

# The human TTAGGG repeat factors 1 and 2 bind to a subset of interstitial telomeric sequences and satellite repeats

Thomas Simonet<sup>1</sup>, Laure-Emmanuelle Zaragosi<sup>2</sup>, Claude Philippe<sup>3</sup>, Kevin Lebrigand<sup>2</sup>, Clémentine Schouteden<sup>1</sup>, Adeline Augereau<sup>1,3</sup>, Serge Bauwens<sup>1</sup>, Jing Ye<sup>1,3</sup>, Marco Santagostino<sup>4</sup>, Elena Giulotto<sup>4</sup>, Frederique Magdinier<sup>1</sup>, Béatrice Horard<sup>1</sup>, Pascal Barbry<sup>2</sup>, Rainer Waldmann<sup>2</sup>, Eric Gilson<sup>1,3,5</sup>

<sup>1</sup>Laboratoire de Biologie Moléculaire de la Cellule-UMR 5239 CNRS/ENS Lyon/ Université Lyon, Ecole Normale Supérieure de Lyon, 46 allée d'Italie, Lyon 69364, France; <sup>2</sup>CNRS and University of Nice Sophia Antipolis, Institut de Pharmacologie Moléculaire et Cellulaire, 06560 Sophia Antipolis, France; <sup>3</sup>Laboratory of Biology and Pathology of Genomes of University of Nice Sophia-Antipolis, CNRS UMR6267/INSERM U998, Faculty of Medicine, Nice, France; <sup>4</sup>Dipartimento di Genetica e Microbiologia Adriano Buzzati-Traverso, Università di Pavia, Pavia, Italy; <sup>5</sup>Department of Medical Genetics, CHU of Nice, Nice, France

**The study of the proteins that bind to telomeric DNA in mammals has provided a deep understanding of the mechanisms involved in chromosome-end protection. However, very little is known on the binding of these proteins to nontelomeric DNA sequences. The TTAGGG DNA repeat proteins 1 and 2 (TRF1 and TRF2) bind to mammalian telomeres as part of the shelterin complex and are essential for maintaining chromosome end stability. In this study, we combined chromatin immunoprecipitation with high-throughput sequencing to map at high sensitivity and resolution the human chromosomal sites to which TRF1 and TRF2 bind. While most of the identified sequences correspond to telomeric regions, we showed that these two proteins also bind to extratelomeric sites. The vast majority of these extratelomeric sites contains interstitial telomeric sequences (or ITSs). However, we also identified non-ITS sites, which correspond to centromeric and pericentromeric satellite DNA. Interestingly, the TRF-binding sites are often located in the proximity of genes or within introns. We propose that TRF1 and TRF2 couple the functional state of telomeres to the long-range organization of chromosomes and gene regulation networks by binding to extratelomeric sequences.**

**Keywords:** telomere; TRF1; TRF2; interstitial telomeric sequence; satellite DNA

*Cell Research* (2011) 21:1028–1038. doi:10.1038/cr.2011.40; published online 22 March 2011

## Introduction

The paramount importance of telomeres to cell fate likely stems from the great diversity in the functions they perform [1, 2]. They control the replication of chromosomal DNA termini, protect chromosome ends from DNA repair and checkpoint activation, control the meiotic spindle, localize the chromosome ends within the nuclear space and regulate long-range chromatin changes as well as gene expression. Telomeres consist of specific nucleoprotein complexes [3]. Telomeric DNA has several distinctive features, including a sequence formed by

repetitions of a small G-rich motif (TTAGGG in mammals) and the presence of a single-stranded tail on the 3'-oriented strand (G tail). Telomeric DNA is transcribed into a UUAGGG repeat-containing RNA called TERRA, which is believed to play fundamental roles in telomere biology [4, 5].

A key component of the mammalian telomere is the shelterin complex, which is composed of six polypeptides: TRF1, TRF2, Rap1, Tin2, TPP1 and Pot1 [5]. Of these, three bind specifically to TTAGGG repeats: TRF1 and TRF2, which recognize the duplex DNA, and Pot1, which binds to the single-stranded 3' overhangs [3, 6]. TRF1 and TRF2 do not exist in budding yeast. Instead, yeast Rap1 acts as an essential capping factor that binds to telomeric DNA, while yeast Cdc13 binds to the 3' overhang and seems to perform functions that are similar to those of Pot1 and TPP1 [7].

Telomeres in yeast and mammals can silence neigh-

Correspondence: Eric Gilson

Tel: +33-472728453; Fax: +33-472728080

E-mail: Eric.Gilson@unice.fr

Received 3 November 2010; revised 9 January 2011; accepted 11 January 2011; published online 22 March 2011

boring genes by exerting telomeric position effect (or TPE) [8-10]. TPE is influenced by telomere length and structure as well as by chromatin-remodeling machineries [11]. Telomeric and subtelomeric chromatin differ from constitutive heterochromatin in terms of structure and dynamics, specificity of DNA sequences, and binding of specific factors [12]. The mechanisms that initiate the formation of heterochromatin at telomeres are unknown but likely involve the binding of specific factors to telomeric DNA. For instance, the N-terminal part of TRF2 may facilitate heterochromatin formation by binding to ORC1 and TERRA [13].

Repetitions of the TTAGGG telomeric unit, called interstitial telomeric sequences, or ITSs, are also present within chromosomes [14]. In humans, three classes of ITSs were identified [15]: (i) subtelomeric ITSs, located within subtelomeric domains and composed of extended arrays (usually several hundreds of base pairs), including many degenerate units; they probably arose from recombination events involving chromosome termini [16]; (ii) short internal ITSs, located away from telomeres and composed of relatively few TTAGGG units; these ITSs are likely to have been generated during the repair of DNA double-strand breaks that occurred during evolution [17]; (iii) one fusion ITS, located in 2q14, derived from the end fusion between the two ancestral chromosomes that gave rise to human chromosome 2 [18]. No clear indication of any particular function of ITSs has been provided so far.

Emerging evidence indicates that the shelterin components have non-telomeric functions in DNA repair [19], Epstein-Barr virus replication [20], transcriptional regulation [21] and NF- $\kappa$ B activation [22]. These non-telomeric functions might be, at least partially, explained by their binding to ITSs. Indeed, there is mounting indication that shelterin components can bind to interstitial DNA sequences: (i) TRF1 and TRF2 bind to the pericentromeric regions of hamster chromosomes containing large blocks of ITSs [23, 24]; (ii) TRF2 and TIN2 bind to an ITS formed by a rare human chromosome rearrangement [25]; (iii) TRF2 binds to a stretch of telomeric sequence that is artificially inserted in the middle of the long arm of chromosome 4 [26]; (iv) Rap1 binds to several ITSs of the mouse genome [27]. However, three naturally occurring ITSs of human chromosomes do not appear to be bound by TRF2 [26]. Therefore, it is still unclear whether TRF1 and TRF2 really bind to the ITSs normally found in human chromosomes or even to unrelated sequences. Moreover, there is evidence that TRF2 modulates gene expression outside from telomeres since it interacts with the repressor element 1-silencing transcription factor (REST), a repressor of genes devoted to

neuronal functions [21].

