


Molecular profiling in diffuse large B-cell lymphoma: why so many types of subtypes?

Ryan D. Morin,^{1,2,3} Sarah E. Arthur^{1,3} and Daniel J. Hodson^{4,5} 

¹Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, BC, ²Genome Sciences Centre, BC Cancer, Vancouver, BC, ³BC Cancer Centre for Lymphoid Cancer, Vancouver, BC, ⁴Wellcome MRC Cambridge Stem Cell Institute, Cambridge Biomedical Campus, Cambridge, and ⁵Department of Haematology, University of Cambridge, Cambridge

Summary

The term diffuse large B-cell lymphoma (DLBCL) includes a heterogeneous collection of biologically distinct tumours. This heterogeneity currently presents a barrier to the successful deployment of novel, biologically targeted therapies. Molecular profiling studies have recently proposed new molecular classification systems. These have the potential to resolve the biological heterogeneity of DLBCL into manageable subgroups of tumours that rely on shared oncogenic programmes. In many cases these biological programmes straddle the boundaries of our existing systems for classifying B-cell lymphomas. Here we review the findings from these major molecular profiling studies with a specific focus on those that propose new genetic subgroups of DLBCL. We highlight the areas of consensus and discordance between these studies and discuss the implications for current clinical practice and for clinical trials. Finally, we address the outstanding challenges and solutions to the introduction of genomic subtyping and precision medicine in DLBCL.

Keywords: lymphomas, diffuse large B-cell lymphoma, cancer genetics, mutation analysis, classifications.

Diffuse large B-cell lymphoma (DLBCL) is an aggressive non-Hodgkin lymphoma and the commonest lymphoid malignancy.¹ First-line therapy with immunochemotherapy regimens such as cyclophosphamide, doxorubicin, vincristine, prednisone + rituximab (R-CHOP) now cures up to 60% of patients. However, those failing first-line therapy present a significant clinical challenge and a majority of these patients still die from their disease. To address this need, the molecular pathogenesis of DLBCL has been the focus of intense study. The extent of new biological and genetic understanding that has been amassed in recent years is remarkable. It has highlighted multiple targetable oncogenic pathways and

launched the development of a plethora of promising novel therapeutic agents. However, this excitement has been tempered by more sobering statistics; since R-CHOP was introduced in 2002,² numerous phase-three trials have examined modifications or additions to R-CHOP as first-line therapy for DLBCL. Many have examined the use of promising novel therapeutic agents.^{3–8} However, none of these trials has met its primary end-point and first-line therapy for DLBCL has remained unchanged for almost two decades.² Although the promise of molecularly targeted therapy in DLBCL may be more elusive than we had hoped, one can also view this as an opportunity to re-assess the design of clinical trials to increase our chances of future improvements to clinical management.

The greatest barrier to the successful introduction of targeted therapy in DLBCL is the diversity of genetic features and phenotypes within this clinical entity. It continues to become clear that DLBCL is not a single disease but a collection of diseases, each with defining molecular and biological features. The full repertoire of distinguishable entities that comprise this spectrum are beginning to be described but remains far from fully resolved. Since each entity may rely on distinct oncogenic pathways it is reasonable to assume that novel therapies designed to inhibit specific oncogenic pathways may have activity only in certain subgroups and therefore may show little detectable activity if used in a blanket approach. Recent molecular profiling studies provide us with a handle to rationalise this biological heterogeneity into more homogeneous subgroups of DLBCL. It is premature to consider that knowledge of these subtypes can direct the optimal treatment for an individual patient. However, it provides a framework on which to design and interpret the results of future clinical trials.

Existing strategies for molecular profiling of DLBCL

The first insight into the molecular heterogeneity of DLBCL arose from the application of gene expression profiling using microarray technology.⁹ This showed that DLBCL could be divided into two groups with distinct gene expression

Correspondence: Daniel J. Hodson, Department of Haematology, University of Cambridge, Cambridge CB20AW, UK.
E-mail: djh1002@cam.ac.uk

patterns; one resembling normal germinal-centre B cells and another sharing features with blood B cells activated *in vitro*. These transcriptional subtypes became known as germinal-centre (GCB) and activated B cell (ABC) respectively. A statistical approach was later developed to rank DLBCLs based on their expression of these genes and assign them probabilistically into one of these two groups, leaving a small number of cases unclassified (UC).¹⁰ Although it had not been established that these classes originated from distinct cell types, this classification system became known as cell of origin (COO). A number of implementations of COO classifiers have since been developed that could be applied to microarray, RNA-Seq or more focused gene expression measurements including NanoString or HTG EdgeSeq technology.^{11–13} The requirement for sophisticated transcriptional profiling technology initially limited the utility of COO profiling to research applications. Attempts to reproduce the COO transcriptional classification in a standard diagnostic setting by scoring surrogate protein markers with immunohistochemistry met with variable success,^{14,15} but at best were able to offer a binary classification of GCB or non-GCB that imperfectly recapitulated the transcriptional classification.

Initial excitement for this stratification related to the observation that patients with ABC DLBCL had, on average, a worse prognosis than GCB DLBCLs.^{11,16} However, perhaps the most significant consequence of the COO classification was to provide a framework on which to build our understanding of DLBCL biology. Indeed, the experimental comparison of ABC and GCB biopsies and cell lines has allowed us to dissect critical oncogenic pathways unique to each subtype.¹⁷ In contrast, its contribution to routine clinical practice has been more limited and COO was not incorporated into the revised WHO classification of DLBCL until 2016.¹⁸ Although there are numerous therapies targeting genetic features that are strongly associated with only one COO, there is currently no compelling evidence to suggest that ABC and GCB patients receive benefit from different treatments up-front or in the relapse setting.

An alternative strategy to subclassify DLBCL is based upon the detection of rearrangement of *MYC*, *BCL2* and/or *BCL6* by fluorescent *in-situ* hybridisation (FISH), a technique routinely available in diagnostic laboratories. These double/triple translocated cases are now assigned to a new diagnostic entity that is currently considered distinct from DLBCL in the 2016 revised WHO classification.¹⁸ This is termed high-grade B-cell lymphoma with *MYC* and *BCL2* and/or *BCL6* rearrangements, or HGBL-DH/TH (also commonly referred to as double-hit or triple-hit lymphoma), which can also include tumours that do not have DLBCL morphology. However, it remains far from clear that the detection of *MYC* and *BCL2/BCL6* rearrangements identifies a biologically homogeneous lymphoma subtype. The impetus for this classification was driven in part by the perceived impact on prognosis; initial retrospective series suggested a dismal survival amongst lymphomas with rearrangement of both *MYC*

and *BCL2*.^{19,20} However, this may partially reflect the initial preferential testing of high-risk cases as subsequent prospective clinical trials have confirmed a definite but more modest negative prognostic impact.^{21,22} The negative prognostic impact of double-hit lymphoma (DHL) may be restricted to those cases with translocation between *MYC* and the immunoglobulin loci rather than other rearrangement partners, a distinction that may not be discernible in all routinely used FISH assays but can be revealed by dual-fusion probe assays.²² This situation is further confounded by the existence of pseudo-double-hit lymphoma, specifically those with t(3;8)(q27;24), which are indistinguishable from separate *BCL6* and *MYC* translocations by break-apart FISH.²³ Many centres have adopted intensified immunochemotherapy approaches for double-hit patients based on retrospective studies.^{24,25} However, this approach has never been tested in a prospective, randomised setting. This widespread adoption now acts as a barrier to a randomised trial and is a situation we should endeavour to avoid repeating as emerging genetic subtypes become incorporated into clinical practice.

More recently, two groups used complementary gene expression approaches to identify *MYC*-driven subtypes of DLBCLs. One of the signatures was derived from the subset of DHL with *MYC* and *BCL2* translocations. These cases were almost all GCB by COO profiling and termed the ‘double-hit signature’ (DHITsig).²⁶ The other study used a Burkitt lymphoma signature to identify ‘Molecular High-Grade’ (MHG) lymphomas.²⁷ Both of these signatures identify *MYC*-driven DLBCL cases that overlap only partially with HGBL-DH/TH. For example, approximately half of tumours classified as DHITsig⁺ lack one of the expected oncogene rearrangements by FISH (e.g., *BCL2* or *MYC*). More thorough genetic characterisation of such cases revealed frequent cryptic rearrangements of the *MYC* or *BCL2* loci.²⁸ These two related transcriptional subgroups have clinical relevance because DHITsig/MHG tend to represent a subgroup of GCB DLBCLs with an inferior survival relative when treated with R-CHOP. Whether intensified or other therapy would benefit this group of patients remains to be determined. Figure 1 provides a high-level overview of the variety of assays in use for research and clinical application for lymphoma diagnosis and subclassification.

Genomic profiling of DLBCL

Initial unbiased genome and exome-wide sequencing studies reaffirmed the biological distinctions between ABC and GCB DLBCLs as well as revealing the genetic similarities between GCB DLBCL and follicular lymphoma.^{29–33} Many recurrently mutated genes were specific to B-cell lymphoma, suggesting biology distinct from that of epithelial malignancies. Investigation of the more frequent mutations revealed lymphoma-specific oncogenic mechanisms such as the mutation of chromatin modifiers in GCB DLBCL and the B-cell receptor (BCR) pathway in ABC DLBCL.^{34,35} As larger numbers of

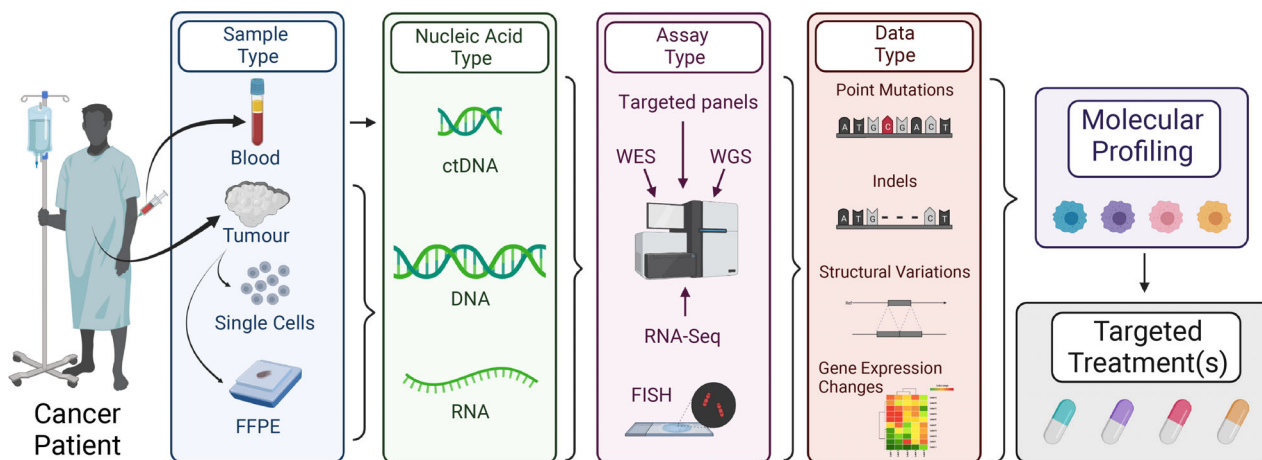


Fig 1. Molecular profiling for personalised medicine. An idealized personalised medicine workflow illustrating the diversity of sample and nucleic acid types that might be subjected to an equally diverse selection of molecular assays with the goal of aiding selection of the most appropriate therapeutic.

patients were sequenced the genomic complexity of DLBCL became increasingly apparent.³⁶ In contrast to other haematological malignancies, DLBCL shows a greater number of mutations per patient and a larger number of recurrently mutated genes, with a long tail of genes mutated in only a small number of cases. Overall, the field has now converged onto approximately 150 protein-coding driver genes that are recurrently mutated or functional targets of somatic copy number alterations in DLBCL. Whilst the majority of these genes are mutated in only a minority of patients it became clear that these mutations were not randomly distributed. For example, cases with mutant *MYD88* were more likely to carry a mutation of *CD79B*. In contrast, cases with *BCL2* translocation are enriched for *CREBBP* and *EZH2* mutations. This suggested that mutations might cluster into functional groups that could represent biological subtypes of DLBCL.

