Supplementary Figures

Transcription Factor Binding Site Position Weight Matrix					
A	0.9	0.33	0.1	0.1
C	0.05	0.33	0.8	0.05
G	0.02	0.33	0.05	0.85
T	0.03	0.01	0.05	0

Standard Logo: *highlights enrichment*

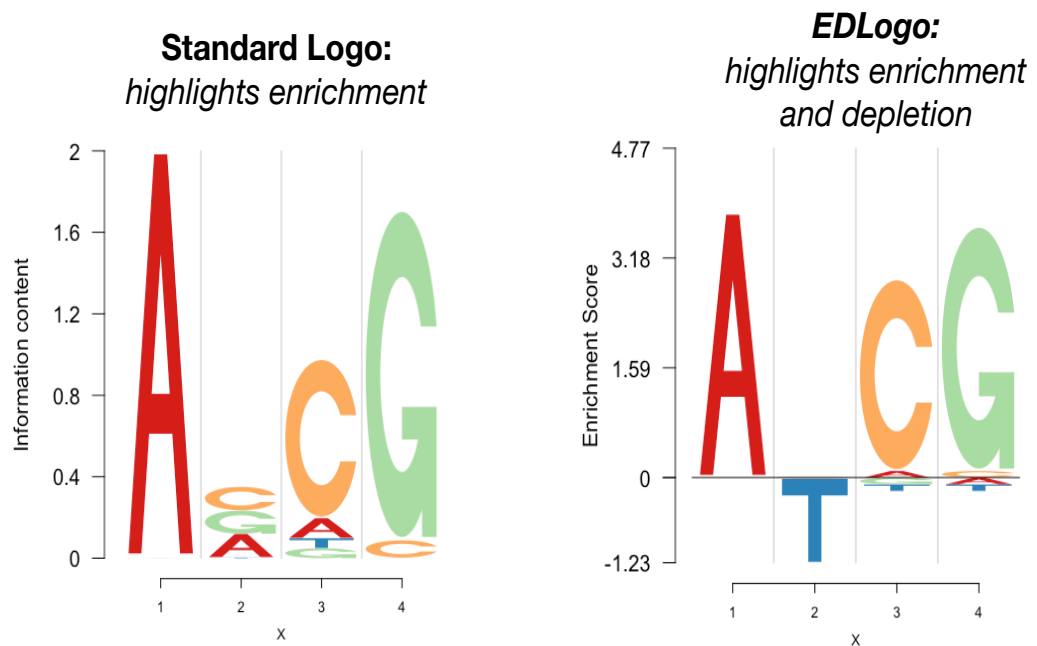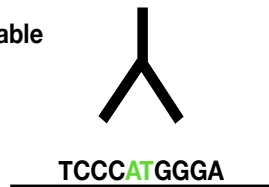**EDLogo:** *highlights enrichment and depletion*

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.
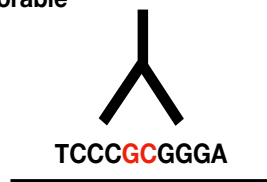
Mechanism of EBF1 (disc1) transcription Factor

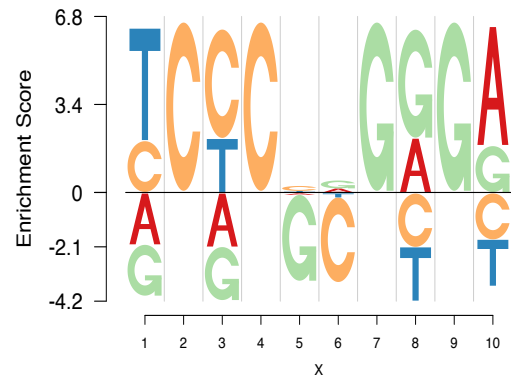**Fig 2. Mechanism behind EBF1-disc1 transcription factor binding**: We present a demo of how the loss of binding affinity of the transcription factor EBF1-disc1 to the binding site in presence of G and C in the middle of the site is reflected as a depletion signal in the *EDLogo* representation.
