Complete Genome analysis: Structural variation detection

SVDetect tutorial

Alban LERMINE – Institut Curie, INSERM U900, Mines ParisTech

institutCurie

Workshop outlines

1) SVDetect tool presentation

Methodology & functionalities

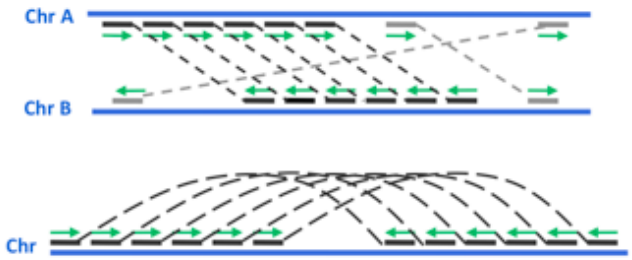
2) SVDetect tutorial on Galaxy

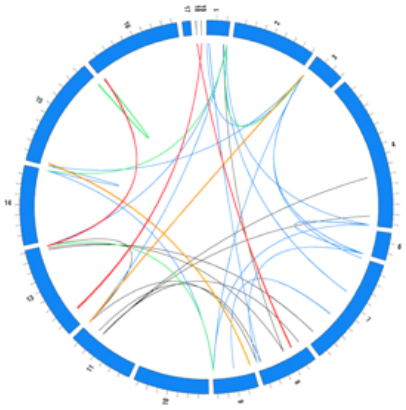
Hands on workshop

SVDetect

SVdetect is a tool dedicated to detection and prediction of inter/intra chromosomal rearrangements from paired-end/mate-pair sequencing data

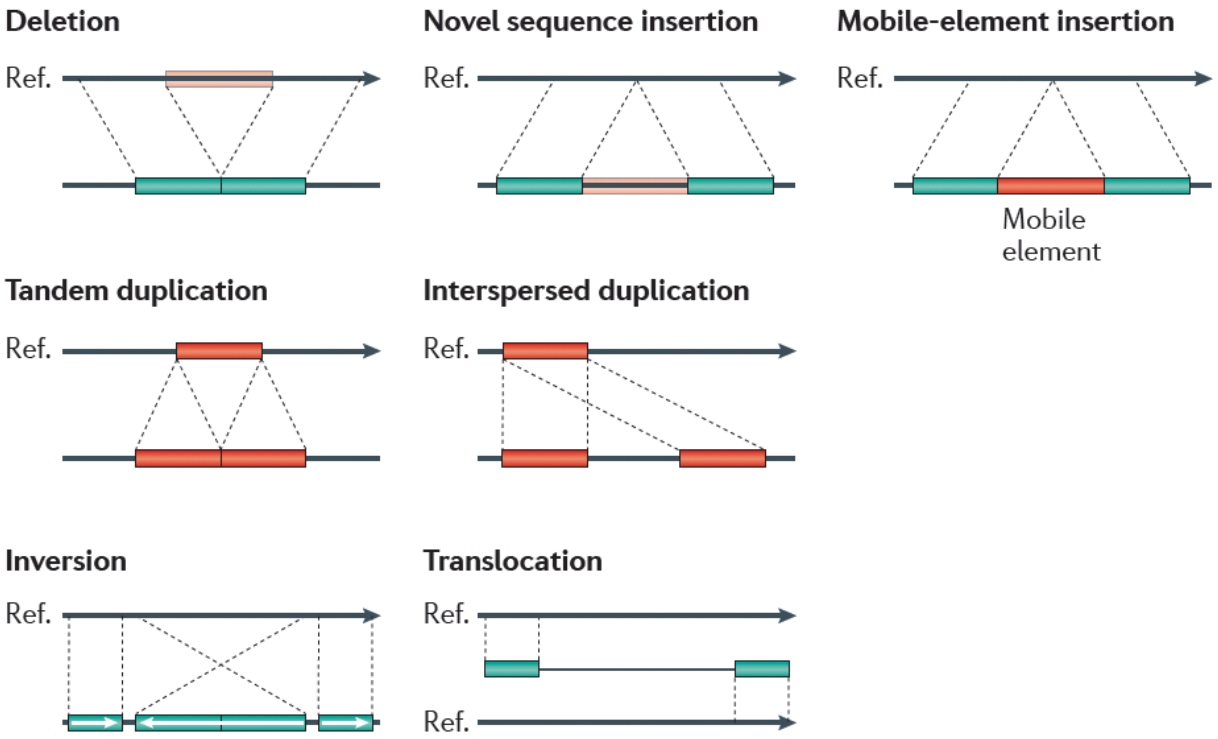
This tool detect structural variation (SV's) by using sliding windows and clustering strategies, and also allows you to visualize them at genome scale





Different classes of SV's

Lot of Tumors have chromosomal mutations in their genome. This can affect the number of chromosomes, and also its structure, implicating large events such as deletions, insertions, duplications, inverisons,translocations, etc.



