TelNet - a database for human and yeast genes involved in telomere maintenance

Delia M. Braun[1], Inn Chung[1], Nick Kepper[1], Katharina I. Deeg[1], & Karsten Rippe[1,*]

[1] Division of Chromatin Networks, German Cancer Research Center (DKFZ) & BioQuant, 69120 Heidelberg, Germany

* Corresponding author: Karsten.Rippe@dkfz.de, Tel: +49-6221-54-51376

Abstract

The ends of linear chromosomes, the telomeres, comprise repetitive DNA sequences that are protected by the shelterin protein complex. Cancer cells need to extend these telomere repeats for their unlimited proliferation, either by reactivating the reverse transcriptase telomerase or by using the alternative lengthening of telomeres (ALT) pathway. The different telomere maintenance (TM) mechanisms appear to involve hundreds of proteins but their telomere repeat length related activities are only partly understood. Currently, a database that integrates information on TM relevant genes is missing. To provide a reference for studies that dissect TM features, we here introduce the *TelNet* database at http://www.cancertelsys.org/telnet/. It offers a comprehensive compilation of more than 2,000 human and over 1,100 yeast genes linked to telomere maintenance. These genes were annotated in terms of TM mechanism, associated specific functions and orthologous genes, a TM significance score and information from peer-reviewed literature. This TM information can be retrieved via different search and view modes and evaluated for a set of genes on a statistics page. With these features *TelNet* can be integrated into the annotation of genes identified from bioinformatics analysis pipelines to determine possible connections with TM networks as illustrated by an exemplary application. We anticipate that *TelNet* will be a helpful resource for researchers that study TM processes.

Keywords

Telomere maintenance, telomerase, alternative telomere lengthening, pathway analysis

Background

Telomeres, the ends of linear chromosomes, consist of repetitive DNA sequences bound by the shelterin protein complex [1, 2]. This protein assembly protects the DNA ends from degradation and accidental recognition as DNA double-strand breaks [3-5]. The progressive shortening of the telomere repeat sequences that accompanies normal replication limits the number of cell divisions. Thus, it needs to be circumvented by cancer cells for unlimited proliferation. This is accomplished by activation of a telomere maintenance (TM) mechanism. It involves either the reactivation of the reverse transcriptase telomerase normally repressed in somatic cells via different mechanisms [6-9], or activation of the alternative lengthening of telomeres (ALT) pathway [10-13]. ALT activity in human cancer cells occurs via DNA repair and recombination pathways but details on the mechanism remain elusive. Thus, TM is a complex process that involves proteins that are located as part of the shelterin complex at telomere repeats [14, 15] or in close proximity [16, 17]. Factors that regulate transcription and functional activity of telomerase are relevant [18, 19]

as well as features of the ALT pathway like PML (promyelocytic leukemia) nuclear bodies along with telomere repeats that are associated with a variety of proteins and referred to as APBs (ALT-associated PML nuclear bodies) [20-23]. Furthermore, investigations of deregulating effects upon telomere shortening have related a number of proteins to telomere crisis [24].

A well-studied model organism for telomere biology is the budding yeast *Saccharomyces cerevisiae* [25]. Several independent deletion screens with subsequent direct measurements of telomere length have identified a comprehensive list of yeast genes involved in telomere length regulation [26-28]. Since telomere structure and function are highly conserved between organisms, mammalian homologues exist for most of the genes identified in the various yeast screens. Thus, it can be informative to relate TM phenotypes found in yeast to human homologues [29]. In *S. cerevisiae*, telomerase is constitutively active and its deletion leads to cellular senescence [30]. Survivor cells that overcome cellular senescence in the absence of telomerase use a mechanism based on homologous recombination for telomere elongation [31]. Interestingly, similar to ALT in human cells, so-called type II survivors are characterized by heterogeneous telomere lengths [32, 33].

The following existing databases provide some telomere-relevant information: *TeloPIN* (Telomeric Proteins Interaction Network, http://songyanglab.sysu.edu.cn/telopin/index.php) is a collection of interaction data in human and mouse cells from available literature and GEO (gene expression omnibus) database [34]. It includes interactions with shelterin proteins and protein interactions of the non-coding telomere repeat-containing RNA (TERRA). It also provides information on the methods used to identify the respective interaction. The *Telomerase database* (http://telomerase.asu.edu/overview.html) is a web-based tool for the study of structure, function, and evolution of the telomerase ribonucleoprotein [35]. It is a comprehensive compilation of information on the telomerase enzyme and its DNA substrate. The previously published *TeCK database* (http://www.bioinfosastra.com/services/teck/index.html) is a collection of telomeric and centromeric sequences as well as telomerase, centromere and kinetochore binding proteins [36] but appears to be no longer available. Finally, *MiCroKiTS* (Midbody, Centrosome, Kinetochore, Telomere and Spindle; http://microkit.biocuckoo.org) provides information on the cellular localization of proteins relevant for cell cycle progression and also includes telomere proteins [37].

All the above-mentioned databases cover telomere related information but lack an annotation of genes with respect to the TM mechanism. Accordingly, we here introduce a compilation of information on telomere maintenance relevant genes via the *TelNet* database. *TelNet* currently comprises more than 2,000 human, and over 1,100 yeast genes that are involved in TM pathways. The annotation of these genes includes the classification of

telomere maintenance mechanisms along with a significance score as well as TM specific functions and homology assignments between different organisms. Furthermore, links to the respective literature sources are given. Thus, *TelNet* provides an integrative platform for dissecting TM networks and elucidating the alternative lengthening of telomeres pathway.

Construction and content

Implementation

The *TelNet* database was constructed using the Filemaker Pro software version 13. It is accessible at http://www.cancertelsys.org/telnet and is distributed via Filemaker server version 15 via its webdirect module. The *TelNet* webpage provides general information about *TelNet* as well as instructions on how to use it. Links to other databases and contact information are provided as well.

