Published in final edited form as:

Nat Cell Biol. 2010 August ; 12(8): 768–780. doi:10.1038/ncb2081.

Mammalian RAP1 controls telomere function and gene expression through binding to telomeric and extra-telomeric sites

Paula Martinez¹, Maria Thanasoula², Ana R. Carlos², Gonzalo Gómez³, Agueda M. Tejera¹, Stefan Schoeftner¹, Orlando Dominguez⁴, David G. Pisano³, Madalena Tarsounas², and Maria A. Blasco^{1,5}

¹Telomeres and Telomerase Group, Molecular Oncology Program, Spanish National Cancer Centre (CNIO), Melchor Fernández Almagro 3, Madrid, E-28029, Spain

²Telomere and Genome Stability Group, The CR-UK/MRC Gray Institute for Radiation Oncology and Biology, Old Campus Road, Oxford OX3 7DQ, U.K.

³Bioinformatics Core Unit, Structural Biology and Biocomputing Program, Spanish National Cancer Centre (CNIO), Melchor Fernández Almagro 3, Madrid, E-28029, Spain

⁴Genomics Core Unit, Biotechnology Program, Spanish National Cancer Centre (CNIO), Melchor Fernández Almagro 3, Madrid, E-28029, Spain

Abstract

Shelterin binds and protects mammalian telomeres. Here, we generated cells and mice conditionally deleted for the shelterin component RAP1. We find that *Rap1* deficiency is dispensable for telomere capping but leads to increased telomere recombination and fragility. Mice with *Rap1* deletion in stratified epithelia are viable but have shorter telomeres and develop skin hyperpigmentation at adulthood. By performing chromatin immunoprecipitation coupled with ultra-highthroughput sequencing, we find that RAP1 binds to telomeres and to extra-telomeric sites through the (TTAGGG)₂ consensus motif. Extra-telomeric RAP1 binding sites are enriched at subtelomeric regions, in agreement with preferential deregulation of subtelomeric genes in *Rap1*-deficient cells. More than 70% of extra-telomeric RAP1 binding sites are at the vicinity of genes and 31% of the genes deregulated in *Rap1*-null cells contain RAP1 binding sites, suggesting a role of RAP1 in transcriptional control. These findings place a shelterin component at the interface between telomere function and transcriptional regulation.

Keywords

telomeres; RAP1; TERT; TPE; subtelomeres; DNA damage; transcription; ChIP-seq; metabolism; conditional *knock out* mice

⁵Correspondence should be addressed to M.A.B

Authors contributions MAB conceived the original idea. MAB and PM designed experiments and wrote the manuscript. PM performed most of the experiments. MTA, ARC and MT contributed with results shown in Fig. 2A-B and Fig. 3A-F. AT performed experiments in Fig. 2C-D and Fig. S.E. SS designed the knockout allele. OD performed DNA methylation analysis, microarray and ChIP-seq experiments. GG and DGP performed microarray and ChIP-seq data analysis.

Competing Interests The authors declare that have no competing financial interests.

Accession Numbers The microarray and ChIP-seq datasets have been deposited in the GEO database under the accession numbers GSE19011 and GSE20867, respectively.

Additional Methods are described in the Supplementary Methods section in Supplementary Information

Introduction

RAP1 (or Terf2ip) is part of the shelterin complex at mammalian telomeres, also encompassing TRF1, TRF2, POT1, TPP1 and TIN2¹. Human RAP1, hRAP1, was identified as a TRF2-interacting protein with homology to budding yeast scRap1^{2,3}. scRap1 binding to telomeres controls telomere length and subtelomeric silencing⁴⁻¹⁰. In addition, scRap1 acts as a transcription factor by controlling the expression of glycolytic enzymes and ribosomal genes^{11,12}. While scRap1 binds telomeric DNA directly, hRAP1 is proposed to be recruited to telomeres by TRF2³. *hRap1* over-expressing cells show continuous telomere lengthening³. Potential roles of mammalian RAP1 in transcriptional regulation and subtelomeric gene silencing are unknown.

Here, we study the role of mouse RAP1 in telomere biology and its impact on gene expression programs and organismal viability by generating *Rap1* tissue-specific conditionally deleted cells and mice. We show that mammalian RAP1 is dispensable for telomere capping but prevents telomere recombination and fragility. Similar findings have been recently reported for an independent conditional knock-out allele of *Rap1*¹³. In contrast to analogous mouse models for TRF1 and TPP1 deficiency^{14,15}, targeted *Rap1* deletion to stratified epithelia does not impact on mouse viability but leads to skin hyperpigmentation at adulthood and presence of shorter telomeres.

Besides its telomeric roles, we demonstrate a role of RAP1 in repression of subtelomeric genes, analogous to that of yeast and *Trypanosoma brucei* RAP1 proteins¹⁶. Moreover, we find that RAP1 binds to both telomeric and extra-telomeric TTAGGG repeats. Extra-telomeric RAP1 binding sites are enriched at subtelomeric regions, in agreement with preferential derepression of subtelomeric genes in *Rap1*-deficient cells. Notably, more than 70% of RAP1 binding sites are found at intragenic positions or at the vicinity of gene-coding chromatin. By performing gene expression studies, we further show that more than a third of the genes significantly deregulated in *Rap1*-deficient cells contain RAP1 binding peaks. In summary, we identify RAP1 as an important factor for telomere integrity and for transcriptional gene regulation in mammals. These findings provide evidence that a shelterin component can bind to non-telomeric regions and have extra-telomeric functions.

Results

Rap1 conditional knockout mice and cells

We generated a conditional *Rap1* knockout mouse, *Rap1^{flox/flox}*, by flanking *Rap1* exon 3 (E3) by *loxP* sites (Fig. 1A). We included *frt* sites to remove the neomycin gene (Fig. 1A)¹⁷. This targeting strategy does not interfere with the essential gene *Kars*¹⁸, which shares promoter with *Rap1*. Deletion of *Rap1* E3 eliminates the *Rap1* telomere localization domain (RCT) and the nuclear localization signal (NLS)(Fig. 1B). Deletion of *Rap1* E3 also disrupts of the 3' UTR and polyadenylation signal.

Wild-type and *Rap1^{flox/flox}* mouse embryonic fibroblasts (MEFs) infected with a pBabe-Cre retrovirus showed poor proliferation *in vitro* (Fig. S1A). Both genotypes proliferated normally by downregulating p53 using a small hairpin RNA (shRNA) against p53 (*Rap1^{+/+}* - shp53-Cre; Fig. 1C and Fig. S1B) or by cancelling both the p53 and Rb pathways using SV40 large T (LT) antigen (*Rap1^{+/+}* -LT-Cre; Fig. 1C and Fig. S1C). *Rap1^{+/+}* -LT-Cre MEFs also showed a similar cell cycle profile to control MEFs (Fig. S1D).

Rap1^{+/+} -LT-Cre cells showed normal expression of *Rap1* exons 1 and 2 and of *Kars* mRNA, while *Rap1* E3 transcripts were undetectable (Fig. 1D). Two antibodies against full-length RAP1 recognized a low abundance protein (25-fold less abundant than full-length RAP1) of

approximately 40 kD in *Rap1*^{-/-}-LT-Cre cells that could correspond to a truncated RAP1 protein (RAP1^{tr}, Fig. 1E). RAP1 antibodies immunoprecipitated RAP1^{tr} only in *Rap1*-null MEFs but not in wild-type controls (Fig. 1F). RAP1^{tr} was undetectable in nuclear extracts (Fig. 1E) or in the chromatin-bound fractions (Fig. 2A) of *Rap1*^{-/-}-LT-Cre cells, in agreement with deletion of the NLS and RCT regions. We confirmed defective RAP1 binding to telomeres in *Rap1*-null cells by immunofluorescence (Fig. 1G; Fig. 2E) and ChIP-sequencing experiments (Fig. 7A). Based on the low and restricted (cytoplasm) expression of RAP1^{tr}, it is unlikely that it might exert significant dominant negative effects at telomeres in *Rap1*-null cells, although potential dominant negative effects on other cellular pathways cannot be discarded.

RAP1 abrogation does not disrupt binding of other shelterins to telomeres

In chromatin fractionation assays, binding of TRF1, TRF2, POT1, TPP1 and TIN2 to chromatin was normal in the absence of RAP1 protein (Fig. 2A,B). As control, histone H3 showed normal binding to chromatin in both genotypes (Fig. 2A). These results were extended to telomeric chromatin by using chromatin immunoprecipitation (ChIP) assays (Fig. 2C,D), as well as by immunofluorescence staining, which showed normal binding of TRF1, TIN2, TRF2 and TPP1 to RAP1-depleted telomeres (Fig. 2E).

