Molecular Cell, Volume *32*

Supplemental Data

A Library of Yeast Transcription Factor Motifs

Reveals a Widespread Function for Rsc3

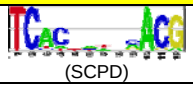
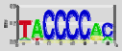
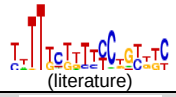
in Targeting Nucleosome Exclusion at Promoters

Gwenael Badis, Esther T. Chan, Harm van Bakel, Lourdes Pena-Castillo, Desiree Tillo, Kyle Tsui, Clayton D. Carlson, Andrea J. Gossett, Michael J. Hasinoff, Christopher L. Warren, Marinella Gebbia, Shaheynoor Talukder, Ally Yang, Sanie Mnaimneh, Dimitri Terterov, David Coburn, Ai Li Yeo, Zhen Xuan Yeo, Neil D. Clarke, Jason D. Lieb, Aseem Z. Ansari, Corey Nislow, and Timothy R. Hughes

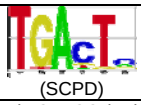
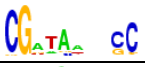
Table S1. List of yeast sequence-specific DNA-binding transcription factors and their binding specificities. We compiled this list by first searching all translated yeast ORFs on PFAM, and identifying those with an established transcription factor DNA-binding domain. Then, we compiled the yeast proteins with GO annotations as "regulation of transcription". For those without a known DNA-binding domain, we searched the literature for evidence that they directly bind specific DNA sequences and regulate transcription. This process has subjective aspects, and the list may not be complete. Also, not all proteins with the indicated domains may be transcription factors; C2H2 zinc fingers, for example, can also have RNA-binding or protein-protein interaction activity.

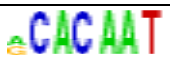
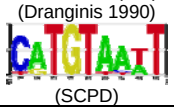
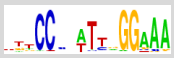
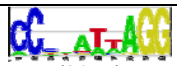
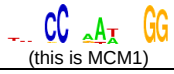
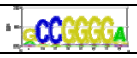
We first compared our motifs to the Harbison and MacIsaac data. In the event of a discrepancy or absence of a motif, we then compared to entries on individual SGD pages, and then search Medline to identify literature motifs. Comparisons were made manually (i.e. by eye). Discrepancies are marked with the gene name in red and additional details are given in Supplementary Table 2. Motifs in grey cells are those we believe are trustworthy. Green = consistent with previous data; Red = discrepancy; Yellow = new but with expected sequence features based on comparisons to proteins with similar DBD sequences; Blue = new and not similar to that of any previously-known motif.

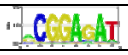
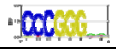
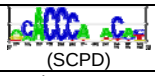
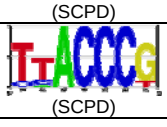
Rap1, Leu3, Pho2, Pho4, and Rox1 were analyzed successfully by Dip-chip. The full data is posted on our project website. As indicated below, the Rap1 and Leu3 motifs are derived from Dip-chip. Gzf3 was analyzed successfully using CSI. The full data is posted on our project website. All other motifs are from PBM data, for consistency, and in the "our motif" column the source of the motif is PBM unless otherwise indicated.

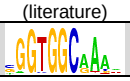
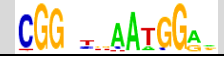
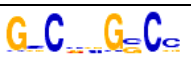
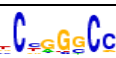
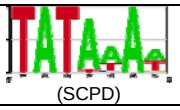
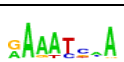
Gene (? Indicates proteins that may not be TFs)	Domain(s)	Our motif	Motif (Harbison)	Motif (MacIsaac)	Motif (SGD/Literature/SCPD)
ABF1	BAF/ABF1				 (SCPD)
ABF2 ?	HMG		Not found	Not found	None
ACA1	BZIP	Not found	Not found	Not found	Likely sim. To ATF/CREB TGACGTCA (Garcia-Gimeno 2000); see ACA2
ACE2	ZnF_C2H2				
ADA2 ?	SANT	Not found	Not found	Not found	
ADR1	ZnF_C2H2		None	 (literature)	GGRGK (Cheng 1994)
AFT1	AFT	Not found			PyPuCACCCPu (Yamaguchi-Iwai, 1996)
AFT2	AFT				
ARG80	MADS	Not found	Not found		De Rijcke 1992: CCTCTAAAGG (binds as heterotrimer with MCM1 and ARG81; responsive sequences are degenerate; see Yoon 2004)
ARG81	GAL4	Not found	Not found		Sim. To ARG80
ARO80	GAL4	Not found	Not found		At least three direct CCG repeats with 7-base spacing for ARO80-dependent element (Iraqui 1999)
ARR1	BZIP	Not found	Not found		One activation site maps to TTAATAAA (Wysocki 2004)  (Tan 2008)
ASG1	GAL4		Not found		None found
ASH1	ZnF_GATA	Not found	Not found	 (this is MCM1)	YTGAT (Maxon 2001)
AZF1	ZnF_C2H2			 (literature)	TTTTTCTT (Newcomb 2002)
BAS1	SANT	Not found			TGACTC (Rolfes 1997)
BDP1 ?	SANT	Not found	Not found	Not found	None found
BUR6 ?	CBFD_NFYB_HMF component	Not found	Not found	Not found	None found
CAD1	BZIP	Not found			TTACTAA (multiple refs)  (Tan 2008)

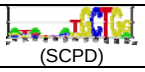
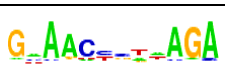
CAT8	GAL4		Not found	Not found	YCCNYTNRKCCG (Roth, 2004); CSRE: NCGGMTNAAHGGRN (Soontorngun 2007)  (SCPD)
CBF1	HLH				
CEF1	SANT	Not found	Not found	Not found	None
CEP3	GAL4		Not found	Not found	CDEIII element contains TTCGGAA (Espelin 1997)
CHA4	GAL4		YLR098C		UASCHA1 contains GCGGAAA; UASCHA2 contains GCGGAGA
CIN5	BZIP				TTACTAA (Fernandez 1997) (prob has flexible spacing)  (Tan 2008)
CRZ1	ZnF_C2H2		Not found	 (literature)	
CST6/ ACA2	BZIP		Not found		Likely sim. To ATF/CREB TGACGTCA (Garcia-Gimeno 2000); sim. To ACA1
CUP2	Copper fist	Not found	Not found	Not found	Binds an extensive palindrome with half-site GTCTTTPyPyGCTGAAC (Buchman, 1990)
CUP9	HOX		Not found	YPL177C	None
DAL80	ZnF_GATA			 (literature)	GATAA (Cunningham, 1994)
DAL81	GAL4	Not found	Not found	 (literature)	Not clear; does not seem to bind the published element
DAL82	None, but binds DNA				Binds DAL-UIS (Dorrington 1993) 
DOT6	SANT		Not found	None	None
ECM22	GAL4		Not found	 (literature)	Binds SRE (TCGTATA) (Vik 2001)
ECM23	ZnF_GATA		Not found	Not found	None found
ECM5 ?	BRIGHT	Not found	Not found	Not found	None found
EDS1	GAL4		Not found	None	None found
FHL1	Forkhead				None found
FKH1	Forkhead	Not found			
FKH2	Forkhead				
FZF1	ZnF_C2H2		Not found	None	Protects sequence CGTATCGTAT AAGGCAACAATAG (Avram 1999)
GAL4	GAL4	Not found			(SCPD)
GAT1	ZnF_GATA				
GAT2	ZnF_GATA	Not found	Not found	Not found	None found

GAT3	ZnF_GATA		Not found		None found
GAT4	ZnF_GATA		Not found	Not found	None found
GCN4	BZIP	Not found			 (SCPD)
GCR1	None, but binds DNA	Not found	Not found		Binds CTTCC (Baker 1991); high affinity to TTTCAG CTTCCTCTAT (Huie 1996)  (SCPD)
GIS1	ZnF_C2H2		Not found	Not found	CCCCT (Jang1999)
GLN3	ZnF_GATA			YER040W	
GZF3	ZnF_GATA	 (PBM and CSI)	YJL110C		
HAA1	Copper fist	Not found	Not found	None	None found
HAC1	BZIP		YFL031W		CAxxxTG (Mori et al., 1998)
HAL9	GAL4		Not found	None	None found
HAP1	GAL4				CGGNNNTAN CGGNNNTA (Ha, 1996)
HAP2	CBF	Not found	Not found		CCAAT-binding complex
HAP3	CBFD_NFYB_HMF component	Not found	Not found	CCAAT	CCAAT-binding complex
HAP4	CBFD_NFYB_HMF component	Not found			CCAAT-binding complex
HAP5	CBFD_NFYB_HMF component	Not found	Not found	CCAAT	CCAAT-binding complex
HCM1	Forkhead		Not found	Not found	None found
HMO1 ?	HMG	Not found	Not found	Not found	None found
HMRA2	HOX		Not found	Not found	None found, but identical to MatAlpha2/ Mata2; matches part of Dranganis 1990 MATALPHA2 binding sequence
HMS1	HLH	Not found	Not found	None	None found; Chua 2006 binds TCACGCAA
HMS2	HSF	Not found	Not found	Not found	None found
HSF1	HSF				recognizes variable heat shock elements (HSEs) consisting of inverted NGAAN repeats (see Yamamoto 2005)
HTA1 ?	CBFD_NFYB_HMF component	Not found	Not found	Not found	(Histone)
HTA2 ?	CBFD_NFYB_HMF component	Not found	Not found	Not found	(Histone)
INO2	HLH	Not found			5'-WYTTCAYR-TGS-3' (Schuller 1995)
INO4	HLH	Not found			5'-WYTTCAYR-TGS-3' (Schuller 1995)
ISW2 ?	SANT	Not found	Not found	Not found	None found
JJJ1 ?	ZnF_C2H2	Not found	Not found	Not found	None found
LEU3	GAL4	 (DIP-chip)			CCGGTACCGG, CCGNNNNCGG (Liu 2005)
LYS14	GAL4		Not found	Not found	TCCRNYGGA (Becker 1998)
MAC1	Copper fist	Not found	Not found		TTTGCTC (Labbe 1997)

MAL13	GAL4	Not found	Not found	Not found	None found
MAL33	GAL4	Not found	Not found	Not found	None found
MAL63	GAL4				MGC-N9-MGS (Sirenko 1995)
MATA1	HOX	Not found	Not found	 (literature)	ANNTACATCA (Dranginis, 1990)
MATALPHA1	HOX	Not found	Not found	Not found	None found
MATALPHA2	HOX	Not found	Not found	Not found	TCATGTNN(A/T) (Dranginis 1990)  (SCPD)
MBF1 ?	HTH-lambda	Not found	Not found	Not found	None found
MBP1	APSES				Well-known
MCM1	MADS	Not found			 (SCPD)
MET28	BZIP	Not found	Not found	CTGTGG (literature)	Not clear that it binds DNA specifically on its own (Kuras 1996)
MET31	ZnF_C2H2		Not found		AAACTGTGG (Blaiseau 1997)
MET32	ZnF_C2H2		Not found		AAACTGTGG (Blaiseau 1997)
MET4 ?	BZIP	Not found			Not clear that it binds DNA on its own; chip-chip motif is that of MET31/32
MGA1	HSF	Not found	Not found	Not found	None found
MIG1	ZnF_C2H2		Not found	 (literature)	Binds GCGGGG (Nehlin 1990)  (SCPD)
MIG2	ZnF_C2H2		Not found	Not found	Sim but not identical to MIG1 (Lutfiyya 1996)
MIG3	ZnF_C2H2		Not found	Not found	Expected, given Mig1/Mig2
MOT3	ZnF_C2H2	Not found	Not found	 (literature)	Most preferred sequence is CAGGCA (Grishin 1998)
MSN2	ZnF_C2H2				AAGGGG (Martinez-Pastor 1996)
MSN4	ZnF_C2H2		Not found		AAGGGG (Martinez-Pastor 1996)
NDT80	NDT80_PhoG	Not found	 (this is MCM1)	Not found	Binds variants of the MSE (VNDNCRCAAW) (Pierce 2003)
NHP10 ?	HMG		Not found	Not found	None found
NHP6A ?	HMG	Not found	Not found	Not found	None found
NHP6B ?	HMG	Not found	Not found	Not found	None found
NRG1	ZnF_C2H2	Not found			ACCC (Park, 1999); not really consistent with Harbison – no GGACCC in UAS
NRG2	ZnF_C2H2	Not found	Not found	Not found	None found
OAF1	GAL4		Not found	YAL051W	Binds part of ORE: CGGN3TN(A/R)N8–12CCG (Gurvitz 2006)
OPI1	OPI1	Not found	Not found		TCGAAYC (SGD cites Harbison)
ORC2 ?	AT_hook	Not found	Not found	Not found	None found

