American Society of Hematology
2021 L Street NW, Suite 900,
Washington, DC 20036
Phone: 202-776-0544 | Fax 202-776-0545
editorial@hematology.org

Telomere content and genomics of myeloid neoplasia by WGS

Tracking no: BLD-2025-028644R3

Luca Guarnera (University of Rome, Italy) adam wahida (Institute of Metabolism and Cell death, Helmholtz Zentrum München, Neuherberg, Germany, Germany) Carmelo Gurnari (Taussig Cancer Institute, Cleveland Clinic, United States) Stephan Hutter (MLL Munich Leukemia Laboratory, Germany) Sabine Stainczyk (Hopp Children's Cancer Center (KiTZ), Germany) Nakisha Williams (Taussig Cancer Institute, Cleveland Clinic, United States) Arda Durmaz (Department of Translational Hematology and Oncology Research, Taussig Cancer Institute, Cleveland Clinic, United States) Yasuo Kubota (The University of Tokyo Hospital, Japan) Carlos Bravo-Perez (Department of Translational Hematology and Oncology Research, Taussig Cancer Institute, Cleveland Clinic, United States) Naomi Kawashima (Taussig Cancer Institute, Cleveland Clinic, United States) Mark Orland (Department of Internal Medicine, Cleveland Clinic, United States) Simona Pagliuca (Centre Hospitalier Régional Universitaire de Nancy, France) Yimin Huang (Case Western Reserve University, United States) Thomas LaFramboise (Case Western Reserve University, United States) Valeria Visconte (Cleveland Clinic, United States) Wencke Walter (Munich Leukemia Laboratory, Germany) Manja Meggendorfer (MLL Munich Leukemia Laboratory, Germany) Wolfgang Kern (MLL Munich Leukemia Laboratory, Germany) Frank Westermann (Hopp Children's Cancer Center (KiTZ), Germany) Lars Feuerbach (Division of Applied Bioinformatics, German Cancer Research Center (DKFZ), German Cancer Consortium (DKTK), Germany) Torsten Haferlach (MLL Munich Leukemia Laboratory, Germany) Jaroslaw Maciejewski (Cleveland Clinic, United States)

Abstract:
Telomere length shortening has been associated with genomic instability and acquisition of molecular lesions, but these processes have not been systematically studied across large cohorts of myeloid neoplasia (MN). As proof of concept for a novel, cross-validated WGS-based method of telomere content (TC) determination combined with mutations, transcriptomics, and functional assays, we studied TC in correlation with specific molecular features of a large cohort (n=1804) of MN patients including acute myeloid leukemia (AML) and myelodysplastic syndrome. When compared to healthy subjects and patients with non-clonal diseases such as persistent polyclonal B cell lymphocytosis, both MN and non-malignant controls with clonal disease, such as paroxysmal nocturnal hemoglobinuria and aplastic anemia, exhibited decreased TC. Furthermore, we show that TC is lowered in adult MN abrogating correlation with age with considerable TC diversification among certain morphologic and molecular subtypes. For instance, AML harbored the lowest TC. Furthermore, MN originating from a more mature cell of origin (e.g., APL), and those characterized by hyperproliferative driver mutations (e.g., RAS pathway genes) had lower TC, possibly indicating a loss of telomere maintenance capacity. In contrast, MN subtypes arising in a context of profound genetic alterations, such as TP53 mutations and complex karyotype, exhibited a relatively higher/preserved TC compared to other mutations. This phenomenon did not involve alternative lengthening processes but was rather consistent with an increased TC due to preserved activity of the telomerase complex. Our results describe a common and genotype-specific telomeric make-up of a large cohort of patients with MN providing a molecular benchmark for future therapeutic targeting of the telomere machinery.

Conflict of interest: No COI declared

COI notes:

Preprint server: No;

Author contributions and disclosures: L.G. data curation and analysis, and wrote the manuscript. A.W. data curation and wrote the manuscript. C.G. helped in data collection and edited the manuscript. S.H. helped in data collection. S.A.S. helped in data collection. N.D.W. helped in data collection. A.D. helped in data analysis. Y.K. helped in data analysis. C.B-P. edited the manuscript. N.K. edited the manuscript. M.O. helped in data collection. S.P. edited the manuscript. Y.H. data collection and data analysis. T.L.F. data collection and data analysis, edited the manuscript. V.V. helped in data collection and edited the manuscript. W.W. data collection. M.M. data collection. W.K. data collection. F.W. data collection. L.F. data collection. T.H. helped in data collection, resources, and funding acquisition. J.P.M. contributed with resources, conceptualization, funding acquisition, and editing.

Non-author contributions and disclosures: No;

Agreement to Share Publication-Related Data and Data Sharing Statement: All data presented in the manuscript are available at: https://github.com/LucaGuarnera/Raw-data_Telomere-Project.git Requests for additional information not present in supplemental material and GitHub should be mailed to the corresponding author: maciejj@ccf.org.

Clinical trial registration information (if any):

Telomere content and genomics of myeloid neoplasia by WGS

Luca Guarnera (1,2) [#], Adam Wahida (3) [#], Carmelo Gurnari (1,2)[#], Stephan Hutter (4)[#], Sabine A. Stainczyk (5,6), Nakisha D. Williams (1), Arda Durmaz (1), Yasuo Kubota (1), Carlos Bravo-Perez (1,7), Naomi Kawashima (1), Mark Orland (1), Simona Pagliuca (1,8), Yimin Huang (9), Thomas LaFramboise (9), Valeria Visconte (1), Wencke Walter (4), Manja Meggendorfer (4), Wolfgang Kern (4), Frank Westermann (5,6), Lars Feuerbach (10), Torsten Haferlach (4)[§]*, Jaroslaw P. Maciejewski (1)[§]*.

(1) Translational Hematology & Oncology Research, Cleveland Clinic, Cleveland, USA
(2) Biomedicine and Prevention, University of Rome Tor Vergata, Rome, Italy
(3) Institute of Metabolism and Cell death, Helmholtz Zentrum München, Neuherberg, Germany
(4) MLL Munich Leukemia Laboratory, Munich, Germany
(5) Hopp Children's Cancer Center (KiTZ), Heidelberg, Germany
(6) German Cancer Research Center (DKFZ), Division of Neuroblastoma Genomics, Heidelberg, Germany
(7) Department of Hematology, Hospital Universitario Morales Meseguer, University of Murcia, IMIB-Pascual Parrilla, CIBERER—Instituto de Salud Carlos III, 30005 Murcia, Spain
(8) Hematology department, Nancy University Hospital & UMR 7365 CNRS University of Lorraine
(9) Department of Genetics and Genome Sciences, Case Western Reserve University, Cleveland, USA
(10) Division of Applied Bioinformatics, German Cancer Research Center (DKFZ), German Cancer Consortium (DKTK), 69120 Heidelberg, Germany

The symbols "#"and "*" indicate equal contribution.

[§]Correspondence: Prof. Torsten Haferlach
MLL Munich Leukemia Laboratory
Max-Lebsche-Platz 31, 81377 Munich, Germany
E-mail: torsten.haferlach@mll.com; Phone: +49 (0)89 99017-100; Fax: +49 (0)89 99017-109

[§]Correspondence: Prof. Jaroslaw P. Maciejewski
Translational Hematology & Oncology Research, Cleveland Clinic, Cleveland, USA
9620 Carnegie Ave n building, Building, NE6-314, Cleveland, OH, USA 44106
E-mail: maciejj@ccf.org; Phone: +1 216 952-5232

Data Sharing Statement: All data presented in the manuscript are available at: https://github.com/LucaGuarnera/Raw-data_Telomere-Project.git Requests for additional information not present in supplemental material and GitHub should be mailed to the corresponding author: maciejj@ccf.org.

Running title: *Telomere content in leukemia*
Abstract: 258/250
Text characters count: 4461/4500
Figures/Tables: 4/1
Supplementary Figures/Tables: 20/11
References: 41

Keywords: Telomeres, myeloid neoplasia, *TP53*.

Key points

- Telomere content (TC) in AML is not correlated with age and presents a broad variability according to cytogenetic and molecular features.
- Increased telomerase complex activity in *TP53*-mutated AML results in higher TC compared to other myeloid neoplasms with other mutations.

