WP3
10.12.21

Task done this week

- Fixed BINDetect error with test data
- Fixed the indexing problem when running TOBIAS with the individual clusters
  - More or less successful runs
- Communication with WP1/2 about input data from them

Problems

- BINDetect did not find any TFBS
  - wrong peak input
    - _summits.bed instead of .narrowPeaks (currently still running)
- File format of previous WPs
  - Barcode not added as a tag but in the first column instead (.bam)

Tasks for next week

- Comparing clusters
- Working with WP1/2s data
  - Adaptation of input file format
    (which WP adjusts?)

```
2DRXX:2:1102:19994:21371 1171    chr1    901767  60      49M     =
 XS:i:19 CR:Z:TTTGGTTGTTCTGAAC   CY:Z:FFFFFFFFFFFFF:FF   CB:Z:TTTGGTTGTTC
2DRXX:1:2131:11803:31923 1171    chr1    901774  60      46M     =
 CR:Z:TCACCTGAGCACCATT   CY:Z:FFFFF:F:FFFFFF:F   CB:Z:TCACCTGAGCACCATT-1
```

  - Making the script work with
    different types of input
    (argparse, i.e. which tag to look for)
    - Have to know the options that can occur beforehand

BAM Barcode Tags

Chromium cellular and molecular barcode information for each read is stored in the following `TAG` fields:

Tag	Type	Description
CB	Z	Chromium cellular barcode sequence that is error-corrected and confirmed against a list of known-good barcode sequences.
CR	Z	Chromium cellular barcode sequence as reported by the sequencer.
CY	Z	Chromium cellular barcode read quality. Phred scores as reported by sequencer.
UB	Z	Chromium molecular barcode sequence that is error-corrected among other molecular barcodes with the same cellular barcode and gene alignment.
UR	Z	Chromium molecular barcode sequence as reported by the sequencer.
UY	Z	Chromium molecular barcode read quality. Phred scores as reported by sequencer.
TR	Z	Trimmed sequence. For the Single Cell 3' v1 chemistry, this is trailing sequence following the UMI on Read 2. For the Single Cell 3' v2 chemistry, this is trailing sequence following the cell and molecular barcodes on Read 1.

The cell barcode `CB` tag includes a suffix with a dash separator followed by a number:

```
AGAATGGTCTGCAT-1
```