Reproducible, Quantitative, and Flexible Molecular Subtyping of Clinical DLBCL Samples Using the NanoString nCounter System

Margaret H. Veldman-Jones1, Zhongwu Lai2, Mark Wappett1, Chris G. Harbron3, J. Carl Barrett2, Elizabeth A. Harrington1, and Kenneth S. Thress2

Abstract

Purpose: Diffuse large B-cell lymphoma (DLBCL) is a heterogeneous disease with distinct molecular subtypes. The most established subtyping approach, the "Cell of Origin" (COO) algorithm, categorizes DLBCL into activated B-cell (ABC) and germinal center B-cell (GCB)-like subgroups through gene expression profiling. Recently developed immunohistochemical (IHC) techniques and other established methodologies can deliver discordant results and have various technical limitations. We evaluated the NanoString nCounter System to address issues with current platforms.

Experimental Design: We devised a scoring system using 145 genes from published datasets to categorize DLBCL samples. After cell line validation, clinical tissue segmentation was tested using commercially available diagnostic DLBCL samples. Finally, we profiled biopsies from patients with relapsed/refractory DLBCL enrolled in the fostamatinib phase IIb clinical trial using three independent RNA expression platforms: NanoString, Affymetrix, and qNPA.

Results: Diagnostic samples showed a typical spread of subtypes with consistent gene expression profiles across matched fresh, frozen, and formalin-fixed paraffin-embedded tissues. Results from biopsy samples across platforms were remarkably consistent, in contrast to published IHC data. Interestingly, COO segmentation of longitudinal fostamatinib biopsies taken at initial diagnosis and then again at primary relapse showed 88% concordance (15/17), suggesting that COO designation remains stable over the course of disease progression.

Conclusions: DLBCL segmentation of patient tumor samples is possible using a number of expression platforms. However, we found that NanoString offers the most flexibility and fewest limitations in regards to robust clinical tissue subtype characterization. These subtype distinctions should help guide disease prognosis and treatment options within DLBCL clinical practice.

Introduction

Diffuse large B-cell lymphoma (DLBCL) is the most common type of lymphoma, accounting for nearly 40% of newly diagnosed cases (1). Although approximately 50% of DLBCLs are curable through standard treatment, there is an urgent need for new therapies as most relapsed patients will eventually die from their disease. DLBCL has been long recognized as a heterogeneous disease with diverse genetic features and variable clinical outcomes. Gene expression profiling (GEP) techniques have been used for over a decade to classify DLBCL into distinct molecular subtypes, some of which carry significant prognostic value (2–5). The most well-established expression signature is the "Cell of Origin" (COO) algorithm, which divides DLBCL into activated B-cell (ABC) and germinal center B-cell (GCB)-like subgroups. The COO signature can be refined to 12 distinct genes capable of accurately subtyping DLBCL with little loss of specificity or sensitivity (4). The continued refinement and potential prognostic value of COO profiling has led to its incorporation into patient selection strategies for emerging targeted therapeutics. Both ibrutinib (Pharmacyics) and bortezomib (Millennium) have applied retrospective ABC/GCB patient profiling to phase I and II clinical trials (6, 7).

COO (and other) expression profiling approaches will undoubtedly lead to continued investment in personalized medicine opportunities in DLBCL (8). An obstacle to more widespread clinical profiling has been the limited availability of snap frozen DLBCL biopsies that yield high quality tumor RNA. Standard clinical practice usually results in collection of formalin-fixed, paraffin-embedded (FFPE) diagnostic tissue that yields fragmented RNA. Such RNA is not the preferred input type for microarray-based approaches (i.e., Affymetrix) historically used for COO profiling. Because of this limitation, alternatives to classic microarrays have emerged to enable COO profiling of diagnostic FFPE specimens. Numerous immunohistochemical (IHC) classifiers have been developed over the last decade (9). The original IHC algorithm is the "Hans" approach based on expression of three protein markers that have been widely adopted within segments of the clinical DLBCL community (10). Being an immunohistochemistry-based assay, the Hans methodology is subject to...
Translational Relevance

