

# Automated ChIP-Seq Analysis and Reporting Pipeline

Stephen Kelly<sup>1,2</sup>, Igor Dolgalev<sup>1-5</sup>, Charalampos Lazaris<sup>3-5</sup> & Aristotelis Tsirigos<sup>1-5</sup>

<sup>1</sup>Applied Bioinformatics Center & <sup>2</sup>Genome Technology Center, NYU School of Medicine, NY 10016, USA, <sup>3</sup>Department of Pathology, <sup>4</sup>NYU Cancer Institute and Helen L. and Martin S. Kimmel Center for Stem Cell Biology, <sup>5</sup>Center for Health Informatics & Bioinformatics,

## Objective

Develop a reproducible ChIP-Seq pipeline that can adapt to varying numbers of samples, treatment groups, and parameter sets while allowing for user expansion with custom programs and analysis tasks.

## Pipeline usage summary

- 1 Create new directory for analysis from a clone of the pipeline repository
- 2 Set input files (fastq, bam)
- 3 Generate sample sheet with pipeline-provided scripts
- 4 (Optional) Modify parameters as needed, add custom tasks
- 5 Run pipeline
- 6 Compile report

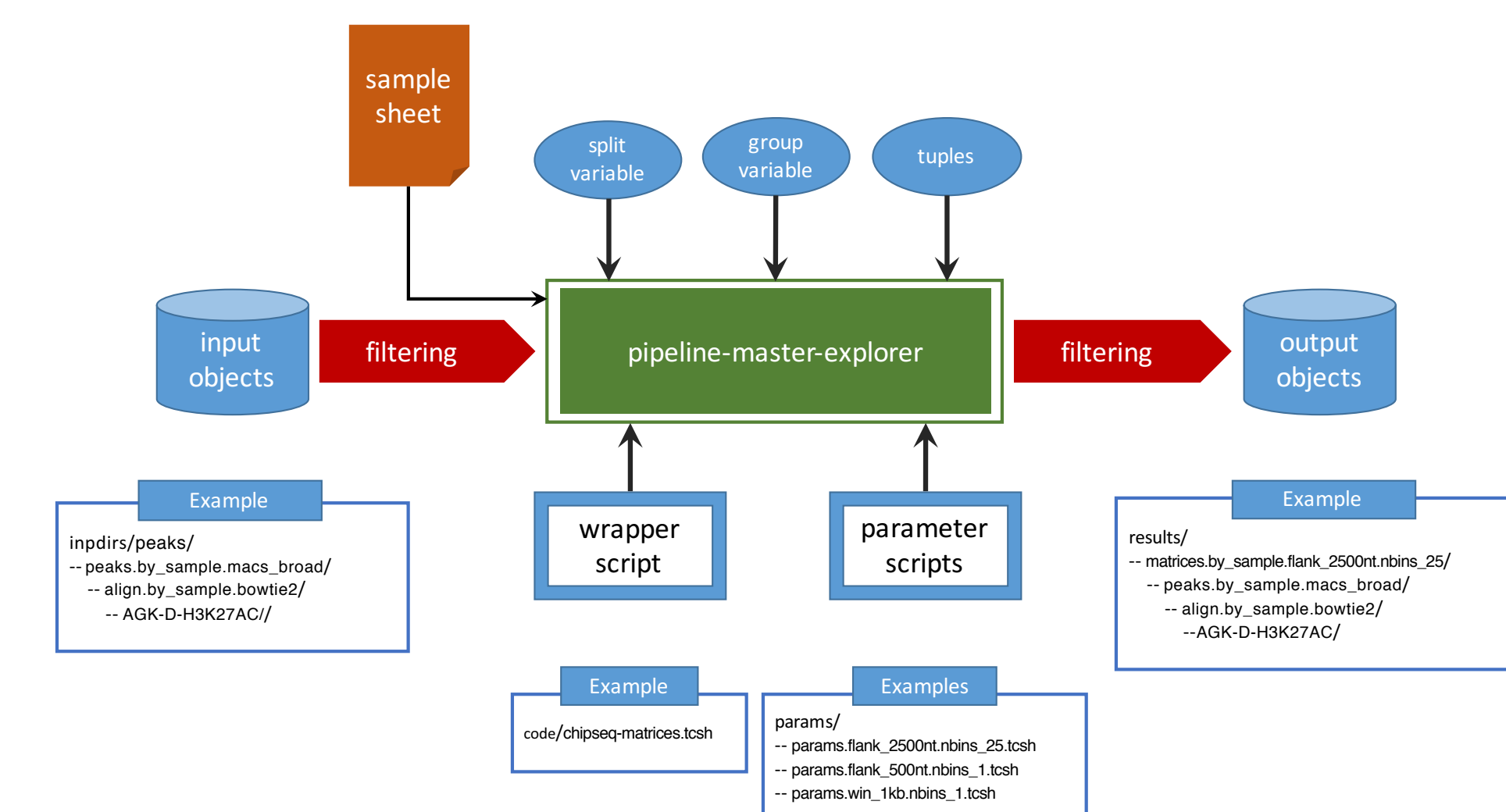


Figure 1: ChIP-Seq pipeline programmatic workflow. All sets of parameters are evaluated for each pipeline task in a combinatorial fashion

ChIP-Seq Pipeline tasks	Programs Used
Alignment	bowtie2
Alignment Stats	bowtie2, <b>R</b>
Quality Control	FastQC, deepTools
Peak Calling	<b>MACS</b>
Matrix Correlation	<b>GenomicTools</b>
Principal Component Analysis	<b>R</b>
Heatmap Clustering	<b>R</b>
Differential Peak Binding	DiffBind
Visualization	<b>R</b>
Automatic Reporting	<b>R, <math>\LaTeX</math> 2<math>\epsilon</math></b>

Table 1: ChIP-Seq pipeline standard components. Internally developed methods listed in bold.

