Diffuse large B-cell lymphoma: sub-classification by massive parallel quantitative RT-PCR

Xuemin Xue,[1,2] Naiyan Zeng,[2] Zifen Gao,[1] Ming-Qing Du[2]

[1]Department of Pathology, Health Science Center, Peking University;

[2]Division of Molecular Histopathology, Department of Pathology, University of Cambridge, UK;

Running title: DLBCL COO-classification by qRT-PCR

Key words: DLBCL, COO-classification, classifier

Corresponding authors:

Professor Ming-Qing Du

Division of Molecular Histopathology,

Department of Pathology,

University of Cambridge,

Level 3 Lab Block, Box 231,

Addenbrooke's Hospital,

Hills Road, Cambridge, CB22 QQ, UK.

Email: mqd20@cam.ac.uk;

Tel.: 00 44 (0)1223 767092;

Fax: 00 44 (0)1223 586670

The research in Du lab was supported by research grants (LLR10006 & LLR13006) from Leukaemia & Lymphoma Research, U.K.. XX was supported by a visiting fellowship from the China Scholarship Council, Ministry of Education, P.R. China.

ABSTRACT

Diffuse large B-cell lymphoma (DLBCL) is a heterogeneous entity with remarkably variable clinical outcome. Gene expression profiling (GEP) classifies DLBCL into activated B-cell like (ABC), germinal centre B-cell like (GCB) and Type-III subtypes, with ABC-DLBCL characterised by a poor prognosis and constitutive NF-κB activation. A major challenge for the application of this cell of origin (COO) classification in routine clinical practice is to establish a robust clinical assay amenable to routine FFPE (formalin-fixed paraffin- embedded) diagnostic biopsies. In this study, we investigated the possibility of COO-classification using FFPE RNA samples by massive parallel quantitative reverse transcription PCR (qRT-PCR). We established a protocol for parallel qRT-PCR using FFPE RNA samples with the Fluidigm BioMark^{TM} HD system, and quantified the expression of the COO classifier genes and the NF-κB targeted genes that characterise ABC-DLBCL in 143 cases of DLBCL. We also trained and validated a series of basic machine learning classifiers and their derived meta classifiers, and identified SimpleLogistic as the top classifier that gave excellent performance across various GEP datasets derived from FF or FFPE tissues by different microarray platforms. Finally, we applied SimpleLogistic to our dataset generated by qRT-PCR, and the ABC and GCB-DLBCL assigned showed the respective characteristics in their clinical outcome and NF-κB target gene expression. The methodology established in this study provides a robust approach for DLBCL sub-classification using routine FFPE diagnostic biopsies in a routine clinical setting.

INTRODUCTION

Diffuse large B-cell lymphoma (DLBCL) is a heterogeneous entity with remarkably variable clinical outcome. Among the many biomarkers investigated so far, only the molecular subtypes by cell of origin (COO) classification, the *MYC* involved chromosome translocation and *TP53* mutation have been consistently shown to bear prognostic value in the setting of rituximab containing chemotherapy regimens. The COO-classification by whole genome expression profiling (GEP) classifies DLBCL into activated B-cell like (ABC), germinal centre B-cell like (GCB) and Type-III (unclassified) subtypes, with the ABC-DLBCL characterised by a poor prognosis and constitutive NF-κB activation.[1-5] The original classification was based on similarity of DLBCL gene expression to the activated peripheral blood B-cells or normal germinal centre B-cells by hierarchical clustering analysis.[1] Subsequently, Wright and colleagues identified 27 genes that were most discriminative in their expression between ABC and GCB-DLBCL, and developed a linear predictor score (LPS) algorithm for COO-classification.[3] These seminal works are entirely based on retrospective investigations of fresh-frozen (FF) lymphoma tissues. A major challenge for the application of this COO-classification in clinical practice is to establish a robust clinical assay amenable to routine FFPE (formalin fixed paraffin embedded) diagnostic biopsies.

Several immunohistochemistry based algorithms have been investigated to recapitulate the COO-classification by GEP, but all suffer from reproducibility, particularly low efficacy in survival separation between the ABC and GCB subtypes.[6-9] Several studies have investigated the possibility of COO-classification of DLBCL using FFPE tissues by quantitative measurement of mRNA expression, including quantitative nuclease protection assay,[10] gene expression profiling (GEP) with the Affymetrix HG U133 Plus 2.0 platform or the Illumina whole genome DASL assay,[11-13] and

NanoString technology.[14] In general, these studies demonstrated high confidence of COO-classification of DLBCL using FFPE tissues and a robust separation in overall survival between ABC and GCB-DLBCL.

Apart from the above technologies for GEP, the expression of a small number of genes can be quantified by high throughput real-time PCR. In comparison with the microarray based approach, high throughput real-time PCR is likely more sensitive and accurate in data acquisition, and the data analysis is expected to be much easier and robust. Real-time PCR has been successfully used for construction of the 6-gene prediction model in DLBCL,[15] but this model, unlike COO classification, does not depict the underlying molecular mechanism and its utility in the context of new therapeutic trials remains unknown. In this study, we have developed and validated a protocol for COO-classification of DLBCL using FFPE tissues by high throughput real time PCR with the Fluidigm BioMark™ HD system together with a newly validated classifier.

MATERIALS AND METHODS

Patient materials and datasets

A total of 152 cases of DLBCL were experimentally investigated in the present study, and all were previously studied for COO classification by Illumina WG-DASL array using FFPE lymphoma tissues.[13,16] In each case, mRNA sample from FFPE lymphoma tissue was available from the previous study. Local ethical guidelines were followed for the use of archival tissues for research with the approval of the ethics committees of the involved institutions.

The following GEP datasets, which had companion clinical follow up data, were retrieved and used for construction and validation of DLBCL classifier: the FF dataset by Lymphochip (http://llmpp.nih.gov/DLBCLpredictor/),[3] the FF dataset (GSE10846: this was further split according to treatment with CHOP or R-CHOP and analysed independently) by Affymetrix U133 plus 2.0,[17] the Monti FF dataset by Affymetrix U133A&B (http://www.ncbi.nlm.nih.gov/gds/),[18] and the FFPE dataset (GEO: GSE32918) by Illumina WG-DASL array (Haematological Malignancy Diagnostic Service, St James Institute of Oncology, Leeds).[13,16]

Primer design and validation

PCR primer pairs were designed for the 20 classifier genes that are commonly present in different microarray platforms, 5 NF-κB target genes that are characteristically enriched in their expression in ABC-DLBCL,[4,5] and 3 reference genes (Table S1). The reference genes were selected as they are stably expressed in lymphoid tissues, but not affected by genomic copy number changes in lymphoma, nor involved in lymphomagenesis. A set of criteria were followed for the primer design and these included: a) targeting a small fragment of the coding sequence with all amplicons in the range of 70-130bp, thus being amenable to FFPE tissues; b) where possible flanking an intron, hence preventing from amplification of any potentially contaminated genomic DNA; c) targeting as many known transcript variants as possible; d) giving a Tm value at or close to 60°C (http://www.oligoevaluator.com); e) avoiding any known SNPs and GC rich sequence region. The specificity of the primers designed and their potential formation of primer dimmers were checked with Primer Blast (www.ncbi.nlm.nih.gov/tools/primer-blast/) and Oligos 9.1, then further assessed by In-Silico PCR package (http://genome.ucsc.edu/cgi-bin/hgPcr?command=start) before purchase from Thermo Fisher Scientific GmbH.