PDC2	HOX-related, CENPB	Not found	Not found	Not found	None found
PDR1	GAL4			 (literature)	CCGCGG (see Akache 2004)  (SCPD)
PDR3	GAL4	Not found	Not found	 (literature)	CCGCGG (see Akache 2004)  (SCPD)
PDR8	GAL4		Not found	Not found	TCCG(A/T/C)GGA (Hikkel, 2003)
PEP7 ?	ZnF_C2H2	Not found	Not found	Not found	None found
PHD1	APSES				None found
PHO2	HOX				
PHO4	HLH				
PIP2	GAL4	Not found	Not found	Not found	Binds part of ORE: CGGN3TN(A/R)N8–12CCG (Gurvitz 2006)
PPR1	GAL4	Not found	Not found	Not found	TTCGG-N6-CCGAA (Liang 1996)
PUT3	GAL4		Not found		CGG-N10-CCG (Siddiqui 1989)
PZF1	ZnF_C2H2	Not found	Not found	Not found	Binds to the 5S rRNA internal control regions; binding is apparently complicated (Rothfels 2007)
RAD18 ?	SAP	Not found	Not found	Not found	None found
RAP1	SANT	 (DIP-chip)			 (SCPD)
RDR1	GAL4		Not found	Not found	Putative consensus sequence from regulated promoters: TTCCGCGGAA (Hellauer 2002)
RDS1	GAL4				None found
RDS2	GAL4		Not found	Not found	CSRE: NCGGMTNAAHGGRN (Soontorngun 2007)  (SCPD)
REB1	SANT				 (SCPD)
REH1	Zfp622	Not found	Not found	Not found	None found
REI1	ZnF_C2H2		Not found	Not found	None found
RFX1	RFX				TCGCCATGGCAAC (Zaim 2005)
RGM1	ZnF_C2H2		Not found	Not found	None found
RGT1	GAL4		Not found		CGGANNA (Kim 2003)
RIM101	ZnF_C2H2		Not found	(literature)	TGCCAAG (Lamb 2003)
RLM1	MADS	Not found	Not found		SCPD (Dodou 1997)

RME1	ZnF_C2H2	Not found	Not found		Nuclease footprint is AAAAGAACCTCAAAAAGT CCA
ROX1	HMG	 (PBM) (Dip-chip)	Not found		 (SCPD)
RPH1	ZnF_C2H2		Not found	(literature)	CCCCTTAAGG (Jang 1999)
RPN4	ZnF_C2H2				GGTGGCAAA (Mannhaupt 1999)
RSC3	GAL4		Not found	Not found	None found
RSC30	GAL4		Not found	Not found	None found
RSC8 ?	SANT	Not found	Not found	Not found	None found
RTG1	HLH	Not found	Not found	Not found	RTG1/3 bind the R box (UASr) GTCAC (Rothermel 1997) Or GGTCAC (Jia 1997)
RTG3	HLH	Not found	Not found		RTG1/3 bind the R box (UASr) GTCAC (Rothermel 1997) Or GGTCAC (Jia 1997)
RTS2	ZnF_C2H2	Not found	Not found	Not found	None found
SEF1	GAL4	Not found	Not found	Not found	None found
SFL1	HSF	Not found	Not found	 (literature)	AGAAAxxT-n-GTTCTT (Conlan and Tzamarias 2001)
SFP1	ZnF_C2H2	Not found			None found
SIG1	RRM, but binds DNA			Not found	None found
SIP4	GAL4				TCCATTSRTCCGR (Roth, 2004)
SIZ1 ?	SAP and zf-MIZ	Not found	Not found	Not found	None found
SKN7	HSF				ATTTGGCYGGSCC (Li 2002)
SKO1	BZIP	Not found	Not found		TGACGTCA (Nehlin 1992)
SMP1	MADS	Not found	Not found	 (literature)	 SCPD (Dodou 1997)
SNF2 ?	AT_hook	Not found	Not found	Not found	None found
SNT1	SANT	Not found	Not found	Not found	None found
SNT2	SANT	Not found			None found
SOK2	APSES				None found
SPT15	TBP	Not found	Not found	Not found	 (SCPD)
SPT21	None, but binds DNA	Not found	Not found	Not found	None found
SPT23	IPT	Not found			None found
SRD1	ZnF_GATA		Not found	Not found	None found
STB4	GAL4				None found

STB5	GAL4				CGGNSNTA (Larochelle 2006)
STE12	STE				 (SCPD)
STP1	ZnF_C2H2	Not found	Not found		RCGGCNNNRCGGC (Nielsen 2001)
STP2	ZnF_C2H2	Not found	Not found	Not found	CGGCTC (de Boer, 2000); various responsive genes contain CGGCNxCGGC (Abdel-Sater 2004)
STP3	ZnF_C2H2		Not found	Not found	None found
STP4	ZnF_C2H2		Not found		None found
SUM1	AT_hook				DSYGWCAYWDW (Pierce 2003)
SUT1	GAL4	Not found			CGCG (Regnacq, 2001)
SUT2	GAL4	Not found	Not found	Not found	None found
SWC4 ?	Sant/Myb/HD-like	Not found	Not found	Not found	None found
SWI4	APSES				Well-known…
SWI5	AT_hook and ZnF_C2H2(2)		Not found		 (SCPD)
TAF3 ?	BTP	Not found	Not found	Not found	None found
TBF1	SANT		Not found	Not found	TTAGGG (Brigati 1993); TAGGGTTGG (Koering 2000)
TBS1	GAL4		Not found	Not found	None found
TEA1	GAL4		Not found	Not found	CGG-N10-CCG (Gray 1996)
TEC1	TEA				CATTCC (Madhani 1997); also same as human TEAD
TFC6 - ?	AT_hook	Not found	Not found	Not found	Together with Tfc3 binds BoxB promoter sites of tRNA and other genes
THI2	GAL4	Not found			None found, although Mojzita 2006 refer to “target genes”
THO1 ?	SAP	Not found	Not found	Not found	None found
TOS8	HOX		Not found	Not found	None found
TYE7	HLH				TCTGGCACACA (Sato 1999)
UGA3	GAL4		Not found	 (literature)	AAAARCCGCSGGCGGSA WT (Talibi 1995), CCGCSSGCGG (Noel 1998), SGCGGNWttt (Idicula 2002)
UME6	GAL4				TCGGCGGCT (Williams 2002)
UPC2	GAL4	Not found	Not found	Not found	Binds SRE (TCGTATA) (Vik 2001)
WAR1	GAL4	Not found	Not found	Not found	CGG-N23-CCG (Kren 2003)
XBP1	APSES		Not found		
YAP1	BZIP	Not found			TTASTMA (Nguyen, 2001) TGACTCA, TTACTAA (Fernandes 1997) (Tan 2008)

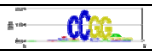
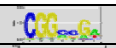
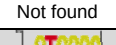
YAP3	BZIP		Not found	 (literature)	TGACTCA, TTACTAA (Fernandes 1997)
YAP5	BZIP	Not found	Not found		Li et al. 2008 claim it binds YAP consensus in CCC1 promoter but do not show data  (Tan 2008)
YAP6	BZIP	Not found	Not found		 (Tan 2008)
YAP7	BZIP	Not found			 (Tan 2008)
YBL054W	SANT		Not found	Not found	None found
YBR239C	GAL4		Not found	Not found	None found
YDR026C	SANT	Not found			None found
YDR049W	Zfp622	Not found	Not found	Not found	None found
YDR520C	GAL4		Not found		None found
YER130C	ZnF_C2H2		Not found	Not found	None found
YER184C	GAL4		Not found	Not found	None found
YFL044C ?	ZnF_C2H2	Not found	Not found	Not found	None found
YFL052W	GAL4	Not found	Not found	Not found	None found
YGR067C	ZnF_C2H2		Not found	Not found	None found
YGR071C	ZnF_BED	Not found	Not found	Not found	None found
YHP1	HOX	Not found	Not found	 (literature)	TAATTG (Kunoh 2000)
YJL103C	GAL4		Not found	Not found	Binds CGGN8CGG and CGGN9CGG (Ho 2006)
YJL206C	GAL4	Not found	Not found	Not found	None found
YKL222C	GAL4		Not found	Not found	None found
YKR064W	GAL4	Not found	Not found	Not found	None found
YLL054C	GAL4		Not found	Not found	None found
YLR278C	GAL4		Not found	Not found	None found
YML081W	ZnF_C2H2		Not found		None found
YNR063W	GAL4		Not found	Not found	None found
YOX1	HOX		Not found		YAATTA (Pramila 2002)
YPL230W	ZnF_C2H2		Not found	Not found	None found
YPR013C	ZnF_C2H2		Not found	Not found	None found
YPR015C	ZnF_C2H2	Not found	Not found	Not found	None found
YPR022C	ZnF_C2H2		Not found	Not found	None found
YPR196W	GAL4		Not found	Not found	None found
YRM1	GAL4		Not found	Not found	None found
YRR1	GAL4		Not found		WCCGYKKWW (Le Crom, 2002)
ZAP1	ZnF_C2H2	Not found			ACCYYNAAGGT (Zhao 1998)
ZMS1	ZnF_C2H2		Not found	Not found	None found

Table S2. Discrepancies between our motifs and known motifs (excluding GAL4-class)

Protein	Domain type(s)	Motif obtained	Previous motif(s)	Comments
SUM1	AT_hook (2)		AGYGWCACAAAAK, GYGWCASWAAW (SGD),  (MacIsaac)	Our motif is for the AT hook domain; the literature motif is from the C-terminus
FHL1	Forkhead		 (MacIsaac)	Human FoxN1 (Schlake et al., 1997): 
SOK2	SANT		 (MacIsaac)	Our motifs for paralogs SOK2 and PHD1 are very similar; the left part of our motifs resembles the reverse complement of the left part of the MacIsaac motif
PHD1	SANT		 (MacIsaac)	See SOK2 above
SIG1	RRM		 (Harbison)	
STP4	ZnF_C2H2		 (MacIsaac)	Our motif matches our Stp3 motif
YML081W	ZnF_C2H2		 (MacIsaac)	Our motif matches our motif for Zms1
GAT3	ZnF_GATA		 (MacIsaac)	Our motif matches our motifs for Ecm23, Srd1, and Gat4

Table S3. Yeast strains used in this study.

