Clinical and Biological Subtypes of B-cell Lymphoma Revealed by Microenvironmental Signatures

Nikita Kotlov¹, Alexander Bagaev¹, Maria V. Revuelta², Jude M. Phillip², Maria Teresa Cacciapuoti³, Zoya Antysheva¹, Viktor Svekolkin¹, Ekaterina Tikhonova¹, Natalia Miheecheva¹, Natalia Kuzkina¹, Grigorii Nos¹, Fabrizio Tabbo³, Felix Frenkel¹, Paola Ghione², Maria Tsiper¹, Nava Almog¹, Nathan Fowler¹, Ari M. Melnick², John P. Leonard², Giorgio Inghirami³, and Leandro Cerchietti²
**ABSTRACT**

Diffuse large B-cell lymphoma (DLBCL) is a biologically and clinically heterogeneous disease. Transcriptomic and genetic characterization of DLBCL has increased the understanding of its intrinsic pathogenesis and provided potential therapeutic targets. However, the role of the microenvironment in DLBCL biology remains less understood. Here, we performed a transcriptomic analysis of the microenvironment of 4,655 DLBCLs from multiple independent cohorts and described four major lymphoma microenvironment categories that associate with distinct biological aberrations and clinical behavior. We also found evidence of genetic and epigenetic mechanisms deployed by cancer cells to evade microenvironmental constraints of lymphoma growth, supporting the rationale for implementing DNA hypomethylating agents in selected patients with DLBCL. In addition, our work uncovered new therapeutic vulnerabilities in the biochemical composition of the extracellular matrix that were exploited to decrease DLBCL proliferation in preclinical models. This novel classification provides a road map for the biological characterization and therapeutic exploitation of the DLBCL microenvironment.

**SIGNIFICANCE:** In a translational relevant transcriptomic-based classification, we characterized the microenvironment as a critical component of the B-cell lymphoma biology and associated it with the DLBCL clinical behavior establishing a novel opportunity for targeting therapies.

---

**INTRODUCTION**

Diffuse large B-cell lymphoma (DLBCL), the most common lymphoid malignancy in adults worldwide, is curable with anti-CD20-based combination immunochemotherapy in just over 60% of patients. Over the past few years, several novel targeted therapies have been developed, but a substantial fraction of patients with relapsed disease continue to die from lymphoma or its complications. Further improvement of treatment outcome relies on elucidating the biology that underlies the clinical behavior of this heterogeneous disease. Gene-expression analysis based on cell-of-origin (COO) classification originally identified two major subgroups of DLBCL: activated B-cell (ABC) and germinal center B-cell (GCB). These subgroups have distinct clinical behavior and molecular features, reflecting differential pathogenesis (1). More recently, studies of genomic alterations in DLBCL cells including mutations, somatic copy-number alterations, and structural variants identified distinct genetic subtypes within the COO subgroups (2–4). Patients with ABC-DLBCL with co-occurring MYD88 and CD79B mutations or harboring NOTCH1 mutations have poorer prognosis than other ABC-DLBCLs (3). Similarly, lymphomas carrying EZH2 mutations and BCL2 translocations are associated with worse outcome within the GCB-DLBCLs (3). Meanwhile, lymphomas harboring MYC and BCL2 and/or BCL6 rearrangements represent a GCB subgroup with distinctly aggressive biology, also referred to as high-grade B-cell lymphoma double-hit or triple-hit (HGBL-DH/TH; refs. 5, 6).

The biology and clinical behavior of DLBCL result not only from the molecular alterations of DLBCL cells but also from their interaction with the microenvironment. Data from patients with lymphoma and animal models indicate that in the lymphoma niche, external stimuli provided by microenvironmental cells and the extracellular matrix (ECM) contribute to disease development, progression, and response to treatment (7–11). Because the interaction of lymphoma cells and the microenvironment is bidirectional (12, 13), it is expected that the DLBCL microenvironment will exhibit molecular and functional complexities that are yet to be defined. Despite the increasingly recognized role of the microenvironment in DLBCL biology (10), the majority of the molecular and therapeutic studies of this disease have been focused on the characterization of the DLBCL cell as an isolated entity (1).

Here, we characterized the DLBCL microenvironment based on the transcriptional footprint of microenvironment cells and processes in a large cohort of patients. We found four distinct microenvironment compositions reflecting unique biological properties and clinical behavior. These newly described categories associate with distinct clinical behavior of genetically similar DLBCLs and provide an array of novel potential targets for innovative therapeutic interventions.

---

**RESULTS**

**The Lymphoma Microenvironment Is Characterized by Four Distinct Categories**

We aimed to define the diversity in the composition and functionality of DLBCL microenvironments based on the statistical power conferred by gene-expression profiles of...
thousands of patients. As a first step toward this end, we generated 25 Functional Gene Expression Signatures (GES) reflecting either distinct cellular subtypes relevant to tumor microenvironment (e.g., cancer-associated fibroblasts and tumor-infiltrating lymphocytes), noncellular components of the tumor microenvironment (e.g., immune-suppressive cytokines and ECM), biological processes (e.g., secretion and proliferation), or canonical signaling pathway activation (e.g., PI3K and NFκB; Fig. 1A; Supplementary Fig. S1A–S1C; Supplementary Table S1). For each GES we calculated an activity score that is directly associated with the magnitude by which a particular effect is present among populations of cells. For example, PI3K and NFκB GES are sensitive enough to detect changes in pathway activity upon their specific inhibition in DLBCL cell lines and mouse models (Supplementary Fig. S1D and S1E). Activity scores of the GES revealed the expected enrichment for the cell types from which they derive (Fig. 1B; Supplementary Fig. S1B and S1C), as well as potentially novel functional interdependencies. For example, although the ECM GES was almost exclusively correlated with fibroblasts, the “ECM remodeling” GES was also correlated with activated macrophages (Fig. 1B), suggesting a potential role for these cells in the ECM microarchitecture. Similarly, they informed on cell states; for example, the biological process “cell proliferation” GES was almost absent in nonproliferating circulating B cells and progressively higher in B-cell lymphomas ranging from low-proliferating indolent follicular lymphomas (FL) to DLBCLs and the highly proliferative Burkitt lymphomas (BL; Fig. 1B).

To validate whether the GES was still informative when applied to gene-expression profiles derived from tissues with mixed cell types, we analyzed the activity score of the “cell proliferation” GES in a cohort of 4,984 B-cell lymphoma gene-expression profiles including FL, DLBCL, and BL and found that, similar to individual cells, the activity score of these GES directly correlated with the proliferation rate reported for these diseases (Fig. 1C). Moreover, as an additional independent validation, the “cell proliferation” GES also correlated with the proliferation rate determined by “gold-standard” Ki-67 IHC of 532 B-cell lymphoma samples (Fig. 1D). Last, we proved that the proportion of “fibroblasts” and “activated macrophages” estimated by GES correlated with the estimation of these cells obtained from anti-SMA and anti-CD68 immunostaining quantification from matched DLBCL tissues (Supplementary Fig. S1F).

We then used the 25 GES to virtually reconstruct the lymphoma microenvironment (LME) of 4,580 DLBCL cases collected from publicly available gene-expression databases and included an additional set of 75 DLBCLs profiled by RNA sequencing (RNA-seq) locally (Supplementary Fig. S2A). We analyzed the correlation between GES among samples by applying an unsupervised community detection algorithm (Fig. 1A) and obtained four major clusters of LMEs (Fig. 1E). These four LMEs reflect distinct GES associations or “communities,” and were present in 15% (LME-1), 33% (LME-2), 25% (LME-3), and 27% (LME-4) of the samples, with similar distribution across all tested data sets (Fig. 1E; Supplementary Fig. S2B–S2F). According to the communities of GES enriched within each microenvironment, the four distinct LMEs were termed, respectively, as “germinal center-like” due to the presence of GES from cell types commonly found in germinal centers (GC; ref. 14); “mesenchymal” (MS) for the abundance of GES from stromal cells and ECM pathways; “inflammatory” (IN) for the presence of GES associated with inflammatory cells and pathways; and finally, a “depleted” (DP) form that, contrasting with the other LMEs, was characterized by an overall lower presence of microenvironment-derived GES (Fig. 1E).

