Gene expression profiling and immunoglobulin stereotypy in Burkitt lymphoma

Dissertation
in fulfilment of the requirements
for the degree "Dr. rer. nat."
of the Faculty of Mathematics
and Natural Sciences
at Kiel University

submitted by
Neus Masqué Soler
Kiel, 2015

First referee:	Professor nat. rer. Holger Kalthoff
Second referee:	Professor nat. rer. Thomas Röder
Date of the oral examination:	14.12.2015
Approved for publication:	02.02.2016
Dean:	Professor Dr. Wolfgang J. Duschl

A l'avi Joan

Contents

List of Figures

List of Tables

List of Abbreviations

Ab	Antibody
ABC DLBCL	Activated B-cell diffuse large B-cell lymphoma
ABS	Antigen-binding sites
Ag	Antigen
AID	Activation-induced deaminase
ANOVA	Analysis of variance
B-AL	Burkitt-acute leukaemia
BCL2	B-cell lymphoma 2
BCL6	B-cell lymphoma 6
BcR	B-cell receptor
BL	Burkitt lymphoma
BM	Bone marrow
BSA	Bovine serum albumin
cDNA	Complementary DNA
CDR	Complementarity determining regions
CHOP	Cyclophosphamide, Hydroxydaunorubicin, Vincistrine, and Prednisolone
CLL	Chronic lymphocytic leukaemia
CNS	Central nervous system
COSY	Correlation spectroscopy
CSR	Class switch recombination
DA	Digital analyzer
DAPI	4',6-diamidino-2-phenylindole
DLBCL	Diffuse large B-cell lymphoma
DMGE	Digital multiplexed gene expression
DNA	Deoxyribonucleic acid
DPBS	Dulbecco's phosphate buffered saline
EBER	Epstein–Barr virus-encoded small RNAs
eBL	Endemic Burkitt lymphoma
EBV	Epstein-Barr Virus
ELISA	Enzyme-linked immunosorbent assay
EtOH	Ethanol
FBS	Foetal bovine serum
FFPE	Formalin-fixed paraffin-embedded
FISH	Fluorescence in situ hybridisation
FL	Follicular lymphoma
FOV	Fields of view
FR	Framework region
FU	Follow up
GC	Germinal centre
GCB DLBC	Germinal centre B-cell-like diffuse large B-cell lymphoma
GEP	Gene expression profiling
GI	Germline identity
HEK	Human embryonic kidney
HRP	Horseradish peroxidase

IG	Immunoglobulin
IGH@	Immunoglobulin heavy chain locus
IGHM	Immunoglobulin heavy chain mu locus
IGHV	Immunoglobulin heavy chain variable region
IGK@	Immunoglobulin light chain kappa locus
IGL@	Immunoglobulin light chain lambda locus
IHC	Immunohistochemistry
IHS	Immunohistochemical staining
LARS	Least-Angle Regression
LB	Lysogeny broth
LDH	Lactate dehydrogenase
LN	Lymph node
mAb	Monoclonal antibody
mBL	Molecular Burkitt lymphoma
MCL	Mantle cell lymphoma
MCS	Multiple cloning site
MMML	Molecular Mechanisms of Malignant Lymphoma
mRNA	Messenger RNA
NHL	Non-Hodgkin lymphoma
NMR	Nuclear magnetic resonance
NOESY	Nuclear Overhauser effect spectroscopy
OD	Optical density
ON	Overnight
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PS	PrepStation
PVDF	Polyvinylidene fluoride
qPCR	Quantitative PCR
RNA	Ribonucleic acid
RPMI medium	Roswell Park Memorial Institute medium
RT	Room temperature
sBL	Sporadic Burkitt lymphoma
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SHM	Somatic hypermutation
SMZL	Splenic marginal zone lymphoma
SpA	*Staphylococcus aureus* protein A
ssDNA	Single stranded DNA
STD	Saturation transfer difference
TBS	Tris-buffered saline
TdT	Terminal deoxynucleotidyl transferase
TMA	Tissue microarrays
TOCSY	Total correlation spectroscopy
TUP	Target-unrelated peptide
WHO	World Health Organisation

Abstract

The present work compilation discusses the creation of improved molecular lymphoma classifiers and studies the immunological properties of lymphoma antibodies.

Lymphatic system cells are grouped in B, T or natural killer cells according to their functions. The B cell population is responsible for humoral immunity by secreting antibodies and presenting antigen fragments in its surface. When malignantly transformed, B cells can develop into what is described as B cell lymphoma.

Burkitt lymphoma (BL) is a rare type of B cell lymphoma characterised by its hallmark chromosomal translocation [t(8;14)] involving the *c-Myc* oncogene and the immunoglobulin heavy chain locus (IGH@). Burkitt lymphoma shares some immunohistochemical (IHC), morphological and genetic characteristics with the most frequent lymphoma form: diffuse large B cell lymphoma (DLBCL). Both diseases are considered aggressive, presenting with severe symptoms and rapid growth. Therefore distinguishing within and between the two diagnostic entities can, at times, be challenging. In the past IHC tools were standardised to overcome the categorisation issue. With the wake of gene expression profile assays lymphoma classification was adapted to newer methods. The first chapter of this thesis presents two improved aggressive B cell lymphoma classifiers based on a novel technology that does not require material amplification, avoiding quantification bias and is able to count actual mRNA molecules to infer genetic expression levels. On one side the classifier distinguishes molecular BL (mBL) from non-BL and on the other the distinction is made between DLBCL subtypes. In chapter two the former classifier is used to understand molecular labelling differences between BCL2-expressing and non-expressing BL. The co-expression of MYC and BCL2 has been associated with poorer outcome in DLBCL. By assessing the molecular labelling of BCL2-positive and -negative BL we determined that an expression of the anti-apoptotic BCL2 protein is not enough to modify the labelling of mBL to another lymphoma subtype and detect no clinical differences between BCL2-expressing and non-expressing cases.

Chapters three and four focus on the humoral immunity aspect of lymphoma B cells and their B cell receptors (BcR). The BcR is a complex formed by a transmembranous immunoglobulin (IG) and the Ig-α/Ig-β heterodimer facing the cytoplasm. The IG molecule is in charge of antigen recognition through its variable domain. A wide range of IG variability is achieved by the so called V(D)J recombination, a procedure that combines gene segments in the IGH@. Lymphomas present a biased repertoire of *IG* gene

recombinations distinct from normal B cells, a phenomenon attributed to antigenic drive. In chapter three the *IG* rearrangement of B-cell lymphomas is reviewed, with a focus on BL. Three V(D)J segments of BL monoclonal antibodies (mAbs) belonging to one same biased subtype are characterised structurally *in silico* and studied for their mutation in the antigen-binding site. By reverting the mutation to its original germline amino acid the antigen-biding pocket is seen to vary in shape. Chapter four learns from these observations and performs experimental procedures with phage display and nuclear magnetic resonance techniques. Four peptides with variable binding affinities are discovered that bind to the BL mAbs previously characterised and are homologous to various proteins of alimentary and pathogenic origin.

Summarizing, this thesis presents an improved tool for BL and DLBCL classification, which can be used for assessing basic science topics as well as avoiding diagnostic pitfalls. The type of IG in tumoral BL cells and their antigen affinity is assessed, pointing to an antigenic aetiology.

Zusammenfassung

Die vorliegende Arbeit stellt eine Zusammenfassung von verbesserten molekularen Lymphom Klassifikatoren und deren Studien über immunologische Eigenschaften von Lymphom Antikörpern dar.

Die Zellen des lymphatischen Systems werden je nach Funktion in drei Gruppen eingeteilt, B-Zellen, T-Zellen oder NK-Zellen (Natürliche Killerzellen). Die B-Zell Population ist für die humorale Immunantwort verantwortlich. B-Zellen produzieren und sekretieren Antikörper und präsentieren nach Antigenkontakt auch Fragmente auf der Oberfläche. Bei einer bösartigen Entartung dieser Zellreihen entwickeln sich B-Zell-Lymphome.

Das Burkitt Lymphom (BL) ist eine seltene Form des B-Zell-Lymphoms, welches durch seine charakteristische Chromosomale Translokation [t(8;14)] mit Beteiligung des c-Myc Onkogens und des IGH Lokus (IGH@) typisiert ist. Es hat zusätzlich auch immunhistochemische (IHC), morphologische und genetische Eigenschaften wie andere, häufigere Lymphome, z. B. das diffuse large B cell lymphoma (DLBCL). Beide Erkrankungen sind sehr aggressiv und durch heftige, schlimme Symptome, sowie durch ein schnelles Wachstum gekennzeichnet. Deshalb ist es manchmal eine Herausforderung zwischen diesen beiden Entitäten zu unterscheiden. In der Vergangenheit wurden IHC Methoden und Antikörperauswahl standardisiert, um dadurch eine bessere Kategorisierung zu erreichen. Mit dem Beginn der Gen Expression Assays wurde die Lymphom Klassifizierung an neuere Methoden angepasst. Das erste Kapitel dieser Promotionsschrift zeigt zwei neue Klassifikatoren von aggressiven B-Zell-Lymphomen die durch die Anwendung einer neuartiger Technologie entwickelt wurden, die keine Vervielfältigung von Material benötigt, dadurch Amplifikationsprobleme vermeidet, in der Lage ist mRNA Moleküle zu zählen, und genaue Rückschlüsse auf das genetische Expressions zulässt. Der Klassifikator unterscheidet molekulare BL (mBL) von non-mBL und zeigt zusätzlich den Unterschied zu DLBCL Subtypen. Im zweiten Kapitel wird beschrieben, wie der Klassifikator verwendet wurde, um molekulare Kennzeichnungen zwischen BCL2 exprimierenden und nicht exprimierenden BL zu unterscheiden. Es ist bekannt, dass die Co-Expression von MYC und BCL2 in DLBCLs mit einer schlechteren Prognose verbunden ist. In unserer Studie sind keine klinischen Unterschiede zwischen den untersuchten BCL2 exprimierenden und nicht exprimierenden Fällen erfunden worden. Bei der Auswertung der molekularen Ergebnisse von BCL2 positiven und negativen BL konnten wir allerdings feststellen, dass die Expression des anti-apoptotischen BCL2

Proteins nicht ausreicht, um die Kennzeichnung eines mBL in einen anderen molekularen Subtypen zu verändern.

Die Schwerpunkte in Kapitel drei und vier beziehen sich auf die humorale Immunantwort bei B-Zell-Lymphomen und den B-Zell-Rezeptoren (BcR). Der BcR ist ein aus einem transmembranären Immunglobulin (IG) und den Ig-alpha/Ig-beta Heterodimer geformter Komplex. Das IG Molekül ist durch seine variable Domäne für die Antigenerkennung hauptverantwortlich. Eine große Reichweite der IG Variabilität wird durch die so genannten V(D)J Rekombination erreicht, ein Ablauf bei dem Gensegmente im IGH@ neu kombiniert werden. Lymphome stellen ein *biased* Repertoire der IG-Gen Rekombination dar, welche sich von normalen B-Zellen unterscheidet, ein Phänomen, welches auf *antigen drive* zurückzuführen ist. In Kapitel drei wurde das IG Rearrangement von B-Zell-Lymphomen, mit dem Schwerpunkt auf BL, untersucht. Immunoglobulin-Gen V(D)J Segmente von drei BL monoklonalen AK (mAK) mit bestimmte Subtypen wurden strukturell *in silico* charakterisiert und auf vorhandene Mutationen in der Antigen-Bindungsstelle geprüft. Beim Zurückleiten auf die ursprüngliche Keimbahnsequenz der Aminosäuren für die Antigen-Bindungstasche ergaben sich leicht veränderte Strukturen. In Kapitel vier werden die experimentellen Verfahren und Durchführungen mit *phage display* und *nuclear magnetic resonance* Technik beschrieben. Dabei wurden vier Peptide mit variablen Bindungsaffinitäten entdeckt, die sich spezifisch an die in Kapitel drei erwähnten BL mAK binden. Die Peptide sind homolog zu verschiedenen Proteinen des Verdauungstrakts und/oder pathogenen Ursprungs.

Zusammenfassend liefert diese Promotionsschrift neue Hilfsmittel für die differentielle BL/DLBCL Klassifikation, welche sowohl in der Grundlagenforschung als auch für die Vermeidung von diagnostischen Fehleinschätzungen genutzt werden kann. Der IG Typ und die entsprechende Affinität wurde in malignen BL Zellen bestimmt und weist auf eine Antigen-vermittelte Ätiologie hin.

General introduction

Cancer epidemiology in Europe

Cancer affects a considerable segment of the population. The latest large study for cancer incidence and mortality reported a total of 355.7 cases per 100 000 inhabitants in data from 2012 and a mortality rate of 1 754 600 in the same year, which accounts for more than a quarter (26.3%) of all deaths of 2011 in the EU-28 countries [1].

Malignant neoplasms show different standardised mortality rates according to gender, being those of men (370.3) significantly higher than those of women (207.1) [1]. Cancers are also of a varied nature and origin, with prostate and breast malignancies being the most frequent in men and women respectively [2], but also include colorectal, lung, lymphoid and hematopoietic cancers.

Lymphomas are classified as Hodgkin (17 600 new cases in Europe in 2012) and non-Hodgkin lymphomas (NHL). A total of 93 500 new NHL cases were diagnosed in Europe in 2012, making it the 11th most common cancer in the region [3]. Lymphomas are a group of cancers originating in the lymphatic system cells, the lymphocytes. Mainly involved in immunological functions, lymphocytes known to be of the B-, T- or natural killer cell type depending on where they mature.

Thus, B-cell lymphomas are malignant neoplasms where the tumoral cell component is of the B-cell lymphocyte population. They are named aggressive if their symptoms are severe and tumor growth is rapid in contrast to the low grade types.

Up to one third of childhood neoplasms are NHL [4]. Burkitt lymphoma (BL) makes up 40% of all paediatric NHL cases and accounts for 1-2% of all lymphomas in adults [5,6] in the western world [7]. In the adult population, however, the most frequent type of NHL is diffuse large B-cell lymphoma (DLBCL).

Burkitt lymphoma

Characteristics and typification

Burkitt lymphoma carries the name of its discoverer, who first described it in 1953 in African children, although Sir Albert Cook wrote about similar clinical traits in children 66 years earlier. The three main types of BL are the endemic (eBL), sporadic (sBL) and the immunodeficient-related variants. All three variants have relevant differences in its presentation and geographical prevalence. The eBL variant is geographically present in what is known as the malaria or equatorial belt and appears in the submandibular area of suffering patients, mostly children and young adults [8]. The sBL variant is located worldwide and appears mainly in the thorax region of children and young adults. The third

variant is explained by the immunodeficiency of its suffering patients [9]. Without treatment the lymphoma courses aggressively, having a tumour size doubling time of as little as 24 hours.

The microscopical morphology of BL is characterized by the starry sky pattern, with more than 90% of the tumoral cells being positive for the proliferation marker Ki67. In some immunohistochemical and clinical characteristics BL can resemble those of DLBCL. Hence, a clear line dividing both lymphoma entities is non-existent. A subgroup of cases with characteristics shared between BL and DLBCL is recognised by the WHO and names it intermediate subtype [10]. Within DLBCL there are three subgroups, namely the activated B-cell (ABC), the germinal centre B-cell-like (GCB), and the unclassifiable subtypes. They are prognostically relevant and their therapeutic regimens are distinctive [11].

Moreover, considerable amounts of archival biopsy material are available in many research and clinical centres. A need for adapting novel profiling techniques to paraffin-embedded biopsies was detected and assessed in our group. The first two chapters of the present thesis aims at improving BL and DLBCL classification while taking advantage of the wide sample collection available in our facilities.

B-cell receptors

Burkitt lymphoma is defined molecularly by its hallmark translocation between chromosome 8 and 14. The t(8;14) locates the *c-Myc* oncogene under control of the immunoglobulin heavy chain locus (IGH@) promoter. This results in an overexpression of the MYC protein and mistakes in the *V(D)J* gene segment assembly [12], responsible to build the antigen-binding site of the B-cell receptors (BcR).

B-cell receptors are transmembranous complexes formed by an immunoglobulin (IG) molecule and the intra-transmembranous heterodimer Ig-α/Ig-β. The latter component of the BcR is involved in intracellular signal transduction. The IG molecule is formed by two heavy and two light chains, named as such due to their length and molecular weight. Each heavy chain binds to a light chain by one disulphide bond and the two heavy chains contact each other via two more disulphide bonds in their hinge regions (the central part of the Y). Each heavy and light chain contains a variable and a constant domain, the latter being longer in heavy chains, explaining its bigger size. The variable domain is made up of three regions, called *variable* (*V*), *diversity* (*D*, only in heavy chains) and *joining* (*J*) gene segments. There are 77 *V(D)J* gene segments in the human IGH@, while IGK@ and IGL@ (the light chain loci) present only V and J segments. One single gene for each segment type is used to build a new variable domain in a genetic procedure called V(D)J or somatic recombination. It occurs in the primary lymphatic organs- that is

the bone marrow for B-lymphocytes. Briefly, the recombination activating genes RAG1 and RAG2 mediate the recombination procedure [13]. RAG1 binds to recombination signal sequences located between each V, D and J segment, and RAG2 is then recruited to form a RAG1/2 complex that will mediate a double-stranded DNA break. Firstly a D and a J segment are randomly selected and joined by a non-homologous end-joining procedure, forming one DJ coding joint and one signal joint or loop, containing the DNA excised from between the two joined segments. The excised loop will be degraded. The same procedure is repeated to join a V to a DJ segment. In case of a light chain recombination, where there are no D segments, the only coding joint created is between a V and a J segment.

Most importantly, IGs are able to recognise antigenic bodies by affinity binding to their antigen-binding sites (ABS). These ABS are composed of complementarity determining regions (CDR) within the variable domains of the IG heavy and light chains. The great sequence variability of IGs is achieved by V(D)J recombination and somatic hypermutation (SHM) of the CDRs [14]. Somatic hypermutation is a genetic alteration mechanism by which mutations are introduced in specific regions of IG genes in order to increase their sequence variability. The procedure is highly regulated by activation-induced deaminase (AID), which transforms the cytosine nucleotides to uracil. The resulting uracil-guanosine nucleotide pair is mutagenic, generating a thymidine in the forthcoming DNA replication. By expressing a wide range of V(D)J segment combinations and enduring SHM BcRs are known to be highly variable in their (appropriately named) variable domains.

Of note, the BcR array expressed in the surfaces of tumoral B-cells is significantly different than normal B-cells. Such biased IG repertoire points to an antigenic drive and the phenomenon has been detected in several lymphoma entities so far. The last two chapters of this thesis focus on this matter with regards to B-cell lymphoma and more particularly the BL framework.

Aims

This thesis combines the study of diagnostic and classificatory tools with the immunological and onset analysis of BL.

- Improvement of existing gene expression profiling classifiers
- Exploitation of archival biopsy material
- Study of B-cell lymphoma cohorts
- Review immunoglobulin bias in lymphoma with focus on Burkitt lymphoma
- Determine Burkitt lymphoma antibody epitopes experimentally

References

[1] Eurostat. Eurostat regional yearbook 2014: Health, 2014.

[2] J. Ferlay, E. Steliarova-Foucher, J. Lortet-Tieulent, S. Rosso, J.W.W. Coebergh, H. Comber, D. Forman, F. Bray. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. European journal of cancer (Oxford, England 1990) 49 (2013) 1374–1403.

[3] Cancer Research UK (2015), Non-Hodgkin lymphoma statistics, Cancer Research UK.

[4] E. Steliarova-Foucher, C. Stiller, B. Lacour, P. Kaatsch. International Classification of Childhood Cancer, third edition. Cancer 103 (2005) 1457–1467.

[5] K.A. Blum, G. Lozanski, J.C. Byrd. Adult Burkitt leukemia and lymphoma. Blood 104 (2004) 3009–3020.

[6] Archibald S. Perkins and Jonathan W. Friedberg. Burkitt lymphoma in Adults. Hematology / the Education Program of the American Society of Hematology. American Society of Hematology. Education Program (2008) 341–348.

[7] S.M. Mbulaiteye, W.F. Anderson, J. Ferlay, K. Bhatia, C. Chang, P.S. Rosenberg, S.S. Devesa, D.M. Parkin. Pediatric, elderly, and emerging adult-onset peaks in Burkitt's lymphoma incidence diagnosed in four continents, excluding Africa. American journal of hematology 87 (2012) 573–578.

[8] Jackson Orem, Edward Katongole Mbidde, Bo Lambert, Silvia de Sanjose, and Elisabete Weiderpass. Burkitt's lymphoma in Africa, a review of the epidemiology and etiology. Afr Health Sci. 3 (2007) 166–175.

[9] A. Gloghini, R. Dolcetti, A. Carbone. Lymphomas occurring specifically in HIV-infected patients: from pathogenesis to pathology. Seminars in Cancer Biology 23 (2013) 457–467.

[10] Swerdlow SH, Campo E, Harris NL, et al. Classification of Tumours of Haematopoietic and Lymphoid Tissues, IARC, 2008.

[11] Ash A. Alizadeh, Michael B. Eisen, R. Eric Davis, Chi Ma, Izidore S. Lossos, Andreas Rosenwald, Jennifer C. Boldrick, Hajeer Sabet, Truc Tran, Xin Yu, John I. Powell, Liming Yang, Gerald E. Marti, Troy Moore, James Hudson Jr, Lisheng Lu, David B. Lewis, Robert Tibshirani, Gavin Sherlock, Wing C. Chan, Timothy C. Greiner, Dennis D. Weisenburger, James O. Armitage, Roger Warnke, Ronald Levy, Wyndham Wilson, Michael R. Grever, John C. Byrd, David Botstein, Patrick O. Brown & Louis M. Staudt. Distinct types of diffuse large B-cell lymphoma identifed by gene expression profling. Nature 403 (2000) 503–511.

[12] F.G. Haluska, S. Finver, Y. Tsujimoto, C.M. Croce. The t(8; 14) chromosomal translocation occurring in B-cell malignancies results from mistakes in V-D-J joining. Nature 324 (1986) 158–161.

[13] M.J. Sadofsky. The RAG proteins in V(D)J recombination: More than just a nuclease. Nucleic acids research 29 (2001) 1399–1409.

[14] R.W. Maul, P.J. Gearhart. AID and Somatic Hypermutation, in: Elsevier, 2010, pp. 159–191.

Chapter 1

1.1 Introduction

Mature aggressive B-cell lymphomas currently present a double challenge when dealt with conventional methodology: correct entity differentiation and molecular subtype classification [1].

Differentiation between two clinical entities is useful to overcome potential diagnostic pitfalls. This can be illustrated with the case of diffuse large B-cell lymphoma (DLBCL) and Burkitt lymphoma (BL). Sharing histopathological features to DLBCL, BL is at times a relevant differential diagnostic entity from the former. However, DLBCL and BL classification is not discrete since a further subentity between the two lymphomas has been characterised, establishing a molecular continuum. This subentity is defined by the World Health Organisation (WHO) as "B-cell lymphoma, unclassifiable, with features intermediate between diffuse large B-cell lymphoma and Burkitt lymphoma" [2] and will be further referred to as intermediate lymphomas. Morphologically, intermediates lie in the grey zone between the classic starry sky appearance and monomorphic medium-sized cells of BL, and the mostly large centroblasts (and sometimes immunoblasts) with broad cytoplasm of DLBCL cells. The distinction would be mainly of academic relevance if it weren't for the fact that patient outcomes are inferior in BL treated with DLBCL regimes instead of an appropriate BL chemotherapy [3–5]. Thus, establishing a correct diagnosis is crucial for optimal therapy.

Molecular subclassification helps distinguish clinical subtypes that are not otherwise easily detected by conventional immunohistochemical (IHC) methods.

In particular, DLBCL can be subclassified in activated B-cell-like type (ABC), germinal centre B-cell-like type (GCB) and the unclassifiable subtype. These subdivisions are relevant as they carry a prognostic value [6] and newly developed therapeutic regimes target these subtype-specific pathomechanisms.

	GCB-DLBCL	*ABC-DLBCL*
Postulated normal counterpart	Germinal center B cell	Post-germinal center B-cell
Clinical outcome (5-year OS)	59%	30%
Oncogenic mechanism	*REL* amplification *BCL2* translocation	Constitutive activation of NF-κB
Chromosomal alterations	Gain 12q12 t(14;18)	Trisomy 3 (*FOXP1?*) Gain 3q Gain 18q21–q22 (*BCL2*) Deletion 6q21–q22 *(BLIMP1)*

Abbreviations: ABC, activated B-cell; DLBCL, diffuse large B-cell lymphoma; GCB, germinal center B; NF-κB, nuclear factor kappa B; OS, overall survival.

Figure 1.1: Clinical and genetic characteristics of ABC and GCB-like DLBCL [7].

It is pivotal that pathologists and clinicians are able to differentiate between aggressive B-cell lymphoma entities and within them, their molecular subtypes. Thus, several efforts have been made in the past to transfer the molecular definitions of lymphoma subtypes, based on fresh-frozen biopsies, into IHC classifiers using formalin-fixed paraffin-embedded (FFPE) biopsies. A summary of previous aggressive B-cell lymphoma classifications in IHC and gene expression profiling (GEP) follows.

Aggressive B-cell lymphoma classifiers - a background

Gene expression profiling has proven to be a very useful tool to understand cancer entities as well as to determine their transcriptional convergences and divergences. Some classifier approaches were based on molecular or GEP, well defining the molecular Burkitt lymphomas (mBL, clinical BL identified molecularly), intermediate B-cell lymphomas between DLBCL and mBL, GCB-like, ABC-like and unclassified DLBCL [4, 8-11]. Even a combination of both IHC and GEP approaches [12] was created by screening their cases with distinct IHC CD21 expression and then clustering them with a GCB/non-GCB gene set.

All these studies used Affymetrix microarray technology and fresh-frozen materials, due to the bad quality of FFPE-derived RNA of the time. Still, fresh-frozen biopsies are not always readily available in diagnostic and research facilities. Furthermore, microarray technology includes a reverse transcription and amplification steps, which can distort real mRNA expression levels.

Since FFPE materials are more readily available, don't need sophisticated storage space and are easy to handle several research groups tried to establish IHC protocols for lymphoma classification. One of the first IHC-based DLBCL classifier that tried to provide pathology labs with an easy decision tree to distinguish molecular subtypes was the one produced by Hans and colleagues [13] in 2004. It was based in tissue microarrays (TMA) of FFPE biopsies and centred on the GCB – non-GCB distinction. Although this was a step forward at the time the classifier was proven to be weak in assigning subgroups a prognostic factor and was not able to distinguish between the ABC and the unclassified subtypes within DLBCL.

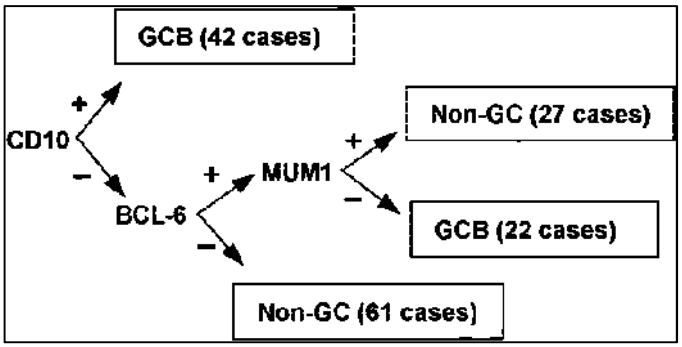


Figure 1.2: Hans et al. decision tree for DLBCL classification[13].

In 2006 an improved classifier by Muris et al. [14] introduced Bcl-2 as a decisive marker. Group 1 was described to have a low clinical risk and group 2 (likely enclosing the non-GCB subtype) presented a high clinical risk. The algorithm was stated to predict the odds for complete remission and relapse. The different naming and slightly different characterization of the groupings makes this classifier not directly comparable to their past and future counterparts.

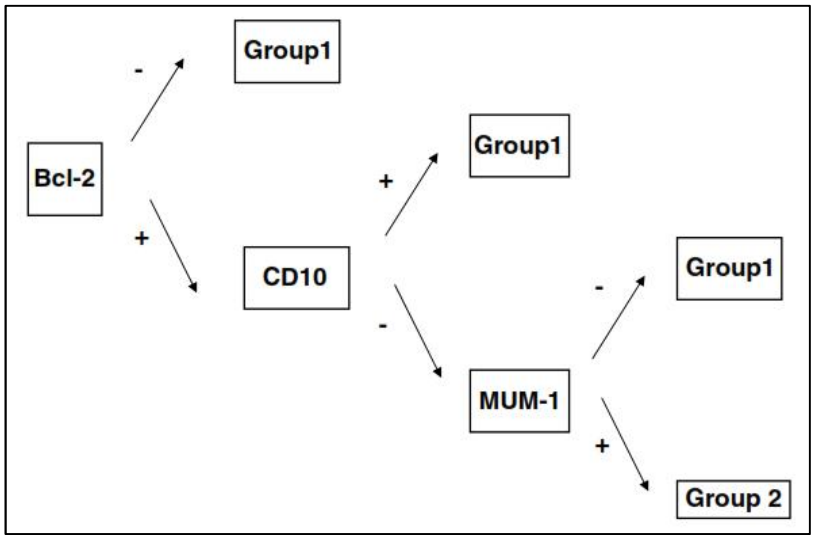


Figure 1.3: Muris et al. classifier tree [14].

Three years on, another publication tackled the IHC stratification of DLBCL. The group from Choi et al [15] created a new algorithm based on 5 IHC markers by validating it with a GEP tool. The complexity of this new approach allowed a distinction between ABC and GCB subtypes for the first time. A prognostic group classification was not included in the algorithm.

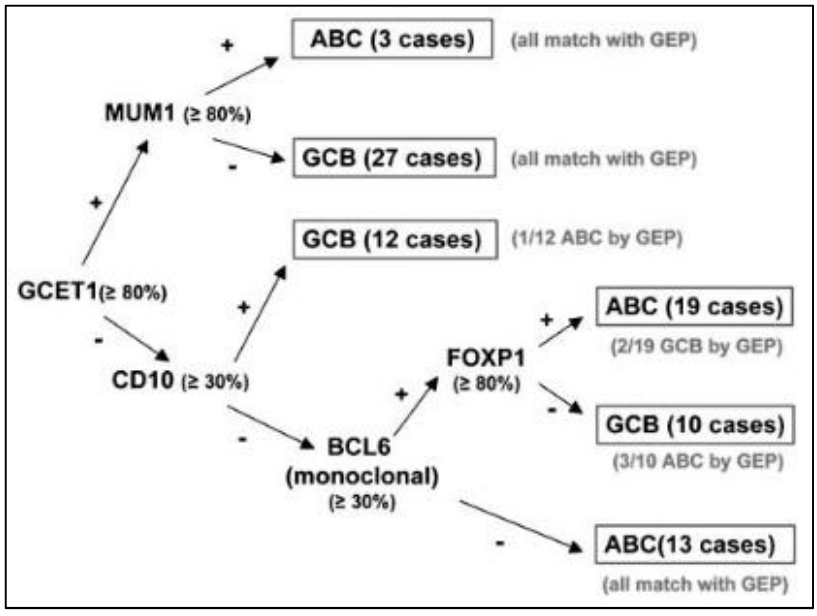


Figure 1.4: Choi et al classifier [15].

These IHC classifiers have in common the omission of BL and thus the lack of the intermediate group assignment. They also didn't assess the unclassified DLBCL subtype and comparison and replication efforts showed their relative unreliability [16].

RNA extraction techniques had lately improved by minimising degradation due to crosslinking of macromolecules in the purification process. We thus set to generate a molecular classifying tool for both BL and DLBCL subtypes from FFPE blocks, a more accessible and obtainable material with a new multiplex gene expression technology [17]. This work is presented in sections 1.2 and 1.3.

References to section 1.1

[1] Chan WC, Armitage JO, Gascoyne R, Connors J, Close P, Jacobs P, Norton A, Lister TA, Pedrinis E, Cavalli F, Berger F, Coiffier B, Ho F, Liang R, Ott G, Schauer A, Hiddemann W, Diebold J, MacLennan KA, Müller-Hermelink HK, Nathwani BN, Weisenburger DD, Harris NL, Anderson JR, and Roy P. A Clinical Evaluation of the International Lymphoma Study Group Classification of Non-Hodgkin's Lymphoma: By The Non-Hodgkin's Lymphoma Classification Project. Blood 89 (1997) 03918.

[2] Swerdlow SH, Campo E, Harris NL, et al. Classification of Tumours of Haematopoietic and Lymphoid Tissues, IARC, 2008.

[3] S. Smeland, A.K. Blystad, S.O. Kvaløy, I.M. Ikonomou, J. Delabie, G. Kvalheim, J. Hammerstrøm, G.F. Lauritzsen, H. Holte. Treatment of Burkitt's/Burkitt-like lymphoma in adolescents and adults: a 20-year experience from the Norwegian Radium Hospital with the use of three successive regimens. Annals of oncology official journal of the European Society for Medical Oncology / ESMO 15 (2004) 1072–1078.

[4] Dave Sandeep S. M.D. Fu Kai M.D., Ph.D. Wright George W. Ph.D. Lam Lloyd T. Ph.D. Kluin Philip M.D. Boerma Evert-Jan B.S. Greiner Timothy C. M.D. Weisenburger Dennis D. M.D. Rosenwald Andreas M.D. Ott German M.D. Müller-Hermelink Hans-Konrad M.D. Gascoyne Randy D. M.D. Delabie Jan M.D. Rimsza Lisa M. M.D. Braziel Rita M. M.D. Grogan Thomas M. M.D. Campo Elias M.D. Jaffe Elaine S. M.D. Dave Bhavana J. Ph.D. Sanger Warren Ph.D. Bast Martin B.S. Vose Julie M. M.D. Armitage James O. M.D. Connors Joseph M. M.D. Smeland Erlend B. M.D., Ph.D. Kvaloy Stein M.D., Ph.D. Holte Harald M.D., Ph.D. Fisher Richard I. M.D. Miller Thomas P. M.D. Montserrat Emilio M.D. Wilson Wyndham H. M.D., Ph.D. Bahl Manisha B.S. Zhao Hong M.S. Yang Liming Ph.D. Powell John M.S. Simon Richard D.Sc. Chan Wing C. M.D. Staudt Louis M. M.D., Ph.D. Molecular Diagnosis of Burkitt's Lymphoma. N Engl J Med (2006).

[5] D.E. Issa, van de Schans, Saskia A M, M.E.D. Chamuleau, H.E. Karim-Kos, M. Wondergem, P.C. Huijgens, J.W.W. Coebergh, S. Zweegman, O. Visser. Trends in incidence, treatment and survival of aggressive B-cell lymphoma in the Netherlands 1989-2010. Haematologica 100 (2015) 525–533.

[6] Ash A. Alizadeh, Michael B. Eisen, R. Eric Davis, Chi Ma, Izidore S. Lossos, Andreas Rosenwald, Jennifer C. Boldrick, Hajeer Sabet, Truc Tran, Xin Yu, John I. Powell, Liming Yang, Gerald E. Marti, Troy Moore, James Hudson Jr, Lisheng Lu, David B. Lewis, Robert Tibshirani, Gavin Sherlock, Wing C. Chan, Timothy C. Greiner, Dennis D. Weisenburger, James O. Armitage, Roger Warnke, Ronald Levy, Wyndham Wilson, Michael R. Grever, John C. Byrd, David Botstein, Patrick O. Brown & Louis M. Staudt. Distinct types of diffuse large B-cell lymphoma identifed by gene expression profling. Nature 403 (2000) 503–511.

[7] P De Paepe and C De Wolf-Peeters. Diffuse large B-cell lymphoma: a heterogeneous group of non-Hodgkin lymphomas comprising several distinct clinicopathological entities: Review. Leukemia 21 (2007) 37–43.

[8] Rosenwald Andreas, Wright George, Chan Wing C., Connors Joseph M., Campo Elias, Fisher Richard I., Gascoyne Randy D., Muller-Hermelink H. Konrad, Smeland Erlend B., Giltnane Jena M., Hurt Elaine M., Zhao Hong, Averett Lauren, Yang Liming, Wilson Wyndham H., Jaffe Elaine S., Simon Richard, Klausner Richard D., Powell John, Duffey Patricia L., Longo Dan L., Greiner Timothy C., Weisenburger Dennis D., Sanger Warren G., Dave Bhavana J., Lynch James C., Vose Julie, Armitage James O., Montserrat Emilio, López-Guillermo Armando, Grogan Thomas M., Miller Thomas P., LeBlanc Michel, Ott German, Kvaloy Stein, Delabie Jan, Holte Harald, Krajci Peter, Stokke Trond, Staudt Louis M. The Use of Molecular Profiling to Predict Survival after Chemotherapy for Diffuse Large-B-Cell Lymphoma.

[9] G. Wright, B. Tan, A. Rosenwald, E.H. Hurt, A. Wiestner, L.M. Staudt. A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. Proceedings of the National Academy of Sciences of the United States of America 100 (2003) 9991–9996.

[10] Hummel Michael, Bentink Stefan, Berger Hilmar, Klapper Wolfram, Wessendorf Swen, Barth Thomas F.E., Bernd Heinz-Wolfram, Cogliatti Sergio B., Dierlamm Judith, Feller Alfred C., Hansmann Martin-Leo, Haralambieva Eugenia, Harder Lana, Hasenclever Dirk, Kühn Michael, Lenze Dido, Lichter Peter, Martin-Subero Jose Ignacio, Möller Peter, Müller-Hermelink Hans-Konrad, Ott German, Parwaresch Reza M., Pott Christiane, Rosenwald Andreas, Rosolowski Maciej, Schwaenen Carsten, Stürzenhofecker Benjamin, Szczepanowski Monika, Trautmann Heiko, Wacker Hans-Heinrich, Spang Rainer, Loeffler Markus, Trümper Lorenz, Stein Harald, Siebert Reiner. A Biologic Definition of Burkitt's Lymphoma from Transcriptional and Genomic Profiling. N Engl J Med 354 (2006) 2419–2430.

[11] Diana Abdueva, Michele Wing, Betty Schaub, Timothy Triche, and Elai Davicioni. Quantitative expression profiling in formalin-fixed paraffin-embedded samples by affymetrix microarrays. The Journal of molecular diagnostics JMD 12 (2010) 409–417.

[12] K. Miyazaki, M. Yamaguchi, M. Suguro, W. Choi, Y. Ji, L. Xiao, W. Zhang, S. Ogawa, N. Katayama, H. Shiku, T. Kobayashi. Gene expression profiling of diffuse large B-cell lymphoma supervised by CD21 expression. British Journal of Haematology 142 (2008) 562–570.

[13] C.P. Hans, D.D. Weisenburger, T.C. Greiner, R.D. Gascoyne, J. Delabie, G. Ott, H.K. Müller-Hermelink, E. Campo, R.M. Braziel, E.S. Jaffe, Z. Pan, P. Farinha, L.M. Smith, B. Falini, A.H. Banham, A. Rosenwald, L.M. Staudt, J.M. Connors, J.O. Armitage, W.C. Chan. Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. Blood 103 (2004) 275–282.

[14] J.J.F. Muris, Meijer, C J L M, W. Vos, van Krieken, J H J M, N.M. Jiwa, G.J. Ossenkoppele, J.J. Oudejans. 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 208 (2006) 714–723.

[15] W.W.L. Choi, D.D. Weisenburger, T.C. Greiner, M.A. Piris, A.H. Banham, J. Delabie, R.M. Braziel, H. Geng, J. Iqbal, G. Lenz, J.M. Vose, C.P. Hans, K. Fu, L.M. Smith, M. Li, Z. Liu, R.D. Gascoyne, A. Rosenwald, G. Ott, L.M. Rimsza, E. Campo, E.S. Jaffe, D.L. Jaye, L.M. Staudt, W.C. Chan. A new immunostain algorithm classifies diffuse large B-cell lymphoma into molecular subtypes with high accuracy. Clinical cancer research an official journal of the American Association for Cancer Research 15 (2009) 5494–5502.

[16] G. Ott, M. Ziepert, W. Klapper, H. Horn, M. Szczepanowski, H.-W. Bernd, C. Thorns, A.C. Feller, D. Lenze, M. Hummel, H. Stein, H.-K. Müller-Hermelink, M. Frank, M.-L. Hansmann, T.F.E. Barth, P. Möller, S. Cogliatti, M. Pfreundschuh, N. Schmitz, L. Trümper, M. Loeffler, A. Rosenwald. Immunoblastic morphology but not the

immunohistochemical GCB/nonGCB classifier predicts outcome in diffuse large B-cell lymphoma in the RICOVER-60 trial of the DSHNHL. Blood 116 (2010) 4916–4925.

[17] G.K. Geiss, R.E. Bumgarner, B. Birditt, T. Dahl, N. Dowidar, D.L. Dunaway, H.P. Fell, S. Ferree, R.D. George, T. Grogan, J.J. James, M. Maysuria, J.D. Mitton, P. Oliveri, J.L. Osborn, T. Peng, A.L. Ratcliffe, P.J. Webster, E.H. Davidson, L. Hood, K. Dimitrov. Direct multiplexed measurement of gene expression with color-coded probe pairs. Nature biotechnology 26 (2008) 317–325.

1.2 To the editor:

Molecular classification of mature aggressive B-cell lymphoma using digital multiplexed gene expression on formalin-fixed paraffin-embedded biopsy specimens

Neus Masqué-Soler[1], Monika Szczepanowski[1], Christian W. Kohler[2], Rainer Spang[2], and Wolfram Klapper[1].

[1]Department of Pathology, Haematopathology Section and Lymph Node Registry, University Hospital Schleswig-Holstein, Campus Kiel/Christian-Albrecht University, Kiel, Germany.

[2] Institute of Functional Genomics, University of Regensburg, Regensburg, Germany.

This piece was published in *Blood*, on the 12th September of 2013, Vol 122, Num 11 (1985-1986).

The most frequent mature aggressive B-cell lymphomas are diffuse large B-cell lymphoma (DLBCL) and Burkitt lymphoma (BL). Patients suffering from molecularly defined BL (mBL) but treated with a regimen developed for DLBCL show an unfavourable outcome compared with mBL treated with chemotherapy regimens for BL [1]. Distinguishing BL from DLBCL by conventional histopathology is challenging in lymphomas that have features common to both diseases (aggressive B-cell lymphoma unclassifiable with features of DLBCL and BL [intermediates]) [2]. Moreover, DLBCLs are a heterogeneous group of lymphomas comprising distinct molecular subtypes: the activated B-cell–like (ABC), the germinal center B-cell–like (GCB), and the unclassifiable subtype as defined by gene expression profiling (GEP) [3]. Attempts to replace GEP with techniques applicable to formalin-fixed paraffin-embedded (FFPE) tissue led to algorithms for immunohistochemical staining (IHS) [4]. Disappointingly, the algorithms yielded conflicting results with respect to their prognostic potential, raising concerns about their validity [5]. Furthermore, IHS algorithms did not provide a fully resolved classification: they did not identify mBL nor did they separate ABC from unclassified DLBCLs [4].

We used digital multiplexed gene expression (DMGE) with FFPE-derived RNA to classify aggressive B-cell lymphomas. Our assay comprised only 30 genes (10 for the detection of mBL and 20 for the detection of ABC and GCB). We chose these genes by reanalysis of the microarray data reported in a previous study [6]. A detailed description of the methods is provided in the supplemental Materials on the *Blood* website. Thirty-nine samples from mature aggressive B-cell lymphomas were analysed using DMGE (nCounter; NanoString Technologies, Seattle, WA; see supplemental Materials for detailed methods) of FFPE- and fresh-frozen–derived RNA. All cases were previously characterized by the Molecular Mechanisms of Malignant Lymphoma [6] consortium using the Affymetrix GeneChip technology (gold standard of classification).

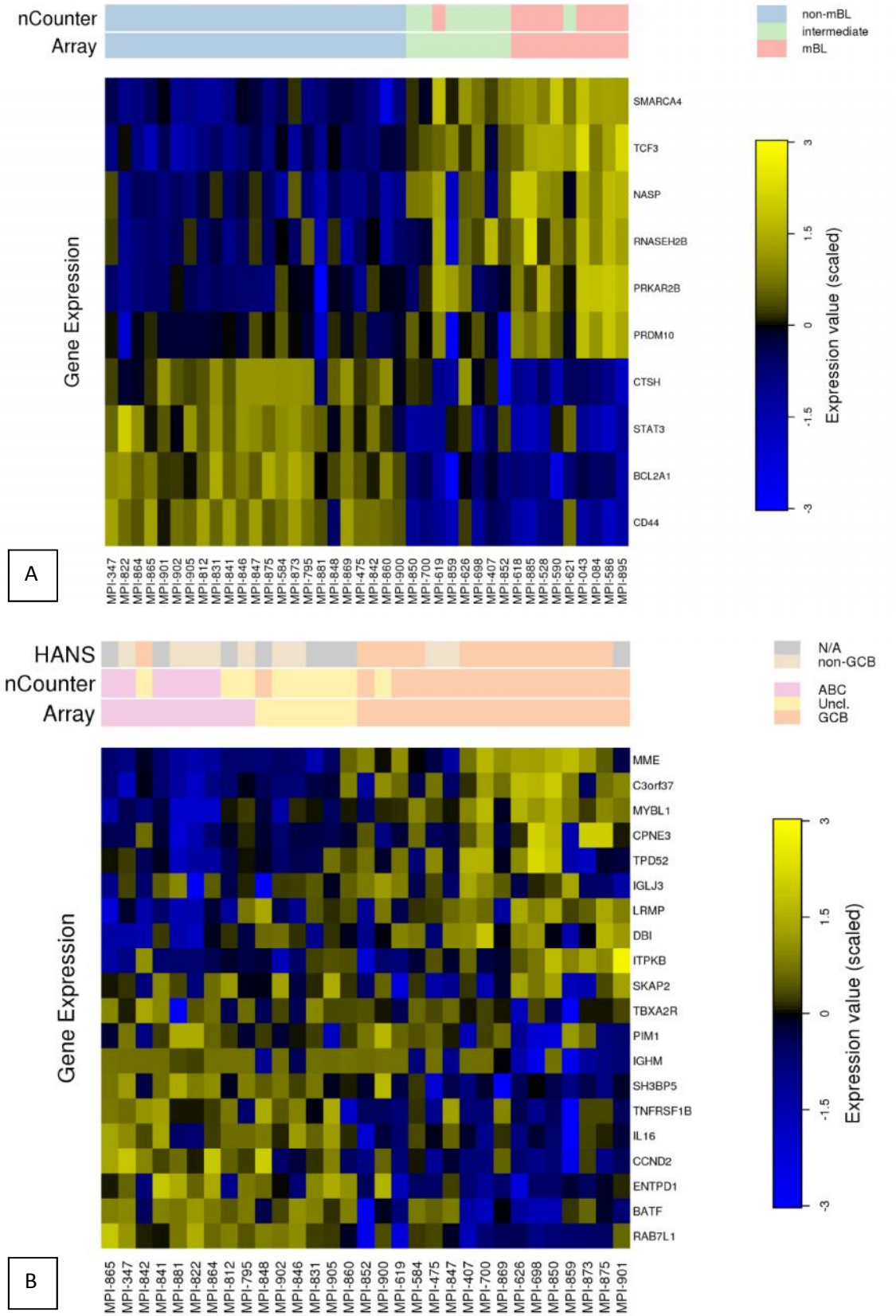


Figure 1.5: Digital multiplexed gene expression of mature aggressive B-cell lymphomas. (A) molecular classification of mature aggressive B-cell lymphomas as mBL, non-mBL, and intermediates using FFPE-derived RNA. The expression of the genes is colour

coded such that high expression is shown in yellow and low expression in blue. The molecular labels assigned by the array using fresh-frozen RNA and DMGE (nCounter) assay using FFPE-derived RNA are indicated as coloured labels in the top bars. (B) Molecular classification of non-mBL using FFPE-derived RNA. The classification according to IHS (Hans), the array using fresh-frozen RNA and the DMGE assay using FFPE-derived RNA are indicated as coloured labels in the top bars.

For FFPE-derived RNA, the classification of only 2 of 39 samples (5%) differed, when comparing DMGE-based predictions with the gold standard (array) (Figure 1A). The 2 divergent classifications were 1 case of mBL that was classified as intermediate and 1 case of an intermediate that was classified as mBL [both lymphomas carried a t(8;14) translocation]. No major mistake (a classification of a non-mBL as mBL or vice versa) was observed. This performance is comparable with that of DMGE data from fresh-frozen tissue blocks (supplemental Materials).Compared with the array, 5 of 31 (16%) non-mBL/DLBCLs received different molecular classifications using DMGE and FFPE-derived RNA (Figure 1B). All discrepancies comprised lymphomas that switched between the unclassified and ABC or unclassified and GCB labels. Again, no major mistake (a classification of a GCB as an ABC or vice versa) was observed. In contrast, the Hans IHS algorithm[7] led to major mistakes (2 GCB classified as non-GCB and 1 ABC as GCB; Figure 1B).

GEP-based molecular classification of mature aggressive B-cell lymphomas is possible with FFPE-derived RNA at a reasonable cost (<100 Euro) and within a reasonable period of time (24 hours). To the best of our knowledge, this is the first molecular method for identifying mBL among mature aggressive B-cell lymphomas using FFPE-derived RNA.

Support

This work was supported by the KinderKrebsInitiative Buchholz/Holm-Seppensen, an intramural grant from the Medical Faculty of the University of Kiel to W.K. (grant F343911), and the Molecular Mechanisms of Malignant Lymphoma with MYC-Dysregulation - A System Biology Approach for Genetics, Evolution, Signaling and Clinical Care (MMML-MYC-SYS) funded by the German Ministry of Health (grant 0316166).

References to section 1.2

[1] Dave Sandeep S. M.D. Fu Kai M.D., Ph.D. Wright George W. Ph.D. Lam Lloyd T. Ph.D. Kluin Philip M.D. Boerma Evert-Jan B.S. Greiner Timothy C. M.D. Weisenburger Dennis D. M.D. Rosenwald Andreas M.D. Ott German M.D. Müller-Hermelink Hans-Konrad M.D. Gascoyne Randy D. M.D. Delabie Jan M.D. Rimsza Lisa M. M.D. Braziel Rita M. M.D. Grogan Thomas M. M.D. Campo Elias M.D. Jaffe Elaine S. M.D. Dave Bhavana J. Ph.D. Sanger Warren Ph.D. Bast Martin B.S. Vose Julie M. M.D. Armitage James O. M.D. Connors Joseph M. M.D. Smeland Erlend B. M.D., Ph.D. Kvaloy Stein M.D., Ph.D. Holte Harald M.D., Ph.D. Fisher Richard I. M.D. Miller Thomas P. M.D. Montserrat Emilio M.D. Wilson Wyndham H. M.D., Ph.D. Bahl Manisha B.S. Zhao Hong M.S. Yang Liming Ph.D. Powell John M.S. Simon Richard D.Sc. Chan Wing C. M.D. Staudt Louis M. M.D., Ph.D. Molecular Diagnosis of Burkitt's Lymphoma. N Engl J Med (2006).

[2] Swerdlow SH, Campo E, Harris NL, et al. Classification of Tumours of Haematopoietic and Lymphoid Tissues, IARC, 2008.

[3] Ash A. Alizadeh, Michael B. Eisen, R. Eric Davis, Chi Ma, Izidore S. Lossos, Andreas Rosenwald, Jennifer C. Boldrick, Hajeer Sabet, Truc Tran, Xin Yu, John I. Powell, Liming Yang, Gerald E. Marti, Troy Moore, James Hudson Jr, Lisheng Lu, David B. Lewis, Robert Tibshirani, Gavin Sherlock, Wing C. Chan, Timothy C. Greiner, Dennis D. Weisenburger, James O. Armitage, Roger Warnke, Ronald Levy, Wyndham Wilson, Michael R. Grever, John C. Byrd, David Botstein, Patrick O. Brown & Louis M. Staudt. Distinct types of diffuse large B-cell lymphoma identifed by gene expression profling. Nature 403 (2000) 503–511.

[4] P.N. Meyer, K. Fu, T.C. Greiner, L.M. Smith, J. Delabie, R.D. Gascoyne, G. Ott, A. Rosenwald, R.M. Braziel, E. Campo, J.M. Vose, G. Lenz, L.M. Staudt, W.C. Chan, D.D. Weisenburger. 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 29 (2011) 200–207.

[5] G. Salles, D. de Jong, W. Xie, A. Rosenwald, M. Chhanabhai, P. Gaulard, W. Klapper, M. Calaminici, B. Sander, C. Thorns, E. Campo, T. Molina, A. Lee, M. Pfreundschuh, S. Horning, A. Lister, L.H. Sehn, J. Raemaekers, A. Hagenbeek, R.D. Gascoyne, E. Weller. Prognostic significance of immunohistochemical biomarkers in diffuse large B-cell lymphoma: a study from the Lunenburg Lymphoma Biomarker Consortium. Blood 117 (2011) 7070–7078.

[6] W. Klapper, M. Kreuz, C.W. Kohler, B. Burkhardt, M. Szczepanowski, I. Salaverria, M. Hummel, M. Loeffler, S. Pellissery, W. Woessmann, C. Schwänen, L. Trümper, S. Wessendorf, R. Spang, D. Hasenclever, R. Siebert. Patient age at diagnosis is associated with the molecular characteristics of diffuse large B-cell lymphoma. Blood 119 (2012) 1882–1887.

1.3 Supplementary material to Molecular classification of mature aggressive B-cell lymphoma using digital multiplexed gene expression on formalin-fixed paraffin-embedded biopsy specimens.

