

Mapping, Advanced & Assembly, basics

Francois Sabot & Guests

2022

DIADÉ unit UM, CIRAD, IRD

Reminders!

- Data from lab human-created

- Data from lab human-created
- Thus errors in exp and design

- Data from lab human-created
- Thus errors in exp and design

• **YOU ARE NOT MAGICIANS**

- Mapping will map only on what exists in reference !

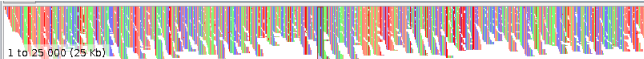
- Mapping will map only on what exists in reference !
- If not existing, mapping on the closest sequence

- Mapping will map only on what exists in reference !
- If not existing, mapping on the closest sequence
- Multiple mappings possible

- Mapping will map only on what exists in reference !
- If not existing, mapping on the closest sequence
- Multiple mappings possible
- Optimal mapping in info terms not the biological relevant one

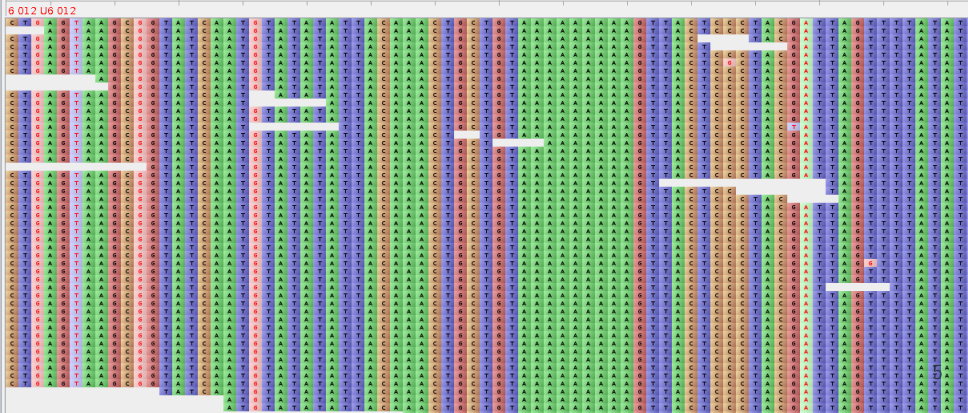
```
francois@xwing:/Data/Loagen/20220826_FS10001149_4_BRD91213-1131$ samtools flagstats 23_S3_L001_vs_Hum
3352415 + 0 in total (QC-passed reads + QC-failed reads)
0 + 0 secondary
152715 + 0 supplementary
0 + 0 duplicates
3144808 + 0 mapped (93.81% : N/A)
3199700 + 0 paired in sequencing
1599850 + 0 read1
1599850 + 0 read2
1450204 + 0 properly paired (45.32% : N/A)
2902572 + 0 with itself and mate mapped
89521 + 0 singletons (2.80% : N/A)
104250 + 0 with mate mapped to a different chr
41312 + 0 with mate mapped to a different chr (mapQ>=5)
```

Working with weird results - Part 2

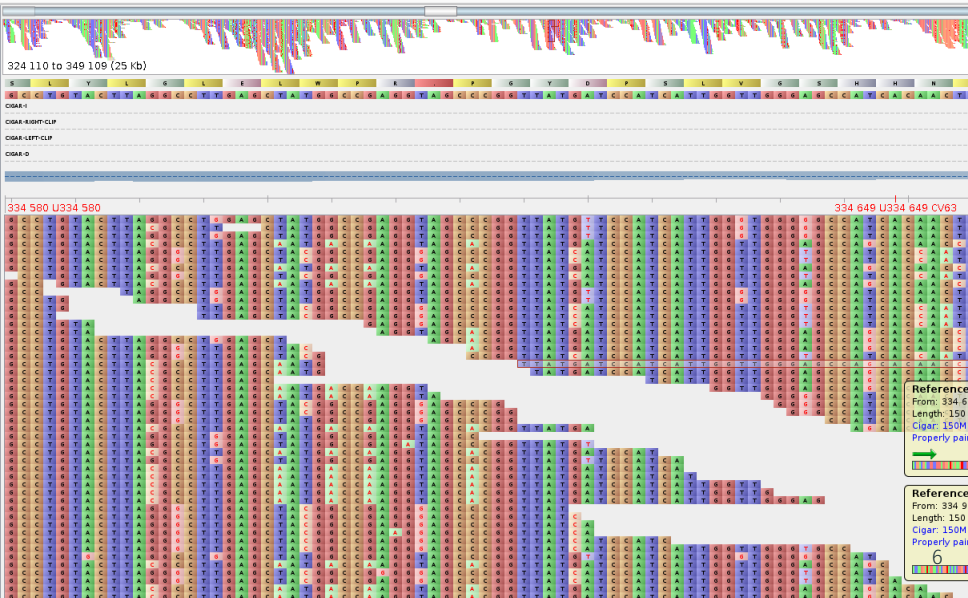


1 to 25 000 (25 kb)

CIGAR-I
CIGAR-LEFT-CLIP
CIGAR-RIGHT-CLIP
CIGAR-D



Institut de Recherche
 pour le Développement
 FRANCE

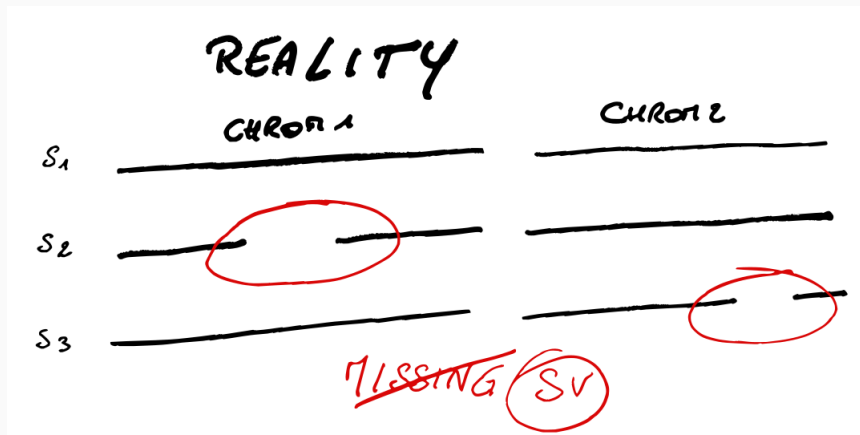


XXCHROM	POS	REF	ALT	-	S1	S2	S3
1	3000	A	T		0/0	0/0	./.
1	5600	G	C		1/1	./.	./.
2	10383	T	C		0/1	0/1	./.

CHROM	POS	REF	ALT	-	S1	S2	S3
1	3000	A	T		0/0	0/0	-/-
1	5600	G	C		1/1	-/-	-/-
2	10383	T	C		0/1	0/1	-/-

INFERENCE ↓ "MAGICAL"

CHROM	POS	REF	ALT	-	S1	S2	S3
1	3000	A	T		0/0	0/0	0/0
1	5600	G	C		1/1	1/1	1/1
2	10383	T	C		0/1	0/1	0/1



- Illumina sequencing Ok
- Mapping BWA - reference Ok
- SNP calling GATK

- Illumina sequencing Ok
- Mapping BWA - reference Ok
- SNP calling GATK
- **BUT** False Positive SNP

- Illumina sequencing Ok
- Mapping BWA - reference Ok
- SNP calling GATK
- **BUT** False Positive SNP
- **WHY ?**