

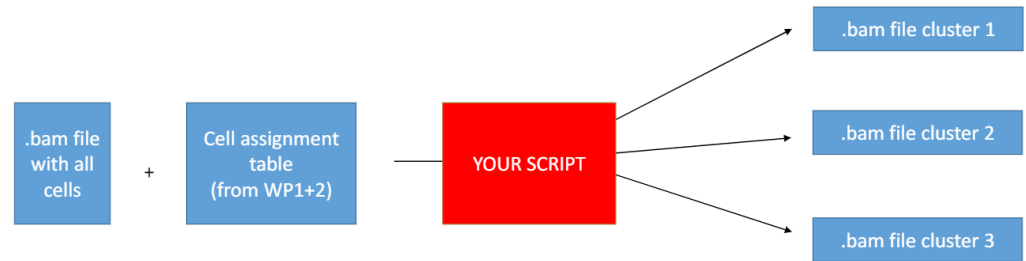
WP3

Approach

- Obtain ATAC data per cluster
 - .bam file with all cells + cell assignment table from WP 1+2
 - bam files for each cluster
- Run TOBIAS footprinting for each cluster
- Comparison between clusters + normalization

Splitting the .bam file into clusters

- Reads have barcodes
- Script to divide the file based on these bar codes and the assignment table
- Input:
 - ATAC of all cells (.bam)
 - barcodes and clusters (.tsv)
- Output: one .bam file per cluster



