DNA methylation signatures define molecular subtypes of diffuse large B cell lymphoma

Rita Shaknovich1,2,*, Huimin Geng1,3,*, Nathalie A. Johnson4, Lucas Tsikitas1, Leandro Cerchietti1, John M. Greally5, Randy D.Gascoyne4**, Olivier Elemento3** and Ari Melnick1,**.

1 Division of Hematology/Oncology, Department of Medicine, Weill Cornell Medical College, New York, NY; 2 Division of Immunopathology, Department of Pathology, Weill Cornell Medical College, New York, NY; 3 Institute for Computational Biomedicine, Weill Cornell Medical College; New York, NY; 4 Department of Pathology, British Columbia Cancer Agency, Vancouver, Canada; 5 Department of Genetics, Albert Einstein College of Medicine, Bronx, NY.

* R.S. and H.G. contributed equally to this study.

Running Title: Epigenetic signatures in DLBCL

Type of article: e-Blood

Scientific Section: Lymphoid Neoplasia

The online version of the article contains a data supplement.

** To whom Correspondence should be addressed:

Ari Melnick, MD
Division of Hematology/Oncology, Department of Medicine
Weill Cornell Medical College
1300 York Ave, New York, NY 10065
email: amm2014@med.cornell.edu

Olivier Elemento, PhD :
Assistant Professor
Weill Cornell Medical College
Institute for Computational Medicine
Department of Physiology and Biophysics
1305 York Avenue, New York, NY, 10021
email: ole2001@med.cornell.edu

Randy D Gascoyne, MD
Research Director, Center for Lymphoid Cancers
BC Cancer Agency & BC Cancer Research Center
675 W 10th Avenue, Vancouver, BC V5Z 1L3
email: rgascoyn@bccancer.bcca

Copyright © 2010 American Society of Hematology
Abstract

Expression profiling has revealed two major and clinically distinct subtypes of diffuse large B-cell lymphomas (DLBCLs): germinal-center B-cell-like (GCB) and activated B-cell-like (ABC) DLBCLs. Further work has shown that these subtypes are partially characterized by distinct genetic alterations and different survival. Here, we show using an assay that measures DNA methylation levels of 50,000 CpGs distributed among >14,000 promoters that these two DLBCL subtypes are also characterized by distinct epigenetic profiles. DNA methylation and gene expression profiling were performed on a cohort of 69 DLBCL patients. After assigning ABC or GCB labels using a Bayesian expression classifier trained on an independent dataset, a supervised analysis identified 311 differentially methylated probesets (263 unique genes) between ABC and GCB DLBCLs. Integrated analysis of methylation and gene expression revealed a core TNFα signaling pathway as the principal differentially perturbed gene network. Sixteen genes overlapped between the core ABC/GCB methylation and expression signatures and encoded important proteins such as IKZF1. This reduced gene set was an accurate predictor of ABC and GCB subtypes. Collectively the data suggest that epigenetic patterning contributes to the ABC and GCB DLBCL phenotypes and could serve as useful biomarker.
Introduction

DLBCL (Diffuse large B-cell lymphoma) is the most common B-cell malignancy and is highly heterogeneous from both clinical and molecular standpoints. Gene expression profiling of primary DLBCL cases identified biologically distinct subtypes of DLBCL. One such approach classified DLBCL into “germinal center B cell-like” (GCB) and “activated B cell-like” (ABC) DLBCL subtypes, based on similarities of the respective gene signatures to normal germinal center B-cells and activated peripheral B-cells, respectively. This subclassification is clinically significant and predicts overall and progression-free survival in patients treated with CHOP and R-CHOP.

Several years after the discovery of these subtypes, the mechanisms controlling gene expression in ABC and GCB DLBCLs are still only partially understood. For example, ABC DLBCLs feature aberrant activity of NFkB signaling, in part due to mutations in upstream components of this pathway. Chromosomal translocations including 3q27 or t(14;18) often do not correlate with the affected protein expression and do not alone define lymphoma phenotypes. Lenz et al used genome-wide copy number analysis to describe 30 recurrent but relatively infrequent chromosomal aberrations with DLBCL subtype-specific frequencies. ABC DLBCLs were characterized by deletion of SPIB, deletion of INK4a/ARF and trisomy 3. GCB DLBCLs showed preferential deletion of PTEN and amplification of the locus encoding the mir17-92 microRNA. In another study, Compagno et al showed that >50% of ABC DLBCLs carry somatic mutations in multiple effectors of NF-kB, which is required for survival of ABC DLBCLs. Despite all of these findings, the biological differences between two subtypes of DLBCLs are not fully understood.

