Application of the LymphGen classification tool to 928 clinically and genetically-characterised cases of diffuse large B cell lymphoma (DLBCL).

Hendrik FP Runge[1,2], Stuart Lacy[3], Sharon Barrans[4], Philip A Beer[5], Daniel Painter[3], Alexandra Smith[3], Eve Roman[3], Cathy Burton[4], Simon Crouch[3], Reuben Tooze[4,6] and Daniel J Hodson[1,2].

1. Wellcome-MRC Cambridge Stem Cell Institute, Jeffrey Cheah Biomedical Centre, Cambridge, CB2 0AW, UK.
2. Department of Haematology, University of Cambridge, Cambridge, CB2 0AW, UK.
3. Epidemiology and Cancer Statistics Group, Department of Health Sciences, University of York, York, YO10 5DD, United Kingdom.
4. Haematological Malignancy Diagnostic Service, St. James's Institute of Oncology, Leeds, LS9 7TF, United Kingdom.
5. Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1SA, United Kingdom.
6. Section of Experimental Haematology, Leeds Institute of Molecular Medicine, University of Leeds, LS2 9JT, United Kingdom.

Address for correspondence:

Dr Daniel Hodson

Wellcome-MRC Cambridge Stem Cell Institute
Jeffrey Cheah Biomedical Centre
Cambridge
CB2 0AW
UK

Phone: 01223 762120
Fax: 01223 336828
Email: djh1002@cam.ac.uk

Application of the LymphGen classification tool to 928 genetically-characterised cases of diffuse large B cell lymphoma (DLBCL).

We recently published results of targeted sequencing applied to 928 unselected cases of DLBCL registered in the Haematological Malignancy Research Network (HMRN) registry (1). Clustering allowed us to resolve five genomic subtypes. These subtypes shared considerable overlap with those proposed in two independent genomic studies(2, 3), suggesting the potential to use genetics to stratify patients by both risk and biology. In the original studies, clustering techniques were applied to sample cohorts to reveal molecular substructure, but left open the challenge of how to classify an individual patient. This was addressed by the LymphGen classification tool (4). LymphGen assigns an individual case to one of six molecular subtypes. The tool accommodates data from exome or targeted sequencing, either with or without copy number variant (CNV) data. Separate gene expression data allows classification of a seventh, MYC-driven subtype defined by a double hit (DHL) or molecular high-grade (MHG) gene expression signature(5-7).

Our large cohort of unselected registry patients, with comprehensive clinical and molecular annotation, provides an opportunity to examine the prognostic implications of the LymphGen classifier and to compare the robustness of cluster assignment across studies. Our sequencing panel provided only limited CNV data; therefore, we chose to enter exclusively mutation data. The LymphGen tool is able to accommodate mutation-only data, recognising that without CNV data the A53 subtype cannot be identified. We previously saw a strong negative prognostic effect of truncating *NOTCH1* mutation(1); we therefore modified our original classification to annotate all patients with truncating exon-34 *NOTCH1* mutation as a distinct subgroup. We compared the classification assigned by our own clustering to that assigned by the LymphGen classifier.

Our original clustering assigned a molecular subtype to 73% of cases. LymphGen assigned a unique classification in 53% (489 cases) (**Figure 1A and Supplementary Table 1**). 46% remained unclassified. 1% were assigned to overlapping categories of uncertain significance. We restricted further analysis to the 477 cases confidently classified in both our study and by the LymphGen classifier to establish the extent of agreement at the level of individual samples (**Figure 1B**).

