

# Recherche de variants génomiques en oncologie clinique

Avec des diapos, données & scripts R de:  
Yannick Boursin, IGR  
Bastien Job, IGR

# Technologies de recherche de variants

- Sanger
  - Toujours utilisé en consultation de génétique
- SNP arrays
  - ~1M SNP (recherche GWAS ou « 23&Me »)
- Panel de gènes
  - Une série d'exons d'intérêt (gènes de cancer= 100kb)
- WES (Exome)
  - Tous les exons du génome (30 Mb)
- WGS
  - Whole genome (3 Gb)

NGS

# Génétique constitutionnelle

At hospital



Blood sample

Sequence  
gene panel



Look for  
specific  
alteration  
(BRCA)

Research



Genotype  
or  
Sequence



Compare  
disease and  
healthy  
cohorts

GWAS studies, 1000 Genome Project...

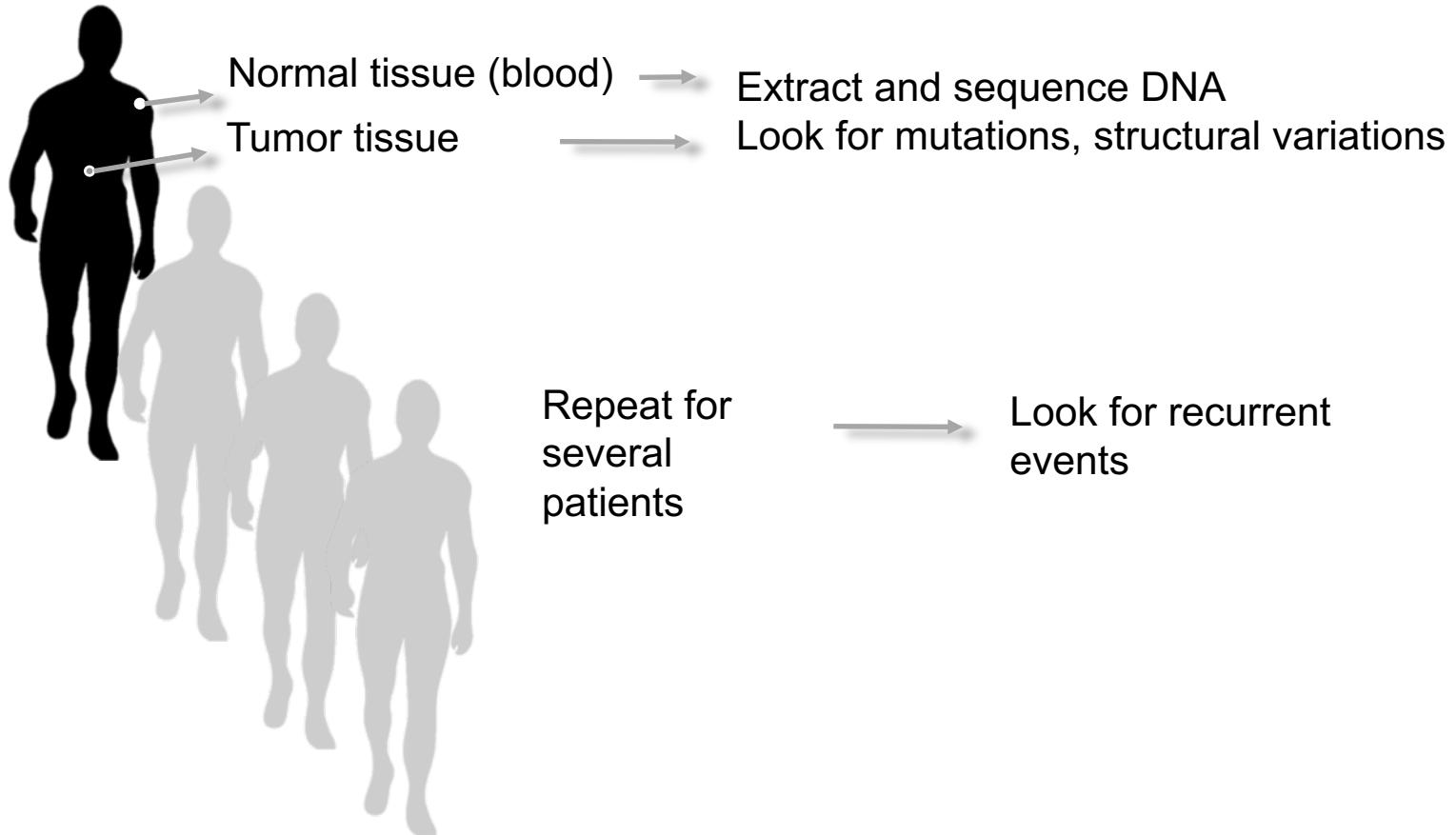
# NGS dans le diagnostic de génétique familiale

Panel de gènes pour cibler:

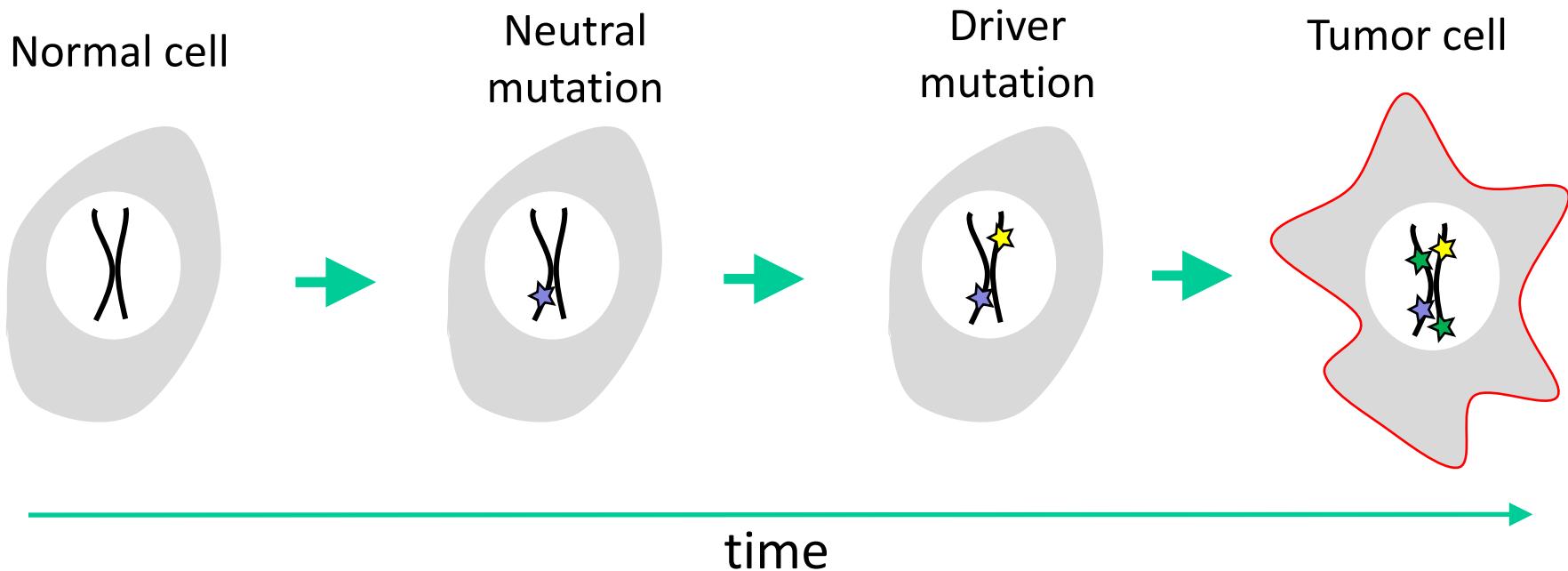
- BRCA1/2 (breast/ovary cancer)
- XPC, XPV.. (melanoma)
- ERCC1 (colorectal cancer)

# En génétique somatique

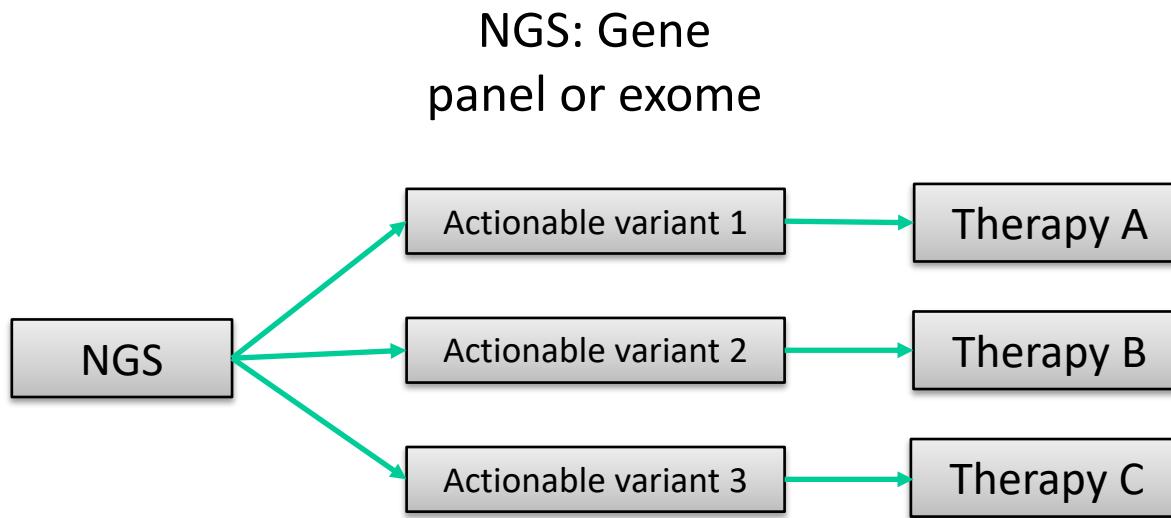
Finding somatic mutations in the tumor genome



# Somatic mutations: clonal evolution of cancer

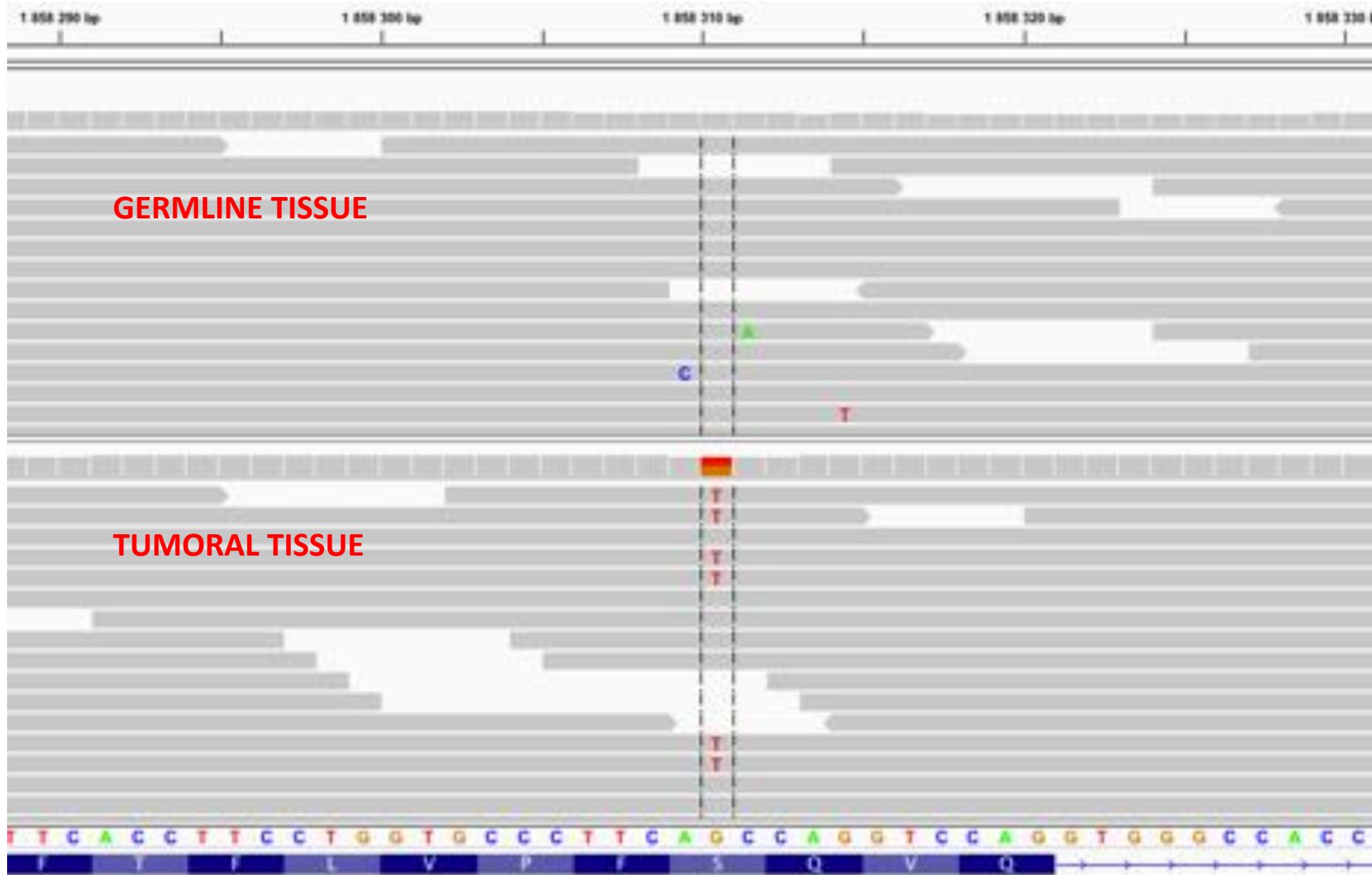


# NGS for precision medicine

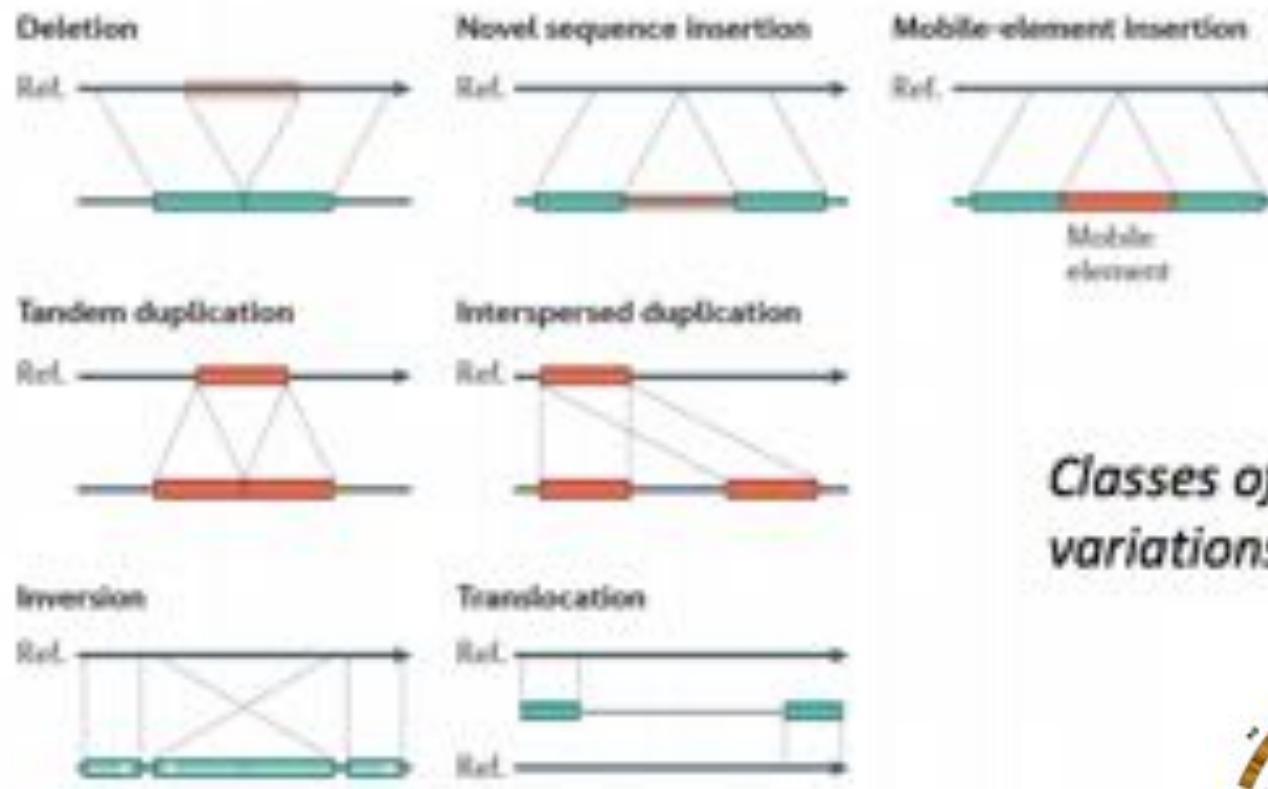


- Clinical trials: MOSCATO (GR), SAFIR (GR), SHIVA (Curie), ...
- Ipilimumab (anti-CTLA4), Nivolumab (anti-PD1), Trastuzumab (anti-HER2), Cetuximab (anti-EGFR)

