 doi:10.20944/preprints202110.0001.v1

Article

Identification of key proteins in the Alternative Lengthening of Telomeres associated Promyelocytic Leukemia Nuclear Bodies

Isaac Armendáriz-Castillo[1,2], Katherine Hidalgo-Fernández[1], Andy Pérez-Villa[1], Jennyfer García-Cárdenas[1,3], Andrés López-Cortés[1]*, and Santiago Guerrero[1,3]*

[1] Latin American Network for the Implementation and Validation of Clinical Pharmacogenomics Guidelines (RELIVAF-CYTED), Madrid, Spain; isaac.arcas@gmail.com, katherine99.hidalgo@gmail.com, andypzvi@gmail.com
[2] Facultade de Ciencias, Universidade da Coruña, 15071, A Coruña, Spain;
[3] Escuela de Medicina, Facultad de Ciencias Médicas de la Salud y de la Vida, Universidad Internacional del Ecuador, 170113 Quito, Ecuador; jennyfergrc7@gmail.com
* Correspondence: sxguerrero@gmail.com, aalc84@gmail.com

Abstract: One of the hallmarks of the Alternative Lengthening of Telomeres (ALT) is the association with Promyelocytic Leukemia (PML) Nuclear Bodies, known as APBs. In the last years, APBs have been described as the main place where telomeric extension occurs in ALT positive cancer cell lines. A different set of proteins have been associated with APBs function, however, the molecular mechanisms behind their assembly, colocalization, and clustering of telomeres, among others, remain unclear. To improve the understanding of APBs in the ALT pathway, we integrated multi-omics analyses to evaluate genomic, transcriptomic and proteomic alterations, and functional interactions of 71 APBs-related genes/proteins in 32 PanCancer Atlas studies from The Cancer Genome Atlas Consortium (TCGA). As a result, we identified 13 key proteins which showed distinctive mutations, interactions, and functional enrichment patterns across all the cancer types and proposed this set of proteins as candidates for future *ex vivo* and *in vivo* analyses that will validate these proteins to improve the understanding of the ALT pathway, fill the current research gap about APBs function and their role in ALT, and be considered as potential therapeutic targets for the diagnosis and treatment of ALT positive cancers in the future.

Keywords: ALT, PML, Telomeres, Pan-Cancer, TCGA

1. Introduction

Telomeres are nucleoprotein complexes composed of tandem repeats of TTAGGG, whose primary function is to protect the ends of chromosomes against end-to-end fusions, chromosomal rearrangements, and genomic instability (1,2). In somatic cells, due to cell division, telomeres shorten, causing senescence or apoptosis (3). To avoid replicative senescence during tumorigenesis, telomerase reactivates in most types of cancer (3). However, 10% to 15% of cancers use a telomerase-independent mechanism to preserve their telomeres, called, Alternative Lengthening of Telomeres (2). Some pathways and molecular mechanisms of ALT are not yet understood, but it has been proposed that it may use dependent or independent mechanisms of homologous recombination (HR) (4).

A particular hallmark of ALT$^+$ cells is the formation of an interactosome with Promyelocytic Leukemia (PML) Nuclear Bodies, known as ALT-associated PML Bodies (APBs) (5,6). PML bodies are membrane-less organelles found in the cell nucleus, which contain small ubiquitin-like modification (SUMO) sites (7) and are formed by PML, Sp100 and SUMO-1/2/3 proteins. Additionally, they use more than 50 proteins such as RAD52, RAD51, RAD50, RPA, BLM, BRCA1, among others, which are involved in different cellular functions, such as tumoral suppression, DNA replication, gene transcription, DNA

 doi:10.20944/preprints202110.0001.v1

Damage Response (DDR), senescence, and apoptosis (8,9). In the course of APBs formation, all of the six subunits (TRF1, TRF2, POT1, TPP1, TIN2 and Rap1) that constitute the shelterin complex detach from the telomeric DNA and are incorporated into the APBs (SUMOylation of shelterin), creating a recombigenic microenvironment that contributes to ALT triggering (8).

Normally, the PML bodies are disassembled when cells enter mitosis, however, due to their hyper-SUMOylated state, APBs have been observed in metaphases of cancer cell lines (7). Recent studies have shown that telomere clustering in tumoral cells promotes ALT through mitotic DNA synthesis (MiDAS) (5). By applying the ATSA (ALT telomere DNA synthesis in APBs) assay, Zhang *et al.* 2019, demonstrated that telomeric DNA synthesis in ALT$^+$ cells take place exclusively in APBs while on the G2 phase of the cell cycle. In addition, the knockdown of the *PML* gene in ALT$^+$ cells has resulted in a reduction of telomeres length and decreased ALT function (10).

Despite having demonstrated that APBs are essential for the ALT pathway, many of the molecular mechanisms for their assembly and how telomeres cluster inside the PML bodies are still unknown (3,7,11). Furthermore, the molecular mechanisms behind the ALT pathway are still poorly understood (10). In a previous research, we identified a group of 20 genes/proteins that could be used as potential molecular markers for the study of ALT (12).

Under this context, the aim of this study is to evaluate the genomic, transcriptomic and proteomic alterations of 71 genes/proteins associated with APBs by using an integrated TCGA PanCancer Atlas and multi-omics analyses in order to improve the understanding of the role of APBs in cancer, their correlation with ALT and their application as potential molecular markers for the diagnosis and treatment of ALT$^+$ cancers.

2. Results

2.1. Gene set and genomic, transcriptomic and proteomic alterations

To evaluate the genomic, transcriptomic, and proteomic alterations of the 71 APBs genes/proteins (Table S1), these were analyzed in the cBioPortal (13,14) by selecting 10,918 samples from 32 studies of the PanCancer Atlas (PCA) (15–24) from The Cancer Genome Atlas consortium (TCGA) (Table 1).

Table 1. List of TCGA PanCancer Atlas studies selected

TCGA study	n
Acute Myeloid Leukemia (LAML)	165
Adrenocortical Carcinoma (ACC)	76
Bladder Urothelial Carcinoma (BLCA)	402
Brain Lower Grade Glioma (LGG)	507
Breast Invasive Carcinoma (BRCA)	994
Cervical Squamous Cell Carcinoma (CESC)	275
Cholangiocarcinoma (CHOL)	36
Colorectal Adenocarcinoma (COAD)	524
Diffuse Large B-cell Lymphoma (DLBC)	39
Esophageal Adenocarcinoma (ESCA)	181
Glioblastoma Multiforme (GBM)	145

 doi:10.20944/preprints202110.0001.v1

Head and Neck Squamous Cell Carcinoma (HNSC)	488
Kidney Chromophobe (KICH)	65
Kidney Renal Clear Cell Carcinoma (KIRC)	352
Kidney Renal Papillary Cell Carcinoma (KIRP)	274
Liver Hepatocellular Carcinoma (LIHC)	348
Lung Adenocarcinoma (LUAD)	503
Lung Squamous Cell Carcinoma (LUSC)	466
Mesothelioma (MESO)	82
Ovarian Serous Cystadenocarcinoma (OV)	201
Pancreatic Adenocarcinoma (PAAD)	168
Pheocromocytoma and Paraganlioma (PCPG)	161
Prostate Adenocarcinoma (PRAD)	488
Sarcoma (SARC)	251
Skin Cutaneous Melanoma (SKCM)	363
Stomach Adenocarcinoma (STAD)	407
Testicular Germ Cell Tumors (TGCT)	144
Thymoma (THYM)	119
Thyroid Carcinoma (THCA)	480
Uterine Carcinosarcoma (UCS)	56
Uterine Corpus Endometrial Carcinoma (UCEC)	507
Uveal Melanoma (UVM)	80

A total of 72,492 alterations were identified and a donut chart was elaborated showing the most frequent alterations after all values were normalized by the number of samples in each study (Table S2). Figure 1a, shows that the most frequent genomic alteration was mRNA high (61%) followed by mRNA low (14.5%), CNV amplifications (8.50%), missense (putative passenger) mutations (6.90%), deep deletions (2.95%), and protein high and low with 1.53% and 1.31%, respectively.

To understand the implication of APBs genes/proteins alterations in cancer progression from primary tumors to metastasis (T1 to T4), genomic, transcriptomic and proteomic alterations were subgrouped for each cancer and metastasis stage from the PCA studies, when available. All alterations were normalized by the number of samples in each stage per each study. No significant differences were found between cancer stage alterations (Figure 1b) or metastasis stage alterations (Figure 1c) after a multiple comparison with the Original False Discovery Rate (FDR) method of Benjamini and Hochberg ($p<0.001$).