Diffuse large B-cell lymphoma (DLBCL) is a heterogeneous disease that can be classified into two major molecular subtypes: activated B-cell–like (ABC) and germinal center B-cell–like (GCB) DLBCL. Clinical trials of targeted therapies are now underway using ABC/GCB segmentation as a patient stratification approach. Both gene expression profiling (GEP) and immunohistochemistry (IHC) have historically been used to segment ABC from GCB-type DLBCL. Although the accurate classification of patients is essential for data interpretation, commonly used GEP and immunohistochemical methodologies suffer from several limitations. We developed a DLBCL assay on the NanoString gene expression system that accurately and reproducibly categorized multiple DLBCL sample types, including both fresh and formalin-fixed patient tumor tissue. After comparing against IHC and several other standard gene expression methodologies, the NanoString offered several key advantages, including sensitivity, flexibility, and clinical applicability. The NanoString platform should be strongly considered for both DLBCL research and patient management activities.

antibody limitations, pathologist subjectivity, and is relatively nonquantitative in its scoring. Not surprisingly, Hans IHC and Affymetrix GEP have been shown to lack full concordance when tested concurrently on a blinded set of biopsies (11). More recently, the quantitative nuclease protection assay (qNPA; HTG Molecular Diagnostics) has emerged as another technology for DLBCL segmentation as it can measure gene expression from FFPE material in a fully quantitative manner (12, 13). While qNPA has some distinct advantages, the current platform is limited in its ability for multiplex analysis of a large number of genes. While NanoString can monitor thousands of genes, qNPA monitors less than 50 in a single run.

The desire to develop an approach that provides a flexible, robust, and fully quantitative method for DLBCL segmentation led us to evaluate the recently described NanoString system for gene expression (14–16). NanoString technology uses digital, color-coded barcodes (codessets) that are attached to sequence specific probes, allowing for fully quantitative and direct measurement of mRNA without amplification. The nonenzymatic nature of NanoString allows for accurate and reproducible quantification of as little as 100 ng of input mRNA from FFPE clinical samples. A key differentiating feature of the NanoString system from IHC and qNPA-based approaches is its ability for multiplex analyses of up to 800 distinct targets. To evaluate the NanoString system for DLBCL subtyping, we designed a custom codeset of more than 300 genes, consisting of 145 COO signature genes and additional genes from alternative DLBCL segmentation signatures. We tested these codesets on the NanoString against DLBCL cell lines with known COO designation and then against a set of commercially procured de novo DLBCL patient samples. Finally, we applied the NanoString codeset to biopsies from an ongoing 60 patient clinical trial evaluating the effectiveness of fostamatinib (Syk inhibitor) in relapsed-refractory DLBCL (17). Through the course of the evaluation, we directly tested many samples against other relevant platforms, including Affymetrix microarray and qNPA, to allow for both relative expression data and COO designation comparisons.

Materials and Methods

Cell line and tissue samples

DLBCL cell lines were purchased from Cambridge Enterprise (DSMZ) in 2010 unless otherwise noted. HBL-1 (provided by Daniel Krappmann, German Research Center for Environmental Health), SU-DHL-10, Karpas 422, RC-K8, and SU-DHL-4 were cultured in RPMI with 10% to 15% FBS and 1% l-glutamine. TMD8 (provided by Shuji Tohda, Tokyo Medical and Dental University) and OCI-Ly19 were cultured in α-Eagle’s minimum essential medium with 10% FBS and 1% l-glutamine. OCI-Ly4 and OCI-Ly10 (both provided by Mark Minden, Ontario Cancer Institute) were cultured in Iscove’s modified Dulbecco’s medium with 20% FBS, 1% l-glutamine, and 50 μmol/l β-mercaptoethanol. Each vial of cells was used for no more than six passages. Cells were maintained in 5% CO₂ at 37°C. All cell lines were tested for authenticity by genotyping in 2010 before use; only those cell lines confirmed to carry the genetic profiles as described previously were used in data presentation.

Commercially available DLBCL tissue samples (RNA; designated as fresh, frozen, and FFPE) were purchased from OriGene. All required consents for these exploratory analyses were acquired. Before processing, each sample was reviewed by an internal certified pathologist to confirm disease diagnosis and verify tumor content.

