Variability in DNA methylation defines novel epigenetic subgroups of DLBCL associated with different clinical outcomes

Nyasha Chambwe, Matthias Kormaksson, Huimin Geng, Subhajyoti De, Franziska Michor, Nathalie A. Johnson, Ryan D. Morin, David W. Scott, Lucy A. Godley, Randy D. Gascoyne, Ari Melnick, Fabien Campagne and Rita Shaknovich
Variability in DNA methylation defines novel epigenetic subgroups of DLBCL associated with different clinical outcomes

Nyasha Chambwe,1,2,3 Matthias Kormaksson,4,5 Huimin Geng,6 Subhajyoti De,7,8,9 Franziska Michor,10,11 Nathalie A. Johnson,12 Ryan D. Morin,13,14 David W. Scott,15 Lucy A. Godley,16 Randy D. Gascoyne,17,18 Ari Melnick,18,19 Fabien Campagne,1,2 and Rita Shaknovich19,20

1The HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, 2Department of Physiology and Biophysics, 3Tri-Institutional Training Program in Computational Biology and Medicine, and 4Department of Public Health, Weill Cornell Medical Center, New York, NY; 5IBM Research-Brazil, Rio de Janeiro, Brazil; 6Department of Laboratory Medicine, University of California, San Francisco, San Francisco, CA; 7Department of Medicine, University of Colorado School of Medicine, Aurora, CO; 8Department of Biostatistics and Informatics, Colorado School of Public Health, Aurora, CO; 9Molecular Oncology Program, University of Colorado Cancer Center, Aurora, CO; 10Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA; 11Department of Biostatistics, Harvard School of Public Health, Boston, MA; 12Department of Medicine, Jewish General Hospital, Montreal, QC, Canada; 13Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, BC, Canada; 14Michael Smith Genome Sciences Centre, and 15Centre for Lymphoid Cancer, British Columbia Cancer Agency, Vancouver, BC, Canada; 16Department of Medicine, The University of Chicago, Chicago, IL; 17Department of Pathology, University of British Columbia, Vancouver, BC, Canada; and 18Department of Pharmacology, 19Division of Hematology and Oncology, and 20Department of Pathology and Laboratory Medicine, Weill Cornell Medical College, New York, NY

**Key Points**

- Unsupervised clustering of DLBCls based on DNA methylation changes identifies 6 novel epigenetic clusters.
- Greater magnitude of methylation changes correlates with worse clinical outcome.

**Diffuse large B-cell lymphoma (DLBCL)** is the most common aggressive form of non-Hodgkin lymphoma with variable biology and clinical behavior. The current classification does not fully explain the biological and clinical heterogeneity of DLBCLs. In this study, we carried out genomewide DNA methylation profiling of 140 DLBCL samples and 10 normal germinal center B cells using the HpaII tiny fragment enrichment by ligation-mediated polymerase chain reaction assay and hybridization to a custom Roche NimbleGen promoter array. We defined methylation disruption as a main epigenetic event in DLBCLs and designed a method for measuring the methylation variability of individual cases. We then used a novel approach for unsupervised hierarchical clustering based on the extent of DNA methylation variability. This approach identified 6 clusters (A-F). The extent of methylation variability was associated with survival outcomes, with significant differences in overall and progression-free survival. The novel clusters are characterized by disruption of specific biological pathways such as cytokine-mediated signaling, ephrin signaling, and pathways associated with apoptosis and cell-cycle regulation. In a subset of patients, we profiled gene expression and genomic variation to investigate their interplay with methylation changes. This study is the first to identify novel epigenetic clusters of DLBCLs and their aberrantly methylated genes, molecular associations, and survival. *(Blood. 2014;123(11):1699-1708)*

