2015•2016
FACULTEIT GENEESKUNDE EN LEVENSWETENSCHAPPEN
master in de biomedische wetenschappen

Masterproef

Molecular Subtype Classification of Diffuse Large B-Cell Lymphoma (DLBCL)

Promotor :
Prof. dr. Luc MICHIELS

Promotor :
Prof. dr. THOMAS TOUSSEYN

Copromotor :
Dr. JULIE MORSCIO

Aran Abengochea Molar

Scriptie ingediend tot het behalen van de graad van master in de biomedische wetenschappen

De transnationale Universiteit Limburg is een uniek samenwerkingsverband van twee universiteiten in twee landen: de Universiteit Hasselt en Maastricht University.

universiteit hasselt
KNOWLEDGE IN ACTION

Universiteit Hasselt | Campus Hasselt | Martelarenlaan 42 | BE-3500 Hasselt
Universiteit Hasselt | Campus Diepenbeek | Agoralaan Gebouw D | BE-3590 Diepenbeek

universiteit hasselt | Maastricht University

2015•2016
FACULTEIT GENEESKUNDE EN LEVENSWETENSCHAPPEN
master in de biomedische wetenschappen

Masterproef

Molecular Subtype Classification of Diffuse Large B-Cell Lymphoma (DLBCL)

Promotor :
Prof. dr. Luc MICHIELS

Promotor :
Prof. dr. THOMAS TOUSSEYN

Copromotor :
Dr. JULIE MORSCIO

Aran Abengochea Molar

Scriptie ingediend tot het behalen van de graad van master in de biomedische wetenschappen

universiteit hasselt | Maastricht University

Aran Abengochea Molar
〔1439066〕

Molecular Subtype Classification of Diffuse Large B-Cell Lymphoma (DLBCL)

SENIOR PRACTICAL TRAINING

MASTER THESIS

Supervision by:
Prof. Thomas Tousseyn (External promoter)
Dr. Julie Morscio (Co-external promoter)
& Luc Michiels (Institutional promoter)

MASTER OF BIOMEDICAL SCIENCES

9th November – 8th June

2015 - 2016

ACKNOWLEDGEMENTS

Firstly, I would like to thank my promoter Prof. Dr. Thomas Tousseyn for bringing me on board and including me on this project. Thank you for your frank supervision, guidance and suggestions.

I would also like to thank to my institutional promoter Luc Michiels for acting as a liaison to UHasselt and the work that entailed. A further thanks for your time in assessing this thesis.

Thanks to all my colleagues at the Pathology Institute at KU Leuven for welcoming me, hosting me and making me feel as part of the team. I could not have hoped for a better internship and I am truly grateful for the experience I gained here.

To Sara, Hanne, Vicky and Wim at the Molecular Diagnostics lab of UZ Leuven, thank you for your direction and assistance in setting up and establishing a working MLPA protocol for my experiments.

Additionally, my thanks to Dr. Sylvain Mareschal for his invaluable help, for answering my questions quickly, thoroughly and articulately, and for enduring my amateur programming and bioinformatics skills. Only through his correspondence was I able to optimise the MLPA software for this project.

Finally, I would like to extend a special thanks to my co-promoter Dr. Julie Morscio. You have been a source of admiration, inspiration and unwavering support during this internship and at every stage of the writing process. I am truly indebted for your time, patience and dedicated involvement even when you already had so much on your plate. You were a cool head when the pressure began to build and you constantly pushed me to improve. I can honestly say this thesis would not have been what it is without you.

RESEARCH PLACEMENT

All research was undertaken at the Tissue and Translational Research group as part of my senior internship. This research unit belongs to the Imaging and Pathology department (Biomedical Sciences group) from KU Leuven University, located on the 5th floor of the UZ Leuven campus Gasthuisberg in Leuven.

UZ Leuven campus Gasthuisberg
Herestraat 49
B-3000 Leuven (BE)

Dr. Julie Morscio (julie.morscio@kuleuven.be), a post-doctoral researcher of the unit, together with Prof. Dr. Thomas Tousseyn (thomas.tousseyn@uzleuven.be) supervised this research project as part of my Senior Practical Training.

TABLE OF CONTENTS

ABSTRACT

Background: Diffuse large B-cell lymphoma (DLBCL) is subdivided into germinal center B-cell-like (GCB) and activated B-cell-like (ABC) DLBCL, associated with different survival. The Epstein-Barr virus (EBV) is found in a fraction of DLBCLs that most commonly arise in immunocompromised individuals, e.g. organ transplant recipients (post-transplant, PT-DLBCL). Most EBV-positive DLBCLs are of ABC origin. Recently, biologically distinct subgroups of EBV-associated DLBCL were discovered. ABC and GCB DLBCL are distinguished using immunostainings (Hans algorithm: CD10, BCL6, MUM1), however this approach is often unreliable. Recently, reverse transcriptase multiplex ligation-dependent probe amplification (RT-MLPA) using eight probes was proposed as a more accurate method to classify DLBCL (PMID: 25891505). We aimed to optimize this assay at UZ Leuven and to explore its utility in EBV-positive and –negative DLBCL.

Materials and Methods: RT-MLPA was performed following a reported protocol (PMID: 25891505) with minor modifications on twelve fresh frozen (FF) and formalin-fixed paraffin-embedded (FFPE) DLBCL samples. A reported model class predictor based on the 8-gene-set was used to discriminate GCB and ABC subgroups. Following optimisation, the RT-MLPA was performed on a series of seventeen PT-DLBC (7 EBV^-, 17 EBV^+) cases and five lymphoblastoid cell lines (LCLs). All cases were stained for CD10, BCL6, MUM1; gene expression profiling data were available for nineteen cases.

Results: In a series of 12 EBV-negative FF IC-DLBCL biopsies, RT-MLPA classified 73% and 80% of the cases using a threshold of 90% and 80% respectively in agreement with the Hans algorithm (70% for FFPE using both thresholds). For EBV-positive DLBCL (n=17) the class-predictor correctly classified only 42% (threshold 80%) and 37% (threshold 90%); 37% and 45% of the cases were misclassified (90% and 80%, respectively). All misclassified cases corresponded to one of the recently identified subgroups within EBV-positive DLBCL. As expected, all LCL were classified as ABC. Four independent repeats for two samples and a half dilution series of three samples demonstrated reproducibility and robustness of the assay.

Conclusion: These results demonstrate the ability of this gene expression-based predictor to classify EBV-negative DLBCLs into clinically distinct subgroups. Caution is required, as the presence of EBV greatly impacts RT-MLPA-based DLBCL classification.

ABBREVIATION LIST

ABC – Activated B-cell
BL – Burkitt Lymphoma
CI – Confidence interval
CO – Competitor oligonucleotide
COO – Cell of Origin
CSR – Class Switch Recombination
CTL – Cytotoxic T-cell Lymphocyte
DLBCL – Diffuse Large B-Cell Lymphoma
EBV – Epstein Barr Virus
EBNA – EBV Nuclear Antigen
EBER – EBV Encoded RNA
FDC – Follicular dendritic cell
FFPE – Formalin-fixed paraffin-embedded
FF – Fresh frozen
GC – Germinal centre
GCB – Germinal centre B-cell
HL – Hodgkin Lymphoma
IC – Immunocompetent
IG – Immunoglobulin
ISH – In situ Hybridisation
IM – Infectious mononucleosis
IHC – Immunohistochemistry
LMP – Latent Membrane Protein
LPD – Lymphoproliferative disorder
LPS – Linear predictor score
NHL – Non-Hodgkin Lymphoma
MLPA – Multiplex Ligation-dependent Probe Amplification
R-CHOP – Rituximab, Cyclophospamide, Hydroxydaunomycin, Oncovin, Prednisone
RIN – RNA Integrity Number
RIS – Reduction of Immunosuppression
RFU – Relative fluorescence units
RT – Reverse Transcriptase
RU – Relative units
PCNSL – Primary Central Nervous System Lymphoma
PT – Post-transplant
PTLD – Post-transplant lymphoproliferative disorder
SHM – Somatic Hyper Mutation
UZ – Universitair Ziekenhuis
UPR – Unfolded Protein Response
WHO – World Health Organization

INTRODUCTION

DIFFUSE LARGE B-CELL LYMPHOMA

The term lymphoma refers to a form of cancer that affects immune cells called lymphocytes, a type of white blood cell. Lymphomas are most commonly of B-cell origin, and are generally classified as either Hodgkin lymphoma (HL) or non-Hodgkin lymphoma (NHL), each of them encompassing several different subtypes depending on the nature of the lymphoid cell affected.

Diffuse large B-cell lymphoma (DLBCL) is the most frequent type of NHL in adults, currently accounting for approximately 40% of B-cell tumours worldwide [1]. DLBCL is an aggressive (fast-growing) type of lymphoma that can develop at either nodal or extra-nodal sites, and morphologically, consists of sheets of clonal B-cells that grow diffusely, partly or completely effacing the typical follicular architecture of the lymphoid tissue (Figure 1).

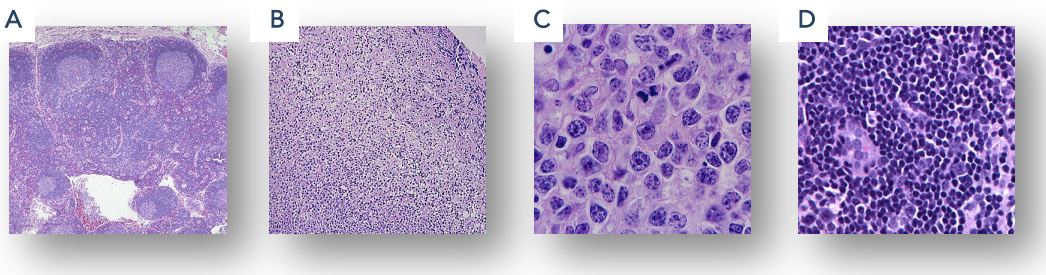


Figure 1. The normal architecture of the lymph node (A) (50x) with germinal centres in the cortex is completely effaced in DLBCL (B) (50x). The neoplastic lymphocytes of DLBCL (C) (200x) are much larger and present a more irregular nucleus and coarse chromatin, than those found in a normal lymph node (D) (200x).

As acknowledged in the 2008 World Health Organization (WHO) classification [2], DLBCL is recognized to include more than one disease entity, as suggested by a marked diversity on clinical, morphological and biological presentation [3]. Based on these features, DLBCLs are further subdivided into several categories, each of them further comprising clinical and pathological variants that make the organization of these neoplasms quite complex.

Despite the continuous development of new treatments and a significant improvement on prognosis over the past decade, durable remissions are achieved in only 40% to 50% of the patients, highlighting that a grand portion of DLBCL patients that either do not respond well to therapies, relapse rapidly or even succumb to the disease [4]. These differences in clinical evolutions can be partly explained by the heterogeneity of this type of tumour. Notably, the cell of origin of DLBCL (GCB versus ABC, see below) has been associated with patient clinical outcome. This stresses the importance of identifying at diagnosis, which patients may benefit from more aggressive or experimental therapies [5, 6].

B-CELL ONTOGENY: RELEVANT FOR UNDERSTANDING OF DLBCL

Our ability to comprehend the pathogenesis of B-cell malignancies relies deeply on the rich and extensive literature regarding normal B cell biology, and in turn, increasing understanding in B-cell differentiation has accelerated our knowledge on lymphomas.

Naïve B-cells are small resting lymphocytes that circulate in the blood to surveil for potential encounters with foreign entities (antigens). Following interaction, B-cell activation can follow two different paths (Figure 2). Some antigens are able to trigger B cells to proliferate and differentiate into antibody-secreting plasma cells without T-cell cooperation (T-cell independent pathway) (Figure 2.B). Once the B-cell reacts with such antigens, it becomes a proliferating activated immunoblast and some of these daughter cells rapidly mature into short-lived plasma cells that fail to induce B-cell memory, affinity maturation, or class switching – all of which require help from T-cells. Therefore, these cells mainly produce low-affinity immunoglobulin-M (IgM) class antibodies resulting in a very quick antibody response but with limited specificity.

In most cases, B cells do require T-cell derived signals to become activated (T-cell dependent pathway) (Figure 2.C). This pathway usually occurs later in the primary response (days or weeks) or in second antigen-encounters. Proliferating IgM^+ B immunoblasts formed from naïve B cells that have encountered an antigen in the T-cell zone migrate into the centre of the primary follicle, forming a germinal centre (GC) reaction that occurs in a lymph node, mucosa-associated lymphoid tissue or spleen. This GC reaction is an effective mechanism to produce B-cells with highly specific antigen receptors and two types of effector cells. Firstly, the immunoblasts differentiate into centroblasts with a high proliferation rate, aided by the repression of cell cycle inhibitors and the expression of cell cycle activators. Secondly, these centroblasts undergo *somatic hypermutation (SHM)* in order to enhance the specificity of the antibodies produced by the cell. Later on, these centroblasts mature into non-proliferating centrocytes and they undergo *heavy-chain class switch recombination (CSR)*, where the IgM constant region changes into IgG, IgA or IgE, resulting in enhanced effector functions by these antibodies. Immunoglobulins expressed on the B-cell surface (B-cell receptors) of different B-cells compete for the interaction with the antigen displayed by follicular dendritic cells (FDC). Only those centrocytes whose immunoglobulin mutations resulted in highly specific B-cell receptors, will survive; whereas those presenting decreased affinity will die by apoptosis. Finally, the GC reaction ends when selected centrocytes differentiate into either plasma cells or memory B cells. B-cells transiting the GC are germinal centre B-cells (GCB) whereas B-cells that have completed the GC reaction are referred as non-GCB cells, activated B-cells (ABC) or post-germinal B-cells.

DLBCL DIAGNOSIS TODAY

With the constant introduction of innovative technologies, heterogeneity in clinical outcomes is commonly attributed to DLBCL tumour biology, and nowadays, as with other organ malignancies, tumour molecular profiling is becoming an increasingly powerful tool to reveal such heterogeneity with regards to prognostic and therapeutic stratification of DLBCL patients [7].

Over fifteen years ago, cDNA microarrays from untreated DLBCL patient samples aimed at creating molecular signatures characteristic of individual DLBCL phenotypes, which resulted in the identification of two major subtypes: germinal center B-cell like (GCB) and activated B-cell like (ABC) DLBCLs. This molecular distinction has striking prognostic implications (with the ABC subtype exhibiting a poorer outcome); and more importantly, they represent lymphomas driven by different oncogenic processes that could be differently exploited for therapeutic opportunities [8, 9]. This phenotype classification defines both DLBCL subtypes based on their derivation from different stages of B-lymphocyte development, and thereby commonly referred as 'cell of origin' (COO) classification. GCB DLBCLs are likely to derive from centroblasts residing in the GC, that bear mutated immunoglobulin genes with ongoing SHM.

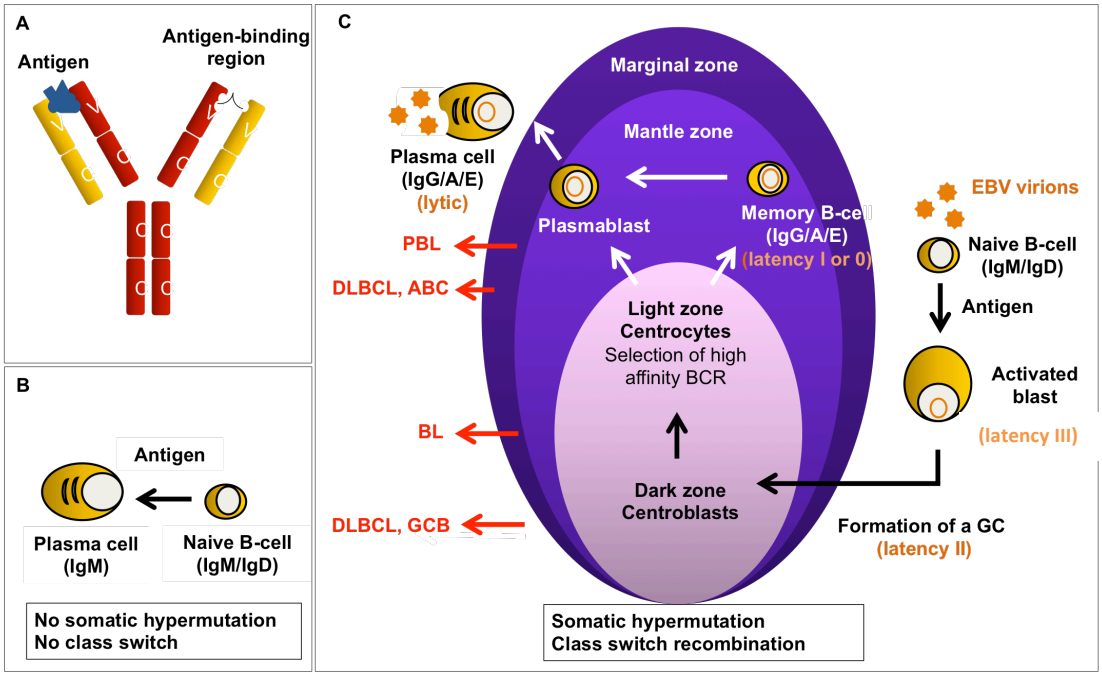


Figure 2. (A) Schematic representation of the antibody structure. During class switch recombination (CSR), the C region of the heavy polypeptide chains is altered and the antibody isotype is changed from one to another (e.g. IgM to IgG). The variable (V) region of both the heavy and the light chain are targeted by somatic hypermutation (SHM). (B-C) Activation of a naïve B cell (that expresses IgM and IgD on its surface) by its cognate antigen results in proliferation and differentiation into a plasma cell or memory B cell. Depending on the nature of the antigen, this activation process can occur through either a T-cell independent (B) or in a T-cell dependent pathway (C). Upon antigen encounter in the first process, IgM-producing plasma cells are rapidly formed, and since no SHM or CSR occurs, these antibodies tend to have lower affinity and specificity towards the antigen. On the contrary, upon antigen encounter through a T-cell dependent manner, the activated B-cell enters a primary follicle in a lymphoid tissue and transforms it into a secondary follicle through a germinal centre (GC) reaction. This GC consists of a light and a dark zone. In the dark zone, the activated B-cell proliferates and downregulates the expression of IgM and IgD allowing SHM and CSR to happen. These B-cells now express high affinity IgG, IgA or IgE antibodies. In the light zone of the GC, those B-cells (centrocytes) with best affinity antibodies are selected and mature into plasma cells or memory cells. (C) Different types of lymphoma arise from different stages on the B-cell development: Germinal centre B-cell like (GCB), diffuse large B-cell lymphoma (DLBCL); Burkitt lymphoma (BL); activated B-cell like (ABC) DLBCL; plasmablastic lymphoma (PBL).

(*See 'The role of EBV in etiology of DLBCL*) > (C) Following the classic B-cell development pathway, once EBV infects a naïve B-cell; the formation of a GC reaction is initiated. In these newly infected (activated) B-cells EBV expresses all nine viral proteins inducing its proliferation. Upon entering the GC, the latency-II profile is displayed and activated centroblasts most likely undergo SHM, conferring selection of highly specific B-cells. Next, these cells either differentiate into plasma cells (activating their lytic cycle) or preferably into memory cells, in which few or no viral proteins are expressed (latency I, or 0).

On the contrary, ABC DLBCLs are believed to arise from B cells at a plastmablastic stage, just prior to GC exit (in the light zone), expressing genes that are often characteristic of mature plasma cells (Figure 3.C) [9, 10].

According to this GEP-based COO classification method, approximately 50% of DLBCLs are classified as GCB subtype, around 30% as ABC DLBCLs and the remaining 20% is often considered unclassifiable, though is commonly grouped together with the ABC subtype (and referred to as non-GCB) since it has similar poor outcome to the ABC DLBCL subtype [7, 11, 12]. More recently, other gene expression profiling (GEP) studies [12-14] assessed by multiple approaches have proposed distinct gene signatures and alternative classifications for DLBCL, e.g. based on the stromal signature [14]. However, no comparative study about which of these is most prognostically significant has yet been performed, and to date, ABC/GCB COO classification is the most validated molecular taxonomy method [10].

DETERMINING CELL OF ORIGIN

Even though this COO classification was described a decade ago, it was not until recently that it became clinically relevant. This mainly due to two factors: (a) The development of new real-time COO assessment methods, and (b) the identification of novel and promising therapeutic agents targeting specific DLBCL subtypes.

Although array-based gene expression profiling is still considered the gold standard method of evaluating COO DLBCL subtypes, the fact that it is currently expensive, time consuming and requires on-going skilled labour, makes it poorly transposable to routine diagnostics in most laboratories. Hence, various immunohistochemistry (IHC) algorithms have taken over as relatively low-cost surrogates for molecular DLBCL subclassification. The Hans algorithm [6] is considered one of the most popular and widely used IHC algorithms and is thought to correlates with GEP-based techniques in around 70-80% of cases (Figure 3). It is worth noting, however, that IHC algorithms do not distinguish 'unclassifiable' cases and thus pools them together with ABC DLBCL cases (referred to as non-GCB).

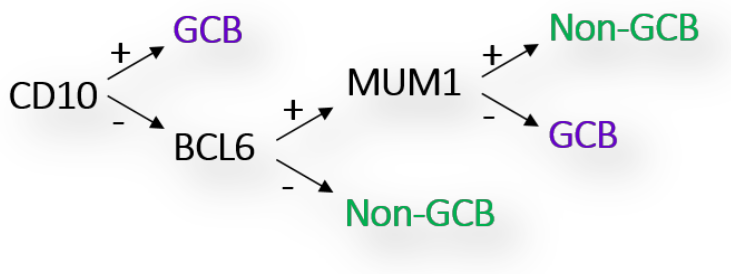


Figure 3. The Hans algorithm applies CD10, BCL6 and MUM1 expression to classify DLBCL into either GCB ($CD10^+$ and/or $BCL6^+$, $MUM1^-$) or non-GCB subtypes with the opposite pattern of staining.

Several other groups have proposed alternative IHC algorithms with minor modifications [15-17], however comparisons between these algorithms has resulted in conflicting results when assigning COO subtypes [18, 19]. Furthermore, IHC-GEP comparisons regarding COO classification and

predictive outcome led to well-correlated results in some reports [6, 17, 19], but were found devoid of prognostic value in others [20, 21]. Such discordant data might be, at least partially, justified by heterogeneity within studied samples (nodal/extra-nodal) and populations (age/gender); technical issues (antibody batch-to-batch differences, antigen retrieval, signal amplification, etc.) and lack of standardised and widely accepted scoring criteria [10, 22].

As generally accepted, no single protein expression method is able to replicate GEP. Accordingly, other GEP techniques based on the evaluation of limited sets of genes have been proposed to address this issue, such as real-time PCR [23, 24], whole genome DASL assay [25], quantitative nuclease protection [26], and NanoString technology [27, 28] among others. Nevertheless, these methods often rely on to dedicated platforms and/or require fresh frozen tissue, which proves prohibitive for most hospitals as diagnostic biopsies are predominantly stored as formalin-fixed paraffin-embedded tissue [7, 22].

