



Figure S2: Biological TF binding sites and random sequences differ in μ^+ and μ^- . In each panel, the horizontal axis shows the fraction p_B of sequence space S ($L = 8$) filled with mouse TF binding sites ('biological sequences') or with the same proportion p_B of randomly and uniformly distributed eight-mers in which all nucleotides are equiprobable ('random sequences'). I refer to such a set of random sequences as r_B . **A)** the vertical axis shows the likelihood μ^- that a single random nucleotide change in a TF binding site destroys this binding site (circles), or that it transforms a random sequence in the set r_B into a sequence outside this set (squares). The analytical approximation is given by $1 - p_B$ (File S2). **B)** the vertical axis shows the likelihood μ^+ that a single random nucleotide change in a sequence from sequence space creates a TF binding site (circles), or that it transforms a sequence outside the set r_B into a sequence inside this set (squares). The analytical approximation is given by p_B (File S2). **C)** the vertical axis shows the fraction of all sequences in sequence space that are in the 1-mutant neighborhood, i.e., only one nucleotide change away, from a biological TF binding site (circles) or from a random sequence in the set r_B . Note that even for $p_B = 0.05$ almost one half of all sequences are in the neighborhood of biological TF binding sites, and this proportion is even higher for random sequences. The analysis is based on three and ten replicate exhaustive analyses of all sequences in sequence space for biological and random sequences, respectively, and on one random nucleotide mutation per sequence in each replicate. Error bars correspond to one standard error of the mean over the replicates.