The PCR primer pairs were experimentally validated for qRT-PCR with RNA samples extracted from FFPE tonsil and lymphoma tissue specimens using the iCycler iQ system (BioRad). The primers failed to give specific results or high efficiency of amplification, or worked inconsistently were rejected, and new primers were designed, until satisfactory primer sets obtained for each of the genes investigated.

RNA extraction, cDNA synthesis and quantitative PCR

RNA was purified from FFPE tissues using the RecoverAll Total Nucleic Acid isolation Kit (Life Technologies), followed by TURBO^{TM}DNase (Life Technologies) treatment to remove genomic DNA. A total of 200ng RNA was subjected for cDNA synthesis in a 10µl reaction mixture with random hexamers using the Superscript III Kit (Invitrogen, Life Technologies) according to the manufacturer's instructions. An aliquot (2µl) of the cDNA was pre-amplified in a 10µl reaction using TaqMan PreAmp mastermix (Life Technologies) with all the 28 pairs of gene specific primers, and the PCR cycle conditions were 95°C for 10 min, followed by 19 cycles of 95°C (15 sec) and 60°C (4 min). This protocol was shown to yield unbiased amplification (Figure S1). The pre-amplified products were treated with exoSAP-IT (Affymetrix), diluted 5 folds using DNA suspension buffer (TEKnova) and stored at -80°C until use for quantitative real time PCR.

For quantitative PCR with iCycler iQ system (BioRad), this was carried out using 2.7µl of the diluted pre-amplified product and SSO Evagreen supermix with low ROX (BioRad) according to the manufacturer's instructions. All reactions were performed in triplicate. Controls included RNA sample to check for genomic DNA contamination and no template cDNA to monitor any across contamination. PCR cycle conditions were 95°C for 1 minute, followed by 30 cycles of 96°C (5 sec) and 60°C (20 sec). Melting curve analysis was routinely preformed to confirm specific amplification.

The expression level of each gene was calculated by the ΔCT method. The amplification efficiency of each primer pair was obtained by quantitative PCR of a serial dilutions of the specific sequence products respectively. The primer pairs validated satisfactorily by experiments with a high amplification efficiency were proceeded to quantitative PCR with the Fluidigm BioMark™ HD system (Fluidigm Corporation, CA, USA).

Massive parallel quantitative PCR with Fluidigm BioMark™ HD system

This was carried out essentially according to the manufacturer's instructions (Figure 1). Briefly, a sample mixture was prepared by mixing 2.7μl of the diluted pre-amplified product, 3μl SSO Evagreen supermix with low ROX (BioRad) and 0.3μl 20 × Sample loading reagent. Each sample was investigated in duplicate. Separately, an assay mixture was prepared for each primer pair and this included 3μl of 10μM forward and reverse primer and 3μl 2× Assay loading reagent. The dynamic array was first primed with control line fluid, and then loaded with the sample and assay mixtures via the appropriate inlets using an IFC controller. The array chip was placed in the BioMark Instrument for PCR at 95°C for 10 min, followed by 30 cycles at 95°C for 15 sec and 60°C for 1 min according to the protocol GE Fast 48x48 PCR + Melt v2.pcl. The data was analyzed with Real-Time PCR Analysis Software in the BioMark™ HD instrument (Fluidigm Corporation, CA, USA).

Normalisation and analysis of Fluidigm qRT-PCR data

This was carried out using the R statistic software and Bioconductor HT-qPCR package. For each sample, the Ct values of the two replicates were averaged and then normalised for each primer pair according to their amplification efficiency. The expression level of each gene was calculated by the ΔCT method using the mean from the three reference genes. Based on the application efficiency in our experimental system, any qPCR reaction with a ΔCt value above the cutoff (25) for linear

amplification was set to 26.[19] For a small proportion of PCR reactions, there was no evidence of amplification at the maximum 30 cycle set by the manufacturer's default protocol, commonly due to low levels of gene expression (for example GCB genes in ABC-DLBCL or vice versa) or rarely as a result of failed amplification. Any cases with more than 15% of targets, i.e. 4/28 genes, showing a negative result were considered unreliable and excluded from data analysis.

Data preparation for DLBCL classifier validation

For each of the GEP datasets by the Affymetrix platform, the probe annotation was updated according to Release 33 (30/10/2012), while for GEP dataset by Illumina WG-DASL assay, the specificity of the classifier gene probes was further checked by search of the NCBI human database. The probe that was found to be non-specific was excluded from the analysis. The median value across the probes for each gene was selected and used for DLBCL classification since it was shown to be more informative in a recent study.[13] The median value for each classifier gene in a given dataset was transformed into a quantile score appose to Z-score as the expression value of a high proportion of the classifier genes in all the datasets tested was not in a normal distribution (Table S2). The quantile transformed values were converted to ARFF files for training and testing with the machine learning classifiers in Weka 3.7.7 (http://www.cs.waikato.ac.nz/ml/weka/).

DLBCL classifier testing and ranking

The LLMPP lymphochip dataset by Wright et al was used for training and selecting the top basic machine learning classifiers using the Weka 3.7.7 package (http://www.cs.waikato.ac.nz/ml/weka/). A total of 26 representative basic machine learning classifiers were trained and cross-validated on the Wright dataset using the default settings (Figure 1). As with the original study, the training and validation were performed in the identical series of 160 and 80 cases respectively.[2,3] The trained

Weka classifiers gave prediction for each of the three classes (i.e. ABC, GCB, Type-III), the class with the highest probability was taken as the predicted class. The performance of these basic classifiers was ranked according to F-Measure and ROC area value as described previously.[13] The resulting top basic machine learning classifiers were systematically combined, trained and cross-validated on the Wright dataset and the resulting best classifier was then tested on the Affymetrix GSE10846 dataset from FF tissues,[17] the Illumina WG-DASL GSE32918 dataset from FFPE tissues,[13,16] and the Affymetrix dataset from FF tissues by Monti et al (http://www.ncbi.nlm.nih.gov/gds/).[18] Finally, the validated best classifier was applied to the qRT-PCR data generated on the FFPE tissues in the present study.