**Introduction**

Diffuse large B-cell lymphoma (DLBCL) is the most common subtype of B-cell non-Hodgkin lymphoma. DLBCLs are highly heterogeneous; only about 60% of patients are responsive to the current standard-of-care chemotherapy: a regimen of rituximab combined with cyclophosphamide, doxorubicin, vincristine, and prednisolone (R-CHOP). The remaining 40% of patients have either primary refractory or relapsed disease with dismal outcome. DLBCLs are also highly heterogeneous at the molecular level. Gene expression profiling studies have defined 3 molecular subtypes: germinal center B-cell–like (GCB) DLBCL, activated B-cell–like (ABC), and primary mediastinal B-cell lymphoma.1,2 These molecular subtypes were shown to have different prognostic outcomes, with the ABC subtype having the most unfavorable outcome. However, some cases of DLBCL cannot be classified according to their gene expression profile, suggesting that DLBCL may harbor more genomic or epigenomic complexity that is not captured by gene expression profiling.3,4 Regulation of gene expression through epigenetic mechanisms such as DNA cytosine methylation is increasingly recognized as a hallmark of cancer.5-7 DNA methylation is involved in critical processes such as normal cell development, cellular differentiation, genome imprinting, and X-chromosome inactivation.8,10 Global DNA hypomethylation in cancer contributes to genomic instability,11 whereas focal hypermethylation at promoters of tumor suppressors is recognized as contributing to neoplastic transformation.12,13 In DLBCLs, promoter hypermethylation in the DNA repair enzyme *MGMT* is significantly associated with prognosis in DLBCL.14,15 Furthermore, the importance...
of DNA methylation in the biology of DLBCLs is underscored by distinct DNA methylation profiles of ABC and GCB DLBCLs.\textsuperscript{16-18} In addition to focal changes in DNA methylation, Hansen et al\textsuperscript{19} reported increased stochastic variation in DNA methylation across solid cancers and suggested that cancer methylomes can be described in terms of their variance from their corresponding cell of origin. Following this observation, De et al\textsuperscript{20} demonstrated extensive intratumor and interpatient variability in DNA methylation in DLBCL. Building on these prior observations, we sought to investigate if DNA methylation differences between patients would help explain the observed biological heterogeneity in DLBCL patient cohorts.

We hypothesized that patterns of DNA methylation could help classify DLBCLs into distinct biologically and clinically relevant subtypes. To test this hypothesis, we carried out genomewide DNA methylation profiling in a cohort of 140 DLBCL cases and 10 normal GCB cell (NGCB) controls. We clustered DLBCL cases based on how their methylome differs from NGCBs. This process defined 6 clusters in this DLBCL cohort. We found that the magnitude of methylation changes from NGCBs associates with survival in patients who have undergone R-CHOP treatment. We also found that changes in DNA methylation at specific loci target important biological processes such as cytokine-mediated signaling, ephrin signaling, and pathways associated with apoptosis and cell-cycle regulation.

Given that lymphomas are characterized by extensive DNA methylation disruption as reported previously,\textsuperscript{20} we hypothesized that clustering DLBCLs based on degree and direction of methylation changes would produce informative biologically distinct subgroups. We quantified DNA methylation disruption in the following way: for each HELP fragment, we calculated the relative methylation difference between each DLBCL case and the mean of NGCB control samples (supplemental Figure 1; see supplemental Methods for statistical details). We estimated a histogram of these methylation differences for each DLBCL case; the histogram counts how many HELP fragments in a DLBCL genome differ from controls at a certain level of methylation change. The spread of a histogram defines the variability between a DLBCL genome and that of NGCB controls. We refer to these histograms as methylation variability profiles (MVP). We defined the sample methylation variability score (MVS) as the difference in area under the curve between a given sample’s MVP and the expected MVP of NGCBs (supplemental Figure 1).

We then carried out unsupervised hierarchical clustering of the DLBCL samples that is conceptually novel in that it uses a similarity metric based on the difference in methylation variability between 2 samples (supplemental Methods). Unsupervised clustering identified 6 DNA methylation-based clusters in this DLBCL cohort (Figure 1A). To confirm that these 6 clusters are stable and reproducible, we performed consensus clustering. Briefly, this method repeats the clustering process on subsets of the complete dataset and checks how consistently samples are clustered together. Consensus clustering confirmed $K = 6$ as an optimal choice for cluster number (supplemental Figure 2A-C).