Abstract

Telomere length shortening has been associated with genomic instability and acquisition of molecular lesions, but these processes have not been systematically studied across large cohorts of myeloid neoplasia (MN). As proof of concept for a novel, cross-validated WGS-based method of telomere content (TC) determination combined with mutations, transcriptomics, and functional assays, we studied TC in correlation with specific molecular features of a large cohort (n=1804) of MN patients including acute myeloid leukemia (AML) and myelodysplastic syndrome. When compared to healthy subjects and patients with non-clonal diseases such as persistent polyclonal B cell lymphocytosis, both MN and non-malignant controls with clonal disease, such as paroxysmal nocturnal hemoglobinuria and aplastic anemia, exhibited decreased TC. Furthermore, we show that TC is lowered in adult MN abrogating correlation with age with considerable TC diversification among certain morphologic and molecular subtypes. For instance, AML harbored the lowest TC. Furthermore, MN originating from a more mature cell of origin (*e.g.*, APL), and those characterized by hyperproliferative driver mutations (*e.g., RAS* pathway genes) had lower TC, possibly indicating a loss of telomere maintenance capacity. In contrast, MN subtypes arising in a context of profound genetic alterations, such as *TP53* mutations and complex karyotype, exhibited a relatively higher/preserved TC compared to other mutations. This phenomenon did not involve alternative lengthening processes but was rather consistent with an increased TC due to preserved activity of the telomerase complex. Our results describe a common and genotype-specific telomeric make-up of a large cohort of patients with MN providing a molecular benchmark for future therapeutic targeting of the telomere machinery.

Introduction

Telomeres are noncoding repeats, majority being but not limited to *TTAGGG* nucleotide sequence, at chromosomal ends that prevent DNA degradation and genomic instability[1]. Increased telomerase activity is one of the essential features of hematopoietic stem cells (HSCs), representing the mechanism countering telomere attrition caused by repeated cell divisions[2]. Indeed, stem cells with short telomeres are more susceptible to cellular senescence, chromosomal instability, and malignant transformation[3,4].

While leukemia-initiating cells may share the ability of telomere maintenance with normal HSCs TC, the bulk of proliferative potential in myeloid neoplasia (MN) is likely secured by ontogenically more mature leukemic precursors lacking this capacity [5,6]. Therefore, excessive telomere shortening or defective telomeric repeats due to replicative stress may also represent a distinctive feature of MN, similar to the observations made in solid cancer genomes[1]. However, telomere machinery expression may vary between different MN subtypes, corresponding to differentiation arrest at various stages of myeloid development.

Albeit the threshold of TC may be highly individual, critical shortening has been associated with genomic instability and other features facilitating the stepwise acquisition of genomic lesions and a more aggressive phenotype[7,8]. Furthermore, inherited telomeropathy traits may be an additional contributing factor to excessive TC lowering, possibly influencing clinical outcomes[9,10]. Hypomorphic alleles of telomere machinery genes have been implicated in various constitutional bone marrow failure (BMF) syndromes as well as in aplastic anemia, including seemingly acquired cases[4]. In some instances, such alterations represent risk factors for leukemia and/or may contribute to accelerated telomeres shortening[4].

Significant differences in telomere length (TL) have been observed among healthy subjects as well. In a recent large study analyzing 462,666 UK Biobank participants WGS-based method of TC measurement has been extensively validated. Using this technology, differential impact of clonal hematopoiesis (CH) gene drivers on TL and *vice versa* was established[11]. The relationship between CH and telomere biology has been also investigated in a cohort of patients with telomere biology disorders. In this setting, CH was observed predominantly in symptomatic patients and was mainly driven by *PPM1D, POT1, TERTp, U2AF1* and *TP53* hits. Indeed, TP53 pathway emerged as a key player in malignant evolution of telomere biology disorders and its mutations predictors of poor prognosis.[12].

In this perspective, despite a thorough investigation of predisposal conditions, a large, systematic study of the relationship between telomere length/content and somatic genomic lesions processes in MN, including acute myeloid leukemia (AML), myelodysplastic syndrome (MDS) and related disorders, has not been conducted.

With the ability to better define somatic defects, it is more likely to find possible correlations between specific molecular alterations and TC. The availability of a new class of drugs targeting the telomerase machinery (*e.g.*,

imetelstat)[13,14] makes studies aiming to precisely estimate TC and its associations in MN very timely, whereas TC may represent a biomarker of resistance or sensitivity to these novel therapeutic agents.

Previously, Southern blot, PCR methods or flow-FISH (fluorescent in situ hybridization with flow cytometry) have been used to assess TC. Still, these methods have inherent technical drawbacks and limitations[15–17]. In addition to the multitude of avenues of research and clinical applications, WGS can also be adopted to measure TC using various bioinformatics pipelines[18,19]. We stipulate that modern sequencing platforms will be helpful to reliably measure TC and thus investigate the biological features associated with decreased TC. To this end, we inferred TC from WGS data, further cross-validating recently implemented methodologies[18,19], and created an experimental set-up allowing for a systematic study of TC in MN, including determining the presence of alternative telomeric repeats.

Taking advantage of a large, well-annotated cohort (n=1804) of patients with MN including AML, MDS, and myeloproliferative neoplasms (MPN), we studied TC, its correlations with gene mutations, clinical phenotypes and prognosis as well as the transcriptomic profile of genes related to the telomere machinery. Finally, we provided mechanistic insights into the relation between *TP53* mutations and telomere biology in MN.

Methods

Patients

Our study cohort (n=1804) included patients with AML (n=730), MDS (n=702) and MPN (n=372, including chronic myeloid leukemia, essential thrombocythemia, polycythemia vera, primary myelofibrosis). Patients with non-malignant clonal disorders such as paroxysmal nocturnal hemoglobinuria (PNH) and aplastic anemia (AA) (n=102, of whom 42 PNH, 40 AA, 20 AA/PNH), patients with persistent polyclonal B cell lymphocytosis (PPBL; n=50) and healthy subjects (n=6) served as controls.
DNA sequencing was performed on banked samples. The review of medical records was approved by the internal Institutional Review Board involved in this study in agreement with the Declaration of Helsinki. Pertinent clinical data including age, gender, diagnosis according to WHO 2016 classification of myeloid disorders[20], cytogenetics, and other clinical parameters were collected. Specimens were obtained in accordance with ethical committee approvals of participating institutions.

Genomic studies

Whole genome sequencing

Libraries for WGS were generated from 1µg of DNA using the TruSeq PCR-Free prep kit, following the manufacturer's recommendations (Illumina, San Diego, CA, USA) and sequenced on NovaSeq6000 or HiSeqX

Illumina instruments following a 2x150bp paired-end reads standard protocol at a mean depth of coverage of >100x.

Sequencing and telomere content estimation

Sequenced reads were mapped to the human hg19/GRCh37 reference genome within Illumina's BaseSpace Sequence Hub infrastructure using Isaac (DNA alignment software) 3.16[21]. Resulting BAM files served as input for estimating TC with TelSeq 0.0.2[19] and TelomereHunter 1.1.0[18]. The two tools were also used to retrieve TC characteristics and telomere repeat heterogeneity via singleton analysis. Results were validated via TC-based methods (See supplementary methods). TC adjustment according to karyotypic anomalies, assessment of chromothriptic events and TC validation by PCR are described in supplementary.

Variant calls and annotations

Variants were annotated using Annovar and had a minimum VAF of 2%, average of >500X coverage and >98% of targeted regions showed over 100X coverage. Variants were classified according to established guidelines[22], and internal bio-analytic pipelines to identify somatic/germline mutations using sequences derived from controls and mutational databases such as dbSNP138, 1000 Genomes or ESP 6500 database, and Exome Aggregation Consortium (ExAC)[23]. In case of mutations not reported in these databases, literature reports and in-silico predictions were used to assess the pathogenicity of the variants.

C-Circle assay

C-Circle assays were performed as described previously[24]. The analysis is based on the detection of the C-Circle, a specific marker of telomere alternative lengthening (ALT). C-Circles are self-primed and present a telomeric G strand GGGATT and a telomeric C strand CCCTAA. After the DNA extraction, the addition of phi29 DNA polymerase, deoxynucleoside triphosphates (with the exception of dCTP and with no primers) will cause an ALT-specific rolling circle amplification (RCA). The RCA product can then be detected by specific probes. For every sample, 30 ng of genomic DNA were used in the assay. A control without polymerase was processed in parallel for every sample. A positive and negative assay control was included in every blot. Products of the rolling circle were blotted onto a nylon membrane using a slot blotter (BioRad). Blots were hybridized with a telomere probe and signals were developed using the TeloTTAGGG kit (Roche) according to the manufacturers' instructions.

