Supplemental Information

Supplemental Methods

Mass spectrometry analysis

In order to identify the components of the purified complexes, bands were cut on the SDS-PAGE gel all along the migration lane of the immunoprecipitated complex, as well as on the control lane. Proteins of the different gel slices were treated with 10mM DTT and 55 mM iodoacetamide for cysteine alkylation, and digested with trypsin (Promega). The resulting peptides were extracted from the gel by successive incubations in 10% formic acid/acetonitrile (1/1) and dried in a speed-vac. Peptides were reconstituted in 5%,acetonitrile, 0.05% trifluoroacetic acid and analyzed by nanoLC-MS/MS using an Ultimate3000 system (Dionex) coupled to a LTQ-Orbitrap mass spectrometer (Thermo Fisher Scientific). Elution from the analytical column (75-µm inner diameter x 15-cm PepMap C18, Dionex) was performed using a 60min acetonitrile gradient at a flow rate of 300 nL/min. The LTQ-Orbitrap was operated in information-dependant acquisition mode with the XCalibur software. Survey scan MS were acquired in the Orbitrap on the 300-2000 m/z range with the resolution set to a value of 60000. The five most intense ions per survey scan were selected for CID fragmentation and the resulting fragments were analyzed in the linear trap (LTQ). Dynamic exclusion was employed within 60 seconds to prevent repetitive selection of the same peptide. The Mascot Daemon software (version 2.2.0, Matrix Science) was used to perform database searches against human entries in Uniprot. Peaklists were created with ExtractMSN (provided with Xcalibur version 2.0 SR2, Thermo Fisher Scientific) with the following parameters: parent ions in the mass range 400-4500, no grouping of MS/MS scans, threshold at 1000. For database search, carbamidomethylation of cysteines was set as a fixed modification, oxidation of methionines, O-GlcNAc glycosylation and phosphorylation of serines and threonines were set as variable modifications, and the mass tolerances in MS and MS/MS were set to 10 ppm and 0.6 Da, respectively. Mascot results were parsed with the in-house developed software MFPaQ version 4.0 and protein hits were automatically validated if were assigned at least one top ranking peptide with a p-value < 0.05. The software was also used to extract MS signal of identified proteins from raw files, and to calculate mean ratio of peptides intensity signal between the lane of immunopurified complex and the control lane. Proteins were considered as potential specific partners if they were exclusively identified by MS/MS sequencing in the lane of the complex with a significant score, or if the ratio of their signal intensity between the complex and control lanes was above 10.

Primer sequences

1) Primers used in ChIP-QPCR experiments :

" -178 " Forward : 5'-CCCACACAAAACATGGTAGCA-3'
Reverse : 5'-ATGGGCCCCATTGGATATGGACATG -3'

" +84 " Forward : 5'-AAAGTGCTGTCTGGCTCCAACT-3'
Reverse : 5'-GACAGAGTGGGAAGGGTTAGGTT-3'

" +2530 " Forward : 5'-GAGGAAGAAGAGAAGGAGAGGAACA-3'
Reverse : 5'-ATCCACACATCAGACATTCATCTCTAT-3'

" p107 " Forward : 5'-CCGGAGGAAAAACGGACTTT-3'
Reverse : 5'-CTGCGGGACGTGTTGTCAT -3'

" Gal4UAS " Forward : 5'-TCATCAATGTATCTTATCATGTCTGGAT-3'
Reverse : 5'-CGGAATGCCAAGCTGGAA -3'

2) Primers used in Q-PCR experiments :

" actin " Forward : 5'-TCCCTGGAGAAGAGCTACGA-3'
Reverse : 5'-AGGAAGGAAGGCTGGAAGAG -3'

" GAPDH " Forward : 5'-GAGTCAACGGATTTGGTCGT-3'

Reverse : 5’-GACAAGCTTCCCGTTCTCAG -3’

Thap1, RRM1, HCF-1, OGT : Quantitect primers assays have been purchased from Qiagen (sequences are not provided by Qiagen)

Supplemental figure legends

Fig. S1. THAP3, HCF-1 and OGT co-sediment on glycerol gradients. (A) Nuclear extracts from Hela cells expressing the human THAP3-Flag/HA protein (t-THAP3) were fractionated on a 10 to 40% glycerol gradient. Fractions were analyzed by SDS-PAGE and western blotting with anti-HA mAb. The migration of molecular weights markers bovine serum albumin (60 kDa) and thyroglobulin (670 kDa) is indicated. Fractions 3-7 corresponding to complex 1 (C1) were pooled for subsequent immunoprecipitation with anti-Flag antibody. (B) HCF-1 and OGT co-sediment with t-THAP3 in low (about 0,6 MDa) and high (about 2MDa) molecular-weight complexes. Nuclear extracts from t-THAP3 Hela cells were fractionated on a 10 to 40% glycerol gradient and collected fractions were immunobloted with anti-OGT and anti-HCF-1 antibodies. (C) t-THAP3 co-immunoprecipitated proteins from uninduced (mock) and induced cells were separated by SDS-PAGE and immunoblotted with the indicated antibodies.

