

## Supplementary File 1

766 features have been used in LogitBoost for feature selection.

Sequence Features	
k-mer frequency	1 1-mer, 10 2-mers, 32 3-mers, 136 4-mers, 512 5-mers Others: TATAAACGCC , GCGCGC, TATAAA/TTTATA (Peckham, et al., 2007)
Poly Attribute (ACGT)	The longest number of A/C/G/T-tracts in the sequence
Poly (dA/dT) Tracts	Similar as k-mer, but only A...A, or T...T. from length 1 to length 20
Repeat Sequence	TGGA, CTG(Peckham, et al., 2007, Lee, et al., 2007)
Motif	ABF1, REB2, STB2 (yeast only)
Structural Features	
Lee, et al., 2007	Persistence Length 1, Persistence Length 2, Enthalpy, Free Energy, Major Groove Mobility, Major Groove Size, Minor Groove Mobility, Minor Groove Size, Nucleosome Probability, Propeller Twist, Rise, Roll, Slide, Tilt, Tip, Twist, Wedge
Abeel, et al., 2008	A-philicity, B-DNA Twist, Disruption Energy Propeller Twist in EP3, Z-DNA Free Energy, CpG Islands