



OPEN Shared genomic features of HIV+ diffuse large B-cell lymphoma in two African cohorts

Sophia M. Roush^{1,10}, Mishalan Moodley^{2,10}, Jenny Coelho¹, Samantha Beck¹, Amon Chirwa³, Edwards Kasonkanji³, Marriam Mponda³, Maurice Mulenga³, Tamiwe Tomoka³, Hanri van Zijl^{4,5}, Katherine Hodgkinson^{4,5}, Arshad Ismail^{2,6,7}, Senzo Mtshali², Jonathan Featherston², Satish Gopal⁸, Matthew S. Painschab⁹, Jenifer Vaughan^{4,5,10} & Yuri Fedoriw^{1,9}✉

The number of people with HIV (PWH) in Africa is rising due to population growth and antiretroviral therapy (ART) availability, with diffuse large B-cell lymphoma (DLBCL) a major cause of mortality. HIV and ART alter DLBCL tumor biology, but few studies of DLBCL include PWH or African patients, limiting translation of emerging treatment strategies. Here, we performed whole exome sequencing of 48 tumors (40 HIV-positive [HIV+]) with paired germline of DLBCL patients from Malawi and South Africa. HIV + DLBCL tumors had distinct mutations depending on ART exposure, and there were several recurrent deleterious variants, with *ANKRD11* mutations being prognostic. One tumor from each cohort had high tumor mutational burden and microsatellite-instability with *PMS2* and *ARID1A* mutation. These findings suggest shared genomic characteristics among HIV + DLBCL in Africa, offering opportunities for tailored biomarkers and therapeutic targets for this underserved population.

Africa faces a rapidly growing cancer burden, driven by both population growth and the high prevalence of HIV¹. By 2030, cancer incidence in Africa is expected to double, and by 2040, low- and middle-income countries are expected to account for 70% of global cancer deaths^{2,3}. As cancer therapies increasingly depend on molecular and genomic characterization, disparities in translational research persist, with significant overrepresentation of patients of European ancestry^{4,5}. These gaps are further exacerbated by the frequent exclusion of people with HIV (PWH) from translational and clinical studies of cancer, despite their growing proportion of cancer patients, particularly in Africa^{6,7}.

In Eastern and Southern Africa, where more than half of the global HIV-positive (HIV+) population resides, the intersection of human immunodeficiency virus (HIV) and cancer has become a pressing public health issue⁷. The introduction of antiretroviral therapy (ART) has significantly improved life expectancy of PWH, yet cancer, particularly aggressive B-cell lymphomas like diffuse large B-cell lymphoma (DLBCL), has emerged as the leading cause of mortality among PWH^{8–10}. Even when well-controlled on ART, PWH are over ten times as likely to develop DLBCL compared to people without HIV^{11–14}. Before ART, survival rates for HIV + DLBCL patients were dismal, with overall survival (OS) rates below 20%, but current treatments combining full-dose chemotherapy and ART have significantly improved survival, making outcomes comparable to those in HIV-negative (HIV-) patients^{11,15}.

Despite these clinical advances, genomic studies of HIV + DLBCL remain limited. While numerous studies of HIV- DLBCL have established genetic and transcriptomic subtypes, molecular characterization efforts in HIV + DLBCL have been constrained by sample numbers, absence of HIV- comparator groups, and restricted gene panel coverage^{16,17}. Prior studies from our Malawi-based cohort provide evidence for impacts of HIV and

¹Department of Pathology and Laboratory Medicine, School of Medicine, University of North Carolina (UNC), Brinkhous-Bullitt Building Rm 822 160 Medical Dr, Chapel Hill, NC 27514, USA. ²Sequencing Core Facility, Division of the National Health Laboratory Service, National Institute for Communicable Diseases, Johannesburg, South Africa. ³UNC Project Malawi, Lilongwe, Malawi. ⁴Department of Molecular Medicine and Haematology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa. ⁵National Health Laboratory Services, Johannesburg, South Africa. ⁶Department of Biochemistry and Microbiology, Faculty of Science, Engineering and Agriculture, University of Venda, Thohoyandou, South Africa. ⁷Institute for Water and Wastewater Technology, Durban University of Technology, Durban 4000, South Africa. ⁸National Cancer Institute Center for Global Health, Rockville, MD, USA. ⁹UNC Lineberger Comprehensive Cancer Center, Chapel Hill, NC, USA. ¹⁰Sophia M. Roush, Mishalan Moodley and Jenifer Vaughan contributed equally to this work. ✉email: yuri.fedoriw@unchealth.unc.edu

ART on the transcriptomic profile, microenvironment, and T-cell receptor repertoire repertoires of DLBCL. Notably, established tumor biomarkers with prognostic significance in HIV- DLBCL, including cell-of-origin (COO), fail to correlate with HIV + DLBCL outcomes in Malawi, but whether this represents distinct tumor biology or reflects the impact of non-biologic factors remains unclear. This underscores the need for more inclusive research and population-specific biomarker validation^{18–21}.

In addition to differences attributable to HIV and/or ART, studying multi-national cohorts may shed light on how ancestral, medical, and environmental factors influence molecular features and outcomes of DLBCL. Black DLBCL patients in the United States have inferior survival compared to white patients, and DLBCL patients of African ancestry are enriched for driver mutations in specific genes^{22,23}. Yet, over 80% of tumor sequencing data originate from white individuals in high-income countries⁴.

To address the gaps in understanding of how HIV, ART exposure, and African ancestry influence DLBCL mutational profile, we performed whole exome sequencing on 48 DLBCL tumors and paired germline samples from cohorts in Malawi (MW) and South Africa (SA), including 40 HIV + cases. We aimed to characterize how HIV and ART shape the mutational profile of DLBCL—an extensively studied malignancy among HIV-individuals in high-income countries. Specifically, we compared clinical and histological features between HIV + and HIV- cohorts, assessed ART's impact on the mutational landscape, and identified recurrent mutations and molecular characteristics in this historically understudied population.

Methods

Cohort characteristics

The MW cohort is composed of patients from the Kamuzu Central Hospital (KCH) Lymphoma Study (NCT02835911), which is based in Lilongwe, MW and began enrolling patients with newly diagnosed lymphoma in 2013^{24,25}. Patients with histologically confirmed DLBCL receive CHOP (cyclophosphamide, doxorubicin, vincristine, and prednisolone) and a subset also received rituximab as part of a clinical trial (NCT02660710)²⁶. Patients are followed for up to five years, with survival data censored on April 26th, 2024 for this study. The KCH Lymphoma Study was approved by the University of North Carolina IRB and Malawi National Health Sciences Research Committee, and all patients provide written informed consent. All studies were performed according to relevant guidelines and regulations. DLBCL patients enrolled from 2013 to 2021 with adequate pre-treatment tumor and paired whole blood samples for DNA extraction were included in this sequencing study.

The SA cohort is composed of patients diagnosed at or referred to Chris Hani Baragwanath Academic Hospital with a primary diagnosis of DLBCL between 2019 and 2022. The study was approved by the human research ethics committee of the University of Witwatersrand (M190709), and ethical clearance for next-generation sequencing was granted under reference number M220965. First-line therapy for patients with DLBCL in SA is CHOP, but several patients received rituximab following a suboptimal response (Table 1). Overall, 13 were treated with CHOP-based therapy, and three received rituximab as part of a second line regimen. Nine were either too ill to receive therapy or were treated with attenuated dose CHOP-based chemotherapy. Paired germline samples were sourced from uninvolved staging bone marrow aspirate slides.

HIV + patients in both cohorts either continued or initiated ART. For the analyses in this study, HIV + patients with over six months of ART prior to study enrollment were considered ART-experienced (ART-exp). If ART duration was unknown or undocumented, HIV viral load < 200 was used to assign a patient as HIV+/ART-exp. Because only CD4 count was available, SA10 and SA15 were not assigned HIV/ART status and were not included in comparisons by HIV/ART status.

Histologic diagnosis and immunohistochemistry

Primary histologic diagnoses were made per standard of care by pathologists and hematopathologists at Chris Hani Baragwanath Academic Hospital and the National Health Laboratory Service in SA, and as previously described at KCH^{20,24,27,28}. COO was assigned for the MW cohort by immunohistochemistry (IHC) using CD10 and BCL6 (Leica Biosystems), and MUM1 (Dako)²⁸. Based on IHC availability, COO was assigned for 15 samples from the SA cohort using CD10 (Roche), BCL6 (Cell Marque) and MUM1 (Dako).