Data source

To compile an initial set of TM relevant genes, we selected screening studies on genes or proteins that play a role in telomere biology (**Figure 1**, **Table 1**) that included the following: (i) Proteins that bound to a telomere probe in an ALT- and a telomerase-positive cell line [14], (ii) proteins from the analysis of telomeric chromatin of telomerase-positive cells [15], (iii) proteins in close proximity to shelterin components [16, 17], (iv) proteins that affected ALT-associated PML nuclear bodies [23, 38], (v) deregulated proteins linked to telomere shortening [24], (vi) genes identified from telomerase activity signatures derived from gene expression data [39], (vii) telomerase regulators identified from a kinase screen or reviewed recently [18, 19] and, (viii) a gene set with potential relevance to telomeres and the ALT pathway [40]. In addition, more than 1,100 budding yeast genes were included in *TelNet*. These genes were identified from the following sources: (i) Deletion screens that identifying telomere length associated genes [26-28], (ii) post-senescent survivor screening after telomerase knockout [41] and, (iii) transcription factors of telomerase [42].

To differentiate the relevance of a gene or corresponding protein we introduced three categories of TM significance: ‘predicted’, ‘screened’ and ‘validated’. The factors collected from the above-mentioned screening or review sources were classified as ‘screened’. Genes with a suggested role in telomere maintenance but lacking experimentally validation were assigned with the TM significance ‘predicted’. Those with detailed experimental evidence for a connection to telomere maintenance were ranked as ‘validated’. Orthologues of gene’s classified as ‘screened’ or ‘validated’ were included in the *TelNet* database as 'predicted', if no further information was available. Moreover, all human and yeast genes with a GO annotation containing the term ‘telo’ [43] were included into the *TelNet* database. In this

manner, we compiled an initial list of human and yeast genes that was further curated and annotated manually.

General information from external databases

For a standardized nomenclature [44], the converter system from DAVID Bioinformatics Resources (https://david.ncifcrf.gov/) [45] or the BioMart tool from Ensembl [46] were used to provide gene and protein identifiers for Entrez, Hugo, Ensembl, Refseq and UniProt. To account for organism specific differences such as the lack of splicing and therefore isoforms in yeast or the absence of locus tags in human, the identifiers were selected differentially for each species. General gene information was retrieved from designated external databases and repositories, such as the National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov) [47], HUGO Gene Nomenclature Committee (HGNC, http://www.genenames.org) [48], Ensembl (http://www.ensembl.org/index.html) [49, 50], and the Saccharomyces Genome Database (SGD, http://www.yeastgenome.org) [51]. The approved gene symbol, full name, and synonyms were taken from NCBI. UniProt [52] and Yeastmine [53] were consulted for the description of the cellular function in human and yeast, respectively, and assignment of orthologues was done with YeastMine. Based on the Gene Ontology (GO, http://www.geneontology.org) annotations [43] and in line with biocuration guidelines [54] we generated a list of cellular functions similar to curation procedures at SGD [55]. Every protein was manually annotated with the respective term that was most representative for its cellular function. In this manner, general information for every gene entry was compiled from a variety of external databases.

Telomere maintenance annotation with literature information and scoring

Genes were further annotated with TM information from peer-reviewed literature for assigning them to functional categories (**Figure 2**). Up to five TM functions of an assembled list that comprises molecular functions as well as cellular processes and structures with regard to TM can be selected. A knock-out or knock-down phenotype related to TM features such as alterations in telomere length, increased or decreased ALT hallmarks, or effects on telomerase was described as free-text in the field ‘TM phenotype’. Details from the literature were summarized in the field ‘TM comment’. To quantify the significance of a given gene for TM we introduced the *TelNet* score ranging from 1 (low) to 10 (high) that is automatically calculated from information entered into the *TelNet* database (**Table 2**). Scoring criteria include the cellular function, number and relevance of assigned TM functions, and the amount of experimental data associated with the TM function of a given gene. Information of the proteins activity in the TM mechanism (TMM) is collected in the ‘TMM annotation’. For

human genes, it is distinguished between alternative lengthening of telomeres (ALT) and telomerase regulation. Yeast genes can be annotated as type I and II survivors or in respect to telomerase regulation. Regulative activities were assigned as ‘repressing’, ‘enhancing’, or ‘conflicting’. The latter refers to cases where literature information was inconsistent or if genes were mentioned in the context of ALT or telomerase without further details of regulation activity. Thus, the annotation of a given factor in *TelNet* informs if it is involved in ALT or a telomerase dependent TM and how it affects this process. Furthermore, the corresponding *TelNet* score provides an assessment of the significance of this assignment.

Utility and discussion

TelNet user interface

On the front layout of *TelNet,* the user must select either *H. sapiens* or *S. cerevisiae* (**Figure 3**). The default selection is *H. sapiens*. All genes can be browsed by clicking on the “Show all” button. The navigation on top allows switching between different views and returning to the front search page. Gene sets can be displayed as a scrollable list and the complete information of an individual gene is given by selecting the “Card view”. A short explanation of each annotation field is given by clicking on the corresponding info button. Orthologous genes are connected via database hyperlinks. Furthermore, every gene is linked to selected publications.

Search and statistics

The *TelNet* database can be used with three different search modes (**Figure 3**) named “Quick search”, “List search” and “Advanced card search”. For a quick search throughout selected fields a keyword can be entered into the search bar. If a user wishes to constrain the found set to e.g. a gene name, the selection of fields can be adapted. By performing an advanced card search, the user can enter different search terms in corresponding fields. Furthermore, a complete list of gene identifiers can also be passed into the list search. The organism and identifier provided are mandatory to perform a list search. Genes found are then listed and can be selected for further analysis and TM network identification within *TelNet* or exported in various file formats. The statistics page gives a graphical overview over the distributions of various *TelNet* annotations such as a histogram of the *TelNet* score, and the distribution of TM significance. Furthermore, *TelNet* statistics can be employed for a more detailed pathway analysis regarding TM functions. A predicted wild-type TMM is computed by evaluation of TMM annotations. The wild-type phenotype of a given gene is used for predicting the likely active TMM for a set of genes. Every protein contributes with its *TelNet* score to one of the groups “ALT”, “telomerase-positive” or “ambiguous”, which refers

to its wild-type form. Thus, a gene that is recurrently mutated in ALT positive tumors like ATRX (alpha thalassemia/mental retardation syndrome X-linked protein) would represent an ALT suppressor. It is classified as “telomerase-positive” for the predicted TMM associated with its wild-type phenotype. The attribute “ambiguous” is used for genes lacking information on TM mechanism as well as genes with conflicting associations. Thus, *TelNet* informs about known and predicted TM features for the genes of interest via its different search and summary analysis tools.

