MASARYKOVA UNIVERZITA

PŘÍRODOVĚDECKÁ FAKULTA

Genomická kompozice leukemických populací u pacientů s víceklonální chronickou lymfocytární leukémií

Diplomová práce

JANA ŠTRIGNEROVÁ

Vedoucí práce: Mgr. Karla Plevová, Ph.D.

Ústav experimentální biologie
Program Molekulární biologie a genetika

Brno 2025

MUNI
SCI

Bibliografický záznam

Autor: Jana Štrignerová
Přírodovědecká fakulta
Masarykova univerzita
Ústav experimentální biologie

Název práce: Genomická kompozice leukemických populací u pacientů s víceklonální chronickou lymfocytární leukémií

Studijní program: Molekulární biologie a genetika

Vedoucí práce: Mgr. Karla Plevová, Ph.D.

Rok: 2025

Počet stran: 138

Klíčová slova: chronická lymfocytární leukémie, B-buněčný antigenní receptor, přestavby genů těžkého řetězce imunoglobulinu, multiklonalita, klonální evoluce, sekvenování nové generace

Bibliographic record

Author: Jana Štrignerová
Faculty of Science
Masaryk University
Department of Experimental Biology

Title of Thesis: Genomic composition of leukemic populations in patients with multiclonal chronic lymphocytic leukemia

Degree Programme: Molecular Biology and Genetics

Supervisor: Mgr. Karla Plevová, Ph.D.

Year: 2025

Number of Pages: 138

Keywords: chronic lymphocytic leukemia, B-cell antigen receptor, immunoglobulin heavy chain gene rearrangements, multiclonality, clonal evolution, next-generation sequencing

Abstrakt

Chronická lymfocytární leukémie (CLL) je nejčastějším typem leukémie u dospělých v západních zemích, postihující především starší osoby. Typicky se jedná o klinicky heterogenní lymfoproliferativní maligní onemocnění charakterizované monoklonální expanzí a progresivní akumulací zralých B lymfocytů se specifickým imunofenotypem. Zatímco většina případů CLL vykazuje jednu produktivní přestavbu genů těžkého řetězce imunoglobulinu (IGH), malá, ale klinicky relevantní podskupina vykazuje více produktivních přestaveb IGH. Toto zjištění naznačuje koexistenci několika klonálních populací B lymfocytů u jednoho pacienta, z nichž každá je definována jedinečnou přestavbou IGH, což představuje další výzvy pro hodnocení prognózy pacientů.

Tato práce se zaměřuje na analýzu přestaveb imunoglobulinových genů v kohortě 76 pacientů s CLL s prokázanou přítomností více produktivních přestaveb IGH v jejich anamnéze. Vzhledem k prognostické a prediktivní hodnotě těchto přestaveb jsme pro jejich detekci srovnali tři komplementární metody: konvenční Sangerovo sekvenování, sekvenování nové generace pomocí genového panelu LYNX a sekvenování nové generace založené na amplikonech. Paralelní využití těchto technologií umožnilo detailně zhodnotit klonální sestavu imunoglobulinových přestaveb. Zde prezentované závěry nabízejí komplexnější pochopení klonální diverzity u multiklonální CLL a zdůrazňují metodologické aspekty, které jsou klíčové pro přesnou detekci klonotypů a stratifikaci onemocnění.

Abstract

Chronic lymphocytic leukemia (CLL) is the most common type of adult leukemia in Western countries, affecting primarily older individuals. It is a clinically heterogeneous lymphoproliferative malignancy characterised typically by the monoclonal expansion and progressive accumulation of mature B-lymphocytes with a specific immunophenotype. While most CLL cases display a single productive immunoglobulin heavy chain (IGH) gene rearrangement, a small but clinically relevant subset exhibits multiple productive IGH rearrangements. This finding reflects the coexistence of distinct clonal B-lymphocyte populations, each defined by a unique IGH rearrangement, within a single patient, posing additional challenges for patient prognostic stratification.

This thesis focuses on the analysis of immunoglobulin gene rearrangements in a cohort of 76 CLL patients with evidence of multiple productive IGH rearrangements in their disease history. Given the prognostic and predictive value of these rearrangements, we compared three orthogonal methodologies for their detection: conventional Sanger sequencing, capture-based next-generation sequencing (LYNX panel), and amplicon-based next-generation sequencing of IGH rearrangements. The parallel use of these techniques enabled a detailed and cross-validated assessment of the IGH clonal landscape. The findings presented here offer a more comprehensive understanding of clonal diversity in multiclonal CLL and highlight the methodological considerations critical for accurate clonotype detection and disease stratification.

MUNI
SCI

MASARYKOVA UNIVERZITA
PŘÍRODOVĚDECKÁ FAKULTA
KOTLÁŘSKÁ 2, 611 37 BRNO
IČ: 00216224
DIČ: CZ00216224

ZADÁNÍ DIPLOMOVÉ PRÁCE

Akademický rok: 2024/2025

Ústav:	Ústav experimentální biologie
Studentka:	Bc. Jana Štrignerová
Program:	Molekulární biologie a genetika
Specializace:	Molekulární biologie a genetika

Ředitel *ústavu* PřF MU Vám ve smyslu Studijního a zkušebního řádu MU určuje diplomovou práci s názvem:

Název práce:	Genomická kompozice leukemických populací u pacientů s víceklonální chronickou lymfocytární leukémií
Název práce anglicky:	Genomic composition of leukemic populations in patients with multiclonal chronic lymphocytic leukemia
Jazyk práce:	angličtina

Oficiální zadání:
U části pacientů s chronickou lymfocytární leukémií (CLL) transformuje souběžně několik klonů B lymfocytů. Větší počet leukemických klonů pak představuje riziko nepříznivého klonálního vývoje, protože každý z klonů může nést odlišné genomické aberace. Cílem diplomové práce bude charakterizovat proporce leukemických klonů u pacientů s víceklonální CLL a jejich genomické změny. Proporce klonů budou stanoveny pomocí detekce klonálních přestaveb imunoglobulinových genů a genomické varianty budou sledovány s využitím sekvenačních metod. Klíčová bude interpretace získaných dat a jejich zasazení do kontextu klonálního vývoje u sledované skupiny pacientů.

Vedoucí práce:	Mgr. Karla Plevová, Ph.D.
Konzultant:	Mgr. Kamila Stránská, Ph.D.
Datum zadání práce:	26. 10. 2023
V Brně dne:	8. 5. 2025

Zadání bylo schváleno prostřednictvím IS MU.

Bc. Jana Štrignerová, 27. 10. 2023
Mgr. Karla Plevová, Ph.D., 30. 10. 2023
RNDr. Pavel Lízal, Ph.D., 9. 11. 2023

Declaration

I declare that I have prepared my thesis independently under the mentorship of the thesis supervisor and using the information sources provided in the thesis.

In Brno, 15 May 2025

……
Jana Štrignerová

Acknowledgements

First of all, I would like to thank my supervisor, Mgr. Karla Plevová, Ph.D., for her guidance, support in the preparation of this thesis, valuable professional advice, constructive feedback, dedicated time, patience, and for the opportunity to work on this topic at the Institute of Medical Genetics and Genomics, Faculty of Medicine, Masaryk University, Brno.

Second, I would like to thank my consultant, Mgr. Kamila Stránská, Ph.D., for her help, willingness, and friendly approach throughout the preparation of this thesis. I would also like to thank Mgr. Jana Bruknerová for her introduction to laboratory practice, professional assistance, and support during the practical implementation of the laboratory experiments. My thanks go to Mgr. Kristýna Taušová for carrying out the amplicon-based IG aNGS experiments, and to Mgr. Tomáš Reigl for providing all the NGS data in a clear and accessible format.

I am also grateful to everyone involved in processing patient samples at the Centre of Molecular Biology and Genetics, Department of Internal Medicine, Hematology and Oncology, University Hospital Brno. Finally, I would like to thank all my colleagues from the Karla Plevová Research Group for creating a pleasant working environment.

The thesis was supported by the project "The Applicability of Comprehensive Genomic Testing – 2wards Better Diagnostics", reg. no. CZ.02.01.01/00/23_020/0008555, and the National Institute for Cancer Research, reg. no. LX22NPO5102, both co-financed by the European Union.

