

Introducción a RNA-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.
6 de Septiembre, 2017

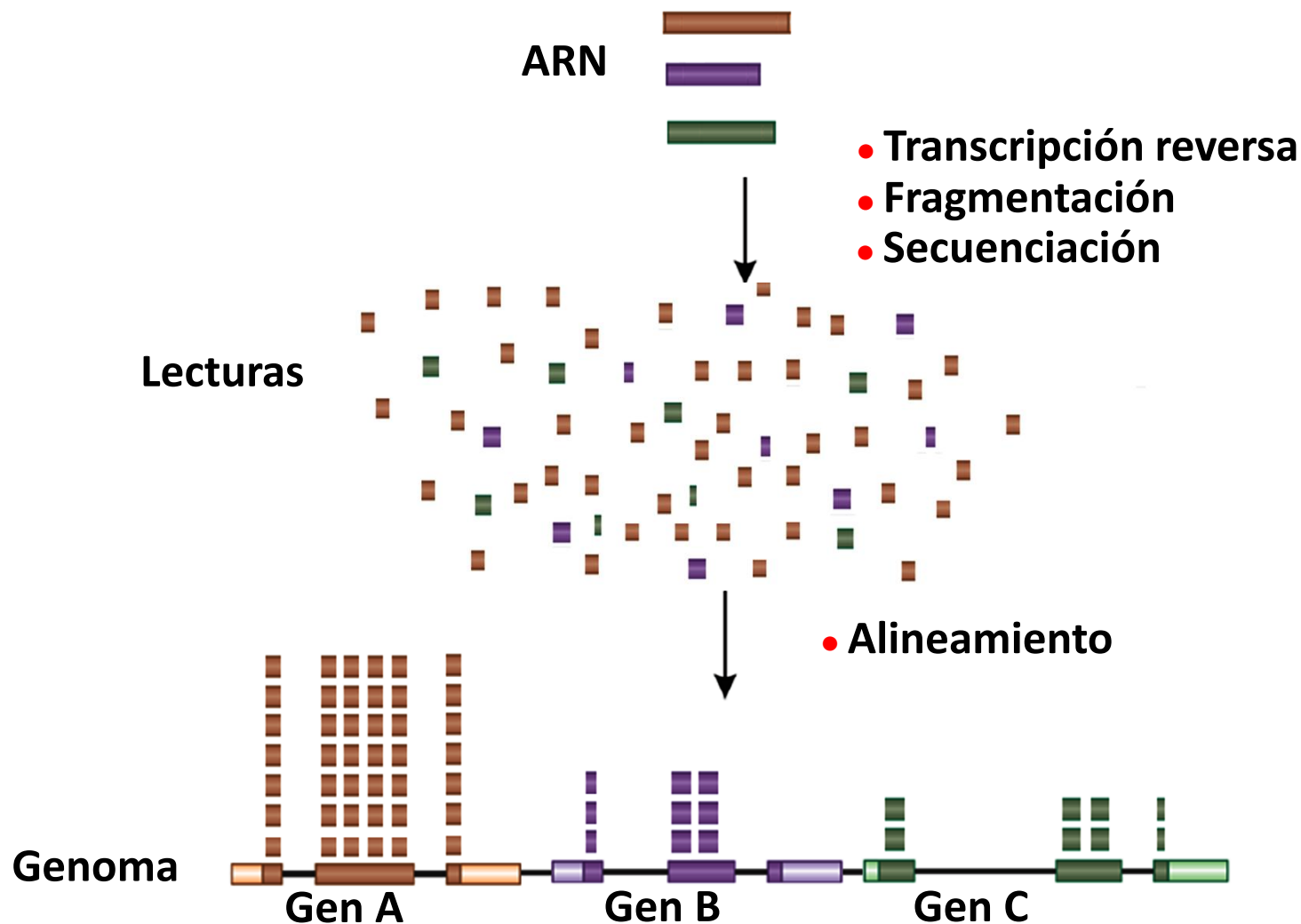
Introducción a RNA-seq

Usos de RNA-seq

- Medir cuantitativamente expresión génica.
- Construir un transcriptoma
- Descubrir nuevos genes.
- Identificar sitios de splicing.