Fostamatinib phase Ib clinical trial (NCT01499303) pretreatment frozen core needle biopsies and archival FFPE tissue sections were shipped from recruitment centers with required consent and handled according to AstraZeneca Human Biological Samples policies and procedures. Before analysis, an hematoxylin and eosin–stained section from each sample was assessed by a certified pathologist to confirm disease diagnoses and verify tumor content. A titration study between two frozen DLBCL biopsies (one of each subtype) and benign lymphoid tissue showed a linear relationship between reduction in tumor content and the effect on the gene signature score (data not shown). These results support an arbitrary minimum cutoff of 70% tumor, with samples deemed unevaluable if containing less than 70% tumor tissue. All clinical samples used throughout this study were determined to contain at least 70% DLBCL.

Sample preparation and RNA extraction

Under material transfer agreement, total RNA from all nine DLBCL cell lines was isolated using standard procedures and provided by colleagues at AstraZeneca Gatehouse Park. Replicate batches of total RNA were isolated from six DLBCL cell lines (HBL-1, TMD8, SU-DHL-4, SU-DHL-10, OCI-Ly19, and Karpas 422) using the RNeasy Mini Kit (Qiagen). Frozen tissues obtained from recruitment centers with required consent and handled according to AstraZeneca Human Biological Samples policies and procedures. Before analysis, an hematoxylin and eosin–stained section from each sample was assessed by a certified pathologist to confirm disease diagnoses and verify tumor content.

A titration study between two frozen DLBCL biopsies (one of each subtype) and benign lymphoid tissue showed a linear relationship between reduction in tumor content and the effect on the gene signature score (data not shown). These results support an arbitrary minimum cutoff of 70% tumor, with samples deemed unevaluable if containing less than 70% tumor tissue. All clinical samples used throughout this study were determined to contain at least 70% DLBCL.
NanoString codeset design and expression quantification
COO and consensus clustering (CC) signature and housekeeping genes were included in the codeset based on previously described publications (Supplementary Table S1). Input total RNA amount was determined by titration studies to show comparable and reliable gene expression data at 100 ng for high quality (cell lines, fresh, and frozen tissue) and at 400 ng for low quality (FFPE tissue) samples based on detection levels, linearity of genes, and binding density (data not shown). Data were normalized through an internally developed Pipeline Pilot Tool (publicly available for use on the Comprehensive R Archive Network, CRAN). In brief, data were log2 transformed after being normalized in two steps: raw NanoString counts were first background adjusted with a Truncated Poisson correction using negative control spikes followed by a technical normalization using positive control spikes. Data were then corrected for input amount variation through a Sigmoid shrunken slope normalization step using the GEO mean expression of housekeeping genes. A transcript was designated as not expressed if the raw count was below the average of the internal negative control raw counts plus two SDs.

Microarray (Affymetrix) quantification
Total RNA was extracted from fresh and frozen samples using a standard protocol. At least 250 ng of total RNA from each sample with OD 260/280 ratio between 1.68 and 2.08 and concentration >50 ng/µL was submitted to Almac Diagnostics. Gene expression profiling was conducted using Affymetrix U133 plus 2.0 chip according to manufacturer recommendations. The CEL files were analyzed using Bioconductor’s Affy package in R. Expression was normalized using MAS5.0 method with scaling factor set to 100. The signals were then log2 transformed before downstream analysis. When a gene had more than one probeset, the probeset with the highest mean signal was selected to represent the gene.

qNPA
Two 5-µm FFPE sections per patient were delivered to HTG Molecular Diagnostics for processing and qNPA analysis. All samples were run against HTG’s 12-gene COO signature array consisting of the following genes: CD10, LRMP, CCND2, ITPKB, PIM1, IL16, IRF4, FUT8, BCL6, LMO2, CD39, and MYBL1. All processing and data analyses were conducted as described (13).

Quantitative real-time RT-PCR
Gene expression assays for 12 genes [FUT8, IL16, IRF4, CCND2, PIM1, CD39 (ENTPD1), ITPKB, LM02, LRMP, CD10 (MME), MYBL1, and BCL6] were ordered from Applied Biosystems (Supplementary Fig. S1A). For normalization purposes, IPO8 was selected from a screen of 16 housekeeping genes based on its robust stability and low SD across a panel of six DLBCL cell lines (Supplementary Fig. S1B and S1C). Reverse transcription of 100 ng RNA was performed using the Superscript Vilo Kit (Invitrogen) and quantitative real-time PCR amplification of cDNA was performed on the 7900HT TaqMan (Applied Biosystems) in 10 µL reactions containing TaqMan Gene Expression master mix and assays (Applied Biosystems). Samples were amplified with three experimental replicates. No template controls were reliably negative.

