CHARACTERIZATION OF GENOMIC ALTERATIONS IN *CIITA* AND THEIR FUNCTIONAL AND CLINICAL IMPLICATIONS IN MALIGNANT LYMPHOMAS

by

Anja Mottok

MD, Johann Wolfgang Goethe-University Frankfurt am Main, 2005

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Pathology and Laboratory Medicine)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

June 2017

Abstract

Emerging evidence that the interplay between tumour cells and reactive immune cells has profound impact on tumour development, evolution and progression inspired the field of cancer research for the last decade. It has become apparent that the evolutionary pressure exerted by the immune system leads to the evolution of various mechanisms by which tumour cells escape immune surveillance. These often include somatically acquired genetic alterations, resulting in disturbed expression of surface molecules or an altered chemokine/cytokine milieu. B cells play an important role in the adaptive immune response and are potent antigen-presenting cells with high expression of major histocompatibility complexes (MHC) I and II. Multiple studies have reported on defective antigen presentation pathways in malignant lymphomas, however, many of the underlying genetic alterations are largely unexplored.

Herein, we applied next generation sequencing techniques and fluorescence *in situ* hybridization to explore the landscape of genetic alterations in *CIITA,* the master transcriptional regulator of MHC class II, in various B cell lymphomas and to determine the spectrum of rearrangement partner genes. The functional impact of these mutations on MHC class II expression and the composition of the tumour microenvironment were subsequently evaluated in cell line model systems, and by immunohistochemistry performed on primary lymphoma specimens. Finally, we integrated our findings with patient outcomes to ascertain the clinical impact.

We discovered that genomic rearrangements and coding sequence mutations in *CIITA* are frequent across B cell lymphoma subtypes and can result in diminished MHC class II expression, coinciding with a lower abundance of CD4- and CD8-positive T cells in the tumour microenvironment. We identified that at least some of the genetic alterations are likely a byproduct of AID-mediated somatic hypermutation, as evidenced by the co-occurrence of mutations in the non-coding region of *CIITA* intron 1. In addition, we described novel translocations involving a broad spectrum of rearrangement partner genes and intra-chromosomal structural variants.

In summary, we established *CIITA* genetic alterations as a frequent immune escape strategy exploited by a variety of malignant B cell lymphomas. These mutations resulted in reduced MHC class II expression and altered microenvironment composition.

Lay Summary

Malignant lymphomas are the fifth most frequent cancer in humans and arise from white blood cells. In certain subgroups of lymphoid cancers, the malignant cells are surrounded by immune cells, yet these cells are unable to muster an effective immune attack against the tumour. In this thesis, we focused on mechanisms how cancer cells escape the immune system by investigating a particular gene, *CIITA.* We found that this gene harbours multiple changes resulting in reduced levels of a molecule on the tumour cell surface, named MHC class II, which is essential for the recognition of cancer cells by the immune system. Loss of MHC class II was accompanied by reduction of immune cells, therefore enabling the tumour to escape the immune attack. Our studies provide information about the prognosis of lymphoma patients and will help develop new drugs leading to higher cure rates for patients suffering from lymphomas.

Preface

The work presented in Chapters 2 and 3 was performed in the Department for Lymphoid Cancer Research at the British Columbia Cancer Agency's Research Centre and were conducted under the auspices of Dr. Christian Steidl. Our work was generously funded by a Program Project Grant from the Terry Fox Research Institute (Grant No. 1023), the Canadian Cancer Society Research Institute, and the Michael Smith Foundation for Health Research. The research presented within this thesis is approved by the BC Cancer Agency – University of British Columbia Research Ethics Board (Certificate Numbers H05-60103, H13-01765, and H14-02304).

This thesis includes published data as indicated in detail in the following paragraph; full references are provided at the end of this preface.

The sequencing analysis of the PMBCL cohort and the *in vitro* functional experiments have been published in entirety. I was involved in designing the research, I analyzed and interpreted data, wrote the manuscript and performed the following experiments: PCR and Sanger sequencing, qRT-PCR, western blots, cloning and ectopic expression of wildtype and mutant *CIITA,* flow cytometry, as well as IHC. The contribution of all other authors is summarized herein: B.W.W. is a co-first author on this paper. He performed PCR, molecular cloning and analyzed sequencing data. F.C.C. and L.C. analyzed data and provided bioinformatics support, F.C.C. helped generating Figure 2.5 and L.C. generated Figure 3.4. E.C. performed library construction and sequencing, S.B.-N. performed FISH, and P.F. evaluated IHC. A.T. was involved in engineering the DEV cell line and performed DNA/RNA extractions, and M.B. helped with flow cytometry experiments and cell sorting. S.R., K.M.T., M.D., D.W.S., R.S. performed experiments. L.M.R. provided conceptual input and R.D.G. and C.S. designed the study and wrote the paper.

The analysis of *CIITA* structural variants in three primary testicular lymphomas is part of a publication with David D. W. Twa, Fong Chun Chan, and me as co-first authors. Each of us contributed equally in designing and performing the research, interpreting the data and writing and editing the manuscript. S.B.-N., K.L.T. and myself performed and analyzed FISH data, while B.W.W. produced and analyzed sequencing data. A.J.M., H.D. and Y.Z. generated BAC capture and sequencing data, which was

also analyzed by R.S.L.. B.H.N., myself and K.M. generated and analyzed IHC data. S.P.S., R.D.M., M.A.M. and D.W.S. provided experimental and editorial input. R.D.G. and C.S. designed the research, analyzed data, wrote the manuscript and C.S. also approved the paper.

Part of the introduction is stemming from a review paper on the genetic basis of immune escape in malignant lymphomas, which was co-authored by me and C.S..

The work on the DLBCL cohort was initiated by Drs. Daisuke Ennishi, David W. Scott and Randy D. Gascoyne. They assembled the patient cohort and curated clinical data. D.E. designed and performed the sequencing experiments and Christoffer Hother performed mutation calling. I was responsible for reviewing cases and confirming pathological diagnoses as well as the assembly of the cases on a tissue microarray. I conducted all pathology-related aspects of this study, performed DNA/RNA extractions, analyzed *CIITA* mutational data and correlated this with IHC and patient outcomes. I generated all figures related to these experiments.

The work on the FL cohort was spearheaded by Drs. Robert Kridel, Fong Chun Chan and Sohrab Shah. R.K. assembled the patient cohort and curated clinical data. R.K. and F.C.C. designed and performed the sequencing experiments and performed mutation calling. S.S. supervised the study. I was responsible for confirming pathological diagnoses and the assembly of the cases on a tissue microarray. I conducted all pathology-related aspects of this study, scored FISH assays, analyzed *CIITA* mutational data and correlated this with IHC and patient outcomes, and generated figures related to these experiments. Lauren Chong helped with the generation of Figures 2.13 and 2.15.

The generation of rearrangement predictions and the graphical visualization (i.e. Figures 2.19 - 2.22) of the oligocapture results (Chapters 2.2.8.2 and 2.3.5) have been performed by Lauren Chong and are included in her master thesis. I contributed to the design of the project, assembled the patient cohort, performed DNA/RNA extractions, as well as validation experiments, and I interpreted the rearrangement predictions.

References:

Mottok A*, Woolcock B*, Chan FC, Tong KM, Chong L, Farinha P, Telenius A, Chavez E, Ramchandani S, Drake M, Boyle M, Ben-Neriah S, Scott DW, Rimsza LM, Siebert R, Gascoyne RD, Steidl C. Genomic Alterations in CIITA Are Frequent in Primary Mediastinal Large B Cell Lymphoma and Are Associated with Diminished MHC Class II Expression. Cell Rep. 2015 Nov 17;13(7):1418-31.

Mottok A, Steidl C. Genomic alterations underlying immune privilege in malignant lymphomas. Curr Opin Hematol. 2015 Jul;22(4):343-54.

Twa DD*, Mottok A*, Chan FC*, Ben-Neriah S, Woolcock BW, Tan KL, Mungall AJ, McDonald H, Zhao Y, Lim RS, Nelson BH, Milne K, Shah SP, Morin RD, Marra MA, Scott DW, Gascoyne RD, Steidl C. Recurrent genomic rearrangements in primary testicular lymphoma. J Pathol. 2015 Jun;236(2):136-41.

* denotes co-first authorship

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AA	amino acid
ABC	activated B cell-like
AID	activation-induced cytidine deaminase
APC	antigen-presenting cell
ba	break-apart
B2M	ß2-microglobulin
BAC	bacterial artificial chromosomes
BCCA	British Columbia Cancer Agency
BCR	B cell receptor
BL	Burkitt lymphoma
BLS	bare lymphocyte syndrome
bp	base pair
CDS	coding sequence
CIITA	class II transactivator
CLC	Centre for Lymphoid Cancer
CN	copy number
CNV	copy number variation
COO	cell-of-origin
CSR	class switch recombination
CTL	cytotoxic T lymphocytes
DLBCL	diffuse large B cell lymphoma
DNA	deoxyribonucleic acid
DSS	disease-specific survival
EBV	Epstein-Barr virus
ER	endoplasmic reticulum
FBS	fetal bovine serum
FC	fold change
FDC	follicular dendritic cell
FDR	false discovery rate

FF	fresh-frozen
FFPE	formalin-fixed, paraffin-embedded
FISH	fluorescence *in situ* hybridization
FL	follicular lymphoma
GC	germinal centre
GCB	germinal centre B cell-like
GEP	gene expression profiling
HL	Hodgkin lymphoma
HLA	human leukocyte antigen
IFN	interferon
Ig	immunoglobulin
IHC	immunohistochemistry
IL	interleukin
kb	kilobase
LCL	lymphoblastoid cell line
MALT	mucosa-associated lymphatic tissue
Mb	megabase
MCL	mantle cell lymphoma
MHC	major histocompatibility complex
mut	mutant
MZL	marginal zone lymphoma
NGS	next generation sequencing
NK	natural killer
NLPHL	Nodular lymphocyte-predominant Hodgkin lymphoma
npFL	non-progressed follicular lymphoma
OS	overall survival
PCR	polymerase chain reaction
PDL	programmed death ligand
pFL	progressed follicular lymphoma
PFS	progression-free survival
PMBCL	primary mediastinal large B cell lymphoma

PTL	primary testicular lymphoma
qRT	quantitative reverse transcriptase
RNA	ribonucleic acid
SHM	somatic hypermutation
SNP	single nucleotide polymorphism
SNV	single nucleotide variant
SV	structural variant
TCR	T cell receptor
TMA	tissue microarray
tFL	transformed follicular lymphoma
TSS	transcriptional start site
TTP	time to progression
TTT	time to transformation
UTR	untranslated region
VAF	variant allelic frequency
WGS	whole-genome sequencing
WHO	World Health Organization
wt	wildtype

Acknowledgements

The work presented herein would not exist without the exceptional research environment that has been established within the Department of Lymphoid Cancer Research and the Centre for Lymphoid Cancer at the BC Cancer Agency Research Centre. I would like to convey my sincere gratitude to my supervisor Christian for his guidance, unconditional support and mentorship over the last 4 years. He has always encouraged me and helped me to grow as a scientist.

I am also grateful to the Chair and Members of my supervisory committee (Dr.'s David Huntsman, Pamela Hoodless, Andrew Weng, and Brad Nelson) for fruitful discussions, their advice and help to stay focused.

I am indebted to the Terry Fox Research Institute, the Canadian Cancer Society Research Institute and the Michael Smith Foundation for Health Research for funding the research presented herein.

I would like to thank my fellow lab members, in particular Bruce Woolcock, Adele Telenius, and Liz Chavez; your expertise and kindness made this work happen. I would like to thank Merrill Boyle, Dr. Barbara Meissner and Susana Ben-Neriah for their help and advice. Drs. Randy Gascoyne, David Scott, Robert Kridel, Pedro Farinha, and Daisuke Ennishi have been great colleagues and approachable mentors, and I am grateful for all the things I have learnt from you. Most of this work would not exist with the exceptional expertise of the bioinformatics team, led by Fong Chun Chan. Especially, I would like to thank Fong and Lauren Chong for their help.

Lastly, I would like to thank my family for their unconditional support and selflessness.

Chapter 1: Introduction

The emergence of the adaptive immune system over 500 million years ago endowed mammals with a highly specific and sophisticated apparatus, not only to resist an immense number of pathogens, but also to develop immunological memory, thereby enabling rapid immune responses and providing the scientific rationale for vaccination [1]. The adaptive immune system relies on three major protein structures, the B cell receptor (BCR), the T cell receptor (TCR) and the major histocompatibility complexes (MHC). Lymphocytes play a crucial role in adaptive immunity, mainly due to their ability to present or recognize antigens and to produce high-affinity antibodies. Those diverse functions necessitate a considerable degree of morphological and functional plasticity, which have arisen over time. The delineation of lymphoid cells into B and T cell compartments, along with the identification of their primary reservoir, were first described about 50 years ago when experiments conducted in chickens revealed the importance of the thymus and the bursa fabricii (the functional, but not anatomical equivalent of the human bone marrow) for development and maturation of T and B lymphocytes, respectively [2,3].

The following paragraphs in this introduction provide an overview of normal B cell maturation, the germinal centre reaction and its importance for the development of B cell malignancies. Furthermore, I review clinical and biological characteristics of B cell lymphomas studied in this thesis and give a summary of current knowledge on MHC class II and its transcriptional activator *CIITA*.

1.1 B cell development and germinal centre reaction

1.1.1 B cell development

Early B cell development from common lymphoid progenitors starts in the fetal liver, whereas later in life this process is mainly confined to the bone marrow. The first phase occurs in an antigen-independent manner, and involves the rearrangement of the immunoglobulin (Ig) loci, which is mediated by RAG1 and RAG2 recombinases [4,5]. This process is of fundamental importance for the generation of Ig diversity, a hallmark of (antibody-producing) B lymphocytes. The genomic region of the Ig loci is uniquely arranged in a way that multiple variable (V) gene segments are located upstream of

diversity (D) and joining (J) gene clusters, a structure termed translocon organization [1]. The variable regions, which eventually interact with antigens, are characterized by conserved segments, that build a framework interrupted by highly variable sequences, also known as “hypervariable” or complementarity-determining regions [6].

The formation of a rearranged Ig is a multi-step process, initiated at the pre-B cell stage, and starts with the rearrangement of the heavy chain locus (*IGH*) on chromosome 14. At first, a D segment is juxtaposed to a J_H gene segment, followed by combination of a V_H segment to DJ_H [6–9]. Subsequently, as a result of a functional rearrangement, a µ-protein is produced. This abandons any further rearrangement at the heavy chain locus and initiates a similar assembly process at the kappa (*IGK*) and lambda (*IGL*) light chain loci on chromosomes 2 and 22, respectively [6,10]. Under normal circumstances, the rearrangement of *IGK* antecedes any potential recombination of the lambda locus, and only if it would render a non-functional protein, *IGL* is rearranged [11].

The multiplicity of V, D and J segments alongside somatic variation, in addition to combinatorial and junctional diversification, creates a nearly indefinite number of possible Ig molecules, thereby overcoming the constraints of limited germline material [12]. It has been estimated, that the process of V(D)J recombination would enable the recognition of 5×10^{13} different molecules [13]. The process of Ig rearrangement needs to be tightly regulated and chronologically orchestrated to ensure that a particular B cell is limited to functional assembly and expression of a BCR with specificity for only one particular antigen, also known as allelic exclusion [9,14–17].

B cells with functional Ig rearrangements and surface expression of a pre-BCR exit the bone marrow as so-called naïve B cells, migrate into the periphery and populate secondary lymphoid organs, such as lymph nodes, spleen and mucosa-associated lymphatic tissues (MALT). Here the B cells normally encounter various antigens, the majority of them presented by T cells, they form primary follicles and eventually germinal centres (GC) to undergo further affinity maturation and functional selection [18,19].

1.1.2 Germinal centre and germinal centre reaction

The germinal centre is an anatomical structure in lymphoid organs and was first described as such in 1884 by Walther Flemming, one of the pioneers in the field of cell biology and cytogenetics [20,21]. Intrigued by the observation of large lymphoid cells and the high frequency of mitotic figures, Flemming thought he had discovered the major source of lymphocytes in the body [22,23]. Although this was eventually disproven, Flemming's initial observations ignited the research in this field, which has subsequently provided profound knowledge about GC physiology and underlying molecular principles.

The GC can be further subdivided into a dark and a light zone based on the histological appearance and functional properties of the respective cells constituting these zones. Zonal separation becomes apparent on day seven after GC formation [24]. The dark zone develops proximal to the lymph node's medullary zone, is comprised of large B cells, so-called centroblasts, which are highly proliferating and undergo genetic alterations, whereas the more peripherally located light zone is dominated by smaller appearing cells. Those cells are termed centrocytes and are intermingled with follicular dendritic cells (FDC) and various T cell subsets, predominantly CD4-positive follicular T helper cells [18,19]. The spatial positioning of the dark and light zones within the lymphoid organs is of immense importance since it strategically places the light zone compartment in close proximity to those anatomical structures, which normally function as access paths for foreign antigens (i.e. sinuses in lymph nodes and spleen or the mucosal surface in tonsils and intestine) [23].

The compartmentalization of the GC into dark and light zones is largely dependent on chemokine and chemokine receptor gradients. CXCR4, a chemokine receptor highly expressed on centroblasts, is thought to be critical for localization of these cells in the dark zone, a concept further substantiated by the higher abundance of the corresponding ligand SDF1 (CXCL12) [24,25]. Centrocytes on the other hand show higher expression of CXCR5 and migrate towards CXCL13. The CXCR5-CXCL13 axis is required for the correct positioning and polarization of the GC and the migration of B cells to the light zone, whereas the formation and segregation of dark and light zones is thought to occur independent of this receptor-ligand interaction [24,26]. Recent

technical advances, real-time imaging experiments and functional studies helped to shed light on the highly dynamic nature of the GC and the biological processes occurring within this anatomic structure (Figure 1.1) [26–29]. A complex network of transcriptional regulators, signaling molecules and cell-cell interactions, subsumed under the term germinal centre reaction, determines the fate of germinal centre B cells; they can undergo cyclic re-entry into the dark zone and experience multiple rounds of receptor editing to eventually pass the selection process in the light zone, which involves FDCs and follicular T helper cells. These positively selected B cells exit the GC as terminally differentiated plasma cells or memory B cells, whereas B cells with reduced affinity of their BCR to a given antigen are predisposed to undergo apoptosis [18,19,30–32].

The germinal centre reaction encompasses two main processes that are required for efficient affinity maturation: class switch recombination (CSR) and somatic hypermutation (SHM), both of which are essential for the maturation of B cells and adaptation to environmental antigens [18,19]. During the process of SHM, point mutations and small indels (insertions/deletions) are introduced into the Ig variable regions, which ultimately form the interface with the antigen and determine the specificity of the antibody. It was estimated that single nucleotide changes occur at a rate of one per 1000 bp per generation [33]. CSR on the other hand is essential for modulating the effector function of an antibody, conferred by the constant region of *IGH,* as well as for the generation of memory responses. The key enzyme to catalyze this “programmed” DNA damage is activation-induced cytidine deaminase (AID; encoded by *AICDA*), a nuclear-cytoplasmic shuttling enzyme which converts deoxycytidine into deoxyuridine [34–36]. AID recognizes and targets cytidine in a specific genomic context and although there have been some controversies about the exact nature of the motif, the majority of the literature acknowledges WRCY/RGWY (W=adenosine or thymidine; R= adenosine or guanosine; C= cytidine and Y= cytidine or thymidine) motifs as AID hotspots [13,37,38]. The process of DNA deamination entails the activation of mismatch repair and base excision repair mechanisms, which result in DNA double-strand breaks and thereby the initiation of CSR and SHM [13].

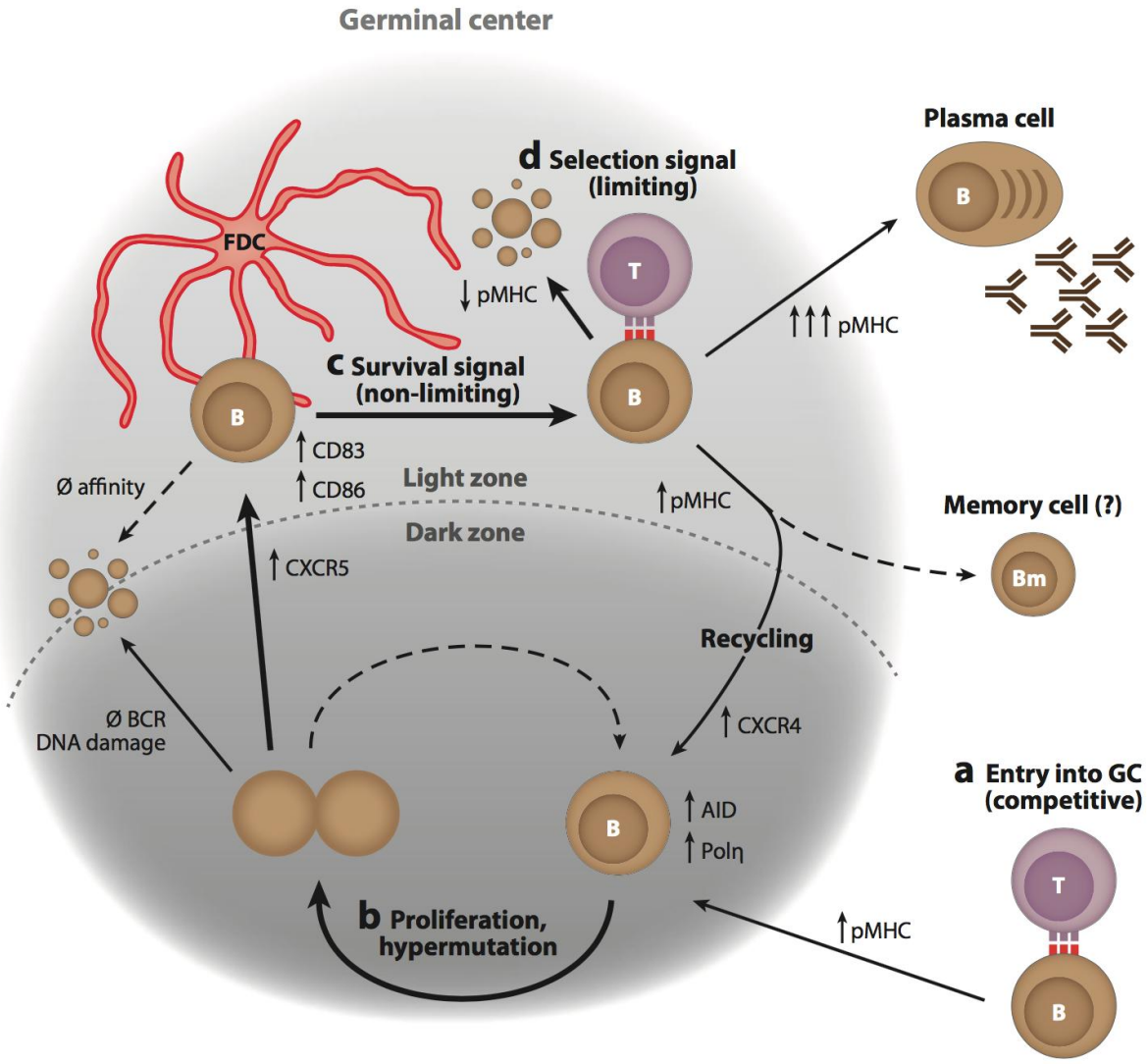

Figure 1.1: Germinal centre reaction. The GC is comprised of a dark zone and a light zone, surrounded by a mantle zone. (a) B cells compete for T cell help at the border of T and B zones and only cells with high affinity are entering the GC. (b) Proliferation and SHM occur primarily in the dark zone, whereas selection and CSR take place in the light zone compartment via interaction with FDCs and follicular T helper cells. (c) Migration of B cells between these two zones is mediated by differential expression of chemokine receptors and their respective ligands. (d) B cells can undergo multiple rounds of receptor editing and affinity selection and leave the GC as plasma cells or memory B cells. Abbreviations: AID, activation-induced cytidine deaminase; BCR, B cell receptor; FDC, follicular dendritic cell; GC, germinal centre; pMHC, peptide-MHC. This figure is reproduced from [19] with permission.

Although the germinal centre reaction is fundamental for the functionality of the adaptive immune response and immunological memory, it also comes with a substantial and potentially dangerous downside as it poses an inherent risk for malignant transformation of germinal centre B cells [18,39,40]. The development of high-throughput sequencing technologies was instrumental in investigating the landscape of

AID off-targets and it has soon been recognized that AID does not only target the Ig heavy and light chain loci, but also several other genes depending on their transcriptional activity and expression at various stages of germinal centre B cell maturation [41–45]. These collateral damage sites rely subsequently on an intact and sufficient DNA damage repair machinery [46], and failure of these mechanisms are believed to contribute to B cell lymphoma pathogenesis [13,18,40,47]. Recent studies have emphasized the importance of structural features, epigenetic accessibility, transcriptional regulation and the presence of super-enhancers for accurate prediction of AID off-targets in a specific cellular context [13].

1.2 B cell lymphomas

The current and recently revised version of the WHO classification of tumours of haematopoietic and lymphoid tissues [48,49] recognizes more than 40 distinct lymphoma entities on the basis of morphology, phenotype, genetic aberrations and clinical features. About 10 % of lymphoma cases belong to the Hodgkin lymphoma (HL) group, whereas the vast majority can be classified as non-Hodgkin lymphomas (NHL). Among all newly diagnosed malignancies, NHL as a group represent the 5th most common cancer type in the U.S. with an incidence of 19.5 cases per 100,000 individuals/year according to the recent Surveillance, Epidemiology and End Results (SEER) report [50]. NHLs can be further subdivided according to their respective lineage in B or T cell lymphomas, with B-NHLs accounting for more than 90 %.

Current pathogenic concepts highlight the acquisition of oncogenic mutations and other potentially transforming events in a definite cell-of-origin (COO) context to give rise to the various malignant phenotypes observed in B cell lymphomas [51,52]. For the majority of described genomic alterations, the germinal centre reaction constitutes the foundation for their occurrence and selection during the process of lymphomagenesis [18,19]. The recognition of distinct entities and molecular subtypes has significantly fostered our understanding of lymphoma biology and is indispensable for clinical decision-making and the development of targeted therapies [53].

1.2.1 B cell lymphomagenesis

B cells at various stages of maturation are characterized by context-dependent gene expression profiles (GEP) and epigenetic programming. Accordingly, initiation and progression of lymphomas are closely related to the microenvironment in which they develop. In light of our knowledge about normal B cell development and the physiological genomic instability of B cells during the germinal centre reaction, it follows that B cell lymphomas can originate from normal counterparts at various developmental stages [51], with the majority stemming from germinal centre B cells (Figure 1.2). Recent gene expression profiling studies suggest that, with the exception of molecular Burkitt lymphomas (BL), most GC-related B cell malignancies originate from cells with a light zone phenotype [39].

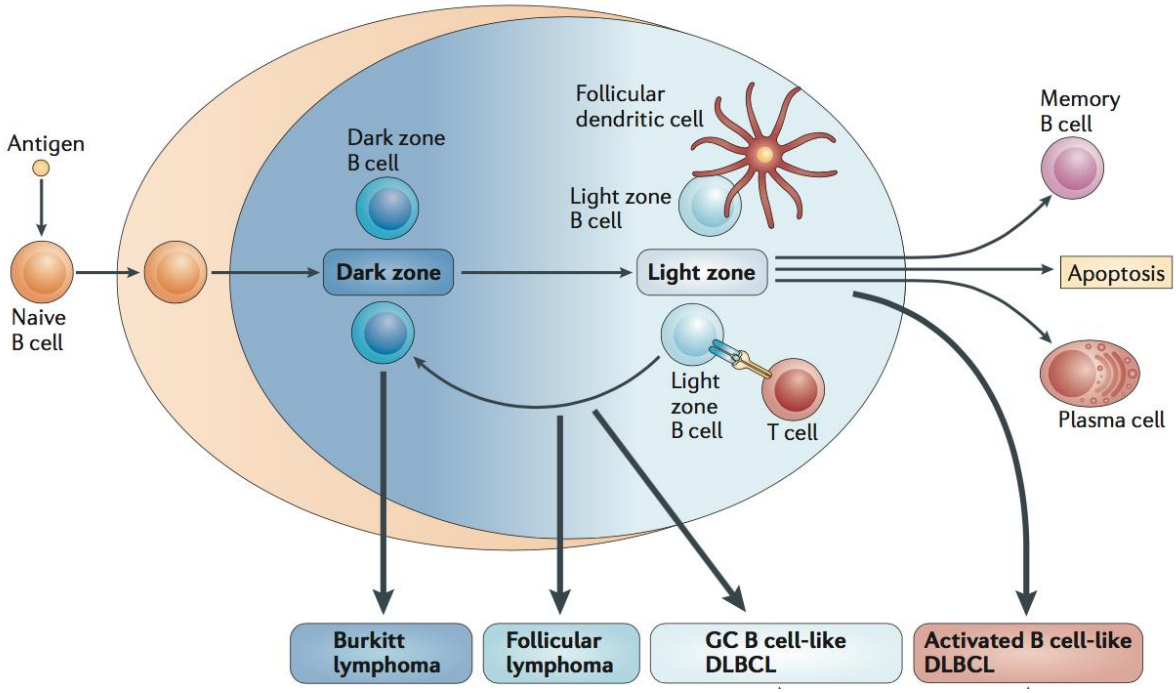

Figure 1.2: Germinal centre-derived B cell malignancies. The biological processes of B cell development and antibody affinity maturation within the GC create an environment for the development of various B cell lymphomas, resembling different developmental stages. Based on GEP, Burkitt lymphoma is closely resembling dark zone B cells, whereas the other major B cell lymphoma subtypes are more related to light zone B cells. Abbreviations: GC, germinal centre; DLBCL, diffuse large B cell lymphoma. Reproduced from [47] with permission.

Similar to solid malignancies, B-NHLs can harbour numerous genetic alterations, such as copy number variations (CNV) or point mutations, which can confer gain- or

loss-of-function phenotypes. As mentioned in section 1.1.1, the processes of V(D)J recombination, SHM and CSR pose a significant risk for malignant transformation of B cells through acquisition of chromosomal translocations and deleterious or oncogenic mutations.

Chromosomal rearrangements are disease-defining biological and diagnostic features in various B cell lymphoma subtypes, e.g. *BCL2* translocations in follicular lymphoma (FL), *CCND1* translocations in mantle cell lymphomas (MCL), *MALT1* translocations in marginal zone lymphomas (MZL), or *MYC* rearrangements in BL [48]. In contrast to other haematological cancers, such as acute leukemias, chromosomal translocations in B-NHLs do not usually lead to the generation of gene fusions and chimeric proteins but rather to a 'promoter swap' [54]. In most cases, the result is the juxtaposition of an oncogene with the promoter or enhancer regions of the Ig loci, leading to dysregulated or exuberant expression of the oncogene. Translocations in B cells can theoretically occur at any time during development, but are more likely to arise as an inadvertent byproduct of the RAG-mediated recombination process (e.g. *IGH-BCL2* translocation t(14;18) in FL) or the AID-dependent mechanisms of CSR and SHM (e.g. *MYC* translocations in BL) [47]. Despite the fact that chromosomal rearrangements in B-NHLs have long been recognized and utilized in diagnostic algorithms in clinical practice, little is known about the exact breakpoint anatomy, and a systematic survey of recurrent chromosomal rearrangements is lacking for the majority of lymphoma entities.

1.2.2 Biological and clinical characteristics of major B cell lymphoma subtypes

1.2.2.1 Diffuse large B cell lymphoma

Diffuse large B cell lymphoma (DLBCL) is the most common lymphoma subtype worldwide and accounts for about 31 % of B-NHLs [55]. It can arise *de novo* or as the result of transformation from a low-grade lymphoma (e.g. chronic lymphocytic leukemia (CLL), MZL or FL). Patients diagnosed with this disease are usually between 60 and 70 years old, however, children and young adolescents can also be affected [55,56]. Histologically, DLBCL is characterized by a diffuse infiltration of lymph nodes or extranodal sites by predominantly large, highly proliferative neoplastic B cells

resembling centroblasts or immunoblasts [48]. Typical extranodal sites include the gastrointestinal tract, the central nervous system, testes, skin and bone [57].

A multitude of studies over the past two decades have demonstrated that the pathogenic mechanisms leading to the development of DLBCL are complex and heterogeneous. The implementation of genome-wide expression profiling and next generation sequencing (NGS) techniques led to: 1) the identification of two distinct molecular subtypes corresponding to different stages of B cell differentiation [58]; and 2) the delineation of the mutational landscape [59–61].

Based on differences in gene expression signatures, the germinal centre B cell-like type (GCB) and a subtype resembling activated B cells (ABC) can be distinguished. Those two subtypes not only differ in their GEP, but are also characterized by hallmark genetic lesions and a remarkably disparate clinical behaviour [62–65]. With the recent developments and emergence of new drugs targeting subtype-specific pathways and mutations [53], the distinction of GCB- and ABC-type DLBCL, commonly referred to as COO-assignment, has become a mandatory task for the diagnosing pathologist with the revision of the WHO classification [49].

The genes most frequently involved in chromosomal translocations in DLBCL are *BCL6* (30 %), followed by *BCL2* (20 %), *MYC* (10 %) and *TBL1XR1*/*TP63* (5 %) [66–69]. In the past decade, a number of high-throughput sequencing studies have shed light on the molecular and genetic alterations contributing to DLBCL pathogenesis. The most commonly mutated genes are involved in epigenetic regulation and chromatin modification and include *EZH2* [70], *CREBBP*, *EP300* and *KMT2D* [59,61,71]. Of crucial importance for DLBCL pathogenesis is the dysregulation of BCL6, either directly through chromosomal translocation and somatic mutations or indirectly via alterations leading to changes in transcriptional activity, such as *MEF2B* mutations [66,72–74].

1.2.2.2 Follicular lymphoma

Follicular lymphoma is the second most common lymphoma subtype, accounting for about 22 % of cases [55], with the highest incidence in North America and Western Europe [48]. FL is considered an indolent, yet mostly incurable cancer, characterized by a slowly progressing disease course in the majority of patients. Similar to DLBCL, FL is

seen mainly in older patients but can occur in children, although pediatric FL differs in terms of histological appearance, pathogenesis, mutational profiles and course of the disease [48,75,76]. As the name implies, FL is a tumour typically growing in follicular structures, which are to some extent reminiscent of normal GCs. The follicles in FL are usually densely-packed and therefore alter the normal lymph node architecture. Furthermore, a mantle zone is frequently absent and the zonal structure, with clearly visible dark and light zones, is disturbed. It has also been demonstrated, by using laser capture microdissection and single cell PCR, that lymphoma cells frequently migrate between neoplastic follicles [77], suggesting that these nodules are rather dynamic structures. The presence of non-neoplastic cells, such as follicular T helper cells, FDCs, macrophages and stromal cells, indicates the dependence of the malignant cell population on external stimuli and survival signals and has been shown to provide prognostic information [78,79]. Indeed, FL is the prototypic example of a cancer achieving ‘re-education’ of the surrounding immune cells to provide a pro-tumoural niche [80] and therefore enabling the lymphoma cells to thrive in this environment.

The translocation t(14;18)(q32;q21), resulting in the juxtaposition of the *BCL2* gene under the control of the *IGH* promoter, is the genetic hallmark of FL and is present in 75-90 % of cases, depending on the detection method employed [81]. Multiple studies have elucidated the mechanisms by which these chromosomal breaks occur and it is established that the *BCL2* translocation in FL is the result of a concerted action involving RAG and AID during early B cell development in the bone marrow [82–84]. The t(14;18) itself is not sufficient for lymphomagenesis, as evidenced by the presence of t(14;18)-positive B cells in otherwise healthy individuals [85,86]. However, the t(14;18) translocation, leading to overexpression of the anti-apoptotic protein BCL2, is pivotal for disease progression and maintenance once an FL has arisen [87–89]. Because the onset of FL cannot be explained by the sole presence of the translocation t(14;18), additional genetic alterations need to occur to drive lymphoma development and progression. NGS techniques have revolutionized the field of cancer genomics and have enhanced our knowledge and understanding of disease biology. Numerous studies have revealed that histone modifiers are among the most frequently mutated genes in FL [59,90,91].

Patients diagnosed with FL are commonly risk-stratified based on clinical observations and measurable parameters, combined together in a prognostic tool called the FL International Prognostic Index (FLIPI) [92]. Recent efforts to refine risk stratification by incorporating mutational signatures of seven genes have led to the development of a new prognosticator, the m7-FLIPI [93].

As mentioned above, FL is an indolent, but incurable disease, meaning that patients affected by this disease will eventually experience relapse, progression or transformation into a high-grade lymphoma. Studies investigating the patterns of progression and transformation by applying conventional cytogenetics or (next-generation) sequencing approaches, often demonstrated divergent clonal evolution to be the most common mechanism [94–98].

1.2.2.3 Primary mediastinal large B cell lymphoma

Primary mediastinal large B cell lymphoma (PMBCL) represents a relatively rare aggressive B cell neoplasm and accounts for only 2-4 % of B-NHLs. Young women in their third to fourth decade of life are frequently affected and, because those patients usually present with bulky tumour masses in the mediastinum, the putative cell of origin is thought to be a thymic medullary B cell [48]. The prototypic PMBCL presents histologically as a tumour of medium to large lymphoid blasts with abundant cytoplasm, some of which can resemble Hodgkin-like cells. Often, the tumour is compartmentalized by delicate collagenous fibrosis and can show extensive areas of necrosis [48].

Global GEP revealed a molecular signature that clearly distinguished PMBCL from other large B cell lymphomas, such as DLBCL, and established a linkage to the nodular sclerosis subtype of classical HL (cHL) [99,100]. Subsequent studies showed that these two entities not only show similarities on the transcriptomic level but also overlapping genomic alterations [101–104]. Moreover, PMBCL and cHL rely on constitutively active signaling pathways, such as the Janus kinase-signal transducer and activator of transcription (JAK-STAT) and nuclear factor-κB (NF-κB) pathway, for proliferation and survival [101,105]. The close relatedness is further substantiated by the introduction of a so-called grey zone category in the 2008 WHO classification to integrate cases which show intermediate features between cHL and PMBCL [48].

From a clinical point of view, patients with PMBCL tend to have better outcomes when treated with multiagent chemotherapy regimens compared to DLBCL [106,107]. More recent therapeutic approaches aim at eliminating the need for mediastinal radiation therapy, which comes with the risk of serious late site effects. A recently conducted phase 2 study using dose-adjusted etoposide, doxorubicin, cyclophosphamide, vincristine, prednisone, and rituximab (DA-EPOCH-R) has demonstrated promising results with event-free survival rates of 93 % [108].

1.2.3 Tumour-microenvironment interactions in malignant B cell lymphomas[1]

Over the past decade we have seen a major shift of focus in cancer research. Newly developed pathogenesis models have deviated from being solely centered on the description of accumulating genetic changes and pathway alterations in malignant cells [109] towards comprehensive consideration of the interactions between tumour cells and non-malignant cells in the tumour microenvironment. The notion of immune cells being significant contributors to cancerogenesis and the importance of cellular crosstalk with malignant cells have led to the recognition of these aspects of tumour biology as an emerging hallmark of cancer ("immune evasion") and enabling characteristic ("tumour promoting inflammation") [110].

Microenvironment-related biology in lymphoid cancers has been explored in a limited number of subtypes with variable contributions of reactive immune cells in the microenvironment. Amongst these, cHL represents the extreme example in a spectrum of diseases that feature a quantitatively dominant microenvironment composed of a variety of non-malignant cell types from both the innate and adaptive immune system. In cHL, these "bystander" cells are believed to be attracted by the Hodgkin- and Reed-Sternberg (HRS) cells as the master recruiters [80,111].

The composition of the tumour microenvironment, and in particular its spatial distribution, can be perceived as a complex interplay of: 1) genetic alterations within the malignant cell population; 2) the extent and dependence on the cellular and molecular crosstalk involving cyto- and chemokines as well as receptor-ligand interactions; and 3) host-specific factors (inflammatory response, systemic immune competence). This

[1] This paragraph has been modified from [111].

results in three highly characteristic blueprints for the microenvironment architecture of malignant lymphomas, termed “re-education”, “recruitment” and “effacement” [80]. FL is the prime example for the ‘re-education’ pattern since the tumour cells are at least to some extent dependent on their microenvironment for proliferation and survival. As described in section 1.2.2.2, the cellular composition of the neoplastic follicles closely resembles reactive GC, with high abundance of FDCs and follicular T helper cells. ‘Recruitment’ of various non-malignant immune cells is a hallmark of cHL. This results in a quantitative dominance of reactive cells over the tumour cells. The other end of the spectrum, ‘effacement’, is best represented by BL, a lymphoma with typically more than 90 % malignant cells. The microenvironment is sparse, in part caused by the ‘independence’ of the lymphoma cells on survival signals provided by the microenvironment [80].

The primary purpose of the immune system is to protect against infectious agents, but it also sufficiently recognizes and eliminates autologous cells displaying non-self antigens or so-called neo-antigens, which – in the case of malignant tumours - are often the result of cancer-specific genetic alterations [112–114]. T cell dependent immune responses involve complex interactions between antigen presenting cells (APCs) and T cells, which engage several stimulatory and inhibitory signaling molecules. In order to avoid non-directional and anomalous reactions that might lead to autoimmunity and excessive tissue damage, the entire process has to be strictly regulated.

The sophisticated apparatus of the adaptive immune response has been exploited by cancer cells and there is ample evidence that interference with anti-tumour immunity is not a passively or randomly occurring process [115–117]. Malignant cells have developed strategies to escape from immune surveillance. These include mutations or deletions of genes involved in antigen presentation or, alternatively, structural rearrangements and copy number alterations (gain/amplification), resulting in overexpression of molecules involved in the induction of peripheral tolerance and T cell exhaustion (so-called immune checkpoint molecules). This ultimately leads to re-programmed and dysfunctional immune cells and, while some of these effects seem to be persistent, a proportion might be reversible and therapeutically targetable [118,119].

Importantly, effective reversal of this altered immune biology in the clinical setting will be accelerated by the identification of genomic and molecular alterations underlying immune privilege and the integration of these findings with clinical and morphological parameters to administer tailored therapy to lymphoma patients.

1.3 Major histocompatibility complexes

The major histocompatibility complexes are glycoproteins on the cell surface and are integral to the normal function of the adaptive immune system. Furthermore, the biology of antigen presentation has major implications for pathologic processes, such as autoimmunity and transplant rejection, but also for tumour immunotherapy [120]. The MHC cluster encompasses a large genomic region harbouring over 100 genes close to 4 Mb in size [121] and can be subdivided into two major groups: MHC class I (with human leukocyte antigen (*HLA)-A*, *-B* and *-C* representing the classical MHC class I genes) and MHC class II (with *HLA-DR, -DP, -DQ* representing the classical MHC class II genes). They have in common the presentation of antigens to effector T cells, however there are also fundamental differences to be recognized. These include tissue-specific patterns of MHC expression, the nature and source of the peptide fragments presented, and the interacting T cell population. MHC class I molecules are expressed on all nucleated cells and present endogenous proteins to cytotoxic, CD8-positive T cells, therefore providing a blueprint of the intracellular proteome and playing an important role for the recognition of ‘self’ and ‘foreign’ [120]. MHC class II molecules are mainly restricted to APCs, they display predominantly exogenous antigens and are required for an immunological response executed by $CD4^+$ T cells [122]. However, under certain circumstances cells deviate from that convention, a phenomenon known as cross-presentation [123]. This process has been described in APCs and leads to the presentation of extracellular antigens (from virus-infected cells or tumour cells) via MHC class I to naïve $CD8^+$ T cells [120,124]. Autophagy, on the other hand, can result in the presentation of cytosolic/endogenous peptides via MHC class II [125–127].

1.3.1 Genomic alterations underlying MHC deficiency

Immune evasion, metastasis and impaired patient outcomes have been, in part, attributed to downregulation of MHC class I and II in a variety of solid and hematological malignancies [128–134] and abnormalities of MHC class I expression represent one of the most frequent changes across different cancer types, allowing tumour cells to avoid destruction by cytotoxic $CD8^+$ T lymphocytes (CTLs) [128,135].

1.3.1.1 MHC class I deficiency[2]

Several mechanisms have been described by which malignant B cells are able to downregulate MHC molecules. Homozygous and heterozygous deletions of the MHC locus on chromosome 6p occur frequently in DLBCL, with a predilection for those arising in 'immune-privileged' sites of the testes or brain [136–139]. Furthermore, the MHC locus is a very common susceptibility locus for the development of a variety of lymphomas, as identified by genome-wide association studies [140–143].

The MHC class I complex is composed of the heavy chain (a transmembrane glycopolypeptide) and the non-covalently bound ß2-microglobulin (B2M) light chain [144,145]. The association with B2M is important for the assembly and stabilization of the entire complex, as well as for maintaining a functionally active conformation in order to present peptides derived from intracellularly degraded proteins [146–148]. Alterations of the *B2M* gene have been described across a variety of solid tumours and malignant lymphomas [149–153]. The mutational pattern with frequent occurrence of loss of the start codon, truncating mutations, deletions and CNVs, as well as biallelic alterations established *B2M* as an important tumour suppressor gene in DLBCL [59,61,154] with a reported frequency of up to 29 %. In contrast, mutations in FL, BL, CLL, MCL and MZL were not or rarely detected [90,154,155]. Recent studies investigating the genetic mechanisms underlying transformation of FL have further demonstrated that *B2M* mutations are enriched in transformed lymphomas with mutation patterns similar to those observed in *de novo* DLBCL, providing evidence for the existence of immune selection pressure during evolution to a high-grade malignancy, and for a linkage between mutations and microenvironment composition [94,96].

[2] This paragraph has been modified from [111].

Interestingly, mutations of *CD58*, a member of the immunoglobulin superfamily and ligand for the CD2 receptor on natural killer (NK) cells [156,157], have been found to co-occur frequently with *B2M* aberrations in *de novo* DLBCL, as well as in transformed FL (tFL), suggesting that *B2M* and *CD58* mutations represent complementary mechanisms to establish immune privilege [94,154]. Specifically, the co-occurrence was attributed to the potentially synergistic effects of reduced recognition by CTLs and inactivating NK cells, since it has been described in earlier studies that escape from CTLs triggers NK cell recognition as part of a compensation mechanism [154,158]. Similar alterations of *B2M* and *CD58* have also been recently identified in HL-derived cell lines and HRS cells from primary tissue biopsies [159–162]. As B2M is indispensable for the assembly of the MHC class I complex, genomic alterations in *B2M* led to concomitant absence of surface HLA-A/B/C staining in mutated DLBCL and HL cases [154,159], a discovery that might also in part explain the reduction of MHC class I expression reported in earlier studies [163,164].

Recently, *NLRC5* (CITA), a new member of the nucleotide-binding domain, leucine-rich repeat protein family, was identified to be involved in the transcriptional control of MHC class I [165,166] along with a potential role in regulating MHC class II transcription [167]. So far, *NLRC5* alterations have been rarely described in malignant lymphomas [59,60,168,169], and functional studies are warranted to provide evidence for a potential link to immune escape in this specific context.

1.3.1.2 MHC class II deficiency

Compared to MHC class I, there has not been nearly as much focus on MHC class II alterations in cancer, probably because of the rather restricted expression in APCs. When interrogating publically available sequencing data across different cancer types, alterations in the classical human MHC class II molecules, encoded by *HLA-DR, HLA-DP* and *HLA-DQ,* are frequently found in glioblastoma multiforme, followed by DLBCL (Figure 1.3).

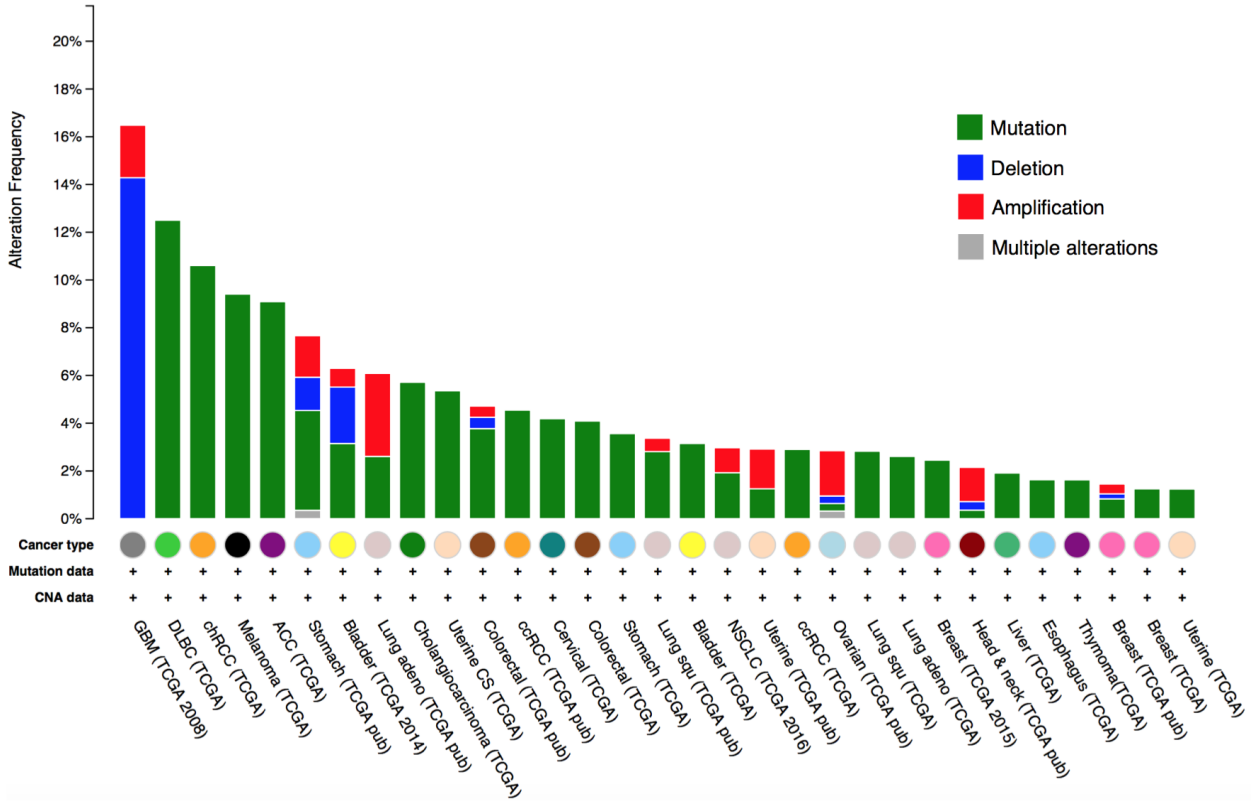

Figure 1.3: Genomic alterations of classical MHC II genes across cancer subtypes. The results shown here are based upon data generated by the TCGA Research Network: http://cancergenome.nih.gov and are visualized using the cBIOPortal for Cancer Genomics http://cbioportal.org/ [170,171].

In their capacity as potent antigen-presenters, B cells normally express MHC class II molecules on their surface to mediate interaction with $CD4^+$ T helper cells. Our group has previously described recurrent structural genomic alterations (specifically, unbalanced chromosomal rearrangements) of *CIITA*, the master regulator of MHC class II transcription, in PMBCL and cHL [172]. These aberrations were proposed to be causative of MHC class II loss, however, a detailed analysis of MHC class II protein expression has not been performed. *CIITA* mutations have also been found in DLBCL [59,61,173] and, furthermore, it appears that concomitantly with *B2M* mutations, and therefore impairment of MHC class I, *CIITA* is frequently mutated in tFL [94]. In contrast, mutations in solid cancers are rarely found (Figure 1.4). Interestingly, the structural genomic aberrations appear to be a unique feature of PMBCL and cHL, since they were rarely observed in other lymphoma subtypes analyzed [172]. A comprehensive

assessment of *CIITA* mutations, as well as correlations with MHC class II expression and microenvironment composition are largely missing.

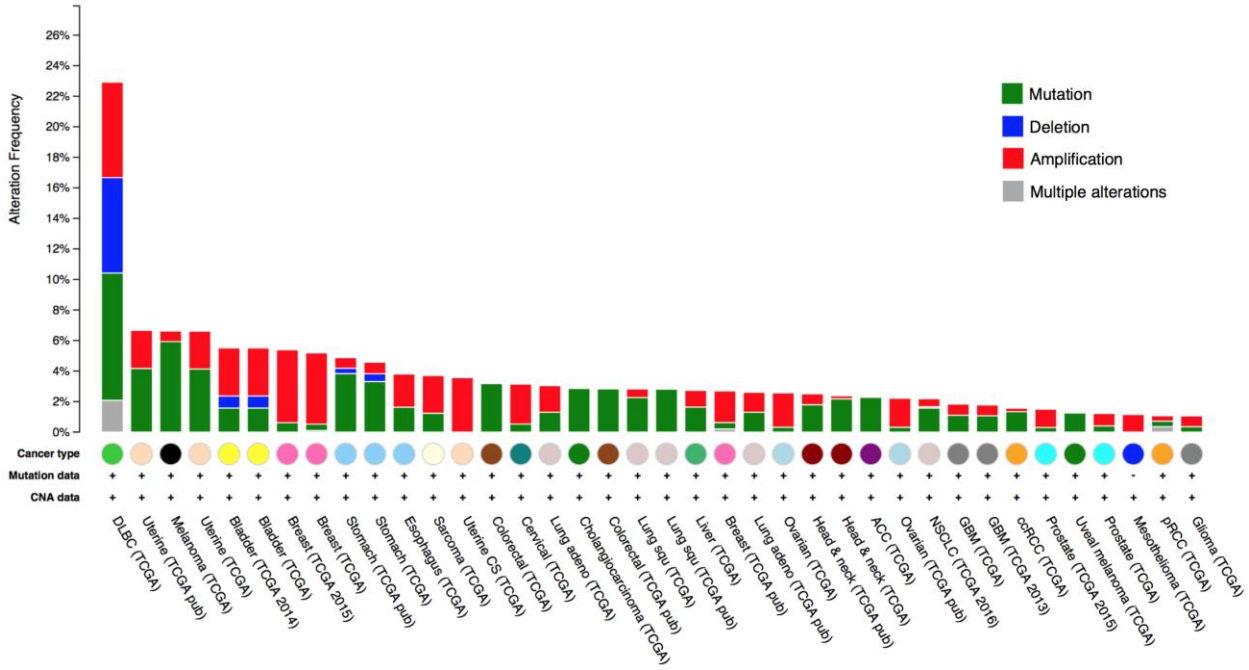

Figure 1.4: Frequency of *CIITA* genomic alterations across different cancer types. The results shown here are based upon data generated by the TCGA Research Network: http://cancergenome.nih.gov and are visualized using the cBIOPortal for Cancer Genomics http://cbioportal.org/ [170,171].

1.3.2 Antigen presentation via MHC class II complexes

Beside the importance of MHC class II-mediated antigen presentation for the protection against pathogens and tumours, it is also essential for selection processes during early T cell development in the thymus and maintenance of self-tolerance [174]. These important immunological functions necessitate tight regulation of MHC class II expression with regards to cell type specificity, activation and developmental stages, to assure an adequate response to pathogens while avoiding any reactivity against the host. MHC class II molecules are typically expressed by thymic epithelial cells and professional APCs, including not only dendritic cells and monocytes/macrophages, but also B cells [122,174,175]. Some studies have highlighted the dependence of MHC class II expression and the $CD4^+$ T cell response on the APC, which is involved in the MHC class II-mediated antigen presentation. Macrophages for example are relying on

activating stimuli to induce MHC class II expression, whereas B cells are characterized by a rather constitutive expression, which can be enhanced by stimulatory signals, such as interleukin (IL) 4 [176,177]. However, the expression of MHC class II within the B cell lineage is strictly regulated. It is absent in early pro-B cells, increases in subsequent developmental stages to reach a maximum in GC B cells, and vanishes in plasma cells [178,179].

The classical human MHC class II molecules are encoded by three genes: *HLA-DR, HLA-DP* and *HLA-DQ*, located on the short arm of chromosome 6. These genes are highly polymorphic and differ in their capability to bind peptides [122]. The MHC class II complex formation starts with the assembly of a heterodimer, consisting of α- and β-chains, which subsequently liaises with the invariant chain (Ii; encoded by *CD74*) in the endoplasmic reticulum (ER). This step is important for stabilization of the entire complex, which gets then transported to the MHC class II compartment (MIIC) [180]. MIIC represents a late endosome and the cellular compartment where peptide loading of MHC class II is taking place. Proteolytic cleavage of the invariant chain results in a small residual peptide (CLIP, class II-associated Ii peptide), that occupies the peptide-binding groove of MHC class II until it is eventually exchanged for a peptide typically derived from the endosomal/endocytic pathway [181]. This step is dependent on the presence of HLA-DM, which functions as an MHC class II chaperone (Figure 1.5) [120,122].

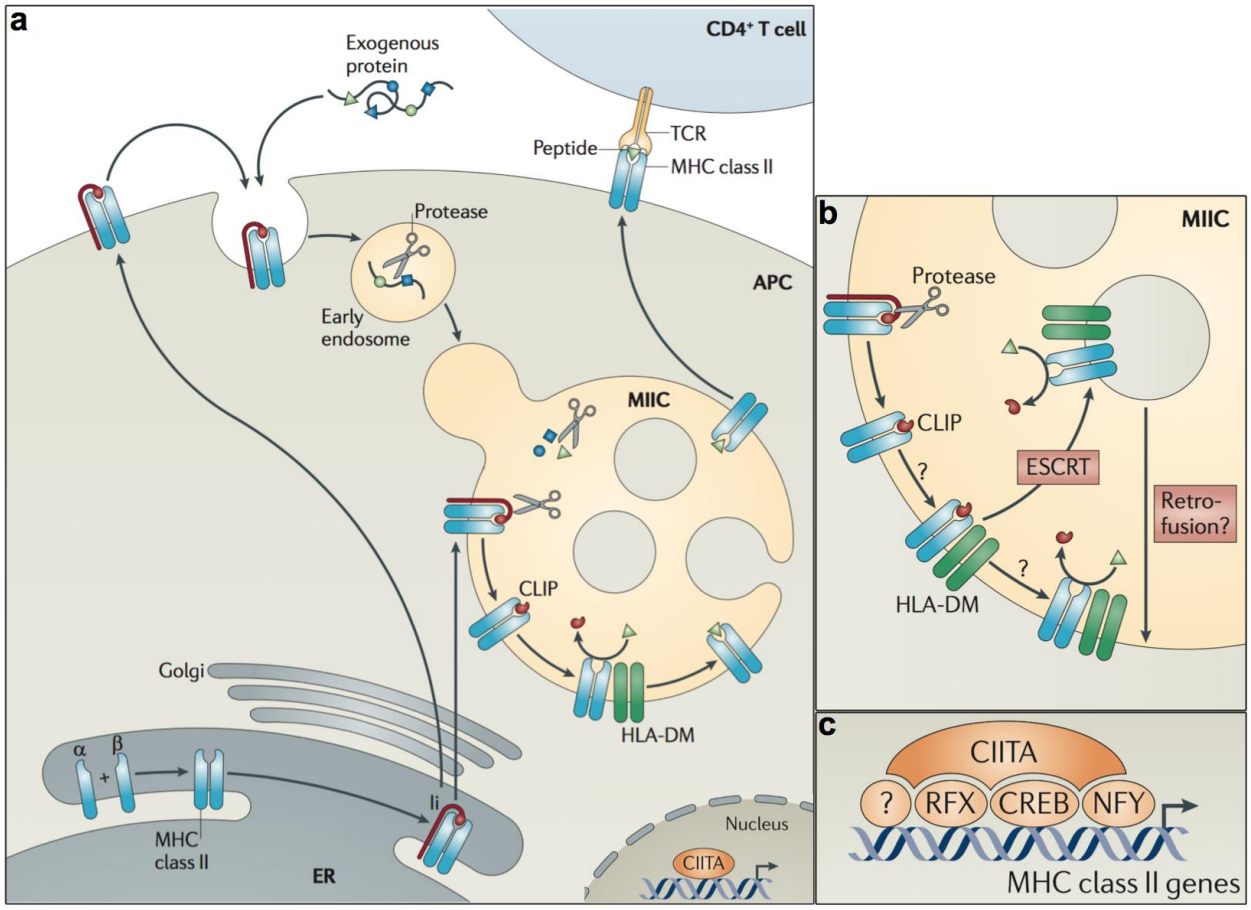

Figure 1.5: Antigen-presentation via the MHC class II complex. (a) The α- and β-chains of the MHC class II complex are assembled in the endoplasmic reticulum (ER) and are transported to the endosome, where, following the removal of CLIP, peptide loading occurs (b). Transcriptional regulation of MHC II is mediated by a multi-protein complex, orchestrated by CIITA, the master transcriptional activator (c). Abbreviations: APC, antigen presenting cell; CLIP, class II-associated Ii peptide; CREB, cAMP-responsive-element-binding protein; CIITA, class II transactivator; ER, endoplasmic reticulum; ESCRT, endosomal sorting complex required for transport; MHC, major histocompatibility complex; MIIC, MHC class II compartment; NFY, nuclear transcription factor Y, RFX, regulatory factor X; TCR, T cell receptor. Adapted from [122] with permission.

1.3.3 Transcriptional regulation of MHC class II

Our understanding of the detailed molecular mechanisms involved in antigen presentation via MHC class II and the transcriptional regulation of the complex itself have been strengthened by studies elucidating the genetic foundation of immunodeficiency [182,183].

The transcription of the MHC class II locus is tightly controlled by a multi-protein complex, involving regulatory factor X (RFX), cyclic AMP-responsive-element-binding protein (CREB), nuclear transcription factor Y (NF-Y) and CIITA, the latter considered to

be the master transcriptional regulator of MHC class II expression [184–186]. Of importance for controlling expression of MHC class II genes is a specific regulatory sequence, located approximately 150-300 bp upstream of the transcription start site (TSS), called SXY module [175,180,183]. This module encompasses four different boxes: W/S (essential for recruitment of CIITA to the promoter region) [187], as well as X1, X2 and Y, which are the binding sites for the heterotrimer RFX, CREB and NF-Y, respectively. Upon binding of these three proteins or protein complexes to the SXY module, the entire structure is termed MHC class II enhanceosome, to which CIITA is ultimately recruited [188]. Similar WXY motifs have been found in the promoter regions of the invariant chain (CD74) and MHC class I, suggesting an important role in regulating antigen presentation pathways. Since CIITA itself is not capable of DNA-binding, the interaction with the aforementioned proteins is indispensable.

CIITA and its pathogenic relevance were first discovered and described in a rare but severe immune disorder, named hereditary, autosomal recessive MHC class II deficiency or type II bare lymphocyte syndrome (BLS) [183]. Subsequent studies have further delineated the underlying molecular defects in specific *trans*-acting factors, based on which patients with BLS can be subdivided in the following four complementation groups: BLS group A patients harbour loss of function mutations within *CIITA* [189], whereas patients in groups B, C and D are characterized by genetic defects of the genes defining the RFX complex, which consists of *RFXANK*, *RFX5* and *RFXAP* [190–193]. Characteristic for the RFX-deficient BLS groups is that the promoters of MHC class II complex genes are unoccupied, highlighting the importance of RFX for the assembly of the MHC class II enhanceosome [180,194]. Mutations in the aforementioned genes often result in a lack of MHC class II protein, and because MHC class II expression is mainly restricted to APCs and is essential for the display of foreign antigens to $CD4^+$ T cells, BLS patients often suffer from multiple infections as a result of an insufficient adaptive immune response.

Cell-type specific expression of CIITA is ensured by usage of different promoter structures [175,180,195]. Four promoters have been described in the literature [196] and transcripts deriving from these promoters differ in their first exon. Promoter I (pI) is used by dendritic cells, promoter III (pIII) is restricted to B cells, whereas CIITA

expression from the promoter IV (pIV) can be induced upon interferon (IFN) γ stimulation in a wide variety of different cell types without antigen presenting function, such as fibroblasts, endothelial cells and epithelium [197–200]. Significant amounts of promoter II (pII)-derived transcripts could not be detected in human tissues or cell lines, hence the biological relevance of this promoter is considered to be rather limited [196].

The CIITA protein consists of three important domains: an amino-terminal acidic domain, including a proline, serine, threonine (PST)-rich segment, a centrally located conserved NACHT-domain, also known as GTP-binding domain, and leucine-rich repeats (LRR) at the carboxyl-terminus [174,175,180]. The NACHT-domain is named after the genes in which it has been found: *Naip, CIITA, HET-E,* and *TP-1* [201]. The N-terminal region is thought to mediate interactions of CIITA with effector proteins, such as transcription factors, chromatin remodeling factors and co-activators [202,203], whereas the GTP-binding domain is important for conformation of the CIITA protein, self-association and localization to the nucleus [204,205]. Similarly, disruption of the LRR region abrogates CIITA activity and results in cytoplasmic accumulation of the protein [206–209].