Name	Genotype	Source	Original Publication
abf1-101	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 abf1-101::KanMX*	Charlie Boone[a,b]	Loo et al., 1995
cep3-1	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 cep3-1::KanMX*	Charlie Boone[a,c]	Strunnikov et al., 1998
mcm1	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 KanMX::mcm1-URA3*	Phil Hieter	Ben-Aroya et al., 2008
rap1-1	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 rap1-1::KanMX*	Charlie Boone[a,d]	Conrad et al., 1990
reb1-212	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 KanMX::reb1-212-URA3*	Phil Hieter	Ben-Aroya et al., 2008
rsc3-1	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 rsc3-1::KanMX*	Charlie Boone[a,e]	Angus-Hill et al., 2001
tbf1	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 KanMX::tbf1-URA3*	Phil Hieter	Ben-Aroya et al., 2008
rsc3-1 Rsc8-TAP	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 rsc3-1::kanMX rsc8-TAP::HIS3MX6*	This study	-
Rsc8-TAP	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 rsc8-TAP::HIS3MX6*	Open Biosystems	Ghaemmaghami, 2003

[a] Reconstructed in BY4741 marked with KanMX
[b] Original source - Jasper Rine
[c] Original source - Douglas Koshland
[d] Original source - Virginia Zakian
[e] Original source - Bradley Cairns

Angus-Hill, ML., Schlichter, A., Roberts, D., Erdjument-Bromage, H., Tempst, P., Cairns, BR. (2001) A Rsc3/Rsc30 zinc cluster dimer reveals novel roles for the chromatin remodeler RSC in gene expression and cell cycle control. Mol Cell. *7*, 741-51.

Ben-Aroya, S., Coombes, C., Kwok, T., O'Donnell, KA., Boeke, JD., Hieter, P. (2008) Toward a comprehensive temperature-sensitive mutant repository of the essential genes of *Saccharomyces cerevisiae*. Mol Cell. *30*, 248-58.

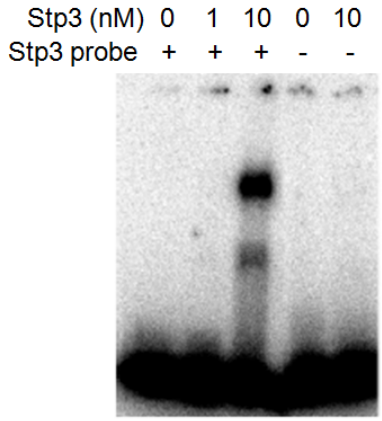
Conrad, MN., Wright, JH., Wolf, AJ., and Zakian, VA. (1990) RAP1 Protein Interacts with Yeast Telomeres In Vivo: Overproduction Alters Telomere Structure and Decreases Chromosome Stability. Cell. *63*,739-750.

Loo, S., Laurenson, P., Foss, M., Dillen., and Rine, J. (1995) Role of ABF1, NPL3, and YCL54 in Silencing in *Saccharomyces cerevisiae.* Genetics. *141*, 889-902.

Strunnikov, AV., Kingsbury, J., Koshland, D. (1998) CEP3 encodes a centromere protein of *Saccharomyces cerevisiae*. J Cell Biol. *128*, 749-60.

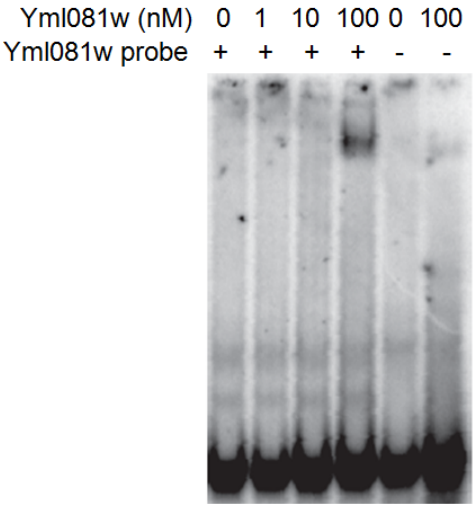
Ghaemmaghami, S., Huh, W.K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., Weissman, J.S. (2003). Global analysis of protein expression in yeast. Nature. *425*, 737-41.

Figure S1. Gel-shift confirmation of motifs that disagree with motifs in the literature. Underlined blue segment shows expected binding sequence, designed to represent the single sequence most closely matching the motif for each protein.



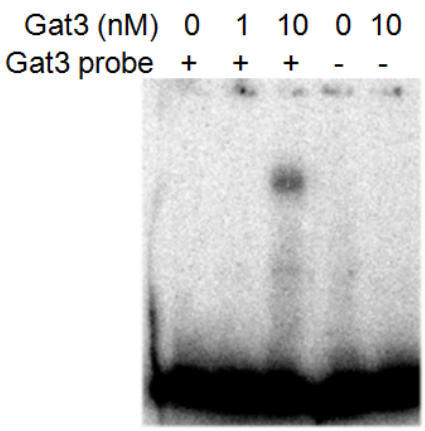
Stp3 probe + GCGTTGATGCGCTAGCGTCGGAC

Stp3 probe - GCGTTGACGCCGTGTCACGCGCGTCGGAC



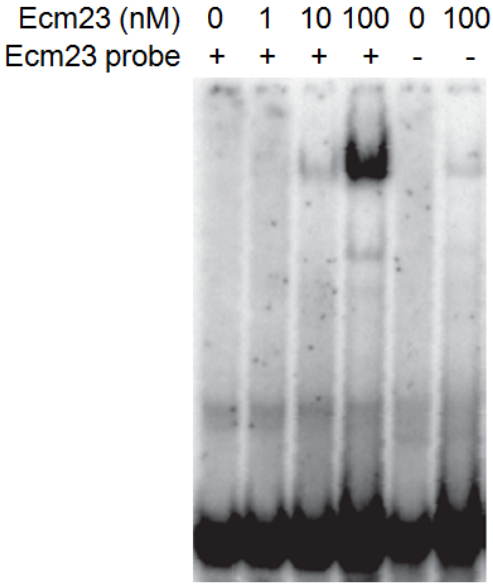
Yml081w probe + GTTCTTCGCCCCGCACGTTCTTCG