Application of *TelNet* for telomere maintenance analysis

In order to illustrate an application of *TelNet*, we performed a pan-cancer correlation analysis of gene expression data with telomere length estimates. It is based on the study of Barthel et al. [39] and uses telomere length data calculated from whole genome sequencing (WGS) data from this reference and gene expression data (stdata_2016_01_28, file uncv2.mRNAseq_RSEM_normalized_log2) downloaded via the Firehose data repository (https://gdac.broadinstitute.org/). A reduced patient data set (n = 281) was compiled that comprised samples where normal control samples of matching tumor tissue were available. In order to normalize for tissue- and age-specific effects, we calculated the ratio of tumor over normal for telomere lengths and the log2 ratio for gene expression. For the two ratios the Spearman correlation coefficient was calculated. A significant correlation of telomere length with gene expression was observed for 87 genes (Rho > 0.186 or Rho < -0.184) and 940 genes were differentially expressed (log2 ratio > 0.852 or log2 ratio < -0.782) (**Figure 4**). It is noted that most of the tumor samples had shorter telomeres than the respective normal control sample. This could be the result of a higher tumor proliferation rate being only partly compensated by the active TMM. This confounding factor as well as the tissue specific expression programs in the different tumor entities are likely to lead to false negative results. For example, TERT (telomerase reverse transcriptase) expression did not show a significant correlation with telomere length. Thus, it might be informative to examine deregulated genes that did not display an (anti-)correlation with telomere length with respect to potential TM activities (**Supplementary Table 1)**.

To further analyze the genes were correlations were detected, we then used *TelNet* to identify TM links. The *TelNet* “List search” option (**Figure 3**) retrieved a set of 132 genes (12.9 %). The following results were obtained: (i) None of the genes with a value of the correlation coefficient above the significance thresholds (Rho > 0.186 or Rho < -0.184) was found to be significantly up- (log2 ratio > 0.852) or downregulated (log2 ratio > 0.782) in tumor samples over normal controls. (ii) ARL4D (ADP ribosylation factor-like GTPase 4D) was downregulated in tumors (log2 ratio = -1.02) and expression levels were positively correlated with telomere elongation (Rho = 0.22). Support for the association of ARL4D with

230 telomere length is found by examining the *TelNet* link to the yeast orthologue ARF1 (ADP
231 ribosylation factor 1), which shows shorter telomeres in the deletion mutant [27, 41].
232 However, the human tumor phenotype appears to be more complex since a downregulation
233 of ARL4D was found in the comparison with healthy tissue.
234 (iii) It appears that from the available pan-cancer data no strong candidate genes are
235 emerging that act as drivers for enhancing or repressing telomere elongation across different
236 tumor-types. Nevertheless, genes like SUMO3 (small ubiquitin-like modifier 3) and ERCC5
237 (excision repair 5 endonuclease) were identified via the *TelNet* analysis as proteins with
238 known TM activities and are among the most highly (anti-)correlated genes with respect to a
239 function in modulating telomere length in tumors (**Table 3**). SUMO3 is attached to key
240 proteins of the ALT pathway and shows a positive correlation with telomere length.
241 Sumoylation of PML and shelterin compounds is essential for the formation of PML nuclear
242 bodies [23, 56]. The ERCC5 endonuclease is involved in DNA recombination and repair by
243 annealing single-stranded DNA. Furthermore, ERCC5 positively interacts with the Werner
244 syndrome helicase (WRN) [57] that is involved in telomeric D-loop digestion in ALT cells
245 [58]. Accordingly, a further investigation of SUMO3 and ERCC5 with respect to their role for
246 ALT in tumor cells might be warranted.
247 In summary, *TelNet* provides a straightforward approach to the annotation of gene lists
248 obtained from different 'omics' approaches as illustrated above for WGS and RNA-seq data
249 to identify candidates involved in telomere maintenance for further analysis.
250

251 **Future development**

252 The *TelNet* database was developed to provide a comprehensive collection of TM relevant
253 genes. As it is an ongoing research project, new information on telomere maintenance will
254 be added as they become available. In addition, an extension of *TelNet* is planned to include
255 also genes from other organisms such as *M. musculus*. We encourage other researchers
256 working on telomeres to communicate their feedback via e-mail to telnet@dkfz.de to further
257 improve the information provided by our database.
258

259 Conclusions

260 The *TelNet* database compiles information on genes associated to telomeres and telomere
261 maintenance. The annotations provided by *TelNet* allow the distinction between different
262 types of TM mechanisms for a gene (set) of interest according to functional terms and a
263 significance ranking. With these features, *TelNet* supports the identification of TM networks
264 in various ways. A gene set derived from a preceding bioinformatics analysis pipeline can be
265 used for a *TelNet* list search to get more detailed insight on the corresponding TM

associated genes. Possible TM links can be explored in an iterative manner. For example, current pan-cancer studies provide a wealth of information on mutated or deregulated genes that can be evaluated in terms of the associated TM mechanisms [39]. Some associations like mutations in *ATRX* and *DAXX* (death-domain associated protein) for ALT as well as *TERT* promoter mutations for telomerase-positive cells are well established. However, in many tumor samples corresponding mutations are absent. One would expect that for these cases the mutation status of a given cancer sample and its active TM are linked via other genes that likely involves a combination of various factors. We anticipate that *TelNet* will prove to be a helpful analysis tool for revealing this type of correlations and to support the identification of active TM networks in different tumor entities.

Availability of data and material

Database homepage: http://www.cancertelsys.org/telnet/. These data are freely available without restrictions for use in academic research.

Abbreviations

ALT, alternative lengthening of telomeres; APB, ALT-associated PML nuclear bodies; TM, telomere maintenance; ARL4D, ADP ribosylation factor-like GTPase 4D; ATRX, alpha thalassemia/mental retardation syndrome X-linked protein; ERCC5, excision repair 5 endonuclease; PML, promyelocytic leukemia; TERT, telomerase reverse transcriptase; SUMO3, small ubiquitin-like modifier 3; TM, telomere maintenance; TMM, telomere maintenance mechanism.

