Frequency and relevance of genetic alterations of the *ID3-TCF3-CCND3* pathway in pediatric mature B-cell Non-Hodgkin lymphoma

Inauguraldissertation
zur Erlangung des Grades eines Doktors der Medizin
des Fachbereichs Medizin
der Justus-Liebig-Universität Gießen

vorgelegt von Marius Rohde
aus Lengerich

Gießen 2013

Aus dem Zentrum für Kinderheilkunde und Jugendmedizin

Abteilung für Pädiatrische Hämatologie und Onkologie

am Fachbereich Medizin der Justus-Liebig-Universität Gießen

Leiter: Prof. Dr. med. A. Reiter

Gutachter: PD Dr. B. Burkhardt

Gutachter: Prof. Dr. M. Rummel

Tag der Disputation: 16.02.2015

Table of content

1 Introduction

The term ‘Non-Hodgkin lymphoma’ (NHL) is a still widely used, but historical collective name for malignant diseases of the lymphoid system, including any kind of lymphoma except Hodgkin lymphomas. There have been different approaches to classify this group of heterogeneous malignancies, like the *Kiel classification* (Stansfeld et al., 1988), the *Working Formulation of Non-Hodgkin’s Lymphoma for Clinical Usage* (Robb-Smith, 1982) and the current *WHO classification of Malignant Lymphoma* (Swerdlow et al., 2008). The WHO classification pursues an approach in which lymphomas are differentiated by the cell type from which they are originating from.

Of all malignancies listed in the annual report of the German Pediatric Cancer Registry, Non-Hodgkin Lymphoma (NHL) (excluding Burkitt leukemia) account for about 5-10% of all malignancies in children under the age of 15 in Germany (GPOH Jahresbericht, 2013). In children and adolescents, about two thirds of all NHL belong to mature B cell lymphomas originating from mature B lymphocytes. Within this heterogeneous subgroup Burkitt lymphoma (BL), including Burkitt leukemia (B-AL), and Diffuse large B-cell lymphoma (DLBCL) are the most common subtypes. (GPOH Jahresbericht, 2013)

1.1 The human Immune System

Functionally, the immune system can be divided in two parts: the innate immune system and the adaptive immune system. The innate immune system comprises generic, non-specific functions of immune response. This includes physical and chemical barriers, inflammatory processes and the complement system; those are immediately available when a pathogen is encountered. In contrast, the adaptive immune system needs more time to establish. It underlies continuous modulation and is trained to recognize specific parts of a pathogen (antigen). Therefore, the adaptive immune response is stronger and more effective. Individual benefits and complex interactions between these two systems ensure an overall powerful immune response at any time point.

Regardless of their later function, cells of the immune system derive from the same pool of pluripotent hematopoietic stem cells (HSC) in the bone marrow. Commitment of HSCs to either a myelopoietic progenitor, later differentiating into granulocytes, thrombocytes, monocytes or erythrocytes, or a lymphoid progenitor, giving rise to B lymphocytes, T lymphocytes and natural killer cells, is initiated by specific environmental signaling. (Murphy et al., 2008)

The effectors of the adaptive immune system can be further divided in cellular and humoral immunity-mediating components. Cellular immunity is mainly mediated by T lymphocytes, humoral immunity by antibodies and their producing cells, the B lymphocytes (Abbas et al., 2012). In mammals, B and T lymphocytes arise from the same lymphoid progenitor in the bone marrow (Abbas et al., 2012). Different signaling pathways lead the progenitor cell to migrate either to the thymus to become a T lymphocyte or to the generative organs (e.g. spleen, lymph nodes) to develop into a B lymphocyte. The latter are named after the Bursa of Fabricius, an organ in birds, which plays an essential role in B cell development in this species (Glick et al., 1955).

1.2 B lymphocytes

B lymphocytes provide their highly specific, humoral immune response by clonal expression of a diverse repertoire of antigen receptors, called B cell receptors (BCR). These BCRs are cell membrane-bound molecules which recognize and bind to specific antigens. If released in the blood from the lymphocytes, BCRs act independently as antibodies. Diversity of antigen receptors is achieved by somatic recombination, a process in which random compositions of genomic DNA segments of the BCR form a new unique antigen receptor coding sequence. The B lymphocyte maturation, starting from a lymphocyte progenitor in the bone marrow to a mature B cell migrating into the lymphoid tissue, is depended on the successful passing of certain developmental stages within a strictly regulated differentiation process (Figure 1).

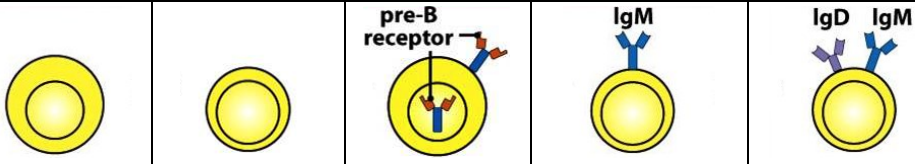
	HSC	pro-B cell	pre-B cell	immature B	mature B
					
H-chain genes	germline DNA	VDJ rearranging	VDJ rearranged	VDJ rearranged	VDJ rearranged
L-chain genes	germline DNA	germline DNA	VJ rearranging	VJ rearranged	VJ rearranged
Surface Ig	absent	absent	pre-BCR	IgM	IgD and IgM

Figure 1: Development of the B cell lineage

Different stages of B cell development are depicted with corresponding status of heavy-chain (H) and light-chain (L) gene rearrangement. In hematopoietic stem cells (HSC) the genetic constitution is equal to germline DNA. During pro-B cell development, H-chain genes located on chromosome 14 are the first to undergo individual rearrangement of variable (V), diversity (D) and joining (J) segments, the so called VDJ rearrangement. This is followed by rearrangement of V and J segments in L-chain genes on either chromosome 2 or 22 during the pre-B cell status. Individually rearranged H- and L-chain genes provide the broad diversity of Immunoglobulins (Ig). The first Igs presented on the cellular surface during the developmental stages of immature and mature B cells are pentameric IgM and monomeric IgD. Modified from Murphy et al. (2008)

1.2.1 Commitment to the B cell lineage

Commitment of common lymphoid progenitors to the B cell lineage depends on regular signaling of Interleukin 7 (IL-7) (Miller et al., 2002) in the bone marrow and expression of Transcription Factor 3 (TCF3) (O'Riordan and Grosschedl, 1999), Early B cell Factor 1 (EBF1) (O'Riordan and Grosschedl, 1999) and Paired Box 5 (PAX5) (Cobaleda et al., 2007). These transcription factors induce cell proliferation and differentiation to provide a large pool of progenitor cells for further maturation. At this time point, the genomic DNA of the progenitors is identical to the germline DNA of the individual (Figure 1) (Abbas et al., 2012).

1.2.2 Somatic recombination

The process of the somatic recombination of the BCR results in a wide repertoire of different antigen-recognizing lymphocytes based on a comprehensible amount of gene segments. On protein level BCRs show a consistent architecture, constituted of two identical heavy chains and two identical light chains. The amino acid sequence of every chain has a constant region, which determines the antigen receptors class, and a specific variable region, which is responsible for antigen binding. On the genomic level the chains are coded in the Immunoglobulin (Ig) gene loci. In human these loci are located on chromosome 14 for the heavy (H) chain and on chromosomes 2 and 22 for the two light (L) chains κ and λ,.

The H-chain locus consists of three different clusters of genes coding for the variable region: starting from the 5’ end with variable (V) segments, followed by diversity (D) segments and ending with joining (J) segments. Further downstream, constant (C) segments are coding for the constant region of the receptor. The κ and λ loci show the same structural pattern of V and J segments, but have no D segments. V(D)J recombination is the process of joining different segments (called V-D-J joining), to form a rearranged gene, coding for the variable region. The rearrangement is catalyzed by lymphoid-specific endonucleases named Recombination-activating gene 1 (Rag-1) and Recombination-activating gene 2 (Rag-2), which are part of the V(D)J recombinases complex. Furthermore, DNA repair enzymes randomly substitute, add and delete nucleotides at the ends of recombined segments. This process also intensifies diversity. Finally, the recombined ending sites are fused by ligating enzymes. (Abbas et al., 2012)

1.2.3 Differentiation stages of B cell maturation

According to different recombination levels of the Ig loci and the expression of the respective BCRs, five developmental stages can be distinguished during B cell maturation. HSCs first undergo VDJ rearrangement of the H-chain locus after commitment to the B cell lineage and are hereinafter called progenitor B cells (pro-B). In the next developmental stage cells start to express the pre-antigen receptor (pre-BCR) requiring successful H-chain locus rearrangement. Cells expressing the pre-BCR are called pre B cells (pre-B). At the same time pre-B cells qualify to undergo additional rearrangement of one of the two L-chains. After successful rearrangement of both Ig loci the so called immature B cell expresses the BCR, which is membrane-bound IgM (Figure 1). The immature B cell is able to leave the bone marrow and migrate into lymphoid tissues, like spleen or lymph nodes. (Abbas et al., 2012)

Self-antigen recognizing lymphocytes represent a risk in the context of autoimmunity. To avoid proliferation of such cells, the affinity to self-antigens of immature B cells is checked in the bone marrow and peripherally in lymphoid tissue. Immature B cells showing a strong recognition of self-antigens are forced to undergo receptor editing, to remove self-reactivity. Therefore, rearrangement of the light chain locus can be repeated or switched to the other remaining light chain allele. Remaining self-antigen recognizing lymphocytes are signaled into apoptosis. Finally, expression of IgM and IgD is upregulated and the B cell matures to respond to antigens with proliferation and differentiation (Figure 1). (Abbas et al., 2012)

1.3 Germinal center reaction

1.3.1 Antigen-activation of B lymphocytes

Before having actively encountered any antigens, B cells with rearranged BCR prior to antigen driven proliferation and further differentiation are termed ‘naïve’. Trigger mechanisms and sites of antigen-activation are diverse. For activation the mature B cell needs to physically meet the antigen directly (soluble antigens) or indirectly (by antigen-presenting cells, e.g. macrophages, dendritic cells, B cells) (Abbas et al., 2012). Both, naïve and antigen-activated B cells are attracted by the chemokine CXCL13 (Legler et al., 1998). CXCL13 is secreted by follicular dendritic cells in follicles and recognized by the B cell’s CXCR5 receptor (Förster et al., 1996). After migration into the follicles, activated B cells establish germinal centers (GC) by proliferation (Rose et al., 1980). Germinal centers are subdivided in a B cell proliferative ‘dark zone’ and high level immunoglobulin-expressing B cell ‘light zone’ (Murphy et al., 2008). Those are

surrounded by resting B cells, forming the ‘mantle zone’ (Murphy et al., 2008). Within the germinal center reaction, antigen-activated B cells undergo the developmental processes of affinity maturation and class switching, which are described in the following.

1.3.2 Affinity maturation

Affinity maturation is a process of enhancing the antigen-binding qualities of the BCR and subsequently produced antibodies. The enzyme Activation-induced cytidine deaminase (AID) is only expressed by activated germinal center B cells (GCB) (Muramatsu et al., 1999). Its expression is upregulated upon CD40-ligand activation on the B cells surface (Fuleihan et al., 1993), which takes place in the germinal center reaction. AID induces random mutations in the variable region by deamination of cytosine bases (Abbas et al., 2012). The residue, uracil, can be substituted by any other base in consequence of DNA repair mechanisms. This process, called somatic hypermutation (SHM), originates GCBs with a diverse repertoire of slightly differing variable regions. GCBs with improved antigen-recognition are selected for proliferation and further differentiation.

1.3.3 Class switching

Class switching recombination (CSR) allows switching between different classes of expressed antibodies (e.g. IgG, IgE, IgA), each having specific effector functions. Class switching depends on recombination of the C_H locus, in which different class-defining C segments are joined next to the previously formed V(D)J gene (Abbas et al., 2012). Again, this is achieved by cleavage of intervening DNA. AID is thought to generate double strand breaks in actively transcribed promoter regions (S regions) upstream of the primary and the designated C segment (Manis et al., 2002). This process forms DNA loops which locate the selected C segment next to the V(D)J region. Finally, double-strand repair mechanisms ligate the switched regions (Murphy et al., 2008).

After antigen-activation, affinity maturation and class switching, the GCBs differentiate into either antibody-producing plasma cells or resting memory cells. (Abbas et al., 2012)

1.3.4 Malignant transformation

The process in which a cell acquires genomic alterations that result in the phenotype of cancer is called malignant transformation. During the different genome-editing events in B cell development, lymphocytes are at risk to gain structural and punctual mutations or aberrations. In B lymphocytes, genomic instabilities might be fostered by either

accidental occasion, malfunction of DNA repair mechanisms or uncontrolled activity of DNA remodeling proteins like AID or Rag-1/-2.

Lymphoid malignancies present with different characteristics according to cellular origin and time point of malignant transformation. For example, acute precursor B cell leukemia (pB-ALL) cells derive form immature progenitor cells, while BL and DLBCL derive from mature germinal center B cells (Figure 2). (Kuppers, 2005)

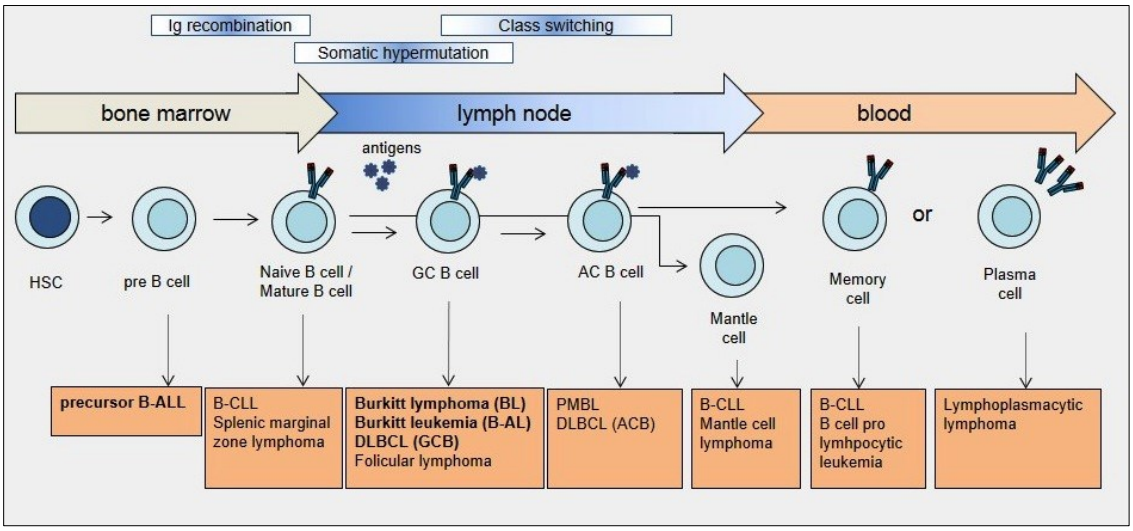


Figure 2: Cellular origin of B cell malignancies

Developmental stages of B cells are depicted starting from hematopoietic stem cells (HSC) in the bone marrow and ending with specialized memory or plasma cells in the blood. Progress of genome-editing events is shown by Ig recombination, Somatic hypermutation and Class switching in a timely manner. Depending on cellular origin and timepoint of malignant transformation arising lymphoid malignancies are listed. Precursor B-ALL, BL, B-AL and DLBCL are highlighted due to their role in the current study. HSC: Hematopoietic Stem Cell, GCB: Germinal Centre B cell, ACB: Activated B cell, B-CLL: B-cell Chronic Lymphoblastic Leukemia, DLBCL: Diffuse large B-cell lymphoma, PMLBL: Primary Mediastinal Large B-cell Lymphoma, Ig: Immunoglobulin. Modified from Kuppers (2005).

1.4 Mature B cell lymphoma

1.4.1 NHL-BFM study group

In Germany, first multicenter trials for patients with pediatric leukemia and lymphoma were established in the 1970s when Hansjörg Riehm (Berlin), Bernhard Kornhuber (Frankfurt) and Günther Schellong (Münster) founded the Berlin-Frankfurt-Münster (BFM) study group. Over the years more than 60 pediatric oncology departments in Germany joined the group. Remarkable increases in patient's survival were achieved by continuous advancement of treatment approaches. In addition the validation of sufficient risk-stratification criteria, the reduction of acute and long-term toxicity and improved diagnostic criteria have been and still are essential objectives for the development of improved treatment strategies. (Rossig et al., 2013)

Today, more than 95% of all pediatric patients diagnosed with mature B cell lymphoma in Germany are registered to the clinical trials of the NHL-BFM group (GPOH Jahresbericht, 2013) and treated according to standardized treatment plans. Risk-stratification is based on the resection status, the stage of disease and the pre-treatment serum LDH level. Staging is assessed according to St Jude's staging system (Murphy, 1980). Bone marrow puncture/biopsy and lumbar puncture are obligatory to evaluate bone marrow and central nervous system (CNS) status. Cases with more than 25% blasts in the bone marrow are called Burkitt leukemia (B-AL). According to NHL-BFM protocols, B cell lymphoma/leukemia patients are stratified into four risk groups (R1 to R4). Treatment intensity is stratified according to the risk group with two to six five-day courses of polychemotherapy. The courses are based on dexamethasone, methotrexate, cyclophosphamide, ifosfamide, cytarabine, etoposid, doxorubicin, vincristine, vindesine and inthrathecal triple therapy. (NHL-BFM Study Group, 2012)

1.4.2 Burkitt Lymphoma

Burkitt lymphoma (BL) is the most common NHL subtype in pediatric patients (Burkhardt et al., 2005). It was first described by Denis Burkitt as a "[…] sarcoma involving the jaws in african children" in 1958 (Burkitt, 1958). Today, the etiology of BL is divided into an Epstein-Bar-Virus (EBV) related endemic group, which is mainly present in tropical Africa, a group related to infections with the Human immunodeficiency virus (HIV) and a sporadic group of cases outside of Africa. The cases investigated in this study belong to the latter sporadic group. In sporadic cases the median age of pediatric BL patients (including the diagnosis of Burkitt leukemia) at diagnosis is 8.4 years and male-female ratio is 4,5:1 (Burkhardt et al., 2005).

Typical Burkitt cells are medium-sized and show a high proliferative rate. Nucleoli present with clumped and dispersed chromatin, the cytoplasm is basophilic and often contains vacuoles. (Swerdlow et al., 2008)

Malignant cells in pleural effusions, ascites and bone marrow typically show L3 morphology according to the French-American-British classification (Bennett et al., 1976). In patients with bone marrow infiltration of more than 25% Burkitt leukemia (B-AL) is diagnosed. In this study the term BL always includes the leukemic variant, unless otherwise stated.

BL cells express IgM on their surface and are usually positive for CD19, CD20, CD22, CD10 and BCL6. BCL2 and TdT are usually negative. (Bennett et al., 1976)

Translocation of *c-MYC* to an immunoglobulin gene locus is virtually present in all cases. The most common translocation t(8;14)(q24;q32) localizes *c-MYC* under the control of the Ig H-chain promoter (80%), whereas translocations to the λ t(8,22)(q24;q11) or κ t(2;8)(p11;q24) light chain loci account for 10% of the cases each (Boxer and Dang, 2001). As Ig loci promoters are highly activated during B cell development, *c-MYC* translocation-carrying B cells are forced to constitutive expression of c-MYC, which is involved in cell cycle regulation, cell growth and cellular metabolism, but also triggers apoptosis (Dang, 1999; Dang et al., 1999; Sanchez-Beato et al., 2003). It was shown in mice that *c-MYC* translocations alone do not necessarily result in malignant transformation (Nepal et al., 2008; Roschke et al., 1997). Furthermore, *c-MYC* translocations were also found in lymphocytes from healthy human individuals at a small number (Muller et al., 1995). This leads to the hypothesis that *c-MYC* positive GCBs require additional molecular alterations, serving as 'second-hit'. To investigate how *c-MYC* positive B cell lymphomas avoid apoptosis, oncogenes like *TP53*, *RB1*, *CDKN2A*, *BAX*, *TP73*, *BCL6* and/or their pathways have been studied and were shown to be altered in BL at small numbers (Bhatia et al., 1992; Capello et al., 1997; Capello et al., 2000; Cinti et al., 2000; Corn et al., 1999; Farrell et al., 1991; Gutierrez et al., 1999; Kalungi et al., 2011; Martinez-Delgado et al., 2002; Wilda et al., 2004; Wiman et al., 1991).

High initial lactate dehydrogenase (LDH) serum levels, as well as more disseminated disease, particularly in form of CNS positive lymphomas and bone marrow (BM) involvement, were shown to be poor prognostic factors (Reiter et al., 1999). However, currently applied combination of intensive chemotherapy regimens reach up to a total of 90% probability of event free survival (pEFS) (Burkhardt et al., 2005; Woessmann et al., 2005). Cases of relapse and progression remain challenging, as they often respond

insufficiently to therapy and present with a significantly worse prognosis (Woessmann, 2013).

Disruption of the ARF–MDM-2–p53 apoptotic pathway was shown for 55% of 24 pediatric BL cases by Wilda and colleagues (Wilda et al., 2004). Preudhomme et al. studied 48 BL/B-AL cases for *TP53* mutations, but there was no significant association with tumor mass or prognosis (Preudhomme et al., 1995). Nelson et al. studied a larger series of pediatric BL/B-AL patients registered on Children's Cancer Group Study CCG-5961 and found significant inferior survival for patients with 13q deletions detected by FISH (Nelson et al., 2010), but no further understanding of the molecular mechanisms of these findings is available so far.

1.4.3 Diffuse large B-cell lymphoma

Diffuse large B-cell lymphoma (DLBCL) is the most common type of B-NHL in adults (The Non-Hodgkin's Lymphoma Classification Project, 1997) and the second most common subtype in children. DLBCL presents as a more heterogeneous disease on morphological, molecular and immunophenotypical aspects and can therefore be divided into several subgroups (Swerdlow et al., 2008). However, the centroblastic (CB) variant is the most frequent in adults as well as in pediatric DLBCL patients. The median age at the time point of diagnosis in pediatric DLBCL-CB cases is 11.4 years and the male to female ratio 1.7:1 (Burkhardt et al., 2005).

All DLBCL are usually positive for B cell markers CD19, CD20, CD22 and CD79a, whereas the expression of BCL6 varies. The cells of the most common centroblastic variant presents with round, medium-sized, vesicular nuclei and amphophilic to basophilic cytoplasma. (Swerdlow et al., 2008)

In contrast to pediatric BL, pediatric DLBCL less often show *c-MYC* translocations; Rates are about 31-38% are reported (Deffenbacher et al., 2012; Gualco et al., 2009; Lu et al., 2011). Gene expression profiling studies established criteria to differentiate two molecular signatures within DLBCL. The classification of activated B cell-like (ACB) and germinal-center B cell-like (GCB) subgroups is especially important in adults, where the latter is associated with a more favorable prognosis. (Rosenwald et al., 2002)

In pediatric DLBCL about 75% of the cases account for the GCB subgroup (Deffenbacher et al., 2012; Miles et al., 2008; Oschlies et al., 2006). Compared to adults, pediatric patients carry *IRF4* translocations (Salaverria et al., 2011) more often and the GCB subtype is more frequent (Klapper et al., 2012).

Like in pediatric BL the prognosis of DLBCL is influenced by clinical characteristics, including stage of disease, LDH serum levels, CNS and BM involvement status. In an integrated analysis of patients enrolled in the NHL-BFM trials 86,90 or 95 probability of event free survival (pEFS) in pediatric DLBCL was 93±2%. However, there was a strong impact on age and gender with respect to the outcome, as the group of adolescent girls had a significantly higher chance to suffer from relapse (pEFS 0.50±16%) (Burkhardt et al., 2011; Burkhardt et al., 2005).

1.4.4 Unclassifiable B-NHL

In some cases discrimination between BL and DLBCL remains unclear, due to ambiguous histologic features. According to the current WHO classification these cases are classified as “B-cell lymphoma, unclassifiable, with features intermediate between diffuse large B-cell lymphoma and Burkitt lymphoma” (Swerdlow et al., 2008). Most common is a presentation of tumors with histomorphological features of BL, but e.g. too low expression of proliferative markers or BCL2 positivity and partial CD10 expression (de Jong, 2009). These cases account for a small number in pediatric B-NHL. In this study the abbreviation ‘B-NHL (DD)’ is used for such cases.

1.5 Recurrently mutated genes in B-NHL

In a recently published study BL tumor samples of four pediatric patients diagnosed and treated according to the NHL-BFM protocols have been analyzed in the ICGC-MMML-seq project (International Cancer Genome Consortium, Molecular Mechanisms in Malignant Lymphoma by Sequencing). Whole-genome, whole-exome and whole-transcriptome sequencing was performed on tumor and control samples to look for somatic alterations that might be involved in malignant transformation and lymphomagenesis. The list of genes that occurred to be affected in at least two out of four analyzed cases included *DDX3X*, *FBXO11*, *MYC*, *RHOA*, *SMARCA4*, *TP53* and *ID3*. (Richter et al., 2012)

Nearly contemporaneous two further NGS studies also reported *ID3* and furthermore *TCF3* and *CCND3* among their lists of candidates being recurrently mutated especially in BL. Furthermore, they provided evidence for functional linkage between ID3, TCF3 and CCND3 on protein level (Love et al., 2012; Schmitz et al., 2012). Hence, these genes represented interesting candidates for further analysis.

1.6 Candidate genes *ID3*, *TCF3* and *CCND3*

The gene Inhibitor of DNA 3 (*ID3*) is one of four helix-loop-helix (HLH) protein coding genes belonging to the Inhibitor of DNA (ID) family. The *ID* genes lack a basic DNA binding site, thus preventing other HLH transcription factors from activating their target genes by heterodimerization (Benezra et al., 1990; Benezra et al., 1990; Murre, 2005). *ID3* is located on the long arm of chromosome 1 and consists of three exons with a coding region (isoform NM_002167.4) of 360 base pairs, resulting in a protein consisting of 120 amino acids (NCBI Gene database, 2013). *ID3* is expressed at constant levels in immature and mature B cells (Becker-Herman et al., 2002; Xu et al., 2007).

During B cell maturation one important role of ID3 is the interaction with Transcription Factor 3 (TCF3). *TCF3* is located on chromosome 19, where alternative splicing results in two isoforms, E12 and E47, the latter showing a higher functional activity in B cell specific target activation (NCBI Gene database, 2013; Sigvardsson, 2000). TCF3 regulates numerous target genes by binding to the respective promoter sites via its HLH domain (reviewed in Murre (2005)). The induction of Rag enzymes for somatic recombination and AID for somatic hypermutation were shown to be TCF3 dependent (Hsu et al., 2003). Sayegh and colleagues also performed retroviral overexpression of ID3 in activated B lymphocytes in mice and showed reduction of AID protein levels, whereas ectopic *TCF3* expression induced AID expression (Sayegh et al., 2003). These findings led to the conclusion that ID3 inhibits TCF3 dependent gene expression in B lymphocytes (Figure 3).

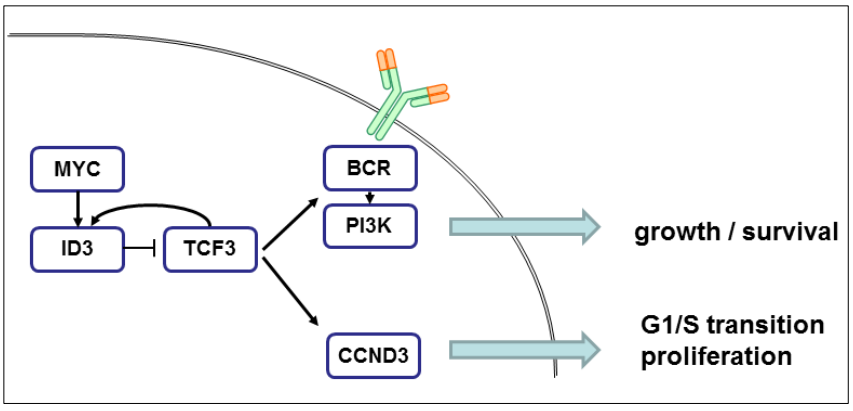


Figure 3: ID3, TCF3, CCND3 pathway

ID3, TCF3 and CCND3 are depicted according to their functional interplay. TCF3 is a potent transcription factor that activates cell-cycle regulating downstream targets like PI3K and CCND3. ID3 is a direct target of MYC and controls TCF3 activity by direct negative feedback regulation. BCR: B-cell receptor. Modified from Campo (2012) and Schmitz et al. (2012).

Moreover TCF3 is an important regulator of BCR signaling and stimulation of downstream targets in the BCR pathway (Verma-Gaur et al., 2012). Activation of the phosphoinositide-3-kinase (PI3K) pathway can be augmented by TCF3 overexpression and promotes cellular growth and survival (Schmitz et al., 2012). Furthermore, TCF3 increases cell proliferation via direct promoter-activation of *CCND3* (Cato et al., 2011; Schmitz et al., 2012; Song et al., 2004). *CCND3* encodes for the G1/S-specific cyclin-D3 protein and is localized on chromosome 6. Cyclin D3 is a subunit that forms activating complexes with cyclin dependant kinases (CDK4/6). CDK4/6 interact with the cell cycle inhibitors p21 and p27 (Sawai et al., 2012; Sherr and Roberts, 2004) causing inactivation of retino blastoma protein (Rb1) by phosphorylation. This subsequently leads to cell cycle progression by G1/S phase transition (Sherr and Roberts, 2004). Cato et al. demonstrated that cyclin D3 expression is required for proliferative expansion of GCB cells (Cato et al., 2011).

Mutations in *ID3*, *TCF3* and *CCND3*

Currently there are three studies available describing mutations of the *ID3* gene in BL. Frequency of *ID3* mutations varied between 34% (Love et al.), 58% (Schmitz et al.) and 68% (Richter et al.) in the analyzed cohorts. Schmitz et al. reported *TCF3* mutations in 11% of the cases and furthermore *CCND3* mutations in 38% of the cases. *CCND3* mutations were also analyzed in the study of Richter et al., also reporting 38% of the cases to display these aberrations. (Love et al., 2012; Richter et al., 2012; Schmitz et al., 2012)

All of these studies described *ID3* mutations to accumulate in the HLH domain and first functional analyses showed ID3 mutant proteins to be less or completely ineffective in inhibiting TCF3, thus forcing increased cell proliferation and survival via PI3K and Cyclin D3 (Love et al., 2012; Richter et al., 2012; Schmitz et al., 2012). While *TCF3* mutations also affected the bHLH domain of its isoform *E47*, TCF3 mutant proteins did not lose their function on downstream targets when compared to wildtype TCF3, but partially reduced ID3/TCF3 interaction and turn them immune to the inhibitory effect of ID3 (Schmitz et al., 2012). CCND3 mutant proteins showed an increase in cell cycle stimulation when compared to unaffected CCND3, thereby indicating a gain of function (Schmitz et al., 2012). All in all, mutations in each of the candidate genes are thought to contribute to cellular growth, cell survival and proliferation (Campo, 2012).

