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

REB1	P-value ≤ 2.47e-131	7.0e-113	1.1e-80	3.15e-89	1e-321
TTACCCG					<u>۲</u>
Consensus motif was			a,]Î ÎVÂÂÂÂ		î <mark>`</mark> ` <b>V</b> V
taken from (3) and data	5 10				1e-280 (fifth best)
from (4)					
	4000 sequences;	100 sequences;	800 sequences;	800 sequences;	800 sequences;
	2,000,000 characters;	50,000 characters;	400,000 characters;	400,000 characters;	400,000 characters;
	1.26 minutes	39.77 minutes	15.58 minutes	14 minutes	5.9 hours
CBF1	P-value ≤ 1.03e-89	2.3e-100	1.1e-39	5.00e-15	1e-321
tCACGTG				₽ <sup>2</sup> ]	
Consensus motif was		CAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	<b>"`ŀĂŸĂĠŶŶŶ</b>	°;] <b>Gaaaaaaaaaaaaaaaa</b>	Î. V
taken from (3) and data		9.1e-070 (second best)	2.0e-023 (forth best)	2.72e-10 (third best)	
from (4)					
		ŢŢŢŢŢŢŢŢŢŢŢŢ	<u>, ĈŇŬĈŎÎŎ</u>		
	4000 sequences;	100 sequences;	800 sequences;	800 sequences;	800 sequences;
	2,000,000 characters;	50,000 characters;	400,000 characters;	400,000 characters;	400,000 characters;
	1.2 minutes	43.3 minutes	29.48 minutes	11 minutes	5.7 hours
UME6	P-value ≤ 2.18e-102	2.0e-114	3.1e-45	3.52e-47	1e-321
taGCCGCCsa					
Consensus motif was		Ĩ <sub>ŧ</sub> Ţ <mark>ĮĄŲŲŲŲ</mark> ĮĄ	<u></u>		<sup>9</sup> <sup>5</sup> <sup>−</sup> <sup>−</sup> <sup>−</sup> <sup>−</sup> <sup>−</sup> <sup>−</sup>
taken from (3) and data			5.1e-030 (third best)		
from (4)					
	4000 sequences:	100 sequences:	800 sequences:	800 sequences:	800 sequences:
	2,000,000 characters;	50,000 characters;	400.000 characters:	400,000 characters;	400,000 characters;
	1.25 minutes	41 minutes	23.16 minutes	13 minutes	5.7 hours
TYE7	P-value ≤ 3.31e-62	7.9e-115	7.3e-43	9.26e-21	1e-321
tCACGTGa		MMALATATIATI TATITAATATITA A TT			2
Consensus motif was			ĨŢĨ <mark>ſŸŸŸŶŶŶŶŶ</mark>	TALALALALALALALALA	
taken from (3) and data	5 10		1.7e-018 (fourth best)	1.11e-10 (third best)	v v
from (4)			<b>CACGTGA</b> E		
	4000 sequences;	100 sequences;	800 sequences:	800 sequences:	800 sequences;
	2,000,000 characters;	50,000 characters;	400,000 characters;	400,000 characters:	400,000 characters;
	1.2 minutes	42.5 minutes	29 minutes	22 minutes	5.7 hours

GCN4	P-value ≤ 2.37e-42	1.3e-85	2.0e-32	4.00e-17	1e-321
TGAsTCa		M AL. A MAR AN			2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
Consensus motif was			<sup></sup> <sup></sup> <sup> </sup>		<sup>8</sup> . Nu <sup>n</sup> u u u u u u u u u u u u u u u u u u u
taken from (3) and data			2.5e-005 (ninth best)		
from (4)			TGAGTCAT		
	4000 sequences;	100 sequences;	800 sequences;	800 sequences;	800 sequences;
	2,000,000 characters;	50,000 characters;	400,000 characters;	400,000 characters;	400,000 characters;
	1.09 minutes	43.15 minutes	20.57 minutes	14 minutes	5.6 hours
INO2	P-value ≤ 1.14e-37	3.1e-88	1.1e-42	2.68e-17	1e-321
CAcaTGc					<u></u>
Consensus motif was		│ <sub>Ţ</sub> ŢŢŢŢŢŢŢŢŢŢŢŢŢŢŢŢŢŢŢŢŢŢŢŢ	<u></u>	°.]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]]	* <mark>````````````````````````````````````</mark>