In this study, we mapped the human chromosomal sites to which TRF1 and TRF2 bind by combining chromatin immunoprecipitation with high-throughput DNA sequencing (ChIP-Seq).

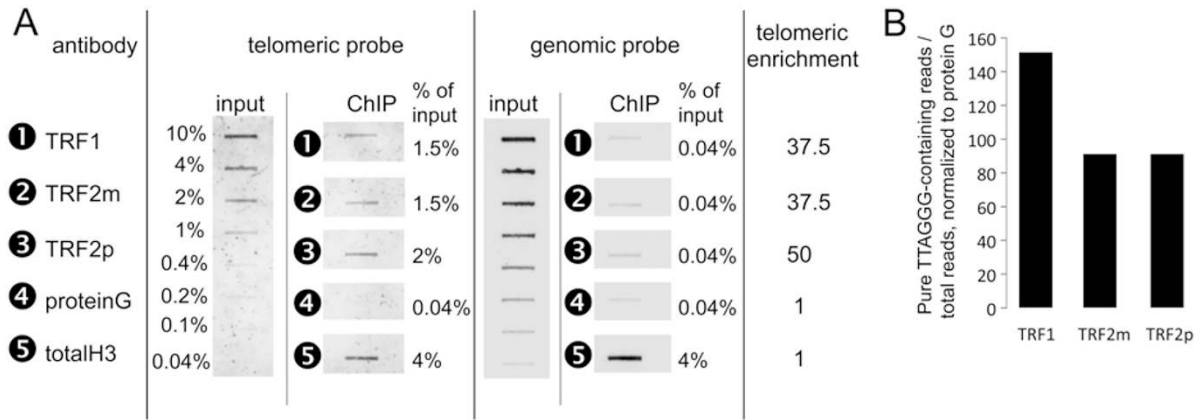
## Results

### *Identification of TRF binding sites by ChIP-Seq analysis*

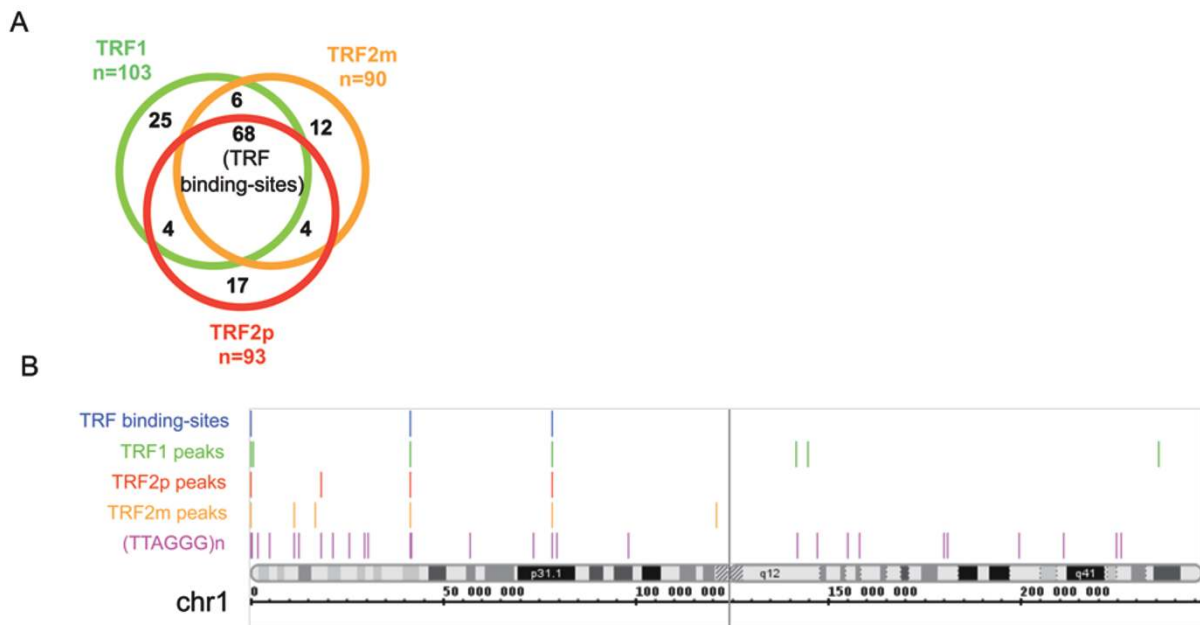
To establish global binding profiles of TRF1 and TRF2 (collectively named the TRF proteins), we performed a ChIP-Seq analysis with one antibody specific for TRF1 and two antibodies specific for TRF2 (one monoclonal or TRF2m, one polyclonal or TRF2p). We used the BJ-HELTRas<sup>mc</sup> tumor cell line because TRF2 is required for tumorigenicity through a pathway that involves uncoupling of telomere protection and the DNA damage response mechanism, suggesting a role for extratelomeric TRF2 binding sites in oncogenesis (Biroccio *et al.*, submitted). The specificity of the anti-TRF1 and anti-TRF2 antibodies was confirmed by slot blot analysis (Figure 1A). We found up to 50-fold enrichment of telomeric sequences in the TRF antibody-immunoprecipitated samples when compared to Protein G-Sepharose-precipitated control samples and total histone H3-immunoprecipitation. This result was confirmed by the analysis of the ChIP-Seq reads. In TRF antibody-immunoprecipitated samples we detected 90 to 150 times more sequences that contain solely the (TTAGGG)<sub>n</sub> motif than in the control samples (Figure 1B).

To identify extratelomeric binding sites for the TRF proteins, we retained only reads that were uniquely aligned on the 2006 Human genome assembly (NCBI36/hg18), and we checked that pure (TTAGGG)<sub>n</sub> reads, which likely originate from telomeric DNA, had been indeed completely discarded. Significantly read-enriched positions or peaks were identified using the SISSR software [28] with a *P* value threshold of 0.001 using protein G immunoprecipitation as background. We further removed the seemingly artifactual (non-specific) peaks through a visual inspection of a density profile of the matched reads (see the example for chromosome 1, shown in Supplementary information, Figure S1). Following this filtering, we identified 68 peaks present in all three TRF ChIP-Seq samples (TRF1, TRF2m and TRF2p) (Figure 2A). Results for chromosome 1 are shown in Figure 2B and those for other chromosomes are shown in Supplementary information, Figure S2.