1.4 Thesis theme and objectives

Malignant lymphomas account for approximately 5 % of all newly diagnosed cancer cases per year and affect patients of all ages [48]. Despite improvements in clinical management of lymphoma patients, a significant proportion still experience relapse or progression and eventually succumb to their disease.

Over the past decade, numerous studies have carved out the importance of the microenvironment for cancer development as well as perpetuation [210,211]. Moreover, there is increasing evidence that tumour cells are orchestrating reactive cells in a targeted manner to escape immune surveillance [111,211]. Several mechanisms by which lymphoma cells can evade host immune response have been described. Expression of surface molecules on malignant cells including overexpression of Galectin-1 and ligands of programmed death receptor 1 (PDL) [212–214] are reported to establish peripheral tolerance in several lymphoma subtypes. Moreover, tumour cells in PMBCL and cHL sometimes lack MHC class II expression, a finding that has been

linked to reduced immunogenicity of tumour cells [133,134,215]. Based on these data, one can reasonably argue that immune escape represents an important pathogenic mechanism in lymphoid cancers. Importantly, although the phenotypic changes related to immune escape are increasingly recognized, most of the underlying genetic mechanisms still need to be uncovered.

In view of our discovery of recurrent structural changes involving the *CIITA* gene locus in a large proportion of PMBCL and cHL cases [172], we hypothesize that genomic alterations of *CIITA* are prevalent in certain lymphoma subtypes and underlie an important mechanism of acquired immune privilege in a subgroup of lymphoid cancers. CIITA functions as the master transcriptional regulator of MHC class II molecules, which are commonly expressed on APCs and are crucial for the interaction with other cellular components of the immune system. In the setting of malignant lymphomas, somatic genetic alterations of *CIITA* could lead to a reduction of MHC class II expression, therefore creating an `immune privilege` phenotype in the malignant population and contributing to disease progression and impaired survival of lymphoma patients. The main objective of this research is to obtain a better understanding of the heterogeneity, molecular mechanisms, as well as the functional consequences, and ultimately the clinical implications, of *CIITA* gene alterations. These insights will be critical to the translation of our discoveries into clinically meaningful progress for affected patients and pave the way for targeted therapies.

1.5 Hypothesis

Genetic alterations of *CIITA* are frequent across a spectrum of B cell lymphomas, result in reduction of MHC class II expression and altered microenvironment composition, thereby conferring an ‘immune privilege’ phenotype and contributing to disease progression and impaired survival of lymphoma patients.

1.6 Research aims and thesis outline

I seek to comprehensively explore mutational patterns of *CIITA* in major B cell lymphoma subtypes (i.e. DLBCL, FL and PMBCL) and to uncover the repertoire of structural rearrangement partner genes. Secondly, I will correlate mutational findings

with pathologic and clinical characteristics and perform *in vitro* functional studies to elucidate consequences of genetic alterations in *CIITA*.

The thesis consists of an introductory part, two chapters describing original research and a discussion with concluding remarks and an outline of ongoing and future research.

The structure of this thesis follows three specific aims:

Aim 1: To characterize genomic alterations in *CIITA* across major subtypes of B cell lymphomas.

Aim 2: To determine the spectrum of novel rearrangements and gene fusion partners of *CIITA* and describe breakpoint anatomy at base pair resolution.

Aim 3: To explore gene expression changes in monoculture and evaluate the functional and clinical impact of *CIITA* alterations in primary lymphoma cases.

Chapter 2: Genomic alterations of *CIITA* in malignant B cell lymphomas

2.1 Introduction

Over the past decade, a lot of emphasis has been put on the refinement of cancer pathogenesis models to acknowledge the contribution of the immune system to disease development and progression. The revised version of the 'hallmarks of cancer' [109] prominently recognizes the importance of the cellular crosstalk of immune cells with the malignant cell population as a significant contributor to cancerogenesis ("tumour promoting inflammation") but, at the same time, introduces the concept of "immune evasion" as one of the emerging hallmarks of cancer [110].

The evolution of NGS techniques had, and still continues to have, a profound impact on the field of cancer genomics, allowing for genome- and transcriptome-wide characterization of structural genomic alterations. By applying these novel techniques, we previously identified and described rearrangements involving *CIITA, CD274* and *PDCD1LG2* as common mutational events in certain lymphoma subtypes, namely PMBCL and cHL, as well as DLBCL cases arising in immune-privileged sites of the body [172,216,217]. Despite the fact that these lymphoma entities differ in their expression of certain B cell lineage markers and transcription factors, the expression of surface MHC class II molecules seems to be defective in a considerable fraction of cases (especially in PMBCL and cHL) - a finding that has been linked to reduced immunogenicity of tumour cells and was shown to correlate with inferior survival [133,134].

The recurrent structural genomic alterations (specifically, unbalanced chromosomal rearrangements) of *CIITA*, the master regulator of MHC class II transcription, were thought to be likely causative of MHC class II loss, however a detailed analysis of MHC class II protein expression and correlations with microenvironment composition and patient outcomes have not been performed. Although these chromosomal aberrations of *CIITA* seem to be critical pathogenic drivers, facilitating immune escape of the malignant B cell population, the full spectrum of genetic changes and related functional phenotypes in lymphomas still remain largely

unexplored. It follows, that the underlying causes of reduced MHC class II expression are only partially understood.

CIITA and its pathogenic relevance were first described in a rare but severe immune disorder, namely hereditary MHC class II deficiency or type II BLS [189]. Patients affected by this heritable disease often harbor loss of function mutations within *CIITA,* resulting in a lack of MHC class II protein expression. Since this is indispensable for the display of foreign antigens to $CD4^+$ T cells, BLS patients often suffer from multiple infections resulting from an insufficient adaptive immune response. CIITA expression is regulated by usage of different promoters in a cell type specific manner. Four promoters have been described in the literature [196] and transcripts deriving from these promoters differ in their first exon. Promoter III driven transcription of *CIITA* is the major source of CIITA expression in B cells, whereas expression from the interferon γ-inducible promoter IV only plays a minor role in this cellular context.

Here, we aimed to describe the full spectrum of genomic alterations of *CIITA* in PMBCL, DLBCL and FL, including chromosomal rearrangements and coding sequence mutations. Furthermore, by applying targeted capture sequencing approaches we wanted to identify and characterize rearrangement partner genes as well as the biological implications of these structural genomic variants.

2.2 Materials and methods

2.2.1 Cell lines and patient cohorts

2.2.1.1 Cell lines

The PMBCL-derived cell lines U2940, MedB-1 and Karpas1106P, as well as the nodular lymphocyte-predominant Hodgkin lymphoma (NLPHL)-derived cell line DEV, were previously selected for and characterized by whole transcriptome paired-end sequencing (RNA-Seq) [104,216]. Karpas1106P was provided by Dr. M. Dyer (University of Leicester, UK) and grown in RPMI-1640 media (Life Technologies) with 20 % heat-inactivated fetal bovine serum (FBS, Life Technologies). The cell line U2940 was obtained from Dr. L. Staudt (National Institute of Health, Bethesda, MD) and propagated in RPMI with 10 % heat-inactivated FBS. MedB-1 was a kind gift from Drs. S. Brüderlein and P. Möller (Department of Pathology, University of Ulm, Germany) and

propagated as published [218]. The cell line DEV was generously provided by Dr. A. Diepstra (University Medical Center Groningen, The Netherlands) and grown in RPMI with 20 % heat-inactivated FBS. The lymphoblastoid cell line (LCL) was derived from Epstein-Barr virus (EBV)-transformed B cells of a healthy male donor and provided by Dr. U. Steidl (Albert Einstein College of Medicine, New York, NY). These cells were grown in RPMI supplemented with 20 % heat-inactivated FBS, 1 % penicillin/streptomycin, 1 % non-essential amino acids and 1 % sodium pyruvate (all Life Technologies).

2.2.1.2 Patient cohorts

2.2.1.2.1 PMBCL cohort

We identified and selected diagnostic pre-treatment specimens from 45 PMBCL patients from our database based on the availability of fresh-frozen (FF) tissue biopsies or cell suspensions. Those were then retrieved from the tissue archives of the Centre for Lymphoid Cancer (CLC) at the British Columbia Cancer Agency (BCCA). Of these 45 cases, seven PMBCL specimens with available matched constitutional DNA were previously used for whole-genome sequencing (WGS) or RNA-Seq analyses [104]. Formalin-fixed, paraffin-embedded (FFPE) tissue specimens from 148 PMBCL cases were assembled on tissue microarrays (TMA), as described previously [104,172].

2.2.1.2.2 DLBCL cohort[3]

Next, we explored *CIITA* mutations in a cohort consisting of tissue biopsies stemming from 347 patients diagnosed with *de novo* DLBCL. All of these patients were uniformly treated with rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone (R-CHOP) and FF tissue or cell suspensions were available for targeted re-sequencing analyses of the protein coding region of 58 genes, known to be important for DLBCL pathogenesis. All tissue biopsies were reviewed by experienced haematopathologists at the BCCA and in addition, clinical parameter and outcome data, as well as COO, determined by the Lymph2Cx assay [65], were available.

[3] This cohort was primarily assembled by Drs. Daisuke Ennishi, David Scott and Randy Gascoyne as part of the TFRI program project grant, subproject 2.

2.2.1.2.3 FL cohort[4]

This cohort represents a spectrum of FL patients with different courses of their disease: 1) patients initially diagnosed with FL who subsequently experienced transformation (tFL) to a high-grade lymphoma (in most cases of the DLBCL subtype); 2) patients with progressive disease (pFL) within 2.5 years after treatment initiation, but without histological evidence of transformed FL; and 3) patients who had no clinical signs of progression (npFL) for at least 5 years. The final cohort used for capture sequencing (see section 2.2.5.3) consisted of 277 FL patients, of which 159 belonged to the transformed group (118 of those had biopsies available from timepoint T1 (per definition, the primary FL diagnosis) and T2 (timepoint of transformation, usually presenting as DLBCL), for 10 patients only the T1 biopsy was available, and 31 patients were represented solely by their T2 biopsy). Forty-one specimens from patients with pFL and 84 biopsies from patients with npFL completed the cohort.

All biopsies were centrally reviewed by expert haematopathologists at the BCCA and Lymph2Cx assay scores were available for transformed lymphomas with classical DLBCL morphology [219].

2.2.2 Cell sorting

Frozen vials of cells obtained from reactive tonsillar specimens were thawed and stained with purified mouse anti-human CD77 antibody (Clone 5B5, BD Pharmingen) for 30 minutes on ice. Cells were then incubated for 15 minutes at 4ºC with rat anti-mouse IgM MicroBeads (Miltenyi) and subsequently separated using the autoMACS Pro cell separator (Miltenyi) in order to enrich for germinal center B cells.

2.2.3 Fluorescence *in situ* hybridization (FISH)

FISH break-apart (ba) assays were performed on TMAs, containing duplicate cores (0.6 or 1.0 mm in diameter) from each specimen. Briefly, TMA sections were stained with green and red labelled bacterial artificial chromosome (BAC) probes

[4] This cohort was primarily assembled by Drs. Robert Kridel and Randy Gascoyne as part of the TFRI program project grant, subproject 1.

flanking the genomic region containing the *CIITA* gene (Figure 2.1,Table 2.1.). DAPI (4,6-Diamidino-2-phenylindole) was used for counterstaining of the cell nuclei.

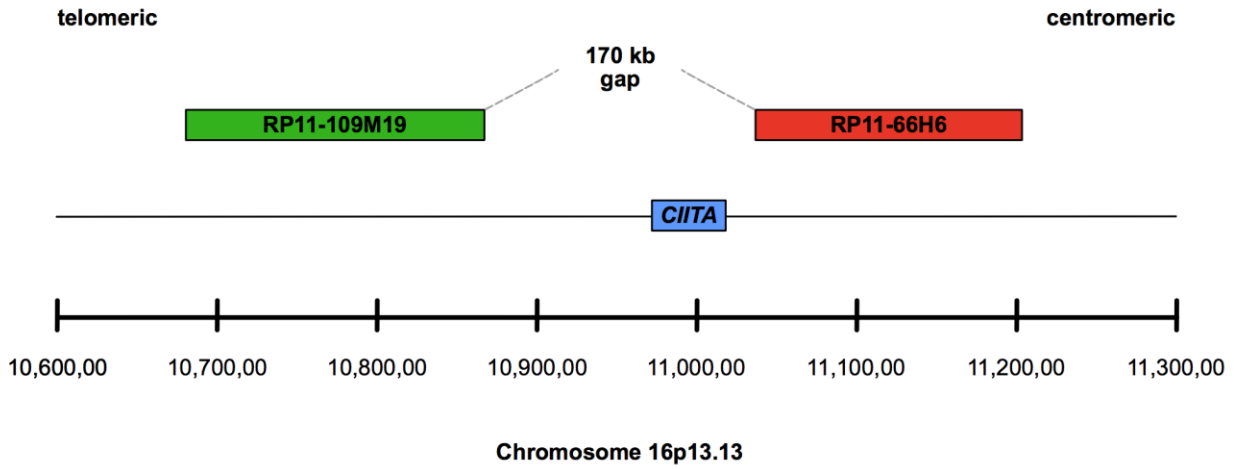

Figure 2.1: Schematic of the *CIITA* FISH assay. Shown at the bottom are the genomic coordinates of the hg19 chr16p13.13 locus, which contains the *CIITA* gene (other genes in this region are not depicted here for visualization purposes). The BACs flanking this region are labelled with either a green or a red fluorescent probe, corresponding to their respective boxes in this picture.

Table 2.1: Specification of the BACs used for assessment of the *CIITA* locus by FISH

BAC/gene name	Genomic coordinates	
	start	end
CIITA	10,971,055	11,018,840
RP11-109M19	10,680,358	10,866,730
RP11-66H6	11,036,514	11,203,598

Genomic coordinates are given according to GRCh37/hg19. BAC: bacterial artificial chromosome.

After hybridization, slides were imaged using a Zeiss AXIO Imager.Z2 microscope, equipped with a Metafer imaging system (MetaSystems). One hundred interphase nuclei were assessed for each specimen and for all cores with examinable patterns, the rearrangement status was determined. Cases were classified as ba-positive if > 5 % of nuclei had split signals or signal patterns indicative of unbalanced rearrangements (e.g. loss of one signal). For CNV, the following cut-offs were applied: 1) loss: > 40 % of nuclei with one fusion signal; 2) gain: > 20 % nuclei with three or four fusion signals; and 3) amplification: > 20 % of nuclei with five or more fusion signals.

2.2.4 Copy number analysis

Affymetrix Human SNP 6.0 microarrays were used to profile the genome-wide copy number (CN) status of the cell lines Karpas1106P, U2940, MedB-1 and DEV following standard protocols according to the manufacturer's instructions. After quality control, SNP 6.0 microarrays were pre-processed using the PennCNV-Affy protocol [220]. OncoSNP (v1.3) was then used to simultaneously segment and predict CN [221]. The CN states of these segments were subsequently projected onto genes using Ensembl (version 72) as the scaffold.

2.2.5 Sequencing of *CIITA* coding sequence and intron 1

Sequencing for the three lymphoma cohorts described in section 2.2.1.2 was performed as part of different subprojects within the TFRI team grant at different timepoints, hence various sequencing methods were applied. Therefore, this section will be subdivided and details will be provided for the different approaches with respect to the corresponding cohort.

2.2.5.1 Sequencing of the PMBCL cohort

Genomic DNA and RNA were extracted using the AllPrep DNA/RNA FFPE kit (Qiagen). *CIITA* coding sequence (CDS) mutations in primary PMBCL tumour specimens were detected by re-analysis of the previously generated WGS and/or RNA-Seq libraries (n = 7) [104]. In 18 cases, Sanger sequencing was performed for all coding exons and 30 cases were analyzed by deep amplicon sequencing (TruSeq Custom Amplicon assay (TSCA), Illumina). One case could not be completely characterized due to insufficient amounts of nucleic acids. For details, see A.1.

TSCA was performed according to the manufacturer's instructions. Briefly, 250 ng genomic DNA was amplified with oligonucleotides for the complete protein-coding sequence of *CIITA* and the alternative exon 1 that is driven by the pIV promoter (details of the oligonucleotide sequences can be found in A.2). Amplicon libraries were sequenced on an Illumina MiSeq instrument using the V2 300-cycles MiSeq reagent kit (Illumina), generating 150 bp paired-end reads. Sequencing reads were aligned to hg19

using Bowtie (v2.1) [222]. VarScan (v2.3.6) was then run using the tumour-normal comparison mode “somatic” for paired tumour-normal specimens. For single tumour samples, VarScan was run using mpileup2snp. Mutational data was then annotated using SnpEff (v3.2) with respect to the GRCh37.70 database. The following filters were applied for the paired tumour-normal sample data following VarScan: 1) filtering of SNPs (dbSNP137); 2) filter mutations with SnpEffs DOWNSTREAM, UPSTREAM, INTERGENIC, and INTRAGENIC; 3) somatic status (SS) == 2; 4) tumour variant read ratio $\geq$ 0.1; 5) normal variant read ratio < 0.01; and 6) number of tumour reads $\geq$ 10. The following filters were applied for the single tumour sample data following VarScan: 1) filtering of SNPs (dbSNP137); 2) filter mutations with SnpEffs DOWNSTREAM, UPSTREAM, INTERGENIC, and INTRAGENIC; 3) tumour variant read ratio $\geq$ 0.05, and 4) number of tumour reads $\geq$ 10.

All mutation predictions resulting from WGS, RNA-Seq and deep amplicon sequencing were subjected to independent validation using Sanger Sequencing (3130 Genetic Analyzer, Applied Biosystems or outsourced to GENEWIZ Inc.). For the remaining specimens, which were not analyzed by NGS techniques, we amplified all *CIITA* coding exons by standard PCR prior to Sanger sequencing. Deletions in *CIITA* intron 1 were detected by amplification of a fragment covering the breakpoint cluster region [172] and aberrant length fragments (< 3 kb) were cloned into the pCR 2.1-TOPO vector (TOPO TA cloning kit, Invitrogen) and fully sequenced.

Primer sets are listed in A.3. Sequence analysis and assembly were performed using Clone Manager software (Scientific & Educational Software). Genomic coordinates are given according to GRCh37/hg19.

2.2.5.2 Sequencing of the DLBCL cohort[5]

Deep amplicon sequencing of the DLBCL cohort was performed on the Illumina TSCA platform. The original panel consisted of 58 genes, which included recurrently mutated genes as described in previous studies [60,61,223], as well as genes of

[5] The study exploring the mutational landscape in DLBCL is spearheaded by Drs. Daisuke Ennishi and David W. Scott. Bioinformatics analyses were performed by Dr. Christoffer Hother, who, together with Dr. Ennishi, provided the initial draft for this methods section. In this thesis, only data on *CIITA* mutations were utilized for further exploration and analyses by myself.

biological interest. Oligo probes mapping to the target regions, which consisted of the entire protein-coding sequence of these genes, were designed using Illumina DesignStudio (Illumina); the coordinates for oligos capturing the *CIITA* gene locus are listed in A.4. Amplicon libraries were constructed according to the manufacturer's protocol (Illumina) for end-repair, A-tailing and adaptor ligation, and 250 ng DNA from FF specimens served as input for the hybridization of the custom-designed oligo pool. Agencourt AMPure XP beads (Beckman Coulter) were used for PCR clean-up. Libraries were normalized according to the manufacturer's instructions and were individually indexed using dual indexing before pooling. Purified pooled libraries were quantified on a real-time PCR instrument using the KAPA Illumina Library Quantification kit (KAPA Biosciences, KK4835) and quality control was performed with the Agilent Bioanalyzer (Agilent DNA 1000 Kit). Library pools were subsequently diluted into HT1 buffer prior to sequencing on the Illumina MiSeq instrument (Illumina), with MiSeq V2 cycle reagent kits and paired-end 2x150 bp reads. Sequencing data were converted and exported as FASTQ files using CASAVA (v1.8.2). Alignment was performed with bwa version 0.7.5a, and Mutascope (v1.02) was used to predict variants. Synonymous mutations and variants outside of the exonic space (e.g. intron, 5' and 3' untranslated region (UTR)) were excluded from further analysis. SNPs were filtered out if they appeared in the NCBI dbSNP137 or 1000 Genomes Project (v3) databases. Variants were not discarded, independent of their presence in other filters, if they were contained within the Catalogue of Somatic Mutations in Cancer (COSMIC) database (v62). Variant annotation and effect predictions were performed using SnpEff (v3.6).

2.2.5.3 Sequencing of the FL cohort[6]

Eighty-six genes were primarily selected for capture-based targeted sequencing of the FL cohort. Those genes were chosen based on three criteria: 1) recurrently mutated in FL with a frequency of > 5 % [90,93]; 2) recurrently mutated in DLBCL > 5 % (our own data) or 3) consistently mutated in BL [168,224,225]. Furthermore, 20 genes,

[6] The study exploring clonal evolution in FL is spearheaded by Drs. Robert Kridel and Fong Chun Chan and has been recently published [96]. Bioinformatics analyses were performed by Fong Chun Chan. For this thesis, only data on *CIITA* mutations obtained from capture-based targeted sequencing were utilized for further exploration and analyses by myself.

including *CIITA,* were selected to assess mutations in known targets of SHM. Libraries were constructed using either 500 ng of genomic DNA extracted from FF or 200 ng of FFPE tissue-derived genomic DNA as an input. Custom SureSelect XT2 baits (Agilent; total capture space 452,129 bp) were used for capturing. Libraries were subsequently pooled (maximum of 46 libraries per pool) and sequenced at Canada's Michael Smith Genome Sciences Centre on one Illumina HiSeq lane per pool, generating 125 bp indexed reads (V4 chemistry). Alignment of sequencing reads to the GRCh37 genome was performed with bwa version 0.7.5a and cases with low mean target coverage (i.e. < 50x) were excluded from subsequent analyses. For single nucleotide variant (SNV) calling, MutationSeq (v4.3.8) [226] was used and the following criteria were applied: mutation probability > 0.8 and coverage $\geq$ 100 or mutation probability > 0.9 and coverage < 100 for FF specimens; for FFPE specimens the requirements were relaxed to meet either a mutation probability > 0.7 and coverage $\geq$ 50 or mutation probability > 0.9 and coverage < 50. For cases with available germline DNA (n = 80), SNVs were filtered out if they were present in the germline specimen. For all other cases, putative SNPs were identified and filtered out if they were present in at least two unrelated germline specimens from the 80 cases mentioned above or in dbSNP version 137. SNVs were annotated with SnpEff. For cases with matching germline specimens, indels were called using Strelka (v1.0.13) and for those without matching germline, indels were called using VarScan and putative germline variants were filtered out using dbSNP (v137) and the 1000 Genomes Project (v3) database.

2.2.6 Analysis of AID target motifs

Potential AID target motifs (WRCY/RGYW) within the first 2 kb downstream of the TSS were identified using the Possum tool (http://zlab.bu.edu/~mfrith/possum/). Cytosine residues within (third position of the 4-base motif) and outside (including cytosine at position 4) of potential AID target motifs were then enumerated with the ratio (within vs outside) of these numbers representing the expected frequency of a cytosine within the motif being affected by mutation. Next, we determined how many of the observed mutations at cytosine residues occurred in WRCY/RGYW motifs and compared this to the number of affected cytosines outside these hotspots.

2.2.7 Immunohistochemistry

Immunohistochemistry (IHC) was performed on 4 μm sections from FFPE tissue specimens arranged on previously constructed TMAs [104,172] or on whole tissue sections in the case where a particular sample was not evaluable or represented on the TMA. Following antigen retrieval, sections were stained with a primary antibody recognizing AID (dilution 1:75, clone 1A9-1, Cell Signaling, #4959) followed by routine protocols for automated IHC on the Ventana Benchmark XT (Ventana Medical Systems). The percentages of positive tumour cells were recorded in 10 % increments.

2.2.8 Characterization of chromosomal rearrangements

Targeted sequencing approaches are ideally suited for research questions focusing on a circumscribed genomic region. Therefore, this technology has become increasingly popular in cancer research since many malignancies are characterized by recurrent mutations and rearrangements in a relatively small number of genomic loci. For our comprehensive assessment of structural variants (SV), we explored two different approaches which are further explained in the sections below.

2.2.8.1 Bacterial artificial chromosome (BAC) capture[7]

To explore novel structural genomic rearrangement partners and breakpoints in primary testicular lymphoma (PTL), a DLBCL arising primarily in the testis, three pre-treatment biopsies, one of which was previously shown to be 9p24.1 rearranged by FISH, were selected for targeted BAC capture high-throughput sequencing [216]. One extranodal DLBCL, known to harbour a *TBL1XR1-TP63* inversion and a *BCL2* rearrangement, was sequenced as a positive control [69]. DNA was extracted from these archival FFPE biopsies using the Qiagen DNA/RNA FFPE Kit as per the manufacturer's protocol and 2 μg were used for library construction. To identify gene fusions and structural genomic alterations, we re-sequenced 7.45 Mb of the human genome with a focus on 12 loci known to be recurrently affected by rearrangements in B-cell lymphomas [59,69,172,216]: *CIITA*, *CD274* (PDL1), *PDCD1LG2* (PDL2), *TP63*, *TBL1XR1*, *BCL2*, *BCL6*, *IGH*, *IGL*, *IGK*, *MYC* and *JAK2.* Custom capture was

[7] A version of this paragraph is included in [217] and is used with permission.

performed in four FFPE specimens using biotinylated RNA capture probes derived from BAC clones. Briefly, 72 BAC clones from the RPCI-11 library were identified as spanning the 12 gene loci from the reference human genome assembly (GRCh37/hg19) displayed in the UCSC Genome Browser. BAC clone DNA was extracted and the identities confirmed by BAC-end Sanger sequencing using T7 and SP6 primers and DNA fingerprinting. Validated clone DNAs were pooled, sonicated, and products in the 75-200 bp range were excised from an 8 % PAGE gel and quantified. Two hundred ng of the size-selected pooled DNA was end-repaired and phosphorylated for A-tailing and adapter ligation. Adapter-ligated products were enriched using eight cycles of PCR. Gel-purified DNA was *in vitro* transcribed using T7 RNA Polymerase. Libraries were captured using a 400:1 ratio of BAC-derived probes to library DNA. Products subsequently underwent 17 cycles of PCR prior to sequencing which was performed on an Illumina HiSeq2000 generating 75 bp paired-end reads. Sequencing data was analysed using the nFuse and deStruct algorithms to generate and cross-corroborate rearrangement predictions, as has been previously published [216,227]. These predictions were subsequently filtered on the basis of the validation probability metric (> 0.75); a score of uniqueness among libraries, quality and number of spanning and split read support, and alignment relative to the capture space. Subsequently, these events were BLAT-assessed and a select number were validated by Sanger sequencing.

2.2.8.2 Capture sequencing[8]

As an alternative to BAC capture sequencing, we used a custom Agilent SureSelect design to enrich for specific genomic regions. Genomic coordinates were chosen based on our genes of interest and all previously described rearrangement breakpoints were taken into consideration. Two regions were located on chromosome 16: one spanning the *CIITA* gene and the other located downstream, spanning the *SOCS1* gene. *SOCS1* is a known tumour suppressor in cHL and PMBCL, which regulates JAK/STAT signaling and is frequently mutated in those lymphoma entities [228]. The third region spanned the adjacent *CD274* and *PDCD1LG2* genes on chromosome 9 and was designed to include all previously described PDL translocation

[8] This section is in part published in [214] and is used with permission.

breakpoints. Table 2.2 displays the exact target coordinates used in the capture design, spanning approximately 0.5 Mb of genomic space in total.

Table 2.2: Custom Agilent SureSelect design used for capture sequencing.

Gene	Chromosome	Captured region		Capture space (bp)
		start	end	
CD274 and *PDCD1LG2*	9	5,449,434	5,573,579	124,146
CIITA	16	10,959,585	11,277,301	317,717
SOCS1	16	11,332,082	11,350,098	18,017

Genomic coordinates are given according to GRCh37/hg19.

The Agilent SureSelect custom kit consisted of labeled RNA probes complementary to the target space (Table 2.2). Probes were designed using 5x tiling and standard boosting. Repetitive regions were not masked to increase the likelihood to identify breakpoints located in repetitive regions, even though this also increased the off-target capture rate.

Cases were selected based on results from previous screening studies using FISH to interrogate the *CIITA, TBL1XR1* and PDL loci. FISH assays were performed as described in section 2.2.3. Specimens were selected if they had a ba-positive status at any of these loci. In addition, five samples were included because they showed over-expression of one or both PDLs by qRT-PCR, but were negative for structural rearrangements or CNV by FISH. Genomic DNA and RNA were extracted from FFPE tissue scrolls using the Qiagen AllPrep DNA/RNA FFPE kit according to the manufacturer's instructions.

Library construction was performed at Canada's Michael Smith Genome Sciences Centre using a 96-well plate protocol suitable for FFPE. All DNA samples were normalized to contain 100 ng in a total volume of 62 μl elution buffer (Qiagen). DNA was subsequently transferred into a microTUBE plate for shearing using an LE220 acoustic sonicator (Covaris) and the following conditions: duty factor = 20 %, peak incident power = 450 W, cycle per burst = 200, duration = 2x 60 seconds with an intervening spin. The expected fragment size for this protocol is 300-400 bp. In order to further improve the quality of the library, a solid phase reversible immobilization (SPRI)

bead-based size selection was performed to remove smaller DNA fragments which may have resulted from fragmentation processes occurring during formalin fixation. DNA damage and end-repair, as well as phosphorylation were performed in a single reaction using an enzyme premix (NEB). Repaired DNA fragments were A-tailed for ligation to paired-end, partial Illumina sequencing adapters and then purified twice with SPRI beads. Eight cycles of PCR were performed to introduce fault-tolerant hexamer "barcodes" to allow multiplexing of the libraries for sequencing. Final library concentration was measured using a high sensitivity Caliper LabChip GX in conjunction with Quant-iT (Invitrogen). Libraries were pooled prior to capture. Sequencing was performed in two batches, the first batch consisted of 16 libraries and was sequenced in a single lane on an Illumina HiSeq 2500 using version 3 chemistry, producing paired-end 100 bp reads. The second batch of the remaining 52 libraries was pooled into two groups of 17 and one group of 18. Sequencing was performed using version 4 chemistry, resulting in paired-end 125 bp reads. Following quality control measures, including filtering and removal of duplicate reads, FASTQ files were aligned to the GRCh37 genome. Quality control and bioinformatics analysis including SV detection were performed by Lauren C. Chong and are described in detail in her Master's thesis (https://dx.doi.org/10.14288/1.0229569).

To validate predicted structural rearrangement events, custom primer sets were designed for a targeted PCR amplification of the respective genomic region (A.3). Primer design was performed using the Primer3 software (version 0.4.0; http://bioinfo.ut.ee/primer3-0.4.0/primer3/). Successfully amplified PCR products were subsequently Sanger sequenced.

2.2.9 Statistical analysis

The Fisher's exact test and χ^2 test were used to compare categorical variables. McNemar's test was used when comparing categorical variables between paired specimens. *P*-values < .05 were considered being significant.

2.3 Results

2.3.1 *CIITA* alterations in PMBCL

2.3.1.1 Biallelic genomic alterations of *CIITA* in PMBCL- and NLPHL-derived cell lines

We have previously demonstrated that genomic rearrangements of *CIITA* occur in various lymphoid malignancies and are highly recurrent in cHL and PMBCL [172]. In order to inform on the prevalence and different types of genomic aberrations affecting *CIITA* in PMBCL, we first re-analyzed previously published whole transcriptome sequencing (RNA-Seq) data of three PMBCL-derived cell lines: Karpas1106P, MedB-1 and U2940 [104]. Since we have reported on a *CIITA-PDCD1LG2* fusion in DEV, an NLPHL-derived cell line [216], this was also included in the analysis. In addition, we performed high-resolution CN analysis (Affymetrix SNP 6.0 arrays), and validated identified SNVs and rearrangements by Sanger sequencing.

No *CIITA* copy number or structural alterations were found in MedB-1, but we identified two SNVs, one (hg19 chr16: g.10989537G>A) resulting in an amino acid exchange (p.71Asp>Asn) in the acidic domain and the other (hg19 chr16: g.35202C>T) altering the amino acid sequence (p.636Thr>Met) in the conserved NACHT-domain of the CIITA protein. Analysis of individual cDNA clones showed that these point mutations occur in *trans* configuration (Figure 2.2).

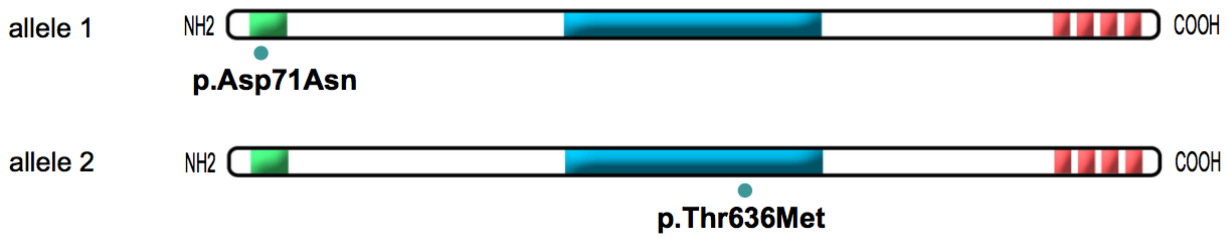

Figure 2.2: *CIITA* genetic alterations in the PMBCL-derived cell line MedB-1. MedB-1 is characterized by two missense mutations occurring in *trans* and leading to amino acid exchanges in functionally relevant protein domains (green: acidic domain; blue: NACHT domain; red: leucine rich repeat (LRR) domain). All genomic coordinates are given according to the hg19 reference genome.

It has been previously described that the tumour suppressor *SOCS1* is inactivated in the Karpas1106P cell line due to a biallelic deletion in this chromosomal region [228,229]. Interestingly, *CIITA* is located just 0.5 Mb telomeric of *SOCS1*. Analysis of our RNA-Seq data revealed a lack of CIITA mRNA expression in this cell line but detected a TXNDC11-EMP2 fusion transcript (exon 3 of *TXNDC11* spliced in frame to exon 2 of *EMP2;* Figure 2.3), which is the result of a genomic deletion encompassing 11 genes including *CIITA, SOCS1* and *CLEC16A,* the latter recently shown to be involved in normal B cell development [230]. The precise coordinates of the deletion were determined as hg19 chr16: g.10641751_11825721del. The SNP array data identified a region in which to perform a "PCR walk" to inform on the breakpoints of the deletion encompassing the 5' region of the other *CIITA* allele, which were in intron 1 of *TVP23A* and intron 1 of *CIITA* (hg19 chr16: g.10899052_10985179del), resulting in the loss of pIII and the associated exon 1 (Figure 2.3). As would be expected, because *TVP23A* and *CIITA* appear in opposite orientations on the chromosome, and the promoter for *TVP23A* is also deleted, no transcript from this fusion was detected in the RNA-Seq data.

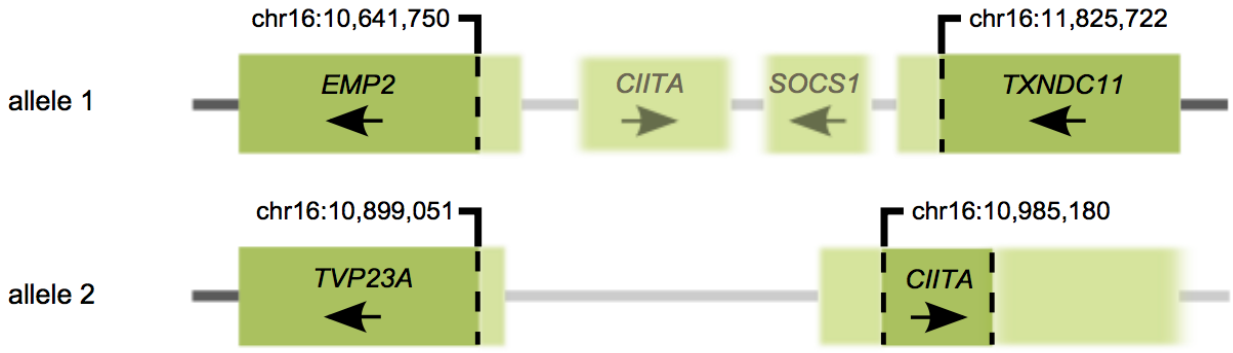

Figure 2.3: *CIITA* genetic alterations in the PMBCL-derived cell line Karpas1106P. Karpas1106P harbours a large deletion encompassing *CIITA* and *SOCS1*, leading to a novel gene fusion between *EMP2* and *TXNDC11*. The other allele is affected by a genomic deletion, resulting in the structural disruption of *CIITA.* All genomic coordinates are given according to the hg19 reference genome.

A similar analysis of U2940 led to the discovery of an in-frame NUBP1-CIITA fusion transcript with the amino terminus of the predicted chimeric protein encoded by the first nine exons of *NUBP1* and the carboxyl terminus encoded by all but the first exon of *CIITA*. The genomic deletion was determined to be hg19 chr16:

g.10858971_10974177del (Figure 2.4). Detailed information on the breakpoint sequence and effect at the transcript level are highlighted in Figure 2.5.

Since transcription of the *NUBP1-CIITA* fusion would be driven by the *NUBP1* promoter, instead of the strong B-cell *CIITA* promoter pIII, expression of CIITA from this allele would be expected to be reduced. Indeed, the majority of transcripts (97 %) detected by RNA-Seq were derived from the second allele, which in this cell line, harbours an SNV (hg19 chr16: g.11017158C>T) that results in the loss of the original stop codon (p.1131STOP>R*35; Figure 2.4).

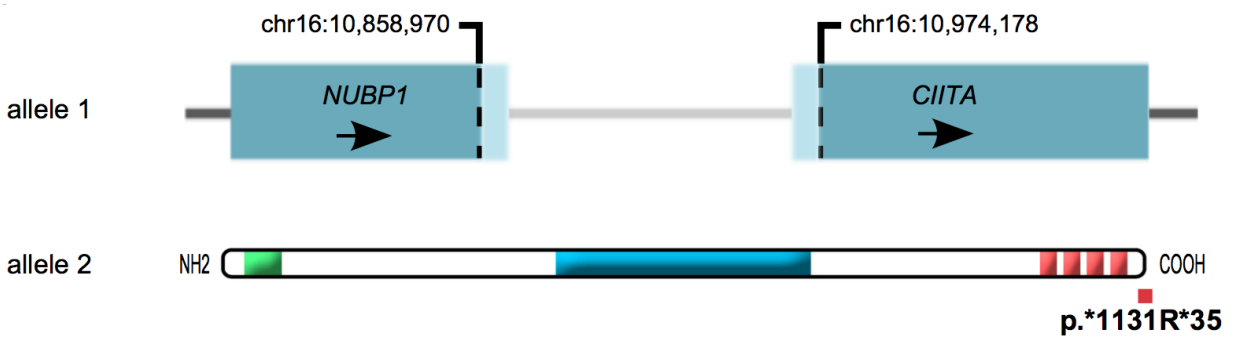

Figure 2.4: *CIITA* genetic alterations in the PMBCL-associated cell line U2940. In U2940, a large deletion results in a NUBP1-CIITA fusion transcript (for details see Figure 2.5). The second allele has acquired a base substitution mutation resulting in the loss of the stop codon. All genomic coordinates are given according to the hg19 reference genome.

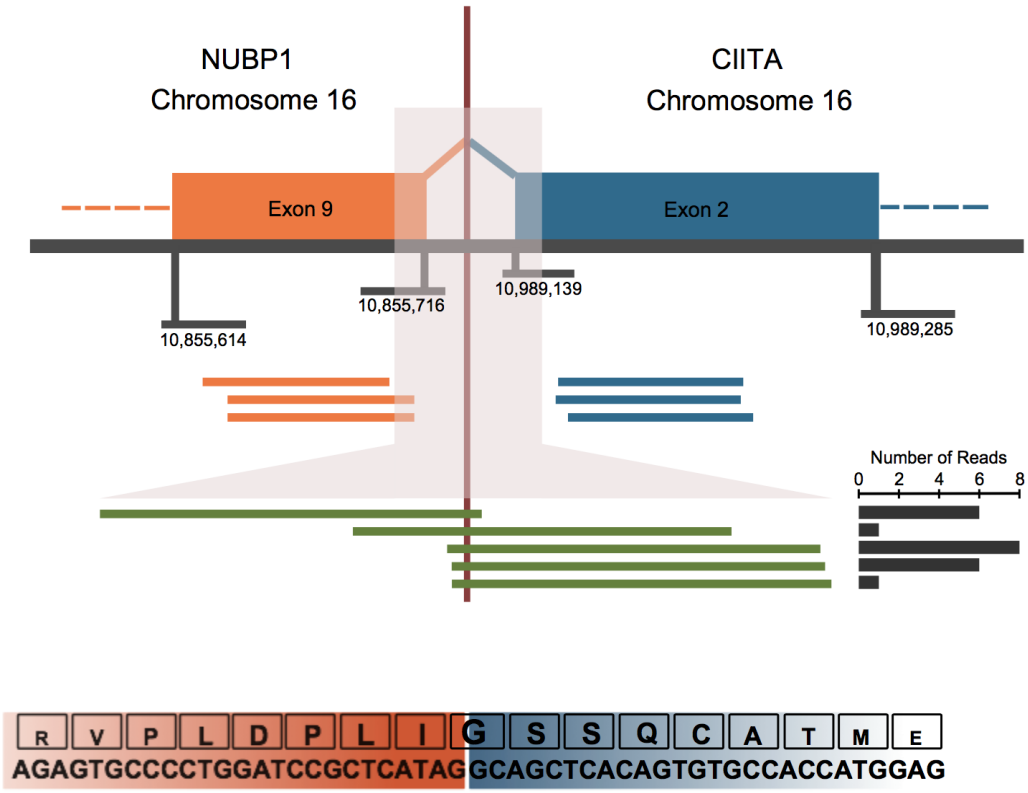

Figure 2.5: *NUBP1-CIITA* fusion observed in U2940 using RNA-Seq. The upper panel shows the genomic structure with paired read sequences aligning on either side of the breakpoint (orange and blue lines). Split reads spanning the breakpoint are highlighted in green with the histogram on the right showing the absolute frequency for each of these reads. The lower panel provides the reading frame at the breakpoint junction and the putative translation. All genomic coordinates are given according to the hg19 reference genome.

By integrating high-resolution CN analysis, FISH mapping and a stepwise PCR approach, we were able to decipher the genomic structure of the *CIITA* rearrangements in the cell line DEV (Figure 2.6). One allele is disrupted by a reciprocal translocation and deletion. The *CIITA-PDCD1LG2* fusion has been previously described [216] and similar fusions have been observed in two primary PMBCL cases [172]. The *PDCD1LG2-CIITA* fusion (t(9;16)(p24.1;p13.13)(chr9:g.5511575::chr16:g.10971955) is linked to a large deletion between *CIITA* intron 1 and *CLEC16A* that includes most of the coding sequence of *CIITA* (hg19 chr16: g.10973066_11275555del). This reciprocal translocation results in *PDCD1LG2* expression driven by the CIITA promoter and a complex *PDCD1LG2-CIITA-CLEC16A* rearrangement failing to encode a functional CIITA protein. Additionally, we identified an inversion involving the second *CIITA* allele

and *SOCS1* (hg19 chr16: g.10972393_11348602inv), that would lead to the functional abrogation of *CIITA*.

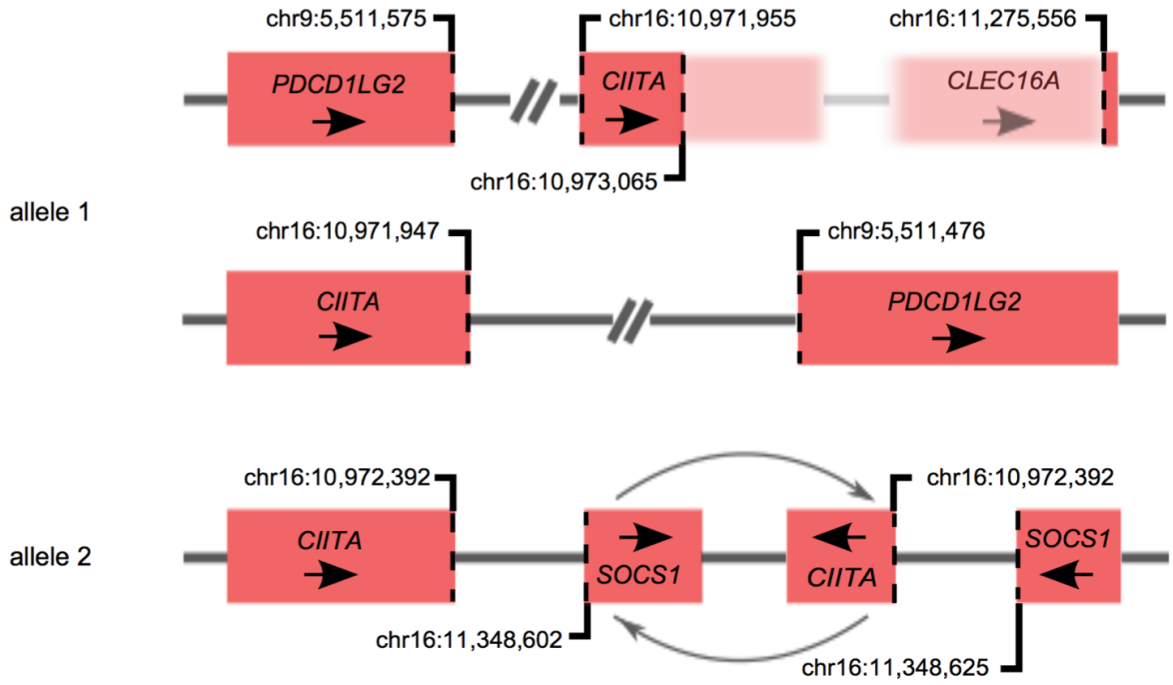

Figure 2.6: *CIITA* genetic alterations in the NLPHL-derived cell line DEV. *CIITA* rearrangements in DEV are complex, involving the PD-1 ligand *PDCD1LG2* and *SOCS1,* respectively. See main text for detailed description. All genomic coordinates are given according to the hg19 reference genome.

2.3.1.2 *CIITA* coding sequence mutations and structural genomic alterations in primary PMBCL cases

Prompted by the observed *CIITA* aberrations in cell lines, we next investigated to what extent such genetic alterations could be detected in primary PMBCL tumours. First, we focused on recently published WGS and RNA-Seq data of seven cases [104]. We identified one case (#19) harbouring a somatic CDS mutation (hg19 chr16: g.11001430T>C, p.694Leu>Ser), which has been validated by deep amplicon sequencing. A second case (#20) revealed a nonsense mutation (hg19 chr16: g.10992550C>T) resulting in the acquisition of a premature stop codon (p.107Glu>Stop). Both mutations were independently validated using Sanger sequencing. Based on these observations, we next screened for mutations within the 19 coding exons of *CIITA* and alterations in the pIII region in a larger cohort of PMBCL

biopsies. In total, we selected 45 cases ("sequencing cohort") from the CLC database at the BCCA based on availability of FF material obtained prior to treatment (A.1). After exclusion of known SNPs and synonymous mutations, we found 16 SNVs within the CDS of 13 clinical specimens, consisting of seven nonsense/frameshift mutations, one SNV resulting in loss of the start codon, four missense mutations and four splice site alterations affecting the exon 1 splice donor site. One of these cases (#16) also harboured a small deletion of 8 bp in the 5' UTR and a partial deletion of exon 1 that extends into the first intron. In addition, we observed another case (#2) with four alterations in the pIII region and a 232 bp deletion that encompasses the entirety of exon 1 from 14 bp upstream of the TSS extending into intron 1. One sample (#4) displayed a single bp substitution in pIII. Overall, alterations affecting the pIII region seem to be rare (five mutations in only two cases), whereas almost half of the CDS mutations (43.8 %) occur in the first exon. The distribution and types of mutations for each exon are shown in Figure 2.7, and a detailed assembly of genomic coordinates and the putative translational impact is given in B.1. For three PMBCL cases we confirmed the identified mutations as being of somatic origin by sequencing constitutional DNA extracted from peripheral blood (for details see B.1).

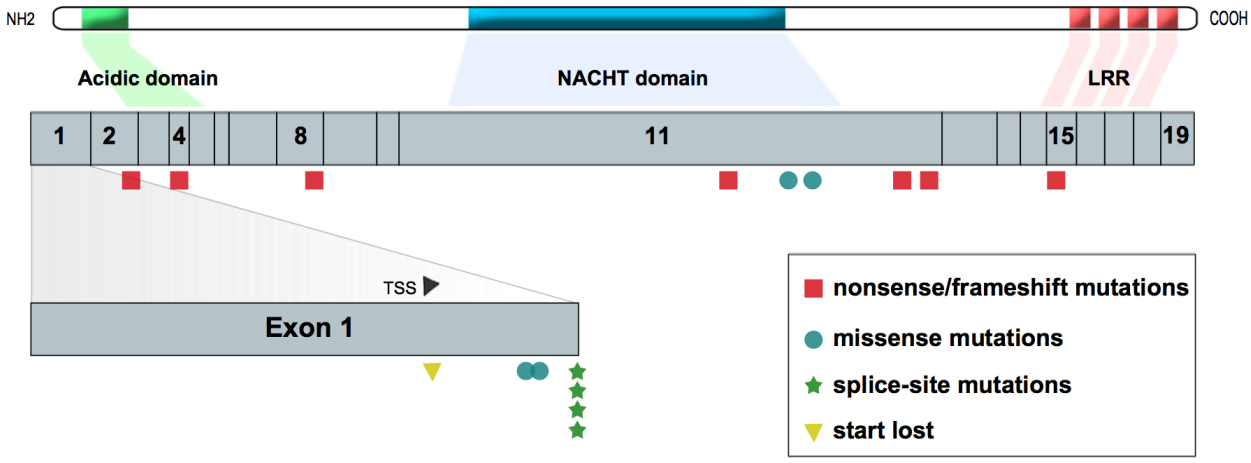

Figure 2.7: Coding sequence mutations in primary PMBCL cases. The distribution of *CIITA* coding sequence mutations in 44 analyzed primary PMBCL specimens identified by WGS, RNA-Seq, TruSeq custom amplicon sequencing and Sanger sequencing. Variations in non-coding regions, silent mutations, larger deletions and known SNPs are not shown. Abbreviations: TSS, translational start site; LRR, leucine-rich repeats.

In order to inform on structural chromosomal alterations in our sequencing cohort, we performed FISH on two TMAs, encompassing 148 PMBCL cases in total, including 41 of the cases in our sequencing cohort. We also performed FISH on whole tissue sections from FFPE tissue biopsies either not represented or initially not evaluable on the TMAs. In total, specimens from 150 individual patients were analyzed and FISH was evaluable in 116 cases (77 %), of which 77 were previously reported [172]. Thirty-nine cases (33.6 %) were scored as *CIITA* ba-positive in the entire cohort (n = 116), including 15 of 41 evaluable samples (36.6 %) in the sequencing cohort. A FISH signal constellation indicative of unbalanced rearrangements and/or deletion of genomic sequence next to the *CIITA* breakpoint was found in 44 % of ba-positive cases.

2.3.1.3 Intron 1 deletions and point mutations in primary PMBCL cases

We have previously shown that the majority of *CIITA* translocation breakpoints in PMBCL cluster within a distinct, 1.6 kb spanning region near the 5' end of intron 1 [172]. Of note, the breakpoints observed in DEV fall in the exact same region, prompting us to examine in detail the genomic sequence of this region in our sequencing cohort. Genomic DNA extending 3 kb from the end of exon 1 into intron 1 was PCR-amplified and visualized after agarose gel electrophoresis. Deletions were clearly evident in 13 of the 45 primary PMBCL samples but were not observed in 20 DLBCL cases, five reactive lymph nodes and three samples of purified tonsillar germinal centre B cells that were analyzed for comparison, suggesting that these deletions might be a specific event in PMBCL pathogenesis. Cloning and sequencing of the aberrant length fragments revealed frequent microdeletions, multiple base substitutions and (more rarely) small insertions, duplications and inversions. As many as 51 alterations were seen per clone and multiple SNVs were detected in some of the alleles lacking visible deletions. In total, 21 cases from the sequencing cohort were found to have sequence alterations within the first 3 kb of intron 1. The allelic origin of a cloned sequence can be inferred from the sequence at heterozygous SNP positions and showed that, at least in some cases, biallelic sequence variants have been acquired. Emphasizing the genomic instability of this region in PMBCL, analysis of 15 PCR clones amplified from purified tonsillar germinal centre B cells of three donors (five clones per donor) revealed only a

total of nine base substitutions among seven clones. A detailed description of the genomic aberrations is given in Figure 2.8 and B.2.

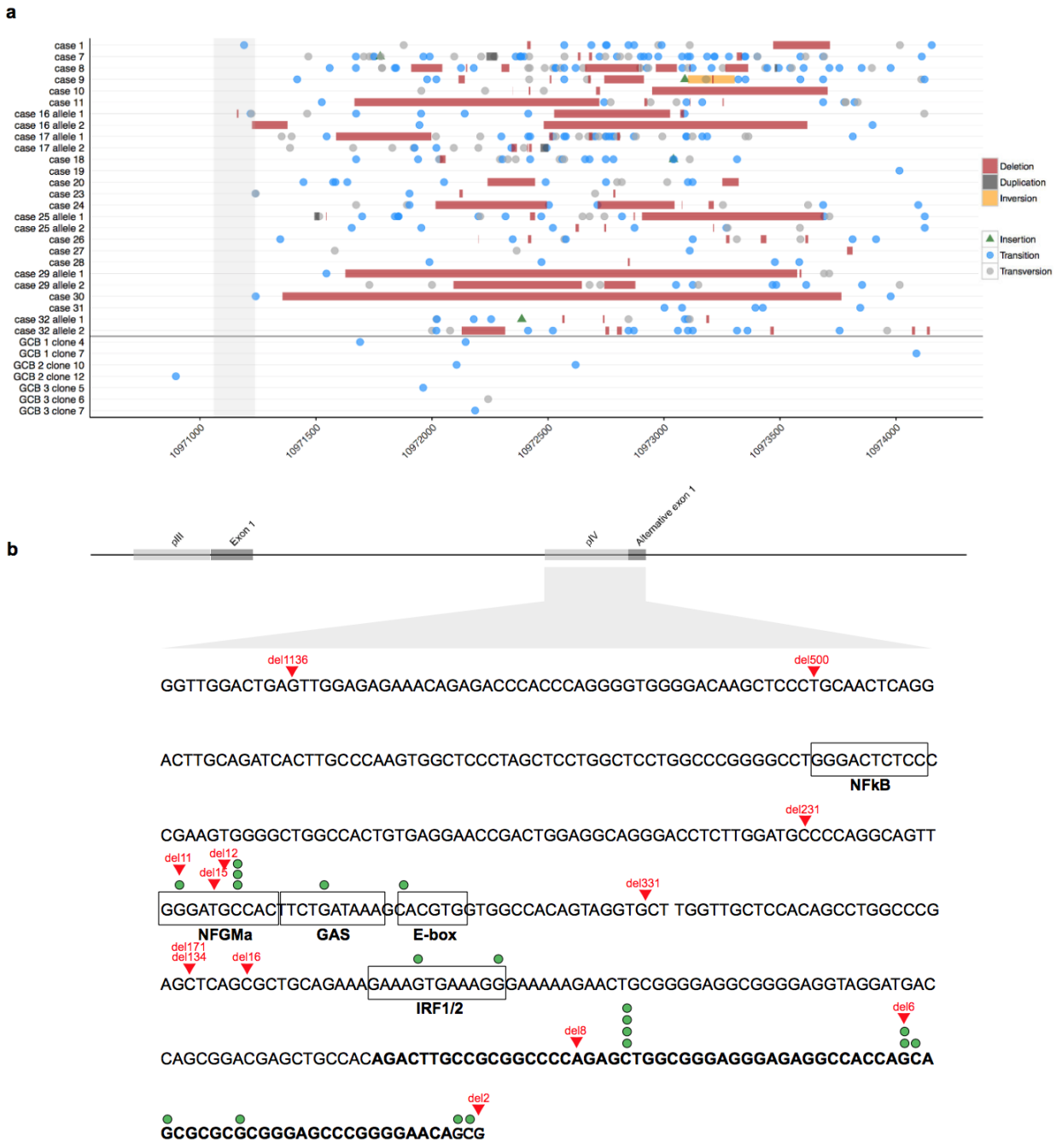

Figure 2.8: *CIITA* intron 1 alterations in primary PMBCL cases. Depicted is the *CIITA* genomic locus with the pIII region and the first 3 kb of intron 1. a) Sequencing analyses revealed multiple genomic aberrations. Point mutations and deletions detected in individual cases are shown. In contrast, sequence alterations were rare in reactive germinal centre B cells (GCB). b) For visualization purposes, only mutations affecting transcription factor binding sites and responsive elements in the pIV region or the alternative exon 1 are shown. Red triangles indicate deletions, green dots SNVs.

Interestingly, deletions and SNVs seemed to be enriched in a region encoding the alternative exon 1 of the *CIITA* transcript that is driven by the pIV promoter and is inducible in many different cell types upon IFN γ stimulation (Figure 2.8). In some primary tumours, we found clear evidence that a stepwise accumulation of genetic alterations occurred, indicating genomic instability and clonal evolution (Figure 2.9).

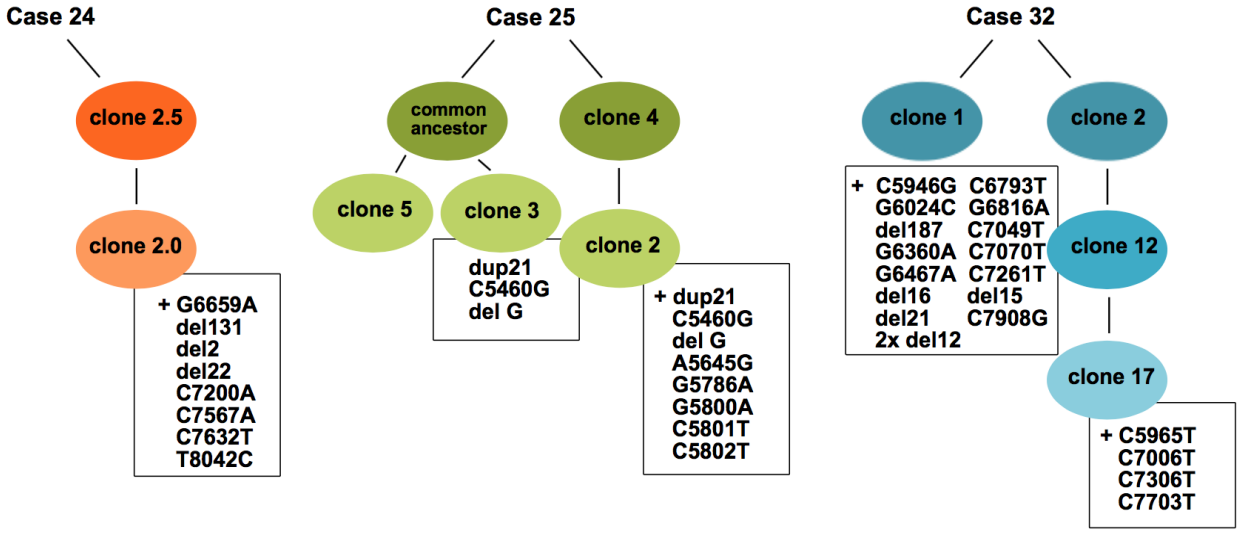

Figure 2.9: Subclonal evolution. Three primary PMBCL cases showed clear evidence for subclonal evolution, likely as the result of ongoing somatic hypermutation at the *CIITA* gene locus. Genetic aberrations which were acquired in individual subclones are shown in boxes adjacent to a particular cloned sequence.

Furthermore, the high prevalence of CDS mutations in exon 1 and alterations within the first 3 kb of intron 1 led us to speculate that these alterations are the result of AID-mediated aberrant SHM. We, therefore, analyzed all SNVs observed within 2 kb from the TSS to determine whether they represent transitions or transversions. In total, 233 transitions and 146 transversions occurred (ratio 1.6:1), which is considerably skewed compared to the theoretically expected transition/transversion ratio (1:2). Furthermore, we found a significant enrichment in mutated cytosine residues within AID target motifs than one would expect by chance (χ^2 test $P < .001$). Specifically, cytosine residues within AGC sequence motifs were most commonly affected (Figure 2.10).

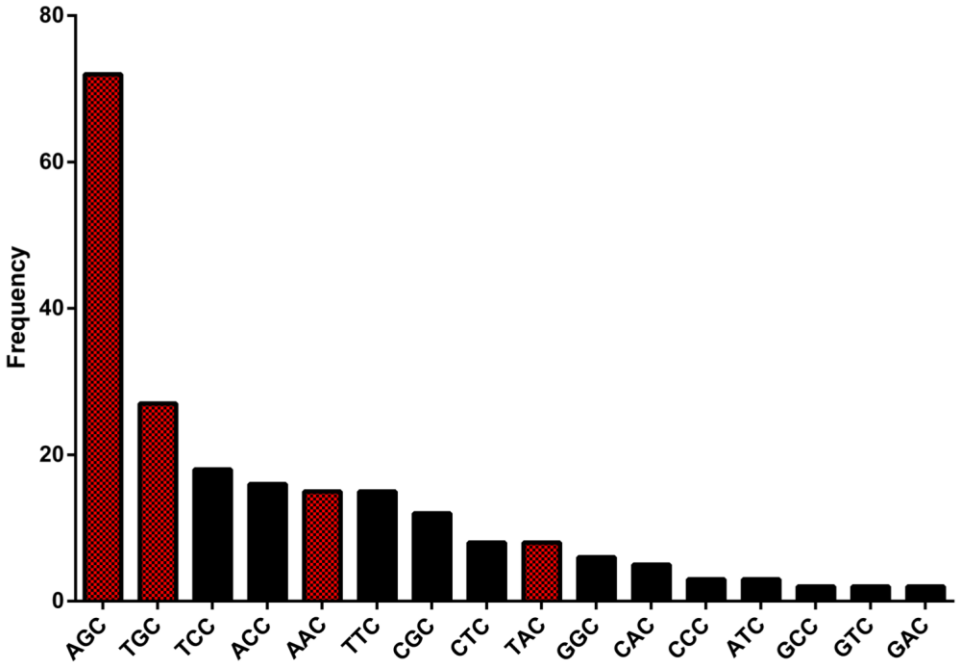

Figure 2.10: AID hotspot targets are frequently mutated in PMBCL. Absolute frequency of mutated cytosine residues within the genomic context of two basepairs upstream. Red bars indicate AID hot spot targets based on RGYW/WRCY motifs.

We then asked the question if AID protein expression is detectable in diagnostic biopsy material of PMBCL cases. Immunohistochemistry was evaluable in 114 cases. 48.2 % (55/114) were considered positive with variable percentages of tumour cells stained. Two representative cases are depicted in Figure 2.11.

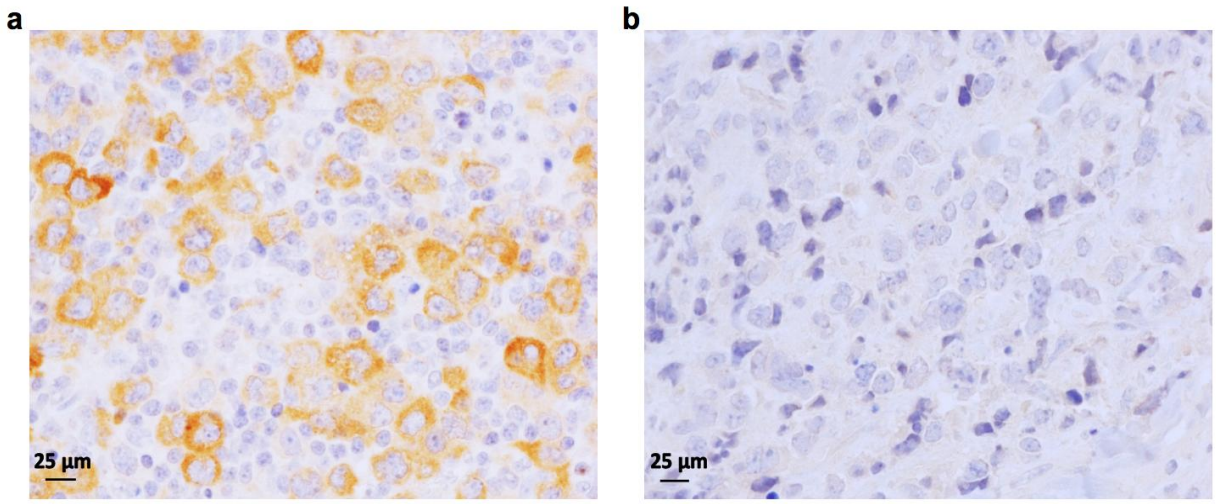

Figure 2.11: AID protein expression in PMBCL cases. The large neoplastic cells in a) show moderate to strong cytoplasmic positivity, whereas the tumour cells in b) are negative. Original magnification: x400.

