Bioinformatics Workflow
- specificity of ATAC-seq -

Carl Herrmann

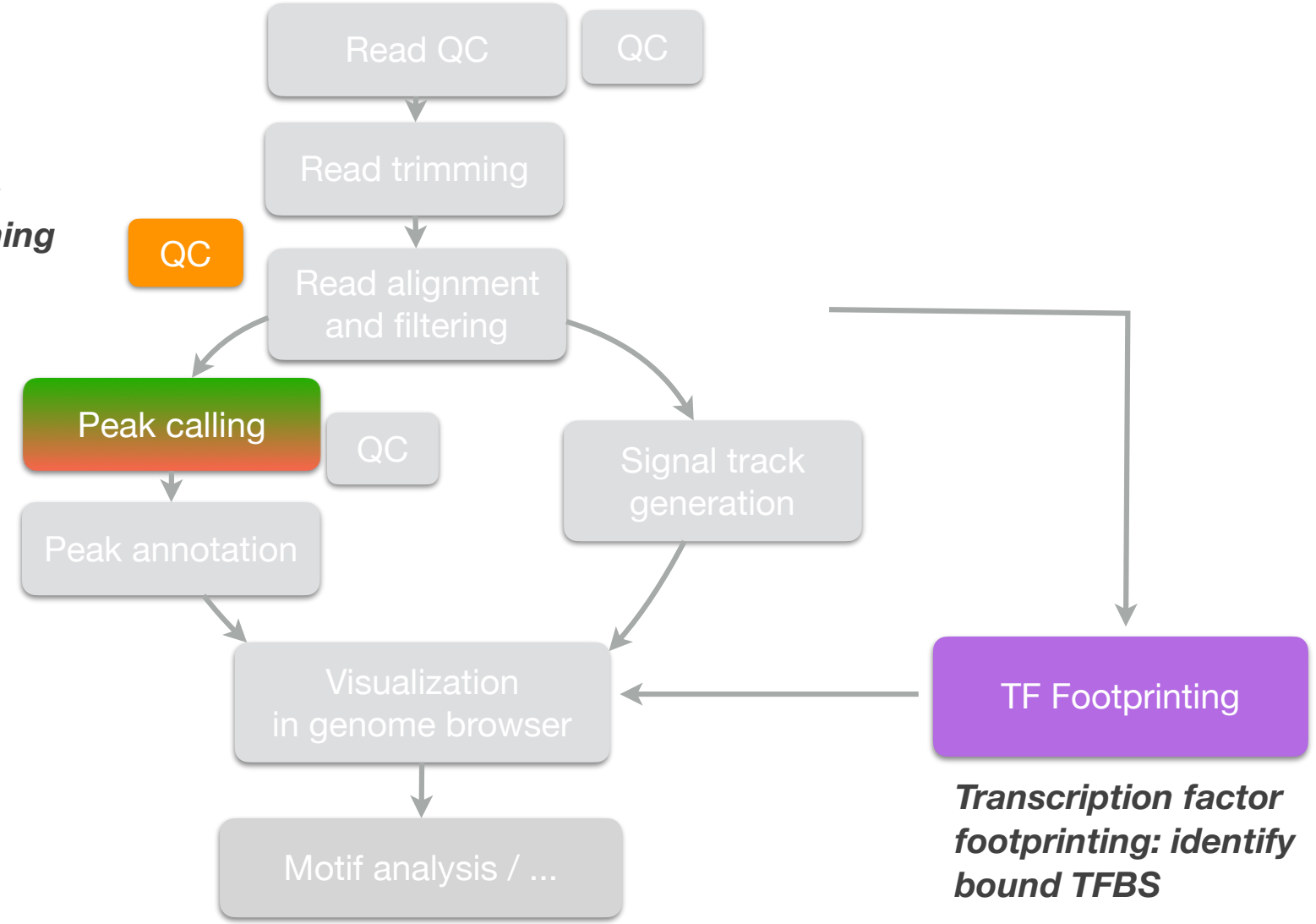
Health Data Science Unit - Medizinische Fakultät Heidelberg

