Supplementary Data

Supplementary Table S1: List of Plasmids

Plasmid	Relevant Description	Strain	Source
pUC19	Amp^R	BL21	New England Biolabs
YCplac33	Amp^R, *URA3*, CEN4	DH10B	(1)
YEplac195	Amp^R, *URA3*, 2µ Ori	DH10B	(1)
YEplac181	Amp^R, *LEU2*, 2µ Ori	DH10B	(1)
pRS423	Amp^R, *HIS3*, 2µ Ori	DH10B	(2)
pRS426	Amp^R, *URA3*, 2µ Ori	DH10B	(2)
pGEX-5X-3	GST fusion vector	BL21	This work
pET28b+	Kan^R	DH10B	Novagen
pMAL-c2x	Amp^R	BL21(DE3)	New England Biolabs
pMS2	*TAF9*-HA_3 on YEplac181	DH10B	This work
pMS3	*TAF9*-HA_3 on YCplac111	DH10B	This work
pMS4	*taf9-tCRD2*-HA_3 on YEplac181	DH10B	This work
pMS5	*taf9-tCRD2*-HA_3 on YCplac111	DH10B	This work
pMS11	*TAF6* in pET28b+	DH10B	This work
pMS31	*taf9-tCRD2*-Myc_{13} in pRS423	DH10B	This work
pMS45	*TAF6* in pGEX-5X-3	DH10B	This work
pMS46	*TAF6* in pGEX-5X-3	BL21(DE3)	This work
pMS53	*TAF9* in pMAL-c2x	DH10B	This work
pMS54	*taf9-tCRD1* in pMAL-c2x	DH10B	This work
pMS55	*taf9-tCRD2* in pMAL-c2x	DH10B	This work
pMS56	*TAF9* in pMAL-c2x	BL21(DE3)	This work
pMS57	*taf9-tCRD1* in pMAL-c2x	BL21(DE3)	This work
pMS58	*taf9-tCRD2* in pMAL-c2x	BL21(DE3)	This work
pSS1	*taf9-tCRD1*-Myc_{13} in YEplac181	DH10B	This work
pSS2	*taf9-tCRD2*-Myc_{13} in YEplac181	DH10B	This work
pSS3	*taf9-tCRD3*-Myc_{13} in YEplac181	DH10B	This work
pSS4	*TAF9* in YEplac181	DH10B	This work
pSS5	*taf9-tCRD2* in YEplac181	DH10B	This work
S10	*TAF9* in YEplac111	DH10B	This work
tCRD2 sc	*taf9-tCRD2* in YCplac111	DH10B	This work
Sp1	*TAF9*-Myc_{13} in YEplac181	DH10B	This work

Plasmid	Relevant Description	Strain	Source
Sp2	*TAF9*-Myc_{13} in YCplac111	DH10B	This work
Sp1-2	Dra III-Bam HI adaptor in place of the *TAF9* ORF in Sp1	DH10B	This work
Ip1	*TAF9*-Myc_{13} in YCplac33	DH10B	This work
Rp1-2	*TAF9*-Myc_{13} in YEplac195	DH10B	This work
p2382	HA_3-*GCN4* in YCp50	DH10B	K. Natarajan
m144	*taf9-m144* in YEplac181	DH10B	This work
m91	*taf9-m91* in YEplac181	DH10B	This work
m99	*taf9-m99* in YEplac181	DH10B	This work
m124	*taf9-m124* in YEplac181	DH10B	This work
m123	*taf9-m123* in YEplac181	DH10B	This work
m149	*taf9-m149* in YEplac181	DH10B	This work
m147	*taf9-m147* in YEplac181	DH10B	This work
m131	*taf9-m131* in YEplac181	DH10B	This work
m111	*taf9-m111* in YEplac181	DH10B	This work
pHYC(14x2)	*CYC1::*UAS_{GCRE}*-lacZ*	DH10B	(3)
pFA6a-13xMyc-HIS3MX6	Myc_{13} tagging construct	DH10B	(4)
p2528	GST-Gcn4 in pGEX-5X-3	BL21	(5)
p2144	GST-Gcn4 10 Ala in pGEX-5X-3	BL21	(5)

