

VARIANTES STRUCTURALES

1. petits indels
2. copy number variation
3. réarrangements

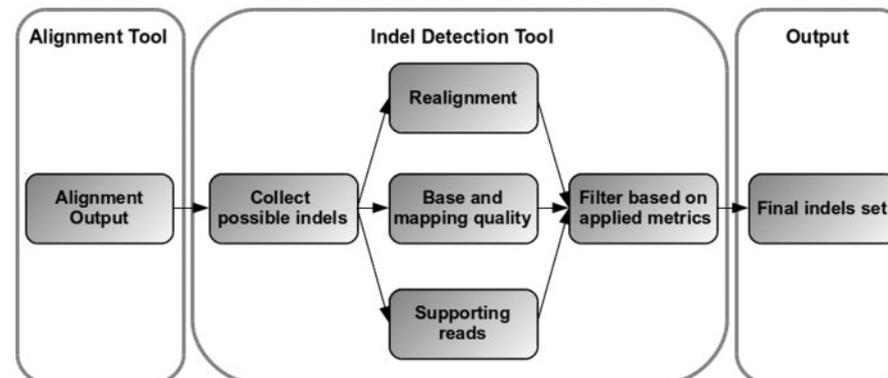
Petits indels

petits indels (1..10 ou ..20 pb) — variantes génomiques 1/8 aussi fréquentes que les SNPs (mais presque même longueur combinée entre SNP et indels)

Table 1. Summary of small INDELS identified from 98 million re-sequencing traces

Traces screened	98,350,511
Bases analyzed (trimmed)	42,000,606,219
Reference human genome (hg18)	3,107,677,273
Cumulative coverage	13.5×
Variants identified	3,400,787
Unique (nonredundant) variants	1,955,656
Insertions	978,721
Deletions	976,935
Total bases affected	11,909,159

- on infère les petits trous à partir des alignements des lectures
- il faut refaire l'alignement pour identifier la variante en commun



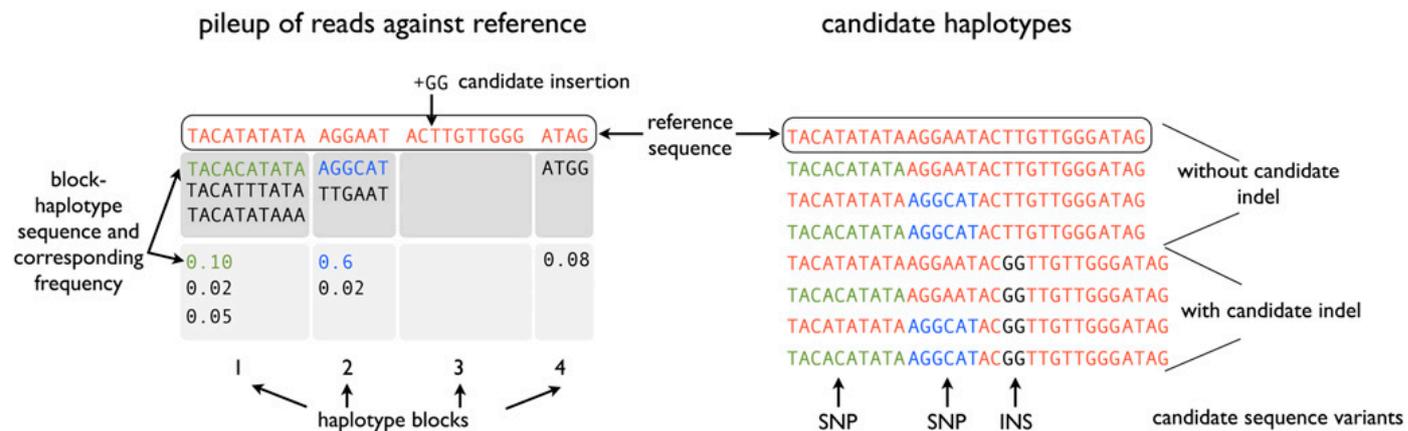
Mills & al *Genome Research* 21 :830 (2011) ; Neuman, Isakov & Shomron *Briefings in Bioinformatics* 14 :46 (2012)

DINDEL

travailler avec des candidats de haplotypes + alignement statistique entre haplotype et lectures