Relevance of *ID3*, *TCF3* and *CCND3* in other malignancies

Besides the novel finding of *ID3*, *TCF3* and *CCND3* mutations in BL by the above mentioned studies, no data on recurrent mutations of these candidate genes is available

for other malignancies. Sequencing of *ID1* was performed in 13 families with a history of malignant melanoma and *ID3* was sequenced in a series of 94 ovarian tumors, but no mutations were found in either case (Arnold et al., 2001; Casula et al., 2003; Perk et al., 2005). Another study investigated the role of genetic alterations in *ID3* by sequencing of sorted T-cell DNA samples from 209 patients suffering from Sjögren's syndrome. The study was set up on the evidence of disease-associating factors in an *ID3*-deficient mouse model, but again no mutations were reported (Sellam et al., 2008).

However, expression levels of the ID proteins have widely been investigated in a vast number of tumour entities, like prostate cancer (Asirvatham et al., 2007; Patel and Chaudhary, 2012; Sharma et al., 2012; Strong et al., 2013), lung cancer (Castanon et al., 2013; Kamalian et al., 2010; Kamalian et al., 2008; Li et al., 2012), medulloblastoma (Phi et al., 2013; Snyder et al., 2013), ovarian cancer (Arnold et al., 2001; Maw et al., 2009) and breast cancer (Gupta et al., 2007; Mern et al., 2010; Noetzel et al., 2008; Wazir et al., 2013). Also clinical relevance was demonstrated in two recent studies, one associating high ID1/ ID3 expression levels with more advanced grade of prostate cancer (Sharma et al., 2012) and the other reporting correlation between ID1 and ID3 co-expression and inferior clinical outcome in non-small cell lung cancer (Castanon et al., 2013). Influence on cell migration was lately reported by Shuno and colleagues, who performed double-knockdown of ID1 and ID3 in pancreatic cancer cell lines and discovered decreased metastatic potential in migration assays (Shuno et al., 2010). Aptamer-based targeting of Id1/3 induced cell-cycle arrest and apoptosis in ovarian and breast cancer cells provided the basis for possible new disease modifying drugs (Mern et al., 2010; Mern et al., 2010), targeting this pathway.

TFC3 is known to form fusion transcripts with *PBX1* in acute lymphoblastic leukemia. Differences in overall survival of *TCF3-PBX1*-negative and –positive cases are discussed in controversy (Burmeister et al., 2010). Also, the role of TCF3 as a tumor promoter in gastric cancer was described previously (Patel and Chaudhary, 2012).

Immunohistochemical expression of CCND3 was investigated in Hodgkin’s lymphoma and showed higher levels in older patients, but didn’t show any correlation with respect to patients outcome (Marnerides et al., 2011).

In sum, ID proteins are functionally involved in a variety of highly aggressive tumor entities, with, in some studies, evidence for prognostic relevance and influence on tumor cell aggressiveness. In the functional context also *TCF3* and *CCND3* have previously shown to be involved in other malignancies.

1.7 Study objectives

Recently published NGS studies on B-NHL noted *ID3* as a recurrently altered gene in Burkitt Lymphoma. Also, *TCF3* and *CCND3* were shown to be affected at a lower frequency, being of special interest regarding their functions on cell cycle regulation in B lymphocytes and *ID3* interaction. Furthermore, previous studies on *ID3* gene expression levels support a clinical relevance in some malignancies.

However, published data on *ID3* mutations in Burkitt lymphoma showed variable mutation rates in the analyzed cohorts, leaving an uncertain picture on its incidence. Also the clinical and prognostic relevance of alterations in the three candidate genes (*ID3, TCF3* and *CCND3)* is unknown.

The aims of this study were

- to analyze somatic variants in the updated set of thirteen NHL-BFM patients analyzed within the ICGC-MMML-seq project
- to assess the frequency of alterations in the genes *ID3*, *TCF3* (isoform E47, exon 16) and *CCND3* (exon 5) in a representative, uniformly diagnosed and treated cohort of pediatric B-NHL patients.
- to investigate mutational patterns and distribution of mutations in the analyzed pathway.
- to analyze the findings regarding clinical and prognostic relevance.
- to investigate the occurrence of *ID3* mutations in cases of pediatric pB-ALL as a control group of a related malignancy.

2 Material and methods

In the following, the materials used in this study are summarized in separate tables for equipment (Table 1), kits (Table 2) and reagents (Table 3).

2.1 Materials

2.1.1 Equipment

Table 1: Equipment	
3130xl Genetic Analyzer	Applied Biosystems
Consort E835, Power supply (Electrophoresis chamber)	Consort NV, Turnhout, Netherlands
DOC-PRINT HOOD - DP-CF-011.C, imaging and documentation system	Vilber Lourmat Deutschland GmbH, Eberhardzell, Germany
DRI-BLOCK DB-2A	Bibby Scientific Limited, Staffordshire, United Kingdom
ED Heating Immersion Circulator	Julabo, Seelbach, Germany
Electrophoresis chamber	von Keutz, Reiskirchen, Germany
GeneAmp PCR System 9600	Perkin Elmer, Waltham, United States
Hettich Mikro 220R Centrifuge	Hettich AG, Bäch, Suisse
Incubator	Memmert, Schwabach, Germany
Incubator shaker	Braun, Melsungen, Germany
NanoDrop™ 1000	PEQLAB Biotechnologie GmbH, Erlangen, Germany
Navigator N28110, Scale	Ohaus Corporation, Parsippany, United States
Small centrifuge, Galaxy Mini	VWR, Darmstadt, Germany
T-Gradient Thermoblock	Biometra, Göttingen, Germany
T-personal Combi	Biometra, Göttingen, Germany
UV-Transilluminator	Ultra Violet Products, Cambridge, United Kingdom
Video Graphic Printer UP897MD	Sony Europe Limited, Surrey, United Kingdom
Vortex Mixer 7-2020	NeoLab, Mannheim, Germany

2.1.2 Kits

Table 2: Kits	
Big Dye Terminator V 3.1 Cycle Sequencing Kit	Life Technologies GmbH, Darmstadt, Germany
E.Z.N.A.® Plasmid Mini Kit I	VWR International GmbH, Darmstadt, Germany
High Pure PCR Template Preparation Kit	Roche Diagnostics GmbH, Mannheim, Germany
High Pure Plasmid Isolation Kit	Roche Diagnostics GmbH, Mannheim, Germany
Illustra™GFX™ PCR DNA and Gel Band Purification Kit	GE Healthcare, Munich, Germany
TOPO® TA Cloning® Kit for subcloning with TOP10 E. coli	Life Technologies GmbH, Darmstadt, Germany

2.1.3 Reagents and Enzymes

Table 3: Reagents and Enzymes	
100% Ethanol	100% Ethanol, Riedel de Haen, Sigma, Seelze, Germany
310 and 31xx Running Buffer, 10X	Life Technologies GmbH, Darmstadt, Germany
Agarose	PeqGOLD Universal Agarose (Peqlab, Erlangen, Germany)
dATP	Roche, Mannheim, Germany
dNTP	Roche, Mannheim, Germany
Fermentas 6X DNA Loading Dye	Fermentas GmbH, St. Leon-Roth, Germany
GeneRuler™ DNA Ladder Mix, ready-to-use	Fermentas GmbH, St. Leon-Roth, Germany
H2O	LiChrosolv® Water for chromatography, Merck KGaA, Darmstadt, Germany
Hi-Di™ Formamide	Life Technologies GmbH, Darmstadt, Germany
Isopropanol	2-Propanol, Riedel de Haen, Sigma, Seelze, Germany

Table 3: Reagents and Enzymes	
Lysogeny Broth (LB) media	Trypton/Pepton from Casein (Roth, Karlsruhe, Germany) Bacto-Yeast-Extract (Becton Dickinson, Heidelberg, Germany) NaCl (Roth, Karlsruhe, Germany)
Polymerase	OneTaq Polymerase 2x MM with Standard Buffer (New England BioLabs GmbH, Frankfurt am Main, Germany)
Polymerase	OneTaq Polymerase 2x MM with GC-Buffer (New England BioLabs GmbH, Frankfurt am Main, Germany)
Polymerase	Taq DNA Polymerase, Recombinant (Invitrogen Life Technologies GmbH, Darmstadt, Germany)
Polymerase	AccuPrime™ Pfx DNA Polymerase (Invitrogen Life Technologies GmbH, Darmstadt, Germany)
POP-7™ Polymer	Life Technologies GmbH, Darmstadt, Germany
Primers (see Table 7)	
Saltfree water	
Sodium Acetate	3M Sodium acetate, Roth, Karlsruhe, Germany
Tris-acetate-EDTA Buffer	50x TAE Tris/Acetic Acid/EDTA Buffer (Bio-Rad Laboratories GmbH, Munich, Germany)

2.2 Methods

2.2.1 Recurrent mutations in 13 pediatric BL in the ICGC-MMML-Seq project

Whole-genome sequencing of pediatric BL tumor and corresponding germline samples was performed within the ICGC-MMML-Seq project (Coordinator: Prof. M. Siebert, Kiel). Sequencing was performed centrally in Kiel (Prof. Dr. P. Rosenstiel). For cases of the NHL-BFM group, the study was approved by the Ethical Advisory Board Gießen (A89/11). As of June 2013 thirteen pediatric BL cases from the NHL-BFM group were analyzed after informed consent of patients and/or guardians (approved by the Ethical Advisory Board of the Faculty of Medicine in Gießen, AZ89/11).

Whole-genome sequencing raw data were analyzed with respect to single nucleotide variants (SNVs) and small insertions/deletions (InDels) by Dr. M. Schlesner (DKFZ, Heidelberg) as previously published in Richter et al. (Richter et al., 2012). Results were available via the internal project database and were furthermore deposited online in the ICGC Data Portal (ICGC Project, 2013).

2.2.2 Validation of *ID3*, *TCF3* and *CCND3* mutations

The validation of candidate genes from the ICGC-MMML-Seq project for patients enrolled in the NHL-BFM protocols was approved by the Ethical Advisory Board of the Faculty of Medicine in Gießen (AZ89/11 – Amendment 06/13).

2.2.3 Patient samples

2.2.3.1 B-NHL samples

The NHL-BFM group collects samples of B-NHL patients in the NHL-BFM laboratory at the Justus-Liebig University Gießen. The inclusion criteria in this study were as follows:

- registered into either the trial NHL-BFM 95 or B-NHL BFM 04 after given informed consent
- date of diagnosis between 01/2000 and 12/2011
- patients diagnosed with “Burkitt lymphoma” or “Burkitt leukemia” or “Diffuse large B-cell lymphoma” or “B-NHL (DD)”
- protocol patient status (no relevant treatment modifications)
- not included in the ICGC-MMML-Seq project

A total of 1127 patients were eligible according to inclusion criteria. Next, the availability of appropriate lymphoma material in the NHL-BFM biobank was reviewed according to the following quality criteria:

- availability of initial frozen tumor cells (tissue, ascites, pleural effusion or bone-marrow only in cases of B-AL)
- for tissue: histological review confirming the diagnosis or cytomorphological review of corresponding tumor touch imprints with the diagnosis of FAB L3 blasts
- for blasts isolated from bone marrow, ascites or pleural effusions: a cytomorphological review confirming at least 75% of FAB L3 blasts.

Reports of reference histology, carried out by the trial's reference pathologists, were available at the NHL-BFM study center.

Reports of reference cytomorphology for tumor touch imprints, effusion cytospins and bone marrow smears were available from the NHL-BFM reference laboratory for Cytomorphology and Genetics, Justus-Liebig University, Gießen.

Reports of FISH examinations were available from the Oncogenetic Laboratory in the Department of Pediatric Hematology and Oncology, Justus-Liebig University, Gießen or from other associated reference centers of the respective trial.

Eighty-four representative cases were randomly selected. In a second step another 10 c-*MYC* rearrangement positive B-NHL cases were chosen in order to enrich the cohort for cases with relapse. Before preparation, samples were stored at -80 °C.

2.2.3.2 pB-ALL samples

For the second part of this study, tumor DNA samples from 96 pediatric pB-ALL patients were kindly provided from the ALL-BFM study center in Kiel (Prof. Dr. M. Schrappe, Dr. A. Möricke). All patients fulfilled the inclusion criteria of the respective ALL-BFM trial, in which they have been registered including reference diagnosis and informed consent.

2.2.4 DNA extraction from B-NHL samples

The High Pure PCR Template Preparation Kit was used to extract high-molecular weight genomic DNA from tumor cells according to the protocols shown in Table 4.

The kit is based on adsorption chromatography, which allows nucleic acids to bind to a glass fiber-matrix, whereas other components do not bind or can be removed by washing and centrifugation.

First, cells and extracellular matrix have to be lysed to release the nucleic acids. Proteins, like DNA cutting nucleases, are digested by the highly proteolytic enzyme Proteinase K. The suspension, containing the nucleic acids, is transferred onto a glass fiber-matrix provided in spin columns. Remaining cellular and extracellular components are removed by washing and centrifugation steps. Finally, the nucleic acids can be released from the glass fiber-matrix by low salt elution.

Table 4: DNA extraction with High Pure PCR Template Preparation Kit

Protocol for tumor cells, malignant effusions (pleural, ascites) and bone marrow

1. Samples were thawed at room temperature and up to 200 µl of sample were transfered to a sterile 1.5 ml microcentrifuge tube.
2. 200 µl Binding buffer and 40 µl Proteinase K solution were added. The solution was vortexed and immediately incubated for 10-30 minutes at 70 °C.
3. After incubation 100 µl Isopropanol was added; the solution was mixed.
4. The solution was transferred to a Spin column which was placed into a sterile collection tube.
5. Centrifugation was carried out for 1 minute at 8000 x g. Spin column was transferred into a new sterile collection tube.
6. 500 µl Inhibitor Removal Buffer were added to the spin column.
7. Centrifugation was carried out for 1 minute at 8000 x g. Spin column was transferred into a new sterile collection tube.
8. First washing: 500 µl Wash Buffer were added to the spin column.
9. Centrifugation was carried out for 1 minute at 8000 x g. Spin column was transferred into a new sterile collection tube.
10. Second washing: 500 µl Wash Buffer were added to the spin column.
11. Centrifugation was carried out for 1 minute at 8000 x g. Spin column was transferred into a new sterile collection tube.
12. Drying: Centrifugation was carried out for 10 seconds at 14000 x g. Spin column was transferred into a sterile 1.5 ml microcentrifuge tube.
13. 200 µl prewarmed (70 °C) Elution Buffer were added to the spin column and incubated for 5 minutes.
14. Centrifugation was carried out for 1 minute at 8000 x g.
15. The eluate was stored at 4 °C until further processing.

Protocol for tumor tissue

1. Samples were thawed at room temperature and 25 mg tissue was taken and sliced into small pieces using a sterile scalpel.

Continue Table 4

2. The sample, 200 µl Tissue Lysis Buffer and 40 µl Proteinase K solution were transferred to a sterile 1.5 ml micocentrifuge tube. The solution was vortexed and incubated at 55 °C overnight.
3. 200 µl Binding Buffer were added and incubated at 70 °C for 10 minutes.
4. Remaining cellular components were drawn out with a pipette tip.
5. Protocol for tumor cells, malignant effusions and bone marrow was followed, starting at step 4.

2.2.5 Polymerase chain reaction

The polymerase chain reaction (PCR) is a widely used method to amplify pieces of DNA. Perquisites for a PCR application are: template DNA, a thermostable DNA-Polymerase, 2`-deoxynucleoside 5`-triphosphates (dNTPs) and two different kinds of oligonucleotides. These specifically bind to unique sequences in the genome thus defining the region to be amplified.

To allow binding of primers DNA has to be single stranded and therefore the double stranded DNA is melted at 94 °C (denaturation). This step is followed by binding of the primers (annealing). The optimal temperature for primer binding depends on the melting temperature of the primer itself and the salt conditions within the approach. Therefore, there are different temperature conditions for each primer pair (Table 7). In a third step, the temperature is raised to 68 °C which is the optimum for the used DNA-polymerase for adding the complementary bases starting at the primer sites (extension). The steps of denaturation, annealing and extension are repeated several times to exponentially amplify the target DNA.

The PCR reaction protocols are shown in Table 5. The PCR program is depicted in Table 6. Primer sequences including annealing temperatures are given in Table 7.

Table 5: Polymerase chain reaction

PCR reaction for direct sequencing	
genomic DNA	1 µl (5 ng/µl)
Forward Primer	1 µl (10 pmol/µl)
Reverse Primer	1 µl (10 pmol/µl)
Polymerase*	12,5 µl
H_2O	10,5 µl
Total reaction volume	25 µl

* OneTaq Hot Start Polymerase 2x MM with GC Buffer was used for *CCND3* amplification. Otherwise OneTaq Hot Start Polymerase 2x MM with Standard Buffer was used.

PCR reaction for subcloning	
AccuPrime™ Pfx DNA Polymerase	1 µl
genomic DNA	1 µl (5 ng/ µl)
dNTPs	1,5 µl
Reaction Buffer, 10x	5 µl
H_2O	40 µl
Forward Primer	0,75 µl (10 pmol/ µl)
Reverse Primer	0,75 µl (10 pmol/ µl)
Total reaction volume	50 µl

Table 6: PCR program

1x	10 minutes 94 °C		
35x	denaturation	annealing	elongation
	45 seconds 94 °C	30 seconds, see Table 7	1 minute 68 °C
1x	10 minutes 68 °C		
	pause at 4 °C		

Table 7: Primer sequences

gene	primer name	forward primer (5'->3') reverse primer (5'->3')	PCR product length (bp)	annealing temperature
ID3	*ID3*_1*	GAACCAGTGTTGGGCTAAAG GCAAGTCACTTGTCCCTCTC	1034	56 °C
ID3	*ID3*_2*	AGGAGCTTTTGCCACTGA CAAAATTCGGCCATGTGCG	1113	56 °C
ID3	*ID3*_K[#]	TCCAGGCAGGCTCTATAAGTG CCGAGTGAGTGGCAATTTTT	694	56 °C
TCF3	*TCF3*[$]	GAAATACAGGAGGCCACACG AGGTGTGTGAGGTGTGGATG	609	68.5 °C
CCND3	*CCND3*[$]	CCATGTGTTGGGAGCTGTC CTGGAGGCAGGGAGGTG	328	57 °C

* self designed, [#] as published in Richter et al. (Richter et al., 2012), [$] as published in Schmitz et al. (Schmitz et al., 2012)

2.2.6 Purification of PCR products

Products of polymerase chain reaction contain residual amounts of dNTPs, primers, salts and genomic DNA. To obtain sufficient quality for the following sequencing reaction, these contaminants have been removed with the Illustra™ GFX™ PCR DNA and Gel Band Purification Kit. The protocol used is depicted in Table 8.

Table 8: Purification of PCR products using the Illustra™ GFX™ PCR DNA and Gel Band Purification Kit

Protocol for purification of PCR products from enzymatic reaction

1. 500 µl Capture buffer type 3 was pipetted onto a MicroSpin column, placed into a collection tube.
2. The complete PCR reaction volume was pipetted to and thoroughly mixed with the Caputer buffer by gently pipetting up and down.
3. Centrifugation was performed for 30 seconds at 16000 x g. The MicroSpin column was transferred into a new sterile collection tube.
4. 500 µl Washing buffer type 1 were added to the MicroSpin column.
5. Centrifugation was performed for 30 seconds at 16000 x g. The MicroSpin column was transferred into a new sterile collection tube.
6. 30 µl Eluation buffer type 6 (sterile, nuclease free water) were added.

Continue Table 8

7. The eluate was stored at 4 °C until further processing.

Protocol for purification of PCR products in gel bands

1. The PCR reaction was subjected to agarose gel electrophoresis. DNA fragments were visualized using UV light and the part of agarose gel containing the amplificate of interest was cut out using a scalpel.
2. The piece of agarose was transferred into a sterile 1.5 ml microcentrifuge tube.
3. 10 µl Capture buffer type 3 were added for each 10 mg of agarose weight. At least 300 µl Capture buffer type 3 were added.
4. The solution was mixed by inversion and incubation was performed at 60 °C for 15-30 minutes, until the agarose was completely melted.
5. Protocol for purification of PCR products from enzymatic reaction was followed, starting at step 3.

2.2.7 Bacterial culture, subcloning and plasmid DNA preparation

Bacterial subcloning allows the separation of a single DNA fragments. This is an important tool for the analysis of e.g. different alleles from a diploid organism. In the context of heterozygous (e.g. frameshift) mutations, analysis of sequencing results is convenient, as there is just one sequence read. In contrast, direct sequencing of PCR ampflificates may produce douple peak signals. Furthermore, after subcloning the allelic state can be identified in cases where more than one mutation is present in an analyzed sequence.

The separation of a single DNA fragment is realized by its integration into a small piece of linearized double strand DNA plasmid, called vector. The ends of the linearized, here used TOPO TA plasmid carry 3' thymidine overhangs. This allows the insertion of one DNA molecule, to which a 3' adenine has been added previously. DNA Topoisomerase I conjugated to the plasmid ligates the vector and the template molecule, resulting in a circular double stranded plasmid. Due to its molecular structure, the plasmid is replicated independently when brought into a convenient host organism like *Escherichia coli*. Usually, each bacterium takes up only one plasmid and therefore, selective picking of plate-grown colonies allows separation of different allele representing plasmids by chance.

Moreover, the plasmid contains an ampicillin resistance cassette and the lacZ gene encoding for beta-galactosidase enzyme. These selective markers are used during clone

selection: Ampicillin containing media will kill non plasmid carrying bacteria, as they are nonresistant to the antibiotic. Bacteria, transformed with a plasmid not containing an insert, will express beta-galactosidase and therefore, turn blue in the presence of 5-bromo-4-chloro-indolyl-β-D-galactopyranoside (X-gal). As the beta-galactosidase encoding gene on the plasmid is disrupted by a successful insertion, these bacteria form white colonies and are easily selected. Bacteria are proliferated in liquid cultures and plasmid DNA can be extracted with spin column kits. The inserted sequence can be analyzed using vector specific sequencing primers (M13).

In the current study cloning was performed using the TOPO® TA Cloning® Kit for subcloning with TOP10 *E. coli*. PCR products for cloning were generated using high fidelity AccuPrime™ Pfx DNA Polymerase (see PCR reaction for cloning, Table 5). To enrich for high quality PCR templates, reaction products were separated by gel electrophoresis (see Table 12). Gel band purification was performed as previously described in Table 8. After DNA concentration measurement, purified PCR products were adenylated following the protocol shown in Table 9. Cloning protocol is shown in Table 10 and DNA isolation from Plasmids in Table 11. Further cultivation of picked colonies was performed, when colony PCR (see PCR reaction for direct sequencing, Table 5) was positive with insert specific primers.

Table 9: Protocol for adenylation of PCR products

Reaction mixture

purified PCR product	8 µl
10x PCR Puffer (BD)	1 µl
dATP	0,5 µl
Taq DNA polymerase, Recombinant	0,5 µl

Protocol

1. Mixing of the reaction mixture
2. Incubation at 72 °C for 10 minutes in a thermocycler.

Table 10: Protocol for ligation and cloning

Protocol

1. Preparation of the ligation reaction mix (1 µl PCR product, 1 µl Salt solution, 1 µl vector, ad 6 µl H_2O), incubation at room temperature for 5 minutes
2. Thaw *E.coli* TOPO 10 cells on ice

Continue Table 10

3. Careful addition of 2,5 µl of ligation mix to the *E.coli* TOP 10 cells
4. Incubation on ice for 10 minutes
5. 30 seconds heat shock at 42 °C prewarmed water bath
6. Set back the reaction on ice
7. Addition of 250 µl SOC-media
8. Incubation with constant shaking at 37 °C for one hour
9. Preparation of LB-plates (10 g Trypton/Pepton from Casein, 5 g Bacto-Yeast-Extract, 10 g NaCl, H_2O ad 1000ml, adjustment of pH 7.0, autoclaved) containing 50µg/ml ampicillin, greasing of 40 µl X-Gal
10. Prewarming of the LB-plates for 30 minutes
11. Greasing of 100 µl of transformed cells per plate
12. Incubation at 37 °C of plates overnight.
13. Selection of white colonies by picking
14. Preparation of colony PCR.
15. In case of successful colony PCR: inoculation of 5 ml of LB medium containing ampicillin (50 µg/ml), incubation at 37 °C shaking overnight, followed by plasmid DNA preparation according to the following protocol (Table 11).

Table 11: Plasmid DNA preparation with High Pure Plasmid Isolation Kit

Protocol

1. The Binding Buffer was placed on ice.
2. Centrifugation of 5 ml previously incubated plasmid carrying *E.coli*-LB media was performed at 6000 x g for 30 seconds. The supernatant was discarded.
3. The bacterial pellet was resuspended with 250 µl Suspension Buffer (RNase). Addition of 250 µl Lysis Buffer. The solution was mixed by inversion.
4. The solution was incubated at room temperature for 5 minutes.
5. Addition of 350µl Binding Buffer. Incubation on ice for 5 minutes.
6. Centrifugation was performed at 13000 x g for 10 minutes.
7. A High Pure Filter Tube was placed into one collection tube. The entire supernatant was transferred to the Filter Tube. Centrifugation was performed at 14000 x g for 1 minute.

Continue Table 11

8. The Filter Tube was placed in a sterile 1.5ml microcentrifuge tube. 700µl Wash Buffer II were added, followed by centrifugation at 14000 x g for 60 seconds.
9. The Filter Tube was transferred to a new sterile 1.5 ml microcentrifuge tube. 100µl Elution Buffer were added, followed by centrifugation at 14000 x g for 60 seconds.
10. The eluate was stored at 4 °C until further preparation.

2.2.8 Agarose gel electrophoresis

For analysis and separation of DNA fragments agarose based gel electrophoresis was used. All gels were prepared using 1% Agarose (w/v) in 1xTAE buffer. For analytical purposes wells were loaded with previously prepared mixtures of 2 µl Loading Buffer and 3 µl of PCR product. Electrophoresis was performed as stated in Table 12. DNA fragments move within the applied voltage according to their size, small fragments move faster than larger ones. Gels were stained in Ethidium bromide solution (10mg/l) for 15 minutes, followed by washing in ddH_2O for 10 minutes. Gels were photographed and analyzed on an UV-transluminator (wavelength 254 nm).

To enrich for high quality PCR templates for subsequent cloning, the total PCR reaction volume of 50 µl was divided in 2x 25 µl. Each was mixed with 5 µl Loading Buffer and loaded to the wells of the gel. Electrophoresis and stain of gels was performed as described described above and in Table 12. Gels were stained in Ethidium bromide solution (10mg/l) for 15 minutes, followed by washing in ddH20 for 10 minutes. Bands were visualized using an UV-transluminator (wavelength 320 nm) and bands of correct size were cut from the gel using a sterile scalpel. DNA extraction from the gel bands was performed as described in Table 8.

Table 12: Protocol for agarose gel electrophoresis

Preparation of agarose gel solutions	
Agarose (w/v) in 1xTAE	1%
Electrophoresis parameters	
Voltage	180 V
Duration	45-75 minutes, depending on gel size

2.2.9 UV spectrometry

After isolation of genomic DNA, purification of PCR products from PCR reactions or gel bands and preparation of plasmid DNA, concentration and quality of the nucleic acids was measured by UV spectrometry using a NanoDrop™ 1000 (PEQLAB Biotechnologie GmbH, Erlangen, Germany). In each case 2 µl of undiluted sample was used for the measurement.

UV spectrometry allows measuring concentration and quality of nucleic acids by using a spectrophotometric instrument. The absorption maximum of DNA is at 260 nm and of RNA at 230 nm; the absorption of a test sample at these wavelengths correlates with the amount of these nucleic acids in the sample. The optical density (OD) of one unit corresponds to a concentration of 50 ng/dl of double stranded DNA (Sambrook and Russel, 2001). To control for protein contamination the absorption at 280 nm was measured. Sufficient purity for subsequent analysis was assumed if the ratio of OD_{260} to OD_{280} was between 1.8 and 2.0.

2.2.10 Cycle sequencing

Cycle sequencing is widely used to analyze the sequence of DNA templates. This method is based on “DNA sequencing with chain-terminating inhibtors” by Sanger et. al published in 1977 (Sanger et al., 1977). In comparison to a normal PCR dideoxynucleotides are added in addition to a taq DNA-polymerase, the deoxynucleotide-mix and one oligonucleotide, serving as a sequencing primer.

First the DNA template is melted at 94 °C (denaturation) to be accessible for the sequencing primer to bind to the single stranded DNA (annealing). Annealing is carried out at 50 °C. Subsequently the temperature is raised again to the optimum of the taq DNA-polymerase to start synthesis along the sequencing primer site. Whenever a dideoxynucleotide is integrated into the new DNA strand further elongation of that particular DNA molecule is disrupted, because it lacks the obligatory hydroxyl group. As the incorporation of dideoxynucleotides and common deoxynucleotides occurs randomly, DNA molecules of different lengths are synthesized. Each of the four different dideoxynucleotides is labeled with a different dye, which can be separately detected by a capillary electrophoresis instrument.

Big Dye Terminator V 3.1 Cycle Sequencing Kit was used to perform cycle sequencing according to the protocol shown in Table 13. Cycle sequencing was performed on a GeneAmp PCR System 9600 (Perkin Elmer, Waltham, USA) running the program shown in Table 14. Generally, each gene was sequenced in forward and reverse direction at

least once per patient. For validation issues either the forward or reverse primer was used.

Table 13: Cycle sequencing with Big Dye Terminator V 3.1 Cycle Sequencing Kit	
Template DNA	100 ng
Ready Reaction Mix	2 µl
Sequencing Buffer (5x)	4 µl
Primer (Forward or Reverse)	1 µl (1 pmol/ µl)
H_2O	up to 20 µl reaction volume
Total reaction volume	20 µl

Table 14: Cycle sequencing program			
1x	1 minute 96 °C		
25x	denaturation	annealing	elongation
	10 seconds 96 °C	5 seconds 50 °C	4 minutes 60 °C
	pause at 4 °C		

2.2.11 Ethanol precipitation

To remove residual amounts of dNTPs, 2`-dideoxynucleoside 5`-triphosphates (ddNTPs) and primers after cycle sequencing, ethanol based precipitation was carried out. The protocol used is shown in Table 15.

Table 15: Protocol for ethanol based precipitation

1. The product of the sequencing reaction (volume 20 µl) is transferred to a sterile 1.5 ml microcentrifuge tube.
2. Addition of 50 µl 100% Ethanol and 2 µl sodium acetate, mixing.
3. Centrifugation at 4 °C for 20 minutes at 14000 x g.
4. Suction of the supernatant with a water-jet pump
5. Addition of 150 µl 70% Ethanol, mixing.
6. Centrifugation at 4 °C for 20 minutes at 14000 x g.
7. Suction of the supernatant with a water-jet pump
8. Drying on a heating block at 37 °C for 10-20 minutes.
9. Storage of the pellet at 4 °C until further processing.

