Supplementary Figures

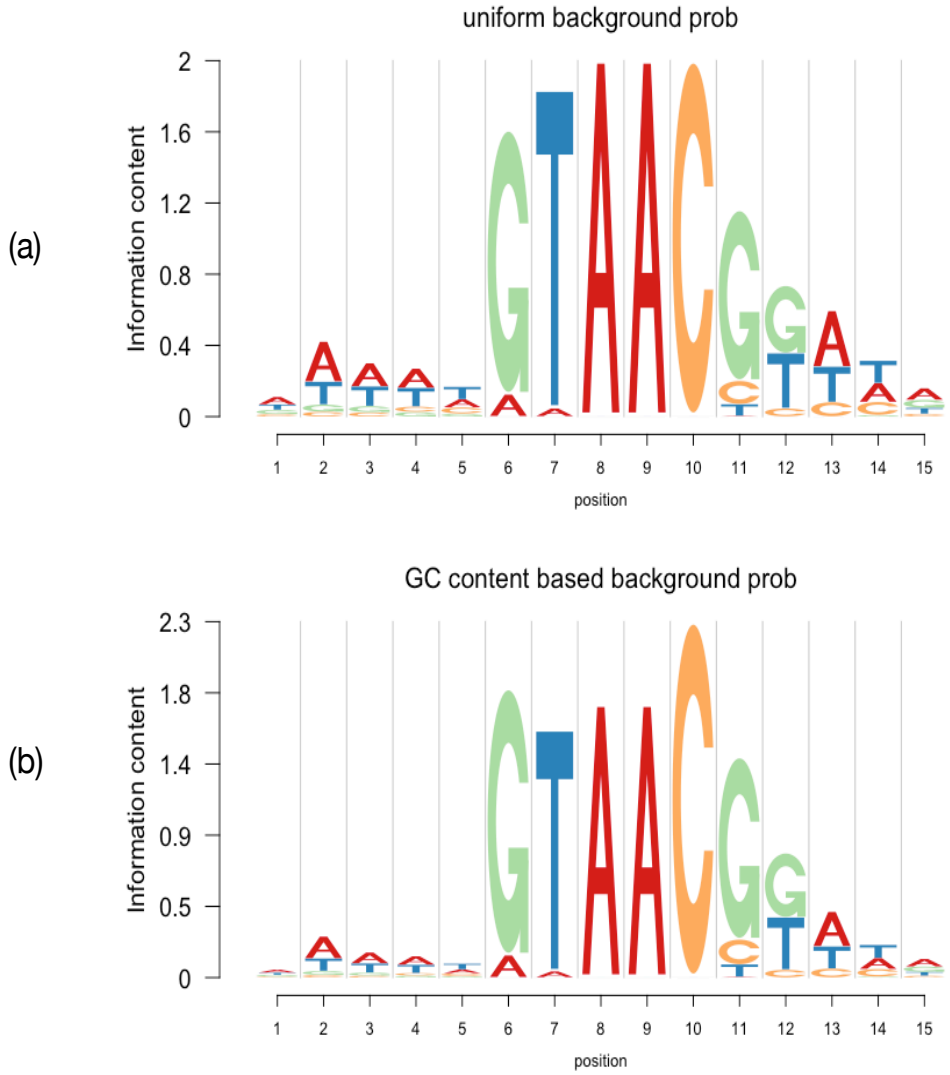

Fig 1. Standard Logo plot comparison under uniform and non-uniform background base probabilities . Logo plot representation of the plant transcription factor Achn021211 (MYB family protein) in *Actinidia chinensis*. The background probability for this species based on GC content is $q = (q_A, q_C, q_G, q_T) = (0.3141, 0.1859, 0.1859, 0.3141)$. The PWM matrix is obtained from PlantTFDB site (`http://planttfdb.cbi.pku.edu.cn/tf.php?sp=Ach&did=Achn021211`). In panel (a), we present the standard logo plot of the PWM matrix with uniform background for all 4 bases. In panel (b), we present the standard logo plot with the above specified background probability.

