Identifying transcriptional profiles and evaluating prognostic biomarkers of HIV-associated diffuse large B-cell lymphoma from Malawi

Yuri Fedoriw[1] · Sara Selitsky[1] · Nathan D. Montgomery[1] · Sviatoslav M. Kendall[2] · Kristy L. Richards[3] · Wei Du[3] · Tamiwe Tomoka[4,5] · Maurice Mulenga[6] · Joel S. Parker[1] · Sandeep S. Dave[2] · Satish Gopal[1,4]

Abstract

Lymphoma incidence in sub-Saharan Africa (SSA) is increasing due to HIV and population aging. Diffuse Large B-cell lymphoma (DLBCL), the most common lymphoma in SSA and worldwide, is highly associated with HIV, but molecular studies of HIV-associated DLBCL are scarce globally. We describe profiling of DLBCL from Malawi, aiming to elucidate tumor biology and identify clinically meaningful biomarkers specifically for SSA. Between June 1, 2013 and June 1, 2016, 59 cases of DLBCL (32 HIV+/27 HIV−) enrolled in the Kamuzu Central Hospital Lymphoma Study were characterized, of which 54 (92%) were negative for Epstein–Barr virus. Gene expression profiling (GEP) by whole transcriptome sequencing was performed on the first 36 cases (22 HIV+/14 HIV−). Immunohistochemistry (IHC) and GEP results were compared with published data and correlated to clinical outcome and pathologic features. Unsupervised clustering strongly segregated DLBCL by HIV status ($p = 0.0003$, Chi-squared test), indicating a marked contribution of HIV to expression phenotype. Pathway analysis identified that HIV-associated tumors were enriched in hypoxia, oxidative stress, and metabolism related gene expression patterns. Cell-of-origin subtype, determined by sequencing and IHC, did not associate with differences in overall survival (OS), while Ki-67 proliferation index ≥80% was associated with inferior OS in HIV+ DLBCL only ($p = 0.03$) and cMYC/BCL2 co-expression by IHC was negatively prognostic across the entire cohort ($p = 0.01$). This study provides among the first molecular characterizations of DLBCL from SSA, demonstrates marked gene expression differences by HIV status, and identifies genomic and immunophenotypic characteristics that can inform future basic and clinical investigations.

Introduction

Lymphoma incidence in sub-Saharan Africa (SSA) is increasing due to epidemic levels of HIV infection,

Deceased: Kristy L. Richards

✉ Yuri Fedoriw
yuri.fedoriw@unchealth.unc.edu

1 University of North Carolina, Chapel Hill, NC, USA

2 Duke Cancer Institute and Center for Genomic and Computational Biology, Durham, NC, USA

3 Cornell University, New York, NY, USA

4 UNC Project-Malawi, Lilongwe, Malawi

5 University of Malawi College of Medicine, Lilongwe, Malawi

6 Kamuzu Central Hospital, Lilongwe, Malawi

population growth, and aging [1–3]. Diffuse Large B-cell Lymphoma (DLBCL), the most common lymphoma worldwide and in SSA, is highly associated with HIV, but thorough studies of HIV-associated DLBCL are globally scarce. While the striking genetic heterogeneity of sporadic DLBCL in HIV-naive patients has been extensively studied [4–10], this work has been challenging to conduct in HIV-positive patients, as prospective, clinically annotated cohorts of HIV-associated lymphoma are uncommon in settings where HIV infection is most frequent. Independent studies in HIV-positive patients may provide unprecedented and generalizable insight into lymphoma biology and inform prevention and treatment strategies regionally and worldwide. Moreover, as treating patients with DLBCL is now often possible and safe in SSA, risk stratification is of paramount importance in a region where supportive care is limited and endemic opportunistic burden is high [11, 12]. Prognostic and predictive biomarkers

of DLBCL that are widely accepted in resource-rich regions and incorporated into the current classification schemes have not been effectively studied in resource-limited settings [13]. When such efforts have been undertaken for lymphoma in SSA, they have been limited by incomplete ascertainment of HIV status, clinical outcomes, and nonstandardized treatment [14]. Whether or not these markers are valid in settings characterized by the distinct genetic, environmental, and socioeconomic pressures of SSA remains uncertain.

Herein we describe whole transcriptome sequencing of DLBCL cases from the ongoing Kamuzu Central Hospital (KCH) Lymphoma Study in Lilongwe, Malawi, where HIV burden is high and DLBCL treatment and follow-up are standardized for enrolled patients. The study affords unique opportunities to investigate genomic differences related to HIV status, and to assess the applicability of well-recognized prognostic biomarkers in the context of regional resource limitations.

Materials and methods

Patient selection and treatment

Patients were enrolled in the KCH Lymphoma Study (NCT02835911) after pathologic diagnosis and clinical screening, as previously described [11, 15]. The prospective observational study enrolls all newly diagnosed patients with confirmed lymphoproliferative disorders at the national teaching hospital in Malawi's capital, Lilongwe. CD4, HIV RNA viral load, and antiretroviral therapy (ART) status were documented for all HIV-infected patients, as were lymphoma-related clinical and laboratory data. Tissue biopsies were performed at KCH and processed in the on-site pathology laboratory, where diagnoses are issued after weekly multidisciplinary telepathology conferences between clinicians and pathologists in Malawi and pathologists at the University of North Carolina (UNC) [15, 16]. After primary diagnosis, the pretreatment, formalin-fixed and paraffin embedded (FFPE) tissue blocks were submitted to UNC for additional assessment by immunohistochemistry (IHC), and gene expression profiling (GEP) by whole transcriptome sequencing (RNA-seq). IHC and GEP results were compared with published expression data and correlated to clinical outcome and pathologic features. Reflecting the regional standard of care for DLBCL in most of SSA, patients were treated with cyclophosphamide, doxorubicin, vincristine, prednisone chemotherapy, and concurrent ART if HIV-positive. Rituximab is not routinely available in the Malawi public sector. All participants were followed until death, or administrative censoring on September 30, 2017. No patients were lost to follow-up.