DLBCL COO subtype classification using signature scores
We calculated a composite score of our COO signature, using a method previously described (18). The unweighted average score was calculated from gene expressions within a cohort after housekeeping gene normalization and log2 transformation. Signature scores were then calculated as the mean expression of genes associated with ABC subtype minus the mean expression of genes associated with GCB subtype; a higher score thus indicating a more ABC-like sample.

We then assessed the ability of our COO signature scores to predict the COO subtypes in 32 DLBCL cell lines and publicly available datasets of DLBCL patient expression profiles. High signature score cell lines were almost exclusively ABC, with TOLEDO the only exception (Supplementary Fig. S2A). We further applied the method to two datasets (GSE10846 and GSE4732, available in GEO), and found them to be well correlated with ABC and GCB subtypes assigned by study authors (Supplementary Fig. S2B and S2C). We also found unclassified DLBCL to be mainly concentrated in the middle. We thus decided to assign an empirical cut off for samples to ABC if the signature scores were above 0.7, to GCB if below 0, and unclassified if between 0 and 0.7.

Results
Definition of NanoString and RT-PCR gene lists for DLBCL segmentation
A comprehensive list of 307 genes was compiled from two literature-established methodologies: COO and CC (Fig. 1). For COO segmentation, multiple predictors have been described in the literature (4, 13, 19). By combining unique predictors from these studies, a list of 51 distinct genes was obtained (Fig. 1A). From four publicly available gene expression profiles of patients with DLBCL with ABC and GCB annotations [Lymphoma (3); Lymphoma 2-GSE4732 (20); Lymphoma-GSE4475 (21); Lymphoma-GSE10874 (5)] in Oncomine (22), we selected 126 genes that were differentially expressed in at least two of the four studies (Fig. 1B). Combined with the 51 predictor genes, we derived 145 unique genes that were associated with either ABC or GCB subtypes. For CC classification, we relied on the original publication (23), which used 150 probes, representing 133 unique genes (Fig. 1C). To correct for batch effects and normalize samples analyzed over time in several different test sets, we included in the codeset 33 housekeeping genes whose expressions were moderate to high and showed little variance across datasets (Fig. 1D). The final NanoString gene list (codeset) contained 307 unique genes (Supplementary Table S1). To compare NanoString with an established quantitative technique, real-time PCR (RT-PCR) primer-probe sets were created for a 12 gene subset of the most refined version of the COO signature, as previously described (4, 13) (Supplementary Table S1 and Fig. S1B).

Validation of NanoString codeset for disease segmentation in DLBCL cell lines and clinical samples
We examined nine DLBCL cell lines with literature-established COO designations (24, 25). To confirm that inhouse cell lines were representative of those in the literature, we used a previously described (18) RT-PCR method and applied a COO signature score (detailed in Materials and Methods); a high scoring sample (>0.7) signified an ABC subtype, low scoring samples (<0) a GCB subtype, and those in between (0 to <0.7) as unclassified. Reverse transcription and RT-PCR was performed using the refined 12-gene COO subset on eight DLBCL cell lines and ΔCt values (mean of triplicate) were scored and ranked. The results confirmed the
between replicates was only slightly inferior (Fig. 2A and B). In addition, agreement was found to be extremely tight, with biologic reproducibility rather than human plasma as previously described (26) could compromise the gene expression signatures generated from NanoString. We then assessed the reproducibility of the NanoString platform.

Replicates of all nine DLBCL cell lines were used to investigate technical reproducibility (intra- and inter-assay) and correlations prepared from fresh DLCBL biopsies (fresh), and 24 flash-frozen excisional DLBCL biopsies (frozen; Fig. 3A). Importantly, these three sample types were almost entirely paired and patient-matched, enabling cross-matrix comparisons. First, we profiled the 36 fresh RNA samples and generated COO signature scores on the NanoString platform. These scores showed the expected subtype distribution of a pretherapy (diagnostic) DLBCL population with 47% GCB, 25% ABC, and 28% unclassified (Fig. 3B). Together with the cell line dataset, this gave confidence in the ability of the NanoString platform to identify and classify DLBCL subgroups using fresh RNA. Patient biopsy material is most often collected as FFPE tissue at clinical sites so we proceeded to profile RNA extracted from FFPE tissue of 13 patients alongside matched fresh RNA from the same patients (Fig. 3C) and five matched frozen biopsies (Supplementary Fig. S3A). The quantitative correlation was shown to be very robust, establishing that NanoString could generate high-quality data on FFPE material to allow for clinically relevant DLBCL segmentation. The same sample set was also evaluated using RT-PCR (12 COO signature genes) with strong correlations observed between the two platforms further validating the NanoString results (data not shown). These data demonstrate that tissue preparation and processing does not compromise the gene expression signatures generated from NanoString.