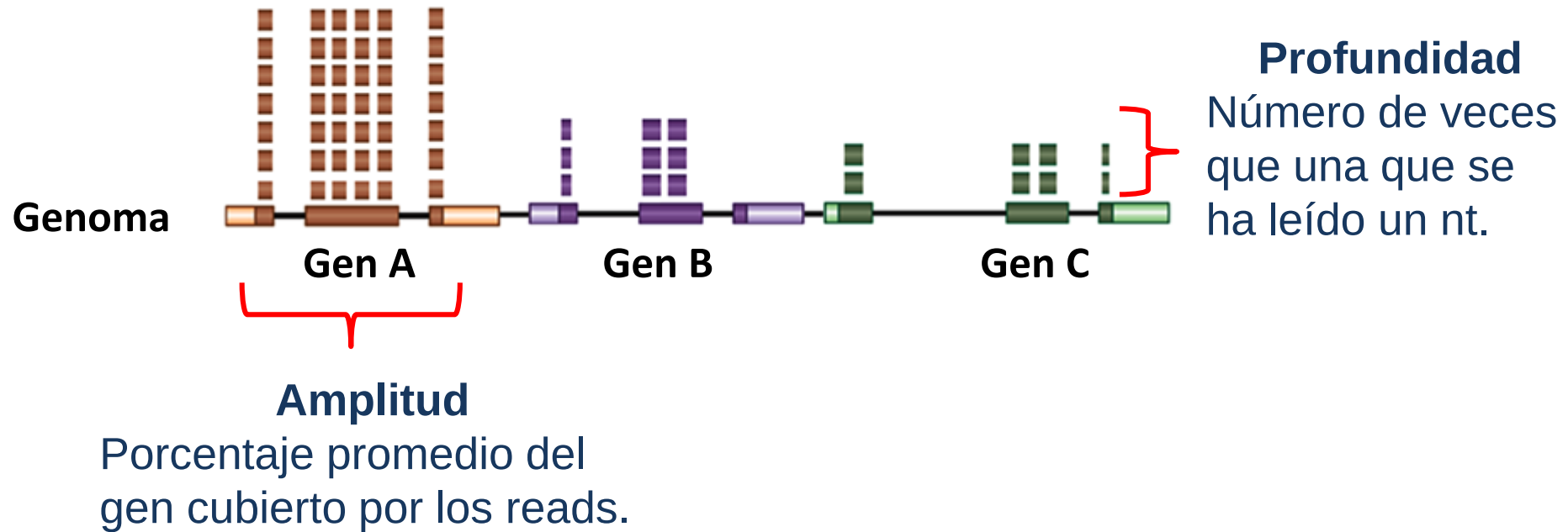
Introducción a RNA-seq

Esquema general RNA-seq



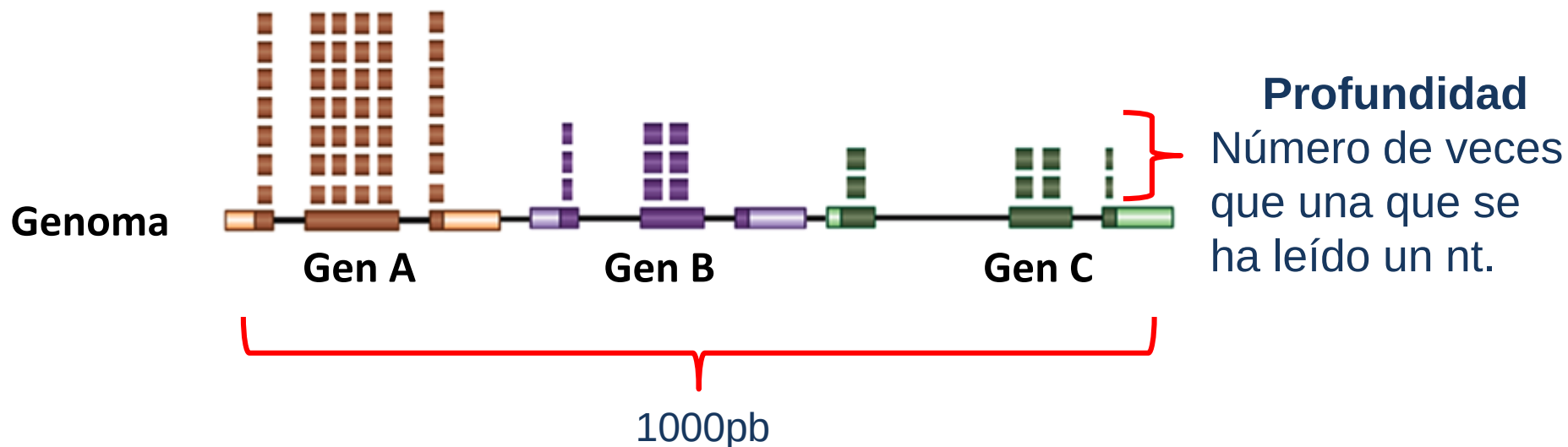
Introducción a RNA-seq

Abundancias



Introducción a RNA-seq

Abundancias



300 lecturas, de 200pb de largo

G: Largo genoma

N: Número de lecturas

L: Largo de lecturas

$$D = N * L / G = 300 * 200 / 1000 = 60x$$

Introducción a RNA-seq

Reads Per Kilobase per Million: RPKM

per base average expression

arbitrary scaling to 1 Million per 1 KB

$$RPKM(G) = \frac{R}{L} \cdot \frac{10^6 \cdot 10^3}{M}$$

R = reads mapped to gene G
L = length of gene G
M = number of reads mapped in Experiment

experiment correction factor

Example:

$$RPKM(G) = 0.5 \cdot \frac{10^6 \cdot 10^3}{10^7} = 50$$

In gene G, 50 reads are found every 1 KB for 1 million mapped reads

“Normaliza” por el largo y el total...

Introducción a RNA-seq

Reads Per Kilobase per Million: RPKM

... RPKM - problems

Condition 1

Gene A 50,000 reads

Gene B 0 reads

$$\begin{aligned} \text{RPKM(A)} &= 50,000 / L * 1e9 / 50,000 \\ &= 1e9 / L \end{aligned}$$