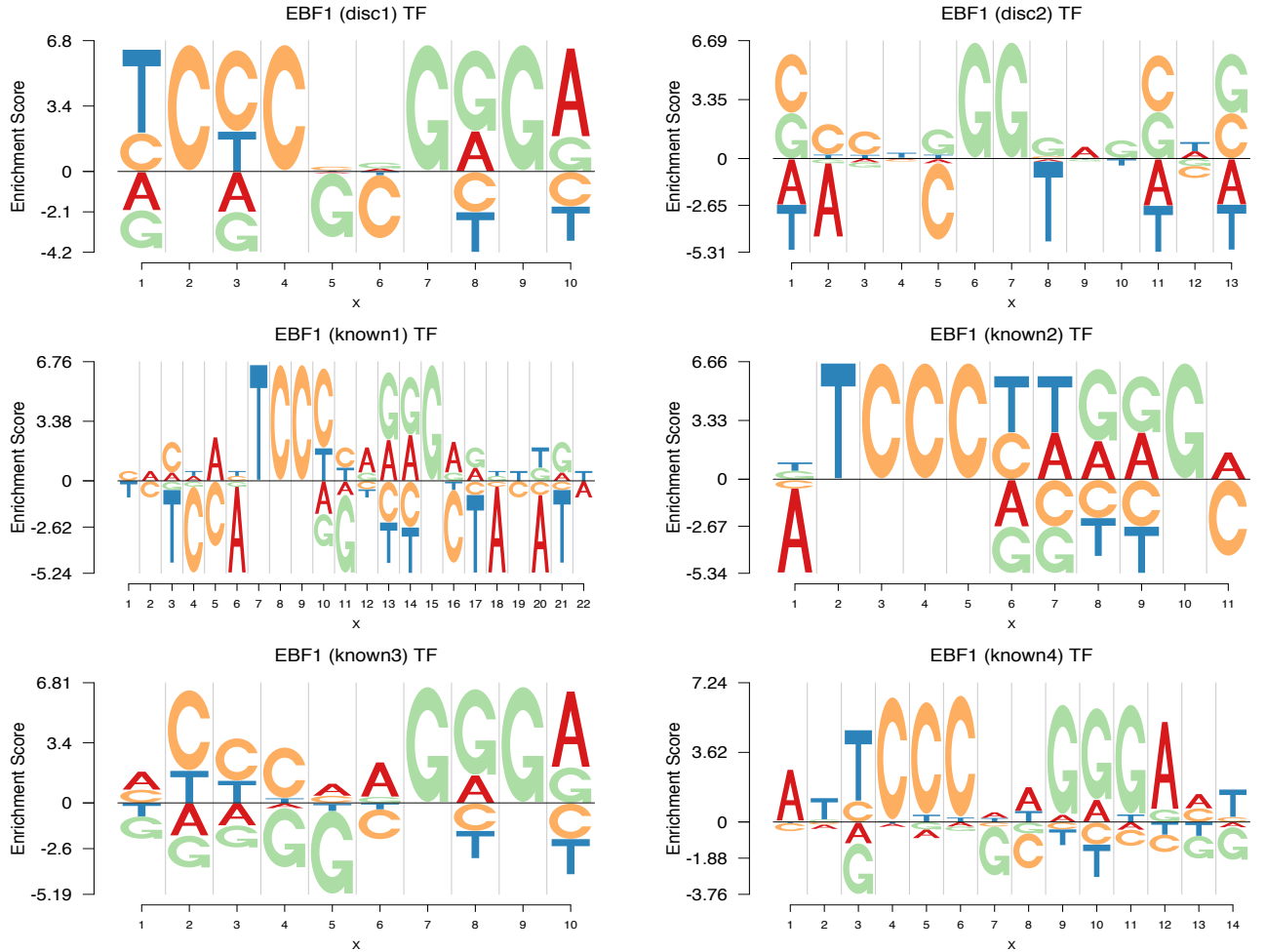
Fig 3. EDlogo representation of the members of the EBF1 family of transcription factors: We present the *EDlogo* representation for the binding sites of 6 transcription factors in the EBF1 family. EBF1-known4 and EBF1-disc1, and also to some extent EBF1-known3 seem to show the depletion of G and C in 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/ [?].