2.3.2 *CIITA* alterations in DLBCL

In the cohort of 347 *de novo* DLBCL cases, we found 36 mutations in 31 individual specimens, resulting in a mutation frequency of 8.9 % for the *CIITA* gene in this lymphoma entity. The majority consisted of missense mutations (29/36; 80.5 %), followed by splice site alterations (4/36; 11.1 %) and frameshift mutations (3/36; 8.3 %). The distribution and types of mutations for each exon are shown in Figure 2.12 and a detailed assembly of genomic coordinates and the putative translational impact is given in Table 2.3. Notably, when integrating data obtained from the molecular subtyping of these cases using the Lymph2Cx NanoString assay [65,231], no significant association with the GCB- or ABC-DLBCL subtype could be observed (Fisher's exact test: P = .38). Compared to the distribution of CDS mutations in PMBCL (see Figure 2.7 in section 2.3.1.2), no obvious enrichment within the first exon of *CIITA* could be observed in DLBCL cases. Instead, most of the mutations seemed to cluster in the gene region encoding for the conserved NACHT-domain (GTP-binding domain) and the LRR-domain at the carboxyl-terminus.

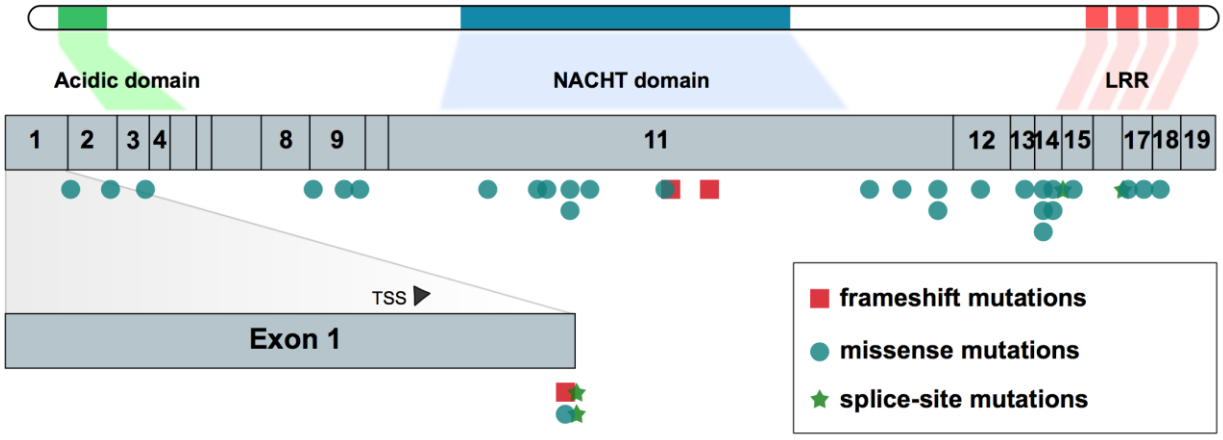

Figure 2.12: Coding sequence mutations in primary DLBCL cases. The distribution of *CIITA* coding sequence mutations in 347 analyzed primary DLBCL specimens identified by TruSeq custom amplicon sequencing. Variations in non-coding regions, silent mutations and known SNPs are not shown. Abbreviations: TSS, translational start site; LRR, leucine-rich repeats.

Table 2.3: CDS mutations in DLBCL.

Case	Gene	Chromosome	Position	Ref	Alt	VAF	Mutation type	cDNA	AA change
DLC006	CIITA	chr16	11001778	T	A	26.03	MISSENSE	2429T>A	Leu810Gln
DLC007	CIITA	chr16	10997738	G	C	50.63	MISSENSE	923G>C	Arg308Pro
DLC018	CIITA	chr16	11001940	G	C	71.35	MISSENSE	2591G>C	Arg864Pro
DLC041	CIITA	chr16	11009508	G	T	40.89	SPLICE_SITE		
DLC041	CIITA	chr16	11016055	G	A	28.84	MISSENSE	3181G>A	Ala1061Thr
DLC041	CIITA	chr16	11009492	C	T	40.89	MISSENSE	2954C>T	Ser985Phe
DLC041	CIITA	chr16	11009491	T	A	40.89	MISSENSE	2953T>A	Ser985Thr
DLC048	CIITA	chr16	11000793	G	A	46.91	MISSENSE	1444G>A	Glu482Lys
DLC073	CIITA	chr16	10971233	C	CCCCA	6.07	FRAME_SHIFT	47_48insCCCA	Gln17Hisfs*22
DLC080	CIITA	chr16	11016023	G	A	5.65	SPLICE_SITE		
DLC092	CIITA	chr16	11000940	G	A	47.08	MISSENSE	1591G>A	Gly531Ser
DLC103	CIITA	chr16	11016097	C	T	43.02	MISSENSE	3223C>T	Arg1075Trp
DLC142	CIITA	chr16	10989151	G	T	62.46	MISSENSE	65G>T	Cys22Phe
DLC142	CIITA	chr16	11009462	A	G	25.24	MISSENSE	2924A>G	Lys975Arg
DLC149	CIITA	chr16	11001295	C	G	45.75	MISSENSE	1946C>G	Ala649Gly
DLC151	CIITA	chr16	10997686	G	A	8.67	MISSENSE	871G>A	Ala291Thr
DLC167	CIITA	chr16	11016287	C	G	11.95	MISSENSE	3257C>G	Ala1086Gly
DLC169	CIITA	chr16	10971240	G	A	49.17	SPLICE_SITE		
DLC173	CIITA	chr16	11000940	G	A	50.91	MISSENSE	1591G>A	Gly531Ser
DLC178	CIITA	chr16	10989612	G	A	61.93	MISSENSE	286G>A	Ala96Thr
DLC183	CIITA	chr16	10971233	C	A	22.51	MISSENSE	46C>A	Pro16Thr
DLC222	CIITA	chr16	11009462	A	G	49.15	MISSENSE	2924A>G	Lys975Arg
DLC251	CIITA	chr16	10997588	G	A	65.76	MISSENSE	773G>A	Gly258Asp
DLC283	CIITA	chr16	11001388	TC	T	28.42	FRAME_SHIFT	2040_2041delC	Pro680Hisfs*3
DLC286	CIITA	chr16	11001304	G	GC	42.15	FRAME_SHIFT	1956_1957insC	Gly655Argfs*92
DLC291	CIITA	chr16	11001843	C	T	49.64	MISSENSE	2494C>T	Arg832Cys
DLC293	CIITA	chr16	11009462	A	G	37.08	MISSENSE	2924A>G	Lys975Arg
DLC305	CIITA	chr16	11000815	A	G	53.12	MISSENSE	1466A>G	Lys489Arg
DLC319	CIITA	chr16	11004092	G	A	46.10	MISSENSE	2864G>A	Arg955Gln
DLC330	CIITA	chr16	11010253	G	T	11.59	MISSENSE	2999G>T	Gly1000Val
DLC353	CIITA	chr16	11001939	C	T	51.11	MISSENSE	2590C>T	Arg864Cys
DLC360	CIITA	chr16	11002983	C	T	26.09	MISSENSE	2755C>T	Pro919Ser
DLC360	CIITA	chr16	11000658	T	A	24.25	MISSENSE	1309T>A	Trp437Arg
DLC366	CIITA	chr16	10989143	C	G	13.03	MISSENSE	57C>G	Ser19Arg
DLC376	CIITA	chr16	10971240	G	A	40.82	SPLICE_SITE		
DLC379	CIITA	chr16	11000992	G	A	51.29	MISSENSE	1643G>A	Arg548Gln

2.3.3 *CIITA* alterations in FL

2.3.3.1 *CIITA* coding sequence mutations and structural genomic alterations in primary FL cases

In the entire FL cohort, consisting of transformed and progressed/non-progressed cases, we detected 28 mutations affecting the CDS of *CIITA.* Similar to the DLBCL cohort, the majority represented missense mutations (19/28; 67.9 %). Furthermore, we found five nonsense/frameshift mutations (5/28; 17.9 %), one case with a small deletion of 3 bp leading to the loss of a codon and three alterations (3/28; 10.7 %) affecting splice sites. The distribution and respective types of mutations for each exon are shown in Figure 2.13 and a detailed assembly of genomic coordinates and the putative translational impact is given in Table 2.4.

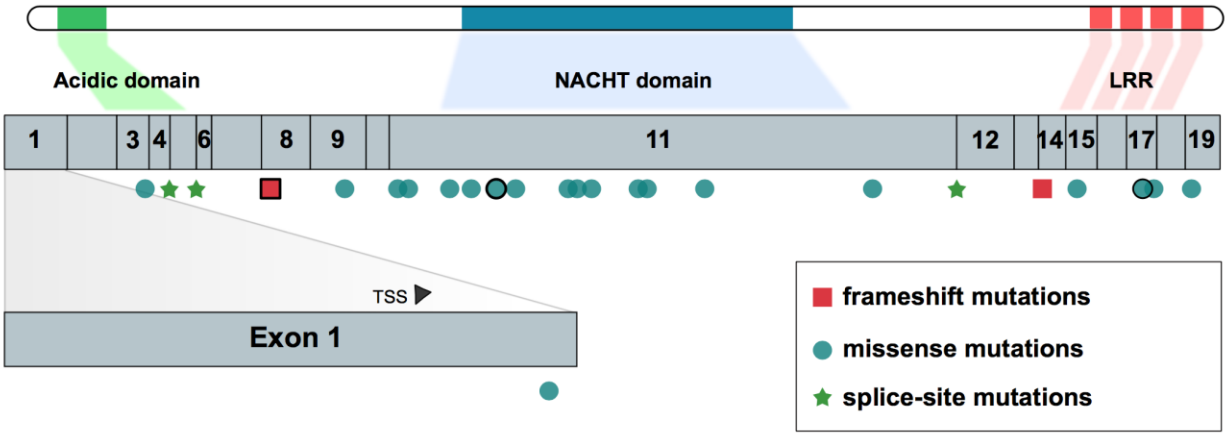

Figure 2.13: Coding sequence mutations in primary FL cases. The distribution of *CIITA* coding sequence mutations in 397 analyzed primary FL specimens identified by targeted capture sequencing. Framed symbols indicate that this alteration was detected in both, the T1 and T2 timepoint of an individual case. Variations in non-coding regions, silent mutations and known SNPs are not shown. Abbreviations: TSS, translational start site; LRR, leucine-rich repeats.

Table 2.4: CDS mutations in FL

Case	Gene	Chromosome	Position	Ref	Alt	VAF	Mutation type	cDNA	AA change
FL1002T1	CIITA	chr16	11000647	T	G	0.13	MISSENSE	1298T>G	Val433Gly
FL1002T2	CIITA	chr16	11000647	T	G	0.37	MISSENSE	1298T>G	Val433Gly
FL1013T1	CIITA	chr16	10996526	AGTTCCTC	A	0.22	FRAME_SHIFT	641_642delGTTCCTC	Ser214*
FL1013T2	CIITA	chr16	10996526	AGTTCCTC	A	0.26	FRAME_SHIFT	641_642delGTTCCTC	Ser214*
FL1014T2	CIITA	chr16	11009456	T	TCC	0.85	FRAME_SHIFT	2919_2920insCC	Phe973Phefs*20
FL1019T2	CIITA	chr16	11001373	T	G	0.29	MISSENSE	2024T>G	Leu675Arg
FL1116T2	CIITA	chr16	11000362	T	C	0.13	MISSENSE	1013T>C	Val338Ala
FL1122T1	CIITA	chr16	11000686	A	T	0.27	MISSENSE	1337A>T	Asp446Val
FL1145T1	CIITA	chr16	11002013	T	G	0.09	SPLICE_SITE_REGION		
FL1178T1	CIITA	chr16	11001155	ACTT	A	0.15	CODON_DELETION	1807_1808delCTT	Leu603X
FL1190T2	CIITA	chr16	10992591	T	G	0.34	SPLICE_SITE_DONOR		
FL1192T2	CIITA	chr16	11001783	T	G	0.34	MISSENSE	2434T>G	Cys812Gly
FL1192T2	CIITA	chr16	11010255	G	T	0.39	MISSENSE	3001G>T	Asp1001Tyr
FL1198T1	CIITA	chr16	11000865	G	A	0.12	MISSENSE	1516G>A	Ala506Thr
FL1204T2	CIITA	chr16	11000848	G	C	0.48	MISSENSE	1499G>C	Gly500Ala
FL1221T2	CIITA	chr16	10995370	G	T	0.14	SPLICE_SITE_ACCEPTOR		
FL1221T2	CIITA	chr16	11000589	C	T	0.09	NONSENSE	1240C>T	Arg414*
FL1221T2	CIITA	chr16	11016104	T	C	0.07	MISSENSE	3230T>C	Met1077Thr
FL1229T2	CIITA	chr16	11001122	A	C	0.10	MISSENSE	1773A>C	Gln591His
FL1252T2	CIITA	chr16	11017098	A	G	0.43	MISSENSE	3331A>G	Thr1111Ala
FL2005T1	CIITA	chr16	11016071	G	A	0.97	MISSENSE	3197G>A	Arg1066His
FL2005T2	CIITA	chr16	11016071	G	A	1.00	MISSENSE	3197G>A	Arg1066His
FL2111T1	CIITA	chr16	11000388	G	A	0.42	MISSENSE	1039G>A	Asp347Asn
FL2111T1	CIITA	chr16	11000895	G	A	0.50	MISSENSE	1546G>A	Gly516Arg
FL2115T1	CIITA	chr16	11000532	G	T	0.44	NONSENSE	1183G>T	Gly395*
FL3013T1	CIITA	chr16	10989621	G	A	0.53	MISSENSE	295G>A	Ala99Thr
FL3140T1	CIITA	chr16	10971192	G	A	0.22	MISSENSE	5G>A	Arg2His
FL3144T1	CIITA	chr16	10997689	C	T	0.25	MISSENSE	874C>T	Pro292Ser

Out of 277 successfully sequenced specimens in the tFL subset of the cohort, six samples which were obtained at the time of the initial FL diagnosis (T1 timepoint; 4.7 %) and 11 samples which were collected at the time of transformation (T2 timepoint; 7.4 %) carried mutations, with only two cases (out of 15) sharing the mutation between T1 and T2. There was no enrichment for *CIITA* mutations in the T2 specimens since four cases had a mutation present at T1 that was not detected at T2, whereas six cases showed mutations at the timepoint of transformation which could not be seen in the respective T1 specimen (McNemar's test: $P = .752$). In addition, three cases with T2 mutations did not have sequencing data from their respective primary diagnostic specimen (T1) available.

Among the early progressers (pFL), 8.1 % showed CDS mutations (3/37), whereas 3.6 % (3/83) in the late/no progressive group (npFL) were found to be mutated. Hence, there was no significant difference between pFL and npFL with regards to the prevalence of *CIITA* CDS mutations (Fisher's exact test: $P = .37$). Likewise, *CIITA* mutations were not more frequently observed in cases, which eventually transformed compared to those which progressed (Fisher's exact test: $P = 1$).

We also performed FISH on all cases that were arranged on TMAs to interrogate the rearrangement status of *CIITA*. Out of 273 tissue specimens in the tFL subset of the cohort, 250 were evaluable (91.6 %). Thirty-five samples which were obtained at the time of the initial FL diagnosis (T1 timepoint; 29.9 %) and 49 samples which were collected at the time of transformation (T2 timepoint; 36.8 %) showed a signal constellation indicative of a *CIITA* ba-status. For 100 cases, information on *CIITA* ba-status was available for both timepoints, with 26 cases sharing the chromosomal rearrangement between T1 and T2. When restricting the analysis to those paired specimens, there was a significant enrichment for *CIITA* alterations in the T2 specimens, since, in addition to the 26 cases with concordant FISH results, five cases were ba-positive at T1 but not at T2, whereas 17 cases showed ba signals at the timepoint of transformation which could not be seen in the respective T1 specimen (McNemar's test: $P = .019$).

Among the early progressers 15.2 % showed *CIITA* ba-positivity (5/33 cases), whereas in the npFL group four out of 74 cases (5.4 %) were translocated. Hence, there

was no significant difference between pFL and npFL with regard to the prevalence of *CIITA* structural chromosomal aberrations (Fisher's exact test: $P = .13$). However, *CIITA* ba-positive cases were significantly enriched in the cohort of FL specimens, which eventually transformed to a high-grade lymphoma compared to those which progressed (Fisher's exact test: $P < .0001$).

2.3.3.2 Intron 1 deletions and point mutations in primary FL cases

The *CIITA* gene was also assessed as a SHM target in the FL capture sequencing panel and, therefore, mutation calls for intron 1 were available and analyzed similar to the PMBCL cohort (see 2.3.1.3).

For the tFL cohort, we focused on the paired specimens, for which we had sequencing data from both timepoints available (n = 118). In 80 cases (67.8 %) intron 1 mutations were detected and, in the vast majority, were seen in the specimens from both timepoints (93.75 %). As many as 46 alterations were seen, most of them confined to the region within 2 kb of the TSS, known to be a hotspot for AID-mediated SHM (Figure 2.14, B.3). We, therefore, analyzed all SNVs observed within 2 kb from the TSS to determine whether they represent transitions or transversions. In total, 384 transitions and 168 transversions occurred (ratio 2.3:1), which is, similar to PMBCL, considerably skewed compared to the theoretically expected ratio (1:2). Again, we found a significant enrichment in mutated cytosine residues within AID target motifs than one would expect by chance (χ^2 test: $P < .001$). Specifically, cytosine residues within AGC sequence motifs were most commonly affected (Figure 2.15).

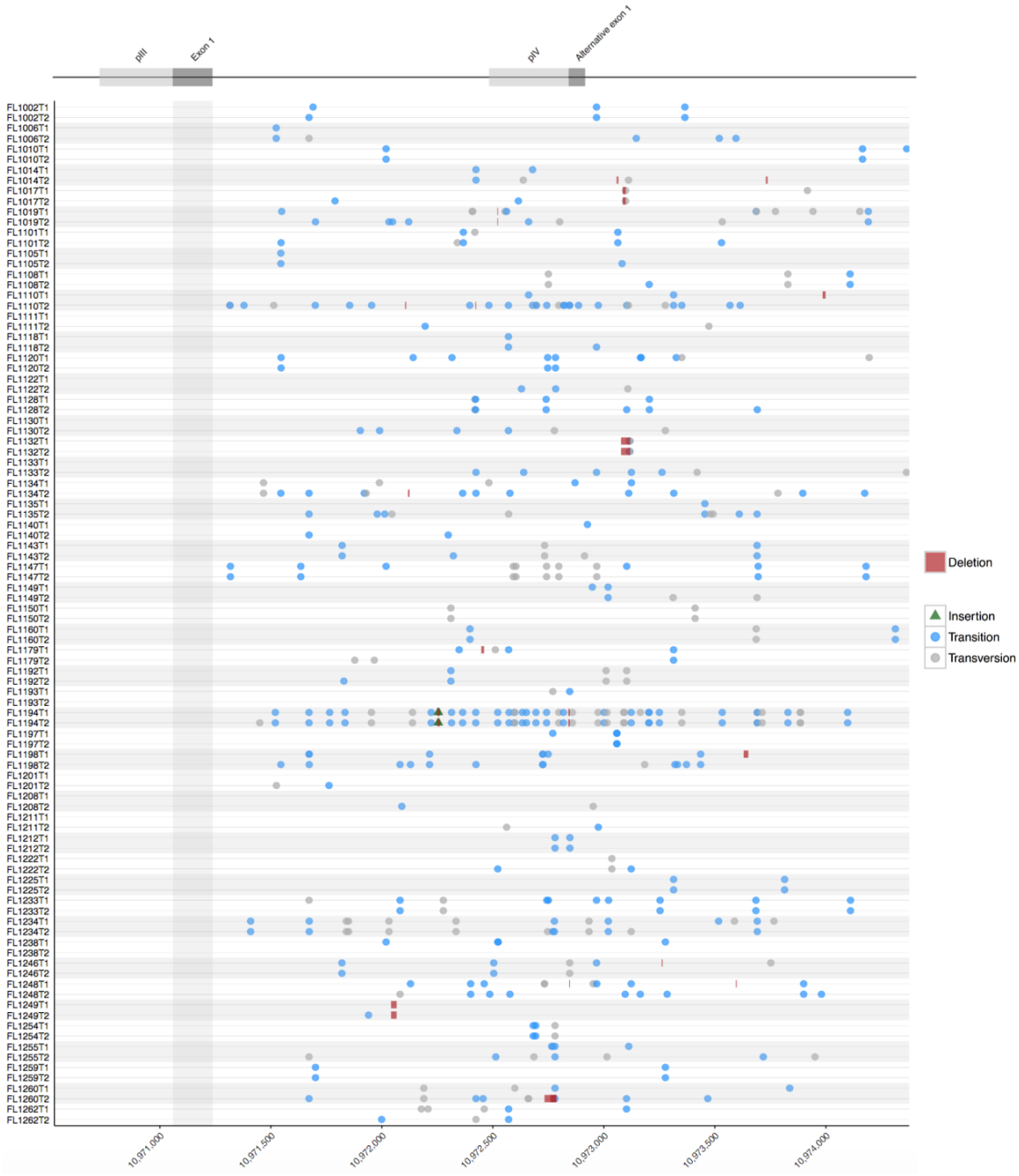

Figure 2.14: *CIITA* intron 1 alterations in tFL cases. Depicted is the genomic locus of *CIITA* encompassing the pIII region and the first 3 kb of intron 1. For visualization purposes, only paired tFL cases and those which harbour at least two alterations are shown. Sequencing analyses revealed multiple genomic aberrations, including point mutations, deletions and insertions within the first 3 kb of intron 1.

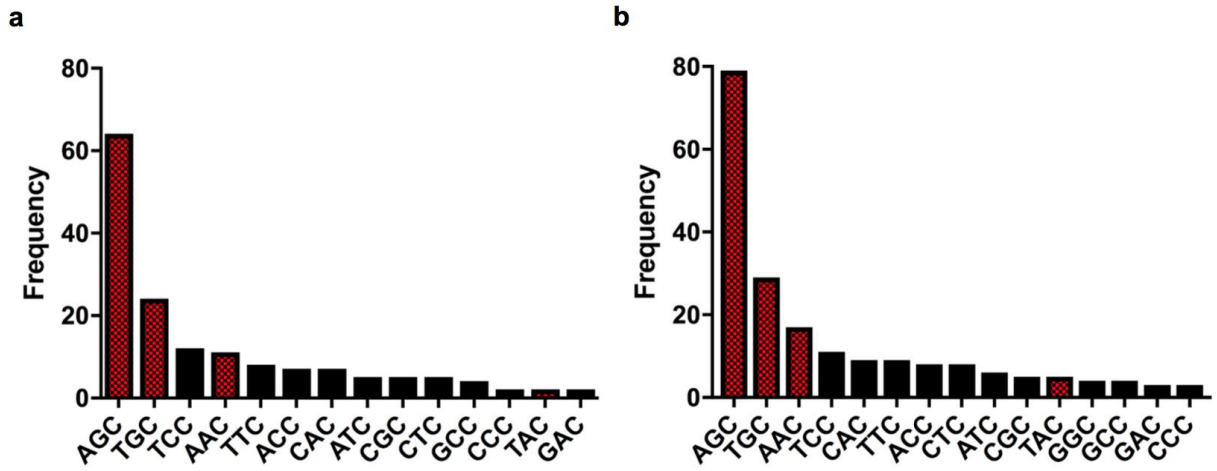

Figure 2.15: AID hotspot targets are frequently mutated in tFL. Absolute frequency of mutated cytosine residues within the genomic context of two basepairs upstream in a) the T1 specimens, and b) the T2 specimens. Red bars indicate AID hot spot targets based on RGYW/WRCY motifs.

A similar analysis was then performed for the pFL and npFL cases (Figure 2.16). As many as 35 individual mutations were seen per case (B.4). In pFL, 56 transitions and 15 transversions occurred (ratio 3.7:1), whereas in npFL, 108 transitions and 55 transversions were observed (ratio 2:1). Again, similar to the distributions in tFL and PMBCL, these ratios are considerably skewed compared to the theoretically expected transition/transversion ratio (1:2). Furthermore, we found a significant enrichment in mutated cytosine residues within AID target motifs than one would expect by chance (χ^2 test $P < .001$) and cytosine residues within AGC sequence motifs were commonly affected (Figure 2.17).

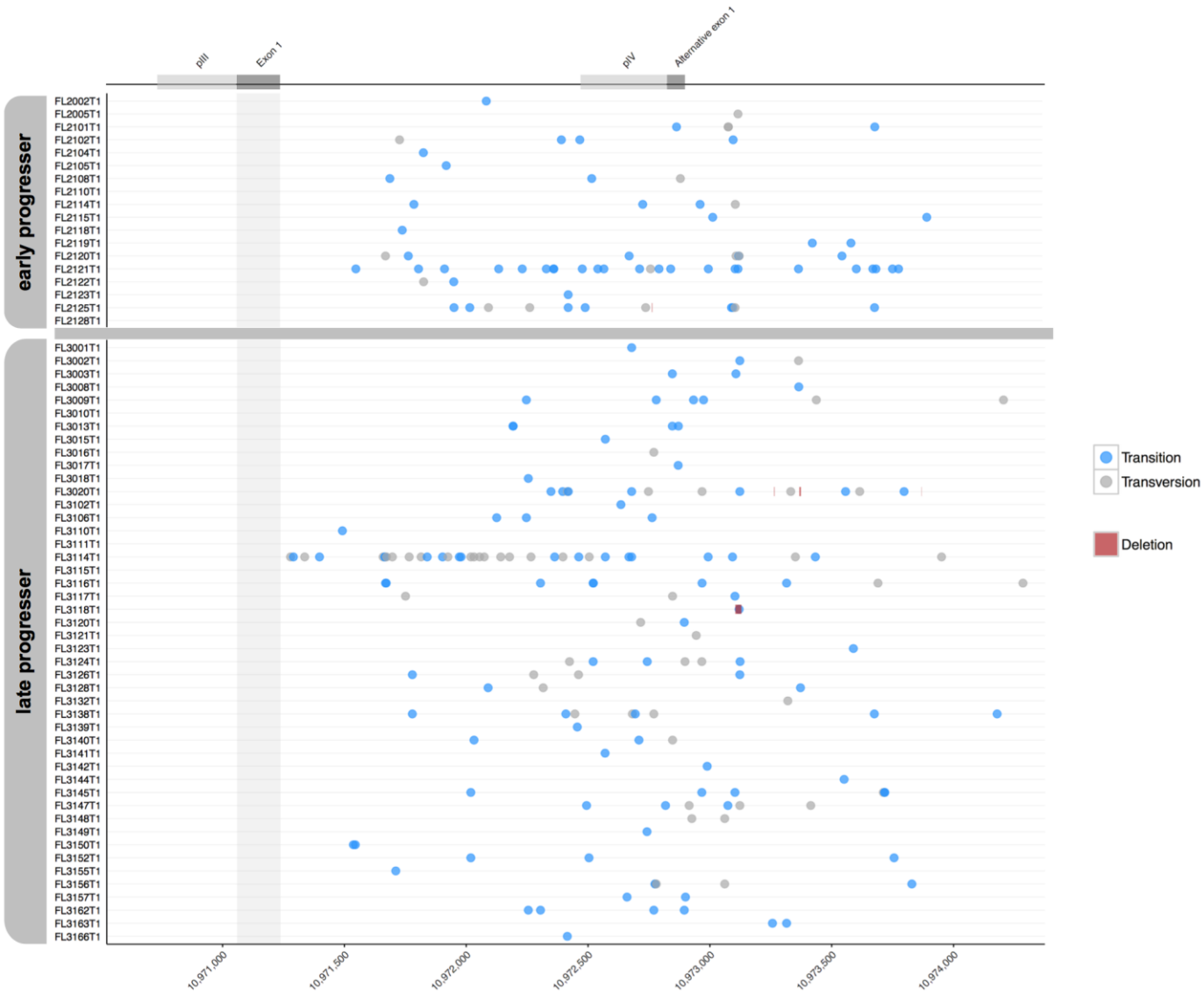

Figure 2.16: *CIITA* intron 1 alterations in early and late progressers. Depicted is the genomic locus of *CIITA* encompassing the pIII region and the first 3 kb of intron 1. Sequencing analyses revealed multiple genomic aberrations, including point mutations and deletions within the first 3 kb of intron 1.

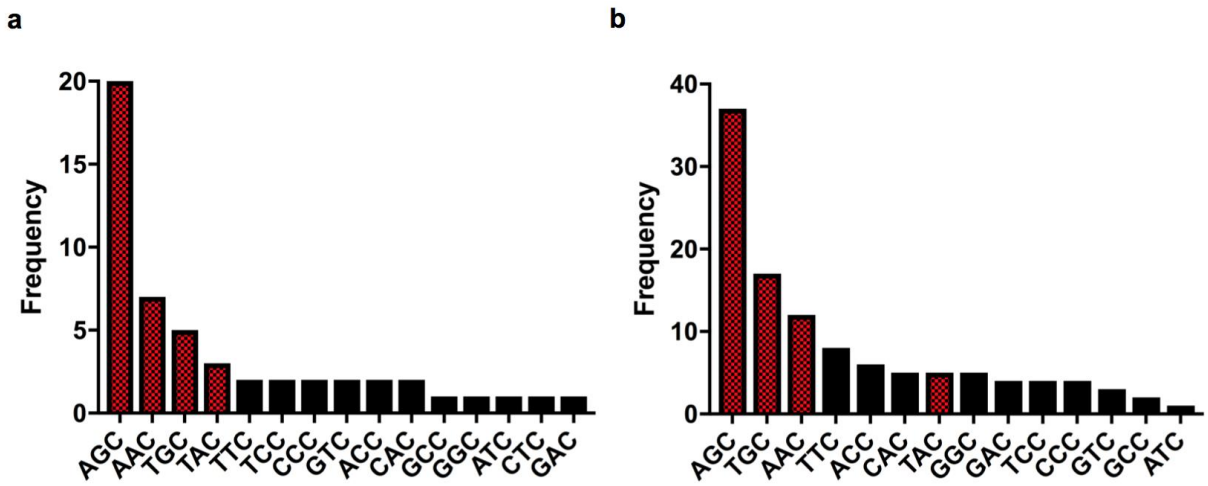

Figure 2.17: AID hotspots targeted by mutations in pFL and npFL. Absolute frequency of mutated cytosine residues within the genomic context of two basepairs upstream in a) early progressers and b) late/no progressers. Red bars indicate AID hot spot targets based on the RGYW/WRCY motifs.

2.3.4 BAC capture

Having *CIITA* firmly established as a recurrent translocation and gene fusion partner in lymphoid malignancies [172] accentuated the need for a comprehensive and detailed description of breakpoint anatomy, as well as the identity of the rearrangement partner genes. First, we explored the utility of BAC capture sequencing in a small pilot project, consisting of four cases with known FISH results for *CIITA*, PD-1 ligands (*CD274* and *PDCD1LG2*) and *TBL1XR1*. BAC capture sequencing of FFPE tissue yielded a mean coverage depth of 82x across the capture space in the four enriched libraries; 16.7 % of reads were non-PCR duplicates that aligned to hg19/GRCh37. Following stringent filtering, a list of 16 genomic rearrangements in the four libraries was generated. In the extranodal DLBCL positive control, we observed a *TBL1XR1-TP63* inversion and a *BCL2-IGHV3-9* translocation, validating previously published FISH data [69]. A complete list with rearrangement partners, genomic breakpoints and putative translation of the other 14 rearrangement events are presented in Table 2.5 and are visualized in the circos plot (Figure 2.18).

.

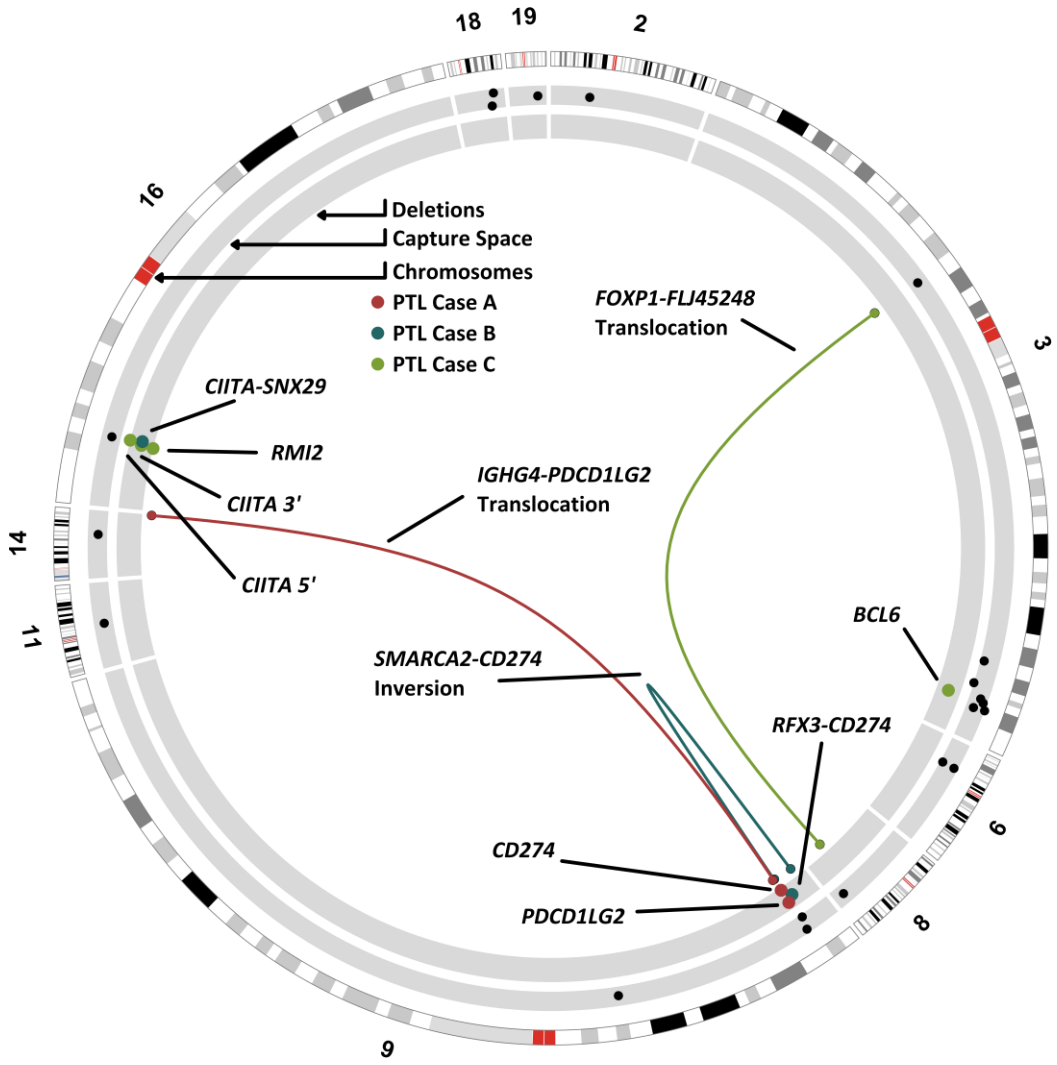

Figure 2.18: BAC capture results for three PTL cases. Circos plot depicting the 20 regions of the BAC capture space spanning 7.45 Mb and the 11 structural genomic events observed in three libraries obtained from PTL archival FFPE tissue. The outer-most concentric track represents the chromosomes, the middle track denotes the capture space, and the inner-most track shows deletion events. The adjoining lines depict rearrangement partners. Events are colour-coded to reflect those which overlap between cases. Only chromosomes relevant to the capture space have been included and chromosomes 3, 9 and 16 have been magnified to improve resolution. The regions of the chromosomes that are highlighted in red are the centromeres. Reproduced with permission from [217].

In the context of this thesis, only alterations involving *CIITA* are discussed. The interested reader is referred to Twa et al. for more details [217]. In case B, we observed an interstitial deletion juxtaposing intron 15 of *CIITA* with intron 14 of *SNX29* (16p13.13). This deletion results in a partial transcriptional loss of the C-terminal LRR domain and putatively produces an out-of-frame fusion leading to a premature stop. It is therefore expected that this may result in a dominant-negative CIITA protein [172]. In

case C, we observed biallelic deletions of *CIITA,* which likely cripple normal function of the protein. The 30 kb deletion on one allele includes both promoter regions (pIII and pIV), crucial for transcription in a B-cell context, while the deletion on the other allele likely results in an in-frame splice variant of exon 1 to exon 10 that lacks the functionally relevant acidic domain. Inexplicably, these events were not sizeable enough to explain the unbalanced signal pattern observed using FISH, suggesting a more complex rearrangement may be present. Extending our study of *CIITA* rearrangements in PTL, we performed FISH on a TMA including additional 85 specimens using the *CIITA* break-apart assay. In total, we observed a ba signal pattern in 8/82 (10 %) evaluable cases, suggesting that *CIITA* rearrangements are recurrent and may be responsible for reduced levels of MHC class II expression in a subset of PTLs [164]

Table 2.5: Structural genomic rearrangements in three PTL cases.

Library	Involved Genes	Structural Rearrangement Type	Genomic Location	Exact Genomic Breakpoints	Putative Translational Significance	Sanger Sequencing Validated
A	*CD274*	Interstitial deletion	Exon 7-3'intergenic	chr9:5468110-chr9:5489373	IE; transcript stability	Y
	PDCD1LG2	Interstitial deletion	Intron 4-Exon 7	chr9:5554986-chr9:5570162	IE; protein solubility	Y
	IGHG4-PDCD1LG2	Translocation	5'intergenic-3'intergenic	chr14:106094861-chr9:5572111	IE	Y*****
B	*CIITA-SNX29*	Interstitial deletion	Intron 15-Intron 14	chr16:11011017-chr16:12330239	DN; reduced MHCII	Y*****
	SMARCA2- CD274	Inversion	5'intergenic-Exon 7	chr9:1948958-chr9:5468076	IE; transcript stability	N
	RFX3-CD274	Interstitial deletion	Exon 16-Exon 7	chr9:3220442-chr9:5468461	IE of PDCD1LG2	Y
C	*CIITA*	Interstitial deletion	5'intergenic-Intron 3	chr16:10960096-chr16:10990709	DE; reduced MHCII	Y
	CIITA	Interstitial deletion	Intron 1-Intron 9	chr16:10988710-chr16:10997807	DN; reduced MHCII	Y
	FOXP1-FLJ45248	Translocation	Intron 11-3'intergenic	chr3:71074272-chr8:103828413	IE; pro-apoptotic repression	N*****
	BCL6	Interstitial deletion	5'intergenic-Intron 1	chr3:187461229-chr3:187463676	IE; cell cycle progression	N
	RMI2	Interstitial deletion	Intron 1-Intron 2	chr16:11389010-chr16:11389154	DE; genomic instability	N

DE, decreased expression; DN, dominant-negative phenotype; IE, increased expression; Y, validated; N, not validated, due to exhaustion of genomic material; *, also observed by FISH. Reproduced from [217] with permission.

2.3.5 Capture sequencing

The high technical demands of the BAC capture technique, as well as the enormous expenditure of time, forced us to explore alternative approaches for the decoding of translocation breakpoint anatomy. Another prerequisite that needed to be fulfilled was the suitability for using FFPE tissue. Targeted capture sequencing is ideally suited for interrogation of a circumscribed genomic region and is applicable to archival tissue specimens. Based on previous FISH results, in total 92 specimens were selected for capture sequencing. Those included 91 clinical cases (41 PMBCLs, 39 DLBCLs and 11 FLs) and a cHL-derived cell line (L1236). Eighty-five of the clinical specimens were FFPE tissues and six were FF cell suspensions. A.5 provides an overview of the specimens with their respective FISH pattern for the loci included in the capture space. Twenty-four specimens failed library construction, either due to insufficient DNA amounts (20 cases) or due to heavily fragmented DNA (four cases). The 68 successful libraries were from 35 PMBCLs, 31 DLBCLs, one FL and the cHL-derived cell line L1236. The results for the PDL loci *CD274* and *PDCD1LG2* have been previously published [214] and are not discussed herein.

At first, quality assessment was performed to inform on the suitability of the sequencing data for further analysis.[9] The mean coverage of each region captured represents the average sequencing depth (at a per-base basis) across all bases in the respective region. PCR duplicates were removed prior to the analysis. The mean coverage was uniform between the targeted regions and exceeded 600x for the *CIITA* and *SOCS1* locus, respectively. However, a per-case analysis revealed a moderate degree of variability between individual specimens (Figure 2.19).

[9] Bioinformatic analyses were performed by Lauren Chong.

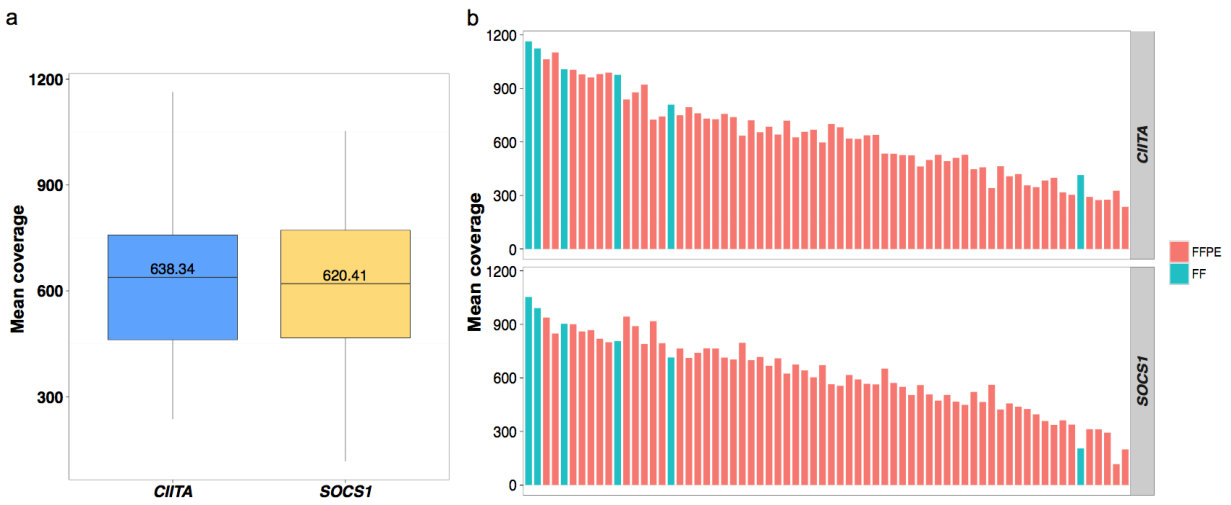

Figure 2.19: Oligocapture target region coverage depth. a) Mean coverage for the two target regions located on the short arm of chromosome 16, *CIITA* and *SOCS1*. b) Mean coverage for *CIITA* (upper panel) and *SOCS1* (lower panel) on a per library basis, each column on the x-axis represents a successfully constructed library. Abbreviations: FFPE; formalin-fixed, paraffin-embedded; FF, fresh-frozen.

Structural variant detection was performed by integrating results from various prediction tools and different trimmed read lengths as described [214].[10] Ultimately, the DELLY and deStruct tools were used because of their high concordance, comprehensive output format and high sensitivity. For the genomic region on chromosome 16, that was included in our capture design, the two SV detection tools resulted in a list of 145 predictions. To further narrow down the list to only include high-confidence SVs, we applied the following filter criteria: 1) read support; a cut-off of $\geq$ 10 reads was applied; and 2) redundancy; predictions had to be obtained in at least two data sets. This reduced the list to 75 unique events, which are listed in detail in B.5. When we restricted our analysis to those predictions with effects on the *CIITA* gene locus, 53 events were identified. Table 2.6 provides a comprehensive overview on exact chromosomal breakpoint locations, SV type and functional impact for all *CIITA*-affecting predictions (including reciprocal events).

[10] These analyses were performed by Lauren Chong and are in detail explained in her MSc thesis. The methodology is described in sufficient detail in the supplement of [214].

Table 2.6: SV predictions for *CIITA* obtained by oligocapture sequencing.

Library ID	SV type	chr1	position1	gene1	chr2	position2	gene2	max spanning reads	num events	interpretation of *CIITA* oligocapture results	Val
A43030	DEL	16	10947070	*CIITA*	16	11116222	*CLEC16A*	38	4	169 kb deletion, breakpoints upstream of *CIITA* and in intron 11 of *CLEC16A,* complete loss of this *CIITA* allele	Y
A43031	INV	16	11013706	*CIITA*	16	11033187	*DEXI*	149	8	19 kb inversion, breakpoints in intron 16 of *CIITA* and intron 1 of *DEXI*, fuses *CIITA* exon 1-16 to *DEXI* exon 2, the latter is non-coding but a putative fusion protein can be derived (B.6)	Y
	INV	16	11013709	*CIITA*	16	11033083	*DEXI*	110	7	reciprocal event, stop codon in exon 1 of *DEXI*, creates no fusion with *CIITA* exon 17-19	
A43036	INV	16	10972119	*CIITA*	16	11349103	*SOCS1*	144	6	377 kb inversion, breakpoints in intron 1 of *CIITA* and exon 2 of *SOCS1*, likely disruptive for both genes	Y
	INV	16	10972127	*CIITA*	16	11349114	*SOCS1*	65	4	reciprocal event	
	TRA	16	10972522	*CIITA*	X	41548791	*GPR34*	68	6	t(X;16), breakpoints in *CIITA* intron 1 and *GPR34* intron 1, putative fusion of *CIITA* exon 1 to *GPR34* exon 2, 51 aa protein predicted (B.7), likely non-functional	Y
	TRA	16	10972530	*CIITA*	X	41548793	*GPR34*	70	2	reciprocal event, t(X;16), small ORFs, non-functional	
	DEL	16	10972770	*CIITA*	16	10973118	*CIITA*	112	4	347 bp deletion *CIITA* intron 1	N

Library ID	SV type	chr1	position1	gene1	chr2	position2	gene2	max spanning reads	num events	interpretation of *CIITA* oligocapture results	Val
A43037	DEL	16	10973706	*CIITA*	16	10972948	*CIITA*	252	6	757 bp deletion intron 1 *CIITA*	Y
A43043	DEL	16	10972316	*CIITA*	16	10972128	*CIITA*	94	4	187 bp deletion *CIITA* intron 1	Y
A43049	INV	16	3056935	*CLDN6*	16	10966595	*CIITA*	21	8	8 Mb inversion, breakpoints upstream of *CLDN6* and upstream of *CIITA*, results in dislocation of the green BAC probe but no structural damage to *CIITA*	Y*
A43050	DEL	16	10972395	*CIITA*	16	10973044	*CIITA*	108	6	648 bp deletion *CIITA* intron 1, FISH not explained but apparently small clone (5 %)	
A43051	DEL	16	7638085	*RBFOX1*	16	10972040	*CIITA*	19	5	3.3 Mb deletion, breakpoints in intron 4 of *RBFOX1* and intron 1 of *CIITA*, leads to fusion of *RFBOX1* exon 4 to exon 2 of *CIITA* (B.8), resulting in a truncated protein	
A43052	TRA	2	61108467	*REL*	16	10974031	*CIITA*	21	8	t(2;16), breakpoints in *CIITA* intron 1 and upstream of *REL*, fusion transcript possible (*CIITA* exon 1 to *REL* exon 2)	Y
	TRA	2	61108477	*REL*	16	10974001	*CIITA*	13	7	reciprocal event	
	TRA	2	89159665	*IGK*	16	10972714	*CIITA*	15	6	t(2;16), breakpoints in *CIITA* intron 1 and centromeric on chromosome 2, close to *IGK* gene region	Y
	DEL	16	10973286	*CIITA*	16	10972769	*CIITA*	13	2	516 bp deletion *CIITA* intron 1	N*

Library ID	SV type	chr1	position1	gene1	chr2	position2	gene2	max spanning reads	num events	interpretation of *CIITA* oligocapture results	Val
A43067	INV	16	10973601	*CIITA*	16	27326617	*IL4R*	39	6	16 Mb inversion, breakpoints in *CIITA* intron 1 and *IL4R* intron 1, same strand direction, therefore no fusion transcript upon inversion, likely disruptive	N*
	INV	16	10973610	*CIITA*	16	27326640	*IL4R*	16	6	reciprocal event	
A43068	DEL	16	10962704	*CIITA*	16	11310352	*CLEC16A*	54	8	348 kb deletion, deletes *CIITA*, *DEXI* and *CLEC16A* entirely	N*
A43069	TRA	16	10973113	*CIITA*	22	39854860	*MGAT3*	55	6	t(16;22), breakpoints in intron 1 of *CIITA* and intron 1 of *MGAT3,* different chromosome arms, same strand direction, therefore no fusion, likely disruptive	Y
	TRA	16	10973113	*CIITA*	22	39854856	*MGAT3*	40	6	reciprocal event	
A43070	DEL	16	10983031	*CIITA*	16	11812699	*TXNDC11*	41	8	830 kb deletion, breakpoints in *CIITA* intron 1 and *TXNDC11* intron 5, different strand directions, therefore no fusion transcript	Y
	TRA	1	2985148	*PRDM16*	16	10972750	*CIITA*	102	5	t(1:16), breakpoints in *CIITA* exon 1 and upstream of *PRDM16*, results in fusion of *CIITA* exon 1 to *PRDM16* exon 2	Y
	TRA	1	2984655	*PRDM16*	16	10972919	*CIITA*	56	8	reciprocal event translocation *CIITA* intron 1 and upstream of *PRDM16*, no promoter swap, disruption of *CIITA* allele	
A43071	DEL	16	10756488	*TEKT5*	16	11339356	*SOCS1*	21	7	417 kb deletion, deletes *CIITA* allele	

Library ID	SV type	chr1	position1	gene1	chr2	position2	gene2	max spanning reads	num events	interpretation of *CIITA* oligocapture results	Val
A43072	DEL	16	10861924	*NUBP1*	16	10996431	*CIITA*	73	8	134 kb deletion, breakpoints in *NUBP1* intron 9 and *CIITA* intron 1, creates in-frame fusion transcript *NUBP1* exon 9 - *CIITA* exon 8	Y
A43075	TRA	16	8762984	*AICDA*	12	10973366	*CIITA*	48	6	t(12;16), breakpoints in *CIITA* intron1 and *AICDA* intron 1, same chromosome arms, different strand direction, therefore no fusion transcript	Y
	TRA	12	8764607	*AICDA*	16	10973178	*CIITA*	133	5	reciprocal event	
A43076	DEL	16	10982313	*CIITA*	16	12374408	*SNX29*	20	6	1.4 Mb deletion, breakpoints in *CIITA* intron 1 and *SNX29* intron 15, putative 29 aa fusion transcript (B.9)	Y
	DUP	16	10972350	*CIITA*	16	10972662	*CIITA*	45	4	300 bp duplication intron 1 *CIITA*	
A43077	DEL	16	10972806	*CIITA*	16	12062507	*TNFRSF17*	81	6	1 Mb deletion, breakpoints in *CIITA* intron 1 and downstream of *TNFRSF17*	Y
A43078	TRA	8	128808741	*PVT1*	16	10972594	*CIITA*	90	4	t(8;16), breakpoints in *CIITA* intron 1 and *PVT1* intron 1, different chromosome arms, same strand direction, therefore no fusion transcript	Y
	TRA	8	128808716	*PVT1*	16	10972569	*CIITA*	39	3	reciprocal event	
	DEL	16	10972600	*CIITA*	16	10972919	*CIITA*	55	2	318 bp deletion *CIITA* intron 1	Y

Library ID	SV type	chr1	position1	gene1	chr2	position2	gene2	max spanning reads	num events	interpretation of *CIITA* oligocapture results	Val
A43079	TRA	10	46794179	*CTSL1P5*	16	10972823	*CIITA*	18	2	t(10;16)	N*
	TRA	10	48989246	*GLUD1P7*	16	10972800	*CIITA*	146	7	t(10;16)	
	DEL	16	10973892	*CIITA*	16	10973408	*CIITA*	189	6	483 bp deletion intron 1 *CIITA*	N
A43080	DEL	16	11348838	*SOCS1*	16	10973122	*CIITA*	195	8	375 kb deletion, breakpoints in intron 1 *CIITA* and *SOCS1* exon 2	Y
	DEL	16	10972446	*CIITA*	16	10971699	*CIITA*	106	8	752 bp deletion intron 1 *CIITA*	Y
A43081	DEL	16	10973504	*CIITA*	16	10971906	*CIITA*	21	5	401 bp deletion intron 1 *CIITA*	
	DEL	16	10972143	*CIITA*	16	10971940	*CIITA*	12	4	202 bp deletion intron 1 *CIITA*	
	INV	16	10972733	*CIITA*	16	10973111	*CIITA*	17	3	367 bp inversion *CIITA* intron 1	
	INV	16	10973005	*CIITA*	16	10972714	*CIITA*	18	3	reciprocal event	
A43082	DEL	16	10972338	*CIITA*	16	10972492	*CIITA*	28	2	153 bp deletion intron 1 *CIITA*	
	DEL	16	10972737	*CIITA*	16	10972561	*CIITA*	29	2	175 bp deletion intron 1 *CIITA*	
	DEL	16	10972598	*CIITA*	16	10972316	*CIITA*	20	2	281 bp deletion intron 1 *CIITA*	
A43084	DEL	16	10972598	*CIITA*	16	10972316	*CIITA*	97	2	281 bp deletion intron 1 *CIITA*	
A43093	DEL	16	10975142	*CIITA*	16	10973558	*CIITA*	80	8	1.5 kb deletion *CIITA* intron 1	
	INV	16	10971536	*CIITA*	16	10972727	*CIITA*	90	6	1.2 kb inversion *CIITA* intron 1	
	INV	16	10971870	*CIITA*	16	10972644	*CIITA*	83	3	773 bp inversion *CIITA* intron 1	
	DUP	16	10972945	*CIITA*	16	10973145	*CIITA*	35	2	199 bp duplication *CIITA* intron 1	

Library ID	SV type	chr1	position1	gene1	chr2	position2	gene2	max spanning reads	num events	interpretation of *CIITA* oligocapture results	Val
A43095	DEL	16	10974070	*CIITA*	16	10972127	*CIITA*	20	6	2 kb deletion *CIITA* intron 1	N*
	DEL	16	10972312	*CIITA*	16	10972015	*CIITA*	76	5	296 bp deletion *CIITA* intron 1	
	INV	16	10973186	*CIITA*	16	10973372	*CIITA*	66	3	185 bp inversion *CIITA* intron 1	Y
	INV	16	10973169	*CIITA*	16	10973376	*CIITA*	37	3	reciprocal event	
	DEL	16	10974698	*CIITA*	16	10974425	*CIITA*	83	2	262 bp deletion *CIITA* intron 1	
A43097	DEL	16	10971998	*CIITA*	16	10971586	*CIITA*	138	5	411 bp deletion *CIITA* intron 1	Y
A43101	INV	16	10971827	*CIITA*	16	10973681	*CIITA*	52	6	1.8 kb inversion *CIITA* intron 1	
	INV	16	10971821	*CIITA*	16	10973687	*CIITA*	36	5	reciprocal event	
A43110	DEL	16	11143548	*CLEC16A*	16	10810673	*NUBP1*	63	7	333 kb deletion of *NUBP1, CIITA, DEXI* and parts of *CLEC16A*	Y
A43115	DEL	16	10973158	*CIITA*	16	10972963	*CIITA*	141	5	194 bp deletion intron 1 *CIITA*	Y
	INV	16	10971282	*CIITA*	16	10971688	*CIITA*	251	3	405 bp inversion in *CIITA* intron 1	N
	INV	16	10971301	*CIITA*	16	10971709	*CIITA*	236	3	reciprocal event	
	TRA	7	128309229	*FAM71F2*	16	10980174	*CIITA*	15	2	t(7:16) breakpoints in *CIITA* intron 1 and upstream of *FAM71F2*, different chromosome arms, same reading direction, no fusion	N*
A43117	DEL	16	10973536	*CIITA*	16	10972620	*CIITA*	349	6	915 bp deletion *CIITA* intron 1	Y

Abbreviations: aa, amino acid; BAC, bacterial artificial chromosome; chr, chromosome; N, not validated; N*, not validated because of limited material or PCR failure; num, number; ORF, open reading frame; SV, structural variant; Val, validation; Y, validated; Y*, validated, but different mapping

Ten translocation events (in eight specimens) were identified, in nine of these, the breakpoint in *CIITA* mapped to the 1.5 kb hotspot region within intron 1, where previously identified breakpoints are located [172,232]. In concordance with our previous observations of a high frequency of large scale and small deletions in that region we found 31 such events in our oligocapture cohort. In addition, 10 inversions and two duplications were detected. A detailed analysis of the translocation events revealed that *CIITA* rearrangements involve multiple different partner genes. Most of these translocations do not result in the generation of fusion transcripts because of strand orientation and/or intergenic location of the second breakpoint. However, since most of the breakpoints in *CIITA* are located within the first 3 kb of *CIITA* intron 1, the majority result in functional abrogation of this gene. An overview of the translocation events and the putative partner genes is provided in the circos plot in Figure 2.20.

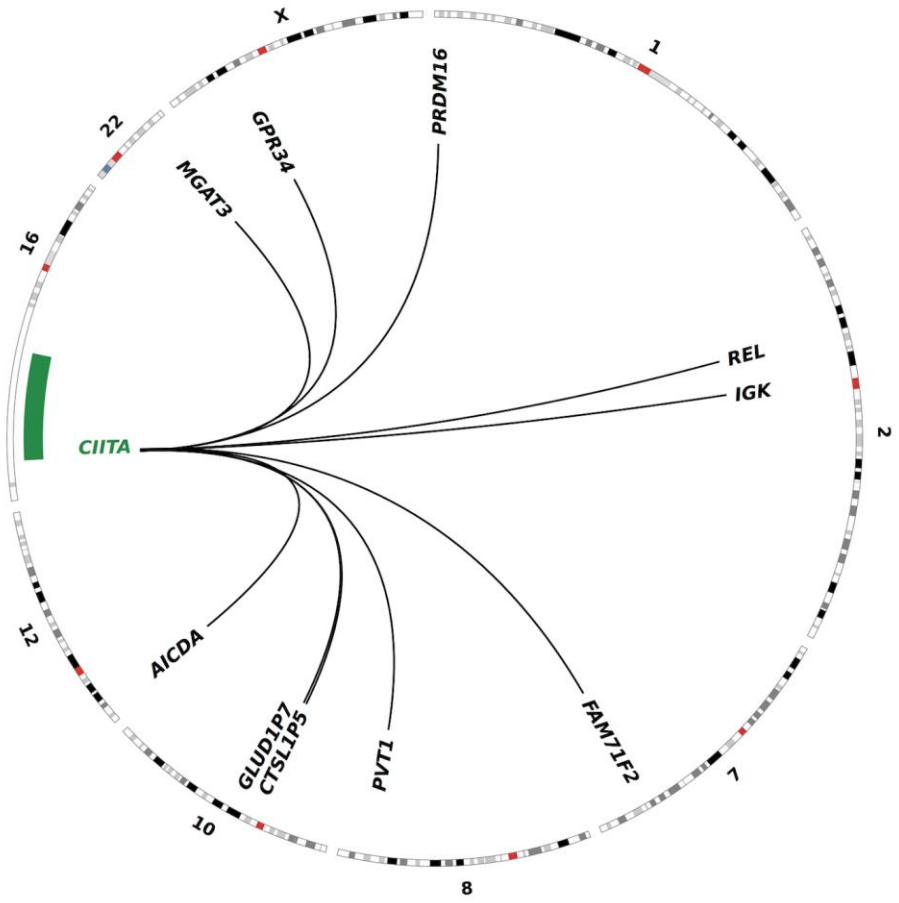

Figure 2.20: Circos plot depicting *CIITA* translocation events.For visualization purposes, only chromosomal regions involved in the predicted translocations are shown and the *CIITA* region is enlarged.

Beside the frequently occurring small deletions in *CIITA* intron 1 we also discovered eight large scale deletions. Of those, four resulted in the deletion of the entire *CIITA* gene locus whereas the remainder led to only a partial deletion. However, since the partial deletions involved either exon 1 or a large proportion of the gene, most of these events would lead to functional abrogation of CIITA. A schematic overview of these events is provided in Figure 2.21.

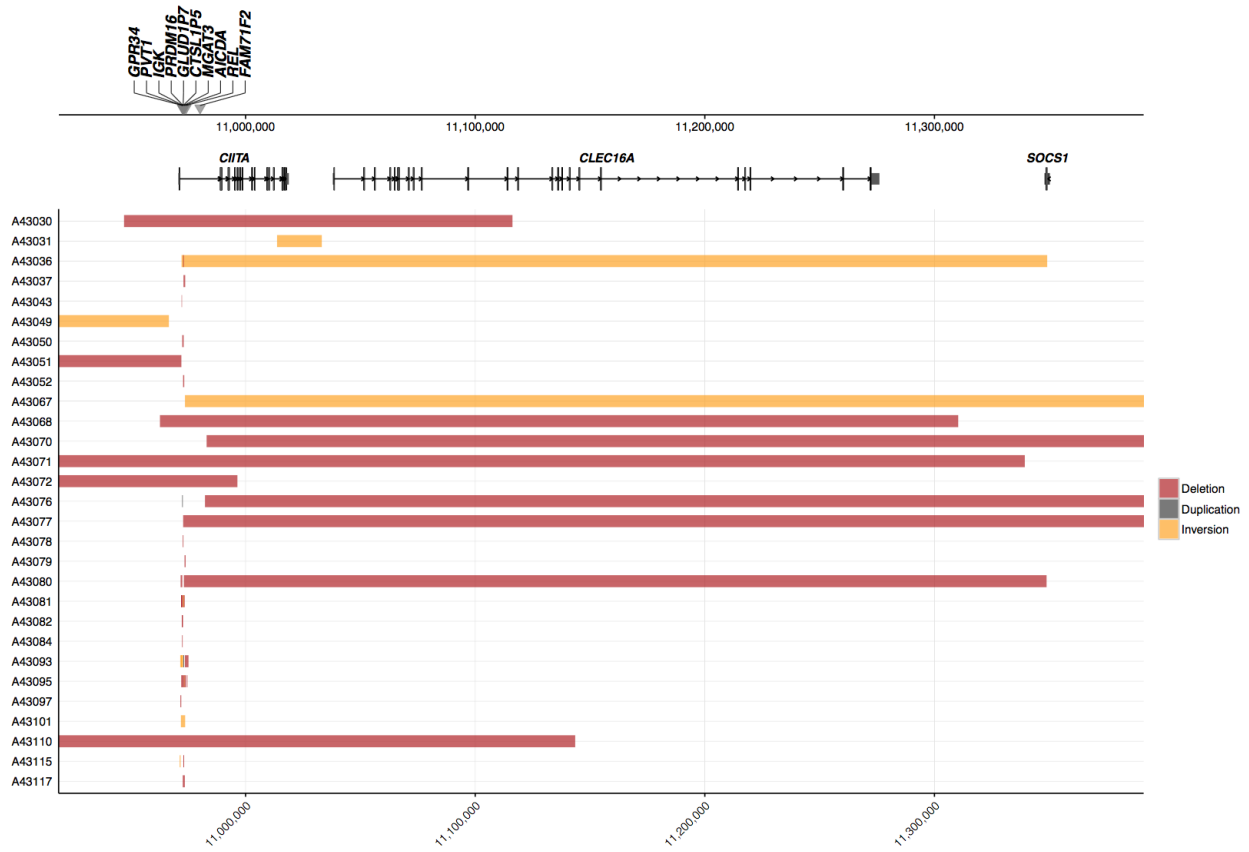

Figure 2.21: Intrachromosomal rearrangements in the chromosome 16 capture space. Only intrachromosomal events involving *CIITA* are depicted. The upper track shows the genomic coordinates according to hg19, as well as the location of the translocation breakpoints and the rearrangement partner genes. The panel underneath provides a schematic of the exon/intron structure of the three main genes included in the chromosome 16 capture space.

Of importance, the vast majority of intrachromosal structural variants involved the cluster breakpoint region in *CIITA* intron 1, which is confined to the first 3 kb of intron 1 and contains the inducible promoter IV with its alternative exon 1 (Figure 2.22).