**LME Categories Represent a Novel Gene Expression-Based DLBCL Classification**

We developed a platform-independent algorithm to classify individual lymphomas according to their ABC-DLBCL versus GCB-DLBCL COO subtype (Supplementary Fig. S3A–S3C; Supplementary Table S2). Among the full cohort (n = 4,655), 59% were GCB-DLBCLs and 41% were ABC-DLBCLs (Supplementary Fig. S3C). These DLBCL subtypes varied in their LME associations. For example, whereas the IN-LME subtype contained a higher proportion of ABC-DLBCLs, the GC-like and MS LMEs were more enriched in GCB-DLBCLs (Fig. 2A, P < 0.001, χ² test, all comparisons). For the subsequent analysis requiring different molecular or clinical annotations, we used the largest possible subset of the data that had the required information (Supplementary Fig. S2A). We investigated the association of LME categories with the Consensus Clustering Classification (CCC; ref. 15) gene-expression subtypes B-cell receptor (BCR), oxidative phosphorylation (OxPhos), and host response (HR). As might be expected, the DP-LME was not associated with HR cases, whereas BCR and OxPhos did show specific LME patterns (Supplementary Fig. S3D). In addition, the MS-LME was enriched in the previously described “stromal-1” and “stromal-2” transcriptional signatures (ref. 10; Supplementary Fig. S3E).

Notably, HGBL-DHs were enriched in DP-LME and secondarily in GC-LME according to the analysis of 1,397 cases with these data available (Supplementary Fig. S3D).
Development of 25 FGES DLBCL sample correlation

Cell types

LEC - helper
B cells
Plasmacytoid dendritic cells
Dendritic cells (DC)
Macrophages
Activated macrophages
Follicular dendritic cells (FDC)
Fibroblasts
Vascular endothelium (VEC)

CD4 T cells
Tregs
Follicular T helper (TFH)
CD8 T cells
Natural killer (NK)
Burkitt lymphoma
DLBCL
Follicular lymphoma

Cell proliferation

Ki-67-positive nuclei (%)

Cell type

FL
DLBCL
BL

FGES score

Spearman’s ρ = 0.58

E

LME

GC-like
Mesenchymal
Inflammatory
Depleted

LEC
VEC
CAF
FRC

ECM remodeling
Granulocyte traffic
IS cytokines
FDC
Macrophages
Activated M1
T-cell traffic
MHC II
MHC I
TFH
Tregs
TIL
IS checkpoints
NK
B-cell traffic
B cells
Cell proliferation

NFκB
PI3K
p53

4,655 DLBCLs

Published OnlineFirst February 4, 2021; DOI: 10.1158/2159-8290.CD-20-0839
A GC-like Mesenchymal Inflammatory Depleted

B MCD N1 AS3 BN2 ST2 MYC MYC' Com Other

C

D Mutation Loss Gain

GCB-DLBCL TNFRSF1 4EBF1 CD83 STAT6

GCB-DLBCL UNC5D TNFAIP3 CARD11 UBE2A ATM BTK CD44 POU2AF1 BIRC3

ARIDSB EP300 TMEM30A ARID1B EZH2 KMT2C

MEF2B XPO1 CD36

MPEG1 PIM2 NOTCH1

PIM1 MYD88 GRHPR CDKN2A MYC PRDM1 TGFAIP3 BCL2 MALT1 SPIB BAX DSEL

E LME

CDKN2A TP53 CCND3

F

P = 0.003 (n = 306)

G LME

B2M CIITA EZH2 Antigen pr.

COO

GNA13 GNAI2 P2RY8 Gα signaling

H Antigen presentation

Gα signaling

P = 0.001 (n = 250)
P < 0.001 (n = 110)
Concurrent rearrangements of MYC and BCL2 are not the only molecular determinants of aggressiveness in DLBCL. Two gene-expression signatures, the molecular high-grade (MHG) and the DHL signature (DHTsig; ref. 16), identified additional patients who behave as HGBL-DH without necessarily having these genomic alterations. We developed and applied the DHTsig classifier (Supplementary Fig. S4A–S4G; Supplementary Table S3) to our collected data set in relation to LME categories and found a significant enrichment of MHG and DHTsig patients in the DP-LME category (Supplementary Fig. S3D). Moreover, the MHG and DHTsig were highly overlapping (Supplementary Fig. S4D), likely reflecting a similar underlying biology. Last, in a mouse model engineered to express Myc and Bcl2 in GC B cells, the resultant B-cell lymphomas harbor an LME that is similar to DP-LME DLBCLs (P < 0.001; Supplementary Fig. S4H).

DLBCLs were recently classified into genetically defined entities with distinct clinical characteristics (2–4), some of them also associated with microenvironmental features (4). One of these efforts (4) grouped DLBCL into the subtypes MCD (co-occurrence of MYD88 and CD79B mutations), N1 (NOTCH1 mutations), EZB (EZH2 mutations and BCL2 translocations), BN2 (BCL6 fusions and NOTCH2 mutations), ST2 (SCL1 and TET2 mutations), and AS (aneuploid with TP53 inactivation). The genetic annotation was available from 737 cases from two cohorts (NCICCR and DLC2), and revealed uneven distribution of the four LME categories (Fig. 2B). For example, the poor-prognostic MCD subtype represented 4% of GC-like, 8% of MS, 17.6% of IN, and 21% of the DP-LME categories (P < 0.001, χ² test; Fig. 2B). The MCD distribution in LMEs was not in full agreement with the proportion of ABC cases. Although a majority of BN2, ST2, and EZB cases had an immune-deserted LME (either MS or DP), in EZB the specific immune-deserted LME category was associated with their MYC status: EZB-MYC+ had a higher proportion of DP-LMEs, whereas EZB-MYC− had more MS LMEs (Fig. 2B). Taken together, these data indicate that DLBCL LME categories provide additional orthogonal information that is not captured by previously reported lymphoma classifications.

**LME Categories Associate with Specific Genomic Alterations in Lymphoma Cells**

The LME is dependent on dynamic interactions between tumor and normal cell types. To better understand the effect of lymphoma cells in driving specific LME patterns, we established 18 patient-derived tumor xenografts (PDTX) by implanting DLBCL tissues from patients into immunosuppressed mice. We then performed RNA-seq in tumors from early passages of each PDTX and used our cell signature deconvolution approach to determine whether murine LME cells were repopulating the PDTX in proportions similar to those from the primary human tissue. Remarkably, the majority of PDTXs maintained similar murine LME categories as were represented in the original primary human lymphoma (Fig. 2C; Supplementary Fig. S5A and S5B). Because recipient mice are immunosuppressed, the immune components of the LME were largely maintained from the human primary tissues in early passages. In contrast, the LME stromal cells, like cancer-associated fibroblasts (CAF) and tumor-associated macrophages (TAM), were recruited from the mouse tissues (Fig. 2C), as has been reported in breast cancer PDTX specimens (17). This suggests an underlying, long-lasting mutual signaling wiring between malignant and LME cells that is maintained in early-passage PDTXs.

To explore mechanisms by which lymphomas might induce specific LMEs, we compared their spectra of mutations and copy-number alterations (CNA). Mutations were obtained from whole-exome sequencing (WES) and/or whole-genome sequencing (WGS) or targeted sequencing from 747 patients, whereas whole-genome CNA were available for 750 patients (Supplementary Fig. S2A). The analysis of the 86 most frequent mutations and the 65 most frequent CNA (Supplementary Fig. S6A and S6B; Supplementary Tables S4 and S5) showed uneven distribution among LME categories with 24 mutations and 22 CNA significantly enriched in a particular LME (Fig. 2D). Importantly, the distribution of LME categories for 9 of these unequally distributed 24 mutations and 6 of the 22 CNA was not entirely explained by their COO designation (Fig. 2D), and so are potentially linked to direct effects of these mutations. Remarkably, some mutations and CNA targeting the same gene converged in a particular LME as was the case for CD44 with the MS-LME and for TMEM30A with the DP-LME (Fig. 2D).

Because CNA and/or mutations targeting functionally related genes may result in similar phenotypes, we analyzed such complementary genomic alterations in relation to LME categories. We considered all genomic aberrations annotated to a pathway independent of their individual statistical significance. We found that genomic alterations resulting in decreased p53 activity and perturbation of cell-cycle regulation were more common in DLBCL with DP-LME (Fig. 2E and F). This was validated in an independent data set of 67
DLBCLs (18), in which 90% of DP-LMEs showed complex CNAs affecting this pathway compared with >60% for the other LMEs (P < 0.04; Supplementary Fig. S3D). Accordingly, DP-LMEs also had a higher proportion of A53 lymphomas (Fig. 2B) and strong enrichment for the “cell proliferation” FGDE5 as later characterized. Recent data (19) showed expression of MHC I and MHC II antigen presentation genes and corresponding sparsity of immune infiltrates in murine lymphoma models expressing EzH2knockout. In agreement with these data, EZH2 genomic alterations as well as mutations in B2M or CIITA that affect MHC I or MHC II, respectively, were enriched in MS-LME (Fig. 2G and H). In addition, MS-LME cases were enriched for mutations in GNA13, GNA2, and P2RY8 (Fig. 2G and H). Mutations in GNA13/I2 and/or P2RY8 disrupt the Gα signaling pathway that is necessary to transduce homing signaling to lymphoma cells from surrounding immune cells (20), which are depleted from this LME, probably suggesting the acquisition of LME-independent growth mechanisms by lymphoma cells.