Supplementary Table S2: List of Yeast Strains

Strain Name	Genotype	Source
BY4741	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*	(6)
BY4742	*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*	(6)
YMS4	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 taf9Δ::kanMX4* [pMS2]	This work
YMS6	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 taf9Δ::kanMX4* [pMS4]	This work
YMS8	*MATα his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 LYS2 RPB3-Myc$_{13}$::HIS3* taf9Δ::kanMX4* [pMS3]	This work
YMS9	*MATα his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 LYS2 RPB3-Myc$_{13}$::HIS3* taf9Δ::kanMX4* [pMS4]	This work
YMS10	*MATα his3Δ1 leu2Δ0 ura3Δ0 MET15 LYS2 SPT7-Myc$_{13}$::HIS3* taf9Δ::kanMX4* [pMS3]	This work
YMS11	*MATα his3Δ1 leu2Δ0 ura3Δ0 MET15 LYS2 SPT7-Myc$_{13}$::HIS3* taf9Δ::kanMX4* [pMS4]	This work
YMS14	*MATa his3Δ1 leu2Δ0 ura3Δ0 MET15 LYS2 TAF11-Myc$_{13}$::HIS3* taf9Δ::kanMX4* [pMS3]	This work
YMS16	*MATa his3Δ1 leu2Δ0 ura3Δ0 MET15 LYS2 TAF11-Myc$_{13}$::HIS3* taf9Δ::kanMX4* [pMS4]	This work
YMS22	*MATa his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 LYS2 SPT15-Myc$_{13}$::HIS3* taf9Δ::kanMX4* [pMS3]	This work
YMS23	*MATa his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 LYS2 SPT15-Myc$_{13}$::HIS3* taf9Δ::kanMX4* [pMS4]	This work
YMS27	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0* [YEplac181]	This work
YMS28	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0* [pSS1]	This work
YMS30	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0* [pSS2]	This work
YMS34	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0* [pSS3]	This work
YMS48	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0* [pRS426]	This work
YMS50	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 spt8Δ::kanMX4* [pRS426]	This work
YMS51	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 spt20Δ::kanMX4* [pRS426]	This work
YMS94	*MATa his3Δ1 leu2Δ0 met15Δ0 lys2Δ0 URA3 GAL11-Myc$_{13}$::HIS3* taf9Δ::kanMX4* [S10]	This work
YMS95	*MATa his3Δ1 leu2Δ0 met15Δ0 lys2Δ0 URA3 GAL11-Myc$_{13}$::HIS3* taf9Δ::kanMX4* [pSS5]	This work
YMS144	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 taf9::kanMX4* [pSS2] [pRS423]	This work
YMS145	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 taf9::kanMX4* [pSS2] [pMS31]	This work
YMS147	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0* [Sp2]	This work
HQY366	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 TBP-Myc$_{13}$::HIS3**	(7)
RPY1	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 taf9::kanMX4* [Ip1]	This work
RPY3	*MATα his3Δ1 leu2Δ0 lys2Δ0 MET15 ura3Δ0 taf9::kanMX4* [Ip1]	This work
RPY66	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 taf9::kanMX4* [Sp2]	This work
RPY67	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 taf9::kanMX4* [Sp1]	This work
RPY72	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 taf9::kanMX4* [Ip1][Sp2]	This work
YSS1	*MATa his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 TAF9-Myc$_{13}$::HIS3**	This work
YSS2	*MATa his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 TAF11-Myc$_{13}$::HIS3**	This work

YSS4	*MATa his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 gcn4Δ-103 TAF9-Myc$_{13}$::HIS3**	This work
YSS26	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 taf9::kanMX4* [pSS2]	This work
YSS18	*MATa his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 lys2Δ0 SPT15-Myc$_{13}$::HIS3** [Ip1]	This work
YSS-C4	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 taf9::kanMX4* [pSS1]	This work
YSS-C5	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 taf9::kanMX4* [pSS2]	This work
YSS-C6	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 taf9::kanMX4* [pSS3]	This work
H3281	*MATa his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 LYS2 RPB3-Myc$_{13}$::HIS3**	(8)
H3235	*MATa his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 LYS2 SPT7-Myc$_{13}$::HIS3**	(8)
H3647	*MATa his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 LYS2 SUA7-Myc$_{13}$::HIS3**	(8)
YSS2-21	*MATa his3Δ1 leu2Δ0 ura3Δ0 MET15 LYS2 TAF11::13Myc::HIS3** [Ip1]	This work
3235-3	*MATα his3Δ1 leu2Δ0 MET15 ura3Δ0 LYS2 taf9Δ::kanMX4 SPT7-Myc$_{13}$::HIS3** [Ip1]	This work
3281-27m	*MATα his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 lys2Δ0 RPB3-Myc$_{13}$::HIS3* taf9Δ::kanMX4* [Ip1]	This work
3647-33	*MATα his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 LYS2 SUA7-Myc$_{13}$::HIS3* taf9Δ::kanMX4* [Ip1]	This work
3647-33 WT$_{NM3}$	*MATα his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 LYS2 SUA7-Myc$_{13}$::HIS3* taf9Δ::kanMX4* [S10]	This work
3647-33 tCRD2$_{NM3}$	*MATα his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 LYS2 SUA7-Myc$_{13}$::HIS3* taf9Δ::kanMX4* [pSS5]	This work
#249	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 gcn4Δ::kanMX4*	Open Biosystems
#7390	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 spt20Δ::kanMX4*	Open Biosystems
#2667	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0 spt8Δ::kanMX4*	Open Biosystems
#20822	*MATa/MATα his3Δ1/his3Δ1 leu2Δ0/leu2Δ0 MET15/met15Δ0 LYS2/lys2Δ0 ura3Δ0/ura3Δ0 TAF9/taf9::kanMX4*	Open Biosystems