Notably, 18 peaks from the TRF2m ChIP (among *n* = 90, 20%) were not found using TRF2p antibody, while 21 peaks identified by the TRF2p ChIP (*n* = 93, 22.5%) were not present in TRF2m ChIP. For most of



**Figure 1 (A)** Slot blot showing the telomeric enrichment of DNA immunoprecipitated with anti-TRF1 or TRF2 antibodies. DNA immunoprecipitated by a total H3 antibody and pulled down by protein G alone was used as a control. Half of the precipitated DNA was loaded, along with an input scale (2 500 ng to 10 ng, corresponding to 10% to 0.04% of the total input), and hybridized sequentially to a telomeric probe and a genomic probe. For each probe, we quantified the fraction of the immunoprecipitated DNA. The ratio of the value obtained for the telomeric probe to the genomic probe is the telomeric enrichment factor. **(B)** Fold enrichment of the fraction of raw reads containing only (TTAGGG)<sub>n</sub> sequences from TRF ChIP-Seq as normalized to the reads obtained through immunoprecipitation of protein G.



**Figure 2 (A)** TRF1, TRF2m, and TRF2p ChIP-Seq peaks. The peaks largely coincide, as shown on the Venn diagram. Assessment of overlaps was performed by visual inspection in the Integrated Genome Browser. **(B)** Visualization of TRF peaks and TRF binding sites. Regions of significant read enrichment ( $P < 0.001$ ) for each ChIP analysis (over the protein G background) are shown for human chromosome 1, along with the (TTAGGG)<sub>n</sub> repeats extracted from RepeatMasker UCSC files [40]. The upper line (TRF binding sites) displays the positions of the common peaks obtained with the three TRF antibodies. The criterion is one peak with a  $P$  value  $< 0.001$  and two peaks with  $P < 0.05$ . For the individual antibodies (TRF1, TRF2p, and TRF2m), only the peaks with a  $P$  value  $< 0.001$  are shown.

these non-overlapping sites, the visual inspection of the read profiles revealed a read enrichment with the other TRF2 antibody, as compared to protein-G ChIP, but not at a level allowing its identification by the statistical parameters used for pSISSRs peak-finder. This is the case for 13 (respectively 17) out of the 18 (respectively 21) TRF2m (respectively TRF2p) peaks not found with TRF2p (respectively TRF2m) (data not shown). Regarding the peaks with no obvious reads enrichment for the ChIP performed with the other antibody (5 out of 90 TRF2m peaks and 4 out of 93 TRF2p peaks), they could correspond to either false-positive peaks or TRF2-DNA complexes exhibiting a differential accessibility to the epitopes recognized by the two types of TRF2 antibodies. Thus, the non-overlapping TRF2p and TRF2m peaks are mainly not antibody-specific, most likely reflecting small variations between ChIP experiments for low-affinity binding sites. More rarely, they can be attributed to differences in epitope exposure and false positivity.

We conclude that the 68 overlapping peaks correspond to a set of *bona fide* TRF binding sites but do not constitute an exhaustive list of extratelomeric TRF binding regions. These 68 peaks will hereafter be referred to as TRF binding sites. The complete list is given in Supplementary information, Table S1.

These ChIP-Seq data have been deposited in NCBI's Gene Expression Omnibus [29] and are accessible through GEO Series accession number GSE26005 (<http://www.ncbi.nlm.nih.gov/gate1.inist.fr/geo/query/acc.cgi?acc=GSE26005>).

#### Validation by ChIP-qPCR

To validate the TRF binding sites identified by ChIP-Seq, we performed independent ChIP experiments with TRF1 and TRF2m antibodies followed by qPCR analysis (ChIP-qPCR) of extratelomeric TRF binding sites identified by ChIP-Seq (Figure 3 and Supplementary information, Table S1). In the TRF1 and TRF2m immunoprecipitates obtained from the same cells as those used for the ChIP-Seq analysis (BJ-HELTRAs<sup>mc</sup>), seven out of seven TRF binding-sites were more enriched than the unrelated *GAPDH* gene ( $P < 0.05$ , Figure 3). Importantly, for two TRF binding sites (Chr6-intron/DNAH8 and Chr10p15-gd), regions 1 000 bp downstream from the binding site were not enriched (Figure 3), indicating that the binding is limited to the peak region. We also tested two sites that were not included in the list of TRF binding sites, because they were identified with only one (Chr.1p36.13) or two (Chr.4p16) of the three anti-TRF antibodies used. They did not appear to be strongly bound by TRF1 and TRF2 in the ChIP-qPCR analyses (Figure 3 and data not shown). We concluded that our criteria for selecting

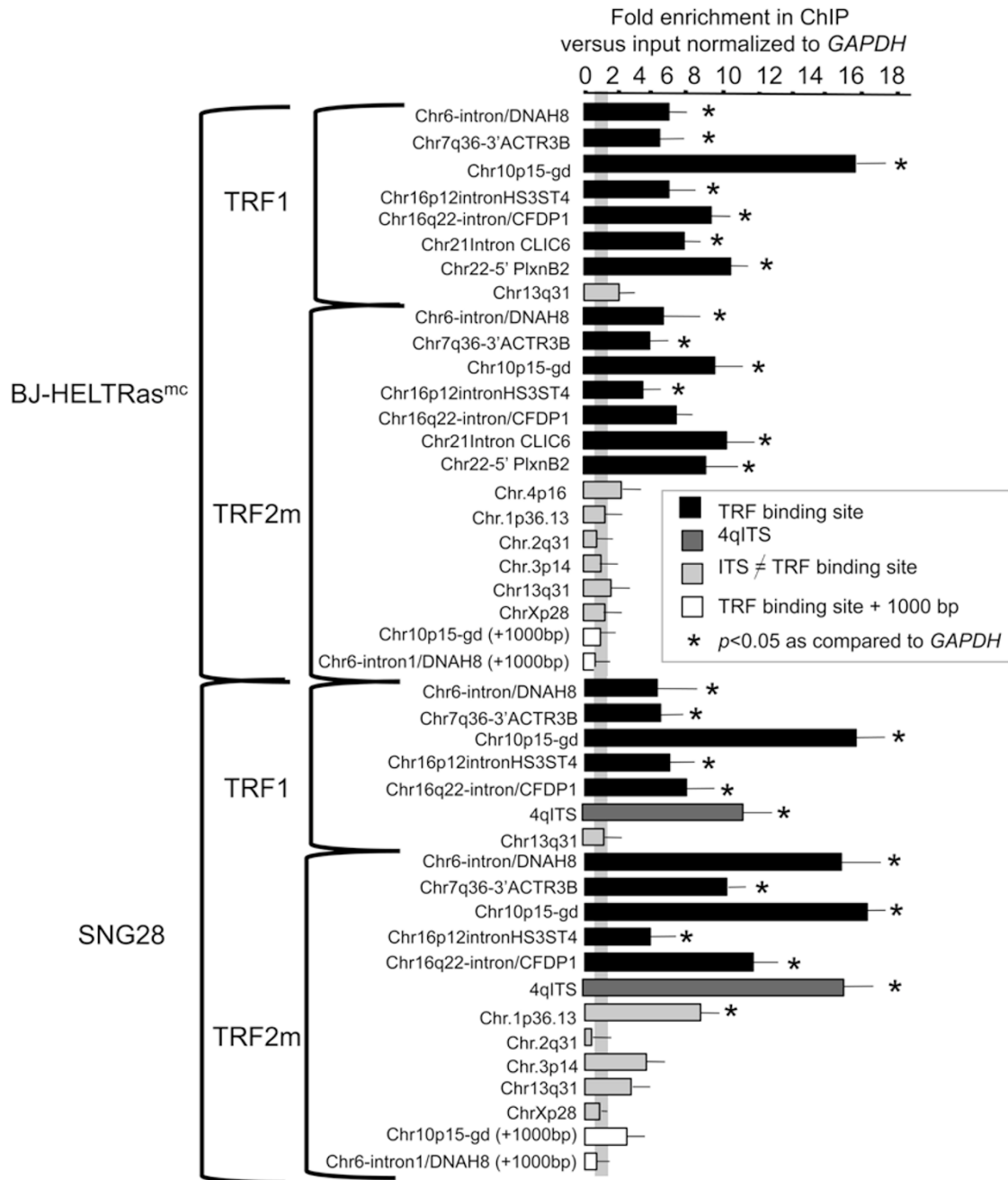
the 68 overlapping TRF binding sites reliably identified binding sites for TRF1 and TRF2, although we cannot exclude the existence of other sites, of lower affinity or less accessible to the TRF antibodies.

