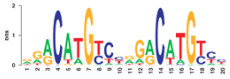
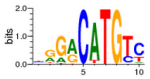
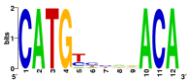
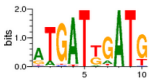
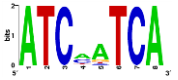
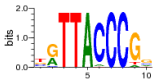
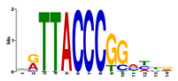
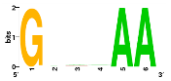
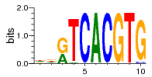
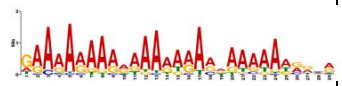
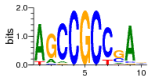
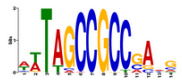
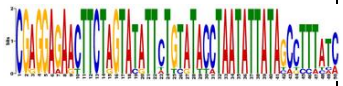
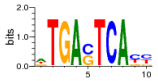
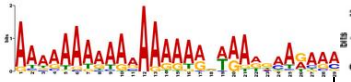
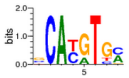
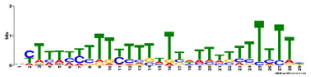
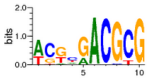
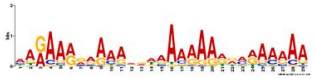
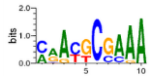
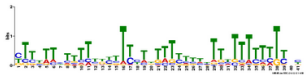
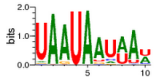
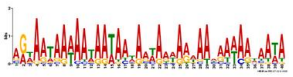
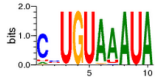
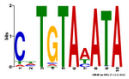
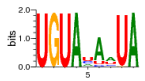
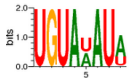
Supplementary Table 1 – Comparison between DRIMUST and three other tools on various DNA and RNA examples

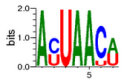
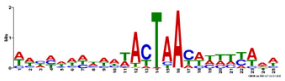
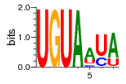
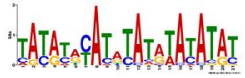
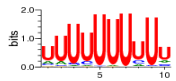
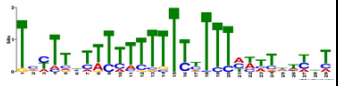
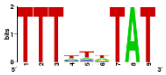
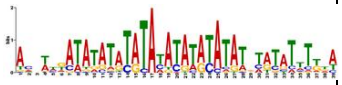
To evaluate the performance of DRIMust in comparison to other state-of-the-art methods we ran DRIMust on 24 examples generated from high throughput binding experiments on 10 transcription factors and 14 RNA-binding proteins and compared the results to those obtained by using four other methods: the standard MEME program, the DREME program, XXmotif and SCOPE. Almost all the input examples comprised ranked lists, except for p53 which comprised target and background sets. Since MEME, DREME, XXmotif and SCOPE expect a target set as input, we converted the ranked lists into target sets by taking the top 100 sequences in the case of MEME (restricted by MEME's limitation of 60,000 characters) and the top 20% sequences for the other tools.

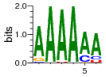
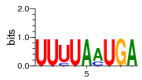
	The protein and its consensus binding motif	DRIMust results	MEME results	DREME results	XXmotif results	SCOPE results
DNA	**p53**  Consensus motif and data were taken from (1)	P-value ≤ 2.22e-266  3092 sequences (half targets and half background); 1,267,720 characters; 0.76 minutes	1.8e-100 100 sequences; 41,000 characters; 28.26 minutes	4.9e-133 1546 sequences; 633,852 characters; 35.66 minutes	1e-490 1546 sequences; 633,852 characters; 46 minutes	1e-321 1e-320 1e-318  1546 sequences; 633,852 characters; 9 hours
	hoxa2 Consensus motif and data were taken from (2)	P-value ≤ 1.70e-80  8245 sequences; 1,649,000 characters; 1.07 minutes	1.6e-81 100 sequences; 20,000 characters; 7.61 minutes	3.0e-120 1649 sequences; 329,800 characters; 13.83 minutes	6.68e-167 1649 sequences; 329,800 characters 13 minutes	1e-377 1e-341 1e-331  1649 sequences; 329,800 characters; 6.1 hours

