

PACMAN: PACific biosciences Methylation ANalyzer

Alexis Loetscher, Laurent Falquet (University of Fribourg and Swiss Institute of Bioinformatics, Switzerland)

Pacific Biosciences sequencing allows for the simultaneous detection of DNA methylation in particular m6A and m4C (Flusberg et al., 2010). The motifs around the methylation sites are provided by the SMRT pipeline in GFF format. However the visualization of these motifs methylated or non-methylated, for example on a bacterial genome, is not part of the pipeline. Circos is a software package for producing publication quality images of large scale data (Krzyszynski et al., 2009), however mastering the numerous configuration files needed requires extra skills not easy to acquire for biologists.

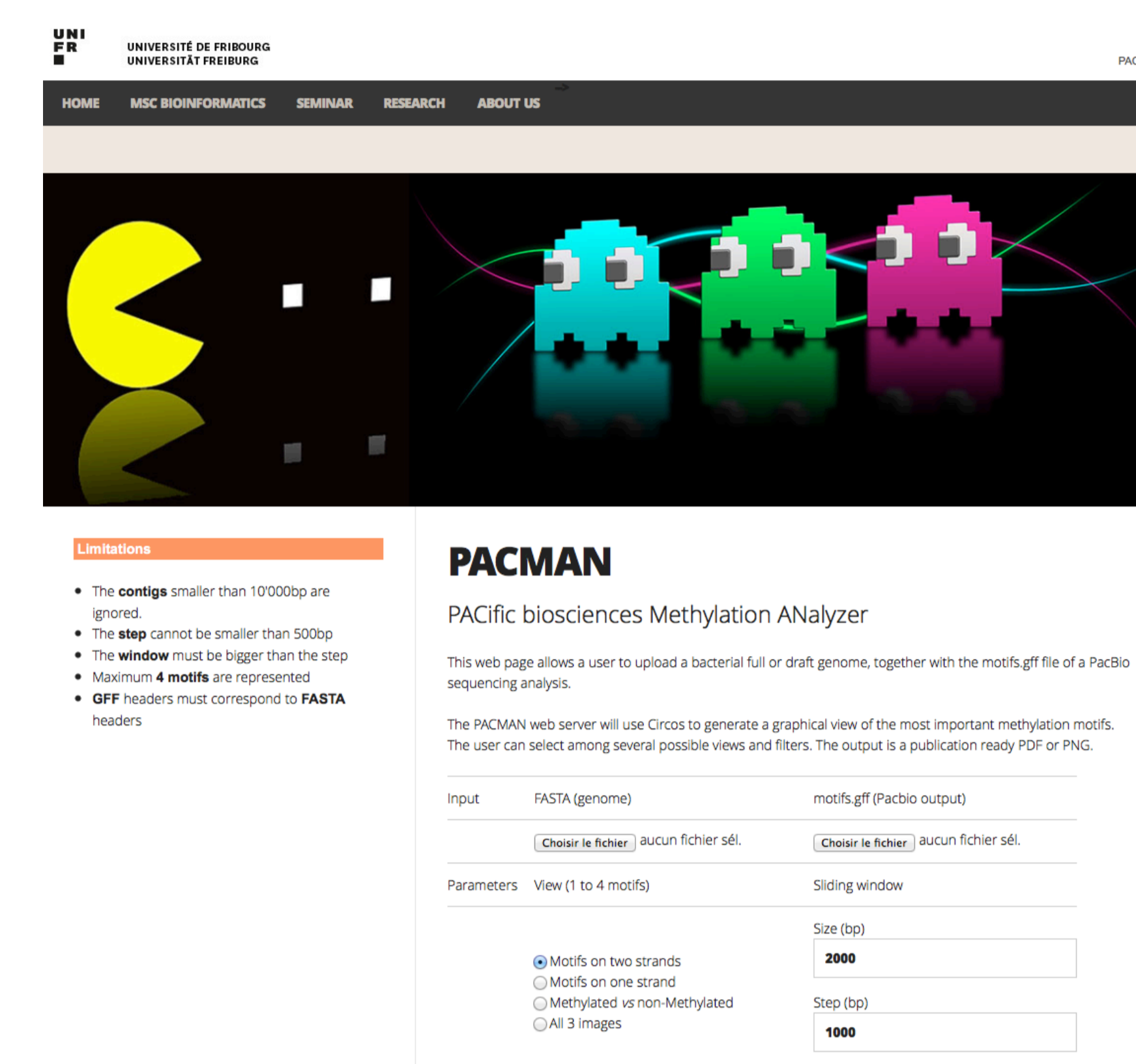
We developed a tool written in Perl and an associated web site called PACMAN (PacBio Methylation Analyzer) allowing users to easily create images with Circos in a user-friendly interface. The required files are 1) the genome or draft in FASTA format and 2) the motifs file in GFF format (from the SMRT pipeline). The counts of each motif are calculated according to a customizable sliding window and normalized by their expected frequency. PACMAN generates a publication quality image of the selected methylated motifs counts, locations and non-methylated locations, on one or both strands of the DNA.

