



Logolas: New Frontiers in Logo Visualization

Enrichment Depletion Logos (EDLogo) | String Logos | Adaptive Scaling of Logos

Kushal K Dey* Dongyue Xie Matthew Stephens*

Department of Statistics, University of Chicago
*Department of Human Genetics, University of Chicago

Project Page

<https://kkdey.github.io/Logolas-pages>

Gallery

<https://kkdey.github.io/Logolas-pages/Gallery.html>

R package

Github Version
kkdey/Logolas

Bioconductor
Logolas

CONTACT

Kushal K Dey
Email: kkdey@uchicago.edu
Github: [kkdey](https://github.com/kkdey)
Twitter: [kanishkadey](https://twitter.com/kanishkadey)
Website: kkdey.github.io

ACKNOWLEDGEMENTS

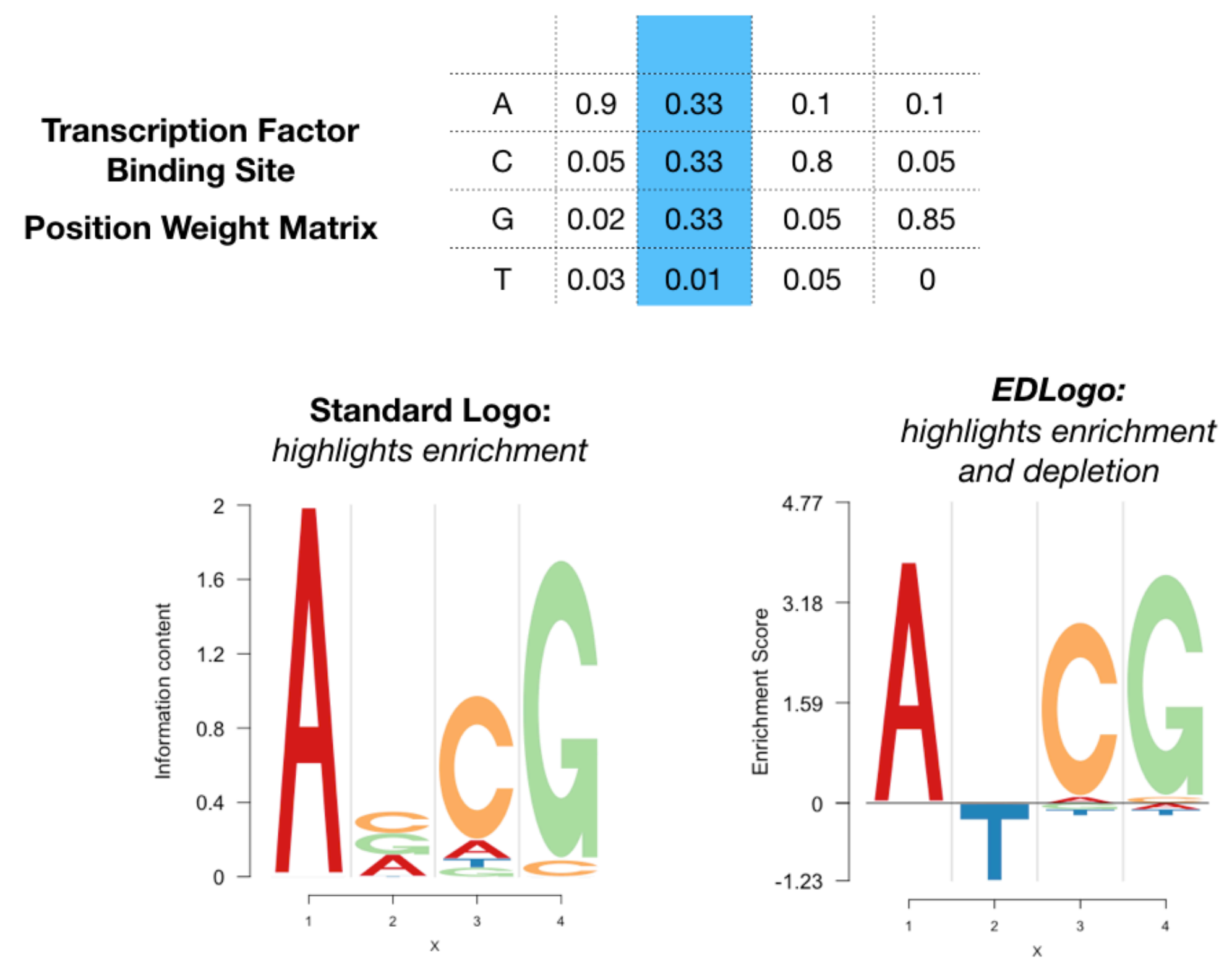
John D. Blischak,
Hussein Al-Asadi,
Alex E. White,
Trevor Price

