Cell Research (2011) 21:1013-1027.
© 2011 IBCB, SIBS, CAS All rights reserved 1001-0602/11 $ 32.00
www.nature.com/cr

ORIGINAL ARTICLE

Human telomeric proteins occupy selective interstitial sites

Dong Yang[2, *], Yuanyan Xiong[1, *], Hyeung Kim[2, *], Quanyuan He[2], Yumei Li[3], Rui Chen[3], Zhou Songyang[1, 2]

[1]State Key laboratory for Biocontrol, Sun Yat-Sen University, Guangzhou 510275, China; [2]Verna and Marrs Mclean Department of Biochemistry and Molecular Biology, [3]Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

Human telomeres are bound and protected by protein complexes assembled around the six core telomeric proteins RAP1, TRF1, TRF2, TIN2, TPP1, and POT1. The function of these proteins on telomeres has been studied extensively. Recently, increasing evidence has suggested possible roles for these proteins outside of telomeres. However, the non-canonical (extra-telomeric) function of human telomeric proteins remains poorly understood. To this end, we systematically investigated the binding sites of telomeric proteins along human chromosomes, by performing whole-genome chromatin immunoprecipitation (ChIP) for RAP1 and TRF2. ChIP sequencing (ChIP-seq) revealed that RAP1 and TRF2 could be found on a small number of interstitial sites, including regions that are proximal to genes. Some of these binding sites contain short telomere repeats, suggesting that telomeric proteins could directly bind to interstitial sites. Interestingly, only a small fraction of the available interstitial telomere repeat-containing regions were occupied by RAP1 and TRF2. Ectopically expressed TRF2 was able to occupy additional interstitial telomere repeat sites, suggesting that protein concentration may dictate the selective targeting of telomeric proteins to interstitial sites. Reducing RAP1 and TRF2 expression by RNA interference led to altered transcription of RAP1- and TRF2-targeted genes. Our results indicate that human telomeric proteins could occupy a limited number of interstitial sites and regulate gene transcription.

***Keywords*:** telomere; extra-telomeric; RAP1; TRF2; ChIP; ChIP-sequencing

Cell Research (2011) **21**:1013-1027. doi:10.1038/cr.2011.39; published online 22 March 2011

Introduction

The ends of human chromosomes – telomeres – consist of long repetitive sequences of (TTAGGG)n that are protected and maintained by the telomerase and networks of protein complexes [1, 2]. Telomere biology is intimately linked to genome stability, aging, and cancer [3, 4]. The dynamic interactions between the telomerase and its associated proteins, core telomere binding factors, and various factors and modification enzymes are key to telomere homeostasis [5-7]. Six core telomeric proteins (RAP1, TRF1, TRF2, TIN2, TPP1, and POT1) have been shown to be essential for maintaining telomere length and protect telomere integrity [8-10]. Of these six proteins, TRF1 and TRF2 were originally cloned as the major telomere double-stranded DNA (dsDNA) binding factors [11-14], and are crucial to maintaining telomere length and end protection [15-17]. In fact, both proteins can homodimerize for telomere DNA binding through their respective myb domains [11]. Mammalian POT1, on the other hand, was cloned based on its sequence homology to yeast cdc13 and shown to specifically bind the 3′ single-stranded telomere overhangs [18, 19]. Through direct interactions with these telomere-binding proteins, RAP1, TPP1, and TIN2 are also targeted to the telomeres for telomere maintenance and regulation [20-33]. Unlike its yeast orthologue, mammalian RAP1 lacks the ability to bind telomere DNA despite the presence of a myb domain [20]. Instead, the telomeric recruitment and stability of RAP1 is dependent on TRF2 [20]. RAP1 is also involved in regulation of telomere length [20, 27, 28]. Interestingly, TRF2, but not RAP1, was required for the inhibition of non-homologous end joining at the telomeres, as depletion of TRF2 leads to telomere fusions [15, 34]. However, RAP1 is essential for the inhibition of homology-directed repair at telomeres [25, 26].

*These three authors contributed equally to this work.
Correspondence: Zhou Songyang
E-mail: songyang@bcm.edu
Received 3 November 2010; revised 4 January 2011; accepted 6 January 2011; published online 22 March 2011

The primary focus for mammalian telomeric proteins has been on their function in telomere maintenance. Increasing evidence, however, suggests a broader role for these proteins outside of the telomeres. For example, human TIN2, TPP1, and POT1 have been shown to localize and interact in the cytoplasm [35]. Furthermore, TRF2 has been implicated in regulating the proliferation and differentiation of neural tumor and stem cells, where TRF2 interacts with the repressor element 1-silencing transcription factor (REST) in PML-nuclear bodies and protects REST from proteosomal degradation [36]. Notably, while the myb domain of TRF2 recognizes telomeric repeat sequences, its basic domain can bind DNA junctions in a telomere sequence-independent manner [37]. In fact, TRF2 appears to play an important role in homologous recombination repair of double-strand breaks at non-telomeric regions [38].

Studies of RAP1 in different organisms have underlined the role of RAP1 in the regulation of gene expression. Budding yeast Rap1 (scRap1) binds directly to telomeric DNA and controls subtelomeric silencing [39, 40]. scRap1 also binds to many gene loci and regulates the transcription of genes that encode proteins from diverse pathways, including ribosomal proteins, glycolytic enzymes, and mating-type factors [41-43]. *Trypanosoma brucei* Rap1 (tpRap1) is a critical transcription suppressor for silencing variant surface glycoprotein genes located within subtelomeric regions [44]. Recently, human RAP1 was found to associate with IκB kinases in the cytoplasm and act as a crucial regulator of NF-κB-modulated gene expression [45]. Furthermore, by comparing the data from RAP1-deficient and wild-type mouse embryonic fibroblasts, chromatin immunoprecipitation (ChIP)-seq analysis revealed that extra-telomeric RAP1 binding sites were enriched in subtelomeric regions and in genes deregulated as a result of RAP1 deletion [26]. However, little was known about the extra-telomeric binding activity of human RAP1.

Here we report our systematic investigation of the binding sites of telomeric proteins RAP1 and TRF2 along human chromosomes. Using anti-RAP1 and TRF2 antibodies, we performed whole-genome ChIP coupled with high-throughput sequencing in human cells. Our analysis found that both TRF2 and RAP1 occupy a limited number of interstitial regions throughout the human genome and regulate gene expression.

Results

Identification of genome-wide RAP1 and TRF2 binding sites by ChIP-seq

To understand the extra-telomeric function of telomeric proteins, we performed ChIP analysis to identify chromosomal binding sites of telomeric proteins in human cells. First, anti-RAP1 and TRF2 antibodies were tested for their ability to pull down telomere protein-DNA complexes in ChIP experiments using HTC75 cells. As shown in Figure 1A, both antibodies were able to specifically and efficiently immunoprecipitate telomere DNA. We then used these antibodies for whole-genome ChIP-seq experiments, where the immunoprecipitated DNA was recovered and sequenced by Solexa technology (Figure 1B).

Our ChIP-seq experiments yielded > 2.0×10^7 uniquely mapped short reads for RAP1, and > 1.4×10^7 reads for TRF2. Short reads from IgG (> 3.1×10^7) were used as controls for RAP1 and TRF2 peak detection (Table 1). Consistent with telomere targeting of RAP1 and TRF2, both RAP1 and TRF2 data sets are highly enriched for telomeric sequences. For example, DNA fragments that contain three or more TTAGGG repeats could be specifically brought down by anti-RAP1 and TRF2 antibodies, and sequences that contain more than six TTAGGG repeats were > 20-fold more abundant compared to IgG controls (Figure 1C). Although the majority of the precipitated DNA could be mapped to telomeric and subtelomeric regions, a small number of interstitial sites were also discovered (Table 2). For example, the gene locus of AC013473.1 appeared to be occupied by both RAP1 and TRF2, but not IgG control (Figure 1D). There were a total of 78 and 77 interstitial sites for RAP1 and TRF2, respectively, indicating that telomeric proteins can indeed bind to chromosomal regions other than the telomeres in human cells.

The majority of the interstitial sites appear to be located within 5 kb of gene loci (73% for RAP1 and 58% for TRF2) (Figure 2A and 2B). The RAP1 sites are situated in the exon (8%), intron (31%), and UTR (34%) regions of the genes (Figure 2A). In total, there are 63 and 50 genes, respectively, in the vicinity of RAP1 (Table 3) and TRF2 binding sites (Table 4). An example is the *CLIC6* gene, where one RAP1 binding site was mapped to its intronic region (~3 kb downstream of the sixth exon) (Figure 2C). Genes from diverse pathways appear to be targets of RAP1 and TRF2, including membrane proteins such as CLIC6 and PLXNB2, and signaling molecules such as PAK2 and VAV3 (Tables 3 and 4). The enrichment of RAP1 and TRF2 binding sites to these gene loci suggests that RAP1 and TRF2 may be targeted to these genes for regulation.