taken from (3) and data			2.2e-007 (eighth best)		
from (4)					
	4000 sequences;	100 sequences;	800 sequences;	800 sequences;	800 sequences;
	2,000,000 characters;	50,000 characters;	400,000 characters;	400,000 characters;	400,000 characters;
	1.1 minutes	40.72 minutes	21.93 minutes	12 minutes	5.6 hours
MBP1	P-value ≤ 1.17e-47	7.9e-97	5.4e-47	1.01e-23	1e-321
ACGCGt Consensus motif was					<b>₽</b> G <b>_</b> AA
taken from (3) and data			7.6e-009 (sixth best)	1.31e-06 (forth best)	
from (4)				≝رً] <u>_ACGCGACGCG</u> T_	
	4000 sequences;	100 sequences;	800 sequences;	800 sequences;	800 sequences;
	2,000,000 characters;	50,000 characters;	400,000 characters;	400,000 characters;	400,000 characters;
	1.2 minutes	39.78 minutes	23.67 minutes	20 minutes	5.6 hours
SWI4	P-value ≤ 2.97e-33	8.3e-120	2.3e-47	2.91e-20	1e-321
CgCsAAA	#1.0 A C A M				<u>н</u>
Consensus motif was			" <sup>™</sup> <sup>™</sup> ไ ฟ ไ ฟ ไ ฟ ╋		<sup>3</sup> . Normal and a second secon
taken from (3) and data					1e-320 (third best)
trom (4)					
	4000 sequences;	100 sequences;	800 sequences;	800 sequences;	800 sequences;
	2,000,000 characters;	50,000 characters;	400,000 characters;	400,000 characters;	400,000 characters;
	1.22 minutes	40 minutes	23.62 minutes	13 minutes	5.6 hours

RNA	puf2	P-value ≤ 4.93e-51	3.2e-8	7.8e-34	8.94e-25	6.3e-322
			* 1524020404040404000000000000000000000000	Ţ <u>Ą</u> ŢĄŢ		, <b>TA A</b>
	Consensus motif was	5 10		8.0e-006 (third best)	2.23e-16 (second best)	6.7e-166
	taken from (5) and data from (6)			<sup>#</sup> . TAATAATA		<b>ATAATA</b>
		4937 sequences;	100 sequences;	988 sequences;	988 sequences;	988 sequences;
		636,234 characters;	16,991 characters;	139,903 characters;	139,903 characters;	139,903 characters;
		0.51 minutes	2.92 minutes	1.9 minutes	9 minutes	2 minutes
	puf3	P-value ≤ 9.44e-153	6.2e-151	7.4e-75	9.05e-65	2.5e-321
	<sup>3</sup> <sup>100</sup> <sup>20</sup> <sup>20</sup> <sup>20</sup> <sup>1</sup> <sup>20</sup> <sup>1</sup> <sup>20</sup>		<sup>a</sup> :C. GAATA	<b>≝</b> ∵ <mark>ŢĠŢĄ<sub>₽</sub>ĄŢ</mark> Ă		<b>₽,</b> ₩ <b>A</b> ₩,₩ <b>T</b> A₩,
	Consensus motif and data					8.7e-170 (third best)
	were taken from (6)					<b>ATGTA</b>
		4970 sequences;	100 sequences;	1007 sequences;	1007 sequences;	1007 sequences;
		640,926 characters;	11,304 characters;	132,972 characters;	132,972 characters;	132,972 characters;
		0.39 minutes	1.45 minutes	1.6 minutes	7 minutes	2 minutes
	puf4	P-value ≤ 5.22e-123	9.0e-50	4.3e-51	8.90e-20	9.6e-322
			* QAX	<sup>z</sup> , TAŢĄŢĄ		A TA A
	Consensus motif and data			4.4e-019 (second best)		6.7e-175
	were taken from (6)					<b>ÇȚĂȚĂ</b>
		4948 sequences;	100 sequences;	1002 sequences;	1002 sequences;	1002 sequences;
		638,135 characters	14,252 characters;	129,950 characters;	129,950 characters;	129,950 characters;
		0.49 minutes	2.16 minutes	1.58 minutes	8 minutes	2 minutes
	PUM2	P-value ≤ 4.9e-324	2.0e-101	2.3e-106	4.10e-148	5.5e-400
		BID CHANNE		<b>₽</b> ₽ <b>₽₽₽₽₽₽₽₽₽₽₽₽₽₽₽₽₽₽₽₽₽₽₽₽₽₽₽₽₽₽₽₽₽</b>		ț <mark>gta, ta</mark>
