blood advances

American Society of Hematology
2021 L Street NW, Suite 900,
Washington, DC 20036
Phone: 202-776-0544 | Fax 202-776-0545
bloodadvances@hematology.org

Longitudinal expression profiling identifies a poor risk subset of patients with ABC-type Diffuse Large B Cell Lymphoma

Tracking no: ADV-2022-007536R1

Findlay Bewicke-Copley (Barts Cancer Institute, Queen Mary University, United Kingdom) Koorosh Korfi (Roche Glycart AG, Switzerland) Shamza Araf (Barts Cancer Institute, Queen Mary University, United Kingdom) Brendan Hodkinson (Janssen Research & Development, United States) Emil Kumar (Queen Mary University of London, United Kingdom) Thomas Cummin (Portsmouth Hospitals NHS Trust, United Kingdom) Margaret Ashton-Key (Southampton University Hospital, United Kingdom) Sharon Barrans (St James's Institute of Oncology, United Kingdom) Suzan van Hoppe (Haematological Malignancy Diagnostic Service, United Kingdom) Cathy Burton (St James's University Hospital, United Kingdom) Mohamed Elshiekh (Imperial College London, United Kingdom) Simon Rule (Plymouth University Medical School, United Kingdom) Nicola Crosbie (University Hospitals Plymouth NHS Trust, United Kingdom) Andrew Clear (Queen Mary University of London, United Kingdom) Maria Calaminici (Centre for Haemato-Oncology; Barts Cancer Institute; Queen Mary University of London, United Kingdom) Hendrik Runge (University of Cambridge, United Kingdom) Robert Hills (University of Oxford, United Kingdom) David Scott (BC Cancer, Canada) Lisa Rimsza (Mayo Clinic, United States) Geetha Menon (Liverpool Clinical Laboratories, United Kingdom) Chulin Sha (Institute of Basic Medicine and Cancer, Chinese Academy of Sciences, China) John Davies (University of Leeds, United Kingdom) Ai Nagano (University College London, United Kingdom) Andrew Davies (University of Southampton, United Kingdom) Daniel Painter (University of York, United Kingdom) Alexandra Smith (University of York, United Kingdom) John Gribben (Barts Cancer Institute, United Kingdom) Kikkeri Naresh (Fred Hutchinson Cancer Research Center, United States) David Westhead (University of Leeds, United Kingdom) Jessica Okosun (Barts Cancer Institute, Queen Mary University of London, United Kingdom) Andrew Steele (Johnson and Johnson, United States) Daniel Hodson (University of Cambridge, United Kingdom) Sriram Balasubramanian (Janssen R&D, United States) Peter Johnson (University of Southampton, United Kingdom) Jun Wang (Barts Cancer Institute, United Kingdom) Jude Fitzgibbon (Queen Mary University of London, United Kingdom)

Abstract:
Despite the effectiveness of immuno-chemotherapy, 40% of patients with diffuse large B-cell lymphoma (DLBCL) experience relapse or refractory disease. Longitudinal studies have previously focused on the mutational landscape of relapse but falling short of providing a consistent relapse-specific genetic signature. In our study, we have focussed attention on the changes in gene expression profile accompanying DLBCL relapse using archival paired diagnostic/relapse specimens from 38 de novo DLBCL patients. Cell of origin remained stable from diagnosis to relapse in 84% of patients, with only a single patient showing COO switching from ABC to GCB. Analysis of the transcriptomic changes that occur following relapse suggest ABC and GCB relapses are mediated via different mechanisms. We developed a 30-gene discriminator for ABC-DLBCLs derived from relapse-associated genes, that defined clinically distinct high and low risk subgroups in ABC-DLBCLs at diagnosis in datasets comprising both population-based and clinical trial cohorts. This signature also identified a population of <60-year-old patients with superior PFS and OS treated with Ibrutinib-R-CHOP as part of the PHOENIX trial. Altogether this new signature adds to the existing toolkit of putative genetic predictors now available in DLBCL that can be readily assessed as part of prospective clinical trials.

Conflict of interest: No COI declared

COI notes:

Preprint server: No;

Author contributions and disclosures: F.B.C., K.K., S.A., J.G., J.O., P.J., J.W. & J.F. conceived and designed the study. F.B.C., J.W. & J.F. wrote the manuscript. K.K., S.A., E.K., A.C., T.C., M.A.K., S.B., S.V.H., C.B., M.E., S.R., N.C., G.M., M.B., A.N., A.D., A.S., K.N.N & M.C. collected samples and clinical information. F.B.C, J.W. & J.F. devised methods for analysis. F.B.C, D.J.H. & J.W. performed the bioinformatic analysis. K.K., S.A., D.W.S, L.M.R, performed experiments. C.S., D.W., D.P., B.H., D.J.H, S.B. & A.S. provided access to other data sets. All authors read, critically reviewed, and approved the manuscript.

Non-author contributions and disclosures: No;

Agreement to Share Publication-Related Data and Data Sharing Statement: All novel data from the publication is avaliable on GEO (GSE193566)

Clinical trial registration information (if any):

Title (120)

Longitudinal expression profiling identifies a poor risk subset of patients with ABC-type Diffuse Large B Cell Lymphoma

Running Title (50):

Longitudinal expression profiling of rrDLBCL

Findlay Bewicke-Copley[1#], Koorosh Korfi[1], Shamza Araf[1], Brendan Hodkinson[2], Emil Kumar[1], Thomas Cummin[3], Margaret Ashton-Key[4], Sharon Barrans[5], Suzan van Hoppe[5], Cathy Burton[5], Mohamed Elshiekh[6], Simon Rule[7], Nicola Crosbie[8], Andrew Clear[9], Maria Calaminici[2], Hendrik Runge[10], Robert K. Hills[11], David W. Scott[12], Lisa M. Rimsza[13], Geetha Menon[14], Chulin Sha[15], John R. Davies[15], Ai Nagano[1], Andrew Davies[3], Daniel Painter[16], Alexandra Smith[16], John Gribben[9], Kikkeri N. Naresh[6], David R. Westhead[15], Jessica Okosun[9], Andrew Steele[17], Daniel J. Hodson[10], Sriram Balasubramanian[17], Peter Johnson[3], Jun Wang[1#], Jude Fitzgibbon[1*#].*

