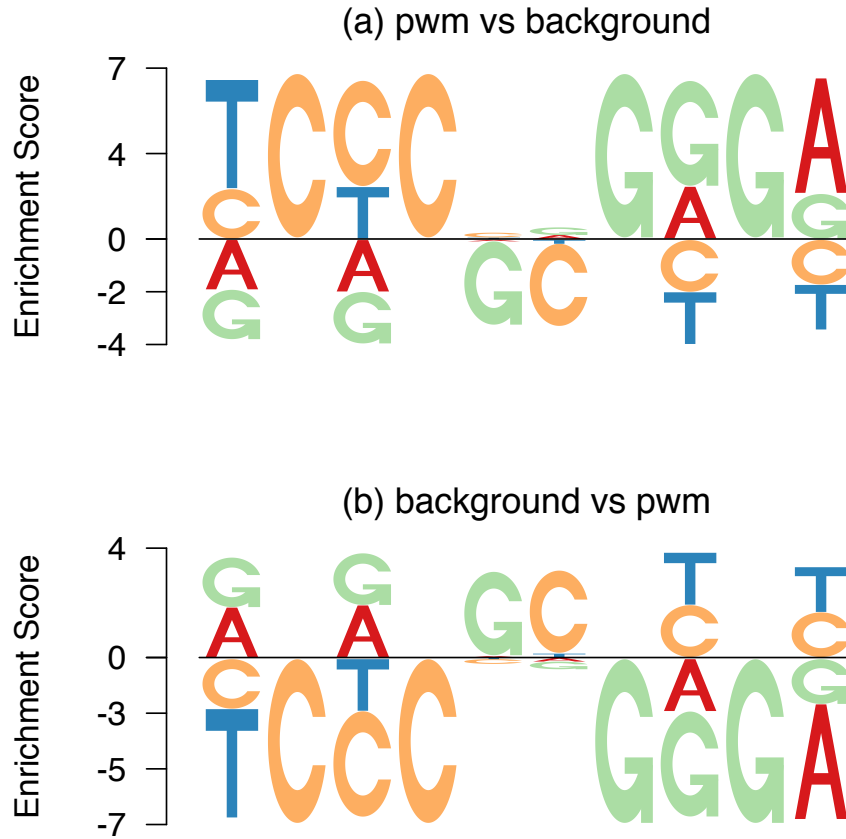# Supplementary Figures

Figure 1: Mirror Property of EDLogo: (panel a) EDLogo plot of the position weight matrix (PWM) of the primary discovered motif *disc1* (in ENCODE [Kheradpour and Kellis(2013)]) of the EBF1 transcription factor against uniform background. (panel b) EDLogo plot of a uniform PWM against the PWM of EBF1 *disc1* motif as background. The figure demonstrates the fact that, under the scoring scheme in Equation 1, the *EDLogo* plot of a position weight vector p with respect to a background weight vector q is the exact mirror image of the *EDLogo* plot of q against p as background.
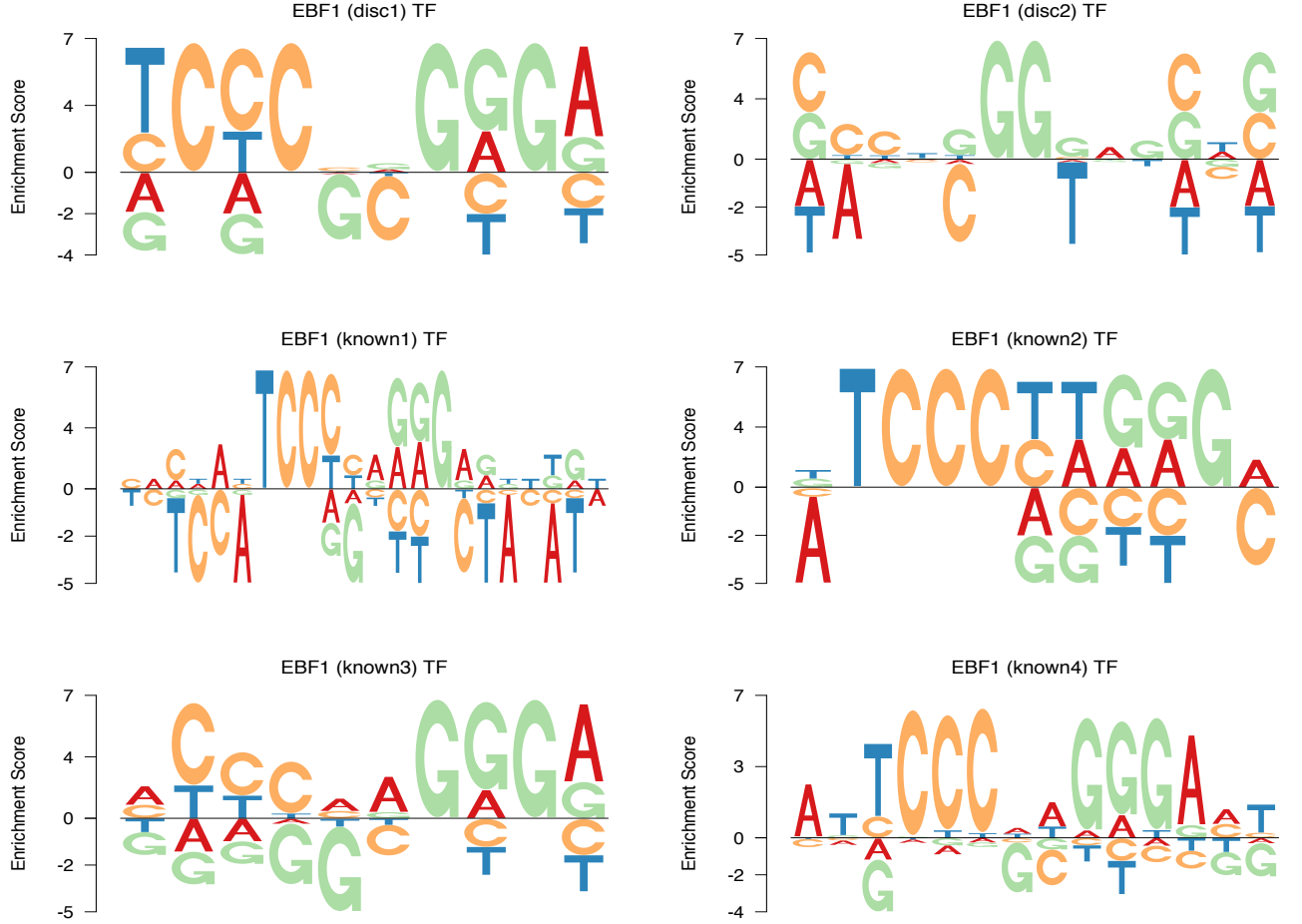
