Table 1: Motif comparison for 30 TFs with six different programs. Table shows the motifs learned by various algorithms used by Harbison *et al*. and those learned by Our algorithm. For comparison, we use the motifs with the top MAP score for AlignACE [6], MEME [1], MDscan [4], MEME_c (modified MEME with conservation information), method by [3], and CONVERGE (a method by Harbison *et al*.), the latter three being based on conservation information) In the ninth column we report the top motif according to our score. We also report the predicted class and the percentage of entries in $\boldsymbol{c}$ contributing to that class. The last column is the literature consensus as used by Harbison *et al*. collected from YPD, SCPD, and TRANSFAC databases at the time the paper was published. The bold sections in the motifs indicate either a match with the literature consensus in the final column or to a motif we found in the literature search we conducted. In cases where the match is not obvious, it is probably because the reverse complement of the sequence matches the literature consensus. Lower case letters in the motifs indicate a weaker preference (less information content at that position). Ambiguity codes: S=C/G, W=A/T, R=A/G, Y=C/T, M=A/C, K=G/T, and '.'= A/C/G/T.

TF	Environ cond.[#]	Harbison *et al*. AlignACE Top MAP	Harbison *et al*. MEME Top MAP	Harbison *et al*. MDscan Top MAP	Harbison *et al*. MEME_c Top MAP	Harbison *et al*. Kellis Top score	Harbison *et al*. CONVERGE Top enrichment[&]	PRIORITY Top MAP	PRIORITY Predicted class	Literature
Basic Leucine Zipper TFs										
Arr1	**YPD**	R.AmA.a.A.A.AmA.A	cAmAcACmcAmAmayrcA	CACACACAC	TkTGtgTkttTstwT	GGA.....GTG	AAACAmAsAmACAMA	YAAACAca	fork 78%	TTACTAA
Cad1	SM	AAraRARAAA..A	AAGrAArArAgR	A**TTA.TmA**	.A**TTAgTMA**kCa	TGC..GGG	sAAAArKTAtTCt.A	c**TKACWAA**	bZip 76%	TTACTAA
	YPD	GtGTGTGkGTGTG	GC**TKACTAA**T.	GKGTGTGK	GC**TkACTAA**T	GCC.....TAT	t..yGm**TtA.tAA**t.	GC**TTACTA**	bZip 73%	
Cin5	H2O2Hi	mR.RAAARAAR	yyyyT.ctTTYTykysT.	TTTtTTTTT	TkTtTt.TyTtktkyyTT	CTGCTTA	gc**TTA..TmA**k....	GCGTG...	bHLH 83%	TTAC[r]TAA *
	H2O2Lo	AARAAAA.AA.A	TTTYyttytTy.ytyYYK	.GSGssgG	AAAAa..AAAraaraAAA	TAC...AGC	.gm**TTAyrTAa**....	**TTAyGTAA**	bZip 94%	
	YPD	RRARAARAAA	GGC.cgGCCGS	**TTAyrTAA**	—	GGG.....CTC	a**TkA.rTaA**gc....	.gcTGG..	bHLH 92%	
Cst6	**YPD**	A.A..rAAA.A..A..a.A	rmAtk.mAwrCRAAaA	AgTY.AsT	rmAtk.mAwrCRAAaA	TTG....ACA	TCCyAkTAYTCTkCA	.ACTGGAC	bZip 80%	—
Gcn4	RAPA	RAAAAARAAr	**TGAGTCA**	.**TGAsTCA**	.**TGAGTCAy**	**rTGAsTCA**	.....**TgAsTCa**Y	**ATGACTCA**	bZip 95%	ArTGACTCw