	Consensus motif and data					2e-398
	were taken from (7)					<b>TGTA_A</b>
		9995 sequences;	100 sequences;	1999 sequences;	1999 sequences;	1999 sequences;
		999,500 characters;	10,000 characters;	199,900 characters;	199,900 characters;	199,900 characters;
		1.35 minutes	1.25 minutes	3.43 minutes	20 minutes	3 minutes

OKI	P-value < 1 9e-321	2 Ne-19	9 80-51	2 550-15	1 20-100
		<sup>1</sup> <del> </del>	", <mark>™</mark> ∜čľ∛ůř		j <sup>z</sup> , , , , , , , , , , , , , , , , , , ,
Consensus motif and data	ŭ				2.5e-356
were taken from (7)					
	9997 sequences;	100 sequences;	2000 sequences;	2000 sequences;	2000 sequences;
	999,700 characters;	10,000 characters;	200,000 characters;	200,000 characters;	200,000 characters;
	1.09 minutes	1.17 minutes	3.03 minutes	25 minutes	10.6 hours
Puf5	P-value ≤ 3.47e-78	3.6e-9	6.8e-42	9.76e-21	3.8e-321
<u>XVUGUADUAD</u>		TATATATATATATATATATAT		ETCA-CCAACTACTC	ŢĂ <sup>Ţ</sup> Ă
Consensus motif and data		3.1e-004	3.1e-012	1.61e-20	4.2e-150
were taken from (8)		TGI AAJAXIA			
	4795 sequences;	100 sequences;	959 sequences;	959 sequences;	959 sequences;
	622,010 characters;	15,234 characters;	134,166 characters;	134,166 characters;	134,166 characters;
	0.43 minutes	2.52 minutes	1.53 minutes	10 minutes	3 minutes
Pub1	P-value ≤ 6.94e-177	7.5e-8	2.1e-36	5.73e-18	1.9e-203
KUUUUUKX			\$ <b>TĂĬŤĬ</b>	IATATATATATATATATA	
Consensus motif and data			2.4e-016	2.48e-05 (4th best)	4.3e-146
were taken from (8)			<sup>#</sup> ŢŢŢŢŢŢŢŢŢ		, TTT <sub>≈∓≈</sub> TAT
	4251 sequences;	100 sequences;	851 sequences;	851 sequences;	851 sequences;
	548,258 characters;	17,020 characters;	133,915 characters;	133,915 characters;	133,915 characters;
	1.03 minutes	2.88 minutes	1.33 minutes	9 minutes	2 minutes
Pab1	No motif was found	8.0e-30	5.4e-45	5.89e-28	7.8e-322
QUAUAUAZ		" ATATATATATATATATATATATATATATATATAT	# <b>ŢĂŤ</b> ŤŤ	#1] ATATATATATATATATA	<sup>3</sup> , <b>T</b> A, <b>A</b> , <b></b>
Consensus motif and data	4142 sequences;	100 sequences;	829 sequences;	829 sequences;	829 sequences;
were taken from (8)	534,415 characters;	12,734 characters	99,365 characters;	99,365 characters;	99,365 characters;
	0.23 minutes	1.96 minutes	1.33 minutes	17 minutes	8.4 hours

Khd1		P-value ≤ 1.39e-16	2.1e-12	1.2e-45	1.24e-26	1.6e-320
<b>XXX</b>	CAUUXY			<sup>≴</sup> ¹ <b>ŢĄ<sub>ŢĄ</sub>ŢĄ</b>	≝ <sup>1</sup> ] <u>tatatatatatatatat</u>	
Consensi were tak	us motif and data en from (8)					7.6e-147
		4773 sequences:	100 sequences:	955 sequences:	955 sequences:	955 sequences
		619.178 characters:	16.243 characters:	137.956 characters:	137.956 characters:	137 956 characters
		0.35 minutes	2.83 minutes	1.28 minutes	9 minutes	8.4 hours
Nab2		P-value ≤ 4.80e-15	1.9e-13	1.0e-39	8.16e-29	1.4e-140
× A	8 <mark>ASG</mark> 88		ATATA ATATA ATATATA			
Consense	us motif and data			2.2e-008	4.16e-9	
were tak	en from (8)			<sup>ª</sup> , <mark>ĂŸŸŸŤŤŤ</mark>		
		4101 sequences;	100 sequences;	821 sequences;	821 sequences;	821 sequences;