1.3.1 Supplementary methods

RNA-extraction and assessment of quality

Formalin-fixed paraffin-embedded material was cut in five 10 μm-thick pieces per sample. Extraction of RNA was done according to the manufacturer's instructions (ExpressArt FFPE Clear RNAready Kit, AmpTec, Hamburg, Germany). Fresh-frozen blocks' RNA was extracted as previously described [1]. To assess the RNA quality all 43 FFPE-extracted and the 2 newly extracted fresh-frozen RNA material was analysed with the Agilent RNA 6000 Nano Chips (Agilent Technologies, Santa Clara, California, USA) following the product's protocol. As expected, the quality of the FFPE samples were considerably degraded due to fixation steps but generally all samples yielded enough RNA with fragments at least longer than 200 nucleotides. A further material quality control is present after RNA processing at the nCounter instrument for digital multiplexed gene expression due to the company's recommendations (see below for summary).

Digital multiplexed gene expression (DMGE)

Straight after a RNA quality measurement and with a known RNA concentration for each sample, 300ng were able to be hybridised with our custom-designed Reporter Probe (each with a specific 5'-end colour code signal) and Capture Probe (marked with biotin at the 3' end). After the overnight hybridisation the samples were ready to load to the Prep Station where they were washed with a two-step magnetic bead-purification process and loaded in the 12-well cartridges. Afterwards the cartridges were loaded to the Digital Analyzer (DA, nCounter, NanoString Technologies Inc., Seattle, WA, USA). The DA resolution was set to high, meaning that 280 fields of view (FOV) were counted per sample. A comma-separated-value (.csv) file was obtained after processing. The first control on data-reliability was done with the FOV-counter to FOV-count ratio, which had to be >80%. Secondly, an acceptable binding density could not lie outside the 0.05-2.25 interval. Afterwards a data normalisation was done: individually adding up the internal positive controls and finding an average value in order to create a normalisation factor for each sample allowed for cartridge-to-cartridge comparison. The third quality control was visible at this point, since reliable samples had to show a normalisation factor between 0.3 and 3. Finally, through each sample's average internal negative control values, a background subtraction was performed. After these correction steps, nCounter data of FFPE and fresh frozen samples were normalized independently from each other using quantile normalization [2] and finally log2 transformed afterwards.

Study cohort

All cases were previously characterized by the Molecular Mechanisms of Malignant Lymphoma (MMML [3]) consortium using the Affymetrix GeneChip® (U133A) technology (referred to as the gold standard classification). We used the same original total RNA used for arrays for 48/50 (96%) samples (11 mBL, 12 intermediates and 27 non-mBL, the latter composed of 10 ABC, 8 GBC, 9 "unclassified"). For 2/50 (4%) cases the original RNA was exhausted. For these cases total RNA was newly extracted from the same fresh-frozen tissue blocks used previously for array based GEP. In addition, total RNA was isolated from corresponding FFPE tissue blocks in 40 cases (ExpressArt FFPE Clear RNAready Kit, AmpTec, Hamburg, Germany): 9 mBL, 8 intermediates, and 23 non-mBL, the latter composed of 9 ABC, 8 GBC, and 6 unclassified. For 39 cases GEPs from both fresh-frozen material and the FFPE tissue blocks were available (Supplementary Figure 1#).

Selection of genes and bioinformatics

A subset of 5/74 probe sets from the original mBL signature (representing the genes *SMARCA4*, *CD44*, *RNASEH2B*, *PRKAR2B*, and *CTSH*) proved to be sufficient to classify mBL and non-mBL cases correctly. However, we also found that the small number of signature genes comes with a price. The robustness of the classification with respect to measurement errors is reduced. We thus selected one back-up gene for all five genes. This was always the gene on the microarray with the highest correlation to the signature gene across all training samples (12). These 5 genes (*TCF3*, *STAT3*, *NASP*, *PRDM10*, and *BCL2A1*) together with the 5 previous genes form the mBL set of the assay. In the same way, we chose 10 genes from the GCB/ABC signature [4] and supplemented them with 10 back-up genes, giving us the ABC/GCB set of the assay. Supplementary table 1* gives a complete list of the classifier's 30 genes with their NCBI reference numbers.

With our final 30 gene assay we obtained GEP for 50 fresh-frozen samples from mature aggressive B-cell lymphomas using DMGE. To obtain molecular classifications of "mBL", "intermediate", and "non-mBL" from DMGE data, we first trained a linear regression function to predict the mBL index [1] of lymphomas using the 10 genes of the mBL set. We used least-angle regression [5] and predicted all mBL-index values in leave-one-out cross-validation. We then calibrated two cut-off values C1 and C2 on the predicted mBL-index values and labelled a case "mBL" if the predicted index was below C1. If it was above C2 we labelled it "non-mBL", and otherwise "intermediate". This procedure was repeated independently for frozen and FFPE derived data.

Currently figure 1.6

* Currently table 1.1

Note: More information on the Code Set gene selection can be found in section 1.4.1

1.3.2 Supplementary figures

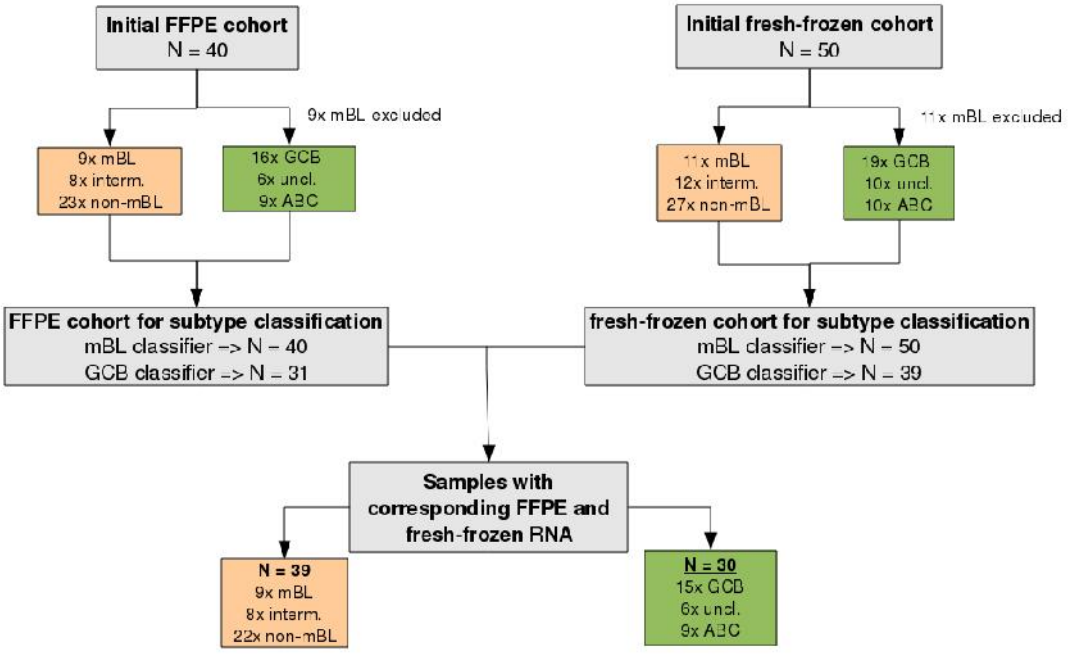


Figure 1.6*: Description of the study cohort.

* Figure 1.6 (in original publication, "Supplementary figure 1") shows the experiment workflow and sample dimensions. Orange boxes hold cohort specifications for the mBL – non-mBL classifier and green boxes do so for the GCB-ABC classifier. A total of 39 sample data was available from the same biopsy in FFPE and fresh-frozen material RNA. They were classified with the mBL - non-mBL algorithm. Thirty samples were clustered with our GCB-ABC classifier and compared between their FFPE and cryo biopsy counterparts. Data from the Molecular Mechanisms of Malignant Lymphoma consortium (see section 1.2, reference 6) was available from the fresh-frozen samples, which had been classified according to data from Affymetrix's GeneChip® processing.

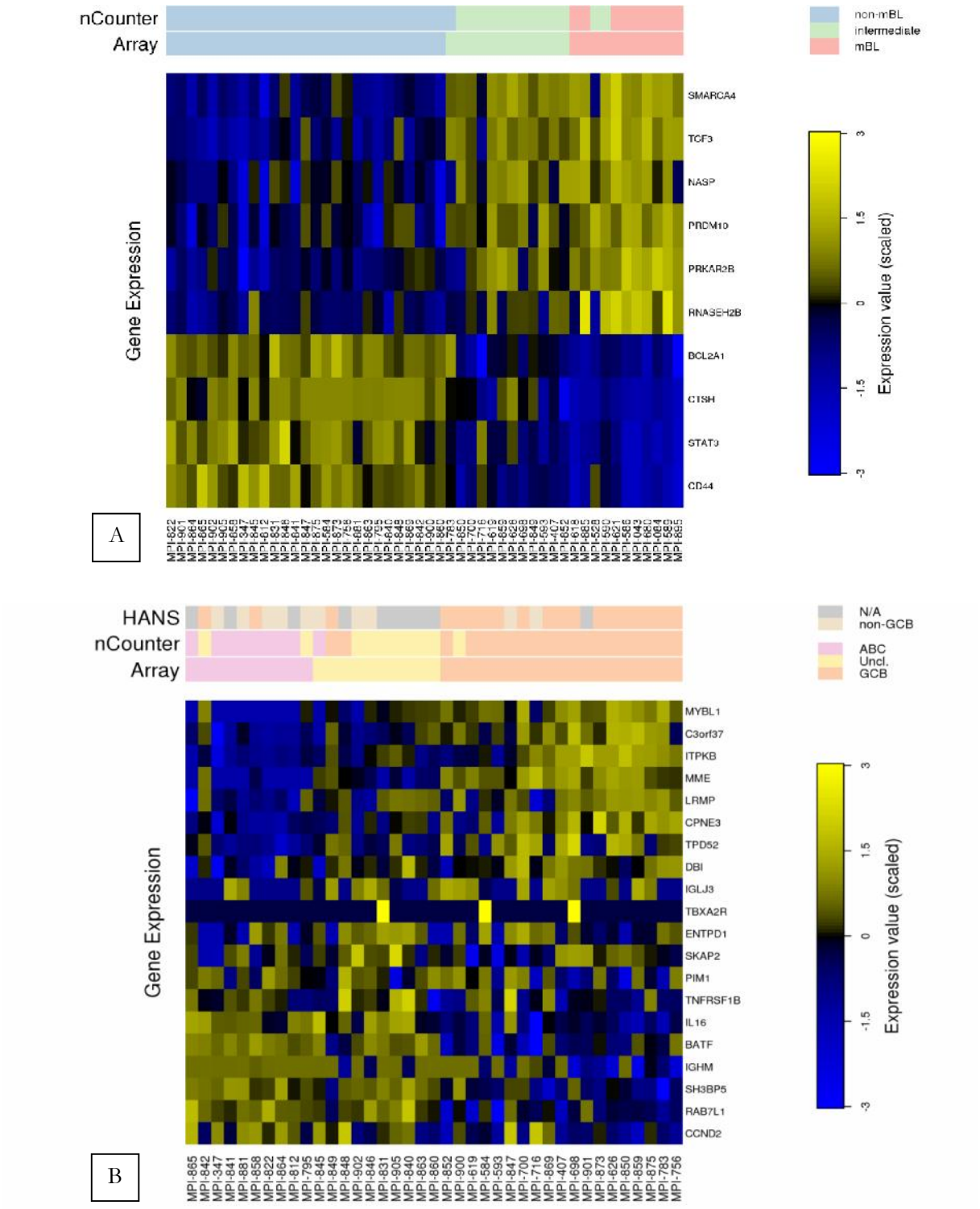


Figure 1.7: (A) Molecular classification of mature aggressive B-cell lymphomas as mBL, non-mBL and intermediates using cryo-block (fresh frozen)-derived RNA. The expression of the genes is displayed in a colour code with high expression in yellow and low expression in blue. The molecular labels assigned by the array using fresh-frozen RNA and DMGE (nCounter) assay using fresh frozen-derived RNA are indicated as coloured labels in the top bars. (B) Molecular classification of non-mBL using fresh frozen-derived RNA. The classification according to the immunohistochemical staining (Hans [6]), the array using fresh-frozen RNA and DMGE assay using fresh frozen-derived RNA are indicated as coloured labels in the top bars.

1.3.3 Supplementary tables

mBL signature	Accession number
BCL2A1	NM_004049.2
CD44	NM_001001392.1
CTSH	NM_148979.2
NASP	NM_172164.1
PRDM10	NM_020228.2
PRKAR2B	NM_002736.2
RNASEH2B	NM_001142279.1
SMARCA4	NM_003072.3
STAT3	NM_139276.2
TCF3	NM_003200.2

GCB/ABC signature	Accession number
BATF	NM_006399.3
C3orf37	NM_020187.2
CCND2	NM_001759.3
CPNE3	NM_003909.2
DBI	NM_001079862.1
ENTPD1	NM_001098175.1
IGHM	ENST00000390559.1
IGLJ3	ENST00000390324.1
IL16	NM_004513.4
ITPKB	NM_002221.3
LRMP	NM_006152.2
MME	NM_000902.2
MYBL1	NM_001080416.2
PIM1	NM_002648.2
RAB7L1	NM_001135664.1
SH3BP5	NM_001018009.2
SKAP2	NM_003930.3
TBXA2R	NM_001060.3
TNFRSF1B	NM_001066.2
TPD52	NM_001025252.1

Endogenous controls	Accession number
GUSB	NM_000181.1
EPHB3	NM_004443.3
INE1	NR_024616.1
EML2	NM_012155.2

Table 1.1*: Genes analysed.

* Table 1.1 (originally "supplementary table 1" in publication) shows the signature genes with their accession numbers. Core genes of each classifier are marked in red and in black are back-up genes.

mBL classifier

	GEP		
SAMPLE	**Array**	**DMGE frozen**	**DMGE FFPE**
MPI-043	mBL	mBL	mBL
MPI-084	mBL	mBL	mBL
MPI-347	non-mBL	non-mBL	non-mBL
MPI-407	intermediate	intermediate	intermediate
MPI-528	mBL	intermediate	mBL
MPI-584	non-mBL	non-mBL	non-mBL
MPI-586	mBL	mBL	mBL
MPI-590	mBL	intermediate	mBL
MPI-618	mBL	mBL	mBL
MPI-619	intermediate	intermediate	mBL
MPI-621	mBL	mBL	intermediate
MPI-626	intermediate	intermediate	intermediate
MPI-698	intermediate	intermediate	intermediate
MPI-700	intermediate	intermediate	intermediate
MPI-795	non-mBL	non-mBL	non-mBL
MPI-812	non-mBL	non-mBL	non-mBL
MPI-822	non-mBL	non-mBL	non-mBL
MPI-831	non-mBL	non-mBL	non-mBL
MPI-841	non-mBL	non-mBL	non-mBL
MPI-842	non-mBL	non-mBL	non-mBL
MPI-846	non-mBL	non-mBL	non-mBL
MPI-847	non-mBL	non-mBL	non-mBL
MPI-848	non-mBL	non-mBL	non-mBL
MPI-850	intermediate	intermediate	intermediate
MPI-852	intermediate	intermediate	intermediate
MPI-859	intermediate	intermediate	intermediate
MPI-860	non-mBL	non-mBL	non-mBL
MPI-864	non-mBL	non-mBL	non-mBL
MPI-865	non-mBL	non-mBL	non-mBL
MPI-869	non-mBL	non-mBL	non-mBL
MPI-873	non-mBL	non-mBL	non-mBL
MPI-875	non-mBL	non-mBL	non-mBL
MPI-881	non-mBL	non-mBL	non-mBL
MPI-885	mBL	mBL	mBL
MPI-895	mBL	mBL	mBL
MPI-900	non-mBL	non-mBL	non-mBL
MPI-901	non-mBL	non-mBL	non-mBL
MPI-902	non-mBL	non-mBL	non-mBL
MPI-905	non-mBL	non-mBL	non-mBL

Table continued on next page

GCB-ABC classifier

		GEP		
SAMPLE	**IHS (Hans)**	**Array**	**DMGE frozen**	**DMGE FFPE**
MPI-347	non-GCB	ABC	ABC	ABC
MPI-407	GCB	GCB	GCB	GCB
MPI-584	GCB	GCB	GCB	GCB
MPI-619	GCB	GCB	GCB	GCB
MPI-626	GCB	GCB	GCB	GCB
MPI-698	GCB	GCB	GCB	GCB
MPI-700	GCB	GCB	GCB	GCB
MPI-795	non-GCB	ABC	unclassified	unclassified
MPI-812	N/A	ABC	ABC	unclassified
MPI-822	non-GCB	ABC	ABC	ABC
MPI-831	N/A	unclassified	unclassified	unclassified
MPI-841	N/A	ABC	ABC	ABC
MPI-842	GCB	ABC	unclassified	unclassified
MPI-846	non-GCB	unclassified	unclassified	unclassified
MPI-847	non-GCB	GCB	GCB	GCB
MPI-848	N/A	unclassified	GCB	GCB
MPI-850	GCB	GCB	GCB	GCB
MPI-852	GCB	GCB	GCB	GCB
MPI-859	GCB	GCB	GCB	GCB
MPI-860	N/A	unclassified	unclassified	unclassified
MPI-864	non-GCB	ABC	ABC	ABC
MPI-865	N/A	ABC	ABC	ABC
MPI-869	GCB	GCB	GCB	GCB
MPI-873	GCB	GCB	GCB	GCB
MPI-875	GCB	GCB	GCB	GCB
MPI-881	non-GCB	ABC	ABC	ABC
MPI-900	GCB	GCB	unclassified	unclassified
MPI-901	N/A	GCB	GCB	GCB
MPI-902	non-GCB	unclassified	unclassified	unclassified
MPI-905	N/A	unclassified	unclassified	unclassified

Table 1.2*: Complete list of classifications according to array, IHS (Hans Algorithm [6]), DMGE using cryo-derived RNA, DMGE using FFPE derived RNA. Lymphomas with divergent classification between the assays are labelled in bold. Only lymphomas of the final cohort for which both array and DMGE data were available are shown. DMGE: digital multiplex gene expression.

* Table 1.2 : Originally "Supplementary Table 2" in publication.

References to section 1.3

[1] Hummel Michael, Bentink Stefan, Berger Hilmar, Klapper Wolfram, Wessendorf Swen, Barth Thomas F.E., Bernd Heinz-Wolfram, Cogliatti Sergio B., Dierlamm Judith, Feller Alfred C., Hansmann Martin-Leo, Haralambieva Eugenia, Harder Lana, Hasenclever Dirk, Kühn Michael, Lenze Dido, Lichter Peter, Martin-Subero Jose Ignacio, Möller Peter, Müller-Hermelink Hans-Konrad, Ott German, Parwaresch Reza M., Pott Christiane, Rosenwald Andreas, Rosolowski Maciej, Schwaenen Carsten, Stürzenhofecker Benjamin, Szczepanowski Monika, Trautmann Heiko, Wacker Hans-Heinrich, Spang Rainer, Loeffler Markus, Trümper Lorenz, Stein Harald, Siebert Reiner. A Biologic Definition of Burkitt's Lymphoma from Transcriptional and Genomic Profiling. N Engl J Med 354 (2006) 2419–2430.

[2] B.M. Bolstad, R. Irizarry, M. Astrand, T.P. Speed. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 19 (2003) 185–193.

[3] W. Klapper, M. Kreuz, C.W. Kohler, B. Burkhardt, M. Szczepanowski, I. Salaverria, M. Hummel, M. Loeffler, S. Pellissery, W. Woessmann, C. Schwänen, L. Trümper, S. Wessendorf, R. Spang, D. Hasenclever, R. Siebert. Patient age at diagnosis is associated with the molecular characteristics of diffuse large B-cell lymphoma. Blood 119 (2012) 1882–1887.

[4] G. Wright, B. Tan, A. Rosenwald, E.H. Hurt, A. Wiestner, L.M. Staudt. A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. Proceedings of the National Academy of Sciences of the United States of America 100 (2003) 9991–9996.

[5] Efron B, Hastie T, Johnstone I and Tibshirani R. Least angle regression. Annals of Statistics 32 (2004) 407–499.

[6] C.P. Hans, D.D. Weisenburger, T.C. Greiner, R.D. Gascoyne, J. Delabie, G. Ott, H.K. Müller-Hermelink, E. Campo, R.M. Braziel, E.S. Jaffe, Z. Pan, P. Farinha, L.M. Smith, B. Falini, A.H. Banham, A. Rosenwald, L.M. Staudt, J.M. Connors, J.O. Armitage, W.C. Chan. Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. Blood 103 (2004) 275–282.

1.4 Additional data

The following materials are intended to be complementary to the above exposed contents and are previously unpublished.

1.4.1 Material and Methods

nCounter technology

Previous methods used for gene expression studies included reverse-transcription and amplification of the analytes. In the original NanoString technology [1] a pair of primer-like probes is built per analysable gene. The probes are named capture and reporter probes. Capture probes are made of (5' to 3'): a 35-50 nucleotide-long sequence homologous to a mRNA of interest, a repeats region, a biotin tag and a repeat sequence at its 3' end (common to all capture probes). Reporter probes are composed of (3' to 5'): 35-50 nucleotides homologous to mRNA fragment, followed by an ssDNA backbone bound to RNA coloured fragments. The coloured fragments build a colour tag that is specific for the sequence that will be recognized by the probe pair. At the 5' of the reporter probes there is also a common repeat sequence to all reporter probes. The reporter and capture probes plus their spiked-in control molecules are named Code Set. Six capture and reporter pair probes are spiked-in for positive control purposes and eight probe pairs are present as negative controls. Positive and negative controls are also useful for normalization purposes, since they are present in known quantities in the Code Set.

The Code Set is hybridised overnight to the isolated mRNA. The capture and reporter probes interact to and bind the homolog sequences found in the mRNA, forming a tripartite molecule. After the hybridization the probe and mRNA mix is placed in the Prep Station (PS), which in turn is loaded with reagent, materials and an acrylic glass cartridge. Each cartridge has twelve lanes, one for each sample to be analysed and is processed individually in the PS. Unhybridized probes are removed in the washing steps through affinity purification. The samples are then loaded by the PS's robot arm to the streptavidin-coated lanes of the cartridge. The biotin molecules of the tripartite structures stick to the cartridge lanes. In each lane the PS applies a 160 V/cm electric current so that the bound molecules lay in the same direction. Finally, biotinylated anti-5' oligonucleotides are introduced in the lanes, binding to the 5' repeats end of the reporter probes (purple strand in figure 1.8-A and -B). This ensures that the tripartite molecules are immobilized horizontally to the lane base. When the current is turned off the molecules lay in the same direction with minimal crisscrossing.

After the samples are loaded in the PS to the cartridge, it is placed into the Digital Analyzer (DA) which scans each lane to retrieve images of the coloured tags. Each coloured tag is a reporter probe attached to a single specific mRNA molecule. The DA is

equipped with a microscopic lens and automatically moves underneath the binding surface of the cartridge through its fields of view (FOV). The more FOV are set up to be counted, the longer it takes for the DA to finish a scan. There are 25, 100, 280 and 555 FOV with GEN2 nCounter machines respectively named low, medium, high and max resolution settings.

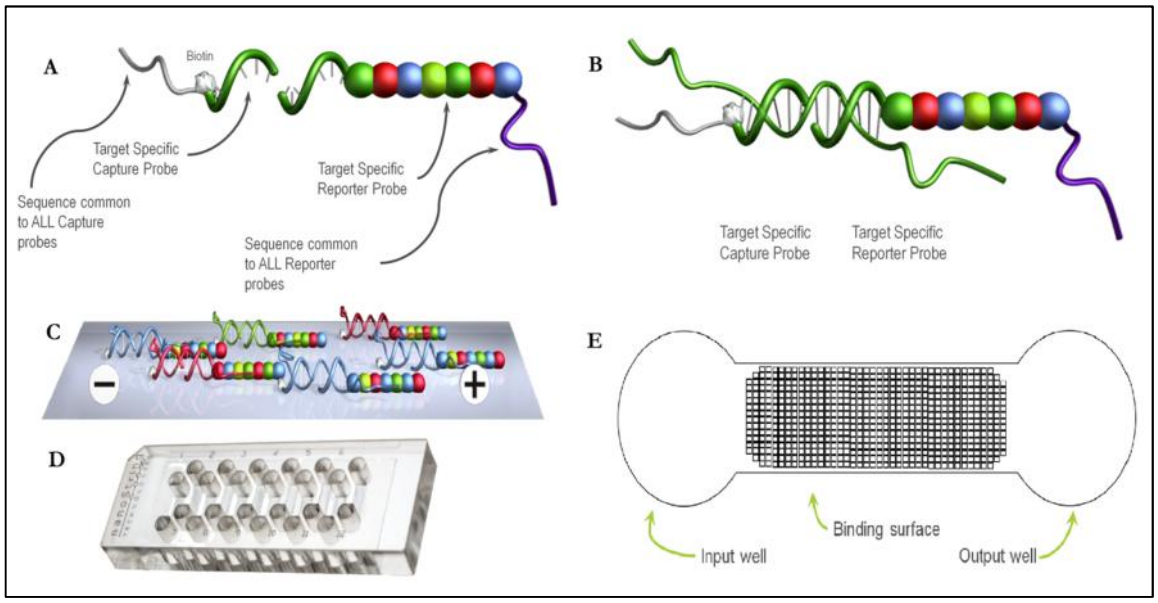


Figure 1.8: **A**: Code Set components, Capture and Reporter probes. **B**: Capture and reporter probes schematized. mRNA is the linear molecule in green. After hybridization tripartite molecules are formed. **C**: Schematic horizontal placement of the tripartite molecules. **D**: Cartridge with its input and output wells per lane. **E**: Graphic representation of a cartridge's lane. Each lane has an input and an output well with a binding surface in between. The samples are distributed in the middle segment of the lane or binding surface. The squares in the middle segment represent the possible regions or fields of view (FOV) that each lane possesses. There are a total of 555 FOV per lane and there are 4 colour channels per region, generating a maximum of 2200 images per lane, or 26400 images per cartridge. Image source: NanoString.

After the scanning, the DA produces a .csv file after the scanning that can be processed either with nCounter's nSolver software or with a standard spreadsheet program. Each output file contains the rough counts of each scan, including the spiked-in positive and negative controls.

Custom Code Set composition and bioinformatics methodology

A few further remarks are made in the following lines to deepen the understanding of the bioinformatics procedure of the initial Code Set construction.

The 48 genes that assemble our custom Code Set are listed in table 1.3 and were selected from existing signatures [3-6]. A selection of a subset of genes was performed for either the mBL or the GCB-ABC classifier with the Least-Angle Regression (LARS) [2] method. LARS is a statistical algorithm that performs variable selection by fitting linear regression models on a given set of predictors. In essence, it is based on sequentially adding the most correlated variables to build a model.

Code Set gene	Accession number	Code Set gene	Accession number
BATF	NM_006399.3	*LMO2*	NM_005574.3
BCL2	NM_000657.2	*LRMP*	NM_006152.2
BCL2A1	NM_004049.2	*MME*	NM_000902.2
BCL6	NM_138931.1	*MMP2*	NM_004530.2
BMP7	NM_001719.1	*MYBL1*	NM_001080416.2
C3orf37	NM_020187.2	*MYC*	NM_002467.3
CCND2	NM_001759.3	*NASP*	NM_172164.1
CCR7	NM_001838.2	*Pecam1 (CD31)*	NM_000442.3
CD44	NM_001001392.1	*PIM1*	NM_002648.2
CEBPA	NM_004364.2	*PRDM10*	NM_020228.2
CPNE3	NM_003909.2	*PRKAR2B*	NM_002736.2
CTSH	NM_148979.2	*RAB7L1*	NM_001135664.1
DBI	NM_001079862.1	*RNASEH2B*	NM_001142279.1
EML2	NM_012155.2	*SCYA3 (CCL3)*	NM_002983.2
ENTPD1	NM_001098175.1	*SH3BP5*	NM_001018009.2
EPHB3	NM_004443.3	*SKAP2*	NM_003930.3
FN1	NM_212482.1	*SMARCA4*	NM_003072.3
GUSB	NM_000181.1	*STAT3*	NM_139276.2
IGHM	ENST00000390559.1	*TBXA2R*	NM_001060.3
IGLJ3	ENST00000390324.1	*TCF3*	NM_003200.2
IL16	NM_004513.4	*TEK*	NM_000459.2
INE1	NR_024616.1	*THBS1*	NM_003246.2
IRF4	NM_002460.1	*TNFRSF1B*	NM_001066.2
ITPKB	NM_002221.3	*TPD52*	NM_001025252.1

Table 1.3: Custom Code Set composition with a total of 48 genes, including mBL- and DLBCL-target genes, and endogenous controls.

Samples from Hummel et al. [3] that had been previously classified with the traditional microarrays were assigned an individual index (single numeric value). The index value represents the distance from such sample to molecular label cut-off. The limit or cut-off is specifically a 73-dimensional hyperplane, in our case, separating the mBL from the non-mBL. The larger a sample's index is, the further it is from the class-separating hyperplane and thus the more defined its molecular diagnosis is. In the mBL gene selection procedure, 73 probeIDs from Hummel et al., [3] were used to run in the LARS algorithm. In a first step, a LARS model was learned on a training cohort, consisting of a subset

(n=113 samples; 25 mBL, 26 intermediate and 62 non-mBLs) from Hummel et al., generated with the old Affymetrix arrays. This model aims to relearn the index for each sample with subsets of the initial 73 probeIDs only. Second, the model was used to predict the individual mBL indexes on the full test cohort from Hummel et al. (n=220; 44, 48 and 128 mBL, intermediate and non-mBLs respectively). Prediction was performed by applying the LARS model from the training stage with probeID subsets of different sizes starting with 5 and ending with 20 genes. At each iteration step, an individual index per sample was predicted and the maximum distance between mBL and non-mBL samples determined respectively. The gene set that resulted in the maximum overall distance was composed of 5 genes for mBL/non-mBL classification.

The GCB-ABC gene selection was done similarly: Twenty-three probeIDs from the same previous publication [3] were used to calculate the individual sample GCB-index on a training set of 92 samples (63 GCBs and 29 ABCs). Following, a test set comprising a total of 246 GCB and ABCs samples were run on a prediction response test using only 10/23 predictor genes. Care was taken to minimise the misclassification between GCB and ABCs.

Ultimately the gene selection from our initial 48-gene Code Set comprised 10 GCB-ABC and 5 mBL genes. In addition, for each of these 10 GCB-ABC and 5 mBL genes a further backup gene was chosen within our Code Set that maximised correlation to its "partner" gene (see table 1.1). Backup genes are useful if data quality of the primary gene is not sufficient, making the analysis more stable.

Two classifiers (from FFPE and frozen materials respectively) generated on the Nanostring platform were learned and are referred to in sections 1.2 and 1.3. The classifier learning procedure is described in section 1.3.1. Figure 1.6 schematizes the two cohorts used for such purposes.

1.4.2 Results and Discussion

Array-nCounter labelling comparison

The comparison between our FFPE RNA-derived data and cryo RNA-derived data in Affymetrix gave a Spearman correlation (rank based) of -0.82 and a total of 2/40 (5%) samples with discrepant clustering (figure 1.9). One of the two cases had been classified previously by the array technology as intermediate but as an mBL by ours. The second misclassified case showed the opposite situation. Both classification shifts were minimal (see heat map in figure 1.5, A), involving neighbouring subtypes. Moreover the two cases carried a t(8;14), which is the hallmark translocation for BL, demonstrating that they rightfully belonged to the group of mBL and intermediates (right side of the heat map).

As well as the mBL classifier, the new FFPE-derived GCB-ABC labelling system was juxtaposed with the gold standard classifier. Figure 1.10 plots 31 FFPE samples and their cryo-preserved counterparts, analysed with our own custom nCounter Code Set and in the Affymetrix array respectively. The rank based Spearman correlation reached 0.88, with 5/31 (16%) of misclassifications. Again, as in the mBL classifier, the different labelling was minor, affecting adjacent DLBCL subtypes. Three of the 5 lymphomas had been classified as ABC subtype by the array and our algorithm gave them the unclassified label. The remaining two cases had been one classified as GCB in the array and as unclassified with our Code Set, and the last case vice versa.

Twenty-three samples of our 31 FFPE-cryo cohort had been also analysed according to the Hans classifier [7] (see figure 1.5, B). The array fresh-frozen samples had 5/23 (22%) misclassifications to the Hans IHC system, containing 1 case from ABC to GCB, two cases from unclassified to non-GCB and two cases from GCB to non-GCB. Three major misclassifications (between non-GCB -or ABC- and GCB) are present. In comparison, in the nCounter-Hans comparison, one can find 7/23 (30%) divergences, being 2 of a major character (GCB in nCounter, and non-GCB in Hans).

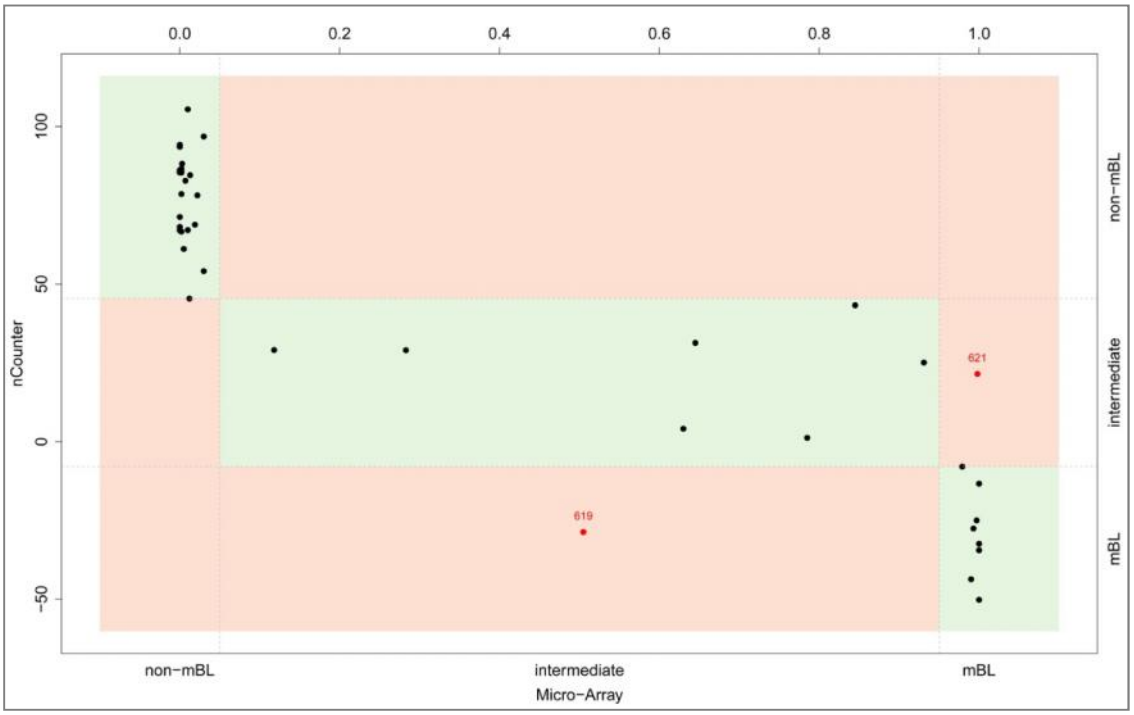


Figure 1.9: mBL-index comparison of 40 FFPE-originated samples analysed in nCounter (x axis) and their counterpart cryo RNA-derived samples analysed in Affymetrix (y axis, as microarray in the graph). Cut-offs for each classifier label are shown (salmon to green colour change).

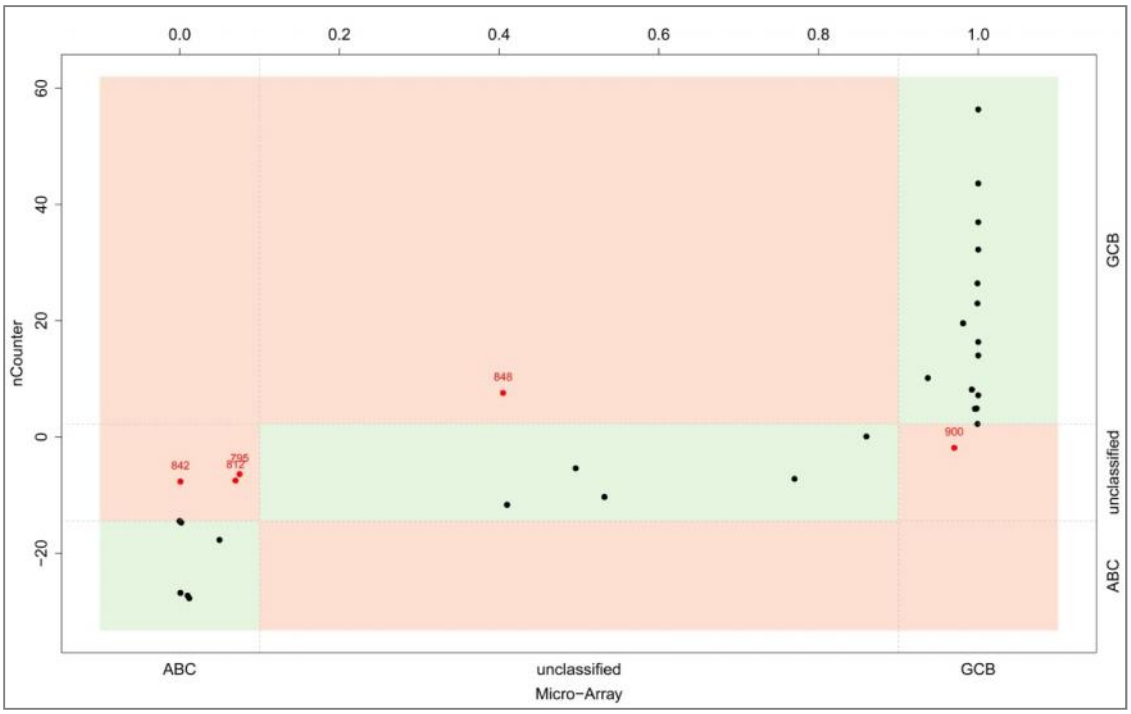


Figure 1.10: GCB/ABC classifier sample clustering representation. Minor misclassifications between nCounter and array data are indicated with red dots. Cut-offs for each classifier label are shown.

By comparing the discrepancies it is clear that greater differences are found comparing the GEP classifiers to the IHC approach than by comparing the GEP classifiers to each other. Between Affymetrix and nCounter results no major misclassifications were generated, showing a robust differentiation of the mBL and DLBCL subtypes.

Several groups have established protocols for GEP on FFPE tissues using different technologies including microarrays [7], whole genome cDNA-mediated Annealing, Selection, 8extension, and Ligation (DASL) assay [9] or ribonuclease protection assays [10]. The technology used here provides several advantages in pathological practice. It does not require reverse transcription and subsequent amplification of the nucleic acids, reducing labour and simplifying standardization and automation. We further showed that a 30 gene assay is sufficient to define mBL and distinguish DLBCL subtypes. In summary, our study demonstrates that a GEP-based molecular classification of mature aggressive B-cell lymphomas is robust on FFPE-derived RNA.

References to section 1.4

[1] G.K. Geiss, R.E. Bumgarner, B. Birditt, T. Dahl, N. Dowidar, D.L. Dunaway, H.P. Fell, S. Ferree, R.D. George, T. Grogan, J.J. James, M. Maysuria, J.D. Mitton, P. Oliveri, J.L. Osborn, T. Peng, A.L. Ratcliffe, P.J. Webster, E.H. Davidson, L. Hood, K. Dimitrov. Direct multiplexed measurement of gene expression with color-coded probe pairs. Nature biotechnology 26 (2008) 317–325.

[2] Efron B, Hastie T, Johnstone I and Tibshirani R. Least angle regression. Annals of Statistics 32 (2004) 407–499.

[3] Hummel Michael, Bentink Stefan, Berger Hilmar, Klapper Wolfram, Wessendorf Swen, Barth Thomas F.E., Bernd Heinz-Wolfram, Cogliatti Sergio B., Dierlamm Judith, Feller Alfred C., Hansmann Martin-Leo, Haralambieva Eugenia, Harder Lana, Hasenclever Dirk, Kühn Michael, Lenze Dido, Lichter Peter, Martin-Subero Jose Ignacio, Möller Peter, Müller-Hermelink Hans-Konrad, Ott German, Parwaresch Reza M., Pott Christiane, Rosenwald Andreas, Rosolowski Maciej, Schwaenen Carsten, Stürzenhofecker Benjamin, Szczepanowski Monika, Trautmann Heiko, Wacker Hans-Heinrich, Spang Rainer, Loeffler Markus, Trümper Lorenz, Stein Harald, Siebert Reiner. A Biologic Definition of Burkitt's Lymphoma from Transcriptional and Genomic Profiling. N Engl J Med 354 (2006) 2419–2430.

[4] I.S. Lossos, D.K. Czerwinski, A.A. Alizadeh, M.A. Wechser, R. Tibshirani, D. Botstein, R. Levy. Prediction of survival in diffuse large-B-cell lymphoma based on the expression of six genes. The New England journal of medicine 350 (2004) 1828–1837.

[5] G. Lenz, G. Wright, S.S. Dave, W. Xiao, J. Powell, H. Zhao, W. Xu, B. Tan, N. Goldschmidt, J. Iqbal, J. Vose, M. Bast, K. Fu, D.D. Weisenburger, T.C. Greiner, J.O. Armitage, A. Kyle, L. May, R.D. Gascoyne, J.M. Connors, G. Troen, H. Holte, S. Kvaloy, D. Dierickx, G. Verhoef, J. Delabie, E.B. Smeland, P. Jares, A. Martinez, A. Lopez-Guillermo, E. Montserrat, E. Campo, R.M. Braziel, T.P. Miller, L.M. Rimsza, J.R. Cook, B. Pohlman, J. Sweetenham, R.R. Tubbs, R.I. Fisher, E. Hartmann, A. Rosenwald, G. Ott, H.-K. Muller-Hermelink, D. Wrench, T.A. Lister, E.S. Jaffe, W.H. Wilson, W.C. Chan, L.M. Staudt. Stromal gene signatures in large-B-cell lymphomas. The New England journal of medicine 359 (2008) 2313–2323.

[6] Dave Sandeep S. M.D. Fu Kai M.D., Ph.D. Wright George W. Ph.D. Lam Lloyd T. Ph.D. Kluin Philip M.D. Boerma Evert-Jan B.S. Greiner Timothy C. M.D. Weisenburger Dennis D. M.D. Rosenwald Andreas M.D. Ott German M.D. Müller-Hermelink Hans-Konrad M.D. Gascoyne Randy D. M.D. Delabie Jan M.D. Rimsza Lisa M. M.D. Braziel Rita M. M.D. Grogan Thomas M. M.D. Campo Elias M.D. Jaffe Elaine S. M.D. Dave Bhavana J. Ph.D. Sanger Warren Ph.D. Bast Martin B.S. Vose Julie M. M.D. Armitage James O. M.D. Connors Joseph M. M.D. Smeland Erlend B. M.D., Ph.D. Kvaloy Stein M.D., Ph.D. Holte Harald M.D., Ph.D. Fisher Richard I. M.D. Miller Thomas P. M.D. Montserrat Emilio M.D. Wilson Wyndham H. M.D., Ph.D. Bahl Manisha B.S. Zhao Hong M.S. Yang Liming Ph.D. Powell John M.S. Simon Richard D.Sc. Chan Wing C. M.D. Staudt Louis M. M.D., Ph.D. Molecular Diagnosis of Burkitt's Lymphoma. N Engl J Med (2006).

[7] C.P. Hans, D.D. Weisenburger, T.C. Greiner, R.D. Gascoyne, J. Delabie, G. Ott, H.K. Müller-Hermelink, E. Campo, R.M. Braziel, E.S. Jaffe, Z. Pan, P. Farinha, L.M. Smith, B. Falini, A.H. Banham, A. Rosenwald, L.M. Staudt, J.M. Connors, J.O. Armitage, W.C. Chan. Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. Blood 103 (2004) 275–282.

[8] P.M. Williams, R. Li, N.A. Johnson, G. Wright, J.-D. Heath, R.D. Gascoyne. A novel method of amplification of FFPET-derived RNA enables accurate disease classification with microarrays. The Journal of molecular diagnostics JMD 12 (2010) 680–686.

[9] S.L. Barrans, S. Crouch, M.A. Care, L. Worrillow, A. Smith, R. Patmore, D.R. Westhead, R. Tooze, E. Roman, A.S. Jack. 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 159 (2012) 441–453.

[10] L.M. Rimsza, G. Wright, M. Schwartz, W.C. Chan, E.S. Jaffe, R.D. Gascoyne, E. Campo, A. Rosenwald, G. Ott, J.R. Cook, R.R. Tubbs, R.M. Braziel, J. Delabie, T.P. Miller, L.M. Staudt. 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. Clinical cancer research an official journal of the American Association for Cancer Research 17 (2011) 3727–3732.

Chapter 2

Clinical and pathological features of Burkitt lymphoma showing expression of BCL2 – an analysis including gene expression in formalin-fixed paraffin-embedded tissue

Neus Masqué-Soler [1*], Monika Szczepanowski [1*] Christian W. Kohler [2*], Sietse Aukema[3], Inga Nagel[3], Julia Richter[3], Reiner Siebert[3], Rainer Spang [2], Birgit Burkhardt[4], Wolfram Klapper[1]

[1]Department of Pathology, Haematopathology Section and Lymph Node Registry, University Hospital Schleswig-Holstein, Campus Kiel/Christian-Albrecht University, Kiel, Germany.
[2] Institute of Functional Genomics, University of Regensburg, Regensburg, Germany.
[3] Institute of Human Genetics. University Hospital Schleswig-Holstein, Campus Kiel / Christian-Albrecht University, Kiel, Germany

[4] Pediatric Hematology and Oncology, University Children's Hospital, Münster, Germany

*These authors contributed equally

Running title: Burkitt lymphoma showing expression of BCL2

This piece was published online in the *British Journal of Haematology*, on the 27th July of 2015.

Summary

The differential diagnosis between Burkitt lymphoma (BL) and diffuse large B-cell lymphoma (DLBCL) can be challenging. BL has been reported to express less BCL2 than DLBCL, but this issue has not been analysed systematically. BL expressing BCL2 can be considered to be MYC/BCL2 co-expressors, a feature that is associated with poorer outcome in DLBCL but that has not been correlated with outcome in BL so far. We analysed the expression of BCL2 in 150 cases of conventionally diagnosed BL using two different BCL2 antibodies. BCL2 expression was detected in 23% of the cases, though the expression varied in intensity and number of positive cells. Using a subgroup of 43 cases for which detailed clinical data were available, we did not detect any relevant differences in clinical presentation and outcome between BCL2-positive and BCL2-negative BL. An independent cohort of 17 BL with expression of BCL2 were analysed molecularly with 13 of 17 cases classified as molecularly defined BL using gene expression profiling on formalin-fixed paraffin-embedded tissues. The four lymphomas diagnosed molecularly as intermediates did not differ in clinical presentation and outcome from molecularly defined BL.

Keywords: Nanostring, nCounter, classifier, Burkitt, diffuse large B-cell lymphoma, MYC

2.1 Introduction

The differential diagnosis of Burkitt lymphoma (BL) versus diffuse large B-cell lymphoma (DLBCL) can be challenging. In daily practice their diagnosis is based on morphological assessment, immunophenotype and detection of translocations of the *MYC* and absence of translocations involving the *BCL2* or *BCL6* gene by fluorescence in situ hybridization (FISH) (conventionally diagnosed BL). The immunophenotype, which is assessed to distinguish BL from DLBCL, frequently includes BCL2 expression [1–3]. The former WHO classification published in 2001 stated that BCL2 is not expressed in BL [4]. However, gene expression profiling identified cases of molecularly defined BL (mBL) with BCL2 expression not caused by translocations of the gene [5,6]. These findings led to a new weighting of the immunophenotypic features of BL in the current WHO classification, which describes BCL2 as being expressed in up to 20% of BL cases, predominantly with a weak expression pattern [7].

BL expressing BCL2 must be distinguished from the group of lymphomas called "B-cell lymphomas unclassified with features intermediate between BL and DLBCL" (conventionally diagnosed intermediates, conventionally diagnosed intermediates), which comprises a heterogeneous group of lymphomas. Conventionally diagnosed intermediate lymphomas show at least some features reminiscent of BL, such as a monomorphic cytological picture with small to medium-sized blasts, a starry sky pattern, high proliferation and frequently translocations involving the *MYC* gene [7]. However, conventionally diagnosed intermediate lymphomas often display immunophenotypic features such as strong BCL2 expression or genetic features like *BCL2* or *BCL6* gene translocations that preclude a BL diagnosis. The presence of both *MYC* and *BCL2* or *BCL6* translocations is called "double hit" and is usually not compatible with the diagnosis of BL [8–10].

Intermediate lymphomas can also be identified by gene expression profiling independent of morphology, immunophenotype and FISH data. Gene expression profiling-based molecular intermediates (m-intermediates) are classified as conventionally diagnosed BL or DLBCL by conventional diagnostics. A subgroup of m-intermediates lack a genetic "double hit" constellation [6]. In children and adolescents translocations of *BCL6* and *BCL2*, and thus "double hit" lymphomas, are exceedingly rare, but m-intermediates in the young age group can be detected by gene expression profiling [11]. Similar to "double hit" lymphomas, m-intermediates identified by gene expression profiling occasionally express BCL2 protein [11]. Currently, gene expression profiling is not available for a routine diagnosis, thus for diagnostic pathologists it is currently unclear whether strong expression of BCL2 justifies the classification of a case with the morphological and immunophenotypical features of a BL as an intermediate lymphoma.

In DLBCL not only the genetic "double hit" with *MYC* and *BCL2* translocations but also the co-expression of both proteins MYC and BCL2 ("co-expressor") characterize

a subgroup of DLBCL with an aggressive course [12–15]. In BL, high levels of MYC protein expression is due to the *MYC* translocations [6]. Whether the co-expression with BCL2 indicates inferior poorer outcome in BL has not yet been studied.

To gain a better understanding of the clinical implications, we systematically analysed BCL2 expression in BL. To this end we screened a large cohort of conventionally diagnosed BL for BCL2 expression using two antibodies that detect BCL2. We correlated the BCL2 expression status with the clinical course in cases with available clinical data. Moreover, we applied digital multiplexed gene expression profiling on formalin-fixed paraffin-embedded (FFPE) tissue specimens from BCL2-positive BCL2-negative conventionally diagnosed BL cases to determine whether they are classified as mBL.

2.2 Materials and methods

2.2.1 Patients and tissue specimens

BL cases from all age groups that were diagnosed between 2001 and 2014 were identified within the files of the Lymph Node Registry Kiel. Cases with sufficient FFPE tissue were arranged in a tissue microarray (TMA) with duplicate cores 0.6 mm in diameter for each lymphoma. We included only lymphomas with morphological features compatible with the diagnosis of conventionally diagnosed BL as judged on full tissue sections. The lymphomas were composed of dense sheets of monomorphic small to medium-sized blasts with a narrow rim of cytoplasm showing cohesive growth and a starry sky pattern in areas of good tissue preservation. The TMA cohort comprised a total of 150 cases of conventionally diagnosed BL from all age groups. A subgroup of cases with specimens of sufficient size and a tumour content of at least 70% were selected for molecular analysis (n=17). Clinical data were available for 43 cases.

2.2.2 Immunohistochemistry and fluorescence in situ hybridization (FISH)

Analysis of BCL2 expression was conducted using clone 100/D5 (DAKO, Glostrup, Denmark) or clone E17 (Zytomed, Berlin, Germany)(Adam et al, 2013). Analysis of CD20, CD10 and Ki67 was performed using an automated stainer (Leica, Wetzlar, Germany). BCL2 immunohistochemistry results were scored by visual inspection in the following categories: negative (no staining in lymphoma cells), 1-25% positive lymphoma cells, 26-50% positive lymphoma cells, 51-75% positive lymphoma cells and >75% positive lymphoma cells. The intensity of BCL2 staining was analysed relative to non-neoplastic T-cells serving as an internal control as weak (weaker than T-cells) or strong (stronger or equal to T-cells). Staining intensity of T-cells also served as a control for immunoreactivity of the tissue. The Ki67 index was assessed in 5% steps, whereas CD20, CD10 and TdT were reported as positive or negative. FISH was conducted as described recently [16].

Break-apart probes for the *MYC* , BCL2 and *BCL6* loci as well as IGH-MYC and IGH-BCL2 fusion probes were applied (Abbott, Abbott Park, Illinois, USA) [17].

2.2.3 RNA extraction and digital multiplexed gene expression

FFPE material was cut into five 10 µm-thick sections per sample. Extraction of RNA was done according to the manufacturer's instructions (ExpressArt FFPE Clear RNAready Kit, AmpTec, Hamburg, Germany), quantified and checked for quality (Agilent RNA 6000 Nano Chips, Agilent Technologies, Santa Clara, California, USA) as previously described [18]. Gene expression analysis was performed using nCounter/Nanostring technology following the previously published protocol [18]. In five specimens with previously visualized BCL2 expression pattern on FFPE sections, BCL2-positive and -negative areas were analogously labelled on unstained tissue sections, manually dissected and separately processed for RNA extraction and gene expression.

2.3 Results

2.3.1 Incidence of BCL2 expression in conventionally diagnosed BL

We analysed 150 cases of conventionally diagnosed BL diagnosed by morphological and immunophenotypical features as well as the presence of MYC translocations. All cases were positive for CD20, CD10 and negative for TdT with a Ki67 index above 90%. All lymphomas were shown by fluorescence in situ hybridization to harbour *MYC* (150/150, 100%) translocations but to lack *BCL2* (0/141, 0%) or *BCL6* (0/137, 0%) translocations. TMAs containing our case cohort were stained by two antibodies directed against BCL2. Applying the most widely used antibody clone 100/D5, 146 cases yielded an interpretable result. No expression was detected in 113/146 cases (77%). BCL2 expression was detected at various levels in 33/146 (23%) of the cases with interpretable results (1-25% positive lymphoma cells in 16/146 (11%), 26-50% in 6/146 (4%), 51-75% in 6/146 (4%) and >75% in 5/146 (3%)). The expression intensity was weak (weaker than non-neoplastic T-cells as internal controls) in the majority of cases (Fig 1) and only two cases showed strong staining intensity equal to or stronger than non-neoplastic T-cells (data not shown).