[1] Centre for Cancer Genomics and Computational Biology, Barts Cancer Institute, Queen Mary University, London, United Kingdom; [2] Oncology Translational Research, Janssen Research & Development, Spring House, PA; [3] Cancer Research UK Centre, University of Southampton, Southampton, United Kingdom; [4] Cellular Pathology, University Hospital Southampton NHS Foundation Trust, Southampton, United Kingdom; [5] Haematological Malignancy Diagnostic Service, St. James's Institute of Oncology, Leeds, United Kingdom; [6] Cellular & Molecular Pathology, Imperial College NHS Trust & Imperial College London, London, United Kingdom; [7] Department of Haematology, Derriford Hospital, University of Plymouth, Plymouth, United Kingdom; [8] Department of Haematology, University Hospitals Plymouth NHS Trust, Plymouth, United Kingdom; [9] Centre for Haemato-Oncology, Barts Cancer Institute, Queen Mary University, London, United Kingdom; [10] Wellcome-MRC Cambridge Stem Cell Institute, University of Cambridge, Cambridge, United Kingdom; [11] Nuffield Department of Population Health, University of Oxford, Oxford, UK; [12] BC Cancer Centre for Lymphoid Cancer and Department of Medicine, University of British Columbia; [13] Department of

Laboratory Medicine and Pathology, Mayo Clinic Arizona. [14] *Haemato-Oncology Diagnostic Service, Liverpool Clinical Laboratories, Liverpool, United Kingdom;* [15] *School of Molecular and Cellular Biology, University of Leeds, Leeds, United Kingdom;* [16] *Epidemiology and Cancer Statistics Group, Department of Health Sciences, University of York, York, United Kingdom;* [17] *Oncology Translational Research, Janssen Research & Development, San Diego, CA*

* *Co-senior authors*

[#] *Corresponding authors*

Findlay Bewicke-Copley: *f.copley@qmul.ac.uk*

Jun Wang*: j.a.wang@qmul.ac.uk*

Jude Fitzgibbon: *j.fitzgibbon@qmul.ac.uk*

Novel data from the publication is avaliable on GEO (GSE193566). Contact the corresponding author for other forms of data sharing: f.copley@qmul.ac.uk.

Competing Interests:

K.K. is an employee and shareholder of Roche

J.F. has provided consultancy and received funding from Epizyme

D.W.S, L.M.R have IP rights to the Lymph2Cx assay

D.W.S. has provided consultancy to Abbvie, AstraZeneca, Celgene, Janssen and Incyte and has received research funding from Janssen, NanoString Technology and Roche/Genentech.

Abstract

Despite the effectiveness of immuno-chemotherapy, 40% of patients with diffuse large B-cell lymphoma (DLBCL) experience relapse or refractory disease. Longitudinal studies have previously focused on the mutational landscape of relapse but fell short of providing a consistent relapse-specific genetic signature. In our study, we have focussed attention on the changes in gene expression profile accompanying DLBCL relapse using archival paired diagnostic/relapse specimens from 38 de novo DLBCL patients. Cell of origin remained stable from diagnosis to relapse in 80% of patients, with only a single patient showing COO switching from ABC to GCB. Analysis of the transcriptomic changes that occur following relapse suggest ABC and GCB relapses are mediated via different mechanisms. We developed a 30-gene discriminator for ABC-DLBCLs derived from relapse-associated genes, that defined clinically distinct high and low risk subgroups in ABC-DLBCLs at diagnosis in datasets comprising both population-based and clinical trial cohorts. This signature also identified a population of <60-year-old patients with superior PFS and OS treated with Ibrutinib-R-CHOP as part of the PHOENIX trial. Altogether this new signature adds to the existing toolkit of putative genetic predictors now available in DLBCL that can be readily assessed as part of prospective clinical trials.

Key Points (140 each):

Cell of origin is stable between diagnosis and relapse

A 30 gene panel of relapse associated genes was able to stratify ABC patient survival at diagnosis.

Introduction

Diffuse large B-cell lymphoma (DLBCL) is a heterogeneous disease encompassing multiple molecular and biological subtypes. Although potentially curable with immuno-chemotherapy, up to 40% of patients will experience relapsed or refractory disease [1,2]. The current standard of care for these lymphomas has not changed in the past 2 decades, and efforts are turning to identification of sub-groups of DLBCL that may demonstrate preferential response to existing or novel therapies [3,4]. Early work focused on DLBCL at diagnosis, using gene expression profiling to delineate the "cell-of origin" (COO) classification system (Germinal Centre B-cell-like (GCB), Activated B-cell-like (ABC) and unclassified) with ABC tumours being linked to poorer outcome [5]. However, attempts to utilise molecular analyses to tailor treatment, and specifically to develop alternative R-CHOP (Rituximab, Cyclophosphamide, Doxorubicin, Vincristine, Prednisolone) regimens to mitigate the poorer outcome of patients in the ABC-DLBCL subgroup, have not, to date, led to significant improvements [6–9].

More recently, attempts to refine the taxonomy of DLBCL through integrative genomic analysis have demonstrated additional heterogeneity not captured by the previous COO classification [10–14]. This has led to the growing realisation that DLBCL encompasses a number of biological entities with distinct oncogenic mechanisms, requiring a more sophisticated approach to patient management and trial design. To date, these studies have predominantly focused on analysing single tumour biopsies at diagnosis, with our understanding of the pre-programmed or acquired mechanisms underpinning relapsed disease hindered by the limited availability of sequential biopsy samples. The majority of longitudinal studies published thus far have focussed on genetic changes between the diagnostic and relapse tumour, providing important confirmation of the clonal relationship between diagnosis and relapse, and describing recurrent relapse-associated genetic aberrations, but fell short of providing a consistent relapse-specific genetic signature [15–22] **(Supplemental Table 1)**. In this study, we sought to utilise gene expression profiling in paired diagnostic and relapse tumours to further understand the mechanisms underpinning treatment failure following immuno-chemotherapy. Using these data, we

demonstrate the stability of COO at relapse in the majority of cases and identify a novel relapse-associated gene expression signature that reliably discriminated two distinct outcome groups within the ABC type of DLBCL at diagnosis.

Methods

Patient Cohort

Ethical approval was obtained from the London Research Ethics Committee (LREC) of the East London and the City Health authority (10/H0704/65 and 06/Q0605/69). Written consent was obtained for the use of specimens for research purposes and samples from collaborating centres had local ethical approval. Paired diagnosis/relapse DLBCL biopsies were collated from 38 patients across 5 centres in the UK. All patients were treated with standard first-line rituximab-based immuno-chemotherapy (e.g. R-CHOP) and achieved either a partial or complete remission **(Figure 1A, Table 1)**. COO was determined using the Lymph2Cx assay on the NanoString platform [23] or the DLBCL Automatic Classifier [24] and all biopsies had >= 19% total B cell content, as estimated by CIBERSORT [25]. Thirty-four biopsies were nodal (15 diagnosis, 19 relapse) and 42 extranodal (23 diagnosis and 19 relapse). The site of the biopsy was identical at diagnosis and relapse for 20 cases (8 nodal, 12 extranodal).