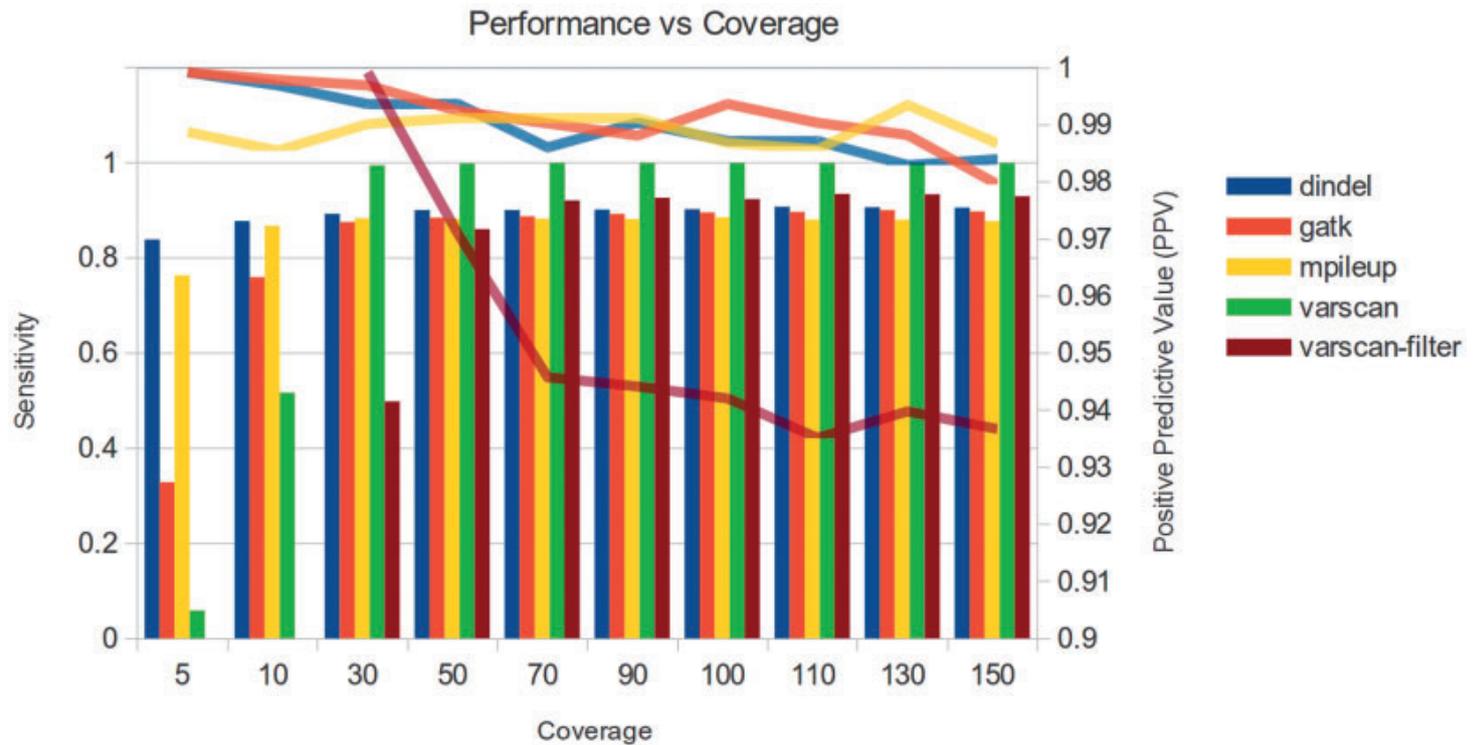
haplotypes : combiner les lectures en blocs de haplotypes

considérer toute combinaison de blocs fréquents + variantes proposées à l'entrée



(≤ 8 haplotypes à la fois)

Détection et couverture



PPV=positive predictive value

$$\text{sensitivity} = \frac{TP}{TP + FN}$$

$$\text{PPV} = \frac{TP}{TP + FP}$$

Neuman, Isakov & Shomron *Briefings in Bioinformatics* 14 :46 (2012)

Meilleures pratiques

Tool name	Advantages	Limitations
GATK	Highly supported with good overall performance	Low sensitivity at very low coverage (<10×; can be improved by less stringent parameters)
Dindel	Best performance at low coverage	Only suitable for Illumina data analysis and has long running time
SAMtools mpileup	High PPV and simple use	Lowest sensitivity at high coverage (>50×)
VarScan	High sensitivity at intermediate/high coverage (>30×) and simple use	Low PPV at default parameter settings and low sensitivity at low coverage (<30×)

