Small B cells in the bone marrow associated with diffuse large B-cell lymphoma and cold agglutinin disease

PhD thesis

Agnieszka Małecka

Department of Pathology
The Norwegian Radium Hospital
Oslo University Hospital

Faculty of Medicine
University of Oslo

UiO : University of Oslo

Oslo University Hospital

Oslo, 2017

Series of dissertations submitted to the
Faculty of Medicine, University of Oslo

ISBN 978-82-8377-181-7

Cover: Hanne Baadsgaard Utigard.
Print production: Reprosentralen, University of Oslo.

Table of Contents

Acknowledgements

The work presented in this thesis was performed at the Department of Pathology, Oslo University Hospital - Norwegian Radium Hospital, with the support from South-Eastern Norway Regional Health Authority, Norwegian Cancer Society and Radiumhospitalets Legater.

I am deeply grateful to my supervisor Jan Delabie for believing in me and for convincing me to start the PhD, for sharing with me his vast knowledge, for his constant scientific and practical support, for his gentle way of giving advices and his friendly attitude. I admire your enthusiasm for research, and your ability to find positive aspects even in negative experimental results.

I am very grateful to my co-supervisor Gunhild Trøen for all the help and guidance, especially after Jan left for Canada, for helping me both with scientific and administrative problems and for simply being there for me. I would not manage to finish this PhD without your help!

I would like to thank my co-supervisor Geir Tjønnfjord for guidance and help with patient samples.

I would like to thank my former co-supervisor Anne Tierens for the planning of the project and the flow cytometry work, for support, encouragement and good words.

I am very grateful to Sigbjørn Berentsen for providing me with CAD patient samples and for sharing with me his enthusiasm about the CAD project.

I would like to thank Harald Holte for planning the project, providing patient samples and contribution to the project.

I would like to thank Ingunn Østlie for help with sample collection, flow cytometry and FACS.

And I would like to thank all co-authors of my papers for their contribution.

I would like to express my deepest gratitude to colleagues from Molecular Pathology Laboratory for every day help, for all the things they have done for me, for help with analysis and lab procedures and with small things (e.g. finding reagents and equipment) that made my work possible, for finding time for me in your busy schedule and for all conversations, lunches and lab parties that helped me both relax and focus.
Special thanks to Helen Vålerhaugen for teaching me the basics of working in the lab, how to do it precisely, well and clean.

I would like to thank my parents for sending me to the University and teaching me the work ethics, my sister Danusia for believing in me and my brother Zbyszek for all the help with computers.
I would like to thank my boys Dorian and Iwo for letting me work in the evenings and weekends. I promise to have more time for you both. And Dorian, I promise to find time from now on to play some computer games with you.
I thank my baby-Kaja for being such a happy and easy-going baby and letting me work, and for not destroying my computer yet. I will let you play with it when you get a bit older.
And last but not least, Jędrek thank you for all the support! Thank you for taking care of kids and letting me work, and also for your intellectual input to my papers. I love you all so much.

Abbreviations

ABC	activated B-cell
AID	activation-induced cytidine deaminase
AP-1	activation protein 1
APCs	antigen presenting cells
B-ALL	B-lymphoblastic leukemia/lymphoma
BCL10	B-cell CLL/lymphoma 10
BCL2	B-cell lymphoma 2
BM	bone marrow
C	constant
CAD	primary cold agglutinin disease
CARD11	caspase recruitment domain family member 11
CBL-MZ	clonal B-cell lymphocytosis of marginal zone origin
CBM	CARD11/BCL10/MALT1 complex
CD	cluster of differentiation
CDKN2A	cyclin-dependent kinase inhibitor 2A
CDR	complementarity determining region
CLL	chronic lymphocytic leukemia
CSR	class switch recombination
CXCR4	C-X-C chemokine receptor type 4
D	diversity
df	degree-of-freedom
DLBCL	diffuse large B-cell lymphoma
FACS	fluorescent activated cell sorting
FL	follicular lymphoma
FR	framework region
GATK	Genome Analysis Toolkit
GC	germinal center
GCB	germinal center B cell
GRCh37	Genome Reference Consortium Human genome build 37
HDAC	histone deacetylase

IG	immunoglobulin
IGH	immunoglobulin heavy chain
IGK	immunoglobulin kappa light chain
IGL	immunoglobulin lambda light chain
IgM IgA, IgG IgE IgD	antibody isotypes
IGV	the Integrative Genomics Viewer
IKK	IκB kinase
indel	insertion/deletion
IRF	interferon response factor
ITAMs	immuno-receptor tyrosine-based activation motifs
J	joining
JNK	c-Jun N-terminal kinase
KMT2D	lysine methyltransferase 2D
LPL	lymphoplasmacytic lymphoma
MALT1	mucosa-associated lymphoid tissue lymphoma translocation gene 1
MAPK	mitogen-activated protein kinase
MBL	monoclonal B-cell lymphocytosis
MCL	mantle cell lymphoma
MGUS	monoclonal gammopathy of uncertain significance
MSBC	monoclonal small B-cells
mTOR	mammalian target of rapamycin
MYD88	myeloid differentiation primary response 88
MZ	marginal zone
MZL	marginal zone lymphoma
NCBI	the National Center for Biotechnology Information
NF-κB	nuclear factor-kappa B
NFAT	nuclear factor of activated T cells
NGS	next-generation sequencing
NK	natural killer
NMZL	nodal marginal zone lymphoma
NOS	not otherwise specified
PE	pair-end reads
pre-BCR	pre-B cell antigen receptor

R-CHOP	(R)ituximab, (C)yclophosphamide, (H)ydroxydaunorubicin, (O)ncovin, (P)rednisone or (P)rednisolone
SHM	somatic hypermutation
SMZL	splenic marginal zone lymphoma
SNPs/SNVs	single nucleotide polymorphisms/variations
SYK	spleen tyrosine kinase
TLR	toll-like receptor
TNFAIP3	tumor necrosis factor, alpha-induced protein 3
TP53	tumor protein p53
TRIF	TIR-domain-containing adaptor protein inducing IFNβ
V	variable
WHO	World Health Organization

List of included papers

Paper 1:

Primary diffuse large B-cell lymphoma associated with clonally-related monoclonal B lymphocytosis indicates a common precursor cell.

Agnieszka Małecka, Anne Tierens, Ingunn Østlie, Roland Schmitz, Gunhild Trøen, Signe Spetalen, Louis M. Staudt, Erlend Smeland, Harald Holte and Jan Delabie

Haematologica October 2015 100: e415-e418

Paper 2:

Immunoglobulin heavy and light chain gene features are correlated with primary cold agglutinin disease onset and activity

Agnieszka Małecka, Gunhild Trøen, Anne Tierens, Ingunn Østlie, Jędrzej Małecki, Ulla Randen, Sigbjørn Berentsen, Geir E. Tjønnfjord and Jan M.A. Delabie

Haematologica September 2016 101: e361-e364

Paper 3:

Frequent somatic mutations of *KMT2D* (*MLL2*) and *CARD11* genes in primary cold agglutinin disease

Agnieszka Małecka, Gunhild Trøen, Anne Tierens, Ingunn Østlie, Jędrzej Małecki, Ulla Randen, Junbai Wang, Sigbjørn Berentsen, Geir E. Tjønnfjord and Jan M.A. Delabie

Manuscript

A version of this manuscript was accepted for publication by British Journal of Haematology on 11th of October 2017.

Introduction

The immune system

The function of the human immune system is primarily defense against foreign microorganisms. However, it also has a function in tissue homeostasis and cancer. It consists of an early response innate immunity and an adaptive, late response, immunity. While the innate immune response reacts very rapidly to infection, it responds in the same way to repeated infections. In contrast, the adaptive immune response increases defensive capabilities with each infection. The main cells participating in innate immunity are dendritic cells, natural killer (NK) cells and phagocytes. Cells active in the adaptive immune response are B and T lymphocytes and antigen-presenting cells (APCs) (Table 1). The immune system consists of primary lymphoid tissues, including bone marrow (BM) and thymus, and secondary lymphoid tissues, including lymph nodes, spleen and mucosal lymphoid tissues. Primary lymphoid tissues are sites where lymphocytes are generated and mature, whereas secondary lymphoid tissues are sites where adaptive responses are induced and regulated. Cells active in the adaptive immune response are predominantly concentrated in secondary lymphoid tissues that allows cellular interactions, necessary for antigen recognition and lymphocyte activation. Secondary lymphoid tissues consist of B-cell zones and T-cell zones, comprising the different subsets of B and T lymphocytes, respectively.

Table 1. Cells of the immune system and their function.[1]

Cell type	Immune response	Function
Phagocytes (including neutrophils and macrophages)	Innate and adaptive (some types of response only)	identify, ingest and destroy microorganisms
Mast cells, basophils eosinophils	innate and adaptive	secrete inflammatory and antimicrobial mediators
Antigen presenting cells (dendritic cells, antigen-presenting cells for effector T lymphocytes, follicular dendritic cells)	link responses of innate immune system to responses of adaptive immune systems	capture antigens for presentation to lymphocytes; stimulate proliferation and differentiation of lymphocytes
B lymphocytes	adaptive	antibody production (humoral immunity); stimulation of phagocytosis; complement activation; neutralization of microbes
$CD4^+$ helper T lymphocytes	adaptive	stimulation of B and T cell proliferation and differentiation, macrophage activation, stimulation of inflammation
$CD8^+$ cytotoxic T lymphocytes	adaptive	killing cells infected with viruses or bacteria; rejection of allografts
regulatory T cells	adaptive	regulation of immune responses and self-tolerance (suppression of other T cells)
γδ T lymphocytes	innate	helper and cytotoxic function
Natural killer cells	innate	cytotoxic killing of viruses infected cells or damaged cells
NKT cells	innate and adaptive	activation and suppression of innate and adaptive response

B lymphocytes

B cell development

B lymphocytes develop from hematopoietic stem cells in the BM. During B cell maturation, the production of an antigen-specific B-cell receptor is a key process. The maturation of the B-cell receptor is linked to the development of distinct B cell subsets (Figure 1). In the bone marrow, immunoglobulin (IG) gene rearrangement starts with

diversity (D)-joining (J) segment rearrangement in pro-B cells followed by variable (V) segment rearrangement. The resulting immunoglobulin heavy chain (IGH) protein is assembled with surrogate IG light chain making the pre-B cell antigen receptor (pre-BCR) complex. The latter is expressed in pre-B cells. The next step is IG light chain gene rearrangement. The IG light chain replaces surrogate IG light chain to give rise to a fully formed IG that is expressed on the surface of immature B cells, both as IgM and IgD isotypes. Subsequently, immature B cells migrate from the bone marrow to the blood and to secondary lymphoid tissues including lymph nodes and spleen. There the cells further differentiate into naïve follicular or marginal zone (MZ) B cells, during a process that is as yet incompletely understood but seems at least partly to be regulated by the strength of antigen binding to the B-cell receptor.

MZ B cells are generated in a T-cell independent rapid immune response and mature to short-lived plasma cells. These cells secrete pentameric natural or so-called non-immune IgM antibodies that play an important role in the defense against pathogens, but importantly, also in the clearance of apoptotic cells from the body.[2] Non-immune pentameric IgM, has unique properties. First, it is polyreactive and can bind multiple epitopes or antigens thereby increasing its binding to pathogens or apoptotic cells. Second, after binding to cell surface antigen, either pathogen or apoptotic cell, IgM binds the complement factor C1q better than any other antibody isotype. C1q is subsequently recognized by phagocytes leading to phagocytosis of the pathogen or apoptotic cell. Further, C1q can recruit other complement factors leading to activation of the classical complement pathway and generation of deposition of CD3b on the cell surface. The latter further enhances internalization by phagocytes. In conclusion, marginal zone cells are important as a first line of defense against pathogens, but also in clearance of apoptotic cells from the body and tissue homeostasis.

Follicular B cells move to the germinal center (GC) of B-cell follicles of secondary lymphoid tissues. There follicular B cells clonally expand while their rearranged immunoglobulin genes undergo somatic hypermutation (SHM) as well as class switch recombination (CSR). SHM is a random process that modifies the expressed immunoglobulin. By antigen-driven selection, B cells displaying an immunoglobulin with the most avid binding to antigen are selected and will further mature. Antigen-induced GC B cell activation is a T cell dependent process. CSR also takes place in the germinal center. During this process, the constant region of immunoglobulin genes are rearranged to produce either immune IgM, IgA, IgG and less frequently IgE. B cells exiting the germinal center mature into long-lived plasma or switched memory B cells. These cells can further circulate and

migrate to the BM or secondary lymphoid organs. Immune IgM, IgA and IgG bind antigens more avidly. The functions of these antibodies are summarized in Table 2.

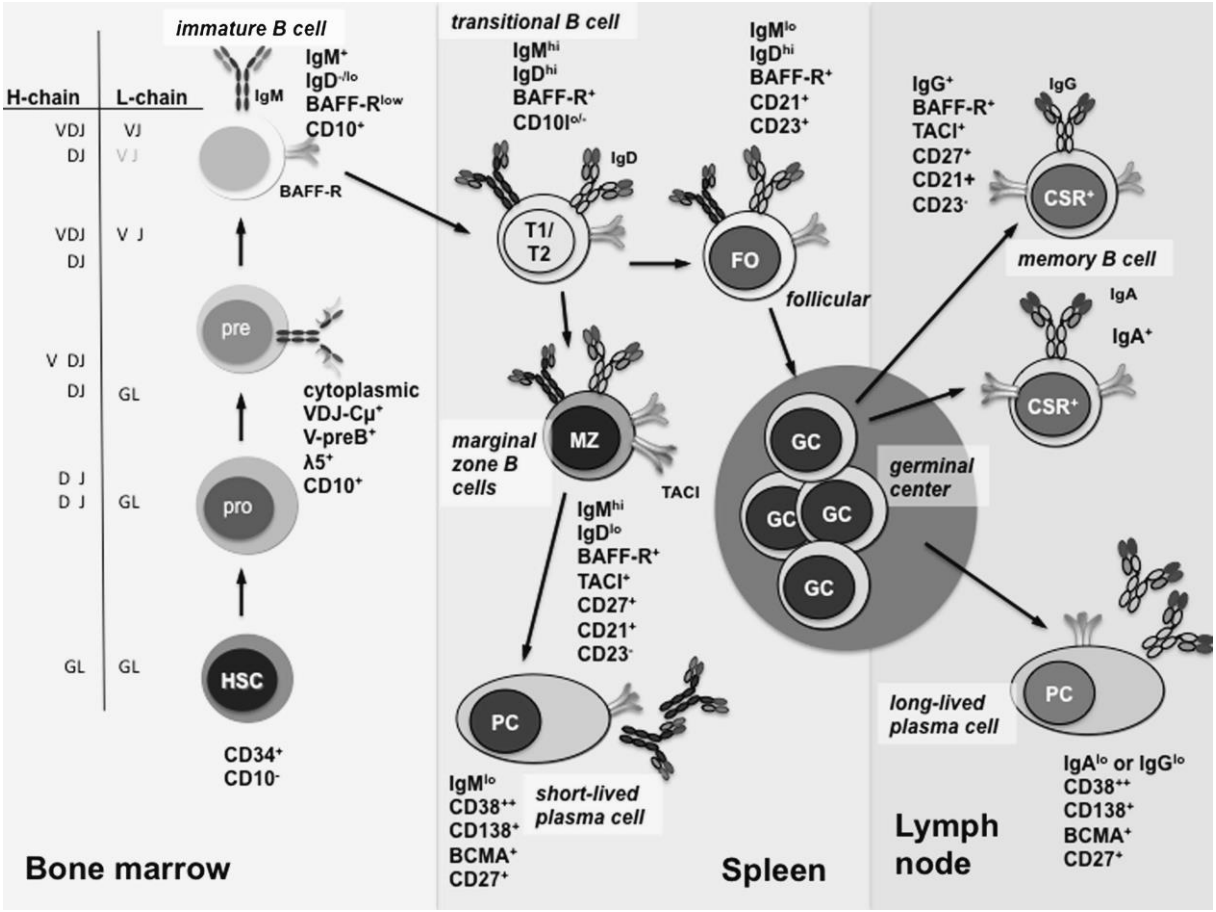


Figure 1. B cell development and B cell subsets (reprinted with permission [3]).

Table 2. Antibody type and function (adapted from [1]).

Antibody isotype	Antibody function
IgM	Complement activation (classical pathway)
IgG	Opsonization of antigens, Complement activation (classical pathway), Neonatal immunity (transfer of maternal antibody), Antibody dependent cell-mediated cytotoxicity, Feedback inhibition of B cell activation
IgD	Naïve B lymphocytes antigen receptor
IgA	Mucosal immunity, Complement activation by the lectin or the alternative pathway
IgE	Defense against parasites, Immediate hypersensitivity

IG structure and diversity

IG consist of two identical heavy chains and two identical light chains (Figure 2). The N-terminal parts of heavy and light chains comprise the antigen-binding site. The C-terminal part of the heavy chain comprises the constant regions that mediate effector functions, as summarized in Table 2.

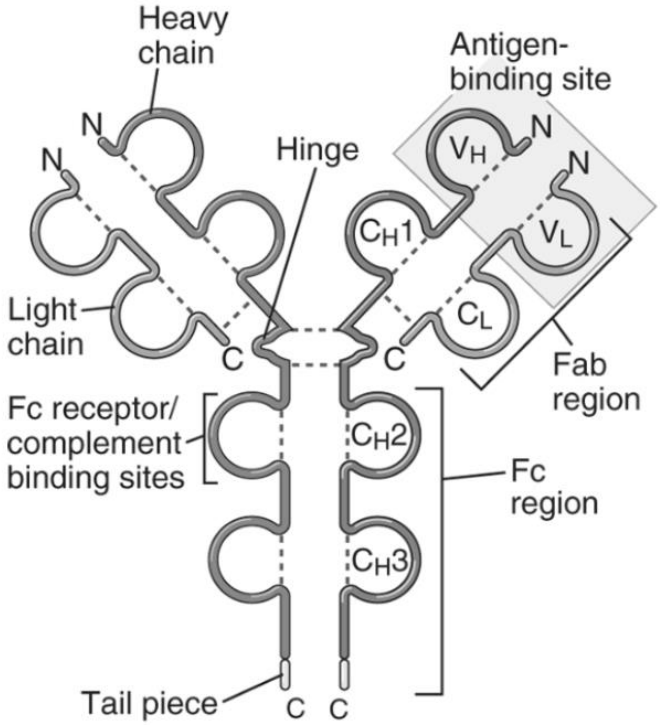
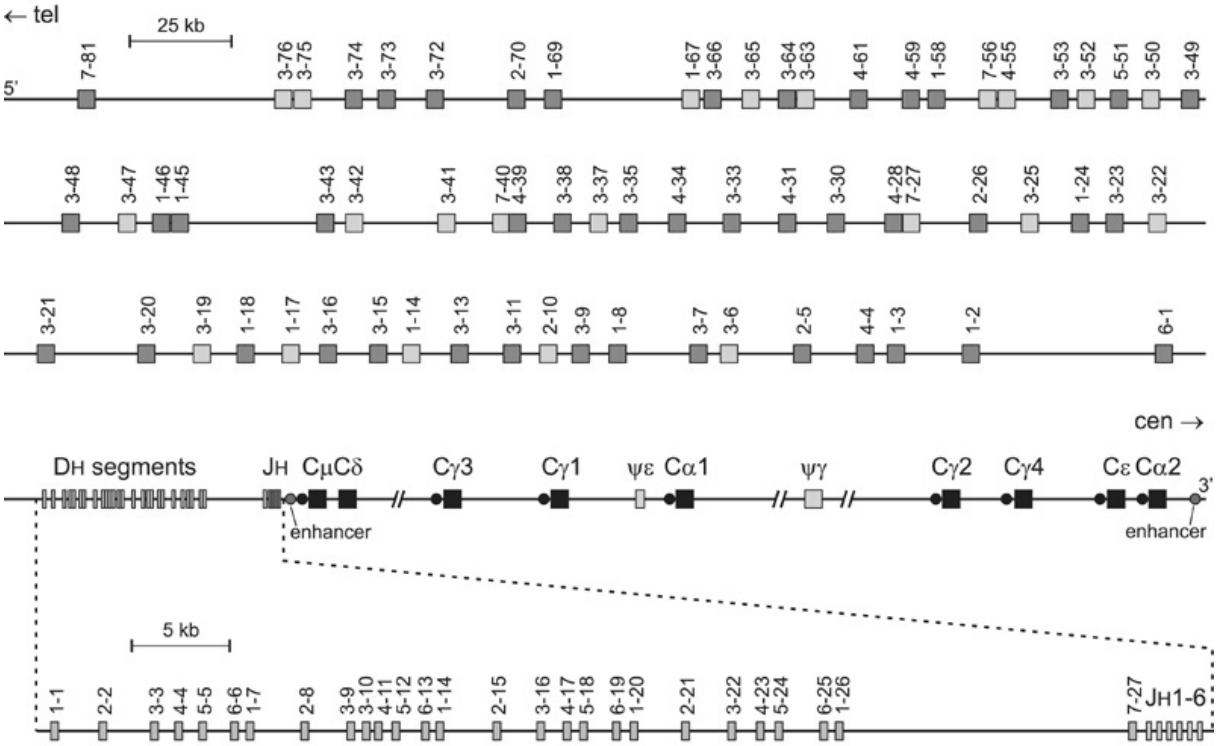


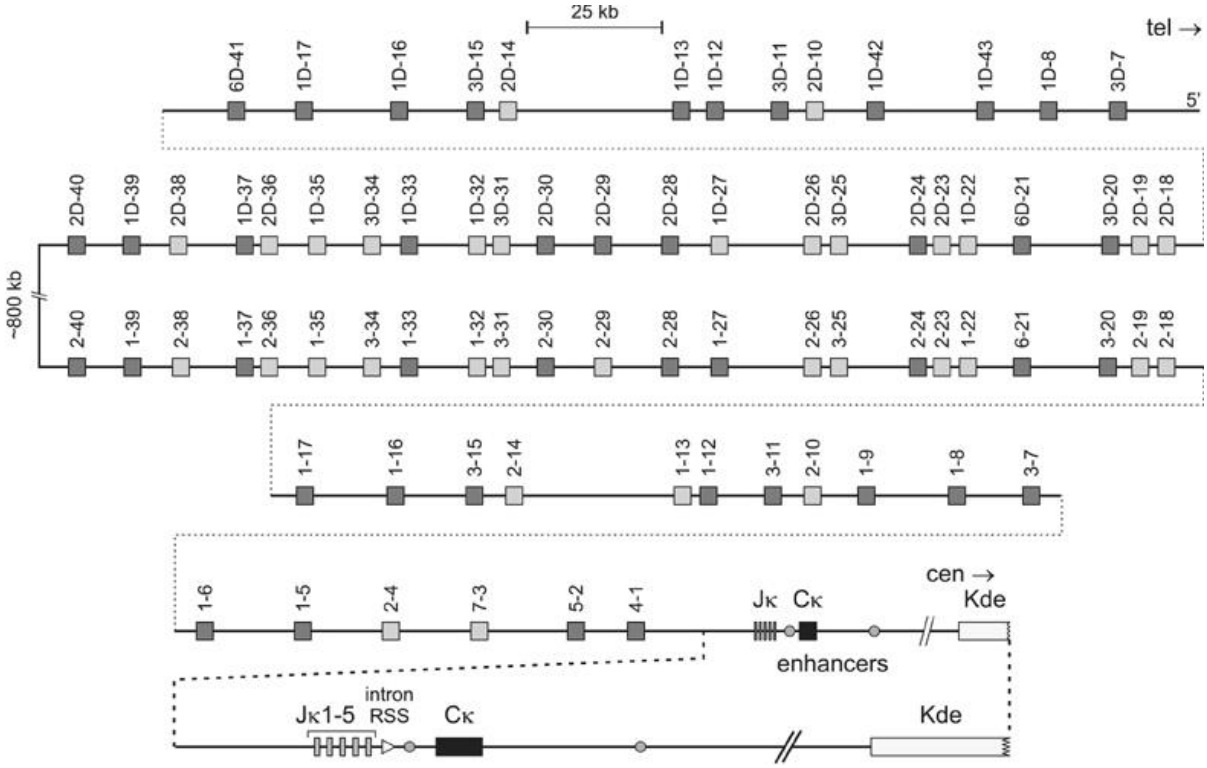
Figure 2. IG structure of secreted IgG (reprinted with permision[1]).

The human *IGH* locus, encoding the IGH protein, is located on chromosome 14q32.3 and covers approximately 1250 kb. It consists of 123 to 129 *IGHV*, 27 *IGHD*, 9 *IGHJ* and 11 *IGH* constant (*C*) genes (Figure 3a).[4,5] Only about half of the *IGHV* genes are functional.
The human *IG* kappa light chain (*IGK*) locus is on chromosome 2p11.2 and spans 1820 kb. There are 76 *IGKV*, 5 *IGKJ* and one *IGKC* genes (Figure 3b).[4,6] Human *IG* lambda light chain (*IGL*) locus is located on chromosome 22q11.2 and spans 1050 kb. It has 73 to 74 *IGLV* genes, 7 to 11 *IGLJ* genes and 7 to 11 *IGLC* genes (Figure 3c).[4,7]

a) *IGH* gene complex (#14q32.3)



b) *IGK* gene complex (#2p11.2)



c) *IGL* gene complex (#22q11.2)

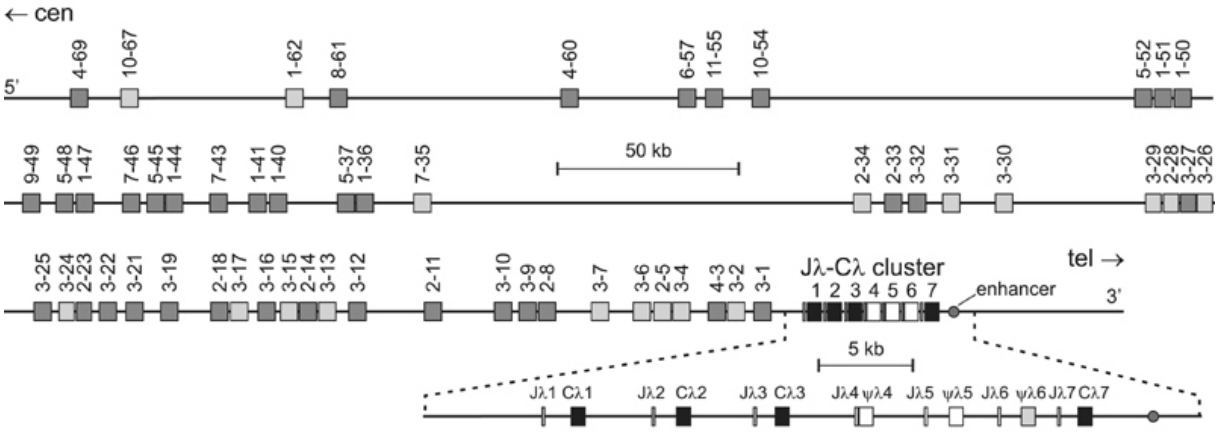


Figure 3. Schematic diagram of IG genes (reprinted with permission [4])

a) *IGH* gene complex on chromosome band 14q32.3

b) *IGK* gene complex on chromosome band 2p11.2

c) *IGL* gene complex on chromosome band 22q11.2

Variable regions of both heavy and light chain genes consist of 3 hypervariable segments called complementarity determining regions 1, 2 and 3 (CDR1, CDR2, CDR3), and 3 relatively conserved regions called framework regions 1, 2 and 3 (FR1, FR2, FR3). The antigen-binding pocket is encoded by the CDR1, CDR2 and CDR3 regions from both heavy and light chains.

IG gene rearrangement creates IG diversity. This process is capable of generating antibodies that recognize more than 5 x 10^{13} antigens. It starts with *IGHD-IGHJ* gene rearrangement followed by *IGHV* gene rearrangement. During the joining of these genes, some base pairs can be introduced or deleted adding further to the diversity. If the rearrangement of the first allele of the *IG* gene does not result in a productive protein, the second allele is rearranged. If the first allele is productive, the expression of the corresponding pre-BCR, composed of the IGH chain and surrogate IG light chain, prevents the second allele from rearranging. This process, called allelic exclusion, results in one cell producing only one antibody. The expression of the pre-BCR initiates rearrangement of the *IG* light chain genes. *IG* light chains show less diversity since antigen-binding regions are only encoded by *V* and *J* genes and not by *D* genes.

After *IG* genes are uniquely rearranged, further diversity is introduced by SHM. By this process, antibodies with low or intermediate affinities can be transformed into high affinity antibodies. SHM occurs in germinal centers where B cells undergo clonal expansion

during a T-cell dependent response. This process is mediated by the enzyme activation-induced cytidine deaminase (AID) and targets hotspots. The main hotspot motifs are RGYW, WRCY and WA (R=A/G, Y=C/T, W=A/T).[8-11]

The effector function of the IG is determined by the constant region. The latter defines the isotype of the immunoglobulin. Constant region genes are rearranged in the germinal center. During the process, the *VDJ* segment is rearranged to either μ, ε, α or γ *IGH* genes, allowing expression of either immune IgM, IgE, IgA or IgG antibodies, respectively. This process is also dependent on the AID enzyme, among others. The various antibody constant chains exert different roles in the humoral immune response as summarized in Table 2.

B lymphocyte activation

B-cell receptor (BCR) signaling

Normal B cells have unique BCRs consisting of dimers of IG heavy and light chains. Upon contact with antigen, antigen-induced aggregation of BCR causes downstream processes that start with phosphorylation of parts of CD79A and CD79B molecules, the so-called immuno-receptor tyrosine-based activation motifs or ITAMs. Spleen tyrosine kinase SYK is subsequently recruited to the ITAMs and activates multiple pathways: mitogen-activated protein kinase (MAPK), nuclear factor of activated T cells (NFAT), AKT/mammalian target of rapamycin (mTOR) and nuclear factor-kappa B (NF-κB) pathways.[12] One of the important genes in BCR signaling to the NF-κB pathway is the scaffold protein caspase recruitment domain family member 11 (CARD11). It is kept in an inactive state by an inhibitory domain until it receives a signal from the BCR. BCR activation induces phosphorylation of CARD11[13] and results in recruitment of B-cell CLL/lymphoma 10 (BCL10) and the paracaspase, mucosa-associated lymphoid tissue lymphoma translocation gene 1 (MALT1). The formation of the CARD11/BCL10/MALT1 (CBM) complex (Figure 4) leads to IκB kinase (IKK) and c-Jun N-terminal kinase (JNK) activation.[14]

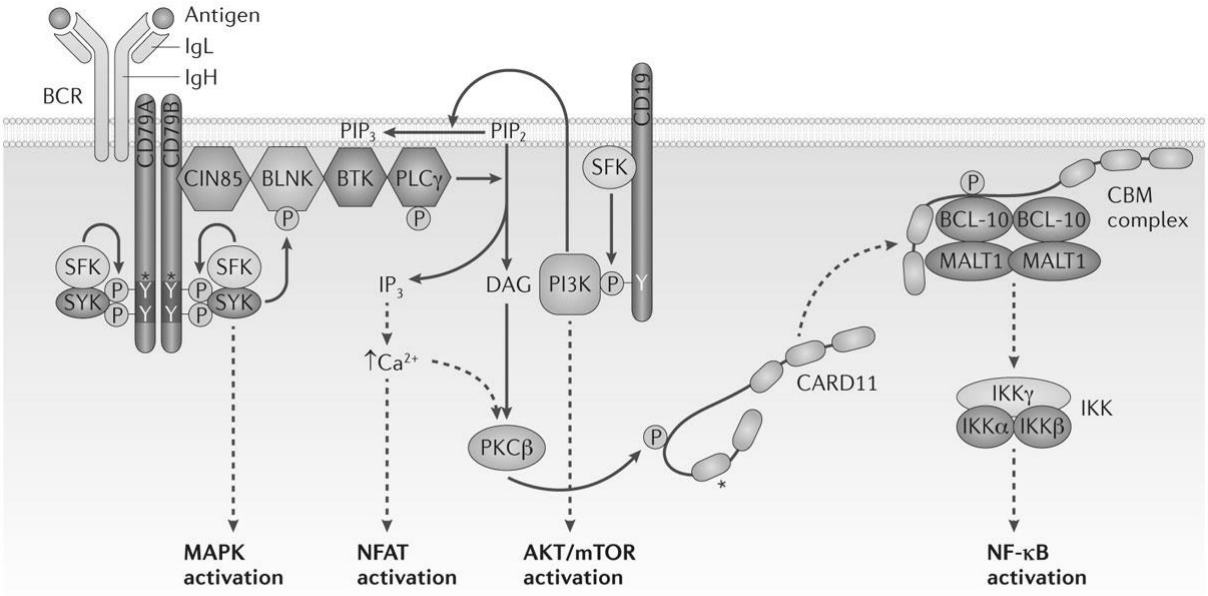


Figure 4. B cell receptor (BCR) signaling. The figure shows the co-receptor CD19 and various signaling intermediates (including CARD11) that are engaged following binding of the BCR to antigen. Several downstream pathways are triggered (reprinted with permission[12]).