Gene Expression Analysis

Gene expression profiling (GEP) of FFPE samples was carried out using the Ion Ampliseq™ Human Gene Expression array, consisting of 20,802 genes. Poorly captured genes (0 reads in ≥1/3 of the cohort) were removed, leaving 15,457 genes. Raw read counts were normalised to $\log_2$ count per million (CPM). Differential expression between matched relapse and diagnostic samples, and gene set enrichment analysis [26] were subsequently performed. The list of differentially expressed (DE) genes were selected for the following gene signature discovery using publicly available datasets. The

expression data and sample information for the rrDLBCL cohort are available from GEO accession GSE193566.

Derivation of a prognostic gene panel

Relapse associated genes found within our paired cohort ($p < 0.05$) were used in conjunction with the Prediction Analysis of Microarrays [27] (PAM) algorithm to define a survival signature for DLBCL. The expression of these genes within a cohort of 262 GCB and 249 ABC diagnostic DLBCL patients [14] (called the Reddy cohort hereafter) was used to train the PAM model. For the validation of the resulting gene signatures, a linear predictor model was constructed based on the prognostic value of each gene in the training dataset and the expression value in the validation dataset. This predictor score was used to stratify patients in three independent GEP cohorts: the REMoDL-B clinical trial [7], the LLMPP (Lymphoma/Leukemia Molecular Profiling Project series [28] and the Haematological Malignancy Research Network (HMRN) population cohort [29] (hereafter referred to as the REMoDL-B, LLMPP , and HMRN cohorts, respectively). All survival analyses were performed using the Cox Proportional Hazards Model in R.

See supplemental methods for a full description of the methods.

Results

COO is stable between diagnosis and relapse

The longitudinal series included 38 paired diagnostic-relapsed DLBCLs (**Figure 1A, Table 1**), all treated at diagnosis with R-CHOP or R-CHOP-like regimens. Cell of origin (COO) calling was successfully completed in both biopsies for 35 cases. COO was stable across 28 patients (80%) and corresponded to 17 ABC-ABC and 11 GCB-GCB pairs, with 2 further cases being unclassified (UNC) at both timepoints (**Figure 1B**). Discordant COO was a feature of just 5 cases (1 ABC-GCB, 2 ABC-UNC, 1 GCB-UNC, 1 UNC-ABC) with a single example of an ABC-GCB transition, suggesting that changes in DLBCL

143 trajectory at relapse, while reported in the literature [30], are uncommon. The median time to relapse
144 was 1.75 years, with 22 patients (58%) relapsing within two years**.** In two cases (1 GCB-GCB and 1
145 ABC-ABC) relapse occurred after more than 10 years (10.2 years and 13.9 years respectively).
146
147 **Deregulated gene expression between diagnosis and relapse**
148 We interrogated whole-transcriptome GEP data from all 76 biopsies with the aim of identifying
149 changes in gene expression associated with DLBCL relapse. Principal component analysis (PCA) based
150 on the full set of profiled genes (n=15,457) did not reveal distinct clustering of the diagnostic or
151 relapse samples (**Figure 1C**). There was no consistent pattern observed in the PCA values within the
152 individual pairs or based on the location of biopsies, nodal/extra-nodal disease, or time to relapse
153 (**Supplemental Figure 1A-C**). As expected, GEP profiles of the samples showed association based on
154 their COO (**Figure 1C**) where differential expression (DE) analyses of the ABC (n = 17) and GCB (n =
155 11) pairs identified unique sets of genes associated with relapse, based on COO (< 4% overlapping DE
156 genes, limma analysis $p < 0.05$) (**Supplemental Figure 1**). This was also supported by gene set
157 enrichment analysis (GSEA) where chromosome maintenance, DNA repair, and rRNA processing were
158 among the top upregulated pathways (FDR < 0.1) in the ABC-ABC series in comparison to adaptive
159 immunity, cytokine signalling and antigen processing and presentation signatures which were unique
160 to GCB-GCB pairs (**Figure 1D-E & Supplemental Tables 2-3**).
161
162 **A 30 gene outcome predictor in ABC DLBCL**
163 We postulated that the expression of these relapse associated genes might hold some prognostic
164 significance in a diagnostic cohort. To this end the PAM algorithm [27] was used to interrogate a total of
165 796 and 387 DE genes ($p < 0.05$) from our ABC and GCB diagnostic-relapse signatures respectively, in
166 the Reddy series of 262 GCB and 249 diagnostic ABC DLBCLs. This analysis identified a 30-gene
167 signature that separated ABC patients into 136 low and 113 high-risk cases with significantly different
168 overall survival (Hazard Ratio (HR)=1.89, 95% CI=1.26-2.83; log-rank p=0.0017; **Figure 1F**). The

majority of the genes in this panel have not previously been implicated in DLBCL pathogenesis, although notable exceptions included *MYC* and *TNFRSF9*, with *MYC* one of five genes demonstrating significant single-gene clinical association, inversely correlated with overall survival ($p < 0.05$; **Figure 1F**). STRING analysis of these 30 genes identified 7 highly interconnected clusters, with MYC at the centre of this protein interaction network (**Supplementary Figure 5**, **Supplementary Table 6**). In contrast to ABC patients, there was no equivalent predictor detected using PAM in the corresponding set of GCB cases. Attempts to define a response signature using the Reddy cohort without the prior enrichment of relapse associated genes were unsuccessful.

Validation of the 30-gene ABC predictor in 3 independent DLBCL series

The reproducibility of this 30 gene outcome predictor was evaluated in 3 separate DLBCL cohorts (REMoDL-B and HMRN – both with RNA profiling achieved using the cDNA-mediated Annealing, Selection, extension and Ligation (DASL) assay, and LLMPP – RNA profiling from an Affymetrix microarray chip [7,28,29]), all treated with R-CHOP (R-CHOP + bortezomib in 126 patients from the REMoDL-B cohort) at diagnosis and comprising 504 ABC cases in total. We evaluated each series separately. Within each cohort, a linear predictor score was calculated for each patient, based on the summation of the expression of 29 or 30 genes (as not all genes were represented on each platform), weighted by their beta coefficients from the training dataset (**Supplemental Table 4**). These linear predictors were standardised using a Z-transformation and each cohort was subdivided into high (standardized linear predictor >0) and low (standardized linear predictor <0) scoring risk groups (see Supplemental Methods). Analysis of the cause of deaths in the HMRN cohort shows that patients with lymphoma associated deaths had a significantly shorter follow up time than patients who died from other causes (*Wilcoxon rank sum* $p < 0.001$, **Supplemental Figure 2D**). Moreover, it was notable that non-lymphoma related deaths increased significantly from 3 years in this series and so we restricted our analysis of overall survival accordingly.

