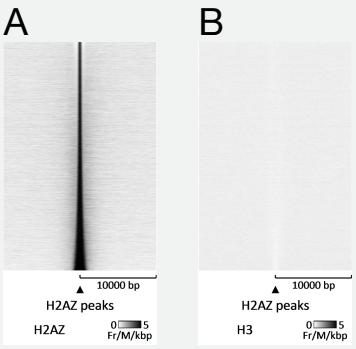
Descriptions of those tools in EaSeq that are not described in the manuscript.

All analysed raw data are from Barski et al (2007), Cell 129, 823-837.

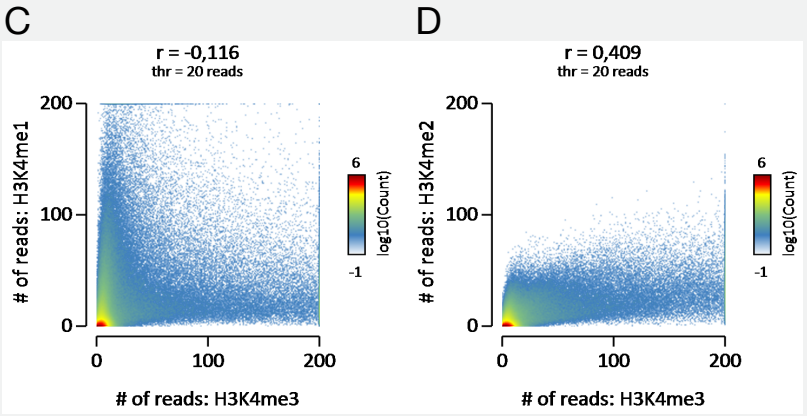
Identify peaks - regions with an enrichment of signal in one dataset

A+B) Heatmaps of peaks found with standard settings in the 'ALT' peak-finding tool. A) shows the positive sample used, and B) shows the negative control. Note the lack of regions with background signal in the neagtive control.



Correlate the signal intensity of data-sets throughout the entire genome

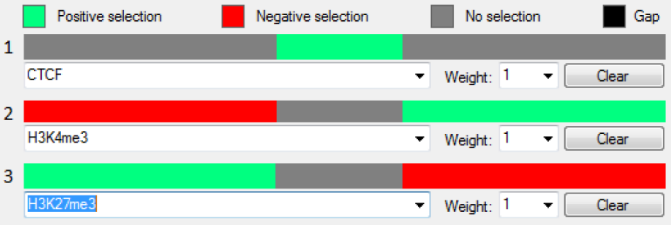
C+D) 2D-histograms of the global levels of H3K4me3 in combination with either H3K4me1 (C) or H3K4me2 (D). Plots and values were obtained using the 'Correlate' tool with the window size set to 1000 bp and max values on axes as 200 counts. r values are genome-wide Pearsson's correlation coefficients for the datasets.



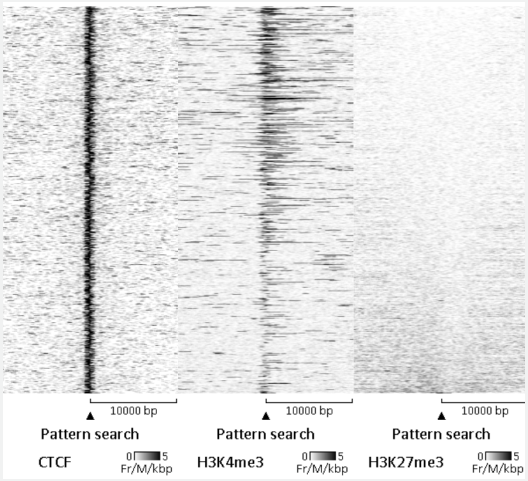
Identify regions that resemble a certain pattern of signals the most

E-H) Patterns search interface (E), as well as heatmaps (F), average tracks (G), and filled tracks from a single locus (H) of CTCF, H3K4me3, and H3K27me3. Regions were identified using the 'Pattern-search' tool (E) in order to find 1000 genomic loci where CTCF-rich regions were flanked by H3K4me3 on the right side, but not the left side and H3K27me3 on the left side but not right side. Although most identified regions only reflected the actual pattern for one or two of the datasets, it was evident on a population level that the sought features were enriched (G).