Acknowledgements

The authors thank all present and former members of the CancerTelSys Consortium especially Brian Luke, André Maicher, Alexandra Poos, and Rainer König for discussion and data contribution. The authors are grateful to Floris Barthel for providing telomere length data and R scripts for performing the correlation analysis.

Funding

This work was supported by the project CancerTelSys [01ZX1302] in the e:Med program of the German Federal Ministry of Education and Research (BMBF).

Authors' contributions

DB developed the database and manually included and curated all entries with input from KR, IC and KID. NK was involved in converting the database to the final webversion. DB and KR wrote the manuscript with input from IC. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

References

1. de Lange T: Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes & Development* 2005, 19(18):2100-2110.
2. Blackburn EH, Epel ES, Lin J: Human telomere biology: A contributory and interactive factor in aging, disease risks, and protection. *Science* 2015, 350(6265):1193-1198.
3. Denchi EL, de Lange T: Protection of telomeres through independent control of ATM and ATR by TRF2 and POT1. *Nature* 2007, 448(7157):1068-1071.
4. Palm W, de Lange T: How shelterin protects mammalian telomeres. *Annu Rev Genet* 2008, 42:301-334.
5. Lazzerini-Denchi E, Sfeir A: Stop pulling my strings - what telomeres taught us about the DNA damage response. *Nat Rev Mol Cell Biol* 2016, 17(6):364-378.
6. Shay JW: Role of Telomeres and Telomerase in Aging and Cancer. *Cancer Discov* 2016, 6(6):584-593.
7. Heidenreich B, Rachakonda PS, Hemminki K, Kumar R: TERT promoter mutations in cancer development. *Curr Opin Genet Dev* 2014, 24:30-37.
8. Peifer M, Hertwig F, Roels F, Dreidax D, Gartlgruber M, Menon R, Kramer A, Roncaioli JL, Sand F, Heuckmann JM *et al*: Telomerase activation by genomic rearrangements in high-risk neuroblastoma. *Nature* 2015, 526(7575):700-704.

9. Valentijn LJ, Koster J, Zwijnenburg DA, Hasselt NE, van Sluis P, Volckmann R, van Noesel MM, George RE, Tytgat GA, Molenaar JJ *et al*: TERT rearrangements are frequent in neuroblastoma and identify aggressive tumors. *Nat Genet* 2015.
10. Cesare AJ, Reddel RR: Alternative lengthening of telomeres: models, mechanisms and implications. *Nat Rev Genet* 2010, 11(5):319-330.
11. Heaphy CM, de Wilde RF, Jiao Y, Klein AP, Edil BH, Shi C, Bettegowda C, Rodriguez FJ, Eberhart CG, Hebbar S *et al*: Altered telomeres in tumors with ATRX and DAXX mutations. *Science* 2011, 333(6041):425.
12. Dilley RL, Greenberg RA: ALTernative Telomere Maintenance and Cancer. *Trends Cancer* 2015, 1(2):145-156.
13. Sobinoff AP, Pickett HA: Alternative Lengthening of Telomeres: DNA Repair Pathways Converge. *Trends Genet* 2017, 33(12):921-932.
14. Dejardin J, Kingston RE: Purification of proteins associated with specific genomic Loci. *Cell* 2009, 136(1):175-186.
15. Grolimund L, Aeby E, Hamelin R, Armand F, Chiappe D, Moniatte M, Lingner J: A quantitative telomeric chromatin isolation protocol identifies different telomeric states. *Nat Commun* 2013, 4:2848.
16. Giannone RJ, McDonald HW, Hurst GB, Shen RF, Wang Y, Liu Y: The protein network surrounding the human telomere repeat binding factors TRF1, TRF2, and POT1. *PLoS ONE* 2010, 5(8):e12407.
17. Lee OH, Kim H, He Q, Baek HJ, Yang D, Chen LY, Liang J, Chae HK, Safari A, Liu D *et al*: Genome-wide YFP fluorescence complementation screen identifies new regulators for telomere signaling in human cells. *Mol Cell Proteomics* 2011, 10(2):M110 001628.
18. Cerone MA, Burgess DJ, Naceur-Lombardelli C, Lord CJ, Ashworth A: High-Throughput RNAi Screening Reveals Novel Regulators of Telomerase. *Cancer research* 2011, 71(9):3328-3340.
19. Ramlee MK, Wang J, Toh WX, Li S: Transcription Regulation of the Human Telomerase Reverse Transcriptase (hTERT) Gene. *Genes (Basel)* 2016, 7(8):50.
20. Yeager TR, Neumann AA, Englezou A, Huschtscha LI, Noble JR, Reddel RR: Telomerase-negative immortalized human cells contain a novel type of promyelocytic leukemia (PML) body. *Cancer Res* 1999, 59(17):4175-4179.
21. Nabetani A, Ishikawa F: Alternative lengthening of telomeres pathway: recombination-mediated telomere maintenance mechanism in human cells. *J Biochem* 2011, 149(1):5-14.
22. Chung I, Osterwald S, Deeg KI, Rippe K: PML body meets telomere: the beginning of an ALTernate ending? *Nucleus* 2012, 3(3):263-275.