General ATAC-seq Workflow

Medizinische Fakultät Heidelberg

Additional QC steps:
- ***mitochondrial reads***
- ***nucleosome patterning***

Different peak calling
- ***paired-end***
- ***no control available!***



Transcription factor footprinting: identify bound TFBS

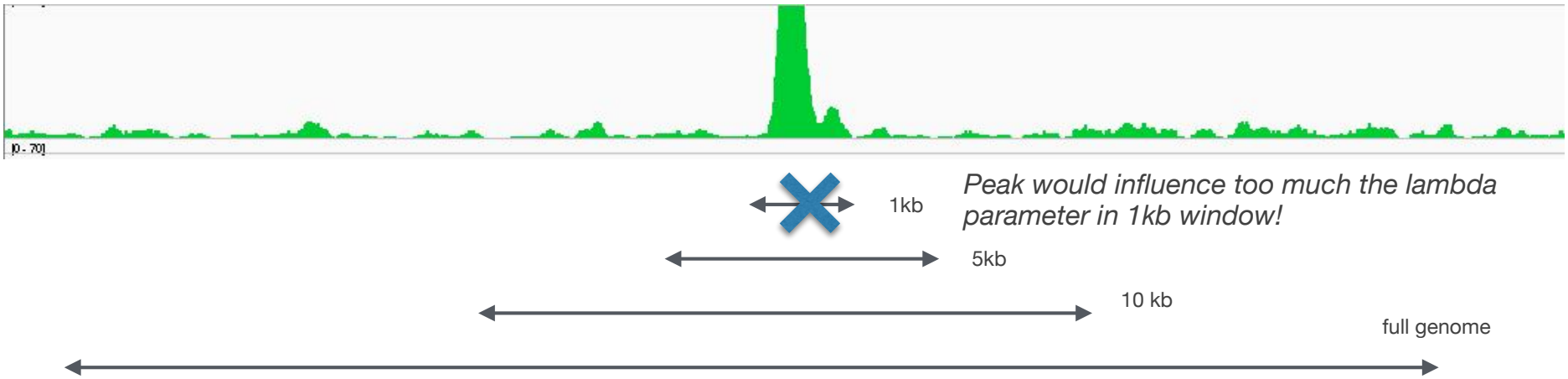
Peak calling

- MACS2 can be used to perform peak calling for ATAC-seq data
- Difference with ChIP-seq
  - ATAC-seq is mostly **paired end**, ChIP-seq (still) mostly single-end
  - ATAC-seq has **no control dataset**!
- Background level must be estimated from the ATAC-seq dataset itself.

Peak calling

Medizinische Fakultät Heidelberg

- Identification of local noise parameter from signal file
  - slide a window of size 2*d across signal
  - at each position, estimate parameter λ_{local} of Poisson distribution **USING A RANGE OF 5kb/10kb**



estimate parameter λ_{local} over different ranges, take max.

[Zhang et al. , 2008]

Peak calling

Medizinische Fakultät Heidelberg

- **MACS2**: typical command for ATAC-seq

```
macs2 callpeak \
--treatment atac.bam \
--name ATAC-Rep1 \
--format BAMPE \
--nomodel \

--keep-dup all \
--gsize 2.7e9 \
--qvalue 0.05 \
--outdir ATAC
```

bam file with IP

name of the experiment (choose freely!)

format of input files (BAM = single-end; BAMPE = paired-end)

do not determine fragment length;
use fragment length from paired-end bam

should duplicate read be kept? (auto / all)
effective (= mappable) genome size

FDR threshold to call a peak

output directory

Hands on:
from reads to peak for ATAC-seq

https://hdsu-bioquant.github.io/chipatac2020/01_ATAC_Intro.html
https://hdsu-bioquant.github.io/chipatac2020/02_ATAC_ReadQC.html
https://hdsu-bioquant.github.io/chipatac2020/03_ATAC_Trimming.html
https://hdsu-bioquant.github.io/chipatac2020/04_ATAC_Alignment.html
https://hdsu-bioquant.github.io/chipatac2020/05_ATAC_PeakCalling.html
https://hdsu-bioquant.github.io/chipatac2020/06_ATAC_PeakAnnotation.html

