Diffuse large B-cell lymphoma (DLBCL) is the most common lymphoma worldwide and consists of a heterogeneous group of cancers classified together on the basis of shared morphology, immunophenotype, and aggressive clinical behavior. It is now recognized that this malignancy comprises at least two distinct molecular subtypes identified by gene expression profiling: the activated B-cell-like (ABC) and the germinal center B-cell-like (GCB) groups—the cell-of-origin (COO) classification. These two groups have different genetic mutation landscapes, pathobiology, and outcomes following treatment. Evidence is accumulating that novel agents have selective activity in one or the other COO group, making COO a predictive biomarker. Thus, there is now a pressing need for accurate and robust methods to assign COO, to support clinical trials, and ultimately guide treatment decisions for patients. The “gold standard” methods for COO are based on gene expression profiling (GEP) of RNA from fresh frozen tissue using microarray technology, which is an impractical solution when formalin-fixed paraffin-embedded tissue (FFPET) biopsies are the standard diagnostic material. This review outlines the history of the COO classification before examining the practical implementation of COO assays applicable to FFPET biopsies. The immunohistochemistry (IHC)-based algorithms and gene expression–based assays suitable for the highly degraded RNA from FFPET are discussed. Finally, the technical and practical challenges that still need to be addressed are outlined before robust gene expression-based assays are used in the routine management of patients with DLBCL.
croarrays, covering a broad range of genes expressed by lymphoid cells and others suspected of being involved in cancer and immunology, to compare the gene expression profiles of lymphoid tumors against profiles from normal lymphoid cells.\(^4\) Using hierarchical clustering (a method of grouping tumors together based on similarity of gene expression), it was shown that there were at least two distinct groups within DLBCL. The first group, which was designated as GCB, was characterized by high expression of genes expressed in normal B cells from the germinal center. The second group showed very low expression of these germinal center B cell genes, but high expression of genes normally expressed by activated B cells and was thus named ABC DLBCL. Importantly, these groupings explained some of the clinical heterogeneity in DLBCL, as patients with ABC DLBCL experienced substantially worse outcomes compared with patients with GCB DLBCL.

This work was expanded by the Leukemia Lymphoma Molecular Profiling Project (LLMPP), by applying the same microarray technology to a larger cohort of patients.\(^5\) Hierarchical clustering, using the 100 genes that were most highly differentially expressed between GCB and ABC, identified three clusters within DLBCL: the aforementioned GCB and ABC, and a third group of cases that had gene expression that fell between these two groups, which they named “type 3.” With cyclophosphamide/doxorubicin/vincristine/prednisone (CHOP) chemotherapy (the standard at that time), this group of patients had outcomes similar to those with ABC DLBCL. This type 3 nomenclature gave the impression that this was a biologically distinct third group of tumors. In hindsight, it is possible this group contained some cases that had low tumor content and some cases that were later identified as primary mediastinal B cell lymphoma.\(^8\) What is clear is that when, in further studies, such cases were carefully removed, about 10% to 15% of DLBCL tumors remained that did not convincingly fall into either ABC or GCB categories—these cases are now labeled “unclassified.”

### The COO Algorithm

The LLMPP rapidly realized that in order for COO to be accurately assigned by other groups, there needed to be shift away from both the vagaries of hierarchical clustering and a microarray chip that was not commercially available. In 2003, Wright et al defined an algorithm that allowed classification of individual cases.\(^9\) This algorithm was based on 27 genes on the Lymphochip microarray that were the most discriminatory between GCB and ABC, with these genes having both high fold-change differences between the two groups and also high variance in expression over the DLBCL population. These characteristics underlie the robustness of this algorithm even when other technology platforms are used to determine gene expression.

The algorithm produces a linear predictor score (LPS) using an equation that sums the gene expression of each gene multiplied by individual gene coefficients. Normal distributions of LPS scores were constructed for the ABC and GCB groups and then, for a given LPS score, a probability is assigned indicating whether that individual case is best allocated to the ABC or GCB group (Fig. 1). If the probability of the case being an ABC is over 90% it is designated ABC, and if the probability of being an ABC is less than 10% (i.e., greater than 90% chance of being a GCB) the case is called a GCB. This approach produces an “unclassified” group (roughly 10% to 15% of the tumors), in which the tumor cannot be assigned to either GCB or ABC with sufficient confidence. Although this would appear to be an undesirable feature of the algorithm, it has a distinct advantage when producing robust assays for COO. In a binary system, where there is no unclassified group, small shifts in the LPS, which could occur on retesting a tumor or testing it with a different technology, would lead to a shift from ABC to GCB, or vice versa. The unclassified group, by producing a buffer zone between ABC and GCB, minimizes the chances of a tumor being frankly misclassified (e.g., an ABC tumor labeled a GCB).