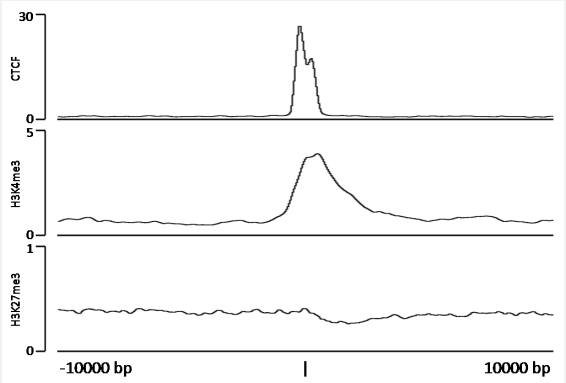
E



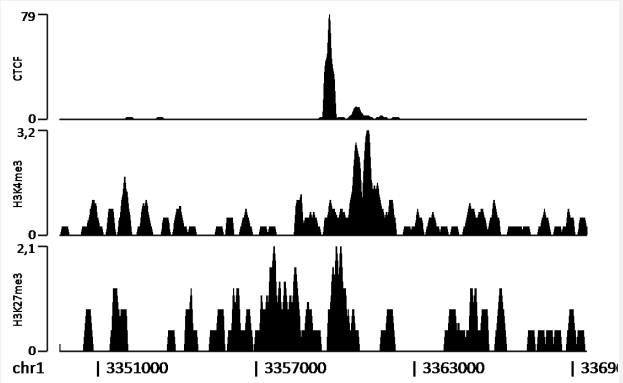
F



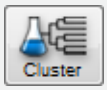
G



H

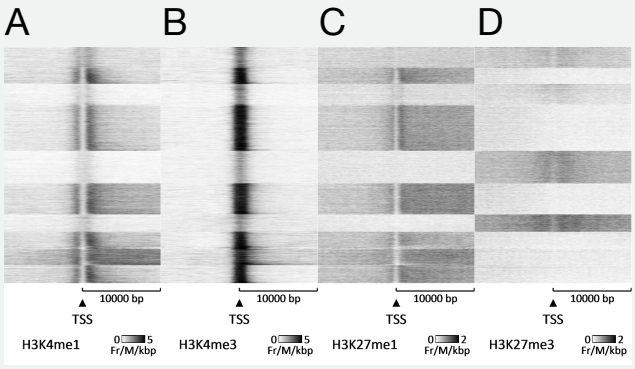


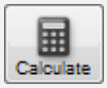
All analysed raw data are from Barski et al (2007), Cell 129, 823-837.



Cluster regions based on dataset signal – Hierarchical or K-means

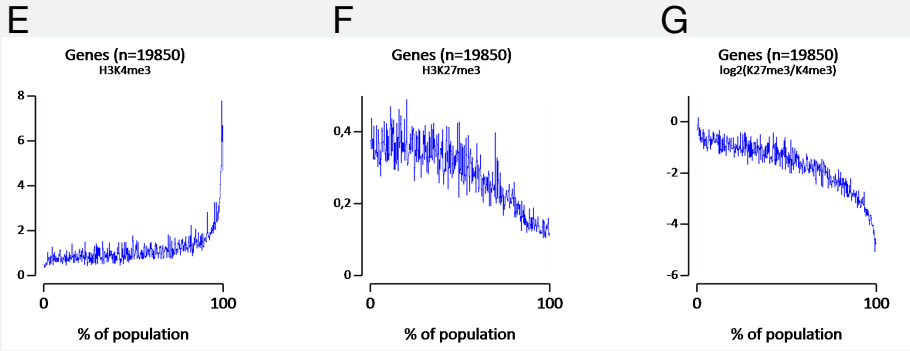
A-D) Heatmaps showing H3K4me1 and -me3 at TSS clustered (k-means) based on H3K4me1 and -me3, H3K27me1 and -me3 profiles at the 20 kbp surrounding TSS.