Taking all the aforementioned limitations into account, Mareschal et al. proposed a simple and cost-effective GEP-based technique that could have wide applicability and practical utility in routine clinical practice. This method is based on a Multiplex Ligation-dependent Probe Amplification (MLPA) assay (schematically described in Supplemental Materials and Methods Figure S.1). Built on a predictive eight- gene panel: 4 GCB-related genes (NEK6, BCL6, MYBL1, LMO2) and 4 ABC-related genes (MUM1, TNFRSF13B, FOXP1 and IGHM), this technique allows for a rapid and accurate classification of GCB and ABC DLBCL phenotypes. Additionally, this gene panel includes five diagnostic and prognostically relevant genes (TNFRSF9, MYC, BCL2, CCND1, CCDN2) and the internal control MS4A1, encoding for CD20). This gene selection was based according to their differential expression in ABC and GCB DLBCL cases in two independent series of Affymetrix U133 Plus 2 arrays, further details explained elsewhere [22].

THE ROLE OF EBV IN ETIOLOGY OF DLBCL

Although the cause of most DLBCLs remains to be elucidated, a distinction should be made between cases arising *de novo* (primary DLBCLs) and those referred to as secondary DLBCLs, representing the progression (transformation) of a less aggressive lymphoma.

Predisposing factors to develop a DLBCL include agents that can cause molecular abnormalities, as well as innate and acquired immunodeficiency states. Moreover, besides evident genetic factors, several studies suggest that some environmental agents might play a crucial role in the pathogenesis of DLBCL [29]. In particular, viruses have become a matter of intensive debate during the last decade. Of interest, DLBCL arising in the setting of a compromised immune system, most often show Epstein Barr-virus (EBV) positivity.

The Epstein-Barr virus (EBV) is a widely spread human herpes virus that benignly infects over 95% of the population, making it one of the most common and successful of all human viruses. It is known, that once EBV infects a person it will remain in the human body for life [30]. Although most people who

carry EBV for a lifetime do not suffer from the viral infection, primary EBV infection can result in self-resolving infectious mononucleosis (most commonly known as kissing disease). However, occasionally, under certain still poorly understood conditions, EBV infection or reactivation may lead to a wide range of life-threating lymphoproliferative disorders (LPDs). These mainly occur in immunocompromised individuals, either in a congenital or acquired context (e.g. AIDS, autoimmune diseases, or under immunosuppressive therapies post-transplantation). Furthermore, EBV is a well-known oncovirus, and as such is associated with several human neoplasms of epithelial origin, such as nasopharyngeal carcinoma, gastric carcinomas and a broad spectrum of lymphomas, of which DLBCL is most frequent [31-33]. Such diversity reflects the intricate interaction between EBV and its human host.

Similarly to other herpes viruses, EBV can alternatively engage two distinct life cycle phases: lytic and latent. In the course of lytic replication (initiated by viral BZLF1/ZEBRA and BRLF1), EBV expresses its full repertory of proteins and new virons are produced. Alternatively, during latency, the virus displays a limited number of its genes which are essentially aimed at maintaining the viral genome (as an episome in the nucleus), at evading the host's immune system [34] and to exploit the normal B-cell activation pathways that drive into memory B-cell differentiation. Three distinct latent phases are displayed based on the expression of nine viral proteins: six EBV nuclear antigens (EBNA1, 2, 3A-C and EBNA leader protein) and three latent membrane proteins (LMP1, 2A-B) (latency III: LMP1+/EBNA2+, latency II: LMP1+/EBNA2-, latency I: LMP1-/EBNA2-) [35, 36]. These latency types are present at specific stages of EBV-infection, each of them being characteristic of different types of tumours [35, 37].

EBV infection *in vivo* is characterised by a lifelong latency profile in memory B-cells rather than active EBV-driven proliferation. This explains why not every infected individual ultimately gets cancer (even when the immune system is compromised). Accordingly, it seems that further aberrations in B-cell or EBV virulent machinery need to occur in order to avoid such latent infection and allow cancer to occur [35, 38].

DLBCL IN IMMUNOCOMPROMISED PATIENTS

The molecular pathogenesis of immunodeficiency-associated DLBCL differs substantially from that of DLBCL in immunocompetent (IC) hosts. In fact, as previously mentioned, EBV infection is present in the tumour cells in a large fraction of immunodeficiency-associated DLBCLs (e.g. in AIDS patients, organ transplant recipients, autoimmune patients treated with immunomodulatory agents), whereas DLBCL of IC hosts are usually EBV-negative [3], except for some cases arising in elderly (>50 years old) individuals (acknowledged by 2008 WHO classification in the provisional entity "EBV^+ DLBCL of the elderly" [2]) in which senescence of the immune system might play a crucial role [39, 40]. Nevertheless, some reports [41-44] have also described EBV^+ DLBCLs in young patients without detected immunodeficiency.

Previous studies from our group [45] and [46] aimed at characterizing the clinic- pathological and molecular characteristics of EBV-driven (EBV^+) and EBV-negative (EBV^-) PT-DLBCL patients and to elucidate whether they are biologically different. Based on microarray data from 48 DLBCL cases, of

which 33 occurred following transplantation (72% EBV$^+$) and in immunocompetent (IC) hosts (EBV$^-$), it was concluded that EBV$^+$ and EBV$^-$ DLBCLs are characterized by a different gene expression profile, illustrating the impact of the EBV infection hallmark on cell signalling and in the pathogenesis of EBV-driven post-transplant lymphoma. Furthermore, EBV$^+$ PT-DLBCL were genetically less complex that EBV$^-$ PT-DLBCL, which harboured similar genetic lesions as EBV$^-$ DLBCL in IC patients [44].

A common feature of all EBV$^+$ DLBCLs is that, according to the BCR SHM status, GEP-based and/or IHC classification methods, they are mostly of non-GCB origin. Nonetheless, a few rare cases lack SHM and thus have been reported. These may derive from naïve pre-GC B-cells or from B-cells that have transited the GC without completing the GC program (i.e without undergoing SHM) [47, 48]. Notably, these tumours correspond to one of two recently discovered subgroups of non-GCB EBV$^+$ PT-DLBCL. Based on IgM staining and SHM analyses from the above-mentioned series of EBV$^+$ non-GCB PT-DLBCL, two subgroups were identified: mutated IgM$^-$ and unmutated IgM$^+$ EBV$^+$ non-GCB PT-DLBCLs. Each group was associated with particular clinical and biological characteristics and prognosis, the latter being worse in the unmutated IgM$^+$ PT-DLBCL group (manuscript submitted). IgM$^-$ tumours arose late following transplantation, mainly in kidney recipients. IgM$^+$ tumours on the other hand arose early, almost exclusively in HSCT. Notably, IgM$^+$ tumours were characterised by plasma cell features (monotypic kappa/lamda expression, high MUM1 expression, partial CD138 expression). Consistent with the plasma cell phenotype, unfolded protein response (UPR) signalling was upregulated.

Moreover, these two non-GCB EBV$^+$ PT-DLBCL subgroups were further correlated with distinct patterns of EBV protein expression: all IgM$^+$ tumours expressed latency III compared to 50% of IgM$^-$ tumours and EBV lytic replication (assessed by means of ZEBRA staining), was most abundant in IgM$^+$ tumours. These findings point to an interaction between the level of immunosuppression and EBV protein expression. Whether these two subgroups also arise in EBV$^-$ PT-DLBCL and/or in lymphomas raised in other immunocompromised situations remains to be investigated.

Although the majority of PT-DLBCLs are associated with EBV infection (60-80% cases), a significant portion is negative for this virus. Our data suggest that, these EBV$^-$ cases do not represent a 'real' PT-DLBCL but are rather coincidental cases of lymphomas that are indistinguishable from DLBCL in an IC-host [35, 49].

CURRENT THERAPEUTIC STRATEGIES

To date, chemotherapy, radiation and the recently incorporated Rituximab (anti-CD20) antibody are the mainstays of lymphoma treatment [50], being the R-CHOP regimen the standard first-line treatment for DLBCL. Whereas approximately 80% of DLBCL patients respond to this treatment, fewer than 40% are likely to be fully cured. This difference in clinical outcome can, as research advances, be explained by the heterogeneity of DLBCL. Enhanced understanding of the tumourgenesis in DLBCL, describing and categorizing it in distinct molecular subgroups, is allowing investigators to move towards stratification of treatment for individual COO phenotypes [7].

Some on-going trials are investigating the effects of several novel agents, as both, in combination with standard R-CHOP therapy or as single agents in a relapse/refractory setting. These agents, whose mechanisms are still not completely understood, present very different activities targeting several pathways that are predominantly involved in non-GCB DLBCLs [51]. Given the persuasive role of the NF-κB pathway in DLBCL, several NF-κB inhibitors are being tested. A promising agent targeting this pathway is the proteasome inhibitor (*Bortezomib*), which is currently being validated in phase 3 clinical trials and that so far has been preferentially active in the ABC DLBCL when combined with chemotherapy. In a similar approach, upstream molecules involved in BCR signalling are being targeted, such as SYK and BTK inhibitors (*Ibrutinib*). Furthermore, other inhibitors target the PI3K/AKT/mTOR pathway (*Rapamycin*) and some explore epigenetic strategies, as histone deacetylase inhibitors. Finally, immunomodulatory agents with pro-apoptotic and anti-angiogenic properties are also being developed in lymphomas [7, 9, 50].

Due to the relative rarity of EBV^+ DLBCL most studies have rather focused on EBV^- cases, and as a result, the management of patients with EBV-driven DLBCLs remain complex and controversial. Chemotherapy and monoclonal antibodies are the therapeutic options most frequently offered to these patients. The administration of Rituximab is largely well tolerated and it induces a rapid depletion of mature B-cells in the peripheral blood, thus reducing the compartment of EBV-infected cells. Currently, some on-going clinical trails are focused on using EBV-specific cytotoxic T-cell lymphocytes (CTLs). Results from these trails are eagerly expected, as they could be of vital importance in the treatment of EBV-positive DLBCL in IC patients [52]. Moreover, the aforementioned *bortezomib* and *rapamycin* may also be valuable on treating EBV-positive DLBCLs as well [53, 54]. On regard of PT-DLBCLs, reduction of immunosuppression (RIS), which allows recovery of EBV-specific CTLs, is considered the first line treatment whenever is possible. Since there is limited lytic EBV replication in these tumours, the use of antivirals for treatment remains debatable [55]. Finally, the development of strategies that reactivate lytic replication in latently infected tumour cells (lytic induction therapy) are of increasing interest as lytic replication promotes death of tumour cells. Nevertheless, today, no approved antiviral drug is available for the management of EBV-associated diseases, partly due to the lack of an efficient replication system.

The molecular diversity in which these lymphomas present is initially daunting, but increasing identification on their shared genetic aberrations and common signaling mechanisms is shedding some light into how to approach them. Ideally, one hopes to identify and exploit these tumours with therapies directed towards oncogenic cells while spearing normal cells.

HYPOTHESIS AND OBJECTIVES

Taken all together, advances in molecular biology hold the potential to address DLBCL heterogeneity and provide precise and clinical relevant sub-classifications, and consequently, ever-increasing demands are placed on pathologists to translate this knowledge into clinical practices. Nonetheless, accurate DLBCL subtyping does have important prognostic and therapeutic implications.

In the current work our first aim was to set up a pilot study to optimize a GEP method by using an MLPA-based assay for predictive markers (MUM1, BCL6, FOXP1, TNFRSF13B, IGHM, MYBL1, NEK6, LMO2) at UZ Leuven. This method, adapted from Mareschal. et al. in [20], would allow us to accurately subdivide DLBCL into GCB/ABC prognostically relevant DLBCL subgroups according to gene expression profiles. Being less affected by RNA degradation, this MLPA assay can also be applied to FFPE biopsies – the main source of human samples in most hospitals.

A fraction of DLBCL harbors the Epstein-Barr virus (EBV). These cases arise mainly in immunocompromised individuals, e.g. following organ transplantation (post-transplant, PT-DLBCL). EBV+ PT-DLBCLs are most frequently of ABC origin and recent studies have suggested that they constitute a distinct DLBCL entity in itself. Additionally, it appears that subgroups within EBV+ PT-DLBCL exist and are, in fact, related to distinct pathogeneses (IgM^+ and IgM^- EBV^+ PT-DLBCL). Given that several predictive markers have shown to be differentially expressed in EBV- and EBV+ PT-DLBCL (subgroups) (e.g. FOXP1, IGHM) we hypothesise that the presence of EBV influences the MLPA-based DLBCL classification.

Currently, DLBCL patients are being treated with a cocktail of chemotherapeutics frequently supplemented with anti-CD20 antibody therapy. This is carried out independently of the COO or the EBV status of the tumour. However, increasing insight into the molecular biology of DLBCL allows for the application of targeted treatments (e.g. ibrutinib to block BCR signaling). Due to fact that the majority of studies focus on EBV^- DLBCL, it is not yet clear whether these compounds will be useful against EBV^+ cases. Thus, our final aim is to evaluate the performance and utility of several drug compounds (ibrutinib, bortezomib, rapamycin) at targeting specifically EBV-driven DLBCL using EBV^+ ABC lymphoblastoid cell lines (LCLs) from several patients as a model.

MATERIALS & METHODS

CASE SELECTION

A total of thirty-six patients with *de novo* DLBCL were included in this study. Of these DLBCL cases, twelve arose in immunocompetent (IC) individuals, and twenty-four of them in transplant recipients in a post-transplant (PT) lymphoproliferative malignancy context. In addition, three patients with benign reactive lymphadenopathy were included as controls. All cases were diagnosed at the University Hospitals of KU Leuven (Leuven, Belgium) following the 2008 WHO criteria by expert pathologists. Formalin-fixed paraffin-embedded (FFPE) and/or fresh-frozen (FF) material was retrieved from the archives of the Pathology department. The Ethical Committee of the University Hospitals of Leuven approved sample collection for this study (s55498), which was performed according to the Declaration of Helsinki.

STUDY DESIGN

A first series comprising twelve IC-DLBCL patients (non-GCB/ABC DLBCL, sample 1-6; GCB-DLBCL, sample 7-12) and four non-neoplastic lymphadenopathy patients (reference samples 1-4) was used to optimize the use of the DLBCL subtype classifier proposed in [22] at UZ Leuven. For all DLBCL patients, the inclusion criteria were a *de novo* EBV-negative DLBCL and paired-matched FFPE and FF lymph node biopsies available at the time of diagnosis. All cases had been previously analysed and classified by IHC following the Hans algorithm [6].

A second series of twenty-four PT-DLBCL patients (7 EBV^- PT-DLBCLs and 17 EBV^+ PT-DLBCLs (9 IgM^- and 8 IgM^+)) was used to further investigate the impact of the EBV when using the same GCB/ABC classifier. An IHC classification was available for all cases and microarray data for 19 of them. These data are available online at GEO GSE38885.

HISTOLOGICAL EVALUATION

IMMUNOHISTOCHEMISTRY

All immunohistochemical stainings were perfomed on 4µm-thick sections on an automated stainer (Bond-Max, Leica Microsystems, Bannockburn, IL, USA). Briefly, Dewax Solution was used to deparaffinise FFPE sections. Following deparaffinisation, Bond Wash Buffer was used, as a standard IHC was buffer. The Bond Peroxidase Block inactivated endogenous peroxidases in all tissue samples, and Heat-Induced Epitope Retrieval (HIER) buffers, comparable to either Citrate Buffer (~pH6) or EDTA Buffer (~pH9), were used for each antibody prior to incubation with primary antibodies (Supplemental Materials and Methods, Table S.1). Antibody detection was performed with a commercial Bond

Polymer Refine Detection System (Leica Biosystems, Newcastle, UK) based on a polymeric horseradish peroxidase (HRP)-linker antibody conjugate system.

Based on CD20 staining and/or haematoxylin-eosin staining the tumoural fraction in each biopsy was estimated. For every immunohistochemical staining, relative protein expression was calculated by multiplying the fraction of positive cells by the tumour fraction.

EBV-ENCODED RNA *IN SITU* HYBRIDISATION

EBV-encoded RNA (EBER) in situ hybridisation is considered the standard for diagnosis of EBV infection and was performed using a 30-mer digoxigenin-labelled oligonucleotide probe (Research Genetics, Huntsville, AL), according to manufacturer's instructions. A control poly-A probe (Ventana Roche, Arizona USA) was applied in order to check for RNA integrity and a proven EBV-driven lymphoma was used as a positive control. Cases were defined as EBV^+ if EBER was expressed in the majority of viable tumour cells.

RNA ISOLATION AND cDNA SYNTHESIS

From all samples included in the study, total RNA from frozen tissue sections was extracted by using the RNeasy Plus Mini Kit according to the instructions of the manufacturer (Qiagen, Hilden, Germany). For the fifteen cases included in the first case series (sample 1—12; reference samples 1,3 and 4), total RNA was isolated from 10µm-thick FFPE tissue sections by using the High Pure FFPET RNA Isolation Kit (Roche, Applied Science, Alameda, USA) also according to manufacturer's protocol. RNA concentrations were measured with a spectrophotometer (Nano-Drop 1000c; Thermo Scientific, Waltham, USA). The purity of RNA samples was determined by OD 260/280 ratio. Samples were stored at -80ºC.

For the reverse transcription (RT), 1 µg of total RNA sample was reverse transcribed into cDNA in a total volume of 20 µl by using a conventional thermal cycle (Applied Biosystems) and the iScript cDNA Synthesis Kit (Bio-Rad, Hercules, CA, USA) following manufacturer's standard protocol. The resulting cDNA samples for all PT-DLBCL cases were further diluted ¼.

RNA INTEGRITY

In order to assess the quality (fragmentation) of the RNA starting material, an electrophoretic RNA integrity assay was performed on total RNA isolated from 6 FF samples (samples 1, 5-7, 12 and reference sample 1) and 6 FFPE samples (1, 5-7, 12 and reference sample 1) using the Agilent RNA 6000 Nano kit and the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Assessment of RNA integrity is based on the RNA Integrity Number (RIN), which removes user-dependent interpretation in RNA quality and takes the entire electrophoretic trace into account. Next, a

RIN software algorithm allows classifying total RNA based on a numbering system from 1 to 10, being 1 the most degraded profile and 10 the most intact [56].

MULTIPLEX LIGATION-DEPENDENT AMPLIFICATION ASSAY

The Multiplex Ligation-dependent amplification assay (MLPA) protocol used in this study is based on the protocol reported by Mareschal et al. in [22] with minor adaptations.

MLPA PROBES DESIGN AND MIX

The reported MLPA probes design and probe mix were used. All MLPA probes consisted of one short synthetic oligonucleotide designed to detect cDNA (Integrated DNA Technologies, Coralville, IA, USA). To avoid contamination of genomic DNA amplifications, all target sequences were designed across exon-exon boundaries. A schematic representation of an MLPA oligonucleotide, together with the gene names and sequences that were used in this study are shown in Figure S.2 and *Table S.2 of Supplemental Materials and Methods*, respectively.

Competitor oligonucleotides (CO) targeting IGHM and NEK6 (*Table S.2 in Supplemental Materials and Methods)* were available to reduce off-scale signal peaks to the level within the dynamic range of the capillary sequencer. The CO for a particular target gene is identical to the 3′ MLPA probe without the PCR-primer binding site, and thus cannot be amplified by PCR. This CO then competes with its corresponding 3′-MLPA probe for the limited number of binding sites on the cDNA. The use of a 1:1 ratio of 3′ MLPA probe and its corresponding competitor usually decreases the probe signal two-fold [57, 58].

A schematic workflow of how the MLPA probe mix was prepared is provided in *Supplemental Materials and Methods Annex 1*. Concisely, 10 µmol/L dilutions of each probe and the competitor were prepared in Tris 10 mmol/L and EDTA 1 mmol/L (TE). The IGHM competitor was mixed with its corresponding IGHM 3′ MLPA probe in a 15:1 ratio. Next, 2 µl of each diluted probe (n=26) and probe + competitor mix (n=1) were pooled together to obtain a volume of 56 µl, to which an equal volume of TE was then added to obtain a final volume of 112 µl (pre-final MLPA mix). Lastly, the final probe mix was obtained by diluting 11,2 µl of the pre-final mix in a final volume of 1 mL of TE buffer.

MLPA OPTIMISED PROTOCOL

A schematic workflow for the optimized MLPA procedure is provided in *Supplemental Materials and Methods Annex 1*. In summary, for each sample, 5 µl of cDNA (250 ng) were transferred into a new tube and denatured at 98ºC for 2 minutes in a thermo cycler. To allow the annealing of the MLPA probes to their targets, 3 µl of the Hybridisation Mix (1,5 µl Final-Probe mix + 1,5 µl SALSA-MLPA buffer) were added to each tube and next incubated at 60ºC for one hour. After cooling the samples down to 56ºC, and without taking the samples out of the thermocycler, 32 µl of the Ligation mix (25 µl RNAse-free water, 3 µl of SALSA-Ligase 65 Buffer A, 3 µl of SALSA-Ligase 65 Buffer B, and 1 µl of SALSA-

Ligase 65) were added to each tube and incubated at 56ºC for 15 minutes to ligate the annealed oligonucleotides. The ligase enzyme was next inactivated with heating, 5 minutes at 98ºC. Next, 5 µl of each ligated product were transferred into a new tube and mixed with 15 µl of PCR mixture (3 µl RNAse-free water, 10 µl Thermo Scientific Extensor Hi-Fidelity PCR Master Mix (ThermoScientific, Marieta, OH, USA), 1 µl of 10 µmmol/L primer U1 and 1 µl of 10 µmmol/L FAM-labelled primer U2 (Integrated DNA Technologies, Coralville, IA, USA)). The PCR amplified products were obtained as followed: 6 minutes at 98ºC; 35 cycles (30" at 94ºC, 30" at 58ºC and 30" at 72ºC); 30 minutes at 72ºC and cooled down to 16ºC.

Further dilutions of the PCR products (1/20 for FF and 1/10 for FFPE samples) were performed before fragment separation and detection. One µl of each diluted PCR product was mixed with 29 µl of detection mix (28 µl of Hi-Di formamide and 1 µl GeneScan-500 HD ROX size standard (Applied Biosystems, Waltham, MA, USA), denatured for 3 minutes at 95ºC. All samples were next loaded onto an ABI 3170XL capillary electrophoresis analyser, a newer device model than the one reported in [22], hence running with slightly different migration settings: Gel type, POP7; temperature, 63ºC; pre-run time, 180 seconds; pre-run voltage, 15 kV; injection time, 8 seconds; injection Voltage, 1,6 kV; run time, 1600 seconds; run voltage, 15 kV.

DATA ANALYSIS

FRAGMENT ANALYSER FILES (.FSA) PROCESSING

The entire analysis of the raw fragment analyser files (*.fsa*) was achieved with the interfaced MLPA package (version MLPA 1.9.1) developed by Mareschal et al., and the exact details on how it is built can be found in here [22]. The software can be freely downloaded and provides raw and normalized gene expression values and graphical representations of the peak profiles for each sample and of the probabilities of belonging to either ABC or GCB DLBCL subgroups based on those profiles (http://bioinformatics.ovsa.fr/70/MLPA_191 last accessed May 2016). For data input, the software makes use of the "Process interface's design file" template (available online), which was adapted according to our protocol (see Supplementary Materials and Methods, Annex 2). To identify the individual MLPA products, the probe signals were aligned with the 50-, 75-, 100-, 139-, 150 and 160-bp size standards, and the time indexes of local maxima were used to fit a linear model that converted time indexes into base pairs. Next, maximal peak heights of the different peaks were measured by looking for the FAM signal in the peak-size intervals that were pre-defined in base pairs in the design file. Normalized expression values were obtained by dividing each gene's expression value by the mean of all genes' expressions measured in the sample, and by calculating the base 2 logarithms of these ratios plus 1.