No fusions but increased telomere fragility and recombination in *Rap1*-null cells

To address DNA damage at telomeres in the absence of RAP1, we determined their co-localization with γH2AX foci, the so-called TIFs^{19, 20}. *Rap1* deletion²¹ resulted in > 70% of *Rap1*^{-/-} cells showing > 3 TIFs per metaphase, while these events were rare in control cells (Fig. 3A-D). We detected phosphorylation of the CHK1 and CHK2 checkpoint kinases in *Rap1*^{-/-} cells but not in wild-type cells (Fig. 3E). As control, *TRF1*-null cells showed robust activation of these kinases¹⁴ (Fig. 3F). Chemical inhibition of CHK1 and CHK2 rescued both cells with DNA damage and cells showing > 3 TIFs (Fig. 3G-I), indicating that both pathways are activated upon *Rap1* deletion.

Rap1 deletion did not lead to telomere fusions in MEFs (Fig. 3J). Instead, *Rap1*-deleted MEFs showed increased frequencies of chromosome ends with more than one telomere signal or multitelomeric signals (MTS; Fig. 3J), recently related to telomere fragility^{14, 22-24}. MTS were increased in wild-type MEFs treated with drugs that induce telomere fragility, such as aphidicolin, and further increased in similarly treated *Rap1*-deficient MEFs (Fig. 3J). Chromosome breaks were also increased in aphidicolin-treated *Rap1*-null MEFs compared to wild-type controls (Fig. 3J). These results indicate that RAP1-depleted telomeres are prone to fragility but not fusion.

Sister telomere recombination (T-SCE) frequencies were also elevated in *Rap1*^{-/-}-Cre-LT MEFs compared to wild-type controls (Fig. 3K,L), supporting the recently proposed role of RAP1 in preventing telomere recombination¹³.

Finally, telomere quantitative FISH (Q-FISH) analysis on metaphases, indicated slightly shorter telomeres in passage 3 *Rap1*^{-/-}-Cre-LT MEFs compared to wild-type controls, including increased abundance of short telomeres and decreased abundance of long telomeres (Fig. S2A-C). This shortening was not associated to decreased TERT mRNA and protein levels, or to lower *in vitro* telomerase activity in *Rap1*-null MEFs (Fig. S2D-F). A number of recently identified TERT target genes²⁵, also showed a similar expression in wild-type and *Rap1*-null MEFs (Fig. S2F).

Mice with Rap1 deletion in stratified epithelia are viable but develop skin hyperpigmentation at adulthood

Next, we targeted *Rap1* deletion to stratified epithelia in *Rap1*^{+/+} *K5-Cre* mice (Fig. 4A)^{26,27}. These mice show excision of *Rap1* E3 in the epidermis (E) but not in the dermis (D) (Fig. 4A) in agreement with *K5-Cre* expression pattern²⁷. Accordingly, RAP1 immunofluorescence was undetectable in *Rap1*^{+/+} *K5-Cre* epidermis (Fig. 4B,C).

Rap1^{+/+} *K5-Cre* mice are viable and survive to adulthood (Fig. 4D), however, they develop skin hyperpigmentation at adulthood (Fig. 4E). In line with this, *Rap1*^{+/+} *K5-Cre* epidermis showed shorter telomeres (Fig. 4F,G) and increased abundance of cells with DNA damage (H2AX-positive) (Fig. 4H) compared to age-matched controls.

Preferential derepression of subtelomeric genes in *Rap1*^{Δ/Δ} cells

To address a role for RAP1 in subtelomeric gene regulation, we performed gene expression analysis of wild-type and *Rap1*-null MEFs. We pre-defined subtelomeric (S) regions as the 3Mbp regions adjacent to telomeres, which allowed us to recruit a robust number of genes located in S regions (n=917) and a total of 40260 genes located in non-subtelomeric (NS) regions. Genes located in S regions showed significant differences in transcriptional activity when compared to genes located in NS regions ($p = 1.044 \times 10^{-5}$, Fig. 5A). In both gene sets (S and NS), the median distribution of $\log_2$ *Rap1*^{+/+}/*Rap1*^{Δ/Δ} values was displaced from zero (Fig. 5A), indicating a genome-wide effect of *Rap1*-deletion on gene expression. This effect was more pronounced in S regions compared to NS regions (median displacement 0.04 versus 0.01, respectively; Fig. 5A). Moreover, S regions in *Rap1*-deficient cells were significantly overexpressed compared to S regions in wild-type cells ($p < 10^{-6}$; Fig. 5A).

Gene set enrichment analysis (GSEA) showed that S regions in chromosomes 1, 2, 11, 15 and 19 were significantly enriched (FDR<0.25) in *Rap1*-deficient cells compared to wild-type cells (Fig. 5B), while only one S region in chromosome 7 showed the inverse pattern (Fig. 5C). This is also illustrated by representing the $\log_2$ *Rap1*^{+/+}/*Rap1*^{Δ/Δ} expression value for each individual S gene at its chromosomal location (Fig. 5D). Thus, *Rap1* deletion results in gene expression changes that are significantly more pronounced at subtelomeric regions, where it primarily leads to augmented gene expression. Finally, this effect was not associated with changes in subtelomeric DNA methylation²⁸ (Fig. S3A,B), global DNA methylation (B1 SINE repeated elements; Fig. S3C), or telomeric transcription (TERRA, Fig. S3D,E)^{29,30}.

RAP1 transcriptional networks include cell adhesion, cancer and metabolism

Apart from changes in subtelomeric gene expression, a stringent (FDR<0.15) differential gene expression analysis revealed 15 genes significantly downregulated and 1 gene significantly upregulated in *Rap1*-null MEFs (FDR<0.15) (Fig. 6A; Tables S1 and S2). All 16 differentially expressed genes (DEGs) were validated by quantitative PCR (q-PCR) (Fig. 6B). *Rap1* was the most downregulated gene (>8-fold by microarray analysis and >200-fold by qPCR analysis) (Fig. 6A,B; Table S1), while the rest of DEGs changed by >4-fold ($-2 > \log_2 FC > 2$; Table S1). Interestingly, 8 out of the 16 DEGs are related to cancer as indicated by GSEA (Fig. 6A-C).

GSEA and Gene Ontology (GO) analysis on genes downregulated by at least 2-fold in *Rap1*-null MEFs ($\log_2 FC < -1$) extended the analysis to a total of 478 genes (Tables S3 and S4-6), including *Igf2* (Fig. 6B). The downregulated genes *H19* and *Igf2* (Fig. 6A,B) are imprinted genes. GSEA for imprinted genes confirmed transcriptional repression of many imprinted genes in *Rap1*-null cells (FDR<0.01) (Fig. S4A; Tables S3 and S7). GSEA analysis also revealed significant downregulation of genes involved in cell adhesion and metabolism in

Rap1-null cells (Fig. S4B-F; Tables S8-11). These included genes of the insulin secretion, PPAR signaling, and growth hormone (GH) pathways, including *Igf2* (Fig. 6B). In turn, *Rap1*-null cells showed significant upregulation of ABC transporters (FDR=0.216) and genes involved in Type II Diabetes (FDR=0.067; Fig. S4D; Table S12), suggesting a negative impact of *Rap1* deletion on metabolism.

Two highly downregulated genes in *Rap1*-null cells, NNMT (Nicotinamide N-methyltransferase), a regulator of lifespan linked to the sirtuin pathway, and CTGF (connective tissue growth factor) (Fig. 6A,B), were recently found upregulated in calorie restricted (CR) mice³¹. Comparison of the CR gene set³¹ to genes downregulated in *Rap1*-null MEFs indicated that CR gene set is significantly downregulated in *Rap1*-null cells (Table S13), with both gene sets showing an inverse pattern (Fig. 6D). This includes PGC1, a regulator of lipid metabolism³², which we further confirmed by qPCR (5-fold decrease in *Rap1*-null MEFs compared to controls, Fig. 6E). Together, these results indicate that absence of *Rap1* leads to gene expression changes with a predicted negative impact on metabolism.