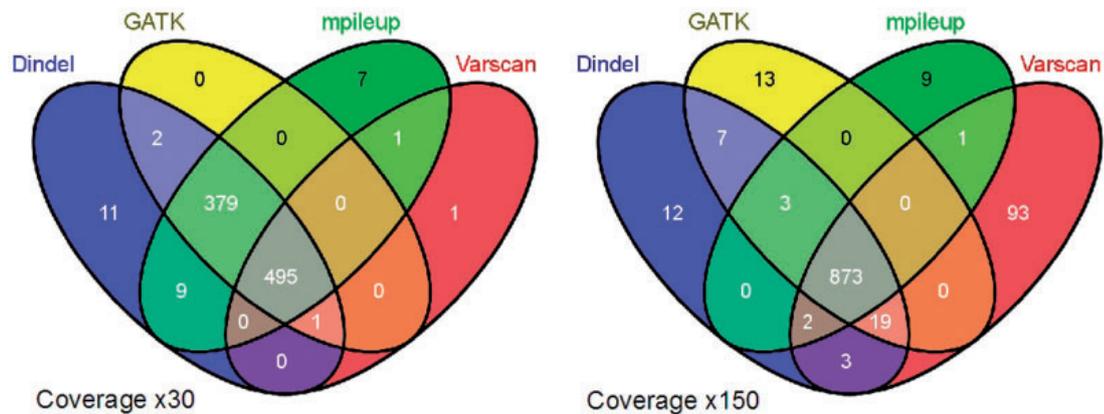
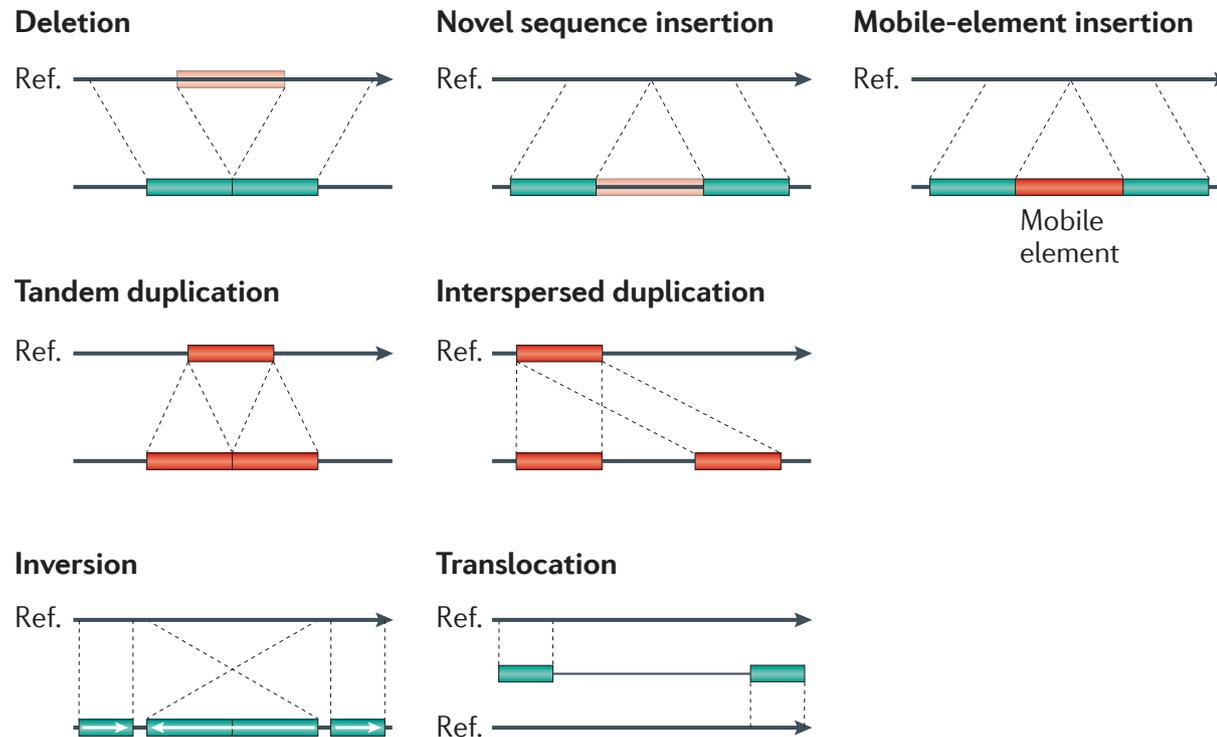