RNA sequencing

RNA was extracted from diagnostic, pretreatment FFPE tumor blocks using the Ribo-Zero kit (Illumina, San Diego, CA) per manufacturer's recommendation. RNA libraries were prepared with the Illumina TruSeq RNA Preparation Kit v2 and sequenced by the Illumina HiSeq2000 and NextSeq. MapSplice v2.0.1.9 [17] was used for RNA read alignment to hg19 and transcript quantification was performed using RSEM v1.1.13 [18]. We used median-adjustment for batch correction, and all data were normalized using upper quartile normalization and log2 transformation.

Cluster assignment and gene expression analysis

Samples were clustered using the 1500 most variable genes with a median normalized count >10 in ConsensusClusterPlus [19], with a maximum cluster assignment (k) of 6, with 50 iterations for 80% of the samples. Based on the consensus cumulative density function, samples were divided into two clusters.

Gene set enrichment analysis was performed using GSVA to compute a module score and then linear regression to test associations, using the Hallmark gene set from Molecular Signatures Database (http://software.broadinstitute.org/gsea/msigdb/index.jsp) [20, 21]. Pathways found to be prognostic in this cohort were evaluated in a recent publication of sequenced de novo DLBCL from Reddy et al. using their processed RNA-seq matrix to calculate gene sets modules using GSVA [10].

We used two algorithms to calculate the cell-of-origin (COO): (1) the algorithm described in Wright et al. [7] which provided categorical assignment as germinal center (GC) or post-germinal center activated B-cell (ABC), and (2) using genes from Wright et al. we subtracted the mean of the median centered genes that were upregulated from the mean of the median centered genes that were down-regulated. The latter provided a continuous value for the COO measure.

Immunohistochemistry and in situ hybridization

Primary diagnosis was aided by manually performed IHC using antibodies available in Lilongwe, Malawi: CD3 (clone PS1), CD20 (clone L26), CD30 (clone 15B3), CD45 (code NCL-L-LCA-RP), CD138 (clone MI15), BCL2 (clone bcl2/100/D5), Ki-67 (Clone MM1), TdT (Clone TdT-338), and HHV8 (NCL-HHV8-LNA), from Leica Biosystems (Buffalo Grove, IL, US). In the United States, additional IHC and in situ hybridization (ISH), when necessary, was performed on a Leica Bond platform (Leica Biosystems) according to manufacturer's instructions. COO

was assigned by IHC using the algorithm described by Hans et al. using CD10 (clone NCL-CD10-270) and BCL6 (code PA0204) from Leica Biosystems, and MUM1 (code M7259m) from Dako (Carpinteria CA, USA) [22]. Expression of BCL2 (clone 124) and cMYC (clone y69) was assessed by IHC using antibodies from Ventana Medical Systems (Tucson, AZ, USA) performed on the Ventana Discovery Ultra. cMYC staining of >40% of neoplastic cells together with BCL2 expression in >70% was interpreted as positive staining and defined the "double-protein expressers" (DPE) [23, 24]. Ki-67 was quantified by light microscopy in 5% increments.

Statistical analysis

Expression cluster assignment association with HIV status was measured by Chi-squared test, and Kaplan–Meier curves were used to assess overall survival (OS) for expression data using Cox proportional hazards model. Patient clinical characteristics and IHC differences between HIV-infected and HIV-uninfected patients were compared by Mann–Whitney U Test (for continuous data) or by Fisher Exact Test (categorical data). The log-rank test, and corresponding hazard ratio (HR) and confidence interval (CI), was used to assess differences in survival between clinical and IHC subgroups using GraphPad Prism 8 (San Diego, CA, USA).

Results

Between June 1, 2013 and June 1, 2016, 59 cases of DLBCL were enrolled in the KCH Lymphoma Study and fully pathologically characterized at UNC; 32 (54%) of these arising in HIV-infected individuals. Clinical characteristics and EBV tumor status as defined by EBER-ISH staining, are listed in Table 1. The International Prognostic Index (IPI) at diagnosis was not different between HIV-positive and HIV-negative patients. EBV was infrequent in the cohort, identified by EBER ISH in two (7%) of HIV-positive DLBCL cases and three (10%) of HIV-negative cases. For HIV-infected patients, median CD4 count was 117.5 cells/μl, with 60% on ART at time of enrollment.

Whole transcriptome and pathway analysis

Resources were available to perform RNA-seq on the first 36 cases of DLBCL. The clinical characteristics of this group did not differ significantly from the larger cohort (Table 1). An unsupervised cluster assignment strongly segregated DLBCL by HIV status (Chi-squared

Table 1 Clinical and pathologic characteristics of study patients and sequenced subset.

	HIV-positive ($n = 32$)	HIV-negative ($n = 27$)	p
Age, years, median (range)	45 (24–63)	55 (11–79)	0.0293
Male, n (%)	21 (66%)	17 (63%)	0.8118
International prognostic index, ≥3, n (%)	9 (28%)	7 (26%)	0.8644
OS, median (months)	11.8	13.7	0.9611
PFS, median (months)	7.4	8.1	0.7943
HIV viral load, $\log_{10}$ copies/ml, median (range)	3.39 (0 to >10,000)	–	–
ART at enrollment, n (%)	19 (60%)	–	–
Duration on ART, years (range)	4.1 (0.1–13.2)	–	–
CD4 count, cells/μl (range)	117.5 (32–1013)	–	–
EBV+, n (%)	2 (6%)	3 (11%)	0.4914
COO IHC, n (%)	32	27	
GC	18 (56%)	16 (60%)	0.7586
Non-GC	14 (44%)	11 (40%)	
Ki-67%	32	26	
≥80%	22 (71%)	13 (50%)	0.1078
<80%	10 (29%)	13 (50%)	
MYC/BCL2	24	17	
DPE+	9 (38%)	3 (18%)	0.1727
DPE−	15 (62%)	14 (82%)	

Sequenced cohort	HIV-positive ($n = 22$)	HIV-negative ($n = 14$)	p
Age, years, median (range)	47 (24–60)	55 (11–77)	0.1326
Male, n (%)	16 (73%)	9 (64%)	0.7159
International prognostic index, >3, n (%)	6 (27%)	5 (36%)	0.5729
OS, median (months)	11.8	11.6	0.3808
PFS, median (months)	8.6	7.9	0.4527
HIV viral load, $\log_{10}$ copies/ml, median (range)	2.15 (0 to >10,000)	–	–
ART at enrollment, n (%)	12 (55%)	–	–
Duration on ART, years (range)	3.3 (0.1–8.3)	–	–
CD4 count, cells/μl (range)	104.5 (48–1013)	–	–
EBV+, n (%)	0 (0%)	3 (21%)	0.0270