Test data

A00519:269:H7M2DRXX:1:2173:13060:9831 7 3745 29406174 0 49M = 29406294 170 CTACTGTAGTAAGAAGAAATCACTTCATCAAACCAACGAAGCATT
 FFFFFFFFHHHFFFFFFFFFFFFFFFFFFFFF FFFFFFFF NM:i:0 MD:Z:49 MC:Z:50M AS:i:149 Xs:i:49 CR:Z:Z:**AAACGAAAGACACGGT** CY:Z::F::FF,FFFF CB:Z:**AAACGAAAGACACGGT**-1
 BC::GGTTGATG QT:Z::FF,FF GP:I:983899258 MQ:i:0 RG:Z:atac_pbm_10k_v1:MissingLibrary:1:H7FM2DRXX:1
 A00519:269:H7M2DRXX:1:2173:13060:9831 107 3745 29406294 0 50M = 29406174 -170 AAACAATCTTCGCAGATCTCGAAGTGATATTTGGCATCTCTCGAGG
 FFFFFFFFHFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF NM:i:2 MD:Z:14T25C9 MC:Z:49M AS:i:40 Xs:i:40 CR:Z:Z:**AAACGAAAGACACGGT** CY:Z::F::FF,FFF,FFFFFF
 CB:Z:**AAACGAAAGACACGGT**-1 BC:Z:GGTTGATG QT:Z::FF,FF GP:I:983899258 MQ:i:0 RG:Z:atac_pbm_10k_v1:MissingLibrary:1:H7FM2DRXX:1
 A00519:269:H7M2DRXX:2:2245:24117:20932 1123 3745 30061645 40 50M = 30061681 265 AGAGTAGGACGCGGAGATAGTAGGGAATCGGACCCTCTCCATCA
 :FFFFFFFH:FFFFFFFHHHFFFFFFFFFFFFFFFFFFFFF NM:i:0 MD:Z:50 MC:Z:50M AS:i:50 Xs:i:50 XA:Z:chr1,-149223366,50M,0; CR:Z:**AAACGAAAGACACGGT**
 CY:Z::F::FFFFFFF:CB:Z:**AAACGAAAGACACGGT**-1 BC:Z:GGTTGATG QT:Z:FFFFFFFH GP:I:984554559 MQ:i:1984554824 MQ:i:45 RG:Z:atac_pbm_10k_v1:MissingLibrary:1:H7FM2DRXX:2
 A00519:269:H7M2DRXX:2:2245:24117:20932 1171 3745 30061681 45 49M = 30061645 -265 GGGACATAAGGCCAGCAAGGCCCTGCTTCCCAAGAACGACC
 FFFFFFFHHHFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF NM:i:0 MD:Z:49 MC:Z:50M AS:i:149 Xs:i:47 CR:Z:**AAACGAAAGACACGGT** CY:Z::F::FFFFFF:FFFF CB:Z:**AAACGAAAGACACGGT**-1
 XA:Z:chr1,-147601177,49M,1:chr1,-143672664,49M,1:chr1,-147834056,43M6S,0:chr1,-143646568,43M6S,0; CR:Z:**AAACGAAAGACACGGT**
 BC:Z:GGTTGATG QT:Z:FFFFFFFH GP:I:984554559 MQ:i:40 RG:Z:atac_pbm_10k_v1:MissingLibrary:1:H7FM2DRXX:2
 A00519:269:H7M2DRXX:1:1228:22643:13191 147 3745 30791687 7 49M = 30791689 -47 AGAGGATGCCATTGGAGGGCAAGTCTGGTCCAGCAGCACGGTAATCT
 FFFFFFFHHHFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF NM:i:1 MD:Z:10C7 MC:Z:6844M AS:i:147 Xs:i:146 CR:Z:**AAACGAAAGACACGGT** CY:Z::F::FFFFFFF
 CB:Z:**AAACGAAAGACACGGT**-1 BC:Z:GGTTGATG QT:Z:FFFFFFFH GP:I:985284650 MQ:i:1985284597 MQ:i:7 RG:Z:atac_pbm_10k_v1:MissingLibrary:1:H7FM2DRXX:1
 A00519:269:H7M2DRXX:1:1228:22643:13191 99 3745 30791689 47 CTTTAGAGGATCGATTGGAGGGCAAGTCTGGTCCAGCAGCACGGTA
 FFFFFFFHHHFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF NM:i:0 MD:Z:44 MC:Z:49M AS:i:144 Xs:i:145 CR:Z:Z:**AAACGAAAGACACGGT** CY:Z::F::FFFFFFF:FFFF CB:Z:**AAACGAAAGACACGGT**-1
 BC:Z:GGTTGATG QT:Z:FFFFFFFH GP:I:985284650 MQ:i:7 RG:Z:atac_pbm_10k_v1:MissingLibrary:1:H7FM2DRXX:1
 A00519:269:H7M2DRXX:1:1133:9064:13463 99 3745 3182129 40 50M = 3182130 50 GGGATGGAATGGGAAGATGATCTTCCCCGGGAGCTTGCATCCAGAGGC
 FFFFFFFHHHFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF NM:i:0 MD:Z:50 MC:Z:49M AS:i:50 Xs:i:50 CR:Z:Z:**AAACGAAAGACACGGT** CY:Z::F::FF:F,F,FFFF CB:Z:**AAACGAAAGACACGGT**-1
 BC:Z:AAGCATAA QT:Z:FFFFFFFH GP:I:985675043 MQ:i:1985675093 MQ:i:0 RG:Z:atac_pbm_10k_v1:MissingLibrary:1:H7FM2DRXX:1
 A00519:269:H7M2DRXX:1:1133:9064:13463 147 3745 3182130 49M = 3182129 -50 GGATGGAATGGGAAGATGATCTTCCCCGGGAGCTTGCATCCAGAGGC
 FFFFFFFHHHFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF NM:i:0 MD:Z:49 MC:Z:50M AS:i:149 Xs:i:149 CR:Z:Z:**AAACGAAAGACACGGT** CY:Z::F::FF,FF,FFFF CB:Z:**AAACGAAAGACACGGT**-1
 BC:Z:AAGCATAA QT:Z:FFFFFFFH GP:I:985675093 MQ:i:1985675043 MQ:i:0 RG:Z:atac_pbm_10k_v1:MissingLibrary:1:H7FM2DRXX:1
 A00519:269:H7M2DRXX:1:1118:8205:32080 99 3745 34975365 40 50M = 34975558 242 CAGAGCCCCGGGAATCTGGAGGGGCGTGGGGCGAGGGGGGGGGCTGC
 FFFFFFFHHHFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF NM:i:0 MD:Z:50 MC:Z:49M AS:i:50 Xs:i:50 CR:Z:**AAACGAAAGACACGGT** CY:Z::F::FF:F,F,FFFF CB:Z:**AAACGAAAGACACGGT**-1
 BC:Z:COCATCGC QT:Z:FFFFFFFH GP:I:989468279 MQ:i:1989468521 MQ:i:160 RG:Z:atac_pbm_10k_v1:MissingLibrary:1:H7FM2DRXX:1
 A00519:269:H7M2DRXX:1:1118:8205:32080 147 3745 34975558 60 49M = 34975365 -242 AAGTGGGGCCGGGAAGTCAGCATGACCTGAGTCCAGCGGAGCAGG
 FFFFFFFHHHFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF NM:i:0 MD:Z:49 MC:Z:50M AS:i:149 Xs:i:10 CR:Z:Z:**AAACGAAAGACACGGT** CY:Z::F::FF:F,F,FFFF CB:Z:**AAACGAAAGACACGGT**-1
 BC:Z:CCCCCTG QT:Z:FFFFFFFH GP:I:989468521 MQ:i:1989468279 MQ:i:160 RG:Z:atac_pbm_10k_v1:MissingLibrary:1:H7FM2DRXX:1
 A00519:269:H7M2DRXX:2:2161:31358:4163 77 0 0 GTCCAGCAGCGCTGCACCACTTCCCTATCTGTAGACAACCCACATACA
 FFF:FFFFFF,FFF:FFFF:,F,F,F,,F,AS:i:Xs:i:0 CR:Z:**AAACGAAAGACACGGT** CY:Z::FFF,FFF:,F,E,,: CB:Z:**AAACGAAAGACACGGT**-1 BC:Z:COCATCGC QT:Z:FFF,,F,F
 GP:i:-1 MP:i:-1 MQ:i:0 RG:Z:atac_pbm_10k_v1:MissingLibrary:1:H7FM2DRXX:2
 A00519:269:H7M2DRXX:2:2161:31358:4163 141 * 0 * 0 CAAGCGGCGATGTTGAGTTGTTCCGCAAGCTCGAATAGTCTCTCC
 ,F,F,FF,,,FF,,,F,F,,,F,F,,,F,AS:i:Xs:i:0 CR:Z:**AAACGAAAGACACGGT** CY:Z::FFF,FFF:,F,E,,: CB:Z:**AAACGAAAGACACGGT**-1 BC:Z:COCATCGC QT:Z:FFF,,F,F
 GP:i:-1 MP:i:-1 MQ:i:0 RG:Z:atac_pbm_10k_v1:MissingLibrary:1:H7FM2DRXX:2
 A00519:269:H7M2DRXX:2:2170:32217:11600 77 * 0 * 0 GTGATCCAGGAGAGACATCAGGCTGACTTTGTTGGTGGCAACAGAA
 FFFFFFFHHHFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF NM:i:0 MD:Z:49 MC:Z:50M AS:i:Xs:i:0 CR:Z:**AAACGAAAGACACGGT** CY:Z::FFFFFFF,FFFF CB:Z:**AAACGAAAGACACGGT**-1 BC:Z:COCATCGC QT:Z:FFFFFFF
 GP:i:-1 MP:i:-1 MQ:i:0 RG:Z:atac_pbm_10k_v1:MissingLibrary:1:H7FM2DRXX:2
 A00519:269:H7M2DRXX:2:2170:32217:11600 141 * 0 * 0 GAGGTAAGGSCACAAAATATCAGAGCTTAATGACACCAATATACAA
 FFFFFFFHHHFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF NM:i:0 MD:Z:49 MC:Z:50M AS:i:Xs:i:0 CR:Z:**AAACGAAAGACACGGT** CY:Z::FFFFFFF,FFFF CB:Z:**AAACGAAAGACACGGT**-1 BC:Z:COCATCGC QT:Z:FFFFFFF
 GP:i:-1 MP:i:-1 MQ:i:0 RG:Z:atac_pbm_10k_v1:MissingLibrary:1:H7FM2DRXX:2