Fourteen of the initially selected 27 genes are present on commercially available Affymetrix microarray chips (Fig. 1) and an adjusted algorithm including just those 14 genes provides very similar results to the full algorithm.\(^9\) The LLMPP have subsequently further refined the algorithm to include approximately 180 genes present on Affymetrix microarrays,\(^9\) although the increase in accuracy over the 14-gene model is incrementally small. GEP on fresh frozen tissue with COO assignment using these algorithms is now widely considered to be the “gold standard” and, because of the robustness of the algorithm, is likely producing acceptable consistency of COO assignment across research groups. The caveat to this is that the accurate deployment of the
Wright algorithm requires that gene expression data are normalized back to that of the original cohort used to define the algorithm, which is a process that requires a large number of tumors that have a similar proportion of ABC and GCB as the original cohort. This error-prone process means that the position of the unclassified group is likely differently located across research groups, producing (subtly) different proportions of ABC, unclassified, and GCB DLBCL.

**COO Defines Biologically Distinct Groups**

Since the observation that gene expression could separate tumors into the two COO groups, ample evidence has emerged that these are, indeed, distinct biologic groups (recently reviewed by Shaffer et al). Targeted genetic examination and subsequent genome-wide studies have revealed an array of genetic aberrations that are heavily enriched, or exclusive to either ABC or GCB DLBCL. It has emerged that ABC DLBCL is characterized by constitutive activation of NF-κB, through chronic active B cell receptor signaling and/or aberrant Toll-like receptor signaling, in concert with a differentiation block preventing maturation beyond the plasmablast stage. Although a straightforward schema is yet to emerge in GCB DLBCL, mechanisms providing a similar block to differentiation are present preventing maturation beyond the germinal center B cell. The GCB and ABC gene expression signatures largely reflect the presence of these differentiation blocks, with the tumor cells continuing to express remnants of the gene expression program of normal germinal center B cells or plasmablasts, respectively.

The separation of DLBCL into ABC and GCB groups explains a substantial proportion of the heterogeneity seen in the genetic mutation landscape of DLBCL. However, it should be appreciated that there is still marked heterogeneity even when ABC and GCB are considered as separate diseases. This is reflected when considering COO as a prognostic biomarker. As a group, patients with GCB DLBCL have superior outcomes following R-CHOP immunotherapy. Hidden heterogeneity is implied by the observation that those patients known to have the poorest prognosis, being those with translocations involving MYC and BCL2 (so-
called double-hit lymphoma) are almost exclusively GCB DLBCL. Correlation between genetic aberrations and outcomes after R-CHOP treatment within ABC and GCB groups is the subject of ongoing research.

**COO as a Predictive Biomarker**

From the clinical perspective, the observations and research linking COO with response to therapeutics are particularly exciting, holding the promise of improved patient outcomes. In some cases, this linkage is based on determining COO retrospectively in clinical trials searching for previously unsuspected differential efficacy across the COO groups, in which the mechanism for these observations are yet to be explained. At the other end of the spectrum are therapeutic agents that are highly likely to only have efficacy in one or other of the COO groups based on their established mechanism of action.

Accurate and reliable assays for COO are critically needed to allow patient selection for clinical trials of these agents, enriching for patients likely to truly benefit. Ultimately these assays will be needed to guide treatment decisions in the standard clinical management of patients with DLBCL. In order for an assay to be suitable for this task, it needs to be applicable to the FFPET biopsies routinely produced in the diagnostic process and have a rapid turnaround time.

**TRANSLATION TO PRACTICAL TECHNOLOGY PLATFORMS**

The gold standard methods originally described to assign COO rely on fresh frozen material and microarray technology. These methods are impractical for routine use in clinical trials, let alone standard clinical practice. Reflecting the importance of reliably assigning COO in the clinic, considerable efforts have been made to approximate the results of the gold standard method using practical technology platforms including IHC and methods for gene expression quantitation using RNA derived from FFPET.

