

ARTICLE OPEN



Genomic characterization of chronic lymphocytic leukemia in patients of African ancestry

Cecilia Bonolo de Campos¹, Chantal E. McCabe², Laura A. Bruins¹, Daniel R. O'Brien², Sochilt Brown¹, Renee C. Tschumper³, Cristine Allmer⁴, Yuan Xiao Zhu¹, Kari G. Rabe⁴, Sameer A. Parikh⁵, Neil E. Kay^{3,5}, Huihuang Yan², James R. Cerhan⁶, John N. Allan⁷, Richard R. Furman⁷, J. Brice Weinberg⁸, Danielle M. Brander⁹, Diane F. Jelinek¹⁰, Marta Chesi¹, Susan L. Slager^{2,5,11}✉ and Esteban Braggio^{1,11}✉

© The Author(s) 2025

Despite the considerable effort to characterize the genomic landscape of chronic lymphocytic leukemia (CLL), published data have been almost exclusively derived from patients of European Ancestry (EA), with significant underrepresentation of minorities, including patients of African Ancestry (AA). To begin to address this gap, we evaluated whether differences exist in the genetic and transcriptomic features of 157 AA and 440 EA individuals diagnosed with CLL. We sequenced 59 putative driver genes and found an increased frequency of high-impact mutations in AA CLL, including genes of the DNA damage repair (DDR) pathway. Telomere erosion was also increased in AA CLL, amplifying the notion of increased genomic instability in AA CLL. Furthermore, we found transcription enrichment of the Tumor Necrosis Factor- α (TNF α) Signaling via NF- κ B pathway in AA CLL compared to EA CLL, suggesting that tumor promoting inflammation plays an important role in AA CLL. In summary, these results suggest that genomic instability and NF- κ B activation is more prevalent in AA CLL than EA CLL.

Blood Cancer Journal (2025)15:14; <https://doi.org/10.1038/s41408-024-01207-3>

INTRODUCTION

Chronic lymphocytic leukemia (CLL) is one of the most common adult leukemias in the United States [1]. The disease is defined by a mature clonal B-cell lymphocyte population $\geq 5 \times 10^9/L$ in the peripheral blood [2] and characterized by a high degree of variability in clinical course [3]. In addition to advanced age, other known risk factors for developing CLL include male gender, family history of lymphoproliferative disorders including CLL, inherited common variants, and European ancestry (EA) [4–9].

In the United States, the incidence rate of CLL among Whites is higher compared to Blacks, yet Black populations have a younger median age of disease onset compared to Whites [10–14]. In addition, differences among clinical characteristics and outcomes exist between self-reported Black and non-Black CLL patients. At time of diagnosis, preliminary reports in African Ancestry (AA) CLL patients showed higher frequency of adverse prognostic factors, including advanced Rai stage, increased $\beta 2$ microglobulin levels, unmutated immunoglobulin heavy-chain variable region (u-IGHV) status, and del(17p) [11, 15, 16]. Moreover, these studies have found that AA CLL patients have shorter time to first treatment (TTT) and reduced overall survival (OS), and African American/Black ancestry has been shown to be an independent predictor of OS [8, 11, 13–15], with a 5-year relative survival of 62%, compared with 76% in non-Hispanic White [17]. Recent studies have found

that the inferior survival among African American/Black ancestry was sustained even in the context of modern targeted agents [13, 14], although the disparity appears to be improving [14]. Nevertheless, the biological and/or socioeconomic features that elicit these differences remain poorly understood.

Over the past decade, the characterization of the CLL tumor genetics has been done through multiple sequencing studies [18–22]. The mutated CLL driver genes have been mostly grouped into Notch signaling, B-cell receptor signaling and differentiation, DNA damage repair and cell cycle control, chromatin modification, RNA and ribosomal processing, apoptosis, NF- κ B signaling, and MAPK-ERK [18, 19, 21]. The total number of recurrently mutated CLL driver genes, as well as the number of altered signaling pathways, have been shown to impact CLL outcome [20, 23–26]. However, these studies have been performed mostly in patients with EA, and minority populations have been underrepresented [27]. Additionally, the importance of increasing the representation of AA populations in genomic studies was recently demonstrated with the discovery of AA-associated driver alterations and clinically actionable biomarkers across multiple cancer types [28]. Therefore, given the lack of genetic information in AA patients with CLL and a nonexistent understanding of the extent to which tumor genetics differs across ancestry among CLL patients, we screened the mutational status of 59 CLL driver genes and performed

¹Division of Hematology/Oncology, Mayo Clinic, Phoenix, AZ, USA. ²Division of Computational Biology, Mayo Clinic, Rochester, MN, USA. ³Department of Immunology, Mayo Clinic, Rochester, MN, USA. ⁴Division of Clinical Trials and Biostatistics, Mayo Clinic, Rochester, MN, USA. ⁵Division of Hematology, Mayo Clinic, Rochester, MN, USA. ⁶Division of Epidemiology, Mayo Clinic, Rochester, MN, USA. ⁷Division of Hematology and Medical Oncology, Weill Cornell Medical College, New York, NY, USA. ⁸Divisions of Hematology and Hematologic Malignancies & Cellular Therapy & VA Medical Center, Durham, NC, USA. ⁹Division of Hematologic Malignancy and Cellular Therapy, Duke Cancer Institute, Durham, NC, USA. ¹⁰Department of Immunology, Mayo Clinic, Scottsdale, AZ, USA. ¹¹These authors contributed equally: Susan L. Slager, Esteban Braggio. ✉email: Slager.susan@mayo.edu; Braggio.esteban@mayo.edu

Received: 24 September 2024 Revised: 26 November 2024 Accepted: 16 December 2024

Published online: 06 February 2025

transcriptome studies in a cohort of CLL patients of genetic AA and compared it with an EA CLL cohort.

METHODS

AA and EA CLL patient cohort

Blood was collected from individuals with CLL ($N=597$). To define genetic ancestry, we used principal components analyses and genotype data from Illumina and Affymetrix arrays; 157 individuals with a median 81% of AA ancestry and 440 individuals with a median 99% of EA ancestry were included. The median time to blood collection from diagnosis was 1.76 years for the AA cohort and 0.03 years for the EA cohort. Baseline clinical characteristics, Rai stage, and *IGHV*-mutation status were abstracted from medical records. The CLL patient cohort was derived from multiple medical research centers located in different regions of the United States.

Peripheral blood mononuclear cells (PBMC) were obtained from all individuals. Extraction of DNA and RNA was performed either directly from PBMCs presenting a tumor cell population >80% or following the enrichment of CD5+/CD19+ clonal cells in cases with purity <80%. DNA and RNA were extracted using the Gentra Puregene Cell Kit and miRNeasy Mini Kit, respectively (Qiagen, Valencia, CA, USA).