Alkan et al 2012

SV's discovery from sequencing data

4 different methods:

- Based on coverage depth differences (Control- FREEC)
- Splitted reads method, local reads realignment around breaking point (Pindel)
- De novo assembly, new sequences - insertions -
- Paired-end/mate-pair sequences constraints (Breakdancer, PEMer, SVDetect)

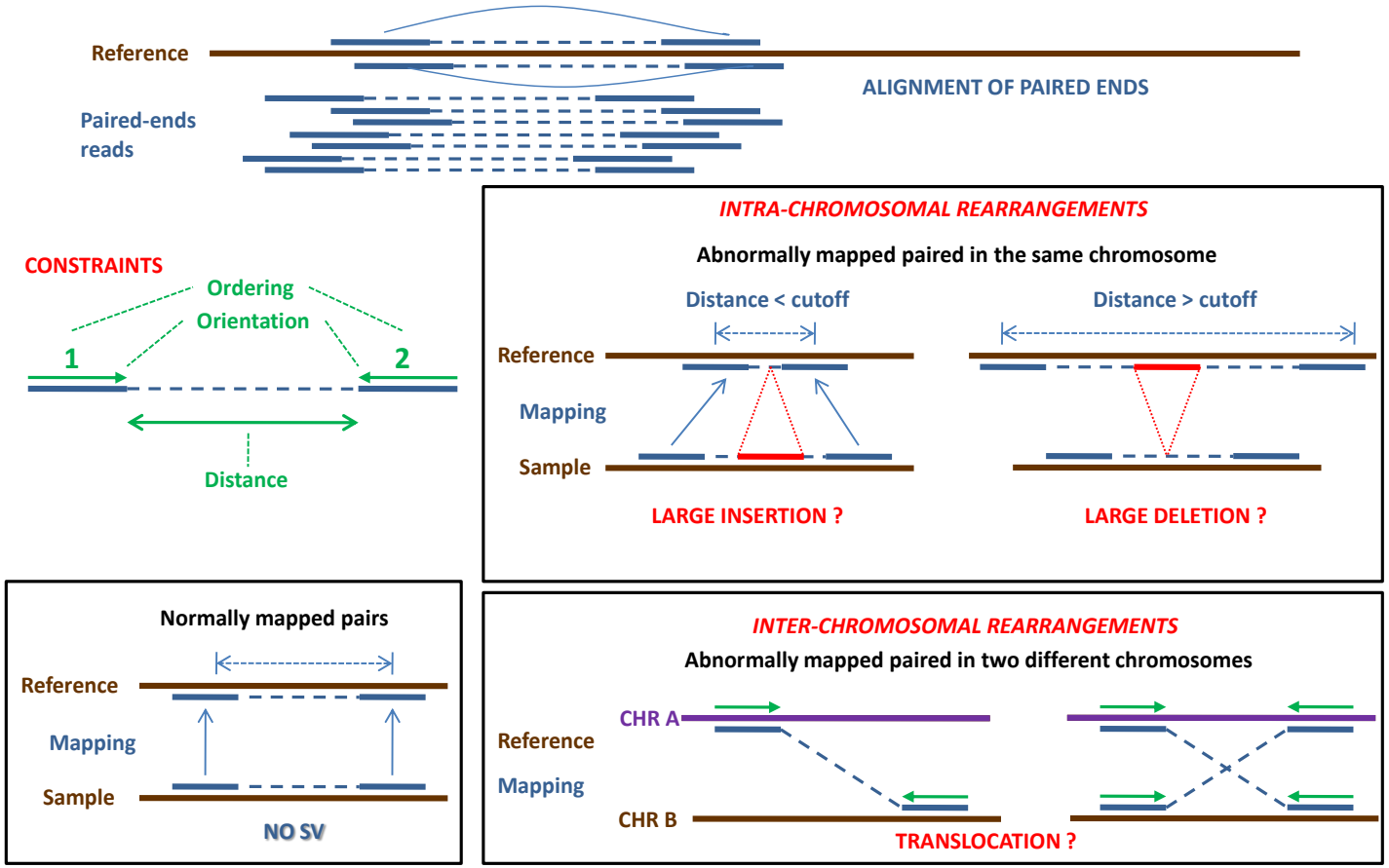
Reviews on thema:

Computational methods for discovering structural variation with next-generation sequencing
Paul Medvedev et al, Nature Methods 6, S13 - S20 (2009)

Genome structural variation discovery and genotyping.
Alkan C et al, Nat Rev Genet. (2012)