Whole exome sequencing

For the samples from MW, DNA was extracted from 40 pre-treatment formalin-fixed, paraffin-embedded (FFPE) DLBCL (GeneRead DNA FFPE Kit) and corresponding whole blood samples (QIAmp Blood Mini Kit) by Novogene Co (South Plainfield, NJ, USA). Target capture was performed using the Agilent SureSelect Human All Exon V6 r2 and the libraries were sequenced on the NovaSeq 6000 platform (Illumina, San Diego, CA). Three FFPE tumors had insufficient DNA quantity extracted (< 10 ng/ul), and seven FFPE tumors and seven blood samples failed Novogene's library preparation or sequencing quality control. For the samples from SA, DNA was extracted from 24 FFPE DLBCL (QiaGen, AllPrep Kit) and corresponding uninvolved bone marrow aspirate (QIAmp Puregene Kit). Target capture was performed using the Agilent SureSelect Human All Exon V8 and the libraries were sequenced on the NextSeq 2000 platform (Illumina, San Diego, CA). One sample failed sequencing with < 10x average depth of coverage and was removed from the rest of the analysis (SA24).

Somatic variant analysis

Tumor and normal FASTQs were processed using the jUNCtion pipeline (somatic-1.0.11, Agilent v6 profile, hg38, UNC Lineberger Bioinformatics Core) via the following steps: Quality control was performed using Picard (RRID: SCR_006525), MultipleMetrics (RRID: SCR_018929), and FastQC (RRID: SCR_014583). BAM files were realigned using coverageBed (RRID: SCR_006646). Variant calling was conducted with Strelka2 (RRID: SCR_005109), Cadabra (RRID: SCR_003277), and Mutect2 (RRID: SCR_000559), with results merged into a

	Malawi		South Africa	Total	
	HIV+ (N=16)	HIV- (N=8)	HIV+ (N=24)	HIV+ (N=40)	HIV- (N=8)
Age (years)					
Median [Min, Max]	40.5 [17.0, 61.0]	46.0 [32.0, 57.0]	40.0 [19.0, 59.0]	40.5 [17.0, 61.0]	46.0 [32.0, 57.0]
Sex					
Female	5 (31.3%)	3 (37.5%)	12 (50.0%)	17 (42.5%)	3 (37.5%)
Male	11 (68.8%)	5 (62.5%)	12 (50.0%)	23 (57.5%)	5 (62.5%)
HIV/ART status					
HIV-	0 (0%)	8 (100%)	0 (0%)	0 (0%)	8 (100%)
HIV+/ART-exp	12 (75.0%)	0 (0%)	12 (50.0%)	24 (60.0%)	0 (0%)
HIV+/ART-naive	4 (25.0%)	0 (0%)	10 (41.7%)	14 (35.0%)	0 (0%)
CD4 count					
Median [Min, Max]	167 [32.0, 387]	NA	120 [23.0, 1120]	140 [23.0, 1120]	NA
HIV viral load					
Median [Min, Max]	0 [0, 37300]	NA	189 [20.0, 8320000]	61.0 [0, 8320000]	NA
B Symptoms					
No	8 (50.0%)	2 (25.0%)	4 (16.7%)	12 (30.0%)	2 (25.0%)
Yes	7 (43.8%)	6 (75.0%)	13 (54.2%)	20 (50.0%)	6 (75.0%)
Stage					
I-II	10 (62.5%)	2 (25.0%)	2 (8.3%)	12 (30.0%)	2 (25.0%)
III-IV	6 (37.5%)	6 (75.0%)	14 (58.3%)	20 (50.0%)	6 (75.0%)
ECOG score					
0-1	13 (81.3%)	3 (37.5%)	8 (33.3%)	21 (52.5%)	3 (37.5%)
2-4	3 (18.8%)	5 (62.5%)	9 (37.5%)	12 (30.0%)	5 (62.5%)
LDH					
Median [Min, Max]	453 [144, 1290]	427 [168, 845]	966 [271, 2500]	683 [144, 2500]	427 [168, 845]
COO (IHC)					
GC	13 (81.3%)	6 (75.0%)	11 (45.8%)	24 (60.0%)	6 (75.0%)
Non-GC	3 (18.8%)	2 (25.0%)	4 (16.7%)	7 (17.5%)	2 (25.0%)
Treatment					
Rituximab	2 (12.5%)	1 (12.5%)	3 (12.5%)	5 (12.5%)	1 (12.5%)
OS (months)					
Median [Min, Max]	26.2 [0.36, 60.0]	26.9 [0.66, 60.0]	2.38 [0.030, 44.0]	6.46 [0.030, 60.0]	26.9 [0.66, 60.0]

Table 1. Clinical and tumor characteristics by HIV status and cohort. Clinical and immunohistochemical data of sequenced tumors stratified by cohort and HIV/ART status. CD4 T-cell count, HIV viral load and lactate dehydrogenase (LDH) laboratory values at diagnosis. ECOG = Eastern Cooperative Oncology Group score. Cell-of-origin: activated B cell (ABC), germinal center B cell (GCB) or unclassified subtype. Overall survival was censored at 60 months.

single VCF file using vcf2maf. QC output is available as recommended per Dreval et al. in Supplemental Table 1²⁹.

Only variants that passed all filters from the individual callers were included. MAfTools (RRID: SCR_024519) was used for further variant filtering and analyses, including oncoplot generation. Only genes mutated in both cohorts were included. Variants were included if they met the following criteria: tumor and germline read depth > 15, germline alternative count < 1, tumor alternative count > 5, and variant allele frequency (VAF) > 5%. Silent, intronic and common variants were excluded. A variant was considered to have an effect if likely deleterious or deleterious by SIFT (RRID: SCR_012813), likely damaging or damaging by PolyPhen (RRID: SCR_013189) and/or HIGH impact by VEP. Finally, TMB was calculated using MAfTools tmb function and reported as number of mutations after standard filtering.

Epstein-Barr virus (EBV) status

EBV-encoded RNA in situ hybridization (EBER-ISH) was performed for MW samples to determine EBV status, with > 10% tumor staining considered positive. As EBER-ISH was not available for the tumors from SA, we performed Kraken taxonomic sequence classification (RRID: SCR_005484) on all tumors and used the MW tumors with both EBER-ISH and Kraken data as a benchmark.

Tumor subtyping

LymphGen (<https://lmpp.nih.gov/lymphgen/index.php>) was applied to classify the tumors according to their genetic characteristics: BN2, EZB, MCD, N1, and ST2¹⁷. BCL2 and BCL6 translocation statuses were not

available, and sequencing depth did not allow for copy number variant consideration (required for A53 subtype assignment).

Microsatellite instability prediction and mismatch repair IHC

Microsatellite instability (MSI) was predicted using Microsatellite Analysis for Normal-Tumor InStability (MANTIS). MANTIS RepeatFinder was used for bed file generation (hg38) and bedtools intersect was used to filter for targets (Agilent v6 for MW samples, Agilent v8 for SA samples). MANTIS was run for all 48 paired tumor/germline samples (min read quality 20, min locus quality 25.0, min locus coverage 30, min repeat reads 1) and Step-Wise Difference above 0.4 was considered microsatellite instability-high³⁰.

Statistical tests

Wilcoxon rank-sum test was used to compare differences by cohort and HIV/ART status for continuous variables. Chi-square test was used to examine associations between categorical variables. Cox regression was used to calculate hazard ratio (HR) and p-value. Fisher's exact test with false discovery rate (FDR) correction was used to determine differentially mutated genes. All statistical analyses were performed using R version 4.4.1.

