Structural variation detection and genotyping by resequencing

Yasmine Nooroya, Jean-Tristan Brandenburg, Maud Tenaillon & Johann Joets

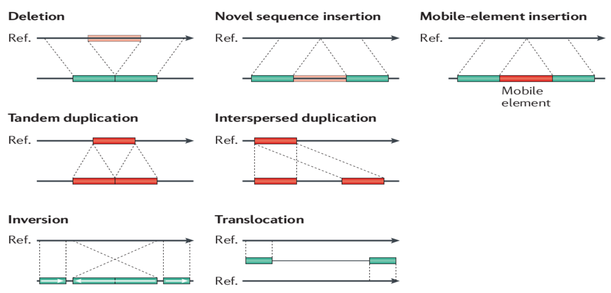
UMR de Génétique Quantitative et Evolution, Ferme du Moulon -91190 Gif-sur-Yvette, France

Contact Yasmine.Nooroya@moulon.inra.fr www http://moulon.inra.fr/

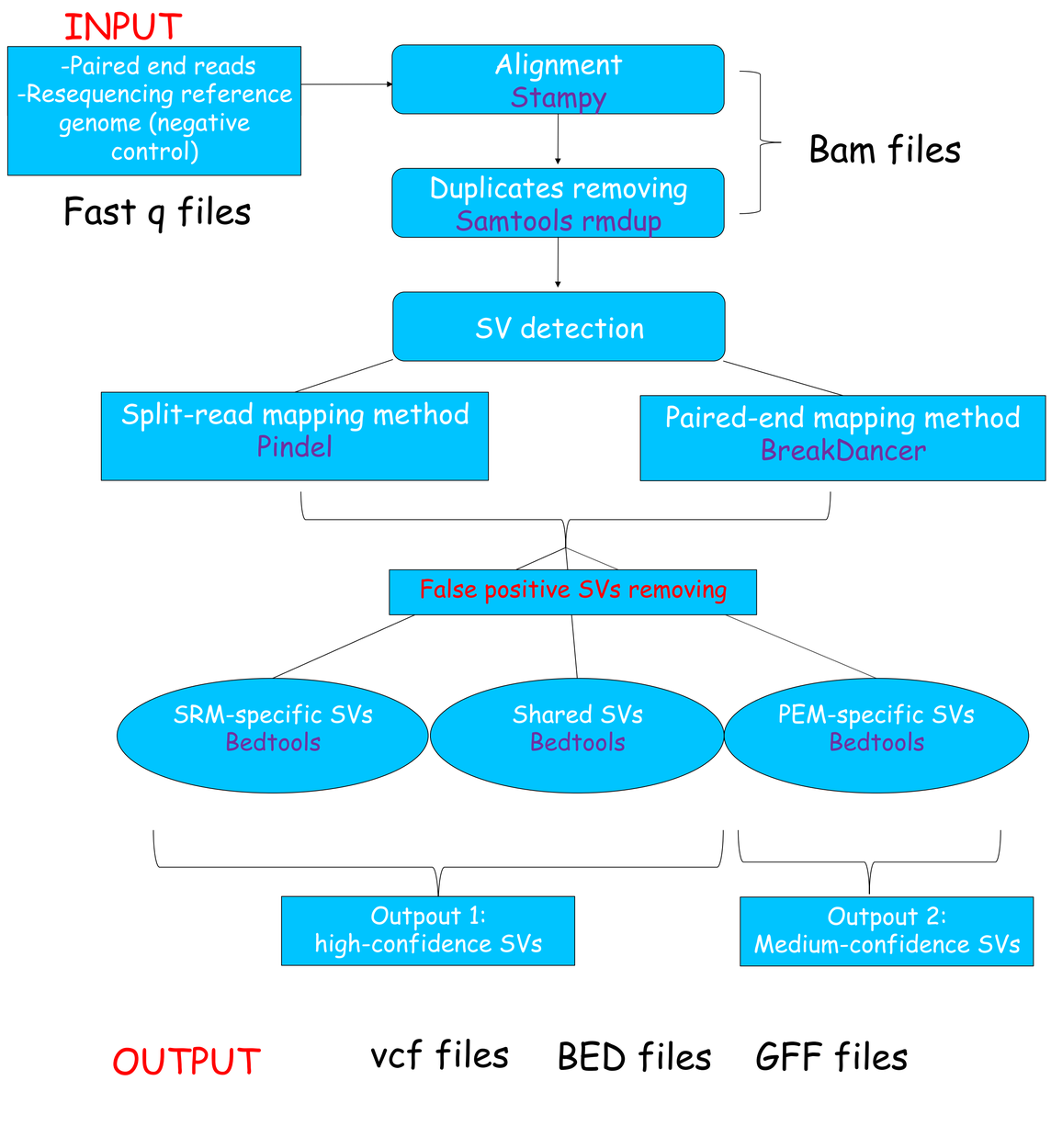
There is a growing interest for the role of genome structural variation in contributing to phenotypic variation within a species. Several characteristics of NGS (Next Generation Sequencing) data can be exploited for identification of structural variation. The challenge now is to discover the full extent of structural variation while controling false positive rate and to be able to genotype it routinely in order to understand its effects on, complex traits and evolution. We apply these methods in the frame on two BASC flagship 2 research programs.

Structural Variation

Structural variation is defined as a genomic variation affecting a sequence of more than 1 Kb in length including large insertions/ deletions, duplications, inversions, and other genome rearrangements.