PACMAN is available on this web site:
<http://www.unifr.ch/bugfri/pacman>

PACMAN web site

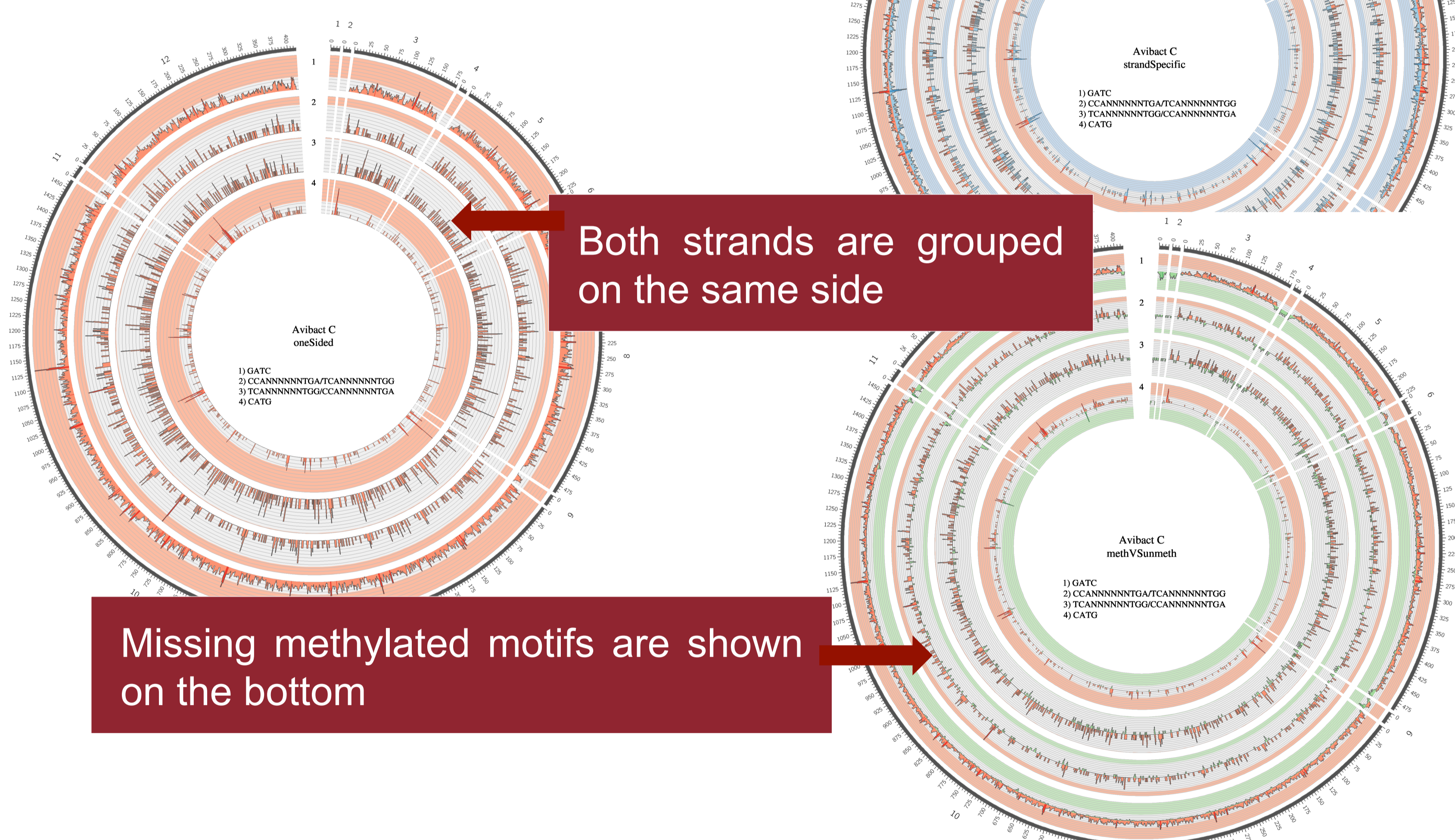
The user-friendly interface is simplified to ease the work of the researcher.