**IHC-Based Assays**

The wide availability and familiarity of pathologists with IHC makes assays based on this technology highly desirable. In 2004, Hans et al developed an IHC-based algorithm applicable to FFPET biopsies that aimed to replicate the COO assignments made using GEP on fresh frozen tissue with the Lymphochip microarray. A tissue microarray (TMA) was produced comprising FFPET from 152 de novo DLBCL tumors, with 142 of these having matching GEP results. Hans et al trained a sequential IHC algorithm (Fig. 2) using FFPET biopsies on TMAs against COO assignments that had been made on matching fresh frozen tissue using the Lymphochip microarray. The assay was then validated using an independent group of 63 tumors that had matching GEP COO assignment with Affymetrix arrays. The reported concordance between the IHC-based assay and GEP in the validation cohort was 88%. Meyer et al recognized that the sequential nature of these algorithms may be discarding important information that could be captured if the results of the stains were tallied up, thus producing the “Tally” algorithm (Fig. 2). A tie-breaker—if the scores supporting ABC were equal to those supporting GCB—was provided by staining for an antibody to LMO2, which is a protein that is highly expressed in germinal center B cells. Trained against COO assigned by GEP using Affymetrix microarrays on matched fresh frozen tissue, they reported a concordance of 93% in the training cohort. Finally, Visco et al trained an IHC-based algorithm using 431 FFPET tumor biopsies on TMA using GEP on FFPET, rather than fresh frozen material, as the gold standard. They also reported a 93% concordance with COO by GEP and, similar to the Tally algorithm, this awaits confirmation in an independent validation cohort. A perceived weakness of these three studies is that they up-front excluded cases that were unclassified by GEP in both the training of the algorithms and calculations of performance. Three other studies have used antibodies that stain proteins that are differentially expressed between ABC and GCB DLBCL to produce assays that are prognostic in DLBCL. It is important that these assays not be considered as COO assays as they were trained against outcome to treatment and not against a COO gold standard.

The described performance of these assays (Fig. 2) appeared very promising. However, a number of studies have highlighted issues regarding reproducibility and accuracy of IHC-based assays for COO. The Lunenburg Lymphoma Biomarker Consortium examined the reproducibility and reporting of a number of IHC stains, including for CD10, BCL6, and MUM1. TMAs were distributed to eight laboratories for COO being assigned, with 142 of these having matching GEP COO assignment with Affymetrix arrays. The reported concordance between the IHC-based assay and GEP in the validation cohort was 88%. Meyer et al recognized that the sequential nature of these algorithms may be discarding important information that could be captured if the results of the stains were tallied up, thus producing the “Tally” algorithm (Fig. 2). A tie-breaker—if the scores supporting ABC were equal to those supporting GCB—was provided by staining for an antibody to LMO2, which is a protein that is highly expressed in germinal center B cells. Trained against COO assigned by GEP using Affymetrix microarrays on matched fresh frozen tissue, they reported a concordance of 93% in the training cohort. Finally, Visco et al trained an IHC-based algorithm using 431 FFPET tumor biopsies on TMA using GEP on FFPET, rather than fresh frozen material, as the gold standard. They also reported a 93% concordance with COO by GEP and, similar to the Tally algorithm, this awaits confirmation in an independent validation cohort. A perceived weakness of these three studies is that they up-front excluded cases that were unclassified by GEP in both the training of the algorithms and calculations of performance. Three other studies have used antibodies that stain proteins that are differentially expressed between ABC and GCB DLBCL to produce assays that are prognostic in DLBCL. It is important that these assays not be considered as COO assays as they were trained against outcome to treatment and not against a COO gold standard.
FIGURE 2. Immunohistochemistry-Based Assays for COO Assignment