***Rap1*^{Δ/Δ}K5-Cre females show increased body weight**

We noted that *Rap1*^{Δ/Δ}K5-Cre female mice weighted approximately 10% more than age-matched wild-type littermates from week 14 of age onwards, a phenotype which was not apparent at birth and was not observed in males (Fig. 4D; Fig. S5A,B). To address whether this could be related to abnormal expression of key metabolic genes upon *Rap1* abrogation, we subjected male and female mice to either a standard diet (SD) or a high fat diet (HFD) (Fig. S5C,D). While wild-type and *Rap1*^{Δ/Δ}K5-Cre male mice gained weight following placement on HFD compared to the same genotypes in SD, *Rap1*^{Δ/Δ}K5-Cre females on a SD gained weight at the same rate than wild-type females on a HFD under conditions of similar food intake (Fig. S6C-E), and this was not increased by placing them on a HFD. Thus, *Rap1* deletion in females phenocopies exposure to a HFD. A similar phenotype of increased body weight in female mice is reported for *PGC1*-deficiency³³.

Mapping *in vivo* Rap1 binding sites by using ChIP-sequencing

To study the role of RAP1 in subtelomeric gene silencing and transcriptional regulation, we set to determine the *in vivo* RAP1 binding loci to the chromatin by using whole genome ChIP-sequencing (ChIP-seq) technology (Fig. S6A). *Rap1*-null MEFs were used as control for RAP1 peak specificity. Over 21 and 17 million of uniquely mapped 36 base reads were collected for wild-type and *Rap1*-null cells, respectively (Table S14). In agreement with RAP1 binding to telomeres, we found a strong overrepresentation of raw 36bp sequences containing the telomeric TTAGGG₅ repeat in WT *Rap1* ChIP-seq compared to *Rap1*-null controls, which showed similar levels to input DNA (Fig. 7A). This difference was specific and not observed with non-telomeric repeats (Table S15; Fig. S6B). We also found perfect permutations of at least 2 telomeric repeats in the sequences of extra-telomeric RAP1 binding sites, in a frequency 32-fold higher (p-value < 2.2e-16) than that found in the mouse genome (Fig. 7B; Fig S6C). In total, between 2%-4% of all the genomic occurrences of telomeric repeats are occupied by RAP1 binding sites (Fig. S6D). When allowing 1 mismatch, the telomeric motif TTAGGG was found at least one time in up to 37.2% of the RAP1 peak sequences (Fig. S6E,F), suggesting that RAP1 binds to non-telomeric regions enriched in telomeric repeats.

In total, 30,398 non-telomeric binding sites (average peak width, 75bp) were identified upon comparing *Rap1* WT and *Rap1*-null ChIP-seq datasets at 10% FDR level (Fig. S7A for peaks at 5% and 1% FDR levels). RAP1 sites were found throughout the genome (Fig. 7C), but showed a noticeable enrichment at the subtelomeric regions (17.45 sites/Mbp) (Fig. 7D)

compared to the genome-wide density (11.71 sites/Mbp). The density of subtelomeric RAP1 peaks decreased proportionally to the distance from the telomere, suggesting a gradient of RAP1 binding from telomeres into the subtelomeric regions (Fig. 7E). The same trend was observed for the genomic distribution of (TTAGGG)₂ repeats (Fig. S7B). Both signals showed a mild correlation ($r=0.46$; Fig. S7C), which was increased when considering regions closer to telomeres ($r=0.89$; Fig. S7C), suggesting a preference for repeat-mediated RAP1 binding at subtelomeres and the possibility of alternative binding mechanisms for the rest of the genome. The X chromosome showed the lowest density of RAP1-binding sites (2.65 RAP1 sites/Mbp; Fig. S7D), which was not attributable to differences in karyotype (all MEFs used were XY). Genes located in the X chromosome suffered less expression changes in *Rap1*-null cells compared to genes in other chromosomes (Fig. S7E), suggesting a correlation between RAP1 binding and gene expression changes.

Up to 22,384 individual RAP1 binding sites (73.63%) could be associated to genes, either mapping inside the gene structure (13,057 sites, 42.95%), or in a 10kbp window surrounding the gene transcription limits (9,327 sites, 30.68%) (Fig. 7F). The remaining 8,014 sites (26.36%) were considered intergenic (Fig. 7F). The increased density of RAP1 peaks around gene-coding regions resembles the genomic location patterns of known transcription factors. Given that coding genes and their 10kbp neighbor regions span about 47% of the mouse genome, this observation discards randomness in the topology of extra-telomeric RAP1 binding distribution ($p\text{-value} < 2.2\text{e-}16$), and supports that RAP1 may influence gene transcription. When exploring a 10kbp window surrounding gene transcription start sites (TSS) sites, up to 6,988 (22.98%) of the RAP1 sites could be associated to genes present in the gene expression array. Despite some strong RAP1 binding sites not associated to genes (Fig. 7H, peak in chromosome 2), 173 (31%) of the 558 deregulated ($|\log\text{FC}| \geq 1$) genes presented 234 RAP1 binding sites mapping to their promoter regions (Table S16; Fig. 7G). When comparing to the 4,980 non-deregulated genes with peaks, we found that the peak presence–gene deregulation dependency is significant ($p\text{-value} = 0.0003$), thus supporting a non-random relationship between RAP1 binding and transcriptional gene regulation. Significantly downregulated genes in *Rap1*-null cells containing RAP1 binding sites included *Rap1* itself, *Hic1* (downregulated 4-fold), *Fgf5* (downregulated 5-fold), and *Ctgf* (downregulated 7-fold) (Fig. 7H; Table S16; Fig. 6A,B), suggesting that RAP1 acts as a transcriptional activator in these genes (see Fig. 8E). Together, these results suggest that RAP1 binding to extra-telomeric sites has an impact on transcriptional gene regulation.

Identification of a consensus RAP1 binding sequence to extra-telomeric sites

To determine the sequence specificity of extra-telomeric mammalian RAP1 binding, we used the de-novo motif discovery algorithm Weeder³⁴ on a subset of the top over-represented RAP1-binding peaks ($n=30$). The highest-ranking motif was (TTAGGG)₂ (Fig. 8A), probably due to presence of at least a perfect TTAGGG motif in up to 2.91% of the RAP1 peak sequences ($n=884$), and in up to 37.28% ($n=11,332$) when allowing 1 mismatch (Fig. S6F). A clear over-representation of the telomeric motif was found in the top 1000 and top 100 ranking peaks (Fig. 8B), with RAP1 peaks in chromosomes 11 and 17 showing more than 20 TTAGGG tandem repeats and the top ranking peak corresponding to a perfect (TTAGGG)₈ repeat at chromosome 2 (Fig. 8B,C). This over-representation is evident in the peak sequences when compared to the aligned WT reads (Fig. S7F,G). When examining strongly deregulated genes, including *Ctgf* and *Fgf5* genes (Table S17), *Olfm1*, *Crabp1* and *Angptl4*, we found the telomeric consensus motif in their RAP1 binding peaks (Table S17; Fig. 8C for *Olfm1*, *Crabp1* and *Angptl4*). We detected imperfect telomeric motifs in the peak sequences associated to half (51.44%) of DEGs (Table S17), a proportion similar to RAP1 peaks in non-deregulated genes ($p=0.31$). Given the significant association between

RAP1 peaks and gene deregulation, these results suggest participation of RAP1 in alternative regulation mechanisms not mediated by telomeric repeat binding.

To explore non TTAGGG-rich putative RAP1 binding motifs we runned Weeder after discarding the top ranking peaks with 3 telomeric repeats. We obtained no clear consensus apart from heterogeneous A/T-rich sequences common for many transcription factors (Table S18).

To validate the ChIP-seq results, we chose a total of 14 RAP1-binding peaks located at both non-coding chromatin (Chrs. 2, 3, 17) and at strongly deregulated genes in wild-type versus *Rap1*-null MEFs (Fig. 8C,D; Table S15). As indicated by ChIP-qPCR, RAP1 showed binding to these regions and this was reduced in *Rap1*-null cells (Fig. 8D). No specific RAP1 binding was detected at chromatin regions that do not contain RAP1 binding sites (NC1 and NC2; Fig. 8D).

To address whether RAP1 binding to genes impacts their transcriptional activity, we cloned genomic fragments containing RAP1 binding sites at the promoter regions of *Ctgf*, *Hic1* and *Angptl4* upstream a minimal promoter driving luciferase expression (Fig. 8E). Upon transfection into wild-type and *Rap1*^{-/-} cells, luciferase activity was significantly decreased in *Rap1*^{-/-} cells compared to wild-type cells for all constructs tested while no differences were observed with the empty vector. These results suggest that these genomic regions have RAP1-dependent enhancer activity, supporting role for RAP1 in transcriptional regulation.