Results

Comparison of clinical characteristics

Patients in the MW cohort had superior OS compared to those in the SA cohort (HR=0.37, $p=0.008$, Cox regression) (Fig. 1A). Further stratification by ART status demonstrated improved survival of HIV+/ART-naïve compared to HIV- patients in the MW cohort (HR=0.58, $p=0.043$, Cox regression), while ART status was not predictive of survival in the SA cohort (Supplemental Fig. 1). The proportion of patients with stage III or IV disease was higher in the SA cohort ($p=0.036$, Chi square), although male/female ratios and the presence/absence of B symptoms were similar between the cohorts (Table 1). Lactate dehydrogenase (LDH) levels and European Cooperative Oncology Group (ECOG) performance status were higher in SA patients compared to MW patients (LDH: median 966 vs. 453, $p=0.001$; ECOG: median 2 vs. 1, $p=0.028$), even when considering only HIV+ patients (LDH: median 982 vs. 453, $p=0.005$; ECOG: median 2 vs. 1, $p=0.007$) (Supplemental Fig. 2). In the MW cohort, there was a trend toward higher LDH levels in HIV+/ART-exp patients compared to HIV+/ART-naïve ($p=0.08$, Fig. 1B). Meanwhile, in the SA cohort, HIV+/ART-naïve patients had higher ECOG scores compared to HIV+/ART-exp (median 3 vs. 1, $p=0.024$).

Among HIV+ patients, CD4 counts were similar between cohorts ($p=0.93$) but HIV viral load was higher in the SA cohort compared to MW cohort (median 188 vs. 0 copies/mL, $p=0.006$) (Supplemental Fig. 2A, B). In the SA cohort, HIV+/ART-exp patients had higher CD4 counts compared to HIV+/ART-naïve ($p=0.05$) and decreased HIV viral load ($p=0.002$) (Fig. 1C, D). Meanwhile, in the MW cohort, only HIV viral load was significantly different by ART status ($p=0.01$) (Fig. 1C, D). PWH in both cohorts received similar ART regimens, the most common being Tenofovir disoproxil (TDF), Lamivudine (3TC), Efavirenz (EFV) (7 MW, 9 SA) and the second most common being TDF/3TC/Dolutegravir (DTG) (3 MW, 5 SA). Three patients from the MW cohort received protease inhibitors (Supplemental Fig. 2).

Commonly mutated genes in both cohorts

The median number of variants per sample was 498, with *TTN*, a large gene known to be frequently mutated in tumor exome studies, being the most recurrently mutated (62% samples).³¹ First, we used a curated lymphoma gene panel to determine the frequency of known DLBCL mutations in our combined cohort. Considering only variants predicted to have an effect with VAF > 10%, *BIRC6*, *KMT2D*, and *CARD11* were the most frequently mutated, with mutations in *ARID1A*, *MYD88*, and *TP53* also being common (Fig. 2A).

Although not validated or specifically developed for HIV+ DLBCL, we applied the LymphGen genomic classification tool to our samples: 4 BN2, 1 BN2/MCD, 3 EZB, 1 EZB/MCD/ST2, 1 EZB/N1, 2 EZB/ST2, 1 MCD, 3 N1, 6 ST2, and the remaining 25 samples (53%) were assigned "Other" (Supplemental Fig. 3). Of genes included in the LymphGen classifier, *EZH2* was mutated in four MW samples (1 HIV-, 2 HIV+/ART-exp, 1 HIV+/ART-naïve), but no SA samples, and *MYC* was mutated in three SA samples (1 HIV+/ART-exp, 2 HIV+/ART-naïve), but no MW samples.

South African HIV+/ART-exp tumors have lower TMB

MW samples had higher tumor mutational burden (TMB) overall (6-fold, $p<0.001$), though samples from SA exhibited the second and third highest TMB values. There was no difference in TMB between HIV+/ART-exp and HIV+/ART-naïve DLBCL ($p=0.94$). Within the MW cohort, HIV+/ART-exp tumors had higher TMB compared to HIV- ($p=0.031$) (Fig. 2B). Comparing the cohorts, SA HIV+/ART-exp had lower TMB than MW HIV+/ART-exp ($p<0.001$) (Fig. 2C).

Differential mutations by HIV/ART status

Gene mutations enriched in HIV+/ART-naïve compared to HIV+/ART-exp DLBCL included: *BOD1L1* and *EML4* (6/14 ART-naïve, 0/23 ART-exp, FDR=0.001), *IGHJ5*, *NCOA6*, *TEP1*, *KCNB2*, (all 5/14 ART-naïve, 1/23 ART-exp, FDR=0.02), among others (Fig. 3A). Meanwhile, HIV+/ART-exp tumors were enriched for mutations in *DUSP2* (8/23 ART-exp, 0/14 ART-naïve, FDR=0.02) and *PIM1* (7/23 ART-exp, 0/14 ART-naïve, FDR=0.03) (Fig. 3A).

EBV-/HIV+ tumors enriched for KMT2D mutations

Computationally assigning EBV status using Kraken, six tumors had EBV read count > 200 and were considered EBV+ for this study (Supplemental Fig. 4). EBV status was not associated with OS ($p=0.25$), though

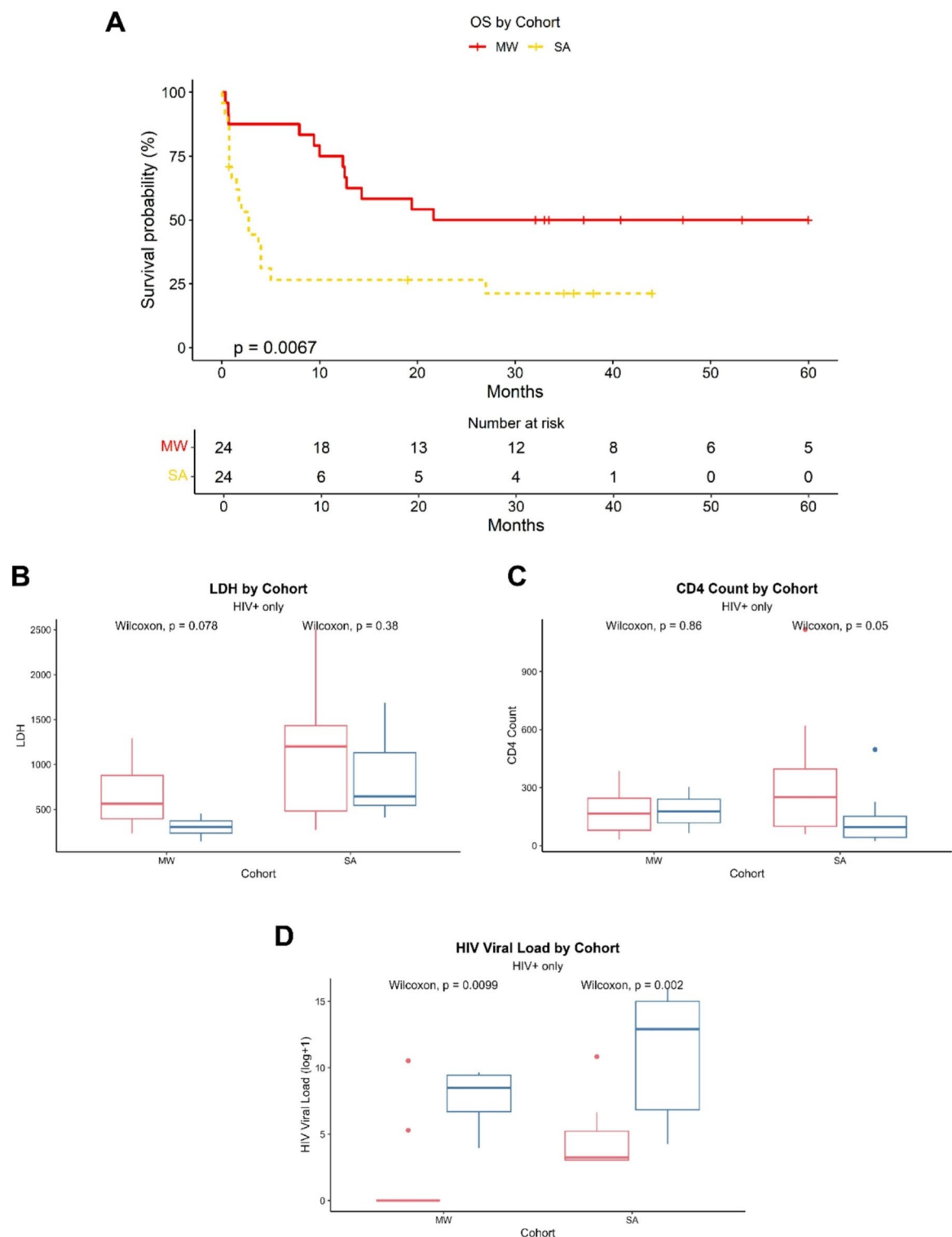


Fig. 1. Clinical characteristics by HIV/ART status and cohort. **(A)** Overall survival (OS) by cohort. **(B)** Lactate Dehydrogenase (LDH) by HIV/ART status and cohort. **(C)** CD4 Count by HIV/ART status and cohort. **(D)** HIV viral load by HIV/ART status and cohort. **(B–D)** Comparisons by Wilcoxon rank-sum test, $n = 40$, HIV+ only. MW Malawi, SA South Africa.