Table S3: Comparison of the TFs bound at Taf9 CTD induced genes under SM

Factor	Total bound at $p \leq 0.01$	Taf9 matches $p \leq 0.01$	% CTD bound	% CTD genes	Factor	Total bound at $p \leq 0.005$	Taf9 matches $p \leq 0.005$	% CTD bound	% CTD genes
ARG81	81	13	16.0	2.8	ARG81	51	9	17.6	1.9
MET32	93	12	12.9	2.6	CAD1	42	7	16.7	1.5
GCN4	345	44	12.8	9.5	GCN4	292	41	14.0	8.9
SIP4	42	5	11.9	1.1	SIP4	30	4	13.3	0.9
CAD1	63	7	11.1	1.5	MET32	72	9	12.5	1.9
GLN3	141	15	10.6	3.2	PHO2	63	7	11.1	1.5
CHA4	162	15	9.3	3.2	GLN3	100	11	11.0	2.4
ARO80	191	17	8.9	3.7	ARO80	118	12	10.2	2.6
DAL82	125	11	8.8	2.4	DAL82	95	9	9.5	1.9
PHO2	115	10	8.7	2.2	CHA4	119	11	9.2	2.4
ARG80	141	12	8.5	2.6	ARG80	58	5	8.6	1.1
CBF1	365	30	8.2	6.5	CBF1	341	28	8.2	6.0
RTG3	93	7	7.5	1.5	PUT3	50	4	8.0	0.9
STP1	178	13	7.3	2.8	STP1	129	10	7.8	2.2
BAS1	70	5	7.1	1.1	DAL81	65	5	7.7	1.1
HAP5	160	11	6.9	2.4	BAS1	40	3	7.5	0.6
MOT3	105	7	6.7	1.5	MOT3	68	5	7.4	1.1
LEU3	138	9	6.5	1.9	RAP1	256	18	7.0	3.9
RAP1	387	25	6.5	5.4	RTG3	62	4	6.5	0.9
PUT3	79	5	6.3	1.1	HAP4	16	1	6.3	0.2
DAL81	181	11	6.1	2.4	HAP5	99	6	6.1	1.3
MET4	106	6	5.7	1.3	MET4	66	4	6.1	0.9
MET31	77	4	5.2	0.9	RCS1	58	3	5.2	0.6
FHL1	287	14	4.9	3.0	MET31	46	2	4.3	0.4
RCS1	86	4	4.7	0.9	GAT1	94	4	4.3	0.9
RTG1	65	3	4.6	0.6	FHL1	263	10	3.8	2.2
GAT1	132	6	4.5	1.3	RTG1	32	1	3.1	0.2
GCR2	114	5	4.4	1.1	RPH1	35	1	2.9	0.2
HAP4	29	1	3.4	0.2	GCR2	90	2	2.2	0.4
SFP1	114	3	2.6	0.6	LEU3	101	2	2.0	0.4
UGA3	54	1	1.9	0.2	SFP1	87	1	1.1	0.2
RPH1	55	1	1.8	0.2	UGA3	34	0	0.0	0.0
ADR1	40	0	0.0	0.0	ADR1	20	0	0.0	0.0
Factor denotes TFs analyzed under SM									
Total bound at p-value: Total target promoters bound by corresponding TF at indicated p-value									
Taf9 matches at p-value: Total number of Taf9 CTD gene promoters for each TF at indicated p-value									
% CTD bound: % of total bound by Taf9 CTD gene promoters for each TF at indicated p-value									
% CTD genes: % of 463 Taf9 CTD gene promoters bound for each TF at indicated p-value									

Table S4. List of TFs bound at p=0.005 from Harbison et al. (9) and statistically enriched at Taf9 CTD induced genes