Fig. S2. O-glycosylation of HCF1. Peptides ions carrying a labile post-translational modification of mass 203.07937, potentially corresponding to an N-acetyl-hexosamine, were detected by Mascot database searching for the protein HCF-1 (Uniprot accession number P51610). The table shows the position of the detected peptides in the protein sequence, and for each peptide ion: the mass to charge ratio (m/z), charge (z), theoretical mass, experimental mass, calculated mass deviation in ppm, Mascot identification score, Mascot rank (sequences of rank 1 correspond to the best matching sequence for a given MS/MS spectrum), peptide sequence, post-traductional modification assigned by Mascot, integrated elution peak area (reflecting the abundance of the peptide ion), and elution time from the nanoLC column. As HCF-1 was identified in various bands cut from the migration lanes of the THAP3 immunopurified complexes, MS/MS identification data and peak area are given only for the band in which a given glycosylated peptide was identified with the best Mascot score. For each detected glycosylated peptide ion, the MS signal of its non-glycosylated counterpart ion was also extracted to evaluate approximately the level of glycosylation. In some cases, the corresponding non-glycosylated peptide ion of same charge was not sequenced (N.S.), but MS/MS identification data is given for ions of lower or higher charge state, if available. Moreover, in some cases, no MS signal could be detected at all for non-glycosylated species (N.D.). Although some species assigned by Mascot as glycosylated peptides were identified with a very low score, sequences shown in this table are proposed to be glycosylated based on the following criteria: i) detection of a peptide ion bearing a modification of 203.07937 Da with high mass accuracy (mass deviation <4ppm) on a high resolution Orbitrap mass spectrometer, ii) detection in the MS/MS spectrum of an intense ion corresponding to a loss of 203.1 Da from the parent ion, characteristic of a labile O-glycosylation lost under CID (collision-induced dissociation) MS/MS sequencing, and iii) detection of the non-glycosylated peptide ion with a shift in elution time of about 1-2 min from the glycosylated peptide, due to different migration behaviour on the reverse-phase nanoLC column. (N.B.: for species assigned by Mascot as doubly glycosylated peptides, two elution times are indicated in some cases due to the detection of two very close but distinct elution peaks, possibly related to heterogeneous patterns of glycosylation on the peptide). Doubly-glycosylated peptides are shown in red, singly-glycosylated peptides in orange, non-glycosylated peptides in yellow.

Fig S3. Evolutionary conservation of the HBMs in THAP-zf proteins. (A) The primary structure of the eight human THAP-zf proteins containing consensus HBMs and their putative orthologues in zebrafish and xenopus is shown above the alignment of their HBMs. The THAP-zf is indicated in grey and the predicted coiled-coil domains in dark grey. Red asterisk indicates the position of the HBM in each primary sequence. Residues underlined in red correspond to the first residues of the predicted coiled-coil domains. (B) Multiple alignment of the HBMs found in THAP-zf proteins. The alignment was generated with Clustal W (http://www.ebi.ac.uk/cgi-bin/clustalw) according to the Blosum matrix and colored with Boxshade

(http://www.ch.embnet.org). Black boxes indicate identical residues, whereas shaded boxes show similar amino-acids. Dashed lines represent gaps introduced to align sequences. The most represented amino acids are shown in the consensus line.

Fig. S4. Identification of human THAP-zf protein cDNAs in large scale two-hybrid screens with an HCF-1 bait. HF7c yeast strain was transformed with pGAL29 encoding the HCF-1 kelch domain (amino acids 3-455) fused to the GAL4 DNA binding domain. The resulting strain was transformed with a HeLa Matchmaker library (Clontech) and plated on SD-Leu/Trp/His media. DNA was isolated from His+/β-galactosidase+ clones and the nucleotide sequence was determined. Of the positive cDNA clones, 14 encoded THAP proteins as detailed in the Table. The number of times a clone was isolated encoding the indicated THAP protein is shown (#). The portion of the THAP-zf protein encoded in various isolated clones is indicated.