DLBCL SUBGROUP CLASSIFIER

Discrimination of GCB and ABC subgroups was performed by using a model class predictor that was also developed by Mareschal et. al. [22], and that followed the linear predictor score (LPS) described by Wright et al. [59] and a Bayesian predictor. In order to build this ABC/GCB classifier, a training series of fifty randomly selected DLBCL patients that had been previously classified as either ABC or GCB (based on microarray data) was used. The classification of samples into one of these two subgroups was based on an eight-gene panel (NEK6, IRF4, IGHM, LMO2, FOXP1, TNFRSF9, BCL6, MYBL1) for each sample. As the ABC/GCB discriminatory power of each gene was different (based on a t test for the difference in expression between the ABC and GCB DLBCL [12, 22, 59], a LPS was then calculated for each sample X, as follows:

$LPS(X) = \sum a_j \cdot x_j$ where x_j represents the gene expression of gene j, and a_j a scaling factor that is applied to balance out the overweighting of the most highly expressed genes.

In the algorithm of Mareschal et al. [22], the relative discriminating power of the different probes is: LMO2 (-11,118) > NEK 6 (-6,074) > MYBL1 (-4,873) > BCL6 (-3,582) and IRF4 (4,837) > TNFRSF13B (4,608) > IGHM (4,476) > FOXP1 (2,740).

To take into account differences in expression ranges in our samples, the scaling factors were further divided by the mean expression of the corresponding genes in the training series. Based on this linear combination of gene expression values, a normal distribution of the scores within the GCB and ABC subgroups was assumed, and the mean and the variance of these distributions could be estimated from all LPS calculated in each subtype. This allowed estimating the likelihood of a new sample to belong in one group or the other. Several confidence intervals to call a sample either GCB or ABC: 95% (applied in [22]), 90% (recommended by Wright et al. [59]), and 80%, -as we considered the first one too strict bearing in mind the thresholds applied in current diagnostic methods (see *Introduction, "Determining Cell of Origin"*).

CELL CULTURE AND TREATMENT

CELL LINES

Lymphoblastoid cell lines derived from five different donors (LCL1-5) were a kind gift from prof. J.A. Martinez-Climent (CIMA, Universidad de Navarra, Pamplona, Spain). The cells were cultured in IMDM medium supplemented with 20% FBS, 2% penicillin/streptomycin and 1% L-glutamine (all from Thermo Fisher Scientific, Waltham, Massachusetts, USA).

Cytospins of each LCL were made following overnight fixation of 500µl cell suspension in an equal volume of cytorich red preservative (BD, Franklin Lakes, New Jersey, USA). Immunohistochemistry (for MUM1, CD10, BCL6) was performed in automated fashion.

MTT ASSAY

LCL cells were plated in duplicate and treated for 24h with *bortezomib* (Velcade; 65nM, dissolved in water), *rapamycin* (Selleckchem; 1.25mM, dissolved in DMSO), *ibrutinib* (Imbruvica; 5mM, dissolved in DMSO) or a combination of two compounds at the same doses. The doses were selected following screening of sensitivity of LCL2 to the individual compounds. The MTT assay was performed following manufacturer instructions using the MTT Cell Growth Assay Kit (Merck Milipore, Billerica, Massachusetts, USA). Absorption was measured at 570nm, of which the background absorption (620nm) was subtracted. For normalization, the resulting value was divided by the absorption in control wells to which water (for *bortezomib*) or DMSO (for all other compounds and combinations) was added.

Because of the limited number of replicates tis study precludes statistical analysis.

STATISTICAL ANALYSIS

The analysis, the organization and presentation of all data were performed with GraphPad Prism 7. Comparisons between two groups were analysed using non-parametric Mann Whitney U tests. Comparisons between matched-pairs were analysed with non-parametric Wilcoxon signed rank tests.

All data are expressed as mean and/or median ± SE, and differences between groups were considered statistically significant at $p < 0.05$.

RESULTS

OPTIMISATION OF MLPA PROTOCOL

Total RNA from matching FFPE and FF tissue from lymph node biopsies of *de novo* DLBCL patients (non-GCB DLBCL: sample 1-6; GCB: sample 7-12) and four non-neoplastic lymphadenopathy patients (reference samples 1-4) was isolated. Results of the RNA yields, purity and quality are summarized in Table 1 Sufficient yields (> 1ug total RNA) for the MLPA assay were obtained from all samples. The 260/280 ratios ranged from 1,91 to 2,17 in all FFPE/FF samples (~ 2.0-2.2 is considered as "pure" RNA); however 260/230 ratios, used as a secondary measure of nucleic acid purity, were markedly lower in FF samples compared to their FFPE counterparts, indicating the presence of contaminants which absorb at 230 nm (e.g. phenols) in the latter. Such reduction is probably due to differences in isolation procedures.

Table 1. Summarized information on isolated RNA from matching FFPE and FF lymph node biopsies.

	Sample	FFPE (10µm)	[Nucleic Acid] (ng/µl)	260/280	260/230	FF	[Nucleic Acid] (ng/µl)	260/280	260/230
non-GBC DLBCL	Sample 1	4 sections (Half tissue)	168,4	2,03	3,49	≈ 20 mg	136,6	2,12	1,94
	Sample 2	3 sections	207,4	2,04	2	≈ 20 mg	79,8	2,15	0,68
	Sample 3	3 sections	234,3	1,99	1,94	≈ 20 mg	132,5	2,15	1,1
	Sample 4	4 sections. (Small tissue)	151,8	1,99	1,87	≈ 20 mg	164,7	2,16	1,19
	Sample 5	4 sections. (Small tissue)	192,4	2,02	2,04	≈ 20 mg	182,2	2,17	1,11
	Sample 6	3 sections	208,1	2	2	≈ 20 mg	216,3	2,13	1,4
GCB DLBCL	Sample 7	3 sections	203,5	2,05	2,04	≈ 20 mg	219,2	2,17	1,02
	Sample 8	3 sections	168,8	2,03	2	≈ 20 mg	150,1	2,15	0,99
	Sample 9	3 sections	134,9	1,99	1,96	≈ 20 mg	192,3	2,17	0,99
	Sample 10	3 sections	116	1,97	1,91	≈ 20 mg	358,8	2,13	1,24
	Sample 11	3 sections	101,7	1,98	1,83	≈ 20 mg	374,5	2,12	1,62
	Sample 12	3 sections	233,6	2,05	2	≈ 20 mg	283,5	2,12	1,44
REFERENCE SAMPLES	RS 1	3 sections	182,6	2,05	3,59	≈ 20 mg	132,6	2,13	1,09
	RS 2	3 sections	62,8	1,91	2,1	≈ 20 mg	88,7	2,11	0,53
	RS 3	3 sections	60,1	2,06	0,66	≈ 20 mg	117,5	2,16	0,74
	RS 4	3 sections	66,5	2,13	0,68	≈ 20 mg	64,7	2,16	0,63

For DLBCL samples (1, 5-7 and 12) and the reference sample 1, a comparison was made between the RIN values (indicating RNA integrity) obtained from FFPE and FF tissue. FFPE storage had a negative impact on RNA quality (RIN N/A) compared to FF (RIN of 10) (Figure 4).

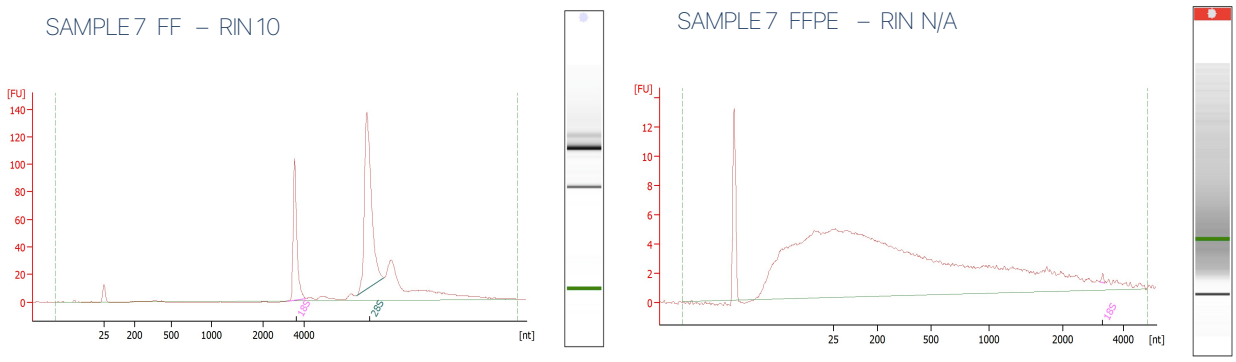


Figure 4. Results of the RNA integrity assay performed on RNA isolated from FF (left panel) and FFPE (right panel). Representative electropherogram profile and gel-like image for sample 7. In the left panel, the two peaks represent intact 28S and 18S ribosomal RNA. The gradual degradation of ribosomal RNA is reflected by a continuous shift towards shorter fragment sizes, also reflected as smearing on the gel image (right panel).

OFF-SCALE SIGNAL INTENSITIES ARE SOLVED BY DILUTING FINAL PCR PRODUCTS

As previously mentioned, our MLPA experiments were performed by adapting the MLPA protocol reported by Mareschal et al. in [22]. The most important modifications we made were: the use of different starting cDNA concentrations, the use of only one competitor oligonucleotide (i.e. IGHM) instead of two; the use of a different size standard, and dilution of the final PCR products before loading onto the capillary electrophoresis analyser which was a newer device model and ran with slightly different migration settings. We also used a different size standard than the one reported. An overview of the optimised MLPA protocol for our experiments is included in *Supplemental Material and Methods, as Annex 1*.

In our experiments, IGHM, CD20, FOXP1 and LMO2 were so highly expressed that, the resulting peak-signals were consistently off-scale (signals above 30000 relative fluorescent units (RFU)) (Figure 5). Accordingly, several optimisation steps were performed in order to bring those off-signal peaks down.

Firstly, the competitor oligonucleotide (CO) and probe ratio for IGHM was increased from 4:1 to 15:1. Conversely, the CO for NEK6 gene was excluded from the MLPA probe mix, as levels of NEK6 were markedly low in all samples (data not shown). Secondly, the initial amount of cDNA in the MLPA reaction was reduced. However, as observed in *Figure 5,* this did not alter the height of the peaks, and in some instances even resulted in increased RFU.

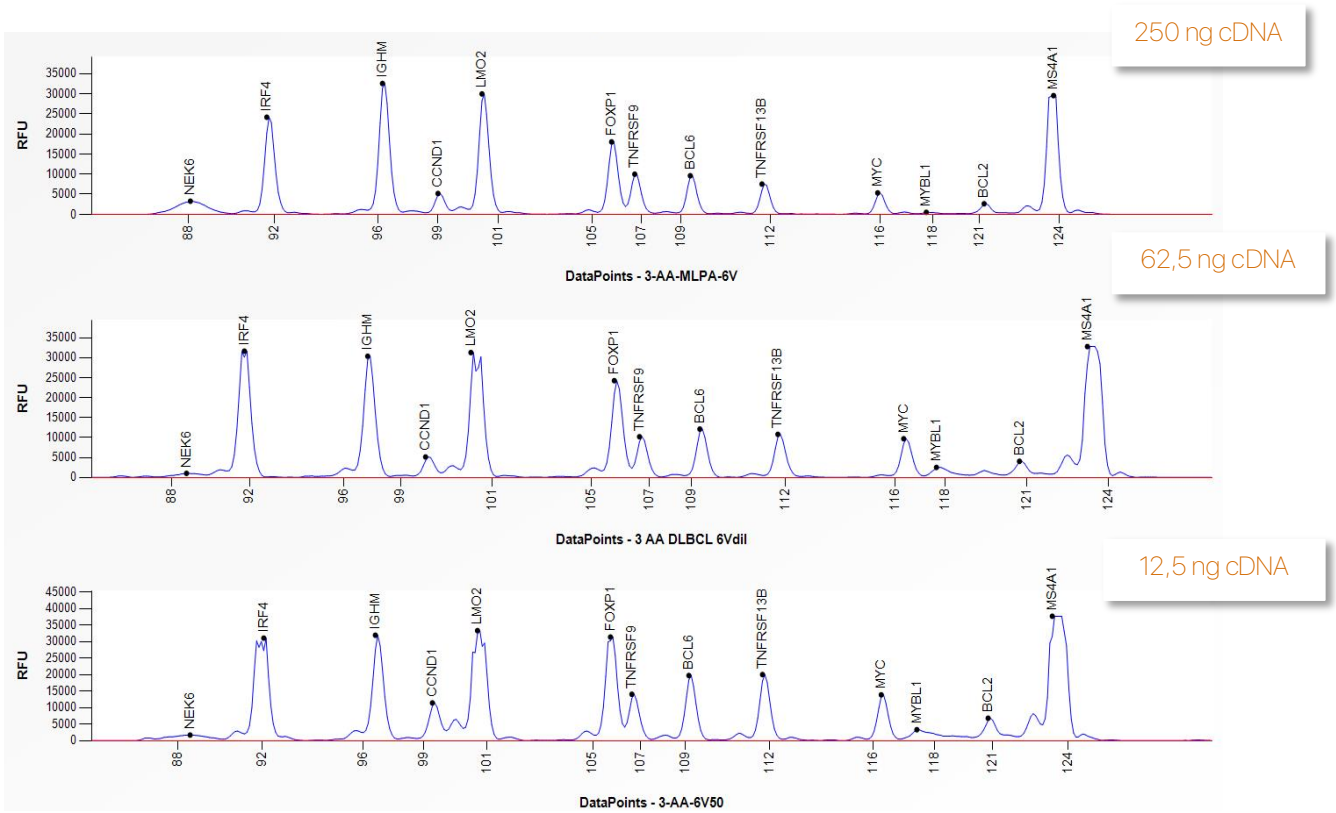


Figure 5. Optimisation step for the MLPA assay. Representative genomic expression profiles of several initial cDNA amounts (250 – 12,5 ng) from FF sample 6. Some genes (IRF4, IGHM, LMO2, FOXP1, MS4A1) show off-scale signal peaks – i.e. signal intensities equal or above 30000 RFU. For all cDNA quantities these genes show either truncated or flattened peaks, indicating that their gene expression is off-limits for the capillary electrophoresis device in use, and thus lowering the amount of initial cDNA amounts fails on reducing peak intensities.

Next, we further diluted the PCR products before loading them onto the fragment analyser. Out of different dilutions (1/2, 1/5, 1/10 and 1/20) (data not shown), diluting the PCR-amplified products to 1/20 and 1/10, for FF and FFPE respectively resulted in detectable gene expression for all genes, whose peak heights were within the range limit of the capillary electrophoresis (Figure 6).

CLASSIFICATION OF THE IC-DLBCL CASE SERIES

HIGHLY DEGRADED RNA SUCCEEDS IN GENERATING MLPA PROBE SIGNALS

As displayed in *Figure 6* for each .fsa sample file, a RT-MLPA profile was provided. Overall, using a CI of 80% ABC/GCB classification could be performed for all FF samples compared to a 67% (8/12) of FFPE samples (Table 2). Since the fragment analysis profile for some of the FFPE samples included in this series failed to generate trustworthy fragment analysis profiles, these were excluded from the study. This was also the case for Reference sample 2 FF. Also when RNA isolations and RT-MLPA reactions were repeated unreliable results were obtained (data not shown). The reason why some FFPE blocks performed better than others remains to be investigated. These results demonstrate that although

MLPA profiles can be obtained for RNA extracted from FFPE tissue, FF is a better source of RNA than FFPE.

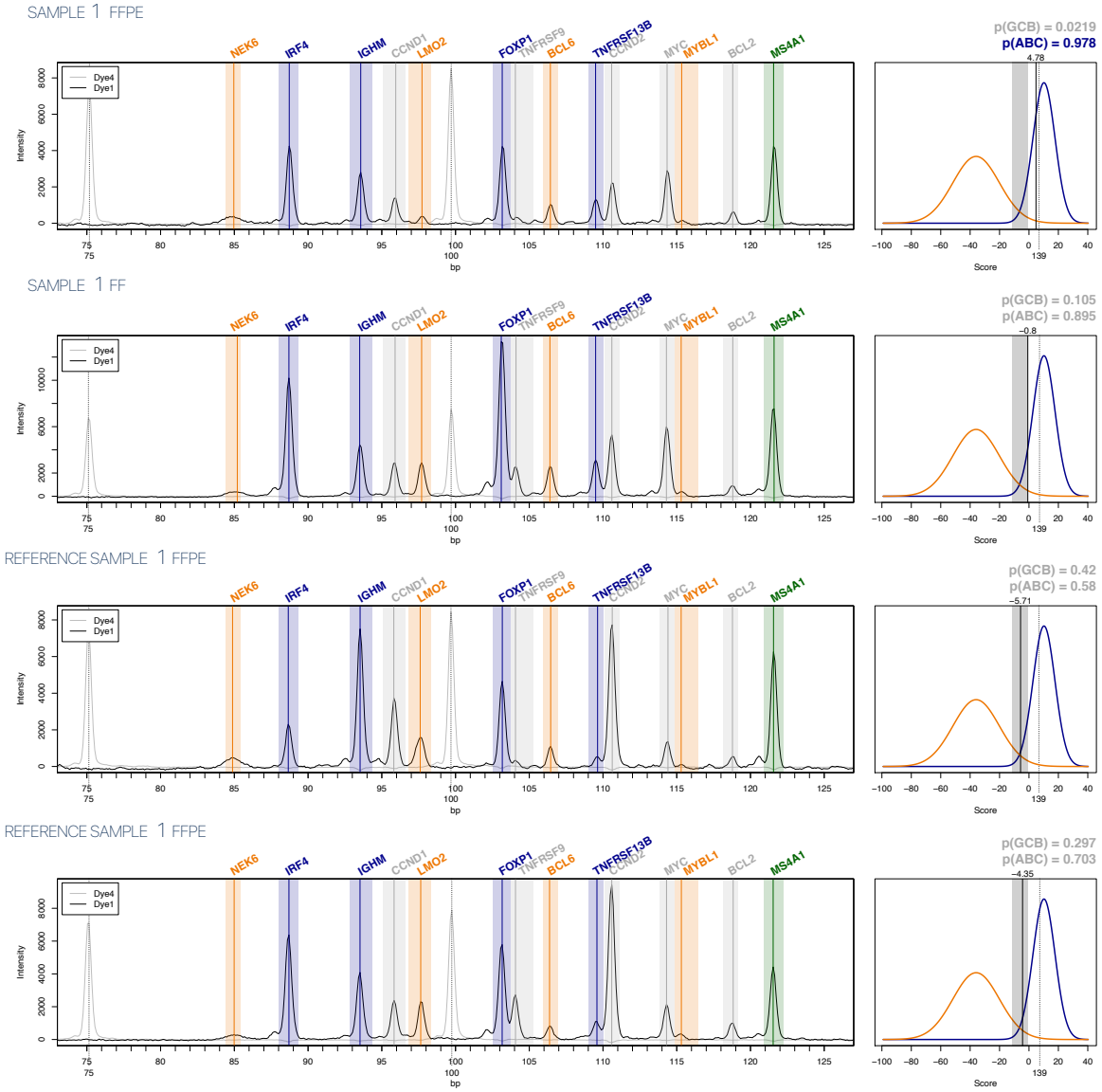


Figure 6. Graphical illustration and GCB/ABC DLBCL classification by MLPA. MLPA profiles from two FFPE/FF paired representative samples (ABC DLBCL, sample 1; a normal (reactive) lymph node, reference sample 1). Left panels: MLPA profiles plotted as intensity of fluorescence as a function of the PCR fragment sizes. Peak intervals are coloured in blue for ABC-related genes, in orange for GCB-related genes and in grey for diagnostic- and prognostically relevant genes. Green highlights the internal control MS4A1 (encoding for CD20). Size marker (Dye 4) is shown with dotted lines, indicating the peak theoretical size. Right panel: Predictive classification of the samples, displaying superposition of the MLPA sample score on the theoretic distributions of scores in both subgroups.

MLPA-BASED DLBCL SUBTYPE CLASSIFICATION CORRELATES WITH ITS IHC SURROGATE

A DLBCL subtype classification previously determined with IHC –based on the Hans algorithm (CD10, MUM1, BCL6, *see Introduction*) was available for IC-DLBCL. *Table 2* collects sample calls and GCB and ABC/non-GCB probabilities according to these two methods. MLPA-based subtype classification was performed with two confidence thresholds, 80% and 90% for better specificity (discarded samples were not included in the subsequent calculations). The predictor classified 73% and 80% of the FF samples in accordance to the Hans algorithm using a threshold of 90% and 80% respectively. A 70% of agreement was reached for FFPE samples when using both thresholds. For those FFPE samples for which RT-MLPA profiles were obtained, similar matching/non-matching results were achieved when compared to their FF counterparts.

The RT-MLPA predictor considered unclassifiable a 13% (2/5) of the FF samples that did belong to one of the DLBCL subtypes according to IHC when using a CI of 90%. Both samples were classified as ABC (one in accordance to IHC) under a CI of 80%. Furthermore, in this FF series a 13% and 20% disagreement (90% and 80% CIs) was reached on samples classified by both techniques. Overall, lowering the confidence threshold from 90% to 80% seemed to increase the probability of agreement between both techniques but also increases the probability of mismatching results.

Table 2. DLBCL subtype classification for each FFPE/FF sample according to RT-MLPA and IHC (Hans algorithm)

Sample	Subtype according to IHC	Subtype according to MLPA 90%		Subtype according to MLPA 80%	
non-GCB IC-DLBCL		FFPE	FF	FFPE	FF
Sample 1	non-GCB	ABC (0,978)	Unclassif. [ABC (0,895)]	ABC (0,978)	ABC (0,895)
Sample 2	non-GCB	ABC (0,99)	ABC (0,994)	ABC (0,99)	ABC (0,994)
Sample 3	non-GCB	ABC (0,971)	ABC (0,908)	ABC (0,971)	ABC (0,908)
Sample 4	non-GCB	Discarded	ABC (0,994)	Discarded	ABC (0,994)
Sample 5	non-GCB	Unclassif. [GCB (0,866)]	GCB (0,983)	GCB (0,866)	GCB (0,983)
Sample 6	non-GCB	GCB (0,983)	GCB (0,999)	GCB (0,983)	GCB (0,999)
GCB IC-DLBCL					
Sample 7	GCB	GCB (1)	GCB (1)	GCB (1)	GCB (1)
Sample 8	GCB	GCB (1)	GCB (1)	GCB (1)	GCB (1)
Sample 9	GCB	Discarded	GCB (1)	Discarded	GCB (1)
Sample 10	GCB	Discarded	GCB (1)	Discarded	GCB (1)
Sample 11	GCB	Discarded	Unclassif. [ABC (0,873)]	Discarded	ABC 0,873
Sample 12	GCB	GCB (1)	GCB (1)	GCB (1)	GCB (1)
Reference sample 1	Unclassified	Unclassif. [ABC(0,58)]	Unclassif. [ABC(0,703)]	Unclassif. [ABC(0,58)]	Unclassif. [ABC(0,703)]
Reference sample 3	Unclassified	Discarded	Unclassif. [ABC(0,796)]	Discarded	Unclassif. [ABC(0,796)]
Reference sample 4	Unclassified	GCB (0,997)]	Unclassif. [ABC(0,524)]	GCB (0,997)]	Unclassif. [ABC(0,524)]
	Matching results	70%	73%	70%	80%
	Unclassified	10%	13%	0%	0%
	Not-matching results	20%	13%	30%	20%

Interestingly, the inclusion of normal (reactive) lymphadenopathy (non-DLBCL) patients in this study resulted as expected, in unclassifiable samples into any of the DLBCL subtypes, with the exception of just one case (reference sample 4) derived from a FFPE block.