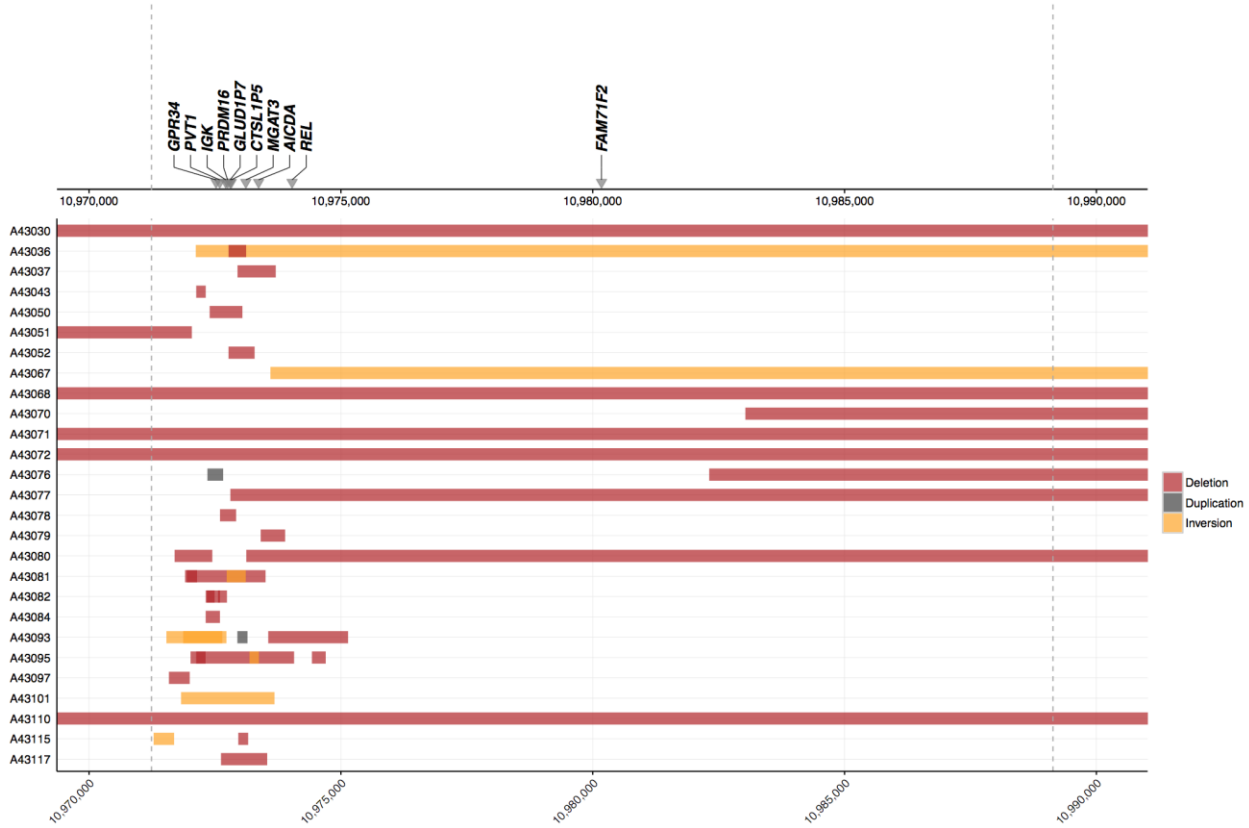

Figure 2.22: Intrachromosomal rearrangements in *CIITA* intron 1. The upper track shows the genomic coordinates according to hg19, as well as the location of the translocation breakpoints and the rearrangement partner genes. The boundaries of *CIITA* intron 1 are indicated with vertical dashed lines.

2.4 Discussion

Here, we performed a comprehensive analysis of genomic alterations in *CIITA* using NGS approaches and FISH across a spectrum of different B cell lymphoma subtypes. We established PMBCL as the entity with the highest frequency of *CIITA* alterations, including novel somatic CDS mutations, as well as deletions and rearrangements and, thereby, we substantially expanded on our previous findings of recurrent *CIITA* rearrangements in this disease [172]. Overall, CDS mutations together with structural alterations of the *CIITA* locus, encompassing deletions and rearrangements, were found in 71 % of PMBCL cases, thus establishing *CIITA* as one of the most frequently altered genes in this lymphoma entity [101]. The pattern of *CIITA* CDS aberrations, predominantly consisting of nonsense and frameshift mutations, as well as small deletions and chromosomal breakpoints, within a hot spot region in the first exon and intron, is indicative of loss of function. Moreover, our findings in PMBCL-

derived cell lines and in a proportion of primary clinical cases establish the pattern of *CIITA* biallelic hits.

The pathogenic contribution of the clustered intron 1 deletions and point mutations still needs to be determined. The majority might be co-selected passenger mutations or more likely the result of AID-mediated aberrant SHM. However, as many of these alterations occur within the pIV region and frequently target transcription factor binding motifs and responsive elements [196] or even entirely delete this promoter and/or its associated alternative first exon, they might be of pathogenic relevance. It has been recently shown in splenic dendritic cells that distal regulatory elements - regions involved in the regulation of *CIITA* transcription - are able to switch to and interact with other downstream promoters if the pI promoter (from where *CIITA* is expressed in dendritic cells) is not accessible [233]. This promoter flexibility could be a potential "rescue mechanism" for B cells to maintain high expression levels of CIITA mRNA despite mutation-associated loss of pIII-driven transcripts. Thus, additional functional abrogation of pIV could be a selective advantage for malignant B cells.

CIITA CDS mutations were less frequently observed in our cohort of *de novo* DLBCL (8.9 %) compared to PMBCL, but the frequency matches the range observed in previous studies [60,61,223,234]. Interestingly, we did not observe an association with a particular COO-subtype. Unfortunately, the design of the sequencing panel and the platform used did not allow the analysis of alterations in the non-coding space, in particular intron 1, in this cohort. However, similar to the rare observation of chromosomal rearrangements in DLBCL (with the exception of PTL), large scale deletions within the first 3 kb of intron 1 could not be detected in a small set of DLBCL cases analyzed by conventional PCR [172,217].

In our highly selected FL cohort, which was enriched for cases with subsequent transformation and progression, we discovered a high frequency of abnormal signal patterns in the FISH study with as much as 36.8 % abnormal cases among the transformed FL specimens (T2 specimens, mostly DLBCL morphology), followed by the respective T1 samples (29.9 %). Specimens within the pFL and npFL cohorts showed structural chromosomal alterations to a much lesser extent with 15.2 % and 5.4 %, respectively, indicating that these alterations might be a surrogate for an imminent

transformation event, when detectable at initial diagnosis. Of note, CDS mutations were not enriched among the FLs which eventually transformed. The analysis of intron 1 mutations in FL showed frequent SNVs and rarely small deletions or insertions.

The clustering of SNVs, deletions and chromosomal breakpoints in the first exon and intron, as well as a significant association with AID target motifs, raises the likelihood that *CIITA* is aberrantly targeted by the AID-mediated SHM process. As it has been shown in previous studies for a variety of genes characteristically expressed in germinal center B cells (e.g. *BCL6*, *MYC*, *PAX5*, *PIM1*, *RHOH, CD95* and *SOCS1)*, mutations in these loci can occur as a side-product of the B cell maturation and antibody diversification process and might play a crucial role in the development of malignant lymphomas of (post)germinal-center B cell origin [41–43,235–237]. Following these studies, genome-wide screens and targeted sequencing approaches in mice and humans have identified a large number of other genes affected by aberrant SHM in both non-malignant and malignant contexts [46,95,98,234,238]. As a result, *CIITA* is considered to be a target of aberrant SHM in DLBCL and FL. We show in our study that AID protein expression is detectable in a substantial fraction of PMBCL tumors. This is in line with previous reports in the literature where the authors observed AID mRNA and protein expression but were limited by analyzing only a small number of tissue samples [239,240].

By using BAC capture and oligocapture sequencing approaches, we were able to provide a comprehensive characterization of breakpoint anatomy and rearrangement partner genes. We further substantiated our notion of a cluster breakpoint region in intron 1 of *CIITA*, identified novel translocations and revealed that *CIITA* rearrangements involve multiple different partner genes. In addition, we identified frequent intrachromosomal SV, most of them resulting in functional abrogation of CIITA.

Chapter 3: Functional and Clinical Relevance of *CIITA* Alterations

3.1 Introduction

Over 20 years ago, *CIITA* was identified as the master transcriptional regulator of MHC class II expression based on studies conducted in patients with a severe immunodeficiency syndrome, namely type II bare lymphocyte syndrome (BLS) [189]. Individuals affected by this disease harbour somatic mutations which ultimately result in the loss of MHC class II expression and as an ultimate consequence patients suffer from multiple infections due to an inadequate adaptive immune response. Further studies uncovered the complex regulation of MHC class II expression, involving several co-factors assembled in a multi-protein complex known as the MHC enhanceosome [188,241]. Expression of CIITA is driven by different promoters which regulate transcription in a cell-type and context-specific manner. Four promoters have been described in the literature and the respective CIITA transcripts are characterized by a different first exon [196]. Of importance for the regulation of *CIITA* transcription in humans are: 1) pI, the promoter used by dendritic cells; 2) pIII, the dominant driver of CIITA-expression in B cells; and 3) pIV, the INF-γ inducible promoter. Significant amounts of pII-derived transcripts could not be detected in human tissues or cell lines, hence the biological relevance of this promoter seems to be rather limited [196].

MHC are glycoproteins on the cell surface and are integral and indispensable for the normal function of the adaptive immune system. MHC class II is predominantly involved in the display of exogenous antigens to $CD4^+$ T cells and expression is mainly restricted to APCs [122]. B cells are potent antigen presenters and are, therefore, characterized by a rather constitutive expression of MHC class II, which can be enhanced by stimulatory signals, such as IL-4 [176,177]. However, the expression of MHC class II within the B cell lineage is restricted and strictly regulated. It is absent in early pro-B cells and then increases in subsequent developmental stages. The expression maximum is reached in germinal centre B cells, whereas terminally differentiated plasma cells are largely negative [178,179].

The loss of surface expression of MHC class II molecules has been described across different lymphoma entities and has been linked to reduced immunogenicity of

tumour cells. While some studies clearly established an association of MHC class II loss with inferior survival in cHL, PMBCL and DLBCL [133,134,215,242,243], others could not demonstrate an impact of MHC class II deficiency on patient outcomes [64,244].

Here we thought to assess the impact of somatic *CIITA* alterations, as identified in Chapter 2, on MHC class II expression in *in vitro* cell line models as well as primary tissue specimens. Furthermore, we aimed at linking *CIITA* mutations and MHC class II loss of expression with the abundance of T cell subsets in the tumour microenvironment and patient outcomes.

3.2 Materials and methods

3.2.1 Flow cytometry

Flow cytometry was used to quantify HLA-DR surface expression in cell lines. Briefly, 0.5 x 10^6 cells were washed once in cold PBS containing 1 % FBS and then incubated for 30 minutes on ice with PerCP/Cy5.5 anti-human HLA-DR (BioLegend, 0.25 µg/test). PerCP/Cy5.5 Mouse IgG2a κ was used as an isotype control. After washing and resuspension, samples were acquired on a FACSCalibur (BD Biosciences). All analyses were done using FlowJo software (version 10.0.7r2, FlowJo, LLC).

3.2.2 Quantitative reverse transcriptase (qRT)-PCR

All qRT-PCR assays were performed on a CFX384 real-time PCR system (Bio-Rad). The efficiencies of target and reference amplifications were verified to be equal, therefore expression levels were quantified using the ΔΔCT method. A TaqMan® RNA-to-C_T™ 1-Step Kit and TaqMan® gene expression assays were used to detect mRNA expression levels of CIITA (Hs00932860_m1) and HLA-DR (Hs00734212_m1). GAPDH expression (GAPDH (DQ) Oligo Mix, Cat. # 4332649, Life Technologies) served as an internal control. Each specimen was assayed in triplicate using 1 ng of total RNA per 10 µl reaction and results are shown relative to the expression in either germinal centre B cells or cells transduced with the empty vector control.

CIITA pIII and pIV specific mRNA transcripts were assayed by a two-step SYBR Green assay (Power SYBR Green PCR Master Mix, Life Technologies). cDNA was

synthesized from 1 μg total RNA using random hexamers and SuperScript III Reverse Transcriptase (Life Technologies). cDNA equivalent to 5 ng RNA was used for each 10 μl reaction and expression levels were normalized to GAPDH.

3.2.3 Western blotting

Total cell lysates were prepared using RIPA lysis and extraction buffer (Thermo Scientific). Cultured cell lines were washed once with PBS and then incubated for 15 minutes on ice in RIPA buffer supplemented with a protease inhibitor cocktail (Sigma-Aldrich). Twenty-five μg of protein lysate was resolved on a 4-12 % NuPAGE Novex Bis-Tris gradient gel (Invitrogen) and transferred to a nitrocellulose membrane by wet transfer. Blots were probed with an anti-CIITA antibody (clone 7-1H, Abcam, ab49989) at a 1:1000 dilution. β-actin (A5441; Sigma-Aldrich) served as an internal loading control (dilution 1:10000). Horseradish peroxidase (HRP) conjugated anti-mouse IgG (W402B, Promega) was used to visualize bands using the enhanced chemiluminescence (ECL) system (Amersham) on a Chemidoc digital imager (Bio-Rad).

3.2.4 Retroviral transduction

Retroviral transductions were performed as previously described [245]. Briefly, the NLPHL-derived cell line DEV was at first transduced with a feline endogenous virus expressing the ecotropic retroviral receptor (mCAT, obtained as supernatant from the stable producer line FLYRD18/mCAT-IRES-Bleo; gift from Dr. Louis Staudt, NIH, Bethesda, MD). Secondly, following a 14-day selection period with Zeocin, cells were infected with an ecotropic retrovirus, expressing the bacterial tetracycline repressor (TetR). Cells expressing TetR were positively selected with Blasticidin over two weeks. Wildtype (wt) or mutant *CIITA* cDNAs were amplified using Phusion DNA polymerase (NEB) and subsequently cloned into the doxycycline-inducible pRetroCMV/TO/GFP vector. The mutant alleles discovered in clinical cases were generated from the vector containing the wt allele by using the QuikChange II XL site-directed mutagenesis kit (Agilent) according to the manufacturer's instructions. The *NUBP1-CIITA* fusion was cloned using the Gibson Assembly Master Mix (NEB). Inserts were verified using Sanger sequencing.

A mixture of *CIITA*-containing plasmid DNA, the mutant ecotropic envelope-expressing plasmid pHIT/EA6x3 and gag-pol expressing plasmid pHIT60 were co-transfected in HEK293T cells using the Lipofectamine 2000 Reagent (Invitrogen). Two days after transfection, retroviral supernatants were collected to infect the engineered target cells in the presence of 8 μg/ml polybrene (Sigma) in a single spin infection (90 minutes at 2500 rpm and 25ºC) and two days after transduction puromycin was added to select for stable integration over 4-6 days. Gene expression was induced by adding 20 ng/ml doxycycline (Sigma) to the media for 48 hours.

3.2.5 RNA-Seq[11]

RNA-Seq was performed as previously described [59] using RNA extracted from DEV cells expressing either the empty pRETRO vector, wt *CIITA* or the p.1131STOP>R*35 mutant. In brief, approximately two million cells were harvested after 48h treatment with 20 ng/ml doxycyclin and RNA was extracted using the RNeasy Mini Kit (Qiagen). RNA was then submitted to the Genome Sciences Centre for further processing, library construction and sequencing. Polyadenylated (polyA+) messenger RNA (mRNA) was purified, ethanol–precipitated, and used to synthesize cDNA using the Maxima H Minus First Strand cDNA Synthesis kit (Thermo-Fisher) and random hexamer primers. The second strand cDNA was synthesized following the Superscript cDNA synthesis protocol by replacing the dTTP with dUTP in the dNTP mix, allowing the second strand to be digested using UNG (Uracil-N-Glycosylase, Life Technologies) and thus achieving strand specificity. cDNA was fragmented by sonication using a Covaris LE220 (Covaris). Libraries with a desired size range were purified and diluted to 8 nM, and then pooled at five per lane and sequenced on the HiSeq 2500 platform, generating 75 bp paired-end reads. Sequencing files were aligned using the STAR aligner (version 2.5.2a) and differentially expressed genes were reported using DESeq2 (version 1.12.2) based on the following criteria: 1) genes with an absolute $\log_2$ fold change (FC) of ≥ 1.0 were considered as being differentially expressed; and 2) false discovery rate (FDR)-adjusted *P*-value (q-value) < .05 was considered significant (Benjamini-Hochberg method).

[11] This section is modified from a method draft provided by Dr. Andrew Mungall.

3.2.6 Immunohistochemistry

IHC was performed on 4 μm sections from FFPE tissue samples arranged on previously constructed TMAs [65,104,172,219] or on whole tissue sections in the case where a particular sample was not evaluable or represented on the TMA. Following antigen retrieval, sections were stained with primary antibodies recognizing HLA-DR/DP/DQ (dilution 1:1000, clone CR3/43, Dako, M0775), CD4 (neat, clone SP35, Ventana 790-4423) and CD8 (dilution 1:50, clone C8/144B, Dako, M710301) followed by routine protocols for automated IHC on the Ventana Benchmark XT (Ventana Medical Systems). MHC class II protein expression was assessed semi-quantitatively as follows: 0 = negative, 1 = membranous staining, 2 = cytoplasmic staining, and the predominant pattern based on evaluation of duplicate TMA cores was assigned to a case. Representative images were taken using a Nikon Eclipse 80i microscope equipped with a Nikon DS-Ri1 camera and NIS Elements imaging software (Version D3.10).

Immunohistochemically stained slides for the T cell markers CD4 and CD8 were scanned with an Aperio ScanScope XT at 20x magnification and analyzed utilizing the Aperio ImageScope viewer (Version 12.1.0; Aperio Technologies). Only cores and areas containing tumour were scored using the Positive Pixel Count algorithm with optimized color saturation thresholds. Any staining was considered positive and the number of positive pixels was divided by the total pixel count. Whenever applicable, scores from both cores were averaged and multiplied by 100 to obtain the percentage of positive pixels.

3.2.7 Statistical and survival analysis

Fisher's exact test and χ^2 test were used to compare categorical variables. McNemar's test was used when comparing categorical variables between paired specimens. The Mann-Whitney U test was applied to compare two groups. Spearman correlation was used to test the association between two continuous variables.

Survival analysis were performed using the Kaplan-Meier method and differences between groups were assessed by applying the log-rank test. Overall survival (OS) was defined as time from diagnosis to death from any cause; disease-

specific survival (DSS) as time from diagnosis to death from lymphoma or acute treatment-related toxicity; progression-free survival (PFS) as time from diagnosis to disease progression/relapse or death from any cause and time to progression (TTP) as time from diagnosis to disease progression/relapse or death from lymphoma or acute treatment-related toxicity. Time to transformation (TTT) was calculated as the time between diagnosis of FL and histological confirmation of large cell transformation. Patients, who did not experience an event, were censored at the time of last follow-up. *P*-values < .05 were considered being significant.

3.3 Results

3.3.1 CIITA and HLA-DR expression in PMBCL- and NLPHL-derived cell lines

Based on our observations of frequent biallelic genomic alterations of *CIITA* in PMBCL-derived cell lines and in DEV and given the functional importance of *CIITA* in the transcriptional regulation of MHC class II genes, we analyzed the expression of CIITA and HLA-DR by qRT-PCR and flow cytometry and compared it to sorted germinal centre B cells and an EBV-transformed lymphoblastoid B cell line (LCL), respectively. In Karpas1106P and in DEV, CIITA mRNA expression was not detectable by qRT-PCR, as a consequence of biallelic rearrangements in these cell lines (see section 2.3.1.1 for details). CIITA expression levels were higher in U2940 compared to germinal centre B cells but markedly reduced in MedB-1 (Figure 3.1).

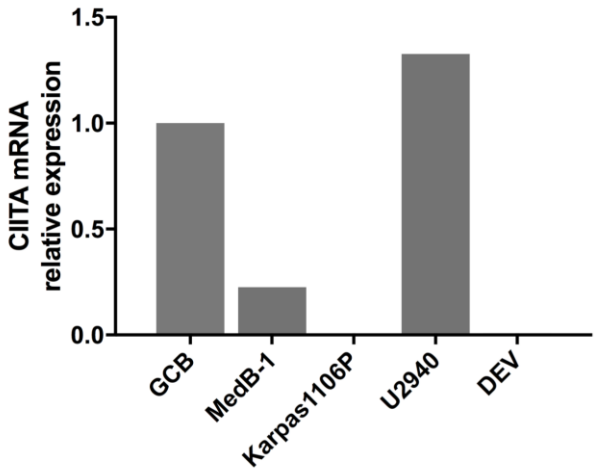

Figure 3.1: CIITA mRNA expression in PMBCL- and NLPHL-derived cell lines.
Shown is the relative expression of CIITA mRNA in four lymphoma cell lines compared to germinal centre B cells (GCB).

We subsequently analyzed HLA-DR protein expression by flow cytometry (Figure 3.2). Karpas1106P and DEV had no detectable surface HLA-DR expression, whereas U2940 showed an intermediate expression level compared to the LCL cell line used as a control. Interestingly, the third PMBCL cell line MedB-1 showed reduced, but remarkably heterogeneous HLA-DR expression.

To explore to which extend CIITA expression can be induced upon IFN γ stimulation in PMBCL cell lines we performed a transcript specific qRT-PCR. As expected, Karpas1106P showed no expression because of the biallelic deletion (see Figure 2.3). The stimulation with IFN γ resulted in a slight increase of the pIII transcript (1.4-fold for MedB-1 and 1.3-fold for U2940). Expression of the pIV transcript increased by 2.3-fold for MedB-1 and 5-fold for U2940. These data are in agreement with the published literature confirming a preferential effect of IFN γ on pIV in B cells [179]. However, no changes in HLA-DR expression were observed, which is expected, given the *CIITA* alterations in these cell lines.

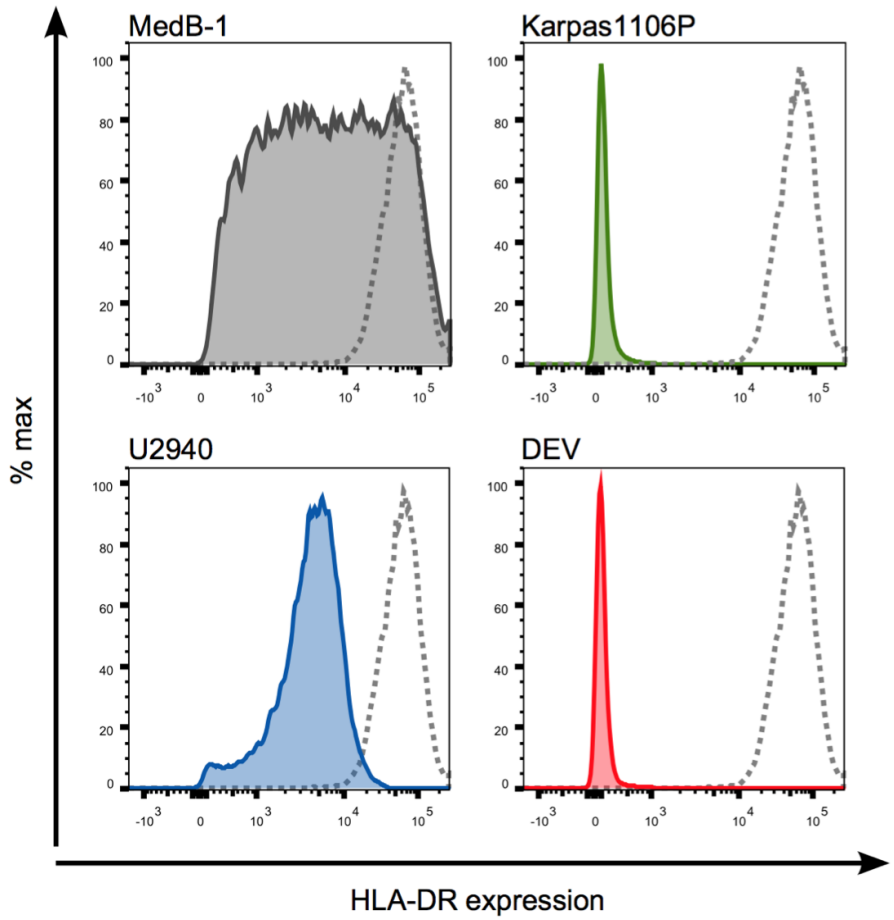

Figure 3.2: HLA-DR expression in PMBCL- and NLPHL-derived cell lines. HLA-DR expression in PMBCL- and NLPHL-derived cell lines was assessed by flow cytometry. Expression levels are compared to LCL, an EBV-transformed lymphoblastoid cell line, which is depicted by the dashed grey line.

3.3.2 Functional implications of *CIITA* mutants in *in vitro* cell line models

To study the functional consequences of *CIITA* mutations, we generated stably transduced cell lines expressing either wt or mutant *CIITA*. The NLPHL-derived cell line DEV was selected for retroviral transduction experiments because it completely lacks surface HLA-DR expression due to biallelic inactivation of *CIITA* (Figure 2.6) and proved to be transducable with the retroviral vector system as described in section 3.2.4.

The ectopic reintroduction of the *CIITA* wt allele into DEV was able to restore CIITA mRNA and protein expression, measured by qRT-PCR and Western blot, and resulted in HLA-DR surface expression by flow cytometry (Figure 3.3 a-g) comparable to levels observed in LCL (see Figure 3.2 for comparison). Next, we evaluated the

effects of several *CIITA* mutants, including the NUBP1-CIITA fusion, which we described in the PMBCL-derived cell lines U2940 and MedB-1 (Figure 2.2 and Figure 2.4), as well as two missense mutations observed in primary patient-derived cases. The hg19 chr16: g.11017158C>T mutation, resulting in a loss of the original stop codon p.1131STOP>R*35, showed CIITA mRNA expression comparable to the wt (Figure 3.3 a), but was not capable of re-establishing HLA-DR expression (Figure 3.3 d, e and g). The NUBP1-CIITA fusion only partially restored HLA-DR levels and therefore explains the intermediate expression levels seen in U2940 (Figure 3.2). Interestingly, the two different mutant alleles which occur in *trans*-configuration in MedB-1 showed approximately equal amounts of CIITA mRNA and protein (Figure 3.3 a and b), but revealed different effects on HLA-DR expression when independently transduced. Specifically, the p.71Asp>Asn mutation restored HLA-DR surface protein levels, whereas the p.636Thr>Met allele resulted in a broad range of expression similar to the parental cell line (Figure 3.3 d, e and Figure 3.2). The p.715Cys>Tyr mutant could only marginally restore HLA-DR expression, however, with the p.11Ser>Pro allele high levels of HLA-DR expression were observed (Figure 3.3 a-g). We noted that, despite similar mRNA expression levels for CIITA, protein expression was lower in cells transduced with wt *CIITA* or the p.1131STOP>R*35 mutant compared to others. We speculate that this may originate from differences in protein stability.

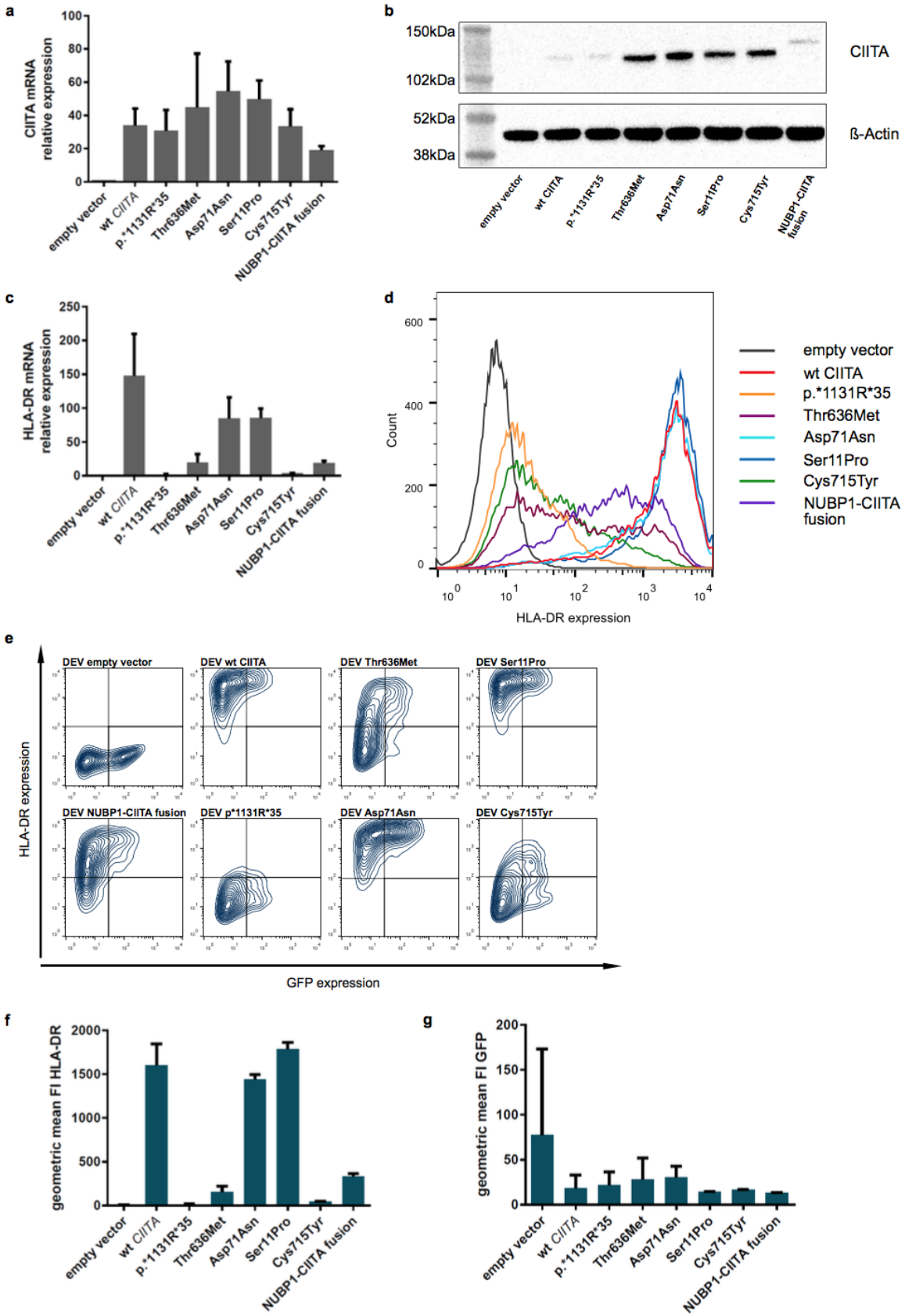
a
CIITA mRNA relative expression
100
80
60
40
20
0
empty vector
wt CIITA
p.*1131R*35
Thr636Met
Asp71Asn
Ser11Pro
Cys715Tyr
NUBP1-CIITA fusion
b
150kDa
102kDa
52kDa
38kDa
CIITA
ß-Actin
empty vector
wt CIITA
p.*1131R*35
Thr636Met
Asp71Asn
Ser11Pro
Cys715Tyr
NUBP1-CIITA fusion
c
HLA-DR mRNA relative expression
250
200
150
100
50
0
empty vector
wt CIITA
p.*1131R*35
Thr636Met
Asp71Asn
Ser11Pro
Cys715Tyr
NUBP1-CIITA fusion
d
Count
600
400
200
0
HLA-DR expression
empty vector
wt CIITA
p.*1131R*35
Thr636Met
Asp71Asn
Ser11Pro
Cys715Tyr
NUBP1-CIITA fusion
e
DEV empty vector
DEV wt CIITA
DEV Thr636Met
DEV Ser11Pro
DEV NUBP1-CIITA fusion
DEV p*1131R*35
DEV Asp71Asn
DEV Cys715Tyr
HLA-DR expression
GFP expression
f
geometric mean FI HLA-DR
2000
1500
1000
500
0
empty vector
wt CIITA
p.*1131R*35
Thr636Met
Asp71Asn
Ser11Pro
Cys715Tyr
NUBP1-CIITA fusion
g
geometric mean FI GFP
200
150
100
50
0
empty vector
wt CIITA
p.*1131R*35
Thr636Met
Asp71Asn
Ser11Pro
Cys715Tyr
NUBP1-CIITA fusion

Figure 3.3: Ectopic expression of *CIITA* wildtype and mutants in DEV. a) No significant differences in CIITA mRNA expression were observed upon retroviral transduction of wt and mutant *CIITA* (number of independent experiments, n = 3). b) CIITA protein expression is restored upon transduction with wt and mutant *CIITA.* As expected, the p.STOP1131R*35 mutant and the NUBP1-CIITA fusion protein have a slightly higher molecular weight compared to the wt protein. c, d) Wt *CIITA* restores HLA-DR mRNA (c; number of independent experiments, n = 3) and surface protein expression (d; representative data from 3 experiments) whereas the p.STOP1131R*35 mutant does not. The missense variants discovered in MedB-1 and two clinical samples have different impact on HLA-DR expression. e) Density plots for GFP and HLA-DR expression in transduced DEV cells. f, g) Geometric mean of the fluorescence intensity (FI) for HLA-DR (f; n = 3) and GFP (g; n = 3). Mean ± SD are shown for all bar graphs. Abbreviations: wt, wildtype.

3.3.3 RNA-Seq analysis

To further evaluate the functional impact of *CIITA* deficiency on a more global level, we performed RNA-Seq experiments to determine genome-wide changes in transcriptional activity. We compared DEV cells transduced with the empty retroviral vector backbone to cells transduced with the vector containing wt *CIITA*. Surprisingly, we found only 26 significant differentially expressed genes (Figure 3.4), of which 12 genes met our criteria (absolute $\log_2$ FC of ≥ 1.0, adjusted *P*-value < .05, Table 3.1).

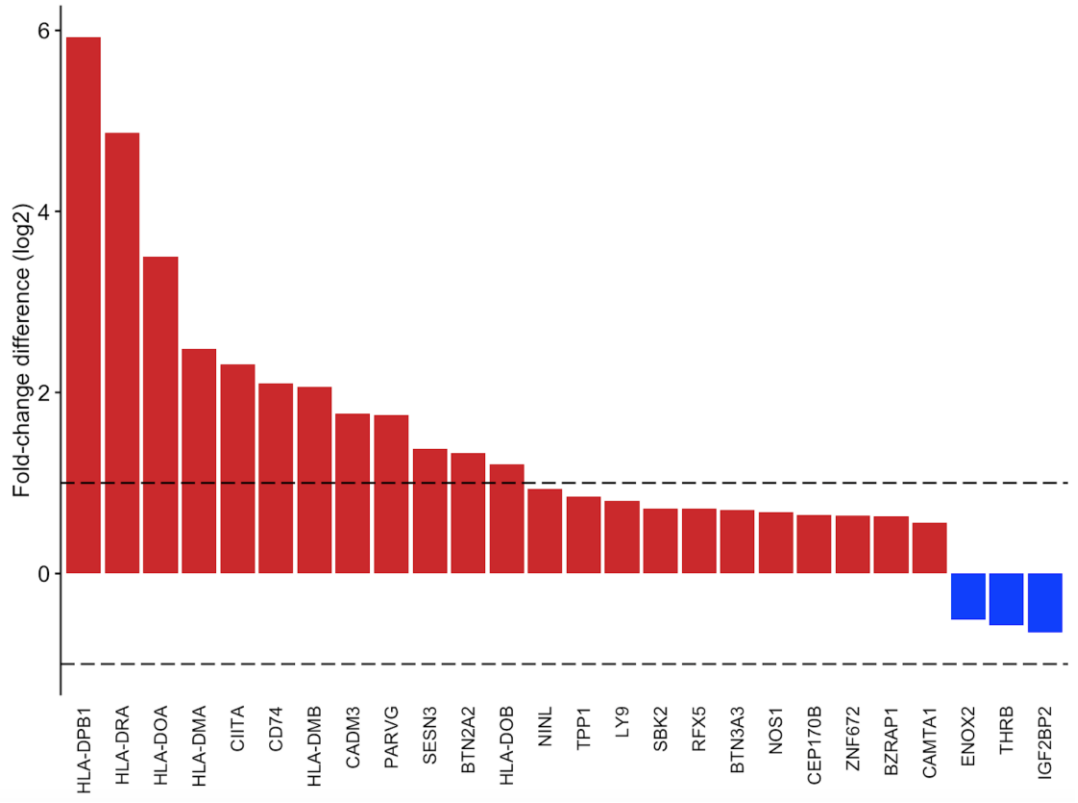

Figure 3.4: Top differentially expressed genes in DEV cells expressing wt *CIITA*. The waterfall plot shows all significant differentially expressed genes (FDR < .05). The dashed line indicates the threshold of an absolute $\log_2$ FC of ≥ 1.0.

Table 3.1: Differentially expressed genes.

gene_name	baseMean	log2FoldChange	pvalue	padj	gene annotation
HLA-DPB1	760.9832721	5.928838653	1.29E-207	1.93E-203	MHC class II
HLA-DRA	5930.725532	4.866696771	4.95E-138	3.72E-134	MHC class II
HLA-DOA	126.9200534	3.496263144	1.57E-72	7.84E-69	MHC class II
HLA-DMA	683.2489663	2.481786444	4.57E-58	1.72E-54	MHC class II
CD74	76631.21505	2.099627182	3.26E-56	9.80E-53	MHC class II (invariant chain)
HLA-DMB	201.4500995	2.064319728	1.36E-37	3.40E-34	MHC class II
CIITA	7354.75111	2.309386328	1.72E-35	3.69E-32	
CADM3	17.69970813	1.76395559	1.00E-27	1.88E-24	cell adhesion molecule 3, member of the nectin family, cytoplasmic region interacts with *EPB41L1*, *DLG3*, *MPP6* and *CASK*
PARVG	29.91164828	1.754111704	4.11E-23	6.86E-20	parvin gamma, focal adhesion molecule, plays a role in the regulation of cell adhesion and cytoskeleton organization
SESN3	117.1944716	1.375663471	1.33E-12	2.00E-09	Sestrin 3, member of the sestrin family of stress-induced proteins, reduces levels of intracellular ROS. The protein is required for normal regulation of blood glucose and plays a role in lipid storage.
BTN2A2	824.6334586	1.331249587	2.81E-12	3.84E-09	encodes a type I receptor glycoprotein involved in lipid, fatty-acid and sterol metabolism
HLA-DOB	39.35284458	1.204205428	2.67E-10	3.34E-07	MHC class II

Beside *CIITA* itself, there were seven more MHC class II complex genes, among those the classical MHC class II complex members, *HLA-DRA* and *HLA-DPB1*, as well as the invariant chain *CD74*, and the chaperones *HLA-DM* and *HLA-DO*.

Similar results were obtained when comparing DEV *CIITA* wt versus DEV cells transduced with the vector containing the STOP-lost mutant as discovered in U2940 (see section 2.3.1.1 and Figure 2.4), underscoring that this mutation indeed abrogates CIITA function.

3.3.4 Correlative studies in primary lymphoma cases

3.3.4.1 PMBCL

In order to inform on the functional implications of *CIITA* genomic alterations, we performed and evaluated MHC class II (HLA-DR/DP/DQ) protein expression by IHC on the aforementioned TMAs (2.2.1.2.1). Only cases with complete information on *CIITA* FISH ba-status and MHC class II protein expression (n = 103) were analyzed further, and mutational data (although only available for 44 cases from the sequencing cohort) were also included. Cases which were scored as ba-positive by FISH and/or harboured *CIITA* CDS mutations were significantly more often negative for MHC class II compared to wt cases (χ^2-test: P = .02). Figure 3.5 shows two representative cases from the sequencing cohort.

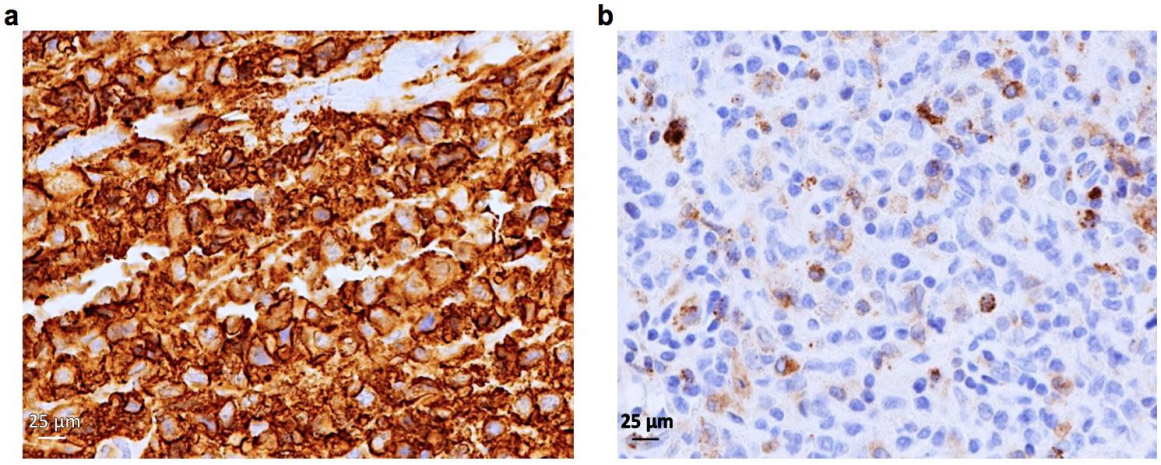

Figure 3.5: MHC class II expression in primary PMBCL cases. Shown are two representative PMBCL cases; a) case #7 and b) case #33 from the sequencing cohort. Case #7 shows strong membranous MHC class II expression in the tumour cell population, whereas in case #33 the neoplastic cells are negative. Small, reactive B cells and macrophages are stained as an internal control. Original magnification: x400.

To explore the potential impact of *CIITA* genetic aberrations or absent MHC class II surface expression on the microenvironment composition and in particular T cell subsets we compared the abundance of CD4- and CD8-positive T cells between the *CIITA* wt and mutated group, as well as between MHC class II positive and negative cases (Figure 3.6). No significant differences in $CD4^+$ and $CD8^+$ T cells were observed when comparing *CIITA* wt versus mutant/ba-positive cases (Mann-Whitney test: $P = .58$ and $P = .63$, respectively). When the same type of comparisons was performed with regard to MHC class II surface expression on the malignant cell population, a significantly higher amount of CD4-expressing T cells was present in MHC class II positive cases (Mann-Whitney test: $P = .008$). A trend was also seen in the CD8-positive subset, although this did not quite reach statistical significance (Mann-Whitney test: $P = .06$). For this cohort of cases the expression of MHC class II had been previously recorded as percentage of positive tumour cells in 10% increments and a histological score was generated, which also considered the staining intensity [232]. Based on these data we were able to perform a Spearman correlation and observed a weak to moderate correlation of MHC class II protein expression and the abundance of CD4- and CD8-positive T cells, respectively (Spearman $r = .224$, $P = .027$ for CD4; $r = .246$, $P = .015$ for CD8).

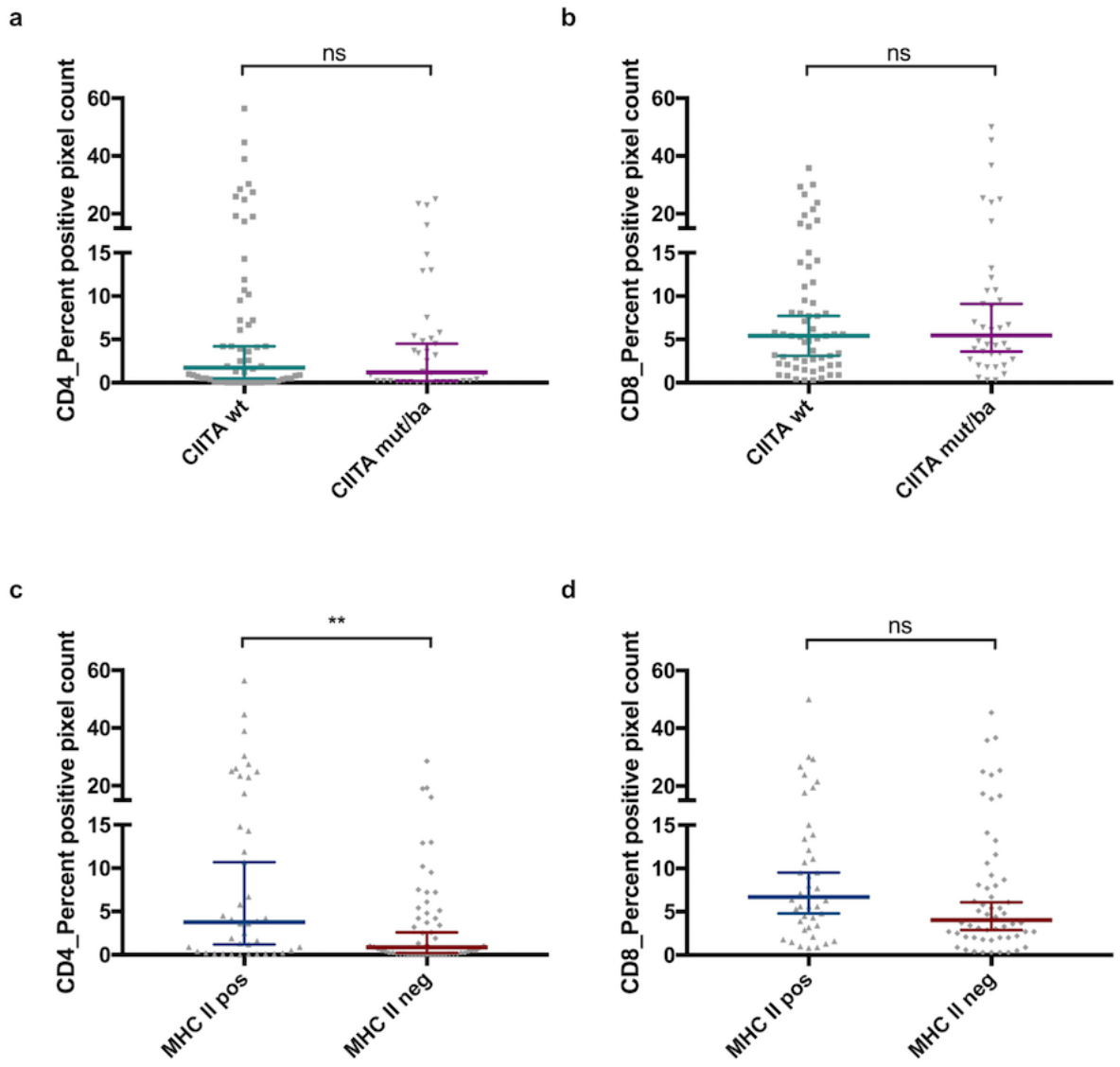

Figure 3.6: Abundance of T cell subsets in primary PMBCL cases. Scatter plots for the percentage of CD4- and CD8-positive pixels, respectively. Coloured horizontal lines show the median and the error bars represent the 95 % confidence interval. Abbreviations: ba, break-apart; mut, mutant; neg, negative; ns, not significant; pos, positive; wt, wildtype. *P*-values: * < .05; ** < .01; *** < .005.

Since it was previously reported, that *CIITA* genomic rearrangements were associated with impaired DSS in PMBCL [172], we next thought to perform outcome analyses in this slightly larger cohort and with the additional information obtained from the sequencing studies. No significant differences in survival (OS, DSS, PFS and TTP) were observed between *CIITA* wt and mutant cases (Figure 3.7). Furthermore, although the outcomes for patients with MHC class II negative tumours were less favourable compared to those with MHC class II positive lymphoma cells, significant differences

could not be observed (Figure 3.8). Similar results were seen when restricting the analyses to only include patients treated with rituximab-containing regimens.

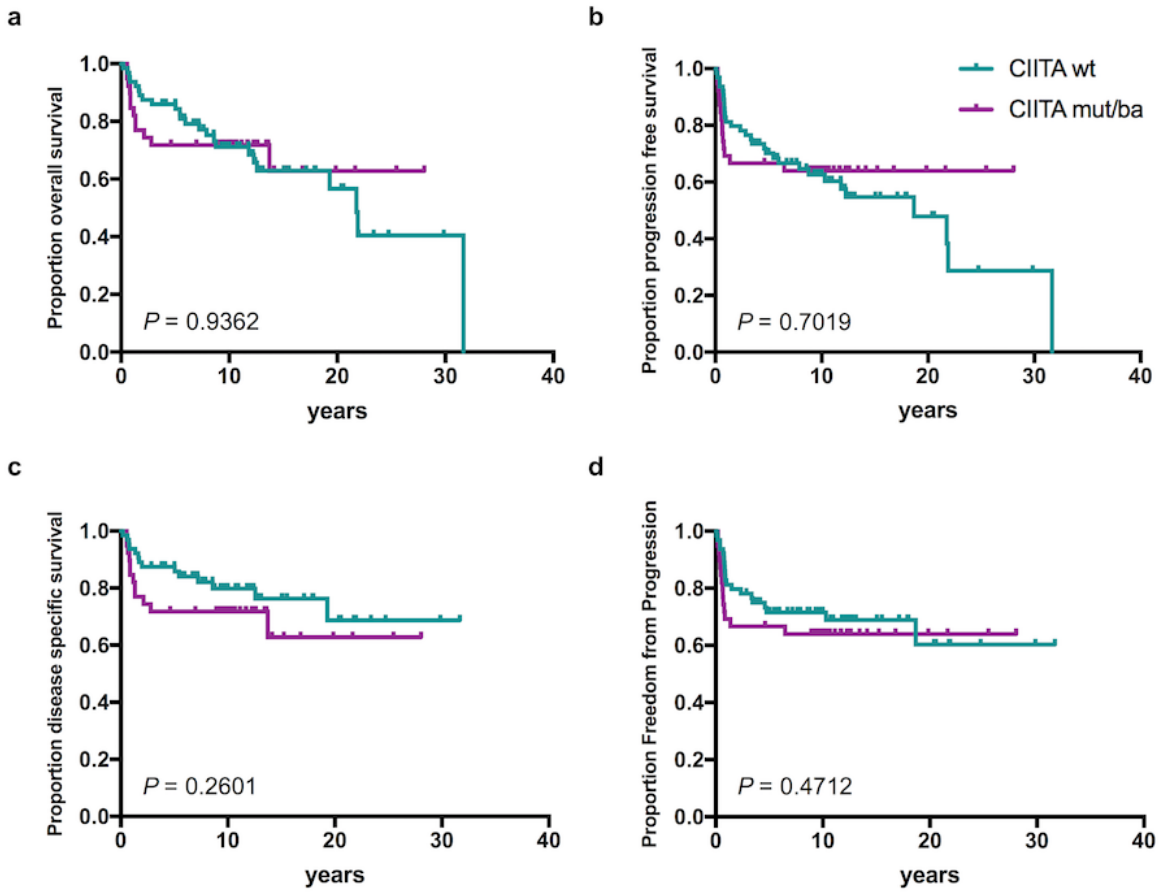

Figure 3.7: Survival of PMBCL patients with *CIITA* wt or mutant tumours. Shown are KM-plots for a) OS, b) PFS, c) DSS, and d) TTP.

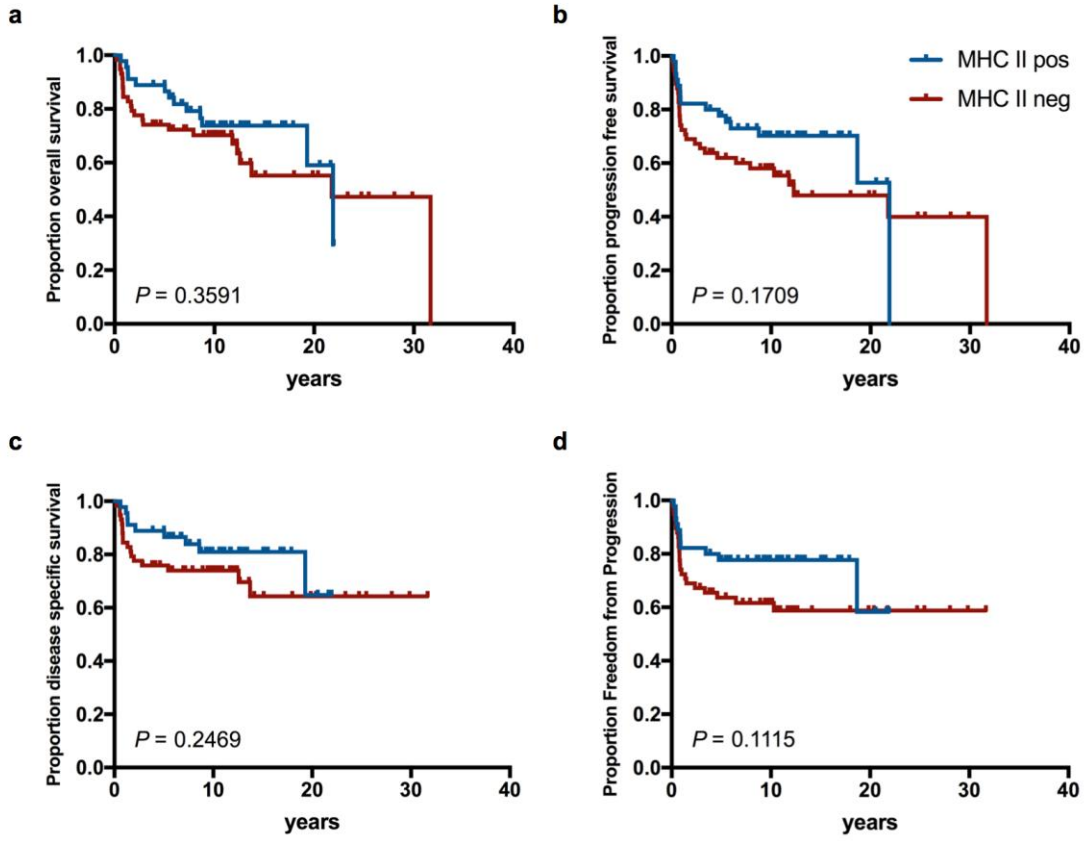

Figure 3.8: Survival of PMBCL patients according to MHC class II expression status. Shown are KM-plots for a) OS, b) PFS, c) DSS, and d) TTP.

3.3.4.2 DLBCL

Of 347 cases, 316 were evaluable for both *CIITA* mutational status and MHC class II expression. Cases harbouring *CIITA* CDS mutations were significantly more often negative for MHC class II compared to wt cases (χ^2-test: P = .002).

Next, we compared the abundance of CD4- and CD8-positive T cells between the *CIITA* wt and mutated group, as well as between MHC class II positive and negative cases (Figure 3.9). No significant differences in the abundance of $CD4^+$ and $CD8^+$ T cells, as inferred from the positive pixel count data, were observed when comparing *CIITA* wt versus mutant cases (Mann-Whitney test: P = .18 and P = .66, respectively; Figure 3.10). When allocating cases to groups based on MHC class II surface expression in the malignant cell population, a significantly higher amount of CD4-expressing T cells was present in MHC class II positive cases (Mann-Whitney test: P =

.03). However, the abundance of CD8-positive T lymphocytes was not significantly different between MHC class II positive and negative cases (Mann-Whitney test: *P* = .45; Figure 3.10).

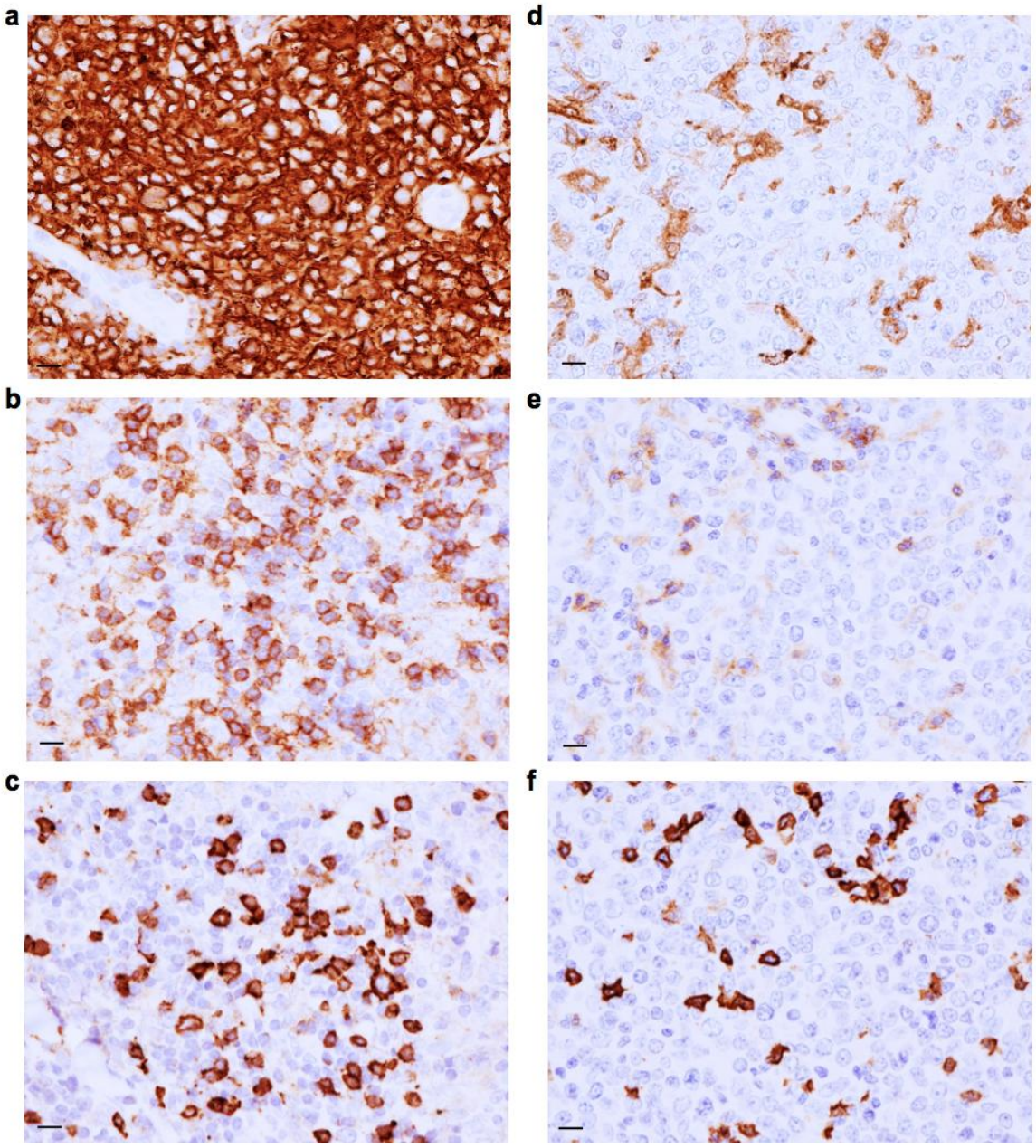

Figure 3.9: MHC class II expression and T cell abundance in primary DLBCL cases. Panels a-c show a *CIITA* wt case with strong membranous MHC class II expression (a) and moderate numbers of tumour infiltrating CD4- (b) and CD8-positive (c) T cells. The case depicted in panels d-f harbours a *CIITA* CDS mutation with subsequent loss of MHC class II expression (d) and, in comparison to the wt case, fewer $CD4^+$ (e) and $CD8^+$ T cells (f). Measurement bars equal 25 μm; original magnification: x400.

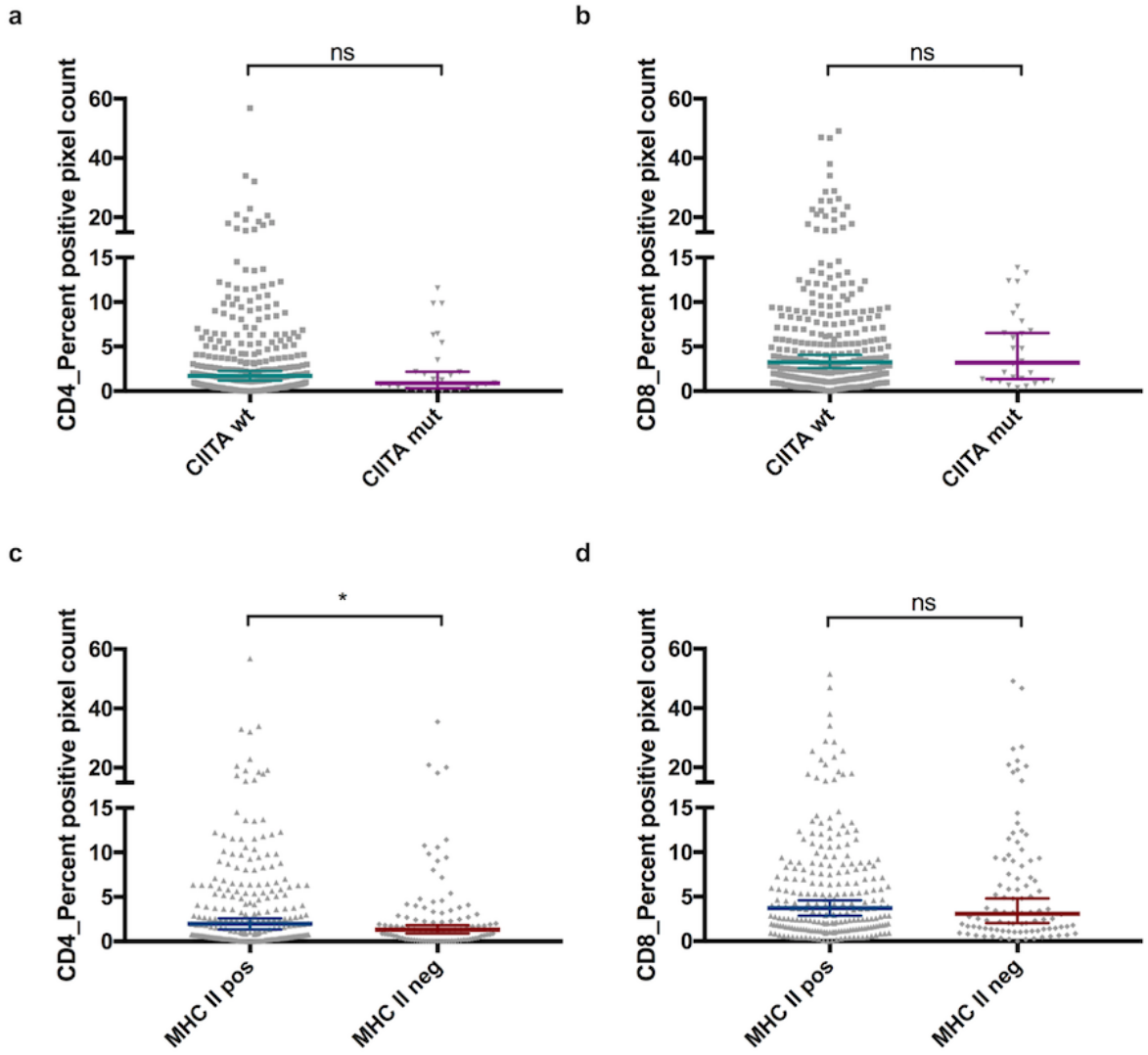

Figure 3.10: Abundance of T cell subsets in primary DLBCL cases. Scatter plots for the percentage of CD4- and CD8-positive pixels, respectively. Coloured horizontal lines show the median and the error bars represent the 95 % confidence interval. Abbreviations: mut, mutant; neg, negative; ns, not significant; pos, positive; wt, wildtype. *P*-values: * < .05; ** < .01; *** < .005

We then performed outcome analyses in this large, uniformly treated DLBCL cohort. No significant differences in OS, DSS, and TTP were observed between *CIITA* wt and mutant cases, but PFS was significantly impaired in patients with *CIITA* mutated DLBCL (Figure 3.11). No differences in survival were seen between MHC class II positive and negative cases (Figure 3.12).

