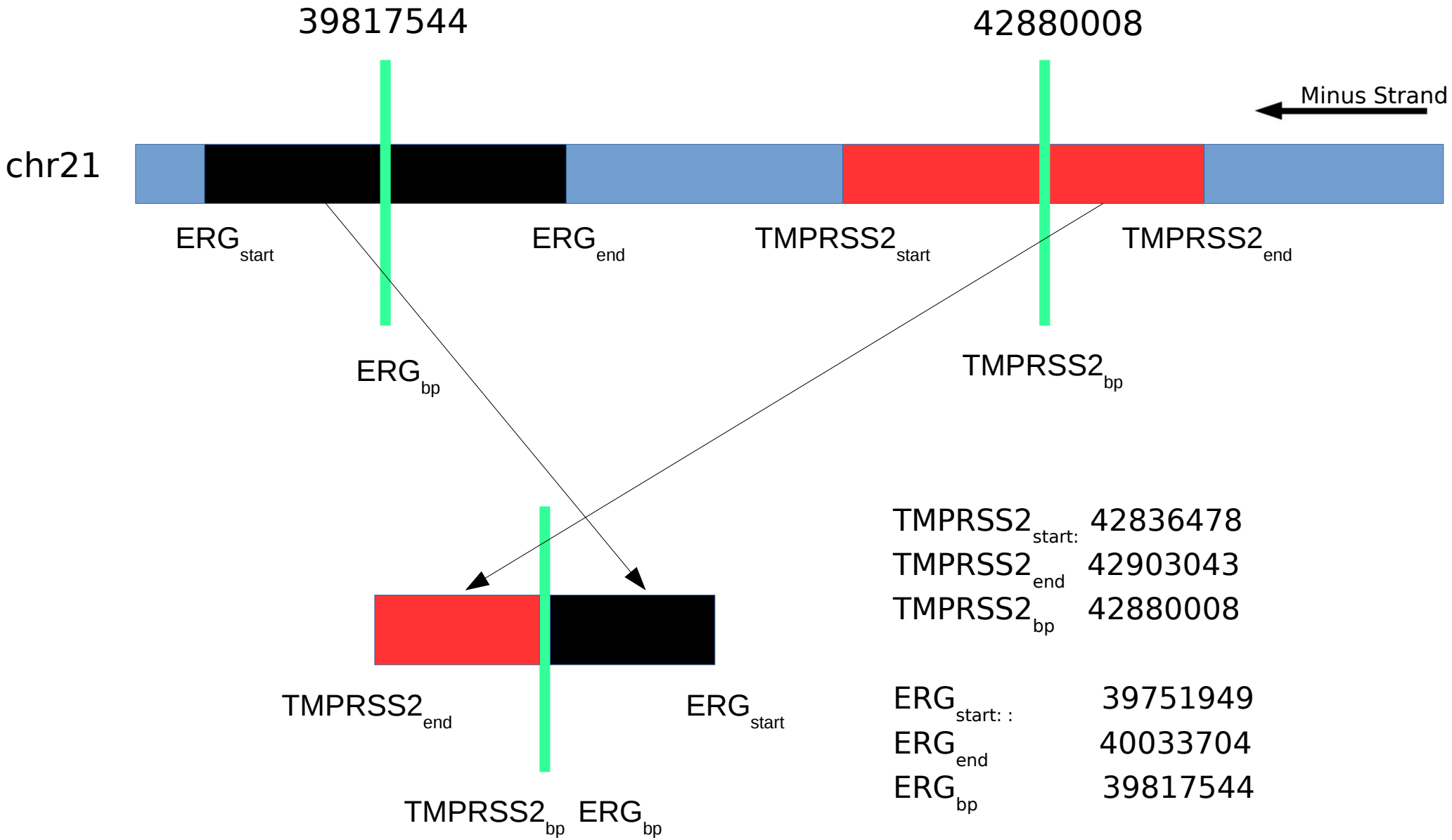
BIOINFORMATICS

LAB 2: 2nd Exercise

PACIELLO Giulia
May, 5-6th 2016

FUSION MECHANISM

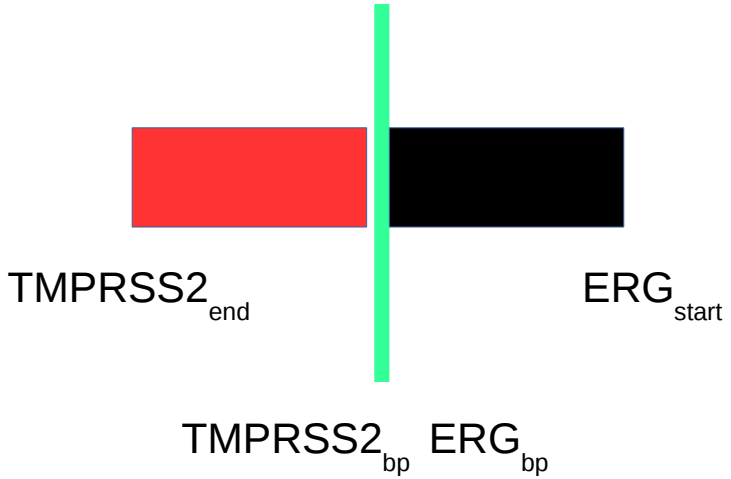


TMPRSS2 start: 42836478
TMPRSS2 end 42903043
TMPRSS2 bp 42880008

ERG start: : 39751949
ERG end 40033704
ERG bp 39817544

EXERCISE OBJECTIVE

1) Build the virtual reference for TMPRSS2-ERG gene fusion;



2) Map with bowtie2 dataset_mate1.fq and dataset_mate2.fq on the virtual reference;

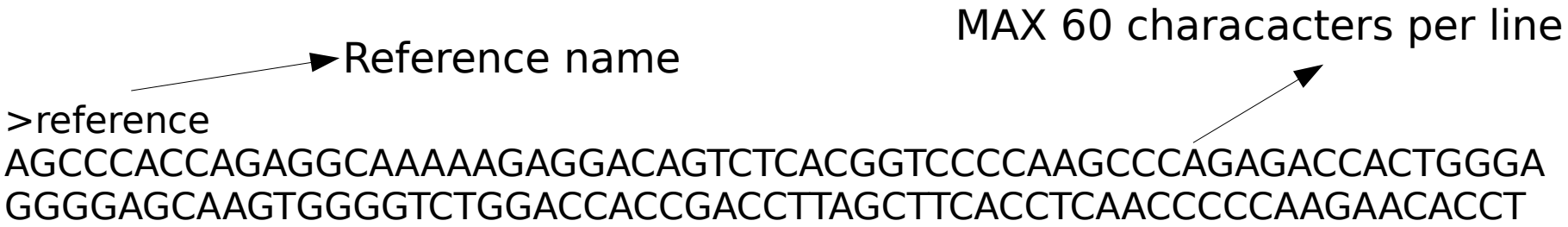
3) Extract mates mapped on TMPRSS2 or ERG AND not spanning the fusion breakpoint.