Table of Contents

List of Figures 11

List of Tables 13

Glossary 14

1 Introduction 17

1.1 Formation of immunoglobulin repertoire of B-lymphocytes 17

1.1.1 B-cell receptor and structure of immunoglobulin 17

1.1.2 Organisation of immunoglobulin loci in the genome 20

1.1.2.1 Immunoglobulin heavy chain gene 21

1.1.2.2 Immunoglobulin light chain kappa gene 22

1.1.2.3 Immunoglobulin light chain lambda gene 23

1.1.3 V(D)J recombination 24

1.1.4 Additional mechanisms increasing immunoglobulin diversity 28

1.1.4.1 Somatic hypermutation 29

1.1.4.2 Class switch recombination 30

1.2 Chronic lymphocytic leukemia 32

1.2.1 Epidemiology 32

1.2.2 Diagnosis 33

1.2.3 Typical molecular markers 34

1.2.3.1 Immunoglobulin heavy chain gene rearrangements 35

1.2.3.2 Stereotyped BCR subsets 35

1.2.3.3 Chromosomal aberrations and gene mutations 36

1.2.4 Prognosis 37

1.3 Multiclonal chronic lymphocytic leukemia 39

1.4 Methods of immunoglobulin gene rearrangements detection 41

2 Aims of the thesis 44

3 Material and methods 45

3.1 Material 45

3.2 Methods 45

3.2.1 Separation of B-lymphocytes from peripheral blood 45
3.2.2 RNA isolation and Sanger sequencing 47
3.2.3 DNA isolation and quality control 47
3.2.4 LYmphoid NeXt-Generation Sequencing Panel 47
3.2.5 MGIEasy Universal Library Conversion 50
3.2.6 Next-Generation Sequencing 51
3.2.7 Bioinformatic analysis of immunoglobulin rearrangements 52

4 Results 53

4.1 Sanger sequencing 53
4.2 Capture-based next-generation sequencing – LYNX panel 58
4.3 Amplicon-based next-generation sequencing – IG aNGS 64
4.4 Differences between monoclonal and multiclonal CLL 70
4.5 Comparison of methods for IGH rearrangement detection 73
4.6 Clonal evolution of IGH rearrangements in multiclonal CLL 84

5 Discussion 89

6 Conclusion 95

Bibliography 96

Appendix A Overview of IGHV genes in CLL 117

Appendix B Overview of IGHD genes in CLL 119

Appendix C Overview of IGHJ genes in CLL 121

Appendix D Overview of IGH rearrangements in multiclonal CLL 122

Appendix E Overview of IGH clonality in multiclonal CLL 133

List of Figures

Fig. 1: Schematic structure of a B-cell antigen receptor. 18
Fig. 2: Two-dimensional model of an immunoglobulin molecule. 19
Fig. 3: Representation of the chromosomal organization of the IGH gene clusters. 21
Fig. 4: Representation of the chromosomal organization of the IGK gene clusters. 22
Fig. 5: Representation of the chromosomal organization of the IGL gene clusters. 23
Fig. 6: V(D)J recombination and the generation of diversity in immunoglobulins. 26
Fig. 7: Mechanism of immunoglobulin heavy chain class switch recombination. 31
Fig. 8: Overview of the immunogenetic properties of major stereotyped subsets. 36
Fig. 9: Gene mutations and recurrent somatic copy number variations in CLL. 37
Fig. 10: The Typical RosetteSep™ Cell Enrichment Protocol. 46
Fig. 11: Illustration of genomic targets and molecular markers in the LYNX panel. 48
Fig. 12: The DNBSEQ™ technology developed by MGI. 52
Fig. 13: Clonal composition of IGH rearrangements analysed by Sanger sequencing. 54
Fig. 14: Distribution of IGHV families analysed by Sanger sequencing. 55
Fig. 15: Distribution of IGHV genes analysed by Sanger sequencing. 55
Fig. 16: Distribution of IGHV mutational status and concordance analysed by Sanger sequencing. 56
Fig. 17: Distribution of IGHD families analysed by Sanger sequencing. 57
Fig. 18: Distribution of IGHD genes analysed by Sanger sequencing. 57
Fig. 19: Distribution of IGHJ families analysed by Sanger sequencing. 58
Fig. 20: Distribution of median sequencing coverage across samples analysed by the LYNX panel. 59
Fig. 21: Variation in the number of detected productive IGH rearrangement fragments in the LYNX panel. 59
Fig. 22: Clonal composition of IGH rearrangements analysed by the LYNX panel. 60
Fig. 23: Distribution of IGHV families analysed by the LYNX panel. 61
Fig. 24: Distribution of IGHV genes analysed by the LYNX panel. 61
Fig. 25: Distribution of IGHD families analysed by the LYNX panel. 62
Fig. 26: Distribution of IGHD genes analysed by the LYNX panel. 63
Fig. 27: Distribution of IGHJ families analysed by the LYNX panel. 63
Fig. 28: Distribution of total read counts across samples obtained from IG aNGS analysis. 64
Fig. 29: Variation in detected reads across productive IGH rearrangements in IG aNGS. 65
Fig. 30: Clonal composition of IGH rearrangements analysed by IG aNGS. 66
Fig. 31: Distribution of IGHV families analysed by IG aNGS. 67
Fig. 32: Distribution of IGHV genes analysed by IG aNGS. 67
Fig. 33: Distribution of IGHV mutational status and concordance analysed by IG aNGS. 68
Fig. 34: Distribution of IGHD families analysed by IG aNGS. 69
Fig. 35: Distribution of IGHD genes analysed by IG aNGS. 69
Fig. 36: Distribution of IGHJ families analysed by IG aNGS. 70
Fig. 37: Distribution of IGH families among monoclonal CLL patients. 71

Fig. 38: Distribution of IGH genes among monoclonal CLL patients. 73
Fig. 39: Relative abundance differences of IGH rearrangements detected by NGS methods. 76
Fig. 40: Comparative analysis of three IGH clonality detection methods in multiclonal CLL. 77
Fig. 41: Comparative detection efficiency of IGHV mutational status in LYNX (A) and IG aNGS (B). 79
Fig. 42: Impact of IGHV mutational identity on detection efficiency in LYNX (A) and IG aNGS (B). 80
Fig. 43: Comparison of capture efficiency for mutated and unmutated IGHV sequences in monoclonal CLL across NGS platforms. 81
Fig. 44: Comparison of capture efficiency for mutated and unmutated IGHV sequences in all detected productive IGH rearrangements across NGS platforms. 82
Fig. 45: Comparison of capture efficiency for mutated and unmutated IGHV sequences in all detected productive IGH rearrangements unaffected by unproductive IGH rearrangements across NGS platforms. 83
Fig. 46: Clonal dynamics of IGH rearrangements in patient BRNO0319 over time. 86
Fig. 47: Clonal dynamics of IGH rearrangements in patient BRNO1048 over time. 87
Fig. 48: Clonal dynamics of IGH rearrangements in patient BRNO1502 over time. 88
Fig. 49: Clonal dynamics of IGH rearrangements in patient BRNO1718 over time. 88

List of Tables

Tab. 1: A comprehensive overview of the IGHV families and their corresponding genes. 117
Tab. 2: A comprehensive overview of the IGHD families and their corresponding genes. 119
Tab. 3: A comprehensive overview of the IGHJ families and their corresponding genes. 121
Tab. 4: A detailed list of all detected productive IGH rearrangements in multiclonal CLL. 122
Tab. 5: A comprehensive overview of IGH clonality results in multiclonal CLL patients. 133

Glossary

AID	–	activation-induced deaminase
ALL	–	acute lymphoblastic leukemia
BCR	–	B-cell receptor
BDL	–	below detection limit
BL	–	Burkitt lymphoma
C	–	immunoglobulin constant domain / gene
cDNA	–	complementary DNA
CDR	–	complementarity-determining region
CLL	–	chronic lymphocytic leukemia
CLL-IPI	–	CLL International Prognostic Index
cnLOH	–	copy neutral loss of heterozygosity
CNV	–	copy number variant
CSR	–	class switch recombination
D	–	immunoglobulin diversity gene
DLBCL	–	diffuse large B-cell lymphoma
DMSO	–	dimethyl sulfoxide
DNA-PK	–	DNA-dependent protein kinase
DNB	–	DNA nanoball
DSB	–	double-strand break
EDTA	–	ethylenediaminetetraacetic acid
FBS	–	fetal bovine serum
FFPE	–	formalin-fixed, paraffin-embedded block
FISH	–	fluorescence in situ hybridisation
FL	–	follicular lymphoma
FR	–	framework region

gDNA	–	genomic DNA
HMG	–	high mobility group protein
IG	–	immunoglobulin
IG aNGS	–	amplicon-based NGS of IGH gene rearrangements
IGH	–	immunoglobulin heavy chain
IGHC	–	immunoglobulin heavy chain constant gene
IGHD	–	immunoglobulin heavy chain diversity gene
IGHJ	–	immunoglobulin heavy chain joining gene
IGHV	–	immunoglobulin heavy chain variable gene
IGK	–	immunoglobulin kappa light chain
IGKC	–	immunoglobulin light chain kappa constant gene
IGKJ	–	immunoglobulin light chain kappa joining gene
IGKV	–	immunoglobulin light chain kappa variable gene
IGL	–	immunoglobulin lambda light chain
IGLC	–	immunoglobulin light chain lambda constant gene
IGLJ	–	immunoglobulin light chain lambda joining gene
IGLV	–	immunoglobulin light chain lambda variable gene
IQR	–	interquartile range
ITAM	–	immunoreceptor tyrosine-based activation motif
J	–	immunoglobulin joining gene
KDE	–	κ-deleting element
LYNX	–	LYmphoid NeXt-Generation Sequencing Panel
M	–	mutated IGHV
MBL	–	monoclonal B-cell lymphocytosis
MCL	–	mantle cell lymphoma
mIg	–	membrane-bound immunoglobulin
MM	–	multiple myeloma