RNA-sequencing

RNA-Seq was performed as previously described by our group[25]. Total RNA (250 ng) from BM specimens was prepared using the Illumina TruSeq Total Stranded RNA library preparation kit. Paired-end reads of 100bp were sequenced with a median depth of 50 million reads/sample. Sequences were aligned to human reference

genome hg19 using STAR 2.5.0 where gene counts were estimated using Cufflinks (v2.2.1)[26]. Expression level was expressed as log2CPM. *p* values were adjusted for multiple hypothesis testing with Bonferroni correction (*Q* values ≤ 0.05 and inclusion level difference ≥5% were considered significant).

A complete description of the methods can be found in Supplemental Material.

Results

TC measurement via WGS – validation and impact on MN features

To comprehensively investigate the impact of changes in TC in MN, we set up an integrated workflow merging transcriptomics and genomics (**Supplementary Figure 1**). We first used two WGS-based methods (TelomereHunter[18] and TelSeq[19]) to independently estimate TC. While the results were grossly comparable (correlation coefficient r^2= 0.3448, p<0.0001) we then selected TelomereHunter for all subsequent analyses given the higher rate of informative TC results using this method. Indeed, TelSeq was unable to produce informative results in 86 cases (4.6% of the population, for whom close to "0" value was rendered) but these cases were easily resolved by TelomereHunter. For the cases that were resolved by both software packages there was a good correlation coefficient of r^2=0.5650, p<0.0001 (**Figure 1A**).

Thus, because of the better performance and the ability to quantify alternative telomere variant repeats, we decided to proceed with the subsequent analyses using readout only from this tool[18,27]. We then studied a cohort of patients including AML (n=730, both primary [n=669] and secondary AML [n=61]), MDS (n= 702), MPN (n=372) and non-malignant controls (n=158), including, a clonal comparator group of PNH and AA (n=102), a non-clonal comparator group of PPBL (n=50) and healthy subjects (n=6) (**Table 1**) with the goal of analyzing the landscape of telomeric footprints across MN. Given the high percentage of aneuploidy among AML and MDS patients (for subsequent analysis referred together as MN), we examined the potential impact of these karyotype abnormalities on TC measurement to assess whether correction factor had to be introduced. Following adjustment according to type and allele frequency of aneuploidies, corrected output and standard TC showed a strong correlation for the overall MN population (p<0.0001, r^2= 0.9918; **Figure 1B**), patients carrying abnormal karyotype (p<0.0001, r^2= 0.97969; **Supplementary figure 2A**), complex karyotype (CK, considered as ≥3 karyotype abnormalities; p<0.0001, r^2= 0.9581. **Supplementary figure 2B**) and the most frequent karyotype aneuploidies (**Supplementary figure 2C-2G**). Accordingly, for subsequent analysis, TC as a surrogate measure for TL was used without adjustments. In addition, we also cross-validated WGS-

based TC analysis with RT-PCR-based TC in 13 matched samples from non-malignant controls, showing a significant correlation of the results obtained via the two methods (p=0.045, r^2= 0.3156; **Figure 1C**).

To gain insight into TC-age relationship, we compared our non-clonal controls (PPBL patients and healthy subjects), whose TC, plotted, reproduced the canonical sigmoid curves of TC gradual erosion over time with MN patients (**Figure 1D**), clonal controls (PNH and AA patients; **Figure 1E**), MDS (**Figure 1F**) and AML (**Figure 1G**) cohorts. Both clonal controls and MN patients showed a decreased TC, even though differences between MDS and AML were noted (**Supplementary table 1**). Furthermore, we observed no age dependent TC erosion in MN cohort (**Supplementary figure 3A**), while blast percentage showed a significant but weak correlation with telomere shortening in both AML and MDS cohorts (p<0.0001, r^2=0.0231 and p=0.0346, r^2=0.0063, respectively; **Supplementary figure 3B-C**). In addition, when we grouped AML by blast morphologic features, a higher median TC in the M6 (erythroblastic) subtype was found when compared to the M0, M1, M3, M4, and M5 groups (**Supplementary figure 4**).

TC-dependent clustering of genetic abnormalities in MN

When TC of MN patients grouped according to various karyotypic abnormalities was analyzed, we detected higher TC in patients with del(5q) (p< 0.0001) and lower TC in patients with t(6;9) (p=0.0002), t(8;21) (p< 0.0001), inv(16) (p=0.0003), t(9;11) (p< 0.0001), t(15;17) (p< 0.0001) when compared to patients with normal karyotype (NK) **(Figure 2A, supplementary table 2; supplementary figure 5**).

The analysis of TC among the whole MN population according to genetic landscape revealed higher TC in patients harboring *ETV6* (p=0.0117), *SF3B1* (p=0.0005), *SRSF2* (p<0.0001) and *TP53* (p=0.0004) mutations and lower TC in patients bearing *FLT3* (p<0.0001), *KRAS* (p<0.0001), *NPM1* (p<0.0001), *NRAS* (p=0.0002) and *WT1* (p<0.0001) mutations when compared to wt counterparts (**Figure 2A; supplementary table 3; supplementary Figure 6**).

In AML cohort, CK (p< 0.0001), inv3/t(3;3) (p=0.0048) and del(5q) (p< 0.0001) cases had higher TC, whereas t(6;9) (p=0.0003), t(8;21) (p=0.0132), t(9;11) (p=0.0099), t(15;17) (p=0008) and tr(1) (p=0.0203) presented lower TC when compared to patients with NK. Furthermore, patients carrying *CEBPA* (p=0.0032), *SF3B1* (p=0.0008), *SRSF2* (p=0.0157) and *TP53* (p<0.0001) had higher TC when compared to wt counterparts (**Supplementary figure 7 A-B**).

In MDS, patients with CK exhibited lower TC when compared to patients with NK (p<0.0001). Moreover, patients with *TP53* mutations had lower TC (p<0.0001), and those harboring *SRSF2* mutations presented higher TC (p=0.0051) when compared to wt counterparts (**Supplementary figure 7 C-D**).

We then performed a multivariate analysis to identify variables predicting TC. While neither the diagnosis (AML vs MDS), nor the age or the karyotype (abnormal vs normal) correlated to TC, we detected the impact of blast percentage (p<0.0001) and *TP53* mutations (p=0.001) (**Figure 2B; supplementary table 4**). The influence of *TP53* mutation on TC was confirmed even considering CK as a variable in the model (**Supplementary table 5**). Considering the functional relationship between *TP53* and *PPM1D*, we then investigated TC in patients harboring *PPM1D* mutations and unmutated for *TP53.* Unlike *TP53*, no significant differences in TC were found when compared TC among *PPM1D* mutants (**Supplementary figure 8**).

To investigate genetic signatures of diverse TC in MN, we analyzed the somatic karyotype abnormalities and mutations in MN presenting a TC over the 90th and below the 10th percentile (**Figure 2C-E**). The first (high TC) cohort, mainly represented by low-risk MDS, demonstrated an enrichment in -7/del(7q), inv(3)/t(3;3) and CK abnormalities as well as *ASXL1*, *SRSF2* and *TP53* mutations. The second (low TC) was mainly represented by AML, with an enrichment in t(15;17) and t(8;21) abnormalities and *KRAS, NPM1, NRAS* and *WT1* mutations. Finally, to investigate potential inter-dependency between somatic mutations and karyotype abnormalities correlating with TC and potential biases due to chromothripsis events, we run another multivariate model, which confirmed the role as independent higher TC predictors of *TP53* and *SRSF2* mutations and chromothripsis. Independent predictors of lower TC were percentage of blasts, t(15;17) and t(8;21) translocations and *KRAS*, *NPM1* and *WT1* hits (**Supplementary table 6**).

When we stratified the chromothriptic lesions according to *TP53* mutational status, we saw a consistent increase in chromothripsis according to the size of the clone expressed as VAF of *TP53* mutation, as shown in **Supplementary Figure 9**. Furthermore, in the *TP53* mutant cohort, chromothripsis events were more common in cases with CK (76% vs 56%, p= 0.0189) and correlated with higher TC (**Supplementary Figure 10**).