As RAP1 cannot bind chromatin directly and its recruited to telomeres by TRF2³, we addressed whether TRF2 was able to bind to TTAGGG-rich extratelomeric RAP1 binding sites, such as those present in top ranking peaks at chromosomes 2 and 17. As determined by ChIP, TRF2 was bound to peaks in chromosomes 2 and 17 and this binding was independent of RAP1 (Fig. 8F). No significant TRF2 binding was detected at the RAP1 peak within the *Crabp1* promoter (Fig. 8F), containing a degenerated telomeric repeat (Fig. 8C). Thus, RAP1 is likely to be recruited to TTAGGG-rich extratelomeric regions by TRF2 (ie, peaks at chromosomes 2 and 17), while other factors may mediate RAP1 recruitment to degenerated telomeric repeats at the *Crabp1* promoter.

Discussion

Here, we describe cells and mice deleted for *Rap1*. While previous studies support a RAP1-dependent pathway for protection against telomere fusions^{35, 36}, our current results suggest that chromatin-bound RAP1 does not significantly inhibit end-to-end fusions in MEFs. Instead, *Rap1*-deficient MEFs show increased frequencies of multitelomeric signals associated to telomere fragility^{14, 24}, and elevated telomere recombination. This is in agreement with a recent report showing a role for RAP1 in preventing telomere recombination but not telomere fusions¹³. In agreement with normal telomere capping, mice with targeted *Rap1* deletion to stratified epithelia are viable and only show mild skin hyperpigmentation at adulthood. This is in marked contrast to 100% perinatal lethality and severe skin morphogenesis defects in similarly generated *TRF1*- and *TPP1*-null mice, with severe telomere capping defects^{14, 15}.

Rap1-deleted MEFs and mice showed a modest telomere shortening compared to controls, which was more apparent in *Rap1*-null proliferating tissues (ie, skin basal layer). This is in analogy to human diseases with presence of short telomeres due to telomerase mutations³⁷⁻⁴¹, also characterized by skin hyperpigmentation. As *Rap1*-null cells have normal telomerase expression, accelerated telomere shortening in *Rap1*-null epidermis may result from increased telomere recombination and fragility, although a role of RAP1 in TERT recruitment to telomeres cannot be discarded.

Yeast scRap1 and *Trypanosoma brucei* tpRap1 silence genes located near the telomeres, a phenomenon known as telomere position effect ^{6, 42}. Here, we show preferential up-regulation of genes located in subtelomeric positions in *Rap1*-deficient cells compared to wild-type controls, thus suggesting a conserved role of RAP1 in subtelomeric gene silencing.

We show that *Rap1*-dependent transcriptional changes also include genes involved in cancer, cell adhesion, and metabolism. In particular, *Rap1* deletion results in gene expression changes with a predicted negative impact on metabolism, in agreement an increased body weight phenotype in *Rap1*-null females. Of note, the transcriptional changes induced by *Rap1* deletion are different from those associated to severe telomere dysfunction which involve activation of a stress response ⁴³, thus although possible, it is unlikely that RAP1-dependent transcriptional effects at subtelomeres and elsewhere in the genome are due to mild telomere dysfunction induced by *Rap1* deletion.

By using ChIP-seq technology, we demonstrate that RAP1 binds to telomeric repeats and to extra-telomeric sites preferentially throughout recognition of the (TTAGGG)₂ consensus motif (Fig. 8E). Extra-telomeric RAP1 binding sites are enriched at subtelomeric regions and this enrichment decreases proportionally to the distance from the telomeres, providing a mechanism by which mammalian RAP1 can influence subtelomeric gene regulation. In addition, RAP1 binding sites are present in a significant percentage (31%) of the differentially regulated genes in *Rap1*-deficient cells, supporting a role of RAP1 in transcriptional gene regulation. RAP1 also binds to intergenic regions that contain many tandem telomeric repeats. The relevance of RAP1 binding to these internal telomeric repeats is unclear, although in analogy to telomeres, RAP1 may prevent recombination and fragility at these regions.

In summary, we describe both telomeric and non-telomeric roles for mammalian RAP1. Although non-telomeric roles have previously shown also for TERT and other shelterins ^{25, 44-48}, our results provide first evidence of a mammalian shelterin component influencing gene transcription through binding to non-telomeric sites. We further show that TRF2 binds to RAP1 peaks containing many telomere repeats but not to those with degenerated repeats, in analogy to RAP1 recruitment to telomeres by TRF2 ³. Future studies warrant identification of factors responsible of RAP1 recruitment to non TTAGGG-rich chromatin regions.

Methods

Construction of the *Rap1* targeting vector

The strategy for disrupting the *Rap1* locus was designed to conditionally delete E3 through a Cre-mediated excision. The targeting vector contains homology regions isogenic with the ES cell line used (129Sv/Pas). The short homology region (SA) harbors a 2.4 kb DNA fragment encompassing E2 and intron 2 (Fig. 1A). The long homology region (LA) is a 5.6 kb DNA fragment downstream the end of E3 (Fig. 1A). The central part contains E3 flanked by two *loxP* sites and a positive selection neomycin gene (PGK-Neo) flanked by two *Frt* sites (Fig. 1A). At the 3'-end of the LA a Diphtheria Toxin (DTA) selection marker was cloned (Fig. 1A). The vector contains a unique *NruI* linearization site. The targeting vector was quality controlled by sequencing of the coding exons, the junctions between the homology arms and the selection cassettes. Sequencing showed no polymorphisms between the C57BL/6 and 129Sv/Pas genetic backgrounds within the isolated *Rap1* sequences. The targeting vector was generated by genOway (www.genoway.com; Lyon, France).

Generation of conditional *Rap1* knockout mice

129Sv/Pas ES cells were transfected with 40 µg of linearized targeting vector. Positive selection was performed by adding 200 µg/mL G418. Approximately 230 positive resistant clones were isolated and PCR screened for homologous recombination first at the 5' end of the *Rap1* locus. Six positive 5' targeted ES cells were further investigated by PCR amplification over the 3' long homology arm to amplify the region of the targeted locus containing the distal *loxP* site. Direct sequencing of the PCR products amplified revealed that 3 of out of 6 clones contained the distal *loxP* site. They were verified by Southern blot analysis of AflIII and PciI restricted genomic DNA using a 5'-internal and a 3'-external probes, respectively (data not shown). Chimeric mice were generated by microinjection of the 3 independently targeted ES clones into C57BL/6J host blastocyst. The resulting offspring showed a high level of chimerism as shown by coat color, and were mated to C57BL/6J mice to assess germ line transmission. The resulting heterozygous *Rap1^{+/flox}-Neo* mice were then bred to transgenic mice expressing the Flpe recombinase¹⁷ to induce excision of the Neo marker. The *Rap1^{+/flox}* heterozygous mice were then intercrossed to generate *Rap1^{flox/flox}*, and *Rap1^{+/flox}* mice. Homozygous *Rap1^{flox/flox}* mice were crossed with transgenic mice expressing the Cre recombinase under the control of the keratin 5 promoter²⁶ (Fig. 4A). Heterozygous *Rap1^{+/flox} K5-Cre* were crossed either to *Rap1^{flox/flox}* or *Rap1^{+/flox}* to generate *Rap1^{-/-} K5-Cre*. The removal of exon 3 by Cre-mediated recombination was confirmed by PCR analysis using primers F and R (Fig. 1A,C,D and Fig. 4A). Amplification of the wild type, flox and knock-out alleles renders a 3.2 kb, 3.3 kb and 0.5 kb fragments, respectively.

The breeding to the F1 generation and characterization of heterozygous *Rap1^{+/flox}-Neo* F1 animals was performed by geneOway (www.genoway.com; Lyon, France). All mice were generated and maintained at the Spanish National Cancer Centre (CNIO) under specific pathogen-free conditions in accordance with the recommendation of the Federation of European Laboratory Animal Science Associations.

Animal experimentation

Mice were fed either with standard chow diet (Harlan Teklad LM-485), or with a high fat diet (Research Diets 12451, 45% kJ from fat) starting at 7 weeks of age. Food intake was monitored by weighting the consumed food every three days in individually caged animals.