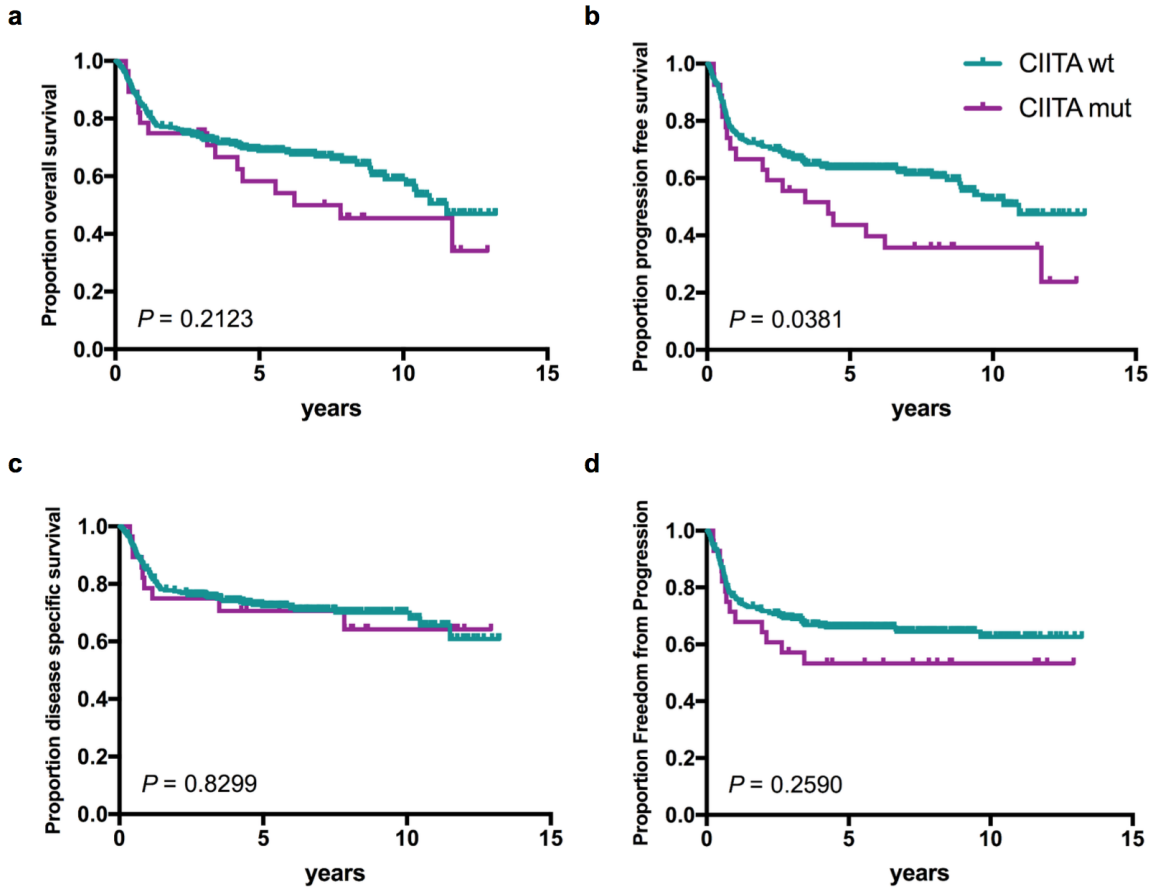

Figure 3.11: Survival of DLBCL patients according to *CIITA* mutational status. Shown are KM-plots for a) OS, b) PFS, c) DSS, and d) TTP.

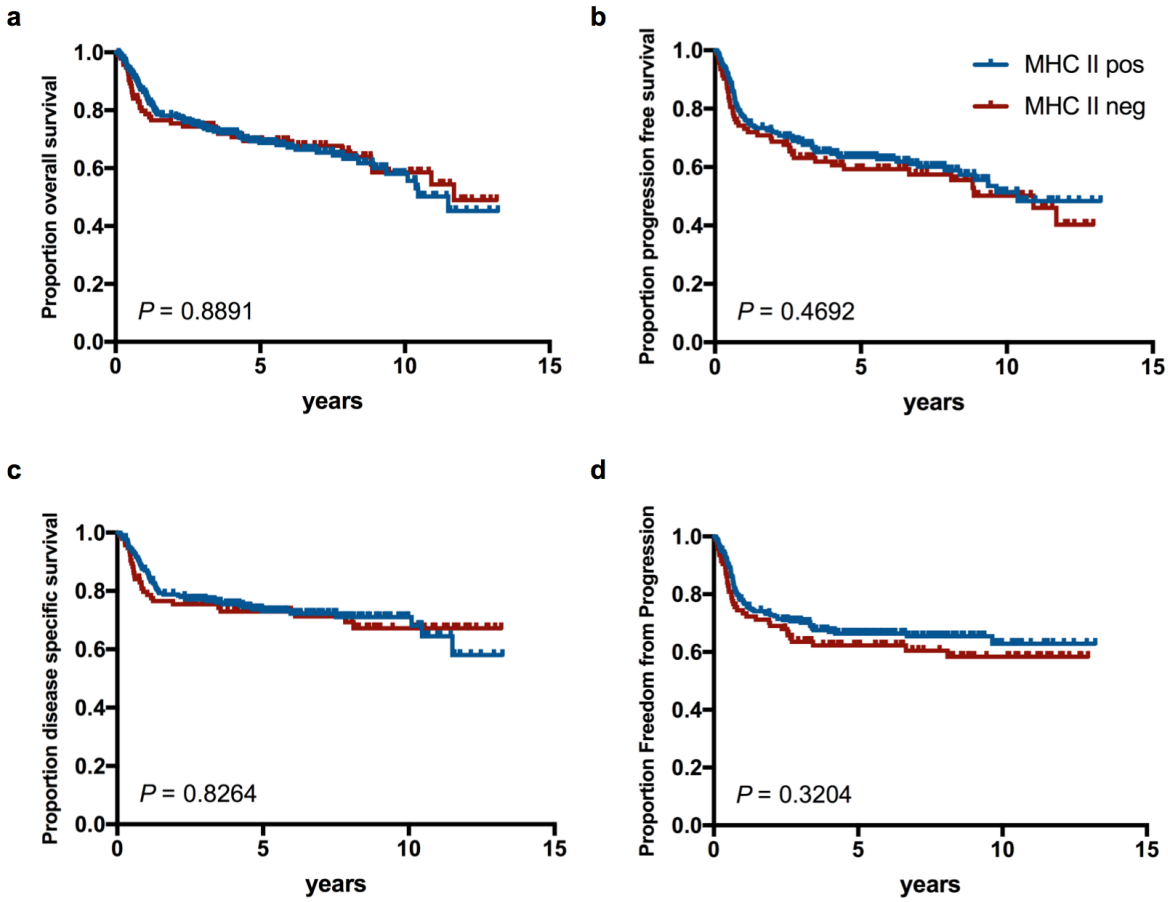

Figure 3.12: Survival of DLBCL patients according to MHC class II surface expression. Shown are KM-plots for a) OS, b) PFS, c) DSS, and d) TTP.

3.3.4.3 FL

3.3.4.3.1 tFL cohort

One hundred and four T1 biopsies were evaluable for both *CIITA* mutational/ba status and MHC class II expression. The Fisher's exact test revealed no significant difference in terms of MHC class II expression between *CIITA* altered and wt cases (*P* = .49). Similar results were obtained for the T2 specimens (130 evaluable cases; Fisher's exact test: *P* = .84).

Since we have observed an enrichment of *CIITA* genetic alterations, specifically structural chromosomal aberrations, in the T2 biopsies compared to their respective T1 specimen (2.3.3.1), we wanted to assess if the T2 specimens were more frequently negative for MHC class II (surface) expression. For 113 tFL pairs we had MHC class II protein expression data available, 85 were MHC class II positive at both timepoints,

seven pairs were concordantly negative, and in three cases tumour cells expressed MHC class II only in the T2 biopsy. In contrast, 18 cases became MHC class II negative at the T2 timepoint (McNemar test, *P* = .002; Figure 3.13). Similar results were obtained when considering surface MHC class II expression (McNemar test, *P* = .001).

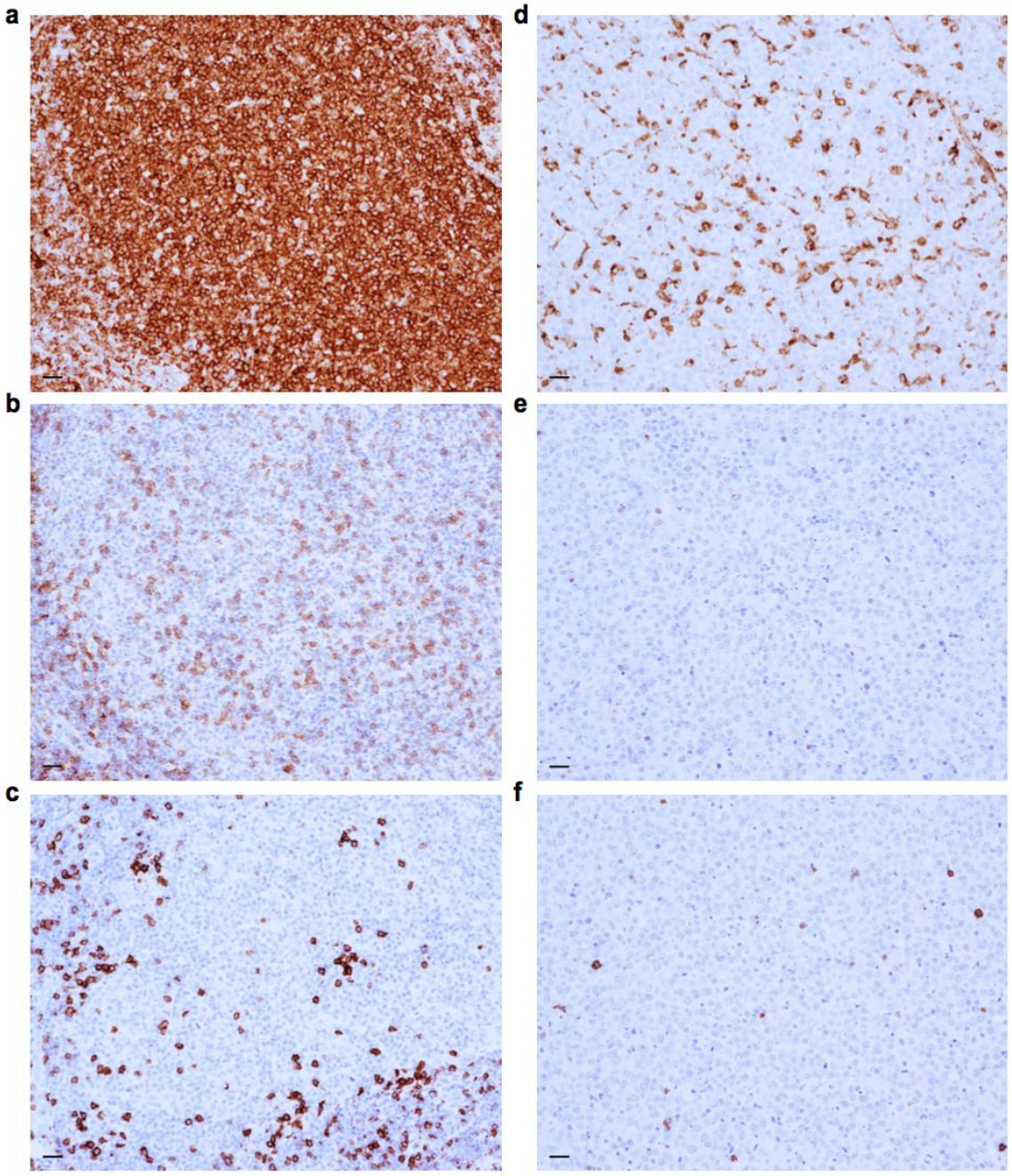

Figure 3.13: MHC class II expression and T cell abundance in tFL. Shown are histological images of two lymphoma biopsies from a patient, who experienced transformation of an FL (panels a-c) into a DLBCL (panels d-f). The initial FL shows tumour cells which are positive for MHC class II (a), whereas the transformed lymphoma specimen is negative (d). In addition, the abundance of $CD4^+$ (b, e) and $CD8^+$ T cells (c, f) is higher in the indolent lymphoma compared to the high-grade lesion. Measurement bars equal 50 μm; original magnification: x200.

Similar to the studies performed in the PMBCL (3.3.4.1) and DLBCL cohort (3.3.4.2), we compared the abundance of CD4- and CD8-positive T cells between the *CIITA* wt and mutated group, as well as between MHC class II positive and negative cases. No significant differences in CD4- and CD8-expressing T cells were observed when comparing *CIITA* wt versus mutant/ba-positive T1 cases (Mann-Whitney test: P = .2 and P = .99, respectively; Figure 3.14), as well as when comparing T2 biopsies with and without genomic *CIITA* alterations (Mann-Whitney test: P = .51 and P = .76, respectively; Figure 3.14).

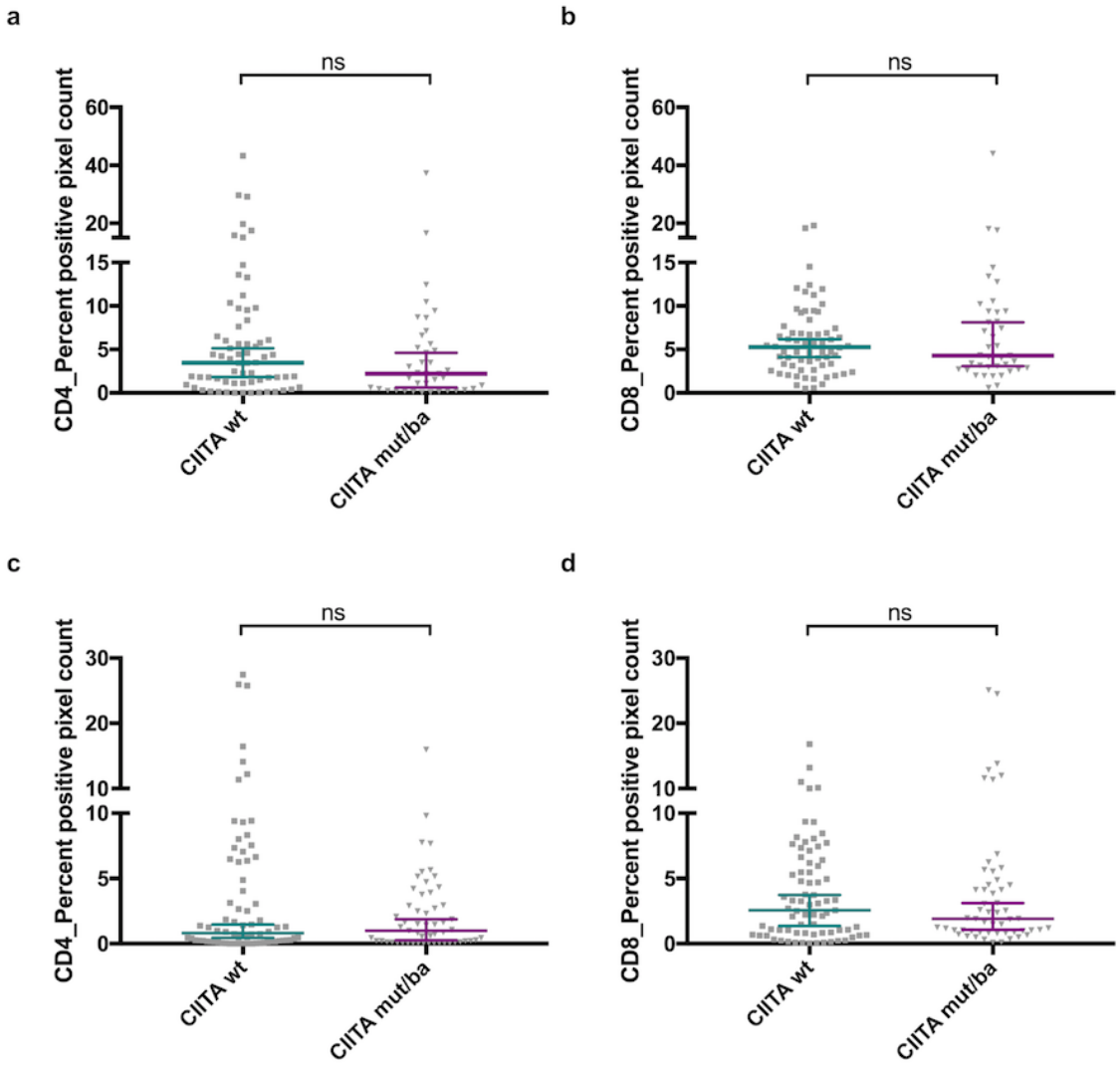

Figure 3.14: Abundance of T cell subsets in tFL according to *CIITA* mutation status. Scatter plots for the percentage of CD4- and CD8-positive pixels. Panel a) and b) show the percentages of CD4- and CD8-positive pixel counts for the T1 specimens, whereas the T2 specimens are shown in panel c) and d).Coloured horizontal lines show the median and the error bars represent the 95 % confidence interval. Abbreviations: ba, break-apart; mut, mutant; neg, negative; ns, not significant; pos, positive; wt, wildtype. *P*-values: * < .05; ** < .01; *** < .005.

When the same type of comparison was performed with regards to MHC class II surface expression on the malignant cell population (Figure 3.15), a significantly higher amount of CD4-expressing T cells was present in MHC class II positive T1 cases (Mann-Whitney test: *P* = .0003) but not in the T2 specimens (Mann-Whitney test: *P* = .35). However, the abundance of CD8-positive T lymphocytes was not significantly different between MHC class II positive and negative T1 cases (Mann-Whitney test: *P* = .07), but was significant in the T2 biopsies (Mann-Whitney test: *P* = .04).

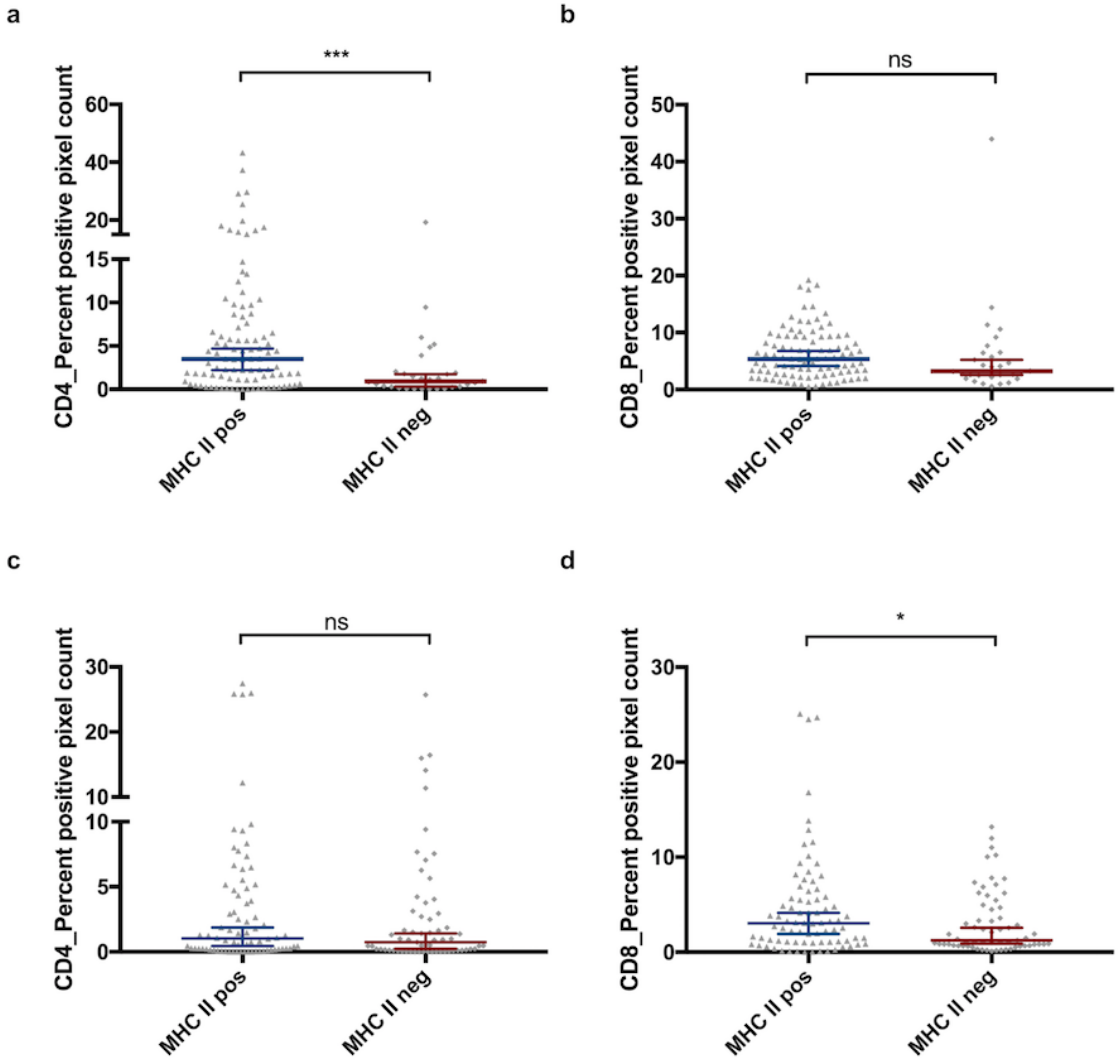

Figure 3.15: Abundance of T cell subsets in tFL according to MHC II expression. Scatter plots for the percentage of CD4- and CD8-positive pixels. Panel a) and b) show the percentages of CD4- and CD8-positive pixel counts for the T1 specimens, whereas the T2 specimens are shown in panel c) and d).Coloured horizontal lines show the median and the error bars represent the 95 % confidence interval. Abbreviations: ba, break-apart; mut, mutant; neg, negative; ns, not significant; pos, positive; wt, wildtype. *P*-values: * < .05; ** < .01; *** < .005.

We also performed outcome analyses in this cohort and were primarily interested in the impact of *CIITA* alterations and MHC class II expression on TTT. No significant differences in TTT were observed between *CIITA* wt and mutant cases, but surprisingly, patients, whose tumours were positive for surface MHC class II, transformed significantly earlier compared to the ones with MHC class II negative tumours (Mann-Whitney test: *P* = .02; Figure 3.16).

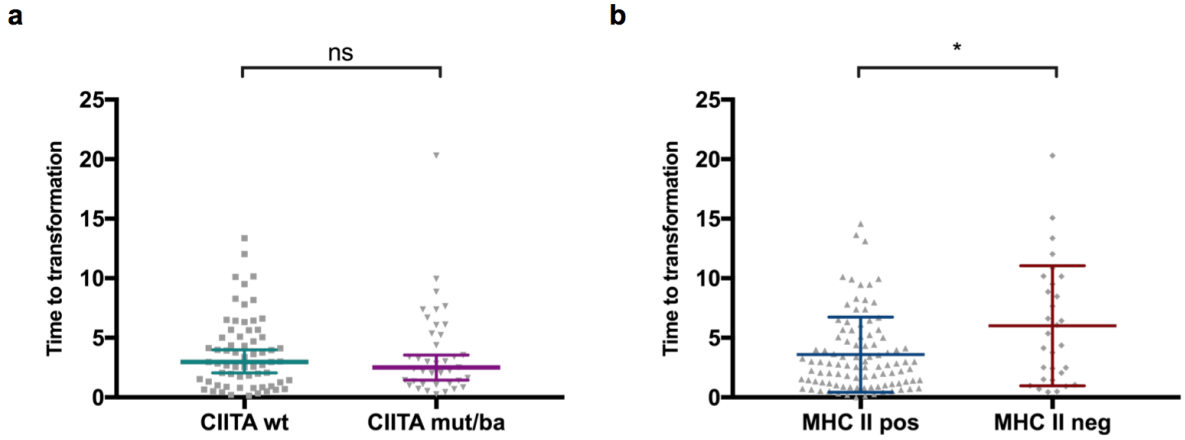

Figure 3.16: TTT analysis in patients with tFL. TTT in patients with tFL comparing *CIITA* wt and mutant cases (a) and MHC class II positive and negative cases (b). *P*-values: * < .05; ** < .01; *** < .005.

No significant differences in OS were seen between *CIITA* wt and mutant cases or MHC class II positive and negative cases (Figure 3.17).

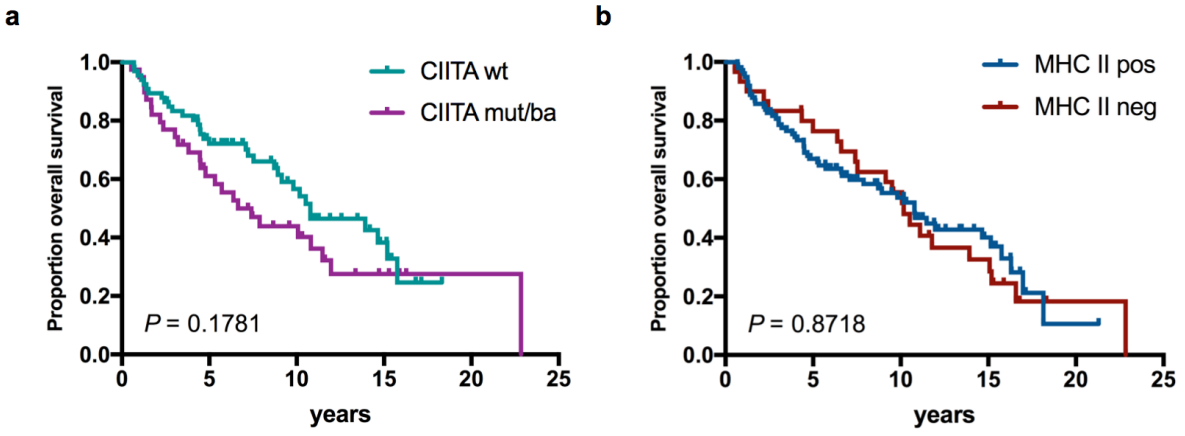

Figure 3.17: OS in patients with tFL. OS in patients with tFL comparing *CIITA* wt and mutant cases (a) and MHC class II positive and negative cases (b).

3.3.4.3.2 pFL/npFL cohort

One hundred five biopsies were evaluable for both, *CIITA* mutational/ba status and MHC class II expression. As in the tFL cohort, there was no significant difference in terms of MHC class II expression status between *CIITA* altered and wt cases (Fisher's exact test, $P = .68$).

The association of *CIITA* mutation status with the abundance of CD4- and CD8-positive T cells (Figure 3.18) revealed no significant differences in CD4 and CD8 T cells between *CIITA* wt and mutant cases (Mann-Whitney test: $P = .61$ and $P = .55$, respectively). When the same type of comparison was performed with regard to MHC class II surface expression (Figure 3.18), again no significant differences were observed (Mann-Whitney test: $P = .08$ and $P = .59$, respectively).

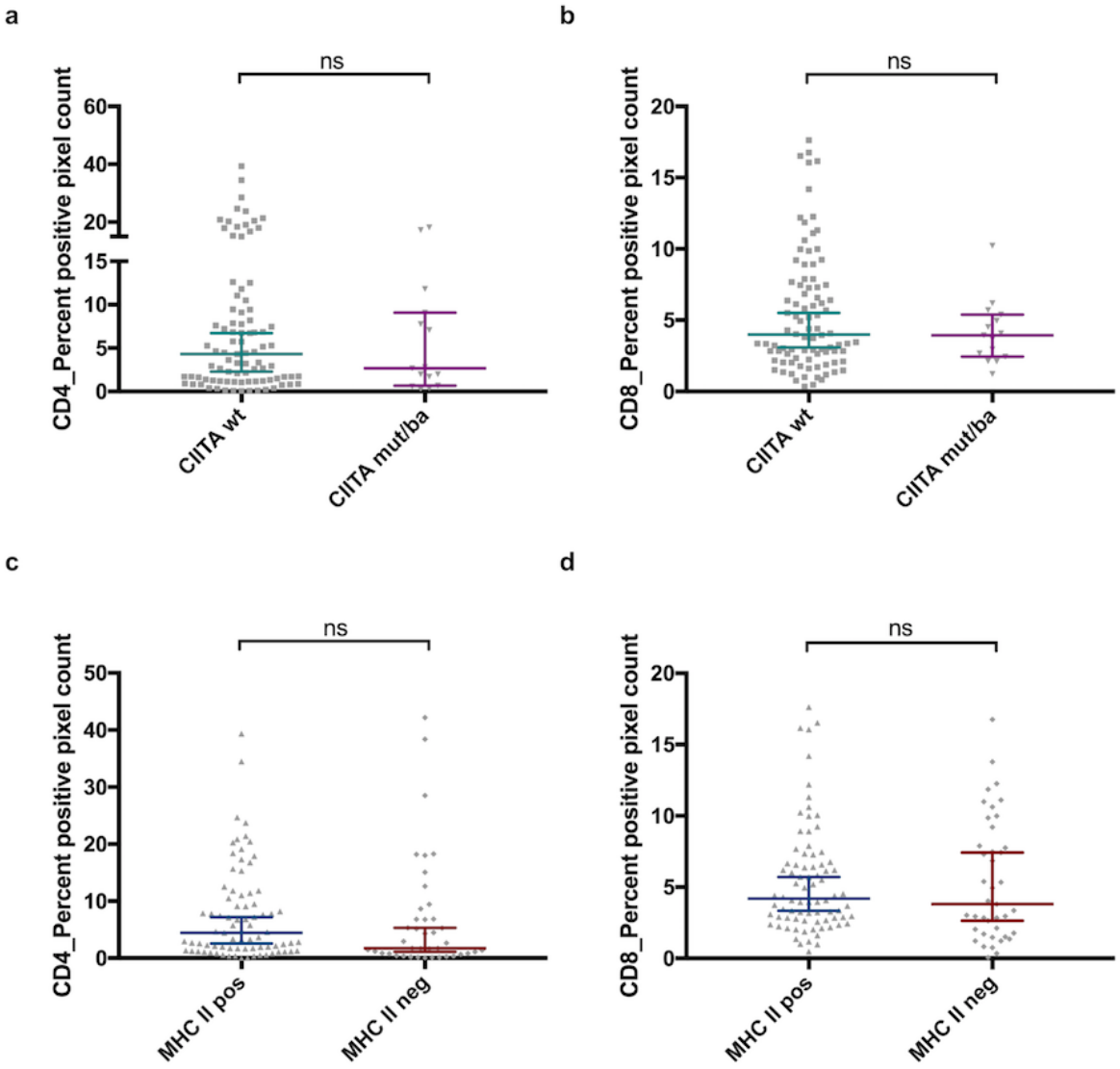

Figure 3.18: Abundance of T cell subsets in primary pFL/npFL cases. Scatter plots for the percentage of CD4- and CD8-positive pixels, respectively. Coloured horizontal lines show the median and the error bars represent the 95 % confidence interval. Abbreviations: ba, break-apart; mut, mutant; neg, negative; ns, not significant; pos, positive; wt, wildtype. *P*-values: * < .05; ** < .01; *** < .005.

Although there seems to be a clear trend towards impaired OS and PFS in patients with *CIITA* mutated tumours (Figure 3.19 panel a and b), similar to what was observed in DLBCL (see section 3.3.4.2), the relatively low numbers preclude us from drawing meaningful conclusions, and further analysis in larger cohorts would be needed to address this question. Differences in survival between MHC class II positive and negative cases were not seen (Figure 3.19 panel c and d).

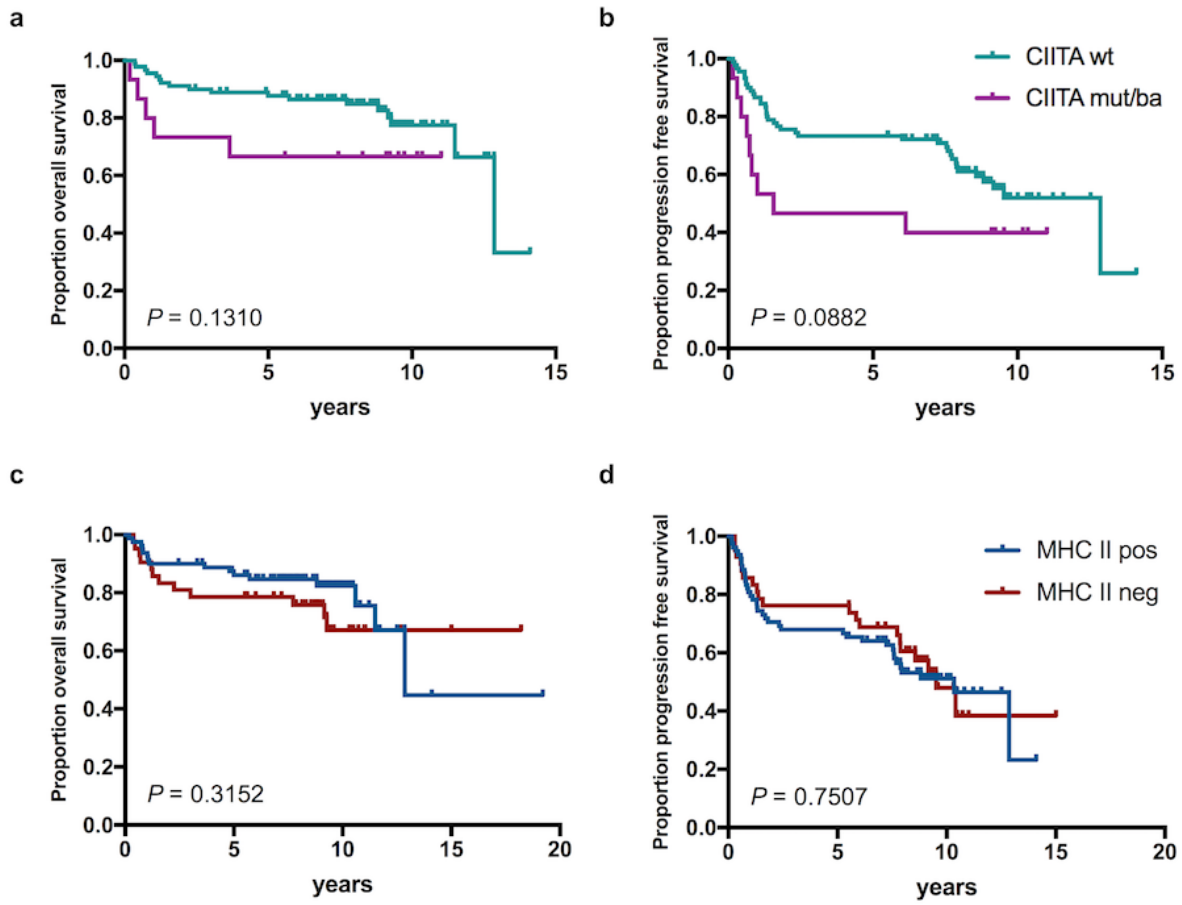

Figure 3.19: Outcomes in patients with pFL and npFL. OS and PFS in patients with pFL/npFL comparing *CIITA* wt and mutant cases (a, b) and MHC class II positive and negative cases (c, d).

3.4 Discussion

In Chapter 2 we have described *CIITA* CDS mutations and structural variants as frequent genetic alterations in PMBCL, DLBCL and FL. Since *CIITA* is known as the master transcriptional regulator of MHC class II, the impaired function would be expected to lead to a decrease in MHC class II expression, which has been linked to reduced tumour cell immunogenicity [215,246]. By performing functional rescue experiments in a cell line model, we were able to demonstrate that at least some of the *CIITA* missense mutants resulted in a functionally deficient protein incapable of inducing HLA-DR surface expression *in vitro*. On the other hand, some altered alleles obviously retain their capability to activate MHC class II transcription. However, some mutations, like the p.636Thr>Met discovered in MedB-1, might lead to a dominant-negative

phenotype, a phenomenon previously seen by others in functional selection studies [206,247] and demonstrated by us for the *CIITA-FLJ27352* gene fusion [172].

In primary PMBCL and DLBCL cases, *CIITA* genomic alterations were significantly associated with loss of MHC class II protein expression, underscoring the role of *CIITA* as the master transcriptional regulator of MHC class II. Interestingly, no such correlation was seen in the FL cohort. In fact, a considerable proportion of cases was characterized by a remarkably heterogeneous MHC class II expression pattern, in the sense that often completely negative areas were admixed with positive follicles or tumour cells with a predominant Golgi-pattern. Future studies would need to explore the underlying genetic/epigenetic mechanisms. Our correlative studies, however, showed that at least in some instances lower numbers of $CD4^+$ and/or $CD8^+$ T cells (as inferred by the percentage of positive pixels, quantified by semi-automated image analysis) were present in the microenvironment of tumour biopsies with low MHC class II expression, whereas no such association was seen with respect to *CIITA* mutation status. This can be explained by our observations in the *in vitro* model, where we were able to demonstrate that some missense mutations do not impair the functional properties of the CIITA protein. Future studies would need to explore changes in T cell proliferation and consequences for their functionality.

Given the semi-quantitative nature of IHC and its low dynamic range, it is difficult to elucidate what the exact consequences of a quantitative reduction might be with respect to tumour-microenvironment interaction. However, our data, together with HLA-DR reduction in *CIITA*-mutated PMBCL cell lines, suggest that *CIITA* mutations may contribute to the immune escape phenomena in these lymphoma subtypes. Since CIITA requires co-factors to induce MHC class II transcription, it is conceivable that mutations of such co-factors may contribute to immune escape. For instance, in a recently published study, Green et al. [90] have shown that *CREBBP* mutant FL samples exhibit lower MHC class II transcript and protein levels. Further studies would be needed to explore how *CREBBP* mutations might be linked to genomic alterations occurring in *CIITA.*

Whereas our data demonstrated the functional relevance and clonal selection of *CIITA* rearrangements and CDS mutations, the pathogenic contribution of the clustered

intron 1 deletions and point mutations still needs to be determined. Of note, intron 1 variants were not correlated with either CIITA or MHC class II expression in our PMBCL sequencing cohort, suggesting that the majority might be co-selected passenger mutations, or more likely the result of AID-mediated aberrant SHM.

In summary, *CIITA* and the subsequent MHC class II abrogation describes an ancillary mechanism that is critically involved in the establishment of an immune privilege phenotype. The identification of acquired immune-privilege properties harbours the potential to develop and pursue targeted treatment approaches, such as PD1- and PDL blockade [248] or cell based vaccination [249], in these molecularly characterized subgroups of B cell lymphoma.

Chapter 4: Conclusion

4.1 Summary of research findings

The main objective of the research presented in this thesis was to obtain a better understanding of the heterogeneity, molecular mechanisms and functional consequences of *CIITA* gene alterations in B cell malignancies. Furthermore, we aimed to inform on the implications for tumour microenvironment composition and patient outcomes in various types of B cell lymphomas. The premise that genetic alterations are frequent across a spectrum of B cell lymphomas constitutes the cornerstone of our hypothesis. To address this, in Chapter 2, we performed targeted sequencing of the *CIITA* locus in large cohorts of PMBCL, DLBCL and (transformed/progressed) FL. FISH was performed on TMAs to broaden our knowledge about the prevalence of *CIITA* structural variants in various B-NHL subtypes [172].

In PMBCL, we discovered that over 70 % of cases harbour at least one genetic hit, establishing *CIITA* as one of the most frequently altered genes in this particular entity. The pattern of *CIITA* aberrations in PMBCL, predominantly consisting of nonsense and frameshift CDS mutations, as well as small deletions and chromosomal breakpoints within a hot spot region confined to the first exon and intron, is indicative of loss of function. Moreover, the pattern of *CIITA* bi-allelic hits in PMBCL-derived cell lines and in a proportion of primary clinical cases, emphasizes a potential tumour suppressor role for *CIITA* in this context.

CIITA CDS mutations were less frequently observed in our cohort of *de novo* DLBCL (8.9 %) compared to PMBCL, but the frequency matches the range observed in previous studies [60,61,223,234]. We did not observe an association with a particular COO-subtype, as determined by gene expression profiling [65]. Interestingly, most of the mutations were missense mutations and clustered in regions encoding functionally important protein domains, such as the conserved NACHT-domain (the GTP-binding domain of CIITA) and the LRR repeats, important for dimerization and nuclear import of the protein.

In our FL cohort, which was enriched for cases with subsequent transformation and progression, we discovered a high frequency of abnormal signal patterns in our FISH study with as much as 36.8 % abnormal cases among the transformed FL

specimens (T2 timepoint), followed by the respective preceding FL samples (T1 timepoint, 29.9 %). Specimens within the pFL and npFL cohorts showed structural chromosomal alterations to a much lesser extent with 15.2 % and 5.4 %, respectively, indicating that these alterations might be a surrogate for an imminent transformation event, when detectable at initial diagnosis. Of note, CDS alterations were not enriched among the FLs which eventually transformed and, similar to the DLBCL cohort, largely consisted of missense mutations affecting crucial protein domains. The analysis of intron 1 mutations in FL showed frequent SNVs, similar to PMBCL, but rarely small deletions or insertions.

To inform comprehensively on novel *CIITA* rearrangements, partner genes and breakpoint anatomy, we applied targeted capture sequencing on 92 lymphoma specimens with known FISH ba-status. We further substantiated our previous finding of a well-defined cluster breakpoint region in intron 1 of *CIITA* and identified 10 novel *CIITA* translocations involving multiple different partner genes, as well as intra-chromosomal structural alterations. Furthermore, we provided evidence that oligocapture sequencing can be successfully performed using FFPE-derived DNA and complements FISH in detecting structural variants [214].

We further hypothesized that the genetic alterations in *CIITA* provide the foundation for the development of an ‘immune privilege’ phenotype as a consequence of loss of MHC class II expression and altered tumour microenvironment composition. By using immunohistochemistry for the exploration of MHC class II expression patterns and automated imaging analysis for the enumeration of T cell subsets, we were able to demonstrate that *CIITA* mutations and structural alterations were associated with a reduction of MHC class II expression in PMBCL and DLBCL, but not in FL. Furthermore, although in most instances the numbers of CD4- and CD8-positive T cells were reduced in *CIITA*-mutated cases compared to wt cases, statistically significant results were only seen when comparisons were centred on surface MHC class II expression.

Lastly, our hypothesis, that *CIITA* alterations and MHC class II expression loss impair patient outcomes was largely disproved. Although the outcomes in *CIITA*-mutated cases were often inferior, a statistically significant result could only be obtained for PFS in the DLBCL cohort.

In conclusion, our findings can be summarized in the following model. Structural genomic alterations, i.e. rearrangements, not only result in the functional abrogation of *CIITA* but also in some cases in overexpression of the respective partner genes such as PD-1 ligands. Therefore, these alterations might present as “double-hits” leading to reduced immunogenicity and the enforcement of an exhausted T cell phenotype. Coding sequence mutations and intron 1 alterations may cause functional impairment of *CIITA*, resulting in reduced levels of MHC class II expression.

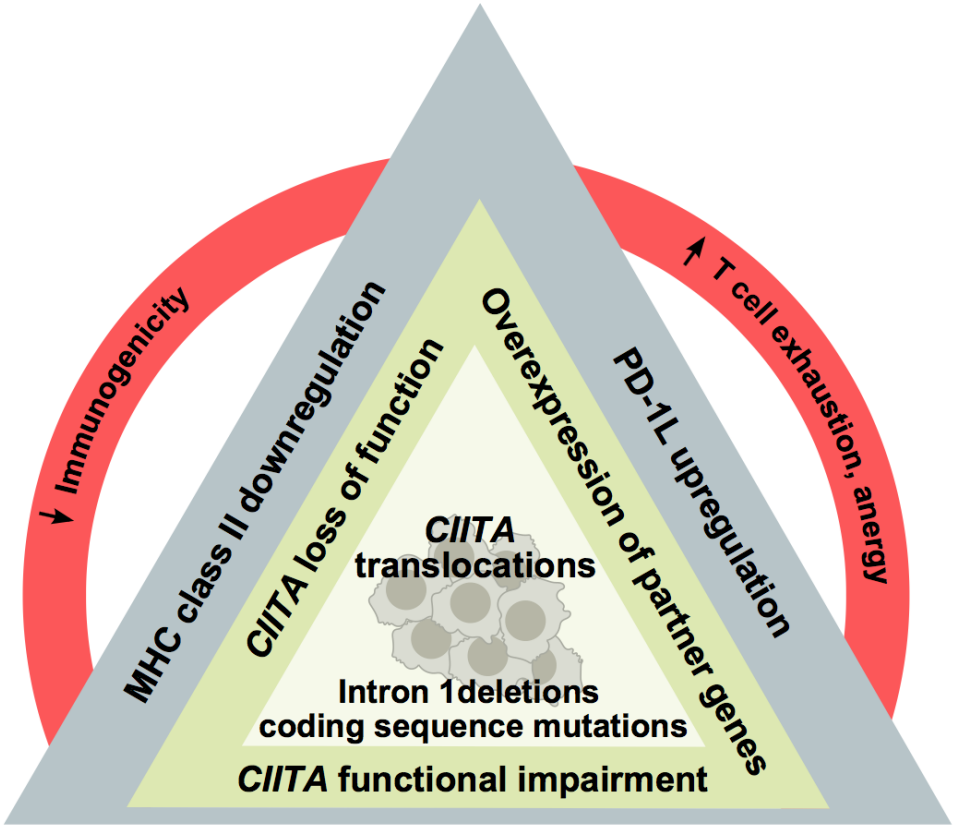

Figure 4.1: Functional impact of *CIITA* alterations in malignant lymphomas.

4.2 Limitations

In the correlative studies performed in this thesis, we largely ignore the fact that - in most instances - tumours represent an aggregation of diverse cell populations with differences in their genetic make-up. This intratumoural heterogeneity is reflected in the existence of subclones, which are exposed to evolutionary pressures, exerted by the tumour niche, the host immune system and therapy [250,251]. In fact, this becomes

clearly evident in our FL cohort, where the allelic frequencies of *CIITA* mutations or the percentage of cells with abnormal FISH-signal patterns were often lower in the initial FL biopsy compared to their respective transformed lymphoma. This likely explains the overt phenotypic heterogeneity with regards to MHC class II protein expression in these specimens.

In addition, malignant tumours represent complex ecosystems and phenotypic traits are often the result of an interaction between multiple genetic alterations, epigenetic regulation and various activated or perturbed signaling pathways [109,110]. Since CIITA requires co-factors to induce MHC class II transcription, it is conceivable that mutations of such co-factors may contribute to immune escape. For instance, in a recently published study, Green et al. [90] have shown that *CREBBP* mutant FL samples exhibit lower MHC class II transcript and protein levels. *CREBBP* is one of the most frequently mutated genes in FL and DLBCL [59,252] and further studies would need to explore how *CREBBP* mutations might be linked to genomic alterations occurring in *CIITA* and how this interplay would determine MHC class II expression. In addition, FOXP1 has recently been proposed to alter MHC class II expression in the ABC-like subtype of DLBCL [253]. Deregulation of pathways such as NFκB and JAK-STAT can result in disturbed expression and abundance of chemokines and cytokines, which in turn might influence expression of CIITA and MHC class II. NFκB and JAK-STAT pathways are constitutively active in DLBCL, PMBCL and, to a lesser extent in FL and likely result in a deregulated micromilieu. The impact on MHC class II expression remains largely unknown. Although in a different cellular context (microglia), it has been demonstrated that IFNγ, IL-3, IL-4 and TNFα can induce CIITA, whereas TGFβ, IL-10 and IL-1b are believed to cause downregulation of CIITA and MHC class II [254,255].

We observed a correlation of MHC class II expression with increased abundance of $CD4^+$ T cells (and in rare instances also with CD8-positive T cells), however, the exact nature and functional phenotype of this population (i.e. Th1, Th2, Th17, regulatory T cells, follicular T helper cells) is unknown. This question would be best addressed using novel technology platforms such as CyTOF, which enables unprecedented simultaneous interrogation of over 40 markers on a single cell level in one experiment. This could be combined with single cell RNA-Seq technologies to further increase the

granularity of information. In addition, functional studies *in vitro,* using co-culture experiments with (HLA-matched) T cells, could provide information on changes in T cell activation status and proliferation.

4.3 Potential applications

A variety of acquired genomic alterations and perturbed signaling pathways can contribute to an immune escape phenotype in malignant lymphoma cells by effectively subverting the ability of the host immune system to target and eliminate tumour cells [111]. Novel therapeutic approaches aiming at a rectification of the deranged anti-tumour immune response have been developed over the past decades to capitalize on the increasing knowledge about immune evasion strategies and mechanisms employed by cancer cells. The compelling success of novel cellular-based therapies, such as CAR-T cells, or immune checkpoint blockade implies a profound knowledge of immune escape mechanisms and synergistic relationships among different pathways to enable the development of predictive biomarkers that can precisely inform on a patient's response to treatment. The potential of determining the genetic basis for clinical response to immune checkpoint inhibitors has been recently demonstrated in patients with malignant melanoma [256]. Preliminary results from phase 1 and 2 clinical trials in lymphoma patients provide evidence for efficacy and safety [248,257,258], but more data are imperatively needed in order to develop reliable prognostic and predictive biomarkers, applicable in routine clinical practice, and to select patients upfront who will benefit from these new therapeutic approaches.

The assessment of MHC deficiency (and underlying genetic mechanisms) will potentially become of greater importance as the field of cancer treatment moves increasingly towards broad implementation of immunotherapeutic strategies for the treatment of cancer patients. A recent study conducted by Johnson et al. [259] has demonstrated the importance of MHC class II expression in patients with malignant melanomas for the response to immune checkpoint blockade with the anti-PD-1 antibodies Nivolumab and Pembrolizumab or anti-PDL1 antibodies. Tumour cells positive for MHC class II (cut-off > 5 %) were more likely to respond to PD-1-blockade, likely due to the contribution of a higher immune cell infiltrate, and had an increased

survival probability. Interestingly, expression of PDL1 itself was relatively rare in those tumours and did not correlate with response to the PD-1/PDL1-targeting approach. The assessment and characterization of defects in antigen presentation can, in theory, help to identify patients which may or may not respond to these immunotherapies. Moreover, with methods available to correct such deficiencies, the response rates could potentially be improved.

An alternative idea is to enhance MHC class II expression for therapeutic purposes in treatment of cancer patients. This approach is largely based on "vaccination" approaches and previous studies have shown that transfection of tumour cells with MHC class II can result in sufficient presentation of cytosolic, tumour-derived endogenous antigens ("neo-antigenes") to $CD4^+$ T cells [260,261]. Therefore, enforced expression of MHC class II in tumour cells poses an opportunity to enhance anti-tumour immune response and tumour rejection [262].

4.4 Ongoing work

We have demonstrated that *CIITA* structural variants occur frequently in tFL and that these genomic alterations might be an indicator of an impending high-grade transformation when detected at initial diagnosis. We also showed that targeted capture sequencing and FISH are complementary techniques to fully capture these alterations, as well as to provide base pair resolution and information on rearrangement partner genes. We will, therefore, apply this technology to our FL cohort, to inform on the exact nature of structural aberrations and to generate hypotheses on functional implications.

Based on prior work of our group in collaboration with Dr. Oliver Weigert and his team (LMU Munich, Germany) on the m7-FLIPI [93], we noticed that the transcription factor FOXP1 is transcriptionally down-regulated in *EZH2* and *MEF2B* mutated FL cases. By using FOXP1 IHC on TMAs from a retrospective series of 107 patients treated with R-CVP at the BCCA, we could demonstrate that FOXP1 protein expression correlated with adverse outcome. We are currently validating the prognostic value of FOXP1 in several TMAs comprised of tissue biopsies from patients enrolled in clinical trials of the German Low Grade Study Group (GLSG). In addition, we will explore this potential biomarker in both, the transformed and progressed FL cohorts. This is not only

of interest in terms of outcome prediction, but also because of the potential link between FOXP1 and MHC class II expression, as recently demonstrated in DLBCL [253]. In this study, Brown et al. showed that expression of FOXP1 was inversely correlated with mRNA expression levels of several MHC class II genes. Moreover, knock-down of *FOXP1* resulted in upregulation of HLA-DRA and CD74. We therefore aim to evaluate if FOXP1 is directly or indirectly involved in regulating expression of CIITA and thereby MHC class II.

4.5 Open questions and future directions

4.5.1 Epigenetic control of MHC II expression

It is well known, that CIITA is involved in the recruitment of chromatin-/histone-modifiers and therefore in the epigenetic control and modulation of MHC class II expression. Of importance for maximal expression of MHC class II is the interaction of the N-terminus of CIITA with CREBBP-p300 [263,264]. As a result, CIITA can be perceived as a bridging molecule, facilitating the interaction of DNA-bound transcription factors, histone acetyltransferases (leading to the acetylation of histones H3 and H4) and the transcription machinery. It is suggested, that the assembly of the DNA-binding proteins RFX, CREB and NF-Y leads to a basal level of H3 and H4 acetylation, and therefore may result in MHC class II expression, even without the presence of CIITA, if exceeding a certain threshold [265]. Furthermore, it has been proposed that MHC class II can be expressed in a CIITA-independent manner in immune-privileged sites, endothelial cells and dendritic cells of *CIITA* knock-out mice, and that this process facilitates the presentation of cytosolic rather than endolysosomal peptides [266–268]. It remains to be shown to which extent these processes occur in normal or malignant B cells in the context of *CIITA*-deficiency. In addition, it needs to be addressed if *CIITA* and *CREBBP* mutations occur frequently together or if they are mutually exclusive in primary FL and DLBCL cases.

Another mechanism of epigenetic regulation of *HLA-DRA* and *CIITA* (in particular of the pIV promoter of *CIITA*) is the introduction of suppressive histone modifications such a H3K27me3, conferred by EZH2, a histone methyltransferase [269,270]. *EZH2* is commonly mutated in DLBCL and FL [70] and further molecular studies need to

elucidate the role of this gene for CIITA and MHC class II expression in B cell malignancies.

4.5.2 Post-translational modification and degradation of CIITA and MHC class II

The N-terminal portion of the CIITA pIII isoform is known to facilitate interaction with cofactors and components of the transcription machinery, and it is also believed to play a role in recruitment to the HLA-DRA promoter [174,175,180]. Recent studies provided evidence, that this same protein domain is also involved in the rapid degradation of CIITA via the ubiquitination pathway [271,272]. Interestingly, Beaulieu et al. demonstrated that SNVs or short deletions within the acidic domain led to increased protein stability accompanied by a loss or reduction of the transactivation potential [272]. Since some of the CDS mutations described in this thesis were found to affect the acidic domain, further studies would be needed to elucidate the consequences for protein degradation and the interaction of CIITA with the basal transcription machinery.

As outlined in section 1.3.2, some APCs, such as macrophages and dendritic cells, but also B cells, can upregulate MHC class II surface expression upon stimulation. Conversely, Thibodeau et al. showed that IL-10 induces transcription of *MARCH1*, a gene encoding a membrane-associated ubiquitin ligase, therefore leading to internalization and degradation of MHC class II protein [273]. To which extend these mechanisms occur in B cell malignancies is largely unexplored.

4.5.3 Interplay between MHC and the PD-1/PDL axis

A recent study provided some evidence that MHC class II expression might be of importance for the response of patients to immune checkpoint blockade [259]. This study was performed in patients with malignant melanomas, a tumour widely considered as being highly immunogenic. We aim on assessing the interplay between MHC and the PD-1/PDL axis in *in vitro* and syngeneic *in vivo* mouse models of malignant lymphomas. Specifically, we plan on generating MHC class I/MHC class II deficient cell lines using the CRISPR/Cas9 genome editing tool combined with ectopic overexpression of PDL1 or PDL2, respectively. *In vitro* co-culture experiments with HLA-matched CD4- or CD8-

positive T cells, as well as *in vivo* mouse models, will be used to measure differences in response to immune checkpoint blockade with Nivolumab.

4.6 Final conclusion

Herein, we described genomic rearrangements and coding sequence mutations in *CIITA* as being frequent across a spectrum of B cell lymphoma subtypes. We demonstrated that these genetic alterations can result in diminished MHC class II expression and subsequently in an altered microenvironment composition with a lower abundance of CD4- and CD8-positive T cells in the tumour microenvironment. We identified that at least some of the genetic alterations are likely a byproduct of AID-mediated somatic hypermutation, as evidenced by the co-occurrence of mutations in the non-coding region of *CIITA* intron 1. In addition, we described novel translocations involving a broad spectrum of rearrangement partner genes and intra-chromosomal structural variants.

Our work is of translational relevance as it provides us with a deeper understanding of molecular mechanisms resulting in MHC class II deficiency. With the emergence of immunotherapeutic approaches targeting the tumour-host interface and the success in certain cancer types (i.e. cHL), we will need reliable biomarkers to better predict response to those novel agents.

A comprehensive assessment of immune escape mechanisms and the concomitant development of therapeutic approaches has the potential to foster the implementation of precision medicine for patients diagnosed with malignant lymphomas. The ultimate goal should be the improvement of cure rates and the reduction of treatment-related toxicities.

Bibliography

1. Flajnik MF, Kasahara M: **Origin and evolution of the adaptive immune system: genetic events and selective pressures**. *Nat. Rev. Genet.* 2009, **11**:47–59.
2. Cooper M, Peterson R, Good R: **Delineation of the thymic and bursal lymphoid systems in the chicken**. *Nature* 1965, **205**:143–146.
3. Cooper M, Peterson RD, South M, Good RA: **The Functions of the Thymus System and the Bursa System in the Chicken**. *J. Exp. Med.* 1966, **123**:75–102.
4. Schatz DG, Oettinger MA, Baltimore D: **The V(D)J recombination activating gene, RAG-1**. *Cell* 1989, **59**:1035–1048.
5. Oettinger MA, Schatz DG, Gorka C, Baltimore D, Oetringer MA: **RAG-1 and RAG-2, Adjacent Genes That Synergistically Activate V(D)J Recombination**. *Science (80-.).* 1990, **248**:1517–1523.
6. Alt FW, Blackwell TK, Yancopoulos GD: **Development of the Primary Antibody Repertoire**. *Science (80-.).* 1987, **238**:1079–1087.
7. Hozumi N, Tonegawa S: **Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions**. *Proc Natl Acad Sci USA* 1976, **73**:3628–3632.
8. Brack C, Hirama M, Lenhard-Schuller R, Tonegawa S: **A complete immunoglobulin gene is created by somatic recombination**. *Cell* 1978, **15**:1–14.
9. Alt FW, Yancopoulos GD, Blackwell TK, Wood C, Thomas E, Boss M, Coffman R, Rosenberg N, Tonegawal S, Baltimore D: **Ordered rearrangement of immunoglobulin heavy chain variable region segments**. *EMBO J.* 1984, **3**:1209–1219.
10. Tonegawa S: **Reiteration frequency of immunoglobulin light chain genes: Further evidence for somatic generation of antibody diversity**. *Proc Natl Acad Sci USA* 1976, **73**:203–207.
11. Alt FW, Enea V, Bothwell ALM, Baltimore D: **Activity of Multiple Light Chain Genes in Murine Myeloma Cells Producing a Single, Functional Light Chain**. *Cell* 1980, **20**:293–301.

12. Hinds-Frey KR, Nishikata H, Litman RT, Litman GW: **Somatic Variation Precedes Extensive Diversification of Germline Sequences and Combinatorial Joining in the Evolution of Immunoglobulin Heavy Chain Diversity**. *J. Exp. Med.* 1993, **178**:815–824.
13. Casellas R, Basu U, Yewdell WT, Chaudhuri J, Robbiani DF, Di Noia JM: **Mutations, kataegis and translocations in B cells: understanding AID promiscuous activity**. *Nat. Rev. Immunol.* 2016, **16**:164–176.
14. Pernis B, Chiappino G, Kelus AS, Gell PGH: **Cellular Localization of Immunoglobulins with Different Allotypic Specificities in Rabbit Lymphoid Tissues**. *J. Clin. Pathol.* 1965, **122**:853–877.
15. Ehlich A, Schaal S, Gu H, Kitamura D, Miiller W, Rajewsky K: **Immunoglobulin Heavy and Light Chain Genes Rearrange Independently at Early Stages of B Cell Development**. *Cell* 1993, **72**:695–704.
16. Löffert D, Ehlich A, Müller W, Rajewsky K: **Surrogate light chain expression is required to establish immunoglobulin heavy chain allelic exclusion during early B cell development**. *Immunity* 1996, **4**:133–144.
17. Mostoslavsky R, Alt FW, Rajewsky K: **The lingering enigma of the allelic exclusion mechanism**. *Cell* 2004, **118**:539–544.
18. Klein U, Dalla-Favera R: **Germinal centres: role in B-cell physiology and malignancy**. *Nat Rev Immunol* 2008, **8**:22–33.
19. Victora GD, Nussenzweig MC: **Germinal Centers**. *Annu. Rev. Immunol* 2012, **30**:429–57.
20. Flemming W: **Beiträge zur Kenntniss der Zelle und Ihrer Lebenserscheinungen**. *Arch. für mikroskopische Anat.* 1880, **18**:151–259.
21. Flemming W: **Zur Kenntnis der Zelle und ihrer Theilungs-Erscheinungen**. *Schriften des Naturwissenschaftlichen Vereins* 1878, **3**:23–27.
22. Niewenhuis P, Opstelten: **Functional anatomy of germinal centers**. *Am. J. Anat.* 1984, **170**:421–435.
23. Allen CDC, Okada T, Cyster JG: **Germinal-Center Organization and Cellular Dynamics**. *Immunity* 2007, **27**:190–202.
24. Allen CDC, Ansel KM, Low C, Lesley R, Tamamura H, Fujii N, Cyster JG:

Germinal center dark and light zone organization is mediated by CXCR4 and CXCR5. *Nat. Immunol.* 2004, **5**:943–952.

25. Bleul CC, Fuhlbrigge RC, Casasnovas JM, Aiuti A, Springer T a: **A highly efficacious lymphocyte chemoattractant, stromal cell-derived factor 1 (SDF-1)**. *J. Exp. Med.* 1996, **184**:1101–1109.
26. Allen CDC, Okada T, Tang HL, Cyster JG: **Imaging of Germinal Center Selection Events During Affinity Maturation**. *Science (80-.).* 2007, **315**:528–531.
27. Victora GD, Schwickert TA, Fooksman DR, Kamphorst AO, Meyer-Hermann M, Dustin ML, Nussenzweig MC: **Germinal center dynamics revealed by multiphoton microscopy with a photoactivatable fluorescent reporter**. *Cell* 2010, **143**:592–605.
28. Schwickert TA, Lindquist RL, Shakhar G, Livshits G, Skokos D, Kosco-Vilbois MH, Dustin ML, Nussenzweig MC: **In vivo imaging of germinal centres reveals a dynamic open structure**. *Nature* 2007, **446**:83–87.
29. Hauser AE, Junt T, Mempel TR, Sneddon MW, Kleinstein SH, Henrickson SE, von Andrian UH, Shlomchik MJ, Haberman AM: **Definition of Germinal-Center B Cell Migration In Vivo Reveals Predominant Intrazonal Circulation Patterns**. *Immunity* 2007, **26**:655–667.
30. Gitlin AD, Shulman Z, Nussenzweig MC: **Clonal selection in the germinal centre by regulated proliferation and hypermutation**. *Nature* 2014, **509**:637–640.
31. Kelsoe G: **Life and death in germinal centers**. *Immunity* 1996, **4**:107–111.
32. Schwickert TA, Victora GD, Fooksman DR, Kamphorst AO, Mugnier MR, Gitlin AD, Dustin ML, Nussenzweig MC: **A dynamic T cell-limited checkpoint regulates affinity-dependent B cell entry into the germinal center**. *J. Exp. Med* 2011, **208**:1243–1252.
33. Berek C, Milstein C: **Mutation Drift and Repertoire Shift in the Maturation of the Immune Response**. *Immunol. Rev.* 1987, **96**:23–41.
34. Muramatsu M, Kinoshita K, Fagarasan S, Yamada S, Shinkai Y, Honjo T: **Class Switch Recombination and Hypermutation Require Activation-Induced**

Cytidine Deaminase (AID), a Potential RNA Editing Enzyme. *Cell* 2000, **102**:553–563.

35. Muto T, Muramatsu M, Taniwaki M, Kinoshita K, Honjo T: **Isolation, Tissue Distribution, and Chromosomal Localization of the Human Activation-Induced Cytidine Deaminase (AID) Gene**. *Genomics* 2000, **68**:85–88.
36. Revy P, Muto T, Levy Y, Geissmann F, Plebani A, Sanal O, Catalan N, Forveille M, Dufourcq-Lagelouse R, Gennery A, et al.: **Activation-Induced Cytidine Deaminase (AID) Deficiency Causes the Autosomal Recessive Form of the Hyper-IgM Syndrome (HIGM2)**. *Cell* 2000, **102**:565–575.
37. Han L, Masani S, Yu K, Frederick Alt by W: **Overlapping activation-induced cytidine deaminase hotspot motifs in Ig class-switch recombination**. *Proc Natl Acad Sci USA* 2011, **108**:11584–11589.
38. Martin A, Bardwell PD, Woo CJ, Fan M, Shulman MJ, Scharff MD: **Activation-induced cytidine deaminase turns on somatic hypermutation in hybridomas**. *Nature* 2002, **415**:802–806.
39. Victora GD, Dominguez-Sola D, Holmes AB, Deroubaix S, Dalla-Favera R, Nussenzweig MC: **Identification of human germinal center light and dark zone cells and their relationship to human B-cell lymphomas**. *Blood* 2012, **120**:2240–2248.
40. Pasqualucci L, Bhagat G, Jankovic M, Compagno M, Smith P, Muramatsu M, Honjo T, Morse III HC, Nussenzweig MC, Dalla-Favera R: **AID is required for germinal center–derived lymphomagenesis**. *Nat. Genet.* 2008, **40**:108–112.
41. Pasqualucci L, Migliazza A, Fracchiolla N, William C, Neri A, Baldini L, Chaganti RSK, Klein U, Küppers R, Rajewsky K, et al.: **BCL-6 mutations in normal germinal center B cells: Evidence of somatic hypermutation acting outside Ig loci**. *Proc. Natl. Acad. Sci.* 1998, **95**:11816–11821.
42. Pasqualucci L, Neumeister P, Goossens T, Nanjangud G, Chaganti RSK, Kuppers R, Dalla-Favera R: **Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas**. *Nature* 2001, **412**:341–346.
43. Migliazza A, Martinotti S, Chen W, Fusco C, Ye BH, Knowles DM, Offit K, Chaganti RS, Dalla-Favera R: **Frequent somatic hypermutation of the 5'**

noncoding region of the BCL6 gene in B-cell lymphoma. *Proc. Natl. Acad. Sci.* 1995, **92**:12520–12524.