NGS	–	next-generation sequencing
NHEJ	–	nonhomologous end-joining
ORF	–	open reading frame
PBS	–	phosphate-buffered saline
PCR	–	polymerase chain reaction
RAG	–	recombination-activating gene
RCR	–	rolling circle replication
RSS	–	recombination signal sequence
S	–	switch region
SHM	–	somatic hypermutation
SNP	–	single-nucleotide polymorphism
TdT	–	terminal deoxynucleotidyl transferase
TE buffer	–	Tris-EDTA buffer
U	–	unmutated IGHV
UNG	–	uracil-DNA glycosylase
V	–	immunoglobulin variable domain / gene
WM	–	Waldenström's macroglobulinemia
X	–	unknown IGHV mutational status

1 Introduction

1.1 Formation of immunoglobulin repertoire of B-lymphocytes

B-lymphocytes can recognize structurally diverse antigens, and their receptors play key roles in B-cell development and activation (McHeyzer-Williams and McHeyzer-Williams, 2005). B-lymphocyte activation is triggered by the binding of specific antigen to the B-cell antigen receptor (BCR), which initiates a cascade of intracellular signals, leading to the internalization of antigen for processing and presentation to T-lymphocytes (Lanzavecchia, 1985). Upon antigen recognition, B-cells can be activated and differentiated into plasma cells that secrete soluble effector molecules – antibodies, which can individually bind to the antigens and neutralize them (McHeyzer-Williams and McHeyzer-Williams, 2005; Schroeder and Cavacini, 2010).

The formation of BCR is a unique and crucial process in B-lymphocytes. Its basic structural unit is the immunoglobulin (IG) molecule expressed on the cell surface (Hombach et al., 1990b). Although the genes encoding IG chains form only a small part of the genome, they can generate a large repertoire of IGs with a wide spectrum of specificity, capable of binding to a virtually unlimited array of antigens (Schroeder and Cavacini, 2010). This diversity is ensured by three main mechanisms ongoing during the B-cell development: V(D)J recombination, class switch recombination (CSR) and somatic hypermutation (SHM). Because of this diversity, organisms can efficiently respond to a broad range of pathogens and protect themselves through adaptive immunity (Pike and Ratcliffe, 2002).

1.1.1 B-cell receptor and structure of immunoglobulin

The BCR is a transmembrane protein complex composed of extracellular membrane-bound monomeric immunoglobulin (mIg), which serves as the central component for antigen recognition, and the Igα/β heterodimer, responsible for signal transduction through the mIg in the BCR (Fig. 1) (Hombach et al., 1990b; Sanchez et al., 1993; Venkitaraman et al., 1991). The mIg and associated Igα/β heterodimer occur in a 1:1 stoichiometry (Dong et al., 2022; Schamel and Reth, 2000; Tolar et al., 2005).

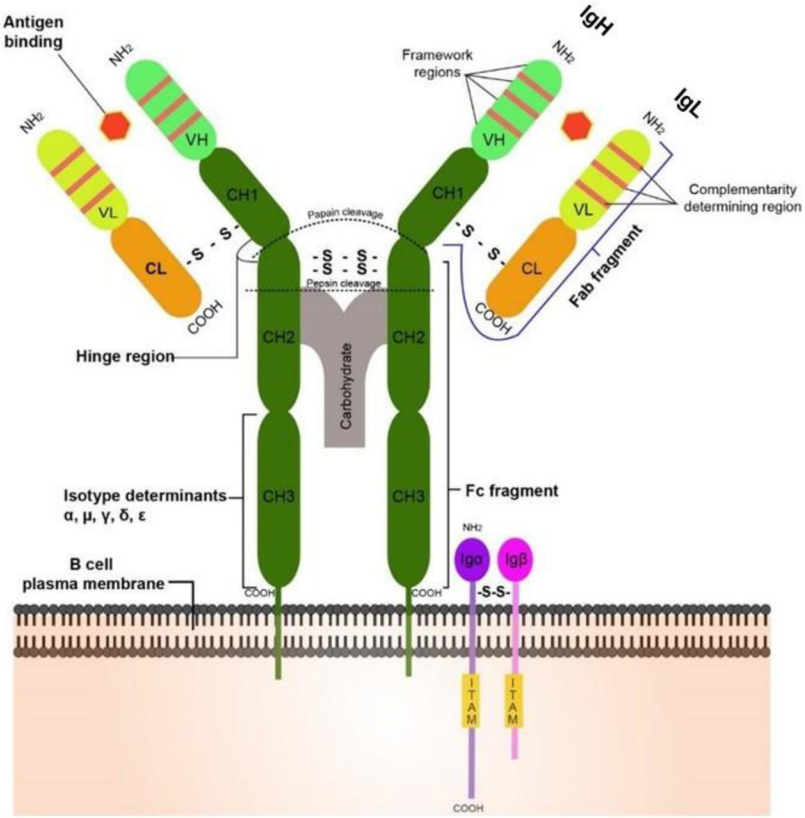


Fig. 1: Schematic structure of a B-cell antigen receptor.

The B-cell receptor consists of a membrane-bound immunoglobulin (mIg) in association with the Igα/β heterodimer. The mIg is composed of two immunoglobulin heavy chains (IGHs) and two immunoglobulin light chains (IGLs). All chains have C-terminal constant (C) domains and N-terminal variable (V) domains, which consists of complementarity-determining regions inserted between framework regions. The IGHs are bound together by disulfide bonds, as are the IGLs, which are bound to the IGHs (Waheed, 2018).

The mIg comprises symmetric homodimers consisting of identical membrane-bound immunoglobulin heavy chains (IGHs) and identical immunoglobulin light chains, which can be either a κ (IGK) or a λ (IGL) chain (Fig. 1) (Torres et al., 2008; Williams and Barclay, 1988). Both IGH and immunoglobulin light chain covalently associate through disulfide bonds, forming an antigen recognition domain with the potential for ligand binding. The IGHs are also covalently linked to each other by disulfide bonds and are anchored to the cytoplasmatic membrane of B-lymphocytes by a short intracellular domain consisting of just three amino acids: lysine, valine, lysine (KVK) (Treanor, 2012). Both polypeptide chains of IGH and immunoglobulin light chain consist of C-terminal constant (C) domains and N-terminal variable (V) domains, which serve distinct functions. The V domains are responsible for antigen binding and determine receptor specificity, while the C domains

determine the IG class and specify effector functions, such as complement activation or binding to Fc receptors (Corfe and Paige, 2009; Schroeder and Cavacini, 2010).

Each V domain is functionally divided into three hypervariable regions, termed complementarity-determining regions (CDRs), which are situated between four conserved framework regions (FRs) that provide structural stability (Fig. 2). The three CDRs of the IGH are paired with the three CDRs of the immunoglobulin light chain to form the antigen-binding site. The V domains of the IGH are formed by the joining of variable (V), diversity (D), and joining (J) genes in a process known as V(D)J recombination. The structure of the immunoglobulin light chain V domains is similar, except that it lacks D genes (Corfe and Paige, 2009). The CDR3 region exhibits the highest variability because it contains gene junction site including random nucleotides (N; V-J gene junction in immunoglobulin light chain, V-D and D-J gene junction in IGH).

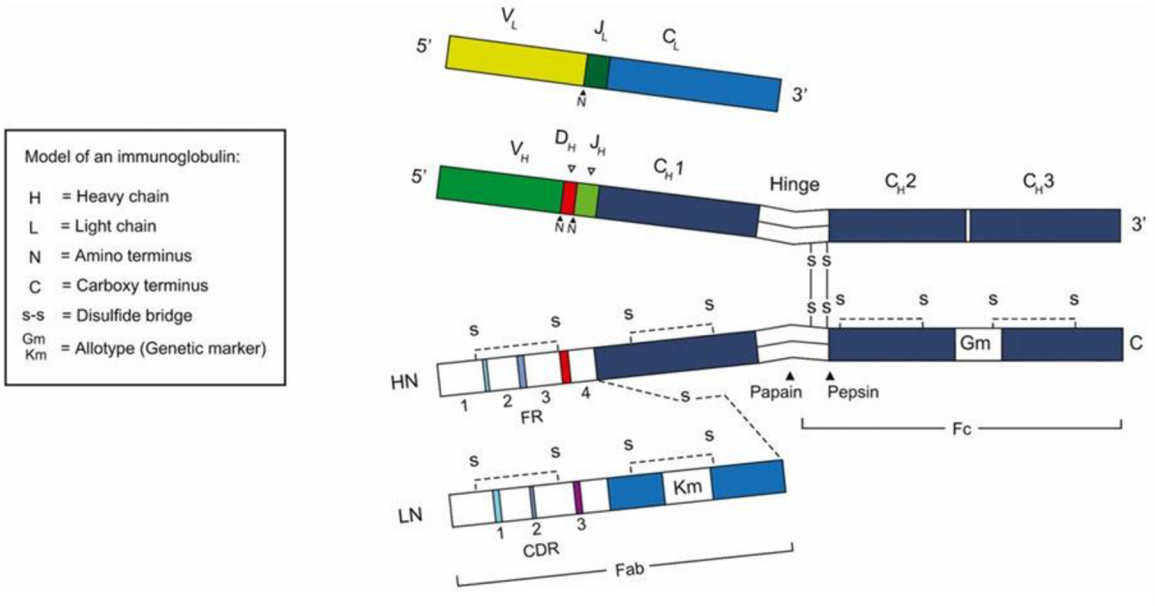


Fig. 2: Two-dimensional model of an immunoglobulin molecule.
The immunoglobulin heavy chain (IGH) and immunoglobulin light chain (IGL) at the top deconstruct the antibody at the nucleotide level, while the chains at the bottom deconstruct the protein sequence. The IGH variable domains are formed by the joining of variable (V), diversity (D), and joining (J) genes, whereas the IGL variable domains are formed only by V and J genes (Schroeder and Cavacini, 2010).