A recent RAP1 ChIP-seq analysis using mouse cells reported 30 398 RAP1 binding sites and 8 687 RAP1 target genes [26]. Surprisingly, we found no overlap in RAP1 sites between the two data sets (Figure 2D). Of

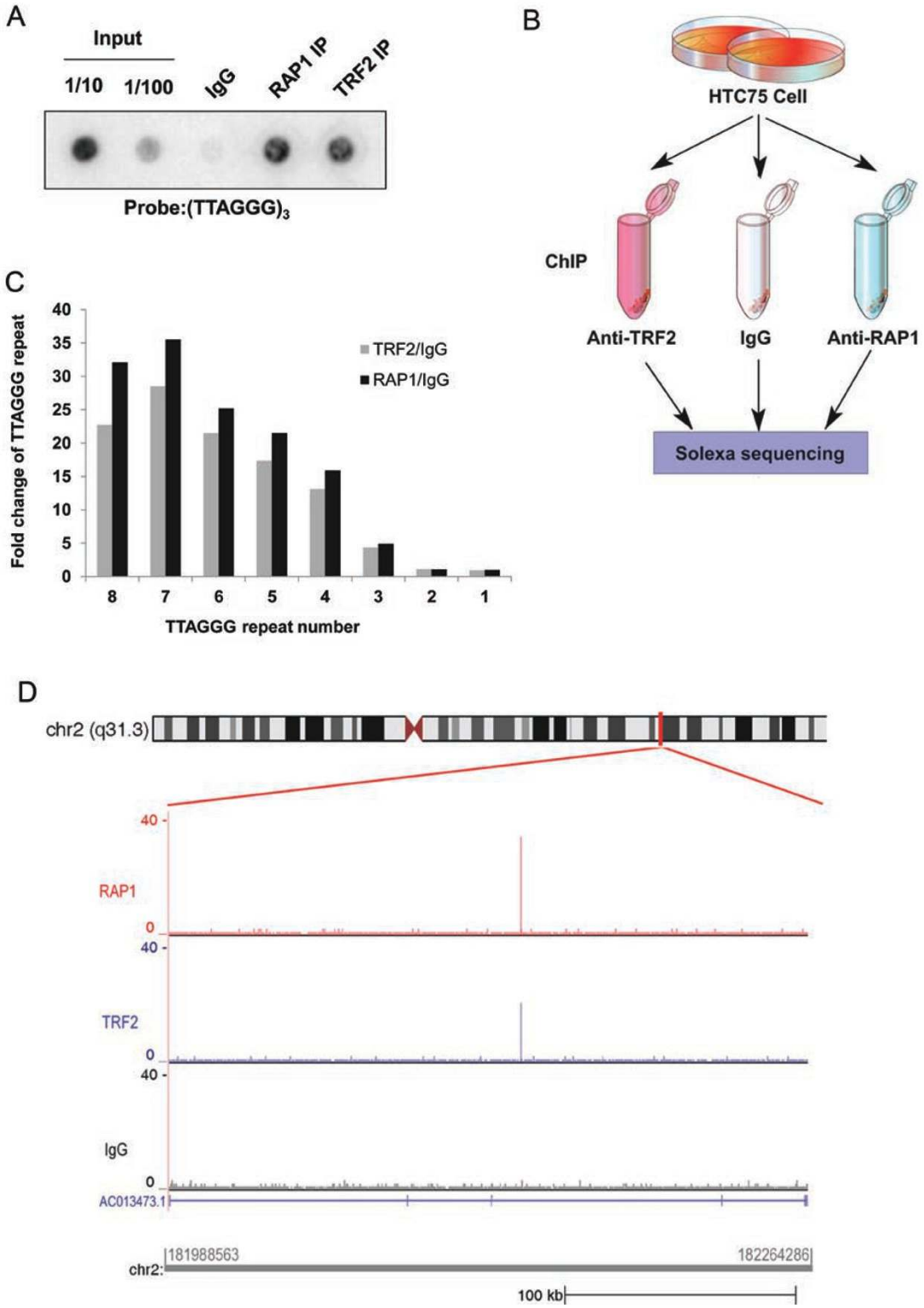


Figure 1 Identification of RAP1 and TRF2 binding sites by ChIP-seq. **(A)** HTC75 cells were crosslinked and immunoprecipitated using anti-RAP1 and TRF2 antibodies. The co-precipitated DNA was analyzed by dot-blotting with the indicated telomere probe. Rabbit IgG was used as negative control. **(B)** Flowchart for whole-genome ChIP-seq experiments. **(C)** Enrichment of telomere sequences. The relative abundance of TTAGGG repeats of various lengths in RAP1 and TRF2 ChIP-seq data was compared to the IgG control. **(D)** A ChIP-seq histogram showing RAP1 and TRF2 interstitial binding sites that were found near gene AC013473.1 on chromosome 2.

Table 1 Summary of ChIP-seq short reads

	IgG (1)*	IgG (2)#	RAP1 (1)*	RAP1 (2)#	TRF2*
Total raw reads	22780234	33033567	20701392	11422526	23050684
Reads mapped to genome	19968347 (87.66%)	17622072 (53.34%)	18294147 (88.37%)	6312555 (55.26%)	16618407 (72.10%)
Reads uniquely mapped to genome	17302199 (75.95%)	14225858 (43.0%)	15700261 (75.84%)	5031792 (44.05%)	14302578 (62.05%)

*Denotes a short read length of 50 bp. # denotes a short read length of 35 bp. Numbers in parentheses indicate independent ChIP experiments.

Table 2 Distribution of RAP1 and TRF2 binding sites

Location	RAP1	TRF2
Subtelomeric	20	15
Subtelo-exon	1	0
Subtelo-intron	3	2
Subtelo-3′UTR	9	5
Subtelo-5′UTR	5	2
Exon	6	4
Intron	24	17
3′UTR	13	15
5′UTR	14	9
Not defined	21	32
SUM	116	101

the potential mouse RAP1 targets, 7 521 are estimated to have human orthologues. However, we found only 16 genes to be human RAP1 targets as well. This number is too small to be of statistical significance, and further indicates the lack of overlap between human and mouse RAP1 sites. Taken together, these findings highlight the differences in RAP1 interstitial binding sites between human and mouse.

Confirmation of the extra-telomeric binding sites for RAP1 and TRF2

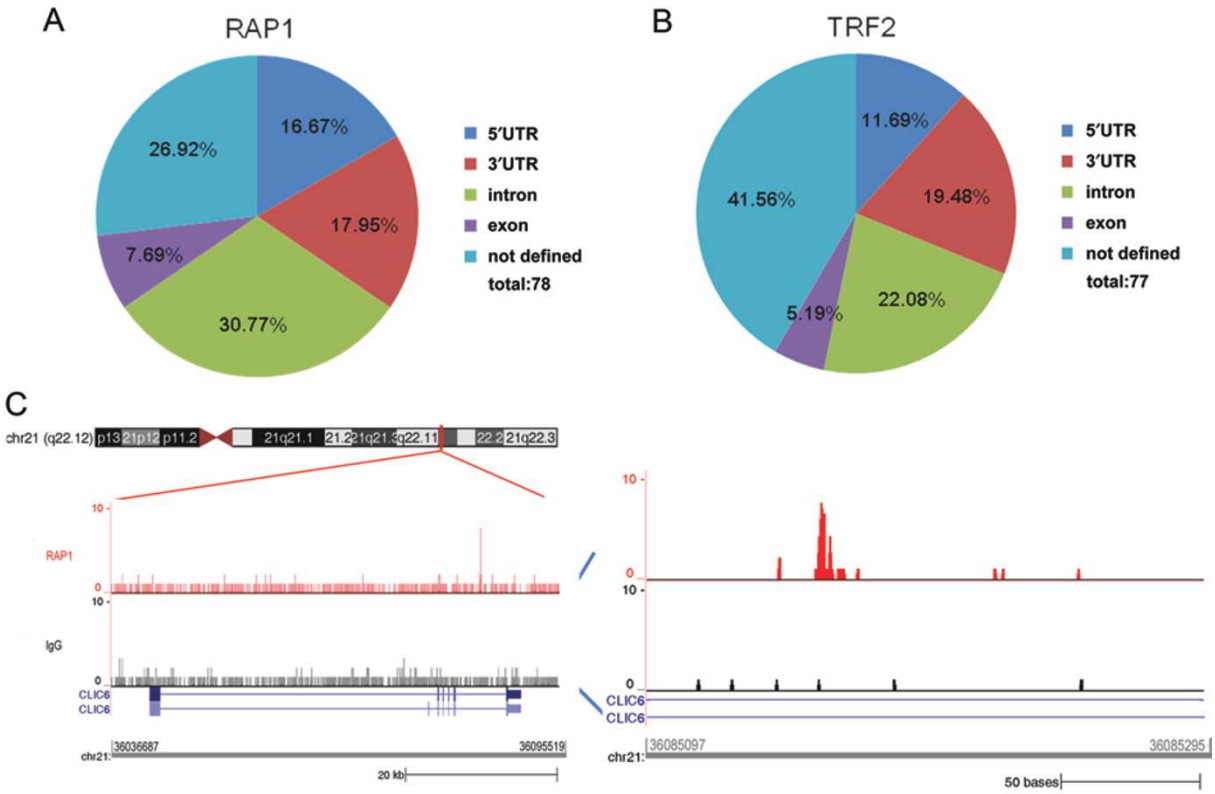
Next, we carried out secondary screens to confirm our ChIP-seq results. To rule out the possibility of antibody cross-reactivity, human HTC75 cells expressing FLAG-tagged RAP1 or TRF2 were generated. The ectopically expressed TRF2 and RAP1 proteins retained their ability to target to the telomeres (Supplementary information, Figure S1). Anti-FLAG ChIP experiments were then carried out using extracts from these cells. The precipitated DNA was then purified for quantitative PCR (qPCR) analysis using primer pairs that were derived from the target sites identified in ChIP-seq experiments (Supplementary information, Table S1). Consistent with our whole-genome ChIP-seq data, FLAG-RAP1 and TRF2 were indeed enriched on the extra-telomeric target sites examined here (Figure 3A and 3B). In comparison, we observed no enrichment at the region that was negative in ChIP-seq analysis (negative chr2, Figure 3A and 3B). Similar experiments were carried out for endogenous RAP1 and the results were consistent as well (data not shown). These findings indicate that RAP1 and TRF2 can associate with specific interstitial sites on human chromosomes.