2.2. TCGA Pan-Cancer studies frequencies and OncoPrint of genomic and proteomic alterations

With the values of genomic, transcriptomic, and proteomic alterations normalized by the number of samples in each study, the highest frequency means of alterations were calculated for the 32 studies and 71 genes/proteins (Table S3). UCS was the cancer type with the highest alteration frequency mean (10.456), followed by ESCA (9.555), UCEC (9.170), SKCM (8.557), BLCA (8.448), ACC (8.152), LUSC (8.016), CESC (8.003), BRCA (7.838), COAD (7.556), STAD (7.461), SARC (7.447), HNSC (7.061), LUAD (7.048), CHOL (6.833), OV (6.663), LIHC (6.653), TGCT (6.188), MESO (6.115), PAAD (6.060), DLBC (6.021), PRAD (5.846), KICH (5.846), KIRP (5.799), PCPG (5.494), LGG (5.484), KIRC (4.846), UVM (4.475), LAML (4.236), THCA (3.886), THYM (3.341) and GBM (2.157) (Figure 2a).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 October 2021 doi:10.20944/preprints202110.0001.v1

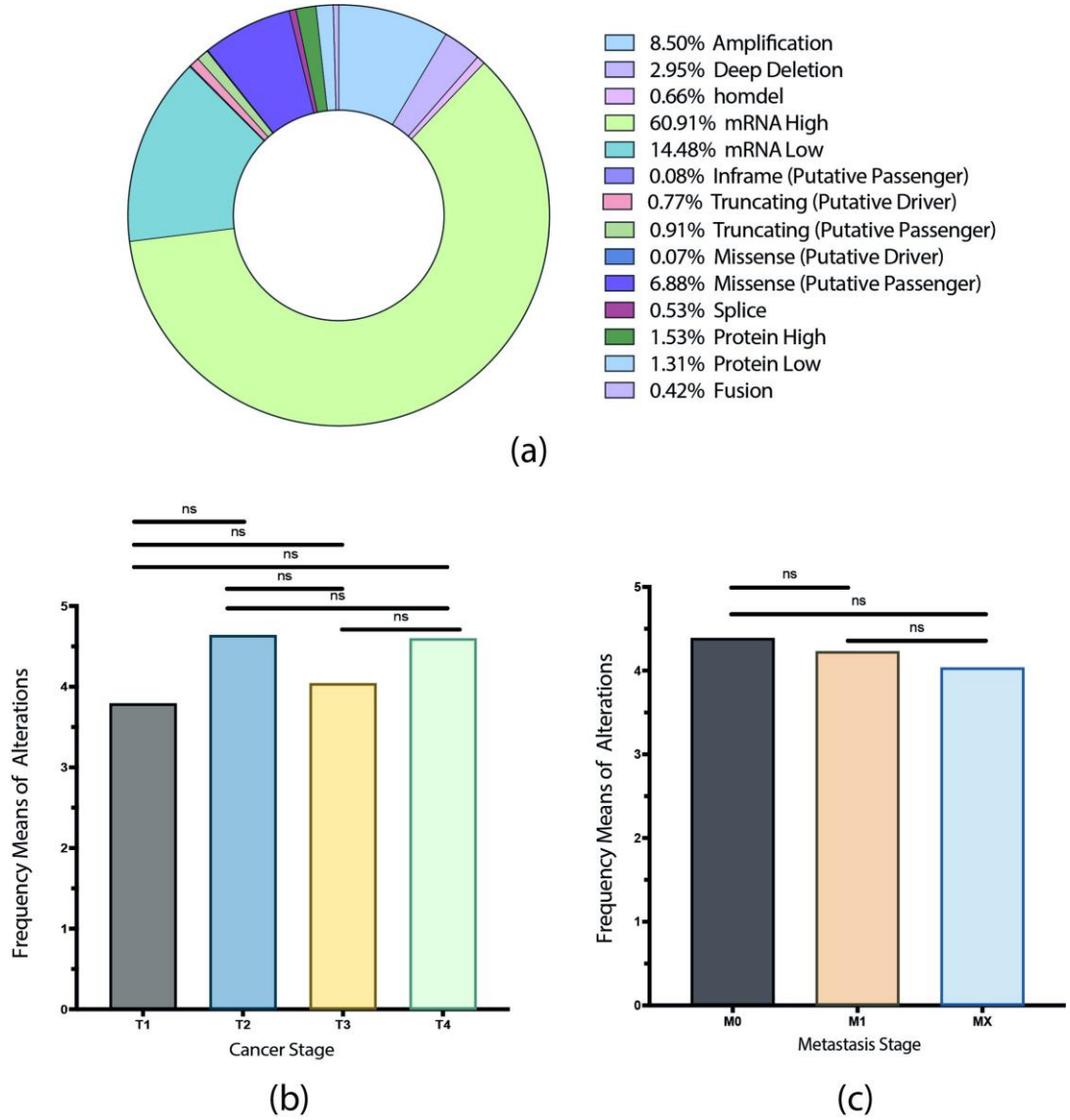


Figure 1. Genomic, transcriptomic, and proteomic alterations. (a) A donut chart with the most frequent alterations identified in the 71 APBs genes/proteins analyzed. (b) (c) Alterations observed in each cancer and metastasis stage of the PCA studies showing no significant difference among alterations in each stage

Consequently, to identify highly altered genes/proteins, the first quartile of the PCA studies with the highest means of alterations were selected to construct a boxplot, and genes and proteins that showed significantly different patterns of alterations were identified by using Tukey's test. Figure 2b shows twelve types of PCA studies in which the APBs genes/proteins present the highest number of frequency means alterations (Table S4). SENP5, TERF1, UPF1, NSMCE2, CDKN1A, SUMO2, ACD, KDM1A, ATM, CBX3, TP53B1, TEP1, NBN, ATR, PIAS4, XRCC6, MRE11, TOP3A, SBTB48, RAD52, HUS1, GNL3L genes/proteins showed significantly higher means of genomic, transcriptomic and proteomic alterations across the twelve studies, therefore, they can be considered as targets of interest for the following analyses.

Finally, an OncoPrint with the first 2 quartiles of the genes/proteins with the highest means of genomic, transcriptomic, and proteomic alterations was constructed by using

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 October 2021 doi:10.20944/preprints202110.0001.v1

the cBioPortal data (https://www.cbioportal.org/) (13,14) (Figure 3a), the most common alteration type observed was mRNA high, followed by mRNA low and CNV amplification (Table S5). In addition, genes/proteins with the highest alteration frequencies were NSMCE2, SENP5 and TERF1 with mRNA high alterations, TERF2, RAD17 and XRCC6 with mRNA low, NSMCE2, SENP5 and NBN with CNV amplification, HMBOX1, WRN and KDM4C with CNV deep deletion, ATM, TEP1 and ATR with missense mutations, ATM, STC2 and ATR with truncating mutations, MRE11, PCNA and RAD50 with protein high, TP53B1, ATM and CDKN1A with protein low and NSMCE2, SENP5 and TERF1 in the overall alterations (Figure 3b).