Ethics approval and consent to participate

This study was approved by the Mayo Clinic Institutional Review Board (15-006439 and 10-005344). All participants provided written informed consent and all methods were performed in accordance with the relevant guidelines and regulations. The study was conducted in accordance with the Declaration of Helsinki.

Custom panel targeted DNA sequencing

The entire coding regions of 59 putative driver genes in CLL were sequenced using a customized SureSelect panel, as described previously [23, 26]. DNA samples were paired-end sequenced (150 bp reads) and run on the Illumina HiSeq 4000 sequencer. The median coverage depth per sample was 1474x, with >99% of the samples having a median coverage depth >1000x per nucleotide, allowing the detection of mutations with variant allelic fraction (VAF) $\geq 1\%$. Somatic mutations were called using MuTect2 in tumor-only mode. Putative germline variants identified in the entire gnomAD dataset and within specific genetic ancestry groups in gnomAD or ESP6500 [29] public databases at a frequency higher than 0.01% (unless present in known CLL hotspots of the COSMIC [30] database), alternative variants with low coverage (<10 reads), and variant artifacts called due to sequencing strand bias were eliminated. The number of mutated genes per sample was then characterized by the presence of high impact (i.e., frameshift, nonsense, and splicing variants) and missense mutations in previously identified CLL hotspots (Supplementary Table 1).

Pipeline for RNA sequencing (MAP-RSeq)

mRNA sequencing was generated using a HiSeq 4000 sequencer for 36 AA CLL patients who were untreated at time of sample collection, 22 AA CLL patients whose samples were collected after treatment, and 292 EA untreated CLL patients. The raw RNA sequencing paired-end reads were processed through the Mayo RNA-Seq bioinformatics pipeline, MAP-RSeq version 3.1.4 [31]. MAP-RSeq employs STAR [32] to align reads to the reference human genome build hg38. Gene and exon expression quantification was performed using the Subread [33] package to obtain both raw and normalized (FPKM—Fragments Per Kilobase of transcript per Million mapped reads) reads. Finally, comprehensive analyses were run on the aligned reads to assess quality of the sequenced libraries. Batch correction for all RNA samples was performed by removing protein coding genes with low counts (i.e., mean and median thresholds below FPKM = 1), for each batch individually, resulting in 9934 protein coding genes. In addition, CombatSeq [34] was run using the annotated reported ancestry as a covariate, with CombatSeq corrected count values used for all downstream analyses. Differentially expressed genes were assessed using the bioinformatics package edgeR 2.6.2 [35] and defined by false discovery rate (FDR) < 0.01 and absolute log2 fold-change (Log2FC) ≥ 1 between groups. Unbiased pathway analysis to identify relevant biological processes was run on Metascape[36] and pathway visualization was performed with Pathview [37].

Telomere length quantification

Telomere length was initially measured through terminal restriction fragment length (TRF), and Southern blot hybridization was performed using the nonradioactive chemiluminescent TeloTAGGG Telomere Length Assay (Roche Diagnostics, Penzberg, Germany), according to the manufacturer's instructions. TRF is the gold standard of telomere length quantification; however, the high quantity of genomic DNA required for the assay (2 μ g) limited its broad use to quantify all study samples. Therefore, we measured the average telomere length through quantitative-PCR (qPCR), which enabled the analysis of large sample sizes using low quantities of genomic DNA [38] and which was recently validated in a large CLL cohort [39]. Briefly, telomere and Beta-Globin (HBB) primers were used to amplify the telomere sequence and the single-copy gene, respectively, in each CLL sample (Supplementary Table 2). Each reaction was performed with 20 ng of DNA in a total volume of 20 μ L in triplicate using the KAPA SYBR® FAST optimized for LightCycler®480 and analyzed on a LightCycler®480 (Roche Diagnostics, Mannheim, Germany). Absolute telomere lengths were then obtained by generating standard curves using synthetic oligonucleotide PCR standards for telomere and HBB, through serial dilution of each standard, as described previously [40] with minor modifications. Five samples with known telomere lengths through TRF analysis were included in every qPCR plate in order to perform a linear regression to account for variability between qPCR plates, as described previously [41]. The qPCR telomere lengths were validated with telomere lengths measured by TRF ($n=28$; $r=0.85$; $p<0.0001$, Supplementary Fig. 1), and short and long telomeres were defined using the median telomere length as a cutoff.

Statistical analyses

Statistical comparisons were done using t-tests, ANOVA, chi-square tests, Pearson correlation coefficients, or non-parametric tests, where appropriate, using GraphPad Prism version 9.4.0. Univariable and multivariable linear regression was used to evaluate associations of telomere length with genetic ancestry; multivariable analyses were adjusted for *IGHV* mutational status, number of mutated genes per sample, and Rai stage.

Data sharing statement

All transcriptome data are available at GEO under accession number GSE246521. The remaining original data are available from the corresponding authors on reasonable request.

RESULTS

Baseline study population characteristics

We sequenced 157 AA CLL patients and 440 EA CLL patients [23]. All but 42 AA CLL were untreated at time of sample collection. The following analyses were based on the untreated samples. The median age at diagnosis among the AA CLL and EA CLL were

Table 1. Demographic and clinical characteristics for AA and EA CLL.

Study Population	AA		EA		<i>p</i> values
	(n = 115)		(n = 440)		
Median Age at Diagnosis (years)	61 (33–94)		61 (27–87)		0.493 ^a
Female	36	34%	121	28%	0.139 ^b
Male	79	66%	319	72%	
Rai 0	71	71%	254	58%	0.027 ^b
Rai I-II	21	21%	153	35%	
Rai III-IV	8	8%	31	7%	<0.0001 ^b
Missing Rai	15		2		
Mutated <i>IGHV</i>	31	28%	213	49%	
Unmutated <i>IGHV</i>	78	72%	218	51%	
Missing <i>IGHV</i>	6		9		

^aMann-Whitney.

^bChi-Square.