Figure 5: Venn diagram depicting the number of indels found for each software with 30× and 150× coverage and read length 72. Inclusion of indels called in any of the software results in a decrease in PPV with only a mild sensitivity improvement. Inclusion of indels supported by at least two software results in a sensitivity improvement for some of the software and a significant increase in PPV, crucial in high coverage data.

Variation structurale

classes principales : suppression, insertion, duplication, inversion/translocation



Alkan, Coe & Eichler *Nature Reviews Genetics* 12 :363 (2011)

Signatures de SV

SV classes	Read pair	Read depth	Split read	Assembly
Deletion				
Novel sequence insertion		Not applicable		
Mobile element insertion		Not applicable		
Inversion		Not applicable		
Interspersed duplication				
Tandem duplication				

Alkan, Coe & Eichler *Nature Reviews Genetics* 12 :363 (2011)

Breakdancer

détection de SV à partir des paires discordantes :

1. paire discordante : orientation incompatible et/ou distance surprenante
2. régions avec trop de paires discordantes
3. liaisons : régions liées par ≥ 2 paires discordantes

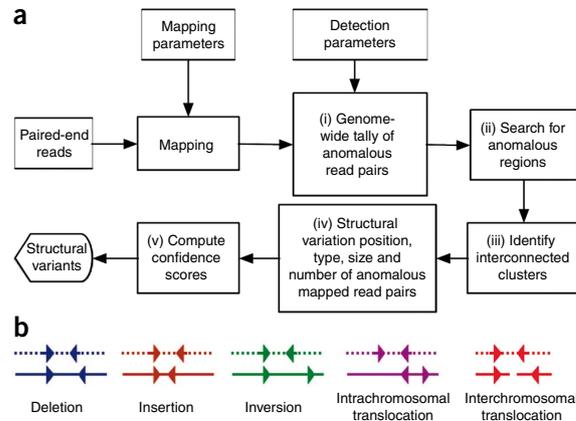
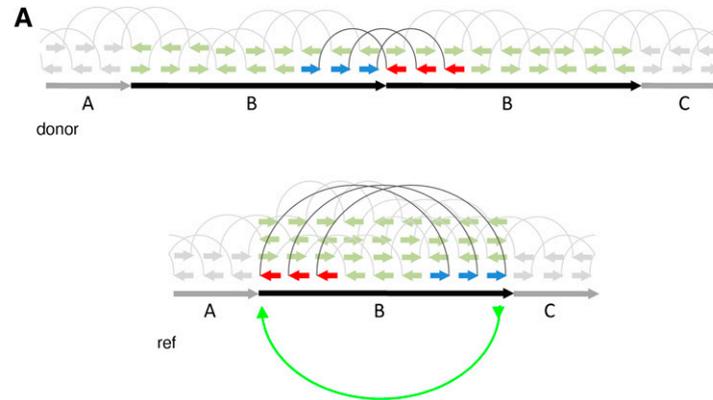


Figure 1 | Overview of BreakDancer algorithm. (a) The workflow. (b) Anomalous read pairs recognized by BreakDancerMax. A pair of arrows represents the location and the orientation of a read pair. A dotted line represents a chromosome in the analyzed genome. A solid line represents a chromosome in the reference genome.