We saw strong consensus amongst these cases, with 86% classified to the analogous LymphGen subtype (**Figure 1B**). In particular, we saw 95% overlap between MYD88 & MCD subgroups and 96% overlap between BCL2 and EZB subgroups. Our SOCS1/SGK1 and TET2/SGK1 clusters represented subdivisions of the ST2 cluster with 89% of ST2 cases corresponding to one of these subgroups. This considerable overlap between separate classification strategies, identified using independent statistical approaches, demonstrates the robust reproducibility amongst the "core" members of these molecular subtypes. However, 47% of our patients did not receive a unique LymphGen classification. In part, this may relate to the lack of CNV data, precluding A53 identification. However, the A53 group represented only 7% of cases in the LymphGen study. Accordingly, even with full CNV data the original LymphGen publication classified only 57% of cases. In contrast, the original Chapuy assigned a classification to 96% of patients. Taken together, we conclude that analogous subgroups identified across studies represent the same robust, biological entities

but that different classifications tolerate differing thresholds of uncertainty when assigning a subtype (**Figure 1C**). That is to say, the main variation between classifications is whether a case is classified at all, rather than the movement of confidently classified cases from one subgroup to another.

We then looked at the prognostic implication of the LymphGen classifier in our cohort of patients. Strengths of our registry cohort include the large patient number, meticulous clinical annotation and comprehensive enrolment of every DLBCL diagnosis without confounding referral bias. The LymphGen classifier suggests use of gene expression to identify a MYC-driven subgroup of the EZB cluster. Since gene expression was not available for every patient, and in an attempt to probe the utility of a mutation-only strategy, we took advantage of a recent observation that *MYC* mutations at codons 57-60 associate strongly with *MYC*-rearranged or MHG DLBCL(6, 8). We used the presence of these mutations to define a MYC-driven subgroup of the EZB cluster.

Our previous analysis emphasised the importance of considering prognostic impact in homogeneously treated patients(1). Therefore, we restricted our analysis to patients receiving full dose R-CHOP. We excluded patients treated with regimens considered R-CHOP-like, who frequently received considerably attenuated chemotherapy and were not equally distributed across genomic subtypes(1)(**Supplementary Table 2**).

We saw poor survival amongst patients assigned to the N1 group, a finding consistent across studies (**Figure 2A,B**). The MYC-EZB subgroup was also associated with poor survival, consistent with *MYC*/*BCL2* rearranged DLBCL(9). In contrast, the ST2 subgroup was associated with favourable outcome. However, the prognostic impact of the remaining subtypes (MCD, EZB, BN2) did not achieve significance in R-CHOP treated cases. Unclassified cases had an intermediate survival (**Figure 2A,B**). Comparison with the international prognostic index (IPI) suggests that clinical factors remain a dominant determinant of survival in DLBCL but that genetic classification provides independent prognostic information over and above the IPI (**Figure 2C,D & Supplementary Table 2**).

In summary, we conclude that mutation-only data from targeted sequencing allows a confident LymphGen classification in just over half of patients. These cases show strong consensus across different classification strategies, reinforcing the robust reproducibility of the core disease subgroups. Identification of the A53 subgroup will require either exome data or a panel specifically designed to provide the required CNV data. Both N1 and MYC-EZB, were associated with markedly inferior prognosis, whilst ST2 showed consistently favourable outcome. We did not observe significant prognostic impact from MCD, EZB and BN2 subgroups in R-CHOP treated patients. Nevertheless, the greatest potential of this classification will be to allow biological stratification of a disease where genetic heterogeneity will otherwise stymie our ability to assess the benefit of biologically targeted therapy, where efficacy may be restricted to specific biological subtypes. Whilst knowledge of the molecular subtype may not yet define the optimal therapy for an individual patient it will allow us to design and interpret clinical trials of these agents in the future.

Acknowledgements

HR was funded by a studentship from the Medical Research Council. DH was supported by a Clinician Scientist Fellowship from the Medical Research Council (MR/M008584/1). The Hodson laboratory receives core funding from Wellcome and MRC to the Wellcome-MRC Cambridge Stem Cell Institute and core funding from the CRUK Cambridge Cancer Centre. HMRN is supported by BCUK 15037 and CRUK 18362.

Author Contributions

HR performed data analysis. SC performed statistical analysis. SL, SB, PB, DP, AS, ER, CB, RT discussed and interpreted data. DH wrote the manuscript with input from HR. All authors reviewed and approved the manuscript before submission.