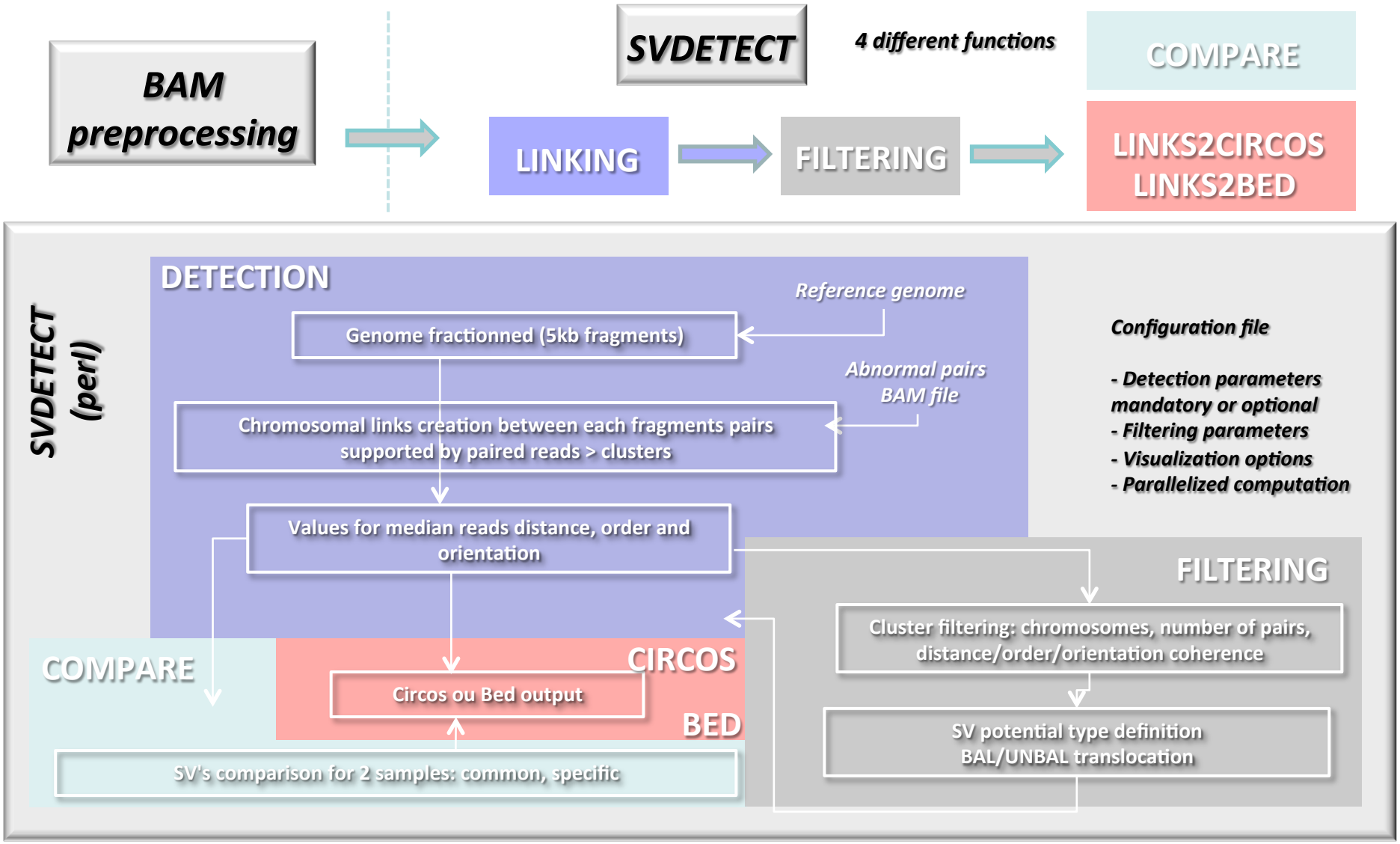
SVDetect - strategy

Paired-end/mate-pair based methods, such as SVDetect, evaluate reads distance, order, and orientation after alignment on a reference genome. Abnormal or discordants pairs are clustered to define their type of SV



Zeitouni et al 2010

SVDetect - implementation



Testing data

Biological data coming from neuroblastoma tumor
Two samples: Tumoral vs Control (Constitutional DNA)

Illumina GAII mate-pair sequencing, 50x50 bp, insert size of 3kb

Alignments files in binary format: BAM

SVDetect need 2 types of files as input:

- Bam file containing all apparied and aligned reads
  > « sample_mates.bam » & « reference_mates.bam »

- LEN file giving reference chromosome size
  > « hs18_chr5_chr11.len »

```
1       chr5            180857966
2       chr11           134452384
```

Let's start hands on SVDetect tutorial!