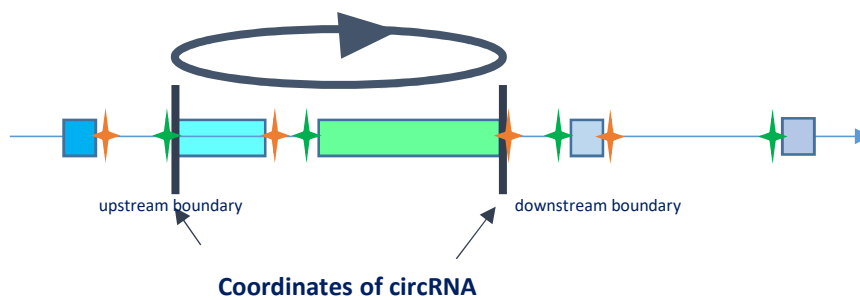
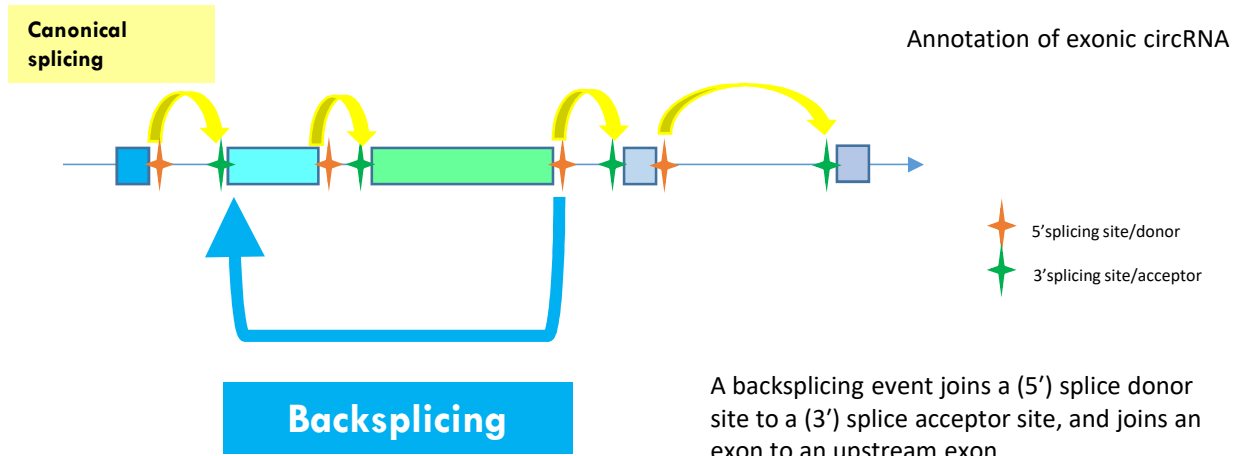
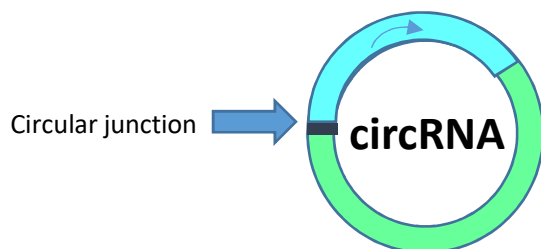


A canonical splicing joins an upstream (5') splice donor site with a downstream (3') splice acceptor site, and joins an upstream exon to a downstream exon

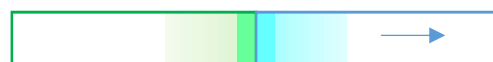


Upstream **exon** that is the “acceptor exon” in the backsplicing

Downstream **exon** that is the “donor exon” in the backsplicing



Circular junction



Circular RNAs : Vocabulary and annotations

➤ *A few vocabulary questions*

- Total-RNAseq: sequence data from total RNA libraries after depletion of ribosomal RNA
- mRNA-Seq: sequencing data from mRNA libraries, i.e. after polyA selection
- circRNA: to design all types of circular RNA supported by reads spanning a circular junction. These reads are aligned to the reference genome sequence as two segments mapping in reverse order.
- Those annotated as “ciRNA” correspond to circRNAs localized entirely in intronic sequences and with the circRNA 5' junction site corresponding to the intron donor site
- Intronic circRNA: intron circle + lariat derived circRNA
- Exonic circRNAs: only circRNAs resulting from backsplicing between two known exons. See also Suppl. Figure S1
- As we considered all circRNAs, not just those derived from backsplicing, it appeared to us that the term BSJ (backsplicing junction) is not appropriate to name the circular junction.