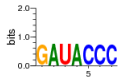
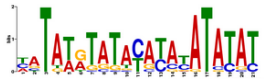
	REB1 TTACCCG Consensus motif was taken from (3) and data from (4)	P-value ≤ 2.47e-131  4000 sequences; 2,000,000 characters; 1.26 minutes	7.0e-113  100 sequences; 50,000 characters; 39.77 minutes	1.1e-80 800 sequences; 400,000 characters; 15.58 minutes	3.15e-89 800 sequences; 400,000 characters; 14 minutes	1e-321  1e-280 (fifth best) 800 sequences; 400,000 characters; 5.9 hours
	CBF1 tCACGTG Consensus motif was taken from (3) and data from (4)	P-value ≤ 1.03e-89  4000 sequences; 2,000,000 characters; 1.2 minutes	2.3e-100  9.1e-070 (second best) 100 sequences; 50,000 characters; 43.3 minutes	1.1e-39 2.0e-023 (forth best) 800 sequences; 400,000 characters; 29.48 minutes	5.00e-15 2.72e-10 (third best) 800 sequences; 400,000 characters; 11 minutes	1e-321 800 sequences; 400,000 characters; 5.7 hours
	UME6 taGCCGCCsa Consensus motif was taken from (3) and data from (4)	P-value ≤ 2.18e-102  4000 sequences; 2,000,000 characters; 1.25 minutes	2.0e-114  100 sequences; 50,000 characters; 41 minutes	3.1e-45 5.1e-030 (third best) 800 sequences; 400,000 characters; 23.16 minutes	3.52e-47 800 sequences; 400,000 characters; 13 minutes	1e-321 800 sequences; 400,000 characters; 5.7 hours
	TYE7 tCACGTGa Consensus motif was taken from (3) and data from (4)	P-value ≤ 3.31e-62 4000 sequences; 2,000,000 characters; 1.2 minutes	7.9e-115  100 sequences; 50,000 characters; 42.5 minutes	7.3e-43 1.7e-018 (fourth best) 800 sequences; 400,000 characters; 29 minutes	9.26e-21 1.11e-10 (third best) 800 sequences; 400,000 characters; 22 minutes	1e-321 800 sequences; 400,000 characters; 5.7 hours

	GCN4 **TGAsTCa** Consensus motif was taken from (3) and data from (4)	P-value ≤ 2.37e-42  4000 sequences; 2,000,000 characters; 1.09 minutes	1.3e-85  100 sequences; 50,000 characters; 43.15 minutes	2.0e-32 2.5e-005 (ninth best) 800 sequences; 400,000 characters; 20.57 minutes	4.00e-17 800 sequences; 400,000 characters; 14 minutes	1e-321 800 sequences; 400,000 characters; 5.6 hours
	INO2 **CAcaTGc** Consensus motif was taken from (3) and data from (4)	P-value ≤ 1.14e-37  4000 sequences; 2,000,000 characters; 1.1 minutes	3.1e-88  100 sequences; 50,000 characters; 40.72 minutes	1.1e-42 2.2e-007 (eighth best) 800 sequences; 400,000 characters; 21.93 minutes	2.68e-17 800 sequences; 400,000 characters; 12 minutes	1e-321 800 sequences; 400,000 characters; 5.6 hours
	MBP1 **ACGCGt** Consensus motif was taken from (3) and data from (4)	P-value ≤ 1.17e-47  4000 sequences; 2,000,000 characters; 1.2 minutes	7.9e-97  100 sequences; 50,000 characters; 39.78 minutes	5.4e-47 7.6e-009 (sixth best) 800 sequences; 400,000 characters; 23.67 minutes	1.01e-23 1.31e-06 (forth best) 800 sequences; 400,000 characters; 20 minutes	1e-321 800 sequences; 400,000 characters; 5.6 hours
	SWI4 **CgCsAAA** Consensus motif was taken from (3) and data from (4)	P-value ≤ 2.97e-33  4000 sequences; 2,000,000 characters; 1.22 minutes	8.3e-120  100 sequences; 50,000 characters; 40 minutes	2.3e-47 800 sequences; 400,000 characters; 23.62 minutes	2.91e-20 800 sequences; 400,000 characters; 13 minutes	1e-321 1e-320 (third best) 800 sequences; 400,000 characters; 5.6 hours