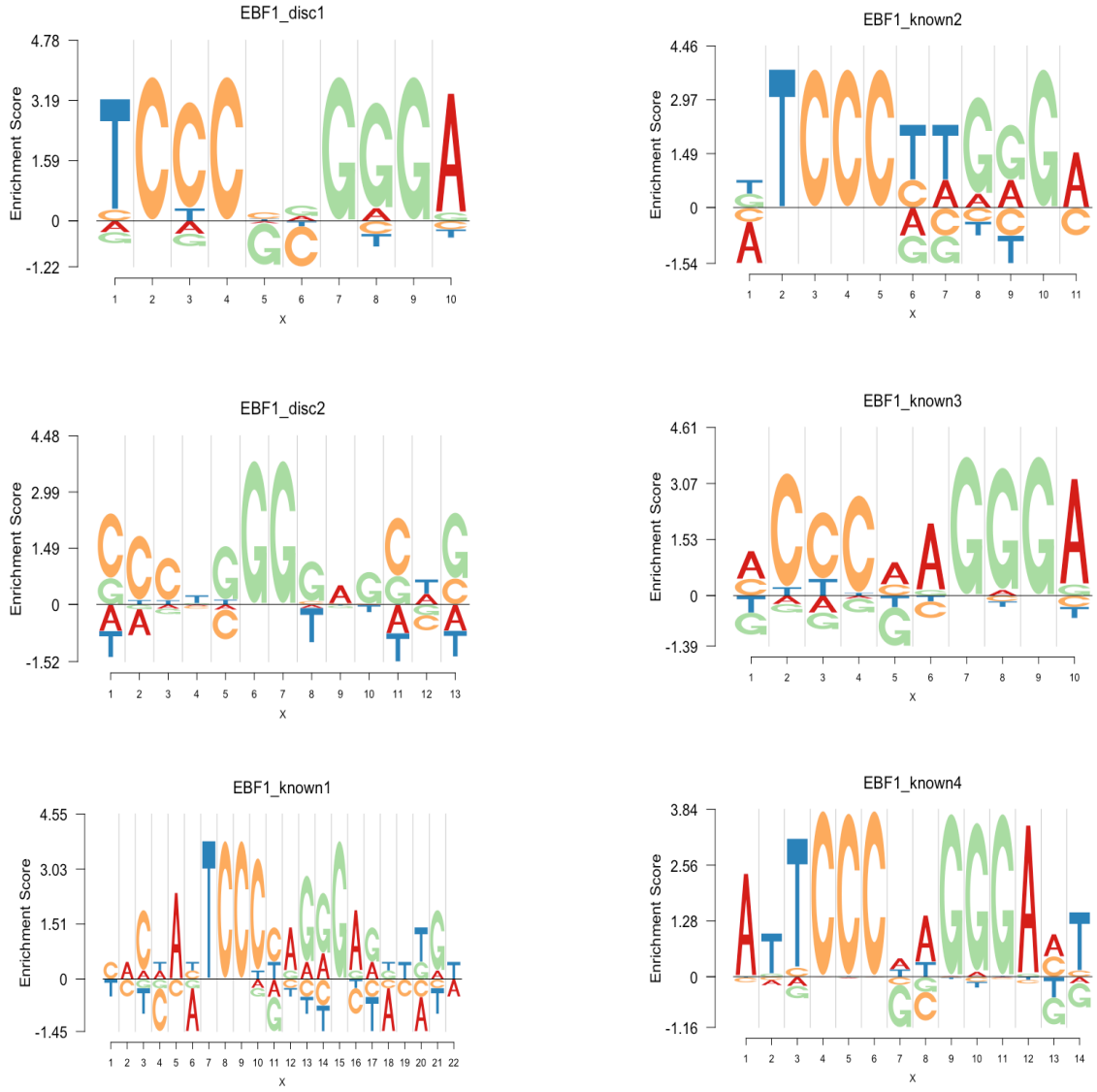

Fig 2. Sparse logo representation of the members of the EBF1 family of transcription factors: We present the sparse logo representation for the binding sites of 6 transcription factors in the EBF1 family. EBF1-known4 and EBF1-disc1, and also to some extent EBF1-known3 seem to show the depletion of G 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/`