To analyse whether BCL2 protein expression might escape detection by the frequently used antibody clone 100/D5, as reported for follicular lymphoma [19], the cases were additionally analysed by the antibody clone E17. However, none of the conventionally diagnosed BL that lacked BCL2 expression using the clone 100/D5 stained positive with the clone E17 (0/104 of cases with interpretable results, Table 2.3, supplement).

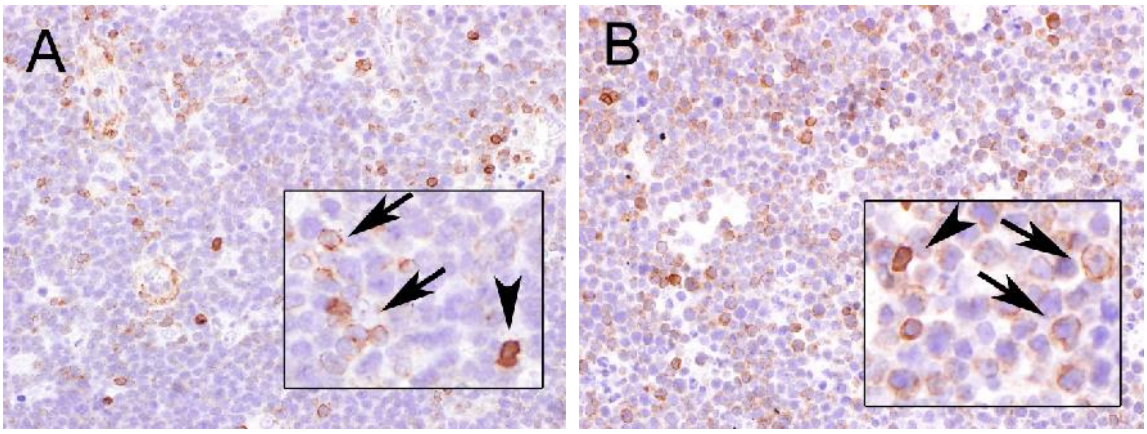


Figure 2.1: Variable expression of BCL2 in BL. Scattered cells comprising about 10% of all lymphoma cells (A, corresponding to Case 9 in Table 2.1) and diffusely dispersed BCL2-positive lymphoma cells comprising about 40% (B, corresponding to Case 12 in Table I). Inserts with high magnifications with arrowheads indicating non-neoplastic T-cells and arrows indicating BCL2-positive lymphoma cells. Original magnification 400x.

2.3.2 Pathological features of conventionally diagnosed BL showing expression of BCL2

A subgroup of 17 patients with conventionally diagnosed BL was further studied on full tissue sections (7 of these patients were also part of the TMA cohort). The patients' clinical features were typical for BL and they were predominantly young (median age 11 years, range 3-24 years) and male (Table 2.1). All but one patient presented with lymphomas with no or limited bone marrow infiltration and they were frequently at a limited stage (Table 2.1). The lymphomas manifested predominantly at extranodal sites with 10/17 in the small or large intestine or as an abdominal or ileocecal mass (Table 2.1).

Histopathology and genetics were compatible with the diagnosis of conventionally diagnosed BL, with small to medium-sized blasts with a narrow rim of cytoplasm. A starry sky pattern was detectable in the majority of cases, at least in areas with good tissue preservation. A plasmacytic differentiation was not detectable. All lymphomas displayed the conventionally diagnosed BL immunophenotype with expression of CD20 and CD10, high proliferation as measured by the Ki67 index and absence of TdT expression (Table 2.1). Only 1/8 cases (13%) was positive for EBV by EBER in situ hybridization (data not shown). The genetic features were compatible with the diagnosis conventionally diagnosed BL too (Table 2.1). The BCL2-expressing cells were either diffusely dispersed over the whole lymphoma/tumour or clustered in restricted areas while leaving other areas BCL2-negative (Figure 2.2). In cases with spatially divergent BCL2 expression, the phenotype and the morphology did not differ between BCL2-positive and BCL2-negative areas of the lymphoma (Figure 2.2). Spatially divergent BCL2 expression occurred in 29% of the cases analysed.

Case	Gender	Age at diagnosis (years)	St. Jude Stage	Relapse	Tumour localization	CD20	CD10	TdT	Ki67 (%)	BCL2% + in lesion	t(8;14)	BCL2-break	BCL6-break	molecular diagnosis by gene expression
1	m	12	II	no	Ileocoecal	pos	pos	neg	>90	60% weak	pos	n.d.	n.d.	mBL
2	m	16	II	no	Mandible	pos	pos	neg	100	60% weak	pos	n.d.	neg	mBL
3	Unknown	12	III	no	ileocoecal	pos	pos	neg	100	30% strong	pos	n.d.	neg	mBL
4	m	4	III	no	Appendix	pos	pos	neg	>95	50% weak	pos	n.d.	neg	mBL
5	m	11	n.d.	no	Meckel-diverticulum	pos	pos	neg	100	5% weak	pos	n.d.	neg	intermediate
6	m	3	III	no	colon	pos	pos	neg	95	80% strong	pos	neg	neg	intermediate
7	m	14	III	no	abdomen	pos	pos	neg	n.d.	20% weak	n.d.	n.d.	n.d.	mBL
8	m	3	II	no	small intestine	pos	pos	neg	>90	30% weak	pos	neg	neg	mBL
9	m	3	I or II	no	unknown	pos	pos	neg	>90	10% weak	n.d.	n.d.	n.d.	mBL
10	m	16	III	no	soft tissue	pos	pos	neg	>95	50% weak	pos	neg	neg	mBL
11	m	24	n.d.	n.d.	small intestine	pos	pos	neg	100	70% weak	pos	neg	neg	mBL
12	m	22	n.d.	n.d.	cervical lymph node	pos	pos	neg	>95	40% weak	pos	neg	neg	mBL
13	m	16	III	no	ileocoecal	pos	pos	neg	100	40% weak	pos	neg	neg	intermediate
14	unknown	8	IV	no	maxilla	pos	pos	neg	100	20% weak	neg*	n.d.	neg	mBL
15	m	9	II	no	ileum	pos	pos	neg	100	20% weak	pos	neg	neg	mBL
16	m	6	n.d.	n.d.	cervical lymph node	pos	pos	neg	95	40% weak	pos	neg	neg	intermediate
17	m	10	B-ALL	no	liver	pos	pos	neg	100	90% strong	pos	neg	neg	mBL

Table 2.1: Clinical, pathological and genetic features of lymphoma cases analysed for molecular diagnosis. * = MYC-translocation with an undetermined translocation partner. Pt =patient number, m= male, n.d.=not determined, pos= positive, neg= negative, mBL= molecular Burkitt lymphoma, intermediate= intermediate between Burkitt and diffuse large B-cell lymphoma. Cases 10, 11, 12, 13, 15, 16 and 17 are also part of the TMA cohort.

2.3.3 BCL2-positive conventionally diagnosed BL show a molecular profile of mBL or m-intermediate

In order to understand whether the expression of BCL2 in conventionally diagnosed BL truly is a BL, we performed a gene expression analysis using RNA obtained from FFPE tissues and digital multiplexed gene expression technology, as previously published [18]. Thirteen of 17 conventionally diagnosed BL (76%) were classified as mBL, confirming the clinicopathological-cytogenetic diagnosis (Table 2.1). In five specimens the BCL2-positive and negative areas of the conventionally diagnosed BL were manually dissected and analysed separately for gene expression. All of these were classified as mBL in both the BCL2-negative and the BCL2-positive areas of the tumour (Case 1, 2, 3, 4 and 11 in Table 2.1).

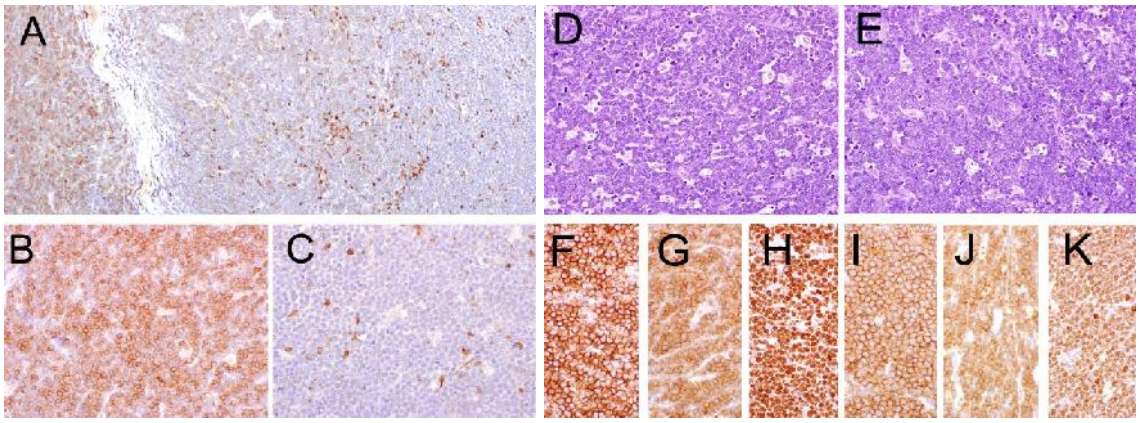


Figure 2.2: Example of a case of BL with differential spatial expression of BCL2. The specimen was obtained from an ileocecal mass in a 12-year-old male child (Case 1 in Table I). The tumour displays heterogeneous expression of BCL2 (A, overview; B and C high magnification). Both areas with and without BCL2 expression show the typical morphology and immunophenotype of BL. BCL2-positive areas: B, BCL2; D, haematoxylin and eosin; F, CD20; G, CD10; H, Ki67. BCL2-negative areas: C, BCL2; E, haematoxylin and eosin; I, CD20; J, CD10; K, Ki67. Original magnification 100x in A and 400x in B-K.

Interestingly, 4/17 conventionally diagnosed BL (24%) were classified as lymphomas intermediate between mBL and non-mBL (m-intermediate, Case 5, 6, 13 and 16 in Table 2.1, Figure 2.3). All m-intermediate lymphomas showed typical features of conventionally diagnosed BL. The patients were under the age of 16 and 3 patients presented with an abdominal mass (Table 2.1). Although the morphology of these lymphomas was ambiguous due to poor tissue quality (Figure 2.3), all lymphomas were positive for t(8;14) and lacked *BCL6* breaks. There were no *BCL2* breaks in two cases with available data. The patients were treated according to the protocols for BL and did not suffer from a relapse.

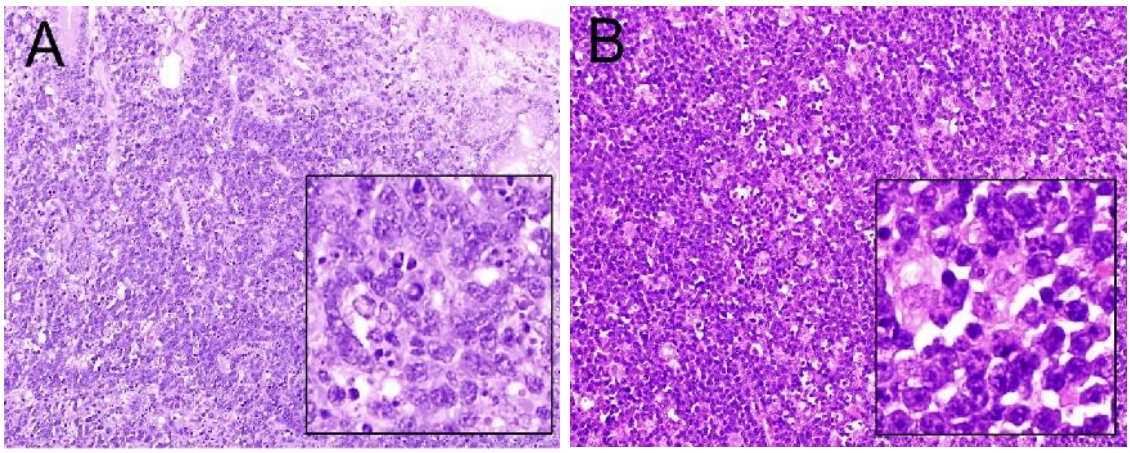


Figure 2.3: Two lymphomas diagnosed as BL by means of morphology, immunophenotype and genetics that were classified as intermediates by gene expression profiling. Panel A corresponds to Case 5 in Table I; panel B corresponds to Case 16 in Table I. All panels are stained with haematoxylin and eosin. Original magnification 200x.

2.3.4 BCL2 expression in paediatric conventionally diagnosed BL is not associated with unfavourable outcome

In DLBCL co-expression of BCL2 and MYC is associated with poorer outcome. In our cohort of 13 mBL based on gene expression profiles and available clinical data the outcome is excellent and none of the patients for whom a molecular diagnosis was available in the current study showed a relapse (Table 2.1). We thus analysed a larger number of conventionally diagnosed BL treated according to protocols of the NHL-BFM study group. Forty three patients, for whom the BCL2 expression status based on whole tissue section staining was available (including three patients that were analysed by gene expression profiling), were registered in the NHL-BFM database, thus providing us with complete clinical data. Seventeen of 43 cases of conventionally diagnosed BL (40%) were reported to show BCL2 expression. There was no major difference in clinical variables between the group with and the group without known BCL2 expression status, with the exception that a lower number of cases with known BCL2-expression status had bone marrow involvement and elevated lactate dehydrogenase (LDH) levels (Supplementary Table 2.4,). A comparison of the clinical variables, gender, age, stage, bone marrow involvement, central nervous system involvement, LDH, B-symptoms, risk groups and event-free survival, did not reveal any difference between 26 BCL2-negative and 17 BCL2-positive cases of conventionally diagnosed BL (Table 2.2).

Characteristic	BCL2 negative (n = 26)		BCL2 positive (n = 17)		P value (Chi)
Gender					
Male	24	92%	15	88%	
Female	2	8%	2	12%	1·00
Age (years)					
<10	13	50%	9	53%	
10–14	9	35%	3	18%	
>14	4	15%	5	29%	0·39
Stage of disease					
I	1	5%	0	0%	
II	5	24%	5	33%	
III	12	57%	8	53%	
IV	1	5%	0	0%	
B-AL	2	10%	2	13%	0·96
BM involvement					
Yes	2	8%	2	12%	1·00
CNS involvement					
Yes	2	8%	1	6%	1·00
LDH (iu/l)					
<500	15	63%	10	59%	
500–1000	6	25%	3	18%	
>1000	3	13%	4	24%	0·74
B symptoms					
Yes	8	31%	3	18%	0·48
Risk group					
R1	0	0%	0	0%	
R2	13	62%	9	60%	
R3	3	14%	2	13%	
R4	5	24%	4	27%	1·00
Event-free survival	96 ± 4%		82 ± 9%		0·15 LR

Table 2.2: Characteristics of BCL2-positive and –negative BL patients followed by the BFM-NHL study group data. B-AL, Burkitt-acute leukaemia; BM, bone marrow; CNS, central nervous system; LDH, lactate dehydrogenase; LR, likelihood ratio.

2.4 Discussion

BCL2 is a member of the BCL2 family of proteins, which are involved in a variety of cellular processes and predominantly act in an anti-apoptotic manner [20]. BCL2 is a promising therapeutic target in lymphomas that express the protein [21]. In this study we show that BCL2 is expressed in a considerable number of conventionally diagnosed BL cases. The percentage of BCL2-positive conventionally diagnosed BL cases in the NHL-BFM database (39%) was higher than the percentage of conventionally diagnosed BL

reported as positive in our tissue microarray study (23%). The difference might be due to the fact that full slides were analysed for BCL2 expression for the NHL-BFM data, which allows areas of local BCL2 reactivity to be detected that might escape identification using a TMA. Translocations involving *BCL2* as a cause of BCL2 expression were ruled out in our study, confirming previous data [5,11,22]. Whether other genetic aberrations, such as mutations in the promoter region or epigenetic mechanisms in BCL2-positive subclones of conventionally diagnosed BL, cause the protein's expression has not yet been elucidated. It is worth mentioning that the widely used antibody clone 100/D5 (DAKO, Glostrup, Denmark) is sufficiently sensitive to identify BCL2 expression in conventionally diagnosed BL. In contrast to follicular lymphoma [19], mutations of BCL2 epitopes do not seem to influence immunoreactivity despite the fact that mutations in *BCL2* occasionally occur in conventionally diagnosed BL [23]. It is important to mention that BCL2 staining suffers from a high inter-observer variability [24].

The expression of BCL2 is associated with poorer survival in DLBCL, especially if co-expressed with MYC protein [14,15]. As all BL cases express high levels of MYC [5,22], the BCL2-positive BL are MYC and BCL2 co-expressors. Our findings show that, in conventionally diagnosed BL or mBL, co-expression of MYC and BCL2 is not associated with a poorer outcome. Certainly, the differences in therapeutic protocols might explain the different prognostic impact of MYC-BCL2 co-expression. All conventionally diagnosed BL and mBL cases in our study were paediatric patients who received intensive chemotherapy according to paediatric protocols. The DLBCL for which the negative prognostic impact of MYC-BCL2 co-expression has been demonstrated were mainly treated with less aggressive immunochemotherapy using a CHOP [cyclophosphamide, hydroxydaunorubicin (doxorubicin), vincristine, prednisolone]-based regimen [25]. Whether intensified chemotherapy might overcome the negative prognostic impact of MYC-BCL2 coexpression in DLBCL is still uncertain. It is noteworthy that a recent study of a young, high-risk cohort of DLBCL treated with intensified immunochemotherapy identified biological risk factors that differ from cohorts of elderly patients with DLBCL [26].

Using our recently developed gene expression-based classifier [18], we were able to apply Nanostring technology to manually dissected areas of the lymphoma. BCL2-positive and BCL2-negative areas of the same lymphomas were independently and consistently classified as mBL, implying that the expression of BCL2 does not indicate a 'transformation' of BL towards an intermediate lymphoma. To our knowledge, this is the first example of gene expression-based molecular diagnosis in independently analysed areas of a lymphoma. The consistency of the results confirms the reliability of our multiplex assay and suggests that the molecular diagnosis based on our classifier is not influenced by intratumoural heterogeneity.

We have previously shown that m-intermediates in children do not harbour a genetic 'double hit' constellation and show more 'Burkitt-ness' than adult m-intermediates [2,11]. This suggests that the majority of m-intermediates in children and young adults are, in fact, BL, which escape classification as mBL due to minor divergence of gene expression or simply the choice of the statistical cut-off in the current algorithms [11]. This study again identified four m-intermediates among 17 lymphomas that were classified as conventionally diagnosed BL on the basis of morphology and immunophenotype. However, the clinical, pathological and genetic features of the m-intermediates, such as age, localization, site of the lymphoma, morphology, immunophenotype, cytogenetics and outcome, do not differ from mBL in this cohort of children and young adults. This finding raises questions as to the potential advantage of molecular over conventional diagnosis for mature aggressive B-cell lymphomas in children and adolescents. Obviously, m-intermediates in young patients do not present a clinically or biologically relevant subgroup. As our cohort of molecularly classified lymphomas is rather small, molecular diagnosis based on gene expression profiling needs to be evaluated in larger cohorts of lymphomas in children and young adults to understand its value as a diagnostic and prognostic tool.

Our data suggest that BCL2 expression is a feature detected to a variable extent in a considerable number of cases of BL. It is important to mention, that even a high expression intensity or a high number of BCL2 positive lymphoma cells does not preclude the diagnosis of BL if all other features are compatible with this diagnosis, especially if BCL2 translocations are absent. BCL2 expression in a lymphoma that clinically, morphologically, immunophenotypically and genetically resembles BL is not sufficient to diagnose an intermediate lymphoma, especially if the BCL2 expression is developed only in a subgroup of lymphoma cells (as was the case in most of the specimens analysed in this study).

In summary, more than 20% of lymphomas with the clinical, morphological, immunophenotypical and genetic features of conventionally diagnosed BL express BCL2. This expression is variable in extent and can occur diffusely or be restricted to certain areas of the lymphoma. BCL2 expression in conventionally diagnosed BL is almost always weaker than in non-neoplastic T-cells. Molecular diagnosis of BCL2-positive conventionally diagnosed BL confirms the diagnosis in the vast majority of cases. The clinical significance of m-intermediates in children is uncertain at the current time. BCL2 expression in conventionally diagnosed BL is not associated with poorer outcome in paediatric patients. Diagnostic pathologists must be aware of the BCL2 expression pattern in conventionally diagnosed BL.

Acknowledgements

The authors would like to thank Olivera Batic, Dana Germer, Charlotte Botz-von Drathen, Claudia Becher, Reina Zühlke-Jenisch and Dorit Schuster for their excellent technical support and Kay Dege for editing the manuscript.

Support

This work was supported by a grant from the German Cancer Aid (1090547), the KinderKrebsInitiative (KKI) Buchholz/Holm-Seppensen, an intramural grant of the Medical Faculty of the University of Kiel to WK (grant number F343911), and the e:MED project MMML-MYC-SYS funded by the German Ministry of Science and Education (BMBF grant number 0316166).

Conflict of interest

The authors have no conflict of interest to disclose.

References

[1] S.B. Cogliatti, U. Novak, S. Henz, U. Schmid, P. Möller, T.F.E. Barth. Diagnosis of Burkitt lymphoma in due time: a practical approach. British Journal of Haematology 134 (2006) 294–301.

[2] I. Salaverria, R. Siebert. The gray zone between Burkitt's lymphoma and diffuse large B-cell lymphoma from a genetics perspective. Journal of clinical oncology official journal of the American Society of Clinical Oncology 29 (2011) 1835–1843.

[3] P. Kluin, E. Schuuring. Molecular cytogenetics of lymphoma: where do we stand in 2010? Histopathology 58 (2011) 128–144.

[4] Jaffe,E., Harris N, Stein H, & Vardiman JW. Pathology and genetics of tumours of haematopoetic and lymphoid tissues, IARC Press, Lyon, France, 2001.

[5] Dave Sandeep S. M.D. Fu Kai M.D., Ph.D. Wright George W. Ph.D. Lam Lloyd T. Ph.D. Kluin Philip M.D. Boerma Evert-Jan B.S. Greiner Timothy C. M.D. Weisenburger Dennis D. M.D. Rosenwald Andreas M.D. Ott German M.D. Müller-Hermelink Hans-Konrad M.D. Gascoyne Randy D. M.D. Delabie Jan M.D. Rimsza Lisa M. M.D. Braziel Rita M. M.D. Grogan Thomas M. M.D. Campo Elias M.D. Jaffe Elaine S. M.D. Dave Bhavana J. Ph.D. Sanger Warren Ph.D. Bast Martin B.S. Vose Julie M. M.D. Armitage James O. M.D. Connors Joseph M. M.D. Smeland Erlend B. M.D., Ph.D. Kvaloy Stein M.D., Ph.D. Holte Harald M.D., Ph.D. Fisher Richard I. M.D. Miller Thomas P. M.D. Montserrat Emilio M.D. Wilson Wyndham H. M.D., Ph.D. Bahl Manisha B.S. Zhao Hong M.S. Yang Liming Ph.D. Powell John M.S. Simon Richard D.Sc. Chan Wing C. M.D. Staudt Louis M. M.D., Ph.D. Molecular Diagnosis of Burkitt's Lymphoma. N Engl J Med (2006).

[6] Hummel Michael, Bentink Stefan, Berger Hilmar, Klapper Wolfram, Wessendorf Swen, Barth Thomas F.E., Bernd Heinz-Wolfram, Cogliatti Sergio B., Dierlamm Judith, Feller Alfred C., Hansmann Martin-Leo, Haralambieva Eugenia, Harder Lana, Hasenclever Dirk, Kühn Michael, Lenze Dido, Lichter Peter, Martin-Subero Jose Ignacio, Möller Peter, Müller-Hermelink Hans-Konrad, Ott German, Parwaresch Reza M., Pott Christiane, Rosenwald Andreas, Rosolowski Maciej, Schwaenen Carsten, Stürzenhofecker Benjamin, Szczepanowski Monika, Trautmann Heiko, Wacker Hans-Heinrich, Spang Rainer, Loeffler Markus, Trümper Lorenz, Stein Harald, Siebert Reiner. A biological definition of Burkitt lymphoma from transcriptional and genomic profiling: Supplementary Appendix.

[7] Swerdlow SH, Campo E, Harris NL, et al. Classification of Tumours of Haematopoietic and Lymphoid Tissues, IARC, 2008.

[8] Sietse M. Aukema, Reiner Siebert, Ed Schuuring, Gustaaf W. van Imhoff, Hanneke C. Kluin-Nelemans, Evert-Jan Boerma, and Philip M. Kluin. Double-hit B-cell lymphoma: Review. Blood 117 (2011) 2319–2331.

[9] S.M. Aukema, M. Kreuz, C.W. Kohler, M. Rosolowski, D. Hasenclever, M. Hummel, R. Küppers, D. Lenze, G. Ott, C. Pott, J. Richter, A. Rosenwald, M. Szczepanowski, C. Schwaenen, H. Stein, H. Trautmann, S. Wessendorf, L. Trümper, M. Loeffler, R. Spang, P.M. Kluin, W. Klapper, R. Siebert. Biological characterization of adult MYC-translocation-positive mature B-cell lymphomas other than molecular Burkitt lymphoma. Haematologica 99 (2014) 726–735.

[10] EG Boerma, R Siebert, PM Kluin and M Baudis. Translocations involving 8q24 in Burkitt lymphoma and other malignant lymphomas: a historical review of cytogenetics in the light of today's knowledge: R eview. Leukemia 23 (2009) 225–234.

[11] W. Klapper, M. Szczepanowski, B. Burkhardt, H. Berger, M. Rosolowski, S. Bentink, C. Schwaenen, S. Wessendorf, R. Spang, P. Möller, M.L. Hansmann, H.-W. Bernd, G. Ott, M. Hummel, H. Stein, M. Loeffler, L. Trümper, M. Zimmermann, A. Reiter, R. Siebert. Molecular profiling of pediatric mature B-cell lymphoma treated in population-based prospective clinical trials. Blood 112 (2008) 1374–1381.

[12] N.A. Johnson, K.J. Savage, O. Ludkovski, S. Ben-Neriah, R. Woods, C. Steidl, M.J.S. Dyer, R. Siebert, J. Kuruvilla, R. Klasa, J.M. Connors, R.D. Gascoyne, D.E. Horsman. Lymphomas with concurrent BCL2 and MYC translocations: the critical factors associated with survival. Blood 114 (2009) 2273–2279.

[13] T.M. Green, K.H. Young, C. Visco, Z.Y. Xu-Monette, A. Orazi, R.S. Go, O. Nielsen, O.V. Gadeberg, T. Mourits-Andersen, M. Frederiksen, L.M. Pedersen, M.B. Møller. Immunohistochemical double-hit score is a strong predictor of outcome in patients with diffuse large B-cell lymphoma treated with rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone. Journal of clinical oncology official journal of the American Society of Clinical Oncology 30 (2012) 3460–3467.

[14] A.M. Perry, Y. Alvarado-Bernal, J.A. Laurini, L.M. Smith, G.W. Slack, K.L. Tan, L.H. Sehn, K. Fu, P. Aoun, T.C. Greiner, W.C. Chan, P.J. Bierman, R.G. Bociek, J.O. Armitage, J.M. Vose, R.D. Gascoyne, D.D. Weisenburger. MYC and BCL2 protein expression predicts survival in patients with diffuse large B-cell lymphoma treated with rituximab. British Journal of Haematology 165 (2014) 382–391.

[15] H. Horn, M. Ziepert, C. Becher, T.F.E. Barth, H.-W. Bernd, A.C. Feller, W. Klapper, M. Hummel, H. Stein, M.-L. Hansmann, C. Schmelter, P. Möller, S. Cogliatti, M. Pfreundschuh, N. Schmitz, L. Trümper, R. Siebert, M. Loeffler, A. Rosenwald, G. Ott. MYC status in concert with BCL2 and BCL6 expression predicts outcome in diffuse large B-cell lymphoma. Blood 121 (2013) 2253–2263.

[16] R.A. Ventura, J.I. Martin-Subero, M. Jones, J. McParland, S. Gesk, D.Y. Mason, R. Siebert. FISH analysis for the detection of lymphoma-associated chromosomal abnormalities in routine paraffin-embedded tissue. The Journal of molecular diagnostics JMD 8 (2006) 141–151.

[17] T.F.E. Barth, L. Floßbach, H.-W. Bernd, R. Bob, M. Buck, S.B. Cogliatti, A.C. Feller, M.L. Hansmann, S. Hartmann, H. Horn, W. Klapper, D. Kradolfer, T. Mattfeldt, P. Möller, A. Rosenwald, H. Stein, C. Thorns, G. Ott. Ringversuch zum Nachweis genomischer Veränderungen bei Non-Hodgkin-Lymphomen mittels In-situ-Hybridisierung. Der Pathologe 34 (2013) 329–334.

[18] N. Masqué-Soler, M. Szczepanowski, C.W. Kohler, R. Spang, W. Klapper. Molecular classification of mature aggressive B-cell lymphoma using digital multiplexed gene expression on formalin-fixed paraffin-embedded biopsy specimens. Blood 122 (2013) 1985–1986.

[19] P. Adam, R. Baumann, J. Schmidt, S. Bettio, K. Weisel, I. Bonzheim, F. Fend, L. Quintanilla-Martínez. The BCL2 E17 and SP66 antibodies discriminate 2 immunophenotypically and genetically distinct subgroups of conventionally BCL2-"negative" grade 1/2 follicular lymphomas. Human Pathology 44 (2013) 1817–1826.

[20] W.A. Siddiqui, A. Ahad, H. Ahsan. The mystery of BCL2 family: Bcl-2 proteins and apoptosis: an update. Archives of toxicology 89 (2015) 289–317.

[21] N. Johnson-Farley, J. Veliz, S. Bhagavathi, J.R. Bertino. ABT-199, a BH3 mimetic that specifically targets Bcl-2, enhances the antitumor activity of chemotherapy, bortezomib and JQ1 in "double hit" lymphoma cells. Leukemia & lymphoma 56 (2015) 2146–2152.

[22] Hummel Michael, Bentink Stefan, Berger Hilmar, Klapper Wolfram, Wessendorf Swen, Barth Thomas F.E., Bernd Heinz-Wolfram, Cogliatti Sergio B., Dierlamm Judith, Feller Alfred C., Hansmann Martin-Leo, Haralambieva Eugenia, Harder Lana, Hasenclever Dirk, Kühn Michael, Lenze Dido, Lichter Peter, Martin-Subero Jose Ignacio, Möller Peter, Müller-Hermelink Hans-Konrad, Ott German, Parwaresch Reza M., Pott Christiane, Rosenwald Andreas, Rosolowski Maciej, Schwaenen Carsten, Stürzenhofecker Benjamin, Szczepanowski Monika, Trautmann Heiko, Wacker Hans-Heinrich, Spang Rainer, Loeffler Markus, Trümper Lorenz, Stein Harald, Siebert Reiner. A Biologic Definition of Burkitt's Lymphoma from Transcriptional and Genomic Profiling. The New England journal of medicine 354 (2006) 2419–2430.

[23] C. Love, Z. Sun, D. Jima, G. Li, J. Zhang, R. Miles, K.L. Richards, C.H. Dunphy, W.W.L. Choi, G. Srivastava, P.L. Lugar, D.A. Rizzieri, A.S. Lagoo, L. Bernal-Mizrachi, K.P. Mann, C.R. Flowers, K.N. Naresh, A.M. Evens, A. Chadburn, L.I. Gordon, M.B. Czader, J.I. Gill, E.D. Hsi, A. Greenough, A.B. Moffitt, M. McKinney, A. Banerjee, V. Grubor, S. Levy, D.B. Dunson, S.S. Dave. The genetic landscape of mutations in Burkitt lymphoma. Nature genetics 44 (2012) 1321–1325.

[24] D. de Jong, A. Rosenwald, M. Chhanabhai, P. Gaulard, W. Klapper, A. Lee, B. Sander, C. Thorns, E. Campo, T. Molina, A. Norton, A. Hagenbeek, S. Horning, A. Lister, J. Raemaekers, R.D. Gascoyne, G. Salles, E. Weller. Immunohistochemical prognostic markers in diffuse large B-cell lymphoma: validation of tissue microarray as a prerequisite for broad clinical applications--a study from the Lunenburg Lymphoma Biomarker Consortium. Journal of clinical oncology official journal of the American Society of Clinical Oncology 25 (2007) 805–812.

[25] K. Zhou, D. Xu, Y. Cao, J. Wang, Y. Yang, M. Huang. C-MYC aberrations as prognostic factors in diffuse large B-cell lymphoma: a meta-analysis of epidemiological studies. PLoS ONE 9 (2014) e95020.

[26] H. Horn, M. Ziepert, M. Wartenberg, A.M. Staiger, T.F.E. Barth, H.-W. Bernd, A.C. Feller, W. Klapper, C. Stuhlmann-Laeisz, M. Hummel, H. Stein, D. Lenze, S. Hartmann, M.-L. Hansmann, P. Möller, S. Cogliatti, M. Pfreundschuh, L. Trümper, M. Loeffler, B. Glass, N. Schmitz, G. Ott, A. Rosenwald. Different biological risk factors in young poor-prognosis and elderly patients with diffuse large B-cell lymphoma. Leukemia 29 (2015) 1564–1570.

2.5 Supplementary materials to **Clinical and pathological features of Burkitt lymphoma showing expression of BCL2 – an analysis including gene expression in formalin-fixed paraffin-embedded tissue**

	E17 BCL2 antibody score				
DAKO BCL2 antibody score	**0**	**1**	**2**	**3**	**4**
0	104/104 (100%)	0%	0%	0%	0%
1	13/15 (87%)	2/15 (13%)	0%	0%	0%
2	3/4 (75%)	0%	1/4 (25%)	0%	0%
3	1/6 (17%)	4/6 (67%)	1/6 (17%)	0%	0%
4	0%	0%	0%	3/5 (60%)	2/5 (40%)

Supplementary Table 2.3: BCL2 expression for n=134 BL in the TMA cohort for which an interpretable result was available for both BCL2 antibodies used in the study. The immunohistochemistry score (see material and methods) for the Dako antibody is indicated by each row. The number and percentage of cases and for Dako score in the second stain for E17 antibody is outlined in the columns.

Characteristics		Patients not analyzed (n=126)		Patients analyzed (n=43)		P value (Chi)
Gender	male	113	89.7%	39	90.7%	
	female	13	10.3%	4	9.3%	0.85
Age	< 10	73	58%	22	51%	
	10-14	40	31.7%	12	28%	
	> 14	13	10.3%	9	21%	0.20
Stage of	I	3	2.6%	1	2.78%	
	II	21	18.4%	10	27.78%	
	III	57	50%	20	55.56%	
	IV	5	4.4%	1	2.78%	
	B-AL	28	24.6%	4	11.1%	
	n.a.	12	-	7	-	0.42
BM involvement	yes	30	23.8%	4	9.3%	0.04
CNS involvement	yes	13	10.3%	3	7%	0.52
LDH	< 500 U/l	48	40.3%	25	61%	
	500-1000 U/l	25	21%	9	22%	
	> 1000 U/l	46	38.7%	7	17%	
	n.a.	7	-	2	-	0.03
B	yes	31	24.6%	11	25.6%	0.90
Risk group	R1	2	1.7%	0	0%	
	R2	41	35.7%	22	61.1%	
	R3	20	17.4%	5	13.9%	
	R4	52	45.2%	9	25%	
	n.a.	11	-	7	-	0.05
EFS		89±3%		90±5%		0.82 LR

Supplementary Table 2.4: Patient characteristics of patients with versus patients without known BCL2 status

2.6 Outlook

The presentation of a new technology or protocol needs of several experiments and approaches to validate it. Chapter 1 [1] presents a method to subtype aggressive B-cell lymphomas using valuable archival FFPE material. On Chapter 2 [2] we apply the new BL classifier to assess a previously unclassified group of BCL2-expressing BL, representing a further validation step for the technology.

Our work has not been alone in trying to improve lymphoma classification. A parallel DLBCL classifier was designed on the same technology and for FFPE materials in 2014 [3], named Lymph2Cx. A total of 15 genes and 5 endogenous genes were used (see table 2.5). Four of the 20 genes overlapped our GCB-ABC and the Lymph2Cx classifier (IRF4 was part of our custom Code Set but not selected as a prediction gene). Unfortunately an algorithm comparison was not possible due to the unavailability of GEO datasets of the Lymph2Cx data.

TNFRSF13B	PIM2	R3HDM1	TRIM56	MAML3
LIMD1	CYB5R2	WDR55	**MME**	ITPKB
IRF4	**RAB7L1**	ISY1	SERPINA9	**MYBL1**
CREB3L2	CCDC50	UBXN4	ASB1	S1PR2

Table 2.5: Genes used in the Lymph2Cx assay. In bold are the overlapping genes with our work.

In a further publication [4] the same group announced that the Lymph2Cx array could assign prognostic indexes to the analysed samples, relative to their subtype or cell of origin labelling (namely ABC, unclassified and GCB). Such developing could indicate that a clinical application based on FFPE materials and multiplex GEP for aggressive B-cell lymphomas could be available in the near future.

A broad study on nCounter technology reliability was published in 2015 [5] and highlights the robustness of the platform by testing multiple variables such as batches of Code Set, instruments, probe sequence design, among others.

However, by comparing Nanostring's-generated data sets to genome scale data (RNA-seq and Illumina arrays) some authors [6] postulate that for research purposes the latter are more informative. It is clear that nCounter focuses on a discrete number of analysable genes, whilst genome scale methods screen genetic phenomena, making it more suitable for basic research. Even so the advantage of Nanostring technology's rapid and unbiased procedure cannot be easily discarded, because it allows for a straightforward analysis of a known question (as disease subtyping). An example of an nCounter-based clinical application is the company's Prosigna™ Breast Cancer Gene Signature Assay, FDA approved for the clinic.

Certainly, nCounter is not alone in the GEP approach for the classification of FFPE material. Other systems for the molecular analysis of lymphoma have been presented lately. One interesting method is the combination of quantitative PCR (qPCR) and capillary electrophoresis separation [7]. With a total of 14 algorithm genes, 2 reference genes and one internal control samples could be assigned to either the ABC or the GCB DLBCL subtype. This assay goes back to the material amplification step of the original gold standard arrays and should be taken with caution when generating new data to avoid biasing errors.

In an era of many available molecular profiling tools the use of one with detriment of another would be altogether scientifically damaging. To prevent this, a combination of techniques could be helpful. Such experiments could be designed to firstly screen a set of samples with high throughput techniques for relevant signature genes; afterwards selected genes of interest could be shaped into a classification or prognostic tool (for example, with GEP or qPCR approaches).

References

[1] N. Masqué-Soler, M. Szczepanowski, C.W. Kohler, R. Spang, W. Klapper. Molecular classification of mature aggressive B-cell lymphoma using digital multiplexed gene expression on formalin-fixed paraffin-embedded biopsy specimens. Blood 122 (2013) 1985–1986.

[2] N. Masqué-Soler, M. Szczepanowski, C.W. Kohler, S.M. Aukema, I. Nagel, J. Richter, R. Siebert, R. Spang, B. Burkhardt, W. Klapper. Clinical and pathological features of Burkitt lymphoma showing expression of BCL2 - an analysis including gene expression in formalin-fixed paraffin-embedded tissue. British Journal of Haematology (2015).

[3] D.W. Scott, G.W. Wright, P.M. Williams, C.-J. Lih, W. Walsh, E.S. Jaffe, A. Rosenwald, E. Campo, W.C. Chan, J.M. Connors, E.B. Smeland, A. Mottok, R.M. Braziel, G. Ott, J. Delabie, R.R. Tubbs, J.R. Cook, D.D. Weisenburger, T.C. Greiner, B.J. Glinsmann-Gibson, K. Fu, L.M. Staudt, R.D. Gascoyne, L.M. Rimsza. Determining cell-of-origin subtypes of diffuse large B-cell lymphoma using gene expression in formalin-fixed paraffin-embedded tissue. Blood 123 (2014) 1214–1217.

[4] D.W. Scott, A. Mottok, D. Ennishi, G.W. Wright, P. Farinha, S. Ben-Neriah, R. Kridel, G.S. Barry, C. Hother, P. Abrisqueta, M. Boyle, B. Meissner, A. Telenius, K.J. Savage, L.H. Sehn, G.W. Slack, C. Steidl, L.M. Staudt, J.M. Connors, L.M. Rimsza, R.D. Gascoyne. Prognostic Significance of Diffuse Large B-Cell Lymphoma Cell of Origin Determined by Digital Gene Expression in Formalin-Fixed Paraffin-Embedded Tissue Biopsies. Journal of clinical oncology official journal of the American Society of Clinical Oncology (2015).

[5] M.H. Veldman-Jones, R. Brant, C. Rooney, C. Geh, H. Emery, C.G. Harbron, M. Wappett, A. Sharpe, M. Dymond, J.C. Barrett, E.A. Harrington, G. Marshall. Evaluating Robustness and Sensitivity of the NanoString Technologies nCounter Platform to Enable Multiplexed Gene Expression Analysis of Clinical Samples. Cancer research 75 (2015) 2587–2593.

[6] C. Sha, S. Barrans, M.A. Care, D. Cunningham, R.M. Tooze, A. Jack, D.R. Westhead. Transferring genomics to the clinic: distinguishing Burkitt and diffuse large B cell lymphomas. Genome medicine 7 (2015) 64.

[7] Angela M. B. Collie, Jork Nölling, Kiran M. Divakar, Jeffrey J. Lin, Paula Carver, Lisa M. Durkin, Brian T. Hill, Mitchell R. Smith, Tomas Radivoyevitch, Lilly I. Kong, Thomas Daly, Gurunathan Murugesan, Jeanna Guenther-Johnson, Sandeep S. Dave, Elena A. Manilich, Eric D. His. Molecular subtype classification of formalin-fixed, paraffin-embedded diffuse large B-cell lymphoma samples on the ICEPlex® system. British Journal of Haematology 167 (2014) 267–289.

Chapter 3

Immunoglobulin rearrangement bias and stereotyped mutations in Burkitt lymphoma

Neus Masqué-Soler[1], Monika Szczepanowski[1], Wolfram Klapper[1]

[1] Department of Pathology, Hematopathology Section and Lymph Node Registry, University Hospital Schleswig Holstein, Campus Kiel / Christian-Albrecht University, Kiel, Germany

Manuscript in preparation.

Abstract

Immunoglobulin (*IG*) gene usage bias and stereotyped mutations in the same genes are present in B-cell lymphoma. Studies in chronic lymphocytic leukaemia (CLL), mantle cell lymphoma, splenic marginal zone lymphoma and lately in follicular and Burkitt lymphoma (BL) have elucidated the most recurrent *IG* gene in each lymphoma entity. Stereotypy has served to classify large cohorts in CLL and even create prognostic relevant groupings. The cause of these stereotyped mutations is thought to be of antigenic origin, although none relevant has been described for sporadic Burkitt lymphoma so far. We review the latest works on *IG* bias and stereotyped mutations in lymphoma and compare it to two unpublished BL data groups. Moreover, the frequently used *V_H4-39* segment in BL is analysed for its stereotyped mutation structurally in three independent BL cases and screened for superantigen binding. As the antigen drive hypothesis in lymphomagenesis evolves, a comprehensive understanding in the structural and molecular antibody-antigen interactions in each lymphoma entity is needed.

Keywords: Immunoglobulin gene usage, stereotyped mutations, Burkitt, B-cell lymphoma.

3.1 Introduction

Immune system B lymphocytes or B cells express B-cell receptors (BcR) on their surface. BcRs are composed of a transmembranous immunoglobulin (IG) molecule that faces extracellularly and an integral membrane heterodimer (CD79a and CD79b). The IG part of the BcR is in charge of antigen recognition and it is composed of two identical heavy and two identical light chains linked by disulfide bonds. Each chain has a variable and constant

domain, being the former the antigen recognising part and the latter the effector and isotype defining part. Four framework regions (FR) and three complementarity determining regions (CDR) are distributed alternately throughout the variable domains. The three CDR regions from a heavy chain form three loops that upon IG conformation are close to three other loops from its corresponding light chain. Together, the six loops will form an antigen binding site (ABS). Each IG molecule possesses two ABS since each IG is built by four IG chains (two heavy and two light).

B cell receptors are able to specifically recognise antigens. By binding unknown bodies an immunological reaction is generated that includes complement and effector cell activation, aiming to control and eliminate a possible pathogen. To perform this function a great variability of antigen recognition sequences need to be expressed in the CDRs. This wide repertoire is achieved by V(D)J recombination. Diverse genes compose the variable domain of the IG chains, named Variable (V), Diversity (D) and Joining (J) segments. There are 44 V, 27 D and 6 J genes in the human IG heavy chain locus. Alternatively, light chains lack D segments but also have multiple V and J genes. When combined, one gene of each segment builds an IG chain. It is accepted that the IG repertoire in B cells is 10^{12} molecule-strong [1].

A biased usage of the immunoglobulin heavy chain variable region (*IGHV*) genes in lymphoma has been reported previously, hypothesising an antigen drive or pathogen-related cause. Although not all lymphomas show the same recurrent *IG* rearrangements there are patterns to be found in each disease type.

The first hint of this phenomenon was detected in chronic lymphocytic leukaemia (CLL) [2]. Particularly, half of the cases in CLL use the *IGHV3-21* gene, frequently in combination with the *IGLV3-21* light chain gene. *IGH* gene rearrangement bias has also been seen in mantle cell lymphoma (MCL) [3], by recurrent use of segments V_H*3-21*, V_H*4-34*, V_H*1-8* and V_H*3-23*. The same group also reported specific usage of *IGHD* or *IGHJ* genes in combination with several *IGHV* sequences. In yet another lymphoma entity, namely splenic marginal zone lymphoma (SMZL), up to 30% of the cases use *IGHV1-2*04* [4,5]. Until recently there were no studies involving *IG* gene repertoire in Burkitt lymphoma (BL). Newly, one group [6], pointed out an expression bias of the *IGHV3* and *IGHV4* segments in this disease.

After the BcR has gone through the V(D)J recombination procedure an affinity maturation process (via somatic hypermutation –SHM- and class switch recombination –CSR-) adds further variability to the IG molecule. B cell receptor stereotypy defines identical mutations detected in independent CDRs of B-cell clones. The probability of such phenomena is very low ($1/10^{-9}$ to $1/10^{-12}$) and it is not typical of physiological SHM. Stereotyped mutations in BcRs are found in B-cell leukaemia and lymphomas with rearranged and clonal BcR in which their variable heavy CDR sequences are restricted or

not similar to healthy patients, where their sequence variability is greater. Although CDR residues only account for 70 of the average 230 residues of the variable fragments [7], they are crucial for antigen-binding [8].

So far, stereotyped or recurrent mutations present in CDR3 sequences have been detected in CLL [2,9], MCL [3], SMZL [5] and BL [6,7]. The biased V_H segments of CLL patients also showed short restricted V_H CDR3s [11]. Up to date it is considered that almost one third of all CLL cases contain stereotyped V_H CDR3 sequences. In MCL, 10% of the 807 analysed cases bore stereotyped V_H CDR3 sequences [3]. With long CDR3 regions and not a strong evidence of SHM, SMZL cases seem to have decisive stereotyped mutations [4].

Biased uses of *IG* gene segments cannot be considered random events, as they seldom occur accidentally. In addition, it has been repeatedly proven that the restricted usage of the V_H domain genes underlie the generation of stereotyped CDR3 motifs of the V_H [12], and thus it can be inferred that stereotypy of CDR motifs in lymphoma populations is associated to *IGH* gene repertoire restriction.

This co-occurrence can be of clinical relevance: studies in CLL have proven that patient stratification according to their stereotyped mutations could generate prognostic-relevant clades [13]. However, the predilection for concrete V_H segments above others in one lymphoma cannot be horizontally translated to other entities, thus being "disease biased". The same can be said for stereotyped mutations in the CDRs of lymphoma BcRs. A need to collect and analyse other lymphoma entities BcRs in sufficient number to elucidate repertoire and mutational patterns has emerged.

With long and detailed studies focusing on the 1D level of stereotyped mutations, incursions in structural characterisation of the mutational effects in affected lymphoma antibodies are lacking. A better understanding of the structural effects the mutational changes have in the antigen-recognising pocket of the antibody might serve to elucidate the nature of these lymphoma antibodies and their tumoral B-cell origin.

Here, we review the latest literature results in BL *IG* loci usage and mutation status. Furthermore three cases of the most prominently stereotyped *IGHV* in BL are characterised in their 3D structures by comparing their originally mutated CDR2 to their unmutated germline sequences.

3.2 Materials and methods

3.2.1 *IG* gene repertoire analysis in B-cell lymphoma

Two unpublished studies to date and a recent publication are reviewed in the first part of this work:

Within the network project MMML (Molecular Mechanisms in Malignant Lymphomas, Deutsche Krebshilfe, 70-3173-Tr3) a total of 60 sporadic BL (sBL) *IGHV* rearrangements were amplified and sequenced [6] using the BIOMED-2 primers and protocols [14]. The obtained sequenced data was run through the IMGT http://www.imgt.org/) database.

Based on next generation sequencing (NGS) and as part of the International Cancer Genome Consortium (ICGC-MMML-Seq, referred here as ICGC) project (https://dcc.icgc.org/) the *IG* loci were RNA-sequenced [15] to assess IG segment usage in different types of B-cell lymphoma [personal communication]. Subtypes of diffuse large B-cell lymphoma (DLBCL) to follicular (FL) and Burkitt lymphomas were analysed.

A further study in BL from Baptista et al [7] was also considered for comparative and validation purposes.

Statistical analysis was done in GraphPad Prism with two-way ANOVA test and a significance level of P=0.05.

3.2.2 Original samples, cDNA extraction and sequencing

Samples R87264/95 (MPI-079), R87184/02 (MPI-589) and R82092/94 (MPI-621) were obtained from fresh-frozen biopsies stored in the Lymph Node Registry, Kiel and had been previously enrolled in the MMML study. From an initial 2μg of RNA, extracted from frozen sections with RNeasy Mini Kit (Qiagen), cDNA was reverse transcribed following the protocol for Invitrogen's SuperScript™ First-Strand Synthesis System (Thermo Fisher Scientific). The resulting material from the reverse transcription procedure was directly amplified through PCR. For set up and reaction composition, see the Buffer and Reaction list (Annex).

A total of 22 primers were tested to obtain the correct IG fragment in the three cases analysed (see Primer Annex, Table P1). Primers were tested in combinations for V_H (forward) and J_H (reverse), V_L and J_L, and V_K with J_K segments (Table P2 in Primer Annex). Electrophoresis gels at 1% agarose were run for 30 min at 120V to separate the PCR products. For each case a band containing the desired IG chain amplicon was generated, cut and separated in microcentrifuge tubes. DNA was isolated from the agarose gel fragments using the NuceloSpin Extract II kit (Macherey&Nagel), and the resulting DNA concentration was measured with NanoDrop (Thermo Fisher Scientific).

With the cDNA a sequencing reaction was set up following ABI Prism's guidelines on DNA quantity for such reactions. The sequencing reaction was set as appears in the Buffer and Reactions list (Appendix). The sequencing reaction purification was done using NucleoSEQ columns (Macherey & Nagel) by following the company's protocol.

Afterwards the sequences were analysed with a Sanger sequencer from a 3500 ABIPrism (Thermo Fischer Scientific) and they were compared to their germline sequences.

3.2.3 Immunoglobulin isotype assessment of the 3 BL-specific monoclonal antibodies

Primers were designed to be specific for each sample (forward) and the two variants, secretory and transmembranous, of the constant IgHμ region (reverse) (see Table P3 in Primer Annex) by guide of existing publications [14] and P01871 (IGHM_HUMAN), isoforms P01871-1 and P01871-2 (UNIPROT: http://www.uniprot.org/). The previously obtained cDNA was amplified to determine the original nature of the expressed Ig.

3.2.4 *In silico* structural analysis

General analysis of the sequenced *IG* genes from our patient material was obtained from IMGT (http://www.imgt.org/IMGT_vquest/vquest) [17], a tool that confirmed their productivity and assessed their germline identity.

In silico modelling was generated of the three BL mAbs for their respective heavy and light chains. The sequences obtained from each case were translated with the translation tool in ExPASy (http://web.expasy.org/translate/) and the resulting amino acid (AA) data was fed to the Antibody section of the web-based tool Rosie (http://rosie.rosettacommons.org/antibody) for modelling. The output from Rosie, a PDB file, was in turn submitted to Paratome (http://www.ofranlab.org/paratome/) [18], a bioinformatic online tool that predicts the ABS of any given antibody structure or sequence [19]. Complementarity determining region estimations were retrieved from IMGT's V-QUEST (http://www.imgt.org/IMGT_vquest/vquest) and Rosie. For CDR2 mutation study the S>N stereotyped mutation was reverted and the resulting modelling was compared to the previous PDB file with the mutation. The online structure alignment tool RCSB PDB Protein Comparison Tool (http://www.rcsb.org/pdb/workbench/workbench.do) was used for this purpose. All modelling PDB files were visualised with Discovery Studio Visualizer 4.1 (Accelrys Software Inc.).

3.3 Results

3.3.1 *IG* gene repertoire analysis in Burkitt lymphoma

The RNA of a cohort of 132 B-cell lymphomas within the ICGC consortium was deep sequenced (RNA-seq) to infer the most prominently expressed IGH@ rearranged loci in the tumour samples. The cohort contained DLBCL, intermediate, BL and FL cases. There were only 29 different V_H segments present in the most prominently expressed IGH alleles,

with an overall bias towards *IGHV3-23*, *3-30*, *3-48* and *4-34* (Figure 3.1). While the *3-23*, *3-48* and *4-34* rearrangements could be attributed to FL cases, there were also biases in other lymphoma entities, being V_H*3-30*, V_H*4-34* recurrent in BL.