# Les mutations somatiques

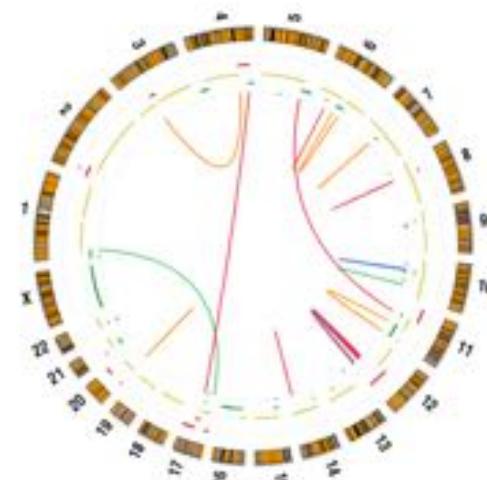


# Variants structuraux



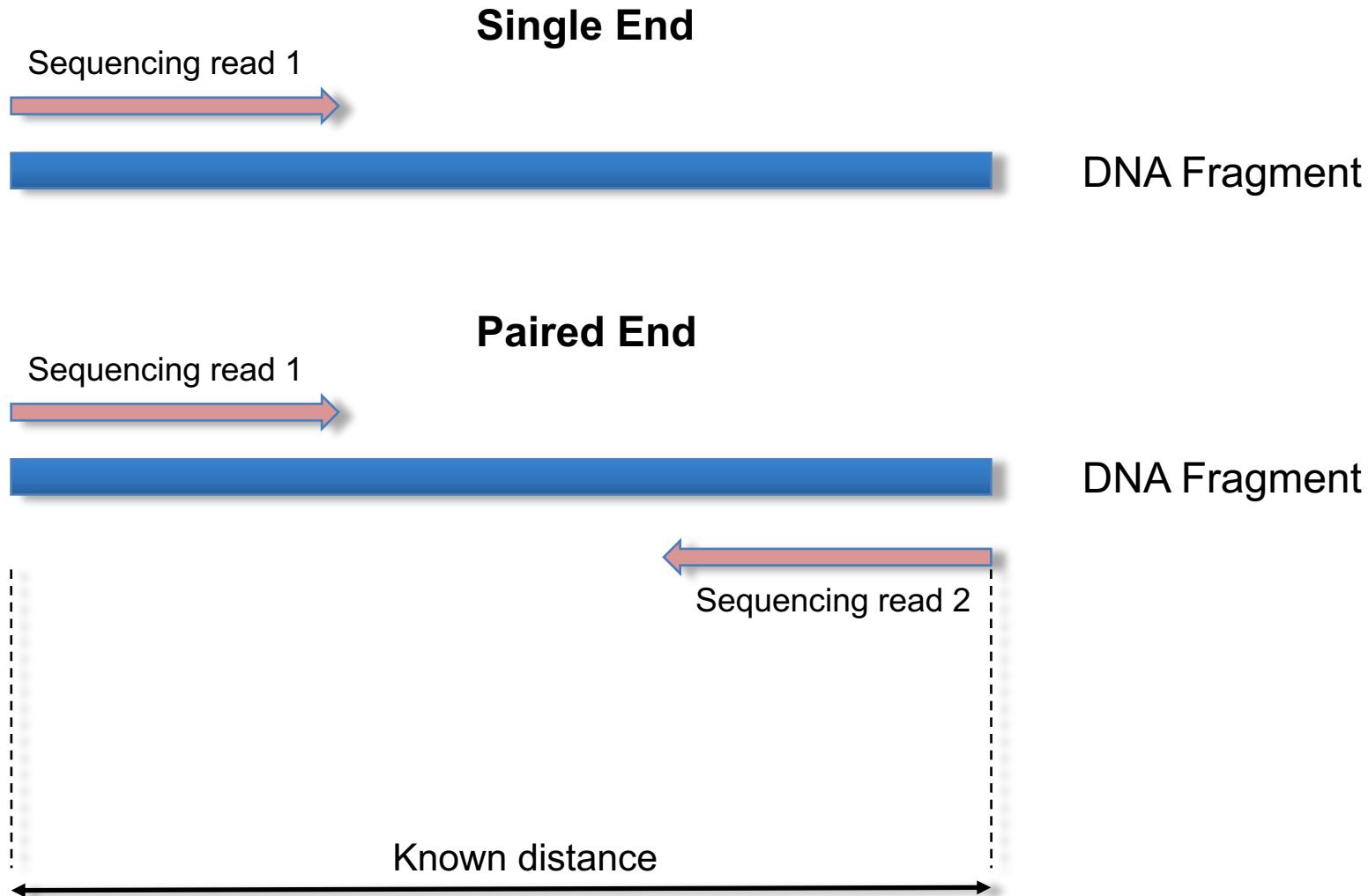
*Classes of structural variations*

Alkan et al 2012



# Intérêt du séquençage paired-end:

## Résolution des repeats et des variants structuraux



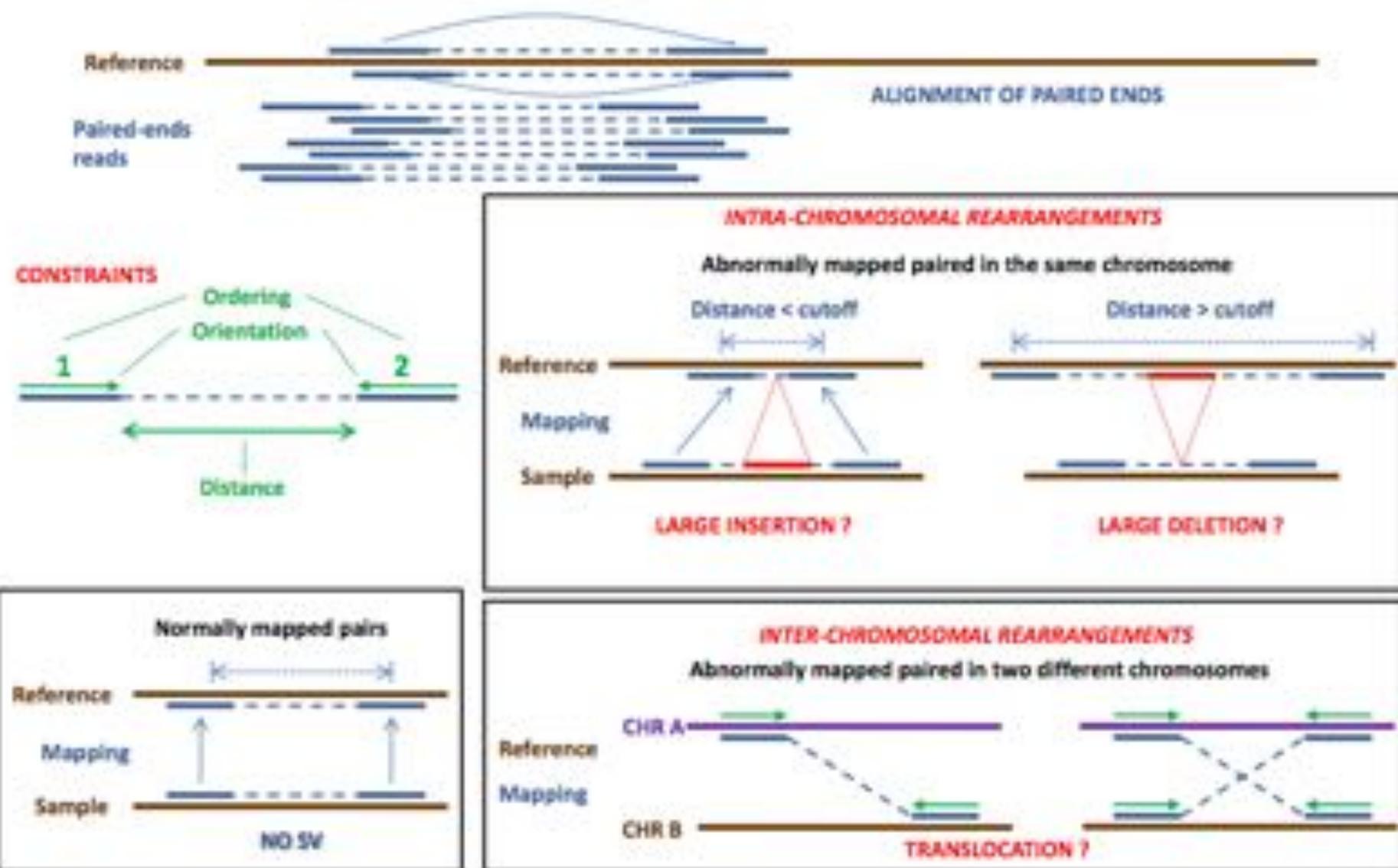
# Exemple

- Single-end alignment – repeated sequence

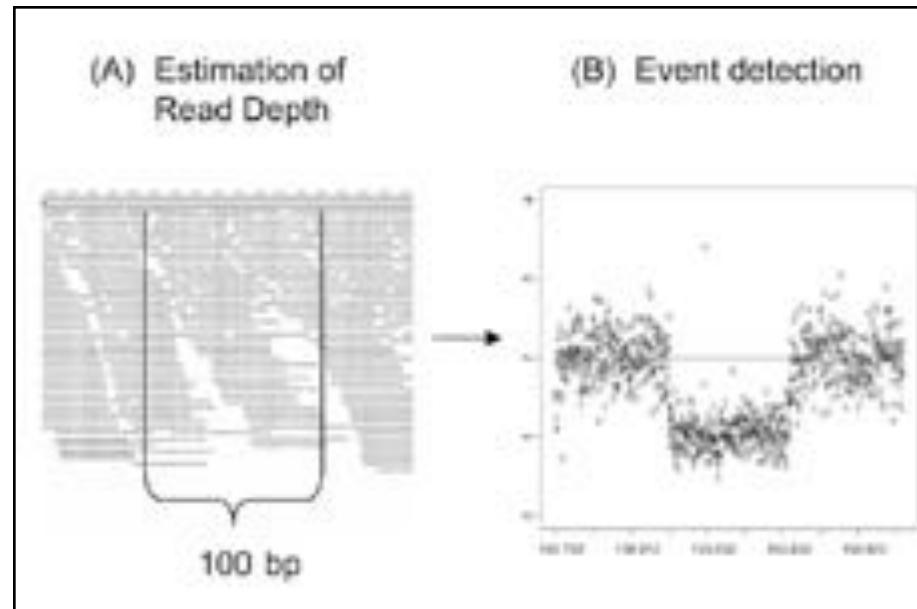
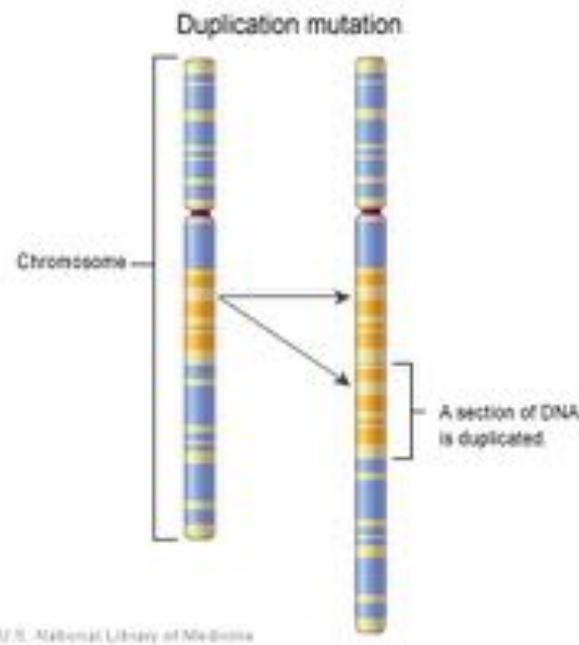


- Paired-end alignment – unique sequence





# Recherche de CNV (copy number variations)



- Attempts to infer variations in copy number from the **local read depth**.
- A strong GC% debiasing is required

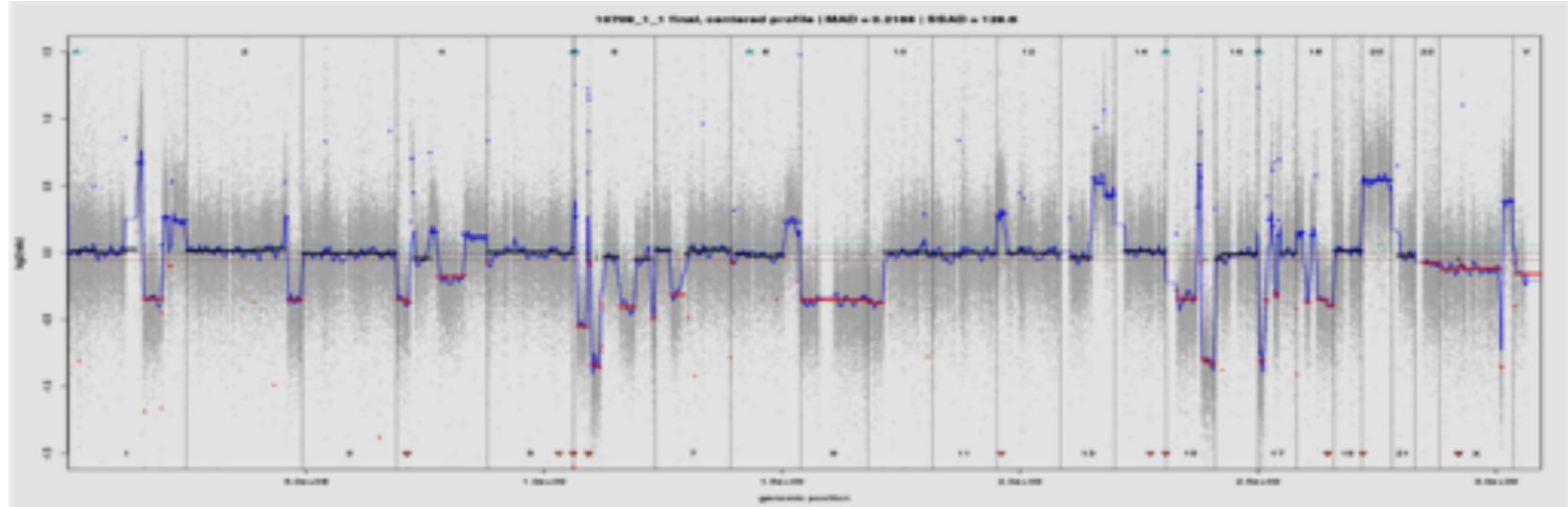
Voir cours Bastien Job

Yoon, 2009

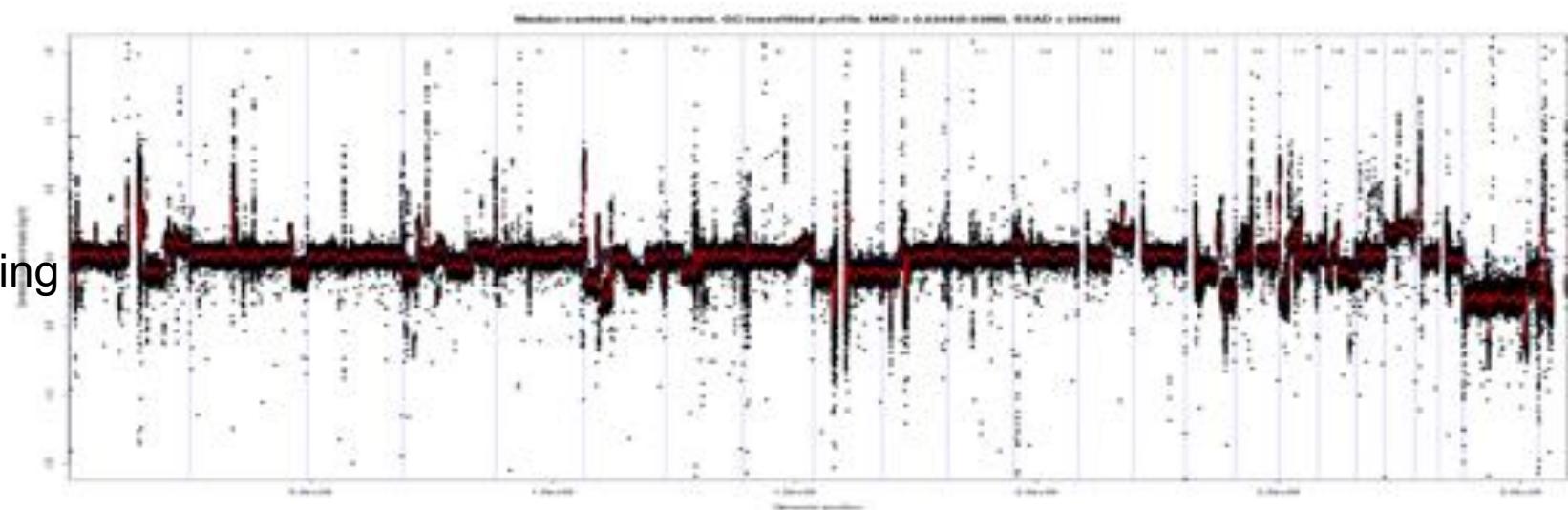
# NGS vs CGH

CGH

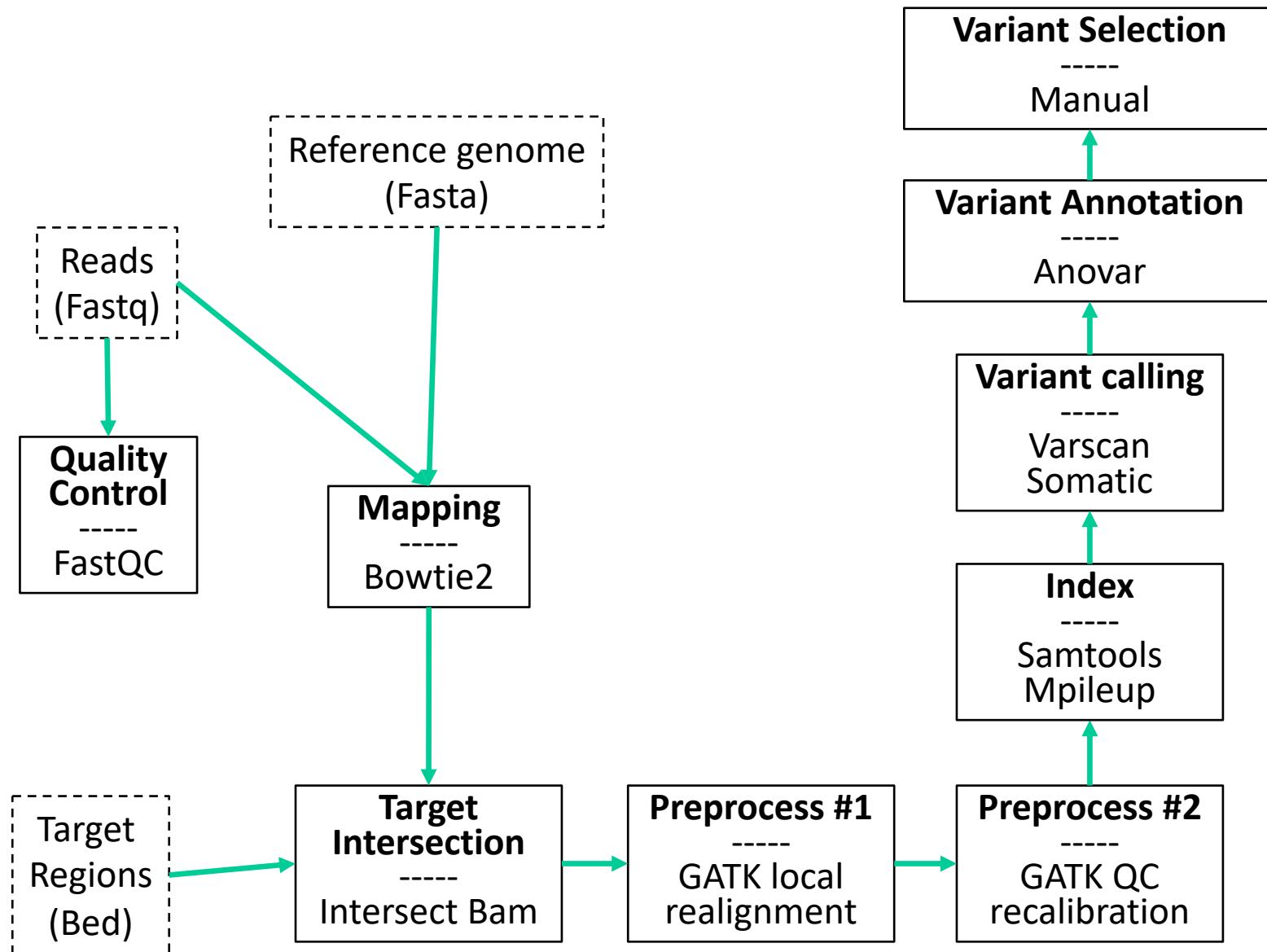
comparative  
genome  
hybridization



Sequencing

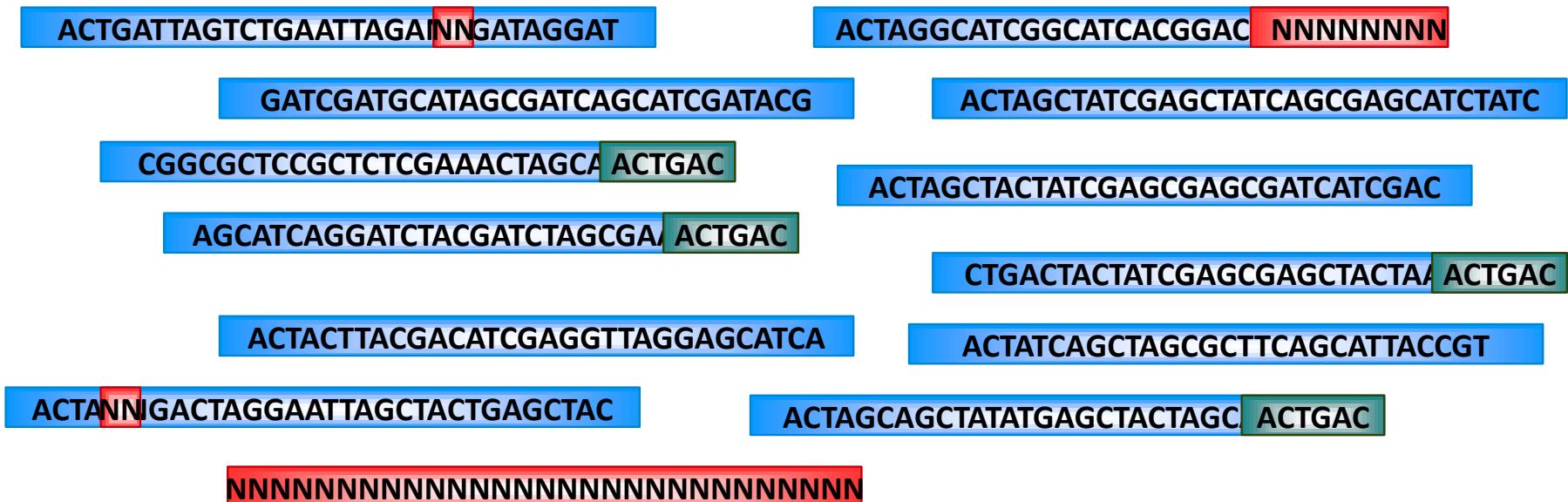


# Un pipeline « variants »



# Quality control

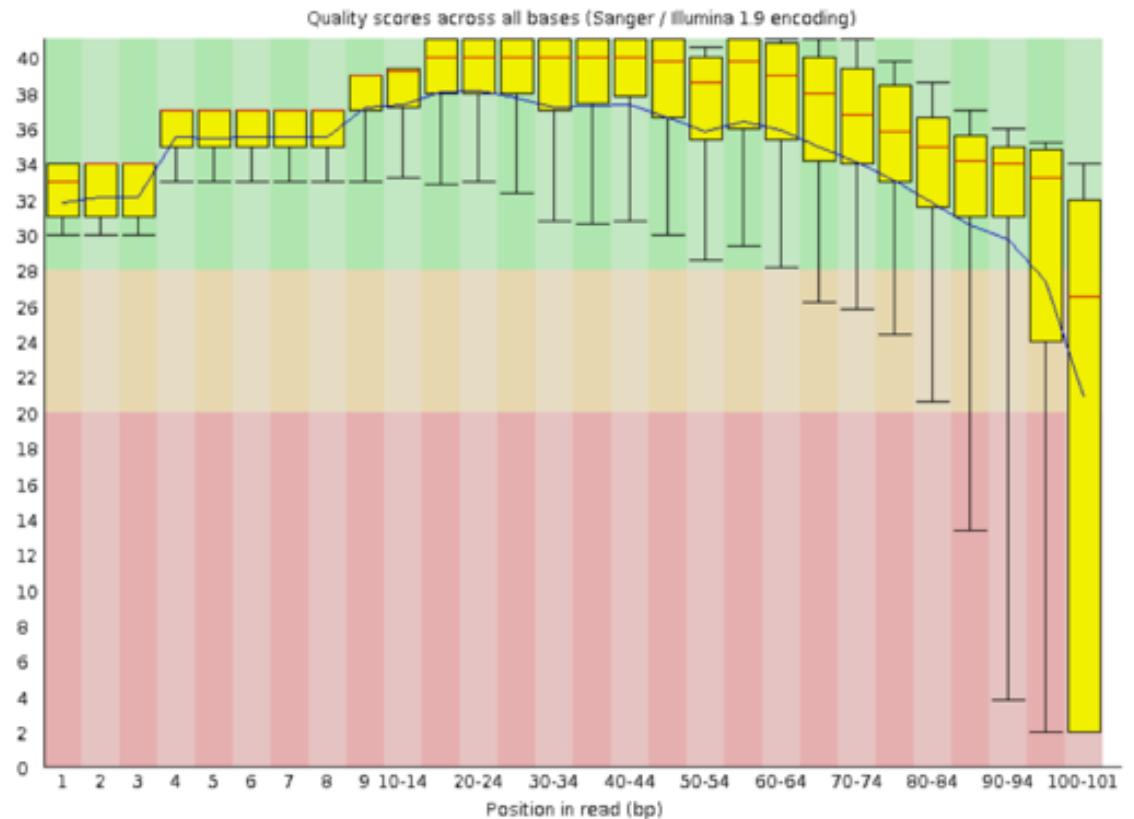
A first Quality Control of raw reads is mandatory and can be established according to the application ('N', low quality sequence, adapter sequences, barcode, contamination, etc.)