To pursue this concept, Staudt and colleagues at the National Cancer Institute (NCI) performed transcriptional profiling, whole-exome sequencing, targeted mutation sequencing and array-based copy number analysis on 574 cases of DLBCL.³⁷ They identified the most enriched combinations of genetic alterations in each transcriptional subtype of DLBCL and found patterns of genetic features that extended beyond the two main COO subgroups. They devised a set of subdivisions or genetic subgroups named after the most common distinguishing features of each group. One of these was characterised by *MYD88* and *CD79B* mutations (MCD) and strongly associated with ABC DLBCLs. Another was enriched for *EZH2* mutation and *BCL2* translocation (EZB) and was prototypical of GCB DLBCL. Stemming from an observation that *BCL6* structural alterations and *NOTCH2* mutations were enriched among cases unclassifiable by COO, a third group (BN2) was proposed. They also noticed a small number of ABC patients with mutations in *NOTCH1* that were mutually exclusive

with other ABC or *NOTCH2* mutations and were considered a separate group (N1). Just over half of patients remained UC suggesting further subtypes remained to be described. A summary of the genetics of the NCI cohort and their relationship to COO is shown in Fig 2.

Concurrent with this work, Shipp and colleagues at Harvard applied a data-driven clustering strategy to mutational and copy number data derived from whole-exome and targeted sequencing of 304 DLBCLs.³⁸ In contrast to the NCI study, the Harvard group relied on an unsupervised consensus clustering method. This resolved patients into five clusters that mathematically shared the most similar repertoires of genetic alterations. These clusters were termed C1–C5. Despite the distinct statistical approach, considerable overlap with the NCI findings were apparent; the C1 cluster, enriched for *BCL6* fusion and *NOTCH2* mutation corresponded clearly to the BN2 group. The C3 cluster, enriched for translocation of *BCL2* and mutation of *CREBBP* and *EZH2* clearly corresponded to the EZB group. The C5 cluster, enriched for *MYD88* and *CD79B* mutations aligned to the MCD group. However, two new clusters emerged. C2 was dominated by mutation of *TP53* and widespread copy number alteration. C4 was enriched for somatic hypermutation of *SGK1* and genes encoding histone linker proteins. A tiny fraction of patients (4%) with no detectable mutations were categorised as C0.

A follow-up paper from the NCI then examined further cases not classified in their original publication.³⁹ They noted that these were enriched for tumours with high levels of aneuploidy and mutation of *TP53*. A second enrichment was seen for cases with mutation of *SGK1* and *TET2*. These genetic features were used to seed the clustering of two new subtypes termed A53 (aneuploidy, *TP53*) and ST2 (*SGK1*, *TET2*). These corresponded closely to the Harvard C2 and C4 clusters. Thus, each of the five Harvard clusters could

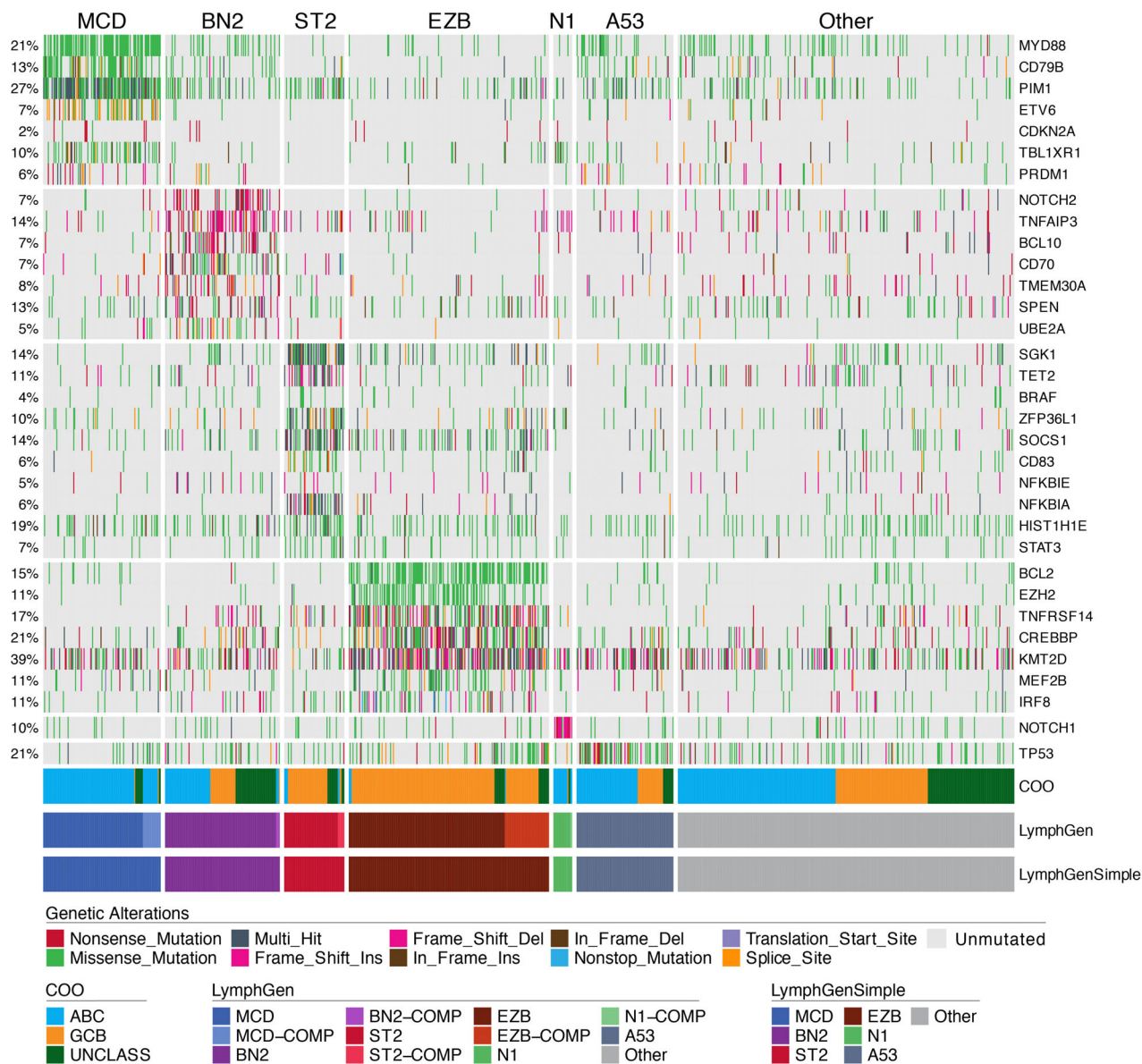


Fig 2. Classification of National Cancer Institute (NCI) cases using LymphGen genetic classifier. This oncoPrint shows the mutation status of patients sequenced in the NCI study.³⁷ Each column represents a single patient. Selected key genes with the greatest impact on classification are shown. Different mutation types are indicated with a distinct colour. The assignment of cell of origin (COO) of cases is shown below the oncoPrint. The LymphGen genetic classifier was used to classify each patient as shown by coloured bars at the figure base. The figure highlights important features of the LymphGen classification system. Firstly, a substantial fraction of cases remains unclassified or are assigned to more than one class. These ‘composite’ cases have a sufficient representation of genetic features from more than one class such that their classification is more ambiguous. A simplified classification (LymphGenSimple) is also presented whereby composite cases are assigned to a single class by selecting only one of the assigned classes using a prioritized set of classes.

now be mapped to one of the NCI genetic subgroups. However, there remained substantial discrepancies in how cases are assigned to each of the corresponding subgroups and a notable difference between studies was that over 40% of cases remained UC or genetically composite in the NCI study. The genetic classification system developed by the NCI became known as LymphGen and was released as a publicly available tool that can be applied to classify an individual case. A

comparison of the genetics and classification strategies of the NCI (LymphGen) and Harvard studies is presented in Fig 3.

Finally, a study from the UK Haematological Malignancy Research Network (HMRN) applied targeted sequencing of 293 genes to DNA extracted from archived formalin-fixed paraffin-embedded (FFPE) biopsies from 928 cases of DLBCL.⁴⁰ Sequencing data were subjected to unsupervised clustering using a Bernoulli mixture modelling strategy.

Clustering was based predominantly on mutation data, with copy number alterations considered only for a small number of genes. Translocation and gene fusion data were not available for all cases and were therefore excluded from the clustering. Therefore, this study differed from the NCI and Harvard studies in several respects: the sequencing strategy used, the types of the genetic data seen by the clustering algorithms, and the statistical approach employed to identify clusters. Despite this, genetic subtypes emerged that could be mapped almost precisely to NCI and Harvard categories. These were named according to the most enriched genetic feature. The group termed MYD88 recapitulated the MCD/C5 clusters. The *BCL2* group corresponded to EZB/C3. A *NOTCH2* group mapped to the BN2/C1 clusters. Finally, an *SGK1* group reproduced the ST2/C4 clusters. Perhaps owing to the greater number of cases in the HMRN study, the *SGK1* group was split into *SOCS1/SGK1* and *TET2/SGK1*. Overall, 27% of cases remained UC. Neither a *NOTCH1* mutant group or a *A53/C2* equivalent group were identified. In fact, the frequency of N1 cases (1.7% of all DLBCL in the NCI study) was too low to be resolvable as a distinct category in unsupervised clustering, whereas limited copy number data precluded identification of an *A53/C2* group in the HMRN study. However, rather than being classified elsewhere, both *TP53* and *NOTCH1* mutant patients were principally enriched amongst the UC cases suggesting their absence should not be considered as evidence against the validity of these groups. A further modification of the HMRN classification used the presence of *MYC* hotspot and *NOTCH1* PEST domain mutations to identify a *NOTCH1* and *BCL2-MYC* subgroups.⁴¹ This modified HMRN classifier showed high concordance when cases were reclassified using the NCI LymphGen classifier (Fig 4).