The C domains, which are the most compact, show very little variation among IGs. Both IGK and IGL contain only one C domain, whereas IGH contains either three or four C domains, depending on the IGH class (Fig. 1, Fig. 2). They typically include a hinge region,

usually between the IGHC1 and IGHC2 domain in IGHs (Torres et al., 2008). In humans, there are five types of IGH C domain-encoding genes (μ, δ, γ, α, ε), which define the IgM, IgD, IgG, IgA, and IgE isotypes. IgG can be further divided into four subclasses – IgG1, IgG2, IgG3, and IgG4, each with its own biologic properties, and IgA can similarly be split into IgA1 and IgA2 (Schroeder and Cavacini, 2010).

Each BCR is composed of an IgM monomer as a mIg, which does not contain any signalling motifs and therefore non-covalently associate with the Igα/β heterodimer (Hombach et al., 1990b; Sanchez et al., 1993; Schamel and Reth, 2000). The Igα/β heterodimer is comprised of Igα (also known as CD79a or MB-1 membrane glycoprotein) and Igβ (also known as CD79b or B-cell-specific glycoprotein B29) which are linked by a disulfide bond (Campbell et al., 1991; Campbell and Cambier, 1990; Hermanson et al., 1988; Hombach et al., 1990a; Sakaguchi et al., 1988). Each chain contains an extracellular Ig domain, a transmembrane region and a cytoplasmic domain (Fig. 1). At the extracellular domain, the Igα/β heterodimer mainly uses Igα to associate with the IGHC3 or IGHC4 domain of one IGH, while leaving the other IGH unbound (Dong et al., 2022). The cytoplasmic domains of both Igα and Igβ carry immunoreceptor tyrosine-based activation motifs (ITAMs); a conserved sequence of four amino acids in which a tyrosine is separated from a leucine or isoleucine by any two amino acids (YxxL/I) (Flaswinkel and Reth, 1994; Reth, 1989).

1.1.2 Organisation of immunoglobulin loci in the genome

IGHs and immunoglobulin light chains are each encoded by separate multigene families located in three unlinked loci. The IGH gene locus is located on chromosome 14, the IGK gene locus is on chromosome 2, and the IGL gene locus is on chromosome 22. The individual V and C domains are each encoded by independent elements: the V domain by V(D)J genes, and the C domains by individual exons. IG genes are spaced along the same chromosome in a way that facilitates recombination, allowing the recombination machinery to bind two coding segments and join them together (Corfe and Paige, 2009).

1.1.2.1 Immunoglobulin heavy chain gene

The IGH locus is located on chromosome 14q32.33, at the telomeric extremity of the long arm, spanning 1250 kb (Kirsch et al., 1982; McBride et al., 1982a). It is considerably more complex than the immunoglobulin light chain clusters and contains 170 – 176 IGH genes per haploid genome, depending on the haplotypes, of which 77 – 84 genes are functional (Honjo, 1983; Lefranc, 2001a). The human IGH locus consists of 123 – 129 IGHV genes (depending on the haplotypes), 27 IGHD genes, 9 IGHJ genes, and, in the most frequent haplotype, 11 IGHC genes, each encoding a separate C domain. The orientation of the locus on the chromosome is reverse. The IGHV genes occupy a position very close to the telomere of the long arm of the chromosome, whereas the IGHC genes are in a more centromeric position. The distribution of all regions of the IGH genes on the chromosome is as follows: telomere – IGHV – IGHD – IGHJ – IGHCμ – IGHCδ – IGHCγ3 – IGHCγ1 – ψIGHCε (pseudogene) – IGHCα1 – ψIGHCγ (pseudogene) – IGHCγ2 – IGHCγ4 – IGHCε – IGHCα2 – centromere (Fig. 3).

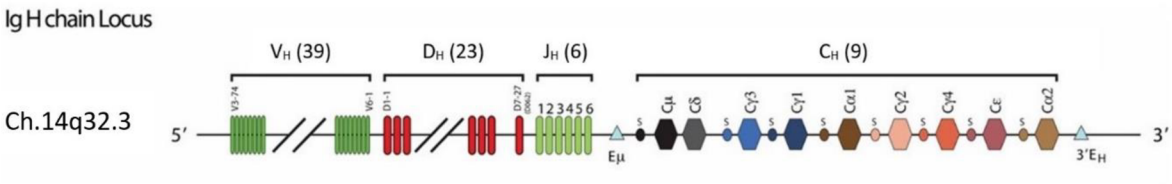


Fig. 3: Representation of the chromosomal organization of the IGH gene clusters.
The immunoglobulin heavy chain (IGH) locus is located on chromosome 14q32.33 and consists of variable (V_H), diversity (D_H), joining (J_H) and constant (C_H) genes. The typical numbers of functional genes present in the locus are shown in brackets (Schroeder and Cavacini, 2010; edited).

Of the 123 IGHV genes, 79 are pseudogenes, leaving 44 IGHV genes with an open reading frame (ORF). Of these, 39 are functional and expressed as heavy chain proteins, and one is transcribed as mRNA (Matsuda et al., 1998). Genetic polymorphisms can result in IG haplotypes with expansion or contraction of the number of IGHV genes (Cook and Tomlinson, 1995). The IGHV genes can be grouped into 7 different families of related genes, IGHV1 – IGHV7 (Lefranc, 2001a).

Of the 27 identified IGHD genes, 23 are functional, and of the 9 IGHJ genes, 6 are functional. Together with IGHV genes, they form CDR3, the most diverse region, which mediates the antigen binding to IG (Schroeder and Cavacini, 2010). Of the 11 IGHC genes,

9 are functional, and 2 are pseudogenes (Matsuda et al., 1998). The IGHC exons can encode two forms of each IGH isotype: an integral membrane protein that anchors the BCR and a soluble protein secreted as an antibody by plasma cells. The choice of the transmembrane or secretory C-terminal exon is regulated at the transcriptional level (Staudt and Lenardo, 1991).

1.1.2.2 Immunoglobulin light chain kappa gene

The IGK locus is located on chromosome 2p11.2 and spans 1820 kb (McBride et al., 1982b). The orientation of the locus on the chromosome is reverse (Lefranc, 2001b). The IGK locus contains two large inverted clusters separated by 800 kb (Ermert et al., 1995). These clusters contain IGKV genes, and the nucleotide sequence homology between these genes is extremely high (98.9% on the average) (Huber et al., 1993; Pargent et al., 1991). The IGKV genes are located upstream of 5 IGKJ genes and a single IGKC exon that encodes the entire C region of the IGK (Fig. 4) (Ermert et al., 1995). The IGK gene cluster also includes a κ-deleting element (KDE) that can rearrange to sequences upstream of IGKC in cells that express IGL, reducing the likelihood of dual IGK and IGL expression (Schroeder and Cavacini, 2010).

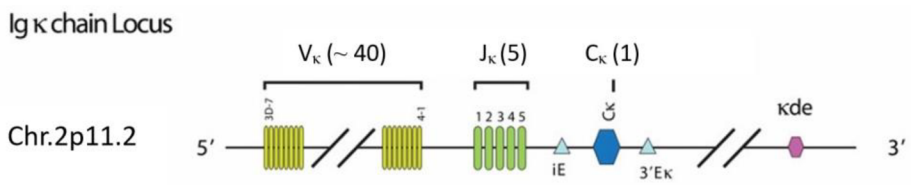


Fig. 4: Representation of the chromosomal organization of the IGK gene clusters.
The immunoglobulin kappa light chain (IGK) locus is located on chromosome 2p11.2 and consists of variable (V_κ), joining (J_κ) and constant (C_κ) genes. The typical numbers of functional genes present in the locus are shown in brackets. The IGK gene cluster also includes a κ-deleting element (KDE) at the end of the locus (Schroeder and Cavacini, 2010; edited).

Initially, 76 IGKV genes were discovered: 36 IGKV genes in the distal cluster (at the 5' end of the locus, in the most centromeric position), spanning 400 kb, and 40 IGKV genes in the proximal cluster (at the 3' end of the locus, closer to IGKC and in the most telomeric position), spanning 600 kb (Ermert et al., 1995). Additionally, one rare IGKV haplotype has been described that contains only the proximal cluster (Lefranc, 2001b). Later, 55 truncated pseudogenes and relics were newly identified (Kawasaki et al., 2001).

Thus, there are 131 IGKV genes in total: 69 in the proximal cluster and 62 in the distal cluster, of which 46 were identified with ORF (Kawasaki et al., 2001). Based on the nucleotide sequence homology, IGKV genes are classified into 7 families, IGKV1 – IGKV7, with the first 5 families containing functional IGKV genes (Lefranc, 2001b).