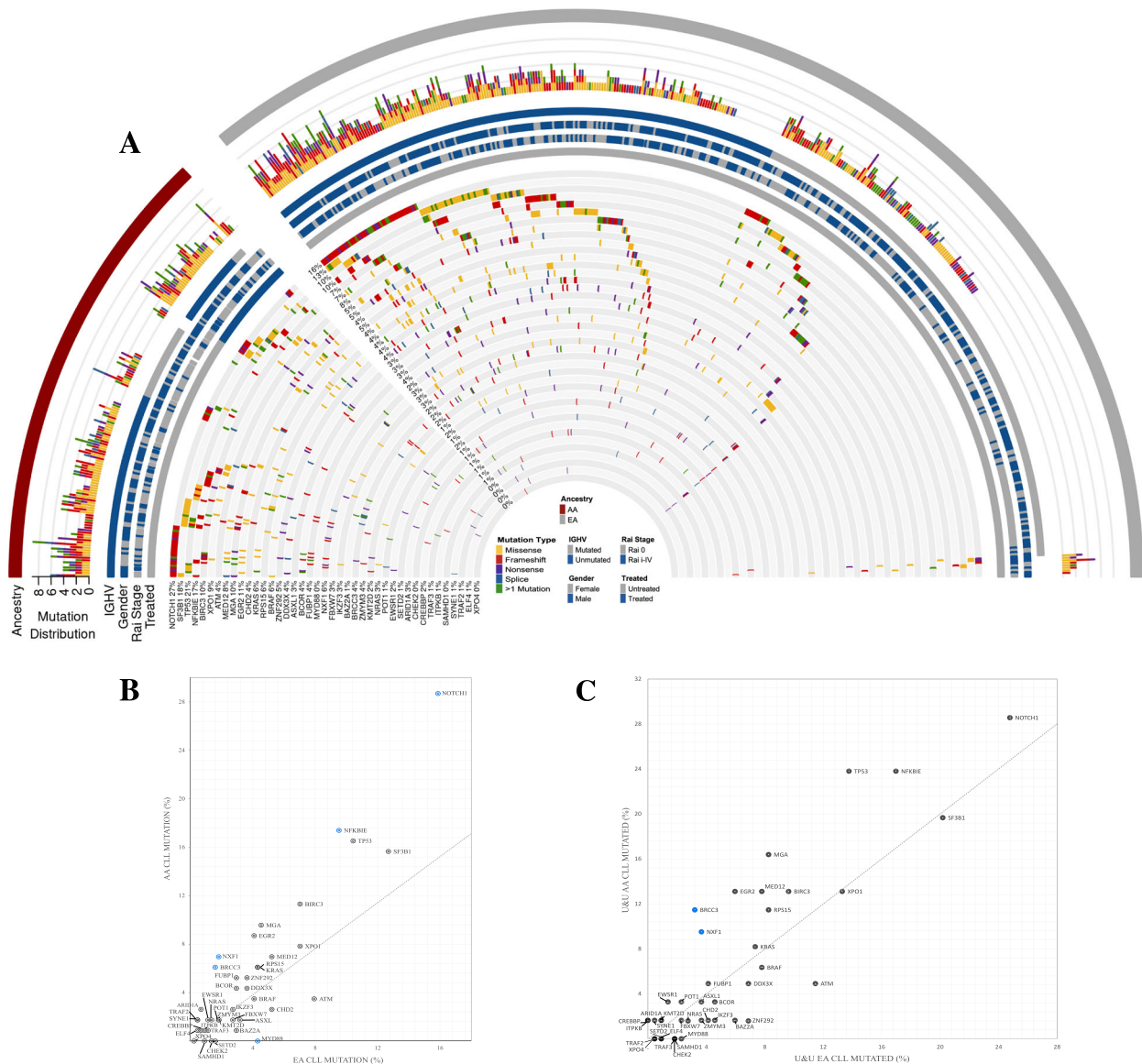


Fig. 1 Identification of recurrent driver mutations in AA and EA CLL. **A** Heatmap of mutated driver genes stratified by ancestry, treatment prior to sample collection, and *IGHV* status. The "Mutation Distribution" histogram represents the number of mutated genes per sample, with the type of mutation color-coded by known hotspot missense (yellow), frameshift (red), nonsense (purple), splice site (blue), and multiple mutations (green). **B** Frequency of mutated CLL driver genes compared between the AA ($n = 115$) and EA ($n = 440$) cohorts (black dot = $p > 0.05$; blue dot = $p \leq 0.05$). **C** Frequency of mutated CLL driver genes compared between the u-*IGHV* AA ($n = 78$) and u-*IGHV* EA ($n = 218$) cohorts (black dot = $p > 0.05$; blue dot = $p \leq 0.05$).

similar (61 years, Table 1) However, compared to EA CLL patients, AA CLL patients presented with lower Rai stage ($p = 0.027$) and higher frequency of unmutated *IGHV* (u-*IGHV*; 72% AA vs 51% EA, $p < 0.0001$; Table 1). Due to the known increased prevalence of somatic mutations in u-*IGHV* CLL compared to m-*IGHV* [18, 19], all analyses were performed in the AA ($N = 115$) and EA ($N = 440$) cohort as well as in the u-*IGHV* AA ($N = 78$) and u-*IGHV* EA ($N = 218$) CLL subsets. Lastly, we characterize the cohort of AA CLL patients whom had been previously treated at the time of sample collection ($N = 42$).

Genetic landscape of AA CLL

The most common mutated gene in EA CLL patients was *NOTCH1* (16%), followed by *SF3B1* (13%), *TP53* (10%), *NFKBIE* (10%), *ATM* (8%), *BIRC3* (7%), and *XPO1* (7%) [23]. In the AA CLL patients, the most commonly mutated gene was also *NOTCH1* (29%), followed by *NFKBIE* (17%), *TP53* (17%), *SF3B1* (16%), *BIRC3* (11%), *MGA* (10%),

EGR2 (9%), *XPO1* (8%), *MED12* (7%), and *NXF1* (7%) (Fig. 1A, B Supplementary Table 3). The type and location of variants were similar in each cohort (Supplementary Fig. 2).

Overall, the number of mutated genes was higher in AA (median of 2, range 0–9) compared to EA CLL (median of 1, range 0–7; $p = 0.069$). However, when we subset to those CLL patients with u-*IGHV* status, the number of mutated genes was similar (median of 2 (0–6) in AA and to 2 (0–7) in EA CLL, $p = 0.23$ Supplementary Table 4). We observed higher mutation prevalence (i.e., >50% increase) in multiple driver genes, specifically *BRCC3*, *NXF1*, *KMT2D*, *SYNE1*, *EGR2*, *FUBP1*, *EWSR1*, *MGA*, *TP53*, and *BIRC3* in u-*IGHV* AA CLL patients when compared to u-*IGHV* EA CLL patients (Fig. 1C).

DNA damage repair impairment and increased telomere erosion in AA CLL

Mutations in *BRCC3* and *TP53* have been associated with impaired DNA damage repair (DDR) and genomic instability [20, 42].