2.2.12 Capillary electrophoresis

Automated analysis of sequences is performed by a capillary electrophoresis instrument and corresponding computer software. The instrument basically consists of a capillary array, two buffer reservoirs and one polymer reservoir, a heating block, an optical detection system, a pump block and a high voltage power supply (Figure 4). A computer is connected via an interface to save and analyze assigned data.

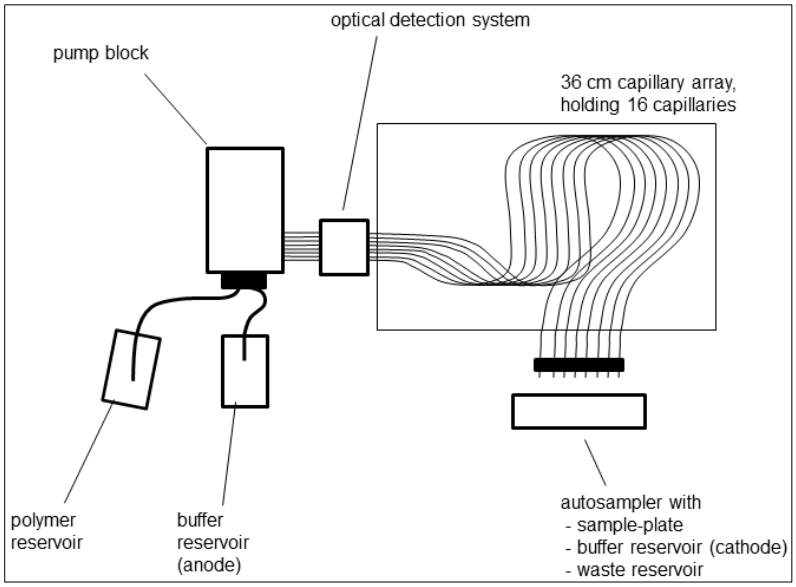


Figure 4: Charting of a capillary sequencer
Essential components of a capillary sequencer contain an autosampler, capillary array, optical detection system, pump block, buffer reservoir and polymer reservoir.

First the pump fills the capillary array with the polymer. In the polymer the cycle sequencing reaction products are separated by length, analog to the separation of DNA fragments in the agarose gel electrophoresis. This is followed by activation of the heating block to warm the polymer. Voltage is applied between the two buffer reservoirs, one therefore acting as anode and the other one being the cathode. The products of cycle sequencing reaction are injected into the capillary array. As DNA fragments are negatively charged, they run towards the positively charged anode. Smaller DNA fragments run faster so the very first fluorescently dye-labeled dideoxynucleotides arriving within the optical detection sytem represent the first nucleotides of the DNA template. The longer the fragment the later it will pass the detection system. Within the detection system a laser beam activates the dyes to fluoresce and the intensity is measured by a CCD-camera. These raw data are forwarded to a computer system which analyzes the digital data and visualizes those as electropherograms (Figure 5).

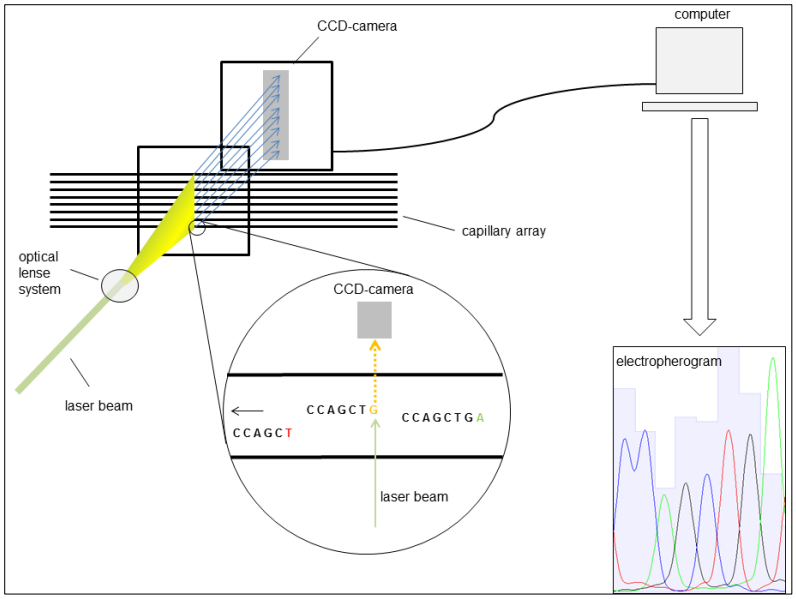


Figure 5: Method of capillary sequencing
In the optical system a laser beam makes dye-labeled nucleotides fluorescent which are floating through the capillary array towards the positively charged anode. A CCD-camera device absorbs light-energy and forwards raw data to a computer. Analysis software allows graphic representation of electropherograms.

A 3130xl Genetic Analyzer (Applied Biosystems) was used to perform capillary electrophoresis in this work. The protocol for sample preparation is shown in Table 16. The settings and conditions of the capillary electrophoresis instrument are shown in Table 17.

Table 16: Sample preparation for capillary electrophoresis

Reagents

Hi-Di™ Formamide

Protocol

1. Addition of 20 µl Hi-Di™ Formamide to the 1.5 ml microcentrifuge tube containing the dried DNA pellet after ethanol precipitation
2. Mix carefully, by pipetting up and down.
3. Transfer of the solved DNA – HiDi mixture into a designated well onto a 96-well plate
4. Transfer of the 96-well plate into the capillary electrophoresis instrument
5. Running of the instrument according to the manual.

Table 17: Capillary electrophoresis instrument settings	
Run Module	
Oven Temperature	60 °C
Poly Fill Vol	6500 steps
Current Stability	5.0 µA
PreRun Voltage	15.0 kV
PreRun Time	180 s
Injection Voltage	1.2 kV
Injection Time	18 s
Voltage Number Of Steps	30
Voltage Step Interval	15 s
Data Delay Time	120 s
Run Voltage	8.5 kV
Run Time	2780 s
Analysis Protocol	3130POP7_BDTv3-KB-Denovo_v5.2 (supplied by manufacturer)
Software	Data Collection Software 3.1.0
File Output Format	AB1

2.2.13 Analysis of electropherograms

The software Geneious (Geneious Software, 2012) was used to load electropherograms and align sequences to the reference sequences (Table 18) downloaded from NCBI website. Previously published short nuclear polymorphisms (SNPs) were downloaded from dbSNP (Build ID: 137, identifiers are also listed in appendix 10.4) (NCBI dbSNP database, 2013) and annotated to the reference sequences. SNPs were excluded from further analysis. Functional domains were annotated according to the data deposited in the Universal Protein Resource (UniProt database, 2013).

Table 18: Reference sequences		
Gene	**Accession number**	**UniProt entry**
ID3	NM_002167.4	Q02535
TCF3 (Exon 16)	NM_001136139.2	P15923-2 (E47)
CCND3 (Exon 5)	NM_001760.3	P30281

All sequence-disagreements within the alignments were reviewed manually. To rule out artificial mistakes or sequencing errors, samples showing mismatches in the alignment were validated separately. This validation was carried out by the criteria shown in Table 19.

Table 19: Validation criteria for samples showing sequence mismatches

Primary finding	**Validation requirement**
frameshift mutation	at least 1x / by cloning
single nucleotide substitution	at least 1x
wildtype	no validation
Cloning	at least 2 out of 10 clones had to show the same allele disambiguates in sequencing

2.2.14 Data analysis and interpretation of sequencing results

Disambiguate results confirmed within a validation experiment were considered to be genomic variants. Variants were described using the reference sequence of the corresponding coding DNA available from the NCBI database. Potential change on protein level was determined by translation using a RNA codon table. The nomenclature provided by the Human Genome Variety Society (den Dunnen and Antonarakis, 2000) (Version 2.0) was applied.

2.2.15 Definition and terminology of mutations

In this study the term ‘mutation’ was used when the genomic variant predicted an amino acid change on protein level. For description of silent changes or changes that occurred in introns, the term ‘genomic variant’ was used.

Mutations were termed heterozygous or homozygous with respect to presentation of peak heights in electropherogram analysis: disambiguates with clearly prominent non-reference peaks were described as ‘homozygous’. However, combination of focal loss of one ID3 allele and occurrence of mutation on the remaining allele must be taken into account in such cases. Regarding the relevance on protein level homozygous presenting cases were considered to feature ‘biallelic involvement’. In the following, five examples of certain mutation types are shown.

Example 1 (case 83):

The *ID3* sequence cutout (Figure 6) shows a nucleotide change from cytosine to thymine that occurred at position c.190. In the middle of the sequence a double peak is visible. The height of the cytosine peak (blue) almost equals the height of the alternate thymine peak (green). This change was considered to be a heterozygous substitution.

Result: c.190C>T

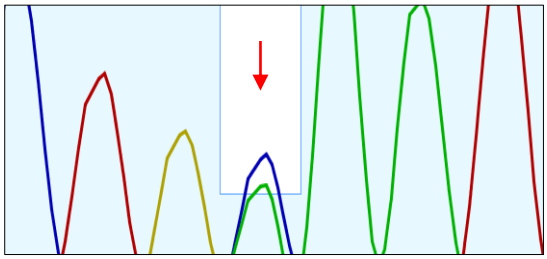


Figure 6: Example 1 – Heterozygous substitution in case 83.

Example 2 (case 55): Figure 7 depicts a sequence cutout with a change at position c.144 resulting in a nucleotide substitution from cytosine to guanine. The cytosine reference peak (blue) was considerably lower compared to the alternate guanine peak (yellow). This change was rated to be a homozygous substitution.

Result: c.[144C>G][(144C>G)]

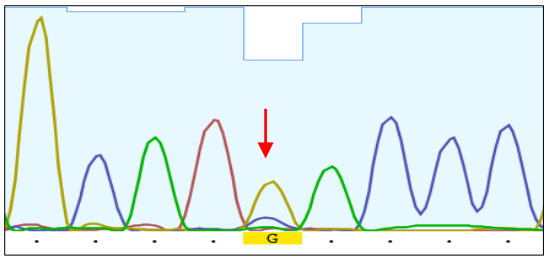


Figure 7: Example 2 – Homozygous substitution in case 55.

Example 3 (case 15): *ID3* forward (above) and reverse (below) sequences are shown as an excerpt (Figure 8) with double peaks starting at position c.120. Double peaks appeared continuous until the end of sequence. Manual read out of the alternative sequence detected a guanine deletion at position c.120, with subsequent shift of the reading frame. This change was rated heterozygous, due to a peak-to-peak ratio of averagely 1:1.

Result: c.120delG

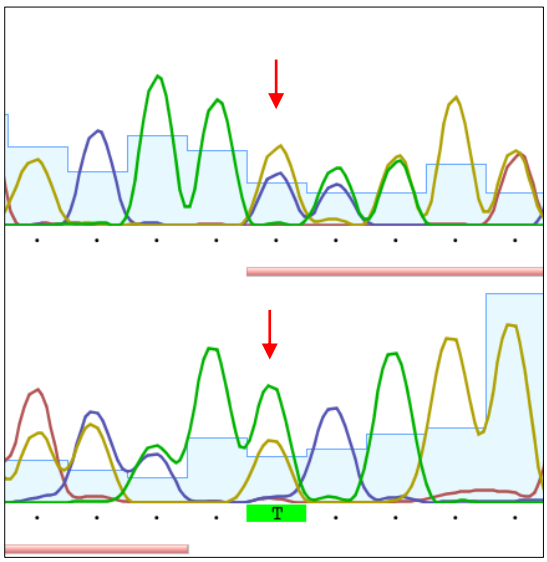


Figure 8: Example 3 – Heterozygous deletion in case 15.

Example 4 (case 91): The *ID3* sequence cutout (Figure 9) presents with a yellow (guanine) extra peak at position c.94, which is not matching with the reference sequence (not shown). As the previous peak is also reporting for a guanine nucleotide at this position, this change is a single nucleotide duplication of position c.93. As shown by the clear ongoing sequence, this change seemed to affect both alleles and was therefore rated to be homozygous.

Results: c.93dupG

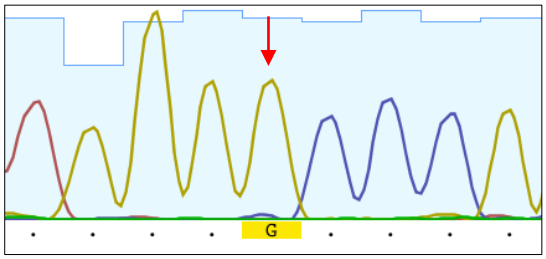


Figure 9: Example 4 – Homozygous duplication in case 91.

Example 5 (case 81): Three different sequence cutouts are shown in Figure 10. In the upper line the forward sequence from direct sequencing is shown, presenting with four double peaks that indicate heterozygous single nucleotide substitutions (red arrows). The middle line presents sequence obtained after cloning (clone number 8), showing that mutations c.[190C>G;206T>C;229A>G] occurred on the same allele. In the third sequence (clone number 9), c.166C>G is present, but not the previous mentioned mutations. In conclusion, the mutations affected different alleles.

Result: c.[166C>G];[190C>G;206T>C;229A>G]

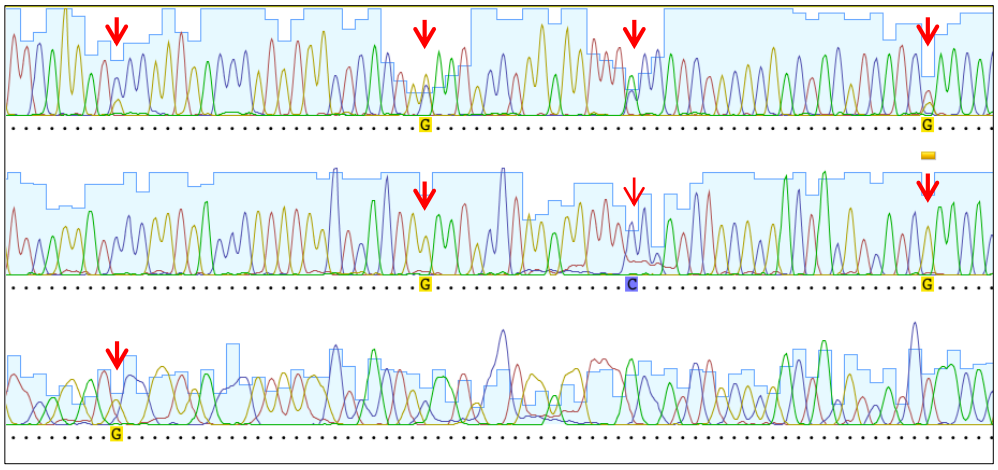


Figure 10: Example 5 – Multiple substitutions and biallelic involvement in case 81.

2.2.16 Statistical analysis

Statistical analysis was performed in order to identify differences in patient characteristics, such as gender, age, stage of disease, BM involvement, CNS involvement, LDH levels, diagnosis, probability of event free survival (pEFS) and overall survival (pOS) according to the mutation status of the analyzed candidate genes. Clinical data for each calculation referred to patients with successful investigation of the respective criteria.

Fisher's exact test (Fisher, 1922) was used to calculate probabilities (p values) in 2x2 contingency tables when less than 10 patients were analyzed in this specific setting. In 2x3, 2x4 and 2x2 tables with total counts of at least 10, Pearson's chi-squared test (Pearson, 1900) was applied.

pEFS was calculated according to Kaplan and Meier (Kaplan and Meier, 1958) under consideration of the time between date of diagnosis and either date of event or date of last follow-up. Patients with a follow-up less than two years were excluded from analysis. Differences in pEFS were compared by the log-rank test (Mantel, 1966).

Significant differences were assumed when the respective p value was lower than 0.05.

Chi-square tests, pEFS and overall survival were calculated by Dr. Martin Zimmermann, statistician of the NHL-BFM group (Pediatric Hematology and Oncology, Hanover Medical School (MHH), Germany). Calculations were conducted using SAS statistical program (SAS-PC, Version 9.3, Cary, NC: SAS Institute Inc.)

Fisher's exact tests and corresponding graphics were calculated using the software Prism 6 for Mac OS X (Version 6.0c Trial), GraphPad Software, San Diego California USA. The two-tailed option was used.

3 Results

3.1 Recurrent mutations in 13 pediatric BL in the ICGC-MMML-Seq project

Malignant mechanisms in mature aggressive B cell Non-Hodgkin lymphoma are under investigation in the ICGC-MMML-Seq project. As of June 2013, thirteen pediatric BL cases of the NHL-BFM study have been analyzed by whole-genome and whole-transcriptome sequencing. Datasets on single nuclear variants (SNVs) and small InDels were available by the projects database platform and are furthermore published online in the ICGC Data Portal (ICGC Project, 2013).

Within the 13 pediatric BL cases a total of 35,208 somatic SNVs were detected. The mean of somatic SNVs per case was 2,708 (range 1,635-4,923) and the median of potentially protein changing mutations was 28 (range 11 to 46). The distribution of SNVs is summarized in Table 20.

Table 20: Overview of SNV counts from 13 pediatric BL cases analyzed in the ICGC-MMML-Seq project

		total of SNVs	range number per patient
all SNVs		35,208	1,635-4,923
exonic		447	14-56
	silent	103	3-13
	non-silent	344	11-46
non-exonic		34,761	1,599-4,867

The list of genes that were affected by exonic, non-silent mutations (either SNVs or small InDels) included 374 different genes. Table 21 presents recurrently affected genes that were altered in more than three of the thirteen cases at a time.

Table 21: Recurrently mutated genes in > 3 pediatric BL cases analyzed in the ICGC-MMML-Seq project

gene	#/13	function	mutations typically found in
MYC	10	cell cycle progression, apoptosis and cellular transformation	Burkitt lymphoma
TP53	9	tumor suppressor protein	various malignancies
CCND3	7	regulation of CDK kinases	
FBXO11	6	adapter protein to p53	
SMARCA4	6	SNF/SWI chromatin remodeling complex	Medulloblastoma
ID3	5	inhibition of HLH proteins	
RHOA	5	cell cycle regulation	
DDX3X	4	cellular growth and division	Medulloblastoma
P2RY8	4	G protein coupled receptor	DLBCL, translocations in ALL

Regarding the pathway of interest in the current study, ten out of 13 cases showed alterations in either ID3, CCND3 and/or TCF3 (Table 22).

Table 22: ID3, TCF3 and CCND3 mutations in 13 pediatric BL cases analyzed in the ICGC-MMML-Seq project

ID in the ICGC database	ID3	TCF3	CCND3
4190495	x	x	
4194218	x		x
4194891	x		x
4189998	x		x
4182393	x		
4193278		x	
4177856			x
4142267			x
4125240			x
4177434			x
4112512			
4119027			
4133511			

x: Mutation in the gene indicated

3.2 Validation of ID3, TCF3 and CCND3 mutations in pediatric B-NHL

3.2.1 Characterization of analyzed samples

3.2.1.1 B-NHL samples of the study cohort

Eighty-four patients were selected for tumor DNA preparation according to the inclusion criteria of the study (see Patients samples). Tumor cell origin of the analyzed samples is shown in Table 23. This set of analyzed cases is referred to as ‘study cohort’.

Table 23: Origin of tumor cell samples in the study cohort

frozen tumor tissue	60
frozen cells from bone-marrow aspirates (buffy coats)	10
frozen cells from malignant pleural effusions	8
frozen cells from malignant ascites	6
total	84

For each type of tumor cell origin one representative micrograph is shown in example indicating the typical appearance of the predominance of malignant cells: tumor tissue imprint (Figure 11), bone-marrow aspirate (Figure 12), pleural effusion (Figure 13) and malignant ascites (Figure 14). As usual for hematologic evaluation smears were stained with a common staining according to Pappenheim.

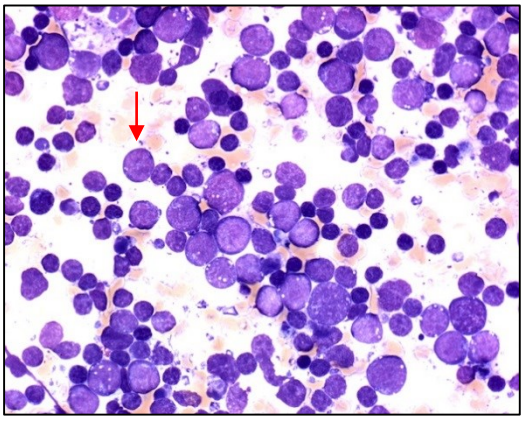


Figure 11: Example for a tumor imprint
Medium-sized cells (red arrow) represent blasts. A minor fraction of cells mostly represents lymphocytes (case 77, BL, Pappenheim staining, 60x magnification).

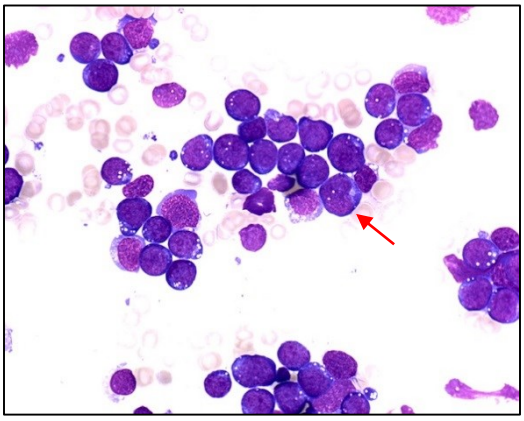


Figure 12: Example for Bone marrow aspirate (smear)
Almost only medium-sized blasts are represented (red arrow). Also few erythrocytes can be seen (case 14, B-AL, Pappenheim staining, 60x magnification)

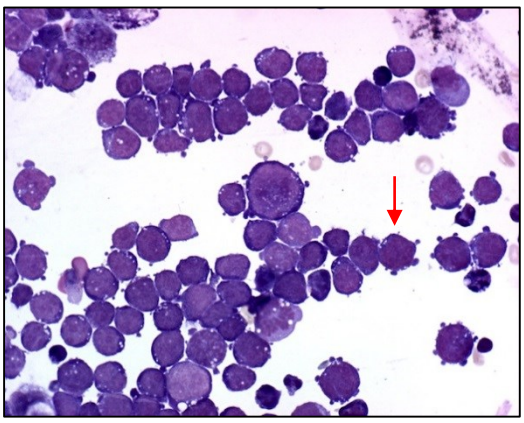


Figure 13: Example for pleural effusion (cytospin)

Almost only medium-sized blasts are represented (red arrow). Few smaller lymphocytes can be seen (case 88, BL, Pappenheim staining, 60x magnification)

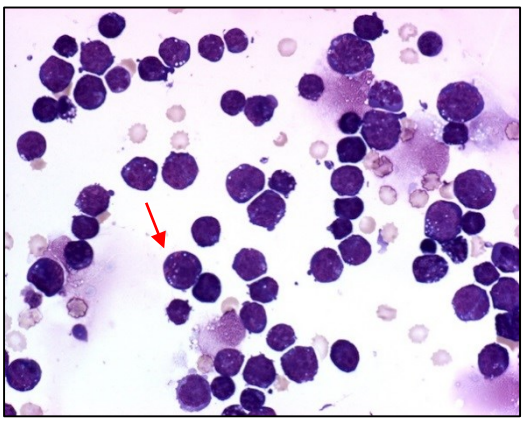


Figure 14: Example for ascites (cytospin)

Most cells represent medium-sized blasts (red-arrow). Also few lymphocytes and erythrocytes can be differentiated (case 31, DLBCL, Pappenheim staining, 60x magnification)

3.2.1.2 B-NHL samples of patients who relapsed

In a second step the study cohort was expanded to disproportionally enrich for patients who suffered from relapse. Importantly the initial material of these cases was analyzed; not the samples at the diagnosis of relapse. These cases are referred to as 'extended cohort'. Tumor cell origin of the selected samples is summarized in Table 24.

Table 24: Origin of tumor cell samples in the extended cohort

frozen tumor tissue	6
frozen cells from bone-marrow aspirates (buffy coats)	2
frozen cells from malignant pleural effusions	1
frozen cells from malignant ascites	1
total	10

3.2.1.3 Acute lymphoblastic leukemia samples

To compare the results obtained on the study cohort, a total of 96 tumor DNA samples from pediatric precursor B-cell acute lymphoblastic leukemia (pB-ALL) patients was also analyzed for *ID3* mutations. pB-ALL samples were kindly provided from the ALL-BFM study group (Prof. Schrappe, Kiel). Tumor DNA was derived from either initial bone marrow aspirates or peripheral blood after review of blast count by the ALL-BFM study center.

3.2.1.4 Patient characteristics of the study cohort

Between January 2000 and December 2012, 1127 eligible patients were registered in the NHL-BFM data center. The patient characteristics from 84 analyzed patients and 1043 not analyzed patients are presented in Table 25. The groups of analyzed and not analyzed patients were similar regarding gender, age, bone marrow (BM) involvement, central nervous system (CNS) involvement, stage of disease, LDH levels, histological B-NHL subtype and outcome (Figure 15, Figure 16). Therefore, the studied cohort featured representative characteristics for pediatric B-NHL patients in the NHL-BFM trials.

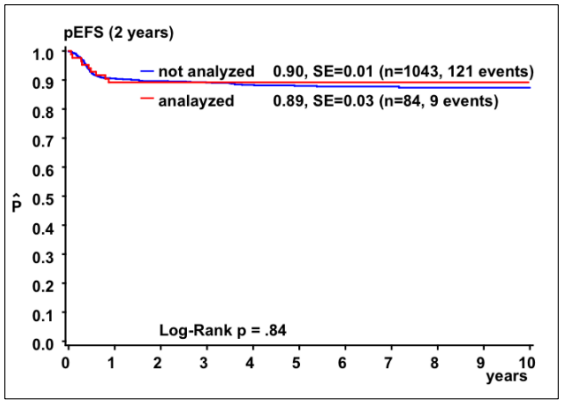


Figure 15: Event-free survival of analyzed and not analyzed patients.

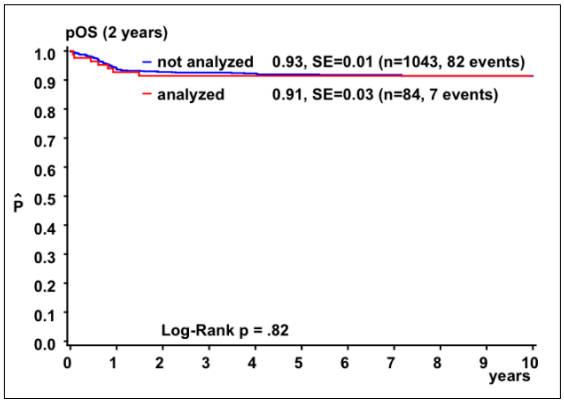


Figure 16: Overall survival of analyzed and not analyzed patients.

Table 25: Patient characteristics of study cohort

Characteristics		Patients not analyzed (n=1043)		Patients analyzed (n=84)		p value
Gender	male	827	80%	69	82%	
	female	216	20%	15	18%	0.53
Age	< 10 y	525	50%	43	51%	
	10-14 y	349	34%	27	32%	
	> 14 y	169	16%	14	17%	0.97
Stage of disease	stage I	98	10%	5	6%	
	stage II	250	25%	14	18%	
	stage III	420	42%	43	55%	
	stage IV	75	8%	3	4%	
	B-AL	163	15%	14	18%	0.17
BM involvement	yes	199	19%	14	17%	0.59
CNS involvement	yes	101	10%	8	10%	0.96
LDH	< 500 U/l	608	59%	39	46%	
	500-1000 U/l	156	15%	16	19%	
	> 1000 U/l	264	26%	29	35%	0.07
Diagnosis	BL	581	56%	49	58%	
	B-AL	163	16%	14	17%	
	DLBCL	267	25%	16	19%	
	B-NHL DD	32	3%	5	6%	0.23
Outcome	pEFS (2y)	90 ± 1%		89 ± 3%		0.84 (LR)
	pOS (2y)	93 ± 1%		91 ± 3%		0.82 (LR)

Data refer to patients with successful investigation of the respective criteria. n: number, y: years, p value: probability value, BM: bone marrow, CNS: central nervous system, LDH: lactate dehydrogenase, U/l: Units per liter, pEFS: probability of event-free survival, pOS: probability of overall survival, LR: log-rank

3.2.2 *ID3*, *TCF3* and *CCND3* sequencing in B-NHL

3.2.2.1 Results of *ID3* sequencing

ID3 mutation status was analyzed in the validation cohort by sequencing of the full *ID3* coding region. *ID3* mutations were found and verified in 56/84 B-NHL samples. The spatial distribution of *ID3* mutations is depicted in Figure 18. A detailed description of the genomic variants including the predicted change on protein level is presented in appendix 10.1.

Frequency of *ID3* mutations according to diagnosis

ID3 mutation frequency according to diagnosis was 73% for BL, 93% for B-AL and 25% for DLBCL. In the subgroup of five analyzed cases with the diagnosis B-NHL (DD) three cases showed mutations in *ID3*. Table 26 reports the absolute counts of cases with ($ID3^{mut}$) and without ($ID3^{wt}$) mutations according to the histological diagnosis. *ID3* mutations occurred at a higher frequency in BL/B-AL cases when compared to DLBCL (Figure 17).

Table 26: *ID3* mutations according to diagnosis

Diagnosis	total	$ID3^{wt}$	$ID3^{mut}$	$ID3^{mut}$ %
BL	49	13	36	73%
B-AL	14	1	13	93%
DLBCL	16	12	4	25%
B-NHL (DD)	5	2	3	60%

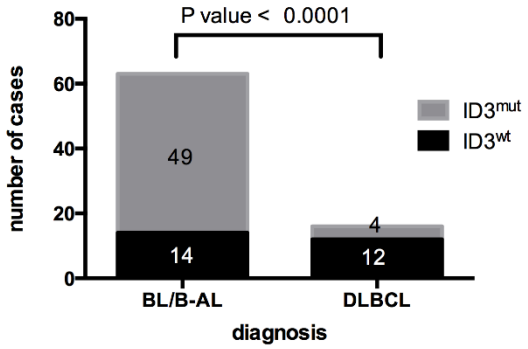


Figure 17: Correlation of $ID3^{wt}$ and $ID3^{mut}$ cases according to diagnosis. $ID3^{mut}$ cases occurred at a higher frequency in BL/B-AL when compared to DLBCL.