Yml081w probe - GTTCTTCCCAGTCTGAAGTTCTTC



Gat3 probe + CGCTTGATAGATCTAGTCGGAC

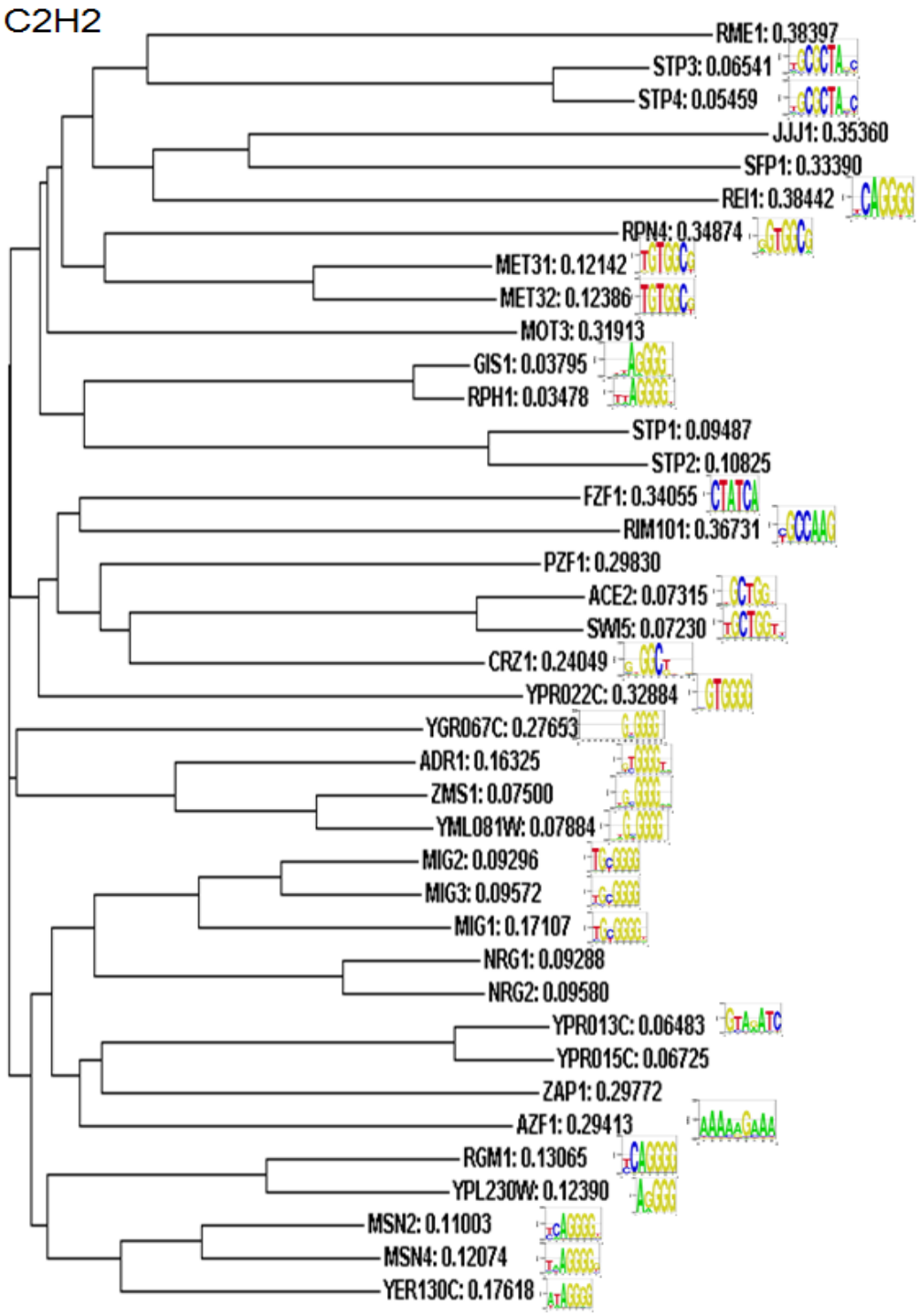
Gat3 probe - CGCTTGAATAACATGGTCGGAC



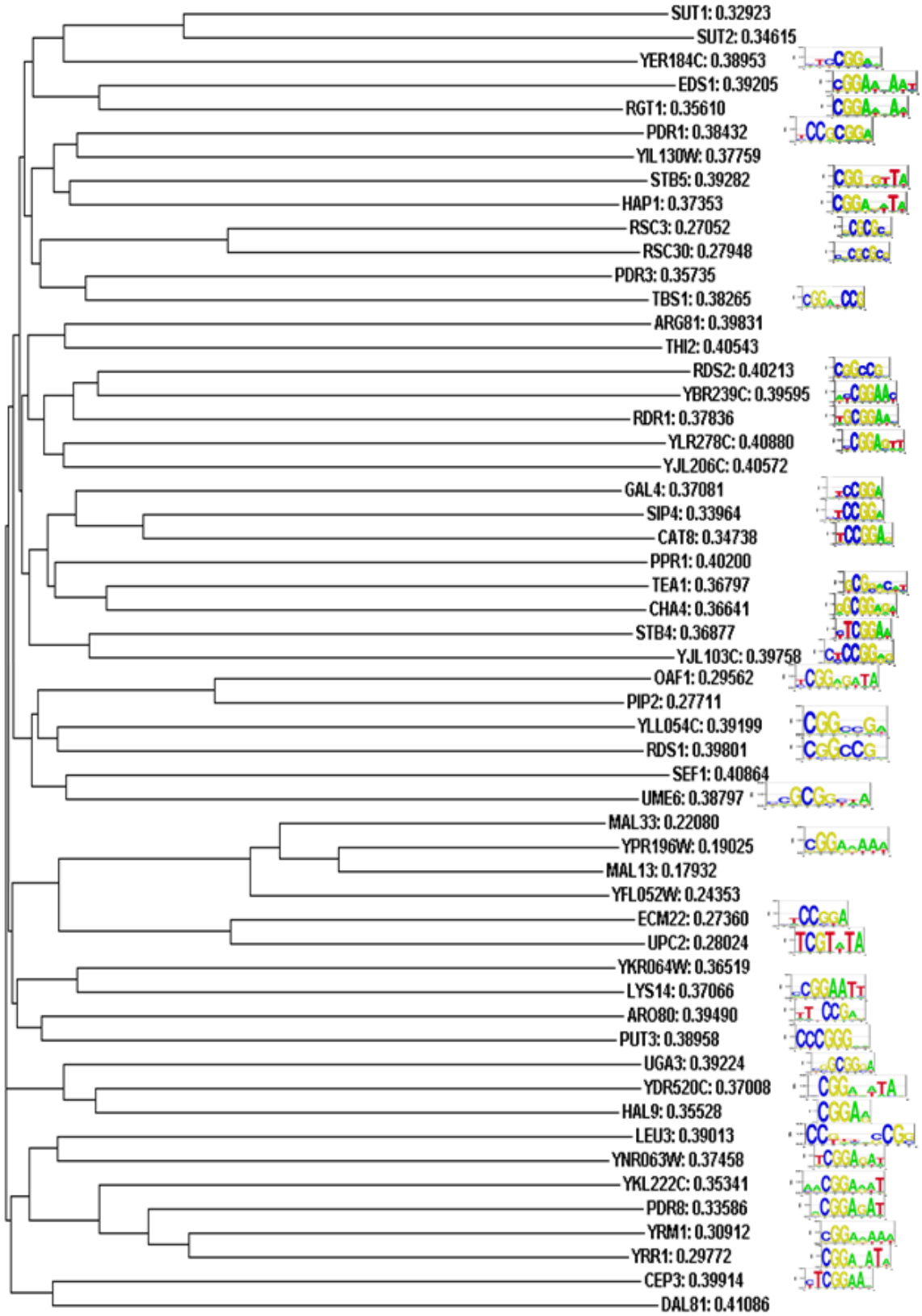
Ecm23 probe + GGTGGTTTAAGATCTTGGTGGTTT

Ecm23 probe - GTTCTTCGGTTCTTCGGTTCTTCG

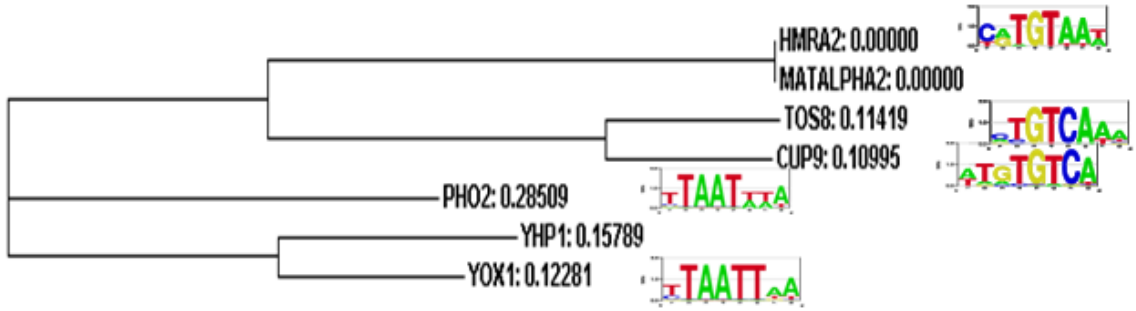
Figure S2 (this and following pages). Clustalw phylograms showing that proteins with similar DBD sequences display related sequence specificities.



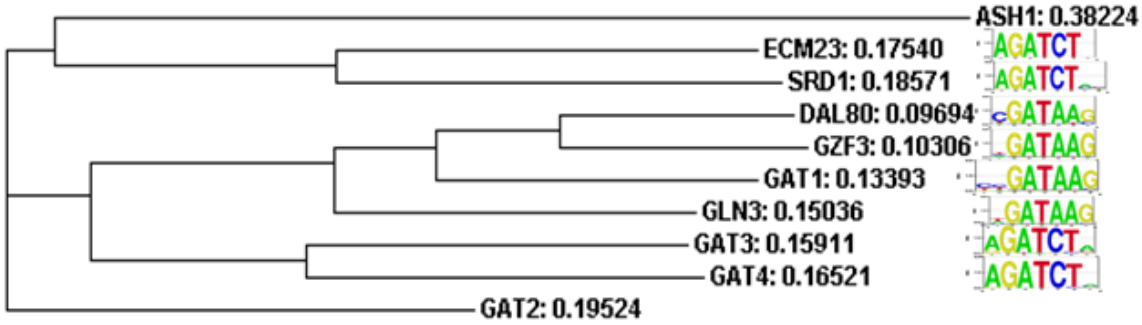
GAL4 zinc fingers



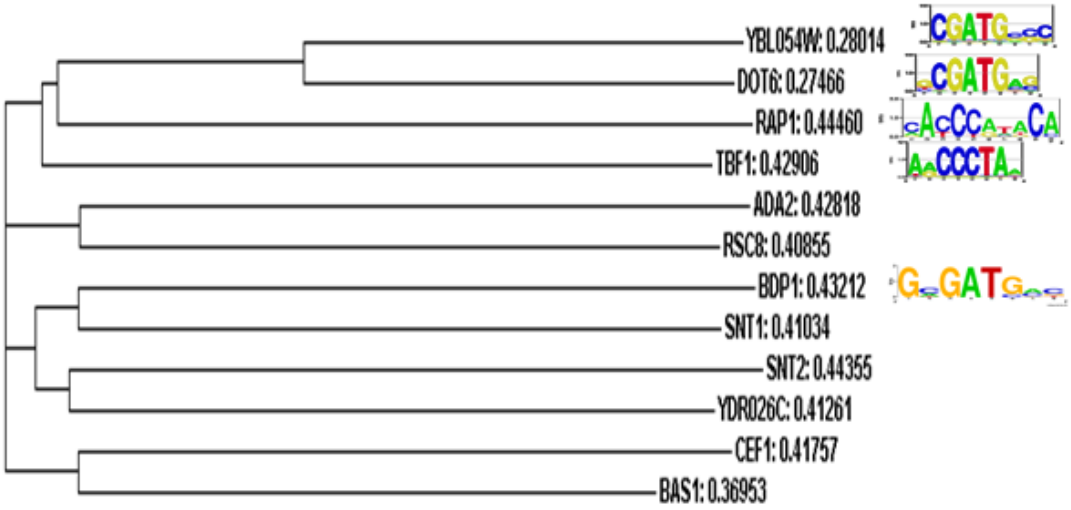
Homeodomain



GATA

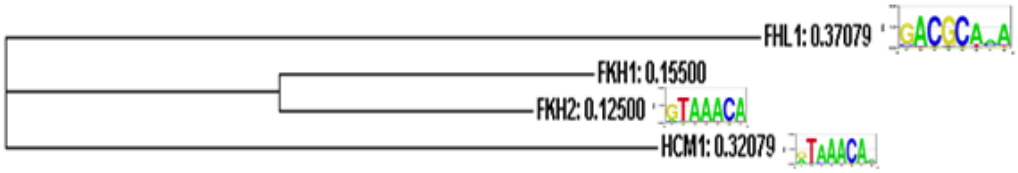


SANT

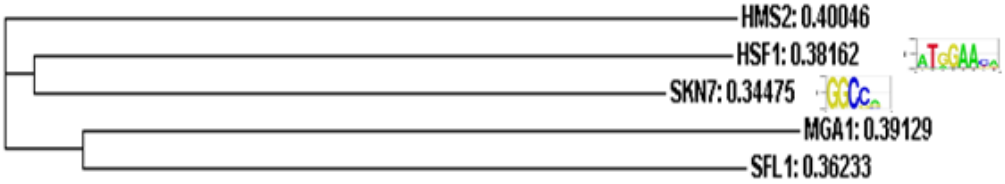


APSES

Forkhead



HSF



HMG

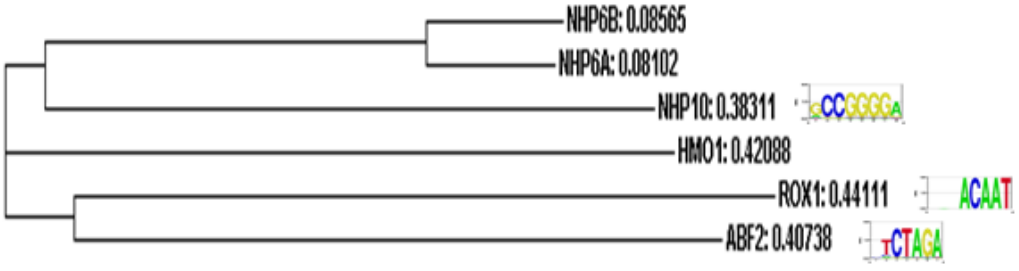
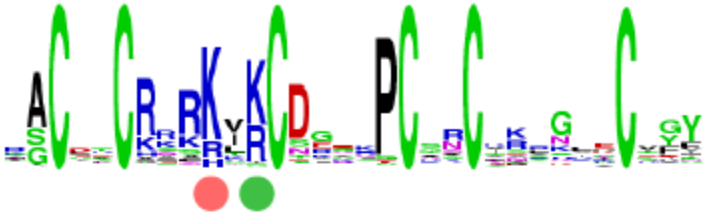
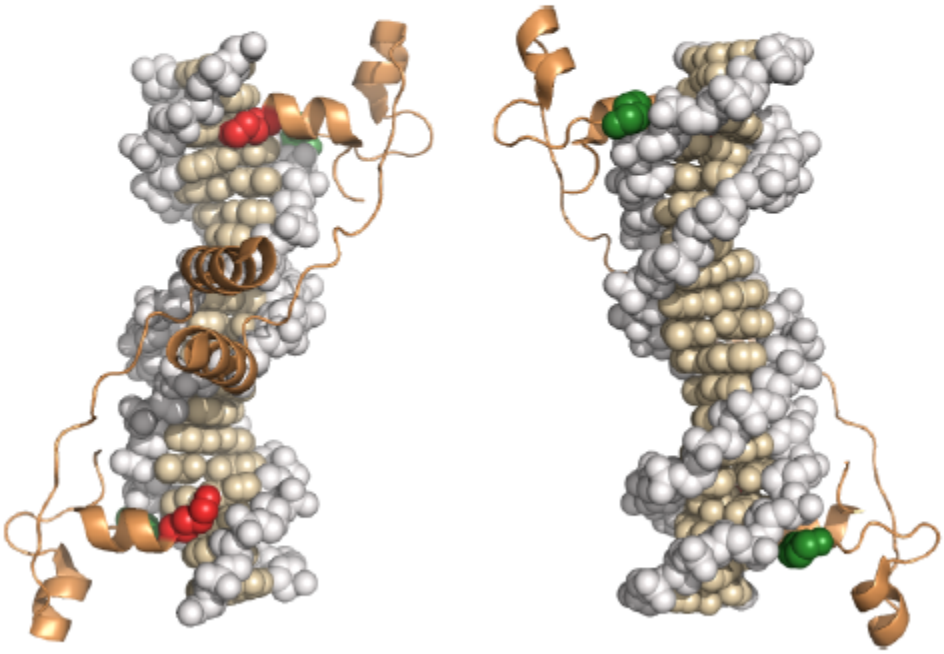


Figure S3. (A) Sequence logo for the zinc cluster family of transcription factors. The residue corresponding to Gal4 K18 is indicated by a red circle; the residue corresponding to Gal4 K20 is indicated by a green circle. (B) Two views of a Gal4-DNA complex. The sidechain for K18 is shown in red, and the sidechain for K20 is shown in green. K18 makes major groove base-specific contacts and is thought to be the main specificity determinant for the canonical Gal4-family half-site. K20 is in close proximity to the backbone, and is replaced by G in Rsc3 and Rsc30.

