Glossary

Like most specialized fields, next-generation sequencing has inspired many an acronym. We are trying to keep track of those abbreviations that we heavily use. Do make us aware if something is unclear: deeptools@googlegroups.com

If you are unfamiliar with the file formats of next-generation sequencing data, do have a look on the next page.

Abbreviations

Acronym	full phrase	Synonyms/Explanation
-seq	-sequencing	indicates that an experiment was completed by DNA sequencing using NGS
ChIP-seq	chromatin immunoprecipitation sequencing	NGS technique for detecting transcription factor binding sites and histone modifications (see entry "Input" for more information)
DNase	deoxyribonuclease	micrococcal nuclease
HTS	high-throughput sequencing	next-generation sequencing, massive parallel short read sequencing, deep sequencing
Input	--	control experiment typically done for ChIP-seq experiments (see above) - while ChIP-seq relies on antibodies to enrich for DNA fragments bound to a certain protein, the input sample should be processed exactly the same way, excluding the antibody. This way, one hopes to account for biases introduced by the sample handling and the general chromatin structure of the cells
MNase	micrococcal nuclease	DNase
NGS	next-generation sequencing	high-throughput (DNA) sequencing, massive parallel short read sequencing, deep sequencing
RPGC	reads per genomic content	used to normalize read numbers (also: normalize to 1x sequencing depth), sequencing depth is defined as: (total number of mapped reads * fragment length) / effective genome size.
RPKM	reads per kilobase per million reads	used to normalize read numbers, the following formula is used by bamCoverage: RPKM (per bin) = number of reads per bin / (number of mapped reads (in millions) * bin length (kb))