The algorithm stratified the 255 ABC REMoDL-B cases, into 108 low and 147 high-risk cases (three year OS; HR=2.04, 95% CI=1.073-3.875; *p*=0.026; **Figure 2A**), the LLMPP series of 93 ABC cases into 44 low and 49 high-risk cases (three year OS; HR=2.3, 95% CI=1.154-4.565; *p*=0.015; **Figure 2B**) and a UK population-based cohort (HMRN) of 156 ABC cases, into 72 low and 84 high-risk cases (three year OS; HR=1.93, 95% CI=1.06-3.522; *p*=0.029; **Figure 2C**). Across all three cohorts, patients with high linear predictor scores (high-risk) showed significant reduction in survival at three years. When later events were included, both the REMoDL-B and LLMPP data showed similar results (OS HR=2.11, 95% CI=1.115-3.993; *p*=0.019 & HR=2.17, 95% CI=1.109-4.242; *p*=0.02; **Supplemental Figure 2A-B** respectively), whilst the HMRN cohort showed a trend for reduced survival in the high-risk group, (HR=1.39, 95% CI=0.917-2.106; *p*=0.12; **Supplemental Figure 2C**) and we have reasoned that the performance of the discriminator may reflect the number of non-lymphoma related deaths in this population-based cohort.

We restricted our multivariate analysis to the REMoDL-B and LLMPP series, accounting for patient age, gender, IPI and stage (where available), and the high-scoring group remained associated with poorer OS (HR=1.95, Wald test *p*=0.042 for REMoDL-B; HR=2.19, *p*=0.023 for LLMPP; **Supplemental Table 5**), suggesting that our linear score offers an additional independent predictor. Whilst there was an over-representation of low IPI (0-2) cases observed in the low-risk group of the REMoDL-B cohort, this was not significant (Fisher's exact test *p*=0.08, **Figure 2D**) and, while case numbers are few, neither did we observe a significant enrichment in the 6 existing genetic subgroups defined by Lacy et al [13] in the HMRN cohort (Lacy subtype available for 63% of samples, Fisher's exact test *p*=0.422, **Figure 2E**).

Previous studies have identified a large number of verifiable random gene signatures, associated with outcome in other cancer types [31,32] so for completeness, we next compared the prognostic ability of our signature against 300,000 random 30 gene panels in the REMoDL-B and LLMPP

datasets, where it outperformed 95.5% of random signatures in the REMoDL-B Data, 98.32% in the LLMPP dataset, and 99.92% in both datasets concurrently (**Supplemental Figure 3**).

Signature predicts superior response to ibrutinib in younger DLBCL patients

We were also intent on testing whether our discriminator could identify populations of ABC patients most likely to respond to COO specific therapies. Wilson and colleagues have recently reported superior outcomes of patients in specific subtypes of DLBCL [33]. We reasoned that our signature may hold relevance for agents postulated to specifically target ABC-subtype DLBCL. The phase III PHOENIX study examined the addition of ibrutinib to R-CHOP in non-GCB DLBCL. Although ibrutinib addition failed to show benefit across the whole intention to treat (ITT) cohort, in younger patients (<60 years), outcomes were indeed superior in the ibrutinib-R-CHOP (I-R-CHOP) arm, with results for older (>60 years) patients seemingly confounded by increased toxicity of the drug. In view of the efficacy in this discrete group of patients, we assessed whether our linear predictor was able to discriminate patients in the PHOENIX cohort with a variable response to ibrutinib, focussing our attention on cases younger than 60 years that were confirmed as ABC sub-type using the HTG EdgeSeq COO Assay (n = 133).

Altogether patients with high linear predictor scores demonstrated poorer PFS compared to patients classified as low-risk in all patients <60 years of age irrespective of treatment (**Figure 3A**, Low-Risk = 57, High-Risk = 76, HR = 2.52, 95% CI = 1.23-5.16, log-rank p = 0.009), although OS was only marginally different (**Supplemental Figure 4A**, HR = 1.46, 95% CI = 0.54-3.95, p = 0.452). We next considered whether the linear predictor behaved differently in I-R-CHOP and R-CHOP treated patients. For ibrutinib-treated patients (n = 55), both PFS and OS were lower in the high versus low-risk group (**Figure 3B & Supplemental Figure 4B** Low-Risk = 26, High-Risk = 29, HR = 11.6, 95% CI = 1.48-90.9, p = 0.003 and p = 0.076 respectively). Indeed, the low-risk group (47%), had strikingly favourable outcomes, with no deaths reported in these 26 patients and only one patient

247 experiencing progression. It is important to note that the control, R-CHOP arm, demonstrated only a
248 trend to inferior outcomes in the high-risk group in PFS, compared to the significant survival
249 differences observed in the LLMPP, REMoDL-B and HMRN datasets (**Figure 3C & Supplemental Figure**
250 **4C**; Low-Risk = 32, High-Risk = 46, PFS: HR = 1.6, 95% CI = 0.727-3.52, *p* = 0.239 and OS: HR = 1.28,
251 95% CI = 0.429-3.82, *p* = 0.656 respectively).
252
253 Finally, we assessed the effect of ibrutinib addition in high and low-risk linear predictor groups
254 separately. Low-risk patients treated with I-R-CHOP had superior PFS and OS than those treated with
255 R-CHOP only (**Figure 3D**, Ibrutinib = 24, Placebo=33, *p* = 0.007 for PFS; **Supplemental Figure 4D**, *p* =
256 0.028 for OS); while in contrast, the high-risk group showed no difference observed between the
257 treatment arms for either PFS (**Figure 3E**, Ibrutinib = 31, Placebo = 45, HR = 0.927, 95% CI = 0.44-
258 1.95, p = 0.841) or OS (**Supplemental Figure 4E**, HR = 0.589, 95% CI = 0.156-2.22, *p* = 0.428). Similar
259 results were shown when examining the non-GCB group of patients. Together, these retrospective
260 data suggest that our gene signature may identify a group of DLBCL patients <60 years who derive
261 benefit from ibrutinib in combination with R-CHOP therapy.

Discussion

DLBCL comprises a molecularly heterogeneous group of lymphomas with different outcomes, linked to a variety of features including COO [25], occurrence of specific translocations [34] and more recently a combination of gene mutation and copy number aberrations [10–13,35]. There are several recently reported discriminators that rely primarily on gene expression, with an emphasis either towards the tumour B cell [3,36,37], or its immune microenvironment [38–41]. However, despite an increased understanding of the biology of these aggressive lymphomas, improvements to the existing standard of care have proven problematic. Altogether, there has been a reliance on the study of the diagnostic biopsy samples, with longitudinal studies typically hindered by the limited availability of sequential biopsy material. Studies comparing mutation status at diagnosis and relapse in paired biopsies, or interrogating independent series of pre-treatment and relapse cases [15–22] (**Supplemental Table 1**) have identified recurrent relapse-associated genes including *TP53* and *MYC*, although alone lack specificity to predict relapse. In this study, we focussed attention on the changes in gene expression profile that accompany DLBCL relapse, to consider whether this approach might offer a novel perspective on the biology of disease resistance. Our new data demonstrate that COO is largely stable between time points, suggest a distinctive pattern of relapse in ABC and GCB lymphomas based on differential gene expression, and resolve a 30-gene discriminator in ABC-DLBCL that defined clinically distinct low- and high-risk subgroups at diagnosis, that was informative both in an independent series of R-CHOP-treated patients, and young patients treated with ibrutinib + R-CHOP in the PHOENIX trial [42].