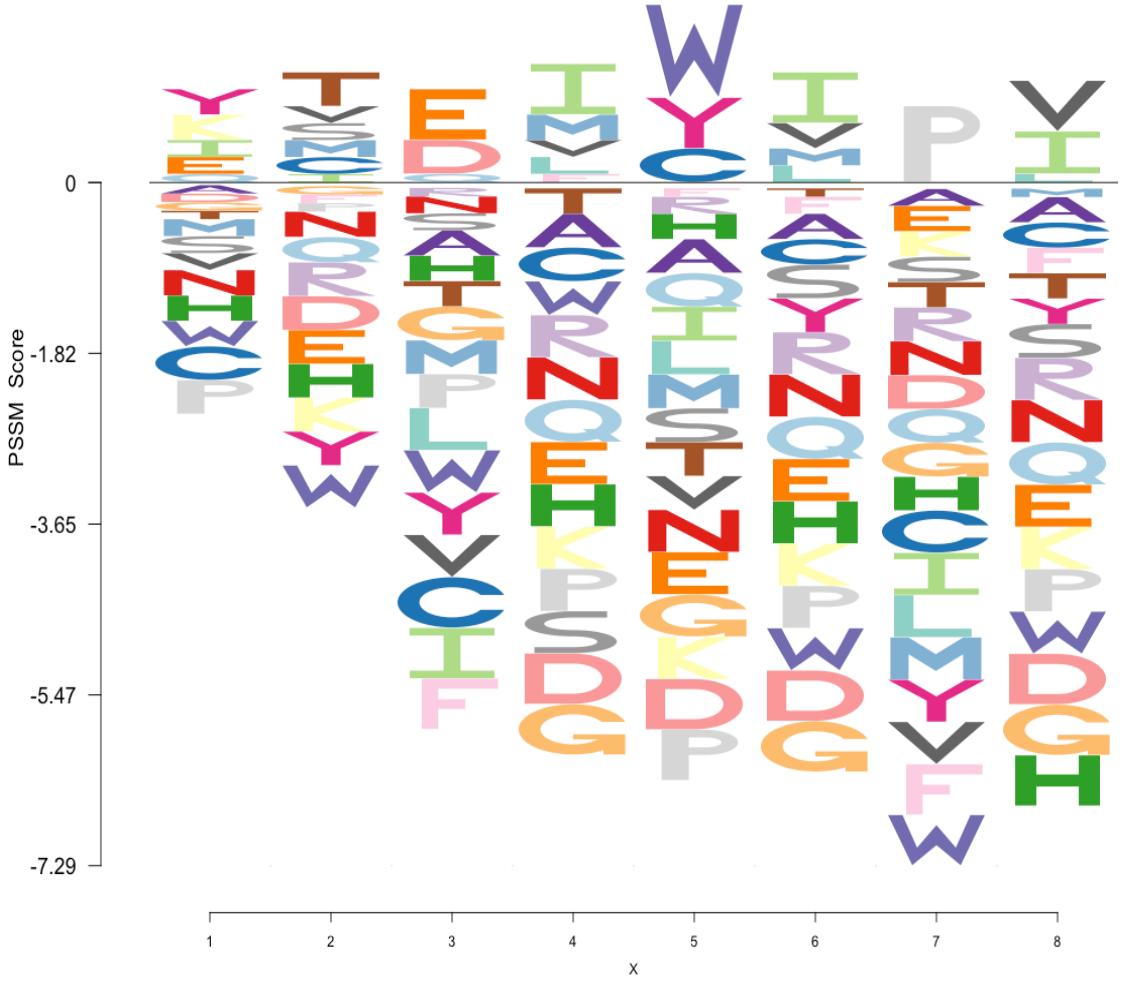

Fig 3. PSSM logo plot for protein sequence motif: The logo representation of the position specific scoring matrix (PSSM) for the Bacterial transcription activator, effector binding domain protein PF06445 (motif 4, Start=153 Length=8). The data is fetched from the 3PFDB website `http://caps.ncbs.res.in/cgi-bin/mini/databases/3pfdb/get_entry.cgi?id=PF06445`