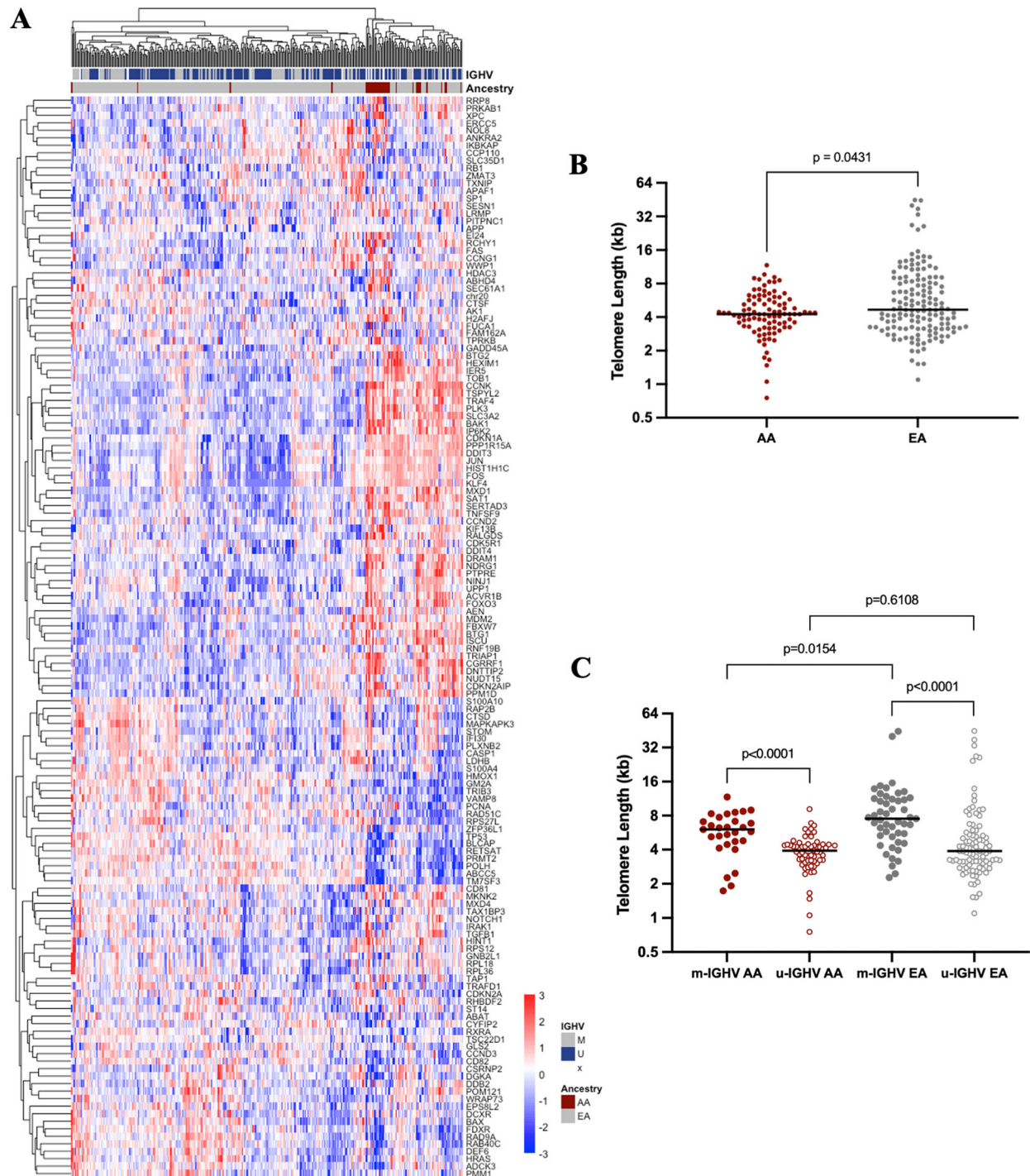


Fig. 2 P53 signaling pathway and telomere length in AA CLL. **A** Heatmap of the genes included in the P53 Pathway from the MSigDB Hallmark gene sets showing unbiased clustering of AA CLL. **B** Telomere length of AA ($n = 95$) and EA ($n = 147$) CLL samples measured by qPCR. **C** Telomere length of u-IGHV ($n = 63$) and m-IGHV ($n = 29$) CLL samples diagnosed in AA and u-IGHV ($n = 94$) and m-IGHV ($n = 49$) EA CLL samples.

Therefore, we evaluated differential gene expression in the MSigDB [43–45] Hallmark gene set P53 Pathway, showing the unbiased clustering of AA CLL (Fig. 2A). Specifically, AA CLL patients presented a significant down-regulation of multiple genes associated with genome stability and cellular DNA damage response (including *TP53* and *BRCC3*; $FDR < 0.05$). These findings persisted when subset the cohort to u-IGHV CLL patients.

To investigate differences in the telomere length between cohorts, we chose 95 AA CLL patients (median age at sample

collection 63 years; range 34–91) and 147 EA CLL patients with frequency-matched based on age (median age at sample collection 63 years; range 31–86, Supplementary Fig. 3). Shorter telomere lengths were observed in the tumor cells of the AA CLL (median 4.25 kb; range 0.75–11.69) compared to EA CLL patients (median 4.68 kb; range 1.1–44.92; $p = 0.0431$; Fig. 2B). Importantly, the differences observed were restricted to the CLL tumor cells, as described previously [39], with no meaningful differences observed in telomere lengths of germline samples collected from

Table 2. Multivariate analyses of the association between telomere length and genetic ancestry adjusting for Rai stage, IGHV status, and the number of mutated genes.

N = 207	Parameter	95% Lower	95% Upper	P
	Estimate	CI	CI	
Intercept	12.45769	9.5	15.41	<0.0001
Ancestry (1 = AA)	-2.68534	-4.56	-0.81	0.0055
Rai stage	-0.62725	-1.65	0.4	0.2312
IGHV(1 = UM)	-2.35662	-4.34	-0.38	0.0207
Number of mutated genes	-0.30345	-0.88	0.27	0.302

AA and EA CLL patients ($p=0.80$, data not shown). When stratifying by IGHV mutation status, shorter telomere length was observed in u-IGHV when compared to m-IGHV among both AA CLL (median 3.92 kb and 6.06, respectively; $p<0.0001$) and EA CLL (median 3.89 kb and 7.53, respectively; $p<0.0001$; Fig. 2C). We also note shorter telomere lengths in AA CLL compared to EA CLL among those with m-IGHV status ($p=0.0154$), but not among those with u-IGHV ($p=0.6108$; Fig. 2C).