Declaration of Interest

PB: consultancy for Karus Therapeutics (Oxford, UK), OncoDNA (Gosselies, Belgium) and Everything Genetic (London, UK).

References

1. 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.
2. 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.
3. 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.
4. 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.
5. 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.
6. 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.
7. Painter D, Barrans S, Lacy S, Smith A, Crouch S, Westhead D, et al. Cell-of-origin in diffuse large B-cell lymphoma: findings from the UK's population-based Haematological Malignancy Research Network. Br J Haematol. 2019;185(4):781-4.
8. 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.
9. 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. Journal of Clinical Oncology. 2019;37(35):3359-68.

Figure Legends

Figure 1
A) Sankey plots showing comparison of HMRN and LymphGen classifications in all 928 cases of DLBCL or **B**) in the 477 cases for whom a unique molecular subtype was assigned by both classifications. **C**) A conceptual, schematic diagram of the molecular classification of DLBCL. The genes most frequently altered in each subtype are indicated. Colour intensity reflects the classification confidence. The LymphGen cluster names and terms used in HMRN and Chapuy publications are indicated in red, black and blue respectively.

Figure 2
A-C) Overall and progression free survival for R-CHOP treated DLBCL patients stratified by **A**) LymphGen classification, **B**) modified HMRN classification (incorporating NOTCH1 and BCL2-MYC subgroups), or **C**) by IPI score. The outcome of unclassified patients (labelled "Other") is shown in grey in **A** and **B**. Statistical analysis in **A** shows log rank p-value for the indicated comparisons. **D**) IPI-adjusted hazard ratios for classified patients treated with R-CHOP are shown relative to the unclassified "Other" patients. Hazard ratios were calculated using a Cox model based on either OS or PFS as indicated.