OS overall survival, *PFS* progression free survival, *ART* antiretroviral therapy, *COO* cell-of-origin by *Hans* classifier, *IHC* immunohistochemistry, *GC* germinal center type, *non-GC* nongerminal center type, *DPE* double-protein co-expressers by IHC.
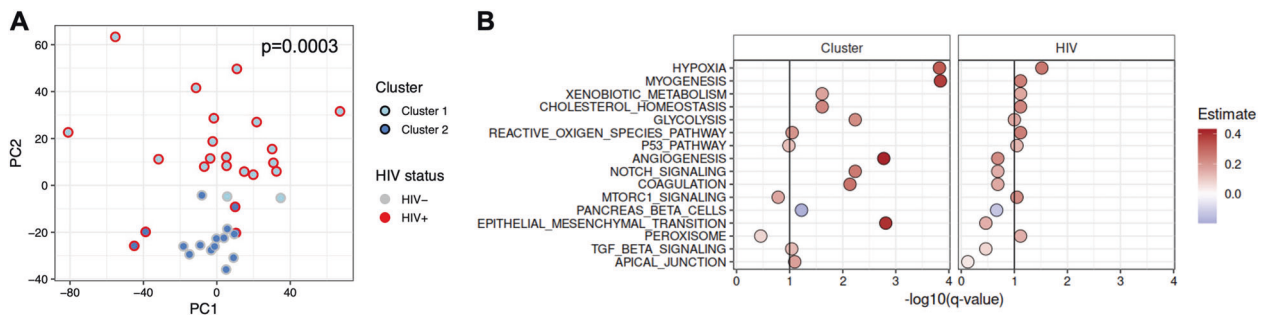
Fig. 1 Transcriptome analysis of DLBCL. a Principle component analysis (PCA) of expression differences between HIV+ and HIV− DLBCL. **b** Gene sets differentially expressed in sequenced DLBCL related to cluster assignment (left) or HIV status (right). Gene set included if it was significant by either HIV status or cluster assignment (q value < 0.1). Color of dot represents the coefficient. Positive is greater in cluster 2 (enriched for HIV+) and negative is greater in cluster 1 (enriched for HIV−).

test, $p = 0.0003$), with 18 of 22 HIV-positive cases (82%) clustering together (Fig. 1a). A total of 2,523 genes were differentially expressed between the clusters with a false discovery rate (FDR) adjusted p value of <0.1. Of note, 3 of 4 HIV-associated DLBCLs that clustered with the HIV-negative cases were on relatively long durations of ART prior to DLBCL diagnosis (range 38–98 months) with suppressed HIV viral loads. The outlying HIV-associated DLBCL patients did not show significant differences with respect to CD4 count, viral load, or other clinical variables.

We performed gene set enrichment using the 50 Hallmark gene sets and found that HIV and cluster assignment had a few differentially regulated modules in common, such as hypoxia and metabolic genes (q values < 0.1 by linear regression, FDR adjustment, Fig. 1b). However, even though only six samples were discordant between HIV status and cluster assignment, there was a much stronger signal for differential regulation by cluster assignment of angiogenesis (HIV status q value = 0.2, cluster assignment q value = 0.002), Notch signaling (HIV status q value = 0.2, cluster assignment q value = 0.006), and epithelial mesenchymal transition (HIV status q value = 0.3, cluster assignment q value = 0.002) gene sets.

DLBCL COO was determined by GEP using the Wright et al. algorithm [7], with 18 cases defined as GC, 13 as ABC, and 5 as "unclassified" (Fig. 2a). We next calculated a single COO score using our own algorithm (see Methods) using the genes described by Wright et al. [7] (Fig. 2b), which enabled us to create a continuous metric, as opposed to a categorical result. By this method, the "unclassified" cases grouped more closely with ABC DLBCL, as previously reported [22, 25]. By GEP, HIV-positive DLBCL cases trended toward the GC-type (GC $n = 13$ vs. $n = 5$ ABC), while HIV-negative DLBCL cases were more evenly distributed (GC $n = 6$ vs. ABC $n = 8$, $p = 0.0934$, Chi-square test HIV+ vs. HIV−, Fig. 2c). Of "unclassified" DLBCL cases, defined by the Wright et al. algorithm, 4 of 5 were HIV-associated.

Clinical outcomes and association with gene expression

IPI was associated with mortality in our cohort (data not shown; $p < 0.0001$), as previously reported for all DLBCLs in the KCH Lymphoma Study [11, 12]. However, neither HIV status, expression cluster, nor COO classification was associated with OS differences (Fig. 3). Hallmark gene expression modules including unfolded protein response, MYC pathways, KRAS signaling, interleukin (IL)-6 and IL- 2 signaling, coagulation, and angiogenesis, were prognostic across all sequenced cases ($p < 0.1$, Cox proportional hazards regression model, Fig. 4). These gene expression modules were also significantly prognostic in a large cohort of de novo DLBCL in the US ($q < 0.05$, FDR multiple testing correction) [10]. Gene expression modules related to interferon gamma (IFNγ) and IFN alpha (IFNα) signaling were positively prognostic in the HIV-positive cases only ($p < 0.05$, Cox proportional hazards regression model), but not in the large cohort of HIV-negative DLBCL or the de novo DLBCL from the United States (Fig. 4).

Clinical outcomes associated with tumor marker expression by immunohistochemistry

Prognostic IHC markers of DLBCL widely used in resource-rich settings for clinical reporting were evaluated using expression data and the larger cohort of Malawi DLBCL. The IHC COO classifier was concordant with our expression algorithm in 14 of 17 (82%) GC cases, and 12 of 17 (71%) of ABC/non-GC cases (overall concordance: 76%, Fig. 5a). The IHC COO classifier was then applied to the larger cohort (Table 1). As with GEP, there were no OS differences related to COO subtype by IHC (Fig. 5b).