Factor	Taf9 matches	Expected	Std dev	Difference	% change	Z score	P-value
GCN4	44	27.202	5.295	16.798	61.755	3.172	0.00215
ARG81	13	5.918	2.529	7.082	119.686	2.8	0.011
MET32	12	6.681	2.685	5.319	79.611	1.981	0.046
GLN3	15	10.499	3.352	4.501	42.873	1.343	0.119
SIP4	5	2.577	1.675	2.423	94.024	1.446	0.128
CAD1	7	4.868	2.297	2.132	43.806	0.928	0.225
CHA4	15	12.026	3.582	2.974	24.73	0.83	0.236
ARO80	17	14.698	3.948	2.302	15.659	0.583	0.311
PHO2	10	8.399	3.005	1.601	19.06	0.533	0.337
DAL82	11	9.354	3.168	1.646	17.603	0.52	0.339
CBF1	30	27.774	5.347	2.226	8.014	0.416	0.362
ARG80	12	11.167	3.455	0.833	7.46	0.241	0.44
SFP1	3	9.735	3.231	-6.735	-69.184	2.085	0.015
FHL1	14	23.098	4.903	-9.098	-39.387	1.856	0.032
ADR1	0	3.436	1.933	-3.436	-100	1.778	0.037
UGA3	1	4.677	2.252	-3.677	-78.618	1.633	0.06
RPH1	1	4.39	2.182	-3.39	-77.223	1.554	0.075
GAT1	6	10.499	3.352	-4.499	-42.851	1.342	0.11
GCR2	5	8.59	3.039	-3.59	-41.793	1.181	0.153
RAP1	25	30.351	5.573	-5.351	-17.631	0.96	0.194
MET4	6	8.972	3.104	-2.972	-33.123	0.957	0.219
RCS1	4	6.395	2.628	-2.395	-37.449	0.911	0.246
DAL81	11	14.221	3.885	-3.221	-22.651	0.829	0.25
MET31	4	6.013	2.55	-2.013	-33.477	0.79	0.294
RTG1	3	4.868	2.297	-1.868	-38.369	0.813	0.295
HAP4	1	2.482	1.644	-1.482	-59.703	0.901	0.304
HAP5	11	12.408	3.637	-1.408	-11.346	0.387	0.421
PUT3	5	6.108	2.569	-1.108	-18.146	0.431	0.435
LEU3	9	10.213	3.307	-1.213	-11.873	0.367	0.437
MOT3	7	8.017	2.937	-1.017	-12.689	0.346	0.456
STP1	13	13.839	3.835	-0.839	-6.065	0.219	0.484
BAS1	5	5.44	2.427	-0.44	-8.094	0.181	0.542
RTG3	7	7.158	2.778	-0.158	-2.212	0.057	0.576

The TFs indicated in Green are statistically overrepresented and those in Red are under-represented with respect to negative control gene sets.

Table S5. List of TFs bound at p=0.01 from Harbison et al. (9) and statistically enriched at Taf9 CTD induced genes

Factor	Taf9 matches	Expected	Std dev	Difference	% change	Z score	P-value
GCN4	44	27.202	5.295	16.798	61.755	3.172	0.00215
ARG81	13	5.918	2.529	7.082	119.686	2.8	0.011
MET32	12	6.681	2.685	5.319	79.611	1.981	0.046
GLN3	15	10.499	3.352	4.501	42.873	1.343	0.119
SIP4	5	2.577	1.675	2.423	94.024	1.446	0.128
CAD1	7	4.868	2.297	2.132	43.806	0.928	0.225
CHA4	15	12.026	3.582	2.974	24.73	0.83	0.236
ARO80	17	14.698	3.948	2.302	15.659	0.583	0.311
PHO2	10	8.399	3.005	1.601	19.06	0.533	0.337
DAL82	11	9.354	3.168	1.646	17.603	0.52	0.339
CBF1	30	27.774	5.347	2.226	8.014	0.416	0.362
ARG80	12	11.167	3.455	0.833	7.46	0.241	0.44
SFP1	3	9.735	3.231	-6.735	-69.184	2.085	0.015
FHL1	14	23.098	4.903	-9.098	-39.387	1.856	0.032
ADR1	0	3.436	1.933	-3.436	-100	1.778	0.037
UGA3	1	4.677	2.252	-3.677	-78.618	1.633	0.06
RPH1	1	4.39	2.182	-3.39	-77.223	1.554	0.075
GAT1	6	10.499	3.352	-4.499	-42.851	1.342	0.11
GCR2	5	8.59	3.039	-3.59	-41.793	1.181	0.153
RAP1	25	30.351	5.573	-5.351	-17.631	0.96	0.194
MET4	6	8.972	3.104	-2.972	-33.123	0.957	0.219
RCS1	4	6.395	2.628	-2.395	-37.449	0.911	0.246
DAL81	11	14.221	3.885	-3.221	-22.651	0.829	0.25
MET31	4	6.013	2.55	-2.013	-33.477	0.79	0.294
RTG1	3	4.868	2.297	-1.868	-38.369	0.813	0.295
HAP4	1	2.482	1.644	-1.482	-59.703	0.901	0.304
HAP5	11	12.408	3.637	-1.408	-11.346	0.387	0.421
PUT3	5	6.108	2.569	-1.108	-18.146	0.431	0.435
LEU3	9	10.213	3.307	-1.213	-11.873	0.367	0.437
MOT3	7	8.017	2.937	-1.017	-12.689	0.346	0.456
STP1	13	13.839	3.835	-0.839	-6.065	0.219	0.484
BAS1	5	5.44	2.427	-0.44	-8.094	0.181	0.542
RTG3	7	7.158	2.778	-0.158	-2.212	0.057	0.576

The TFs indicated in Green are statistically overrepresented and those in Red are under-represented with respect to negative control gene sets.