Gene expression patterning is also affected by epigenetic modifications such as methylation of CpG dinucleotides. In normal development and homeostasis, cytosine methylation mediates gene imprinting, X chromosome inactivation, tissue specific gene expression, and silencing of parasitic DNA elements. Aberrant distribution of cytosine methylation is a hallmark of tumors, and involves aberrant hypermethylation and hypomethylation of promoters, as well as redistribution of intergenic DNA methylation. Aberrant DNA methylation of specific gene loci has been reported in DLBCL. For example, the MGMT promoter is hypermethylated in 39% of DLBCLs and is associated with favorable prognosis. On a more global level, Rahmatpaha et al. studied gene methylation patterns in 43 small B-cell lymphomas using differential methylation
hybridization and showed that 256 genes are differentially methylated between Small lymphocytic lymphoma, Mantle cell lymphoma and Follicular lymphoma. Martin Subero et al. examined a set of 1,505 CpGs across 807 genes using the Illumina GoldenGate Methylation Cancer panel in a set of 83 mature aggressive B-cell Lymphomas, and identified using supervised analysis, a group of 56 hypermethylated genes in lymphoma. They subsequently showed that these genes are enriched in target genes of Polycomb (PcG) complex in stem cells, thus suggesting interplay between these two types of epigenetic repressive mechanisms in lymphoma cells. Importantly for the present study, the authors did not detect any difference in DNA methylation pattern between ABCs vs GCB cell-of-origin subtypes of DLBCL. In a more recent study, Pike et al. was able to identify 15 differentially methylated genes (e.g., FLJ21062, GNMT, ONECUT2, CYP27B1, DRD1, KL, MINT2, and NEUROG1) between ABC and GCB DLBCLs using a CpG island microarray with 4,395 probes in 27 cases. Since these studies were limited in both sample size and scope, we wondered whether a more comprehensive DNA methylation platform might still distinguish more extensive aberrant epigenetic patterning associated with the ABC and GCB DLBCL subtypes. Using the HELP assay, we examined the status of 50,000 CpGs distributed among 14,000 promoters in a cohort of patients with DLBCL who were uniformly treated with R-CHOP, and integrated these data with gene expression profiles obtained from the same samples. We find that ABC and GCB DLBCLs do indeed display distinctive epigenetic profiles, involving biological pathways of likely significance, and that there is a partial overlap between ABC and GCB DNA methylation and gene expression signatures. Collectively these data suggest that aberrant epigenetic patterning contributes to the phenotype of these two major DLBCL subtypes.
Materials and methods

DLBCL samples
Specimens were obtained at diagnosis from 69 patients with de novo DLBCL in Vancouver at the British Columbia Cancer Agency. Cases were selected based on the presence of at least 80% of the neoplastic cells within the tumor section. The use of human tissue was approved by the research ethics board of the Vancouver Cancer Center/University of British Columbia and Weill Cornell Medical Center. Patients were selected based on the availability of tissue, and independent of the outcome. All patients were treated with R-CHOP (Supplemental Table 1). The primary end-points of the study were the overall survival (OS) and progression-free survival (PFS).

High molecular weight DNA extraction
100 mg or 3 mm$^3$ of frozen tissue was cut into small pieces and submerged in liquid nitrogen, followed by pestle pulverization. Liquid nitrogen was allowed to evaporate and the powder was quickly collected and transferred to the Eppendorf tube on ice. DNA purification was done using the Qiagen Puregene Gentra cell kit (Qiagen, Valencia, CA). DNA was diluted in water or 10mM Tris-HCl pH 8.0 and the quality was assessed in 1% agarose gel.

Array-based methylation analysis using HELP
The HELP assay was performed as previously published. $^{18,19}$ One microgram of high molecular weight DNA was digested overnight with isoschizomer enzymes HpaII and MspI respectively (NEB, Ipswich, MA). DNA fragments were purified using phenol/chloroform, resuspended in 10mM Tris-HCl pH 8.0, and used immediately to set up the ligation reaction with MspI/HpaII-compatible adapters and T4 DNA ligase. Ligation-mediated PCR was performed with enrichment for the 200 to 2000 base pair (bp) products, and was submitted for hybridization to Roche NimbleGen, Inc. (Madison, WI). We used the HG_17 human promoter custom array covering 25,626 HpaII amplifiable fragments within the promoters of the genes. Data quality control (QC) and analysis was performed as described previously$^{20}$, using R software and Bioconductor package (http://www.bioconductor.org/). Probesets with intensity <2.5 mean absolute deviation (MAD) of the random probesets on the array were marked as missing values. After QC processing, a median normalization was performed on each array by subtracting the median log-ratio (HpaII/MspI) of that array (resulting in median log-ratio of 0 for each array).
Differentially methylated probesets between the ABC and GCB DLBCLs were determined by the t-test. The significance threshold was set to p<0.001, and concurrently false discovery rate was determined using the Benjamini and Hochberg method. In addition, we required that to be considered differentially methylated, the methylation difference for each gene between the means of ABC and GCB groups had to be greater than 20% (i.e. ratio of the mean methylation in the two groups >1.5 from HELP, which corresponds to 20% difference by MASSArray EpiTYPING, see Supplemental Figure 2).