369 23. Osterwald S, Deeg KI, Chung I, Parisotto D, Wörz S, Rohr K, Erfle H, Rippe K: PML
370 induces compaction, TRF2 depletion and DNA damage signaling at telomeres and
371 promotes their alternative lengthening. *J Cell Sci* 2015, 128(10):1887-1900.
372 24. Uziel O, Yosef N, Sharan R, Ruppin E, Kupiec M, Kushnir M, Beery E, Cohen-Diker T,
373 Nordenberg J, Lahav M: The effects of telomere shortening on cancer cells: a network
374 model of proteomic and microRNA analysis. *Genomics* 2015, 105(1):5-16.
375 25. Kupiec M: Biology of telomeres: lessons from budding yeast. *FEMS microbiology*
376 *reviews* 2014, 38(2):144-171.
377 26. Askree SH, Yehuda T, Smolikov S, Gurevich R, Hawk J, Coker C, Krauskopf A, Kupiec
378 M, McEachern MJ: A genome-wide screen for Saccharomyces cerevisiae deletion
379 mutants that affect telomere length. *Proceedings of the National Academy of Sciences*
380 *of the United States of America* 2004, 101(23):8658-8663.
381 27. Gatbonton T, Imbesi M, Nelson M, Akey JM, Ruderfer DM, Kruglyak L, Simon JA,
382 Bedalov A: Telomere length as a quantitative trait: genome-wide survey and genetic
383 mapping of telomere length-control genes in yeast. *PLoS Genet* 2006, 2(3):e35.
384 28. Ungar L, Yosef N, Sela Y, Sharan R, Ruppin E, Kupiec M: A genome-wide screen for
385 essential yeast genes that affect telomere length maintenance. *Nucleic Acids Research*
386 2009, 37(12):3840-3849.
387 29. Lippuner AD, Julou T, Barral Y: Budding yeast as a model organism to study the effects
388 of age. *Fems Microbiology Reviews* 2014, 38(2):300-325.
389 30. Lundblad V: Telomere maintenance without telomerase. *Oncogene* 2002, 21(4):522-
390 531.
391 31. Lundblad V, Blackburn EH: An alternative pathway for yeast telomere maintenance
392 rescues est1- senescence. *Cell* 1993, 73(2):347-360.
393 32. Chen Q, Ijpma A, Greider CW: Two survivor pathways that allow growth in the absence
394 of telomerase are generated by distinct telomere recombination events. *Molecular and*
395 *Cellular Biology* 2001, 21(5):1819-1827.
396 33. Teng SC, Zakian VA: Telomere-telomere recombination is an efficient bypass pathway
397 for telomere maintenance in Saccharomyces cerevisiae. *Molecular and Cellular Biology*
398 1999, 19(12):8083-8093.
399 34. Luo Z, Dai Z, Xie X, Feng X, Liu D, Songyang Z, Xiong Y: TeloPIN: a database of
400 telomeric proteins interaction network in mammalian cells. *Database (Oxford)* 2015,
401 2015(0):bav018-bav018.
402 35. Podlevsky JD, Bley CJ, Omana RV, Qi X, Chen JJ: The telomerase database. *Nucleic*
403 *Acids Res* 2008, 36(Database issue):D339-343.

404 36. Gowthaman R, Krishnamoorthy S, Nandakumar RD, Ayyarappan V: TeCK database: a
405 comprehensive collection of telomeric and centromeric sequences with their associated
406 proteins. *Bioinformation* 2007, 2(2):73-75.
407 37. Huang Z, Ma L, Wang Y, Pan Z, Ren J, Liu Z, Xue Y: MiCroKiTS 4.0: a database of
408 midbody, centrosome, kinetochore, telomere and spindle. *Nucleic Acids Res* 2015,
409 43(Database issue):D328-334.
410 38. Jiang WQ, Zhong ZH, Henson JD, Reddel RR: Identification of candidate alternative
411 lengthening of telomeres genes by methionine restriction and RNA interference.
412 *Oncogene* 2007, 26(32):4635-4647.
413 39. Barthel FP, Wei W, Tang M, Martinez-Ledesma E, Hu X, Amin SB, Akdemir KC, Seth S,
414 Song X, Wang Q *et al*: Systematic analysis of telomere length and somatic alterations in
415 31 cancer types. *Nat Genet* 2017:published online 30 January 2017.
416 40. Lovejoy CA, Li W, Reisenweber S, Thongthip S, Bruno J, de Lange T, De S, Petrini JH,
417 Sung PA, Jasin M *et al*: Loss of ATRX, genome instability, and an altered DNA damage
418 response are hallmarks of the alternative lengthening of telomeres pathway. *PLoS*
419 *Genet* 2012, 8(7):e1002772.
420 41. Hu Y, Tang HB, Liu NN, Tong XJ, Dang W, Duan YM, Fu XH, Zhang Y, Peng J, Meng
421 FL *et al*: Telomerase-null survivor screening identifies novel telomere recombination
422 regulators. *PLoS Genet* 2013, 9(1):e1003208.
423 42. Teixeira MC, Monteiro PT, Guerreiro JF, Goncalves JP, Mira NP, dos Santos SC,
424 Cabrito TR, Palma M, Costa C, Francisco AP *et al*: The YEASTRACT database: an
425 upgraded information system for the analysis of gene and genomic transcription
426 regulation in Saccharomyces cerevisiae. *Nucleic Acids Res* 2014, 42(Database
427 issue):D161-166.
428 43. Gene Ontology C: Gene Ontology Consortium: going forward. *Nucleic Acids Res* 2015,
429 43(Database issue):D1049-1056.
430 44. Klionsky DJ, Bruford EA, Cherry JM, Hodgkin J, Laulederkind SJ, Singer AG: In the
431 beginning there was babble. *Autophagy* 2012, 8(8):1165-1167.
432 45. Huang da W, Sherman BT, Lempicki RA: Systematic and integrative analysis of large
433 gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009, 4(1):44-57.
434 46. Kinsella RJ, Kahari A, Haider S, Zamora J, Proctor G, Spudich G, Almeida-King J,
435 Staines D, Derwent P, Kerhornou A *et al*: Ensembl BioMarts: a hub for data retrieval
436 across taxonomic space. *Database (Oxford)* 2011, 2011:bar030.
437 47. Brown GR, Hem V, Katz KS, Ovetsky M, Wallin C, Ermolaeva O, Tolstoy I, Tatusova T,
438 Pruitt KD, Maglott DR *et al*: Gene: a gene-centered information resource at NCBI.
439 *Nucleic Acids Res* 2015, 43(Database issue):D36-42.