REFERENCES

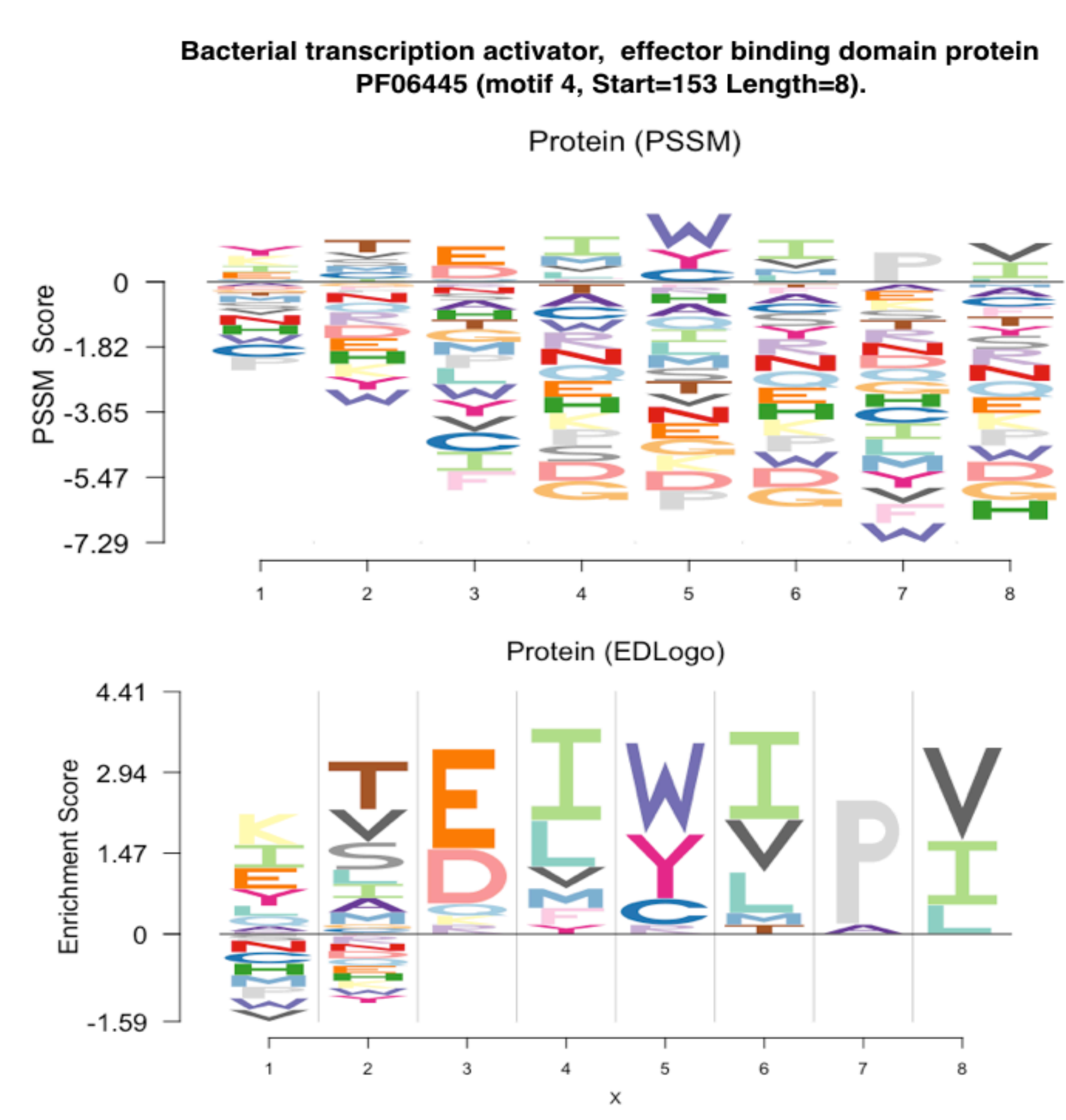
Stephens, M.: False discovery rates: a new deal. *Biostatistics* 18 (2), 275–294 (2016)
Koch, C.M., et al.: The landscape of histone modifications across 1in five human cell lines. *Genome Research* 17(6), 691–707 (2007)
Alexandrov, L., Nik-Zainal, G., Wedge, D., Campbell, P., Stratton, M.: Deciphering signatures of mutational processes operative in human cancer. *Cell Reports* 3(1), 246–259 (2013)
Kheradpour, P., Kellis, M.: Systematic discovery and characterization of regulatory motifs in encode tf binding experiments. *Nucleic Acids Research*, 1–12 (2013)
Joseph, A.P., Shingate, P., Upadhyay, A.K., Sowdhamini, R.: 3pfdb+: improved search protocol and update for the identification of representatives of protein sequence domain families. *Database (Oxford)* bau026 (2014)