Semiquantitative assessment of tumor cell proliferation by Ki-67 and cMYC/BCL2 co-expression was assessed on available cases using IHC (Table 1). Among HIV-positive
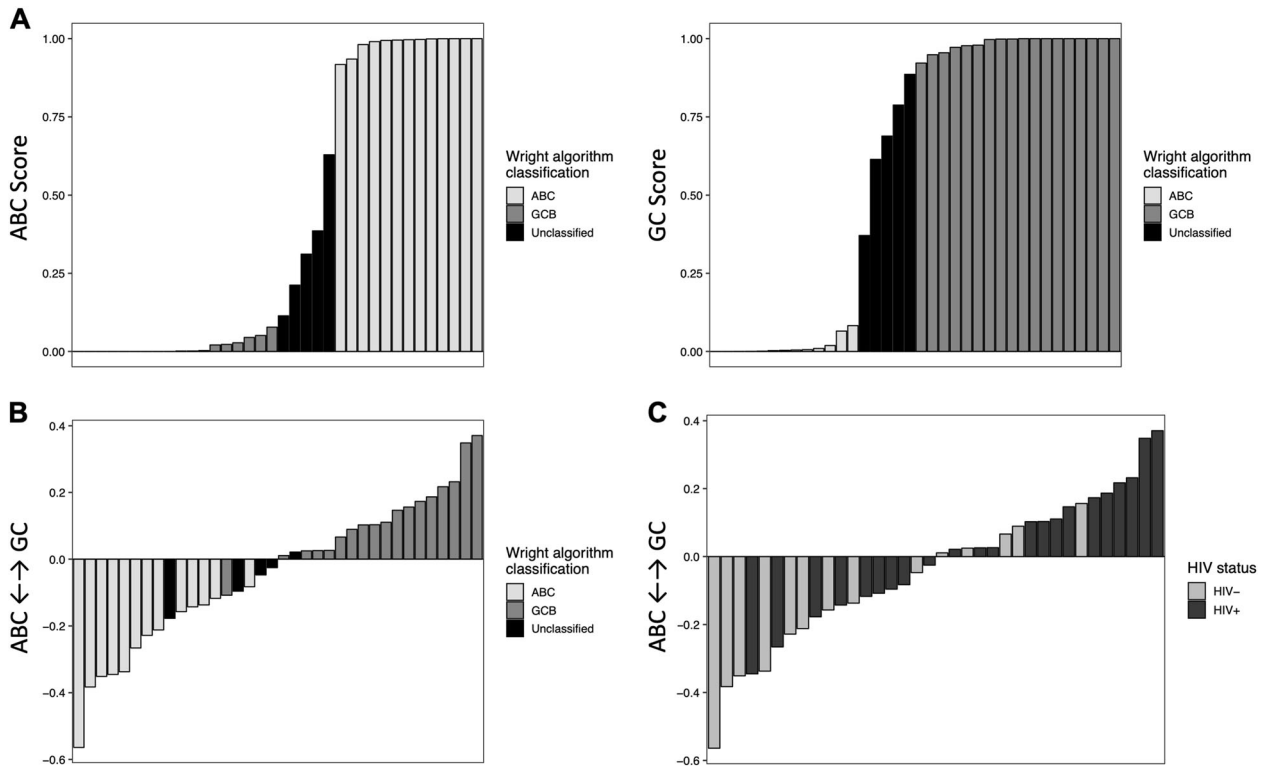
Fig. 2 Cell-of-origin (COO) analysis by transcriptional profile.
a Analysis by conventional Wright et al. algorithm (left: activated B-cell (ABC) score; right: germinal center (GC) score; "unclassified"

in black). **b** Waterfall plot analysis of ABC and GC gene expression and overlay from Wright et al. classifier. **c** Overlay of HIV status on to COO expression category.
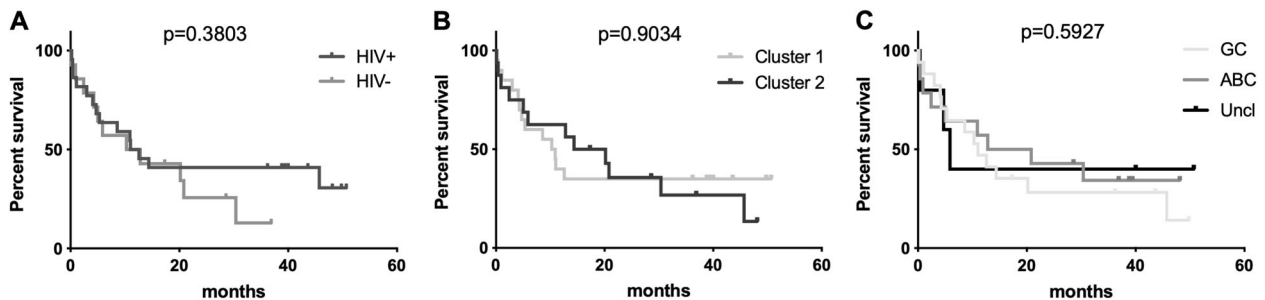

Fig. 3 Outcome associated with HIV status, expression cluster, and COO. Kaplan–Meier survival curves of DLBCL associated with HIV status (**a**), cluster assignment (**b**), and COO by GEP (**c**).

DLBCL cases, Ki-67 staining fractions of ≥80% was associated with inferior OS (median survival 8.57 months vs. not reached, $p = 0.03$; HR 2.845 with 95% CI 1.085–7.508, Fig. 5c). In the HIV-negative cohort, a similar trend related to proliferation index by IHC was identified ($p = 0.1$). Cases with cMYC/BCL2 co-expression (DPE by IHC assessment), representing 29% of all cases, also showed an inferior OS compared with those that did not co-express cMYC and/or BCL2 irrespective of HIV status (1.70 vs. 20.17 months; $p = 0.012$; HR 3.558 with 95% CI 1.322–9.576, Fig. 4d). DPE status did not associate with IHC COO ($p = 0.161$) (Fig. 5d).

