

## Pipeline Overview

The DNase-seq pipeline inputs gzipped DNA-sequencing reads in the form of fastqs and a genome index. These fastqs come from technical replicates<sup>1</sup> that are processed with bwa to generate unfiltered and filtered alignments (bam format). The filtered alignments, combined with a DNase mappability reference (<https://www.encodeproject.org/references/ENCSR792JJZ/>) depending on assembly, provides additional outputs:

- Merging one or more technical replicate bams into a biological replicate level unfiltered bam file; subsequently, filtering and UMI deduplication will produce a filtered bam file.
- Hotspot calling. The filtered bam files will produce hotspots and peak files (bed and bigBed format), and signal files (bigWig format)
  - inputs: filtered\_biorep.bam; outputs: hotspots.bed/bb, peaks.bed/bb, signal.bw
- IDR of 2 replicate peak sets. inputs: peaks.bed; outputs: idr.bed/bb, qc

The pipeline does not have any restrictions towards its inputs, however it is recommended to have at least two biological replicates<sup>2</sup> and the read length be 36nt.

**For single-end library preparation:** [View pipeline](#)

**For paired-end library preparation:** [View pipeline](#)

## Genomic References Used in this Pipeline

Unmodified Genome References and Chromosome Sizes:  
[encodeproject.org/ENCSR425FOI](http://encodeproject.org/ENCSR425FOI)

## References

- Lazarovici A, Zhou T, Shafer A, Machado AC, Riley TR, Sandstrom R, Sabo PJ, Lu Y, Rohs R, Stamatoyannopoulos JA, Bussemaker HJ. Probing DNA shape and methylation state on a genomic scale with DNase I. *Proceedings of the National Academy of Sciences*. 2013 Apr 16;110(16):6376-81.
- Sabo PJ, Hawrylycz M, Wallace JC, Humbert R, Yu M, Shafer A, Kawamoto J, Hall R, Mack J, Dorschner MO *et al*. Discovery of functional noncoding elements by digital analysis of chromatin structure. *Proc Natl Acad Sci U S A*. 2004 Nov 30;101(48):16837-42.
- Sabo PJ, Kuehn MS, Thurman R, Johnson BE, Johnson EM, Cao H, Yu M, Rosenzweig E, Goldy J, Haydock A *et al*. Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. *Nat Methods*. 2006 Jul;3(7):511-8.

---

<sup>1</sup> ENCODE definition of “technical replicates” can be found at ([www.encodeproject.org/data-standards/terms/#replication](http://www.encodeproject.org/data-standards/terms/#replication))

<sup>2</sup> ENCODE definition of “biological replicates” can be found at ([www.encodeproject.org/data-standards/terms/#replication](http://www.encodeproject.org/data-standards/terms/#replication))