Condition 2

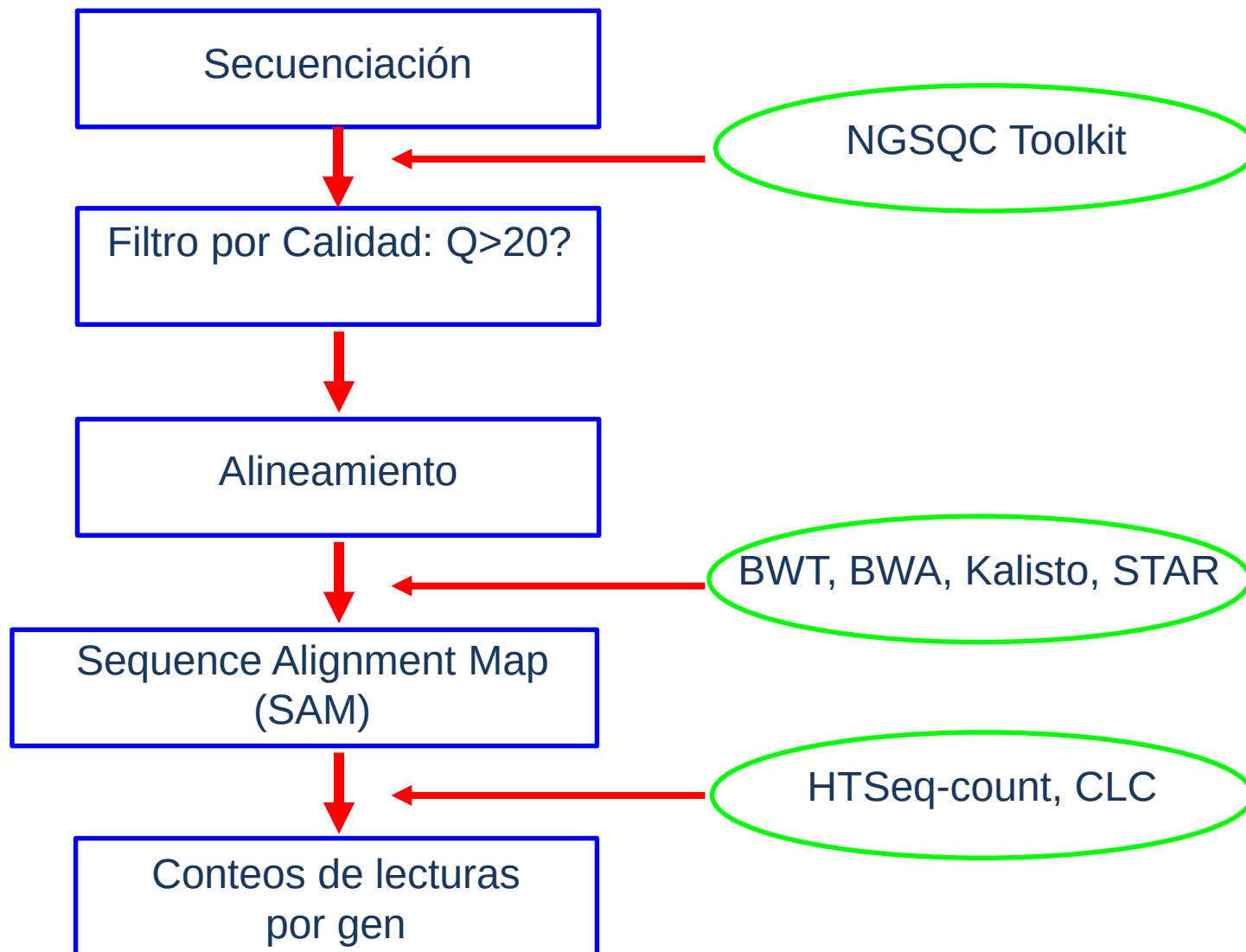
Gene A 50,000 reads

Gene B 10,000 reads

$$\begin{aligned} \text{RPKM(A)} &= 50,000 / L * 1e9 / 60,000 \\ &= 5/6 1e9 / L \end{aligned}$$