440 48. Yates B, Braschi B, Gray KA, Seal RL, Tweedie S, Bruford EA: Genenames.org: the
441 HGNC and VGNC resources in 2017. *Nucleic Acids Res* 2017, 45(D1):D619-D625.
442 49. Aken BL, Achuthan P, Akanni W, Amode MR, Bernsdorff F, Bhai J, Billis K, Carvalho-
443 Silva D, Cummins C, Clapham P *et al*: Ensembl 2017. *Nucleic Acids Res* 2017,
444 45(D1):D635-D642.
445 50. Aken BL, Ayling S, Barrell D, Clarke L, Curwen V, Fairley S, Fernandez Banet J, Billis K,
446 Garcia Giron C, Hourlier T *et al*: The Ensembl gene annotation system. *Database*
447 *(Oxford)* 2016, 2016.
448 51. Cherry JM: The Saccharomyces Genome Database: A Tool for Discovery. *Cold Spring*
449 *Harb Protoc* 2015, 2015(12):pdb top083840.
450 52. The UniProt C: UniProt: the universal protein knowledgebase. *Nucleic Acids Res* 2017,
451 45(D1):D158-D169.
452 53. Balakrishnan R, Park J, Karra K, Hitz BC, Binkley G, Hong EL, Sullivan J, Micklem G,
453 Cherry JM: YeastMine--an integrated data warehouse for Saccharomyces cerevisiae
454 data as a multipurpose tool-kit. *Database (Oxford)* 2012, 2012:bar062.
455 54. Poux S, Gaudet P: Best Practices in Manual Annotation with the Gene Ontology.
456 *Methods Mol Biol* 2017, 1446:41-54.
457 55. Hong EL, Balakrishnan R, Dong Q, Christie KR, Park J, Binkley G, Costanzo MC,
458 Dwight SS, Engel SR, Fisk DG *et al*: Gene Ontology annotations at SGD: new data
459 sources and annotation methods. *Nucleic Acids Res* 2008, 36(Database issue):D577-
460 581.
461 56. Chung I, Leonhardt H, Rippe K: De novo assembly of a PML nuclear subcompartment
462 occurs through multiple pathways and induces telomere elongation. *J Cell Sci* 2011,
463 124(Pt 21):3603-3618.
464 57. Trego KS, Chernikova SB, Davalos AR, Perry JJ, Finger LD, Ng C, Tsai MS, Yannone
465 SM, Tainer JA, Campisi J *et al*: The DNA repair endonuclease XPG interacts directly
466 and functionally with the WRN helicase defective in Werner syndrome. *Cell Cycle* 2011,
467 10(12):1998-2007.
468 58. Opresko PL, Otterlei M, Graakjaer J, Bruheim P, Dawut L, Kolvraa S, May A, Seidman
469 MM, Bohr VA: The Werner syndrome helicase and exonuclease cooperate to resolve
470 telomeric D loops in a manner regulated by TRF1 and TRF2. *Mol Cell* 2004, 14(6):763-
471 774.
472

473 **Table 1. Screening studies included in *TelNet* for identification of TM genes**

Method	Organism / cell line	# genes / proteins	Ref.
Proteomics of isolated chromatin segments (PICh)	human / Wi38-VA13 HeLa 1.2.11	296	[14]
Quantitative telomeric chromatin isolation protocol (QTIP)	human / HeLa	34	[15]
Protein network analysis surrounding telomere repeat binding factors, TRF1, TRF2, and POT1 using dual-tag affinity purification in combination with multidimensional protein identification technology liquid chromatography - tandem mass spectrometry (MudPIT LC-MS/MS)	human / 293T	211	[16]
Protein complementation assay (PCA/bimolecular fluorescent complementation (BiFC)) of shelterin compounds and ~12.000 candidate genes; GST-pulldown of FLAG-tagged genes	human / HTC75	339	[17]
siRNA mediated knockdown; APB formation	human / U2OS	29	[23, 38]
Proteomic analysis of deregulated genes upon telomere shortening caused by the telomerase inhibitor GRN163	human / SK-N-MC	99	[24]
QTRAP: kinase library screen	human / HeLa	109	[18]
Telomerase regulators affecting its transcription	human / various	53	[19]
Gene set with potential relevance to telomeres and the ALT pathway	human / various	297	[40]
Telomerase activity signature	human	43	[39]
GO annotation containing the term ‘telo’	human	245	[43]
Haploid deletion screen, telomere length	yeast	166	[26]
Telomere length-variation screen in deletion strains	yeast	138	[27]
Screen of DAmP collection	yeast	77	[28]
Telomerase null screen of yeast mutants	yeast	270	[41]
Telomerase regulators	yeast	35	[42]
GO annotation containing the term ‘telo’	yeast	192	[43]

474

475 Selected screening studies and other sources with information on TM genes are shown. The
476 method used for the respective investigation as well as the organism and if available the
477 human cell line is indicated. Also, the number of genes, added from the given literature
478 source is listed.

479 **Table 2. Calculation of the *TelNet* score**

Condition	Score
Identification in a TM related screen (added up if present in multiple screens)	1
TM significance is 'validated'	3
If TM significance is 'predicted', the gene yields 5% of the orthologue's TelNet score	0-1
Cellular function contains the term 'telomere'	1.5
A TM function is assigned (added up for multiple functions)	0.5
If a TM function is 'shelterin' or 'telomerase'	0.5-2

480

481 The different criteria of evaluated and associated *points* are shown. The *TelNet* score is
482 calculated as the sum of the different entries and can reach a maximum value of 10 points.

483

484

485

486 **Table 3. *TelNet* genes (anti-)correlating with telomere length (p < 0.01) changes**
487 **without altered expression**

Gene	Correlation coefficient Rho	p-value (Rho)	log2 expr. ratio (tumor/ normal)	p-value (expr.)	TM information
PRPS2	0.25	4.7E-04	0.21	4.4E-01	
GATA3	0.23	1.1E-03	-0.26	3.6E-01	
ARL4D	0.22	1.8E-03	-1.02	8.6E-04	
PDK3	0.22	2.4E-03	0.21	4.5E-01	
FAM58A	0.20	4.2E-03	0.35	2.3E-01	
MRPL34	0.20	5.4E-03	-0.25	3.8E-01	
TSPYL5	0.19	7.6E-03	-0.63	3.5E-02	
SUMO3	0.19	7.8E-03	0.16	5.4E-01	ALT enhancing; Telomerase conflicting
GIGYF2	-0.19	6.6E-03	-0.08	7.2E-01	
PPM1D	-0.20	5.9E-03	-0.30	2.9E-01	
CTNNA1	-0.21	3.6E-03	-0.14	5.9E-01	
ERCC5	-0.21	3.1E-03	-0.18	5.0E-01	ALT conflicting

488

489 Significantly (anti-)correlated genes, that are included in the *TelNet* database are shown.
490 Correlation coefficients Rho and the log2 ratio of tumor over normal gene expression are
491 displayed with the respective p-values. Telomere maintenance information from the TMM
492 annotation fields are shown, if available.