Toll-like Receptor (TLR) signaling

TLRs are pattern recognition receptors that recognize molecules expressed by microbial cells or endogenous molecules that indicate cell damage. TLRs play an important role in inflammatory responses and adaptive immune responses (Figure 5). TLRs can be located on the cell surface or on the surface of endosomes within the cell (TLR 3, 7 and 9). The TLR signaling pathways activate diverse transcription factors: NF-κB, activation protein 1 (AP-1), interferon response factor 3 (IRF3) and 7 (IRF7). All TLRs with the exception of TLR3 signal through myeloid differentiation primary response 88 (MYD88) and can activate NF-κB and induce an inflammatory response. TLR3 signals through TIR-domain-containing adaptor protein inducing IFNβ (TRIF), activates IRF3 and induces expression of type I interferons. TLR4 can signal through both MYD88 and TRIF and is able to induce both responses.[1] TLR signaling is involved in B-cell activation and differentiation into plasma cells. TLR signaling also plays a role in autoantibody production. Synergistic effect of TLR9 and BCR signaling can activate autoreactive cells in vitro.[15] Deficiency of MYD88, TLR7 and TLR9 can result in reduced autoantibody production.[16]

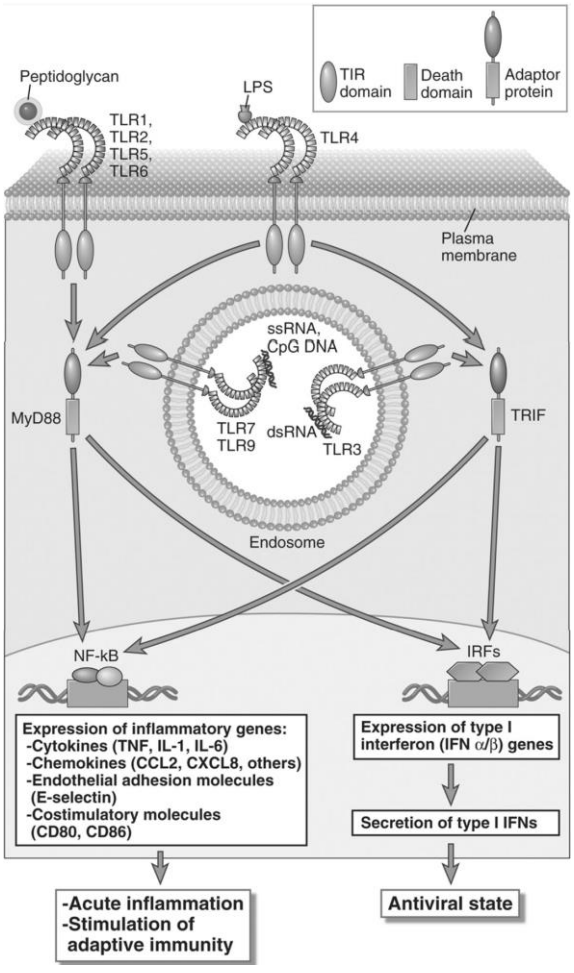


Figure 5. Signaling functions of TLRs (reprinted with permission[1]).

B-cell lymphoma

B-cell lymphomas are tumors that arise from B lymphocytes and include many different disease entities. These entities are described by the World Health Organization (WHO), most recently in the 2016 classification of Tumours of Hematopoietic and Lymphoid Tissues.[17] These tumors have unique biologies and by various mechanisms abnormally activate signaling pathways used by normal B lymphocytes, resulting in neoplastic tumour growth. The diagnosis of these diseases is complex and is made by integration of results obtained by histology, immunophenotypic features, clinical features and genetic features. Immature B-cell neoplasms and mature B-cell neoplasms are summarized in new WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues (Revised 4th edition).[17]

B-cell lymphomas may be thought of as originating from different stages of B cell development.[18] One of the first stages of B cell development, pre-B cells are the cell of origin of B-lymphoblastic leukemia/lymphoma (B-ALL). B-ALL may display a variety of gene translocations or mutations. As B cells mature, they become naïve B cells. These B cells are the cell of origin of chronic lymphocytic leukemia (CLL) and mantle cell lymphoma (MCL). Mature B cells can then differentiate into marginal zone B cells or germinal center B cells. Marginal zone B cells give rise to marginal zone lymphoma (MZL). Several types of MZL are recognized and have a predilection for different organs. Splenic marginal zone lymphoma (SMZL) arises in the spleen; mucosa-associated lymphoid tissue (MALT) lymphoma arise in mucosal tissues and nodal marginal zone lymphoma (NMZL) in lymph nodes. Germinal center B cells are the cell of origin for diverse lymphoma types: follicular lymphoma (FL), diffuse large B-cell lymphoma (DLBCL) and Burkitt lymphoma. Memory B cells are the cell of origin for CLL displaying immunoglobulin genes that are somatically mutated. Plasma cells are the cell of origin of multiple myeloma.

Recurrent genetic alterations characterize many types of B-cell lymphoma. Gene translocations often involving both immunoglobulin genes and oncogenes, as well as oncogene mutations (Table 3) are typically seen.

Table 3. Frequently mutated genes in small B-cell lymphomas and DLBCL.

Lymphoma type	**Frequently mutated genes (frequency of mutations - %)**	**References**
Chronic lymphocytic leukemia/small lymphocytic lymphoma (CLL/SLL)	*SF3B1* (15–20%) *NOTCH1* (Up to 15%) *TP53* (7–15%) *ATM* (9–12%) *MYD88* (3–10%) *FBXW7* (4%)	19-24
Follicular lymphoma (FL)	*KMT2D* (*MLL2*) (80–90%) *BCL2* (hypermutated in 76%) *CREBBP* (30–70%) *TNFRSF14* (32%) *EZH2* (7–22%) *ARID1A* (15%) *MEF2B* (10–15%) *EP300* (9–14%) *CARD11* (12%) *STAT6* (11%) *FOXO1*	24-32
Splenic marginal zone lymphomas (SMZL)	*NOTCH2* (20-25%) *TNFAIP3* (*A20*) (13%) *BIRC3* (11%) *TRAF3* (10%) *NIK* (8%)	33-36
Nodal marginal zone lymphoma (NMZL)	*KMT2D* (*MLL2*) (34%) *PTPRD* (20%) *NOTCH2* (20%) *KLF2* (17%)	37
Marginal zone lymphomas (MZLs)	*TNFAIP3* (*A20*) (19%)	38
Lymphoplasmacytic lymphoma (LPL)	*MYD88* (80-90%) *CXCR4* (27%) *ARID1A* (17%)	24,39-42
Mantle cell lymphoma (MCL)	*CCND1* (35-65%) *ATM* (43%) *TP53* (15%) *KMT2D* (*MLL2*) (14%) *NOTCH1* (5-12%) *WHSC1* (10%)	36,43-48
Diffuse large B-cell lymphoma (DLBCL) (subtypes: activated B-cell (ABC) and germinal center B-cell (GCB)) (* % according to Reddy et al., 2017) [60]	*KMT2D* (*MLL2*) (25%)* *BCL2, MYD88, HIST1H1E, PIM1* (17%)* *CREBBP, CARD11* (11%)* *SPEN, TP53, ARID1A* (10%)* *TNFRSF14, SOCS1, CDKN2A* (9%)* *NOTCH2, ARID1B, SETD1B* (8%)* *GNA13, SMARCA4, SGK1, MGA* (8%)* *CREBBP* (~32% of GCB) *EZH2* (22% of GCB) *TNFAIP3* (A20) (23-54% of ABC) *MYD88* (30-40% of ABC) *PRDM1*/*BLIMP1* (~20-30% of ABC) *CD79A*/*B* (~21-23% of ABC) *CARD11* (~8-10 of ABC)	24,28,29,31,49-60

Pre-malignant B-cell lymphoma lesions

Increasingly, pre-malignant B-cell non-Hodgkin lymphoma lesions have been recognized and have since been included in the WHO 2016 classification. The lesions include: monoclonal B-cell lymphocytosis (MBL); in situ follicular lymphoma; in situ mantle cell lymphoma; monoclonal gammopathy of uncertain significance (MGUS) and, IgM MGUS.

MBL and clonal B-cell lymphocytosis of marginal zone origin (CBL-MZ)

MBL is defined as small clonal B-cells detected at low levels (<5 x10^9/L) in the blood of otherwise healthy people.[17,61] MBL has been detected in the blood of elderly patients, with an incidence rate of more than 10%, dependent on the sensitivity of the flow cytometry analysis used.[62] MBL with a CLL immunophenotype has been demonstrated to be a precursor state of CLL.[63,64] Not surprisingly, MBL shows similar genetic lesions as also seen in CLL/SLL such as del(13q) and trisomy 12. MBL probably precedes all cases of CLL/SLL.[63] However, most patients with MBL of CLL-type will never develop CLL. The revised 2016 WHO classification emphasizes the difference between low count MBL (<0.5 x 10^9/L) and high count MBL. Low count MBL has significant differences from CLL, and it does not require follow up since it has very low risk of progression.[65,66]

MBL with a non-CLL immunophenotype, recently renamed to clonal B-cell lymphocytosis of marginal zone origin (CBL-MZ)[67] has not been demonstrated to be a precursor state of CLL/SLL. CBL-MZ is mostly an indolent proliferation that rarely seems to progress to lymphoma. However, rare cases do progress, predominantly to splenic marginal zone lymphoma.[67] More studies are needed to confirm this. It is of interest that CBL-MZ predominantly rearranges the *IGHV4-34* gene, in contrast to SMZL, but more in line with splenic diffuse red pulp lymphoma.[68] Mutational analysis identified mutations of *NOTCH2*, *CD79b*, *TNFAIP3* and *MYD88*, indicating similarities to SMZL and lymphoplasmacytic lymphoma (LPL).[69] Larger series will need to be studied to better understand disease progression in CBL-MZ.

Patients with MBL almost invariably show similar cells in the bone marrow.[70] A high incidence of monoclonal small B-cells (MSBC) consistent with MBL and CBL-MZ was reported by Tierens et al. in the bone marrow of patients with DLBCL.[71] Of interest, activated B-cell (ABC) DLBCL showed a higher frequency of MSBC in the bone marrow than germinal center B-cell (GCB) DLBCL, 28,2% versus 3,7%, respectively (p=0,0002). In

addition, other DLBCL types such as DLBCL, leg type and primary DLBCL of the central nervous system, that are also of ABC origin, showed a high incidence of MSBC.[71] Whether MSBC and DLBCL were clonally related, i.e. whether MSBC may be considered a precursor lesion for DLBCL, was not analyzed, except for one case, in the retrospective series by Tierens et al. and has been a topic of investigation in this thesis.

In situ follicular lymphoma

Circulating B lymphocytes with the *BCL2* gene translocation, typical of FL, are detected at high frequency in the elderly population. Most of the patients do not progress to FL. Equally, B lymphocytes with *BCL2* translocation can be detected in germinal centers of otherwise reactive lymphoid tissues. This is called 'in situ' follicular lymphoma. Circulating B cells with *BCL2* translocation and in situ follicular lymphoma are considered a precursor lesion of FL, with a very low incidence of progression to overt lymphoma.[72]

In situ mantle cell lymphoma

In situ mantle cell lymphoma is defined by the minimal presence of cells with the typical t(11:14) with cyclin D1 overexpression in lymphoid tissues in patients with no clinical evidence of lymphoma. It is mostly an incidental finding in lymphoid tissues investigated in the course of other diseases. It is a rare occurrence and seems to have a higher degree of progression to clinically overt mantle cell lymphoma.[73]

MGUS and smoldering multiple myeloma

MGUS and smoldering multiple myeloma are asymptomatic plasma cell neoplasias that may progress to clinically overt multiple myeloma. MGUS progresses to multiple myeloma at 1% per year, whereas smoldering multiple myeloma, characterized by a higher number of monoclonal plasma cells has a higher incidence of transformation to multiple myeloma (reviewed by Korde et al.[74]).

IgM MGUS and smoldering lymphoplasmacytic lymphoma

IgM MGUS and smoldering lymphoplasmacytic lymphoma are characterized by a low level monoclonal IgM peak in the serum and a low level of lymphoma involvement in the bone marrow. These conditions, show the typical mutations in *MYD88* and *CXCR4* genes as seen in LPL. IgM MGUS progresses to overt LPL, at a rate similar to that of MGUS and smoldering multiple myeloma progression to myeloma (reviewed by Mailankody et al.[75]) .

Transformation of low grade B-cell lymphoma to high-grade B-cell lymphoma

Histological transformation of low-grade lymphoma into high grade lymphoma is a well-known occurrence (Table 4).[76] Histological transformation of FL to DLBCL has been most studied. It occurs at a relatively high frequency, with a risk of about 30% at 10 years. Pasqualucci et al.[77] as well as others [78] have shown that transformed FL arises from common precursor cells as FL, but does not directly develop from FL. Precursor cells usually have mutations in epigenetic modifiers (e.g. Lysine Methyltransferase 2D (*KMT2D*)) and antiapoptotic genes and share the t(14;18) with transformed lymphoma. Precursor cells acquire further mutations with time. Of interest, transformed FL shows similar mutations as the ones also identified in de novo GC DLBCL.[77]

Transformation from CLL to DLBCL, or so-called Richter syndrome, is also well-known. It occurs in about 2% - 10% of CLL patients. Important genetic events upon transformation involve *CDKN2A* loss, *TP53* disruption, *C-MYC* activation, and *NOTCH1* mutations.[79,80] MZL may also transform to DLBCL. With regard to extranodal MZL of the stomach, Starostik et al.[81,82] showed that t(11;18)(q21;q21)-positive lymphoma, in contrast to t(11;18)(q21;q21)-negative lymphoma, does not transform to DLBCL. The reasons for this are not entirely clear, but may involve dependence on continued antigen-stimulation in the latter, but not in the former. Finally, transformation of LPL and nodular lymphocyte predominant Hodgkin lymphoma to DLBCL have also been described.[83]

Table 4. Histologic Transformation in B-Cell Lymphoma (adapted from Montoto et al.[83]).

Indolent Lymphoma	**Transformed Lymphoma**
Follicular lymphoma	Diffuse large B-cell lymphoma
Follicular lymphoma	High grade B-cell lymphoma with *MYC* and *BCL2* or *BCL6* translocation
Small lymphocytic lymphoma/chronic lymphocytic leukemia	Diffuse large B-cell lymphoma
Small lymphocytic lymphoma/chronic lymphocytic leukemia	Hodgkin lymphoma
Lymphoplasmacytic lymphoma	Diffuse large B-cell lymphoma
Mucosa-associated lymphoid tissue lymphoma	Diffuse large B-cell lymphoma
Nodular lymphocyte-predominant Hodgkin's lymphoma	Diffuse large B-cell lymphoma

DLBCL

DLBCL accounts for about 30% of all adult non-Hodgkin lymphomas. Distinct clinical and genetic entities are recognized.[61] The most common type is DLBCL, not otherwise specified (NOS). DLBCL, NOS is further divided in two subtypes according to cell of origin, either from activated B cells or from germinal center B cells.[84] DLBCL ABC has a worse overall and failure-free survival in comparison with GCB DLBCL.[85] DLBCL ABC is characterized by constitutive activation of NF-κB through activating mutations in the B-cell receptor and Toll-like receptor pathways (reviewed by Lenz and Staudt [86] and Ngo et al.[57]). Therefore, patients with DLBCL ABC may profit from novel treatment modalities targeting B-cell receptor cell signaling, such as Bruton kinase inhibitors.[87]

A high incidence of MSBC/MBL in the bone marrow of patients with DLBCL, especially DLBCL ABC, was reported by Tierens et al.[71] It was also demonstrated that the immunophenotype of the MSBC/MBL in DLBCL most frequently is that of non-CLL type than of CLL type, the latter being the most frequent MBL type in the general population.[88]

Primary cold agglutinin disease (CAD)

CAD is a hemolytic anemia mediated by monoclonal IgM anti-I autoantibodies. I antigen is expressed on most cells, including red blood cells. CAD represents 15% of all cases of autoimmune hemolytic anemia with an incidence of 1×10^{-6} per year. Patients suffer from anemia as well as circulatory problems, although the severity of disease varies greatly between patients as recently reviewed[89]. Anemia is caused by binding of the antibodies to erythrocyte cell surface I antigens at low temperatures, subsequent activation of complement with C3b complement deposition on the surface and ultimate destruction of the erythrocytes in liver and spleen. [90,91] Previously, it was shown that patients with CAD have an underlying clonal B-cell disorder,[92,93] and recently Randen et al. demonstrated that CAD is caused by a low grade B-cell lymphoproliferative disease of the bone marrow with a typical histology that is different from LPL and, accordingly, does not display the *MYD88* L265P mutation.[94] CAD is therefore unlike IgM MGUS. The mutational landscape of CAD has not been studied except for showing the absence of *MYD88* L265P mutation.[95] The molecular changes underlying most B-cell lymphoproliferative diseases have now been extensively mapped, largely through recent genome-wide genetic studies.[96] Of interest, many of these changes can be specifically targeted by novel drugs.[36] Since we ascertained that a clonal B-cell

lymphoproliferative disease is the cause of CAD, it is reasonable to assume that genetic changes may be underlying this disease. This study was part of this thesis project.

Almost all patients display circulating monoclonal antibodies encoded by *IGHV4-34*. The framework region 1 (FR1) of IGHV4-34 encodes Gln^{6}-Trp^{7} (QW) and Ala^{24}-Val^{25}-Tyr^{26} (AVY) sequences that determine binding to I antigen.[97] Irrespective of this common finding, patients present with varying severity of disease. Cold agglutinin disease activity does not correlate with antibody titers, but seems to be determined by the thermal amplitude of the agglutinin, i.e., the highest temperature at which the cold agglutinin binds to I antigen. This cannot be explained by binding of I antigen to the FR1 of the *IGHV4-34* encoded antibody since it is common to all patients. The other molecular features that may influence antigen binding were hitherto unknown and were therefore studied during this thesis project.

Current treatment of CAD depends on the severity of the disease, and consists of avoidance of exposure to low temperatures, transfusions when necessary and pharmacotherapy. The latter includes rituximab, combined with fludarabine for those patients not responding to rituximab monotherapy. However, fludarabine treatment results in significant toxicity.[98] Chemoimmunotherapy with rituximab and bendamustin is better tolerated and is at least as effective as rituximab and fludarabine.[99]

Treatment of lymphoma

More than sixty distinct lymphoma entities with a different biology and clinical outcome are recognized. Standard therapy of these diseases comprises radiotherapy, chemotherapy and immunotherapy, depending on the type and stage of the disease. High-dose chemotherapy with stem cell rescue is also used for some types of clinically aggressive or recurrent disease. One of the most frequent chemo-immunotherapy regimens used in lymphoma is R-CHOP ((R)ituximab, (C)yclophosphamide, (H)ydroxydaunorubicin, (O)ncovin, (P)rednisone or (P)rednisolone).[100]

More recently, many new therapy modalities have been developed including novel chemotherapy, monoclonal antibody and antibody–drug conjugates therapy, radioimmunotherapy, treatment with small-molecule inhibitors targeting cell signaling pathways, inducers of apoptosis and histone deacetylase (HDAC) inhibitors (Figure 6).[100] Newer immunotherapies, such as treatments that target the immune system checkpoints or using T-cells with altered antigen receptors, have also shown promise for treatment of lymphomas.[101,102] Many of the novel therapies are being studied in clinical trials.

Development of novel drugs has been greatly facilitated by the study of lymphomas by next-generation sequencing (NGS).[96] Drugs have been developed to target specific mutations or pathways that are altered by these mutations.[36] One of the examples is MALT1 inhibitors for treatment of ABC DLBCL. ABC DLBCL is addicted to NF-κB signaling through the CARD11-BCL10-MALT1 complex.[103] Another example are HDAC inhibitors, counteracting the effects of histone methyltransferase malfunction.[104] Inactivating mutations in *KMT2D* gene, encoding a histone methyltransferase, are frequent in FL, NMZL and DLBCL.[29,49,77,105]

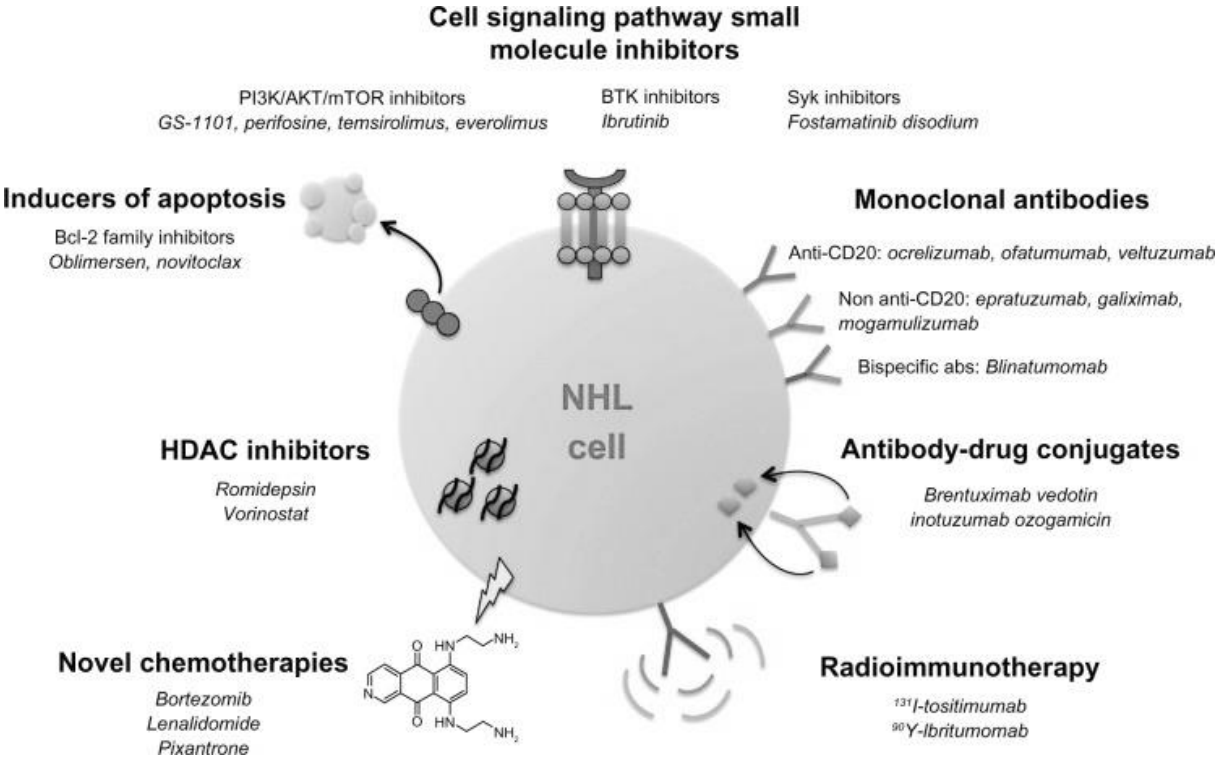


Figure 6. New therapeutic options in NHLs (reprinted with permission[100]).

Aims

The overall aim was to better characterize clonal small B cells in the bone marrow of patients with DLBCL and CAD that did not represent secondary involvement by small B-cell lymphoma.

1. To study whether clonal small B cells in the bone marrow in patients diagnosed with DLBCL were clonally related, and may therefore potentially represent precursor cells.
2. To study whether clonal small B cells in CAD are the source of auto-immune antibody production and whether the features of the immunoglobulin correlate with disease activity.
3. To study clonal small B cells in CAD by gene mutation analysis and thereby to investigate whether these B cells derive from known B-cell lymphoma types such as LPL or represent a novel B-cell lymphoproliferative disease.

Methodological considerations

Patients samples

Diffuse large B-cell lymphoma (paper 1)

We prospectively collected blood and bone marrow samples from patients with primary DLBCL and without histological BM infiltration with large B-cell lymphoma. All patients were diagnosed and treated at the Oslo University Hospital, Oslo, Norway. The patient characteristics are provided in Table 5. For all patients, 20ml blood and 10ml bone marrow were collected. Diagnostic formalin-fixed lymphoma tissue was available for all patients. Snap-frozen lymphoma tissue, stored at -80°C, was available for only one patient. In addition, a bone marrow trephine biopsy was procured for staging purposes. The study was approved by the Regional Committee for Medical and Health Research Ethics of South-East Norway (REK SØ 2010/3241).

Table 5. DLBCL patient characteristics

Patient	Age	LDH[1]	Number of extranodal sites	Stage	WHO performance status	IPI[2]	Biopsy site
1	84	0.69	1	IEA	1	1	Rectal mucosa
2	60	0.89	1	IEA	1	0	Gastric mucosa
3	82	0.91	1	IIEA	1	1	Gastric mucosa
4	68	0.96	1	IVA	1	2	Gastric mucosa
5	80	1.2	1	IVA	1	3	Lymph node
6	75	1.18	2	IVA	0	4	Gingiva

[1]fraction of upper normal limit; [2]IPI: International Prognostic Index

Cold agglutinin disease (paper 2 and 3)

We collected bone marrow and blood from 27 patients with well-documented primary CAD. This series included 8 patients previously published by our group.[94] Clinical data were available for most of the patients and included hemoglobin-, lactate dehydrogenase- (LD), bilirubin- and IgM-levels, leukocyte-, lymphocyte-, reticulocyte, and thrombocyte counts, cold agglutinin (CA) titer, sex and age at diagnosis. Clinical data are provided in Table 6.

The project was approved by the Regional Committee for Medical and Health Research Ethics of South-East Norway (REK SØ 2012/131).

Table 6. Summary of clinical data.

Clinical characteristics	No. of patients*	median	range
Hemoglobin (g/dL)	26	9,0	4,9 – 13,9
Leukocytes (10^9/L)	26	6,8	3,1 – 20,7
Lymphocytes (10^9/L)	26	2,8	0,4 – 4,8
Reticulocytes (10^9/L)	24	140,5	76 - 256
Thrombocytes (10^9/L)	26	208,5	108 - 506
CA titer	24	2048**	128 – 256000**
IgM g/L	21	6,9	1,8 – 51,5
LD U/L	26	308,5	222 - 475
Bilirubin µmol/L	26	49,0	10 - 72
Age at diagnosis	25	70	52 - 84

* Clinical data for some of the patients were not available.

**Some laboratories did not report results above titer 2048.

Histology and immunohistology

Hematoxylin and eosin-stained sections of formalin-fixed lymphoma tissue and Zinc-formalin-fixed, formic-acid decalcified bone marrow trephine biopsies were made for routine diagnoses. All sections have been reviewed for the studies. Immunohistochemical analysis of lymphoma and bone marrow trephine sections were performed in all cases. Paraffin blocks were cut at 4–6 µm, dried overnight at 60°C and dewaxed in xylene prior to immunohistochemical staining. The following antibodies were used: antibodies against CD20, MUM1, BCL6, Ki67 (all from Dako Cytomation, Glostrup, Denmark), CD5, CD21, CD23, CD10, BCL2 (all from Novocastra, Newcastle, U.K.), CD3, cyclin D1 (Lab Vision/NeoMarkers, Fremont, CA), CD138 (Serotec, Kidlington, U.K.) and PAX-5 (Becton Dickinson, Franklin Lakes, NJ). Visualization was performed using the EnVision® detection system (Dako Cytomation) according to the manufacturer's instructions. Appropriate positive and negative controls were used. Lymphomas were diagnosed according to the WHO classification[61] and the DLBCL cell of origin was studied using the Hans algorithm.[106]

Flow cytometry of blood and bone marrow samples

Eight-color flow cytometry analysis was used with the following antibody combinations labeled with Pacific Blue/ e450 (PB/e450), Krome Orange (KO), FITC/ Pe / PercPCy5.5/ Phycoerithrin cyanine 7(PeCy7)/APC/ APC Hilite7 or APC/cyanine7 (APCH7/cy7): (1) CD20+CD4/CD45/CD8+Igλ/CD56+/Igκ/CD5/CD19+TCRγδ/CD38; (2) CD20/CD45/CD23/CD10/CD79b/CD19/CD200/CD43. Anti-CD56, anti-CD5, anti-CD3 and anti-CD79b were purchased from Becton-Dickinson (San José, CA, USA); anti-CD23 from Dako; anti-CD200 from eBioscience (San Diego, CA); anti-CD8, anti-Igk and anti-Igl from Cytognos (Salamanca, Spain) and the remaining of the antibodies from Beckman Coulter (Brea, CA). Flow cytometry analysis was performed on a LSRII instrument (Becton-Dickinson), using FACSDiva software (Becton-Dickinson).

Fluorescent activated cell sorting (FACS)

Mononuclear cell suspensions were made of bone marrow and blood samples using Leucosep® tubes (Greiner Bio-One North America, Inc.) according to manufacturer's recommendations. Cells were resuspended in PBS supplemented with 1% FCS and 10% DMSO and were subsequently frozen using an isopropanol chamber and stored in liquid nitrogen until FACS analysis.

For FACS analysis, the mononuclear cell suspensions were thawed and divided in aliquots of 0,5-1,0 x 10^6 cells/tube. The cells were washed with 2000 µl PBS with 0,5 % BSA (PAA laboratories GmbH, Austria) and stained for surface antigens with the following antibodies: anti-CD45 (clone J.33, Beckman Coulter), anti -CD20 (clone B9E9(HRC20), Beckman Coulter), anti-CD19 (clone J3-119, Beckman Coulter), anti-CD5 (clone L17F12, Becton-Dickinson (San Jose, CA)) and anti-CD10 (clone HI10a, Becton Dickinson) anti-λ and anti-κ (polyclonal antibodies, Cytognos (Salamanca, Spain)). Antibodies were conjugated to either fluoresceine thyocyanate (FITC), phycoerythrine (Pe), peridinin chlorophyll protein-cy5.5 (PerCP-Cy5.5), phycoerythrine cyanine 7 (PeCy7), allophycocyanin (APC), Pacific Blue or Krome Orange. After staining, the cell suspensions were incubated for 15 minutes in the dark at room temperature and washed with 2 ml PBS supplemented with 0,5 % BSA. Tubes with cell suspensions from the same patients were pooled and filtered through a 70 µm filter.

Stained samples were sorted with high-pressure settings using a FACS Aria IIu High speed sorter (Becton Dickinson) equipped with a 408 nm, 488 nm and a 633 nm laser.
Selection of MSBC/MBL and CAD monoclonal B cells for sorting was performed using Becton Dickinson FACSDiva software, starting with gating of viable cells using the forward scatter versus side scatter dot plot. Subsequently, T cells and B cells were gated out using a CD5 versus CD19 dot plot.

MSBC/MBL were separated from polyclonal B cells taking advantage of the aberrant B-cell phenotypes identified by flow cytometry analysis. The marker combination used for sorting are indicated for each patient in Table 7. Samples with very low cell numbers were sorted directly into RLT plus lysis buffer (Qiagen, Germany) to prevent loss of cells during centrifugation. Samples with relatively high cell numbers were sorted into PBS, then centrifuged and suspended into RLT plus lysis buffer before DNA extraction.

Table 7. Markers used for FACS

SAMPLE	Markers used for FACS
Patient 1 MSBC/MBL	CD19+, CD20dim, CD5+, IgL+
Patient 2 MSBC/MBL	CD19+, CD20+, CD5-, IgK+
Patient 3 MSBC/MBL	CD19+, CD20+, CD5-, IgL+
Patient 4 MSBC/MBL	CD19+, CD20dim, CD5+, IgK+
Patient 5 MSBC/MBL	CD19+, CD20+, CD5dim, IgK+
Patient 6 MSBC/MBL	CD19+, CD20+, CD5-, CD10+,

CAD monoclonal B cells were separated from the polyclonal B cells using the immunoglobulin light chain gate, taking advantage of the fact that B-cell clones show either κ or λ immunoglobulin light chain restriction (Figure 7). Additionally, T cells were collected as normal control for NGS analysis for each patient. Samples were sorted into PBS or RLT plus lysis buffer as explained for MSBC/MBL sorting.