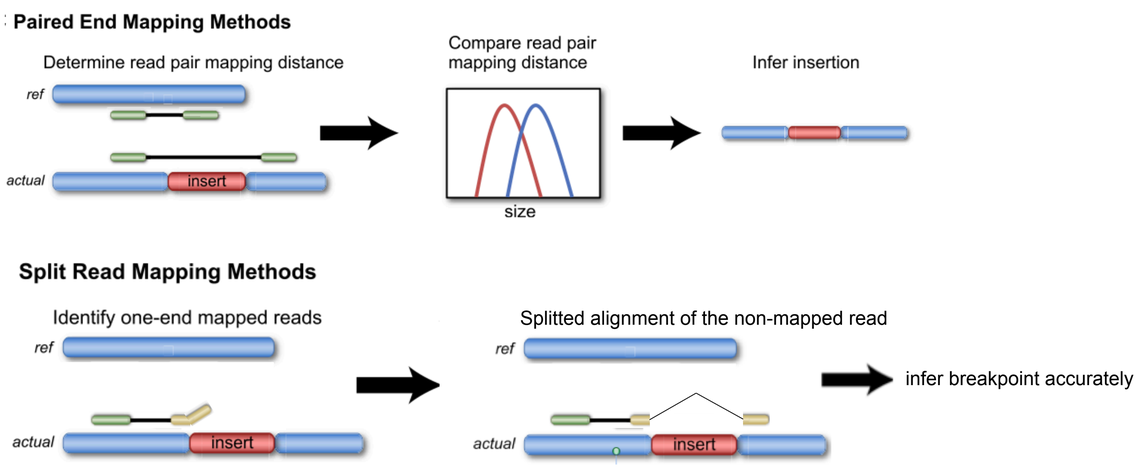


Bioinformatics approaches



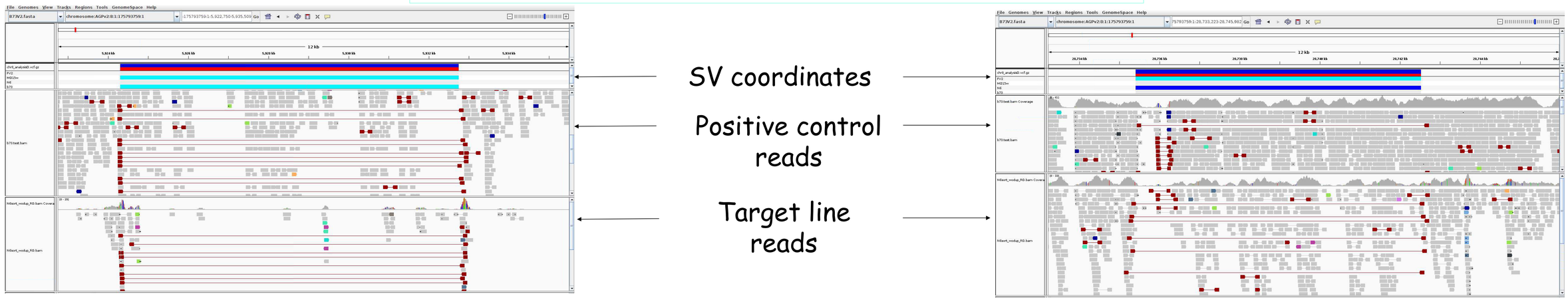
SV method detection

Two general NGS-based approaches are used to detect structural variation: the paired-end mapping method (PEM) and the split read mapping method (SRM).



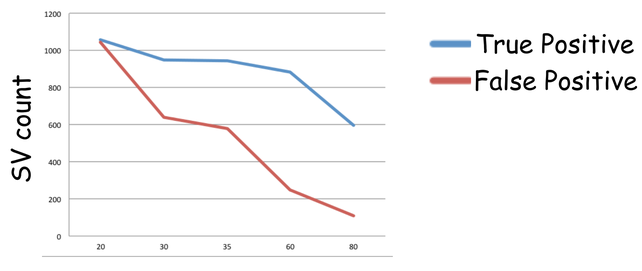
False positive SV control

The visualization of mapped reads is performed with IGV tool.



In this example, B73 is used as negative control to discard False positive SVs

Impact of software settings on false positive rate



Increasing mapping quality threshold value of breakDancer reduces FPs without severely impairing TPs discovery

Main origins of FPs are

- Reference genome assembly issues
- Detection method artefacts

➔ Using resequencing of the reference genome as a negative control help discarding FPs

Ongoing projects:

67 lines of Maize, Jean-Tristan Brandenburg and Maud Tenaillon, UMR GQE

30 strains of M Graminicola, Anne Genissel, BIOGER

1.Benjamin J. Raphael, Chapter 6 Structural variation and Medical Genomics. Plos Computational Biology –Dec 27 2012
2.Abel HJ,Duncavage EJ, Detection of structural DNA variation from next generation sequencing data/ a review of informatic approaches. Cancer genetic 432-40 – Dec 2013
3.Can Alkan, Bradley P.Coe & Evan E.Eicher , Genome structural variation discovery and genotyping. Nature Reviews Genetics 12,363-376 –May 2011
4. Genome sequencing and structural variation. http://mi.fu-berlin/wiki/pub/ABI/GenomicsLecture10Materiels/structural-variation.pdf
5. Ye k, Schulz MH, Long Q, Apweiler R, Ning Z, Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics 2865-71 –Nov 2009
6. Chen K, Wallis JW, McLellan MD, BreakDancer: an algorithm for high-resolution mapping of genomc structural variation. Nature Methods 6, 667-681 – August 2009