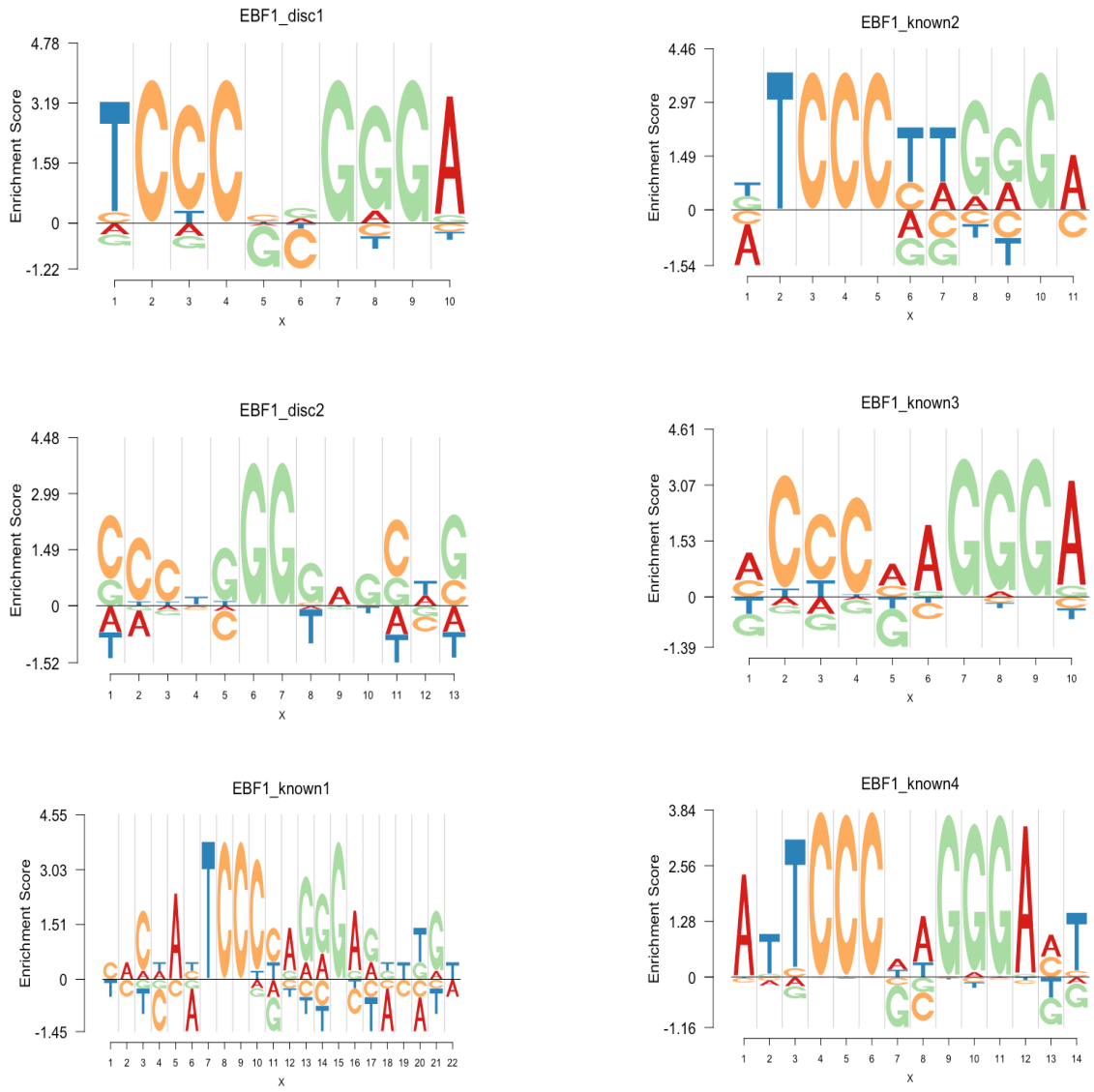

Fig 4. Various approaches of sparse logo representations for a transcription factor : The sparse logo representation under various stack height and stack composition methods - *log*, *log-odds*, *ratio*, *ic-log*, *ic-log-odds* and *ic-ratio* for the Early B cell factor 1 disc 1 (EBF1-disc1) transcription factor. The data is fetched from the CompBio website of MIT `http://compbio.mit.edu/encode-motifs/`