Next, we tested whether we could confirm this enrichment profile in a second cell line. We performed TRF1 and TRF2m ChIP-qPCR in the SNG28 human squamous carcinoma cell line. This cell line contains an artificially integrated 800-bp telomeric sequence in the middle of the long arm of chromosome 4 (named 4qITS) [26, 30], which serves as a positive control for the immunoprecipitations. We observed a clear TRF1 and TRF2 enrichment for five out of five identified TRF binding sites and also of the 4qITS sequence (Figure 3). Thus, the TRF binding-sites appear to be well conserved in different cell lines. Interestingly, the Chr.1p36.13 peak, which was detected only in the TRF2p ChIP-Seq data, and which is not bound by TRF2 in BJ-HELTRAs<sup>mc</sup> cells (based on ChIP-qPCR results obtained with TRF2m antibodies, Figure 3), is well enriched after TRF2 immunoprecipitation in SNG28 cells (Figure 3). Thus, although the TRF binding-sites profile defined in BJ-HELTRAs<sup>mc</sup> cells seems to be largely conserved in SNG28, some difference exists, suggesting that the specific cellular context may determine the ability of TRF1 and TRF2 to bind to certain regions of the genome. In addition, the length polymorphism that is known to characterize both intrachromosomal and subtelomeric loci [31] could influence TRF binding.

#### Most of the TRF binding sites correspond to ITSs

To determine the type of extratelomeric DNA bound by TRF1 and TRF2, the TRF binding sites were analyzed with a *de novo* consensus motif prediction software (MEME). The consensus sequence TTAGGGTTAGG was identified in 59 of the 68 TRF binding-sites (Figure 4A, Supplementary information, Table S1). This sequence is a nearly perfect concatenation of two telomeric TTAGGG motifs, and thus represents an ITS. As illustrated in Figure 4B, reads are equally distributed around the identified ITSs, indicating that TRF proteins directly bind to the ITSs identified in this study (Figure 4B). An example of reads around a TRF-unbound ITS is also shown (Figure 4B).

A detailed analysis of the TRF binding sites containing an ITS revealed that they were present in 48 different loci (boxed in Supplementary information, Table S1): 17 were subtelomeric regions (sequences less than 100 kb from a chromosome end), 30 were short internal ITSs and one corresponded to the 2q14 fusion ITS. These TRF-bound ITSs account for only 8% of the 714 human ITSs listed in the RepeatMasker files from UCSC. The

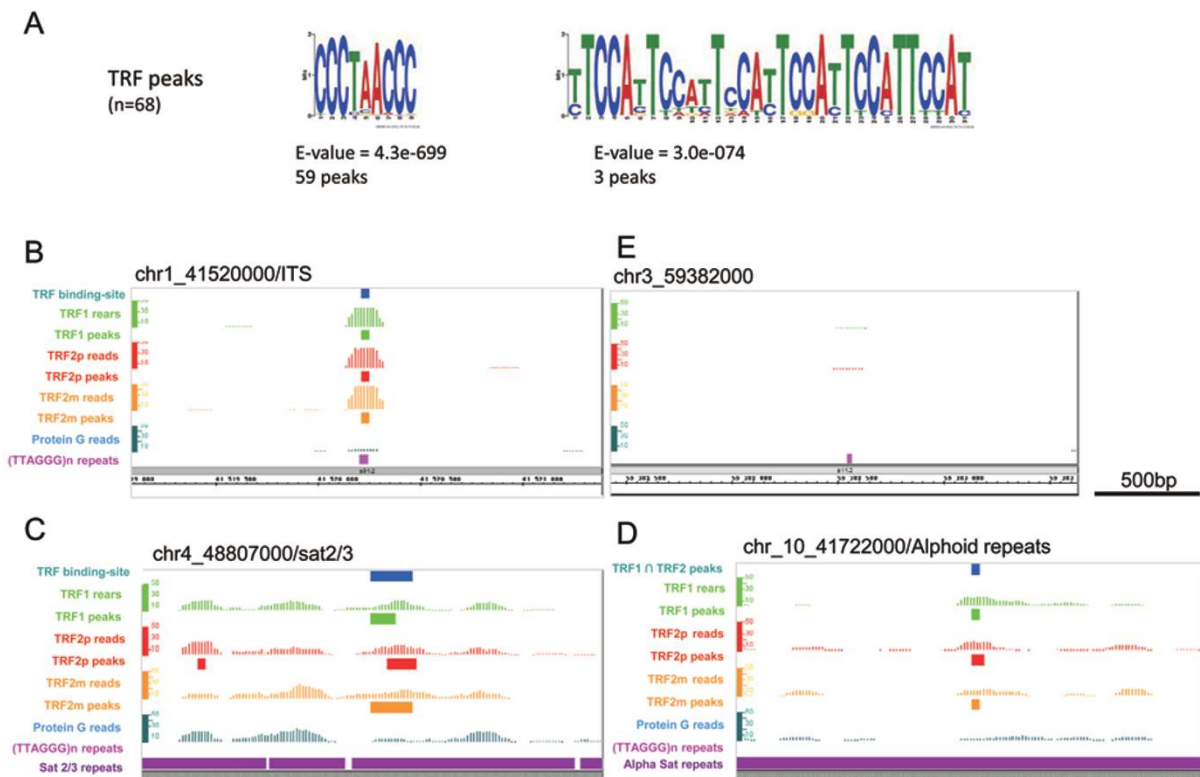


**Figure 3** Validation of the TRF binding sites by ChIP-qPCR with TRF1 and TRF2m antibodies. Enrichment (quantified as the IP/input ratio minus the background ratio (obtained from the protein G ChIP analysis)) of the different loci was normalized to the value for a *GADPH* gene sequence. Three ChIP-qPCR analyses were performed using BJ-HELTRas<sup>mc</sup> cells and SNG28 cells.

non-binding or poor binding of TRF1 and TRF2 to a vast majority of ITSs is in agreement with the results of ChIP-qPCR (Figure 3). This suggests that TRF proteins have a high affinity for only a subset of ITSs, at least in the cell lines used in this study.