# FastQC Metrics

- Look at the different metrics for both reads
- **Problem:** the per base sequence quality of the Read2 are quite low towards the end

✖ Per base sequence quality



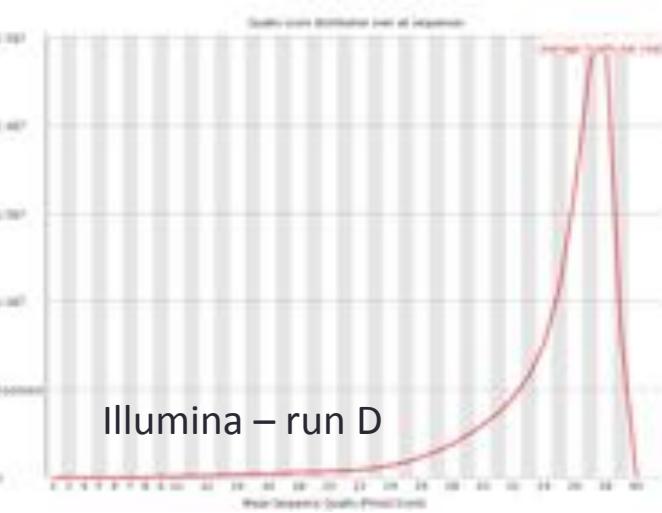
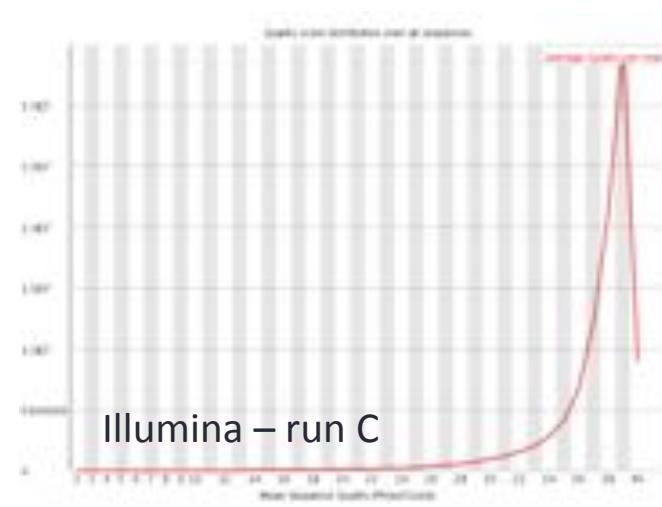
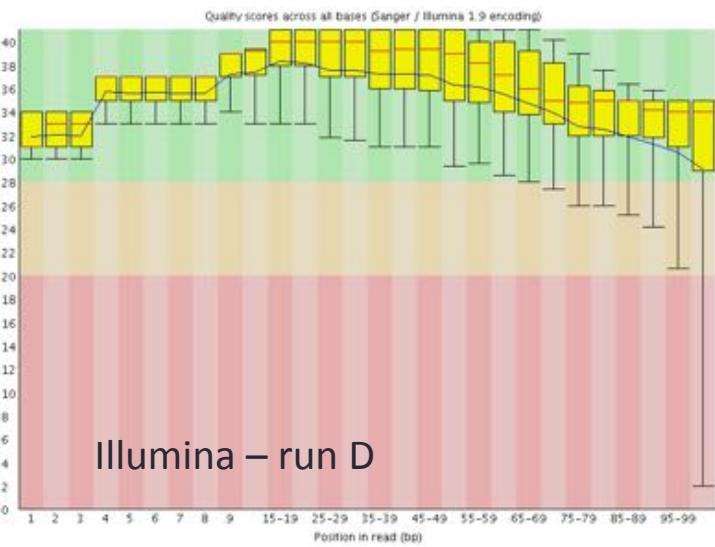
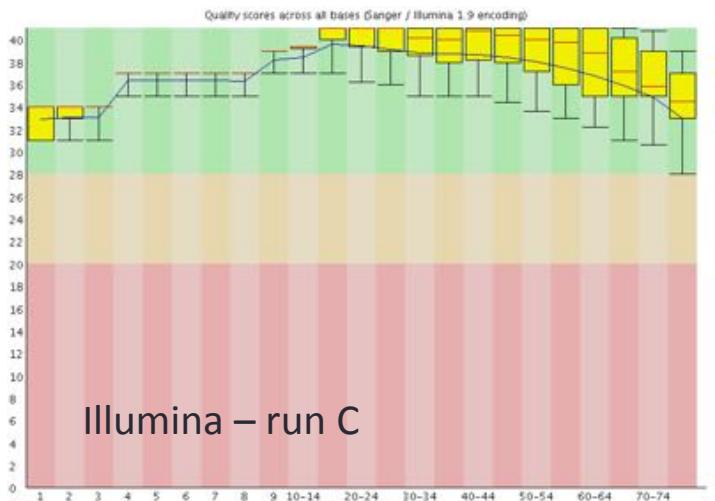
## Solution:

Trim low quality bases from the 3' end of the reads

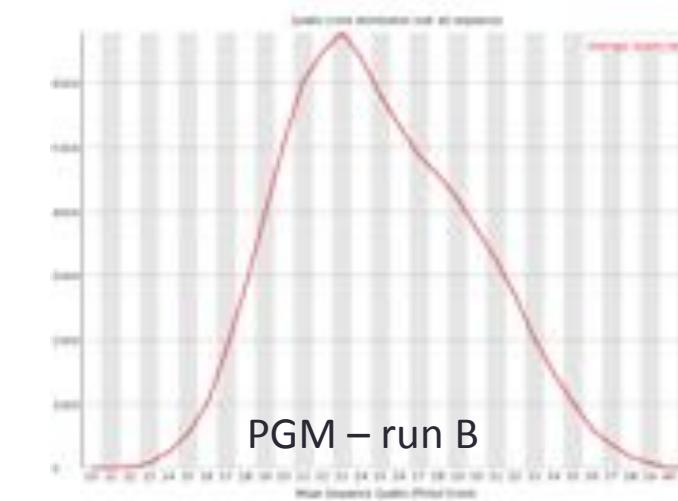
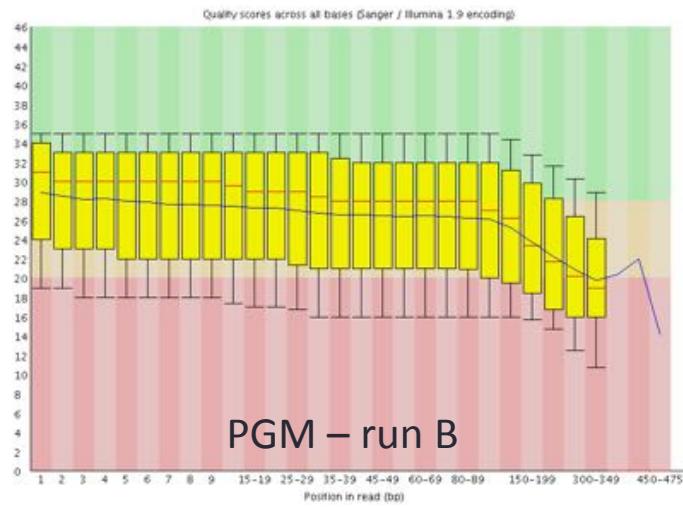
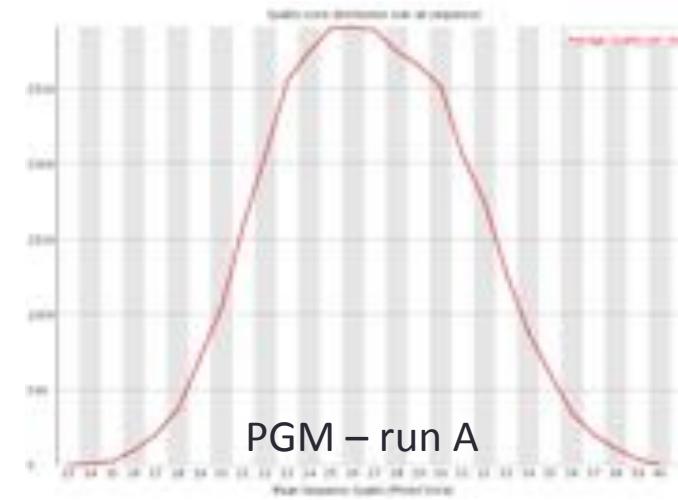
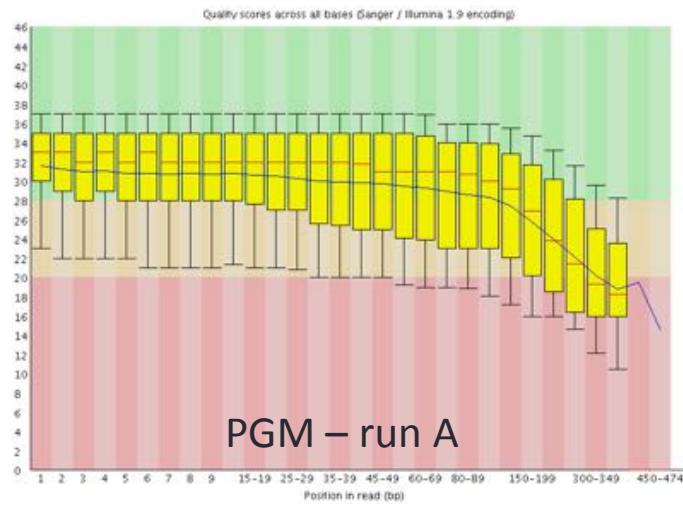
➤ Higher confidence in the sequenced information

(Trimmomatic)

# Illumina

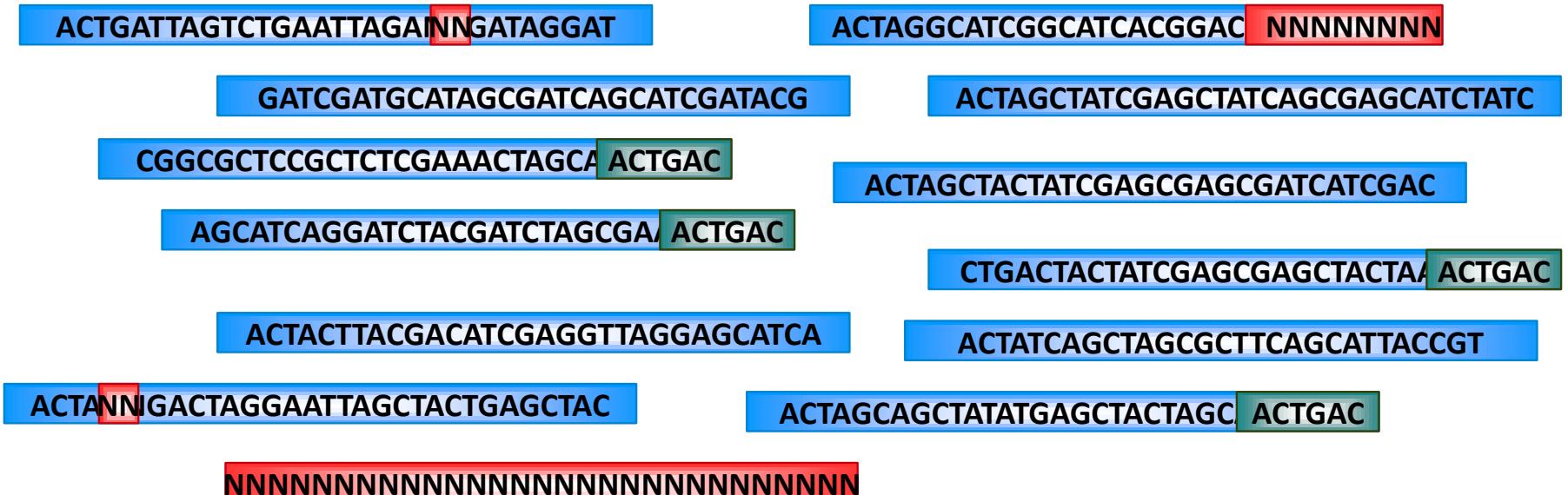


# PGM

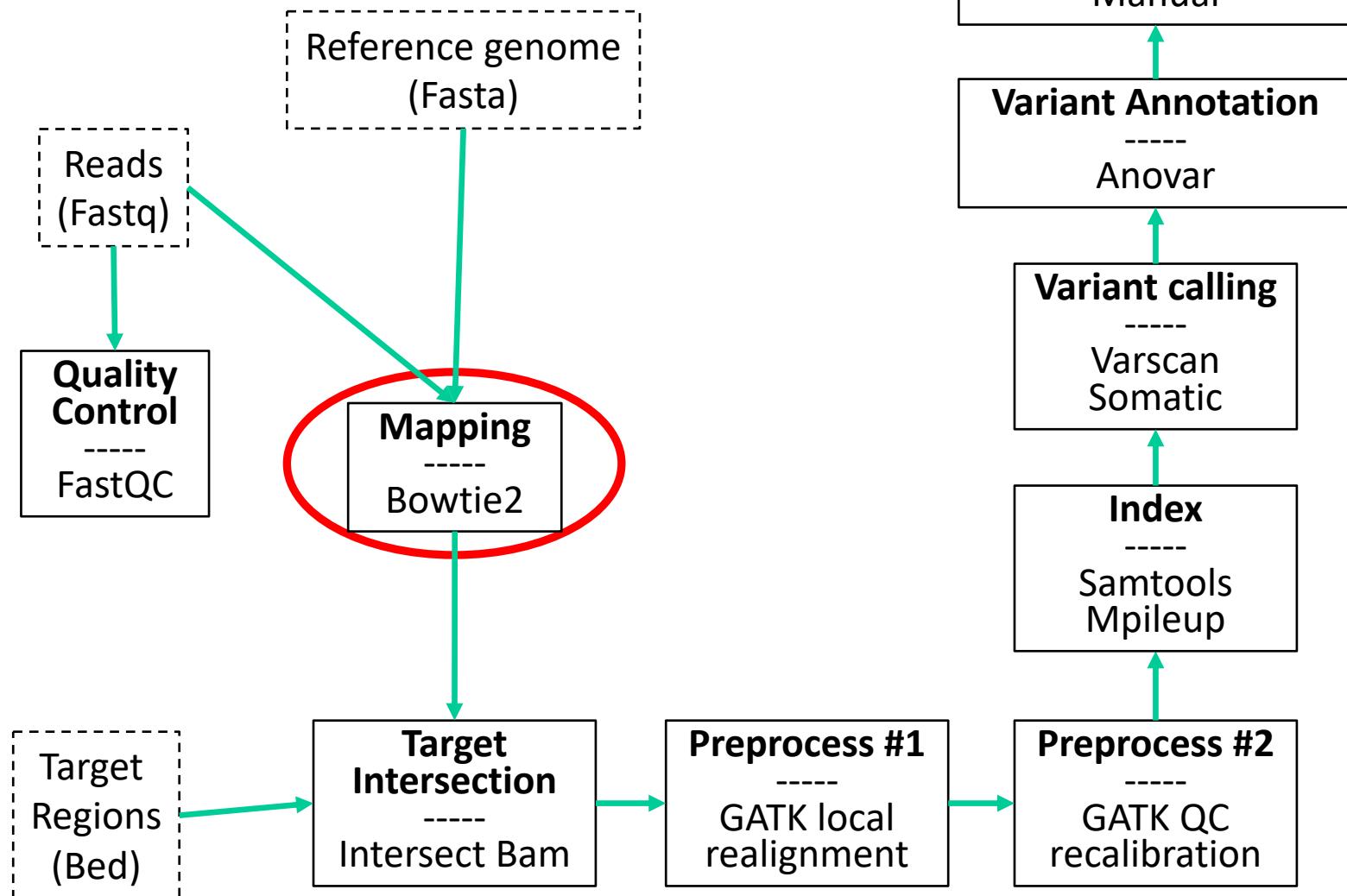


# Trimming and discarding low quality reads...

Popular Software=trimmomatic, cutadapt



Processed reads: blue parts are to be kept, green and red parts to be removed



# Alignment des reads sur le génome de référence



# Algorithme en $O(mn)$

ACGTTACCGAATCGATCAAGTCGA  
TAC



OK pour 1 read:  $O(3.10^9 \times 100)$   
Mais pour  $1^8$  reads???

« supercalifragilis-ticexpialidocious »

←   →

Préfixe                                      Suffixe

## Suffix array

“GOOGOL”

Tableau trié de tous les suffixes  
d'une chaîne de caractères

0 GOOGOL\$		6 \$	
1 OOGOL\$		3 GOL\$	
2 OGOL\$		0 GOOGOL\$	
3 GOL\$	→	5 L\$	→ (6,3,0,5,2,4,1)
4 OL\$		2 OGOL\$	
5 L\$		4 OL\$	
6 \$		1 OOGOL\$	

Propriété: toutes les occurrences d'une même chaîne  
sont regroupées.

# Suffix arrays

Exemple: trouver la chaîne **GO**

0 GOOGOL\$		6 \$
1 OOGOL\$		3 <b>G</b> O L\$
2 OGOL\$		0 <b>GO</b> OGOL\$
3 GOL\$	→	5 L\$
4 OL\$		2 OGOL\$
5 L\$		4 OL\$
6 \$		1 OOGOL\$

# Most popular aligners for variant analysis

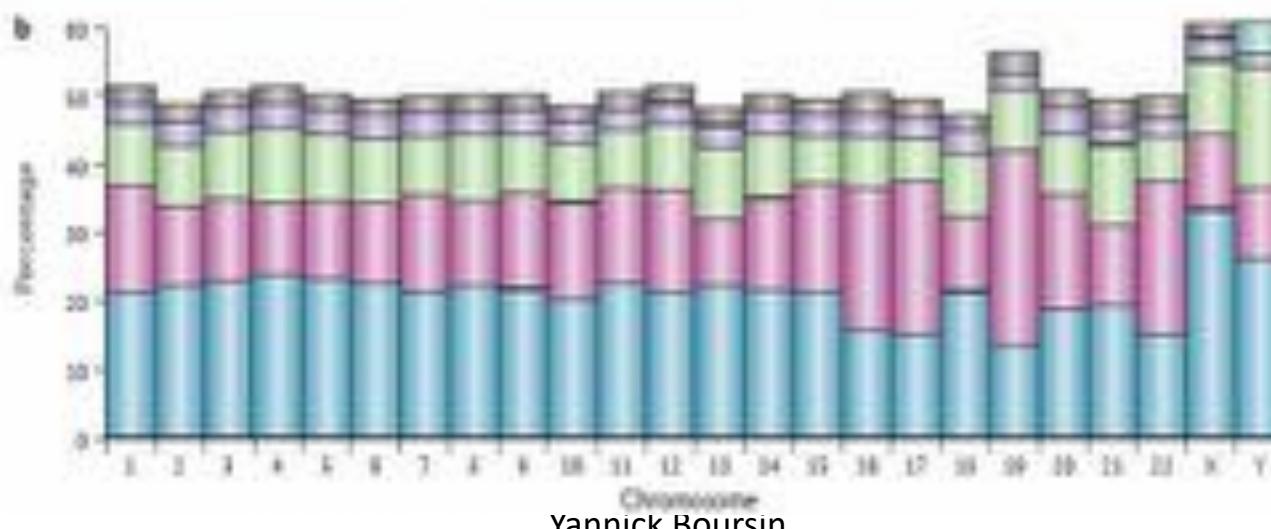
(support mismatched, gapped, paired-end alignment)

- BWA
  - Li H. and Durbin R. (2009)
- Bowtie2
  - Langmead B, Salzberg S (2012)

# Alignment key parameters - Repeats

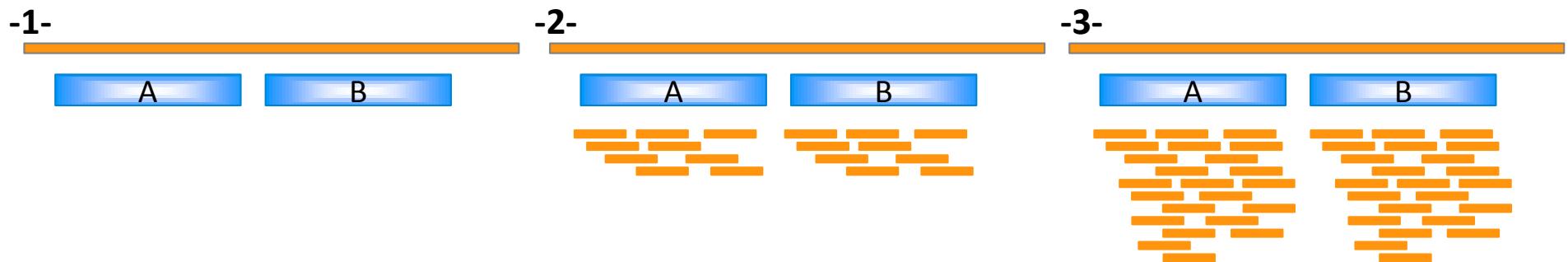
Approximately **50%** of the human genome is comprised of repeats

Repeat class	Repeat type	Number (bp/2L)	Ctg.	Length (bp)
Microsatellite, minisatellite or satellite	Tandem	425,518	1%	2-100
LINE	Interspersed	1,747,375	11%	100-1000
Alu	Interspersed	463,778	2%	200-2000
LTR retrotransposon	Interspersed	718,125	2%	300-5,000
LINE	Interspersed	1,508,345	21%	500-10,000
rDNA (18S, 28S, 5.8S and 25S)	Tandem	601	0.01%	1,000-41,000
Segmental duplications and other classes	Tandem or interspersed	2,779	0.2%	1,000-100,000