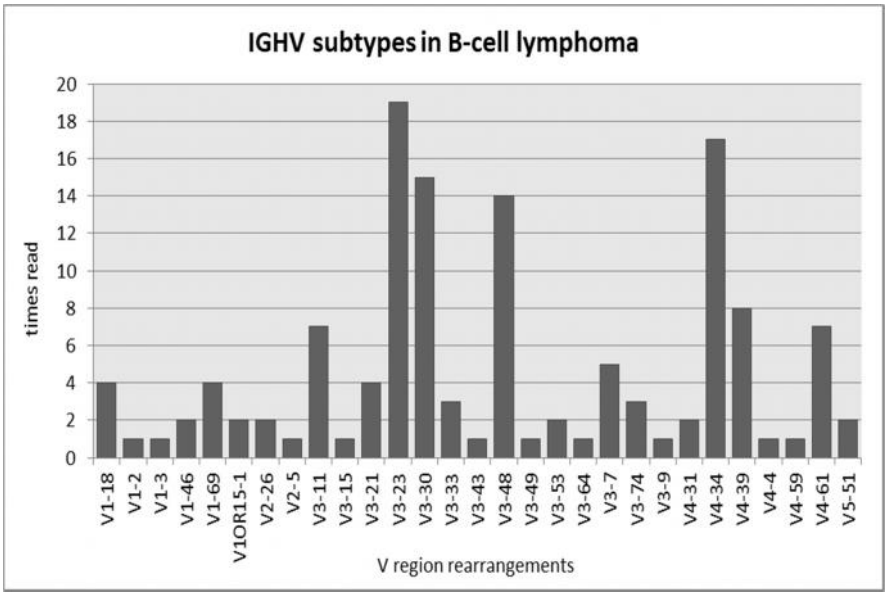


Figure 3.1: *IGHV* rearrangement spectrum in B-cell lymphoma, ICGC.

Nineteen of the 132 lymphoma samples were diagnosed as BL. As three BL cases had been previously included in the MMML study they are not further considered here, making a total of 16 final BL. Their predominant *IGHV* segments were V_H4 (43.8%, with 7/16 samples) and V_H3 (37.5%, with 6/16 samples). More specifically, the most recurrent subtypes were *IGHV4-34* (4/16, 25%) and *IGHV3-30* (2/16, 12.5%). Overall the 16 BL cases had a clear bias towards *IGHJ4* segment usage, being this segment present in 10/16 (62.5%) of the cases. For detailed subtype information see supplementary Table 3.5.

The unpublished work by Trautmann and colleagues within the MMML joint project [6] amplified and analysed the productive VDJ rearrangements of 54 BL (supplementary Table 3.7). Strikingly, a total of 42.6% (23/54) *IGHV* sequences belonged to the *IGHV4* subgroup, and 39% (21/54) to the *IGHV3*. Compared to the 54-BL of the MMML cohort the *IGH* gene rearrangement distribution for the V region in ICGC is similar (Figure 3.2, Table 3.1). Looking at patient age, *IGHV* subtype *4-34* and *4-39* presented a bias to paediatric cases, being significantly more frequent in underage patients (7/11 and 5/6 respectively). With regard to the *IGHV4-39* variant, a bias towards the *IGHJ3* gene usage was also seen, being present in 4/6 of the V_H*4-39* cases. In the ICGC cohort the case presenting the V_H*4-39* segment showed an *IGHJ4* gene.

In another study [7] with a total of 72 BL cell lines (35/72) and BL patient biopsies (37/72) were screened for V segment usage. The most prevalent family rearrangements in

the heavy chain *IG* gene were *IGHV3* (47.9%) and *IGHV4* (35.6%), with the other families being comparatively underrepresented (Table 3.1). This group could also demonstrate that the usage of the *IGHV* genes was statistically different to those of normal B-cells and particularly the event of *IGHV4-34* subtype usage was significantly higher to that of non-tumoral cells. With a high mutational load of the series, Baptista et al considered that practically all cases had gone through SHM. As previously reported [20,21] IgM expression was confirmed and no bias towards kappa or lambda light chain was seen.

	MMML			**Baptista et al**			**ICGC**		
	Variants	n	Percentage	Variants	n	Percentage	Variants	n	Percentage
IGHV1	4	6	11.1	6	7	9.6	1	1	6.3
IGHV2	1	2	3.7	2	2	4.1	1	1	6.3
IGHV3	11	21	38.9	13	35	47.9	5	6	37.5
IGHV4	6	23	42.6	7	26	35.6	4	7	43.8
IGHV5	1	1	1.9	2	2	2.7	1	1	6.3
IGHV6	1	1	1.9	0	0	0.0	0	0	0.0
Total		54			72			16	

Table 3.1: IGHV rearrangement distribution in three independent BL cohorts.

While statistically comparing the most recurrent *IGHV* BL rearrangements in all three cohorts (*V3-23*, *V3-30*, *V4-34* and *V4-39*) the results were negative, showing no significant differences in the frequency of the most recurrent allele recombinations. A preference in light chain usage (lambda or kappa) was detected in none of the three groups

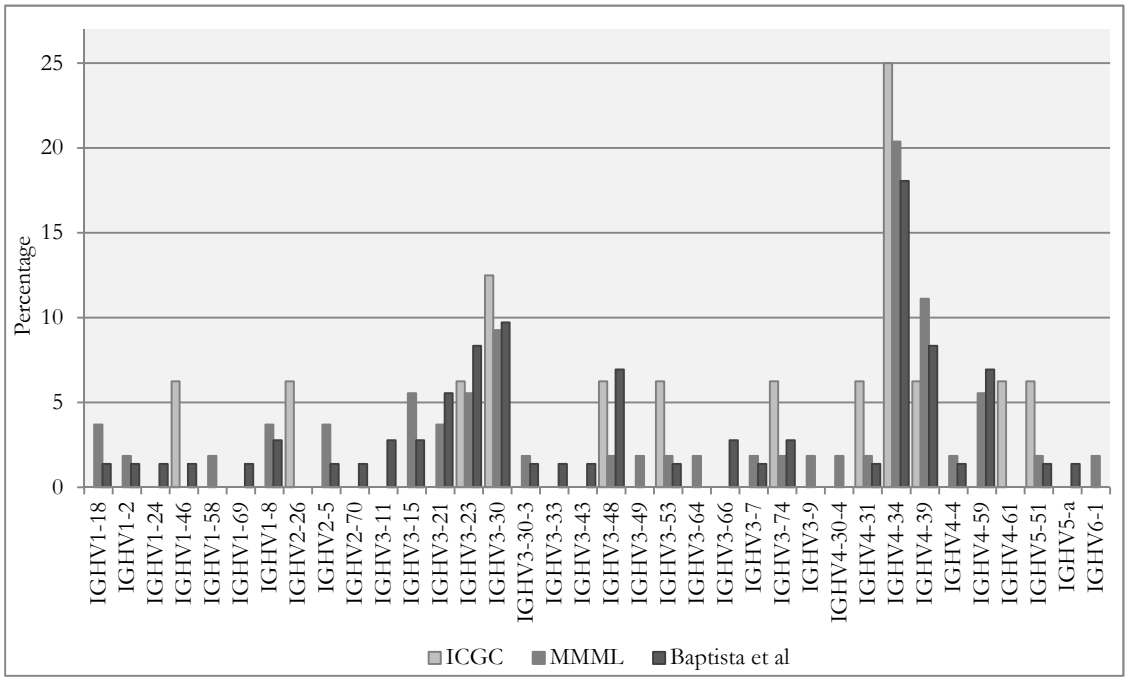


Figure 3.2: *IGHV* rearrangement frequency percentages in Burkitt lymphoma corresponding to the ICGC, MMML and Baptista et al cohorts.

Interestingly, while stereotyped mutations in CDR3 were not detected, they were found in the CDR1 and CDR2 (MMML cohort, [6]). Table 3.2 summarises the recurrent AA changes observed in the *IGHV4-34* and *4-39* subtypes. Some of the recurrent mutations were localised in the framework regions (FR) but some were seen in the CDR1 for *IGHV4-34* cases and in the CDR2 for *IGHV4-39* cases.

IGHV4-34 sequences	**AA change**	**BL**	**CLL***
IMGT-HCDR1, codon 36	G-to-D	5/11	27/135
IMGT-HFR3, codon 80	V-to-L	3/11	7/135
IMGT-HFR2, codon 40	S-to-T	3/11	43/135
IMGT-HFR2, codon 45	P-to-S	3/11	30/135
IMGT-HFR3, codon 69	P-to-S	2/11	1/135
IGHV4-39 sequences	**AA change**	**BL**	**CLL**
IMGT-HCDR2, codon 64	S-to-N	4/6	2/45

Table 3.2: Stereotyped mutations in *IGHV4-34* and *IGHV4-39* (MMML, [6]). *Results from Murray F et al. Blood 2008.

CDR-located mutations are of particular interest due to their being part of the ABS. We focused on the serine-to-asparagine (S>N) AA change (AGC to AAC) in the CDR2, which was detected in 4/6 (66%) of the *IHGV4-39* samples and is also seen in CLL. Three BL paediatric cases presenting both the *IGHV4-39* rearrangement and the stereotyped CDR2 S>N mutation at codon 64 from the MMML [6] study were recovered and further analysed.

3.3.2 Clinical information

The three paediatric patients source for the RNA material and subsequently their mAb had all been diagnosed with BL by a panel of pathologists within the MMML study and later assigned their molecular Burkitt lymphoma label. The cases presented typical morphological and immunohistochemical characteristics for BL, such as expression of CD10, CD20, BCL6, Ki67 (more than 90%), and IgM as well as presenting the BL hallmark translocation t(8;14). The three cases were negative for TdT, EBV and BCL2. Being paediatric patients, all three were treated with the NHL-BFM protocol and none suffered a relapse (see Supplementary Table 3.9 for clinical data).

3.3.3 BL mAbs characterisation

By sequencing the amplified *IGH* and *IGK* or *IGL* loci of the specific 3 paediatric patients and analysing the sequences with the online IMGT/V-QUEST tool the prominent IG rearrangement for each of the 3 BL samples could be determined (Table 3.3).

Heavy chains				Light chains		
CASE	**IGHV**	**IGHD**	**IGHJ**	**CASE**	**IGLV**	**IGLJ**
079	IGHV4-39*01	IGHD2-2*02	IGHJ3*01	**079**	IGLV3-12*01 IGLV1-51*02	IGLJ3*02 IGLJ3*01
589	IGHV4-39*01	IGHD6-13*01	IGHJ6*04	**589**	IGKV3-20*01	IGKJ2*01
621	IGHV4-39*07	IGHD4-4*01	IGHJ3*01	**621**	IGLV1-47*02	IGLJ1*01

Table 3.3: *IG* heavy and light chain gene usage of the chosen BL patients. All three rearrangements presented the stereotyped mutation in CDR2 from serine (Ser, with codon AGC) to asparagine (Asn, with codon AAC) in codon 64.

All three cases showed the same V segment of the heavy chain (*V_H4-39*), only with case 621 having a distinct sub-variant (*07 instead of *01). The D segments were not redundant but at the J segment case 079 and 621 converged again by using J3*01, having an identical sub-variant. On the light chain side, case 079 and 621 displayed a predominant lambda allele and case 589's was a kappa. Two light chain alleles were sequenced in preliminary amplifications of case 079, being named 079A (*IGLV3-12*01*) and 079B (*IGLV1-51*02*). However, only 079B was successfully amplified and referred to as 079 for further purposes.

The *IGHM* locus amplification yielded several bands in the electrophoresis gel due to the IGHM primer structure and the fact that the amplification used part of the constant region sequence. However, only the primer combinations with secretory primers gave positive results. This confirmed that the type of Ig cloned is originally an IgHμ and of the secretory type.

Figure 3.3: *IGHM* locus amplification electrophoresis gel. Odd-numbered lanes contain cDNA material amplified with primers for the membranous region and even-numbered lanes contain secretory-amplified *IGHM*. Lanes 1 and 2 represent case 079, lanes 3 and 4 case 589, and lanes 5 and 6 show case 621.

Germline sequence comparison or germline (GI) identity calculation of the three IG cases helped to classify two of them (079 and 589) as mutated (<98% GI), with 93.47%

and 94.50% GI respectively; and one (621) as minimally mutated (98-99% GI), with a 98.97% GI. This could be a sign of ongoing SHM [5]. However, with lack of subclonal analysis of our patient's HCDR3 this cannot be fully confirmed.

Binding motives for staphylococcal protein A (SpA) and carbohydrate I/i are scanned in our 3 mAb sequences since they have been postulated to be relevant superantigens [21,22]. SpA binding motifs have been mainly described for germline sequences of the *IGHV3* family subtypes while the hydrophobic patch in heavy chain framework region 1 (HFR1) of the *IGHV4* family (germline) has been related to carbohydrate I/i.

	HFR1							HCDR2	HFR3								
Codon	7	16	18	20	24	**25**	26	**65**	**67**	**72**	74	**75**	**77**	**79**	90	**92**	**93**
SpA		G	S	R/K				K/I/**T**	**Y**	**K**	G	**R**	**T**	**S**	Q	**N**/G	**S**
079	S	S	T	S	S		S				S				K	T	
589	S	S	T	S	S		S	A			S				K	S	
621	S	S	T	S	T		S				S				K		
Carbohydrate I/i	W				A	**V**	Y										

Table 3.4: Binding motifs for the SpA and carbohydrate I/i hydrophobic patch in the cloned BL antibodies. Empty cells in the antibody rows show a match in the superantigen binding sequences, which in turn are marked in bold letters in the superantigen rows.

In our 3 mAbs, all of the *IGHV4-39* segment family, there should be no expected conserved superantigen binding sites, however, there were up to 8/13, 7/13 and 9/13 of the residues in respectively each case 079, 589 and 621. The conserved binding residues were mainly concentrated in the HFR3 area.

3.2.4 *In silico* structural analysis

Both variable domains of the heavy and light chains were modelled in silico, containing the three complementarity determining regions, which span throughout the V(D)J regions. The resulting Paratome-defined ABS (Supplementary Figure 3.9) showed distinct pocket volumes throughout the 3 BL mAbs. Case 079 had a mostly narrow and deep ABS pocket, being 1.6nm x 0.6nm x 1.55nm (wide x long x deep). Being also slightly narrow, case 589 was not as deep and had the following dimensions: 1.65nm x 1.2nm x 0.9nm. Monoclonal antibody 621 had a wider and shallower ABS, with 1.63nm x 1.62nm x 0.85nm in size.

Comparing the CDR areas estimated by IMGT and Rosie servers to the ABS determination from Paratome [18] there were some discrepancies to be found. CDR and ABS are not identical as it has depends on the numbering system used, modelling methods and data available at analysis (Supplemental Figures 3.7 and 3.8). Of note, Paratome is able to define Ag binding Paratome-unique residues, unidentified with previous CDR detection

methods, without losing CDR-relevant residues [19]. The differences in CDR and ABS are mostly due to the approaches used in residue numbering. Since Kabat's first approach in defining CDRs [9], followed by Chothia's contribution [24], to the more recent IMGT numbering system [25], there have been many improvements on the subject. The latest incursion in redefining the antibody-antigen (Ab-Ag) contact regions defines Antigen Binding Regions (ABS) by analysing Ab-Ag complexes, showing that Ag binding residues belong to structural consensus regions [19].

Figure 3.4: Variable regions of the 3 BL IgMs analysed. Continuous lines mark the CDR sequences detected by IMGT and discontinuous lines mark the CDRs recognised by Rosie. Letters in colour mark the ABR as detected by Paratome as coded in the figure legend. A: 079B heavy chain, B: 079B light chain, C: 589 heavy chain, D: 589 light chain, E: 621 heavy chain and F: 621 light chain.

To assess whether the stereotyped mutation in the CDR2 of our 3 BL mAbs provoked a substantial change in the ABS structure, the stereotyped mutation from serine to asparagine was reverted and the two modellings were compared with the structure alignment tool RCSB PDB Protein Comparison Tool. The modelled comparisons of case 079 yielded 3 regions of structural difference, one including the mutated AA and two regions not directly in contact. The smallest backbone shift occurred in the I29-Y33 AA, belonging to the first ABS of the light chain. There were two more slightly larger backbone shifts of 0.12nm in its maximum displacement between M162-T168 (ABS2 of the heavy chain, where the CDR2 mutation is located) and S210-F218 (ABS3 of the heavy chain).

Case 589 displayed a maximum of 0.133nm of backbone separation after mutation reversion. There were four structural changes detected: between D137-Y144 in ABS1, S166-Y170 in ABS2 and I210-V217 in ABS3 of the heavy chain and between G93-Y99 in the ABS3 of the light chain. In case 621 there was a maximum backbone displacement of 0.233nm, with 3 main backbone angle shifts: in ABS1, comprising S141-W146, ABS2 in S162-T169 and ABS3 in G210-D214 of the heavy chain.

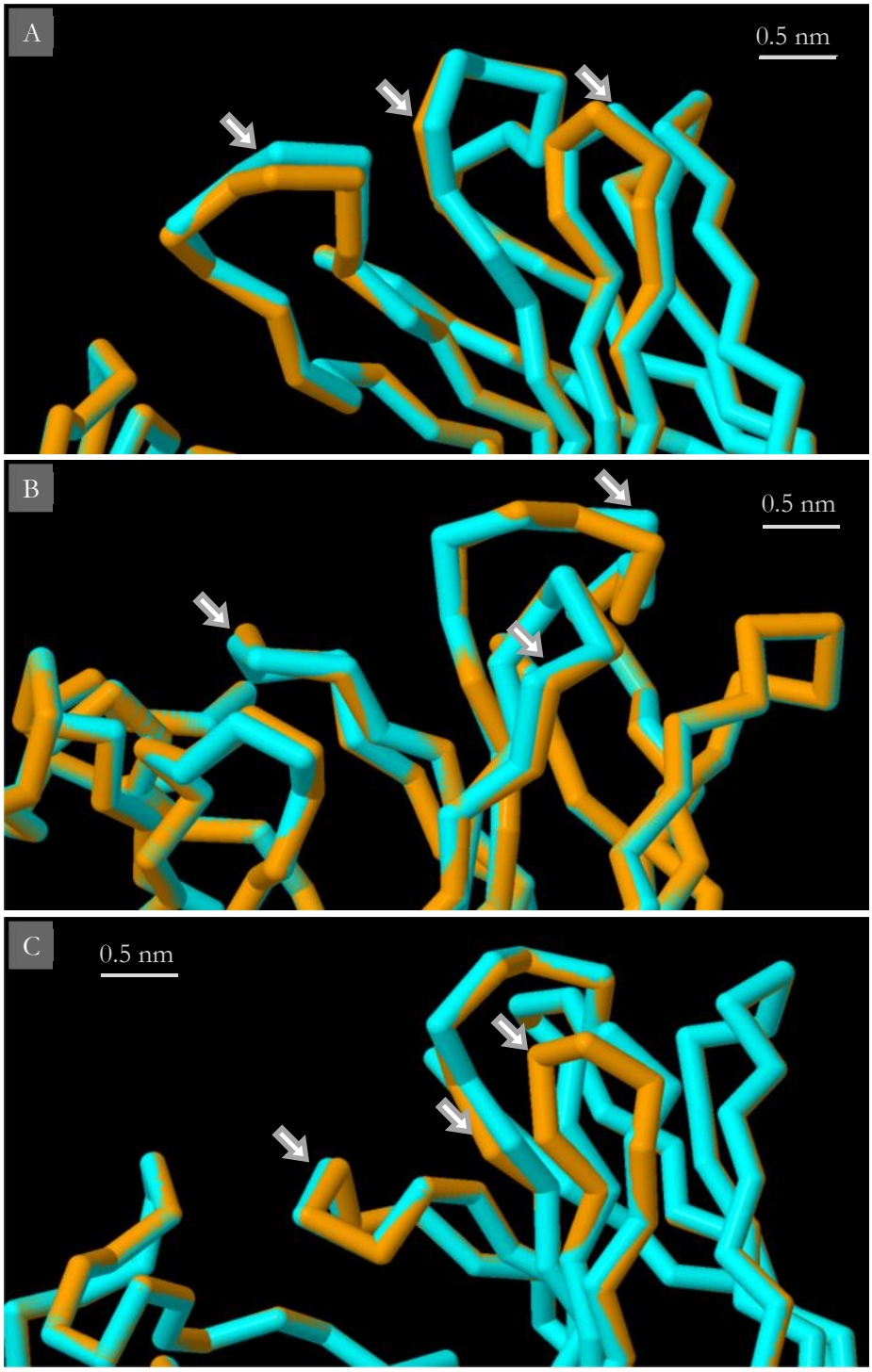


Figure 3.5: Antibody backbone alignment of 079 (A), 589 (B) and 621 (C). The arrows mark the main structural shifts.

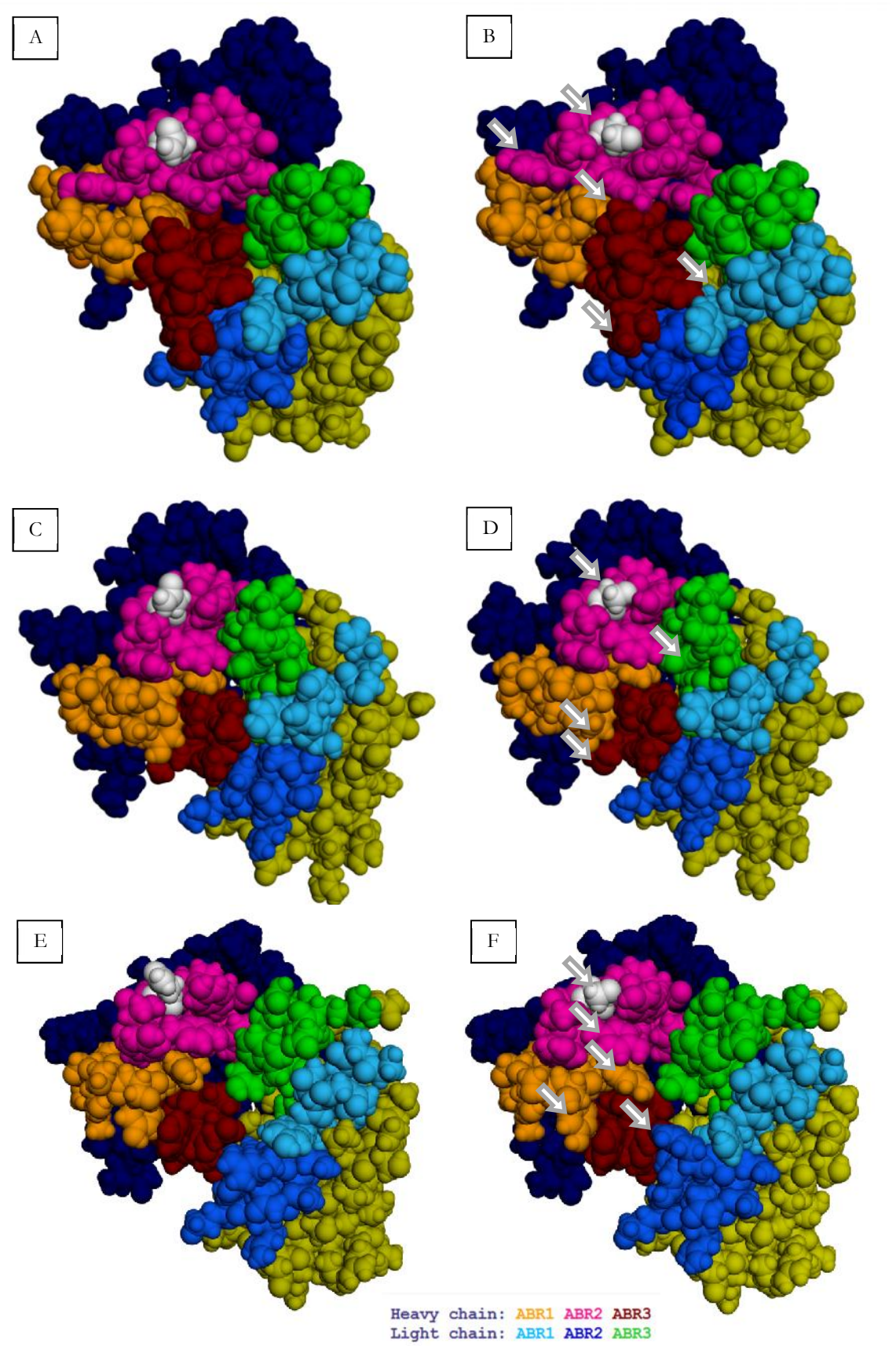


Figure 3.6: 3D modelling of the three BL mAbs. Complementarity determining regions are defined by Paratome and colour labelled accordingly. Case 079: A-B, case 589: C-D,

and case 621: E-F. Figures on the left are the strereotyped mutated antibodies as extracted from patients' cDNA. Right figures had their codon 64 mutation reverted. Arrows mark the main structure differences upon CDR2 mutation reversion.

3.4 Discussion

Lymphoma cells, as their normal B cell counterparts, are dependent on their surface-expressed BcRs for tonic survival signalling. As they originate from differentiated B cells, this fact is not surprising but shows the importance of the BcR expression in lymphoma cell populations [26].

By reviewing *IGHV* gene usage in Burkitt lymphoma it has been shown that the most recurrent segment in this lymphoma entity is V_H*4-34* (in 18-25% of the cases). Table 3.5 summarises the most used segments detected in different B-cell lymphoma entities. This particular segment is also biased in FL (ICGC, in preparation), MCL in 14.5% of cases [3] CLL in variable frequencies [27] and DLBCL in 31% [40]. The second most prominent *IGHV* gene in Burkitt lymphoma is V_H*3-30*, being consistently seen in around 9% of the cases, and also present in 8% of FL. Another *IGHV* gene with overrepresentation in BL is also detected in other lymphoma populations in a biased fashion: V_H*3-23*. This segment is present in FL (22.5%) (ICGC, in preparation), MCL (7.5%) [3] and CLL (5-12%) [13,27,28]. However, it has been demonstrated that gene V_H*3-23* is also frequently used by normal $CD5^+$ B cells [29,30]. With presence in 8.5% of the cases, gene V_H*4-39* is uniquely biased in BL. Both *IGHV4-34* and *IGHV4-39* are used by only 3-5% of normal B cells [31] and the same rearrangements were seen to be biased towards paediatric cases over adults (MMML, [6]).

CLL	%	MCL	%	SMZL	%	FL	%	BL	%	DLBCL	%
V3-21	11.7	V3-21	16.5	V1-2	24.9	V3-23	22.5	V4-34	21.1	V4-34	31
V1-69	7.2-11.9	V4-34	14.5	V4-34	12.8	V3-48	14.3	V3-30	9.1	V3-7	7.8
V4-34	11.9-18	V1-8	8	V3-23	8.1	V4-34	14.3	V4-39	8.5		
V3-7	12.1	V3-23	7.5			V3-30	8.2	V3-23	6.7		
V3-23	4.8-7-11.9					V3-11	8.2				
						V3-7	6.1				
[10,12,27,29]		[3]		[4]		[ICGC, p.c.]		[6,7, ICGC, p.c.]		[40]	

Table 3.5. Summary of biased IGHV usage in B-cell lymphoma. References are in brackets for each column and entity. P.c: personal communication.

A B cell differentiation starts after exiting the bone marrow, having already gone through *IG* loci recombination. The naïve B cells circulate the bloodstream, where they will

be exposed to antigens. They can either become regulatory B cells, short-lived plasma cells (expressing IgM or IgG) or enter the germinal centre (GC) where they will undergo affinity maturation through SHM, and CSR [32]. In the case of BL lymphocytes, the expression of CD10 and BCL6, the mature phenotype of the tumoral B cells as well as intraclonal diversity and acquisition of glycosylation sites [7] hints that they might have gone through SHM. Ongoing SHM is further confirmed with the germline identity (GI) being lower than 100% in the 3 BL mAbs analysed in this work and elsewhere [7]. Burkitt lymphoma is considered one of the GC-derived lymphomas together with FL and the GCB subtype of DLBCL, due to the fact that their expression profiles resemble normal GC B cells [33]. IgM expressing B cells tend to induce the generation of mitogenic signals, not of differentiation [33]. Taken together, BL cells might have been exposed to an antigen, and having gone through an incomplete or monoallelic CSR process as supposed by their IgM expression[34].

Moreover, BL cases expressing the *V_H4-39* gene have been described to possess a mutation in codon 64 (CDR2), which is also seen, in much less frequency, in CLL [35]. Upon CDR2 mutation reversion we could detect a structural shift in all 3 antibody backbones, not only restricted to the mutated area but also affecting other portions of the ABS. Whether this mutation and the subsequent backbone shifts are enough to disturb a hypothetical antigen binding will be addressed in Chapter 4.

The stereotyped mutations appearing in the CDRs of B cells in lymphoma have been postulated to be a sign of antigen triggering of the disease, by proliferation stimulation [12]. Could the biased repertoire of *IG* gene expression and the subsequent stereotyped mutations be of ontogeny relevance in Burkitt lymphoma? The comparison to carbohydrate I/i binding site produced no relevant conserved residues in the 3 analysed BL antibodies. Yet, it has been previously demonstrated that the staphylococcal protein A (SpA, widely used for antibody purification) has a Fab fragment specificity to the V_H3 family that can "stimulate a large fraction of B cells, contributing to lymphocyte clonal selection" [22]. In our 3 analysed BL antibodies there was a partially conserved SpA binding site located in the CDR2 and HFR3. It has been noted that the localisation of SpA binding residues are located in areas not touched by SHM, suggesting a superantigenic interaction [35]. Even if the BL antibodies were of the V4 family, the SpA binding residues were conserved to those of the V_H3 family in the binding positions, explaining the cross-family interaction.

Another possible antigenic driver in BL is the Epstein-Barr virus (EBV). In EBV^+ BL there has been evidence of the virus driving BL pathogenesis by inducing B lymphocyte proliferation through CD40 signal mimicking [35]. While a correlation with EBV infection and Burkitt lymphoma in endemic cases has been shown to be 100%, sporadic BLs have a much lower EBV coexisting infection (10%) [41] and their aetiology is not yet understood.

Since there is no detected pathogen in BL cases the antigen driver for EBV^- cases remains unknown, leaving the antigenic hit and run hypothesis still open. In this light, a pathogen that would no longer be present in a full blown BL could already have manipulated B cell function, a mechanism by which B cells accumulate oncogenic characteristics [32]. B cells do not only offer a reservoir for the pathogens, but they can also alter B cell maturation and survival. A particular approach is that of antibody dilution by stimulating the production of short-lived plasma cells that produce nonspecific antibodies, effectively diluting the specific antibody concentration in the organism [32]. This mechanism, combined with the fact that BL tumoral cells seem not to be fully matured and an antigen promoting B cell overproliferation would further explain the IgM nature of the expressed B cell antibodies. Thus, there is a risk that the secreted BL antibodies would have immature and unspecific epitopes, making a causing antigen definition through epitope mapping a difficult task.

Regarding aetiology, a risk factor prediction for BL negatively correlated a history of allergy or asthma to the lymphoma occurrence in younger patients (below 50 years old) [37]. The same study revealed an association of eczema to an elevated risk of BL in younger patients and to a history of hepatitis C virus seropositivity in older patients. Another group detected a negative association between atopic disease and non-Hodgkin lymphoma (NHL) in a pooled data analysis [38], meaning a significant lower risk of presenting NHL in a population of asthma, hay fever and specific allergy (excluding eczema) sufferers. These results reinforce a hit and run pathogen hypothesis or chronic immune activation for some cases in the older age wave of sporadic BL and define an immature or partially disfunctional immune system in younger BL patients.

To conclude, we review B-cell lymphoma IGH expression bias and describe the 3D ABS structures from 3 BL mAbs with stereotyped CDR2 mutations. The triggering pathogenic effects summed to the already described mutations in BL explain the final appearance of the lymphoma as we know it [39]. However, many fronts remain open in the initial BL agent's, autologous or exogen, discovery.

Acknowledgments

The authors would like to thank Olivera Batic, Steffen Möller, Matthias Peipp and the ICGC and MMML consortiums for data usage allowance.

Web tool references:

AbYsis: http://www.bioinf.org.uk/abysis/about/definitions/definitions.html

IMGT V-QUEST: http://www.imgt.org/IMGT_vquest/vquest

UNIPROT: http://www.uniprot.org/

Translate tool: http://web.expasy.org/translate/

Rosetta Antibody Protocol: http://rosie.rosettacommons.org/antibody

Paratome: http://www.ofranlab.org/paratome/

RCSB PDB Protein Comparison Tool: http://www.rcsb.org

References

[1] Ian Magrath. The Lymphoid Neoplasms: 3ed, CRC Press, 2010.

[2] N. Chiorazzi, M. Ferrarini. B cell chronic lymphocytic leukemia: lessons learned from studies of the B cell antigen receptor. Annual review of immunology 21 (2003) 841–894.

[3] A. Hadzidimitriou, A. Agathangelidis, N. Darzentas, F. Murray, M.-H. Delfau-Larue, L.B. Pedersen, A.N. Lopez, A. Dagklis, P. Rombout, K. Beldjord, A. Kolstad, M.H. Dreyling, A. Anagnostopoulos, A. Tsaftaris, P. Mavragani-Tsipidou, A. Rosenwald, M. Ponzoni, P. Groenen, P. Ghia, B. Sander, T. Papadaki, E. Campo, C. Geisler, R. Rosenquist, F. Davi, C. Pott, K. Stamatopoulos. Is there a role for antigen selection in mantle cell lymphoma? Immunogenetic support from a series of 807 cases. Blood 118 (2011) 3088–3095.

[4] V. Bikos, N. Darzentas, A. Hadzidimitriou, Z. Davis, S. Hockley, A. Traverse-Glehen, P. Algara, A. Santoro, D. Gonzalez, M. Mollejo, A. Dagklis, F. Gangemi, D.S. Bosler, G. Bourikas, A. Anagnostopoulos, A. Tsaftaris, E. Iannitto, M. Ponzoni, P. Felman, F. Berger, C. Belessi, P. Ghia, T. Papadaki, A. Dogan, M. Degano, E. Matutes, M.A. Piris, D. Oscier, K. Stamatopoulos. Over 30% of patients with splenic marginal zone lymphoma express the same immunoglobulin heavy variable gene: ontogenetic implications. Leukemia 26 (2012) 1638–1646.

[5] S. Zibellini, D. Capello, F. Forconi, P. Marcatili, D. Rossi, S. Rattotti, S. Franceschetti, E. Sozzi, E. Cencini, R. Marasca, L. Baldini, A. Tucci, F. Bertoni, F. Passamonti, E. Orlandi, M. Varettoni, M. Merli, S. Rizzi, V. Gattei, A. Tramontano, M. Paulli, G. Gaidano, L. Arcaini. Stereotyped patterns of B-cell receptor in splenic marginal zone lymphoma. Haematologica 95 (2010) 1792–1796.

[6] Heiko Trautmann, Anastasia Hadzidimitriou, Nikos Darzentas, Monika Szczepanowski, Hilmar Berger, Claudia Philipp, Stefan Bentink, Birgit Burkhardt, Michael Hummel, Markus Loeffler, Alfred Reiter, Rainer Spang, Harald Stein, Lorenz Trümper, Michael Kneba, Reiner Siebert, Wolfram Klapper, Ralf Küppers, Kostas Stamatopoulos, Christiane Pott. Evidence for antigen-driven development of sporadic burkitt lymphomas, Manuscript in preparation.

[7] M.J. Baptista, E. Calpe, E. Fernandez, L. Colomo, T.M. Cardesa-Salzmann, P. Abrisqueta, F. Bosch, M. Crespo. Analysis of the IGHV region in Burkitt's lymphomas supports a germinal center origin and a role for superantigens in lymphomagenesis. Leukemia Research 38 (2014) 509–515.

[8] R.M. MacCallum, A.C. Martin, J.M. Thornton. Antibody-antigen interactions: contact analysis and binding site topography. Journal of molecular biology 262 (1996) 732–745.

[9] Wu T. T. and Kabat E. A. An Analysis of the Sequences of the Variable Regions of Bence Jones Proteins and Myeloma Light Chains and Their Implications for Antibody Complementarity. J Exp Med 132 (1970) 211–250.
[10] K. Stamatopoulos, C. Belessi, A. Hadzidimitriou, T. Smilevska, E. Kalagiakou, K. Hatzi, N. Stavroyianni, A. Athanasiadou, A. Tsompanakou, T. Papadaki, G. Kokkini, G. Paterakis, R. Saloum, N. Laoutaris, A. Anagnostopoulos, A. Fassas. Immunoglobulin light chain repertoire in chronic lymphocytic leukemia. Blood 106 (2005) 3575–3583.
[11] G. Tobin, U. Thunberg, A. Johnson, I. Eriksson, O. Söderberg, K. Karlsson, M. Merup, G. Juliusson, J. Vilpo, G. Enblad, C. Sundström, G. Roos, R. Rosenquist. Chronic lymphocytic leukemias utilizing the VH3-21 gene display highly restricted Vlambda2-14 gene use and homologous CDR3s: implicating recognition of a common antigen epitope. Blood 101 (2003) 4952–4957.
[12] N. Darzentas, K. Stamatopoulos. Stereotyped B cell receptors in B cell leukemias and lymphomas. Methods in molecular biology (Clifton, N.J.) 971 (2013) 135–148.
[13] Terry J. Hamblin, Zadie Davis, Anne Gardiner, David G. Oscier, and Freda K. Stevenson. Unmutated Ig VH Genes Are Associated With a More Aggressive Form of Chronic Lymphocytic Leukemia. Blood 94 (1999) 1848–1854.
[14] van Dongen, J J M, A.W. Langerak, M. Brüggemann, P.A.S. Evans, M. Hummel, F.L. Lavender, E. Delabesse, F. Davi, E. Schuuring, R. García-Sanz, van Krieken, J H J M, J. Droese, D. González, C. Bastard, H.E. White, M. Spaargaren, M. González, A. Parreira, J.L. Smith, G.J. Morgan, M. Kneba, E.A. Macintyre. Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. Leukemia 17 (2003) 2257–2317.
[15] J. Richter, M. Schlesner, S. Hoffmann, M. Kreuz, E. Leich, B. Burkhardt, M. Rosolowski, O. Ammerpohl, R. Wagener, S.H. Bernhart, D. Lenze, M. Szczepanowski, M. Paulsen, S. Lipinski, R.B. Russell, S. Adam-Klages, G. Apic, A. Claviez, D. Hasenclever, V. Hovestadt, N. Hornig, J.O. Korbel, D. Kube, D. Langenberger, C. Lawerenz, J. Lisfeld, K. Meyer, S. Picelli, J. Pischimarov, B. Radlwimmer, T. Rausch, M. Rohde, M. Schilhabel, R. Scholtysik, R. Spang, H. Trautmann, T. Zenz, A. Borkhardt, H.G. Drexler, P. Möller, R.A.F. MacLeod, C. Pott, S. Schreiber, L. Trümper, M. Loeffler, P.F. Stadler, P. Lichter, R. Eils, R. Küppers, M. Hummel, W. Klapper, P. Rosenstiel, A. Rosenwald, B. Brors, R. Siebert. Recurrent mutation of the ID3 gene in Burkitt lymphoma identified by integrated genome, exome and transcriptome sequencing. Nature genetics 44 (2012) 1316–1320.
[16] C. Vollmers, L. Penland, J.N. Kanbar, S.R. Quake. Novel exons and splice variants in the human antibody heavy chain identified by single cell and single molecule sequencing. PLoS ONE 10 (2015) e0117050.
[17] X. Brochet, M.-P. Lefranc, V. Giudicelli. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. Nucleic acids research 36 (2008) W503-8.
[18] V. Kunik, S. Ashkenazi, Y. Ofran. Paratome: an online tool for systematic identification of antigen-binding regions in antibodies based on sequence or structure. Nucleic acids research 40 (2012) W521-4.
[19] V. Kunik, B. Peters, Y. Ofran. Structural consensus among antibodies defines the antigen binding site. PLoS computational biology 8 (2012) e1002388.
[20] Jan Erikson, Janet Finan, Peter C. Nowell, and Carlo M. Croce. Translocation of immunoglobulin VH genes in Burkitt lymphoma. Proceedings of the National Academy of Sciences of the United States of America 79 (1982) 5611–5615.
[21] Swerdlow SH, Campo E, Harris NL, et al. Classification of Tumours of Haematopoietic and Lymphoid Tissues, IARC, 2008.
[22] Marc Graille, Enrico A. Stura, Adam L. Corper, et al. Crystal structure of a Staphylococcus aureus protein A domain complexed with the Fab fragment of a human

IgM antibody: Structural basis for recognition of B-cell receptors and superantigen activity. Proceedings of the National Academy of Sciences 97 (2000) 5399–5404.

[23] K.N. Potter, P. Hobby, S. Klijn, F.K. Stevenson, B.J. Sutton. Evidence for Involvement of a Hydrophobic Patch in Framework Region 1 of Human V4-34-Encoded Igs in Recognition of the Red Blood Cell I Antigen. The Journal of Immunology 169 (2002) 3777–3782.

[24] Cyrus Chothia and Arthur M. Lesk. Canonical Structures for the Hypervariable Regions of Immunoglobulins. J Mol Biol 196 (1987) 901–917.

[25] Marie-Paule Lefranc, Christelle Pommié, Manuel Ruiz, Véronique Giudicelli, Elodie Foulquier, Lisa Truong, Valérie Thouvenin-Contet, Gérard Lefranc. IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. Developmental & Comparative Immunology 27 (2003) 55–77.

[26] R. Küppers. Mechanisms of B-cell lymphoma pathogenesis. Nature reviews. Cancer 5 (2005) 251–262.

[27] L.-A. Sutton, E. Kostareli, A. Hadzidimitriou, N. Darzentas, A. Tsaftaris, A. Anagnostopoulos, R. Rosenquist, K. Stamatopoulos. Extensive intraclonal diversification in a subgroup of chronic lymphocytic leukemia patients with stereotyped IGHV4-34 receptors: implications for ongoing interactions with antigen. Blood 114 (2009) 4460–4468.

[28] F. Fais, F. Ghiotto, S. Hashimoto, B. Sellars, A. Valetto, S.L. Allen, P. Schulman, V.P. Vinciguerra, K. Rai, L.Z. Rassenti, T.J. Kipps, G. Dighiero, H.W. Schroeder, M. Ferrarini, N. Chiorazzi. Chronic lymphocytic leukemia B cells express restricted sets of mutated and unmutated antigen receptors. The Journal of clinical investigation 102 (1998) 1515–1525.

[29] Hans-Peter Brezinschek, Sandra J. Foster, Ruth I. Brezinschek, Thomas Dörner, Rana Domiati-Saad, and Peter E. Lipsky. Analysis of the Human VH Gene Repertoire: Differential Effects of Selection and Somatic Hypermutation on Human Peripheral CD5+/IgM+ and CD5-/IgM+ B cells. J. Clin. Invest. 99 (1997) 2488–2501.

[30] L.-A. Sutton, A. Agathangelidis, C. Belessi, N. Darzentas, F. Davi, P. Ghia, R. Rosenquist, K. Stamatopoulos. Antigen selection in B-cell lymphomas--tracing the evidence. Seminars in Cancer Biology 23 (2013) 399–409.

[31] Y.-C. Wu, D. Kipling, H.S. Leong, V. Martin, A.A. Ademokun, D.K. Dunn-Walters. High-throughput immunoglobulin repertoire analysis distinguishes between human IgM memory and switched memory B-cell populations. Blood 116 (2010) 1070–1078.

[32] Katharina Nothelfer, Philippe J. Sansonetti, and Armelle Phalipon. Pathogen manipulation of B cells: the best defence is a good offence. Nature reviews. Microbiology 13 (2015) 173–184.

[33] A.L. Shaffer, R.M. Young, L.M. Staudt. Pathogenesis of human B cell lymphomas. Annual review of immunology 30 (2012) 565–610.

[34] J.E.J. Guikema, C. de Boer, E. Haralambieva, L.A. Smit, van Noesel, Carel J M, E. Schuuring, P.M. Kluin. IGH switch breakpoints in Burkitt lymphoma: exclusive involvement of noncanonical class switch recombination. Genes, chromosomes & cancer 45 (2006) 808–819.

[35] F. Murray, N. Darzentas, A. Hadzidimitriou, G. Tobin, M. Boudjogra, C. Scielzo, N. Laoutaris, K. Karlsson, F. Baran-Marzsak, A. Tsaftaris, C. Moreno, A. Anagnostopoulos, F. Caligaris-Cappio, D. Vaur, C. Ouzounis, C. Belessi, P. Ghia, F. Davi, R. Rosenquist, K. Stamatopoulos. Stereotyped patterns of somatic hypermutation in subsets of patients with chronic lymphocytic leukemia: implications for the role of antigen selection in leukemogenesis. Blood 111 (2008) 1524–1533.

[36] C. Gutzeit, N. Nagy, M. Gentile, K. Lyberg, J. Gumz, H. Vallhov, I. Puga, E. Klein, S. Gabrielsson, A. Cerutti, A. Scheynius. Exosomes derived from Burkitt's lymphoma cell lines induce proliferation, differentiation, and class-switch recombination in B cells. Journal of immunology (Baltimore, Md. 1950) 192 (2014) 5852–5862.

[37] S.M. Mbulaiteye, L.M. Morton, J.N. Sampson, E.T. Chang, L. Costas, S. de Sanjosé, T. Lightfoot, J. Kelly, J.W. Friedberg, W. Cozen, R. Marcos-Gragera, S.L. Slager, B.M. Birmann, D.D. Weisenburger. Medical history, lifestyle, family history, and occupational risk factors for sporadic Burkitt lymphoma/leukemia: the Interlymph Non-Hodgkin Lymphoma Subtypes Project. Journal of the National Cancer Institute. Monographs 2014 (2014) 106–114.

[38] C.M. Vajdic, M.O. Falster, S. de Sanjose, O. Martínez-Maza, N. Becker, P.M. Bracci, M. Melbye, K.E. Smedby, E.A. Engels, J. Turner, P. Vineis, A.S. Costantini, E.A. Holly, E. Kane, J.J. Spinelli, C. La Vecchia, T. Zheng, B.C.-H. Chiu, L. Dal Maso, P. Cocco, M. Maynadié, L. Foretova, A. Staines, P. Brennan, S. Davis, R. Severson, J.R. Cerhan, E.C. Breen, B. Birmann, W. Cozen, A.E. Grulich. Atopic disease and risk of non-Hodgkin lymphoma: an InterLymph pooled analysis. Cancer research 69 (2009) 6482–6489.

[39] R. Schmitz, M. Ceribelli, S. Pittaluga, G. Wright, L.M. Staudt. Oncogenic mechanisms in Burkitt lymphoma. Cold Spring Harbor Perspectives in Medicine 4 (2014).

[40] R.M. Young, T. Wu, R. Schmitz, M. Dawood, W. Xiao, J.D. Phelan, W. Xu, L. Menard, E. Meffre, W.-C.C. Chan, E.S. Jaffe, R.D. Gascoyne, E. Campo, A. Rosenwald, G. Ott, J. Delabie, L.M. Rimsza, L.M. Staudt. Survival of human lymphoma cells requires B-cell receptor engagement by self-antigens. Proceedings of the National Academy of Sciences of the United States of America (2015).

[41] G. Brady, G.J. MacArthur, P.J. Farrell. Epstein-Barr virus and Burkitt lymphoma. Postgraduate Medical Journal 84 (2008) 372–377.

3.5 Supplementary data to Immunoglobulin rearrangement bias and stereotyped mutations in Burkitt lymphoma

	ICGC cohort					
	Heavy			Light		
	V gene	D gene	J gene	V gene	J gene	
1	IGHV3-23	IGHD3-3	IGHJ4	IGKV3-20	IGKJ1	IGK
2	IGHV3-30	IGHD6-19	IGHJ2	IGKV1D-13	IGKJ3	IGK
3	IGHV4-39	IGHD7-27	IGHJ3	IGLV1-51	IGLJ2	IGL
4	IGHV1-46	IGHD3-10	IGHJ4	IGKV4-1	IGKJ2	IGK
5	IGHV4-34	IGHD3-3	IGHJ5	IGLV1-40	IGLJ1	IGL
6	IGHV4-31	IGHD3-22	IGHJ6	IGLV2-14	IGKJ1	IGK
7	IGHV3-30	IGHD3-16	IGHJ2	IGLV1-40	IGLJ3	IGL
8	IGHV2-26	IGHD2-2	IGHJ6	IGKV1-5	IGKJ1	IGK
9	IGHV3-53	IGHD3-9	IGHJ4	IGKV2-24	IGKJ2	IGK
10	IGHV4-34	IGHD3-10	IGHJ6	IGLV1-47	IGLJ3	IGL
11	IGHV4-61	IGHD3-3	IGHJ5	IGKV3-20	IGKJ1	IGK
12	IGHV4-34	IGHD3-22	IGHJ4	IGLV1-40	IGLJ3	IGL
13	IGHV4-34	IGHD1-26	IGHJ4	IGKV1-12	IGKJ2	IGK
14	IGHV1-46	IGHD1-26	IGHJ5	IGLV2-14	IGLJ3	IGL
15	IGHV3-74	IGHD2-15	IGHJ4	IGKV2D-28	IGKJ4	IGK
16	IGHV3-30	IGHD2-2	IGHJ4	IGLV1-40	IGKJ1	IGL
17	IGHV3-48	IGHD3-22	IGHJ4	IGKV2-40	IGKJ2	IGK
18	IGHV4-39	IGHD3-22	IGHJ4	IGLV1-51	IGLJ3	IGL
19	IGHV5-51	IGHD5-24	IGHJ4	IGKV3-20	IGKJ1	IGK

Supplementary Table 3.6: List of VDJ rearrangements for the main gene of the heavy and light chain detected in Burkitt lymphoma RNA-seq cases. Cases 2, 3 and 14 had been previously included in the MMML study.

LOCUS	MMML		ICGC		Baptista et al	
	n	%	n	%	n	%
IGHV1-18	2	3.70			1	1.39
IGHV1-2	1	1.85			1	1.39
IGHV1-24					1	1.39
IGHV1-46			1	6.25	1	1.39
IGHV1-58	1	1.85				
IGHV1-69					1	1.39
IGHV1-8	2	3.70			2	2.78
IGHV2-26			1	6.25		
IGHV2-5	2	3.70			1	1.39
IGHV2-70					1	1.39
IGHV3-11					2	2.78
IGHV3-15	3	5.56			2	2.78
IGHV3-21	2	3.70			4	5.56
IGHV3-23	3	5.56	1	6.25	6	8.33
IGHV3-30	5	9.26	2	12.50	7	9.72
IGHV3-30-3	1	1.85			1	1.39
IGHV3-33					1	1.39
IGHV3-43					1	1.39
IGHV3-48	1	1.85	1	6.25	5	6.94
IGHV3-49	1	1.85				
IGHV3-53	1	1.85	1	6.25	1	1.39
IGHV3-64	1	1.85				
IGHV3-66					2	2.78
IGHV3-7	1	1.85			1	1.39
IGHV3-74	1	1.85	1	6.25	2	2.78
IGHV3-9	1	1.85				
IGHV4-30	1	1.85				
IGHV4-31	1	1.85	1	6.25	1	1.39
IGHV4-34	11	20.37	4	25.00	13	18.06
IGHV4-39	6	11.11	1	6.25	6	8.33
IGHV4-4	1	1.85			1	1.39
IGHV4-59	3	5.56			5	6.94
IGHV4-61			1	6.25		
IGHV5-51	1	1.85	1	6.25	1	1.39
IGHV5-a					1	1.39
IGHV6-1	1	1.85				

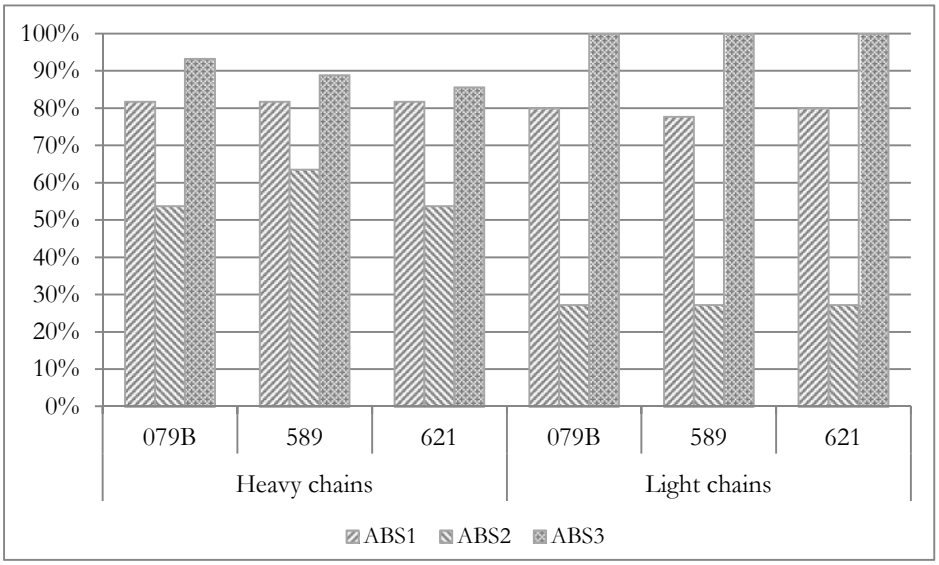
Supplementary Table 3.7: IGHV rearrangements detected in three study cohorts for Burkitt lymphoma.