	SM	rAAAAARAAa	yTyTyyTyyYTyTTTc	**TGAsTCAt**	.**tGAGTCAy**	**.rTGACTCw**T	**rTGACTca**	**ATGACTCA**	bZip 96%	
	YPD	rAAAAArAAA	.**TGAGTCAy**	.**TGAsTCA**	**TGAGTCAy**	**rTGACTCA**.	**.rTGAsTCA**.....	.**TGACTCA**	bZip 91%	
Hac1	**YPD**	A..rAA..MAAARA	TrCSTSkccwywtmM	**TAcGTGkC**	TrCSTSkccwywtmM	CTC.....CAT	amAAyGAksAgCAAA	**ACGTGTCA**	bZip 76%	kGmCA[G]CGTGTC **
Met28	**YPD**	a.A.A.....AAAA.AAA	TkyTTTTkskssskcTTw	ATrTAyAT	CGGCCTC	ATA.CTT	mAmASsrYGkGTrTr	SKAAACYG	bZip 75%	—
Met4	**SM**	AA.AR.RARAAA	AArAAMmmRmAA	TATATATAT	ASTGTGGySGyS	TCT.....TAC	MMAAsTGTGGySsCr	TGTCACGt	bHLH 80%	—
Sko1	**YPD**	mAAA.RARr.AA	TkTTkyyykTTTkyKKCk	sSgtacSs	kwTyTTkTyyTyktTTkT	CTG.....AGG	A**TGACGT**r	.t**ACGTCA**	bZip 72%	ACGTCA
Yap1	**H2O2Lo**	AwMArrAAr..A	ssTTTyCrT	**TTA.TAA**k	**TTAGTmA**Gc	CCG..TTT	c**TkACtAA**	**TKACTAA**w	bZip 87%	TTAsTAA
	YPD	ARrAAAArAR	yTyy.TkC.TKkyyttmt	CAC.CasA	TTt..tTtt.Tt.tktTT	TAGTCAGC	a**TTAgTmA**	..YAAAC.	fork 87%	
Yap3	**YPD**	A.A...A.A.A.A.A.A.Amr	T.kyttcTT.mTTkTT	CACACACAC	gkkkkCkykmTtrT..	TTC.....AAC	rAAArsCATGTATyT	.CTAAaTS	bZip 65%	TTACTAA
Yap5	**YPD**	GTGTG.GTGTG	GTCGAgsAgAAcsAgGAt	CACACmCAC	ATGTAyggrtg	TTG.CGG	yGmrsAGmACsAGrA	CGTGKGYG	bHLH 93%	TTACTAA
Yap6	H2O2Hi	ARrRAAA.AAR	s.TtTyytykYc.yt.YT	cySggcCG	ssGgCsGAgcss.s.m	TTT.TCC	wAtmTGKwCCTrYGs	CGTGGG.g	bHLH 87%	TTACTAA
	H2O2Lo	RARARRA.AAA	A.r.ArAArrrrrAAA	scyGCRgs	yTTTtCtTyTTytyCsTt	ATC.CTT	ARartTtA	..CGTGG.	bHLH 91%	
	YPD	RArAArAAAA	magAAA.rrrAArArR	smYGCAks				..CGTGG.	bHLH 91%	
Yap7	**H2O2Hi**	AAAARRAAAR	m**TTAsTmA**kc..	**TKAsTMA**k	raA.aArrAAaarGAAAA	GAC.AAG	AAcyYYra	**TkASTAA**K	bZip 82%	TTACTAA
	H2O2Lo	RRAAA.ARAAR	TtT..yyttTctyTTyyT	**TTAsTMA**k	Gc**TKAsTAA**	TTA...AGC	.y..**TkAsTaA**k...	A**TKASTAA**	bZip 92%	
Forkhead TFs										
Fhl1	RAPA	RTGTayGGrTg	.t.tacAyCCrTACAyyy	yCCrtACA	rrrrTGtayGGrTgta.a	TGT.TGGrT	rrTGYayGGrTGtw.	TGTAYGGg	other 91%	—
	SM	RTGTayGGrTg	.trCAyCCrTRCAyyy	GTayGGrT	rrrrTGTAyGGrTgta.a	TGT.TGGrT	CAyCCrtACAyyy..	TGTAYGGt	other 94%	
	YPD	RTGTayGGrTG	.t.taCAyCCrTACAYyy	TGyryGGr	rrrrTGTayGGrTgta.a	TGT.TGGrT	rrRTGTayGGrTGtw	TGTAYGGR	other 94%	
Fkh1	**YPD**	RAR.ARA.RAA.A	aaa.**rtAAACAa**..r.a	**tTGTTTAC**	.k...**tTGTTTAC**.tTtk	**.TTGTTTAC**	.raa.**gTAAACAa**..	**GTAAACAA**	fork 95%	GGTAAACAA
Fkh2	H2O2Hi	RRaARR.AAA.R	TTttyc.TTYyyty.TTT	T.TKTmCm	gg.aAAa.**GTAAAcAr**	**TTGTTTAC**	gg.aAaa.GTAAAcA	.**GTAAACA**	fork 92%	GGTAAACAA
Fkh2	H2O2Lo	ArMArr.AAR.A	gg.AAAA.**GTMAACA.**ya	S.AgAgAG	gG.rAaw.**gTMAACA**	ACG.....ATT	gG.AAaa.GTMAACA	**GTAAACAA**	fork 80%	
Fkh2	**YPD**	rRaAAR.AAA.R	AArrr.rAAAaa.r.AAA	**.GtAAACAA**	g.rAaa.**gTAAACAa**.r.	**TTGTTTAC**..TT	.raaa.**gtAAAcAa**.	**GTAAACAA**	fork 94%	