# Alignment key parameters – Repeats – 3 strategies

- 1- Report only unique alignment
- 2- Report best alignments and randomly assign reads across equally good loci
- 3- Report all (best) alignments



Treangen T.J. and Salzberg S.L. 2012. Nature review Genetics 13, 36-46

# Reads alignés: le format BAM/SAM



# Format BAM/SAM

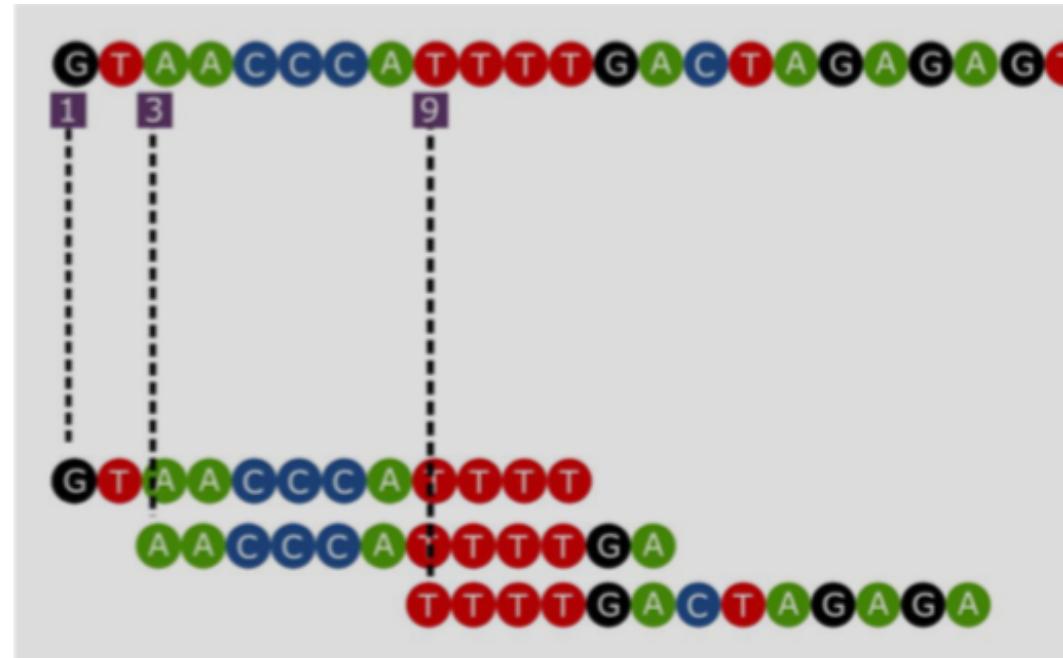
Contient les séquences alignées sur le génome

Concept:

chr7 1324324 ACGTGCCTTCGCGT

chr8 1424324 GCGTGATGCGTAAG

chr8 1724354 GTATGTTATATGTA



# Format SAM/BAM

- A real SAM:

```
@RG  ID:group1  SM:1425_CD34  PL:ILLUMINA  LB:lib1 PU:unit1
@PG  ID:bwa  PN:bwa  VN:0.7.12-r1039 CL:bwa mem -M -t 2 -A 2 -E 1 -R @RG\tID:group1\tSM:1425_CD34\tPL:ILLUMINA\tLB:lib1\tPU:unit1 /root/myd
ERR166338.13782800  83  chr13  32890449  60  101M  =  32890343  -207  GGGACTGAATTAGAACAAATTTCAGCGCTT
ERR166338.13782800  163  chr13  32890343  60  75M  =  32890449  207  CACTAGCCACGTTCGAGTGCTTAATGTGGCTAGTGGC
ERR166338.26716588  99  chr13  32890406  60  101M  =  32890553  222  AATGTTCCCACCTCACAGTAAGCTGTTACCGTTCCAG
ERR166338.26716588  147  chr13  32890553  60  75M  =  32890406  -222  TTGCAGACTTACCAAGCATTGGAGGAATATCGTA
ERR166338.27259961  99  chr13  32890496  60  101M  =  32890558  137  ACCTCAGTCACATAATAAGGAATGCATCCCTGTGTAAG
ERR166338.27259961  147  chr13  32890558  60  75M  =  32890496  -137  GACTTATTTACCAAGCATTGGAGGAATATCGTAGGTAA
ERR166338.63037998  99  chr13  32890496  60  101M  =  32890558  137  ACCTCAGTCACATAATAAGGAATGCATCCCTGTGTAAG
ERR166338.63037998  147  chr13  32890558  60  75M  =  32890496  -137  GACTTATTTACCAAGCATTGGAGGAATATCGTAGGTAA
```

↑  
read ID

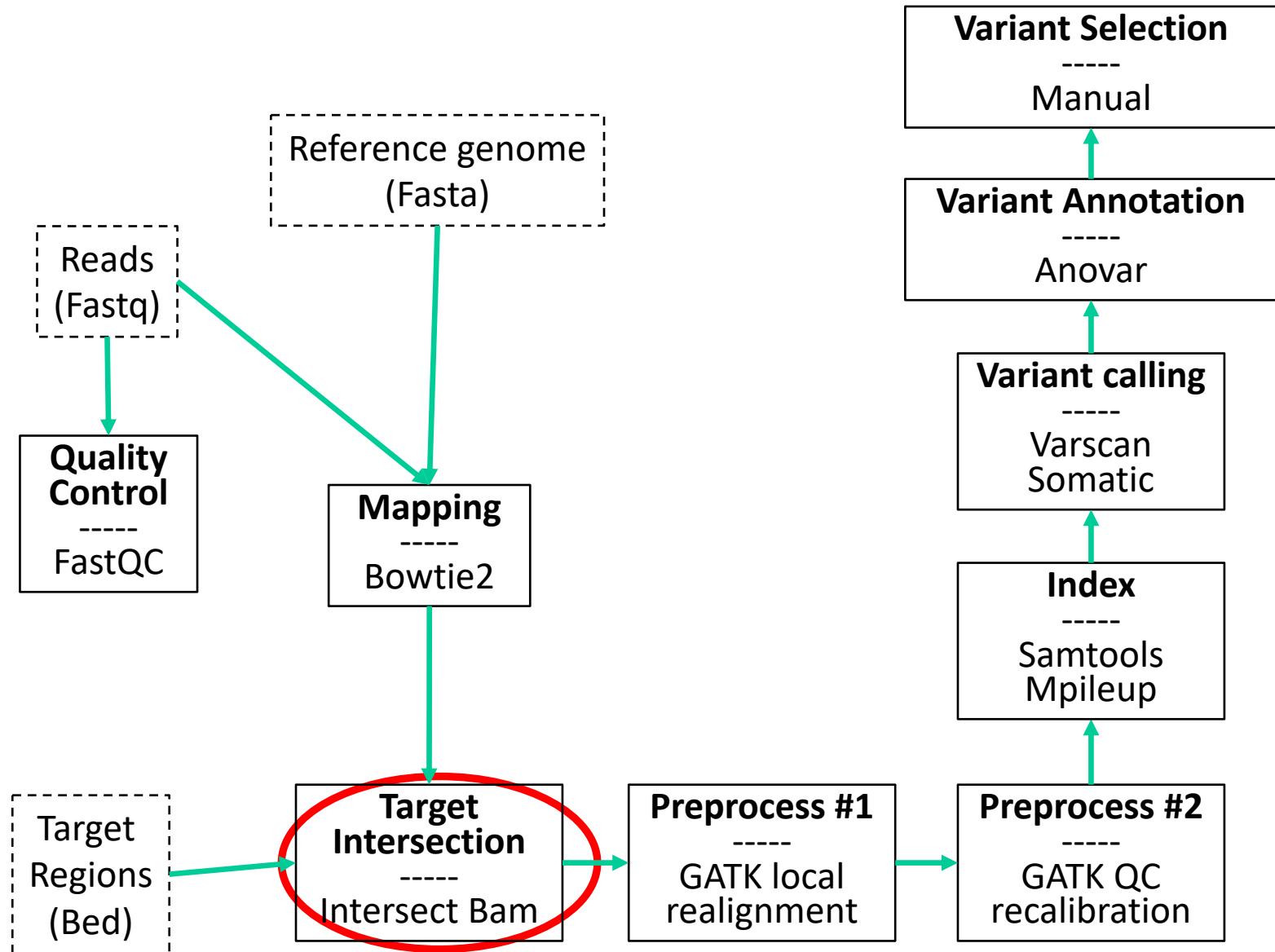
↑  
position

↑  
CIGAR

↑  
mapping qual.

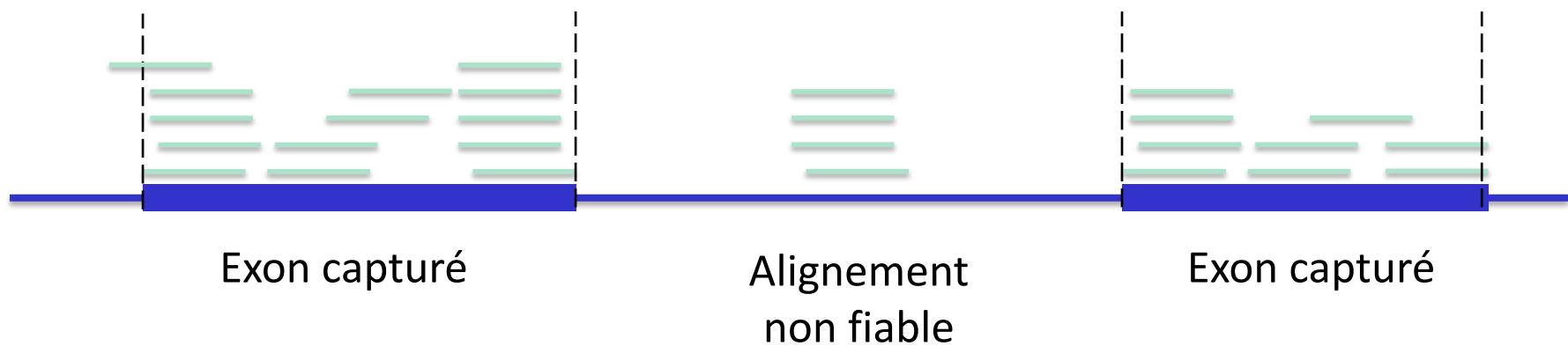
↑  
mate info

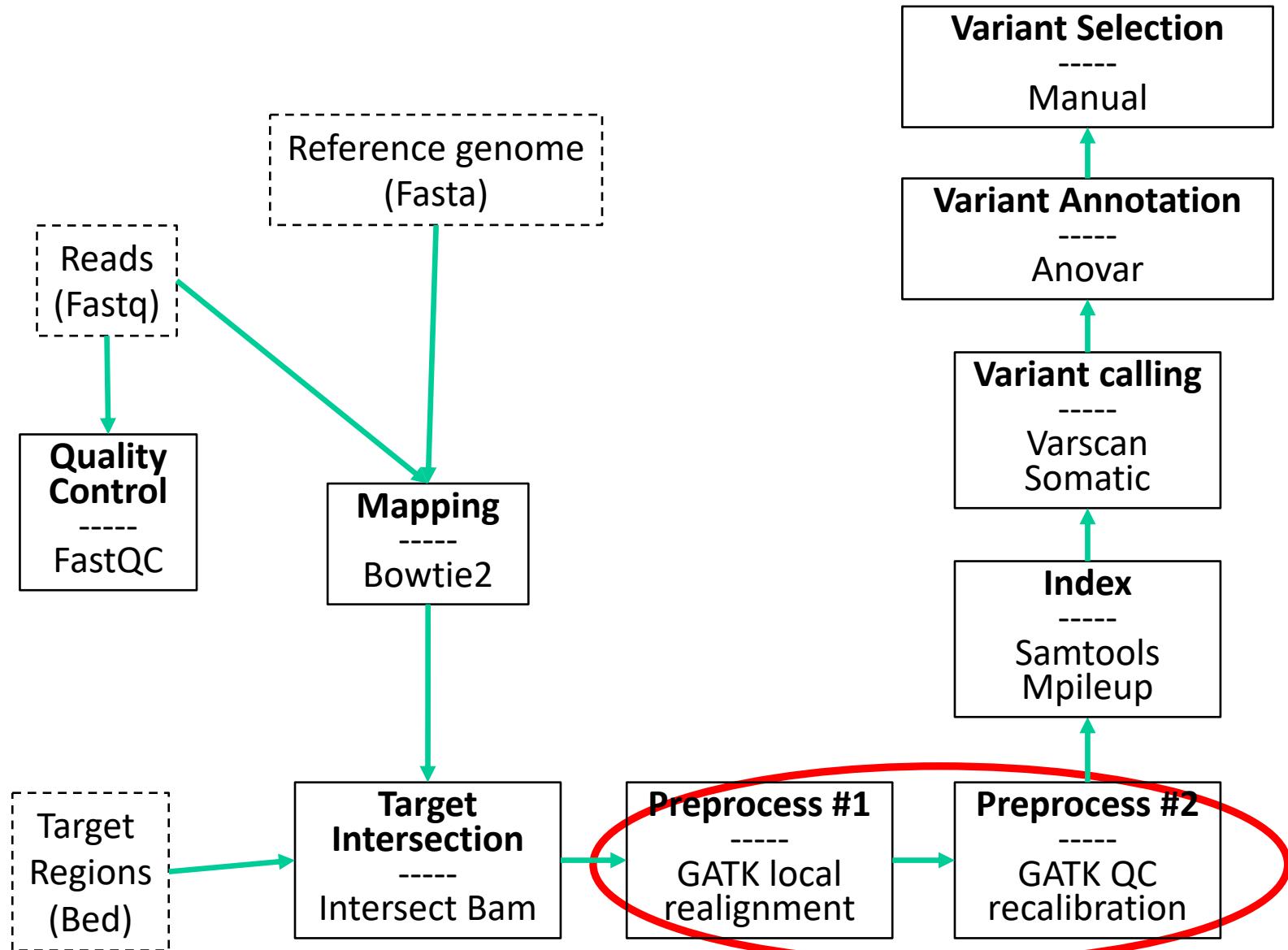
↑  
flag



# Target intersection

- Comparer l'alignement obtenu à la liste des positions visées par le protocole de capture





# Why realign around indels ?

- Small Insertion/deletion (Indels) in reads (especially near the ends) can trick the mappers into wrong alignments
  - Alignment scoring – cheaper to introduce multiple Single Nucleotide Variants (SNVs) than an indel: induce a lot of false positive SNVs
- ➔ artifactual mismatches
- **Realignment around indels helps improve the downstream processing steps**

# Wrong alignment near indels: a common problem in repeats/microsatellites

*Genome*

GACGTGCCAGCGAGTTACT

*Sequence with 2  
deletions*

GACGT -- CAGCGAGTTACT

GACGTCA**GCGAGTTACT**

*Cost: 2 indels*

*Cost: 8 mismatches*

preferred ali

*Genome: microsat.*

GACGTAGAGAGAGAGAGCG

*Sequence with 2  
deletions*

GACGT -- AGAGAGAGAGCG

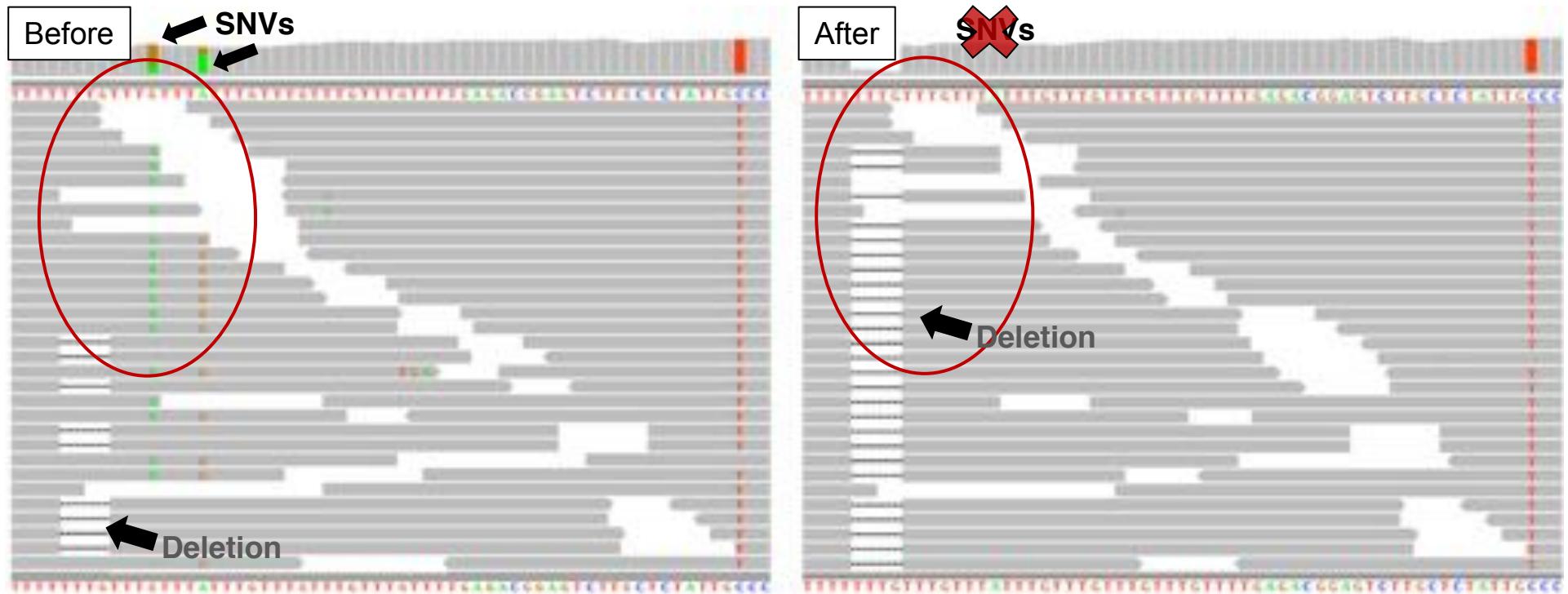
GACGTAGAGACAGAG**CG**

*Cost: 2 indels*

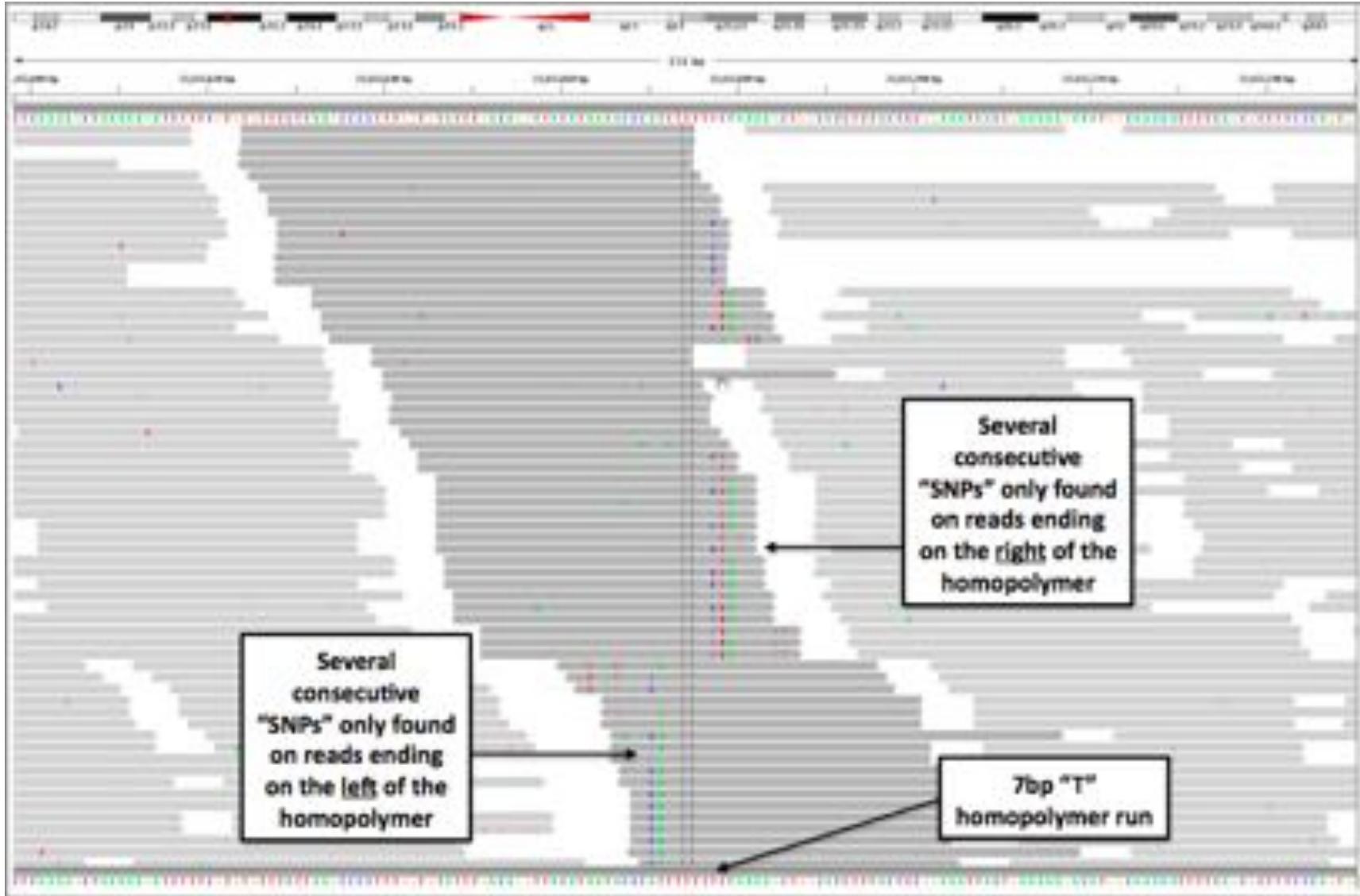
*Cost: 1 mismatch*

preferred ali

# Local realignment around indels



# Local realignment around indels



# Local realignment around indels



# Indel realignment in 2 steps

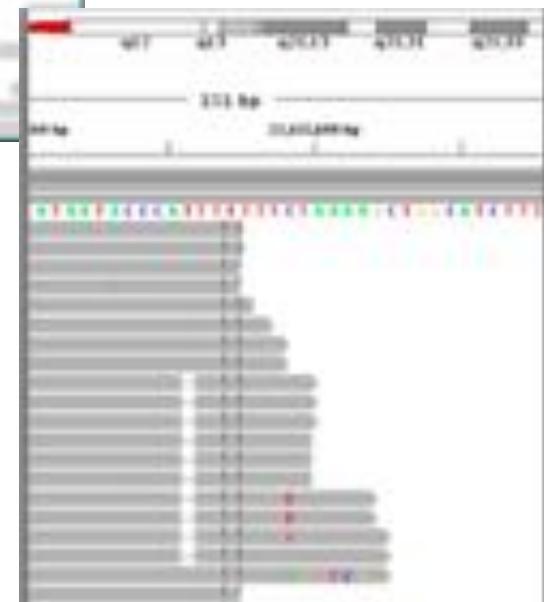
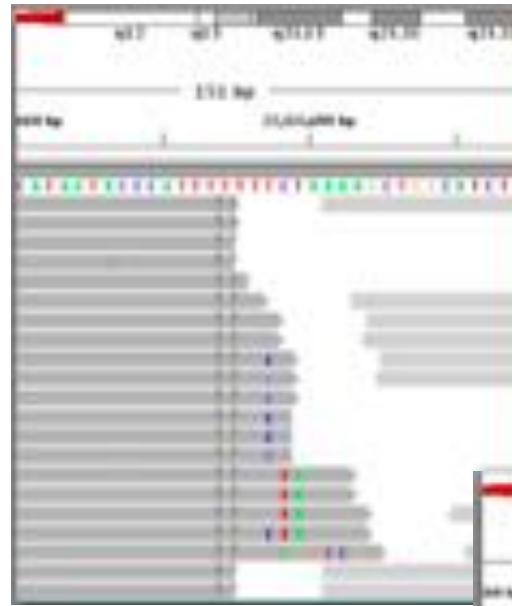
1. Identify what regions need to be realigned