The accrual of paired material of suitable quality for analysis was challenging. From a large initial series of FFPE paired biopsies obtained from multiple UK institutions, suitably paired data was retrieved from 38 *de novo* DLBCL patients, constituting one of the largest published cohorts of paired diagnosis-relapse samples to date. Regardless, it is important to acknowledge the heterogeneity of the cohort; site of the biopsy differs between the diagnosis and relapse in 18 of the

38 pairs; the time to relapse varied across the series and samples demonstrated variable tumour content. Irrespective of these potential confounding effects, we have been able to make some robust observations shedding new light onto the evolution of DLBCL. We had initially sought to recover both DNA and RNA from these specimens, to facilitate a parallel analysis of mutation and gene expression, but this proved technically unfeasible in the majority of cases, highlighting the challenges in collating paired material of sufficient quantity and quality for multi-omic analyses. Our subsequent studies focused exclusively on generating gene expression data, through global GEP and a COO analysis. Comparison of paired biopsies confirmed what has long been assumed, but not formally shown, that COO is stable in most paired diagnostic/relapse cases, ruling out a simple switch in COO as the dominant mechanism underlying disease relapse and R-CHOP failure. Indeed, while changes in COO accompanying DLBCL relapse were observed in 5 cases, this included just a single example of ABC-GCB switching, where biopsies were excised from different locations 1.5 years apart (**Table 1**). While this example is reminiscent of a recent study demonstrating spatial and temporal heterogeneity in a case of DLBCL manifesting as site-discordant COO and response to immuno-chemotherapy [43], these data confirm that such discordant cases represent the exception rather than the rule.

We noted minimal overlap in DE genes between COO groups, with GSEA suggesting that relapse is likely mediated by different mechanisms depending on the tumour's COO. Tumour growth and proliferation signatures were enriched in ABC relapses, while adaptive immunity-related signatures were a feature of GCB type lymphomas. Consequently, we considered ABC (n=17) and GCB (n=11) lymphomas separately for subsequent analysis. We next tested whether these relapse associated genes held prognostic significance in a diagnostic cohort. Using the PAM algorithm we resolved a 30-gene signature that divided ABC cases in into low- and high-risk groups. Critically, this expression signature was validated using a linear score in 3 independent GEP datasets derived using different platforms and comprising both population-based and clinical trial cohorts.

Going forward, it will be important to prospectively validate individual signatures, as well as benchmark them against each other, to determine their relative merits and application in real world patients. While it is reassuring to note in three recent mutation-focussed studies [10,12,13,35] the significant overlap and consensus across classifications based on gene mutation, it remains to be seen whether the various emerging gene expression-based signatures similarly resolve identical groups of DLBCLs, or rather each identify distinct high-risk groups. Moreover, combined mutation and gene expression data from the HMRN dataset demonstrated that high and low risk patients from our ABC discriminator arose independently of the groups reported by Lacy et al [13]. In contrast, there was limited overlap of patients classified in the Phoenix trial using the LymphGen to allow a direct comparison with our high-low risk patient groups. This data suggests that gene expression profiling imparts important information independent of mutation and CNA-based classifications.

There is a recognition that genetic signatures, rather than informing clinical decisions based on outcome prediction, may offer instead a tool to identify discrete populations of patients who may benefit from specific precision-based approaches to treatment. It was of interest in our study, that our ABC-discriminator resolved patients with particularly favourable outcome following ibrutinib + R-CHOP in the PHOENIX study within ABC-subtype patients diagnosed at <60 years, albeit in a small retrospective cohort. Importantly, however, in this cohort the discriminator was unable to identify groups with different outcomes in the R-CHOP arm. Ideally, this observation will undergo prospective validation in patients on the upcoming combination study of the BTK inhibitor acalabrutinib with R-CHOP for untreated DLBCL (REMoDL-A: clinicaltrials.gov/ct2/show/NCT04546620) as part of the UK PMAL programme.

There are certain limitations in our study. Overall, the cohort sizes are small, particularly in the example of GCB-GCB relapse pairs, which may explain the inability to generate a prognostic

discriminator for this group of patients. Furthermore, while we employed a biologically agnostic approach to our discriminator discovery, so as not to overlook the impact of unappreciated gene interactions or biology, the resulting discriminator by its nature lacks an immediately apparent biological rationale. However, an interaction network revealed 7 biologically distinct clusters of protein interactions containing several enriched pathways with potential relevance to disease progression, including RNA transport, protein processing and immune pathways. The notable presence of *MYC* at the centre of the interaction network highlights the role of *MYC* in disease aggressiveness and reinforces the need to develop *MYC*-directed therapies.

The future utility of the many emerging genetic discriminators requires independent validation as part of prospective clinical trials and highlight the need for comprehensive and multi-omic profiling of these cohorts. There are currently limitations in performing direct comparisons between existing GEP studies, e.g. the use of different discovery platforms, and it is possible that fluctuations in the proportion of specific subgroups observed may reflect the unpredictable nature of real-world studies (HMRN) compared with clinical trials (REMoDL-B). Indeed, the inclusion of patients for analysis in many biological studies, are typically dependent on a confirmed lymphoma diagnosis, their treatment, and having sufficient residual material for molecular analysis. In addition, whilst various candidates are being investigated to augment the efficacy of R-CHOP, the performance of the proposed predictive signatures will require re-appraisal in the context of any new standard of care.

In summary, we have leveraged one of the largest cohorts of paired diagnosis-relapse series in DLBCL demonstrating the stability of COO and derived a 30-gene signature that robustly distinguished low- and high-risk subgroups of ABC patients. This signature also identifies patients who derive benefit from BTK inhibition in combination with R-CHOP adding to the existing toolkit of putative genetic predictors now available in DLBCL that can be readily assessed as part of prospective clinical trials.