RAP1 and TRF2 can occupy sites with or without telomere repeat sequences

RAP1 and TRF2 can bind to each other and form a heterodimer that is anchored on telomeres through the direct binding of TRF2 to telomere dsDNA [13, 20]. We therefore analyzed whether interstitial RAP1 and TRF2 sites contain similar sequence motifs. Indeed, one class of the interstitial site motifs contains the TTAGGG repeat (Figure 4A). Here, 12 of the 78 RAP1 sites match the telomere repeats, suggesting direct binding of the RAP1-TRF2 heterodimer to these sites. When the TTAGGG repeat-containing sites were excluded from the analysis, novel motifs for RAP1 and TRF2 binding emerged. For instance, 15 RAP1-binding sites share the motif CC[AC]T[TG][CT]C[AT]T[TC]CC (Figure 4B), while a closely related motif CCATTCC[AT]TTCC is shared by 14 of the 77 TRF2 sites (Figure 4C), suggesting possible co-occupancy of both RAP1 and TRF2 on these sites. Interestingly, a fraction of TRF2 sites do not overlap with RAP1 sites, raising the possibility that RAP1 and TRF2 can be recruited to distinct regions of chromosomes. The appearance of non-TTAGGG-containing motifs for RAP1 and TRF2 further suggests that TRF2 and RAP1 either possess binding capacity for non-telomeric DNA sequences, or can localize to interstitial sites through additional proteins.

Selective occupancy of interstitial telomere repeat-containing sites by TRF2

Our results thus far indicate that RAP1 and TRF2 are



D

	Binding sites (#)	Associated gene loci (#)
mouse	30398	8687
human	116	63
both	0	16

Figure 2 Human RAP1 and TRF2 are able to bind interstitial sites. Gene location and distribution of RAP1 **(A)** and TRF2 **(B)** binding sites. Interstitial binding sites were categorized based on their relative locations to gene loci. UTR is defined as sequences within 5 kb upstream of the first exon or 5 kb downstream of the last exon. Not defined, sites > 5 kb away from annotated genes. **(C)** RAP1 binding site on *CLIC6*. Left, RAP1 and IgG ChIP-seq reads were mapped to *CLIC6* on chromosome 2. Right, RAP1 binding sites in the intron region of *CLIC6*. **(D)** Differential binding of RAP1 in human vs. mouse cells. RAP1 ChIP-seq data from human and mouse were compared.

capable of occupying interstitial sites that contain TTAGGG repeat sequences. The small number of binding sites identified here prompted us to examine the abundance of such binding sites in the human genome. Based on the target site sequences for RAP1 and TRF2, we deduced a minimal telomere-containing pattern that was named T2N100. T2N100 represents two TTAGGGTTAG sequences that are separated by 0-100 nucleotides. This motif is consistent with the structure of the TRF2 myb domain in complex with telomere DNA [46], in which a single TRF2 myb domain recognizes TTAGGGTTA. It also took into consideration findings that TRF1 and TRF2 can bind telomere DNA as homodimers where they recognize two TTAGGGTTA sites that are separated by a spacer of varying length [11, 47, 48]. Scanning of the human genome using the T2N100 motif revealed ~300 potential binding sites for TRF2, an example of which is shown in Figure 4D. This number is much higher than the number of interstitial telomeric-repeat sites we observed in the RAP1 and TRF2 ChIP-seq experiments, suggesting a selective targeting mechanism for RAP1 and TRF2 extra-telomeric binding in human cells.

Selective association of RAP1 and TRF2 to interstitial sites may be achieved through a number of mechanisms.

Table 3 List of RAP1 binding sites (a total of 63 genes, including subtelomeric regions)

Ensemble gene ID	ChrID	Gene name	Gene Biotype	Peak location
ENSG00000125510	chr20	OPRL1	protein_coding	subtelo-intro: 62728335-62728451
ENSG00000120645	chr12	IQSEC3	protein_coding	subtelo-intro: 217211-217300
ENSG00000181031	chr17	RPH3AL	protein_coding	subtelo-intro: 165372-165400
ENSG00000237516	chr4	AC126281.1	protein_coding	subtelo-exon: 190988742-190988776
ENSG00000225143	chr12	AC215219.3	protein_coding	subtelo-5UTR: 95014-95744; subtelo-5UTR: 94640-94669; subtelo-5UTR: 95571-95611
ENSG00000188076	chr11	SCGB1C1	protein_coding	subtelo-5UTR: 189694-189886
ENSG00000232152	chr15	AC140725.2	protein_coding	subtelo-5UTR: 102521051-102521256
ENSG00000227232	chr1	WASH5P	protein_coding	subtelo-3UTR: 9938-10596
ENSG00000232852	chr20	AL137028.1	pseudogene	subtelo-3UTR: 62918210-62918436; subtelo-3UTR: 62918640-62918668
ENSG00000236438	chr3	FAM157A	protein_coding	subtelo-3UTR: 197900131-197900291; subtelo-3UTR: 197901195-197901472; subtelo-3UTR: 197900771-197900911
ENSG00000148408	chr9	CACNA1B	protein_coding	subtelo-3UTR: 141023379-141023650
ENSG00000181404	chr9	XXyac-YRM2039.2	protein_coding	subtelo-3UTR: 10151-10441
ENSG00000230388	chr7	AC215522.1	protein_coding	subtelo-3UTR: 10014-10151
ENSG00000155256	chr10	ZFYVE27	protein_coding	intro: 99509351-99509411
ENSG00000141568	chr17	FOXK2	protein_coding	intro: 80544211-80544240
ENSG00000157637	chr17	SLC38A10	protein_coding	intro: 79223169-79223198
ENSG00000162341	chr11	TPCN2	protein_coding	intro:68850331-68850371
ENSG00000145536	chr5	ADAMTS16	protein_coding	intro: 5259211-5259271
ENSG00000196576	chr22	PLXNB2	protein_coding	intro: 50744871-50745111
ENSG00000237562	chr10	BX322613.1	protein_coding	intro: 42394243-42394271
ENSG00000235358	chr1	AC093151.1	processed_transcript	intro: 41747558-41747760
ENSG00000105221	chr19	AKT2	protein_coding	intro: 40788271-40788331
ENSG00000159212	chr21	CLIC6	protein_coding	intro: 36085097-36085295
ENSG00000078900	chr1	TP73	protein_coding	intro: 3596004-3596033
ENSG00000204590	chr6	GNL1	protein_coding	intro: 30523591-30523651
ENSG00000108576	chr17	SLC6A4	protein_coding	intro: 28542411-28542571
ENSG00000080608	chr9	KIAA0020	protein_coding	intro: 2823981-2824104
ENSG00000182601	chr16	HS3ST4	protein_coding	intro: 25771377-25771488; intro: 25839491-25839519
ENSG00000205212	chr17	CCDC144NL	protein_coding	intro: 20771351-20771391
ENSG00000180370	chr3	PAK2	protein_coding	intro: 196521571-196521651
ENSG00000234663	chr2	AC013473.1	processed_transcript	intro: 182140515-182140671
ENSG00000004399	chr3	PLXND1	protein_coding	intro: 129296751-129296791
ENSG00000107902	chr10	LHPP	protein_coding	intro: 126283791-126283851
ENSG00000227483	chr5	AC114291.1	protein_coding	intro: 1258251-1258291
ENSG00000151929	chr10	BAG3	protein_coding	intro: 121419531-121419591
ENSG00000236138	chr3	RP11-413E6.1	protein_coding	exon: 75718321-75718349
ENSG00000111077	chr12	TENC1	protein_coding	exon: 53453591-53453711
ENSG00000198250	chr10	ANTXRL	processed_transcript	exon: 47667271-47667311
ENSG00000163814	chr3	CDCP1	protein_coding	exon: 45135071-45135131
ENSG00000237645	chr3	AC092967.1	pseudogene	exon: 173362411-173362471
ENSG00000074660	chr17	SCARF1	protein_coding	exon: 1543731-1543811

Table 3 List of RAP1 binding sites (continued)