Supplementary Method

Taf6-Taf9 interaction assay:

Construction of GST-TAF6 clones:

Plasmid pMS11 was constructed by PCR amplification of the *TAF6* ORF from S288C genomic DNA and cloned into pET28b+ vector. The *TAF6* ORF from pMS11 was subcloned into pGEX-5X-3 vector to generate GST-*TAF6*. The construct was transformed into *E.coli* BL21 strain, grown and induced at 25°C with 1mM IPTG for 5 hours. Lysates were prepared by sonication. Details of plasmid construction are available on request.

Construction of MBP-TAF9 Clones:

The *TAF9*, *taf9-tCRD1* and *taf9-tCRD2* inserts were PCR amplified using primers ON368/ON369 (*TAF9),* ON368/ON371 (*taf9-tCRD1)* or ON368/ON370 (*taf9-tCRD2*), digested with EcoRI and XhoI and cloned into pMAL-c2x to obtain MBP fusions. The positive clones were named pMS53 (*TAF9*), pMS54 (*taf9-tCRD1*) and pMS55 (*taf9-tCRD2*). The constructs were transformed into BL21 (DE3) and fusion proteins induced at 37°C with 0.3mM IPTG and 0.2% glucose for 3 hrs.

GST pull down assay:

The sonicated whole cell extracts were prepared, separated on an SDS-PAGE gel and specific protein concentrations were estimated by visual quantification using Low-range protein marker (BioRad). *E. coli* lysate containing ~1nmole (83µl) of the GST-TAF6 protein was mixed with *E. coli* lysates containing 10 nmole or 30 nmole each of Taf9, Taf9-tCRD1 or Taf9-tCRD2 MBP fusion proteins, or MBP alone as control. The GSH-Sepharose 4B (GE Healthcare) beads (10µl beads per reaction) were added, reaction mixture incubated at 4°C for 2 h, washed three times with GST-binding buffer as per manufacturer's instructions and fusion protein eluted with 10mM

reduced glutathione. The eluates and analysed on SDS-PAGE gel along with input samples, gel stained with Coomassie Brilliant Blue to visualize the protein bands.

Meta-analysis to predict TFs bound to genes induced in a Taf9-CTD dependent manner.

We obtained the lists of gene promoters bound by each of the 34 TFs under SM conditions at $p \leq 0.01$ and $p \leq 0.005$ identified by ChIP-on-chip analysis by the Young lab {Harbison, 2004 #4994}. We then determined the overlap between the gene promoters bound by the 34 TFs and the 463 genes induced in a Taf9 CTD-dependent manner. To evaluate the significance of the overlap between genes induced in a Taf9 CTD-dependent manner and the genes bound by 34 TFs, the following test was conducted.

We considered a "negative list" of 4848 genes, consisting of all ORFs in the yeast genome except for the 463 Taf9 CTD-induced genes, 400 genes that were down-regulated, and those marked dubious. The negative set was used to compute, in a typical set of 463 genes, how many genes $\langle n \rangle$ are expected to be targets for a given transcription factor; what is the standard deviation σ for this expectation; by how many standard deviations (z-score) does the observed number of targets n differ from $\langle n \rangle$ and what is the corresponding (one-sided) p-value (probability of seeing an observation with this z-score or higher under the "null hypothesis" that there is no special property of the 463 induced genes.

The statistical error model employed is described as follows. We assumed that the probability of a region bound by a given TF being *p,* and different regions being independently bound (so that these are Bernoulli trials, with outcomes given by the Binomial distribution). We have upstream regulatory regions for *N* genes, of which *n* are bound by that factor. If we knew *p*, we could determine whether *n* is significant or not, as follows: the expected number of bound regions is *Np*, and the expected standard deviation is $\sqrt{Np(1-p)}$. An observation *n* that

differed from *Np* by more than, say, two standard deviations could be deemed significant.

Since *p* is unknown, we constructed a negative set with upstream sequence for *M* genes, chosen to be statistically similar to the "positive" set of *N* genes. We find *m* of these are bound by that factor. If *M* is large enough, we could estimate *p = m/M* and proceed as above. However, we preferred the exact expression below. Given that out of *M* regulatory regions in the negative set, *m* are bound by a factor, the probability that *n* out of *N* in the positive set are also bound by that factor is

$$P(n,N|m,M) = \frac{\binom{N}{m}\binom{M+2c-1}{m+c-1}}{\binom{N+M+2c-1}{n+m+c-1}} \frac{(M-m+c)}{(N-n+M-m+c)}$$

This formula can be derived as follows: *P(n, N | m, M) = P(n+m, N+M)/P(m, M).* That is, the conditional probability is the ratio of the joint probability for seeing *(m+n)* bound regions in the total set of *(M+N)* regions, to the probability of seeing *m* bound regions in the original set of *M* regions. For a given *p* each of these is given by the binomial distribution. Since we don't know *p*, we integrate (numerator and denominator separately) over *p*, assuming a "Beta prior" for *p*:

$P(p) \propto p^{c-1}(1-p)^{c-1}$

We take *c*=1, which yields a "uniform prior" (all values of *p* are equally probable *a priori*). The formula above is not very sensitive to *c*, provided *c* is small. Also, if *M* and *m* are much larger than *N* and *n*, then the formula reduces to the binomial expression

$P(n,N) = \binom{N}{n} p^n (1-p)^{N-n}$ with *p = m/M*.