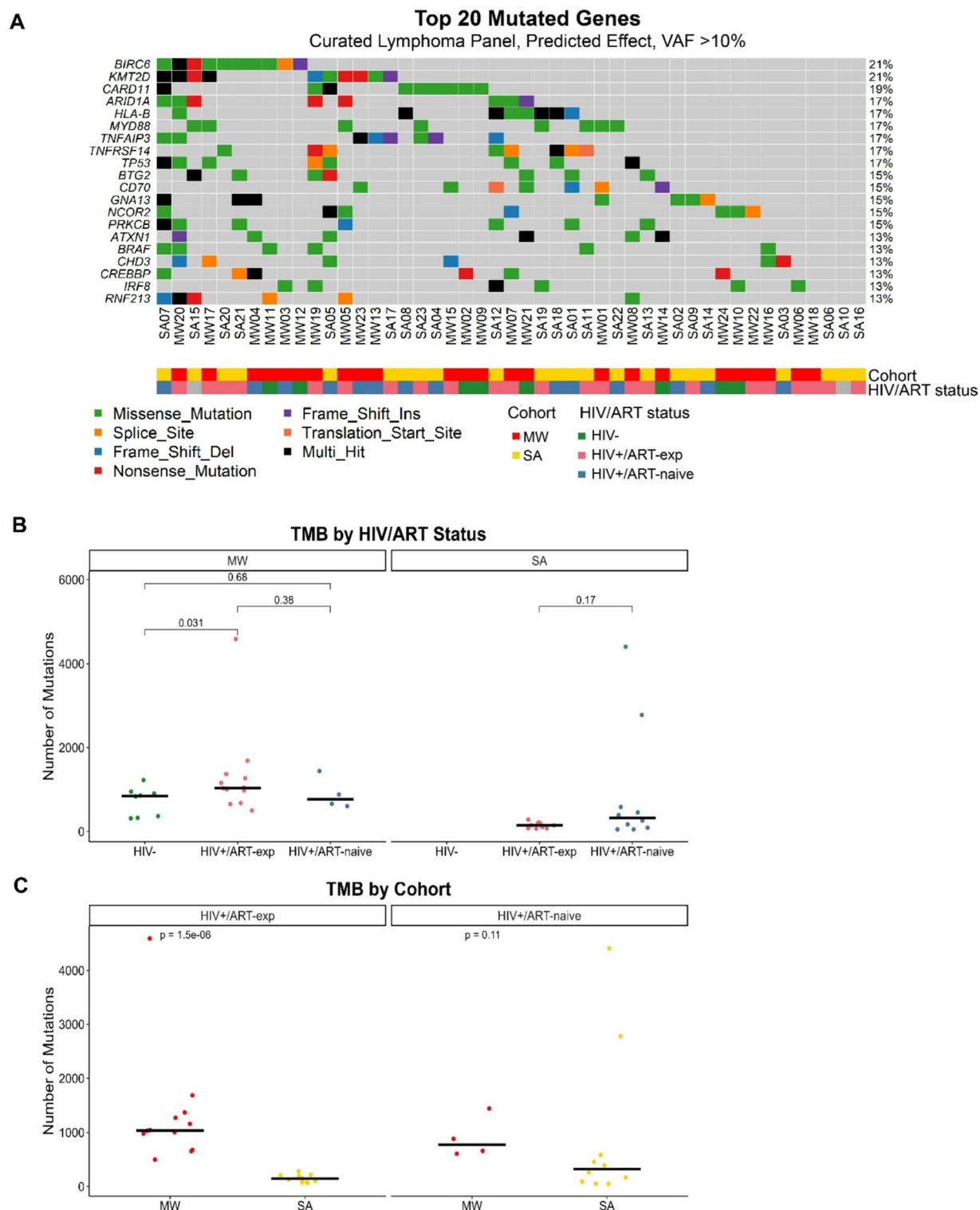


Fig. 2. Recurrently mutated genes. **(A)** OncoPrint of the top 20 mutated genes across the combined cohort. Considering only genes from curated lymphoma panel, only variants predicted to have an effect with VAF > 10%³². **(B)** TMB by HIV/ART status, stratified by cohort. Pairwise Wilcoxon rank-sum test. **(C)** TMB by cohort, stratified by HIV/ART status. HIV+ tumors only. Pairwise Wilcoxon rank-sum test. **(B, C)** Whole exome, VAF > 5%, $n = 38$ HIV+ tumors with known ART status.

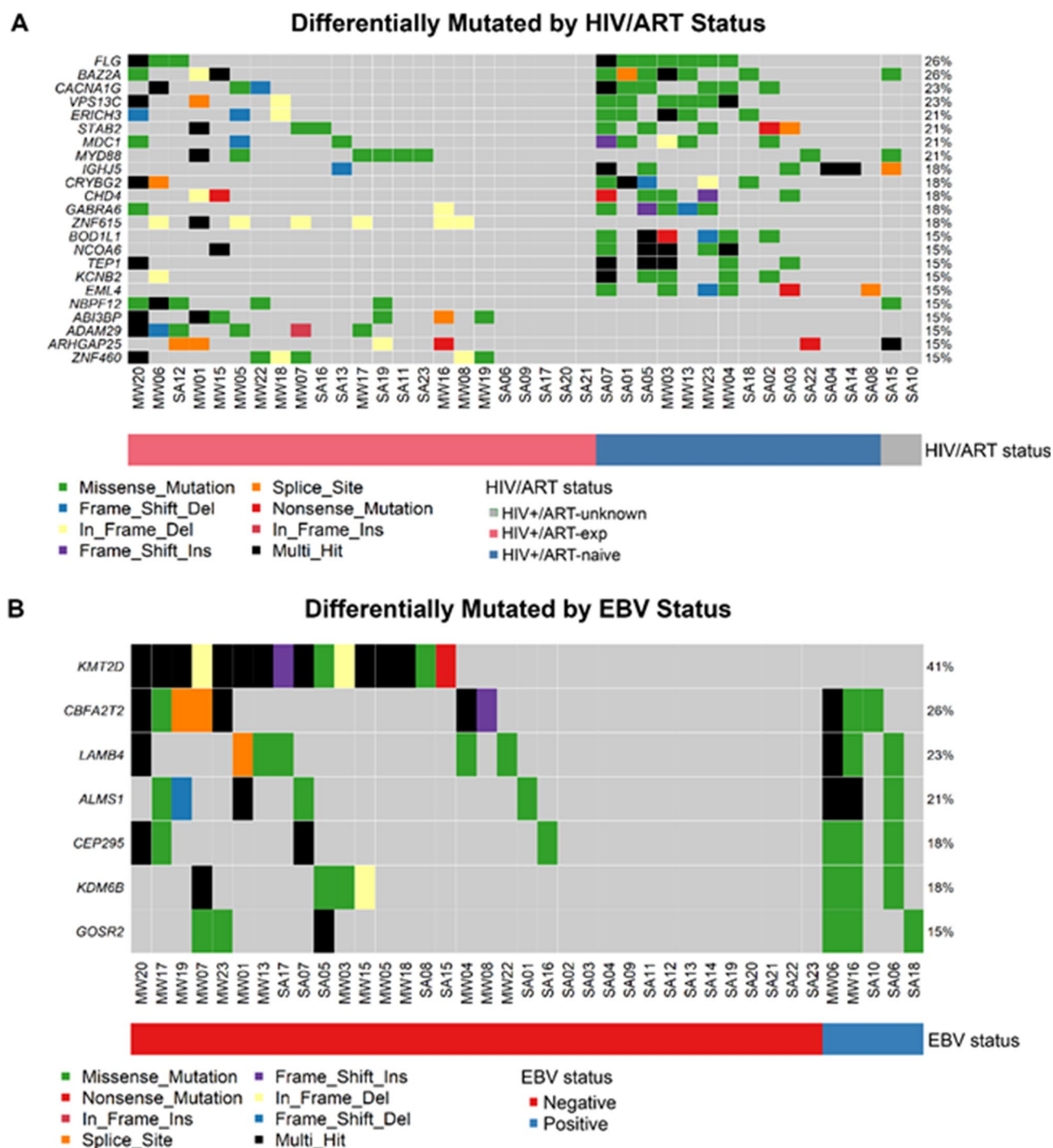


Fig. 3. Differentially mutated genes by HIV/ART status and EBV status. **(A)** Differentially mutated genes by antiretroviral therapy (ART) status (FDR < 0.05, Fisher's exact test). **(B)** Differential mutational profile by Epstein-Barr virus (EBV) status, as determined by Kraken (FDR < 0.1, Fisher's exact test). **(A, B)** HIV+ tumors only, whole exome, VAF > 5%.