TF	Environ cond.[#]	Harbison *et al.* AlignACE Top MAP	Harbison *et al.* MEME Top MAP	Harbison *et al.* MDscan Top MAP	Harbison *et al.* MEME_c Top MAP	Harbison *et al.* Kellis Top score	Harbison *et al.* CONVERGE Top enrichment[&]	PRIORITY Top MAP	PRIORITY Predicted class	Literature
Basic Helix Loop Helix TFs										
Cbf1	**SM**	rAAAAARAAR	**RTCACGTGm**	**kCACGTGm**	—	**CACGTGr**y	**CACGTGA**y	**rTCACGTG**	bHLH 97%	rTCACrTGA
	YPD	**CACGTGA**cc.	**CACGTGA**CCm	**CACGTGay**	gG**TCACGTG**	TCT....AGC	**CACGTGAC**cm..wk.	**GTCACGTG**	bHLH 67%	
Ino2	**YPD**	rARARARR.AA	s**CAYsTGMw**.a	k**CAsrTG**c	.s**CATGTGma.**A	**TTCACAT**GC	g**CATGTGa**	**TCACAT**GC	bHLH 86%	ATTTCACATC
Ino4	**YPD**	A...A.AARrArAR	t**TTYCACATG**s	**CAygTGma**	t**TTTCACATG**s	T**.TTCACATG**	**.CATGTGa**	**CATGTGAA**	bHLH 85%	CATGTGAAAT
Phd1	**BUT90**	rRAAARrRAA	rAaA.grAaA.RrRaA	.SSSsSSS	TT.TTTtT.TTtkTytTT	CGT.....TTT	RAA.rrrm.armrAA	kCGTGsc.	bHLH 95%	—
	YPD	RARARRR.AAA	yyTtk..yyycykTTkyC	sctGCags	GMAGGCACg	GCA..GAT	.mAGGCAc	..GTGCCT	bHLH 92%	
Pho4	**Pi-**	AAAAArAAA..A	**sCACGTg**	**sCACGTGs**	g.**CACGTGs**	C**CACGTG**	**cAcGTGss**	**CACGTG**cs	bHLH 81%	wcacgtk.g
	YPD	r.AAR.RRAAAA	yTgyTkTtcYTtwT.Y	GTGTGtgT	AaAakAgAaAAaggawAr	CCT.CCT	TSCkYGAGYAsTWAR	**YRTGKG**.G	bHLH 89%	
Sok2	**BUT14**	ARAArRAAA.R	ArrM..AAAmr.RrAA	SSss.sSG	AaaAr.AAAaAgAaAA	CGT.....TTT	yAGGsAm.	.sCGTG..	bHLH 93%	—
Tye7	**YPD**	rRAARAraAAAYs	rY**CACsTG**AYg	T**CACGTG**A	rY**CACsTG**Ayg	**CACGTG**A	T**CACGTG**A	**CAsGTG**AT	bHLH 92%	CA..TG
TFs belonging to structural classes other than the three modeled explicitly										
Leu3	**SM**	aA.AAAaAAA...A	**gCCsGtacCGS**wc	**CGgtacCG**	g**CSgGTacCGs**	**CCGGT.CCGG**	**cSGgTwcC**	**GgtACyGG**	other 80%	yGCCGGTACCGGyk