**Single locus quantitative DNA methylation assays.**

EpiTYPER assays (Sequenom, CA) were performed on bisulfite-converted DNA, as previously described. EpiTYPER primers were designed to cover the flanking HpaII sites of selected HpaII Amplifiable Fragments (HAF), as well as any other HpaII sites found up to 2,000 bp upstream of the downstream site and up to 2,000 bp downstream of the upstream site, in order to cover all possible alternative sites of digestion. For the biological validation of the 16 gene overlap signature, primers were designed to cover CpG dense areas of interest associated with the respective HAFs. The primers were designed using Sequenom EpiDesigner beta software (http://www.epidesigner.com/). The primer sequences are available in Supplemental Table 2.

**Quantitative Real Time PCR**

Total RNA was extracted from 10^7 cells using RNeasy mini kit from Qiagen (Valencia, CA), and eluted in RNAse-free water. cDNA synthesis was done using the Superscript III First Strand Kit from Invitrogen (Carlsbad, CA). All primer sequences are available in Supplemental Table 3.

**Gene expression profiling and data processing**

Total RNA was extracted from 69 fresh frozen tissue using the All Prep kit (Qiagen, CA), was reverse transcribed and hybridized to Affymetrix HG U133 plus 2.0 arrays according to the manufacturers protocol (Affymetrix, CA). CEL files were processed with Microarray Suite version 5.0 (MAS 5.0) using Affymetrix default analysis settings and global scaling as normalization method. The trimmed mean target intensity of each array was set to 500.

**Expression-based classification of ABC and GCB subtypes**

In order to assign gene expression based ABC and GCB labels to our 69 DLBCLs, we used the gene expression microarray data of 203 DLBCLs (GEO accession number GSE11318) with
known ABC and GCB labels as the training data, and the 185 Affymetrix probesets as the ABC and GCB gene expression signature in a Bayesian predictor. We normalized the gene expression data of our 69 DLBCL cases with the 203 training DLBCL cases using BRB-ArrayTools with the median array as a reference array (http://linus.nci.nih.gov/BRB-ArrayTools.html). Among 203 DLBCLs, there were 74 ABC DLBCLs, 72 GCB DLBCLs, 31 PMBLs and 26 unclassified DLBCLs (Supplemental Figure 1). We only used ABC and GCB DLBCLs to train the Bayesian predictor. We replicated the procedure of Bayesian predictor as described in Wright et al. for ABC/GCB classification of DLBCLs. A tumor is classified as ABC or GCB subtype if the probability that it belongs to the ABC or GCB subgroup is greater than 0.9; otherwise it is unclassifiable (Supplemental Table 4). Differentially expressed probesets between our ABC and GCB DLBCLs were determined by the t-test with p<0.001.

Pathway analysis
We explored the association of gene sets with Gene Ontology (GO) terms, canonical pathways (KEGG and Biocarta), and lymphoid-specific gene expression signatures curated by the laboratory of Dr. Staudt (Supplemental Tables 5 and 6). Fisher’s Exact test was used to calculate enrichment p values and the Benjamini-Hochberg (BH) method was used for the multittest adjustment and False Discovery Rate (FDR) control. We also used the Ingenuity Pathway Analysis software (IPA) (Redwood City, CA) to identify deregulated gene networks. We first integrated the most differentially expressed and methylated genes in ABC vs. GCBs present on both arrays: out of 263 differentially methylated genes 239 were present on both platforms, and out of 622 differentially expressed genes, 411 were present on both platforms (for the lists, see supplemental information).

Motif analysis
Finding Informative Regulatory Elements (FIRE) software was used to find consensus motifs among the differentially methylated probes in the 311 methylation signature.
Results

ABC and GCB DLBCLs have distinct epigenetic signatures.