EBV+ patients in the MW cohort had comparatively poor survival (< 10 months) (Supplemental Fig. 4). EBV-/HIV+ tumors were enriched for *KMT2D* mutations (0/5 EBV+, 16/34 EBV-, FDR = 0.07). Meanwhile, EBV+/HIV+ tumors were enriched for several genes compared to EBV-/HIV+ tumors including *GOSR2* (3/5 EBV+, 3/34 EBV-, FDR = 0.0192), *CEP295* (3/5 EBV+, 4/34 EBV-, FDR = 0.0321), and *KDM6B* (3/5 EBV+, 4/34 EBV-, FDR = 0.0321) (Fig. 3B). Of note, *STAT3* mutations were present exclusively in EBV-/HIV+ tumors.

Shared MYD88 variants among HIV+/ART-exp DLBCL across both cohorts

We identified shared pathogenic variants, defined as those predicted to be deleterious by SIFT or PolyPhen, with a variant allele frequency (VAF) exceeding 20% and excluding in-frame and splice site alterations, present in at least two tumor samples (Fig. 4A). Of these, five were present in at least one tumor from each cohort: *IGLL5* p.L28M, *MTMR1* p.R389Q, *MYD88* p.S206C, *PABPC3* p.K231E, *MYD88* p.S206C was found exclusively in HIV+/ART-exp tumors across the combined cohort, and *MYD88* p.S238N was found in two MW HIV+/ART-exp tumors.

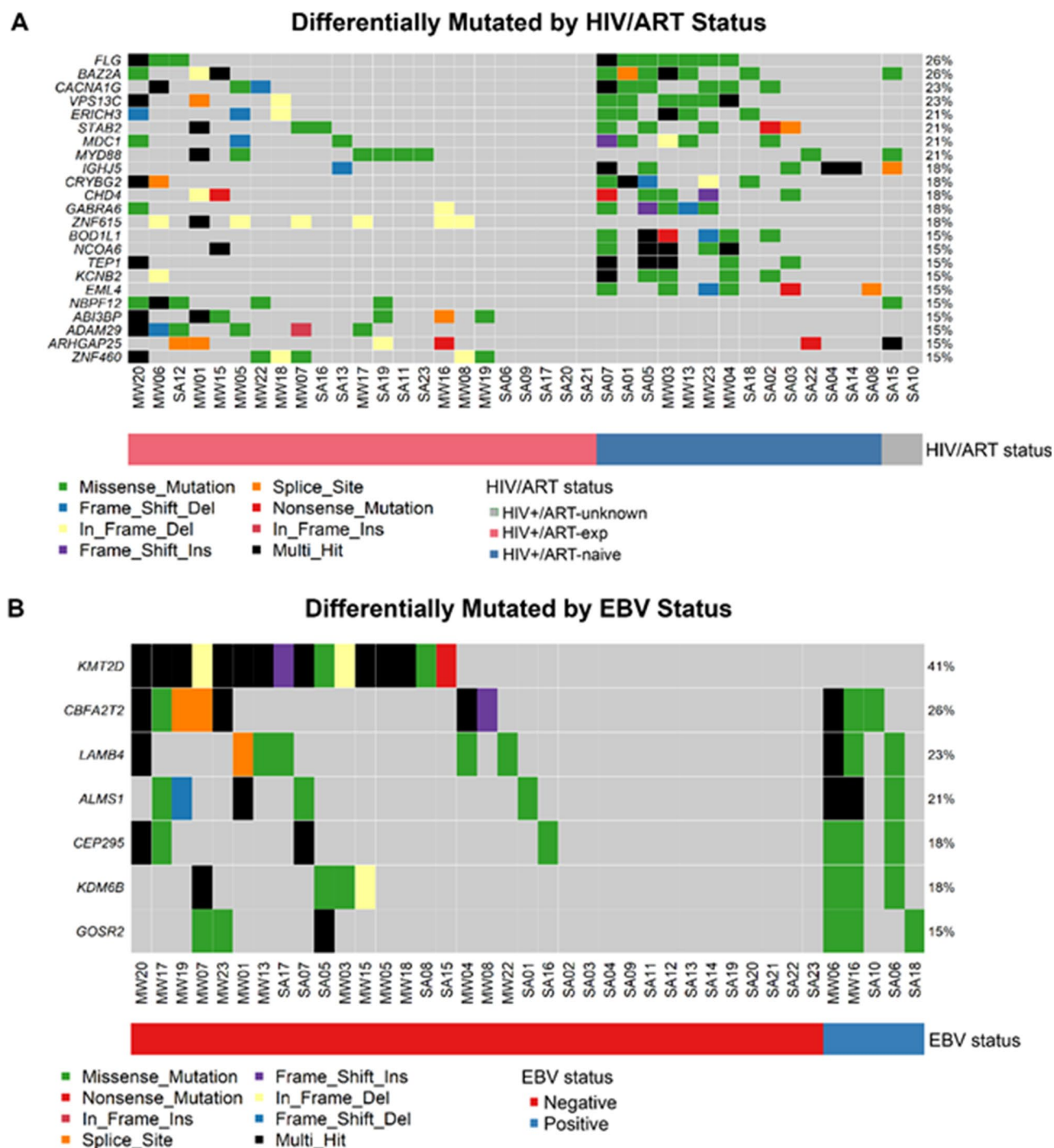


Fig. 4. Shared and prognostic mutations among HIV + tumors. (A) Shared high variant allele frequency (VAF), predicted deleterious variants by antiretroviral therapy (ART) status. (B) Kaplan-Meier curve depicting overall survival (OS) by *ANKRD11* mutation status and cohort. (C) Kaplan-Meier curve depicting OS by *ARID1B* mutation status and cohort.

ANKRD11 mutations associated with improved survival in HIV + DLBCL

We investigated prognostic associations of genes mutated in at least five HIV + patients. Ten genes showed significant associations with OS ($p < 0.05$). Of these, five were mutated in two or more samples from each cohort: *MUC4*, *TTN*, *ANKRD11*, *ARID1B*, *RYR3*. Only *ANKRD11* (HR=3.34, $p=0.054$) and *ARID1B* (HR=2.58, $p=0.086$) mutations remained marginally prognostic after adjusting for cohort (Fig. 4B, C). Both *ANKRD11* (4.8-fold, $p=0.0087$) and *ARID1B* (6.4-fold change, $p < 0.001$) mutation associated with higher TMB compared to non-mutated HIV + DLBCL samples.

High mutational burden phenotype with microsatellite instability and PMS2 and ARID1A mutation

Given several tumors exhibited over five times the median TMB, we assessed microsatellite instability (MSI) for all samples, and eight samples (6 MW, 2 SA) were predicted to be MSI-high, including the three tumors with the highest TMB (MW20, SA07 and SA05). Using a continuous metric, MSI correlated with TMB ($R=0.76$, $p < 0.001$, Pearson correlation) (Fig. 5A). MSI-high tumors were enriched for *TTF2* mutations (6/8 MSI-high vs. 0/8 MSI-low, FDR<0.001). We detected mutations in mismatch repair genes in 7/8 MSI-high samples. Interestingly, both MW20 and SA07 had *PMS2* (MW20: p.T89P VAF 63%, SA07: p.L239P VAF 44%) and *ARID1A* mutations (MW20: p.Y948H VAF 42% and p.Y518H VAF 9%, SA07: p.Y1226H VAF 36%) (Fig. 5B, C). Of note, both tumors harbored *ARID1A* mutations altering tyrosine to histidine and *PMS2* mutations changing to proline. MSI-high samples as well as additional samples with mutations in DNA mismatch repair genes or *ARID1A* are displayed in Table 2.