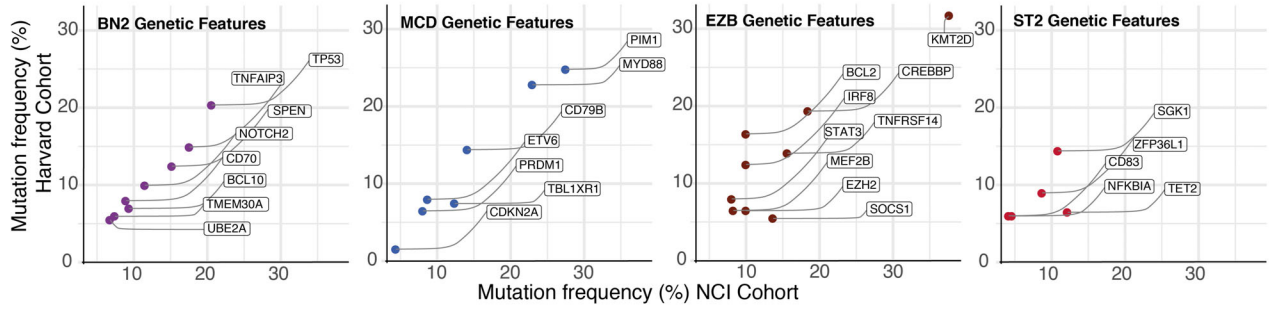
It is encouraging that studies with different sequencing and computational approaches independently converge on a remarkably similar system for the genetic classification of DLBCL. The degree of consensus provides strong evidence for the validity of genetically defined, biological subtypes of DLBCL. Gene expression profiling offers further support that each subtype relies on fundamentally different biological pathways.^{37,40} A comparison across studies and the salient feature of each subtype are summarised in Table I. For consistency we will adopt the NCI (LymphGen) nomenclature throughout the remainder of this study, unless referring to subgroups from a specific study. Ultimately, genetic classification provides a means to rationalise the genetic heterogeneity of DLBCL into subgroups that share a common biological pathogenesis and may therefore respond similarly to specific therapies. We anticipate that robust publicly available methods to assign DLBCLs to these subgroups will emerge as their definition continues to be refined. This, therefore, begins to address one of the biggest barriers to the introduction of targeted therapies or precision medicine to the treatment of DLBCL.

Superimposing the biology of genetic subtypes onto other known lymphoma entities

Examining the mutational repertoires of individual subtypes provides clues to the biology and reveals unexpected overlap with other categories of lymphoma distinct from DLBCL (Figs 5 and 6, Table I). MCD tumours have the strongest ABC expression signature and are characterised by mutations that activate BCR and toll-like receptor pathways, both of which converge onto increased nuclear factor kappa B (NFkB) activity.^{42–44} Other frequent genetic alterations include amplification of the *BCL2* locus, deletion of the cell cycle negative regulator *CDKN2A*, and mutations that converge upon immune evasion.³⁹ Almost all cases show the transcriptional profile of ABC DLBCL including increased signatures of NFkB and *MYC* activity.^{37,40} The genetic features of MCD overlap strongly with those reported in the extranodal lymphomas, including primary central nervous system lymphoma (PCNSL), primary breast lymphoma and primary testicular lymphoma (PTL).^{45–47} Indeed, cases of PCNSL and PTL included in the above clustering studies almost all clustered into the MCD subtype (Fig 5B). Recent single-cell analysis of normal lymph nodes suggests a substantial proportion of MCD lymphoma may arise from a distinct pre-memory B-cell stage of post-germinal-centre B-cell development (Fig 6).⁴⁸ This finding resonates with the frequent mutation of *TBL1XR1* in MCD DLBCL, since *TBL1XR1* mutations in mouse models promote memory B-cell expansion and extranodal lymphoma.⁴⁹ Overall, this suggests MCD DLBCL is a distinct form of DLBCL that, from a biological perspective, overlaps more strongly with PCNSL and PTL than it does with other forms of DLBCL not otherwise specified (NOS). The recurrent mutation of BCR and immune pathway genes within this subtype leads to hypotheses about potential therapeutic vulnerabilities that may be tested in future clinical trials of BCR or immune checkpoint inhibition.

At the other end of the biological spectrum, EZB tumours are strongly enriched for GCB DLBCLs. They are characterised by translocation of *BCL2* into the immunoglobulin locus, and mutation of histone modifiers such as *KMT2D*, *EZH2*, *CREBBP* and *EP300*. Mouse models have shown how mutations of these genes lead to a block of differentiation and sustained expression of the germinal-centre transcriptional programme, which co-operate with *BCL2* to drive lymphomagenesis.^{50–58} The mutation profile of EZB matches closely to that seen in follicular lymphoma (FL).^{29,33,59,60} The HMRN study showed that DLBCLs that had transformed from FL were strongly enriched in this cluster (Fig 5B).⁴⁰ The shared genetic features of transformed FL and EZB DLBCL is consistent with the notion that so-called *de novo* cases of EZB DLBCL and FL may both arise from a common origin (Fig 6).⁴⁰ Indeed, the HMRN study revealed that 27% of cases assigned to this subtype had evidence of previously

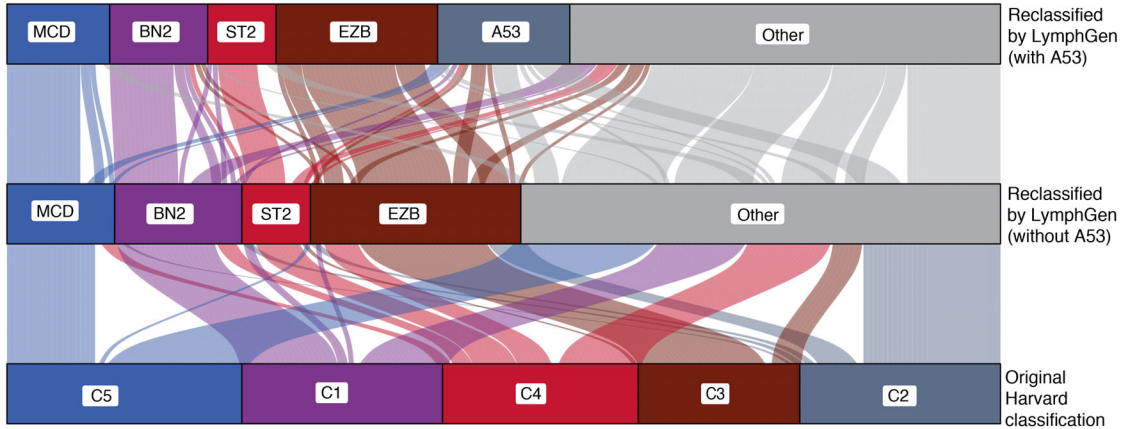
(A)



(B)



(C)



undiagnosed FL discovered on lymph node or trephine biopsy concurrent with the diagnosis of DLBCL.⁴⁰ This leads us to speculate that a substantial proportion of *de novo* DLBCLs of the EZB class may arise via transformation from an occult FL.

The EZB subtype also appears to be the primary genetic background of DHL. The HMRN study showed how HGBL-DH/TH cases (identified by FISH) were found predominantly within this subtype, as were the transcriptionally identified MHG cases.⁴⁰ Similarly, the NCI study used the closely related DHITsig transcriptional signature to reveal enrichment of DHITsig⁺ cases within the EZB subtype.³⁹ This suggests the existence of a subgroup of aggressive, MYC-driven lymphomas within the EZB cluster that represents a distinct disease entity that may arise from germinal-centre dark-zone centroblasts (Fig 6). It also suggests that MYC or BCL2 single or doubly rearranged cases from other genetic subtypes may not necessarily have the same clinical or biological implications. In the NCI LymphGen classification system, EZB is further subdivided into EZB-MYC⁺ and EZB-MYC⁻ to differentiate cases with and without either MYC rearrangement or the DHITsig gene expression signature. Similarly, in a modification to the original HMRN classification (modified HMRN), the presence of MYC hotspot mutations, which are known to correlate strongly with rearrangement status,⁶¹ were used to define MYC-driven cases within the BCL2 subgroup.⁴¹ Notably, tissue collection in most of the above studies occurred prior to the WHO revision to lymphoid classification which introduced the category of high-grade B-cell lymphoma with MYC and BCL2 and/or BCL6 rearrangements. These new genetic findings challenge the utility of this WHO disease category as a meaningful, homogeneous biological entity and motivate the revision of this definition.

The BN2 subgroup does not have unifying gene expression features and represents a mix of ABC, GCB and UC cases. BN2 is enriched for BCL6 fusion and mutation of NOTCH2 and other NOTCH pathway genes. Interestingly, transcriptional signatures of NOTCH activation were not identified in these tumours,^{37,40} suggesting that the activating NOTCH2 mutations may exert their effect at a specific stage of lymphoma development. Other mutations enriched in the BN2 cluster appear to activate NFkB; these include loss of

TNFAIP3, gain of BCL10 and 3'-untranslated region (3'UTR) mutation (leading to enhanced expression) of NFkBIZ.⁶² Many of these genetic alterations, especially NOTCH2 mutation, are reminiscent of marginal-zone lymphoma (MZL).^{63,64} It is tempting to speculate that these lymphomas arise from transformation of an underlying MZL (Fig 6). However, evidence of pre-existing MZL was not identified in the clustering studies.^{38,40} Nevertheless, it is clear that a shared biological programme and likely a shared cellular origin may link these two diseases.

The ST2 subtype also shares genetic similarity to indolent lymphoma — in this case nodular lymphocyte-predominant Hodgkin lymphoma (NLPHL)⁶⁵ but direct evidence of transformation from the latter remains to be identified. Mutations in this subtype (including SOCS1, DUSP2, STAT3 and BRAF) may lead to activation of JAK/STAT and ERK signalling pathways, a suggestion supported by gene expression signatures.⁴⁰ SGK1 mutations may lead to hyperstable protein isoforms that act in parallel to AKT in the PI3K pathway.⁶⁶ In the HMRN study the SGK1 subtype was subdivided into SOCS1/SGK1 and TET2/SGK1. The former shares genetic overlap with primary mediastinal B-cell lymphoma, including SOCS1, ITPKB, NFkBIE and CIITA.^{67,68} The SOCS1/SGK1 subtype was enriched for cases of PMBCL (Fig 5B); however, the remaining cases did not show preferential mediastinal involvement. This suggests that PMBCL, a tumour defined in part by its restricted anatomical involvement, may share considerable biological overlap with this subtype of nodal DLBCL NOS. This supports previous descriptions of non-mediastinal DLBCLs with gene expression features and genetics reminiscent of PMBCL and we speculate that these contribute, in part, to the SOCS1/SGK1 subgroup.^{69,70}

The N1 subtype is dominated by mutations that remove the degradation domain, and thereby activate the oncogene NOTCH1. These mutations are rare in DLBCL overall (1.7% and 2.4% in the NCI and HMRN studies respectively) but are common in chronic lymphocytic leukaemia (CLL) and in Richter's syndrome.^{71,72} N1 cases reported in the clustering studies do not appear to represent overt transformation of CLL but seem to share common oncogenic programmes with these diseases.