We found a large MVS for DLBCLs, indicative of methylation changes of larger magnitude in DLBCL samples. Changes of greater magnitude are visible in the heavier left and right tails for DLBCL MVPs compared with the average NGCB MVP (Figure 1B). Clustering of the samples shows that DLBCL samples can be grouped by magnitude of methylation changes compared with controls (Figure 1B-C). DLBCL clusters were labeled A through F based on increasing magnitude of methylation changes from NGCBs, with cluster A having the smallest magnitude of methylation changes compared with NGCB and cluster F the largest. Clusters B, D, and E show a tendency toward hypomethylation in DLBCL (heavier right tail of the profiles, Figure 1B). Clusters A and C have a tendency toward hypermethylation. Cluster F shows the largest methylation changes with almost equal proportion of methylation gain and loss in different parts of the genome.

To test whether these changes occurred throughout the DLBCL genome, we assayed genomewide 5-methylcytosine (5-mC) content by liquid chromatography mass spectrometry in a subset of DLBCL tumors. We observed a global hypomethylation (mean 5-mC 4.9%) in DLBCLs compared with NGCBs (mean 5-mC 12.08%, supplemental Figure 3). However, we found that genomewide 5-mC content was similar across DLBCL clusters, ranging from around 5% for clusters A-D to 3.73% for cluster F. Therefore, global differences in genomewide content of 5-mC cannot explain the pattern of gain and loss of methylation we observed in promoter regions with the HELP assay (Figure 1B). The comparison of HELP assay results and genomewide results would suggest that the global loss of 5-mC content in DLBCLs occurs primarily in the intergenic or coding sequence areas of the genome.

The magnitude of DNA methylation changes predicts survival

We assessed the association of the DNA methylation based clusters of DLBCL with survival outcomes. Cluster identity alone did not predict

---

**Materials and methods**

**Sample collection**

A total of 140 diagnostic de novo DLBCL samples were collected from individuals with de novo DLBCL at the British Columbia Cancer Agency, Canada. Supplemental Table 1 (available on the Blood Web site) presents characteristics of the study cohort. NGCBs were obtained from leftover tonsilllectomy specimens at New York Presbyterian Hospital. All tissue collection was approved by the Weill Cornell Medical College Institutional Review Board and in accordance with the stipulations of the Declaration of Helsinki treaties.

**HELP assay and data analysis**

We measured DNA methylation using the published HpaII tiny fragment enrichment by ligation-mediated polymerase chain reaction (HELP) assay.\textsuperscript{21,22} The microarray design is documented in the Gene Expression Omnibus accession number GPL6604. Data from this study are publicly available by accessing Gene Expression Omnibus accession number GSE54200. HELP data were processed using standard pipeline as outlined in the HELP analysis.\textsuperscript{23} We estimated a histogram of these methylation differences for each HELP genomic fragment measured by the assay.

---

**Results**

**Identification of DNA methylation-based clusters in DLBCL**

We profiled DNA methylation in 140 DLBCL cases and 10 NGCB cell samples using the HELP assay and hybridization to a custom-designed Roche NimbleGen array. This array represents approximately 50,000 CpGs favoring promoter regions of 14,000 genes. We carried out data processing, quality control analysis, and quantile normalization of these data and obtained the relative methylation signal ($\log_2(HpaII/MspI)$) for each HELP genomic fragment measured by the assay.
survival outcomes (log-rank test: overall survival [OS] \( P = .375 \), progression-free survival [PFS] \( P = .139, n = 124 \); supplemental Figure 4), possibly reflecting insufficient number of patients in each cluster. We tested the prognostic significance of the IPI, a widely accepted standard prognostication model in DLBCL. IPI was significantly associated with OS but not PFS (log-rank test OS \( P = .089 \), PFS \( P = .259 \), Figure 2A) in our cohort. We also studied clinical outcomes by dividing patients based on the median MVS. The high-risk group is composed patients with MVS above the median; the low-risk group is composed of patients with an MVS below the median. We observed a statistically significant difference in survival between high- and low-risk groups (log-rank test OS \( P = .036 \), PFS \( P = .023 \),
Patients with a larger magnitude of methylation changes compared with NGCB display poorer survival outcomes compared with patients with smaller magnitude of methylation changes. The univariate Cox proportional hazard model shows that the MVS is moderately predictive of OS ($P = .072$) and predicts PFS ($P = .029$) (Table 1). We performed a multivariate Cox analysis for OS and PFS using IPI and MVS as predictors. After accounting for IPI, MVS is a significant predictor of PFS ($P = .03$) and is a moderately significant predictor of OS ($P = .07$) (Table 2). These findings suggest that classifying patients according to the extent of their methylation divergence from normal B cells is a useful factor in building prognostic models for DLBCL because it performs comparably with IPI in univariate analysis and remains significant in a multivariate model with both factors.