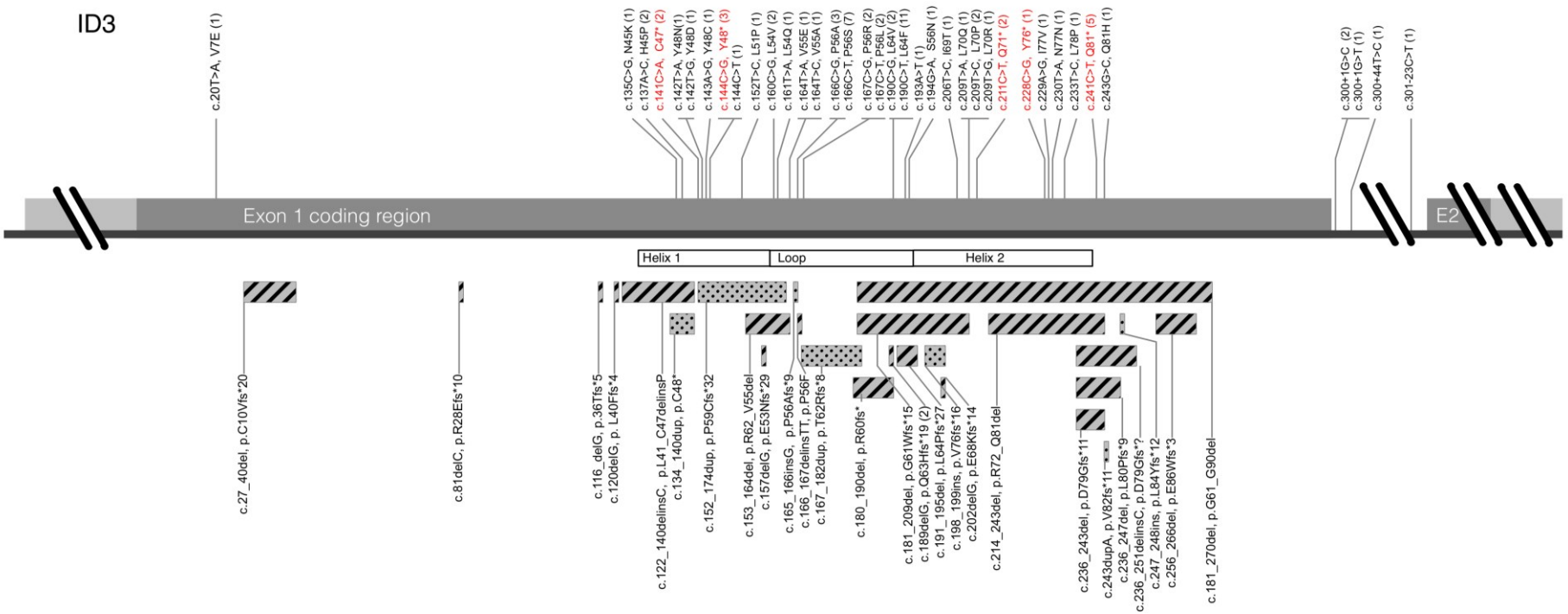


Figure 18: ID3 mutations in the study cohort.

The ID3 coding region of exon 1 is illustrated with single base pair substitutions on the upper and more complex alterations (insertions, deletions, InDels, duplications) on the lower site. Substitutions resulting in a nonsense mutation are depicted in red. Hatched bars delineate deletions and InDels, dotted bars characterize insertions and duplications. Each mutation is labeled with correspondent description on genomic and protein level, as well as the absolute number of occurrence in brackets. The functional helix-loop-helix domain is mapped according to UniProt entry Q02535 (UniProt database, 2013).

Number and type of ID3 mutations

Of the 56 B-NHL cases with *ID3* mutations, 31 cases presented with multiple mutations in *ID3*:

- 26 cases with two *ID3* mutations
- 4 with three *ID3* mutations
- one with four *ID3* mutations

In the remaining 25 cases, one *ID3* mutation each was detected.

Figure 19 depicts the distribution of ID3 mutations in the 56 B-NHL cases with identified alterations.

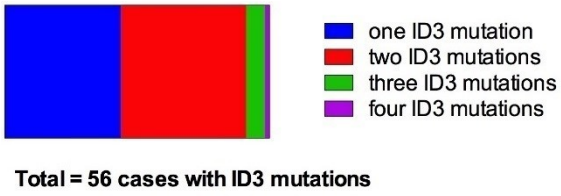


Figure 19: Distribution of the number of ID3 mutations in 56 B-NHL cases

Ninety-three *ID3* mutations were found in the 56 B-NHL cases, including 66 substitutions, 17 deletions, 4 duplications, 3 insertions and 3 InDels (deletion followed by an insertion) (Figure 20).

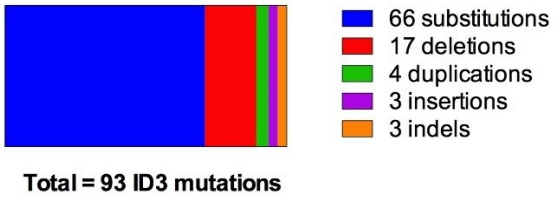


Figure 20: Mutation types found in 56 B-NHL cases

Five mutations presented with homozygous characteristics in sequence analysis: three substitutions and two deletions.

Four disambiguates were genomic variants that were not predicted to result in change on amino acid level: c.144C>T, c.193A>T, c.300+44T>C and c.301-23C>T. However, all these cases also harbored at least one other *ID3* mutation.

Concerning the 66 substitutions, mutations affected the nucleotide base pairs G/C in 74% (49/66) and A/T in 27% (18/66) of the cases. The substitutions occurred in the genomic region between c.135 and c.300+1, which contains a G/C-content of 64%. Assuming a random chance for mutations to affect nucleotides in this selected region, a

distribution of 44 hits in G/C base pairs and 25 in A/T base pairs would have been expected. There was no significant difference between these expected values and the findings, indicating no certain preference of the underlying mutation mechanism on affecting G/C or A/T base pairs.

Biallelic involvement and homozygous mutations

To investigate whether mutations occurred on the same or different alleles, ten out of 31 cases with the finding of multiple mutations in *ID3* where selected for cloning. Sequence analysis of different clones from the same sample showed different alleles to be affected in all analyzed cases (Table 27).

In the group of five homozygous mutations two were also analyzed by cloning. In both cases, the mutation was present in at least 80% of the clones, supporting evidence of biallelic involvement. This finding could indicate either a mutation affecting both alleles or combined focal ID3 loss and occurrence of a mutation on the remaining allele. However, in either case this would functionally result in expression of only altered or impaired ID3 protein and therefore this effect was considered as biallelic involvement with respect to the effect on protein level.

Table 27 summarizes sequencing results of the 10 heterozygous and 2 homozygous cloned cases, whose DNA has been subcloned.

Table 27: Results of ID3 clone sequencing of selected cases

case	allele 1 (number of clones)	allele 2 (number of clones)	effect
35	c.[167C>T] (3)	c.[181_209del] (2)	biallelic
11	c.[144C>G] (6)	c.[300+1G>C] (4)	biallelic
23	c.[236_251delinsC] (3)	c.[300+1G>C] (4)	biallelic
73	c.[81delC] (2)	c.[247_248ins] (6)	biallelic
80	c.[160C>G] (2)	c.[165_166insG] (2)	biallelic
1	c.[198_199insCTAAG] (3)	c.[194G>A] (4)	biallelic
81	c.[166C>G] (5)	c.[190C>G;206T>C;229A>G] (2)	biallelic
8	c.[144C>G;243G>C] (5)	c.[236_243del] (4)	biallelic
10	c.[153_164del] (3)	c.[256_266del] (4)	biallelic
53	c.[134_140dup] (5)	c.[209T>C] (2)	biallelic

Continue Table 27

case	allele 1 (number of clones)	allele 2 (number of clones)	effect
54	c.[214_243del] (9)	wildtype (1)	biallelic
92	c.[122_130del] (5)	wildtype (0)	biallelic

Focal ID3 loss in cases with homozygous *ID3* mutation

In order to investigate the likelihood of focal *ID3* loss in cases that presented with homozygous characteristics in electropherogram analysis, the occurrence of genomic variants at SNP positions was assessed for all cases analyzed in the study cohort and the extended cohort. This analysis was based on the assumption that cases with heterozygous SNPs are less likely presenting allelic loss, while cases with only homozygous SNPs in *ID3* could represent focal loss of one allele. SNPs within the *ID3* coding region that were investigated for this analysis are listed in appendix 10.4.

In total, 56 cases showed at least one heterozygous SNP and 38 cases showed only homozygous SNPs. Interestingly, none of the cases with apparent homozygous *ID3* mutation presented with a heterozygous SNP simultaneously. Frequency of cases without any heterozygous SNPs was remarkably higher in cases without homozygous *ID3* mutation (Table 28).

Table 28: Correlation of cases with heterozygous SNPs and *ID3* mutations presenting homozygous

Heterozygous SNP	total	*all other cases of study cohort and extended cohort*	homozygous $ID3^{mut}$	p value (Fisher)
yes	56	56	0	
no	38	31	7	
total	94	87	7	0.0012

Mutational hotspots

The dataset of *ID3* mutations in the study cohort showed a number of recurrent changes affecting the same nucleotide position. For this analysis also deletions, insertions, duplications and InDels have been taken into account by counting the first nucleotide affected, respectively.

Changes of c.C190 were most often (13), followed by c.C166 (10) and c.C241 (5). Table 29 gives an overview on genomic *ID3* nucleotide positions that occurred to be affected in more than two cases.

Table 29: *ID3* hotspot mutations (nucleotide positions affected in > 2 cases)

nucleotide and position	number of cases affected
c.C190	13
c.C166	10
c.C241	5
c.C167	4
c.C144	3
c.T209	3
c.A236	3
c.G300+1	3

Mutations affecting the functional *ID3* Helix-Loop-Helix domain

The helix-loop-helix (HLH) domain of ID3 is the proteins functional site of interaction with other HLH proteins. On the genomic level 77 of 93 (83%) mutations directly occurred within the HLH coding region. The remaining 18 mutations were allocated either close to the splice-site of exon 1 (4 cases), upstream (8 cases) or downstream (5 cases) of the HLH domain or in the intronic region between exon 1 and 2 (1 case). Of notice, all cases with a mutation not directly affecting the HLH domain or the splice-site had at least a second mutation in the HLH domain (Figure 18).

3.2.2.2 Results of *TCF3* sequencing

The coding region of exon 16 nearly covers the full HLH domain coding sequence. Furthermore it was shown by the index studies that mutations in TCF3 always occurred in this exon. Therefore, exon 16 was analyzed for *TCF3* mutations in the study cohort. *TCF3* mutations were identified in eight out of 84 cases. All changes were heterozygous substitutions and there was no case presenting multiple mutations in *TCF3*.

Figure 21 shows the distribution of mutations in *TCF3* exon 16. The detailed description of genomic variants including the predicted change on protein level is presented in appendix 10.1.

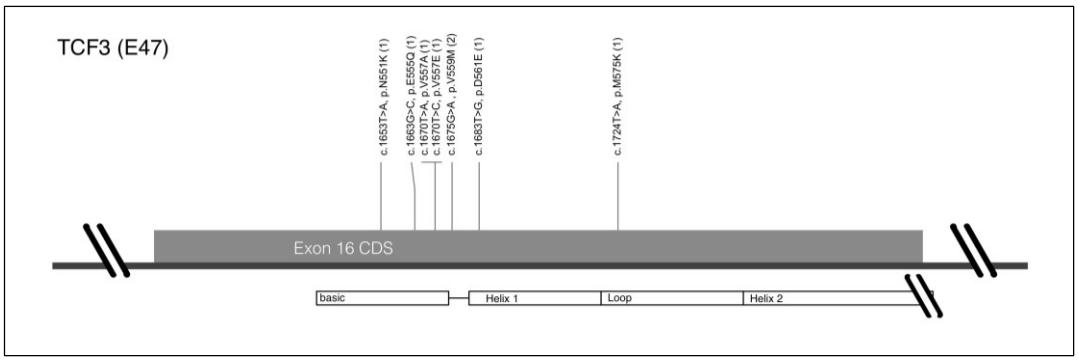


Figure 21: TCF3 mutations in the study cohort
The coding region of TCF3 exon 16 is illustrated. Each substitution is labeled with correspondent description on genomic and protein level, as well as the absolute frequency of occurrence in brackets. The functional basic-Helix-Loop-Helix domain is mapped according to the description of the functional sites in UniProt entry P15923-2 (UniProt database, 2013).

Frequency of *TCF3* mutations according to diagnosis

The overall finding of mutations in *TCF3* was considerably lower compared to the high frequency of *ID3* mutated cases. In BL/B-AL the frequency of *TCF3* mutations was 13% (8/63). There were no mutations of the *TCF3* gene in cases with the diagnosis of DLBCL or B-NHL (DD). Absolute counts of cases with (*TCF3*^{mut}) and without (*TCF3*^{wt}) mutations in *TCF3* are presented in Table 30.

Table 30: *TCF3* mutations according to diagnosis

Diagnosis	total	*TCF3*^{wt}	*TCF3*^{mut}	*TCF3*^{mut} %
BL	49	43	6	12%
B-AL	14	12	2	14%
DLBCL	16	16	0	0%
B-NHL (DD)	5	5	0	0%

Number and types of mutation, mutation hotspots

Due to the small number of *TCF3* mutations, it was not possible to gain informative results regarding different types of mutation and recurrently affected nucleotides. One mutation, G1675A, was present in two cases.

Mutations affected the *TCF3* bHLH domain in all cases

All eight mutations occurred in or in between the coding region of the bHLH binding domain of *TCF3* (Figure 21).

3.2.2.3 Results of *CCND3* sequencing

CCND3 exon 5 was sequenced to analyze for *CCND3* mutations in the study cohort. Exon 5 was previously shown to represent a mutation hotspot in BL by the index studies. *CCND3* mutations were present in 26/84 (31%) cases. The findings are depicted in Figure 22. The complete description of genomic variants including the predicted change on protein level is presented in appendix 10.1.

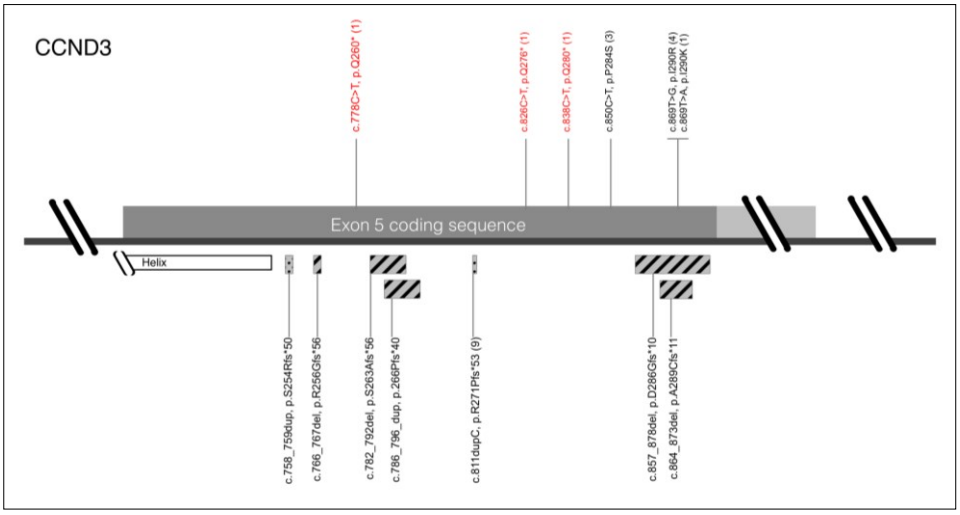


Figure 22: CCND3 mutations in the study cohort.
The coding region of CCND3 exon 5 is illustrated, with substitutions on the upper and more complex alterations (insertions, deletions, indels, duplications) on the lower site. Substitutions resulting in a nonsense mutation are depicted in red. Hatched bars delineate deletions and indels, dotted bars characterize insertions and duplications. Each mutation is labeled with correspondent description on genomic and protein level, as well as the absolute frequency of occurrence in brackets. A part of a functional helix domain is mapped according to the description of the functional sites in UniProt entry P30281.

Frequency of *CCND3* mutations according to diagnosis

In the study cohort, 26 cases harbored *CCND3* mutations. Twenty-three out of 63 BL/B-AL cases (37%) presented with a mutation in *CCND3*. In DLBCL there were two out of 16 cases affected. One mutation was present in a case with B-NHL (DD) diagnosis. Table 31 summarizes absolute counts of *CCND3* mutations according to diagnosis.

Table 31: *CCND3* mutations according to diagnosis

Diagnosis	total	*CCND3*^{wt}	*CCND3*^{mut}	*CCND3*^{mut} %
BL	49	34	15	31%
B-AL	14	6	8	57%
DLBCL	16	14	2	13%
B-NHL (DD)	5	4	1	20%

Number and types of mutation

In the analysis of the *CCND3* gene, only heterozygous mutations were detected and no cases presenting with more than one mutation.

The 26 cases with *CCND3* mutation included 5 deletions, 11 duplications and 10 single nucleotide substitutions (Figure 23).

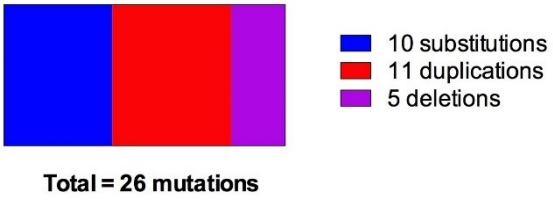


Figure 23: CCND3 Mutation types found in 26 B-NHL cases

Mutation hotspots

Mutations affected nucleotide c.C811 with a cytosine duplication in 9 cases, resulting in a protein elongating frameshift. Four cases presented with c.T869G substitution and three cases had c.C580T mutation.

3.2.2.4 Mutual relationship of mutations in different candidate genes and correlation with MYC rearrangement

ID3, *TCF3* and *CCND3* sequencing results were integrated to analyze mutual relationship. The majority of cases with mutations had an exclusive mutation of *ID3* (51%). The combination of *ID3*^{mut} and *CCND3*^{mut} was present in 31% of the cases. There were two cases (case 12, case 33) that had mutations in all three genes.

Table 32 and Figure 24 summarize and present different combinations of simultaneous occurrence of *ID3*, *TCF3* and *CCND3* mutations in non-wildtype cases. Table 34 gives an overview on mutual relationship of overall sequencing results including diagnosis and MYC rearrangement.

Table 32: Mutual relationship of mutations in different candidate genes

	number of cases	% of cases with mutation
exclusive *ID3*^{mut}	32	51%
exclusive *TCF3*^{mut}	2	3%
exclusive *CCND3*^{mut}	4	6%

Continue Table 32		
exclusive *ID3*^{mut} and *TCF3*^{mut}	3	5%
exclusive *ID3*^{mut} and *CCND3*^{mut}	19	31%
exclusive *TCF3*^{mut} and *CCND3*^{mut}	1	2%
ID3^{mut} and *TCF3*^{mut} and *CCND3*^{mut}	2	3%
total	63	100%

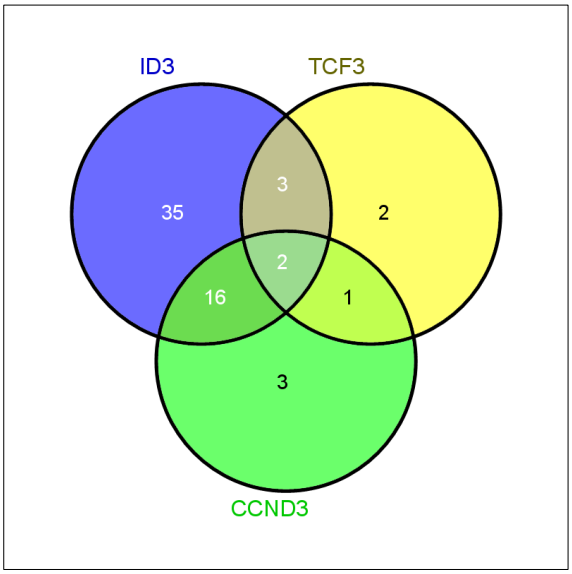


Figure 24: Distribution of cases with certain combinations of ID3, TCF3 and CCND3 mutations. Each circles' number represents the number of cases with mutations in the respective gene. Numbers in overlapping areas show summarize cases with mutations in multiple genes. This diagram was created using Venny (Oliveros, 2007).

Correlation with *MYC* rearrangement

Results of Fluorescence-in-situ-Hybridization (FISH) for detection of *MYC* rearrangements were available for 78 cases of the studied cohort. Fifty-eight of 66 *MYC* rearrangement positive cases (88%) had at least one mutation in *ID3* and/or *TCF3* and/or *CCND3*. In contrast, within 12 *MYC* rearrangement negative patients only one harbored an *ID3* mutation.

Interestingly, the *MYC* FISH report of the case with *ID3* mutation but no *MYC* rearrangement (case 68) states that 15% of the analyzed cells presented with an additional signal of *MYC*. In the report it was speculated that this observation might be the result of either a translocation with a proximal break-point of *MYC* or a result of a *MYC* duplication. As the *MYC* status of the respective case remained unclear, this case was not included in further *MYC* correlating analyses.

Table 33 shows the correlation between *MYC* rearrangement positive cases and cases with mutation in at least one of the three analyzed genes (p value < 0.0001).

Table 33: Correlation of mutation status with MYC rearrangement status

	total	*ID3*^{wt} and *TCF3*^{wt} and *CCND3*^{wt}	*ID3*^{mut} and/or *TCF3*^{mut} and/or *CCND3*^{mut}	p value (Fisher)
MYC rearrangement positive	66	8	58	
MYC rearrangement negative	12	11	1	
total	78	19	59	< 0.0001

Table 34: Overview of B-NHL patient sequencing results on *ID3*, *TCF3* and *CCND3*

Specimen	case 33	case 12	case 66	case 32	case 23	case 57	case 1	case 71	case 44	case 5	case 55	case 50	case 76	case 59	case 64	case 9	case 48	case 24	case 54	case 14	case 7	case 43	case 40	case 16	case 49	case 15	case 53	case 61	case 13	case 67	case 25	case 8	case 11	case 10	case 84	case 79	case 37	case 81	case 83	case 75	case 35	case 36	case 26	case 45	case 69	case 39	case 73	case 78	case 72
Diagnosis	BL	BL	B-AL*	BL	BL	BL	BL	BL*	B-AL*	BL	BL	B-AL*	B-AL*	BL	B-AL*	BL	B-AL*	BL	BL*	B-AL	B-AL	BL	B-AL*	B-AL	BL	BL	BL*	BL*	BL	BL	BL	BL	BL*	BL*	BL	B-AL	BL	B-AL*	BL	BL*	BL	BL*	BL	BL	BL*	BL	B-AL	BL
MYC rear.	+	+	+	+	+	o	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+*	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ID3																																																	
TCF3																																																	
CCND3																																																	

Specimen	case 2	case 52	case 63	case 74	case 38	case 22	case 82	case 29	case 3	case 77	case 20	case 65	case 34	case 4	case 80	case 68	case 27	case 19	case 42	case 31	case 56	case 70	case 47	case 60	case 6	case 28	case 18	case 58	csae 51	case 46	case 62	case 17	case 21	case 41	case 30
Diagnosis	BL	B-AL*	BL	BL*	BL	BL	BL	BL	BL*	BL	BL	BL	BL	BL	B-NHL (DD)	B-NHL (DD)	B-NHL (DD)	B-NHL (DD)	B-NHL (DD)	DLBCL**	DLBCL	DLBCL**	DLBCL	DLBCL	DLBCL	DLBCL	DLBCL	DLBCL	DLBCL	DLBCL	DLBCL	DLBCL	DLBCL	DLBCL	DLBCL
MYC rear.	+	+	+	+	+	+	-	-	+	+	+	+	+	+	o	-	+	+	-	+	+	o	+	o	o	-	o	-	-	-	-	-	-	-	+
ID3																																			
TCF3																																			
CCND3																																			

*Reference diagnosis according to NHL-BFM study: Burkitt lymphoma (BL), Burkitt leukemia (B-AL), Diffuse large B-cell lymphoma (DLBCL), B-NHL unclassifiable (B-NHL DD), * no reference pathology review available; ** no reference diagnosis available; MYC status: + MYC rearrangement positive, - MYC rearrangement negative, o MYC status unknown, +* MYC-Ig PCR positive; ID3, TCF3 (Exon 16) and CCND3 (Exon 5, coding region) mutation status: red block = mutated, black block = wildtype*

3.2.3 Clinical characteristics and outcome according to *ID3*, *TCF3* and *CCND3* mutation status

3.2.3.1 Correlation of *ID3*, *TCF3* and *CCND3* results with clinical characteristics

Based on the high correlation between cases with mutation and concurrent *MYC* rearrangement clinical characteristics and outcome were analyzed within the group of 66 *MYC* rearrangement positive B-NHL patients.

However, there was no significant difference comparing the clinical features between 66 *MYC* rearrangement positive B-NHL patients with and without *ID3* mutation with respect to gender, age, stage of disease, BM involvement, CNS involvement and LDH levels. This was also the case for the comparison between *TCF3*^{mut} and *TCF3*^{wt}, as well as *CCND3*^{mut} and *CCND3*^{wt} cases. The clinical characteristics are summarized in Table 35.

Regarding pEFS and pOS there were no significant differences between wildtype and mutated cases in the study cohort, as shown in the Kaplan-Meier plots for *ID3* (Figure 25, Figure 26), *TCF3* (Figure 27, Figure 28) and *CCND3* (Figure 29, Figure 30).

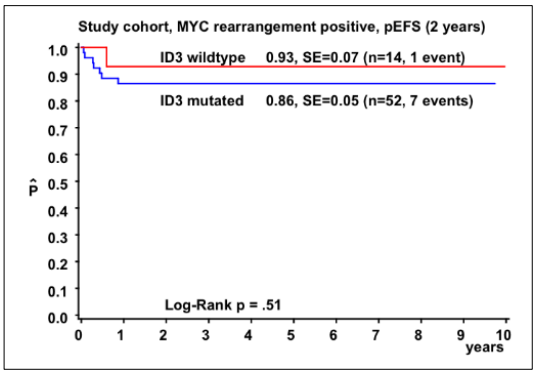


Figure 25: Event-free survival (2 years) for ID3 mutated and non-mutated cases

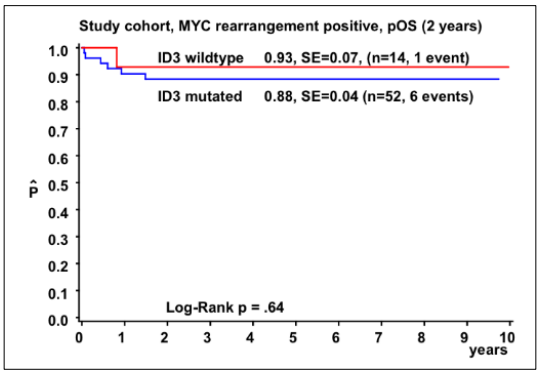


Figure 26: Overall survival (2 years) for ID3 mutated and non-mutated cases

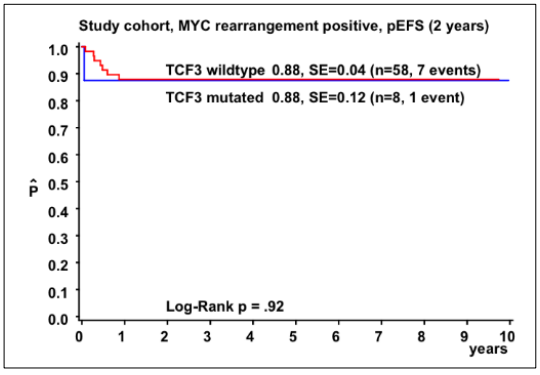


Figure 27: Event-free survival (2 years) for TCF3 mutated and non-mutated cases

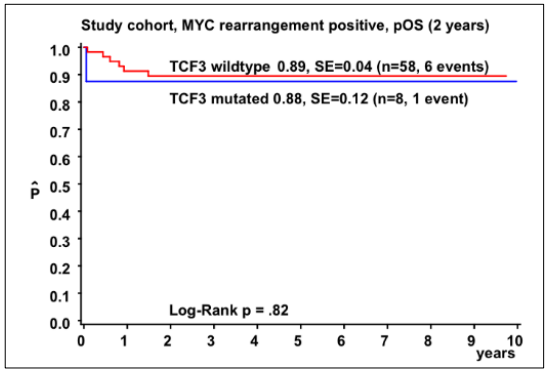


Figure 28: Overall survival (2 years) for TCF3 mutated and non-mutated cases

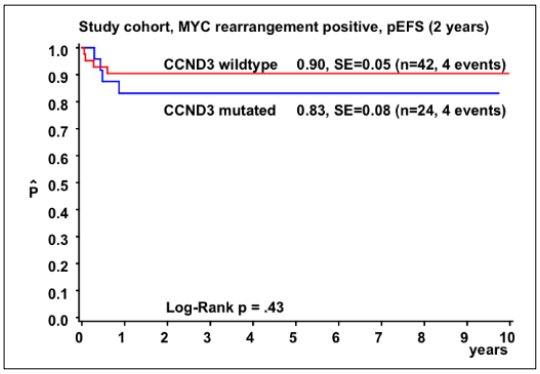


Figure 29: Event-free survival (2 years) for CCND3 mutated and non-mutated cases

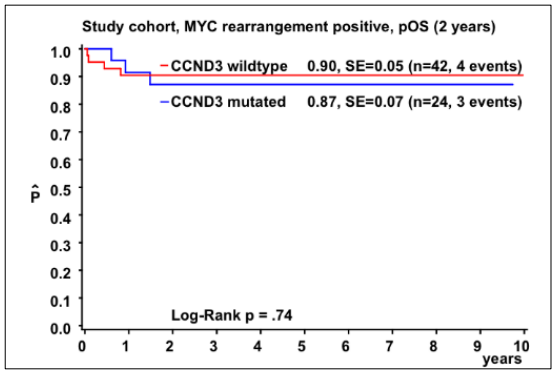


Figure 30: Overall survival (2 years) for CCND3 mutated and non-mutated cases

Table 35: Clinical characteristics of 66 *MYC* rearrangement positive patients with and without mutation in *ID3*, *TCF3* and *CCND3*

Characteristics		$ID3^{mut}$		$ID3^{wt}$		p value	$TCF3^{mut}$		$TCF3^{wt}$		p value	$CCND3^{mut}$		$CCND3^{wt}$		p value
Gender	male	44	85%	13	93%		6	75%	51	88%		19	79%	38	90%	
	female	8	15%	1	7%	0.67	2	25%	7	12%	0.30	5	21%	4	10%	0.27
Age	< 10	31	60%	7	50%		7	88%	31	53%		16	67%	22	52%	
	10-14	14	27%	6	43%		1	12%	19	33%		5	21%	15	36%	
	> 14	7	7%	1	7%	0.48	0	0%	8	14%	0.18	3	12%	5	12%	0.44
Stage of	I	1	2%	2	14%		0	0%	3	6%		1	3%	2	5%	
disease	II	8	17%	3	21%		4	50%	7	13%		5	21%	6	16%	
	III	25	52%	7	50%		2	25%	30	56%		8	34%	24	63%	
	IV	1	2%	1	7%		0	0%	2	4%		2	8%	0	0%	
	B-AL	13	27%	1	7%	0.18	2	25%	12	22%	0.12	8	34%	6	16%	0.09
BM	yes	13	25%	1	7%	0.27	2	25%	12	21%	0.67	8	33%	6	14%	0.12
CNS	yes	6	12%	1	7%	1.00	1	13%	5	7%	0.72	5	21%	2	5%	0.09
LDH	< 500 U/l	17	33%	9	64%		4	50%	22	39%		8	33%	18	43%	
	500-1000	11	21%	1	7%		2	25%	9	16%		5	21%	7	17%	
	> 1000 U/l	24	46%	4	29%	0.09	2	25%	26	46%	0.24	11	46%	17	40%	0.74
Diagnosis	BL	35	67%	11	79%		6	75%	40	69%		14	58%	32	14%	
	B-AL	13	25%	1	7%		2	25%	12	21%		8	33%	6	76%	
	DLBCL	3	6%	1	7%		0	0%	4	7%		1	4%	3	7%	
	B-NHL	1	2%	1	7%	0.41	0	0%	2	3%	0.82	1	4%	1	2%	0.30
Outcome	pEFS (2y)	86 ± 5%		93 ± 7%		0.51 (LR)	88 ± 12%		88 ± 4%		0.92	83 ± 8%		90 ± 5%		0.43 (LR)
	pOS (2y)	88 ± 4%		93 ± 7%		0.64 (LR)	88 ± 12%		89 ± 4%		0.82	87 ± 7%		90 ± 5%		0.74 (LR)

Data refer to patients with successful investigation of the respective criteria. mut: mutated, wt: wildtype,, y: years, p value: probability value, BM: bone marrow, CNS: central nervous system, LDH: lactate dehydrogenase, U/l: Units per liter, pEFS: probability of event-free survival, pOS: probability of overall survival, LR: log-rank

3.2.3.2 Mutual relationship of mutations in the candidate genes and correlation with clinical characteristics

To investigate differences between cases with certain combinations of *ID3*, *TCF3* and *CCND3* mutations, each combination that occurred in the mutual relationship analysis was evaluated separately with respect to clinical relevance.