Fig. S5. Endogenous THAP1 recruits endogenous HCF-1 to the *RRM1* promoter in proliferating HUVECs. (A) Schematic representation of the human *RRM1* promoter. The binding sites for THAP1 (THABS) and E2Fs are indicated. The position of the DNA fragments analyzed in ChIP-qPCR assays are shown. (B) Knockdown of THAP1 reduces recruitment of THAP1 and HCF1 to the *RRM1* promoter *in vivo*. ChIP-qPCR assays with anti-THAP1, anti-HCF1 or ChIP control (Abcam) rabbit antibodies were performed using proliferating primary HUVECs treated with control or THAP1 siRNAs. Immunoprecipitated DNA was quantified in triplicate by qPCR using the % of input method. In brief, the amount of genomic DNA co-precipitated with anti-THAP1 antibodies was calculated as a % of total input the following way: $\Delta CT=CT(input)-CT(THAP1\text{-}IP)$, $\%input=2^{\Delta CT}x0.25$ (0.25% of total input was used). These independent experiments provide additional evidence for an important role of THAP1 in HCF-1 recruitment to the *RRM1* promoter *in vivo* (see Figure 5 of the manuscript).

Fig. S6. Knockdown of endogenous OGT in primary HUVECs does not modify *RRM1* mRNA levels. (A,B) Knockdown of endogenous OGT was performed with 4 individual ON-TARGET-plus OGT siRNAs. (A) OGT and Tubα (loading control) expression levels were analyzed by western blot. (B) RNA was isolated from cells transfected with individual OGT siRNAs, 48h after siRNA transfection, and used for qPCR analysis with the indicated human gene primers (*OGT, RRM1, HCF-1, THAP1* and control gene *GAPDH*). *Actin* was used as a control gene for normalization. Results are shown as means with s.d. from 3 separate datapoints.

Figure S1

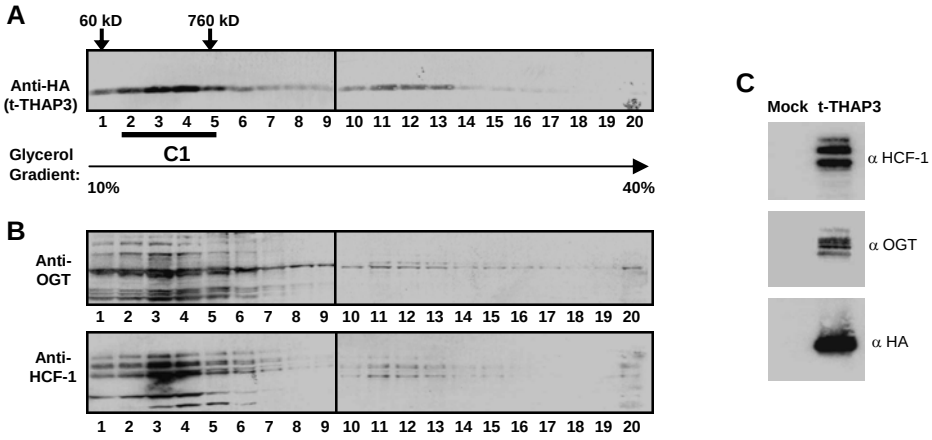

A
60 kD
760 kD
Anti-HA
(t-THAP3)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
C1
Glycerol
Gradient:
10%
40%
B
Anti-
OGT
Anti-
HCF-1
C
Mock t-THAP3
α HCF-1
α OGT
α HA