**Characteristics of epigenetic clusters**

We then investigated how each of the 6 DLBCL clusters differed from controls. We carried out differential methylation analysis between each DLBCL cluster and NGCBs. This analysis produced the signatures presented in Figure 3A (supplemental Table 2). In line with the extent of methylation disruption shown in Figure 1B, we observed increasing amounts of methylation changes: from cluster A...
with 49 fragments (47 genes) to cluster F with 9114 fragments (7361 genes) (Figure 3A). Cluster B (74%), D (79%), and E (70%) signatures showed predominant hypomethylation in DLBCL (Figure 3B). Fifty-five percent of cluster A and 84% of cluster C signature fragments were significantly hypermethylated (Figure 3B). Cluster F showed extensive methylation changes that affected 9114 of 25,625 fragments (48% fragments were hypomethylated and 52% hypermethylated).

Interestingly, we found that clusters B, D, and E have a substantial overlap in aberrantly methylated fragments: 53 of 65 signature fragments from cluster B are also aberrantly methylated in cluster D and 408 of 439 signature fragments from cluster D are also aberrantly methylated in cluster E (Figure 3C). These results suggest a possible progressive accumulation of aberrant methylation in the genome of some DLBCL patients. Although these results are suggestive, definitive demonstration of progressive accumulation of changes would require measuring methylation over time for the same patients.

We performed technical validation of methylation levels for a subset of genes in these signatures using Sequenom MassARRAY Epityper as an orthogonal method. We selected 10 fragments and epityped 10 DLBCL cases at these locations. HELP and Sequenom MassARRAY estimates are highly correlated ($r^2 = 0.7$) (supplemental Figure 5). We also confirmed methylation status of some of the biologically important signature genes. p15/CDKN2B was hypermethylated, whereas BTG2 was hypomethylated in clusters B, C, D, and E (Figure 4A-B). We confirmed hypermethylation in the promoter region of CCR6 in cluster A (Figure 4C), RUNX1 (Figure 4D), and WNT2 (Figure 4E).

The gene expression–based molecular subtypes ABC and GCB DLBCL are well characterized and validated. We investigated how the ABC/GCB classification was related to the DNA methylation–based clusters (supplemental Figure 6). We found that DNA methylation clusters are not exclusive of a particular gene-expression subtype. Clusters A, B, C, and D with lower methylation disruption have higher frequencies of GCB DLBCLs (60%, 50%, 83%, and 56%, respectively), whereas cluster E has the highest frequency of ABC-DLBCL (78%)—much higher than the overall cohort ABC-DLBCL frequency of 30%. Limited conclusions can be made for cluster F with 3 samples in the cluster.

We then investigated which biological functions were over-represented in the genes that compose the different DLBCL cluster signatures (Figure 5). Cluster A’s differentially methylated genes were involved in the cytokine–mediated signaling pathway (STAT3, TNFRSF1A, and KRAS, supplemental Figure 7). Cluster B signature was enriched in genes contributing to multicellular organismal homeostasis (e.g., CALD1, GIMAP), and T-cell activation (CD3D, CD3G) (supplemental Figure 8). Cluster C was characterized by hypermethylation of many important developmental genes, particularly homeobox and forkhead box family genes (supplemental Figure 9). We found that the tricarboxylic acid cycle is one of the top canonical pathways in cluster D (supplemental Figure 10). Of note, IDH2 belongs to this pathway and is significantly hypomethylated in clusters D, E, and F. IDH1 and IDH2 mutations in acute myeloid leukemia (AML) are associated with hypermethylation.24 Aberrant methylation of the ephrin signaling pathway was a hallmark of cluster E (supplemental Figure 11), with aberrant hypermethylation of EPHA5 and PIK3CG, hypermethylation of EPHB1, the tyrosine-protein kinase FYN, GRB7, GNAO1, and PXN, and ephexin. Many processes that contribute to a malignant phenotype are enriched in cluster F, such as regulation of apoptotic processes and aberrant methylation of cell-cycle genes as well as many signal transduction pathways associated with cancer (protein kinase B signaling, inhibition of extracellular signal-regulated kinase, or 5′ adenosine monophosphate-activated protein kinase signaling, supplemental Figure 12). Additional details about pathway analysis can be found in supplemental Results.