The association of genetic ancestry on telomere erosion held in multivariable regression analysis, after accounting for Rai stage, IGHV status, and number of mutated CLL driver genes ($n=207$; Table 2). Telomere length was 2.69 kb shorter among AA compared to EA ($P=0.0055$). Similarly, those with u-IGHV had a significantly shorter telomere length compared to m-IGHV ($P=0.02$). Previous studies have shown that the acquisition of different mutations (e.g., *TP53* and *ATM* checkpoint variants) can be permissive to progressive cellular proliferation and critical telomere shortening [46]. However, most associations between telomere length and mutations of individual driver genes did not hold after adjusting for genetic ancestry, Rai stage, and IGHV status. We only found a significant increase in telomere length associated to *XPO1* and *IKZF3*, although these results were limited by sample size (Supplementary Table 5).

NF- κ B signaling pathway activation in AA CLL

Next, we performed an unbiased pathway enrichment analysis of differentially expressed genes. We identified the MSigDB [43–45] Hallmark gene set Tumor Necrosis Factor- α (TNF α) Signaling via the NF- κ B pathway as the most enriched pathway in AA CLL compared to EA CLL (Fig. 3A). The NF- κ B signaling pathway regulates immune response and inflammation, apoptosis and proliferation [47], and our gene expression analysis further demonstrated an overall enrichment of multiple additional pathways associated with immune response and inflammation in AA CLL. NF- κ B pathway enrichment was identified when all cases were analyzed (Fig. 3B), as well as when the u-IGHV subset of patients was compared (Fig. 3C). Specifically, we observed an upregulation of key mediators of the canonical and non-canonical NF- κ B pathway in AA CLL, including *NFKB1*, *RELA*, *NFKB2*, and *RELB* (FDR < 0.01; Supplementary Figs. 4, 5).

Treated AA CLL samples

Lastly, we evaluated a cohort of AA CLL samples collected from patients previously treated prior to sample collection ($N=42$). Prior treatment history of the treated AA CLL cohort was available for 33 patients and included 46% treated with chemotherapy regimens, 36% treated with chemoimmunotherapy, 15% with single-agent immunotherapy, and 3% with a novel agent associated with immunotherapy. Collectively, 60% of the treated AA CLL were male, median age at diagnosis was 52 years

(range 25–79), Rai Stage at diagnosis was 0 in 56%, I-II in 33%, and III-IV in 11%, and 86% were u-IGHV.

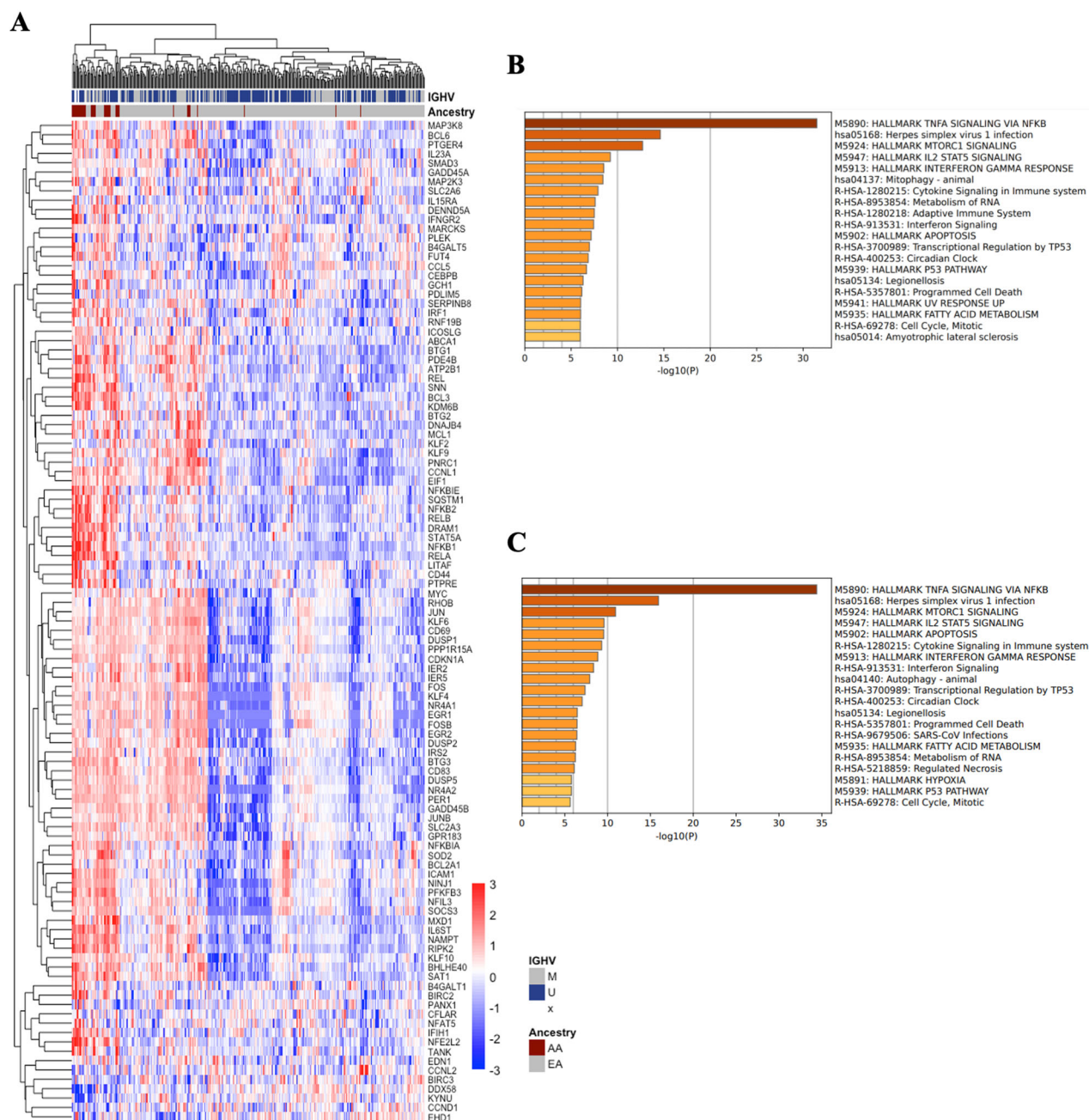
Unsurprisingly, *TP53* was identified as the most commonly mutated gene, found in 33% of cases. Then, *NOTCH1* (24%), *SF3B1* (24%), *NFKB1E* (17%), *EGR2* (17%), *BRAF* (12%), *MED12* (12%), *MGA* (12%), and *XPO1* (12%). However, the overall transcriptional program between the treated AA CLL ($n=22$) and the untreated AA CLL ($n=36$) was very similar, with only one gene presenting an absolute log2FC higher than 1 (*NKG7*; log2FC = -1.031; $p<0.05$). Finally, previous exposure to CLL treatment was associated to increased telomere erosion in an age and IGHV status frequency-matched comparison (Supplementary Fig. 6).