Ensemble gene ID	ChrID	Gene name	Gene Biotype	Peak location
ENSG00000222855	chr21	SSU_rRNA_5	rRNA	5UTR: 9826612-9826664; 5UTR: 9826885-9826914
ENSG00000224764	chr9	AL353768.1	processed_transcript	5UTR: 97700213-97700241
ENSG00000231692	chr10	BX322613.7	protein_coding	5UTR: 42362527-42362593
ENSG00000084636	chr1	COL16A1	protein_coding	5UTR: 32172011-32172071
ENSG00000207985	chr20	hsa-mir-663	miRNA	5UTR: 26190281-26190392; 5UTR: 26188869-26189052; 5UTR: 26189843-26189871
ENSG00000222865	chr17	SSU_rRNA_5	rRNA	5UTR: 22020559-22020850
ENSG00000213123	chr3	TCTEX1D2	protein_coding	5UTR: 196047091-196047291
ENSG00000231938	chr8	AC105118.1	protein_coding	5UTR: 144462611-144462651
ENSG00000227984	chrY	AC069323.1	protein_coding	5UTR: 13142957-13142986
ENSG00000198763	chrM	AF347015.27	protein_coding	5UTR: 0-400; 3UTR: 8305-8334
ENSG00000143429	chr2	AC027612.8	pseudogene	3UTR: 91819811-91819931; intro: 91844791-91844831
ENSG00000130589	chr20	RP4-697K14.1	protein_coding	3UTR: 62186051-62186091
ENSG00000234054	chr4	AC119751.11	protein_coding	3UTR: 49659848-49659876
ENSG00000237511	chr4	AC119751.6	protein_coding	3UTR: 49659125-49659153
ENSG00000235525	chr20	AL162458.5	pseudogene	3UTR: 44600651-44600711
ENSG00000234738	chr10	BX322613.10	protein_coding	3UTR: 42376194-42376358
ENSG00000236623	chr10	AL590623.6	protein_coding	3UTR: 39108051-39108111
ENSG00000210195	chrM	AF347015.22	Mt_tRNA	3UTR: 16039-16067; 3UTR: 16257-635
ENSG00000206163	chrY	AC134882.1	protein_coding	3UTR: 13454205-13454234
ENSG00000146556	chr2	WASH2P	pseudogene	3UTR: 114360771-114360991
ENSG00000134215	chr1	VAV3	protein_coding	3UTR: 108113091-108113119
ENSG00000217801	chr1	RP11-465B22.2	pseudogene	3UTR: 1008991-1009031

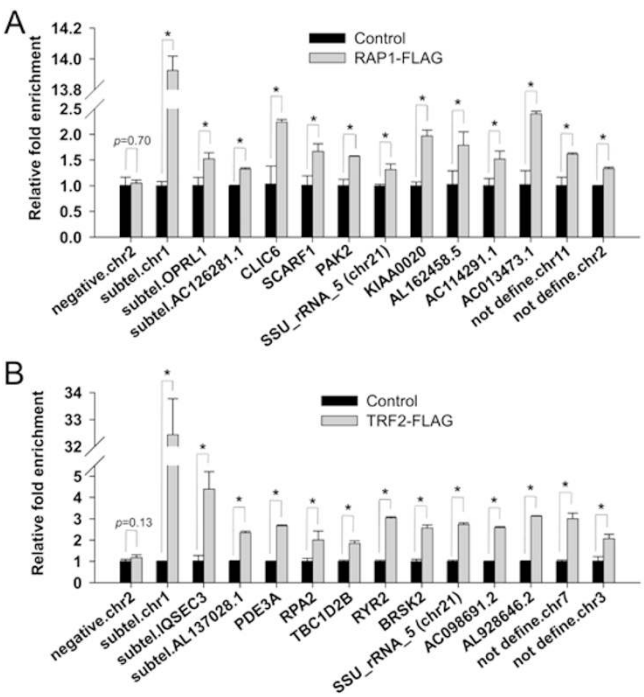


Figure 3 ChIP-qPCR analysis to confirm RAP1 and TRF2 binding sites. Anti-FLAG ChIP was carried out using HTC75 cells expressing **(A)** RAP1-FLAG or **(B)** TRF2-FLAG, and the precipitated DNA was analyzed by qPCR using primer pairs derived from the indicated RAP1 or TRF2 binding sites identified by ChIP-seq (Supplementary information, Table S1). Rabbit IgG served as control. Error bars indicate standard errors (n = 3). P values were calculated by Student's t test and * indicates P < 0.05.

Interaction with additional proteins may determine whether RAP1 and TRF2 are targeted to certain sites. Alternatively, the amount of RAP1 and TRF2 proteins may be a limiting factor. If the latter is true, increasing RAP1 or TRF2 expression should increase the number of interstitial binding sites targeted by these proteins. To test this possibility, we utilized the cells that overexpressed FLAG-tagged TRF2 and determined its occupancy by ChIP coupled with qPCR on a subset of binding sites that

Table 4 List of TRF2 binding sites (a total of 50 genes including subtelomeric regions)

Esemble Gene ID	ChrID	Gene name	Gene Biotype	Peak location
ENSG00000120645	chr12	IQSEC3	Protein_coding	subtelo-intro: 217158-217360
ENSG00000215585	chr18	AP001005.2	Protein_coding	subtelo-intro: 110421-110449
ENSG00000226880	chr10	AL732375.1	Pseudogene	subtelo-5UTR: 135500746-135500774
ENSG00000232152	chr15	AC140725.2	Protein_coding	subtelo-5UTR: 102521191-102521262
ENSG00000227232	chr1	WASH5P	Protein_coding	subtelo-3UTR: 9976-10335
ENSG00000232852	chr20	AL137028.1	Pseudogene	subtelo-3UTR: 62918584-62918780
ENSG00000236438	chr3	FAM157A	Protein_coding	subtelo-3UTR: 197900771-197900891
ENSG00000230388	chr7	AC215522.1	Protein_coding	subtelo-3UTR: 10015-10151
ENSG00000233991	chr2	AC116050.1	Pseudogene	intro: 91785991-91786031
ENSG00000141568	chr17	FOXK2	Protein_coding	intro: 80544181-80544210
ENSG00000167202	chr15	TBC1D2B	Protein_coding	intro: 78360751-78360871
ENSG00000137571	chr8	SLCO5A1	Protein_coding	intro: 70602322-70602510
ENSG00000223342	chr6	AL158817.1	Processed_transcript	intro: 6916511-6916911
ENSG00000179242	chr20	CDH4	Protein_coding	intro: 60507571-60507631
ENSG00000147853	chr9	AK3	Protein_coding	intro: 4713989-4714030
ENSG00000234501	chr10	AL590623.7	Protein_coding	intro: 39090771-39090891
ENSG00000198626	chr1	RYR2	Protein_coding	intro: 237766291-237766557
ENSG00000172572	chr12	PDE3A	Protein_coding	intro: 20704338-20704471
ENSG00000234663	chr2	AC013473.1	Processed_transcript	intro: 182140551-182140591
ENSG00000152061	chr1	RABGAP1L	Protein_coding	intro: 174766851-174767091
ENSG00000174672	chr11	BRSK2	Protein_coding	intro: 1472351-1472391
ENSG00000238159	chr1	AL583842.6	Protein_coding	intro: 142542047-142542075
ENSG00000181234	chr12	TMEM132C	Protein_coding	intro: 128958071-128958131
ENSG00000133808	chr11	MICALCL	Protein_coding	intro: 12377971-12378051
ENSG00000086848	chr11	FDXACB1	Protein_coding	intro: 111692639-111692720
ENSG00000222855	chr21	SSU_rRNA_5	rRNA	exon: 9827231-9827731
ENSG00000156127	chr14	BATF	Protein_coding	exon: 76013091-76013131
ENSG00000227124	chr3	ZNF717	Protein_coding	exon: 75788151-75788191
ENSG00000226958	chrX	AL928646.2	Pseudogene	exon: 108297476-108297505
ENSG00000232673	chr1	AC098691.2	Protein_coding	5UTR: 91852726-91852755
ENSG00000236839	chr7	AC069280.1	Processed_transcript	5UTR: 68654875-68654903
ENSG00000234739	chr4	AC119751.4	Protein_coding	5UTR: 49634368-49634396
ENSG00000236506	chr4	AC118282.15	Protein_coding	5UTR: 49105971-49105999
ENSG00000117748	chr1	RPA2	Protein_coding	5UTR: 28242561-28242652
ENSG00000100031	chr22	GGT1	Protein_coding	5UTR: 25001633-25001833
ENSG00000161055	chr5	SCGB3A1	Protein_coding	5UTR: 180021773-180021931
ENSG00000234643	chr21	AP003900.2	Pseudogene	5UTR: 11186419-11186447
ENSG00000237143	chr2	AC233264.9	Protein_coding	3UTR: 89878745-89878850
ENSG00000237511	chr4	AC119751.6	Protein_coding	3UTR: 49657799-49657828
ENSG00000026559	chr20	KCNG1	Protein_coding	3UTR: 49615591-49615631
ENSG00000229193	chr4	AC118282.22	Protein_coding	3UTR: 49117196-49117359
ENSG00000185186	chr21	C21orf84	Processed_transcript	3UTR: 44887751-44887791
ENSG00000236331	chr10	BX322613.9	Protein_coding	3UTR: 42366325-42366479
ENSG00000224084	chr10	AL590623.4	Protein_coding	3UTR: 39125526-39125554
ENSG00000200002	chr16	SSU_rRNA_5	rRNA	3UTR: 33963631-33964091
ENSG00000201966	chr19	5_8S_rRNA	rRNA	3UTR: 24184134-24184163
ENSG00000230803	chr2	AC097532.5	Processed_transcript	3UTR: 133038911-133039025
ENSG00000221288	chr2	hsa-mir-663b	miRNA	3UTR: 133012932-133013291
ENSG00000227984	chrY	AC069323.1	Protein_coding	3UTR: 13137708-13137809
ENSG00000229366	chrY	AC134882.3	Pseudogene	3UTR: 13488788-13488817