Finally, the A53 subtype is defined by widespread copy number variation. Mutation or deletion of TP53 is enriched

Fig 3. Comparison of the National Cancer Institute (NCI) and Harvard clusters. (A) The frequency of mutations in key genes associated with diffuse large B-cell lymphoma (DLBCL) genetic subgroups are compared between the NCI and Harvard cohorts for the four largest subgroups. Whilst strong agreement is seen for some genes, such as those defining the BN2/C1 group, there is more variable consensus across other subtypes, exemplified by genes such as TET2, BCL2 and SOCS1. (B) This oncoprint shows the genetic features of each class across three DLBCL cohorts including the Harvard,³⁸ NCI³⁷ and a cohort of patients from British Columbia.⁷⁴ LymphGen was applied to the genetic data from all three cohorts and the classifications are shown with and without composite labels — LymphGen and LymphGenSimple respectively. (C) The alluvial plot shows the relative proportion of cases from the Harvard cohort assigned to each class when reclassified by LymphGen. This reclassification was done in two ways; with or without the A53 option enabled in the LymphGen classifier. Vertical ribbons represent individual cases and can be followed from top to bottom. Notably, when the A53 class is available, a large number of cases switch classification from one of the core classes to A53. The plot demonstrates relatively high consensus between the two classification systems when considering the core classes, but weaker consensus over patients classified into the A53 and C2 clusters.

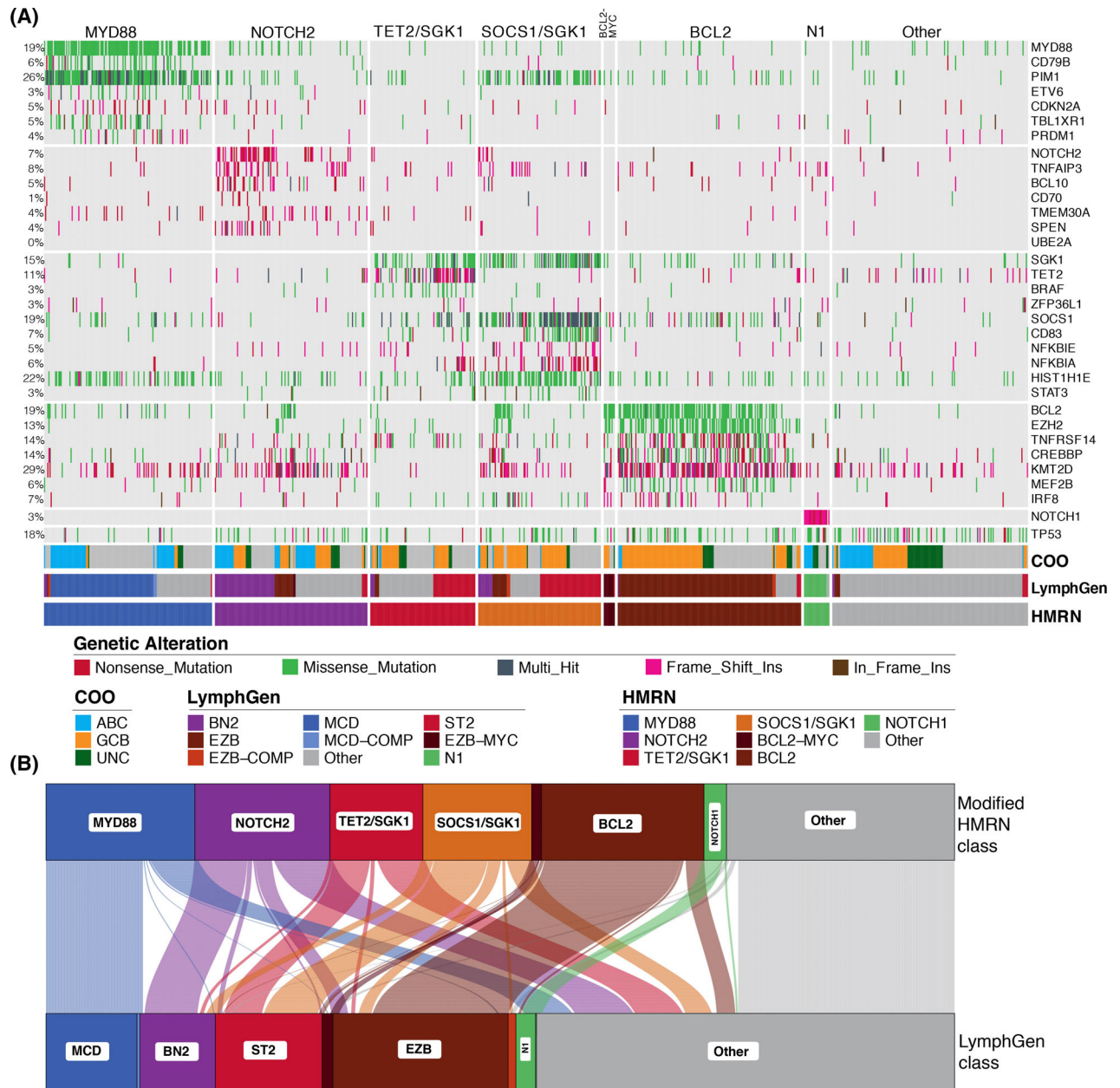


Fig 4. Comparison of the Haematological Malignancy Research Network (HMRN) and LymphGen classifications. (A) This oncoprint shows patients from the HMRN study.⁴⁰ The LymphGen classifier was applied to the same data, with the A53 option disabled due to lack of sufficient copy number information. LymphGen and HMRN classifications are shown in the lower bar. The HMRN classification was modified as described in Runge *et al.*⁴¹ to identify BCL2-MYC and NOTCH1 groups. (B) Alluvial plot showing comparison of modified HMRN and LymphGen classifications for individual cases from the HMRN study. It can be seen that cases classified by LymphGen are predominantly assigned to their equivalent HMRN group. The main distinctions between LymphGen and HMRN are that the latter further subdivides ST2 into the TET2/SGK1 and the SOCS1/SGK1 classes and has a higher classification, leaving fewer cases classified as ‘other’.

but not exclusive to this subtype. Beyond *TP53*, very few other coding mutations were enriched within this subtype and tumours with prototypical mutations characteristic of the other groups are commonly assigned A53 or composite classes including A53. Other than sharing aneuploidy resulting from the loss of *TP53*, it is unclear whether a unifying

biology exists within this group. When Harvard cases are reclassified using the LymphGen classifier it becomes clear that the A53 and C2 classes show much less overlap than do the other equivalent classes. The observation that many of the cases classified as A53 are reclassified to one of the other classes indicates that this classification may mask biology that

Table 1. Summary of key features of each molecular subtype.

LymphGen Cluster name	Mutated genes	LymphGen Wright <i>et al.</i> ³⁹	Harvard Chapuy <i>et al.</i> ³⁸	UK-HMRN Lacy <i>et al.</i> ⁴⁰	COO	Related lymphomas	Oncogenic pathways
MCD	<i>MYD88 (L265P)</i>	x	x	x	Almost all ABC	Extranodal lymphomas including PCNSL, primary testicular and primary breast lymphomas	B cell receptor signaling NFKB activation
	<i>CD79B</i>	x	x	x			Cell cycle dysregulation
	<i>PIM1</i>	x	x	x			MYC activation
	<i>ETV6</i>	x	x	x			Immune escape
	<i>CDKN2A mut/del</i>	x	x	x		Similar mutation profile to lymphoplasmacytic lymphoma	
	<i>BCL2 amp</i>	x	x	x			
	<i>TBL1XR1</i>	x	x	x			
	<i>PRDM1 mut/del</i>	x	x	x			
	<i>BCL2 translocation</i>	x	x	x			
	<i>BCL2</i>	x	x	x	Almost all GCB. Includes most cases with MHG or DHTTsig ⁺ gene expression signature	Closely related to follicular lymphoma	Histone modification BCL6 target repression PI3K activation
BN2	<i>EZH2</i>	x	x	x		Transformed FL and cases of DLBCL with a concurrent diagnosis of FL almost all fall into this cluster	
	<i>TNFRSF14</i>	x	x	x			
	<i>CREBBP</i>	x	x	x			
	<i>KMT2D</i>	x	x	x			
	<i>MEF2B</i>	x	x	x			
	<i>IRF8</i>	x	x	x			
	<i>BCL6 fusion</i>	x	x	x			
	<i>NOTCH2</i>	x	x	x	Most unclassified but includes ABC and some GCB cases	Genetic similarity to marginal-zone lymphoma, although no evidence of direct transformation from MZL	NFKB activation NOTCH2 activation Aberrant B cell development Immune evasion
	<i>TNFAIP3 mut/del</i>	x	x	x			
	<i>BCL10</i>	x	x	x			
ST2	<i>CD70</i>	x	x	x			
	<i>TMEM30A</i>	x	x	x			
	<i>SPEN</i>	x	x	x			
	<i>UBE2A</i>	x	x	x			
	<i>SGK1</i>	x	x	x	Almost all GCB	Similar mutations to nodular lymphocyte presomniant Hodgkin lymphoma	JAK/STAT activation PI3K activation ERK activation
	<i>TET2</i>	x	x	x		The SOCS1/SGK1 subtype shares many genetic features with primary mediastinal B-cell lymphoma. Most cases of PMBCL are classified into this cluster	
	<i>BRAF</i>	x	x	x			
	<i>ZFP36L1</i>	x	x	x			
	<i>SOCS1</i>	x	x	x			
	<i>CD83</i>	x	x	x			
<i>NFKBIE</i>	x	x	x				
<i>NIK6IA</i>	x	x	x				
<i>HIST1H1E</i>	x	x	x				
<i>STAT3</i>	x	x	x				

Table 1. (Continued)

LymphGen Cluster name	Mutated genes	LymphGen Wright <i>et al.</i> ³⁹	Harvard Chapuy <i>et al.</i> ³⁸	UK-HMRN Lacy <i>et al.</i> ⁴⁰	COO	Related lymphomas	Oncogenic pathways
A53	<i>TP53</i> <i>17p del</i>	x x	x x		No COO skew		Widespread aneuploidy
N1	<i>NOTCH1</i> <i>ID3</i>	x x			Mostly ABC	Similarities with Richter's transformation but no direct evidence that these cases represent transformed CLL	NOTCH activation Quiescence signature

The key genetic and gene expression signatures, likely oncogenic pathways and related lymphoid malignancies, are summarised for each of the molecular subtypes.