BUILD THE VIRTUAL REFERENCE

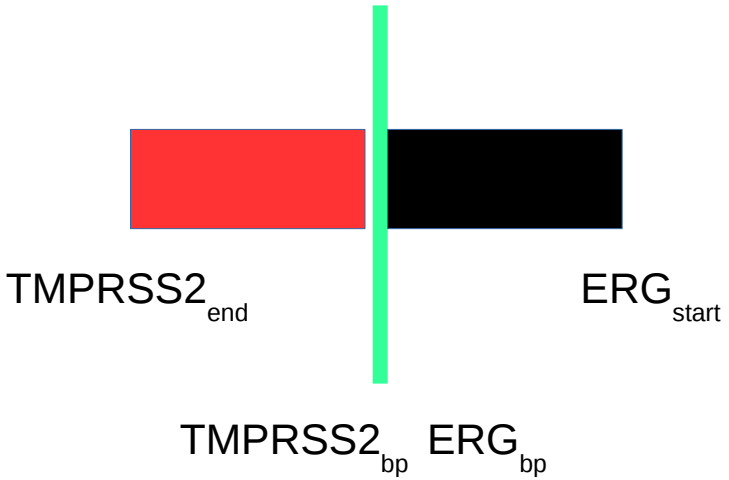
1) Extract from align_index.fa the sequences of the two partner gens:

- Extract chr21 sequence by identifying in the file '>chr21';
- For TMPRSS2 extract the portion of chr21 from the BP (42880007) to the end of the gene (42903043). Reverse and complement end to end this sequence;
- For ERG extract the portion of chr21 from the start of the gene (39751948) to the BP (39817544). Reverse and complement end to end this sequence;
- Concatenate TMPRSS2 and ERG sequences.