**LME Categories Associate with Distinct Disease Courses**

LME-GC-like and LME-DP had, respectively, the highest proportion of low (0–1) and high (4–5) international prognostic index (IPI) scores (Supplementary Fig. S7A). To determine the association of the LME categories with response to R-CHOP chemoinmunotherapy, we compiled and performed RNA-seq of a cohort of 105 newly diagnosed patients with DLBCL who were balanced between responders (55%) and nonresponders (45%) including refractory (i.e., never achieved complete response) and relapsed (i.e., free from disease < 2 years). The number of responders was highest in GC-like– and lowest in DP-LME patients (Fig. 3A; Z tests P = 0.030 for GC-like vs. DP and P = 0.022 for GC-like + MS vs. IN + DP). In contrast, DP-LME had the highest proportion of nonresponders and specifically of relapsed cases (Fig. 3A; 40.5% vs. 16.2% for DP-LME vs. all other LMEs, respectively; Z test P = 0.011). These associations were validated in an independent cohort of 1,349 unselected patients with DLBCL who received chemoinmunotherapy and were evaluated according to the REICIL criteria. The proportion of nonresponders (SD + PD) was 4.5%, 7.3%, 11.1%, and 12.7% for GC-like, MS, IN, and DP-LMEs, respectively (Supplementary Fig. S7B; Z test P = 5 × 10−4 for GC-like + MS vs. IN + DP-LMEs).

We further analyzed the association of LME categories with overall survival (OS) and progression-free survival (PFS) in patient cohorts of 2,646 and 2,189 DLBCLs, respectively. All patients received rituximab-based chemoinmunotherapy resulting in typical OS and PFS Kaplan–Meier curves (Supplementary Fig. S7C and S7D). Analysis of OS by LME indicated significant differences in prognosis from better to poor as follows: GC-like, MS (P = 0.03, vs. GC-like), IN (P = 0.0008, vs. MS), and DP (P = 0.007, vs. IN) LMEs (Fig. 3B), whereas GC-like and MS have similarly favorable PFS curves (P = 0.9) followed by IN (P < 0.001) and DP (P = 0.03, vs. IN), which presented with the poorest PFS (Fig. 3C). Bonferroni correction for multiple testing indicates that GC-like and MS are in the same strata for OS and PFS, as well as IN and DP for PFS. LME categories associated with different proportions of cases achieving OS and PFS at 24 months (OS24 and PFS24, respectively; Fig. 3D; Supplementary Fig. S7E).

When segregated by COO, the DP-LME retained the poorest OS and PFS in both COO subtypes. However, in ABC-DLBCL, the LME with the best PFS and OS was MS, and in GCB-DLBCL, the best LME was GC-like (Fig. 3E; Supplementary Fig. S7F), suggesting that the biological impact of the microenvironment may be different depending on lymphoma subtype (21). Accordingly, individual LME cell subtypes contributed differently to OS in ABC- and GCB-DLBCLs (Fig. 3F). Overall, the combination of GCB-DLBCL with GC-like LME and ABC-DLBCL with DP-LME resulted in the most favorable and poorest clinical outcome, respectively (Fig. 3E; Supplementary Fig. S7F). In multivariate Cox-proportional hazard analysis for PFS (n = 2,024) and OS (n = 2,052) controlled for COO and IPI, the LME subtypes remained informative, with GC-like and MS LMEs associated with better outcome (Fig. 3G; Supplementary Fig. S7G).

Finally, HGBL-DH, which carry significantly lower PFS and OS than DLBCL with neither MYC nor BCL2 rearrangements (Supplementary Fig. S7H), modified their outcome when LME was considered. HGBL-DH with the favorable prognosis LMEs GC-like and MS had significantly better PFS and OS than HGBL-DHs with the unfavorable prognosis LMEs IN and DP (Fig. 3H; Supplementary Fig. S7H). In multivariate Cox-proportional hazard analysis for OS (n = 486) controlled for COO, IPI, and MYC or BCL2 rearrangements, the LME subtypes remained informative, with GC-like and MS LMEs associated with better outcome (Supplementary Fig. S7I). The same was true among the DHITsig-positive DLBCLs, i.e., poor-prognostic GCB-DLBCLs with high-grade molecular features including concurrent MYC and BCL2 rearrangements (ref. 16; Fig. 3I and J; Supplementary Fig. S7J and S7K). Collectively, these data underline the impact of the LME category on outcome regardless of COO subgroup or genetic subtype, suggesting that these host factors make truly critical and independent contributions to lymphoma biology and clinical presentation.
Microenvironmental Signatures in Lymphoma

A. Response to RCHOP

<table>
<thead>
<tr>
<th>LME</th>
<th>GC</th>
<th>MS</th>
<th>IN</th>
<th>DP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohort</td>
<td>105</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sensitve | Relapsed | Refractory

B. Probability of OS

Comparison | P value | 4-way | 5 × 10^{-14}
GC-like vs. MS | 0.03
GC-like vs. IN | 5 × 10^{-6}
GC-like vs. DP | 3 × 10^{-11}
MS vs. IN | 3 × 10^{-8}
MS vs. DP | 0.007
IN vs. DP |

No. at risk
GC-like | 396 360 314 238 190 137 97
MS | 932 816 674 509 377 284 209
IN | 649 558 450 341 249 185 140
DP | 669 545 420 320 254 189 118

C. Probability of PFS

Comparison | P value | 4-way | 6 × 10^{-13}
GC-like vs. MS | 0.0002
GC-like vs. IN | 5 × 10^{-4}
MS vs. IN | 6 × 10^{-4}
MS vs. DP | 2 × 10^{-3}
IN vs. DP | 0.03

No. at risk
GC-like | 332 290 239 177 138 95 68
MS | 780 627 525 386 288 219 162
IN | 536 403 329 255 183 137 100
DP | 561 401 312 241 182 129 76

D. PFS 24% (%)

| LME | GC | MS | IN | DP |

E. Probability of OS

Comparison | P value | 4-way | 6 × 10^{-5}
GC-like vs. MS | 0.02
MS vs. IN | 0.004
MS vs. DP | 3 × 10^{-4}
IN vs. DP | 0.04

No. at risk
GC-like | 105 88 66 51 35 26 21
MS | 224 178 148 108 87 69 50
IN | 326 234 186 142 98 71 57
DP | 258 178 126 93 69 50 31

F. ABC-DLBCL

- Macrophages M2
- Other LME cells
- Tregs
- CD8+ PD-1+
- Plasma B cells
- Macrophages M1
- Monocytes
- Neutrophils
- B cells
- VEC
- NK cells
- CD4+ T cells
- Fibroblasts

G. GC-like-LME

- GC
- GCB
- GC-like-MS-LME
- MS-LME
- IPI intermediate
- IPI low

H. GC-like + MS

I. GC-like + MS

J. DHIITsig-pos

- IN-LME
- MS-LME
- GC-like-LME
- GCB
- IPI intermediate
- IPI low

Published OnlineFirst February 4, 2021; DOI: 10.1158/2159-8290.CD-20-0839

CANCER DISCOVERY | 0F8

JUNE 2021

American Association for Cancer Research

Downloaded from cancerdiscovery.aacrjournals.org on May 16, 2021. © 2021 American Association for Cancer Research.
LME Categories Are Composed of Distinct Cellular Communities

To further understand the nature of each LME category, we expanded the community analysis of F^{GES} signatures by estimating cell percentages in the LME using a RNA-seq deconvolution algorithm (22). This allowed us to more precisely define cell subpopulations and estimate their abundance. We validated this approach to cellular composition analysis by comparing and contrasting the results with mass cytometry performed on reactive tonsillar tissues (Supplementary Fig. S8A). In addition to cellular composition, we also investigated the enrichment of the LME-dependent signaling and transcriptional pathways hypoxia-inducible factor (HIF), TGFβ/SMAD, JAK/STAT, and TNF.