The genotypes were assessed for significant correlations with clinical relevance with respect to gender, age, stage of disease, CNS involvement, BM involvement, LDH levels, diagnosis, risk group and pEFS and survival. Results are summarized in Table 36. The complete data set of correlations between certain mutation patterns and clinical data is available in appendix 10.2.

Table 36: Correlation according to mutual relationship

Selection	Number of cases	Comparison of clinical characteristics according to mutated and wildtype status
ID3	52	no statistically significant results
TCF3	8	no statistically significant results
CCND3	24	no statistically significant results
ID3 exclusive	29	no statistically significant results
TCF3 exclusive	2	no statistically significant results
CCND3 exclusive	3	no statistically significant results
ID3 and/or *TCF3*	55	LDH, risk group
ID3 and/or *CCND3*	56	no statistically significant results
TCF3 and/or *CCND3*	29	gender, BM involvement, CNS involvement
ID3 and/or *TCF3* and/or *CCND3*	58	no statistically significant results
ID3 and *TCF3*	5	no statistically significant results
ID3 and *CCND3*	20	stage, BM involvement
TCF3 and *CCND3*	3	no statistically significant results
ID3 and *TCF3* exclusive	3	gender
ID3 and *CCND3* exclusive	18	stage, BM involvement, diagnosis
TCF3 and *CCND3* exclusive	1	no statistically significant results
ID3 and *TCF3* and *CCND3*	2	no statistically significant results

Results of statistically significant differences are presented in the following:

Selection: Cases with *ID3* and/or *TCF3* mutation

- Frequency of *ID3* and/or *TCF3* mutations was higher in patients with LDH over 500 U/l compared to patients with lower LDH levels (Table 37).
- Frequency of *ID3* and/or *TCF3* mutations was higher in patients belonging to risk group R3 or R4 compared to patients with lower risk group (Table 38).

Table 37: Correlation of cases with *ID3* and/or *TCF3* mutation and LDH level

LDH level (U/l)	total	$ID3^{wt}$ and $TCF3^{wt}$	$ID3^{mut}$ and/or $TCF3^{mut}$	p value (Fisher)
LDH <500	26	8	18	
LDH >=500	40	3	37	
total	66	11	55	0.019

Table 38: Correlation of cases with *ID3* and/or *TCF3* mutation and risk group

Risk group	total	$ID3^{wt}$ and $TCF3^{wt}$	$ID3^{mut}$ and/or $TCF3^{mut}$	p value (Fisher)
R1 or R2	24	8	16	
R3 or R4	40	3	37	
total	64	11	53	0.014

Selection: Cases with *TCF3* and/or *CCND3* mutation

- Frequency of *TCF3* and/or *CCND3* mutations was higher in female patients compared to male patients (Table 39).
- Frequency of *TCF3* and/or *CCND3* mutations was higher in patients with BM involvement compared to patients without BM involvement (Table 40).
- Frequency of *TCF3* and/or *CCND3* mutations was higher in patients with CNS disease compared to patients without CNS disease (Table 41).

Table 39: Correlation of cases with *TCF3* and/or *CCND3* mutation and gender

Gender	total	*TCF3*^{wt} and *CCND3*^{wt}	*TCF3*^{mut} and/or *CCND3*^{mut}	p value (Fisher)
male	57	35	22	
female	9	2	7	
total	66	38	28	0.036

Table 40: Correlation of cases with *TCF3* and/or *CCND3* mutation and BM involvement

BM involvement	total	*TCF3*^{wt} and *CCND3*^{wt}	*TCF3*^{mut} and/or *CCND3*^{mut}	p value (Fisher)
no	52	33	19	
yes	14	4	10	
total	66	37	29	0.032

Table 41: Correlation of cases with *TCF3* and/or *CCND3* mutation and CNS disease

CNS disease	total	*TCF3*^{wt} and *CCND3*^{wt}	*TCF3*^{mut} and/or *CCND3*^{mut}	p value (Fisher)
no	59	36	23	
yes	7	1	6	
total	66	37	29	0.038

Selection: Cases with exclusive combination of *ID3* and *TCF3* mutation

- Frequency of exclusive combination of *ID3* and *TCF3* mutation was higher in female patients compared to male patients (Table 42).

Table 42: Correlation of cases with exclusive *ID3* and *TCF3* mutation and gender

Gender	total	Rest of the cohort	*ID3*^{mut} and *TCF3*^{mut} only	p value (Fisher)
male	57	56	1	
female	9	7	2	
total	63	63	3	0.047

Selection: Cases with *ID3* and *CCND3* mutation

- Frequency of cases with *ID3* and *CCND3* mutation was higher in patients with BM involvement compared to patients without BM involvement (Table 43).
- Frequency of cases with *ID3* and *CCND3* mutation was higher in patients with stage B-AL compared to patients with other stages (Table 44).

Table 43: Correlation of cases with *ID3* and *CCND3* mutation and BM involvement

BM involvement	total	Rest of the cohort	*ID3*^{mut} and *CCND3*^{mut}	p value (Fisher)
no	52	40	12	
yes	14	6	8	
total	66	46	20	0.021

Table 44: Correlation of cases with *ID3* and *CCND3* mutation and stage

Stage	total	Rest of the cohort	*ID3*^{mut} and *CCND3*^{mut}	p value (Fisher)
I, II, III, IV	48	36	12	
B-AL	14	6	8	
total	62	42	20	0.048

Selection: Cases with exclusive combination of *ID3* and *CCND3* mutation

- Frequency of exclusive combination of *ID3* and *CCND3* mutation was higher in patients with stage B-AL compared to patients with other stages (Table 45).
- Frequency of exclusive combination of *ID3* and *CCND3* mutation was higher in patients with BM involvement compared to patients without BM involvement (Table 46).

Table 45: Correlation of cases with exclusive *ID3* and *CCND3* mutation and stage

Stage	total	Rest of the cohort	*ID3*^{mut} and *CCND3*^{mut} only	p value (Fisher)
I, II, III, IV	48	38	10	
B-AL	14	6	8	
total	62	43	18	0.017

Table 46: Correlation of exclusive *ID3* and *CCND3* mutation and BM involvement

BM involvement	total	Rest of the cohort	*ID3*^{mut} and *CCND3*^{mut} only	p value (Fisher)
no	52	42	10	
yes	14	6	8	
total	66	48	18	0.014

- The occurrence of cases with mutation in two or more genes was higher in B-AL (9/14) when compared to BL (14/46) (Figure 31). This was also true when cases with *ID3* and *CCND3* mutations in combination were considered. These mutations were more often observed in B-AL compared to BL (BL 11/46, B-AL 8/14; Fishers exact: p value < 0,045).

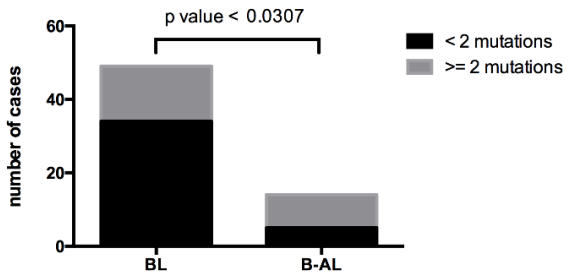


Figure 31: Correlation between the number of affected genes and diagnosis

3.2.3.3 Number and location of *ID3* mutations

To investigate the impact of different *ID3* mutations affecting specific functional regions of the ID3 protein on clinical characteristics, *ID3* coding region was divided into four sections (Table 47). Sections were built upon the mutational distribution as presented in appendix 10.3, with section two and three representing mutations close to helix 1 and helix 2, respectively, and section one and four covering the up and downstream region of the functional domain. In cases with more complex alterations, like deletions, duplications and InDels, first affected amino acid residues were counted, respectively.

Table 47: *ID3* mutation cluster analysis

Section	Covered protein sequence (amino acid positions)	Number of mutations in this section
1	1 to 40	3
2	40 to 59	28
3	60 to 80	36
4	81 and above	5

There was no significant difference between mutations in different sections with respect to clinical characteristics and outcome (data not shown).

Similar analysis was performed for the three most recurrently altered nucleotide positions, *ID3* c.190, c.166 and c.241. There was no significant correlation for c.190 and c.166 cases.

Cases with ID3 mutations that affected nucleotide position c.241

- Frequency of *ID3* mutations affecting nucleotide position c.241 was higher in patients with B-AL compared to BL stages I-IV.
- Frequency of *ID3* mutations affecting nucleotide position c.241 was higher in patients with bone marrow involvement compared to patients without bone marrow involvement (Table 49).

Table 48: Correlation of *ID3* mutations affecting c.241 and stage

Stage	total	Rest of the cohort	*c.241 ID3*^{mut}	p value (Fisher)
I, II, III, IV	48	47	1	
B-AL	14	10	4	
total	62	57	5	0.008

Table 49: Correlation of *ID3* mutations affecting c.241 and BM involvement

BM involvement	total	Rest of the cohort	*c.241 ID3*^{mut}	p value (Fisher)
no	52	51	1	
yes	14	10	4	
total	66	61	5	0.006

There was no difference between cases without, cases with one and cases with more than one mutation in *ID3* regarding clinical relevance (data not shown).

3.3 *ID3*, *TCF3* and *CCND3* mutations in B-NHL patients who relapsed

3.3.1 Patient characteristics

To further investigate the prognostic role of *ID3*, *TCF3* and *CCND3* mutations, ten additional *MYC* rearrangement positive B-NHL patients were identified who suffered relapse. The initial sample not the relapse sample was analyzed. Cases were analyzed together with seven additional cases also suffering from relapse already included in the study cohort.

3.3.2 Results of *ID3*, *TCF3* and *CCND3* sequencing

In 17 relapsed/progressed *MYC* rearrangement positive B-NHL cases, occurrence of *ID3*, *TCF3* and *CCND3* mutations was 14/17, 1/17 and 6/17, respectively and therefore not significantly different from non-relapse cases with respect to frequency.

Due to the sufficiently high number of *ID3* mutations the pattern of *ID3* mutations between cases with relapse/progress and cases with freedom of progression was investigated, to analyze for commonly affected mutation sites. However, there was no clear difference between the two groups, as visualized in Figure 32.

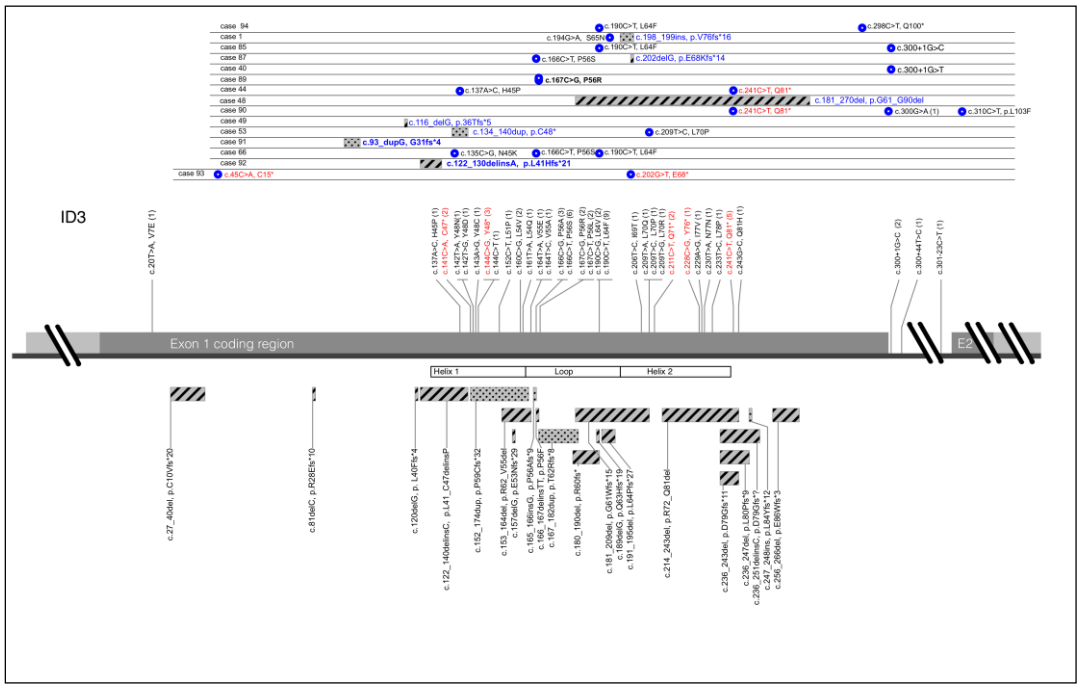


Figure 32: Comparison of ID3 mutations between relapse and non-relapse cases.
This figure refers to Figure 18. In addition, mutations found in relapsed cases are represented separately at the top of the figure. The respective case numbers are indicated. The blue dots delineate substitutions of relapsed cases. In the direct comparison there is no clear mapping of certain hotspot regions within relapse cases and no difference of mutational patterns compared to non-relapse cases.

3.4 *ID3* sequencing in pB-ALL

DNA samples from 96 pediatric pB-ALL patients were analyzed for *ID3* mutations. Two out of 96 (2%) patients presented with genomic variants of *ID3*. Table 50 summarizes pB-ALL cases with genomic variants.

pB-ALL-42 had one silent mutation in the coding sequence and duplication in the intronic region, likely without changes on protein level. Also pB-ALL-79 presented with a homozygous substitution in the intronic region.

Table 50: Genomic variants of *ID3* in pB-ALL cases

pB-ALL-42	c.[240G>C(;)300+31_32dupACC]
pB-ALL-79	c.[301-49G>C]

Unfortunately, corresponding germline material was not available for these cases to investigate the chance of germline mutations. However, remaining material from pB-ALL-79 was available to perform *MYC*-FISH analysis showing an additional third *MYC* signal in the clones (Figure 33).

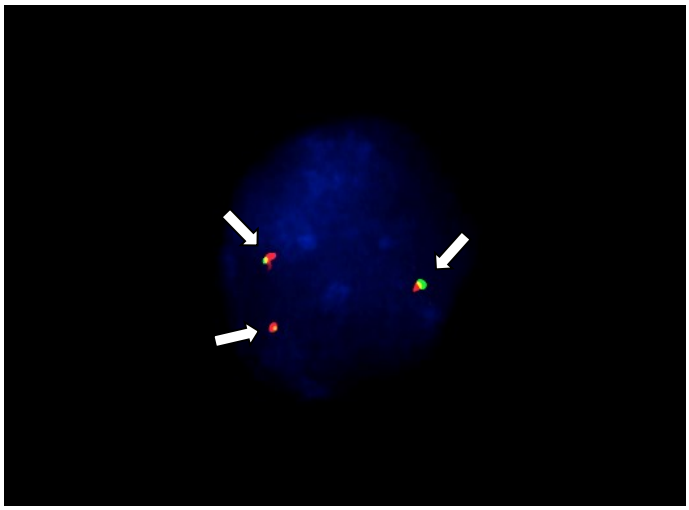


Figure 33: MYC FiSH of pB-ALL case pB-ALL-79

The image shows an interphase nucleus after hybridisation with fluorescent LSI MYC Dual Color Break Apart Rearrangement Probes (Abbott). Three MYC loci are represented by pairs of red and green labelled MYC specific probes (marked by arrows), indicating presence of an additional, third MYC locus (courtesy of Dr. A. Teigler-Schlegl, Oncogenetic Laboratory, University Gießen).

4 Discussion

4.1 Patients, samples and methods

Patients

This study aimed to assess the frequency and relevance of *ID3*, *TCF3* and *CCND3* mutations in pediatric B-NHL using a well-defined patient cohort that was uniformly diagnosed, treated and evaluated. Adequate material for tumor DNA extraction was available for 84 B-NHL patients representing the study cohort. In general the availability of tumor material for retrospective studies is limited and based on remaining samples after all diagnostic procedures. However, the number of samples analyzed in the study cohort corresponds approximately to the annual totals of pediatric B-NHL in Germany (GPOH Jahresbericht, 2013) and represents the so far largest cohort of pediatric B-NHL cases that has been analyzed and furthermore investigated for frequency and relevance of *ID3*, *TCF3* and *CCND3* mutations.

All patients enrolled in this study were previously registered in the NHL-BFM study center. One of the aims of this infrastructure is to support and assure diagnosis and therapy on highest standard of care. In each case the final diagnosis was centrally assessed in the study center after careful review of histological, cytomorphological and genetic reports from associated reference institutions. The analyzed cohort of B-NHL patients featured the characteristics of the overall cohort and might therefore be seen as representative for pediatric B-NHL patients in Germany.

Samples

Tumor samples were available from the NHL-BFM biobank where the samples had been stored at minus 80 °C. DNA from pB-ALL cases was kindly provided from the ALL-BFM study center in Kiel (A. Möricke, M. Stanulla, M. Schrappe). The current study can serve as an informative example how the infrastructure of the clinical trials of the GPOH (Gesellschaft für Pädiatrische Onkologie und Hämatologie) supports research activities.

For investigations of genomic alterations the tumor cell count in the samples used for DNA preparation is an important key point. In cases of very low tumor cell counts, mutations might be missed due to low representation of affected cells in the analyzed sample. For this reason the cytomorphological reviews of available liquid materials (effusion, ascites, bone marrow) were checked carefully and a sufficient percentage of blasts was confirmed before inclusion into this analysis. For the use of tumor tissue we checked the histology review report to confirm a sufficient tumor cell count. For a few

cases where no reference histology report was available, local pathology reports were used.

Methods

The experimental setup was tested and optimized before patient sample analysis. Capillary sequencing is a widely used and robust method for detection of genomic variants in respective DNA templates of interest. When samples showed disambiguates compared to the reference sequence they were analyzed again to confirm the finding. In all cases results were reproducible in validation experiments, giving evidence for robust functionality of the established experimental setup.

In this study, tumor DNA from pediatric B-NHL samples was analyzed for mutations in different candidate genes. A certain limitation of the study approach lies in exclusive analysis of DNA extracted from the malignant cells missing the corresponding germline material. Therefore potential non-somatic mutations might be overseen by simply assuming somatic origin of the mutations found in tumor DNA. However, by the proof of somatic acquisition of mutations in *ID3*, *TCF3* and *CCND3* by the primary studies (Love et al., 2012; Richter et al., 2012; Schmitz et al., 2012) there is strong support to expect mutations of these genes to be acquired somatically. Also, in the previous investigation of ID3 variants in 209 cases of Sjögren's syndrome, neither unknown variants nor germline mutations were reported (Sellam et al., 2008). In this context alterations detected in this study were generally considered to be somatic mutations.

In the analyzed cohort of Richter et al. there were two cases presenting with focal loss of *ID3* gene in the context of structural genomic alterations detected by single nucleotide polymorphism arrays (SNP-Array) (Richter et al., 2012). By the methods applied in the current study such genetical aberrations might not have been detected because PCR reactions only covered the genomic region of interest. Cases that presented with homozygous mutations might in fact be cases with complete loss of one allele and one remaining allele with mutation. And also for cases without mutations focal loss of *ID3* cannot be excluded. We performed detailed analyses to investigate whether biallelic *ID3* alterations differ from heterozygous alterations and could not support that hypothesis. Therefore the methodological risk to underestimate the rate of biallelic alterations seems acceptable assuming total disruption of *ID3* gene function in either case. The uncertainty whether the rate of cases without any *ID3* mutation might be overestimated and the clinical consequences from that risk must remain unsolved.

4.2 Frequency of *ID3*, *TCF3* and *CCND3* mutations in Burkitt lymphoma

Recurrent genetic alterations in BL have recently been reported by three large scale next-generation-sequencing projects, in which all studies reported *ID3* mutations to occur at high frequencies and almost exclusively in BL (Love et al., 2012; Richter et al., 2012; Schmitz et al., 2012). While these studies were the first to describe genomic alterations of *ID3* at all and equally evaluated these finding as hallmark in BL, there were striking differences regarding the *ID3* mutation frequency in the studied cohorts, ranging between 34% (Love et al.), 58% (Schmitz et al.) and 68% (Richter et al.). This observation requires a more detailed insight into the studied cohorts, especially with respect to sample inclusion criteria and patient characteristics, to further evaluate potential causes for these inconsistencies.

Schmitz and colleagues analyzed a collection of 78 sporadic BL samples obtained from pre-treatment patient biopsies and additional eleven EBV-negative BL cell lines. They performed Sanger sequencing of *ID3*, *TCF3* exon 16 and *CCND3* exon 5 and reported mutations in 58%, 11% and 38% of the cases, respectively. About one third of the analyzed samples were obtained from patients aged 18 or younger at diagnosis and the male:female ratio was 3:1. There were no information about assessment and confirmation of diagnosis (Schmitz et al., 2012).

In the study of Richter et al. an extended collection of 100 *MYC*-rearrangement positive B cell lymphomas was analyzed by Sanger sequencing, showing 42 *ID3* mutated cases. When using a previously established gene-expression classifier for selection of molecular BL (Hummel et al., 2006), 36 of 53 cases harbored *ID3* mutations (68%). Regarding the molecular BL cases 65% of the analyzed samples were from patients aged 18 or younger at the time of diagnosis and the male:female ratio was 3:1. Histologic diagnosis was previously assessed in panel review. (Richter et al., 2012)

In a third study, Love and colleagues reported *ID3* mutations in 20 cases within a cohort of 59 *MYC*-rearrangement positive BL samples (34%), including 51 initial patient biopsies and eight cell lines. The analyzed cohort included 13 pediatric patients, however, information regarding age, gender and further characteristics was not available in the majority of cases. (Love et al., 2012)

The current study presents the so far largest analysis of *ID3*, *TCF3* and *CCND3* mutation status in pediatric B-NHL and the analyzed cases were shown to feature representative characteristics for pediatric B-NHL in Germany (Table 25). *ID3* mutations had an incidence of 77% in pediatric BL (including B-AL), most supporting the data from Richter

et al., but clearly diverging from the other two studies. In the following, selected aspects that contribute to these differences are discussed in more detail.

1) Differences in the applied methods: While all primary studies used different next-generation-sequencing technologies to analyze index cases of BL, validation of *ID3* mutations and mutation frequency assessment was performed in extended validation cohorts by Sanger sequencing. Primers used for PCR product amplification varied, but equivalently covered the coding region of exon 1 including the HLH coding region. Hence, differences in the applied methods are not sufficient to explain the variability of the reported *ID3* mutation rates.
2) Inclusion criteria of samples / diagnostic quality: The inclusion criteria for evaluation in the validation cohorts varied in between the presented studies (see above) and information on diagnostic quality was limited especially in the studies of Schmitz and Love. Attempts to determine the frequency of *ID3* mutations in BL are further complicated by the diagnostic difficulty of BL. Despite enormous efforts to standardize histopathological differentiation of BL and DLBCL, this distinction is still vulnerable. Therefore careful histopathological review and description of the evaluated cases is necessary to allow the comparison of identified *ID3* mutations rates. This is of exceptional importance in our study, as the rate of *ID3* mutations in DLBCL is reported to be very low. Therefore a (un-)known mix of BL and DLBCL will not allow a robust estimate of the *ID3* mutation rate in BL. Interestingly, Love et al. restricted their analyses to cases with proven *MYC* rearrangement. The rationale for that might be the attempt to enrich the analyzed cohort for BLs and to exclude DLBCLs, as the latter more infrequently harbor a *MYC* rearrangement. Despite these efforts, the rate of *ID3* mutations in *MYC* positive lymphoma remains significantly different between the cohort reported by Love and colleagues and our cohort. Also not exactly known for the Love cohort, we expect that our cohort comprises much more pediatric B-NHL cases. Therefore the detected difference in the *ID3* mutation rate might hint at a certain impact of patient age on the presence of absence of this mutation. In consequence, these data support the hypothesis of pathogenetic and biological differences between pediatric and adult BL.
3) Biological differences according to patient characteristics: Primary studies included patients with varying distribution of age and there is strong evidence for differences in molecular presentation of BL between pediatric and adult patients. Trautmann et al. analyzed Ig repertoires and SHM status in molecularly classified BL and reported age biased differences in affected Ig heavy chain genes, with the mutation load being significantly higher in older BL patients (Trautmann et al., 2009). Furthermore,

McClure and colleagues studied morphologic and immunophenotypic features, as well as *MYC* and BCL2 status in Burkitt-like adult B-cell lymphoma and proposed a distinct biology for adult BL (McClure et al., 2005).

These observations lead to the conclusion that *ID3* mutation frequency might in fact be associated with more homogeneously presenting pediatric BL and occur less often in the more heterogeneous group of Burkitt-like adult B-cell lymphoma.

The frequency of 13% *TCF3* and 38% *CCND3* mutations were consistent with the findings of Schmitz et al., who reported frequencies of 11% and 38% mutated cases, respectively (Schmitz et al., 2012).

4.3 Functional background of *ID3*, *TCF3* and *CCND3* mutations

ID3 encodes for a small protein that functionally belongs to the family of dominant-negative HLH transcription regulators and therefore lacks a basic functional site for DNA binding. The inhibitory effect on other HLH proteins is based on the formation of heterodimers via its HLH domain, causing subsequent interference on DNA binding capabilities to the partners transcription targets (Murre, 2005; Perk et al., 2005). In B-cells isoform E47 of TCF3 is one of the partners interacting with ID3. Both, *ID3* and *TCF3* are expressed at high levels in BL (Richter et al., 2012).

The accumulation of *ID3* mutations affecting the HLH domain was reported by the NGS studies. Furthermore, biallelic involvement and the occurrence of multiple mutations were indicated (Love et al., 2012; Richter et al., 2012; Schmitz et al., 2012). Several genomic mutants have been modeled and functionally analyzed in cell proliferation assays of *ID3* mutant transfected BL cell lines and were shown to enhance G1 to S-phase transition (Love et al., 2012; Schmitz et al., 2012).

These findings are supported by the current study, showing that the majority of *ID3* mutations occurred in the crucial *ID3* HLH-coding domain (Figure 18). Furthermore we found biallelic involvement in all tested cases with more than one mutation (Table 27). Besides large genomic deletions and insertions - partly resulting in new, early ending reading frames or complete loss of whole functional helices - more than two thirds of determined mutations were single nucleotide substitutions (Figure 20). Virtually all substitutions affected the HLH domain and resulted in either nonsense or missense mutations. The only three cases (case 5, case 56, case 78) harboring silent nucleotide changes had further concomitant non-silent mutations. In conclusion, all cases that were

affected by *ID3* alterations had at least one mutation resulting in a change of ID3 amino-acid sequence.

As *ID3* seems to be the most recurrently mutated gene in *MYC* rearrangement positive B-NHL and mutations have so far not been described in other malignancies, there is good evidence for a crucial role of *ID3* in malignant transformation of these lymphomas. In this context the exceptional number and distribution of different mutations along the *ID3* gene characterizes the fragility of ID3 protein structure with regard to successful function.

The much lower frequency of *TCF3* mutations promotes an inferior role in B-NHL compared to *ID3* mutations. However, *TCF3* mutations also affected its bHLH domain (Figure 21). Most mutations (N551K, E555Q, V557A, V557E, V559M) occurred in the coding region of the basic site, which is responsible for DNA binding on TCF3 targets. Other mutations affected the first helix (D561E) and the loop region (M575K). While the latter can easily be imagined to disrupt binding capabilities with ID3 and therefore might rescue TCF3 from ID3 inhibition, the effect of mutations affecting the DNA binding domain is less self-explanatory. Schmitz and colleagues showed that even mutations affecting the DNA binding site, like TCF3(V557E), resulted in failure of TCF3/ID3 interaction while still being successful in binding to TCF3 targets. In general they reported isoform E47 to be higher expressed in *TCF3* mutated cases, indicating a gain of function for TCF3 mutants (Schmitz et al., 2012). In this regard, *ID3* and *TCF3* mutations might both cause the same effect with respect to increased activation of TCF3 transcription targets.

CCND3 was shown to be a direct target of transcriptional activation by TCF3 (Schmitz et al., 2012). Certain *CCND3* mutant proteins have proliferative advantages by increased accumulation and activation of downstream targets in BL cell lines (Schmitz et al., 2012). The most common mutation in the current study was a frame-shift causing cytosine-duplication at position c.811 (R271Pfs) that was also previously reported in other studies (Richter et al., 2012; Schmitz et al., 2012). Interestingly there were no cases with mutations affecting T283 in the cohort of Richter et al. and the current study, while Schmitz et al. reported recurrent mutations of this position in BL and DLBCL. This might indicate differences in the mutation pattern of *CCND3* with respect to certain histology, as different rates of *CCND3* mutations were reported for sporadic BL (38%), HIV-associated BL (67%) and endemic BL (1.8%) (Schmitz et al., 2012). In conclusion these results for different groups of BL need further confirmation and evaluation in well-defined cohorts.