Bioinformatics Workflow
- footprinting ATAC-seq -

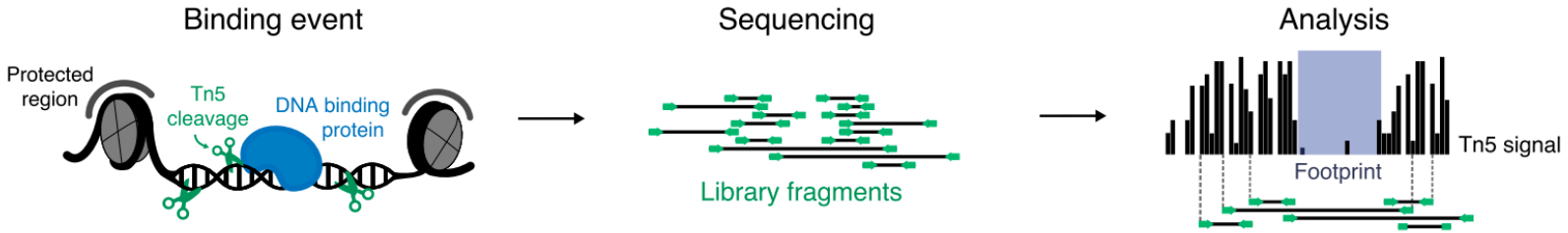
Carl Herrmann

Health Data Science Unit - Medizinische Fakultät Heidelberg

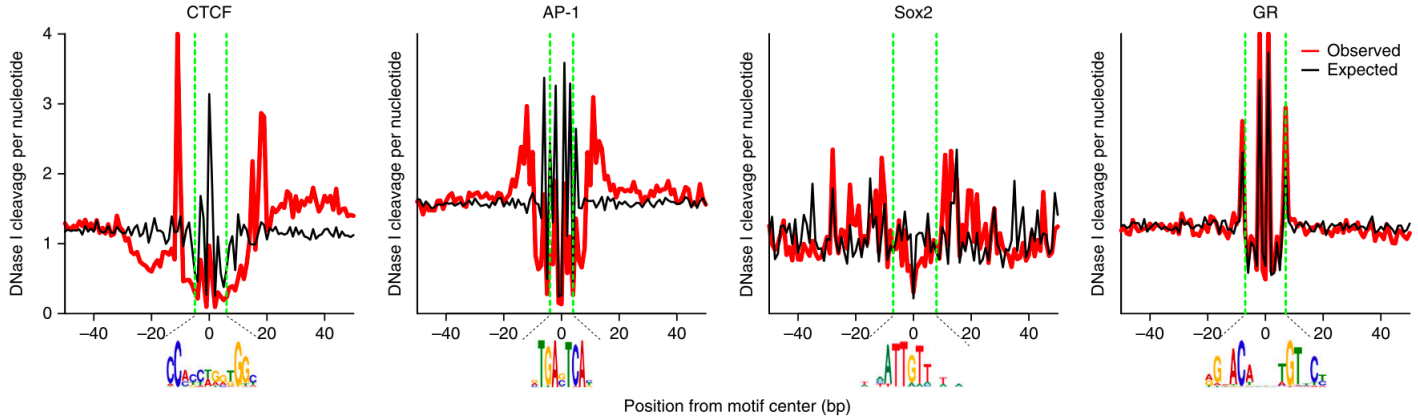
ATAC-seq footprinting

Medizinische Fakultät Heidelberg

- Use ATAC-seq profile to highlight specific TF binding events



[Bentsen, Nature Com., 2020]



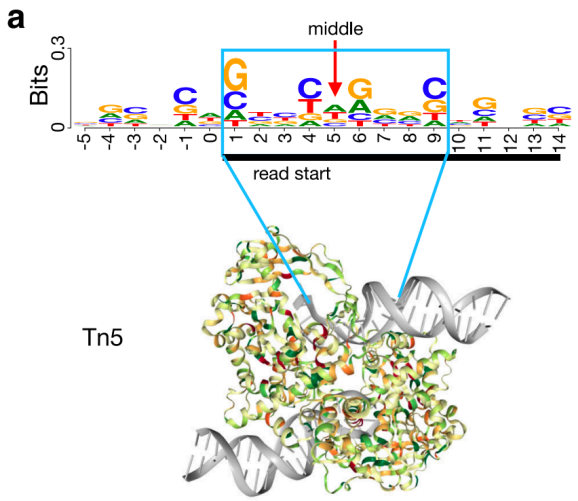
[Sung, Nature Method, 2016]