Local alignments of TRF-bound and -unbound ITSs

showed that the bound sequences were significantly longer and more conserved than the unbound ones (Supplementary information, Figure S3). An analysis restricted to the well-conserved ITSs (containing at least four TTAGGG units and less than one mismatch per unit, see Supplementary information, Table S1) also revealed a



**Figure 4** (A) Motif prediction analysis of the 68 TRF binding sites, performed using MEME software. The telomeric (TTAGGG)<sub>n</sub> motif and the (ATTC)n motif present in satellite DNA families 2 and 3 were identified. (B) An example of a TRF peak associated with a (TTAGGG)<sub>n</sub> repeat. (C) An example of a TRF peak associated with a Satellite 2/3 sequence. (D) An example of a TRF peak associated with an alphoid satellite sequence. (E) An example of a (TTAGGG)<sub>n</sub> repeat not enriched by Chip-Seq analysis performed using the three anti-TRF antibodies.

statistically higher sequence conservation for the TRF-bound as compared to the -unbound ITSs, both for the subtelomeric and for the internal TRF binding sites (Supplementary information, Figure S4). These results indicate that the primary sequence plays an important role in the ability of ITSs to bind to TRF proteins. In agreement with this conclusion, we previously showed that a 0.8-kb stretch of perfect telomeric repeats inserted artificially into the middle of chromosome 4 was efficiently bound by TRF1 and TRF2 [26] (Figure 3B). However, this might not be the only determinant of the ability of TRF1 and TRF2 to bind to an ITS since some TRF-unbound ITSs display only few mismatches compared to the exact (TTAGGG)<sub>n</sub> array (of those longer than 30 bp, 137 have fewer than 12% mismatches, and 11 of them have none) while some TRF-bound ITSs are highly degenerated (Supplementary information, Figure S4). This suggests that the ability of an ITS to bind TRF proteins is also determined by features other than sequence conservation, such as cell type and chromosomal environment. In agreement with this hypothesis, we found that

one ITS to which TRF2 did not bind in BJ-HELTRas<sup>mc</sup> cells (according to the results of ChIP-qPCR analysis), Chr.1p36.13, was bound by TRF2 in SNG28 cells (Figure 3). Moreover, 38% of the TRF binding sites are located in subtelomeric regions (sequences less than 100 kb from a chromosome end). This indicates a marked preference for these locations, since globally only 10% of all ITSs which are present in RepBase are located in these regions ( $P = 7.72 \times 10^{-10}$ ).

*A subset of TRF binding sites correspond to nontelomeric satellite DNA repeats*

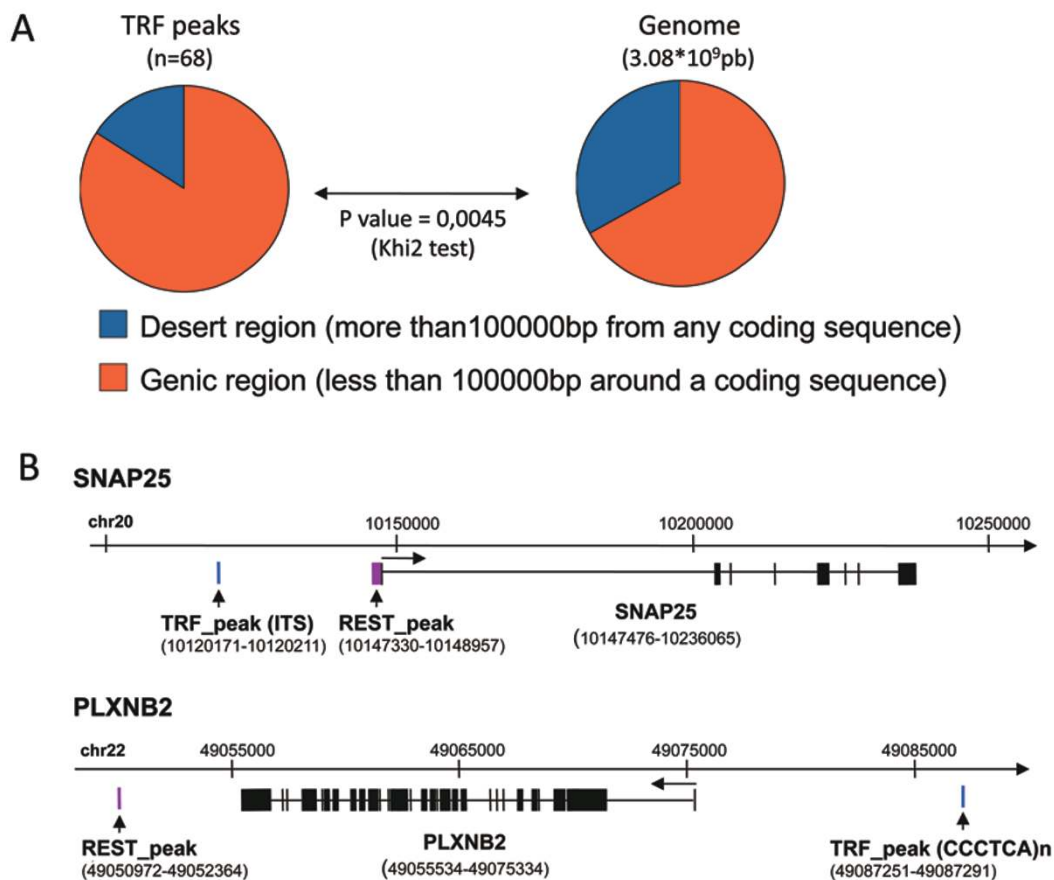
The same motif prediction analysis also identified sequences derived from a consensus consisting of repetitions of the CCATT pentamer [32], which is found in human peri-centromeric *satellite 2/3* sequences, and were identified in three additional peaks (Figure 4A and 4C, Supplementary information, Table S1). The three remaining peaks represented two alphoid satellite sequences and one LINE L1 sequence (Figure 4D, Supplementary information, Table S1). These data show that the TRF pro-

teins can bind to repeated sequences other than telomeric DNA. Interestingly, only a small subset of repetitive DNA regions interacts with TRF1 and TRF2, suggesting that, as observed for ITSS, primary sequence recognition is not the sole determinant of binding.