Previous reports had shown that lymphomas aberrantly methylate a subset of the targets of the PRC2 polycomb complex (EZH2 is the catalytic subunit of the complex).25,26 We found that differentially methylated genes were enriched for targets of EZH2 with statistically significant enrichment in clusters C, D, and E (supplemental Figure 13, hypergeometric test q value <0.05). More than 60% of the EZH2 targets present in each cluster signature were hypermethylated compared with NGCBs. Examples of hypermethylated EZH2 targets included CDKN2A, CDKN2B, NID1, HOXA9, HOXD8, ERICH1, and EPHA5 (supplemental Table 3).

We also investigated if there were genes that were aberrantly methylated in all DLBCLs and thus defined a lymphoma-specific methylation signature. We found 200 differentially methylated genes when comparing DLBCL and NGCBs (supplemental Figure 14). These commonly aberrantly methylated genes were enriched in cell-adhesion genes—in particular, proto-cadherins. Interestingly, CDKN2B was hypermethylated in all DLBCL clusters except A, suggesting a possible early event in lymphomagenesis. Epigenetic deregulation of the INK4A-ARF cluster appearance to be a common and progressive event in lymphomagenesis because more deregulated clusters B and D displayed hypermethylation of CDKN2B and CDKN2B-AS1 and the most deregulated clusters C and E in addition to hypermethylating CDKN2B and CDKN2B-AS1 also displayed hypermethylation of CDKN2A. Most deregulated cluster F also had hypermethylation of other cell-cycle regulators such as CDKN1A, CDKN1B, CDKN2D, and CDKN2AIP. Our data show that the INK4A, CDKN1A, and CDKN1B genes display aberrant methylation in DLBCLs.

We asked if DNA methylation changes could correlate with genomic changes in the samples. To this end, we measured and analyzed copy number changes using single nucleotide polymorphism data from a subset of analyzed DLBCLs. We identified broad regions of genomic amplification and deletion in this cohort using the GISTIC algorithm (supplemental Table 4).27 3q, 7q, 11q, and 18q amplifications and 9q deletions were the most frequently observed genomic changes in this cohort (supplemental Figure 15). The 3q amplification has been reported before and contains the NFKBIZ gene.28 NFKBIZ can bind to NFKB and activate downstream signaling of NFKB, resulting in upregulation of interleukin-6 among other targets.29 Activation of NFKB and IL-6 signaling through STAT3 both contribute to the

### Table 1. Univariate Cox proportional hazards models for OS and PFS

<table>
<thead>
<tr>
<th>Variable</th>
<th>n</th>
<th>Hazard ratio</th>
<th>95% CI</th>
<th>P</th>
<th>Hazard ratio</th>
<th>95% CI</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPI</td>
<td>102</td>
<td>1.37</td>
<td>1.05-1.79</td>
<td>0.02*</td>
<td>1.28</td>
<td>1.01-1.62</td>
<td>0.04*</td>
</tr>
<tr>
<td>MVS</td>
<td>124</td>
<td>151.70</td>
<td>0.64-35 848</td>
<td>0.07</td>
<td>219.40</td>
<td>1.74-27 630</td>
<td>0.03*</td>
</tr>
</tbody>
</table>

*Statistically significant.