Figure 2: EDlogo plot of the different motifs of the EBF1 transcription factor: *EDlogo* plot is presented for 6 reported motifs of the transcription factor Early B cell Factor 1 (EBF1) in ENCODE project - 4 of which are previously known from literature (*known1* and *known2* from TRANSFAC database [Wingender et al.(2000)Wingender, Chen, Hehl, and et al.], *known3* from JASPAR database [Sandelin et al.(2004)Sandelin, Wynand, Engstrom, W.W., and Lenhard] and *known4* from [Jolma et al.(2013)Jolma, Yan, Whitington, Toivonen, Nitta, Rastas, Morgunova, Enge, Tai and 2 are discovered (*disc1* and *disc2*) by the ENCODE project [Kheradpour and Kellis(2013)]. Two of the known EBF1 motifs (*known3* and *known4*), along with the primary discovered motif *disc1*, showed depletion of G and C in the middle of the binding site.
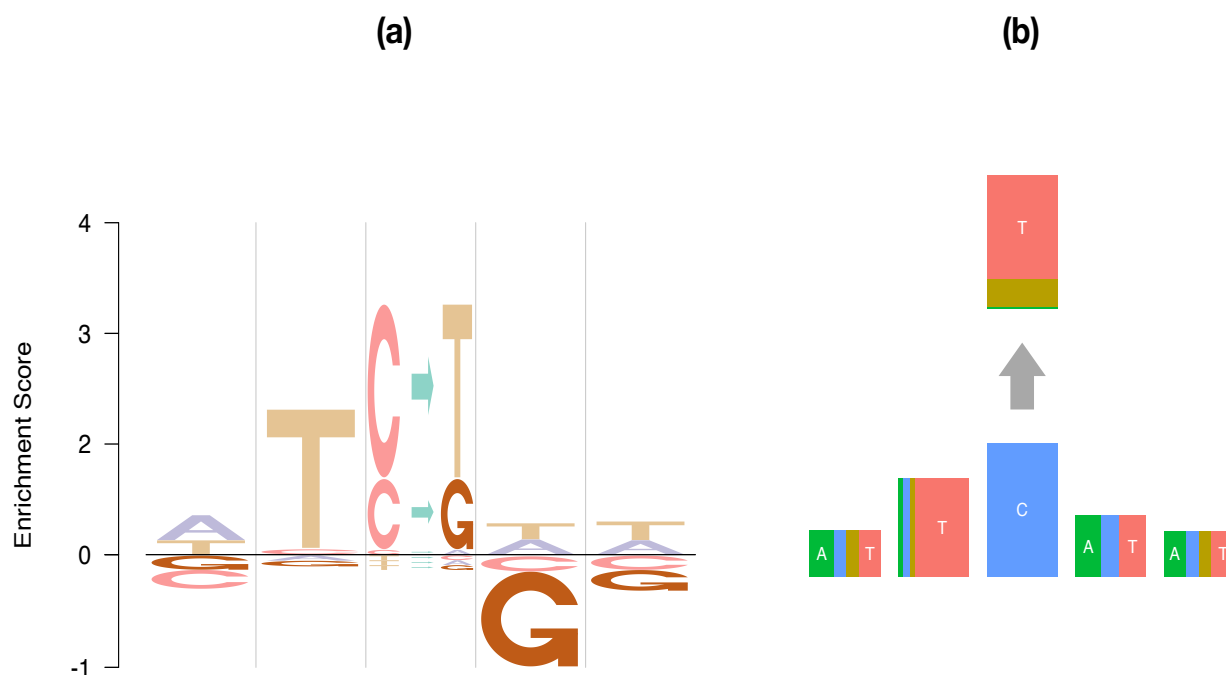
Figure 3: Comparison of the *EDLogo* plot with pmsignature plot for visualizing cancer mutation signatures: *EDLogo* plot is compared with the *pmsignature* representation due to Shiraishi et al (2015) [Shiraishi et al.(2015)Shiraishi, Tremmel, Miyano, and Stephens] for visualizing the cancer mutation signature of lymphoma B cell [Alexandrov et al.(2013)Alexandrov, Nik-Zainal, Wedge, Campbell, and Stratton]. The *EDLogo* plot shows the depletion of G at the right flanking base more clearly and is arguably more visually appealing in highlighting the overall patterns of the signature compared to the *pmsignature* plot.