Discussion

This study offers the first multinational somatic mutational profiling of HIV + DLBCL from Africa, advancing our understanding of a patient population historically underrepresented in genomic profiling studies. By integrating data from patient cohorts in MW and SA, we identified shared genomic variants and microsatellite instability in HIV + DLBCL, and further suggest that ART exposure may impact the mutational profile. These findings add to biological understanding of lymphomagenesis in the context of HIV and ART, and are essential for ensuring PWH benefit from emerging therapies minimize cytotoxicity and improve outcomes for patients with relapsed/refractory disease.

Clinical parameters differed significantly between the two cohorts, with SA patients showing poorer HIV control, elevated LDH, higher ECOG scores at diagnosis, and inferior survival. Although this study did not have an epidemiological focus, these observed disparities in clinical outcomes between the cohorts likely stem from variations in access to HIV care, healthcare infrastructure, and time to diagnosis and treatment. Previous studies have shown contrasting survival patterns: MW HIV+/ART-naïve patients demonstrated better survival compared to HIV+/ART-exp patients, while SA data showed the opposite pattern^{19,33}. It is possible that the analyzed MW cohort may be enriched for HIV+/ART-naïve individuals with favorable biologic characteristics who survive long enough to access care. Additionally, exclusive presence of *EZH2* mutations in the MW cohort and *MYC* mutations in the SA cohort may provide further insight or support for these survival disparities and warrant further investigation³⁴.

The differential mutation patterns between HIV+/ART-exp and HIV+/ART-naïve tumors supports the hypothesis that ART exposure shapes the mutational profile of HIV + DLBCL. HIV+/ART-exp tumors were enriched for *DUSP2* and *PIM1* mutations, among others. In contrast, HIV+/ART-naïve tumors displayed a distinct profile with mutations in *BOD1L1* and *EML4*, both DNA damage response genes. These findings support prior evidence that HIV infection promotes lymphomagenesis through increased DNA damage, highlighting a potential mechanistic link between HIV and lymphoma development³⁵. Although EBV is commonly assumed to be implicated in many HIV + DLBCL cases, only 15% of HIV + tumors were EBV + in this study³⁶. Among the HIV + tumors, EBV- tumors were enriched for *KMT2D* mutations, consistent with a prior study identifying histone-modifying gene mutations in EBV-/HIV + DLBCL.³² However, unlike previous reports, *STAT3* mutations in our cohort occurred exclusively in EBV- cases³².

Our analysis also identified several shared, high VAF deleterious variants across both cohorts, including *IGLL5* (p.L28M), *MYD88* (p.S206C), and *MTMR1* (p.R389Q) which only occurred in HIV+/ART-exp tumors. These shared variants suggest there are common mutational drivers in HIV + DLBCL in southern Africa, which may represent therapeutic targets or biomarkers for this population. Additionally, *ANKRD11* mutations were positively prognostic in our study, consistent with findings from a study of EBV + HIV- DLBCL³⁷. Interestingly, while the latter study found *ANKRD11* mutations in 32% of EBV + HIV- DLBCL, none of the *ANKRD11* mutations in our cohorts were found in EBV+ tumors, indicating dysregulation of a more general immune mechanism, rather than an EBV-specific one³⁷.

Finally, the identification of two HIV + tumors across the two cohorts, both characterized by high TMB and MSI, with similar predicted pathogenic mutations in *ARID1A* and *PMS2*—genes associated with DNA mismatch repair deficiency—is particularly intriguing^{38,39}. MW20 was HIV+/ART-exp with 37 months of ART exposure prior to DLBCL diagnosis, and SA07 was HIV+/ART-naïve but defaulted on their ART regimen prior to diagnosis, indicating they might have had substantial exposure to ART. These findings are consistent with previous work linking DNA damage to chronic immune dysregulation as driver for lymphomagenesis, and define a subset of HIV + DLBCL patients that may benefit from targeted therapies^{40–43}.

In conclusion, this study begins to fill the critical gap in our understanding of the molecular mechanisms driving HIV + DLBCL in Africa. Our findings highlight the importance of ART exposure in shaping the mutational landscape of HIV + DLBCL, with potential implications for developing tailored therapeutic strategies. Understanding the role of DNA repair defects in HIV + DLBCL could inform strategies for selecting patients who may benefit from therapies targeting DNA damage, such as PARP inhibitors or other DNA repair

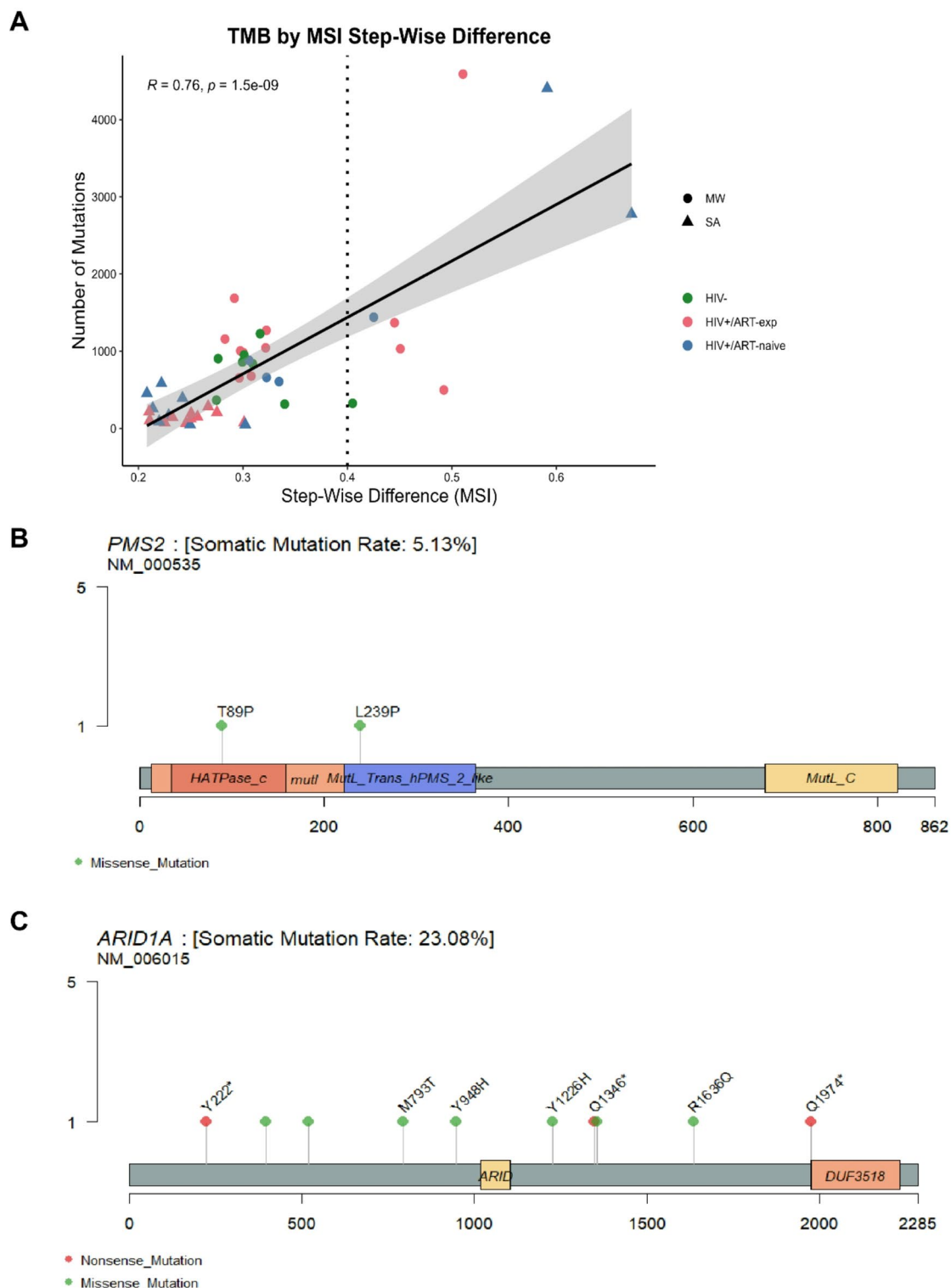


Fig. 5. Evidence of microsatellite-instability among HIV + DLBCL. **(A)** Tumor mutational burden (TMB) by continuous microsatellite instability (MSI) metric using Pearson correlation. Dotted line indicates MSI-high (step-wise difference ≥ 0.4). Shape indicates cohort and color indicates HIV/ART status. $n = 46$, samples with known HIV/ART status. **(B)** Lollipop plot of *PMS2* mutations. **(C)** Lollipop plot of *ARID1A* mutations. **(B, C)** All labelled variants occurred at a VAF $> 10\%$ and were predicted to have an effect.