- RealignerTargetCreator + known sites

Intervals  
↓

2. Perform the actual realignment (BAM output)

- IndelRealigner

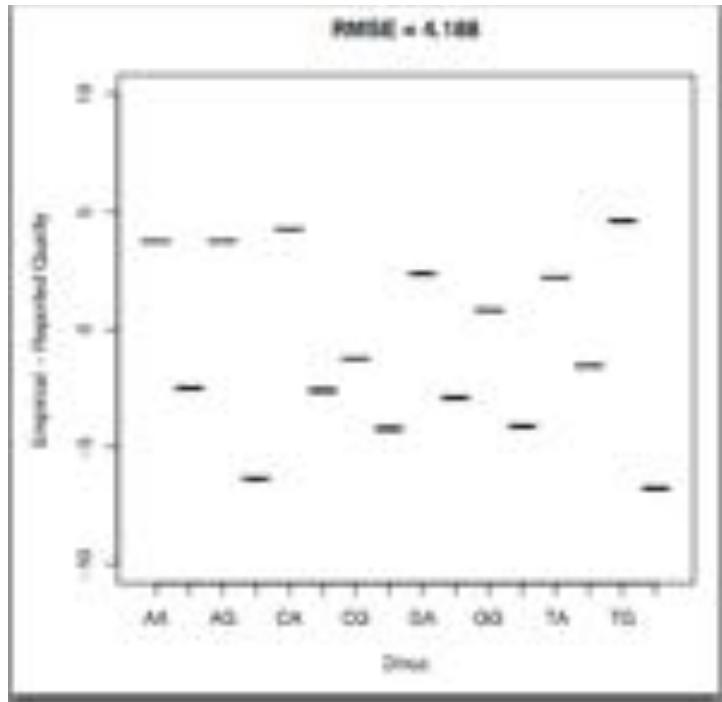


# Types of realignment targets

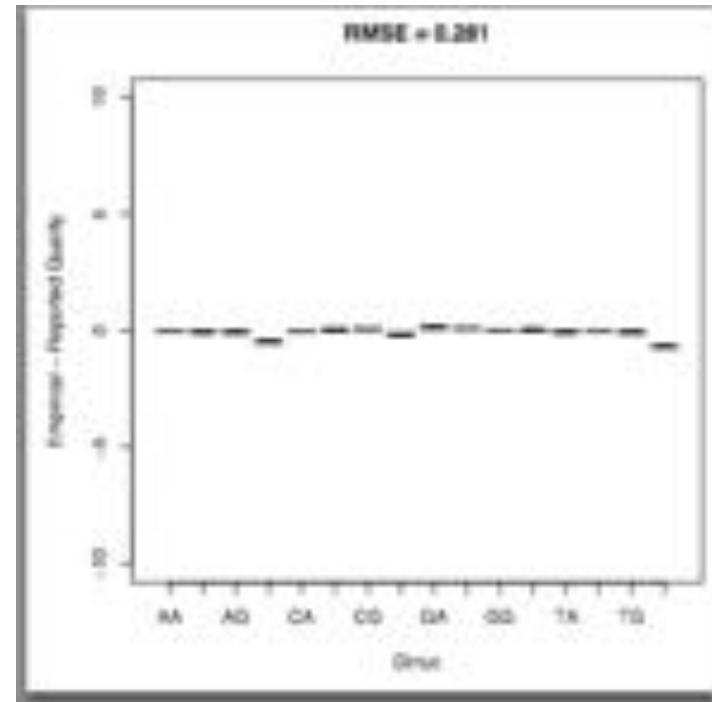
1. Indels seen in original alignments (in CIGAR, indicated by I for Insertion or D for Deletion)
2. Sites where evidences suggest a hidden indel (SNV abundance)
3. Known sites:
  - Common polymorphisms: dbSNP, 1000Genomes

# The quality scores issued by sequencers are biased

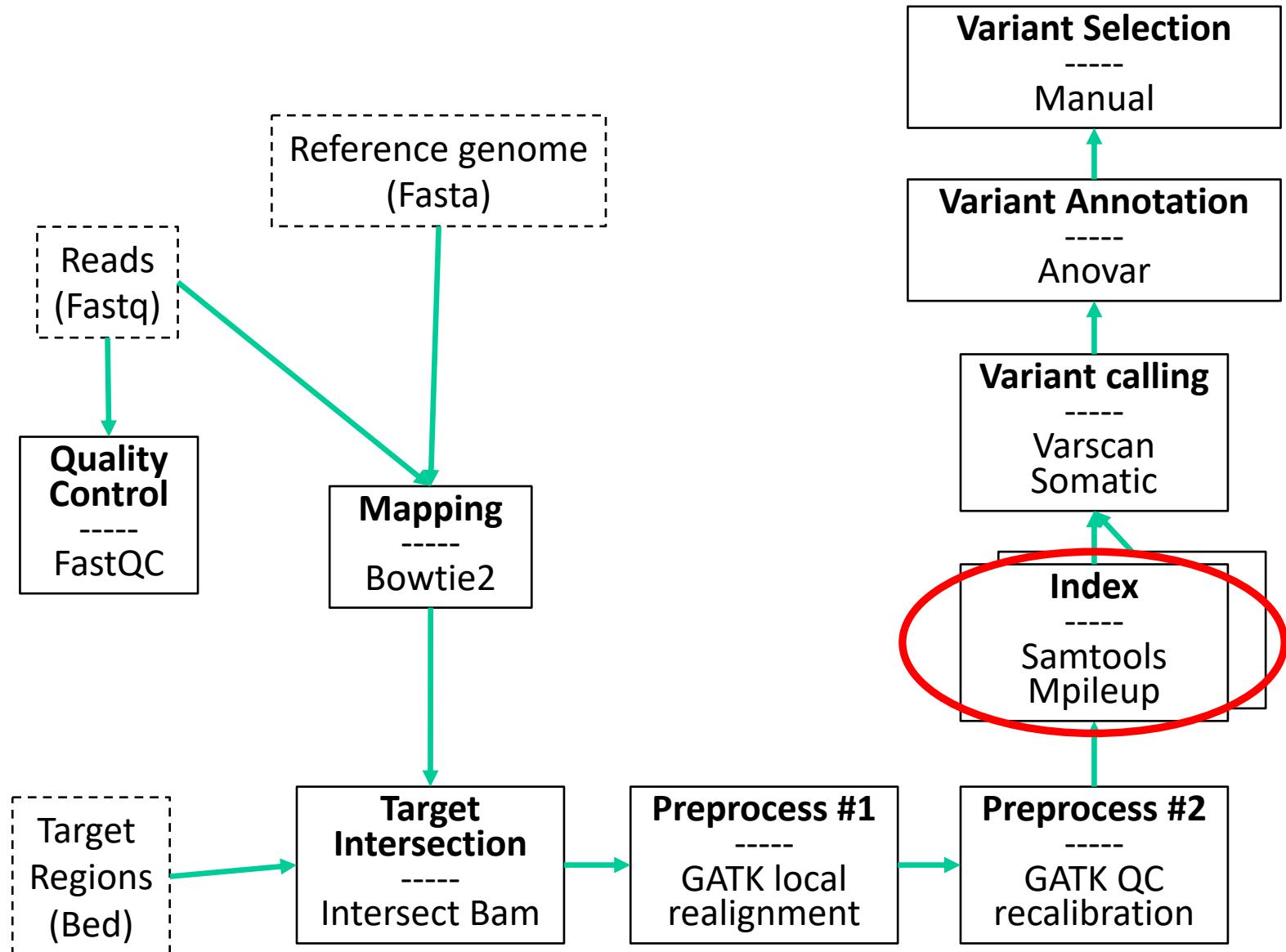
- Quality scores are critical for all downstream analysis
- Systematic biases are a major contributor to bad calls
- Example of sequence context bias in the reported qualities:



before



after



# PileUp:

# Pourquoi un nouveau format de fichier?

Rappel BAM:

```
@RG  ID:group1  SM:1425_CD34  PL:ILLUMINA  LB:lib1 PU:unit1
@PG  ID:bwa  PN:bwa  VN:0.7.12-r1039 CL:bwa mem -M -t 2 -A 2 -E 1 -R @RG\tID:group1\tSM:1425_CD34\tPL:ILLUMINA\tLB:lib1\tPU:unit1 /root/myd
ERR166338.13782800  83  chr13  32890449  60  101M  =  32890343  -207  GGGACTGAATTAGAACAAATTTCAGCGCTT
ERR166338.13782800  163  chr13  32890343  60  75M  =  32890449  207  CACTAGCCACGTTCGAGTGCTTAATGTGGCTAGTGGC
ERR166338.26716588  99  chr13  32890406  60  101M  =  32890553  222  AATGTTCCCACCTCACAGTAAGCTGTTACCGTTCCAG
ERR166338.26716588  147  chr13  32890553  60  75M  =  32890406  -222  TTGCAGACTTACCAAGCATTGGAGGAATATCGTAA
ERR166338.27259961  99  chr13  32890496  60  101M  =  32890558  137  ACCTCAGTCACATAATAAGGAATGCATCCCTGTGTAAG
ERR166338.27259961  147  chr13  32890558  60  75M  =  32890496  -137  GACTTATTTACCAAGCATTGGAGGAATATCGTAGGTAA
ERR166338.63037998  99  chr13  32890496  60  101M  =  32890558  137  ACCTCAGTCACATAATAAGGAATGCATCCCTGTGTAAG
ERR166338.63037998  147  chr13  32890558  60  75M  =  32890496  -137  GACTTATTTACCAAGCATTGGAGGAATATCGTAGGTAA
```

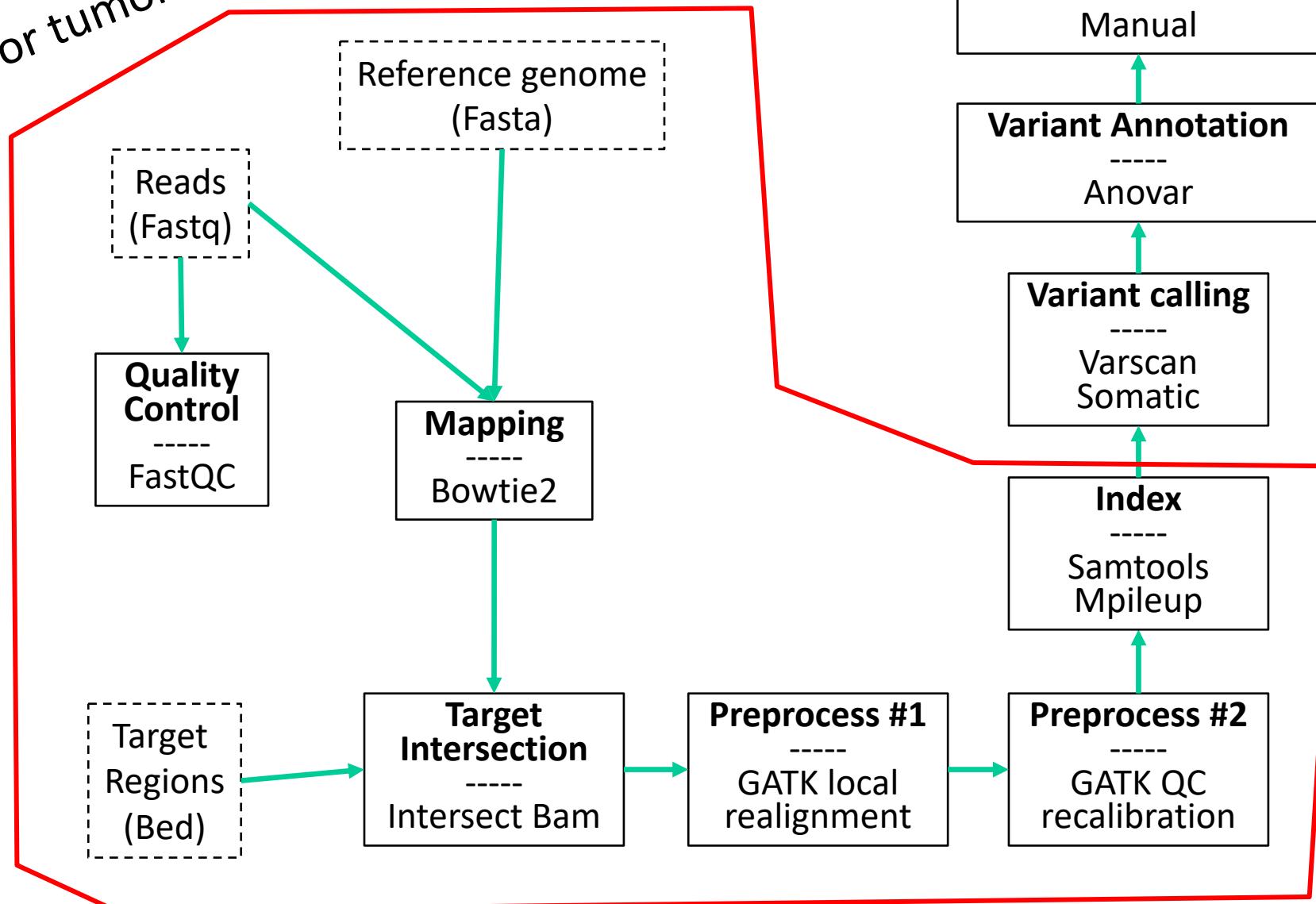


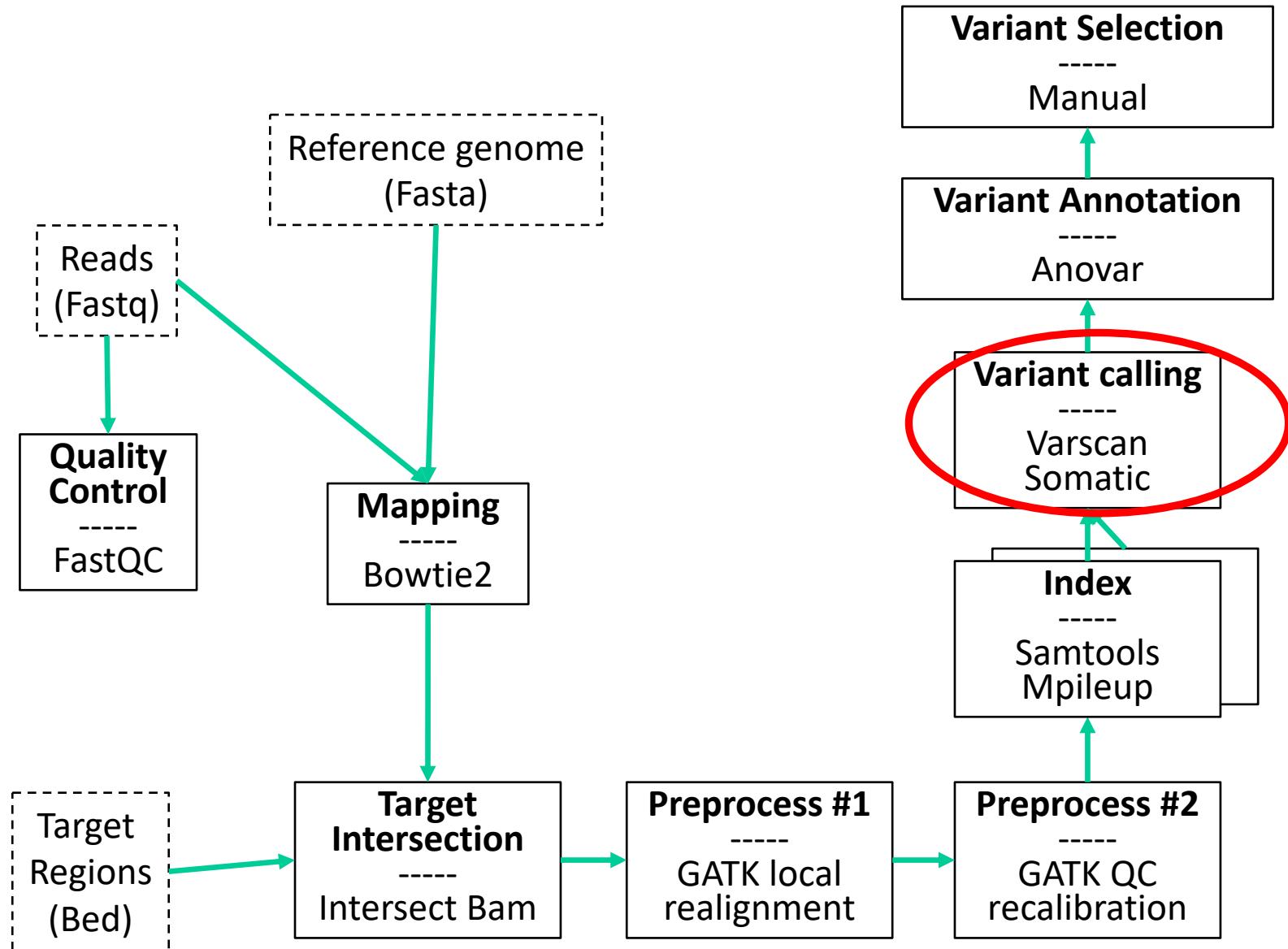
# Pileup format

Describes base-pair information at each position

		Reference base	Base qualities
chr12	112888238	A 108	=4 =??????@??@@@ @=@ ??@ ? @? ? ??< ? ??@?????? ? @??? ? ??@?? A???@ @ @@@???AB????= ? @ @@??@ @?@ A 00
chr12	112888239	C 108	.\$t,,,..T,tT.,,T.,,t.tTtt.tTT,t.T,tTt.tT,T,,.t TtTttt.,,Tt.ttt.,T,,tT,,T,T.,tT,,t,TttTtT,T. @6????<?=6 66= ? ??? 6=7??=???<8@7=7=? ,tt,,T,T,tttt^F.^F, 936 78??6??6 45<875? ??? ?@6 77??7?8 ??78??7????? 8 <8??88 9?8 ?0048
Number of reads covering the site (total depth)		Read bases:	
		<p>. / , = match on forward/reverse strand ACGTN / acgtN = mismatch on forward/reverse strand '-\+[0-9]+\[ACGTNacgtN]+\'' indicates an indel</p>	