DISCUSSION

Minorities have been consistently underrepresented in somatic sequencing studies, including CLL, and, to the best of our knowledge, this is first study describing the leukemic genetic and transcriptomic characteristics of AA CLL patients. Importantly, AA CLL was characterized by increased genomic complexity when compared to EA CLL, with increased prevalence of mutations in *BRCC3*, *NOTCH1*, *NXF1*, and *NFKB1E*. We also found enrichment of NF- κ B signaling transcription and increased telomere erosion in AA CLL compared to EA CLL. Although our study was limited to 115 untreated AA CLL patients, our findings offer an inaugural overview of the genomic landscape in AA CLL and underscore the critical need to improve representation of AA patients in clinical trials, observational studies, and patient-reported outcome measures to better address the unique challenges faced by this population.

The findings identified in the present study indicate an augmented impaired DNA damage response (DDR) in AA CLL. DDR deficiencies accelerate the accumulation of chromosomal aberrations with progressive loss of genomic stability, recurrently identified in various cancers [48] and shown to negatively affect prognosis [20, 42]. Studies specifically evaluating the genomic alterations of many cancer types in the context of genomic ancestry reported higher genomic instability [49], homologous recombination deficiency, and aggressive molecular features in African Americans compared to European Americans [50]. However, these differences may be cancer type specific, highlighting the importance for increased inclusion of AA patients in future genomic studies [28].

Considering telomere dysfunction as a consequence of impaired DDR, we investigated telomere length in AA and EA CLL [51]. As expected, we found shorter telomere lengths among AA CLL compared to EA CLL patients, which was not limited to samples harboring *TP53* and *BRCC3* mutations. A recent study demonstrated highly variable telomere lengths in EA CLL, with shorter telomeres associated with a more aggressive CLL phenotype characterized by advanced Rai stage, u-IGHV, and *TP53* and *SF3B1* mutations. Additionally, shorter telomeres were also associated with higher genomic complexity, clonal evolution, and reduced progression free survival and overall survival [39]. The association between shorter telomeres in u-IGHV CLL has been previously documented, explained by the elongation of telomeres in the post-germinal center that occurs in m-IGHV CLL [39, 52]. Short telomeres also potentially generate a selective pressure for loss of cell-cycle checkpoint genes (including *TP53* and *ATM*), becoming permissive to further cell proliferation and telomere shortening [53]. These alterations promote the continuation of the disordered processes, progressively accumulating genomic lesions [46]. Taken together, the impaired DDR, increased frequency of mutated genes, and shorter telomere lengths identified herein indicate higher genomic instability in AA CLL. Nevertheless, future studies should also evaluate the impact of lifestyle and environmental factors on telomere length.



The increased activation of the NF- κ B signaling pathway, related to inflammation and proliferation, identified in AA CLL compared to EA CLL patients validates recent publications demonstrating an enrichment of immune- and inflammation-related genes and signaling pathways in tumor tissue-types from AA individuals [54–58]. Our results also corroborate findings of upregulation of NF- κ B and inflammatory cytokines, including increased expression of *IL6* in AA prostate cancer [59, 60]. Specifically, we suggest that tumor promoting inflammation plays an important role in AA CLL. Additional studies with larger sample sizes should be conducted to confirm the germline upregulation of the NF- κ B signaling pathway in normal B-cells of AA patients, evaluating its potential impact in the tumorigenesis and progression of CLL.

SPRINGER NATURE

landscape of AA CLL in an unbiased manner, as well as further characterize AA CLL in the context of the heterogeneous ancestry within Africa.

In summary, we have begun the characterization of the AA CLL genomic landscape, with evidence of higher genomic instability among AA CLL compared to EA CLL. Although the increased prevalence of driver mutations in the AA population is certainly driven by the higher frequency of u-IGHV cases [18, 19] and the longer median time to blood collection following diagnosis, we highlight the enrichment of crucial disease-related signaling pathways in our entire AA CLL cohort but also when evaluating solely the u-IGHV AA CLL individuals. Future studies are needed to evaluate whether the genomic and transcriptomic differences identified in this study of AA CLL impact clinical outcomes of these patients, providing novel insights into the racial/ethnic health disparities found in the disease. Further explorations of CLL in minority populations are necessary to better understand differences in the disease across populations, aiming towards the identification of novel vulnerabilities.