4.4 Clinical relevance of *ID3*, *TCF3* and *CCND3* mutations in *MYC* rearrangement positive B-NHL

Differential expression of ID proteins has previously been shown in a number of malignancies (Castanon et al., 2013; Kamalian et al., 2010; Kamalian et al., 2008; Li et al., 2010; Li et al., 2012; Sharma et al., 2012; Snyder et al., 2013; Strong et al., 2013) and lately evidence was provided for association between increased ID expression levels and more disseminated disease in prostate cancer, medulloblastoma and small lung cell cancer (Gupta et al., 2007; Maw et al., 2009; Phi et al., 2013). *ID3* is also expressed at high levels in BL when compared to adult DLBCL (Richter et al., 2012), but due to its designated role as a tumor suppressor regarding TCF3 inhibition, its functional role in B-NHL might be distinct from other malignancies. Mutations of *ID3*, *TCF3* and *CCND3* were shown to be potent drivers of cell proliferation and cell survival in mutant transfected cell lines (Love et al., 2012; Schmitz et al., 2012).

First analysis of patient characteristics with respect to *ID3* mutations was performed in the validation cohort of Richter et al. and tendencies towards lower age and favorable prognosis were stated for *ID3* mutated cases (Richter et al., 2012). It is however difficult to compare pediatric and adult patients with respect to clinical characteristics and prognosis as treatment regimens are generally different and outcome is considerably worse in adult patients (Perkins and Friedberg, 2008).

In the current study the clinical relevance of *ID3*, *TCF3* and *CCND3* mutations in pediatric B-NHL was investigated using the well-defined study cohort. As mutations in the three candidate genes were virtually exclusive in *MYC* rearranged cases ($ID3^{mut}$±$TCF3^{mut}$±$CCND3^{mut}$: *MYC* rearranged 58/66 vs. *MYC* not rearranged 1/12; p value < 0.0001), relevance of the mutation status was tested within the group of 66 *MYC* rearrangement positive cases. While there was no obvious association between clinical characteristics and mutation status when investigating the candidate genes separately, certain mutational patterns in the pathway showed correlation with clinical criteria.

Of particular note was the combination of *ID3* and/or *TCF3* mutations, which were discussed to be equivalent with respect to their function earlier. These cases were associated with higher LDH levels ($ID3^{mut}$±$TCF3^{mut}$: LDH <500 U/l 18/26 vs. LDH >=500 U/l 37/40; p value 0.019), which is also reflected in an increased frequency of these cases in higher risk groups ($ID3^{mut}$±$TCF3^{mut}$: R1/R2 16/24 vs. R3/R4 37/40; p value 0.014). Also, in patients with simultaneous *ID3* and *CCND3* mutations, the frequency of BM involvement and in consequence higher stage of disease was increased (exclusive $ID3^{mut}$+$CCND3^{mut}$: BM involvement yes 8/14 vs. no 10/52; p value 0.014). These findings

may reflect advantages in proliferation and migration for respectively affected lymphoma cells, with the clinical correlate of advanced and more disseminated disease. However, biological interpretation of these results remains speculative for this time and must be evaluated in further functional analysis.

Male patients are affected by BL up to five times more often compared to females (Burkhardt et al., 2005) and so far all experimental efforts to reveal the pathogenetic mechanism in this gender-specific shift failed. The number of female patients analyzed in this study was rather low in absolute count and therefore the interpretation of gender-specific analyses must be performed carefully. While *TCF3* and *CCND3* mutations did not associate to gender directly, the frequency of cases having either *TCF3* mutation, *CCND3* mutation or both were higher in female patients (*TCF3*^{mut}±*CCND3*^{mut}: female 7/9 vs. male 22/57; p value 0.036). This trend must be confirmed in a larger series of female patients before drawing any conclusions. The provoking hypothesis of such a study could be the assumption that BL in females is related to different pathogenic events as BL in males.

Regarding *ID3* mutation hotspots it is of notice that mutations affecting c.241, resulting in Q81* nonsense mutation on protein level, accumulated in cases with B-AL (*ID3* c.241^{mut}: B-AL 4/14 vs. other diagnosis 1/52; p value 0.006) compared to stage I-IV disease. But after the inclusion of another non B-AL case in the extended cohort, the association turned under the level of statistical significance (data not shown). Further investigated *ID3* hotspots at c.190 and c.166 did not show association to any clinical criteria.

In our study cohort of only pediatric cases there was no association between *ID3* mutation status and age (Table 35). pEFS was 86±5% in *ID3*^{mut} cases compared to 93±7% in *ID3*^{wt} cases (log rank p value 0.5) and; rates for survival were *ID3*^{mut} 88±4% vs. *ID3*^{wt} 93±7%; (log rank p value 0.64). In certain sense, these findings represent a contrast to the data of Richter et al., where age-dependent *ID3* mutation frequency and superior outcome for cases with *ID3* mutations were reported. Again, one possible explanation is the more heterogeneous cohort analyzed by Richter et al. and the aforementioned general differences between BL in pediatric and adult patients.

The detailed analyses of *ID3* mutations, frequencies and patterns showed no significant difference comparing *ID3* mutations in patients with or without relapse. This became more obvious after the addition of relapsed cases to the study cohort (Figure 32). None of the investigated genes seemed to have prognostic relevance with respect to pEFS and pOS (Table 35). These observations are quite comprehensible, as the high rate of

ID3 mutations did not offer much but the chance, that *ID3* mutated cases are at a higher risk of relapse. The high number of *ID3* mutations and recurrent involvement of its partners suggests an essential role of these alterations for BL lymphomagenesis, rather than a role for disease recurrence in a small subgroup of patients.

4.5 *ID3* mutations in pB-ALL and the role of *MYC* rearrangement

Mature B cells have undergone Ig rearrangement and somatic hypermutation during B cell maturation. Initiation of malignant transformation caused by such mechanisms has been widely discussed and investigated in a number of studies: Goosens et al. provided evidence for the involvement of somatic hypermutation in the generation of chromosomal translocations including the Ig loci, like BL hallmark Ig/*MYC* (Goossens et al., 1998). However, Nepal and colleagues analyzed the contribution of somatic hypermutation-mediating AID and Ig rearrangement-initiating RAGs to lymphomagenesis in *MYC* transgenic mice and did not find any direct effect between AID/RAG expression and subsequent lymphoma development (Nepal et al., 2008). In contrast, Pasqualluci et al. suggested somatic hypermutation to be the major contributor for aberrant translocations and mutations outside the intended locations in lymphomagenesis of DLBCL (Pasqualucci et al., 2001). Regarding *ID3* mutations, Richter et al. showed enrichment of mutations in the RGYW-motif that is favorably affected by the AID (Richter et al., 2012).

While the actual roles of AID and RAG remain unclear with respect to malignant transformation and in particular generation of *ID3* mutations, the analysis of pB-ALL cases in this study was thought to investigate the occurrence of *ID3* mutations in a related malignancy, whose cells are pre-germinal centre cells and have not undergone somatic hypermutation. Assuming that the mutational mechanisms resulting in *ID3* mutations are based on somatic hypermutation, any such mutations are unexpected in pB-ALL. And in fact, the vast majority of pB-ALL cases did not show any *ID3* mutation and this finding therefore serves as next indication for a link between somatic hypermutation and *ID3* affection. However, in the cohort of 96 pB-ALL cases, two harbored genomic alterations in *ID3*. With regard to the effect on protein level, both cases presented with either silent or intronic alterations. Several assumptions and hypothesis might be derived from these findings:

a) The two pB-ALL cases harboring *ID3* variants showed genomic variants that are also present in germline.

b) The two cases might represent transitional malignancies between immature and mature B cell lymphoma. Cases of *MYC* rearrangement positive pB-ALL are reported at a small number (Greer et al., 2003; Hirzel et al., 2013; Meeker et al., 2011).
c) The underlying mechanism responsible for mutagenesis in B-NHL was also activated in these cases, but did not cause protein changing mutations and therefore remained phenotypically unobtrusive.
d) Non-silent *ID3* mutations require the environment of *MYC* activation, otherwise the *ID3* mutated cells have no survival benefit and die.

Interpretation of this finding is speculative and therefore demands further investigation of *ID3* mutated pB-ALL cases. However, for one of the affected pB-ALL cases (pB-ALL-79), it was possible to perform *MYC*-FISH analysis. While not observing any classical *MYC* rearrangement, the case presented with an additional signal of *MYC* in a lower number of cells. The third *MYC* signal indicates a third copy of the *MYC* locus and represents an interesting finding with respect to the concurrent finding of a genomic variant in *ID3*. Interestingly there was a similar case in the B-NHL cohort which had an *ID3* mutation but was reported to have no *MYC* rearrangement but a third *MYC* signal copy (case 68). These findings raise the question whether additional *MYC* copies similar to *MYC* rearrangements support the occurrence of *ID3* mutations. A future analysis of selected pB-ALL cases and other malignancies with additional *MYC* copies might shed light into mutual influence and dependency between *ID3* mutations and *MYC* alterations.

4.6 Consequence of *ID3*, *TCF3* and *CCND3* mutations in pediatric BL

Burkitt lymphoma has been shown to be a homogenous malignancy in gene expression profiles, especially in comparison to the related group of DLBCL (Hummel et al., 2006; Lenze et al., 2011). Lately, results from whole-genome-sequencing of thirteen pediatric BL of the NHL-BFM group supported these observations on genomic level, showing a median of only 28 somatic mutations per tumor and a high frequency of recurrently affected genes even in the small number of twelve cases (Table 20, Table 21) (Rohde et al., 2013).

In the current study, 88% of the analyzed *MYC* rearrangement positive cases had mutations in at least one of the three investigated candidate genes, representing affection of the *ID3*-*TCF3*-*CCND3* pathway in the vast majority of pediatric BL cases (Figure 34).

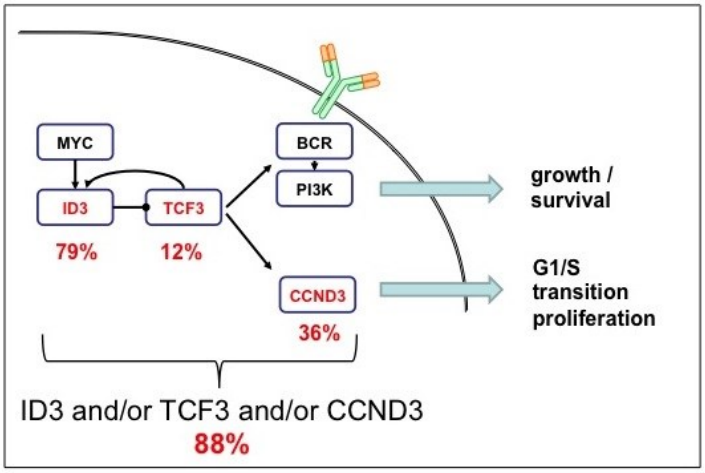


Figure 34: ID3, TCF3, CCND3 pathway with frequency of respective mutations in MYC rearrangement positive B-NHL.
Modified from Schmitz and Campo (Campo, 2012; Schmitz et al., 2012).

In conclusion with the homogenous genomic landscape in pediatric BL and the so far exclusive finding of *ID3* mutations in BL, these finding stress the evident relevance of this pathway especially in pediatric cases. The absence of clear associations to clinical characteristics and prognosis may even imply an essential function of pathway-disruption for BL lymphomagenesis. Cases presenting without mutations might still be affected by focal loss of *ID3* or mutations in other functional partners that are involved up- or downstream within the same pathway, likely in the BCR pathway, PI3K, CDK4/6 and all their regulating genes. However, additional candidates will less likely present at similar high frequencies, as NGS studies so far should have covered most of the highly recurrent genomic events in BL. Investigation of more BL cases by NGS will be helpful to also cover mutations in genes that are distributed around the candidates analyzed in this study - that might be as relevant for lymphomagenesis - but only occur at much lower frequencies.

These results demand a further functional evaluation of this pathway in many respects. Functional dependency on *MYC* alterations has to be investigated to better understand the causes for malignant transformation and uncover underlying mutational mechanisms. Furthermore, the overall high number of affected cases asks for therapeutic-targeting of this pathway. There is first promising evidence for successful application and efficacy of orally available CDK4/6 inhibitor PD 0332991, shown for tumor mass reduction in a BL-mouse model by Schmitz et al. (Schmitz et al., 2012). CDK4/6 inhibitors have recently also been shown to be effective in renal cell carcinoma cell lines and breast cancer cell line (Finn et al., 2009; Logan et al., 2013) and are in preparation for clinical phase-I and II studies in breast cancer patients (U.S. National

Institutes of Health, 2013). Further investigation of this pathway will shed more light on molecular processes in BL and hopefully reveal more specific therapeutic options.

4.7 Conclusion of the study

In the current study, frequency and relevance of *ID3*, *TCF3* and *CCND3* mutations were studied in a uniformly diagnosed and treated cohort of pediatric B-NHL patients. The analyzed cohort had representative characteristics for pediatric B-NHL patients in Germany with respect to diagnosis, clinical features and outcome. *ID3* mutations were present in 77% of BL/B-AL cases, *TCF3* and *CCND3* showed mutations in 13% and 37% of the cases, respectively. As a result, we demonstrated that the occurrence of mutations in any of the analyzed genes was positively correlated with *MYC* rearrangement and 88% of *MYC* rearrangement positive cases had at least one mutation in one of the investigated genes. Regarding the clinical characteristics in MYC rearrangement positive B-NHL, there was no relevant association of the *ID3*, *TCF3* and *CCND3* mutation status, with gender, age, stage of disease, BM involvement, CNS involvement, and LDH levels. Outcome was not associated with the mutation status. However, there was a tendency towards advanced stages and more disseminated disease in cases with certain concurrent mutations that demands further investigation of the respective patient groups.

In the context of relatively homogeneous genomic alterations in pediatric Burkitt lymphoma, the high number of *ID3* mutations found in this study of pediatric B-NHL patients suggests an essential role for this pathway with respect to lymphomagenesis and the phenotype of Burkitt lymphoma.

5 Summary (English/German)

B-cell Non-Hodgkin Lymphoma (B-NHL) is the most common type of Non-Hodgkin Lymphoma in childhood and adolescent cancer patients. B-NHL can be further classified into subtypes, with Burkitt lymphoma (BL) being the most common entity in pediatric patients. Recently published large-scale next-generation sequencing studies unveiled sets of recurrently mutated genes in tumor cells of pediatric and adult B-NHL patients and introduced functionally related Inhibitor of DNA 3 (*ID3)*, Transcription Factor 3 (*TCF3)* and Cyclin D3 (*CCND3)* as potential drivers of BL lymphomagenesis. However, validation of these findings showed inconsistent mutation rates and assessment of clinical relevance was limited.

In the present study frequency and relevance of mutations in *ID3*, *TCF3* and *CCND3* were analyzed within a well-defined cohort of 84 uniformly diagnosed and treated pediatric B-NHL patients. Mutation frequency was 77% (*ID3*), 13% (*TCF3*) and 37% (*CCND3*) in BL (including Burkitt leukemia) and mutations remained almost exclusive for MYC rearrangement positive cases. *ID3* mutations were detected in remarkably higher frequency than previously published. There was no clear association between mutation status and outcome, but certain concurrent mutations where enriched in more advanced stages of the disease.

As a control group of a related malignancy, 96 samples from precursor B-cell leukemia (pB-ALL) patients were also analyzed for *ID3* mutations. As expected, no mutations were found, but two cases showed genomic variants in *ID3*. Interestingly, for one of these cases it was possible to show a genetic alteration involving *MYC*, which is usually a key feature of BL.

We conclude, that almost 90% of *MYC* rearrangement positive B-NHL harbored mutations in at least one of the investigated genes and therefore this study promotes the corresponding pathway to play an essential role in BL and especially in pediatric cases.

Zusammenfassung

B-Zell Non-Hodgkin Lymphome (B-NHL) sind der häufigste Typ von Non-Hodgkin Lymphomen bei Kindern und Jugendlichen mit malignen Krebserkrankungen. B-NHL können weiter in Subtypen klassifiziert werden, in denen das Burkitt Lymphom (BL) die häufigste Entität bei Kindern und Jugendlichen darstellt. In der jüngsten Zeit haben Ganzgenomsequenzierungen von Tumorproben eine Liste von besonders häufig mutierten Genen in B-NHL von pädiatrischen und erwachsenen Patienten gezeigt. Unter anderem wurden die funktionell zusammenhängenden Kandidaten *ID3*, *TCF3* und *CCND3* als mögliche Antreiber der Lymphompathogenese bei BL beschrieben. Allerdings zeigten sich in verschiedenen Studien inkonsistente Mutationsfrequenzen und die Untersuchung der klinischen Bedeutung war limitiert.

In dieser Studie wurden die Frequenz und Relevanz von Mutationen in den drei Kandidatengenen *ID3*, *TCF3* und *CCND3* in einer klar definierten Kohorte von 84 einheitlich diagnostizierten und behandelten pädiatrischen Fällen mit B-NHL untersucht. In BL (inklusive Burkitt Leukämie) waren die Mutationsfrequenzen 77% (*ID3*), 13% (*TCF3*) und 37% (*CCND3*) und in Bezug auf die Gesamtkohorte kamen Mutationen in den untersuchten Kandidatengenen fast ausschließlich in Fällen mit gleichzeitiger *MYC*-Translokation vor. Die Frequenz der *ID3* Mutationen war unerwartet höher als bisher publiziert. Es zeigte sich kein klarer Zusammenhang zwischen dem Mutationsstatus und der Prognose. Es ergaben sich jedoch Hinweise für einen Zusammenhang von bestimmten Mutationskombinationen mit höheren Krankheitsstadien.

In einer Kontrollgruppe von 96 pädiatrischen Fällen mit einer akuten Vorläufer-B-Zell lymphoblastischen Leukämie ließen sich keine *ID3* Mutationen feststellen, sondern lediglich zwei Fälle mit nichtkodierenden Varianten. Interessanterweise gelang es, für einen dieser Fälle eine genetische Aberration des *MYC*-Gens nachzuweisen, dessen Translokation eigentlich die typische molekulare Veränderung für BL ist.

Zusammenfassend lässt sich festhalten, dass nahezu 90% der untersuchten *MYC*-rearrangierten pädiatrischen B-NHL eine Mutation in mindestens einem der drei untersuchten Kandidatengene aufwiesen und damit unterstreicht diese Arbeit die besondere Rolle der beschriebenen Gene, insbesondere in pädiatrischen Burkitt Lymphomen.

6 List of abbreviations

A/T	Adenine/Thymine
ACB	Actibated B cell
AID	Activation-Induced cytidine Deaminase (protein)
ALL-BFM	Acute lymphoblastic leukemia - Berlin Frankfurt Münster study group
ARF	CDKN2A gene, alternate reading frame (gene)
B-AL	Burkitt Leukemia
BAX	bcl-2-like protein 4 (gene)
BCL2	B-cell lymphoma 2 (gene)
BCL6	B-cell lymphoma 6 (gene)
B-CLL	B-cell chronic lymphocytic leukemia
BCR	B cell receptor
bHLH	basic-Helix-loop-helix
BL	Burkitt Lymphoma
BM	bone marrow
B-NHL	B-cell Non-Hodgkin lymhoma
B-NHL (DD)	B-cell Non-Hodgkin lymphoma, unclassifiable between DLBCL and BL
bp	Base-pairs
CCD	Charge coupled device
CCG	Children's cancer group
CCND3	G1/S-specific cylin-D3 (gene)
CD	Cluster of Differentiation
CDK4/6	Cyclin-dependent kinase 4/6 (protein)
CDKN2A	Cyclin-dependent kinase inhibitor 2A (gene)
c-MYC	cellular - myelocytomatosis viral oncogene
CNS	Central nervous system
CSR	Class Swichting Recombination
CXCL13	B lymphocyte chemoattractant (protein)
CXCR5	Receptor
Cyclin D3	G1/S-specific cyclin-D3 (protein)
dATP	Deoxyadenosine triphosphate

dbSNP	Single Nucleotide Polymorphism Database
ddH2O	double distilled Water
ddNTPs	Dideoxynucleotides
DDX3X	ATP-depentdent RNA helicase DDX3X (gene)
DKFZ	Deutsches Krebsforschungszentrum
DLBCL	Diffuse Large B cell Lymphoma
DLBCL-CB	Diffuse large B-cell lymphoma - centroblastic variant
DNA	Deoxyribonucleic acid
dNTP	2-Desoxyribonucleosid–5-triphosphate
dNTPs	Deoxyribonucleotides
E.coli	Escherichia coli
e.g.	exempli gratia
E12	Isoform of TCF3
E47	Isoform of TCF3
EBF1	Early B cell Factor 1 (gene)
EBV	Epstein-Bar-Virus
EDTA	Ethelenediaminetetraacetic acid
FAB	French-American-British classification
FBXO11	F-box only protein 11 (gene)
FISH	Fluorescence in situ hybridization
g	Gram
G/C	Guanine/Cytosine
GC	germinal centers
GCB	Germinal Center B cell
GPOH	Gesellschaft für Pädiatrische Onkologie und Hämatologie
H2O	Water
H-chain	Heavy chain
HIV	Human immunodeficiency virus
HLH	Helix-loop-helix
HSC	Hematopoetic stemm cell
ICGC	International Cancer Genome Consortium
ICGC-MMML-Seq	International Cancer Genome Consortium - Molecular Mechanisms of Malignant Lymphoma by Sequencing
ID1	Inhibitor of DNA 1

ID2	Inhibitor of DNA 2
ID3	Inhibitor of DNA 3
Ig	Immunoglobulin
Il	Interleukin
InDel	Deletion followed by an Insertion
IRF4	Interferon regulatory factor 4 (gene)
kb	Kilo base-pairs
L-chain	Light chain
LDH	Lacate dehydrogenase
LR	Log rank test
M13	M13 primer
MDM-2	Mouse double minute 2 homolog (gene)
mut	mutated
MYC	see c-MYC
NCBI	National Center for Biotechnology Information
NGS	Next-generation sequencing
NHL	Non-Hodgkin Lymphoma
NHL-BFM	Non-Hodgkin lymphoma - Berlin Frankfurt Münster study group
OD	Optical density
p value	probability value
p21	p21/WAF1 (cyclin-dependent kinase inhibitor 1) (protein)
p27	cyclin-dependent kinase inhibitor 1B (protein)
P2RY8	P2Y purinoceptor 8 (gene)
p53	protein 53 (protein)
PAX5	Paired Box 5 (gene)
pB-ALL	Precursor B-cell acute lymphoblastic leukemia
PBX1	Pre-B-cell leukemia transcription factor 1 (gene)
PCR	Polymerase chain reaction
pEFS	probability of event free survival
PI3K	Phosphatidylinositide 3-kinases
PMBL	Primary mediastinal B-cell lymphoma
pOS	probability of overall survival
pre-B	pre-B cell

pre-BCR	pre-B cell receptor
pro-B	pro-B cell
R	Risk group
Rag-1/-2	Recombination-activating gene 1/2 (genes)
RB1	retinoblastoma 1 (gene)
Rb1	Retinoblastoma protein
RNA	Ribonucleic acid
SHM	Somatic Hypermutation
SMARCA4	ATP-dependent helicase SMARCA4 (gene)
SNP	Single Nucleotide Polymorphism
SNP-Array	Single Nucleotide Polymorphism-Array
SNV	Single Nucleotide Variation
TAE	Tris base, acetic acid, EDTA buffer
TCF3	Transcription Factor 3 (gene)
TdT	Terminal deoxynucleotidyl transferas
TP53	tumor protein 53 (gene)
TP73	tumor protein 73 (gene)
U	Units
UV	Ultraviolett (light)
V	Volt
VDJ	Variable, Diversity, Joinging segments
WHO	World Health Organization
wt	wildtype
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

7 List of figures

8 List of tables

9 References

Abbas, A. K., A. H. Lichtman and S. Pillai (2012). Cellular and molecular immunology. Philadelphia, Elsevier/Saunders.

Arnold, J. M., S. C. Mok, D. Purdie and G. Chenevix-Trench (2001). "Decreased expression of the Id3 gene at 1p36.1 in ovarian adenocarcinomas." Br J Cancer 84(3): 352-359.

Asirvatham, A. J., J. P. Carey and J. Chaudhary (2007). "ID1-, ID2-, and ID3-regulated gene expression in E2A positive or negative prostate cancer cells." Prostate 67(13): 1411-1420.

Becker-Herman, S., F. Lantner and I. Shachar (2002). "Id2 negatively regulates B cell differentiation in the spleen." J Immunol 168(11): 5507-5513.

Benezra, R., R. L. Davis, A. Lassar, S. Tapscott, M. Thayer, D. Lockshon and H. Weintraub (1990). "Id: a negative regulator of helix-loop-helix DNA binding proteins. Control of terminal myogenic differentiation." Ann N Y Acad Sci 599: 1-11.

Benezra, R., R. L. Davis, D. Lockshon, D. L. Turner and H. Weintraub (1990). "The protein Id: a negative regulator of helix-loop-helix DNA binding proteins." Cell 61(1): 49-59.

Bennett, J. M., D. Catovsky, M. T. Daniel, G. Flandrin, D. A. Galton, H. R. Gralnick and C. Sultan (1976). "Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group." Br J Haematol 33(4): 451-458.

Bhatia, K. G., M. I. Gutierrez, K. Huppi, D. Siwarski and I. T. Magrath (1992). "The pattern of p53 mutations in Burkitt's lymphoma differs from that of solid tumors." Cancer Res 52(15): 4273-4276.

Boxer, L. M. and C. V. Dang (2001). "Translocations involving c-myc and c-myc function." Oncogene 20(40): 5595-5610.

Burkhardt, B., I. Oschlies, W. Klapper, M. Zimmermann, W. Woessmann, A. Meinhardt, E. Landmann, A. Attarbaschi, F. Niggli, M. Schrappe and A. Reiter (2011). "Non-Hodgkin's lymphoma in adolescents: experiences in 378 adolescent NHL patients treated according to pediatric NHL-BFM protocols." Leukemia 25(1): 153-160.

Burkhardt, B., M. Zimmermann, I. Oschlies, F. Niggli, G. Mann, R. Parwaresch, H. Riehm, M. Schrappe and A. Reiter (2005). "The impact of age and gender on biology, clinical features and treatment outcome of non-Hodgkin lymphoma in childhood and adolescence." Br J Haematol 131(1): 39-49.

Burkitt, D. (1958). "A sarcoma involving the jaws in african children." British Journal of Surgery 46(197): 218-223.

Burmeister, T., N. Gokbuget, S. Schwartz, L. Fischer, D. Hubert, A. Sindram, D. Hoelzer and E. Thiel (2010). "Clinical features and prognostic implications of TCF3-PBX1 and ETV6-RUNX1 in adult acute lymphoblastic leukemia." Haematologica 95(2): 241-246.

Campo, E. (2012). "New pathogenic mechanisms in Burkitt lymphoma." Nat Genet 44(12): 1288-1289.

Capello, D., A. Carbone, C. Pastore, A. Gloghini, G. Saglio and G. Gaidano (1997). "Point mutations of the BCL-6 gene in Burkitt's lymphoma." Br J Haematol 99(1): 168-170.

Capello, D., U. Vitolo, L. Pasqualucci, S. Quattrone, G. Migliaretti, L. Fassone, C. Ariatti, D. Vivenza, A. Gloghini, C. Pastore, C. Lanza, J. Nomdedeu, B. Botto, R. Freilone, D. Buonaiuto, V. Zagonel, E. Gallo, G. Palestro, G. Saglio, R. Dalla-Favera, A. Carbone and G. Gaidano (2000). "Distribution and pattern of BCL-6 mutations throughout the spectrum of B-cell neoplasia." Blood 95(2): 651-659.

Castanon, E., J. Bosch-Barrera, I. Lopez, V. Collado, M. Moreno, J. M. Lopez-Picazo, L. Arbea, M. D. Lozano, A. Calvo and I. Gil-Bazo (2013). "Id1 and Id3 co-expression correlates with clinical outcome in stage III-N2 non-small cell lung cancer patients treated with definitive chemoradiotherapy." J Transl Med 11: 13.

Casula, M., P. A. Ascierto, A. Cossu, M. C. Sini, S. Tore, M. Colombino, M. P. Satta, A. Manca, C. Rozzo, S. M. Satriano, G. Castello, A. Lissia, F. Tanda and G. Palmieri (2003). "Mutation analysis of candidate genes in melanoma-prone families: evidence of different pathogenetic mechanisms at chromosome 9P21." Melanoma Res 13(6): 571-579.

Cato, M. H., S. K. Chintalapati, I. W. Yau, S. A. Omori and R. C. Rickert (2011). "Cyclin D3 is selectively required for proliferative expansion of germinal center B cells." Mol Cell Biol 31(1): 127-137.

Cinti, C., P. P. Claudio, C. M. Howard, L. M. Neri, Y. Fu, L. Leoncini, G. M. Tosi, N. M. Maraldi and A. Giordano (2000). "Genetic alterations disrupting the nuclear localization of the retinoblastoma-related gene RB2/p130 in human tumor cell lines and primary tumors." Cancer Res 60(2): 383-389.

Cobaleda, C., A. Schebesta, A. Delogu and M. Busslinger (2007). "Pax5: the guardian of B cell identity and function." Nat Immunol 8(5): 463-470.

Corn, P. G., S. J. Kuerbitz, M. M. van Noesel, M. Esteller, N. Compitello, S. B. Baylin and J. G. Herman (1999). "Transcriptional silencing of the p73 gene in acute lymphoblastic leukemia and Burkitt's lymphoma is associated with 5' CpG island methylation." Cancer Res 59(14): 3352-3356.

Dang, C. V. (1999). "c-Myc target genes involved in cell growth, apoptosis, and metabolism." Mol Cell Biol 19(1): 1-11.

Dang, C. V., L. M. Resar, E. Emison, S. Kim, Q. Li, J. E. Prescott, D. Wonsey and K. Zeller (1999). "Function of the c-Myc oncogenic transcription factor." Exp Cell Res 253(1): 63-77.

de Jong, D. (2009). "Novel lymphoid neoplasms--the borderland between diffuse large B-cell lymphoma and Burkitt's lymphoma." Haematologica 94(7): 894-896.

Deffenbacher, K. E., J. Iqbal, W. Sanger, Y. Shen, C. Lachel, Z. Liu, Y. Liu, M. S. Lim, S. L. Perkins, K. Fu, L. Smith, J. Lynch, L. M. Staudt, L. M. Rimsza, E. Jaffe, A. Rosenwald, G. K. Ott, J. Delabie, E. Campo, R. D. Gascoyne, M. S. Cairo, D. D. Weisenburger, T. C. Greiner, T. G. Gross and W. C. Chan (2012). "Molecular distinctions between pediatric and adult mature B-cell non-Hodgkin lymphomas identified through genomic profiling." Blood 119(16): 3757-3766.

den Dunnen, J. T. and S. E. Antonarakis (2000). "Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion." Hum Mutat 15(1): 7-12.