The four immunohistochemistry-based algorithms that were trained against gold standard gene expression profiling are shown. Accuracy is defined as the concordance with the gold standard. Sensitivity and specificity are shown for the ABC subtype—as they are binary assays, the specificity for ABC is the sensitivity for GCB, whereas the sensitivity for ABC is the specificity for GCB. The performance values are determined from independent validation cohorts, except where marked with an *, where they are from the training cohort and thus may overestimate the performance of the assay.
pathologist improved. However, the level of agreement with regards to COO was still suboptimal at 77% (κ statistic of 0.62). This level of agreement was improved if stains were excluded if no positive internal control was seen, but this resulted in a substantial rise in the failure rate of the assays. In response to the technical issues identified with BCL6, it has been suggested that the Hans and Choi classifiers be modified to exclude the BCL6 antibody. More recently, Coutinho et al examined the level of agreement across the IHC-based assays showing that the concordance between the assays was generally low. The study of Gutierrez-Garcia et al directly addressed the accuracy of the IHC-based assays comparing the COO assignments by the Hans, Choi, and Tally algorithms with those from GEP on matched fresh frozen tissue using Affymetrix microarrays. The results were disappointing, with concordance rates between 59% and 65% for the three IHC-based assays compared with the gold standard. The issues with reproducibility and the variable agreement with the gold standard COO likely play major roles in the observation that, although COO by GEP is prognostic in the R-CHOP era, the literature is highly discordant when COO is assigned by IHC-based assays. A meta-analysis of the IHC-based studies and the GEP-based studies has recently been published illustrating this point.44

In aggregate, despite the careful studies that defined the assays and the wide availability of IHC, the accumulated evidence indicates that IHC-based methods are not ready to guide clinical care at this time. If IHC-based assays are to be used to make management decisions, there will have to be a shift away from individual laboratory–developed tests and toward defined kits, where standardized antibodies, laboratory techniques, and agreement on scoring may improve reproducibility and accuracy, with reproducible observer scoring representing the most formidable obstacle to standardize. This approach is currently being used with the Hans algorithm providing the patient selection for a phase III, randomized, controlled trial assessing the efficacy of adding ibrutinib to R-CHOP in non-GCB DLBCL (ClinicalTrials.gov Identifier NCT01324596). The randomization to the treatment arms, stratified by COO to ensure equal proportions of patients with ABC and GCB in each arm, occurs after one cycle of R-CHOP to allow patients to commence treatment in a timely fashion while awaiting the GEP results.

The LLMP used the NanoString platform to develop a parsimonious gene expression–based COO assay for RNA from FFPET. We took advantage of the ability of this platform to quantitate up to 800 RNA species in multiplex to build a model that most accurately replicated the larger GEP model described in Lenz et al. In brief, 93 genes that were most differentially expressed between ABC and GCB on Affymetrix microarray data from fresh frozen tissue were quantified in RNA extracted from matched FFPET biopsies on the NanoString platform. Genes were then selected based on the degree of correlation of the expression between the two platforms (and across the sources of RNA). In total, 15 genes (Fig. 3A) were sufficient to accurately replicate the model from Lenz et al with the addition of further genes resulting in negligible increase in assay accuracy. The low-density assay, comprising only these 15 genes and five house-keeping genes, was applied to samples that contributed to the training of the Lenz et al model, the final model was built and then “locked.” This assay, named the Lymph2Cx, was then tested in a large cohort of patients with de novo DLBCL treated with R-CHOP, with the expected survival difference observed between the GCB and ABC groups.

A range of other technology platforms have subsequently been used as the basis of COO assays, including quantitative nuclease protection assay, multiplex reverse transcription (RT)–polymerase chain reaction (PCR), DASL, and NanoString. In studies in which the comparator was the gold standard of GEP in fresh frozen tissue, concordance was typically high at approximately 95%, when unclassified cases are removed. Two studies will be discussed in more detail below as these assays are now supporting phase III randomized controlled trials.

Barrans et al have developed an assay for DLBCL COO based on the cDNA-mediated Annealing, Selection, extension, and Ligation (DASL) platform (Illumina, San Diego, CA), which provides genome-wide GEP in degraded RNA samples. This technology, in combination with a new tool for COO assignment based on balanced voting between four machine-learning tools, was used to assign COO in a population-based cohort of 172 patients with DLBCL. The COO assignments in this population were 48% GCB, 31% ABC, and 21% type 3 (presumably equivalent to unclassified). Although there was no comparison with the gold standard COO assignment from fresh frozen material, the characteristics of the ABC and GCB groups were consistent with previous reports, with inferior outcomes following treatment in the ABC group. This assay is currently being used to stratify patients for a phase III, randomized, controlled trial assessing the efficacy of adding bortezomib to R-CHOP in patients with DLBCL (ClinicalTrials.gov Identifier NCT01324596). The randomization to the treatment arms, stratified by COO to ensure equal proportions of patients with ABC and GCB in each arm, occurs after one cycle of R-CHOP to allow patients to commence treatment in a timely fashion while awaiting the GEP results.