Discussion

B-cell lymphomas in patients with HIV arise in the unique and heterogeneous context of varying degrees of ongoing HIV replication, immune dysregulation, and concurrent environmental pressures that are permissive for the acquisition of genetic lesions, transformation, and proliferation of malignant B cells [26, 27]. However, the molecular characterization of HIV-associated lymphomas has been a global challenge, and such methods to date have not been widely applied to HIV-positive patients. To our knowledge, this represents among the first published RNA-seq investigations

Fig. 4 Clinical outcomes associated with gene sets. Forest plots depicting hazard ratios and 95% confidence intervals for Hallmark gene expression modules with a nominal *p* value < 0.05 in all Malawi DLBCL (light gray) or Malawi DLBCL HIV-positive cases only (dark gray). These were compared with previously published de novo DLBCL [10] (black). The shape of the dot represents the *p* value status: *x* ≥ 0.1, square ≤ 0.01, diamond ≤ 0.05, and triangle ≤ 0.1.
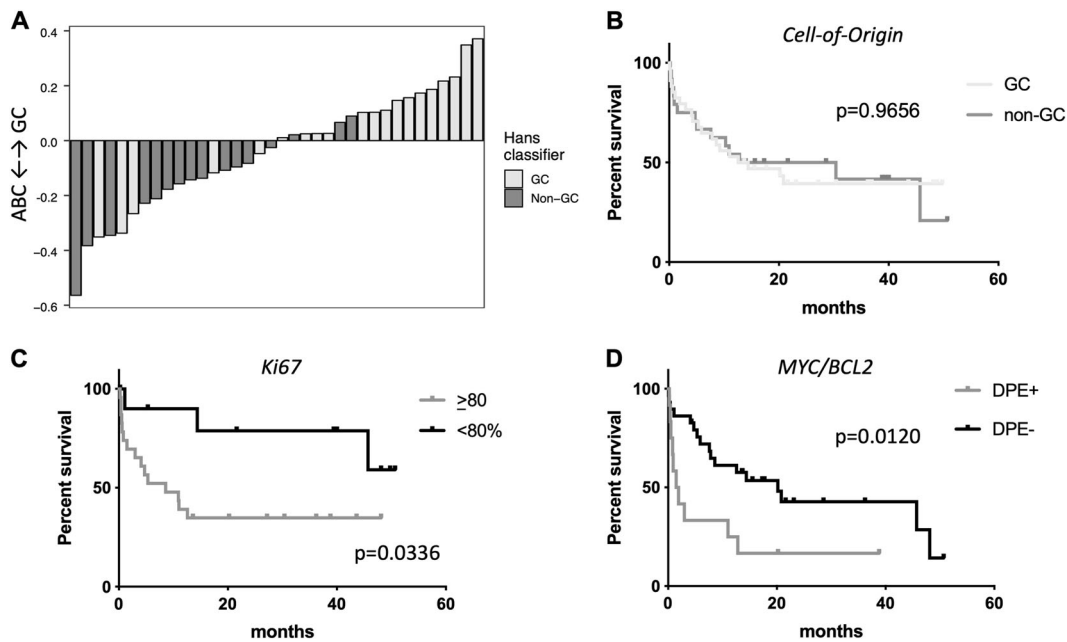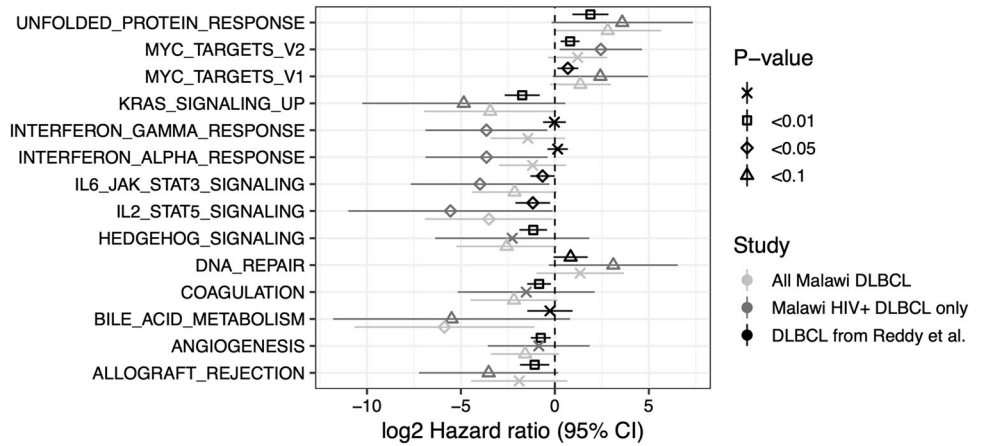

Fig. 5 Clinical outcomes associated with immunohistochemical markers. a Overlay of conventional Wright et al. COO algorithm with immunohistochemical assignment by Hans et al.; **b** overall survival (OS) by immunohistochemical COO classifier by Hans et al.; **c** OS of HIV+ DLBCL by Ki-67 staining fraction; **d** OS of all DLBCL by MYC/BCL2 double-protein co-expression (DEP) by IHC.

of HIV-associated DLBCL worldwide, remarkably from one of the economically poorest countries in the world.

In resource-rich countries, such studies are difficult to conduct as HIV prevalence is low, and large, prospective, clinically annotated cohorts of HIV-associated lymphomas with appropriate biospecimens are relatively uncommon. Furthermore, matched HIV-negative DLBCL control cases with otherwise similar patient characteristics apart from HIV have not been established. Conversely, in areas with high HIV burden, diagnostic, clinical, and research infrastructure often preclude accurate diagnosis, treatment standardization, measurement of known prognostic factors, and long-term follow-up to determine outcomes.

To address these gaps, and building on years of lymphoma research capacity investments in Malawi, we uniquely performed whole transcriptome analysis of prospectively enrolled HIV-positive and HIV-negative DLBCL cases from SSA who received standardized treatment with longitudinal follow-up, and validated biomarkers used for risk assessment from resource-rich regions. The confounder of EBV on lymphomagenesis, particularly in immuno-compromised hosts, was limited in this study, as all sequenced HIV-positive DLBCL cases were EBV-negative by EBER ISH. As previous molecular characterizations demonstrated significant genomic differences related to EBV infection, this cohort identifies transcriptional

differences in EBV-negative DLBCL in the HIV-infected population [28].