The GC-like LME resembled the cellular constitution of the GC microenvironment (14) since it contained a relatively higher proportion of follicular dendritic cells, lymphatic endothelial cells, total T cells, and several CD4+ T-cell subpopulations including regulatory T cells (Treg) and Th cells (Fig. 4A and B). The ratio of malignant to normal B cells was lower in this subtype. The F^{GES} for cell proliferation activity that primarily captures the activity of malignant cells was higher in the GC subtype (Fig. 4A and B). The ratio of malignant to normal B cells was lower compared with other LME categories (Supplementary Fig. S8B). These lymphomas may thus represent transformed variants of indolent, GC-derived FLs. Further supporting this, among the unfavorable LME categories, the IN-LME was highly enriched among primary mediastinal B-cell lymphomas versus other LME subtypes (n = 287 total cohort, P = 0.003; Supplementary Fig. S8C).

Among the unfavorable LME categories, the IN-LME was enriched in neutrophils F^{GES} associated with poor outcome (26), macrophages (likely M1; Supplementary Fig. S8F), CD8+ T cells (over total T cells or over CD4+ T cells; Supplementary Fig. S8F), and the subset of CD8+ T cells with high PD-1 expression (Fig. 4F and G). The proportion of natural killer (NK) cells was not significantly different among LMEs; however, the NK activity by F^{GES} was significantly higher in IN-LME (Supplementary Fig. S8F). This finding plus an enrichment of cytotoxicity activity as determined by the cytolytic score (ref. 27; Fig. 4H) and presence of T-cell activation markers suggests a certain degree of antilymphoma immunity. Indeed, DLBCL with IN-LME presented with the lowest relative number of malignant cells and lower mutational load (Fig. 4I), potentially representing lymphoma cells that may be able to escape tumor immunity (28, 29). The IN-LME had features that may contribute to decrease lymphoma immunity such as enrichment of immune-suppressive and prolymphoma cytokines (including the neutrophils attractant CXCL8, ref. 30; as well as IL10, refs. 31–34; Fig. 4J), and expression of the immune-checkpoint molecule PD-L1 (33, 34) and the tryptophan catabolic enzyme indoleamine 2,3-dioxygenase 1 (IDO1; refs. 35, 36; Fig. 4K). As might be expected from this inflammatory milieu given the presence of macrophages and neutrophils F^{GES} in particular (37), there was a significantly higher activity of the signaling pathways NFKB (ref. 38; only partially associated with the COO subgroup; Supplementary Fig. S8G), JAK/STAT (34, 38), and TNF (39) compared with the other LME categories (Fig. 4L).

DNA Methylation Patterning of Lymphoma Cells Contributes to a Depleted Microenvironment

DLBCLs with a DP-LME were characterized by a minimal presence of F^{GES} derived from microenvironmental cells as well as by higher proportions of proliferating tumor cells (Fig. 5A) and clonal tumor cells (Fig. 5B; Supplementary Fig. S9A). This category of LME was also common in patients with BL (Supplementary Fig. S8C). DP-LME is also characterized by activation of the PI3K signaling pathway (Supplementary Fig. S9B), suggesting potential therapeutic vulnerabilities. Higher tumor cell proliferation was insufficient to account for the scarcity of F^{GES} derived from microenvironmental cells. Indeed, a majority of the tumor cells stained negative for MHC I and II (by IHC) than any other LME (n = 177, P = 0.04; Fig. 5C), although there was no enrichment of MHC mutations in this group. We and others have previously reported (23, 24) that aberrant DNA hypermethylation of the SMAD1 promoter allows DLBCL cells to escape from the antiproliferative effects of LME-derived TGFβ ligands. Indeed, analyzing our cohort and the TCGA cohort of 100 DLBCLs, we found that the degree of global DNA hypermethylation and specific SMAD1 promoter methylation were significantly higher among DP-LME (Fig. 5D; Supplementary Fig. S9C). Accordingly, the SMAD1 expression and TGFβ pathway activity were also lower (Fig. 5E; Supplementary Fig. S9D and S9E). In the full data set, DLBCL with DP-LME had lower SMAD1 expression and TGFβ pathway activity than any other LME (Fig. 5F).

To determine the contribution of aberrant DNA hypermethylation to this microenvironmental scenario, we used the A20 syngeneic murine B-cell lymphoma model, which we characterized as having a similar LME constitution as patients with DP-LME DLBCL (Fig. 5G). Once tumors developed in the treatment cohort (n = 12), we exposed the mice to the DNA hypomethylating agent azacitidine or vehicle for four consecutive days and then performed RNA-seq and DNA methylation sequencing of the tumors (Fig. 5H). Azacitidine treatment resulted primarily in gene upregulation and hypomethylation (Supplementary Fig. S9F and S9G).

Pathway analysis of significantly hypomethylated gene promoters and upregulated genes revealed regaining of “cellular response to cytokines,” “signal transduction,” and immune response including “activation of T cells” (Fig. 5I; Supplementary Table S6). There was upregulation of Smad1 as well as increased TGFβ pathway activity (Fig. 5I). In addition, there was induction of MHC class I and II expression and an increased proportion of CD4+ T cells by transcriptome deconvolution (Fig. 5I) and increase in the distribution of tumor-infiltrating T cells by IHC analysis (4.7% (q25–75: 2.9–7.3) vs. 18% (q25–75: 14.65–18.9), vehicle vs. azacitidine, P < 0.001
Mann–Whitney test, Fig. 5J). Overall, these data indicate that cytosine methylation patterning of lymphoma cells may contribute to a depleted LME.

**LME Heterogeneity Decreases with Disease Progression**

To explore the notion that LME cellular composition changes during evolution of lymphomas, we first established a series of temporally ordered conditions representing early and late stages of lymphoma development. In each sample, we established the LME composition by transcriptome signature deconvolution and measured its complexity by Shannon entropy metric, which has two components: the number of unique cell subtypes (richness) and their equality of distribution (evenness). The Shannon metric directly cor-

**Figure 4.** Distinct cellular communities define DLBCL LME categories. A, Proportion of LME cells significantly enriched in the GC-like LME obtained by cell deconvolution algorithms or FGES. B, Schematic representation of selected features of GC-like-LME. FDC, follicular dendritic cells; LEC, lymphatic endothelial cells; TH, T helper cells. C, Proportion of LME cells significantly enriched in the MS-LME obtained by cell deconvolution algorithms or FGES. D, Schematic representation of selected features of MS-LME. VEC, vascular endothelial cells; FRC, fibroblastic reticular cells; CAF, cancer-associated fibroblasts. E, HIF1 (hypoxia) and TGFβ pathways activity. F, Proportion of LME cells significantly enriched in the IN-LME obtained by cell deconvolution algorithms. G, Schematic representation of selected features of IN-LME. TAM, tumor-associated macrophages. H, Cytolytic score in the IN-LME versus other LMEs. I, Lymphoma cellularity (proportion of malignant cells) and mutational load (number of mutations per cell) in DLBCL containing IN-LME versus other LMEs. J, Immune-suppressive/prolymphoma cytokines F[**] in the IN-LME versus other LMEs. K, Expression of PD-L1 and IDO1 in lymphomas with IN-LME versus other LMEs. L, NFκB, JAK/STAT, and TNFα pathway activity. ***, P<0.01; ***, P<0.001.
microenvironmental signatures in lymphoma

100

trajectory. using the space of the 25 f...

we applied a pseudotime algorithm to reconstruct lme...

relapsed dlbcl (40).

with recent data demonstrating a significant reduction in...

the decreasing intratumoral content of t cells, which agrees...

0.001, paired...

of progression by approximately 40% (95% ci, 20%–60%,...}

Shannon entropy, decreases with lymphoma development...

Bcl2, Ezh2 engineered mouse models expressing in gcb cells...

late murine lymphoma progression (1), indolent fl and transformed (t) fl (1).

and with the epigenetic age (41, 42) of the lymphoma (supplementary fig. 6C; supplementary fig. 7D).

LME pseudotime also demonstrated an association between...

LME Shannon entropy for individual dlbcl samples...

significantly correlated with the estimation of CAFs by f...

the ECM (matrisome; fig. 7A).

To characterize the tendency of the LME to immune desertion...}

tumor microenvironment, the ECM is primarily produced and remodeled by the balanced activity of TAMS and CAFs. We thus determined whether the ratio of Fges...

in DP-LME with an HR: 4.4 (fig. 7A). To determine which ECM components are more likely to be associated with the activity of CAFs versus TAMS, we conducted unbiased proteomic analysis of the ECM (matrisome; fig. 7B) of 18 primary dlbcls, using the matrisome of inflamed tonsils as a control for normalization. We had enough material in 14 of these dlbcls to conduct RNA-seq that we used to estimate tam/CAF, LME, and COO. We identified 131 proteins in the dlbcl matrisome distributed as glycoproteins, ECM regulators, collagens, ECM-affiliated, secreted factors, and proteoglycans (supplementary table S7), clustering in three matrotyps (fig. 7C).

