analyze more, process less

Tools for the inspection, correction, analysis and visualization of deep sequencing data

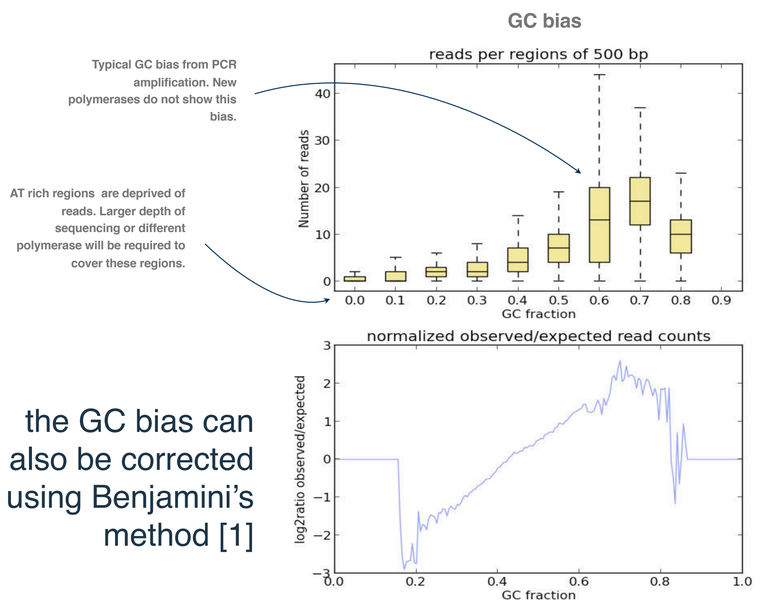
Fidel Ramírez, Friederike Mühlpfordt, Sarah Diehl and Thomas Manke

Max Planck Institute of Immunobiology & Epigenetics
Stübeweg 51, D-79108 Freiburg, Germany,
E-mail: ramirez@ie-freiburg.mpg.de

deepTools is suite of software tools for the post-processing and analysis of aligned reads with a special emphasis on ChIP-seq, MeDIP-seq and MNase-seq data. **deepTools** comprises efficient and reliable programs that enable users to generate files of **normalized read coverages** with different scaling approaches depending on its data's quality and biases. In addition, **deepTools** allows for the computation of various mathematical operations (e.g. difference, log2(ratio)) of two aligned read files, typically a ChIP and an input sample. Furthermore, it can be used for the computation and **visualization** of average and individual signal profiles for large numbers of genomic regions which are commonly applied in down-stream analyses of deeply sequenced data. **deepTools** takes advantage of the **multicore nature** of today's computer servers and can be used either within the **Galaxy** framework or as a stand-alone tool so that inexperienced as well as bioinformatically proficient users can profit from it.

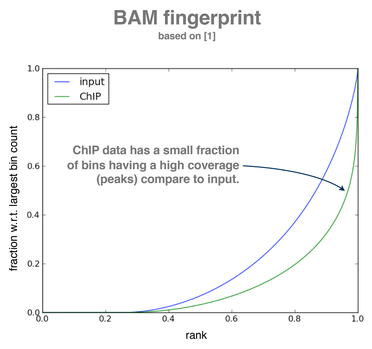
- Diagnostic tools for quality control.
- Multicore features for quick processing.
- Galaxy wrappers.
- Options for GC bias visualization and correction [1].
- Advanced normalizations for ChIP vs. input comparison [2].
- Daily usage in our pipelines. Over a year of development guarantees a set of reliable tools.

Quality Control

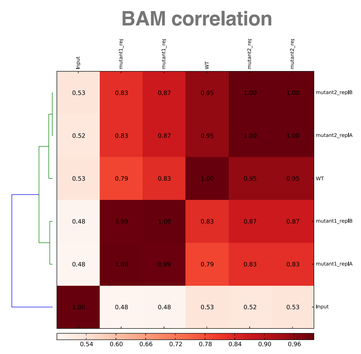


the GC bias can also be corrected using Benjamini's method [1]

```
$ computeGCBias -b chip.bam --biasPlot plot.pdf
```

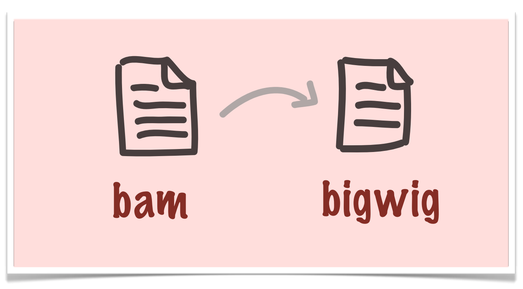


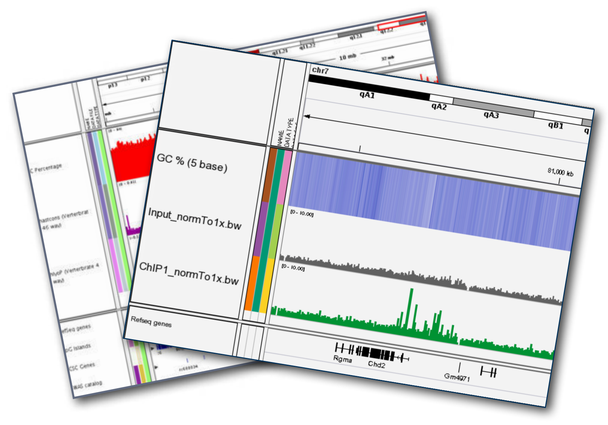

```
$ bamFingerprint -b input.bam
  chip.bam
```




```
$ bamCorrelate -b WT.bam
  input.bam etc...
```