Farrell, P. J., G. J. Allan, F. Shanahan, K. H. Vousden and T. Crook (1991). "p53 is frequently mutated in Burkitt's lymphoma cell lines." EMBO J 10(10): 2879-2887.

Finn, R. S., J. Dering, D. Conklin, O. Kalous, D. J. Cohen, A. J. Desai, C. Ginther, M. Atefi, I. Chen, C. Fowst, G. Los and D. J. Slamon (2009). "PD 0332991, a selective cyclin D kinase 4/6 inhibitor, preferentially inhibits proliferation of luminal estrogen receptor-positive human breast cancer cell lines in vitro." Breast Cancer Res 11(5): R77.

Fisher, R. A. (1922). "On the Interpretation of χ2 from Contingency Tables, and the Calculation of P." Journal of the Royal Statistical Society 85(1): 87-94.

Förster, R., A. E. Mattis, E. Kremmer, E. Wolf, G. Brem and M. Lipp (1996). "A Putative Chemokine Receptor, BLR1, Directs B Cell Migration to Defined Lymphoid Organs and Specific Anatomic Compartments of the Spleen." Cell 87(6): 1037-1047.

Fuleihan, R., N. Ramesh, R. Loh, H. Jabara, R. S. Rosen, T. Chatila, S. M. Fu, I. Stamenkovic and R. S. Geha (1993). "Defective expression of the CD40 ligand in X chromosome-linked immunoglobulin deficiency with normal or elevated IgM." Proc Natl Acad Sci U S A 90(6): 2170-2173.

Geneious Software. (2012). "Geneious Pro trial 5.6.5 created by Biomatters. Available from http://www.geneious.com/."

Glick, B., T. S. Chang and R. G. Jaap (1955). "The Bursa of Fabricius and Antibody Production." Poult. Sci.

Goossens, T., U. Klein and R. Kuppers (1998). "Frequent occurrence of deletions and duplications during somatic hypermutation: implications for oncogene translocations and heavy chain disease." Proc Natl Acad Sci U S A 95(5): 2463-2468.

GPOH Jahresbericht. (2013). "Jahresbericht 2010 - Übersichtsdarstellungen." Retrieved 18.02.2013, 2013, from http://www.kinderkrebsregister.de/extern/veroeffentlichungen/jahresberichte/jb2010.

Greer, W. L., C. L. Lee, M. B. Callanan, E. Zayed and I. Sadek (2003). "Case of acute lymphoblastic leukemia presenting with t(14;18)/BCL2, t(8;14)/cMYC, and t(1;2)/FCGR2B." Am J Hematol 74(2): 112-118.

Gualco, G., L. M. Weiss, W. J. Harrington, Jr. and C. E. Bacchi (2009). "Nodal diffuse large B-cell lymphomas in children and adolescents: immunohistochemical expression patterns and c-MYC translocation in relation to clinical outcome." Am J Surg Pathol 33(12): 1815-1822.

Gupta, G. P., J. Perk, S. Acharyya, P. de Candia, V. Mittal, K. Todorova-Manova, W. L. Gerald, E. Brogi, R. Benezra and J. Massague (2007). "ID genes mediate tumor reinitiation during breast cancer lung metastasis." Proc Natl Acad Sci U S A 104(49): 19506-19511.

Gutierrez, M. I., B. Cherney, A. Hussain, H. Mostowski, G. Tosato, I. Magrath and K. Bhatia (1999). "Bax is frequently compromised in Burkitt's lymphomas with irreversible resistance to Fas-induced apoptosis." Cancer Res 59(3): 696-703.

Hirzel, A. C., A. Cotrell, R. Gasparini and V. Sriganeshan (2013). "Precursor B-Cell Acute Lymphoblastic Leukemia/Lymphoma with L3 Morphology, Philadelphia Chromosome, MYC Gene Translocation, and Coexpression of TdT and Surface Light Chains: A Case Report." Case Rep Pathol 2013: 679892.

Hsu, L. Y., J. Lauring, H. E. Liang, S. Greenbaum, D. Cado, Y. Zhuang and M. S. Schlissel (2003). "A conserved transcriptional enhancer regulates RAG gene expression in developing B cells." Immunity 19(1): 105-117.

Hummel, M., S. Bentink, H. Berger, W. Klapper, S. Wessendorf, T. F. Barth, H. W. Bernd, S. B. Cogliatti, J. Dierlamm, A. C. Feller, M. L. Hansmann, E. Haralambieva, L. Harder, D. Hasenclever, M. Kuhn, D. Lenze, P. Lichter, J. I. Martin-Subero, P. Moller, H. K. Muller-Hermelink, G. Ott, R. M. Parwaresch, C. Pott, A. Rosenwald, M. Rosolowski, C. Schwaenen, B. Sturzenhofecker, M. Szczepanowski, H. Trautmann, H. H. Wacker, R. Spang, M. Loeffler, L. Trumper, H. Stein and R. Siebert (2006). "A biologic definition of Burkitt's lymphoma from transcriptional and genomic profiling." N Engl J Med 354(23): 2419-2430.

ICGC Project. (2013). "ICGC Data Portal." http://dcc.icgc.org, from http://dcc.icgc.org.

Kalungi, S., H. Wabinga and L. Bostad (2011). "The RB (pRb2/p16) and p53 (p14/p53/p21) tumor-suppressor pathways in endemic Burkitt lymphoma." J Pediatr Hematol Oncol 33(2): e54-59.

Kamalian, L., S. S. Forootan, Z. Z. Bao, Y. Zhang, J. R. Gosney, C. S. Foster and Y. Ke (2010). "Inhibition of tumourigenicity of small cell lung cancer cells by suppressing Id3 expression." Int J Oncol 37(3): 595-603.

Kamalian, L., J. R. Gosney, S. S. Forootan, C. S. Foster, Z. Z. Bao, C. Beesley and Y. Ke (2008). "Increased expression of Id family proteins in small cell lung cancer and its prognostic significance." Clin Cancer Res 14(8): 2318-2325.

Kaplan, E. L. and P. Meier (1958). "Nonparametric Estimation from Incomplete Observations." Journal of the American Statistical Association 53(282): 457-481.

Klapper, W., M. Kreuz, C. W. Kohler, B. Burkhardt, M. Szczepanowski, I. Salaverria, M. Hummel, M. Loeffler, S. Pellissery, W. Woessmann, C. Schwanen, L. Trumper, S. Wessendorf, R. Spang, D. Hasenclever and R. Siebert (2012). "Patient age at diagnosis is associated with the molecular characteristics of diffuse large B-cell lymphoma." Blood 119(8): 1882-1887.

Kuppers, R. (2005). "Mechanisms of B-cell lymphoma pathogenesis." Nat Rev Cancer 5(4): 251-262.

Legler, D. F., M. Loetscher, R. S. Roos, I. Clark-Lewis, M. Baggiolini and B. Moser (1998). "B Cell–attracting Chemokine 1, a Human CXC Chemokine Expressed in Lymphoid Tissues, Selectively Attracts B Lymphocytes via BLR1/CXCR5." The Journal of Experimental Medicine 187(4): 655-660.

Lenze, D., L. Leoncini, M. Hummel, S. Volinia, C. G. Liu, T. Amato, G. De Falco, J. Githanga, H. Horn, J. Nyagol, G. Ott, J. Palatini, M. Pfreundschuh, E. Rogena, A. Rosenwald, R. Siebert, C. M. Croce and H. Stein (2011). "The different epidemiologic subtypes of Burkitt lymphoma share a homogenous micro RNA profile distinct from diffuse large B-cell lymphoma." Leukemia 25(12): 1869-1876.

Li, J., T. Maruyama, P. Zhang, J. E. Konkel, V. Hoffman, B. Zamarron and W. Chen (2010). "Mutation of inhibitory helix-loop-helix protein Id3 causes gammadelta T-cell lymphoma in mice." Blood 116(25): 5615-5621.

Li, X. J., C. D. Zhu, W. Yu, P. Wang, F. F. Chen, X. Y. Xia and B. Luo (2012). "Overexpression of Id3 induces apoptosis of A549 human lung adenocarcinoma cells." Cell Prolif 45(1): 1-8.

Logan, J. E., N. Mostofizadeh, A. J. Desai, V. O. N. E. E, D. Conklin, V. Konkankit, H. Hamidi, M. Eckardt, L. Anderson, H. W. Chen, C. Ginther, E. Taschereau, P. H. Bui, J. G. Christensen, A. S. Belldegrun, D. J. Slamon and F. F. Kabbinavar (2013). "PD-0332991, a Potent and Selective Inhibitor of Cyclin-dependent Kinase 4/6, Demonstrates Inhibition of Proliferation in Renal Cell Carcinoma at Nanomolar Concentrations and Molecular Markers Predict for Sensitivity." Anticancer Res 33(8): 2997-3004.

Love, C., Z. Sun, D. Jima, G. Li, J. Zhang, R. Miles, K. L. Richards, C. H. Dunphy, W. W. Choi, G. Srivastava, P. L. Lugar, D. A. Rizzieri, A. S. Lagoo, L. Bernal-Mizrachi, K. P. Mann, C. R. Flowers, K. N. Naresh, A. M. Evens, A. Chadburn, L. I. Gordon, M. B. Czader, J. I. Gill, E. D. Hsi, A. Greenough, A. B. Moffitt, M. McKinney, A. Banerjee, V. Grubor, S. Levy, D. B. Dunson and S. S. Dave (2012). "The genetic landscape of mutations in Burkitt lymphoma." Nat Genet, 10.1038/ng.2468.

Lu, B., C. Zhou, W. Yang, H. Huang, Z. Gao, Y. He, H. Liu, X. Zhou and L. Gong (2011). "Morphological, immunophenotypic and molecular characterization of mature aggressive B-cell lymphomas in Chinese pediatric patients." Leuk Lymphoma 52(12): 2356-2364.

Manis, J. P., M. Tian and F. W. Alt (2002). "Mechanism and control of class-switch recombination." Trends Immunol 23(1): 31-39.

Mantel, N. (1966). "Evaluation of survival data and two new rank order statistics arising in its consideration." Cancer Chemother Rep 50(3): 163-170.

Marnerides, A., T. P. Vassilakopoulos, E. Boltetsou, G. Levidou, M. K. Angelopoulou, I. Thymara, M. C. Kyrtsonis, V. Pappi, O. Tsopra, P. Panayiotidis, G. A. Pangalis, P. Beris, E. Patsouris and P. Korkolopoulou (2011). "Immunohistochemical expression and prognostic significance of CCND3, MCM2 and MCM7 in Hodgkin lymhoma." Anticancer Res 31(10): 3585-3594.

Martinez-Delgado, B., B. Melendez, M. Cuadros, M. Jose Garcia, J. Nomdedeu, C. Rivas, J. Fernandez-Piqueras and J. Benitez (2002). "Frequent inactivation of the p73 gene by abnormal methylation or LOH in non-Hodgkin's lymphomas." Int J Cancer 102(1): 15-19.

Maw, M. K., J. Fujimoto and T. Tamaya (2009). "Overexpression of inhibitor of DNA-binding (ID)-1 protein related to angiogenesis in tumor advancement of ovarian cancers." BMC Cancer 9: 430.

McClure, R. F., E. D. Remstein, W. R. Macon, G. W. Dewald, T. M. Habermann, A. Hoering and P. J. Kurtin (2005). "Adult B-cell lymphomas with burkitt-like morphology are phenotypically and genotypically heterogeneous with aggressive clinical behavior." Am J Surg Pathol 29(12): 1652-1660.

Meeker, N. D., A. M. Cherry, C. D. Bangs and J. K. Frazer (2011). "A pediatric B lineage leukemia with coincident MYC and MLL translocations." J Pediatr Hematol Oncol 33(2): 158-160.

Mern, D. S., J. Hasskarl and B. Burwinkel (2010). "Inhibition of Id proteins by a peptide aptamer induces cell-cycle arrest and apoptosis in ovarian cancer cells." Br J Cancer 103(8): 1237-1244.

Mern, D. S., K. Hoppe-Seyler, F. Hoppe-Seyler, J. Hasskarl and B. Burwinkel (2010). "Targeting Id1 and Id3 by a specific peptide aptamer induces E-box promoter activity, cell cycle arrest, and apoptosis in breast cancer cells." Breast Cancer Res Treat 124(3): 623-633.

Miles, R. R., M. Raphael, K. McCarthy, A. Wotherspoon, M. A. Lones, M. J. Terrier-Lacombe, C. Patte, M. Gerrard, A. Auperin, R. Sposto, V. Davenport, M. S. Cairo and S. L. Perkins (2008). "Pediatric diffuse large B-cell lymphoma demonstrates a high proliferation index, frequent c-Myc protein expression, and a high incidence of germinal center subtype: Report of the French-American-British (FAB) international study group." Pediatr Blood Cancer 51(3): 369-374.

Miller, J. P., D. Izon, W. DeMuth, R. Gerstein, A. Bhandoola and D. Allman (2002). "The Earliest Step in B Lineage Differentiation from Common Lymphoid Progenitors Is Critically Dependent upon Interleukin 7." Journal of Experimental Medicine 196(5): 705-711.

Muller, J. R., S. Janz, J. J. Goedert, M. Potter and C. S. Rabkin (1995). "Persistence of immunoglobulin heavy chain/c-myc recombination-positive lymphocyte clones in the blood of human immunodeficiency virus-infected homosexual men." Proc Natl Acad Sci U S A 92(14): 6577-6581.

Muramatsu, M., V. S. Sankaranand, S. Anant, M. Sugai, K. Kinoshita, N. O. Davidson and T. Honjo (1999). "Specific Expression of Activation-induced Cytidine Deaminase (AID), a Novel Member of the RNA-editing Deaminase Family in Germinal Center B Cells." Journal of Biological Chemistry 274(26): 18470-18476.

Murphy, K., P. Travers and M. Walport (2008). Janeway's Immunobiology. New York, Garland Science, Taylor & Francis Group, LLC.

Murphy, S. B. (1980). "Classification, staging and end results of treatment of childhood non-Hodgkin's lymphomas: dissimilarities from lymphomas in adults." Semin Oncol 7(3): 332-339.

Murre, C. (2005). "Helix-loop-helix proteins and lymphocyte development." Nat Immunol 6(11): 1079-1086.

NCBI dbSNP database. (2013). "Database of Single Nucleotide Polymorphisms (dbSNP Build ID: 137)." http://www.ncbi.nlm.nih.gov/SNP/, from http://www.ncbi.nlm.nih.gov/SNP/.

NCBI Gene database. (2013). Retrieved 24.11.2013, 2013, from http://www.ncbi.nlm.nih.gov/gene/.

Nelson, M., S. L. Perkins, B. J. Dave, P. F. Coccia, J. A. Bridge, E. R. Lyden, N. A. Heerema, M. A. Lones, L. Harrison, M. S. Cairo and W. G. Sanger (2010). "An increased frequency of 13q deletions detected by fluorescence in situ hybridization and its impact on survival in children and adolescents with Burkitt lymphoma: results from the Children's Oncology Group study CCG-5961." Br J Haematol 148(4): 600-610.

Nepal, R. M., A. Zaheen, W. Basit, L. Li, S. A. Berger and A. Martin (2008). "AID and RAG1 do not contribute to lymphomagenesis in Emu c-myc transgenic mice." Oncogene 27(34): 4752-4756.

NHL-BFM Study Group (2012). NHL-BFM Registry 2012, Gesellschaft für Pädiatrische Hämatologie und Onkologie (GPOH).

Noetzel, E., J. Veeck, D. Niederacher, O. Galm, F. Horn, A. Hartmann, R. Knuchel and E. Dahl (2008). "Promoter methylation-associated loss of ID4 expression is a marker of tumour recurrence in human breast cancer." BMC Cancer 8: 154.

O'Riordan, M. and R. Grosschedl (1999). "Coordinate Regulation of B Cell Differentiation by the Transcription Factors EBF and E2A." Immunity 11: 21-31.

Oliveros, J. C. (2007). "VENNY. An interactive tool for comparing lists with Venn Diagrams." Retrieved 21.11.2013, 2013, from http://bioinfogp.cnb.csic.es/tools/venny/index.html.

Oschlies, I., W. Klapper, M. Zimmermann, M. Krams, H. H. Wacker, B. Burkhardt, L. Harder, R. Siebert, A. Reiter and R. Parwaresch (2006). "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-Munster) Multicenter Trial." Blood 107(10): 4047-4052.

Pasqualucci, L., P. Neumeister, T. Goossens, G. Nanjangud, R. S. Chaganti, R. Kuppers and R. Dalla-Favera (2001). "Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas." Nature 412(6844): 341-346.

Patel, D. and J. Chaudhary (2012). "Increased expression of bHLH transcription factor E2A (TCF3) in prostate cancer promotes proliferation and confers resistance to doxorubicin induced apoptosis." Biochem Biophys Res Commun 422(1): 146-151.

Pearson, K. (1900). "X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling." Philosophical Magazine Series 5 50(302): 157-175.

Perk, J., A. Iavarone and R. Benezra (2005). "Id family of helix-loop-helix proteins in cancer." Nat Rev Cancer 5(8): 603-614.

Perkins, A. S. and J. W. Friedberg (2008). "Burkitt lymphoma in adults." Hematology Am Soc Hematol Educ Program, 10.1182/asheducation-2008.1.341: 341-348.

Phi, J. H., S. A. Choi, S. H. Lim, J. Lee, K. C. Wang, S. H. Park and S. K. Kim (2013). "ID3 contributes to cerebrospinal fluid seeding and poor prognosis in medulloblastoma." BMC Cancer 13: 291.

Preudhomme, C., I. Dervite, E. Wattel, M. Vanrumbeke, M. Flactif, J. L. Lai, B. Hecquet, M. C. Coppin, B. Nelken, B. Gosselin and et al. (1995). "Clinical significance of p53 mutations in newly diagnosed Burkitt's lymphoma and acute lymphoblastic leukemia: a report of 48 cases." J Clin Oncol 13(4): 812-820.

Reiter, A., M. Schrappe, M. Tiemann, W. D. Ludwig, E. Yakisan, M. Zimmermann, G. Mann, A. Chott, W. Ebell, T. Klingebiel, N. Graf, B. Kremens, S. Muller-Weihrich, H. J. Pluss, F. Zintl, G. Henze and H. Riehm (1999). "Improved treatment results in childhood B-cell neoplasms with tailored intensification of therapy: A report of the Berlin-Frankfurt-Munster Group Trial NHL-BFM 90." Blood 94(10): 3294-3306.

Richter, J., M. Schlesner, S. Hoffmann, M. Kreuz, E. Leich, B. Burkhardt, M. Rosolowski, O. Ammerpohl, R. Wagener, S. H. Bernhart, D. Lenze, M. Szczepanowski, M. Paulsen, S. Lipinski, R. B. Russell, S. Adam-Klages, G. Apic, A. Claviez, D. Hasenclever, V. Hovestadt, N. Hornig, J. O. Korbel, D. Kube, D. Langenberger, C. Lawerenz, J. Lisfeld, K. Meyer, S. Picelli, J. Pischimarov, B. Radlwimmer, T. Rausch, M. Rohde, M. Schilhabel, R. Scholtysik, R. Spang, H. Trautmann, T. Zenz, A. Borkhardt, H. G. Drexler, P. Moller, R. A. Macleod, C. Pott, S. Schreiber, L. Trumper, M. Loeffler, P. F. Stadler, P. Lichter, R. Eils, R. Kuppers, M. Hummel, W. Klapper, P. Rosenstiel, A. Rosenwald, B. Brors and R. Siebert (2012). "Recurrent mutation of the ID3 gene in Burkitt lymphoma identified by integrated genome, exome and transcriptome sequencing." Nat Genet, 10.1038/ng.2469.

Robb-Smith, A. H. T. (1982). "U.S. NATIONAL CANCER INSTITUTE WORKING FORMULATION OF NON-HODGKIN'S LYMPHOMAS FOR CLINICAL USE." The Lancet 320(8295): 432-434.

Rohde, M., M. Schlesner, J. Richter, M. Szczepanowski, J. Lisfeld, A. Claviez, A. Borkhardt, B. Brors, R. Eils, S. Eberth, M. Hummel, W. Klapper, J. O. Korbel, M. Loeffler, P. Rosenstiel, A. Rosenwald, L. Truemper, A. Reiter, R. Siebert and B. Burkhardt (2013). "Pediatric Burkitt Lymphoma of the NHL-BFM group analyzed within the ICGC-MMML-SEQ: Whole genome sequencing data from 12 cases." Hematol Oncol 31 Suppl 1(Abstract number 081): 96-150.

Roschke, V., E. Kopantzev, M. Dertzbaugh and S. Rudikoff (1997). "Chromosomal translocations deregulating c-myc are associated with normal immune responses." Oncogene 14(25): 3011-3016.

Rose, M. L., M. S. C. Birbeck, V. J. Wallis, J. A. Forrester and A. J. S. Davies (1980). "Peanut lectin binding properties of germinal centers of mouse lymphoid tissue." Nature 284(5754): 364-366.

Rosenwald, A., G. Wright, W. C. Chan, J. M. Connors, E. Campo, R. I. Fisher, R. D. Gascoyne, H. K. Muller-Hermelink, E. B. Smeland, J. M. Giltnane, E. M. Hurt, H. Zhao, L. Averett, L. Yang, W. H. Wilson, E. S. Jaffe, R. Simon, R. D. Klausner, J. Powell, P. L. Duffey, D. L. Longo, T. C. Greiner, D. D. Weisenburger, W. G. Sanger, B. J. Dave, J. C. Lynch, J. Vose, J. O. Armitage, E. Montserrat, A. Lopez-Guillermo, T. M. Grogan, T. P. Miller, M. LeBlanc, G. Ott, S. Kvaloy, J. Delabie, H. Holte, P. Krajci, T. Stokke and L. M. Staudt (2002). "The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma." N Engl J Med 346(25): 1937-1947.

Rossig, C., H. Juergens, M. Schrappe, A. Moericke, G. Henze, A. von Stackelberg, D. Reinhardt, B. Burkhardt, W. Woessmann, M. Zimmermann, H. Gadner, G. Mann, G. Schellong, C. Mauz-Koerholz, U. Dirksen, S. Bielack, F. Berthold, N. Graf, S. Rutkowski, G. Calaminus, P. Kaatsch and U. Creutzig (2013). "Effective childhood cancer treatment: The impact of large scale clinical trials in Germany and Austria." Pediatr Blood Cancer, 10.1002/pbc.24598.

Salaverria, I., C. Philipp, I. Oschlies, C. W. Kohler, M. Kreuz, M. Szczepanowski, B. Burkhardt, H. Trautmann, S. Gesk, M. Andrusiewicz, H. Berger, M. Fey, L. Harder, D. Hasenclever, M. Hummel, M. Loeffler, F. Mahn, I. Martin-Guerrero, S. Pellissery, C. Pott, M. Pfreundschuh, A. Reiter, J. Richter, M. Rosolowski, C. Schwaenen, H. Stein, L. Trumper, S. Wessendorf, R. Spang, R. Kuppers, W. Klapper and R. Siebert (2011). "Translocations activating IRF4 identify a subtype of germinal center-derived B-cell lymphoma affecting predominantly children and young adults." Blood 118(1): 139-147.

Sambrook, J. and D. W. Russel (2001). Commonly Used Techniques in Molecular Cloning. New York, Cold Spring Harbor Laboratory Press.

Sanchez-Beato, M., A. Sanchez-Aguilera and M. A. Piris (2003). "Cell cycle deregulation in B-cell lymphomas." Blood 101(4): 1220-1235.

Sanger, F., S. Nicklen and A. R. Coulson (1977). "DNA sequencing witH-chain-terminating inhibitors." Proc Natl Acad Sci 74(12): 5463-5467.

Sawai, C. M., J. Freund, P. Oh, D. Ndiaye-Lobry, J. C. Bretz, A. Strikoudis, L. Genesca, T. Trimarchi, M. A. Kelliher, M. Clark, J. Soulier, S. Chen-Kiang and I. Aifantis (2012). "Therapeutic targeting of the cyclin D3:CDK4/6 complex in T cell leukemia." Cancer Cell 22(4): 452-465.

Sayegh, C. E., M. W. Quong, Y. Agata and C. Murre (2003). "E-proteins directly regulate expression of activation-induced deaminase in mature B cells." Nat Immunol 4(6): 586-593.

Schmitz, R., R. M. Young, M. Ceribelli, S. Jhavar, W. Xiao, M. Zhang, G. Wright, A. L. Shaffer, D. J. Hodson, E. Buras, X. Liu, J. Powell, Y. Yang, W. Xu, H. Zhao, H. Kohlhammer, A. Rosenwald, P. Kluin, H. K. Muller-Hermelink, G. Ott, R. D. Gascoyne, J. M. Connors, L. M. Rimsza, E. Campo, E. S. Jaffe, J. Delabie, E. B. Smeland, M. D. Ogwang, S. J. Reynolds, R. I. Fisher, R. M. Braziel, R. R. Tubbs, J. R. Cook, D. D. Weisenburger, W. C. Chan, S. Pittaluga, W. Wilson, T. A. Waldmann, M. Rowe, S. M. Mbulaiteye, A. B. Rickinson and L. M. Staudt (2012). "Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics." Nature 490(7418): 116-120.

Sellam, J., C. Miceli-Richard, J. E. Gottenberg, A. Proust, M. Ittah, F. Lavie, P. Loiseau and X. Mariette (2008). "Is Inhibitor of differentiation 3 involved in human primary Sjogren's syndrome?" Rheumatology (Oxford) 47(4): 437-441.

Sharma, P., D. Patel and J. Chaudhary (2012). "Id1 and Id3 expression is associated with increasing grade of prostate cancer: Id3 preferentially regulates CDKN1B." Cancer Med 1(2): 187-197.

Sherr, C. J. and J. M. Roberts (2004). "Living with or without cyclins and cyclin-dependent kinases." Genes Dev 18(22): 2699-2711.

Shuno, Y., N. H. Tsuno, Y. Okaji, T. Tsuchiya, D. Sakurai, T. Nishikawa, N. Yoshikawa, K. Sasaki, K. Hongo, G. Tsurita, E. Sunami, J. Kitayama, K. Tokunaga, K. Takahashi and H. Nagawa (2010). "Id1/Id3 knockdown inhibits metastatic potential of pancreatic cancer." J Surg Res 161(1): 76-82.

Sigvardsson, M. (2000). "Overlapping expression of early B-cell factor and basic helix-loop-helix proteins as a mechanism to dictate B-lineage-specific activity of the lambda5 promoter." Mol Cell Biol 20(10): 3640-3654.

Snyder, A. D., A. N. Dulin-Smith, R. H. Houston, A. N. Durban, B. J. Brisbin, T. D. Oostra, J. T. Marshall, B. M. Kahwash and C. R. Pierson (2013). "Expression Pattern of Id Proteins in Medulloblastoma." Pathol Oncol Res, 10.1007/s12253-012-9599-4.

Song, S., J. Cooperman, D. L. Letting, G. A. Blobel and J. K. Choi (2004). "Identification of cyclin D3 as a direct target of E2A using DamID." Mol Cell Biol 24(19): 8790-8802.

Stansfeld, A. G., J. Diebold, H. Noel, Y. Kapanci, F. Rilke, G. Kelenyi, C. Sundstrom, K. Lennert, J. A. van Unnik, O. Mioduszewska and et al. (1988). "Updated Kiel classification for lymphomas." Lancet 1(8580): 292-293.

Strong, N., A. C. Millena, L. Walker, J. Chaudhary and S. A. Khan (2013). "Inhibitor of differentiation 1 (Id1) and Id3 proteins play different roles in TGFbeta effects on cell proliferation and migration in prostate cancer cells." Prostate 73(6): 624-633.

Swerdlow, S. H., E. Campo, N. L. Harris, E. S. Jaffe, S. A. Pileri, H. Stein, J. Thiele and J. W. Vardiman (2008). WHO Classification of Tumours of Haematopoietic and Lymphoid Tissue.

The Non-Hodgkin's Lymphoma Classification Project (1997). "A clinical evaluation of the International Lymphoma Study Group classification of non-Hodgkin's lymphoma. The Non-Hodgkin's Lymphoma Classification Project." Blood 89(11): 3909-3918.

Trautmann, H., A. Hadzidimitriou, N. Darzentas, W. Klapper, R. Siebert, H. Berger, M. Szczepanowski, M. Kneba, C. Philipp, R. Kuppers, K. Stamtaopoulos and C. Pott (2009). "Evidence for Antigen-Driven Development of Molecularly Classified Burkitt Lymphomas." Blood (ASH Annual Meeting Abstracts) 114(22): Abstract Number 317.

U.S. National Institutes of Health. (2013). "Study Of Letrozole With Or Without PD 0332991 For The First-Line Treatment Of Hormone-Receptor Positive Advanced Breast Cancer." http://clinicaltrials.gov/show/NCT00721409, from http://clinicaltrials.gov/show/NCT00721409.

UniProt database (2013). "Update on activities at the Universal Protein Resource (UniProt) in 2013." Nucleic Acids Res 41(Database issue): D43-47.

Verma-Gaur, J., J. Hauser and T. Grundstrom (2012). "Negative feedback regulation of antigen receptors through calmodulin inhibition of E2A." J Immunol 188(12): 6175-6183.

Wazir, U., W. G. Jiang, A. K. Sharma, R. F. Newbold and K. Mokbel (2013). "The mRNA expression of inhibitors of DNA binding-1 and -2 is associated with advanced tumour stage and adverse clinical outcome in human breast cancer." Anticancer Res 33(5): 2179-2183.

Wilda, M., J. Bruch, L. Harder, D. Rawer, A. Reiter, A. Borkhardt and W. Woessmann (2004). "Inactivation of the ARF-MDM-2-p53 pathway in sporadic Burkitt's lymphoma in children." Leukemia 18(3): 584-588.

Wiman, K. G., K. P. Magnusson, T. Ramqvist and G. Klein (1991). "Mutant p53 detected in a majority of Burkitt lymphoma cell lines by monoclonal antibody PAb240." Oncogene 6(9): 1633-1639.

Woessmann, W. (2013). "How to treat children and adolescents with relapsed Non-Hodgkin lymphoma." Hematol Oncol 31 Suppl 1: 64-68.

Woessmann, W., K. Seidemann, G. Mann, M. Zimmermann, B. Burkhardt, I. Oschlies, W. D. Ludwig, T. Klingebiel, N. Graf, B. Gruhn, H. Juergens, F. Niggli, R. Parwaresch, H. Gadner, H. Riehm, M. Schrappe and A. Reiter (2005). "The impact of the methotrexate administration schedule and dose in the treatment of children and adolescents with B-cell neoplasms: a report of the BFM Group Study NHL-BFM95." Blood 105(3): 948-958.