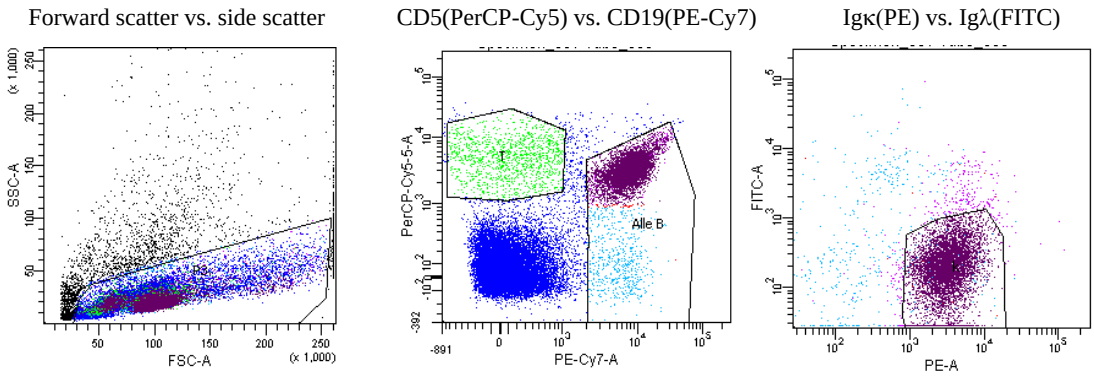


Figure 7: Sorting strategy for isolation of monoclonal B cells from bone marrow by flow cytometry. Step one: selection of lymphocytes by forward scatter vs side scatter; step two: separation of B cells and T cells by CD5 vs CD19 gating; step 3: selection of monoclonal B cells using the immunoglobulin light chain gate κ+.

DNA extraction and whole genome amplification

DNA from sorted cells was extracted using Qiagen AllPrep DNA/RNA Micro kit (Germany) according the instructions of the manufacturer with minor modifications. Since the amount of DNA was very limited additional incubation time and additional elution steps were added to recover all DNA. Due to limited amount of DNA for PCR analysis and Sanger sequencing genomic DNA was subsequently amplified using illustra Ready-To-Go GenomiPhi V3 DNA Amplification Kit (GE Healthcare Life Sciences, U.K.).

DNA from formalin-fixed paraffin-embedded tissue and fresh frozen tissue of DLBCL samples was extracted using Qiagen AllPrep DNA/RNA FFPE Kit and AllPrep DNA/RNA Mini kit according to manufacturer's recommendations.

The concentration of extracted nucleic acid was measured using a NanoDrop 2000 spectrophotometer (Thermo Scientific, Waltham, MA) and Qubit (Life Technologies).

For exome sequencing (CAD samples), five samples with sufficient DNA were used without any further amplification, while one sample with a low DNA yield, was amplified using Illustra Ready-To-Go GenomiPhi V3 DNA Amplification Kit (GE Healthcare Life Sciences, U.K.). For targeted sequencing (validation samples), 9 of 10 samples required whole genome amplification. We chose Illustra Ready-To-Go GenomiPhi V3 DNA Amplification Kit because it supposed to amplify DNA evenly and to cover the whole genome; also, samples amplified this way are suitable for exome sequencing. DNA was also amplified, as described above, for all samples for the purpose of verification of results by

Sanger sequencing. The low yield of DNA is due to the scant clonal B-cell infiltrate in the bone marrow, characteristic of most patients with CAD.[94]

Sequence analysis of rearranged immunoglobulin genes and proteins

Rearranged *IGH* genes were studied in order to analyze the clonal relationship between MSBC/MBL and DLBCL paired samples, and in order to better characterize these genes in CAD patients.

Rearranged *IGH* genes from all bone marrow B cell samples were amplified from DNA using the IGH Somatic Hypermutation Assay v2.0 (Invivoscribe Inc., San Diego, CA). The PCR products were subsequently sequenced using the BigDye® Terminator v1.1 Cycle Sequencing Kit (Life Technologies, Carlsbad, CA) and the primers from the IGH Somatic Hypermutation Assay v2.0 kit (Invivoscribe Inc.). The International Immunogenetics Information System web-based software (www.imgt.org) was used to analyze the rearranged *IGH* sequences. The entire analysis was repeated twice. In addition, sequencing was repeated with *IGHV* family-specific primers (Figure 8).[4]

Rearranged *IGH* genes from formalin-fixed paraffin embedded tissue DLBCL samples were analyzed using primers complimentary to *IGHV* framework 1, 2 and 3 as described before (Figure 8).[4] Sequencing and analysis of amplified *IGH* genes was as described above. A repeated analysis was subsequently performed using family-specific primers. The PCR products of this amplification were purified by agarose gel electrophoresis, followed by bacterial cloning (TOPO® TA Cloning -Life Technologies). Bacterial colonies were directly sequenced using vector-specific primers. At least 10 colonies per sample were analyzed.

Immunoglobulin light chain genes were amplified by an in-house diagnostic protocol using Biomed-2 primers[4] (Figure 9) and then sequenced. For cases in which rearranged *IGKV3* family genes were detected, additional *IGKV3* family-specific primers, designed to acquire longer PCR products, were used to confirm the findings.[107] All sequences were analyzed using the IMGT database (www.imgt.org).

To compare our results with those published in the literature, re-analysis of previously published sequences was necessary, because the nomenclature of immunoglobulin genes has changed during the last decades.[108,109-113] On-line protein analysis software IMGT (http://www.imgt.org/3Dstructure-DB/cgi/DomainGapAlign.cgi) and IgBLAST (http://www.ncbi.nlm.nih.gov/igblast/) were used for this purpose.

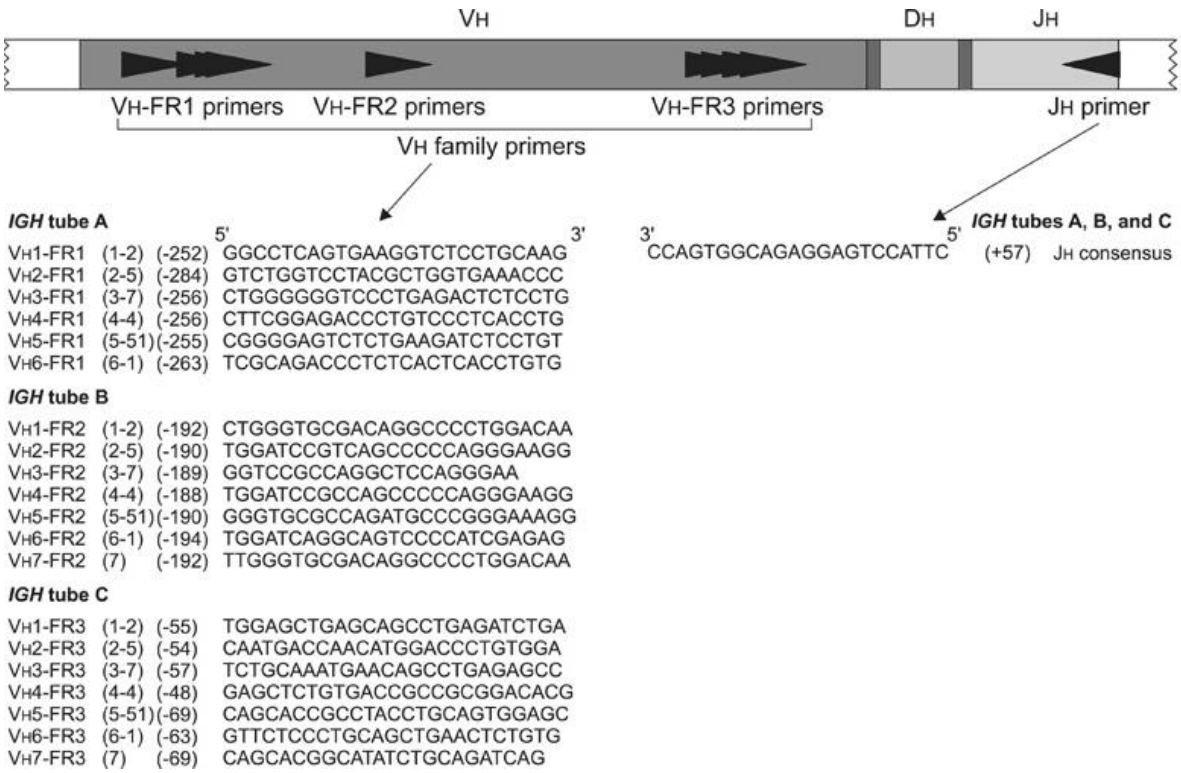


Figure 8. PCR analysis of *IGH* (*IGHV–IGHJ*) gene rearrangements (reprinted with permision[4]). Shown diagram of *IGHV–IGHJ* gene rearrangement with 3 sets of *IGHV* primers (one set for FR1, FR2 and FR3) and one *IGHJ* consensus primer.

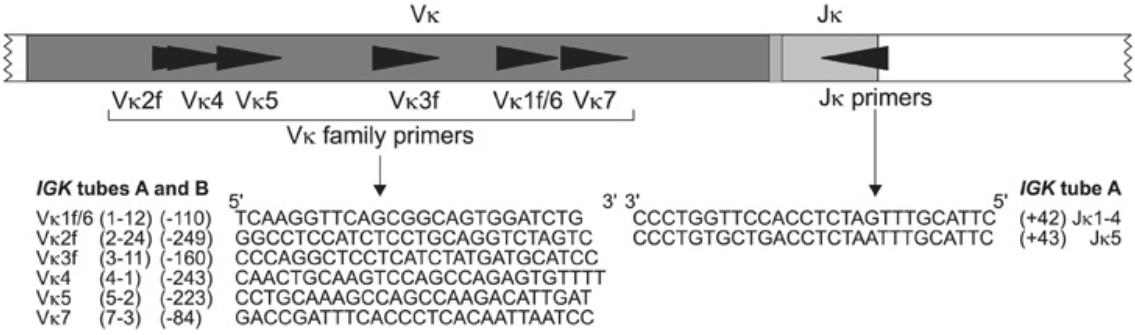


Figure 9. PCR analysis of *IGK* (*IGKV-IGKJ*) gene rearrangements (reprinted with permision[4]). Shown diagram of *IGKV–IGKJ* rearrangement with 6 *IGKV* primers and 2 *IGKJ* primers.

MYD88 L265P mutation analysis

A single nucleotide polymorphism, at amino acid position 265 of the *MYD88* gene (NM_002468) was detected using PCR and a SNaPshot mini-sequencing assay (Life Technologies). PCR was carried out using Phusion hot start DNA polymerase (Life Technologies) according to the supplier's instructions with the following PCR primers: 5'-TGC AGG TGC CCA TCA GAA GCG-3' and 5'-CAG ACA GTG ATG AAC CTC AGG ATG C-3'. Then a single nucleotide extension reaction was performed, according to the instructions of the manufacturer. The extension primers are as follows: 5'-CCC CCC CCC CAG GTG CCC ATC AGA AGC GAC-3' and 5'-CCT TGT ACT TGA TGG GGA TC-3'. PCR products were fractionated by capillary electrophoresis using a 3130 Genetic Analyzer and GeneMapper v.4.1 Software (Life Technologies) (Figure 10). The sensitivity of this *MYD88* L265P mutation analysis is 3%, determined by using dilution series of DNA from the *MYD88* L265P positive ABC DLBCL cell line OCI-Ly10 in DNA from normal blood.

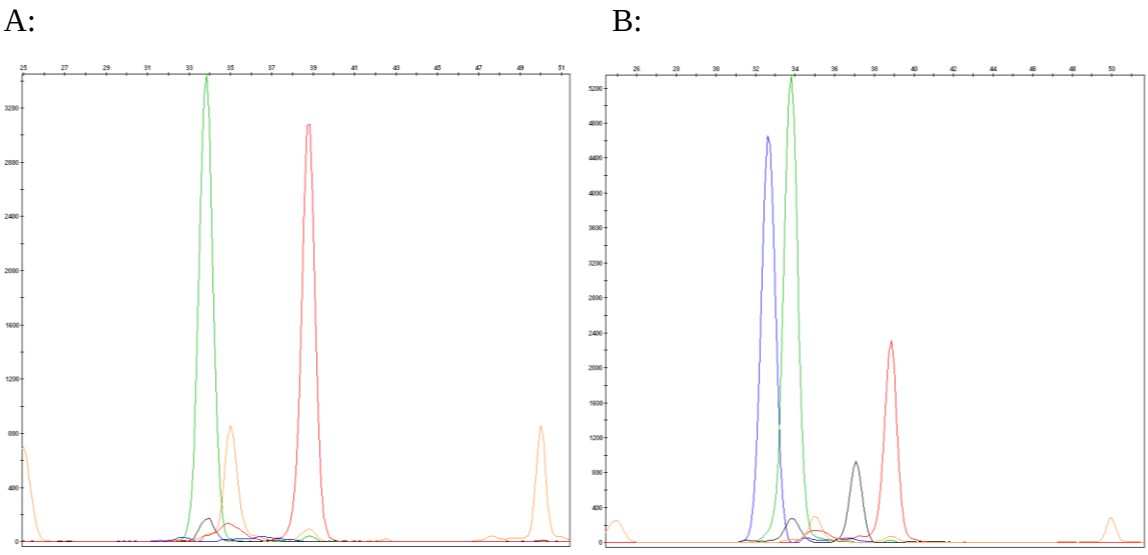


Figure 10. Electropherograms of SNaPshot products illustrating the loci detected in the *MYD88* gene (L265P). The plot shows the relative fluorescence intensity versus the measured size (in nucleotides) of the products relative to the GeneScan-120 LIZ internal size standard (orange peaks). Bases are represented by the following colors: T = red, wildtype; C = black, mutated; A = green, wildtype G = blue, mutated. A: A non-mutated case where both products are wildtype. B: A *MYD88* L265P heterozygote case where both wildtype and mutated products are detected.

Next-generation sequencing and analysis

We performed exome sequencing of 6 CAD cases (CAD-2, 5, 7, 19, 20, 22) using FACS-purified monoclonal B cells as test samples, and paired T cells as normal control. This was followed by targeted sequencing of 10 additional cases (CAD-1, 3, 4, 6, 10, 12, 13, 14, 15, 18) using only monoclonal B cells.

The use of FACS-purified cells for this analysis assured that mutations stem from tumor cells. It facilitated also distinguishing real mutations from sequencing errors. When mixed cell populations are used, such as is the case when analyzing frozen tissues, DNA derives from both normal and tumor cells and thus mutations may be discarded as sequencing errors in samples with a predominance of normal cells.

Whole-exome sequencing was performed at BGI Tech Solutions (Hong Kong) using the Agilent SureSelect Human All Exon V4 Reagent Kit and Illumina HiSeq technology. The method normally requires input of 3 µg DNA but also works well with input amounts down to 200 ng, an advantage for our study using low cell numbers for analysis. Five of the cases had at least 200 ng DNA, and for one case amplified genomic DNA had to be used. The exome sequencing procedure included untranslated regions that gave both coding and regulatory regions with 50x coverage, pair-end reads and about 100 bp long reads. FACS- sorted clonal B cells were analyzed using paired sorted T cells as the normal control. Since pure cell populations (FACS acquired) were analyzed, a 50x coverage was deemed sufficient.

Bioinformatics analysis was carried out by use of software programs that are free for academic use and work in an UNIX environment. Analysis of sequencing data included pre-processing of raw data (removal of adapters and trimming of low quality bases), followed by quality control using the FastQC software (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Then, sequence alignment was performed using BWA 0.7.8 (http://bio-bwa.sourceforge.net/)[114], a well-known short reads alignment tool. Reads were aligned to human reference genome (Genome Reference Consortium GRCh37 released in Feb 2009). The raw BAM alignment files were treated by a standard Genome Analysis Toolkit (GATK) variant calling pipeline. This included two Picard tools 1.113 (https://broadinstitute.github.io/picard/): FixMateInformation and MarkDuplicates, followed by two programs from the GATK 3.1-1 (https://www.broadinstitute.org/gatk/): IndelRealigner and BaseRecalibrator. After marking of duplicates, realignment around indels is necessary since that area is more prone to noise. Since base quality score is a critical factor for variant detection we performed base

recalibration. This is necessary to minimize artifacts that might affect the variant calling procedure. Subsequently, detection of somatic variants was performed by two programs: Strelka 1.0.14 (https://sites.google.com/site/strelkasomaticvariantcaller/)[115] and MuTect v1 (http://gatkforums.broadinstitute.org/gatk/categories/mutect), which simultaneously analyze tumor-normal paired samples. MuTect is able to detect single nucleotide polymorphisms/variations (SNPs/SNVs), while Strelka detects both SNPs/SNVs and small insertions/deletions (indels). In order to analyze the data for large indels and breakpoints, the Pindel program (http://gmt.genome.wustl.edu/packages/pindel/)[116] was used. This program uses a pattern growth approach to identify the breakpoints in paired-end short reads. Annotation was performed by the SnpEff program (http://snpeff.sourceforge.net/).[117] Somatic mutations were considered as relevant and chosen for further analysis if identified concurrently by two programs: SNPs/SNVs by both MuTect and Strelka; indels by both Strelka and Pindel. In addition, somatic mutations identified with a high quality score by only one of the programs were manually verified using the Integrative Genomics Viewer (IGV) 2.3.34 browser (https://www.broadinstitute.org/igv/)[118]. Such manual verification allowed to distinguish obvious technical errors from real mutations. Although laborious, this is necessary when analyzing NGS data from samples with low DNA quantities.

Targeted NGS of the genes with recurrent mutations was performed on 10 additional CAD samples. Patient samples for targeted sequencing, were chosen based on DNA quality and availability. Targeted NGS was performed only for genes that were found mutated by exome sequencing in at least two CAD cases, and were classified as ‘high’ or ‘moderate’ impact by SnpEff. This selection was a challenge because our samples had low DNA content resulting in coverage problems. Some genes of interest may therefore not have been detected. A solution would have been to repeat sequencing multiple times, but DNA was not available for this. Since only six cases were analyzed initially, we may have overlooked less frequent mutations in CAD. Our group is still collecting additional CAD samples with relatively high DNA quantity for further exome sequencing to detect less frequent mutations.

All cases except one (CAD-1) were amplified using Illustra Ready-To-Go GenomiPhi V3 DNA Amplification Kit, prior to targeted sequencing. Sequencing was done at Novogene (Hong Kong) on an Illumina HiSeq instrument with a coverage of about 200x and pair-end reads (PE) of 150 bp long. Coverage for all the samples was satisfactory (Figure 11). Samples were analyzed in the same way as the exome data except for Mutect, and Strelka that could not be used since targeted sequencing for normal controls was not performed. Instead variant calling from GATK was used. Further, somatic or germline status was identified for each

mutation using Sanger sequencing of normal controls. All mutations were verified by Sanger sequencing (Figure 12). Primers were design using the National Center for Biotechnology Information (NCBI) tool (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). For PCR reaction AmpliTaq Gold® 360 Master Mix (Thermo Fisher Scientific) was used, followed by sequencing using BigDye® Terminator v1.1 Cycle Sequencing Kit (Thermo Fisher Scientific) and the ABI Prism 3130 sequencer.

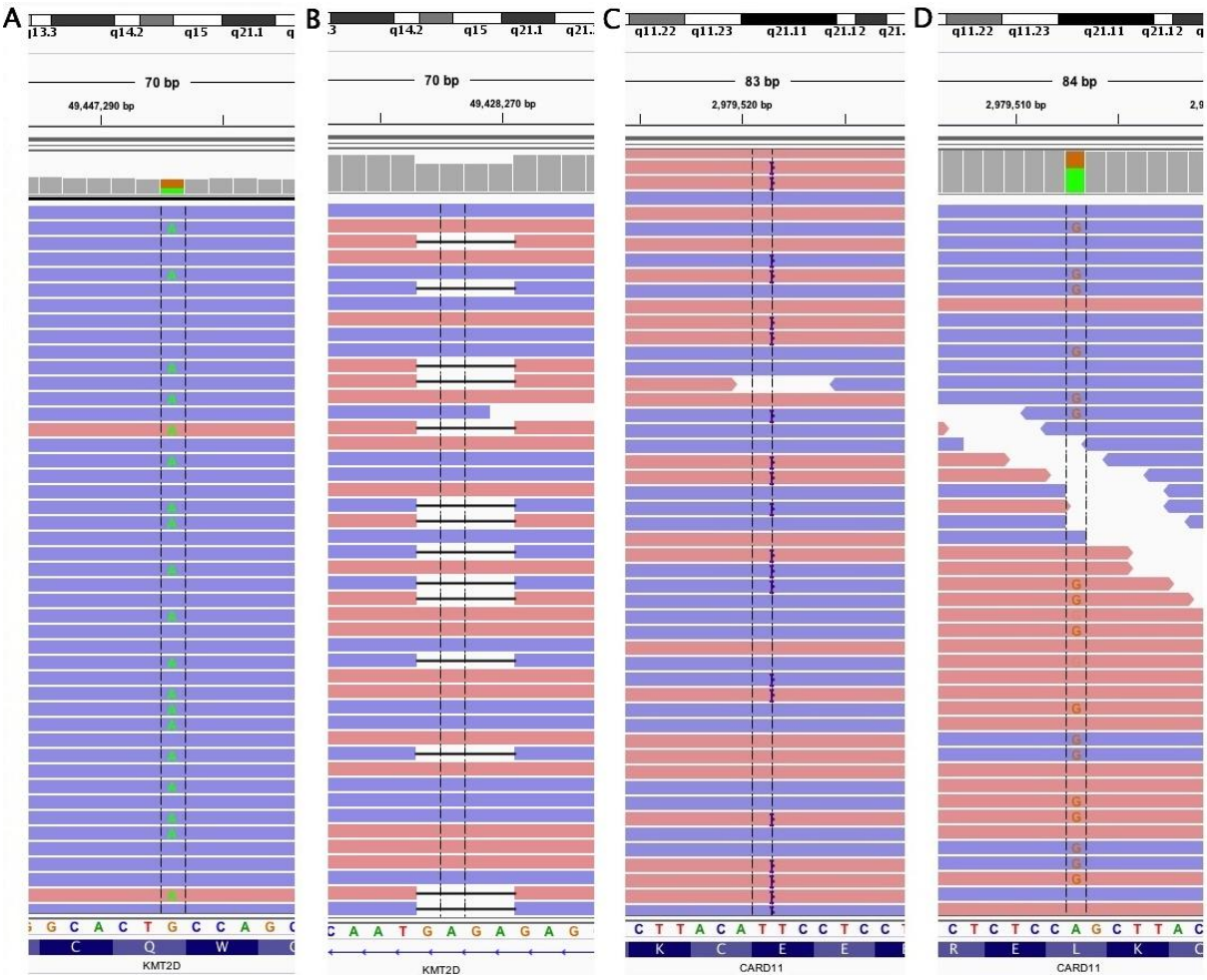


Figure 11. Examples of mutations detected by exome or targeted sequencing of clonal B cells. A) Nonsense mutation in *KMT2D* from CAD-20 detected by exome sequencing. B) 4 bp deletion in *KMT2D* from CAD-13 detected by targeted sequencing. C) 3 bp in-frame insertion in *CARD11* from CAD-6 detected by targeted sequencing. D) Missense mutation in *CARD11* from CAD-20 detected by exome sequencing. Sequences are displayed in IGV browser, with aligned reads colored by strand (red/blue). Point mutations and indels in A-D are detected in approximately 40-60% of reads, and are present on both strands.

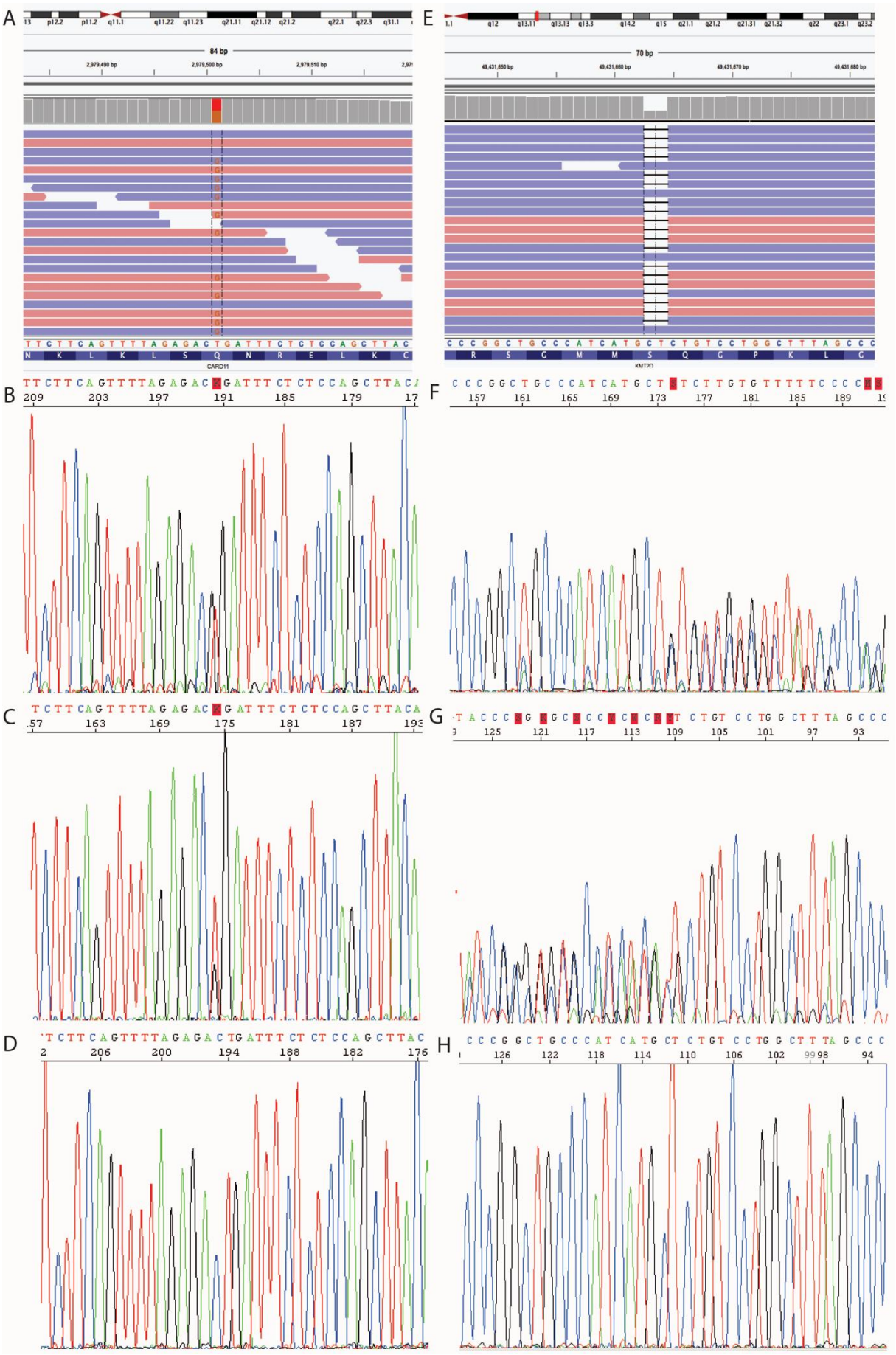

A
84 bp
CARD11
E
70 bp
KMT2D
B
C
D
F
G
H

Figure 12. Examples of NGS analysis and Sanger sequencing verification of mutations in *CARD11* and *KMT2D* genes. A-C) missense mutation of *CARD11* in clonal B cells from CAD-22, as determined by exome sequencing (A) and Sanger sequencing using forward (B) and reverse (C) primers. The visible double peak in Sanger sequencing matches the missense mutation detected by NGS. D) Sanger sequencing of the same region of *CARD11* in T cells (control) from CAD-22 detects presence of only germline sequence, without mutation (no double peak). Sequences A-D are aligned to simplify the comparison. E-G) 2 bp deletion within *KMT2D* in clonal B cells from CAD-1, as determined by targeted NGS (E) and Sanger sequencing using forward (F) and reverse (G) primers. Sanger sequencing results show presence of single peaks up to the site of deletion, with double peaks starting from the deletion site. The site of deletion, detected by Sanger sequencing, matches the position of deletion determined by NGS. H) Sanger sequencing of the same region of *KMT2D* in T cells (control) from CAD-1 detects the presence of only germline sequence, without mutations (no double peaks). Sequences E-H are aligned to simplify the comparison.

Statistical analysis

We used the Pearson correlation two-tailed test to unravel connections between molecular features and clinical data. For $r > |0.4|$, data were examined manually, and degree-of-freedom (df) and p-values were calculated. Only statistically significant results ($p < 0.05$) are reported.

Results

Diffuse large B-cell lymphoma associated with monoclonal B-cell lymphocytosis (paper 1)

Prospectively collected bone marrow and blood samples of patients presenting with primary DLBCL were analyzed and sorted for MBL/MSBC when present. Six of the 19 patients with DLBCL showed MBL/MSBC. Five of six patients had DLBCL-ABC and one had DLBCL-GCB (Table 8). MSBC in the bone marrow was associated with MBL in the six patients (Table 8). The bone marrow trephine biopsies showed one to three foci with small B lymphoid cells with infiltration patterns previously described in patients with MBL.[70] Infiltration with large atypical cells was not seen. In addition, MBL/MSBC identified by flow cytometry showed a forward scatter overlapping with that of small polytypic B-lymphocytes in the same sample. Taken together, these results indicate that the MBL/MSBC were small clonal B-cells and did not correspond to minimal infiltration of the bone marrow by DLBCL.

The specific rearranged *IGHV* sequences detected in MBL/MSBC and DLBCL samples revealed a clonal relationship in three of six paired samples, as indicated by identical *IGH* CDR3 sequences (Table 8). The samples of patients one and two showed 94.27% and 94.59 % homology to germline *IGHV* genes, respectively (Table 8). The somatic mutations were identical or partially identical in paired MBL/MSBC and DLBCL samples that were clonally related. Of interest, patient one showed MBL/MSBC with a CLL immunophenotype. The paired DLBCL showed an ABC immunophenotype without CLL immunophenotype. Patient two showed a non-CLL-type MBL/MSBC with a DLBCL with an ABC immunophenotype. No differences in marker expression between MBL/MSBC and DLBCL was observed in this patient. Patient six showed shared as well as divergent somatic mutations in MBL/MSBC and DLBCL, with the latter showing more mutations. The MBL/MSBC in patient six showed a GCB immunophenotype as did the paired DLBCL. Histologic review of the DLBCL sample in this patient did not reveal concurrent FL. Of interest, only one focal paratrabecular small B-cell aggregate was seen in the bone marrow. While these findings were not diagnostic of concurrent FL because of the very limited infiltrate, it is of note that the infiltration pattern of the MBL/MSBC in the bone marrow is as observed for that of FL. The somatic mutation rate of the paired samples was also considerably higher in this patient than observed for the other patients. This is consistent with a GCB origin.

A common clonal origin of MBL/MSBC and DLBCL could not be demonstrated for

the other three cases. However, *IGHV* gene usage was either the same or belonged to the same *IGHV* gene family for paired MBL/MSBC and DLBCL of the three cases without clonal relationship.

Table 8. Immunoglobulin sequence characteristics of MSBCL/MBL and DLBCL

Patient	Sample	Immunophenotype	COO	*IGHV*	V-region homology	CDR3 amino acid sequence
Patient 1	MSBC/MBL	CD19+, CD20dim, CD79b-, CD5+, CD23dim, IgL+	CLL-like	3-53	94.27%	CARGDCSSTTCNILAVW
	DLBCL	CD20+, CD10-, BCL6+, MUM1+, CD5-, CD23-, EBER-	ABC	3-53	94.27%	CARGDCSSTTCNILAVW
Patient 2	MSBC/MBL	CD19+, CD20+, CD79b+, CD5-, CD10-, CD23-, IgK+	Non-CLL	4-34	94.59%	CARGPRDDMAVALDNW
	DLBCL	CD20+, CD10-, BCL6+, MUM1+, CD5-, EBER-	ABC	4-34	94.59%	CARGPRDDMAVALDNW
Patient 3	MSBC/MBL	CD19+, CD20+, CD79b+, CD5-, CD23-, IgL+	Non-CLL	4-59	91.61%	CARAGQYYYDSSGYYAYAFDIW
	DLBCL	CD20+, CD10-, BCL6-, MUM1-, CD5-, EBER-	ABC	4-34	96.13%	CARGCSVYGLVYW
Patient 4	MSBC/MBL	CD19+, CD20dim, CD79b-, CD5+, CD23dim, IgK+	CLL-like	3-7	90.00%	CARYGRAVSIDYW
	DLBCL	CD20+, CD10-, BCL6-, MUM1-, CD5-, EBER-	ABC	3-23	94.38%	CAKECSNSWDTW
Patient 5	MSBC/MBL	CD19+, CD20+, CD79b+, CD5dim, CD23-, IgK+	Non-CLL	2-5	95.48%	CAHIPYFHTSGRIFDYW
	DLBCL	CD20+, CD10-, MUM1+, CD5-, CD23dim, EBER-	ABC	2-5	100.00%	CARPKKSLL*WW*LLFNV#FDYW
Patient 6	MSBC/MBL	CD19+, CD20+, CD79b+, CD5-, CD23-, CD10+, IgK+	GCB-like	3-11	83.12%	CARIYRHSLDIW
	DLBCL	CD20+, CD10+, MUM1-, CD5-, EBER-	GCB	3-11	77.50%	CARFYRHAFDIW

IGHV: immunoglobulin heavy chain variable gene; COO: cell of origin; CDR3: complementarity determining region 3; MSBC/MBL: monoclonal small B-cell infiltrate/monoclonal B-cell lymphocytosis; DLBCL: diffuse large B-cell lymphoma; CLL: chronic lymphocytic leukemia; ABC: activated B-cell type; GCB: germinal center cell type

Primary cold agglutinin disease (CAD) (paper 2 and 3)

Immunoglobulin gene analysis and correlation with clinical data

Immunoglobulin heavy and light chain gene sequences of 27 patients with well-characterized CAD were studied to find features that may affect I antigen binding, and therefore might explain heterogeneity in clinical presentation and disease activity. These molecular features were subsequently correlated with the clinical features. Productive *IGHV4-34* gene rearrangements were identified in 81% (22/27) of patients (Table 9). One patient showed *IGHV3-23* gene rearrangement while clonal rearrangements were not detected or unproductive in the other patients. The N-glycosylation sequon, located within CDR2 of the *IGHV4-34* gene, was variably mutated (Table 9; Figure 13A). Since this N-glycosylation site is located within the antigen-binding pocket (Figure 13B), inactivating mutations preventing attachment of glycans likely modulate binding. Patients with inactivating mutations of the sequon had a significantly lower hemoglobin level ($p = 0.036$). In addition, a mutation hotspot within FR3 of IGH was identified at the germline amino acid sequence Lys^{90}-Leu^{91}-Ser^{92} (KLS). Only 13,5% of patients did not have any mutation in this hotspot (Table 9; Figure 13A). Patients with more mutations had lower hemoglobin levels ($p = 0.030$) (Table 9). The presence of mutations in the N-glycosylation sequon and the number of mutations in KLS mutation hotspot were also analyzed in combination. Mutations at both sites were even more strongly correlated with anemia ($p = 0.0055$; Table 9).