Figure S2

Position	m/z	z	Theo. Mass	Exp. Mass	delta (ppm)	Score	Rank	Sequence	PTM	Peak Area	elution time
400-426	959.8239	3	2876.4444	2876.4499	1.91	30	1	YDIPATAATATSPTPNPVPSVPANPPK	HexNAc	64176	25.2
	892.1309	3	2673.3650	2673.3709	2.22	N.S.	N.S.	YDIPATAATATSPTPNPVPSVPANPPK		978981	26.8
	1337.6920	2	2673.3650	2673.3694	1.67	63	1	YDIPATAATATSPTPNPVPSVPANPPK			
489-511	830.0963	3	2487.2639	2487.2671	1.26	29	1	VTGPQATTGTPLVTMoxRPASQAGK	HexNAc	142617	19.8
	762.4044	3	2284.1846	2284.1866	0.88	18	1	VTGPQATTGTPLVTMoxRPASQAGK		375210	20.3
512-524	735.9127	2	1469.8090	1469.8108	1.27	9	1	APVTVTSLPAGVR	HexNAc	302271	21.5 / 21.8
	634.3732	2	1266.7296	1266.7302	0.51	51	1	APVTVTSLPAGVR		1121128	23.1
579-594	1006.0110	2	2010.0079	2010.0074	-0.22	9	1	TMoxAVTPGTTTLPATVK	2 HexNAc	191864	21.7 / 22.7
	904.4720	2	1806.9285	1806.9294	0.52	46.01	1	TMoxAVTPGTTTLPATVK	HexNAc	641204	23.3
	802.9332	2	1603.8491	N.D.	N.D.	N.S.	N.S.	TMoxAVTPGTTTLPATVK		N.D.	N.D.
612-637	1002.1888	3	3003.5361	3003.5452	3.03	9	4	TAAAQVGTSVSSATNTSTRPIITVHK	2 HexNAc	138768	18.1 / 19.2
	934.4946	3	2800.4567	2800.4620	1.89	N.S.	N.S.	TAAAQVGTSVSSATNTSTRPIITVHK	HexNAc	160607	18.8 / 19.8
	866.8010	3	2597.3773	2597.3812	1.49	N.S.	N.S.	TAAAQVGTSVSSATNTSTRPIITVHK		57376	20.6
638-659	1268.6730	2	2535.3280	2535.3314	1.35	45	1	SGTVTVAQQAQVVTTVVGGVTK	2 HexNAc	31763	21.9
	1167.1316	2	2332.2486	N.D.	N.D.	N.S.	N.S.	SGTVTVAQQAQVVTTVVGGVTK	HexNAc	N.D.	N.D.
	1065.5919	2	2129.1693	N.D.	N.D.	N.S.	N.S.	SGTVTVAQQAQVVTTVVGGVTK		N.D.	N.D.
683-713	1135.2730	3	3402.7917	3402.7972	1.62	58	1	VMoxSVVQTKPVQTSAVTGQASTGPVTQIIQTK	HexNAc	66822	24.9
	1067.5810	3	3199.7123	3199.7212	2.78	58	1	VMoxSVVQTKPVQTSAVTGQASTGPVTQIIQTK		3782	25.8
771-793	1210.6340	2	2419.2516	2419.2534	0.76	57	1	TIPMoxSAIITQAGATGVTSSPGIK	HexNAc	44428	24.9
	1109.0934	2	2216.1722	N.D.	N.D.	N.S.	N.S.	TIPMoxSAIITQAGATGVTSSPGIK		N.D.	N.D.
794-802	588.8403	2	1175.6649	1175.6660	0.96	5	4	SPITIITTK	HexNAc (T)	1854574	25.0
	487.2999	2	972.5855	972.5852	-0.30	44	1	SPITIITTK		114340	27.4
856-875	1113.1740	2	2224.3294	2224.3334	1.81	17	1	LVTPVTVSAVKPAVTTLVVK	HexNAc	182824	28
	1011.6345	2	2021.2500	2021.2544	2.18	N.S.		LVTPVTVSAVKPAVTTLVVK		39126	30.5
	674.7590	3	2021.2500	2021.2552	2.54	44	1	LVTPVTVSAVKPAVTTLVVK			
1233-1244	496.5697	3	1486.6834	1486.6873	2.58	9	2	HSHAVSTAAMoxTR	HexNAc	42294	8.9
	428.8763	3	1283.6041	1283.6071	2.34	N.S.		HSHAVSTAAMoxTR		86875	9.3
	642.8110	2	1283.6041	1283.6074	2.64	46	1	HSHAVSTAAMoxTR			
1483-1500	985.0272	2	1968.0416	1968.0398	-0.88	34	1	AVTTVTQSTPVPGPSVPK	HexNAc (S)	107656	22.7
	883.4901	2	1764.9622	1764.9656	1.92	38	1	AVTTVTQSTPVPGPSVPK		300096	24.2