IGHV GENE	n	%
IGHV1-18	2	3.70
IGHV1-2	1	1.85
IGHV1-58	1	1.85
IGHV1-8	2	3.70
IGHV2-5	2	3.70
IGHV3-15	3	5.56
IGHV3-21	2	3.70
IGHV3-23	3	5.56
IGHV3-30	4	7.41
IGHV3-30-3	2	3.70
IGHV3-48	1	1.85
IGHV3-49	1	1.85
IGHV3-53	1	1.85
IGHV3-64	1	1.85
IGHV3-7	1	1.85
IGHV3-74	1	1.85
IGHV3-9	1	1.85
IGHV4-30-4	1	1.85
IGHV4-31	1	1.85
IGHV4-34	11	20.37
IGHV4-39	6	11.11
IGHV4-4	1	1.85
IGHV4-59	3	5.56
IGHV5-51	1	1.85
IGHV6-1	1	1.85
Total	**54**	**100**

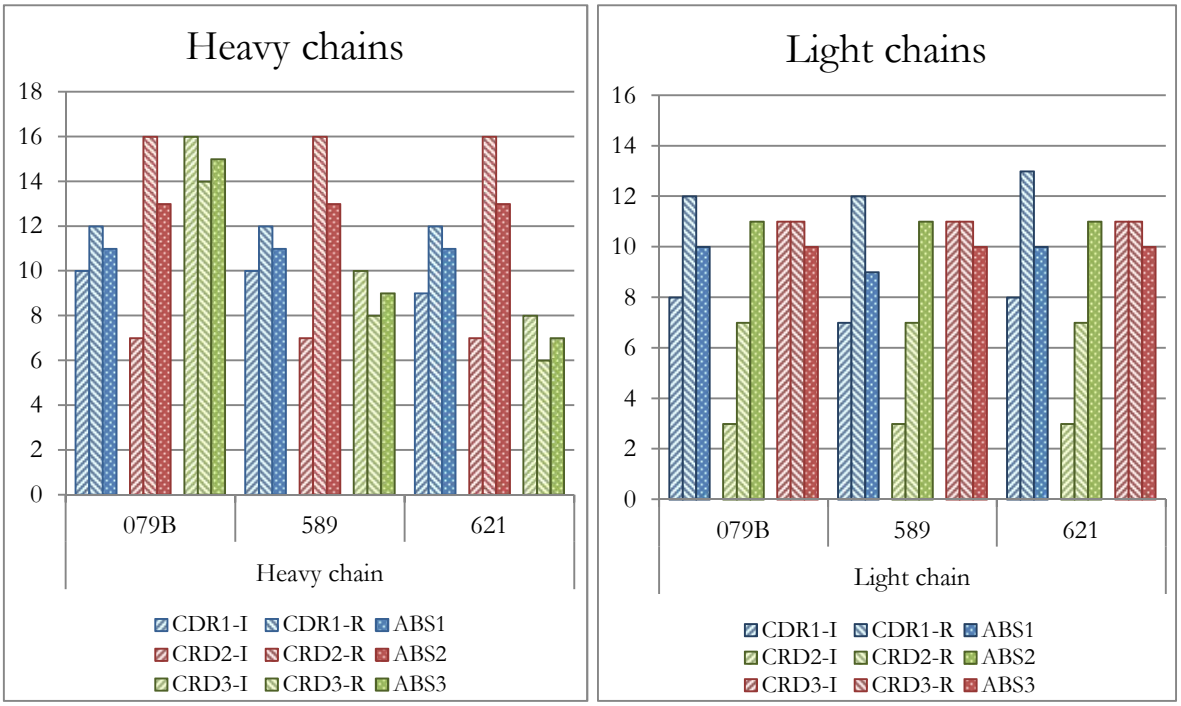
Supplementary Table 3.8: Rearrangements found in 54 BL cases. MMML.

	079	589	621
Gender	Female	Male	Male
Ref. diagnostic	sBL	High-grade nos	BL-like
Mol. diagnostic	mBL	mBL	mBL
CD20	+	+	+
CD10	+	+	+
BCL2	-	-	-
BCL6	+	+	+
KI67	90%	100%	95%
MUM1	+	-	-
TdT	-	-	-
t(8:14)	+	+	+
T cells	5%	n.a.	n.a.
Light chain	k	k	l
Localisation	Intraabd. LN	Left maxillary	Cervical LN
Hepatomegaly	-	-	-
Splegnomegaly	+	-	-
EBV	-	-	-
IgM	n.a.	+	+
Staging	II	IV	II
Treatment	NHL-BFM 95	NHL-BFM 95	NHL-BFM 90
FU time (days)	3538	911	4936
Relapse	No	No	No

Supplementary Table 3.9: Patient's clinical data. Ref. diagnostic: reference diagnostic, sBL: sporadic Burkitt lymphoma, nos: not otherwise specified, Mol. diagnostic: molecular diagnostic, mBL: molecular Burkitt lymphoma, n.a.: not available, k: kappa light chain, l: lambda light chain, Intraabd.: Intraabdominal, LN: lymph node, Staging: Murphy staging, FU time: follow up time.

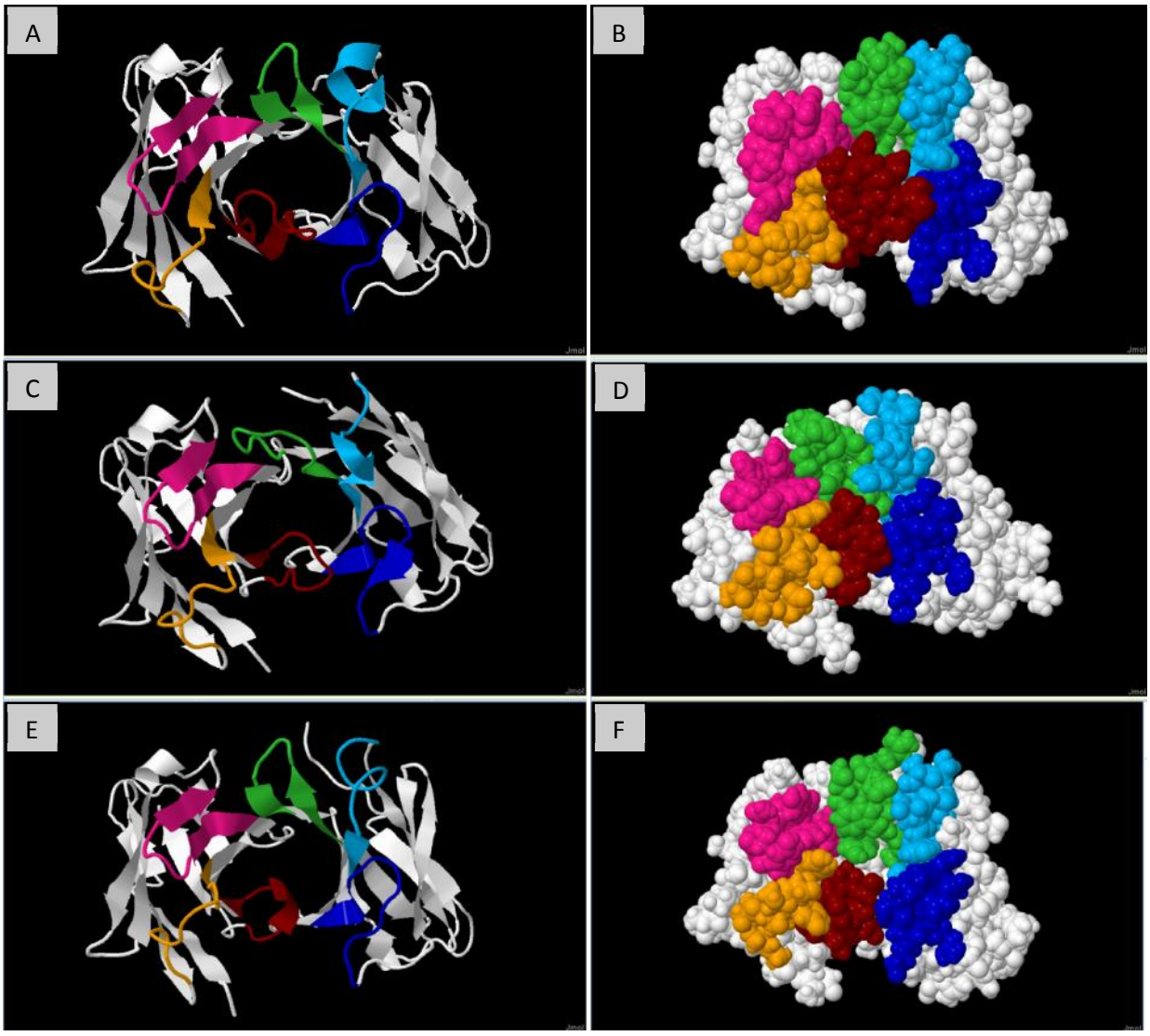


Supplementary Figure 3.7: Overall overlap of the CDRs to the detected Paratome ABS. The overlap differences are consistent between mAbs. There are similarities across antibodies but differences in the consensus level between regions. Antigen binding site 3 in the light chains correlates very well with the CDR regions detected by both IMGT and Rosie. The least similar comparison is seen in the ABS2 in the light chain. Overall there is generally more than 50% overlap in all regions.



Supplementary Figure 3.8: Length of the 3 BL antibody CDR and ABS. This graph shows the length variation through different CDR and ABS annotating methods. The annotations from IMGT are marked with a left-down to right-top diagonal, the ones made by Rosie are marked with opposite diagonal lines and the bars with points on them represent the

ABS region lengths. They are grouped by the region and antibody, so the differences between numbering techniques can be compared. Overall CDR2 seems to be the less homolog numbering in the heavy and light chains. CDR1 has also some significant discordances, mainly in the light chains.



Supplementary Figure 3.9: Top views of the ABS definition by Paratome. A-B: 079, C-D: 589 and E-F: 621. Ribbon (left) and spacefill (right) ABS representation

Chapter 4

Putative antigenic epitopes of CDR2 stereotyped-mutated Burkitt lymphoma antibodies

Neus Masqué-Soler[1], Monika Szczepanowski[1], Andreas Recke[2], Frank D. Sönnichsen[3], Wolfram Klapper[1].

[1] Department of Pathology, Haematopathology Section and Lymph Node Registry, University Hospital Schleswig-Holstein, Christian-Albrecht University, Kiel, Germany.
[2] Department of Dermatology, Allergology and Venereology, University of Lübeck, Lübeck, Germany; Institute of Experimental Dermatology, University of Lübeck, Lübeck, Germany.
[3] Otto Diels Institute of Organic Chemistry, Otto-Hahn-Platz 3/4, Christian-Albrecht University, Kiel, Germany.

Manuscript in preparation.

Abstract

Sporadic Burkitt lymphoma (BL) aetiology is still eluding discovery. Several B-cell lymphomas have been related to auto or superantigenic drive or pathogen influence. While antigenic screening is demanding and can result in unspecific matches, we choose an antibody side approach to determine possible antigenic epitopes. We analyse three paediatric BL antibodies from patients expressing the same biased immunoglobulin gene family. By using phage display techniques and *in silico* modelling methods we describe BL binding peptides. Better understanding of lymphomagenesis can help improve prevention and treatment procedures for these diseases.

Keywords: Burkitt lymphoma, antigenic drive, antibody stereotypy.

4.1 Introduction

Burkitt lymphoma (BL) is a non-Hodgkin lymphoma of mature B-cells considered as aggressive. It affects a paediatric to young adult population with a later middle-aged patient wave. Patient age is a factor upon treatment decision [1,2] but also defines some pathogenic characteristics. The hallmark translocation between chromosomes 8 and 14, involving the *C-MYC* oncogene and the immunoglobulin heavy chain locus (IGH@) is common in both age subgroups. Yet, up to 31% of paediatric diffuse large B-cell lymphoma (DLBCL) cases were reclassified as molecular BL in recent gene expression profiling studies [3]. Paediatric BL also expresses the IGHV4-34 and IGHV4-39 rearrangements more frequently than adult BL, the same types that have been seen to carry stereotyped mutations (see Chapter 3 of this thesis). Moreover, epidemiologic studies have demonstrated a negative correlation

between allergy (and other over-activating immunological diseases) and BL [7]. The same study concluded that adult and paediatric BL might be aetiologically different.

Of the three main types of BL (endemic –eBL-, sporadic –sBL- and immunodeficiency-associated), the sporadic version is the one where a possible pathogenic aetiology is still eluding discovery. The endemic variant is present in the equatorial or malaria belt zones, where it is strongly associated to the Epstein-Barr virus (EBV) infection. The same broad geographical distribution is shared by *Plasmodium falciparum* infection rates [4] but a much clearer causal relationship is found with EBV association. The third variant is explained by the immunodeficiency of its suffering patients and, again, by mostly co-occurring with EBV infection [5].

Even though the similarities between BL subtypes are subtle, some of their characteristics point to interesting questions. Why is the location of sBL and eBL restricted mostly to abdominal and maxillary areas respectively? Indeed, an exposure to the *Euphorbia sp.* plant has been linked to activate EBV replication [6], which would explain antigenic pressure in eBL. As the appendix is the analogous sBL location of maxillary eBL it has been postulated that the intestinal Peyer patches could play an important role in early pathogenic stages of sBL. Peyer patches are lymphoid tissue aggregates of the small intestine mucosa, where sBL arises.

Yet, sBL pathogenic aetiology is unknown. With little to none infectious agents found to correlate with the disease (EBV infection co-occurs in only 10% of sBL), extraneous or autologous superantigens such as protein A of *Staphylococcus aureus* (SpA) and blood antigen I/i could be a possible explanation. These same superantigens have been seen to bind CLL antibodies [8] and are now postulated to play a role in the lymphoma onset.

The biased immunoglobulin (IG) repertoire and stereotyped mutations found in BL could be a mark of antigen drive [9] as postulated for CLL [8,10]. We chose a set of three antibody sequences of the IGHV4-39 family expressed in paediatric BL patients. By producing them *in vitro* we could analyse their putative epitopes and thus infer possible pathogenically relevant antigens.

With a better understanding of lymphoma aetiology new approaches in prevention [11] and treatment could result in less malignancy events and better treatment options.

4.2 Materials and methods

4.2.1 Original samples and cDNA

Samples and cDNA generation were processed as previously [Chapter 3, Section 3.2.2].

4.2.2 Vector cloning

Vector pCR2.1 from the Topo TA cloning kit was used (Thermo Fisher Scientific) to amplify the immunoglobulin sequences needed in TOP 10 *E.coli* competent cells (Thermo Fisher Scientific) following the company's manual. After clone amplification and plasmid extraction the plasmids were sequenced, controlled and amplified via PCR.

The vectors to be used for mammalian cell transfection, were amplified in One Shot TOP10 Chemically Competent E.coli (Thermo Fisher Scientific), their DNA extracted with PureLink® Quick Plasmid Miniprep Kit (Thermo Fisher Scientific), and measured with NanoDrop and fotometer. The vectors were modified in order to make insertion to the transfection vector (pFUSE) easier.

To know the quantity of each vector to be used in each ligation reaction a molarity ratio had to be calculated. The vectors were ligated and used to transform OneShot Chemically Competent Cells (Thermo Fisher Scientific) according to the company's guidelines. The transformed bacteria were amplified, their DNA extracted and sequenced. Finally the multiple cloning sites were deleted out of the pFUSE vectors.

Full protocols can be found in the Supplementary Methods section.

4.2.3 Transfection experiments and protein purification

HEK-293 cells (ACC305, DSMZ) were cultured in complete medium (RPMI, 10% FBC, S/P) at 37°C, and 5% CO_2 and used for transfection. Transfection tests were made with a GFP-containing plasmid (pmaxGFP®, Lonza) and to assess the following methods: lipofection with TransIT (Mirus) and Lipofectamine 2000 (Thermo Fisher Scientific), and electroporation with Amaxa's Nucleofector (Lonza) following the manufacturer's protocols. Cell viability and transfection performance were assessed via DAPI-stained cytospins.

Further transfection tests were tested (H:L:pAdvantage): 1:4:1, 2:3:1, 1:4, 1:4Ø, in complete RPMI medium [10% Ultra Low IgG FBS (Thermo Fisher Scientific), without antibiotics]. Cells and supernatants were collected for protein purification.

Cell pellets were processed for protein extraction, and IgG purification was carried out for all supernatants. IgG fractions were dialysed and concentrated. All supernatants and protein extracts were measured for protein concentration. For extended protocols please see the Supplementary Methods and Buffer list section.

4.2.4 Western blots

Reducing Western blots were made with protein extracts and supernatants from the transfected HEK-293 cells. Native Western blots were performed for the 1:4:1 and 72h

ratio and protein G-purified supernatants except for case 079A due to lack of detectable protein in the supernatants, for such case protein extracts were used.

Detailed protocols can be found in the Supplementary Methods and Buffer list section.

4.2.5 Cell blocks and IHC stainings

The cyto block (or cell block) protocol aims at fixing the cultured cells and embedding them in paraffin. Cell blocks were generated from all the transient transfection variations (079B, 589, 621 and mock transfection with empty vectors) and from other lymphoma cell lines: L428 (Hodgkin lymphoma), BL-2 and EB-1 (BL), Oci-ly7 and Rck-8 (diffuse large B-cell lymphoma, DLBCL) and MLMA (hairy cell lymphoma). All cell block slides were cut at 5μm thickness and stained with the supernatants of a 72h-old transfection, including the supernatant of the empty vector transfection as control. To confirm *IG* expression the transfected cell blocks were also stained with an anti IgG_{2a} antibody (Invitrogen) and an anti-IgL antibody (Abcam).

HEK-293 transfected cells were grown on sterile coverslips, which were attached to glass slides and fixed with acetone. After allowing drying they were stained with the purified IgG fractions of all transfection variants to screen for auto or cross reactivity. Each coverslip was stained with a total of 50μl of approximately 350μg/ml concentrated antibody (17.5μg of antibody per slide). Control stainings were made with anti-mouse IgG_{2a} antibody (Thermo Fisher Scientific) at 1:100 dilution.

Three brain biopsies from DLBCL and Peyer patches present in six formalin-fixed paraffin embedded (FFPE) biopsies were stained with supernatants obtained from the transient transfections.

In cryo preserved material we could stain the original patient's biopsies with their newly generated BL mAbs, including cross-combinations between patient mAbs, and with μ, λ (lambda) and κ (kappa) chain antibodies. An unrelated appendix, one small intestine and one tonsil biopsy were also stained with the transfection supernatants.

All stainings included a secondary antibody control. Staining protocols can be found in the Supplementary Material and Methods.

4.2.6 Phage display

Phage display is established with PhD12 library (NewEngland Biolabs). For full method specifications refer to the Supplementary Material and Methods section. For each mAb two phage display experiments were performed as repeats. Each experiment was followed by an

enzyme-linked immunosorbent assay (ELISA), where the reactivity between the mAbs and the selected phages was tested. The resulting positive wells after the ELISA were extracted for DNA and further sequenced. Some negative wells were processed as well as controls. Sequencing reactions were performed with the M13 96-primer from the PhD12 kit. Phage DNA sequences were reviewed and the peptide coding regions were translated *in silico* (Expassy, http://web.expasy.org/translate/) to obtain the peptide each sequence codes for. All sequences found are listed in the Supplementary Table 4.2.

Recurrent peptides were selected to be ordered in a synthetic primer facility (Eurogentec, Belgium). All peptides were labelled with $CONH_2$ in the C-terminus to mimic internal peptide structure. Aliquotes of the same peptides with biotin marking were used for ELISAs. A linker (Ahx, an inert six carbon linear aminohexanoic chain) between the peptide sequence and the biotin mark was used to minimise possible folding interferences. Random peptide QSHWFVPWQDRW was generated by randomising the most recurrent peptide sequence VHWDFRQWWQPS to serve as a negative control in binding assays.

4.2.7 *In silico* docking and structural analysis

The online tool pBLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) was used to perform alignments. For phage protein-peptide alignments the threshold was set to 100 and the option for short alignment optimisation was activated. Other online resources are also used, such as: ExPASy (http://web.expasy.org/translate/) for nucleotide-to-amino acid translation, and IMGT/V-QUEST (http://www.imgt.org/IMGT_vquest/share/textes/) for productivity, structure and sequence determination of our antibody sequences.

All peptides identified in the phage display technique were analysed for selection advantages and / or randomness through information indexes by the info program (http://www.northeastern.edu/xray/downloads/info-program/) [12]. Two files were submitted: an unselected phage display result containing random peptide sequences and the sequences obtained with a selective procedure. The program was able to assign a random probability number to each peptide sequence. The probability was then transformed to an associated information parameter. The greater the information content measure, the more probable is that the sequence was detected due to a real target binding. A further control step was done by submitting the 12mer peptide sequences to the MimoDB (http://immunet.cn/mimodb/) [13] database, where peptides were checked for known motives or sequences, and thus for possible contaminations of the phage display experiments. The sequences were also screened for other known motives by reviewing published works [14,15]. Peptide sequences are screened for DNA and protein motifs with the MEME Suite, version 4.10.2 (http://meme-suite.org/)

Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) was used to align the various sequences obtained to each other and to other published Burkitt lymphoma-related epitopes.

Rosetta (http://rosie.rosettacommons.org/antibody) [16] was used to model the 3 BL mAbs. The resulting .pdb files were fed to CABSdock (http://biocomp.chem.uw.edu.pl/CABSdock), a web-based tool [17] to model the three BL mAbs binding to the main peptides found with a recurrence in the phage display experiments. The contact cut-off for the analysed peptide-antibody contacts was set at 4.5 Å. CABSdock groups modelling trajectories by hierarchical clustering. The models are divided into 10 groups according to their structure similarities. For each cluster an average sample is chosen by the program as the representative model. Clusters were evaluated for their density, which is the number of elements divided by average cluster root mean square deviation (RMSD). The average distance of the Cα atoms in all models of a cluster defines RMSD. Thus, the greater a cluster density is, the more homogenous its model components are. Therefore cluster density and binding site per peptide were taken into account upon model reviewing. To evaluate realistic binding sites models where only one chain was interacting with the peptide were excluded, as well as models where interactions were located opposite from the ABS.

Two visualising software packages were used to interpret the .pdf output files of the different modelling experiments: The PyMOL Molecular Graphics System, Version 1.7.4.5 (Schrödinger, LLC) and Discovery Studio Version 4.1 (Accelrys, Biovia). PyMOL was also used to mark the conserved binding residues for the superantigen staphylococcal protein A (SpA).

4.2.8 mAb-peptide ELISAs

To prove that the synthetic peptides bound in a specific way to the mAbs an affinity ELISA per interaction was done. Cases 079 and 589 were hybridised with peptides VH, GL and QS, while case 621 was tested with peptides GL and QS. For protocol specifications refer to Supplemental Material and Methods.

4.2.9 Nuclear Magnetic Resonance

Purified mAbs binding to synthetic peptides were analysed with NMR [18]. For detailed methodology see Supplemental Material and Methods. Broadly, mAb and peptide solutions were irradiated to selectively saturate mAb resonances. The principle is based in irradiating the mAb within the peptide solution to liberate a small part of the peptide, should the latter be bound to the mAb. The disruption can be detected by comparing the pre- and post-irradiation spectra with saturation transfer difference (STD) experiments.

4.2.10 Statistical tests

Statistical experiments were carried out with the transfection experiment data. One-way ANOVA (Kruskal-Wallis test) was done to compare the three different methods in cell viability and positivity. Two-tailed Mann Whitney tests were done to compare effectiveness in pairs of transfection methods against each other as well as region of preference in peptide-to-antibody binding. For ABS-FR binding comparisons Mann Whitney tests were performed for the mutated and the germline populations independently. The confidence interval was always 95% and software used was GraphPad Prism, version 5.03.

4.3 Results

4.3.1 Lipofection transfection is more efficient than electroporation

Three different transient transfection reagents were tested in the mammalian cell line HEK-293 T (Figure 4.1). There were statistical differences in the transfection efficacy between both the lipofection and the electroporation method: Amaxa vs TransIT p= 0.0022, Amaxa-Lipofectamine p=0.0411, Lipofectamine-TransIT p=0.6991. When comparing all three methods for cell viability there were no significant differences observed (p=0.9897), while there were differences in positively transfected cell quantity, or efficacy (p=0.0148).

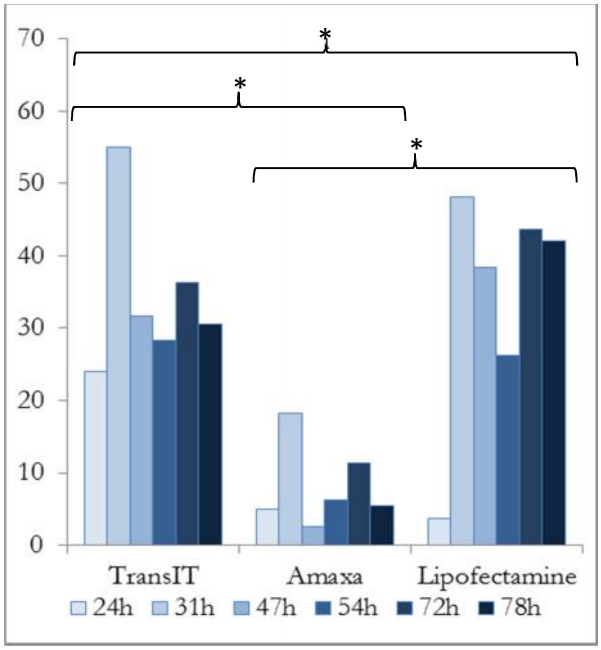


Figure 4.1: Transient transfection comparison of the three protocols assessed.

Expression of *IG* was confirmed with SDS- and native PAGE (Supplementary Figure 4.11) from both protein extracts and transfection supernatants of HEK-293 cells. While case 079A was producing IG it was not being secreted and the light chain

production was considerably less than the other cases. Cases 079B, 589 and 621 could secrete IG as it could be detected in their supernatants. The most productive ratio and incubation time was chosen to be 1:4:1 after 72h of transfection after Western blot and protein concentration measurement.

4.3.2 Burkitt lymphoma antibodies show cross- and self-recognition in immunohistochemistry

Heavy and light chain production was further proven with cell block staining of the transfected HEK-293 cells. All tested blocks were positive for both the heavy and light chain expression, with the exception of the empty-plasmid transfected cells (data not shown). Original cryo material was positive for IgM in the three cases, case 079 and 621 expressed λ and 589 showed κ expression. When stained with the transfection supernatants the 3 BL patient original material showed a cross-positivity on the biopsy material from cases 589 and 621 when stained with supernatant from the 589 transfection (Figure 4.2).

None of the analysed cell line blocks of four different lymphoma entities presented binding when stained with the transfection supernatants of either case. However, there was some cytoplasmic positivity detected in the transfected HEK-293 cell blocks when stained with their own and the other cases' transfection supernatants. Untransfected cell blocks remained negative when stained with transfection supernatants of all three cases (Supplementary Figure 4.12).

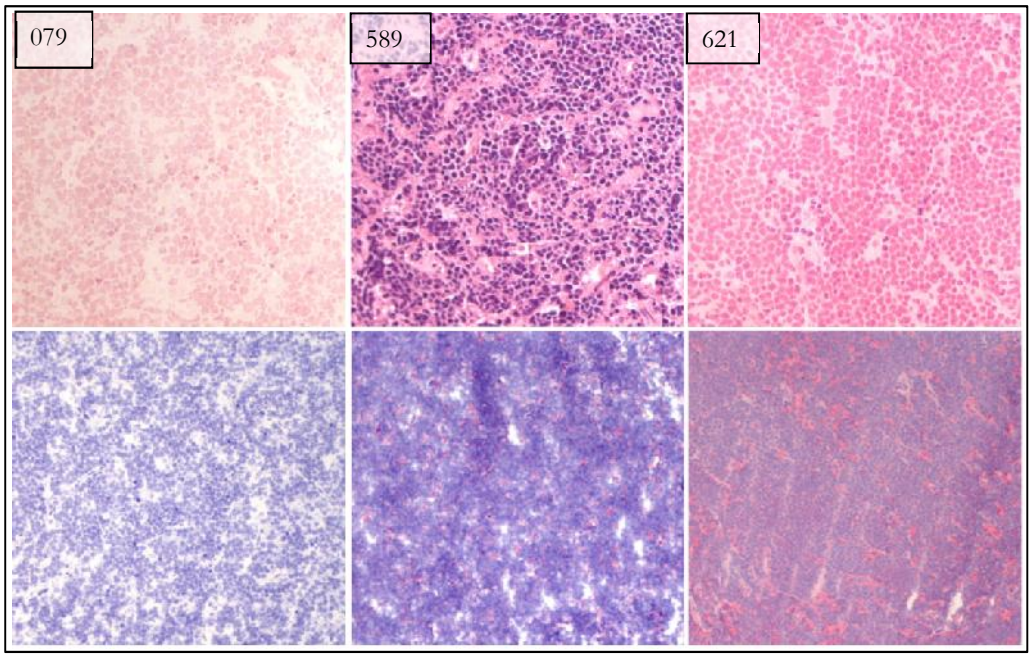


Figure 4.2. HE stainings of the original cryo BL biopsies are represented in the first row. Underneath each HE image the corresponding case stained with supernatant of the 589 transfection is represented. All images are at 10x.

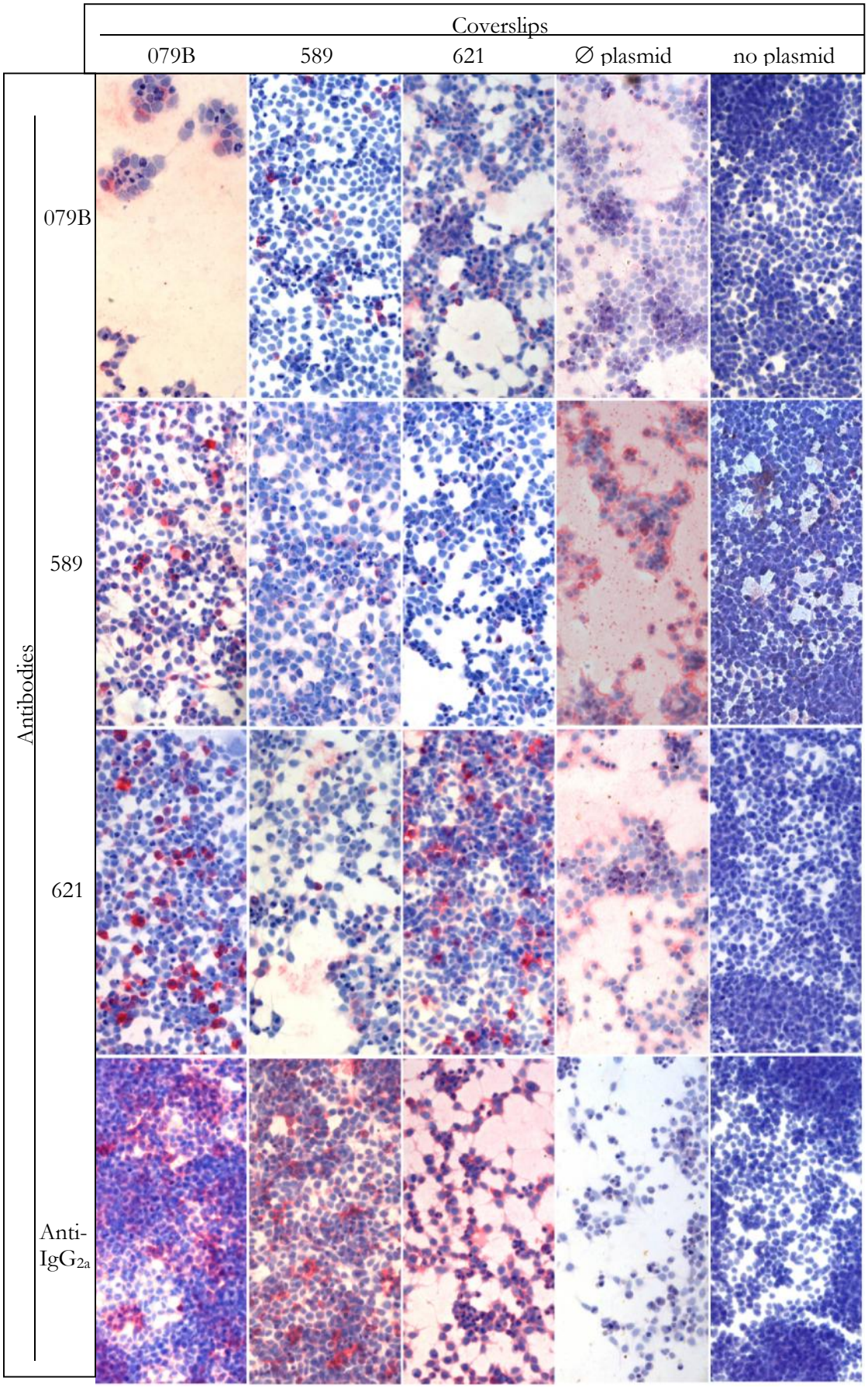


Figure 4.3. Rows represent the purified IgG antibodies used to stain the corresponding coverslips containing fixed transfected cells (columns). Al pictures at 20x.

The transfected HEK-293 cells grown in coverslips and stained with transfection supernatants showed cross-reactivity between the three BL mAbs clearer (Figure 4.3). All three mAb presented cross- and self-reactivity when applied to the transfected cells in the coverslips in variable levels throughout. Of note, the transfection supernatants stained cells that had been only transfected with an empty vector, only containing a mouse constant region (Figure 4.3, ∅ plasmid column). Cells that had been transfected with no plasmid (mock control) showed no positivity when stained with any of the 3 purified mAbs. Anti-IgG_{2a} staining showed antibody presence in membrane and cytoplasm.

The transfection supernatants were used to stain other tissues like gastrointestinal biopsies (6 Peyer patches in FFPE, 1 fresh-frozen appendix, and 1 fresh-frozen small intestine), one tonsil in fresh-frozen state and 3 brain biopsies corresponding to two different patients. The gastrointestinal biopsies of Peyer Patches showed some unspecific nuclear reactivity with 621 (Supplementary Results, Figure 4.13), also present in villi epithelia in the same sample. Otherwise, no Peyer patch lymphocytes were specifically recognised by any of our 3 mAbs. While the fresh-frozen tonsils, appendix and small intestine showed some unspecific positivity, it was only when stained with the 589 supernatant and in the small intestine sample where it was deemed specific by our pathology reviewer. The reactivity seen in the small intestine biopsy was located in neuronal cells (Figure 4.4, A). To confirm the nervous cell affinity of the 589 mAb three brain DLBCL cases were stained with the 3 mAb supernatants. Again only sample 589 showed affinity to the tissues (Figure 4.4, B).

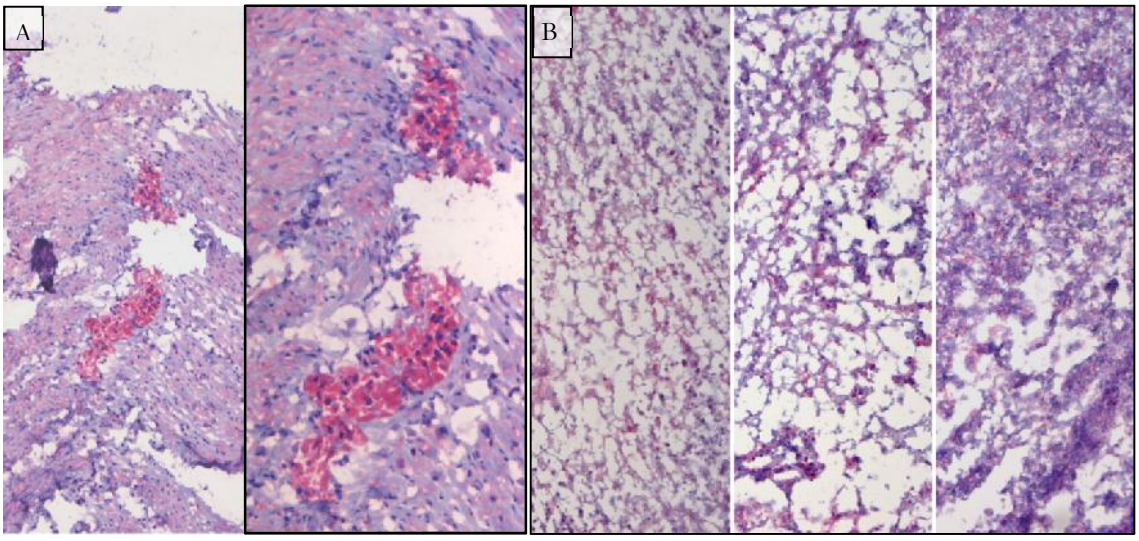


Figure 4.4. 589 transfection supernatant staining of different cryo preparations. **A**: Small intestine cryo section. The main image is at 10x and the framed detail framed is at 20x. **B**: Brain biopsies of DLBCL patients. All images are at 10x.

4.3.3 Recurrently interacting peptides to Burkitt lymphoma antibodies

A total of 104 sequences were read throughout the 6 phage display experiments, amounting to a total of 59 different peptide sequences. The majority of the sequenced phages were chosen out of positive ELISA wells (82/104, 79%, Figure 4.5, A) but the phages in 22 negative wells (21%) were also sequenced to control due to signals being weak. Supplementary Table 4.2 and Figure 4.14 summarize all peptides translated from the 104 sequencing reactions performed.

Target-specific binders are molecules that bind specifically to the antigen-recognising part of a particular antibody. Target-unrelated peptides (TUP) bind to other antibody regions instead, such as their constant fragments, or to elements of the screening system (plates, streptavidin, protein G, etc.). On that note, the resulting phage display 12mer sequences were screened for known motives such as plastic binders and selective advantage sequences by reviewing literature on the matter [14,15]. Motive SSL, a TUP, was found within the peptide sequences, such as SGV**SSL**RGNDVR and HT**SSL**WHLFRST. Peptide SGVSSLRGNDVR was interestingly sequenced out of a negative ELISA well, while peptide HTSSLWHLFRST had a low recurrence, having being detected once in a negative and once in a positive ELISA well. Plastic binder motives were also detected in the overall 12mer sequences, mainly those containing a tryptophan (W) pair, separated by one, two or three random amino acids. Specifically, motives with two W separated by one (WXW) or two (WXXW) different amino acids were found in six independent sequences (Supplementary Table 4.2).

The total 59 different sequences collected in the six phage display rounds were screened with BLAST for phage protein detection. The peptide sequences were also specifically BLASTed against the nine M13 phage proteins to control for biased phage selection. Of concern were the phage coating proteins or presented in its exterior (proteins III, VI, VII, VIII and IX). Only the alignment to proteins I, III and IV gave matches of any kind. While proteins I and IV gave short hits (5 or less amino acids in common and non-consecutive), protein III presented two interesting alignments to peptides **FIP**FD**P**M**SM**RWE and D**PFWSH**PCVLSW. Although the latter only appears once and is not further considered peptide FIPFDPMSMRWE is one of the recurrent ones, possibly being its selection due to phage binding, even if only partially.

Information content of only the positive ELISA sequences from the phage display experiments were generated (Figure 4.5, B) with the INFO software, which estimates the likelihood of random occurrence of given sequences. It yielded a normal distribution and overall a high average information content, pointing to a correct and specific phage display screening. Contamination though the different phage display runs was thus ruled out. A further control procedure was performed with MimoDB database, which aims to be a virtual tool for phage display experiment control. Its database consists of peptide sequences

with known targets. At the time of analysis Version 4.3 contained 22418 peptide sequences, 2463 mimotope sets, as well as unique peptides, targets, templates and libraries. No sequences retrieved in the phage display experiments for the 3 BL mAbs gave back a known peptide of relevance.

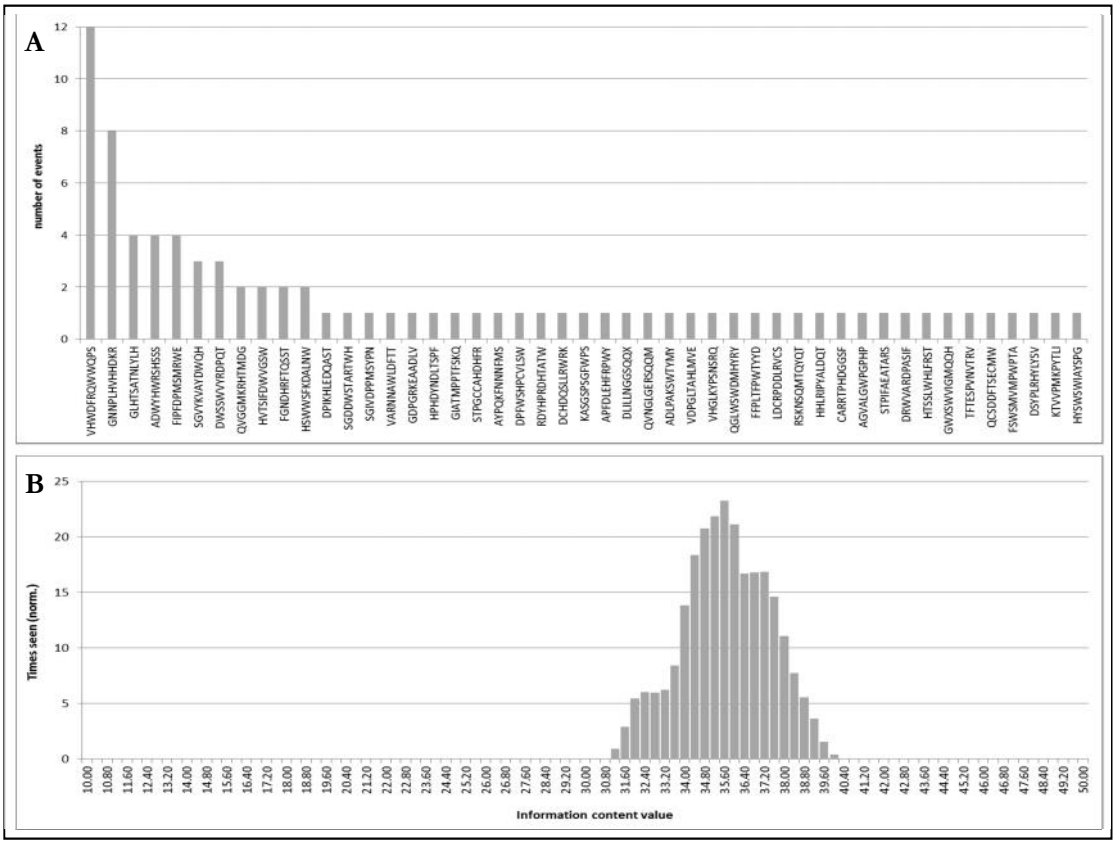
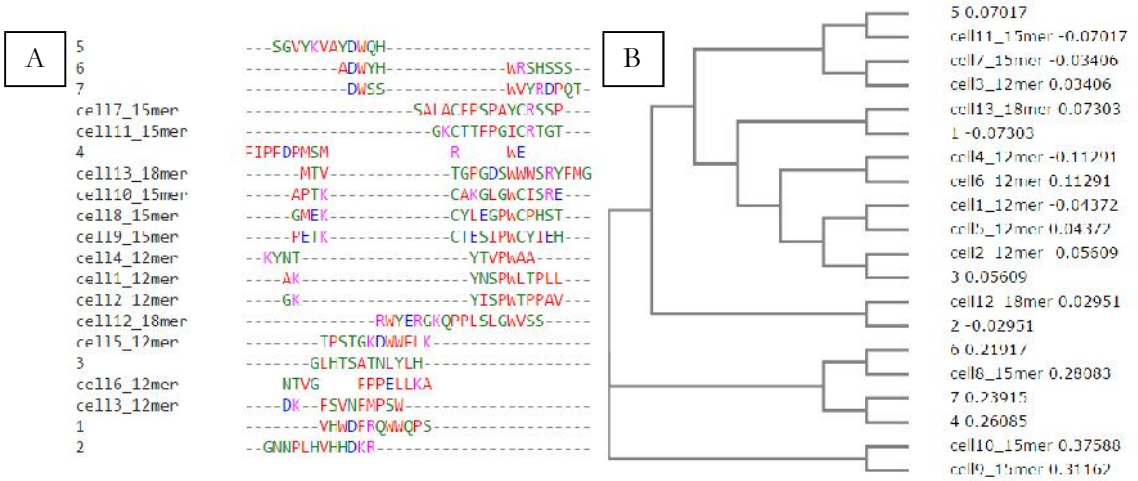


Figure 4.5. **A**: Peptide sequences retrieved from positive post-phage display ELISAs. **B**: Information content distribution of the peptides found in positive ELISA wells. The resulting histogram is normalised to 100 peptides.

All 59 peptide sequences were clustered in ClustalOmega including those found by another group studying BL cell line epitopes [19]. The resulting multiple sequence alignment showed no relevant similarities, either in sequence or amino acid properties (not shown). Another alignment is performed with only the most recurrent peptides found with interaction to the 3 analysed BL mAbs (Figure 4.6) and it is coloured by ClustalOmega's amino acid coding. The small and hydrophobic residues are marked with red colour (Als, Val, Phe, Pro, Met, Ile, Leu, and Trp). In blue are the acidic amino acids (Asp and Glu) and in magenta the basic ones (Arg and Lys). Finally, glycine and the amino acids containing amine, hydroxyl and sulfhydryl groups are labelled in green (Ser, Thr, Tyr, His, Cys, Asn and Gln). After domain reviewing of all peptide sequences with the MEME tool (http://meme-suite.org/) no conclusive results were produced, with only a vague DNA motif and several infrequent amino acid domains detected (a representation can be found in Supplementary Figure 4.15, A and B).



C

	Sequence	Occurrence	079	589	621
1	VHWDFRQWWQPS	12	✓	✓	
2	GNNPLHVHHDKR	8			✓
3	FIPFDPMSMRWE	4		✓	✓
4	GLHTSATNLYLH	4	✓	✓	✓
Random1	QSHWFVPWQDRW	-			

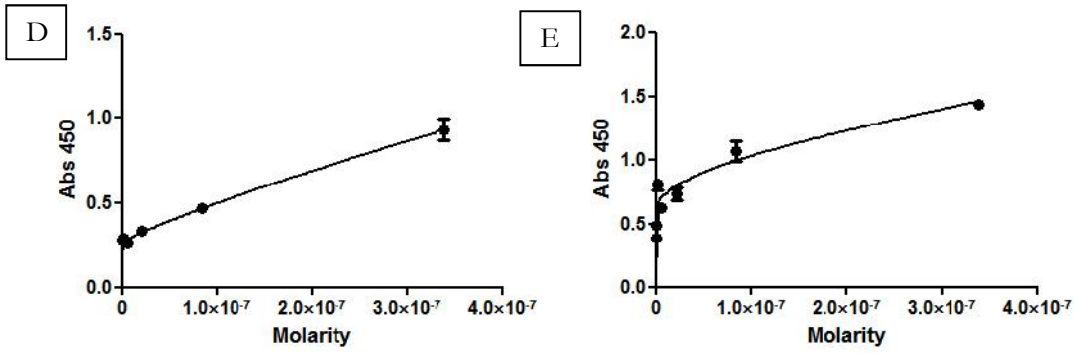


Figure 4.6. Peptide sequence alignments and bindings **A**: Peptide alignments and **B**: cladogram of recurrent putative epitopes of BL, named "cell-" [19]. Numbers 1-7 mark the most recurrent peptides in the phage display experiments performed in this study. **C**: All synthetic peptides tested either in NMR or affinity binding in ELISA and their occurrence in the phage display experiments. Peptide named random1 is the randomised peptide from number 1, used as control. **D** and **E**: Specific binding for peptide-589 ELISA results for peptides VH and GL respectively.

The four most recurrent peptides that were not suspected of having suffered a selection bias (number 1-4 in Figure 4.6, C) were analysed in pBLAST with non-redundant protein sequence databases (including PDB, SwissProt, PIR and PRF). As the peptides are only 12 amino acids in length the E value had to be taken as a broad estimate of match significance. Randomised peptide QSHWFVPWQDRW (QS) showed BLAST results of E value >100 after already the fourth hit, therefore in the subsequent analysed peptides 100 was taken as an ultimate maximum for hits to be considered. In the case of peptide

VHWDFRQWWQPS (VH) there was a predominance of putative human pathogens like Enterobacteriaceae, Pseudomonadaceae, Streptomycetaceae, Mycobacteriaceae, Lactobacillaceae, Aeromonadaceae and Vibrionaceae, all with an <50 E value (Supplementary Table 4.3). The majority of hits represented cytoplasmic proteins (data not shown). There were hits for specific proteins of the regular melon (*Cucumis melo*), the cotton plant (*Gossypium raimondii*) and the common housefly (*Musca domestic*a). Results for peptide GNNPLHVHHDKR yielded high E values (>81) from the first hit (Supplementary Table 4.4). However, nuclear and cytosolic proteins of known human pathogens appeared in the list, such as *Microbacterum sp.*, *Cedecea sp.* (Enterobacteriaceae), *Streptococcus sp.*, *Streptomyces sp.* and *Candida albicans*. For peptide FIPFDPMSMRWE proteins present in gastrointestinal pathogens and parasites such as Gammaproteobacteria, *Trypanosoma cruzi*, *Fusobacterium varium* and *Leptospira sp.* were matched with <60 E values, however overall higher than those of the VH peptide (Supplementary Table 4.5). Of note, the F-box/kelch-repeat protein of tomato (*Solanum lycopersicum*) and the alcohol dehydrogenase of the castor oil plant (*Ricinus communis*) were also within the most relevant hits.

Peptide GLHTSATNLYLH (GL) was commonly detected in all three mAbs. It was the third most recurrent peptide sequenced throughout the six phage display experiments and its BLAST output contained mostly proteins of alimentary-related products, such as sugarcane (bZIP transcription factor 29), millet (basic leucine zipper 9-like), corn, and apple (Supplementary Table 4.6). The proteins related to human pathogens had E values higher than 200 and included *Mycobacterium*, *Firmicutes*, *Acetinobacter* and *Clostridium spp.*. Strikingly, peptide GL was also detected in a MCL sample from an independent laboratory (personal communication). When aligning the MCL IG sequence to those of the 3 BLs between 45-48% identity percentage in their heavy chains and a 68-83% similarity through the light chains was seen. Belonging to a different IGHV family (V5), converging residues between the MCL and the 3 BL mAb sequences were mostly located in the FR rather than their CDRs (data not shown).

To further prove the binding of the mAbs to the VH and GL peptides ELISAs were performed. Each antibody was tested with both peptides and the QS control, except for antibody 621, which didn't produce VH sequences in the phage display experiments (Supplementary Figure 4.15, C-D). Binding was weak but an increasing signal for specific binding could be detected with higher antibody molarity in all peptide-mAb tested combinations. Of note, peptide GL produced complete binding curve when combined to antibody 589 (Figure 4.6, E).

4.3.4 Structural 3D modelling of specific peptides and superantigen SpA to BL antibodies

Binding or docking models were generated with the online tool CABSdock. Each docking assessment yielded a total of 10 possible models, which were evaluated for ABS or FR

binding and cluster density. Upon three-dimensional modelling of the 3 clones mAb there was a trend of some antibody to bind tested peptides more frequently either in ABS or in FR. Cases 079 and 621 mostly bound peptides in their ABS, while mAb 589 contacted them most frequently in its FRs. Comparing 13 recurrent bindings (079-VH, 079-GL, 079-SG, 589-VH, 589-DW, 589-FI, 589-GL, 589-AD, 621-DW, 621-FI, 621-GN, 621-GL and 621-SG) no preference was seen to either binding ABS or FR peptide-wise.

A localisation analysis of the structurally conserved motifs for staphylococcal protein A (SpA) present in our 3 BL mAbs was performed (Supplementary Figure 4.16). They were localised mainly in the heavy chain FR3 of all 3 mAbs.

4.3.5 NMR assessment of peptide-antibody binding

Nuclear magnetic resonance is able to characterize peptide sequences with the use of several resonance spectra. In our case we characterized each peptide tested for binding and thus we were able to specify which amino acid was interacting with the mAb in the test. Supplementary Table 4.7 summarizes all interactions attributed to the peptide amino acids when bound to each mAb. Moreover, by resonance properties observed in the characterisation (2D COSY and NOESY) experiments of peptide GL a tendency to form a β-turn between the $T_7N_8L_9$ ($Amino\_acid_{position\_in\_peptide}$) portion was seen. The same molecule showed spatial interactions between residues $A_6T_7N_8$ and between $Y_{10}L_{11}$. Peptide GL, thus, seems to have a light structural tendency, possibly belonging to a conformational, as opposed to a linear, epitope.

Analysis of the peptide-mAb binding with nuclear magnetic resonance could confirm some results seen in the ELISA experiments. Saturation transfer difference plots confirmed binding of the GL peptide with all three studied BL mAbs (Figure 4.7). Peptide GN with mAbs 079 and 589, as expected, didn't show binding; while binding to mAb 621 could not be confirmed. Binding of peptide FI to mAbs 589 and 621 resulted with low signals in the peptide aromatic residues, but consistent binding throughout the peptide length could not be confirmed with NMR in neither (Supplementary Figure 4.17, B, C). Peptide VH was seen to be unstable in H_2O solution by showing degenerated proton spectra after ON storage. Accordingly, fresh solutions were prepared before each analysis. VH binding to mAb 079, albeit weak, was confirmed; but to mAb 589 it was not (see Supplementary Figure 4.17, A). However, the unstable characteristic of this peptide in its linear form could be a sign for a nonlinear or conformational epitope containing part or all VH residues.

Peptide QS, formed by the same scrambled residues of peptide VH was tested with mAb 589 (Supplementary Figure 4.17, D). Some saturation was observed in the aromatic

group region, possibly due to the irradiation transmission or some residues interacting to the mAb unspecifically.

