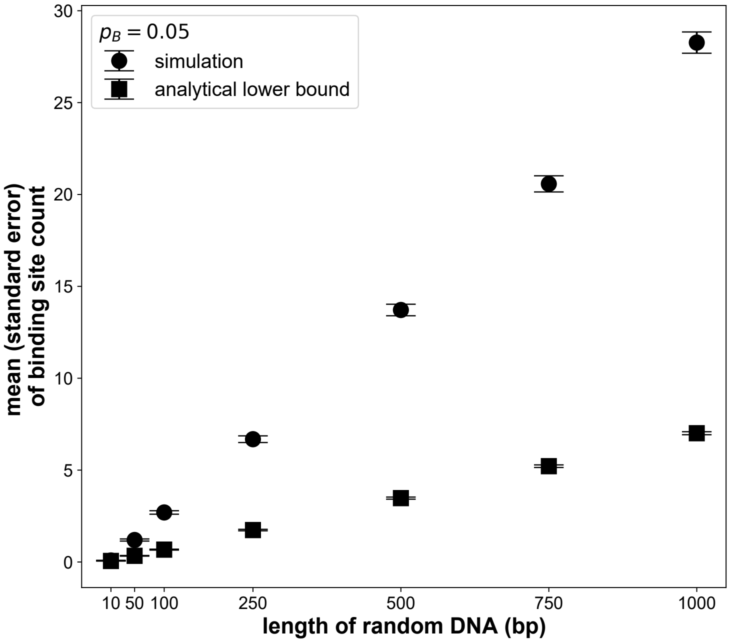

Figure S8: Even at low p_B, a short stretch of random DNA can contain multiple TF binding sites. The horizontal axis shows the length of a stretch of random DNA in which all four nucleotides are equiprobably and independently distributed. The vertical axis shows the number of TF binding sites of length $L = 8$ bp that are found in such a stretch of DNA. Simulation data is based on mouse TF binding sites that fill a fraction $p_B \approx 0.05$ of sequence space. I obtained such a set of binding sites as described in Methods, by creating the union of multiple sets of binding sites for different, randomly chosen mouse TFs until the size of this union, expressed as a fraction of the sequence space size 4^L, exceeded the threshold of p_B. (The actual value for the data shown in the plot is $p_B = 0.056$, because the size of this union is not exactly equal to 0.05.) I then created $n = 1000$ random DNA strings of a given length L_{prom}. For each such string and for each binding site in the set B of these mouse binding sites, I counted the number of times the string contained the binding site (not allowing binding sites to overlap). I then added up these counts for all binding sites. The figure ('simulation') shows the mean counts over all n random DNA strings of a given length, as well as the standard error of the mean. The analytical lower bound is obtained by a simple binomial approximation, based on the number n_b of DNA fragments of length L into which a larger string of length L_{prom} can be partitioned, i.e., $n_b = \lfloor L_{prom}/L \rfloor$, where $\lfloor x \rfloor$ indicates the largest integer that is smaller than or equal to x. Specifically, the number of binding sites under this approximation is binomially distributed with mean $n_B p_B$ and variance $n_B p_B(1-p_B)$. The figure ('analytical') shows the mean and the standard deviation of this binomial distribution. The actual binding site counts are substantially greater than this estimate, because many individual binding sites will span the boundaries of any such partition.