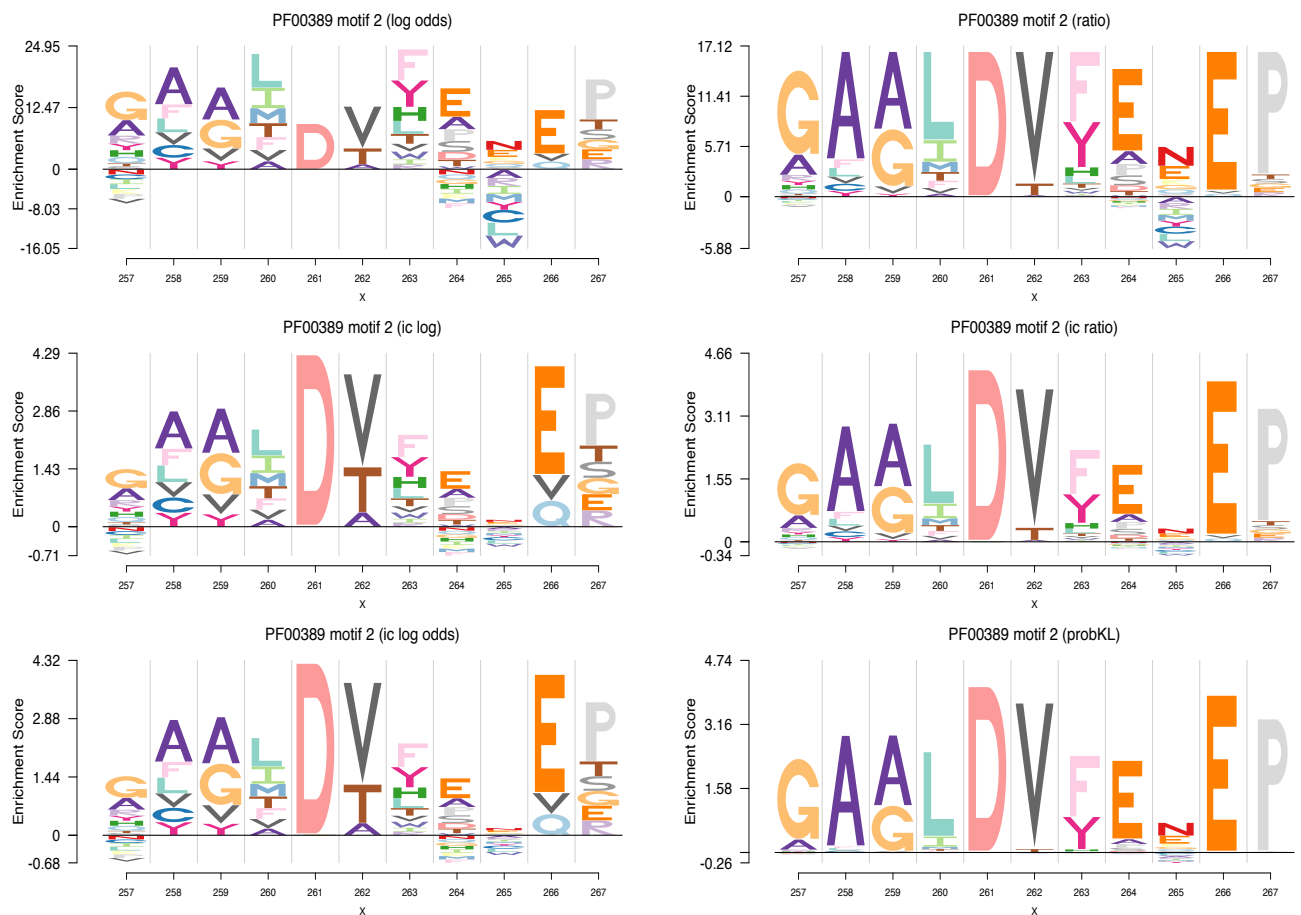
Fig 4. Different options for EDLogo representation - Protein example: We present the *EDLogo* representation of the binding motif (Motif2 Start=257 Length=11) of the protein *D-isomer specific 2-hydroxyacid dehydrogenase, catalytic domain (IPR006139)* under several other scoring schemes (*log-odds*, *ratio*, *ic-log*, *ic-ratio*, *ic-log odds* and *probKL*) besides the log based scoring used in Figure 1 Panel (B).