## Auto-report Sample Output

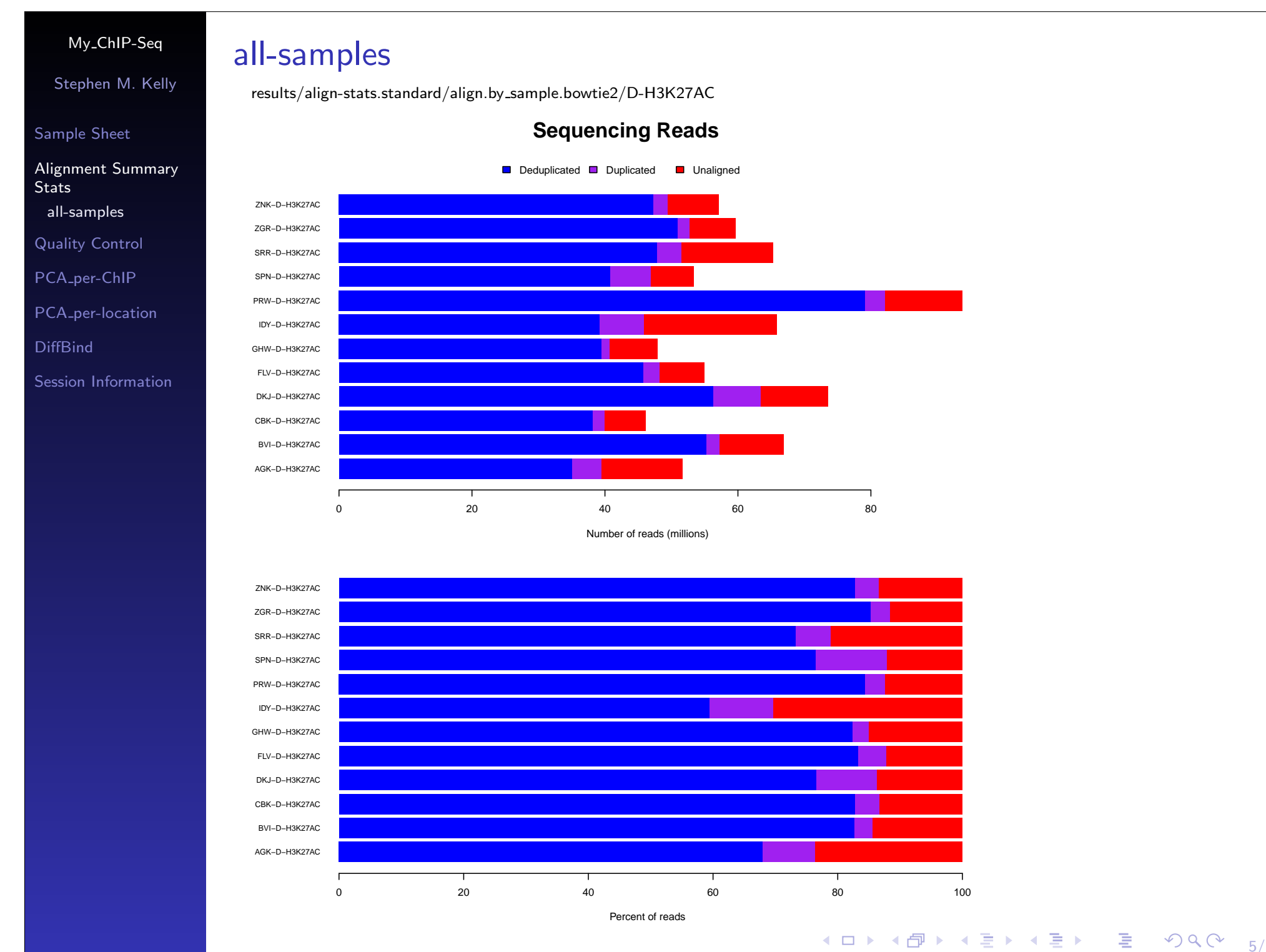


Figure 2: Alignment summary statistics

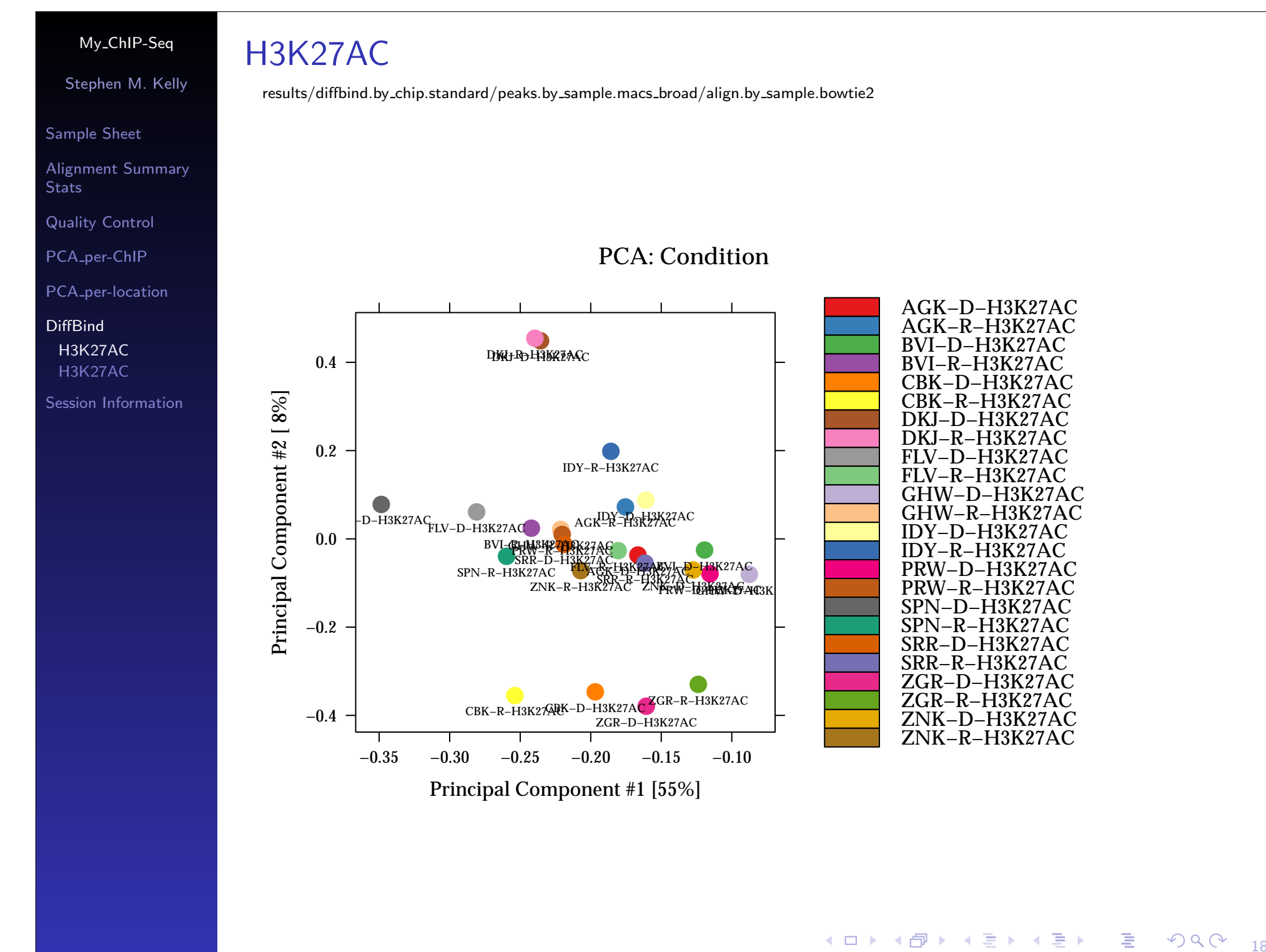


Figure 4: Principal component analysis

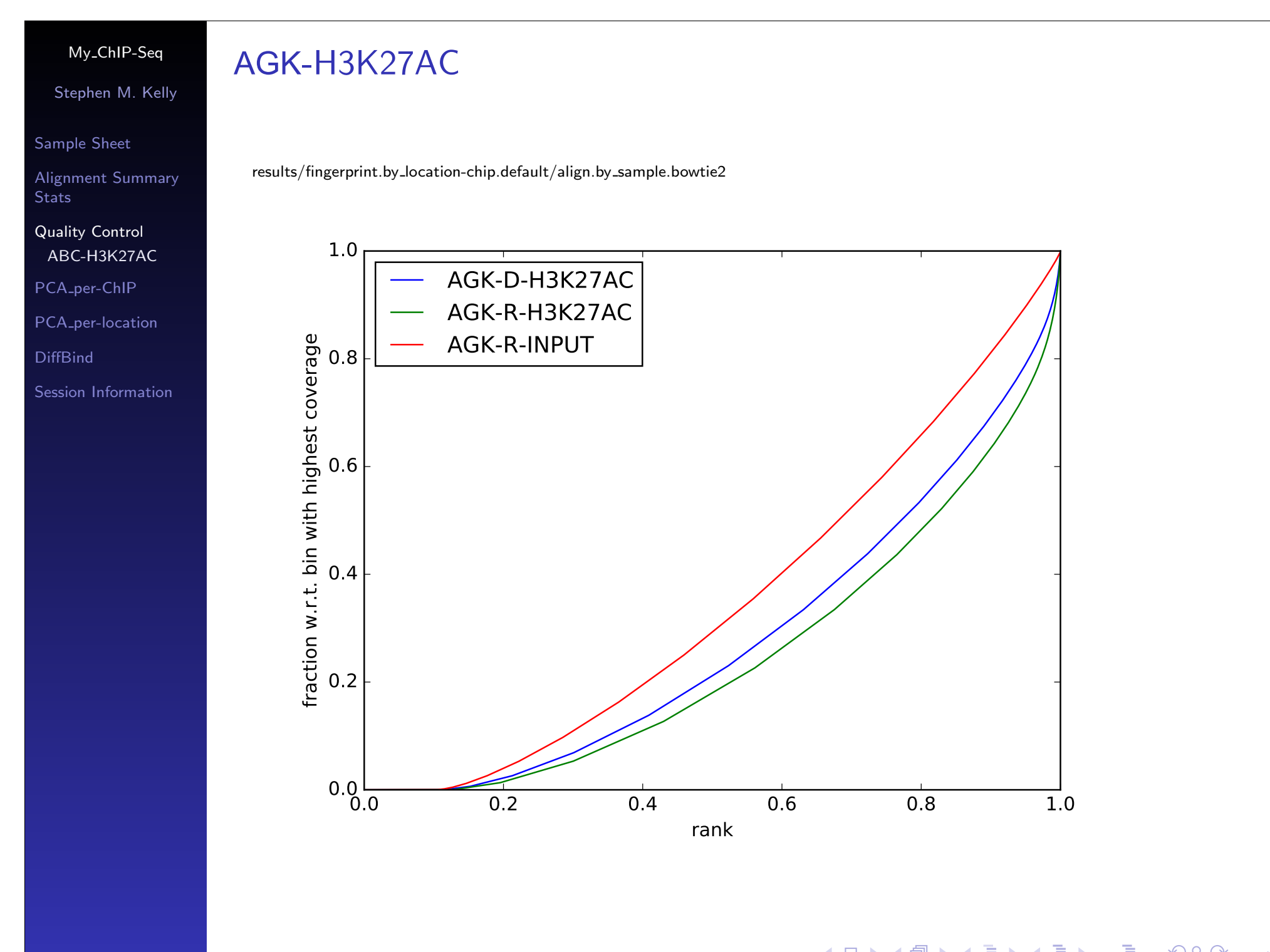


Figure 3: Quality control metrics