➤ *Presentations of CD*

The source code of the CD is available from <https://github.com/GenEpi-GenPhySE/circRNA.git>.

CD identifies reads containing a circular junction within those reads that STAR calls “chimeric reads” (CR) from the tabular file (chimeric.out.junction) provided by STAR: Subsequently, these reads will be called “circular chimeric reads” (CCRs). The main CD-output file (detection.bed) consists of a list of all circular RNAs and their associated number of CCRs, each circular RNA is being defined by the coordinates of the circular junction (chromosome:start-end|strand).

For the detection step, the user can select a threshold x to retain only circRNAs characterized by at least x CCRs and a minimal genomic size for circRNAs (distance between the two feet of the circRNA). In this study, we have chosen not to consider non-redundant or sporadic circularization events. Several studies have shown the value of excluding such events [Gruhl et al. 2021 ; Xu et al. 2021]. The detection of circRNAs was performed individually for each dataset by CD with a threshold of 5 CCRs.

The annotation is made to the precision of the base. CD does not tolerate any differences from the gtf file and does not perform any grouping. If there are three circRNAs, after annotation, it will keep three. There is one parameter that can be modified to enable grouping, but we strongly advise against using it.

This circRNA tool was specifically designed for intronic circRNA characterization [Robic et al. 2022 ; Robic et al. 2021]. The identification of the reads supporting the circular junction is performed from the single end alignments of the PE reads, and the compatibility of the second read with this characterization is not verified. Given the difficulty of sequencing the 2'-3' circular junction, this is an undeniable advantage for the characterization of intronic circRNAs [Robic et al. 2020a ; Robic et al. 2020b].

➤ ***Annotation of exonic circRNAs***

Both junctions correspond to exonic boundaries from a single gene located on the same strand as circRNA. Consequently, the circRNAs must satisfy the three following rules