Figure S3

B

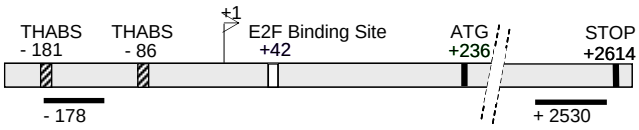
```
hThap1       118 --LLMPPLQTPVNLSVFCDHNYTVEDTMHQ--- 145
xThap1       118 --LLLPPLYTPSHIAVICDHNYTVEDTVHQ--- 145
zThap1       129 -EQCVPPGGQEAG-AVSCDHNYTLEDCVQQ--- 156
zThap3       133 --NNEVADCQSTKVFAAADHTYALSDACTT--- 160
zThap6       135 -VNKETTQKQETKPLPNVDHSYALPADLPGL-- 164
hThap11      227 -PQLVVVGEEGFPD-TGSDHSYSLSSGTTEEEL 253
xThap11      212 --DEPYPSADCWPSPLCSDHSYSVSSGTTTEEL 242
zThap11      167 AVHESCSVAD-FAL-CGADHSYSLTTGTTSTEL 197
hThap7       112 -SAYMLRLPPPAGAYIQNEHSYQVGSALLW--- 140
xThap7       343 -SVYMMRLPQPPGAYIQNEHSYHVGNALLWKR- 373
zThap7       195 -SCYMKRLPPPPGFYLPKEHSYAQLCPLVWRK- 225
hThap5       305 -DSLYKDVDYGT-EVLQIEHSYCRQDINKEHL- 334
xThap5       314 -LGFENDAQSTS-EILESEHSYCRQTTDRHYL- 343
hThap9       106 -KILKQPLPDNSQEVATEDHNYSLKTPLTIG-- 135
hThap3       161 --PTGPAGLRRTPNKQPSDHSYALLDLDSL--- 188
zTHAP2       273 ----TPATEPDPLESSDSDHPYALNPVKAK--- 298
zThap9       174 -SYATLSSTHETPKFIKEDHSYNLASPRSLK-- 203
hThap2       109 -SNLKSNISSQQ---VLLEHSYAFRNPMEA--- 134
hThap6       122 -LVGASSCIEEFQSQFIFEHSYSVMDSPKKL-- 151
consensus                      v  dHsY v
```

Figure S4

Protein	#	amino acids
THAP3	3	128-239 153-239
THAP4	1	291-577
THAP7	5	223-309
THAP11	5	73-313 203-313 206-313 216-313