Given their prognostic relevance and the tight correlation of *TP53* mutations with TC independently of karyotype abnormalities, genetic mutations, clinical and demographic features and chromothripsis events (**Figure 2B; supplementary table 4-6**), we performed a more in-depth TC analysis of *TP53*-mutated MN. First, we investigated the allelic status of *TP53* alterations grouping the cases according to our recently proposed scheme for *TP53* mutational configuration[28] (**Figure 2F**). When TC was ranked, the distribution of TC across *TP53* wt/mut populations showed difference (**Figure 2G**): a proportionally higher presence of *TP53*-mutated cases vs *TP53* wild type (wt) was seen in patients with TC>90th (14.4% vs 9.5%, respectively; p=0.06). Furthermore, although not quite significant (p=0.231), allelic status did seem to affect TC in the overall cohort of *TP53*-mutated patients (16% probably monoallelic, 23% probably biallelic, 60% biallelic) vs *TP53*-mutated patients presenting high TC (above 90th) (0% probably monoallelic, 25% probably biallelic, 75% biallelic) (**Figure 2H**).

Grouping patients according to *TP53* mutational configuration and TC allowed for construction of KM curves. *TP53*-mutant cases presented an inferior overall survival (OS) both in high TC (above 90th percentile) and low TC (below 90th percentile) patients (**Figure 2I**). KM curves showed a significant separation when MN were grouped according to TC ranks (**Supplementary figure 11**). However, TC did not retain independent prognostic impact, when accounting for strong and well-established patient- (age, gender) and disease-related variables (*TP53* mutations, blast percentage, cytogenetics) in both AML/MDS together and separately (**Supplementary tables 7-9; Supplementary Figures 12-14)**. *TP53* mutations conferred different prognosis according to the allelic state of the hits, being biallelic the one characterized by worse outcome. (**Supplementary figure 15A**).

When we focused on the characteristics of *TP53* mutations and their correlations with TC, biallelic *TP53*-mutated MNs demonstrated a significantly increased TC when compared to cases with *TP53* wt and "likely monoallelic" *TP53*-mutated MNs (**Figure 2J**). In contrast, the clonal burden did not correlate with TC (**Supplementary figure 15A**). Of note is that, analyzing the phenotype of *TP53*-mutated MN according to allelic distribution, we observed a significant enrichment of AML in biallelic mutations and MDS in probably monoallelic ones (p<0.001; **supplementary figure 15B**), which thus explain the correlation between *TP53* mutations and high TC in AML and low TC in MDS (**supplementary figure 7**). Patients harboring insertion/deletion mutations showed high TC, and, among missense mutations, the highest TC was observed in cases carrying R175H (**Supplementary Figure 15C-D**). Finally, when divided according to the functional impact of *TP53* variants (according to Dutta et al.[29]), we found no differences in terms of TC between disruptive mutations and non-disruptive mutations (**Supplementary figure 15E**).

An increased telomerase activity but not alternative lengthening mechanisms defines TP53-mutated MN

Pursuant the higher TC in *TP53* mutant MN, we then investigated the possible mechanisms of telomere elongation in high TC cases. We first compared TC measured in blasts and matched lymphocytes in AML patients and observed a statistically significant difference in blast/lymphocyte ratio in samples with blasts characterized by low TC (below 10th percentile) vs high TC (over 90th percentile) (p=0.0034, **Figure 3A**). The uncoupling of blast and lymphocyte TC was also confirmed by PCR, with blast/lymphocyte ratios ranging from 0.02 to 2.45 (**Figure 3B**).

We then addressed the issue of telomere elongation in *TP53*-mutated MN by comparing the enrichment of singletons (a specific type of telomere repeats that can be identified by the TelomereHunter software) in our population grouped by *TP53* mutational status and TC. *TP53*-mutated MN presented an increase in canonical telomere repeat variants (chiefly TTTGGG, p<0.0001), arguing against the involvement of alternative

lengthening of telomeres (ALT), a process characterized by a different singletons signature[1] (**Supplementary figure 16**). This observation was confirmed by comparing high TC *TP53*-Mutated MN vs high TC *TP53* wt MNs (p=0.0141, **Figure 3C**). In support of these notions, we analyzed the presence of germline mutations in genes involved in telomere machinery (**Supplementary table 10**), which could be responsible for alternative TC sequences and ALT phenotypes (e.g. *DAXX* or *ATRX*-mediated ALT, commonly observed in solid cancers[30,31]). A total of 18 pathogenic/likely pathogenic variants were identified (**Supplementary table 11**), of which 2 co-occurred with *TP53* mutation (*ATRX*, c.6332G>C and *TERT*, c.1807del). However, these mutations did not show a correlation with TC (**Supplementary figure 17**).

We also studied the association of *TP53* hits with ALT as a possible telomere maintenance mechanism in MN. C-Circle assays[30] were performed for representative samples with high and low TC, both in *TP53*-mutant and wt samples (**Figure 3D-E; Supplementary figure 18-20, Supplementary tables 12-18**). In line with our previous results, none of the subgroups exhibited significant signs of ALT. Once the lack of contribution of ALT to telomere extension was established, we sought to determine whether increased TC may be due to higher telomerase activity imprinted by *TP53* mutations.

Differently expressed telomere machinery genes in long telomeres-MN

Thus, we analyzed the differential expression of telomere machinery genes in MNs. In cases with increased TC MNs (above 90th percentile), increased expression of *RAP1A*, *TERC*, *TINF2*, *TPP1*, *CTC1* and a downregulation of *TERF1* was detected (**Figure 4A**). In *TP53*-mutated MN, *DKC1*, *TERT* and *TERC* were significantly overexpressed as compared to wt cases, while *TEP1* was downregulated. Furthermore, the expression of *DKC1* significantly correlated with high TC whereas *TEP1* correlated with low TC (**Figure 4B-E**). We observed no difference in *TP53* expression between high- and low- TC MNs (**Supplementary figure 21**). Of note, when stratified according to *TP53* allelic status, we observed no differences in terms of expression of telomerase machinery genes with the exception of *TINF2*, which was downregulated in patients harboring biallelic mutations (**Supplementary figure 22)**. To sum up, these findings may suggest a role of telomerase machinery genes in TP53-mediated TC elongation. Functional studies to correlate the expression of telomerase machinery genes and telomerase activity are warranted for further validation but could not be performed here due to the lack of material.

Discussion

The increased use of NGS approaches, as well as the availability of more efficient techniques, led to a broad application of comprehensive whole genome and transcriptomic sequencing for both research and diagnostic purposes[32,33]. Our study demonstrates another useful application of WGS, namely for the determination of TC, which can potentially complement complex data sets that include copy number changes, somatic, and germline alterations, among others. In this study, we applied WGS to a large cohort of patients with MN to

explore TC dynamics, its associations with clinical and molecular phenotypes, and determinants of TC. Using this bioinformatic framework, extensively validated in recently published large cohorts which applied analogous bioanalytic strategy[11], we found that TC lowering represents an overarching feature of hematologic disorders of the myeloid compartment, inherent to their clonal nature and accounting for age factor. Furthermore, we showed that significantly lower TC is inherent to excessive proliferation of leukemic progenitor cells as a much stronger factor than aging. Most importantly, specific behaviors of different cell types, such as the permanent self-renewal capability of HSCs in the bone marrow, could skew telomere shortening. This finding is consistent with the difference in the telomere biology of progenitors rather than HSCs. For instance, in AML, the bulk leukemic cells may lack the ability to extend telomeres due to an excessive proliferative drive and clonal outgrowth from a single initiating cell. According to this notion, our results suggest that a bulk leukemic population occurs at the level of progenitor cells [*e.g*., promyelocytes in AML harboring t(15;17)] which lose the telomere elongation activity, whereas the one characterized by a profound genomic alteration (*e.g*., harboring *TP53* mutation or CK) maintain it. In addition, exuberant blast proliferation *e.g.* with *NRAS, KRAS* mutations) might exceed the compensatory capacity of normal telomere lengthening.

As confirmation of the importance of disease's features on telomere characteristics, TC was correlated to blast percentage and *TP53* mutation irrespective of age. Furthermore, TC correlated with the allelic status of the mutation (biallelic/probable biallelic vs probable monoallelic and wt).

Large studies focused on the relationship between TC and *TP53* mutations: Gutierrez-Rodriguez et al. analyzed a case series of patients with telomere biology disorders (characterized by short or very short TL) and focused on somatic rescue and risk of secondary malignancies, detecting an enrichment in *TP53* and *PPM1D* mutations[34]. Burren et al. analyzed, within UK Biobank individuals, a large population of healthy subjects with CHIP via WGS, showing a correlation between *TP53* mutation and shortened TL[11]. In contrast, Nakao et al., using the UK Biobank and TOPMed databases, performed bidirectional Mendelian randomization studies and demonstrated that longer TL increased propensity to develop CHIP, while CHIP, in turn shows enhanced TL shortening[35]. However, in contrast to our analysis, in these studies TC was determined in non-clonal cells, constituting most of the cells contributing DNA, rather than at that point tiny *TP53* mutant clones. Indeed, studies on MN produced contrasting evidence: Myllymäki and colleagues, in a cohort of 1267 MDS patients (262 harboring *TP53* mutation), did not find a significant correlation between *TP53* mutation and TC (*TP53* mutations were enriched in the first quartiles of patients stratified for TC without, though, reaching the significance threshold)[8]. In agreement with our results in a much larger and diverse cohort of patients, Abel and colleagues found a correlation between *TP53* mutation and higher TC analyzing 42 *TP53*-mutated MDS/AML patients compared to a cohort of patients with core-binding AML[36].