- Average profiles around predicted TFBS
- Different profiles depending of TF (shallow / deep)

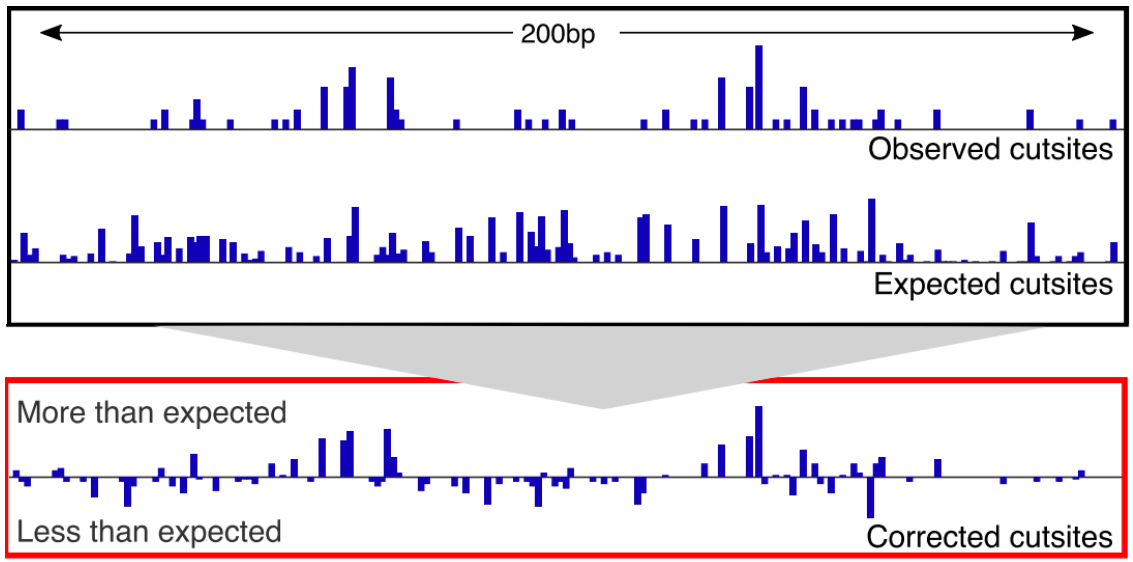
ATAC-seq footprinting

Medizinische Fakultät Heidelberg

- **Challenge**
  - the Tn5 has an insertion bias which needs to be corrected for; if not, false positive/negative predictions!



[Li et al., Genome Biol. 2019]



[Bentsen, Nature Com., 2020]

ATAC-seq footprinting

Medizinische Fakultät Heidelberg

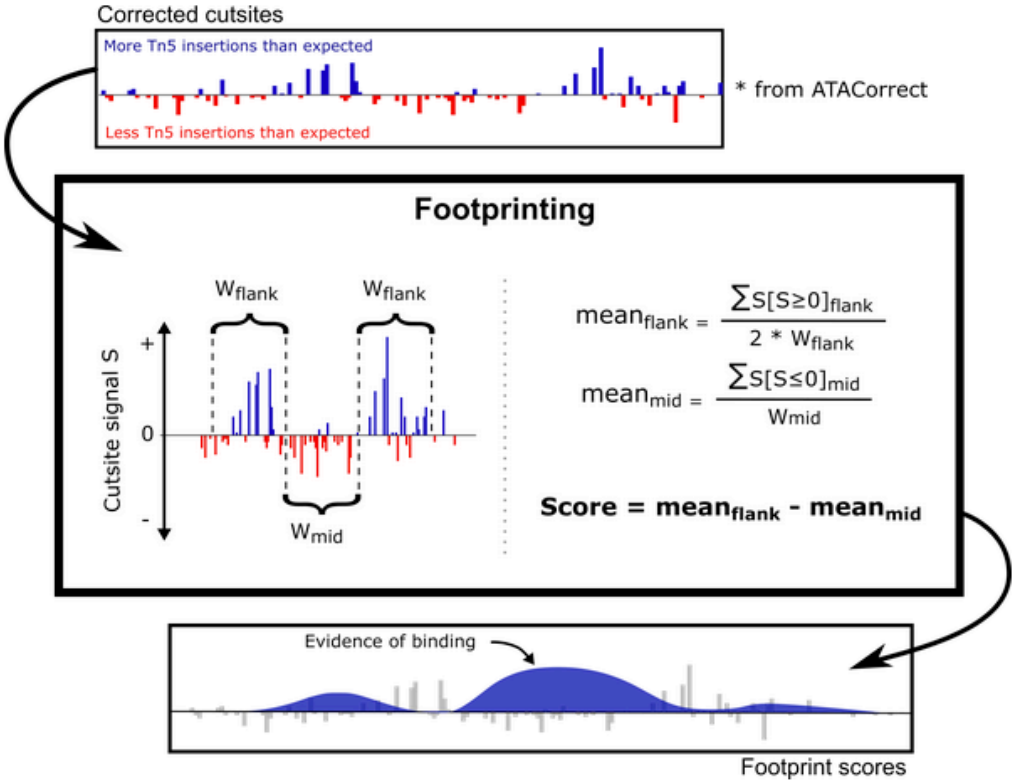
- Many methods published
  - HINT-ATAC [Li. et al, Genome Biology 2019]
  - PIQ [Sherwood et al., Nature Biotech, 2014]
  - TOBIAS [Bentsen et al., Nature Comm. 2020]
  - DeFCOM [Quach, Furey, Bioinformatics 2016]