ABC, activated B cell; CLL, chronic lymphocytic leukaemia; COO, cell of origin; FL, follicular lymphoma; GCB, germinal-centre B cell; HMRN, haematological malignancy research network; MZL, marginal-zone lymphoma; NFKB, nuclear factor kappa B; PCNSL, primary central nervous system lymphoma.

would otherwise be revealed from other genetic features (Fig 3). This suggests that the genetic identity A53/C2 is the least robust of the genetic groups (Fig 3B).

The genetic heterogeneity within the diagnosis of DLBCL NOS suggests it represents an assortment of diseases. However, comparison with other lymphoma types suggests that the biology may straddle the boundaries between DLBCL NOS and other lymphoma types defined by the current World Health Organization classification system. Ultimately, a biology-focused classification of aggressive B-cell lymphoma may redefine many of these boundaries.

Implications for prognosis

The molecular classification systems discussed above are based principally upon grouping tumours with shared biology. However, a separate question is whether genetic classifications provide us with prognostic information that could be useful in guiding patient management. It is clear that clinical factors remain a dominant factor determining prognosis. It is also clear that clinical factors are not independently distributed across the genetic subtypes.⁴⁰ Thus, care must be taken when inferring the independent prognostic impact of each subtype.

Whilst genetic subgroups may show differential responses to targeted therapy in the future, all current information relates to conventional immunochemotherapy. Each of the studies discussed above examined the impact of genetic subtypes on patient outcome. It is important to acknowledge differences in the patient cohorts and the types of treatment used. The Harvard study reported clinical outcomes on 259 patients treated with R-CHOP-like therapy and derived from a combination of archived biopsy collections and the RICOVER-60 trial of elderly DLBCL. The NCI study reported clinical outcomes on 240 patients, enriched for ABC DLBCL, treated with R-CHOP-like therapy and derived from a combination of archived biopsy collections and the CALGB 50303 clinical trial. The HMRN study reported clinical data on 690 patients from the HMRN registry, which prospectively tracks outcomes of every new haematological malignancy diagnosis made at a regional diagnostic referral centre. The latter may escape the inevitable recruitment and selection biases implicit in clinical trial or pathological archives. Finally, distinct from the previous studies, the HMRN study also reported outcome for the 579 DLBCL patients treated with full-dose R-CHOP. Comparison to patients treated with R-CHOP-like regimens in the same study reveals the importance of this subtle distinction.⁴⁰ A summary of five-year overall survival (OS) outcomes by subtype across studies is shown in Fig 7.

When comparing across studies some clear conclusions can be drawn. First is the association of the ST2 subgroup with a favourable outcome. It was associated with superior survival in the NCI study (five-year OS 84%) and the Harvard study (five-year OS approximately 75%). In the HMRN

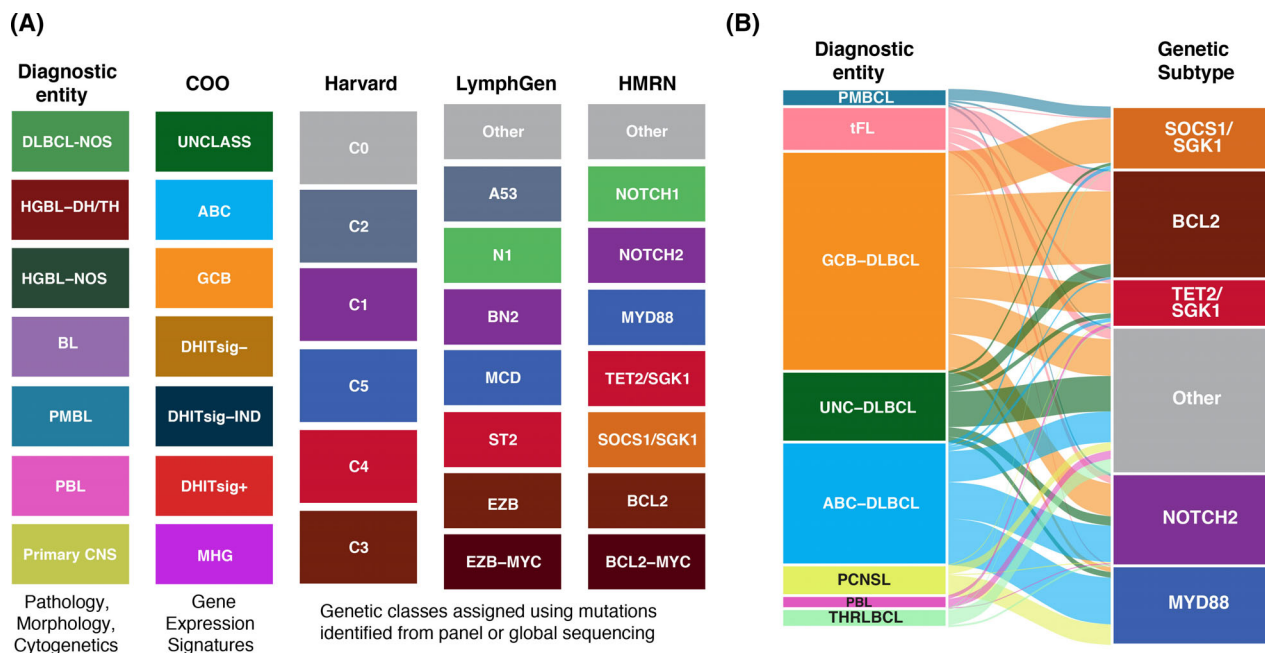


Fig 5. A guide and unified colour scheme for the plethora of diffuse large B-cell lymphoma (DLBCL) classification systems. (A) The diagnosis of DLBCL has grown progressively more complex as we continue to refine methods to delineate cases with distinct pathobiology. Several clinical entities that share features with DLBCL have been re-defined as distinct entities (column 1). The cell of origin (COO) system has been refined further to separate cases with gene expression profiles resembling double-hit lymphomas (DHITsig+) or Burkitt lymphoma (MHG) (Column 2). Columns 3–5 show the relationship between subgroups defined in the three systems discussed in the text. Each of the LymphGen [National Cancer Institute (NCI)] and revised Haematological Malignancy Research Network (HMRN) systems include a subgroup defined by activating mutations in *NOTCH1*. Both the NCI and Harvard systems (but not HMRN) accommodate aneuploid cases with (or without) *TP53* mutations. This figure uses a colour palette that was carefully designed to allow for the consistent representation of the main diagnostic entities, molecular and genetic classifications for DLBCL. The colour codes and utilities to allow the use of this palette are available as an open-source package at <https://github.com/morinlab/ggsci>. (B) Using this colour scheme, we have re-illustrated the data from the HMRN study to show the relationship between the lymphoma diagnostic entities and the HMRN genetic classes.

study, SOCS1/SGK1 was the subtype associated with the highest OS (five-year OS 80%) in R-CHOP-treated patients and the lowest hazard ratio, when adjusted for International Prognostic Index (IPI). Therefore, clear consensus exists around the favourable outcome in this genetic subtype. Conversely, consensus also exists around the poor outcome of patients with the N1 subtype; five-year OS was 27% in the NCI study³⁹ and 40% using a modified HMRN classification.⁴¹ Similarly poor survival was seen for EZB-MYC patients identified by expression profiling in the NCI study (five-year OS 48%)³⁹ or BCL2-MYC cases in the HMRN study (five-year OS 40%).⁴¹

The potential prognostic impact of the other groups is not as clear. EZB patients had an intermediate outcome (five-year OS 70%) in the NCI study, a good outcome in the HMRN study (five-year OS 82%) but one of the poorest survivals (five-year OS 60%) in the Harvard study (Fig 7). The MCD subtype had an extremely poor survival in the NCI study (five-year OS 40%). However, in the Harvard study the MCD equivalent (C5) showed outcomes identical to the EZB/C3 subgroup (five-year OS 60%). Interestingly, the HMRN study showed a poor outcome amongst curatively treated (R-CHOP-like) patients, but this effect was greatly reduced when the analysis was restricted to cases treated with

full-dose R-CHOP (Fig 7). Thus, the negative outcome in this subtype may in part reflect an overrepresentation of older or comorbid patients who are unable to tolerate full-dose treatment and are therefore generally excluded from clinical studies. The BN2 subgroup shows an intermediate outcome in the NCI study (five-year OS 67%), an excellent outcome in the Harvard study (five-year OS 80%), but a poor outcome in the HMRN study (five-year OS 55%). The A53 group showed an intermediate prognosis in both the NCI and Harvard studies (five-year OS 65%).

In summary, whilst the effect of some groups is clear (ST2, N1, EZB-MYC), the true prognostic impact of the other groups remains to be determined in prospective trials. However, the real value of a genetic classification system does not lie in its ability to predict response to R-CHOP. Rather it will be to identify homogeneous groups of tumours with shared biology that may respond similarly to specific targeted therapies.

Challenges of implementing molecular subtyping

Implementing a genomic classifier in clinical practice will be associated with a number of challenges. Important