REFERENCES

- American Cancer Society. Cancer facts & figures 2022. Atlanta, Ga: American Cancer Society; 2022. <https://doi.org/10.3238/arztebl.2008.0255>.
- Hallek M, Cheson BD, Catovsky D, Caligaris-Cappio F, Dighiero G, Döhner H, et al. iwCLL guidelines for diagnosis, indications for treatment, response assessment, and supportive management of CLL. *Blood*. 2018;131:2745–60.
- Kipps TJ, Stevenson FK, Wu CJ, Croce CM, Packham G, Wierda WG, et al. Chronic lymphocytic leukaemia. *Nat Rev Dis Primers*. 2017;3:16096.
- Slager SL, Benavente Y, Blair A, Vermeulen R, Cerhan JR, Costantini AS, et al. Medical history, lifestyle, family history, and occupational risk factors for chronic lymphocytic leukemia/small lymphocytic lymphoma: the InterLymph non-Hodgkin lymphoma subtypes project. *J Natl Cancer Inst Monogr*. 2014;2014:41–51.
- Berndt SI, Skibola CF, Joseph V, Camp NJ, Nieters A, Wang Z, et al. Genome-wide association study identifies multiple risk loci for chronic lymphocytic leukemia. *Nat Genet*. 2013;45:868–76.
- Law PJ, Berndt SI, Speedy HE, Camp NJ, Sava GP, Skibola CF, et al. Genome-wide association analysis implicates dysregulation of immunity genes in chronic lymphocytic leukaemia. *Nat Commun*. 2017;8:14175.
- Berndt SI, Camp NJ, Skibola CF, Vijai J, Wang Z, Gu J, et al. Meta-analysis of genome-wide association studies discovers multiple loci for chronic lymphocytic leukemia. *Nat Commun*. 2016;7:10933.
- Coombs CC, Rassenti LZ, Falchi L, Slager SL, Strom SS, Ferrajoli A, et al. Single nucleotide polymorphisms and inherited risk of chronic lymphocytic leukemia among African Americans patients ascertained at Duke University. *Blood*. 2012;120:1687–91.
- Berndt SI, Vijai J, Benavente Y, Camp NJ, Nieters A, Wang Z, et al. Distinct germline genetic susceptibility profiles identified for common non-Hodgkin lymphoma subtypes. *Leukemia*. 2022;36:2835–44.
- National Cancer Institute. SEER Cancer stat facts: leukemia—chronic lymphocytic leukemia (CLL). 2022. <https://seer.cancer.gov/statfacts/html/clyl.html>.
- Zhao Y, Wang Y, Ma S. Racial differences in four leukemia subtypes: comprehensive descriptive epidemiology. *Sci Rep*. 2018;8:548.
- Ayers AA, Lyu L, Dance K, Ward KC, Flowers CR, Koff JL, et al. Characterizing lymphoma incidence and disparities for a cancer center catchment region. *Clin Lymphoma Myeloma Leuk*. 2019;19:699–708.
- Kittai AS, Huang Y, Bhat SA, Paskett ED, Rogers KA, Barrientos JC, et al. Racial and socioeconomic disparities in CLL/SLL: analysis of SEER data from 2006 to 2019. *Blood Adv*. 2023;7:2575–9.
- Vardell VA, Ermann DA, Fitzgerald LA, Shah HR, Hu B, Stephens DM. Influence of racial and ethnic identity on overall survival in patients with chronic lymphocytic leukemia. *Am J Hematol*. 2023;98:E172–E174.
- Falchi L, Keating MJ, Wang X, Coombs CC, Lanasa MC, Strom S, et al. Clinical characteristics, response to therapy, and survival of African American patients diagnosed with chronic lymphocytic leukemia. *Cancer*. 2013;119:3177–85.
- Coombs CC, Falchi L, Weinberg JB, Ferrajoli A, Lanasa MC. Chronic lymphocytic leukemia in African Americans. *Leuk Lymphoma*. 2012;53:2326–9.
- Shenoy PJ, Malik N, Sinha R, Nooka A, Nastoupil LJ, Smith M, et al. Racial differences in the presentation and outcomes of chronic lymphocytic leukemia and variants in the United States. *Clin Lymphoma Myeloma Leuk*. 2011;11:498–506.
- Puente XS, Beà S, Valdés-Mas R, Villamor N, Gutiérrez-Abril J, Martín-Subero JL, et al. Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature*. 2015;526:519–24.
- Landau DA, Tausch E, Taylor-Weiner AN, Stewart C, Reiter JG, Bahlo J, et al. Mutations driving CLL and their evolution in progression and relapse. *Nature*. 2015;526:525–30.
- Brieghel C, da Cunha-Bang C, Yde CW, Schmidt AY, Kinalis S, Nadeu F, et al. The number of signaling pathways altered by driver mutations in chronic lymphocytic leukemia impacts disease outcome. *Clin Cancer Res*. 2020;26:1507–15.
- Knisbacher BA, Lin Z, Hahn CK, Nadeu F, Duran-Ferrer M, Stevenson KE, et al. Molecular map of chronic lymphocytic leukemia and its impact on outcome. *Nat Genet*. 2022;54:1664–74.
- Robbe P, Ridout KE, Vavoulis DV, Dréau H, Kinnersley B, Denny N, et al. Whole-genome sequencing of chronic lymphocytic leukemia identifies subgroups with distinct biological and clinical features. *Nat Genet*. 2022;54:1675–89.
- Kleinstern G, O'Brien DR, Li X, Tian S, Kabat BF, Rabe KG, et al. Tumor mutational load predicts time to first treatment in chronic lymphocytic leukemia (CLL) and monoclonal B-cell lymphocytosis beyond the CLL international prognostic index. *Am J Hematol*. 2020;95:906–17.
- Yu L, Kim HT, Kasar SN, Benien P, Du W, Hoang K, et al. Survival of Del17p CLL depends on genomic complexity and somatic mutation. *Clin Cancer Res*. 2017;23:735–45.
- Leeksa AC, Taylor J, Wu B, Gardner JR, He J, Nahas M, et al. Clonal diversity predicts adverse outcome in chronic lymphocytic leukemia. *Leukemia*. 2019;33:390–402.
- Kleinstern G, Boddicker NJ, O'Brien DR, Allmer C, Rabe KG, Norman AD, et al. Tumor mutational load is prognostic for progression to therapy among high-count monoclonal B-cell lymphocytosis. *Blood Adv*. 2024;8:2118–29.
- Popejoy AB, Fullerton SM. Genomics is failing on diversity. *Nature*. 2016;538:161–4.
- Jiagge E, Jin DX, Newberg JY, Perea-Chamblee T, Pekala KR, Fong C, et al. Tumor sequencing of African ancestry reveals differences in clinically relevant alterations across common cancers. *Cancer Cell*. 2023;41:1963–71.
- Exome Variant Server, NHLBI Exome Sequencing Project (ESP), Seattle, WA. 2011. <http://evs.gs.washington.edu/EVS/>.
- Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res*. 2019;47:D941–D947.
- Kalari KR, Nair AA, Bhavsar JD, O'Brien DR, Davila JI, Bockel MA, et al. MAP-RSeq: Mayo analysis pipeline for RNA sequencing. *BMC Bioinformatics*. 2014;15:224–34.
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29:15–21.
- Liao Y, Smyth G, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res*. 2013;41:e108.
- Zhang Y, Parmigiani G, Johnson WE. ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genom Bioinform*. 2020;2:lqaa078.
- Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010;26:139–40.
- Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun*. 2019;10:1523.
- Luo W, Pant G, Bhavnasi YK, Blanchard SG, Brouwer C. Pathview Web: user friendly pathway visualization and data integration. *Nucleic Acids Res*. 2017;45:W501–W508.
- Cawthon RM. Telomere measurement by quantitative PCR. *Nucleic Acids Res*. 2002;30:e47.
- Jebaraj BMC, Tausch E, Landau DA, Bahlo J, Robrecht S, Taylor-Weiner AN, et al. Short telomeres are associated with inferior outcome, genomic complexity, and clonal evolution in chronic lymphocytic leukemia. *Leukemia*. 2019;33:2183–94.
- O'Callaghan NJ, Dhillon VS, Thomas P, Fenech M. A quantitative real-time PCR method for absolute telomere length. *BioTechniques*. 2008;44:807–9.
- Jebaraj BMC, Stilgenbauer S. Telomere dysfunction in chronic lymphocytic leukemia. *Front Oncol*. 2021;10:1–11.
- Hernández-Sánchez M, Rodríguez-Vicente AE, González-Gascón y Marín I, Quijada-Alamo M, Hernández-Sánchez JM, Martín-Izquierdo M, et al. DNA damage response-related alterations define the genetic background of patients with chronic lymphocytic leukemia and chromosomal gains. *Exp Hematol*. 2019;72:9–13.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. 2005;102:15545–50.
- Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. 2011;27:1739–40.
- Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database hallmark gene set collection. *Cell Syst*. 2015;1:417–25.
- Bloehdorn J, Braun A, Taylor-Weiner A, Jebaraj BMC, Robrecht S, Krzykalla J, et al. Multi-platform profiling characterizes molecular subgroups and resistance networks in chronic lymphocytic leukemia. *Nat Commun*. 2021;12:1–18.
- Dolcet X, Llobet D, Pallares J, Matias-Guiu X. NF- κ B in development and progression of human cancer. *Virchows Arch*. 2005;446:475–82.
- Minten EV, Yu DS. DNA repair: translation to the clinic. *Clinical Oncology*. 2019;31:303–10.