Two matrotyps (termed Fm1 and Fm2) associated with a higher proportion of CAFs and had a higher proportion of collagens and proteoglycans (fig. 7C), whereas the third matrotype containing a higher proportion of ECM-affiliated and secreted proteins had a higher expression of TAMS (thus termed Mm; fig. 7C). Reflecting on the low expression of Fges, DP-LME DLBCL had relatively low ECM abundance (fig. 7C). To dissect cellular contributors to the DLBCL ECM, we correlated the matrisome genes with their median expression in 15 LME cell subtypes and DLBLC cells to obtain their probability of expression according to cell type (supplementary fig. S11A; representative examples). The DLBCL matrisome genes were then ranked from the highest probability of expression in fibroblasts versus macrophages to obtain their respective contributions to the ECM (fig. 7D; supplementary fig. S11B). Collagens and the proteoglycans decorin (DCN) and biglycan (BGN) were among the strongly associated with the CAFs (fig. 7D). Conversely, the ECM protein abundance of these collagen proteoglycans significantly correlated with the estimation of CAFs by Fges (R = 0.88, P = 0.001, n = 9 patients) and fibroblast proportion by signature deconvolution (R = 0.9, P < 0.001, n = 9 patients; supplementary fig. S11C). Accordingly, CAF-associated proteins were enriched in the portion of the DLBCL matrisome that correlated with a lower TAM/CAF ratio in 14 patients (fig. 7E). Last, and similar to CAF Fges, the expression of the CAF's ECM proteoglycans DCN and BGN was associated with favorable prognosis, particularly with DLBCL LME-DP patients (fig. 7F and G).

By a combination of short- and long-range effects (43), microenvironmental proteoglycans have the potential to influence tumor growth. To determine their effects in DLBCL, we first developed two PDX models from patients with DP-LME ABC-DLBC that conserved these characteristics in the mouse (fig. 7H; supplementary fig. S11D). Once the PDXs were fully established, we injected the mice...
intraperitoneally with vehicle, DCN, or BGN and followed the tumor growth curve. We found that both proteoglycans significantly delayed the progression of the tumor versus vehicle ($P = 0.009$ and $P = 0.039$ for BGN and DCN, respectively, in PDTX1 and $P = 0.0004$ and $P = 0.01$ for BGN and DCN, respectively, in PDTX2, Fig. 7I; Supplementary Fig. S11E). Proteoglycans did not show a statistically significant antiproliferative effect in isolated PDX-derived lymphoma cells. Although there was no evidence of direct incorporation of the administered proteoglycans into the PDTX ECM, BGN administration resulted, as previously described (44), in a 20% increase in DCN content ($P = 0.032$, vs. vehicle) in the PDTX2 ECM. Overall, this suggests that the antiproliferative effect of these proteoglycans results from both short- and long-range interactions.

**DISCUSSION**

We characterized the microenvironment of DLBCL by analyzing a large collection of gene-expression profiles and technical innovations, including development of microenvironment-derived F$_{GES}$, analysis of the ECM composition by proteomics, and establishment of PDTX models. Similar to cell deconvolution algorithms, our methodology reflects the cellular composition of the microenvironment but also provides valuable information on its functionality. This methodology
allowed us to describe four basic categories of DLBCL LME with distinct clinical and biological connotations. In addition, by overlaying mutations and epigenetic alterations in lymphoma cells with the temporal ordering of LME changes, we were able to infer potential mechanisms of cancer cell-LME coevolution. Lastly, we identified novel potential therapeutic targets present in the LME, including ECM proteins exemplified by the proteoglycans DCN and BGN.

Our study represents a comprehensive analysis of the LME in DLBCL that simultaneously integrates characteristics of microenvironmental and malignant cell populations into the prognosis of the disease. Pioneering transcriptomic studies that examined host factors and the stroma in DLBCL identified the microenvironment as a relevant component of disease biology (10, 15, 45). These initial studies relied on identifying differences in gene-expression profiles of tumor samples (10, 15, 45). Our study refined this analysis by extracting microenvironmental signatures from the confounding transcriptome background of the usually most abundant lymphoma cells, which enabled the identification and quantification of microenvironmental cells within the bulk tissue transcriptome (46). We further optimized this technique to include functional signatures that include the activity of certain cell types such as T cell and NK cell cytotoxicity as well as pathways not uniquely associated with a cell type such as ECM remodeling and cytokine secretion. The association of these functional signatures suggested the presence of functional cellular ecosystems within the LME and offered a comprehensive understanding of the role of the microenvironment in this disease. This approach provided more biologically relevant evidence than that obtained from the quantification of individual cell populations. Remarkably, the LME mirroring the GC microenvironment confers a better prognosis than the LME depleted of microenvironmental cells, suggesting that, analogous to the molecular checkpoints operating in centroblasts (47, 48), the microenvironment may provide primordial mechanisms to avoid lymphogenesis. In turn, lymphoma cells develop and/or select genetic and epigenetic traits that contribute to the evasion of immune microenvironmental constraints (19). Our data suggest that for initially immune-rich DLBCLs, disease progression tends to associate with an immune-deserted LME. More broadly, this notion suggests that the coevolution of lymphoma cells and LME cells (likely also involving non-immune LME cells) progresses toward a state with lower heterogeneity in these two compartments, a concept that remains to be experimentally validated as more LME cell subpopulations are described.

In addition to the restriction imposed by the immune cells to lymphoma development (49), our work suggests that the ECM also has a critical role. The ECM is a network of physically and biochemically distinct macromolecules, comprised of mostly fibrillar proteins, proteoglycans, and glycoproteins, that are central to the structural integrity of tissues and the behavioral regulation of malignant and microenvironmental cells. We described that LMEs presenting with a higher proportion of ECM components derived from CAFs (i.e., a CAF matreotype) rather than from TAMs have a favorable prognosis. Some ECM-derived proteins, particularly the small leucine rich proteoglycans (50) BGN and/or DCN, exert short- and long-range effects that ultimately affect lymphoma cell proliferation, as demonstrated in our PDTX experiments. Because the PDTXs were developed in immunocompromised mice, the mechanism appears independent of the establishment of an adaptive lymphoma immune response. Along these lines, previous work in DLBCL has established that CAFs were associated with a possible better outcome (51). In addition, through the secretion into the interstitial matrix of soluble factors such as TGFB ligands (52), fibroblasts may contribute to curtailing lymphoma cell progression. A mechanism that can be overcome in DLBCL cells by the epigenetic repression of TGFB transducers such as SMAD1 (23, 24). The antilymphoma role of CAFs, and of TGFB, in DLBCL is in contrast with their protumorigenic role described for the majority of solid tumors (52, 53), underlining the need to study the microenvironment in a tumor-specific manner.

In cohort analysis, HGBL-DH patients harboring translocations of MYC and BCL2 as well as DLBCL expressing high-molecular-grade transcriptional signatures have unfavorable prognosis after first-line therapy, which often justifies the implementation of intensive chemotherapy regimens (16, 54). However, at the individual level, long-term responders to standard chemoimmunotherapy do exist, suggesting the presence of a subgroup of less aggressive lymphomas. Although this phenomenon could be explained by the presence of additional genomic and epigenetic alterations (16), our data indicate that distinct biological behavior associates with unique LME categories. HGBL-DH and other MHG DLBCLs exhibit a significantly better outcome within GC-like and MS LMEs than within IN and DP-LMEs. In light of the chemoresistance associated with this lymphoma subtype, new approaches are considering the addition of targeted therapies such as BCL2 inhibitors (54). The results of our studies suggest that, in clinical trials, targeted agents should be considered not only in the context of particular genetic subtypes but also in consideration of LME categories.

In summary, we defined the LME into four major transcriptionally defined categories with distinct biological properties and clinical behavior, which complements gene-expression subgroups and genetic subtypes of DLBCL in guiding the development of rational therapeutic approaches for these patients.

**METHODS**

**Human Subjects**

All patients providing samples gave written informed consent. Molecular and clinical data acquisition and analysis and PDTX establishment were approved and carried out in accordance with Declaration of Helsinki and were approved by Institutional Review Boards of the New York Presbyterian Hospital, Weill Cornell Medicine (WCM), New York, NY, and Ospedale San Giovanni Battista delle Molinette, Turin, Italy.