44. Kleinstein M Tomayko SH, Shlomchik MJ, Schatz DG, Duke JL, Liu M, Yaari G, Khalil AM: **Predict AID Targeting in Non-Ig Genes Multiple Transcription Factor Binding Sites [Internet]**. *J Immunol* 2013, **190**:3878–3888.
45. Yamane A, Resch W, Kuo N, Kuchen S, Li Z, Sun H, Robbiani DF, McBride K, Nussenzweig MC, Casellas R: **Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes**. *Nat. Immunol.* 2011, **12**:62–69.
46. Liu M, Duke JL, Richter DJ, Vinuesa CG, Goodnow CC, Kleinstein SH, Schatz DG: **Two levels of protection for the B cell genome during somatic hypermutation**. *Nature* 2008, **451**:841–845.
47. Basso K, Dalla-Favera R: **Germinal centres and B cell lymphomagenesis. [Internet]**. *Nat. Rev. Immunol.* 2015, **15**:172–84.
48. Swerdlow SH, Campo E, Harris N, Jaffe ES, Pileri SA, Stein H, Thiele J, Vardiman JW: *WHO classification of tumours of haematopoeitic and lymphoid tissues*. IARC; 2008.
49. Swerdlow SH, Campo E, Pileri SA, Lee Harris N, Stein H, Siebert R, Advani R, Ghielmini M, Salles GA, Zelenetz AD, et al.: **The 2016 revision of the World Health Organization classification of lymphoid neoplasms**. *Blood* 2016, **127**:2375–2390.
50. Howlader N, Noone A, Krapcho M, Miller D, Bishop K, Altekruse S, Kosary C, Yu M, Ruhl J, Tatalovich Z, et al.: *SEER Cancer Statistics Review, 1975-2013, National Cancer Institute*. 2016.
51. Kuppers R: **Mechanisms of B-cell lymphoma pathogenesis**. *Nat Rev Cancer* 2005, **5**:251–262.
52. Carbone A, Gloghini A, Kwong Y-L, Younes A: **Diffuse large B cell lymphoma: using pathologic and molecular biomarkers to define subgroups for novel therapy**. *Ann. Hematol.* 2014, **93**:1263–1277.
53. Roschewski M, Staudt LM, Wilson WH: **Diffuse large B-cell lymphoma-treatment approaches in the molecular era.** *Nat. Rev. Clin. Oncol.* 2014,

11:12–23.

54. Lieber MR: **Mechanisms of human lymphoid chromosomal translocations**. *Nat. Rev. Cancer* 2016, **16**:387–398.
55. The Non-Hogkin's Lymphoma Classification Project: **A Clinical Evaluation of the International Lymphoma Study Group Classification of Non-Hodgkin's Lymphoma**. *Blood* 1997, **89**:3909–3918.
56. Oschlies I, Klapper W, Zimmermann M, Krams M, Wacker H, Harder L, Siebert R, Reiter A, Parwaresch R, Burkhardt B: **Diffuse large B-cell lymphoma in pediatric patients belongs predominantly to the germinal-center type B-cell lymphomas : a clinicopathologic analysis of cases included in the German BFM (Berlin-Frankfurt-Münster) Multicenter Trial**. *Blood* 2011, **107**:4047–4052.
57. Jaffe ES, Harris NL, Vardiman JW, Campo E, Arber DA: *Hematopathology*. 2011.
58. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, et al.: **Distinct types of diffuse large B-cell lymphoma identified by gene expression profling**. *Nature* 2000, **403**:503–511.
59. Morin RD, Mendez-Lago M, Mungall AJ, Goya R, Mungall KL, Corbett R, Johnson NA, Severson TM, Chiu R, Field M, et al.: **Frequent mutation of histone modifying genes in non-Hodgkin lymphoma**. *Nature* 2011, **476**:298–303.
60. Morin RD, Mungall K, Pleasance E, Mungall AJ, Goya R, Huff RD, Scott DW, Ding J, Roth A, Chiu R, et al.: **Mutational and structural analysis of diffuse large B-cell lymphoma using whole-genome sequencing**. *Blood* 2013, **122**:1256–1265.
61. Pasqualucci L, Trifonov V, Fabbri G, Ma J, Rossi D, Chiarenza A, Wells VA, Grunn A, Messina M, Elliot O, et al.: **Analysis of the coding genome of diffuse large B-cell lymphoma**. *Nat Genet* 2011, **43**:830–837.
62. Rosenwald A, Wright G, Chan W, Connors JM, Campo E, Fisher RI, Gascoyne RD, Muller-Hermelink H, Smeland EB, Giltnane J, et al.: **The use of molecular profiling to predict survival after chemotherapy for diffuse large B-cell lymphoma.** *N. Engl. J. Med.* 2002, **346**:1937–1947.
63. Lenz G, Wright GW, Emre NCT, Kohlhammer H, Dave SS, Davis RE, Carty S, Lam LT, Shaffer AL, Xiao W, et al.: **Molecular subtypes of diffuse large B-cell**

lymphoma arise by distinct genetic pathways. *Proc Natl Acad Sci USA* 2008, **105**:13520–13525.

64. Lenz G, Wright GW, Dave SS, Xiao W, Powell J, Zhao H, Xu W, Tan B, Goldschmidt N, Iqbal J: **Stromal gene signatures in large-B-cell lymphomas**. *N. Engl. J. Med.* 2008, **359**:2313–2323.
65. Scott DW, Mottok A, Ennishi D, Wright GW, Farinha P, Ben-Neriah S, Kridel R, Barry GS, Hother C, Abrisqueta P, et al.: **Prognostic significance of diffuse large B-cell lymphoma cell of origin determined by digital gene expression in formalin-fixed paraffin-embedded tissue biopsies**. *J. Clin. Oncol.* 2015, **33**:2848–2856.
66. Lo Coco F, Ye BH, Corradini P, Offit K, Knowles DM, Chaganti RSK, Dalla-Favera R: **Rearrangements of the BCL6 Gene in Diffuse Large Cell Non-Hodgkin's Lymphoma**. *Blood* 1994, **83**:1757–1759.
67. Ladanyi M, Offit K, Jhanwar SC, Filippa DA, Chaganti RSK: **MYC Rearrangement and Translocations Involving Band 8q24 in Diffuse Large Cell Lymphomas**. *Blood* 1991, **77**:1057–1063.
68. Kramer MH, Hermans J, Wijburg E, Philippo K, Geelen E, van Krieken JH, de Jong D, Maartense E, Schuuring E, Kluin PM: **Clinical relevance of BCL2, BCL6, and MYC rearrangements in diffuse large B-cell lymphoma [Internet]**. *Blood* 1998, **92**:3152–3162.
69. Scott DW, Mungall KL, Ben-Neriah S, Rogic S, Morin RD, Slack GW, Tan KL, Chan FC, Lim RS, Connors JM, et al.: **TBL1XR1 / TP63 : a novel recurrent gene fusion in B-cell non-Hodgkin lymphoma**. *Blood.* 2012, **119**:4949–4952.
70. Morin RD, Johnson NA, Severson TM, Mungall AJ, An J, Goya R, Paul JE, Boyle M, Woolcock BW, Kuchenbauer F, et al.: **Somatic mutation of EZH2 (Y641) in Follicular and Diffuse Large B-cell Lymphomas of Germinal Center Origin [Internet]**. *Nat. Genet.* 2010, **42**:181–185.
71. Ortega-Molina A, Boss IW, Canela A, Pan H, Jiang Y, Zhao C, Jiang M, Hu D, Agirre X, Niesvizky I, et al.: **The histone lysine methyltransferase KMT2D sustains a gene expression program that represses B cell lymphoma development**. *Nat. Med.* 2015, **21**:1199–1208.

72. Pasqualucci L, Migliazza A, Basso K, Houldsworth J, Chaganti RSK, Dalla-Favera R: **Mutations of the BCL6 proto-oncogene disrupt its negative autoregulation in diffuse large B-cell lymphoma**. *Blood* 2003, **101**:2914–2923.
73. Cattoretti G, Pasqualucci L, Ballon G, Tam W, Nandula S V., Shen Q, Mo T, Murty V V., Dalla-Favera R: **Deregulated BCL6 expression recapitulates the pathogenesis of human diffuse large B cell lymphomas in mice**. *Cancer Cell* 2005, **7**:445–455.
74. Ying CY, Dominguez-Sola D, Fabi M, Lorenz IC, Hussein S, Bansal M, Califano A, Pasqualucci L, Basso K, Dalla-Favera R: **MEF2B mutations lead to deregulated expression of the BCL6 oncogene in Diffuse Large B cell Lymphoma [Internet]**. *Nat Immunol* 2013, **14**:1084–1092.
75. Schmidt J, Gong S, Marafioti T, Mankel B, Gonzalez-Farre B, Balagú O, Mozos A, Cabeçadas JE, van der Walt J, Hoehn D, et al.: **Genome-wide analysis of pediatric-type follicular lymphoma reveals low genetic complexity and recurrent alterations of TNFRSF14 gene**. *Blood* 2016, **128**:1101–1111.
76. Louissaint A, Schafernak KT, Geyer JT, Kovach AE, Ghandi M, Gratzinger D, Roth CG, Paxton CN, Kim S, Namgyal C, et al.: **Pediatric-type nodal follicular lymphoma: A biologically distinct lymphoma with frequent MAPK pathway mutations**. *Blood* 2016, **128**:1093–1100.
77. Oeschger S, Bräuninger A, Küppers R, Hansmann M-L: **Tumor cell dissemination in follicular lymphoma [Internet]**. *Blood* 2002, **99**:2192–2198.
78. Dave SS, Wright G, Tan B, Rosenwald A, Gascoyne RD, Chan WC, Fisher RI, Braziel RM, Rimsza LM, Grogan TM, et al.: **Prediction of Survival in Follicular Lymphoma Based on Molecular Features of Tumor-Infiltrating Immune Cells**. *N. Engl. J. Med.* 2004, **351**:2159–2169.
79. Umetsu DT, Esserman L, Donlon TA, DeKruyff RH, Levy R: **Induction of proliferation of human follicular (B type) lymphoma cells by cognate interaction with CD4+ T cell clones. [Internet]**. *J. Immunol.* 1990, **144**:2550.
80. Scott DW, Gascoyne RD: **The tumour microenvironment in B cell lymphomas**. *Nat. Rev. Cancer* 2014, **14**:517–534.
81. Horsman DE, Gascoyne RD, Coupland RW, Coldman AJ, Adomat SA:

Comparison of Cytogenetic Analysis, Southern Analysis, and Polymerase Chain Reaction for the Detection of t(14; 18) in Follicular Lymphoma [Internet]. *Am. J. Clin. Pathol.* 1995, **103**:472–478.

82. Swaminathan S, Klemm L, Park E, Papaemmanuil E, Ford A, Kweon S, Trageser D, Hasselfeld B, Henke N, Mooster J, et al.: **Mechanisms of clonal evolution in childhood acute lymphoblastic leukemia**. *Nat. Immunol.* 2015, **16**:766–774.
83. Tsujimoto Y, Gorham J, Cossman J, Jaffe E, Croce CM: **The t(14;18) Chromosome Translocations Involved in B-Cell Neoplasms Result from Mistakes in VDJ Joining [Internet]**. *Science (80-.).* 1985, **229**:1390–1393.
84. Cui X, Lu Z, Kurosawa A, Klemm L, Bagshaw AT, Tsai AG, Gemmell N, Müschen M, Adachi N, Hsieh C-L, et al.: **Both CpG Methylation and Activation-Induced Deaminase Are Required for the Fragility of the Human bcl-2 Major Breakpoint Region: Implications for the Timing of the Breaks in the t(14;18) Translocation**. *Mol. Cell. Biol.* 2013, **33**:947–957.
85. Limpens J, Stad R, Vos C, De Viaam C, De Jong D, Van Ommen G-JB, Schuuring E, Kluin PM: **Lymphoma-Associated Translocation t(14; 18) in Blood B Cells of Normal Individuals**. *Blood* 1995, **85**:2528–2536.
86. Schüler F, Dölken L, Hirt C, Kiefer T, Berg T, Fusch G, Weitmann K, Hoffmann W, Fusch C, Janz S, et al.: **Prevalence and frequency of circulating (14;18)-MBR translocation carrying cells in healthy individuals**. *Int. J. Cancer* 2009, **124**:958–963.
87. Sungalee S, Mamessier E, Morgado E, Gregoire E, Brohawn PZ, Morehouse CA, Jouve N, Monvoisin C, Menard C, Debroas G, et al.: **Germinal center reentries of BCL2-overexpressing B cells drive follicular lymphoma progression**. *J. Clin. Invest.* 2014, **124**:5337–5351.
88. Gaulard P, D'Agay M-F, Peuchmaur M, Brousse N, Gisselbrecht C, Solal-Celigny P, Diebold J, Mason DY: **Expression of the bcl-2 Gene Product in Follicular Lymphoma**. *Am. J. Pathol.* 1992, **140**:1089–1095.
89. Tsujimoto Y, Ikegaki N, Croce CM: **Characterization of the protein product of bcl-2, the gene involved in human follicular lymphoma [Internet]**. *Oncogene* 1987, **2**:3–7.

90. Green MR, Kihira S, Liu CL, Nair R V, Salari R, Gentles AJ, Irish J, Stehr H, Vicente-Dueñas C, Romero-Camarero I, et al.: **Mutations in early follicular lymphoma progenitors are associated with suppressed antigen presentation.** *Proc Natl Acad Sci USA* 2015, **112**:E1116–E1125.
91. Li H, Kaminski M, Li Y, Yildiz M, Ouilette P, Jones S, Fox H, Jacobi K: **Mutations in linker histone genes HIST1H1 B, C, D, and E; OCT2 (POU2F2); IRF8; and ARID1A underlying the pathogenesis of follicular lymphoma**. *Blood* 2014, **123**:1487–1498.
92. Solal-Céligny P, Roy P, Colombat P, White J, Armitage JO, Arranz-Saez R, Au WY, Bellei M, Brice P, Caballero D, et al.: **Follicular lymphoma international prognostic index**. *Blood* 2004, **104**:1258–1265.
93. Pastore A, Jurinovic V, Kridel R, Hoster E, Staiger AM, Szczepanowski M, Pott C, Kopp N, Murakami M, Horn H, et al.: **Integration of gene mutations in risk prognostication for patients receiving first-line immunochemotherapy for follicular lymphoma: a retrospective analysis of a prospective clinical trial and validation in a population-based registry.** *Lancet. Oncol.* 2015, **2045**:1–12.
94. Pasqualucci L, Khiabanian H, Fangazio M, Vasishtha M, Messina M, Holmes A, Ouillette P, Trifonov V, Rossi D, Tabbò F, et al.: **Genetics of Follicular Lymphoma Transformation**. *Cell Rep.* 2014, **6**:130–140.
95. Loeffler M, Kreuz M, Haake A, Hasenclever D, Trautmann H, Arnold C, Winter K, Koch K, Klapper W, Scholtysik R, et al.: **Genomic and epigenomic co-evolution in follicular lymphomas**. *Leukemia* 2015, **29**:456–463.
96. Kridel R, Chan FC, Mottok A, Boyle M, Farinha P, Tan K, Meissner B, Bashashati A, McPherson A, Roth A, et al.: **Histological Transformation and Progression in Follicular Lymphoma: A Clonal Evolution Study**. *PloS Med.* 2016, **13**:e1002197.
97. Eide MB, Liestøl K, Lingjærde OC, Hystad ME, Kresse SH, Meza-Zepeda L, Myklebost O, Trøen G, Aamot HV, Holte H, et al.: **Genomic alterations reveal potential for higher grade transformation in follicular lymphoma and confirm parallel evolution of tumor cell clones**. *Blood* 2010, **116**:1489–1497.

98. Okosun J, Bödör C, Wang J, Araf S, Yang C-Y, Pan C, Boller S, Cittaro D, Bozek M, Iqbal S, et al.: **Integrated genomic analysis identifies recurrent mutations and evolution patterns driving the initiation and progression of follicular lymphoma.** *Nat. Genet.* 2014, **46**:176–81.
99. Rosenwald A, Wright G, Leroy K, Yu X, Gaulard P, Gascoyne RD, Chan WC, Zhao T, Haioun C, Greiner TC, et al.: **Molecular Diagnosis of Primary Mediastinal B Cell Lymphoma Identifies a Clinically Favorable Subgroup of Diffuse Large B Cell Lymphoma Related to Hodgkin Lymphoma Molecular Diagnosis of Primary Mediastinal B Cell Lymphoma [Internet]**. *J. Exp. Med.* 2003, **198**:851–86217.
100. Savage KJ, Monti S, Kutok JL, Cattoretti G, Neuberg D, De Leval L, Kurtin P, Dal Cin P, Ladd C, Feuerhake F, et al.: **The molecular signature of mediastinal large B-cell lymphoma differs from that of other diffuse large B-cell lymphomas and shares features with classical Hodgkin lymphoma**. *Blood* 2003, **102**:3871–3879.
101. Steidl C, Gascoyne RD: **The molecular pathogenesis of primary mediastinal large B-cell lymphoma.** *Blood* 2011, **118**:2659–69.
102. Wessendorf S, Barth TFE, Viardot A, Mueller A, Kestler H, Kohlhammer H, Lichter P, Bentz M, Döhner H, Möller P, et al.: **Further delineation of chromosomal consensus regions in primary mediastinal B-cell lymphomas: an analysis of 37 tumor samples using high-resolution genomic profiling (array-CGH)**. *Leukemia* 2007, **21**:2463–2469.
103. Schmitz R, Hansmann ML, Bohle V, Martín-Subero JI, Hartmann S, Mechtersheimer G, Klapper W, Vater I, Giefing M, Gesk S, et al.: **TNFAIP3 (A20) is a tumor suppressor gene in Hodgkin lymphoma and primary mediastinal B cell lymphoma [Internet]**. *J. Exp. Med* 2009, **206**:981–989.
104. Gunawardana J, Chan FC, Telenius A, Woolcock B, Kridel R, Tan KL, Ben-Neriah S, Mottok A, Lim RS, Boyle M, et al.: **Recurrent somatic mutations of PTPN1 in primary mediastinal B cell lymphoma and Hodgkin lymphoma**. *Nat. Genet.* 2014, **46**.
105. Dunleavy K, Steidl C: **Emerging Biological Insights and Novel Treatment**

Strategies in Primary Mediastinal Large B-Cell Lymphoma. *Semin. Hematol.* 2015, **52**:119–125.

106. Savage KJ, Al-Rajhi N, Voss N, Paltiel C, Klasa R, Gascoyne RD, Connors JM: **Favorable outcome of primary mediastinal large B-cell lymphoma in a single institution: The British Columbia experience**. *Ann. Oncol.* 2006, **17**:123–130.
107. Rieger M, Österborg A, Pettengell R, White D, Gill D, Walewski J, Kuhnt E, Loeffler M, Pfreundschuh M, Ho AD: **Primary mediastinal B-cell lymphoma treated with CHOP-like chemotherapy with or without rituximab: Results of the Mabthera International Trial Group study**. *Ann. Oncol.* 2011, **22**:664–670.
108. Dunleavy K, Pittaluga S, Maeda LS, Advani R, Chen CC, Hessler J, Steinberg SM, Grant C, Wright G, Varma G, et al.: **Dose-Adjusted EPOCH-Rituximab Therapy in Primary Mediastinal B-Cell Lymphoma**. *N Engl J Med* 2013, **368**:1408–16.
109. Hanahan D, Weinberg RA: **The Hallmarks of Cancer**. *Cell* 2000, **100**:57–70.
110. Hanahan D, Weinberg RA: **Hallmarks of cancer: The next generation**. *Cell* 2011, **144**:646–674.
111. Mottok A, Steidl C: **Genomic alterations underlying immune privilege in malignant lymphomas**. *Curr. Opin. Hematol.* 2015, **22**.
112. Robbins PF, Lu Y-C, El-Gamil M, Li YF, Gross C, Gartner J, Lin JC, Teer JK, Cliften P, Tycksen E, et al.: **Mining exomic sequencing data to identify mutated antigens recognized by adoptively transferred tumor-reactive T cells**. *Nat Med* 2013, **19**:747–752.
113. Matsushita H, Vesely MD, Koboldt DC, Rickert CG, Uppaluri R, Magrini VJ, Arthur CD, White JM, Chen Y-S, Shea LK, et al.: **Cancer exome analysis reveals a T-cell-dependent mechanism of cancer immunoediting**. *Nature* 2012, **482**:400–404.
114. Dupage M, Mazumdar C, Schmidt LM, Cheung AF, Jacks T: **Expression of tumour-specific antigens underlies cancer immunoediting**. *Nature* 2012, **482**:405–409.
115. Zou W: **Immunosuppressive networks in the tumour environment and their therapeutic relevance**. *Nat Rev Cancer* 2005, **5**:263–274.

116. Khong HT, Restifo NP: **Natural selection of tumor variants in the generation of "tumor escape" phenotypes.** *Nat. Immunol.* 2002, **3**:999–1005.
117. Pardoll D: **Does the immune system see tumors as foreign or self?** *Annu. Rev. Immunol.* 2003, **21**:807–839.
118. Wherry EJ: **T cell exhaustion**. *Nat. Immunol.* 2011, **131**:492–499.
119. Speiser DE, Utzschneider DT, Oberle SG, Münz C, Romero P, Zehn D: **T cell differentiation in chronic infection and cancer: functional adaptation or exhaustion?** *Nat. Rev. Immunol.* 2014, **14**:768–774.
120. Rock KL, Reits E, Neefjes J: **Present Yourself! By MHC Class I and MHC Class II Molecules [Internet]**. *Trends Immunol.* 2016, **0**:724–737.
121. The MHC sequencing consortium: **Complete sequence and genemap of a human major histocompatibility complex**. *Nature* 1999, **401**:921–923.
122. Neefjes J, M Jongsma ML, Paul P, Bakke O: **Towards a systems understanding of MHC class I and MHC class II antigen presentation**. *Nat. Rev. Immunol.* 2011, **11**:823–836.
123. Kurts C, S Robinson BW, Knolle PA: **Cross-priming in health and disease**. *Nat. Rev. Immunol.* 2010, **10**:403–414.
124. Rock KL, Shen L: **Cross-presentation: underlying mechanisms and role in immune surveillance**. *Immunol. Rev.* 2005, **207**:166–183.
125. Crotzer VL, Blum JS: **Autophagy and Its Role in MHC-Mediated Antigen Presentation [Internet]**. *J. Immunol.* 2009, **182**:3335–3341.
126. Crotzer VL, Glosson N, Zhou D, Nishino I, Blum JS: **LAMP-2-deficient human B cells exhibit altered MHC class II presentation of exogenous antigens**. *Immunology* 2010, **131**:318–330.
127. Crotzer VL, Blum JS: **Autophagy and adaptive immunity**. *Immunology* 2010, **131**:9–17.
128. Garrido F, Algarra I, García-Lora A: **The escape of cancer from T lymphocytes: immunoselection of MHC class I loss variants harboring structural-irreversible "hard" lesions**. *Cancer Immunol. Immunother.* 2010, **59**:1601–1606.
129. Bernsen M, Håkansson L, Gustafsson B, Krysander L, Rettrup B, Ruiter D,

Håkansson A: **On the biological relevance of MHC class II and B7 expression by tumour cells in melanoma metastases**. *Br. J. Cancer* 2003, **88**:424–431.

130. Atkins D, Breuckmann A, Schmahl GE, Binner P, Ferrone S, Krummenauer F, Störkel S, Seliger B: **MHC class I antigen processing pathway defects, RAS mutations and disease stage in colorectal carcinoma**. *Int. J. Cancer* 2004, **109**:265–273.
131. Ferrone S, Marincola FM: **Loss of HLA class I antigens by melanoma cells: molecular mechanisms, functional significance and clinical relevance**. *Immunol. Today* 1995, **16**:487–494.
132. Angell TE, Lechner MG, Jang JK, LoPresti JS, Epstein AL: **MHC class I loss is a frequent mechanism of immune escape in papillary thyroid cancer that is reversed by interferon and selumetinib treatment in Vitro**. *Clin. Cancer Res.* 2014, **20**:6034–6044.
133. Roberts RA, Wright G, Rosenwald A, Jaramillo MA, Grogan TM, Miller TP, Frutiger Y, Chan WC, Gascoyne RD, Ott G, et al.: **Loss of major histocompatibility class II gene and protein expression in primary mediastinal large B-cell lymphoma is highly coordinated and related to poor patient survival.** *Blood* 2006, **108**:311–8.
134. Diepstra A, van Imhoff GW, Karim-Kos HE, van den Berg A, te Meerman GJ, Niens M, Nolte IM, Bastiaannet E, Schaapveld M, Vellenga E, et al.: **HLA class II expression by Hodgkin Reed-Sternberg cells is an independent prognostic factor in classical Hodgkin's lymphoma.** *J. Clin. Oncol.* 2007, **25**:3101–8.
135. Seliger B: **Molecular mechanisms of MHC class I abnormalities and APM components in human tumors**. *Cancer Immunol. Immunother.* 2008, **57**:1719–1726.
136. Riemersma SA, Jordanova ES, Haasnoot GW, Drabbels J, Schuuring E, Schreuder GMT, Kluin PM: **The Relationship Between HLA Class II Polymorphisms and Somatic Deletions in Testicular B Cell Lymphomas of Dutch Patients**. *Hum. Immunol.* 2006, **67**:303–310.
137. Booman M, Douwes J, Glas AM, Riemersma SA, Jordanova ES, Kok K, Rosenwald A, de Jong D, Schuuring E, Kluin PM: **Mechanisms and effects of**

loss of human leukocyte antigen class II expression in immune-privileged site-associated B-cell lymphoma. *Clin. Cancer Res.* 2006, **12**:2698–2705.

138. Riemersma SA, Jordanova ES, Schop RF, Philippo K, Looijenga LH, Schuuring E, Kluin PM: **Extensive genetic alterations of the HLA region, including homozygous deletions of HLA class II genes in B-cell lymphomas arising in immune-privileged sites.** *Blood* 2000, **96**:3569–3577.
139. Jordanova ES, Riemersma SA, Philippo K, Giphart-Gassler M, Schuuring E, Kluin PM: **Hemizygous deletions in the HLA region account for loss of heterozygosity in the majority of diffuse large B-cell lymphomas of the testis and the central nervous system**. *Genes, Chromosom. Cancer* 2002, **35**:38–48.
140. Vijai J, Wang Z, Berndt SI, Skibola CF, Slager SL, de Sanjose S, Melbye M, Glimelius B, Bracci PM, Conde L, et al.: **A genome-wide association study of marginal zone lymphoma shows association to the HLA region**. *Nat Commun* 2015, **6**.
141. Conde L, Halperin E, Akers NK, Brown KM, Smedby KE, Rothman N, Nieters A, Slager SL, Brooks-Wilson A, Agana L, et al.: **Genome-wide association study of follicular lymphoma identifies a risk locus at 6p21.32**. *Nat. Genet.* 2010, **42**:661–664.
142. Cerhan JR, Berndt SI, Vijai J, Ghesquieres H, McKay J, Wang SS, Wang Z, Yeager M, Conde L, de Bakker PIW, et al.: **Genome-wide association study identifies multiple susceptibility loci for diffuse large B cell lymphoma**. *Nat. Genet.* 2014, **46**:1233–1238.
143. Enciso-Mora V, Broderick P, Ma Y, Jarrett RF, Hjalgrim H, Hemminki K, van den Berg A, Olver B, Lloyd A, Dobbins SE, et al.: **A genome-wide association study of Hodgkin's lymphoma identifies new susceptibility loci at 2p16.1 (REL), 8q24.21 and 10p14 (GATA3)**. *Nat Genet* 2010, **42**:1126–1130.
144. Cresswell P, Turner MJ, Strominger JL: **Papain-Solubilized HL-A Antigens from Cultured Human Lymphocytes Contain Two Peptide Fragments**. *Proc Natl Acad Sci USA* 1973, **70**:1603–1607.
145. Grey HM, Kubo RT, Colon SM, Poulik MD, Cresswell P, Springer T, Turner M,

Strominger JL: **The small subunit of HL-A antigens is ß2-microglobulin**. *J. Exp. Med.* 1973, **138**:1608–1612.

146. D'Urso CM, Wang ZG, Cao Y, Tatake R, Zeff RA, Ferrone S: **Lack of HLA class I antigen expression by cultured melanoma cells FO-1 due to a defect in B2m gene expression.** *J. Clin. Invest.* 1991, **87**:284–292.
147. Edidin M, Achilles S, Zeff R, Wei T: **Probing the stability of class I major histocompatibility complex (MHC) molecules on the surface of human cells**. *Immunogenetics* 1997, **46**:41–45.
148. Wang Z, Arienti F, Parmiani G, Ferrone S: **Induction and functional characterization of β 2-microglobulin (β2-μ)-free HLA class I heavy chains expressed by β2-μ-deficient human FO-1 melanoma cells**. *Eur. J. Immunol.* 1998, **28**:2817–2826.
149. Maleno I, Aptsiauri N, Cabrera T, Gallego A, Paschen A, López-Nevot MA, Garrido F: **Frequent loss of heterozygosity in the β2-microglobulin region of chromosome 15 in primary human tumors**. *Immunogenetics* 2011, **63**:65–71.
150. Kloor M, Michel S, Buckowitz B, Rüschoff J, Büttner R, Holinski-Feder E, Dippold W, Wagner R, Tariverdian M, Benner A, et al.: **Beta2-microglobulin mutations in microsatellite unstable colorectal tumors**. *Int. J. Cancer* 2007, **121**:454–458.
151. Koopman LA, Corver WE, van der Slik AR, Giphart MJ, Fleuren GJ: **Multiple Genetic Alterations Cause Frequent and Heterogeneous Human Histocompatibility Leukocyte Antigen Class I Loss in Cervical Cancer**. *J. Exp. Med.* 2000, **191**:961–976.
152. Chang C-C, Campoli M, Restifo NP, Wang X, Ferrone S: **Immune Selection of Hot-Spot β2-Microglobulin Gene Mutations, HLA-A2 Allospecificity Loss, and Antigen-Processing Machinery Component Down-Regulation in Melanoma Cells Derived from Recurrent Metastases following Immunotherapy**. *J. Immunol.* 2005, **174**:1462–1471.
153. Jordanova ES, Riemersma SA, Philippo K, Schuuring E, Kluin PM: **Beta2-microglobulin aberrations in diffuse large B-cell lymphoma of the testis and the central nervous system.** *Int. J. Cancer* 2003, **103**:393–8.

154. Challa-Malladi M, Lieu YK, Califano O, Holmes AB, Bhagat G, Murty V V, Dominguez-Sola D, Pasqualucci L, Dalla-Favera R: **Combined genetic inactivation of β2-Microglobulin and CD58 reveals frequent escape from immune recognition in diffuse large B cell lymphoma.** *Cancer Cell* 2011, **20**:728–40.
155. Beà S, Valdés-Mas R, Navarro A, Salaverria I, Martín-Garcia D, Jares P, Giné E, Pinyol M, Royo C, Nadeu F, et al.: **Landscape of somatic mutations and clonal evolution in mantle cell lymphoma.** *Proc Natl Acad Sci USA* 2013, **110**:18250–18255.
156. Shaw S, Ginther Luce GE, Quinones R, Gress RE, Springer TA, Sanders ME: **Two antigen-independent adhesion pathways used by human cytotoxic T-cell clones**. *Nature* 1986, **323**:262–264.
157. Petrányi GG, Pócsik E, Kotlán B, Görög G, Benczur M: **Regulatory function of cell surface molecules CD2-, LFA- and β2-microglobulin in natural killer cell activity**. *Mol. Immunol.* 1986, **23**:1275–1279.
158. Karre K, Ljunggren HG, Piontek G, Kiessling R: **Selective rejection of H-2-deficient lymphoma variants suggests alternative immune defence strategy**. *Nature* 1986, **319**:675–678.
159. Reichel J, Chadburn A, Rubinstein PG, Giulino-Roth L, Tam W, Liu Y, Gaiolla R, Eng K, Brody J, Inghirami G, et al.: **Flow sorting and exome sequencing reveal the oncogenome of primary Hodgkin and Reed-Sternberg cells**. *Blood* 2014, **125**:1061–1072.
160. Liu Y, Abdul Razak FR, Terpstra M, Chan FC, Saber A, Nijland M, van Imhoff G, Visser L, Gascoyne R, Steidl C, et al.: **The mutational landscape of Hodgkin lymphoma cell lines determined by whole-exome sequencing**. *Leukemia* 2014, **28**:2248–2251.
161. Schneider M, Schneider S, Zühlke-Jenisch R, Klapper W, Sundström C, Hartmann S, Hansmann ML, Siebert R, Küppers R, Giefing M: **Alterations of the CD58 gene in classical Hodgkin lymphoma**. *Genes Chromosom. Cancer* 2015, **54**:638–645.
162. Abdul Razak F, Diepstra A, Visser L, van den Berg A: **CD58 mutations are**

common in Hodgkin lymphoma cell lines and loss of CD58 expression in tumor cells occurs in Hodgkin lymphoma patients who relapse. *Genes Immun.* 2016, **17**:363–366.

163. Möller P, Herrmann B, Moldenhauer G, Momburg F: **Defective expression of MHC class I antigens is frequent in B-cell lymphomas of high-grade malignancy**. *Int. J. Cancer* 1987, **40**:32–39.
164. Riemersma SA, Oudejans JJ, Vonk MJ, Dreef EJ, Prins FA, Jansen PM, Vermeer MH, Blok P, Kibbelaar RE, Muris JJF, et al.: **High numbers of tumour-infiltrating activated cytotoxic T lymphocytes, and frequent loss of HLA class I and II expression, are features of aggressive B cell lymphomas of the brain and testis**. *J. Pathol.* 2005, **206**:328–336.
165. Meissner TB, Liu Y-J, Lee K-H, Li A, Biswas A, van Eggermond MCJ a, van den Elsen PJ, Kobayashi KS: **NLRC5 cooperates with the RFX transcription factor complex to induce MHC class I gene expression.** *J. Immunol.* 2012, **188**:4951–8.
166. Neerincx A, Rodriguez GM, Steimle V, Kufer T a: **NLRC5 controls basal MHC class I gene expression in an MHC enhanceosome-dependent manner.** *J. Immunol.* 2012, **188**:4940–50.
167. Neerincx A, Jakobshagen K, Utermöhlen O, Büning H, Steimle V, Kufer TA: **The N-Terminal Domain of NLRC5 Confers Transcriptional Activity for MHC Class I and II Gene Expression**. *J. Immunol.* 2014, **193**:3090–3100.
168. Schmitz R, Young RM, Ceribelli M, Jhavar S, Xiao W, Zhang M, Wright G, Shaffer AL, Hodson DJ, Buras E, et al.: **Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics.** *Nature* 2012, **490**:116–20.
169. Scholtysik R, Kreuz M, Hummel M, Rosolowski M, Szczepanowski M, Klapper W, Loeffler M, Trümper L, Siebert R, Küppers R, et al.: **Characterization of genomic imbalances in diffuse large B-cell lymphoma by detailed SNP-chip analysis**. *Int. J. Cancer* 2015, **136**:1033–1042.
170. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, et al.: **The cBio Cancer Genomics Portal: An**

Open Platform for Exploring Multidimensional Cancer Genomics Data [Internet]. *Cancer Discov.* 2012, **2**:401–404.

171. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, et al.: **Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal [Internet]**. *Sci. Signal.* 2013, **6**:pl1.
172. Steidl C, Shah SP, Woolcock BW, Rui L, Kawahara M, Farinha P, Johnson NA, Zhao Y, Telenius A, Ben-Neriah S, et al.: **MHC class II transactivator CIITA is a recurrent gene fusion partner in lymphoid cancers**. *Nature* 2011, **471**:377–380.
173. de Miranda NFCC, Georgiou K, Chen L, Wu C, Gao Z, Zaravinos A, Lisboa S, Enblad G, Teixeira MR, Zeng Y, et al.: **Exome sequencing reveals novel mutation targets in diffuse large B-cell lymphomas derived from Chinese patients**. *Blood* 2014, **124**:2544–2553.
174. Reith W, LeibundGut-Landmann S, Waldburger JM: **Regulation of MHC class II gene expression by the Class II Transactivator**. *Nat. Rev. Immunol.* 2005, **5**:793–806.
175. Ting JP-Y, Trowsdale J: **Genetic Control of MHC Class II Expression**. *Cell* 2002, **109**:S21–S33.
176. Noelle R, Krammer PH, Ohara J, Uhr JW, Vitetta ES: **Increased expression of Ia antigens on resting B cells: An additional role for B-cell growth factor**. *Proc Natl Acad Sci USA* 1984, **81**:6149–6153.
177. Pai R, Askew D, Boom H, Harding C V: **Regulation of Class II MHC Expression in APCs: Roles of Types I, III, and IV Class II Transactivator [Internet]**. *J Immunol* 2002, **169**:1326–1333.
178. Halper J, Fu SM, Wang CY, Winchester R, Kunkel HG: **Patterns of Expression of Human "Ia-Like" Antigens during the Terminal Stages of B Cell Development [Internet]**. *J. Immunol.* 1978, **120**:1480.
179. Piskurich JF, Gilbert CA, Ashley BD, Zhao M, Chen H, Wu J, Bolick SC, Wright KL: **Expression of the MHC class II transactivator (CIITA) type IV promoter in B lymphocytes and regulation by IFN-γ**. *Mol. Immunol.* 2006, **43**:519–528.

180. Boss JM, Jensen PE: **Transcriptional regulation of the MHC class II antigen presentation pathway**. *Curr. Opin. Immunol.* 2003, **15**:105–111.
181. Gosh P, Amaya M, Mellins E, Wiley DC: **The structure of an intermediate in class II MHC maturation: CLIP bound to HLA-DR3**. *Nature* 1995, **378**:457–462.
182. Klein C, Lisowska-Grospierre B, LeDeist F, Fischer A, Griscelli C: **Major histocompatibility complex class deficiency: Clinical manifestations, immunologic features, and outcome**. *J. Pediatr.* 1993, **123**:921–928.
183. Reith W, Mach B: **THE BARE LYMPHOCYTE SYNDROME AND THE REGULATION OF MHC EXPRESSION**. *Annu. Rev. Immunol* 2001, **19**:331–373.
184. Scholl T, Mahanta SK, Strominger JL: **Specific complex formation between the type II bare lymphocyte syndrome-associated transactivators CIITA and RFX5**. *Proc. Natl. Acad. Sci.* 1997, **94**:6330–6334.
185. Moreno CS, Beresford GW, Louis-Plence P, Morris AC, Boss JM: **CREB Regulates MHC Class II Expression in a CIITA-Dependent Manner**. *Immunity* 1999, **10**:143–151.
186. Mantovani R: **The molecular biology of the CCAAT-binding factor NF-Y [Internet]**. *Gene* 1999, **239**:15–27.
187. Muhlethaler-Mottet A, Krawczyk M, Masternak K, Spilianakis C, Kretsovali A, Papamatheakis J, Reith W: **The S box of major histocompatibility complex class II promoters is a key determinant for recruitment of the transcriptional co-activator CIITA**. *J. Biol. Chem.* 2004, **279**:40529–40535.
188. Masternak K, Muhlethaler-Mottet A, Villard J, Zufferey M, Steimle V, Reith W: **CIITA is a transcriptional coactivator that is recruited to MHC class II promoters by multiple synergistic interactions with an enhanceosome complex**. *Genes Dev.* 2000, **14**:1156–1166.
189. Steimle V, Otten LA, Zufferey M, Mach B: **Complementation cloning of an MHC class II transactivator mutated in hereditary MHC class II deficiency (or bare lymphocyte syndrome)**. *Cell* 1993, **75**:135–146.
190. Nagarajan UM, Louis-Plence P, Desandro A, Nilsen R, Bushey A, Boss JM: **RFX-B Is the Gene Responsible for the Most Common Cause of the Bare Lymphocyte Syndrome, an MHC Class II Immunodeficiency**. *Immunity* 1999,

10:153–162.

191. Masternak K, Barras E, Zufferey M, Conrad B, Corthals G, Aebersold R, Sanchez J-C, Hochstrasser DF, Mach B, Reith W: **A gene encoding a novel RFX-associated transactivator is mutated in the majority of MHC class II deficiency patients**. *Nat. Genet.* 1998, **20**:273–277.
192. Steimle V, Durand B, Barras E, Zufferey M, Hadam MR, Mach B, Reith W: **A novel DNA-binding regulatory factor is mutated in primary MHC class II deficiency (bare lymphocyte syndrome)**. *Genes Dev.* 1995, **9**:1021–1032.
193. Durand B, Sperisen P, Emery P, Barras E, Zufferey M, Mach B, Reith W: **RFXAP, a novel subunit of the RFX DNA binding complex is mutated in MHC class II deficiency**. *EMBO J.* 1997, **16**:1045–1055.
194. Kara CJ, Glimcher LH: **In vivo footprinting of MHC class II genes: bare promoters in the bare lymphocyte syndrome.** *Science (80-.).* 1991, **252**:709–712.
195. LeibundGut-Landmann S, Waldburger JM, Krawczyk M, Otten LA, Suter T, Fontana A, Acha-Orbea H, Reith W: **Specificity and expression of CIITA, the master regulator of MHC class II genes**. *Eur. J. Immunol.* 2004, **34**:1513–1525.
196. Muhlethaler-Mottet A, Otten LA, Steimle V, Mach B: **Expression of MHC class II molecules in different cellular and functional compartments is controlled by differential usage of multiple promoters of the transactivator CIITA**. *EMBO J.* 1997, **16**:2851–2860.
197. Geppert TD, Lipsky PE: **Antigen presentation by interferon-gamma-treated endothelial cells and fibroblasts: differential ability to function as antigen-presenting cells despite comparable Ia expression. [Internet]**. *J. Immunol.* 1985, **135**:3750.
198. Waldburger J-M, Suter T, Fontana A, Acha-Orbea H, Reith W: **Selective Abrogation of Major Histocompatibility Complex Class II Expression on Extrahematopoietic Cells in Mice Lacking Promoter IV of the Class II Transactivator Gene [Internet]**. *J. Exp. Med* 2001, **83931400**:393–406.
199. Mulder DJ, Pooni A, Mak N, Hurlbut DJ, Basta S, Justinich CJ: **Antigen Presentation and MHC Class II Expression by Human Esophageal Epithelial**

Cells: Role in Eosinophilic Esophagitis. *Am. J. Pathol.* 2011, **178**:744–753.

200. Kreisel D, Richardson SB, Li W, Lin X, Kornfeld CG, Sugimoto S, Hsieh C-S, Gelman A, Krupnick AS: **Cutting Edge: MHC Class II Expression by Pulmonary Nonhematopoietic Cells Plays a Critical Role in Controlling Local Inflammatory Responses**. *J. Immunol.* 2010, **185**:3809–3813.
201. Fritz JH, Ferrero RL, Philpott DJ, Girardin SE: **Nod-like proteins in inflammation and disease**. *Nat. Immunol.* 2006, **7**:1250–1257.
202. Zika E, Ting JPY: **Epigenetic control of MHC-II: Interplay between CIITA and histone-modifying enzymes**. *Curr. Opin. Immunol.* 2005, **17**:58–64.
203. Mudhasani R, Fontes JD: **The Class II Transactivator Requires brahma-Related Gene 1 To Activate Transcription of Major Histocompatibility Complex Class II Genes**. *Mol. Cell. Biol.* 2002, **22**:5019–5026.
204. Harton JA, Cressman DE, Chin K-C, Der CJ, Ting JP-Y: **GTP Binding by Class II Transactivator : Role in Nuclear Import**. *Science (80-.).* 1998, **285**:1402–1405.
205. Bewry NN, Bolick SCE, Wright KL, Harton JA: **GTP-dependent recruitment of CIITA to the class II major histocompatibility complex promoter**. *J. Biol. Chem.* 2007, **282**:26178–26184.
206. Camacho-Carvajal MM, Wollscheid B, Aebersold R, Steimle V, Schamel W: **Importance of class II transactivator leucine-rich repeats for dominant-negative function and nucleo-cytoplasmic transport**. *Int. Immunol.* 2004, **16**:65–75.
207. Harton JA, O'Connor W, Conti BJ, Linhoff MW, Ting JPY: **Leucine-rich repeats of the class II transactivator control its rate of nuclear accumulation**. *Hum. Immunol.* 2002, **63**:588–601.
208. Linhoff MW, Harton JA, Cressman DE, Martin BK, Ting JP-Y: **Two Distinct Domains within CIITA Mediate Self-Association : Involvement of the GTP-Binding and Leucine-Rich Repeat Domains**. *Mol. Cell. Biol.* 2001, **21**:3001–3011.
209. Hake SB, Masternak K, Kammerbauer C, Janzen C, Reith W, Steimle V: **CIITA Leucine-Rich Repeats Control Nuclear Localization, In Vivo Recruitment to the Major Histocompatibility Complex (MHC) Class II Enhanceosome, and**

MHC Class II Gene Transactivation. *Mol. Cell. Biol.* 2000, **20**:7716–7725.

210. Steidl C, Lee T, Shah SP, Farinha P, Han G, Nayar T, Delaney A, Jones SJ, Iqbal J, Weisenburger DD, et al.: **Tumor-Associated Macrophages and Survival in Classic Hodgkin's Lymphoma**. *N. Engl. J. Med.* 2010, **362**:875–885.
211. Aldinucci D, Celegato M, Casagrande N: **Microenvironmental interactions in classical Hodgkin lymphoma and their role in promoting tumor growth, immune escape and drug resistance**. *Cancer Lett.* 2016, **380**:243–252.
212. Green MR, Monti S, Rodig SJ, Juszczynski P, Currie T, O'Donnell E, Chapuy B, Takeyama K, Neuberg D, Golub TR, et al.: **Integrative analysis reveals selective 9p24.1 amplification, increased PD-1 ligand expression, and further induction via JAK2 in nodular sclerosing Hodgkin lymphoma and primary mediastinal large B-cell lymphoma**. *Blood* 2010, **116**:3268–3277.
213. Juszczynski P, Ouyang J, Monti S, Rodig SJ, Takeyama K, Abramson J, Chen W, Kutok JL, Rabinovich GA, Shipp MA: **The AP1-dependent secretion of galectin-1 by Reed–Sternberg cells fosters immune privilege in classical Hodgkin lymphoma**. *Proc Natl Acad Sci USA* 2007, **104**:13134–13139.
214. Chong LC, Twa DDW, Mottok A, Ben-Neriah S, Woolcock BW, Zhao Y, Savage KJ, Marra MA, Scott DW, Gascoyne RD, et al.: **Comprehensive characterization of programmed death ligand structural rearrangements in B-cell non-Hodgkin lymphomas**. *Blood* 2016, **128**:1206–1213.
215. Rimsza LM, Roberts RA, Miller TP, Unger JM, LeBlanc M, Braziel RM, Weisenberger DD, Chan WC, Muller-Hermelink HK, Jaffe ES, et al.: **Loss of MHC class II gene and protein expression in diffuse large B-cell lymphoma is related to decreased tumor immunosurveillance and poor patient survival regardless of other prognostic factors: a follow-up study from the Leukemia and Lymphoma Molecular**. *Blood* 2004, **103**:4251–8.
216. Twa DDW, Chan FC, Ben-Neriah S, Woolcock BW, Mottok A, Tan KL, Slack GW, Gunawardana J, Lim RS, McPherson AW, et al.: **Genomic rearrangements involving programmed death ligands are recurrent in primary mediastinal large B-cell lymphoma**. *Blood* 2014, **123**:2062–2065.
217. Twa DDW, Mottok A, Chan FC, Ben-Neriah S, Woolcock BW, Tan KL, Mungall

AJ, McDonald H, Zhao Y, Lim RS, et al.: **Recurrent genomic rearrangements in primary testicular lymphoma**. *J. Pathol.* 2015, **236**:136–141.

218. Bentz M, Barth TFE, Brüderlein S, Bock D, Schwerer MJ, Baudis M, Joos S, Viardot A, Feller AC, Müller-Hermelink H-K, et al.: **Gain of chromosome arm 9p is characteristic of primary mediastinal B-cell lymphoma (MBL): Comprehensive molecular cytogenetic analysis and presentation of a novel MBL cell line**. *Genes, Chromosom. Cancer* 2001, **30**:393–401.
219. Kridel R, Mottok A, Farinha P, Ben-Neriah S, Ennishi D, Zheng Y, Chavez EA, Shulha HP, Tan K, Chan FC, et al.: **Cell of origin of transformed follicular lymphoma**. *Blood* 2015, **126**:2118–2127.
220. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SFA, Hakonarson H, Bucan M: **PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data**. *Genome Res.* 2007, **17**:1665–1674.
221. Yau C, Mouradov D, Jorissen R, Colella S, Mirza G, Steers G, Harris A, Ragoussis J, Sieber O, Holmes C: **A statistical approach for detecting genomic aberrations in heterogeneous tumor samples from single nucleotide polymorphism genotyping data**. *Genome Biol.* 2010, **11**:R92.
222. Langmead B, Schatz M, Lin J, Pop M, Salzberg S: **Searching for SNPs with cloud computing**. *Genome Biol.* 2009, **10**:R134.
223. Lohr JG, Stojanov P, Lawrence MS, Auclair D, Chapuy B, Sougnez C, Cruz-Gordillo P, Knoechel B, Asmann YW, Slager SL, et al.: **Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing**. *Proc Natl Acad Sci USA* 2012, **109**:3879–3884.
224. Love C, Sun Z, Jima D, Li G, Zhang J, Miles R, Richards KL, Dunphy CH, Choi WWL, Srivastava G, et al.: **The genetic landscape of mutations in Burkitt lymphoma.** *Nat. Genet.* 2012, **44**:1321–5.
225. Richter J, Schlesner M, Hoffmann S, Kreuz M, Leich E, Burkhardt B, Rosolowski M, Ammerpohl O, Wagener R, Bernhart SH, et al.: **Recurrent mutation of the ID3 gene in Burkitt lymphoma identified by integrated genome, exome and**

transcriptome sequencing. *Nat. Genet.* 2012, **44**:1316–20.

226. Ding J, Bashashati A, Roth A, Oloumi A, Tse K, Zeng T, Haffari G, Hirst M, Marra MA, Condon A, et al.: **Feature-based classifiers for somatic mutation detection in tumour-normal paired sequencing data**. *Bioinformatics* 2012, **28**:167–175.
227. McPherson A, Wu C, Wyatt AW, Shah S, Collins C, Sahinalp SC: **nFuse: Discovery of complex genomic rearrangements in cancer using high-throughput sequencing**. *Genome Res.* 2012, **22**:2250–2261.
228. Mestre C, Rubio-Moscardo F, Rosenwald A, Climent J, Dyer MJS, Staudt L, Pinkel D, Siebert R, Martinez-Climent JA: **Homozygous deletion of SOCS1 in primary mediastinal B-cell lymphoma detected by CGH to BAC microarrays**. *Leukemia* 2005, **19**:1082–1084.
229. Melzner I, Weniger MA, Bucur AJ, Brüderlein S, Dorsch K, Hasel C, Leithäuser F, Ritz O, Dyer MJS, Barth TFE, et al.: **Biallelic deletion within 16p13.13 including SOCS-1 in Karpas1106P mediastinal B-cell lymphoma line is associated with delayed degradation of JAK2 protein.** *Int. J. cancer* 2006, **118**:1941–4.
230. Li J, Jorgensen SF, Maggadottir SM, Bakay M, Warnatz K, Glessner J, Pandey R, Salzer U, Schmidt RE, Perez E, et al.: **Association of CLEC16A with human common variable immunodeficiency disorder and role in murine B cells**. *Nat Commun* 2015, **20**:6804.
231. Scott DW, Wright GW, Williams PM, Lih C-J, Walsh W, Jaffe ES, Rosenwald A, Campo E, Chan WC, Connors JM, et al.: **Determining cell-of-origin subtypes of diffuse large B-cell lymphoma using gene expression in formalin-fixed paraffin-embedded tissue**. *Blood* 2014, **123**:1214–1217.
232. Mottok A, Woolcock B, Chan FC, Tong KM, Chong L, Farinha P, Telenius A, Chavez E, Ramchandani S, Drake M, et al.: **Genomic Alterations in CIITA Are Frequent in Primary Mediastinal Large B Cell Lymphoma and Are Associated with Diminished MHC Class II Expression**. *Cell Rep.* 2015, **13**:1418–1431.
233. Lohsen S, Majumder P, Scharer C, Barwick B, Austin J, Zinzow-Kramer W, Boss

J: **Common distal elements orchestrate CIITA isoform-specific expression in multiple cell types**. *Genes Immun.* 2014, **15**:543–555.

234. Khodabakhshi A, Morin RD, Fejes AP, Mungall AJ, Mungall KL, Bolger-Munro M, Johnson NA, Connors JM, Gascoyne RD, Marra MA, et al.: **Recurrent targets of aberrant somatic hypermutation in lymphoma**. *Oncotarget* 2012, **3**:1308–1319.
235. Müschen M, Re D, Jungnickel B, Diehl V, Rajewsky K, Küppers R: **Somatic Mutation of the CD95 Gene in Human B Cells as a Side-Effect of the Germinal Center Reaction**. *J. Exp. Med.* 2000, **192**:1833–1840.
236. Mottok A, Renné C, Willenbrock K, Hansmann M-L, Bräuninger A: **Somatic hypermutation of SOCS1 in lymphocyte-predominant Hodgkin lymphoma is accompanied by high JAK2 expression and activation of STAT6**. *Blood* 2007, **110**:3387–3390.
237. Mottok A, Renné C, Seifert M, Oppermann E, Bechstein W, Hansmann M-L, Küppers R, Bräuninger A: **Inactivating SOCS1 mutations are caused by aberrant somatic hypermutation and restricted to a subset of B-cell lymphoma entities**. *Blood* 2009, **114**:4503–4506.
238. Jiang Y, Soong TD, Wang L, Melnick AM, Elemento O: **Genome-wide detection of genes targeted by non-Ig somatic hypermutation in lymphoma.** *PLoS One* 2012, **7**:e40332.
239. Bödör C, Bognár Á, Reiniger L, Szepesi Á, Tóth E, Kopper L, Matolcsy A: **Aberrant somatic hypermutation and expression of activation-induced cytidine deaminase mRNA in mediastinal large B-cell lymphoma**. *Br. J. Haematol.* 2005, **129**:373–376.
240. Popov SW, Moldenhauer G, Wotschke B, Brüderlein S, Barth TF, Dorsch K, Ritz O, Möller P, Leithäuser F: **Target sequence accessibility limits activation-induced cytidine deaminase activity in primary mediastinal B-cell lymphoma.** *Cancer Res.* 2007, **67**:6555–64.
241. Masternak K, Reith W: **Promoter-specific functions of CIITA and the MHC class II enhanceosome in transcriptional activation**. *EMBO J.* 2002, **21**:1379–1388.

242. Bernd HW, Ziepert M, Thorns C, Klapper W, Wacker HH, Hummel M, Stein H, Hansmann ML, Ott G, Rosenwald A, et al.: **Loss of HLA-DR expression and immunoblastic morphology predict adverse outcome in diffuse large B-cell lymphoma - Analyses of cases from two prospective randomized clinical trials**. *Haematologica* 2009, **94**:1569–1580.
243. Higashi M, Tokuhira M, Fujino S, Yamashita T, Abe K, Arai E, Kizaki M, Tamaru J: **Loss of HLA DR expression is related to tumor microenvironment and predicts adverse outcome in diffuse large B cell lymphoma**. *Leuk. Lymphoma* 2016, **57**:161–166.
244. Medeiros LJ, Gelb AB, Wolfson K, Doggett R, Mcgregor B, Cox RS, Horning SJ, Warnket RA: **Major Histocompatibility Complex Class I and Class II Antigen Expression in Diffuse Large Cell and Large Cell Immunoblastic Lymphomas: Absence of a Correlation between Antigen Expression and Clinical Outcome**. *Am. J. Pathol.* 1993, **143**:1086–1097.
245. Ngo VN, Davis RE, Lamy L, Yu X, Zhao H, Lenz G, Lam LT, Dave S, Yang L, Powell J, et al.: **A loss-of-function RNA interference screen for molecular targets in cancer**. *Nature* 2006, **441**:106–110.
246. Cycon KA, Clements JL, Holtz R, Fuji H, Murphy SP: **The immunogenicity of L1210 lymphoma clones correlates with their ability to function as antigen-presenting cells**. *Immunology* 2009, **128**:e641–e651.
247. Bontron S, Ucla C, Mach B, Steimle V: **Efficient repression of endogenous major histocompatibility complex class II expression through dominant negative CIITA mutants isolated by a functional selection strategy.** *Mol. Cell. Biol.* 1997, **17**:4249–4258.
248. Armand P, Nagler A, Weller EA, Devine SM, Avigan DE, Chen Y-B, Kaminski MS, Kent Holland H, Winter JN, Mason JR, et al.: **Disabling Immune Tolerance by Programmed Death-1 Blockade With Pidilizumab After Autologous Hematopoietic Stem-Cell Transplantation for Diffuse Large B-Cell Lymphoma: Results of an International Phase II Trial**. *J. Clin. Oncol.* 2013, **31**:4199–4206.
249. Mortara L, Frangione V, Castellani P, De Lerma Barbaro A, Accolla RS:

Irradiated CIITA-positive mammary adenocarcinoma cells act as a potent anti-tumor-preventive vaccine by inducing tumor-specific CD4+ T cell priming and CD8+ T cell effector functions. *Int. Immunol.* 2009, **21**:655–665.

250. Greaves MF, Maley CC: **Clonal evolution in cancer**. *Nature* 2012, **481**:306–313.
251. Andor N, Graham TA, Jansen M, Xia LC, Athena Aktipis C, Petritsch C: **Pan-cancer analysis of the extent and consequences of intratumor heterogeneity**. *Nat. Med.* 2015, **22**:105–113.
252. Pasqualucci L, Dominguez-Sola D, Chiarenza A, Fabbri G, Grunn A, Trifonov V, Kasper LH, Lerach S, Tang H, Ma J, et al.: **Inactivating mutations of acetyltransferase genes in B-cell lymphoma [Internet]**. *Nature* 2011, **471**:189–195.
253. Brown P, Wong K, Felce S, Lyne L, Spearman H, Soilleux E, Pedersen L, Møller M, Green T, Gascoyne D, et al.: **FOXP1 suppresses immune response signatures and MHC class II expression in activated B-cell-like diffuse large B-cell lymphomas**. *Leukemia* 2016, **30**:605–616.
254. O'Keefe GM, Nguyen VT, Tang LP, Benveniste EN: **IFN-γ Regulation of Class II Transactivator Promoter IV in Macrophages and Microglia: Involvement of the Suppressors of Cytokine Signaling-1 Protein**. *J. Immunol.* 2001, **166**:2260–2269.
255. O 'Keefe GM, Nguyen VT, Benveniste EN: **Class II transactivator and class II MHC gene expression in microglia: modulation by the cytokines TGF-I , IL-4, IL-13 and IL-10**. *Eur. J. Immunol.* 1999, **29**:1275–1285.
256. Snyder A, Makarov V, Merghoub T, Yuan J, Zaretsky JM, Desrichard A, Walsh LA, Postow MA, Wong P, Ho TS, et al.: **Genetic Basis for Clinical Response to CTLA-4 Blockade in Melanoma**. *N. Engl. J. Med.* 2014, **371**:2189–2199.
257. Ansell SM, Lesokhin AM, Borrello I, Halwani A, Scott EC, Gutierrez M, Schuster SJ, Millenson MM, Cattry D, Freeman GJ, et al.: **PD-1 Blockade with Nivolumab in Relapsed or Refractory Hodgkin's Lymphoma**. *N. Engl. J. Med.* 2014, **372**:311–319.
258. Westin JR, Chu F, Zhang M, Fayad LE, Kwak LW, Fowler N, Romaguera J, Hagemeister F, Fanale M, Samaniego F, et al.: **Safety and activity of PD1**

blockade by pidilizumab in combination with rituximab in patients with relapsed follicular lymphoma: a single group, open-label, phase 2 trial. *Lancet Oncol.* 2014, **15**:69–77.

259. Johnson DB, Estrada M V, Salgado R, Sanchez V, Doxie DB, Opalenik SR, Vilgelm AE, Feld E, Johnson AS, Greenplate AR, et al.: **Melanoma-specific MHC-II expression represents a tumour-autonomous phenotype and predicts response to anti-PD-1/PD-L1 therapy**. *Nat. Commun.* 2016, **7**:10582.
260. Armstrong TD, Clements VK, Martin BK, Ting JP-Y, Ostrand-Rosenberg S: **Major histocompatibility complex class II-transfected tumor cells present endogenous antigen and are potent inducers of tumor-specific immunity**. *Proc Natl Acad Sci USA* 1997, **94**:6886–6891.
261. Armstrong TD, Clements VK, Ostrand-Rosenberg S: **Class II-Transfected Tumor Cells Directly Present Endogenous Antigen to CD4+ T Cells In Vitro and Are APCs for Tumor-Encoded Antigens In Vivo**. *J Immunother* 1998, **21**:218–224.
262. Meazza R, Comes A, Orengo AM, Ferrini S, Accolla RS: **Tumor rejection by gene transfer of the MHC class II transactivator in murine mammary adenocarcinoma cells**. *Eur. J. Immunol.* 2003, **33**:1183–1192.
263. Fontes JD, Kanazawa S, Jean D, Matija Peterlin B: **Interactions between the Class II Transactivator and CREB Binding Protein Increase Transcription of Major Histocompatibility Complex Class II Genes**. 1999, **19**:941–947.
264. Kretsovali A, Agalioti T, Spilianakis C, Tzortzakaki E, Merika M, Papamatheakis J: **Involvement of CREB Binding Protein in Expression of Major Histocompatibility Complex Class II Genes via Interaction with the Class II Transactivator**. *Mol. Cell. Biol.* 1998, **18**:6777–6783.
265. Beresford GW, Boss JM: **CIITA coordinates multiple histone acetylation modifications at the HLA-DRA promoter**. *Nat. Immunol.* 2001, **2**:652–657.
266. Arancibia-Cárcamo C V., Osawa H, Arnett HA, Háskova Z, George AJT, Ono SJ, Ting JPY, Streilein JW: **A CIITA-independent pathway that promotes expression of endogenous rather than exogenous peptides in immune-privileged sites**. *Eur. J. Immunol.* 2004, **34**:471–480.

267. Collinge M, Pardi R, Bender JR: **Cutting Edge: Class II Transactivator-Independent Endothelial Cell MHC Class II Gene Activation Induced by Lymphocyte Adhesion**. *J. Immunol.* 1998, **161**:1589–1593.
268. Williams GS, Malin M, Vremec D, Chang C-H, Boyd R, Benoist C, Mathis D: **Mice lacking the transcription factor CIITA— a second look**. *Int. Immunol.* 1998, **10**:1957–1967.
269. Truax AD, Thakkar M, Greer SF: **Dysregulated recruitment of the histone methyltransferase EZH2 to the class II transactivator (CIITA) promoter IV in breast cancer cells**. *PLoS One* 2012, **7**:e36013.
270. Holling TM, van Eggermond MCJA, Jager MJ, van den Elsen PJ: **Epigenetic silencing of MHC2TA transcription in cancer**. *Biochem. Pharmacol.* 2006, **72**:1570–1576.
271. Schnappauf F, Hake SB, Camacho Carvajal MM, Bontron S, Lisowska-Grospierre B, Steimle V: **N-terminal destruction signals lead to rapid degradation of the major histocompatibility complex class II transactivator CIITA**. *Eur. J. Immunol.* 2003, **33**:2337–2347.
272. Beaulieu YB, Machado JAL, Ethier S, Gaudreau L, Steimle V: **Degradation, promoter recruitment and transactivation mediated by the extreme N-terminus of MHC Class II transactivator CIITA isoform III**. *PLoS One* 2016, **11**:e0148753.
273. Thibodeau J, Bourgeois-Daigneault MC, Huppé G, Tremblay J, Aumont A, Houde M, Bartee E, Brunet A, Gauvreau ME, de Gassart A, et al.: **Interleukin-10-induced MARCH1 mediates intracellular sequestration of MHC class II in monocytes**. *Eur. J. Immunol.* 2008, **38**:1225–1230.