### Table 2. Multivariate Cox proportional hazards models for OS and PFS

<table>
<thead>
<tr>
<th>Variable</th>
<th>n</th>
<th>Hazard ratio</th>
<th>95% CI</th>
<th>P</th>
<th>Hazard ratio</th>
<th>95% CI</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPI</td>
<td>102</td>
<td>1.38</td>
<td>1.05-1.60</td>
<td>0.02*</td>
<td>1.28</td>
<td>1.01-1.62</td>
<td>0.04*</td>
</tr>
<tr>
<td>MVS</td>
<td>102</td>
<td>471.77</td>
<td>0.63-35 500</td>
<td>0.07</td>
<td>718.79</td>
<td>2.29-225 400</td>
<td>0.03*</td>
</tr>
</tbody>
</table>

*Statistically significant.
proliferative potential of DLBCLs. The 18q amplification was present in all clusters with greater frequency in clusters E and F. This amplification has been reported previously to be more prevalent in ABC-type DLBCLs, which is consistent with our data.\textsuperscript{28} BCL2, an antiapoptotic protein playing a pivotal role in the pathogenesis of many lymphoma subtypes, was reported to be the most overexpressed gene as a result of this amplification.\textsuperscript{28} Interestingly the 6q deletion is found in less than 10% of cluster A and is absent in cluster F cases. The deleted arm of 6q contains the candidate tumor suppressor gene PRDM1, which is crucial for plasmacytic differentiation.\textsuperscript{30}

Similar to the magnitude of methylation difference from NGCBs, genomic instability increases from cluster A to cluster F. The 3q, 7pq, 11q, and 18q amplifications are enriched with increasing frequency from cluster A to F (2-sided Fisher exact test, \( P \leq 0.01 \)). We sought to rule out that genomic aberrations alone could explain the patterns of methylation variability observed in this cohort. To this end, we identified the genomic regions where no significant amplifications or deletions were detected by GISTIC. We calculated the MVS based on the HELP fragments that map to these regions and observed a similar pattern of increased MVS from clusters A to F that we observed when all fragments were used (supplemental Figure 16). These results show that both genomic aberrations and DNA methylation changes compared with normal increase from patients in cluster A to patients in cluster F. Additionally, we ruled out that sample purity is not the cause for different methylation variability between the clusters (supplemental Figure 17).

Concordant changes in DNA methylation and gene expression

We integrated DNA methylation and gene expression data to look for genes whose regulation could be associated with DNA methylation status. Gene expression was assayed in 52 samples spanning each DNA methylation cluster. We determined genes that were significantly up or downregulated in DLBCL clusters compared with normal. For these differentially expressed genes, we examined whether methylation was perturbed in those genes for each cluster. This analysis showed that 14% of cluster A and 11% of cluster B RefSeq transcripts show an inverse correlation with expression, whereas for all other clusters less than 5% of the methylation signature falls in this category (supplemental Tables 5-6). Inversely correlated between methylation and expression across clusters are genes such as CD3D, NMB, GZMK, and VSTM3 (Table 3). These genes have immune

![Heat map representation for the HELP fragments that are differentially methylated between NGCB cells and DLBCL cases in each cluster (moderated Student t test \( q \) value \(<0.05 \) and log fold change \( \geq 1.5 \)). Each row represents a single HELP fragment (probe set) and each column a single patient/normal sample. Yellow represents highly methylated (hypermethylated) fragments and blue represents fragments with lower methylation (hypomethylated). (B) Bar plot showing relative abundance of methylation gain and loss for each cluster signature. (C) Venn diagram depicting the overlap of differentially methylated HELP fragments among clusters B, D, and E.
functions such as lymphocyte activation and T-cell activation. Enzymes that act against guanosine triphosphate in the immunity-associated protein family (GIMAPs) are known to regulate lymphocyte survival. Here we find that GIMAP1 and GIMAP5 are hypomethylated and overexpressed in DLBCLs. We found that ASXL1 was hypermethylated and downregulated in cluster E and F DLBCLs. ASXL1 is a tumor suppressor gene that is associated with the repressive polycomb complex PRC2.