Clonal rearrangement of the *IGKV3-20* gene was detected in 16/27 (59%) of patients, and clonal rearrangement of *IGKV3-15* gene was identified in 4/27 (15%) of patients (Table 9). Other unique clonal *IGKV* gene rearrangements were found in 4/27 patients (15%), whereas no *IGKV* gene rearrangements were found in three patients (11%). Reanalysis of previously published data showed that the *IGKV3-20* gene was used in most CAD cases (12/16). In one case (1/16) the *IGK3-15* gene was used, and in three cases (3/16) other *IGKV* genes were used.[109-113] Since *IGKV3-20* was used in most of our cases as well as published cases, it is likely that this gene has been selected for. Of interest, the IG light chain CDR3 region was highly homologous in a subgroup of patients (Figure 13C), another strong indication of antigen selection. The group of patients with homologous IG light chain CDR3 regions were younger at diagnosis ($p = 0.011$). As expected in view of the *IGHV* sequence data, somatic hypermutations were also demonstrated in the rearranged *IG* light chain genes

of the patients. A low level of mutations in *IGKV3-20* was correlated with younger age at diagnosis (p=0.012).

Table 9. Analysis of *IGH* and *IGL* sequences and correlation with clinical features.

Analysis of *IGH* and *IG* light chain gene sequences	**% of patients**
Rearranged *IGH* genes	
Productive *IGHV4-34*	81%
Unproductive/not detected	15%
Productive *IGHV3-23*	4%
Mutations in N-glycosylation sequence in *IGHV4-34* CDR2	
No mutations	36,5%
Mutations in flanking residues only	27%
Mutations in the core region (inactivating mutations)	36,5%
Mutations in Lys-Leu-Ser (KLS) sequence in *IGHV4-34* FR3	
No mutations	13.5%
1 mutation	32%
2 mutations	50%
3 mutations	4.5%
Rearranged *IGL* genes	
Clonal *IGKV3-20*	59%
Clonal *IGKV3-15*	15%
Other *IG* light chain genes	15%
Not detected	11%
Amino acid sequence of *IGKV3-20* CDR3 region	
Highly homologous	56%
Mutated	38%
Correlation between molecular findings and clinical features	**P-value**
Mutations in IGHV4-34 CDR2 N-glycosylation site with hemoglobin levels	p = 0.036
Mutations in IGHV4-34 FR3 KLS sequence with hemoglobin levels	p = 0.030
Mutations in IGHV4-34 CDR2 N-glycosylation site + FR3 KLS sequence with hemoglobin levels	p = 0.005
Stereotyped IGK CDR3 region with lower age	p = 0.011
% of mutations in IGKV3-20 gene with age	p = 0.012

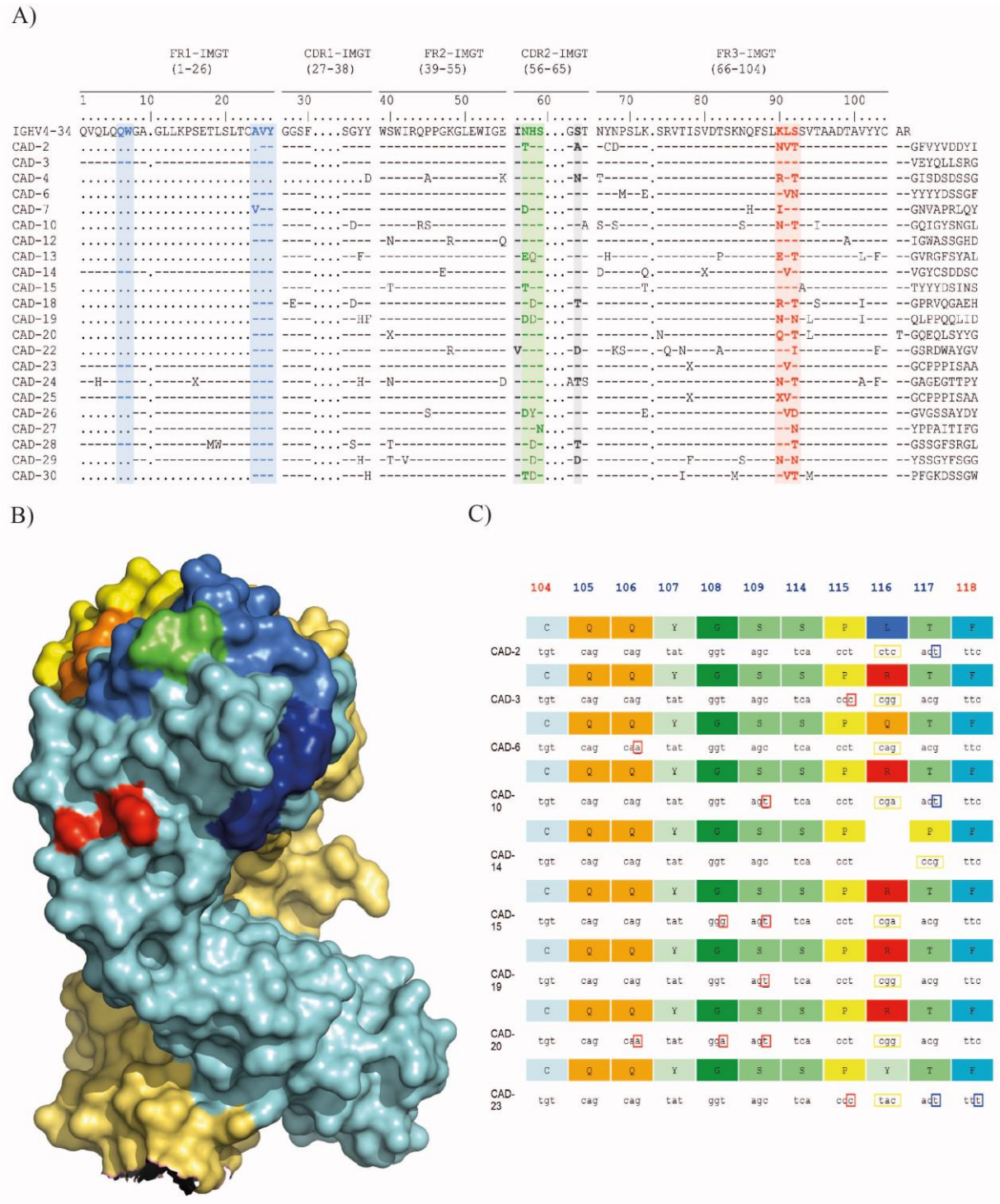


Figure 13. Molecular features of the *IG* in CAD patients. (A) Alignment of amino acid sequences of IGH chain from CAD patients expressing the *IGHV4-34* gene**.** The core region of the N-glycosylation site (Asn–X–Ser/Thr) is marked in green whereas the flanking residues are in bold.[119] A mutation hotspot within FR3 is marked in red. (B) Schematic representation of the crystal structure of the FAB fragment from a human IG cold agglutinin. The three-dimensional immunoglobulin structure was generated by the PyMOL program (PDB 1DN0).[110] *IGHV4-34* encoded heavy chains are colored light blue whereas *IGKV3-20*

encoded light chains are colored yellow. Regions of interest are marked with different colors (coloring corresponds to A): hydrophobic patch QW + AVY (dark blue), N-glycosylation site NHS (green), KLS sequence (red), IGK CDR3 (orange), IGK CDR1 and CDR2 (bright yellow), IGH CDR1, CDR2 and CDR3 (blue). (C) Alignment of highly homologous *IGKV3-20* CDR3 sequences of nine CAD patients. Somatic mutations are marked by rectangles colored red for the variable gene, blue for the joining gene and yellow for the intervening non-templated sequence. Numbering and coloring of amino acids are according to IMGT.

MYD88 mutation analysis

Bone marrow samples of 27 patients with well-characterized CAD were studied for *MYD88* L265P mutation. A single nucleotide polymorphism, at amino acid position 265 of the *MYD88* gene was analyzed using PCR and a SNaPshot mini-sequencing assay. *MYD88* L265P mutation, which is typical for LPL, was not detected in any of the analyzed CAD samples.

Next generation sequencing and targeted sequencing

Prospectively collected bone marrow samples of 16 CAD patients, enrolled in a clinical trial (CAD5), were studied for unknown gene variants. Six of CAD cases were exome sequenced and findings were confirmed in ten additional cases using targeted sequencing. For this analysis, bone marrow fluorescent activated cell sorted clonal B cells and normal T cells were used.

Recurrently mutated genes detected in CAD patients by exome sequencing are summarized in Table 10. The only significant recurrent mutations were found in *KMT2D* and *CARD11* genes using criteria as described in Methodological considerations (Table 11). For that reason, 10 additional cases were screened for mutations in these two genes with targeted deep sequencing.

In total, somatic mutations of the *KMT2D* gene were detected in 11 out of 16 cases (69%). Seven of these mutations were classified as high impact by SnpEff, and consisted of: 2 out-of-frame deletions, 1 out-of-frame insertion, 3 stop codon gained and 1 splice donor variant. These mutations result in C-terminally truncated proteins that lack the SET-domain and, therefore the proteins are enzymatically inactive. In two other cases, we identified missense mutations located in the C-terminal domain, which are expected to impair activity of

the SET domain or result in loss-of-function (Figure 14). Interestingly, similar mutations have previously been identified in patients with Kabuki syndrome. Finally, two patients showed *KMT2D* mutations causing splice region variants classified by SnpEff as of low impact. The mutated fraction of *KMT2D* was between 40-60% for all patients, as detected by NGS, suggesting that only one allele was mutated.

CARD11 was somatically mutated in 5 out of 16 cases (31%). All mutations were classified as of moderate impact by the SnpEff program. Three samples had a missense mutation, of which two had exactly the same mutation, one sample had a disruptive in-frame deletion and one sample had a disruptive in-frame insertion (both mutations were in the same position). All five mutations were located within a 20bp stretch of exon 6, coding for the BAR domain of the coiled-coil region of CARD11 (Figure 15). Mutations localized in BAR domain and coiled-coil region of CARD11 have previously been demonstrated in DLBCL and shown to induce constitutive activation of NF-kB pathway. It is therefore likely that *CARD11* mutations identified in CAD have a similar impact on NF-kB pathway activation. The mutated fraction of *CARD11* was between 40-60% for the five patients, suggesting that only one allele was mutated.

Other recurrent mutations were not detected using criteria as described in Methodological considerations.

Table 10. Recurrently mutated genes detected in CAD patients by exome sequencing.

Sample	Chr	Chr position (GRCh37)	Ref allele	Alt allele	Mutect quality score	Strelka quality score	Indels detected by Pindel	Annotation[a]	Impact[a]	Gene name[a]
CAD-2	11	134133723	A	T	32	QSS_NT=58	-	sequence feature	LOW	*ACAD8*
CAD-7	11	134131103	T	G	13	QSS_NT=15	-	sequence feature	LOW	*ACAD8*
CAD-22	*7*	*2979501*	*T*	*G*	*122*	*QSS_NT=98*	-	*missense variant*	*MODERATE/ MISSENSE*	*CARD11*
CAD-20	*7*	*2979513*	*A*	*G*	*122*	*QSS_NT=130*	-	*missense variant*	*MODERATE/ MISSENSE*	*CARD11*
CAD-2	X	21609164	G	T	212	QSS_NT=232	-	missense variant	MODERATE/ MISSENSE	*CNKSR2*
CAD-5	X	21458930	T	A	173	QSS_NT=67	-	sequence feature	LOW	*CNKSR2*
CAD-22	8	113236173	G	T	62	QSS_NT=93	-	3 prime UTR variant	MODIFIER	*CSMD3*
CAD-2	8	114110894	C	T	14	not detected	-	sequence feature	LOW	*CSMD3*
CAD-7	8	113246694	C	A	70	QSS_NT=124	-	missense variant	MODERATE/ MISSENSE	*CSMD3*
CAD-22	2	136875581	A	T	82	QSS_NT=96	-	sequence feature	LOW	*CXCR4*
CAD-19	2	136875517	C	T	49	QSS_NT=17	-	sequence feature	LOW	*CXCR4*
CAD-19	2	136872550	T	TG	not analyzed	QSI_NT=64	Pindel	frame shift variant	HIGH	*CXCR4*
CAD-20	4	126413695	T	A	82	QSS_NT=109	-	3 prime UTR variant	MODIFIER	*FAT4*
CAD-19	4	126412847	C	G	38	QSS_NT=56	-	missense variant	MODERATE/ MISSENSE	*FAT4*
CAD-19	4	126412848	A	C	33	QSS_NT=53	-	synonymous variant	LOW/ SILENT	*FAT4*

CAD-2	4	126314969	A	T	18	not detected	-	sequence feature	LOW	*FAT4*
CAD-2	4	162402468	A	T	23	not detected	-	intron variant	MODIFIER	*FSTL5*
CAD-2	4	162402469	T	G	23	not detected	-	intron variant	MODIFIER	*FSTL5*
CAD-7	4	162305216	G	T	42	QSS_NT=71	-	3 prime UTR variant	MODIFIER	*FSTL5*
CAD-22	5	161530818	T	G	23	not detected	-	sequence feature	LOW	*GABRG2*
CAD-19	5	161578674	A	AT	not analyzed	not detected	Pindel	intron variant	MODIFIER	*GABRG2*
CAD-19	14	106329406	A	T	16	QSS_NT=67	-	upstream gene variant	MODIFIER	*hsa-mir-4538*
CAD-7	14	106329371	T	C	21	QSS_NT=21	-	upstream gene variant	MODIFIER	*hsa-mir-4538*
CAD-22	15	81591803	C	A	27	QSS_NT=56	-	missense variant	MODERATE/ MISSENSE	*IL16*
CAD-2	15	81603315	C	G	63	QSS_NT=67	-	3 prime UTR variant	MODIFIER	*IL16*
CAD-20	2	121108403	G	A	60	QSS_NT=78	-	3 prime UTR variant	MODIFIER	*INHBB*
CAD-5	2	121108198	A	G	8	QSS_NT=22	-	3 prime UTR variant	MODIFIER	*INHBB*
CAD-5	4	88083130	T	G	213	QSS_NT=164	-	3 prime UTR variant	MODIFIER	*KLHL8*
CAD-7	4	88084185	T	A	62	QSS_NT=114	-	3 prime UTR variant	MODIFIER	*KLHL8*
CAD-22	*12*	*49431178*	*G*	*A*	*32*	*QSS_NT=21*	-	*stop gained*	*HIGH/ NONSENSE*	*KMT2D*
CAD-20	*12*	*49447293*	*G*	*A*	*127*	*QSS_NT=92*	-	*stop gained*	*HIGH/ NONSENSE*	*KMT2D*
CAD-19	*12*	*49415905*	*C*	*A*	*110*	*QSS_NT=129*	-	*missense variant*	*MODERATE/ MISSENSE*	*KMT2D*
CAD-2	*12*	*49427849*	*C*	*T*	*25*	*not detected*	-	*splice donor variant*	*HIGH*	*KMT2D*
CAD-22	4	62814038	T	C	23	QSS_NT=15	-	downstream gene variant	MODIFIER	*LPHN3*
CAD-5	4	62758504	C	A	115	QSS_NT=88	-	synonymous variant	LOW/ SILENT	*LPHN3*
CAD-7	4	62758392	C	A	31	QSS_NT=47	-	missense variant	MODERATE/ MISSENSE	*LPHN3*
CAD-19	12	85492623	A	T	14	QSS_NT=22	-	sequence feature	LOW	*LRRIQ1*
CAD-7	12	85517943	G	A	29	QSS_NT=55	-	missense variant	MODERATE/ MISSENSE	*LRRIQ1*
CAD-20	22	29886274	T	A	19	not detected	-	missense variant	MODERATE/ MISSENSE	*NEFH*
CAD-5	22	29877002	G	A	20	not detected	-	missense variant	MODERATE/ MISSENSE	*NEFH*
CAD-20	9	33797783	A	T	26	not detected	-	intron variant	LOW	*PRSS3*
CAD-20	9	33799026	C	G	not detected	QSS_NT=25	-	missense variant	MODERATE/ MISSENSE	*PRSS3*
CAD-7	9	33798574	G	A	20	not detected	-	missense variant	MODERATE/ MISSENSE	*PRSS3*
CAD-20	3	196198131	C	A	15	QSS_NT=20	-	3 prime UTR variant	MODIFIER	*RNF168*
CAD-7	3	196197249	A	C	15	QSS_NT=16	-	3 prime UTR variant	MODIFIER	*RNF168*
CAD-19	3	78646365	A	G	94	QSS_NT=41	-	downstream gene variant	MODIFIER	*ROBO1*
CAD-7	3	78646415	A	C	38	QSS_NT=50	-	3 prime UTR variant	MODIFIER	*ROBO1*
CAD-5	8	119203304	C	T	8	QSS_NT=19	-	3 prime UTR variant	MODIFIER	*SAMD12*
CAD-19	8	119634213	CGGC TCGG	C	not analyzed	QSI_NT=109	Pindel	frame shift variant	HIGH	*SAMD12*
CAD-2	7	97369304	C	T	79	QSS_NT=94	-	3 prime UTR variant	MODIFIER	*TAC1*
CAD-7	7	97361850	G	C	109	QSS_NT=167	-	upstream gene variant	MODIFIER	*TAC1*
CAD-20	8	35648680	A	G	37	QSS_NT=71	-	3 prime UTR variant	MODIFIER	*UNC5D*
CAD-19	8	35652157	A	T	152	QSS_NT=124	-	3 prime UTR variant	MODIFIER	*UNC5D*
CAD-22	16	27738431	T	G	not detected	QSS_NT=48	-	upstream gene variant	MODIFIER	*Y_RNA*
CAD-5	12	123830128	A	G	48	QSS_NT=57	-	downstream gene variant	MODIFIER	*Y_RNA*
CAD-5	12	123830136	A	C	45	QSS_NT=57	-	downstream gene variant	MODIFIER	*Y_RNA*
CAD-5	8	77616593	C	A	281	QSS_NT=203	-	synonymous variant	LOW/ SILENT	*ZFHX4*
CAD-7	8	77765298	A	C	10	QSS_NT=18	-	synonymous variant	LOW/ SILENT	*ZFHX4*

Analyzed with the SnpEff program; Chr: chromosome; Chr position (GRCh37): chromosome positions according to human reference genome Genome Reference Consortium GRCh37 released in Feb 2009; Ref: reference; Alt: alternative.

In italic: genes chosen for NGS panel sequencing, i.e. genes with mutations present in at least two cases and, each mutation was identified by two programs or have been given a high quality score by either of these programs and, the impact as classified by SnpEff was 'high' or 'moderate'.

Table 11. Mutations in *KMT2D* and *CARD11* in CAD patients.

Sample name	Gene	Effect	Mutation[a]
CAD-1	*KMT2D*	Frameshift variant	c.9475_9476delAG
CAD-2	*KMT2D*	Splice donor variant	c.10741G>A
CAD-4	*KMT2D*	Frameshift variant	c.1035_1036delCT
CAD-10	*KMT2D*	Splice region variant	c.16413C>G
CAD-12	*KMT2D*	Frameshift variant	c.661_662insG
CAD-13	*KMT2D*	Splice region variant	c.10441_10444delTCTC
CAD-14	*KMT2D*	Missense variant	c.15143G>A
CAD-15	*KMT2D*	Stop gained	c.4265G>A
CAD-19	*KMT2D*	Missense variant	c.16442G>T
CAD-20	*KMT2D*	Stop gained	c.805C>T
CAD-22	*KMT2D*	Stop gained	c.9961C>T
CAD-1	*CARD11*	Disruptive inframe deletion	c.723_725delGGA
CAD-6	*CARD11*	Disruptive inframe insertion	c.725_726insGGA
CAD-13	*CARD11*	Missense variant	c.734T>C
CAD-20	*CARD11*	Missense variant	c.734T>C
CAD-22	*CARD11*	Missense variant	c.746A>C

Numbering of nucleotides is according to reference sequences: ENST00000301067.11 (*KMT2D*) and ENST00000396946.8 (*CARD11*).

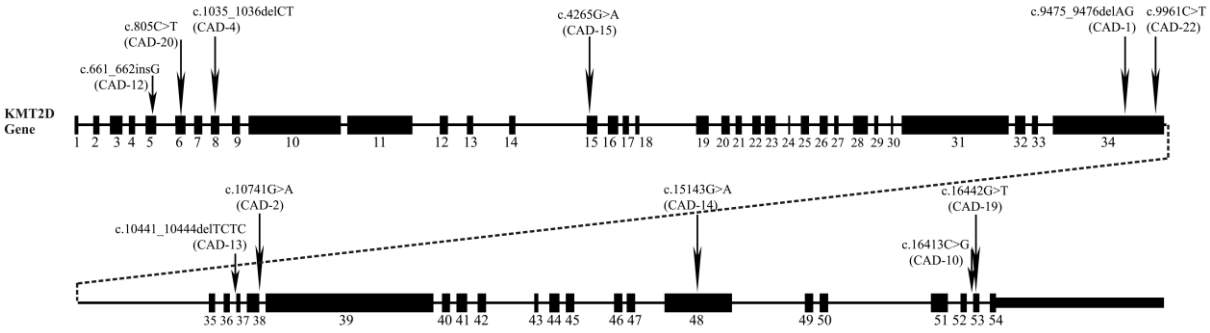


Figure 14. Schematic representation of the human *KMT2D* gene (ENST00000301067.11). The gene organization is adapted from the UCSC browser (hg19). Vertical arrows show the approximate position of mutations. CAD patient numbers are indicated in parenthesis.
ins: insertion; del: deletion; fs: frameshift; STOP: nonsense mutation.

```
WT       231  QLKHRLNKME EECKLERNQS LKLKNDIENR  260
CAD-1         QLKHRLNKME ECKLERNQS LKLKNDIENR
CAD-6         QLKHRLNKME EEECKLERNQS LKLKNDIENR
CAD-13        QLKHRLNKME EECKPERNQS LKLKNDIENR
CAD-20        QLKHRLNKME EECKPERNQS LKLKNDIENR
CAD-22        QLKHRLNKME EECKLERNPS LKLKNDIENR
```

Figure 15. Human CARD11 amino acid 231-260 (NP_115791.3) sequence. Mutations are underlined. WT: wild-type.

Discussion

Diffuse large B-cell lymphoma associated with monoclonal B-cell lymphocytosis

The close association between MSBC and MBL has previously been demonstrated.[70,120] MBL/MSBC is present at a high frequency in patients with DLBCL, especially DLBCL ABC.[71] We have prospectively collected blood and bone marrow samples in patients with primary DLBCL at diagnosis to study the clonal relationship of MBL/MSBC with the paired DLBCL. MBL/MSBC were detected in 6/19 patients, of whom 5 with DLBCL with ABC origin and one with GCB origin, in accordance with a previous study.[71] MSBC in the bone marrow was associated with MBL in the six patients. The bone marrow trephine biopsies of the six patients showed one to three foci with small B lymphoid cells with infiltration patterns previously described in patients with MBL.[70,120]

MBL/MSBC were clonally related to the paired DLBCL in 3/6 patients as demonstrated by identically rearranged immunoglobulin heavy chain gene sequences in the paired samples. Patient one showed only shared somatic mutations in MBL/MSBC and DLBCL, while patient two showed shared somatic mutations and one divergent but silent somatic mutation in the CDR3 region. Patient six showed shared as well as divergent somatic mutations in MBL/MSBC and DLBCL with the latter showing more mutations. We had previously also shown, in one patient, that the MBL/MSBC clone was clonally related to the paired DLBCL in one patient.[71]

Patient one showed MBL/MSBC with a CLL immunophenotype. The paired DLBCL showed an ABC immunophenotype without CLL immunophenotype. One case of CLL-like MBL/MSBC that was clonally related to a paired DLBCL-ABC was also demonstrated in the earlier publication of our group.[71] Of interest, the MBL/MSBC in patient six showed a GCB immunophenotype as did the paired DLBCL. The GCB immunophenotype is consistent with the high mutation rate and the presence of divergent mutations, suggesting on-going mutation. The presence of common as well as divergent mutations in MBL/MSBC and DLBCL in this patient, suggest parallel evolution of both clones. Parallel development of DLBCL from a common progenitor cell has previously been demonstrated for FL.[77] Case 6 in our series may represent a variant of the same process.

IGHV gene usage was the same or belonged to the same *IGHV* gene family for paired MBL/MSBC and DLBCL of the three cases without clonal relationship. Whether this

indicates that MBL/MSBC and DLBCL arose through common antigen-stimulation with subsequent selection of *IGHV* genes in these cases is an interesting hypothesis but is difficult to ascertain. Preferential *IGHV* gene usage has previously been demonstrated in DLBCL and may support this hypothesis.[121,122]

Taken together, a common clonal origin of MBL/MSBC and DLBCL was demonstrated in 3/6 cases and suggest that DLBCL may arise from small precursor cells, with either a CLL, non-CLL or GC immunophenotype. More cases will have to be prospectively collected to confirm the findings and to perform molecular genetic studies. The latter may further elucidate the time sequence of genetic changes in DLBCL pathogenesis.

Primary cold agglutinin disease

Immunoglobulin heavy and light chain gene features correlate with primary cold agglutinin disease onset and activity

We demonstrated that mutations at specific sites in the immunoglobulin heavy chain gene *IGHV4-34* correlate with severity of anemia in CAD. The N-glycosylation site, located within the CDR2 of *IGHV4-34* gene, is variably mutated in CAD, and patients with inactivating mutations have significantly lower hemoglobin levels ($p = 0.036$). It has previously been suggested that an intact N-glycosylation site, allowing the addition of bulky glycans, diminishes specific binding of the antigen-binding site of the antibody.[119] It has previously been demonstrated that I antigen binding, the direct cause of CAD, is also mediated by the antigen-binding site of the antibody although it is not an absolute prerequisite.[97,108,123] The observation that patients with mutations at the N-glycosylation site have more severe anemia is consistent with these data. A recent study also showed that reintroduction of the N-glycosylation site in mutated *IGHV4-34* caused a moderate (20%), but reproducible, impairment of B cell survival in cells that are dependent on the binding of the B-cell receptor to auto-antigens.[124] Somatic hypermutation of the site, rendering it non-functional, might thus be advantageous for cell survival in neoplastic B cells that have escaped the normal immune control of auto-reactivity. Somatic hypermutation of the N-glycosylation site in *IGHV4-34* in normal memory B cells may play a similar role although in those cells, increased accessibility of the antigen-binding site to foreign antigens instead of I antigen is likely important for cell selection and survival.[119]

In addition to mutations in the N-glycosylation site in CDR2 of *IGHV4-34*, we found that the KLS amino acid sequence in FR3 of *IGHV4-34* was mutated in almost all CAD

patients. Increased mutation level of this sequence correlated with lower hemoglobin levels ($p = 0.030$). We investigated mutations in this sequence in previously published sequences as well and found it broadly mutated in normal B cells as well as B-cell lymphoma, most frequently at the Ser (S).[119,125] This region seems to be highly conserved in *IGHV4* family germline genes. Analysis of the protein structure[110] shows that KLS amino acids are located on the protein surface, relatively close to the antigen-binding region at a distance similar to the AVY hydrophobic patch that is necessary for I antigen binding. The KLS amino acid sequence contains five known hotspot motifs characteristic for somatic hypermutation (SHM).[8-11] Clinical data suggest that a mutated KLS sequence might enhance I antigen binding since the number of mutations in this sequence correlated with lower hemoglobin levels. Of interest, combining the mutation levels at the N-glycosylation and KLS sites gave a very strong correlation ($p = 0.0055$) with decreased hemoglobin levels, stressing the likely importance of these sites for I antigen binding. The overall mutation rate of the rearranged *IGHV* gene was not correlated with clinical parameters.

Although CAD antibodies are almost exclusively encoded by *IGHV4-34*, one patient in our study (CAD-5) shows cold agglutinins encoded by *IGHV3-23* and *IGKV3-20*. This patient displayed typical clinical characteristics of CAD. Previously, it was shown that naturally occurring anti-I cold agglutinins may also be encoded by *IGHV3* family genes.[126] This has also been demonstrated by Marks et al.[127] who isolated a human *IGHV3-23* antibody with specificity against I antigen. From our series, use of *IGHV3* family genes in CAD seems to be an exception.

Our study, including the reanalysis of older published data of smaller series, shows that the immunoglobulin light chain is encoded by the *IGKV3-20* or *IGKV3-15* gene in more than 80% of CAD patients, indicating that the light chain equally contributes to I antigen binding. The highly restricted usage of these light chain genes provides a rationale for using 'off-the-shelf' anti-immunoglobulin light chain vaccination as part of the treatment for CAD. Anti-immunoglobulin light chain vaccination has been proposed for treating other B-cell lymphomas.[128]

There was a strong correlation between younger age at diagnosis and the presence of a highly homologous or stereotyped *IGKV3-20* CDR3 region ($p = 0.011$). Nine of the patients showed almost identical *IGKV3-20* CDR3 amino acid sequences. This homology is a strong argument for antigen selection during the pathogenesis of CAD. Whether this antigen is I antigen or a related exogenous, perhaps bacterial antigen is unknown. Since both *IGHV4-34* and *IGKV3-20* are known to encode for antibodies that bind infectious antigens, it is possible

that infection is the source of antigen selection that triggers CAD.[129] In that respect it is of interest that transient cold agglutinins may arise after bacterial infection, most often mycoplasma infection.[89,130]

We had previously demonstrated that *MYD88* mutation is not a characteristic of CAD-associated lymphoproliferative disease.[94] The absence of the *MYD88* L265P mutation in CAD-associated lymphoproliferative bone marrow disease has recently been questioned in a review article[131], based on the studies of others[40] (27% (14/53) of *MYD88*-mutated Waldenstrom macroglobulinemia patients had associated cold agglutinin disease). However, our present findings in a large number of CAD patients confirm the absence of *MYD88* L265P mutation in bone marrow samples using a sensitive technique. Discordances of our findings, based on the study of patients with CAD diagnosed according to strict criteria (study criteria available at www.clinicaltrials.gov, NCT02689986) with the limited findings in the literature may be due to the use of less strict criteria for diagnosis in previous studies.

Frequent somatic mutations of *KMT2D* (*MLL2*) and *CARD11* genes in primary cold agglutinin disease

We have demonstrated that *KMT2D* and *CARD11* are frequently mutated in CAD. Other recurrent gene mutations have not been identified. *KMT2D* and *CARD11* mutations, combined with previously described distinctive histology and immune phenotypic findings[94,95], establishes CAD-associated lymphoproliferative disease as a unique disease, distinct from other B-cell lymphoproliferative diseases. Of note, our findings conclusively establish that CAD-associated B-cell lymphoproliferative disease is distinct from LPL with which it has been confounded previously. The latter lymphoma shows *MYD88* mutation in more than 90% of cases, and does not show *KMT2D* or *CARD11* mutation.[42]

Most of the *KMT2D* mutations in CAD inactivate the SET enzymatic domain of the protein. *KMT2D* inactivating mutations are not unique to CAD-associated lymphoproliferative disease, but are frequent in FL, NMZL and DLBCL.[29,49,77,105] It has been hypothesized that *KMT2D* mutation along with other chromatin modulating genes may be an early event, with other genetic changes occurring later.[30,132] We have not identified recurrent mutations in CAD other than *KMT2D* and *CARD11*. Studies of larger case series of this rare disease will be necessary to discover less frequent recurrent mutations, if any.

