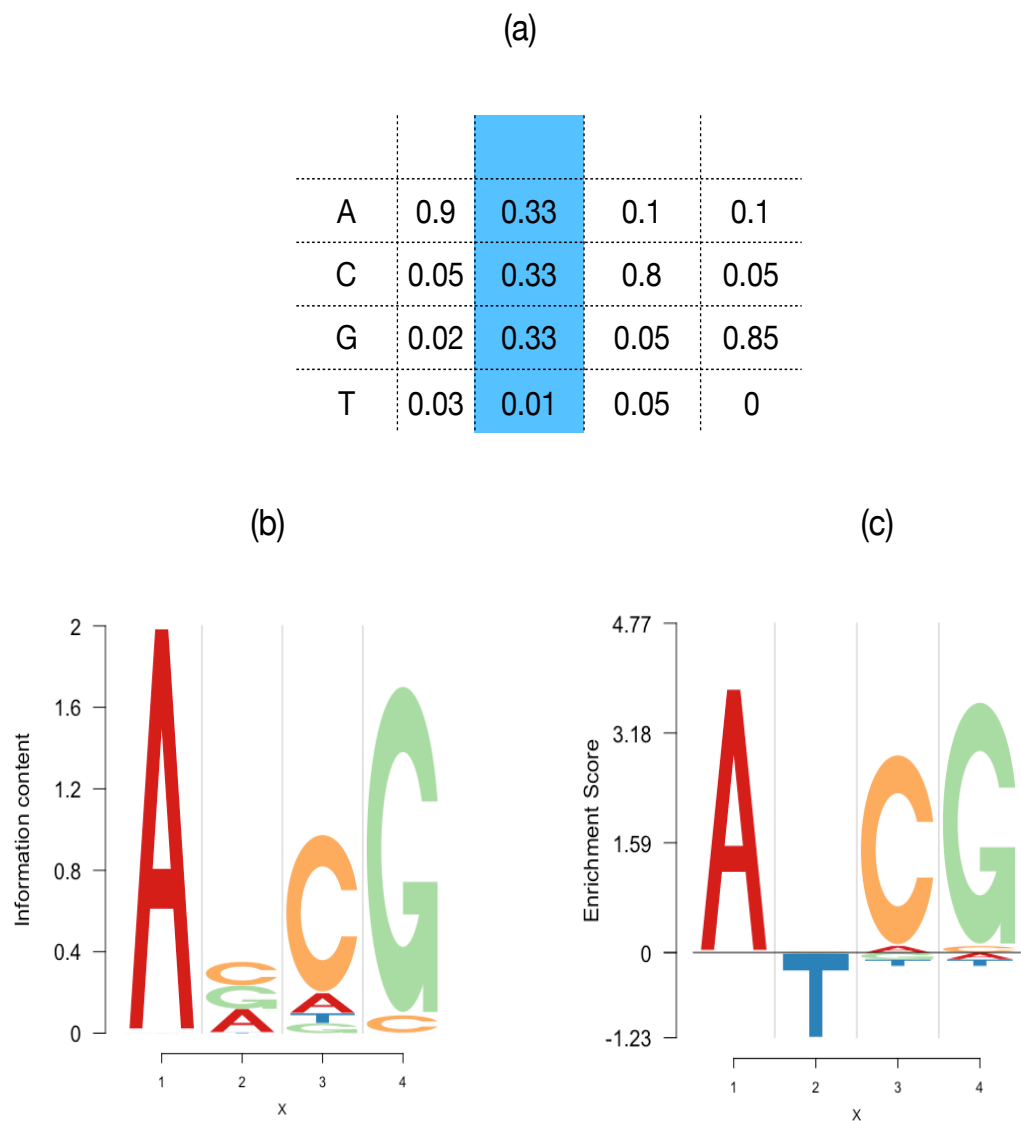
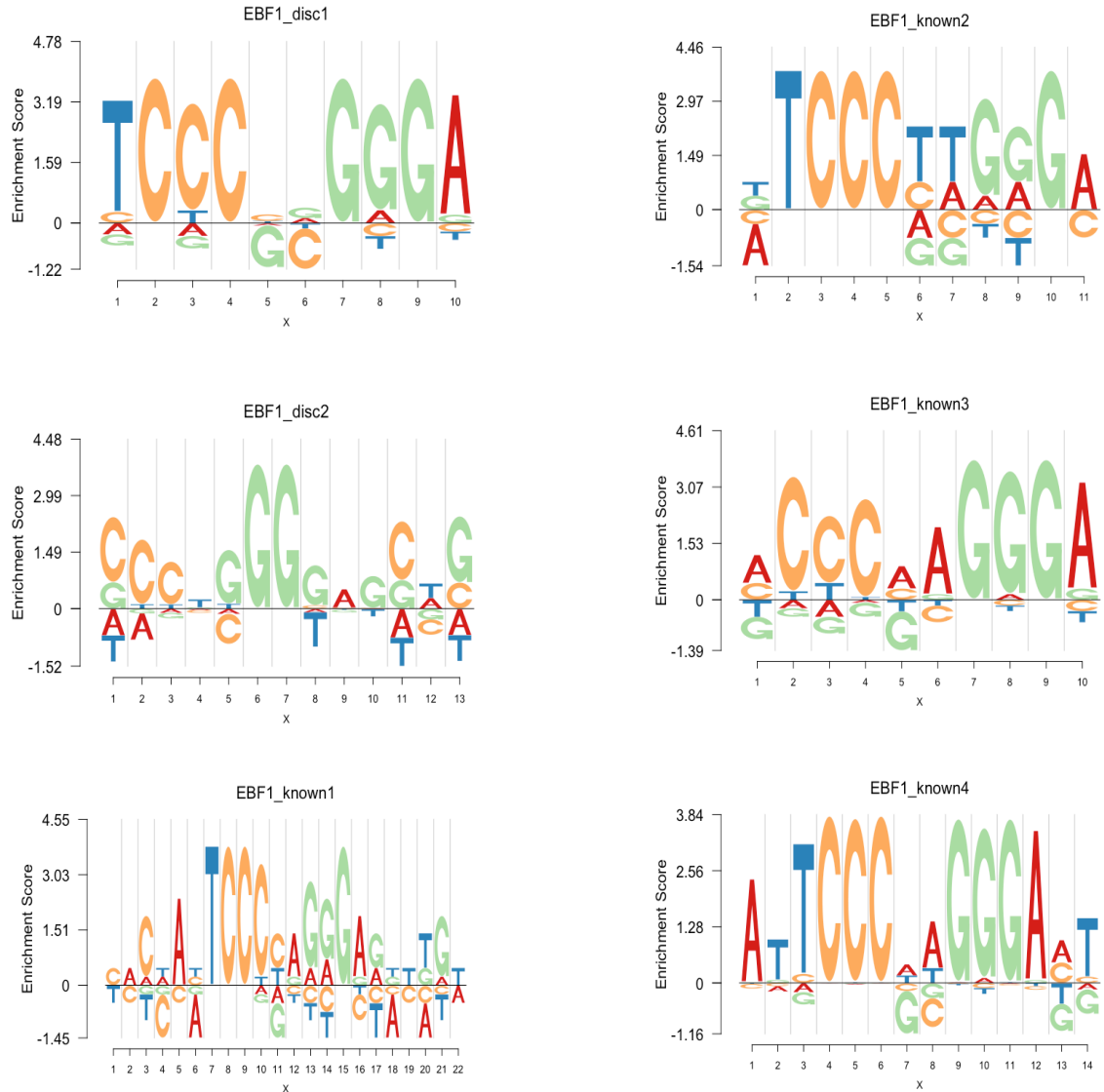


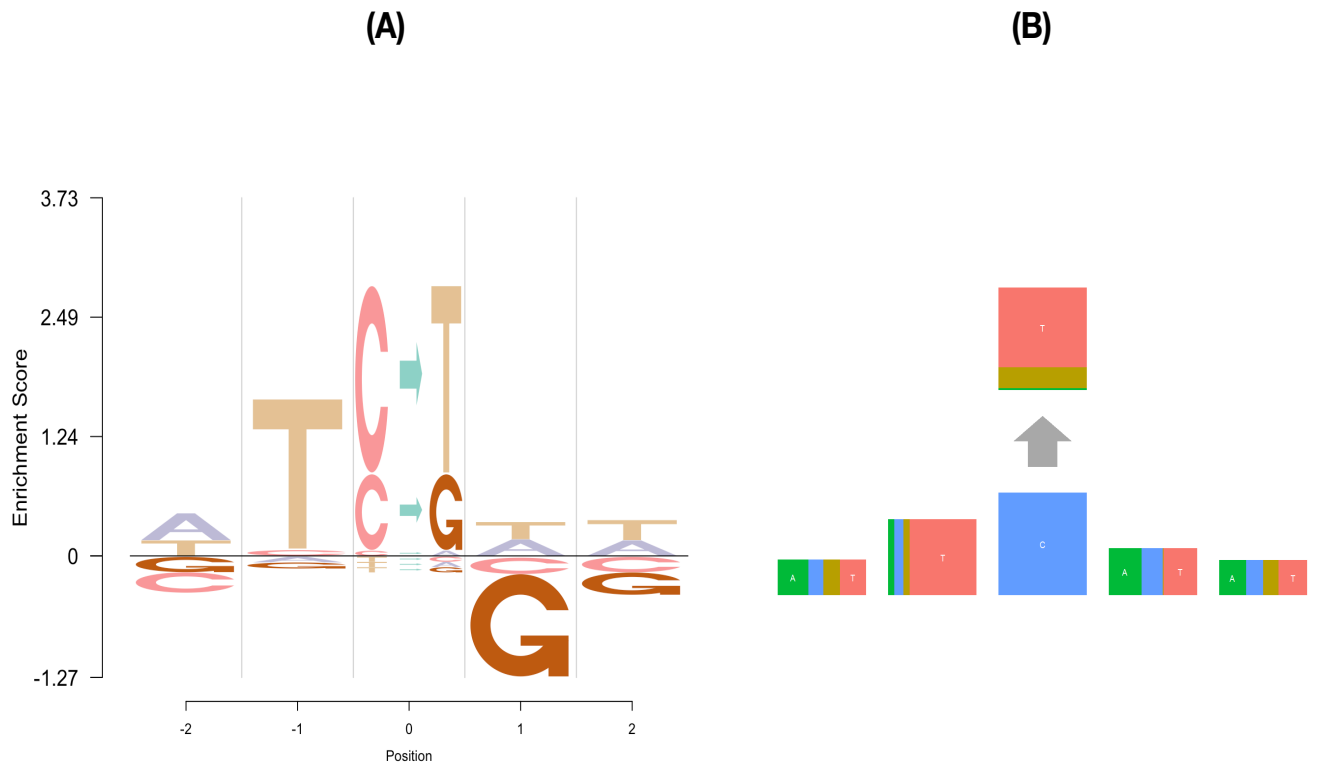
## Supplementary Figures



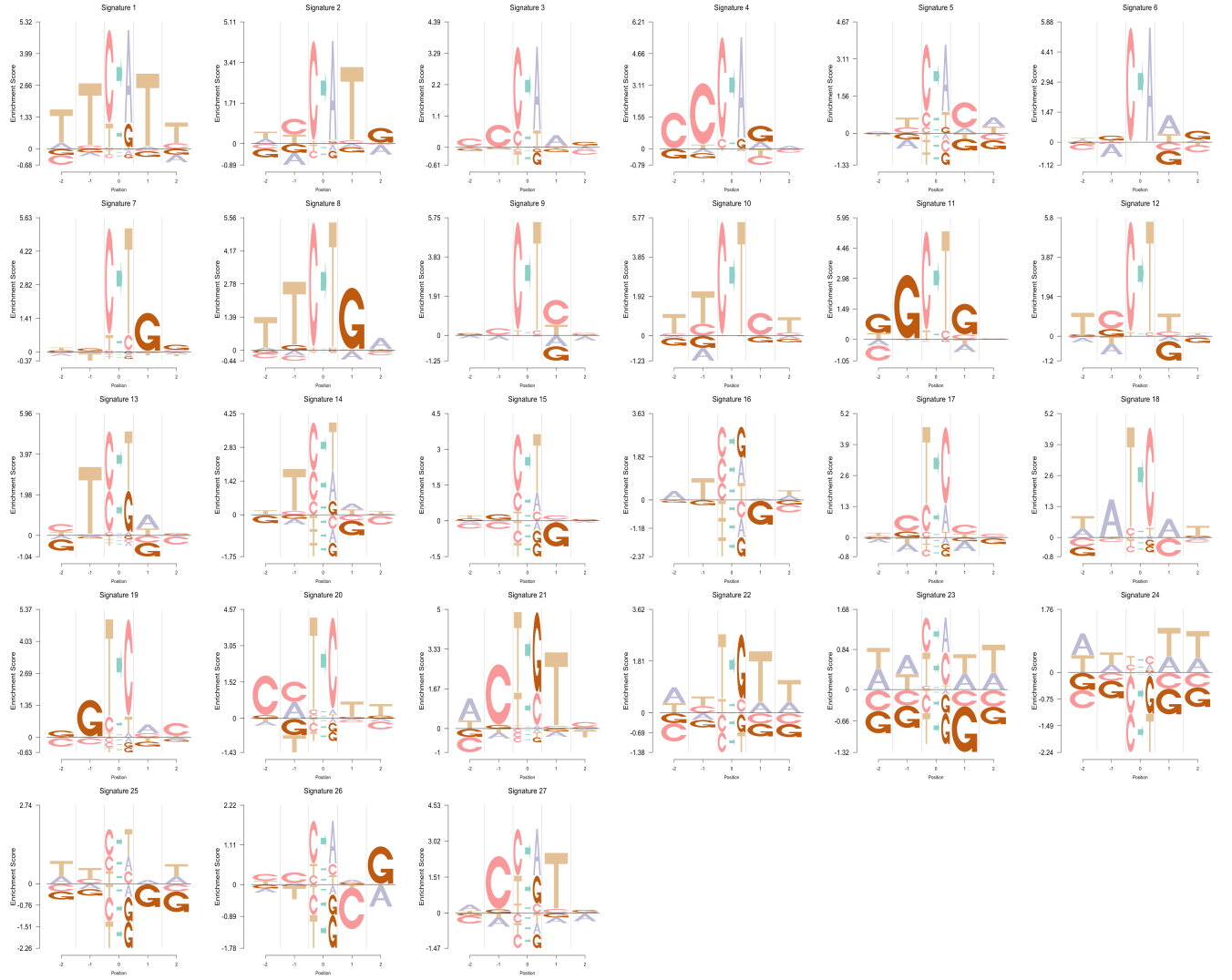
**Fig 1. Illustration of the EDLogo representation.** We present an illustration of how *EDLogo* representation accounts for depletion signal and provides a more informative visualization of the sequence motif. In panel (a), we present a position weight matrix with the position weight vector at the second position having a depletion of *T*, but is flanked by enrichments around it. In panel (b), we present the corresponding standard logo plot representation of the PWM matrix in panel (a). The signal at the second position gets swamped by the bias towards enrichment signals flanking it. In panel (c), we present the *EDLogo* representation of the PWM matrix, where both the enrichment signals as well as the depletion signal at position 2 are clearly observed.