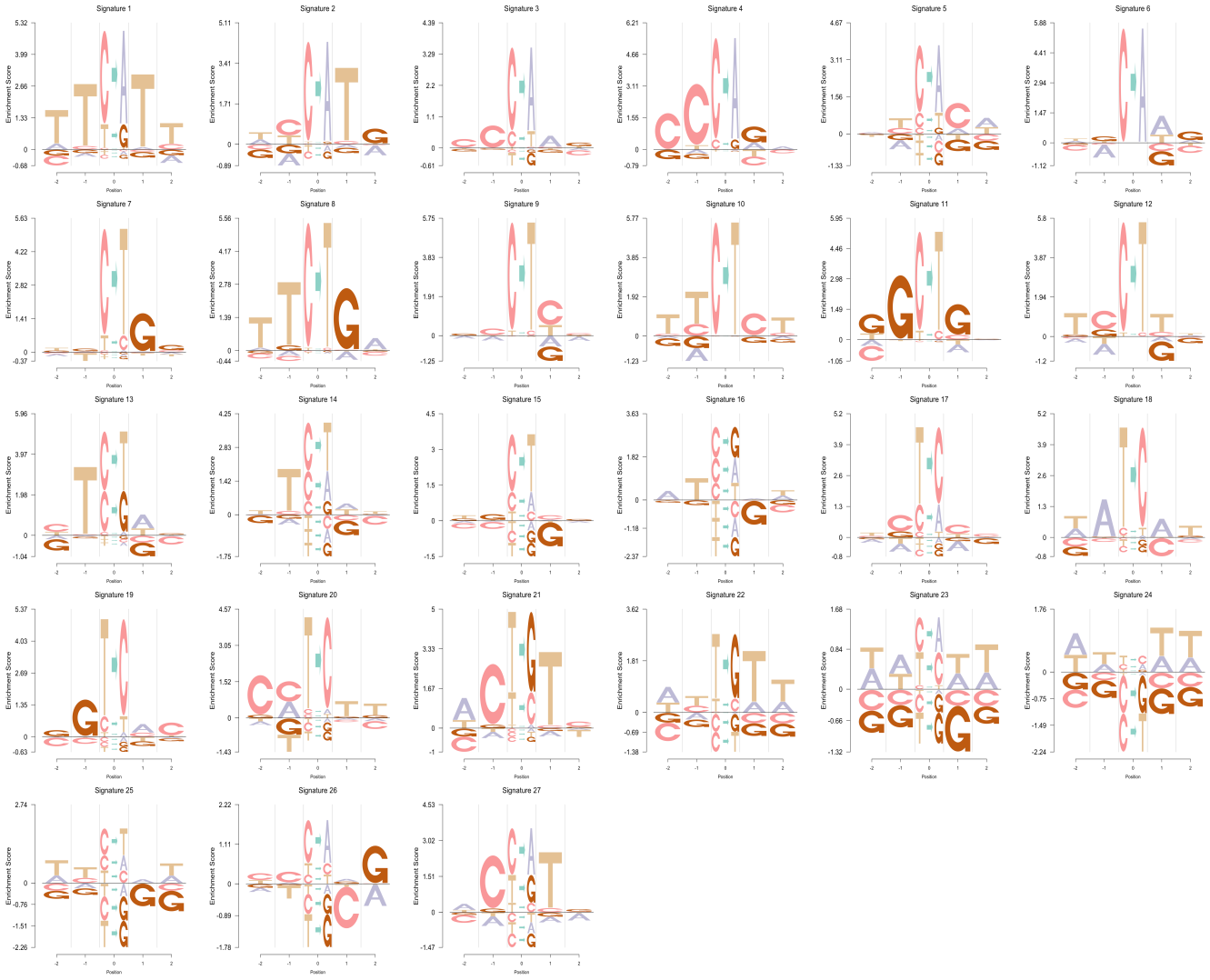

Fig 5. Logolas plots for the mutational signature profiles for 27 clusters in Shiraishi et al (2015): We present the sparse logo representations (ratio) method for the 27 cluster signature profiles obtained from fitting a grade of membership model on the cancer mutational signature data across 30 cancer types by Shiraishi et al (2015) [**?**]. This plot is an alternative logo plot based representation of Figure 4 in Shiraishi et al (2015) [**?**]