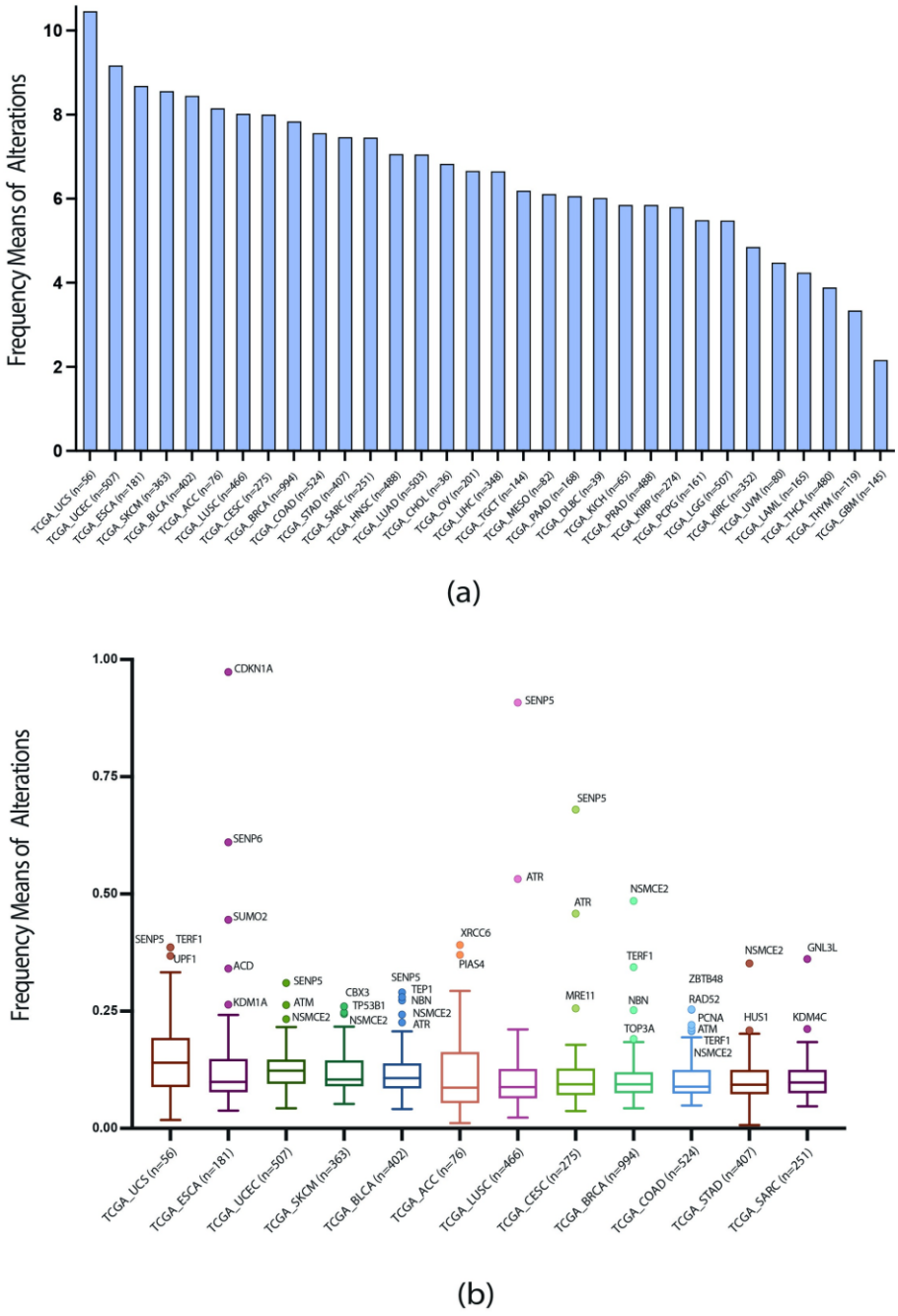


Figure 2. Frequency of genomic, transcriptomic and proteomic alterations per TCGA Pan-Cancer type. (a) Frequency means of alterations of the 71 APBs genes/proteins across the 32 PCA studies. (b) Boxplot showing genes/proteins which showed significant differences in alterations patterns in twelve PCA studies.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 October 2021 doi:10.20944/preprints202110.0001.v1

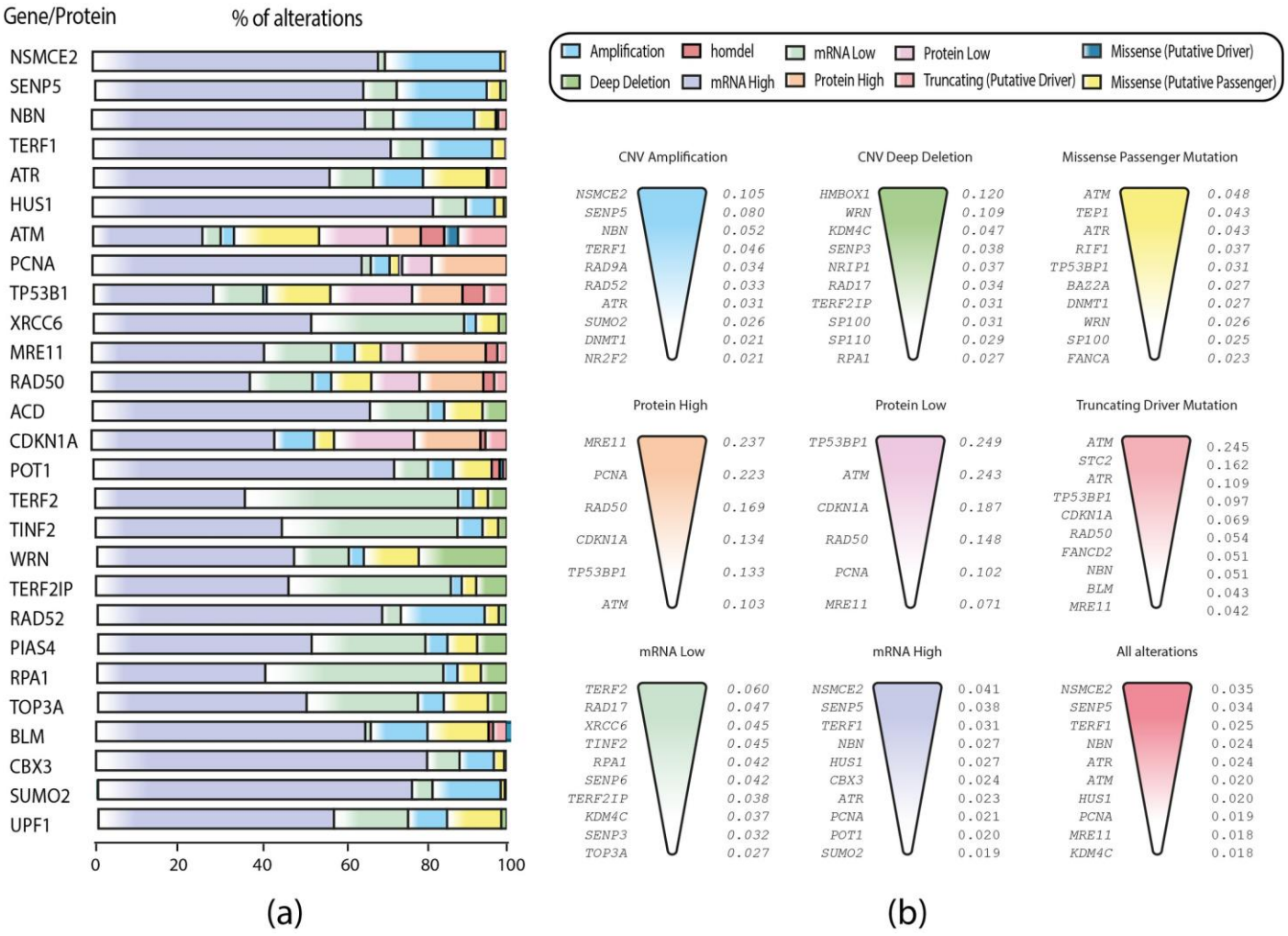
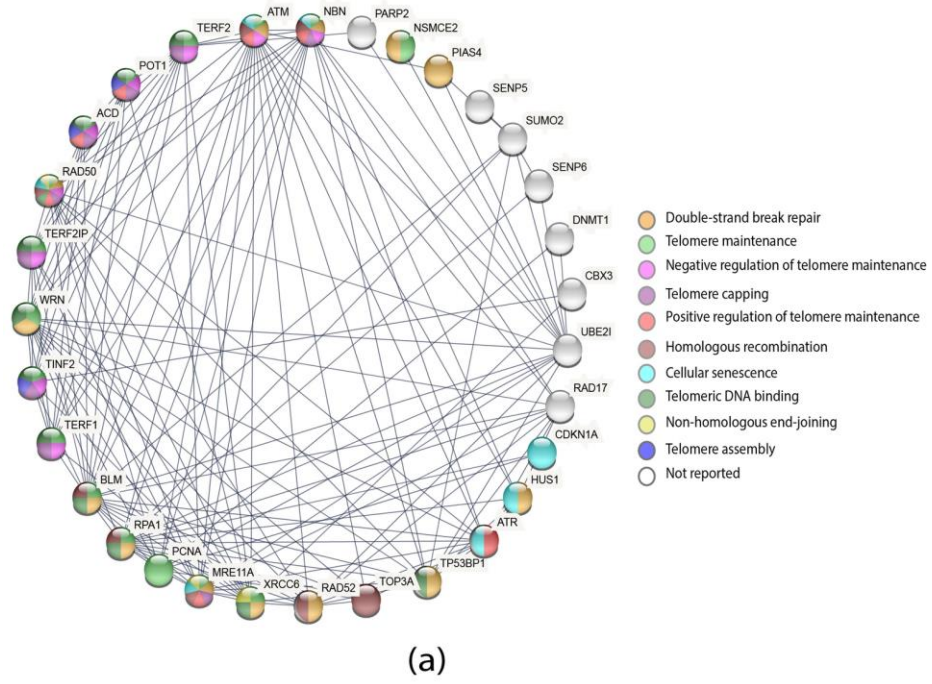


Figure 3. OncoPrint and ranking of the genes/proteins with the highest means of alterations across the PCA studies. (a) OncoPrint of genomics, transcriptomics and proteomic alterations across 32 TCGA Pan-Cancer studies. (b) Ranking of the most altered genes/proteins per alteration type.