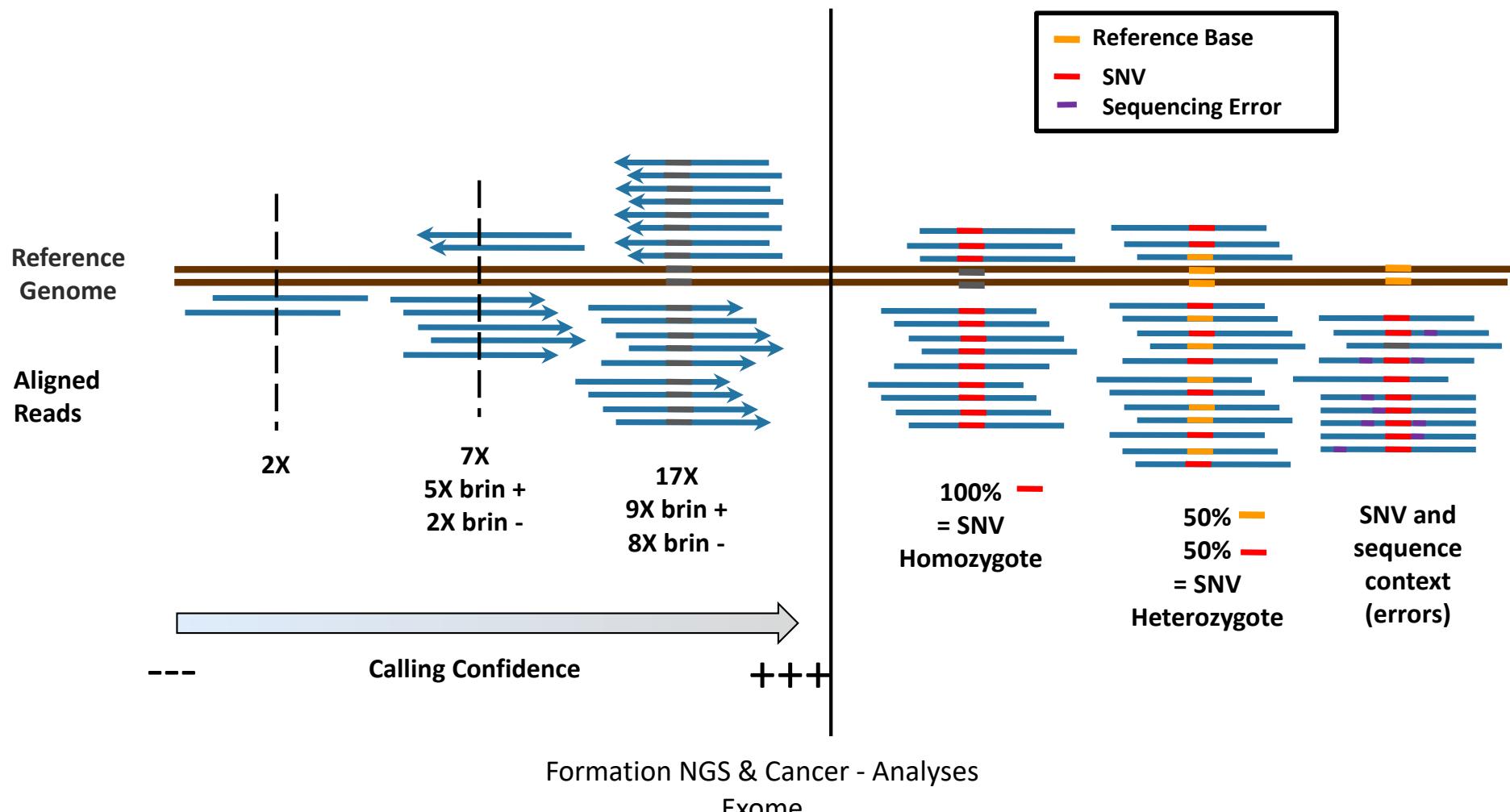
X2 for tumor & normal





# Variant calling criteria

Depth of Coverage = number of reads supporting one position  
ex: 1X, 5X, 100X... >1000X



# Factors to consider for Variant Calling

- Calling a SNV:
  - Base call qualities of each supporting base (base quality)
  - Proximity to small indels, or homopolymer run
  - Mapping qualities of the reads supporting the SNP
  - Sequencing **depth**: >=30x for constit ; >=100 for tumor
  - SNVs position within the reads: Higher error rate at the reads ends
  - Look at strand bias (SNVs supported by only one strand are more likely to be artifactual)
  - **Allelic frequency**: Tumor cellularity will reduce the % of an heterozygous variant
- Calling an indel:
  - Higher stringency (and Sanger validation often needed)

# VarScan2

- Mutation caller written in **Java** (portable)
- Works with **Pileup files** of Targeted, Exome, and Whole-Genome sequencing data (DNAseq or RNAseq)
- **Multi-platforms:** Illumina, SOLiD, Life/PGM, Roche/454
- Germline mode:
  - Variants in individual samples
  - Multi-sample variants **shared or private** in multi-sample datasets
- Somatic Mode: **Tumor/Normal pairs:**
  - Somatic, germline, LOH events
  - Somatic copy number alterations (CNAs)

# VarScan2 Performance

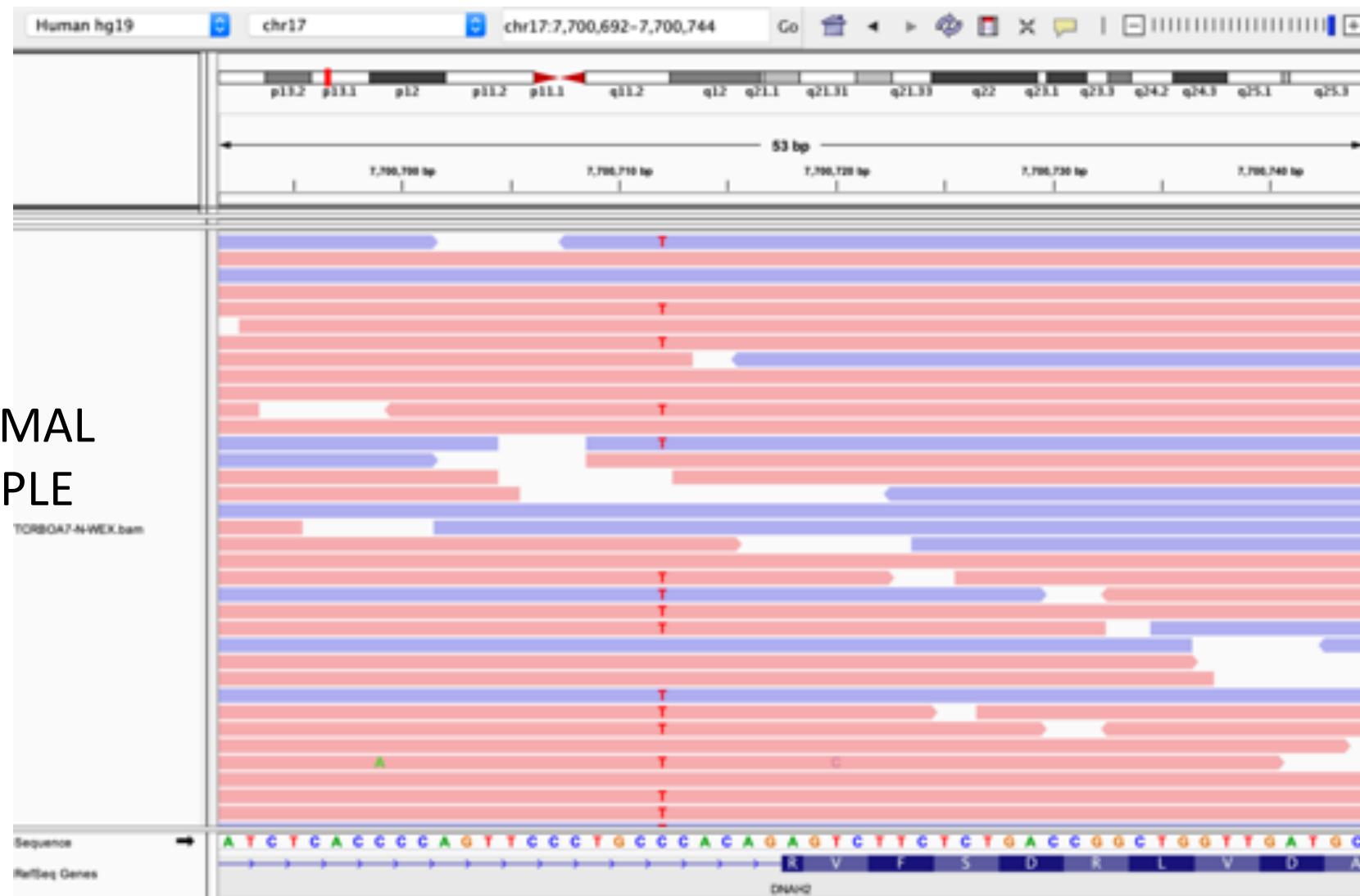
- VarScan uses a robust **heuristic/statistic** approach to call variants that meet desired thresholds for read depth, base quality, variant allele frequency, and statistical significance
- Stead *et al.* (2013) compared 3 different **somatic callers** : MuTect, Strelka, VarScan2
  - **VarScan2 performed best** overall with sequencing depths of 100x, 250x, 500x and 1000x required to accurately identify variants present at 10%, 5%, 2.5% and 1% respectively
- Other widely used tool: **GATK**

# Somatic variant calling

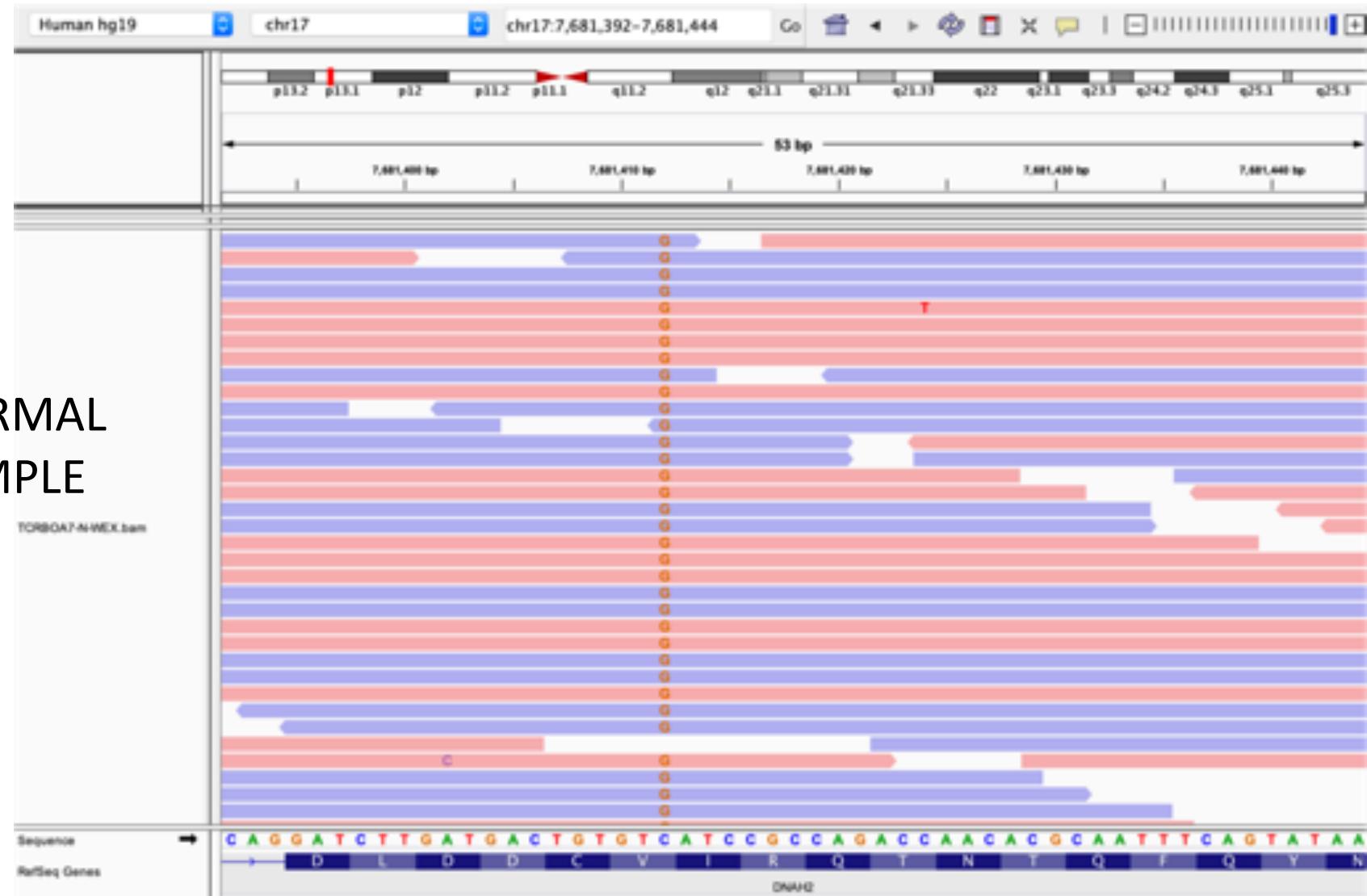
# Fréquence/fraction allélique

- Termes
  - Germline/population: Minor Allele frequency (**MAF**). Par ex. dans données 1000Genomes.
  - Somatic: Variant Allele frequency (**VAF**) ou B-Allele Frequency (**BAF**) (or Allelic Fraction)
- Où trouver l'info?
  - Colonne info#AF dans VCF

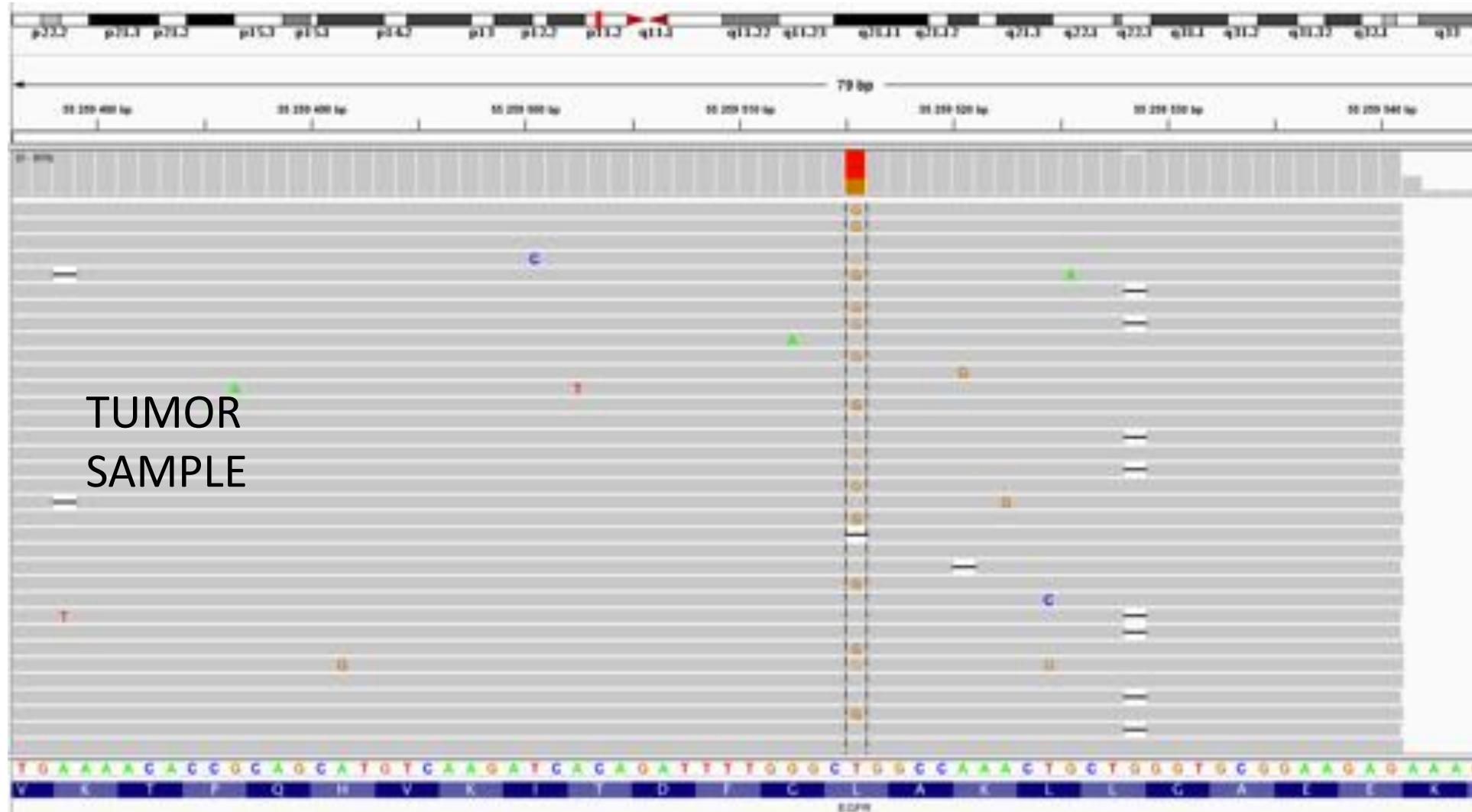
# Un polymorphisme (SNP) hétérozygote



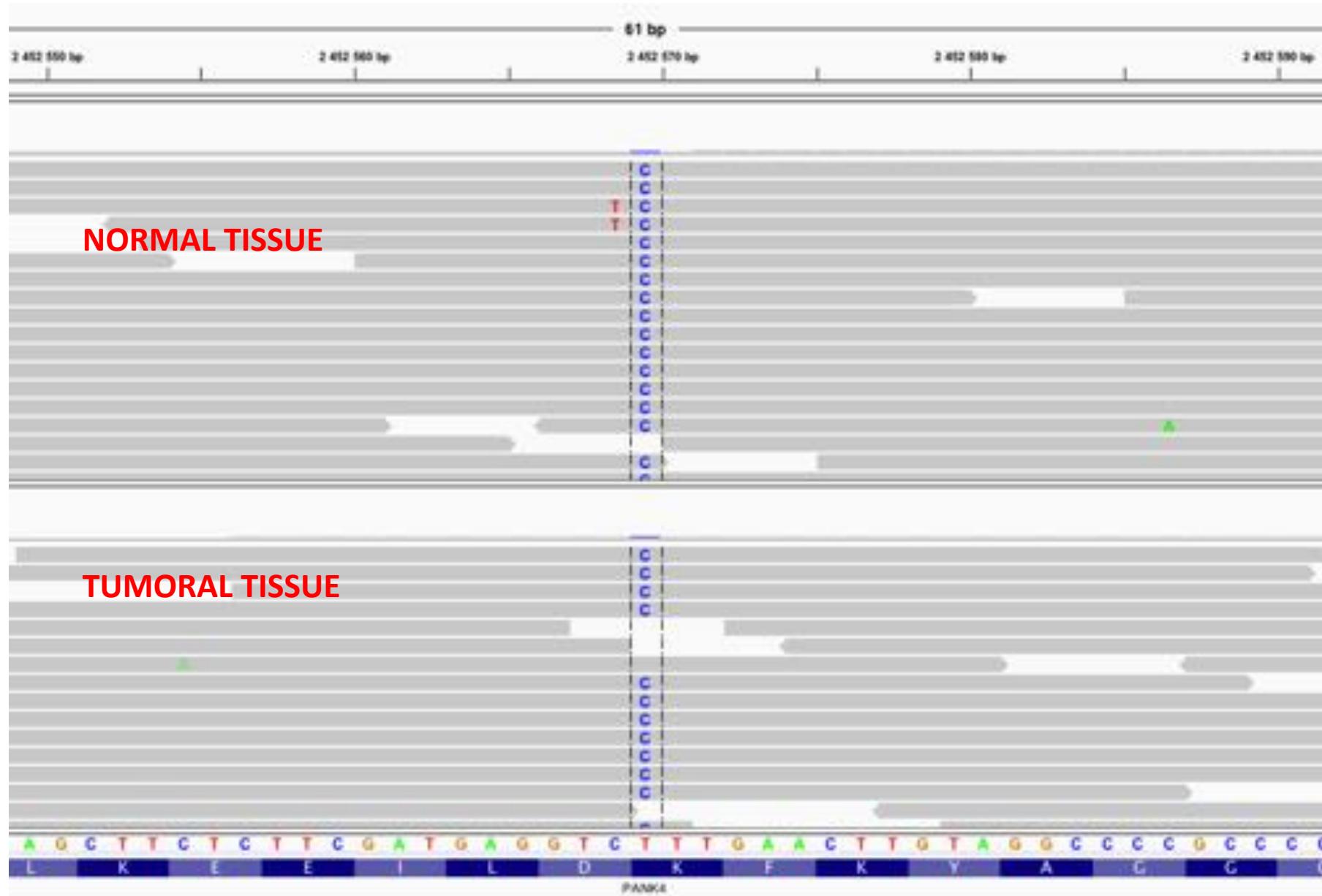
# Un polymorphisme (SNP) homozygote



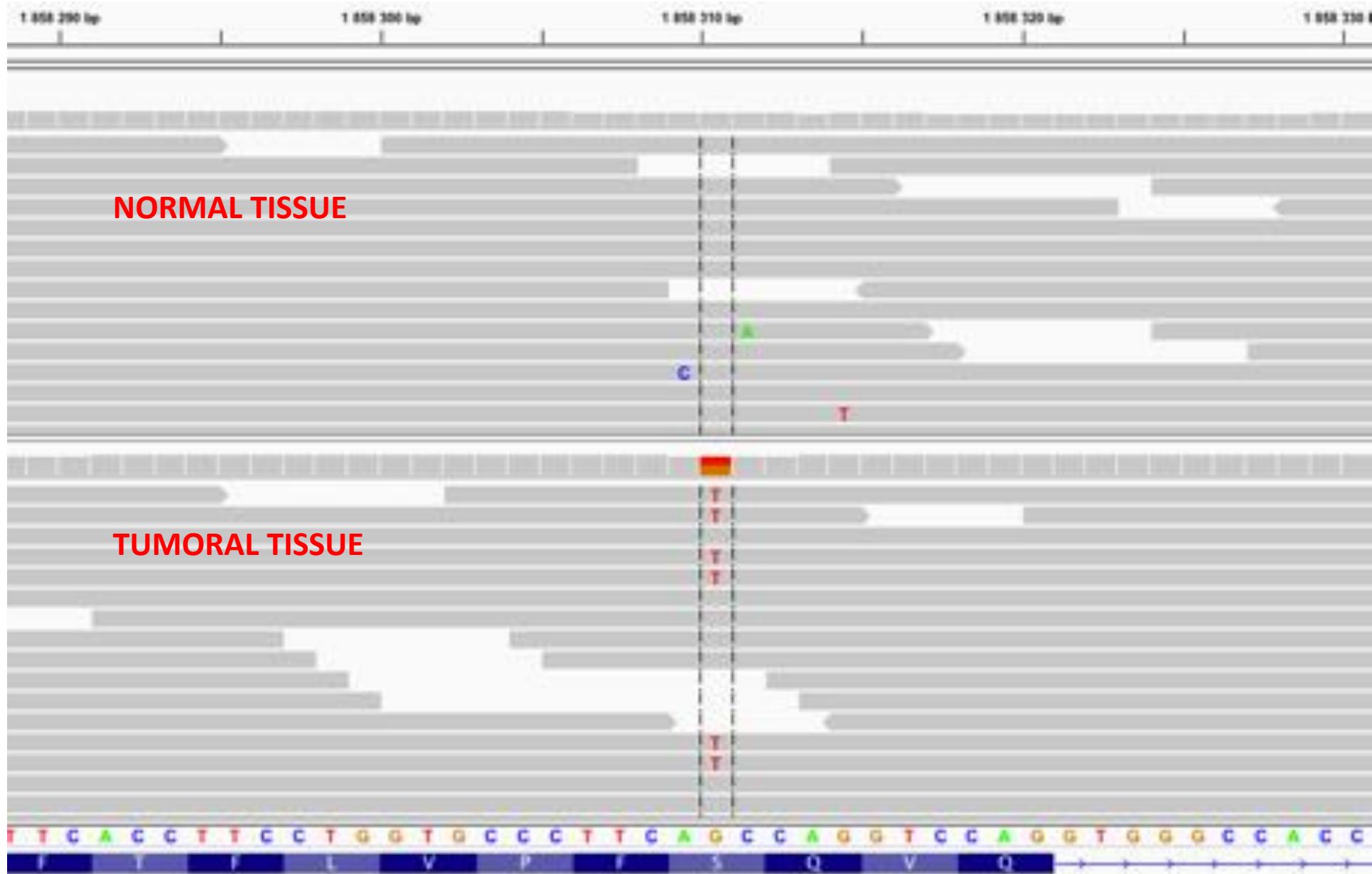
# Un variant tumoral: polymorphisme ou mutation?