Duplications

duplications : vérifier paires et couverture par lectures

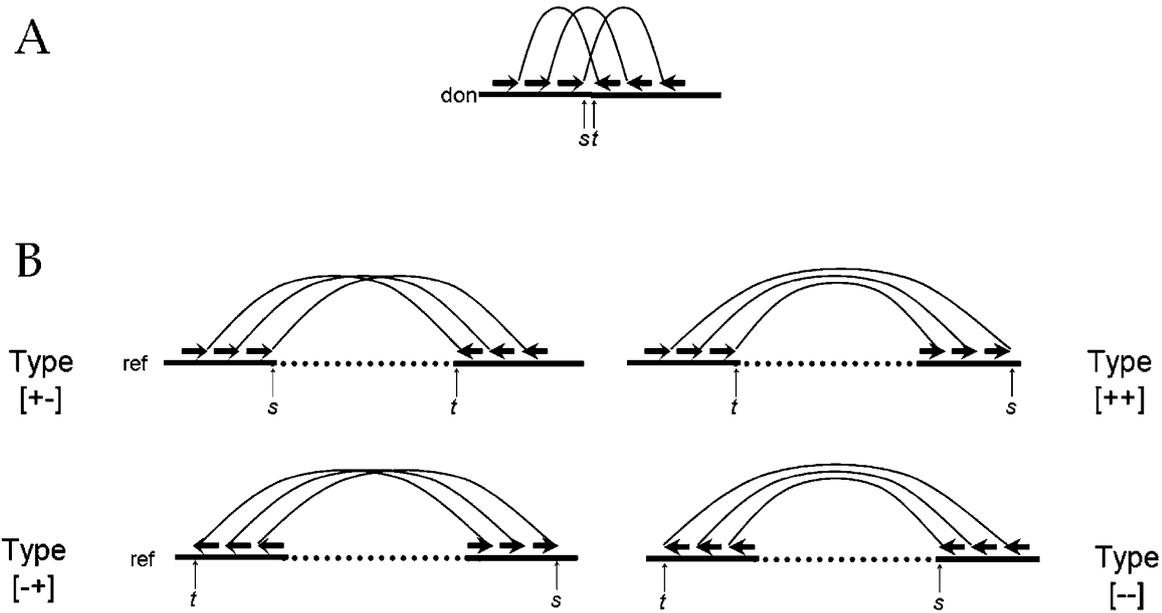


1. mapping : identifier des lectures trop fréquentes et discordantes (mauvaise orientation ou distance $> \mu + 3\sigma$)

Medvedev & al. *Genome Research* 20 :1613 (2010)

CNVer

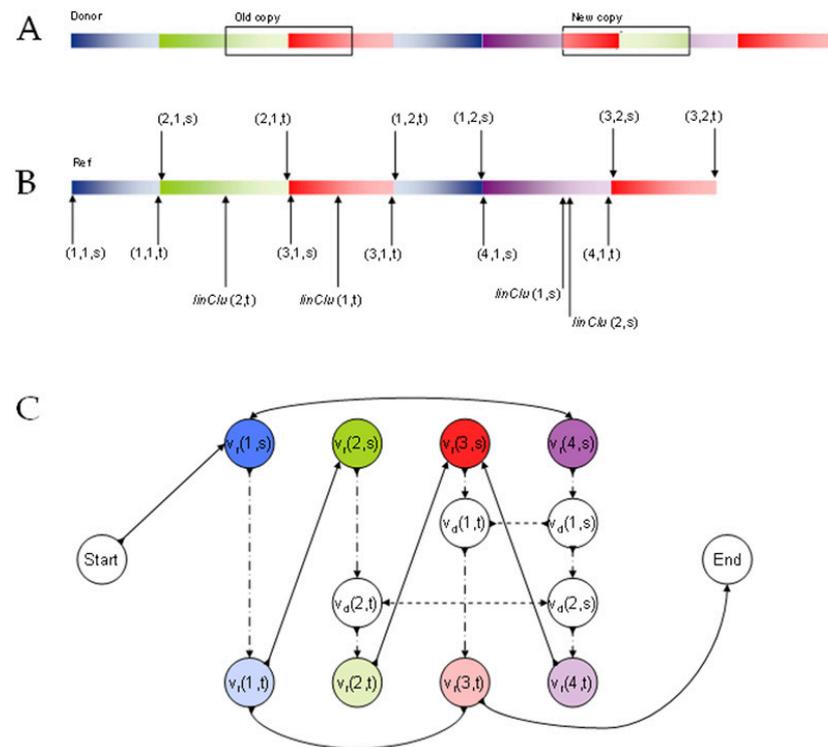
2. groupage de paires discordantes, calculer positions (s, t) sur la référence pour la partie intérieure



Medvedev & al. *Genome Research* 20 :1613 (2010)

CNVer (cont.)

3. “Donor graph” : représenter la référence par blocs uniques (compression de régions dupliquées), y ajouter les clusters de paires discordantes, arêtes bi-orientées



4. optimisation de flot dans le réseau (sans assemblage)