- With sufficiently high sequencing depth, footprinting can be used as a proxy for TF specific ChIP-seq
- Recommended minimal number of aligned reads
  - Peak calling: 20 million reads
  - Footprinting: 50-100 million reads (depending on the TF)

ATAC-seq footprinting

Medizinische Fakultät Heidelberg

- **TOBIAS workflow**
  - determine insertion preference matrix; determine expected cut-sites; correct (tool: **ATACorrect**)
  - Score corrected profile for probability of TF binding (tool: **ScoreBigWig**)

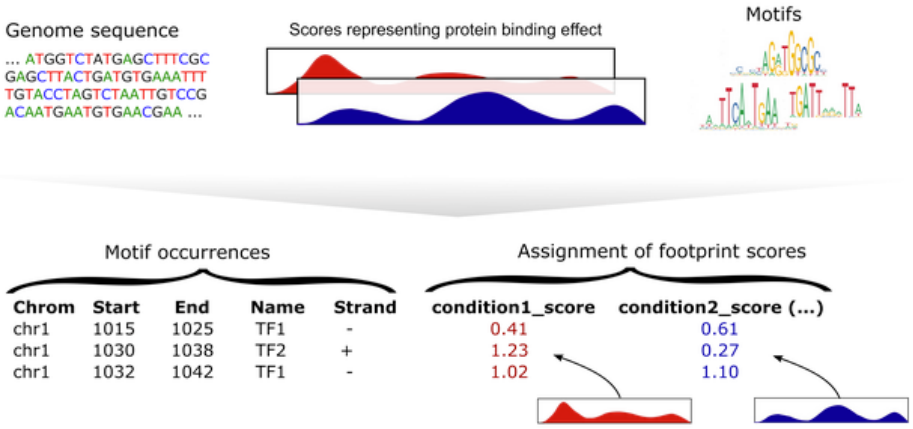


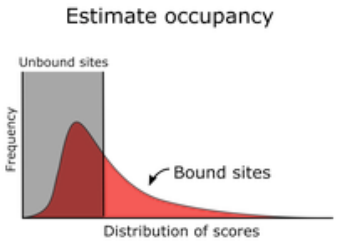
[Bentsen, Nature Com., 2020]

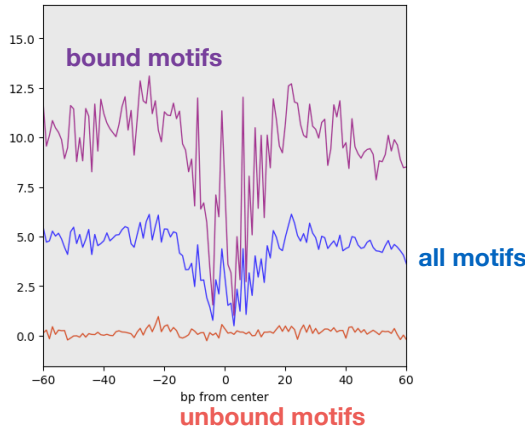
ATAC-seq footprinting

Medizinische Fakultät Heidelberg

- **TOBIAS workflow**
  - Score motif occurences for each transcription factor; distinguish bound/unbound motif occurences (tool: **BINDetect**)
  - Produce aggregate profile plots for each TF (tool: **PlotAggregate**)