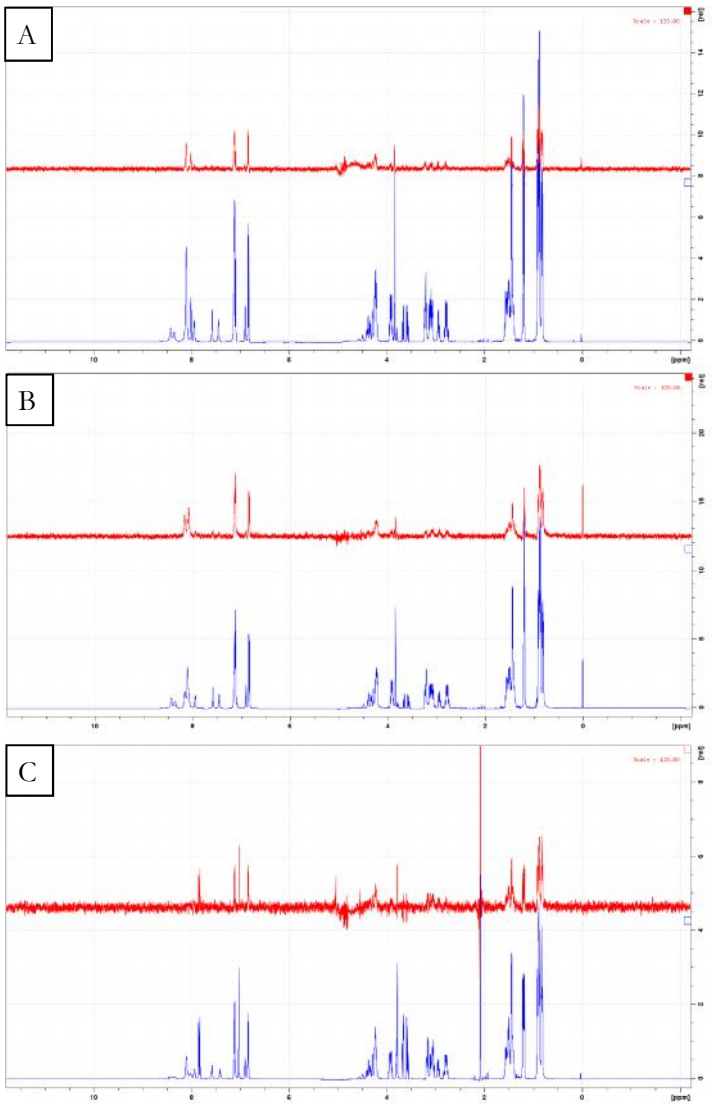


Figure 4.7. In red, STD spectra of peptide GL with **A**: 079, **B**: 589 and **C**: 621 mAbs after irradiation at 5952 or 7110Hz are shown. For each STD graph a proton spectrum (blue) is represented underneath, showing the amide, aromatic groups of the side chains (between 6-9ppm) and aliphatic groups (containing Cα and methyl groups) of the analysed peptide (between 0-5ppm).

4.3.6 BcR CDR2 stereotyped mutation is relevant in recurrent peptide binding

The three studied BL antibodies contained a stereotyped CDR2 mutation in codon 64. It is unclear, but suspected, whether unique and recurring mutations in lymphoma signal a role of antigens in its pathogenesis. To assess if the S>N mutation in CDR2 was determinant in antigen-binding peptide sequences obtained through phage display experiments were docked *in silico* using CABSdock. For each antibody a mutated (containing asparagine) and germline (with serine) version in codon 64 were docked to several phage display peptides.

For each peptide-mAb studied interaction (079-VH, 079-GL, 079-SG, 589-VH, 589-DW, 589-FI, 589-GL, 589-AD, 621-DW, 621-FI, 621-GN, 621-GL and 621-SG) the number of detected ABS and FR interactions was measured and statistical differences were calculated between the mutated and the germline populations. Only in the ABS binding comparison the difference was statistically significant between the mutated and the germline groups ($p=0.017$), showing that there was indeed an effect in reverting the stereotyped mutation to its germline sequence original (Figure 4.8).

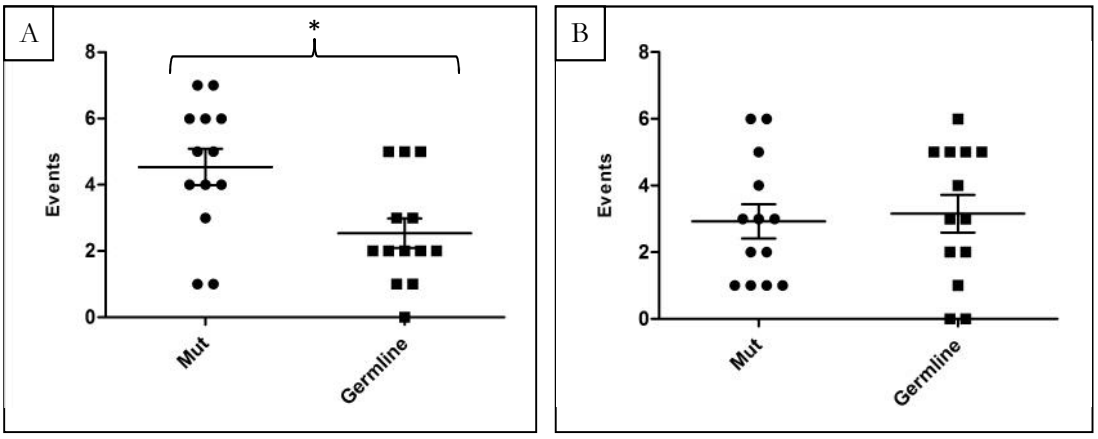


Figure 4.8. Number of times the sequences of the three BL mAbs have bound relevant phage display peptides in their ABS (**A**) or outside (**B**) of it.

The mutation reversion effect translated in different *in silico* binding of the mAbs to the peptide sequences selected by phage display. Comparison between the amount of peptide binding to the mAb (number of amino acids contacting the antibody or number of contacts between peptide and antibody) between the CDR2 mutated sequences and the models containing the germline codon 64 found no statistical differences ($p=0.368$ and $p=0.731$ respectively, not shown). This shows that while there was no evidence that the binding of the peptide to the mAb was weakened or strengthened by the CDR2 mutation the areas where the peptides made contact were preferably the ABS for the mutated in comparison to the ones carrying the germline amino acid.

4.4 Discussion

In our IHC analysis an inter-sample cross- and auto-reactivity was detected by at least one of the produced mAbs from the most productive transfection technique. The monoclonal antibody (mAb) from sample 589 presented auto-reactivity on the original biopsy slides when stained with the transfection supernatants and also stained the original patient cryo material of case 621. The same mAb had specificity for neuronal tissue of the intestine and the brain. There was a more broad interaction of all 3 BL mAbs, which reacted against cells transfected with a vector containing their own IG sequence and against the other two

cases. This cross-reactivity seen in acetone-fixed transfected HEK cells was not only restricted to cells with full-insert plasmids, but was also seen in cells that had been transfected with a vector containing only the mouse constant region. However, HEK-293 cells that were mock transfected (with no plasmid) were not stained with any mAb supernatant. Most probably, thus, the mAbs were either reacting against themselves and/or against the mouse IG constant region present in the culture supernatant, suggesting broad antibody affinity or lack of antibody maturation.

The peptides encountered in phage display experiments showed two types of possible antigen origin: pathogenic or alimentary. While peptide GNNPLHVHHDKR was only homologous to proteins of known human pathogens, peptides VHWDFRQWWQPS and FIPFDPMSMRWE hit both pathogen and alimentary-related proteins. The third most recurrent peptide found in all 6 phage display experiments with the BL mAbs was GLHTSATNLYLH. Not only was this peptide also present in independent MCL phage display experiments, but its matching proteins were mostly alimentary. As epidemiological studies have demonstrated there is a negative correlation between the risk of suffering BL and having allergy [7]. The fact that one transversal possible epitope (namely peptide GL) had sequence homologies with nutritional elements could signal an undeveloped food allergy or intolerance.

Moreover, conserved residues between the BL and MCL mAbs recognising GL are located in the FR region. Since FRs are less affected by somatic hypermutation events [20] this was expected, but it could also be a sign of FR binding, and therefore unspecific or inmature mAb recognition. Another sign of possible unspecific FR binding is the localisation of *Staphylococcus aureus* protein A (SpA) binding motifs in the 3 BL mAbs. SpA has been related to IGHV3 and V4 family binding, stimulating the survival of those B-cells expressing them [9,21]. The special property of these IG rearrangements would consist in superantigen binding through subgroup-specific residues outside the usual ABS [22].

A previously was the case with biased IG gene repertoire, stereotyped mutation studies are also lead by chronic lymphocytic leukemia (CLL) study groups [22,23] but not much is known for BL. Our 3 studied BL cases were chosen due to containing the same stereotyped S>N mutation in codon 64 of the CDR2 region. Having detected putative antigenic peptides for all 3 mAbs through phage display, docking experiments were performed to assess if the CDR2 mutation affected peptide binding. Mutation reversion from the patient's original arginine to the germline serine showed significant differences in the location of peptide binding *in silico*, being the mutated antibody version more efficient in ABS binding than the germline version. Since tethering of an antigen to the antibody ABS is considered to be a sign of specific binding, the validity of the peptide sequences discovered by phage display is reinforced.

Furthermore, the stereotyped mutation present in the 3 BL mAbs could be a sign of an antigenic exposure prior to lymphomagenesis. However, not a sign towards a recurrent motif throughout different BL samples and cell lines was found (see phage display epitope clustering and MEME results). There doesn't seem to be a clear linear or structural pattern by aligning the sequences found by us and other groups working with Burkitt lymphoma materials. In the immunohistochemical stainings we were able to show cross-reactivity between the three studied BL samples but also cross-tissue affinity of at least one mAb sample. This heterogeneous reactivity can be explained by the existence of a nonlinear epitope (or several mimotopes) or even an unspecific autoreactive mAb nature. The entropic or sticky binding site would make defining single epitopes difficult. That being said, an initial triggering event for the three studied BL cases doesn't seem to have been related to autoreactivity *per se*, since no human proteins resulted from the BLAST screening of the phage display sequences. Even in such a small sample experiment it can already be pointed out that BL mAbs can recognise peptides with homology to different origin proteins, which would advocate for a unspecific or inmature nature of BL BcRs. The multiple antigenic pseudo-affinity phenomenon is seen in also CLL [8], where many antigens have been postulated to be related to the lymphoma pathogenesis. From autogenous potential antigens such as non-muscle myosin heavy chain IIA present in apoptotic bodies [24], or the i and I blood carbohydrates [25], to exogenous molecules like fungal polysaccharides [26] and viral infections such as Hepatitis C virus [27], CLL has a wide array of antigenic candidates. Thus, like CLL, BL could also have in antigens a key component in its pathogenesis.

With several protein-origins as antigenic candidates for the analysed BL mAbs there probably was not a single organism driving or carrying the hit-and-run event prior to their lymphoma onset.

What remains to be fully understood is the missing link between antigenic contact and B-cell malignancy. Several studies have focused to elucidate the molecular causes behind B-cell lymphoma [28,29] and Burkitt lymphoma in particular [30,31]. A very interesting development was that of Muppidi et al. regarding mechanisms by which a signalling disruption of the membrane receptor Gα13 pathway promotes loss of confinement in germinal center of B cells, leading to dissemination within the LN, lymph and blood. This is due to loss of function mutations of the receptor itself (GNA13, encoded by Gα13), of its coupling receptors P2RY8 and S1PR2, and of its effector ARHGEF1 [32]. Furthermore, Peyer's patches involvement in early BL has been seen in a new and promising mouse model for the disease [33]. We could not show reactivity to heterologous Peyer's patches with our mAbs but a screening for such human tissues with antigenic markers would be an interesting test.

As investigations on BcR signalling and nature develop [34] we gain insight in how pathogenic events may unfold. Yet, a full picture that shows the main hypothesis from antigenic hit-and-run to B-cell activation, linked to secondary mutagenic events leading to lymphomagenesis is still unpainted. What seems clear until now is that the characteristics and signalling function of the BcR in lymphoma are pivotal to both understand ontogeny and improve treatment options.

Acknowledgments

The authors would like to thank Olivera Batic, Charlotte Botz-von-Drathen, Gitta Kohlmeyer, Tanja Engel, and Dana Germer for excellent technical support. Advice from Dr. Matthias Peipp, Dr. Andreas Recke, Dr. Steffen Möller, Michael Fitchner and Sven Schuchardt is much appreciated.

Web tool references:

Translate tool: http://web.expasy.org/translate/

INFO software: http://www.northeastern.edu/xray/downloads/info-program/

MimoDB: http://immunet.cn/mimodb/

Clustal Omega: http://www.ebi.ac.uk/Tools/msa/clustalo/

MEME Suite: http://meme-suite.org/

CABS-dock: http://biocomp.chem.uw.edu.pl/CABSdock

Rosetta Antibody Protocol: http://rosie.rosettacommons.org/antibody

References

[1] K.A. Blum, G. Lozanski, J.C. Byrd. Adult Burkitt leukemia and lymphoma. Blood 104 (2004) 3009–3020.

[2] Nader Kim El-Mallawany, and Mitchell S. Cairo. Advances in the Diagnosis and Treatment of Childhood and Adolescent B-Cell Non-Hodgkin Lymphoma. Clinical Advances in Hematology & Oncology 13 (2015) 113–123.

[3] W. Klapper, M. Szczepanowski, B. Burkhardt, H. Berger, M. Rosolowski, S. Bentink, C. Schwaenen, S. Wessendorf, R. Spang, P. Moller, M.L. Hansmann, H.-W. Bernd, G. Ott, M. Hummel, H. Stein, M. Loeffler, L. Trumper, M. Zimmermann, A. Reiter, R. Siebert. Molecular profiling of pediatric mature B-cell lymphoma treated in population-based prospective clinical trials. Blood 112 (2008) 1374–1381.

[4] Jackson Orem, Edward Katongole Mbidde, Bo Lambert, Silvia de Sanjose, and Elisabete Weiderpass. Burkitt's lymphoma in Africa, a review of the epidemiology and etiology. Afr Health Sci. 3 (2007) 166–175.

[5] A. Gloghini, R. Dolcetti, A. Carbone. Lymphomas occurring specifically in HIV-infected patients: from pathogenesis to pathology. Seminars in Cancer Biology 23 (2013) 457–467.

[6] Mizuno F., Koizumi S., Osato T., Kokwaro J.O., Ito Y. Chinese and african Euphorbiaceae plant extracts: Markedly enhancing effect on Epstein-Barr virus-induced transformation. Cancer Letters (1983) 199–205.

[7] S.M. Mbulaiteye, L.M. Morton, J.N. Sampson, E.T. Chang, L. Costas, S. de Sanjosé, T. Lightfoot, J. Kelly, J.W. Friedberg, W. Cozen, R. Marcos-Gragera, S.L. Slager, B.M. Birmann, D.D. Weisenburger. Medical history, lifestyle, family history, and occupational risk factors for sporadic Burkitt lymphoma/leukemia: the Interlymph Non-Hodgkin Lymphoma Subtypes Project. Journal of the National Cancer Institute. Monographs 2014 (2014) 106–114.

[8] L.-A. Sutton, A. Agathangelidis, C. Belessi, N. Darzentas, F. Davi, P. Ghia, R. Rosenquist, K. Stamatopoulos. Antigen selection in B-cell lymphomas--tracing the evidence. Seminars in Cancer Biology 23 (2013) 399–409.

[9] M.J. Baptista, E. Calpe, E. Fernandez, L. Colomo, T.M. Cardesa-Salzmann, P. Abrisqueta, F. Bosch, M. Crespo. Analysis of the IGHV region in Burkitt's lymphomas supports a germinal center origin and a role for superantigens in lymphomagenesis. Leukemia Research 38 (2014) 509–515.

[10] N. Darzentas, K. Stamatopoulos. Stereotyped B cell receptors in B cell leukemias and lymphomas. Methods in molecular biology (Clifton, N.J.) 971 (2013) 135–148.

[11] J.-K. Oh, E. Weiderpass. Infection and cancer: global distribution and burden of diseases. Annals of global health 80 (2014) 384–392.
[12] S. Mandava, L. Makowski, S. Devarapalli, J. Uzubell, D.J. Rodi. RELIC-a bioinformatics server for combinatorial peptide analysis and identification of protein-ligand interaction sites. Proteomics 4 (2004) 1439–1460.
[13] J. Huang, B. Ru, P. Zhu, F. Nie, J. Yang, X. Wang, P. Dai, H. Lin, F.-B. Guo, N. Rao. MimoDB 2.0: a mimotope database and beyond. Nucleic acids research 40 (2012) D271-7.
[14] A. Menendez, J.K. Scott. The nature of target-unrelated peptides recovered in the screening of phage-displayed random peptide libraries with antibodies. Analytical biochemistry 336 (2005) 145–157.
[15] M. Vodnik, U. Zager, B. Strukelj, M. Lunder. Phage display: selecting straws instead of a needle from a haystack. Molecules (Basel, Switzerland) 16 (2011) 790–817.
[16] S. Lyskov, F.-C. Chou, S.Ó. Conchúir, B.S. Der, K. Drew, D. Kuroda, J. Xu, B.D. Weitzner, P.D. Renfrew, P. Sripakdeevong, B. Borgo, J.J. Havranek, B. Kuhlman, T. Kortemme, R. Bonneau, J.J. Gray, R. Das. Serverification of molecular modeling applications: the Rosetta Online Server that Includes Everyone (ROSIE). PLoS ONE 8 (2013) e63906.
[17] M. Kurcinski, M. Jamroz, M. Blaszczyk, A. Kolinski, S. Kmiecik. CABS-dock web server for the flexible docking of peptides to proteins without prior knowledge of the binding site. Nucleic acids research 43 (2015) W419-24.
[18] M. Pellecchia. Solution nuclear magnetic resonance spectroscopy techniques for probing intermolecular interactions. Chemistry & biology 12 (2005) 961–971.
[19] C. Wehr, F. Müller, J. Schüler, T. Tomann, C. Nitschke, H. Seismann, E. Spillner, K. Klingner, T. Schneider-Merck, M. Binder, H.-H. Fiebig, R. Mertelsmann, M. Trepel. Anti-tumor activity of a B-cell receptor-targeted peptide in a novel disseminated lymphoma xenograft model. International journal of cancer. Journal international du cancer 131 (2012) E10-20.
[20] Hans-Peter Brezinschek, Sandra J. Foster, Ruth I. Brezinschek, Thomas Dörner, Rana Domiati-Saad, and Peter E. Lipsky. Analysis of the Human VH Gene Repertoire: Differential Effects of Selection and Somatic Hypermutation on Human Peripheral CD5+/IgM+ and CD5-/IgM+ B cells. J. Clin. Invest. 99 (1997) 2488–2501.
[21] M. Montesinos-Rongen, F.G. Purschke, A. Brunn, C. May, E. Nordhoff, K. Marcus, M. Deckert. Primary Central Nervous System (CNS) Lymphoma B Cell Receptors Recognize CNS Proteins. Journal of immunology (Baltimore, Md. 1950) 195 (2015) 1312–1319.
[22] K. Stamatopoulos, C. Belessi, C. Moreno, M. Boudjograh, G. Guida, T. Smilevska, L. Belhoul, S. Stella, N. Stavroyianni, M. Crespo, A. Hadzidimitriou, L. Sutton, F. Bosch, N. Laoutaris, A. Anagnostopoulos, E. Montserrat, A. Fassas, G. Dighiero, F. Caligaris-Cappio, H. Merle-Béral, P. Ghia, F. Davi. Over 20% of patients with chronic lymphocytic leukemia carry stereotyped receptors: Pathogenetic implications and clinical correlations. Blood 109 (2007) 259–270.
[23] A. Agathangelidis, A. Vardi, P. Baliakas, K. Stamatopoulos. Stereotyped B-cell receptors in chronic lymphocytic leukemia. Leukemia & lymphoma 55 (2014) 2252–2261.
[24] C.C. Chu, R. Catera, L. Zhang, S. Didier, B.M. Agagnina, R.N. Damle, M.S. Kaufman, J.E. Kolitz, S.L. Allen, K.R. Rai, N. Chiorazzi. Many chronic lymphocytic leukemia antibodies recognize apoptotic cells with exposed nonmuscle myosin heavy chain IIA: implications for patient outcome and cell of origin. Blood 115 (2010) 3907–3915.
[25] N.M. Bhat, M.M. Bieber, C.J. Chapman, F.K. Stevenson, N.N. Teng. Human antilipid A monoclonal antibodies bind to human B cells and the i antigen on cord red blood cells. J Immunol 151 (1993) 5011–5021.
[26] R. Hoogeboom, van Kessel, Kok P M, F. Hochstenbach, T.A. Wormhoudt, R.J.A. Reinten, K. Wagner, A.P. Kater, J.E.J. Guikema, R.J. Bende, van Noesel, Carel J M. A mutated B cell chronic lymphocytic leukemia subset that recognizes and responds to fungi. The Journal of experimental medicine 210 (2013) 59–70.

[27] E. Kostareli, M. Gounari, A. Janus, F. Murray, X. Brochet, V. Giudicelli, S. Pospisilova, D. Oscier, L. Foroni, P.F. Di Celle, B. Tichy, L.B. Pedersen, J. Jurlander, M. Ponzoni, A. Kouvatsi, A. Anagnostopoulos, K. Thompson, N. Darzentas, M.-P. Lefranc, C. Belessi, R. Rosenquist, F. Davi, P. Ghia, K. Stamatopoulos. Antigen receptor stereotypy across B-cell lymphoproliferations: the case of IGHV4-59/IGKV3-20 receptors with rheumatoid factor activity. Leukemia 26 (2012) 1127–1131.

[28] R. Küppers. Mechanisms of B-cell lymphoma pathogenesis. Nat Rev Cancer 5 (2005) 251–262.

[29] J.I. Martin-Subero, M. Kreuz, M. Bibikova, S. Bentink, O. Ammerpohl, E. Wickham-Garcia, M. Rosolowski, J. Richter, L. Lopez-Serra, E. Ballestar, H. Berger, X. Agirre, H.-W. Bernd, V. Calvanese, S.B. Cogliatti, H.G. Drexler, J.-B. Fan, M.F. Fraga, M.L. Hansmann, M. Hummel, W. Klapper, B. Korn, R. Kuppers, MacLeod, R. A. F., P. Moller, G. Ott, C. Pott, F. Prosper, A. Rosenwald, C. Schwaenen, D. Schubeler, M. Seifert, B. Sturzenhofecker, M. Weber, S. Wessendorf, M. Loeffler, L. Trumper, H. Stein, R. Spang, M. Esteller, D. Barker, D. Hasenclever, R. Siebert. New insights into the biology and origin of mature aggressive B-cell lymphomas by combined epigenomic, genomic, and transcriptional profiling. Blood 113 (2009) 2488–2497.

[30] R. Schmitz, R.M. Young, M. Ceribelli, S. Jhavar, W. Xiao, M. Zhang, G. Wright, A.L. Shaffer, D.J. Hodson, E. Buras, X. Liu, J. Powell, Y. Yang, W. Xu, H. Zhao, H. Kohlhammer, A. Rosenwald, P. Kluin, H.K. Müller-Hermelink, G. Ott, R.D. Gascoyne, J.M. Connors, L.M. Rimsza, E. Campo, E.S. Jaffe, J. Delabie, E.B. Smeland, M.D. Ogwang, S.J. Reynolds, R.I. Fisher, R.M. Braziel, R.R. Tubbs, J.R. Cook, D.D. Weisenburger, W.C. Chan, S. Pittaluga, W. Wilson, T.A. Waldmann, M. Rowe, S.M. Mbulaiteye, A.B. Rickinson, L.M. Staudt. Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics. Nature 490 (2012) 116–120.

[31] A. Greenough, S.S. Dave. New clues to the molecular pathogenesis of Burkitt lymphoma revealed through next-generation sequencing: NO FREE ACCESS. Curr. Opin. Hematol. (2014).

[32] J.R. Muppidi, R. Schmitz, J.A. Green, W. Xiao, A.B. Larsen, S.E. Braun, J. An, Y. Xu, A. Rosenwald, G. Ott, R.D. Gascoyne, L.M. Rimsza, E. Campo, E.S. Jaffe, J. Delabie, E.B. Smeland, R.M. Braziel, R.R. Tubbs, J.R. Cook, D.D. Weisenburger, W.C. Chan, N. Vaidehi, L.M. Staudt, J.G. Cyster. Loss of signalling via Gα13 in germinal centre B-cell-derived lymphoma. Nature 516 (2014) 254–258.

[33] S. Sander, D.P. Calado, L. Srinivasan, K. Köchert, B. Zhang, M. Rosolowski, S.J. Rodig, K. Holzmann, S. Stilgenbauer, R. Siebert, L. Bullinger, K. Rajewsky. Synergy between PI3K Signaling and MYC in Burkitt Lymphomagenesis. Cancer Cell 22 (2012) 167–179.

[34] V. Seda, M. Mraz. B-cell receptor signalling and its crosstalk with other pathways in normal and malignant cells. European journal of haematology 94 (2015) 193–205.

4.5 Supplementary data to **Putative antigenic epitopes of CDR2 stereotyped-mutated Burkitt lymphoma antibodies**

4.5.1 Supplementary Material and Methods

- Vector cloning for plasmid expansion

Vector pCR2.1 from the Topo TA cloning kit is used (Thermo Fisher Scientific) to amplify the immunoglobulin sequences needed, a procedure that already ads the poly-A tails needed for the cloning. When the plasmids are transformed to TOP 10 *E.coli* competent cells (Thermo Fisher Scientific) following the company's manual they are spread to LB agarKan plates to grow overnight (ON) at 37°C.

After the ON incubation the plates are placed at 4°C for further clone amplification. Three milliliters of LBKan medium is used to generate minipreps that will contain 1 picked colony. Per case and chain 7 minipreps are generated and incubated ON at 37°C. The minipreps are centrifuged for 10minutes at 5000g and the supernatant is discarded. The resulting pellets are frozen at -20°C. Bacterial pellets are purified using peqGOLD Plasmid Miniprep Kit II (PeqLab) following the company's guidelines. Plasmid DNA concentration is measured with NanoDrop.

To control for the correct insert introduction to the vector an enzymatic digestion is prepared as appears in the Buffer and Reactions List. The reaction is maintained at 37°C for 2h and then an electrophoresis gel at 1% agarose is run for 20 minutes at 120V with the restriction digestion material. Tubes containing clones with a reliable band pattern are chosen and sequencing reactions are prepared with the M13 forward and reverse primers (TOPO® TA Cloning Kit, Thermo Fisher Scientific). The sequencing reaction products are cleaned and sequenced as before. After an extensive sequence analysis the original sequences previously sequenced by another group are compared to the recent ones. The sequences are blasted to the IgBlast (http://www.ncbi.nlm.nih.gov/igblast/) and IMGT/V-QUEST (http://www.imgt.org/IMGT_vquest/share/textes/) databases to be checked for productivity and compared to their germline sequences. The samples that show no sequence changes are chosen as reference material for the future.

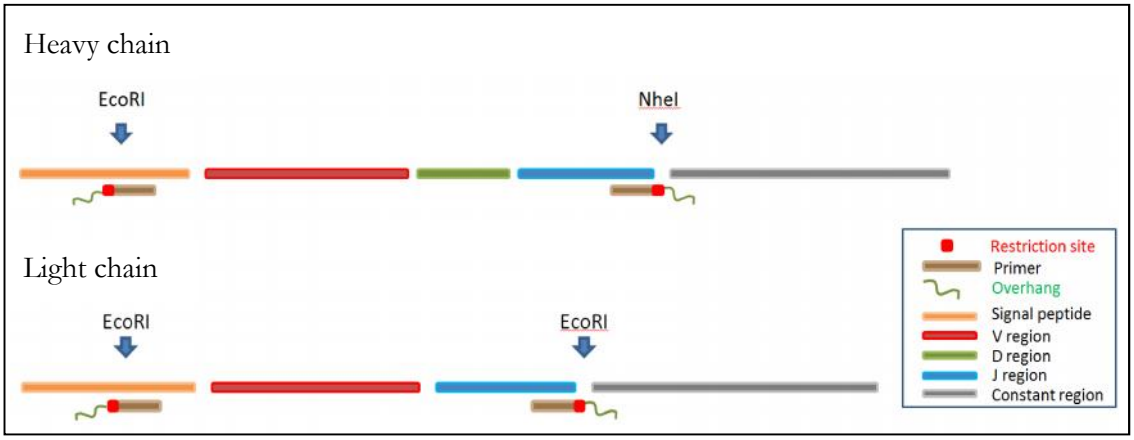
The pCR2.1 TOPO TA cloning products are amplified via PCR. A gel electrophoresis at 1% agarose is done for size control. The right bands are cut and DNA is isolated from them with NucleoSpin® Gel and PCR Clean-up kit (Macherey & Nagel). DNA concentrations are measured with NanoDrop.

- Vector cloning for transfection preparation

The vectors to be used for mammalian cell transfection, pFUSE2-CLIg-mI1 and pFUSE-CHIg-mG2a, are amplified in One Shot TOP10 Chemically Competent *E.coli* (Thermo Fisher Scientific), extracted with PureLink® Quick Plasmid Miniprep Kit (Thermo Fisher

Scientific), and their DNA concentration measured with NanoDrop and fotometer. In order to control the clones a restriction reaction using EcoRI for 2h at 37°C and a subsequent 1% agarose gel electrophoresis is done.

To make insertion to pFUSE vector easier extra restriction sites and overhangs (Supplementary Figure 4.9) are cloned inside the insert sequences. Overhangs are 6nt long and are generated randomly. Some nucleotides are added to complete codons at the 3' end of some sequences, to avoid frame shifts in the mouse constant region of the vector. Primers used are listed in Table P4 of the Primer Annex.



Supplementary Figure 4.9. pFUSE vector modification strategy to obtain new restriction sites in the sequences of interest.

The modified inserts are controlled with an electrophoresis gel and the extracted bands are purified with the NucleoSpin Extract II kit (Macherey & Nagel). Then, a digestion with the appropriate restriction enzymes is performed at 37°C ON (heavy chains with EcoRI and NheI and light chains with EcoRI) to prepare the inserts to be attached to their vector. After digestion the inserts are purified with the QIA Quick PCR Purification Kit (QIAgen).

On the vector side, the receiving pFUSE vectors are digested as their inserts. The digested vectors are dephosphorylated at 37°C for 30 minutes. The dephosphorylated vector is cleaned with a phenol-chloroform protocol: the dephosphorylated samples are brought to 200µl volume with distilled water and a further 200µl of phenol:chloroform:isoamyl alcohol (25:24:1) is added. The mixture is vortexed and centrifuged for 10 minutes at 14000 rpm. The resulting upper phase is brought to a fresh tube, to which 600µl of 100% EtOH is added. After mixing, 20µl of 3M sodium acetate (pH5.2) is added to the mix and all is vortexed before centrifuging it again for 15 minutes at 14000 rpm. The supernatant is discarded, leaving the pellet to be washed with 500µl of 70% EtOH. Finally, the pellet is left to dry and 30µl of distilled water is added to the tubes to resuspend the resulting DNA.

Then, samples are controlled via gel electrophoresis and their DNA concentration is measured in Qubit (Thermo Fisher Scientific). The quantity of each vector to be used in each ligation reaction is calculated with the following formula:

$$\text{Insert quantity (ng)} = \frac{\text{Vector quantity (ng)} \cdot \text{Insert size (Kb)}}{\text{Vector size (Kb)} \cdot \text{Molar ratio (insert: vector)}}$$

The molar ratio used is 1:3 (insert:vector) and the vector quantity is 40ng. The ligation reaction is performed at 16°C for 2 hours with T4 ligase and its buffer (New England Biolabs). And right after the ligated products are transformed to OneShot Chemically Competent Cells (Thermo Fisher Scientific) according to the companys guidelines.

The transformed cells are plated to either Zeocine (heavy chain pFUSE) or Blasticidin (light chain pFUSE) LB plates. In further minipreps and maxiprep specifications the right antibiotic is used in each respective case. Per case a total of 5 clones are selected and brought to minipreps as previously described. After the ON incubation plasmids are extracted with peqGOLD Plasmid Miniprep Kit II (PeqLab) and plasmid concentration is measured in Qubit (Thermo Fisher Scientific). Once again, an enzyme restriction with EcoRI and NheI as described above and gel electrophoresis is done for controlling purposes.

Right clones are chosen by sequencing and the use of consensus sequencing primers for both heavy and light chains. The primers bind up to the vector's original sequence, so that all insert is readable (see primers in Table P5 of the Primer Annex). The sequencing procedure is carried out as previously. After assessing the sequences the mini preps that are equal to the original samples are chosen.

To amplify the plasmids an OneShot TOP10 (Thermo Fisher Scientific) reaction is done for each heavy and light chain. Backup glycerol stocks (25% final glycerol concentration) are stored per sample. They are amplified with minipreps ON and in maxi preps afterwards.

Since the multiple cloning site (MCS) of the pFUSE vectors is quite large and it is located between the variable and the constant region (see vector maps in Annex) it is mutated away to avoid antibody conformation problems. In case of the light chain vector it is both mutated and digested due to its length. Primers used are listed in Table P6 of the Primer Annex. The mutations are carried out by following the QuikChangeII protocol (Agilent Technologies). After, the DpnI digestion (1 hour at 37°C) ensures that the template material, which is methylated, is degraded leaving only the amplification product. Then TOP10 cells are transformed following the provider's manual. Minipreps, glycerol stocks and plasmid isolation follow as described before. DNA concentration is measured with Qubit® dsDNA BR Assay Kit (Thermo Fisher Scientific), and sequencing reactions

are prepared. Control sequencing reactions are done as described above and maxipreps are stored per case.

- Transfection experiments

HEK-293 cells (ACC305, DSMZ) are cultured in complete medium (RPMI, 10% FBC, S/P) at 37°C and 5% CO2 before transfection. In order to decide with transient transfection is most suitable a comparative set-up is made where a GFP-containing plasmid (pmaxGFP®, Lonza) is transfected to 6-well plates containing HEK-293 cells. Three extra wells are taken as control, with empty plasmid (pFUSE). Three methods are compared: lipofection with TransIT (Mirus) and Lipofectamine 2000 (Thermo Fisher Scientific), and electroporation with Amaxa's Nucleofector (Lonza) following the manufacturer's protocols.

For the comparative assessment the supernatants of the three transfections are collected 24, 31, 47, 54, 72 and 78h after transfection. A fraction of the collected cells is used to assess its viability and the rest to generate cytospins. Cytospins are done by washing the cells off of the well surface with 1ml of 1xPBS. According to the cell density between 250µl and 500µl is applied to each funnel in a Cytospin3 machine (Shandon) to generate at least 2 cytospins per culture well. Loading of the cytospin machine is done according to the company's guidelines. The cells are centrifuged at 1000rpm for 5 minutes. The resulting cytospin slides are left to dry ON and stored at -80°C before DAPI staining. With an Axioplan 2 fluorescent microscope (Zeiss) pictures of the cytospin slides are made and analysed using the TissueQuest v3.0 (TissueGnostics) image analysis program.

Further transfection tests are designed to assess the best molarity ratio between the heavy chain, light chain and the pAdvantage plasmids. Molarity calculation for vector quantities has to be taken into account upon doing transfections. Three ratios are tested (H:L:pAdvantage): 1:4:1, 2:3:1, 1:4, 1:4Ø. The transfections of cases 079A, 079B, 589 and 621 are done with complete RPMI medium [10% Ultra Low IgG FBS (Thermo Fisher Scientific), without antibiotics]. Six 6-well plates are used to test different plasmid ratios and collection times in triplicates. After collecting the supernatants the cells of two wells from each triplicate are resuspended with 1ml of culture medium and further processed into cytoblocks. The remaining well of the triplicates is resuspended in 1ml of culture medium, centrifuged at 1000rpm for 2min, its supernatant is discarded and the remaining pellet is then placed in liquid N_2 and later at -80°C for storage for protein extraction.

Transfection is scaled up to reach significantly higher protein quantities for all cases.

- Protein extraction and purification

Protein extraction is performed to the available cell pellets (cases 079A, 079B and 621): for each well of collected cells 200µl of lysis buffer (see Buffer and Reactions List) and 30 µl of the cOmplete protease inhibitor (Roche) is mixed until the cell pellet is resuspended. The

mixtures are incubated on ice for 30 minutes and centrifuged at 14000rpm and 4°C for 10 minutes. The resulting supernatant is pipetted in a clean microtube and freeze-shocked in liquid N_2. Finally the protein extracts can be stored at -20 or -80°C.

IgG purification with protein (MAb Trap Kit, GE Healthcare) is carried out for all cases' supernatants following the suppier's guidelines. On upscaled transfections the MAb Trap 1ml columns are substituted with the 5ml HiTrap columns (GE Healthcare). Buffers used are listed in the Buffer and Reactions List.

After independently purifying the secreted Ig from the transient transfection experiments the fractions that result with higher absorvance at 260nm (but lower than 0.01 at 320nm) on a fotometer are brought together. Dialysis in 5 litres of 1x PBS throughout 24h with a minimum of 2 buffer changes is followed. Afterwards the purified IgGs are concentrated with Amicon® Ultra-4 Centrifugal Filter Units 30k (Merck-Millipore).

All supernatants and protein extracts are measured with the Qubit Protein Assay Kit (Thermo Fisher Scientific) for protein concentration measurement.

- Western blots

Reducing Western blots are made with NuPAGE® Novex® 4-12% Bis-Tris Protein Gels (Thermo Fisher Scientific) with protein extracts and supernatants from the transfected HEK-293 cells. On the gels the MultiMark® Colored Standard (Thermo Fisher Scientific) is used as ladder, with NuPAGE MES as running buffer (Thermo Fisher Scientific). Gels are pre-run for 20 minutes at 40V after rinsing the wells with running buffer without samples. Samples are complemented with the 1x loading dye and denatured for 5 minutes at 95°C before being loaded to the gel. IgG_{2a} mouse isotype control (BioLegend) is used as a positive control on a 1:10 dilution. Membrane transfer is carried out in cold conditions for 1h or ON at either 100 or 20V respectively on Amersham Hybond-P PVDF membranes (GE Healthcare). Simply Blue SafeStain (Thermo Fisher Scientific) is used to ensure product transferred, following the supplier's guidelines. Membrane blocking is done in blocking buffer (see Buffer and Reactions List) during 1h at RT with light shaking. Blocking is followed by a wash in 1xTBST during 5 minutes and 2h incubation with 1:5000 in blocking buffer of the primary antibody (except at 1:2500 for the anti-bovine antibody). Next, three washes of 15 minutes each in 1xTBST and the secondary antibody incubation (at 1:10000) for 45 minutes in blocking buffer is done. Two more washes in 1xTBST and one in 1xTBS at 15 minutes each follow. Membranes are then processed with Amersham ECL Prime Western Blotting Detection Reagent (GE Healthcare) according to the company's guidelines. Stripping of the PVDF membrane is done with 100ml of the stripping buffer and 700µl of β-mercaptoethanol (see Buffer and Reactions List) at 50°C for 30 minutes with light shaking. The membrane is subsequently washed three times for 10 minutes with 1xTBST.

Native Western blots are performed for the 1:4:1 and 72h ratio and proteinG-purified supernatants except for case 079A where no protein is being detected in the supernatants, for such case the protein extracts are used. The protocol used is the one supplied by NativePAGE™ Novex® 4-16% Bis-Tris Protein Gels (Thermo Fisher Scientific). Materials used are listed in the Materials table. Specifically, Light Blue Cathode Buffer (containing a 0.002% of G-250 Coomassie) is used in the runs and is prepared as described in the company's protocol. The native gels are run at 150V for 120 minutes, at RT. After the gel run the Coomassie staining protocol in the manual is performed, followed by the Western blotting per se. The transference is done with the transfer buffer used for the denaturing Westerns, without methanol at 75V for 1h, starting at 130mA and finishing at 100mA. After blotting the membrane is stained with the Ponceau S solution for control. The membrane is blocked and stained as detailed in the denaturing version.

- Cyto blocks

The cells are detached from the plate after 1xPBS wash and centrifuged for 10 minutes at 2000rpm. The resulting supernatant is discarded while the pellet is resuspended with 2ml of 4% formaldehyde. This fixation step is left 15 minutes at RT and it is centrifuged for 1 minute at 2000rpm. Then, the supernatant is discarded and 3 drops of protein glycerol and 2ml of 96% isopropanol are added to the pellet. The mixture is centrifuged for 5 minutes at 2000rpm. The pellet is scraped with a spatula and placed in a filter paper in a tissue capsule. The fixed material is then processed as standard biopsy material and embedded in a paraffin block.

- Staining protocols

Staining protocol for FFPE immunology

The formalin-fxed paraffin-embedded (FFPE) slides are left to dry after cutting at 37°C ON. They are then incubated in xylol for 20 minutes and afterwards apply shortly the slides in 3 different cuvettes of xylol sequentially. The slides are then placed through an alcohol line (100% three times, 96% two times and 70% once). The endogenous peroxidase is blocked by incubating the slides in for 10 minutes in a methanol H2O2 solution. The next step will unmask the antigen through 3 minutes pressure-cooking of the slides in citrate buffer (Merck), pH 6.0.

The slides are then washed in regular water and in WB. After washing between 150µl of the 1:10 transfection supernatant is applied to each slide, and left for 1h at RT covered to avoid drying out. The antibody CD20 (Dako) is used in the control slides at 1:600. The slides are washed in regular water three times and once in WB.

The secondary antibody (ab97245 or ab74316, Abcam) is applied at a 1:200 dilution in PBS, 150µl to each slide and left to incubate for 30 minutes in a wet

chamber to hinder drying out. The slides are washed as previously. Next, a total of 100µl of Dako-DAB is applied to each slide and left to react for 5-7 minutes or until the brown colour develops. Then the slides are washed with regular water. The materials are stained with Mayer's hemalaun solution (Merck) at 1:4 dilution, washed with regular water and left for 10 minutes in a clean water cuvette. The slides are passed through an alcohol line (70%, 1x; 96%, 2x; and 100%, 3x) and 2 times with a 100% solution of alcohol-xylol (1:1). Finally, submerge the slides 4 times in xylol and cover with Pertex (Medite).

All materials and providers are listed in the Material List.

Staining for fresh-frozen material, cytospins and cell coverslips

The fresh material and cytospins are fixed for 10minutes in acetone and left to dry. They are washed with wash buffer (WB).

The target (10x) is prepared at 1:50 dilution, ON at RT or 1h at 95°C. Slides are pre-processed with the target (Dako) either 7 minutes at 95°C or 3 minutes at RT. After incubation, the slides are washed with WB.

After this step the protocol is valid also for the coverslips containing cultured cells, although the washing is for this material more delicate, by use of a pipette and always washed with WB.

The slides are incubated for 1h with the transfection supernatant (1:10) and washed as previously. The primary antibody enhancer is applied in 3-4 drops per slide and left to incubate 20 minutes. Washed with WB, three times with regular water (not for coverslips) and one further time with WB, the slides receive each 3-4 drops of AP polymer to each slide and are washed as before. The colour developing is performed by applying 100µl of the appropriate reagent of the Lab Vision™ UltraVision™ kit to the slides, and they are left to incubate 10 minutes. They are washed once with regular water (except for coverslips, in WB) and stained with Mayer's hemalaun solution (Merck) in 1:4 dilution for 3-6 minutes. After staining, the slides are washed with regular water for 10 minutes, this time for all types of slides, but carefully and with a pipette for the coverslips.

At this stage the coverslips are ready to be left to dry before covering them with Aquatex (Merck) and protected with a further coverslip on top. The rest of the materials will be further processed in an alcohol line (30%, 70%, 95%, 95% and 100%), once submerged in xylol and covered in Pertex (Medite).

All materials and providers are listed in the Material List.

DAPI staining

Leave cryo slides or cytospins to dry and place 10 minutes in acetone to fix the tissue. Let dry and submerge in PBS. In each slide apply 100µl of DAPI, 1:5000, during 2 minutes. Wash 3 times in PBS and cover with mounting medium, a coverslip and seal with transparent nail polish.

All materials and providers are listed in the Material List.

- Phage display

Phage display is established with PhD12 library (NewEngland Biolabs). Each run is done separately to avoid cross-contamination from different antibodies and the reagents are all prepared afresh per experiment. The purified IgGs are biotinylated using and following the manual of EZ-Link™ Micro Sulfo-NHS-LC-Biotinylation Kit (Thermo Fisher Scientific). Once biotinylated they are desalted through the Zeba Spin Desalting Columns within the EZ-Link kit. Protein concentration is measured in Qubit (Thermo Fisher Scientific).

The K12 strain of *Escherichia coli* (ER2738) is used for phage amplification and maintenance. Selection is done on IPTG/Xgal agar plates and bacteria are grown in LB^{Tet}. A panning round is followed by an amplification round with their respective titrations. Each titration is used to control for the appropriate number of phages present in the eluates. Two rounds of panning and amplification are done before the 3rd panning, where its titration will be used to independently amplify selected clones.

Specifically a panning round consists of preparing one 20ml LB culture of ER2738 from a plate in an Erlenmeyer and a 50ml-Falcon with 5ml of LB^{Tet} of ER2738 from a plate (not older than a week). The *E.coli* cultures are placed in the incubator, at 37°C, shaking and at 200rpm until the Erlenmeyer reaches an OD of 0,01-0,05 and the Falcon tube 0,5. A 96-well panning plate (Nunc Immobilizer Streptavidin F96 Clear, Thermo Fisher Scientific) is sealed and 4 wells are opened. The 4 wells are activated with 350ml of 1xDPBS each, they are washed with a pipette and fresh 1xDPBS is applied and left for 15 minutes. The panning mix contains 0,1µg of the biotinylated protein, 1µg of IgG_{2a} mouse isotype control, 20µl of the phage library (to reach $2x10^{11}$ phages) and the necessary volume of 1xDPBS to reach 400µl. The mixture is kept on ice. Under a fume hood the DPBS is discarded and 100µl/well of the panning mix is applied to each of the 4 wells and left to incubate at RT for 10 minutes. One microliter of biotin is applied per well and incubated a further 5 minutes at RT. The liquid is discarded and the plate is slapped down to cellulose tissues to dry the wells. The used wells are washed five times with 1xTBST and the plate is slapped face-down to some clean tissues each time. The bound phages are eluted by pipetting 100µl of triethylamine (100mM) per well and left 10 minutes at RT with light shaking. After the incubation the liquid is placed in a microtube that will quickly receive a

further 200µl of 1M TrisHCl (pH9) to neutralize the acidic eluate. The tube with the panning eluate is vortexed and placed on ice.

When the Erlenmeyer culture reaches an OD of 0,01-0,05 the phage eluate can be inoculated for amplification. A small quantity of the panning eluate is kept for titration. Infected bacterial culture is left in the incubator at 37°C for 4.5h shaking at 200rpm so that the phages can be reproduced. After the amplification time is over the 20ml culture is poured into a 50ml-Falcon. The culture is centrifuged during 15 minutes at 4000g at 4°C and the supernatant is transferred to a clean tube and re-spinned, while the pellet is discarded. The resulting upper 80% of the supernatant is transferred into a new tube, where 1/6 volume of 20% PEG/2.5M NaCl is added. This solution will precipitate the phages at 4°C ON.

When the Falcon tube reaches an OD of 0.5 the output titration can be done in the range of 10-10^4. The titration protocol from the phage library manual is followed.

The next day the titration results are evaluated by counting the blue plaques in each culture plate to calculate the plaque forming units per microlitre (pfu/µl). The amount of transparent plaques (if any) has also to be taken into account, since that indicates wild type phages and thus contamination. Only plates with 100 or more plaques can be assessed. The ON PEG mixture is centrifuged at 4000g for 30 minutes at 4°C. The supernatant is discarded and the tube is re-spinned shortly to remove the remaining supernatant with a pipette. The phage pellet is resuspended in 1ml of TBS and placed in a microtube. The tube is centrifuged at 14000rpm for 5 minutes at 4°C to pellet residual cells. The phage-containing supernatant is then pipetted to a clean microtube and 1/6 of its volume of 20% PEG/2 .5 M NaCl is added to it to precipitate the phages. The tube is incubated on ice for 15-60 minutes and then centrifuged at 14000rpm for 10 minutes at 4°C. The supernatant is discarded, the tube re-spinned briefly and the remaining supernatant is removed by pipetting. Two-hundred microliters of TBS are used to re-suspend the phage pellet and finally the mixture is centrifuged for another 1 minute. The supernatant is placed in a microcentrifuge tube and can be stored a maximum of 3 weeks at 4°C. This purified amplification is titrated taking into account the panning titration. The supplier guidelines are taken into account at this point. Normally 5 titration dilutions are made and left to incubate at 37°C ON.

The second panning and amplification are done as the first round but with more stringent washing of the panning step: 10 washes with TBST (0.05% tween).

The third panning round is done as the two previous but with 20 washes of TBST (0.05% tween) and without a subsequent amplification. On the titration the plates with an expected suitable plaque number will be duplicated in order to have enough plaques to pick and amplify.

ELISA plates (Nunc maxisorp) are washed with sterile PBS and then coated with the appropriate purified antibody. An antibody solution at 10µg/ml is generated in carbonate

buffer (see Buffer and Reactions List). The plate is left covered at 4°C ON. Two ELISA plates are prepared per phage display round and antibody: one with the antibody of interest and another with a mouse IgG_{2a} isotype control (ab18414, Abcam), in the same conditions.

The following day an amplification plate is prepared. A 1:100 dilution in sufficient volume is generated out of an ON ER2738 culture grown in 10ml of LB^{Tet} at 37°C. When the OD of the diluted bacterial culture reaches 0.1 some 220µl of the culture are pipetted to each ELISA plate well. No bacteria are placed in the last 3 wells of the last row. Next, a plaque is picked from the titration plates, not older than 3 days, and inoculated to each well with *E.coli*. The plate is incubated at 37°C, shaking at 200rpm for 4h. After the incubation time the plate is centrifuged at 4000rpm for 5 minutes and 50µl of the upper supernatant is transferred to the washed ELISA plate. The upper 80% of the remaining volume is stored on a clean standard plate.

After the ON incubation of the ELISA plates the liquid is blotted out of the wells and the plates are washed. A wash round consists of a submersion of the plate in washing buffer (see Buffer and Reactions List), decantation and blotting repeated four times. Then, 400µl of blocking buffer is pipetted to the plate and left at RT until the amplification plate is ready. Another wash is performed and 50µl of the supernatant from the last amplification is applied to the respective wells. In the last 3 wells a total of two positive and one negative control are established. The first positive control and instead of the bacterial culture contains 30µl of the purification of 2nd amplification round and the second positive control receives 50µl of the 3rd panning elution. The negative control has only fresh LB medium. The ELISA plate is sealed and incubated at RT for 1.5h with light shaking. A further wash is performed and the detection antibody is applied to each well with a quantity of 100µl and 1:5000 ratio in blocking buffer. The second positive control receives anti-mouse-HRP in the same conditions instead. The plate is sealed and incubated 1h at RT with light shaking. After one last wash the developing of the plate can be carried out following the 1-Step™ Ultra TMB-ELISA Substrate Solution supplier's instructions (Thermo Fisher Scientific). Readings are taken at 650nm before stopping the development and at 450nm after adding the H_2SO_4. The reference reading is the latter. Both antibody and its corresponding anti-IgG_{2a} mouse isotype control plate are processed in parallel in the same day.

After the last amplification of the selected phages and the subsequent ELISA assay the positive ELISA and a number of negative samples are selected on the amplified phages plate and a DNA extraction of the existing phages is performed. Forty µl of PEG is added to each tube containing phage in solution. Tubes are left at RT for 15 minutes and afterwards centrifuged at 14000rpm for 10 minutes at 4°C. The generated supernatants are discarded and the tubes re-spinned briefly. The resulting pellets are re-suspended in 20µl of iodide buffer (see Buffer and Reactions List) and the tubes are tapped vigorously. Next, 50µl of ethanol is added to each tube and incubated for 15 minutes at RT. The samples are

centrifuged at 14000rpm for 10 minutes at 4°C and their supernatants are discarded. A 100µl of cold ethanol (kept at -20°C) is added to each resulting pellet and then centrifuged as before once again. Supernatants are discarded and pellets are left to dry. Finally, phage DNA pellets are re-suspended in 15µl of TE buffer and measured in the NanoDrop (Thermo Fisher Scientific) for DNA concentration estimation.

Sequencing reactions are performed with the M13 96-primer from the PhD12 kit (New England Biolabs, sequence: 5'-CCCTCATAGTTAGCGTAACG-3'). Forty nanograms of phage DNA are used per sequencing reaction, which is performed as previously described but with 1µM of 96-primer. Sequence reaction purification is performed as before and read in the ABI 3500 Sanger machine.

- mAb-peptide ELISAs

Biotinylated synthetic peptides (Eurogentec) are re-suspended in ddH2O to generate 1mg/ml concentrations. NeutrAvidin-coated 96-well plates (Thermo Fisher Scientific) are washed three times with wash buffer (WB, see Buffer and Reactions List).

	Coating peptide	mAb	Detection Ab
Control +	Y	Y	Y,HRP-streptavidin
Control -1	Y	Y	N
Control -2	Y	N	N
Control -3	N	Y, 079	Y, HRP-anti-mouse
Control -4	N	Y, 589	Y, HRP-anti-mouse
Control -5	N	Y, 621	Y, HRP-anti-mouse

Supplementary Table 4.1. Positive and negative controls performed in the peptide-mAb ELISAs with their respective mAb application (primary Ab) and use or absence of the detection Ab.

After adjusting the pH to reach 6-7 the peptides are allowed to coat the plates in 1xPBS ON at 4°C. A total of 100µl of peptide at a 30µg/ml concentration is pipetted in each corresponding well. Control (VHWDFRQWWQPS randomized, named QS) and test peptides are placed in the same plate to be able to compare and subtract unspecific binding (represented by the random peptide). All seven mAb dilutions are tested in duplicates, from 0.35uM to 84pM, separated by 1:4 dilution steps. The following controls are applied: a positive control, and 5 negative controls (see Supplementary Table SX for their specifications).