A



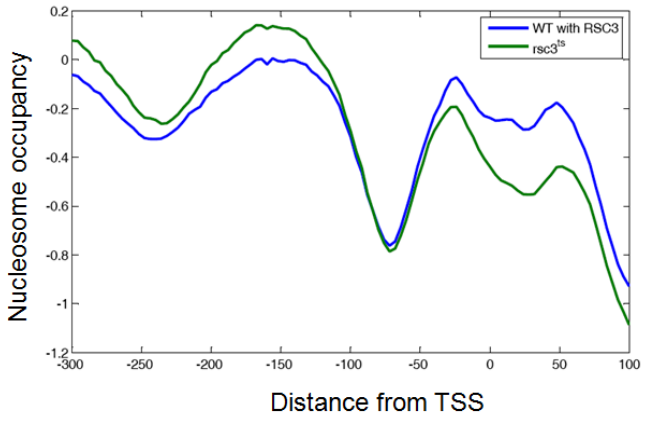
B



Gal4 (PDB:1D66)
front

Gal4 (PDB:1D66)
back

Figure S4. Impact of the *rsc3-1* mutation on nucleosome occupancy and Rsc8 occupancy at tRNAs. Top, average nucleosome occupancy profile over 275 tRNA genes. Bottom, profiles for individual tRNA genes: nucleosome occupancy (left) and Rsc8 occupancy (right). Color scale as in Figures 5 and 6.



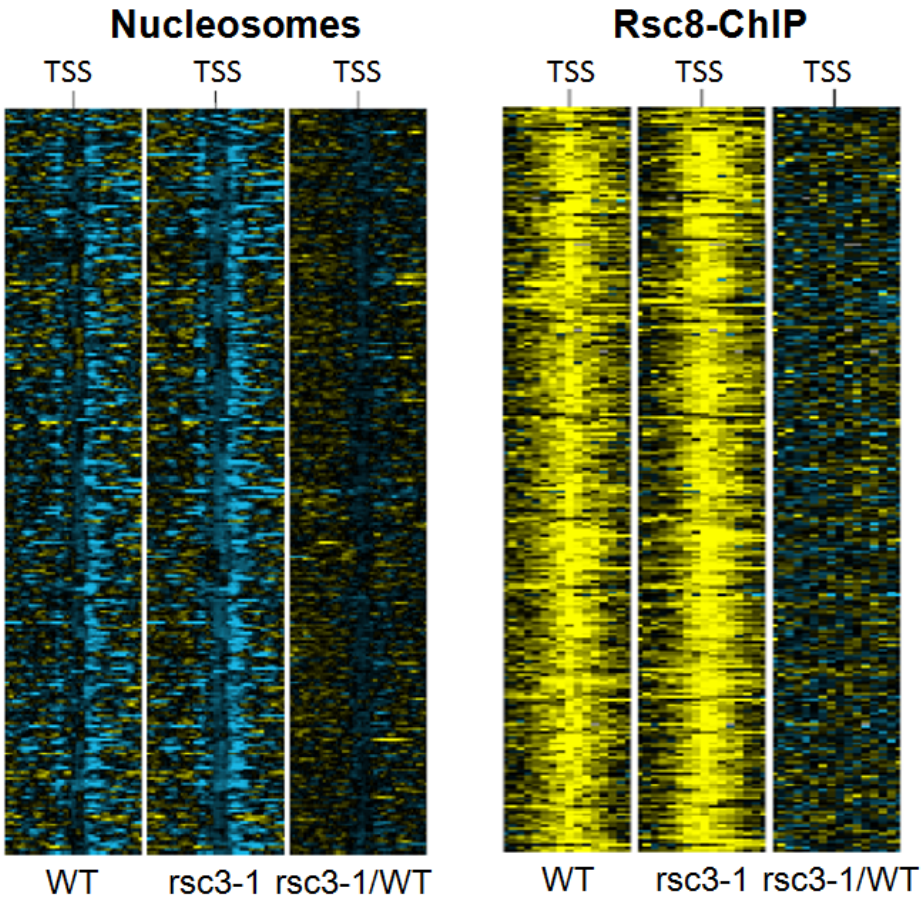
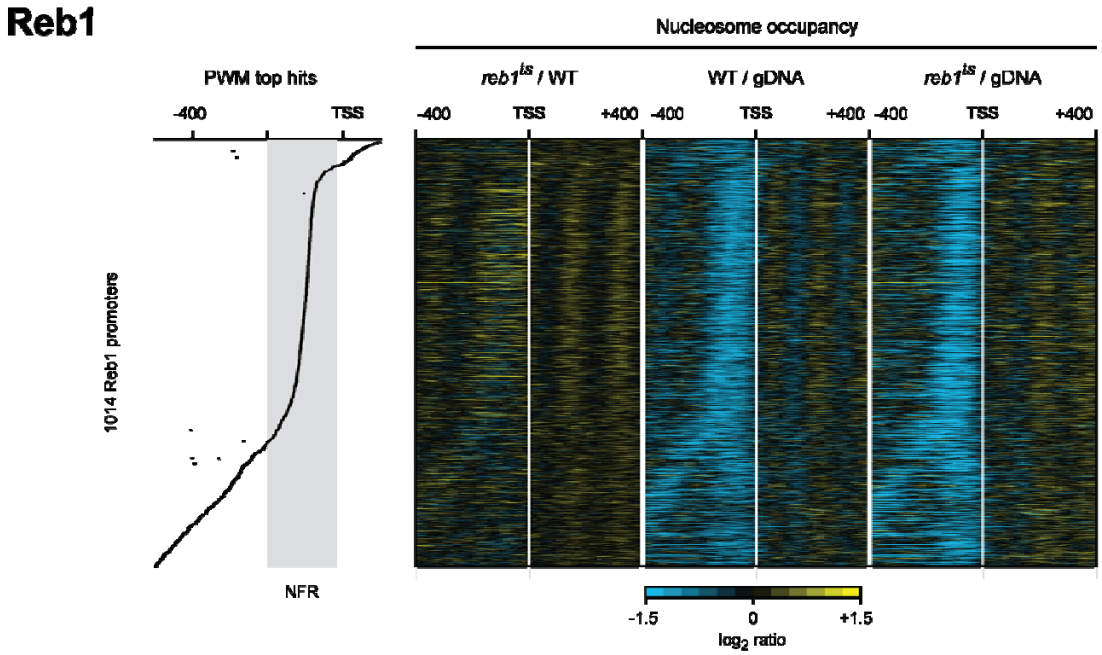
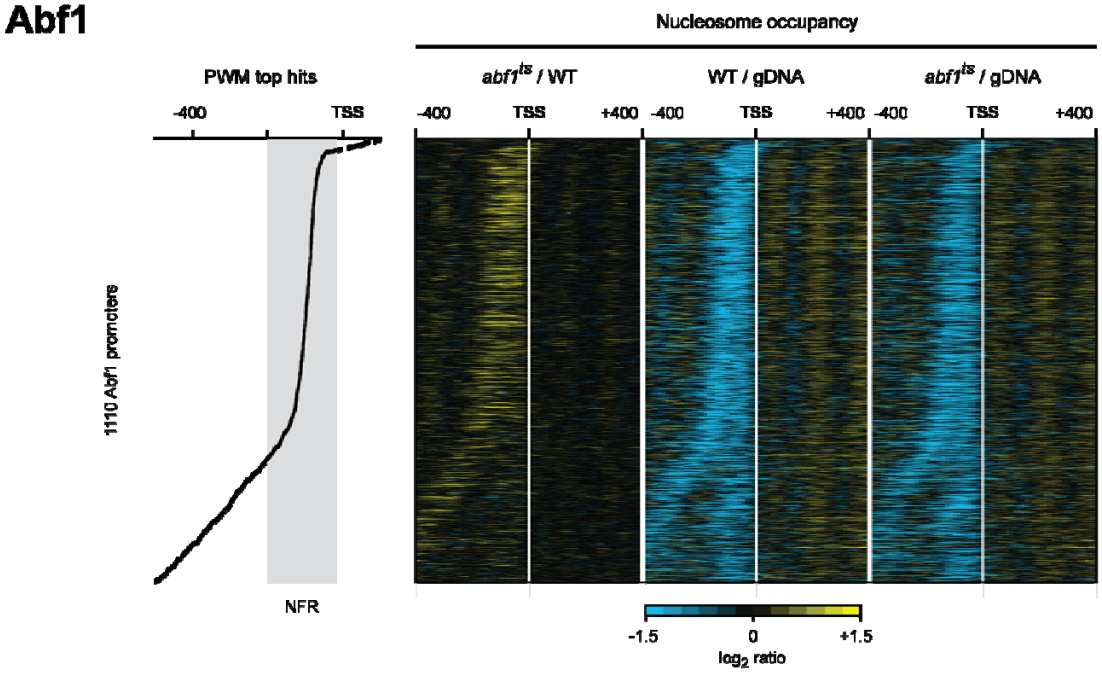
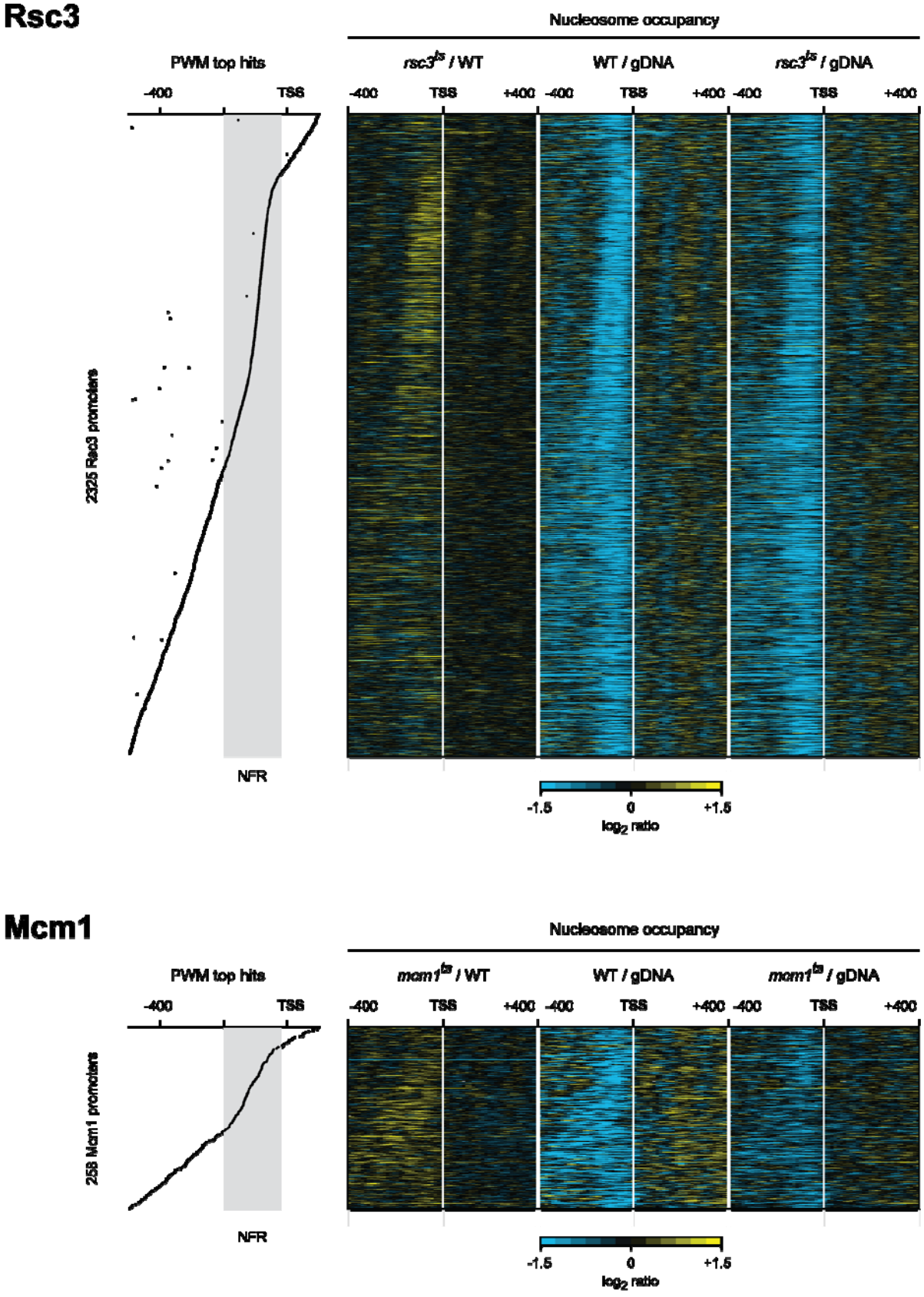


Figure S5 (this and following pages). Nucleosome profiles for mutants in essential TFs with known binding sequences, sorted by location of binding sites in promoters. These figures were generated in the same manner as those in Figure 5.