Xu, Z., E. J. Pone, A. Al-Qahtani, S. R. Park, H. Zan and P. Casali (2007). "Regulation of aicda expression and AID activity: relevance to somatic hypermutation and class switch DNA recombination." Crit Rev Immunol 27(4): 367-397.

10 Appendix

10.1 Overall sequencing results on B-NHL patients

Overall ID3, TCF3 and CCND3 sequencing results on 84 B-NHL patients from the study cohort and 10 B-NHL patients from the extended cohort

case	diagnosis	MYC rearr.	DNA origin	c.ID3	p.ID3	c.TCF3	p.TCF3	c.CCND3	p.CCND3
1	BL	yes	ascites	c.[198_199insCTAAG];[c.194G>A]	p.[V67fs*16];[S65N]	wt	wt	c.[850C>T]	p.[P284S]
2	BL	yes	tissue	wt	wt	c.[1724T>A]	p.[M575K]	c.[811dupC]	p.[R271Pfs*53]
3	BL	yes	tissue	wt	wt	wt	wt	wt	wt
4	BL	yes	tissue	wt	wt	wt	wt	wt	wt
5	BL	yes	tissue	c.[141C>A(;)144C>T(;)166C>T]	p.[C47*(;)P56S]	wt	wt	c.[758_759dupAG]	p.[S254Rfs*50]
6	DLBCL	N/A	tissue	wt	wt	wt	wt	wt	wt
7	B-AL	yes	tissue	c.[241C>T]	p.[Q81*]	wt	wt	c.[869T>G]	p.[I290R]
8	BL	yes	tissue	c.[144C>G;243G>C];[236_243delACCTGCAG]	p.[Y48*;Q81H];[D79GfsSPGRASPWTP*]	wt	wt	wt	wt
9	BL	yes	tissue	c.[180_190delAGGCACTCAGC]	p.[R60Sfs*]	wt	wt	c.[811dupC]	p.[R271Pfs*53]
10	BL	yes	pleura	c.[153_164delGCGGGAACTGGT];[256_266delGAGCCAGCCCC]	p.[R52_V55del];[E86WfsTP*]	wt	wt	wt	wt
11	BL	yes	ascites	c.[144C>G];[300+1G>C]	p.[Y48*];[V72_Q100del]	wt	wt	wt	wt
12	BL	yes	tissue	c.[236_247delACCTGCAGGTAG]	p.[L80PfsGRASPWTP*]	c.[1675G>A]	p.[V559M]	c.[857_878delATGTCACAGCCATACACCTGTA]	p.[D286GfsPGEALWSGH*]
13	BL	yes	ascites	c.[142T>A]	p.[Y48N]	wt	wt	wt	wt
14	B-AL	yes	bm	c.[241C>T]	p.[Q81*]	wt	wt	c.[778C>T]	p.[Q260*]
15	BL	yes	tissue	c.[120delG(;)166C>T]	p.[L40FfsWTT*(;)P56S]	wt	wt	wt	wt
16	B-AL	yes	tissue	c.[211C>T(;)241C>T]	p.[Q71*(;)Q81*]	wt	wt	wt	wt
17	DLBCL	no	tissue	wt	wt	wt	wt	wt	wt
18	DLBCL	N/A	tissue	wt	wt	wt	wt	wt	wt
19	B-NHL	yes	tissue	wt	wt	wt	wt	c.[786_796dup]	p.[A266Pfs*40]
20	BL	yes	tissue	wt	wt	wt	wt	wt	wt
21	DLBCL	no	tissue	wt	wt	wt	wt	wt	wt

case	diagnosis	MYC rearr.	DNA origin	c.ID3	p.ID3	c.TCF3	p.TCF3	c.CCND3	p.CCND3
22	BL	yes	tissue	wt	wt	wt	wt	wt	wt
23	BL	yes	tissue	c.[236_251delinsC];[300+1G>C]	p.[D79Gfs*?];[V72_Q100del]	c.[1663G>C]	p.[E555Q]	wt	wt
24	BL	yes	tissue	c.[190C>T]	p.[L64F]	wt	wt	c.[811dupC]	p.[R271Pfs*53]
25	BL	yes	tissue	c.[143A>G]	p.[Y48C]	wt	wt	wt	wt
26	BL	yes	tissue	c.[189delG]	p.[Q63HfsLARWKSYSASSTTFSTCR*]	wt	wt	wt	wt
27	B-NHL	yes	tissue	c.[152T>C(;)228C>G]	p.[L51P(;)Y76*]	wt	wt	wt	wt
28	DLBCL	no	tissue	wt	wt	wt	wt	wt	wt
29	BL	no	tissue	wt	wt	wt	wt	wt	wt
30	DLBCL	yes	tissue	wt	wt	wt	wt	wt	wt
31	DLBCL	yes	ascites	c.[166C>T(;)211C>T]	p.[P56S(;)Q71*]	wt	wt	c.[811dupC]	p.[R271Pfs*53]
32	BL	yes	tissue	c.[190C>T]	p.[L64F]	c.[1653T>A]	p.[N551K]	wt	wt
33	BL	yes	tissue	c.[190C>T]	p.[L64F]	c.[1675G>A]	p.[V559M]	c.[766_767delAG]	p.[R256Gfs*56]
34	BL	yes	tissue	wt	wt	wt	wt	wt	wt
35	BL	yes	tissue	c.[167C>T];[181_209delGGCACTCAGCTTAGCCAGGTGGAAATCCT]	p.[P56L];[G61WfsKSYSASSTTFSTCR*]	wt	wt	wt	wt
36	BL	yes	pleura	c.[167C>T(;)190C>T]	p.[P56L(;)L64F]	wt	wt	wt	wt
37	BL	yes	tissue	c.[166_167delinsTT]	p.[P56F]	wt	wt	wt	wt
38	BL	yes	tissue	wt	wt	wt	wt	c.[869T>G]	p.[I290R]
39	BL	yes	pleura	c.[191_195delTTAGC]	p.[L64PfsGGNPTARHRLHSRPAGSPGRASPWTP*]	wt	wt	wt	wt
40	B-AL	yes	bm	c.[300+1G>T]	p.[V72_Q100del]	wt	wt	c.[782_792delCCAGCTCCAGC]	p.[S263Afs*56]
41	DLBCL	no	tissue	wt	wt	wt	wt	wt	wt
42	B-NHL	no	tissue	wt	wt	wt	wt	wt	wt
43	BL	yes	tissue	c.[27_40delCTGCTACGAGGCGG;202delG];[202delG];[27_40delCTGCTACGAGGCGG]	p.[C10VfsLPVGTQSGHRPGPREGPGS*;E68KfsSYSASSTTFSTCR*];[E68KfsSYSASSTTFSTCR*];[C10VfsLPVGTQSGHRPGPREGPGS*]	wt	wt	c.[811dupC]	p.[R271Pfs*53]
44	B-AL	yes	bm	c.[137A>C(;)241C>T]	p.[H45P(;)Q81*]	wt	wt	c.[869T>A]	p.[I290K]

case	diagnosis	MYC rearr.	DNA origin	c.ID3	p.ID3	c.TCF3	p.TCF3	c.CCND3	p.CCND3
45	BL	yes	tissue	c.[190C>T(;)233T>C]	p.[L64F(;)L78P]	wt	wt	wt	wt
46	DLBCL	no	tissue	wt	wt	wt	wt	wt	wt
47	DLBCL	yes	tissue	c.[241C>T]	p.[Q81*]	wt	wt	wt	wt
48	B-AL	yes	bm	c.[181_270del]	p.[G61_G90del]	wt	wt	c.[850C>T]	p.[P284S]
49	BL	yes	tissue	c.[116_delG]	p.[S36Tfs*5]	wt	wt	wt	wt
50	B-AL	yes	tissue	c.[166C>T(;)209T>C]	p.[P56F(;)L70P]	wt	wt	c.[869T>G]	p.[I290R]
51	DLBCL	no	tissue	wt	wt	wt	wt	wt	wt
52	B-AL	yes	bm	wt	wt	c.[1670T>C]	p.[V557A]	wt	wt
53	BL	yes	tissue	c.[134_140dupACCACTG];[209T>C]	p.[C48*];[L70P]	wt	wt	wt	wt
54	BL	yes	pleura	c.[214_243del;(214_243del)]	p.[R72_Q81del;(R72_Q81del)]	wt	wt	c.[811dupC]	p.[R271Pfs*53]
55	BL	yes	tissue	c.[144C>G;(144C>G)]	p.[Y48*;(Y48*)]	wt	wt	c.[811dupC]	p.[R271Pfs*53]
56	DLBCL	yes	tissue	c.[137A>C(;)161T>A(;)300+44T>C(;)300+85C>T]	p.[H45P(;)L54Q]	wt	wt	wt	wt
57	BL	N/A	tissue	c.[190C>T(;)230T>A]	p.[L64F(;)I77N]	wt	wt	c.[811dupC]	p.[R271Pfs*53]
58	DLBCL	no	tissue	wt	wt	wt	wt	wt	wt
59	BL	yes	tissue	c.[167C>G]	p.[P56R]	wt	wt	c.[850C>T]	p.[P284S]
60	DLBCL	N/A	tissue	wt	wt	wt	wt	c.[864_873delAGCCATACAC]	p.[A289CfsSPGEALWSGH*]
61	BL	yes	pleura	c.[141C>A(;)190C>T]	p.[C47*(;)L64F]	wt	wt	wt	wt
62	DLBCL	no	tissue	wt	wt	wt	wt	wt	wt
63	BL	yes	ascites	wt	wt	c.[1683T>G]	p.[D561E]	wt	wt
64	B-AL	yes	bm	c.[167C>G]	p.[P56R]	wt	wt	c.[869T>G]	p.[I290R]
65	BL	yes	tissue	wt	wt	wt	wt	wt	wt
66	B-AL	yes	bm	c.[135C>G(;)166C>T(;)190C>T]	p.[N45K(;)P56S(;)L64F]	c.[1670T>A]	p.[V557E]	wt	wt
67	BL	yes	tissue	c.[142T>G(;)166C>G(;)190C>G]	p.[Y48D(;)P56A(;)L64V]	wt	wt	wt	wt
68	B-NHL	no	pleura	c.[20T>A(;)164T>A]	p.[V7E(;)V55E]	wt	wt	wt	wt
69	BL	yes	tissue	c.[190C>T]	p.[L64F]	wt	wt	wt	wt
70	DLBCL	N/A	tissue	c.[152_174dup(;)209T>G]	p.[P59Cfs*32(;)L70R]	wt	wt	wt	wt
71	BL	yes	ascites	c.[122_140delTGGACGACATGAACCACTG_insC;(122_140delTG	p.[L41Pfs*73;(L41Pfs*73]	wt	wt	c.[826C>T]	p.[Q276*]

case	diagnosis	MYC rearr.	DNA origin	c.ID3	p.ID3	c.TCF3	p.TCF3	c.CCND3	p.CCND3
				GACGACATGAACCACTG_insC)]					
72	BL	yes	ascites	c.[243dupA]	p.[V82Gfs*11]	wt	wt	wt	wt
73	BL	yes	pleura	c.[81delC];[247_248insTCTACAGCGCGTCATCGACTACATTCTCGACCTGCAGGTAG]	p.[R28EfsGRARQLRSR*];[L84Yfs*12]	wt	wt	wt	wt
74	BL	yes	pleura	wt	wt	wt	wt	c.[838C>T]	p.[Q280*]
75	BL	yes	tissue	c.[166C>T]	p.[P56S]	wt	wt	wt	wt
76	B-AL	yes	bm	c.[167_182dupCCGGAGTCCCGAGAGG]	p.[T62RfsSPERHSA*]	wt	wt	c.[811dupC]	p.[R271Pfs*53]
77	BL	yes	tissue	wt	wt	wt	wt	wt	wt
78	B-AL	yes	bm	c.[189dedelG;193A>T];[190C>T]	p.[Q63HfsFARWKSYSASSTTFSTCR*];[L64F]	wt	wt	wt	wt
79	B-AL	yes	bm	c.[160C>G(;)164T>C]	p.[L54V(;)V55A]	wt	wt	wt	wt
80	B-NHL	N/A	tissue	c.[160C>G];[165_166insG]	p.[L54V];[P56AfsRSPERHSA*]	wt	wt	wt	wt
81	B-AL	yes	bm	c.[166C>G];[190C>G;206T>C;229A>G]	p.[P56A];[L64V(;)I69T(;)I77V]	wt	wt	wt	wt
82	BL	no	tissue	wt	wt	wt	wt	wt	wt
83	BL	yes	tissue	c.[166C>T(;)190C>T]	p.[P56S(;)L64F]	wt	wt	wt	wt
84	BL	yes	tissue	c.[157delG(;)166C>G(;)209T>A]	p.[P56A(;)E53NfsWYPESREALSLARWKSYSASSTTFSTCR*(;)L70Q]	wt	wt	wt	wt
extended cohort (cases 85-94)									
case ID	**diag.**	**myc rearr.**	**tumor origin**	**ID3 (coding)**	**ID3 (protein)**	**TCF3 (coding)**	**TCF3 (protein)**	**CCND3 (coding)**	**CCND3 (protein)**
85	BL	yes	ascites	c.[190C>T(;)300+1G>C]	p.[L64F(;)V72_Q100l]	wt	wt	wt	wt
86	BL	yes	tissue	wt	wt	wt	wt	wt	wt
87	BL	yes	tissue	c.[166C>T(;)202lG]	p.[P56S(;)E68KfsSYSASSTTFSTCR*]	wt	wt	wt	wt
88	BL	yes	pleura	wt	wt	wt	wt	wt	wt
89	BL	yes	ascites	c.[167C>G;(167C>G)]	p.[P56R;(P56R)]	wt	wt	wt	wt
90	BL	yes	tissue	c.[241C>T(;)300G>A(;)310C>T]	p.[Q81*(;)Q100Q(;)L103F]	c.[1675G>A]	p.[V559M]	wt	wt

case	diagnosis	MYC rearr.	DNA origin	c.ID3	p.ID3	c.TCF3	p.TCF3	c.CCND3	p.CCND3
91	BL	yes	tissue	c.[93_dupG;(93_dupG)]	p.[G31fsPGS*;(G31fsPGS*)]	wt	wt	c.[856_866dupGA]	p.[D286Efs*18]
92	B-NHL	yes	tissue	c.[122_130lTGGACGACinsA;(122_130lTGGACGACinsA)]	p.[L41Hfs*21;(L41Hfs*21)]	wt	wt	wt	wt
93	B-AL	yes	bm	c.[45C>A(;)202G>T]	p.[C15*(;)E68*]	wt	wt	wt	wt
94	B-AL	yes	tissue	c.[190C>T(;)298C>T]	p.[L64F(;)Q100*]	wt	wt	c.[811dupC]	p.[R271Pfs*53]

wt: wildtype, bm: bone marrow, MYC rearr.: MYC-rearrangment, N/A: not available

10.2 Correlation between certain mutational patterns and clinical data

Correlation between certain mutational patterns within 66 MYC rearrangement positive cases and clinical data.

		ID3 ± CCND3			**ID3 ± TCF3**			**TCF3 ± CCND3**			**ID3 ± TCF3 ± CCND3**		
		mutated	wildtype	p value	mutated	wildtype	p value	mutated	wildtype	p value	mutated	wildtype	p value
Gender	male	48	9		47	10		22	35		50	7	
	female	8	1	1.00	8	1	1.00	7	2	0.04	8	1	1.00
Age	< 10	34	4		34	4		21	17		36	2	
	10-14	15	5		14	6		5	15		15	5	
	> 14	7	1	0.34	7	1	0.16	3	5	0.81	7	1	0.09
Stage of disease	I	2			1	2		1	2		2	1	
	II	10	1		9	2		7	4		10	1	
	III	25	1		26	6		9	23		26	6	
	IV	2	7		1	1		2	0		2	0	
	B-AL	13	1	0.55	14	0	0.05	10	4	0.02	14	0	0.33
BM involvement	yes	13	1	0.68	14	0	0.10	10	4	0.03	14	0	0.19
CNS involvement	yes	7	0	0.58	6	1	1.00	6	1	0.04	7	0	0.58
LDH level (U/l)	< 500	20	6		18	8		10	16		20	6	
	500-1000	11	1		11	1		6	6		11	1	
	> 1000	25	3	0.34	26	2	0.05	13	15	0.75	27	1	0.08
Diagnosis	BL	38	8		37	9		17	29		39	7	
	B-AL	13	1		14	0		10	4		14	0	
	DLBCL	3	1		3	1		1	3		3	1	
	B-NHL (DD)	2	0	0.67	1	1	0.18	1	1	0.12	2	0	0.36
Outcome	pEFS (2y)	87 ± 4%	90 ± 9%	0.8 (LR)	87 ± 5%	91 ± 9%	0.71 (LR)	83 ± 7%	92 ± 4%	0.27 (LR)	88 ± 4%	88 ±12%	0.99 (LR)
	pOS (2y)	89 ± 4%	90 ± 9%	0.95 (LR)	89 ± 4%	91 ± 9%	0.86 (LR)	86 ± 7%	92 ± 4%	0.47 (LR)	90 ± 4%	88 ±12%	0.84 (LR)

		ID3 exclusive			TCF3 exclusive			CCND3 exclusive			ID3 + TCF3		
		mutated	wildtype	p value	mutated	wildtype	p value	mutated	wildtype	p value	mutated	wildtype	p value
Gender	male	28	29		2	55		3	54		3	54	
	female	1	8	0.07	0	9	1.00	0	9	1.00	2	7	0.13
Age	< 10	15	23		2	36		2	36		4	34	
	10-14	10	10		0	20		1	19		1	19	
	> 14	4	4	0.7	0	8	0.47	0	8	0.8	0	8	0.52
Stage of disease	I	1	2		0	3		1	2		0	3	
	II	3	8		0	11		1	10		3	8	
	III	17	15		1	31		0	32		1	31	
	IV	0	2		0	2		1	1		0	2	
	B-AL	4	10	0.27	1	13	0.87	0	14	0.002	1	13	0.14
BM involvement	yes	4	10	0.24	1	13	0.38	0	14	1.00	1	13	11.00
CNS involvement	yes	1	6	0.12	0	7	1	1	6	0.29	1	6	0.44
LDH level (U/l)	< 500	10	16		0	26		2	24		3	23	
	500-1000	5	7		0	12		0	12		2	10	
	> 1000	14	14	0.68	2	26	0.25	1	27	0.54	0	28	0.12
Diagnosis	BL	22	24		1	45		2	44		4	42	
	B-AL	4	10		1	13		0	14		1	13	
	DLBCL	2	2		0	4		0	4		0	4	
	B-NHL (DD)	1	1	0.63	0	2	0.78	1	1	0.16	0	2	0.9
Outcome	pEFS (2y)	93 ± 5%	84 ± 6 %	0.27 (LR)	no events	87 ± 4%	0.61 (LR)	no events	87 ± 4%	0.52 (LR)	80 ± 18%	88 ± 4%	0.50 (LR)
	pOS (2y)	93 ± 5%	86 ± 6%	0.4 (LR)	no events	89 ± 4%	0.63 (LR)	no events	89 ± 4%	0.55 (LR)	80 ± 18%	90 ± 4%	0.42 (LR)

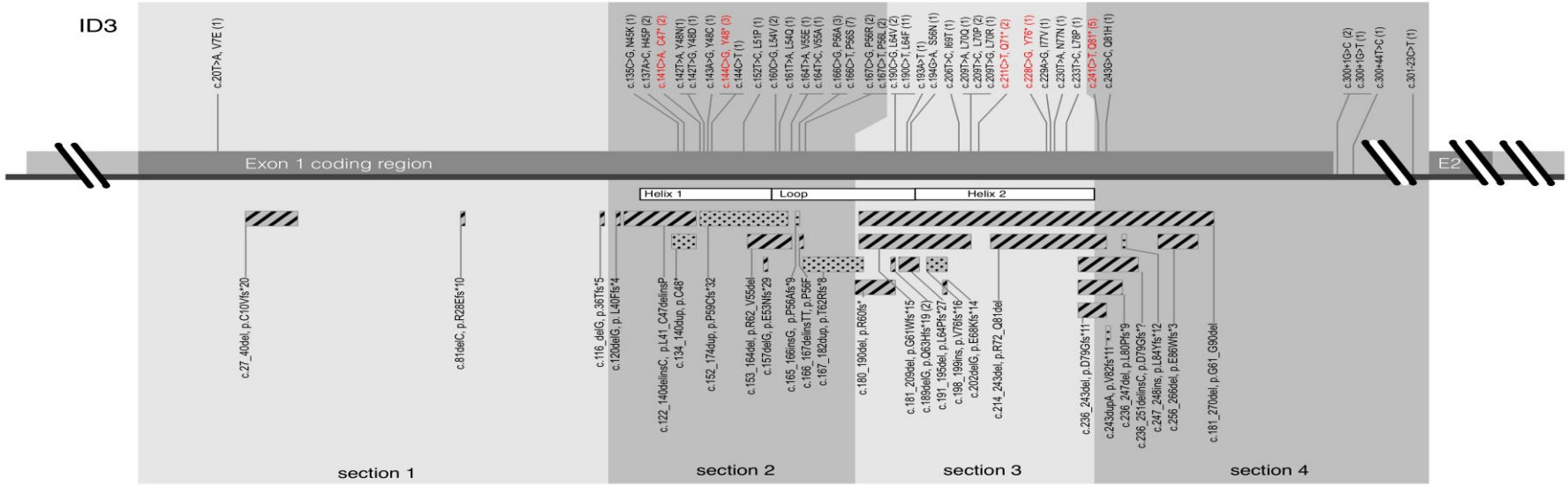
		ID3 + CCND3			TCF3 + CCND3			ID3 + TCF3 + CCND3			ID3 + TCF3 exclusive		
		mutated	wildtype	p value	mutated	wildtype	p value	mutated	wildtype	p value	mutated	wildtype	p value
Gender	male	15	42		3	54		2	55		1	56	
	female	5	4	0.11	0	9	1.00	0	9	1.00	2	7	0.05
Age	< 10	13	25		2	36		1	37		3	35	
	10-14	4	16		1	19		1	19		0	20	
	> 14	3	5	0.48	0	8	0.8	0	8	0.77	0	8	0.31
Stage of disease	I	0	3		0	3		0	3		0	3	
	II	3	8		2	9		1	10		2	9	
	III	8	24		1	31		1	31		0	32	
	IV	1	1		0	2		0	2		0	2	
	B-AL	8	6	0.16	0	14	0.25	0	14	0.76	1	13	0.18
BM involvement	yes	8	6	0.02	0	14	1.00	0	14	1.00	1	13	0.52
CNS involvement	yes	4	3	0.19	0	7	1.00	0	7	1.00	1	6	0.29
LDH level (U/l)	< 500	5	21		2	24		1	25		2	24	
	500-1000	5	7		1	11		1	11		1	11	
	> 1000	10	18	0.27	0	28	0.31	0	28	0.35	0	28	0.31
Diagnosis	BL	11	35		3	43		2	44		2	44	
	B-AL	8	6		0	14		0	14		1	13	
	DLBCL	1	3		0	4		0	4		0	4	
	B-NHL (DD)	0	2	0.09	0	2	0.71	0	2	0.82	0	2	0.92
Outcome	pEFS (2y)	80 ± 9%	91 ± 4%	0.22 (LR)	no events	87 ± 4%	0.52 (LR)	no events	87 ± 4%	0.61 (LR)	67 ± 27%	89 ± 4%	0.16 (LR)
	pOS (2y)	84 ± 8%	91 ± 4%	0.48 (LR)	no events	89 ± 4%	0.55 (LR)	no events	89 ± 4%	0.63 (LR)	67 ± 27%	90 ± 4%	0.11 (LR)

		ID3 + CCND3 exclusive			TCF3 + CCND3 exclusive		
		mutated	wildtype	p value	mutated	wildtype	p value
Gender	male	13	44		1	56	
	female	5	4	0.05	0	9	1.00
Age	< 10	12	26		1	37	
	10-14	3	17		0	20	
	> 14	3	5	0.31	0	8	0.69
Stage of disease	I	0	3		0	3	
	II	2	9		1	10	
	III	7	25		0	32	
	IV	1	1		0	2	
	B-AL	8	6	0.77	0	14	0.32
BM involvement	yes	8	6	0.01	0	14	1.00
CNS involvement	yes	4	3	0.08	0	7	1.00
LDH level (U/l)	< 500	4	22		1	25	
	500-1000	4	8		0	12	
	> 1000	10	18	0.21	0	28	0.46
Diagnosis	BL	9	37		1	45	
	B-AL	8	6		0	13	
	DLBCL	1	3		0	4	
	B-NHL (DD)	0	2	0.04	0	2	0.93
Outcome	pEFS (2y)	77 ± 10%	92 ± 4%	0.14 (LR)	no events	88 ± 4%	0.72 (LR)
	pOS (2y)	83 ± 9%	92 ± 4%	0.35 (LR)	no events	89 ± 4%	0.74 (LR)

All data refer to patients with successful investigation of the respective criteria. mut: mutated, wt: wildtype,, y: years, p value: probability value, BM: bone marrow, CNS: central nervous system, LDH: lactate dehydrogenase, U/l: Units per liter, pEFS: probability of event-free survival, pOS: probability of overall survival, LR: log-rank

10.3 ID3 gene segmented into sections

The ID3 gene plot was segmented to four sections that were separately analyzed for relevance with respect to clinical characteristics.



10.4 List of SNP identifiers

This list presents identifiers of Single nucleotide polymorphisms (SNPs) according to the dbSNP database (NCBI dbSNP database, 2013) that were taken into account in sequence analysis and data interpretation.

dbSNP identifier (ID3)
rs201368186
rs81216731
rs184371695
rs146711026
rs41268121
rs150987910
rs11574
rs202157091
rs139428797
rs201697476
rs201313484
rs147380623
rs26739388
rs147934602
rs74062742
rs201701997
rs190449947
rs11542315
rs146163818
rs199661785
rs142720912
rs61749352
rs138633497
rs201168120
rs146156581
rs2067053
rs200199499
rs34462407
rs142466418
rs146830460
rs140027264
rs148870313
rs148996867
rs138679775
rs11542319
rs180772482
rs141280661

dbSNP identifier (ID3)
rs147689987
rs11542317
rs201488898
rs81216730

dbSNP identifier (TCF3)
rs187316941
rs200962332
rs201939465
rs200433738
rs182233043

dbSNP identifier (CCND3)
rs1051130
rs201507759
rs200046302
rs147958536
rs202197675
rs142862109
rs3218101
rs3218103
rs3218102
rs33966734

11 List of publications

Recurrent mutation of the ID3 gene in Burkitt lymphoma identified by integrated genome, exome and transcriptome sequencing.

Nature Genetics, 2012 (Richter J, Schlesner M, Hoffmann S, Kreuz M, Leich E, Burkhardt B, Rosolowski M, Ammerpohl O, Wagener R, Bernhart SH, Lenze D, Szczepanowski M, Paulsen M, Lipinski S, Russell RB, Adam-Klages S, Apic G, Claviez A, Hasenclever D, Hovestadt V, Hornig N, Korbel JO, Kube D, Langenberger D, Lawerenz C, Lisfeld J, Meyer K, Picelli S, Pischimarov J, Radlwimmer B, Rausch T, **Rohde M**, Schilhabel M, Scholtysik R, Spang R, Trautmann H, Zenz T, Borkhardt A, Drexler HG, Möller P, MacLeod RA, Pott C, Schreiber S, Trümper L, Loeffler M, Stadler PF, Lichter P, Eils R, Küppers R, Hummel M, Klapper W, Rosenstiel P, Rosenwald A, Brors B, Siebert R; ICGC MMML-Seq Project.)

Incidence and prognostic relevance of genetic variations in T-cell lymphoblastic lymphoma in childhood and adolescence.

Blood, 2013 (Bonn BR, **Rohde M**, Zimmermann M, Krieger D, Oschlies I, Niggli F, Wrobel G, Attarbaschi A, Escherich G, Klapper W, Reiter A, Burkhardt B.)

Mature B-Cell Lymphoma and Leukemia in Children and Adolescents - Review of Standard Chemotherapy Regimen and Perspectives.

Pediatric Hematology and Oncology, 2013 (Worch J, **Rohde M**, Burkhardt B.)

Recurrent RHOA mutations in pediatric Burkitt lymphoma treated according to the NHL-BFM protocols

Genes, Chromosomes, Cancer, *2014* **(Rohde M**, Richter J, Schlesner M, Betts MJ, Claviez A, Bonn BR, Zimmermann M, Damm-Welk C, Russel RB, Borkhardt A, Eils R, Hoell JI, Szczepanowski M, Oschlies I, Klapper W, Burkhardt B, Siebert R)

Parts of this study have been presented in oral presentations:

Pediatric Burkitt Lymphoma of the NHL-BFM group analyzed within the ICGC-MMML-Seq: Whole genome sequencing data from 12 cases

Given at the 12th International Conference on Malignant Lymphoma, 19-22 June 2013, Lugano, Switzerland; Hematol Oncol 31 Suppl 1 (Abstract number 081): 96-150., 2013 **(Rohde, M**., M. Schlesner, J. Richter, M. Szczepanowski, J. Lisfeld, A. Claviez, A. Borkhardt, B. Brors, R. Eils, S. Eberth, M. Hummel, W. Klapper, J. O. Korbel, M. Loeffler, P. Rosenstiel, A. Rosenwald, L. Truemper, A. Reiter, R. Siebert and B. Burkhardt)

Burkitt-Lymphome: Ergebnisse der Whole-Genome-Sequenzierung und Validierung

Rohde M; at the BFM Plenartagung 2013, 24-26 October 2013, Würzburg, Germany

12 Declaration (in German)

„Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und ohne unzulässige Hilfe oder Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten oder nichtveröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht. Bei den von mir durchgeführten und in der Dissertation erwähnten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der „Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis“ niedergelegt sind, eingehalten sowie ethische, datenschutzrechtliche und tierschutzrechtliche Grundsätze befolgt. Ich versichere, dass Dritte von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen, oder habe diese nachstehend spezifiziert. Die vorgelegte Arbeit wurde weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde zum Zweck einer Promotion oder eines anderen Prüfungsverfahrens vorgelegt. Alles aus anderen Quellen und von anderen Personen übernommene Material, das in der Arbeit verwendet wurde oder auf das direkt Bezug genommen wird, wurde als solches kenntlich gemacht. Insbesondere wurden alle Personen genannt, die direkt und indirekt an der Entstehung der vorliegenden Arbeit beteiligt waren. Mit der Überprüfung meiner Arbeit durch eine Plagiatserkennungssoftware bzw. ein internetbasiertes Softwareprogramm erkläre ich mich einverstanden.“

_____ _____

Datum, Ort Unterschrift

13 Acknowledgment (in German)

This page is not available in the online version of this document.

14 Curriculum vitae

This page is not available in the online version of this document.