49. Yuan J, Hu Z, Mahal BA, Zhao SD, Kensler KH, Pi J, et al. Integrated analysis of genetic ancestry and genomic alterations across cancers. *Cancer Cell*. 2018;34:549–60.
50. Sinha S, Mitchell KA, Zingone A, Bowman E, Sinha N, Schäffer AA, et al. Higher prevalence of homologous recombination deficiency in tumors from African Americans versus European Americans. *Nat Cancer*. 2020;1:112–21.
51. Sieverling L, Hong C, Koser SD, Ginsbach P, Kleinheinz K, Hutter B, et al. Genomic footprints of activated telomere maintenance mechanisms in cancer. *Nat Commun*. 2020;11:733.
52. Roos G, Kröber A, Grabowski P, Kienle D, Bühler A, Döhner H, et al. Short telomeres are associated with genetic complexity, high-risk genomic aberrations, and short survival in chronic lymphocytic leukemia. *Blood*. 2008;111:2246–52.
53. Rossi D, Bodoni CL, Genuardi E, Monitillo L, Drandi D, Cerri M, et al. Telomere length is an independent predictor of survival, treatment requirement and Richter's syndrome transformation in chronic lymphocytic leukemia. *Leukemia*. 2009;23:1062–72.
54. Hardiman G, Savage JS, Starr Hazard E, Wilson RC, Courtney SM, et al. Systems analysis of the prostate transcriptome in African–American men compared with European–American men. *Pharmacogenomics*. 2016;17:1129–43.
55. Klopot A, Baida G, Kel A, Tsoi LC, Perez White BE, Budunova I. Transcriptome analysis reveals intrinsic pro-inflammatory signaling in healthy African American skin. *J Invest Dermatol*. 2022;142:1360–71.
56. Singh U, Hernandez KM, Aronow BJ, Wurtele ES. African Americans and European Americans exhibit distinct gene expression patterns across tissues and tumors associated with immunologic functions and environmental exposures. *Sci Rep*. 2021;11:1–14.
57. Thames AD, Irwin MR, Breen EC, Cole SW. Experienced discrimination and racial differences in leukocyte gene expression. *Psychoneuroendocrinology*. 2019;106:277–83.
58. Marczyk M, Qing T, O'Meara T, Yagahoobi V, Pelekanou V, Bai Y, et al. Tumor immune microenvironment of self-identified African American and non-African American triple negative breast cancer. *NPJ Breast Cancer*. 2022;8:88.
59. Powell IJ, Dyson G, Land S, Ruterbusch J, Bock CH, Lenk S, et al. Genes associated with prostate cancer are differentially expressed in African American and European American men. *Cancer Epidemiol Biomarkers Prev*. 2013;22:891–7.
60. Powell IJ, Chinni SR, Reddy SS, Zaslavsky A, Gavande N. Pro-inflammatory cytokines and chemokines initiate multiple prostate cancer biologic pathways of cellular proliferation, heterogeneity and metastasis in a racially diverse population and underlie the genetic/biologic mechanism of racial disparity: update. *Urol Oncol*. 2021;39:34–40.

ACKNOWLEDGEMENTS

The authors would like to thank the individuals who participated in our study. This work could not have been done without their valuable time and effort. We also like to acknowledge the greatly enlightening exchanges on telomere length quantification methods with Drs. Billy Michael Chelliah Jebaraj and Stephan Stilgenbauer. This work was supported by the National Institutes of Health grants R01CA235026 and R01CA254951.

AUTHOR CONTRIBUTIONS

Concept and Study Design was performed by CBC, SLS, and EB. Acquisition of Data was performed by CBC, LAB, and CA. Analysis of Data was performed by CBC, CEM,

DRO, KGR, SLS, and EB. Interpretation of Data was performed by CBC, SLS, and EB. Manuscript was drafted by CBC, SLS, and EB. Manuscript was reviewed and approved by CBC, CEM, LAB, DRO, SB, RCT, CA, YXZ, KGR, SAP, NEK, HY, JRC, JNA, RRF, JBW, DMB, DFJ, MC, SLS, and EB.

COMPETING INTERESTS

SAP: Research funding has been provided to the institution from Janssen, AstraZeneca, Merck, and Genentech for clinical studies in which SAP is a principal investigator. Honoraria has been provided to the institution from Pharmacyclics, Merck, AstraZeneca, Janssen, Genentech, Amgen, MingSight Pharmaceuticals, TG Therapeutics, Novalgen Limited, Kite Pharma, and AbbVie for SAP's participation in consulting activities/advisory board meetings. NEK: Advisory Board for AbbVie, AstraZeneca, Beigene, Behring, Boehringer Ingelheim Pharmaceuticals, Inc., Dava Oncology, Janssen, Juno Therapeutics, and Pharmacyclics. Data Safety Monitoring Committee for Agios Pharm, AstraZeneca, BMS – Celgene, and Dren Bio Janssen. Research funding from AbbVie, Acerta Pharma, Bristol Meyer Squib, Celgene, Genentech, Pharmacyclics, Sunesis, and Vincex. JNA: Consulting: Abbvie, BeiGene, Genentech, Janssen, Pharmacyclics. Research: BeiGene, Genentech, Janssen. Honoraria: Abbvie, AstraZeneca, BeiGene, Genentech, Janssen, Pharmacyclics

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41408-024-01207-3>.

Correspondence and requests for materials should be addressed to Susan L. Slager or Esteban Braggio.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025