Standard sequence logos tend to highlight enrichment primarily.

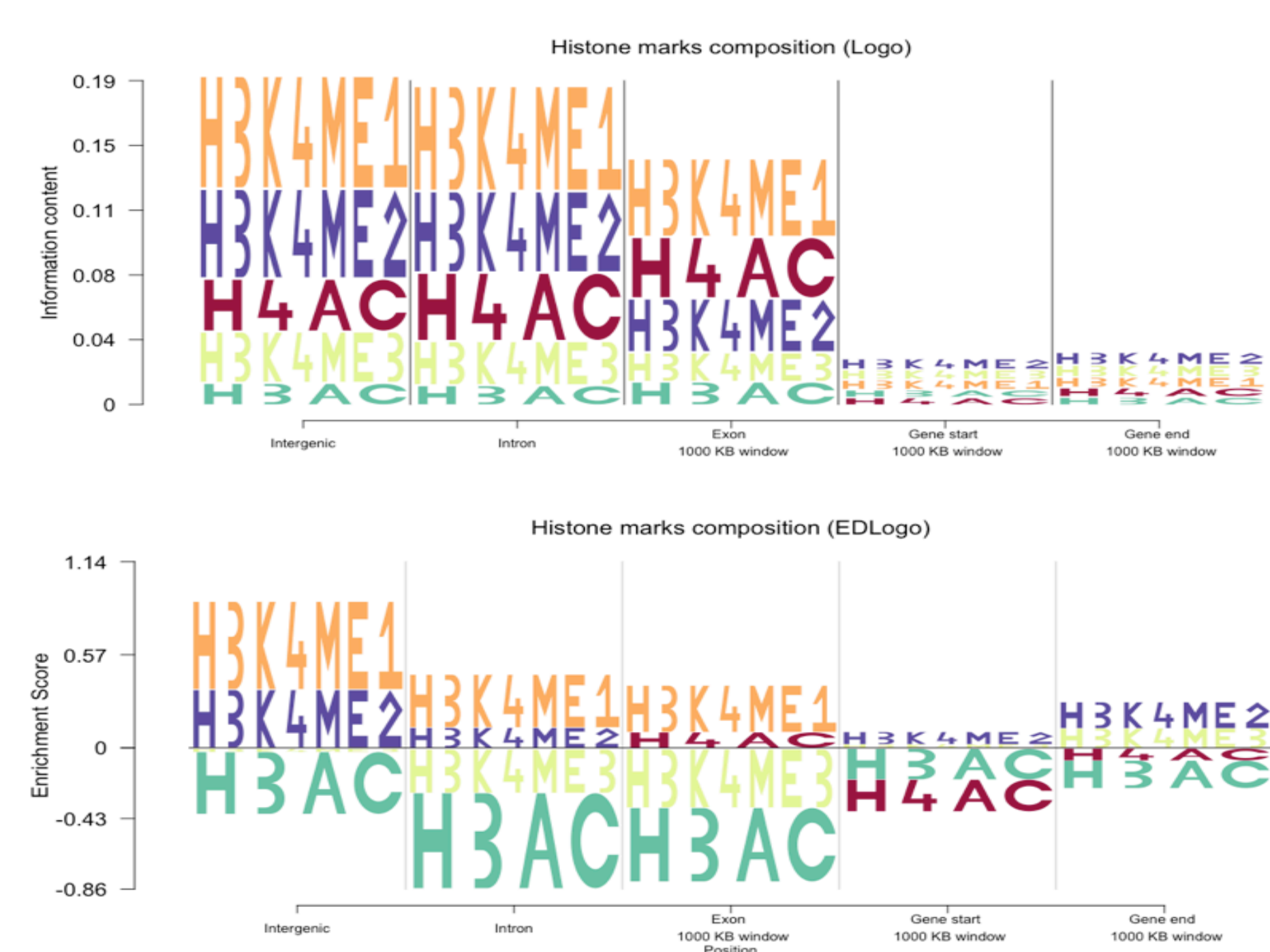
We propose *EDLogo* plots that highlight both enrichment and depletion at a position.