[Bentsen, Nature Com., 2020]

Hands-on:
ATAC-seq footprinting for CTCF

https://hdsu-bioquant.github.io/chipatac2020/07_ATAC_Footprinting.html

Bioinformatics Workflow
- ATAC-seq specific QC -

Carl Herrmann

Health Data Science Unit - Medizinische Fakultät Heidelberg

ATAC-seq specific QC

Medizinische Fakultät Heidelberg

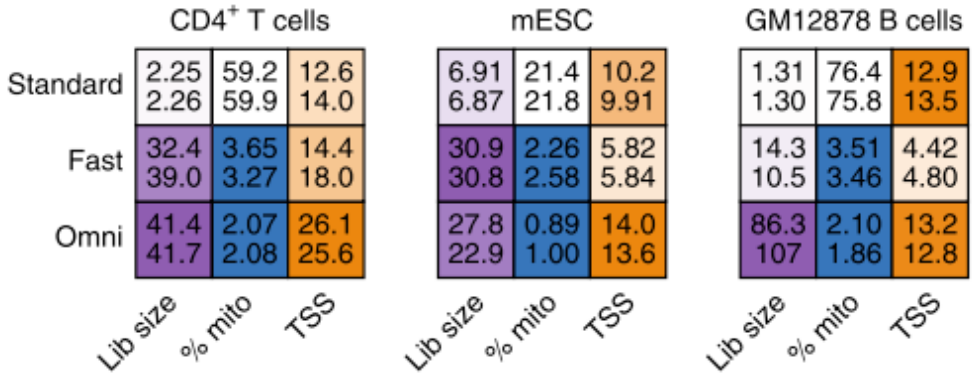
- Some QC measures from the ChIP-seq analysis can be used
  - QC for sequenced reads (**FastQC**)
  - **fingerprinting** (= Lorenz Curve): shows how the signal is restricted to specific regions or spread across the genome
  - Fraction of reads in peaks (FRiP)

- Other QC measures are specific to ATAC-seq:
  - proportion of **mitochondrial** reads
  - **nucleosome** patterning

ATAC-seq specific QC

Medizinische Fakultät Heidelberg

- **Mitochondrial read contamination**
- Initial protocols of ATAC-seq had contamination of mitochondrial reads up to 80%
- Hyper-accessibility of mitochondrial DNA (no chromatin packaging!)
- Recent protocols have improved this a lot used different lysis conditions
  - **Omni-ATAC** [Corces et al., Nature methods 2017]
  - **fast-ATAC** [Corces et al. Nature Genetics 2016]
- Fraction of mt-reads to total library size is an important QC parameter!