Comparison of NF-κB target gene expression between ABC and GCB-DLBCL

The expression of NF-κB target genes, *BCL2*, *CCDN2*, *CCR7*, *CD44*, *cFLIP*, *IκBα* and *IRF4*, was compared between the assigned ABC and GCB subgroups using non-parametric Mann-Whitney U test.

RESULTS

Identification of the best classifier amenable to datasets from both FF and FFPE tissues

The DLBCL automatic classifier (DAC) was developed for COO classification by Illumina WG-DASL profiling using FFPE tissues.[13] The meta-classifier DAC utilised a balanced voting between the individual classifiers LMT, J48, RF100 and SMO, and was shown highly confidence in classification of GEP data from both FF and FFPE tissues.[13] In the initial analysis, we applied DAC to our qRT-PCR data, but found classification unsatisfactory in comparison with those obtained by WG-DASL.[13]

While exploring the data analysis, we found that the expression value of a high proportion of classifier genes in the published datasets was not in a normal distribution (Table S2), thus not suitable for Z-score transformation. In addition, we found that the models generated by logistic based classifiers, such as Logistic, SimpleLogistic, LMT and FT, were not always compatible with the expected function of individual classifier genes in ABC and GCB assignment. Depending on individual classifiers, a proportion of the ABC classifier genes were given a coefficient favouring GCB rather ABC class assignment or vice versa (Table S3). This is most likely caused by multicollinearity due to high correlation of expression of the classifier genes.

To circumvent the above issues, we first used quantile score for data transformation, which is amenable to quantitative data regardless their distribution. To overcome multicollinearity, we converted the expression of 20 individual classifier genes into two variable indices, by summing the quantile score of all ABC or GCB genes in each case respectively. Based on the characterised expression pattern of the classifier genes, one would expect that ABC-DLBCL is featured by high ABC but low GCB index, GCB-DLBCL by high GCB but low ABC index, and Type-III DLBCL by both low ABC and GCB indices.

Based on the above principle, we systematically trained and validated a series of basic and meta machine learning classifiers and identified the classifier that gave the best performance across different microarray platforms and datasets. Of the 27 classifier genes defined by Wright et al,[3] only 20 were commonly present among the Lymphochip, Affymetrix and Illumina WG-DASL platforms. A previous study showed that the FFPE dataset was best classified using the 20 classifier genes with the LLMPP lymphochip dataset by Wright et al for training.[13] Accordingly, we calculated the ABC and GCB indices based on the same 20 classifier genes and used the Wright dataset for initial

classifier training (n=160 cases) and validation (n=80 cases). We first tested a total of 26 representative basic machine learning classifiers and ranked them according to F-Measure and ROC area value.[13] There were 7 basic classifiers showing F-Measure and ROC area value above 0.80 and 0.90 respectively, and 6 of these classifiers, namely NaiveBayes, Logistic, FT, MultilayerPerceptron, RBFnetwork, SimpleLogistic, yielded Type-III class at a relatively low frequency. These 6 classifiers were systematically combined and trained, and validated on the Wright dataset.

All the meta classifiers except one showed F-Measure and ROC area values similar to the top 6 basic classifiers (Table-S5). The performance of the 6 basic and 16 of their derived meta classifiers were ranked according to survival separation between the assigned ABC and GCB subtypes and the least number of cases assigned to Type-III. This identified SimpleLogistic as the best choice, and we then further tested this classifier on the Affymetrix GSE10846 dataset based on FF tissues (including two cohorts: one treated with CHOP and the other treated with R-CHOP),[17] and the Illumina WG-DASL GSE32918 dataset based on FFPE tissues.[13,16] SimpleLogistic consistently gave excellent performance for all three datasets, overall better than the original class assignment as measured by overall survival separation between the ABC and GCB groups, and the least number of cases assigned to Type-III (Figure 2, Table S6).

We further tested SimpleLogistic on the dataset by Monti et al.[18] The Monti dataset lacked 4 of the 20 classifier genes including absence of GCET1. Based on the 16 classifier genes, SimpleLogistic also showed a comparable ABC/GCB/Type-III assignment and survival separation between the ABC and GCB class (Table S6), confirming robustness of the classifier.

Finally, to further testify the ABC/GCB class assigned by SimpleLogistic, we compared the NF-κB target gene expression between the two classes by SimpleLogistic and the original classifier. For each of the three datasets investigated, SimpleLogistic and the original classifier showed nearly identical results on the NF-κB target gene expression between the ABC and GCB groups assigned, with the expression of *IRF4*, *CCND2*, *CD44*, *cFLIP*, *BCL2* and *CCR7* being significantly higher in ABC than GCB-DLBCL, but no difference in the *NFKB1A* expression between the two groups (Figure 2). These results are very similar to the NF-κB target gene signature in DLBCL ABC/GCB subtypes, which was defined in the original studies.[4,5]

DLBCL COO classification by qRT-PCR with Fluidigm Dynamic Array

In total, 143 of the 152 cases included were successfully investigated by qRT-PCR with the Fluidigm BioMark™ HD system, with the remaining 9 cases failed due to insufficient quantity of RNA or incomplete data acquisition. We applied the SimpleLogistic classifier to this qRT-PCR dataset as described above. The distribution of ABC (28.7%), GCB (53.1%) and Type-III subtype assigned by the qRT-PCR/SimpleLogistic classifier was nearly identical to those by Illumina WG-DASL/DAC classifier.

Of the 143 cases successfully investigated by qRT-PCR, 120 were treated with R-CHOP and were further analysed for correlation between treatment outcome and COO subtype. As shown in Figure 3B, there was a significant difference in the overall survival between the ABC and GCB-DLBCL assigned by the qRT-PCR/SimpleLogistic classifier. As expected, the expression of NF-κB target genes *IRF4*, *CCND2*, *CD44*, *cFLIP* and *CCR7,* with the exception of *BCL2*, was significantly higher in ABC than GCB-DLBCL (Figure 3).

DISCUSSION

COO-classification of DLBCL has two significant clinical implications. First, the classification divides DLBCL into different prognostic subgroups with ABC-DLBCL showing worse survival than GCB-DLBCL in both CHOP and R-CHOP treatment settings; Second, ABC-DLBCL is characterised by constitutive NF-κB activation, and may be treated by inhibitors of the NF-κB pathway. For example, addition of bortezomib to chemotherapy significantly improved the treatment response and overall survival of the patients with ABC-DLBCL, but not those with GCB-DLBCL.[20] Currently, a prospective phase-III randomised controlled clinical trial, known as REMoDL-B, is being conducted to assess the clinical efficacy between R-CHOP and bortezomib plus R-CHOP according to ABC and GCB molecular subtype. Therefore, the survival separation between the ABC and GCB subtype and their difference in NF-κB activities are the key parameters in assessment of methodologies for COO-classification of DLBCL. There are many factors affecting the performance of COO-classification and the critical elements include the nature of lymphoma specimen, experimental methods for data collection, data normalisation and transformation, classifier and the level of probability used for subtype assignment.