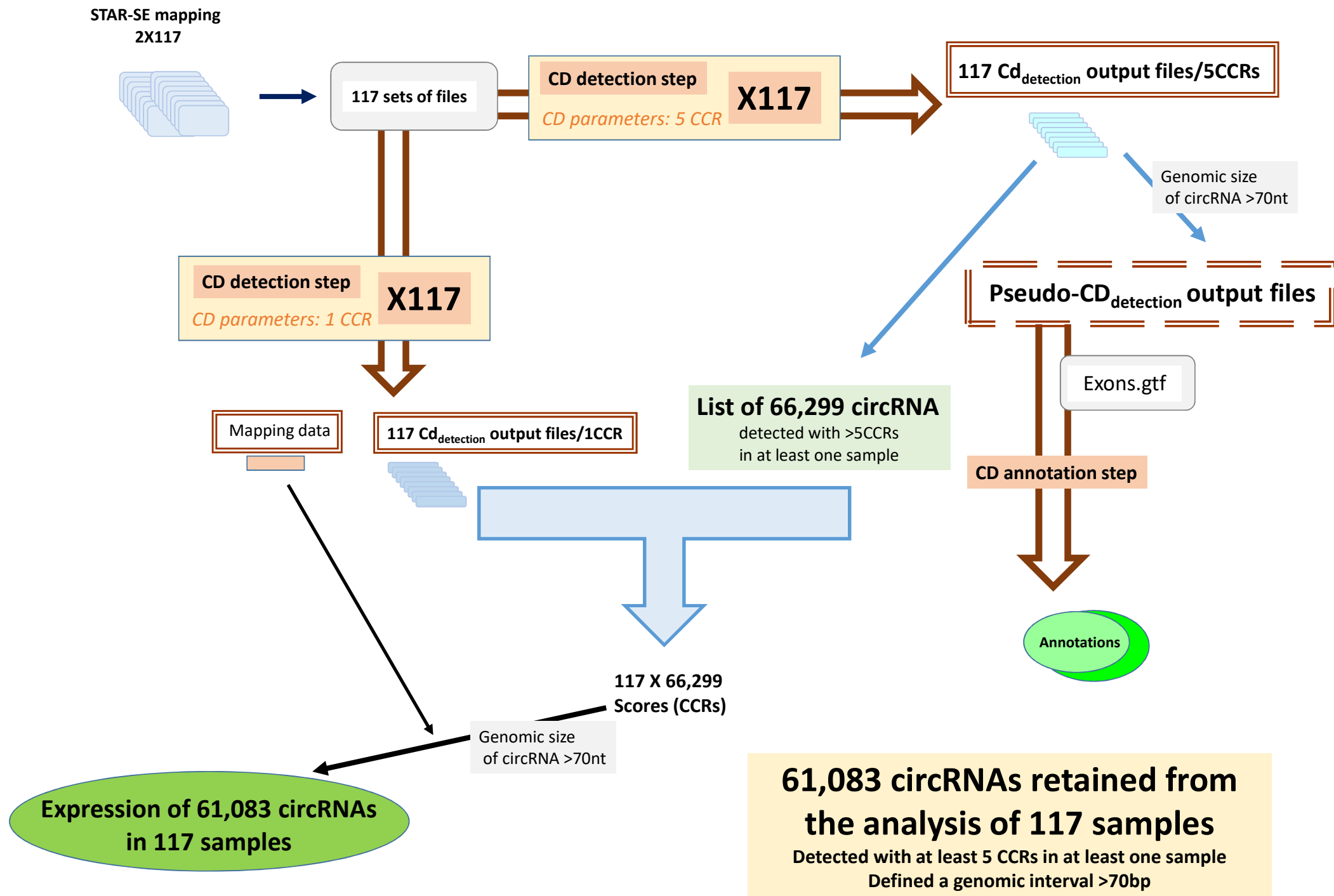
- The 3' junction of a circRNA must precisely correspond to an exon donor site (3' end of an exon, ie 5' donor site of the next intron) from a gene located on the same strand as circRNA
- The 5' junction must precisely correspond to an upstream exon acceptor site (5' end of an exon, ie 3' acceptor site of the previous intron) from a gene located on the same strand as circRNA
- The exon donor and the exon acceptor are associated to a common gene