The mutated allelic fraction was between 40-60% and only one *KMT2D* mutation was found per patient, suggesting that only one allele is mutated. This is in keeping with mutations

found in other lymphoma types as well as in Kabuki syndrome, a congenital autosomal dominant disease caused by monoallelic *KMT2D* mutation[133,134] Monoallelic *KMT2D* mutations inactivating the SET domain cause partial loss of protein expression and increased B cell proliferation.[132] *KMT2D* mutation also impedes class switch recombination.[135] This is consistent with the observation that patients with Kabuki syndrome with *KMT2D* constitutional mutation show reduced levels of class-switched B cells.[136] *KMT2D* has therefore been suggested to be a tumor suppressor gene, the function of which is compromised when haplo-insufficient.[135] These data point to the importance of *KMT2D* mutation for B-cell lymphoproliferative disease. *KMT2D* mutation might act in concert with B-cell survival signals induced by stimulation of the auto-reactive *IGHV4-34*-encoded immunoglobulin receptor expressed on the surface of CAD B cells. A growth advantage through auto-antigen stimulation of the *IGHV4-34*-encoded B-cell receptor was recently demonstrated for a subset of DLBCL.[124]

Five out of the 16 CAD patients did not show somatic *KMT2D* mutation. However, KMT2D protein expression may also be diminished or lost through epigenetic silencing.[132] Whether this is the case for the few patients in our series without *KMT2D* mutation needs as yet to be investigated.

CARD11 was somatically mutated in 5 of 16 CAD cases. Interestingly, all 5 mutations were located in a short region, within 20 bp, in exon 6 of the coiled-coil domain. *CARD11* coiled-coil domain mutations were previously detected in DLBCL of ABC origin.[54] Mutations result in constitutive NF-kB activation and enhanced NF-kB activity upon antigen receptor stimulation.[54] Mono-allelic *CARD11* coiled-coil domain mutations are not oncogenic *per se* in mice, but result in B-cell proliferation and auto-antibody production.[137] These findings are consistent with the recent discovery of *CARD11* constitutional mutations in a rare syndrome, called BENTA (B-cell Expansion with NF-kB and T-cell Anergy).[138] This rare disorder is characterized by polyclonal B-cell lymphocytosis from birth and carries an increased risk for lymphoma development. That four of the five *CARD11* mutations were detected in patients with a concurrent *KMT2D* mutation, may equally suggest that *CARD11* gain-of-function mutation on its own is not sufficient for lymphomagenesis, but suggests that both genes act in concert. Whether patients with *CARD11* mutation have more severe CAD than patients without this mutation is of interest, but needs to be studied in larger patient series.

Diagnostic testing for *MYD88*, *KMT2D* and *CARD11* mutations may help in diagnosis of CAD, and to exclude LPL.

Concluding remarks and further perspectives

In conclusion, a common clonal origin of MBL/MSBC and DLBCL was demonstrated in 3/6 cases and suggest that DLBCL may arise from small precursor cells, with either a CLL, non-CLL or GC immunophenotype.

More paired MSBC and DLBCL samples and corresponding normal controls (T-cells) should prospectively be collected, for the analysis of mutations by NGS. Since MSBCs and DLBCLs may stem from a common precursor cell, it is expected that these precursor cells have some shared genetic aberrations. If found, it will be of interest whether these genetic aberrations are different from regular MBL and whether these mutations predict for DLBCL development. In case the genetic aberrations would not be different from MBL, this information would still be important. Recently, Barrio et al[139] have found that MBL shows mutations in important driver genes several years before progression to CLL, as evaluated by targeted sequencing. Most of these mutated genes, were detected at similar frequency as in CLL. Additionally, MBLs that show subclonal expansion on sequential analysis had shorter progression time to CLL and shorter time-to treatment. These findings indicate that early detection of mutations in MBL may be important for future clinical decisions. Diagnosis is moving rapidly forward, and it is already possible to diagnose DLBCL using circulating DNA from cell-free serum samples.[140] Circulating cell-free DNA can be detected at a high sensitivity by droplet PCR or by using NGS on a selected gene panel.[141] If these new diagnostic methods should be used in a routine diagnostic laboratory in the future, mutations in precursor lesions should be distinguished from those in high grade lesions, to properly diagnose and treat patients.

Our data show that in addition to *IGHV*, *IG* light chain usage is highly restricted in CAD. The data indicate that multiple regions within the immunoglobulin heavy chain, as well as immunoglobulin light chain, contribute to I antigen binding and may determine activity of the disease. Of practical consequence, the highly restricted usage of IGKV3-20 provides a rationale for ‘off-the-shelf’ vaccination with IGKV3-20 proteins, known to be immunogenic and being considered for treatment in other lymphoproliferative diseases.[128]

KMT2D and *CARD11* are the genes that are most recurrently mutated in CAD patients. It is necessary to confirm whether the detected mutations have an effect on protein expression and function. *KMT2D* expression will be analyzed on the RNA and protein level. First, we will extract RNA and perform qRT-PCR and use normal B cells as control. In addition, for samples with enough material we will run a semi-quantitative analysis of

KMT2D expression by Western blot.[132] Normal B cells of the patient will be used as control. We will investigate the impact of *CARD11* mutations by using a functional assay that measures constitutive activation of NF-κB pathway in cells, similarly as described for DLBCLs.[54] This will be done by expressing mutated CARD11 proteins in Jurkat JPM50.6 cell line that does not express endogenous CARD11 and contains a NF-kB–driven enhanced green fluorescent protein (EGFP) reporter and results will be compared to cells expressing wild-type CARD11 as control.

The finding of recurrent *KMT2D* and *CARD11* mutations in CAD, suggests that demonstration of these mutations may be used to properly diagnose the disease and distinguish it from LPL. CAD-associated lymphoproliferative disease is an indolent disease[89,130], that does not progress to systemic lymphoproliferation and does not transform to clinically aggressive lymphoma, as seen in LPL and other small B-cell lymphoma types. Also, the demonstration of recurrent *KMT2D* and *CARD11* mutations in CAD suggests that targeted treatment might be attempted instead of chemoimmune therapy to provide a potentially less toxic treatment of the disease. HDAC inhibitors, counteracting the effects of histone methyl transferase malfunction, have been used as novel agents for many cancers, including lymphoma and myeloma.[104] In addition, HDAC inhibitors have recently been tested in models of Kabuki syndrome.[142] Our findings provide a rationale for future testing of HDAC inhibitors in treatment of CAD. Targeted therapy to counteract the effect of *CARD11* gain-of-function mutations has more recently been developed and is currently being evaluated for use in diffuse large B-cell lymphoma.[103] Perhaps, such therapy could also be considered in future clinical trials for CAD.

In order to find less recurrent mutations and pathways affected in CAD, exome sequencing needs to be performed in a larger number of cases. Our group is planning to sequence 10 more exomes followed by targeted sequencing and functional studies depending on the nature of the findings. Samples for such analysis are being collected as a part of the CAD5 clinical trial. After the trial is concluded and analyzed, the molecular findings could also be compared with the treatment outcome. Since not all patients respond to therapy, finding markers that predict response to therapy is of interest.

The study of the mutations in CAD combined with our previously published histology data, indicate that CAD is different from other small B-cell lymphoproliferative diseases and that CAD is a separate disease entity. We propose that disease name should be changed to Cold Agglutinin associated B-cell Lymphoproliferative Disease or CALD.

References

1. Abbas AK, Lichtman AH, Pillai S. Cellular and molecular immunology (ed 7th). Philadelphia: Elsevier/Saunders; 2012.
2. Ehrenstein MR, Notley CA. The importance of natural IgM: scavenger, protector and regulator. *Nat Rev Immunol*. 2010;10(11):778-786.
3. Pieper K, Grimbacher B, Eibel H. B-cell biology and development. *J Allergy Clin Immunol*. 2013;131(4):959-971.
4. van Dongen JJ, Langerak AW, Bruggemann M, et al. Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. *Leukemia*. 2003;17(12):2257-2317.
5. Lefranc MP. Nomenclature of the human immunoglobulin heavy (IGH) genes. *Exp Clin Immunogenet*. 2001;18(2):100-116.
6. Lefranc MP. Nomenclature of the human immunoglobulin kappa (IGK) genes. *Exp Clin Immunogenet*. 2001;18(3):161-174.
7. Lefranc MP. Nomenclature of the human immunoglobulin lambda (IGL) genes. *Exp Clin Immunogenet*. 2001;18(4):242-254.
8. Milstein C, Neuberger MS, Staden R. Both DNA strands of antibody genes are hypermutation targets. *Proc Natl Acad Sci U S A*. 1998;95(15):8791-8794.
9. Rogozin IB, Diaz M. Cutting edge: DGYW/WRCH is a better predictor of mutability at G:C bases in Ig hypermutation than the widely accepted RGYW/WRCY motif and probably reflects a two-step activation-induced cytidine deaminase-triggered process. *J Immunol*. 2004;172(6):3382-3384.
10. Rogozin IB, Kolchanov NA. Somatic hypermutagenesis in immunoglobulin genes. II. Influence of neighbouring base sequences on mutagenesis. *Biochim Biophys Acta*. 1992;1171(1):11-18.
11. Rogozin IB, Pavlov YI, Bebenek K, Matsuda T, Kunkel TA. Somatic mutation hotspots correlate with DNA polymerase eta error spectrum. *Nat Immunol*. 2001;2(6):530-536.
12. Young RM, Staudt LM. Targeting pathological B cell receptor signalling in lymphoid malignancies. *Nat Rev Drug Discov*. 2013;12(3):229-243.
13. Sommer K, Guo B, Pomerantz JL, et al. Phosphorylation of the CARMA1 linker controls NF-kappaB activation. *Immunity*. 2005;23(6):561-574.
14. Shinohara H, Yasuda T, Aiba Y, et al. PKC beta regulates BCR-mediated IKK activation by facilitating the interaction between TAK1 and CARMA1. *J Exp Med*. 2005;202(10):1423-1431.
15. Leadbetter EA, Rifkin IR, Hohlbaum AM, Beaudette BC, Shlomchik MJ, Marshak-Rothstein A. Chromatin-IgG complexes activate B cells by dual engagement of IgM and Toll-like receptors. *Nature*. 2002;416(6881):603-607.
16. Green NM, Marshak-Rothstein A. Toll-like receptor driven B cell activation in the induction of systemic autoimmunity. *Semin Immunol*. 2011;23(2):106-112.
17. Swerdlow SH, Campo E, Harris NL, et al. WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues (Revised 4th edition): IARC; Lyon 2017.
18. Rickert RC. New insights into pre-BCR and BCR signalling with relevance to B cell malignancies. *Nat Rev Immunol*. 2013;13(8):578-591.

19. Fabbri G, Rasi S, Rossi D, et al. Analysis of the chronic lymphocytic leukemia coding genome: role of NOTCH1 mutational activation. *J Exp Med*. 2011;208(7):1389-1401.
20. Landau DA, Carter SL, Stojanov P, et al. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell*. 2013;152(4):714-726.
21. Puente XS, Pinyol M, Quesada V, et al. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature*. 2011;475(7354):101-105.
22. Quesada V, Conde L, Villamor N, et al. Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nat Genet*. 2011;44(1):47-52.
23. Wang L, Lawrence MS, Wan Y, et al. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *N Engl J Med*. 2011;365(26):2497-2506.
24. Ho C, Kluk MJ. Molecular Pathology: Predictive, Prognostic, and Diagnostic Markers in Lymphoid Neoplasms. *Surg Pathol Clin*. 2016;9(3):489-521.
25. Bodor C, O'Riain C, Wrench D, et al. EZH2 Y641 mutations in follicular lymphoma. *Leukemia*. 2011;25(4):726-729.
26. Green MR, Gentles AJ, Nair RV, et al. Hierarchy in somatic mutations arising during genomic evolution and progression of follicular lymphoma. *Blood*. 2013;121(9):1604-1611.
27. Loeffler M, Kreuz M, Haake A, et al. Genomic and epigenomic co-evolution in follicular lymphomas. *Leukemia*. 2015;29(2):456-463.
28. Morin RD, Johnson NA, Severson TM, et al. Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nat Genet*. 2010;42(2):181-185.
29. Morin RD, Mendez-Lago M, Mungall AJ, et al. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature*. 2011;476(7360):298-303.
30. Okosun J, Bodor C, Wang J, et al. Integrated genomic analysis identifies recurrent mutations and evolution patterns driving the initiation and progression of follicular lymphoma. *Nat Genet*. 2014;46(2):176-181.
31. Pasqualucci L, Dominguez-Sola D, Chiarenza A, et al. Inactivating mutations of acetyltransferase genes in B-cell lymphoma. *Nature*. 2011;471(7337):189-195.
32. Pastore A, Jurinovic V, Kridel R, et al. Integration of gene mutations in risk prognostication for patients receiving first-line immunochemotherapy for follicular lymphoma: a retrospective analysis of a prospective clinical trial and validation in a population-based registry. *Lancet Oncol*. 2015;16(9):1111-1122.
33. Kiel MJ, Velusamy T, Betz BL, et al. Whole-genome sequencing identifies recurrent somatic NOTCH2 mutations in splenic marginal zone lymphoma. *J Exp Med*. 2012;209(9):1553-1565.
34. Rossi D, Deaglio S, Dominguez-Sola D, et al. Alteration of BIRC3 and multiple other NF-kappaB pathway genes in splenic marginal zone lymphoma. *Blood*. 2011;118(18):4930-4934.
35. Rossi D, Trifonov V, Fangazio M, et al. The coding genome of splenic marginal zone lymphoma: activation of NOTCH2 and other pathways regulating marginal zone development. *J Exp Med*. 2012;209(9):1537-1551.
36. Vaque JP, Martinez N, Batlle-Lopez A, et al. B-cell lymphoma mutations: improving diagnostics and enabling targeted therapies. *Haematologica*. 2014;99(2):222-231.

37. Spina V, Khiabanian H, Messina M, et al. The genetics of nodal marginal zone lymphoma. *Blood*. 2016;128(10):1362-1373.
38. Novak U, Rinaldi A, Kwee I, et al. The NF-{kappa}B negative regulator TNFAIP3 (A20) is inactivated by somatic mutations and genomic deletions in marginal zone lymphomas. *Blood*. 2009;113(20):4918-4921.
39. Hunter ZR, Xu L, Yang G, et al. The genomic landscape of Waldenstrom macroglobulinemia is characterized by highly recurring MYD88 and WHIM-like CXCR4 mutations, and small somatic deletions associated with B-cell lymphomagenesis. *Blood*. 2014;123(11):1637-1646.
40. Poulain S, Roumier C, Decambron A, et al. MYD88 L265P mutation in Waldenstrom macroglobulinemia. *Blood*. 2013;121(22):4504-4511.
41. Treon SP, Cao Y, Xu L, Yang G, Liu X, Hunter ZR. Somatic mutations in MYD88 and CXCR4 are determinants of clinical presentation and overall survival in Waldenstrom macroglobulinemia. *Blood*. 2014;123(18):2791-2796.
42. Treon SP, Xu L, Yang G, et al. MYD88 L265P somatic mutation in Waldenstrom's macroglobulinemia. *N Engl J Med*. 2012;367(9):826-833.
43. Bea S, Valdes-Mas R, Navarro A, et al. Landscape of somatic mutations and clonal evolution in mantle cell lymphoma. *Proc Natl Acad Sci U S A*. 2013;110(45):18250-18255.
44. Camacho E, Hernandez L, Hernandez S, et al. ATM gene inactivation in mantle cell lymphoma mainly occurs by truncating mutations and missense mutations involving the phosphatidylinositol-3 kinase domain and is associated with increasing numbers of chromosomal imbalances. *Blood*. 2002;99(1):238-244.
45. Fang NY, Greiner TC, Weisenburger DD, et al. Oligonucleotide microarrays demonstrate the highest frequency of ATM mutations in the mantle cell subtype of lymphoma. *Proc Natl Acad Sci U S A*. 2003;100(9):5372-5377.
46. Greiner TC, Moynihan MJ, Chan WC, et al. p53 mutations in mantle cell lymphoma are associated with variant cytology and predict a poor prognosis. *Blood*. 1996;87(10):4302-4310.
47. Jares P, Campo E. Advances in the understanding of mantle cell lymphoma. *Br J Haematol*. 2008;142(2):149-165.
48. Kridel R, Meissner B, Rogic S, et al. Whole transcriptome sequencing reveals recurrent NOTCH1 mutations in mantle cell lymphoma. *Blood*. 2012;119(9):1963-1971.
49. Pasqualucci L, Trifonov V, Fabbri G, et al. Analysis of the coding genome of diffuse large B-cell lymphoma. *Nat Genet*. 2011;43(9):830-837.
50. Compagno M, Lim WK, Grunn A, et al. Mutations of multiple genes cause deregulation of NF-kappaB in diffuse large B-cell lymphoma. *Nature*. 2009;459(7247):717-721.
51. Davis RE, Ngo VN, Lenz G, et al. Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma. *Nature*. 2010;463(7277):88-92.
52. Honma K, Tsuzuki S, Nakagawa M, et al. TNFAIP3/A20 functions as a novel tumor suppressor gene in several subtypes of non-Hodgkin lymphomas. *Blood*. 2009;114(12):2467-2475.
53. Kato M, Sanada M, Kato I, et al. Frequent inactivation of A20 in B-cell lymphomas. *Nature*. 2009;459(7247):712-716.
54. Lenz G, Davis RE, Ngo VN, et al. Oncogenic CARD11 mutations in human diffuse large B cell lymphoma. *Science*. 2008;319(5870):1676-1679.

55. Lohr JG, Stojanov P, Lawrence MS, et al. Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. *Proc Natl Acad Sci U S A*. 2012;109(10):3879-3884.
56. Mandelbaum J, Bhagat G, Tang H, et al. BLIMP1 is a tumor suppressor gene frequently disrupted in activated B cell-like diffuse large B cell lymphoma. *Cancer Cell*. 2010;18(6):568-579.
57. Ngo VN, Young RM, Schmitz R, et al. Oncogenically active MYD88 mutations in human lymphoma. *Nature*. 2011;470(7332):115-119.
58. Pasqualucci L, Compagno M, Houldsworth J, et al. Inactivation of the PRDM1/BLIMP1 gene in diffuse large B cell lymphoma. *J Exp Med*. 2006;203(2):311-317.
59. Tam W, Gomez M, Chadburn A, Lee JW, Chan WC, Knowles DM. Mutational analysis of PRDM1 indicates a tumor-suppressor role in diffuse large B-cell lymphomas. *Blood*. 2006;107(10):4090-4100.
60. Reddy A, Zhang J, Davis NS, et al. Genetic and Functional Drivers of Diffuse Large B Cell Lymphoma. *Cell*. 2017;171(2):481-494.e415.
61. Swerdlow SH, International Agency for Research on Cancer., World Health Organization. WHO classification of tumours of haematopoietic and lymphoid tissues (ed 4th). Lyon, France: International Agency for Research on Cancer; 2008.
62. Nieto WG, Almeida J, Romero A, et al. Increased frequency (12%) of circulating chronic lymphocytic leukemia-like B-cell clones in healthy subjects using a highly sensitive multicolor flow cytometry approach. *Blood*. 2009;114(1):33-37.
63. Landgren O, Albitar M, Ma W, et al. B-cell clones as early markers for chronic lymphocytic leukemia. *N Engl J Med*. 2009;360(7):659-667.
64. Rawstron AC, Bennett FL, O'Connor SJ, et al. Monoclonal B-cell lymphocytosis and chronic lymphocytic leukemia. *N Engl J Med*. 2008;359(6):575-583.
65. Rawstron AC, Shanafelt T, Lanasa MC, et al. Different biology and clinical outcome according to the absolute numbers of clonal B-cells in monoclonal B-cell lymphocytosis (MBL). *Cytometry B Clin Cytom*. 2010;78 Suppl 1:S19-23.
66. Vardi A, Dagklis A, Scarfo L, et al. Immunogenetics shows that not all MBL are equal: the larger the clone, the more similar to CLL. *Blood*. 2013;121(22):4521-4528.
67. Xochelli A, Oscier D, Stamatopoulos K. Clonal B-cell lymphocytosis of marginal zone origin. *Best Pract Res Clin Haematol*. 2017;30(1-2):77-83.
68. Xochelli A, Kalpadakis C, Gardiner A, et al. Clonal B-cell lymphocytosis exhibiting immunophenotypic features consistent with a marginal-zone origin: is this a distinct entity? *Blood*. 2014;123(8):1199-1206.
69. Bruscaggin A, Monti S, Arcaini L, et al. Molecular lesions of signalling pathway genes in clonal B-cell lymphocytosis with marginal zone features. *Br J Haematol*. 2014;167(5):718-720.
70. Randen U, Tierens AM, Tjonnfjord GE, Delabie J. Bone marrow histology in monoclonal B-cell lymphocytosis shows various B-cell infiltration patterns. *Am J Clin Pathol*. 2013;139(3):390-395.
71. Tierens AM, Holte H, Warsame A, et al. Low levels of monoclonal small B cells in the bone marrow of patients with diffuse large B-cell lymphoma of activated B-cell type but not of germinal center B-cell type. *Haematologica*. 2010;95(8):1334-1341.
72. Jegalian AG, Eberle FC, Pack SD, et al. Follicular lymphoma in situ: clinical implications and comparisons with partial involvement by follicular lymphoma. *Blood*. 2011;118(11):2976-2984.

73. Carvajal-Cuenca A, Sua LF, Silva NM, et al. In situ mantle cell lymphoma: clinical implications of an incidental finding with indolent clinical behavior. *Haematologica*. 2012;97(2):270-278.
74. Korde N, Kristinsson SY, Landgren O. Monoclonal gammopathy of undetermined significance (MGUS) and smoldering multiple myeloma (SMM): novel biological insights and development of early treatment strategies. *Blood*. 2011;117(21):5573-5581.
75. Mailankody S, Landgren O. Monoclonal gammopathy of undetermined significance and Waldenstrom's macroglobulinemia. *Best Pract Res Clin Haematol*. 2016;29(2):187-193.
76. Cullen MH, Lister TA, Brearley RI, Shand WS, Stansfeld AG. Histological transformation of non-Hodgkin's lymphoma: a prospective study. *Cancer*. 1979;44(2):645-651.
77. Pasqualucci L, Khiabanian H, Fangazio M, et al. Genetics of follicular lymphoma transformation. *Cell Rep*. 2014;6(1):130-140.
78. Eide MB, Liestol K, Lingjaerde OC, et al. Genomic alterations reveal potential for higher grade transformation in follicular lymphoma and confirm parallel evolution of tumor cell clones. *Blood*. 2010;116(9):1489-1497.
79. Villamor N, Conde L, Martinez-Trillos A, et al. NOTCH1 mutations identify a genetic subgroup of chronic lymphocytic leukemia patients with high risk of transformation and poor outcome. *Leukemia*. 2013;27(5):1100-1106.
80. Chigrinova E, Rinaldi A, Kwee I, et al. Two main genetic pathways lead to the transformation of chronic lymphocytic leukemia to Richter syndrome. *Blood*. 2013;122(15):2673-2682.
81. Starostik P, Greiner A, Schultz A, et al. Genetic aberrations common in gastric high-grade large B-cell lymphoma. *Blood*. 2000;95(4):1180-1187.
82. Starostik P, Patzner J, Greiner A, et al. Gastric marginal zone B-cell lymphomas of MALT type develop along 2 distinct pathogenetic pathways. *Blood*. 2002;99(1):3-9.
83. Montoto S, Fitzgibbon J. Transformation of indolent B-cell lymphomas. *J Clin Oncol*. 2011;29(14):1827-1834.
84. Alizadeh AA, Eisen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*. 2000;403(6769):503-511.
85. Lenz G, Wright G, Dave SS, et al. Stromal gene signatures in large-B-cell lymphomas. *N Engl J Med*. 2008;359(22):2313-2323.
86. Lenz G, Staudt LM. Aggressive lymphomas. *N Engl J Med*. 2010;362(15):1417-1429.
87. Roschewski M, Staudt LM, Wilson WH. Diffuse large B-cell lymphoma-treatment approaches in the molecular era. *Nat Rev Clin Oncol*. 2014;11(1):12-23.
88. Rawstron AC, Green MJ, Kuzmicki A, et al. Monoclonal B lymphocytes with the characteristics of "indolent" chronic lymphocytic leukemia are present in 3.5% of adults with normal blood counts. *Blood*. 2002;100(2):635-639.
89. Berentsen S, Randen U, Tjonnfjord GE. Cold agglutinin-mediated autoimmune hemolytic anemia. *Hematol Oncol Clin North Am*. 2015;29(3):455-471.
90. Jaffe CJ, Atkinson JP, Frank MM. The role of complement in the clearance of cold agglutinin-sensitized erythrocytes in man. *J Clin Invest*. 1976;58(4):942-949.
91. Shi J, Rose EL, Singh A, et al. TNT003, an inhibitor of the serine protease C1s, prevents complement activation induced by cold agglutinins. *Blood*. 2014;123(26):4015-4022.

92. Berentsen S, Bo K, Shammas FV, Myking AO, Ulvestad E. Chronic cold agglutinin disease of the "idiopathic" type is a premalignant or low-grade malignant lymphoproliferative disease. *APMIS*. 1997;105(5):354-362.
93. Berentsen S, Ulvestad E, Langholm R, et al. Primary chronic cold agglutinin disease: a population based clinical study of 86 patients. *Haematologica*. 2006;91(4):460-466.
94. Randen U, Troen G, Tierens A, et al. Primary cold agglutinin-associated lymphoproliferative disease: a B-cell lymphoma of the bone marrow distinct from lymphoplasmacytic lymphoma. *Haematologica*. 2014;99(3):497-504.
95. Malecka A, Troen G, Tierens A, et al. Immunoglobulin heavy and light chain gene features are correlated with primary cold agglutinin disease onset and activity. *Haematologica*. 2016;101(9):e361-364.
96. Shaffer AL, 3rd, Young RM, Staudt LM. Pathogenesis of human B cell lymphomas. *Annu Rev Immunol*. 2012;30:565-610.
97. Potter KN, Hobby P, Klijn S, Stevenson FK, Sutton BJ. Evidence for involvement of a hydrophobic patch in framework region 1 of human V4-34-encoded Igs in recognition of the red blood cell I antigen. *J Immunol*. 2002;169(7):3777-3782.
98. Berentsen S, Randen U, Vagan AM, et al. High response rate and durable remissions following fludarabine and rituximab combination therapy for chronic cold agglutinin disease. *Blood*. 2010;116(17):3180-3184.
99. Berentsen S, Randen U, Oksman M, et al. Bendamustine plus rituximab for chronic cold agglutinin disease: Results of a Nordic prospective multicenter trial. *Blood*. 2017.
100. Chao MP. Treatment challenges in the management of relapsed or refractory non-Hodgkin's lymphoma - novel and emerging therapies. *Cancer Manag Res*. 2013;5:251-269.
101. Rouce RH, Sharma S, Huynh M, Heslop HE. Recent advances in T-cell immunotherapy for haematological malignancies. *Br J Haematol*. 2017;176(5):688-704.
102. Younes A, Ansell S, Fowler N, et al. The landscape of new drugs in lymphoma. *Nat Rev Clin Oncol*. 2016.
103. Young RM, Staudt LM. A new "brew" of MALT1 inhibitors. *Cancer Cell*. 2012;22(6):706-707.
104. Imai Y, Maru Y, Tanaka J. Action mechanisms of histone deacetylase inhibitors in the treatment of hematological malignancies. *Cancer Sci*. 2016.
105. Spina V, Khiabanian H, Messina M, et al. The genetics of nodal marginal zone lymphoma. *Blood*. 2016.
106. Hans CP, Weisenburger DD, Greiner TC, et al. Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. *Blood*. 2004;103(1):275-282.
107. Payne K, Wright P, Grant JW, et al. BIOMED-2 PCR assays for IGK gene rearrangements are essential for B-cell clonality analysis in follicular lymphoma. *Br J Haematol*. 2011;155(1):84-92.
108. Li Y, Spellerberg MB, Stevenson FK, Capra JD, Potter KN. The I binding specificity of human VH 4-34 (VH 4-21) encoded antibodies is determined by both VH framework region 1 and complementarity determining region 3. *J Mol Biol*. 1996;256(3):577-589.
109. Capra JD, Kehoe JM, Williams RC, Jr., Feizi T, Kunkel HG. Light chain sequences of human IgM cold agglutinins (variable-region subgroups amino-acid sequence-kappa light chain-N-terminal). *Proc Natl Acad Sci U S A*. 1972;69(1):40-43.

110. Cauerhff A, Braden BC, Carvalho JG, et al. Three-dimensional structure of the Fab from a human IgM cold agglutinin. *J Immunol*. 2000;165(11):6422-6428.
111. Gergely J, Wang AC, Fudenberg HH. Chemical analyses of variable regions of heavy and light chains of cold agglutinins. *Vox Sang*. 1973;24(5):432-440.
112. Ruzickova S, Pruss A, Odendahl M, et al. Chronic lymphocytic leukemia preceded by cold agglutinin disease: intraclonal immunoglobulin light-chain diversity in V(H)4-34 expressing single leukemic B cells. *Blood*. 2002;100(9):3419-3422.
113. Silberstein LE, Jefferies LC, Goldman J, et al. Variable region gene analysis of pathologic human autoantibodies to the related i and I red blood cell antigens. *Blood*. 1991;78(9):2372-2386.
114. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754-1760.
115. Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics*. 2012;28(14):1811-1817.
116. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics*. 2009;25(21):2865-2871.
117. Cingolani P, Platts A, Wang le L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)*. 2012;6(2):80-92.
118. Robinson JT, Thorvaldsdottir H, Winckler W, et al. Integrative genomics viewer. *Nat Biotechnol*. 2011;29(1):24-26.
119. Sabouri Z, Schofield P, Horikawa K, et al. Redemption of autoantibodies on anergic B cells by variable-region glycosylation and mutation away from self-reactivity. *Proc Natl Acad Sci U S A*. 2014;111(25):E2567-2575.
120. Nelson BP, Abdul-Nabi A, Goolsby C, Winter J, Peterson L. Characterization of tissue findings in bone marrow biopsy specimens with small monoclonal B-cell populations. *Am J Clin Pathol*. 2014;141(5):687-696.
121. Hsu FJ, Levy R. Preferential use of the VH4 Ig gene family by diffuse large-cell lymphoma. *Blood*. 1995;86(8):3072-3082.
122. Sebastian E, Alcoceba M, Balanzategui A, et al. Molecular characterization of immunoglobulin gene rearrangements in diffuse large B-cell lymphoma: antigen-driven origin and IGHV4-34 as a particular subgroup of the non-GCB subtype. *Am J Pathol*. 2012;181(5):1879-1888.
123. Potter KN. Molecular characterization of cold agglutinins. *Transfus Sci*. 2000;22(1-2):113-119.
124. Young RM, Wu T, Schmitz R, et al. Survival of human lymphoma cells requires B-cell receptor engagement by self-antigens. *Proc Natl Acad Sci U S A*. 2015;112(44):13447-13454.
125. Zhu D, McCarthy H, Ottensmeier CH, Johnson P, Hamblin TJ, Stevenson FK. Acquisition of potential N-glycosylation sites in the immunoglobulin variable region by somatic mutation is a distinctive feature of follicular lymphoma. *Blood*. 2002;99(7):2562-2568.
126. Jefferies LC, Carchidi CM, Silberstein LE. Naturally occurring anti-i/I cold agglutinins may be encoded by different VH3 genes as well as the VH4.21 gene segment. *J Clin Invest*. 1993;92(6):2821-2833.