Immunohistochemical study of the surrogate protein markers and several algorithms have been extensively investigated to recapitulate the COO-classification by GEP, but all suffered from reproducibility and low efficacy in survival separation between the ABC and GCB-DLBCL.[8,9] GEP using RNA samples from FF tissues has been shown highly consistent in COO-classification, and this approach has also been applied to FFPE tissues with very encouraging results. Nonetheless, the drawback of the GEP approach is cumbersome in experimental setup and data analyses, and additionally not cost effective. For clinical application, a quantitative measurement of the

expression of the classifier genes, rather than the whole genome, would be preferable. In a recent study, Care et al showed that a panel of 20 of the 27 classifier genes defined by Wright et al gave the best performance for COO-classification after testing a series of classifier genes.[3,13]

To develop a tailor-made clinical assay for COO-classification of DLBCL, we established a robust protocol to measure the expression of the 20 classifier genes together with the NF-κB target genes characteristically over-expressed in ABC-DLBCL, using RNA samples from FFPE tissues by parallel qRT-PCR with the Fluidigm BioMark™ HD system. We demonstrated that the expression of classifier genes could be reproducibly measured using the protocol established. In comparison with the microarray based GEP, the qRT-PCR based approach offers several notable advantages including high sensitivity and reproducibility, easy to perform and cheap to run, and a short turnaround time.

Apart from high quality of data acquisition, generation of an accurate classifier is another challenge for application of COO-classification to qRT-PCR dataset from FFPE tissue as which classifier to use appears to depend on the method/platform of data acquisition and the nature (FF or FFPE) of the tissue specimen used. For example, the LPS algorithms successfully used on the LLMPP FF dataset by Lymphochip was not suitable for the FFPE dataset by Illumina WG-DASL.[13] Similarly, we found that the meta classifier DAC developed based on the FFPE dataset by Illumina WG-DASL was not amenable to our qRT-PCR dataset. The potential reasons accounting for these incompatibilities are many and the critical ones may include variations in data distribution, normalisation and transformation, as well as classifier selection and validation.

There are many machine learning classifiers and a proportion of them are based on logistic model, which is potentially suitable for assessing the relationship between categorical variables (DLBCL

subtypes) and their dependent variables (expression of classifier genes). However, their utility in COO-classification of DLBCL is restricted by the presence of multicollinearity among the expression of the classifier genes (Table S3). To overcome this, we combined all the ABC and GCB gene expression values into two indices respectively, thus making the data amenable to a wide range of machine learning classifiers.

To develop a classifier that is amenable to the qRT-PCR based dataset, we have systematically trained and validated 26 basic machine learning classifiers and their derived meta classifiers using the LLMPP lymphochip dataset by Wright et al.[2,3] As with the study by Care et al,[13] the selection of top classifier was based on the significance of survival separation between the ABC and GCB groups, and the least number of cases assigned to Type-III. This combined assessment avoided selection bias toward classifiers that gave significant survival separation between the ABC and GCB groups, but at the expense of the numbers of cases successfully assigned to these biological subtypes. The top classifier identified was then tested with additional GEP datasets from both FF (by Affymetrix platform) and FFPE lymphoma specimens (Illumina WG-DASL). Analyses of survival separation and NF-κB target gene expression between the ABC and GCB groups assigned, and the distribution of the three molecular subtypes consistently showed that SimpleLogistic, the top classifier identified, gave a similar or better performance than the respective original classifier. Finally, the assured top classifier was applied to our qRT-PCR dataset. As the final classifier used for our qRT-PCT dataset was sequentially validated using a series of published datasets derived from both FF and FFPE specimens by different microarray platforms, the above process of classifier generation has little risk of overfitting, a common issue in classifier generation.

Despite that SimpleLogistic and original classifier yielded similar performance including survival separation and differential NF-κB target gene expression between the ABC and GCB subtypes in a range of datasets, the concordant rate in class assignment between these different classifiers was at 80% across all datasets investigated. These findings are very similar to those found in previous studies irrespective of the nature (fresh frozen or FFPE) of DLBCL specimens investigated.[10,13,14,21] In general, the samples showing discrepant classes by different classifiers had a low confidence score for either ABC or GCB class assignment.[13] It is possible that these cases are in a "molecular grey zone", and their class cannot be accurately defined no matter which classifiers used. In support of this speculation, there is emerging evidence showing overlap in somatic mutation profiles between ABC and GCB DLBCL.[22,23] Nonetheless, it remains to be investigated whether the cases not amenable for an accurate COO classification represent an intermediate subset with overlapping genetic changes.

In summary, we have developed a robust protocol for COO-classification of DLBCL using RNA samples from FFPE tissues by qRT-PCR using the Fluidigm BioMark™ HD system with SimpleLogistic classifier. The ABC and GCB-DLBCL assigned show the respective characteristics in their clinical outcome and NF-κB target gene expression. The methodology may provide a robust approach for DLBCL sub-classification using routine FFPE diagnostic biopsies in a routine clinical setting.

Acknowledgements: The research in Du lab was supported by grants from Leukemia & Lymphoma Research, U.K.. We are grateful to Dr Sharon Barrans, Dr Lisa Worrillow and Professor Andrew Jack for providing RNA samples of DLBCL used in this study. We thank Prof. Berthold Göttgens, Prof. Tony Green and Dr Brian Huntly for access of their Fluidigm BioMark™ HD system, and Victoria

Moignard for technical help. We also thank Dr Matthew Care and Dr Reuben Tooze for discussing the data analyses.

Authors contributions: XX designed and carried out all bench experiments and data analyses, and prepared the figures and tables; NZ helped initial qPCR experiments; ZG and MQD supervised the study; MQD designed the study, analysed data and wrote the manuscript. All the authors commented on the manuscript and approved its submission.

References:

1. Alizadeh AA, Eisen MB, Davis RE et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000;403:503-511.

2. Rosenwald A, Wright G, Chan WC et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med* 2002;346:1937-1947.

3. Wright G, Tan B, Rosenwald A et al. A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. *Proc Natl Acad Sci U S A* 2003;100:9991-9996.

4. Davis RE, Brown KD, Siebenlist U, Staudt LM. Constitutive nuclear factor kappaB activity is required for survival of activated B cell-like diffuse large B cell lymphoma cells. *J Exp Med* 2001;194:1861-1874.