Correlation between NanoString quantification and alternative platforms using OriGene samples

There are several existing and emerging methodologies for COO classification of clinical DLBCL tissue (Fig. 4A). After validating the NanoString platform as another viable approach, we compared NanoString outputs to other established technologies using the OriGene clinical DLBCL matched sample set. The Hans IHC algorithm, the most widely used clinical COO segmentation tool, is based on three antibodies (CD10, Bcl-6,
Mum1) and uses FFPE diagnostic tissue (10). We profiled the FFPE samples from 10 different patients through both the Hans IHC (Phenopath Inc.) and NanoString platforms and observed a 90% concordance rate of finalized COO signature calls (Supplementary Fig. S3B).

Although not often employed for clinical segmentation, Affymetrix GEP is the most well-established methodology for COO subtyping and is an important comparator for NanoString evaluation (28). Thirty-four fresh RNA samples were profiled concurrently across NanoString and Affymetrix using the full NanoString DLBCL codeset and the U133 plus 2.0 array (Affymetrix). In contrast to Hans IHC, both Affymetrix and NanoString provide fully quantitative outputs, thus allowing for more robust analyses. The COO score correlations derived from the same 145 COO genes were impressive across the two platforms (Fig. 4B), as were the gene expression correlations and COO designations (Supplementary Fig. S3A and S3C).

A qNPA from HTG Molecular Diagnostics is another COO segmentation methodology that has been described in the literature (12). Because qNPA has already been used to segment DLBCL patients in a clinical trial, we wished to directly compare the two methodologies (13, 29–31). Thirty-one samples (18 fresh, 13 FFPE), representing 26 distinct patients, were profiled through HTG’s COO array consisting of the same 12 COO genes used for the RT-PCR profiling described previously. The quantitative correlations between NanoString and qNPA datasets were very robust and, importantly, tight correlations were observed for both FFPE and fresh RNA samples (Fig. 4C and Supplementary Fig. S3D).

Any misclassifications between platforms can be explained by the use of different classifiers to create COO scores (12 vs. 145...
genes). As expected, the correlation is better between matching and larger classifiers (qNPA vs. RT-PCR, NanoString vs. Affymetrix). In addition, the target RNA sequences used were not identical between platforms, which can lead to different expression patterns that affect the COO designation. Of note, many discordant cases were due to those classifications bordering the cut-off criteria.

Application of NanoString-based DLBCL segmentation to fostamatinib relapsed-refractory DLBCL phase IIb trial samples

The NanoString platform and DLBCL codeset were applied to biopsy samples from an ongoing clinical trial of the Syk inhibitor, fostamatinib. The randomized, double-blind phase IIb study, is a trial of 60 patients designed to evaluate the efficacy of fostamatinib in patients with relapsed or refractory DLBCL (31). Based on emerging preclinical data, an objective of the study was to explore whether DLBCL subtype might predict response to fostamatinib. Flash-frozen core needle tumor biopsies were collected from all relapsed/refractory participants before fostamatinib dosing, as well as the original diagnostic (FFPE) biopsy. Because predose biopsies would best represent the patient’s tumor biology at the start of fostamatinib therapy (following R-CHOP relapse), RNA was first extracted from the evaluable core-needle biopsy samples (n = 59) and subsequently profiled on NanoString to enable COO segmentation. The COO signature scores of these patients with relapsed/refractory DLBCL (Fig. 5A) displayed a distribution similar to that seen from diagnostic (pretherapy) biopsies described in the literature (32). The same RNA from 48 patients that passed QC criteria and profiled across the Affymetrix U133 2.0 gene chip displayed strong quantitative and COO score concordance with the NanoString output (Fig. 5B). Diagnostic tumor material (FFPE slide or block) was available for a subset of the patients on the fostamatinib study providing the opportunity to explore whether COO designation might evolve during the course of R-CHOP therapy in patient-matched tissue. Eighteen patients provided evaluable diagnostic (FFPE) and fresh fostamatinib biopsy material that were run on the qNPA and NanoString platforms, respectively. An 88% concordance (15/17) was observed in the COO calls between the two sample types (Fig. 5C), suggesting that COO designation remains stable from initial
### Methodology Input Vendor

