

Table S6. Average information content from different data sources

<i>Data Set</i>	<i>Organism</i>	<i>Mean (\pm Standard Deviation) Information Content (I)</i>	<i>Number of Motifs</i>
RegTransBase [16]**	Bacteria (531 species)	25.1 (\pm 5)	94
Regulon DB [5]	<i>E. coli</i>	20.0 (\pm 9.2)	72
DP Interact [20]*	<i>E. coli</i>	24.3 (\pm 11.6)	68
Average	Bacteria	23.3	
Zhu, et al. [18]	Yeast	13.8 (\pm 2.6)	89
Harbison, et al. [21]*	Yeast	14.1 (\pm 5.7)	102
MacIsaac, et al. [22]*	Yeast	12.5 (\pm 5.1)	124
Average	Yeast	13.4	
JASPER (core) [19]	Multicellular eukaryotes	12.1 (\pm 4.5)	123
FlyReg [23]*	Fly	12.5 (\pm 5.5)	75
Bergman, et al.*	Fly	13.4 (\pm 4.4)	62
Badis, et al. [24]	Mouse (primary motifs)	12.3 (\pm 2.0)	104
Average	Multicellular eukaryotes	12.8	

*These data sets were accessed from STAMP [25]

**Motifs constructed from fewer than 8 sample binding sites were eliminated.

5. Gama-Castro, S., *et al.* (2008) RegulonDB (version 6.0): gene regulation model of *Escherichia coli* K-12 beyond transcription, active (experimental) annotated promoters and Textpresso navigation. *Nucleic acids research* 36, D120-124

16. Kazakov, A.E., *et al.* (2007) RegTransBase--a database of regulatory sequences and interactions in a wide range of prokaryotic genomes. *Nucleic acids research* 35, D407-412

18. Zhu, C., *et al.* (2009) High-resolution DNA-binding specificity analysis of yeast transcription factors. *Genome Res* 19, 556-566

19. Vlieghe, D., *et al.* (2006) A new generation of JASPAR, the open-access repository for transcription factor binding site profiles. *Nucleic acids research* 34, D95-97

20. Robison, K., *et al.* (1998) A comprehensive library of DNA-binding site matrices for 55 proteins applied to the complete Escherichia coli K-12 genome. *J Mol Biol* 284, 241-254
21. Harbison, C.T., *et al.* (2004) Transcriptional regulatory code of a eukaryotic genome. *Nature* 431, 99-104
22. MacIsaac, K.D., *et al.* (2006) An improved map of conserved regulatory sites for Saccharomyces cerevisiae. *BMC bioinformatics* 7, 113
23. Bergman, C.M., *et al.* (2005) Drosophila DNase I footprint database: a systematic genome annotation of transcription factor binding sites in the fruitfly, Drosophila melanogaster. *Bioinformatics* 21, 1747-1749
24. Badis, G., *et al.* (accepted) Diversity and complexity in DNA recognition by transcription factors. *Science*
25. Mahony, S., and Benos, P.V. (2007) STAMP: a web tool for exploring DNA-binding motif similarities. *Nucleic Acids Res* 35, W253-258