**Discussion**

Extensive gene expression profiling studies of DLBCLs resulted in identification of several molecular subgroups of clinical significance, including ABC-like, GCB-like, and primary mediastinal B-cell lymphoma subtypes. The biology of these subgroups is not entirely explained by genomic events and transcriptional programs, suggesting an additional layer of regulation. Recently, somatic mutations have been identified in components of the epigenetic machinery—such as EZH2, CBP/p300, and MLL2—that shed light on the significance of epigenetic regulation in normal B-cell development and in lymphomagenesis. In addition to histone modifiers and small noncoding RNAs, chemical modifications of DNA such as cytosine methylation emerged recently as paramount in regulating genome stability and gene expression. Targeted studies identified several loci with altered DNA methylation in DLBCL, including INK4A, MGMT, and BCL2. Following these observations, we asked whether such changes are widespread in the genome of DLBCL.
patients and used a genomewide approach to measure DNA methylation at more than 14,000 promoters.

A key finding in our study is that the magnitude of methylation changes and the number of gene promoters perturbed in DLBCLs compared with NGCBs correlates with clinical outcome. The magnitude of methylation changes is related to the concept of epigenetic variability. Epigenetic variability has been detected in other cancers such as colon, breast, and lung, and results in the loss of the bimodal distribution of methylation that is normally observed in normal healthy tissues. This feature so far has not been described in other hematologic malignancies, which are characterized by aberrant methylation of a specific set of genes. In AML, epigenetic signatures define most cytogenetic AML subtypes. The mechanisms that are implicated in aberrant DNA methylation in other cancers such as AML and pre-B–acute lymphoblastic lymphoma such as mutations in DNMT3A, IDH1/2, and TET1/2 have not been identified in DLBCLs, whereas changes in the level of expression of methyltransferases have been, setting this subtype of B-cell non-Hodgkin lymphoma apart. We also proposed in our earlier work that factors such as AID and CTCF may play a role in creating methylation variability in DLBCLs. Here, we defined MVP and MVS as novel quantitative measures of methylation disruption that can be applied to other tumor types. In this study, the MVP and MVS measures specifically address the extent of intrasample heterogeneity. We found that the magnitude of DNA methylation changes across the genome defines 6 clusters among 140 patients.

The underlying cause for increased magnitude of methylation changes in DLBCLs may lie in their cell of origin. The NGCB cell at the origin of DLBCL tumors is known to possess increased genomic and epigenomic mutability because of its ability to suppress DNA repair mechanisms to allow physiologic somatic hypermutation and class switch recombination. This phenomenon of epigenetic variability in DLBCLs may be an underlying cause of clonal evolution and chemoresistance. Technical approaches measuring intrasample variability will be necessary to determine the contribution of DNA methylation to clonal evolution in these tumors.

ABC DLBCLs have been shown to have poorer prognosis compared with GCB DLBCL. Here we show that DLBCLs with high levels of methylation disruption compared with NGCBs have poorer survival outcomes and are enriched in ABC DLBCLs. Based on these data, we can postulate that extensive methylation disruption and the ABC signature are associated and result in more aggressive forms of DLBCL.

We confirmed the hypermethylation of EZH2 targets in our cohort. This finding has been reported in smaller cohorts of patients and reflects aberrant colocalization of these methylation marks with H3K7me3 on targets that are normally repressed by EZH2 and PRC2 complex in embryonic stem cells. Deleterious consequences of a common EZH2 mutation resulting inmarkedly upregulated H3K27me3 in DLBCLs are further enhanced by colocalizing the inhibitory DNA methylation mark. Our data revealed that the most deregulated clusters E and F have hypermethylation and downregulation of another member of PRC2 complex tumor suppressor gene ASXL1. Mutations in ASXL1 are associated with poor outcomes in hematopoietic malignancies such as in AML. Loss of ASXL1 through mutation results in impaired PRC2 function; thus, H3K27me3 is depleted. As a result, DNA methylation may represent an alternative pathway to repress ASXL1 as seen in DLBCL clusters E and F.