1.1.2.3 Immunoglobulin light chain lambda gene

The IGL locus is located on chromosome 22q11.2, spanning 1050 kb, and is situated 6 Mb from the centromere (Dunham et al., 1999; Emanuel et al., 1985). The orientation of the locus on the chromosome is forward (Lefranc, 2001c). The most 5' IGLV genes occupy the more centromeric position and form a V-cluster, while the IGLC genes, located at the 3' end, are the most telomeric genes in IGL locus. Together with the IGLJ genes, they form a J-C-cluster, where each IGLJ gene is associated with its corresponding IGLC gene (Fig. 5).

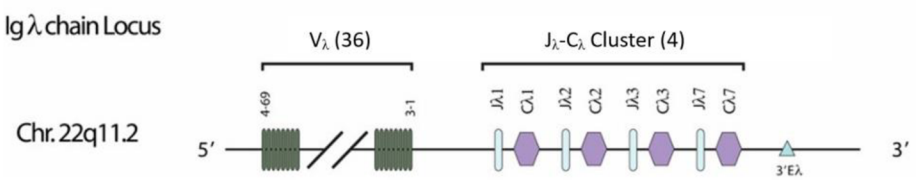


Fig. 5: Representation of the chromosomal organization of the IGL gene clusters.
The immunoglobulin lamda light chain (IGL) locus is located on chromosome 22q11.2 and consists of variable (V_λ), joining (J_λ) and constant (C_λ) genes. Each C_λ exon is associated with its corresponding J_λ gene. The typical numbers of functional genes present in the locus are shown in brackets (Schroeder and Cavacini, 2010; edited).

Initially, 69 IGLV genes were discovered: 36 potentially active genes and 33 pseudogenes (Kawasaki et al., 1997). Later, extensive characterization of 50 new IGLV pseudogenes was made, bringing the total number of IGLV genes in the IGL locus to 119 (Kawasaki et al., 2000). The IGLV genes are arranged in three distinct clusters on the chromosome: A, B and C, each containing members of different IGLV families (Frippiat et al., 1995; Williams et al., 1996). We distinguish 11 IGLV families, IGLV1 – IGLV11, based on the nucleotide sequence homology. Cluster A (nearest the IGLJ genes) contains the IGLV2 and IGLV3 families, as well as one member of the IGLV4 family. Cluster B (in the middle of the IGL locus) contains the IGLV1, IGLV5, IGLV7 and IGLV9 families. Cluster C (furthest from the IGLJ genes) contains the two other members of the IGLV4 family and the IGLV6, IGLV8, IGLV10 and IGLV11 families (Lefranc, 2001c).

The IGLV genes are followed by 7 – 11 IGLJ genes and 7 – 11 IGLC genes, depending on the haplotypes. Each IGLC gene is preceded by one IGLJ gene at the 5' end (Hieter et al., 1981; Vasicek and Leder, 1990). Of these, 4 – 5 IGLJ and 4 – 5 IGLC genes are functional: IGLJ1 + IGLC1, IGLJ2 + IGLC2, IGLJ3 + IGLC3, IGLJ7 + IGLC7 (Lefranc, 2001c; Schroeder and Cavacini, 2010).

1.1.3 V(D)J recombination

The primary repertoire of IGs is created during lymphocyte development by the joining of V, D, and J genes from both IGH and immunoglobulin light chain. This process, known as V(D)J recombination, occurs in lymphocyte precursors and is one of the fundamental processes that generate diverse antigen receptor repertoires (Bassing et al., 2002; Ramsden et al., 1997). V(D)J recombination is a site-specific process that occurs between genes flanked by conserved recombination signal sequences (RSSs). Each RSS consists of a highly conserved palindromic heptamer (5'-CACAGTG) and an AT-rich nonamer (5'-ACAAAAACC), separated by non-conserved 12- or 23-bp spacers (Tonegawa, 1983). V(D)J recombination occurs only between genes with RSS containing different spacers lengths. A 12-bp spacer (12-RSS) preferentially pairs with a 23-bp spacer (23-RSS), in accordance with the 12/23 rule, facilitating efficient recombination and thereby preventing wasteful V-V or J-J rearrangements (Steen et al., 1996; van Gent et al., 1996).

V(D)J recombination is initiated by the introduction of DNA double-strand breaks (DSBs) between the termini of the rearranging gene and its adjacent RSS. This process is catalysed by the recombination-activating genes RAG1 and RAG2, which are predominantly expressed in developing lymphocytes (Oettinger et al., 1990; Schatz et al., 1989). The initial step involves the recognition of a 12-RSS by RAG1. Subsequently, RAG2 associates with RAG1 and the heptamer sequence to form a synaptic complex (Hiom and Gellert, 1998; van Gent et al., 1996). Binding of a second RAG1-RAG2 complex to a 23-RSS allows interaction between the two synaptic complexes, resulting in the formation of a paired complex. This assembly is facilitated by DNA-bending accessory factors known as high mobility group proteins HMG1 and HMG2 (West and Lieber, 1998). Following paired complex formation, the RAG proteins introduce single-strand break on the top strand

between the gene segment and the first nucleotide of the heptamer. In the subsequent hairpin formation step, the bottom strand is cleaved, creating a hairpin structure at the coding end and a blunt signal end (McBlane et al., 1995).

The RAG-initiated DSBs are then repaired through the nonhomologous end-joining (NHEJ) pathway, a DNA repair mechanism common to all cells (Taccioli et al., 1993). In this process, the nuclease Artemis opens the hairpin structure at the coding end, allowing the binding of other NHEJ proteins, including DNA-dependent protein kinase (DNA-PK) catalytic subunit along with its regulatory subunits Ku70 and Ku80, protein XRCC4, and DNA ligase IV, which ligates the genes to form a new coding sequence (Critchlow et al., 1997; Grawunder et al., 1998, 1997; Khanna and Jackson, 2001; Li et al., 1995; Ma et al., 2002; Moshous et al., 2001).

The NHEJ process creates precise joins between the RSS-bearing ends, while the joins between the coding ends are imprecise, introducing further diversity at the junction between the coding ends. The coding ends are modified through processes that involve nucleotide loss or addition (Tonegawa, 1983). Single-stranded DNA may contain a few nucleotides from the coding sequence, along with the complementary nucleotides from the opposite DNA strand, forming short palindromic sequences of P nucleotides added to the ends of the genes (Janeway et al., 2001). Then, terminal deoxynucleotidyl transferase (TdT), which is expressed only in lymphocytes, can replace or add non-germline-encoded nucleotides (N nucleotides) to the coding ends of the recombination product (Cabaniols et al., 2001; Komori et al., 1993). The terminus of each rearranging gene can also undergo a loss of one to five nucleotides during the recombination process, likely due to the other isoform of TdT, yielding additional junctional diversity (Thai et al., 2002; Thai and Kearney, 2004; Troshchynsky et al., 2015).

The rearrangement of IG genes is an ordered, sequential process, typically commencing at the IGH locus (Alt et al., 1984; Korsmeyer et al., 1981). Each IGHV and IGHJ gene is associated with a 23-RSS, which prevents direct V-J joining. A pair of 12-RSS flanks each IGHD segment (Tonegawa, 1983). Recombination begins with the joining of a IGHD to a IGHJ gene, followed by the joining of a IGHV gene to the amino-terminal end of the IGHD-IGHJ intermediate (Fig. 6). Only one-third of rearrangements are in frame and,

therefore, productive (Bassing et al., 2002). In such case, the completion of a productive rearrangement leads to the inhibition of rearrangements on the second allele through allelic exclusion (Löffert et al., 1996). However, progenitor cells that first make unproductive rearrangements go on to rearrange their second allele, thus, most of differentiating lymphocytes achieve productive rearrangements. In a small subset of cells, this control fails, leading to allelic inclusion (incomplete allelic exclusion or allelic escape), and B-cells express dual (or multiple) BCRs (Shi et al., 2019; Zhu et al., 2023). If the rearrangement process fails, B lineage cells undergo apoptosis and are rapidly engulfed by resident macrophages (Osmond et al., 1994).