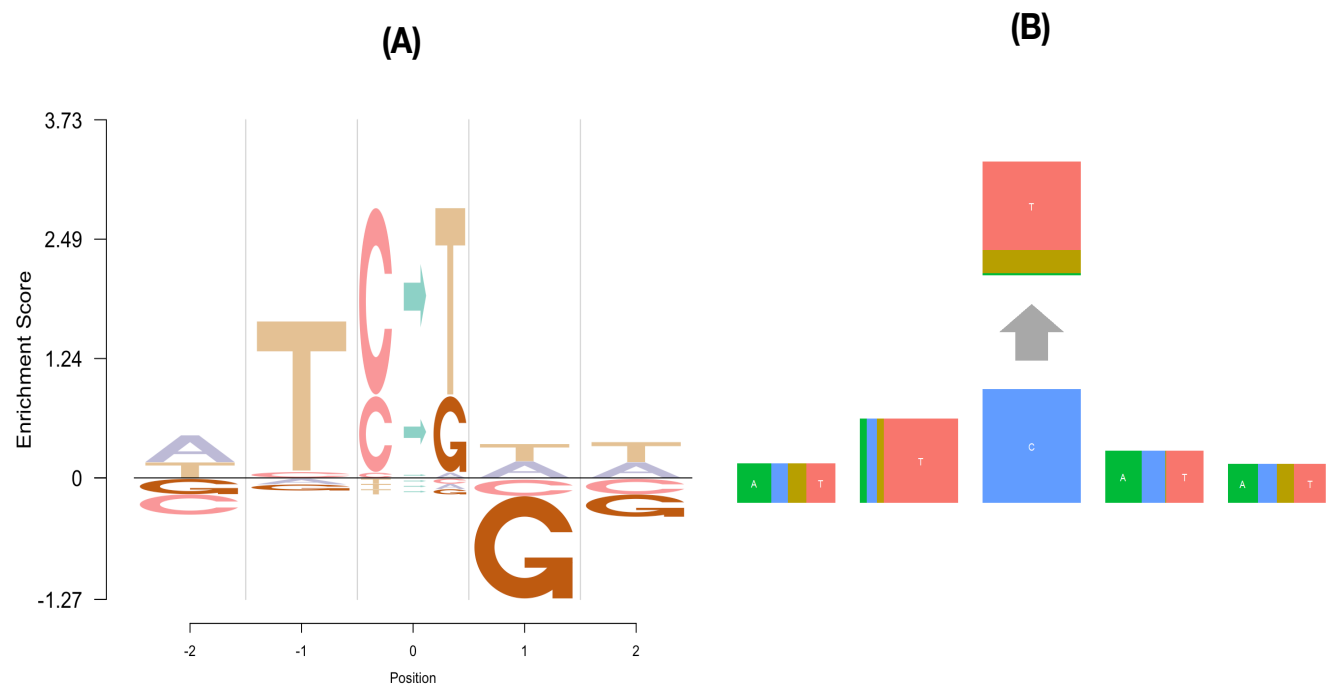
Fig 5. Comparison of Logolas EDLogo plot with pmsignature representation for cancer mutation signatures: We compare the *EDLogo* plot representation and the *pmsignature* representation due to Shiraishi et al (2015) [**?**] for mutation signature profile of lymphoma B cell from Alexandrov et al 2013 [**?**]. 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.