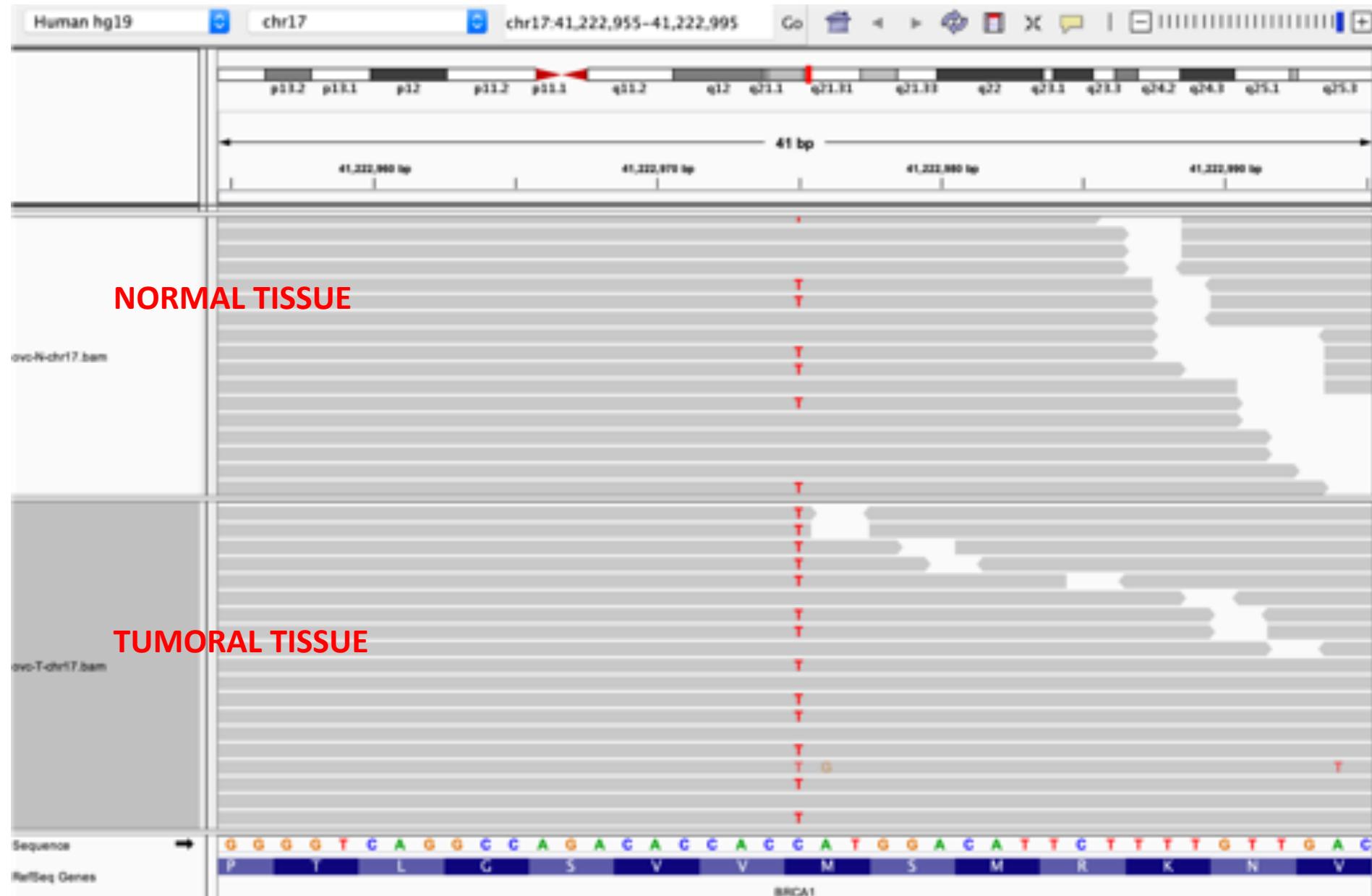
# Un polymorphisme vu dans N et T



# Une mutation somatique



# Une LOH (loss of heterozygosity)



# Varscan's Somatic P-value

## Variant Calling and Comparison

At every position where both normal and tumor have sufficient coverage, a comparison is made. First, normal and tumor are called independently using the germline consensus calling functionality. Then, their genotypes are compared by the following algorithm:

Calculate significance of allele frequency difference by Fisher's Exact Test

**If difference is significant (p-value < threshold):**

If normal matches reference

==> Call Somatic

Else If normal is heterozygous

==> Call LOH

Else normal and tumor are variant, but different

==> Call IndelFilter or Unknown

**If difference is not significant:**

==> Call Germline

The diagram illustrates the logic for determining if a variant is somatic or LOH based on allele counts in normal (N) and tumor (T) samples.

**Alleles**

	Ref	Var
N	8	0
T	6	7

**Somatic Call:** A red arrow points from the table to the text "Somatic".

	Ref	Var
N	4	4
T	8	1

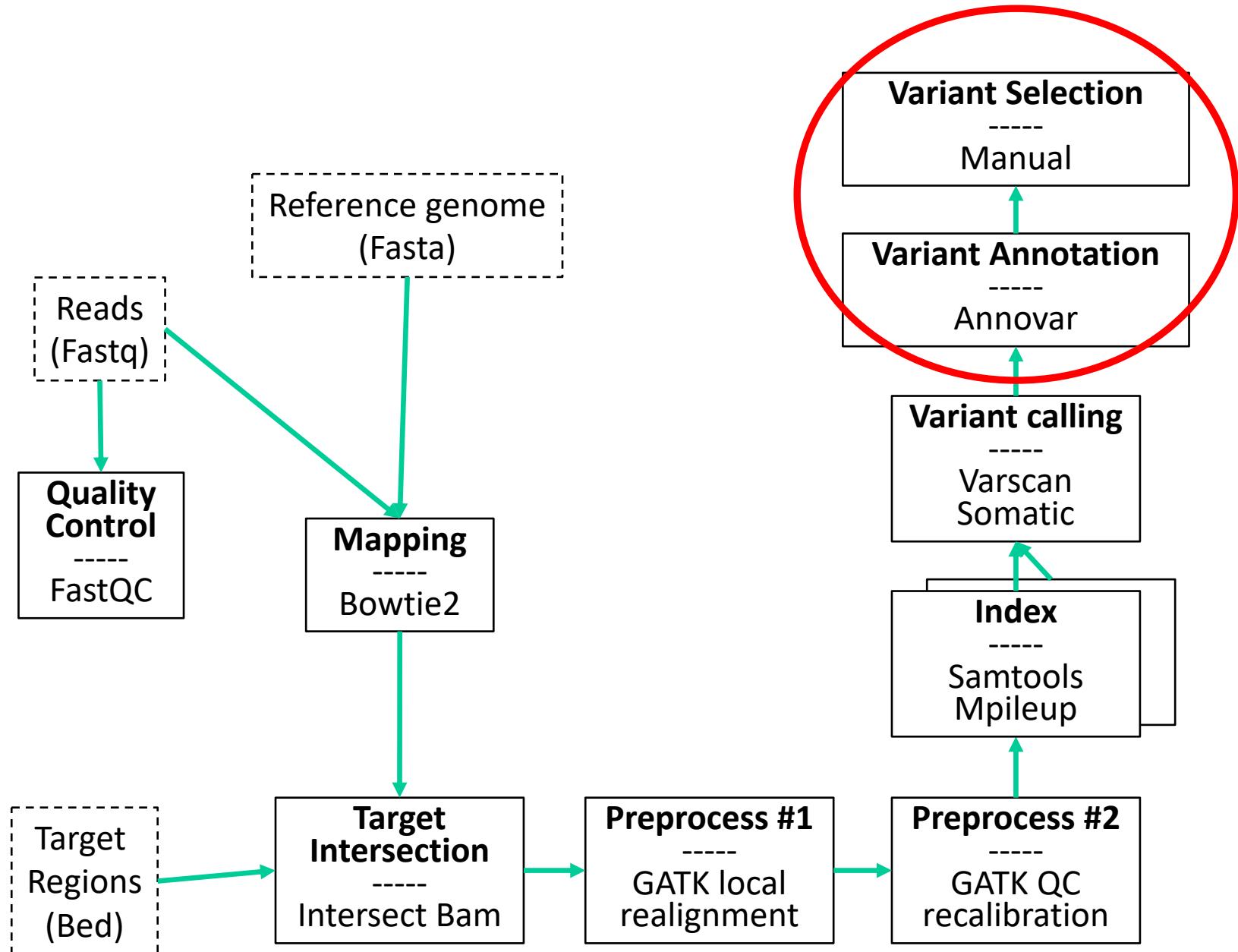
**LOH Call:** A red arrow points from the table to the text "LOH".





# Fields in Varscan "native format"

Native Output Field	VCF Field	Description
chrom	CHROM	Chromosome or reference name
position	POS	Position from pileup (1-based)
ref	REF	Reference base at this position
var	ALT	Variant base seen in tumor
normal_reads1	RD (col 10)	Reads supporting reference in normal
normal_reads2	AD (col 10)	Reads supporting variant in normal
normal_var_freq	FREQ (col 10)	Variant allele frequency in normal
normal_gt	GT (col 10)	Consensus genotype call in normal
tumor_reads1	RD (col 11)	Reads supporting reference in tumor
tumor_reads2	AD (col 11)	Reads supporting variant in tumor
tumor_var_freq	FREQ (col 11)	Variant allele frequency in tumor
tumor_gt	GT (col 11)	Consensus genotype call in tumor
somatic_status	SS (col 8)	Somatic status call (Germline, Somatic, LOH, or Unknown)
variant_p_value	GPV (col 8)	Variant p-value for Germline events
somatic_p_value	SPV (col 8)	Somatic p-value for Somatic/LOH events
tumor_reads1_plus	DP4 (col 11)	Tumor reference-supporting reads on + strand
tumor_reads1_minus	DP4 (col 11)	Tumor reference-supporting reads on - strand
tumor_reads2_plus	DP4 (col 11)	Tumor variant-supporting reads on + strand
tumor_reads2_minus	DP4 (col 11)	Tumor variant-supporting reads on - strand
normal_reads1_plus	DP4 (col 10)	Normal reference-supporting reads on + strand
normal_reads1_minus	DP4 (col 10)	Normal reference-supporting reads on - strand
normal_reads2_plus	DP4 (col 10)	Normal variant-supporting reads on + strand
normal_reads2_minus	DP4 (col 10)	Normal variant-supporting reads on - strand





# Resources dedicated to human genetic variation

- dbSNP and 1000-genomes
  - Population-scale DNA polymorphisms
- COSMIC
  - Catalogue Of Somatic Mutations In Cancer
- Non synonymous SNVs predictions
  - SIFT, Polyphen2 (damaging impact)... PhyloP, GERP++ (conservation)



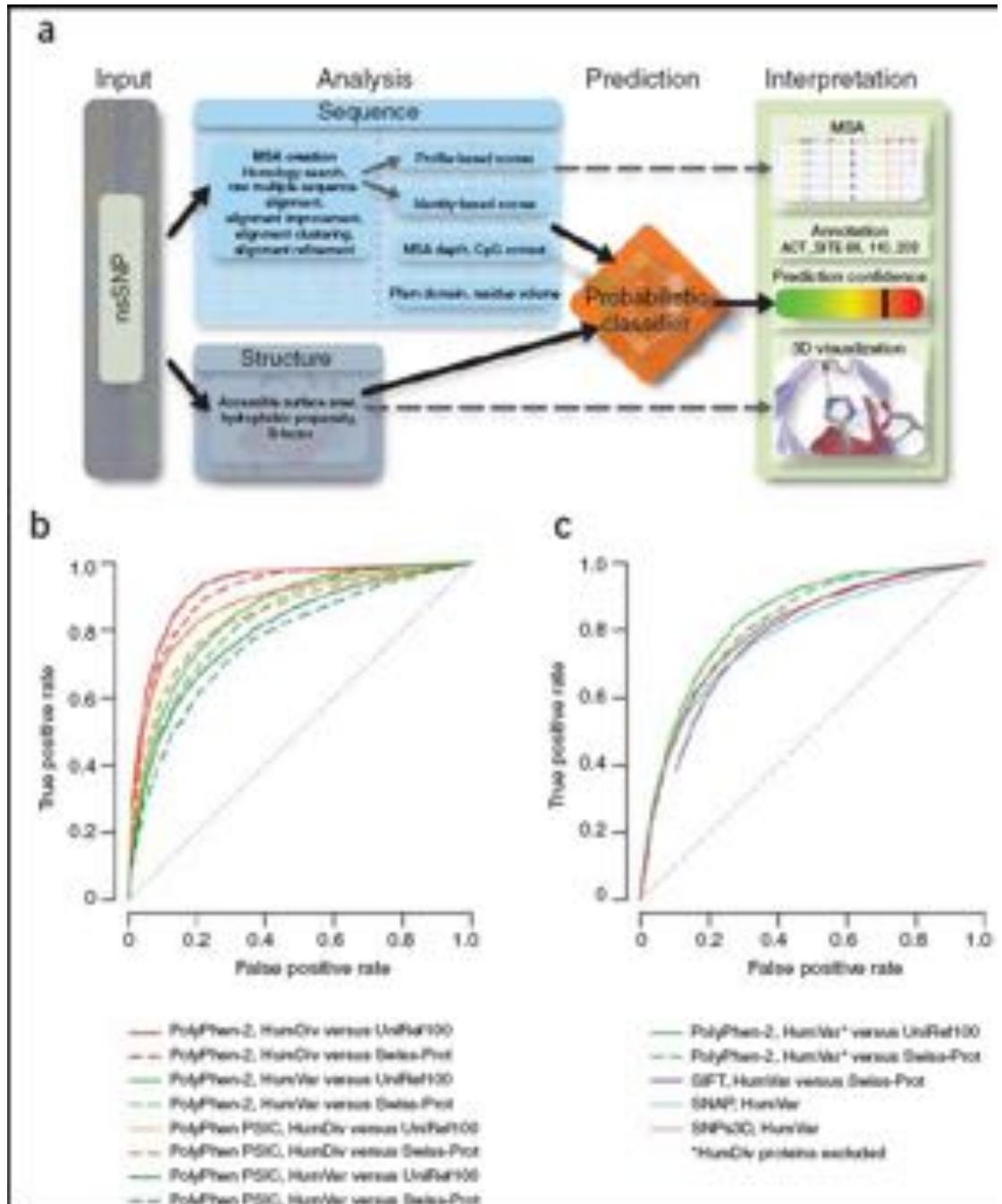


# PolyPhen2

Adzhubei et al. *Nature Methods* 2010.

**Probabilistic classifier:**  
Estimates the probability of the missense mutation being damaging based on a combination of seq+struct properties.

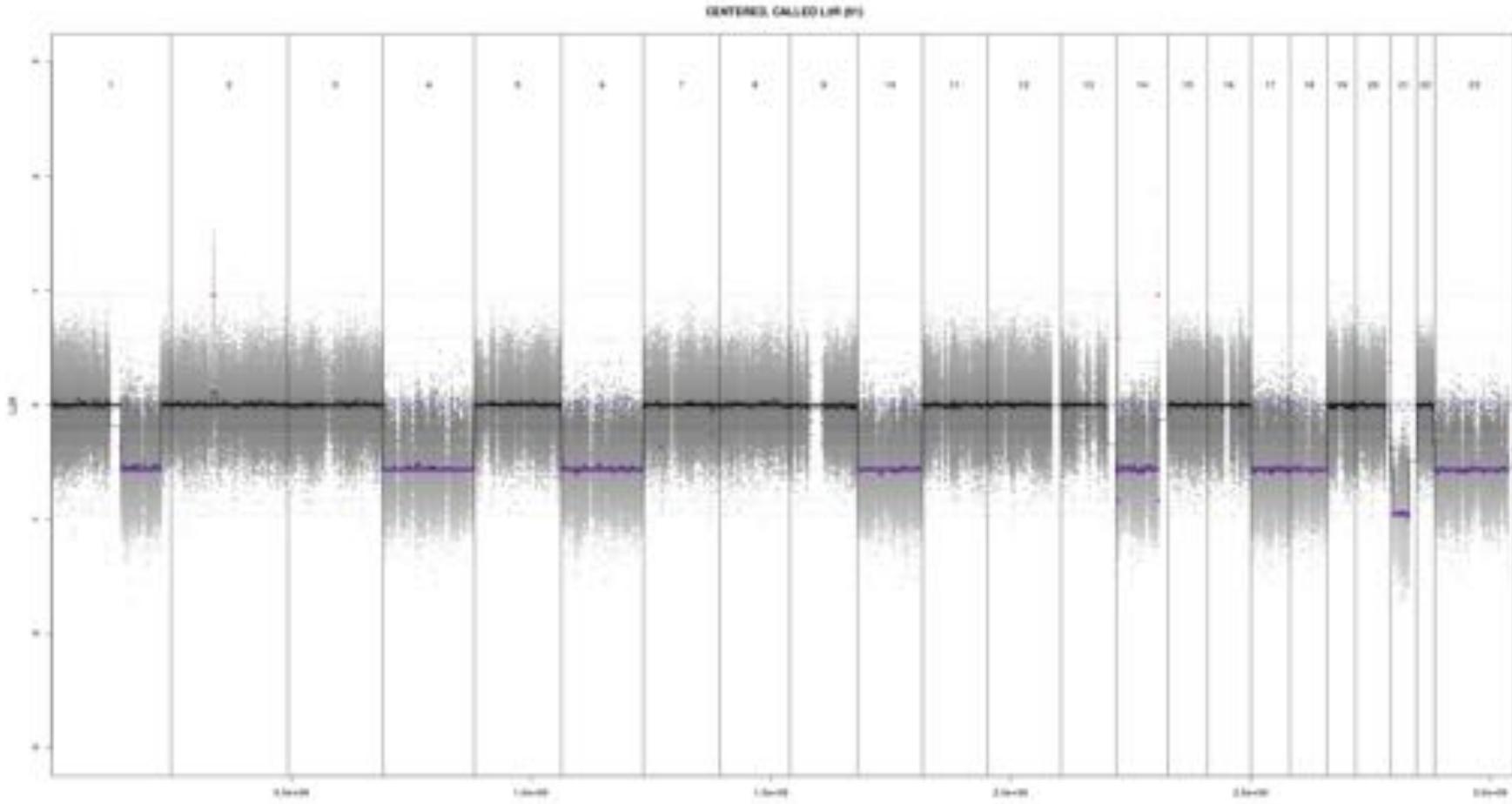
Classe en: **Benign**,  
**Possibly damaging**, or  
probably **Damaging**