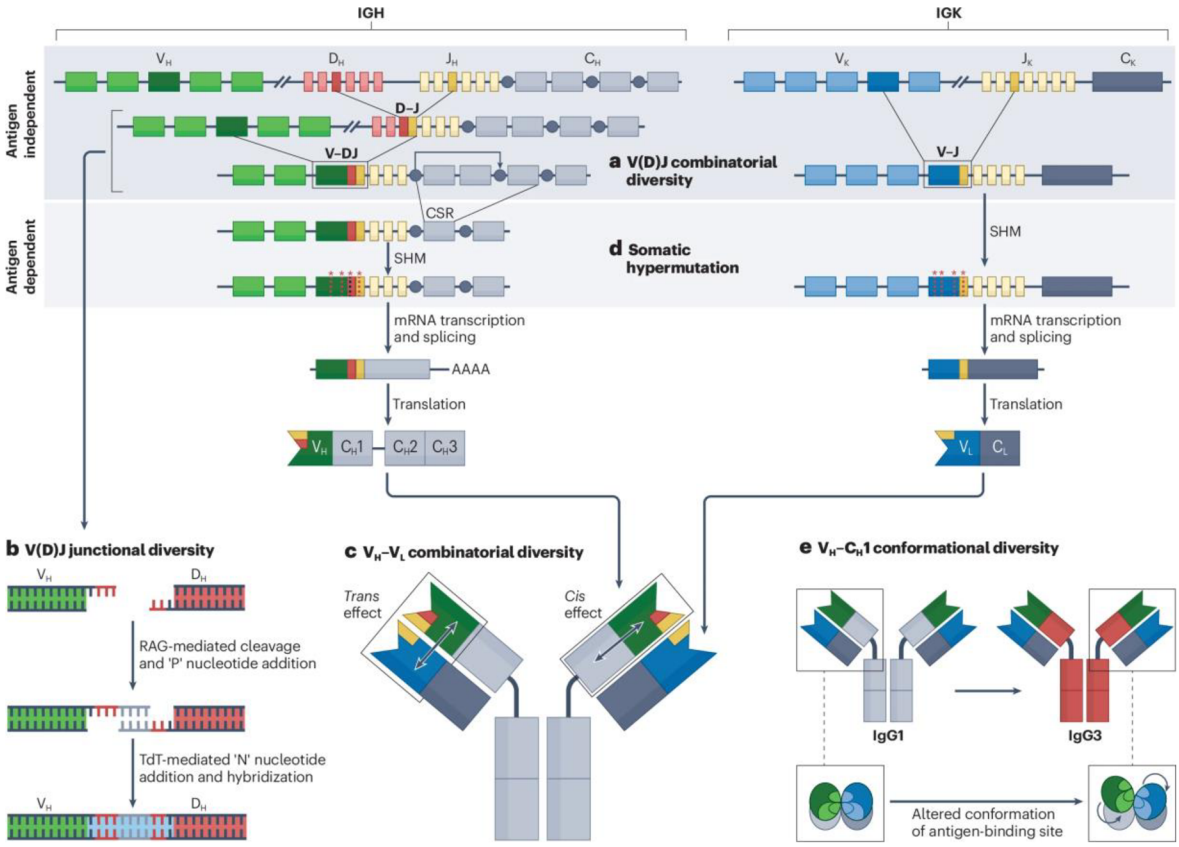


Fig. 6: V(D)J recombination and the generation of diversity in immunoglobulins.
In naive pre-B-cells, genetic rearrangements of the immunoglobulin heavy chain (IGH) and immunoglobulin light chain kappa (IGK) loci occur at the DNA level. For the IGH locus, diversity (D_H) and joining (J_H) genes recombine (D–J), followed by the variable (V_H) gene joining the DJ rearrangement (V–DJ). In contrast, for the IGK locus, a single variable and joining gene rearrangement (V_K–J_K) occurs. In both cases, V(D)J combinatorial diversity is introduced because of the large number of possible permutations of V(D)J gene combinations (**a**). Additionally, V(D)J junctional diversity is introduced during rearrangement through stochastic additions of 'P' nucleotides (shown in red) or 'N' nucleotides (shown in grey) during the genetic

fusion of the V(D)J segments (**b**). After recombination, genes are removed from the chromosome, and the entire gene is transcribed into mRNA, where subsequent RNA splicing removes the unused flanking genes. The mRNA transcripts are then translated to produce IGH and IGK, which are paired in the endoplasmic reticulum and expressed as membrane-bound B-cell receptor (BCR). At this stage, V_H–V_L combinatorial diversity is introduced by the pairing of different IGH and light chain (IGL) variable regions. This diversity is further expanded by the selection of either kappa or lambda IGL. This process concludes if the naive IGH and IGL rearrangements yield a functional BCR that is stably expressed on the cell surface and is not negatively selected due to self-antigen recognition (**c**). When a B-cell encounters an antigen, additional diversity is introduced in immunoglobulin sequences by a process known as somatic hypermutation (SHM), in which random mutations (indicated by red stars) are introduced to each variable region gene (**d**). In addition, antigen-experienced B-cells receive cytokine signals from T helper cells, leading to class switch recombination (CSR), in which the isotype of the soluble expressed antibody is switched via genetic alterations of the IGH constant (C) region at the DNA level. The alteration of antibody isotype has been shown to subtly modify the conformation of the variable-region antigen-binding site through the influence of the C_H1 domain (referred to as V_H–C_H conformational diversity), potentially resulting in altered antibody specificity (**e**). RAG, recombination activating gene; TdT, terminal deoxynucleotidyl transferase (McConnell and Casadevall, 2025).

The rearrangement activity then shifts to the immunoglobulin light chain loci. The formation of the immunoglobulin light chain requires only a single joining reaction between the V and J genes (Fig. 6). Expression of immunoglobulin light chains is not equally represented, IGK rearrangement typically precedes IGL rearrangement, with IGK rearrangements forming 65% of immunoglobulin light chains rearrangements (Corfe and Paige, 2009; Korsmeyer et al., 1981). The completion of a productive rearrangement also leads to the inhibition of rearrangements on the second allele through allelic exclusion. In addition, rearrangements are halted at the other immunoglobulin light chain locus in a process called isotype exclusion, ensuring the expression of only either IGK or IGL (Gorman and Alt, 1998). If a productive rearrangement is not successful on the first allele of the IGK locus, rearrangements are initiated on the second allele. If this also fails, rearrangements proceed at the IGL locus. In the case of only unproductive immunoglobulin light chains are produced, or if the productive immunoglobulin light chains cannot pair with the pre-existing IGH or are self-reactive, the B-cell can undergo apoptosis or be rescued via receptor editing (Gay et al., 1993; Radic et al., 1993; Tiegs et al., 1993). Receptor editing is a process by which B-cells undergo secondary rearrangements of their immunoglobulin light chain genes, without further rearrangement of the IGH (Nemazee and Weigert, 2000). It was also shown that receptor editing is more difficult for IGH but still possible, adding further complexity to the BCR (Radic and Zouali, 1996; Wilson et al., 2000).

However, the IG gene rearrangement order IGH – IGK – IGL is not inviolate, and IGK gene rearrangement can precede IGH rearrangement (Ehlich et al., 1993; Kubagawa et al., 1989; Novobrantseva et al., 1999). Thus, the production of a functional IGH is not an absolute prerequisite for immunoglobulin light chain gene rearrangement, nor is the failure of rearrangement of both IGK alleles a prerequisite for IGL gene rearrangement (Burrows et al., 2004; Chen et al., 1993).

The diversity of the IG repertoire observed in BCRs results from several layers of complexity (Fugmann et al., 2000). The first, germline-encoded combinatorial diversity, is achieved by the random selection of different V, D, and J genes, with each different V(D)J combination yielding a unique receptor specificity (Fig. 6a). Compared with immunoglobulin light chains, IGHs, which include D gene elements, generate considerably more combinatorial diversity (Nemazee, 2006). There are approximately 6 300 possible combinations of the gene elements that comprise IGH chains. In contrast, immunoglobulin light chains, which are encoded by two independent loci, IGK and IGL, have only 350 possible combinations, which is ~ 95% fewer than the number of possible IGH chains.

The second, junctional diversity, is a major source of variation in the preimmune repertoire (Schroeder and Cavacini, 2010). It arises from the imprecise joining of the V, D, and J genes, along with the addition of non-templated (N) nucleotides by TdT (Fig. 6b). This process alters the specificity of receptors encoded by identical V, D, and J genes and allows the generation of CDR3s of varying length and structure. As a result, more than 10^6 IGHs and more than 10^5 immunoglobulin light chains can be generated at the IG rearrangements. Taken as a whole, combinatorial rearrangement of individual genes, junctional diversity leading to variation in CDR3, and the combinatorial association between different immunoglobulin light chains and IGH chains (Fig. 6c) can yield a potential preimmune IG repertoire of greater than 10^{12} different IGs.

1.1.4 Additional mechanisms increasing immunoglobulin diversity

Moreover, IG loci are subject to additional genomic alterations that further expand the IG repertoire in the periphery (Tonegawa, 1983). This occurs through the affinity maturation in response to repeated immunization or antigen exposure (Berek and Milstein, 1987;

Schroeder and Cavacini, 2010; Steiner and Eisen, 1967). After undergoing V(D)J recombination, B-cells migrate into B-cell follicles of secondary lymphatic organs, where they rapidly expand and establish germinal centres (Jacobs and Bross, 2001; Rajewsky, 1996). Here, B-cells modify their rearranged IG genes to enhance both the ability of IGs to recognize and bind to foreign antigens and the effector capacities of the expressed IGs (Chi et al., 2020). This is due to two genetic modifications that B-cells undergo: somatic hypermutation (SHM) and IGH class switch recombination (CSR) (Jacobs and Bross, 2001; Manis et al., 2002). The molecular mechanisms of SHM and CSR share many similarities and several enzymes, but the key factor necessary for both is activation-induced deaminase (AID) (Dudley et al., 2005; Neuberger, 2008). AID is a single-strand DNA cytidine deaminase that converts cytosine bases (C) to uracils (U) and can be expressed in an activated germinal centre (Chaudhuri et al., 2003; Muramatsu et al., 2007).