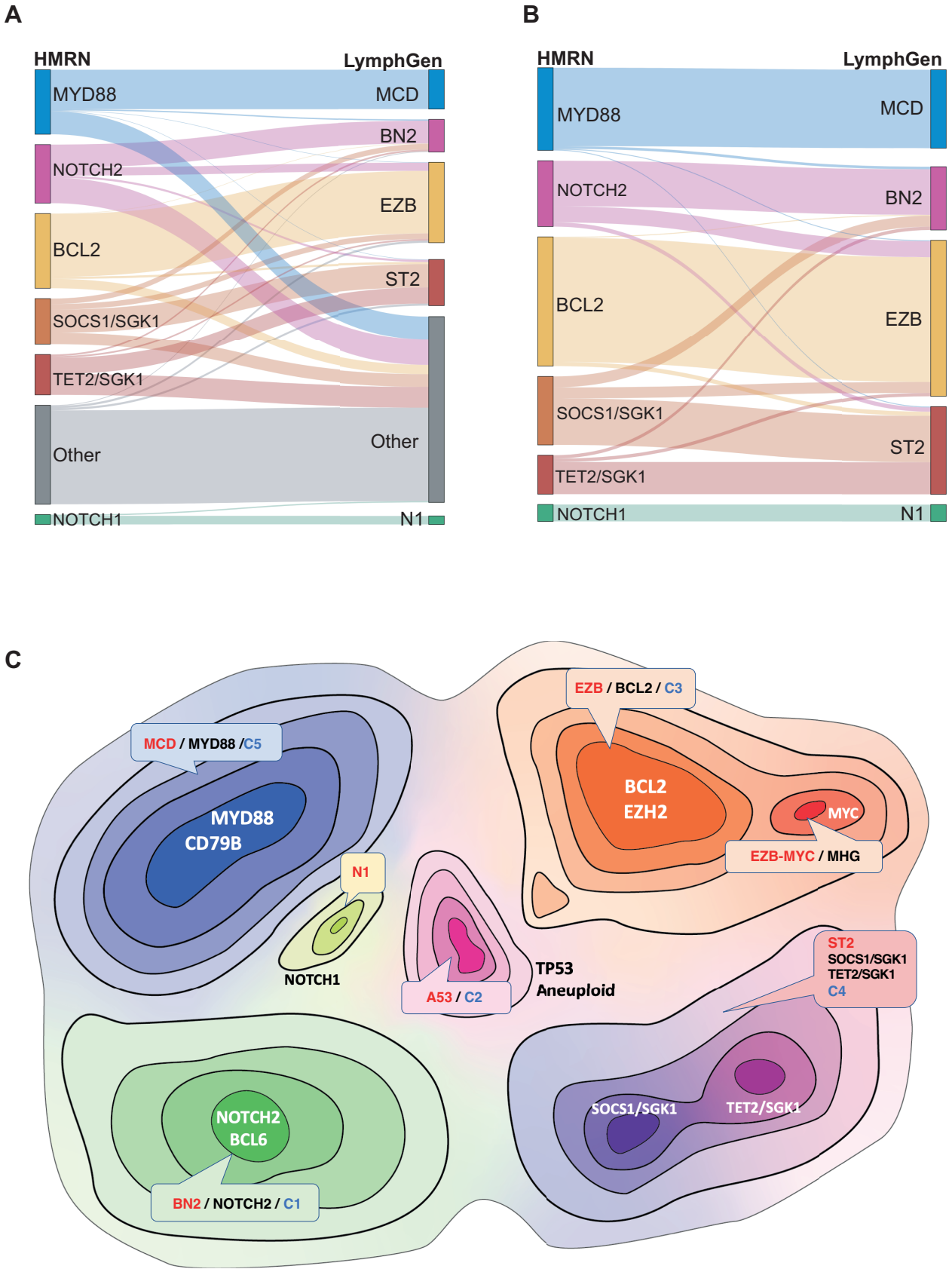
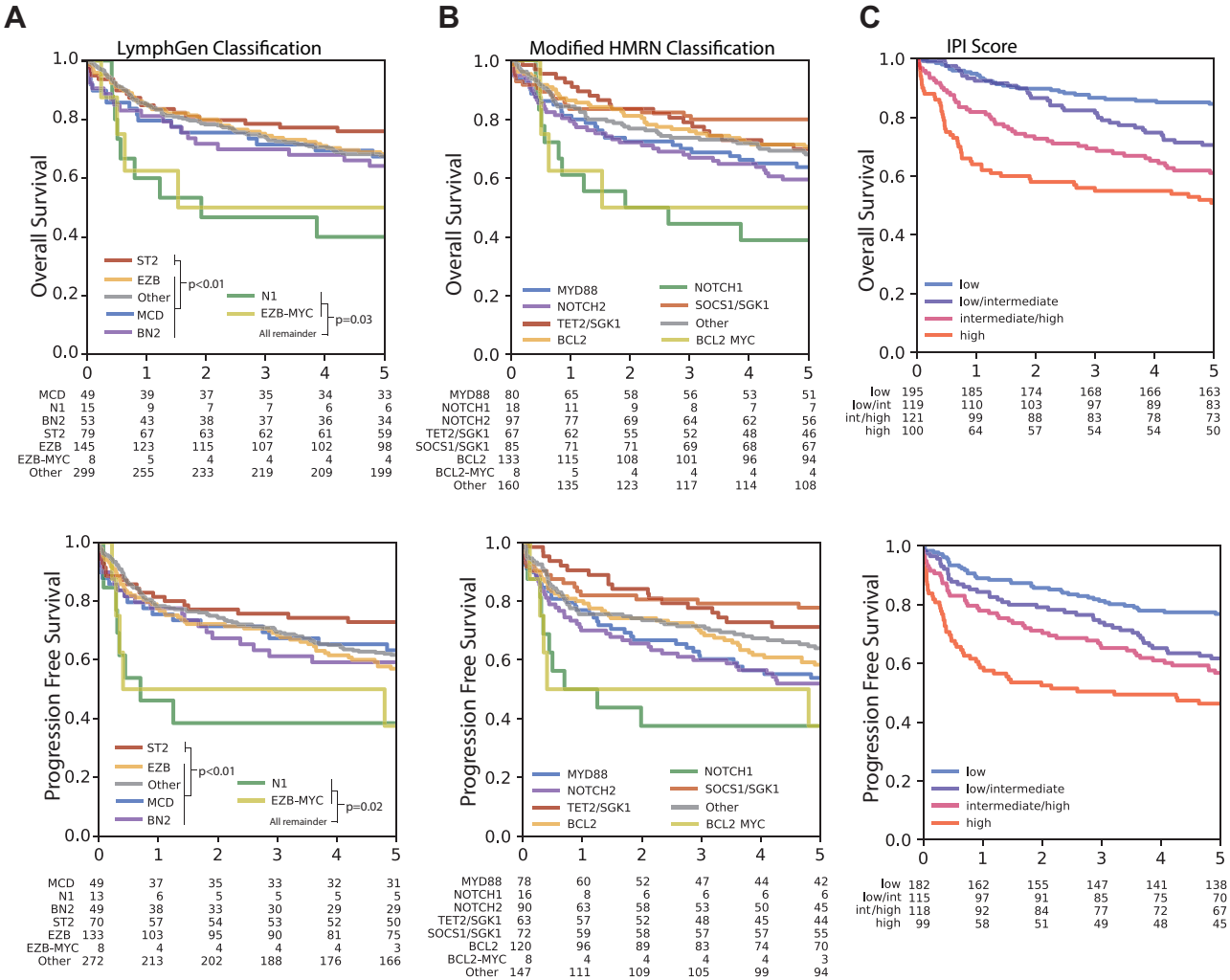