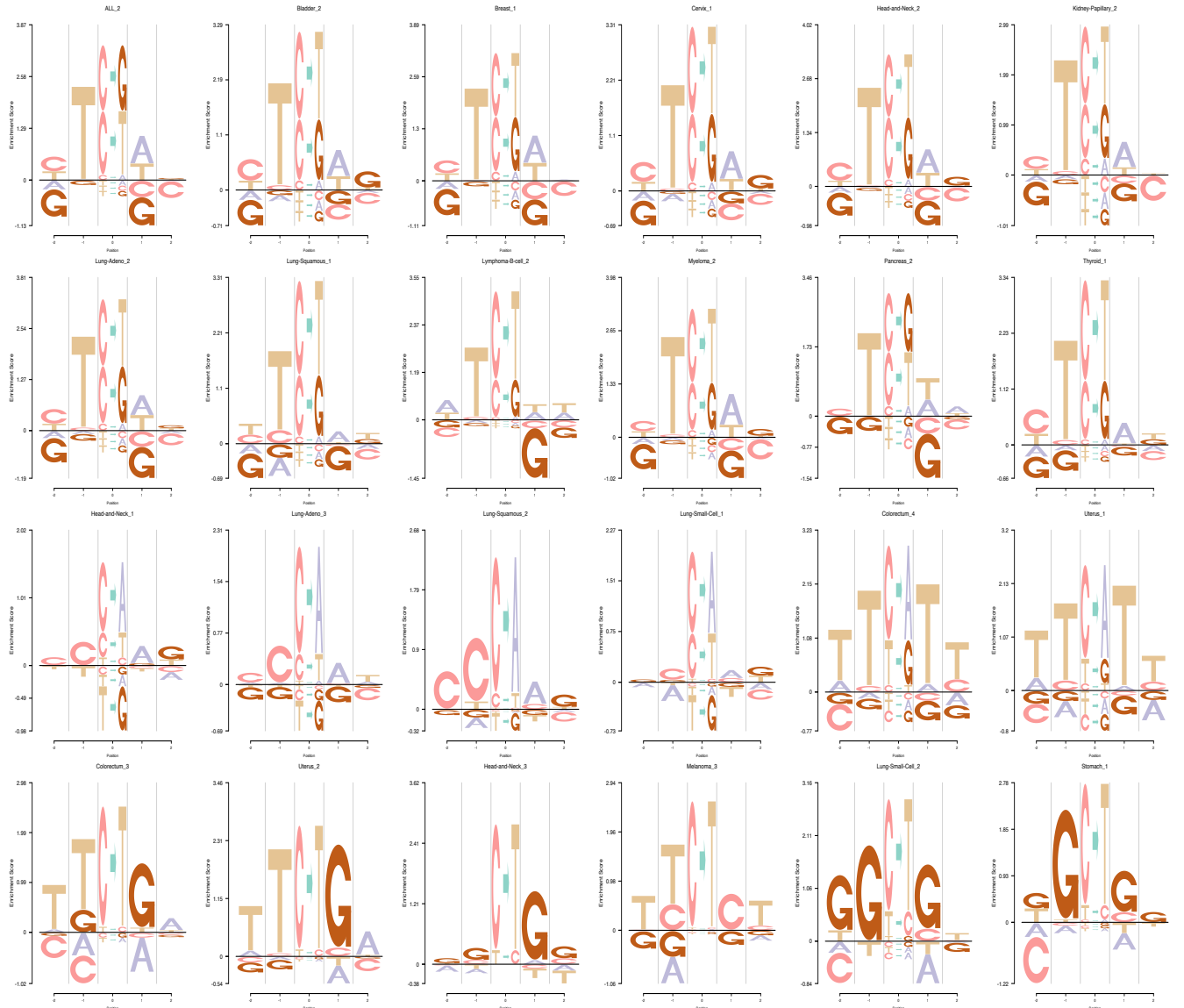
Fig 6. EDLogo plots for the mutational signature profiles of 30 cancer types in Alexandrov et al (2013): We present the sparse logo representations of the cancer mutational signature profiles across a number of tissues where the mutational signature data has been collected from 7042 cancers by Alexandrov et al (2013) [**?**]. Each mutational signature profile is represented by the mutation type at the center and the two bases flanking it to the left and two bases to the right.