Rsc3
Nucleosome occupancy
PWM top hits
rsc3ts / WT
WT / gDNA
rsc3ts / gDNA
-400
TSS
+400
2325 Rsc3 promoters
NFR
-1.5
0
+1.5
log2 ratio
Mcm1
Nucleosome occupancy
PWM top hits
mcm1ts / WT
WT / gDNA
mcm1ts / gDNA
258 Mcm1 promoters
NFR
log2 ratio

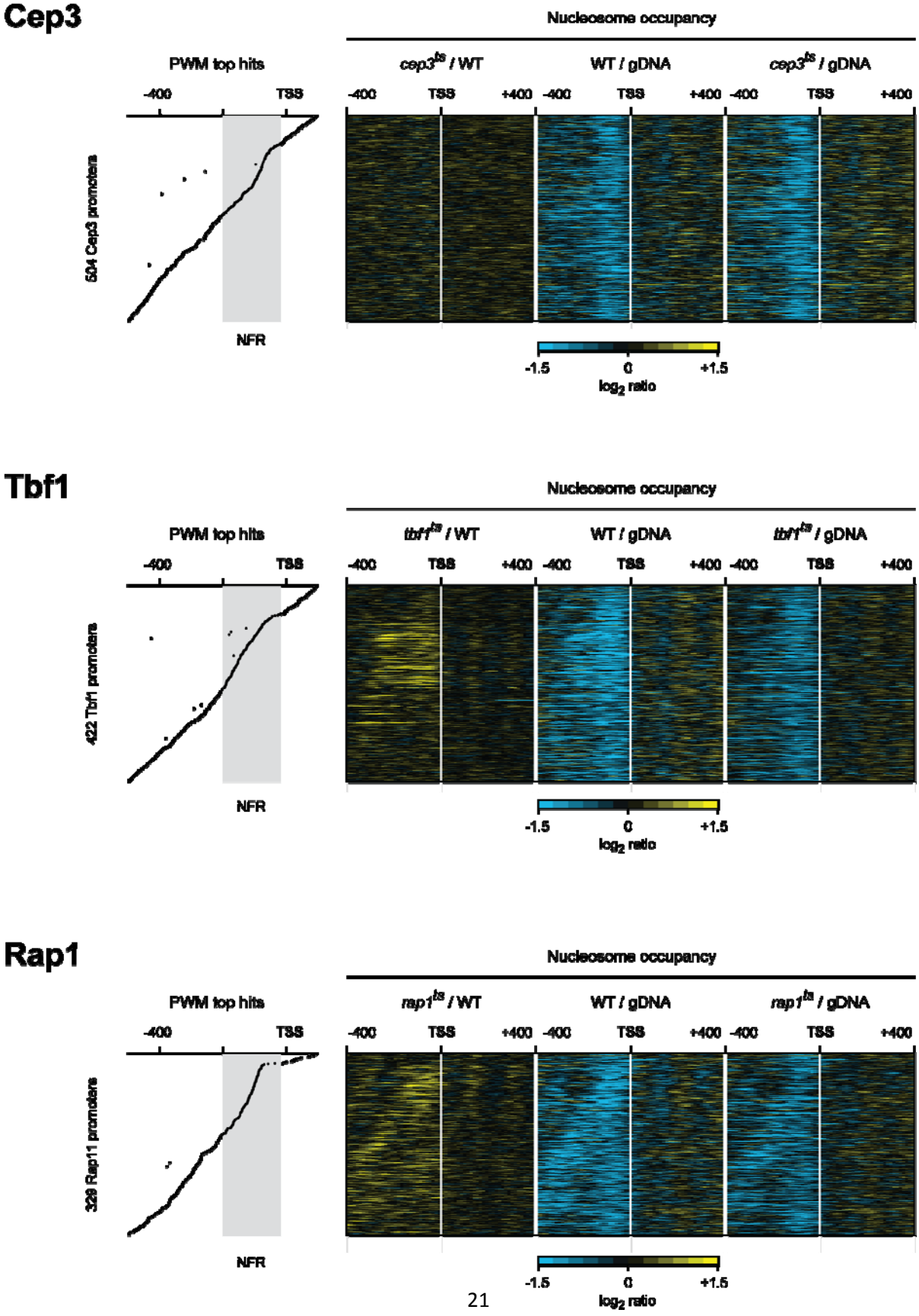
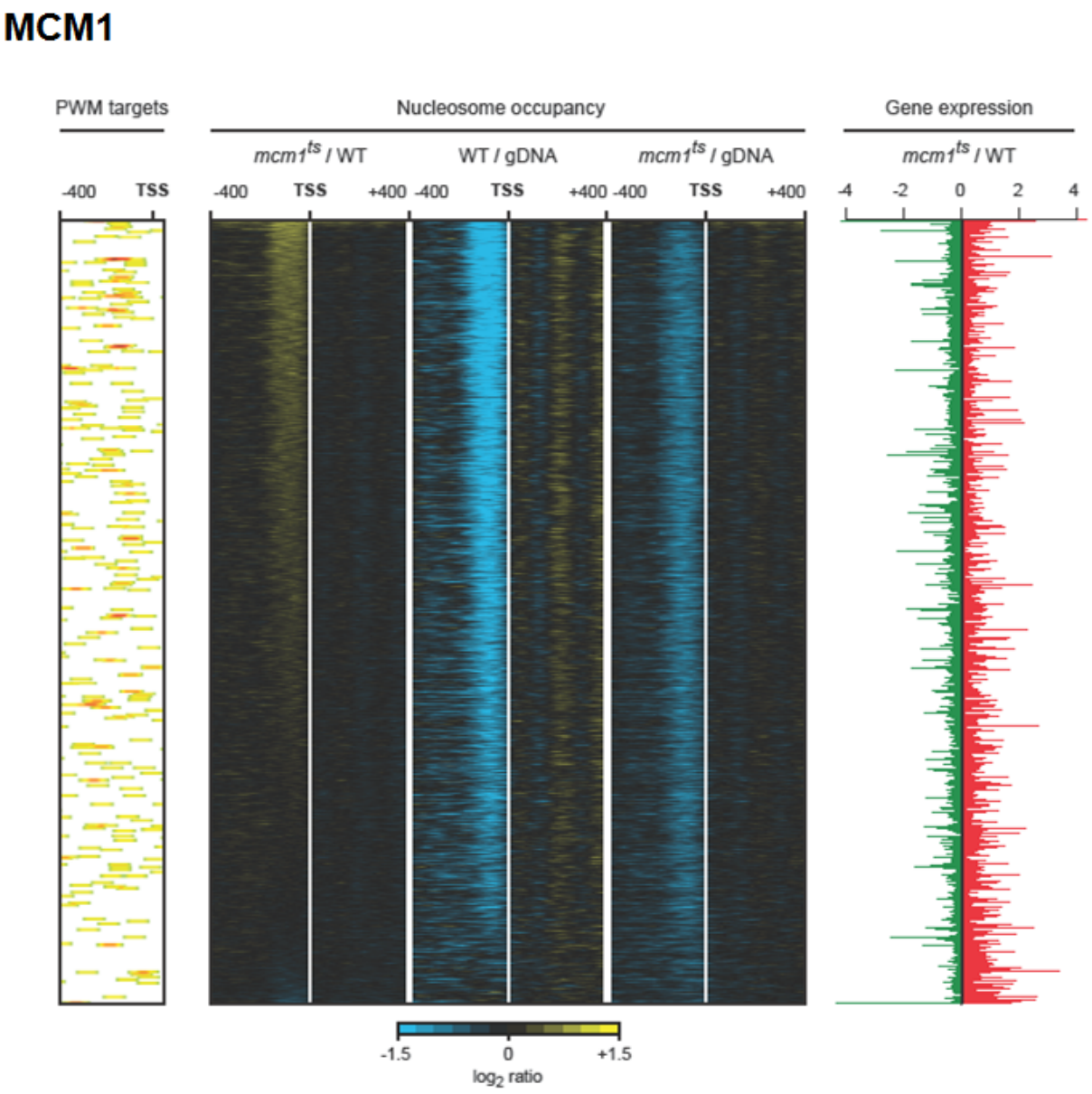
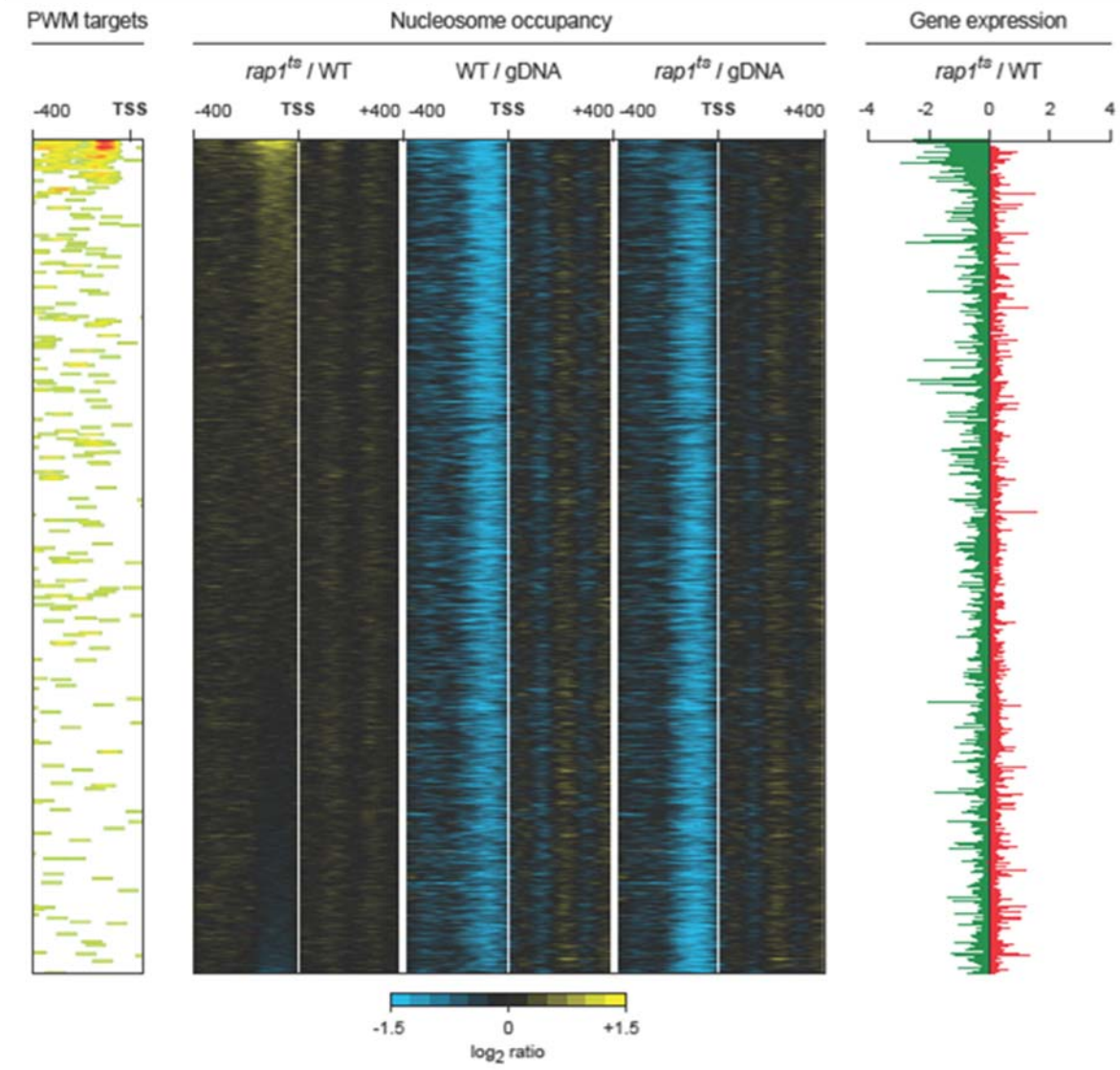