A
HMRN
LymphGen
MYD88
NOTCH2
BCL2
SOCS1/SGK1
TET2/SGK1
Other
NOTCH1
MCD
BN2
EZB
ST2
Other
N1
B
HMRN
LymphGen
MYD88
NOTCH2
BCL2
SOCS1/SGK1
TET2/SGK1
NOTCH1
MCD
BN2
EZB
ST2
N1
C
MCD / MYD88 /C5
MYD88
CD79B
N1
NOTCH1
EZB / BCL2 / C3
BCL2
EZH2
MYC
EZB-MYC / MHG
A53 / C2
TP53
Aneuploid
ST2
SOCS1/SGK1
TET2/SGK1
C4
SOCS1/SGK1
TET2/SGK1
NOTCH2
BCL6
BN2 / NOTCH2 / C1

Figure 1


A
LymphGen Classification
Overall Survival
ST2
EZB
Other
MCD
BN2
p<0.01
N1
EZB-MYC
All remainder
p=0.03
B
Modified HMRN Classification
Overall Survival
MYD88
NOTCH2
TET2/SGK1
BCL2
NOTCH1
SOCS1/SGK1
Other
BCL2 MYC
C
IPI Score
Overall Survival
low
low/intermediate
intermediate/high
high
Progression Free Survival
p<0.01
p=0.02
All remainder