2) Build a fasta file by formatting TMPRSS2-ERG sequence:



FROM GENOMIC TO VIRTUAL COORDINATES

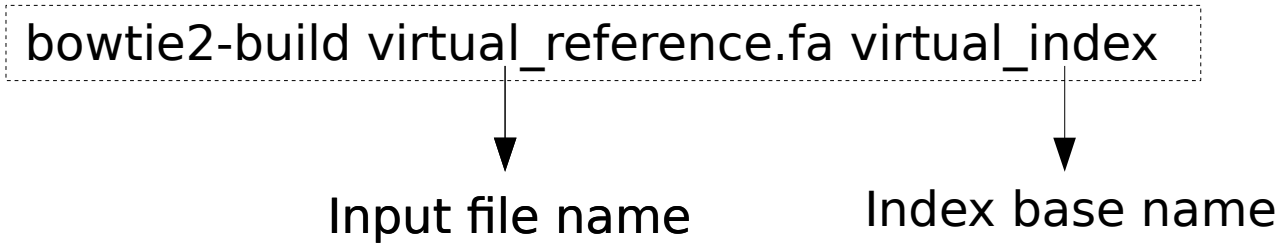


$TMPRSS2_{end}$ -$TMPRSS2_{bp}$ = 42903043-42880007 = 23036

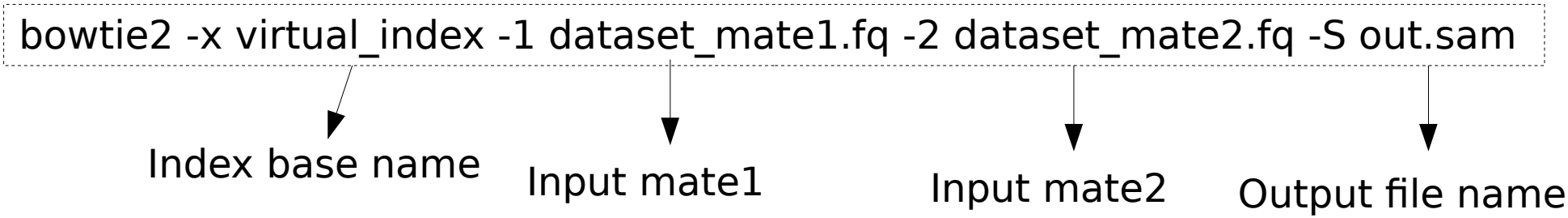
The breakpoint is located at coordinate 23036 on your virtual reference

MAP READS USING BOWTIE2

1) Build the index of the virtual reference using bowtie2 indexer bowtie2-build:



2) Run bowtie2 with default parameters:



http://bowtie-bio.sourceforge.net/bowtie2/manual.shtml#the-bowtie2-aligner

MAPPED READS ELABORATION (1)

1) Look at the out.sam file:

chr10-94220 89 reference 17854 0 7M1D46M = 17854 0
TATTTTTGTAGAGACGGGGTTTCACCATGTCGGCCAGGCTGGTCTCAAACCCC @.5CDE1,?
7?CB)/=D/B@?-@A?6(6>=&B6;@A:A?@>=>=1&=6=>&>= AS:i:-25 XN:i:0
XM:i:5 XO:i:1 XG:i:1 NM:i:6 MD:Z:7^A24A1T11T0G3T2 YT:Z:UP
chr10-94220 133 reference 17854 0 * = 17854 0
CAGCAAGACCGGGCCAAATTCATCATCAGCAGAAACCTGCTTCCTTGATGGAA :=.;;9':)2?
@.(A?6C=->AA@EB?B,A?;8@EB*'3//,'4.@A7?591) YT:Z:UP

MAPPED

chr10-94330 77 * 0 0 * * 0 0
ACGTGAAAGCGCAGAAAGGGAACAACAACAACAGCAACTCCTGCACATAATGC
9;===4=@(?>?'B@C??@7@9?B?),?B;A'A?>;@B:@@.>?,;??)8C@= YT:Z:UP
chr10-94330 141 * 0 0 * * 0 0
TAATATCACTATTGCTAAAACATCTTAACTAGCTTGTTCTTAGTTTGAAACGG
;=.;=;91<;@EA;?@@==BD1&=B@>6?7B'&E=@>@C)/B)'/?@B,1?,7 YT:Z:UP

UNMAPPED

MAPPED READS ELABORATION (2)

1) Convert out.sam to out.bam:

```
samtools view -bS out.sam > out.bam
```

2) Sort out.bam:

```
samtools sort out.bam out_sorted
```

3) Create out.bam index file

```
samtools index out_sorted.bam out_sorted.bai
```

MAPPED READS ELABORATION (3)

1) Extract thanks to samtools view program all the mates mapped on to TMPRSS2 and ERG not spanning the fusion breakpoint.