

Distance distribution of randomly selected 5 mers (pages 2-15). Each graph is a histogram that represents the number of occurrences of each motif (every 200 nt stretch) in the first 50 kb upstream region of protein-coding genes. In stark contrast to the top-predicted motifs, the control motifs do not have a strong preference to occur in the neighborhood of the transcription start sites of protein-coding genes.





