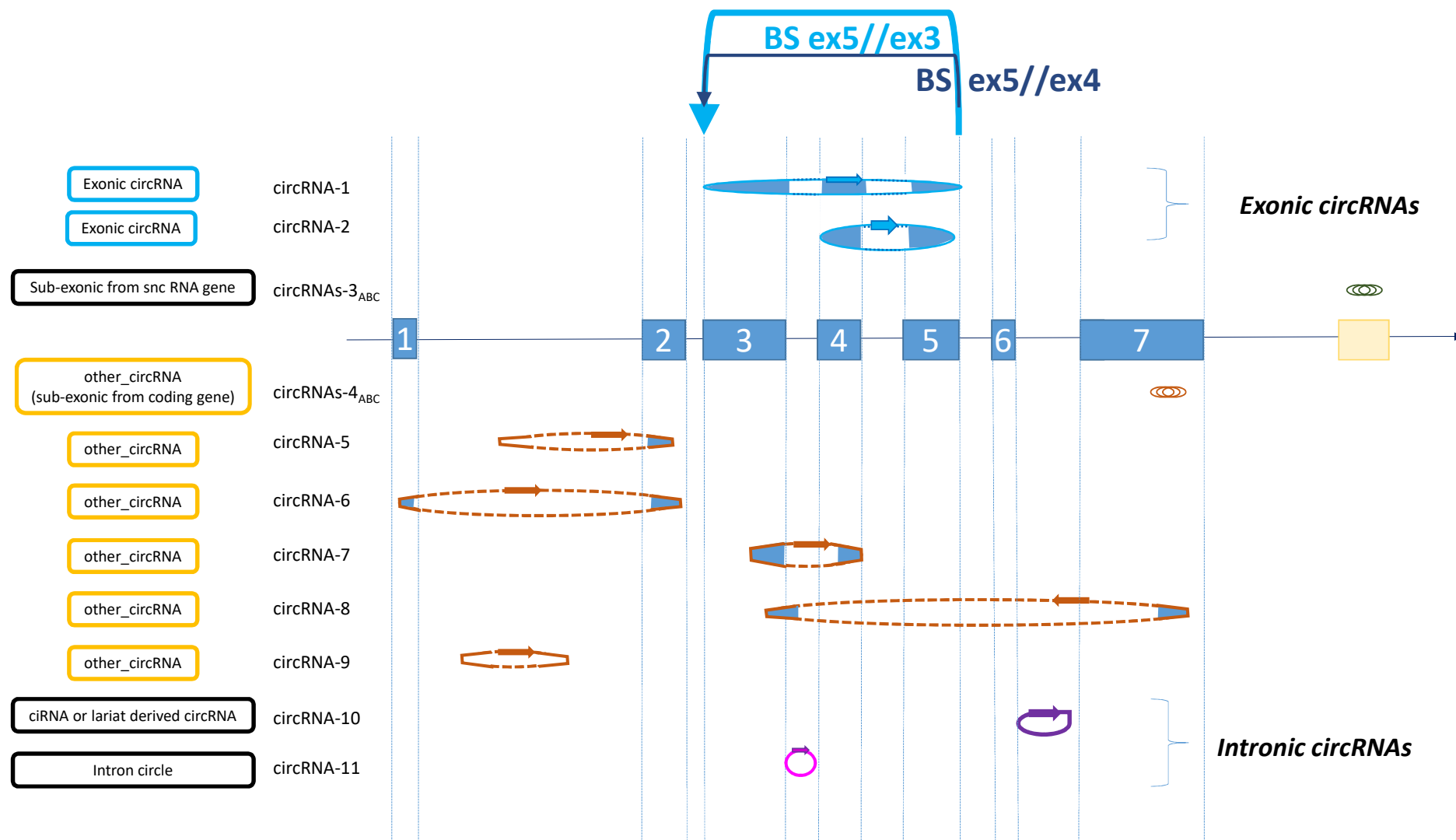
➤ ***Annotation of intronic circRNAs: lariat-derived intronic circRNA (or ciRNA) and intron circle***

- both junctions are located within a single intron
- the 5' junction must precisely correspond to the 5' intron donor site
- the 3' junction must be compatible with a circularization event limited by the branch point (12-60 base pair away from the 3' intron acceptor site) for lariat derived circRNAs
- the 3' junction must be compatible with a circularization of the entire intron (-5/5 base pair away from the 3' intron acceptor site) for intron circle

➤ ***Annotation of sub-exonic circRNAs from snc genes***

- Both junctions are located within a single exon
- This class contains no circRNA classified as Exonic
- Only the ones that are associated to a gene not reported as lnc, coding gene or pseudo-gene





The circRNA-1 and -2 are exonic circRNAs from the blue gene. The set of circRNAs-3 groups sub-exonic circRNAs from a snc gene and they can be in sense or antisense. Only circRNA-10 and -11 are intronic circRNAs (purple and pink). The first is a lariat-derived circRNA (ciRNA) and the second is an intron circle and they contain only intronic sequences. We chose to classify those that are not identified by CD as being exonic or intronic circRNAs or sub-exonic circRNAs originating from the snc gene as other_circRNAs. It is sort of a provisional class and our goal is to sort it out. circRNA-5, -6, -7, -8 and -9 and the set of circRNAs-4_{ABC} are other_circRNAs (in brown). The classification exonic/other_cirRNA/miscellaneous appears as a colored frame (blue/orange/black) on the left.