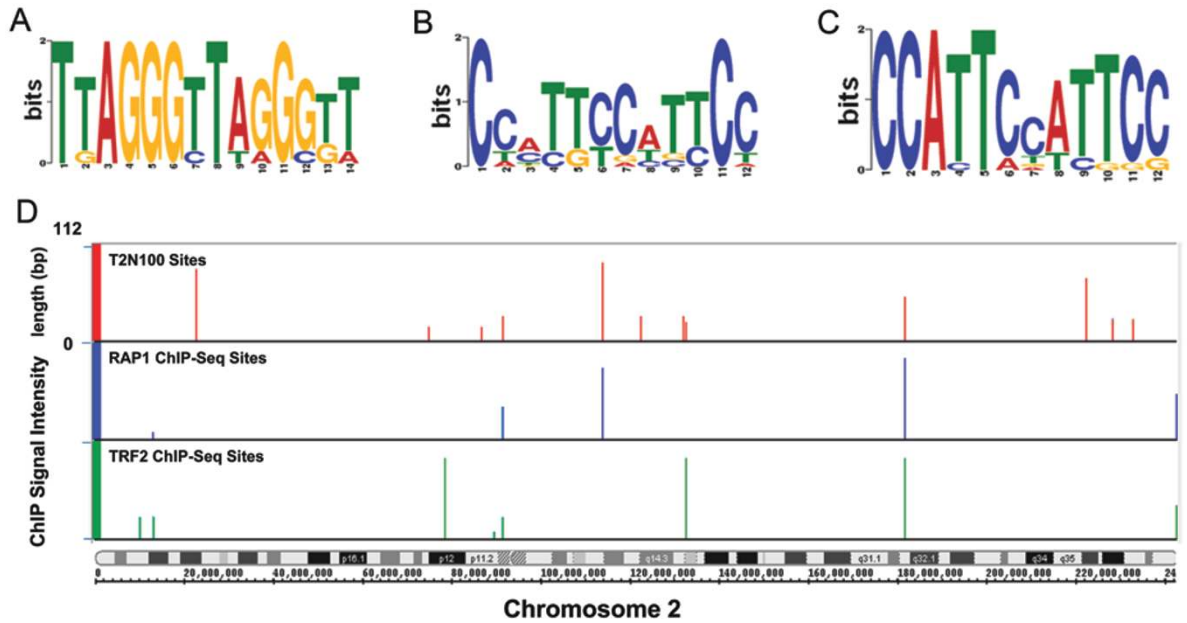


Figure 4 RAP1 and TRF2 occupy selective telomere-repeat-containing sites. **(A)** Motif search using the MEME and MAST software identified a highly conserved motif $(TTAGGG)_2$. **(B)** MEME and MAST were used to identify additional motifs for RAP1 binding sites after sequences containing the $(TTAGGG)_2$ motif were removed (E-value less than 0.1). **(C)** A similar analysis was done to obtain additional motifs for TRF2 binding sites (E-value less than 0.1). **(D)** RAP1 and TRF2 occupy a subset of the predicted binding sites on chromosome 2. Potential interstitial binding sites containing telomeric repeats were identified by scanning the human genome using T2N100 (TTAGGGTTAG{0, 100}TTAGGGTTAG). The predicted sites (red) were compared to RAP1 (blue) and TRF2 (green) sites found in ChIP-seq analysis.

were predicted to contain the T2N100 motif (Figure 5). When TRF2 was overexpressed (Figure 5C), association of TRF2 could be seen in 22 out of 26 predicted T2N100 sites (Figure 5A), but the majority of these sites were not occupied by endogenous TRF2 in control cells (Figure 5B). In comparison, overexpression of FLAG-RAP1 (Figure 5C) did not increase binding of RAP1 to the T2N100 sites tested (Figure 5D and 5E). Interestingly, overexpression of TRF2 also failed to bring RAP1 to the T2N100 sites examined (Supplementary information, Figure S2), suggesting that TRF2 may occupy T2N100 sites independently of RAP1. Our results point to the importance of cellular TRF2 concentration in determining TRF2 occupancy of interstitial sites that contain telomere repeats.

Regulation of gene transcription by telomeric proteins

Our findings of preferential binding of RAP1 and TRF2 to interstitial sites proximal to gene loci suggest that these telomeric proteins may regulate gene transcription. To understand whether RAP1 and TRF2 can affect the expression of their target genes, we generated HTC75 cells whose endogenous RAP1 or TRF2 was knocked down by RNA interference (RNAi). As shown in Figure 6A, RAP1 mRNA levels were reduced by ~80% using two different siRNA sequences. This reduction in RAP1 message level was accompanied by significant decreases in RAP1 protein levels as well (Figure 6B). We then compared the expression of RAP1 target genes in these cells by qRT-PCR. While the expression of a subset of the genes examined exhibited no change upon RAP1 knockdown, altered expression was observed in a number of RAP1 target genes (Figure 6D). For example, while *CLIC6* expression increased in RAP1 knockdown cells, the level of *RPH3AL* expression decreased. For TRF2, knocking down TRF2 by two different shRNAs in HTC75 cells (Figure 6C) resulted in decreased binding of TRF2 to its target genes (Supplementary information, Figure S3). Consequently, TRF2 knockdown reduced TRF2-target gene *PDE3A* expression while increasing the expression of *RPA2* (Figure 6E). These findings support the hypothesis that telomeric proteins can regulate the transcription of their target genes located in the interstitial regions.