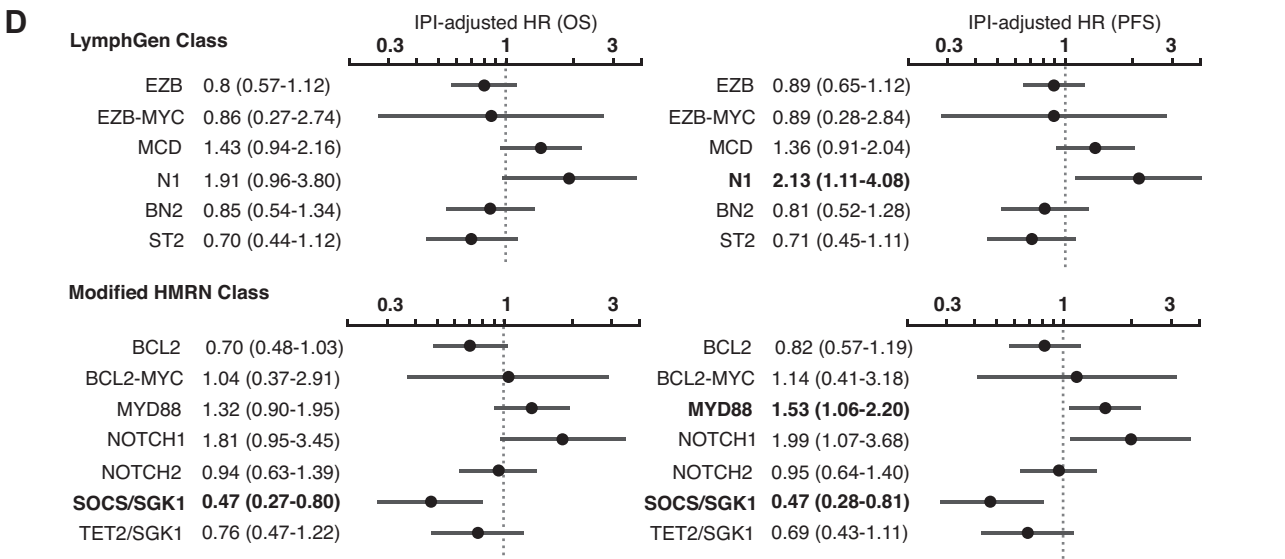

D
LymphGen Class
IPI-adjusted HR (OS)
EZB 0.8 (0.57-1.12)
EZB-MYC 0.86 (0.27-2.74)
MCD 1.43 (0.94-2.16)
N1 1.91 (0.96-3.80)
BN2 0.85 (0.54-1.34)
ST2 0.70 (0.44-1.12)
IPI-adjusted HR (PFS)
EZB 0.89 (0.65-1.12)
EZB-MYC 0.89 (0.28-2.84)
MCD 1.36 (0.91-2.04)
N1 2.13 (1.11-4.08)
BN2 0.81 (0.52-1.28)
ST2 0.71 (0.45-1.11)
Modified HMRN Class
BCL2 0.70 (0.48-1.03)
BCL2-MYC 1.04 (0.37-2.91)
MYD88 1.32 (0.90-1.95)
NOTCH1 1.81 (0.95-3.45)
NOTCH2 0.94 (0.63-1.39)
SOCS/SGK1 0.47 (0.27-0.80)
TET2/SGK1 0.76 (0.47-1.22)
BCL2 0.82 (0.57-1.19)
BCL2-MYC 1.14 (0.41-3.18)
MYD88 1.53 (1.06-2.20)
NOTCH1 1.99 (1.07-3.68)
NOTCH2 0.95 (0.64-1.40)
SOCS/SGK1 0.47 (0.28-0.81)
TET2/SGK1 0.69 (0.43-1.11)