	YPD	rARArrAARAA	m.GmaAsaaaArkaAArm	**CGgtacCG**	m.GmaAsaaaArkaAArm	**CCGGT.CCGG**	rATkyrTcAkTMATA	**CGRTWcCG**	other 86%	
Nrg1	**H2O2Hi**	rAAAAArAAR	srAarmSrAAA	gGA**CCCT**k	GGA**CCCT**k	TAAACGA	**.aGGG**tcc	.GGA**CCCT**	other 89%	CCCT
	H2O2Lo	AArAARAaAR	rrAArakArAAmsraAAr	ggsGgsGG	rAaAr.gAAAAAsaAAAr	TTC.....GGC	mAGGcAcr	CGTGCmTS	bHLH 86%	
	YPD	rAARARrAAR	rAAA.rrrAA.RaAAR	CACACACA	CssAr.CsCmcS.s.m	CTT.....ATG	AAGgAAAA	G.GTGTGt	bHLH 90%	
Rap1	**YPD**	**rTGYayGGrTg**	**..grTGyayGGrTGyr**	**yCCRtrCM**	**rrTGyayGGrTG**ya	**TGT.TGGGT**	**.wrCAyCCrtrCAyc**	**ACCCRTAC**	other 90%	wrmACCCATACAyy
Reb1	H2O2Hi	k**sCGGGTAA**y	.r**TTACCCG.**myt	r**TTACCCG**	r**TTACCCG.**m	**TTACCCG**	**sCGGGTAA**	m**TTACCCG**	other 86%	TTACCCGG
	YPD	k**sCGGGTAA**y	...k**sCGGGTAA**y.	r**TTACCCG**	.r**TTACCCGs**m.	G**TTACCCG**	r**TTACCCGs**m.....	m**TTACCCG**	other 93%	
Ste12	**Alpha**	AAAArRAAA..R	gAaACa..t.**TgAaACa**	**tGTTTCA.**	wTTtwwAmwwkwwTta	**TGAAAC**r	Aaacr....**yrAaAc**	tt**TGAAAC**	fork 95%	ATGAAAC
	BUT14	RAARAA.RAAA	tTTtyyyTtyt.k.YTTY	t**GTTTCA.**	GCAraarM.rcAagAmsm	**TGAAAC**A	rcATTcyy	AaGAATCT	other 91%	
	BUT90	rAr.AARrAAA	smaaA..A.r.Ar.Rrra	**.TgAAAC**A	CCsrAw..GGaA	**TGAAAC**A	..r..t**GTTTCA.**m.	.**TGAAACA**	other 94%	
	YPD	AArAaAAAAA	AmAaRAAgsAArA.aaaa	**.TrAAAC**A	aAACA.tT.**TGAAAC**m	**TGAAAC**	AAacr..t.**yrAAAc**	.**TgAAACA**	other 78%	
Ume6	H2O2Hi	**TsgGCGGCT**A	ww**TAGCCGCcsa**..	T**AGCCGCC**	aww**TAgCCGCcsa**..	T**AGCCGCC**s.	.....wt**TAGCCGCc**	T**AGCCGCC**	other 92%	wGCCGCCGw
	YPD	**TsGGCGGCT**A	ww**TAGCCGCCsA**.s	T**AGCCGCC**	ww**TAGCCGCCsA**.s	.w**TAGCCGCC**sA..	s.**tsgGCGGCT**Aww.	**AGCCGCCs**	other 86%	