		528,461 characters;	11,514 characters;	111,334 characters;	111,334 characters;	111,334 characters;
		0.29 minutes	1.67 minutes	1.58 minutes	10 minutes	3 minutes
Vts1		No motif was found	4.5e-5	1.0e-14	1.08e-8	2.1e-75
USX	CGSGGĕ		*ATATATATATATATATATA	\$Ţ <b>Ą</b> ŢĄŢ		
Consensi	us motif and data	1787 sequences;	100 sequences;	358 sequences;	358 sequences;	358 sequences;
were tak	en from (8)	234,740 characters;	13,720 characters;	45,130 characters;	45,130 characters;	45,130 characters;
		0.09 minutes	2.23 minutes	0.38 minutes	7 minutes	2 minutes
Pin4		P-value ≤ 3.72e-9	1.3	3.1e-51	1.65e-28	2.2e-167
ĂNĂ	AA¥GA			<b>ŢĂĬŶĬ</b>	≝ <mark>] IATATATATATATATAT</mark>	
Consense	us motif and data	4261 sequences;	100 sequences;	853 sequences;	853 sequences;	853 sequences;
were tak	en from (8)	549,860 characters;	12,003 characters;	109,847 characters;	109,847 characters;	109,847 characters;
		0.25 minutes	1.76 minutes	1.04 minutes	11 minutes	7.5 hours
Nrd1		P-value ≤ 6.69e-13	6.6e-8	4.1e-44	4.15e-20	2.2e-143
₽ncr	JUG⊻⊉		TereIe AlereIAIA AIAreIeIeIe <sup>*</sup>	<sup>z</sup> , ŢĄŢĄŢĄ		<b>ATGTA</b>
Consensi	us motif and data			1.1e-004 (third best)		
were tak	en from (8)					
		3947 sequences;	100 sequences;	790 sequences;	790 sequences;	790 sequences;
		507,022 characters;	14,906 characters;	104,759 characters;	104,759 characters;	104,759 characters;
		0.24 minutes	2.48 minutes	1.28 minutes	10 minutes	7.5 hours

YLL032C	P-value ≤ 5.84e-7	2.3e-11	3.4e-24	9.54e-14	4.6e-69
A U A C C S	ag in GAUACCC	IATA A A A A A A A A A A A A A A A A A A	ŧŢĂŢĂŢ		
Consensus motif and data	2286 sequences;	100 sequences;	458 sequences;	458 sequences;	458 sequences;
were taken from (8)	295,474 characters;	14,249 characters;	58,789 characters;	58,789 characters;	58,789 characters;
	0.11 minutes	2.38 minutes	0.48 minutes	8 minutes	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 <u>http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2441782/</u>
- 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 <a href="http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3351182/">http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3351182/</a>
- 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 <u>http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3006441/</u>
- 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 <a href="http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3006441/">http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3006441/</a>
- 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 <a href="http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3153972/">http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3153972/</a>
- 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 <u>http://www.ncbi.nlm.nih.gov/pmc/articles/PMC368173/</u>
- 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 <u>http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2861495/</u>
- 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 <u>http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2573929/</u>