Intriguingly, our data would suggest that the lengthening mechanisms would act independently of ALT. Multiple lines of evidence support this notion: i) our in-depth singleton repeat analysis, ii) the paucity of mutations in telomere machinery genes potentially responsible for ALT phenotype (*e.g.*, *ATRX* or *DAXX*), and iii) the results from C-Circle assays.

While the use of bulk WGS for TC adds another useful application of this technology, there are several limitations. The measures obtained correspond to the average length across various cell types both clonal and non-clonal. In cases with relatively small clonal population, the TL corresponds to the normal polyclonal hematopoiesis, while is highly clonal conditions the measurements will be reflective of neoplastic cells and likely decoupled from the chronologic telomere attrition. Nevertheless, once WGS is obtained there may be many useful applications of TC determination in both malignant and non-clonal conditions (*e.g.*, hereditary telomeropathies).

The implications of our study extend beyond the realm of telomere biology and leukemia pathophysiology. Recently, therapies targeting the telomere machinery, *e.g.*, the 13-mer inhibitor imetelstat, have shown encouraging response rates in various myeloproliferative entities and MDS[14,37–39]. Our findings allow us to hypothesize that the inhibition of telomerase in patients with shorter telomeres (not-*TP53*-mutated MN) may lead to their senescence and apoptosis while cases with high TC have a larger “telomeric reserve”, fostering drug resistance. Our hypothesis is supported by the observation of higher expression of *TERC* in *TP53*-mutated cases (**Figure 4C**), which may cause a reduction of inhibition via imetelstat, and thus less susceptibility of these patients to the drug irrespective of the TC. Several ongoing studies are currently investigating the use of imetelstat in other MN. We stipulate that TC could further help stratify the patients most amenable to be treated with such approaches and guide therapeutic strategy.

It is important to acknowledge that long-read WGS enabling quantitation of neotelomeres may add another important aspect of TC quantitated in our study by short read NGS[40,41]. Indeed, neotelomeres at sites of double-strand DNA breaks (DBSs) potentially enable escape from bridge-fusion-breakage cycles. As neotelomeres may have similar length to normal chromosomal caps and chromothripsis can involve up to several hundred DSBs[40,41], thus the resultant neotelomeres may significantly contribute to the total TC. Short-read NGS used in our study precluded exploration of this mechanism; unfortunately, even newer bioinformatic tools such as TelFuse cannot resolve neotelomere length from short read NGS input. This shortcoming of our results will be mitigated by future long-read WGS applications in MN.

Additional limitations of the study include the lack of investigation of intrachromosomal telomerase repeat insertions. These stretches of telomeric repeats were observed in about 1/3 of *TP53*-mutant cases in the study of Abel and colleagues[36] and might potentially contribute to enhanced TC.

In summary, our findings show (1) a genomic rationale for the use of WGS to investigate TC; (2) the fundamental discovery that telomere shortening is common in MN; (3) *TP53* mutations being characterized by increased TC; (4) the exploitation of TC to identify patient subgroups eligible for imetelstat treatment. We also stipulate that TC measurements using WGS will help investigate pathophysiological features associated with TC lowering, contributing to our understanding of the mechanisms of TC maintenance in leukemic genome. The availability of therapies targeting telomere machinery may offer an opportunity for personalized therapy beyond MPN, their current indication. It remains to be tested whether high TC associated with *TP53* mutations can serve as a marker of sensitivity or resistance to these agents. The findings described herein will have to be further clarified on a more mechanistic level to identify the contributors to the lower-than-expected TC attrition in *TP53* mutant MN.

Acknowledgements

We thank The Leukemia & Lymphoma Society TRP Award 6645-22 (to J.P.M.) and R35 R35HL135795 (to J.P.M.). C.B-P. has a postdoctoral fellowship from Instituto de Salud Carlos III (JR22/00041). We thank The Torsten Haferlach Leukämiediagnostik Stiftung to support this work. This work was supported by a grant from the Edward P. Evans Foundation (to C.G.).

We thank prof. Rodrigo T. Calado for the invaluable insights.

Author contributions

L.G. data curation and analysis and wrote the manuscript. A.W. data curation and wrote the manuscript. C.G. helped with data collection and edited the manuscript. S.H. helped in data collection. S.A.S. helped in data collection. N.D.W. helped with data collection. A.D. helped with data analysis. Y.K. helped with data analysis. C.B-P. edited the manuscript. N.K. edited the manuscript. M.O. helped with data collection. S.P. edited the manuscript. Y.H. data collection and data analysis. T.L.F. data collection and data analysis, edited the manuscript. V.V. helped with data collection and edited the manuscript. W.W. data collection. M.M. data collection. W.K. data collection. F.W. data collection. L.F. data collection. T.H. helped in data collection, resources, and funding acquisition. J.P.M. contributed with resources, conceptualization, funding acquisition, and editing.

Conflict of Interest Statement

The authors declare no competing interests.

References

1. Sieverling, L. *et al.* Genomic footprints of activated telomere maintenance mechanisms in cancer. *Nat Commun* **11**, 733 (2020).
2. Allsopp, R. C., Morin, G. B., DePinho, R., Harley, C. B. & Weissman, I. L. Telomerase is required to slow telomere shortening and extend replicative lifespan of HSCs during serial transplantation. *Blood* **102**, 517–520 (2003).
3. Townsley, D. M., Dumitriu, B. & Young, N. S. Bone marrow failure and the telomeropathies. *Blood* **124**, 2775–2783 (2014).
4. Calado, R. T. & Young, N. S. Telomere diseases. *N Engl J Med* **361**, 2353–2365 (2009).
5. Trumpp, A. & Haas, S. Cancer stem cells: The adventurous journey from hematopoietic to leukemic stem cells. *Cell* **185**, 1266–1270 (2022).
6. Guarnera, L. *et al.* Microenvironment in acute myeloid leukemia: focus on senescence mechanisms, therapeutic interactions, and future directions. *Exp Hematol* S0301-472X(23)01705–8 (2023) doi:10.1016/j.exphem.2023.09.005.
7. Calado, R. T. *et al.* Short telomeres result in chromosomal instability in hematopoietic cells and precede malignant evolution in human aplastic anemia. *Leukemia* **26**, 700–707 (2012).
8. Myllymäki, M. *et al.* Short telomere length predicts nonrelapse mortality after stem cell transplantation for myelodysplastic syndrome. *Blood* **136**, 3070–3081 (2020).
9. Reilly, C. R. *et al.* The clinical and functional effects of TERT variants in myelodysplastic syndrome. *Blood* **138**, 898–911 (2021).
10. Gurnari, C. *et al.* A study of Telomerase Reverse Transcriptase rare variants in myeloid neoplasia. *Hematol Oncol* **40**, 812–817 (2022).
11. Burren, O. S. *et al.* Genetic architecture of telomere length in 462,666 UK Biobank whole-genome sequences. *Nat Genet* (2024) doi:10.1038/s41588-024-01884-7.