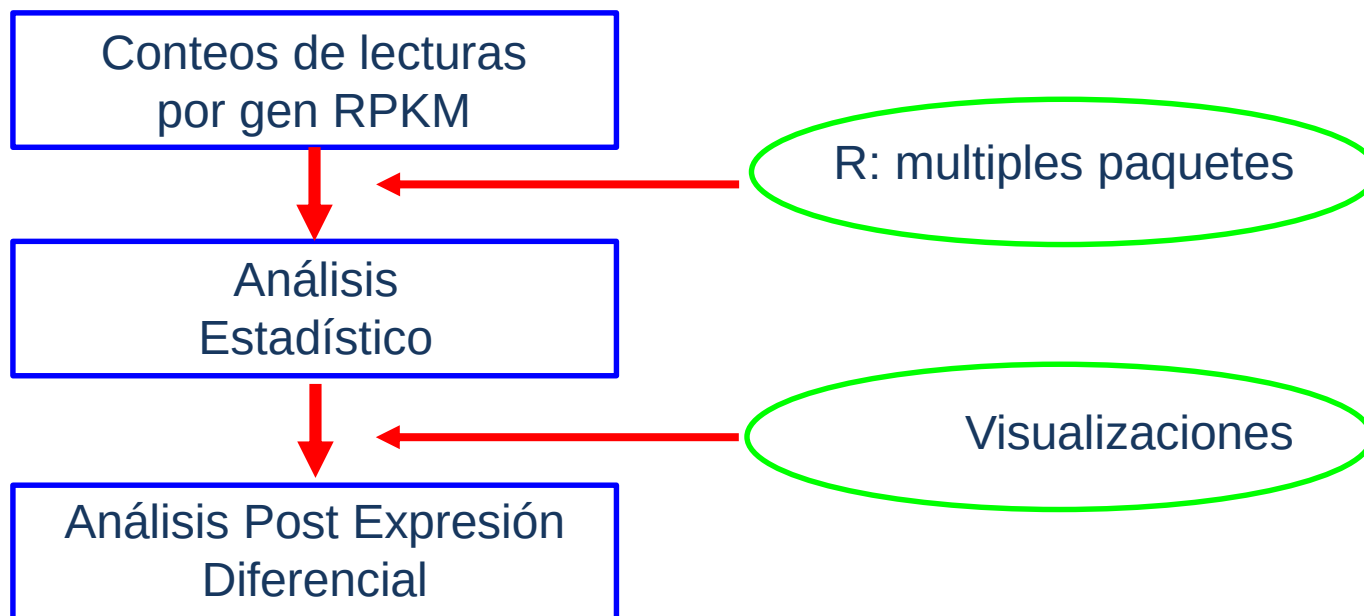
Introducción a RNA-seq

Flujo de trabajo con RNA-seq



Introducción a RNA-seq

Flujo de trabajo con RNA-seq

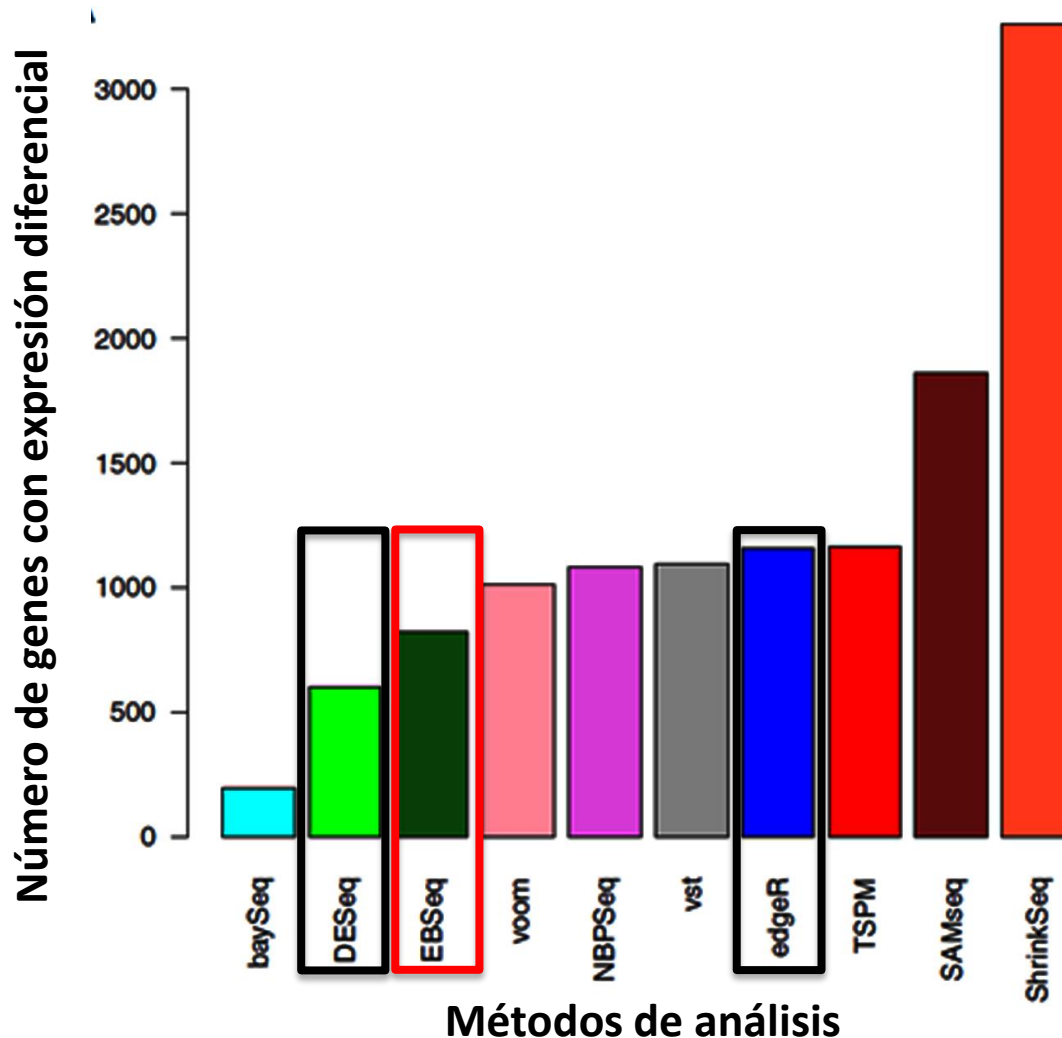


Introducción a RNA-seq

Métodos de Análisis Estadísticos de RNA-seq



Detección genes con expresión diferencial

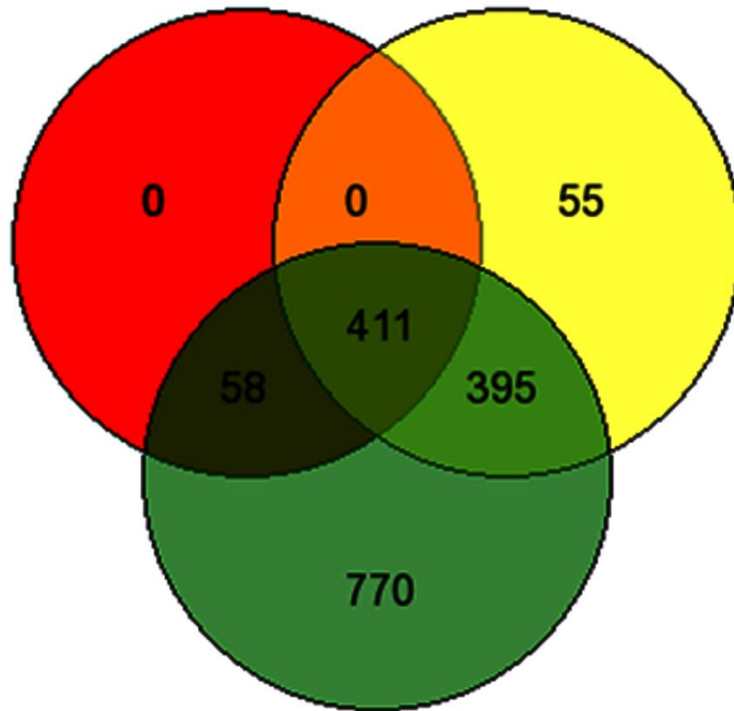


Soneson, 2013.



DESeq2 (469)

EdgeR(861)



Ebseq (1634)

Diagrama de Venn de genes diferencialmente expresados determinados por los 3 métodos estadísticos al comparar la cepa parental versus Δ MIG1 crecidas en presencia de glucosa.



Pruebas estadísticas:

- T-test.
- F-test.
- Exact Test.
- Alternativas Bayesianas

$$H_0 : \mu_{1j} - \mu_{2j} = 0$$

EdgeR

$$E(y_{gi}) = \mu_{gi} = N_i \pi_{gi}$$

DEseq2

$$K_{ij} \sim \text{NB}(\mu_{ij}, \alpha_i)$$

$$\mu_{ij} = s_j q_{ij}$$

$$\log_2(q_{ij}) = x_j \cdot \beta_i$$

EBseq

$$X_{gi,s}^C | r_{gi,s}, q_{gi}^C \sim \text{NB}(r_{gi,s}, q_{gi}^C)$$

$$q_{gi}^C : q_{gi}^C | \alpha, \beta^{I_g} \sim \text{Beta}(\alpha, \beta^{I_g})$$

Introducción a RNA-seq

Métodos de Análisis Estadísticos de RNAseq

- Inferencia Clásica

EdgeR

DEseq2

$$H_0 : \mu_{1j} - \mu_{2j} = 0$$

- Valores p (ajustado)

- Inferencia Bayesiana

EBseq

$$H_1 \text{ (DE)} : q_{g_i}^{C1} \neq q_{g_i}^{C2}.$$

- Probabilidades a Posteriori

Introducción a RNA-seq

Práctico R

- Análisis Expresión Diferencial mediante Inferencia frecuentista y Bayesiana para levaduras wild type y mutante.
- Introducción al uso de Biomart
- Manejos de data.frames y visualizaciones en ggplot2