FFPE TISSUE PROCESSING AFFECTS THE DETECTION OF RT-MLPA GENES

To assess the effects of the FFPE procedure (fixation, paraffin-embedding and storage) on the expression profile of the individual MLPA probes, gene expression values from the matched-paired FF/FFPE samples were plotted. Since many of the RT-MLPA profiles that failed were part of the GCB group, in Figure 7 only FF/FFPE gene expression values for the ABC/non-GCB group are displayed. We found the expression profiles obtained from FF and FFPE to be highly correlated (non of the comparisons were significantly different), however the sample size was small. Overall, FF gene expression values were slightly higher when compared to their FFPE equivalents. Looking at individual samples (dots), expression values for some genes (i.e. BCL6, MYBL1, IRF4, IGHM, TNFRSF13B, CCND2, BCL3, MS4A1) were only detectable at the FF level (apparent as single dots with no connecting lines). The paired study design was evident for some FF/FFPE pairs that clustered together, especially for those genes whose expression levels were generally low (NEK6, BCL6, MYBL1, BCL2), whereas other pairs showed a greater spread (LMO2, IGHM, CCND2).

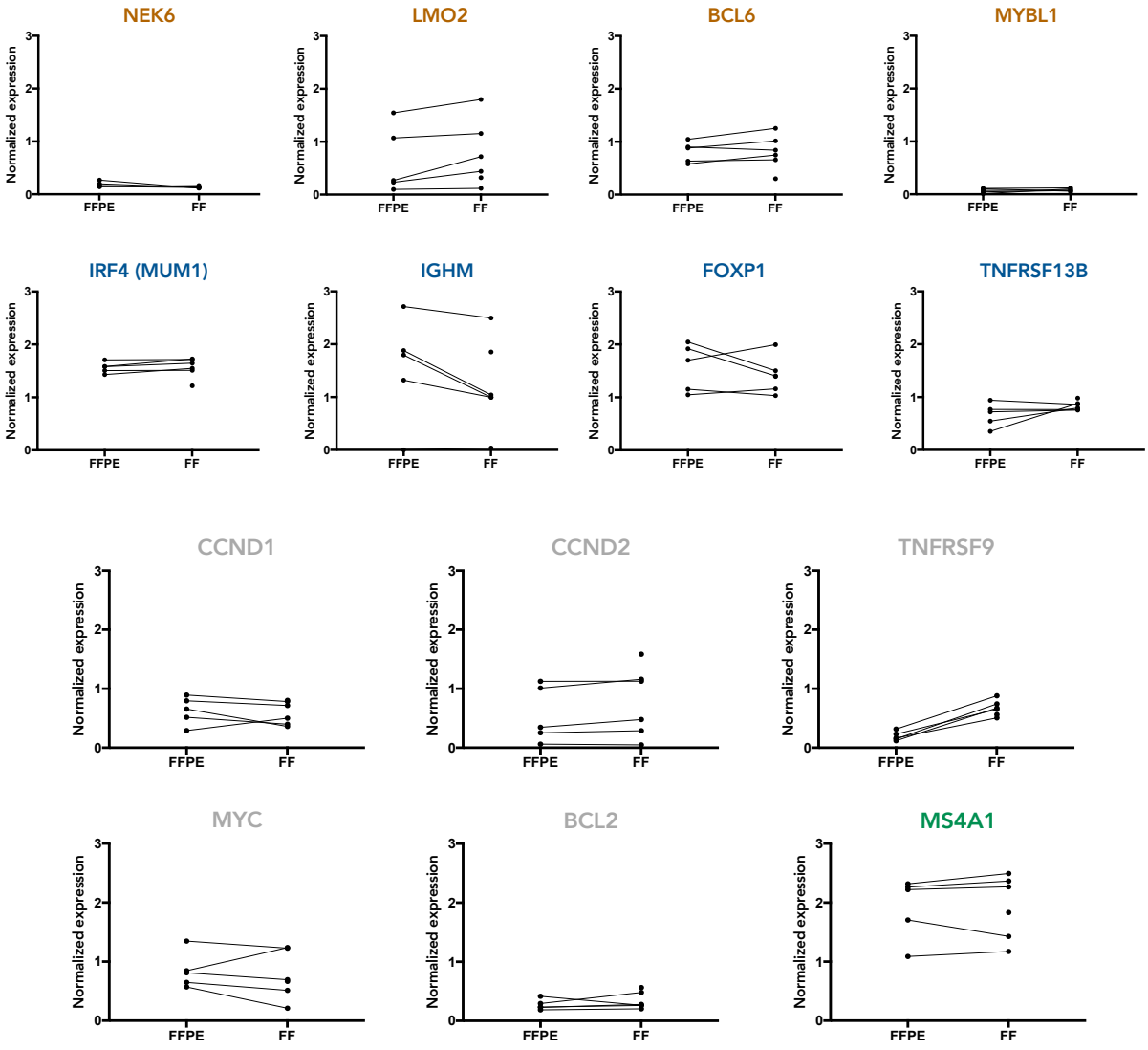


Figure 7. Normalized peak heights (expression values) for the 14 genes included in the RT-MLPA predictor for paired FFPE/FF samples from 6 ABC (non-GCB) DLBCL patients. Gene names are coloured in blue for ABC-related genes, in orange for GCB-related genes, in grey for diagnostic- and prognostically relevant genes and in green for the internal control MS4A1 (encoding for CD20). The lines connect matchi ng FF/FFPE samples.

MUM1, TNFRSF13B AND MYBL1 EXPRESSION IS SIGNIFICANTLY ASSOCIATED WITH THE ABC/GCB SUBGROUPS

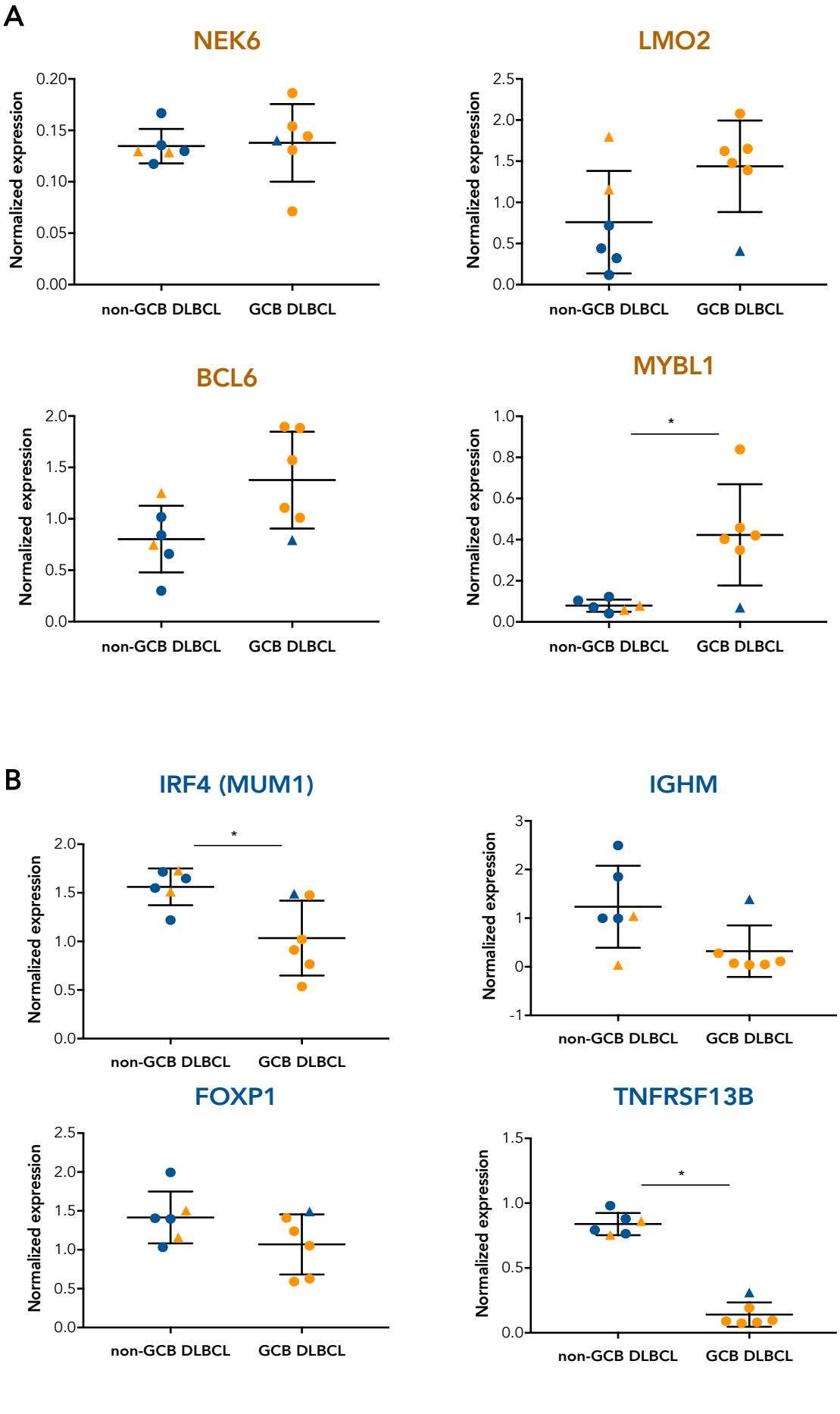
Analysis of the individual RT-MLPA gene expression values in FF samples was performed to validate the association of the probes with either the ABC or GCB profile. Single gene normalized peak height values for all FF samples are displayed in Figure 8, and the statistical p-values obtained are collected in Table 3.

In our series, three out of fourteen genes reached statistical significance: two ABC DLBCL related genes (MUM1, TNFRSF13B) and one GCB-related gene (MYBL1), although BCL6 showed a considerable trend towards lower expression in the non-GCB DLBCL group (p-value 0,0649). Similar statistical differences were found in the FFPE counterparts (IRF4, p-value 0,0357; TNFRSF13B, p-value 0,0159 and MYBL1, p-value 0,057). A possible explanation of why not all classifier probes were significantly differentially expressed between these two groups might be the small sample size of the study.

Table 3. Statistical comparison between ABC and GCB molecular subtypes from FF samples of IC-DLBCL patients. P-values < 0,05.

ABC vs GCB DLBCL	NEK6	LMO2	BCL6	MYBL1	MUM1	IGHM	FOXP1
	0,3095	0,132	0,0649	0,026	0,0087	0,132	0,3095
	TNFRSF13B	CCND1	TNFRSF9	CCND2	MYC	BCL2	CD20
	0,0022	0,6991	0,6991	0,6991	0,5887	0,1797	0,2403

Notably, in the non-GCB (based on IHC) the highest LMO2 expression was detected in sample 5 and 6, classified as GCB based on MLPA. Similarly, sample 11, designated as GCB based on IHC but as ABC based on MLPA was characterised by low expression of GCB-related genes (LMO2, BCL6, MYBL1), and high expression of ABC-related genes (MUM1, IGHM, FOXP1, TNFRSF13B) (Figure 8.A-B). When comparing these two DLBCL subtypes among diagnostic- and prognostically relevant genes, no major differences were found (Figure 8.C).


A
NEK6
Normalized expression
0.20
0.15
0.10
0.05
0.00
non-GCB DLBCL
GCB DLBCL
LMO2
Normalized expression
2.5
2.0
1.5
1.0
0.5
0.0
non-GCB DLBCL
GCB DLBCL
BCL6
Normalized expression
2.0
1.5
1.0
0.5
0.0
non-GCB DLBCL
GCB DLBCL
MYBL1
Normalized expression
1.0
0.8
0.6
0.4
0.2
0.0
*
non-GCB DLBCL
GCB DLBCL
B
IRF4 (MUM1)
Normalized expression
2.0
1.5
1.0
0.5
0.0
*
non-GCB DLBCL
GCB DLBCL
IGHM
Normalized expression
3
2
1
0
-1
non-GCB DLBCL
GCB DLBCL
FOXP1
Normalized expression
2.5
2.0
1.5
1.0
0.5
0.0
non-GCB DLBCL
GCB DLBCL
TNFRSF13B
Normalized expression
1.5
1.0
0.5
0.0
*
non-GCB DLBCL
GCB DLBCL

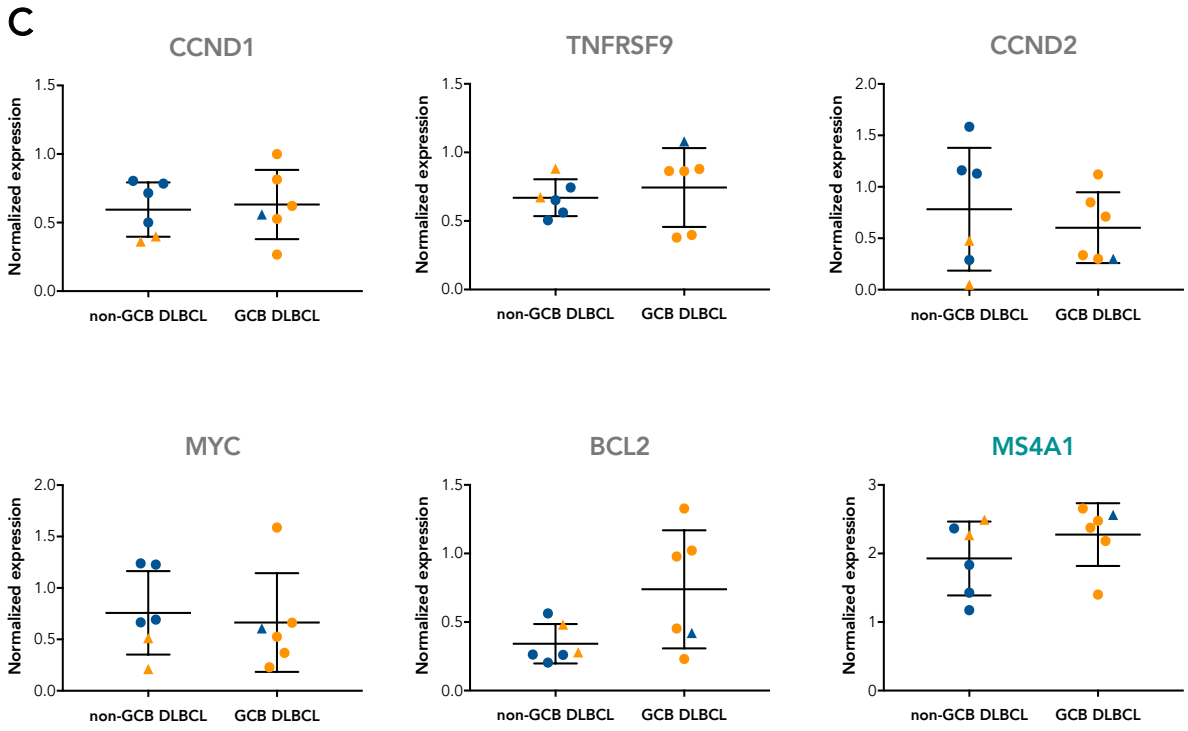


Figure 8. Normalized peak heights (expression values) for the 14 genes included in the RT-MLPA predictor for FF samples from 6 ABC (non-GCB) and 6 GCB DLBCL patients. Gene names are coloured in orange for GCB-related genes (A), in blue for ABC-related genes (B), in grey for diagnostic- and prognostically relevant genes and in green for the internal control MS4A1 (encoding for CD20) (C). Samples are represented as dots when the RT-MLPA subtype classification (confidence threshold of 90%) matches the previously available IHC classification. Samples represented as triangles indicate mismatched RT-MLPA / IHC-based DLBCL subtype classification. Dots/triangles coloured in orange indicate GCB sample calls, whereas blue colour indicates non-GCB DLBCL subgroup.
Data are shown as individual normalized expression values ± SE (n=12). (Mann Whitney U test). Labelling with *, $p<0,05$ (comparison between DLBCL subgroups)

RT-MLPA SHOWN TO BE ROBUST AND REPRODUCIBLE

To assess the reproducibility of the RT-MLPA assay, we performed four independent repeats for two paired FFPE/FF samples (sample 3 and sample 7, ABC and GCB, respectively). The calculated scores remained highly stable for frozen (mean ± SD, -40 ± 1,59 for sample 7; 0,083 ± 0,37 for sample 3) as well as for paraffin (mean ± SD, -46 ± 1,94 for sample 7, 1,96 ± 1,22 for sample 3) (Data not shown).

To further address the robustness of the assay, we also tested nine serial one-half RNA dilutions from 250 ng to < 1 ng RNA of the three-paired FFPE/FF samples (sample 1 and sample 3 being ABC samples, and sample 7 as GCB sample). The normalized peak height values for 14 genes (Figure 9.A) remained very stable from 250 ng to approximately 15,6 ng of RNA for FF samples, whereas greater variability was observed in both FFPE samples. Nevertheless, the resulting RT-MLPA scores and associated GCB/ABC probabilities for these three samples (for both paraffin and frozen) were stable for an RNA amount of as low as 7,8 ng (Figure 9.B). Taken together, these results confirmed that RT-MLPA method for gene expression profiling is robust and reproducible.

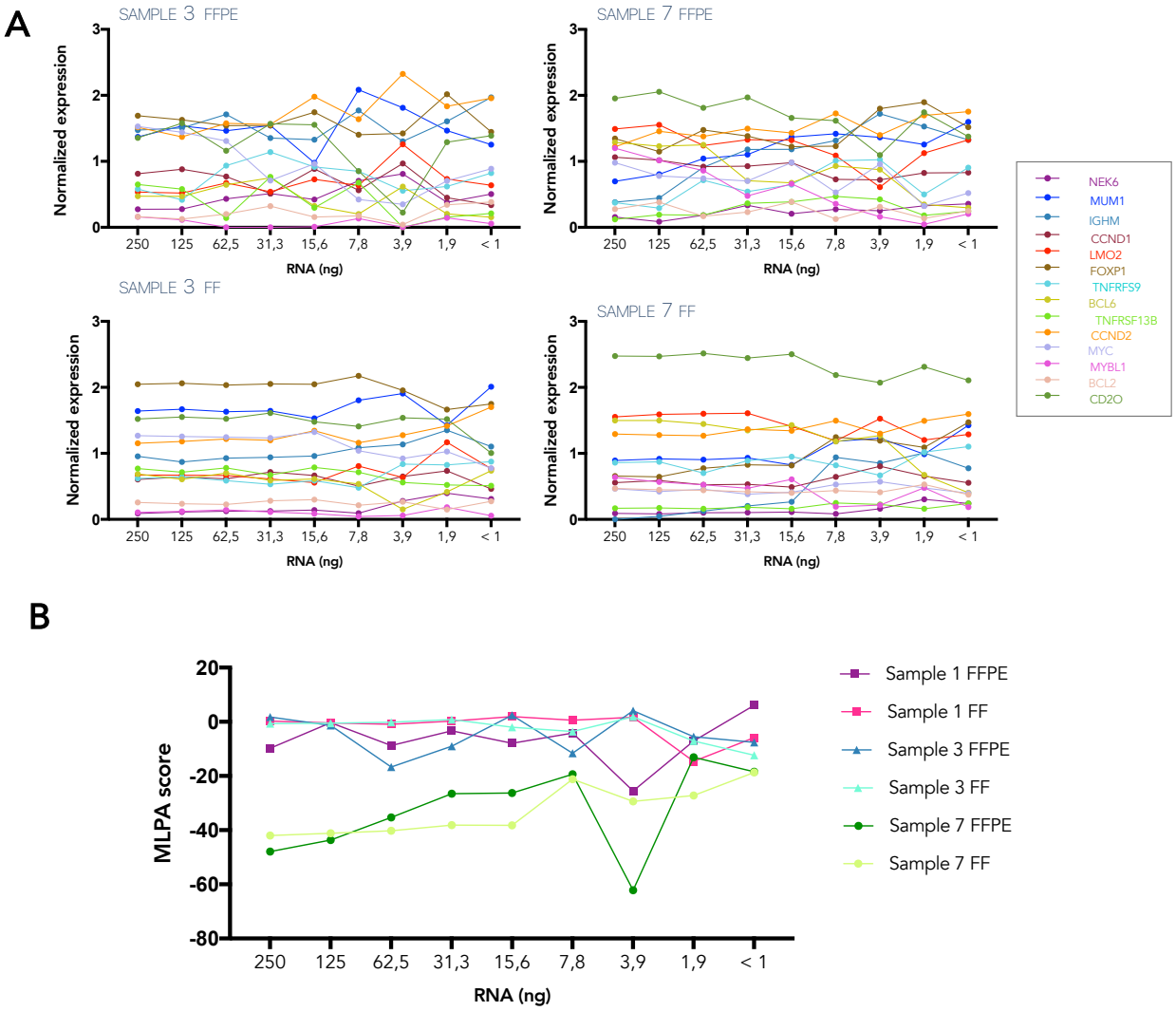


Figure 9. Robustness of the RT-MLPA predictor. (A) Normalized peak heights for the fourteen genes included in the RT-MLPA predictor in the nine serial dilutions of paired FFPE/FF samples 3 and 4 are illustrated. (B) RT-MLPA scores for the RNA dilution series of samples 1,3 (both ABC) and 7 (GCB).

GENE AND PROTEIN EXPRESSION CORRELATED FOR MUM1, BCL6, FOXP1 AND IGHM

The next step was to evaluate the correlation between the level of gene expression (assed by means of RT-MLPA) and the level of protein expression (of MUM1, BCL6, LMO2, FOXP1, IGHM, BLC2, MYBL1, MYC) within the tumour area in those samples (relative protein expression). The CD10 immunostaining (a strong GCB marker) was performed for each sample as part of the Hans algorithm applied in the original case selection (data not shown). The mean and median of relative protein expression percentages of all samples are collected in *Table 4* below. Detailed absolute and relative protein percentages for each sample are gathered in *Supplemental Data, Table S.1.*

Table 4. Protein expression of a subset of GCB and ABC related genes included in the RT-MLPA classifier for which staining could be performed.