Using the exact formula for *P(n,N|m,M)* with *c*=1, we calculate the mean, standard deviation, *z*-score, *p*-value, and two-sided *p*-value for each *n* in the positive set, given the corresponding *m* and *M* in the negative set. The *p* value is defined as the probability of seeing data as extreme as, or more extreme than, the observed *n* on the same side of the mean. So if *n* is greater than the

mean, the *p*-value is $\sum_{n\prime n} P\,(n', N|m, M)$.

By this approach, we identified TFs that were overrepresented in the 463 Taf9 CTD gene list in relation to the random control gene lists.

Estimating the significance of the gene overlap in Venn diagram

To evaluate the statistical significance of the overlap between genes induced in a Taf9 CTD-dependent manner and genes induced in a Gcn4-dependent or Spt20-dependent manner, we designated *N* to be the "pool" of genes to draw from (where N = 5900, the number of non-dubious ORFs annotated in *S. cerevisiae*). Let *m* and *n* be the number of genes in the first list and the second list, and let the overlap of these be *k*. We assumed that *m* genes fall in the overlapped category, and if we draw *n* genes at random (without replacement) from the pool of *N* genes, we determined the probability that *k* of them will be in that same category. This was determined by the hypergeometric distribution.

$$P(k; n \vee m; N) = \frac{\binom{n}{k}\binom{N-n}{m-k}}{\binom{N}{m}}$$

where the bracketed factors are binomial coefficients

$$\binom{N}{n} = \frac{N!}{n!\,(N-n)!}$$

The hypergeometric distribution is symmetric on exchanging *n* and *m*, as can be verified on expanding: $P(k; n \vee m; N) = P(k; m \vee n; N)$(that is, it does not matter which is the first list and which the second). It vanishes if *k* is greater than either *m* or *n*.

We assumed that the probability that a random draw of *n* genes would have an overlap of *k or higher*. So the *p*-value for the data is

$p = \sum_{k\prime=k}^{min(m,n)} \frac{\binom{n}{k}\binom{N-n}{m-k}}{\binom{N}{m}}$.

If we assume N=5900, then for Fig. 3A (n=227, m=463, k=134), we get $p = 2.12 \times 10^{-94}$, and for Fig 4A (with n=515, m=463, k=175) we get $p = 1.98 \times 10^{-75}$, which is vanishingly small values suggesting a highly significant overlap. If we assume that for biological reasons the "relevant" pool of genes is much smaller than this – say, roughly one-third, i.e. N=2000 – we still get tiny p-values of $p = 3.37 \times 10^{-11}$ and $p = 9.83 \times 10^{-36}$ respectively. The "expected" overlap is given approximately by nm/N and, for figures 4a and 4c, is respectively 40 and 18 if we take the pool size N=5900. With N=2000, the expected overlaps are 119 and 53 respectively.