127. Marks JD, Ouwehand WH, Bye JM, et al. Human Antibody Fragments Specific for Human Blood Group Antigens from a Phage Display Library. *Nat Biotech*. 1993;11(10):1145-1149.
128. Martorelli D, Guidoboni M, De Re V, et al. IGKV3 proteins as candidate "off-the-shelf" vaccines for kappa-light chain-restricted B-cell non-Hodgkin lymphomas. *Clin Cancer Res*. 2012;18(15):4080-4091.
129. Henry Dunand CJ, Wilson PC. Restricted, canonical, stereotyped and convergent immunoglobulin responses. *Philos Trans R Soc Lond B Biol Sci*. 2015;370(1676).
130. Berentsen S, Tjonnfjord GE. Diagnosis and treatment of cold agglutinin mediated autoimmune hemolytic anemia. *Blood Rev*. 2012;26(3):107-115.
131. Swerdlow SH, Kuzu I, Dogan A, et al. The many faces of small B cell lymphomas with plasmacytic differentiation and the contribution of MYD88 testing. *Virchows Arch*. 2015.
132. Zhang J, Dominguez-Sola D, Hussein S, et al. Disruption of KMT2D perturbs germinal center B cell development and promotes lymphomagenesis. *Nat Med*. 2015;21(10):1190-1198.
133. Ng SB, Bigham AW, Buckingham KJ, et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet*. 2010;42(9):790-793.
134. Courtens W, Rassart A, Stene JJ, Vamos E. Further evidence for autosomal dominant inheritance and ectodermal abnormalities in Kabuki syndrome. *Am J Med Genet*. 2000;93(3):244-249.
135. Ortega-Molina A, Boss IW, Canela A, et al. The histone lysine methyltransferase KMT2D sustains a gene expression program that represses B cell lymphoma development. *Nat Med*. 2015;21(10):1199-1208.
136. Lindsley AW, Saal HM, Burrow TA, et al. Defects of B-cell terminal differentiation in patients with type-1 Kabuki syndrome. *J Allergy Clin Immunol*. 2016;137(1):179-187 e110.
137. Jeelall YS, Wang JQ, Law HD, et al. Human lymphoma mutations reveal CARD11 as the switch between self-antigen-induced B cell death or proliferation and autoantibody production. *J Exp Med*. 2012;209(11):1907-1917.
138. Snow AL, Xiao W, Stinson JR, et al. Congenital B cell lymphocytosis explained by novel germline CARD11 mutations. *J Exp Med*. 2012;209(12):2247-2261.
139. Barrio S, Shanafelt TD, Ojha J, et al. Genomic characterization of high-count MBL cases indicates that early detection of driver mutations and subclonal expansion are predictors of adverse clinical outcome. *Leukemia*. 2017;31(1):170-176.
140. Scherer F, Kurtz DM, Newman AM, et al. Distinct biological subtypes and patterns of genome evolution in lymphoma revealed by circulating tumor DNA. *Sci Transl Med*. 2016;8(364):364ra155.
141. Wan JC, Massie C, Garcia-Corbacho J, et al. Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nat Rev Cancer*. 2017;17(4):223-238.
142. Bjornsson HT, Benjamin JS, Zhang L, et al. Histone deacetylase inhibition rescues structural and functional brain deficits in a mouse model of Kabuki syndrome. *Sci Transl Med*. 2014;6(256):256ra135.

I

II

III

Frequent somatic mutations of *KMT2D* (*MLL2*) and *CARD11* genes in primary cold agglutinin disease

Agnieszka Małecka[1,2], Gunhild Trøen[1], Anne Tierens[3], Ingunn Østlie[1], Jędrzej Małecki[4], Ulla Randen[1], Junbai Wang[1], Sigbjørn Berentsen[5], Geir E. Tjønnfjord[6] and Jan M.A. Delabie[3]

[1]Department of Pathology, Oslo University Hospital, Oslo, Norway; [2]Faculty of Medicine, University of Oslo, Oslo, Norway; [3]Laboratory Medicine Program, University Health Network and University of Toronto, Toronto, ON, Canada; [4]Department of Biological Sciences, University of Oslo, Oslo, Norway, [5]Department of Research and Innovation, Haugesund Hospital, Helse Fonna, Haugesund, Norway; [6]Department of Haematology, Oslo University Hospital and Institute of Clinical Medicine, University of Oslo, Oslo, Norway

Running short title:

***KMT2D (MLL2)* and *CARD11* mutations in cold agglutinin disease**

Corresponding Author:

Agnieszka Małecka

Department of Pathology, Oslo University Hospital, The Norwegian Radium Hospital

P.O.Box 4950, Nydalen, NO-0424 Oslo, Norway

E-mail: a.m.malecka@medisin.uio.no, machynia@yahoo.com

A version of this manuscript was accepted for publication by British Journal of Haematology on 11th of October 2017.

Abstract

Primary cold agglutinin disease (CAD), a rare hemolytic anemia caused by monoclonal IgM anti-I autoantibodies, was previously demonstrated to be associated with a small B-cell lymphoproliferative disorder of the bone marrow. This B-cell lymphoproliferative disorder has histopathological characteristics different from lymphoplasmacytic lymphoma and lacks the *MYD88* L265P mutation. Using flow cytometry-assisted cell sorting of bone marrow, exome sequencing and targeted sequencing of clonal B cells and control T cells, we searched for recurring mutations in CAD. A total of 16 well-characterized patients were analysed. Recurrent mutations of *KMT2D* (11/16, 69%) and *CARD11* (5/16, 31%) were found. The mutations were concurrent in four patients. These findings confirm that lymphoproliferative disease associated with primary CAD is a unique small B-cell lymphoproliferative disease of the bone marrow. The identification of the mutations may be useful to properly diagnose the disease and discern it from other small B-cell lymphomas affecting the bone marrow. Whether targeted treatment for CAD, in view of the high frequency of *KMT2D* mutation is an option, remains to be demonstrated.

Keywords:

Cold agglutinin disease, hemolytic anemia, lymphoproliferative disorder, KMT2D, CARD11.

Introduction

Primary chronic cold agglutinin disease (CAD) is a rare type of hemolytic anemia mediated by monoclonal IgM anti-I autoantibodies. The antibodies bind to erythrocytes at low temperatures, mostly in acral parts of the body, causing agglutination and complement activation (Berentsen*, et al* 2015, Berentsen and Tjonnfjord 2012). CAD represents 15% of all cases of autoimmune hemolytic anemia and has an incidence of 1 per million per year. Typically, patients display circulating monoclonal antibodies encoded by the immunoglobulin heavy chain gene *IGHV4-34*. The framework region 1 (FR1) of *IGHV4-34* encodes for an AVY sequence that determines binding to I antigen (Potter*, et al* 2002). The antibody light chain is mostly encoded by *IGKV3-20* and displays highly homologous antigen-binding regions in a subgroup of patients, indicating that the light chain also contributes to I-antigen binding (Malecka*, et al* 2016). Although CAD is characterized by monoclonal gammopathy, patients typically display no signs or symptoms of lymphoma or myeloma at diagnosis. In addition, CAD does not progress to myeloma or lymphoma upon long term follow-up (Berentsen*, et al* 2015). However, clonal B-cells are demonstrated in the bone marrow of most patients. The bone marrow disease was previously thought to be lymphoplasmacytic lymphoma or marginal zone lymphoma (Berentsen*, et al* 2006). However, more recent studies, including flow cytometry analysis demonstrated a homogeneous histology and immunophenotype of the bone marrow B-cell lymphoproliferative disease (de Tute, et al , Randen, et al 2014). The bone marrow usually shows a patchy intraparenchymatous B-cell infiltrate with a CD20+, sIgM+, sIgD+, CD5-/+, CD10-, CD11c-, CD23-, CD27+, CD38- immunophenotype. In addition, the MYD88 L265P mutation, typical of lymphoplasmacytic lymphoma, is not present in CAD (Randen*, et al* 2014, Schmidt*, et al* 2015). Together, the clinical and pathology findings suggest that CAD patients have a unique low-grade clonal B-cell lymphoproliferative disease of the bone marrow. The molecular changes underlying most

frequent lymphoproliferative diseases have now extensively been mapped by genome-wide genetic studies (Rosenquist, *et al* 2016). The current study was undertaken to find molecular changes in CAD-associated lymphoproliferative disease, potentially useful for diagnosis or targeted treatment.

Materials and Methods

We performed exome sequencing using monoclonal B cells as test samples and T cells as normal control in a test cohort followed by targeted sequencing in a validation cohort.

Patient materials

The patients included in this study were enrolled in the prospective CAD-5 trial (NTC02689986). Bone marrow samples were collected for diagnosis and for flow cytometry-assisted cell sorting (FACS) for the purpose of this study. In total, samples of 19 patients were prospectively collected for this study. These patients have been included in a previous study that also included retrospectively collected bone marrow samples. The retrospective samples were not suited for FACS analysis and hence could not be included in the present study (Malecka, *et al* 2016). CAD was diagnosed according to strict criteria (Berentsen, *et al* 2010). The clinical features have previously been reported using the same patient reference codes (Malecka, *et al* 2016). Clinical data are summarized in Supplementary Tab I. None of the patients showed extramedullar involvement with lymphoma. Bone marrow samples consisted of either a bone marrow trephine or a bone marrow aspirate or both. Bone marrow trephines were routinely processed for histological and immunohistochemical analysis. The aspirate was used partly for morphological and flow cytometric analyses. Another part of the aspirate was used and processed within 24 hours for FACS analysis. All cases demonstrated a clonal B cell population in the bone marrow. The morphologic and immunophenotypic findings have been described before (Malecka, *et al* 2016, Randen, *et al* 2014). The histology of a typical

case is illustrated in Supplementary Fig 1, the flow cytometry findings are given in Supplementary Tab II. Briefly, intraparenchymatous non-paratrabecular aggregates of B cells were found in all cases documented by a bone marrow trephine biopsy. Infiltration with B cells ranged between 0.5 and 70% (median 15%).

The project was approved by the Regional Committee for Medical and Health Research Ethics of South-East Norway (REK SØ 2012/131).

DNA extraction and whole genome amplification

Monotypic B cells and control T cells were isolated by FACS as previously described (Supplementary Fig 2) (Malecka*, et al* 2016). FACS resulted in over 90% pure monotypic B cells and control T cells, respectively. Monotypic B cells and control T cells were subsequently analysed by next generation sequencing (NGS). The analysis of pure cell populations by NGS greatly reduces the risk of detecting false positive and false negative somatic mutations. Additionally, it facilitates interpretation of mutation frequency in the clonal B-cell population. DNA from purified cells was extracted using Qiagen AllPrep DNA/RNA Micro kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The concentration of extracted nucleic acid was measured using a NanoDrop 2000 spectrometer (Thermo Scientific, Waltham, MA) and Qubit (Life Technologies). Infiltration of bone marrow with CAD-associated lymphoproliferative disease is typically variable and ranged between 0.5 and 70%. Hence, the number of B cells isolated from these patients, as well as DNA extracted from these cells, was limited in some patients and varied from patient to patient. Samples with the highest DNA content were used for exome sequencing (CAD-2, -5, -7, -19, -20 and -22), whereas samples with the lowest DNA content (CAD-1, -3, -4, -6, -10, -12, -13, -14, -15, -18) were used for targeted sequencing. Samples from three of the 19 patients did not yield enough DNA for further analysis.

For exome sequencing, five sample pairs (monotypic B cells/control T cells), with sufficient DNA, were used without amplification, while one sample pair, with a lower DNA yield, was amplified using Illustra Ready-To-Go GenomiPhi V3 DNA Amplification Kit (GE Healthcare Life Sciences, U.K.). For targeted sequencing 9 of 10 samples required whole genome amplification.

Next generation sequencing and Sanger sequencing validation of NGS results

DNA from monotypic B cells, as well as paired control T cells, was analysed by exome sequencing, which was performed at BGI Tech Solutions (Hong Kong) using the Agilent SureSelect Human All Exon V4 Reagent Kit and Illumina HiSeq technology. Depending on DNA availability, between 200 ng and 3 μg DNA was analysed. Amplified genomic DNA was used in one case. The exome sequencing procedure included UTR that gave both coding and regulatory regions with 50x coverage, pair-end reads (PE) and about 100 bp long reads. Since pure cell populations (FACS acquired) were analysed, a 50x coverage was deemed sufficient.

Analysis of sequencing data included pre-processing of raw data (removal of adapters and trimming of low quality bases), followed by quality control using the FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Next, sequence alignment was performed using BWA 0.7.8 (http://bio-bwa.sourceforge.net/)(Li and Durbin 2009), a well-known short reads alignment tool. Reads were aligned to human reference genome (Genome Reference Consortium GRCh37; released in Feb 2009). The raw BAM alignment files were treated by a standard GATK variant calling pipeline (verify mate-pair information between mates, mark duplicated reads, local sequence realignment, and base quality score recalibration). Analysis was performed by two Picard tools 1.113 (https://broadinstitute.github.io/picard/): FixMateInformation and MarkDuplicates, followed by two programs from the Genome Analysis Toolkit 3.1-1 (GATK)

(https://www.broadinstitute.org/gatk/): IndelRealigner and BaseRecalibrator. Subsequently, detection of somatic variants was performed by two programs: Strelka 1.0.14 (https://sites.google.com/site/strelkasomaticvariantcaller/)(Saunders, *et al* 2012) and MuTect v1 (http://gatkforums.broadinstitute.org/gatk/categories/mutect), which simultaneously analyse tumor-normal paired samples. Somatic variants that passed through their stringent filters were taken for further analysis (Cibulskis, *et al* 2013, Saunders, *et al* 2012). MuTect is able to detect single nucleotide polymorphisms/variations (SNPs/SNVs), while Strelka detects both SNPs/SNVs and small insertions/deletions (indels). In order to analyse the data for large indels and breakpoints, the Pindel program (http://gmt.genome.wustl.edu/packages/pindel/)(Ye, *et al* 2009) was used. This program uses a pattern growth approach to identify the breakpoints in paired-end short reads. Annotation was performed by the SnpEff program (http://snpeff.sourceforge.net/) (Cingolani, *et al* 2012). Somatic mutations were considered relevant and chosen for further analysis if identified concurrently by two programs: SNPs/SNVs by both MuTect and Strelka; indels by both Strelka and Pindel. In addition, somatic mutations identified with a high quality score (Supplementary Tab III) by only one of the programs were included after manual verification using the IGV 2.3.34 browser (https://www.broadinstitute.org/igv/)(Robinson, *et al* 2011). Such manual verification allowed to distinguish obvious technical errors from real mutations. Although laborious, this is necessary when analysing NGS data from samples with low DNA quantities.

Targeted NGS of genes with recurrent mutations was performed on 10 additional CAD samples. Targeted NGS was performed only for genes that were found mutated by exome sequencing in at least two CAD cases. Only non-synonymous mutations in coding part of the genome that were classified as ‘high’ or ‘moderate’ impact by SnpEff, were considered (Supplementary Tab III). Prior to targeted NGS, DNA from all CAD cases, except one (CAD-

1), was amplified using Illustra Ready-To-Go GenomiPhi V3 DNA Amplification Kit. Sequencing was done at Novogene (Hong Kong) on an Illumina HiSeq instrument with a coverage of about 200x and pair-end reads (PE) of 150 bp long. Raw sequencing data were processed in the same way as the exome sequencing data, except for MuTect and Strelka programs that could not be used since targeted NGS of normal T cell controls was not performed. Instead, variant calling from GATK was used.

All recurrent mutations, detected by NGS in B cells, were verified by Sanger sequencing for all CAD cases, and their somatic status was additionally verified by Sanger sequencing of paired normal T cell controls. For each mutation, individual primers were designed using an NCBI tool (http://www.ncbi.nlm.nih.gov/tools/primer-blast/). For PCR reaction, AmpliTaq Gold® 360 Master Mix (Thermo Fisher Scientific) was used, followed by sequencing using BigDye® Terminator v1.1 Cycle Sequencing Kit (Thermo Fisher Scientific) and ABI Prism 3130 sequencer.

Results

Recurrent *KMT2D* mutations

Somatic mutations of the *KMT2D* gene were detected in 11 out of 16 cases (69%) (Tab I; Fig 1). Seven of these mutations were classified as high impact by SnpEff, and consisted of: 2 out-of-frame deletions, 1 out-of-frame insertion, 3 stop codons gained and 1 splice donor variant. These mutations occurred N-terminally from the SET domain, which is located at the C-terminus of KMT2D and confers its lysine- (K) specific methyltransferase (KMT) activity. These mutations will result in a C-terminally truncated protein that lacks the SET-domain and, therefore, is enzymatically inactive. In two other cases, we identified missense mutations located in the C-terminal domain, which may also impair the activity of the SET domain or result in KMT loss-of-function (Zhang*, et al* 2015). Ng*, et al* (2010) found several pathogenic

missense variants in the C-terminal domain of KMT2D associated with Kabuki syndrome. Interestingly, missense mutations at Arg5048 and Cys5481 residues, found in two CAD patients (CAD-14 and CAD-19), were previously identified also in patients with Kabuki syndrome (Banka*, et al* 2012, Makrythanasis*, et al* 2013, Miyake*, et al* 2013), thus strongly indicating the importance of these residues for proper KMT2D function. Finally, two patients showed *KMT2D* mutations causing splice region variants classified by SnpEff as of low impact. Functional tests are therefore needed to determine their effect on protein function.

Since pure cell populations (FACS acquired) were analysed, mutations affecting only one allele of the gene are expected to be present in about 40-60% of reads, as detected by NGS, while mutations affecting both alleles would be expected in about 90-100% of reads. The variant allele frequency of *KMT2D* was approximately 40-60% for all patients (Supplementary Fig 3 and 4), thus indicating that only one allele was mutated.

Recurrent *CARD11* mutations

CARD11 was somatically mutated in 5 out of 16 cases (31%). All mutations were classified as of moderate impact by the SnpEff program. Three samples had a missense mutation, of which two had exactly the same mutation, one sample had a 3 bp in-frame deletion and one sample had a 3 bp in-frame insertion. The latter mutations were in the same position. All 3 missense mutations are classified as pathogenic by Cosmic (the Catalogue Of Somatic Mutations In Cancer; http://cancer.sanger.ac.uk/cosmic) (Tab I). The five CARD11 mutations were located within a 20bp stretch of exon 6, coding for the BAR domain of the coiled-coil region of CARD11 (Tab I; Fig 2). Proper formation of alpha-helical coiled-coil regions is strictly dependent on the precise spatial distribution of hydrophobic and charged residues, arranged in heptad repeats within the primary structure (Lupas*, et al* 1991, Parry*, et al* 2008). This arrangement allows formation of a hydrophobic ‘stripe’ on one side of the alpha-helix, required for their coiling. Therefore, disruption of the heptad repeats, caused by

in-frame insertions or deletions as seen in two patients, are likely to interrupt the formation of coiled-coil. Mutations substituting a residue for a proline, were seen in three patients. Proline is a well-known alpha-helix disruptor, likely interfering with coiled-coil formation in these patients. Mutations localized in BAR domain and coiled-coil region of CARD11, including in-frame preserving indels and missense mutations to proline, were previously demonstrated in diffuse large B-cell lymphoma and shown to induce constitutive activation of the NF-kB pathway (Lenz*, et al* 2008). It is therefore likely that *CARD11* mutations identified in CAD have a similar impact on NF-kB pathway activation. The variant allele frequency of *CARD11* was approximately 40-60% for the five patients (Supplementary Fig 3 and 4), indicating that only one allele of the gene was mutated.

Other recurrent mutations were not detected using criteria as described in Materials and Methods.

Discussion

We have demonstrated that *KMT2D* and *CARD11* are frequently mutated in CAD. Other recurrent gene mutations have not been identified. *KMT2D* and *CARD11* mutations, combined with previously described distinctive histology and immune phenotypic findings (Malecka*, et al* 2016, Randen*, et al* 2014), establishes CAD-associated B-cell lymphoproliferative disease as a unique disease, distinct from other B-cell lymphoproliferative diseases. Of note, our findings conclusively establish that CAD-associated B cell lymphoproliferative disease is distinct from lymphoplasmacytic lymphoma with which it is often confounded. The latter lymphoma shows *MYD88* L265P mutation in more than 90% of cases, and does not show *KMT2D* or *CARD11* mutation (Treon*, et al* 2012). *MYD88* mutation is not found in CAD-associated lymphoproliferative disease, as demonstrated in this study and our previous study (Malecka*, et al* 2016).

Most of the *KMT2D* mutations in CAD result in truncated KMT2D protein that lacks the enzymatic SET domain. *KMT2D* SET domain-inactivating mutations are not unique to CAD-associated lymphoproliferative disease, but are frequent in follicular lymphoma, nodal marginal zone lymphoma and diffuse large B cell lymphoma (Morin*, et al* 2011, Pasqualucci*, et al* 2014, Pasqualucci*, et al* 2011, Spina*, et al* 2016). It has been suggested that *KMT2D* mutation along with other chromatin modulating genes may be an early event, with other genetic changes occurring later (Okosun*, et al* 2014, Zhang*, et al* 2015). Recurrent mutations other than *KMT2D* and *CARD11* could not be demonstrated in our study. Larger case series of this rare disease will be necessary to study less frequent recurrent mutations, if any.

The variant allele frequency was approximately 40-60% and only one *KMT2D* mutation was found per patient, suggesting that only one allele is mutated. This is in keeping with monoallelic *KMT2D* mutations found in other lymphoma types. Monoallelic *KMT2D* mutations are also associated with Kabuki syndrome, a congenital autosomal dominant disorder characterized by facial abnormalities, multiorgan anomalies and mental impairment (Courtens*, et al* 2000, Ng*, et al* 2010). Monoallelic *KMT2D* mutations inactivating the SET domain cause partial loss of protein expression and increased B cell proliferation (Zhang*, et al* 2015). *KMT2D* mutation also impedes class switch recombination (Ortega-Molina*, et al* 2015). This is consistent with the observation that patients with Kabuki syndrome with *KMT2D* constitutional mutation show reduced levels of class-switched B-cells (Lindsley*, et al* 2016). *KMT2D* has therefore been suggested to be a tumor suppressor gene, the function of which is compromised when haplo-insufficient (Ortega-Molina*, et al* 2015). These data combined suggest that *KMT2D* mutation is a driver mutation for B-cell lymphoproliferative disease. *KMT2D* mutation might act in concert with B cell survival signals induced by stimulation of the auto-reactive *IGHV4-34*-encoded immunoglobulin receptor expressed on the surface of CAD B cells (Young*, et al* 2015). A growth advantage through auto-antigen

stimulation of the *IGHV4-34*-encoded B-cell receptor was recently demonstrated for a subset of diffuse large B-cell lymphoma, suggesting a similar role in CAD (Young*, et al* 2015).

Five out of the 16 CAD patients did not show somatic *KMT2D* mutation. However, whether other mechanisms affecting KMT2D protein expression, such as epigenetic silencing may be present in these cases, needs as yet to be investigated (Zhang*, et al* 2015).

CARD11 was somatically mutated in 5 of 16 CAD cases. Interestingly, the 5 mutations were located in a short 20 bp region of the coiled-coil domain in exon 6. *CARD11* coiled-coil domain mutations were previously detected in diffuse large B-cell lymphomas of activated B cell origin (Lenz*, et al* 2008). These mutations result in constitutive NF-kB activation and enhanced NF-kB activity upon antigen receptor stimulation (Lenz*, et al* 2008). Mono-allelic *CARD11* coiled-coil domain mutations are not oncogenic *per se* in mice, but result in B-cell proliferation and auto-antibody production (Jeelall*, et al* 2012). These findings are consistent with the recent discovery of *CARD11* constitutional mutations in a rare syndrome, called BENTA (B-cell Expansion with NF-kB and T-cell Anergy) (Snow*, et al* 2012). This rare disorder is characterized by polyclonal B cell lymphocytosis from birth and carries an increased risk for lymphoma development. Four of the five *CARD11* mutations were detected in patients with a concurrent *KMT2D* mutation, suggesting that *CARD11* gain-of-function mutation on its own is not sufficient for lymphomagenesis, and that both genes may act in concert. Whether patients with *CARD11* mutation have more severe CAD than patients without this mutation is of interest but needs to be studied in larger patient series.

The finding of recurrent *KMT2D* and *CARD11* mutations in CAD, suggests that demonstration of these mutations may be used to properly diagnose the disease and distinguish it from lymphoplasmacytic lymphoma. CAD-associated lymphoproliferative disease is an indolent disease (Berentsen*, et al* 2015, Berentsen and Tjonnfjord 2012), which does not progress to systemic lymphoproliferative disease and does not transform to

aggressive lymphoma, as seen in lymphoplasmacytic lymphoma and other small B-cell lymphoma types. Therefore, the term CAD-associated B-cell lymphoproliferative disease instead of CAD-associated B-cell lymphoma is suggested. Also, the demonstration of recurrent *KMT2D* and *CARD11* mutations in CAD suggests that targeted treatment might be attempted, resulting in potentially less toxic treatment of the disease. Histone deacetylase (HDAC) inhibitors, counteracting the effects of histone methyl transferase malfunction, have been used as novel agents for many cancers, including lymphoma and myeloma (Imai*, et al* 2016). In addition, HDAC inhibitors have recently been tested in models of Kabuki syndrome (Bjornsson*, et al* 2014). Our findings provide a rationale for future testing of HDAC inhibitors in the treatment of CAD. Targeted therapy to counteract the effect of *CARD11* gain-of-function mutations has more recently been developed and is currently being evaluated for use in diffuse large B cell lymphoma (Young and Staudt 2012). Perhaps, such therapy could also be considered in future clinical trials for CAD.

Acknowledgements

This study was supported by Health Region Authority South-East Norway and by the Norwegian Cancer Society.

Authorship

Contribution: AM, GT, AT, SB, GET and JD designed the study. AM, GT, IØ, JM and JW performed the analyses. AT, SB, GET and JD supervised the study. AT, UR and JD reviewed the diagnostic patient samples, GET and SB collected the clinical data. AM, GT, JM and JD prepared the manuscript. All authors have critically read the manuscript.

Competing interests Statement

Competing interests: the authors have no competing interests.

References

Banka, S., Veeramachaneni, R., Reardon, W., Howard, E., Bunstone, S., Ragge, N., Parker, M.J., Crow, Y.J., Kerr, B., Kingston, H., Metcalfe, K., Chandler, K., Magee, A., Stewart, F., McConnell, V.P., Donnelly, D.E., Berland, S., Houge, G., Morton, J.E., Oley, C., Revencu, N., Park, S.M., Davies, S.J., Fry, A.E., Lynch, S.A., Gill, H., Schweiger, S., Lam, W.W., Tolmie, J., Mohammed, S.N., Hobson, E., Smith, A., Blyth, M., Bennett, C., Vasudevan, P.C., Garcia-Minaur, S., Henderson, A., Goodship, J., Wright, M.J., Fisher, R., Gibbons, R., Price, S.M., D, C.d.S., Temple, I.K., Collins, A.L., Lachlan, K., Elmslie, F., McEntagart, M., Castle, B., Clayton-Smith, J., Black, G.C. & Donnai, D. (2012) How genetically heterogeneous is Kabuki syndrome?: MLL2 testing in 116 patients, review and analyses of mutation and phenotypic spectrum. *European Journal of Human Genetics,* **20,** 381-388.

Berentsen, S., Randen, U. & Tjonnfjord, G.E. (2015) Cold agglutinin-mediated autoimmune hemolytic anemia. *Hematology/Oncology Clinics of North America,* **29,** 455-471.

Berentsen, S., Randen, U., Vagan, A.M., Hjorth-Hansen, H., Vik, A., Dalgaard, J., Jacobsen, E.M., Thoresen, A.S., Beiske, K. & Tjonnfjord, G.E. (2010) High response rate and durable remissions following fludarabine and rituximab combination therapy for chronic cold agglutinin disease. *Blood,* **116,** 3180-3184.

Berentsen, S. & Tjonnfjord, G.E. (2012) Diagnosis and treatment of cold agglutinin mediated autoimmune hemolytic anemia. *Blood Reviews,* **26,** 107-115.

Berentsen, S., Ulvestad, E., Langholm, R., Beiske, K., Hjorth-Hansen, H., Ghanima, W., Sorbo, J.H. & Tjonnfjord, G.E. (2006) Primary chronic cold agglutinin disease: a population based clinical study of 86 patients. *Haematologica,* **91,** 460-466.

Bjornsson, H.T., Benjamin, J.S., Zhang, L., Weissman, J., Gerber, E.E., Chen, Y.C., Vaurio, R.G., Potter, M.C., Hansen, K.D. & Dietz, H.C. (2014) Histone deacetylase inhibition rescues structural and functional brain deficits in a mouse model of Kabuki syndrome. *Science Translational Medicine,* **6,** 256ra135.

Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E.S. & Getz, G. (2013) Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature Biotechnology,* **31,** 213-219.

Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X. & Ruden, D.M. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin),* **6,** 80-92.

Courtens, W., Rassart, A., Stene, J.J. & Vamos, E. (2000) Further evidence for autosomal dominant inheritance and ectodermal abnormalities in Kabuki syndrome. *American Journal of Medical Genetics,* **93,** 244-249.

de Tute, R., Rawstron, A., Evans, P. & Owen, R. Cold agglutinin disease is a phenotypically distinct clonal B-cell disorder. *Clinical Lymphoma, Myeloma and Leukemia,* **15,** e184.

Geer, L.Y., Domrachev, M., Lipman, D.J. & Bryant, S.H. (2002) CDART: protein homology by domain architecture. *Genome Research,* **12,** 1619-1623.

Imai, Y., Maru, Y. & Tanaka, J. (2016) Action mechanisms of histone deacetylase inhibitors in the treatment of hematological malignancies. *Cancer Science*.

Jeelall, Y.S., Wang, J.Q., Law, H.D., Domaschenz, H., Fung, H.K., Kallies, A., Nutt, S.L., Goodnow, C.C. & Horikawa, K. (2012) Human lymphoma mutations reveal CARD11 as the switch between self-antigen-induced B cell death or proliferation and autoantibody production. *Journal of Experimental Medicine,* **209,** 1907-1917.

Lenz, G., Davis, R.E., Ngo, V.N., Lam, L., George, T.C., Wright, G.W., Dave, S.S., Zhao, H., Xu, W., Rosenwald, A., Ott, G., Muller-Hermelink, H.K., Gascoyne, R.D., Connors, J.M., Rimsza, L.M.,

Campo, E., Jaffe, E.S., Delabie, J., Smeland, E.B., Fisher, R.I., Chan, W.C. & Staudt, L.M. (2008) Oncogenic CARD11 mutations in human diffuse large B cell lymphoma. *Science,* **319,** 1676-1679.

Li, H. & Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics,* **25,** 1754-1760.

Lindsley, A.W., Saal, H.M., Burrow, T.A., Hopkin, R.J., Shchelochkov, O., Khandelwal, P., Xie, C., Bleesing, J., Filipovich, L., Risma, K., Assa'ad, A.H., Roehrs, P.A. & Bernstein, J.A. (2016) Defects of B-cell terminal differentiation in patients with type-1 Kabuki syndrome. *Journal of Allergy and Clinical Immunology,* **137,** 179-187 e110.

Lupas, A., Van Dyke, M. & Stock, J. (1991) Predicting coiled coils from protein sequences. *Science,* **252,** 1162-1164.

Makrythanasis, P., van Bon, B.W., Steehouwer, M., Rodriguez-Santiago, B., Simpson, M., Dias, P., Anderlid, B.M., Arts, P., Bhat, M., Augello, B., Biamino, E., Bongers, E.M., Del Campo, M., Cordeiro, I., Cueto-Gonzalez, A.M., Cusco, I., Deshpande, C., Frysira, E., Izatt, L., Flores, R., Galan, E., Gener, B., Gilissen, C., Granneman, S.M., Hoyer, J., Yntema, H.G., Kets, C.M., Koolen, D.A., Marcelis, C., Medeira, A., Micale, L., Mohammed, S., de Munnik, S.A., Nordgren, A., Psoni, S., Reardon, W., Revencu, N., Roscioli, T., Ruiterkamp-Versteeg, M., Santos, H.G., Schoumans, J., Schuurs-Hoeijmakers, J.H., Silengo, M.C., Toledo, L., Vendrell, T., van der Burgt, I., van Lier, B., Zweier, C., Reymond, A., Trembath, R.C., Perez-Jurado, L., Dupont, J., de Vries, B.B., Brunner, H.G., Veltman, J.A., Merla, G., Antonarakis, S.E. & Hoischen, A. (2013) MLL2 mutation detection in 86 patients with Kabuki syndrome: a genotype-phenotype study. *Clinical Genetics,* **84,** 539-545.

Malecka, A., Troen, G., Tierens, A., Ostlie, I., Malecki, J., Randen, U., Berentsen, S., Tjonnfjord, G.E. & Delabie, J.M. (2016) Immunoglobulin heavy and light chain gene features are correlated with primary cold agglutinin disease onset and activity. *Haematologica,* **101,** e361-364.