In order to determine whether GCB and ABC DLBCLs feature unique DNA methylation signatures, we selected a cohort of 69 cases of primary DLBCL uniformly treated with R-CHOP and with available high quality tissue for microarray analysis (For patient characteristics, see Supplemental Table 1). Prior to determining whether GCB and ABC DLBCLs had distinct DNA methylation signatures it was necessary to identify each case as belonging to the ABC or GCB subtype based on their gene expression signature. Gene expression profiles were obtained using Affymetrix HG U133 plus 2.0 microarrays. We used a Bayesian predictor and the published 185-gene expression signature from Lenz et al to assign ABC or GCB labels to our DLBCL cases. The Bayesian predictor was trained on a publicly available gene expression dataset consisting of 203 DLBCL cases with known ABC and GCB labels (Supplemental Figure 1). The predictor was then used to divide our patients into 20 ABC, 40 GCB DLBCLs with a probability greater than 0.9 (Figure 1A and Supplemental Table 4). Nine cases were unclassifiable at this threshold. Kaplan-Meier estimates in these patients demonstrated an overall five-year survival of 85% for GCB and 67% for ABC cases (log rank p = 0.043) and progression-free five-year survival of 83% for GCB and 37% for ABC (log rank p=0.0026) (Figure 1B), consistent with findings by other groups.

DNA methylation profiles were determined on the same 69 patients using the HELP assay and a NimbleGen microarray representing >50,000 CpGs contained within the promoters of 14,000 genes. After data processing and normalization to establish the relative methylation level of each probeset, we performed a series of technical validations to confirm the reliability of HELP in determining percent methylation of CpGs. Five randomly selected high variance genes (p53AiP1, S100A9, B2M, CSF2, TREML2) in four randomly selected DLBCL cases were assessed by a direct quantitative DNA methylation sequencing method, MassARRAY EpITYPING. MassARRAY and HELP displayed high correlation (R=0.91) (Supplemental Figure 2), indicating that HELP values accurately reflect the CpG methylation status of the various genes.

We next performed a supervised analysis on the methylation profiles of the 60 DLBCLs labeled as ABC or GCB based on gene expression profiling to determine whether any genes were differentially methylated between the two subtypes. We identified 311 differentially methylated
probesets corresponding to 263 unique genes using t-test with p<0.001 (which corresponds to FDR<0.064 after multitest adjustment) and a methylation difference greater than 20% (i.e. ratio of the mean methylation in the two groups >1.5 from HELP, which corresponds to 20% difference by MASSArray Epityping based on the curve in Supplemental Figure 2) between ABC and GCB DLBCLs (Figure 2A). This 311-probeset methylation signature assigned ABC and GCB labels to DLBCL cases with the predictive accuracy of 91%, when using the Bayesian predictor and a probability greater than 0.8. This is comparable to the performance of the gene expression signature. When considering ABC and GCB labels assigned according to methylation profiles, the progression-free five-year survival of GCB cases was 83% and ABC cases 57% (log rank p=0.015) (Figure 2B). Collectively, these data show that ABC and GCB DLBCLs feature specific and distinct DNA methylation profiles, suggesting that aberrant DNA methylation contributes to the biology of these disease subtypes.

**Differentially methylated genes between ABC and GCB DLBCLs involve specific biological pathways.**

We next determined whether these differentially methylated genes were associated with specific biological functions. The most enriched gene ontogeny (GO) terms included regulation of protein metabolic processes, adaptive immune response, and metallopeptidase activity (Supplemental Table 5). The top scoring KEGG pathway was antigen processing and presentation (Supplemental Table 6). Significantly enriched Biocarta pathways were cytokine and inflammatory signaling, which remained statistically significant after correction for multiple testing (Supplemental Table 6).

As a more functional approach we examined enrichment within publicly available gene sets defined by manipulating specific transcription factors using RNAi technology, dominant-negative inhibition or drug treatments and published gene profiling studies \(^5, 23, 26\) (Supplemental Table 6). The most significantly enriched pathways included NFKB pathway as defined after treatment with an IKK beta inhibitor in the K1106 PMBL cell line \(^27\) and GCB genes differentially expressed between ABC and GCB DLBCLs measured by LymphoChip. \(^28\) Although enrichment of some of these pathways was not significant when corrected for multiple testing, their representation among methylated genes is suggestive of their potential biological impact. This analysis also confirmed involvement of pathways previously defined by gene expression analysis as differentially regulated in ABC and GCB lymphomas underscoring the likely relevance of DNA methylation in gene regulation in these DLBCL subtypes.
Integration of DNA methylation with gene expression profiling was shown to be more informative than either platform alone in capturing differentially involved gene networks between tumor subtypes. The 263 differentially methylated genes at p<0.001 with >20% methylation difference were analyzed together with the 622 genes differentially expressed between the ABC and GCB subtypes at p<0.001 (t-test). All of the genes were categorized based on whether their relative methylation and expression levels were greater in ABC DLBCLs than GCB DLBCLs. Using Ingenuity Pathway Analysis with this combined set of differentially methylated and/or expressed genes we identified the heavy involvement of one particular gene network centered around TNFα, which was hypermethylated in ABC DLBCLs, along with the TAP1, SOX9, PSMB9, IL21, SQLE, LTA and MUC1 genes (Figure 3). Downstream of TNFα several other genes were also relatively repressed in ABCs (including MAP2K4, IRS1, ALDH2, ZNF318, DUSP10); while the subsequent more peripheral genes were more highly expressed (IRF4, IL10, IL6, CCND2, CCL18, OAS1). These data are consistent with the KEGG and BioCarta pathway analysis in being enriched for inflammatory and cytokine pathways and collectively implicating these cytokine networks in playing central and distinct roles in GCB and ABC DLBCL.