Figure 2

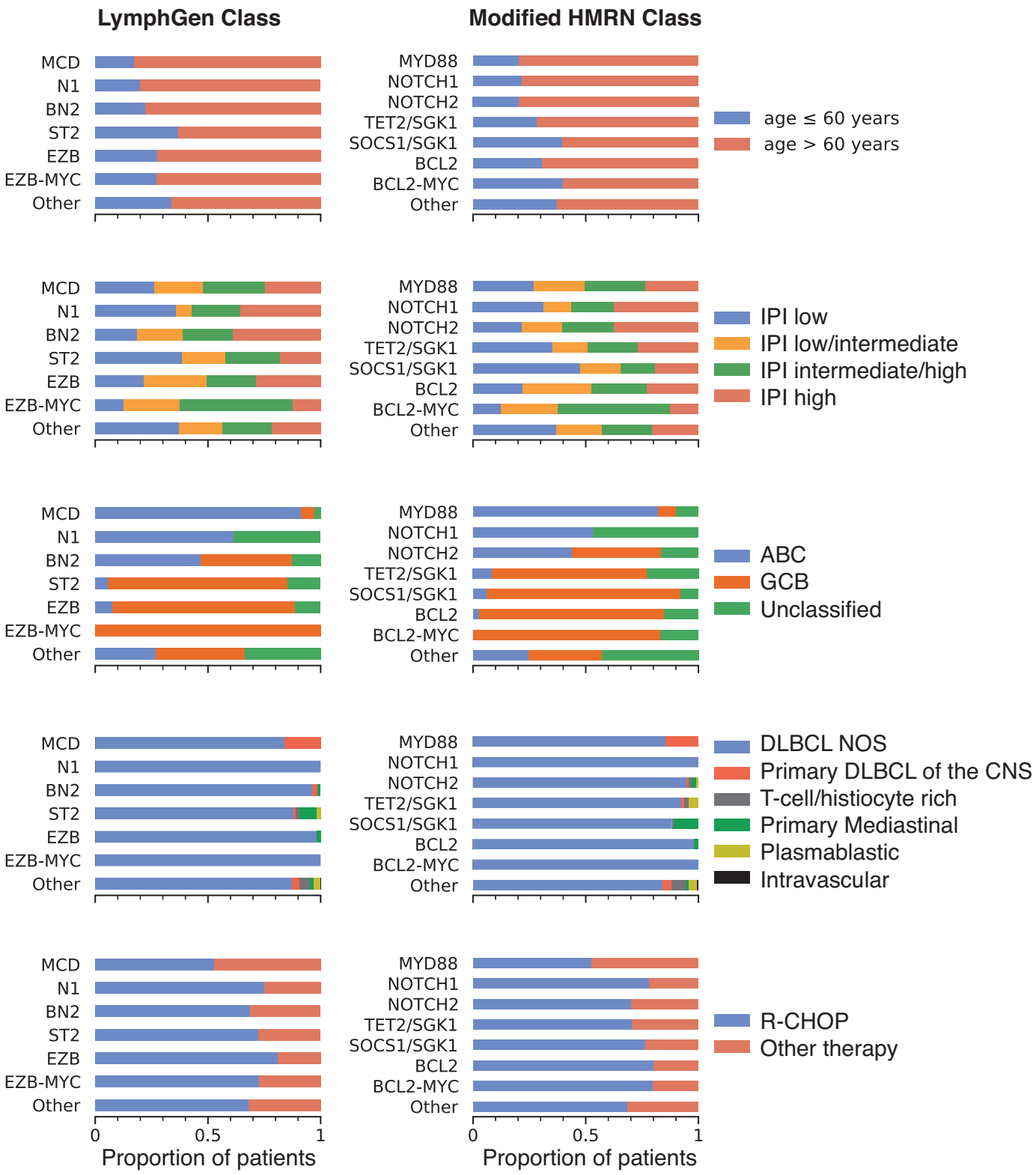
Supplementary Figure Legends

Supplementary Figure 1
The distribution of age, IPI score, COO, tumour histology and treatment regimen is shown for each of the subgroups classified by LymphGen or the modified HMRN criteria.

Supplementary Tables

Supplementary Table 1
Table showing the molecular subgroups assigned to all 928 patients in the original HMRN study (Lacy et al) compared to that from the LymphGen classifier.

Supplementary Table 2
Unadjusted and IPI-adjusted hazard ratios for R-CHOP-treated patients classified by either Modified HMRN or the LymphGen classifier relative to the unclassified “Other” patients. Hazard ratios were calculated by Cox model for both OS and PFS as indicated.



Supplementary Figure 1

Supplementary Table 2 Unadjusted and IPI-adjusted Hazard Ratios

Modified_HMRN_Class, Overall Survival. Baseline Class = "Other".
N = 648 (312 events) Unadjusted, N = 535 (248 events) Adjusted.