*TRF binding sites are preferentially located in genic regions*

TRF binding sites were preferentially located less than 100 kb from coding sequences (genic regions) (Figure 5A). The 43 genes located proximal to or containing these peaks can be considered as potential targets of the TRF proteins. Although the size of the gene sample was small, gene ontology (GO) annotation using the Database for Annotation, Visualization and Integration Discovery (DAVID; Supplementary information, Table S2) and Ingenuity Pathway Analysis (Supplementary information, Table S3) revealed a significant over-representation of

genes involved in specific biological functions. These functions include vesicular transport (*SNAP25*, *ARF-GAP3*, and *PACSIN2*) and ion transport (*CACNA1B*, *CLIC6*, and *LCN2*), as well as axon growth (*PLXNB2*, *EHD4*, and *VCAN*). Although the biological significance of these observations remains unclear, it is worth noting that TRF2 is reportedly overexpressed during neuronal differentiation, and that TRF2-REST interaction modulates neuronal gene silencing [21]. We therefore explored whether REST binding sites occur in proximity to TRF binding sites, using chip-seq REST peaks identified by Johnson *et al.* [33]. This hypothesis was confirmed in two cases: TRF-bound ITSS were found 27 kb upstream of the REST binding site in the *SNAP25* gene and 10 kb upstream of the coding sequence of the *PLXNB2* gene, which harbors a REST site in its 3' region (Figure 5B). Interestingly, *SNAP25* expression is up-regulated in cells expressing a dominant-negative *TRF2* allele [21], sug-



**Figure 5 (A)** Classification of the peaks according to their location relative to genic sequences. Note the significant bias in the location of the TRF peaks, and more generally that of the ITSS, such that they tended to occur in genic regions of the genome (defined as sequences located less than 100 kb from any gene) as opposed to gene desert regions (sequences located more than 100 kb from any gene). **(B)** Schematic representations of the *SNAP25* and *PLXNB2* gene regions showing the TRF and REST peaks.

gesting that a TRF2-mediated synergistic interaction between the ITS and the REST sites represses *SNAP25* expression.

## Discussion

In this study, we established the genome-wide DNA-binding profiles for TRF1 and TRF2 to identify the potential target genes and regulatory elements controlled by these telomeric proteins. In order to determine specific sites for TRF proteins, we analyzed statistically significant peaks in three independent ChIP-Seq experiments performed with one TRF1-specific antibody and two different anti-TRF2 antibodies. The results of these three ChIP-Seq experiments overlapped remarkably and allowed the identification of 68 extratelomeric binding sites for TRF1 and TRF2 (Figure 2A). A subset of these sites (10%) was confirmed by independent ChIP-qPCR (Figure 3), further validating the reliability of this list. These TRF binding sites largely, but not exclusively, comprise ITSs (Figure 4 and Supplementary information, Table S1). Their occupancy by TRF proteins is observed in two different tumour cell lines. Whether TRF1 and TRF2 also bind extratelomeric sites in normal healthy cells remains to be determined.

Strikingly, TRF1 and TRF2 bind *in vivo* to only a small fraction of previously reported ITSs. This is in agreement with another ChIP-Seq analysis in human cells for TRF2 and Rap1 [34]. Sequence alignments of bound and unbound ITSs suggest that TRF1 and TRF2 discriminate between different ITSs on the basis of their length and sequence (Supplementary information, Figures S3 and S4). It is likely that other features such as accessibility and/or the chromatin structure of the DNA region surrounding the ITS influence TRF binding. Thus, additional ITSs might be bound if the TRF protein concentration and/or chromatin context is altered. In fact, we noted, for some unbound ITSs, an accumulation of reads that did not satisfy the statistical requirements to be scored as a peak but that can reveal a TRF binding with a low affinity (data not shown).

One unexpected finding of this study was the identification of non-ITS binding sites centered on (ATTCC)<sub>n</sub> satellite 2/3 repeats or alphoid DNA satellite sequences, which form part of the most prominent autosomal heterochromatin blocks. This suggests that a part of TRF binds to extratelomeric heterochromatin regions of the genome. Given the recently reported role of TRF1 and TRF2 in the control of replication fork progression through telomeric chromatin [26, 35, 36], it is possible that these shelterin components play a similar role in other regions of DNA that are difficult to replicate, such as those pack-

aged as heterochromatin.

Our GO data suggest that a large subset of TRF binding sites are biologically relevant because they occur more frequently within or in close proximity to genes than what would be expected by chance. TRF binding sites are frequently located in intronic regions or distant from promoters. Thus, TRF1 and TRF2 possibly regulate gene expression through looping mechanisms or by modifying the chromatin landscape. It is possible that cellular levels of TRF proteins influence their binding to the ITSs, and thus the expression of neighboring genes.

Telomeric factors have long been known to play a role in binding at internal chromosomal locations. The first example of this kind was yeast Rap1, which specifically binds to telomeric DNA and which was identified, at first, as a general regulatory factor. Interestingly, in yeast, telomere alterations can lead to the delocalization from telomeres of Rap1-associated heterochromatin factors that are able to operate at interstitial genomic sites [37, 38]. Based on these yeast results, it is tempting to propose that TRF1 and TRF2 are released from the telomeres after telomere shortening or alteration and subsequently relocalized to ITSs, where they modify the cellular transcriptional program. In mammals, Rap1 does not bind to telomeric DNA directly but does so through an interaction with the protein TRF2. Interestingly, recent analyses revealed numerous Rap1 binding sites throughout the human and mouse genome, which appear to regulate gene expression [27, 34]. Whether these sites are also bound by TRF2 remains unknown.

Overall, our results reveal that TRF1 and TRF2 bind to a number of ITSs and non-telomeric heterochromatin-like repeats of the human genome. These results shed new light on the role of these proteins in the mediation of long-range interactions between telomeres and gene networks, which likely contribute to the control of cell fate by telomeres.

## Materials and Methods

### Chromatin immunoprecipitation

Trypsinized cells were collected in culture medium, washed once in PBS and cross-linked through incubation with formaldehyde (final concentration of 1%) for 10 min. The formaldehyde was quenched with glycine (final concentration 0.125 M), and the cells were washed twice with cold PBS. The cells were disrupted with a Dounce homogenizer. After incubation for 20 min in hypotonic buffer (50 mM Tris, 10 mM KCl, 2 mM EDTA, 0.5% NP40, 0.1% DOC, proteases inhibitors), the pellets were resuspended and sonicated in nucleus lysis buffer (50 mM Tris, 10 mM EDTA, 1% SDS), using a Bioruptor sonicator, until the average fragment size reached 250 bp. After centrifugation at 14 000 r.p.m., the supernatants were transferred and diluted 10-fold to produce the following final concentration of the ChIP buffer: 50 mM Tris, 150 mM NaCl,



2 mM EDTA, 1% Triton, and 0.1% SDS. The precleared sonicates were incubated overnight with the primary antibody. Protein G-Sepharose beads (GE Healthcare) pre-coated with 0.1% BSA were added for a further 2 h. The beads were washed twice with ChIP buffer, twice with high-salt buffer (50 mM Tris, 500 mM NaCl, 2 mM EDTA, 1% Triton, 0.1% SDS), and twice with LiCl buffer (50 mM Tris, 250 mM LiCl, 2 mM EDTA, 0.5% NP40, 0.5% DOC). Chromatin was eluted by vortexing twice with 250  $\mu$ l 1% SDS in 0.1 M NaHCO<sub>3</sub>, followed by incubation for 15 min at 65 °C, and then cross-link was reversed through an overnight incubation at 65 °C in the following buffer: Tris (final concentration 20 mM), NaCl (200 mM), EDTA (2 mM), RNase A (100  $\mu$ g/ml). DNA was purified by incubation with proteinase K (Sigma, 50  $\mu$ g/ml final concentration) for 1 h at 45 °C, followed by classic phenol-chloroform purification and ethanol precipitation steps. We used the following antibodies: TRF1 (abcam 10579, mouse monoclonal), TRF2m (Imgenex 124A, mouse monoclonal), TRF2p (Imgenex 148A, goat polyclonal) and total H3 (abcam 1791, rabbit polyclonal).