5. Feuerhake F, Kutok JL, Monti S et al. NFkappaB activity, function, and target-gene signatures in primary mediastinal large B-cell lymphoma and diffuse large B-cell lymphoma subtypes. *Blood* 2005;106:1392-1399.

6. Hans CP, Weisenburger DD, Greiner TC et al. Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. *Blood* 2004;103:275-282.

7. Choi WW, Weisenburger DD, Greiner TC et al. A new immunostain algorithm classifies diffuse large B-cell lymphoma into molecular subtypes with high accuracy. *Clin Cancer Res* 2009;15:5494-5502.

8. Gutierrez-Garcia G, Cardesa-Salzmann T, Climent F et al. Gene-expression profiling and not immunophenotypic algorithms predicts prognosis in patients with diffuse large B-cell lymphoma treated with immunochemotherapy. *Blood* 2011;117:4836-4843.

9. Coutinho R, Clear A, Owen A et al. Poor concordance among nine immunohistochemistry classifiers of cell-of-origin for Diffuse Large B-cell Lymphoma: implications for therapeutic strategies. *Clin Cancer Res* 2013;19:6686-6695.

10. Rimsza LM, Wright G, Schwartz M et al. Accurate classification of diffuse large B-cell lymphoma into germinal center and activated B-cell subtypes using a nuclease protection assay on formalin-fixed, paraffin-embedded tissues. *Clin Cancer Res* 2011;17:3727-3732.

11. Linton K, Howarth C, Wappett M et al. Microarray gene expression analysis of fixed archival tissue permits molecular classification and identification of potential therapeutic targets in diffuse large B-cell lymphoma. *J Mol Diagn* 2012;14:223-232.

12. 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;159:441-453.

13. Care MA, Barrans S, Worrillow L et al. 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:e55895.

14. Scott DW, Wright GW, Williams PM et al. Determining cell-of-origin subtypes of diffuse large B-cell lymphoma using gene expression in formalin-fixed paraffin-embedded tissue. *Blood* 2014;123:1214-1217.

15. Malumbres R, Chen J, Tibshirani R et al. Paraffin-based 6-gene model predicts outcome in diffuse large B-cell lymphoma patients treated with R-CHOP. *Blood* 2008;111:5509-5514.

16. Barrans S, Crouch S, Smith A et al. Rearrangement of MYC is associated with poor prognosis in patients with diffuse large B-cell lymphoma treated in the era of rituximab. *J Clin Oncol* 2010;28:3360-3365.

17. Lenz G, Wright G, Dave SS et al. Stromal gene signatures in large-B-cell lymphomas. *N Engl J Med* 2008;359:2313-2323.

18. Monti S, Savage KJ, Kutok JL et al. Molecular profiling of diffuse large B-cell lymphoma identifies robust subtypes including one characterized by host inflammatory response. *Blood* 2005;105:1851-1861.

19. Moignard V, Macaulay IC, Swiers G et al. Characterization of transcriptional networks in blood stem and progenitor cells using high-throughput single-cell gene expression analysis. *Nat Cell Biol* 2013;15:363-372.

20. Dunleavy K, Pittaluga S, Czuczman MS et al. Differential efficacy of bortezomib plus chemotherapy within molecular subtypes of diffuse large B-cell lymphoma. *Blood* 2009;113:6069-6076.

21. Abramson JS, Shipp MA. Advances in the biology and therapy of diffuse large B-cell lymphoma: moving toward a molecularly targeted approach. *Blood* 2005;106:1164-1174.

22. Pasqualucci L, Trifonov V, Fabbri G et al. Analysis of the coding genome of diffuse large B-cell lymphoma. *Nat Genet* 2011;43:830-837.

23. Zhang J, Grubor V, Love CL et al. Genetic heterogeneity of diffuse large B-cell lymphoma. *Proc Natl Acad Sci U S A* 2013;110:1398-1403.

FIGURE LEGENDS:

Figure 1. Outline of classifier training, validation and testing, and application of the best classifier to qRT-PCR data. The LLMPP lymphochip dataset by Wright et al was used for training and selecting the top basic machine learning classifiers using the Weka 3.7.7 package (http://www.cs.waikato.ac.nz/ml/weka/). A total of 26 representative basic machine learning classifiers were trained and cross-validated on the Wright dataset using the default settings. The trained Weka classifiers gave prediction for each of the three classes (i.e. ABC, GCB, Type-III), the

class with the highest probability was taken as the predicted class. The performance of these basic classifiers was ranked according to F-Measure and ROC area value. The resulting 6 top basic machine learning classifiers were systematically combined, trained and cross-validated on the Wright dataset. All the meta classifiers except one yielded F-Measure and ROC area values similar to the top 6 basic classifiers. The performance of these meta classifiers together with the 6 top basic classifiers were ranked according to survival separation between the assigned ABC and GCB cases and the least number of cases assigned to Type 3. The resulting best classifier, SimpleLogistic, was then tested on the Affymetrix GSE10846 dataset from FF tissues,[17] the Illumina WG-DASL GSE32918 dataset from FFPE tissues,[13,16] and the Affymetrix dataset from FF tissues by Monti et al (http://www.ncbi.nlm.nih.gov/gds/).[18] Finally, the validated best classifier was applied to the qRT-PCR data generated on the FFPE tissues in the present study.

Figure 2. Testing of SimpleLogistic classifier for COO classification of DLBCL using published datasets. The probability of ABC/GCB/Type-III assignment is shown on the top of heatmap illustration in each dataset. Sum_ABC: the summed ABC gene score; Sum_GCB: the summed GCB gene score. The middle panel shows comparison of overall survival among the ABC, GCB and Type-III groups. The table on the right displays the significant difference in the expression of the NF-κB target genes between ABC and GCB-DLBCL assigned. *Benjamini and Hochberg method was used to correct multiple testing problems.

Figure 3. COO classification of DLBCL using RNA samples from FFPE tissues by parallel qRT-PCR and SimpleLogistic classifier. A) Heatmap shows the level of expression of the 20 classifier genes used for classification; B) Difference in overall survival between ABC, GCB and Type3 DLBCL in 120 cases

of DLBCL treated with R-CHOP; C) Difference in NF-κB target gene expression between ABC and GCB-DLBCL. *Benjamini and Hochberg method was used to correct multiple testing problems.