Several elements can be defined in a region of chromosome A by their coordinates (chr: Start,End | strand):

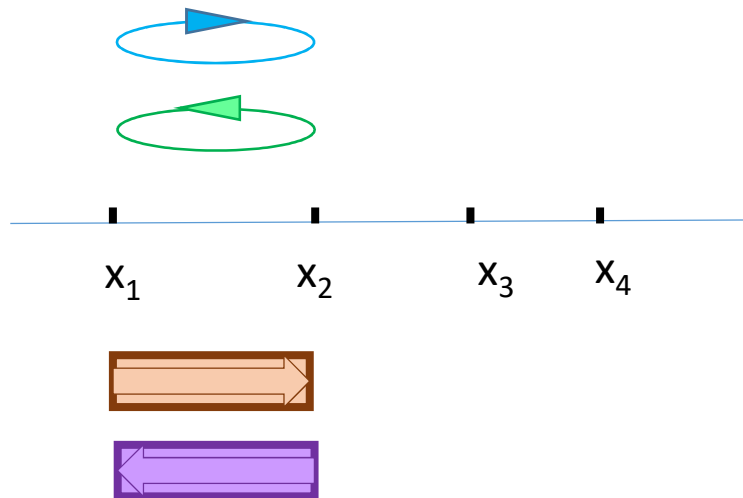
circ-1 is defined by chrA: x_1, x_2 | +

circ-2 is defined by chrA: x_1, x_2 | -

exon-1 is defined by chrA: x_1, x_2 | +

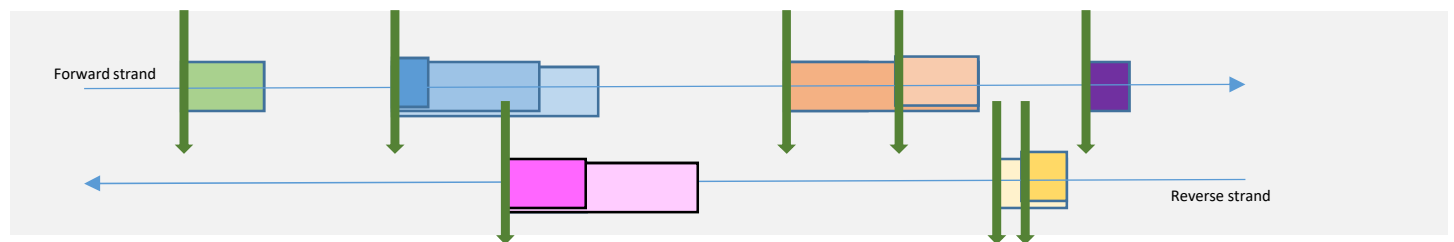
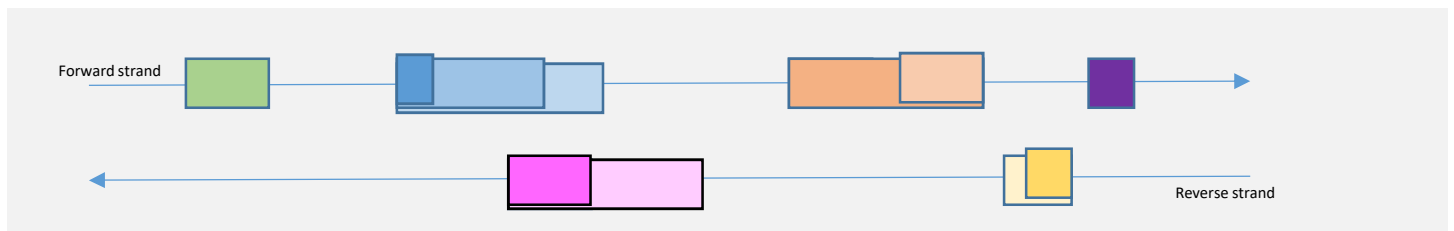
exon-2 is defined by chrA: x_1, x_2 | -

In every case x_1 is the first genomic coordinate and x_2 the second.