Appendices

Appendix A - Supplementary methods

A.1 Assays and methodology applied to the PMBCL sequencing cohort

Case #	FISH CIITA ba	WGS	WTS	TSCA	Sanger exon coverage	pIII	Intron1 PCR	Intron1 sequencing	Intron1 subcloning
1	x			x	x	x	x	x	x
2	x					x	x		
3	x			x		x	x		
4	x			x		x	x		
5	x			x		x	x		
6	x			x	x	x	x		
7	x			x	x	x	x	x	x
8	x		x			x	x	x	x
9	x	x	x	x		x	x	x	x
10	x			x	x	x	x	x	x
11	x				ND	x	x	x	x
12	x		x			x	x		
13	x			x		x	x		
14	x			x	x	x	x		
15	x			x		x	x		
16	x				x	x	x	x	x
17	x			x		x	x	x	x
18	x			x	x	x	x	x	x
19	x	x	x	x	x	x	x	x	x
20	ND		x		x	x	x	x	x
21	x				x	x	x	x	
22	x			x		x	x		
23	x			x		x	x	x	
24	x		x			x	x	x	x
25	x			x		x	x	x	x
26	x				x	x	x	x	x
27	x			x		x	x	x	
28	x			x		x	x	x	
29	x				x	x	x	x	x
30	x				x	x	x	x	x
31	x			x	x	x	x	x	
32	x			x		x	x	x	x
33	x			x		x	x		
34	x				x	x	x		
35	x				x	x	x		
36	x			x		x	x	x	
37	x				x	x	x	x	
38	x			x		x	x	x	
39	x			x		x	x	x	
40	x			x	x	x	x	x	

Case #	FISH CIITA ba	WGS	WTS	TSCA	Sanger exon coverage	pIII	Intron1 PCR	Intron1 sequencing	Intron1 subcloning
41	x			x		x	x		
42	x		x			x	x	x	
43	NE			x		x	x	x	
44	NE			x		x	x	x	
45	ND			x		x	x		

Abbreviations: ba, break-apart; WGS, whole genome sequencing; WTS, whole transcriptome sequencing (RNA-Seq); TSCA, TrueSeq custom amplicon; ND, not done; NE, not evaluable

A.2 TSCA oligos for the PMBCL cohort

Amplicon name	ULSO sequence	DLSO sequence	comment
CIITA + CIITA_UserDefined (5254060)_5553285	TGGGAGTCCTGGAAGACATACTGGTC	CTCAAGAAATGTTTGTTGAATGAATGAATG	exon 4 +5
CIITA + CIITA_UserDefined (5254060)_5553286	CTTCAGTTAGACCTTGTTGATTGACTGC	CCTGCATTTCCTGCCTTGTTCCCT	exon 4 +5
CIITA + CIITA_UserDefined (5254060)_5553287	TGAAAGCCCAAGGTGAGTCTCTATTG	CAAACTTACTGAAAATGTCCTTGCTCAG	exon 4 +5
CIITA + CIITA_UserDefined (5254060)_5553288	ATTCATTGATGGGCAGTCAGACCC	GGCAGAAAAGTCAGAAAAGACGTGAG	exon 4 +5
CIITA + CIITA_UserDefined (5254060)_5553289	TTTTAAAGGGCCTCCCAACCAGACAGGAC	AACTTCTGCTGGCATCTCCATAC	exon 4 +5
CIITA + CIITA_UserDefined (5254061)_5553290	TCTCGGCTCCCACGTCGCAGATGCAG	GGGGGAAGGAATCAATATTTATTGCACAAC	exon 17 + 18
CIITA + CIITA_UserDefined (5254061)_5553291	CACAGGCCTCCAATCCCTCCCCCT	GGGCGGTGGGTGGCTCAGCCCGGGGTGGGA	exon 17 + 18
CIITA + CIITA_UserDefined (5254061)_5553292	GCAAGAGAAACTCACCTTGGGGC	ACACTCACTCCATCACCCGGAGGGAC	exon 17 + 18
CIITA + CIITA_UserDefined (5254061)_5553293	GAGCTGGGGAGTCCCAAGGGCCA	GAGAACAACTCACTCCCCAGGCGTGT	exon 17 + 18
CIITA_Exon (8620300)_5553342	AAGTGTCAAGTGAATGAGCAATGTGAA	ATGGAGAAGCAGGTGCCAGATTTAGG	exon 10
CIITA_Exon (8620301)_5553343	AAGCACACAGCCTCATCACTAGCCTC	GTTTCTGAACACCCTCTAATTTTACCAC	exon 1
CIITA_Exon (8620301)_5553344	TTTGCATGTTGGCTTAGCTTGGC	CCTGGCTGGGATTCCTACACAAT	exon 1

Amplicon name	ULSO sequence	DLSO sequence	comment
CIITA_Exon (8620301)_5553345	CAGAATGGTTTCTCTGTTTATCTGGAATGG	AGCAGCTCCCGGAGTCTGGCAGC	exon 1
CIITA_Exon (8620302)_5553346	ATGCACCATCCCCATCAGACTTGGGCC	AAAACATGTGATCAGCTGCCCCAGGG	exon 6
CIITA_Exon (8620303)_5553347	ACTTTGGGGGCCCGATTCAGCAGGAA	AATCAGATGGGGGCCATCAGCTAGCG	exon 15
CIITA_Exon (8620304)_5553348	AGCAGTCGCTCACTGGTCTCACTAG	TTGTCATCTTCTCAGCCCTGGCTGC	exon 7
CIITA_Exon (8620304)_5553349	CCCCCACTGTGGTGACTGGCAGTCT	GAGAGAGTGGGCTTTCTCCCTCTT	exon 7
CIITA_Exon (8620305)_5553350	ACTGGGAGGGGGTACTTGGCTGGCCT	GAGAGAAGAGAGTAGAACTTCCAAAGGAA	exon 9
CIITA_Exon (8620305)_5553351	TATTCTCACACCACTCTCCACCCCCAAT	CTGCCCAGCATGCCTGAACCTG	exon 9
CIITA_Exon (8620305)_5553352	ATGACCTGTTGTCCCTACAGGCAGCTTT	TCAGTGGCTGATGGAGCGAAGGGGC	exon 9
CIITA_Exon (8620318)_5553353	TGCAGGGACTCCCACAGCGCCA	AGCTCAACCTCTACCTTTCCCAGAAA	exon 12
CIITA_Exon (8620318)_5553354	CACCGTGCCTGGGTCTGAGGCCCT	CCTGAAGGATGTGGAAGACCTGGGAAA	exon 12
CIITA_Exon (8620318)_5553355	ATGGGGTGTCCCAGAGGACAGGGGGC	ACTTGGCTTTGAAAGGCTCGATG	exon 12
CIITA_Exon (8620312)_5553356	GGAGATTCAGGCAGCTCAACGAG	TCCCTAAGAGCAGTAGCTGTTTCTGT	exon 8
CIITA_Exon (8620312)_5553357	TCTCTTGCAGTGCCTTTCTCCAGTT	GGCCTGGCTCCCCGACCACCTCT	exon 8
CIITA_Exon (8620313)_5553358	TGCGGGCTCGGCACCATACGTG	ACATCTGTTCCCCACACAGTTTTT	exon 11
CIITA_Exon (8620313)_5553359	CGCCCCTGGCCTTTGCAGAGCC	TGGAGCGGGAACTGGCCACCCCGGACT	exon 11
CIITA_Exon (8620313)_5553360	ACAGCAATCACTCGTGTCTCACGCG	TGGCCTGCACCAGATCCACCTCC	exon 11
CIITA_Exon (8620313)_5553361	GCCCAAGGAGGCCTGGCTGAGGTGCTGTT	GCCGGCTTCCCCAGTACGACTTTGTCTT	exon 11
CIITA_Exon (8620313)_5553362	GCCACGAGTGGCTGTGGGCCCA	AGCCCAATAGCTCTTGCCCTGACC	exon 11

Amplicon name	ULSO sequence	DLSO sequence	comment
CIITA_Exon (8620313)_5553363	TGCTTGAACCGTCCGGGGGATGCCTAT	TGGAAGCGCAAGATGGCTTCCT	exon 11
CIITA_Exon (8620313)_5553364	GGTGCAACCTCGGAGCAGCTTC	AACGCGGTCAGGTCTCTTCAAGATGT	exon 11
CIITA_Exon (8620313)_5553365	CGGACCGGCACCGGCGGAGCCCTGCTCCCT	GAGCTGTCCGGCTTCTCCATGGA	exon 11
CIITA_Exon (8620313)_5553366	TGAGAAGAAGTGGCCGGTCCCGGAGG	TCAGGCTCTGGACCAGGCGGCCCCGG	exon 11
CIITA_Exon (8620313)_5553367	GCGCTACTTTGAGAGCTCAGGGATGA	TGGGGAGGACGCCAAGCTGCCCTCCA	exon 11
CIITA_Exon (8620313)_5553368	TACTTTGATGTCTGCGGCCCAGCTCC	CAGGGCCTCTGAGAGCTGGCAC	exon 11
CIITA_Exon (8620313)_5553369	GGCAGAGCTGGCCAAGCTGGCCT	GCCTTCCCCAGCTTCCTCCTGCAAT	exon 11
CIITA_Exon (8620313)_5553370	TGGCACGCCCTCCAGCCAGTTGT	AGCTCGGACTCTGCGGCCCGCGGTGGG	exon 11
CIITA_Exon (8620313)_5553371	TTGACCCCAAGGAAGAAGAGGCCCTAT	TGCTTGCGAGGTACCTGAAGCG	exon 11
CIITA_Exon (8620313)_5553372	AGAAAAGAGAGGCGGCCGGGGAGCT	TTCTGCTTCCTGTCCACCGAGGCA	exon 11
CIITA_Exon (8620313)_5553373	TGGAATTTGGCAGCACGTGGTACA	CCCCTCTGGATTGGGGAGCCTC	exon 11
CIITA_Exon (8620313)_5553374	CGGCTGCTCTGCATACTAAAAGAGA	AAATGCCAGTGCTGCGGAGGTCCAG	exon 11
CIITA_Exon (8620314)_5553375	ATGTGGGTTCCCTGCGCTCTGCAGCCCC	ACACAGTGAGGGGGAGGGCTCAGGAC	exon 16
CIITA_Exon (8620315)_5553376	CAAGATGTGGTTCATTCCGCAGC	GGAGGAGCCAGGAGAGGGGGTT	exon 20
CIITA_Exon (8620316)_5553377	TACAAGCCCAGCTAATGCTGCAGGGGA	AATGTTAGGGGGAGCAGGCACTGCTGT	exon 13
CIITA_Exon (8620306)_5553378	TGGAGGTCTTACCCTTGCTCTTT	TCTGGGTGCAGTGCTGTGATCATA	exon 14
CIITA_Exon (8620308)_5553379	TGTTAAGAAGCTCCAGGTAGCCAC	CTGGGGATGAGAGGAGTGAATAAAAGC	exon 2
CIITA_Exon (8620308)_5553380	CACCATGGAGTTGGGGCCCCTAGAA	ATCCCCCACCCCTCAGCTTGCTGTAGA	exon 2

Amplicon name	ULSO sequence	DLSO sequence	comment
CIITA_Exon (8620309)_5553381	CAGCAAAGAACTCTTGCCCTTGATTGT	AGTCCCTTGGATGAAGAAGGAAATTTC	exon 3
CIITA_Exon (8620310)_5553382	ACAAGGACACTGCCCCCAACCCACTG	GGGGATGGGACTCAGAGCCAGG	exon 19
CIITA pIV exon1_UserDefined (8658714)_5553515	TTGAGCAAGTAGCTGACAGTCTCGGA	AGCTCGTCCGCTGGTCATCCTAC	pIV exon 1

A.3 Primer sets

Primer name	Primer sequence (5'-3')	comment
-21M13F	TGTAAAACGACGGCCAGT	sequencing primer and tag
-27M13R	CAGGAAACAGCTATGAC	sequencing primer and tag
CIITA pIII F1	CAAGGGTACCATATTTGGGTTA	promoter 3 region + exon 1
CIITA pIII R1	GACTCCTGTTCCCATCCTCA	promoter 3 region + exon 1
CIITA F5	GTCCTACCTGTCAGAGCCCCAAG	Intron 1
CIITA I1R1	TGGAGTCACCTGGGATAGATGGT	Intron 1
CIITA ISEQ F1	AAGGCAGCATGGCAGCTA	sequencing primer Intron 1
CIITA ISEQ F2	CCTAGGGCCAGCATCAGA	sequencing primer Intron 1
CIITA ISEQ F2.5	GCTCCCTGCAACTCAGGA	sequencing primer Intron 1
CIITA ISEQ F3	GGCTCCGAGACTGTCAGC	sequencing primer Intron 1
CIITA ISEQ R1	CTTTCCCACAGGTCCATTG	sequencing primer Intron 1
CIITA ISEQ R2	CCGTGAAAGTGGCAAACC	sequencing primer Intron 1
CIITA ISEQ R3	GAGTCGTTGCGGGGATG	sequencing primer Intron 1
CIITA ISEQ R4	ATTCCGGCTTTCCTGGAC	sequencing primer Intron 1
CIITAe1F1	CCTGGCTGGGATTCCTACAC	CDS exon 1
CIITA e1R1	CTCTGACAGGTAGGACCCAGCA	CDS exon 1
CIITA_e2e3F1	TCAATTTTCTGCCTCTTTCCA	CDS exon 2+3
CIITA_e2e3R1	GACGTGGCTCATGATGAATG	CDS exon 2+3

Primer name	Primer sequence (5'-3')	comment
CIITA_e4e5F1	CACACAGTGGGCCTTCAGTT	CDS exon 4+5
CIITA_e4e5R1	CAGGCTTTGGAGTCAAGGAA	CDS exon 4+5
CIITA_e6e7F1	GGCCGTATAGCCTGCTAGAGT	CDS exon 6+7
CIITA_e6e7R1	CCTGACAGTCCCTGCCTTAG	CDS exon 6+7
CIITA_e8F1	AATCGCAAACACAGGTGCTA	CDS exon 8
CIITA_e8R1	GCAGTCAGGTAGGAGGGAGA	CDS exon 8
CIITA_e9F1	GGTGACCCAAGTGCATTTCT	CDS exon 9
CIITA_e9R1	GCCTCGATCCTGCTTCTAGT	CDS exon 9
CIITA_e10F1	GGGGTAACCCTCACCCTAAA	CDS exon 10
CIITA_e10R1	GATATGGGCTTCCATCTCCA	CDS exon 10
CIITA_e11F1	GTAAATGATGGTGGCAGTGCT	CDS exon 11
CIITA e11F2	GCTCTGAGTGGCGAAATCA	CDS exon 11
CIITA_e11R1	CTTCCTTGGGGTCAATGCTA	CDS exon 11
CIITA_e11R2	GCACCAAGACCGCAGTTAAT	CDS exon 11
CIITA_e13F1	GGTAAGGGCTCAGTGACAGC	CDS exon 13
CIITA_e13R1	GGTGGTACAAGCCCAGCTAA	CDS exon 13
CIITA_e14F1	GAATGAGGGGCTGTGACTGT	CDS exon 14
CIITA_e14R1	GAGGCTGCAGTGAGCTATGA	CDS exon 14
CIITA_e15F1	GTCAGATGGCCCCAGGAC	CDS exon 15
CIITA_e15R1	ACAGCGAGCCCTTGTCTAAA	CDS exon 15
CIITA_e16F1	CATGCAAGTTTGGTCCTGAG	CDS exon 16
CIITA_e16R1	GCTGTGGTGATTGCTTCTAGG	CDS exon 16
CIITA_e17e18F1	AGTGGGACCAGGCTTTTTCT	CDS exon 17+18
CIITA_e17e18R1	TGGGGAGTGAGTTGTTCTCC	CDS exon 17+18
CIITA_e19F1	CGCCCTCTCTCCTCTAACC	CDS exon 19
CIITA_e19R1	AAGGACACTGCCCCCAAC	CDS exon 19
CIITA DEV INV F	AAGTCTCATGCCTTGGAGGA	DEV inversion CIITA, SOCS1
CIITA DEV INV R	TGGGTGGGTCTCTGTTTCTC	DEV inversion CIITA, SOCS1
SOCS1 DEV INV F	AGGTAGGAGGTGCGAGTTCA	DEV inversion CIITA, SOCS1
SOCS1 DEV INV R	GTCCTCCGCGACTACCTGAG	DEV inversion CIITA, SOCS1

Primer name	Primer sequence (5'-3')	comment
CLEC16A R1	GACAAGGGTCAGCGATTCGAC	DEV deletion CIITA, CLEC16A
CLEC16A R2	AGAGGAATCTAGCGCGTCTG	DEV deletion CIITA, CLEC16A
EMP2 R1	TGATGACTGTGCAATTCGTG	Karpas1106P TXNDC11-EMP2 fusion
EMP2 I1 R4na	CTGGGCGTGGTCGTCATAC	Karpas1106P TXNDC11-EMP2 fusion
EMP2 I1 R1	AGCAGGAATGGCATATGAGC	Karpas1106P TXNDC11-EMP2 fusion
EMP2 I1 R2	CAAGTCTAAGCCGCATCCTC	Karpas1106P TXNDC11-EMP2 fusion
EMP2 I1 R3	TCTGTTGCAGAACCCCAAAC	Karpas1106P TXNDC11-EMP2 fusion
EMP2 I1R5	CTGCAGAGGTGATGTGGAAG	Karpas1106P TXNDC11-EMP2 fusion
TXNDC11 E3 F1	GTGCTCTGTAACGGGATGGT	Karpas1106P TXNDC11-EMP2 fusion
TXNDC11 I3F2	CACATTCCATCCCATTTCC	Karpas1106P TXNDC11-EMP2 fusion
TXNDC11 I3F1	ACAAAGAGGGCTGAAAAACG	Karpas1106P TXNDC11-EMP2 fusion
TXNDC11 F2	GGGGAAATGCAGAAAACAGA	Karpas1106P TXNDC11-EMP2 fusion
TXNDC11 F1	AATACCAGCAAAGCCACCTG	Karpas1106P TXNDC11-EMP2 fusion
TXNDC11 I3F5	TTGAAGGGACCTCAAAGGAA	Karpas1106P TXNDC11-EMP2 fusion
FAM18A I2 F1	GGTCCCAGGCTTACTTGAGGAGA	TVP23A
FAM18A I2 F2	TCCAATTCAACGTATCTTGCTGCT	TVP23A
FAM18A I2 F2.2	AAAAACATTAGCCGGTGTGG	TVP23A
FAM18A I2 F3	GGCACACAAAGGTCCTTGAAAAT	TVP23A
NUBP1_5UTR_F1	GTTCCGGTGACCACGAAG	U2940 NUBP1-CIITA fusion
NUBP1_e9_F1	ACTTGGAGGTCCCTCTCCTC	U2940 NUBP1-CIITA fusion
NUBP1_I1F1	TGGAGCATCTCAGTCTGCTG	U2940 NUBP1-CIITA fusion
NUBP1_I9_F2	GTGCCATGCTCGTCTCAGT	U2940 NUBP1-CIITA fusion
CIITA xpr F2	CACCATGCGTTGCCTGGCTCC	cDNA cloning
XhoI-CIITA xprR1*	CCTAAGATCTCGAGTCATCTCAGGCTGATCCGTGAATC	cDNA cloning
XhoI-U2940 CIITA R3*	CCTAAGATCTCGAGTCATGGGATACAGCCTGGAGAAGAGC	cDNA cloning
pCMV F	CGCAAATGGGCGGTAGGCGTG	Expression clone sequencing
PGK-R	CAGAGGCCACTTGTGTAG	Expression clone sequencing
CIITA cD-F2 na	GCTCTGGCAAATCTCTGAGG	Expression clone sequencing
CIITA cD-F2.5na	GCTTCCCCAGTACGACTTTG	Expression clone sequencing
CIITA cD-F3na	AGAAGAAGCTGCTCCGAGGT	Expression clone sequencing

Primer name	Primer sequence (5'-3')	comment
CIITA cD-F4 na	GCAGCACGTGGTACAGGAG	Expression clone sequencing
CIITA cD-R1na	CACTGGGAGGGGGTACTTG	Expression clone sequencing
CIITA cD-R2na	AAGCCGGACAGCTCAAATAG	Expression clone sequencing
CIITA cD-R3na	GCCCAGTACATGTGCATCAG	Expression clone sequencing
CIITA F5	GTCCTACCTGTCAGAGCCCCAAG	qRT-PCR
CIITA E2R1	AGCCAGGTCCATCTGGTCATAGA	qRT-PCR
Retro_NUBP1	CCGGACTCTAGCGTTTGGAATGGAGGAGGTGCC	U2940 NUBP1-CIITA fusion Gibson Assembly
Retro_CIITA_end	GCTCGGTACCAAGCTTAAGTTTTCATCTCAGGCTGATCC	U2940 NUBP1-CIITA fusion Gibson Assembly
GAPDH_fwd	CATGAGAAGTATGACAACAGCCT	qRT-PCR
GAPDH R	AGTCCTTCCACGATACCAAAGT	qRT-PCR
A43030_del_CIITA_F	CAGGAGGATTGGAGGATCAC	oligocapture validation
A43030_del_CIITA_R	GGACAGGAGAACATGGCTTC	oligocapture validation
A43031_inv_CIITA_F	CAGATGTCGTTCTTGCTTTTG	oligocapture validation
A43031_inv_CIITA_R	TCCAAGCCTCTGTCACCTCT	oligocapture validation
A43036_inv_F	GGGGTGGTATCCCTTTTCTC	oligocapture validation
A43036_inv_R	CTAAGGGCGAAAAAGCAGTT	oligocapture validation
A43036_tX;16_F	CAGGAGCTAGGGAGCCACTT	oligocapture validation
A43036_tX;16_R	TTTTGCAGATCTACTTGCATGA	oligocapture validation
A43049_inv_CIITA_F	AGGCCCTTTTATCAAGTGAGG	oligocapture validation
A43049_inv_CIITA_R	CATCAGCAGGTCCAGGTTCT	oligocapture validation
A43051_del_CIITA_F	CTGGGTTCAAAGCAAACCAT	oligocapture validation
A43051_del_CIITA_R	AGGAGCTCAAGAGCAACCTG	oligocapture validation
A43052_CIITA-LINC01185_F	AGCCCGGGAACCTACATC	oligocapture validation
A43052_CIITA-LINC01185_R	CCGCTCCTGAACTTTAAACC	oligocapture validation
A43052_LINC01185-CIITA_F	GACGCAGCAACCCTCACC	oligocapture validation
A43052_LINC01185-CIITA_R	CGCAGCCTGGAGTGTCTAAC	oligocapture validation
A43052_IgK-CIITA_F	GCATGATACAGAAAAGTGGAAAA	oligocapture validation

Primer name	Primer sequence (5'-3')	comment
A43052_IgK-CIITA_R	CTCCCTAGCTCCTGGCTCCT	oligocapture validation
A43052_del_CIITA_F	TGGTTAAAGGTTTGGCTCCC	oligocapture validation
A43052_del_CIITA_R	AGTTGGGATGCCACTTCTGA	oligocapture validation
A43067_inv_IL4R_F	CATCTTGGCGAAGGTGTGTG	oligocapture validation
A43067_inv_IL4R_R	TCAGGTGTAGGTTTGGCCAG	oligocapture validation
A43068_del_CIITA_F	GGGGCCTAATGTTGTCCTCT	oligocapture validation
A43068_del_CIITA_R	AACTCTGTTTCCTCCCTCGG	oligocapture validation
A43069_MGAT3_F	CCTCTGCCCCGTTTTCATCT	oligocapture validation
A43069_MGAT3_R	GGTCCAAAGTCTGGCGTACA	oligocapture validation
A43070_del_TXNDC11_F	AGAGGCAGCAGCCTTTTTCT	oligocapture validation
A43070_del_TXNDC11_R	TGGGGTAATGCCTACGAGTC	oligocapture validation
A43070_CIITA-PRDM16_F	GATGCTGGAAACGAGGTGTT	oligocapture validation
A43070_CIITA-PRDM16_R	GGAAGAGGAAGCGTCTGGTC	oligocapture validation
A43072_del_NUBP1_F	GGAGATTCAGGCAGCTCAAC	oligocapture validation
A43072_del_NUBP1_R	CCAGCCACGTTAGCCTACAG	oligocapture validation
A43075_AID-CIITA_F	TCCATTTGTTCAGACGTAGCTT	oligocapture validation
A43075_AID-CIITA_R	AACACCTCGTTTCCAGCATC	oligocapture validation
A43075_CIITA-AID_F	CGCCAAGATGTTCAACGAG	oligocapture validation
A43075_CIITA-AID_R	TTGAACTCCAGGGCTCAAG	oligocapture validation
A43076_CIITA_SNX29_F	TGCAGTAAATGCCGTTTGAG	oligocapture validation
A43076_CIITA_SNX29_R	CACCATCTGAGCTCCATTCA	oligocapture validation
A43077_delCIITA_F	TTGCAACAACCTCAGTGGAG	oligocapture validation
A43077_delCIITA_R	GTTGGGATGCCACTTCTGAT	oligocapture validation
A43078_MYC/PVT1-CIITA_F	ACAAAGCTGCCAGAGAAACG	oligocapture validation
A43078_MYC/PVT1-CIITA_R	CCACCACGTGCTTTATCAGA	oligocapture validation
A43080_delCIITA_F	AGAGCTTCGACTGCCTCTTC	oligocapture validation
A43080_delCIITA_R	AACACCTCGTTTCCAGCATC	oligocapture validation
A43080_del_F	TGGGTGGGTCTCTGTTTCTC	oligocapture validation

Primer name	Primer sequence (5'-3')	comment
A43080_del_R	GCCAGACTCCACTCCATACC	oligocapture validation
A43095_CIITAinv_F	CAGGGAAAGTGAAGCTCAGG	oligocapture validation
A43095_CIITAinv_R	AAGGTTTGGCTCCCTACTGC	oligocapture validation
A43110_del_CIITA_F	GCCATCACCTCACTGAACCT	oligocapture validation
A43110_del_CIITA_R	GTACACCCTCTGCGGTATGG	oligocapture validation
A43115_CIITAinv_F	CTGTCAGAGCCCCAAGGTAA	oligocapture validation
A43115_CIITAinv_R	TACAGGCTTTCCCACAGGTC	oligocapture validation
A43115_tra7_16_F	AGCGAGACTCCGTCTCAAAA	oligocapture validation
A43115_tra7_16_R	CTCCTGACCTCAGGTGATCC	oligocapture validation

A.4 TSCA design (*CIITA*) for the DLBCL cohort

Amplicon name	ULSO Sequence	DLSO Sequence
CIITA + CIITA_UserDefined (15231807)	AAGACCTCCCCACCCACCACAAACTTAC	CTCAAGAAATGTTTGTTGAATGAATGAAT
CIITA + CIITA_UserDefined (15231807)	AGCTGGAGGGCCTGAGCAAGGACATTTT	CATAGGACCAGATGAAGTGATCGGTGA
CIITA + CIITA_UserDefined (15231807)	TGTTGGCTTTTAAAGGGCCTCCC	GCTCTACTTTGAGAAAAACCAGAGACC
CIITA + CIITA_UserDefined (15969338)	TGGTCATAGAAGTGGTAGAGGCACA	GAATAAAAGCGCTCATTCAGCACCTCT
CIITA + CIITA_UserDefined (15969338)	GAGCTTCTTAACAGCGATGCTGA	GGACGCTCTCTGCAGATGGGGATGAT
CIITA + CIITA_UserDefined (15969338)	CTTGATTGTCCTTTTCTGGGCTCAG	ACAGACGTGGGAGCTGTCCGTGGTG
CIITA_Exon (15192912)	GGAGGTCTTACCCTTGCTCTTTGC	ACAGGTGCATGCTACAGTGCCCAGCAA
CIITA_Exon (16954886)	GCTGAGACTGCACGCTAAATTAAGATG	AGTTGGGAGCCCGCCAAGCTAAGCCA
CIITA_Exon (16756956)	TTTGTTCAGGGCTGTGGTGATTGCTTC	ACACAGTGAGGGGGAGGGCTCAGGACC
CIITA_Exon (15728519)	TACATGTCCTCAACCTGCATGGCGTGA	TCAGACCCAGGCACGGTGACCAG
CIITA_Exon (17651572)	ATGAGGGGATGTCTCTGATGACGCAT	GGGGGATGGGACTCAGAGCCAG
CIITA_Exon (17652495)	AAGTGCATGGAGTATGGGGAGGATGA	AAAGGCAGTGAGGTGGGATCTTGCAT
CIITA_Exon (14742566)	TGATGGAGCGAAGGGGCTGGTGGAGC	GGAGAGAAGAGAGTAGAACTTCCAAAG

Amplicon name	ULSO Sequence	DLSO Sequence
CIITA_Exon (14742566)	TCCCAACATCTCCAGACCGGCCA	AGTCGGGACAGGGAGGGTCTCCA
CIITA_Exon (16530899)	TAAATAAATGAGTGTGTGAGCCAACAA	AGAAGCAGGTGCCAGATTTAGGGTGAG
CIITA_Exon (15777327)	CTGCTGGAAGCTATTTCCAAAGTGGT	AATGTTAGGGGGAGCAGGCACTGCTG
CIITA_Exon (15009032)	TTAGAGGACTCTAAGGGACCCCAAGCT	AAAACATGTGATCAGCTGCCCCAGGGA
CIITA_Exon_1928651_UserDefined (16481070)	TGGTTGAACAGCGCAGGCAGTGG	TTGATCAGCAACTGCTCTGTGCCAG
CIITA_Exon_1928651_UserDefined (16481070)	GACTGCTCCACCCTGCCCTGCCT	AGCCAGTTTTATCCTTGGGGCC
CIITA_Exon_1929553_UserDefined (15485636)	AAAACTGTGTGGGGAACAGATGTAAATG	GCCAGGCTGGAGAGGAGCAGCAGCAA
CIITA_Exon_1929553_UserDefined (15485636)	CCGCGTGAGACACGAGTGATTG	ACTCGTGGCGGCCGATGAGGTTTTCA
CIITA_Exon_1929553_UserDefined (15485636)	GGCTGTGAGGAGGAGGGTGCAA	CTGTGGGCCCAGGGAGAAGAGC
CIITA_Exon_1929553_UserDefined (15485636)	CTTTTCCAGAAGAAGCTGCTCCGA	GGCCACTTCTTCTCAGTCACAGC
CIITA_Exon_1929553_UserDefined (15485636)	TTTGATGTCTGCGGCCCAGCTCCCAGGCCA	TCCCGGAGGAGCGTCAGGGCTC
CIITA_Exon_1929553_UserDefined (15485636)	GGGCCCTGGCAGAGCTGGCCAA	CTCTGAGTGGCGAAATCAAGGACAA
CIITA_Exon_1929553_UserDefined (15485636)	TTCAGGTACCTCGCAAGCACCTTCTG	AGCCACAGGGCCCCCAGGAAGC
CIITA_Exon_1929553_UserDefined (15485636)	ATCGGCGGCTGCCTCGGTGGACAGGAA	CCTTGGAGGCGGCGGGCCAAGACTTCT
CIITA_Exon_1929553_UserDefined (15485636)	AATCCATTCTGCCCCACCCGGCTGCTCT	CTTGCCCAGTACATGTGCATCAG
CIITA_Exon_1928561_UserDefined (17619073)	CAGATGGCCCCAGGACGCTAGCTGAT	TTTTTAGACAAGGGCTCGCTGTGTCA

A.5 Cases selected for oligocapture sequencing

BCCRC_ID	Library ID	Lymphoma subtype	Localization	CIITA ba	PDL1-2 ba	TP63 ba	TBL1XR1 ba
BC_001 B		PMBCL	mediastinum	1	0	0	0
BC_002	A43049	PMBCL	mediastinum	1	0	0	0
BC_003	A43050	PMBCL	mediastinum	1	0	0	0
BC_004	A43051	DLBCL	GI	1	0	0	0
BC_005	A43052	DLBCL	tonsil	1	0	0	0
BC_006	A43053	DLBCL	testis	1	0	0	0
BC_007		FL	LN	1	0	0	0
BC_008		FL	LN	1	0	0	0
BC_009		FL	LN	1	0	0	0
BC_010		FL	LN	1	0	0	0
BC_011		FL	LN	1	0	0	0
BC_012		FL	LN	1	0	0	0
BC_013		FL	LN	1	0	0	0
BC_014		FL	LN	1	0	0	1
BC_015		FL	BM	1	0	0	0
BC_016		FL	LN	1	0	0	0
BC_017		PMBCL	LN	1	0	NA	NA
BC_018 B		PMBCL	LN	1	0	NA	NA
BC_019		PMBCL	LN	1	0	NA	NA
BC_020	A43067	PMBCL	LN	1	0	NA	NA
BC_021	A43068	PMBCL	LN	1	0	NA	NA
BC_022	A43069	PMBCL	mediastinum	1	0	NA	NA
BC_023	A43070	PMBCL	LN	1	0	NA	NA
BC_024	A43071	PMBCL	mediastinum	1	1	NA	NA
BC_025	A43072	PMBCL	LN	1	1	NA	NA
BC_026	A43073	PMBCL	mediastinum	1	0	NA	NA
BC_027 B	A43074	PMBCL	mediastinum	1	1	NA	NA
BC_028	A43075	PMBCL	mediastinum	1	0	NA	NA
BC_029	A43076	PMBCL	mediastinum	1	0	NA	NA
BC_030	A43077	PMBCL	mediastinum	1	0	NA	NA
BC_031 B	A43078	PMBCL	mediastinum	1	0	NA	NA
BC_032 B	A43079	PMBCL	LN	1	1	NA	NA
BC_033	A43080	PMBCL	LN	1	0	NA	NA
BC_034 *	A43081	PMBCL	mediastinum	1	0	NA	NA

BCCRC_ID	Library ID	Lymphoma subtype	Localization	CIITA ba	PDL1-2 ba	TP63 ba	TBL1XR1 ba
BC_035	A43082	PMBCL	LN	0	1	0	0
BC_036	A43083	DLBCL	GI	0	1	0	0
BC_037	A43084	PMBCL	mediastinum	0	1	0	0
BC_038	A43085	DLBCL	LN	0	1	0	0
BC_039	A43086	DLBCL	LN	0	1	0	0
BC_040	A43087	DLBCL	LN	0	1	0	0
BC_041	A43088	DLBCL	GI	0	1	0	0
BC_042	A43089	DLBCL	testis	0	1	0	0
BC_043	A43090	PMBCL	LN	0	1	NA	NA
BC_044 *	A43091	PMBCL	LN	0	1	NA	NA
BC_045	A43092	PMBCL	mediastinum	0	1	NA	NA
BC_046	A43093	PMBCL	mediastinum	0	1	NA	NA
BC_047	A43094	PMBCL	mediastinum	0	1	NA	NA
BC_048	A43095	PMBCL	soft tissue	0	1	NA	NA
BC_049 I	A43096	PMBCL	LN	0	1	NA	NA
BC_050	A43097	PMBCL	mediastinum	0	1	NA	NA
BC_051 B		PMBCL	LN	0	1	NA	NA
BC_052	A43099	DLBCL	LN	0	1	NA	NA
BC_053	A43100	PMBCL	mediastinum	0	up	NA	NA
BC_054	A43101	PMBCL	LN	0	up	NA	NA
BC_055B	A43102	DLBCL	CNS	0	0	1	0
BC_056	A43103	DLBCL	LN	0	up	NA	NA
BC_057	A43104	DLBCL	LN	0	0	0	1
BC_058	A43105	DLBCL	LN	0	0	0	1
BC_060		DLBCL	soft tissue	0	0	1	1
BC_061	A43106	DLBCL	ENT	0	0	1	1
BC_062	A43107	DLBCL	ENT	0	0	1	1
BC_063	A43108	DLBCL	ENT	0	0	0	1
BC_064	A43109	DLBCL	LN	0	0	1	0
BC_065	A43110	DLBCL	sinus	0	1	1	1
BC_066	A43111	DLBCL	testis	0	0	1	0
BC_069 B	A43030	PMBCL	mediastinum	1	1	0	0
BC_071	A43117	HL (L1236)	cell line	0	1	NA	NA
BC_072		PMBCL	mediastinum	1	0	NA	NA
BC_073		DLBCL	testis	0	1	0	0
BC_074		DLBCL	testis	1	0	0	0

BCCRC_ID	Library ID	Lymphoma subtype	Localization	CIITA ba	PDL1-2 ba	TP63 ba	TBL1XR1 ba
BC_075		DLBCL	testis	1	0	0	0
BC_076		DLBCL	LN	0	0	1	1
BC_077		DLBCL	testis	1	0	0	0
BC_078	A43029	DLBCL	GI	1	0	0	0
BC_079	A43031	DLBCL	testis	1	0	0	0
BC_080	A43032	DLBCL	testis	1	0	0	0
BC_081	A43033	DLBCL	testis	1	0	0	0
BC_082	A43034	DLBCL	CNS	1	0	0	0
BC_083	A43038	DLBCL	testis	0	1	0	0
BC_084	A43042	DLBCL	CNS	0	1	0	0
BC_085		DLBCL	ENT	0	up	NA	1
BC_086	A43045	DLBCL	LN	0	0	0	1
BC_087	A43046	DLBCL	LN	0	0	0	1
BC_088	A43047	DLBCL	bladder	0	0	1	1
BC_089	A43119	FL	LN	1	NA	NA	NA
BC_091		DLBCL	testis	0	1	0	0
BC_092 B	A43041	DLBCL	testis	0	1	0	0
BC_093 I	A43036	PMBCL	LN	1	0	NA	NA
BC_094 I	A43037	PMBCL	mediastinum	1	0	NA	NA
BC_095 I	A43043	PMBCL	mediastinum	0	1	NA	NA
BC_097 I	A43115	PMBCL	LN	1	1	NA	NA
BC_099 I	A43118	PMBCL	mediastinum	0	up	NA	NA

Abbreviations: BM, bone marrow; CNS, central nervous system; DLBCL, diffuse large B cell lymphoma; ENT, enteric: FL, follicular lymphoma; GI, gastro-intestnal; LN, lymph node; NA, not available; PMBCL, primary mediastinal large B cell lymphoma

Appendix B - Supplementary results

B.1 CDS mutations and promoter III alterations in primary PMBCL specimens

Case #	Genomic alteration	Putative translational impact	Allelic frequency	Comments
1	G5136A	p.MET1?	0.34	SNV is in *cis* with *CIITA-PDL2* translocation; infer "intron 1 deletion" from 2nd allele because translocation breakpoint renders that allele incapable of amplification
2	T4582G	pIII		two alleles are affected by structural variations, all SNVs are linked to pIII/exon 1 deletion
	T4649C	pIII		
	G4970T	pIII		
	4582_4583insT	pIII		
	4986-5217del	deletes pIII and exon 1		
3	G5186A	p.Gly18Asp*9	0.28	
4	A4605T	pIII		
5	G44204T	p.Glu1002*	0.23	
6	G35439A ‡	p.Cys715Tyr	0.65	
16	T5164C	p.Ser11Pro		T5164C, A5168C and 5105_5112del are linked but in trans to p.Pro16Aspfs*65
	A5168C	p.Tyr12Ser		
	5105_5112del	5' UTR		
	5171_5323del	p.Pro16Aspfs*65		
17	35796delTT	p.Phe835Serfs*9	0.31	biallelic intron 1 mutations; linkage of del to which intron1 allele is unknown
18	35897delA ‡	p.Ile868Phefs*22	0.01	
19	T35376C ‡	p.Leu694Ser	0.35	
20	C26496T	p.Glu107*		
21	C23229G	p.Ser66*		
22	A5183T	p.Gln17Leu		
23	G5185A	p.Gly18Thrfs*9	0.24	
	G5186C	p.Gly18Thrfs*9		
	35250_35251insC	p.Gly655Argfs*60	0.26	
25	30588_30591dup	p.Phe254Tyrfs*7	0.14	

‡ somatic origin; genomic coordinates according to LRG49

B.2 Intron 1 alterations in primary PMBCL cases

Case #	SNV or indel Position Start (hg19)	SNV or indel Position End (hg19)	ref	alt	type of genetic alteration
case 1	10971190	10971190	G	A	transition
case 1	10971878	10971878	C	G	transversion
case 1	10972410	10972425		del 16	deletion

Case #	SNV or indel Position Start (hg19)	SNV or indel Position End (hg19)	ref	alt	type of genetic alteration
case 1	10972570	10972570	G	A	transition
case 1	10972687	10972687	G	A	transition
case 1	10972750	10972750	C	T	transition
case 1	10972753	10972753	C	T	transition
case 1	10972847	10972847	C	T	transition
case 1	10972871	10972871	C	T	transition
case 1	10972975	10972975	C	G	transversion
case 1	10972990	10972990	C	T	transition
case 1	10973111	10973111	G	C	transversion
case 1	10973471	10973715		del 245	deletion
case 1	10974017	10974017	A	C	transversion
case 1	10974154	10974154	C	T	transition
case 7	10971467	10971467	G	T	transversion
case 7	10971672	10971672	G	A	transition
case 7	10971706	10971706	G	T	transversion
case 7	10971730	10971730	C	G	transversion
case 7	10971744	10971744	T	A	transversion
case 7	10971751	10971751	G	A	transition
case 7	10971777	10971778		ins 11	insertion
case 7	10971779	10971779	G	T	transversion
case 7	10971963	10971963	C	T	transition
case 7	10971990	10971990	C	T	transition
case 7	10972096	10972096	G	C	transversion
case 7	10972212	10972212	A	C	transversion
case 7	10972235	10972280		dup	duplication
case 7	10972269	10972269	G	C	transversion
case 7	10972361	10972361	G	A	transition
case 7	10972377	10972377	C	A	transversion
case 7	10972379	10972379	C	T	transition
case 7	10972382	10972382	C	T	transition
case 7	10972383	10972383	C	T	transition
case 7	10972385	10972385	C	T	transition
case 7	10972397	10972397	C	T	transition
case 7	10972419	10972419	G	C	transversion
case 7	10972571	10972571	C	G	transversion
case 7	10972629	10972640		del 12	deletion
case 7	10972678	10972689		del 12	deletion
case 7	10972722	10972722	T	C	transition
case 7	10972747	10972747	G	C	transversion
case 7	10972797	10972797	G	A	transition
case 7	10972811	10972811	C	A	transversion
case 7	10972847	10972847	C	A	transversion
case 7	10972900	10972900	G	A	transition
case 7	10972913	10972913	G	A	transition
case 7	10972933	10972933	G	A	transition

Case #	SNV or indel Position Start (hg19)	SNV or indel Position End (hg19)	ref	alt	type of genetic alteration
case 7	10972975	10972975	C	T	transition
case 7	10973078	10973078	G	A	transition
case 7	10973102	10973102	G	A	transition
case 7	10973112	10973112	C	T	transition
case 7	10973186	10973186	G	A	transition
case 7	10973217	10973217	C	T	transition
case 7	10973218	10973218	C	T	transition
case 7	10973280	10973280	C	T	transition
case 7	10973314	10973336		del 23	deletion
case 7	10973347	10973347	C	T	transition
case 7	10973351	10973351	C	T	transition
case 7	10973402	10973402	G	A	transition
case 7	10973403	10973403	G	T	transversion
case 7	10973588	10973588	G	A	transition
case 7	10973610	10973610	G	C	transversion
case 7	10973686	10973686	C	T	transition
case 7	10973690	10973690	G	C	transversion
case 7	10974111	10974111	G	A	transition
case 8	10971559	10971559	C	T	transition
case 8	10971672	10971672	G	A	transition
case 8	10971784	10971784	G	T	transversion
case 8	10971785	10971785	C	G	transversion
case 8	10971840	10971840	G	A	transition
case 8	10971845	10971845	G	A	transition
case 8	10971911	10972044		del	deletion
case 8	10972125	10972125	G	A	transition
case 8	10972147	10972152		del	deletion
case 8	10972180	10972180	C	T	transition
case 8	10972300	10972333		del	deletion
case 8	10972415	10972415	C	G	transversion
case 8	10972421	10972421	T	G	transversion
case 8	10972486	10972486	G	T	transversion
case 8	10972522	10972522	C	G	transversion
case 8	10972532	10972532	C	T	transition
case 8	10972534	10972534	C	T	transition
case 8	10972554	10972554	C	A	transversion
case 8	10972570	10972570	G	A	transition
case 8	10972619	10972619	G	A	transition
case 8	10972651	10972651	C	G	transversion
case 8	10972661	10972891		del	deletion
case 8	10972897	10972897	G	T	transversion
case 8	10972898	10972898	C	A	transversion
case 8	10972899	10972900		del	deletion
case 8	10972926	10972926	C	T	transition
case 8	10972933	10972933	G	C	transversion

Case #	SNV or indel Position Start (hg19)	SNV or indel Position End (hg19)	ref	alt	type of genetic alteration
case 8	10972965	10973055		del	deletion
case 8	10973079	10973079	G	A	transition
case 8	10973103	10973103	C	G	transversion
case 8	10973110	10973113		del	deletion
case 8	10973123	10973123	G	A	transition
case 8	10973205	10973205	C	T	transition
case 8	10973253	10973253	C	G	transversion
case 8	10973265	10973363		del	deletion
case 8	10973436	10973436	G	T	transversion
case 8	10973445	10973445	C	T	transition
case 8	10973478	10973487		dup	duplication
case 8	10973495	10973495	G	A	transition
case 8	10973544	10973544	C	T	transition
case 8	10973551	10973551	G	C	transversion
case 8	10973602	10973602	G	A	transition
case 8	10973713	10973713	C	T	transition
case 8	10973756	10973756	G	A	transition
case 8	10973852	10973852	C	T	transition
case 8	10973896	10973896	G	C	transversion
case 8	10973977	10973977	C	T	transition
case 9	10971419	10971419	A	G	transition
case 9	10971979	10971979	C	T	transition
case 9	10972019	10972019	C	T	transition
case 9	10972115	10972141		del 27	deletion
case 9	10972419	10972419	G	C	transversion
case 9	10972508	10972514		del 7	deletion
case 9	10972570	10972570	G	A	transition
case 9	10972668	10972668	C	A	transversion
case 9	10972674	10972684		del 11	deletion
case 9	10972743	10972913		del 171	deletion
case 9	10973090	10973090		ins A	insertion
case 9	10973104	10973304		inv 201	inversion
case 9	10973208	10973213		del 6	deletion
case 9	10973179	10973179	C	T	transition
case 9	10973184	10973184	G	A	transition
case 9	10973320	10973320	C	T	transition
case 9	10973351	10973351	C	T	transition
case 9	10973588	10973588	G	A	transition
case 9	10973686	10973686	C	T	transition
case 9	10974109	10974109	C	G	transversion
case 9	10974123	10974123	G	A	transition
case 10	10971954	10971954	C	A	transversion
case 10	10972229	10972229	G	C	transversion
case 10	10972349	10972349		del C	deletion
case 10	10972417	10972424		del 8	deletion

Case #	SNV or indel Position Start (hg19)	SNV or indel Position End (hg19)	ref	alt	type of genetic alteration
case 10	10972482	10972482	G	C	transversion
case 10	10972708	10972724		del 17	deletion
case 10	10972949	10973705		del 757	deletion
case 11	10971525	10971525	C	T	transition
case 11	10971666	10972721		del 1056	deletion
case 11	10972742	10972742	G	A	transition
case 11	10972770	10972770	G	T	transversion
case 11	10972916	10972927		del 12	deletion
case 11	10972942	10972942	G	C	transversion
case 11	10973054	10973054	G	C	transversion
case 11	10973097	10973097	C	T	transition
case 11	10973112	10973114		del 3	deletion
case 11	10973165	10973165	C	T	transition
case 11	10973253	10973257		del 5	deletion
case 11	10973686	10973686	C	T	transition
case 11	10973777	10973777	C	T	transition
case 11	10973781	10973781	G	C	transversion
case 11	10973784	10973784	G	T	transversion
case 11	10973821	10973821	G	A	transition
case 11	10973835	10973835	C	A	transversion
case 16 allele 1	10971159	10971166		del 8	deletion
case 16 allele 1	10971218	10971218	T	C	transition
case 16 allele 1	10971222	10971222	A	C	transversion
case 16 allele 1	10971462	10971462	T	A	transversion
case 16 allele 1	10971673	10971673	C	T	transition
case 16 allele 1	10971953	10971953	G	A	transition
case 16 allele 1	10972141	10972141	G	A	transition
case 16 allele 1	10972415	10972415	C	T	transition
case 16 allele 1	10972527	10973026		del 500	deletion
case 16 allele 1	10973071	10973085		del 15	deletion
case 16 allele 1	10973089	10973089	C	T	transition
case 16 allele 1	10974122	10974122	A	C	transversion
case 16 allele 2	10971225	10971377		del 153	deletion
case 16 allele 2	10971946	10971946	G	A	transition
case 16 allele 2	10972482	10973617		del 1136	deletion
case 16 allele 2	10973899	10973899	A	G	transition
case 17 allele 1	10971352	10971352	G	C	transversion
case 17 allele 1	10971395	10971395	G	C	transversion
case 17 allele 1	10971546	10971546	C	T	transition
case 17 allele 1	10971587	10971997		del 411	deletion
case 17 allele 1	10972017	10972017	T	A	transversion
case 17 allele 1	10972208	10972208	G	C	transversion
case 17 allele 1	10972299	10972299	G	A	transition
case 17 allele 1	10972415	10972415	C	T	transition
case 17 allele 1	10972423	10972423	G	A	transition

Case #	SNV or indel Position Start (hg19)	SNV or indel Position End (hg19)	ref	alt	type of genetic alteration
case 17 allele 1	10972506	10972520		del 15	deletion
case 17 allele 1	10972522	10972522	C	T	transition
case 17 allele 1	10972534	10972534	C	G	transversion
case 17 allele 1	10972570	10972570	G	A	transition
case 17 allele 1	10972577	10972577	G	A	transition
case 17 allele 1	10972634	10972634	C	G	transversion
case 17 allele 1	10972674	10972674	G	C	transversion
case 17 allele 1	10972677	10972691		del 15	deletion
case 17 allele 1	10972694	10972694	C	G	transversion
case 17 allele 1	10972704	10972704	C	T	transition
case 17 allele 1	10972719	10972719	G	C	transversion
case 17 allele 1	10972747	10972747	G	C	transversion
case 17 allele 1	10972750	10972750	C	T	transition
case 17 allele 1	10972763	10972763	G	A	transition
case 17 allele 1	10972777	10972777	G	A	transition
case 17 allele 1	10972786	10972786	G	A	transition
case 17 allele 1	10972797	10972797	G	C	transversion
case 17 allele 1	10972799	10972810		del 12	deletion
case 17 allele 1	10972870	10972870	G	A	transition
case 17 allele 1	10972879	10972879	G	C	transversion
case 17 allele 1	10972993	10972993	G	C	transversion
case 17 allele 1	10973096	10973096	C	T	transition
case 17 allele 1	10973097	10973097	C	T	transition
case 17 allele 1	10973103	10973103		del C	deletion
case 17 allele 1	10973123	10973123	G	T	transversion
case 17 allele 1	10973165	10973165	C	T	transition
case 17 allele 1	10973186	10973186	G	A	transition
case 17 allele 1	10973351	10973351	C	G	transversion
case 17 allele 1	10973364	10973364	G	C	transversion
case 17 allele 1	10973814	10973814	G	A	transition
case 17 allele 2	10971388	10971388	T	A	transversion
case 17 allele 2	10971660	10971660	C	A	transversion
case 17 allele 2	10971829	10971829	C	G	transversion
case 17 allele 2	10971921	10971921	G	C	transversion
case 17 allele 2	10971925	10971925	G	A	transition
case 17 allele 2	10972019	10972019	C	T	transition
case 17 allele 2	10972094	10972094	G	T	transversion
case 17 allele 2	10972200	10972200	C	G	transversion
case 17 allele 2	10972342	10972342	G	A	transition
case 17 allele 2	10972343	10972365		del 23	deletion
case 17 allele 2	10972396	10972396	G	C	transversion
case 17 allele 2	10972411	10972414		del 4	deletion
case 17 allele 2	10972419	10972429		del 11	deletion
case 17 allele 2	10972468	10972501		dup 34	duplication
case 17 allele 2	10972494	10972494	C	T	transition

Case #	SNV or indel Position Start (hg19)	SNV or indel Position End (hg19)	ref	alt	type of genetic alteration
case 18	10971673	10971673	C	T	transition
case 18	10971940	10971940	G	A	transition
case 18	10972031	10972031	G	A	transition
case 18	10972032	10972032	G	C	transversion
case 18	10972035	10972058		del	deletion
case 18	10972293	10972293	G	C	transversion
case 18	10972304	10972304	G	A	transition
case 18	10972327	10972327	G	C	transversion
case 18	10972361	10972361	G	T	transversion
case 18	10972423	10972423	G	A	transition
case 18	10972550	10972550	C	G	transversion
case 18	10972562	10972562	G	A	transition
case 18	10972570	10972570	G	C	transversion
case 18	10972660	10972660	G	A	transition
case 18	10972679	10972679	C	T	transition
case 18	10972749	10972749	G	A	transition
case 18	10972782	10972782	G	A	transition
case 18	10973040	10973040	A	C	transversion
case 18	10973040	10973040	A	C	transversion
case 18	10973041	10973042		ins 11	insertion
case 18	10973042	10973042	A	G	transition
case 18	10973111	10973111	G	C	transversion
case 18	10973315	10973315	C	T	transition
case 19	10974015	10974015	A	G	transition
case 20	10971446	10971446	T	C	transition
case 20	10971574	10971574	A	G	transition
case 20	10971584	10971584	G	A	transition
case 20	10971635	10971635	G	A	transition
case 20	10972052	10972052	T	C	transition
case 20	10972240	10972444		del 205	deletion
case 20	10972490	10972490	G	A	transition
case 20	10972750	10972750	C	T	transition
case 20	10972798	10972798	G	T	transversion
case 20	10972818	10972818	G	C	transversion
case 20	10973014	10973014	G	T	transversion
case 20	10973103	10973103	C	T	transition
case 20	10973123	10973123	G	A	transition
case 20	10973252	10973322		del 71	deletion
case 23	10971239	10971239	G	A	transition
case 23	10971240	10971240	G	C	transversion
case 23	10971903	10971903	C	T	transition
case 23	10972119	10972132		del 14	deletion
case 23	10972703	10972703	G	C	transversion
case 23	10972782	10972790		del 9	deletion
case 24	10971673	10971673	C	A	transversion

Case #	SNV or indel Position Start (hg19)	SNV or indel Position End (hg19)	ref	alt	type of genetic alteration
case 24	10971892	10971892	C	A	transversion
case 24	10971903	10971903	C	T	transition
case 24	10972016	10972495		del 480	deletion
case 24	10972504	10972504	C	T	transition
case 24	10972713	10972713	G	A	transition
case 24	10972715	10973045		del 331	deletion
case 24	10973078	10973079		del 2	deletion
case 24	10973193	10973214		del 22	deletion
case 24	10973254	10973254	C	A	transversion
case 24	10973621	10973621	C	A	transversion
case 24	10973686	10973686	C	T	transition
case 24	10974096	10974096	T	C	transition
case 25 allele 1	10971494	10971514		dup 21	duplication
case 25 allele 1	10971514	10971514	C	G	transversion
case 25 allele 1	10971545	10971545		del G	deletion
case 25 allele 1	10971699	10971699	A	G	transition
case 25 allele 1	10971840	10971840	G	A	transition
case 25 allele 1	10971854	10971854	G	A	transition
case 25 allele 1	10971855	10971855	C	T	transition
case 25 allele 1	10971856	10971856	C	T	transition
case 25 allele 1	10972200	10972200	C	T	transition
case 25 allele 1	10972208	10972208	G	C	transversion
case 25 allele 1	10972317	10972317	C	T	transition
case 25 allele 1	10972423	10972444		del 22	deletion
case 25 allele 1	10972471	10972471	G	A	transition
case 25 allele 1	10972649	10972649	A	C	transversion
case 25 allele 1	10972679	10972679	C	G	transversion
case 25 allele 1	10972743	10972743	C	A	transversion
case 25 allele 1	10972819	10972819	C	T	transition
case 25 allele 1	10972870	10972875		del 6	deletion
case 25 allele 1	10972906	10973687		del 782	deletion
case 25 allele 1	10973691	10973691	C	T	transition
case 25 allele 1	10973713	10973713	C	G	transversion
case 25 allele 1	10974124	10974124	G	A	transition
case 25 allele 2	10971654	10971654	A	G	transition
case 25 allele 2	10971954	10971954	C	T	transition
case 25 allele 2	10972521	10972521	G	A	transition
case 25 allele 2	10972619	10972633		del 15	deletion
case 25 allele 2	10972743	10972750		del 8	deletion
case 25 allele 2	10972873	10972873	G	A	transition
case 25 allele 2	10973212	10973215		del 4	deletion
case 25 allele 2	10973269	10973269	C	T	transition
case 25 allele 2	10973274	10973274	C	A	transversion
case 25 allele 2	10973586	10973586	A	C	transversion
case 25 allele 2	10974124	10974124	G	A	transition

Case #	SNV or indel Position Start (hg19)	SNV or indel Position End (hg19)	ref	alt	type of genetic alteration
case 26	10971346	10971346	G	A	transition
case 26	10972201	10972201		del C	deletion
case 26	10972348	10972348	C	T	transition
case 26	10972414	10972428		del 15	deletion
case 26	10972570	10972570	G	C	transversion
case 26	10972577	10972577	G	A	transition
case 26	10973124	10973124	C	T	transition
case 26	10973268	10973282		del 15	deletion
case 26	10973314	10973314	G	C	transversion
case 26	10973418	10973441		del 24	deletion
case 26	10973588	10973588	G	T	transversion
case 26	10973610	10973621		del 12	deletion
case 26	10973813	10973813	A	G	transition
case 26	10973914	10973914	G	A	transition
case 27	10971581	10971581	G	C	transversion
case 27	10972365	10972365	C	G	transversion
case 27	10973111	10973111	G	A	transition
case 27	10973789	10973813		del 25	deletion
case 28	10971989	10971989	G	A	transition
case 28	10972472	10972472	G	A	transition
case 28	10972845	10972852		del 8	deletion
case 28	10973478	10973478	G	A	transition
case 28	10973583	10973583	G	A	transition
case 29 allele 1	10971545	10971545	G	A	transition
case 29 allele 1	10971627	10973574		del 1948	deletion
case 29 allele 1	10973585	10973592		del 8	deletion
case 29 allele 1	10973690	10973690	G	C	transversion
case 29 allele 1	10973712	10973712	G	C	transversion
case 29 allele 2	10971730	10971730	C	G	transversion
case 29 allele 2	10972000	10972000	C	G	transversion
case 29 allele 2	10972093	10972646		del 554	deletion
case 29 allele 2	10972679	10972679	C	G	transversion
case 29 allele 2	10972725	10972725	T	A	transversion
case 29 allele 2	10972743	10972876		del 134	deletion
case 29 allele 2	10973052	10973052	C	T	transition
case 29 allele 2	10973124	10973124	C	T	transition
case 29 allele 2	10973148	10973148	G	T	transversion
case 29 allele 2	10973468	10973468	G	A	transition
case 29 allele 2	10973484	10973484	T	C	transition
case 29 allele 2	10973613	10973613	T	C	transition
case 29 allele 2	10973854	10973854	G	A	transition
case 29 allele 2	10974016	10974016	C	G	transversion
case 30	10971240	10971240	G	A	transition
case 30	10971356	10973764		del 2409	deletion
case 30	10973977	10973977	C	T	transition

Case #	SNV or indel Position Start (hg19)	SNV or indel Position End (hg19)	ref	alt	type of genetic alteration
case 31	10973002	10973002	G	A	transition
case 31	10973078	10973078	G	A	transition
case 31	10973394	10973394	C	T	transition
case 31	10973426	10973426	C	T	transition
case 31	10973846	10973846	C	T	transition
case 32 allele 1	10972019	10972019	C	A	transversion
case 32 allele 1	10972021	10972021	G	A	transition
case 32 allele 1	10972180	10972180	C	T	transition
case 32 allele 1	10972255	10972255	C	T	transition
case 32 allele 1	10972387	10972388		ins 4	insertion
case 32 allele 1	10972562	10972571		del 10 ins T	deletion
case 32 allele 1	10972737	10972743		del 7	deletion
case 32 allele 1	10972847	10972847	C	G	transversion
case 32 allele 1	10973090	10973090	C	T	transition
case 32 allele 1	10973103	10973103	C	T	transition
case 32 allele 1	10973111	10973111	G	C	transversion
case 32 allele 1	10973183	10973194		del 12	deletion
case 32 allele 2	10972000	10972000	C	G	transversion
case 32 allele 2	10972019	10972019	C	T	transition
case 32 allele 2	10972078	10972078	G	C	transversion
case 32 allele 2	10972129	10972315		del 187	deletion
case 32 allele 2	10972414	10972414	G	A	transition
case 32 allele 2	10972521	10972521	G	A	transition
case 32 allele 2	10972748	10972763		del 16	deletion
case 32 allele 2	10972798	10972818		del 21	deletion
case 32 allele 2	10972847	10972847	C	T	transition
case 32 allele 2	10972870	10972870	G	A	transition
case 32 allele 2	10973060	10973060	C	T	transition
case 32 allele 2	10973103	10973103	C	T	transition
case 32 allele 2	10973124	10973124	C	T	transition
case 32 allele 2	10973315	10973315	C	T	transition
case 32 allele 2	10973360	10973360	C	T	transition
case 32 allele 2	10973459	10973473		del 15	deletion
case 32 allele 2	10973757	10973757	C	T	transition
case 32 allele 2	10973962	10973962	C	G	transversion
case 32 allele 2	10974070	10974081		del 12	deletion
case 32 allele 2	10974134	10974145		del 12	deletion

B.3 Intron 1 alterations in tFL cases

Case #	SNV or indel Position Start (hg19)	SNV or indel Position End (hg19)	ref	alt	type of genetic alteration

Case #	SNV or indel Position Start (hg19)	SNV or indel Position End (hg19)	ref	alt	type of genetic alteration
FL1001T1	10973365	10973365	G	A	transition
FL1001T2	10973365	10973365	G	A	transition
FL1002T1	10971690	10971690	G	A	transition
FL1002T1	10972967	10972967	G	A	transition
FL1002T1	10973365	10973365	G	A	transition
FL1002T1	10974732	10974732	G	A	transition
FL1002T2	10971672	10971672	G	A	transition
FL1002T2	10972967	10972967	G	A	transition
FL1002T2	10973365	10973365	G	A	transition
FL1002T2	10974732	10974732	G	A	transition
FL1004T1	10972128	10972128	G	A	transition
FL1004T2	10972128	10972128	G	A	transition
FL1006T1	10971524	10971524	G	A	transition
FL1006T2	10971524	10971524	G	A	transition
FL1006T2	10971672	10971672	G	T	transversion
FL1006T2	10973146	10973146	C	T	transition
FL1006T2	10973520	10973520	C	T	transition
FL1006T2	10973595	10973595	C	T	transition
FL1010T1	10972019	10972019	C	T	transition
FL1010T1	10974165	10974165	G	A	transition
FL1010T1	10974364	10974364	C	T	transition
FL1010T2	10972019	10972019	C	T	transition
FL1010T2	10974165	10974165	G	A	transition
FL1011T1	10972864	10972865		ins	insertion
FL1011T1	10974856	10974856	A	C	transversion
FL1011T1	10974902	10974902	G	A	transition
FL1011T1	10974905	10974905	T	A	transversion
FL1011T1	10974906	10974906	G	A	transition
FL1011T1	10974918	10974918	G	C	transversion
FL1011T1	10974928	10974928	C	T	transition
FL1011T1	10974960	10974960	A	G	transition
FL1011T2	10972864	10972865		ins	insertion
FL1014T1	10972424	10972424	C	T	transition
FL1014T1	10972679	10972679	C	T	transition
FL1014T2	10972424	10972424	C	T	transition
FL1014T2	10972637	10972637	C	A	transversion
FL1014T2	10973058	10973065		del	deletion
FL1014T2	10973111	10973111	G	C	transversion
FL1014T2	10973730	10973737		del	deletion
FL1016T2	10971426	10971426	C	T	transition
FL1017T1	10973085	10973098		del	deletion
FL1017T1	10973098	10973098	C	A	transversion
FL1017T1	10973917	10973917	C	G	transversion
FL1017T2	10971789	10971789	G	A	transition