Supplementary Figure S1: Confirmation of uniformity of pre-amplification for target enrichment prior to qPCR with the Fluidigm BioMark™ HD system. To increase the sensitivity for qRT-PCR using FFPE tissues, a pre-amplification of cDNA sample with gene specific primers was performed before qPCR with the Fluidigm BioMark™ HD system. To confirm the pre-amplification step did not introduce any bias, we designed the experiment outlined in panel A and compared the qPCR measurements among the three protocols on three representative samples. As shown in panel B, △△Ct_1: BioRad(original) –FluidigmBioMark(with_preamp), △△Ct_2: BioRad (with_preamp) – FluidigmBioMark(with preamp), and △△Ct_3: BioRad(original) – FluidigmBioMark(with preamp) for each of the genes investigated were within the range of 1.5 and -1.5 , with the majority being close to 0. There is a strong correlation among the data from the three protocols. Taken together, the data indicate that no biased amplification was introduced by this pre-amplification step.

Preamp: pre-amplification.

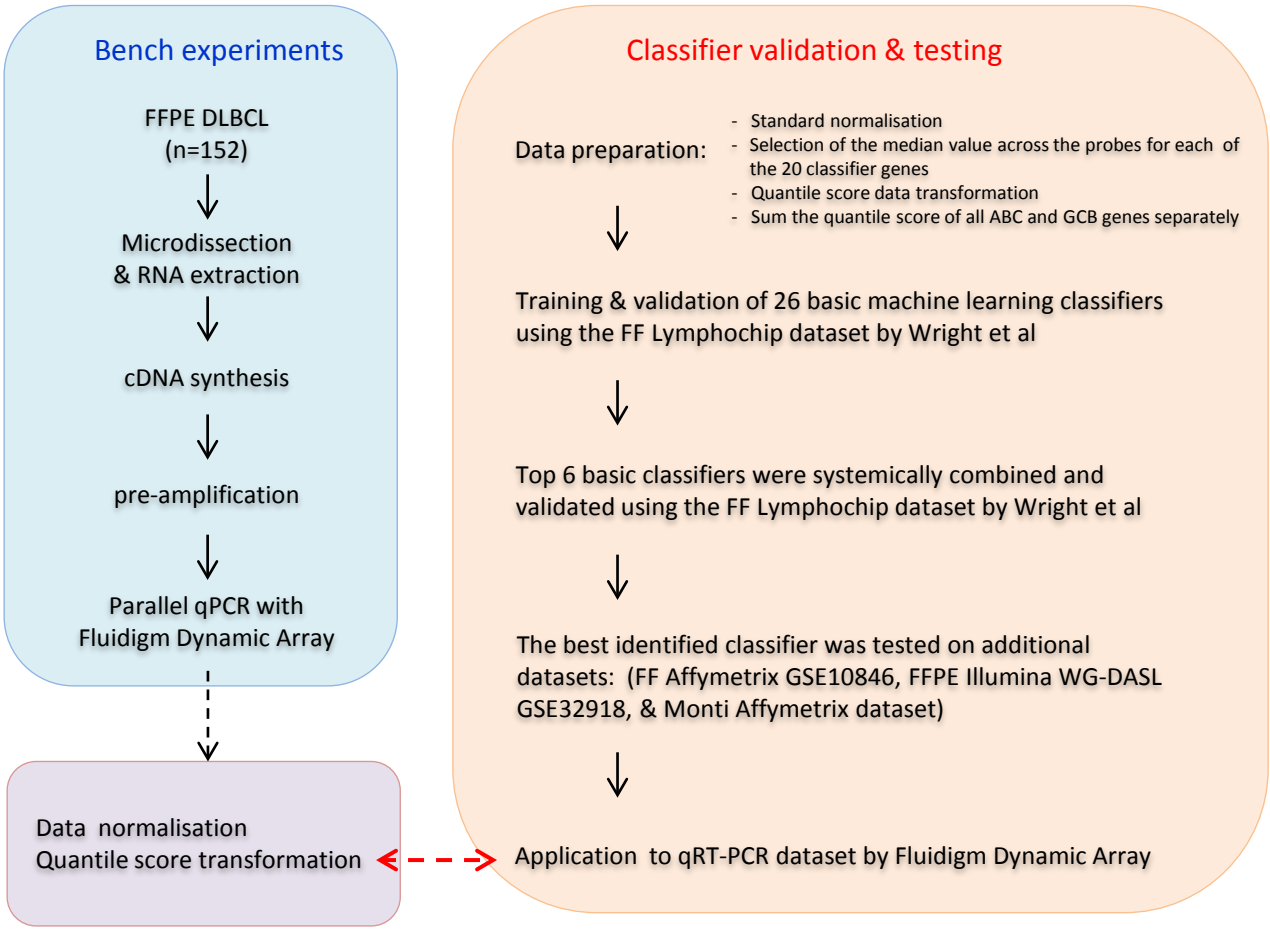


Figure 1. Outline of classifier training, validation and testing, and application of the best classifier to qRT-PCR data. The LLMPP lymphochip dataset by Wright et al was used for training and selecting the top basic machine learning classifiers using the Weka 3.7.7 package (http://www.cs.waikato.ac.nz/ml/weka/). A total of 26 representative basic machine learning classifiers were trained and cross-validated on the Wright dataset using the default settings. The trained Weka classifiers gave prediction for each of the three classes (i.e. ABC, GCB, Type-III), the class with the highest probability was taken as the predicted class. The performance of these basic classifiers was ranked according to F-Measure and ROC area value. The resulting 6 top basic machine learning classifiers were systematically combined, trained and cross-validated on the Wright dataset. All the meta classifiers except one yielded F-Measure and ROC area values similar to the top 6 basic classifiers. The performance of these meta classifiers together with the 6 top basic classifiers were ranked according to survival separation between the assigned ABC and GCB cases and the least number of cases assigned to Type 3. The resulting best classifier, SimpleLogistic, was then tested on the Affymetrix GSE10846 dataset from FF tissues,[17] the Illumina WG-DASL GSE32918 dataset from FFPE tissues,[13,16] and the Affymetrix dataset from FF tissues by Monti et al (http://www.ncbi.nlm.nih.gov/gds/).[18] Finally, the validated best classifier was applied to the qRT-PCR data generated on the FFPE tissues in the present study.