<table>
<thead>
<tr>
<th>Vendor</th>
<th>Input</th>
<th>Methodology</th>
<th>Preferred Sample Type</th>
<th>Markers Genes Examined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affymetrix</td>
<td>mRNA</td>
<td>GEP</td>
<td>FF</td>
<td>10,000+</td>
</tr>
<tr>
<td>NanoString</td>
<td>mRNA</td>
<td>GEP</td>
<td>FF and FFPE</td>
<td>800</td>
</tr>
<tr>
<td>PhenoPath</td>
<td>Protein</td>
<td>IHC</td>
<td>FFPE</td>
<td>3</td>
</tr>
<tr>
<td>HTG</td>
<td>mRNA</td>
<td>qNPA</td>
<td>FF and FFPE</td>
<td>48</td>
</tr>
</tbody>
</table>

#### Figure 4

A. Comparison of platforms used for COO profiling. FF, fresh frozen; GEP, gene expression profiling. B. Correlation plot between NanoString and Affymetrix platforms based on 145-gene COO scores calculated from 34 fresh RNA samples. Pearson correlation is 0.993. C. Correlation plot between NanoString and qNPA platforms based on 12-gene COO scores calculated from 31 OriGene samples. Connected lines indicate samples were derived from the same patient. Pearson correlation is 0.956.
Figure 5
A, waterfall plot of NanoString data illustrating 145-gene COO scores and subtype designations of 59 patients. Arbitrary scoring cut-off was set at >0.7 = ABC, <0.0 = GCB, and 0.0-0.7 = unclassified. B, correlation between NanoString and Affymetrix platforms based on 145-gene COO scores calculated from 48 patients. Pearson correlation is 0.995. C, COO designation is maintained over the course of standard DLBCL therapy. Patients marked in bold show discordant COO designation and patients marked in italics could not be correlated due to sample QC issues. ND, not determined.
diagnosis through primary relapse. In addition to COO profiling, specific 5′ and 3′ NanoString probes were designed to detect the presence of BCL2-IGH t(14:18) gene fusions known to occur in DLBCL (33). Using the OCI-Ly8 cell line with known BCL2-IGH fusion as a positive control, we detected putative BCL2 gene fusion events in 23% of fostamatinib DLBCL samples (Fig. 6; Supplementary Table S2).

Discussion

The heterogeneous nature of DLBCL has prompted many efforts at disease segmentation to inform prognosis or predict efficacy of specific therapies. The original description of the COO algorithm, indicating that patients with GCB showed significantly better overall survival than patients with ABC, led to numerous reports aimed at confirming and extending the original findings (3). In that time, several COO profiling methodologies have emerged that have influenced both research and clinical practice. Use of microarrays is well established at the research level but high quality microarray data require RNA isolated from frozen tissue. Because the majority of clinical DLBCL diagnostic tissues are FFPE samples, microarray-based COO profiling can be challenging. As a way to enable widespread clinical utility, many IHC-based approaches for COO segmentation have been developed and adopted by clinical practitioners. IHC is attractive as a rapid, cost-effective, and accessible platform, commonly available in most clinical centers. While IHC algorithms for COO segmentation have shown reasonable concordance with microarray-based approaches (10, 34, 35), the subjective nature of IHC scoring allows for inherent variability. Indeed, recent data have shown that various iterations of nine related IHC algorithms do not correlate well with one another (9). A robust and reliable methodology for COO profiling, applicable to both research and clinical samples, is required to enable the successful discovery and development of targeted therapies for DLBCL.

In this study, we evaluated the NanoString gene expression platform for the molecular classification of DLBCL specimens. The NanoString system generated high quality, reproducible, and fully quantitative results on a range of samples, including cell lines and clinical specimens. Importantly, inclusion of paired sample sets allowed us to demonstrate a strong concordance between patient-matched frozen and FFPE material, showing the applicability of the NanoString platform to the most commonly available type of DLBCL patient tissue. Sample processing and turnaround time were user friendly; RNA input requirements were minimal and achievable from a single 10 μm clinical section.