#### Library construction and sequencing

For each ChIP sample, 100 ng of DNA was used for library construction. DNA was sheared using the Covaris S2 System to reduce the fragment size down to 60 to –110 bp.

Sheared DNA was end-repaired with an End-It Kit (Epicentre) according to the manufacturer's instructions. Fragments were ligated using the Quick Ligase Kit (NEB), and the P1 and P2 adapters were supplied with the SOLiD Library Oligos kit. Ligated fragments were size-selected on 8% TBE acrylamide/bis-acrylamide gels. After elution, they were nick-translated and amplified using Invitrogen AmpliTaq and pfu DNA polymerase following the manufacturer's instructions. The following conditions were applied: 72 °C for 20 min; 95 °C for 5 min; then 10 cycles of 95 °C for 15 s, 62 °C for 15 s, and 70 °C for 1 min; and finally 70 °C for 5 min. Amplified fragments (150 to 200 bp) were purified on 2% SizeSelect gels (Invitrogen) and quantified on a Bioanalyzer High Sensitivity DNA Chip (Agilent).

Fragment sequencing was achieved through emulsion PCR, bead deposition, and ligation-based sequencing, performed using a SOLiD 3 sequencer according to the manufacturer's manual.

#### Matching

Reads (1.5–2  $\times$  10<sup>7</sup> per sample) were matched against the Human Genome 18, using Corona SOLiD software. Alignments were performed using 50 bp of the reads, then only the first 45 bp from the 5' end for the unplaced reads, and so on, down to 25 bp. Five mismatches were allowed for the 50-bp matching, four for the 35–45-bp matching, three for the 30-bp matching, and two for the 25-bp matching.

#### Peak analysis

We employed SISSR software [28] using uniquely placed reads, with the following settings: *P*-value threshold 0.001 or 0.05, with no more than 1 read per location (“-a” option), a default fragment size of 150 bp, enrichment at both sides of a site required (w/o “-U” option). We refined the peak selection by removing peaks associated with obvious background, by visual inspection of a density profile of the uniquely matched reads, summed in 150-bp windows (a window size that corresponds to the average frag-

ment size) sliding by a 15-bp step. For this, we computed the start position of the reads aligned in both directions, retaining no more than one read per position for each strand. The sums were normalized to the total number of reads for each sequencing reaction, and visualized with the Integrated Genome Browser. By doing so, we removed, respectively, 28%, 32%, and 30% of TRF1, TRF2m, and TRF2p peaks.

Finally, we selected the peaks common to the two TRF2 ChIP analyses and the TRF1 ChIP analysis using the following statistical criteria: at least one peak was identified with a *P* value of < 0.001, and the other two had *P*-values of < 0.05.

#### Sequence and functional genomics analysis

We searched for motifs shared by the TRF peaks using the MEME 4.4 software [39] (options: -mod anr, -nmotifs 10, -evt 1, -minw 6, -maxw 100, -maxsites 1 000 -revcomp). We retrieved the coordinates and the alignment features of the ITSs, sat2/3 and alpha satellite repeats from the repeat masker file from UCSC. We identified the positions of the 68 TRF peaks relative to the genes and to the REST peaks (after coordinates conversion to hg18) using SoleSearch software [40]. Fifty-seven peaks fell within the coding regions or putative regulatory regions (within 100 kb of the CDS), in a total of 43 genes. The repeat coordinates and alignment values were extracted from RepeatMasker files from UCSC (AFA Smit, R Hubley & P Green, *RepeatMasker v3.2.7*) [41].

We analyzed the putative functions of these 43 genes associated with TRF peaks through GO analysis, performed using DAVID version 6.7 [42]. Of these 43 genes, 34 were associated with a GO term. The Ingenuity Pathway Analysis 8.7 (Ingenuity Systems, Inc., Redwood City, CA, USA) was also used to analyze the list of 43 genes. The functional analysis provides the most significant functions and/or diseases in the gene list and the biological categories in which they are classified. *P*-value was calculated using Fisher's exact test. It determines the probability that a specific biological function and/or disease associated to the gene list was due to chance alone.

#### ChIP validation

For slot blotting, purified DNA was denatured in SSC2X by heating at 100 °C for 10 min, before being spotted on Hybond N+ membrane (GE Healthcare) using the Bio-Dot SF system (Biorad, Ivry, France), and crosslinked at 80 °C for 2 h. Membranes were incubated overnight at 65 °C in hybridization buffer (0.5 M NaPO<sub>4</sub> pH 7.2, 7% SDS, 0.1% BSA, 10 M EDTA) containing DIG-labeled (DIG-High Prime kit, Roche Applied Bioscience) telomeric, 400 bp of repeated C3TA2 motif (5'-T2AG3-3' motif), and washed for 30 min in wash buffer 1 (200 mM NaPi, 1% SDS, 1 mM EDTA) and 4 times for 30 min in wash buffer 2 (40 mM NaPi, 1% SDS, 1 mM EDTA) at 65 °C. After exposure, the membrane was stripped in boiling 0.5% SDS for 20 min, and re-probed with DIG-labeled sonicated input DNA representing a non-selective “genomic” probe.

Precipitated and input DNA from independent experiments were quantified by qPCR using primers targeted to unique sequences bordering (1) the TRF binding sites; (2) other peaks identified using one or two antibodies raised against TRF1 and/or TRF2, and ITS; (3) ITSs not associated with peaks or ITSs located 1 000 bp from the nearest TRF binding site. The results were normalized to the value obtained from a region upstream of *GAPDH*

(ENSG00000111640). Primer sequences can be provided upon request.

## Acknowledgments

We thank Marie-Joseph Giraud-Panis (CNRS UMR6267/INSERM U998) for critical reading. We are also grateful to Zhou Songyang (Baylor College of Medicine) for sharing unpublished results. This work was supported by grants from the Association de la recherche contre le Cancer (ARC), the Institut National du Cancer (program TELOFUN), ANR (program TELOREP and INNATELO) and the European Community (TELOMARKER Health-F2-2007-200950). TS and LEZ thank the Fondation de la Recherche Médicale (FRM) and the ARC, respectively, for their fellowships.

