














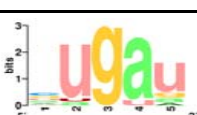





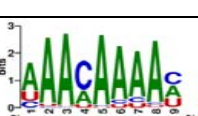

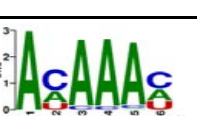













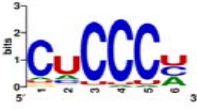


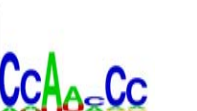



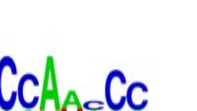





































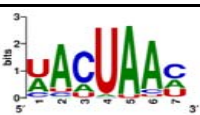

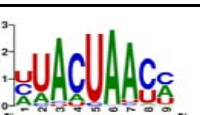



















































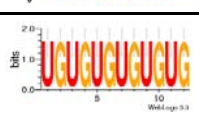



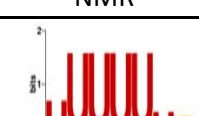
Supplementary Table S1

















Protein	Source	Motif 1	p-value L	Motif 2	p-value L	Motif 3	p-value L	<i>In vivo</i> Motif *	Structure Info**
EIG4G2 eCLIP K562-01 h38	ENCODE		2.00e-308 7		2.75e-308 5		1E-282 6		
ELAV1 PARCLIP HEK293 h19	doRiNa		6.78E-08 7						single stranded
EWSR1 PARCLIP HEK293-01 h19	CLIPdb		8.78E-30 7		2.34E-23 5		2.69E-19 6		
EWSR1 eCLIP K562-02 h38	ENCODE		3.64E-166 6		4.3E-135 7		3.3E-128 5		
FIP1L1 PARCLIP HEK293 h19	CLIPdb		5.35E-74 4		1.57E-63 5		4.5E-49 6		
FMR1 PARCLIP HEK293 h19	doRiNa		2E-213 6		1.4E-174 4		2.6E-151 5		
FUBP3 eCLIP K562-01 h38	ENCODE		9.4E-133 6		1.3E-102 5		1.12E-96 7		
FUS PARCLIP HEK293 h19	CLIPdb		1.24E-53 7		1.92E-41 6		7.43E-38 5		SELEX

hnRNPA1 eCLIP K562-02 h38	ENCODE		3.54E-56	5		3.23E-53	7		9.56E-51	6	
hnRNPC iCLIP HeLA h19	doRiNa		3.6E-06	7		3.86E-06	5				
hnRNPC eCLIP HepG2-01 h38	ENCODE		1.1E-232	6		9.5E-153	5		1.57E-43	4	
hnRNPK eCLIP HepG2-01 h38	ENCODE		2.00e-308	5		2.00e-308	6		2.00e-308	7	
hnRNPK eCLIP K562-02 h38	ENCODE		2.9E-156	7		2.9E-146	6		2.1E-113	5	
hnRNPL eCLIP HepG2-01 h38	ENCODE		2.00e-308	5		2.00e-308	6		2.00e-308	7	
hnRNPL eCLIP K562-01 h38	ENCODE		2.00e-308	5		2.00e-308	6		2.00e-308	7	
hnRNPM eCLIP HepG2-01 h38	ENCODE		3.73E-96	5		4.49E-93	6		4.33E-46	7	
hnRNPM eCLIP K562-01 h38	ENCODE		3.37E-68	5		3.8E-56	7		8.32E-39	6	
hnRNPU eCLIP HepG2-01 h38	ENCODE		2.01E-54	5		1.22E-50	6		5.31E-47	7	

hnRNPU eCLIP K562-01 h38	ENCODE		1E-127	5		2.2E-115	6		4.3E-115	7	
IGF2BP1 PARCLIP HEK293 h19	CLIPdb		5.42E-59	7		6.14E-44	5		3.58E-29	6	
IGF2BP1 eCLIP K562-01 h38	ENCODE		2.13E-67	5		6.81E-53	7		2.33E-42	6	
IGF2BP2 PARCLIP HEK293 h19	CLIPdb		1.78E-95	6		4.33E-87	7		2.39E-58	5	
IGF2BP2 eCLIP K562-02 h38	ENCODE		4.4E-138	7		1.4E-133	5		5E-129	6	
IGF2BP3 PARCLIP HEK293 h19	CLIPdb		2.94E-97	7		9.36E-97	6		6.28E-96	5	
IGF2BP3 eCLIP HepG2-02 h38	ENCODE		1.5E-142	7		3.7E-126	6		3.9E-119	5	
KHDRBS1 eCLIP K562-01 h38	ENCODE		2.00e-308	7		7.2E-305	6		1E-296	5	
KHSRP eCLIP K562-01 h38	ENCODE		2.5E-145	6		9.1E-115	5		3.3E-100	7	
PCBP2 eCLIP HepG2-02 h38	ENCODE		2.00e-308	5		2.00e-308	7		2.00e-308	6	