Table 3. RefSeq transcripts inversely correlated between DNA methylation and expression

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Hypermethylated underexpressed in DLBCL</th>
<th>Examples</th>
<th>Hypermethylated overexpressed in DLBCL</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>FLJ40869, UBE2J1</td>
<td>4</td>
<td>CD3D, GIMAP1, NMB, VSTM3, FLJ40869</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>—</td>
<td>6</td>
<td>CALD1, CD3D, GIMAP1, NMB, RHOBTB3, VSTM3</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>CBX5, PIK3CG, PPAT</td>
<td>1</td>
<td>GZMK</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>COPG2, UBE2J1</td>
<td>9</td>
<td>CALD1, CD3D, DPT, GIMAP5, GUCY1B3, GZMK, NMB, PLSCR4, S100A9</td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>ASXL1, CYSL1, EDEM1, FBXL7, PIK3CG, SHROOM2, UBE2J1</td>
<td>30</td>
<td>CALD1, CD3D, DPT, GIMAP5, GUCY1B3, GZMK, NENF, NMB, RHOBTB3, SERPING1, TNFRSF1B</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>ASXL1, BCL7A, CDK1, CHEK1, ETS1, MS12, SERPING1, PPAT, RECLQ4, SMARCD2, SOX5, STAT5B, STIL, TAF15, THRAP3, UHFR2</td>
<td>87</td>
<td>BTN3A1, CALD1, CD37, CD3D, CD63, CXCL13, FLT3, FXYD2, GIMAP7, GTF3A, HLA-A, SERPING1, TSC22D4, VSTM3</td>
<td></td>
</tr>
</tbody>
</table>
A common aberrant epigenetic event in DLBCLs also observed here is the aberrant methylation of the \textit{INK4B-ARF-INK4A} locus. This appears to be a progressive oncogenic event that is more common in more aggressive DLBCLs (clusters E and F). Prior reports highlighted frequent deletion of the \textit{INK4B-ARF-INK4A} locus in patients with DLBCL\textsuperscript{35,36,50,51} and suggested that an alternative mechanism of gene inactivation through aberrant hypermethylation also exists and cumulatively with deletions may affect between one-third and one-half of patients with DLBCLs. In addition, we demonstrated aberrant hypermethylation of \textit{CDKN1A} and \textit{CDKN1B}, which is a novel finding. Correlation of lower expression of \textit{CDKN1A} and \textit{CDKN1B} in lymphomas with higher proliferative capacity has been reported before without addressing the mechanism\textsuperscript{52,53}. Methylation of tumor suppressor genes that have a cell-cycle regulatory role in DLBCLs may provide a rationale for treatment with demethylating agents.

Our clustering study suggests a model for the pathogenesis of DLBCLs and identifies DNA methylation–based molecular states that underlie this process. Functional clustering based on the magnitude of methylation disruption underscores the existence of several subtypes of DLBCL with variable patterns and magnitude of DNA methylation change compared with the normal cell of origin (NGCB in this instance). Our data suggest that some epigenetic subtypes may be interrelated and may result from progressive accumulation of aberrant epigenetic changes (such as subtypes B, D, and E). Other subtypes may arise independently and possibly with different lead time to diagnosis, but eventually ending up in certain predictable aberrant methylation states. These aberrant states of methylation must be predicated on the underlying molecular defects, which are still under investigation.

In summary, we defined novel epigenetic subgroups of DLBCLs and analyzed their unique biological features, deregulated signature genes, and revealed potential novel therapeutic targets. We also developed a method to measure methylation disruption in lymphomas that could be useful for risk stratification.

### Acknowledgments

The authors thank Dr Anja Mottok for contributing her expertise in pathology to this study.

This work was supported by grants from the Tri-Institutional Training Program in Computational Biology and Medicine (N.C.), the National Institutes of Health Clinical Investigator Award (K08 CA127353), and The Leukemia & Lymphoma Society (6304-11) (R.S.) (U54CA143798) (S.D. and F.M.).

### Authorship


**Conflict-of-interest disclosure:** The authors declare no competing financial interests.

Correspondence: Rita Shaknovich, Weill Cornell Medical College, 1300 York Ave, Room 620C, New York, NY 10065; e-mail: ris9004@med.cornell.edu; and Fabien Campagne, Weill Cornell Medical College, 1300 York Ave, Box 140, New York, NY 10065; e-mail: fac2003@campagnelab.org.

### References

25. Martin-Subero JI, Kreuz M, Bibikova M, et al; Molecular Mechanisms in Malignant Lymphomas Network Project of the Deutsche Krebsfelle. New...