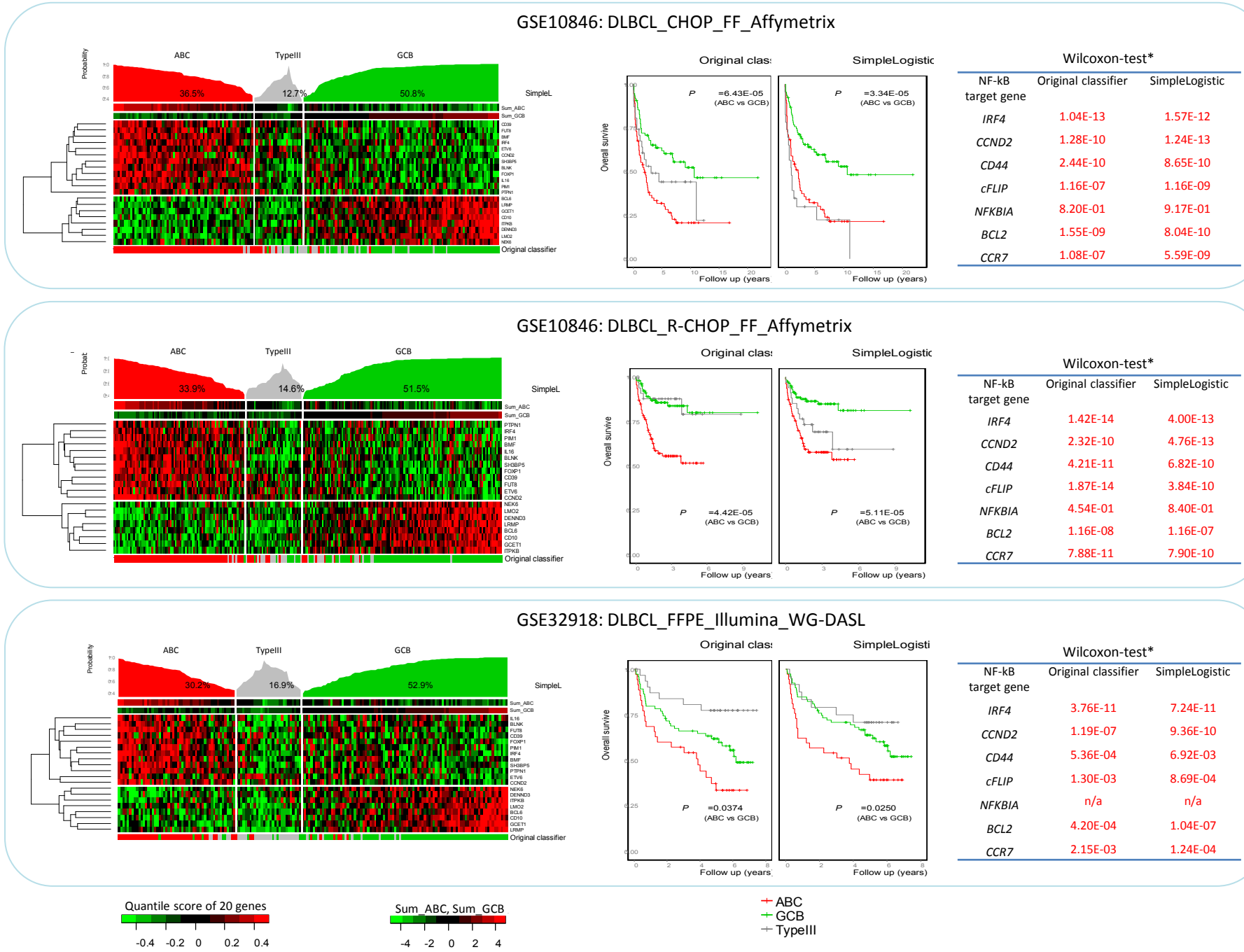


Figure 2. Testing of SimpleLogistic classifier for COO classification of DLBCL using published datasets. The probability of ABC/GCB/Type-III assignment is shown on the top of heatmap illustration in each dataset. Sum_ABC: the summed ABC gene score; Sum_GCB: the summed GCB gene score. The middle panel shows comparison of overall survival among the ABC, GCB and Type-III groups. The table on the right displays the significant difference in the expression of the NF-κB target genes between ABC and GCB-DLBCL assigned. *Benjamini and Hochberg method was used to correct multiple testing problems.