	CD20		MUM1		BCL6		LMO2		FOXP1		IGHM		BCL2		MYBL1		MYC	
	MD %	MN %	MD %	MN %	MD %	MN %	MD %	MN %	MD %	MN %	MD %	MN %	MD %	MN %	MD %	MN %	MD %	MN %
non-GCB IC- DLBCL	80	77,5	66	56,33	57	48,67	54	48,6	77,5	64	68	58,5	65	56,67	8	9,8	18	25,2
GCB IC- DLBCL	90	83,33	1,75	6	49,5	55,17	90	82	90	83,33	0	15	33,5	38,17	13,5	9,667	12	20,33

p-value									
non-GCB IC DLBCL vs GCB IC-DLBCL	0,2424	0,0498	0,9069	0,0317	0,1515	0,1126	0,2879	0,9286	0,6558

In line with the results obtained in the gene expression experiments, MUM1 (IRF4) was higher in the non-GCB group at the protein level (p-value 0,0498). Conversely, MYBL1 that reached statistical significance at the gene expression level was similarly expressed at the protein level in both groups. Based on the immunostaining, LMO2 was markedly higher in the GCB group, and its staining pattern was mostly nuclear compared to a consistent reduced cytoplasmic staining in the non-GCB group. It is nonetheless noteworthy that LMO2 and MYBL1 are particularly difficult to evaluate by IHC. No TNFRSF13B antibody was available to correlate the gene expression difference observed to that at the protein level. The rest of the immunostainings assessed displayed similar positivity in both DLBCL subgroups. The relative IHC staining percentages of all samples are displayed in *Figure 10*.

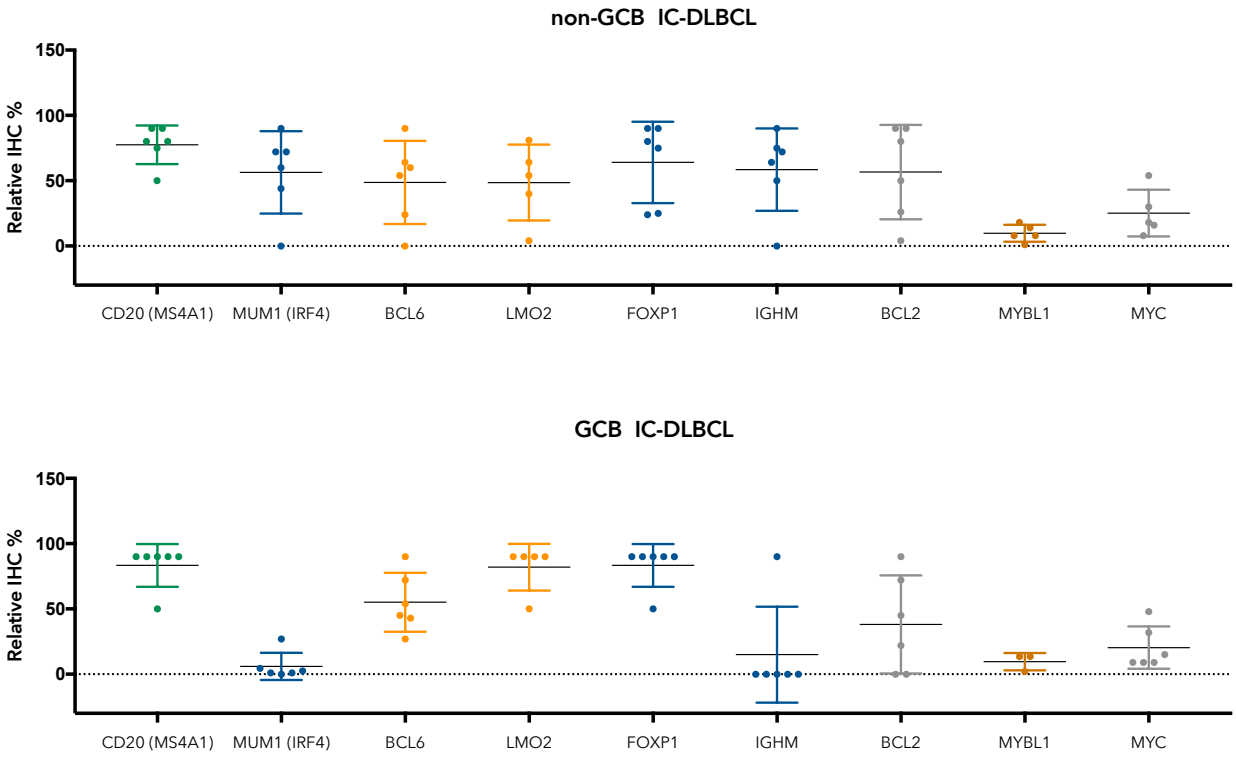


Figure 10. Immunohistochemical analysis of the relative protein expression within tumoural areas of several ABC/GCB markers that are included in the RT-MLPA gene panel. Samples are coloured according to their gene groups: orange for GCB-related genes, in blue for ABC-related genes, in grey for diagnostic- and prognostically relevant genes and in green for the internal control MS4A1 (encoding for CD20). Since protein expression in some cases is very low, dotted line marks the Y axe at point 0 allowing the a better interpretation of the SE.

CLASSIFICATION OF THE PT-DLBCL SERIES

EBV+ IGM- (ABC)-PT-DLBCL IS CONSISTENLY CLASSIFIED WITHIN THE GCB DLBCL SUBTYPE ACCORDING TO THE RT-MLPA CLASSIFIER

Following optimisation of the MLPA assay, a positive selection of twenty-four PT-DLBCL patients was used to evaluate how the presence of EBV in the tumour cells influences the molecular DLBCL subtype classification of these samples based on the MLPA predictor. ABC/GCB subgroups determined with IHC–based on the Hans algorithm was available for all samples, and additional microarray data was available for eighteen of the samples. In accordance with the literature, all EBV+ PT DLBCL cases except one (case 2.3) were ABC-derived DLBCLs [52] (Table 5).

GCB and ABC sample calls and the probabilities of belonging to one group or the other are gathered in the *Table* 5 below. This RT-MLPA classification was based upon two confidence thresholds (80% and 90%). For this series, using a CI of 90% the predictor only classified a 37% of the samples within their expected subgroups, failing to classify a 27% of the cases and misclassifying the other 37%. When using a 80% CI, the percentages of matching results as well as of mismatches increased, being the latter greater.

EBV^+ PT-DLBCLs are most frequently of ABC origin and within this group it appears that two subgroups exist: IgM^+ and IgM^- EBV^+ PT-DLBCL (*see Introduction*). Remarkably, most sample miscalls were found grouped together within the IgM^- EBV^+ PT-DLBCL, suggesting that these samples present extremely high probabilities of belonging to the GCB group instead of to the ABC group. Moreover, a few strong mismatches are also found within the EBV^- PT-DLBCL subgroup, in a similar ratio as EBV- ABC DLBCL in IC patients (50% versus 67% respectively).

To define the cause of these discordant classifications, we evaluated the gene expression profile of each sample for every gene included in the series. Statistical tests for significantly affected genes between these three groups were conducted. For each individual gene, normalized peak height values for each sample are plotted in *Figure 11*, and the statistical p-values obtained are collected in *Table 6*.

Table 5. DLBCL subtype classification according to RT-MLPA, IHC (Hans algorithm) and microarray GEP data

Sample	Subtype according to microarray	Subtype according to IHC	Subtype according to MLPA	
EBV- PT-DLBCL			90%	80%
Sample 2.1	ABC	non-GCB	GCB (1)	GCB (1)
Sample 2.2	ABC	non-GCB	GCB (0,957)	GCB (0,957)
Sample 2.3	GCB	GCB	GCB (1)	GCB (1)
Sample 2.4	ABC	non-GCB	ABC (0,969)	ABC (0,969)
Sample 2.5	ABC	non-GCB	ABC (0,994)	ABC (0,994)
Sample 2.6	ABC	non-GCB	Unclass. GCB (0,593)	Unclass. GCB (0,593)
Sample 2.7	ABC	non-GCB	GCB (0,928)	GCB (0,928)
EBV+ IgM- PT-DLBCL				
Sample 2.8	ABC	non-GCB	GCB (0,999)	GCB (0,999)
Sample 2.9	ABC	non-GCB	GCB (0,996)	GCB (0,996)
Sample 2.10	ABC	non-GCB	GCB (0,999)	GCB (0,999)
Sample 2.11	ABC	non-GCB	Unclass. ABC (0,895)	ABC (0,895)
Sample 2.12	ABC	non-GCB	GCB (0,959)	GCB (0,959)
Sample 2.13	ABC	non-GCB	GCB (1)	GCB (1)
Sample 2.14	ABC	non-GCB	GCB (0,90)	GCB (0,90)
Sample 2.15	ABC	non-GCB	Unclass. GCB (0,869)	GCB (0,869)
Sample 2.16	ABC	non-GCB	Unclass. ABC (0,727)	Unclass. ABC (0,727)
EBV+ IgM+ PT-DLBCL				
Sample 2.18	ABC	non-GCB	ABC (0,985)	ABC (0,985)
Sample 2.19	ABC	non-GCB	ABC (0,985)	ABC (0,985)
Sample 2.21	ABC	non-GCB	ABC (0,938)	ABC (0,938)
Sample 2.22	ABC	non-GCB	ABC (0,996)	ABC (0,996)
Sample 2.23	ABC	non-GCB	Unclassif. [GCB (0,892)]	GCB (0,892)
Sample 2.24	ABC	non-GCB	ABC (0,984)	ABC (0,984)
Sample 2.25	ABC	non-GCB	Unclass. [GCB(0,64)]	Unclass. [GCB(0,64)]
Sample 2.26	ABC	non-GCB	ABC (0,994)	ABC (0,994)
		Matching results	37,5%	41,6%
		Unclassified	25%	12,5%
		Not-matching results	37,5%	45%

Overall, by looking at the statistical comparison between EBV$^-$ PT-DLBCL and each of the EBV$^+$ PT-DLBCLs subgroups separately, no common significant differences were found. As expected, the amount of IGHM expression within the EBV$^+$ IgM$^+$ PT-DLBCLs was significantly higher than in the EBV$^+$ IgM$^-$ and the EBV$^-$ groups, in which expression levels were similar. Opposite results wwere observed for FOXP1.

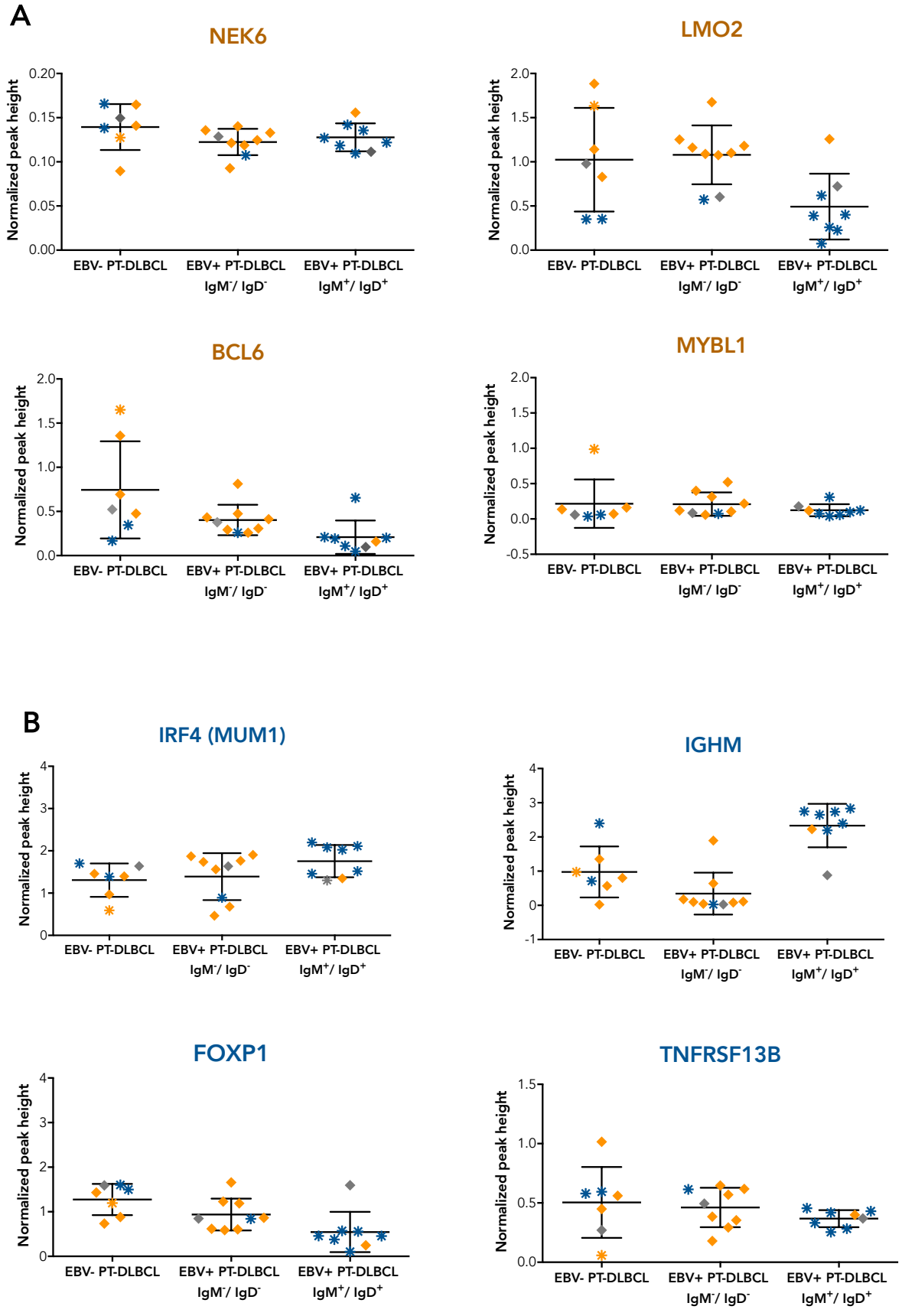
In line with the RT-MLPA based classification results, the most striking differences that we found were between EBV$^+$ PT DLBCL subgroups. Apart from the differences in ABC-related IGHM (p-value 0,0002) and FOXP1 (p-value 0,005), GCB-related LMO2 (p-value 0,0206) and BCL6 (p-value 0,014) were more highly expressed in the IgM$^-$ when compared to IgM$^+$-cases. These expression patterns may explain the GCB MLPA classification of non-GCB EBV$^+$ IgM$^-$ PT-DLBCL (Figure 11). For example, in the EBV$^+$ subgroups (all non-GCB based on IHC) LMO2 normalized expression values for ABC-classified samples by MLPA (blue *-marked) were consistently low among the three groups when compared to mismatches samples GCB-classified by MLPA (orange-coloured and diamond-shaped).

However, this did not seem to be the case when looking at the same samples among ABC-related genes that inclined to be clustered more closely together. This could be due to the fact that LMO2 is thought to be a marker restricted to GC and had the most discriminatory weight among GCB-related genes in the algorithm (see *Introduction*). This high LMO2 gene expression level plus the fact that EBV^+ IgM^- PT-DLBCL samples presented global low levels of IGHM (ABC-related gene) may have been enough to interpret these samples as GCB instead of ABC.

Similarly, marked differences also existed in diagnostic- and prognostically relevant genes (CCND1, p-value < 0,002; TNFRFS9 p-value < 0,0079; and BCL2, p-value 0,0206). This difference in BCL2 gene expression was also noticeable when comparing it to the EBV^- subgroup. High levels of BCL2 have been correlated with a poorer prognosis, suggesting so in this EBV^+ IgM^- PT-DLBCL subgroup (Figure 11.C).

Table 6. Statistical comparison between genes among EBV-, EBV+ IgM- and EBV+ IgM+ PT-DLBCLs . Significant difference p < 0,05

	NEK6	LMO2	BCL6	MYBL1	MUM1	IGHM	FOXP1
EBV^- vs EBV^+ IgM^-	0,1217	0,6065	0,1416	0,351	0,4079	0,0907	0,0907
EBV^- vs EBV^+ IgM^+	0,1893	0,0939	0,014	0,8665	0,152	0,0093	0,0093
EBV^- IgM^- vs EBV^+ IgM^+	0,743	0,0206	0,0055	0,3213	0,2766	0,0002	0,0055
	TNFRSF13B	TNFRFS9	CCND1	CCND2	MYC	BCL2	CD20
EBV^- vs EBV^+ IgM^-	>0,9999	0,9182	0,0418	0,0549	0,4079	0,0033	0,7577
EBV^- vs EBV^+ IgM^+	0,1893	0,014	0,1893	0,2319	0,3969	0,4634	0,0289
EBV^- IgM^- vs EBV^+ IgM^+	0,2359	0,0079	0,0016	0,3704	0,6058	0,0206	0,3704


A
NEK6
Normalized peak height
0.20
0.15
0.10
0.05
0.00
EBV- PT-DLBCL
EBV+ PT-DLBCL IgM-/ IgD-
EBV+ PT-DLBCL IgM+/ IgD+
LMO2
Normalized peak height
2.0
1.5
1.0
0.5
0.0
EBV- PT-DLBCL
EBV+ PT-DLBCL IgM-/ IgD-
EBV+ PT-DLBCL IgM+/ IgD+
BCL6
Normalized peak height
2.0
1.5
1.0
0.5
0.0
EBV- PT-DLBCL
EBV+ PT-DLBCL IgM-/ IgD-
EBV+ PT-DLBCL IgM+/ IgD+
MYBL1
Normalized peak height
2.0
1.5
1.0
0.5
0.0
-0.5
EBV- PT-DLBCL
EBV+ PT-DLBCL IgM-/ IgD-
EBV+ PT-DLBCL IgM+/ IgD+
B
IRF4 (MUM1)
Normalized peak height
4
3
2
1
0
EBV- PT-DLBCL
EBV+ PT-DLBCL IgM-/ IgD-
EBV+ PT-DLBCL IgM+/ IgD+
IGHM
Normalized peak height
4
3
2
1
0
-1
EBV- PT-DLBCL
EBV+ PT-DLBCL IgM-/ IgD-
EBV+ PT-DLBCL IgM+/ IgD+
FOXP1
Normalized peak height
4
3
2
1
0
EBV- PT-DLBCL
EBV+ PT-DLBCL IgM-/ IgD-
EBV+ PT-DLBCL IgM+/ IgD+
TNFRSF13B
Normalized peak height
1.5
1.0
0.5
0.0
EBV- PT-DLBCL
EBV+ PT-DLBCL IgM-/ IgD-
EBV+ PT-DLBCL IgM+/ IgD+

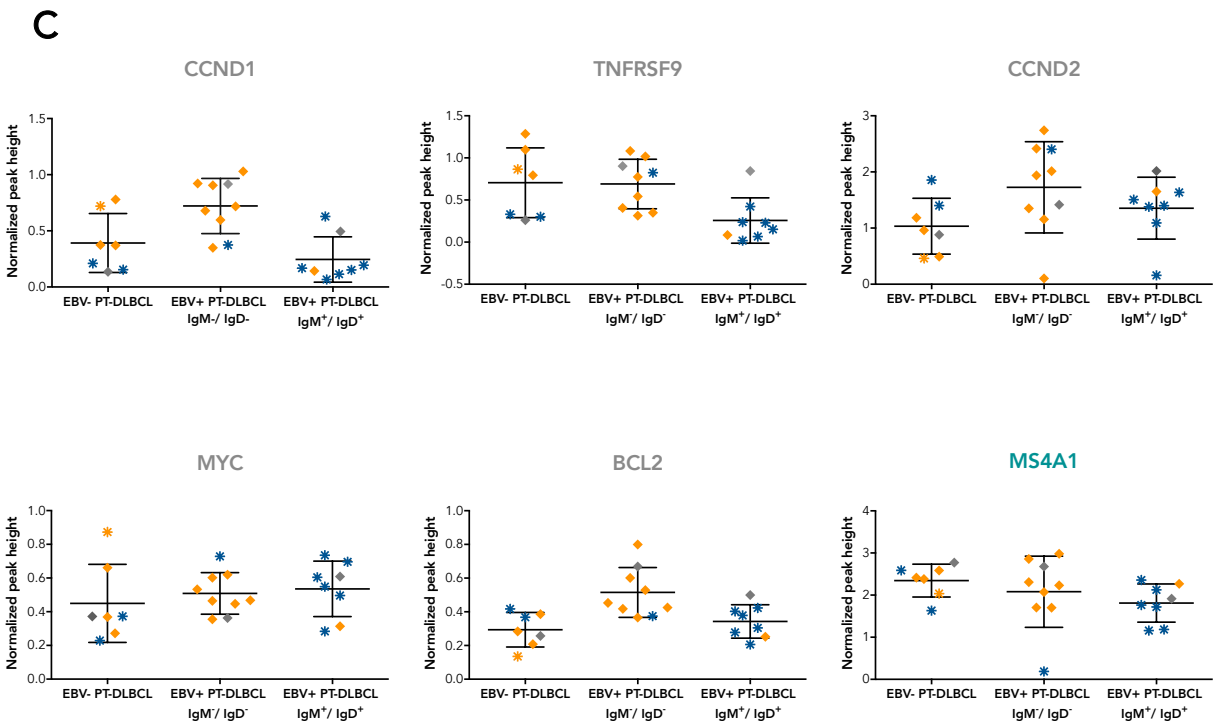


Figure 11. Normalized peak heights (expression values) for the 14 genes included in the RT-MLPA predictor from 7 EBV⁻ PT-DLBCL, 9 EBV⁺ IgM⁻ PT-DLBCL and 8 9 EBV⁺ IgM⁺ PT-DLBCL patients. Gene names are coloured in orange for GCB-related genes (A), in blue for ABC-related genes (B), in grey for diagnostic- and prognostically relevant genes and in green for the internal control MS4A1 (encoding for CD20) (C). Samples are represented as * when the RT-MLPA subtype classification (confidence threshold of 90%) matches the previously available IHC, and microarray data when available, classifications. Samples represented as diamonds indicate mismatched RT-MLPA / IHC DLBCL subtype classification. Dots/diamonds coloured in orange indicate GCB sample calls, whereas blue colour indicates non-GCB DLBCL subgroup. // Data are shown as individual normalized expression values ± SE (n=12).

GENE AND PROTEIN EXPRESSION CORRELATED FOR MUM1 AND IGHM

Alike it was performed with the IC-DLBCL series, we further aimed to correlate these gene expression findings at the protein level for those antibodies that were available to us. The mean and median of relative protein expression percentages of all samples are collected in Table 7 below. Detailed absolute and relative protein percentages for each sample are gathered in Supplemental Data, Table S.2.

Table 7. Immunochemistry for some relevant GCB and ABC related genes included in the RT-MLPA classifier.

	CD20		MUM1		BCL6		LMO2		FOXP1		IGHM		BCL2		MYC	
	MD %	MN %	MD %	MN %	MD %	MN %	MD %	MN %	MD %	MN %	MD %	MN %	MD %	MN %	MD %	MN %
EBV- PT-DLBCL	80	67,14	8,5	18,83	16	24	80	73,67	49,5	49,5	60	53	46,5	46,5	4,18	5,56
EBV+ PT-DLBCL IgM- / IgD-	70	67,78	32	33,57	0	1	70	67,5	0	8,714	0	0	42	42	12	15,56
EBV+ PT-DLBCL IgM+ / IgD+	80	80	80	62	0	0	80	78,5	80	65	80	70	80	80	8	15

p-values								
EBV- PT-DLBCL vs EBV+ PT-DLBCL IgM- / IgD-	0,5251	0,2453	0,0430	0,5416	0,7273	0,0002	>0.9999	0,8646
EBV- PT-DLBCL vs EBV+ PT-DLBCL IgM+ / IgD+	0,6556	0,0355	0,0210	0,5931	0,0431	0,4394	0,5063	0,6126
EBV+ PT-DLBCL IgM- / IgD- vs EBV+ PT-DLBCL IgM+ / IgD+	0,0668	0,0728	>0.9999	0,1588	0,0146	0,0076	0,0146	0,8404

Besides the obvious IGHM differences that define EBV$^+$ subgroups in these series, in concordance with the results obtained in the gene expression experiments, BCL6 protein expression was also significantly higher in the EBV$^-$ PT-DLBCLs grouping compared to the EBV$^+$ IgM$^+$ group. Moreover, BCL6 at the protein level also achieved a significant difference between the EBV$^-$ and the IgM$^-$ groups, which was not observable at the gene expression level. In fact, the difference in BCL6 gene expression between both EBV$^+$ factions completely disappeared at the protein level and this was steadily negative throughout both groups. However, this significance might be not completely real, as within this group, there is a truly EBV$^-$ GCB DLBCL and two ABC samples misclassified as GCB, that together they might pull up such difference that way. Inversely, MUM1 was significantly lowered in EBV$^-$ PT-DLBCLs when compared to its firmly high protein expression in EBV$^+$ IgM$^+$. MUM1 expression within the IgM$^-$ was somewhat more variable.