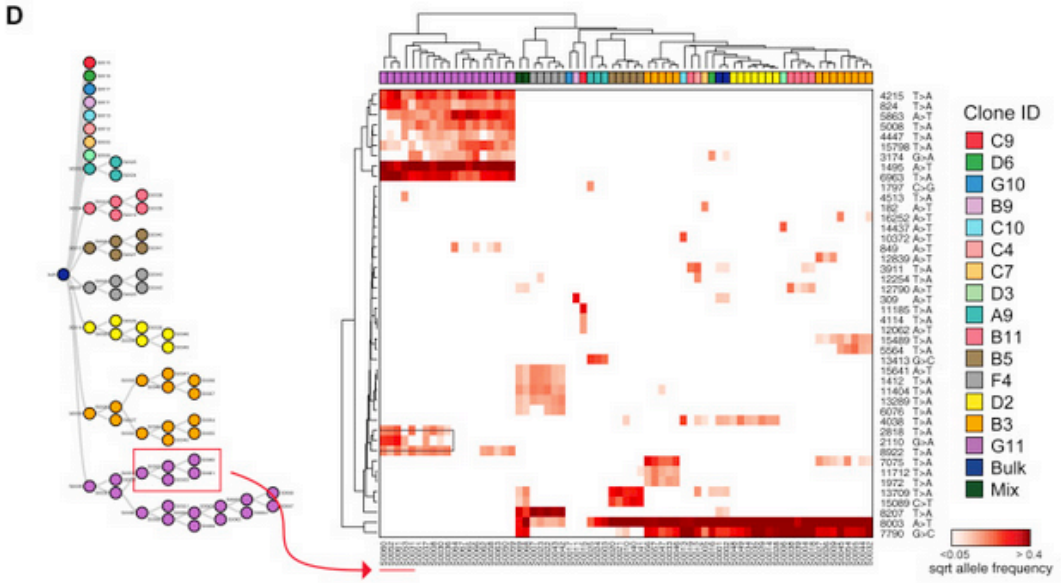


[Corces et al., Nature Methods 2017]

ATAC-seq specific QC

Medizinische Fakultät Heidelberg

- **Nice example of data recycling!**
- Some papers have use the mitochondrial reads obtained as a side-product from ATAC-seq sequencing to determine mitochondrial mutation rates
- Infer cell lineages and clonality in 65 sub-clonal populations from TF1 cell line

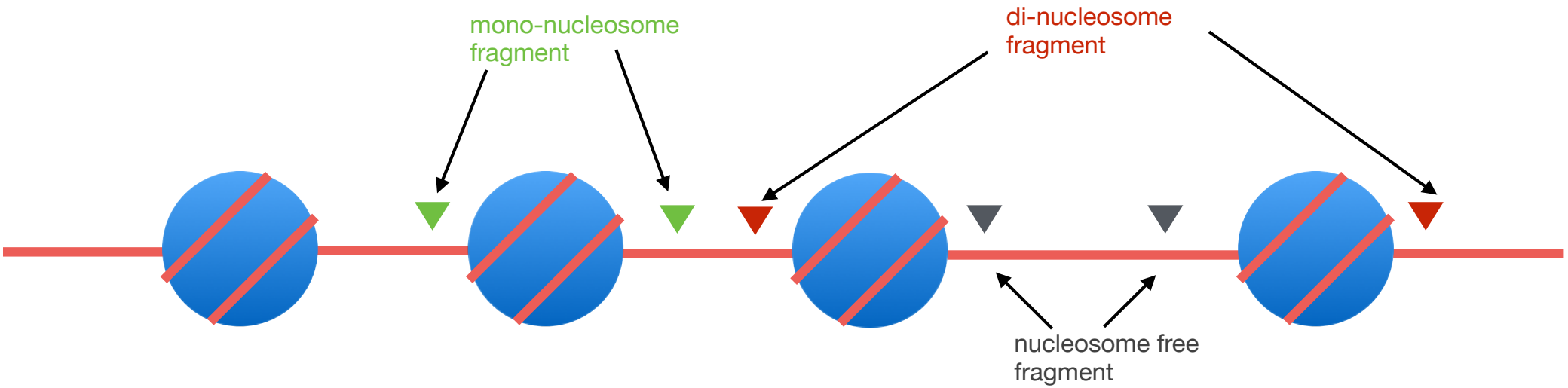


[Ludwig et al., Cell 2019]

