Mapping



Requirements

Before diving into this slide deck, we recommend you to have a look at:

• Galaxy introduction



? Questions

- What two things are crucial for a correct mapping?
- What is BAM?



Objectives

- You will learn what mapping is
- A genome browser is shown that helps you to understand your data



What is mapping?



Mapping

• Search the position of a read in the genome





Sequence alignment

A A C C G C C T T
A G G G G C C T T
1 2 3 4 5 6 7 8 9



Sequence alignment



Looks easy...

Lego time: Who wants to volunteer?



- Know your data!
- Maybe not the "best" result is the one you are searching for
- Try different mappers



Tools, tools, tools

- Bowtie
- HISAT
- MAQ
- BWA
- Eland
- SOAP
- STAR
- BISMARCK

And many more!



SAM/BAM file format

```
chr1 497 37
                                             37M
1:497:R:-272+13M17D24M 113
                                                      15
                                                                100338662 0
CGGGTCTGACCTGAGGAGAACTGTGCTCCGCCTTCAG
                                    0;==-==9;>>>>>>>>>>>>>>>
                                                                        XT:A:U
                                                                                  NM: i:0
                                                                MD:Z:37
SM:i:37 AM:i:0
                                             X0:i:0
19:20389:F:275+18M2D19M
                                    chr1
                                             17644
                                                                37M
                                                                                  17919
         TATGACTGCTAATAATACCTACACATGTTAGAACCAT
                                             >>>>>>>>>>>>>
RG:Z:UM0098:1
                  XT:A:R
                           NM:i:0
                                    SM:i:0
                                             AM: i:0
                                                      X0:i:4
                                                                X1:i:0
                                                                        XM:i:0
                                                                                  X0:i:0
XG:i:0
         MD:Z:37
19:20389:F:275+18M2D19M
                            147
                                    chr1
                                             17919
                                                                18M2D19M =
17644
         -314
                  GTAGTACCAACTGTAAGTCCTTATCTTCATACTTTGT
                                                      ;44999;499<8<8<<<8<<><<7<;<<<>><
XT:A:R
         NM:i:2
                  SM:i:0
                           AM: i:0
                                    X0:i:4
                                             X1:i:0
                                                      XM:i:0
                                                               X0:i:1 XG:i:2
18^CA19
9:21597+10M2I25M:R:-209
                            83
                                    chr1
                                             21678
                                                                8M2I27M =
                                                      <;9<<5><<<>>>>>>
21469
         -244
                  CACCACATCACATATACCAAGCCTGGCTGTCTTCT
XT:A:R
         NM:i:2
                  SM:i:0
                           AM:i:0
                                    X0:i:5
                                                      XM:i:0
                                                               X0:i:1
                                                                        XG:i:2
                                             X1:i:0
```



SAM/BAM file format

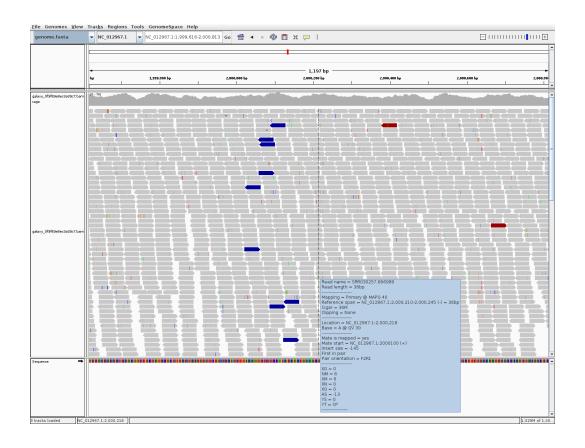
```
Read ID Read sequence
                                  Read position
                                      37M
                                              15
                                                      100338662 0
                              NM: i:0
                                      17644
                                                                     17919
       TATGACTGCTAATAATACCTACACATGTTAGAACCAT
                                                     ><<>>><<>><
RG:Z:UM0098:1
               XT:A:R
                                      AM:i:0
                                             X0:i:4
                                                     X1:i:0
                                                                     X0:i:0
               D19M 147 chr1 17919
GTAGTACCAACTGTAAGTCCTTATCTTCATACTTTGT
19:20389:F:275+18M2D19M
                                                     18M2D19M
                                              ;44999;499<8<8<<<8<<>><<<>><7<;<<<>><<
18^CA19
9:21597+10M2I25M:R:-209
                       83
                              chr1
                                      21678
                                                     8M2I27M
               XT:A:R NM:i:2
                                      X1:i:0
                                             XM:i:0
```

SAM: Sequence Alignment/Map format

BAM: Binary SAM



IGV Browser -- Hands-on





Key points

- Know your data!
- Mapping is not trivial
- There are many mapping algorithms, it depends on your data which one to choose



Thank you!

This material is the result of a collaborative work. Thanks the Galaxy Training Network and all the contributors (Joachim Wolff)!



Found a typo? Something is wrong in this tutorial? Edit it on GitHub