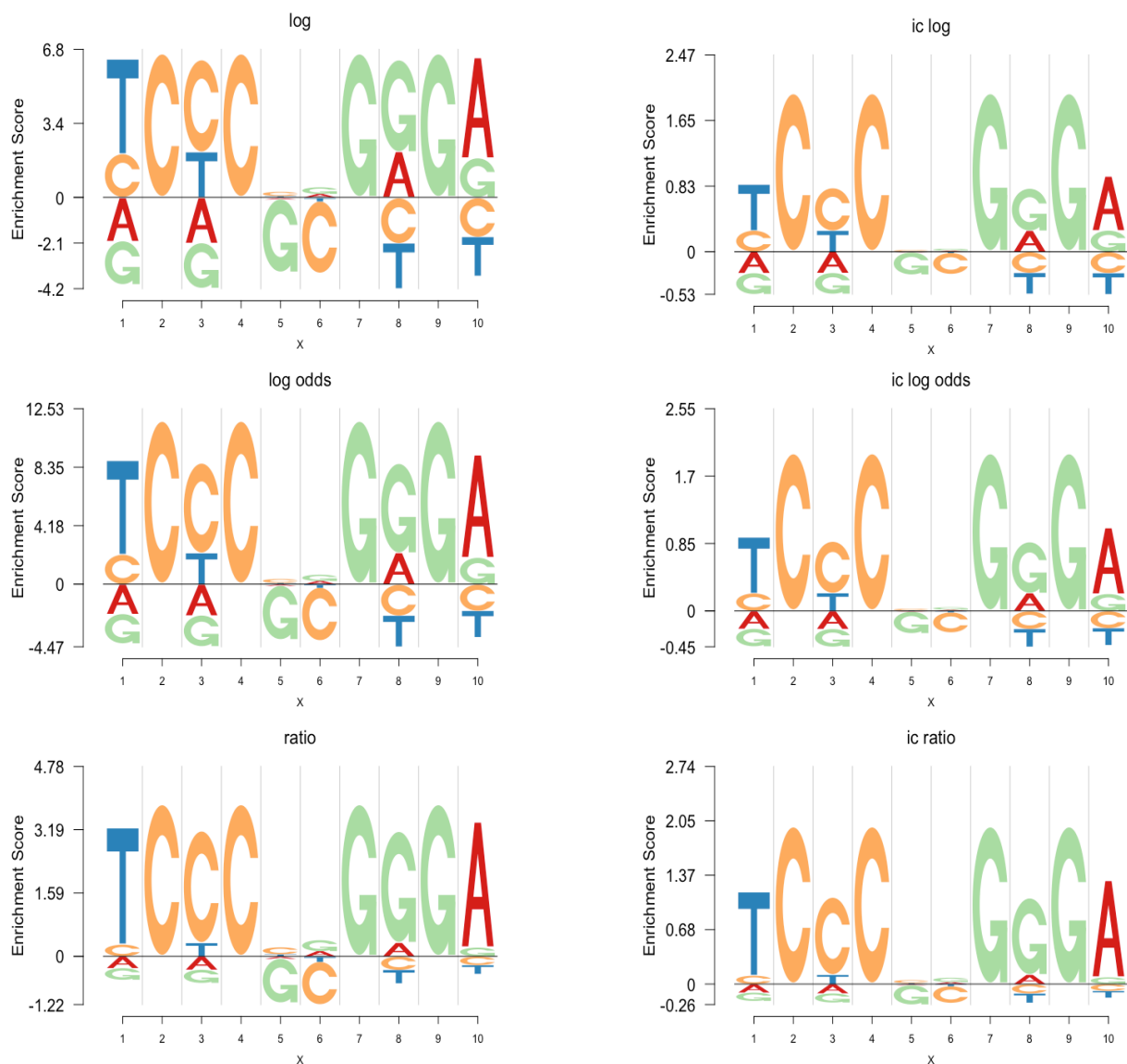
**Fig 2. EDlogo representation of the members of the EBF1 family of transcription factors:** We present the *EDlogo* representation for the binding sites of 6 transcription factors in the EBF1 family. EBF1-known4 and EBF1-disc1, and also to some extent EBF1-known3 seem to show the depletion of G and C in the middle of the binding site. The PWM data for all the transcription factors have been obtained from the ENCODE TF Chip-seq datasets and are hosted on the webpage <http://compbio.mit.edu/encode-motifs/> [?].