Coverage and ratios



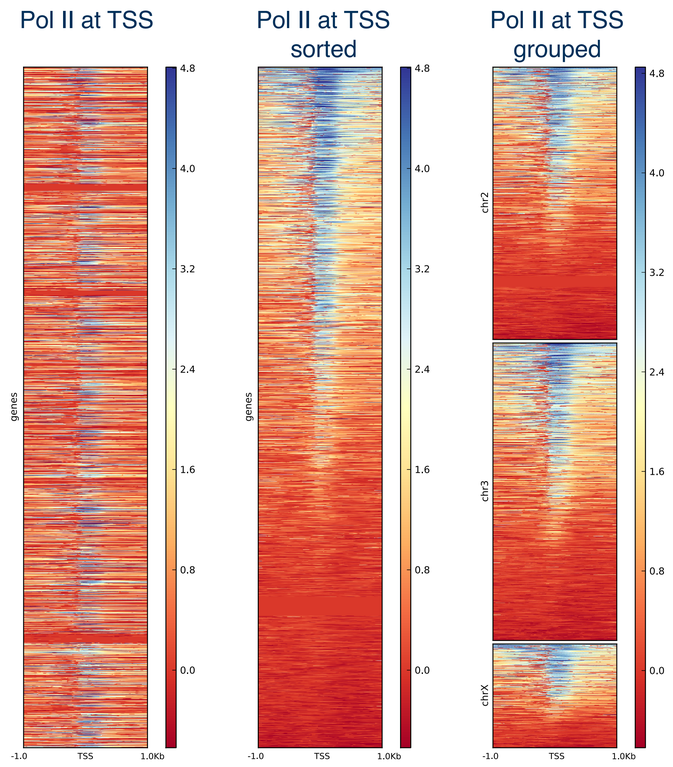



```
$ bamCoverage -b chip.bam -o chip.bw --normalizeUsingRPKM
```

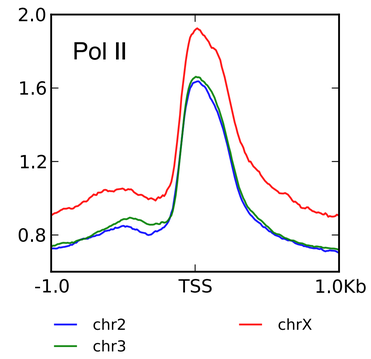
```
$ bamCompare -b1 chip.bam -b2 input.bam --ratio log2
  -o chip_vs_input.bw
```

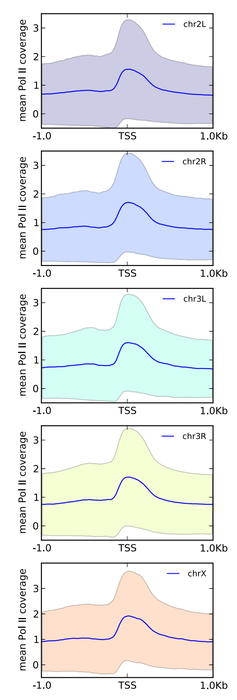
- Easy computation of coverage files with options for filtering reads and normalizing the counts.
- Easy comparison of bam files to produce ratios, log2ratios, sum, diff or fold change.

Heatmaps and profiles



- Large range of options for sorting, thresholding, filtering, binning etc.
- Options to export underlying data for further analysis.
- Two plotting modes: around a reference-point (e.g. TSS) or fitting different region lengths to one size.






```
$ computeMatrix -S PolII.bw -R dm3_genes.bed -o PolII.mat
$ heatmapper -m PolII.mat -o heatmap.pdf
```

```
$ profiler -m PolII.mat -o profile.pdf [--onePlotPerGroup]
```

references

1. Benjamini, Y., & Speed, T. P. (2012). Summarizing and correcting the GC content bias in high-throughput sequencing. Nucleic acids research, 40(10), e72.

2. Diaz, A., Park, K., Lim, D. A., & Song, J. S. (2012). Normalization, bias correction, and peak calling for ChIP-seq. Statistical applications in genetics and molecular biology, 11(3). doi:10.1515/1544-6115.1750

funding

This work is supported by a BMBF grant for the German Epigenome Programme (DEEP) and a DFG grant for Medical Epigenetics (MEDEP).

https://github.com/fidelram/deepTools