## Table S13: Correspondence between unsupervised promoter classification and sigma factor binding sites from DBTBS



(a) Motif cluster identified by the unsupervised sequence clustering algorithm (Table S4). Each up-shift is associated with a cluster according to the maximum posterior probability rule. M21 corresponds to an absence of a motif, i.e. sequence modeled as emitted by the background model. (b) up-shifts classified as belonging to this cluster that do not correspond to a sigma factor binding site reported in DBTBS (using a 22bp distance cut-off). (c) up-shifts classified as belonging to this cluster that correspond to a sigma factor binding site in DBTBS (closest DBTBS site at a distance ≤22bp).