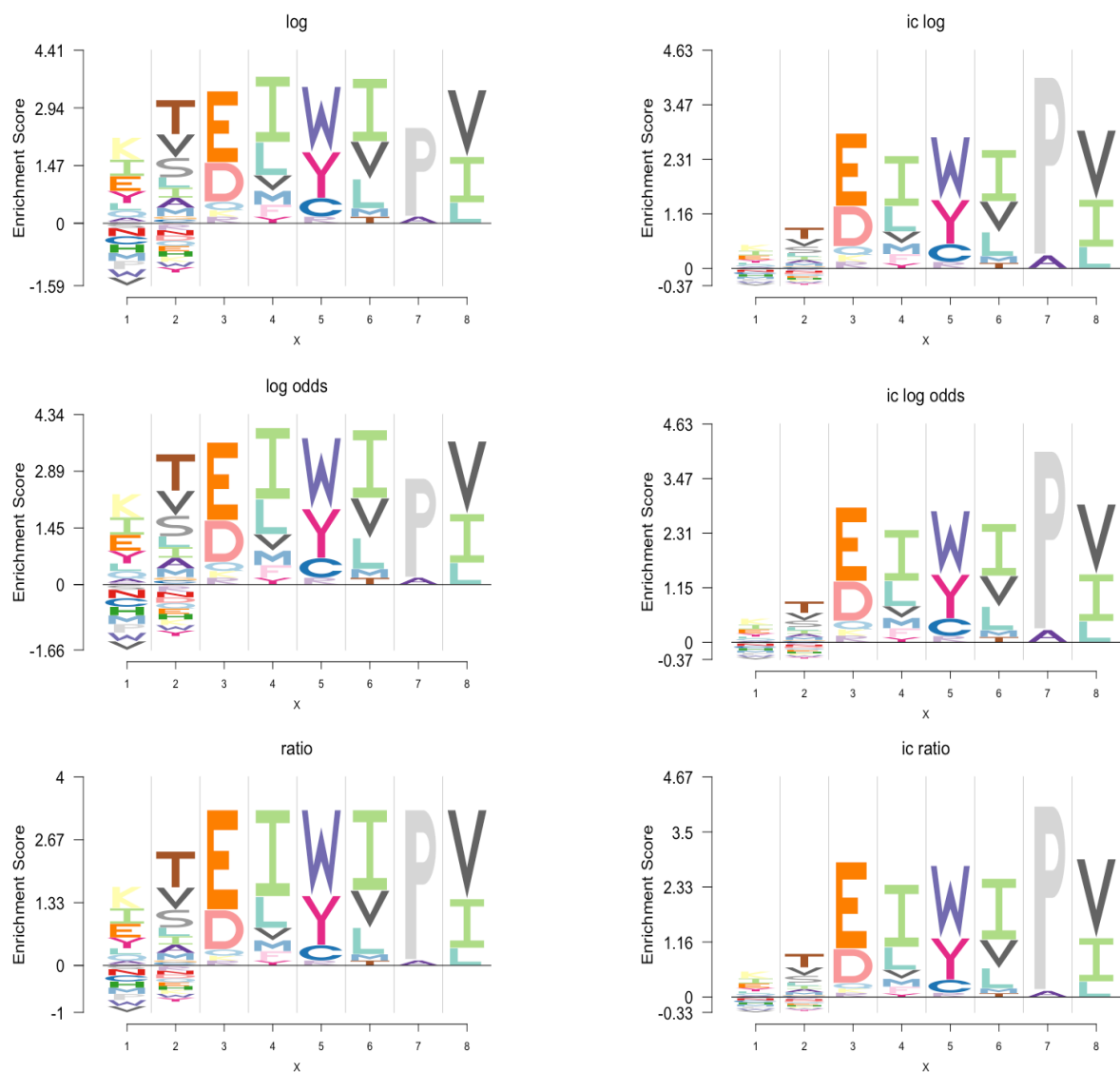
**Fig 3. Comparison of Logolas EDLogo plot with pmsignature representation for cancer mutation signatures:** We compare the *EDLogo* plot representation and the *pmsignature* representation due to Shiraishi et al (2015) [?] for mutation signature profile of lymphoma B cell from Alexandrov et al 2013 [?]. The position 0 corresponds to the mutation. Positions  $-1$  and  $-2$  correspond to the the two left flanking bases with respect to the mutation. Positions 1 and 2 correspond to the the two right flanking bases with respect to the mutation. Clearly, *EDLogo* representation shows the depletion of G at the right flanking base more clearly and is more interpretable and visually appealing in highlighting the overall mutation signature patterns compared to the *pmsignature* plot.