Distinct results were observed regarding BCL2. While such gene expression was much higher in EBV$^+$ IgM$^-$ grouping, BCL2 at the protein level was consistently positive in all tumoral cells within the EBV$^+$ IgM$^+$ group. No useful comparison to the EBV$^-$ PT-DLBCL group was achieved as many of the BCL2 stainings within this group were missing.

The most remarkable finding concerned FOXP1. Whereas this marker was steadily low in all EBV$^+$ IgM$^+$ PT-DLBCL cases, opposing results were observed in the immunostaings, showing strongly nuclear positivity in all tumoral cells for most of the samples within this group. Inversely, EBV$^+$ IgM$^-$ cases showed no (or very vague) FOXP1 positivity while its gene expression levels much higher.

Finally, all differences observed on BCL2 from the RT-MLPA profiles were no longer appreciable at protein level. Overall protein expression of each assessed antibody is displayed per grouping as relative IHC percentages in *Figure 12*.

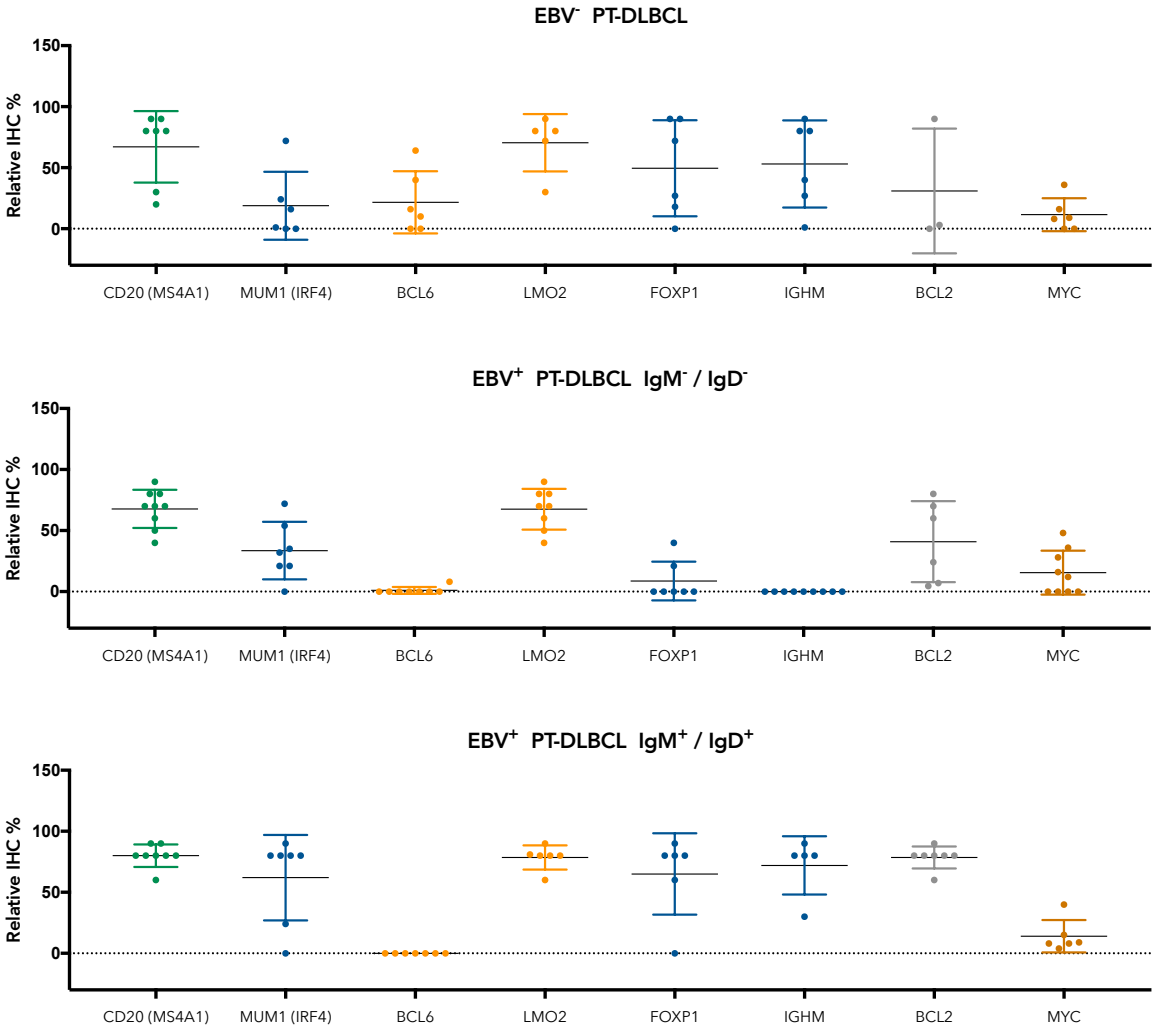


Figure 12. Immunohistochemical analysis of the relative protein expression within tumoural areas of several antibodies that are included in the RT-MLPA gene panel. Samples are coloured according to their gene groups: orange for GCB-related genes, in blue for ABC-related genes, in grey for diagnostic- and prognostically relevant genes and in green for the internal control MS4A1 (encoding for CD20) (**C**). Since protein expression in some cases is very low, dotted line marks the X axe at point 0 allowing the a better interpretation of the SE.

LYMPHOBLASTOID CELL LINES (LCLs)

LCLs SHOWED SENSITIVITY TO TREATMENT COMBINATIONS OF BORTEZOMBIB, RAPAMYCIN AND IBRUTINIB

Based on the Hans algorithm and RT-MLPA profiles, all five LCL were shown to be of ABC origin (whose MLPA scores were: LCL2, 9,61; LCL4, 11,7; LCL7, 11,7; LCL9, 17,4; LCL10, 9,3). The LCLs were treated with bortezomib, rapamycin and ibrutinib holding much promise towards targeted therapies for the treatment of ABC DLBCL in the general population.

In all five LCLs, combination treatment markedly decreased cell viability relative to either compound alone (Figure 13).

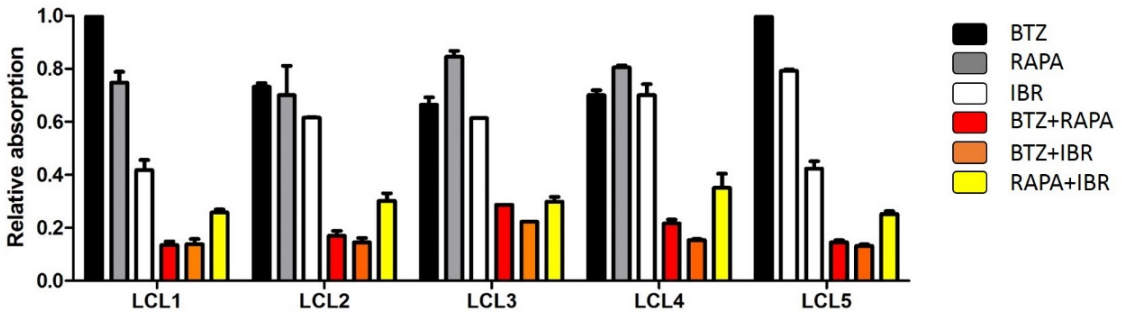


Figure 13. Five LCL were submitted to a 24h treatment with bortezomib (BTZ; 65 nM), rapamycin (RAPA; 1,25 mM), ibrutinib (IBR, 5mM) or a combination of two compounds. In all cases, combination of bortezomib with either rapamycin or ibrutinib was most potent.

DISCUSSION

Clinically and biologically distinct molecular phenotypes with distinct pathogenic driving mechanisms are now well recognised in DLBCL tumour biology and carry much promise, pointing towards individualised therapy based on molecular characteristics. GCB and ABC molecular DLBCL subgroups have been recognised as DLBCL subtypes in the revised 2016 World Health Organization classification of lymphoid neoplasms [60] underlining the need for reliable tools to accurately determine the DLBCL COO.

Although GEP-based methods remain the gold standard in determining the COO, to date, they still remain highly inaccessible, pricey and generally require dedicated platforms and/or fresh frozen material, which makes them poorly applicable to routine diagnosis as most hospitals mainly store patient biopsies as formalin-fixed paraffin-embedded tissue. To address this issue, several IHC surrogate methods have been proposed, for which several conflicting results can be found in the literature. Whereas several large analyses do find decent IHC-GEP correlations, others found them devoid of standardisation, precision and reliability, which dissuades against their routine clinical use. To this end, Mareschal et al. in a recent publication [22] described a robust and cost-effective RT-MLPA assay that could challenge IHC and other quantitative GEP methods to facilitate the stratification of patients according to their molecular DLBCL phenotypes. This method enables a quick evaluation of the relative expression of a fourteen gene predictive panel in a single reaction.

Accordingly, in this work we took as a first step to set up a pilot study to optimise this GEP-based technique in our lab, with the aim to eventually be implemented as a novel diagnostic assay at UZ Leuven. Although we followed their protocol closely, the variability in RNA expression levels combined with a relatively low-dynamic range of the RT-MLPA [61] necessitated that we adjust the protocol according to our conditions. As it turned out, an MLPA probe set that performs well on a given tissue or under certain devices/settings may require further adjustments when using other conditions.

One of the claimed advantages of the RT-MLPA over other mRNA expression profiling techniques is the possibility of using highly degraded RNA, such as RNA derived from FFPE tissue [62]. Whereas RNA isolations from FFPE biopsies resulted in as high purity ratios, these RNAs were completely fragmented in comparison to the intactness of the FF-derived RNAs. Some FFPE samples were excluded from the study since their RT-MLPA profiles were not reliable. The reason why some FFPE blocks perform better than others remains to be investigated. Two possible explanations are the variability in tissue collection and processing, or/and the amount of storage time. While it is difficult to estimate the expected quality of extracted RNA based on the available information for a given FFPE block, two reports in the literature showed that longer storage periods are irrelevant to the quality of the RNA [63, 64]. This should be further validated for RT-MLPA analysis.

Nevertheless, in most cases FFPE samples did succeed in generating MLPA probe signals that in fact, in some genes were slightly higher than the ones observed in their FF-counterparts. For those paired FF/FFPE samples included in the study, similar RT-MLPA profiles were obtained, and most frequently

classified (or unclassified) within the same DLBCL subgroup. These results confirm that small stretches of cDNA are sufficient to generate RT-MLPA probe signals, since probe-specific primers can partly overlap with their corresponding probes and need to be elongated by merely fifty nucleotides.

No statistical differences in individual gene expressions were found between paired FF/FFPE samples. When analyzing these results, it is important to keep in mind that FF/FFPE samples are indeed sampled from two different tissue locations, and that as a result, tissue heterogeneity (percentage of tumour cells, percentage of necrosis and other differences due to tumour-microenvironment) can generate background variation, hence differences between FF and FFPE are expected [63]. Overall, the data in this pilot study come from a small sample size, and should naturally be considered as preliminary. Decisions regarding how much data deviation can be tolerated when comparing FFPE and FF tissues should also be decided.

The RT-MLPA predictor classified 73% and 80% of the frozen cohorts within their expected subgroups –based on the IHC Hans algorithm (CI of 90% and 80%, respectively). These proportions are in line with the 82% (also based on IHC) originally described in Mareschal et al., but lower than the 90% and 93% – according to GEP-based DASL data (from two separate validation series), for which a CI of 95% was used [22]. Nevertheless, these proportions are consistent with the performances obtained with other gene expression-based methods: 76% [65] or 84% [27] (cut-offs 90% and 80% respectively). We observed a 13% (CI, 90%) of unclassifiable samples that was slightly lower than the 17% (CI, 95%) observed in and Mareschal's publication (also grounded on IHC data). These results are in agreement with what is typically observed between Affymetrix classifications and Hans algorithm, which identifies GCB cases but do not differentiate unclassified cases from ABC cases (both referred as non-GCB cases) [6, 22]. Further, performance of the RT-MLPA predictor was supported by the non-neoplastic (reference) samples included in the study that as we expected were considered unclassifiable according to both cut-off thresholds, 80% and 90%.

In our IC-DLBCL series, discordances between IHC and RT-MLPA classifications were observed. Within the non-GCB DLBCL subgroup, two cases were designated as GCB probably due to their high levels of LMO2 –the most discriminatory GCB-related gene included in the predictive panel, and also of BCL6. Reassessment of the stainings confirmed non-GCB classification based on high levels of MUM1. Similarly, a designated GCB based on IHC but as ABC on MLPA was characterised by low expression of GCB-related genes and high expression of ABC-related genes. Based on the immunostainings, this sample was CD10 positive, described as a restrict marker for GCB-DLBCL. These IHC-MLPA discordances may have been due to the presence of normal lymph node tissue and/or the amount of stromal tissue present in the biopsies. If there is a significant amount of non-tumour tissue present, gene expression data not only reflect the gene expression profile of the tumour and thus influence their molecular sub-classification.

In this regard, immunohistochemistry does allow a more direct visualisation and co-localisation of the neoplastic cells (start with protein). According to antibody availability, correlations between relative gene and protein expressions were only possible for nine out the fourteen genes included in the MLPA. In this IC-DLBCL series, MUM1 (IRF4) was the only marked difference that correlated at the protein-gene level and opposing results were found for two GCB markers, LMO2 and MYBL1. It is though

noteworthy that we found these two markers particularly tricky to evaluate by IHC. On one hand, MYBL1 staining has not been used in DLBCL diagnosis and –apart from tumour cells– it seemed to also stain (stromal) macrophages. On the other hand, LMO2 has been constantly characterised by strong nuclear staining [15, 66], however in both our case series (IC and PT-DLBCL) LMO2 was also expressed in the cytoplasm. Besides, LMO2 has extensively described as to be restricted to GC B cells at gene and protein level [8, 23, 24]; and fully in keeping with these gene-expression profiling studies, Natkunam et al. showed that LMO2 protein is expressed in several GC-derived B-cell lymphomas (i.e. follicular, Burkitt, DLBCL) [15]. In our study, LMO2 at gene expression level showed no significant differences between ABC and GCB DLBCLs, and to some extent, non-GCB DLBCL cases showed some LMO2 positivity at the protein level. As argued in [15], that could be attributed to an "unrestricted" expression of GC markers in a subset of non-GC-derived DLBCLs due to a constant B-cell differentiation where specific GC markers are most likely to be up- or down- regulated at any particular stage. However, this could be the case for any described GC marker when defining DLBCL phenotypes. Alternatively, this could suggest a misclassified proportion of DLBCL cases by the Hans algorithm. In fact, as previously mentioned, their original publication misclassifies an approximate 20% of the cases in respect to the gold standard assessed on the same cases [6]. Importantly, Natkunam et al. even considers the possibility that the GCB-DLBCL subgroup comprises more than one molecular signature.

The molecular pathogenesis of immunodeficiency-associated DLBCL (e.g. PT-DLBCL) differs extensively from that of DLBCL in IC hosts. In fact, a large fraction of these DLBCLs arising in immunocompromised individuals harbour EBV. Accordingly, as the second goal of this project we applied the RT-MLPA classifier to a PT-DLBCL to evaluate whether EBV positivity in the tumoral cells influence the MLPA-based DLBCL subgroup classification.

Whereas EBV^+ PT-DLBCLs are most frequently of ABC origin, in this series the RT-MLPA predictor misclassified a 37% of our EBV^+ (ABC) PT-DLBCLs samples as GCB DLBCL phenotype; failed to classify a 27% of the samples and only classified as expected the other 37%. Surprisingly, and we believe that not coincidentally, according to the recent discovery of two EBV^+ PT-DLBCLs subgroups: IgM^+ and IgM^- EBV^+ PT-DLBCL *(see Introduction*), most samples miscalls were found within the same grouping, EBV^+ IgM^- PT-DLBCL. As expected, IGHM gene expression in EBV^+ IgM^+ was significantly higher than in the EBV^+ IgM^- and the EBV^- groups. However, as compared to most IGHM negative GCB-DLBCLs in IC individuals (and classified as such by the MLPA), this phenomenon alone cannot explain these misclassifications. Although IGHM is highly expressed in most EBV^- ABC-DLBCL in IC patients, an interesting observation would be to look at the IGHM levels found in EBV-driven ABC-DLBCL in IC patients. Such comparison is, however, highly unfeasible as DLBCLs in IC are mostly EBV^-.

Taken all together: with very low IGHM, decreased MUM1 expression and significantly increased LMO2 and BCL6 expression in all IgM^- cases in comparison to those in the IgM^+ group, it is not surprising that the RT-MLPA predictor practically classified all IgM^- samples as GCB instead of as ABCs. It is important to recall that slightly higher levels of IGHM gene expression would have been present in this IgM^- without the use of a 15:1 competitor/probe ratio for that gene. Furthermore, we could argue that since the RT-MLPA classifier was built using an EBV^- DLBCL training series, this would no longer valid for valid

for classifying EBV-driven DLBCLs. However, to date, EBV$^+$ and EBV$^+$ DLBCLs are not officially considered as different entities, nor the EBV-status was mentioned as exclusion criteria in the original publication [22]. Overall, these results suggest that caution is required when classifying DLBCL lymphomas as the presence of EBV greatly impacts RT-MLPA-based DLBCL classification, and highlights the importance of identifying whether the lymphoma is EBV-driven.

Similarly, marked differences also existed in diagnostic and prognostically relevant genes. Gene expression levels of BCL2 reached statistical significance when EBV$^-$ and EBV$^+$ IgM$^+$ PT-DLBCL subgroups were compared to EBV$^+$ IgM$^-$ PT-DLBCL, being in the latter much higher. This is in line with few reports in the literature that suggest that EBV-driven lymphomas present a poorer clinical and prognostic outcome than EBV-negative ones [35, 45]. When correlating BCL2 relative gene and protein expressions, these differences were no longer appreciable.

The most remarkable finding concerned FOXP1. Most of EBV$^+$ IgM$^-$ cases were consistently negative at the protein level in comparison to the strongly nuclear positivity in all tumoral cells observed in most samples of the EBV$^+$ IgM$^+$ group. IHC-MLPA correlations were completely reversed. FOXP1 gene expression levels where significantly higher in IgM$^-$ than in IgM$^+$ cases. Taking into account that not all mRNA expression finally gets translated into protein, plus the fact that tumour heterogeneity is evident in most tumour biopsies, it is possible that FOXP1 protein expression in the IgM$^-$ group undergoes strict regulation at a post-transcriptional level. Whereas, in the IgM$^+$ group, it is possible that a different variant (a) particular alternative splicing variant(s) of FOXP1 is expressed at the protein level than the one represented in the MLPA probe [67, 68]. Furthermore, FOXP1 is also expressed in normal reactive lymphocytes present in the microenvironment of the tumour. Expression in these subsets was not taken into account when evaluating the staining, however it is picked up by the MLPA.

The impact of the presence of EBV on RT-MLPA-based classification of DLBCL in this study supports the concept that non-GCB EBV$^+$ and EBV$^-$ are biologically different entities. Furthermore, within the group of EBV$^+$ DLBCL, at least two distinct subtypes seem to exist. Apart from the importance for diagnosis, the different biology of these tumours allows patient stratification and the use of subtype-specific targeted therapies. Targeting of B-cell receptor (ibrutinib), mTORC1 (rapamycin) and NFKB signaling (bortezomib) are currently, extensively researched strategies for the treatment of ABC DLBCL. In this study LCL were used to determine the potential utility of these compounds in EBV$^+$ DLBCL. Since all LCL were IgM-negative and harboured somatically hypermutated immunoglobulin genes (unreported findings), they represent mutated IgM$^-$ EBV$^+$ PT-DLBCL. It remains to be investigated whether LCL mimicking unmutated IgM$^+$ EBV$^+$ PT-DLBCL LCL can be established. All LCL responded (to some extent) to rapamycin and in particular to ibrutinib monotherapy. In contrast, two LCL were resistant to bortezomib, potentially reflecting variable treatment response for different patients. As in DLBCL in the general population, bortezomib may have to be combined with chemotherapy to achieve a treatment effect [69]. As gene expression profiling demonstrated upregulation of unfolded response signalling, a pathway targeted with bortezomib, in IgM$^+$ compared to IgM$^-$ EBV$^+$ PT-DLBCL (manuscript submitted) we speculate that IgM$^+$ tumours will be more sensitive to bortezomib than IgM-negative tumours.

Strikingly, combination of two compounds significantly increased cytotoxicity in all LCL suggesting these are interesting approaches for management of EBV+ DLBCL. Notably, also in EBV– ABC DLBCL, ibrutinib and rapamycin was identified as a potent combination therapy [70].

CONCLUSION & SYNTHESIS

DLBCL is an aggressive type of lymphoma, that although its prognosis has improved significantly over the past decade, predicting patient prognosis still remains devious. This is can be, at least partially, explained by its tumour heterogeneity with no clear histologic criteria for subdivision. Gene profiling expression methods have demonstrated their utility in predicting patient survival, however this is not practical for the analysis of routine patient samples. To address this issue, efforts have been made to translate these data into using simpler and more globally accessible techniques.

In this project we optimised the previously described GEP-based technique RT-MLPA assay, that allows for an accurate classification of GCB and ABC DLBCLs samples based on the relative expression of fourteen genes in a single reaction. In fact, this method not only classifies but predicts the probability to belong to one group or the other. The method proved to be robust, reproducible and sensitive, generating stable signals with amounts of cDNA as low as 15,6 ng. Since only short cDNA fragments are necessary for the generation of MLPA signals, testing in highly RNA-degraded samples such as FFPE biopsies was possible. In fact, matching gene expression profiles to their FF counterparts were achieved in some cases. However, not all FFPE samples performed as well; suggesting that tissue processing and storage periods might influence their performance on MLPA. Overall, the method classified 70-80% of the samples according to their previously determined molecular phenotype. These results are in line with some IHC-based classification methods, and slightly lower than the gold standard GEP methods. Overall, this method could challenge IHC and other quantitative GEP methods to allow the stratification of patients who could ultimately benefit from prospective clinical trials and more aggressive and targeted therapies.

However, when applied to a series of EBV-driven DLBCL cases, the success rate dropped to 37%. This indicates that carefulness is required when classifying DLBCL lymphomas as EBV-positivity in tumoral cells most probably influences this RT-MLPA based classification and highlights the importance of EBV-status at time of diagnosis. To some end, this even extends to the recently identified IgM^+ and IgM^- DLBCL subgroups found within EBV^+ PT-DLBCLs. In that regard, as a suggested direction for the future research, we could try to identify whether IgM^- EBV^+ DLBCL subgroups are also the case in other populations, such as primary CNS lymphoma patients, DLBCL of the elderly and autoimmune diseased patients under treatment with immunomodulatory agents, for which preliminary data seem to indicate that this might be the case.