Discussion

In this study, we analyzed genome-wide chromatin-

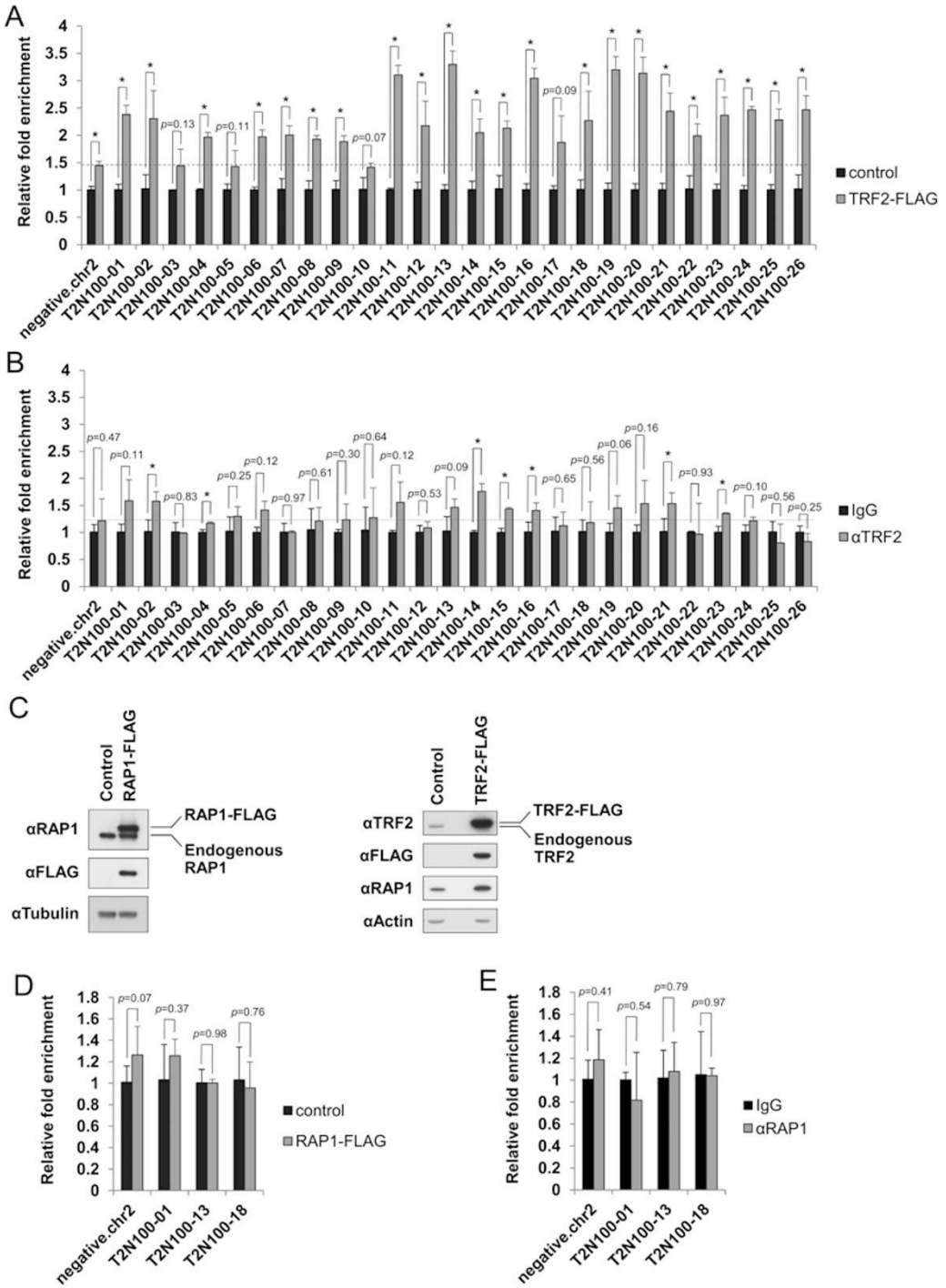


Figure 5 Selective occupancy of interstitial sites. **(A)** ChIP-qPCR analysis using anti-FLAG antibodies using HTC75 cells expressing TRF2-FLAG or vector alone (control). **(B)** ChIP-qPCR analysis of HTC75 cells using anti-TRF2 antibodies or IgG. Primer pairs were derived from the predicted T2N100 sites and are listed in Supplementary information, Table S1. Sequences from chromosome 2 were used as negative control for qPCR. Error bars indicate standard error (n = 3). P values were calculated by Student's t test and * indicates $P < 0.05$. **(C)** HTC75 cells expressing FLAG-tagged RAP1 (RAP1-FLAG) or TRF2 (TRF2-FLAG) were analyzed by western blotting using the indicated antibodies. Anti-tubulin or anti-actin antibodies were used as loading control. **(D)** Vector control or RAP1-FLAG-expressing HTC75 cells were analyzed by ChIP-qPCR using anti-FLAG antibodies. **(E)** HTC75 cells were analyzed by ChIP-qPCR using IgG or anti-RAP1 antibodies. Primer pairs were derived from the predicted T2N100 sites and are listed in Supplementary information, Table S1. Sequences from chromosome 2 were used as negative control for qPCR. Error bars indicate standard error (n = 3). P values were calculated by Student's t test.

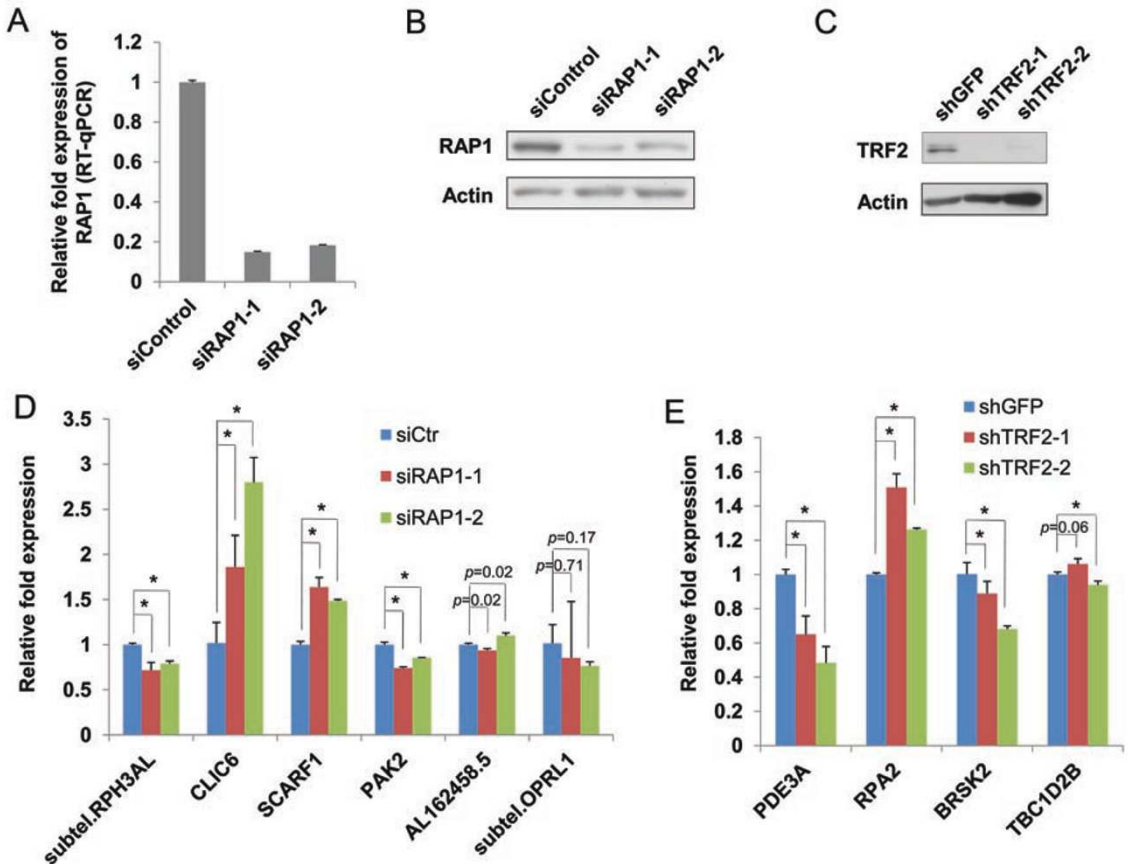


Figure 6 Telomeric proteins regulate gene transcription. HTC75 cells transfected with two siRNA oligos were harvested for **(A)** qRT-PCR analysis for mRNA levels and **(B)** western blotting for protein levels. siControl, control siRNA oligos. Anti-actin antibodies were used as loading control. **(C)** HTC75 cells stably expressing shRNA sequences against TRF2 were analyzed by western blotting. shRNA sequences against GFP were used as negative controls (shGFP). Anti-actin antibodies were used as loading control. **(D)** The expression of RAP1 target genes was examined in RAP1 knockdown cells by qRT-PCR. Error bars indicate standard error (n = 3). P values were calculated by Student's t test and * indicates $P < 0.05$. **(E)** The expression of TRF2 target genes was examined in TRF2 knockdown cells by qRT-PCR. Error bars indicate standard error (n = 3). P values were calculated by Student's t test and * indicates $P < 0.05$.

binding patterns of two telomeric proteins RAP1 and TRF2, and found these proteins to associate with interstitial sites. In budding yeast, the scRAP1 protein has long been implicated in global gene regulation [39, 41-43, 49, 50]. However, mammalian RAP1 proteins exhibit many differences from yeast RAP1, including the inability to bind directly to telomere DNA. Our findings provide evidence for an extra-telomeric function of human telomeric proteins, and suggest that such function is conserved through evolution. Recent findings of mouse RAP1 also support this notion [26]. Given the association of multiple core telomeric proteins that can form high-molecular-weight complexes on the telomeres [10], other telomeric proteins such as TIN2 and POT1 may also co-occupy RAP1 and TRF2 target sites outside of telomeres. Such possibilities warrant further investigation.

Work on RAP1 from mice and yeast found numerous extra-telomeric sites for RAP1 [26, 41-43], in direct contrast to our findings. In this study, we identified a limited number of interstitial binding sites for RAP1 and TRF2 in human cells. Our results appear to be consistent with another study, where ChIP-seq using anti-TRF1 and anti-TRF2 antibodies revealed restricted abilities of TRF1 and TRF2 in binding extra-telomeric sites in the human genome [51]. It is possible that the number of RAP1 and TRF2 binding sites may have been underestimated, due to potential antibody-epitope access problems in our ChIP experiments. Or the binding of interstitial sites in human cells may be more tightly controlled than in other species. Here, we provide evidence that the cellular concentration of TRF2 may play a role in selective binding of TRF2 to its target sites. Using stringent criteria,

we could predict ~ 300 potential TRF2 binding sites that contain telomeric repeats. TRF2 overexpression was sufficient to target TRF2 to these predicted interstitial sites, indicating that these sites can indeed be targeted by TRF2. Telomeres in *Mus musculus* are at least 10 times longer than in humans, and the concentration of mouse TRF2 and RAP1 proteins is likely higher as well. This may explain why a large number of interstitial sites were bound by RAP1 in mouse cells. We also observed that overexpression of TRF2 did not lead to an increase in the binding of endogenous RAP1 at T2N100 sites, possibly due to marginal enhancement of RAP1 level in TRF2-overexpression cells (Figure 5C). It is equally possible that TRF2 may not associate with RAP1 at the T2N100 sites tested. We speculate that the difference in telomeric protein concentration may at least in part account for the difference in binding site numbers observed between human and mouse. Taken together, our observations raise the intriguing possibility that the concentration of telomeric proteins such as TRF2 may vary in different cell types, resulting in distinct chromatin-binding profiles and biological outcomes.

TRF2 and RAP1 have overlapping but clearly distinct interstitial binding sites. The overlap is expected given that these two proteins form heterodimers on telomere repeats. Indeed, the major DNA motif common to both RAP1 and TRF2 binding sites is the TTAGGG repeat. Surprisingly, we found additional non-telomere-repeat motifs that are also shared by RAP1 and TRF2. Whether these sequences can be recognized by RAP1 and TRF2 *in vitro* and *in vivo* would be interesting to pursue. In addition, we also identified unique bindings for RAP1 and TRF2, respectively, indicating TRF2-independent function of RAP1. While the binding of RAP1 to TTAGGG repeat most likely occurs through its interaction with TRF2, the mechanism by which RAP1 associates with non-telomere-repeat sequences remains unclear and warrants further investigation. One possibility is that RAP1 may interact with other chromatin-associated factors (Songyang, unpublished data).