After the ON incubation of the peptides the plates are washed with WB three times and they are blocked with 300µl of 1xTBS/3% milk powder at RT for 2 hours. Plates are washed three times again with WB before 150µl of each mAb dilution step is pipetted to each corresponding well in the plate. An incubation at RT for 30 minutes is followed by another wash and the detection antibodies are applied at 100µl per well. The detecting antibody HRP-anti-mouse (Jackson ImmunoResearch) is used at a 1:1000 in WB and HRP-

streptavidin is used at a 1:4000in 1xPBS. Plates are incubated in the dark with light shaking for 30 minutes and then washed two times with WB and the last time with 1xTBS. Developing of the signals is done as previously detailed.

To evaluate binding the results from the plate wells containing peptide QS are used as unspecific binding and GraphPad's protocol for Two sites Specific binding is followed to generate graphical results.

- Nuclear magnetic resonance

Two milligram of peptide were dissolved in 550µl of distilled H_2O and 50µl of D_2O (containing a small amount of DSS for calibration purposes). Once the solution was fully resolved pH was adjusted to 5-6 by adding NaOD at 0.01N (5-10µl). The NMR equipment used was a Brunker Avance 600 MHz spectrometer, with triple channel cryoprobe head. Firstly, peptide proton nuclear magnetic resonances were assigned by double quantum filtered COSY (correlation spectroscopy), 2D-NOESY (nuclear Overhauser effect spectroscopy) and 2D-TOCSY (total correlation spectroscopy) experiments. Mixing times of TOCSY were 60ms and 300ms for NOESY. Measurements were carried out at 280°K (6.85°C) and 300°K (RT). Excitation sculpting with gradients was used for the water suppression in all experiments and results were visualised with the TopSpin 3.2 (Bruker) software.

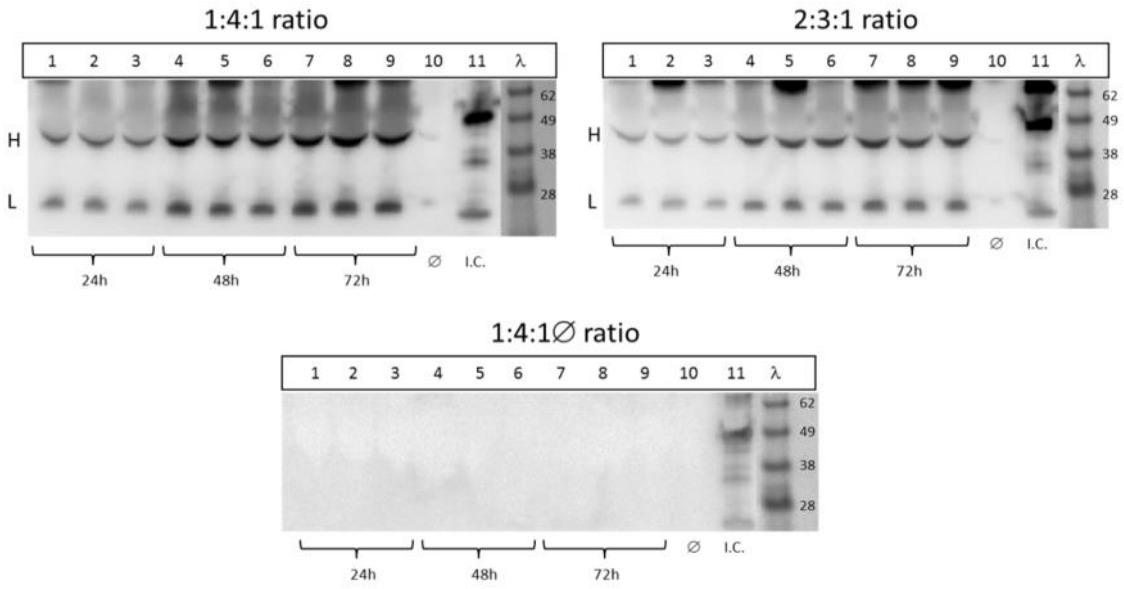
Once peptides were deemed pure approximately 500pmol of mAb was added to the 2500µM peptide solution, making an end molarity of 0.8µM. The molar ratio used in the mAb: peptide experiments was thus 1:3000. The mAb and peptide solution was irradiated at +11.85ppm (7110Hz) and -0.9ppm (-540Hz) to selectively saturate the mAb resonances. Gauss pulses for 12.5ms were used, without delay, 4s of saturation and 2s of acquisition time. All STDs were either carried out at 1200, 2500 or 4800 scans for need of extra resolution.

4.5.2 Supplementary Results

- mAb production

The pFUSE plasmids (see Chapter 3) were used for the heavy and light chain sequence transfections. Additionally an adjuvant plasmid (pAdvantage, Promega) was also used to improve protein production. Two molar ratios between the three transfection plasmids were tested, namely 1:4:1 and 2:3:1, representing "heavy chain : light chain : pAdvantage". An extra combination with empty plasmids (that is, plasmids without insert cloned in them) was used, with the molecular ratio of 1:4:1, and labelled 1:4:1∅. Three collection time points were also considered to detect the most effective production time, at 24, 48 and 72h after transfection. The selected ratio and time point for supernatant collection was 1:4:1, 72h post transfection (Supplementary Figure 4.10).

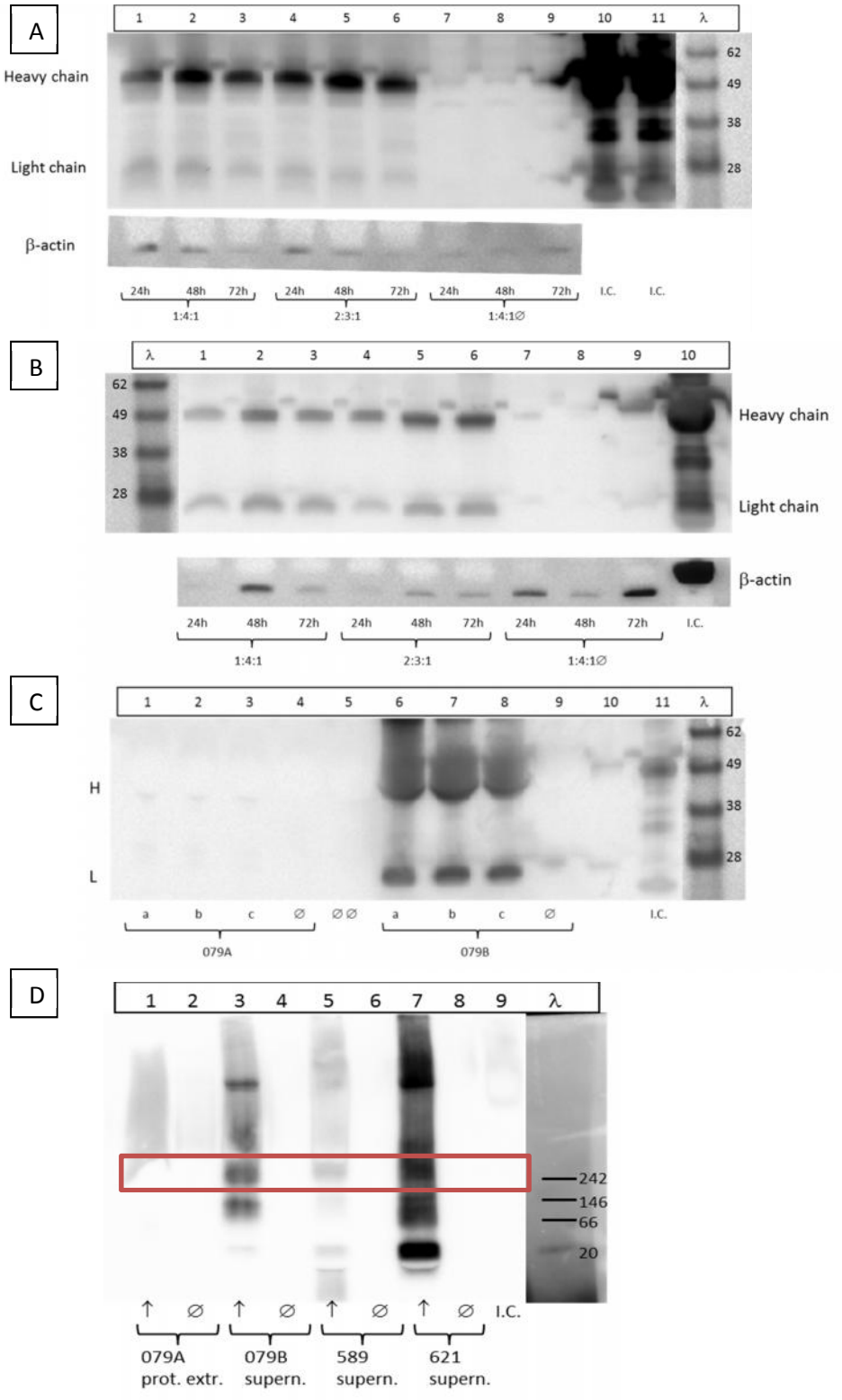
When dealing with 079A, which included variant IGLV3-12*01 and IGLJ3*02 in the light chain it was seen that while the protein extraction material contained a heavy chain product the secreted, the light chain band was very weak. In the direct comparison of the transfection supernatant products between case 079A (IGLV3-12*01 and IGLJ3*02) and case 079B (IGLV1-51*02 and IGLJ3*01) it was seen that only cells transfected with case 079B secreted a heavy and light chain peptide (Supplementary Figure 4.11, C).



Supplementary Figure 4.10. Ratio and collection time experiment for case 589 in triplicates. I.C.: isotype control, λ: size marker, ∅: empty pFUSE plasmid supernatant, H: heavy chain, L: light chain.

Further SDS-PAGE blots were done to check for minimal quantities of bovine proteins that could be present in purified fractions of the mAbs. When loading the gel with BSA the IgG heavy and light detecting antibodies do not react to it (not shown). Also, the loading of complete medium from a mock transfection (with ultra-low IgG FBS) shows that while in the Ponceau staining there is a band see in the medium-loaded well, the IgG

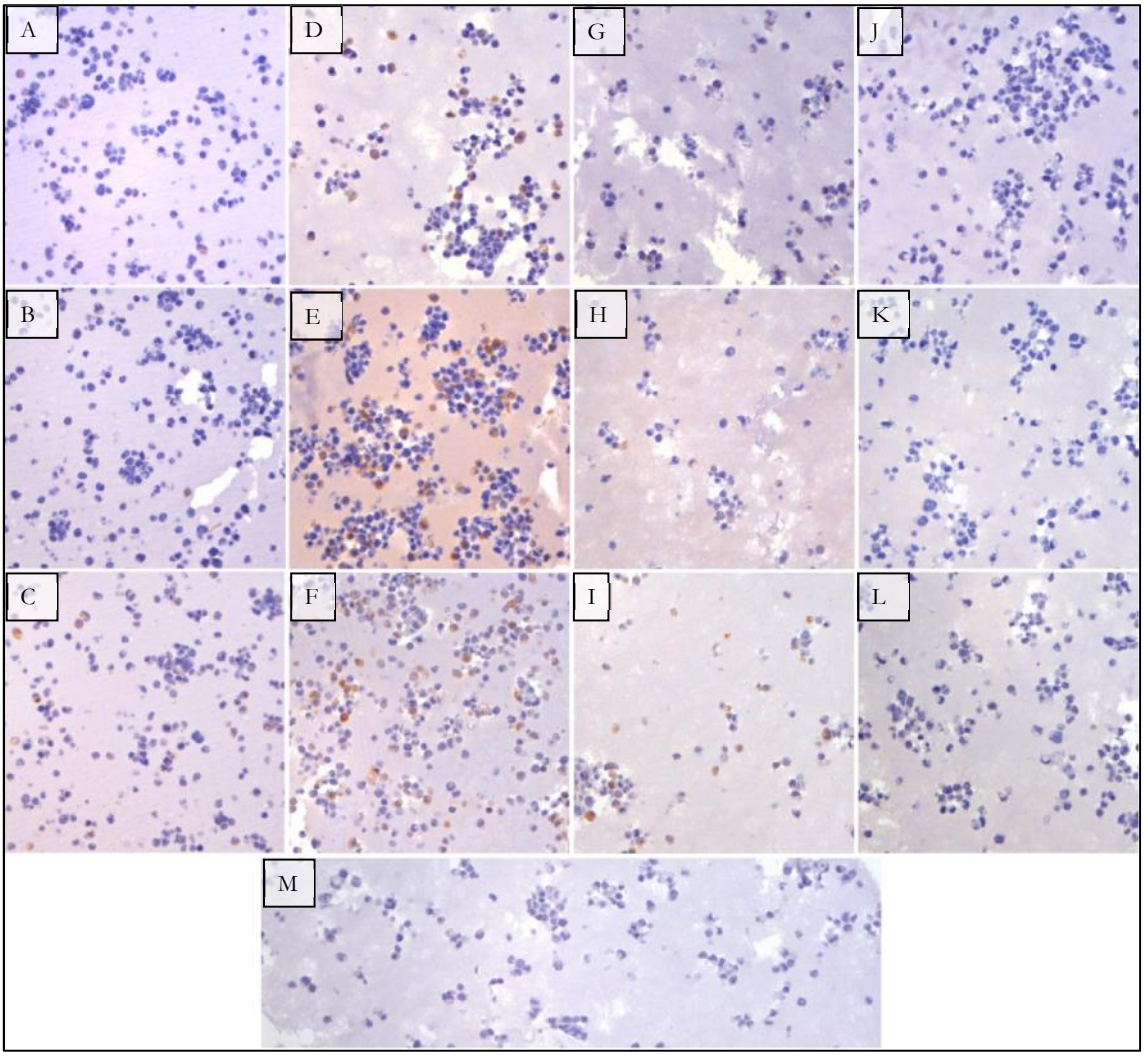
detecting antibodies do generate a signal in the same lane, so there is no risk of false-positive band detection.



Supplementary Figure 4.11. A-C, SDS-PAGE of **A**: 079A protein extraction of the three different plasmid ratios; **B**: 621 in protein extraction of the three different plasmid ratios tested: 1:4:1, 2:3:1 and 1:4:1∅ (with empty plasmids); **C**: 079A and 079B transfection

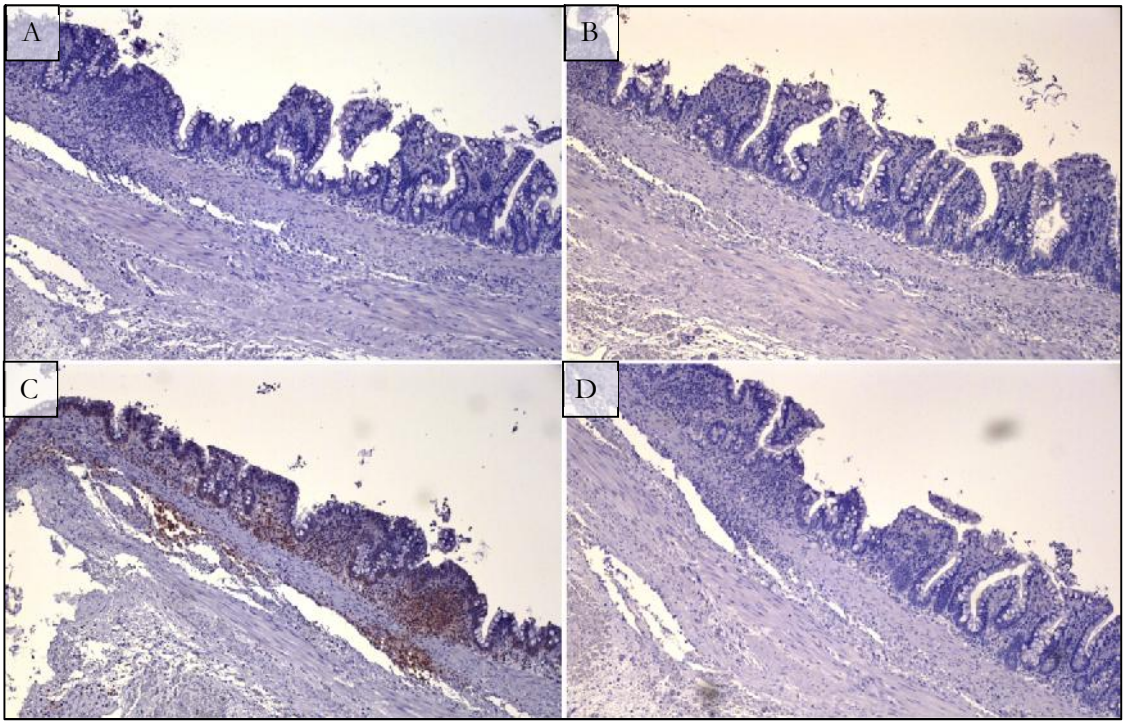
supernatants, with pAdvantage on a ratio of 1:4:1 (H:L:pAdvantage) and 72h post transfection. **D**: Native-PAGE the 079A protein extraction and 079B, 589 and 621 transfection supernatants. The red square marks where the band for the recombinant IgG shoud be. The arrows mark the lanes with the supernatant or protein extract materials. a, b, c: triplicates; I.C.: isotype control, λ: size marker, ∅: empty pFUSE plasmid supernatant, ∅∅: no plasmid supernatant, H: heavy chain, L: light chain.

- Burkitt lymphoma antibodies show cross- and self-recognition



Supplementary Figure 4.12. HEK-293 cells transfected with the 3 constructs (heavy and light and adjuvant pAdvantage plasmids), embedded in paraffin blocks. **A-C**: cells transfected with 079, **D-F**: cells transfected with 589, **G-I**: cells transfected with 621, **J-L**: untransfected cells, **M**: untransfected cells stained with their own supernatant. The first image row represents slides stained with supernatant of the 079 transfection, the second of the 589 transfection and the third row of images are cells stained with the 621 transfection supernatant. All pictures are at 20x.

Peyer patches positivity was seen in one antibody of the 3 mAbs. However, it was deemed not specific due to it staining also epithelial villi and showing a slight nuclear specificity (Supplementary Figure 4.13).

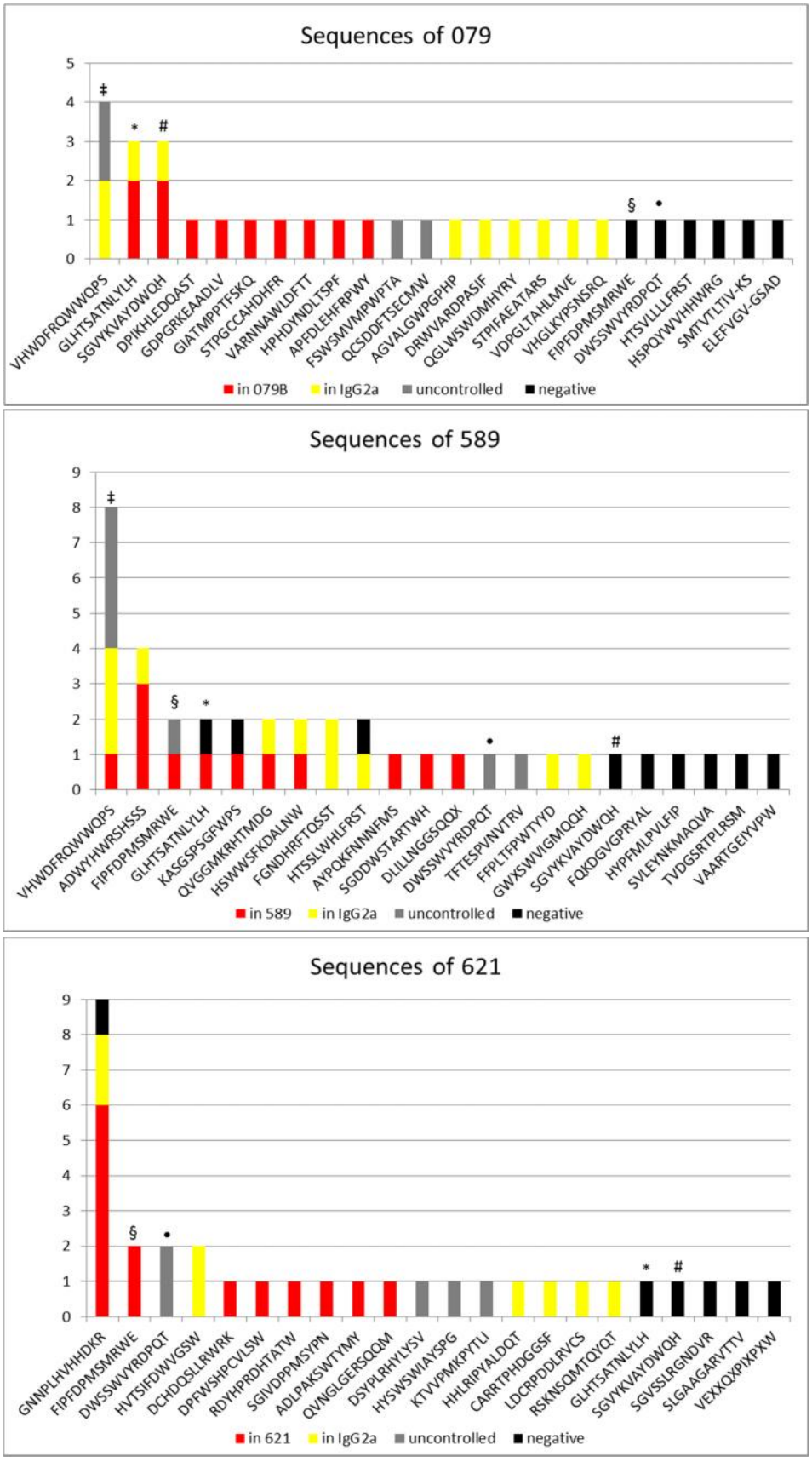


Supplementary Figure 4.13. Intestinal biopsy showing a Peyer patch area. The same biopsy is stained with antibody-containing supernatant from cells transfected with constructs **A**: 079B, **B**: 589, **C**: 621 and **D**: no construct (mock control). All pictures are at 10x.

- Recurrent interacting peptides to Burkitt lymphoma antibodies

Two sets of phage display experiments were run per mAb. In the first round of phage display the corresponding ELISA plate was only coated with the antibody of interest. Sequences obtained with this procedure are marked in grey (Suplementary Figure 4.14) and labelled as "uncontrolled". On the repeated round of phage display two ELISA plates per experiment were coated: one with the mAb of interest and another with an IgG_{2a} antibody (isotype control, ab18414, Abcam) to compare the positivity. The phages generated in the last amplification round were distributed to the two plates and ELISA was performed as described above. Well positivities were then assigned relative to the mAb (a well is only positive in the mAb plate), the IgG_{2a} mouse antibody (a well is only positive in the IgG_{2a} plate, yellow in Supplementary Figure 4.14).

Of the 104 sequences analysed 59 different peptides were detected. Some sequences came from ELISA wells that didn't develop colour above background. Within these sequences there were some recurrent peptides that had shown positivity in the other two antibodies and are marked with the same symbol across Supplementary Figure 4.14.



Supplementary Figure 4.14. All sequences read in the six phage display experiments. Sequences found in positive ELISA wells are in red (for the mAb plate) and yellow (for the

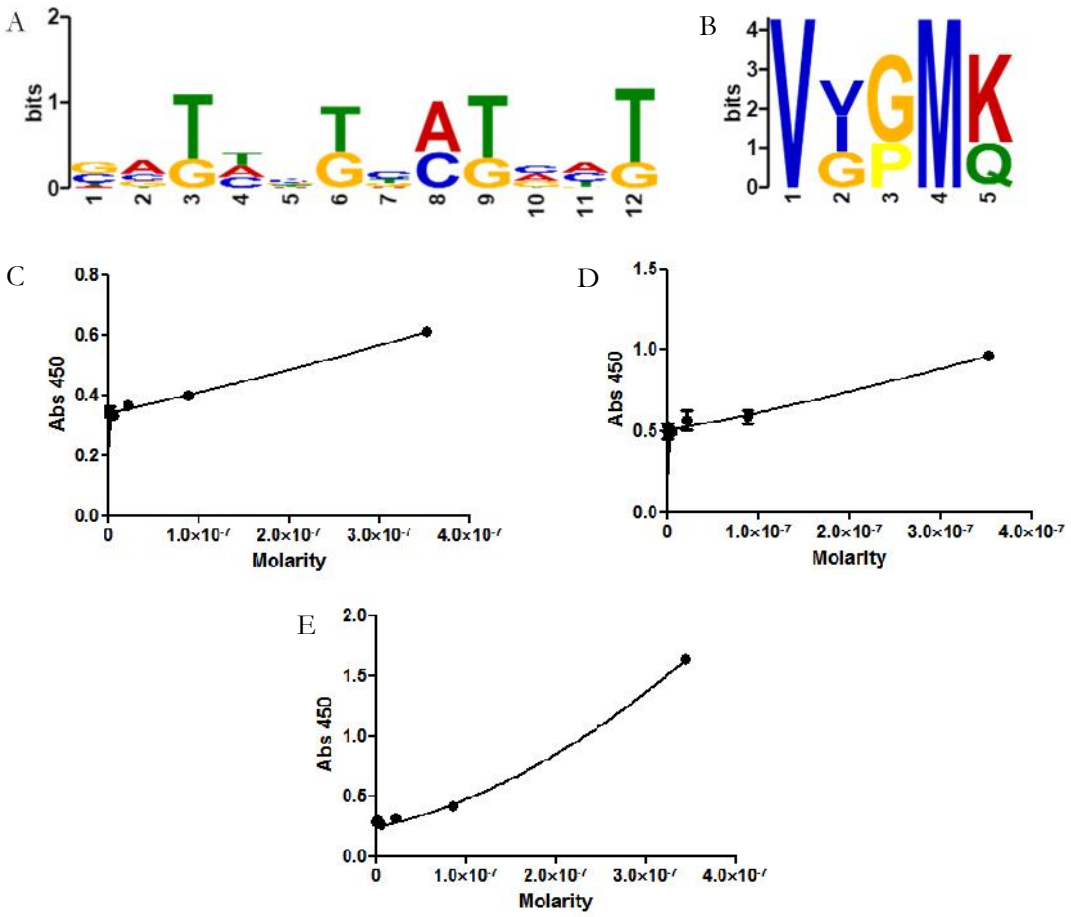
IgG_{2a} plate). Results from negative ELISA wells are in black. The sequences found in the first round of phage display, where the second parallel ELISA with the IgG_{2a} coating was not performed are represented in grey.

				In positive, of which				
	Sequences	**In positive**	**In negative**	**Ab plate**	**Unattri-butable**	**IgG plate**	**Unknown**	**Possible binding to**
1	VHWDFRQWWQPS	12		1		5	6	
2	GNNPLHVHHDKR	8	1	3	3	2		
3	GLHTSATNLYLH	4	2	3		1		
4	AD**W**YH**W**RSHSSS	4		2	1	1		plastic
5	FIPFDPMSMRWE	4	1		1		3	
6	SGVYKVAYDWQH	3	2	1	1	1		
7	D**W**SS**W**VYRDPQT	3	1				3	plastic
8	QVGGMKRHTMDG	2		1		1		
9	HVTSIFD**W**VGS**W**	2				2		plastic
10	FGNDHRFTQSST	2				2		
11	HSWWSFKDALNW	2			1	1		
12	DPIKHLEDQAST	1		1				
13	SGDDWSTARTWH	1		1				
14	SGIVDPPMSYPN	1		1				
15	VARNNAWLDFTT	1		1				
16	GDPGRKEAADLV	1		1				
17	HPHDYNDLTSPF	1		1				
18	GIATMPPTFSKQ	1		1				
19	STPGCCAHDHFR	1		1				
20	AYPQKFNNNFMS	1		1				
21	DPFWSHPCVLSW	1		1				
22	RDYHPRDHTATW	1	1	1				
23	DCHDQSLLRWRK	1		1				
24	KASGSPSGFWPS	1	1		1			
25	APFDLEHFRPWY	1			1			
26	DLILLNGGSQQX	1			1			
27	QVNGLGERSQQM	1			1			
28	ADLPAKSWTYMY	1			1			
29	VDPGLTAHLMVE	1				1		

Table continued on the next page

30	VHGLKYPSNSRQ	1				1		
31	QGL**W**S**W**DMHYRY	1				1		plastic
32	FFPLTFPWTYYD	1				1		
33	LDCRPDDLRVCS	1				1		
34	RSKNSQMTQYQT	1				1		
35	HHLRIPYALDQT	1				1		
36	CARRTPHDGGSF	1				1		
37	AGVALGWPGPHP	1				1		
38	STPIFAEATARS	1				1		
39	DRWVARDPASIF	1				1		
40	HT**SSL**WHLFRST	1	1			1		unspecific
41	G**W**XS**W**VIGMQQH	1				1		plastic
42	TFTESPVNVTRV	1					1	
43	QCSDDFTSECMW	1					1	
44	FSWSMVMPWPTA	1					1	
45	DSYPLRHYLYSV	1					1	
46	KTVVPMKPYTLI	1					1	
47	HYS**W**S**W**IAYSPG	1					1	plastic
48	VEXXQXPIXPXW		1					
49	SLGAAGARVTTV		1					plastic
50	FQKDGVGPRYAL		1					
51	HYPFMLPVLFIP		1					
52	HSPQY**W**VHH**W**RG		1					plastic
53	TVDGSRTPLRSM		1					
54	VAARTGEIYVPW		1					
55	SVLEYNKMAQVA		1					
56	HTSVLLLLFRST		1					
57	ELEFVGV-GSAD		1					
58	SMTVTLTIV-KS		1					
59	SGV**SSL**RGNDVR		1					unspecific
	TOTAL	104						

Supplementary Table 4.2. All sequences read in the six phage display experiments. Results from exclusively negative ELISA wells are in red. Residues that account for target-unspecific binding are in bold letters, and the rightmost column gives details to which type of binding was detected. The sequences found in the first round of phage display, where the second parallel ELISA with the IgG_{2a} coating was not performed are represented in the column labelled "Unknown".



Supplementary Figure 4.15. DNA (**A**) and protein (**B**) motifs found within the 59 different phage display-obtained sequences. Motif A, albeit weak, was found in 58/59 (98%) samples. The following motifs detected were only found in 2 and 3 sequences. Protein motif B was found in 3/59 (5%) samples. Peptide-mAb binding analysis was performed by ELISA for peptides VH and GL and scrambled peptide QS as the unspecific binding control **C**: Specific binding of mAb 079 to peptide VH. **D**: Specific binding of mAb 079 to peptide GL. **E**: Specific binding of mAb 621 to peptide GL.

	Description	Organism	E value	Accession	Characteristics
1	ParB-like protein	*Providencia alcalifaciens R90-1475*	0.73	EUD09004.1	Associated with diarrhea in children and travelers. It is found in the gastrointestinal tract.
2	transcriptional regulator	*Providencia alcalifaciens Dmel2*	0.74	EKT62626.1	
3	hypothetical protein PTSG_05644	*Salpingoeca rosetta*	1.4	XP_004993512.1	Choanoflagellate
4	putative ADP-ribosylation/crystallin J1	*Cronobacter phage S13*	4.7	AIA64848.1	Phage
5	methionyl tRNA synthetase, cytoplasmic	*Hymenolepis microstoma*	6.8	CDS25457.1	Intestinal dwelling parasite in the bile duct and small intestines of mice and rats, rarely infects humans
6	RND transporter	*Pseudomonas sp. LAIL14HWK12:I7*	9.2	WP_027594488.1	The best studied species include *P. aeruginosa* in its role as an opportunistic human pathogen
7	putative methionine-tRNA synthetase	*Schistosoma mansoni*	9.2	CCD78761.1	Significant parasite of humans, a trematode
8	PREDICTED: methionine--tRNA ligase, cytoplasmic	*Musca domestica*	9.4	XP_005182375.1	Housefly
9	particulate methane monooxygenase protein A	*Crenothrix polyspora*	11	ABC59824.1	Filamentous bacteria which utilize iron in their metabolism, and cause staining, plugging and taste and odor problems in water systems.
10	hypothetical protein	*Streptomyces sp. NRRL B-24484*	12	WP_030262153.1	Found predominantly in soil and decaying vegetation. They are infrequent pathogens, though infections in humans, such as mycetoma, can be caused by *S. somaliensis* and *S. sudanensis*
11	diguanylate cyclase	*Mycobacterium sp. MCS*	12	ABG06477.1	Includes pathogens known to cause serious diseases in mammals, including tuberculosis and the classic Hansen's strain of leprosy.
12	planctomycete cytochrome C domain protein	*Rhodopirellula sp. SWK7*	13	WP_009103602.1	Marine bacterium
13	beta-lactamase	*Enterobacter cloacae*	15	WP_044858668.1 and others	Member of the normal gut flora of many humans and is not usually a primary pathogen. It is sometimes associated with urinary tract and respiratory tract infections.
14	NDP-glucose dehydratase	*Streptomyces sp. 275*	16	AAR23325.1	Found predominantly in soil and decaying vegetation. They are infrequent pathogens, though infections in humans, such as mycetoma, can be caused by *S. somaliensis* and *S. sudanensis*
15	hypothetical protein	*Burkholderia glathei*	16	WP_035939563.1 KDR39482.1	Gram- soil bacterium. Pathogen for Asian Rice
16	beta-lactamase	*Enterobacter sp.*	17	WP_042717670.1 and others	Several strains of these bacteria are pathogenic and cause opportunistic infections in immunocompromised hosts
17	NDP-4-6-dehydratase	*Streptomyces sp. 275*	17	AGG12562.1	Found predominantly in soil and decaying vegetation. They are infrequent pathogens, though infections in humans, such as mycetoma, can be caused by *S. somaliensis* and *S. sudanensis*
18	beta-lactamase	*Enterobacter ludwigii Enterobacter asburiae*	17	WP_040017771.1 KJP74682.1 and others	Close to *E.cloacae* but separate strains from clinical specimens. Hospital-acquired infections
19	beta-lactamase	*Enterobacter mori*	17	WP_041951418.1	Plant pathogen
20	general substrate transporter	*Ophiostoma piceae UAMH 11346*	17	EPE06190.1	Fungus
21	OPT family small oligopeptide transporter	*Mortierella verticillata NRRL 6337*	17	KFH66298.1	Soil fungi. Mortierella wolfii (only), causes bovine abortion, pneumonia and systemic mycosis
22	putative glycosyltransferase	*Rubidibacter lacunae*	23	WP_022604159.1	Unicellular, phycoerythrin-containing cyanobacterium
23	hypothetical protein B456_005G0190002	*Gossypium raimondii*	23	KJB27967.1	Cotton plant
24	PREDICTED: protein trichome birefringence-like 3	*Cucumis melo*	23	XP_008464204.1	Melon
25	hypothetical protein	*Lactobacillus fabifermentans*	24	WP_024625239.1	Often involved in food fermentation. Lactobacillus has been reported to be the causative pathogen in many types of infection
26	hypothetical protein B456_005G018900	*Gossypium raimondii*	24	KJB27963.1	Cotton plant

Supplementary Table 4.3. Main proteins hit by peptide VH in pBLAST.

	Description	Organism	E value	Accession	Characteristics
1	hypothetical protein	*Actinoplanes sp. N902-109*	81	WP_041833668.1 AGL16941.1	Produce pharmaceutically important compounds
2	PREDICTED: contactin-6	*Eptesicus fuscus*	83	XP_008153220.1	Big brown bat
3	DNA-binding protein	*Microbacterium ketosireducens*	101	WP_045276536.1 WP_018170430.1	Found in human clinical specimens
4	glucose-6-phosphate 1-dehydrogenase	*Corynebacterium maris*	151	WP_020934819.1	Marine bacterium
5	hypothetical protein HMPREF0623_0441	*Pediococcus acidilactici DSM 20284*	171	EFL96390.1	Homofermentative bacterium
6	transcriptional regulator	*Lactobacillus farciminis*	184	WP_010020995.1	Potential for bacteriocin and antibiotic assay
7	methyltransferase	*Cedecea neteri*	189	WP_039296343.1	Isolated in clinical specimens
8	hypothetical protein	*Nitratireductor aquibiodomus*	190	WP_036540293.1	alpha-proteobacterium from the marine denitrification
9	hypothetical protein	*Alcaligenes faecalis*	194	WP_035272769.1	Commonly found in the environment. Oportunist infections and generally considered nonpathogenic
10	cupin	*Alcaligenes faecalis*	196	WP_042481935.1	
11	hypothetical protein	*Pseudomonas thermotolerans*	199	WP_040640672.1	Not known to be pathogenic
12	HIT family hydrolase	*Streptococcus oralis*	202	WP_033630192.1	Commensal bacteria (belongs to the Mitis group which contains the human pathogen S. pneumoniae), and is found in the human oral cavity. Opportunist pathogen
13	MULTISPECIES: hypothetical protein	*Streptomyces*	203	WP_030357946.1	Infections in humans can be caused
14	hypothetical protein ACD_77C00140G0015	*uncultured bacterium*	204	EKD32280.1	
15	HIT family hydrolase	*Streptococcus infantis*	204	WP_033685884.1 and others	Belongs to the Mitis group. Opportunist.
16	hypothetical proteins	*Candida albicans*	205	KGR12292.1 and others	Fungus. Opportunistic oral and genital infections in humans
17	histidine triad protein	*Streptococcus infantis SK1076*	205	EGL88114.1	Belongs to the Mitis group. Opportunist.
18	major facilitator transporter	*Nocardioides sp. URHA0032*	206	WP_028636268.1	*Vibrio cholerae* control
19	asparagine synthase	*Streptomyces sp. NRRL F-5727*	206	WP_031010442.1	Infections in humans can be caused
20	HIT family hydrolase	*Streptococcus mitis*	209	WP_004252756.1	Inhabits the human mouth
21	HIT family hydrolase	*Streptococcus peroris*	209	WP_006145489.1	Belongs to the Mitis group. Opportunist. Clinical specimens
22	hypothetical protein SK141_0919	*Streptococcus oralis*	209	KEQ46442.1	Commensal bacteria (belongs to the Mitis group which contains the human pathogen S. pneumoniae), and is found in the human oral cavity. Opportunist pathogen
23	histidine triad protein	*Streptococcus sp. SK643*	209	EIF39158.1	Infections in humans can be caused
24	hypothetical protein Desmer_2846	*Desulfosporosinus meridiei DSM 13257*	265	AFQ44752.1	Sulfate-reducing bacteria, found in soil
25	DeoR faimly transcriptional regulator	*Pseudomonas fluorescens*	275	WP_042732208.1	Pathogenesis not known
26	predicted protein	*Thalassiosira pseudonana CCMP1335*	277	XP_002291965.1	Marine centric diatom

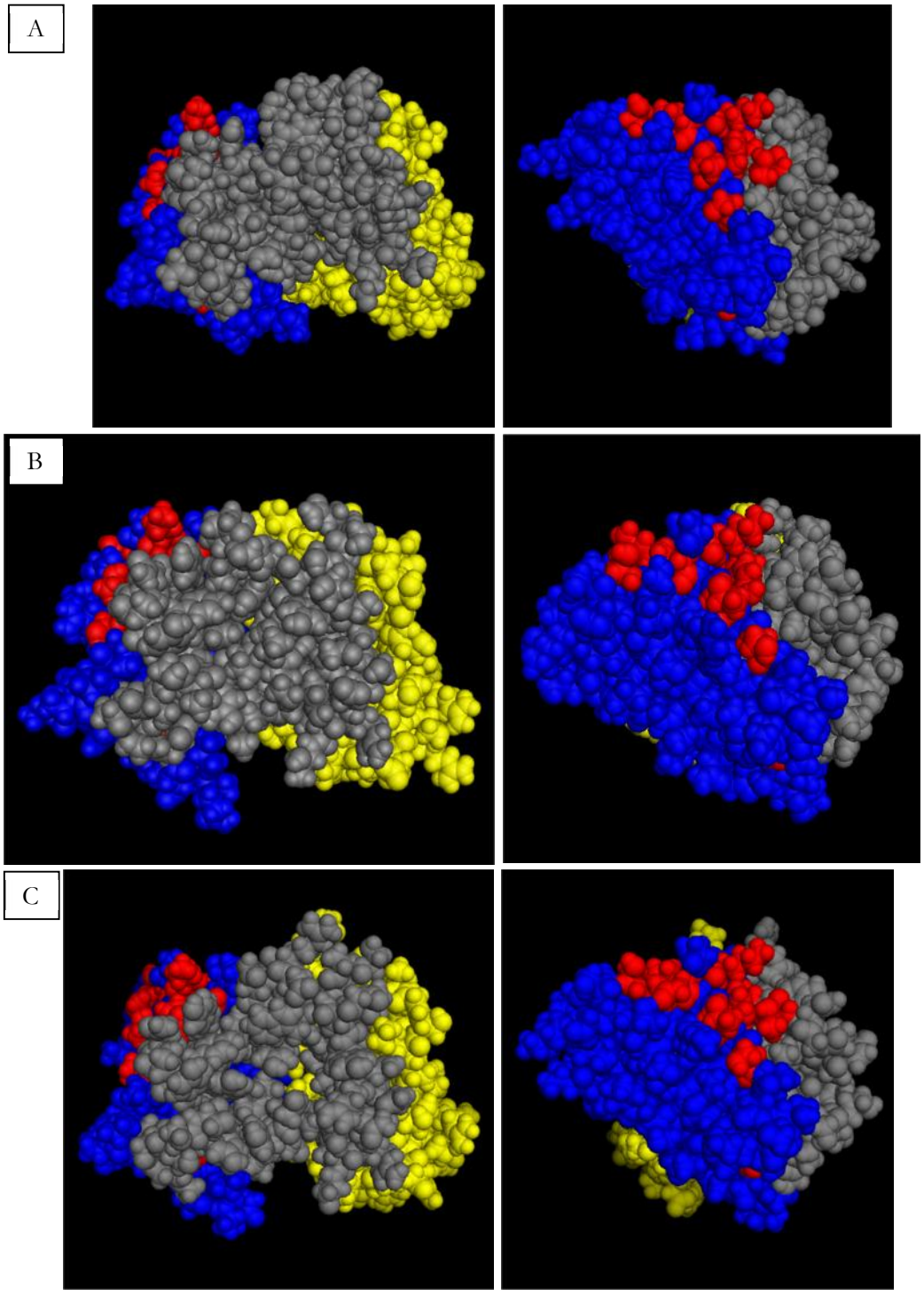
Supplementary Table 4.4. Main proteins hit by peptide GN in pBLAST.

	Description	Organism	E value	Accession	Characteristics
1	stage V sporulation protein AF	*Tuberibacillus calidus*	1.9	WP_027724421.1	Soil bacteria
2	hypothetical protein B456_007G065400	*Gossypium raimondii*	12	KJB40466.1	Cotton
3	alcohol dehydrogenase	*Burkholderia glathei*	12	WP_035934187.1	Asian rice pathogen. Other species are human pathogens
4	hypothetical protein	*Gammaproteobacteria bacterium MFB021*	17	WP_035470415.1	Pathogenic species within the group
5	hypothetical protein MGU_06668	*Metarhizium guizhouense ARSEF 977*	17	KID86235.1	Fungus
6	2'-5' RNA ligase	*Selenomonas sp. FC4001*	21	WP_037352693.1	Gastrointestinal bacteria. Pathogenicity not proven
7	chlorophyll a/b-binding protein	*Coccomyxa subellipsoidea C-169*	22	XP_005644739.1	Microalgae
8	histidine kinase	*Micromonospora sp. ATCC 39149*	23	WP_007071575.1	Bacterium used to fight fungal infections in plants
9	metal ABC transporter ATPase	*Endozoicomonas numazuensis*	24	WP_034833803.1	Bacteria isolated from marine sponges
10	hypothetical protein PIIN_01893	*Piriformospora indica DSM 11827*	24	CCA68026.1	Cultivable Plant-Growth-Promoting Root Endophyte
11	PREDICTED: F-box/kelch-repeat protein At1g22040	*Solanum lycopersicum*	32	XP_004230575.1	Tomato
12	hypothetical protein	*Rhodopirellula maiorica*	32	WP_008700650.1	Marine bacterium
13	hypothetical protein H696_00155	*Fonticula alba*	42	XP_009492265.1	Slime mold
14	Alcohol dehydrogenase, putative	*Ricinus communis*	43	XP_002537752.1	Castor oil plant
15	hypothetical protein	*Trypanosoma cruzi strain CL Brener*	45	XP_809502.1 XP_804727.1	Parasitic euglenoid protozoan. Causes Chagas' disease and sleeping sickness
16	hypothetical protein	*Pseudomonas amygdali*	46	WP_010207465.1	Almond tree pathogen
17	hypothetical protein	*Fusobacterium varium*	55	WP_005948124.1	Localized in the colonic mucosa of patients with ulcerative colitis
18	co-chaperone GrpE	*Leptospira kirschneri*	56	WP_020762800.1 and others	The species is pathogenic and can cause leptospirosis, most commonly in pigs
19	co-chaperone GrpE	*Leptospira santarosai*	56	WP_020766940.1	Pathogenic
20	co-chaperone GrpE	*Leptospira interrogans*	57	WP_017854508.1 and others	Humans are accidental hosts
21	protein GrpE	*Leptospira interrogans*	57	WP_000829531.1 WP_000852393.1	
22	co-chaperone GrpE	*Leptospira borgpetersenii*	57	WP_002748244.1	Humans are accidental hosts. Leptospirosis
23	abhydrolase domain-containing protein 4	*Clonorchis sinensis*	59	GAA32290.2	Chinese liver fluke. Parasites human livers.
24	hypothetical protein	*Lactobacillus crispatus*	60	WP_035163946.1	Beneficial micro biota species
25	hypothetical protein T265_10435	*Opisthorchis viverrini*	60	XP_009175071.1	Southeast Asian liver fluke. Causes opisthorchiasis.
26	hypothetical protein	*Fusobacterium ulcerans*	60	WP_008695352.1	Cause tropical ulcers.

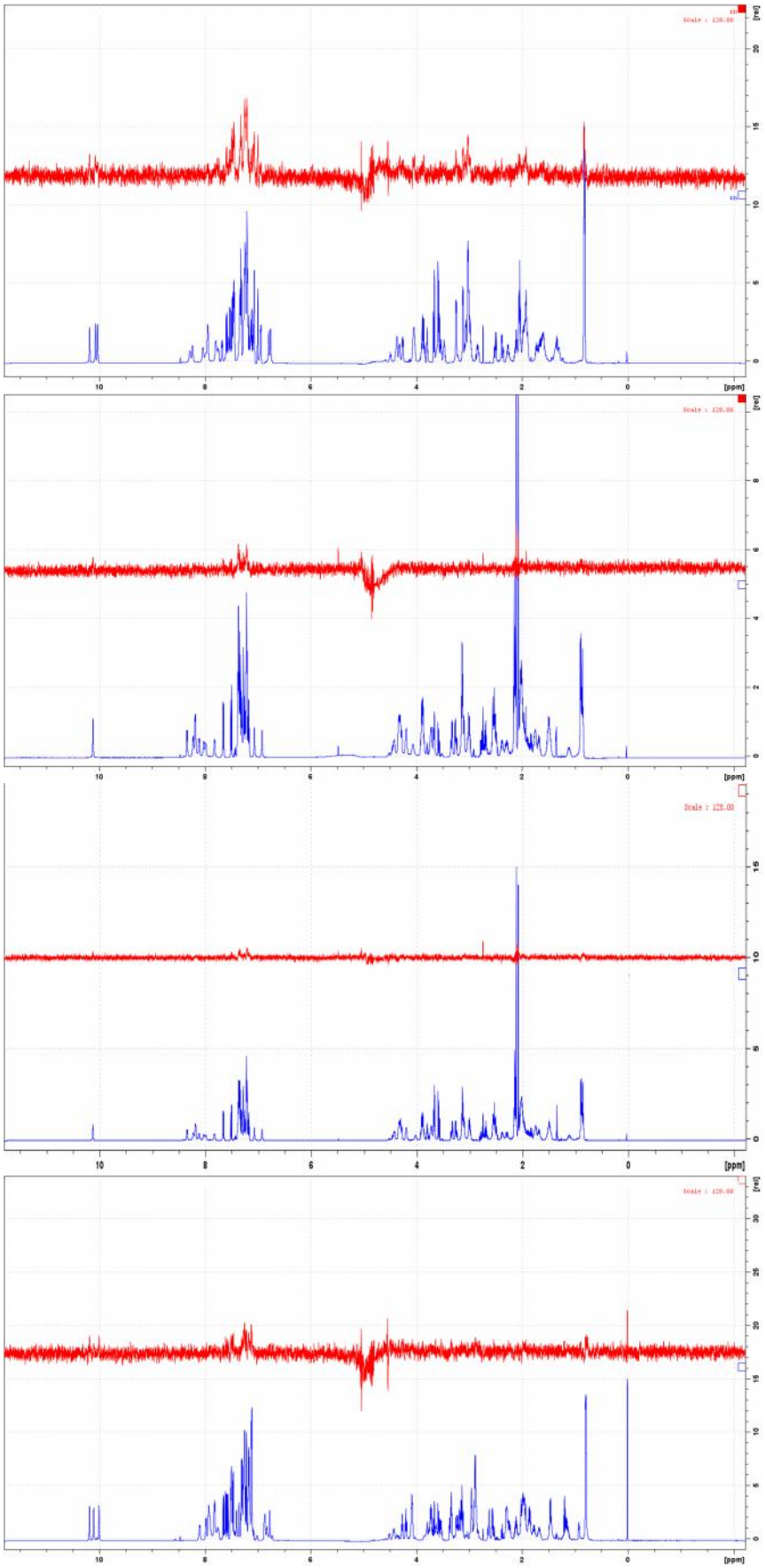
Supplementary Table 4.5. Main proteins hit by peptide FI in pBLAST.

	Description	Organism	E value	Accession	Characteristics
1	hypothetical protein	*Hymenobacter norwichensis*	1.7	WP_022823267.1	Newly defined bacteria isolated from the air in Norwich, England
2	PREDICTED: uncharacterized protein LOC103374163	*Stegastes partitus*	30	XP_008302425.1	Fish
3	hypothetical protein	*Oceanobacillus iheyensis*	33	WP_011064616.1	Isolated from deep sea sediment
4	hypothetical protein	*Oceanobacillus kimchii*	33	WP_017795246.1	N/A
5	hypothetical protein SAMD00019534_081110	*Acytostelium subglobosum LB1*	51	GAM24936.1	Slime mold
6	hypothetical protein Psta_1654	*Pirellula staleyi DSM 6068*	105	ADB16329.1	Water bacterium
7	putative ATPase (AAA+ superfamily)	*Spirochaeta africana*	110	WP_014455122.1	Non-pathogenic
8	hypothetical protein DFA_02459	*Dictyostelium fasciculatum*	111	XP_004366844.1	Soil-living amoeba. Slime mold
9	bZIP transcription factor 29	*Saccharum hybrid cultivar CoC 671*	136	AHZ96966.1	Sugarcane
10	hypothetical protein CAPTEDRAFT_84087	*Capitella teleta*	147	ELU15435.1	Polychaete worm
11	hypothetical protein PGTG_12583	*Puccinia graminis f. sp. tritici CRL 75-36-700-3*	148	XP_003332555.1	Plant pathogen. Fungi. Possible allergen.
12	PREDICTED: basic leucine zipper 9-like	*Setaria italica*	148	XP_004951741.1 XP_004951740.1 XP_004951739.1 and others	Millet (cultured plant)
13	PREDICTED: uncharacterized protein LOC100192650 isoform X1	*Zea mays*	148	XP_008681056.1 NP_001131334.1	Corn
14	BZO2H2	*Zea mays*	148	ACG31513.1	
15	hypothetical protein SEPMUDRAFT_76258	*Sphaerulina musiva SO2202*	152	EMF16749.1	Fungi. Plant pathogen
16	hypothetical protein ACD_75C02113G0003	*uncultured bacterium*	153	EKD34949.1	N/A
17	PREDICTED: uncharacterized protein LOC103449949	*Malus domestica*	187	XP_008387494.1	Apple
18	hypothetical protein	*Nocardia jiangxiensis*	196	WP_040823595.1	Soil actinomycetes. Some Nocardia cause infections in humans.
19	GTP cyclohydrolase	*Streptomyces cellulosae*	196	WP_030663047.1	Inhibits Gram positive bacteria
20	aminoglycoside phosphotransferase	*Mycobacterium sp. JLS*	203	WP_011855250.1 WP_011559205.1	Includes pathogens known to cause tuberculosis and the Hansen's strain of leprosy.
21	restriction modification system DNA specificity domain protein	*Desulfitobacterium hafniense DCB-2*	204	ACL21964.1	Environmental bacteria
22	hypothetical protein L798_03101	*Zootermopsis nevadensis*	207	KDR07320.1	Termite
23	predicted protein	*Fibroporia radiculosa*	208	CCM03831.1	Fungus
24	flagellar motor protein MotB	*Microbulbifer variabilis*	210	WP_040628820.1	Bacteria that produce anticancer antibiotics
25	glycoside/pentoside/hexuronide:cation symporter GPH family	*Firmicutes bacterium CAG:882*	279	WP_022369591.1	Some are pathogenic
26	hypothetical protein BAUCODRAFT_78015	*Baudoinia compniacensis UAMH 10762*	280	XP_007680480.1	Sac fungus

Supplementary Table 4.6. Main proteins hit by peptide GL in pBLAST.



Supplementary Figure 4.16. Burkitt lymphoma mAbs 3D frontal and lateral representation. Heavy (blue) and light (yellow) chains, ABS (grey) and SpA (red) binding residues are represented for **A**: case 079, **B**: case 589 and **C**: case 621. Left pictures are from the ABS perspective and right pictures are lateral to the heavy chain.



Supplementary Figure 4.17. **A**: Peptide VH with mAb 079, **B**: Peptide FI with mAbs 589 and **C**: 621, **D**: Peptide QS with mAb 589.

By nuclear magnetic resonance the amino acids that had contacted the mAbs in each test could be determined. Below is a summary of all contacts detected.

