An introduction to the nuCpos package

Hiroaki Kato*, Takeshi Urano

April 15, 2025

1 About nuCpos

nuCpos, a derivative of *NuPoP*, is an R package for predicting **nuc**leosome **pos**itions. *nuCpos* calculates local and whole nucleosomal HBA scores for a given 147-bp sequence. This package was designed to demonstrate the use of chemical maps in prediction. As the parental package *NuPoP* now provides chemical-map-based prediction, the function for dHMM-based prediction was removed from this package. *nuCpos* continues to provide functiojns for HBA calculation. The models are based on chemical maps of nucleosomes from budding yeast *Saccharomyces cerevisiae* (Brogaard et al. (2012)), fission yeast *Schizosaccharomyces pombe* (Moyle-Heyrman et al. (2012)), or embryonic stem cells of house mouse *Mus musculus* (Voong et al. (2016)).

The parental package *NuPoP*, licensed under GPL-2, was developed by Ji-Ping Wang and Liqun Xi. Please refer to Xi et al. (2010) and Wang et al. (2008) for technical details of *NuPoP*. Their excellent codes were adapted in *nuCpos* to demonstrate the usefulness of chemical maps in prediction.

Note that when *nuCpos* was released, *NuPoP* only used an MNase-seq-based map of budding yeast nucleosomes to train a duration hidden Markov model. However, as *NuPoP* now provides chemical map-based prediction, users are encouraged to use *NuPoP* functions to conduct dHMM-based prediction in their original way.

2 nuCpos functions

nuCpos has two functions: `HBA`, and `localHBA`.

The functions `HBA` and `localHBA` receive a sequence of 147-bp DNA and calculate whole nucleosomal and local HBA scores. These functions invoke core Fortran codes for HBA calculation that were adapted from the excellent dHMM code of *NuPoP*.

nuCpos requires the *Biostrings* package, especially when DNA sequences are given as DNAString objects to the functions `HBA`, and `localHBA`. These functions can also receive DNA sequences as simple character string objects without loading the *Biostrings* package. Note: *nuCpos* requires the *NuPoP* package to perform some example runs.

Load the *nuCpos* package as follows:

```
> library(nuCpos)
```

---

*hkato@med.shimane-u.ac.jp

3 Histone binding affinity score calculation with HBA

HBA score can be calculated for a given 147-bp sequence with the `HBA` function. In the examples bellow, a character string object `inseq` and a DNAString object `INSEQ` with the same 147-bp DNA sequences are given to `HBA`. Note: the *Biostrings* package is required for the latter case.

```
> load(system.file("extdata", "inseq.RData", package = "nuCpos"))
> HBA(inseq = inseq, species = "sc")


      HBA
-2.460025


> for(i in 1:3) cat(substr(inseq, start = (i-1)*60+1,
+     stop = (i-1)*60+60), "\n")

ATCGAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCTCTAGCACCGCTTAA
ACGCACGTACGCGCTGTCCCCCGCGTTTTAACCGCCAAGGGGATTACTCCCTAGTCTCCA
GGCACGTGTCAGATATATACATCCGAT

> load(system.file("extdata", "INSEQ_DNAString.RData",
+     package = "nuCpos"))
> INSEQ

147-letter DNAString object
seq: ATCGAGAATCCCGGTGCCGAGGCCGCTCAATTGGTC...TAGTCTCCAGGCACGTGTCAGATATATACATCCGAT

> HBA(inseq = INSEQ, species = "sc")


      HBA
-2.460025
```

The argument *inseq* is the character string object to be given. Alternatively, a DNAString object can be used here. The length of DNA must be 147 bp. The argument *species* can be specified as follows: mm = *M. musculus*; sc = *S. cerevisiae*; sp = *S. pombe*.

4 Local histone binding affinity score calculation with localHBA

Local HBA scores are defined as HBA scores for 13 overlapping subnucleosomal segments named A to M. They can be calculated for a given 147-bp sequence with the `localHBA` function. Like `HBA`, this function can receive either a character string object or a DNAString object. The segment G corresponds to the central 21 bp region, in which the dyad axis passes through the 11th base position. This means that the local HBA score for the G segment implies the relationship between DNA and histone proteins at around superhelical locations -0.5 and +0.5. The neighboring F segment, which is 20 bp in length, is for SHLs -1.5 and -0.5. The result of example run shown below suggests that subsequence of `inseq` around SHL -3.5 and -2.5 is suitable for nucleosome formation.

```
> localHBA(inseq = inseq, species = "sc")
```

```
    lHBA_A      lHBA_B      lHBA_C      lHBA_D      lHBA_E      lHBA_F
-1.56140949 -1.62502354  0.48885990  2.37615568  2.90458625 -1.35195919
    lHBA_G      lHBA_H      lHBA_I      lHBA_J      lHBA_K      lHBA_L
-3.13228907 -0.32208031  0.27650871  0.01922002  0.49787625 -0.17151500
    lHBA_M
-1.27186158


> barplot(localHBA(inseq = inseq, species = "sc"),
+     names.arg = LETTERS[1:13], xlab = "Nucleosomal subsegments",
+     ylab = "local HBA", main = "Local HBA scores for inseq")
```

5 Acknowledgements

References

Wang JP, Fondufe-Mittendorf Y, Xi L, Tsai GF, Segal E and Widom J (2008). Preferentially quantized linker DNA lengths in *Saccharomyces cerevisiae*. *PLoS Computational Biology*, 4(9):e1000175.

Xi L, Fondufe-Mittendorf Y, Xia L, Flatow J, Widom J and Wang JP (2010). Predicting nucleosome positioning using a duration hidden markov model. *BMC Bioinformatics*, 11:346.

Brogaard K, Xi L, and Widom J (2012). A map of nucleosome positions in yeast at base-pair resolution. *Nature*, 486(7404):496-501.

Moyle-Heyrman G, Zaichuk T, Xi L, Zhang Q, Uhlenbeck OC, Holmgren R, Widom J and Wang JP (2013). Chemical map of *Schizosaccharomyces pombe* reveals species-specific features in nucleosome positioning. *Proc. Natl. Acad. Sci. U. S. A.*, 110(50):20158-63.

Ichikawa Y, Morohoshi K, Nishimura Y, Kurumizaka H and Shimizu M (2014). Telomeric repeats act as nucleosome-disfavouring sequences in vivo. *Nucleic Acids Res.*, 42(3):1541-1552.

Voong LN, Xi L, Sebeson AC, Xiong B, Wang JP and Wang X (2016). Insights into Nucleosome Organization in Mouse Embryonic Stem Cells through Chemical Mapping. *Cell*, 167(6):1555-1570.

Fuse T, Katsumata K, Morohoshi K, Mukai Y, Ichikawa Y, Kurumizaka H, Yanagida A, Urano T, Kato H, and Shimizu M (2017). Parallel mapping with site-directed hydroxyl radicals and micrococcal nuclease reveals structural features of positioned nucleosomes in vivo. *Plos One*, 12(10):e0186974.