Gene Expression–Based Assays

Over the last decade, technologies have been developed that allow GEP employing the highly degraded RNA that is extracted from FFPET. Since 2010, a number of GEP-based assays have been described capitalizing on these advances. The generally high degree of accuracy of these assays is a testament to the very distinct gene expression patterns between ABC and GCB, and the robustness of the COO algorithm on which most of these assays are based.

The first demonstration of robust COO assignment using GEP on RNA from FFPET came from techniques that make this degraded RNA suitable for GEP using standard microarray technology. Williams et al compared the results of applying the Wright algorithm to GEP using RNA from FFPET compared with that of matched fresh frozen tissue. When unclassifiable cases were excluded, there was one frank misclassification out of 44 cases, giving a concordance of 98%.
in an independent cohort of 68 FFPET biopsies from the validation cohort of Lenz et al cases that have never contributed to COO model building. Compared with the gold standard, only one case was frankly misclassified (ABC to GCB) and the assay had a similar rate of unclassified cases to the gold standard. The Lymph2Cx was tested across two independent laboratories and showed 95% agreement of COO assignment, with discordant cases having shifted from a definitive assignment to unclassified, or vice versa (Fig. 3B). To the author’s knowledge, this is the only published data addressing interlaboratory performance of the gene expression–based COO assays. The Lymph2Cx has subsequently been applied to FFPET biopsies from a population-based cohort of patients with de novo DLBCL, displaying a low failure rate of 1%, an unclassified rate of 11%, and substantial separation of outcomes following R-CHOP between the ABC and GCB groups. The assay has now been developed into a companion diagnostic by NanoString and is used to select patients for inclusion in a phase III, randomized, controlled trial assessing the efficacy of adding lenalidomide to R-CHOP in patients with ABC DLBCL (ClinicalTrials.gov Identifier NCT02285062). The rapid turnaround time of the assay is allowing randomization before starting treatment.

The unclassified category identifies tumors in which COO cannot be assigned with sufficient confidence. It is anticipated that clinical trials testing novel agents will be designed to include or exclude these patients by weighing the potential, albeit uncertain, benefits against the expected toxicities of the therapeutic. However, it is possible that patients with unclassified DLBCL will be left in a therapeutic wasteland,
where no agents have been specifically trialed in this group. This will particularly be an issue with assays that assign larger proportion of tumors in this category. It is hoped that the use of assays that provide more uniform and consistent COO assignment that we will be able to determine the genetic mutation landscape of unclassified cases and ascertain whether they represent a unique biologic group.

Ongoing Challenges for Gene Expression-Based Assays
In order for gene expression–based assays to be broadly and reliably deployed, a number of challenges need to be addressed. The first is the interlaboratory performance of the assays, which has only been demonstrated for the Lymph2Cx assay. The second is the provision of adequate tissue for these assays. With the growing trend toward core needle biopsies for diagnosis of lymphoma, in many cases the biopsy material is exhausted in the routine diagnostic work-up. For these assays to be applicable to all patients, there will need to be a move back toward excisional biopsies and/or dedicated cores taken for molecular studies. Finally, most of these assays require the purchase and standardization of specialized equipment. It is anticipated that substantive evidence demonstrating the clinical and financial benefit of accurately determining DLBCL COO will be needed for the widespread adoption of these technologies.

ENVISING THE NEAR FUTURE
Although there is an expectation that eventually treatment of DLCBL will be guided by a comprehensive analysis of the genetic aberrations found in the tumor cells, the huge complexity that has been uncovered by genome-wide studies indicates that this will not be realized in the short-term or maybe even the medium-term. The first step toward precision medicine in DLBCL is the recognition that it comprises at least two distinct entities, as identified 15 years ago. The accurate and reliable ascertainment of COO in this disease will provide important enrichment for patients that will benefit from targeted therapeutics. It is anticipated that further genetic analyses will be performed on patients that do not respond to these agents to identify determinants of drug resistance.

At this point in time, there are no assays for COO assignment that can reliably inform treatment decisions outside of clinical trials. However, I firmly believe that by the time there is substantive evidence that new targeted agents improve patient outcomes, robust gene expression–based assays will be in place to identify the patients that will benefit.

ACKNOWLEDGMENT
Many thanks to Drs. Randy Gascoyne and Joseph Connors for their constructive comments. DWS is supported by the BC Cancer Foundation.

Disclosures of Potential Conflicts of Interest


References