In order to determine whether differentially methylated genes shared any specific DNA motifs we used the FIRE program to study sequences up to 1kb upstream of the reported transcription start sites. This analysis identified that a motif resembling the Sp1 transcription factor binding site is significantly enriched in the genes that were relatively highly methylated in ABC as compared to GCB and the control genes (all the rest of the genes). Sp1 was detected in 67% of the methylated sequences in ABC, while it was found in 40% of the methylated sequences in GCB and 31% of the control sequences (Supplemental Figure 3).

Sixteen genes overlap and are inversely correlated between methylation and gene expression signatures.

To identify a core set of functionally relevant differentially methylated genes we explored the overlap of the differentially regulated gene sets captured by gene expression (411 genes) and DNA methylation (239 genes) and represented on both arrays. A set of 16 genes was present in both signatures. While this number is relatively low, it is still much greater than expected by chance (p=0.005 by Fisher’s Exact test; Figure 4A) given the number of genes present in both platforms. All of these genes displayed inverse correlation between DNA methylation and gene expression, suggesting that their DNA methylation is functionally significant. The 16 differentially methylated genes predicted GCB vs. ABC labels with 92% accuracy using Bayesian predictor
with probability greater than 0.8 (as expected, since the gene set was derived from the same patients, Figure 4B). Moreover expression of the same 16 genes also had 98% accuracy in predicting ABC and GCB labels in an independent cohort of 203 DLBCL from Lenz et al.\textsuperscript{11} (Figure 4C). Included among these 16 genes were \textit{IKZF1}, which is a tumor suppressor in B-ALL\textsuperscript{30}, and also \textit{IL-12}, \textit{JDP2}, \textit{PAK1} and \textit{PMM2}. To further confirm differential methylation of these genes we performed single locus quantitative methylation sequencing assays (MassARRAY EpiTYPER) in a set of five randomly selected GCB and five ABC patients. These assays covered most CpGs at each locus up to 3kb upstream of the transcriptional start site, providing a more inclusive and extensive coverage than HELP (Figure 5A and supplemental Figure 4). In 13 of the 16 genes, Epityping confirmed significant differences in methylation between ABC and GCB DLBCLs observed in HELP (p<0.005 for most genes except for one with p<0.1). For three genes (\textit{JDP2}, \textit{SORL1}, \textit{ARHGAP17}) we failed to detect marked difference in overall promoter methylation (Supplemental Figure 4C). Further comparison of the methylation profiles revealed that all three genes have more focal differences (only selected CpGs, p<0.01) in methylation between ABC and GCB groups. We also validated differential expression of a subset of the 16 genes (IKZF1, GALNS and PMM2) in five randomly selected cases of ABC and five GCB primary DLBCLs using QPCR (patients for Epityping validation and Q-PCR validation were not matched due to the lack of corresponding sample material). The Q-PCR showed a similar trend of differential expression in ABC and GCB to that predicted by the expression arrays, but not statistically significant likely due to the limited sample size (Figure 5B). This core methylation signature between ABC and GCB DLBCLs thus contains genes potentially contributing to biological differences between two subtypes of lymphoma and which may serve as potential biomarkers.
Discussion

It is increasingly clear that aberrant epigenetic regulation of gene expression is a hallmark of cancer. As such we reasoned that examination of DNA methylation profiles could help to understand the unique biological properties of ABC and GCB DLBCLs. Since genetic lesions do not fully explain the differences between these DLBCL subtypes, it is reasonable to postulate that epigenetic programming might also contribute to the phenotype of these tumors. Accordingly, by analyzing the DNA methylation status of over 50,000 CpGs distributed among 14,000 gene promoters we identified 263 genes (311 probesets) that are differentially methylated among ABC and GCB DLBCLs. Previous studies of DNA methylation in DLBCL either failed to find differentially methylated genes between ABC and GCB DLBCLs or identified only a small set of genes. The most likely reason for this was the methylation platforms interrogating smaller numbers of genes and CpGs. In our study the major gene pathways affected by differential methylation between GCB and ABC DLBCLs were cytokine signaling, GC B-cell and NFkB regulated genes, which in fact are consistent with the known biological differences between the GCB and ABC type DLBCLs. ABC DLBCLs are known to harbor mutations in genes within the NFkB pathway and to display prominent STAT3 signaling. These data confirm the relevance of these signaling pathways to DLBCL pathophysiology and show that DNA methylation may contribute together with genetic lesions to their deregulation in ABC vs. GCB lymphomas.

