

Mads Lerdrup, Jens Vilstrup Johansen, Shuchi Agrawal-Singh, and Klaus Hansen:

# An interactive environment for agile analysis and visualization of ChIP-seq sequencing data

## Motivation

- Chromatin Immuno-Precipitation sequencing (ChIP-seq) and related methods are becoming widely used.
- With lowered sequencing costs and kits for sample preparation, data analysis and comprehension has become the limiting factor for feasibility.
- Genome browsers are popular tools for visualization of individual loci, but offers limited means for genome-wide abstraction.
- Comprehensive, integrated, and user-friendly tools with ample visualization and data-interaction will improve comprehension and democratize analysis.

## EaSeq

### Made for bioscientists

- Graphical user interface
- Point and click with pervasive visualization
- Runs on an ordinary Windows PC
- Self-contained
- Versatile
- Free

### Tightly integrated

- Build up figures that link to genomic loci.
- Organize data into exchangeable session files 1 that ties raw data 2 to plots 3.
- Fully integrated tools 4 accelerates work.
- Limits file -reformatting and -export.

### Autodocumenting

- EaSeq automatically records each step 5.
- Exported figures 6 can be enumerated
- Autogenerated legends and methods 7 are exported along.

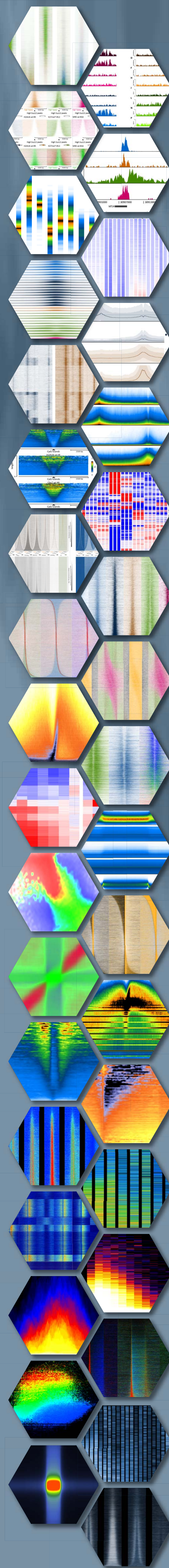
### Highly interactive

- Loci from heatmaps 8 or plots 9 can be visualized in other plots and tracks 10.
- Subsets of regions can be gated into new sets 11.
- Lists 12 and heatmaps 13 can be used to inspect individual loci ensuring that interpretations match underlying signals 14.

## Comprehensive toolset

Visualization		Graph of quantified values in sorted order		Simple open track of locus/loci that updates to show selection		Pie chart showing sums of multiple parameters	
		Graph of average quantified values in sorted order		Overlaid track of multiple average tracks		Heatmap of multiple loci in sorted order	
		Density plot showing distributions of multiple parameters		2D-histogram / scatter plot		Ratiometric heatmap of multiple loci in sorted order	
		Simple line plot showing averages from multiple parameters		2D-histograms showing distribution as expected by chance		Heatmap with loci put on y-axis according to quantified parameter	
		Box plot showing fractiles from multiple parameters		2D-histogram pseudocolored for values of a third parameter		Normalized heatmap with loci put on y-axis according to parameter	
		Bar plot showing averages from multiple parameters		1D-histogram		Simple plot of values of a parameter in sorted order	
		Simple filled track of locus/loci that updates to show selection		Annotation track to show genes and coordinates		Simple heatmap of values of a parameter in sorted order	
		Simple average track of loci that updates to show selection					
	Analysis and processing		Calculate "Phantom peak coefficient" to measure dataset quality		Calculate numbers in two parameters, e.g. ChIP-seq signals		Find the subset of two sets of regions that has similar qualities
			Identify peaks - regions with an enrichment of signal in a dataset		Normalize parameters - linear, quantile, or LOESS		Depict the co-distribution of two sets of regions relative to a third
		Identify regions with the highest resemblance to a pattern of signals throughout the entire genome		Quantify the amount of dataset signals at a set of regions		Colocalize a set of regions with features from a set of genes	
		Correlate the signal of datasets throughout the entire genome		Generate randomized parameters		Find the overlap between two or three sets of genes	
		Sort the set of regions based on one or more parameters		Score members of a set of regions positive / negative for a parameter		Annotate a set of regions with the name of the nearest gene	
		Cluster regions based on dataset signal - Hierarchical or K-means		Colocalize two sets of regions		Extract positions of e.g. TSS from a list of genes	
		Gate a subset of regions based on one or more parameters		Make random control regions matched to a known genomic feature		Pool, shrink, shift or calibrate datasets	
		Cluster the regions based on parameters - Hierarchical or K-means		Calculate the overlap between sets of regions		Merge, overlay, exclude, or fragment sets of regions	

## Output



## Applications