Barcode	Cluster	
AAACGAAAGACACGGT-1		2
AAACGAAAGAGGTGGG-1		10
AAACGAAAGCACGTTAG-1		14
AAACGAAAGCGCTCTAC-1		3
AAACGAAAGACTTTCCG-1		2
AAACGAAAGGCGTCTCT-1		1
AAACGAAAGGCTTTATC-1		4
AAACGAAAGTGTTATG-1		14
AAACGAAAGTTTGGA-1		2
AAACGAACAAACGACG-1		2
AAACGAACAATTGCCA-1		4
AAACGAACACTCGTGT-1		12
AAACGAACACGGTGTC-1		14
AAACGAACATCGAGAA-1		12
AAACGAAGTCACTGT-1		6
AAACGAAGTGATTCCA-1		10
AAACGAAGTTTGACCA-1		11
AAACGAATCCAACGCG-1		2
AAACGAATCCATCTAT-1		2
AAACGAATCCCTTAAG-1		3
AAACGAATCTCTCAAGA-1		11
AAACGAATCGCATAAC-1		11
AAACGAATCGGGACAA-1		9
AAACGAATCTGAACTG-1		2
AAACGAATCTGTCGTC-1		7
AAACTCGAGAAAGCAG-1		2
AAACTCGAGAGCTGCC-1		13
AAACTCGAGAGTGGTC-1		7
AAACTCGAGCGAATAT-1		4
AAACTCGAGCGTGT-1		9
AAACTCGAGGTACATA-1		2
AAACTCGAGTCGCGTC-1		4
AAACTCGAGTCTTAGC-1		4
AAACTCGAGTCTCAAGA-1		12

Running TOBIAS

- Requirements
 - .bam files of each cluster
 - Depending on tool:
 - reference genome (.fasta)
 - peaks (.bed)
 - motifs (.pfm / .jaspar / .meme)
 - output of previous tools in the pipeline
- Main output: footprint score → heatmap to compare clusters