Figure S5

A


+1
THABS
- 181
THABS
- 86
E2F Binding Site
+42
ATG
+236
STOP
+2614
- 178
+ 2530

B

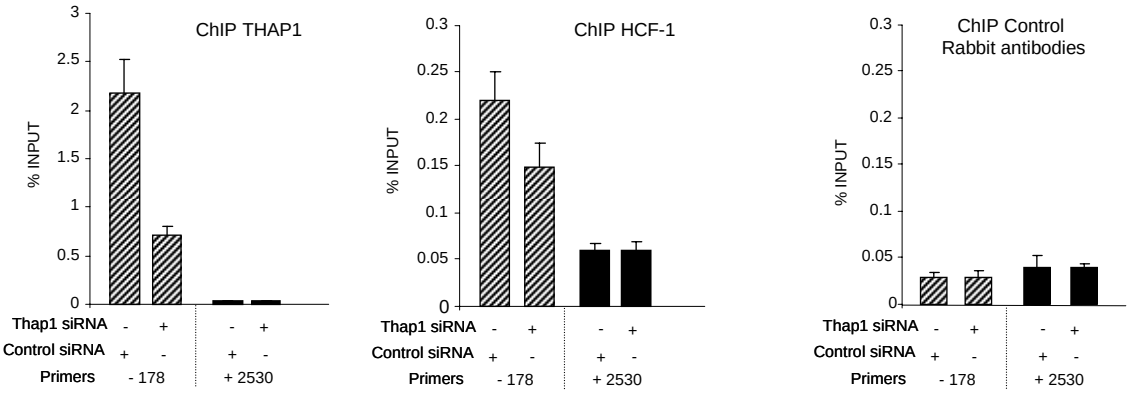
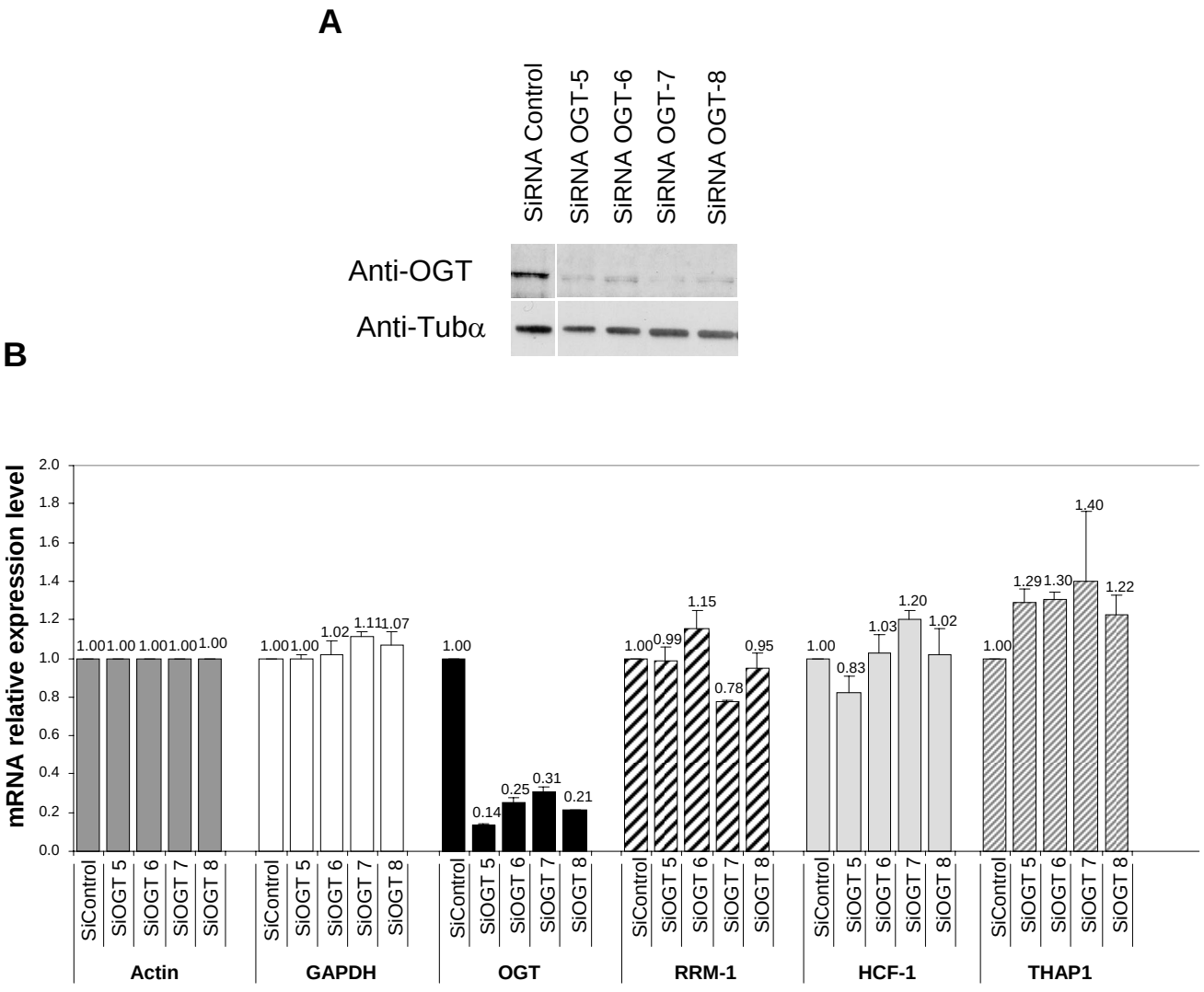

ChIP THAP1
% INPUT
3
2.5
2
1.5
1
0.5
0
Thap1 siRNA - + - +
Control siRNA + - + -
Primers - 178 + 2530
ChIP HCF-1
% INPUT
0.3
0.25
0.2
0.15
0.1
0.05
0
Thap1 siRNA - + - +
Control siRNA + - + -
Primers - 178 + 2530
ChIP Control
Rabbit antibodies
% INPUT
0.3
0.25
0.2
0.15
0.1
0.05
0
Thap1 siRNA - + - +
Control siRNA + - + -
Primers - 178 + 2530

Figure S6


A
SiRNA Control
SiRNA OGT-5
SiRNA OGT-6
SiRNA OGT-7
SiRNA OGT-8
Anti-OGT
Anti-Tubα
B
mRNA relative expression level
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
1.00 1.00 1.00 1.00 1.00
1.00 1.00 1.02 1.11 1.07
1.00 0.14 0.25 0.31 0.21
1.00 0.99 1.15 0.78 0.95
1.00 0.83 1.03 1.20 1.02
1.00 1.29 1.30 1.40 1.22
SiControl
SiOGT 5
SiOGT 6
SiOGT 7
SiOGT 8
Actin
GAPDH
OGT
RRM-1
HCF-1
THAP1