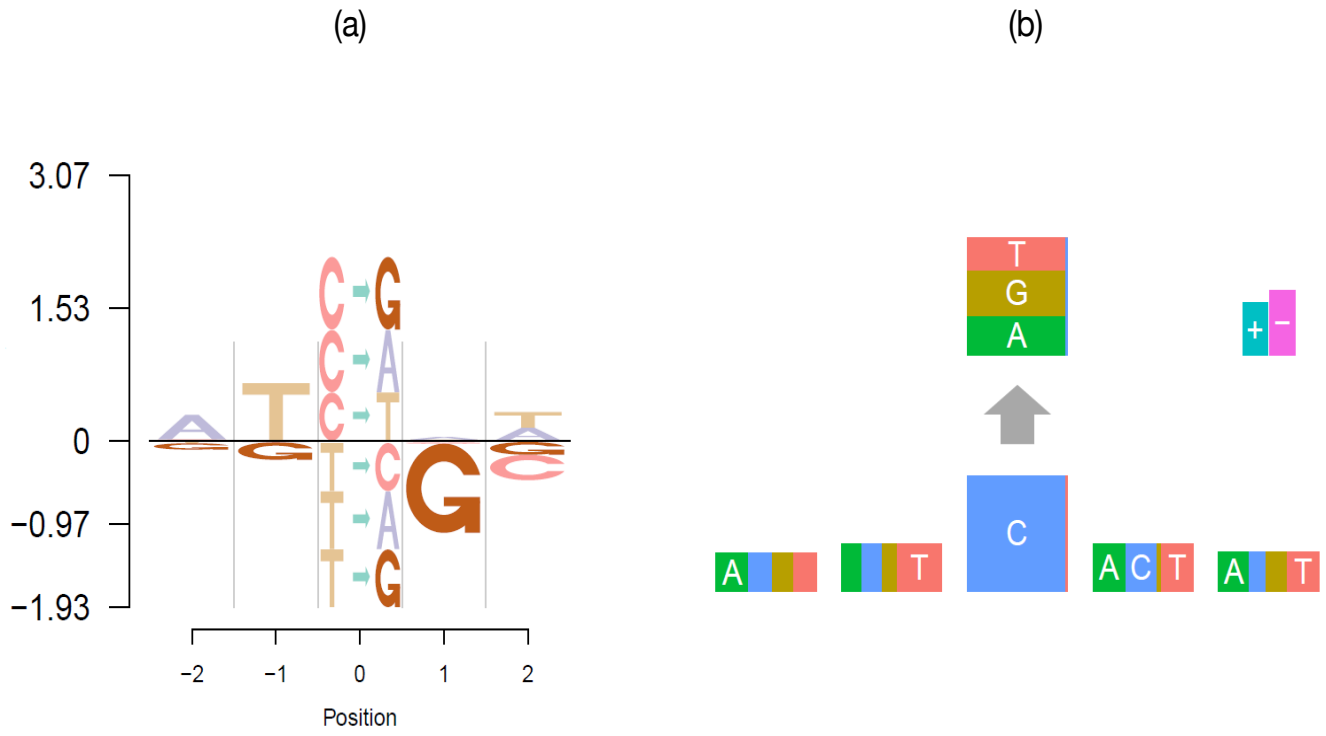

Fig 6. Comparison of Logolas sparse logo plot with pmsignature representation for cancer mutation signatures: We compare the sparse logo plot representation and the pmsignature representation due to Shiraishi et al (2015) [**?**] for mutation signature profile of cluster 16 in their paper. The position 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, the logo plot representation shows the depletion of G at the right flanking base more clearly than the pmsignature plot. Also, overall, the logo plot representation is more interpretable and visually appealing in highlighting the mutation signature patterns compared to the pmsignature plot.