PTB1 eCLIP HepG2-01 h38	ENCODE		2.00e-308	5		2.00e-308	7		2.00e-308	6		single stranded
PTB1 eCLIP K562-01 h38	ENCODE		2.00e-308	5		2.00e-308	7		2.00e-308	6		single stranded
PUM2 PARCLIP HEK293 h19	doRiNa		1.09E-22	6		6.39E-19	5		1E-16	7		
PUM2 eCLIP K562-02 h38	ENCODE		1E-261	7		1.4E-210	6		1.1E-205	5		
QKI PARCLIP HEK293 h19	ENCODE		2.00e-308	5		3.1E-289	6		1.2E-267	7		single stranded
QKI eCLIP K562-01 h38	ENCODE		2.00e-308	5		2.00e-308	6		2.00e-308	7		single stranded
RBPF0X2 eCLIP HepG2-02 h38	ENCODE		1.1E-175	6		4.5E-163	5		1.3E-158	7		
RBPF0X2 eCLIP K562-02 h38	ENCODE		7.5E-101	5		3.29E-70	6		2.69E-68	7		
SLBP eCLIP K562-01 h38	ENCODE		1.6E-39	5		2.11E-36	6		2.38E-33	7	AAAggcucuUUUC from SOLVED STRUCTURE	stem-loop
SRSF1 eCLIP HepG2-02 h38	ENCODE		2E-88	6		8.43E-84	5		9.08E-80	4		single stranded



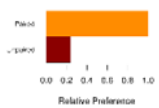




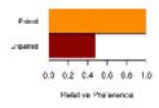




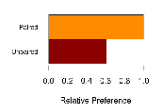




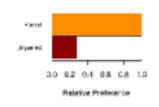




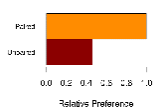




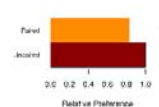


SRSF1 eCLIP K562-02 h38	ENCODE		9.9E-137	4		7.6E-117	5		1.5E-102	6		single stranded
SRSF1 HITSCLIP MEF mm9	doRiNa		5.16E-32	4		8.13E-27	5		1.85E-16	6		single stranded
SRSF2 HITSCLIP MEF mm9	doRiNa		1.29E-36	4		7.78E-27	5		2.62E-26	6		
SRSF7 eCLIP K562-02 h38	ENCODE		1.1E-165	5		2.1E-155	6		3.3E-150	7		
SRSF9 eCLIP HepG2-02 h38	ENCODE		1.1E-107	6		2.1E-107	5		4.8E-107	7		
TAF15 PARCLIP HEK293-01 h19	CLIPdb		9.51E-12	7		2.68E-09	6		7.6E-08	5		
TAF15 eCLIP HepG2-02 h38	ENCODE		3E-179	5		1.4E-175	6		2.1E-168	7		
TARDBP iCLIP SH-SY5Y h19	doRiNa		1.8E-301	7		4.6E-276	5		6.9E-271	6		
TARDBP eCLIP K562-01 h38	ENCODE		2.00e-308	5		2.00e-308	6		2.00e-308	7		NMR
TIA1 iCLIP HeLa h19	doRiNa		6.9E-21	7		3.64E-15	6		3.67E-12	5		

TIA1 eCLIP HepG2-01 h38	ENCODE		7.8E-194	5		2.3E-106	7		1.3E-105	6	
TRA2A eCLIP K562 h38	GEO		7E-220	6		2.7E-215	5		1.7E-213	7	
U2AF2 eCLIP HepG2-01 h38	ENCODE		2.5E-110	5		7.03E-98	6		2.11E-66	4	
U2AF2 eCLIP K562-02 h38	ENCODE		1.31E-55	4		8.98E-54	5		1.05E-53	6	