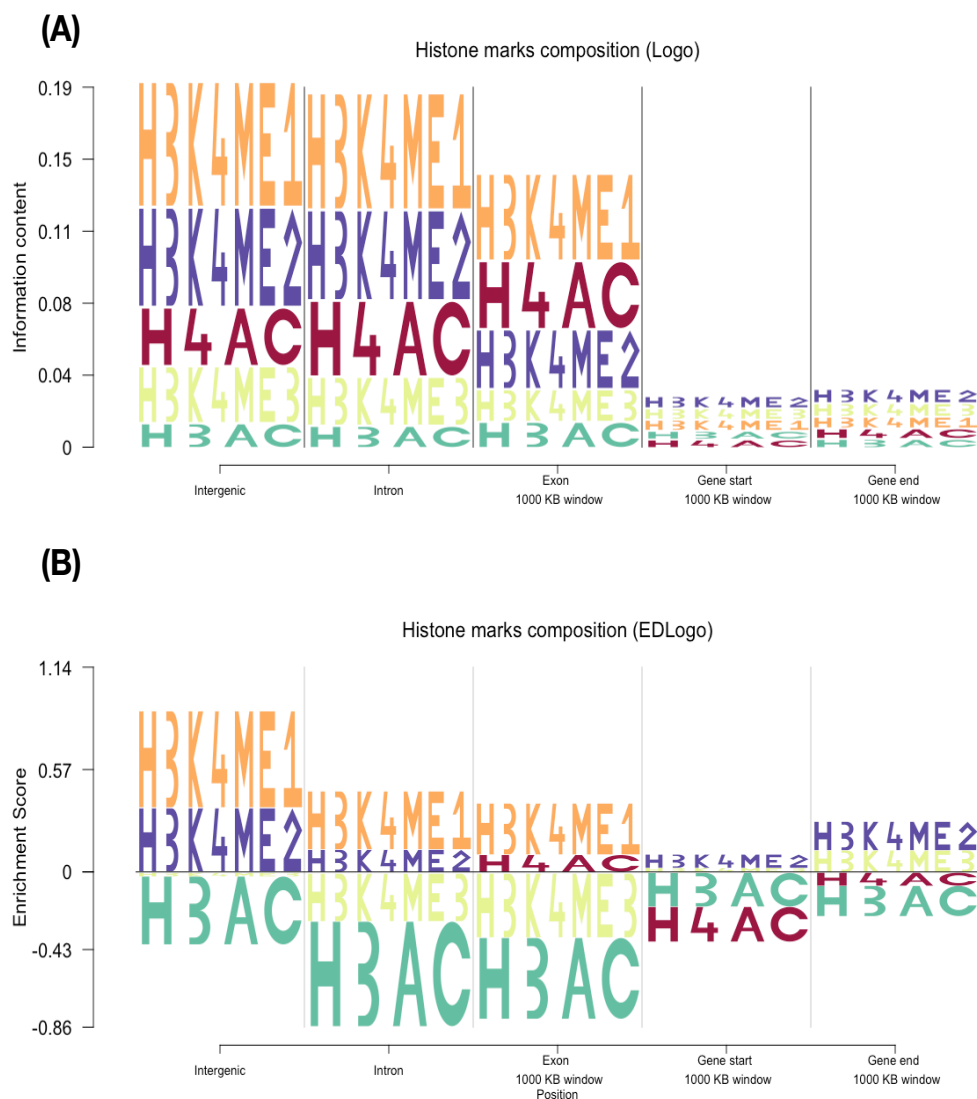
**Fig 4. EDLogo plots for the mutational signature profiles for 27 clusters in Shiraishi et al (2015):** We present the *EDlogo* representations (ratio) method for the 27 cluster signature profiles obtained from fitting a grade of membership model on the cancer mutational signature data across 30 cancer types by Shiraishi et al (2015) [?]. This plot is an alternative logo plot based representation of Figure 4 in Shiraishi et al (2015) [?].



**Fig 5. Different options for EDLogo representation - TFBS example:** We present the visualizations corresponding to the different options of the *EDLogo* plot for the EBF1-disc1 transcription factor. The details of the methods behind the various options of *EDLogo* are presented in Supplementary Methods.



**Fig 6. Different options for EDLogo representation - Protein example:** We present the visualizations corresponding to the different options of the *EDLogo* plot for the binding sequence of the Bacterial transcription activator, effector binding domain protein PF06445 (motif 4, Start=153 Length=8).



**Fig 7. Example applications of string logo plots.** We present the standard and the *EDlogo* plot representations of the composition of histone modification types that overlap with an intergenic region, intron, exon, gene start or gene end for the lymphoblastoid cell line GM06990 as reported in Koch et al 2007 [?]. These plots use alphanumeric string symbols as characters in the plot - a special feature of the package *Logolas*.