12. Gutierrez-Rodrigues, F. *et al.* Clonal landscape and clinical outcomes of telomere biology disorders: somatic rescuing and cancer mutations. *Blood* blood.2024025023 (2024) doi:10.1182/blood.2024025023.
13. Tefferi, A. *et al.* A Pilot Study of the Telomerase Inhibitor Imetelstat for Myelofibrosis. *N Engl J Med* **373**, 908–919 (2015).
14. Platzbecker, U. *et al.* Imetelstat in patients with lower-risk myelodysplastic syndromes who have relapsed or are refractory to erythropoiesis-stimulating agents (IMerge): a multinational, randomised, double-blind, placebo-controlled, phase 3 trial. (2023) doi:https://doi.org/10.1016/S0140-6736(23)01724-5.
15. Kimura, M. *et al.* Measurement of telomere length by the Southern blot analysis of terminal restriction fragment lengths. *Nat Protoc* **5**, 1596–1607 (2010).
16. Lin, J., Smith, D. L., Esteves, K. & Drury, S. Telomere length measurement by qPCR - Summary of critical factors and recommendations for assay design. *Psychoneuroendocrinology* **99**, 271–278 (2019).
17. Wand, T. *et al.* Telomere content measurement in human hematopoietic cells: Comparative analysis of qPCR and Flow-FISH techniques. *Cytometry A* **89**, 914–921 (2016).
18. Feuerbach, L. *et al.* TelomereHunter - in silico estimation of telomere content and composition from cancer genomes. *BMC Bioinformatics* **20**, 272 (2019).
19. Ding, Z., Mangino, M., Aviv, A., Spector, T. & Durbin, R. Estimating telomere length from whole genome sequence data. *Nucleic Acids Res* **42**, e75 (2014).
20. Arber, D. A. *et al.* The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* **127**, 2391–2405 (2016).
21. Raczy, C. *et al.* Isaac: ultra-fast whole-genome secondary analysis on Illumina sequencing platforms. *Bioinformatics* **29**, 2041–2043 (2013).
22. Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* **17**, 405–424 (2015).
23. Adema, V. *et al.* Pathophysiologic and clinical implications of molecular profiles resultant from deletion 5q. *EBioMedicine* **80**, 104059 (2022).

24. Henson, J. D. *et al.* DNA C-circles are specific and quantifiable markers of alternative-lengthening-of-telomeres activity. *Nat Biotechnol* **27**, 1181–1185 (2009).
25. Hershberger, C. E. *et al.* Complex landscape of alternative splicing in myeloid neoplasms. *Leukemia* **35**, 1108–1120 (2021).
26. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**, 511–515 (2010).
27. Ferrer, A., Stephens, Z. D. & Kocher, J.-P. A. Experimental and Computational Approaches to Measure Telomere Length: Recent Advances and Future Directions. *Curr Hematol Malig Rep* **18**, 284–291 (2023).
28. Bahaj, W. *et al.* Novel Scheme for Defining the Clinical Implications of TP53 Mutations in Myeloid Neoplasia. *Research square* rs.3.rs-2656206 Preprint at https://doi.org/10.21203/rs.3.rs-2656206/v1 (2023).
29. Dutta, S. *et al.* Functional Classification of TP53 Mutations in Acute Myeloid Leukemia. *Cancers (Basel)* **12**, (2020).
30. Noureen, N. *et al.* Integrated analysis of telomerase enzymatic activity unravels an association with cancer stemness and proliferation. *Nat Commun* **12**, 139 (2021).
31. Heaphy, C. M. *et al.* Altered telomeres in tumors with ATRX and DAXX mutations. *Science* **333**, 425 (2011).
32. Voso, M. T. & Gurnari, C. Have we reached a molecular era in myelodysplastic syndromes? *Hematology Am Soc Hematol Educ Program* **2021**, 418–427 (2021).
33. Duncavage, E. J. *et al.* Genome Sequencing as an Alternative to Cytogenetic Analysis in Myeloid Cancers. *N Engl J Med* **384**, 924–935 (2021).
34. Gutierrez-Rodrigues, F. *et al.* Clonal landscape and clinical outcomes of telomere biology disorders: somatic rescue and cancer mutations. *Blood* **144**, 2402–2416 (2024).
35. Nakao, T. *et al.* Mendelian randomization supports bidirectional causality between telomere length and clonal hematopoiesis of indeterminate potential. *Sci Adv* **8**, eabl6579 (2022).

36. Abel, H. J. *et al.* Genomic landscape of TP53-mutated myeloid malignancies. *Blood Adv* **7**, 4586–4598 (2023).
37. Steensma, D. P. *et al.* Imetelstat Achieves Meaningful and Durable Transfusion Independence in High Transfusion-Burden Patients With Lower-Risk Myelodysplastic Syndromes in a Phase II Study. *J Clin Oncol* **39**, 48–56 (2021).
38. Kuykendall, A. T. *et al.* Favorable overall survival with imetelstat in relapsed/refractory myelofibrosis patients compared with real-world data. *Ann Hematol* **101**, 139–146 (2022).
39. Baerlocher, G. M. *et al.* Telomerase Inhibitor Imetelstat in Patients with Essential Thrombocythemia. *N Engl J Med* **373**, 920–928 (2015).
40. Kinzig, C. G., Zakusilo, G., Takai, K. K., Myler, L. R. & de Lange, T. ATR blocks telomerase from converting DNA breaks into telomeres. *Science* **383**, 763–770 (2024).
41. Tan, K.-T. *et al.* Neotelomeres and telomere-spanning chromosomal arm fusions in cancer genomes revealed by long-read sequencing. *Cell Genom* **4**, 100588 (2024).

Tables

Table 1. Table illustrating composition of the study cohort, demographic data and telomere content.

Population (n)		Age (Years), median (Range)	M/F ratio	TC (*), median
Myeloid malignancies (1804)	AML (730)	67.7 (17.8-93.1)	1.1	352.9 (123.8-2605.4)
	pAML (669)	67.9 (17.8-93.1)	1.1	349.9 (123.8-2605.4)
	sAML (61)	63.5 (18.1-87.6)	1.6	369.1 (181.6-935.2)
	MDS (702)	73.1 (23.3-93.1)	1.4	480.8 (184.8-1350.5)
	LR-MDS (404)	73.2 (23.3-93.1)	1.2	498.1 (184.8-1350.5)
	HR-MDS (292)	72.9 (31.8-90.8)	1.6	455.7 (188.9-1208.8)
	MPN (372)	64.2 (15.7-87.3)	1.4	406.4 (162-1034)
Non-malignant controls (158)	PNH/AA (102)	38.5 (20-60)	0.9	429.2 (29.4-673.6)
	PPBL (50)	49.3 (22.5-68.5)	0.3	462.6 (283.4-764.1)
	HS (6)	33 (20-60)	NA	639.23 (519-747)

AA: aplastic anemia; AML: acute myeloid leukemia; HR-MDS: high risk-myelodysplastic syndrome; HS: healhty subjects; LR-MDS: low risk-myelodysplastic syndrome; pAML: primary AML; PNH: paroxysmal nocturnal hemoglobinuria; TC: telomere content; MPN: myeloproliferative neoplasms; PPBL: persistent polyclonal B cell lymphocytosis; sAML: secondary AML.

* Intratelomeric reads normalized by the number of reads of comparable GC content (48–52%) and multiplied by 10^6

Figure legends

Figure 1. Validation of telomere content using WGS-based and PCR-based methods and correlations of TC with age. A Correlation between TelSeq and Telomere Hunter, software programs able to measure TC from whole genome sequencing. When comparing the two bioinformatic pipelines, including all results, we obtained a correlation coefficient r^2=0.3448 (black line). However, Telseq was unable to produce informative results in 86 cases (4.4% of the population, for whom close to zero value was rendered; red dots), which were easily resolved by TelomereHunter. For the cases that were resolved by both software packages there was a good correlation coefficient of r^2=0.5650 (gray line). **B** Impact of TC adjustment, as described in methods section, in MN patients. **C** Correlation between Telomere Hunter-based and PCR-based TC. To investigate the correlation between age and TC across disease phenotypes and clonality, we projected the interquartile (25^{th}-75^{th} percentile) range of TC according to age ranges of healthy subject and non-clonal controls (PPBL patients). We then plotted TC according to age ranges of our study cohort [MN population (**D**), non-malignant clonal controls (including AA and PNH, **E**), AML (**F**) and MDS (**G**)] and highlighted the range of patients exhibiting TC lower and higher than 25^{th} and 75^{th} percentile for age ranges (in the histogram) and the overall TC split according to percentiles (pie chart). For visualization purposes and to respect the granularity of the data, percentiles of non-clonal controls were shown according to 10-years age ranges, whereas the remaining study cohort according to 5-years age ranges. Refer to **supplementary table 1** for additional information.

AA: aplastic anemia; AML: acute myeloid leukemia; MDS: myelodysplastic syndrome; ns: not significant; PNH: paroxysmal nocturnal hemoglobinuria; TC: telomere content; MPN: myeloproliferative neoplasms; PPBL: persistent polyclonal B cell lymphocytosis; WGS: whole genome sequencing.