**Clinical Characteristics**

All the patients with DLBCL in the WCM cohort were treated with R-CHOP. Response categories were defined as complete response when patients remained free of disease at five years, nonresponders...
Microenvironmental Signatures in Lymphoma

refractory when complete response was never achieved, and nonresponders relapsed when they remained free of disease for less than two years. All the biopsies for molecular analysis were taken before any treatment at the moment of diagnosis. From the publicly available cohorts, we considered for outcome analysis all gene-expression DLBCL data sets with annotated clinical and outcome data, including response by RECIL criteria (55), PFS, and/or OS of patients treated with rituximab-containing chemotherapy regimens. We excluded data sets in which the OS and/or PFS Kaplan–Meier curves were not representative of unselected patients with DLBCL treated with standard-of-care chemotherapy and/or data sets with fewer than 30 samples. Survival analysis was conducted using the Lifelines Python package.

**RNA-seq**

RNA was extracted from fresh-frozen DLBCL samples maintained in TRIzol (Invitrogen), and libraries (PE 50 or 100 bp) were prepared using the TruSeq RNA sample kit and validated using the Agilent Technologies 2100 Bioanalyzer. Library preparation, sequencing, and post-processing of the raw data were conducted at the WCM Epigenomics Core Facility on Illumina HiSeq2500. This data set (WCM cohort) was combined for normalization and aggregated into training and validation cohorts as described in the specific analysis.

**Compilation of a Transcriptomic LME Data Set**

Publicly available DLBCL gene-expression (GE) data sets and our own data set were compiled into a unique data set (LME data set) comprising a total of 4,655 samples (Supplementary Fig. S2A). Only cohorts with 30 or more lymphoma samples were included. Because our goal was to obtain information on the tumor microenvironment, GE data obtained from selected or isolated lymphoma cells were excluded from this analysis. For comparison, GE cohorts containing HGHL, FL, BL, and primary mediastinal B-cell lymphoma samples were also considered. Clinical annotation, pathologic information, and molecular data were curated and harmonized among all included data sets. Only patients with DLBCL and HGHL treated with rituximab-containing regimens were considered for response and outcome comparisons. Our WCM cohort included a balanced number of R-CHOP responsive and nonresponsive patients.

**RNA-seq Data Processing.** New and deposited RNA-seq reads were processed using a unified pipeline (56). Reads were aligned using Kallisto v0.42.4 to GENCODE v23 transcripts 69 with default parameters. Protein coding, IGH/K/L- and TCR-related transcripts were retained, whereas noncoding, histone- and mitochondrial-related transcripts were removed, resulting in 20,062 analyzed transcripts. Gene expression was quantified as transcripts per million and log_{2} transformed. Almost all RNA-seq cohorts were sequenced using poly-A enrichment and were combined into a single LME data set.

**GE Microarray Data Processing.** Raw and processed data were downloaded from the Gene Expression Omnibus (GEO). When possible, expression was reprocessed from raw files. All Affymetrix data sets with available CEL files were renormalized using the gcRMA package with default parameters. Illumina array probes were converted into genes using one probe with the highest mean values in the cohort per gene. Data sets with more than one platform were split by platforms. Samples with low Pearson correlation with other samples in the space of all genes (< 0.8 for Affymetrix platforms and < 0.7 for Illumina platforms) for each cohort were excluded as well as outliers according to the principal component analysis (PCA) projection. Microarray GE data sets were combined into the LME data set if they were obtained in the same institution, on the same platform, had no significant batch effects on PCA projection in the LME signature space, and had no significant differences in outcomes.

**Purified Cell Type Compendium Collection**

We collected 5,069 RNA-seq gene-expression data sets from purified cell populations including normal and lymphoma cells (DLBCL, FL, and BL) from public data sources (NCBI GEO, ref. 57; SRA, ref. 58; ENA, Array Express, Protein Atlas, ref. 59; BluePrint, and ImmPort) to create a cell gene-expression compendium. We included data sets using the following criteria: isolated from human tissue, poly-A or total RNA-seq performed with read length higher than 31 bp, having at least 4M of coding read counts, passed quality control by FASTQC, and no contamination with other cell types was detected (<2%).

**Development of LME GE Signatures**

We developed F\(_{\text{GES}}\) of tumor microenvironmental cells, cellular states, physiologic and pathologic processes, and signaling pathways using a combination of GE signatures and literature curation. For cell types were obtained from purified GE cell population signatures manually curated to include only those genes that are exclusively expressed in the defined cell type or specifically associated with particular biological processes using data from more than 500 publications. The association of GE profiles from 5,009 cell populations with the curated signatures was determined by tSNE projections and Mann–Whitney tests. Signature scores were calculated using in-house python implementation of the single sample gene set enrichment analysis (ssGSEA; ref. 60). Some cellular F\(_{\text{GES}}\) were also validated using publicly available single-cell RNA-seq data (61). The activity scores of biological pathways were calculated using PROGENy (Pathway RespOnsive GENes; ref. 62). As a result, 25 F\(_{\text{GES}}\) (including three signaling pathways) were developed (Supplementary Table S1).

**LME Clustering.** F\(_{\text{GES}}\) signatures were used to identify microenvironmental patterns among DLBCL GE samples by unsupervised dense clustering using the Louvain method for community detection (63). F\(_{\text{GES}}\) intensities were median-transformed within each cohort. Non-DLBCL samples were also transformed using DLBCL samples’ median and MAD values. Intersample similarity was calculated using...
Development of a COO Classifier

Training samples \((n = 1,968)\) were obtained from the data sets GSE117556, Leipzig Lymphoma (64), GSE13152, GSE10846, GSE87371, GSE11318, GSE62918, GSE23501, LLMPP, and GSE93984. For each data set, we selected samples having COO labeling followed by another round of random selection to obtain a balanced COO ABC:GCB ratio of 40:60 per data set. Validation samples \((n = 928)\) with COO labeling were selected from the data sets GSE34171 (GPL96 + GPL97), BCA (DCL1 + DCL2; ref. 19), GSE22898, GSE64555, WCM (GSE145043), GSE19246, and the NCICCR data sets. A final data set of samples originally considered COO unknown or unclassified \((n = 1,169)\) was aggregated from COO-unlabeled samples included in the previous data sets plus the GSE69015, GSE69049, ETA-TBM-346, GSE68895, GSE8202, GSE12195, ICGC_MALY_DE, and NCICCGCI data sets. To eliminate platform and data set batch effects without introducing scaling batch effects, we applied rank transformation to each sample independently as follows: Expression Geneset \(A = [x_1, x_2, x_3, \ldots, x_{N-1}, x_N]\) sorted ascending \([x_2, x_15, x_{N-1}, \ldots, x_1, x_N]\) \([1, 2, 3 \ldots N-1, N]\). Missing genes were omitted in the rank transformation, and genes with equal values were assigned with the average rank. To perform binary classification, we used a gradient boosting decision tree classifier in LightGBM. To perform feature selection, we estimated the feature importance to the model using boosting decision tree classifier in LightGBM. To perform binary classification, we used a gradient boosting decision tree classifier in LightGBM. To perform feature selection, we estimated the feature importance to the model using boosting decision tree classifier in LightGBM. To perform feature selection, we estimated the feature importance to the model using boosting decision tree classifier in LightGBM. To perform feature selection, we estimated the feature importance to the model using boosting decision tree classifier in LightGBM. To perform feature selection, we estimated the feature importance to the model using boosting decision tree classifier in LightGBM.

Deconvolution of Cell Percentages from Bulk DLBCL Transcriptomes

A recently described algorithm was used for cell deconvolution (22). Briefly, the sorted cell population compendium was used to develop a machine learning–based cell deconvolution algorithm to calculate the percentage of different cell types from bulk RNA-seq mixtures based on the minor difference between cell subpopulations. We used a two-stage hierarchical learning procedure for gradient boosting a LightGBM model that included training on artificial RNA-seq mixtures of different cell types including immune and stromal cell populations. Artificial RNA-seq mixtures were created by admixing different data sets of sorted cells together in various cell proportions, and the LightGBM model was trained to predict the admixed cell percentage. Then, the model was used to reconstruct proportions of cell subpopulations using the information from the proportion of the major cell populations and subpopulations. The algorithm estimates the RNA proportion of a cell type in bulk RNA-seq mix of a sample, which could be converted into cell percentage if the RNA concentration of a cell type was known (68). Total RNA abundance in isolated cells was quantified using Qubic.