Although the GCB lymphomas are often thought of as arising from germinal center B cells and the ABC lymphomas from B-cells of late germinal center cells or plasmablasts, the cells of origin of these tumors have not been strictly defined functionally. Emerging data suggest that at least some of the gene expression differences of these tumors could have a root in the distinct biology of the respective normal cell of origin which might reflect the differential ability of genetic or epigenetic alterations to contribute to lymphomagenesis. We hypothesize that both genetic and epigenetic factors contribute to establishing ABC and GCB signatures. It is increasingly clear that DNA methylation patterns vary among tissues and with differentiation, and in other studies we have found that naïve B-cells display differences in DNA methylation compared to germinal center centroblasts. As future studies more clearly delineate the cell of origin for these DLBCLs it will be interesting to evaluate whether the stage-specific DNA methylation distribution of B-cell precursors contributes to the epigenetic differences between ABC and GCB DLBCLs.
An alternative hypothesis explaining aberrant DNA methylation in DLBCLs is based on the observation reported by Martin-Subero et al suggesting that de novo methylated genes in lymphomas are enriched for Polycomb targets and reflect acquisition of “stemness” during lymphomagenesis, which may represent the early event in neoplastic transformation of B cells. The same group demonstrated that different forms of non-Hodgkin lymphomas have distinct epigenetic profiles. Along these lines, it is possible that alterations of certain transcriptional or epigenetic regulatory proteins might induce specific epigenetic patterns in lymphoma. For example, Morin et al discovered frequent point mutations in the Polycomb protein and histone methyltransferase EZH2, more frequently occurring in GCB type DLBCLs. Given that the interplay between DNA methylation and histone modifications is possible such lesions could induce specific alternations in DNA methylation.

Another possible mechanistic clue can be derived from our finding that an Sp1 motif is enriched in differentially hypermethylated genes in ABC DLBCLs. It is possible that levels of Sp1 determine the occupancy of the binding motifs and thus the degree to which those are protected from methylation. Since Sp1 binding may be affected by methylation status of the cytosines in its binding motif (5’-GGGCGG) aberrant methylation of Sp1 sites in ABC DLBCLs may lead to dysregulation in Sp1 binding and affected transcriptional regulation of its downstream targets. Alternatively, lower levels of Sp1 binding in ABC precursor cells might facilitate aberrant methylation of these sites.

We find that integration of the differentially expressed and differentially methylated gene signatures between ABC and GCB patients captured a specific gene network centered around TNFα, suggesting that this network could play a significant role in explaining the biological difference between these tumors. The genes involved include TNFα, SOX9, MUC1, IL21, IL4, IL10, LTA etc and reflect a possible functional attenuation of TNFα pathway in ABC, as compared to GCB DLBCLs. While the DNA methylation status of these genes was inversely correlated with gene expression, they did not reach our cutoff for statistical significance and so were not captured as part of the sixteen gene overlap signature. Nonetheless it is reasonable to propose that this gene network be further explored since gain or loss of function of components of this pathway have been linked to lymphomagenesis. TNFα has autocrine and paracrine effects and can mediate activation of NF-kB pathway with anti-apoptotic effects, activation of MAPK pathway with proliferation and differentiation effects, and induction of death.
signaling with activation of caspase-8 and induced apoptosis. \(^42\) Stimulation of TNF\(\alpha\) occurs during inflammatory disease, and a link between infection, chronic antigenic stimulation and lymphomagenesis has been reported as relevant to DLBCL pathogenesis. De Vita et al reported expression of TNF\(\alpha\) in premalignant lymphoproliferative lesions in the setting of Sjogren's disease and during progression to B-cell non-Hodgkin Lymphoma. \(^38\) Elevated secretion of TNF\(\alpha\) in combination with H. pylori –specific TH-1 response predisposes to a more adverse outcome of gastritis including peptic ulcer and higher risk of MALT-derived lymphomas. \(^43\) The InterLymph consortium reported that polymorphisms of TNF\(\alpha\), LTA and IL10 are associated with a higher risk of developing DLBCL. \(^44, 45\) Elevated serum levels of TNF\(\alpha\) and IL10 were also associated with inferior survival in DLBCL patients, although no distinction was made regarding whether these were ABC or GCB tumors. \(^46, 47\) On the other hand, a link between treatment with TNF\(\alpha\) suppressing drugs used for chronic inflammatory diseases and lymphomagenesis has been reported. More detailed mechanistic studies of this pathway are clearly warranted, specifically in the context of GCB vs. ABC DLBCLs. However, one potential therapeutic implication of the involvement of this network relates to use of steroids, which are included in most DLBCL regimens, and which function in part through a TNF\(\alpha\) related pathway. Perturbation of this network might influence the therapeutic actions of steroids.

