**Supplementary Figures**



Figure S1. Enrichments of spurious binding sites in promoters of highly expressed, lowly expressed and unexpressed genes in human cell lines (A), worm (B) and fly (C). The number of spurious sites is normalized by the length of promoter regions. The promoter region of a gene is the 2kb up and down streams of the transcription start sites.



Figure S2. Comparing binding sites detected from IP experiments using DNA input and mock IP controls. Mock IP controls tend to detect fewer sites than DNA input controls. And the reduction of binding sites is correlated with the abundance of spurious sites estimated from mock IP with DNA input control (A). As expected, the binding sites overlapped with spurious sites tend to be removed by mock IP, compared to DNA input (B).



Figure S3. Motif enrichments in binding peaks from fly (A) and worm (B). The binding peaks are detected and ranked by four methods: (1) using probability as the scoring metric and DNA input as control; (2) using probability as the scoring metric and mock IP as control; (3) using SPP signal and mock IP as control; (4) using probability from the Bayesian model with both DNA input and