ChIP-sequencing analyses and identification of a putative RAP1-binding consensus sequence

Biological duplicates of ChIP samples (see above) were independently processed into sequencing libraries with a ChIP-Seq sample prep kit (Illumina) by following manufacturer instructions with some minor modification⁴⁹. Libraries were prepared from 85-135bp DNA fractions (excluding adaptor length). Input samples from both *Rap1*-null and wild type MEFs were pooled and sequenced as a single library. Libraries were sequenced in an Illumina Genome Analyzer IIx (GA2) single 36-base read run. Primary data was obtained by Pipeline 1.4 analysis (PL1.4, Illumina). Raw sequences were defined as reads passing purity filter before the genome alignment.

Genome alignment was performed with PL1.4 versus the latest mouse assembly (NCBI m37/mm9, April 2007) under default settings. These settings exploit the maximum mapping specificity allowed by the aligning algorithm. PL1.4 permits alignments with more than 2 errors for 36 base reads, but with no more than 2 errors in the first 32 bases. The best alignment among alternate candidate positions is eventually chosen based on quality scores.

The experimental settings, sequences and analysis protocols of the ChIP-seq experiment have been deposited in GEO under the accession number GE20867.

Only the reads having a unique alignment in the reference genome were used for the peak detection which was performed using CisGenome v1.2⁵⁰. Briefly, uniquely aligned 36bp-length reads obtained in two independent GA2 runs were pooled into 3 datasets, corresponding to the Rap1 wild-type (WT) MEFs, Rap1-null MEFs (KO), and IP mockup. The software pipeline to analyze two-sample ChIP-seq experimental designs was applied WT reads as sample set and KO reads as negative control set using a 100bp-sliding window and a 10% FDR. To look for known genes in the neighborhood of the RAP1 binding sites, CisGenome searching method was executed against a database of mouse annotations, using a symmetrical window of 10kbp surrounding the TSS. The gene-annotated sites were cross-related with the gene labels in the microarray expression experiment by merging their expression profiles using BioConductor. Those merged genes with average $|\log_{2}(\text{FC}_{\text{WTvsKO}})| \geq 1$ were considered as deregulated in the peak association experiments.

The dataset of ChIP-seq peaks was strongly filtered ($\text{minFDR} = 0.0001$, $\text{max}|\text{FC}| \leq 6$) to select the binding sites that were clearly represented. To derive any putative consensus motifs, the resulting 30 genomic sequences corresponding to the filtered peaks were processed with the de-novo motif discovery tool Weeder 1.3³⁴, as described in^{51, 52}. The motifs identified by the algorithm as highest ranking were selected, and the STAMP online tool⁵³ was used to represent the sequence logos⁵⁴ of the consensus sequences.

The pattern matching algorithms *fuzznuc*⁵⁵ and *oligoMatch*⁵⁶ were used to scan ChIP-seq binding sequences and mouse genomic sequence for permutations of telomeric repeats. To derive any putative non-TTAGGG consensus motifs, we discarded the peaks with 3 or more occurrences of the telomeric repeats. We used STAMP⁵³ to determine the best match between the obtained motifs and known JASPAR and TRANSFAC v11.3 matrices.

Validation of ChIP-seq results

The Chip-seq results were validated by q-PCR on pulled down DNA from an independently performed chip with a rabbit polyclonal anti-RAP1 (a gift from Dr. West, CRUK, UK). Oligos were designed to amplify the DNA fragment containing the peaks corresponding to *Peak Rank 1* (chr2: 57,482,074-57,482,124), *Peak Rank 4* (chr2: 28,040,118-28,040,149), *Peak Rank 9* (chr3: 8,246,273-8,246,509), *Peak Rank 14* (chr17: 53,291,729-53,291,957), *Peak Rank 27* (chr17: 22,362,598-22,362,692), *Peak Rank 41* (chr11: 3,091,929-3,092,050), *Peak Rank 260* (chr17: 33,920,665-33,920,732), *Peak Rank 561* (chr1: 40,496,711-40,496,796), *Peak Rank 707* (chr13: 16,114,008-16,114,120), *Peak Rank 764* (chr1: 12,701,710-12,701,939), *Peak Rank 841* (chr19: 38,200,547-38,200,616), *Peak Rank 2675* (chr10: 24,308,726-24,308,800), *Peak Rank 8218* (chr11: 28,766,736-28,766,797), *Peak Rank 29521* (chr9: 54,613,901-54,614,084). Genomic regions containing telomeric repeats that did not render any hit in Chip-seq analysis were chosen as negative controls; negative control 1 (chr1:121,265,202-121,265,386) and negative control 2 (chr15:54,909,216-54,909,375). The primer sequences are available upon request. The amplification levels of the above-mentioned fragments were analyzed in the input DNA in each case for normalization and in the pulled down DNA. The relative level of each fragment was determined by calculating ΔCt values between the levels obtained in input DNA and that of the pulled down DNA. The results were normalized to wild type levels.

TRF2 binding to *Peak Rank 1*, *Peak Rank 14* and to *Peak Rank 29521* corresponding to *Crabp1* promoter was performed as described above using pulled down DNA with polyclonal TRF2 antibody (a gift from Dr. West, CRUK, UK). The results were normalized to input levels.

Luciferase assay

DNA fragments (~300 bp) harbouring Peak Rank 2675 (chr10: 24,308,726-24,308,800), Peak Rank 19335 (chr11: 74,972,250-74,972,307) and Peak Rank 260 (chr17: 33,920,665-33,920,732) located at *CTGF*, *HIC1* and *ANGPTL4* promoter region, respectively, were PCR amplified from mouse genomic DNA. Forward and reverse primers contained a XhoI and a HindIII and their sequence are available upon request. The fragments were cloned into XhoI-HindIII sites of pGL4.28 vector containing a minimal promoter driving *luciferase* (Promega) and sequence verified. LT-sv40 immortalized *Rap1*^{+/+} and *Rap1*^{-/-} MEF were transfected with the reporter constructs by using Eugene (Roche). A plasmid (pGL4.75, promega) containing a CMV promoter driving *Renilla luciferase* was cotransfected as an internal control. Cells were harvested 48 hours after transfection, and the luciferase activities of cell lysates were measured by using Dual-luciferase Reporter Assay System (Promega).

Statistical analysis

A t-student test was used to calculate the statistical significance of the observed differences in the percentage of RAP1 positive cells, H2AX foci, 53BP1 foci and chromosomal aberrations. Wilcoxon-Mann-Whitney rank sum test was used to calculate statistical significance of the observed differences in the mean telomere length. A two-tailed Wilcoxon-rank test was used to calculate the statistical significance of the differences in gene expression levels amongst subtelomeric and non-subtelomeric genes. To determine the gene expression levels trend in subtelomeric regions between classes under comparison we applied a one-tailed Wilcoxon-rank test. P-values<0.05 were considered statistically significant. Fisher Exact test was used to estimate the differences in the presence of repetitive sequences in the raw reads. To test whether is possible to find a binding peak located in or near the genic regions just by random chance we performed an exact binomial test. Pearson's Chi-square test was used to check the dependency between the presence of RAP1 binding peaks and expression deregulation of their associated genes. Freely distributed R software was employed for these calculations (<http://www.r-project.org/>).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are indebted to L. Harrington for *Tert*-deficient MEFs and S. West for RAP1 antibodies, and R. Serrano for animal care. P.M. is funded by a "Ramón y Cajal" from the Spanish Ministry of Innovation and Science. M. A. Blasco's laboratory is funded by the Spanish Ministry of Innovation and Science, the European Union (FP7-Genica, Telomarker), the European Research Council (ERC Advance Grants), the Spanish Association Against Cancer (AECC), and the Körber European Science Award to M.A. Blasco. The work in M. Tarsounas Labroatory is funded by Cancer Research UK.