Cep3
Nucleosome occupancy
PWM top hits
cep3ts / WT
WT / gDNA
cep3ts / gDNA
-400
TSS
+400
504 Cep3 promoters
NFR
-1.5
0
+1.5
log2 ratio
Tbf1
Nucleosome occupancy
PWM top hits
tbf1ts / WT
WT / gDNA
tbf1ts / gDNA
422 Tbf1 promoters
NFR
Rap1
Nucleosome occupancy
PWM top hits
rap1ts / WT
WT / gDNA
rap1ts / gDNA
329 Rap11 promoters
NFR

Figure S6 (this and following pages). Nucleosome profiles for mutants in essential TFs with known binding sequences, sorted by change in promoter nucleosome occupancy in the mutant. These figures were generated in the same manner as those in Figure 6 of the main text.



RAP1


PWM targets
Nucleosome occupancy
Gene expression
rap1ts / WT
WT / gDNA
rap1ts / gDNA
rap1ts / WT
-400
TSS
+400
-4
-2
0
2
4
-1.5
0
+1.5
log2 ratio

TBF1

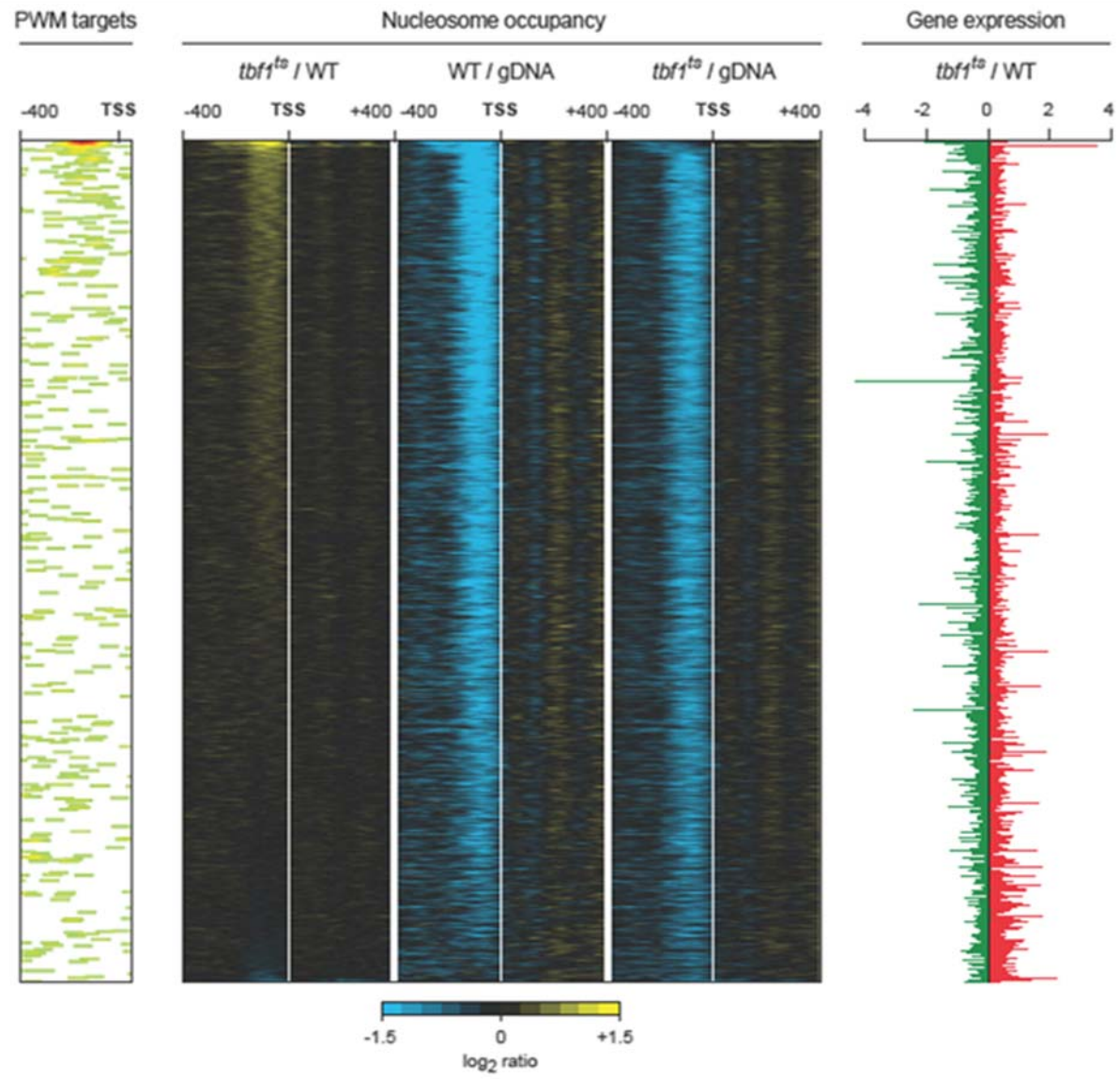
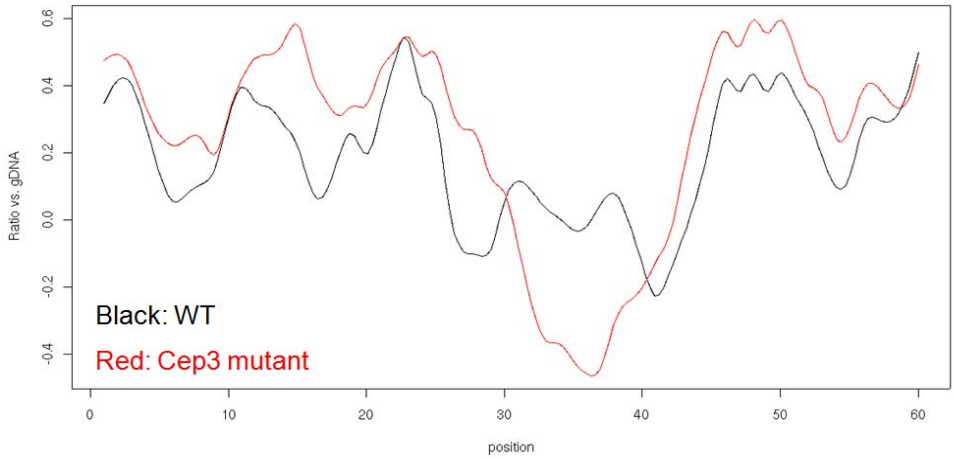
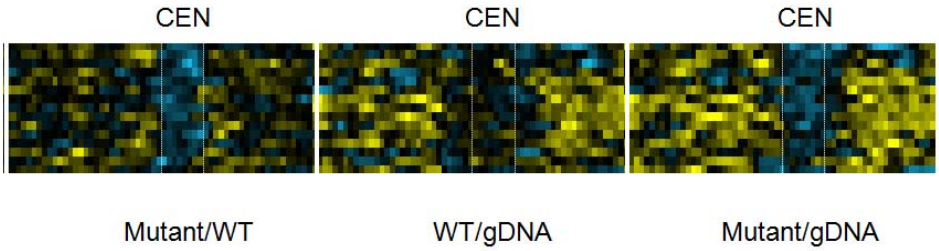

PWM targets
Nucleosome occupancy
Gene expression
tbf1ts / WT
WT / gDNA
tbf1ts / gDNA
tbf1ts / WT
-400
TSS
+400
-4
-2
0
2
4
-1.5
0
+1.5
log2 ratio

Figure S7. Effect of the *cep3-1* mutation on yeast centromeres. Top, average profile over all centromeres; Middle, data for individual centromeres; Bottom, location of match to the Cep3 motif is found in the invariant region of CPEIII of all sixteen yeast chromosomes.






```
CEN10          ATGTTTATGATTTCCGAACCTAAATA 26
CEN11          ATGTTCATGATTTCCGAACGTATAAA 26
CEN13          ATGTGTATGCGTTCCGAACTTTAAAT 26
CEN15          ATGTATATGACTTCCGAAAATATAT  26
CEN4           TTGTTTATGATTACCGAAACATAAAA 26
CEN7           TTGTTTTTGCCTTCCGAAAGAAAAT  26
CEN3           GTGTATTTGATTTCCGAAAGTTAAAA 26
CEN5           CAGTATTAGATTTCCGAAAAGAAAAA 26
CEN14          ATGTATTTGTCTTCCGAAAAGTAAAA 26
CEN12          TTGTATTTGTTATCCGAACAATAAAA 26
CEN2           ATGTTTTTGTTTTCCGAAAAAGAAAA 26
CEN1           ATGTTTTTGTTTTCCGAAGCAGTCAA 26
CEN6           TAGTTTTTGTTTTCCGAAGATGTAAA 26
CEN9           ATGGTTTTGTTTTCCGAAATGTTTTT 26
CEN8           TGGGTTTTGTGTTCCGAACTTAGAAA 26
CEN16          TTGGTTAAGATTTCCGAAAATAGAAA 26
                 *      *    *****
```

Supplemental Experimental Procedures

PBM analyses. PBM arrays and assays were as described (Berger et al., 2006; Mintseris and Eisen, 2006). We relied on two different transformations of the PBM data to estimate relative preference for each of 32,896 8-mers. First, we took the median signal intensity across the array from the 32 spots containing each 8-mer and expressed this as a Z-score. Second, we calculated an "E-score" value (for Enrichment) for each 8-mer (Berger et al., 2006). The E-score (for enrichment) is a variation on AUC (Area under the ROC curve), in which the value represents the relative ability of an 8-mer sequence to predict the rank order of the 35-mer intensities. Previous analyses have established that E-scores above 0.45 can generally be taken as a success criterion for PBM experiments (Berger et al., 2008). Using a cross-validation regime, highly-bound 35-mers and 8-mers for each experiment (defined by the inflection point in sorted Z-scores, or 0.45 for E-scores) were selected from the distribution and input into a panel of motif-finding tools, for which GOMER (Granek and Clarke, 2005) scores for all possible 8-mers and 35-mer probe sequences were calculated against all motifs, and those with the highest correlation to the input data were retained. We considered an experiment successful if (a) it contained 8-mers with E-scores above 0.45, and (b) a motif could be obtained from either the 35-mers or 8-mers in which the PWM scores of 8-mers scale with the original Z-transformed data. Only one TF (Abf1), containing a large gapped binding site, yielded a motif that predicted 35-mer scores while failing to predict 8-mer scores; hence, our primary criterion was ability to predict 8-mer scores. Additional details will be described elsewhere (Chan, Peña-Castillo et al., in preparation).