Sample	TMB	Variant (VAF)	HIV/ART status
MW20*	4591	<i>PMS2</i> p.T89P (62%) [#] <i>ARID1A</i> p.Y948H (41%) [#] <i>ARID1A</i> p.Y518H (9%) [#]	HIV+/ART-exp
SA07*	4407	<i>PMS2</i> p.L239P (44%) [#] <i>ARID1A</i> p.Y1226H (36%) [#]	HIV+/ART-naive
SA05*	2778	<i>MLH1</i> p.A29G (63%) [#] <i>MSH2</i> p.A636Hfs*49 (25%) [#]	HIV+/ART-naive
MW01	1686	<i>ARID1A</i> p.G397R (8%) [#]	HIV+/ART-exp
MW03*	1441	<i>MLH1</i> p.X196_splice (12%) [#]	HIV+/ART-naive
MW07*	1369	<i>ARID1A</i> p.R1636Q (21%) [#] <i>MSH6</i> p.R298* (6%) [#]	HIV+/ART-exp
MW05	1271	<i>ARID1A</i> p.Q1346* (16%) [#]	HIV+/ART-exp
MW06*	1031	<i>MSH2</i> p.Q170E (44%)	HIV+/ART-exp
MW19	980	<i>ARID1A</i> p.Y222* (12%) [#]	HIV+/ART-exp
MW24	951	<i>MSH6</i> p.L743Ffs*13 (35%) [#]	HIV-
MW12	860	<i>MSH2</i> p.R96H (7%) [#]	HIV-
MW11	837	<i>ARID1A</i> p.T1359_Q1367del (5%)	HIV-
MW16	678	<i>MSH2</i> (12%) [#]	HIV+/ART-exp
MW04	660	<i>ARID1A</i> p.F1354S (7%) [#]	HIV+/ART-naive
MW21*	325	<i>MSH6</i> p.A587P (40%) [#] <i>ARID1A</i> p.Q507Vfs*117 (24%) [#]	HIV-
SA12	283	<i>ARID1A</i> p.M793T (62%) [#]	HIV+/ART-exp
SA15	165	<i>ARID1A</i> p.Q1974* (35%) [#]	HIV+/ART-unknown

Table 2. Samples with mutations in DNA mismatch repair genes or *ARID1A*. *MSI-high by MANTIS. [#]Variant predicted to have an effect.

modulators. However, the relatively small number of cases, differences in health systems, and potential selection biases limit the generalizability of our results to all HIV + DLBCL patients across the continent. Broader studies including additional geographic regions and larger sample sizes will be essential to confirm these observations and assess their relevance to the wider population. As the burden of HIV in Africa remains high, these insights are crucial for ensuring that emerging therapies are accessible and effective for HIV + DLBCL patients in low-resource settings.

Data availability

Filtered variants are included as supplemental data, genomic data from the Malawi and South Africa cohorts are available in the BioProject repositories, PRJNA1203246 and PRJNA1281316, respectively. Code available by request from corresponding author.

Received: 12 February 2025; Accepted: 3 July 2025

Published online: 09 July 2025

References

- Sharma, R. et al. Mapping Cancer in africa: A comprehensive and comparable characterization of 34 Cancer types using estimates from GLOBOCAN 2020. *Front. Public. Health.* **10**, 839835. <https://doi.org/10.3389/fpubh.2022.839835> (2022).
- Larkin, H. D. Cancer deaths May double by 2030 in Sub-Saharan Africa. *JAMA* **327** (23), 2280. <https://doi.org/10.1001/jama.2022.10019> (2022).
- Gopal, S. & Sharpless, N. E. Cancer as a global health priority. *JAMA* **326** (9), 809. <https://doi.org/10.1001/jama.2021.12778> (2021).
- Nugent, A. et al. Reporting of race in genome and exome sequencing studies of cancer: a scoping review of the literature. *Genet. Sci.* **21** (12), 2676–2680. <https://doi.org/10.1038/s41436-019-0558-2> (2019).
- Fatumo, S. et al. A roadmap to increase diversity in genomic studies. *Nat. Med.* **28** (2), 243–250. <https://doi.org/10.1038/s41591-021-01672-4> (2022).
- Lurain, K. Treating Cancer in people with HIV. *JCO* **41** (21), 3682–3688. <https://doi.org/10.1200/JCO.23.00737> (2023).
- UNAIDS & Global, A. I. D. S. : The path to ending AIDS as a public health threat by 2030. UNAIDS. Published online July 22, 2024. (2024). https://www.unaids.org/en/resources/presscentre/pressreleaseandstatementarchive/2024/july/20240722_global-aid-s-update
- Harries, A. D. et al. Act local, think global: how the Malawi experience of scaling up antiretroviral treatment has informed global policy. *BMC Public. Health.* **16** (1), 938. <https://doi.org/10.1186/s12889-016-3620-x> (2016).
- Trickey, A. et al. Estimation of improvements in mortality in spectrum among adults with HIV receiving antiretroviral therapy in High-Income countries. *JAIDS J. Acquir. Immune Defic. Syndr.* **95** (1S), e89–e96. <https://doi.org/10.1097/QAI.0000000000003326> (2024).
- Horner, M. J., Shiels, M. S., Pfeiffer, R. M. & Engels, E. A. Deaths attributable to Cancer in the US human immunodeficiency virus population during 2001–2015. *Clin. Infect. Dis.* **72** (9), e224–e231. <https://doi.org/10.1093/cid/ciaa1016> (2021).
- Yarchoan, R. & Uldrick, T. S. HIV-Associated cancers and related diseases. *N Engl. J. Med.* **378** (11), 1029–1041. <https://doi.org/10.1056/NEJMra1615896> (2018).