RNA	**puf2** Consensus motif was taken from (5) and data from (6)	P-value ≤ 4.93e-51  4937 sequences; 636,234 characters; 0.51 minutes	3.2e-8  100 sequences; 16,991 characters; 2.92 minutes	7.8e-34 8.0e-006 (third best) 988 sequences; 139,903 characters; 1.9 minutes	8.94e-25 2.23e-16 (second best) 988 sequences; 139,903 characters; 9 minutes	6.3e-322 6.7e-166 988 sequences; 139,903 characters; 2 minutes
	puf3 Consensus motif and data were taken from (6)	P-value ≤ 9.44e-153  4970 sequences; 640,926 characters; 0.39 minutes	6.2e-151  100 sequences; 11,304 characters; 1.45 minutes	7.4e-75 1007 sequences; 132,972 characters; 1.6 minutes	9.05e-65 1007 sequences; 132,972 characters; 7 minutes	2.5e-321 8.7e-170 (third best) 1007 sequences; 132,972 characters; 2 minutes
	puf4 Consensus motif and data were taken from (6)	P-value ≤ 5.22e-123  4948 sequences; 638,135 characters 0.49 minutes	9.0e-50 100 sequences; 14,252 characters; 2.16 minutes	4.3e-51 4.4e-019 (second best) 1002 sequences; 129,950 characters; 1.58 minutes	8.90e-20 1002 sequences; 129,950 characters; 8 minutes	9.6e-322 6.7e-175 1002 sequences; 129,950 characters; 2 minutes
	PUM2 Consensus motif and data were taken from (7)	P-value ≤ 4.9e-324  9995 sequences; 999,500 characters; 1.35 minutes	2.0e-101 100 sequences; 10,000 characters; 1.25 minutes	2.3e-106 1999 sequences; 199,900 characters; 3.43 minutes	4.10e-148 1999 sequences; 199,900 characters; 20 minutes	5.5e-400 2e-398 1999 sequences; 199,900 characters; 3 minutes

	QKI Consensus motif and data were taken from (7)	P-value ≤ 4.9e-324  9997 sequences; 999,700 characters; 1.09 minutes	2.0e-19  100 sequences; 10,000 characters; 1.17 minutes	9.8e-51 2000 sequences; 200,000 characters; 3.03 minutes	2.55e-45 2000 sequences; 200,000 characters; 25 minutes	4.2e-400 2.5e-356 2000 sequences; 200,000 characters; 10.6 hours
	Puf5 Consensus motif and data were taken from (8)	P-value ≤ 3.47e-78  4795 sequences; 622,010 characters; 0.43 minutes	3.6e-9  3.1e-004 100 sequences; 15,234 characters; 2.52 minutes	6.8e-42 3.1e-012 959 sequences; 134,166 characters; 1.53 minutes	9.76e-21 1.61e-20 959 sequences; 134,166 characters; 10 minutes	3.8e-321 4.2e-150 959 sequences; 134,166 characters; 3 minutes
	Pub1 Consensus motif and data were taken from (8)	P-value ≤ 6.94e-177  4251 sequences; 548,258 characters; 1.03 minutes	7.5e-8  100 sequences; 17,020 characters; 2.88 minutes	2.1e-36 2.4e-016 851 sequences; 133,915 characters; 1.33 minutes	5.73e-18 2.48e-05 (4th best) 851 sequences; 133,915 characters; 9 minutes	1.9e-203 4.3e-146  851 sequences; 133,915 characters; 2 minutes
	Pab1 Consensus motif and data were taken from (8)	No motif was found 4142 sequences; 534,415 characters; 0.23 minutes	8.0e-30  100 sequences; 12,734 characters 1.96 minutes	5.4e-45 829 sequences; 99,365 characters; 1.33 minutes	5.89e-28 829 sequences; 99,365 characters; 17 minutes	7.8e-322 829 sequences; 99,365 characters; 8.4 hours

	Khd1 Consensus motif and data were taken from (8)	P-value ≤ 1.39e-16 4773 sequences; 619,178 characters; 0.35 minutes	2.1e-12 100 sequences; 16,243 characters; 2.83 minutes	1.2e-45 955 sequences; 137,956 characters; 1.28 minutes	1.24e-26 955 sequences; 137,956 characters; 9 minutes	1.6e-320 7.6e-147 955 sequences; 137,956 characters; 8.4 hours
	Nab2 Consensus motif and data were taken from (8)	P-value ≤ 4.80e-15  4101 sequences; 528,461 characters; 0.29 minutes	1.9e-13 100 sequences; 11,514 characters; 1.67 minutes	1.0e-39 2.2e-008 821 sequences; 111,334 characters; 1.58 minutes	8.16e-29 4.16e-9 821 sequences; 111,334 characters; 10 minutes	1.4e-140 821 sequences; 111,334 characters; 3 minutes
	Vts1 Consensus motif and data were taken from (8)	No motif was found 1787 sequences; 234,740 characters; 0.09 minutes	4.5e-5 100 sequences; 13,720 characters; 2.23 minutes	1.0e-14 358 sequences; 45,130 characters; 0.38 minutes	1.08e-8 358 sequences; 45,130 characters; 7 minutes	2.1e-75 358 sequences; 45,130 characters; 2 minutes
	Pin4 Consensus motif and data were taken from (8)	P-value ≤ 3.72e-9  4261 sequences; 549,860 characters; 0.25 minutes	1.3 100 sequences; 12,003 characters; 1.76 minutes	3.1e-51 853 sequences; 109,847 characters; 1.04 minutes	1.65e-28 853 sequences; 109,847 characters; 11 minutes	2.2e-167 853 sequences; 109,847 characters; 7.5 hours
	Nrd1 Consensus motif and data were taken from (8)	P-value ≤ 6.69e-13 3947 sequences; 507,022 characters; 0.24 minutes	6.6e-8 100 sequences; 14,906 characters; 2.48 minutes	4.1e-44 1.1e-004 (third best) 790 sequences; 104,759 characters; 1.28 minutes	4.15e-20 790 sequences; 104,759 characters; 10 minutes	2.2e-143 790 sequences; 104,759 characters; 7.5 hours