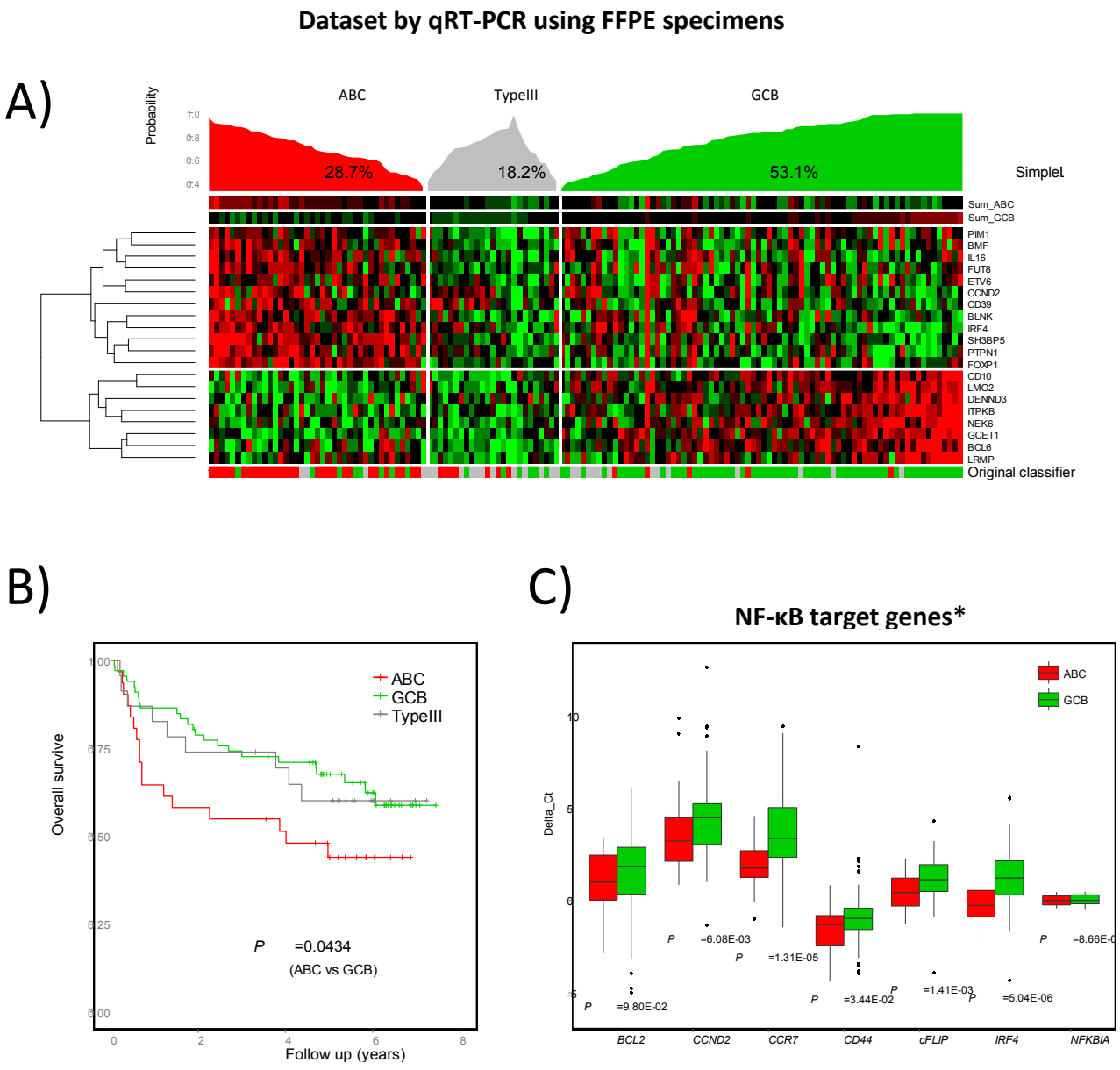
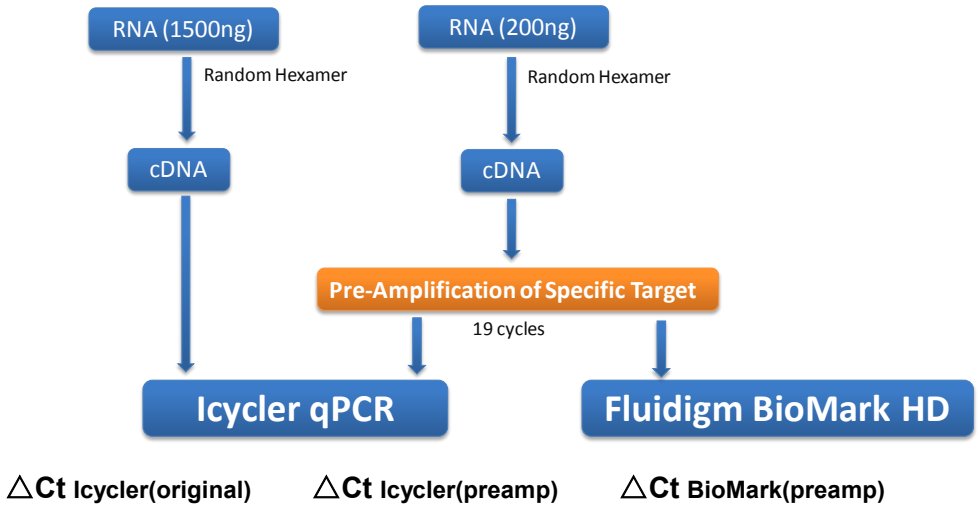
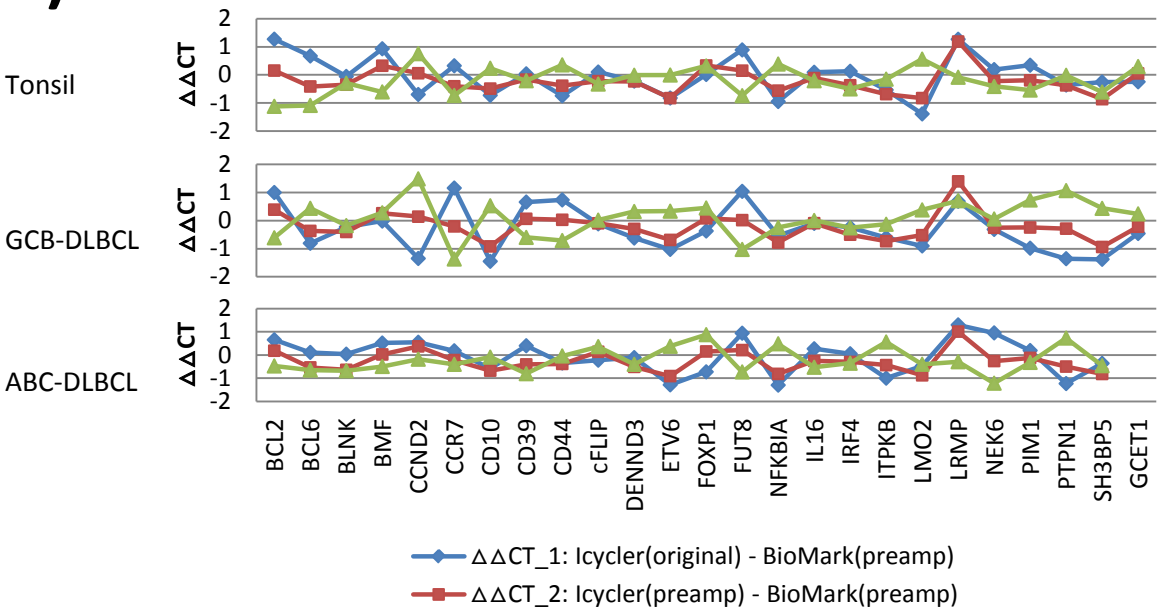


Figure 3. COO classification of DLBCL using RNA samples from FFPE tissues by parallel qRT-PCR and SimpleLogistic classifier. A) Heatmap shows the level of expression of the 20 classifier genes used for classification; B) Difference in overall survival between ABC, GCB and Type3 DLBCL in 120 cases of DLBCL treated with R-CHOP; C) Difference in NF-κB target gene expression between ABC and GCB-DLBCL. *Benjamini and Hochberg method was used to correct multiple testing problems.

A) Testing uniformity of pre-amplification



B)



Correlation analysis among the 3 experiments

	Pearson correlation (r)	*P* value	Spearman correlation (ρ)	*P* value
△Ct Icycler(original) vs △Ct Icycler(preamp)	0.9409327	p< 2.2e-16	0.9431618	p<2.2e-16
△Ct Icycler(original) vs △Ct BioMark(preamp)	0.8696793	p< 2.2e-16	0.9503272	p<2.2e-16
△Ct Icycler(preamp) vs △Ct BioMark(preamp)	0.9014614	p< 2.2e-16	0.8852869	p<2.2e-16

Supplementary Figure S1: Confirmation of uniformity of pre-amplification for target enrichment prior to qPCR with the Fluidigm BioMark™ HD system. To increase the sensitivity for qRT-PCR using FFPE tissues, a pre-amplification of cDNA sample with gene specific primers was performed before qPCR with the Fluidigm BioMark™ HD system. To confirm the pre-amplification step did not introduce any bias, we designed the experiment outlined in panel A and compared the qPCR measurements among the three protocols on three representative samples. As shown in panel B, △△Ct_1: BioRad(original) – FluidigmBioMark(with_preamp), △△Ct_2: BioRad (with_preamp) – FluidigmBioMark(with preamp), and △△Ct_3: BioRad(original) – FluidigmBioMark(with preamp) for each of the genes investigated were within the range of 1.5 and -1.5 , with the majority being close to 0. There is a strong correlation among the data from the three protocols. Taken together, the data indicate that no biased amplification was introduced by this pre-amplification step.