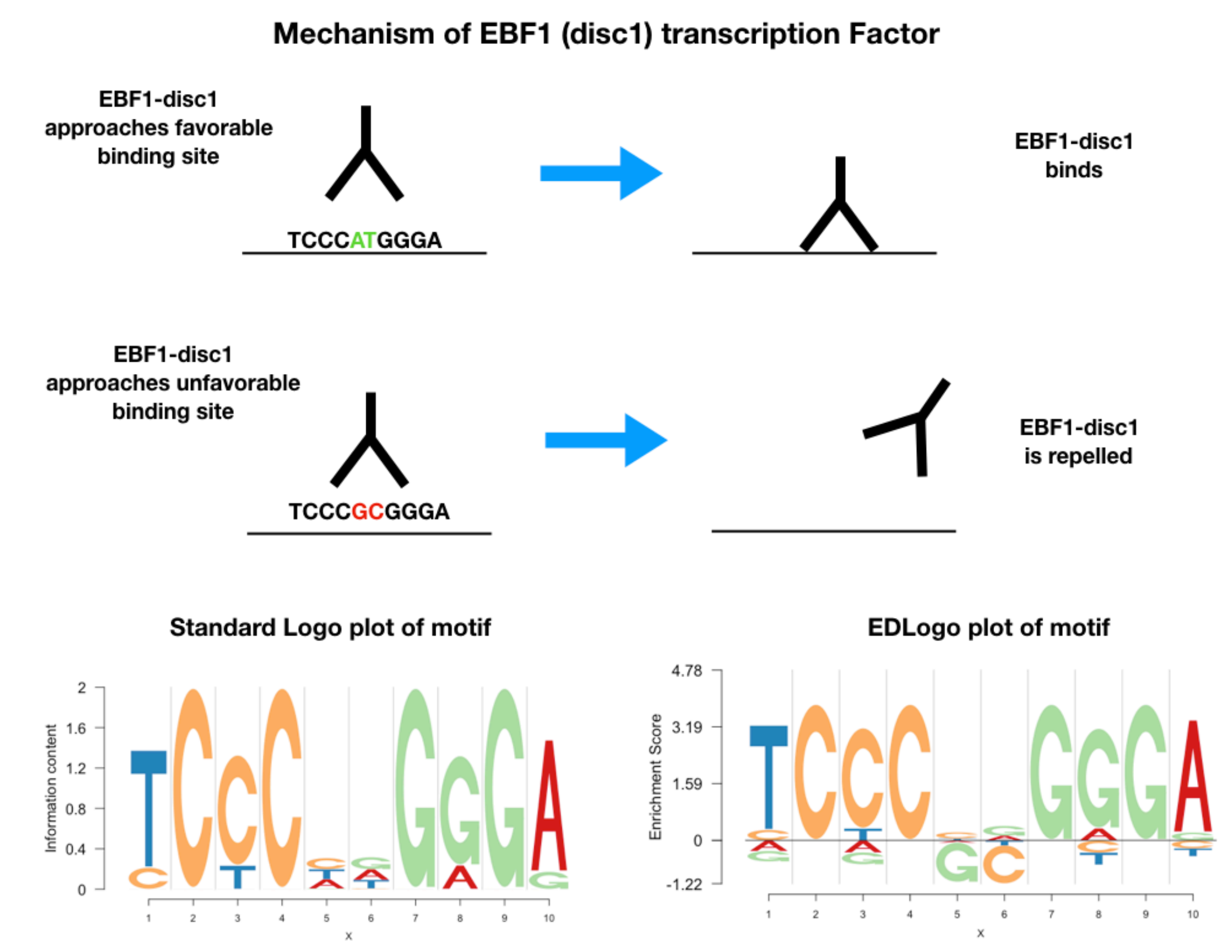
PSSM logo plots are often used to visualize protein sequence logos, but *EDLogo* representation is more parsimonious and interpretable than the PSSM logo plot.



An application of alphanumeric string symbols in *Logolas* is in modeling the composition of histone marks in different regions of the genome (sue to Koch et al 2007).