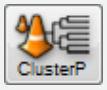




Calculate numbers based on two parameters, e.g. log2 fold change

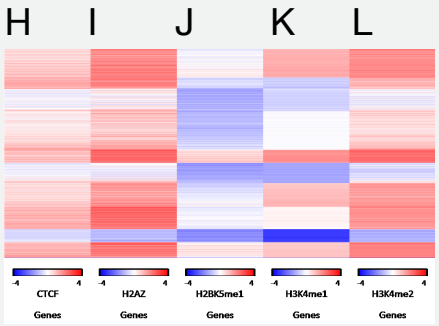
E-G) Graphs showing the levels of H3K4me3 (E) and H3K27me3 (F) quantified at the gene bodies. The graph in G shows the genome-wide result of calculating the log2 ratio between the two signals using the 'Calculate' tool.

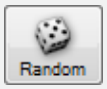




Cluster the regions based on parameters – Heirachical or K-means

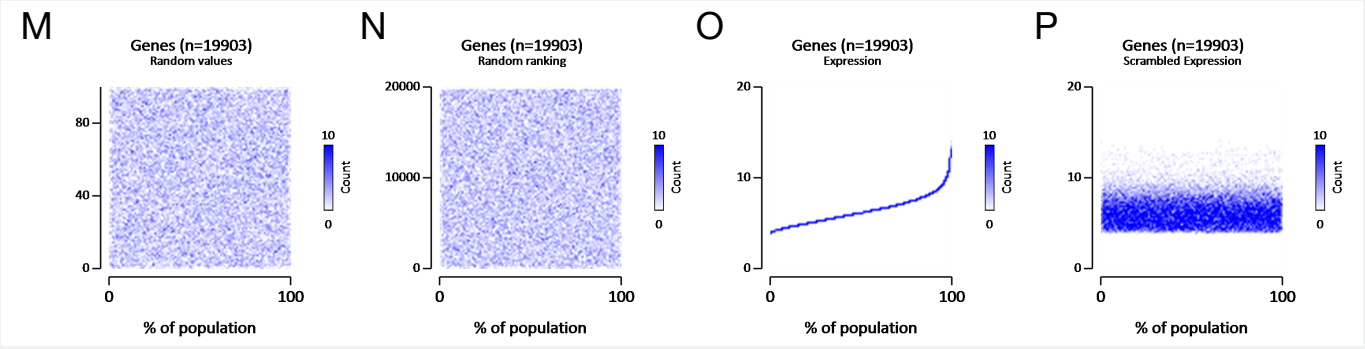
H-L) Heatmaps showing parameters from several histone marks that were quantified in gene bodies and clustered (k-means) based on their levels.

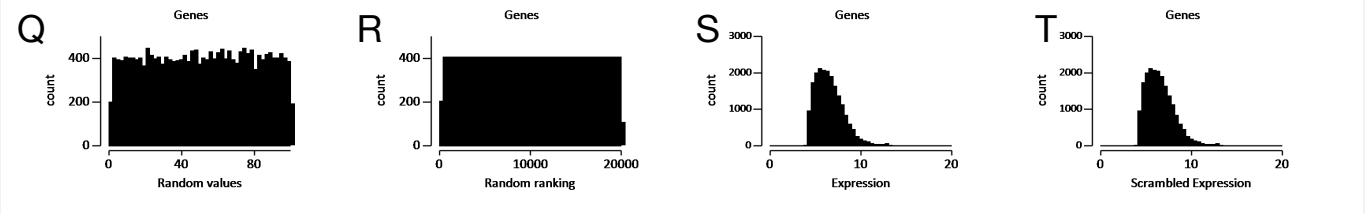


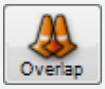


Generate randomized parameters

M-T) Graphs (M-P) and Histograms (Q-T) of values generated using the 'Random'-tool. The tool can generate completely random values (M+Q), randomly ranked numbers from 1 to n (N+R), or scramble the values from an existing parameter (O+S) in a random manner (P+T).

