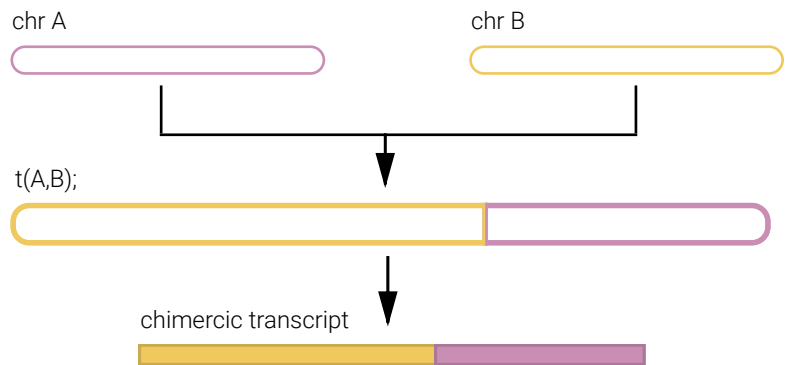


Anatomy of a chimeric junction identified with CRAC algorithm

Input data

Chimeric event
Chromosome A and B are translocated together. A chimeric transcript is produced from the fusion of two genes, one from chromosome A and the other from chromosome B.



chimeric read

A read overlapping the chimeric junction.

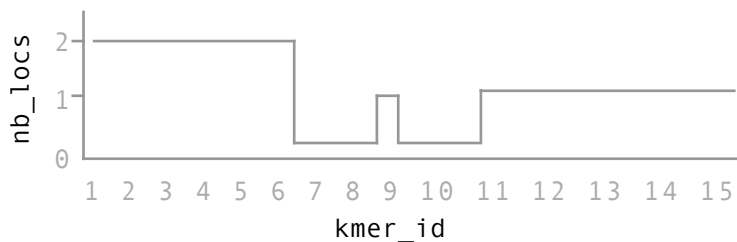
ATGCTGTAGG GCTGATCGAC

CRAC algorithm

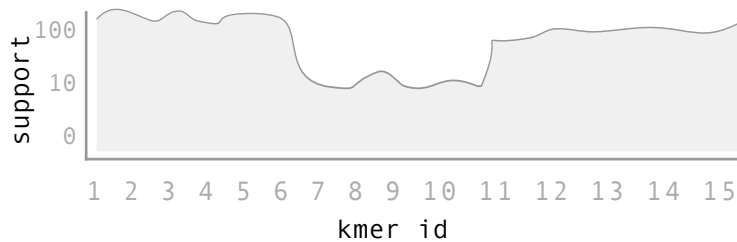
CRAC k-mer analysis
CRAC procedure to analyse reads relies on the analysis of k-mer profiles.



localization profile
Number of time each k-mer is found on the reference genome.



support profile
Number of time each k-mer is found in the input dataset.



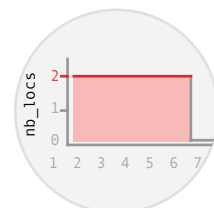
ML features

score break length
Measure of distance between the observed break length and the theoretical break length.

$$\text{score_break_length} = \text{observed_break_length}() / (k - 1) = 1$$

score is duplicated
If the k-mer location around the break is ambiguous.

score_is_duplicate = true



score coefficient variation
Coefficient variation of the support profile.

$$\text{score_coeff_var} = \text{std_dev}(\text{support}[]) / \text{mean}(\text{support}[])$$