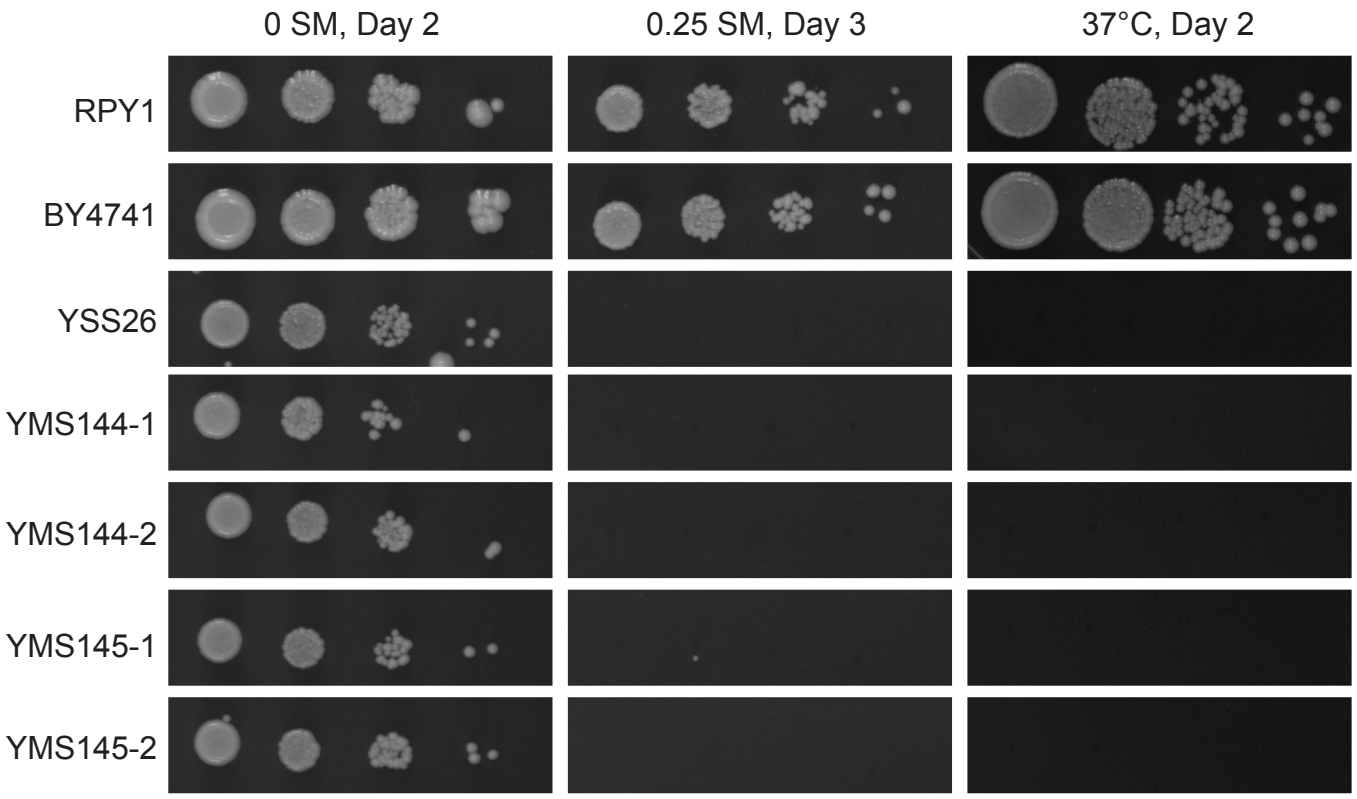
Supplemental References

1. **Gietz RD, Sugino A.** 1988. New yeast-Escherichia coli shuttle vectors constructed with in vitro mutagenized yeast genes lacking six-base pair restriction sites. Gene **74:**527-534.
2. **Sikorski RS, Hieter P.** 1989. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. Genetics **122:**19-27.
3. **Hinnebusch AG, Lucchini G, Fink GR.** 1985. A synthetic *HIS4* regulatory element confers general amino acid control on the cytochrome *c* gene *CYC1* of yeast. Proc. Natl. Acad. Sci. U S A **82:**498-502.
4. **Longtine MS, McKenzie A, 3rd, Demarini DJ, Shah NG, Wach A, Brachat A, Philippsen P, Pringle JR.** 1998. Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. Yeast **14:**953-961.
5. **Drysdale CM, Jackson BM, McVeigh R, Klebanow ER, Bai Y, Kokubo T, Swanson M, Nakatani Y, Weil PA, Hinnebusch AG.** 1998. The Gcn4p activation domain interacts specifically in vitro with RNA polymerase II holoenzyme, TFIID, and the Adap-Gcn5p coactivator complex. Mol. Cell. Biol. **18:**1711-1724.
6. **Brachmann CB, Davies A, Cost GJ, Caputo E, Li J, Hieter P, Boeke JD.** 1998. Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. Yeast **14:**115-132.
7. **Qiu H, Hu C, Yoon S, Natarajan K, Swanson MJ, Hinnebusch AG.** 2004. An array of coactivators is required for optimal recruitment of TATA binding protein and RNA polymerase II by promoter-bound Gcn4p. Mol. Cell. Biol. **24:**4104-4117.
8. **Qiu H, Hu C, Zhang F, Hwang GJ, Swanson MJ, Boonchird C, Hinnebusch AG.** 2005. Interdependent recruitment of SAGA and Srb mediator by transcriptional activator Gcn4p. Mol. Cell. Biol. **25:**3461-3474.
9. **Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne JB, Reynolds DB, Yoo J, Jennings EG, Zeitlinger J, Pokholok DK, Kellis M, Rolfe PA, Takusagawa KT, Lander ES, Gifford DK, Fraenkel E, Young RA.** 2004. Transcriptional regulatory code of a eukaryotic genome. Nature **431:**99-104.

Fig. S1. Additional *taf9-tCRD2* does not rescue mutant phenotype. (**Top**) Spot assay analysis of YSS26 strain bearing high-copy *taf9-tCRD2::Myc*$_{13}$ (YMS145) or control pRS423 (YMS144). Strains bearing *TAF9::Myc*$_{13}$ (RPY1), untagged *TAF9* (BY4741) and *taf9-tCRD2::Myc*$_{13}$ (YSS26) are shown as controls. (**Bottom**) Western blot analysis of WT and *taf9-tCRD2* proteins probed with anti-Myc antibody from indicated strains. Anti-G6PDH was used as loading control.