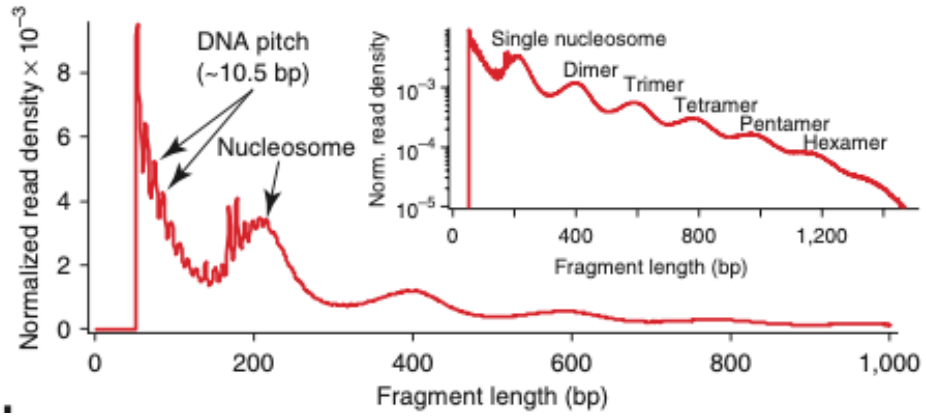
ATAC-seq specific QC

Medizinische Fakultät Heidelberg

- **Nucleosome patterning**



Clear periodicity of fragment length 150-200bp should be observable

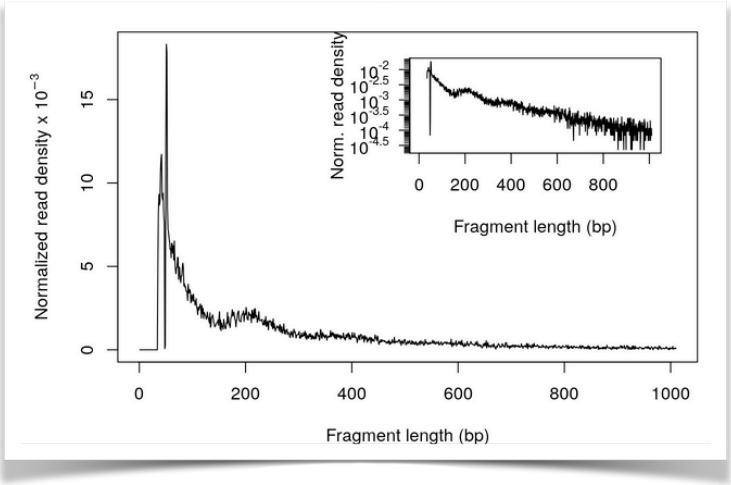


[Buenrostro et al., Nature Methods 2013]

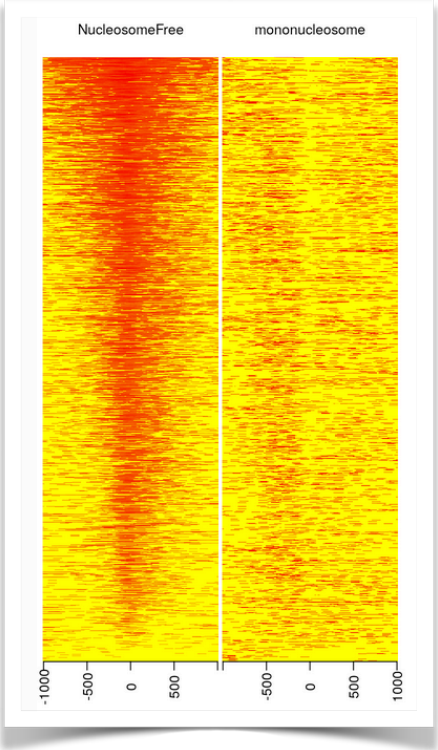
ATAC-seq specific QC

Medizinische Fakultät Heidelberg

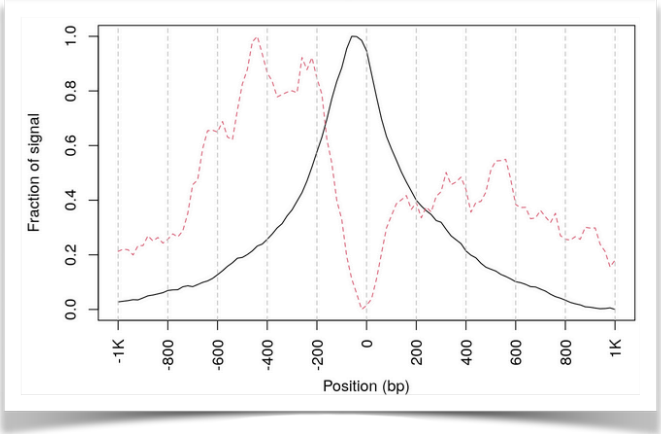
- R/Bioconductor package to perform ATAC-seq QC: **ATACseqQC**
- Check the vignette **here**



Nucleosome patterning



Heatmaps



Profile plots

Hands-on:
ATAC-seq QC using ATACseqQC

https://hdsu-bioquant.github.io/chipatac2020/08_ATAC_QC.html