Affymetrix microarray profiling is the most well-established COO subclassification methodology and considered the gold standard. To compare NanoString with Affymetrix, we profiled a set of 34 well-annotated DLBCL clinical samples, with resulting COO scores showing remarkable concordance. We extended this comparative analysis by profiling 48 additional patient samples from the fostamatinib clinical trial across both platforms with similarly concordant COO scores. Any discordant COO designations were due to small changes in the COO score around the cutoffs and highlights the importance of considering the score in the context of patient clinical characteristics.

Figure 6.
Identification of potential BCL2 translocation events from 59 patients in a fostamatinib phase IIb trial using NanoString technology. Two probes spanning common BCL2 fusion breakpoints were designed to identify patients with potential BCL2 translocation, which typically results in the loss of the BCL2 3′ UTR. The scatter plot shows the expression of these two probes, where the x-axis is the probe upstream of the breakpoints, and the y-axis is the probe at the 3′ UTR downstream of the breakpoints. The solid line is the correlation between these two probes when there is no translocation. The dashed lines indicate 4-fold deviation from the regression line. Each point represents a patient. Patients at the lower right outside the dashed line are potential BCL2 translocation candidates. Cell line OCI-Ly8, which is known to carry the BCL2-IGH translocation, is highlighted. Thirteen of the 15 BCL2 translocations are classified as GCB subtype, which is consistent with the literature.
conjunction with the designation to distinguish between clear and marginal GCB/ABC classifications. These data are in close agreement with a recent report using a small 20 gene NanoString code set (36).

The frequent clinical use of the Hans IHC algorithm prompted us to compare a subset of 10 FFPE samples across both NanoString and IHC. The Hans COO designations showed a 90% concordance with designations from NanoString profiling of 145 genes. While concordance in this small sample set is high, it is important to recognize differences in the data output between these two platforms despite the identical sample input demand (one FFPE section). The Hans approach delivered qualitative expression on just three proteins while NanoString returned fully quantitative data on more than 300 genes. Our cross-platform results comparing NanoString with IHC and Affymetrix are consistent with a recently published report using a smaller DLBCL sample set (37).

NanoString is not the only quantitative, FFPE-applicable technology that has been used for COO segmentation of DLBCL. The qNPA platform uses a 12-gene COO panel and requires similar sample inputs as NanoString without needing RNA extraction (13). To compare qNPA and NanoString directly, 31 clinical samples (frozen and FFPE) were evaluated. COO scores were tightly correlated in both FFPE and frozen tissue and unlike the Affymetrix platform, relative gene expression data also showed strong correlation, even for genes at lower expression level (data not shown). While qNPA data compared favorably with NanoString, the setup of the current qNPA array is limited in its ability to multiplex large numbers of genes. At present, single qNPA arrays are restricted to a maximum of 48 genes while NanoString’s upper limit is 800 and Affymetrix arrays can multiplex thousands. Depending on the application, the ability to profile large numbers of genes may be an advantage of NanoString, especially for exploratory approaches. As COO signatures have already been refined multiple times over the past years, this particular feature may be less important for DLBCL segmentation. To ensure increased coverage of biology, we have chosen to be inclusive with our 145-gene COO NanoString code set. Scott and colleagues have recently published a 20-gene version of a NanoString code set for COO segmentation (36) and our analyses of 14 overlapping genes common to both code sets showed highly correlated COO scores (Supplementary Fig. S4), suggesting our 145-gene code set could be further refined.