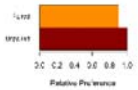




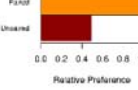
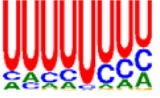

* *In-vivo motifs were extracted from CISBP-RNA <http://cisbp-rna.cabr.utoronto.ca/> or from AtTRACT <http://attract.cnic.es>.*

** *Structure information was extracted from RNAcompete-S (Cook et al., 2017)*

Supplementary Table S2

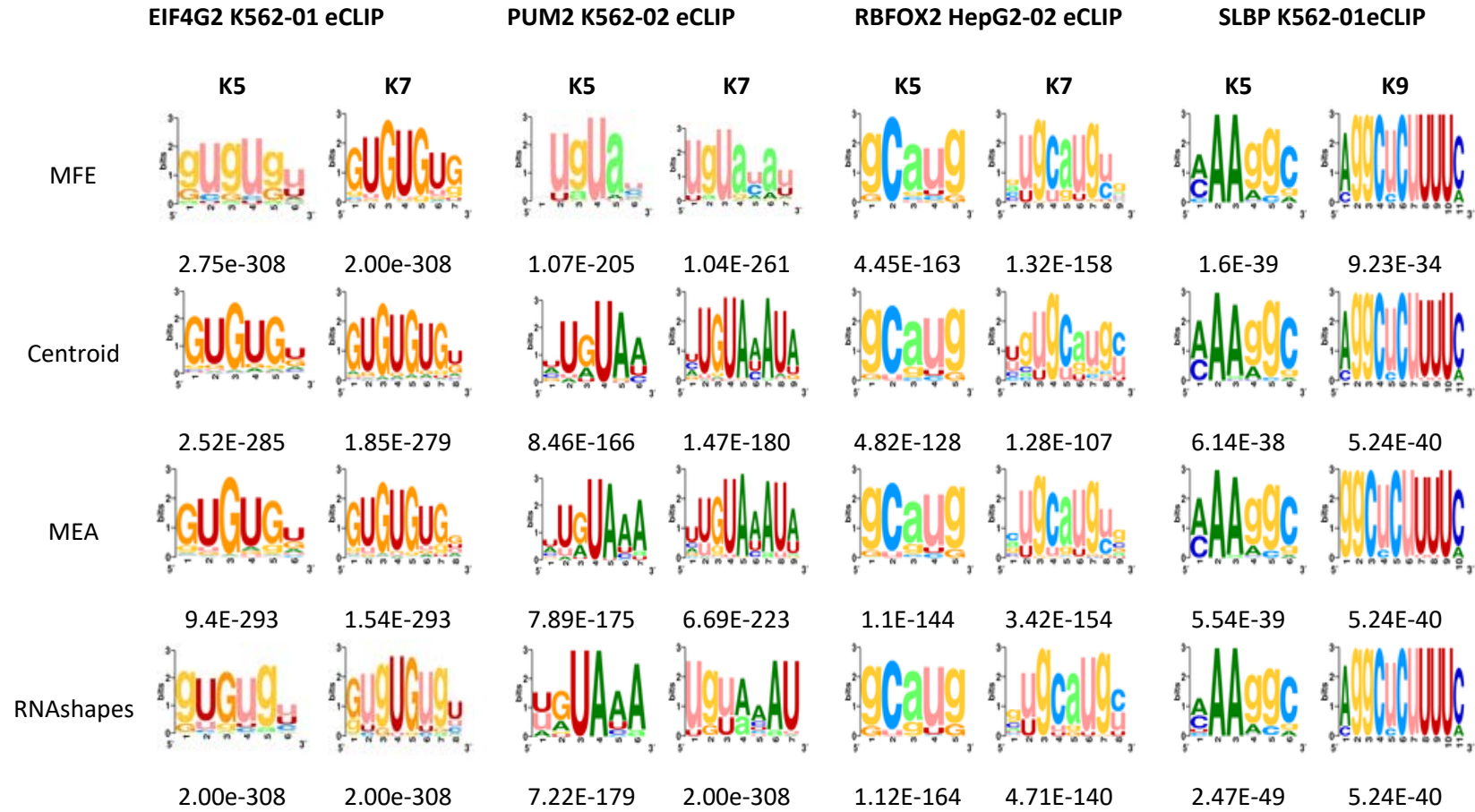
RBP+source	SMARTIV *	RBPMotif		GraphProt	
	Seq+Struct	Seq	Struct	Seq	Struct
EIF4G2 eCLIP K562_01 hg38					
EWSR1 eCLIP K562_01 hg38					
FUBP3 eCLIP HepG2-02 h38					
hnRNPA1 eCLIP K562-02 h38					
hnRNPM eCLIP HepG2-01 h38					
KHSRP eCLIP K562-01 h38					