References

[Alexandrov et al.(2013)Alexandrov, Nik-Zainal, Wedge, Campbell, and Stratton] L. Alexandrov, G. Nik-Zainal, D. Wedge, P. Campbell, and M. Stratton. Deciphering signatures of mutational
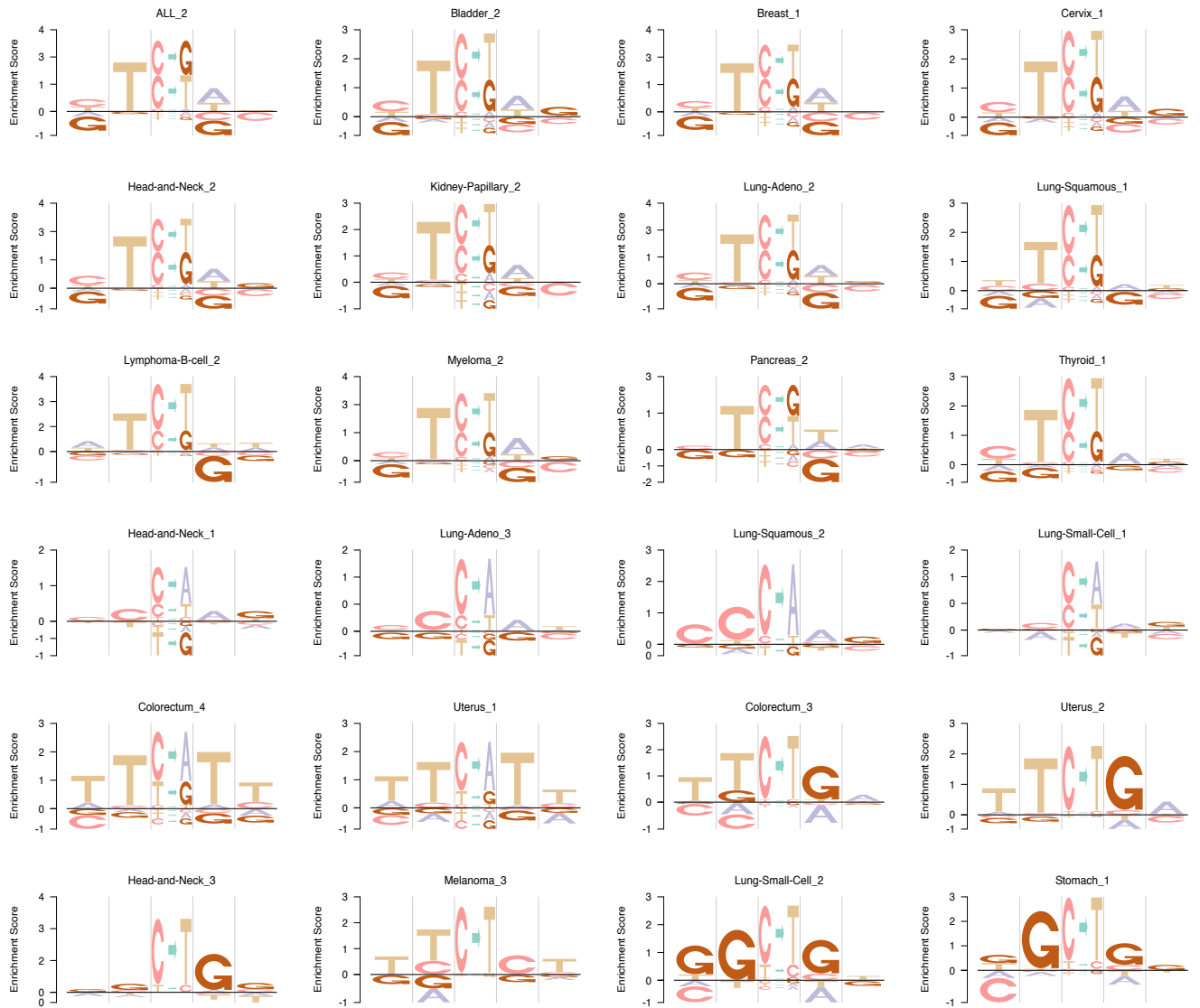
Figure 4: EDLogo plots for the mutation signature profiles of different cancer types in Alexandrov et al (2013): *EDLogo* plots of the cancer mutational signature profiles for different cancer types collected from across 7042 cancers by Alexandrov et al (2013) [Alexandrov et al.(2013)Alexandrov, Nik-Zainal, Wedge, Campbell, and Stratton].

processes operative in human cancer. *Cell Reports*, 3(1):246–259, 2013.

[Jolma et al.(2013)Jolma, Yan, Whitington, Toivonen, Nitta, Rastas, Morgunova, Enge, Taipale, Wei, and et al.] A. Jolma, J. Yan, T. Whitington, J. Toivonen, KR. Nitta, P. Rastas, E. Morgunova, M. Enge, M. Taipale, G. Wei, and et al. Dna-binding specificities of human transcription factors. *Cell*, 152: 327–339, 2013.