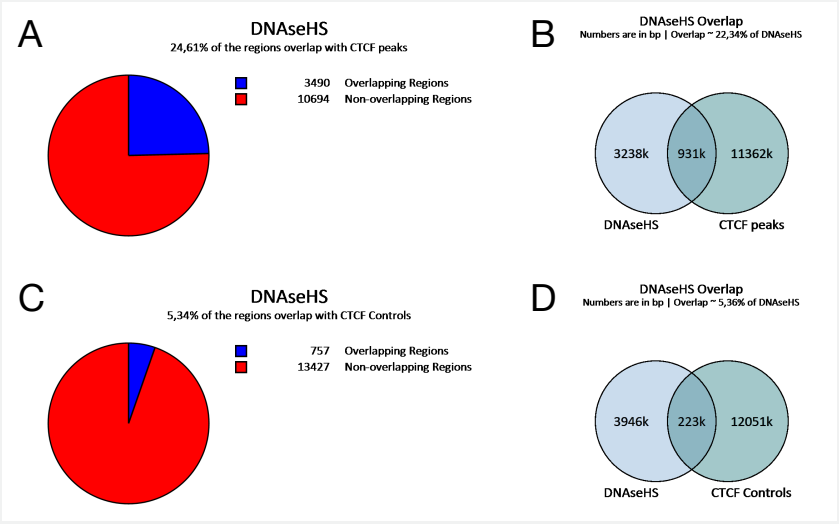


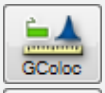


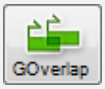


Calculate the overlap between sets of regions

A-D) Pie charts (A+C) and crude Venn diagrams (B+D) showing the number of overlapping regions (A+C) and overlap in bp (B+D) calculated for DNAse hypersensitive sites and CTCF peaks or matched controls for CTCF. Charts and calculations were done using the 'Overlap' tool.



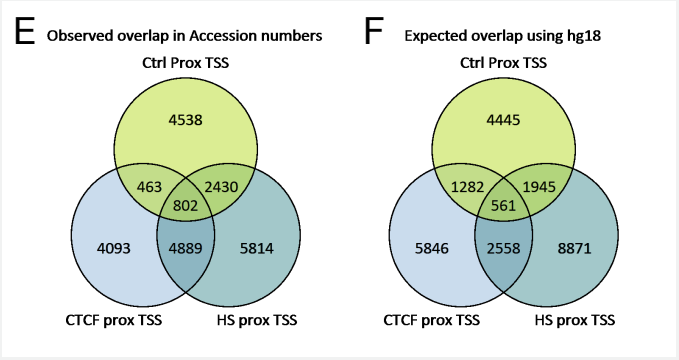




Colocalize a set of regions with features from a set of genes

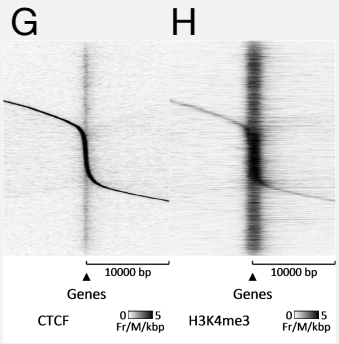
Find the overlap between two or three sets of genes

E+F) Crude Venn diagrams showing the actual (E) and expected (F) overlap between three sets of target genes that were compiled by having TSS within 1kbp of the sets of regions. Diagrams were created using the 'GOverlap' tool and Genesets were created using the 'GColoc'-tool.



Colocalize two sets of regions

G+H) Heatmaps showing CTCF (G) and H3K4me3 (H) signal genome-widely at TSS. Genes were sorted according the distance to the nearest CFCT peak (using the 'Sort'-tool, and this distance was calculated using the 'Coloc'-tool.



Depict the co-distribution of two sets of regions relative to a third

I+L) Heatmaps showing the genome-wide cooccurrence of CTCF peaks and DNAse hypersensitive sites at varying distances relative to TSS in the genome. Heatmaps shows the cooccurrence in absolute numbers (I), expected numbers based on the prevalence of each set of regions at the locations relative to TSS (K), and the Observed/Expected ratio with the density of the stain controlled by either the absolute number (J) or log10-transformed (L) number of observations.