12. Gibson, T. M., Morton, L. M., Shiels, M. S., Clarke, C. A. & Engels, E. A. Risk of non-Hodgkin lymphoma subtypes in HIV-infected people during the HAART era: a population-based study. *AIDS* **28** (15), 2313–2318. <https://doi.org/10.1097/QAD.0000000000000428> (2014).
13. Blattner, W. A. & Nowak, R. G. Epidemiology of AIDS-Defining malignancies. In: (eds Hope, T. J., Richman, D. D. & Stevenson, M.) *Encyclopedia of AIDS*. Springer New York; :503–513. doi:https://doi.org/10.1007/978-1-4939-7101-5_20 (2018).
14. Huguet, M., Navarro, J. T., Moltó, J., Ribera, J. M. & Tapia, G. Diffuse large B-Cell lymphoma in the HIV setting. *Cancers (Basel)* **15** (12), 3191. <https://doi.org/10.3390/cancers15123191> (2023).
15. Noy, A. Optimizing treatment of HIV-associated lymphoma. *Blood* **134** (17), 1385–1394. <https://doi.org/10.1182/blood-2018-01-791400> (2019).
16. Chapuy, B. et al. Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. *Nat. Med.* **24** (5), 679–690. <https://doi.org/10.1038/s41591-018-0016-8> (2018).
17. Wright, G. W. et al. A probabilistic classification tool for genetic subtypes of diffuse large B cell lymphoma with therapeutic implications. *Cancer Cell* **37** (4), 551–568.e14. <https://doi.org/10.1016/j.ccell.2020.03.015> (2020).
18. Chaddburn, A. et al. Immunophenotypic analysis of AIDS-related diffuse large B-cell lymphoma and clinical implications in patients from AIDS malignancies consortium clinical trials 010 and 034. *J. Clin. Oncol.* **27** (30), 5039–5048. <https://doi.org/10.1200/JCO.2008.20.5450> (2009).
19. Coelho, J. et al. HIV and prior exposure to antiretroviral therapy alter tumour composition and tumour: T-cell associations in diffuse large B-cell lymphoma. *Br. J. Haematol.* **205** (1), 194–206. <https://doi.org/10.1111/bjh.19531> (2024).
20. Fedoriv, Y. et al. Identifying transcriptional profiles and evaluating prognostic biomarkers of HIV-associated diffuse large B-cell lymphoma from Malawi. *Mod. Pathol.* **33** (8), 1482–1491. <https://doi.org/10.1038/s41379-020-0506-3> (2020).
21. Roush, S. M. et al. HIV infection and ART exposure affect tumor TCR repertoire of diffuse large B cell lymphoma. *JCI Insight* **9** (13), e180771. <https://doi.org/10.1172/jci.insight.180771> (2024).
22. Lee, M. J. et al. Genome-defined African ancestry is associated with distinct mutations and worse survival in patients with diffuse large B-cell lymphoma. *Cancer* **126** (15), 3493–3503. <https://doi.org/10.1002/cncr.32866> (2020).
23. Shenoy, P. J. et al. Racial differences in the presentation and outcomes of diffuse large B-cell lymphoma in the united States. *Cancer* **117** (11), 2530–2540. <https://doi.org/10.1002/cncr.25765> (2011).
24. Gopal, S. et al. Early experience after developing a pathology laboratory in malawi, with emphasis on cancer diagnoses. *PLoS One* **8** (8), e70361. <https://doi.org/10.1371/journal.pone.0070361> (2013).
25. Painschab, M. S. et al. Mature outcomes and prognostic indices in diffuse large B-cell lymphoma in malawi: a prospective cohort. *Br. J. Haematol.* **184** (3), 364–372. <https://doi.org/10.1111/bjh.15625> (2019).
26. Kimani, S. et al. Safety and efficacy of rituximab in patients with diffuse large B-cell lymphoma in malawi: a prospective, single-arm, non-randomised phase 1/2 clinical trial. *Lancet Global Health* **9** (7), e1008–e1016. [https://doi.org/10.1016/S2214-109X\(21\)00181-9](https://doi.org/10.1016/S2214-109X(21)00181-9) (2021).
27. Gondwe, Y. et al. Comparison of baseline lymphoma and HIV characteristics in Malawi before and after implementation of universal antiretroviral therapy. *PLoS One* **17** (9), e0273408. <https://doi.org/10.1371/journal.pone.0273408> (2022).
28. Hans, C. P. Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. *Blood* **103** (1), 275–282. <https://doi.org/10.1182/blood-2003-05-1545> (2004).
29. Dreval, K., Boutros, P. C. & Morin, R. D. Minimal information for reporting a genomics experiment. *Blood* **140** (24), 2549–2555. <https://doi.org/10.1182/blood.2022016095> (2022).
30. Kautto, E. A. et al. Performance evaluation for rapid detection of pan-cancer microsatellite instability with MANTIS. *Oncotarget* **8** (5), 7452–7463. <https://doi.org/10.18632/oncotarget.13918> (2017).
31. Shyr, C. et al. FLAGS, frequently mutated genes in public exomes. *BMC Med. Genomics* **7** (1), 64. <https://doi.org/10.1186/s12920-014-0064-y> (2014).
32. Chapman, J. R. et al. EBV-positive HIV-associated diffuse large B cell lymphomas are characterized by JAK/STAT (STAT3) pathway mutations and unique clinicopathologic features. *Br. J. Haematol.* **194** (5), 870–878. <https://doi.org/10.1111/bjh.17708> (2021).
33. Vaughan, J., Perner, Y. & Wiggill, T. Diffuse large B-Cell lymphoma in the Public-Sector of Johannesburg, South Africa, in the era of widescale antiretroviral therapy use. *JAIDS J. Acquir. Immune Defic. Syndr.* **91** (4), 335–342. <https://doi.org/10.1097/QAI.0000000000003069> (2022).
34. Isshiki, Y. et al. EZH2 Inhibition enhances T cell immunotherapies by inducing lymphoma immunogenicity and improving T cell function. *Cancer Cell Published Online Dec.* **2024**:S1535610824004410. <https://doi.org/10.1016/j.ccell.2024.11.006>
35. Lurain, K., Ramaswami, R. & Yarchoan, R. The role of viruses in HIV-associated lymphomas. *Semin Hematol.* **59** (4), 183–191. <https://doi.org/10.1053/j.seminhematol.2022.11.002> (2022).
36. Shindiapina, P., Ahmed, E. H., Mozhenkova, A., Abebe, T. & Baiocchi, R. A. Immunology of EBV-Related lymphoproliferative disease in HIV-Positive individuals. *Front. Oncol.* **10**, 1723. <https://doi.org/10.3389/fonc.2020.01723> (2020).
37. Gebauer, N. et al. Genomic insights into the pathogenesis of Epstein-Barr virus-associated diffuse large B-cell lymphoma by whole-genome and targeted amplicon sequencing. *Blood Cancer J.* **11** (5), 102. <https://doi.org/10.1038/s41408-021-00493-5> (2021).
38. Ge, H. et al. Mismatch repair deficiency is associated with specific morphologic features and frequent loss of ARID1A expression in ovarian clear cell carcinoma. *Diagn. Pathol.* **16** (1), 12. <https://doi.org/10.1186/s13000-021-01071-w> (2021).
39. Sun, D. et al. ARID1A deficiency associated with MMR deficiency and a high abundance of tumor-infiltrating lymphocytes predicts a good prognosis of endometrial carcinoma. *Transl Oncol.* **33**, 101685. <https://doi.org/10.1016/j.tranon.2023.101685> (2023).
40. Tian, T. et al. Microsatellite instability and its associations with the clinicopathologic characteristics of diffuse large B-cell lymphoma. *Cancer Med.* **9** (7), 2330–2342. <https://doi.org/10.1002/cam4.2870> (2020).
41. Han, B. et al. An unusual case of microsatellite instability-high/deficient mismatch repair (MSI-H/dMMR) diffuse large B-cell lymphoma revealed by targeted gene sequencing. *J. Pathol. Transl Med.* **56** (2), 92–96. <https://doi.org/10.4132/jptm.2021.10.15> (2022).
42. Duval, A. et al. The mutator pathway is a feature of immunodeficiency-related lymphomas. *Proc. Natl. Acad. Sci. USA* **101** (14), 5002–5007. <https://doi.org/10.1073/pnas.0400945101> (2004).
43. de Miranda, N. F. C. C. et al. DNA repair genes are selectively mutated in diffuse large B cell lymphomas. *J. Exp. Med.* **210** (9), 1729–1742. <https://doi.org/10.1084/jem.20122842> (2013).

Acknowledgements

Thank you to the patients and their families for participating in the research studies. This work was made possible through the support of Kamuzu Central Hospital, the Malawi Ministry of Health, and the University of North Carolina Project-Malawi, as well as Chris Hani Baragwanath Academic Hospital. Funding for this research was provided by the National Institutes of Health (CA016086-46S3 to UNC/YF, UNC-Integrated Translational Oncology Program T32-CA244125 to UNC/SMR, U54CA254564 to UNC/YF/SMR), UNC Lineberger Comprehensive Cancer Center (Developmental Funding Award to YF) and The National Institute For Communicable Diseases (NICD) of South Africa (Funding provided by Arshad Ismail).

Author contributions

SMR, MiMo, JV and YF conceived and designed the analyses; SMR and MiMo analyzed and interpreted the data with assistance from JC and SB and guidance from JV and YF; SMR drafted the manuscript, and JC, MiMo, JV and YF revised the manuscript; AC, MP and EK acquired clinical data for the Malawi cohort; JV acquired clinical data for the South African cohort; TT, MaMp, MaMu and YF acquired pathologic data for the Malawi cohort; JV acquired pathologic data for the South African cohort; SM and HvZ performed sequencing of South African cohort; all authors provided edits and final approval of the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-10529-6>.

Correspondence and requests for materials should be addressed to Y.F.

Reprints and permissions information is available at 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 4.0 International License, which permits use, sharing, adaptation, 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 changes were made. 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/4.0/>.

© The Author(s) 2025