The user can choose the type of image (each strand alone, both strands together, or meth vs unmeth) and the filtering parameters (IPD ratio, coverage, idQv). The top 4 most frequent motifs are displayed.



Examples images

Both strands are shown separately (top and bottom)



Both strands are grouped on the same side

Missing methylated motifs are shown on the bottom

Pacific Biosciences methylation detection

The SMRT® sequencing method of PacBio not only provides extremely long reads, but also kinetic information of the base incorporation. The DNA methylation detection relies on differences in kinetics when the base is modified compared to the canonical base called IPDratio (interpulse duration ratio). Using IPDratio measurement and context specific based identification, the SMRT pipeline allows for the identification of m6A and m4C. Additional steps in the library preparation are required for the detection of m5C.

PACMAN uses the motifs.gff file from the SMRT pipeline and the genome in FASTA format as input.

Circos image generation

Circos is a very powerful package for high quality image creation (<http://www.circos.ca>). However, the task of generating a publication quality image with Circos requires mastering a large number of text configuration files and a sometimes long back and forth process.

PACMAN eases the creation of high quality images by reducing the burden of manipulating configuration files. We developed a Perl script to convert the gff and FASTA files to a nice Circos image.

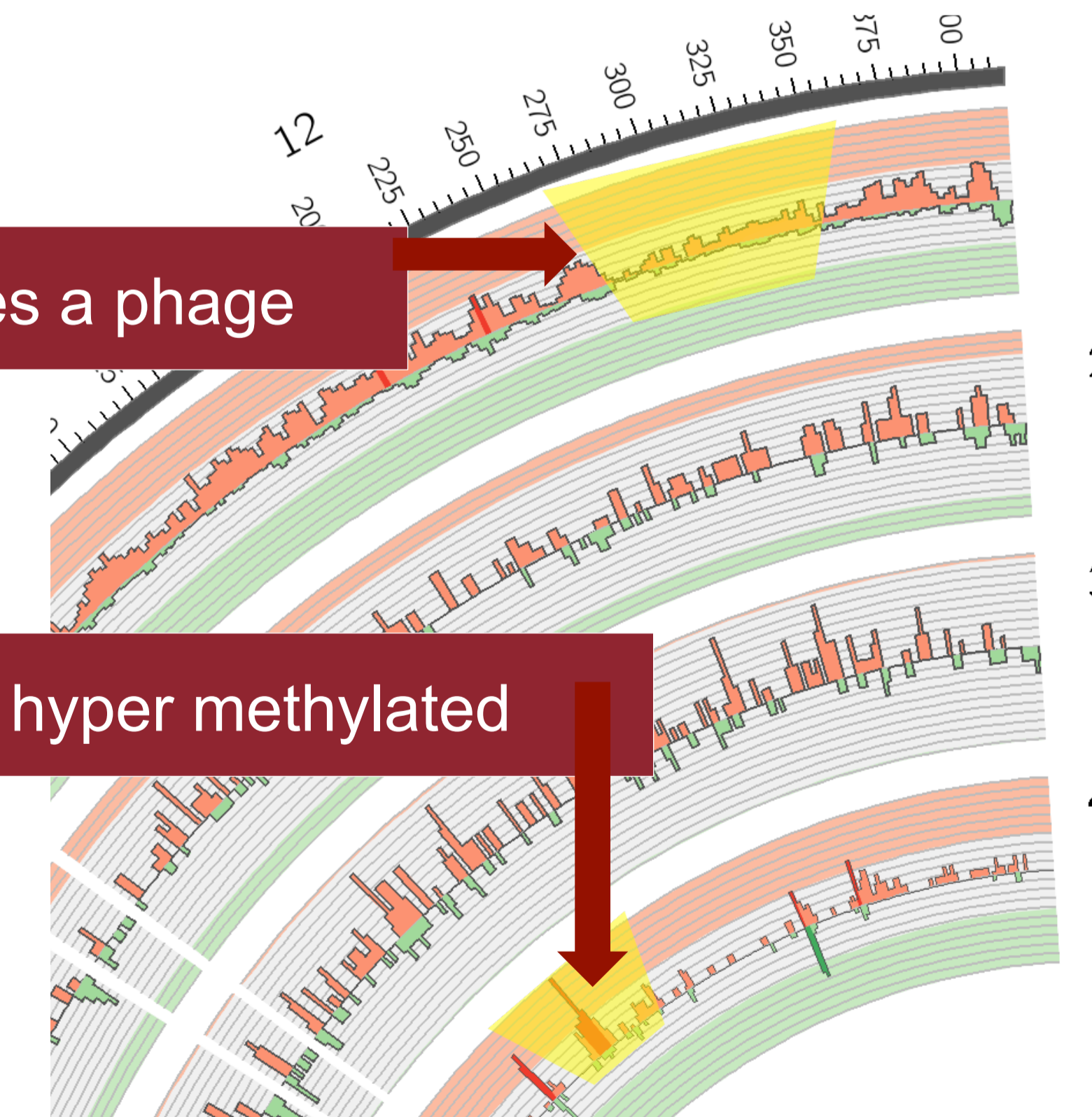
References

- Flusberg, B.A., Webster, D.R., Lee, J.H., Travers, K.J., Olivares, E.C., Clark, T.A., Korlach, J., and Turner, S.W. (2010). Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* 7, 461–465.
- Krzyszynski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: An information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645.
- Storno GD (2000). DNA binding sites: representation and discovery. *Bioinformatics* 16, 1, 16-23.

A real case :

A hypo methylated region hides a phage

A toxin operon is hyper methylated



Future developments

Improved images and filtering capabilities

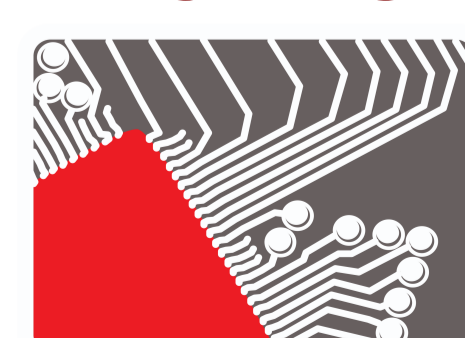
- allow zooming on specific regions

IPDratio outliers

- Highlight IPDratio outliers



Powered by



Vital-IT
 High Performance Computing Center

Contact
bugfri@unifr.ch
www.unifr.ch/bugfri