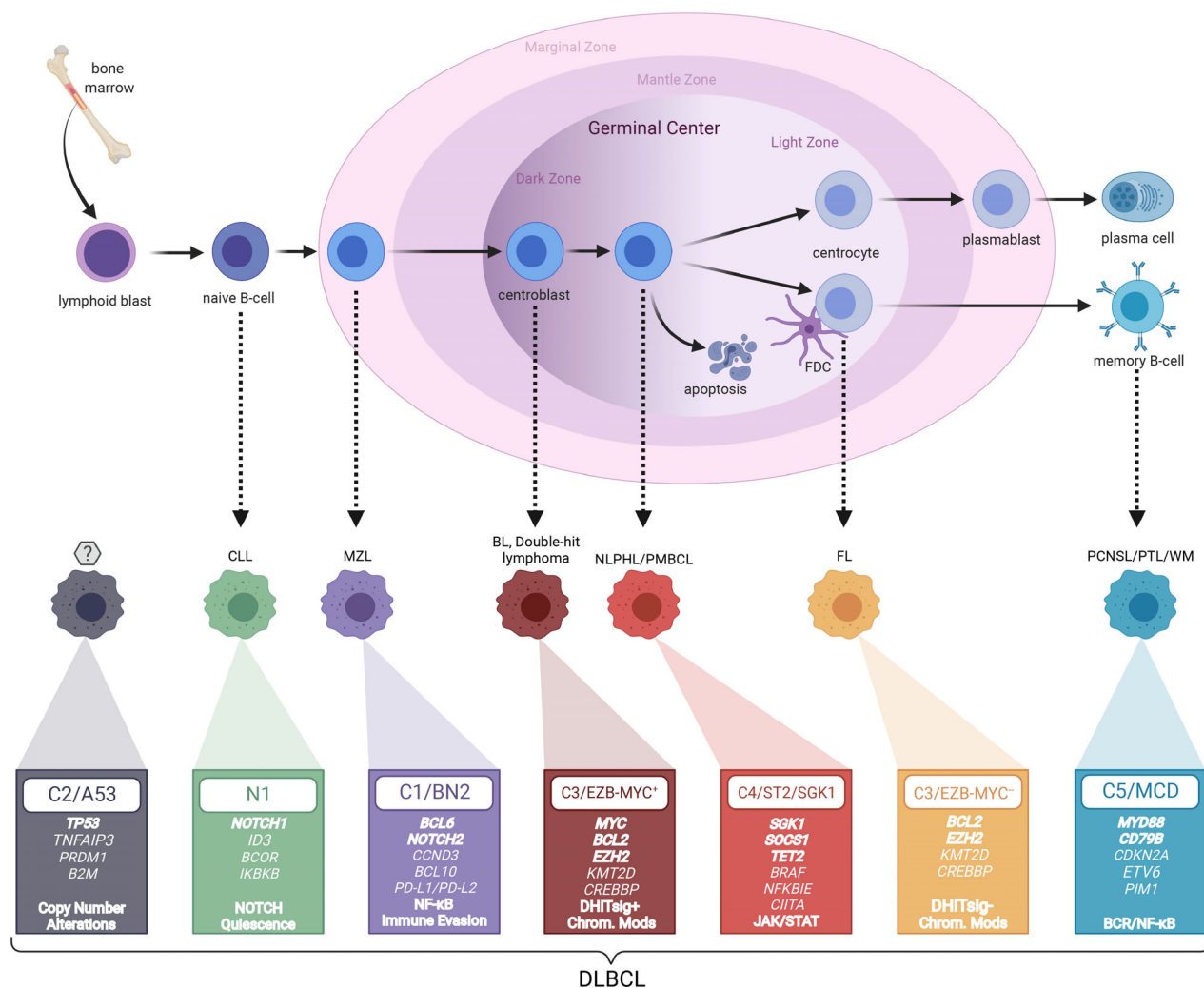


Fig 6. Distinct stages of B-cell differentiation hijacked by subtypes of diffuse large B-cell lymphoma (DLBCL) and related lymphoid malignancies. The progression of a B cell through the germinal centre is shown including lymphoma types that may originate from each of these stages of differentiation. The molecular subclassifications described by the LymphGen, Harvard and HMRN groups are shown, associated with the alternative lymphoid malignancies they most closely resemble. These include primary central nervous system lymphoma (PCNSL), primary testicular lymphoma (PTL), and Waldenström macroglobulinaemia (WM), which most likely derive from a precursor memory B cell and share features with the MCD subtype. The characteristics of BN2 cases suggest similarity to marginal zone lymphoma (MZL). EZB tumours recapitulate the genetics of follicular lymphoma (FL), and likely derive from light-zone centrocytes. In contrast, EZB-MYC⁺ cases most likely arise from dark-zone centroblasts. *NOTCH1* mutations, whilst rare in DLBCL, suggest a possible link to chronic lymphocytic leukaemia (CLL), with either a naïve B cell or precursor memory B cell (MBC) origin. ST2 cases have a genetic signature similar to nodular lymphocyte-predominant Hodgkin lymphoma (NLPHL) and primary mediastinal large B cell lymphoma (PMBCL) and arise from germinal-centre B cells. The biology and cellular origin of the A53/C2 subtype remain unclear. The most enriched gene mutations and gene expression signatures are indicated below each subtype.

considerations include the type of sequencing required and choice of classifier to be used (Fig 1). The NCI and Harvard studies used whole-exome sequencing to identify mutation and copy number alteration in a combination of fresh frozen and FFPE biopsies. The HMRN study used a targeted sequencing panel applied to FFPE biopsies, an approach that might be more suited to a diagnostic laboratory. Indeed, for all classifiers the vast majority of information comes from the genetic status of about 100–150 genes. However, the absence of genome-wide copy number precludes identification of the

A53 group. Targeted copy number assays or shallow whole-genome sequencing may ultimately become cost-efficient ways to identify the A53 group. However, for now, for individual research groups, the added value of identifying the A53 group (6.6% of patients) will need to be weighed against the significant increase in sequencing required for A53 identification.

The choice of classifier is also important. Initial studies relied upon clustering of large numbers of patients. Translating this to a single patient is now possible using the NCI

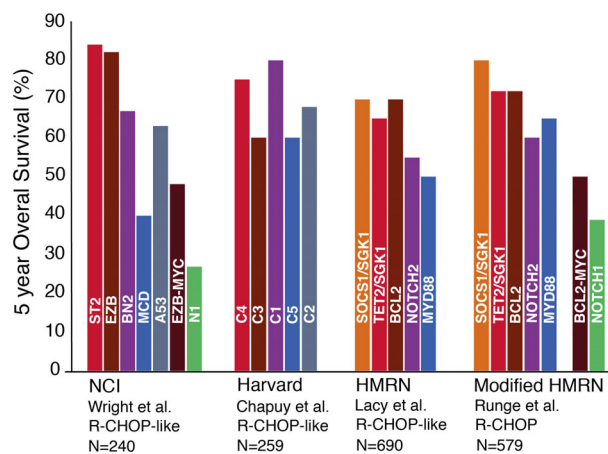


Fig 7. Comparison of survival outcomes by genetic subtype across studies. The five-year overall survival is shown for each subtype as described across their respective publications. Equivalent subtypes share the colour scheme described in Fig 5. The classification system used, the relevant published study, treatment received [(cyclophosphamide, doxorubicin, vincristine, prednisone + rituximab) (R-CHOP) *versus* R-CHOP-like], and number of patients contributing to survival analysis in each study are shown below the graph.

LymphGen classifier³⁹ or the code linked to the HMRN study.⁴⁰ Importantly the LymphGen classifier is designed to work with 'imperfect data' meaning it can be used for both targeted or whole-exome sequencing. A further classifier from the Harvard group has been described in abstract that relies on just 22 genetic features but is not publicly available. A recent publication applied the LymphGen classifier to cases classified in the HMRN study and showed high concordance with the original HMRN assignment.⁴¹ Indeed, the greatest source of variability was the number of cases classified at all rather than the movement between subtypes (Fig 5B). LymphGen assigned a unique classification to 53% of cases, whereas HMRN classified 73% of cases.⁴¹ In contrast, the Harvard classifier, according to their original clustering paper, assigns a classification to 96% of cases. These numbers are clearly very different and raise the important question of how to judge which provides the most correct answer. Whilst the general description of individual molecular subgroups is an area of agreement, precisely where to place the boundaries around them is not. In the absence of a gold standard reference this may prove challenging to resolve. The answer will ultimately come from the ability of a classifier to identify patient groups that respond to targeted therapies. Until consensus is reached it would seem prudent for clinical trials to capture as much genetic information as is reasonably practicable in order to employ both existing and future classifiers.

An important source of variability between studies is the strategy for calling genetic variants. Current classifications have been constructed based upon a binary call of mutant versus not-mutant for each gene. However, it is not always straightforward to determine the significance of an individual

mutation, especially in the absence of germline DNA. Whilst this is straightforward for driver genes with well-established hotspot codons like *MYD88* L265P, it is more challenging to interpret the significance of scattered missense mutations in genes that are subject to somatic hypermutation (SHM). Whilst not all of these have a lymphoma driver function it seems clear that patterns of hypermutated genes differ across subtypes, suggesting these mutations may represent useful classification markers.⁷³ This question of whether to report SHM as a genetic marker independent of a driver mutation role is approached differently in different studies. Examples of genes with widely different mutation frequencies include *SOCS1*, *TET2* and *BCL2* (Fig 3A). This most likely reflects different variant filtering strategies. Thus, different variant calling strategies may provide different outputs even when the same classifier is used. This may become a more significant problem as classifiers attempt to focus onto a smaller number of classifier genes, where the opportunity to spread the risk of classifying individual genes is reduced.

It remains likely that a significant number of cases will not fit into the current genetic subtypes and that further subtypes remain to be discovered. Given the number of cases of DLBCL that have been subjected to exome sequencing it seems less likely that new driver mutations will be discovered in protein-coding genes. However, the next wave of whole-genome sequencing may reveal previously undiscovered alterations in non-coding and regulatory regions. A recent example is the frequent mutation of the 3'UTR of *NFKBIZ*, leading to elevated protein expression in ABC DLBCL.⁷⁴ Further such discoveries may reveal new subgroups within the UC cases or may refine the current classification in a manner similar to how gene expression profiling has identified a MYC-driven subgroup of EZB. Advances in technology for proteomic quantification, assessment of host immune status or the involvement of viral pathogens in driving lymphomagenesis may all contribute new understanding to the challenge of defining molecular subtypes of DLBCL in coming years.

Finally, there are logistical considerations associated with the need for sufficient quantity and quality of biopsy tissue, and the ability to return sequencing data in a clinically meaningful timeframe. The recent UK REMoDL-B study overcame similar challenges for microarray assays and returned gene expression data within a three-week window for nationally recruited patients analysed at a central diagnostic laboratory.⁶ We believe a similar approach could be applied to genetic profiling. Furthermore, advances in circulating tumour DNA technology may ultimately allow DLBCL genotyping to be performed on a plasma sample.⁷⁵

Why now is the time to invest in genomic profiling for DLBCL

It is important to be clear about the potential value of any form of molecular profiling. Broadly there are three reasons

to do this: (i) to inform on prognosis; (ii) to allow selection of optimal therapy; and (iii) to provide a biology-based framework on which to design and interpret clinical trials. We have discussed how genomic profiling provides some, although limited, prognostic information on R-CHOP-treated DLBCL. Our current understanding of the underlying biology already allows hypotheses to be generated about which subgroups may respond to specific therapy. However, at present these remain hypotheses and it would be premature to suggest that genomics subtypes can currently be used to select therapy. Instead, the true value of genomics subtyping lies in the provision of a biology-based framework on which to design and interpret DLBCL trials. As discussed at the start of the review, the most significant barrier to the introduction of new and targeted therapies to first-line DLBCL is the considerable biological heterogeneity within the umbrella of DLBCL. The molecular subtypes discussed above allow the grouping of cases that share similar biology, depend on similar oncogenic pathways and therefore may respond in a similar way to targeted therapies. Stratifying patients into molecular subtypes will allow us to detect responses to therapy that may only be seen in a small but defined molecular subgroup. This may require that we rethink our clinical trial strategy. We envisage that future trials may follow an adaptive design where novel agents are initially screened across all DLBCL types but subsequently focused onto molecular subgroups where a potential response signal is observed. Our current biological understanding may allow us to narrow the therapeutic focus already; however there will always be surprises. An example of this is the REMoDL-B trial, which randomised first-line DLBCL patients to R-CHOP plus or minus proteasome inhibition with bortezomib based on the hypothesis that proteasome inhibition might benefit ABC DLBCL patients.⁶ Whilst improved outcome was not seen overall, or in the hypothesised subtype, a trend towards improved progression-free survival (PFS) was unexpectedly seen in the MHG subgroup identified by transcriptional profiling²⁷. This result should now be tested in a formal prospective trial in MHG patients.