Acknowledgements:

The authors thank the patients and their families for donating specimens for research in this study as well as the Haemato-Oncology Tissue Bank at the BCI. Blood Cancer UK through funding of the Precision Medicine for Aggressive Lymphoma (PMAL) consortium. Cancer Research UK Centre of Excellence Award to Barts Cancer Centre [C16420/A18066]. JO is funded by Cancer Research UK (C57432/A22742). Haematological Malignancy Research Network is funded by Cancer Research UK, grant number 29685; and Blood Cancer UK, grant number 15037. We want to thank Dr George Wright for earlier discussions on the LLMPP series of cases.

Authorship Contributions:

F.B.C., K.K., S.A., J.G., J.O., P.J., J.W. & J.F. conceived and designed the study.

F.B.C., J.W. & J.F. wrote the manuscript.

K.K., S.A., E.K., A.C., T.C., M.A.K., S.B., S.V.H., C.B., M.E., S.R., N.C., G.M., M.B., A.N., A.D., A.S., K.N.N & M.C. collected samples and clinical information.

F.B.C, J.W. & J.F. devised methods for analysis.

F.B.C, D.J.H. & J.W. performed the bioinformatic analysis.

K.K., S.A., D.W.S, L.M.R, performed experiments.

C.S., D.W., D.P., B.H., D.J.H, S.B. & A.S. provided access to other data sets.

All authors read, critically reviewed, and approved the manuscript.

Competing Interests:

J.F. has provided consultancy and received funding from Epizyme

K.K. is an employee and shareholder of Roche

D.W.S, L.M.R have IP rights to the Lymph2Cx assay

D.W.S. has provided consultancy to Abbvie, AstraZeneca, Celgene, Janssen and Incyte and has received research funding from Janssen, NanoString Technology and Roche/Genentech.

References

1. Coiffier B, Lepage E, Brière J, 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):235–242.
2. Rovira J, Valera A, Colomo L, et al. Prognosis of patients with diffuse large B cell lymphoma not reaching complete response or relapsing after frontline chemotherapy or immunochemotherapy. *Ann. Hematol.* 2015;94(5):803–812.
3. Sha C, Barrans S, Cucco F, 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):202–212.
4. Ma Z, Niu J, Cao Y, et al. Clinical significance of 'double-hit' and 'double-expression' lymphomas. *J. Clin. Pathol.* 2019;jclinpath-2019-206199.
5. Liu Y, Barta SK. Diffuse large B-cell lymphoma: 2019 update on diagnosis, risk stratification, and treatment. *Am. J. Hematol.* 2019;94(5):604–616.
6. Younes A, Sehn LH, Johnson P, 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):1285–1295.
7. Davies A, Cummin TE, Barrans S, 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;1–14.
8. Gopal AK, Schuster SJ, Fowler NH, et al. Ibrutinib as treatment for patients with relapsed/refractory follicular lymphoma: Results from the open-label, multicenter, phase II DAWN study. *J. Clin. Oncol.* 2018;36(23):2405–2412.
9. Nowakowski GS, Chiappella A, Gascoyne RD, 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):1317–1328.
10. Schmitz R, Wright GW, Huang DW, et al. Genetics and Pathogenesis of Diffuse Large B-Cell

Lymphoma. *N. Engl. J. Med.* 2018;378(15):1396–1407.

11. Wright GW, Huang DW, Phelan JD, et al. A Probabilistic Classification Tool for Genetic Subtypes of Diffuse Large B Cell Lymphoma with Therapeutic Implications. *Cancer Cell*. 2020;37(4):551-568.e14.

12. Chapuy B, Stewart C, Dunford AJ, et al. Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. *Nat. Med.* 2018;24(5):679–690.

13. Lacy SE, Barrans SL, Beer PA, et al. Targeted sequencing in DLBCL, molecular subtypes, and outcomes: A Haematological Malignancy Research Network report. *Blood*. 2020;135(20):1759–1771.

14. Reddy A, Zhang J, Davis NS, et al. Genetic and Functional Drivers of Diffuse Large B Cell Lymphoma. *Cell*. 2017;171(2):481-494.e15.

15. Morin RD, Assouline S, Alcaide M, et al. Genetic Landscapes of Relapsed and Refractory Diffuse Large B-Cell Lymphomas. *Clin. Cancer Res.* 2016;22(9):2290–2300.

16. Melchardt T, Hufnagl C, Weinstock DM, et al. Clonal evolution in relapsed and refractory diffuse large B-cell lymphoma is characterized by high dynamics of subclones. *Oncotarget*. 2016;7(32):.

17. Juskevicius D, Lorber T, Gsponer J, et al. Distinct genetic evolution patterns of relapsing diffuse large B-cell lymphoma revealed by genome-wide copy number aberration and targeted sequencing analysis. *Leukemia*. 2016;30(12):2385–2395.

18. Jiang Y, Redmond D, Nie K, et al. Deep sequencing reveals clonal evolution patterns and mutation events associated with relapse in B-cell lymphomas. *Genome Biol.* 2014;15(8):1–17.

19. Mareschal S, Dubois S, Viailly P-J, et al. Whole Exome Sequencing of Relapsed/Refractory Patients Expands the Repertoire of Somatic Mutations in Diffuse Large B-Cell Lymphoma. *Genes Chromosom. Cancer*. 2016;55:251–267.

20. Rushton CK, Arthur SE, Alcaide M, et al. Genetic and evolutionary patterns of treatment

resistance in relapsed B-cell lymphoma. *Blood Adv.* 2020;4(13):2886–2898.

21. Nijland M, Seitz A, Terpstra M, et al. Mutational Evolution in Relapsed Diffuse Large B-Cell Lymphoma. *Cancers (Basel).* 2018;10(11):459.

22. Greenawalt DM, Liang WS, Saif S, et al. Comparative analysis of primary versus relapse/refractory DLBCL identifies shifts in mutation spectrum. *Oncotarget*. 2017;8(59):99237–99244.

23. Scott DW, Mottok A, Ennishi D, 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):2848–2856.

24. Barrans SL, Crouch S, Care MA, et al. Whole genome expression profiling based on paraffin embedded tissue can be used to classify diffuse large B-cell lymphoma and predict clinical outcome. *Br. J. Haematol.* 2012;159(4):441–453.

25. Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods*. 2015;12(5):453–457.

26. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* 2005;102(43):15545–15550.

27. Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc. Natl. Acad. Sci.* 2002;99(10):6567–6572.

28. Lenz G, Wright G, Dave SS, et al. Stromal Gene Signatures in Large-B-Cell Lymphomas. *N. Engl. J. Med.* 2008;359(22):2313–2323.

29. Smith A, Howell D, Crouch S, et al. Cohort Profile: The Haematological Malignancy Research Network (HMRN): a UK population-based patient cohort. *Int. J. Epidemiol.* 2018;47(3):700-700g.

30. Araf S, Korfi K, Bewicke-Copley F, et al. Genetic heterogeneity highlighted by differential FDG-PET response in diffuse large B-cell lymphoma. *Haematologica*. 2020;105(6):318–321.

