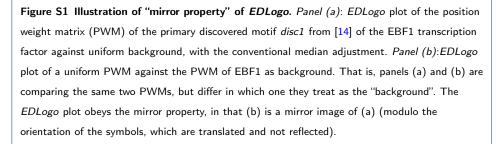


421 Supplementary Figures



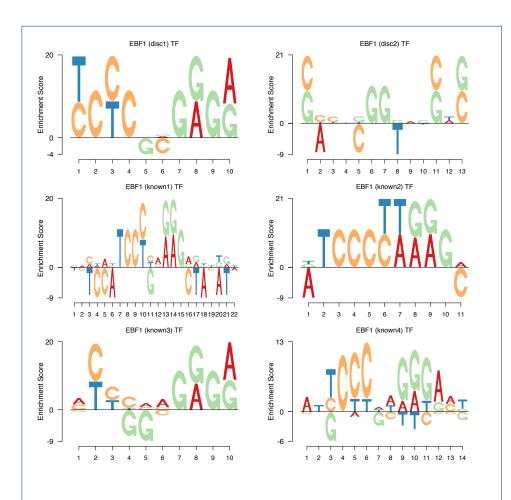
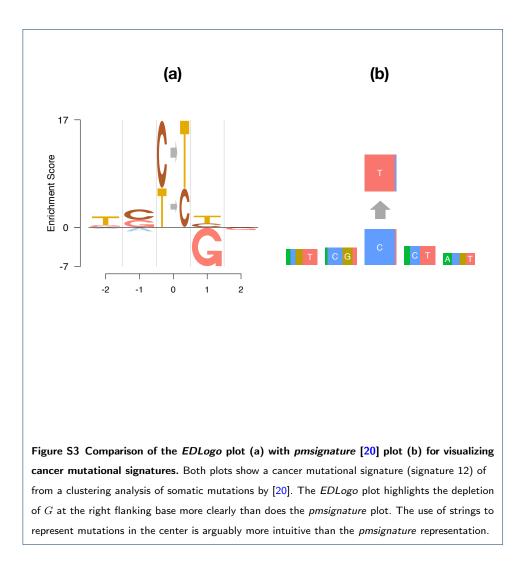


Figure S2 *EDLogo* plots for six different motifs of the EBF1 transcription factor. The PWMS for *known1* and *known2* come from the TRANSFAC database [17]; *known3* from the JASPAR database [16]; *known4* from [18]; *disc1* and *disc2* were discovered by the ENCODE project [14]. Three of the motifs (*known3, known4* and *disc1*) show depletion of G and C in the middle of the binding site.



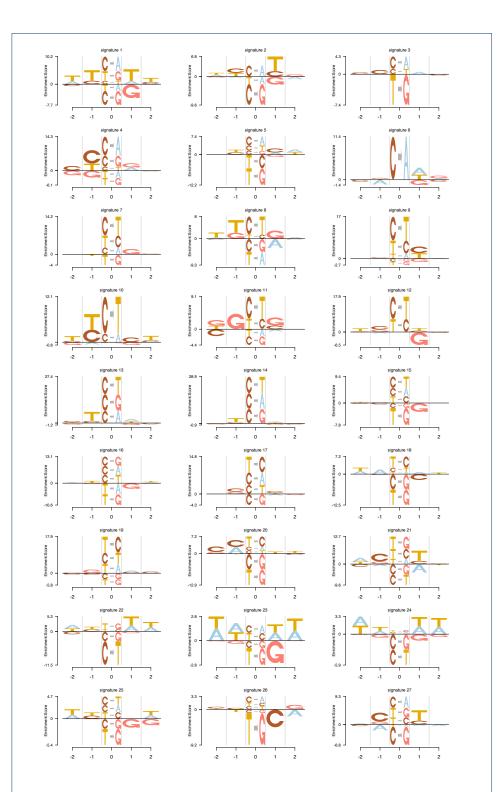
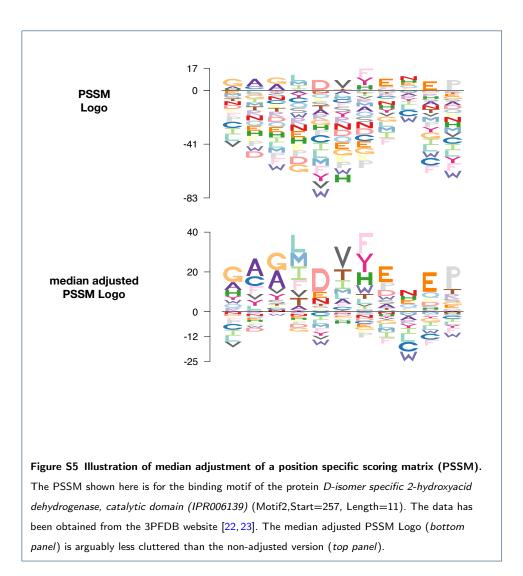


Figure S4 Illustration of *EDLogo* for all mutation signatures from Shiraishi et al: *EDLogo* plots for the 27 mutation signature profiles estimated by [20] using data from different cancer types. The heights of the strings in the center of each plot $(C \rightarrow G, C \rightarrow T, \text{ etc}$ at position 0 on x axis) reflect the relative frequency of each substitution type among somatic mutations contributing to the signature profile, while the heights of the bases at flanking positions on either side reflect the relative frequency of each base at these flanking positions.



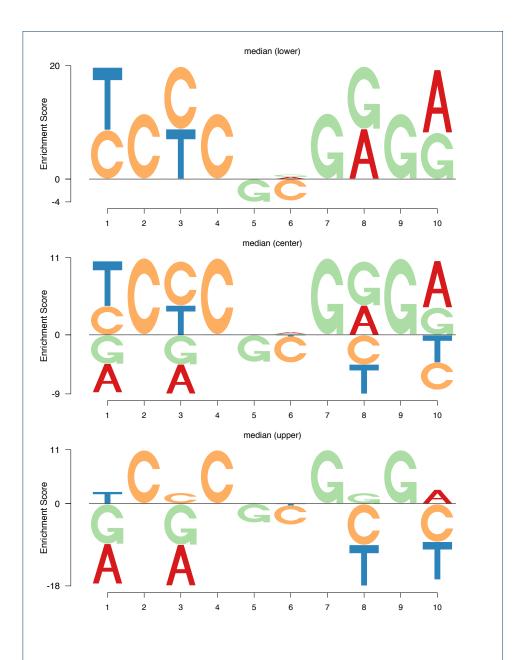


Figure S6 Choice of median. An illustration of how the choice of median value used for centering the \tilde{r}_i when the median is an interval (for an even number of characters/classes) can change the *EDLogo* representation of the EBF1-disc1 transcription factor binding site example from Figure 2 (panel a). In general, choosing the smallest median value favors enrichment of symbols (*top*), whereas choosing the largest median value favors depletion (*bottom*) and choosing the mid-point of the interval treads a common ground between enrichment and depletion (*middle*). As default option in our software and for all the *EDLogo* plots in this paper, we use the smallest median centering.