	YLL032C Consensus motif and data were taken from (8)	P-value ≤ 5.84e-7  2286 sequences; 295,474 characters; 0.11 minutes	2.3e-11  100 sequences; 14,249 characters; 2.38 minutes	3.4e-24 458 sequences; 58,789 characters; 0.48 minutes	9.54e-14 458 sequences; 58,789 characters; 8 minutes	4.6e-69 458 sequences; 58,789 characters; 7.4 hours

References

1. Smeenk, L., van Heeringen, S.J., Koeppel, M., van Driel, M.A., Bartels, S.J.J., Akkers, R.C., Denissov, S., Stunnenberg, H.G. and Lohrum, M. (2008) Characterization of genome-wide p53-binding sites upon stress response. *Nucleic Acids Research*, **36**, 3639-3654. doi: 10.1093/nar/gkn232 http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2441782/
2. Donaldson, I.J., Amin, S., Hensman, J.J., Kutejova, E., Rattray, M., Lawrence, N., Hayes, A., Ward, C.M. and Bobola, N. (2012) Genome-wide occupancy links Hoxa2 to Wnt–β-catenin signaling in mouse embryonic development. *Nucleic Acids Research*, **40**, 3990-4001. doi: 10.1093/nar/gkr1240 http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3351182/
3. Harbison, C.T., Gordon, D.B., Lee, T.I., Rinaldi, N.J., Macisaac, K.D., Danford, T.W., Hannett, N.M., Tagne, J.-B., Reynolds, D.B., Yoo, J. *et al.* (2004) Transcriptional regulatory code of a eukaryotic genome. *Nature*, **431**, Supplementary Table 2. doi: 10.1038/nature02800 http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3006441/
4. Harbison, C.T., Gordon, D.B., Lee, T.I., Rinaldi, N.J., Macisaac, K.D., Danford, T.W., Hannett, N.M., Tagne, J.-B., Reynolds, D.B., Yoo, J. *et al.* (2004) Transcriptional regulatory code of a eukaryotic genome. *Nature*, **431**, 99-104. doi: 10.1038/nature02800 http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3006441/
5. Yosefzon, Y., Koh, Y.Y., Chritton, J.J., Lande, A., Leibovich, L., Barziv, L., Petzold, C., Yakhini, Z., Mandel-Gutfreund, Y., Wickens, M. *et al.* (2011) Divergent RNA binding specificity of yeast Puf2p. *RNA*, **17**, 1479-1488. doi: 10.1261/rna.2700311 http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3153972/
6. Gerber, A.P., Herschlag, D. and Brown, P.O. (2004) Extensive Association of Functionally and Cytotopically Related mRNAs with Puf Family RNA-Binding Proteins in Yeast. *PLoS Biol*, **2**, e79. doi: 10.1371/journal.pbio.0020079 http://www.ncbi.nlm.nih.gov/pmc/articles/PMC368173/
7. Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano Jr, M., Jungkamp, A.-C., Munschauer, M. *et al.* (2010) Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP. *Cell*, **141**, 129-141. doi: 10.1016/j.cell.2010.03.009 http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2861495/
8. Hogan, D.J., Riordan, D.P., Gerber, A.P., Herschlag, D. and Brown, P.O. (2008) Diverse RNA-Binding Proteins Interact with Functionally Related Sets of RNAs, Suggesting an Extensive Regulatory System. *PLoS Biol*, **6**, e255. doi: 10.1371/journal.pbio.0060255 http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2573929/