References

1. de Lange T. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev.* 2005; 19:2100–2110. [PubMed: 16166375]
2. Shore D, Nasmyth K. Purification and cloning of a DNA binding protein from yeast that binds to both silencer and activator elements. *Cell.* 1987; 51:721–732. [PubMed: 3315231]
3. Li B, Oestreich S, de Lange T. Identification of human Rap1: implications for telomere evolution. *Cell.* 2000; 101:471–483. [PubMed: 10850490]
4. 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. [PubMed: 8319907]

5. Marcand S, Gilson E, Shore D. A protein-counting mechanism for telomere length regulation in yeast. *Science*. 1997; 275:986–990. [PubMed: 9020083]
6. Hecht A, Laroche T, Strahl-Bolsinger S, Gasser SM, Grunstein M. Histone H3 and H4 N-termini interact with SIR3 and SIR4 proteins: a molecular model for the formation of heterochromatin in yeast. *Cell*. 1995; 80:583–592. [PubMed: 7867066]
7. Tanny JC, Dowd GJ, Huang J, Hilz H, Moazed D. An enzymatic activity in the yeast Sir2 protein that is essential for gene silencing. *Cell*. 1999; 99:735–745. [PubMed: 10619427]
8. Carmen AA, Milne L, Grunstein M. Acetylation of the yeast histone H4 N terminus regulates its binding to heterochromatin protein SIR3. *J Biol Chem*. 2002; 277:4778–4781. [PubMed: 11714726]
9. Imai S, Armstrong CM, Kaerberlein M, Guarente L. Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase. *Nature*. 2000; 403:795–800. [PubMed: 10693811]
10. Conrad MN, Wright JH, Wolf AJ, Zakian VA. RAP1 protein interacts with yeast telomeres in vivo: overproduction alters telomere structure and decreases chromosome stability. *Cell*. 1990; 63:739–750. [PubMed: 2225074]
11. Buchman AR, Kimmerly WJ, Rine J, Kornberg RD. Two DNA-binding factors recognize specific sequences at silencers, upstream activating sequences, autonomously replicating sequences, and telomeres in *Saccharomyces cerevisiae*. *Mol Cell Biol*. 1988; 8:210–225. [PubMed: 3275867]
12. Capieaux E, Vignais ML, Sentenac A, Goffeau A. The yeast H⁺-ATPase gene is controlled by the promoter binding factor TUF. *J Biol Chem*. 1989; 264:7437–7446. [PubMed: 2523395]
13. 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. [PubMed: 20339076]
14. Martinez P, et al. Increased telomere fragility and fusions resulting from TRF1 deficiency lead to degenerative pathologies and increased cancer in mice. *Genes Dev*. 2009; 23:2060–2075. [PubMed: 19679647]
15. Tejera A, et al. TPP1 is required for TERT recruitment, telomere elongation during nuclear reprogramming, and normal skin development in mice. *Developmental Cell*. 2010; 18:775–789. [PubMed: 20493811]
16. 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. [PubMed: 19345190]
17. Rodriguez CI, et al. High-efficiency deleter mice show that FLPe is an alternative to Cre-loxP. *Nat Genet*. 2000; 25:139–140. [PubMed: 10835623]
18. Tan M, Wei C, Price CM. The telomeric protein Rap1 is conserved in vertebrates and is expressed from a bidirectional promoter positioned between the Rap1 and KARS genes. *Gene*. 2003; 323:1–10. [PubMed: 14659874]
19. d'Adda di Fagagna F, et al. A DNA damage checkpoint response in telomere-initiated senescence. *Nature*. 2003; 426:194–198. [PubMed: 14608368]
20. Takai H, Smogorzewska A, de Lange T. DNA damage foci at dysfunctional telomeres. *Curr Biol*. 2003; 13:1549–1556. [PubMed: 12956959]
21. Silver DP, Livingston DM. Self-excising retroviral vectors encoding the Cre recombinase overcome Cre-mediated cellular toxicity. *Mol Cell*. 2001; 8:233–243. [PubMed: 11511376]
22. Blanco R, Munoz P, Flores JM, Klatt P, Blasco MA. Telomerase abrogation dramatically accelerates TRF2-induced epithelial carcinogenesis. *Genes Dev*. 2007; 21:206–220. [PubMed: 17234886]
23. Munoz P, Blanco R, Flores JM, Blasco MA. XPF nuclease-dependent telomere loss and increased DNA damage in mice overexpressing TRF2 result in premature aging and cancer. *Nat Genet*. 2005; 37:1063–1071. [PubMed: 16142233]
24. Sfeir A, et al. Mammalian telomeres resemble fragile sites and require TRF1 for efficient replication. *Cell*. 2009; 138:90–103. [PubMed: 19596237]
25. Park JI, et al. Telomerase modulates Wnt signalling by association with target gene chromatin. *Nature*. 2009; 460:66–72. [PubMed: 19571879]

26. Tarutani M, et al. Tissue-specific knockout of the mouse Pig-a gene reveals important roles for GPI-anchored proteins in skin development. *Proc Natl Acad Sci U S A*. 1997; 94:7400–7405. [PubMed: 9207103]
27. Ramirez A, Bravo A, Jorcano JL, Vidal M. Sequences 5' of the bovine keratin 5 gene direct tissue- and cell-type-specific expression of a lacZ gene in the adult and during development. *Differentiation*. 1994; 58:53–64. [PubMed: 7532601]
28. Benetti R, Garcia-Cao M, Blasco MA. Telomere length regulates the epigenetic status of mammalian telomeres and subtelomeres. *Nat Genet*. 2007; 39:243–250. [PubMed: 17237781]
29. Azzalin CM, Reichenbach P, Khoraiuli L, Giulotto E, Lingner J. Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science*. 2007; 318:798–801. [PubMed: 17916692]
30. Schoeftner S, Blasco MA. Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. *Nat Cell Biol*. 2008; 10:228–236. [PubMed: 18157120]
31. Estep PW 3rd, Warner JB, Bulyk ML. Short-term calorie restriction in male mice feminizes gene expression and alters key regulators of conserved aging regulatory pathways. *PLoS One*. 2009; 4:e5242. [PubMed: 19370158]
32. Crowley VE, Yeo GS, O'Rahilly S. Obesity therapy: altering the energy intake-and-expenditure balance sheet. *Nat Rev Drug Discov*. 2002; 1:276–286. [PubMed: 12120279]
33. Leone TC, et al. PGC-1 α deficiency causes multi-system energy metabolic derangements: muscle dysfunction, abnormal weight control and hepatic steatosis. *PLoS Biol*. 2005; 3:e101. [PubMed: 15760270]
34. Pavesi G, Pesole G. Using Weeder for the discovery of conserved transcription factor binding sites. *Curr Protoc Bioinformatics*. 2006; 11. Chapter 2, Unit 2. [PubMed: 18428764]
35. Bae NS, Baumann P. A RAP1/TRF2 complex inhibits nonhomologous end-joining at human telomeric DNA ends. *Mol Cell*. 2007; 26:323–334. [PubMed: 17499040]
36. Sarthy J, Bae NS, Scrafford J, Baumann P. Human RAP1 inhibits non homologous end joining at telomeres. *EMBO J*. 2009; 28:3390–3399. [PubMed: 19763083]
37. Armanios MY, et al. Telomerase mutations in families with idiopathic pulmonary fibrosis. *N Engl J Med*. 2007; 356:1317–1326. [PubMed: 17392301]
38. Mitchell JR, Wood E, Collins K. A telomerase component is defective in the human disease dyskeratosis congenita. *Nature*. 1999; 402:551–555. [PubMed: 10591218]
39. Tsakiri KD, et al. Adult-onset pulmonary fibrosis caused by mutations in telomerase. *Proc Natl Acad Sci U S A*. 2007; 104:7552–7557. [PubMed: 17460043]
40. Vulliamy T, et al. The RNA component of telomerase is mutated in autosomal dominant dyskeratosis congenita. *Nature*. 2001; 413:432–435. [PubMed: 11574891]
41. Yamaguchi H, et al. Mutations in TERT, the gene for telomerase reverse transcriptase, in aplastic anemia. *N Engl J Med*. 2005; 352:1413–1424. [PubMed: 15814878]
42. Moretti P, Freeman K, Coodly L, Shore D. Evidence that a complex of SIR proteins interacts with the silencer and telomere-binding protein RAP1. *Genes Dev*. 1994; 8:2257–2269. [PubMed: 7958893]
43. Schoeftner S, et al. Telomere shortening relaxes X chromosome inactivation and forces global transcriptome alterations. *Proc Natl Acad Sci U S A*. 2009; 106:19393–19398. [PubMed: 19887628]
44. 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. [PubMed: 15665826]
45. Deng Z, Atanasiu C, Burg JS, Broccoli D, Lieberman PM. Telomere repeat binding factors TRF1, TRF2, and hRAP1 modulate replication of Epstein-Barr virus OriP. *J Virol*. 2003; 77:11992–12001. [PubMed: 14581536]
46. 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. [PubMed: 11938440]
47. Smogorzewska A, et al. Control of human telomere length by TRF1 and TRF2. *Mol Cell Biol*. 2000; 20:1659–1668. [PubMed: 10669743]