[Kheradpour and Kellis(2013)] P. Kheradpour and M. Kellis. Systematic discovery and characterization
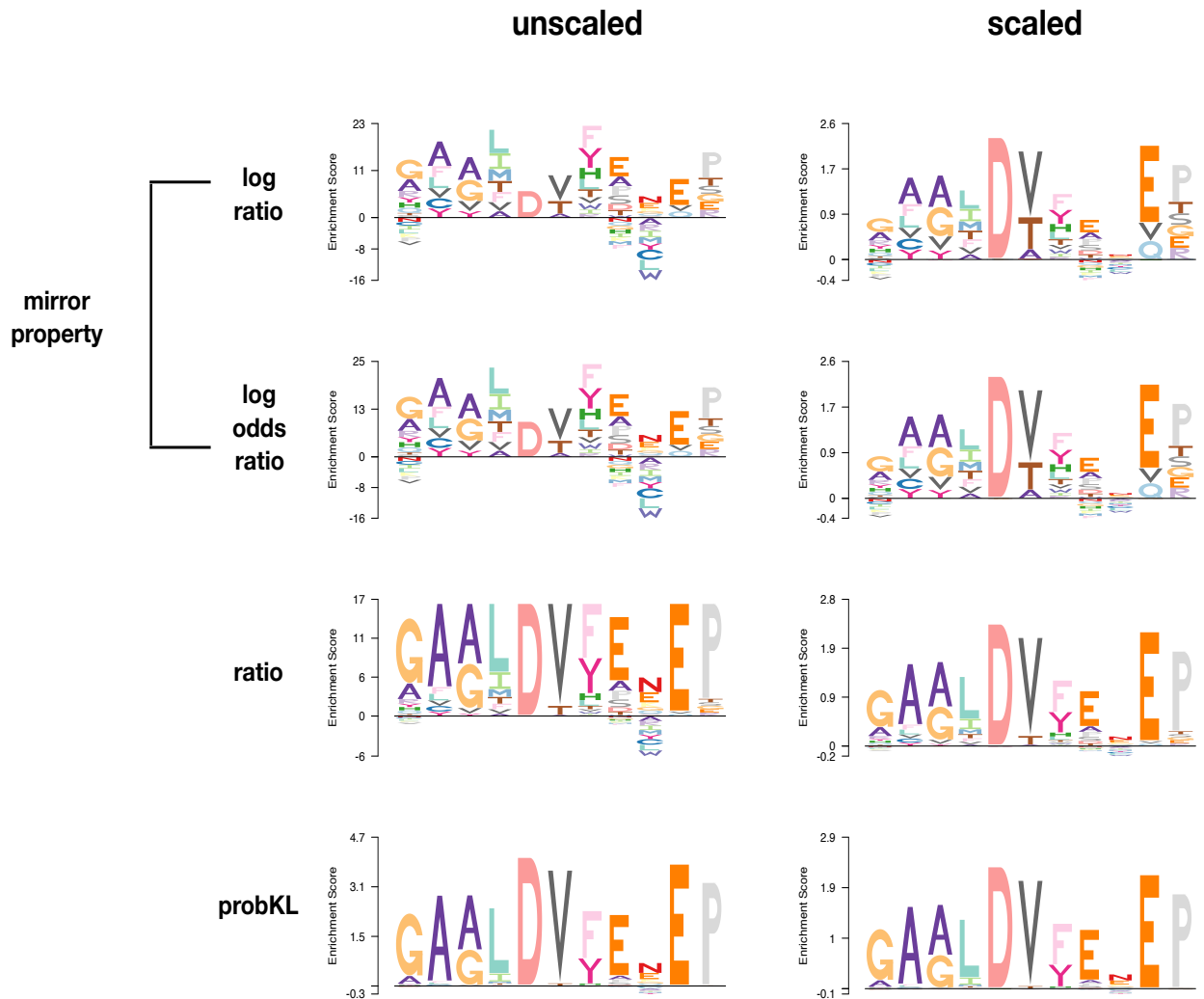
Figure 5: Different options for EDLogo plot - Protein example: *EDLogo* representation of the binding motif (Motif2 Start=257 Length=11) of the protein *D-isomer specific 2-hydroxyacid dehydrogenase, catalytic domain (IPR006139)* under several other scoring schemes (*log ratio, log odds ratio, ratio* and *probKL*) with and without the scaling by symmetric Kullback-Leibler divergence against an uniform background. The *EDLogo* plots for *log ratio* and *log odds ratio* scoring schemes show the "mirror property" with or without the scaling.

of regulatory motifs in encode tf binding experiments. *Nucleic Acids Research*, pages 1–12, 2013.