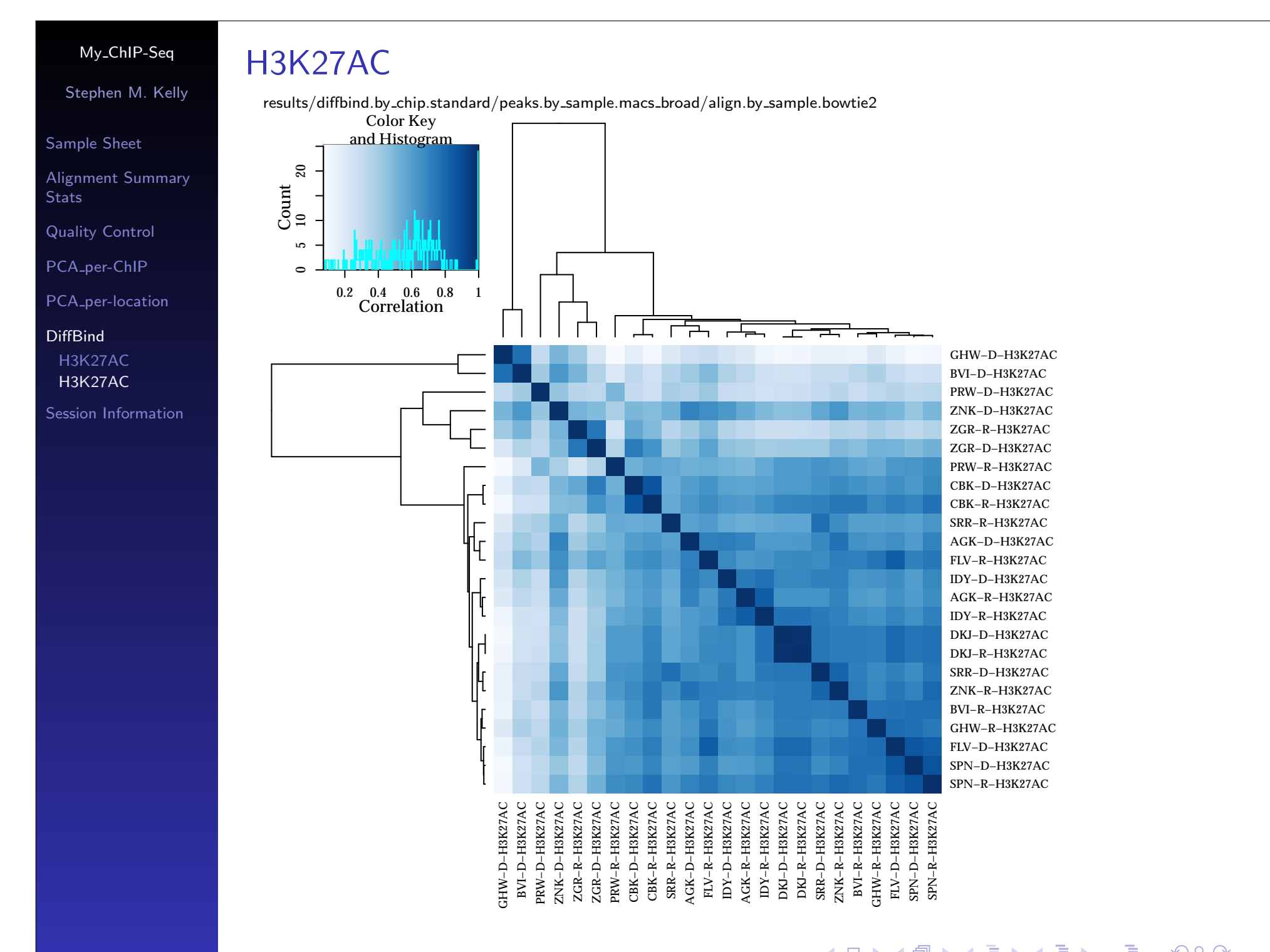


Figure 5: Differential binding heatmaps

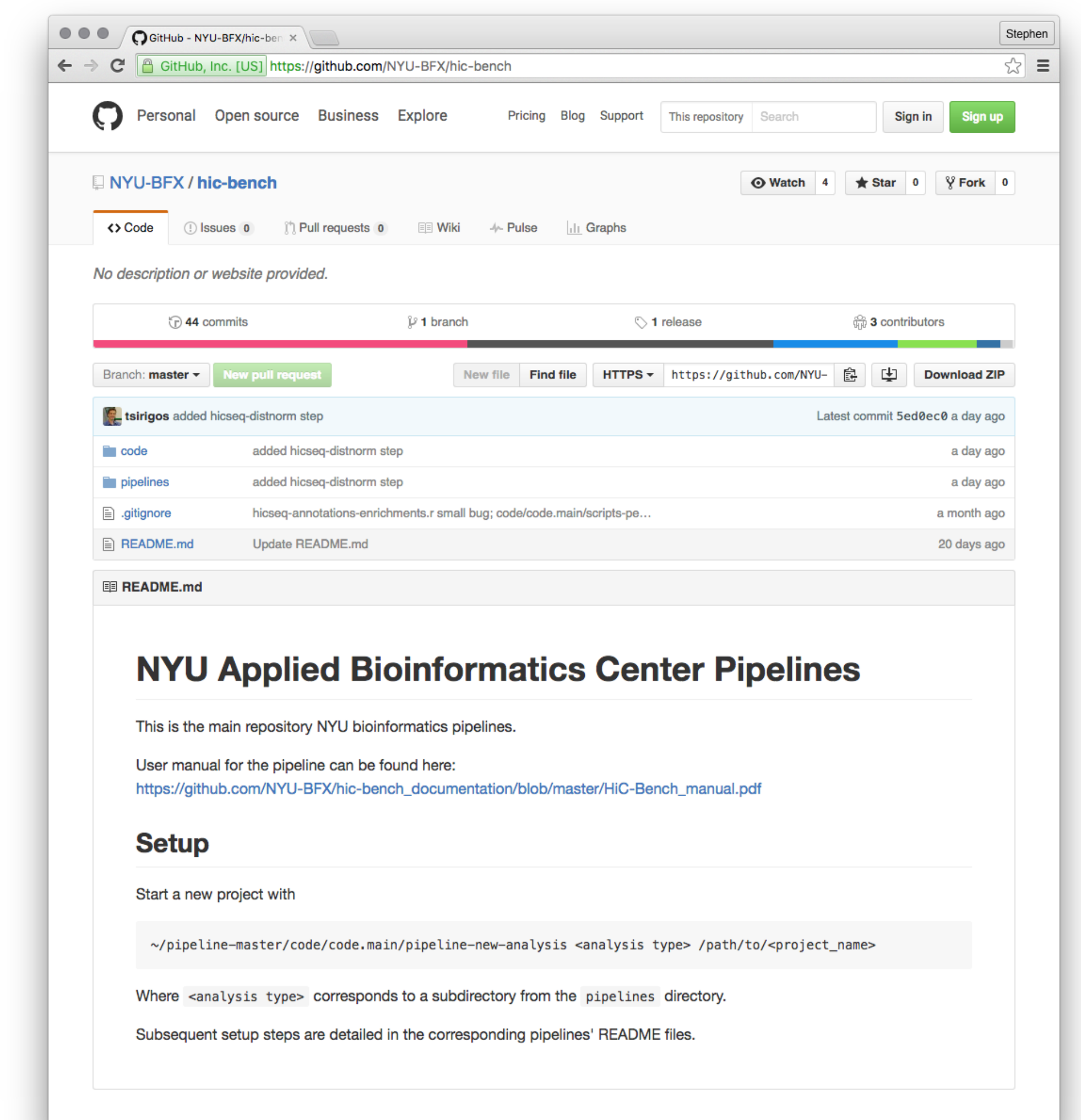


Figure 6: Our ChIP-Seq pipeline is part of the HiC-Bench software package, available on GitHub

## Future Developments

- Automated motif analysis
- Peak overlapping and filtering

## Acknowledgements

This work used computing resources at the Laura and Isaac Perlmutter Cancer Center, which is supported by Cancer Center Support Grant P30CA016087. A. Tsirigos was supported by a Research Scholar Grant, RSG-15-189-01-RMC from the American Cancer Society. This work also used computing resources at the High Performance Computing Facility of the Center for Health Informatics and Bioinformatics at the NYU Langone Medical Center.

## Software

- Web: <http://www.med.nyu.edu/ocs/applied-bioinformatics-center>
- GitHub: <https://github.com/NYU-BFX/hic-bench>
- Zenodo: <https://zenodo.org/record/47676>
- Contact: [stephen.kelly@nyumc.org](mailto:stephen.kelly@nyumc.org), [aristotelis.tsirigos@nyumc.org](mailto:aristotelis.tsirigos@nyumc.org)