In this cohort, unsupervised hierarchical clustering of RNA-seq data demonstrated a strong contribution of HIV status to DLBCL expression phenotype, with the majority of HIV-positive DLBCL cases (82%) clustering together. Mechanisms underlying this phenomenon are unclear but may reflect systemic or microenvironmental pressures on lymphoma development or evolution in the unique setting of HIV infection. Compared with HIV-negative cases, HIV-associated DLBCL was enriched for hypoxia-induced genes and expression modules related to oxidative stress, and the expression cluster heavily enriched for HIV DLBCL also showed significant differences related to angiogenesis. These findings are in keeping with previously published histologic and phenotypic observations, demonstrating stromal and vascular differences in HIV-associated lymphoma [29].

While the strong association of expression profile with HIV status is itself remarkable, three of four HIV-positive DLBCL cases that clustered with the HIV-negative cases had relatively long durations of ART prior to lymphoma diagnosis, and cluster assignment, rather than HIV status, showed stronger differential regulation of gene sets associated with angiogenesis, Notch signaling, and epithelial mesenchymal transition gene sets. This suggests significant tumor microenvironment differences related to the immunologic and virologic environments in which the DLBLC occurs and raises the possibility that DLBCLs differ within the HIV-infected cohort based on the length of preceding HIV treatment. Thus, tumors developing in the context of long-term ART may have more in common with those arising in HIV-negative, than HIV-positive patients. The pattern of dysregulation supports a primitive wound-healing microenvironment in the HIV+ DLBCL expression cluster, characterized by fibrosis, hypoxia, and angiogenesis, that requires further investigation.

By univariate analysis, there were no survival differences with respect to HIV status or expression cluster, although the median OS in our setting is lower than in resource-rich settings [11, 12]. While validity of clinical prognostic scores (IPI) has been demonstrated previously in our cohort, prognostic biomarkers of disease have not been validated. As treatment resources for DLBCL become increasingly available in SSA where HIV infection rates are high, biomarker validation for DLBCL in SSA is critical to guide therapy, minimize treatment-related morbidity, appropriately allocate scarce resources, and direct future clinical trials and translational research.

By RNA-seq, we identified prognostic signatures related to unfolded protein response, MYC pathways, KRAS signaling, IL-6 and IL-2 signaling, coagulation and angiogenesis, which were similarly prognostic in a large study of de novo DLBCL of primarily US origin [10]. This association reinforces the global applicability of these studies and highlights the biological overlap of disease across geographic regions. Amongst HIV-positive DLBCL in our cohort, IFNγ and IFNα signaling was positively prognostic, suggesting that in HIV infection, higher IFN response associates with better DLBCL outcomes. Larger cohort studies are necessary to validate these expression pattern differences and further assess functional pathways.

Molecular profiling of sporadic, HIV-negative DLBCL has identified clinically meaningful prognostic expression signatures [5, 6, 8]. Nearly two decades ago, the COO subtypes in HIV-negative DLBCL were identified, differing in their genetic alterations, signaling pathways, and outcomes [4, 7, 9]. The ABC subset showed an inferior survival compared with GC-type DLBCL, but differentially altered pathways have highlighted potential therapeutic targets to improve outcomes for ABC-type DLBCL [30–33]. As comprehensive expression profiling of clinical tumor samples is not yet universally applied, immunohistochemical algorithms are often used as surrogates for DLBCL COO subtypes [22, 34–36] and may be more amenable to application in resource-limited settings. A common such classifier, originally published by Hans et al. shows acceptable correlation with GEP and is independently prognostic [22]. Studies of HIV-associated DLBCL have shown variable associations with COO subtypes, but the high prevalence of EBV in DLBCL of immunocompromised populations associates strongly with ABC subtype and confounds analyses of DLBCL in HIV-infected individuals [37, 38].

Although the cases in our cohort can be effectively stratified by COO subtype, both by GEP and IHC algorithms, COO status was not prognostic. This may reflect differences in underlying biology, but it is also likely that the risk imparted by COO subtype is confounded by non-biological patient and health system factors that influence survival in resource-limited settings more than in resource-rich environments. Of note, however, COO subtype using the IHC classifier was similarly not shown to associate with OS for HIV-associated DLBCL patients treated in United States AIDS Malignancy Consortium trials [37]. Moreover, when expression of genes that define COO are plotted as a continued variable, there appears to be an even distribution between those with a "high-ABC" to "high-GC" score, irrespective of HIV status (Fig. 2b). This gradient in COO expression may have biologic or treatment implications beyond simple dichotomization that requires further study.

Additional genomic alterations associate with COO and other DLBCL subtypes. DLBCL harboring rearrangements of *BCL2* and *cMYC* showed dramatically worse survival compared with cases without this genomic "double hit" [36], despite being GC-type [39, 40]. More recently,

DLBCL expressing both cMYC and BCL2 by IHC show a similarly poor prognosis, independent of rearrangements identified by fluorescence in situ hybridization (FISH) studies [23, 41]. The majority of these cMYC/BCL2 "double-protein expressers" (DPE) show an ABC immunophenotype [23]. In our cohort, DPE was associated with inferior OS irrespective of HIV status. While there was no association of DPE status with COO, the sample size is limited. Resource, technical and tissue limitations remain significant obstacles, and precluded FISH evaluation of *cMYC* and *BCL2* translocation status.

Finally, the prognostic significance of proliferative capacity, as measured by Ki-67 (MIB-1) IHC, has also been extensively investigated. While some have shown inferior outcomes associated with increased Ki-67 staining fractions [42–44], others, including analysis of HIV-associated DLBCL, have demonstrated the opposite [37, 45, 46]. The reasons for these differences across studies remain unclear. In our cohort, high Ki-67 proliferative index (≥80%) was associated with an inferior prognosis only for HIV-associated DLBCL. This finding differs from analyses of AIDS Malignancy Consortium trial participants in the United States [37], among whom improved survival was associated with high Ki-67 staining. However, there are many notable differences between the United States and Malawi, including treatment of many of the AIDS Malignancy Consortium patients with rituximab and/or continuous infusion chemotherapy regimens developed specifically to more effectively treat highly proliferative B-cell lymphomas [47–49].

