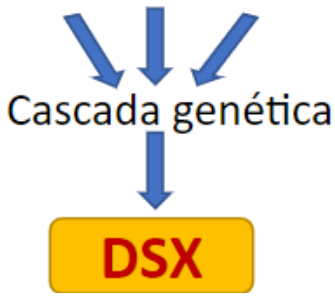
Sistema de estudio del TP

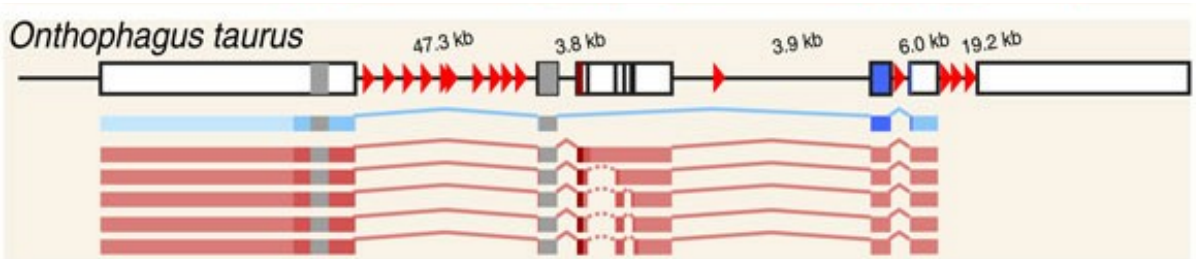
Dimorfismo sexual *Onthophagus taurus*

Sistema de estudio del TP

Dimorfismo sexual *Onthophagus taurus*

Determinación de sexo genética (XY)





Sistema de estudio del TP

Dimorfismo sexual *Onthophagus taurus*

Determinación de sexo genética (XY)

↓

Cascada genética

↓

DSX

↓

Desarrollo de distintos tejidos y órganos

↓

Diferencias entre individuos de sexos distintos

DIMORFISMO SEXUAL

Control (femenino)

Control (masculino)

Sistema de estudio del TP

Diseño experimental

Sistema de estudio del TP

Diseño experimental

x2 Sexos

F
M

Sistema de estudio del TP

Diseño experimental

x2 Sexos

F
M

x4 Tejidos

BRN
CEH
TEH
GEN

Sistema de estudio del TP

Diseño experimental

x2 Sexos

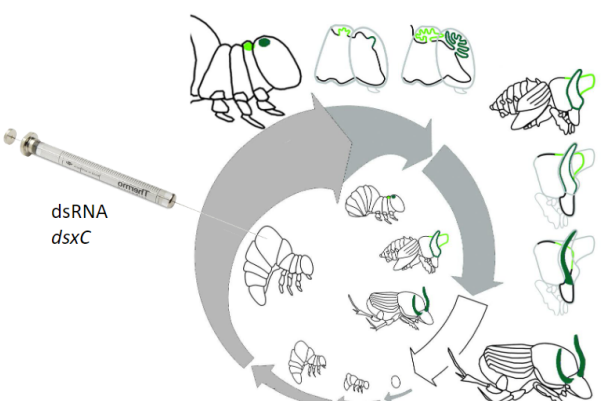
F
M

x4 Tejidos

BRN
CEH
TEH
GEN

x2 Tratamientos

Dsx RNAi
Control



Sistema de estudio del TP

Diseño experimental

x2 Sexos

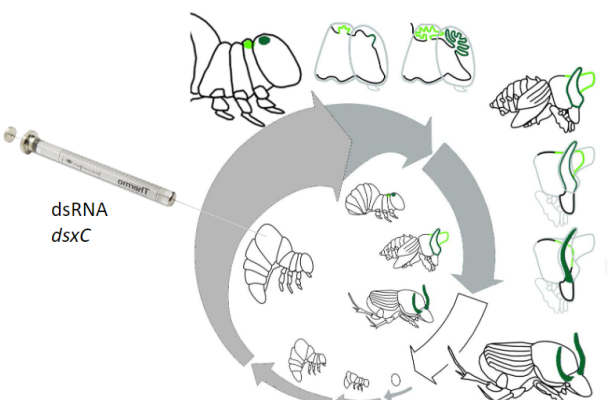
F
M

x4 Tejidos

BRN
CEH
TEH
GEN

x2 Tratamientos

Dsx RNAi
Control



x6 Réplicas

N = 96