48. Zhang P, et al. Nontelomeric TRF2-REST interaction modulates neuronal gene silencing and fate of tumor and stem cells. *Curr Biol.* 2008; 18:1489–1494. [PubMed: 18818083]
49. Quail MA, et al. A large genome center's improvements to the Illumina sequencing system. *Nat Methods.* 2008; 5:1005–1010. [PubMed: 19034268]
50. Ji H, et al. An integrated software system for analyzing ChIP-chip and ChIP-seq data. *Nat Biotechnol.* 2008; 26:1293–1300. [PubMed: 18978777]
51. Chen CC, et al. Acetylated lysine 56 on histone H3 drives chromatin assembly after repair and signals for the completion of repair. *Cell.* 2008; 134:231–243. [PubMed: 18662539]
52. Long JD, et al. Evidence Review of Technology and Dietary Assessment. *Worldviews Evid Based Nurs.* 2009
53. Mahony S, Benos PV. STAMP: a web tool for exploring DNA-binding motif similarities. *Nucleic Acids Res.* 2007; 35:W253–258. [PubMed: 17478497]
54. Schneider TD, Stephens RM. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* 1990; 18:6097–6100. [PubMed: 2172928]
55. Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* 2000; 16:276–277. [PubMed: 10827456]
56. Kent WJ, et al. The human genome browser at UCSC. *Genome Res.* 2002; 12:996–1006. [PubMed: 12045153]

Figure 1. Generation of *Rap1*^{-/-} cells

a Schematic outline of wild-type (*Rap1*⁺), floxed (*Rap1*^{lox-neo}), and deleted (*Rap1*^Δ) loci. The *Kars* gene is indicated in blue. Transgenic expression of Flp and Cre recombinases excise the Neo and the *Rap1*-E3 cassettes, respectively, giving rise to the *Rap1* null allele. The forward (F) and reverse (R) primers used for PCR-genotyping are indicated with red arrows. **b** Schematic outline of RAP1 protein domains. BRCT, BRCA1 C terminus; Myb-like, Myb-related HTH motif; Coil, predicted coil domain; RCT, homology to the RAP1 C terminus; NLS, nuclear localization signal. The amino acid positions are indicated. Primers used for q-PCR are depicted. **c** Deletion of *Rap1* was induced in wildtype and *Rap1*^{lox/lox} MEF by retroviral infection with the Cre recombinase. Cells were simultaneously infected with either shp53 or SV40-LT. As control, cells were infected with empty pBabe. *Rap1* deletion was confirmed by PCR amplification of *Rap1* alleles using the F and R primers (a) (Fig. S8A for uncropped image of the gel). **d** Quantitative qRT-PCR of *Rap1* and *Kars* mRNA levels. Three different primers pairs were used to quantify transcripts levels corresponding to *Rap1* E1-E2 (F1-R1, blue arrows), *Rap1* E3 (F2-R2, green arrows), and *Kars*. mRNA expression levels were normalized to wild-type. Error bars indicate standard deviation. The right panel shows a *Kars* Northern blot and the expression level quantification is indicated below. **e** Expression of RAP1 protein in cytoplasmic (C) and nuclear (N) fractions of *Rap1*^{+/+} and *Rap1*^{Δ/Δ} cells by Western blot. SA1 and tubulin were used as nuclear and cytoplasmic markers, respectively. Note that a faint band of approximately 40 kD is detected in the cytoplasmic extracts of *Rap1*^{Δ/Δ} cells (*), but is not

present in the nuclear fractions. Molecular weight ladder is shown. (Fig. S8B for uncropped image of the blots). *f.* RAP1 protein was immunoprecipitated from total cellular extracts and immunoblotted as shown. PI, preimmune. Molecular weight ladder is shown. Fig. S8C for uncropped image of the blots. *g.* Representative RAP1 immunofluorescence showing no detectable signal in *Rap1*-null MEFs. The scale bars represent 10 μ m.

Figure 2. Rap1 deletion does not affect the binding of the other shelterin components to telomeres

a. Subcellular fractionation of the indicated MEFs. Note absence of full-length RAP1 in Rap1-null MEFs. A putative truncated for of RAP1 (*) is restricted to cytoplasmic extracts. Tubulin is used as loading control for the cytoplasmic fraction and histone H3 for the chromatin-bound fraction (Fig. S8E for uncropped image of the blots). **b.** Quantification of protein levels relative to H3. **c.** Chromatin immunoprecipitation (ChIP) experiments using the indicated antibodies. **d.** Quantification of ChIP values for telomere repeats after normalization by the input. n=MEFs used. Bars, average between replicates; standard error is shown. A Student's t-test was used to calculate statistical significance. **e.** Representative images showing normal telomeric localization of the indicated shelterin components in *Rap1*-null MEFs. Telomeric foci were detected by co-staining with a mouse anti-TRF1 antibody in combination with rabbit polyclonal antibodies against the different shelterins. The scale bars represent 5 μ m.

G

H

I

Figure 3. *Rap1*-deleted MEFs show increased telomere damage and fragility but no fusions
a,b Western blot and immunofluorescence of RAP1 protein in the indicated MEFs (Fig.S8F for uncropped images). Nuclei were stained with anti-RAP1 antibody (red) and DAPI (blue). Scale bars represent 10 μ m. **c.** H2AX immunofluorescence (green) combined with telomere FISH (red). Scale bars represent 10 μ m. **d.** Percentage of metaphases with <3 or >3 gH2AX foci from a total of 200 metaphases per genotype prepared as in c. **e.** Western blot of phospho-CHK1 and CHK2 in the indicated MEFs (Fig. S8G for uncropped images). Wild-type MEFs treated with 0 or 10 Gy of ionizing radiation are controls for checkpoint activation. Tubulin is a loading control. **f.** Fold increase in p-CHK1 and p-CHK2 levels in *Rap1*-deleted cells compared to controls. Comparison with *TRF1*-null MEFs illustrates a milder DDR induced by *Rap1* abrogation. **g.** Rescue of 53BP1 DNA damage foci in *Rap1*-depleted MEFs treated with 10 μ M KU55933 (ATM inhibitor) or 5 μ M caffeine (ATM-ATR inhibitor). Mean values relative to wild-type $\pm$ standard error are shown. Statistical significance was determined by Student's t-test. **h.** Percentage of damaged cells (53BP1 positive) with >3 telomere induced foci (TIFs) in non-treated and treated cells as in f. **i.** Representative 53BP1 positive cells without or with TIFs (white arrow head). Scale bars represent 5 μ m. **j.** Frequency of chromosome aberrations. Cells were treated with aphidicolin when indicated. Magnifications of the indicated aberrations are shown below the bar graphs. Scale bars represent 1 μ m. Two MEFs were used per genotype and condition and total number of cells scored is indicated. Bars, standard error. Statistical comparisons are done

using the Student's t-test and p-values are indicated. **k.** Percentage of T-SCE events. (n) chromosomes used for the analysis from 2 MEFs per genotype. Bars, standard error. The Chi-square test was used for statistical analysis, p-value is indicated. **l.** Representative CO-FISH images showing the leading (green) and lagging (red) telomere strands. T-SCE events are indicated with arrows. A T-SCE was considered positive when involves a reciprocal exchange of telomere signal. Scale bars represent 1 μm .

Figure 4. Deletion of *Rap1* in stratified epithelia leads to telomere shortening and skin hyperpigmentation at adulthood

a. PCR amplification of *Rap1* alleles in dermis (D) and in epidermis (E) of the indicated genotypes using the F and R primers depicted in Fig. 1A (Fig. S8D for uncropped image of the gel). **b.** Representative images of tail skin sections showing no detectable RAP1 immunofluorescence in *Rap1*^{+/+} K5-Cre epidermis. The scale bars represent 10 μm. **c.** Quantification of RAP1-positive cells in the epidermis of wild-type and *Rap1*^{+/+} K5-Cre mice. Bars, standard error. Student's t-test was used for statistical analysis and p-values are indicated. **d.** *Rap1*^{+/+} K5-Cre mice reach adulthood and show no defects in hair growth or skin morphogenesis. Note that the *Rap1* mutant females are bigger than the wild-type controls. **e.** Representative images of tails in *Rap1*^{+/+} K5-Cre and wild type littermates at 1 and 10 weeks after birth. Note hyperpigmented skin in *Rap1*^{+/+} K5-Cre 10-week old mice but not in the 1-week olds. **f,g.** Telomere (f) and centromere (g) fluorescence distribution of individual telomere and centromere (major satellite repeats) dots in tail skin sections of mice of the indicated genotype as determined by Q-FISH analysis. Note that average telomere length decreases by 26% in *Rap1*^{+/+} K5-Cre mice. Mean fluorescence and standard error values are shown. The total number of mice and telomeres analyzed per genotype is indicated. The Wilcoxon-Mann-Whitney rank sum test was used for statistical comparisons and p-values are indicated. **h.** Percentage of γH2AX positive cells (> 3 foci). Total number of mice and cells scored per genotype is indicated. Bars, standard error. Student's t-test was used for statistical analysis and p-values are indicated.