	Signal location (ppm)	Irradiation at -540 Hz Extinction (%)	Irradiation at 7110 Hz Extinction (%)	Amino acid
VH + 079	10.185	0.71	0.54	9 W
	10.075	0.39	0.42	3 W
	7.209	0.44	0.37	9 W / 3 W
	7.074	0.59	0.44	
	3.143	0.44	0.29	3 W
	3.018	0.41	0.33	9 W
	0.819	0.34	0.27	1 V
VH + 589	7.455	0.71	0.54	9 W
	7.259	0.39	0.42	5 F
	7.209	0.44	0.37	9 W
	7.074	0.59	0.44	9 W/3 W
	3.143	0.44	0.29	3 W
	3.018	0.41	0.33	9 W
	0.819	0.34	0.27	1 V
FI + 079	7.649	0.29	0.22	11 W
	7.512	0.42	0.20	
	7.498	0.42	0.20	
	7.358	0.20	0.15	1 F / 4 F
	7.346	0.27	0.15	
	7.334	0.32	0.15	
	7.257	0.43	0.24	11 W
	7.243	0.49	0.24	
	7.206	0.23	0.17	1 F / 11 W
	7.177	0.40	0.13	
	7.164	0.41	0.22	
	2.746	0.18	0.33	5 D
	2.114	0.38	0.08	7 M
	2.080	0.23	0.05	9 M
	1.935	0.07	0.40	9 M / 12 E
	0.882	0.51	0.09	2 I
	0.858	0.660	0.06	
FI + 589	7.663	0.23	0.18	11 W
	7.649	0.15	0.24	
	7.512	0.15	0.24	
	7.498	0.17	0.12	
	7.369	0.19	0.14	
	7.346	0.19	0.14	
	7.218	0.12	0.12	
	7.206	0.18	0.15	
	0.903	0.39	n.a.	2 I
	0.862	0.56	n.a.	
FI + 621	10.123	0.18	0.24	11 W
	7.498	0.13	0.10	
	7.346	0.13	0.10	
	7.220	0.08	0.10	
	7.206	0.18	0.15	
	0.888	0.34	n.a.	02 I

Table is continued on next page

	Signal location (ppm)	Irradiation at -540 Hz Extinction (%)	Irradiation at 7110 Hz Extinction (%)	Amino acid
GL + 079	8.085	0.15	0.22	10Y/9L
	7.986	0.22	0.31	11L
	7.105	0.24	0.22	10Y
	6.818	0.27	0.26	
	4.205	0.24	0.17	7T / 9L
	3.814	0.12	0.10	1G
	3.068	0.17	0.10	3H / 12H /10Y
	2.918	0.26	0.22	10Y
	2.768	0.23	0.18	8N
	1.493	0.35	0.16	2L / 9L /11L
	1.427	0.24	0.12	6A
	1.191	0.25	0.13	4T / 7T
	0.886	0.36	0.18	11L
	0.867	0.39	0.18	2L / 9L
	0.806	0.41	0.18	9L
GL + 589	8.163	0.13	1.02	4 T
	8.119	0.17	0.33	7 T
	8.077	0.20	1.14	10 Y / 9 L
	7.584	0.08	0.20	8 N
	7.121	0.37	0.51	10 Y
	6.847	0.40	0.49	
	6.833	0.44	0.47	
	4.395	0.28	0.29	2 L
	4.297	0.40	0.30	7 T
	4.246	0.43	0.32	11 L
	3.926	0.27	0.21	5 S
	3.849	0.22	0.16	1 G
	3.217	0.19	0.12	3 H
	3.140	0.20	0.20	12 H
	3.078	0.36	0.23	10 Y
	2.948	0.43	0.36	
	2.807	0.33	0.21	08 N
	1.573	0.48	0.19	2 L
	1.503	0.62	0.25	
	1.455	0.45	0.21	9 L / 11 L
	1.443	0.46	0.20	6 A
	1.220	0.52	0.21	4 T / 7 T
	1.208	0.46	0.18	
	1.197	0.40	0.16	
	0.925	0.66	0.26	2 L
	0.895	0.76	0.30	11 L
	0.885	0.67	0.27	2 L
	0.845	0.81	0.36	11 L
	0.816	0.81	0.32	

Table is continued on next page

	Signal location (ppm)	Irradiation at -540 Hz Extinction (%)	Irradiation at 7110 Hz Extinction (%)	Amino acid
GL + 621	7.857	0.51	0.45	11 L
	7.831	0.48	0.50	
	7.131	0.53	0.47	10 Y
	7.117	0.47	0.43	
	7.034	0.52	0.44	8 N
	6.849	0.57	0.50	10 Y
	6.835	0.61	0.48	1
	4.238	0.44	0.36	11 L
	3.791	0.37	0.30	1 G
	3.162	0.35	0.25	3 H / 12 H
	1.570	0.56	0.32	2 L
	1.506	0.53	0.27	
	1.447	0.51	0.31	11 L / 9 L
	1.222	0.49	0.27	4 T/ 7 T
	1.188	0.48	0.33	
	0.916	0.58	0.30	2 L
	0.887	0.61	0.32	11 L
	0.832	0.65	0.37	
QS + 589	7.615	0.28	0.24	1 Q / 9 Q/ 4 W/ 8 W (Aromatic)
	7.601	0.24	0.14	
	7.582	0.24	0.22	
	7.569	0.25	0.19	
	7.521	0.33	0.16	4 W/ 8 W/ 12 W/ 5 F/ NH_2 (end) (Aromatic)
	7.500	0.32	0.19	
	7.480	0.28	0.21	
	7.466	0.35	0.21	
	7.311	0.24	0.14	
	7.254	0.31	0.22	
	7.234	0.37	0.28	
	7.217	0.24	0.16	
	7.175	0.28	0.15	
	7.128	0.26	0.18	
	7.117	0.17	0.13	

Supplementary Table 4.7. Amino acid contact summary of each peptide and antibody NMR experiment. Signal location represents the x axis position of the signal peak in the proton spectrum. The next two columns show the relative signal height (in percentage) for each irradiation side (left at 7110Hz and right at -540 Hz). Amino acid notation is composed of its position in number followed by its universal one-letter symbol, where 2 L is a leucine in position 2 in the peptide.

Final conclusions

This written compilation presents a series of scientific approaches to understand several aspects of Burkitt lymphoma (BL) pathogenesis, with a focus on paediatric cases.

The first part reports on a new method for gene expression profiling of aggressive B-cell lymphomas. The innovative aspect of the presented methodology is the fact that instead of the usual and scarce fresh-frozen biopsy materials the more readily available and numerous paraffin-embedded biopsy blocks can be used. Together with the latter, the protocol described avoids amplification bias of genetic material and is able to count individual molecules in an absolute way, where previously mRNA had to be amplified and its quality inferred. A further validation of the newly generated B-cell lymphoma classifiers is manifested in the second publication of this thesis, which focuses on uncommon BCL2-expressing BL. We describe a good commercially available immunohistochemical marker for BCL2 and assess the amount of BL cases presenting the protein's expression. More related to other lymphomas, BCL2 expression is confirmed in BL, where its presence is not enough to change the molecular labelling of a BL expressing it. We confirm that a poorer outcome is not seen in BCL2-expressing BL as in other BCL2-positive lymphoma entities.

B-cell lymphoma immunoglobulin expression bias is reviewed, including previously unpublished data of such phenomenon in BL. Moreover, a specific mutation occurring in codon 64 of the heavy chain sequence in BL BcRs is modelled in 3D. The conformational changes of the antigen-binding site of the antibody upon reversion of said mutation to its germline sequence are presented. Possible antigenic peptides of the 3 analysed BL mAb are newly determined and characterised as of putative alimentary or pathogenic origin. By reversion of the CDR2 mutation it is demonstrated that the stereotyped change present in BL cases is relevant for antigenic binding localisation.

Burkitt lymphoma survival in the western world is overall 70-80% with today therapeutic regimens, but relapsed cases tend to have a much worse prognosis, with a five-year overall survival of 21%-42%[1]. In the last decades a substantial improvement in outcome has been seen with the implementation of monoclonal antibody therapies (like rituximab). With outcomes starting to reach a cure plateau and expected to even slightly improve in the near future with personalised therapy, a fresh interest in prevention and aetiology in B-cell lymphoma is being developed [2]. Targeted therapies are still a hope for BL patients in Africa, where intensive chemotherapeutic regimens are hard to implement [3].

For BL in particular and B-cell lymphomas in general, new efforts in risk assessment combined with aetiology studies and improved prevention policies could bend the actual static rate of newly diagnosed lymphatic neoplasms per year.

References

[1] H. Kim, E.S. Park, S.H. Lee, H.H. Koo, H.S. Kim, C.J. Lyu, S.E. Jun, Y.T. Lim, H.J. Baek, H. Kook, J.W. Lee, H.J. Kang, K.D. Park, H.Y. Shin, H.S. Ahn. Clinical outcome of relapsed or refractory burkitt lymphoma and mature B-cell lymphoblastic leukemia in children and adolescents. Cancer research and treatment official journal of Korean Cancer Association 46 (2014) 358–365.

[2] S.M. Mbulaiteye, L.M. Morton, J.N. Sampson, E.T. Chang, L. Costas, S. de Sanjosé, T. Lightfoot, J. Kelly, J.W. Friedberg, W. Cozen, R. Marcos-Gragera, S.L. Slager, B.M. Birmann, D.D. Weisenburger. Medical history, lifestyle, family history, and occupational risk factors for sporadic Burkitt lymphoma/leukemia: the Interlymph Non-Hodgkin Lymphoma Subtypes Project. Journal of the National Cancer Institute. Monographs 2014 (2014) 106–114.

[3] R. Schmitz, R.M. Young, M. Ceribelli, S. Jhavar, W. Xiao, M. Zhang, G. Wright, A.L. Shaffer, D.J. Hodson, E. Buras, X. Liu, J. Powell, Y. Yang, W. Xu, H. Zhao, H. Kohlhammer, A. Rosenwald, P. Kluin, H.K. Müller-Hermelink, G. Ott, R.D. Gascoyne, J.M. Connors, L.M. Rimsza, E. Campo, E.S. Jaffe, J. Delabie, E.B. Smeland, M.D. Ogwang, S.J. Reynolds, R.I. Fisher, R.M. Braziel, R.R. Tubbs, J.R. Cook, D.D. Weisenburger, W.C. Chan, S. Pittaluga, W. Wilson, T.A. Waldmann, M. Rowe, S.M. Mbulaiteye, A.B. Rickinson, L.M. Staudt. Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics. Nature 490 (2012) 116–120.

Annex

Primer Tables

Amplification primers

HEAVY CHAINS

MPI-079 (IGV4-39*01; IGHD2-2*02; IGHJ3*01)

VH4-39-for1:	5'-GGTCCAACTCATAAGGGAAATG-3'
VH4-39-for2:	5'-CTTAAATTCAGGTCCAACTCATA-3'
VH4-39-for3:	5'-GGGAAATGCTTTCTGAGAGTCA-3'
JH3-01-rev1:	5'-CTGAAGAGACGGTGACCATTG-3'

MPI-589 (IGV4-39*01; IGHD6-13*01; IGHJ6*04)

VH4-39-for1:	5'-GGTCCAACTCATAAGGGAAATG-3'
VH4-39-for2:	5'-CTTAAATTCAGGTCCAACTCATA-3'
VH4-39-for3:	5'-GGGAAATGCTTTCTGAGAGTCA-3'
JH6-04-rev1:	5'-TCTTACCTGAGGAGACGGTGACCG-3'
JH6-04-rev2:	5'-TCTTACCTGAGGAGACGGTGA-3'
JH6-04-rev3:	5'-TCTTACCTGAGGAGACGGTG-3'

MPI-621 (IGV4-39*07; IGHD4-4*01; IGHJ3*01)

VH4-39-for1:	5'-GGTCCAACTCATAAGGGAAATG-3'
VH4-39-for2:	5'-CTTAAATTCAGGTCCAACTCATA-3'
VH4-39-for3:	5'-GGGAAATGCTTTCTGAGAGTCA-3'
JH3-01-rev1:	5'-CTGAAGAGACGGTGACCATTG-3'

LIGHT CHAINS

MPI-079 (IGLV3-12*01 and IGLV1-51*02)

VL3-12-for1:	5'-GCCTCAGCCATGGCCTGGA-3'
VL1-51-for1:	5'-CAGGACTCGGGACAATCTTCA-3'
JLcons-rev1:	5'-CTAGGACGGTCAGCTTGGTCCC-3'

MPI-589 (IGKV3-20*01)

VK-3-20-for1:	5'- CTGCTCAGTTAGGACCCAGA-3'
JK2-01-rev1:	5'- GTTTGATCTCCAGCTTGGTCCC-3'

MPI-621 (IGLV1-47*02)

VL1-47-for1:	5'-AGGATTCAGGACAATCTCCAGC-3'
VL1-47-for2:	5'-GATTCAGGACAATCTCCAGC-3'
JLcons-rev1:	5'-CTAGGACGGTCAGCTTGGTCCC-3'

Table P1. Amplification primers for the three Burkitt lymphoma cases analysed.

Primer combinations

HEAVY	MPI-079	VH4-39-for1	JH3-01-rev1
		VH4-39-for2	JH3-01-rev1
		VH4-39-for3	JH3-01-rev1
	MPI-589	VH4-39-for1	JH6-04-rev1
		VH4-39-for2	JH6-04-rev1
		VH4-39-for3	JH6-04-rev1
		VH4-39-for1	JH6-04-rev2
		VH4-39-for2	JH6-04-rev2
		VH4-39-for3	JH6-04-rev2
		VH4-39-for1	JH6-04-rev3
		VH4-39-for2	JH6-04-rev3
		VH4-39-for3	JH6-04-rev3
	MPI-621	VH4-39-for1	JH3-01-rev1
		VH4-39-for2	JH3-01-rev1
		VH4-39-for3	JH3-01-rev1
LIGHT	MPI-079	VL3-12-for1	JLcons-rev1
		VL1-51-for1	JLcons-rev1
	MPI-589	VK-3-20-for1	JK2-01-rev1
	MPI-621	VL1-47-for1	JLcons-rev1
		VL1-47-for2	JLcons-rev1

Table P2. Primer combinations used to produce an amplified IGH cDNA fragment.

Ig subtype primers

Primer name	Sequence
FW 079 amplification	5'-AAGCTGACGTCTGTGACTGC-3'
FW 589+621 amplification	5'-ACACGTCCAAGAACCAGTTC-3'
RW membrane amplification	5'-CTTGAACAAGGTGACGGTGG-3'
RW secretion amplification	5'-AGTAGCAGGTGCCAGCTGTG-3'
FW 079 sequencing	5'-GACACGGCTGTGTATTACTG-3'
FW 589 sequencing	5'-GAGGCTGAGCTCTGTGACC-3'
FW 621 sequencing	5'-TGAAGCTGAACTCTGTGACC-3'
RW membrane sequencing	5'-AGAAGAGGCTCAGGAGGAAG-3'
RW secretion sequencing	5'-CATGACCAGGGACACGTTG-3'

Table P3. Primers used to determine the 3 studied BL isotype.

Overhang primers

	Heavy chain	Light chain
079 forward	5'-GCTGAGGAATTCGAGA GTCATGGATCTCATGTG-3'	5'-GTTGACGAATTCTCAT CATGACCTGCTCCCCT-3'
079 reverse	3'-ACCAGTGGCAGAGGAG TCCACGATCGGAGATG-5'	3'-TCGACTGGCAGGATCC AGTCCTTAAGTCAACC-5'
589 forward	5'-GCTGAGGAATTCGAGA GTCATGGATCTCATGTG-3'	5'-CGACACGAATTCGGAA CCATGGAAACCCCAGCG-3'
589 reverse	3'-ACCAGTGGCAGAGAAG TCCACGATCGGTGAGG-5'	3'-CCTGGTTC**G**ACCTCTAG TTTGCTTAAGGTGAAC-5'
621 forward	5'-GCTGAGGAATTCGAGA GTCATGGACCTCCTGT-3'	5'-AAGTGCGAATTCCCAG CATGGCCGGCTTCCCT-3'
621 reverse	3'-ACCAGTGGCAGAGAAG TCCACGATCGGTGAGG-5'	3'-CAGTCGCAGGATCCAG TTCCTCTTAAGTCAACC-5'

Table P4. Green nucleotides (nt) mark the overhangs to be inserted. Red are new restriction sites. Blue mark ATG start codons for translation. In orange are added nt to maintain an appropriate reading frame.

Sequencing primers

Primer name	Sequence
Heavy sequencing forward	5'-GCGCCGTTACAGATCCAAGC-3'
Heavy sequencing reverse	5'-ACAGGGGCCAGTGGATAGAC-3'
Light sequencing forward	5'-GTTACAGATCCAAGCTGTGAC-3'
Light sequencing reverse	5'-GAAGGTGGAAACAGGGTGAC-3'

Table P5. Primers used for general vector sequencing

MCS mutation primers

	Heavy chains
079 forward	5'-GGTCACCGTCTCCTCAGCTAAAACAACAGC-3'
079 reverse	5'-GCTGTTGTTTTAGCTGAGGAGACGGTGACC-3'
589 and 621 forward	5'-GGTCACCGTCTCTTCAGCTAAAACAACAGC-3'
589 and 621 reverse	5'-GCTGTTGTTTTAGCTGAAGAGACGGTGACC-3'

	Light chains
079 forward	5'-CAAGCTGACCGTCCTAGGCCAGCCCAAGTCTTC-3'
079 reverse	5'-GAAGACTTGGGCTGGCCTAGGACGGTCAGCTTG-3'
589 forward	5'-CAAGCTGGAGATCAAAGGCCAGCCCAAGTCTTC-3'
589 reverse	5'-GAAGACTTGGGCTGGCCTTTGATCTCCAGCTTG-3'
621 forward	5'-CAAGGTCAGCGTCCTAGGCCAGCCCAAGTCTTC-3'
621 reverse	5'-GAAGACTTGGGCTGGCCTAGGACGCTGACCTTG-3'

Table P6. Multiple cloning site mutation primers

Buffer and reactions list

PCR reaction set-up
5µl of 5x Kapa HiFi buffer
0.5µl of dNTPs (10mM each)
0.5µl of primer FW (10pmol)
0.5µl of primer RW (10pmol)
0.5µl of Kapa HiFi Polymerase
200ng of cDNA template
Add H_2O to 25µl of total volume

Sequencing reaction set-up
4µl of 5x sequencing buffer
2µl of premix
0.5µl of 10pmol/µl forward or reverse primer
40ng of the cleaned PCR sample
Add H_2O to 20µl of total volume

Temp.	Time	
96°C	20sec	x 27
60°C	4min	
10°C	∞	

Enzimatic restriction reaction
0,3µl of restriction enzyme at 20U/µl
1,5µl of 10x restriction buffer
1µl of 20mg/ml BSA
500ng of plasmid DNA
Add H_2O to 15µl

Lysis buffer for protein extraction
100ml of PBS
2ml of triton
0.4ml of 0.25 M EDTA

IgG purification buffers
Binding buffer
20mM of monosodium phosphate (NaH_2PO_4) at pH7

Elution buffer
100mM of glycine ($C_2H_5NO_2$) at pH2.7

Neutralisation buffer
1M of TrisHCl at pH9.0
It is used in 1/10 of the final elution volume.

Western blots
Loading dye, 6x
0.35 M Tris-HCl
10.28% SDS
36% glycerol
5% β-mercaptoethanol
0.012% bromophenol blue

10x TBS
24.2gr trizma
80gr NaCl
$H_2O_{dest.}$ to 1l, pH7.6

TBST
1x TBS
0.01% tween-20

Transfer buffer
14.4gr glycine
3gr trizma
200ml methanol
$H_2O_{dest.}$ to 1l

Blocking buffer
1x TBST
5% milk powder

Stripping buffer
7.57gr Tris
20gr SDS
$H_2O_{dest.}$ to 1l, pH 6.7. Store without β–mercaptoethanol.
Add 6.98 ml of β-mercaptoethanol when stripping.

Phage display
PEG/NaCl solution
14.6gr NaCl (2.5M final)
20gr PEG 8000 (20% final)
100ml of H_2O, autoclave

Iodide buffer
10mM of TrisHCl, pH8
1mM of EDTA
29,92gr of NaI (4M)
Autoclave and store at RT, in the dark

Tris-HCl solution
1M TrisHCl, pH9

IPTG/Xgal
1.25gr IPTG
1gr Xgal
25ml N,N-Dimethylformamide
Filtered through 0.22μm, stored at -20°C
1ml of this stock is used per liter of LB agar plates

Wash solution
1x TBST
Tween quantities: 0.1% for 1st and 0.5% for the 2nd and 3rd pannings

ELISA

Carbonate buffer, 0.05M
5.3gr of Na_2CO_3
5.04gr of $NaHCO_3$
100ml H_2O
pH 9.3-9.6

Wash buffer
1x PBS
0.05% tween-20
Filtered through 0.22μm membrane

Blocking buffer
1x PBS
0.05% tween-20
5% milk powder

Stopping reagent
H_2SO_4 in 0.5M

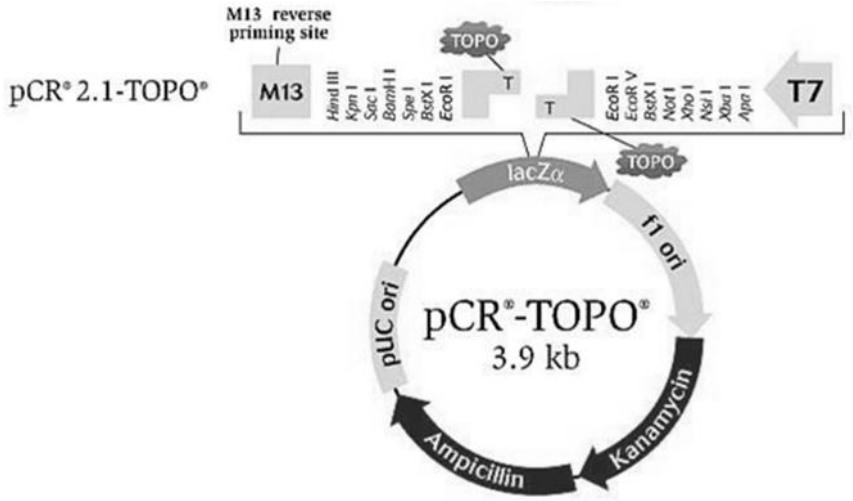
For immunohistochemical stainings

Wash buffer
9,0 gr Tris-Base (Sigma)
68,5 gr Tris HCl (Sigma)
87,8 gr NaCl (Merck)
ad 10 l dest. H_2O
and 5 ml Tween 20(Merck)

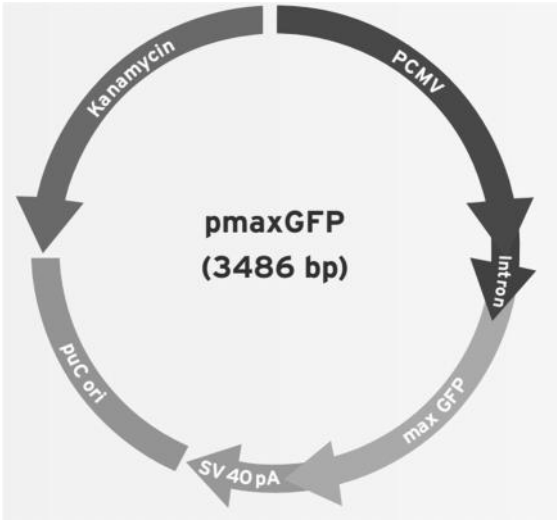
Citrate buffer
10mM, pH 6.0
10,5 gr citric acid monohydrate
mix with ca. 65 ml NaOH (2N)

Methanol H_2O_2
200ml Methanol with 6-8 ml H_2O_2 dest (1x)

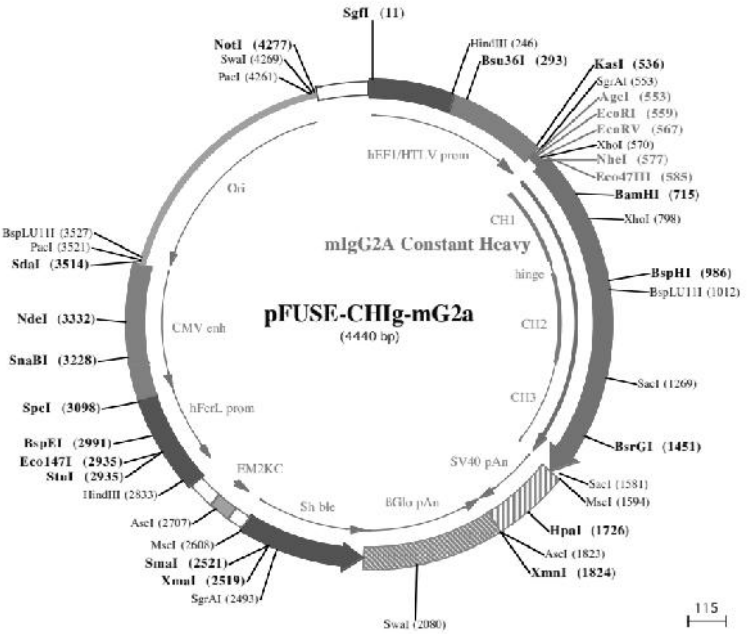
Vectors



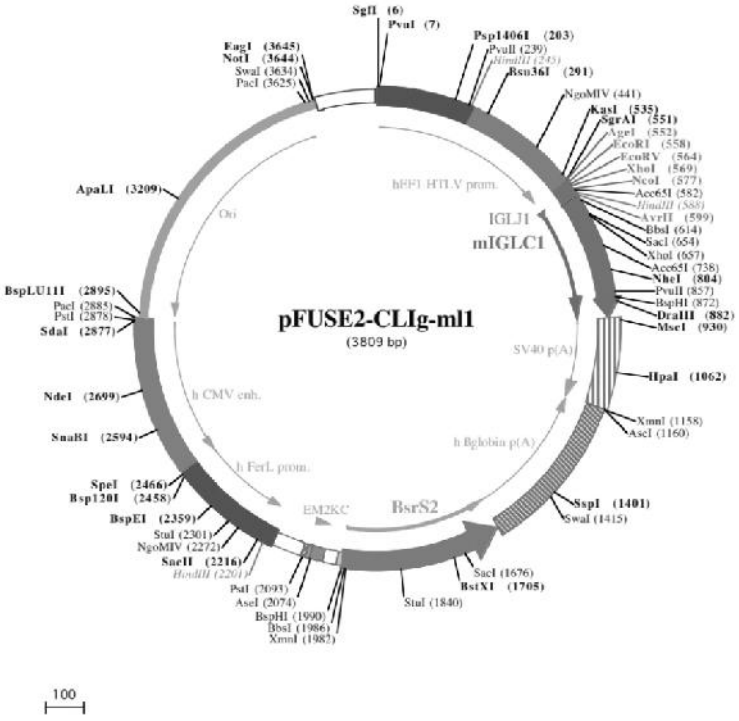
pCR2.1 vector



pmaxGFP vector



pFUSE-CHIg-mG2a vector



pFUSE-CLIg-ml1 vector

Material List

Product	Type	Reference	Company	Company location	Country
10x DPBS	M	14200-067	Thermo Fischer Scientific	Waltham	USA
10X NEBuffer	M	B7002S	New England Biolabs	Ipswich	USA
10x TAE buffer for	M	CL86.1	Carl Roth GmBH	Karlsruhe	Germany
1-Step™ Ultra TMB-ELISA Substrate Solution	M	34028	Thermo Fischer Scientific	Waltham	USA
5x sequencing buffer for BigDye Terminator	M	4339843	Thermo Fischer Scientific	Waltham	USA
6-well culture plates	M	83.392	Sarstedt	Nümbrecht	Germany
ABI PRISM® Genetic Analyzer	E	3500	Life Technologies	Carlsbad	USA
Agar powder	M	05040-1-KG	Sigma-Aldrich	Saint Louis, MO	USA
Agarose (electrophoresis grade)	M	15510-027	Thermo Fischer Scientific	Waltham	USA
Agilent RNA 6000 Nano Kit	M	5067-1511	Agilent Technologies	Santa Clara, CA	USA
Alkaline Phosphatase	M	M0290S	New England Biolabs	Ipswich	USA
Amaxa electroporator machine	E	Nucleofector I	Lonza	Basel	Switzerland
Amaxa® Cell Line Nucleofector® Kit V	M	VCA-1003	Lonza	Basel	Switzerland
Amersham Hybond-P PVDF membrane	M	10600021	GE Healthcare	Little Chalfont	UK
Amicon® Ultra-4 Centrifugal Filter Units	M	UFC803024	Merck-Millipore	Darmstadt	Germany
Ampicilin sodium salt	M	A9518-5G	Sigma-Aldrich	Saint Louis, MO	USA
Anti-Bcl-2 antibody	M	Clone 100-D5	Dako, Agilent Technologies	Santa Clara, CA	USA
Anti-Bcl-2 antibody	M	Clone E17	Zytomed	Berlin	Germany
Anti-CD10	M	PA0270	Leica	Wetzlar	Germany
Anti-CD20	M	CD20-L26-L-CE	Leica	Wetzlar	Germany
Anti-Ki67	M	PA0230	Leica	Wetzlar	Germany
Anti-lambda light chain antibody	M	ab74316	Abcam	Cambridge	UK
Aquatex	M	1085620050	Merck	Kenilworth, NJ	USA
Bacterial culture plates	M	821,472,001	Sarstedt	Nümbrecht	Germany
BigDye® Terminator v1.1 Cycle Sequencing Kit	M	4337450	Thermo Fischer Scientific	Waltham	USA
Bio Fotometer 8.5mm	E	RS323C	Eppendorf	Hamburg	Germany
Bioanalyzer	E	2100	Agilent Technologies	Santa Clara, CA	USA
BioRad Universal Hood for gel pictures	E	SN75S	BioRad	Hercules, CA	USA
Blasticidin powder	M	ant-bl-10p	InvivoGen	San Diego	USA
β-mercaptoethanol	M	M6250-250ML	Sigma-Aldrich	Saint Louis, MO	USA
Bromophenol blue	M	114391-25G	Sigma-Aldrich	Saint Louis, MO	USA
BSA, molecular biology grade	M	B9000S	New England Biolabs	Ipswich	USA
C-Chip Neubauer Improved	M	DHC-NO1	PeqLab	Erlangen	Germany
Cell culture incubator with copper chambers	E	Heracell™ 150i	Thermo Fischer Scientific	Waltham	USA
Centrifuge	E	Varifuge RF	Heraeus	Hanau	Germany
Centrifuge eppendorf microtubes 0.5ml	M	72.735.002	Sarstedt	Nümbrecht	Germany
Centrifuge eppendorf microtubes 1.5ml	M	72.690.001	Sarstedt	Nümbrecht	Germany
Centrifuge eppendorf microtubes 2.0 ml	M	72.695.200	Sarstedt	Nümbrecht	Germany
Centrifuge for 500ml bottles	E	J-6B	Beckman Coulter	Brea, CA	USA
Citric acid monohydrate	M	1.00244.100	Merck	Kenilworth, NJ	USA
CodeSet, custom	M	N/A	NanoString Technologies Inc	Seattle	USA
cOmplete Protease Inhibitor Cocktail Tablets	M	11697498001	Roche	Basel	Switzerland
Coverslips 22mm, sterile	M	83.1840.001	Sarstedt	Nümbrecht	Germany
Culture bottle large	M	83.3912.002	Sarstedt	Nümbrecht	Germany
Culture bottle medium	M	83.1813.002	Sarstedt	Nümbrecht	Germany
Culture bottle small	M	83.1810.002	Sarstedt	Nümbrecht	Germany
Culture plates	M	83.3902	Sarstedt	Nümbrecht	Germany
CutSmart® Buffer	M	B7204S	New England Biolabs	Ipswich	USA
Cycler 5402	E	N/A	Eppendorf	Hamburg	Germany

Product	Type	Reference	Company	Company location	Country
Cytospin machine, Shandon	E	Cytospin3	Thermo Fischer Scientific	Waltham	USA
DAB+, Liquid	M	K3468	Dako, Agilent Technologies	Santa Clara, CA	USA
DAPI	M	D1306	Thermo Fischer Scientific	Waltham	USA
Dialysis tubing cellulose	M	D9527-100FT	Sigma-Aldrich	Saint Louis, MO	USA
Digital Analyzer 2GEN	E	N/A	NanoString Technologies Inc	Seattle	USA
Discovery Studio Visualizer	S	v 4.1.0.14169	Accelrys Software Inc.-Biovi	San Diego, CA	USA
DMSO (Dimethyl Sulphoxide)	M	D2438	Sigma-Aldrich	Saint Louis, MO	USA
DpnI restriction enzyme	M	ER1701	Thermo Fischer Scientific	Waltham	USA
E. coli K12 ER2738	M	N/A	New England Biolabs	Ipswich	USA
EcoRI restriction enzyme	M	R0101S	New England Biolabs	Ipswich	USA
Ethylenediaminetetraacetic acid (EDTA)	M	E9884-500G	Sigma-Aldrich	Saint Louis, MO	USA
ExpressArt FFPE Clear Total RNAready kit FFPE	M	9009-A100	AmpTec	Hamburg	Germany
EZ-Link™ Micro Sulfo-NHS-LC-Biotinylation Kit	M	21935	Thermo Fischer Scientific	Waltham	USA
Falcon tubes, 15ml	M	62.554.502	Sarstedt	Nümbrecht	Germany
Falcon tubes, 50ml	M	62.547.004	Sarstedt	Nümbrecht	Germany
FBS (Fetal Bovine Serum)	M	10270-098	Thermo Fischer Scientific	Waltham	USA
Fisherbrand™ Filter Cards for	M	22-030-410	Thermo Fischer Scientific	Waltham	USA
Fixogum	M	290117000	Marabu	Stuttgart	Germany
Fluorescent microscope	E	Axioplan 2	Carl Zeiss	Oberkochen	Germany
Folded Filters	M	10311643	Sigma-Aldrich	Saint Louis, MO	USA
Glycerol	M	G5516-500ML	Sigma-Aldrich	Saint Louis, MO	USA
Glycine ($C_2H_5NO_2$)	M	50046	Sigma-Aldrich	Saint Louis, MO	USA
Goat Anti-Mouse IgG2a heavy chain (HRP)	M	ab97245	Abcam	Cambridge	UK
Goat anti-mouse lambda chain	M	A90-121A	Bethyl Laboratories Inc.	Montgomery, TX	USA
GraphPad Prism	S	v 5.03	GraphPad Softeare, Inc.	La Jolla, CA	USA
Greiner CELLSTAR® 96 well	M	M3562-60EA	Sigma-Aldrich	Saint Louis, MO	USA
H_2SO_4	M	258105-500ML-	Sigma-Aldrich	Saint Louis, MO	USA
HEK-293 cells	M	ACC 305	DSMZ	Braunschweig	Germany
HiTrap columns, 5ml columns	M	29-0588-06	GE Healthcare	Little Chalfont	UK
HRP anti-M13 monoclonal conjugate	M	27-9421-01	GE Healthcare	Little Chalfont	UK
HRP-goat anti-rabbit IgG (H+L)	M	111-035-003	Jackson ImmunoResearch	West Grove, PA	USA
HRP-rabbit anti-goat IgG (H+L)	M	305-035-003	Jackson ImmunoResearch	West Grove, PA	USA
HRP-rabbit anti-mouse IgG+IgM (H+L)	M	315-035-044	Jackson ImmunoResearch	West Grove, PA	USA
IgG2a isotype control	M	ab18414	Abcam	Cambridge	UK
IgG_{2a} mouse isotype control	M	401501	BioLegend	Sant Diego, CA	USA
IHC Select ® Citrate Buffer pH 6.0	M	21545	Merck	Kenilworth, NJ	USA
IPTG Dioxane free, 2gr	M	AM9462	Thermo Fischer Scientific	Waltham	USA
Isopropanol	M	I9516	Sigma-Aldrich	Saint Louis, MO	USA
Kanamycin	M	K4378	Sigma-Aldrich	Saint Louis, MO	USA
Kapa HiFi PCR Kit	M	07-KK2100-01	Peqlab	Erlangen	Germany
Lab Vision™ UltraVision™ LP Detection	M	TL-060-AL	Thermo Fischer Scientific	Waltham	USA
LB powder	M	L3022-1KG	Sigma-Aldrich	Saint Louis, MO	USA
Lipofectamine 2000 Reagent	M	11668-030	Thermo Fischer Scientific	Waltham	USA
Luminometer	E	Synergy HT	Biotek	Winooski, VT	USA
Mab Trap kit, 1ml columns	M	17-1128-01	GE Healthcare	Little Chalfont	UK
Mayer's hemalum solution	M	1092490500	Merck	Kenilworth, NJ	USA
Methanol	M	8045	JT Baker	Center Valley, PA	USA
Microscope slides	M	1000612	Marienfeld	Lauda Königshofen	Germany
Milk Powder, Blotting Grade	M	170-6404	BioRad	Hercules, CA	USA
Millex-GP Syringe Filter Unit,	M	SLGP033RS	Merck-Millipore	Darmstadt	Germany
Monoclonal Mouse Anti-Human. CD20	M	M075529-2	Dako, Agilent Technologies	Santa Clara, CA	USA
Monosodium phosphate	M	S5011-500G	Sigma-Aldrich	Saint Louis, MO	USA
Mounting medium	M	S3023	Dako, Agilent Technologies	Santa Clara, CA	USA
Mouse monoclonal anti-beta	M	A5441-2ml	Sigma-Aldrich	Saint Louis, MO	USA
MultiMark® Multi-Colored	M	LC5725	Thermo Fischer Scientific	Waltham	USA
N,N-Dimethylformamide (DMF)	M	D4551-500ML	Sigma-Aldrich	Saint Louis, MO	USA

Product	Type	Reference	Company	Company location	Country
NanoDrop 2000	E	ND-2000c-PC	Thermo Fischer Scientific	Waltham	USA
NativeMark™ Unstained Protein Standard	M	LC0725	Thermo Fischer Scientific	Waltham	USA
NativePAGE™ 20x Cathode Buffer Additive	M	BN2002	Thermo Fischer Scientific	Waltham	USA
NativePAGE™ 20x Running	M	BN2001	Thermo Fischer Scientific	Waltham	USA
NativePAGE™ Novex® 4-16% Bis-Tris Protein Gels	M	BN1002BOX	Thermo Fischer Scientific	Waltham	USA
NativePAGE™ Running Buffer Kit	M	BN2007	Thermo Fischer Scientific	Waltham	USA
NativePAGE™ Sample Buffer	M	BN2003	Thermo Fischer Scientific	Waltham	USA
Natrium hydroxide 2N	M	1091361000	Merck	Kenilworth, NJ	USA
NeutrAvidin 96-well plate	M	15129	Thermo Fischer Scientific	Waltham	USA
NheI restriction enzyme	M	R0131S	New England Biolabs	Ipswich	USA
NucleoSEQ® columns	M	740523250	Macherey-Nagel	Düren	Germany
NucleoSpin® Extract II Kit	M	636972	Macherey-Nagel	Düren	Germany
NucleoSpin® Gel and PCR Clean-up kit	M	740.609.250	Macherey-Nagel	Düren	Germany
Nunc Immobilizer Streptavidin F96 Clear	M	436014	Thermo Fischer Scientific	Waltham	USA
Nunc MaxiSorp® flat-bottom 96 well plate	M	44-2404-21	Affymetrix	Santa Clara, CA	USA
NuPAGE Bis-Tris Mini 4-12% gels	M	NP0321BOX	Thermo Fischer Scientific	Waltham	USA
NuPAGE® MES SDS Running Buffer (20X)	M	NP0002	Thermo Fischer Scientific	Waltham	USA
One Shot® TOP10 Chemically Competent *E. coli*	M	C4040-03	Thermo Fischer Scientific	Waltham	USA
Opti-MEM® I Reduced Serum	M	31985-070	Thermo Fischer Scientific	Waltham	USA
pAdVAntage™ Vector	M	E1711	Promega	Madison	USA
Penicillin-Streptomycin	M	15140-122	Thermo Fischer Scientific	Waltham	USA
peqGOLD Plasmid Miniprep Kit II	M	12-6945-02	Peqlab	Erlangen	Germany
Permanent AP Red Kit, colour development	M	ZuC001-125	Zytomed	Berlin	Germany
Pertex	M	41-4041-00	Medite	Orlando, FL	USA
pFUSE2-CLIg-mI1	M	pfuse2-mcll1	InvivoGen	San Diego	USA
pFUSE-CHIg-mG2a	M	pfuse-mchg2a	InvivoGen	San Diego	USA
PhD12 phage library kit	M	E8110S	New England Biolabs	Ipswich	USA
Phenol – chloroform – isoamyl alcohol mixture	M	77617-500ML	Sigma-Aldrich	Saint Louis, MO	USA
Pipette tips low protein bind, 0.5-10ul	M	732824	Brand GmBH	Wertheim	Germany
Pipette tips low protein bind, 50-1000ul	M	732834	Brand GmBH	Wertheim	Germany
Pipettes, 10ml	M	86.1284.001	Sarstedt	Nümbrecht	Germany
Pipettes, 25ml	M	86.1685.001	Sarstedt	Nümbrecht	Germany
Pipettes, 5ml	M	86.1253.001	Sarstedt	Nümbrecht	Germany
pmaxGFP®	M	N/A	Lonza	Basel	Switzerland
Polyethylene glycol 8000 (PEG)	M	1546605-1G	Sigma-Aldrich	Saint Louis, MO	USA
Polyvalent centrifuge 5430R	E	N/A	Eppendorf	Hamburg	Germany
PonceauS solution	M	P7170-1L	Sigma-Aldrich	Saint Louis, MO	USA
Premix Buffer Concentrate	M	401446	Thermo Fischer Scientific	Waltham	USA
PrepStation 2GEN	E	N/A	NanoString Technologies Inc	Seattle	USA
Primers (general)	M	N/A	Eurogentec	Liège	Belgium
Protein glycerol	M	P049.1	Carl Roth GmBH	Karlsruhe	Germany
Protein LoBind tubes, 1.5ml	M	30108.116	Eppendorf	Hamburg	Germany
PureLink® Quick Plasmid Miniprep Kit	M	K2100-10	Thermo Fischer Scientific	Waltham	USA
PyMOL	S	v 1.7.4.5	Schrödinger, LLC	New York, NY	USA
QIAquick PCR Purification Kit	M	28106	QIAgen	Venlo	Netherlands
Qubit asssay tubes	M	Q32856	Thermo Fischer Scientific	Waltham	USA
Qubit fluorometer	E	v 2.0	Thermo Fischer Scientific	Waltham	USA
Qubit® dsDNA BR Assay Kit	M	Q32850	Thermo Fischer Scientific	Waltham	USA

Product	Type	Reference	Company	Company location	Country
Qubit® Protein Assay Kit	M	Q33211	Thermo Fischer Scientific	Waltham	USA
QuikChange II Site-Directed Mutagenesis Kit	M	200523	Agilent Technologies	Santa Clara, CA	USA
Rabbit anti-bovine IgG, IgM, IgA	M	SA1-36042	Thermo Fischer Scientific	Waltham	USA
Rabbit anti-mouse IgG_{2a} heavy chain, 1ml	M	61-0200	Thermo Fischer Scientific	Waltham	USA
RNAse free tubes, 0.5ml	M	AM12300	Thermo Fischer Scientific	Waltham	USA
RNAse free tubes, 1.5ml	M	AM12400	Thermo Fischer Scientific	Waltham	USA
RNAse free tubes, 2.0ml	M	AM12475	Thermo Fischer Scientific	Waltham	USA
RNeasy Mini Kit	M	74104	QIAgen	Venlo	Netherlands
Rotilabo®-syringe filters PVDF, sterile	M	P666.1	Carl Roth GmBH	Karlsruhe	Germany
RPMI medium	M	21875-034	Thermo Fischer Scientific	Waltham	USA
Scrapers	M	83.1831	Sarstedt	Nümbrecht	Germany
SimplyBlue SafeStain	M	LC6060	Thermo Fischer Scientific	Waltham	USA
SnakeSkin Dialysis Tubing Clips	M	68011	Thermo Fischer Scientific	Waltham	USA
Sodium acetate	M	S2889-250G	Sigma-Aldrich	Saint Louis, MO	USA
Sodium bicarbonate (NaHCO3)	M	S6014-500G	Sigma-Aldrich	Saint Louis, MO	USA
Sodium carbonate (Na2CO3)	M	223484-500G	Sigma-Aldrich	Saint Louis, MO	USA
Sodium chloride	M	S3014-5KG	Sigma-Aldrich	Saint Louis, MO	USA
Sodium dodecyl sulfate (SDS)	M	L3771	Sigma-Aldrich	Saint Louis, MO	USA
Sodium iodide (NaI)	M	383112-500G	Sigma-Aldrich	Saint Louis, MO	USA
Stericup-GV, 0.22 µm	M	SCGVU05RE	Merck-Millipore	Darmstadt	Germany
SuperScript® First-Strand Synthesis System for RT-PCR	M	11904-018	Thermo Fischer Scientific	Waltham	USA
Synthetic peptides	M	N/A	Eurogentec	Liège	Belgium
Syringes, 5ml	M	302187	BD Plastipak	Plymouth, MI	USA
T4 DNA Ligase Reaction Buffer	M	B0202S	New England Biolabs	Ipswich	USA
T4 ligase	M	M0202S	New England Biolabs	Ipswich	USA
Target(10x)	M	S1699	Dako, Agilent Technologies	Santa Clara, CA	USA
TC-plate 96 well, Suspens. K	M	83.3926.500	Sarstedt	Nümbrecht	Germany
TE buffer pH8.0	M	AM9849	Thermo Fischer Scientific	Waltham	USA
Tetracyclin	M	87128-25G	Sigma-Aldrich	Saint Louis, MO	USA
TissueQuest	S	v3.0	TissueGnostics	Vienna	Austria
TOPO TA Cloning® Kit (pCR™2.1 Vector and One Shot® *E. coli)*	M	K2000-01	Thermo Fischer Scientific	Waltham	USA
TransIT®-293 Transfection	M	MIR 2705	Mirus	Madison, WI	USA
Triethylamine [(C2H5)3N]	M	90335-250ML	Sigma-Aldrich	Saint Louis, MO	USA
Tris	M	T1503	Sigma-Aldrich	Saint Louis, MO	USA
Tris-HCl	M	PHG0002-100G	Sigma-Aldrich	Saint Louis, MO	USA
Triton, for molecular biology	M	T8787-250ML	Sigma-Aldrich	Saint Louis, MO	USA
Trizma® base	M	T1503-1KG	Sigma-Aldrich	Saint Louis, MO	USA
Trypan Blue Solution, 50ml	M	93595	Sigma-Aldrich	Saint Louis, MO	USA
Turbo Blotter GB-Set Blotting	M	55482300	Sigma-Aldrich	Saint Louis, MO	USA
Tween-20	M	P2287-500ML	Sigma-Aldrich	Saint Louis, MO	USA
Ultra Low IgG FBS (Fetal Bovine Serum)	M	16250-078	Thermo Fischer Scientific	Waltham	USA
VWR Disposable Spatulas	M	231-0103	VWR	Tingalpa	Australia
X-Gal 99% 5g	M	2315.4	Carl Roth GmBH	Karlsruhe	Germany
Zeocin	M	R25001	Thermo Fischer Scientific	Waltham	USA

M, Material; E, Equipment; S, Software

Affidavit

Declaration of authorship for publications with other authors

Chapters 1 and 2 of this thesis are published works. Chapters 3 and 4 have not yet been presented elsewhere, either in written or spoken form. The distribution of the specific tasks within the Chapters is as follows:

Chapter 1.

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

Introduction and Additional Data sections not appearing in the original manuscript were written by Ms Masqué-Soler. She performed, coordinated DMGE analysis, interpreted results and collaborated in the manuscript´s writing. Christian Kohler and Reiner Spang performed bioinformatic analysis. Monika Szczepanowski interpreted results and provided guidance. Wolfram Klapper designed the research, provided funding and wrote the manuscript.

Chapter 2.

> N. Masqué-Soler, M. Szczepanowski, C.W. Kohler, S.M. Aukema, I. Nagel, J. Richter, R. Siebert, R. Spang, B. Burkhardt, W. Klapper. Clinical and pathological features of Burkitt lymphoma showing expression of BCL2 - an analysis including gene expression in formalin-fixed paraffin-embedded tissue. British Journal of Haematology 171 (2015) 501–508.

Ms Masqué-Soler performed, coordinated DMGE analysis, collaborated in the manuscript´s writing and wrote the outlook part independently. Monika Szczepanowski collaborated in results interpretation and provided guidance. Christian Kohler and Reiner Spang performed bioinformatic analysis. Inga Nagel, Sietse M. Aukema and Rainer Siebert analysed genetic data and interpreted results. Birgit Burkhardt provided clinical data. Wolfram Klapper designed the research, provided funding and wrote the manuscript.

Chapter 3.

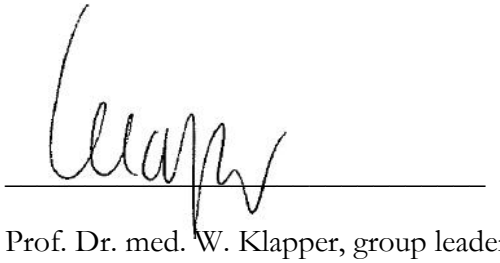
Ms Masqué-Soler wrote the manuscript, performed analysis and assessed the data. Monika Szczepanowski and Wolfram Klapper provided guidance.

Chapter 4.

Manuscript writing, laboratory experiments (other than NMR) and data evaluation was done by Ms Masqué-Soler. Andreas Recke provided phage display guidance. Gitta Kohlmeyer performed the NMR experiments, Gitta Kohlmeyer, Frank Sönnichsen and Ms Masqué-Soler analysed NMR data. Wolfram Klapper evaluated immunohistochemical stainings. Monika Szczepanowski and Wolfram Klapper provided guidance.

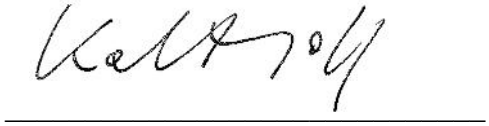
The thesis has been prepared subject to the Rules of Good Scientific Practice of the German Research Foundation.

Kiel, 3rd of November, 2015.



Prof. Dr. med. W. Klapper, group leader of the Haematopathology Section

As supervisor of the dissertation, I confirm the given declaration regarding Ms Neus Masqué Soler authorship.



Prof. Dr. rer.nat. H. Kalthoff, thesis supervisor

Eidesstattliche Erklärung

Hiermit erkläre ich, daß die von mir für das Promotionsverfahren vorgelegte Abhandlung "Gene expression profiling and immunoglobulin stereotypation in Burkitt lymphoma", abgesehen von der Beratung der Betreuer nach Inhalt und Form, meine eigene Arbeit ist. Zwei Kapitel wurden schon vorher veröffentlicht, wie in Kapitel eins (*Blood. 2013 Sep 12;122(11):1985-6*) und Kapitel zwei (*Br J Haematol. 2015 Nov;171(4):501-8*) gezeigt wurde. Der Rest der Arbeit wurde noch nicht an anderer Stelle veröffentlicht und/oder im Rahmen eines Prüfungsverfahrens vorgelegt.

Ich erkläre zusätzlich, dass die Arbeit unter Einhaltung der Regeln guter wissenschaftlicher Praxis der Deutschen Forschungsgemeinschaft entstanden ist.

Kiel, den

_____

Neus Masqué Soler

Acknowledgements

Professor Holger Kalthoff, Professor Wolfram Klapper, and Professor Thomas Röder are much appreciated for taking part and mentoring my thesis procedure.

At the LKR I would like to thank Dr. Monika Szczepanowski, Olivera Batic, Charlotte Botz-von-Drathen, Dr. Artur Gontarewicz, Tanja Engel and Dana Germer for everyday scientific and laboratory support. Also thank you to Dr. Mahsa Kanhlari, Dr. Niklas Vohgt, and Dr. Dmitri Abramov, for letting me learn within their projects. Thank you to Suzanne Griep and Suzanne Pietz. Help from the KinderKrebsInitiative Buchholz/Holm-Seppensen was very valuable. A special mention must go to the late Christiane Mäder, who should be everybody's role model.

At the Pathology Insitute I would like to acknowledge Dr. Jochen Haag, for always listening and coming up with good advice, Sven Thormälen, Dr. Christian Vokuhl, Professor Ivo Leuschner, Dr. Eva Simon, Dr. Franziska Wilhelm, Michael Weiss, Hanna Mindak and Sandra Krüger.

There are people outside the everyday workplace who nevertheless helped me and I would like to name them here: Professor Sönnichsen and Mis Gitta Kohlmeyer-Yilmaz for the NMR analysis, Professor Matthias Peipp, Britta von Below, Heidi Bosse and Dr. Christian Kellner in Kiel; Dr. Steffen Möller and Dr. Andreas Recke in Lübeck; and Professor Rainer Spang, Dr. Christian Kohler and Dr. Michael Altenbuchinger in Regensburg. I wouldn't have been able to do this without your expertise.

To all the other scientists and staff which I have had the chance of collaborating with during my PhD process, thank you.

On the music department, I would like to mention: Reggie; Fred, John, Roger and Brian; Norah; Mark; Idina, Kristin and Norbert; Sara, Tim, David, Eric and John; Bebe and Ann; Kate T; Kate N; Billy, Michael and Haydn; Chris, Matt and Dominic; Mark, Howard, Jason, Rob, and Gary; Rufus; Guy, Craig, Pete, Mark and Richard; Tom, Russell, Edward and Chris; Dave, Abi, Boyan, Russell, Paul and Sean; Jack and Meg; Tom, Jonny, Ed, Phil and Colin; Lluís; Michael; Dave, Taylor and Nate; Florence, Isabella, Robert, Tom, Chris and Mark; Hannah, Dan and Dot; Jonathan, Jeremy, Alex and Michael. Thank you for making everything better.

I would like to thank the friends I have met in Kiel, who slowly became family.

Finally, to my father Josep, my brother Joan, my grandparents Angelina, Joan, Valentí and Àngela, and to Santi, for journeying throughout.