Cytometry by Time of Flight and IHC Analysis

Inflamed tonsils were procured from the WCM–New York Presbyterian Hospital as pathology discards. Tonsils were mechanically dissociated and incubated in a digestion buffer (25 mg/ml collagenase A, 25 mg/ml dispase II, 250 mg/ml DNase II in a solution of 140 mmol/L NaCl, 5 mmol/L KCl, 2.5 mmol/L phosphate buffer saline pH 7.4, 10 mmol/L HEPES, 2 mmol/L CaCl\(_2\), and 1.3 mmol/L MgCl\(_2\)) until a single-cell suspension was obtained. Homogenous fresh cell suspensions were divided for RNA-seq and Cytometry by Time of Flight (CyTOF) analyses. Briefly, cells for CyTOF analysis were barcoded and pooled for incubation with FcR blocking reagent (Miltenyi Biotec), stained with 300 μL of the antibody panel per 10\(^6\) cells, and resuspended in nucleic acid IR-intercalator (Fluidigm). Samples were analyzed at the Mass Cytometry Core of WCM. Data were processed and analyzed as previously described (69). Clustering analysis was performed using the Python implementation of Phenograph run on all samples simultaneously and then manually curated into cell populations. CAFs and TAMs were identified by IHC using anti-SMA and anti-CD68 antibodies, respectively, and quantified using QuPath (70).

Genetic Alterations Analysis

WES and WGS next-generation sequencing quality control analysis was performed using FastQC v0.11.5, FastQ Screen v0.11.1, and MultiQC v1.6. Sample correspondence was checked using HLA comparison and the CompAsign software (71). For WES, low-quality reads were filtered using FilterByTile/BBMap v37.90 and aligned to human reference genome GRCh38 (GRCh38.d1.v1.d1 assembly) using
**Microenvironmental Signatures in Lymphoma**

BWA v0.7.17. Duplicate reads were removed using MarkDuplicates v2.6.0. indels were realigned by IndelRealigner and recalibrated by BaseRecalibrator (both of GATK v3.8.1).

**Variant Calling.** Germline and somatic single-nucleotide variations, small insertions, and deletions were detected using Strelka v2.9 and annotated using Variant Effect Predictor v92.1. CNA were estimated by customized version of Sequenza v2.1.2. Translocations from WGS samples were called using DELLY (v0.8.1; ref: 72) and annotated with AnnotSV v2.2; ref: 73).

**Public Genomic Data.** Whole-genome mutations were available for NCICCR, WCM, and GSE98888/hs000573 cohorts. Targeted mutations were available for the GSE117556 (70 genes), GSE22898 (322 genes), BCA/DLC2 (144 genes), and GSE87371 (34 genes) cohorts. Whole-genome CNA calls were available for NCICCR, WCM, GSE34171_1, and GSE98888 cohorts.

**Association of LME with Specific Genetic Alterations.** NCICCR, BCA and WCM cohorts were used to analyze disbalances in the proportion of mutations in relation to LME categories. NCICCR, WCM, and GSE87371 cohorts were used to analyze disbalances in the proportion of CNAs in relation to LME categories. Ploidy was accessed using weighted mean and reported as decimal number. The gene-level CNVs were obtained intersecting segments with hg19 gene coordinates using BEDTools (74). Gene CNVs were analyzed relative to ploidy levels. Additionally, we curated mutations and CNAs for selected genes including MYD88, CD79B, EZH2, CREBBP, EP300, KMT2-D, TP53, 2BM, CITA, GNA13, CCND3, GNAI2, P2RY8, CD70, and CDKN2A, across all data sets and platforms.

**Tumor Clonality**

To process immune data from RNA sequences into quantitated clonotypes, we applied MiXCR v2.1.7 (75). Single clonotypes were grouped into clones with unique VDJ combination and identical CDR3 nucleotide sequences. For B cells, the clones were further aggregated into clone groups if the VDJ combination was the same and CDR3 nucleotide sequences differed no more than 1 nucleotide. The biggest clone group was assigned as tumor if the absolute clonotype counts > 20, the relative clonotype counts > 5%, the ratio of the second biggest group to the first < 0.6 and the group contains an enriched clone > 25%. The tumor light chain was called if there was an enriched clonotype in one of the light chains. In cases with an enriched clone in both chains, the biggest by absolute counts was selected.

**DNA Methylation Analysis**

Library preparation, sequencing, and post-processing of the raw data were performed at the WCM Epigenomics Core Facility following the enhanced reduced representation bisulfite sequencing method (eRRBS; ref: 76). Briefly, DNA was digested and then ligated with 5-methylcytosine-containing indexed Illumina adaptors. Adapter-ligated DNA fragments were size selected and bisulfite converted using the EZ DNA Methylation Kit (Zymo Research). Amplification was performed on size-selected DNA fractions (150–250 bp and 250–400 bp), and purification steps were done using Agencourt AMPure XP (Beckman Coulter) beads. Libraries were sequenced on Illumina HiSeq2500. Bisulfite reads were aligned to the bisulfite-converted hg19 reference genome using Bismark (77). In addition to the WCM cohort (GSE145043), two publicly available data sets were used in this analysis (TCGA and GSE23967). To estimate CIMP values, we selected probes located in CpG islands in gene promoters (i.e., regions TSS200, TSS1500, 5’UTR, and first exon) excluding the genes located in chromosomes X and Y. For CIMP scores, we selected unmethylated probes (i.e., 95th percentile < 0.2) from 60 normal immune cell subtypes from data set GSE35069. Probes that were methylated (beta-value > 30%) in less than one third of tumor samples were also excluded, yielding 8,137 informative probes that were used to compute CIMP scores as their mean methylation values for samples included in the data sets TCGA and WCM. The data sets GSE23501 (gene expression) and GSE23967 (HELP methylation assay) were used to validate SMAD1 promoter methylation on one available probe and TGFβ pathway activity.

**Mouse Experiments**

All the experiments were conducted under approval of the WCM Institutional Animal Care and Use Committee at the Research Animal Resource Center. PDTX were established in female and male NOD.B2M mice (NOD.Cg-B2Mm1虔rPrkdckod Il2γ−/−;SjJ) and then serially propagated into NSG mice (NOD.Cg-Prkdckod Il2γ−/−;SjJ, The Jackson Laboratory). PDTX were analyzed by RNA-seq to determine whether molecular features from the original lymphoma were maintained in the xenografted tissue. As part of the PDTX characterization, PDTX implanted in NSG mice were treated with CHOP chemotherapy to determine whether the original patient response was reproduced in the mice. Clonal correspondence was proved via BCR rearrangement analyses. Molecular and therapeutically characterized PDTX were subsequently implanted in NSG mice to assess the antitumor effect of proteoglycans. Once tumors reached a palpable size, mice were randomized to receive vehicle (PBS), decorin (R&D Biosystems) 0.4 mg/kg intraperitoneally, or biglycan (R&D Biosystems) 0.4 mg/kg intraperitoneally according to the schedule shown in the figures. Mice were treated daily in a weekend-off schedule, and tumor sizes were recorded every other day. Body weight was used as a surrogate for drug toxicity. PDTX establishment, characterization, and treatment were conducted at the WCM, Mayer Cancer Center, PDTX Core Facility.

**Mouse Lymphoma Model.** Mouse B-cell lymphoma A20 cells were subcutaneously implanted in the inguinal lymph node pad into female and male BALB/c mice (at least 8-week-old), and once tumors were fully established and palpable, mice were randomized to receive either vehicle (saline solution) or azacitidine 12.5 mg/kg/day by subcutaneous injection for four consecutive days.Authenticated A20 cells were purchased from the ATCC repository (TIB-208). We routinely conducted analysis of Mycophilus sp. in cellular cultures by PCR. At the end of the treatment, mice were sacrificed and tumors extracted for analysis.