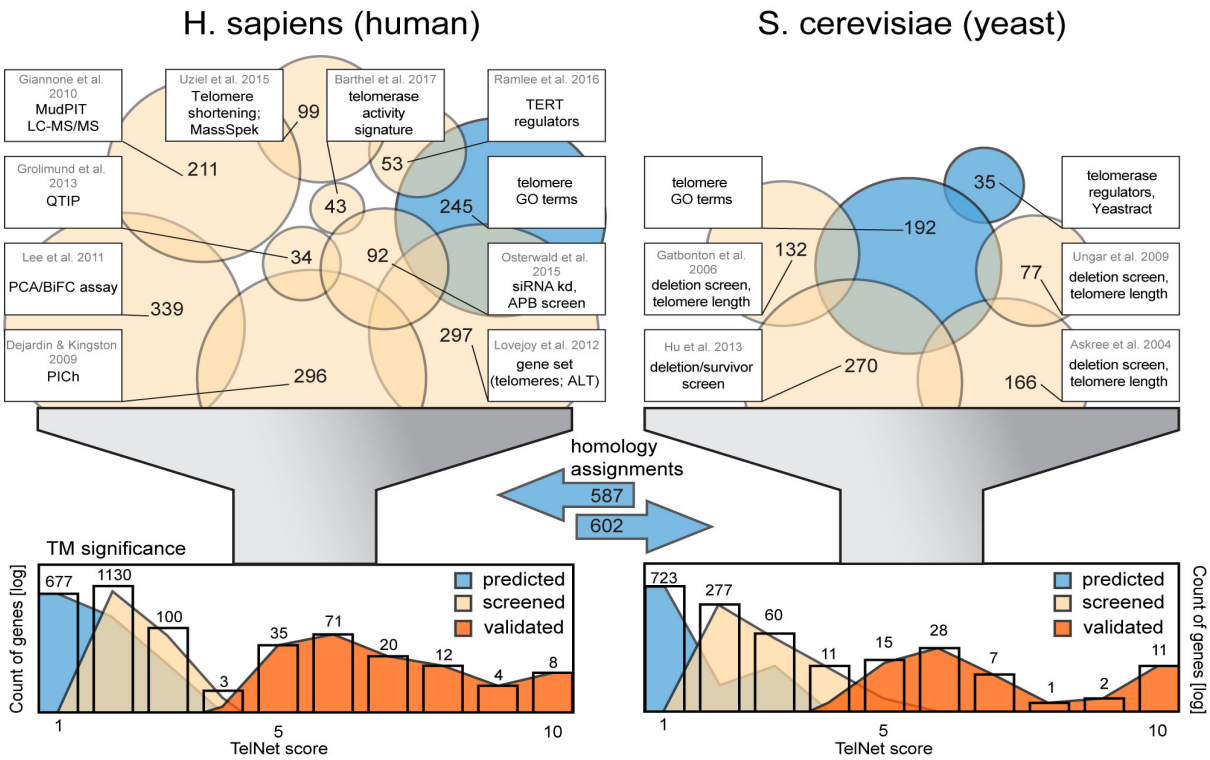

Figure 1. Data sources of TM genes for *TelNet*

Selected screening studies and other references that serve as sources for TM genes are indicated. In total over 2,000 human genes and more than 1,100 budding yeast genes were included in the *TelNet*. Histograms of the *TelNet* scores are displayed for the complete gene sets per species, colored by TM significance. Color scheme: blue, predicted TM genes; orange, genes from screening studies; red, validated genes.

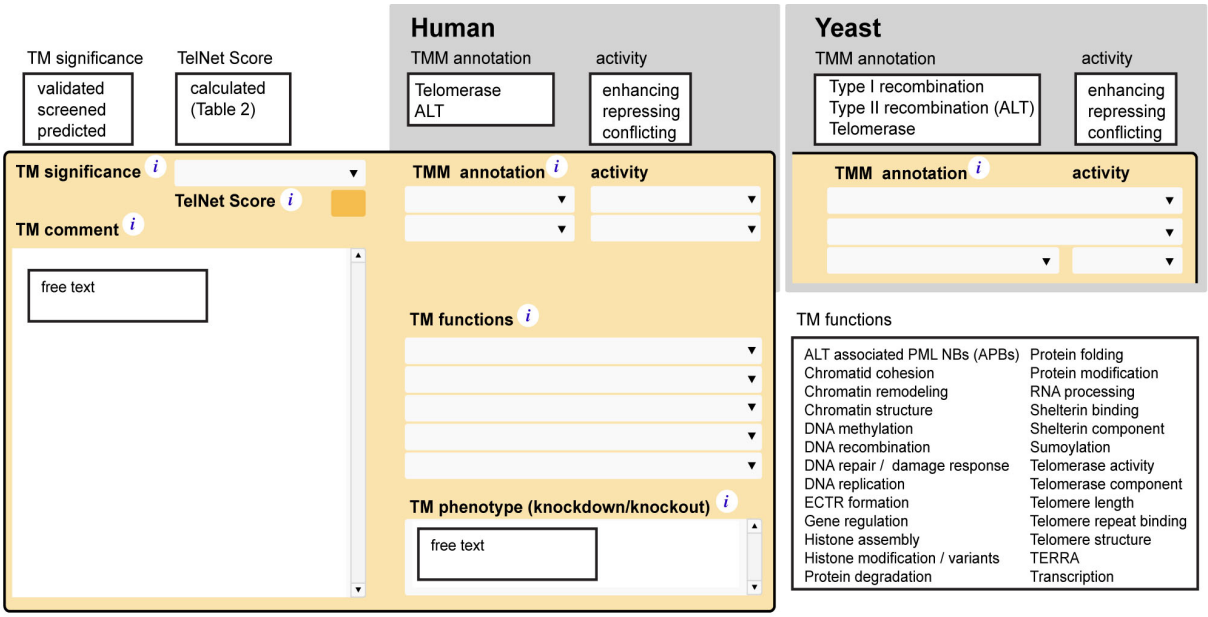

Figure 2. Gene card view of the telomere maintenance information

Annotation fields and possible entries for TM significance, associated *TelNet* score, TM comment, TMM annotation, TM functions and TM phenotype.