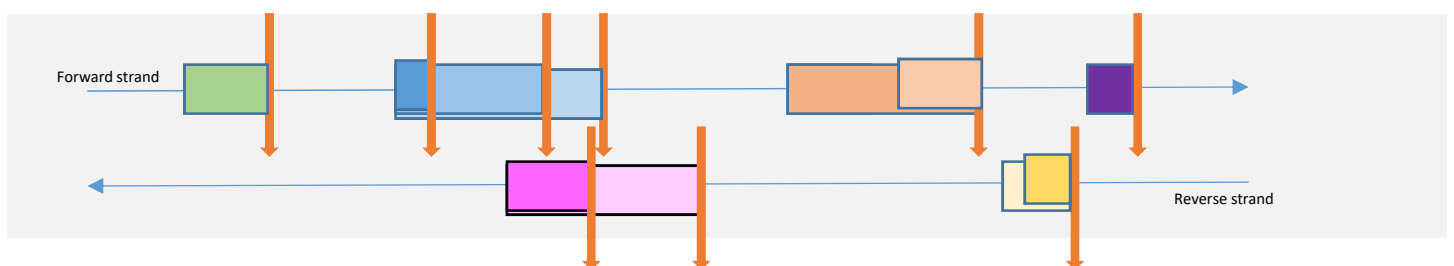
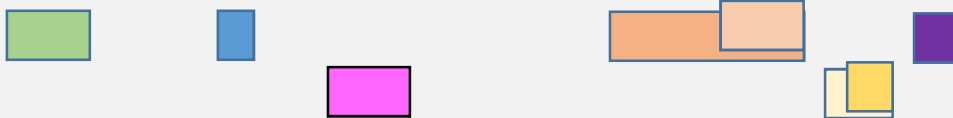
In this example, we can annotate **circ-1** and **circ-2** as a **single exon circRNA** from **exon-1** and **exon-2**, respectively

To perform what we call a minimal_annotation of exonic circRNAs, we created two sub-lists (Left_exons and Right_exons) from the list of all BovReg exons.

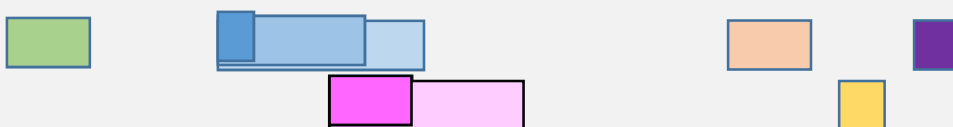


To constitute the list of Left_exons, we selected exons according to their unique first genomic coordinate keeping only the exon with the smallest size in case of multiple exons with the same first coordinate. This first coordinate (indicated by a green arrow) corresponds to the 5' coordinate when the exon is defined on forward strand and to the 3' coordinate for exons defined on reverse strand.