**Transcription and DNA Methylation Analysis of Tumors.** RNA was isolated according to the kit manufacturer’s instructions (Zymo Quick-RNA Miniprep) from cryopreserved A20 tumors and from biobanked cryopreserved whole tissues (spleen and tumors) for genetically engineered lymphoma mouse models. The Illumina TruSeq Stranded poly-A Library Prep Kit was used for cDNA preparation. Sequencing was performed using an Illumina NovaSeq 6000 SP Flowcell at WCM Genomics Resources Core Facility. Transcripts were quantified against Gencode mm38 genome using Salmon (78). Transcript abundances were summarized with tximport in R version 3.6.2. DESeq2 (79) was used for differential expression analyses. DNA was isolated according to the kit manufacturer’s instructions (Zymo Quick-DNA Miniprep). Library preparation, sequencing, and post-processing of the raw data were performed at the WCM Epigenomics Core Facility following the eRRBS method. Libraries were sequenced on Illumina NovaSeq 6000 SP Flowcell. Bisulfite reads were aligned to the bisulfite-converted mm10 reference genome using Bismark (77) and differential methylation was assessed using MethylKit (80). Briefly, promoter regions were defined as 2.5 kb upstream and 0.5 kb downstream of the transcription start site,
and differential methylation was calculated on those regions, with a minimum number of covered bases of 3. Tumor tissues were processed for paraffin embedding and CD3 IHC analysis at the WCM Comparative Pathology Core Facility. CD3-stained cells were quantified using ImageJ software.

Analysis of Murine LME Signatures

We used Kallisto v0.42.4 (81) to align RNA-sequencing reads to the transcriptome reference GRCm38.6. Transcript annotation (protein coding and noncoding), transcript to gene mapping, and annotation to human homologs for murine genes were retrieved from the Ensembl database (release 96; ref. 82). Only protein-coding genes with human homologs were used in subsequent analysis. To measure the similarity between the human lymphoma sample microenvironment and a particular murine lymphoma phenotype, we developed the microenvironment similarity (MES) metric, which is a reversed Euclidean distance between estimated percentages of different cell types in murine and human samples as follows: MES (P, i) = \left( \sum_{j} (p_{i}^{m} - p_{j}^{h})^{2} \right)^{1/2}, where P is a murine lymphoma phenotype, i is a human lymphoma sample, phumani is the estimated cell percentage of cell type i in the sample, pfire is the median estimated cell percentage of cell type i among murine samples with phenotype P.

Proteomics and Matrisome Analysis

Lymphoma tissues were processed based on a modified decellularization protocol (83) using 3% peracetic acid and incubated in a 1% Triton X-100, 2 mM/L EDTA solution for 27 hours. Then, tissues were successively incubated in distilled water for 24 hours followed by 600 U/mL DNase in PBS for 24 hours and followed by distilled water for 72 hours. Tissue fragments were then dried using a vacuum filtration system and digested and prepared for tandem mass tag (TMT) mass spectrometry analysis using a modify protocol (84). Briefly, samples were resuspended in 8 M urea and subjected to reduction with DTT and alkylation with iodoacetamide. After that, samples were diluted to 2 M urea with 25 mM/L ammonium bicarbonate and incubated with PNGase F for 2 hours at 37°C. Lys-C were added for 2-hour incubation at 37°C after deglycosylation. Trypsin was added for overnight digestion at 37°C. The resulting peptides were desalted by C18 Stage tips. Lyophilized peptides were then labeled with the TMT reagents (Thermo Fisher Scientific) according to the manufacturer’s protocol. Labeled peptides were mixed, lyophilized, and desalted prior to LC/MS analysis. An EASY-nLC 1200 coupled on-line to a Fusion Lumos mass spectrometer was used (Thermo Fisher Scientific). Buffer A (0.1% formic acid in water) and buffer B (0.1% formic acid in 80% acetonitrile) were used as mobile phases for gradient separation. A 75-µm PicoTip emitter (New Objective) was packed in-house for peptide separation. Peptides were separated with a gradient of 5%-10% buffer B over 1 minute, 10%-42% buffer B over 231 minutes, and 42%-100% B over 5 minutes at a flow rate of 300 nL/minute. Full MS scans were acquired in the Orbitrap mass analyzer over a range of 400–1,500 m/z with resolution 60,000 at m/z 200. Top 15 most abundant precursors were selected with an isolation window of 0.7 Thomson and fragmented by higher-energy collisional dissociation with normalized collision energy of 40. MS-MS scans were acquired in the Orbitrap mass analyzer. For protein identification, raw files were processed using the MaxQuant computational proteomics platform version 1.5.5.1 (Max Planck Institute) and the fragmentation spectra were used to search the UniProt human protein database. Oxidation of methionine and protein N-terminal acetylation were used as variable modifications for database searching. Both peptide and protein identifications were filtered at 1% false discovery rate based on decoy search using a database with the protein sequences reversed. To identify the abundance of ECM proteins, we used the matrisome database to find the overlapping proteins between the ones identified from the tissues and the ones present in the database of identified human ECM proteins (http://matrisomeproject.mit.edu). As a control to normalize between subsequent mass spectrometry runs, we used decellularized human tonsils from one individual and computed the ratios of the abundance of proteins found in the DLBCL tissues to the abundance found in the tonsils. Experiments were conducted at the WCM Metabolomics and Proteomics Core Facility. For samples with matched matrisome-transcriptome data, the RNA was extracted using a Qiagen RNeasy Kit following the manufacturer’s instructions. The sequencing was performed using an exome capture protocol, and libraries were constructed using Illumina RNA-Prep with enrichment (tagmentation kit with dual indexes) according to the manufacturer’s protocol and sequenced on Novaseq 6000 SP Flowcell.

Data Sharing

WCM transcriptomics and DNA methylation samples (n = 184) and PDXT transcriptomics (n = 18) are deposited in the Gene Expression Omnibus (GSE145043).

Authors’ Disclosures

N. Koletov reports personal fees from BostonGene outside the submitted work; in addition, N. Koletov has a patent for BostonGene pending and issued. A. Bagaev reports personal fees from BostonGene outside the submitted work; in addition, A. Bagaev has a patent for BostonGene pending and issued to BostonGene. Z. Antysheva reports personal fees from BostonGene outside the submitted work; in addition, Z. Antysheva has a patent for BostonGene pending to BostonGene. V. Svekolkin reports personal fees from BostonGene outside the submitted work; in addition, V. Svekolkin has a patent for BostonGene pending and issued to BostonGene. E. Tikhonova reports personal fees from BostonGene outside the submitted work. M. Ilieva reports personal fees from BostonGene outside the submitted work. N. Kuzkina reports personal fees from BostonGene outside the submitted work. F. Frenkel reports personal fees from BostonGene Corp outside the submitted work. N. Almog reports personal fees from BostonGene outside the submitted work. N. Almog reports personal fees from BostonGene outside the submitted work. N. Almog reports personal fees from BostonGene outside the submitted work. A. Fowler reports other support from BostonGene during the conduct of the study. J.P. Leonard reports personal fees from Sutro, Miltenyi, AstraZeneca, BMIS/Celgene, Regeneron, Bayer, Gilead/Kite, Karyopharm, GenMab, AbbVie, and Incyte and grants and personal fees from Epizyme and Genentech/Roche outside the submitted work. L. Cerchietti reports grants from NIH–NICI and Leukemia and Lymphoma Society during the conduct of the study; grants from Bristol Myers Squibb/Celgene outside the submitted work. No other disclosures were reported.

Authors’ Contributions

N. Koletov: Data curation, software, supervision, validation, visualization, and methodology. N. Kuzkina: Software and investigation.


A. Bagaev: Software, supervision, and methodology.

G. Inghirami: Resources, supervision, and investigation.

L. Cerchietti: Conceptualization, data analysis, formal analysis, investigation, funding acquisition, investigation, writing–original draft, editing. M.V. Ruvuelta: Formal analysis, investigation, and visualization.

J.M. Phillips: Formal analysis, investigation, and methodology.


Downloaded from cancerdiscovery.aacrjournals.org on May 16, 2021. © 2021 American Association for Cancer Research.
Microenvironmental Signatures in Lymphoma


Acknowledgments

Funding for this work was provided by the Leukemia and Lymphoma Society (LLS-SCOR 7012-16 and LLS TRP R6510-19), Italian Association for Cancer Research (AIRC 20198), and the NIH–NCI (R01CA242069).

Received June 12, 2020; revised November 24, 2020; accepted January 21, 2021; published first February 4, 2021.

REFERENCES


Clinical and Biological Subtypes of B-cell Lymphoma Revealed by Microenvironmental Signatures

Nikita Kotlov, Alexander Bagaev, Maria V. Revuelta, et al.

Cancer Discov  Published OnlineFirst February 4, 2021.

Updated version
Access the most recent version of this article at:
doi:10.1158/2159-8290.CD-20-0839

Supplementary Material
Access the most recent supplemental material at:
http://cancerdiscovery.aacrjournals.org/content/suppl/2021/01/26/2159-8290.CD-20-0839.DC1

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, use this link http://cancerdiscovery.aacrjournals.org/content/early/2021/05/14/2159-8290.CD-20-0839. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.