Using stringent criteria we identified a minimal set of sixteen genes that were significantly differentially methylated and inversely correlated with expression between ABC and GCB DLBCLs and which readily distinguished ABC from GCB cases. This set of genes includes known tumor suppressor (such as \(IKZF1\) \(^48\)) or pro-oncogenic (such as \(SOX9\) and \(PAK1\)) functions.\(^49, 50\) Further functional study of these genes is warranted to define whether they contribute to lymphomagenesis. This gene set might also serve as a potential clinically useful biomarker, as suggested by the fact that expression of these genes was 98% accurate in distinguishing ABC and GCB patients in an independent cohort. Given the stability of DNA methylation in clinical specimens and the relatively technical ease with which DNA methylation abundance can be determined, it is possible that methylation biomarkers such as these could eventually be used either alone or in combination with other methods to distinguish ABC from GCB cases in the clinical setting. Prospective validation studies will help to define the utility of this approach.
Acknowledgements

RS is supported by NCI K08 CA127353. This research was supported by LLS SCOR grant 7017-09 to AM and AM is a Leukemia and Lymphoma Society Scholar. RDG receives support through the National Cancer Institute of Canada (NCIC), Terry Fox Foundation Program Project grant #019001. NAJ was supported by the Terry Fox Foundation (019005), the Michael Smith Foundation for Health Research (ST-PDF-01793) and the Canadian Institute of Health Research (STP-53912).
Authorship

Contribution: R.S. performed HELP profiling, analyzed methylation data and Q-PCR data, wrote the manuscript; H.G. performed gene expression analysis, statistical and integrative analysis, wrote the manuscript; L.T. performed Q-PCR, DNA extraction; L.C. performed gene expression analysis, revised manuscript; N.J. collected patient samples and performed gene expression analysis; R.G., O.E. and A.M. designed and supervised the research.

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

Correspondence: Ari Melnick, Division of Hematology/Oncology, Department of Medicine Weill Cornell Medical College, 1300 York Ave, New York, NY 10065, e-mail: amm2014@med.cornell.edu; Olivier Elemento, Institute for Computational Medicine, Department of Physiology and Biophysics, Weill Cornell Medical College, 1305 York Avenue, New York, NY, 10021, email: ole2001@med.cornell.edu; Randy D Gascoyne, Research Director, Center for Lymphoid Cancers, BC Cancer Agency & BC Cancer Research Center, 675 W 10th Avenue, Vancouver, BC V5Z 1L3, email: rgascoyn@bccancer.bcca.
References


Figure Legends

Figure 1. ABC/GCB labels of our 69 DLBCLs based on the gene expression signature. (A) Heatmap reveals assignment of labels to our 69 cases of DLBCLs on the basis of the gene expression signature using Bayesian predictor with cutoff value of 0.9. (B) Kaplan-Meier estimates of overall survival and progression free survival based on case labels assigned on the basis of the gene expression signature show that patients with GCB DLBCL have higher probability of overall survival (left) and progression free survival (right) than patients with ABC DLBCL.

Figure 2. ABC and GCB subgroups of DLBCLs have distinct methylation signatures. (A) 69 cases of DLBCLs were studied using HG_17 human promoter array from Roche NimbleGen. The heatmap illustrates methylation values of 311 probe sets in 69 cases. Using leave one out cross validation method we demonstrated that 311-probe signature is able to correctly assign ABC and GCB labels to the DLBCL cases using Bayesian predictor with probability cutoff of 0.8 (plot above the heatmap) and predictive accuracy of 91%. (B). Kaplan-Meier estimates of progression free survival in ABC and GCB subgroups, classified on the basis of 311 methylation signature, demonstrate that patients with GCB DLBCL had a higher probability of progression free survival than patients with ABC DLBCL.

Figure 3. Tumor necrosis factor \(\alpha\) network is key epigenetically disregulated network between ABC and GCB DLBCLs. Integration of 263 differentially methylated genes with 622 differentially expressed genes allowed identification of one main differentially regulated network using Ingenuity Pathway Analysis software: Tumor Necrosis Factor \(\alpha\) Pathway. Relative expression and methylation is represented by the ratio of ABC/GCB: overmethylated genes have greater methylation in ABC than GCB DLBCLs, and overexpressed genes have greater expression in ABC than GCB DLBCLs.