Miyake, N., Koshimizu, E., Okamoto, N., Mizuno, S., Ogata, T., Nagai, T., Kosho, T., Ohashi, H., Kato, M., Sasaki, G., Mabe, H., Watanabe, Y., Yoshino, M., Matsuishi, T., Takanashi, J., Shotelersuk, V., Tekin, M., Ochi, N., Kubota, M., Ito, N., Ihara, K., Hara, T., Tonoki, H., Ohta, T., Saito, K., Matsuo, M., Urano, M., Enokizono, T., Sato, A., Tanaka, H., Ogawa, A., Fujita, T., Hiraki, Y., Kitanaka, S., Matsubara, Y., Makita, T., Taguri, M., Nakashima, M., Tsurusaki, Y., Saitsu, H., Yoshiura, K., Matsumoto, N. & Niikawa, N. (2013) MLL2 and KDM6A mutations in patients with Kabuki syndrome. *American Journal of Medical Genetics. Part A,* **161A,** 2234-2243.

Morin, R.D., Mendez-Lago, M., Mungall, A.J., Goya, R., Mungall, K.L., Corbett, R.D., Johnson, N.A., Severson, T.M., Chiu, R., Field, M., Jackman, S., Krzywinski, M., Scott, D.W., Trinh, D.L., Tamura-Wells, J., Li, S., Firme, M.R., Rogic, S., Griffith, M., Chan, S., Yakovenko, O., Meyer, I.M., Zhao, E.Y., Smailus, D., Moksa, M., Chittaranjan, S., Rimsza, L., Brooks-Wilson, A., Spinelli, J.J., Ben-Neriah, S., Meissner, B., Woolcock, B., Boyle, M., McDonald, H., Tam, A., Zhao, Y., Delaney, A., Zeng, T., Tse, K., Butterfield, Y., Birol, I., Holt, R., Schein, J., Horsman, D.E., Moore, R., Jones, S.J., Connors, J.M., Hirst, M., Gascoyne, R.D. & Marra, M.A. (2011) Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature,* **476,** 298-303.

Ng, S.B., Bigham, A.W., Buckingham, K.J., Hannibal, M.C., McMillin, M.J., Gildersleeve, H.I., Beck, A.E., Tabor, H.K., Cooper, G.M., Mefford, H.C., Lee, C., Turner, E.H., Smith, J.D., Rieder, M.J., Yoshiura, K., Matsumoto, N., Ohta, T., Niikawa, N., Nickerson, D.A., Bamshad, M.J. & Shendure, J. (2010) Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nature Genetics,* **42,** 790-793.

Okosun, J., Bodor, C., Wang, J., Araf, S., Yang, C.Y., Pan, C., Boller, S., Cittaro, D., Bozek, M., Iqbal, S., Matthews, J., Wrench, D., Marzec, J., Tawana, K., Popov, N., O'Riain, C., O'Shea, D., Carlotti, E., Davies, A., Lawrie, C.H., Matolcsy, A., Calaminici, M., Norton, A., Byers, R.J., Mein, C., Stupka, E., Lister, T.A., Lenz, G., Montoto, S., Gribben, J.G., Fan, Y., Grosschedl, R., Chelala, C. & Fitzgibbon, J. (2014) Integrated genomic analysis identifies recurrent mutations and

evolution patterns driving the initiation and progression of follicular lymphoma. *Nature Genetics,* **46,** 176-181.

Ortega-Molina, A., Boss, I.W., Canela, A., Pan, H., Jiang, Y., Zhao, C., Jiang, M., Hu, D., Agirre, X., Niesvizky, I., Lee, J.E., Chen, H.T., Ennishi, D., Scott, D.W., Mottok, A., Hother, C., Liu, S., Cao, X.J., Tam, W., Shaknovich, R., Garcia, B.A., Gascoyne, R.D., Ge, K., Shilatifard, A., Elemento, O., Nussenzweig, A., Melnick, A.M. & Wendel, H.G. (2015) The histone lysine methyltransferase KMT2D sustains a gene expression program that represses B cell lymphoma development. *Nature Medicine,* **21,** 1199-1208.

Parry, D.A., Fraser, R.D. & Squire, J.M. (2008) Fifty years of coiled-coils and alpha-helical bundles: a close relationship between sequence and structure. *Journal of Structural Biology,* **163,** 258-269.

Pasqualucci, L., Khiabanian, H., Fangazio, M., Vasishtha, M., Messina, M., Holmes, A.B., Ouillette, P., Trifonov, V., Rossi, D., Tabbo, F., Ponzoni, M., Chadburn, A., Murty, V.V., Bhagat, G., Gaidano, G., Inghirami, G., Malek, S.N., Rabadan, R. & Dalla-Favera, R. (2014) Genetics of follicular lymphoma transformation. *Cell Reports,* **6,** 130-140.

Pasqualucci, L., Trifonov, V., Fabbri, G., Ma, J., Rossi, D., Chiarenza, A., Wells, V.A., Grunn, A., Messina, M., Elliot, O., Chan, J., Bhagat, G., Chadburn, A., Gaidano, G., Mullighan, C.G., Rabadan, R. & Dalla-Favera, R. (2011) Analysis of the coding genome of diffuse large B-cell lymphoma. *Nature Genetics,* **43,** 830-837.

Potter, K.N., Hobby, P., Klijn, S., Stevenson, F.K. & Sutton, B.J. (2002) Evidence for involvement of a hydrophobic patch in framework region 1 of human V4-34-encoded Igs in recognition of the red blood cell I antigen. *Journal of Immunology,* **169,** 3777-3782.

Randen, U., Troen, G., Tierens, A., Steen, C., Warsame, A., Beiske, K., Tjonnfjord, G.E., Berentsen, S. & Delabie, J. (2014) Primary cold agglutinin-associated lymphoproliferative disease: a B-cell lymphoma of the bone marrow distinct from lymphoplasmacytic lymphoma. *Haematologica,* **99,** 497-504.

Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G. & Mesirov, J.P. (2011) Integrative genomics viewer. *Nature Biotechnology,* **29,** 24-26.

Rosenquist, R., Rosenwald, A., Du, M.Q., Gaidano, G., Groenen, P., Wotherspoon, A., Ghia, P., Gaulard, P., Campo, E., Stamatopoulos, K., European Research Initiative on, C.L.L. & the European Association for, H. (2016) Clinical impact of recurrently mutated genes on lymphoma diagnostics: state-of-the-art and beyond. *Haematologica,* **101,** 1002-1009.

Saunders, C.T., Wong, W.S., Swamy, S., Becq, J., Murray, L.J. & Cheetham, R.K. (2012) Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics,* **28,** 1811-1817.

Schmidt, J., Federmann, B., Schindler, N., Steinhilber, J., Bonzheim, I., Fend, F. & Quintanilla-Martinez, L. (2015) MYD88 L265P and CXCR4 mutations in lymphoplasmacytic lymphoma identify cases with high disease activity. *British Journal of Haematology,* **169,** 795-803.

Snow, A.L., Xiao, W., Stinson, J.R., Lu, W., Chaigne-Delalande, B., Zheng, L., Pittaluga, S., Matthews, H.F., Schmitz, R., Jhavar, S., Kuchen, S., Kardava, L., Wang, W., Lamborn, I.T., Jing, H., Raffeld, M., Moir, S., Fleisher, T.A., Staudt, L.M., Su, H.C. & Lenardo, M.J. (2012) Congenital B cell lymphocytosis explained by novel germline CARD11 mutations. *Journal of Experimental Medicine,* **209,** 2247-2261.

Spina, V., Khiabanian, H., Messina, M., Monti, S., Cascione, L., Bruscaggin, A., Spaccarotella, E., Holmes, A.B., Arcaini, L., Lucioni, M., Tabbo, F., Zairis, S., Diop, F., Cerri, M., Chiaretti, S., Marasca, R., Ponzoni, M., Deaglio, S., Ramponi, A., Tiacci, E., Pasqualucci, L., Paulli, M., Falini, B., Inghirami, G., Bertoni, F., Foa, R., Rabadan, R., Gaidano, G. & Rossi, D. (2016) The genetics of nodal marginal zone lymphoma. *Blood,* **128,** 1362-1373.

Treon, S.P., Xu, L., Yang, G., Zhou, Y., Liu, X., Cao, Y., Sheehy, P., Manning, R.J., Patterson, C.J., Tripsas, C., Arcaini, L., Pinkus, G.S., Rodig, S.J., Sohani, A.R., Harris, N.L., Laramie, J.M., Skifter, D.A., Lincoln, S.E. & Hunter, Z.R. (2012) MYD88 L265P somatic mutation in Waldenstrom's macroglobulinemia. *New England Journal of Medicine,* **367,** 826-833.

Ye, K., Schulz, M.H., Long, Q., Apweiler, R. & Ning, Z. (2009) Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics,* **25,** 2865-2871.

Young, R.M. & Staudt, L.M. (2012) A new "brew" of MALT1 inhibitors. *Cancer Cell,* **22,** 706-707.

Young, R.M., Wu, T., Schmitz, R., Dawood, M., Xiao, W., Phelan, J.D., Xu, W., Menard, L., Meffre, E., Chan, W.C., Jaffe, E.S., Gascoyne, R.D., Campo, E., Rosenwald, A., Ott, G., Delabie, J., Rimsza, L.M. & Staudt, L.M. (2015) Survival of human lymphoma cells requires B-cell receptor engagement by self-antigens. *Proceedings of the National Academy of Sciences of the United States of America,* **112,** 13447-13454.

Zhang, J., Dominguez-Sola, D., Hussein, S., Lee, J.E., Holmes, A.B., Bansal, M., Vlasevska, S., Mo, T., Tang, H., Basso, K., Ge, K., Dalla-Favera, R. & Pasqualucci, L. (2015) Disruption of KMT2D perturbs germinal center B cell development and promotes lymphomagenesis. *Nature Medicine,* **21,** 1190-1198.

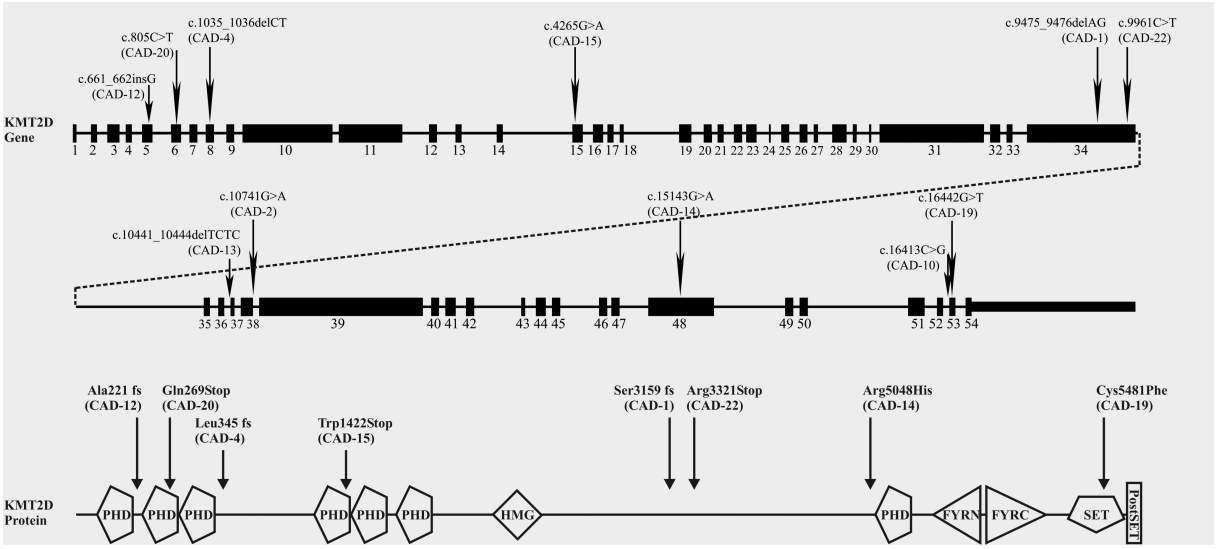


Figure 1. Schematic representation of the human *KMT2D* gene (ENST00000301067.11, upper panel) and protein (NP_003473.3, lower panel). The gene organization is adapted from the UCSC browser (hg19). Exons are represented by black rectangles. Protein domains are indicated according to the NCBI database. Vertical arrows show the approximate position of mutations. CAD patient numbers are indicated in parenthesis. ins: insertion; del: deletion; fs: frameshift; STOP: nonsense mutation; PHD: plant homeodomain finger; HMG: high mobility group box; FYRN: F/Y rich N-terminus; FYRC: F/Y rich C-terminus; SET: SET domain; PostSET: post-SET domain.

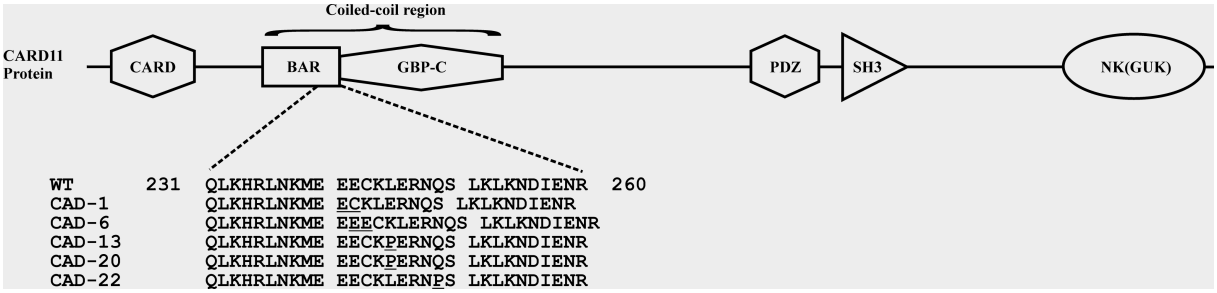


Figure 2. Schematic representation of the human CARD11 protein (NP_115791.3). Domains indicated according to the NCBI conserved domain architecture retrieval tool (CDART) (Geer*, et al* 2002). Mutations are localized in the C-terminal part of the BAR domain within the coiled-coil region, coded by exon 6 (ENST00000396946.8). The protein sequences (amino acids 231-260) of the wild-type (WT) and mutated CARD11 proteins are shown. Mutations are underlined. CARD: caspase activation and recruitment domain; BAR: Bin/Amphiphysin/Rvs domain; GBP-C: guanylate-binding protein C-terminal domain; PDZ: PDZ domain; SH3: Src Homology 3 domain; NK(GUK): nucleoside/nucleotide kinase (guanylate kinase) domain.

Table I. Mutations in *KMT2D* and *CARD11* in CAD patients.

Sample name	Gene	Effect[a]	Impact[a]	Amino acid change[b]	Mutation[c]	COSMIC ID[d]
CAD-1	*KMT2D*	Frameshift variant	HIGH	p.Ser3159fs	c.9475_9476delAG	NA
CAD-2	*KMT2D*	Splice donor variant	HIGH		c.10741G>A	NA
CAD-4	*KMT2D*	Frameshift variant	HIGH	p.Leu345fs	c.1035_1036delCT	NA
CAD-10	*KMT2D*	Splice region variant	LOW		c.16413C>G	NA
CAD-12	*KMT2D*	Frameshift variant	HIGH	p.Ala221_Ala222fs	c.661_662insG	NA
CAD-13	*KMT2D*	Splice region variant	LOW		c.10441_10444delTCTC	NA
CAD-14	*KMT2D*	Missense variant	MODERATE/MISSENSE	p.Arg5048His	c.15143G>A	COSM2006665
CAD-15	*KMT2D*	Stop gained	HIGH/NONSENSE	p.Trp1422*	c.4265G>A	COSM6030602
CAD-19	*KMT2D*	Missense variant	MODERATE/MISSENSE	p.Cys5481Phe	c.16442G>T	NA
CAD-20	*KMT2D*	Stop gained	HIGH/NONSENSE	p.Gln269*	c.805C>T	NA
CAD-22	*KMT2D*	Stop gained	HIGH/NONSENSE	p.Arg3321*	c.9961C>T	COSM221061
CAD-1	*CARD11*	Disruptive inframe deletion	MODERATE	p.Glu241_Glu242del	c.723_725delGGA	NA
CAD-6	*CARD11*	Disruptive inframe insertion	MODERATE	p.Glu242_Cys243insGluGlu	c.725_726insGGA	NA
CAD-13	*CARD11*	Missense variant	MODERATE/MISSENSE	p.Leu245Pro	c.734T>C	COSM1580650
CAD-20	*CARD11*	Missense variant	MODERATE/MISSENSE	p.Leu245Pro	c.734T>C	COSM1580650
CAD-22	*CARD11*	Missense variant	MODERATE/MISSENSE	p.Gln249Pro	c.746A>C	COSM133703

a. Annotations are according to the SnpEff program. b. Numbering of amino acids in proteins (p.) according to reference sequences: NP_003473.3 (KMT2D) and NP_115791.3 (CARD11). c. Numbering of nucleotides in the coding strand (c.) according to reference sequences: ENST00000301067.11 (*KMT2D*) and ENST00000396946.8 (*CARD11*). d. the Catalogue Of Somatic Mutations In Cancer (COSMIC) – Mutation Id. Localization of mutations is graphically shown in Fig. 2-3. fs: frameshift, ins: insertion; del: deletion; *: nonsense mutation, NA: not available.

SUPPLEMENTARY APPENDIX

Frequent somatic mutations of *KMT2D* (*MLL2*) and *CARD11*

Agnieszka Małecka[1,2], Gunhild Trøen[1], Anne Tierens[3], Ingunn Østlie[1], Jędrzej Małecki[4], Ulla Randen[1], Junbai Wang[1], Sigbjørn Berentsen[5], Geir E. Tjønnfjord[6] and Jan M.A. Delabie[3]

[1]Department of Pathology, Oslo University Hospital, Oslo, Norway; [2]Faculty of Medicine, University of Oslo, Oslo, Norway; [3]Laboratory Medicine Program, University Health Network and University of Toronto, Toronto, ON, Canada; [4]Department of Biological Sciences, University of Oslo, Oslo, Norway, [5]Department of Research and Innovation, Haugesund Hospital, Helse Fonna, Haugesund, Norway; [6]Department of Haematology, Oslo University Hospital and Institute of Clinical Medicine, University of Oslo, Oslo, Norway

Supplementary table I. Clinical characteristics

Sample	Hemo globin (g/dL)	Leukocytes/ lymphocytes (10^9/L)	Reticulo cytes (10^9/L)	Thrombo cytes (10^9/L)	CA titer (4°C)[a]	IgM (g/L)	LD (U/L)	Bilirubin (µmol/L)	Circulatory symptoms	age at diagnosis	Year of birth	Sex
CAD-1	8,4	7,6 / 3,3	160	171	1024	4,9	295	39	Yes, slight	75	1934	F
CAD-2	7,2	5,5 / 2,0	155	205	128	4,4	340	62	Yes, slight	64	1940	F
CAD-3	9,5	6,1/2,4	198	381	128	NA	250	32	No	80	1930	F
CAD-4	8,2	10,3 / 4,6	96	211	4096	6,9	264	48	Yes, slight	72	1934	M
CAD-5	9,9	7,0 / 4,1	139	195	2048	NA	248	29	Yes, slight	NA	1943	M
CAD-6	10	6,4 / 3,0	118	186	1024	NA	293	28	Yes, slight	56	1956	M
CAD-7	11,3	6,0/2,54	120	199	NA	22,6	222	49	Yes, prominent	67	1934	F
CAD-10	9,1	20,7/3,4	76	384	16384	8,8	293	54	No	71	1938	F
CAD-12	8,8	8,7 / 3,6	153	194	2048	5,1	287	38	Yes, slight	83	1929	M
CAD-13	8,2	6,7 / 2,9	97	200	2048	NA	307	43	Yes, slight	81	1931	M
CAD-14	13,9	4,9 / 2,3	107	241	512	7,4	275	24	Yes, prominent	69	1942	M
CAD-15	10,2	9,8/2,5	110	276	1024	2	253	10	yes	about 70	1939	F
CAD-18	8,6	4,9 / 3,0	131	197	8192	5,5	301	33	Yes, slight	84	1928	M
CAD-19	7	6,8 / 2,9	205	163	64000	7,9	346	61	Yes, prominent	68	1943	F
CAD-20	8,6	6,0 / 2,7	164	194	1024	4	310	51	Yes, slight	68	1938	F
CAD-22	9,1	11,2 / 4,0	256	271	256	8,4	313	54	No	76	1936	M

a: CA titer: some laboratories did not report results above titer 2048; NA: not available; ND: no data.

Supplementary table II. Flow cytometry analysis of bone marrow and blood samples

CAD patient	Samples[a]	Immunophenotype of the neoplastic B cell population
CAD-1	B, BM	CD45+, CD19+, CD20+, CD79b+, CD200+, CD5+, CD10-, CD23+, CD43-, CD38+, CD38+/-, IGK+, IGL-
CAD-2	B, BM	CD45+, CD19+, CD20+, CD79b+, CD200+, CD5+, CD10-, CD23-/+, CD43-, CD38 weak+, IGK+, IGL-
CAD-3	BM	CD45+, CD19+, CD20+, CD79b+, CD200+, CD5+, CD10-, CD23-, CD43-, IGK+, IGL-
CAD-4	BM	CD45+, CD19+, CD20+, CD79b+, CD200+, CD5+, CD10-, CD23-, CD38+, IGK+, IGL-
CAD-5	BM	CD45+, CD19+, CD20+, CD22+, CD10-, CD5+, IGK+, IGL-
CAD-6	BM	CD45+, CD19+, CD20+, CD79b+, CD200+, CD5+, CD10-, CD23-, CD43+, IGK+, IGL-
CAD-7	BM	CD45+, CD19+, CD20+, CD79b+, CD200+, CD5+/-, CD10-, CD23-, CD38-/+, CD43-, IGK+, IGL-
CAD-10	BM	CD45+, CD19+, CD20+, CD79b+, CD200+, CD5+, CD10-, CD23-, CD43-, IGK+, IGL-
CAD-12	B, BM	CD45+, CD19+, CD20-, CD79b+, CD200+, CD5-, CD10-, CD23-, CD43-, IGK+, IGL-
CAD-13	B, BM	CD45+, CD19+, CD20+, CD79b+, CD200-, CD5-, CD10-, CD23-, IGK+, IGL-
CAD-14	BM	CD45+, CD19+, CD20+, CD79b-, CD200+, CD5+, CD10-, CD23-, CD43-, IGK+, IGL-
CAD-15	BM	CD45+, CD19+, CD20+, CD79b+, CD200+, CD5-, CD10-, CD23-/+, CD43-, IGK+, IGL-
CAD-18	B, BM	CD45+, CD19+, CD20+, CD79b-, CD200+, CD5-, CD10-, CD23-, CD43-, IGK+, IGL-
CAD-19	BM	CD45+, CD19+, CD20+, CD79b+, CD200+, CD5-/+, CD10-, CD23-/+, CD43-, IGK+, IGL-
CAD-20	B, BM	CD45+, CD19+, CD20+, CD79b+, CD200-, CD5-, CD10-, CD23-, CD43-, IGK+, IGL-
CAD-22	B, BM	CD45+, CD19+, CD20+, CD79b+, CD200+, CD5+, CD10-, CD23-, CD43-, CD38-/+, IGK+, IGL-

a: B: blood; BM: bone marrow

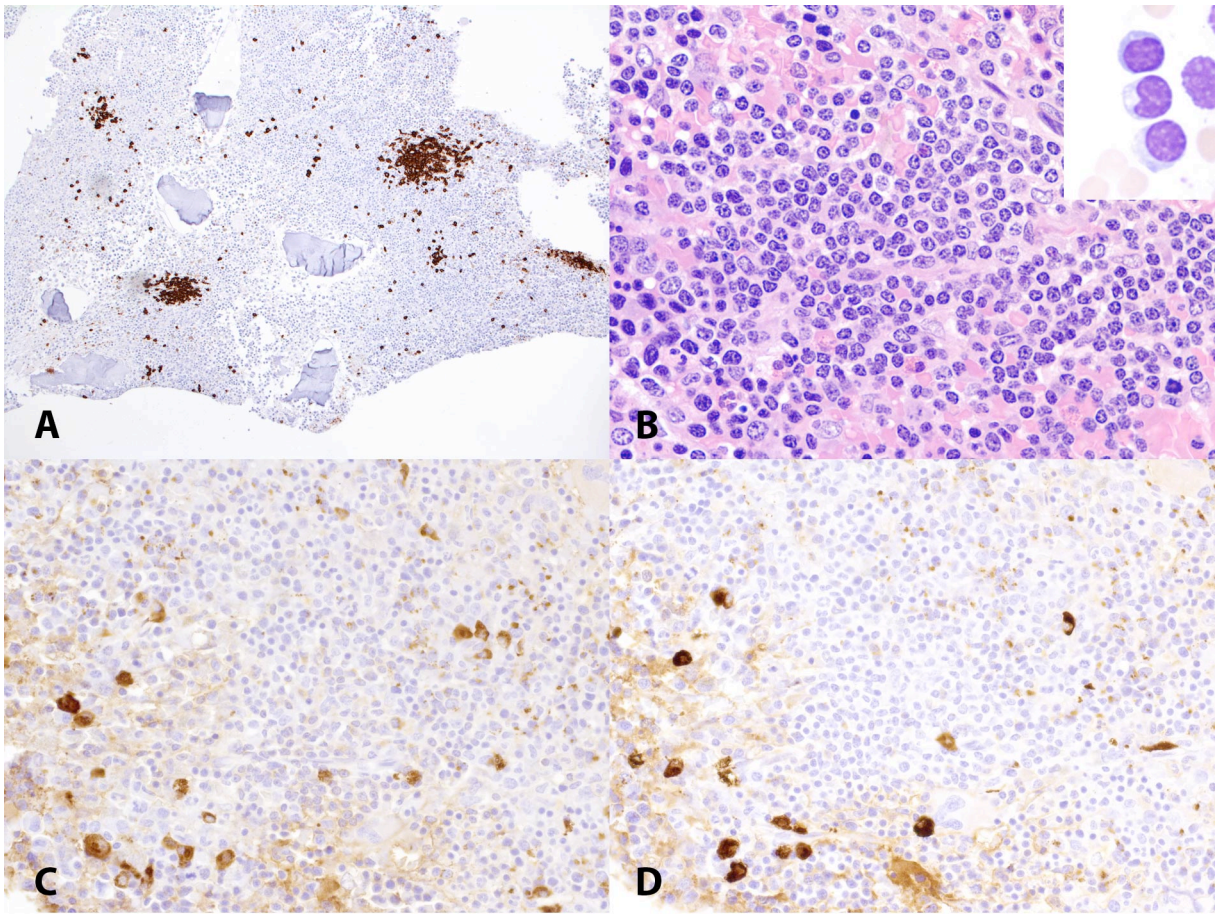
Supplementary table III. Non-synonymous mutations in CAD patients detected by exome sequencing.

Sample	Chr	Chr position (GRCh37)	Ref allele	Alt allele	Mutect quality score t_lod_fstar	Strelka quality score QSS_NT	Indels detected by Pindel	Annotation (a)	Impact (a)	Impact(a)	Gene name (a)
CAD-2	2	209358053	C	T	12,63051	QSS_NT=27		missense_variant	MODERATE	MISSENSE	**PTH2R**
CAD-2	3	33434869	G	A	72,729761	QSS_NT=73		missense_variant	MODERATE	MISSENSE	**UBP1**
CAD-2	6	56883333	A	G	70,990072	QSS_NT=43		missense_variant	MODERATE	MISSENSE	**BEND6**
CAD-2	7	100551799	A	G	53,315691	not detected		missense_variant	MODERATE	MISSENSE	**MUC3A**
CAD-2	11	30921189	G	A	96,992526	QSS_NT=87		missense_variant	MODERATE	MISSENSE	**DCDC1**
CAD-2	11	33054314	T	C	73,823243	QSS_NT=55		missense_variant	MODERATE	MISSENSE	**DEPDC7**
CAD-2	11	101762177	A	C	14,544456	QSS_NT=22		missense_variant	MODERATE	MISSENSE	**ANGPTL5**
CAD-2	12	49427849	C	T	25,465121	not detected		splice_donor_variant	HIGH		**KMT2D**
CAD-2	13	95095830	G	T	183,809821	QSS_NT=146		missense_variant	MODERATE	MISSENSE	**DCT**
CAD-2	17	37009959	G	C	31,935311	QSS_NT=46		missense_variant	MODERATE	MISSENSE	**RPL23**
CAD-2	18	31319730	G	A	169,240519	QSS_NT=175		missense_variant	MODERATE	MISSENSE	**ASXL3**
CAD-2	X	21609164	G	T	211,655086	QSS_NT=232		missense_variant	MODERATE	MISSENSE	**CNKSR2**
CAD-2	X	35966451	C	T	102,076562	QSS_NT=122		missense_variant	MODERATE	MISSENSE	**CXorf22**
CAD-5	1	11177061	C	T	41,314294	QSS_NT=63		missense_variant	MODERATE	MISSENSE	**MTOR**
CAD-5	1	16475403	A	T	18,507037	QSS_NT=34		missense_variant	MODERATE	MISSENSE	**EPHA2**
CAD-5	1	109793096	C	T	108,784161	QSS_NT=100		missense_variant	MODERATE	MISSENSE	**CELSR2**
CAD-5	2	186654807	A	T	16,331333	QSS_NT=58		missense_variant	MODERATE	MISSENSE	**FSIP2**
CAD-5	3	190039796	A	G	153,562483	QSS_NT=101		missense_variant	MODERATE	MISSENSE	**CLDN1**
CAD-5	4	157771515	TA	T	not analysed	QSI_NT=54	Pindel	frameshift_variant	HIGH		**PDGFC**
CAD-5	4	22425833	C	A	76,017811	QSS_NT=74		missense_variant	MODERATE	MISSENSE	**GPR125**
CAD-5	4	47427892	C	T	159,468787	QSS_NT=128		missense_variant	MODERATE	MISSENSE	**GABRB1**
CAD-5	4	69535903	T	G	158,648262	QSS_NT=148		missense_variant	MODERATE	MISSENSE	**UGT2B15**
CAD-5	4	114824781	G	T	206,243051	QSS_NT=164		missense_variant	MODERATE	MISSENSE	**ARSJ**
CAD-5	5	139743047	T	A	10,403573	QSS_NT=23		missense_variant	MODERATE	MISSENSE	**SLC4A9**
CAD-5	6	40999801	C	T	15,665352	QSS_NT=36		missense_variant	MODERATE	MISSENSE	**UNC5CL**
CAD-5	12	70963524	G	T	9,457459	QSS_NT=16		missense_variant	MODERATE	MISSENSE	**PTPRB**
CAD-5	13	21721477	A	T	84,331275	QSS_NT=105		missense_variant	MODERATE	MISSENSE	**SAP18**
CAD-5	14	80669368	A	T	88,363322	QSS_NT=107		missense_variant	MODERATE	MISSENSE	**DIO2**
CAD-5	16	5077892	C	T	40,08399	QSS_NT=39		missense_variant	MODERATE	MISSENSE	**NAGPA**
CAD-5	16	31476143	T	G	9,52186	QSS_NT=16		missense_variant	MODERATE	MISSENSE	**ARMC5**
CAD-5	17	7832644	C	G	18,208955	QSS_NT=24		missense_variant	MODERATE	MISSENSE	**KCNAB3**
CAD-5	20	60791447	C	T	32,468538	QSS_NT=74		missense_variant	MODERATE	MISSENSE	**HRH3**
CAD-5	22	29877002	G	A	20,241988	not detected		missense_variant	MODERATE	MISSENSE	**NEFH**
CAD-7	1	7723803	T	G	19,31326	QSS_NT=29		missense_variant	MODERATE	MISSENSE	**CAMTA1**
CAD-7	1	12336739	A	T	12,532276	QSS_NT=27		missense_variant	MODERATE	MISSENSE	**VPS13D**
CAD-7	1	209803131	G	C	11,098536	QSS_NT=28		missense_variant	MODERATE	MISSENSE	**LAMB3**
CAD-7	2	39893305	A	G	11,182137	QSS_NT=24		missense_variant	MODERATE	MISSENSE	**TMEM178A**
CAD-7	3	49145852	A	C	6,533404	QSS_NT=18		missense_variant	MODERATE	MISSENSE	**USP19**
CAD-7	4	62758392	C	A	31,082173	QSS_NT=47		missense_variant	MODERATE	MISSENSE	**LPHN3**
CAD-7	4	85600332	T	G	11,297767	QSS_NT=28		missense_variant	MODERATE	MISSENSE	**WDFY3**
CAD-7	4	90857068	A	G	44,339525	QSS_NT=70		missense_variant	MODERATE	MISSENSE	**MMRN1**
CAD-7	5	78985848	G	A	12,267337	QSS_NT=19		missense_variant	MODERATE	MISSENSE	**CMYA5**
CAD-7	5	128956447	C	A	21,310713	QSS_NT=30		missense_variant	MODERATE	MISSENSE	**ADAMTS19**
CAD-7	6	12749755	A	C	8,082223	QSS_NT=30		missense_variant	MODERATE	MISSENSE	**PHACTR1**
CAD-7	7	5266930	T	C	6,849513	QSS_NT=16		missense_variant	MODERATE	MISSENSE	**WIPI2**
CAD-7	7	70229797	A	C	11,598763	QSS_NT=30		missense_variant	MODERATE	MISSENSE	**AUTS2**
CAD-7	7	70229862	A	C	11,331324	QSS_NT=30		missense_variant	MODERATE	MISSENSE	**AUTS2**
CAD-7	8	113246694	C	A	70,156638	QSS_NT=124		missense_variant	MODERATE	MISSENSE	**CSMD3**
CAD-7	9	116930710	T	G	10,356544	QSS_NT=18		missense_variant	MODERATE	MISSENSE	**COL27A1**
CAD-7	10	11505707	G	C	12,454193	QSS_NT=32		missense_variant	MODERATE	MISSENSE	**USP6NL**
CAD-7	11	66411332	A	C	15,143618	QSS_NT=15		missense_variant	MODERATE	MISSENSE	**RBM14-RBM4**
CAD-7	11	85375043	T	G	14,112163	QSS_NT=16		missense_variant	MODERATE	MISSENSE	**CREBZF**
CAD-7	11	117340691	C	T	40,197289	QSS_NT=40		missense_variant	MODERATE	MISSENSE	**DSCAML1**
CAD-7	11	128934794	C	T	40,779947	QSS_NT=75		missense_variant	MODERATE	MISSENSE	**ARHGAP32**
CAD-7	11	1018559	A	C	20,035152	not detected		missense_variant	MODERATE	MISSENSE	**MUC6**
CAD-7	12	248369	T	G	8,144599	QSS_NT=23		missense_variant	MODERATE	MISSENSE	**IQSEC3**
CAD-7	12	7045848	A	C	13,065088	QSS_NT=38		missense_variant	MODERATE	MISSENSE	**ATN1**
CAD-7	12	83251115	T	G	24,463482	QSS_NT=31		missense_variant	MODERATE	MISSENSE	**TMTC2**
CAD-7	12	85517943	G	A	29,082721	QSS_NT=55		missense_variant	MODERATE	MISSENSE	**LRRIQ1**
CAD-7	12	121437352	A	C	12,770379	QSS_NT=35		missense_variant	MODERATE	MISSENSE	**HNF1A**
CAD-7	12	122492816	C	A	42,566442	QSS_NT=46		missense_variant	MODERATE	MISSENSE	**BCL7A**
CAD-7	14	53619314	A	C	15,895701	QSS_NT=22		missense_variant	MODERATE	MISSENSE	**DDHD1**
CAD-7	14	91975963	A	C	12,522442	QSS_NT=15		missense_variant	MODERATE	MISSENSE	**SMEK1**
CAD-7	14	105353974	A	G	10,668532	QSS_NT=23		missense_variant	MODERATE	MISSENSE	**CEP170B**
CAD-7	14	37050561	A	C	31,300026	not detected		missense_variant	MODERATE	MISSENSE	**NKX2-8**
CAD-7	15	93555608	A	C	15,286906	QSS_NT=31		missense_variant	MODERATE	MISSENSE	**CHD2**
CAD-7	16	2816112	G	A	11,410676	QSS_NT=22		stop_gained	HIGH	NONSENSE	**SRRM2**
CAD-7	16	86545091	A	C	8,970869	QSS_NT=17		missense_variant	MODERATE	MISSENSE	**FOXF1**