A

B

Subtelomeric genes enriched in *Rap1^{4/4}*

C

Subtelomeric genes enriched in *Rap1^{+/-}*

D

Figure 5. Role for RAP1 in subtelomeric silencing

a. Frequency distribution of Log_2 fold change ($Rap1^-/-/Rap1^{+/+}$) of genes mapped in subtelomeric regions (S) and non-subtelomeric regions (NS) (orange and blue, respectively). n =number of genes in each set. The median of both distributions is shown. A non-parametric two-tailed Wilcoxon-rank test shows that gene expression levels in S and NS regions are statistically different; p -value is shown. Bottom part, S and NS distributions showing that the medians of logFC are biased towards higher expression in $Rap1^-/-$ cells, being this difference higher in S than in NS-regions (0.04 vs 0.01). A one-tailed Wilcoxon-rank test further indicates that S regions are significantly over-expressed in $Rap1^-/-$ cells. **b,c.** GSEA for genes located 3 Mb from the telomeres at the indicated chromosomes. Microarray genes are ranked based on the two-tailed t -statistic obtained from the $Rap1^-/-$ versus $Rap1^{+/+}$ pairwise. The red-blue bar represents the ranked list (red= $Rap1^-/-$, blue= $Rap1^{+/+}$). Gene showing higher expression levels for each genotype are located in the edges of the bar. The black lines represent the subtelomeric genes (hits) location in the whole ranked list. The green line and the grey distribution represent the enrichment score and the ranking metric score, respectively. The distribution of these genes was found biased to $Rap1^-/-$ cells in chromosome 1, 2, 11, 15 and 19 (b), while in chromosome 7 was biased to $Rap1^{+/+}$ cells (c). FDR values are indicated. The genes are listed to the right of the plots. **d.** 3Mb of subtelomeric regions in the indicated chromosomes. Red bars, genes upregulated in $Rap1^-/-$ cells; Green bars, genes downregulated. The bar position reflects the gene location on the chromosome. The height of the bars is proportional to the log_2 ratio of expression between $Rap1^-/-/Rap1^{+/+}$.

Figure 6. Differentially expressed genes upon *Rap1* abrogation

a. Signal $\log_2$ expression of downregulated (blue) and upregulated (red) genes in *Rap1*-deleted compared to wild-type MEFs. The colour intensity is proportional to the signal $\log_2$ intensity. Biological processes are indicated to the right. All genes had an FDR < 0.15 except for *Igf2* (FDR=0.3). Values for 3 independent MEFs per genotype are shown. **b.** q-PCR validation of the DEGs. Fold change with respect to wild-type levels is represented. Two independent MEFs were analyzed and the analysis was performed in triplicate. Bars, standard deviation. **c.** GSEA for cancer-related genes. Red color indicates gene-expression changes validated by qPCR. **d.** Comparison of DEGs induced by CR and *Rap1* deletion. Left column, $\log_2$ fold change between high calorie (“HighC”) and low calorie (“CR”) diets. Red, genes upregulated under CR. Right column, $\log_2$ fold change between *Rap1*^{-/-}-LT-Cre and *Rap1*^{+/+}-LT-Cre MEFs. In all GSEA, microarray genes were ranked based on the two-tailed t-statistic obtained from the *Rap1*^{-/-} versus *Rap1*^{+/+} pair-wise comparison. DEGs are listed to the right. Red color indicates gene-expression changes validated by qPCR. **e.** qPCR shows decreased PGC1- α expression in *Rap1*-null cells relative to wild-type cells. Two independent MEFs were analyzed and the analysis was performed in triplicate with two different primer pairs.

D

E

F

G

H

Figure 7. Genome-wide mapping of RAP1 binding sites

a. Percentage of 36bp raw ChIP-seq reads (before genome alignment) containing perfect TTAGGG₅/CCCTAA₅ repeats in the indicated conditions. These reads represent RAP1 binding to telomeric sequences. **b.** Percentage of RAP1 peaks after genome alignment (excluding telomeric regions) with at least one permutation of 2 consecutive telomeric repeats compared to the same permutations along the mouse genome, forcing perfect matches. Data correspond to the median of the 6 possible permutations of the TTAGGG sequence. **c.** Genome-wide distribution of 30,398 RAP1 binding sites in the mouse genome. Each bar shows the number of sites in a 1 Mbp window. Colored part of the bar represents

the sites that are in the neighborhood of a gene. **d.** Average density of RAP1 binding sites (in peaks per Mbp) in the subtelomeric (3Mbp region adjacent to telomeres), non-subtelomeric (rest of the genome) and whole genome fractions. **e.** Genomic density of RAP1 binding sites expressed as occupancy per percentage of chromosome length, and ordered by ascending distance to the telomere. When defining the subtelomere as a 3Mbp region downstream the telomere, this represents from 1.5% (chromosome 1) to 4.9% (chromosome 19) of chromosome length. RAP1 sites are enriched at the subtelomeric regions common to all chromosomes (red dots) and their density decreases as the distance to the telomere increases. **f.** Distribution of RAP1 sites according to their vicinity to known genes. A mock gene is shown indicating the regions in the distribution. **g.** Overlapping between the ChIP-seq and the expression array experiments. Up to 6,988 RAP1 binding sites are correlated with genes in the expression array, including 234 sites associated with 173 distinct differentially expressed genes (2x fold change). **h.** Examples of RAP1 sites. Significant RAP1 peaks, as detected by the peak finding method and after position correction, are represented in the WT sample in dark orange. The rest of the sequencing reads are shown in grey for the KO and in green for the WT. The corresponding Refseq gene is shown in blue. Scale represents the number of raw sequencing reads in each sample.

A

Identified consensus binding motif in extra-telomeric RAP1-binding peaks

B

Figure 8. De novo identification of a consensus RAP1 binding site

a. RAP1 binding matrix predicted by the Weeder algorithm represented as a sequence logo. **b.** Best RAP1 binding loci are saturated of TTAGGG repeats. Telomeric repeats in the best ranking 100 and 1,000 RAP1 binding peaks. When considering perfect matches (PM, red) and mismatches (MM, blue), the best peak has up to 26 repeats. **c.** Representative RAP1 binding loci indicating the identified sequence. Nucleotides underlined correspond to GGGTTA/TAACCC repeats, and those in bold indicate perfect matches with the repeat sequences. All genes shown are deregulated ($|\log_{2}FC(WT/KO)| \geq 1$) in the microarray experiment. **d.** Validation of individual RAP1-binding peaks using ChIP with anti-RAP1 antibodies followed by qPCR. Primers used are indicated by blue arrows in part c. The Peak

Rank of the regions tested and the associated genes are indicated in the X-axis. Four peaks within non-coding regions at chromosome 2, 3 and 17 were tested. As negative controls (NC), two non-coding regions at chromosome 1 and 15 containing telomeric repeats that did not show reads in Chip-seq analysis were analyzed. The results were normalized to wild-type levels. Note decreased RAP1 binding to RAP1 peaks in *Rap1*-null MEFs. *e.* Genomic fragments associated to *Ctcf*, *Hic1* and *Angptl4* show RAP1-dependent enhancer activity. The results are normalized to wild-type cells. The LogFC (*Rap1*^{-/-} / *Rap1*^{+/+}) values obtained in the gene expression experiment are also shown. *f.* Analysis of TRF2 binding to the indicated peaks by using ChIP with anti-TRF2 antibodies followed by qPCR. The results are normalized to input levels. *g.* Summary of results. RAP1 binds to both telomeric and non-telomeric chromatin, thereby exerting dual role in telomere function (control of telomere length and prevention of telomere fragility and recombination) as well as transcriptional gene regulation, including subtelomeric gene silencing. RAP1 also binds to non-coding regions enriched in TTAGGG repeats, where it may also have a role in preventing fragility.