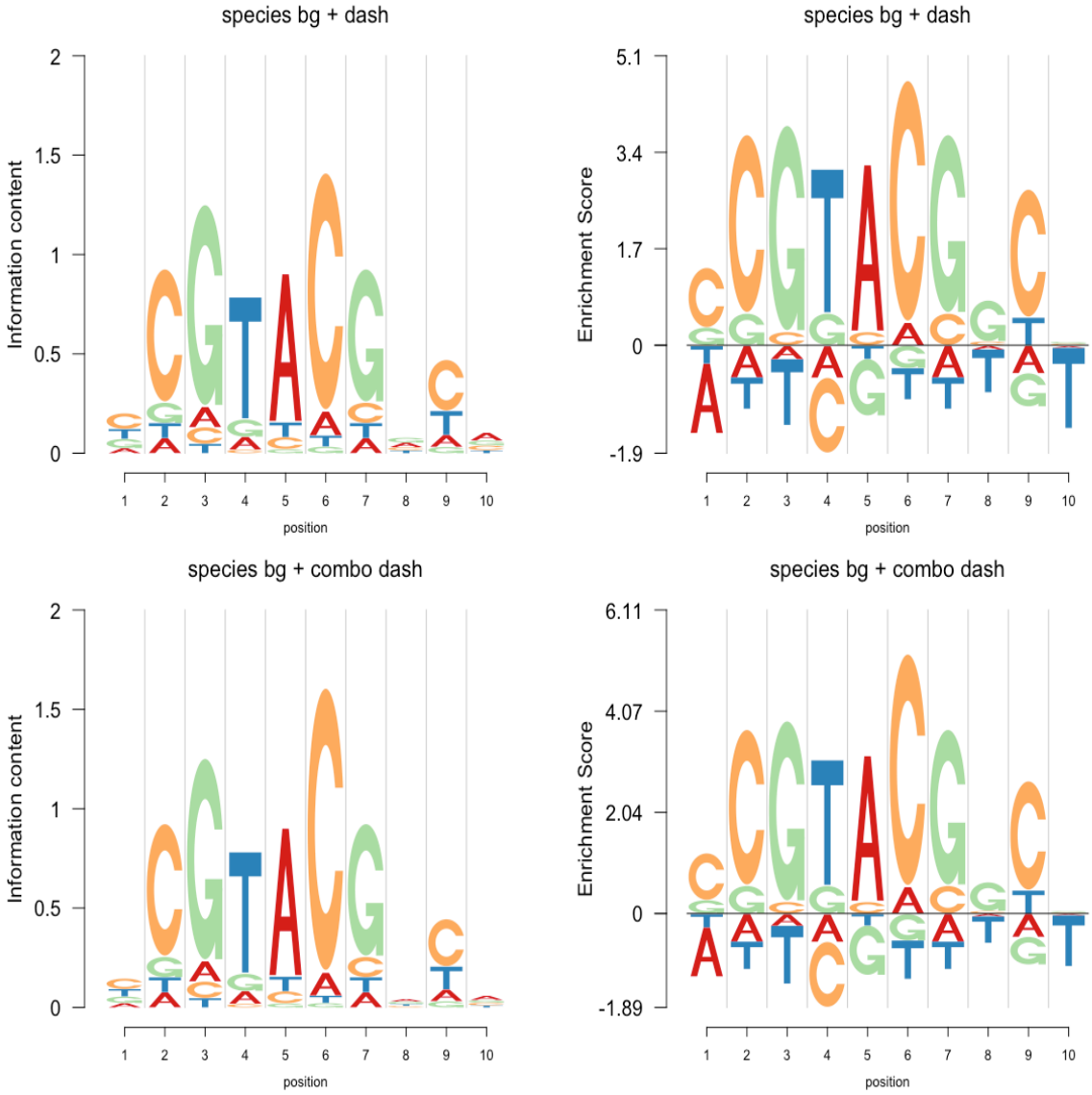

Fig 7. Dirichlet Adaptive Shrinkage (dash) training on combined transcription factor data for a species: We compare the two versions of Dirichlet Adaptive Shrinkage (dash) applied to the SBP protein transcription factor Achn185791 in *Actinidia chinensis*. In one case, the parameters of the *dash* model are learnt from the positional frequency data from Achn185791, while in the other case, the the parameters are learnt from the pooled positional frequency data across all 290 transcription factors of *Actinidia chinensis*, which we refer to as *combo dash* in this plot. The background probability of the bases for this species are $q = (q_A, q_C, q_G, q_T) = (0.3141, 0.1859, 0.1859, 0.3141)$. The transcription factor data for *Actinidia chinensis* along with the background probability information are derived from the PlantTFDB v4.0 database `http://planttfdb.cbi.pku.edu.cn/index.php?sp=Ach`.

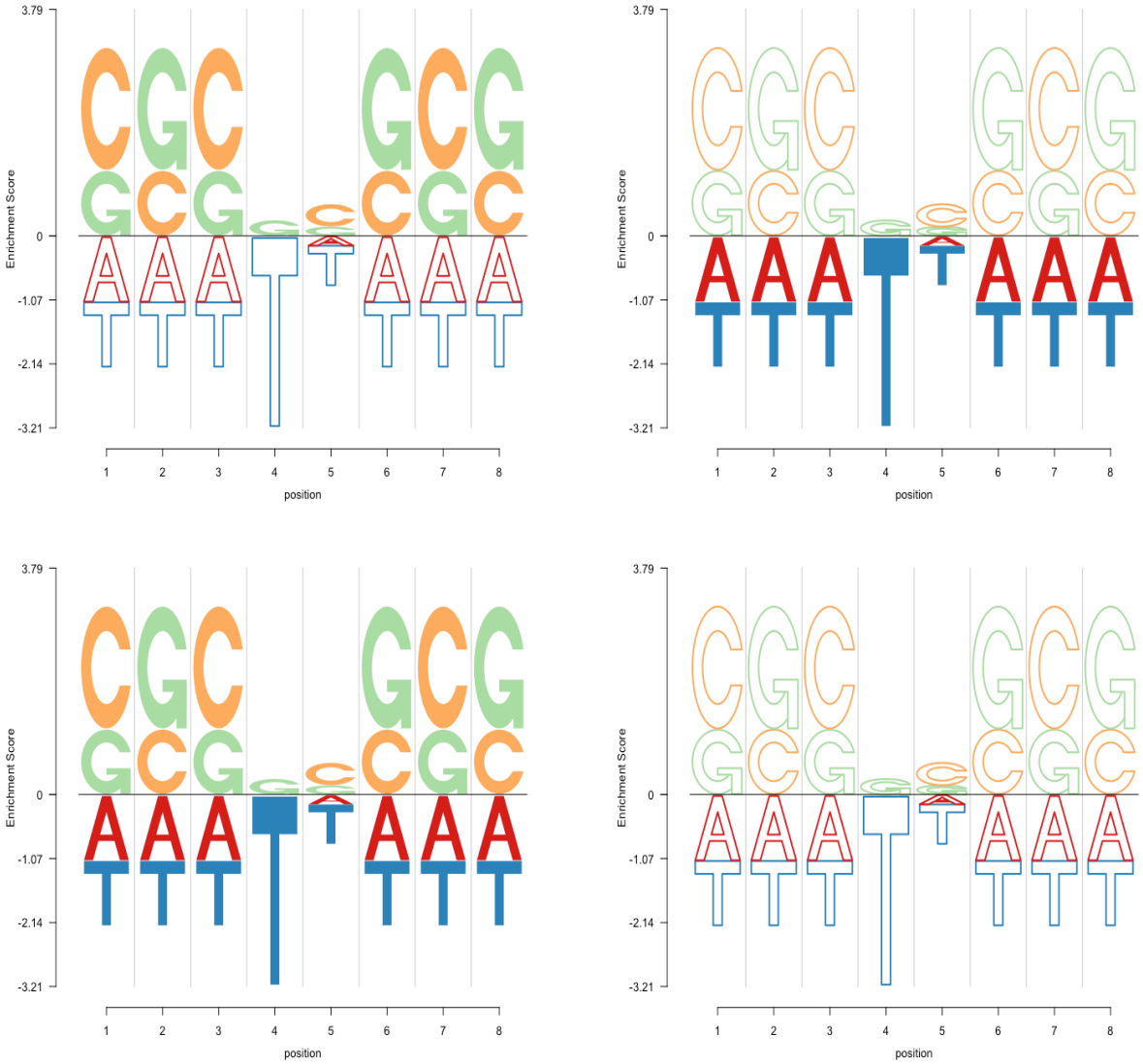

Fig 8. Fill and border styles in Logolas.: A demonstration of how fill and border styles can be used to distinguish between the enrichment and depletion of symbols at a position in a sparse logo plot.