Fig. S2. TAF6-TAF9 interaction is intact in a Taf9 CTD mutant strain in vitro**.** An *E. coli* cell extract containing 1nmole of GST-Taf6 was mixed with *E. coli* cell extracts containing three or nine-fold molar excess of MBP-Taf9, MBP-tCRD1, MBP-tCRD2 or MBP alone as control. The reactions were incubated for 2h at 4°C, and the GST-Taf6 was pulled down using GSH-Sepharose beads, loaded on a 10% SDS-PAGE gel and stained with coomassie brilliant blue. For each pull down sample, the entire eluate was used for gel analysis. *E. coli* lysates 0.1µl each containing GST-TAF6 (12pmol), MBP-TAF9 (14pmol), MBP-tCRD1 (40pmol), MBP-tCRD2 (40pmol) fusion proteins or MBP alone (75pmol) were also loaded as control.

Figure S1


0 SM, Day 2
0.25 SM, Day 3
37°C, Day 2
RPY1
BY4741
YSS26
YMS144-1
YMS144-2
YMS145-1
YMS145-2

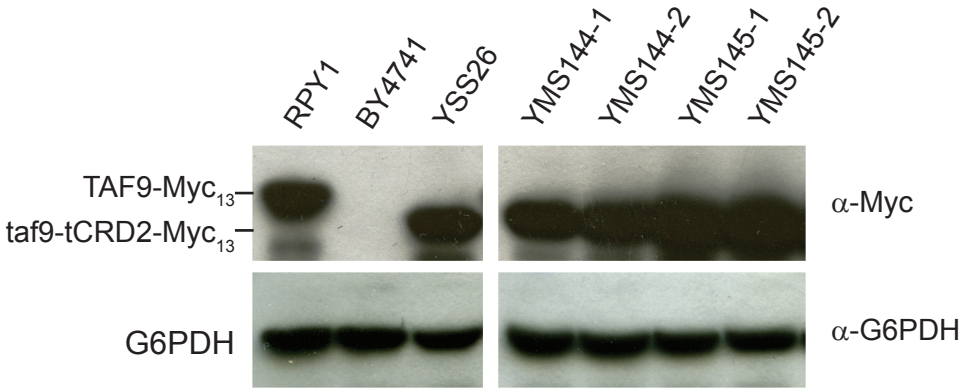
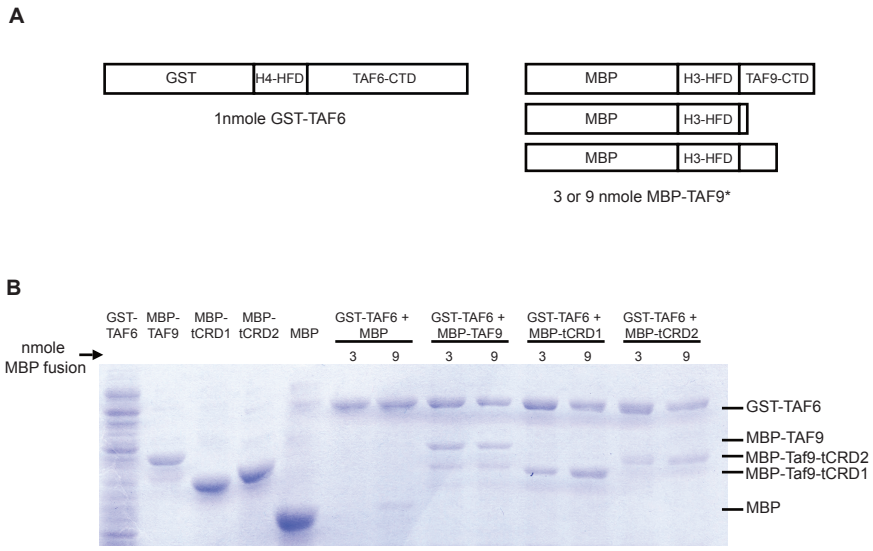

RPY1
BY4741
YSS26
YMS144-1
YMS144-2
YMS145-1
YMS145-2
TAF9-Myc13
taf9-tCRD2-Myc13
α-Myc
G6PDH
α-G6PDH

Figure S2


A
GST
H4-HFD
TAF6-CTD
1nmole GST-TAF6
MBP
H3-HFD
TAF9-CTD
MBP
H3-HFD
MBP
H3-HFD
3 or 9 nmole MBP-TAF9*
B
GST-TAF6
MBP-TAF9
MBP-tCRD1
MBP-tCRD2
MBP
GST-TAF6 + MBP
GST-TAF6 + MBP-TAF9
GST-TAF6 + MBP-tCRD1
GST-TAF6 + MBP-tCRD2
3
9
nmole MBP fusion
GST-TAF6
MBP-TAF9
MBP-Taf9-tCRD2
MBP-Taf9-tCRD1
MBP