Finally, the enrichment of TRF2 and RAP1 binding in gene loci suggests a role of these telomeric proteins in gene regulation. Our data indicate that the transcription of a subset of the identified target genes was modulated by RAP1 and TRF2. Reducing RAP1 or TRF2 level led to altered target gene expression. This finding is in line with the repressor/activator role of scRAP1 [41-43]. Recently, RAP1 and TRF2 have been reported to associate with RNA splicing factors as well as proteins that regulate DNA recombination and replication [52, 53]. It will be important to sort out the various regulatory pathways impacted by telomeric proteins at the extra-telomeric sites, and to uncover the differences in the underlying mechanisms utilized by telomeric proteins in different species. Our work underscores the importance of further investigation into the non-canonical activities of telomeric proteins.

Materials and Methods

Vectors, cell line, and antibodies

cDNAs encoding human TRF2 and RAP1 were cloned into a pCL-based retroviral vector that allows for FLAG epitope tagging and stable expression in HTC75 cells. For western blotting and ChIP experiments, the following antibodies were used: mouse monoclonal anti-TRF2 antibody for western blotting (Calbiochem), rabbit polyclonal anti-TRF2 antibody for ChIP analysis (Bethyl Laboratories), rabbit polyclonal anti-RAP1 antibody (Bethyl Laboratories), rabbit polyclonal anti-FLAG antibody (Sigma), goat polyclonal anti-actin antibody (Santa Cruz), HRP-conjugated goat anti-mouse and goat anti-rabbit antibodies (Bio-Rad), and HRP-conjugated anti-FLAG M2 antibody (Sigma).

RNAi knockdown

Control Stealth RNAi siRNA duplexes (siControl, cat# 12935-300) and those against RAP1 were obtained from Invitrogen. siRAP1-1: AUUAACUGCCGAAUGAUCUUAAUGG, anti-sense; and siRAP1-2: ACGAACAGAGUCGAGGAAUGGGUGG, anti-sense. HTC75 cells were transfected with the siRNA duplexes (60 pmol each) using RNAiMAX (5 μl) (Invitrogen) in six-well plates, and analyzed 48 h later. For knockdown of TRF2, two shRNA sequences against TRF2, shTRF2-1 (5′-AGGAAATGGTGAAGTCTAT-3′) and shTRF2-2 (5′-GAGCATGGTTCCTAATAAT-3′), were cloned into a retroviral vector as previously described [54]. A shRNA sequence against GFP (shGFP) (5′-CACAAGCTGGAGTACAACT-3′) was the negative control. HTC75 cells stably expressing these sequences were selected in puromycin for 3 days before further analysis.

ChIP assay

ChIP assays were performed essentially as described previously [55], with slight modifications. Sonicated lysates (from ~5 × 10^6 cells) were pre-cleared with 50 μl protein A beads, 2 μg rabbit IgG (Sigma), 10 μl of 5% BSA, and 5 μg of sheared *E. coli* DNA at 4 °C for 2 h. For immunoprecipitation, pre-cleared lysates were incubated with 3 μg of antibodies (rabbit IgG, polyclonal anti-TRF2, anti-RAP1, or anti-FLAG), 1 μl of 5% BSA, 25 μg of sheared *E. coli* DNA, and 50 μl of Protein A agarose beads. The precipitated DNA was then eluted and analyzed by sequencing, dot-blotting, or qPCR. A radiolabeled oligonucleotide $(TTAGGG)_3$ was used to detect telomeric DNA in dot-blotting.

Real-time qPCR

Isolated total RNA (RNeasy Mini Kit, Qiagen) was reverse transcribed using the iScript cDNA Synthesis Kit (Bio-Rad). Real-time qPCR was carried out using an ABI StepOnePlus real-time PCR system and the SYBR green master mix (Applied Biosystems). qPCR was also done to validate candidates and T2N100 sites found in ChIP and ChIP-seq experiments. For specific prim-

ers used in this study, please see Supplementary information, Table S1.

ChIP sequencing

The ChIP DNA libraries were prepared using DNA sample kit (Illumina) following the manufacturer's protocols. Briefly, 10 ng of ChIP DNA was end-repaired and ligated to Illumina adaptors. DNA samples were then amplified using adaptor-specific primers for 21 cycles and fragments of ~ 150 bp were isolated from the agarose gel. The quantity and size distribution of sequencing libraries were determined using the PicoGreen fluorescence assay and the Agilent 2100 Bioanalyzer, respectively. The quality of the library was assessed by checking the enrichment for known targeted regions with real-time PCR. Sequencing of the library was carried out on the Illumina Genome Analyzer system according to the manufacturer's specifications.

Data analysis

Alignment of short reads to the human genome (hg19 release, USCS) was carried out using SOAP [56], allowing for a maximum of two mismatches. Mitochondria sequences are included. For reads that were longer than 35 bp and could not be mapped to the genome, only the first 35 bp were used for alignment. All short reads that were uniquely aligned were used for further analysis. The programs SISSRs v.1.4 [57] and CisGenome v. 1.2 [58] were used for peak detection under default settings (0.001 FDR for SISSRs and 0.1 FDR for CisGenome). RAP1 or TRF2 ChIP libraries were individually pooled as the sample set and the IgG library was pooled as the negative control set for the two-sample ChIP-seq analysis. All of the peaks detected by the two algorithms were combined and merged (overlap or within 400 bp).

Annotation of the identified peaks was carried out using human genomic sequences (hg19) and gene loci information based on UCSC exon locus annotation (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/database/) and BioMart (www.biomart.org/ version 57). For exon annotation, the longest transcript was considered for each gene. Regions within 5 kb up- or downstream of annotated transcripts were referenced as 5'UTR or 3'UTR, respectively. Subtelomeric regions are defined as 500-kb regions adjacent to the terminal fragment of each chromosome [59].

For comparative analysis of RAP1-binding sites between human and mouse, the 30 398 mouse RAP1 binding sites identified by Cisgenome (10% FDR) were obtained [26]. Pair-wise alignment files of human (hg19) and mouse (mm9) genome sequences were obtained from UCSC, and human and mouse orthologue information was retrieved from Ensembl v.59 (www.ensembl.org). Genomic sequences at the corresponding peaks from ChIP-seq experiments were used for motif search with the Multiple Expectation Maximization for Motif Elicitation (MEME) and Motif Alignment and Search Tool (MAST) software [60].

Acknowledgment

We would like to thank Dr Maria Blasco (Spanish National Cancer Center) for sharing with us the mouse RAP1 ChIP-seq data. We thank Yi Zhang (Baylor College of Medicine) for technical help with TRF2 knockdown and Dr Dan Liu (Baylor College of Medicine) for critical reading of this manuscript. This work is supported by NCI CA133249, NIGMS GM081627, the Welch Foundation Q-1673, and 2010CB945400. This work is also supported by 1S10RR026550-01 and 5U54HG003273 (RC). ZS is a Leukemia and Lymphoma Society Scholar.

References

1 Blackburn EH, Collins K. Telomerase: An RNP Enzyme Synthesizes DNA. *Cold Spring Harb Perspect Biol* 2010 Jul 21. doi:10.1101/cshperspect.a003558
2 Palm W, de Lange T. How Shelterin Protects Mammalian Telomeres. *Annu Rev Genet* 2008; **42**:301-334.
3 Djojosubroto MW, Choi YS, Lee HW, Rudolph KL. Telomeres and telomerase in aging, regeneration and cancer. *Mol Cells* 2003; **15**:164-175.
4 Wright WE, Shay JW. Telomere biology in aging and cancer. *J Am Geriatr Soc* 2005; **53**:S292-294.
5 Venteicher AS, Abreu EB, Meng Z, *et al.* A human telomerase holoenzyme protein required for Cajal body localization and telomere synthesis. *Science* 2009; **323**:644-648.
6 Teixeira MT, Arneric M, Sperisen P, Lingner J. Telomere length homeostasis is achieved via a switch between telomerase- extendible and -nonextendible states. *Cell* 2004; **117**:323-335.
7 Seeler JS, Bischof O, Nacerddine K, Dejean A. SUMO, the three Rs and cancer. *Curr Top Microbiol Immunol* 2007; **313**:49-71.
8 de Lange T. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev* 2005; **19**:2100-2110.
9 Xin H, Liu D, Songyang Z. The telosome/shelterin complex and its functions. *Genome Biol* 2008; **9**:232.
10 Liu D, O'Connor MS, Qin J, Songyang Z. Telosome, a mammalian telomere-associated complex formed by multiple telomeric proteins. *J Biol Chem* 2004; **279**:51338-51342.
11 Broccoli D, Smogorzewska A, Chong L, de Lange T. Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2. *Nat Genet* 1997; **17**:231-235.
12 Zhong Z, Shiue L, Kaplan S, de Lange T. A mammalian factor that binds telomeric TTAGGG repeats *in vitro*. *Mol Cell Biol* 1992; **12**:4834-4843.
13 Bilaud T, Brun C, Ancelin K, Koering CE, Laroche T, Gilson E. Telomeric localization of TRF2, a novel human telobox protein. *Nat Genet* 1997; **17**:236-239.
14 Bilaud T, Koering CE, Binet-Brasselet E, *et al.* The telobox, a Myb-related telomeric DNA binding motif found in proteins from yeast, plants and human. *Nucleic Acids Res* 1996; **24**:1294-1303.
15 van Steensel B, Smogorzewska A, de Lange T. TRF2 protects human telomeres from end-to-end fusions. *Cell* 1998; **92**:401-413.
16 van Steensel B, de Lange T. Control of telomere length by the human telomeric protein TRF1. *Nature* 1997; **385**:740-743.
17 Zhou XZ, Perrem K, Lu KP. Role of Pin2/TRF1 in telomere maintenance and cell cycle control. *J Cell Biochem* 2003; **89**:19-37.
18 Baumann P, Cech TR. Pot1, the putative telomere end-binding protein in fission yeast and humans. *Science* 2001; **292**:1171-1175.
19 Lei M, Podell ER, Baumann P, Cech TR. DNA self-recognition in the structure of Pot1 bound to telomeric single-strand-