To conclude, the unbiased bulk tumor transcriptomic analysis of DLBCL cases from Malawi uniquely identifies marked expression differences between HIV-positive and HIV-negative DLBCL. Prognostic differences related to cMYC/BCL2 co-expression and Ki-67 staining in HIV-positive DLBCL patients were identified, but COO status did not associate with outcome. These findings underscore the need for validation of HIV-specific and region-specific prognostic markers to inform clinical care. This work also suggests that greater understanding of unique aspects of lymphoma biology for HIV-infected patients in SSA is possible and should be an important regional research priority moving forward.

Compliance with ethical standards

References

1. Gopal S, Wood WA, Lee SJ, Shea TC, Naresh KN, Kazembe PN, et al. Meeting the challenge of hematologic malignancies in sub-Saharan Africa. Blood. 2012;119:5078–87.
2. Parkin DM, Nambooze S, Wabwire-Mangen F, Wabinga HR. Changing cancer incidence in Kampala, Uganda, 1991–2006. Int J Cancer. 2010;126:1187–95.
3. Chokunonga E, Borok MZ, Chirenje ZM, Nyakabau AM, Parkin DM. Trends in the incidence of cancer in the black population of Harare, Zimbabwe 1991–2010. Int J Cancer. 2013;133:721–9.
4. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature. 2000;403:503–11.
5. Shipp MA, Ross KN, Tamayo P, Weng AP, Kutok JL, Aguiar RC, et al. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. Nat Med. 2002;8:68–74.
6. Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, Fisher RI, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. N Engl J Med. 2002;346:1937–47.
7. Wright G, Tan B, Rosenwald A, Hurt EH, Wiestner A, Staudt LM. A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. Proc Natl Acad Sci USA. 2003;100:9991–6.
8. Lenz G, Wright G, Dave SS, Xiao W, Powell J, Zhao H, et al. Stromal gene signatures in large-B-cell lymphomas. N Engl J Med. 2008;359:2313–23.
9. Lenz G, Wright GW, Emre NC, Kohlhammer H, Dave SS, Davis RE, et al. Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways. Proc Natl Acad Sci USA. 2008;105:13520–5.

10. Reddy A, Zhang J, Davis NS, Moffitt AB, Love CL, Waldrop A, et al. Genetic and functional drivers of diffuse large B cell lymphoma. Cell. 2017;171:481–94.e15.

11. Gopal S, Fedoriw Y, Kaimila B, Montgomery ND, Kasonkanji E, Moses A, et al. CHOP Chemotherapy for Aggressive Non-Hodgkin Lymphoma with and without HIV in the Antiretroviral Therapy Era in Malawi. PLoS ONE. 2016;11:e0150445.

12. Painschab MS, Kasonkanji E, Zuze T, Kaimila B, Tomoka T, Nyasosela R, et al. Mature outcomes and prognostic indices in diffuse large B-cell lymphoma in Malawi: a prospective cohort. Br J Haematol. 2019;184:364–72.

13. Swerdlow S, Campo E, Harris N, Jaffe E, Pileri S, Stein H, et al. WHO classification of tumors of haematopoietic and lymphoid tissues. Revised 4th ed. Lyon: International Agency for Research on Cancer (IARC); 2017.

14. Morgan EA, Sweeney MP, Tomoka T, Kopp N, Gusenleitner D, Redd RA, et al. Targetable subsets of non-Hodgkin lymphoma in Malawi define therapeutic opportunities. Blood Adv. 2016;1: 84–92.

15. Montgomery ND, Liomba NG, Kampani C, Krysiak R, Stanley CC, Tomoka T, et al. Accurate real-time diagnosis of lymphoproliferative disorders in Malawi through clinicopathologic teleconferences: a model for pathology services in sub-Saharan Africa. Am J Clin Pathol. 2016;146:423–30.

16. Gopal S, Krysiak R, Liomba NG, Horner MJ, Shores CG, Alide N, et al. Early experience after developing a pathology laboratory in Malawi, with emphasis on cancer diagnoses. PLoS ONE. 2013;8:e70361.

17. Wang K, Singh D, Zeng Z, Coleman SJ, Huang Y, Savich GL, et al. MapSplice: accurate mapping of RNA-seq reads for splice junction discovery. Nucleic Acids Res. 2010;38:e178.

18. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinform. 2011;12:323.

19. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics. 2010;26:1572–3.

20. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database (MSigDB) hallmark gene set collection. Cell Syst. 2015;1:417–25.

21. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinform. 2013;14:7.

22. Hans CP, Weisenburger DD, Greiner TC, Gascoyne RD, Delabie J, Ott G, et al. Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. Blood. 2004;103:275–82.

23. Hu S, Xu-Monette ZY, Tzankov A, Green T, Wu L, Balasubramanyam A, et al. MYC/BCL2 protein coexpression contributes to the inferior survival of activated B-cell subtype of diffuse large B-cell lymphoma and demonstrates high-risk gene expression signatures: a report from The International DLBCL Rituximab-CHOP Consortium Program. Blood. 2013;121:4021–31.

24. Green TM, Young KH, Visco C, Xu-Monette ZY, Orazi A, Go RS, et al. Immunohistochemical double-hit score is a strong predictor of outcome in patients with diffuse large B-cell lymphoma treated with rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone. J Clin Oncol. 2012;30:3460–7.

25. Scott DW, Wright GW, Williams PM, Lih CJ, Walsh W, Jaffe ES, et al. Determining cell-of-origin subtypes of diffuse large B-cell lymphoma using gene expression in formalin-fixed paraffin-embedded tissue. Blood. 2014;123:1214–7.

26. Pantanowitz L, Carbone A, Dolcetti R. Microenvironment and HIV-related lymphomagenesis. Semin Cancer Biol. 2015;34: 52–7.

27. Carbone A. Emerging pathways in the development of AIDS-related lymphomas. Lancet Oncol. 2003;4:22–9.

28. Capello D, Scandurra M, Poretti G, Rancoita PM, Mian M, Gloghini A, et al. Genome wide DNA-profiling of HIV-related B-cell lymphomas. Br J Haematol. 2010;148:245–55.