As genomic profiling becomes standard in DLBCL we should heed the lessons learned from our experiences with COO and FISH profiling. This includes the need for a harmonised approach, and to resist the temptation for oversimplified assays that provide a poor proxy for the true classification. It also highlights the importance of prospective studies to determine the true prognostic impact, and for randomised studies to avoid the premature adoption of fashionable but unproven subtype-directed therapies.

We consider it is essential to now include molecular profiling in all prospective drug trials in DLBCL. Whether this should include comprehensive exome and RNA sequencing or a more focused sequencing strategy is a debate that will continue. But without some form of genomic testing to resolve individual molecular subtypes it is unlikely we will ever deploy novel therapies in DLBCL to their greatest

advantage. As our understanding of the genomics and biology of DLBCL progresses we can expect that classification systems will continue to evolve. The broadest possible molecular profiling may therefore future-proof trials against evolution in genetic subtyping. However, we should not let this delay starting. Indeed, one important aspect of introducing molecular profiling now will be to overcome the logistical and infrastructure barriers to returning genomic data in a clinically meaningful timeframe. To stand a chance of being effective, biologically-targeted therapies need to be deployed in a biologically targeted manner. Molecular subtyping of DLBCL is the next step in this precision medicine journey.

Acknowledgements

DJH is supported by a CRUK Senior Research Fellowship (RCCFEL100072). The Hodson laboratory receives core funding from Wellcome (203151/Z/16/Z) and MRC to the Wellcome-MRC Cambridge Stem Cell Institute, and from the CRUK Cambridge Centre (A25117) and the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014). RDM is a Scholar of the Michael Smith Foundation for Health Research.

Author contributions

RDM, SEA and DJH all contributed to the writing of this manuscript and approved the final version.

Conflicts of interest

DJH: Research funding from Gilead Sciences and Astra Zeneca. RDM: Consulting for Celgene. Royalties for patents for genetic subclassification of DLBCL.

References

- Smith A, Crouch S, Lax S, Li J, Painter D, Howell D, et al. Lymphoma incidence, survival and prevalence 2004–2014: sub-type analyses from the UK's Haematological Malignancy Research Network. *Br J Cancer*. 2015;**112**(9):1575–84.
- Coiffier B, Lepage E, Briere J, Herbrecht R, Tilly H, Bouabdallah R, et al. CHOP chemotherapy plus rituximab compared with CHOP alone in elderly patients with diffuse large-B-cell lymphoma. *N Engl J Med*. 2002;**346**(4):235–42.
- Vitolo U, Trnėný M, Belada D, Burke JM, Carella AM, Chua N, et al. Obinutuzumab or rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone in previously untreated diffuse large B-cell lymphoma. *J Clin Oncol*. 2017;**35**(31):3529–37.
- Crump M, Leppä S, Fayad L, Lee JJ, Di Rocco A, Ogura M, et al. Randomized, double-blind, phase III trial of enzastaurin versus placebo in patients achieving remission after first-line therapy for high-risk diffuse large B-cell lymphoma. *J Clin Oncol*. 2016;**34**(21):2484–92.
- Seymour JF, Pfreundschuh M, Trnėny M, Sehn LH, Catalano J, Csinady E, et al. R-CHOP with or without bevacizumab in patients with previously untreated diffuse large B-cell lymphoma: final MAIN study outcomes. *Haematologica*. 2014;**99**(8):1343–9.

6. Davies A, Cummin TE, Barrans S, Maishman T, Mamot C, Novak U, et al. Gene-expression profiling of bortezomib added to standard chemoimmunotherapy for diffuse large B-cell lymphoma (REMoDL-B): an open-label, randomised, phase 3 trial. *Lancet Oncol.* 2019;**20**(5):649–62.
7. Nowakowski GS, Chiappella A, Gascoyne RD, Scott DW, Zhang Q, Jurczak W, et al. ROBUST: a phase III study of lenalidomide plus R-CHOP versus placebo plus R-CHOP in previously untreated patients with ABC-type diffuse large B-cell lymphoma. *J Clin Oncol.* 2021;**39**(12):1317–28.
8. Younes A, Sehn LH, Johnson P, Zinzani PL, Hong X, Zhu J, et al. Randomized phase III trial of ibrutinib and rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone in non-germinal center B-cell diffuse large B-cell lymphoma. *J Clin Oncol.* 2019;**37**(15):1285–95.
9. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature.* 2000;**403**(6769):503–11.
10. Wright G, Tan B, Rosenwald A, Hurt EH, Wiestner A, Staudt LM. A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. *Proc Natl Acad Sci USA.* 2003;**100**(17):9991–6.
11. Scott DW, Mottok A, Ennishi D, Wright GW, Farinha P, Ben-Neriah S, et al. Prognostic significance of diffuse large B-cell lymphoma cell of origin determined by digital gene expression in formalin-fixed paraffin-embedded tissue biopsies. *J Clin Oncol.* 2015;**33**(26):2848–56.
12. Ahmed S, Glover P, Taylor J, Sha C, Care MA, Tooze R, et al. Comparative analysis of gene expression platforms for cell-of-origin classification of diffuse large B-cell lymphoma shows high concordance. *Br J Haematol.* 2021;**192**(3):599–604.
13. Care MA, Barrans S, Worrillow L, Jack A, Westhead DR, Tooze RM. A microarray platform-independent classification tool for cell of origin class allows comparative analysis of gene expression in diffuse large B-cell lymphoma. *PLoS One.* 2013;**8**(2):e55895.
14. Meyer PN, Fu K, Greiner TC, Smith LM, Delabie J, Gascoyne RD, et al. Immunohistochemical methods for predicting cell of origin and survival in patients with diffuse large B-cell lymphoma treated with rituximab. *J Clin Oncol.* 2011;**29**(2):200–7.
15. Read JA, Koff JL, Nastoupil LJ, Williams JN, Cohen JB, Flowers CR. Evaluating cell-of-origin subtype methods for predicting diffuse large B-cell lymphoma survival: a meta-analysis of gene expression profiling and immunohistochemistry algorithms. *Clin Lymphoma Myeloma Leuk.* 2014;**14**(6):460–7.e2.
16. Lenz G, Wright G, Dave SS, Xiao W, Powell J, Zhao H, et al. Stromal gene signatures in large-B-cell lymphomas. *N Engl J Med.* 2008;**359**(22):2313–23.
17. Shaffer AL 3rd, Young RM, Staudt LM. Pathogenesis of human B cell lymphomas. *Annu Rev Immunol.* 2012;**30**:565–610.
18. Swerdlow SH, Campo E, Pileri SA, Harris NL, Stein H, Siebert R, et al. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood.* 2016;**127**(20):2375–90.
19. Johnson NA, Savage KJ, Ludkovski O, Ben-Neriah S, Woods R, Steidl C, et al. Lymphomas with concurrent BCL2 and MYC translocations: the critical factors associated with survival. *Blood.* 2009;**114**(11):2273–9.
20. Niitsu N, Okamoto M, Miura I, Hirano M. Clinical features and prognosis of de novo diffuse large B-cell lymphoma with t(14;18) and 8q24/c-MYC translocations. *Leukemia.* 2009;**23**(4):777–83.
21. Copie-Bergman C, Cuillière-Dartigues P, Baia M, Briere J, Delarue R, Canioni D, et al. MYC-IG rearrangements are negative predictors of survival in DLBCL patients treated with immunochemotherapy: a GELA/LYSA study. *Blood.* 2015;**126**(22):2466–74.
22. Rosenwald A, Bens S, Advani R, Barrans S, Copie-Bergman C, Elsensohn M-H, et al. Prognostic significance of MYC rearrangement and translocation partner in diffuse large B-cell lymphoma: a study by the Lunenburg Lymphoma Biomarker Consortium. *J Clin Oncol.* 2019;**37**(35):3359–68.
23. Johnson SM, Umakanthan JM, Yuan JJ, Fedoriv Y, Bociek RG, Kaiser-Rogers K, et al. Lymphomas with pseudo-double-hit BCL6-MYC translocations due to t(3;8)(q27;q24) are associated with a germinal center immunophenotype, extranodal involvement, and frequent BCL2 translocations. *Hum Pathol.* 2018;**80**:192–200.
24. Petrich AM, Gandhi M, Jovanovic B, Castillo JJ, Rajguru S, Yang DT, et al. Impact of induction regimen and stem cell transplantation on outcomes in double-hit lymphoma: a multicenter retrospective analysis. *Blood.* 2014;**124**(15):2354–61.
25. Oki Y, Noorani M, Lin P, Davis RE, Neelapu SS, Ma L, et al. Double hit lymphoma: the MD Anderson Cancer Center clinical experience. *Br J Haematol.* 2014;**166**(6):891–901.
26. Ennishi D, Jiang A, Boyle M, Collinge B, Grande BM, Ben-Neriah S, et al. Double-hit gene expression signature defines a distinct subgroup of germinal center B-cell-like diffuse large B-cell lymphoma. *J Clin Oncol.* 2019;**37**(3):190–201.
27. Sha C, Barrans S, Cucco F, Bentley MA, Care MA, Cummin T, et al. Molecular high-grade B-cell lymphoma: defining a poor-risk group that requires different approaches to therapy. *J Clin Oncol.* 2019;**37**(3):202–12.
28. Hilton LK, Tang J, Ben-Neriah S, Alcaide M, Jiang A, Grande BM, et al. The double-hit signature identifies double-hit diffuse large B-cell lymphoma with genetic events cryptic to FISH. *Blood.* 2019;**134**(18):1528–32.
29. Morin RD, Mendez-Lago M, Mungall AJ, Goya R, Mungall KL, Corbett RD, et al. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature.* 2011;**476**(7360):298–303.
30. Pasqualucci L, Trifonov V, Fabbri G, Ma J, Rossi D, Chiarenza A, et al. Analysis of the coding genome of diffuse large B-cell lymphoma. *Nat Genet.* 2011;**43**(9):830–7.
31. Lohr JG, Stojanov P, Lawrence MS, Auclair D, Chapuy B, Sougnez C, et al. Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. *Proc Natl Acad Sci USA.* 2012;**109**(10):3879–84.
32. Zhang J, Grubor V, Love CL, Banerjee A, Richards KL, Mieczkowski PA, et al. Genetic heterogeneity of diffuse large B-cell lymphoma. *Proc Natl Acad Sci USA.* 2013;**110**(4):1398–403.
33. Morin RD, Johnson NA, Severson TM, Mungall AJ, An J, Goya R, et al. Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nat Genet.* 2010;**42**(2):181–5.
34. Young RM, Staudt LM. Targeting pathological B cell receptor signalling in lymphoid malignancies. *Nat Rev Drug Discov.* 2013;**12**(3):229–43.
35. Jiang Y, Dominguez PM, Melnick AM. The many layers of epigenetic dysfunction in B-cell lymphomas. *Curr Opin Hematol.* 2016;**23**(4):377–84.
36. Reddy A, Zhang J, Davis NS, Moffitt AB, Love CL, Waldrop A, et al. Genetic and functional drivers of diffuse large B cell lymphoma. *Cell.* 2017;**171**(2):481–94.e15.
37. Schmitz R, Wright GW, Huang DW, Johnson CA, Phelan JD, Wang JQ, et al. Genetics and pathogenesis of diffuse large B-cell lymphoma. *N Engl J Med.* 2018;**378**(15):1396–407.
38. Chapuy B, Stewart C, Dunford AJ, Kim J, Kamburov A, Redd RA, et al. Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. *Nat Med.* 2018;**24**(5):679–90.
39. Wright GW, Huang DW, Phelan JD, Coulibaly ZA, Roulland S, Young RM, et al. A probabilistic classification tool for genetic subtypes of diffuse large B cell lymphoma with therapeutic implications. *Cancer Cell.* 2020;**37**(4):551–68.e14.
40. Lacy SE, Barrans SL, Beer PA, Painter D, Smith AG, Roman E, et al. Targeted sequencing in DLBCL, molecular subtypes, and outcomes: a Haematological Malignancy Research Network report. *Blood.* 2020;**135**(20):1759–71.
41. Runge HFP, Lacy S, Barrans S, Beer PA, Painter D, Smith A, et al. Application of the LymphGen classification tool to 928 clinically and genetically-characterised cases of diffuse large B cell lymphoma (DLBCL). *Br J Haematol.* 2021;**192**(1):216–20.
42. Davis RE, Ngo VN, Lenz G, Tolar P, Young RM, Romesser PB, et al. Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma. *Nature.* 2010;**463**(7277):88–92.