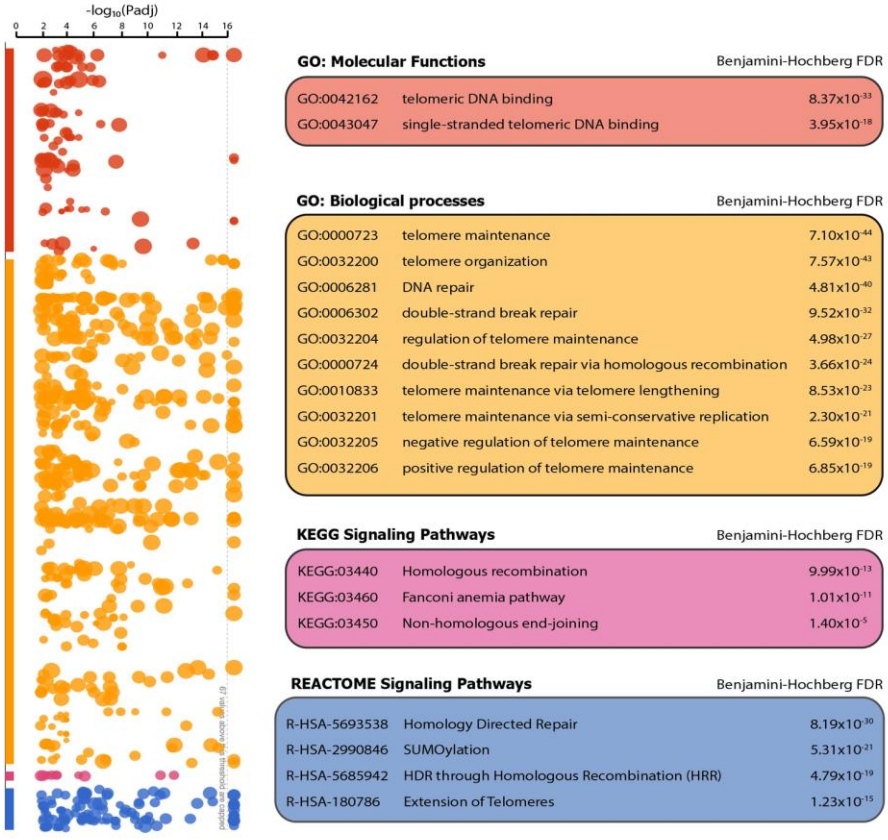
2.3. Protein-Protein interaction (PPi) network and functional enrichment analysis

PPi netwroks are fundamental resources to understand protein interactions among diseases (25). Thus,we analyzed the 71 APBs-related proteins selected for our study by querying the STRING database (26). After selecting the interaction score of the highest confidence (0.900) (27), according to the level of evidence of interactions, we obtained a network with 31 proteins interacting at the highest level of evidence, of which, 23 are involved in pathways significantly related ($p<0.001$) to the molecular functions linked to the mechanism of formation and function of APBs. Figure 4a shows interactions between proteins which are differentiated by colored nodes according to the most significant pathway in which each one is intervening; 70% of these are involved in telomeric DNA binding, 56% are involved in double-strand break repair (DSBR) and in Telomere Maintenance, 40% in telomere capping and positive and negative regulation of telomere maintenance and around 30% in homologous recombination and non-homologous end-joining (NHEJ) mechanisms.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 October 2021 doi:10.20944/preprints202110.0001.v1



(a)



(b)

Figure 4. Protein-protein interaction network and functional enrichment analysis. (a) PPi network showing APBs proteins interactions. Proteins are colored according to the most significant pathway they are involved (b) Functional enrichment analysis showing the most significant GO: biological processes, GO: molecular functions, and KEGG and REACTOME signaling pathways according to Benjamini-Hochberg FDR ($p<0.001$); data was visualized using the g:profiler software.

 doi:10.20944/preprints202110.0001.v1

Additionally, a functional enrichment analysis of the 71 proteins was performed using the g:profiler software (28) (Table S6). Figure 4b shows a Manhattan plot of the most significant GO: Molecular functions, GO: Biological processes and KEGG and REACTOME (29) signaling pathways with Benjamini-Hochberg FDR ($p<0.001$), which gives us a clearer idea about the function of the studied proteins and will allow us to understand and discuss the consequences of their genomic alterations in the different types of cancer.

2.4. Correlation between ALT and APBs in the TCGA PCA studies

In our previous study (12), we classified the PCA tumors based on literature reports and *in-silico* analyses as: frequent, rare and non-ALT tumors. With the help of a Venn diagram (Figure 5a), a correlation was made among the ALT-related tumors from our previous study and the most altered tumors across the 32 PCA types of the present study with APBs. Consequently, it is observed that SARC, SKCM, UCS, ACC and STAD (ALT frequent tumors) and BLCA, UCEC, ESCA, COAD, LUSC and BRCA (ALT rare tumors) show a high frequency of genomic, transcriptomic and proteomic alterations of APBs-related genes/proteins.

As a result, the most altered proteins from the ALT frequent and rare tumors mentioned above were selected and a PPi was performed with the most altered APBs proteins from the same PCA studies by using the same criteria of evidence and interactions used in the construction of the previous network. Figure 5b shows the interactions between the APBs-related and ALT-related proteins from the frequent ALT tumors, and Figure 5c shows the interactions between the same groups of proteins but from the rare ALT tumors. Thus, the interaction of ALT and APBs proteins and their genomic, transcriptomic and proteomic alterations can be correlated to improve the understanding of their association in the activation of telomerase-independent telomere maintenance mechanisms in cancer.

Then, to prioritize and identify a set of key proteins from the APBs, we integrated the most significant proteins from the networks in Figures 4a, 5b, 5c and the OncoPrint analysis in Figure 3a and 3b. As a consequence, Figure 5d shows a Venn diagram with the integrated analysis of the most significant APBs proteins from the different in-silico approaches applied, resulting in 13 key proteins.

Finally, Figure 6 shows a heatmap with the most significant ($p<0.001$) GO processes, functions, and signaling pathways where these 13 APBs proteins are interacting according to the PPi network and functional enrichment analysis performed in previous steps. A total of 21 pathways related to TM mechanisms were selected and all of them are related to telomere maintenance mechanisms.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 October 2021 doi:10.20944/preprints202110.0001.v1