31. Venet D, Dumont JE, Detours V. Most Random Gene Expression Signatures Are Significantly Associated with Breast Cancer Outcome. *PLoS Comput. Biol.* 2011;7(10):e1002240.
32. Shimoni Y. Association between expression of random gene sets and survival is evident in multiple cancer types and may be explained by sub-classification. *PLOS Comput. Biol.* 2018;14(2):e1006026.
33. Wilson WH, Wright GW, Huang DW, et al. Effect of ibrutinib with R-CHOP chemotherapy in genetic subtypes of DLBCL. *Cancer Cell*. 2021;1–11.
34. Hilton LK, Tang J, Ben-Neriah S, et al. The double-hit signature identifies double-hit diffuse large B-cell lymphoma with genetic events cryptic to FISH. *Blood*. 2019;134(18):1528–1532.
35. Runge HFP, Lacy S, Barrans S, 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):216–220.
36. Dubois S, Tesson B, Mareschal S, et al. Refining diffuse large B-cell lymphoma subgroups using integrated analysis of molecular profiles. *EBioMedicine*. 2019;48:58–69.
37. Ennishi D, Jiang A, Boyle M, 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.
38. Kotlov N, Bagaev A, Revuelta M V., et al. Clinical and biological subtypes of B-cell lymphoma revealed by microenvironmental signatures. *Cancer Discov.* 2021;11(6):candisc.0839.2020.
39. Matias Autio, Suvi-Katri Leivonen, Oscar Brück, et al. Immune cell constitution in the tumor microenvironment predicts the outcome in diffuse large B-cell lymphoma. *Haematologica*. 2020;106(3):718–729.
40. Merdan S, Subramanian K, Ayer T, et al. Gene expression profiling-based risk prediction and profiles of immune infiltration in diffuse large B-cell lymphoma. *Blood Cancer J.* 2021;11(1):2.
41. Staiger AM, Altenbuchinger M, Ziepert M, et al. A novel lymphoma-associated macrophage interaction signature (LAMIS) provides robust risk prognostication in diffuse large B-cell

lymphoma clinical trial cohorts of the DSHNHL. *Leukemia*. 2019;

42. Johnson P, Balasubramanian S, Hodkinson B, et al. Clinical Impact of Ibrutinib with R-CHOP in Untreated Non-GCB DLBCL Co-Expressing BCL2 and MYC Genes in the Phase 3 Phoenix Trial. *Blood*. 2019;134(Supplement_1):354.

43. Araf S, Wang J, Korfi K, et al. Genomic profiling reveals spatial intra-tumor heterogeneity in follicular lymphoma. *Leukemia*. 2018;32(5):1258–1263.

44. Wang HI, Smith A, Aas E, et al. Treatment cost and life expectancy of diffuse large B-cell lymphoma (DLBCL): a discrete event simulation model on a UK population-based observational cohort. *Eur. J. Heal. Econ.* 2017;18(2):255–267.

Figure Legends

Figure 1: Gene expression profiles of paired diagnosis and relapse DLBCL biopsies.

*38 patients who underwent relapse were included in the study, the clinical features of these patients are shown in **A**. COO remained stable in the majority of cases **(B).** Gene expression profiling was carried out using Ion AmpliSeq™ Transcriptome Human Gene Expression Kit. Principal Component Analysis carried out on these samples suggested poor separation based on timepoint, with a greater degree of separation observed in COO **(C).** Diagnosis = green, relapse = red**,** ABC = blue, GCB = orange, unclassified = grey, NA=black**.** Differential gene expression was carried out separately for the stable ABC and GCB cohorts and gene set enrichment analysis was performed, with the number of genes sets dysregulated (FDR <= 0.1) at relapse are shown in **D**. Heatmaps of normalised enrichment score for examples of the dysregulated gene sets are shown in **E.** A 30 gene panel capable of stratifying ABC-DLBCL patients from a training cohort (Reddy et al 2017) [14] into two risk groups with different overall survival was discovered using PAMR **(F)**. Red = High Risk, Blue = Low Risk, ** p <= 0.01, * p <= 0.05, . p <= 0.1*

Figure 2: Validation of 30 gene risk model for ABC-DLBCL in population and clinical trial cohorts. *The risk model was tested with survival restricted to 3 years. The risk model, based on the 30 gene panel, was also able to separate data from the REMoDL-B clinical trial [7] **(A)**, the R-CHOP arm of the LLMPP cohort (2008) [28] **(B)**, and the HMRN population study [44] **(C).** Red = High Risk, Blue = Low Risk. **D)** Comparison of IPI scores and the risk groups defined using the linear predictor in the REMoDL-B cohort**. E)** Comparison of genetic sub-categories described by Lacy [13] with risk groups defined using the linear predictor in the HMRN cohort. Of the 156 ABC cases in the HMRN data, the genomic sub-groups were avaliable for 98.*

Figure 3: Prognostic ability of the linear predictor in the PHOENIX trial cohort. *The GEP data from the ABC patients <60 years old in the PHOENIX trial were used to generate linear scores for each*

patient. These scores where then used to stratify the patients into high and low risk cohorts. Kaplan Meier plots of the PFS of these patient subgroups is shown. ***(A)*** *Both treatment arms combined, only patients designated ABC by GEP. The PFS of these subgroups was also examined in each arm separately,* ***(B)*** *Ibrutinib and* ***(C)*** *placebo. Red = High Risk, Blue = Low Risk. Finally, the effect of the drugs on PFS within the subgroups was assessed,* ***(D)*** *low risk and* ***(E)*** *high risk. Green = R-CHOP + placebo, Purple = R-CHOP + Ibrutinib.*