The robust pilot data and promising comparative findings of NanoString versus other established methodologies gave us the confidence to profile patient samples from an active DLBCL clinical trial. The clinical activity of fostamatinib is being assessed in patients with DLBCL who have progressed on R-CHOP therapy in a 60 patient phase IIb study (17). As pre-clinical data suggested that ABC-type cell lines may be preferentially sensitive to fostamatinib, COO segmentation of patient DLBCL tissue was considered key to a possible patient selection strategy (ref. 25, AstraZeneca unpublished data). Fresh, predose DLBCL tumor biopsies (core needle) were successfully collected from nearly all patients during prescreening. These flash-frozen tissues were processed for NanoString and Affymetrix analyses. Fifty-nine samples were successfully profiled on NanoString using the complete 307-gene code set. The NanoString-derived COO designations of these 59 relapsed/refractory samples showed a 31% ABC and 56% GCB split, numbers comparable with those previously reported for diagnostic (pre-R-CHOP) DLBCL tissue. To our knowledge, these represent the largest COO dataset reported in tumor tissue from second line R-CHOP relapsed DLBCL. With the caveats associated with a sample set of only 59 patients, these data could suggest that R-CHOP treatment may not selectively enrich for ABC-type disease. Additional profiling of relapsed biopsies will be needed to corroborate this hypothesis. Furthermore, by comparing diagnostic tissue (pre-R-CHOP) from 18 patients to patient-matched rebiopsies (post-R-CHOP), we were able to show that nearly 90% maintained a stable COO designation during the course of extended R-CHOP therapy. This observation could have implications for future targeted agents that may intend to selectively treat ABC or GCB-type patients upon relapse to R-CHOP (i.e., ibrutinib). In this setting, patient rebiopsies may be requested. Our data suggest that diagnostic tissue may represent a suitable alternative to an invasive rebiopsy, at least in regards to COO classification.

In addition to expression profiling, the flexibility of the NanoString system allows for detection of disease-relevant gene fusions/translocations. Several groups have reported successful fusion detection using NanoString with 5’ and 3’ probes that span the specific breakpoint of the gene under investigation (38, 39). Using a similar strategy, we designed NanoString probes aimed at detecting BCL2 translocation events within the fostamatinib clinical trial sample set. The BCL2-IGH t (14;18) fusion is known to occur in DLBCL with an incidence of up to 30%; the majority are associated with the GCB subtype (33, 40). We detected putative BCL2 gene fusion events in 25% of fostamatinib-treated DLBCL samples, relative to that previously reported for BCL2-IGH (Fig. 6). All but two of these occurred in a sample with a GCB subtype, again consistent with the literature. The specific BCL2 NanoString assay used to detect BCL2 in the fostamatinib samples was designed to detect only major break points within the 3′/ITR that result in truncated BCL2 transcripts. Because the assay was not designed to detect intermediate and minor break points, we cannot rule out the possibility that other BCL2 translocations may have gone undetected. While additional profiling is underway to confirm the specific BCL2 fusion partner and expand the analyses, these observations demonstrate the flexibility of the NanoString platform compared with other methodologies.

Additional NanoString analyses from fostamatinib biopsies, including COO subtype, BCL2 fusion incidence and/or baseline gene expression correlation to clinical response are ongoing and will be reported separately.

The findings from this study indicate that the NanoString system is a robust platform for molecular classification of DLBCL. NanoString offers several advantages to established techniques for COO segmentation, including multiplexing a large number of genes, low RNA input requirements, good reproducibility, complex genomic analysis (fusions) and sample type flexibility (FFPE and frozen) giving access to greater numbers of clinical samples. We have successfully applied NanoString to a large set of relapsed/refractory DLBCL biopsies obtained from an ongoing clinical trial. The resulting data confirmed the robust COO classification outputs from NanoString and demonstrated the promise of the system to generate data not otherwise achievable using other techniques. The NanoString system should be strongly considered alongside other established approaches for the molecular characterization of DLBCL.
Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Authors’ Contributions
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): M.H. Veldman-Jones, K.S. Thress
Writing, review, and/or revision of the manuscript: M.H. Veldman-Jones, Z. Lai, C.G. Harbron, J.C. Barrett, E.A. Harrington, K.S. Thress
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): M.H. Veldman-Jones, K.S. Thress

References


Reproducible, Quantitative, and Flexible Molecular Subtyping of Clinical DLBCL Samples Using the NanoString nCounter System


Updated version
Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-14-0357

Supplementary Material
Access the most recent supplemental material at:
http://clincancerres.aacrjournals.org/content/suppl/2014/10/11/1078-0432.CCR-14-0357.DC1

Cited articles
This article cites 34 articles, 16 of which you can access for free at:
http://clincancerres.aacrjournals.org/content/21/10/2367.full#ref-list-1

Citing articles
This article has been cited by 5 HighWire-hosted articles. Access the articles at:
http://clincancerres.aacrjournals.org/content/21/10/2367.full#related-urls

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, contact the AACR Publications Department at permissions@aacr.org.