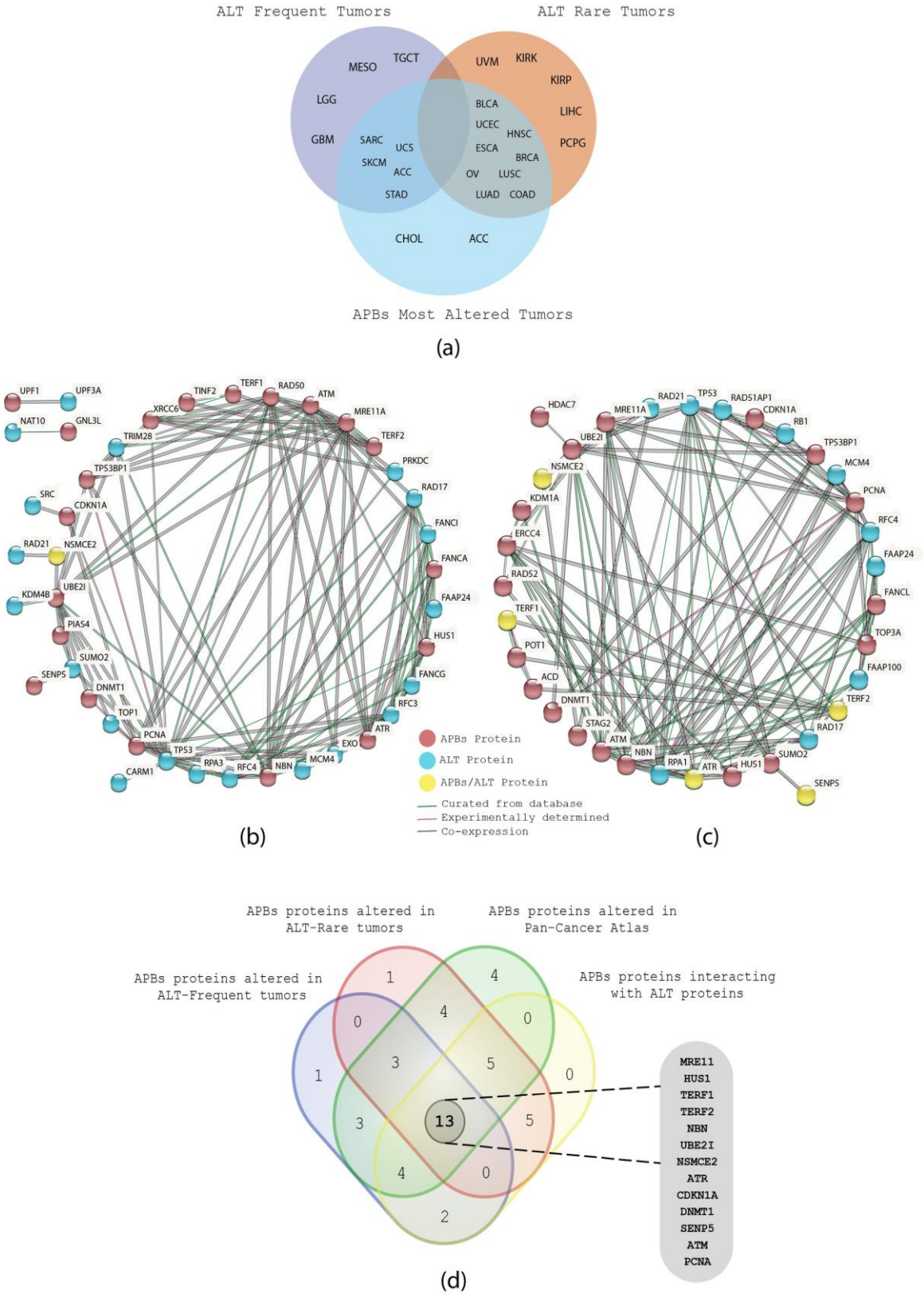


Figure 5. Correlation of APBs-related and ALT-related proteins. (a) Venn diagram is showing the association of ALT frequent and related tumors with the PCA tumors with the highest means of alterations in the APBs proteins. (b) PPi network is showing the interaction among ALT-frequent tumors proteins and APBs proteins. (c) PPi network is showing the interaction among ALT-rare tumors proteins and APBs proteins. (d) Venn diagram is showing an integrative analysis of different in-silico approaches which resulted in the obtention of the 13 most relevant proteins of the study.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 1 October 2021 doi:10.20944/preprints202110.0001.v1

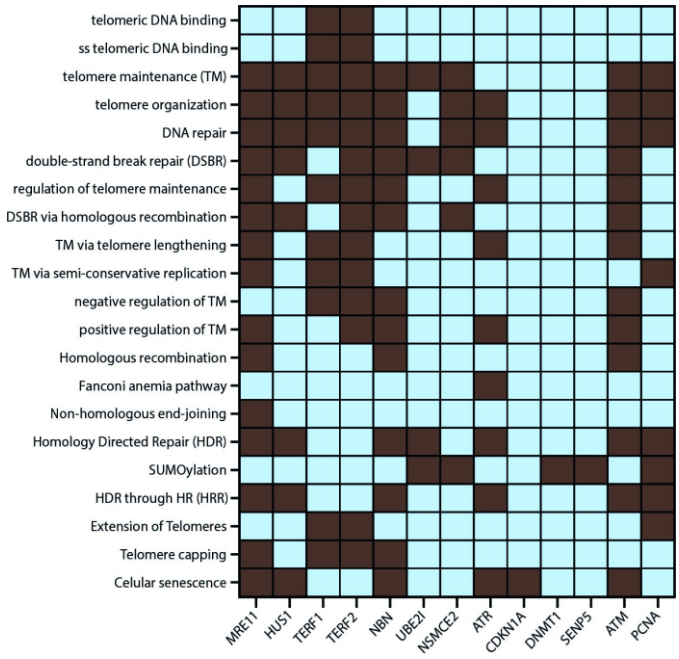


Figure 6. Heatmap showing the 21 most significant GO processes, functions and signaling pathways (dark boxes) in which the 13 most relevant APBs proteins obtained in this study are involved.

3. Discussion

The alternative lengthening of telomeres mechanism is a Break Induced Replication (BIR)-based process (11), through which, some cancer cells elongate their telomeres without the need of telomerase (30). Although ALT has been widely studied and described in the last years, the mechanism through which this is activated and most of its pathways are still poorly understood (6). One of the hallmarks of ALT is its association with promyelocytic leukemia nuclear bodies, better known as, APBs (31). APBs formation is driven by liquid-liquid phase separation with an environment marked by high levels of SUMOylated proteins, that bring telomeres together allowing ALT to occur; however, the way APBs assemble or how they promote ALT remains unclear (32). As a consequence, for this study we used a multi-omics approach to identify the genomic, transcriptomic and proteomic mutations of 71 APBs-related genes in 32 cancer types from the TCGA PanCancer Atlas; as a result, we proposed 13 key proteins which in addition to the 20 ALT-related proteins proposed in a past study (12), represent the best *in silico* evidence so far for the study of the ALT mechanism in cancer.

A total of 72,492 alterations were identified (Figure 1) being mRNA alterations and CNV amplifications the predominant mutations; in cancer, there is a close correlation between CNV and differential gene expression at a transcriptional level (33), therefore, the correlation of these alterations in the APBs-related genes may help to explain its function in the ALT pathway. ALT is known to commonly occur only in 10% to 15% of cancers, most of them from mesenchymal origin (34); however, in the last years, ALT$^+$ cells have been observed in a wide variety of epithelial tumors, and there is strong evidence of switching from a telomerase-mediated telomeric extension to ALT, as a consequence of anti-telomerase and radiation-based therapies which can trigger the accumulation of DNA damage response (DDR) factors in telomeres that can lead to ALT activation (35,36).

With the first quartile of PCA studies exhibiting the highest frequencies of alterations, we constructed a boxplot applying the Tukey´s test and as a consequence, the

 doi:10.20944/preprints202110.0001.v1

genes/proteins with different patterns of genomic, proteomic, and transcriptomic alterations were determined. Figure 2b shows genes/proteins SENP5, TERF1, UPF1, NSMCE2, CDKN1A, SUMO2, ACD, KDM1A, ATM, CBX3, TP53B1, TEP1, NBN, ATR, PIAS4, XRCC6, MRE11, TOP3A, SBTB48, RAD52, HUS1, and GNL3L to have the highest frequencies of alterations among the different PCA studies. Then, we wanted to observe the predominant mutations of each gene/protein, therefore, an OncoPrint with the first quartile of genes/proteins with the highest frequencies of alteration was constructed with the aid of alterations data from the cBioPortal (13,14); Figure 3a shows the predominant alterations of each gene/protein, where mRNA high has the highest percentage of alterations, nevertheless, XRCC6, TERF2, ATM, TP53B1, MRE11, RAD50, CDKN1A, TINF2, WRN, and RPA1 are genes/proteins with a different pattern of alterations, hence, we decided to rank each gene/protein per alteration; Figure 3b shows each genomic, transcriptomic or proteomic alteration with a ranking of the most altered gene/protein for each one. This ranking will allow us to elucidate the role of each gene/protein in APBs. For instance, NSMCE2 is highly amplified and overexpressed across the PCA studies (Figure 3b), this gene encodes a protein of the small ubiquitin-related modifier (SUMO) and it is part of the SMC5/6 complex, which is crucial for the SUMOylation of proteins (37) that is a hallmark of the APBs environment, additionally, the knockdown of NSMCE2 in ALT$^+$ cell lines had led to a reduction of telomere length (38).

