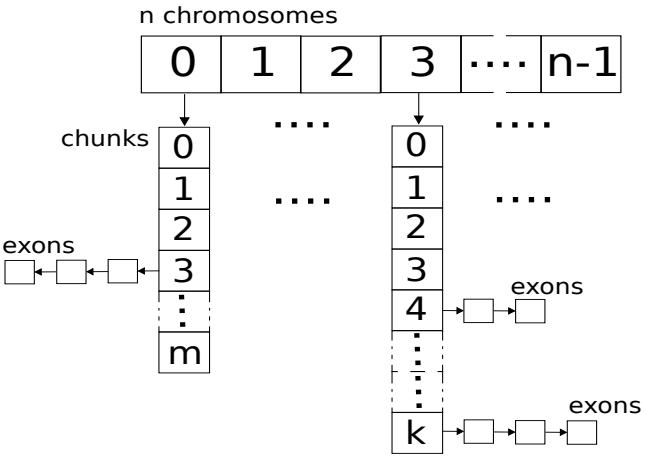
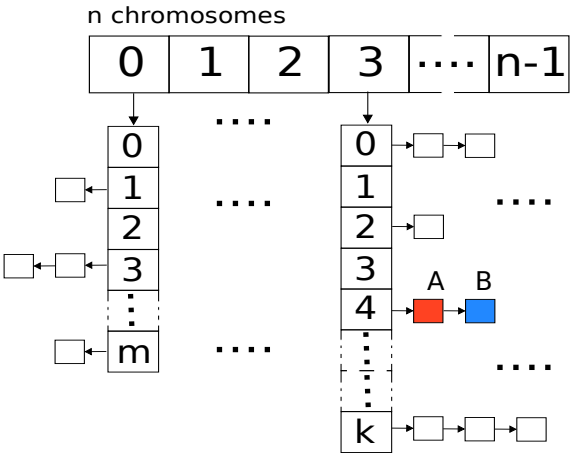
Highly sensitive and ultrafast read mapping for RNA-seq analysis

Medina, I., Tárraga, J., Martínez, H., Barrachina, S., Castillo, M.I., Paschall, J., Salavert-Torres, J., Blanquer-Espert, I., Hernández-García, V., Quintana-Ortí, E.S. & Dopazo, J.

Supplementary Figure 1



(a)



(b)

Metaexon data structure

The metaexon data structure consists of an array of chromosomes, each one with an array of chunks (a). Exons are stored in a sorted list of a given chunk according to the exon chromosome and coordinates.

How the metaexon structure improves alignment accuracy by reducing soft-clipped bases

1)

Reads partially aligned, e.g., 92M8S, don't modify the metaexon structure, they are stored in a temporal file to be solved after the whole fastq file has been processed.

2)

Reads completely aligned, e.g., 42M627N58M, update the metaexon structure. In our example, two exons A and B are added to the metaexon structure sorted by their positions (). The first 42 bases correspond to exon A and the last 58 bases to exon B. In the metaexon structure, we save the end position of exon A, the start positon of exon B and the distance between them, 627bp.

3)

Once the whole fastq file has been processed, the metaexon structure is completely full and the partially aligned reads, stored in a temporal file, can be completed.

In our previous alignment 92M8S, considering that the first 92 bases mapped at the end of exon A, then we can mapped the last 8 bases to the beginning of exon B, providing a much more accurated alignment, 92M627N8M.