## References

- 1 Blackburn EH. Telomere states and cell fates. *Nature* 2000; **408**:53-56.
- 2 Segal-Bendirdjian E, Gilson E. Telomeres and telomerase: from basic research to clinical applications. *Biochimie* 2008; **90**:1-4.
- 3 Giraud-Panis MJ, Pisano S, Poulet A, Le Du MH, Gilson E. Structural identity of telomeric complexes. *FEBS Lett* 2010; **584**:3785-3799.
- 4 Azzalin CM, Reichenback P, Khoriantou L, Giulotto E, Lingner J. Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science* 2007; **318**:798-801.
- 5 Schoeftner S, Blasco MA. Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. *Nat Cell Biol* 2008; **10**:228-236.
- 6 Celli GB, de Lange T. DNA processing is not required for ATM-mediated telomere damage response after TRF2 deletion. *Nat Cell Biol* 2005; **7**:712-718.
- 7 Giraud-Panis MJ, Teixeira MT, Geli V, Gilson E. CST meets shelterin to keep telomeres in check. *Mol Cell* 2010; **39**:665-676.
- 8 Gottschling DE, Aparicio OM, Billington BL, Zakian VA. Position effect at *S. cerevisiae* telomeres: reversible repression of Pol II transcription. *Cell* 1990; **63**:751-762.
- 9 Baur JA, Zou Y, Shay JW, Wright WE. Telomere position effect in human cells. *Science* 2001; **292**:2075-2077.
- 10 Koering CE, Pollice A, Zibella MP, et al. Human telomeric position effect is determined by chromosomal context and telomeric chromatin integrity. *EMBO Rep* 2002; **3**:1055-1061.
- 11 Ottaviani A, Gilson E, Magdinier F. Telomeric position effect: from the yeast paradigm to human pathologies? *Biochimie* 2008; **90**:93-107.
- 12 Blasco MA. The epigenetic regulation of mammalian telomeres. *Nat Rev Genet* 2007; **8**:299-309.
- 13 Deng Z, Norseen J, Wiedmer A, Riethman H, Lieberman PM. TERRA RNA binding to TRF2 facilitates heterochromatin formation and ORC recruitment at telomeres. *Mol Cell* 2009; **35**:403-413.
- 14 Ruiz-Herrera A, Nergadze SG, Santagostino M, Giulotto E. Telomeric repeats far from the ends: mechanisms of origin and role in evolution. *Cytogenet Genome Res* 2008; **122**:219-228.
- 15 Azzalin CM, Nergadze SG, Giulotto E. Human intrachromosomal telomeric-like repeats: sequence organization and mechanisms of origin. *Chromosoma* 2001; **110**:75-82.
- 16 Ambrosini A, Paul S, Hu S, Riethman H. Human subtelomeric duplicon structure and organization. *Genome Biol* 2007; **8**:R151.
- 17 Nergadze SG, Santagostino MA, Salzano A, Mondello C, Giulotto E. Contribution of telomerase RNA retrotranscription to DNA double-strand break repair during mammalian genome evolution. *Genome Biol* 2007; **8**:R260.
- 18 Ijdo JW, Baldini A, Ward DC, Reeders ST, Wells RA. Origin of human chromosome 2: an ancestral telomere-telomere fusion. *Proc Natl Acad Sci USA* 1991; **88**:9051-9055.
- 19 Bradshaw PS, Stavropoulos DJ, Meyn MS. Human telomeric protein TRF2 associates with genomic double-strand breaks as an early response to DNA damage. *Nat Genet* 2005; **37**:193-197.
- 20 Deng Z, Lezina L, Chen CJ, et al. Telomeric proteins regulate episomal maintenance of Epstein-Barr virus origin of plasmid replication. *Mol Cell* 2002; **9**:493-503.
- 21 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.
- 22 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.
- 23 Smogorzewska A, van Steensel B, Bianchi A, et al. Control of human telomere length by TRF1 and TRF2. *Mol Cell Biol* 2000; **20**:1659-1668.
- 24 Krutilina RI, Smirnova AN, Mudrak OS, et al. Protection of internal (TTAGGG)<sub>n</sub> repeats in Chinese hamster cells by telomeric protein TRF1. *Oncogene* 2003; **22**:6690-6698.
- 25 Mignon-Ravix C, Depetris D, Delobel B, Croquette MF, Mattei MG. A human interstitial telomere associates *in vivo* with specific TRF2 and TIN2 proteins. *Eur J Hum Genet* 2002; **10**:107-112.
- 26 Ye J, Lenain C, Bauwens S, et al. TRF2 and apollo cooperate with topoisomerase 2alpha to protect human telomeres from replicative damage. *Cell* 2010; **142**:230-242.
- 27 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.
- 28 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.
- 29 Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 2002; **30**:207-210.
- 30 Desmaze C, Alberti C, Martins L, et al. The influence of interstitial telomeric sequences on chromosome instability in human cells. *Cytogenet Cell Genet* 1999; **86**:288-295.
- 31 Mondello C, Pirzio L, Azzalin CM, Giulotto E. Instability of interstitial telomeric sequences in the human genome. *Genomics* 2000; **68**:111-117.
- 32 Lee C, Wevrick R, Fisher RB, Ferguson-Smith MA, Lin CC. Human centromeric DNAs. *Hum Genet* 1997; **100**:291-304.

- 33 Johnson DS, Mortazavi A, Myers RM, Wold B. Genome-wide mapping of *in vivo* protein-DNA interactions. *Science* 2007; **316**:1497-1502.
- 34 Yang D, Xiong Y, Kim H, *et al.* Human telomeric proteins occupy selective interstitial sites. *Cell Res* 2011 Mar 22. doi:10.1038/cr.2011.39
- 35 Sfeir A, Kosiyatrakul ST, Hockemeyer D, *et al.* Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication. *Cell* 2009; **138**:90-103.
- 36 Martínez P, Thanasoula M, Muñoz P, *et al.* Increased telomere fragility and fusions resulting from TRF1 deficiencies lead to degenerative pathologies and increased cancer in mice. *Genes Dev* 2009; **23**:2060-2075.
- 37 Maillet L, Boscheron C, Gotta M, *et al.* Evidence for silencing compartments within the yeast nucleus: a role for telomere proximity and Sir-protein concentration in silencer-mediated repression. *Genes Dev* 1996; **10**:1796-1811.
- 38 Marcand S, Buck SW, Moretti P, Gilson E, Shore D. Silencing of genes at nontelomeric sites in yeast is controlled by sequestration of silencing factors at telomeres by Rap1 protein. *Genes Dev* 1996; **10**:1297-1309.
- 39 Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* 1994; **2**:28-36.
- 40 Blahnik KR, Dou L, O'Geen H, *et al.* Sole-Search: an integrated analysis program for peak detection and functional annotation using ChIP-seq data. *Nucleic Acids Res* 2010; **38**:e13.
- 41 Jurka J, Kapitonov VV, Pavlicek A, *et al.* Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* 2005; **110**:462-467.
- 42 Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009; **4**:44-57.

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