There are some limitations inherent in these studies: For the first case series only IHC-based classifications were available compared to the more reliable GEP-microarray standards. Moreover, correlations between IHC and GEP methods are not that straightforward. Not only the sampling comes from different locations, but also non-tumoral microenvironment is present in the sample. As a consequence, to date, both methods complement each other.

In summary, the results presented in these studies so far, are not decisive, and therefore validation work on this area is needed to provide a clearer insight on the matter. With this we hope to clearly understand the pathogenesis of DLBCL in an effort to discover novel, more targeted therapeutic.

REFERENCES

1. Fisher SG, Fisher RI: The epidemiology of non-Hodgkin's lymphoma. Oncogene 2004, 23(38):6524-6534.

2. Campo E, Swerdlow SH, Harris NL, Pileri S, Stein H, Jaffe ES: The 2008 WHO classification of lymphoid neoplasms and beyond: evolving concepts and practical applications. Blood 2011, 117(19):5019-5032.

3. Rossi D, Gaidano G: Molecular heterogeneity of diffuse large B-cell lymphoma: implications for disease management and prognosis. Hematology 2002, 7(4):239-252.

4. Coiffier B, Thieblemont C, Van Den Neste E, Lepeu G, Plantier I, Castaigne S, Lefort S, Marit G, Macro M, Sebban C et al: Long-term outcome of patients in the LNH-98.5 trial, the first randomized study comparing rituximab-CHOP to standard CHOP chemotherapy in DLBCL patients: a study by the Groupe d'Etudes des Lymphomes de l'Adulte. Blood 2010, 116(12):2040-2045.

5. Coiffier B: Diffuse large cell lymphoma. Curr Opin Oncol 2001, 13(5):325-334.

6. Hans CP, Weisenburger DD, Greiner TC, Gascoyne RD, Delabie J, Ott G, Muller-Hermelink HK, Campo E, Braziel RM, Jaffe ES et al: Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. Blood 2004, 103(1):275-282.

7. Barton S, Hawkes EA, Wotherspoon A, Cunningham D: Are we ready to stratify treatment for diffuse large B-cell lymphoma using molecular hallmarks? Oncologist 2012, 17(12):1562-1573.

8. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X et al: Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 2000, 403(6769):503-511.

9. Sehn LH, Gascoyne RD: Diffuse large B-cell lymphoma: optimizing outcome in the context of clinical and biologic heterogeneity. Blood 2015, 125(1):22-32.

10. Gifford GK, Gill AJ, Stevenson WS: Molecular subtyping of diffuse large B-cell lymphoma: update on biology, diagnosis and emerging platforms for practising pathologists. Pathology 2016, 48(1):5-16.

11. Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, Fisher RI, Gascoyne RD, Muller-Hermelink HK, Smeland EB, Giltnane JM et al: The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. N Engl J Med 2002, 346(25):1937-1947.

12. Lenz G, Wright G, Dave SS, Xiao W, Powell J, Zhao H, Xu W, Tan B, Goldschmidt N, Iqbal J et al: Stromal gene signatures in large-B-cell lymphomas. N Engl J Med 2008, 359(22):2313-2323.

13. Lenz G, Wright GW, Emre NC, Kohlhammer H, Dave SS, Davis RE, Carty S, Lam LT, Shaffer AL, Xiao W et al: Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways. Proc Natl Acad Sci U S A 2008, 105(36):13520-13525.

14. Monti S, Savage KJ, Kutok JL, Feuerhake F, Kurtin P, Mihm M, Wu B, Pasqualucci L, Neuberg D, Aguiar RC et al: Molecular profiling of diffuse large B-cell lymphoma identifies robust subtypes including one characterized by host inflammatory response. Blood 2005, 105(5):1851-1861.

15. Natkunam Y, Zhao S, Mason DY, Chen J, Taidi B, Jones M, Hammer AS, Hamilton Dutoit S, Lossos IS, Levy R: The oncoprotein LMO2 is expressed in normal germinal-center B cells and in human B-cell lymphomas. Blood 2007, 109(4):1636-1642.

16. Muris JJ, Meijer CJ, Vos W, van Krieken JH, Jiwa NM, Ossenkoppele GJ, Oudejans JJ: Immunohistochemical profiling based on Bcl-2, CD10 and MUM1 expression improves risk stratification in patients with primary nodal diffuse large B cell lymphoma. The Journal of pathology 2006, 208(5):714-723.

17. Choi WW, Weisenburger DD, Greiner TC, Piris MA, Banham AH, Delabie J, Braziel RM, Geng H, Iqbal J, Lenz G et al: A new immunostain algorithm classifies diffuse large B-cell lymphoma into molecular subtypes with high accuracy. Clin Cancer Res 2009, 15(17):5494-5502.

18. Coutinho R, Clear AJ, Owen A, Wilson A, Matthews J, Lee A, Alvarez R, Gomes da Silva M, Cabecadas J, Calaminici M 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(24):6686-6695.

19. Meyer PN, Fu K, Greiner TC, Smith LM, Delabie J, Gascoyne RD, Ott G, Rosenwald A, Braziel RM, Campo E et al: Immunohistochemical Methods for Predicting Cell of Origin and Survival in Patients With Diffuse Large B-Cell Lymphoma Treated With Rituximab. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 2011, 29(2):200-207.

20. Ott MM, Horn H, Kaufmann M, Ott G: The Hans classificator does not predict outcome in diffuse large B cell lymphoma in a large multicenter retrospective analysis of R-CHOP treated patients. Leukemia research 2012, 36(5):544-545.

21. Gutierrez-Garcia G, Cardesa-Salzmann T, Climent F, Gonzalez-Barca E, Mercadal S, Mate JL, Sancho JM, Arenillas L, Serrano S, Escoda L 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(18):4836-4843.

22. Mareschal S, Ruminy P, Bagacean C, Marchand V, Cornic M, Jais JP, Figeac M, Picquenot JM, Molina TJ, Fest T et al: Accurate Classification of Germinal Center B-Cell-Like/Activated B-Cell-Like Diffuse Large B-Cell Lymphoma Using a Simple and Rapid Reverse Transcriptase-Multiplex Ligation-Dependent Probe Amplification Assay: A CALYM Study. J Mol Diagn 2015.

23. Alizadeh AA, Gentles AJ, Alencar AJ, Liu CL, Kohrt HE, Houot R, Goldstein MJ, Zhao S, Natkunam Y, Advani RH et al: Prediction of survival in diffuse large B-cell lymphoma based on the expression of 2 genes reflecting tumor and microenvironment. Blood 2011, 118(5):1350-1358.

24. Lossos IS, Czerwinski DK, Alizadeh AA, Wechser MA, Tibshirani R, Botstein D, Levy R: Prediction of survival in diffuse large-B-cell lymphoma based on the expression of six genes. N Engl J Med 2004, 350(18):1828-1837.

25. Barrans SL, Crouch S, Care MA, Worrillow L, Smith A, Patmore R, Westhead DR, Tooze R, Roman E, Jack AS: Whole genome expression profiling based on paraffin embedded tissue can be used to classify diffuse large B-cell lymphoma and predict clinical outcome. British journal of haematology 2012, 159(4):441-453.

26. Rimsza LM, Wright G, Schwartz M, Chan WC, Jaffe ES, Gascoyne RD, Campo E, Rosenwald A, Ott G, Cook JR 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(11):3727-3732.

27. Masque-Soler N, Szczepanowski M, Kohler CW, Spang R, Klapper W: Molecular classification of mature aggressive B-cell lymphoma using digital multiplexed gene expression on formalin-fixed paraffin-embedded biopsy specimens. Blood 2013, 122(11):1985-1986.

28. Scott DW, Wright GW, Williams PM, Lih CJ, Walsh W, Jaffe ES, Rosenwald A, Campo E, Chan WC, Connors JM 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(8):1214-1217.

29. Hogfeldt T, Bahnassy AA, Kwiecinska A, Osterborg A, Tamm KP, Porwit A, Zekri AR, Lundahl J, Khaled HM, Mellstedt H et al: Patients with activated B-cell like diffuse large B-cell lymphoma in high and low infectious disease areas have different inflammatory gene signatures. Leuk Lymphoma 2013, 54(5):996-1003.

30. Nourse JP, Jones K, Gandhi MK: Epstein-Barr Virus-related post-transplant lymphoproliferative disorders: pathogenetic insights for targeted therapy. American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons 2011, 11(5):888-895.

31. Cesarman E: Gammaherpesviruses and lymphoproliferative disorders. Annu Rev Pathol 2014, 9:349-372.

32. Cesarman E: Gammaherpesvirus and lymphoproliferative disorders in immunocompromised patients. Cancer Lett 2011, 305(2):163-174.

33. Martelli M, Ferreri AJ, Agostinelli C, Di Rocco A, Pfreundschuh M, Pileri SA: Diffuse large B-cell lymphoma. Critical reviews in oncology/hematology 2013, 87(2):146-171.

34. Thorley-Lawson DA: Epstein-Barr virus: exploiting the immune system. Nat Rev Immunol 2001, 1(1):75-82.

35. Morscio J, Dierickx D, Tousseyn T: Molecular pathogenesis of B-cell posttransplant lymphoproliferative disorder: what do we know so far? Clin Dev Immunol 2013, 2013:150835.

36. da Silva SR, de Oliveira DE: HIV, EBV and KSHV: viral cooperation in the pathogenesis of human malignancies. Cancer Lett 2011, 305(2):175-185.

37. de la Cruz-Merino L, Lejeune M, Nogales Fernandez E, Henao Carrasco F, Grueso Lopez A, Illescas Vacas A, Pulla MP, Callau C, Alvaro T: Role of immune escape mechanisms in Hodgkin's lymphoma development and progression: a whole new world with therapeutic implications. Clin Dev Immunol 2012, 2012:756353.

38. Thorley-Lawson DA: EBV the prototypical human tumor virus--just how bad is it? J Allergy Clin Immunol 2005, 116(2):251-261; quiz 262.

39. Oyama T, Ichimura K, Suzuki R, Suzumiya J, Ohshima K, Yatabe Y, Yokoi T, Kojima M, Kamiya Y, Taji H et al: Senile EBV+ B-cell lymphoproliferative disorders: a clinicopathologic study of 22 patients. The American journal of surgical pathology 2003, 27(1):16-26.

40. Oyama T, Yamamoto K, Asano N, Oshiro A, Suzuki R, Kagami Y, Morishima Y, Takeuchi K, Izumo T, Mori S et al: Age-related EBV-associated B-cell lymphoproliferative disorders constitute a distinct clinicopathologic group: a study of 96 patients. Clin Cancer Res 2007, 13(17):5124-5132.

41. Beltran BE, Morales D, Quinones P, Medeiros LJ, Miranda RN, Castillo JJ: EBV-positive diffuse large b-cell lymphoma in young immunocompetent individuals. Clin Lymphoma Myeloma Leuk 2011, 11(6):512-516.

42. Cohen M, De Matteo E, Narbaitz M, Carreno FA, Preciado MV, Chabay PA: Epstein-Barr virus presence in pediatric diffuse large B-cell lymphoma reveals a particular association and latency patterns: analysis of viral role in tumor microenvironment. Int J Cancer 2013, 132(7):1572-1580.

43. Cohen M, Narbaitz M, Metrebian F, De Matteo E, Preciado MV, Chabay PA: Epstein-Barr virus-positive diffuse large B-cell lymphoma association is not only restricted to elderly patients. Int J Cancer 2014, 135(12):2816-2824.

44. Uccini S, Al-Jadiry MF, Scarpino S, Ferraro D, Alsaadawi AR, Al-Darraji AF, Moleti ML, Testi AM, Al-Hadad SA, Ruco L: Epstein-Barr virus-positive diffuse large B-cell lymphoma in children: a disease reminiscent of Epstein-Barr virus-positive diffuse large B-cell lymphoma of the elderly. Human pathology 2015, 46(5):716-724.

45. Morscio J, Dierickx D, Ferreiro JF, Herreman A, Van Loo P, Bittoun E, Verhoef G, Matthys P, Cools J, Wlodarska I et al: Gene expression profiling reveals clear differences between EBV-positive and EBV-negative posttransplant lymphoproliferative disorders. American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons 2013, 13(5):1305-1316.

46. Ferreiro JF, Morscio J, Dierickx D, Vandenberghe P, Gheysens O, Verhoef G, Zamani M, Tousseyn T, Wlodarska I: EBV-Positive and EBV-Negative Posttransplant Diffuse Large B Cell Lymphomas Have Distinct Genomic and Transcriptomic Features. Am J Transplant 2016, 16(2):414-425.

47. Capello D, Cerri M, Muti G, Berra E, Oreste P, Deambrogi C, Rossi D, Dotti G, Conconi A, Vigano M et al: Molecular histogenesis of posttransplantation lymphoproliferative disorders. Blood 2003, 102(10):3775-3785.

48. Brauninger A, Spieker T, Mottok A, Baur AS, Kuppers R, Hansmann ML: Epstein-Barr virus (EBV)-positive lymphoproliferations in post-transplant patients show immunoglobulin V gene mutation patterns suggesting interference of EBV with normal B cell differentiation processes. European journal of immunology 2003, 33(6):1593-1602.

49. Leblond V, Davi F, Charlotte F, Dorent R, Bitker MO, Sutton L, Gandjbakhch I, Binet JL, Raphael M: Posttransplant lymphoproliferative disorders not associated with Epstein-Barr virus: a distinct entity? Journal of clinical oncology : official journal of the American Society of Clinical Oncology 1998, 16(6):2052-2059.

50. Shaffer AL, 3rd, Young RM, Staudt LM: Pathogenesis of human B cell lymphomas. Annu Rev Immunol 2012, 30:565-610.

51. Dunleavy K, Pittaluga S, Czuczman MS, Dave SS, Wright G, Grant N, Shovlin M, Jaffe ES, Janik JE, Staudt LM et al: Differential efficacy of bortezomib plus chemotherapy within molecular subtypes of diffuse large B-cell lymphoma. Blood 2009, 113(24):6069-6076.

52. Castillo JJ, Beltran BE, Miranda RN, Paydas S, Winer ES, Butera JN: Epstein-barr virus-positive diffuse large B-cell lymphoma of the elderly: what we know so far. Oncologist 2011, 16(1):87-96.

53. Vaysberg M, Balatoni CE, Nepomuceno RR, Krams SM, Martinez OM: Rapamycin inhibits proliferation of Epstein-Barr virus-positive B-cell lymphomas through modulation of cell-cycle protein expression. Transplantation 2007, 83(8):1114-1121.

54. Zou P, Kawada J, Pesnicak L, Cohen JI: Bortezomib induces apoptosis of Epstein-Barr virus (EBV)-transformed B cells and prolongs survival of mice inoculated with EBV-transformed B cells. Journal of virology 2007, 81(18):10029-10036.

55. Petrara MR, Giunco S, Serraino D, Dolcetti R, De Rossi A: Post-transplant lymphoproliferative disorders: from epidemiology to pathogenesis-driven treatment. Cancer Lett 2015, 369(1):37-44.

56. RNA Integrity Number (RIN), Agilent Technologies: https://www.agilent.com/cs/library/applications/5989-1165EN.pdf [Consulted June 2016]

57. Eldering E, Spek CA, Aberson HL, Grummels A, Derks IA, de Vos AF, McElgunn CJ, Schouten JP: Expression profiling via novel multiplex assay allows rapid assessment of gene regulation in defined signalling pathways. Nucleic Acids Res 2003, 31(23):e153.

58. Designing Synthetic Probes: https://www.mlpa.com [Last consulted June 2016]

59. Wright G, Tan B, Rosenwald A, Hurt EH, Wiestner A, Staudt LM: A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. Proc Natl Acad Sci U S A 2003, 100(17):9991-9996.

60. Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM, Bloomfield CD, Cazzola M, Vardiman JW: The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood 2016, 127(20):2391-2405.

61. Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G: Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Res 2002, 30(12):e57.

62. MLPA Procedure: https://www.mlpa.com [Last consulted June 2016]

63. Hedegaard J, Thorsen K, Lund MK, Hein AM, Hamilton-Dutoit SJ, Vang S, Nordentoft I, Birkenkamp-Demtroder K, Kruhoffer M, Hager H et al: Next-generation sequencing of RNA and DNA isolated from paired fresh-frozen and formalin-fixed paraffin-embedded samples of human cancer and normal tissue. PloS one 2014, 9(5):e98187.

64. Kokkat TJ, Patel MS, McGarvey D, LiVolsi VA, Baloch ZW: Archived formalin-fixed paraffin-embedded (FFPE) blocks: A valuable underexploited resource for extraction of DNA, RNA, and protein. Biopreserv Biobank 2013, 11(2):101-106.

65. Rimsza LM, Wright G, Schwartz M, Chan WC, Jaffe ES, Gascoyne RD, Campo E, Rosenwald A, Ott G, Cook JR 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(11):3727-3732.

66. Shams TM: High expression of LMO2 in Hodgkin, Burkitt and germinal center diffuse large B cell lymphomas. J Egypt Natl Canc Inst 2011, 23(4):147-153.

67. Rouhigharabaei L, Finalet Ferreiro J, Tousseyn T, van der Krogt JA, Put N, Haralambieva E, Graux C, Maes B, Vicente C, Vandenberghe P et al: Non-IG aberrations of FOXP1 in B-cell malignancies lead to an aberrant expression of N-truncated isoforms of FOXP1. PloS one 2014, 9(1):e85851.

68. Dekker JD, Park D, Shaffer AL, 3rd, Kohlhammer H, Deng W, Lee BK, Ippolito GC, Georgiou G, Iyer VR, Staudt LM et al: Subtype-specific addiction of the activated B-cell subset of diffuse large B-cell lymphoma to FOXP1. Proc Natl Acad Sci U S A 2016, 113(5):E577-586.

69. Dunleavy K, Grant C, Wilson WH: Using biologic predictive factors to direct therapy of diffuse large B-cell lymphoma. Therapeutic advances in hematology 2013, 4(1):43-57.

70. Mathews Griner LA, Guha R, Shinn P, Young RM, Keller JM, Liu D, Goldlust IS, Yasgar A, McKnight C, Boxer MB et al: High-throughput combinatorial screening identifies drugs that cooperate with ibrutinib to kill activated B-cell-like diffuse large B-cell lymphoma cells. Proc Natl Acad Sci U S A 2014, 111(6):2349-2354.

SUPPLEMENTARY MATERIAL AND METHODS

SUPPLEMENTARY TABLES

Table S.1. Antibodies and staining conditions

Antibody	Clone	Source	Antigen Retrieval	Dilution
CD10	56C6	Dako	High	Ready to use
CD20	L26	Dako	High	Ready to use
MUM1	MUM1p	Dako	High	Ready to use
IGHM	R1/69	Dako	High	1/500
BCL6	PG-B6p	Dako	High	Ready to use
FOXP1	Ab16645	Dako	Low	1/200
LMO2	Ab1491-1	Abcam	High	—
BCL2	124	Abcam	High	Ready to use
MYBL1	ARP32707_T100	Dako	Low	1/100
MYC	Y69	Epitomics, Burlingame, CA, USA	High	1/100
ZEBRA	BZ1	Santa Cruz	Low	1/50

Oligonucleotide	Sequence	Oligonucleotide	Sequence
5′ RT-MLPA probes		3′ RT-MLPA probes	
NEK6_E2L	5′-GTGCCAGCAAGATCCAATCTAGACCTGTGCATCCTCCTGACCCACAG-3′	NEK6_E3R	5′-Pho-AGGCATCCCAACACGCTGTCTTTTCCAACCCTTAGGGAACCC-3′
IRF4_E6L	5′-GTGCCAGCAAGATCCAATCTAGATCTGCCGAAGCCTTGGCGTTCTCAG-3′	IRF4_E7R	5′-Pho-ACTGCCGGCTGCACATCTGCCTGTATCCAACCCTTAGGGAACCC-3′
IGHM_E2L	5′-GTGCCAGCAAGATCCAATCTAGATGCGTCCTCCATGTGTGGCCCCG-3′	IGHM_E3R	5′-Pho-ATCAAGACACAGCCATCCGGGTCTTCTACTATCCAACCCTTAGGGAACCC-3′
CCND1_E3L	5′-GTGCCAGCAAGATCCAATCTAGATACCTTCGTTGCCCTCTGTGCCACAG-3′	CCND1_E4R	5′-Pho-ATGTGAAGTTCATTTCCAATCCGCCCTTACTTCCAACCCTTAGGGAACCC-3′
LMO2_E5L	5′-GTGCCAGCAAGATCCAATCTAGACGGAAGCTCTGCCGGAGAGACTATCTCAG-3′	LMO2_E6R	5′-Pho-GCTTTTTGGGCAAGACGGTCTCTGCTACTATCCAACCCTTAGGGAACCC-3′
FOXP1_E10L	5′-GTGCCAGCAAGATCCAATCTAGACCCTTCCCCTTCAACCTCTTGCTCAAG-3′	FOXP1_E11R	5′-Pho-GCATGATTCCAACAGAACTGCAGCAGCTACTACTACTCCAACCCTTAGGGAACCC-3′
TNFRSF9_E2L	5′-GTGCCAGCAAGATCCAATCTAGATACGGACCTGTGACATATGCAGGCAGTGTAAAG-3′	TNFRSF9_E3R	5′-Pho-GTGTTTTCAGGACCAGGAAGGAGTGTTCCTACTCCAACCCTTAGGGAACCC-3′
BCL6_E3L	5′-GTGCCAGCAAGATCCAATCTAGATACTACTCATAAAACGGTCCTCATGGCCTGCAG-3′	BCL6_E4R	5′-Pho-TGGCCTGTTCTATAGCATCTTTACAGACCAGTTGTCCAACCCTTAGGGAACCC-3′
TNFRSF13B_E2L	5′-GTGCCAGCAAGATCCAATCTAGATACTACTACTAGCGCACCTGTGCAGCCTTCTGCA-3′	TNFRSF13B_E3R	5′-Pho-GGTCACTCAGCTGCCGCAAGGAGCTACTACTACTACTCCAACCCTTAGGGAACCC-3′
CCND2_E1L	5′-GTGCCAGCAAGATCCAATCTAGATACTACTGACCTTCATTGCTCTGTGTGCCACCG-3′	CCND2_E2R	5′-Pho-ACTTTAAGTTTGCCATGTACCCACCGTCGATACTACTATCCAACCCTTAGGGAACCC-3′
MYC_E1L	5′-GTGCCAGCAAGATCCAATCTAGATACTACTACTTCGGGTAGTGGAAAACCAGCAGCCTC-3′	MYC_E2R	5′-Pho-CCGCGACGATGCCCCTCAACGTTATACTACTACTACTATCCAACCCTTAGGGAACCC-3′
MYBL1_E10L	5′-GTGCCAGCAAGATCCAATCTAGACCAGAATTTGCAGAGACTCTAGAACTTATTGAATCT-3′	MYBL1_E11R	5′-Pho-GATCCTGTAGCATGGAGTGACGTTACCAGTTTTTACTACTTCCAACCCTTAGGGAACCC-3′
BCL2_E1L	5′-GTGCCAGCAAGATCCAATCTAGATACTACTACTACCCTGGATCCAGGATAACGGAGGCTGG-3′	BCL2_E2R	5′-Pho-GATGCCTTTGTGGAACTGTACGGCCTACTACTACTACTACTTCCAACCCTTAGGGAACCC-3′
MS4A1_E5F*	5′-GTGCCAGCAAGATCCAATCTAGATACTACTACTATTCTTCATGAGGGAATCTAAGACTTTGGGG-3′	MS4A1_E6R	5′-Pho-GCTGTCCAGATTATGAATGGGCTCTTCCACTACTACTACTATCCAACCCTTAGGGAACCC-3′
		IGHM_comp	5′-ATCAAGACACAGCCATCCGGGTCTTC-3′

Table S.2. Sequences of the RT-MLPA oligonucleotides, All 5′ MLPA probes had an identical PCR-primer binding tail sequence (GTGCCAGC-AAGATCCAATCTAGA) at the 5′ end, complementary to the universal primer U1; whereas all 3′ MLPA probes had a different common PCR-primer binding tail (TCCAACCCTTAGGGAACCC), located at the 3′end and complementary to the universal primer U2. The latter had phosphorylated 5′ ends to allow for the ligation reactions.