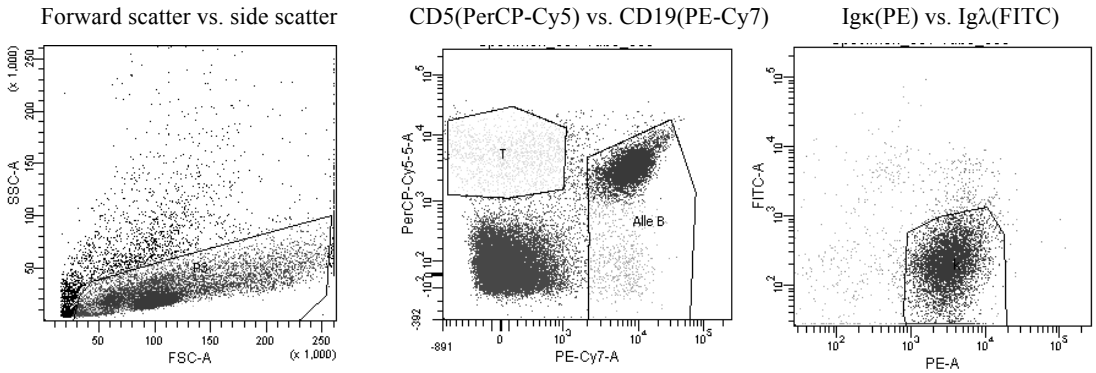
CAD-7	17	4837166	A	C	6,943214	QSS_NT=15		missense_variant	MODERATE	MISSENSE	**GP1BA**
CAD-7	17	62291327	T	G	17,442185	QSS_NT=17		missense_variant	MODERATE	MISSENSE	**TEX2**
CAD-7	19	42753892	T	C	10,32186	QSS_NT=18		splice_acceptor_variant	HIGH		**ERF**
CAD-7	19	33792461	A	C	12,037664	QSS_NT=16		missense_variant	MODERATE	MISSENSE	**CEBPA**
CAD-7	19	35556823	C	T	22,876063	QSS_NT=58		missense_variant	MODERATE	MISSENSE	**HPN**
CAD-7	19	39218616	T	G	28,461261	not detected		missense_variant	MODERATE	MISSENSE	**ACTN4**
CAD-7	19	19408000	T	G	20,502516	not detected		missense_variant	MODERATE	MISSENSE	**SUGP1**
CAD-7	21	43166260	A	G	8,40535	QSS_NT=19		missense_variant	MODERATE	MISSENSE	**RIPK4**
CAD-7	22	39381985	A	C	32,026145	not detected		missense_variant	MODERATE	MISSENSE	**APOBEC3B**
CAD-7	22	36894159	T	G	20,263608	not detected		missense_variant	MODERATE	MISSENSE	**FOXRED2**
CAD-7	GL000205.1	117481	G	A	9,731223	QSS_NT=17		missense_variant	MODERATE	MISSENSE	**AC011841.1**
CAD-7	X	101396080	T	C	13,799014	QSS_NT=25		missense_variant	MODERATE	MISSENSE	**TCEAL6**
CAD-7	X	105280854	T	C	36,369715	QSS_NT=49		missense_variant	MODERATE	MISSENSE	**SERPINA7**
CAD-7	X	132161884	G	T	13,075079	QSS_NT=65		missense_variant	MODERATE	MISSENSE	**USP26**
CAD-19	1	226924565	C	A	114,88076	QSS_NT=97		stop_gained	HIGH	NONSENSE	**ITPKB**
CAD-19	1	181706792	T	G	25,614133	not detected		splice_donor_variant	HIGH		**CACNA1E**
CAD-19	1	23769083	C	A	72,099952	QSS_NT=80		missense_variant	MODERATE	MISSENSE	**ASAP3**
CAD-19	1	40928811	T	A	14,191706	QSS_NT=44		missense_variant	MODERATE	MISSENSE	**ZFP69B**
CAD-19	1	240370628	C	A	96,389872	QSS_NT=134		missense_variant	MODERATE	MISSENSE	**FMN2**
CAD-19	1	39322724	T	A	not detected	QSS_NT=32		missense_variant	MODERATE	MISSENSE	**RRAGC**
CAD-19	1	39322725	C	G	not detected	QSS_NT=32		missense_variant	MODERATE	MISSENSE	**RRAGC**
CAD-19	1	39322726	T	A	not detected	QSS_NT=29		missense_variant	MODERATE	MISSENSE	**RRAGC**
CAD-19	2	20194131	GC	G	not analysed	QSI_NT=109	Pindel	frameshift_variant	HIGH		**MATN3**
CAD-19	2	136872550	T	TG	not analysed	QSI_NT=64	Pindel	frameshift_variant	HIGH		**CXCR4**
CAD-19	2	201470292	C	A	29,750013	QSS_NT=56		missense_variant	MODERATE	MISSENSE	**AOX1**
CAD-19	3	121413569	T	G	130,904682	QSS_NT=117		missense_variant	MODERATE	MISSENSE	**GOLGB1**
CAD-19	3	126261141	G	A	16,349848	QSS_NT=20		missense_variant	MODERATE	MISSENSE	**CHST13**
CAD-19	4	185559554	C	T	24,907904	not detected		splice_donor_variant	HIGH		**CASP3**
CAD-19	4	42067366	G	T	39,23867	QSS_NT=39		missense_variant	MODERATE	MISSENSE	**SLC30A9**
CAD-19	4	126412847	C	G	38,23074	QSS_NT=56		missense_variant	MODERATE	MISSENSE	**FAT4**
CAD-19	5	9043059	T	A	23,238354	QSS_NT=22		missense_variant	MODERATE	MISSENSE	**SEMA5A**
CAD-19	5	140573299	C	A	162,025206	QSS_NT=121		missense_variant	MODERATE	MISSENSE	**PCDHB10**
CAD-19	5	178634657	C	T	90,702575	QSS_NT=92		missense_variant	MODERATE	MISSENSE	**ADAMTS2**
CAD-19	6	47649116	G	T	88,950251	QSS_NT=95		missense_variant	MODERATE	MISSENSE	**GPR111**
CAD-19	7	41729763	G	A	70,251035	QSS_NT=102		missense_variant	MODERATE	MISSENSE	**INHBA**
CAD-19	8	119634213	CGGCTCGG	C	not analysed	QSI_NT=109	Pindel	frameshift_variant	HIGH		**SAMD12**
CAD-19	8	144621189	C	G	165,98318	QSS_NT=56		missense_variant	MODERATE	MISSENSE	**ZC3H3**
CAD-19	9	139848658	C	A	10,735878	QSS_NT=22		missense_variant	MODERATE	MISSENSE	**LCN12**
CAD-19	11	55563658	G	C	194,836038	QSS_NT=145		missense_variant	MODERATE	MISSENSE	**OR5D14**
CAD-19	11	57982515	C	A	48,784006	QSS_NT=86		missense_variant	MODERATE	MISSENSE	**OR1S1**
CAD-19	12	49415905	C	A	110,414153	QSS_NT=129		missense_variant	MODERATE	MISSENSE	**KMT2D**
CAD-19	13	52516566	G	A	11,394642	QSS_NT=22		missense_variant	MODERATE	MISSENSE	**ATP7B**
CAD-19	14	26917592	G	T	40,493139	QSS_NT=60		missense_variant	MODERATE	MISSENSE	**NOVA1**
CAD-19	16	4432263	G	A	104,721895	QSS_NT=102		missense_variant	MODERATE	MISSENSE	**VASN**
CAD-19	17	5984160	C	T	52,271673	QSS_NT=40		missense_variant	MODERATE	MISSENSE	**WSCD1**
CAD-19	18	28586988	A	T	108,810421	QSS_NT=86		stop_gained	HIGH	NONSENSE	**DSC3**
CAD-19	19	38024338	G	A	37,589309	QSS_NT=58		missense_variant	MODERATE	MISSENSE	**ZNF793**
CAD-19	X	148690496	A	G	46,556141	QSS_NT=36		missense_variant	MODERATE	MISSENSE	**TMEM185A**
CAD-20	1	228345706	C	T	52,299437	QSS_NT=71		stop_gained	HIGH	NONSENSE	**GJC2**
CAD-20	1	236399092	G	A	36,702134	QSS_NT=49		stop_gained	HIGH	NONSENSE	**ERO1LB**
CAD-20	1	20879617	C	T	31,870577	QSS_NT=46		missense_variant	MODERATE	MISSENSE	**FAM43B**
CAD-20	1	38171157	G	A	46,911992	QSS_NT=66		missense_variant	MODERATE	MISSENSE	**CDCA8**
CAD-20	1	75005997	G	A	40,222042	QSS_NT=68		missense_variant	MODERATE	MISSENSE	**FPGT-TNNI3K**
CAD-20	2	220343909	G	A	28,575592	not detected		missense_variant	MODERATE	MISSENSE	**SPEG**
CAD-20	3	130110149	G	C	75,593128	QSS_NT=108		missense_variant	MODERATE	MISSENSE	**COL6A5**
CAD-20	3	138176461	C	T	64,147161	QSS_NT=86		missense_variant	MODERATE	MISSENSE	**ESYT3**
CAD-20	3	176767804	A	T	76,659224	QSS_NT=74		missense_variant	MODERATE	MISSENSE	**TBL1XR1**
CAD-20	4	10445228	G	A	102,176031	QSS_NT=123		missense_variant	MODERATE	MISSENSE	**ZNF518B**
CAD-20	5	13839571	C	T	40,986992	QSS_NT=55		missense_variant	MODERATE	MISSENSE	**DNAH5**
CAD-20	5	35873601	C	A	65,243499	QSS_NT=66		missense_variant	MODERATE	MISSENSE	**IL7R**
CAD-20	5	160721423	C	T	18,665325	QSS_NT=28		missense_variant	MODERATE	MISSENSE	**GABRB2**
CAD-20	6	394954	A	T	30,419449	QSS_NT=51		missense_variant	MODERATE	MISSENSE	**IRF4**
CAD-20	6	4892232	G	A	19,442444	QSS_NT=58		missense_variant	MODERATE	MISSENSE	**CDYL**
CAD-20	6	17130867	C	A	13,665206	QSS_NT=25		missense_variant	MODERATE	MISSENSE	**STMND1**
CAD-20	6	27277395	G	A	55,665293	QSS_NT=88		missense_variant	MODERATE	MISSENSE	**POM121L2**
CAD-20	6	27861559	C	G	35,100504	QSS_NT=39		missense_variant	MODERATE	MISSENSE	**HIST1H2BO**
CAD-20	6	35392496	G	A	37,43368	QSS_NT=26		missense_variant	MODERATE	MISSENSE	**PPARD**
CAD-20	6	39835443	G	T	51,268279	QSS_NT=73		missense_variant	MODERATE	MISSENSE	**DAAM2**
CAD-20	6	125284151	G	T	121,949523	QSS_NT=97		missense_variant	MODERATE	MISSENSE	**RNF217**
CAD-20	6	157525017	C	T	81,636535	QSS_NT=78		missense_variant	MODERATE	MISSENSE	**ARID1B**
CAD-20	7	127222086	C	T	79,776183	QSS_NT=100		stop_gained	HIGH	NONSENSE	**GCC1**
CAD-20	7	2979513	A	G	121,552369	QSS_NT=130		missense_variant	MODERATE	MISSENSE	**CARD11**
CAD-20	7	44796131	A	G	29,357601	QSS_NT=21		missense_variant	MODERATE	MISSENSE	**ZMIZ2**

CAD-20	7	47408433	C	A	14,102582	QSS_NT=27		missense_variant	MODERATE	MISSENSE	**TNS3**
CAD-20	7	121653362	C	T	152,239856	QSS_NT=112		missense_variant	MODERATE	MISSENSE	**PTPRZ1**
CAD-20	7	149422992	C	G	25,846339	not detected		missense_variant	MODERATE	MISSENSE	**KRBA1**
CAD-20	8	56864534	CTCTG	C	not analysed	QSI_NT=134	Pindel	frameshift_variant	HIGH		**LYN**
CAD-20	8	59411027	G	T	37,491413	QSS_NT=65		missense_variant	MODERATE	MISSENSE	**CYP7A1**
CAD-20	8	110520455	T	G	54,615008	QSS_NT=79		missense_variant	MODERATE	MISSENSE	**PKHD1L1**
CAD-20	9	130742356	G	C	52,802234	QSS_NT=69		missense_variant	MODERATE	MISSENSE	**FAM102A**
CAD-20	9	33799026	C	G	not detected	QSS_NT=25		missense_variant	MODERATE	MISSENSE	**PRSS3**
CAD-20	11	70052261	CAT	C	not analysed	QSI_NT=60	Pindel	frameshift_variant	HIGH		**FADD**
CAD-20	11	73801954	C	T	48,951166	QSS_NT=93		missense_variant	MODERATE	MISSENSE	**C2CD3**
CAD-20	11	118499019	C	T	115,862987	QSS_NT=95		missense_variant	MODERATE	MISSENSE	**PHLDB1**
CAD-20	12	49447293	G	A	126,647048	QSS_NT=92		stop_gained	HIGH	NONSENSE	**KMT2D**
CAD-20	12	7060840	G	A	33,133168	QSS_NT=47		missense_variant	MODERATE	MISSENSE	**PTPN6**
CAD-20	12	21448582	A	C	13,853342	QSS_NT=25		missense_variant	MODERATE	MISSENSE	**SLCO1A2**
CAD-20	12	122265657	T	G	80,438358	QSS_NT=98		missense_variant	MODERATE	MISSENSE	**SETD1B**
CAD-20	12	1748963	C	T	23,626812	not detected		missense_variant	MODERATE	MISSENSE	**WNT5B**
CAD-20	13	39454844	A	G	7,311015	QSS_NT=24		missense_variant	MODERATE	MISSENSE	**FREM2**
CAD-20	14	21992885	C	T	39,170248	QSS_NT=71		missense_variant	MODERATE	MISSENSE	**SALL2**
CAD-20	14	75142994	C	T	51,14386	QSS_NT=55		missense_variant	MODERATE	MISSENSE	**AREL1**
CAD-20	14	86089537	G	A	76,161526	QSS_NT=72		missense_variant	MODERATE	MISSENSE	**FLRT2**
CAD-20	15	72874494	C	T	25,583236	QSS_NT=24		stop_gained	HIGH	NONSENSE	**ARIH1**
CAD-20	16	3788646	A	C	34,692234	QSS_NT=21		missense_variant	MODERATE	MISSENSE	**CREBBP**
CAD-20	16	16142146	T	C	194,192327	QSS_NT=108		missense_variant	MODERATE	MISSENSE	**ABCC1**
CAD-20	16	68010603	C	T	52,561013	QSS_NT=52		missense_variant	MODERATE	MISSENSE	**DPEP3**
CAD-20	16	72991548	T	C	152,677082	QSS_NT=124		missense_variant	MODERATE	MISSENSE	**ZFHX3**
CAD-20	17	61877828	G	T	78,76772	QSS_NT=58		stop_gained	HIGH	NONSENSE	**DDX42**
CAD-20	17	8638733	C	T	24,347789	QSS_NT=35		missense_variant	MODERATE	MISSENSE	**CCDC42**
CAD-20	19	4511829	C	T	102,865403	QSS_NT=83		missense_variant	MODERATE	MISSENSE	**PLIN4**
CAD-20	19	36884584	G	T	86,194672	QSS_NT=104		missense_variant	MODERATE	MISSENSE	**ZFP82**
CAD-20	19	36113852	G	A	21,696697	not detected		missense_variant	MODERATE	MISSENSE	**HAUS5**
CAD-20	X	48935736	G	A	17,086652	QSS_NT=16		stop_gained	HIGH	NONSENSE	**WDR45**
CAD-20	X	50378494	G	C	52,912663	QSS_NT=92		missense_variant	MODERATE	MISSENSE	**SHROOM4**
CAD-20	X	106016240	A	C	62,869598	QSS_NT=77		missense_variant	MODERATE	MISSENSE	**RNF128**
CAD-22	1	117617846	A	G	68,978912	QSS_NT=97		missense_variant	MODERATE	MISSENSE	**TTF2**
CAD-22	2	29404559	G	C	50,926303	QSS_NT=57		missense_variant	MODERATE	MISSENSE	**CLIP4**
CAD-22	2	54570988	G	T	64,024451	QSS_NT=84		missense_variant	MODERATE	MISSENSE	**C2orf73**
CAD-22	2	64113553	T	C	13,11984	QSS_NT=23		missense_variant	MODERATE	MISSENSE	**UGP2**
CAD-22	3	111766669	A	G	119,899463	QSS_NT=110		missense_variant	MODERATE	MISSENSE	**TMPRSS7**
CAD-22	5	96124250	C	T	32,358925	QSS_NT=42		missense_variant	MODERATE	MISSENSE	**ERAP1**
CAD-22	6	26156926	G	A	35,412599	QSS_NT=53		missense_variant	MODERATE	MISSENSE	**HIST1H1E**
CAD-22	7	2979501	T	G	122,453474	QSS_NT=98		missense_variant	MODERATE	MISSENSE	**CARD11**
CAD-22	7	117232563	A	T	93,290635	QSS_NT=91		missense_variant	MODERATE	MISSENSE	**CFTR**
CAD-22	8	38271461	C	T	57,040245	QSS_NT=59		missense_variant	MODERATE	MISSENSE	**FGFR1**
CAD-22	9	78796450	A	C	94,530691	QSS_NT=72		missense_variant	MODERATE	MISSENSE	**PCSK5**
CAD-22	10	99148260	A	G	9,801711	QSS_NT=37		missense_variant	MODERATE	MISSENSE	**RRP12**
CAD-22	11	3239235	G	A	12,799793	QSS_NT=17		missense_variant	MODERATE	MISSENSE	**MRGPRG**
CAD-22	12	49431178	G	A	31,566738	QSS_NT=21		stop_gained	HIGH	NONSENSE	**KMT2D**
CAD-22	14	88450836	C	T	27,755409	not detected		missense_variant	MODERATE	MISSENSE	**GALC**
CAD-22	15	81591803	C	A	26,681518	QSS_NT=56		missense_variant	MODERATE	MISSENSE	**IL16**
CAD-22	16	48149423	G	A	11,263155	QSS_NT=20		missense_variant	MODERATE	MISSENSE	**ABCC12**
CAD-22	17	3634434	G	A	38,127981	QSS_NT=36		missense_variant	MODERATE	MISSENSE	**ITGAE**
CAD-22	19	44605379	G	A	114,45607	QSS_NT=139		splice_donor_variant	HIGH		**ZNF224**

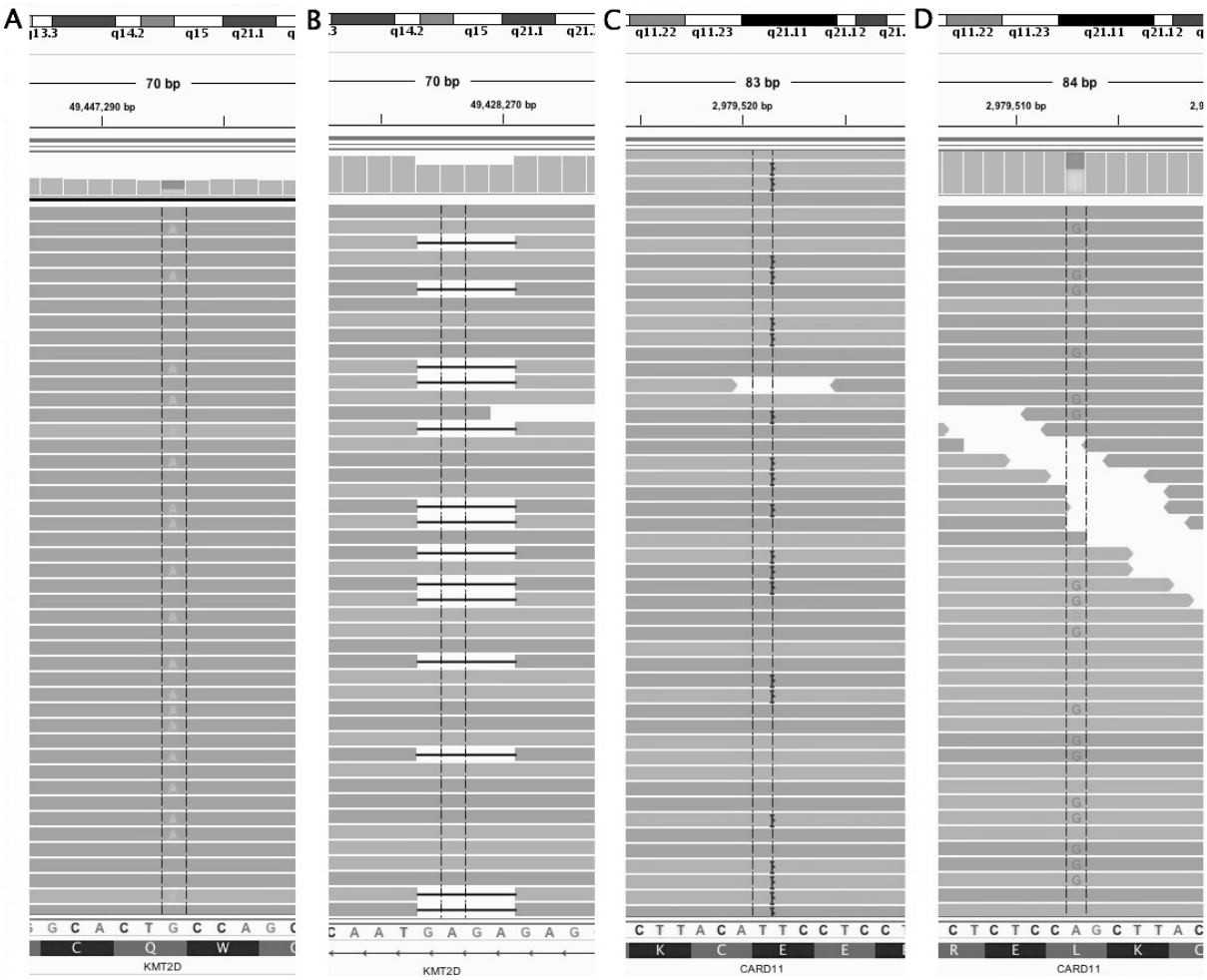
Genes mutated in at least 2 samples (marked as green) were taken for further analysis. (a). according to the SnpEff program; Chr - chromosome; Chr position (GRCh37) – chromosome positions according to human reference genome Genome Reference Consortium GRCh37 released in Feb 2009; Ref- reference; Alt – alternative; t_lod_fstar - CORE STATISTIC: Log of (likelihood tumor event is real / likelihood event is sequencing error; QSS_NT - Quality score reflecting the joint probability of a somatic variant and NT (QSS - Quality score for any somatic snv, ie. for the ALT allele to be present at a significantly different frequency in the tumor and normal; NT - Genotype of the normal in all data tiers, as used to classify somatic variants).



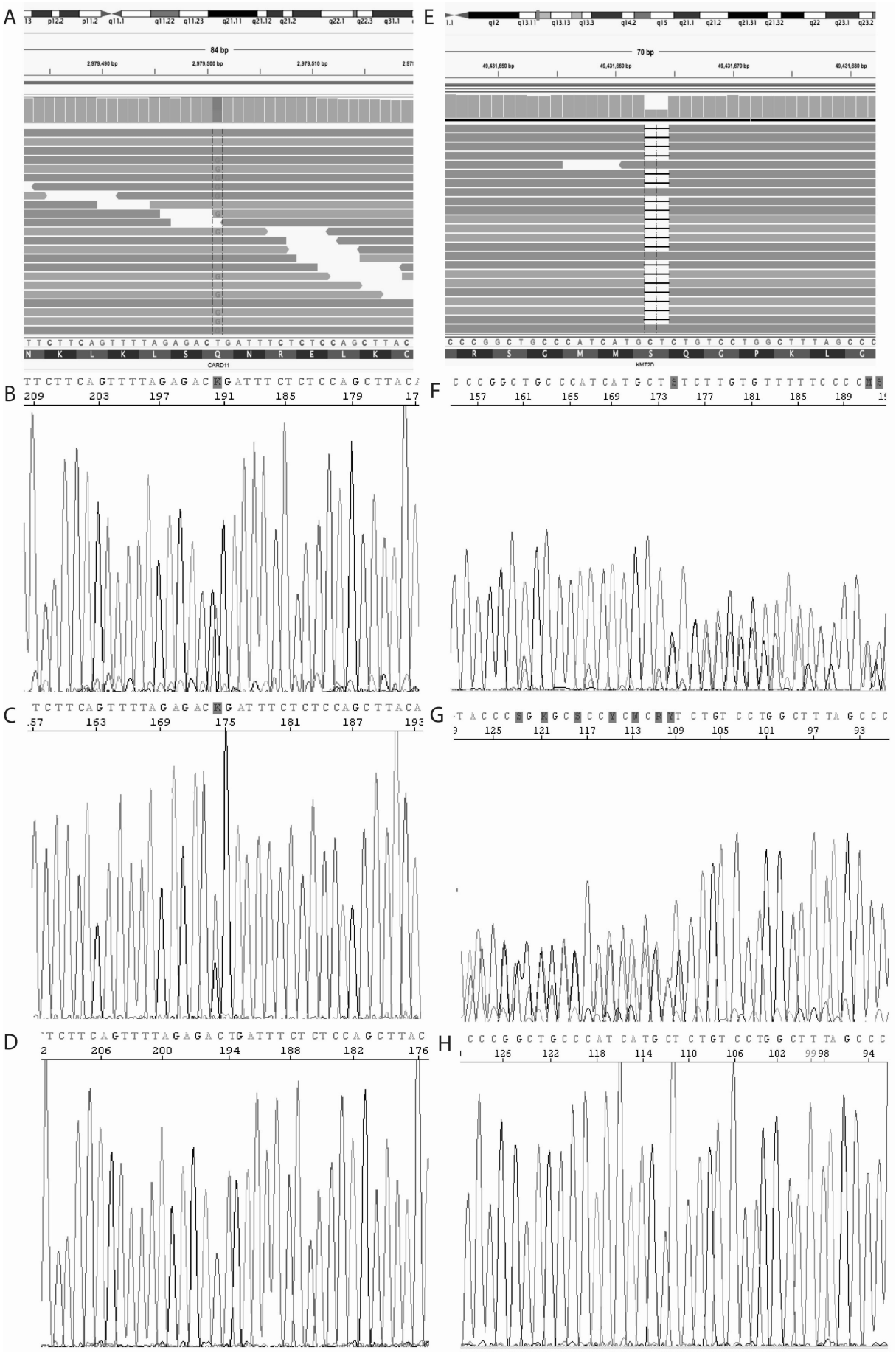
Supplementary figure 1. The figure illustrates the typical pathology of CAD-associated B-cell lymphoproliferative disease. A. Multiple small intraparenchymatous B-cell infiltrates are seen in the bone marrow trephine (anti-CD20 immunostaining, 100X). B. The lymphoid cells are small, have round nuclei with fine chromatin without prominent nucleoli and display some cytoplasm (H&E section, 400X, inset: bone marrow smear, 630X). C and D. Few plasma cells are identified surrounding the lymphoid infiltrates. The plasma cells may not show immunoglobulin light chain restriction, as illustrated here, but may show restriction in some cases. The small lymphoid cells do not show intracytoplasmic immunoglobulin expression, consistent with the lack of plasmacytoid cell differentiation (anti-IGK and anti-IGL, respectively, 400X).



Supplementary figure 2: Sorting strategy for isolation of monoclonal B cells from bone marrow by flow cytometry. Step one: selection of lymphocytes by forward scatter vs side scatter; step two: separation of B cells and T cells by CD5 vs CD19 gating; step 3: selection of monoclonal B cells using the immunoglobulin light chain gate κ+.



Supplementary figure 3. Examples of mutations detected by exome or targeted sequencing of clonal B cells. A) Nonsense mutation in *KMT2D* from CAD-20 detected by exome sequencing. B) 4 bp deletion in *KMT2D* from CAD-13 detected by targeted sequencing. C) 3 bp in-frame insertion in *CARD11* from CAD-6 detected by targeted sequencing. D) Missense mutation in *CARD11* from CAD-20 detected by exome sequencing. Sequences are displayed in IGV browser, with aligned reads colored by strand (red/blue). Point mutations and indels in A-D are detected in approximately 40-60% of reads, and are present on both strands.


A
84 bp
CARD11
B
C
D
E
70 bp
F
G
H

Supplementary figure 4. Examples of NGS analysis and Sanger sequencing verification of mutations in *CARD11* and *KMT2D* genes. A-C) missense mutation of *CARD11* in clonal B cells from CAD-22, as determined by exome sequencing (A) and Sanger sequencing using forward (B) and reverse (C) primers. The visible double peak in Sanger sequencing matches the missense mutation detected by NGS. D) Sanger sequencing of the same region of *CARD11* in T cells (control) from CAD-22 detects presence of only germline sequence, without mutation (no double peak). Sequences A-D are aligned to simplify the comparison. E-G) 2 bp deletion within *KMT2D* in clonal B cells from CAD-1, as determined by targeted NGS (E) and Sanger sequencing using forward (F) and reverse (G) primers. Sanger sequencing results show presence of single peaks up to the site of deletion, with double peaks starting from the deletion site. The site of deletion, detected by Sanger sequencing, matches the position of deletion determined by NGS. H) Sanger sequencing of the same region of *KMT2D* in T cells (control) from CAD-1 detects the presence of only germline sequence, without mutations (no double peaks). Sequences E-H are aligned to simplify the comparison.