[Sandelin et al.(2004)Sandelin, Wynand, Engstrom, W.W., and Lenhard] A. Sandelin, A. Wynand, Par Engstrom, Wasserman W.W., and B. Lenhard. Jaspar: an open-access database for eukaryotic tran-
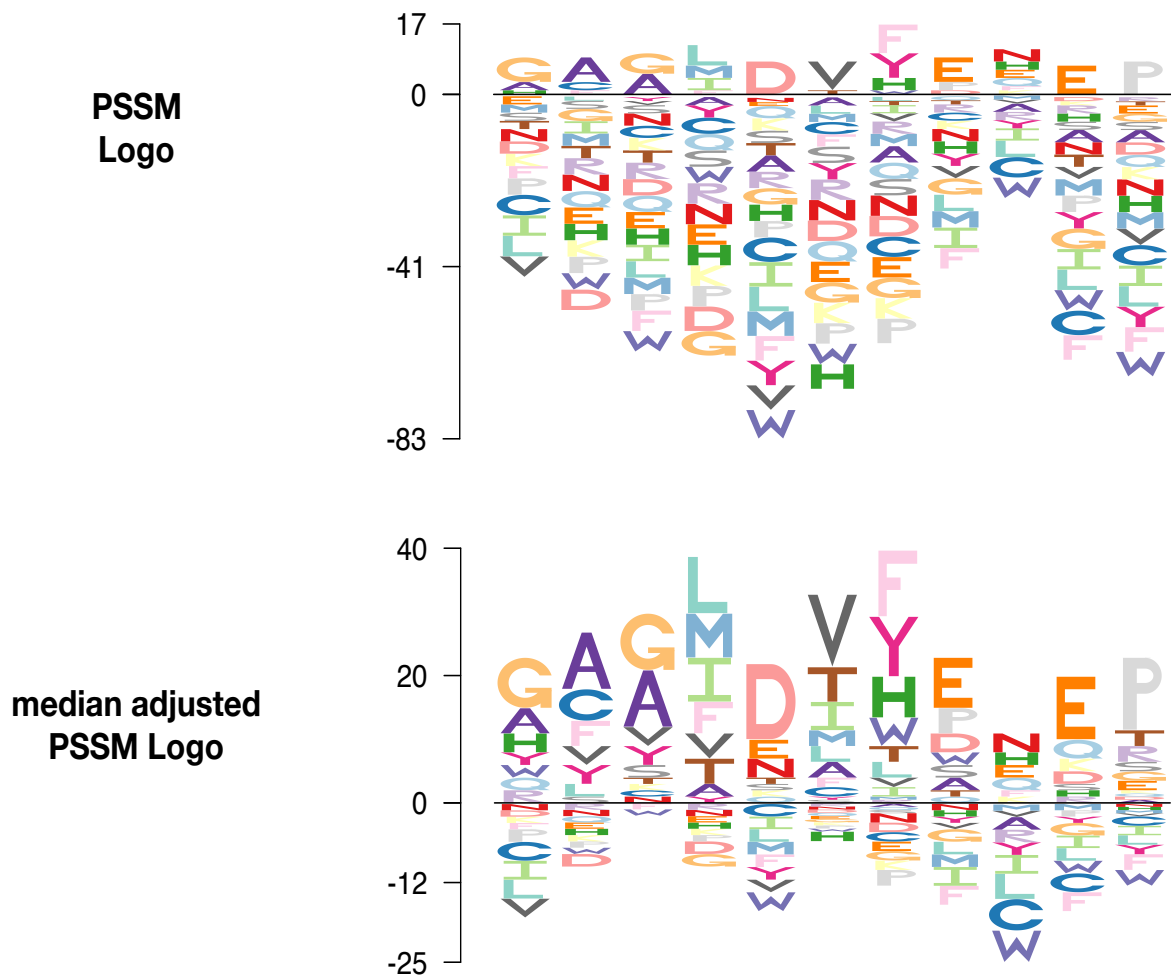
Figure 6: Logo representation of position specific scoring matrix (PSSM): A demonstration of how the median adjustment of position specific scores can reduce visual clutter in logo plot using the example of the binding motif (Motif2 Start=257 Length=11) of the protein *D-isomer specific 2-hydroxyacid dehydrogenase, catalytic domain (IPR006139)*.

scription factor binding profiles. *Nucleic Acids Research*, 32 (Database issue):D91–D94, 2004.

[Shiraishi et al.(2015)Shiraishi, Tremmel, Miyano, and Stephens] Y. Shiraishi, G. Tremmel, s. Miyano, and M. Stephens. A simple model-based approach to inferring and visualizing cancer mutation signatures. *PLoS Genetics*, 11(12):e1005657, 2015.

[Wingender et al.(2000)Wingender, Chen, Hehl, and et al.] E. Wingender, X. Chen, R. Hehl, and et al. Transfac: an integrated system for gene expression regulation. *Nucleic Acids Research*, 28(1):316–319,

2000.