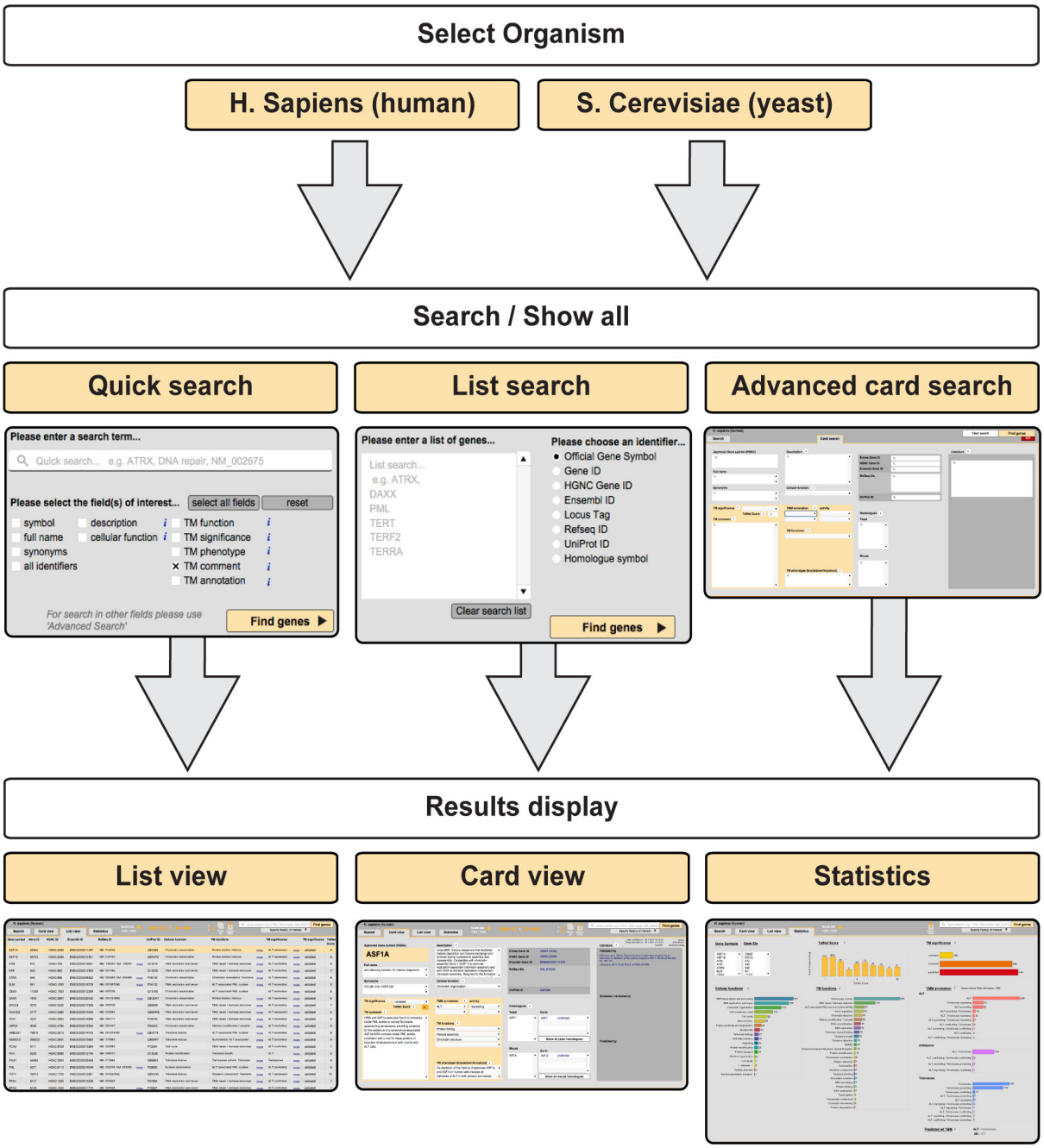

Figure 3. Typical *TelNet* workflow

Top: On the front page, the organism is selected. *Middle:* Three different kinds of search tools, namely “quick search”, “list search” and “advanced card search”, are available to retrieve a set of genes. *Bottom:* The resulting genes can be displayed as a scrollable list or as a series of single gene cards. An overview of the associated TM annotations is obtained from the statistics page.

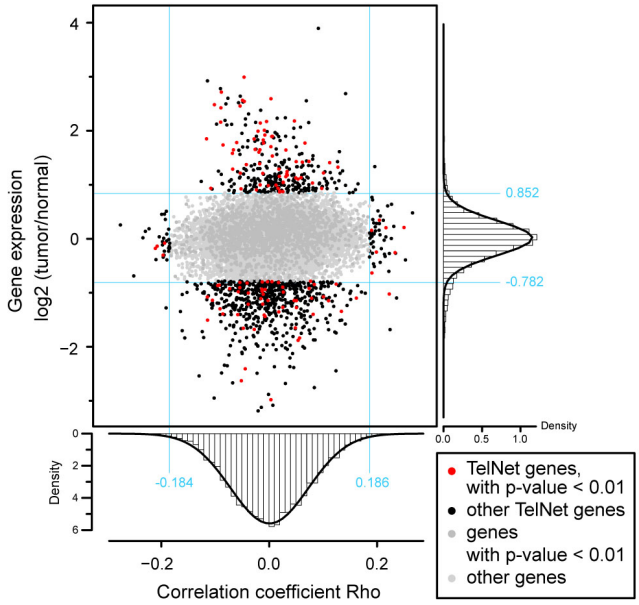

Figure 4. Application of *TelNet* to correlation analysis of telomere length and gene expression

Scatter plot showing the log2 ratio (tumor/normal) of gene expression versus the Spearman correlation coefficient Rho for gene expression and telomere length. For histograms of Rho and log2 ratio a Gaussian fit is shown with significance values defined from the 1%-tail of the fit. Genes that were either significantly ($p < 0.01$) up- (log2 ratio > 0.852) or downregulated (log2 ratio < -0.782) or significantly ($p < 0.01$) correlating (Rho > 0.186) or anti-correlating (Rho < -0.184) were colored in black. Genes above the significance thresholds that were present in the *TelNet* database are shown in red color.