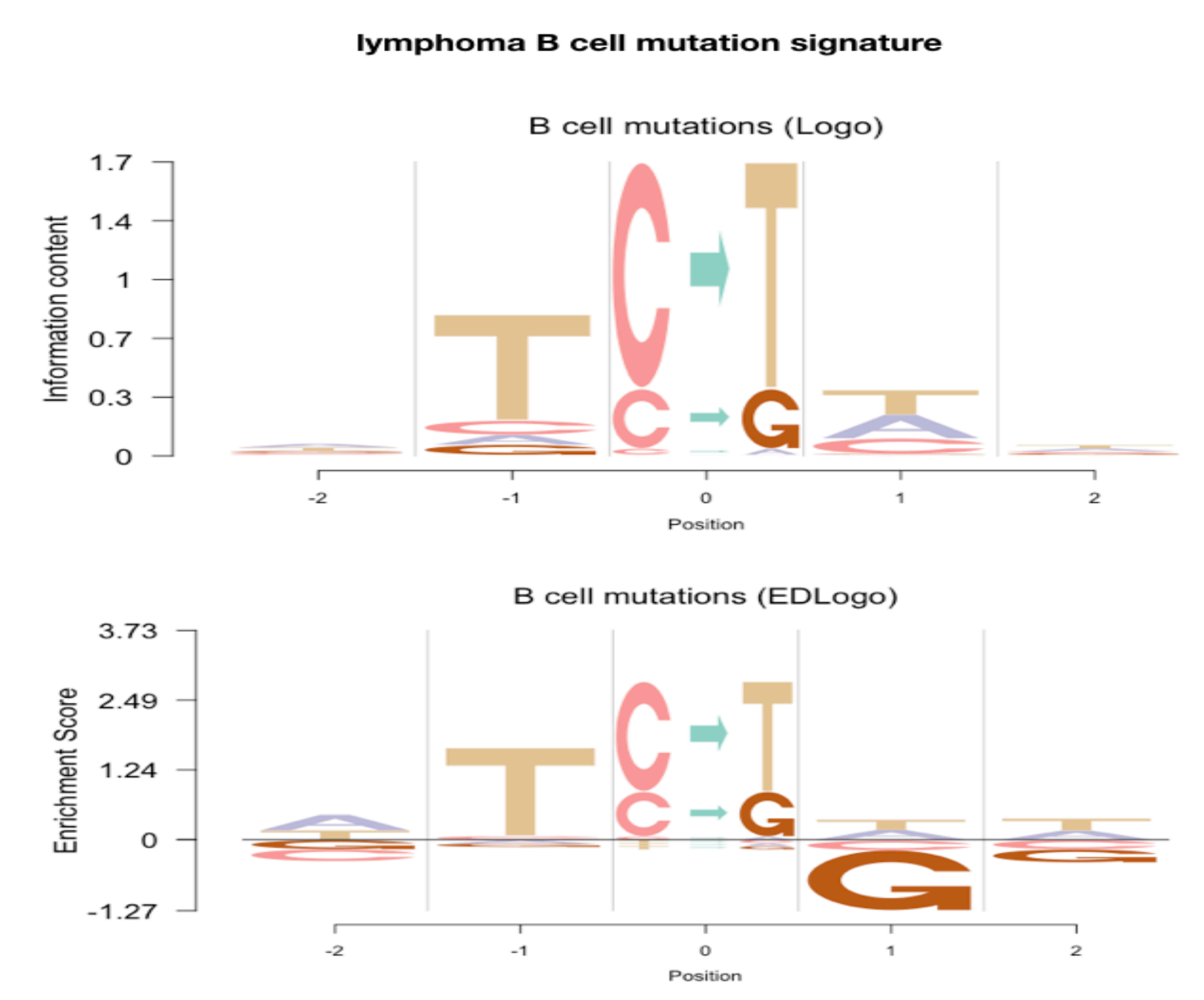


Depletion is a biological phenomenon. Several transcription factors of EBF-1 family show base depletion in the middle of the motif. An example provided below for EBF1-disc1 transcription factor.



Logolas allows the user to plot string logos. An example application of such string logos is in visualizing mutational signature profiles.

EDLogo representation can detect depletion of G to the right of C mutation dominant signatures, possibly reflecting CpG site bias. We illustrate using an example from Alexandrov et al (2013).



Logolas provides an adaptive scaling of the Position Weight Matrix based on number of aligned sequences.

This method is called Dirichlet adaptive shrinkage or *dash*.