SUPPLEMENTARY FIGURES

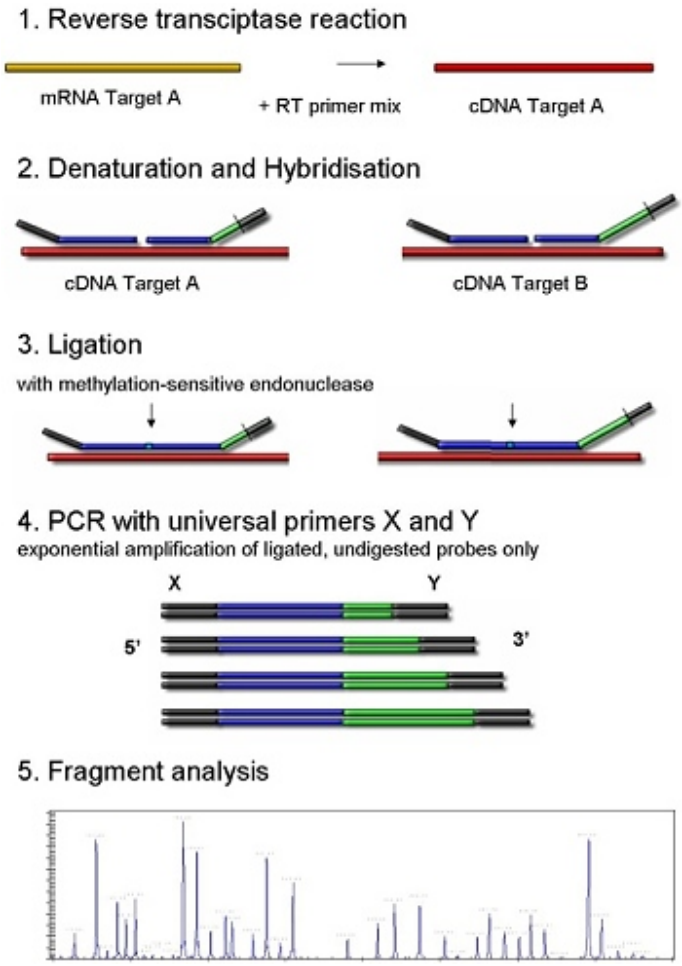


Figure S.1. Schematic representation of the mRNA MLPA – RT-MLPA technique. The RT-MLPA reaction comprises four main steps. (1). Reverse transcriptase reaction that transcribes the mRNA target into cDNA. (2) Denaturation and Hybridisation of each probes with its corresponding cDNA target. (3) Ligation. It is only when the two probe oligonucleotides are both hybridised to their target that they can be ligated into a single probe, containing both the forward and reverse primer sequence. (4) While these ligated probes are amplified exponentially during the PCR reaction, the individual non-ligated probe oligonucleotides are not. The number of probe ligation products of one probe therefore depends on the number of target sequences in the sample. (5). Detection and fragment analysis. The resulting amplification products are separated by capillary electrophoresis. An MLPA probe set is designed so that the length of each of its amplification products is unique.
Source from MLPA website [62].

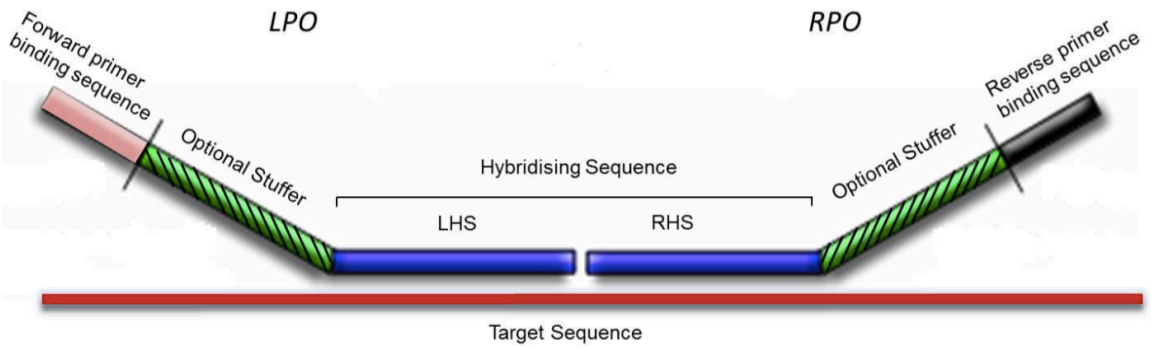


Figure S.2. Terminology of probe components. Each MLPA probe consists of a gene-specific region complementary to the cDNA target (the hybridisation sequence, HS), a PCR-primer binding tail and a TAC-repeat stuffer that is inserted between the PCR-primer binding tail and the gene-specific HS, which allows for the separation and identification of the PCR product according to its size. *Source from MLPA website [62].*

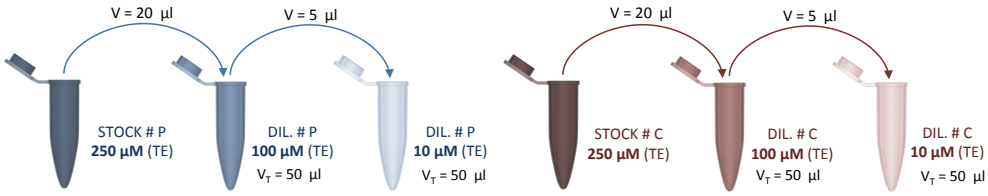
SUPPLEMENTARY ANNEXES

ANNEX 1. OVERVIEW OF THE MLPA PROBE MIX AND WORKFLOW

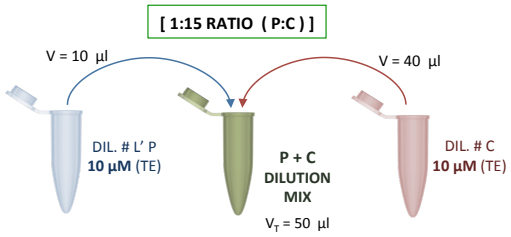
MLPA PROBE MIX

- There are 14 probes (P) (5′ and 3′ → **28 probes**) + **1 competitors** (CO)
- For each P and C prepare two stock dilutions from 250 µM
- Prepare further dilutions: 100 µM and 10 µM

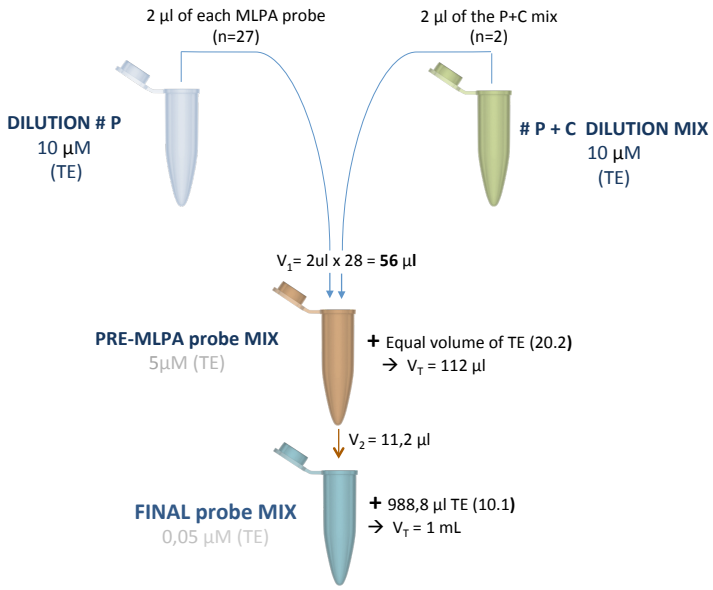
EXAMPLE:



- Mix the competitor with its corresponding R′ probe in a 15:1 ratio (C/P)

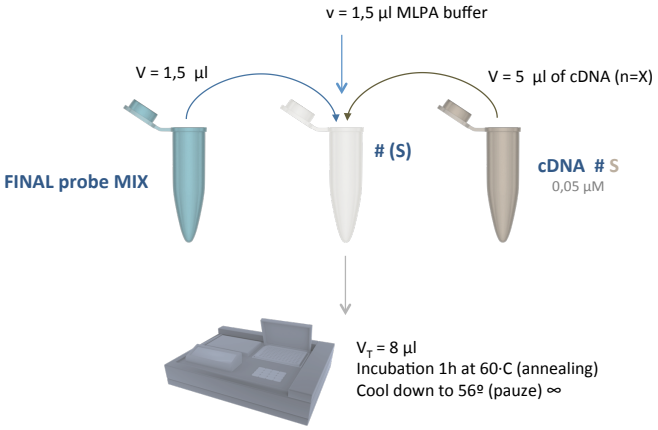


MLPA PROBE MIX



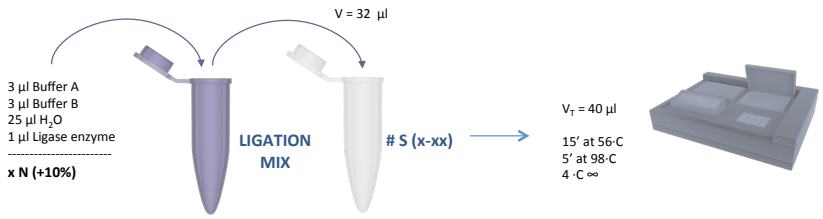
1. HYBRIDISATION

- Pipette in a new tube 5µl cDNA #S
- Denaturation of each cDNA tube #S for 2' at 98·C
- Then, for each sample (S):



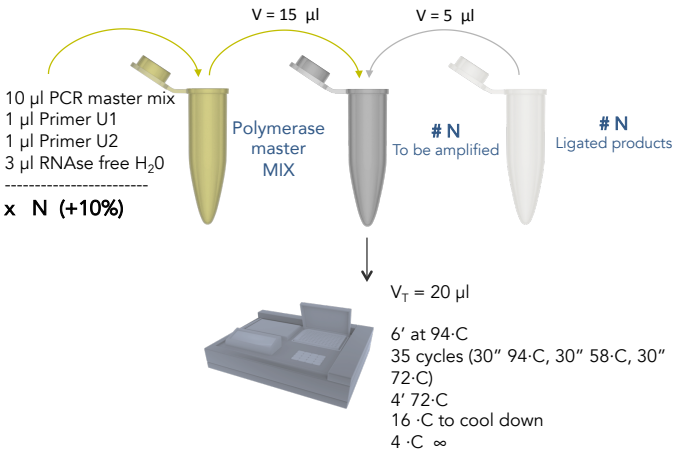
2. LIGATION

- Prepare a LIGATION MIX for all samples



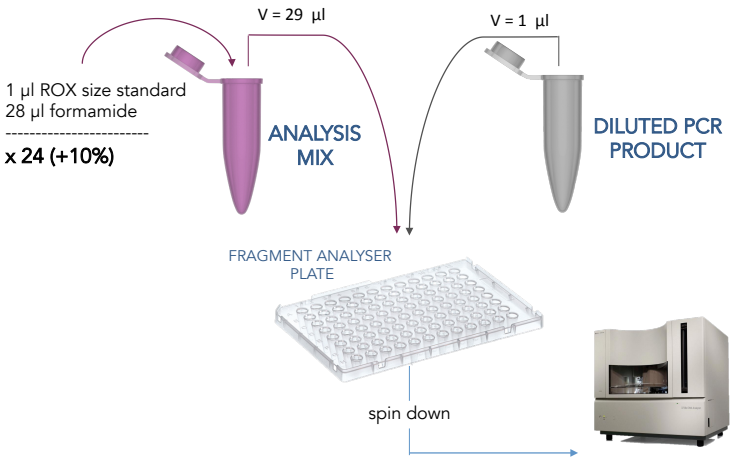
3. PCR AMPLIFICATION

- Prepare **PCR mix** for N samples



4. FRAGMENT ANALYSER

- DILUTIONS of the PCR products → 1/10 for FFPE and 1/20 FF
- Prepare an DETECTION MIX for all samples



ANNEX 2. ADAPTED PROCESS INTERFACE DESIGN FILE

```
[DESIGN]
# Design description
author    Sylvain Mareschal
purpose GCB-ABC classification model, as published in JMD (PMID: 25891505)
MLPA    1.9.0
updated 2015-07-03

[MLPA.process]
# Arguments to be passed to the R function, see the Reference Manual for further details

[read.fsa]
# Arguments to be passed to the R function, see the Reference Manual for further details

[align.fsa]
# Arguments to be passed to the R function, see the Reference Manual for further details
rescue    FALSE
channel  Dye4
fullLadder        35      50      75      100     139     150     160     200     250     300
          340     350     400     450     490     500
useLadder         50      75      100     139     150     160
outThreshold      0.6
maskOffScale      FALSE
noiseLevel        400
trim      backward

[peaks.fsa]
# Arguments to be passed to the R function, see the Reference Manual for further details
logTransform      TRUE
lowThreshold      400
channelsDye1

[plot.fsa]
# Arguments to be passed to the R function, see the Reference Manual for further details
channelsDye4      Dye1
chanColors        grey      black
xlim      75      125
ylim      NA


[model]
# Arguments to be passed to the R function, see the Reference Manual for further details
threshold         0.90
groupNames        ABC       GCB
groupColors       darkbluedarkorange2
groupMeans        10.21696-35.92091
```

groupSDs 7.51672 15.8074
geneNames LMO2 NEK6 MYBL1 TNFRSF13B BCL6 IRF4 IGHM FOXP1
geneTs -11.11728 -6.07617 -4.86899 4.4753 -3.59697 4.82121 4.51134 2.69062
geneMs 0.68433 0.38955 0.31225 0.46633 0.48262 1.48637 1.70233 1.352
[classify]
Arguments to be passed to the R function, see the Reference Manual for further details
plot TRUE

[PEAKS]
Intervals in which look for a gene-related signal

name	channel	size.min	size.max	color
NEK6	Dye1	84.44	85.44	darkorange2
IRF4	Dye1	88.02	89.32	darkblue
IGHM	Dye1	92.86	94.36	darkblue
CCND1	Dye1	95.11	96.61	darkgrey
LMO2	Dye1	96.85	98.35	darkorange2
FOXP1	Dye1	102.57	103.75	darkblue
TNFRSF9	Dye1	103.75	105.25	darkgrey
BCL6	Dye1	105.93	106.93	darkorange2
TNFRSF13B	Dye1	109.04	110.04	darkblue
CCND2	Dye1	110.11	111.11	darkgrey
MYC	Dye1	113.86	114.86	darkgrey
MYBL1	Dye1	114.87	116.47	darkorange2
BCL2	Dye1	118.18	119.18	darkgrey
MS4A1	Dye1	120.93	122.23	darkgreen

SUPPLEMENTARY DATA

SUPPLEMENTARY TABLES

Table S.1. Absolute and relative protein expression percentages for IC-DLBCL

				Stainings																											
		Tumor % in the sample	CD20	%	R%	CD10	MUM1	%	R%	BCL6	%	R%	LMO2	%	R%	FOXP1	%	R%	IGHM	%	R%	BCL2	%	R%	MYBL1	%	R%	MYC	%	R%	Remarks
EBV- IC-DLBCL (non-GCB)	Sample 1	80	+	100	80	-	+	50-60	44	+	30	24	C+	80	64	N+	100	80	+	80	64	+	100	80	+	10	8	-	20	16	
	Sample 2	90	+	100	90	-	+	> 80	72	+	60	54	C+	90	81	N+	100	90	+	100	90	+	100	90	-	1	1	+	60	54	
	Sample 3	80	+	100	80	-	+	90	72	+	80	64	C-	< 5	4	C-	<30	24	+	90	72	-	<5	4	+	15-20	14	-	< 10	8	
	Sample 4	50	+	100	50	-	-	0	0	-	0	0	C+	80	40	+	50	25	+	100	50	+	100	50	+	15-20	8				
	Sample 5	90	+	100	90	-	+	100	90	+	100	90	C+ > N+	60	54	+	100	90	-	0	0	+	100	90	C+	> 20	18	-	20	18	Part of the tissue not involved
	Sample 6	75	+	100	75	-	+	80	60	+	80	60	??	Most cells		N+	100	75	+	100	75	-	30-40	26	??	??		+	40	30	
EBV- IC-DLBCL (GCB)	Sample 7	90	+	100	90	-	-	0	0	+	100	90				N+	100	90	-	0	0	+	> 50	45	-	≅ 2	2	+	30-40	32	
	Sample 8	50	+	100	50	+	-	< 5	2,5	+	85	43	C+ > N+	100	50	N+	100	50	-	0	0	-	0	0				-	<30	15	
	Sample 9	90	+	100	90	+	-	≅ 1	1	+	50	45	N+ > C+	100	90	N+	90	90	-	0	0	-	20-30	22	C			-	<10	9	
	Sample 10	90	+	100	90	+	-	1	1	+	80	72	N+ > C+	100	90	N+	90	90	-	0	0	-	0	0	C			-	<10	9	
	Sample 11	90	+	100	90	+	+	30	27	+	30	27	C- > N-	??	90	N+	100	90	+	100	90	+	100	90	C+++	15	13,5	-	≅10	9	
	Sample 12	90	+	100	90	+	-	< 5	4,5	+	60	54	N ≅ C	100	90	N+	100	90	-	0	0	+	80	72	C+	15	13,5	+	≅50-60	48	

Table S.2 Absolute and relative protein expression percentages for PT-DLBCL

			Stainings																											
		Tumor % in the sample	CD20	%	R%	CD10	%	R%	MUM1	%	R%	BCL6	%	R%	LMO2	%	R%	FOXP1	%	R%	IGHM	%	R%	BCL2	%		MYC	%	R%	Remarks
EBV- PT-DLBCL	Sample 2-1	30	+	100	30	-	0	0	-	0	0	-	0	0	C+	100	30	NC+	90	27	+	90	27	-	< 10	3	+	30	9	
	Sample 2-2	80	+	100	80	-	0	0	-	0	0	-	<20	16	C++	90	72	-	0	0	+	50	40	-			-	<10	8	Strong LMO2 Cyto. / IgM is Mb, not clear
	Sample 2-3	20	+	100	20	+	70	14	-	< 5	1	+	50	10	?	?	?	N+C+	90	18	-	> 5%	1				-	0	0	
	Sample 2-4	80	+	100	80	-	0	0	-	≅ 20	16	+	80	64	C+	100	80	N+	90	72	+	100	80	-	?	?	-	0	0	
	Sample 2-5	90	+	100	90	-	0	0	+	> 80	72	-	0	0	C+	100	90	N+	100	90	+	100	90	+	100	90	+	40	36	
	Sample 2-6	80	+	100	80	-	0	0	-	< 30	24	+	50	40	C+	100	80	N+	100	90	+	100	80	-	0	0	-	< 20	16	MUM1 is very weak
	Sample 2-7	90	+	100	90	-			+			+			-			+						+						
EBV+ PT-DLBCL IgM-/IgD-	Sample 2-8	70	+	100	70	-	0	0	-	< 30	21	-	0	0	C+	100	70	-	<30	21	-	0	0	+	100	70	-	0	0	
	Sample 2-9	80	+	100	80	-	0	0	+	90	72	-	< 10	8	C+	100	80				-	0	0	+	100	80	+	60	48	LMO2 is weak but very specific
	Sample 2-10	50	+	100	50				-	0	0	-	0	0	C+	100	50	-	0	0	-	0	0				-	0	0	FOXP1 is very weak.
	Sample 2-11	90	+	100	90				+	60	54	-	0	0	C+	100	90	-	0	0	-	0	0	-	5	4,5	+	40	36	
	Sample 2-12	70	+	100	70				-	< 30	21				?	?	?	-	0	0	-	0	0	-	10	7	+	40	28	
	Sample 2-13	40	+	100	40							-	0	0	C+	100	40	-	0	0	-	0	0	+			-	0	0	
	Sample 2-14	70	+	100	70				+	50	35	-	0	0	C+	100	70	-	0	0	-	0	0				-	0	0	
	Sample 2-15	80	+	100	80	-	0	0	+	40	32	-	0	0	C+	90	80	+	50	40	-	0	0	-	30	24	-	20	16	LMO2 is weak. FOXP1 weak
	Sample 2-16	60	+	100	60	-	0	0	+	?	?	-	0	0	C+	100	60	-	100		-	0	0	-	100	60	-	20	12	MUM1 cytopl. 30-40% / FOXP1 100% cytopl.
EBV+ PT-DLBCL IgM+/IgD+	Sample 2-18	60	+	100	60	-	0	0	+	40	24	-	0	0	C+	100	60	+	100	60	+	50	30	+	100	60	-	25	15	
	Sample 2-18	90	+	100	90										C+	90	81													
	Sample 2-21	80	+	100	80	-	0	0	+	100	80	-	0	0	NR	NR	NR	+	NR	NR	NR	NR	NR	+	100	80	-	< 5	4	Very necrotic tissue. 80% viable cells are tumour
	Sample 2-22	90	+	100	90	-	0	0	+	100	90	-	0	0	C+	100	90	N+	100	90	+	100	90	+	100	90		10-50	9-45	
	Sample 2-23	80	+	100	80	-	0	0	-	0	0	-	0	0				C-	100	0				+	100	80	-	< 10	8	The report says that FOXP1 is positive. Thus repeat staining?
	Sample 2-24	80	+	100	80	-			+	100	80	-	0	0	C+	100	80	N+ C+	100	80	+	100	80	+	100	80	-	< 10	8	FOXP1 Nucl. And cytop.
	Sample 2-25	80	+	100	80	-	0	0	+	100	80	-	0	0	C+	100	80	C+	100	80	+	100	80	+	100	80				100% cytopl.
	Sample 2-26	80	+	100	80	-	0	0	+	100	80	-	0	0	C+	100	80	C+>N+	100	80	+	100	80	+	100	80	+	50	40	Not as clear as 2-22

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
Molecular Subtype Classification of Diffuse Large B-Cell Lymphoma (DLBCL)

Richting: **master in de biomedische wetenschappen-klinische moleculaire wetenschappen**
Jaar: **2016**

in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Voor akkoord,

Abengochea Molar, Aran

Datum: **8/06/2016**