29. Liapis K, Clear A, Owen A, Coutinho R, Greaves P, Lee AM, et al. The microenvironment of AIDS-related diffuse large B-cell lymphoma provides insight into the pathophysiology and indicates possible therapeutic strategies. Blood. 2013;122:424–33.

30. Hernandez-Ilizaliturri FJ, Deeb G, Zinzani PL, Pileri SA, Malik F, Macon WR, et al. Higher response to lenalidomide in relapsed/refractory diffuse large B-cell lymphoma in nongerminal center B-cell-like than in germinal center B-cell-like phenotype. Cancer. 2011;117:5058–66.

31. Dunleavy K, Pittaluga S, Czuczman MS, Dave SS, Wright G, Grant N, et al. Differential efficacy of bortezomib plus chemotherapy within molecular subtypes of diffuse large B-cell lymphoma. Blood. 2009;113:6069–76.

32. Leonard JP, Kolibaba KS, Reeves JA, Tulpule A, Flinn IW, Kolevska T, et al. Randomized phase II study of R-CHOP with or without bortezomib in previously untreated patients with non-germinal center B-Cell-like diffuse large B-cell lymphoma. J Clin Oncol. 2017;35:3538–46.

33. Wilson WH, Young RM, Schmitz R, Yang Y, Pittaluga S, Wright G, et al. Targeting B cell receptor signaling with ibrutinib in diffuse large B cell lymphoma. Nat Med. 2015;21:922–6.

34. Meyer PN, Fu K, Greiner TC, Smith LM, Delabie J, Gascoyne RD, et al. Immunohistochemical methods for predicting cell of origin and survival in patients with diffuse large B-cell lymphoma treated with rituximab. J Clin Oncol. 2011;29:200–7.

35. Choi WW, Weisenburger DD, Greiner TC, Piris MA, Banham AH, Delabie J, et al. A new immunostain algorithm classifies diffuse large B-cell lymphoma into molecular subtypes with high accuracy. Clin Cancer Res. 2009;15:5494–502.

36. Muris JJ, Meijer CJ, Vos W, van Krieken JH, Jiwa NM, Ossenkoppele GJ, et al. Immunohistochemical profiling based on Bcl-2, CD10 and MUM1 expression improves risk stratification in patients with primary nodal diffuse large B cell lymphoma. J Pathol. 2006;208:714–23.

37. Chadburn A, Chiu A, Lee JY, Chen X, Hyjek E, Banham AH, et al. Immunophenotypic analysis of AIDS-related diffuse B-cell lymphoma and clinical implications in patients from AIDS Malignancies Consortium clinical trials 010 and 034. J Clin Oncol. 2009;27:5039–48.

38. Montes-Moreno S, Odqvist L, Diaz-Perez JA, Lopez AB, de Villambrosia SG, Mazorra F, et al. EBV-positive diffuse large B-cell lymphoma of the elderly is an aggressive post-germinal center B-cell neoplasm characterized by prominent nuclear factor-kB activation. Mod Pathol. 2012;25:968–82.

39. Li S, Lin P, Fayad LE, Lennon PA, Miranda RN, Yin CC, et al. B-cell lymphomas with MYC/8q24 rearrangements and IGH@BCL2/t(14;18)(q32;q21): an aggressive disease with heterogeneous histology, germinal center B-cell immunophenotype and poor outcome. Mod Pathol. 2012;25:145–56.

40. Johnson NA, Savage KJ, Ludkovski O, Ben-Neriah S, Woods R, Steidl C, et al. Lymphomas with concurrent BCL2 and MYC translocations: the critical factors associated with survival. Blood. 2009;114:2273–9.

41. Johnson NA, Slack GW, Savage KJ, Connors JM, Ben-Neriah S, Rogic S, et al. Concurrent expression of MYC and BCL2 in diffuse large B-cell lymphoma treated with rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone. J Clin Oncol. 2012;30:3452–9.

42. Yoon DH, Choi DR, Ahn HJ, Kim S, Lee DH, Kim SW, et al. Ki-67 expression as a prognostic factor in diffuse large B-cell

lymphoma patients treated with rituximab plus CHOP. Eur J Haematol. 2010;85:149–57.

43. Sanchez E, Chacon I, Plaza MM, Munoz E, Cruz MA, Martinez B, et al. Clinical outcome in diffuse large B-cell lymphoma is dependent on the relationship between different cell-cycle regulator proteins. J Clin Oncol. 1998;16:1931–9.

44. Broyde A, Boycov O, Strenov Y, Okon E, Shpilberg O, Bairey O. Role and prognostic significance of the Ki-67 index in non-Hodgkin's lymphoma. Am J Hematol. 2009;84:338–43.

45. Wilson WH, Teruya-Feldstein J, Fest T, Harris C, Steinberg SM, Jaffe ES, et al. Relationship of p53, bcl-2, and tumor proliferation to clinical drug resistance in non-Hodgkin's lymphomas. Blood. 1997;89:601–9.

46. Jerkeman M, Anderson H, Dictor M, Kvaloy S, Akerman M, Cavallin-Stahl E, et al. Assessment of biological prognostic factors provides clinically relevant information in patients with diffuse large B-cell lymphoma–a Nordic Lymphoma Group study. Ann Hematol. 2004;83:414–9.

47. Bartlett NL, Wilson WH, Jung SH, Hsi ED, Maurer MJ, Pederson LD, et al. Dose-adjusted EPOCH-R compared with R-CHOP as frontline therapy for diffuse large B-cell lymphoma: clinical outcomes of the phase III intergroup trial alliance/CALGB 50303. J Clin Oncol. 2019;37:1790–2799.

48. Dunleavy K, Pittaluga S, Shovlin M, Steinberg SM, Cole D, Grant C, et al. Low-intensity therapy in adults with Burkitt's lymphoma. N Engl J Med. 2013;369:1915–25.

49. Wilson WH, Grossbard ML, Pittaluga S, Cole D, Pearson D, Drbohlav N, et al. Dose-adjusted EPOCH chemotherapy for untreated large B-cell lymphomas: a pharmacodynamic approach with high efficacy. Blood. 2002;99:2685–93.