	Unadjusted Hazard Ratio (95% CI)	p-value	IPI Adjusted Hazard Ratio (95% CI)	p-value
BCL2	0.86 (0.62, 1.19)	0.36	0.70 (0.48, 1.03)	0.069
BCL2-MYC	1.37 (0.55, 3.38)	0.50	1.04 (0.37, 2.91)	0.94
MYD88	1.26 (0.89, 1.80)	0.20	1.32 (0.90, 1.95)	0.16
NOTCH1	**1.92 (1.07, 3.45)**	**0.029**	1.81 (0.95, 3.45)	0.07
NOTCH2	1.13 (0.80, 1.59)	0.49	0.94 (0.63, 1.39)	0.75
SOCS1/SGK1	**0.42 (0.26, 0.68)**	**0.00037**	**0.47 (0.27, 0.80)**	**0.0054**
TET2/SGK1	0.74 (0.48, 1.15)	0.18	0.76 (0.47, 1.22)	0.26

Modified_HMRN_Class, Progression Free Survival. Baseline Class = "Other".
N = 594 (317 events) Unadjusted, N = 514 (269events) Adjusted.

	Unadjusted Hazard Ratio (95% CI)	p-value	IPI Adjusted Hazard Ratio (95% CI)	p-value
BCL2	0.94 (0.67, 1.30)	0.70	0.82 (0.57, 1.19)	0.30
BCL2-MYC	1.35 (0.55, 3.33)	0.52	1.14 (0.41, 3.18)	0.80
MYD88	**1.45 (1.03, 2.03)**	**0.034**	**1.53 (1.06, 2.20)**	**0.022**
NOTCH1	**2.19 (1.22, 3.93)**	**0.0088**	**1.99 (1.07, 3.68)**	**0.029**
NOTCH2	1.21 (0.86, 1.70)	0.28	0.95 (0.64, 1.40)	0.79
SOCS1/SGK1	**0.43 (0.26, 0.70)**	**0.00067**	**0.47 (0.28, 0.81)**	**0.0061**
TET2/SGK1	0.68 (0.44, 1.06)	0.091	0.69 (0.43, 1.11)	0.13

Lymphgen_Class_MYC, Overall Survival. Baseline Class = "Other".
N = 648 (312 events) Unadjusted, N = 535 (248 events) Adjusted.

	Unadjusted Hazard Ratio (95% CI)	p-value	IPI Adjusted Hazard Ratio (95% CI)	p-value
BN2	1.10 (0.73, 1.64)	0.66	0.85 (0.54, 1.34)	0.48
EZB	0.97 (0.73, 1.29)	0.83	0.80 (0.57, 1.12)	0.20
EZB-MYC	1.49 (0.61, 3.63)	0.38	0.86 (0.27, 2.74)	0.80
MCD	1.39 (0.95, 2.05)	0.094	1.43 (0.94, 2.16)	0.095
N1	1.84 (0.97, 3.49)	0.063	1.91 (0.96, 3.80)	0.064
ST2	**0.55 (0.36, 0.86)**	**0.0085**	0.70 (0.44, 1.12)	0.13

Lymphgen_Class_MYC, Progression Free Survival. Baseline Class = "Other".
N = 594 (317events) Unadjusted, N = 514 (269 events) Adjusted.

	Unadjusted Hazard Ratio (95% CI)	p-value	IPI Adjusted Hazard Ratio (95% CI)	p-value
BN2	1.05 (0.70, 1.59)	0.80	0.81 (0.52, 1.28)	0.37
EZB	1.05 (0.79, 1.39)	0.73	0.89 (0.65, 1.22)	0.47
EZB-MYC	1.39 (0.57, 3.40)	0.47	0.89 (0.28, 2.84)	0.85
MCD	1.33 (0.91, 1.95)	0.14	1.36 (0.91, 2.04)	0.13
N1	**2.11 (1.11, 4.00)**	**0.023**	**2.13 (1.11, 4.08)**	**0.023**
ST2	**0.60 (0.39, 0.92)**	**0.020**	0.71 (0.45, 1.11)	0.13