Figure 2. Telomere content and disease features. A The histogram shows the percentage of patients with TC higher or lower the 25^{th} and 75^{th} percentile of MN cohort according to karyotypic abnormalities and gene

mutations. For a better visualization only somatic lesions with higher TC and detected in at least 1% of MN population were shown (Refer to **supplementary tables 2 and 3** for a comprehensive landscape of TC across somatic mutations and karyotype abnormalities). The asterisks highlight genotypes correlated to significantly higher or lower TC when compared to the remaining cohort (See additional information in **supplementary figures 5 and 6**). **B** Multivariate analysis for TC. Statistically significant features are indicated by asterisks. **C** MN population stratified according to TC. Disease phenotype, karyotype abnormalities and somatic lesions in MN presenting a TC above the 90th(**D**) and below the 10th (**E**) percentile. **F** Overview of *TP53* classification scheme proposed by Bahaj et al. **G** Distribution of TC in *TP53* wt and *TP53* mutated MN. **H** Distribution of *TP53* mutated MN in the overall cohort (outer circle) and high TC MN (Over 90th percentile; inner circle) according to the classification by Bahaj et al. **I** Kaplan-Meier curve for OS in MN patients grouped by *TP53* mutational status and TC (High: above 90th percentile, low: below 90th percentile). **J** Correlation between TC and *TP53* mutational and allelic status.

AML: acute myeloid leukemia; B: biallelic; MDS: myelodysplastic syndrome; MN: myeloid neoplasia; ns: not significant; PB: probable biallelic; PM: probable monoallelic; TC: telomere content; WT: wild type.

Figure 3. Mechanisms of telomere elongation. A Blasts TC distribution of AML patients for whom both blasts and lymphocytes were available for TC measurements. A statistically significant difference was observed in blast/lymphocyte ratio in samples with blast characterized by low TC (below 10th percentile) vs high TC (over 90th percentile). **B** The uncoupling of blast and lymphocytes TC was confirmed with PCR. **C** Evaluation, through singletons analysis, of canonic telomerase-dependent elongation (TTTGGG repeats) vs alternative lengthening (ALT) in high and low TC MN grouped by *TP53* mutational status; refer to supplementary figure 9 for the comprehensive singleton distribution across *TP53* mutated MN, *TP53* wt MN and non-malignant controls. **D** Schematic C-Circle analysis overview. Refer to methods for further details. **E** C-Circle assays for representative samples of different subgroups based on TC and *TP53* mutation status: *TP53* mutated and high TC, *TP53* wt and high TC, *TP53* mutated and low TC, *TP53* wt and low TC. For every

sample, a control without polymerase (-Pol) was compared to the test sample with polymerase (+Pol). See supplementary figures 19-20 and supplementary tables 12-18 for row data.

AML: acute myeloid leukemia; MN: myeloid neoplasia; TC: telomere content; wt: wild type.

Figure 4. Telomeres machinery genes expression analysis. A Volcano plot representing differential gene expressions between MNs whose TC is above (on the right) or under (on the left) the 90th percentile. The telomeres machinery genes are highlighted, the downregulated ones in blue, the upregulated ones in red. In **B-E** correlations between TC and genes with differential expression in *TP53* mutated and *TP53* wt MN: *DKC1* (**B**), *TEP1* (**C**), *TERC* (**D**), *TERT* (**E**).

MN: myeloid neoplasia; ns: not significant; TC: telomere content; wt: wild type.

Telomere content and genomics of myeloid neoplasia by WGS

Context of Research

Telomere length shortening has been associated with genomic instability and acquisition of molecular lesions, but these processes have not been systematically studied in large cohorts of myeloid neoplasia (MN) patients.

Methods

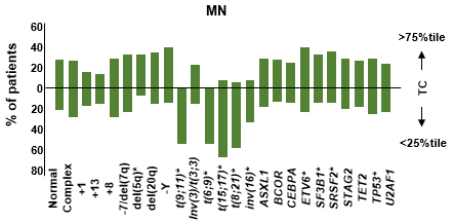
We adopted whole genome sequencing (WGS)-based pipelines to measure telomere content (TC) in a genotypically characterized and well-annotated cohort of MN patients (n=1804) and 158 non-malignant controls.

RNA sequencing was performed to investigate the expression of telomerase machinery genes.

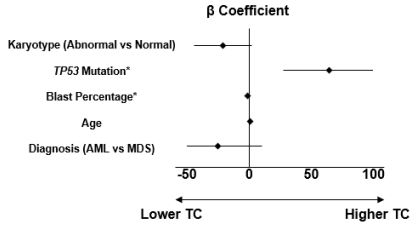
c-circle assays were performed in exemplary cases to establish the mechanisms of telomere elongation.

Main Findings

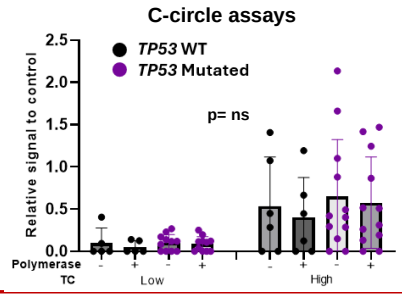
- General decrease in TC in MDS/AML irrespective of age.



- Higher TC in del(5q) and *ETV6*, *SF3B1*, *SRSF2* and *TP53* mutant cases. Lower TC in balanced translocations and *FLT3*, *KRAS*, *NPM1* and *NRAS* mutant cases.



- *TP53* mutation independently correlated with higher TC.



- Higher TC in *TP53*-mutated MN is not due to alternative telomere elongation.

Conclusions: WGS-based pipelines are reliable tools to measure TC. *TP53* mutant myeloid neoplasms are characterized by high TC.