1.1.4.1 Somatic hypermutation

SHM introduces mutations in and around the V region of IGH and immunoglobulin light chain genes, primarily in the CDRs (Fig. 6d) (Kocks and Rajewsky, 1989; Pech et al., 1981). These mutations are confined to the hypermutation domain, which starts about 150 nucleotides downstream of the IG promoter and extends over approximately 2 kb (Neuberger and Milstein, 1995; Rajewsky, 1996). The mutation rate has been estimated to approach 10^{-3} base pairs per generation, which is six orders of magnitude higher than that of spontaneous mutations (Allen et al., 1987; Berek et al., 1991; McKean et al., 1984). Although most mutations are substitutions, it is known that deletions occur at a frequency of approximately 4 – 7%, and duplications at a frequency of about 1% (Goossens et al., 1998). All four bases can be targeted for mutation, with C:G pairs and A:T pairs being targeted with approximately equal frequency (Neuberger and Milstein, 1995; Rajewsky, 1996).

SHM is correlated with the transcription of the IG locus and is frequently targeted to hot spots on both DNA strands (Dörner et al., 1998; Milstein et al., 1998). These hot spots occur within the RGYW motif (where R = A or G; Y = C or T; and W = A or T), which is the preferred target for the mutation machinery (Rogozin and Kolchanov, 1992). Although most major hot spots conform to the RGYW consensus, not all consensus sequences within

the V domain are mutational hot spots. This is likely dependent on the surrounding sequence and may also involve secondary structures. The nature of the substitutions is also non-random, with transitions (the exchange of a purine for a purine or of a pyrimidine for a pyrimidine) being more frequent than transversions (the exchange of a purine for a pyrimidine or vice versa) (Neuberger and Milstein, 1995).

SHM occurs in two phases: the first phase targets C:G pairs, while the second phase introduces substitutions at A:T pairs (Di Noia and Neuberger, 2007). In the first phase, AID deaminates C in C:G pairs, transforming them into U:G mismatch pairs (Muramatsu et al., 1999). AID-generated U:G lesions or the abasic sites formed by the excision of uracil through the action of uracil-DNA glycosylase (UNG) lead to transitions and transversions at C:G pairs during DNA synthesis (Di Noia and Neuberger, 2007). In the second phase, the U:G mismatch pairs are recognized by the mismatch recognition heterodimer MSH2/MSH6, triggering mutagenic patch repair that results in mutations at A:T pairs (Rada et al., 1998; Wiesendanger et al., 2000; Wilson et al., 2005). Although UNG- and MSH2/MSH6-mediated repair usually leads to error-free DNA repair, in antibody diversification, this process triggers mutagenic resolution.

1.1.4.2 Class switch recombination

CSR of IGHs is the genetic process by which a B-cell switches from producing IgM to producing other IGH isotypes (Honjo, 1983; Manis et al., 2002). CSR occurs after immunization or infection, where antigen-activated, IgM-expressing B-cells undergo a change in the C domain under the influence of cytokines secreted by T-cells and other related cells. This process involves recombinational deletion, which removes the initially expressed IGHCμ (IgM) exons and replaces them with an alternative set of downstream IGHC exons, such as IGHCγ (IgG), IGHCε (IgE), or IGHCα (IgA), without changing the rearranged V(D)J segment. As a result, CSR enables the expression of IGs with the same antigen specificity but a different IGH isotype, thereby conferring different effector functions and contributing to the organism's adaptive immune response. However, recent studies have shown that alteration in the IGH isotype can modify the conformation of the V-region antigen-binding site through the influence of the IGHC1 domain (referred to as IGHV-IGHC

conformational diversity), potentially resulting in altered antibody specificity (Fig. 6e) (McConnell and Casadevall, 2025).

Each IGHC gene, except IGHCδ (which is expressed through alternative RNA splicing), is preceded by specialized DNA sequences called switch (S) regions at the 5' end (Honjo, 1983; Manis et al., 2002). The S regions consist of highly repetitive sequences ranging from 1 to 10 kb, with G-rich non-template strands (Dunnick et al., 1993; Stavnezer, 1996). The S region repeats can be divided into two general categories: Sμ, Sα and Sε are composed of variations of pentameric sequences, whereas Sγ1, Sγ2, Sγ3 and Sγ4 mostly contain repeats that are 49 – 52 bp in length. In response to cytokines, transcription through the S regions initiates the CSR process, rendering these regions substrates for AID modification (Fig. 7) (Chaudhuri et al., 2003; Muramatsu et al., 2000; Shinkura et al., 2003). During the repair of uracil residues introduced by AID, single-stranded DNA breaks are generated and subsequently converted into DNA DSBs (Liu et al., 2024). The formation of two DSBs in different S regions leads to DNA recombination. These DSBs facilitate the switch-associated deletion of the intervening DNA region, including the IGHCμ exons (Chaudhuri and Alt, 2004). CSR is completed by NHEJ between donor and acceptor S regions, resulting in the expression of a rearranged V(D)J segment with a different downstream IGHC gene (Chi et al., 2020; Liu et al., 2024). Afterward, mature B-cells have the potential to generate any IGH isotype, and the direction of conversion is determined by the accessibility of the S region.

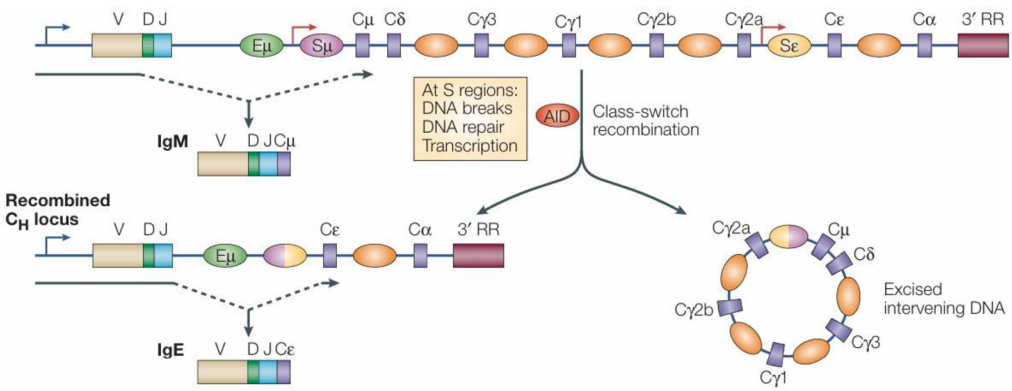


Fig. 7: Mechanism of immunoglobulin heavy chain class switch recombination.
Transcription across the immunoglobulin heavy chain (IGH) locus is initiated by a promoter located upstream of the rearranged V(D)J segment (blue arrow). Initially, the first constant (C) gene, Cμ, is transcribed and translated, leading to the production of IgM. It associates with an immunoglobulin light chain to form a

complete IgM molecule, which is displayed on the surface of a B-cell. Subsequently, secondary isotypes are produced through class switch recombination (CSR), a process that replaces the Cμ gene of the IGH with one of the downstream IGHC genes (CSR to Cε, which leads to the production of IgE, is shown). Cytokines stimulate transcription (red arrows) through the target IGHC gene and determine the immunoglobulin isotype to which the B-cell will switch. This deletional-recombination reaction requires the activation-induced cytidine deaminase (AID) and involves the generation of DNA breaks at switch (S) regions, which precede each IGHC gene, followed by DNA repair through nonhomologous end-joining (NHEJ). This leads to a rearranged IGHC locus and deletion of the intervening sequence as an episomal circle (Chaudhuri and Alt, 2004; edited).

1.2 Chronic lymphocytic leukemia

High demands on the variety of IGs, driven by constant antigenic stimulation and somatic hypermutation processes, often lead to genomic instability that has the potential to initiate the malignant transformation of B-cells. Chronic lymphocytic leukemia (CLL) represents a heterogeneous lymphoproliferative malignant disease characterised by the monoclonal expansion and progressive accumulation of mature B-lymphocytes with a typical immunophenotype in the blood, bone marrow, and secondary lymphoid tissues (Hallek et al., 2018).

1.2.1 Epidemiology

CLL is a disease primarily affecting older adults and is the most common type of adult leukemia in Western countries (Chiorazzi et al., 2021; Hallek, 2025). It generally accounts for more than 30% of all leukemia cases (Panovská et al., 2010). Approximately 0.6% of individuals will be diagnosed with CLL at some point during their lifetime (Hallek, 2025). In the United States, the incidence of CLL is 4.6 per 100,000 inhabitants per year, which is similar to the incidence in Europe at 4.9 per 100,000 inhabitants per year (Sant et al., 2010). According to an older study conducted between years 2006 and 2007, the incidence of CLL was slightly higher in the Czech Republic, ranging from 5.8 to 6.2 per 100,000 inhabitants per year (Panovská et al., 2010). While the incidence of CLL has remained stable over the past two decades, mortality rates have been continuously declining (Chiorazzi et al., 2021; Hallek, 2025). The five-year relative survival of patients with CLL has also steadily increased during this period.