* The motif with the inserted r was experimentally confirmed by Harbison *et al*. after they conducted a gel-shift assay to verify the authenticity of the motif they obtained by their *in silico* analysis for Cin5.

** Harbison *et al*. report the longer motif with the central G as literature consensus for HAC1, but in a literature search we conducted, we found that a new binding site TYACGTGTYM without the central G has been experimentally confirmed by [5] using gel-retardation assays.

[#] Regulators were profiled in one or more specific environments by Harbison *et al*. The condition in bold (one per TF) is the condition used by Harbison *et al*. in their final motif, which we have used in the main text of the paper. In cases where they do not report a final motif, we arbitrarily select for the main text of the paper the condition which yields the largest number of probes. Here, we list all. Environmental conditions in brief are Alpha: mating, BUT14/BUT90: filamentation, H2O2Hi: highly peroxide, H2O2Lo: mildly peroxide, Pi-: phosphate deprived, RAPA: nutrient deprived, SM: amino acid starved, YPD: rich medium.

[&] Although all other programs mentioned in this table report a score for the discovered motif, we did not find a score attached to CONVERGE from the results reported by Harbison *et al*.. Here, we use the reported motif with the top enrichment score. It is thus important to note that this is a post-processed motif, giving CONVERGE a slight benefit over all other programs in the table.

Table 1 reports results for all TFs considered in the main text of the paper, and in all environmental conditions used by Harbison *et al*. Sequence sets contained at least 10 probes, each bound with a p-value< 0.001 by the respective TF. There are 54 total sequence sets covering the 30 TFs. Known binding sites are reported in the literature for effectively 45 of these sequence sets (assuming the same binding mechanism for a TF in all conditions, which may not be necessarily true). Out of these 45 sequence sets, PRIORITY finds the correct binding site motif in **33**, AlignACE finds it in **6**, MEME finds it in **19**, MDscan finds it in **29**, MEME_c finds it in **24**, Kellis finds it in **22**, and CONVERGE finds it in **29**.

In 10 of the 12 cases that PRIORITY does not find a motif matching the literature, it predicts a class other than the class of the TF being profiled. This is helpful in diagnosing that the reported motif may not be the correct motif for the profiled TF.

Among the 9 sequence sets with no known binding site in the literature, PRIORITY predicts a motif of the correct class in the sequence sets of bZips Cst6 and Met4, so these might be good candidates for experimental validation. The cases of Met4, Fhl1 (where PRIORITY consistently predicts a class of "other" in all sequence sets), Phd1 (where PRIORITY finds a similar bHLH motif in both sequence sets), and Sok2 have all been discussed in detail in the main text of the paper.

References

[1] Bailey,T. and Elkan,C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers, *ISMB '94*, AAAI Press, Menlo Park, California, pp. 28–36.

[2] Harbison,C., *et al*. (2004) Transcriptional regulatory code of a eukaryotic genome, *Nature*, 431:99–104.

[3] Kellis,M., Patterson,N., Endrizzi,M., Birren,B., and Lander,E. (2003) Sequencing and comparison of yeast species to identify genes and regulatory elements, *Nature*, 432:241–254.

[4] Liu,X., Brutlag,D., and Liu,J. (2002) An algorithm for finding protein-DNA binding sites with applications to chromatin immunoprecipitation microarray experiments, *Nature Biotech.*, 20:835–839.

[5] Patil,C., Li,H., and Walter P. (2004) Gcn4p and novel upstream activating sequences regulate targets of the unfolded protein response, (*PLoS Biology*), 2(8):E246.

[6] Roth,F., Hughes,J., Estep,P., and Church,G. (1998) Finding DNA regulatory motifs within unaligned non-coding sequences clustered by whole-genome mRNA quantitation, *Nature Biotech.*, 16:939–945.