Figure 1

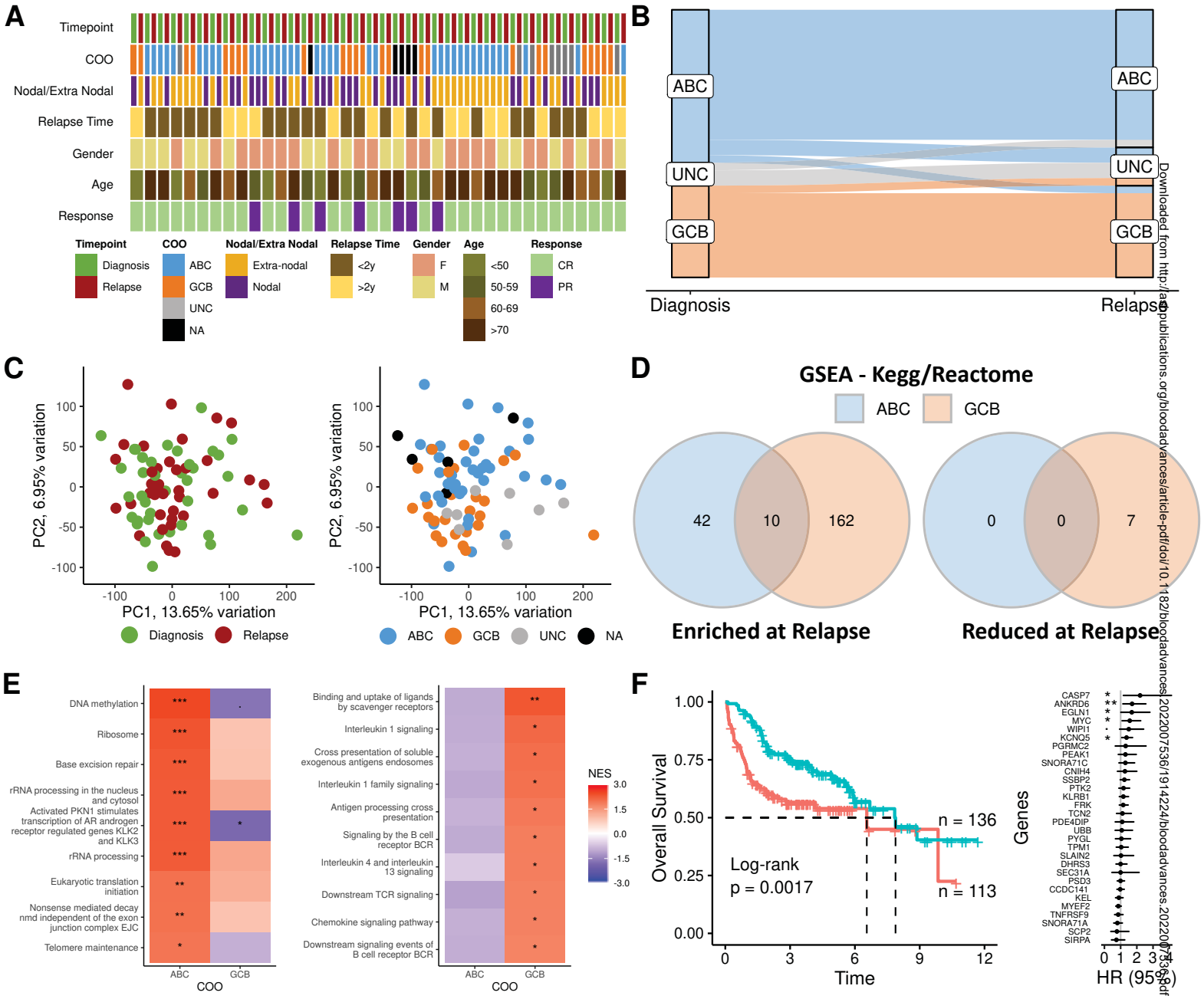

Figure 2

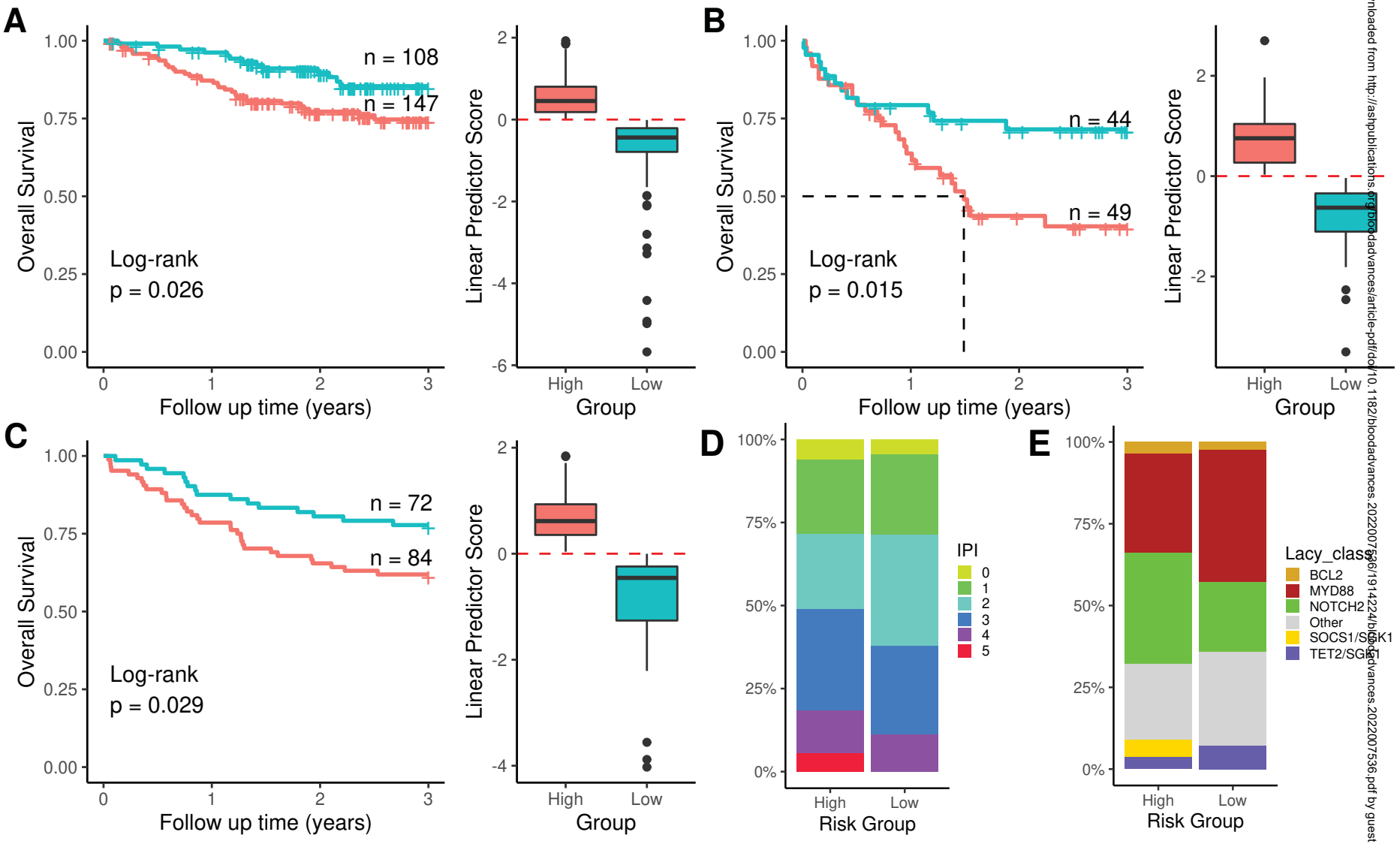

Figure 3