43. Ngo VN, Young RM, Schmitz R, Jhavar S, Xiao W, Lim K-H, et al. Oncogenically active MYD88 mutations in human lymphoma. *Nature*. 2011;**470**(7332):115–9.
44. Phelan JD, Young RM, Webster DE, Roulland S, Wright GW, Kasbekar M, et al. A multiprotein supercomplex controlling oncogenic signalling in lymphoma. *Nature*. 2018;**560**(7718):387–91.
45. Chapuy B, Roemer MGM, Stewart C, Tan Y, Abo RP, Zhang L, et al. Targetable genetic features of primary testicular and primary central nervous system lymphomas. *Blood*. 2016;**127**(7):869–81.
46. Cao XX, Li J, Cai H, Zhang W, Duan MH, Zhou DB. Patients with primary breast and primary female genital tract diffuse large B cell lymphoma have a high frequency of MYD88 and CD79B mutations. *Ann Hematol*. 2017;**96**(11):1867–71.
47. Taniguchi K, Takata K, Chuang S-S, Miyata-Takata T, Sato Y, Satou A, et al. Frequent MYD88 L265P and CD79B mutations in primary breast diffuse large B-cell lymphoma. *Am J Surg Pathol*. 2016;**40**(3):324–34.
48. Holmes AB, Corinaldesi C, Shen Q, Kumar R, Compagno N, Wang Z, et al. Single-cell analysis of germinal-center B cells informs on lymphoma cell of origin and outcome. *J Exp Med*. 2020;**217**(10). <https://doi.org/10.1084/jem.20200483>.
49. Venturutti L, Teater M, Zhai A, Chadburn A, Babiker L, Kim D, et al. TBL1XR1 mutations drive extranodal lymphoma by inducing a protumorigenic memory fate. *Cell*. 2020;**182**(2):297–316.e27.
50. Zhang J, Dominguez-Sola D, Hussein S, Lee J-E, Holmes AB, Bansal M, et al. Disruption of KMT2D perturbs germinal center B cell development and promotes lymphomagenesis. *Nat Med*. 2015;**21**(10):1190–8.
51. Zhang J, Vlasevska S, Wells VA, Nataraj S, Holmes AB, Duval R, et al. The CREBBP acetyltransferase is a haploinsufficient tumor suppressor in B-cell lymphoma. *Cancer Discov*. 2017;**7**(3):322–37.
52. Béguelin W, Popovic R, Teater M, Jiang Y, Bunting K, Rosen M, et al. EZH2 is required for germinal center formation and somatic EZH2 mutations promote lymphoid transformation. *Cancer Cell*. 2013;**23**(5):677–92.
53. Ortega-Molina A, Boss IW, Canela A, Pan H, Jiang Y, Zhao C, et al. The histone lysine methyltransferase KMT2D sustains a gene expression program that represses B cell lymphoma development. *Nat Med*. 2015;**21**(10):1199–208.
54. Jiang Y, Ortega-Molina A, Geng H, Ying HY, Hatzl K, Parsa S, et al. crebbp inactivation promotes the development of HDAC3-dependent lymphomas. *Cancer Discov*. 2017;**7**(1):38–53.
55. Caganova M, Carrisi C, Varano G, Mainoldi F, Zanardi F, Germain P-L, et al. Germinal center dysregulation by histone methyltransferase EZH2 promotes lymphomagenesis. *J Clin Invest* 2013;**123**(12):5009–22.
56. García-Ramírez I, Tadros S, González-Herrero I, Martín-Lorenzo A, Rodríguez-Hernández G, Moore D, et al. Crebbp loss cooperates with Bcl2 overexpression to promote lymphoma in mice. *Blood*. 2017;**129**(19):2645–56.
57. Horton SJ, Giotopoulos G, Yun H, Vohra S, Sheppard O, Bashford-Rogers R, et al. Early loss of Crebbp confers malignant stem cell properties on lymphoid progenitors. *Nat Cell Biol*. 2017;**19**(9):1093–104.
58. Bődör C, Grossmann V, Popov N, Okosun J, O’Riain C, Tan K, et al. EZH2 mutations are frequent and represent an early event in follicular lymphoma. *Blood*. 2013;**122**(18):3165–8.
59. Okosun J, Bődör C, Wang J, Araf S, Yang C-Y, Pan C, et al. Integrated genomic analysis identifies recurrent mutations and evolution patterns driving the initiation and progression of follicular lymphoma. *Nat Genet*. 2014;**46**(2):176–81.
60. Pasqualucci L, Dominguez-Sola D, Chiarenza A, Fabbri G, Grunn A, Trifonov V, et al. Inactivating mutations of acetyltransferase genes in B-cell lymphoma. *Nature*. 2011;**471**(7337):189–95.
61. Cucco F, Barrans S, Sha C, Clipson A, Crouch S, Dobson R, et al. Distinct genetic changes reveal evolutionary history and heterogeneous molecular grade of DLBCL with MYC/BCL2 double-hit. *Leukemia*. 2020;**34**(5):1329–41.
62. Arthur SE, Gao J, Healy S, Rushton CK, Thomas N, Hilton LK, et al. Non-coding NFKBIZ 3’ UTR mutations promote cell growth and resistance to targeted therapeutics in diffuse large B-cell lymphoma. *bioRxiv*. 2021. <https://doi.org/10.1101/2021.05.22.445261>.
63. Rossi D, Trifonov V, Fangazio M, Bruscazzin A, Rasi S, Spina V, et al. The coding genome of splenic marginal zone lymphoma: activation of NOTCH2 and other pathways regulating marginal zone development. *J Exp Med*. 2012;**209**(9):1537–51.
64. Spina V, Khiabani H, Messina M, Monti S, Cascione L, Bruscazzin A, et al. The genetics of nodal marginal zone lymphoma. *Blood*. 2016;**128**(10):1362–73.
65. Hartmann S, Schuhmacher B, Rausch T, Fuller L, Döring C, Weniger M, et al. Highly recurrent mutations of SGK1, DUSP2 and JUNB in nodular lymphocyte predominant Hodgkin lymphoma. *Leukemia*. 2016;**30**(4):844–53.
66. Gao J, Sidiropoulou E, Walker I, Krupka JA, Mizielinski K, Usheva Z, et al. SGK1 mutations in DLBCL generate hyperstable protein neoisoforms that promote AKT independence. *Blood*. 2021. <https://doi.org/10.1182/blood.2020010432>.
67. Mottok A, Hung SS, Chavez EA, Woolcock B, Telenius A, Chong LC, et al. Integrative genomic analysis identifies key pathogenic mechanisms in primary mediastinal large B-cell lymphoma. *Blood*. 2019;**134**(10):802–13.
68. Chapuy B, Stewart C, Dunford AJ, Kim J, Wienand K, Kamburov A, et al. Genomic analyses of PMBL reveal new drivers and mechanisms of sensitivity to PD-1 blockade. *Blood*. 2019;**134**(26):2369–82.
69. Yuan JI, Wright G, Rosenwald A, Steidl C, Gascoyne RD, Connors JM, et al. Identification of primary mediastinal large b-cell lymphoma at non-mediastinal sites by gene expression profiling. *Am J Surg Pathol*. 2015;**39**(10):1322–30.
70. Duns G, Viganò E, Ennishi D, Sarkozy C, Hung SS, Chavez E, et al. Characterization of DLBCL with a PMBL gene expression signature. *Blood*. 2021;**138**(2):136–148.
71. Fabbri G, Rasi S, Rossi D, Trifonov V, Khiabani H, Ma J, et al. Analysis of the chronic lymphocytic leukemia coding genome: role of NOTCH1 mutational activation. *J Exp Med*. 2011;**208**(7):1389–401.
72. Rossi D, Rasi S, Fabbri G, Spina V, Fangazio M, Forconi F, et al. Mutations of NOTCH1 are an independent predictor of survival in chronic lymphocytic leukemia. *Blood*. 2012;**119**(2):521–9.
73. Alkodsí A, Cervera A, Zhang K, Louhimo R, Meriranta L, Pasanen A, et al. Distinct subtypes of diffuse large B-cell lymphoma defined by hypermutated genes. *Leukemia*. 2019;**33**(11):2662–72.
74. Arthur SE, Jiang A, Grande BM, Alcaide M, Cojocaru R, Rushton CK, et al. Genome-wide discovery of somatic regulatory variants in diffuse large B-cell lymphoma. *Nat Commun*. 2018;**9**(1):4001.
75. Scherer F, Kurtz DM, Newman AM, Stehr H, Craig AFM, Esfahani MS, et al. Distinct biological subtypes and patterns of genome evolution in lymphoma revealed by circulating tumor DNA. *Sci Transl Med*. 2016;**8**(364):364ra155.