*Guarnera et al. DOI: 10.xxxx/**blood**.2025xxxxxx*

Figure 1

Figure 1

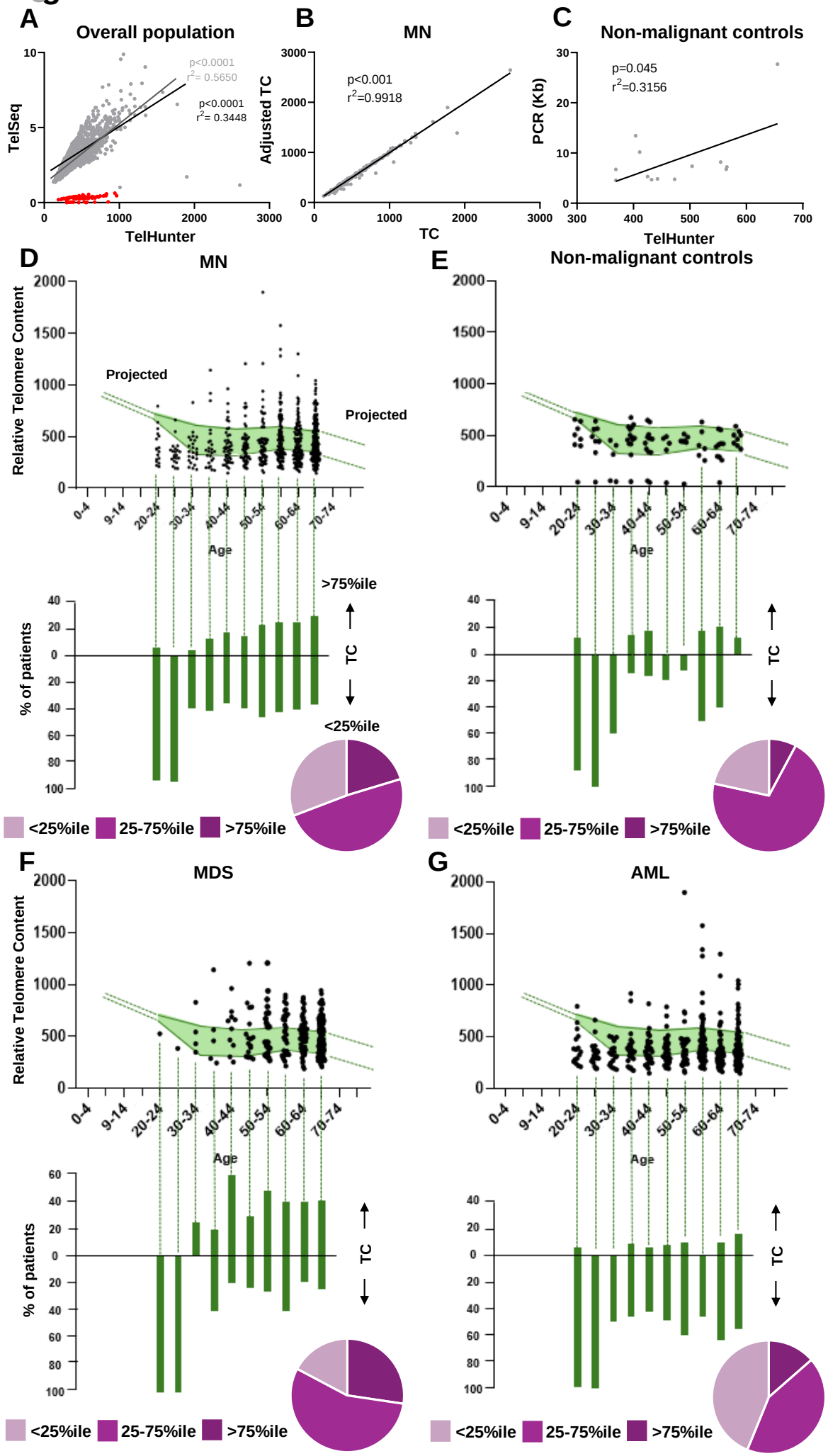


Figure 2

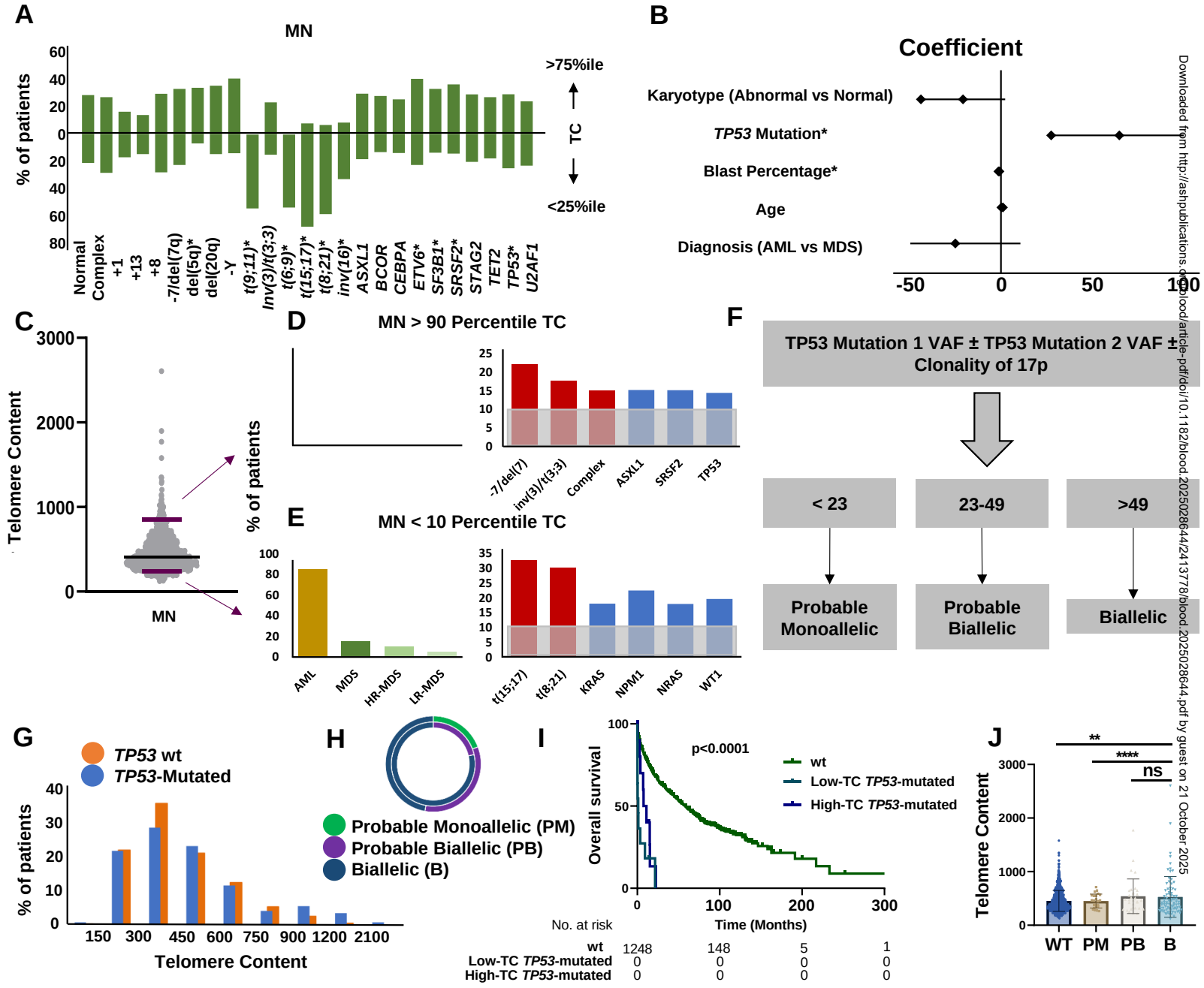


Figure 3

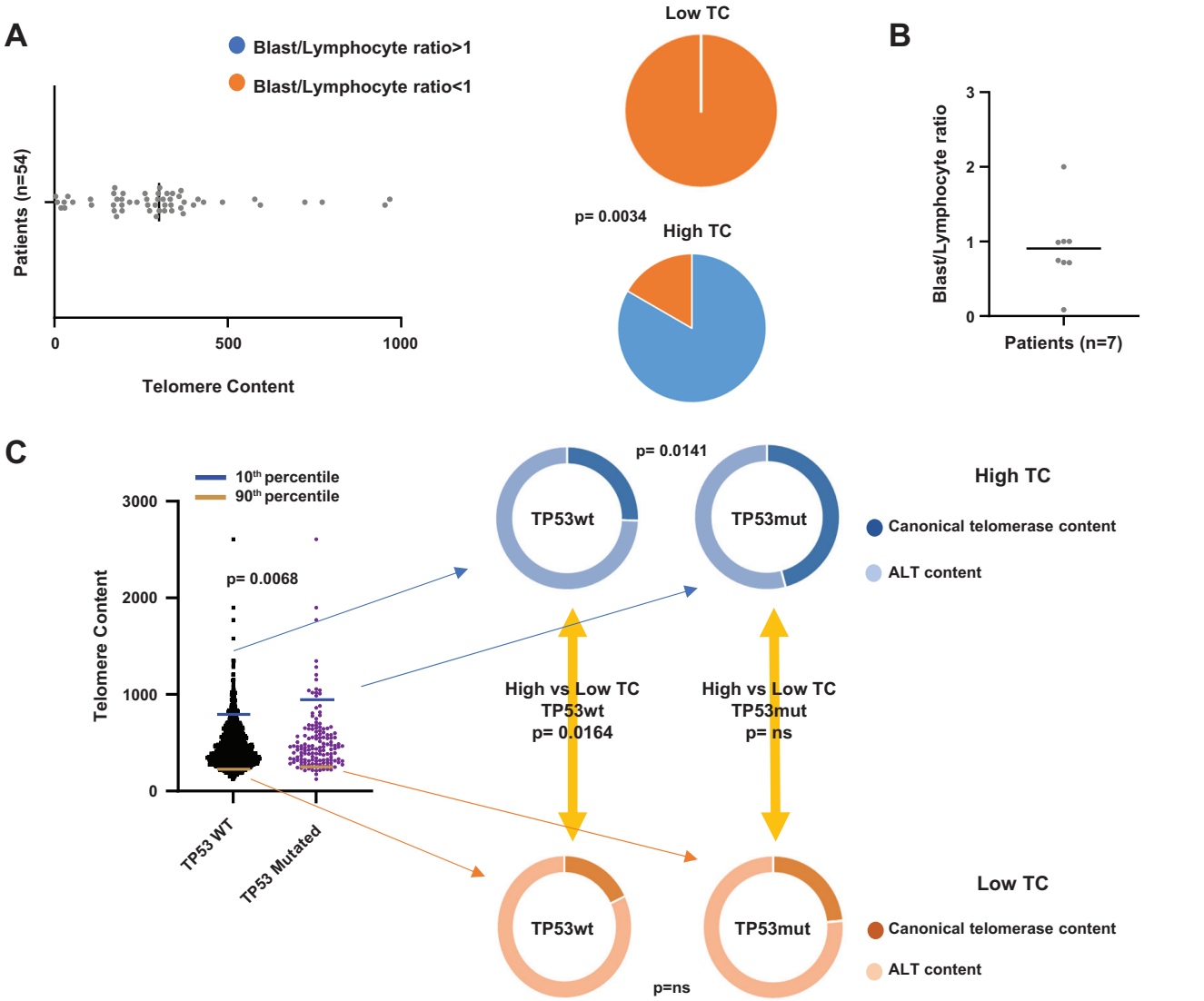


Figure 4