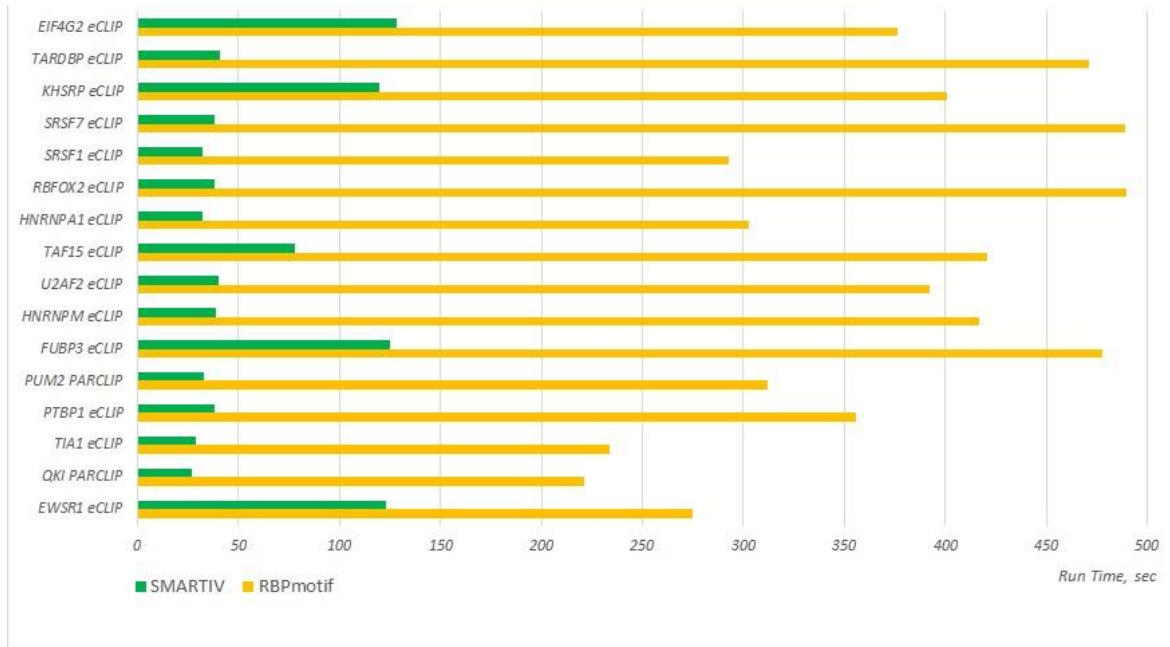
PTB1 eCLIP K562-02 h38					
PUM2 PARCLIP HEK293 h19					
QKI PARCLIP HEK293 h19					
RBFOX eCLIP HepG2-02 h38					
SRSF1 eCLIP HepG2-01 h38					
SRSF7 eCLIP K562-02 h38					
TAF15 eCLIP HepG2-02 h38					
TARDB eCLIP K562-01 h38					

TIA1 eCLIP K562-02 h38					
U2AF2 eCLIP HepG2-02 h38					

* For comparison SMARTIV was run on a 1000 sequence on the top of the list and 1000 on the bottom

Supplementary Table S3





Supplementary Figure S1: SMARTIV vs RBPmotif runtimes

SMARTIV and RBPmotifs were tested on 16 different dataset (as detailed in Supplementary Table S2) using default parameters. For comparison with RBPmotifs, we run both SMARTIV and RBPmotifs on 2000 sequences per each dataset (top 1000 and bottom 1000 sequences extracted from the ranked sequence list). As shown SMARTIV run time is between 30 second to 3 minutes, depending on the total sequence length and the GC content of the sequences. RBPmotif runtime was usually 5 to 10 fold slower.