ed DNA. *Nature* 2003; **426**:198-203.
20 Li B, Oestreich S, de Lange T. Identification of human Rap1: implications for telomere evolution. *Cell* 2000; **101**:471-483.
21 Kim SH, Kaminker P, Campisi J. TIN2, a new regulator of telomere length in human cells. *Nat Genet* 1999; **23**:405-412.
22 Liu D, Safari A, O'Connor MS, *et al.* PTOP interacts with POT1 and regulates its localization to telomeres. *Nat Cell Biol* 2004; **6**:673-680.
23 Ye JZ, Hockemeyer D, Krutchinsky AN, *et al.* POT1-interacting protein PIP1: a telomere length regulator that recruits POT1 to the TIN2/TRF1 complex. *Genes Dev* 2004; **18**:1649-1654.
24 Houghtaling BR, Cuttonaro L, Chang W, Smith S. A dynamic molecular link between the telomere length regulator TRF1 and the chromosome end protector TRF2. *Curr Biol* 2004; **14**:1621-1631.
25 Sfeir A, Kabir S, van Overbeek M, Celli GB, de Lange T. Loss of Rap1 induces telomere recombination in the absence of NHEJ or a DNA damage signal. *Science* 2010; **327**:1657-1661.
26 Martinez P, Thanasoula M, Carlos AR, *et al.* Mammalian Rap1 controls telomere function and gene expression through binding to telomeric and extratelomeric sites. *Nat Cell Biol* 2010; **12**:768-780.
27 O'Connor MS, Safari A, Liu D, Qin J, Songyang Z. The human Rap1 protein complex and modulation of telomere length. *J Biol Chem* 2004; **279**:28585-28591.
28 Li B, de Lange T. Rap1 affects the length and heterogeneity of human telomeres. *Mol Biol Cell* 2003; **14**:5060-5068.
29 Ye JZ, Donigian JR, van Overbeek M, *et al.* TIN2 binds TRF1 and TRF2 simultaneously and stabilizes the TRF2 complex on telomeres. *J Biol Chem* 2004; **279**:47264-47271.
30 Kim SH, Beausejour C, Davalos AR, Kaminker P, Heo SJ, Campisi J. TIN2 mediates functions of TRF2 at human telomeres. *J Biol Chem* 2004; **279**:43799-43804.
31 Xin H, Liu D, Wan M, *et al.* TPP1 is a homologue of ciliate TEBP-beta and interacts with POT1 to recruit telomerase. *Nature* 2007; **445**:559-562.
32 Wang F, Podell ER, Zaug AJ, *et al.* The POT1-TPP1 telomere complex is a telomerase processivity factor. *Nature* 2007; **445**:506-510.
33 Abreu E, Aritonovska E, Reichenbach P, *et al.* TIN2-tethered TPP1 recruits human telomerase to telomeres *in vivo*. *Mol Cell Biol* 2010; **30**:2971-2982.
34 Bae NS, Baumann P. A RAP1/TRF2 complex inhibits nonhomologous end-joining at human telomeric DNA ends. *Mol Cell* 2007; **26**:323-334.
35 Chen LY, Liu D, Songyang Z. Telomere maintenance through spatial control of telomeric proteins. *Mol Cell Biol* 2007; **27**:5898-5909.
36 Zhang P, Pazin MJ, Schwartz CM, *et al.* Nontelomeric TRF2-REST interaction modulates neuronal gene silencing and fate of tumor and stem cells. *Curr Biol* 2008; **18**:1489-1494.
37 Fouche N, Cesare AJ, Willcox S, Ozgur S, Compton SA, Griffith JD. The basic domain of TRF2 directs binding to DNA junctions irrespective of the presence of TTAGGG repeats. *J Biol Chem* 2006; **281**:37486-37495.
38 Mao Z, Seluanov A, Jiang Y, Gorbunova V. TRF2 is required for repair of nontelomeric DNA double-strand breaks by homologous recombination. *Proc Natl Acad Sci USA* 2007; **104**:13068-13073.
39 Kyrion G, Liu K, Liu C, Lustig AJ. RAP1 and telomere structure regulate telomere position effects in *Saccharomyces cerevisiae*. *Genes Dev* 1993; **7**:1146-1159.
40 Marcand S, Gilson E, Shore D. A protein-counting mechanism for telomere length regulation in yeast. *Science* 1997; **275**:986-990.
41 Buchman AR, Lue NF, Kornberg RD. Connections between transcriptional activators, silencers, and telomeres as revealed by functional analysis of a yeast DNA-binding protein. *Mol Cell Biol* 1988; **8**:5086-5099.
42 Shore D. RAP1: a protean regulator in yeast. *Trends Genet* 1994; **10**:408-412.
43 Kurtz S, Shore D. RAP1 protein activates and silences transcription of mating-type genes in yeast. *Genes Dev* 1991; **5**:616-628.
44 Yang X, Figueiredo LM, Espinal A, Okubo E, Li B. RAP1 is essential for silencing telomeric variant surface glycoprotein genes in Trypanosoma brucei. *Cell* 2009; **137**:99-109.
45 Teo H, Ghosh S, Luesch H, *et al.* Telomere-independent Rap1 is an IKK adaptor and regulates NF-kappaB-dependent gene expression. *Nat Cell Biol* 2010; **12**:758-767.
46 Hanaoka S, Nagadoi A, Nishimura Y. Comparison between TRF2 and TRF1 of their telomeric DNA-bound structures and DNA-binding activities. *Protein Sci* 2005; **14**:119-130.
47 Bianchi A, Stansel RM, Fairall L, Griffith JD, Rhodes D, de Lange T. TRF1 binds a bipartite telomeric site with extreme spatial flexibility. *EMBO J* 1999; **18**:5735-5744.
48 Bianchi A, Smith S, Chong L, Elias P, de Lange T. TRF1 is a dimer and bends telomeric DNA. *EMBO J* 1997; **16**:1785-1794.
49 Lieb JD, Liu X, Botstein D, Brown PO. Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. *Nat Genet* 2001; **28**:327-334.
50 Morse RH. RAP, RAP, open up! New wrinkles for RAP1 in yeast. *Trends Genet* 2000; **16**:51-53.
51 Simonet T, Zaragosi LE, Philippe C, *et al*. The human TTAGGG Repeat Factors 1 and 2 bind to a subset of interstitial telomeric sequences and satellite repeats. *Cell Res* 2011 Mar 22. doi:10.1038/cr.2011.40
52 Giannone RJ, McDonald HW, Hurst GB, Shen RF, Wang Y, Liu Y. The protein network surrounding the human telomere repeat binding factors TRF1, TRF2, and POT1. *PLoS One* 2010; **5**.
53 Svendsen JM, Smogorzewska A, Sowa ME, *et al.* Mammalian BTBD12/SLX4 assembles a Holliday junction resolvase and is required for DNA repair. *Cell* 2009; **138**:63-77.
54 O'Connor MS, Safari A, Xin H, Liu D, Songyang Z. A critical role for TPP1 and TIN2 interaction in high-order telomeric complex assembly. *Proc Natl Acad Sci USA* 2006; **103**:11874-11879.
55 Liang J, Wan M, Zhang Y, *et al.* Nanog and Oct4 associate with unique transcriptional repression complexes in embryonic stem cells. *Nat Cell Biol* 2008; **10**:731-739.
56 Li R, Yu C, Li Y, *et al.* SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* 2009; **25**:1966-1967.
57 Jothi R, Cuddapah S, Barski A, Cui K, Zhao K. Genome-wide identification of *in vivo* protein-DNA binding sites from

ChIP-Seq data. *Nucleic Acids Res* 2008; **36**:5221-5231.
58 Ji H, Jiang H, Ma W, Johnson DS, Myers RM, Wong WH. An integrated software system for analyzing ChIP-chip and ChIP-seq data. *Nat Biotechnol* 2008; **26**:1293-1300.
59 Riethman H, Ambrosini A, Castaneda C, *et al.* Mapping and initial analysis of human subtelomeric sequence assemblies. *Genome Res* 2004; **14**:18-28.
60 Bailey TL, Williams N, Misleh C, Li WW. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res* 2006; **34**:W369-373.

(**Supplementary information** is linked to the online version of the paper on the *Cell Research* website.)