To further understand the role of the proteins analyzed in this study, we constructed a protein interactome among the 71 APBs-related proteins by using the highest confidence score of 0.9 according to co-expression, curated from the STRING database and experimentally determined parameters; as a result in Figure 4a, a protein-protein interaction (PPi) network with 31 proteins is observed, the most significant ($p<0.001$) pathways were DNA double-strand breaks (DDSB), telomere maintenance (TM), negative and positive regulation of telomere maintenance, non-homologous end joining (NHEJ) and homologous recombination (HR). ALT is a BIR-related process, triggered by oxidative stress and due to cancer treatment; all NHEJ and HR, are BIR-related pathways that are proposed as the main way by which ALT$^+$ cells extend their telomeres (39), in fact, BIR-induced replication stress with the SUMOylation of key proteins initiates the recruitment of DNA damage response (DDR) factors in APBs of ALT$^+$ cell lines (40). Moreover, a functional enrichment analysis was performed according to Benjamini-Hochberg FDR score ($p<0.001$); Figure 4b shows a Manhattan plot with the most relevant molecular functions (telomeric DNA binding), biological processes (TM, DSB repair) and signaling pathways (HR, Fanconi anemia pathway, NHEJ, SUMOylation); Fanconi anemia pathway proteins like FANCD2, are highly correlated with ALT(41); one report suggests that its depletion can lead to a high telomeric extension in APBs and the over expression of H2AX and TP53B1 proteins which are positive regulators of ALT (40); in Figure 3b, FANCD2 is ranked in the truncating mutations groups, which is favorable for ALT activation.

In our last study, we classified the PCA studies according to literature reports and alteration frequencies as ALT-frequent tumors, ALT-rare tumors, and not reported (12). With the aim to observe in which category the PCA studies with the highest frequencies of genomic, transcriptomic and proteomic alterations of this study are, we grouped them in a Venn diagram which can be observed in Figure 5a, showing that SARC, SKCM, UCS, ACC and STAD are ALT-frequent tumors and BLCA, UCEC, ESCA, OV, HNSC, BRCA, LUSC, LUAD and COAD are in the ALT-rare tumors. Then, we constructed two protein interactomes applying the same criteria used for the previous PPi; Figure 5b shows an interactome of ALT-related proteins and APBs- related proteins which have high alteration frequencies in ALT-frequent tumors, while Figure 5c shows the interactions of ALT-related proteins and APBs-related proteins with the highest alteration frequencies in ALT-rare tumors. These interactomes can improve the understanding of how PMLs associate with ALT$^+$ cell lines for the assembling of APBs; for instance, HDAC7 is believed to

 doi:10.20944/preprints202110.0001.v1

promote PML protein SUMOylation (4), however, how the HDAC family interacts with PMLs is still unclear; in the interactome in Figure 5c, HDAC7 is interacting directly with UBE2I which is a highly expressed protein in ALT+ cancers like osteosarcoma (42,43).

Moreover, we integrated all the multi-omics approaches used in the study: PCA genomic, transcriptomic and proteomic alterations, protein interactomes, enrichment functional analysis and gene ontology in a single Venn diagram (Figure 5d) and identified 13 proteins that have significative alterations and interactions. MRE11 and NBN that are part of the MRN complex (MRE11/RAD50/NBN) which is predicted to be a key step in APBs formation (44); TERF1 and TERF2, components of the shelterin complex, are believed to be recruited by APBs with the aid of the MRN complex and PML protein and are predicted to be SUMOylated by NSMCE2 in ALT+ cell lines (45–47). Furthermore, TERF1 inhibition has been associated with the disassembly of APBs and a phosphorylated form of TERF1 has been detected in the APBs of some ALT+ cell lines (48–50) however, the mechanism behind TERF1/2 interaction with the PML protein in the assemble of APBs remains unclear (46). CDKN1A protein knockdown is associated with activation of ALT (51), Figure 3b, shows CDKN1A to have low protein expression in about 20% of the PCA studies and truncating mutations in 10% of cancers. HUS1 is part of the 911 (RAD9-RAD1-HUS1) checkpoint which activates ATR and protects telomere integrity during DNA damage response and oxidative stress (41,52). SENP5 is a key protein in the SUMOylation pathway (42,53), UBE2I is a positive regulator of APBs (42,43), DNMT1 depletion is related to ALT activation (41) and ATM and PCNA have been observed to colocalize with APBs during HR and HDR (54,55). Finally, to summarize the role of the 13 key APBs-related proteins, a heatmap was elaborated showing the 21 most significant pathways in which they are interacting (Figure 6).

This work identified 13 key APBs-related proteins which after a series of integrated *in-silico* and multi-omics analyses showed distinctive genomic, transcriptomic and proteomic alterations, significant protein-protein interaction patterns and to be involved in significant pathways related to telomere maintenance through APBs. This protein set in addition to 20 ALT-related proteins identified in a previous study (12), represents so far to our knowledge the most complete *in-silico* evidence of potential molecular targets for the study of the ALT pathway. Further *ex vivo* analyses will support this *in silico* results and the addition of this set of proteins to fill the knowledge gap that currently exists in the ALT pathway research. Bioinformatics techniques through the aid of computational biology models have proven to be valuable tools to prioritize proteins that could improve future *in vivo* research in different hallmarks of cancer progression as the APBs mediated telomere maintenance.

4. Materials and Methods

4.1. Gene/protein set

TelNet (**http://www.cancertelsys.org/telnet/**) is a database that groups more than 2000 human telomere maintenance (TM) genes. All genes are annotated according to their classification of telomere maintenance mechanism, telomere maintenance function, and a significance score given by the evidence of gene function in telomeres (56). The database shows the role of each gene in ALT and telomerase-mediated mechanisms. Therefore, the TelNet database was downloaded and manually filtered resulting in a set of 71 genes that are related to PMLs and APBs.

 doi:10.20944/preprints202110.0001.v1

4.2. TCGA PanCancer studies frequencies and OncoPrint of genomic and proteomic alterations

After selecting the set of APBs-related genes, we analyzed their genomic, transcriptomic and proteomic alterations in 32 cancer studies from the PanCancer Atlas (PCA) project which is part of The Cancer Genome Atlas (TCGA) consortium (15,20). With the aid of the cBio Portal database (**http://www.cbioportal.org/**) (13,14), a total of 10,918 samples were selected from the 32 PCA studies: LAML, ACC, BLCA, LGG, BRCA, CESC, CHOL, COAD, DLBC, ESCA, GBM, HNSC, KICH, KIRC, KIRP, LIHC, LUAD, LUSC, MESO, OV, PAAD, PCPG, PRAD, SARC, SKCM, STAD, TGCT, THYM, THCA, UCS, UCEC and UVM (15–24). The cBioPortal uses data from the GISTIC2.0 computational approach which facilitates sensitive and confident localization of CNV amplifications and deep deletions in human cancers (57); additionally, it identifies inframe, truncating, and missense mutations through whole-exome sequencing; mRNA up and down-regulation are analyzed through RNA sequencing V2 RSEM by comparing the expression Z-scores of tumor samples to the logarithmic expression of mRNA of adjacent normal samples (58) and the up and down expression of protein are measured by reverse-phase protein arrays (RPPA) (59).

To calculate the frequency means of every genomic, transcriptomic, and proteomic alteration and construct the OncoPrint we: 1) filtered and calculated the number of alterations per gene and per PCA type; 2) calculated the frequency of alteration of each gene through normalization by dividing the number of alterations by the number of individuals of each cancer study; 3) identified the most altered APBs genes/proteins with the aid of a boxplot by using Tukey´s test; 4) validated the most significantly altered genes/proteins with a multiple comparison test by using the Original False Discovery Rate (FDR) method of Benjamini and Hochberg using the GraphPad Prism v9.1.1 software ($p< 0.01$) (60)

4.3. Protein-protein interaction network

In order to predict the most essential protein interactions, and APBs protein-protein interactome (PPi) network was constructed with the aid of the STRING database (**https://string-db.org/**). An interaction score of 0.9 (highest confidence) was set according to co-expression, curated from the database, and experimentally determined (26,61). The most significant signaling pathways ($p< 0.001$) related to APBs were selected and differentiated by colors in the network.

4.4. Functional enrichment analysis

An enrichment analysis gives curated signatures of protein sets generated from omics experiments (62). Thus, we performed the analysis of the 71 APBs proteins by using the g:Profiler tool version e104_eg51_p15_3922dba (**https://biit.cs.ut.ee/gprofiler/gost**) (28). The most significant annotations were selected after Benjamini-Hochberg and False Discovery Rate (FDR) corrections ($p< 0.001$), based on Gene Ontology (GO), Molecular Function (MF), Biological Process (BP), the Kyoto Encyclopedia of Genes and Genomes (KEGG) and REACTOME signaling pathways (28,63).

4.5. Correlation between ALT and APBs in the PanCancer studies

In our previous research (12) ALT tumors were classified as frequent ALT tumors and rare ALT tumors. For this study, we wanted to determine whether the key APBs proteins identified in this study have a correlation with the ALT proteins identified in our latter study. Under this context we: 1) elaborated a Venn diagram to correlate the type of

ALT tumor with the PCA studies with the most significant patterns of genomic, transcriptomic, and proteomic alterations; 2) constructed PPi networks to predict the interactions among the APBs proteins with the highest frequencies of alterations and the most altered proteins from each ALT-frequent and ALT-rare groups determined in our previous study; and 3) with the aid of a Venn diagram we integrated four different approaches in order to predict key proteins from the APBs pathway.