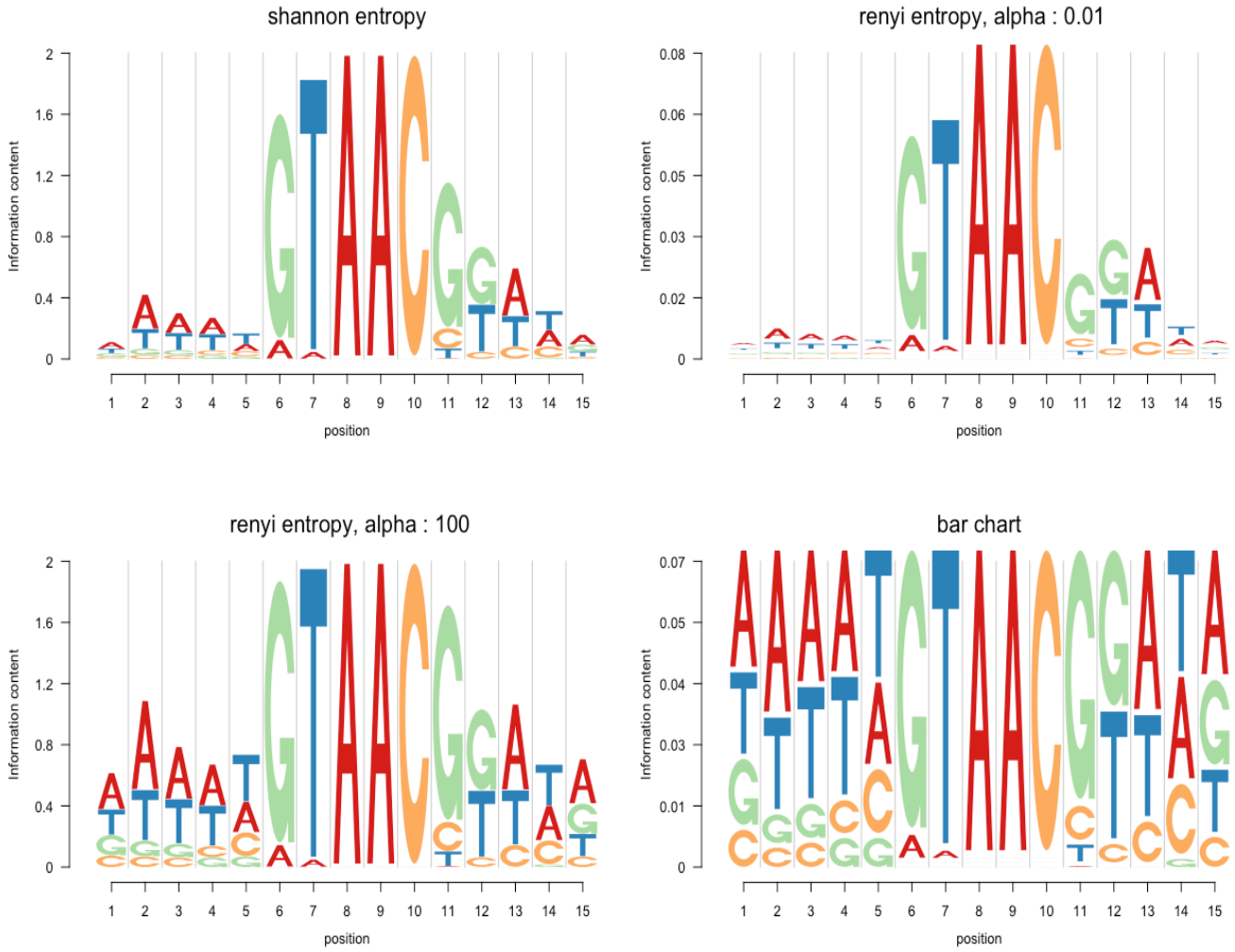

Fig 9. Stack heights in Logolas. : A demonstration of how stack height for a position in a standard logo plot can be determined in various ways in *Logolas*. In panel (a), the standard Shannon entropy based Information content is used to determine the height of the stack of symbols at each position. For panels (b) and (c), we use Renyi entropy based information content for two levels of tuning parameter α, one when $\alpha = 0.1$ is small and the other when $\alpha = 100$ is large. In panel (d), a relative frequency based stacked bar chart representation using logos is implemented. All these options can be passed as input arguments and control arguments to the *logomaker* functon in *Logolas*.