CSI analyses. CSI methods essentially followed (Warren et al., 2006). Double-stranded hairpin microarrays were synthesized by NimbleGen. Hairpins were induced by 30min incubation at

65°C with 7M urea in 1x PBS (13mM NaCl, 2.7mM KCl, 10.1mM Na2HPO4, 1.8mM KH2PO4, pH 7.4) shaking every 10min, followed by a 15min incubation at 65°C in 1xPBS. The induction step was completed with a 5min wash in NimbleGen's Non Stringent Wash Buffer (6x Saline-Sodium Phosphate-EDTA with 0.01% v/v Tween-20). Hairpinned arrays were washed one time with 1xPBS then blocked for one hour with 2% w/v non-fat dried milk. Blocking solution was washed away with 1xPBS and the array was then incubated with the TF (100nM GST-TF, 2% non-fat dried milk, 51.3ng/μL salmon sperm DNA, 0.2μg/μL bovine serum albumin, and 50μM zinc acetate) for one hour. Arrays were washed with PBS then bound TF was detected by 0.05mg/mL anti-GST Alexa Fluor 488 antibody in 50μM zinc acetate, 2% non-fat dried milk, in 1xPBS. The array was washed with PBS, dried by centrifugation, and scanned at 488nm at a 2μm resolution.

For each replicate, global mean normalization was used to ensure the mean intensity of each microarray was the same. Local mean normalization was then used to ensure that the intensity was evenly distributed throughout each sector of the microarray surface. Outliers were detected by calculating a coefficient variance (stdev/mean) and those over 0.75 were deleted. The replicates were then quantile-normalized to account for any possible nonlinearity between arrays. The median of the averaged features was subtracted to account for background. Z scores were calculated as |signal – median|/standard deviation. Because of the right-handed tail effect, standard deviation of the background signal was on the basis of the standard deviation from the median of all signals less than the median.

DIP-chip and motif discovery. DIP-chip was carried out as described previously (Liu et al., 2005) and the resulting DNA was hybridized to NimbleGen microarrays covering the yeast

genome at 32bp resolution. Peaks were identified by ChIPOTle (Buck et al., 2005) using a 200bp window and 50bp step size. DNA sequences under peaks (Bonferonni corrected $p < 1x10^{-3}$) were used as input for BioProspector and MDscan (Liu et al., 2002). The top 5 motifs returned by each program were then scored for their ability to predict the DIP-ChIP results by GOMER (Granek and Clarke, 2005). The motif with the highest ROC-AUC was reported.

Electrophoretic Mobility Shift Analyses. EMSA probes for Stp3+, Stp3-, Gat3+ and Gat3- were labeled with γP^{32}ATP and mixed with the cold reverse complement primer to have a final concentration of 0.1 mM of each oligo. Primers were denatured 10 min at 65 degrees, cooled down 10 min at room temperature. Each binding reaction contained 1x binding buffer (10mM Hepes pH7.8, 75mM KCl, 2.5mM MgCl2, 1mM DTT, 3% Ficoll), and variable concentrations of proteins (0, 1 or 10 nM). Binding reactions were mixed to radiolabeled primers at a final concentration of 5pM and incubated 15 minutes at room temperature, before being loaded on a 5 % non denaturing acrylamide gel and ran at 4 degrees during 50 minutes at 100V. Gels were dried on a Whatman paper and exposed overnight on a Phosphorimager screen.

For Ecm23 and Yml081w, twenty-four base single stranded DNA oligonucleotides were annealed and labeled with 32P using standard protocols. Reactions using less than 1nM dsDNA were performed in binding buffer (50mM Tris (pH 7.5), 25mM NaCl, 10mM KCl, 10% glycerol). Binding reactions were incubated on ice for 1h before loading onto the gel. The reactions were resolved through a prerun 10% acrylamide/3% glycerol gel in 0.5x TBE (45 mM Tris/32.3 mM boric acid/1.75 mM EDTA, pH 8.3) at 4°C. The gels were dried, exposed to a phosphorimager screen overnight and visualized using a Typhoon imager. Gels were analyzed using ImageQuant 5.2.

Nucleosome and mRNA tiling array analyses. Temperature-sensitive mutants (Supplementary Table 3) were grown at 22°C (permissive) until mid-log phase, then an equal volume of hot medium was added to equilibrate the culture to 37°C (restrictive). Cultures were grown a further 3 to 7 hours until a difference in OD between the mutant strain and its corresponding wildtype control became apparent. Extraction of nucleosomal DNA from the samples and hybridization onto the yeast tiling array was performed according to (Lee et al., 2007). Isolation of total RNA and hybridization onto the tiling arrays followed (Juneau et al., 2007), except that Actinomycin D was added in a final concentration of 6 µg/ml during cDNA synthesis to prevent antisense artefacts (Perocchi et al., 2007). Tiling arrays were quantile-normalized with the Affymetrix Tiling Analysis Software (TAS) v1.1 using perfect-match probes only and a bandwidth of 20. Raw data from nucleosomal DNA hybridizations from mutant strains were normalized against either nucleosomal DNA from the wildtype control (mutant/WT) or MNase treated genomic DNA (mutant/gDNA). For the total RNA hybridizations, raw data from the mutant strains were normalized against the corresponding wildtype control and a log2 expression difference was calculated for each gene by averaging across sense-probes mapping to the ORF.

ChIP-chip analysis. ***Culture and Crosslinking.*** Rsc8-TAP tagged strains in a wild-type or rsc3-1 background (Supplementary Table 3) were cultured as described in 'Nucleosome and mRNA tiling array analyses' above. After 7 hours of growth at restrictive conditions, cells were crosslinked by adding formaldehyde to a final concentration of 1%. After 20 min, crosslinking was stopped with glycine (final concentration 300 mM). Cells (200 ml) were spun down for 10

min at 4,000 rpm at 4°C, washed twice in 50 ml ice cold TBS pH 7.5 (Tris 10 mM, NaCl 150 mM, PMSF 1 mM) and once in 25 ml FA lysis buffer (50 mM Hepes pH 7.5, 150 mM NaCl, 1mM EDTA, 1% Triton, 0.1% Sodium deoxycholate, 0.1% SDS, and 1 pill of protease inhibitors, Roche, per 50 ml; PI). Pellets were resuspended in 2ml FA lysis buffer and centrifuged for 1 min at 14,000 rpm at 4°C. After removal of the supernatant, pellets were frozen in liquid nitrogen and stored at -80°C until further processing. ***Chromatin extraction.*** Pellets were resuspended in 2 ml FA lysis buffer plus PI and transferred to 2 ml screw-cap tubes with 0.5 ml zirconia/silica beads (diameter 0.5 µm, BioSpec Products). Cells were homogenized in a mini bead-beater for 7 cycles of 2 minutes, keeping the tubes on ice for 2 min between each cycle. The recovered whole cell extracts were transferred to 2 ml eppendorf tubes and centrifuged for 15 min at 14,000 rpm at 4°C. After removal of the supernatant, pellets were washed once by resuspending in 2 ml FA lysis buffer plus PI and incubating at 4°C for 30 mins on a rotator. The extracts were then centrifuged for 15 mins at 14,000 rpm at 4°C and pellets were taken up in 3 ml FA lysis buffer plus PI. Extracts were sonicated for 8 cycles of 25 seconds (0.5 sec on, 0.5 sec off; power setting 3) on a Branson Sonifier 450, leaving the samples on ice for 2 min between each cycle, to shear the chromatin to an average size of ~500 bp. After sonication, samples were centrifuged for 20 min at 14,000 rpm at 4°C. The supernatant, containing the chromatin-enriched extract (CE), was frozen in liquid nitrogen and stored at 80°C. ***Chromatin immunoprecipitation and array analysis.*** 700 µl of CE was incubated with 50 µl IgG sepharose beads (GE Healthcare) and 18 µl BSA (25 mg/ml) for 1.5 hours at RT. Beads were then washed two times with 1 ml FA lysis buffer plus PI, two times with 1 ml wash buffer 1 (FA lysis buffer with 500 mM NaCl), two times with 1 ml wash buffer 2 (10 mM Tris pH 8, 250 mM LiCl, 0.5% Nonidet P-40, 0.5% Sodium deoxycholate, 1 mM EDTA) and rinsed once in

1 ml TE 50/1 pH 7.5. Bound complexes were eluted twice with 50 µl of TE 50/1 pH 7.5, SDS 1% at 65°C for 10 mins. After elution, crosslinks were reversed by incubating O/N at 65°C in the presence of RNase (20 U RNase H, Epicentre; 60 U of RNase Cocktail, Ambion). Proteins were digested by adding 3.5 ul CaCl2 300 mM, 2 U proteinase K (Fermentas) and TE 10/1 to a final volume of 200 µl, and incubating for 20 mins at 55°C. DNA was isolated using standard phenol extraction, further purified using columns from the Qiagen PCR purification kit and eluted in a volume of 40 µl. The DNA was purified by Zymo column and amplified by LM-PCR as in (Ercan et al., 2007) with the exception that amplification was carried out in a 50µL reaction volume for 7 cycles, then 15µL of this initial amplification reaction was transferred to a new 50µL reaction for an additional 25 cycles. DNA was then cleaned up on a Zymo column and eluted into 18µL dH20. 1 µg of this DNA was then labeled as per NimbleGen's instructions and hybridized to 385k whole-genome S. cerevisiae chips (NimbleGen #C4214-00-01), scanned, and images processed using GenePix Pro 4.0.

Supplemental References

Berger, M.F., Badis, G., Gehrke, A.R., Talukder, S., Philippakis, A.A., Pena-Castillo, L., Alleyne, T.M., Mnaimneh, S., Botvinnik, O.B., Chan, E.T.*, et al.* (2008). Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. Cell *133*, 1266-1276.

Berger, M.F., Philippakis, A.A., Qureshi, A.M., He, F.S., Estep, P.W., 3rd, and Bulyk, M.L. (2006). Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. Nat Biotechnol *24*, 1429-1435.

Buck, M.J., Nobel, A.B., and Lieb, J.D. (2005). ChIPOTle: a user-friendly tool for the analysis of ChIP-chip data. Genome Biol *6*, R97.

Ercan, S., Giresi, P.G., Whittle, C.M., Zhang, X., Green, R.D., and Lieb, J.D. (2007). X chromosome repression by localization of the C. elegans dosage compensation machinery to sites of transcription initiation. Nat Genet *39*, 403-408.

Granek, J.A., and Clarke, N.D. (2005). Explicit equilibrium modeling of transcription-factor binding and gene regulation. Genome Biol *6*, R87.

Juneau, K., Palm, C., Miranda, M., and Davis, R.W. (2007). High-density yeast-tiling array reveals previously undiscovered introns and extensive regulation of meiotic splicing. Proc Natl Acad Sci U S A *104*, 1522-1527.

Lee, W., Tillo, D., Bray, N., Morse, R.H., Davis, R.W., Hughes, T.R., and Nislow, C. (2007). A high-resolution atlas of nucleosome occupancy in yeast. Nat Genet *39*, 1235-1244.

Liu, X., Noll, D.M., Lieb, J.D., and Clarke, N.D. (2005). DIP-chip: rapid and accurate determination of DNA-binding specificity. Genome Res *15*, 421-427.

Liu, X.S., Brutlag, D.L., and Liu, J.S. (2002). An algorithm for finding protein-DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments. Nat Biotechnol *20*, 835-839.

Mintseris, J., and Eisen, M.B. (2006). Design of a combinatorial DNA microarray for protein-DNA interaction studies. BMC Bioinformatics *7*, 429.

Perocchi, F., Xu, Z., Clauder-Munster, S., and Steinmetz, L.M. (2007). Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin D. Nucleic Acids Res *35*, e128.

Warren, C.L., Kratochvil, N.C., Hauschild, K.E., Foister, S., Brezinski, M.L., Dervan, P.B., Phillips, G.N., Jr., and Ansari, A.Z. (2006). Defining the sequence-recognition profile of DNA-binding molecules. Proc Natl Acad Sci U S A *103*, 867-872.