Supplementary Materials: Table S1: Gene/protein set. Table S2: Genomic, transcriptomic and proteomic alterations across the 32 PCA studies. Table S3: TCGA Study vs Alteration (normalized). Table S4: TCGA Study vs gene (normalized). Table S5: OncoPrint. Table S6: Functional enrichment analysis.

Author Contributions: Conceptualization, Isaac Armendáriz-Castillo; Data curation, Isaac Armendáriz-Castillo and Katherine Hidalgo-Fernández; Methodology, Isaac Armendáriz-Castillo and Andrés López-Cortés; Supervision, Andrés López-Cortés and Santiago Guerrero; Writing – original draft, Isaac Armendáriz-Castillo and Katherine Hidalgo-Fernández; Writing – review & editing, Andy Pérez-Villa, Jennyfer García-Cárdenas, Andrés López-Cortés and Santiago Guerrero.

Funding: This research received no external funding

Data Availability Statement: All data generated for this study is included in the manuscript and its supplementary files.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mentegari E, Bertoletti F, Kissova M, Zucca E, Galli S, Tagliavini G, et al. A Role for Human DNA Polymerase λ in Alternative Lengthening of Telomeres. Int J Mol Sci. 2021 Feb;22(5).
2. Yang C-W, Hsieh M-H, Sun H-J, Teng S-C. Nuclear envelope tethering inhibits the formation of ALT-associated PML bodies in ALT cells. Aging (Albany NY). 2021 Apr;13(7):10490–516.
3. Zhang H, Zhao R, Tones J, Liu M, Dilley RL, Chenoweth DM, et al. Nuclear body phase separation drives telomere clustering in ALT cancer cells. Mol Biol Cell. 2020 Aug;31(18):2048–56.
4. Jamiruddin MR, Kaitsuka T, Hakim F, Fujimura A, Wei F-Y, Saitoh H, et al. HDAC9 regulates the alternative lengthening of telomere (ALT) pathway via the formation of ALT-associated PML bodies. Biochem Biophys Res Commun. 2016 Dec;481(1–2):25–30.
5. Li Y, Ma X, Wu W, Chen Z, Meng G. PML Nuclear Body Biogenesis, Carcinogenesis, and Targeted Therapy. Trends in cancer. 2020 Oct;6(10):889–906.
6. Graham MK, Kim J, Da J, Brosnan-Cashman JA, Rizzo A, Baena Del Valle JA, et al. Functional Loss of ATRX and TERC Activates Alternative Lengthening of Telomeres (ALT) in LAPC4 Prostate Cancer Cells. Mol Cancer Res. 2019 Dec;17(12):2480–91.
7. Min J, Wright WE, Shay JW. Clustered telomeres in phase-separated nuclear condensates engage mitotic DNA synthesis through BLM and RAD52. Genes Dev. 2019 Jul;33(13–14):814–27.
8. Fonin A V, Silonov SA, Shpironok OG, Antifeeva IA, Petukhov A V, Romanovich AE, et al. The Role of Non-Specific Interactions in Canonical and ALT-Associated PML-Bodies Formation and Dynamics. Int J Mol Sci. 2021 May;22(11).
9. Salmina K, Bojko A, Inashkina I, Staniak K, Dudkowska M, Podlesniy P, et al. "Mitotic Slippage" and Extranuclear DNA in Cancer Chemoresistance: A Focus on Telomeres. Int J Mol Sci. 2020 Apr;21(8).
10. Zhang J-M, Yadav T, Ouyang J, Lan L, Zou L. Alternative Lengthening of Telomeres through Two Distinct Break-Induced Replication Pathways. Cell Rep. 2019 Jan;26(4):955-968.e3.
11. Zhang JM, Yadav T, Ouyang J, Lan L, Zou L. Alternative Lengthening of Telomeres through Two Distinct Break-Induced

Replication Pathways. Cell Rep [Internet]. 2019;26(4):955-968.e3. Available from: https://doi.org/10.1016/j.celrep.2018.12.102

12. Armendáriz-Castillo I, López-Cortés A, García-Cárdenas J, Guevara-Ramírez P, Leone PE, Pérez-Villa A, et al. TCGA Pan-Cancer Genomic Analysis of Alternative Lengthening of Telomeres (ALT) Related Genes. Genes (Basel). 2020 Jul;11(7).
13. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013 Apr;6(269):pl1.
14. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. Cancer Discov [Internet]. 2012 May 1;2(5):401 LP – 404. Available from: http://cancerdiscovery.aacrjournals.org/content/2/5/401.abstract
15. Hoadley KA, Yau C, Hinoue T, Wolf DM, Lazar AJ, Drill E, et al. Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. Cell. 2018 Apr;173(2):291-304.e6.
16. Berger AC, Korkut A, Kanchi RS, Hegde AM, Lenoir W, Liu W, et al. A Comprehensive Pan-Cancer Molecular Study of Gynecologic and Breast Cancers. Cancer Cell [Internet]. 2018;33(4):690-705.e9. Available from: http://www.sciencedirect.com/science/article/pii/S1535610818301193
17. Liu Y, Sethi NS, Hinoue T, Schneider BG, Cherniack AD, Sanchez-Vega F, et al. Comparative Molecular Analysis of Gastrointestinal Adenocarcinomas. Cancer Cell [Internet]. 2018;33(4):721-735.e8. Available from: http://www.sciencedirect.com/science/article/pii/S1535610818301144
18. Campbell JD, Yau C, Bowlby R, Liu Y, Brennan K, Fan H, et al. Genomic, Pathway Network, and Immunologic Features Distinguishing Squamous Carcinomas. Cell Rep [Internet]. 2018;23(1):194-212.e6. Available from: http://www.sciencedirect.com/science/article/pii/S2211124718304248
19. Ricketts CJ, De Cubas AA, Fan H, Smith CC, Lang M, Reznik E, et al. The Cancer Genome Atlas Comprehensive Molecular Characterization of Renal Cell Carcinoma. Cell Rep [Internet]. 2018;23(1):313-326.e5. Available from: http://www.sciencedirect.com/science/article/pii/S2211124718304364
20. Huang K, Mashl RJ, Wu Y, Ritter DI, Wang J, Oh C, et al. Pathogenic Germline Variants in 10,389 Adult Cancers. Cell [Internet]. 2018;173(2):355-370.e14. Available from: http://www.sciencedirect.com/science/article/pii/S0092867418303635
21. Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, et al. Comprehensive Characterization of Cancer Driver Genes and Mutations. Cell [Internet]. 2018;173(2):371-385.e18. Available from: http://www.sciencedirect.com/science/article/pii/S009286741830237X
22. Gao Q, Liang W-W, Foltz SM, Mutharasu G, Jayasinghe RG, Cao S, et al. Driver Fusions and Their Implications in the Development and Treatment of Human Cancers. Cell Rep [Internet]. 2018;23(1):227-238.e3. Available from: http://www.sciencedirect.com/science/article/pii/S2211124718303954
23. Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, et al. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. Cell [Internet]. 2018;173(2):400-416.e11. Available from: http://www.sciencedirect.com/science/article/pii/S0092867418302290
24. Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC, et al. Oncogenic Signaling Pathways in The Cancer Genome Atlas. Cell [Internet]. 2018;173(2):321-337.e10. Available from: http://www.sciencedirect.com/science/article/pii/S0092867418303593
25. López-Cortés A, Guevara-Ramírez P, Guerrero S, Ortiz-Prado E, García-Cárdenas JM, Zambrano AK, et al. Metastatic signaling of hypoxia-related genes across TCGA Pan-Cancer types. bioRxiv [Internet]. 2020 Jan 1;2020.02.01.930479. Available from: http://biorxiv.org/content/early/2020/02/02/2020.02.01.930479.abstract
26. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015 Jan;43(Database issue):D447-52.
27. García-Cárdenas JM, Guerrero S, López-Cortés A, Armendáriz-Castillo I, Guevara-Ramírez P, Pérez-Villa A, et al. Post-

transcriptional Regulation of Colorectal Cancer: A Focus on RNA-Binding Proteins [Internet]. Vol. 6, Frontiers in Molecular Biosciences . 2019. p. 65. Available from: https://www.frontiersin.org/article/10.3389/fmolb.2019.00065

28. Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res [Internet]. 2019 May 8;47(W1):W191–8. Available from: https://doi.org/10.1093/nar/gkz369
29. Fabregat A, Sidiropoulos K, Garapati P, Gillespie M, Hausmann K, Haw R, et al. The Reactome pathway Knowledgebase. Nucleic Acids Res. 2016 Jan;44(D1):D481-7.
30. Lovejoy CA, Takai K, Huh MS, Picketts DJ, de Lange T. ATRX affects the repair of telomeric DSBs by promoting cohesion and a DAXX-dependent activity. PLoS Biol. 2020 Jan;18(1):e3000594.
31. Chang HR, Munkhjargal A, Kim M-J, Park SY, Jung E, Ryu J-H, et al. The functional roles of PML nuclear bodies in genome maintenance. Mutat Res. 2018 May;809:99–107.
32. Wee Y, Liu Y, Lu J, Li X, Zhao M. Identification of novel prognosis-related genes associated with cancer using integrative network analysis. Sci Rep. 2018 Feb;8(1):3233.
33. Shao X, Lv N, Liao J, Long J, Xue R, Ai N, et al. Copy number variation is highly correlated with differential gene expression: a pan-cancer study. BMC Med Genet. 2019 Nov;20(1):175.
34. Sobinoff AP, Pickett HA. Mechanisms that drive telomere maintenance and recombination in human cancers. Curr Opin Genet Dev. 2020 Feb;60:25–30.
35. De Vitis M, Berardinelli F, Sgura A. Telomere length maintenance in cancer: At the crossroad between telomerase and alternative lengthening of telomeres (ALT). Int J Mol Sci. 2018;19(2).
36. Hu Y, Shi G, Zhang L, Li F, Jiang Y, Jiang S, et al. Switch telomerase to ALT mechanism by inducing telomeric DNA damages and dysfunction of ATRX and DAXX. Sci Rep. 2016;6(August):1–10.
37. Alhendi ASN, Royle NJ. The absence of (TCAGGG)(n) repeats in some telomeres, combined with variable responses to NR2F2 depletion, suggest that this nuclear receptor plays an indirect role in the alternative lengthening of telomeres. Sci Rep. 2020 Nov;10(1):20597.
38. Zhang J-M, Zou L. Alternative lengthening of telomeres: from molecular mechanisms to therapeutic outlooks. Cell Biosci [Internet]. 2020;10(1):1–9. Available from: https://doi.org/10.1186/s13578-020-00391-6
39. Hussain SS, Majumdar R, Moore GM, Narang H, Buechelmaier ES, Bazil MJ, et al. Measuring nonhomologous end-joining, homologous recombination and alternative end-joining simultaneously at an endogenous locus in any transfectable human cell. Nucleic Acids Res. 2021 Jul;49(13):e74.
40. Zhang J-M, Genois M-M, Ouyang J, Lan L, Zou L. Alternative lengthening of telomeres is a self-perpetuating process in ALT-associated PML bodies. Mol Cell. 2021 Mar;81(5):1027-1042.e4.
41. Osterwald S, Deeg KI, Chung I, Parisotto D, Worz S, Rohr K, et al. PML induces compaction, TRF2 depletion and DNA damage signaling at telomeres and promotes their alternative lengthening. J Cell Sci. 2015 May;128(10):1887–900.
42. Meng Y, Li X. Expression and Prognosis Analysis of SUMOylation Regulators in Oral Squamous Cell Carcinoma Based on High-Throughput Sequencing. Front Genet. 2021;12:671392.
43. Yang H, Gao S, Chen J, Lou W. UBE2I promotes metastasis and correlates with poor prognosis in hepatocellular carcinoma. Cancer Cell Int. 2020;20:234.
44. Jiang W-Q, Zhong Z-H, Henson JD, Reddel RR. Identification of candidate alternative lengthening of telomeres genes by methionine restriction and RNA interference. Oncogene. 2007 Jul;26(32):4635–47.
45. Yu J, Lan J, Wang C, Wu Q, Zhu Y, Lai X, et al. PML3 interacts with TRF1 and is essential for ALT-associated PML bodies assembly in U2OS cells. Cancer Lett. 2010 May;291(2):177–86.
46. Feng X, Luo Z, Jiang S, Li F, Han X, Hu Y, et al. The telomere-associated homeobox-containing protein TAH1/HMBOX1

participates in telomere maintenance in ALT cells. J Cell Sci. 2013 Sep;126(Pt 17):3982–9.

47. Hsu JK, Lin T, Tsai RYL. Nucleostemin prevents telomere damage by promoting PML-IV recruitment to SUMOylated TRF1. J Cell Biol. 2012 May;197(5):613–24.
48. Ho A, Wilson FR, Peragine SL, Jeyanthan K, Mitchell TRH, Zhu X-D. TRF1 phosphorylation on T271 modulates telomerase-dependent telomere length maintenance as well as the formation of ALT-associated PML bodies. Sci Rep. 2016 Nov;6:36913.
49. Wang C, Songyang Z, Huang Y. TRIM28 inhibits alternative lengthening of telomere phenotypes by protecting SETDB1 from degradation. Cell Biosci. 2021 Jul;11(1):149.
50. Wilson FR, Ho A, Walker JR, Zhu X-D. Cdk-dependent phosphorylation regulates TRF1 recruitment to PML bodies and promotes C-circle production in ALT cells. J Cell Sci. 2016 Jul;129(13):2559–72.
51. Min J, Wright WE, Shay JW. Alternative Lengthening of Telomeres Mediated by Mitotic DNA Synthesis Engages Break-Induced Replication Processes. Mol Cell Biol [Internet]. 2017 Oct 15;37(20):e00226-17. Available from: http://mcb.asm.org/content/37/20/e00226-17.abstract
52. Tan J, Wang X, Hwang B-J, Gonzales R, Konen O, Lan L, et al. An ordered assembly of MYH glycosylase, SIRT6 protein deacetylase, and Rad9-Rad1-Hus1 checkpoint clamp at oxidatively damaged telomeres. Aging (Albany NY). 2020 Sep;12(18):17761–85.
53. Loe TK, Li JSZ, Zhang Y, Azeroglu B, Boddy MN, Denchi EL. Telomere length heterogeneity in ALT cells is maintained by PML-dependent localization of the BTR complex to telomeres. Genes Dev. 2020 May;34(9–10):650–62.
54. Shen M, Young A, Autexier C. PCNA, a focus on replication stress and the alternative lengthening of telomeres pathway. DNA Repair (Amst). 2021 Apr;100:103055.
55. Xu M, Qin J, Wang L, Lee H-J, Kao C-Y, Liu D, et al. Nuclear receptors regulate alternative lengthening of telomeres through a novel noncanonical FANCD2 pathway. Sci Adv. 2019 Oct;5(10):eaax6366.
56. Braun DM, Chung I, Kepper N, Deeg KI, Rippe K. TelNet - a database for human and yeast genes involved in telomere maintenance. BMC Genet. 2018;19(1):1–10.
57. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol. 2011;12(4):R41.
58. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011 Aug;12:323.
59. Boellner S, Becker K-F. Reverse Phase Protein Arrays-Quantitative Assessment of Multiple Biomarkers in Biopsies for Clinical Use. Microarrays (Basel, Switzerland). 2015 Mar;4(2):98–114.
60. López-Cortés A, Paz-y-Miño C, Guerrero S, Cabrera-Andrade A, Barigye SJ, Munteanu CR, et al. OncoOmics approaches to reveal essential genes in breast cancer: a panoramic view from pathogenesis to precision medicine. Sci Rep [Internet]. 2020;10(1):5285. Available from: https://doi.org/10.1038/s41598-020-62279-2
61. Guerrero S, López-Cortés A, García-Cárdenas JM, Armendáriz-Castillo I, Zambrano AK, Indacochea A, et al. In silico analyses reveal new putative Breast Cancer RNA-binding proteins. bioRxiv [Internet]. 2020 Jan 1;2020.01.08.898965. Available from: http://biorxiv.org/content/early/2020/01/09/2020.01.08.898965.abstract
62. Reimand J, Isserlin R, Voisin V, Kucera M, Tannus-Lopes C, Rostamianfar A, et al. Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. Nat Protoc. 2019 Feb;14(2):482–517.
63. López-Cortés A, Abarca E, Silva L, Velastegui E, León-Sosa A, Karolys G, et al. Identification of key proteins in the signaling crossroads between wound healing and cancer hallmark phenotypes. Sci Rep. 2021 Aug;11(1):17245.