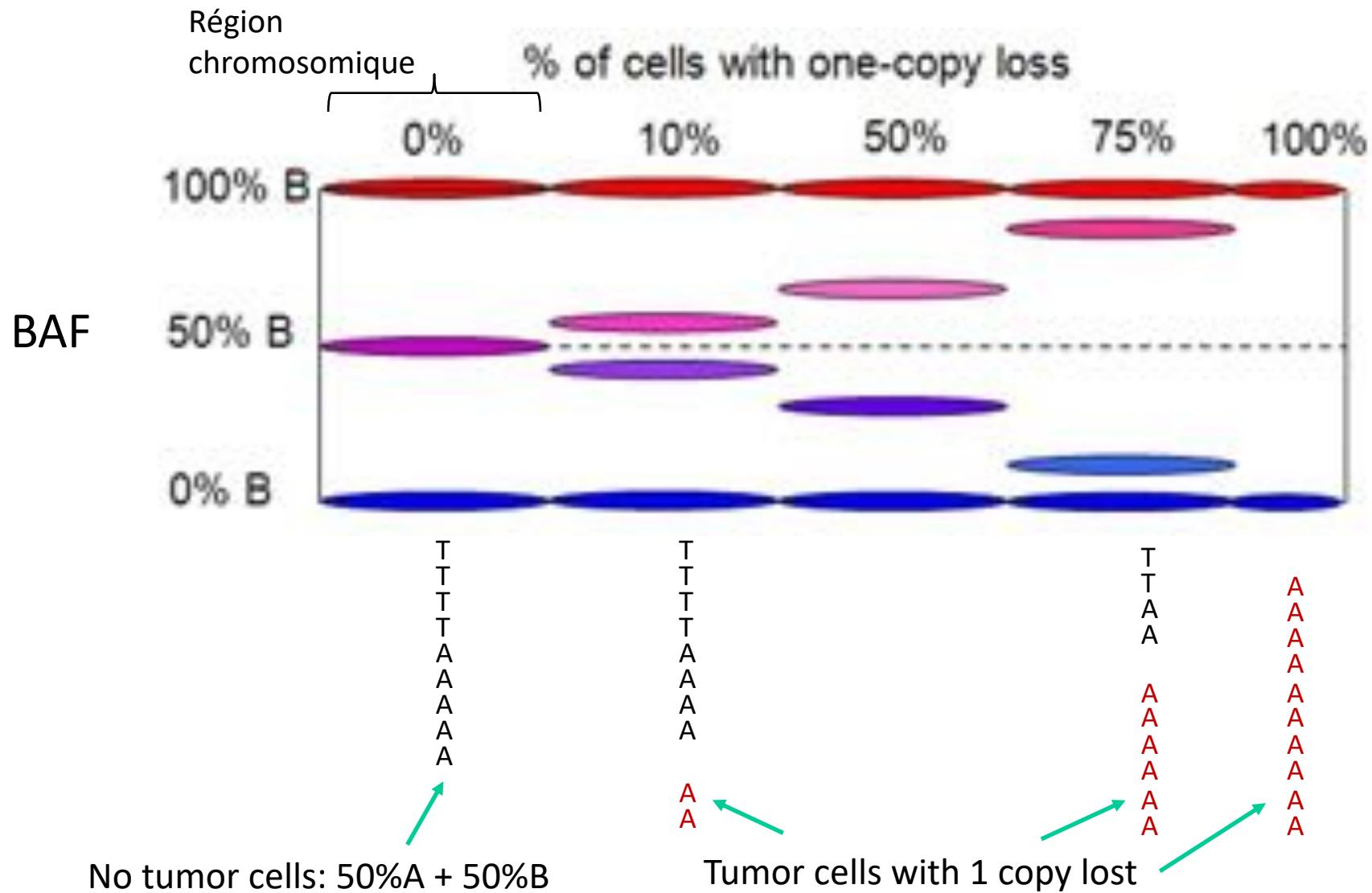
# Coverage & Allelic Frequencies For CNV detection

# Detection of copy-number variations

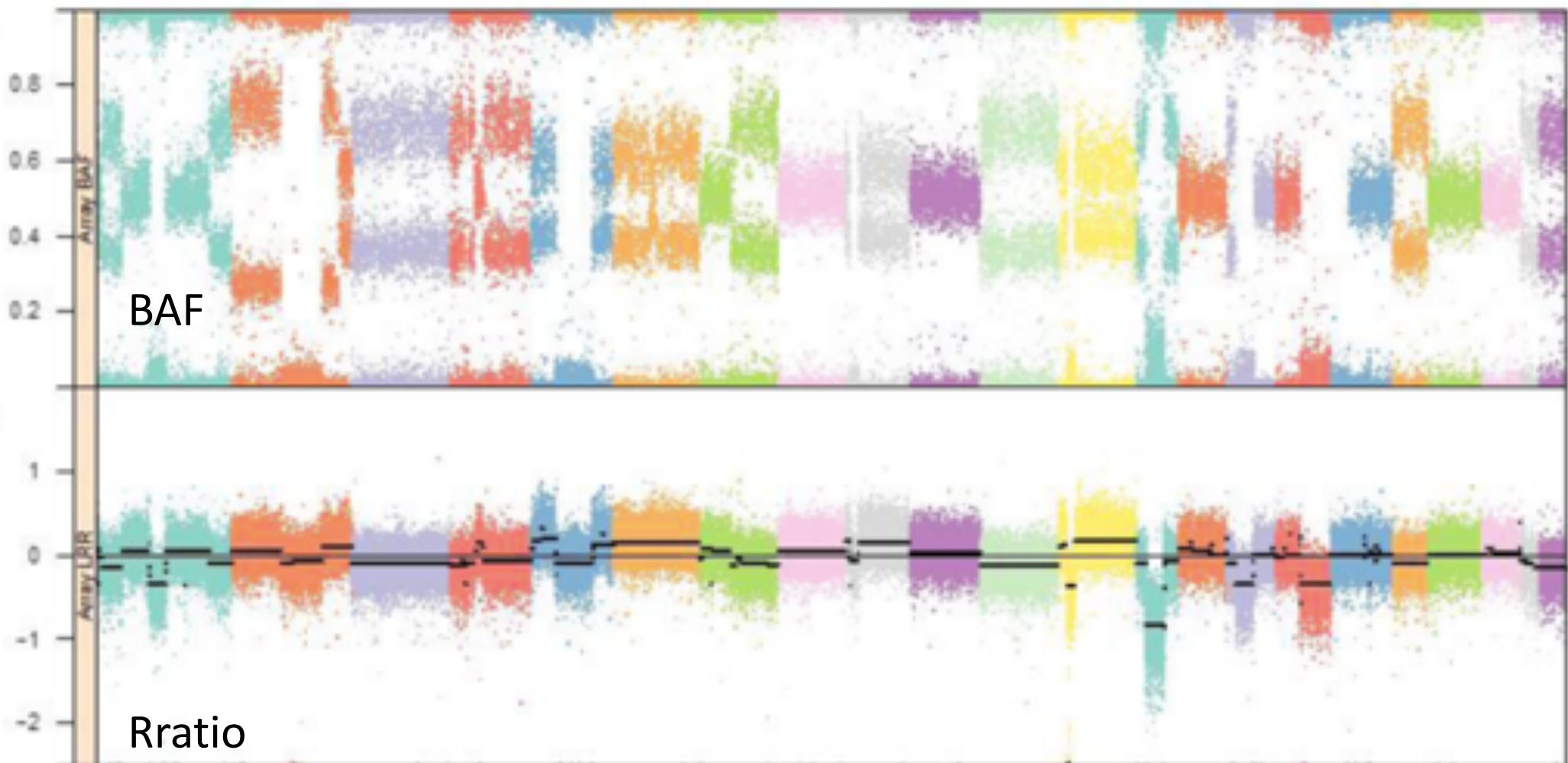
Are there any copy-number alteration (gain or loss of chromosomal regions, amplifications ...) that could explain tumorigenesis ?



# Cellularité et Fréquence Allelique



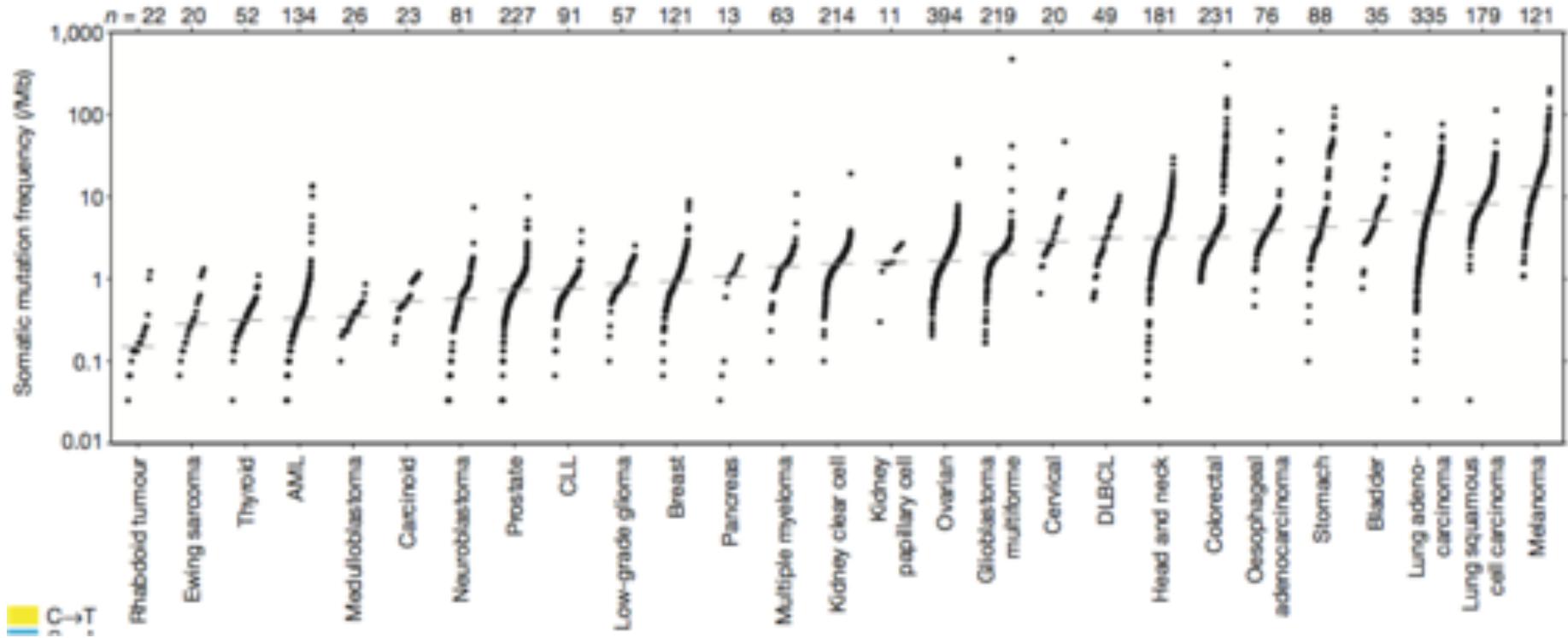
# Segmentation et fréquence allélique



R ratio=utilisé en CGH, =couverture en NGS

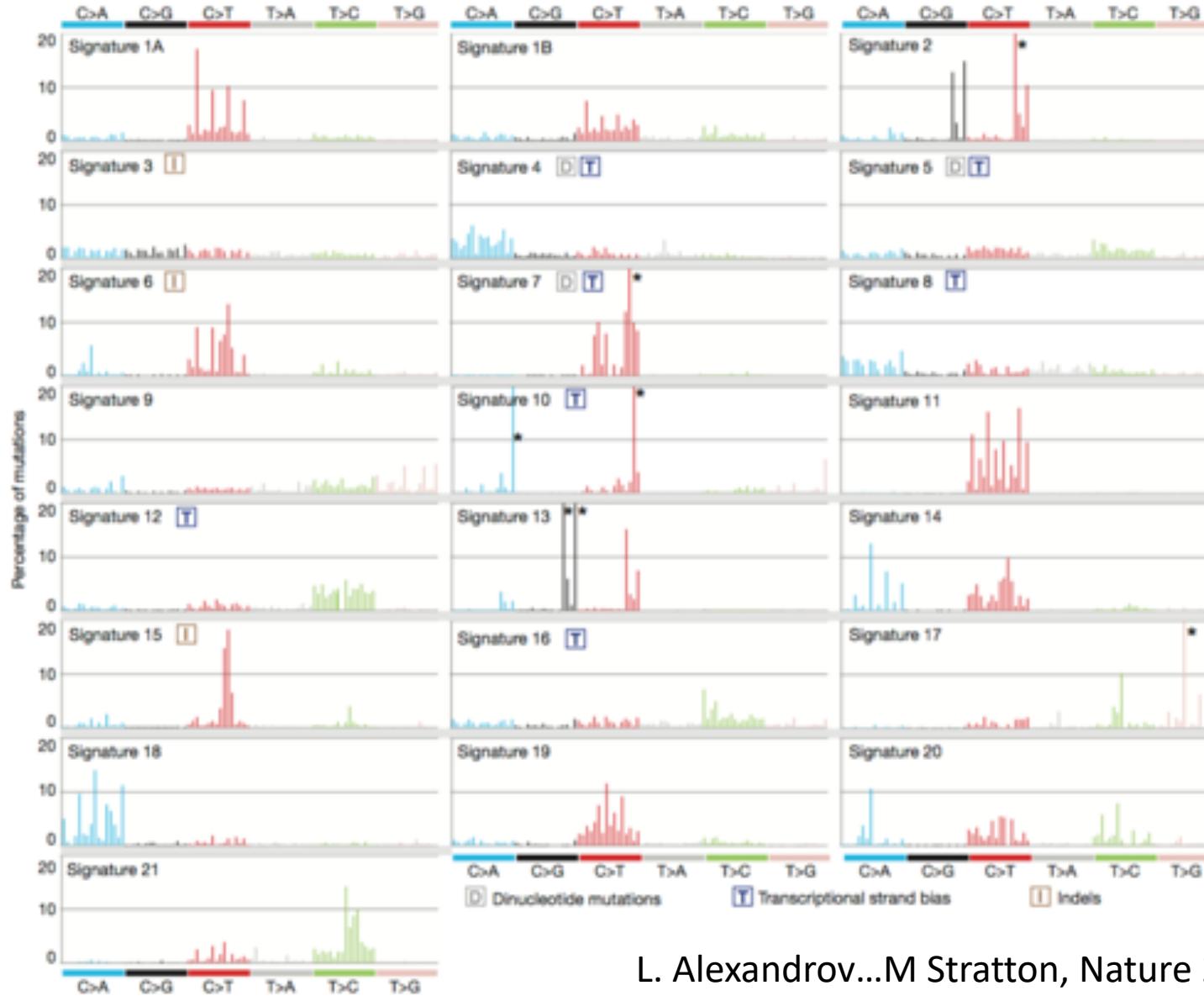
Scott et al. Gene 2014

# 2013: premières études pan-cancer WGS



Alexandrov...M Stratton, Nature 2013.  
Lawrence et al. Nature 2013.

# Les Signatures Mutationnelles





# Annexes

- Formats de fichiers NGS

# Format fasta

**\*.fa , \*.fasta**

```
>identifiant1 commentaire libre
CAGCATCGATCGTCGGCGATGCATGCGGATGCTAGCTGATCACGATGC
CGCATGCTAGTCAGGCAGGGATATTATTAGCAGGTATCGGATGA
CAGCATTACGGCGGGAGTGCTATTATTATGAGCGCGAT
>identifiant2 commentaire libre
CAGGCAGGTTCTTATTATCGGCGGGCGGAGGCAGGATGCATC
CAGTGCAGTACGCTAGTCAGCGATGCATTATGACTGACTCAGTTT
CCCGCTAGCTATGCTATGCTATTGATCGATTGAGCTGATCTGGC
CAGCTATGCTTAGTA
```

# Format fastq

Descriptif du read (position sur la piste de séquençage, taille,...)

```
@SRR001666.1 071112_SLXA-EAS1_s_7:5:1:817:345 length=36
GGGTGATGGCCGCTGCCGATGGCGTCAAATCCCACC
+SRR001666.1 071112_SLXA-EAS1_s_7:5:1:817:345 length=36
IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII9IG9IC
```

Qualité (probabilité que la base soit correcte) encodé par code ASCII



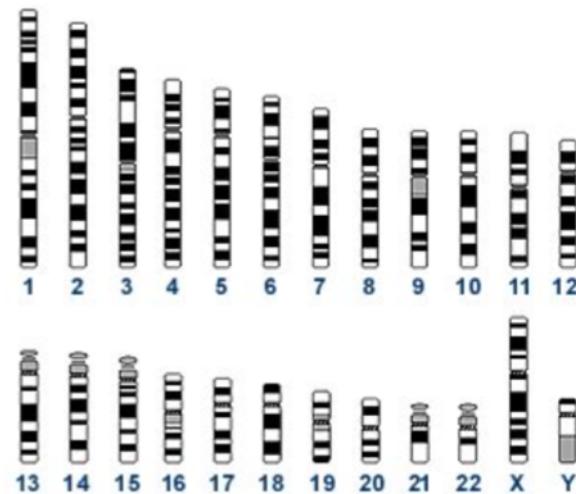
# Fichiers de régions

Coordonnées génomiques indiquant une région du génome

<chromosome>:<start>-<end>  
chr7:117465784-117715971

Formats de régions

- BED
- GTF/GFF: annotation de features dans le génome
- SAM/BAM: alignement de reads de séquence sur le génome
- VCF: variant calling file



# Format bed

obligatoire		<i>name</i>	<i>score</i>	<i>strand</i>	<i>Thick start</i>	<i>Thick end</i>	<i>color</i>	
chr7	127471196	127472363	Pos1	0	+	127471196	127472363	255,0,0
chr7	127472363	127473530	Pos2	0	+	127472363	127473530	255,0,0
chr7	127473530	127474697	Pos3	0	+	127473530	127474697	0,255,0
chr7	127474697	127475864	Pos4	0	+	127474697	127475864	255,0,255

## Attention

Le premier nucléotide est numéroté 0.

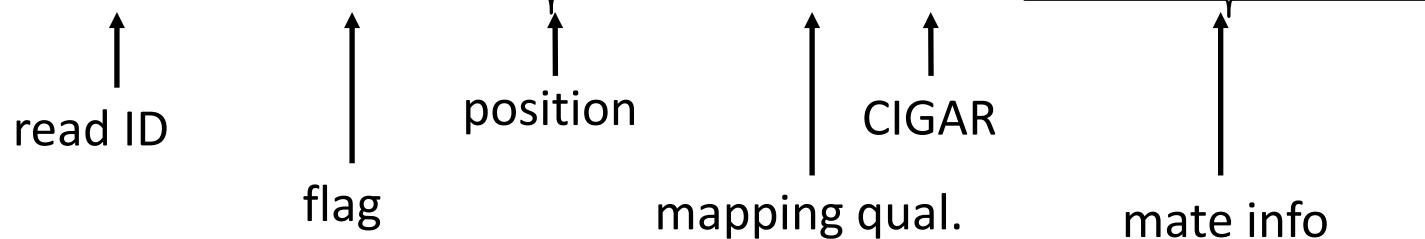
end - start = taille de la séquence



# Format SAM/BAM

Rappel BAM:

```
@RG  ID:group1  SM:1425_CD34  PL:ILLUMINA  LB:lib1 PU:unit1
@PG  ID:bwa  PN:bwa  VN:0.7.12-r1039 CL:bwa mem -M -t 2 -A 2 -E 1 -R @RG\tID:group1\tSM:1425_CD34\tPL:ILLUMINA\tLB:lib1\tPU:unit1 /root/myd
ERR166338.13782800  83  chr13  32890449  60  101M  =  32890343  -207  GGGACTGAATTAGAACAAATTTCAGCGCTT
ERR166338.13782800  163  chr13  32890343  60  75M  =  32890449  207  CACTAGCCACGTTCGAGTGCTTAATGTGGCTAGTGGC
ERR166338.26716588  99  chr13  32890406  60  101M  =  32890553  222  AATGTTCCCACCTCACAGTAAGCTGTTACCGTTCCAG
ERR166338.26716588  147  chr13  32890553  60  75M  =  32890406  -222  TTGCAGACTTACCAAGCATTGGAGGAATATCGTAA
ERR166338.27259961  99  chr13  32890496  60  101M  =  32890558  137  ACCTCAGTCACATAATAAGGAATGCATCCCTGTGTAAG
ERR166338.27259961  147  chr13  32890558  60  75M  =  32890496  -137  GACTTATTTACCAAGCATTGGAGGAATATCGTAGGTAA
ERR166338.63037998  99  chr13  32890496  60  101M  =  32890558  137  ACCTCAGTCACATAATAAGGAATGCATCCCTGTGTAAG
ERR166338.63037998  147  chr13  32890558  60  75M  =  32890496  -137  GACTTATTTACCAAGCATTGGAGGAATATCGTAGGTAA
```



# Le champ CIGAR

Example:

52M36890N45M3S

REF : chr20



## All Cigar operations

Op	BAM	Description
M	0	alignment match (can be a sequence match or mismatch)
I	1	insertion to the reference
D	2	deletion from the reference
N	3	skipped region from the reference
S	4	soft clipping (clipped sequences present in SEQ)
H	5	hard clipping (clipped sequences NOT present in SEQ)
P	6	padding (silent deletion from padded reference)
=	7	sequence match
X	8	sequence mismatch

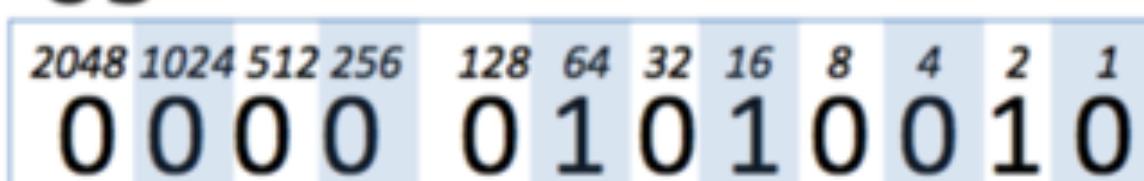
# Les Flags SAM

## Example:

- Decimal Flag Value

83

- Binary Flag Value



- To each bit corresponds a meaning

Bit	Description
1	0x1 template having multiple segments in sequencing
2	0x2 each segment properly aligned according to the aligner
4	0x4 segment unmapped
8	0x8 next segment in the template unmapped
16	0x10 SEQ being reverse complemented
32	0x20 SEQ of the next segment in the template being reverse complemented
64	0x40 the first segment in the template
128	0x80 the last segment in the template
256	0x100 secondary alignment
512	0x200 not passing filters, such as platform/vendor quality controls
1024	0x400 PCR or optical duplicate
2048	0x800 supplementary alignment