Retained exons for the list of Left_exons

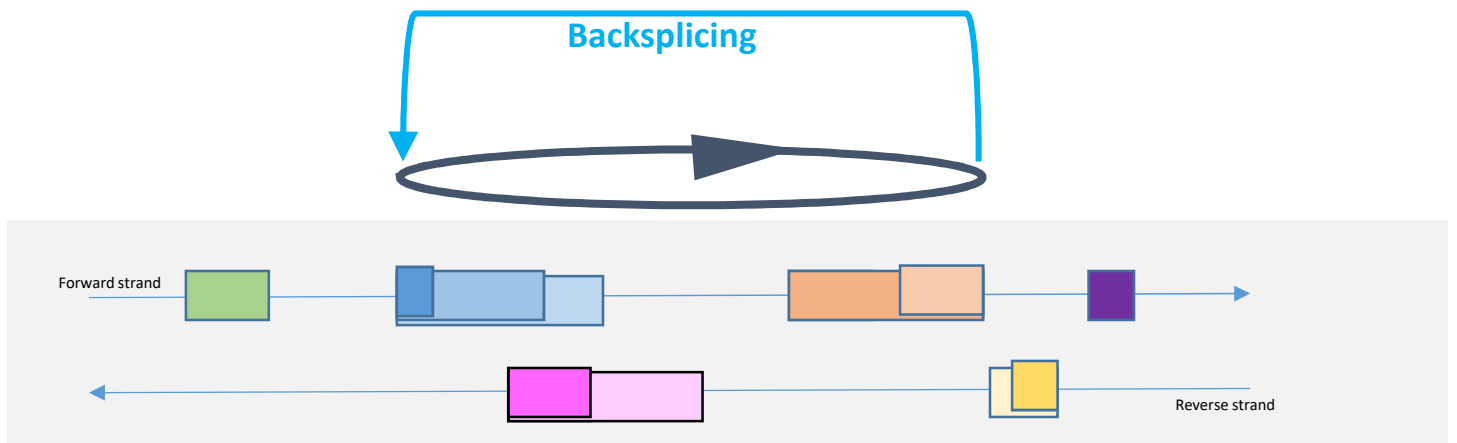


Retained exons for the list of Right_exons

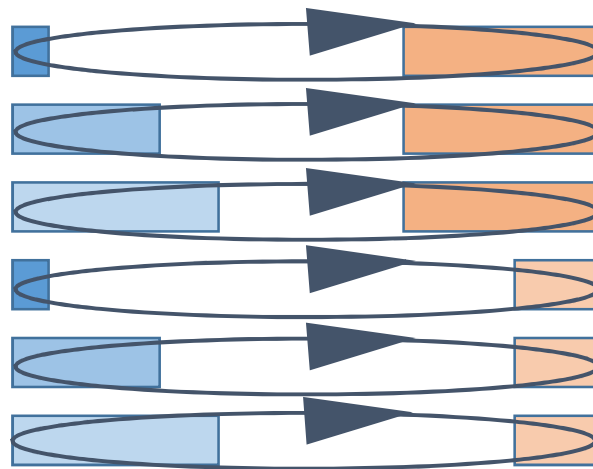


M&M_Adoc-6A

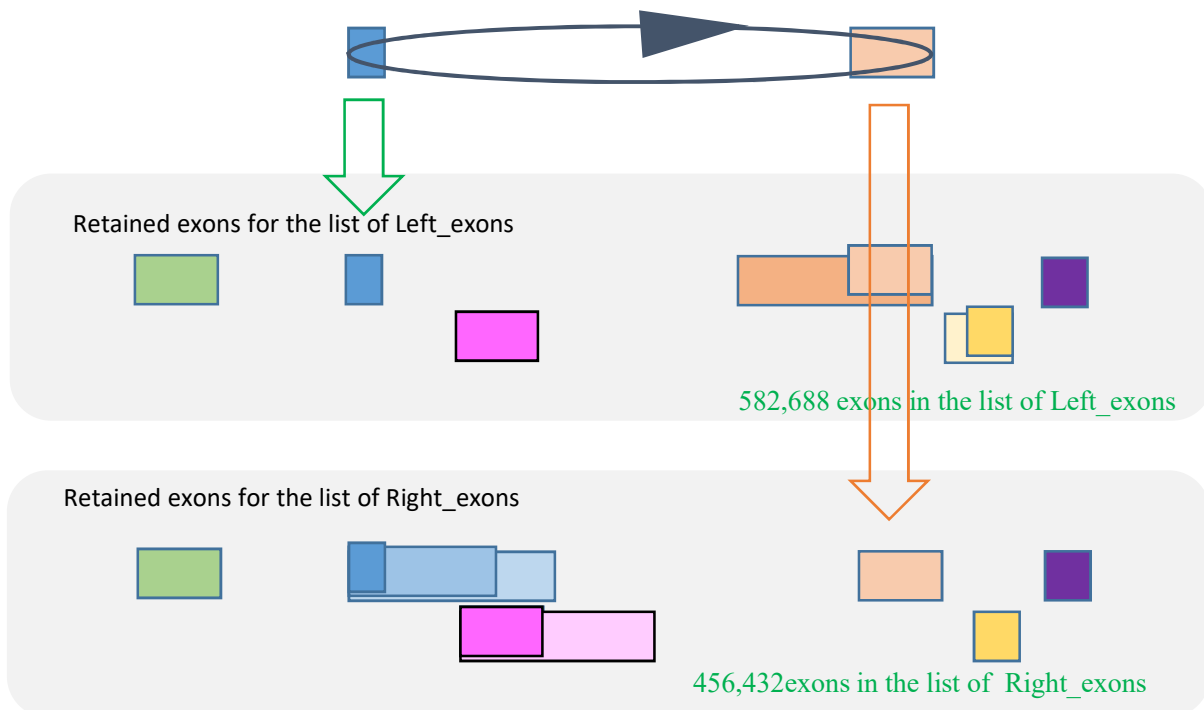
**The constitution of the two sub-lists
to perform the minimal annotation**



Six distinct annotations of this exonic circRNA were possible



Annotation retained as Minimal_annotation



M&M_Adoc-6B

The constitution of the two sub-lists
to perform the minimal annotation