Case #	SNV or indel Position Start (hg19)	SNV or indel Position End (hg19)	ref	alt	type of genetic alteration
FL1017T2	10972615	10972615	G	A	transition
FL1017T2	10973085	10973098		del	deletion
FL1017T2	10973098	10973098	C	A	transversion
FL1018T1	10972743	10972743	C	G	transversion
FL1018T2	10971672	10971672	G	A	transition
FL1019T1	10971549	10971549	G	A	transition
FL1019T1	10972407	10972407	G	C	transversion
FL1019T1	10972409	10972409	G	T	transversion
FL1019T1	10972521	10972523		del	deletion
FL1019T1	10972555	10972555	C	A	transversion
FL1019T1	10972562	10972562	G	A	transition
FL1019T1	10973686	10973686	C	T	transition
FL1019T1	10973687	10973687	T	A	transversion
FL1019T1	10973773	10973773	A	C	transversion
FL1019T1	10973942	10973942	C	G	transversion
FL1019T1	10974153	10974153	G	C	transversion
FL1019T1	10974191	10974191	C	T	transition
FL1019T2	10971701	10971701	C	T	transition
FL1019T2	10972032	10972032	G	A	transition
FL1019T2	10972048	10972048	C	T	transition
FL1019T2	10972121	10972121	G	A	transition
FL1019T2	10972521	10972523		del	deletion
FL1019T2	10972661	10972661	C	T	transition
FL1019T2	10972801	10972801	G	C	transversion
FL1019T2	10973533	10973533	C	G	transversion
FL1019T2	10974191	10974191	C	T	transition
FL1101T1	10972367	10972367	G	A	transition
FL1101T1	10972419	10972419	G	T	transversion
FL1101T1	10973063	10973063	C	T	transition
FL1101T2	10971546	10971546	C	T	transition
FL1101T2	10972340	10972340	C	G	transversion
FL1101T2	10972367	10972367	G	A	transition
FL1101T2	10973063	10973063	C	T	transition
FL1101T2	10973530	10973530	G	A	transition
FL1102T1	10972939	10972939	G	T	transversion
FL1102T1	10989650	10989650	A	G	transition
FL1102T2	10972939	10972939	G	T	transversion
FL1103T1	10972749	10972749	G	A	transition
FL1103T2	10972749	10972749	G	A	transition
FL1105T1	10971545	10971545	G	A	transition
FL1105T2	10971545	10971545	G	A	transition
FL1105T2	10973082	10973082	C	T	transition
FL1106T1	10972397	10972397	C	T	transition
FL1106T2	10972397	10972397	C	T	transition

Case #	SNV or indel Position Start (hg19)	SNV or indel Position End (hg19)	ref	alt	type of genetic alteration
FL1107T2	10975028	10975028	T	C	transition
FL1108T1	10972750	10972750	C	G	transversion
FL1108T1	10973829	10973829	C	G	transversion
FL1108T1	10974109	10974109	C	T	transition
FL1108T1	10975020	10975020	C	T	transition
FL1108T2	10972750	10972750	C	G	transversion
FL1108T2	10973204	10973204	G	A	transition
FL1108T2	10973829	10973829	C	G	transversion
FL1108T2	10974109	10974109	C	T	transition
FL1108T2	10975028	10975028	T	C	transition
FL1108T2	10975061	10975061	A	G	transition
FL1110T1	10972661	10972661	C	T	transition
FL1110T1	10973314	10973314	G	A	transition
FL1110T1	10973986	10973998		del	deletion
FL1110T1	10974928	10974928	C	T	transition
FL1110T2	10971315	10971315	C	G	transversion
FL1110T2	10971316	10971316	C	T	transition
FL1110T2	10971379	10971379	A	G	transition
FL1110T2	10971513	10971513	A	T	transversion
FL1110T2	10971700	10971700	G	A	transition
FL1110T2	10971855	10971855	C	T	transition
FL1110T2	10971954	10971954	C	T	transition
FL1110T2	10972105	10972109		del	deletion
FL1110T2	10972396	10972396	G	A	transition
FL1110T2	10972424	10972424	GC	G	deletion
FL1110T2	10972482	10972482	G	A	transition
FL1110T2	10972570	10972570	G	A	transition
FL1110T2	10972679	10972679	C	T	transition
FL1110T2	10972693	10972693	G	T	transversion
FL1110T2	10972696	10972696	C	T	transition
FL1110T2	10972743	10972743	C	T	transition
FL1110T2	10972797	10972797	G	C	transversion
FL1110T2	10972819	10972819	C	T	transition
FL1110T2	10972822	10972822	C	T	transition
FL1110T2	10972844	10972844	G	A	transition
FL1110T2	10972846	10972846	G	A	transition
FL1110T2	10972886	10972886	C	T	transition
FL1110T2	10972975	10972975	C	T	transition
FL1110T2	10973103	10973103	C	T	transition
FL1110T2	10973111	10973111	G	T	transversion
FL1110T2	10973277	10973277	G	C	transversion
FL1110T2	10973314	10973314	G	A	transition
FL1110T2	10973351	10973351	C	T	transition
FL1110T2	10973568	10973568	C	T	transition

Case #	SNV or indel Position Start (hg19)	SNV or indel Position End (hg19)	ref	alt	type of genetic alteration
FL1110T2	10973614	10973614	G	A	transition
FL1111T2	10972195	10972195	C	T	transition
FL1111T2	10973473	10973473	G	T	transversion
FL1112T1	10973124	10973124	C	T	transition
FL1114T2	10973728	10973728	G	T	transversion
FL1116T1	10972423	10972423	G	A	transition
FL1116T2	10972423	10972423	G	A	transition
FL1117T1	10971672	10971672	G	A	transition
FL1117T2	10971672	10971672	G	A	transition
FL1118T1	10972570	10972570	G	A	transition
FL1118T2	10972570	10972570	G	A	transition
FL1118T2	10972967	10972967	G	A	transition
FL1119T1	10972397	10972397	C	T	transition
FL1120T1	10971546	10971546	C	T	transition
FL1120T1	10972141	10972141	G	A	transition
FL1120T1	10972316	10972316	G	A	transition
FL1120T1	10972747	10972747	G	A	transition
FL1120T1	10972782	10972782	G	A	transition
FL1120T1	10973165	10973165	C	T	transition
FL1120T1	10973168	10973168	C	T	transition
FL1120T1	10973326	10973326	C	T	transition
FL1120T1	10973351	10973351	C	G	transversion
FL1120T1	10974195	10974195	C	G	transversion
FL1120T2	10971546	10971546	C	T	transition
FL1120T2	10972747	10972747	G	A	transition
FL1120T2	10972782	10972782	G	A	transition
FL1122T2	10972629	10972629	G	A	transition
FL1122T2	10972783	10972783	C	T	transition
FL1122T2	10973108	10973108	C	G	transversion
FL1127T1	10974148	10974148	G	C	transversion
FL1127T1	10974384	10974384	G	A	transition
FL1127T1	10974905	10974905	T	A	transversion
FL1127T1	10974918	10974918	G	C	transversion
FL1127T1	10975051	10975051	A	G	transition
FL1128T1	10972419	10972419	G	C	transversion
FL1128T1	10972422	10972422	A	G	transition
FL1128T1	10972740	10972740	G	A	transition
FL1128T1	10973205	10973205	C	T	transition
FL1128T2	10972419	10972419	G	C	transversion
FL1128T2	10972422	10972422	A	G	transition
FL1128T2	10972740	10972740	G	A	transition
FL1128T2	10973103	10973103	C	T	transition
FL1128T2	10973205	10973205	C	T	transition
FL1128T2	10973691	10973691	C	T	transition

Case #	SNV or indel Position Start (hg19)	SNV or indel Position End (hg19)	ref	alt	type of genetic alteration
FL1130T2	10971903	10971903	C	T	transition
FL1130T2	10971990	10971990	C	T	transition
FL1130T2	10972338	10972338	C	T	transition
FL1130T2	10972570	10972570	G	A	transition
FL1130T2	10972777	10972777	G	C	transversion
FL1130T2	10973277	10973277	G	C	transversion
FL1132T1	10973115	10973115	T	C	transition
FL1132T1	10973078	10973118		del	deletion
FL1132T1	10973118	10973118	C	G	transversion
FL1132T1	10974841	10974841	C	T	transition
FL1132T2	10973115	10973115	T	C	transition
FL1132T2	10973078	10973118		del	deletion
FL1132T2	10973118	10973118	C	G	transversion
FL1132T2	10974841	10974841	C	T	transition
FL1134T1	10971467	10971467	G	C	transversion
FL1134T1	10971989	10971989	G	C	transversion
FL1134T1	10972482	10972482	G	C	transversion
FL1134T1	10972870	10972870	G	A	transition
FL1134T1	10973124	10973124	C	T	transition
FL1134T2	10971467	10971467	G	C	transversion
FL1134T2	10971545	10971545	G	A	transition
FL1134T2	10971672	10971672	G	A	transition
FL1134T2	10971921	10971921	G	A	transition
FL1134T2	10971929	10971929	G	C	transversion
FL1134T2	10972118	10972124		del	deletion
FL1134T2	10972364	10972364	C	T	transition
FL1134T2	10972424	10972424	C	T	transition
FL1134T2	10972577	10972577	G	A	transition
FL1134T2	10973112	10973112	C	T	transition
FL1134T2	10973315	10973315	C	T	transition
FL1134T2	10973784	10973784	G	T	transversion
FL1134T2	10973896	10973896	G	A	transition
FL1134T2	10974175	10974175	G	A	transition
FL1135T1	10973455	10973455	G	A	transition
FL1135T2	10971672	10971672	G	A	transition
FL1135T2	10971979	10971979	C	T	transition
FL1135T2	10972013	10972013	G	A	transition
FL1135T2	10972045	10972045	T	A	transversion
FL1135T2	10972571	10972571	C	G	transversion
FL1135T2	10973455	10973455	G	A	transition
FL1135T2	10973479	10973479	T	G	transversion
FL1135T2	10973492	10973492	G	T	transversion
FL1135T2	10973610	10973610	G	A	transition
FL1135T2	10973690	10973690	G	A	transition

Case #	SNV or indel Position Start (hg19)	SNV or indel Position End (hg19)	ref	alt	type of genetic alteration
FL1135T2	10974565	10974565	C	T	transition
FL1136T1	10973021	10973021	C	T	transition
FL1136T2	10973021	10973021	C	T	transition
FL1136T2	10975009	10975009	A	G	transition
FL1140T1	10972926	10972926	C	T	transition
FL1140T2	10971672	10971672	G	A	transition
FL1140T2	10972299	10972299	G	A	transition
FL1143T1	10971821	10971821	C	T	transition
FL1143T1	10972733	10972733	C	A	transversion
FL1143T1	10973690	10973690	G	A	transition
FL1143T2	10971821	10971821	C	T	transition
FL1143T2	10972322	10972322	G	A	transition
FL1143T2	10972733	10972733	C	A	transversion
FL1143T2	10972913	10972913	G	T	transversion
FL1143T2	10973690	10973690	G	A	transition
FL1145T1	10973257	10973257	C	T	transition
FL1145T1	11002015	11002015	A	G	transition
FL1147T1	10971318	10971318	T	C	transition
FL1147T1	10971635	10971635	G	A	transition
FL1147T1	10972019	10972019	C	T	transition
FL1147T1	10972592	10972592	C	A	transversion
FL1147T1	10972604	10972604	C	A	transversion
FL1147T1	10972742	10972742	G	C	transversion
FL1147T1	10972797	10972797	G	C	transversion
FL1147T1	10972968	10972968	C	G	transversion
FL1147T1	10973103	10973103	C	T	transition
FL1147T1	10973695	10973695	G	A	transition
FL1147T1	10974181	10974181	G	A	transition
FL1147T2	10971318	10971318	T	C	transition
FL1147T2	10971635	10971635	G	A	transition
FL1147T2	10972592	10972592	C	A	transversion
FL1147T2	10972604	10972604	C	A	transversion
FL1147T2	10972742	10972742	G	C	transversion
FL1147T2	10972797	10972797	G	C	transversion
FL1147T2	10972968	10972968	C	G	transversion
FL1147T2	10973695	10973695	G	A	transition
FL1147T2	10974181	10974181	G	A	transition
FL1148T1	10974902	10974902	G	A	transition
FL1148T1	10974905	10974905	T	A	transversion
FL1148T1	10974918	10974918	G	C	transversion
FL1148T1	10974928	10974928	C	T	transition
FL1148T1	10975061	10975061	A	G	transition
FL1148T2	10974905	10974905	T	C	transition
FL1148T2	10975061	10975061	A	G	transition

Case #	SNV or indel Position Start (hg19)	SNV or indel Position End (hg19)	ref	alt	type of genetic alteration
FL1149T1	10972948	10972948	C	T	transition
FL1149T1	10973019	10973019	T	C	transition
FL1149T2	10973019	10973019	T	C	transition
FL1149T2	10973312	10973312	G	C	transversion
FL1149T2	10973690	10973690	G	C	transversion
FL1150T1	10972311	10972311	C	A	transversion
FL1150T1	10973411	10973411	G	T	transversion
FL1150T2	10972311	10972311	C	A	transversion
FL1150T2	10973411	10973411	G	T	transversion
FL1153T1	10971989	10971989	G	A	transition
FL1153T2	10971989	10971989	G	A	transition
FL1160T1	10972397	10972397	C	T	transition
FL1160T1	10973686	10973686	C	G	transversion
FL1160T1	10974313	10974313	G	A	transition
FL1160T2	10972397	10972397	C	T	transition
FL1160T2	10973686	10973686	C	G	transversion
FL1160T2	10974313	10974313	G	A	transition
FL1162T1	10972846	10972846	G	A	transition
FL1162T1	10974901	10974901	C	T	transition
FL1162T2	10972846	10972846	G	A	transition
FL1170T1	10975005	10975005	G	T	transversion
FL1170T1	10975028	10975028	T	C	transition
FL1175T1	10972719	10972719	G	A	transition
FL1179T1	10972348	10972348	C	T	transition
FL1179T1	10972448	10972460		del	deletion
FL1179T1	10972511	10972511	G	C	transversion
FL1179T1	10972571	10972571	C	T	transition
FL1179T1	10973314	10973314	G	A	transition
FL1179T2	10971878	10971878	C	G	transversion
FL1179T2	10971966	10971966	T	G	transversion
FL1179T2	10973314	10973314	G	A	transition
FL1188T1	10971525	10971525	C	T	transition
FL1188T1	10998557	10998557	G	A	transition
FL1188T2	10971470	10971470	C	T	transition
FL1188T2	10998557	10998557	G	A	transition
FL1189T2	10974008	10974008	G	A	transition
FL1190T1	10972847	10972847	C	G	transversion
FL1190T1	10974348	10974348	G	A	transition
FL1192T1	10972311	10972311	C	T	transition
FL1192T1	10973010	10973010	G	C	transversion
FL1192T1	10973103	10973103	C	A	transversion
FL1192T2	10971829	10971829	C	T	transition
FL1192T2	10972311	10972311	C	T	transition
FL1192T2	10973010	10973010	G	C	transversion

Case #	SNV or indel Position Start (hg19)	SNV or indel Position End (hg19)	ref	alt	type of genetic alteration
FL1192T2	10973103	10973103	C	A	transversion
FL1193T1	10972770	10972770	G	T	transversion
FL1193T1	10972846	10972846	G	A	transition
FL1194T1	10971520	10971520	G	A	transition
FL1194T1	10971672	10971672	G	A	transition
FL1194T1	10971765	10971765	G	A	transition
FL1194T1	10971834	10971834	C	T	transition
FL1194T1	10971953	10971953	G	C	transversion
FL1194T1	10972138	10972138	G	C	transversion
FL1194T1	10972222	10972222	G	A	transition
FL1194T1	10972252	10972253		ins	insertion
FL1194T1	10972254	10972255		ins	insertion
FL1194T1	10972257	10972258		ins	insertion
FL1194T1	10972251	10972260		del	deletion
FL1194T1	10972314	10972314	C	T	transition
FL1194T1	10972364	10972364	C	T	transition
FL1194T1	10972423	10972423	G	A	transition
FL1194T1	10972522	10972522	C	T	transition
FL1194T1	10972573	10972573	C	T	transition
FL1194T1	10972596	10972596	G	A	transition
FL1194T1	10972599	10972599	C	G	transversion
FL1194T1	10972633	10972633	C	T	transition
FL1194T1	10972651	10972651	C	T	transition
FL1194T1	10972694	10972694	C	T	transition
FL1194T1	10972742	10972742	G	A	transition
FL1194T1	10972797	10972797	G	C	transversion
FL1194T1	10972818	10972818	G	A	transition
FL1194T1	10972840	10972846		del	deletion
FL1194T1	10972858	10972858	G	C	transversion
FL1194T1	10972974	10972974	G	C	transversion
FL1194T1	10973001	10973001	G	A	transition
FL1194T1	10973014	10973014	G	T	transversion
FL1194T1	10973089	10973089	C	A	transversion
FL1194T1	10973092	10973092	G	T	transversion
FL1194T1	10973123	10973123	G	A	transition
FL1194T1	10973164	10973164	G	C	transversion
FL1194T1	10973202	10973202	G	A	transition
FL1194T1	10973204	10973204	G	A	transition
FL1194T1	10973250	10973250	G	A	transition
FL1194T1	10973351	10973351	C	G	transversion
FL1194T1	10973533	10973533	C	T	transition
FL1194T1	10973690	10973690	G	C	transversion
FL1194T1	10973691	10973691	C	T	transition
FL1194T1	10973713	10973713	C	G	transversion

Case #	SNV or indel Position Start (hg19)	SNV or indel Position End (hg19)	ref	alt	type of genetic alteration
FL1194T1	10973829	10973829	C	T	transition
FL1194T1	10973884	10973884	T	A	transversion
FL1194T1	10973886	10973886	C	G	transversion
FL1194T1	10974098	10974098	G	A	transition
FL1194T1	10974424	10974424	C	G	transversion
FL1194T2	10971450	10971450	G	C	transversion
FL1194T2	10971520	10971520	G	A	transition
FL1194T2	10971672	10971672	G	A	transition
FL1194T2	10971765	10971765	G	A	transition
FL1194T2	10971834	10971834	C	T	transition
FL1194T2	10971953	10971953	G	C	transversion
FL1194T2	10972138	10972138	G	C	transversion
FL1194T2	10972222	10972222	G	A	transition
FL1194T2	10972252	10972253		ins	insertion
FL1194T2	10972254	10972255		ins	insertion
FL1194T2	10972257	10972258		ins	insertion
FL1194T2	10972251	10972260		del	deletion
FL1194T2	10972314	10972314	C	T	transition
FL1194T2	10972364	10972364	C	T	transition
FL1194T2	10972423	10972423	G	A	transition
FL1194T2	10972522	10972522	C	T	transition
FL1194T2	10972573	10972573	C	T	transition
FL1194T2	10972596	10972596	G	A	transition
FL1194T2	10972599	10972599	C	G	transversion
FL1194T2	10972633	10972633	C	T	transition
FL1194T2	10972651	10972651	C	T	transition
FL1194T2	10972694	10972694	C	T	transition
FL1194T2	10972742	10972742	G	A	transition
FL1194T2	10972797	10972797	G	C	transversion
FL1194T2	10972818	10972818	G	A	transition
FL1194T2	10972840	10972846		del	deletion
FL1194T2	10972858	10972858	G	C	transversion
FL1194T2	10972974	10972974	G	C	transversion
FL1194T2	10973001	10973001	G	A	transition
FL1194T2	10973014	10973014	G	T	transversion
FL1194T2	10973089	10973089	C	A	transversion
FL1194T2	10973092	10973092	G	T	transversion
FL1194T2	10973123	10973123	G	A	transition
FL1194T2	10973202	10973202	G	A	transition
FL1194T2	10973204	10973204	G	A	transition
FL1194T2	10973250	10973250	G	A	transition
FL1194T2	10973351	10973351	C	G	transversion
FL1194T2	10973533	10973533	C	T	transition
FL1194T2	10973690	10973690	G	C	transversion

Case #	SNV or indel Position Start (hg19)	SNV or indel Position End (hg19)	ref	alt	type of genetic alteration
FL1194T2	10973691	10973691	C	T	transition
FL1194T2	10973713	10973713	C	G	transversion
FL1194T2	10973829	10973829	C	T	transition
FL1194T2	10973884	10973884	T	A	transversion
FL1194T2	10973886	10973886	C	G	transversion
FL1194T2	10974098	10974098	G	A	transition
FL1194T2	10974424	10974424	C	G	transversion
FL1197T1	10972770	10972770	G	A	transition
FL1197T1	10973058	10973058	C	T	transition
FL1197T1	10973059	10973059	C	T	transition
FL1197T2	10973058	10973058	C	T	transition
FL1197T2	10973059	10973059	C	T	transition
FL1198T1	10971672	10971672	G	A	transition
FL1198T1	10971673	10971673	C	T	transition
FL1198T1	10972215	10972215	C	T	transition
FL1198T1	10972724	10972724	C	T	transition
FL1198T1	10972726	10972726	C	T	transition
FL1198T1	10972749	10972749	G	A	transition
FL1198T1	10973436	10973436	G	A	transition
FL1198T1	10973630	10973650		del	deletion
FL1198T2	10971545	10971545	G	A	transition
FL1198T2	10971673	10971673	C	T	transition
FL1198T2	10972082	10972082	G	A	transition
FL1198T2	10972129	10972129	C	T	transition
FL1198T2	10972215	10972215	C	T	transition
FL1198T2	10972424	10972424	C	T	transition
FL1198T2	10972724	10972724	C	T	transition
FL1198T2	10972726	10972726	C	T	transition
FL1198T2	10973184	10973184	G	C	transversion
FL1198T2	10973320	10973320	C	T	transition
FL1198T2	10973332	10973332	G	A	transition
FL1198T2	10973372	10973372	G	A	transition
FL1198T2	10973436	10973436	G	A	transition
FL1202T1	10973677	10973677	G	A	transition
FL1202T2	10973677	10973677	G	A	transition
FL1211T2	10972562	10972562	G	T	transversion
FL1211T2	10972975	10972975	C	T	transition
FL1212T1	10972780	10972780	C	T	transition
FL1212T1	10972847	10972847	C	T	transition
FL1212T2	10972780	10972780	C	T	transition
FL1212T2	10972847	10972847	C	T	transition
FL1212T2	10974960	10974960	A	G	transition
FL1216T1	10974901	10974901	C	T	transition
FL1216T2	10974905	10974905	T	C	transition

Case #	SNV or indel Position Start (hg19)	SNV or indel Position End (hg19)	ref	alt	type of genetic alteration
FL1218T1	10972968	10972968	C	T	transition
FL1218T2	10972968	10972968	C	T	transition
FL1219T2	11016378	11016378	A	G	transition
FL1222T1	10973036	10973036	G	T	transversion
FL1222T1	10998557	10998557	G	A	transition
FL1222T2	10972522	10972522	C	T	transition
FL1222T2	10973036	10973036	G	T	transversion
FL1222T2	10973123	10973123	G	A	transition
FL1222T2	10998557	10998557	G	A	transition
FL1223T2	10974491	10974491	C	G	transversion
FL1225T1	10973314	10973314	G	A	transition
FL1225T1	10973814	10973814	G	A	transition
FL1225T1	10974901	10974901	C	T	transition
FL1225T1	10974905	10974905	T	C	transition
FL1225T1	10975005	10975005	G	T	transversion
FL1225T1	10975036	10975036	T	C	transition
FL1225T2	10973314	10973314	G	A	transition
FL1225T2	10973814	10973814	G	A	transition
FL1233T1	10971672	10971672	G	C	transversion
FL1233T1	10972082	10972082	G	A	transition
FL1233T1	10972277	10972277	C	A	transversion
FL1233T1	10972743	10972743	C	T	transition
FL1233T1	10972750	10972750	C	T	transition
FL1233T1	10972967	10972967	G	A	transition
FL1233T1	10973021	10973021	C	T	transition
FL1233T1	10973253	10973253	C	T	transition
FL1233T1	10973685	10973685	G	A	transition
FL1233T1	10974111	10974111	G	A	transition
FL1233T1	10974397	10974397	G	A	transition
FL1233T2	10972082	10972082	G	A	transition
FL1233T2	10972277	10972277	C	A	transversion
FL1233T2	10973253	10973253	C	T	transition
FL1233T2	10973685	10973685	G	A	transition
FL1233T2	10974111	10974111	G	A	transition
FL1234T1	10971409	10971409	C	T	transition
FL1234T1	10971673	10971673	C	T	transition
FL1234T1	10971840	10971840	G	T	transversion
FL1234T1	10971850	10971850	G	C	transversion
FL1234T1	10972032	10972032	G	C	transversion
FL1234T1	10972334	10972334	G	T	transversion
FL1234T1	10972777	10972777	G	A	transition
FL1234T1	10972933	10972933	G	T	transversion
FL1234T1	10973020	10973020	G	A	transition
FL1234T1	10973517	10973517	G	A	transition

Case #	SNV or indel Position Start (hg19)	SNV or indel Position End (hg19)	ref	alt	type of genetic alteration
FL1234T1	10973588	10973588	G	T	transversion
FL1234T1	10973691	10973691	C	T	transition
FL1234T1	10973766	10973766	G	C	transversion
FL1234T2	10971409	10971409	C	T	transition
FL1234T2	10971673	10971673	C	T	transition
FL1234T2	10971840	10971840	G	T	transversion
FL1234T2	10971850	10971850	G	C	transversion
FL1234T2	10972032	10972032	G	C	transversion
FL1234T2	10972334	10972334	G	T	transversion
FL1234T2	10972747	10972747	G	C	transversion
FL1234T2	10972770	10972770	G	A	transition
FL1234T2	10972777	10972777	G	A	transition
FL1234T2	10972933	10972933	G	T	transversion
FL1234T2	10973020	10973020	G	A	transition
FL1234T2	10973123	10973123	G	C	transversion
FL1234T2	10973691	10973691	C	T	transition
FL1240T1	10974905	10974905	T	C	transition
FL1246T1	10971820	10971820	G	A	transition
FL1246T1	10972504	10972504	C	T	transition
FL1246T1	10972846	10972846	G	C	transversion
FL1246T1	10972967	10972967	G	A	transition
FL1246T1	10973261	10973264		del	deletion
FL1246T1	10973752	10973752	G	C	transversion
FL1246T2	10971820	10971820	G	A	transition
FL1246T2	10972504	10972504	C	T	transition
FL1246T2	10972846	10972846	G	C	transversion
FL1248T1	10972129	10972129	C	T	transition
FL1248T1	10972400	10972400	C	T	transition
FL1248T1	10972461	10972461	C	T	transition
FL1248T1	10972732	10972732	C	A	transversion
FL1248T1	10972733	10972733	C	A	transversion
FL1248T1	10972846	10972846		Del G	deletion
FL1248T1	10972952	10972952	G	T	transversion
FL1248T1	10972968	10972968	C	T	transversion
FL1248T1	10973123	10973123	G	A	transition
FL1248T1	10973595	10973598		del	deletion
FL1248T1	10973900	10973900	G	A	transition
FL1248T2	10972082	10972082	G	C	transversion
FL1248T2	10972400	10972400	C	T	transition
FL1248T2	10972486	10972486	G	A	transition
FL1248T2	10972577	10972577	G	A	transition
FL1248T2	10973096	10973096	C	T	transition
FL1248T2	10973164	10973164	G	A	transition
FL1248T2	10973285	10973285	G	A	transition

Case #	SNV or indel Position Start (hg19)	SNV or indel Position End (hg19)	ref	alt	type of genetic alteration
FL1248T2	10973900	10973900	G	A	transition
FL1248T2	10973980	10973980	A	G	transition
FL1249T1	10972042	10972066		del	deletion
FL1249T2	10971940	10971940	G	A	transition
FL1249T2	10972042	10972066		del	deletion
FL1250T1	10973103	10973103	C	T	transition
FL1250T2	10973103	10973103	C	T	transition
FL1254T1	10972682	10972682	C	T	transition
FL1254T1	10972693	10972693	G	A	transition
FL1254T1	10972780	10972780	C	G	transversion
FL1254T2	10972682	10972682	C	T	transition
FL1254T2	10972693	10972693	G	A	transition
FL1254T2	10972780	10972780	C	G	transversion
FL1255T1	10972765	10972765	G	A	transition
FL1255T1	10972770	10972770	G	A	transition
FL1255T1	10972780	10972780	C	T	transition
FL1255T1	10973112	10973112	C	T	transition
FL1255T2	10971672	10971672	G	C	transversion
FL1255T2	10972513	10972513	G	A	transition
FL1255T2	10972685	10972685	C	A	transversion
FL1255T2	10972780	10972780	C	T	transition
FL1255T2	10973014	10973014	G	T	transversion
FL1255T2	10973718	10973718	G	A	transition
FL1255T2	10973951	10973951	C	G	transversion
FL1256T1	10972927	10972927	C	T	transition
FL1256T2	11009406	11009406	C	T	transition
FL1258T1	10973310	10973314		del	deletion
FL1258T2	10973310	10973314		del	deletion
FL1259T1	10971701	10971701	C	T	transition
FL1259T1	10973277	10973277	G	A	transition
FL1259T2	10971701	10971701	C	T	transition
FL1259T2	10973277	10973277	G	A	transition
FL1260T1	10972189	10972189	C	G	transversion
FL1260T1	10972599	10972599	C	G	transversion
FL1260T1	10972780	10972780	C	T	transition
FL1260T1	10973837	10973837	G	A	transition
FL1260T1	11010189	11010189	G	A	transition
FL1260T2	10971672	10971672	G	A	transition
FL1260T2	10972189	10972189	C	G	transversion
FL1260T2	10972424	10972424	C	T	transition
FL1260T2	10972456	10972456	G	A	transition
FL1260T2	10972659	10972659	T	A	transversion
FL1260T2	10972661	10972661	C	A	transversion
FL1260T2	10972780	10972780	C	T	transition

Case #	SNV or indel Position Start (hg19)	SNV or indel Position End (hg19)	ref	alt	type of genetic alteration
FL1260T2	10972759	10972785		del	deletion
FL1260T2	10972733	10972786		del	deletion
FL1260T2	10973102	10973102	G	A	transition
FL1260T2	10973468	10973468	G	A	transition
FL1260T2	11010189	11010189	G	A	transition
FL1261T1	10973686	10973686	C	T	transition
FL1261T2	10974319	10974319	G	A	transition
FL1262T1	10972178	10972178	A	C	transversion
FL1262T1	10972208	10972208	G	C	transversion
FL1262T1	10972461	10972461	C	A	transversion
FL1262T1	10972571	10972571	C	T	transition
FL1262T1	10973102	10973102	G	A	transition
FL1262T2	10971999	10971999	G	A	transition
FL1262T2	10972424	10972424	C	G	transversion
FL1262T2	10972571	10972571	C	T	transition

B.4 Intron 1 alterations in pFL and npFL

Case #	SNV or indel Position Start (hg19)	SNV or indel Position End (hg19)	ref	alt	type of genetic alteration
FL2002T1	10972087	10972087	G	A	transition
FL2002T1	10974524	10974536		del	deletion
FL2005T1	10973123	10973123	G	C	transversion
FL2101T1	10972870	10972870	G	A	transition
FL2101T1	10973082	10973082	C	G	transversion
FL2101T1	10973084	10973084	A	C	transversion
FL2101T1	10973686	10973686	C	T	transition
FL2102T1	10971730	10971730	C	A	transversion
FL2102T1	10972396	10972396	G	A	transition
FL2102T1	10972472	10972472	G	A	transition
FL2102T1	10973103	10973103	C	T	transition
FL2104T1	10971828	10971828	G	A	transition
FL2105T1	10971922	10971922	C	T	transition
FL2108T1	10971690	10971690	G	A	transition
FL2108T1	10972521	10972521	G	A	transition
FL2108T1	10972886	10972886	C	A	transversion
FL2110T1	11012426	11012426	G	C	transversion
FL2114T1	10971789	10971789	G	A	transition
FL2114T1	10972731	10972731	G	A	transition
FL2114T1	10972967	10972967	G	A	transition
FL2114T1	10973112	10973112	C	G	transversion
FL2115T1	10973019	10973019	T	C	transition
FL2115T1	10973900	10973900	G	A	transition
FL2118T1	10971741	10971741	G	A	transition

Case #	SNV or indel Position Start (hg19)	SNV or indel Position End (hg19)	ref	alt	type of genetic alteration
FL2119T1	10973429	10973429	G	A	transition
FL2119T1	10973588	10973588	G	A	transition
FL2120T1	10971672	10971672	G	C	transversion
FL2120T1	10971766	10971766	G	A	transition
FL2120T1	10972675	10972675	G	A	transition
FL2120T1	10973115	10973115	T	G	transversion
FL2120T1	10973126	10973126	C	T	transition
FL2120T1	10973128	10973128	G	T	transversion
FL2120T1	10973551	10973551	G	A	transition
FL2121T1	10971550	10971550	G	A	transition
FL2121T1	10971808	10971808	G	A	transition
FL2121T1	10971915	10971915	G	A	transition
FL2121T1	10972138	10972138	G	A	transition
FL2121T1	10972235	10972235	G	A	transition
FL2121T1	10972334	10972334	G	A	transition
FL2121T1	10972364	10972364	C	T	transition
FL2121T1	10972365	10972365	C	T	transition
FL2121T1	10972482	10972482	G	A	transition
FL2121T1	10972546	10972546	A	G	transition
FL2121T1	10972571	10972571	C	T	transition
FL2121T1	10972718	10972718	T	C	transition
FL2121T1	10972763	10972763	G	C	transversion
FL2121T1	10972798	10972798	G	A	transition
FL2121T1	10972846	10972846	G	A	transition
FL2121T1	10973001	10973001	G	A	transition
FL2121T1	10973111	10973111	G	A	transition
FL2121T1	10973123	10973123	G	A	transition
FL2121T1	10973372	10973372	G	A	transition
FL2121T1	10973610	10973610	G	A	transition
FL2121T1	10973678	10973678	C	T	transition
FL2121T1	10973691	10973691	C	T	transition
FL2121T1	10973759	10973759	C	T	transition
FL2121T1	10973784	10973784	G	A	transition
FL2122T1	10971829	10971829	C	G	transversion
FL2122T1	10971953	10971953	G	A	transition
FL2123T1	10972424	10972424	C	T	transition
FL2125T1	10971954	10971954	C	T	transition
FL2125T1	10972019	10972019	C	T	transition
FL2125T1	10972096	10972096	G	T	transversion
FL2125T1	10972266	10972266	C	A	transition
FL2125T1	10972424	10972424	C	T	transition
FL2125T1	10972494	10972494	C	T	transition
FL2125T1	10972743	10972743	C	A	transversion
FL2125T1	10972769	10972770		del 2	deletion
FL2125T1	10973096	10973096	C	T	transition

Case #	SNV or indel Position Start (hg19)	SNV or indel Position End (hg19)	ref	alt	type of genetic alteration
FL2125T1	10973102	10973102	G	A	transition
FL2125T1	10973111	10973111	G	C	transversion
FL2125T1	10973685	10973685	G	A	transition
FL2128T1	10975051	10975051	A	G	transition
FL3001T1	10972679	10972679	C	T	transition
FL3003T1	10972846	10972846	G	A	transition
FL3003T1	10973107	10973107	G	A	transition
FL3008T1	10973365	10973365	G	A	transition
FL3009T1	10972247	10972247	G	A	transition
FL3009T1	10972780	10972780	C	T	transition
FL3009T1	10972933	10972933	G	A	transition
FL3009T1	10972974	10972974	G	A	transition
FL3009T1	10973437	10973437	G	C	transversion
FL3009T1	10974205	10974205	T	A	transversion
FL3010T1	10974928	10974928	C	T	transition
FL3010T1	10974997	10974997	A	G	transition
FL3013T1	10972192	10972192	C	T	transition
FL3013T1	10972193	10972193	C	T	transition
FL3013T1	10972846	10972846	G	A	transition
FL3013T1	10972871	10972871	C	T	transition
FL3015T1	10972571	10972571	C	T	transition
FL3016T1	10972770	10972770	G	T	transversion
FL3017T1	10972870	10972870	G	A	transition
FL3018T1	10972255	10972255	C	T	transition
FL3020T1	10972348	10972348	C	T	transition
FL3020T1	10972396	10972396	G	A	transition
FL3020T1	10972415	10972415	C	G	transversion
FL3020T1	10972419	10972419	G	A	transition
FL3020T1	10972679	10972679	C	T	transition
FL3020T1	10972748	10972748	C	A	transversion
FL3020T1	10972968	10972968	C	G	transversion
FL3020T1	10973123	10973123	G	A	transition
FL3020T1	10973264	10973265		del 2	deletion
FL3020T1	10973332	10973332	G	C	transversion
FL3020T1	10973368	10973373		del	deletion
FL3020T1	10973557	10973557	G	A	transition
FL3020T1	10973615	10973615	C	G	transversion
FL3020T1	10973797	10973797	C	T	transition
FL3020T1	10973869	10973869		del G	deletion
FL3020T1	10974854	10974854	G	A	transition
FL3102T1	10972635	10972635	G	A	transition
FL3106T1	10972125	10972125	G	A	transition
FL3106T1	10972247	10972247	G	A	transition
FL3106T1	10972763	10972763	G	A	transition
FL3110T1	10971492	10971492	C	T	transition

Case #	SNV or indel Position Start (hg19)	SNV or indel Position End (hg19)	ref	alt	type of genetic alteration
FL3111T1	10974538	10974538	C	T	transition
FL3114T1	10971279	10971279	C	G	transversion
FL3114T1	10971291	10971291	T	C	transition
FL3114T1	10971337	10971337	G	C	transversion
FL3114T1	10971398	10971398	T	C	transition
FL3114T1	10971659	10971659	G	C	transversion
FL3114T1	10971666	10971666	G	A	transition
FL3114T1	10971667	10971667	G	A	transition
FL3114T1	10971673	10971673	C	G	transversion
FL3114T1	10971697	10971697	A	C	transversion
FL3114T1	10971766	10971766	G	T	transversion
FL3114T1	10971815	10971815	A	C	transversion
FL3114T1	10971840	10971840	G	A	transition
FL3114T1	10971903	10971903	C	T	transition
FL3114T1	10971925	10971925	G	T	transversion
FL3114T1	10971972	10971972	A	G	transition
FL3114T1	10971979	10971979	C	T	transition
FL3114T1	10972019	10972019	C	G	transversion
FL3114T1	10972032	10972032	G	C	transversion
FL3114T1	10972055	10972055	C	G	transversion
FL3114T1	10972074	10972074	T	A	transversion
FL3114T1	10972142	10972142	C	G	transversion
FL3114T1	10972178	10972178	A	T	transversion
FL3114T1	10972266	10972266	C	G	transversion
FL3114T1	10972363	10972363	A	G	transition
FL3114T1	10972397	10972397	C	G	transversion
FL3114T1	10972462	10972462	C	T	transition
FL3114T1	10972505	10972505	C	G	transversion
FL3114T1	10972571	10972571	C	T	transition
FL3114T1	10972668	10972668	C	T	transition
FL3114T1	10972679	10972679	C	T	transition
FL3114T1	10972993	10972993	G	A	transition
FL3114T1	10973093	10973093	G	A	transition
FL3114T1	10973351	10973351	C	A	transversion
FL3114T1	10973433	10973433	A	G	transition
FL3114T1	10973951	10973951	C	G	transversion
FL3115T1	10975028	10975028	T	C	transition
FL3115T1	10975051	10975051	A	G	transition
FL3115T1	10975072	10975072	T	G	transversion
FL3116T1	10971669	10971669	G	A	transition
FL3116T1	10971672	10971672	G	A	transition
FL3116T1	10972305	10972305	C	T	transition
FL3116T1	10972521	10972521	G	A	transition
FL3116T1	10972522	10972522	C	T	transition
FL3116T1	10972968	10972968	C	T	transition

Case #	SNV or indel Position Start (hg19)	SNV or indel Position End (hg19)	ref	alt	type of genetic alteration
FL3116T1	10973315	10973315	C	T	transition
FL3116T1	10973690	10973690	G	C	transversion
FL3116T1	10974285	10974285	C	G	transversion
FL3117T1	10971751	10971751	G	C	transversion
FL3117T1	10972847	10972847	C	G	transversion
FL3117T1	10973103	10973103	C	T	transition
FL3117T1	10995609	10995609	G	C	transversion
FL3118T1	10973120	10973120	A	G	transition
FL3118T1	10973105	10973129		del	deletion
FL3120T1	10972716	10972716	C	G	transversion
FL3120T1	10972895	10972895	C	T	transition
FL3121T1	10972944	10972944	G	C	transversion
FL3123T1	10973589	10973589	C	T	transition
FL3124T1	10972424	10972424	C	G	transversion
FL3124T1	10972521	10972521	G	A	transition
FL3124T1	10972743	10972743	C	T	transition
FL3124T1	10972898	10972898	C	A	transversion
FL3124T1	10972967	10972967	G	C	transversion
FL3124T1	10973124	10973124	C	T	transition
FL3126T1	10971779	10971779	G	A	transition
FL3126T1	10972277	10972277	C	A	transversion
FL3126T1	10972461	10972461	C	A	transversion
FL3126T1	10973123	10973123	G	A	transition
FL3128T1	10972090	10972090	C	T	transition
FL3128T1	10972316	10972316	G	T	transversion
FL3128T1	10973372	10973372	G	A	transition
FL3132T1	10973320	10973320	C	G	transversion
FL3138T1	10971779	10971779	G	A	transition
FL3138T1	10972409	10972409	G	A	transition
FL3138T1	10972446	10972446	C	G	transversion
FL3138T1	10972682	10972682	C	G	transversion
FL3138T1	10972694	10972694	C	T	transition
FL3138T1	10972770	10972770	G	T	transversion
FL3138T1	10973675	10973675	G	A	transition
FL3138T1	10974179	10974179	C	T	transition
FL3138T1	10974620	10974620	G	A	transition
FL3139T1	10972456	10972456	G	A	transition
FL3140T1	10972032	10972032	G	A	transition
FL3140T1	10972709	10972709	G	A	transition
FL3140T1	10972847	10972847	C	G	transversion
FL3141T1	10972570	10972570	G	A	transition
FL3142T1	10972989	10972989	G	A	transition
FL3144T1	10973551	10973551	G	A	transition
FL3145T1	10972019	10972019	C	T	transition
FL3145T1	10972967	10972967	G	A	transition

Case #	SNV or indel Position Start (hg19)	SNV or indel Position End (hg19)	ref	alt	type of genetic alteration
FL3145T1	10973103	10973103	C	T	transition
FL3145T1	10973712	10973712	G	C	transversion
FL3145T1	10973717	10973717	G	A	transition
FL3145T1	10973718	10973718	G	A	transition
FL3147T1	10972494	10972494	C	T	transition
FL3147T1	10972818	10972818	G	A	transition
FL3147T1	10972915	10972915	C	G	transversion
FL3147T1	10973074	10973074	C	T	transition
FL3147T1	10973123	10973123	G	C	transversion
FL3147T1	10973414	10973414	C	G	transversion
FL3147T1	10974506	10974506	G	A	transition
FL3148T1	10972926	10972926	C	A	transversion
FL3148T1	10973061	10973061	C	G	transversion
FL3148T1	10974617	10974617	G	A	transition
FL3149T1	10972742	10972742	G	A	transition
FL3150T1	10971537	10971537	G	A	transition
FL3150T1	10971545	10971545	G	A	transition
FL3152T1	10972019	10972019	C	T	transition
FL3152T1	10972504	10972504	C	T	transition
FL3152T1	10973756	10973756	G	A	transition
FL3155T1	10971711	10971711	C	T	transition
FL3156T1	10972775	10972775	A	G	transition
FL3156T1	10972776	10972776	A	G	transition
FL3156T1	10972780	10972780	C	A	transversion
FL3156T1	10973061	10973061	C	G	transversion
FL3156T1	10973829	10973829	C	T	transition
FL3157T1	10972660	10972660	G	A	transition
FL3157T1	10972900	10972900	G	A	transition
FL3162T1	10972255	10972255	C	T	transition
FL3162T1	10972305	10972305	C	T	transition
FL3162T1	10972770	10972770	G	A	transition
FL3162T1	10972895	10972895	C	T	transition
FL3162T1	10974631	10974631	G	T	transversion
FL3163T1	10973257	10973257	C	T	transition
FL3163T1	10973315	10973315	C	T	transition
FL3166T1	10972415	10972415	C	T	transition

B.5 High confidence predictions for the chromosome 16 capture space

Library ID	SV type	chr1	position1	gene1	chr2	position2	gene2	max spanning reads	num events	interpretation of chr16 oligocapture results
A43029	INV	16	11201073	*CLEC16A*	16	11785708	*TXNDC11*	67	8	584 kb inversion, breakpoints in intron 18 of *CLEC16A* and exon 8 of *TXNDC11*, likely disruptive, functional impact unclear
A43030	DEL	16	10947070	*CIITA*	16	11116222	*CLEC16A*	38	4	169 kb deletion, breakpoints upstream of *CIITA* and in intron 11 of *CLEC16A,* complete loss of this *CIITA* allele
	INV	16	11348882	*SOCS1*	16	11351572	*SOCS1*	14	3	breakpoint in exon 2 of *SOCS1* and upstream of the *SOCS1* gene, disrupts *SOCS1*
	TRA	14	106326048	*IGH*	16	11349135	*SOCS1*	19	2	t(14;16), breakpoints in *IGH* intronic region and *SOCS1* exon 2 , strand direction different, therefore no fusion, disruption of *SOCS1*, effect on *IGH* locus unclear
A43031	INV	16	11013706	*CIITA*	16	11033187	*DEXI*	149	8	19 kb inversion, breakpoints in intron 16 of *CIITA* and intron 1 of *DEXI*, fuses *CIITA* exon 1-16 to *DEXI* exon 2, the latter is non-coding but a putative fusion protein can be derived (B.6)
	INV	16	11013709	*CIITA*	16	11033083	*DEXI*	110	7	reciprocal event, stop codon in exon 1 of *DEXI*, creates no fusion with *CIITA* exon 17-19

Library ID	SV type	chr1	position1	gene1	chr2	position2	gene2	max spanning reads	num events	interpretation of chr16 oligocapture results
A43036	INV	16	10972119	*CIITA*	16	11349103	*SOCS1*	144	6	377 kb inversion, breakpoints in intron 1 of *CIITA* and exon 2 of *SOCS1*, likely disruptive for both genes
	INV	16	10972127	*CIITA*	16	11349114	*SOCS1*	65	4	reciprocal event
	TRA	16	10972522	*CIITA*	X	41548791	*GPR34*	68	6	t(X;16), breakpoints in *CIITA* intron 1 and *GPR34* intron 1, putative fusion of *CIITA* exon 1 to *GPR34* exon 2, 51 aa protein predicted (B.7), likely non-functional
	TRA	16	10972530	*CIITA*	X	41548793	*GPR34*	70	2	reciprocal event, t(X;16), small ORFs, non-functional
	DEL	16	10972770	*CIITA*	16	10973118	*CIITA*	112	4	347 bp deletion *CIITA* intron 1
A43037	DEL	16	10973706	*CIITA*	16	10972948	*CIITA*	252	6	757 bp deletion intron 1 *CIITA*
	INV	16	11062836	*CLEC16A*	16	57168099	*CPNE2*	234	6	nested inversion, pericentric, breakpoints in exon 4 of *CLEC16A* and intron 12 of *CPNE2,* strand direction different, therefore no fusion
	INV	16	11063029	*CLEC16A*	16	30683452	*FBRS*	148	6	nested inversions
	INV	16	10537483	*ATF7IP2*	16	11080769	*CLEC16A*	36	6	nested inversions
	INV	16	11110612	*CLEC16A*	16	28448129	*EIF3C*	217	3	nested inversions
	INV	16	11059965	*CLEC16A*	16	21428484	*NPIPL3*	18	2	nested inversions
	DUP	16	11106289	*CLEC16A*	16	21748652	*OTOA*	10	2	duplication, functional impact unclear
A43043	DEL	16	10972316	*CIITA*	16	10972128	*CIITA*	94	4	187 bp deletion *CIITA* intron 1
A43049	INV	16	3056935	*CLDN6*	16	10966595	*CIITA*	21	8	8 Mb inversion, breakpoints upstream of *CLDN6* and upstream of *CIITA*, results in dislocation of the green BAC probe but no structural damage to *CIITA*

Library ID	SV type	chr1	position1	gene1	chr2	position2	gene2	max spanning reads	num events	interpretation of chr16 oligocapture results
A43050	DEL	16	10972395	*CIITA*	16	10973044	*CIITA*	108	6	648 bp deletion *CIITA* intron 1, FISH not explained but apparently small clone (5 %)
A43051	DEL	16	7638085	*RBFOX1*	16	10972040	*CIITA*	19	5	3.3 Mb deletion, breakpoints in intron 4 of *RBFOX1* and intron 1 of *CIITA*, leads to fusion of *RFBOX1* exon 4 to exon 2 of *CIITA* (B.8), resulting in a truncated protein
A43052	TRA	2	61108467	*REL*	16	10974031	*CIITA*	21	8	t(2;16), breakpoints in *CIITA* intron 1 and upstream of *REL*, fusion transcript possible (*CIITA* exon 1 to *REL* exon 2)
	TRA	2	61108477	*REL*	16	10974001	*CIITA*	13	7	reciprocal event
	TRA	2	89159665	*IGK*	16	10972714	*CIITA*	15	6	t(2;16), breakpoints in *CIITA* intron 1 and centromeric on chromosome 2, close to *IGK* gene region
	DEL	16	10973286	*CIITA*	16	10972769	*CIITA*	13	2	516 bp deletion *CIITA* intron 1
A43067	INV	16	10973601	*CIITA*	16	27326617	*IL4R*	39	6	16 Mb inversion, breakpoints in *CIITA* intron 1 and *IL4R* intron 1, same strand direction, therefore no fusion transcript upon inversion, likely disruptive
	INV	16	10973610	*CIITA*	16	27326640	*IL4R*	16	6	reciprocal event
A43068	DEL	16	10962704	*CIITA*	16	11310352	*CLEC16A*	54	8	348 kb deletion, deletes *CIITA*, *DEXI* and *CLEC16A* entirely
	TRA	14	106211708	*IGHG1*	16	11348887	*SOCS1*	28	3	t(14;16), breakpoints in I*GH* intron 1 and *SOCS1* exon 2, different chromosome arms, same strand direction, therefore no fusion, disrupts *SOCS1*, effect on *IGH* locus unclear
	TRA	14	106213380	*IGHG1*	16	11348899	*SOCS1*	106	3	reciprocal event

Library ID	SV type	chr1	position1	gene1	chr2	position2	gene2	max spanning reads	num events	interpretation of chr16 oligocapture results
A43069	TRA	16	10973113	*CIITA*	22	39854860	*MGAT3*	55	6	t(16;22), breakpoints in intron 1 of *CIITA* and intron 1 of *MGAT3,* different chromosome arms, same strand direction, therefore no fusion, likely disruptive
	TRA	16	10973113	*CIITA*	22	39854856	*MGAT3*	40	6	reciprocal event
A43070	DEL	16	10983031	*CIITA*	16	11812699	*TXNDC11*	41	8	830 kb deletion, breakpoints in *CIITA* intron 1 and *TXNDC11* intron 5, different strand directions, therefore no fusion transcript
	TRA	1	2985148	*PRDM16*	16	10972750	*CIITA*	102	5	t(1:16), breakpoints in *CIITA* exon 1 and upstream of *PRDM16*, results in fusion of *CIITA* exon 1 to *PRDM16* exon 2
	TRA	1	2984655	*PRDM16*	16	10972919	*CIITA*	56	8	reciprocal event translocation *CIITA* intron 1 and upstream of *PRDM16*, no promoter swap, disruption of *CIITA* allele
	DEL	16	11215236	*CLEC16A*	16	11480301	*RMI2*	26	3	265 kb deletion, breakpoints in intron 19 of *CLEC16A* and downstream of *RMI2*
	DEL	16	11215210	*CLEC16A*	16	11480299	*RMI2*	24	2	reciprocal event
	INV	16	11215206	*CLEC16A*	16	11480274	*RMI2*	19	2	265 kb inversion, breakpoints in intron 19 of *CLEC16A* and downstream of *RMI2*
A43071	DEL	16	10756488	*TEKT5*	16	11339356	*SOCS1*	21	7	417 kb deletion, deletes *CIITA* allele
A43072	DEL	16	10861924	*NUBP1*	16	10996431	*CIITA*	73	8	134 kb deletion, breakpoints in *NUBP1* intron 9 and *CIITA* intron 1, creates in-frame fusion transcript *NUBP1* exon 9 - *CIITA* exon 8

Library ID	SV type	chr1	position1	gene1	chr2	position2	gene2	max spanning reads	num events	interpretation of chr16 oligocapture results
A43075	DEL	16	11349240	*SOCS1*	16	11348417	*SOCS1*	90	7	800 bp deletion *SOCS1*
	TRA	16	8762984	*AICDA*	12	10973366	*CIITA*	48	6	t(12;16), breakpoints in *CIITA* intron1 and *AICDA* intron 1, same chromosome arms, different strand direction, therefore no fusion transcript
	TRA	12	8764607	*AICDA*	16	10973178	*CIITA*	133	5	reciprocal event
A43076	INV	16	11037301	*CLEC16A*	16	11352433	*RMI2*	40	7	315 kb inversion involves *CLEC16A*
	DEL	16	10982313	*CIITA*	16	12374408	*SNX29*	20	6	1.4 Mb deletion, breakpoints in *CIITA* intron 1 and *SNX29* intron 15, putative 29 aa fusion transcript (B.9)
	DUP	16	10972350	*CIITA*	16	10972662	*CIITA*	45	4	300 bp duplication intron 1 *CIITA*
A43077	DEL	16	10972806	*CIITA*	16	12062507	*TNFRSF17*	81	6	1 Mb deletion, breakpoints in *CIITA* intron 1 and downstream of *TNFRSF17*
A43078	TRA	8	128808741	*PVT1*	16	10972594	*CIITA*	90	4	t(8;16), breakpoints in *CIITA* intron 1 and *PVT1* intron 1, different chromosome arms, same strand direction, therefore no fusion transcript
	TRA	8	128808716	*PVT1*	16	10972569	*CIITA*	39	3	reciprocal event
	DEL	16	10972600	*CIITA*	16	10972919	*CIITA*	55	2	318 bp deletion *CIITA* intron 1

Library ID	SV type	chr1	position1	gene1	chr2	position2	gene2	max spanning reads	num events	interpretation of chr16 oligocapture results
A43079	TRA	10	46794179	*CTSL1P5*	16	10972823	*CIITA*	18	2	t(10;16)
	TRA	10	48989246	*GLUD1P7*	16	10972800	*CIITA*	146	7	t(10;16)
	DEL	16	10973892	*CIITA*	16	10973408	*CIITA*	189	6	483 bp deletion intron 1 *CIITA*
	TRA	14	106325853	*IGH*	16	11349095	*SOCS1*	29	4	t(14;16), breakpoints in *IGH* and *SOCS1* exon 2, different chromosome arms, same strand direction, therefore no fusion, disrupts *SOCS1,* effect on *IGH* locus unclear
A43080	DEL	16	11348838	*SOCS1*	16	10973122	*CIITA*	195	8	375 kb deletion, breakpoints in intron 1 *CIITA* and *SOCS1* exon 2
	DEL	16	10972446	*CIITA*	16	10971699	*CIITA*	106	8	752 bp deletion intron 1 *CIITA*
A43081	DEL	16	10973504	*CIITA*	16	10971906	*CIITA*	21	5	401 bp deletion intron 1 *CIITA*
	DEL	16	10972143	*CIITA*	16	10971940	*CIITA*	12	4	202 bp deletion intron 1 *CIITA*
	INV	16	10972733	*CIITA*	16	10973111	*CIITA*	17	3	367 bp inversion *CIITA* intron 1
	INV	16	10973005	*CIITA*	16	10972714	*CIITA*	18	3	reciprocal event
A43082	DEL	16	10972338	*CIITA*	16	10972492	*CIITA*	28	2	153 bp deletion intron 1 *CIITA*
	DEL	16	11349550	*SOCS1*	16	11349119	*S*	45	6	175 bp deletion intron 1 *CIITA*
	DEL	16	10972737	*CIITA*	16	10972561	*CIITA*	29	2	281 bp deletion intron 1 *CIITA*
	DEL	16	10972598	*CIITA*	16	10972316	*CIITA*	20	2	281 bp deletion intron 1 *CIITA*
A43084	DEL	16	10972598	*CIITA*	16	10972316	*CIITA*	97	2	752 bp deletion intron 1 *CIITA*
A43090	DEL	16	11180645	*CLEC16A*	16	11179301	*CLEC16A*	70	8	1.3 kb deletion *CLEC16A* intron 18
	DEL	16	11348970	*SOCS1*	16	11348503	*SOCS1*	16	4	~470 bp deletion *SOCS1* exon 2 - 3'UTR

Library ID	SV type	chr1	position1	gene1	chr2	position2	gene2	max spanning reads	num events	interpretation of chr16 oligocapture results
A43092	TRA	8	128749207	*MYC*	16	11349139	*SOCS1*	23	7	translocation intron 1 *MYC*, exon 2 *SOCS1*, likely no splice donor/acceptor, disruption of *SOCS1*
	TRA	8	128749148	*MYC*	16	11349126	*SOCS1*	13	4	reciprocal event
A43093	DEL	16	10975142	*CIITA*	16	10973558	*CIITA*	80	8	1.5 kb deletion *CIITA* intron 1
	INV	16	10971536	*CIITA*	16	10972727	*CIITA*	90	6	1.2 kb inversion *CIITA* intron 1
	INV	16	10971870	*CIITA*	16	10972644	*CIITA*	83	3	773 bp inversion *CIITA* intron 1
	DUP	16	10972945	*CIITA*	16	10973145	*CIITA*	35	2	199 bp duplication *CIITA* intron 1
A43094	DEL	16	11348834	*SOCS1*	16	9992019	*GRIN2A*	111	7	del 16p 1.3 Mb, intron 3 of *GRIN2A* to exon 2 of *SOCS1*
A43095	DEL	16	10974070	*CIITA*	16	10972127	*CIITA*	20	6	2 kb deletion *CIITA* intron 1
	DEL	16	11349373	*SOCS1*	16	11348681	*SOCS1*	85	5	296 bp deletion *CIITA* intron 1
	DEL	16	10972312	*CIITA*	16	10972015	*CIITA*	76	5	185 bp inversion *CIITA* intron 1
	INV	16	10973186	*CIITA*	16	10973372	*CIITA*	66	3	reciprocal event
	INV	16	10973169	*CIITA*	16	10973376	*CIITA*	37	3	262 bp deletion *CIITA* intron 1
	DEL	16	10974698	*CIITA*	16	10974425	*CIITA*	83	2	2 kb deletion *CIITA* intron 1
A43097	DEL	16	10971998	*CIITA*	16	10971586	*CIITA*	138	5	411 bp deletion *CIITA* intron 1
A43101	INV	16	10971827	*CIITA*	16	10973681	*CIITA*	52	6	~1.8 kb inversion *CIITA* intron 1
	INV	16	10971821	*CIITA*	16	10973687	*CIITA*	36	5	reciprocal event
A43110	DEL	16	11143548	*CLEC16A*	16	10810673	*NUBP1*	63	7	333 kb deletion of *NUBP1, CIITA, DEXI* and parts of *CLEC16A*

Library ID	SV type	chr1	position1	gene1	chr2	position2	gene2	max spanning reads	num events	interpretation of chr16 oligocapture results
A43115	DEL	16	10973158	*CIITA*	16	10972963	*CIITA*	141	5	194 bp deletion intron 1 *CIITA*
	INV	16	10971282	*CIITA*	16	10971688	*CIITA*	251	3	405 bp inversion in *CIITA* intron 1
	INV	16	10971301	*CIITA*	16	10971709	*CIITA*	236	3	reciprocal event
	DEL	16	11349494	*SOCS1*	16	11349402	*SOCS1*	25	3	
	TRA	7	128309229	*FAM71F2*	16	10980174	*CIITA*	15	2	t(7:16) breakpoints in *CIITA* intron 1 and upstream of *FAM71F2*, different chromosome arms, same reading direction, no fusion
A43117	DEL	16	10973536	*CIITA*	16	10972620	*CIITA*	349	6	915 bp deletion *CIITA* intron 1

Abbreviations: aa, amino acid; BAC, bacterial artificial chromosome; chr, chromosome; N, not validated; N*, not validated because of limited material or PCR failure; num, number; ORF, open reading frame; SV, structural variant; Val, validation; Y, validated; Y*, validated, but different mapping

B.6 Putative fusion transcript A43031

MRCLAPRPAG SYLSEPQGSS QCATMELGPL EGGYLELLNS DADPLCLYHF YDQMDLAGEE
EIELYSEPDT DTINCDQFSR LLCDMEGDEE TREAYANIAE LDQYVFQDSQ LEGLSKDIFK
HIGPDEVIGE SMEMPAEVGQ KSQKRPFPEE LPADLKHWKP AEPPTVVTGS LLVGPVSDCS
TLPCLPLPAL FNQEPASGQM RLEKTDQIPM PFSSSSLSCL NLPEGPIQFV PTISTLPHGL
WQISEAGTGV SSIFIYHGEV PQASQVPPPS GFTVHGLPTS PDRPGSTSPF APSATDLPSM
PEPALTSRAN MTEHKTSPTQ CPAAGEVSNK LPKWPEPVEQ FYRSLQDTYG AEPAGPDGIL
VEVDLVQARL ERSSSKSLER ELATPDWAER QLAQGGLAEV LLAAKEHRRP RETRVIAVLG
KAGQGKSYWA GAVSRAWACG RLPQYDFVFS VPCHCLNRPG DAYGLQDLLF SLGPQPLVAA
DEVFSHILKR PDRVLLILDG FEELEAQDGF LHSTCGPAPA EPCSLRGLLA GLFQKKLLRG
CTLLLTARPR GRLVQSLSKA DALFELSGFS MEQAQAYVMR YFESSGMTEH QDRALTLLRD
RPLLLSHSHS PTLCRAVCQL SEALLELGED AKLPSTLTGL YVGLLGRAAL DSPPGALAEL
AKLAWELGRR HQSTLQEDQF PSADVRTWAM AKGLVQHPPR AAESELAFPS FLLQCFLGAL
WLALSGEIKD KELPQYLALT PRKKRPYDNW LEGVPRFLAG LIFQPPARCL GALLGPSAAA
SVDRKQKVLA RYLKRLQPGT LRARQLLELL HCAHEAEEAG IWQHVVQELP GRLSFLGTRL
TPPDAHVLGK ALEAAGQDFS LDLRSTGICP SGLGSLVGLS CVTRFRAALS DTVALWESLQ
QHGETKLLQA AEEKFTIEPF KAKSLKDVED LGKLVQTQRT RSSSEDTAGE LPAVRDLKKL
EFALGPVSGP QAFPKLVRIL TAFSSLQHLD LDALSENKIG DEGVSQLSAT FPQLKSLETL
NLSQNNITDL GAYKLAEALP SLAASLLRLR PAPLKSFVPT ERFLFEPRRV EERLGLSAEV
GKAPAPTEGG TQEASQDLIC LQCQIDGLAL ASDYLQLRWM FTGTQPEFAS LHFIPERTCF
PHFTFGEDTS NCGHTQKRLP APFDVSSTOP

B.7 Putative fusion transcript A43036

MRCLAPRPAG SYLSEPQAVN AILKHALNRF SHYQVKLKTV KGCDYYQIGK ISTOP

B.8 Putative fusion transcript A43051

MLASQGVLLH PYGVPMIVPA APYLPGLIQG NQEAAAAPDT MAQPYASAQF APPQNGIPAE
YTAPHPHPAP EYTGQTTVPE HTLNLYPPAQ THSEQSPADT SAQTVSGTAT QTDDAAPTDG
QPQTQPSENT ENKSQPKRLH VSNIPFRFRD PDLRQMFGQF GKILDVEIIF NERGSKAAHS
VPPWSWGPSTOP

B.9 Putative fusion transcript A43076

MRCLAPRPAG SYLSEPQGGR DAWRADSTOP