Figure 4. Methylation and/or expression level of 16 genes can accurately differentiate between ABC and GCB DLBCLs. (A) 16 genes represent an overlap between differentially methylated (p<0.001) and most differentially expressed (p<0.001) genes between ABC and GCB DLBCLs. This set of genes represents genes potentially regulated by methylation with inverse methylation and expression levels (the inverse correlation is not random with p<0.005 in Fisher’s Exact test). (B-C) The mean log-ratio levels in ABC vs GCB DLBCLs of methylation
and expression values of the 16 signature genes are depicted according to color scale in the heatmaps. The prediction probability graphs (on top of heatmaps) reveal that 16-gene methylation level can assign ABC and GCB labels with 92% accuracy using Bayesian predictor at probability cutoff of 0.8 (panel B); while gene expression level can assign ABC and GCB labels with 98% accuracy using Bayesian predictor at probability cutoff of 0.8 using an independent set of 203 DLBCL cases from Lenz et al (panel C).

**Figure 5.** Single genes of 16-gene methylation signature can differentiate between ABC and GCB subtypes of primary DLBCLs across platforms. (A) Heatmaps represent EpiTYPER results for three of the 16 genes: IKZF1, GALNS and PMM2 (the rest of 16 genes see in supplemental Figure 4), performed in five randomly selected ABC and five GCB primary DLBCL cases. Rows of the heatmap represent individual CpGs in the promoter regions with color reflecting methylation value; while columns represent individual cases with class label on the bottom. T-test using methylation values from all tested CpGs was performed between ABC and GCB subtypes, with p value represented below each heatmap. Panels in the middle show the methylation levels for each CpG averaged in ABC and GCB cases. Panels on the right show the average methylation level of all CpGs in ABC and GCB cases with error bar for standard deviation. (B) Q-PCR was performed in five ABC and five GCB DLBCLs with primers specific for *IKZF1, GALNS* and *PMM2*. The amount of transcript was calculated by normalizing to internal control gene RPIL3 and is shown as an average within the subtype. It reveals the trend for greater expression in ABC than GCB DLBCLs, which is inversely correlated with greater DNA methylation of the corresponding promoters.
Figure 2

A

B

Log rank p < 0.015
Figure 4

(A) Methylation and expression scale log ratio ABC vs GCB

- IKZF1: IKAROS family zinc finger 1 (Ikaros)
- ASPHD2: aspartate beta-hydroxylase domain containing 2
- PMM2: phosphomannomutase 2
- PAK1: p21/Cdc42/Rac1-activated kinase 1 (STE20 homolog, yeast)
- LYAR: Ly1 antibody reactive homolog (mouse)
- JDP2: Jun dimerization protein 2
- FGD2: FYVE, RhoGEF and PH domain containing 2
- GALNS: galactosamine (N-acetyl)-6-sulfate sulfatase (Morquio syndrome, mucopolysaccharidosis type IVA)
- IL12A: interleukin 12A (natural killer cell stimulatory factor 1, cytotoxic lymphocyte maturation factor 1, p35)
- ARHGAP17: Rho GTPase activating protein 17
- SORL1: sortilin-related receptor, L(DLR class) A repeats-containing
- KIAA0746: KIAA0746 protein
- LANCL1: LanC lantibiotic synthetase component C-like 1 (bacterial)
- KCNK12: potassium channel, subfamily K, member 12
- SOX9: SRY (sex determining region Y)-box 9 (camptomelic dysplasia, autosomal sex-reversal)
- CXorf57: chromosome X open reading frame 57

(B) ABC and GCB probability

(C) ABC and GCB probability

Methylation scale log ratio ABC vs GCB

Expression scale log ratio ABC vs GCB
Figure 5

A

IKZF1

CpG

GCB  ABC  GCB  ABC

p=0.0004

Average Methylation

IKZF1 Normalized to RPL3

GCB  ABC

B

GALNS

CpG

GCB  ABC  GCB  ABC

p=2e-5

Average Methylation

GALNS Normalized to RPL3

GCB  ABC

PMM2

CpG

GCB  ABC  GCB  ABC

p=0.04

Average Methylation

PMM2 Normalized to RPL3

GCB  ABC

Percentage of methylation
DNA methylation signatures define molecular subtypes of diffuse large B cell lymphoma

Rita Shaknovich, Huimin Geng, Nathalie A. Johnson, Lucas Tsikitas, Leandro Cerchietti, John M. Greally, Randy D. Gascoyne, Olivier Elemento and Ari Melnick