CLL typically affects elderly individuals over the age of 60, with a median age at diagnosis of 70 years (Hallek, 2025). Fewer than 10% of CLL patients are younger than 45

years. Interestingly, younger patients often exhibit a more adverse clinical and molecular risk profile compared to their older counterparts, including a higher frequency of high-risk genetic aberrations and a more aggressive disease course (Parikh et al., 2014). Incidence rates also increase with age, rising to more than 30 per 100,000 inhabitants per year among individuals over the age of 80 (Eichhorst et al., 2021).

The incidence of CLL also varies by race and ethnicity, with the highest rates observed in individuals of Caucasian descent, followed by African American, Hispanic, and Asian/Pacific Islander populations (Bassig et al., 2016; Chiorazzi et al., 2021; Shenoy et al., 2011). African American patients with CLL tend to present at a younger age and with more advanced disease compared to Caucasian patients, and they generally experience poorer overall survival outcomes.

Additionally, CLL incidence also varies by gender, being approximately twice as common in males compared to females, and this gender disparity appears to be consistent across all ethnic groups (Chiorazzi et al., 2021; Hallek, 2025; Molica, 2006). Male patients with CLL tend to present with a more advanced form of the disease and typically experience a worse prognosis compared to female patients (Damle et al., 1999).

1.2.2 Diagnosis

CLL diagnosis is based on a combination of clinical, morphological, and immunophenotyping examinations (Matutes et al., 2007). In most cases, the diagnosis is established through differential blood counts, a blood smear, and immunophenotyping (Hallek et al., 2018) To confirm the diagnosis in cases with atypical morphological features or an unusual immunophenotype, additional molecular genetic analyses may be conducted. CLL is usually a slow-growing disease and is often diagnosed at an asymptomatic stage during a routine medical checkup, typically due to the detection of an elevated lymphocyte count in a complete blood count (Chiorazzi et al., 2021). Additionally, various non-specific B symptoms, such as fever, nausea, weight loss, or night sweats, may be observed in some patients (Hallek et al., 2018).

The diagnosis of CLL requires the presence of $\geq 5 \times 10^9$/l B-lymphocytes in the peripheral blood, sustained for at least three months (Hallek et al., 2018). The leukemia cells

found in the blood smear are characteristically small, mature lymphocytes with a narrow border of cytoplasm and a dense nucleus lacking discernable nucleoli and partially aggregated chromatin. These malignant B-cells co-express the surface antigen CD5 along with the B-cell antigens CD19, CD20, and CD23. The levels of mIg, CD20, and CD79b are characteristically low compared to those found on normal B-cells (Hulkkonen et al., 2002). The clonality of the circulating B-lymphocytes must be confirmed by demonstrating immunoglobulin light chain restriction using flow cytometry (Hallek et al., 2018). Clinically, CLL frequently manifests with lymphadenopathy, hepatosplenomegaly, and/or cytopenias.

If the B-lymphocyte count in peripheral blood is less than 5×10^9/l, but the cells exhibit the same immunophenotype as those seen in CLL, and lymphadenopathy or splenomegaly is present, the disease is classified as small lymphocytic lymphoma (SLL) (Eichhorst et al., 2021; Hallek et al., 2018). The diagnosis should be confirmed by histopathological examination of a lymph node. In contrast, if there is no lymphadenopathy, organomegaly, cytopenia, or clinical symptoms, and monoclonal B-lymphocytes in the blood remain below 5×10^9/l, the condition is classified as monoclonal B-cell lymphocytosis (MBL) (Marti et al., 2005). MBL represents a premalignant stage of CLL and can be detected in up to 3% of individuals with a normal blood count. Its prevalence increases with age, affecting more than 5% of adults aged 60 years or older. Progression to CLL occurs in approximately 1 – 2% of MBL cases per year (Rawstron et al., 2008).

1.2.3 Typical molecular markers

In CLL, cytogenetic and molecular markers play a crucial role in prognostic assessment and therapeutic decision-making. These markers enable the stratification of patients into distinct risk categories based on the anticipated disease course and its biological aggressiveness. Collectively, they offer essential insights not only into disease behaviour but also into the likely response of individual patients to specific treatments. Consequently, these biomarkers are central to the advancement of personalised treatment strategies in CLL, allowing clinicians to tailor therapy according to each patient's unique genetic and molecular profile (Agathangelidis et al., 2021; Hallek et al., 2018).

1.2.3.1 Immunoglobulin heavy chain gene rearrangements

Each case of CLL can be characterised by a unique set of IG gene rearrangements that encode the BCR. Since all leukemic cells originate from a single mature B-cell that has completed IG loci rearrangements, they uniformly express an identical BCR structure (Melchers et al., 2000). The predominantly monoclonal character of CLL allows the use of IG gene rearrangements as stable molecular markers, making them highly reliable for long-term disease monitoring. Among these, the SHM status of the clonotypic IGHV gene serves as a critical biomarker for prognostic stratification and prediction of therapeutic responses (Agathangelidis et al., 2022; Sutton et al., 2017). Importantly, this biomarker remains stable over time, making it highly reliable for long-term disease monitoring.

The clinical cut-off value is defined as 98% sequence identity to the closest germline IGHV gene (Ghia et al., 2007). This threshold enables a clinically relevant distinction between unmutated CLL, which shows no or limited SHM, and mutated CLL, characterised by a significant SHM load. Patients with mutated IGHV typically exhibit a more indolent disease course and better clinical outcomes, whereas those with unmutated IGHV tend to have more aggressive disease behaviour, require earlier treatment intervention and have shorter overall survival (Hamblin et al., 1999; Tobin et al., 2005).

1.2.3.2 Stereotyped BCR subsets

Although IG gene rearrangements can theoretically generate up to 10^{12} distinct IG genes, the IG repertoire observed in CLL is notably restricted, with a preferential usage of specific combinations of IGHV, IGHD, and IGHJ genes. Similarly, the CDR3 of the IGH, which is expected to exhibit high variability, displays significant sequence homology in approximately one-third of CLL cases (Agathangelidis et al., 2012). These highly similar or even identical CDR3 sequences are referred to as stereotyped, and the corresponding BCRs are classified as stereotypical. CLL cases expressing such stereotypical BCRs are grouped into distinct stereotyped subsets.

Currently, 29 major subsets have been identified, each with a frequency of at least 0.2%, collectively accounting for approximately 13.5% of all CLL cases (Agathangelidis et al., 2021). Patients belonging to the same stereotyped subsets often share distinct clinical and biological features (Bomben et al., 2009; Stamatopoulos et al., 2007). Among these,

subsets #1 and #2 are the most well-characterised aggressive variants, while subset #4 is typically associated with an indolent clinical course. Subset #8 is of particular interest due to its strong association with the highest risk of Richter's transformation (K. Stamatopoulos et al., 2017). The immunogenetic characteristics of these four paradigmatic major subsets are illustrated in Fig. 8.

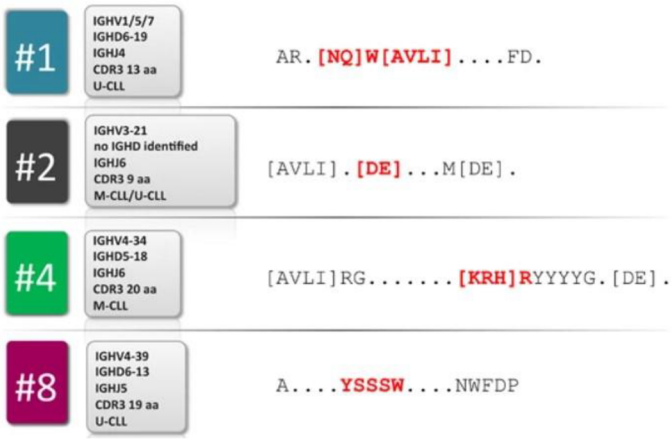


Fig. 8: Overview of the immunogenetic properties of major stereotyped subsets.

Subset #1 is characterised by the utilisation of IGHV genes from IGHV1/5/7 family, in combination with IGHD6-19 and IGHJ4. It features a CDR3 that is 13 amino acids in length and typically includes a QWL motif. Subset #2 is defined by the consistent usage of the IGHV3-21 and IGHJ6 genes; however, the rearranged IGHD gene cannot be reliably identified. The CDR3 is 9 amino acids long and is notable for the presence of an acidic residue at position 3, most commonly aspartic acid (D), and less frequently glutamic acid (E). Subset #4 is characterised by the expression of the IGHV4-34 gene paired with IGHD5-18 and IGHJ6. It has a 20 amino acid-long CDR3 containing two basic residues, either lysine (K) or arginine (R) at position 11, and arginine (R) at position 12. Subset #8 utilises the IGHV4-39 gene in combination with IGHD6-13 and IGHJ5, giving rise to a CDR3 of 19 amino acids in length. M-CLL; mutated IGHV; U-CLL, unmutated IGHV (Stamatopoulos et al., 2017).

1.2.3.3 Chromosomal aberrations and gene mutations

Leukemic cells in the majority of patients with CLL exhibit chromosomal abnormalities at some point during the disease course (Chiorazzi et al., 2021). To date, several recurrent somatic copy number variations have been identified in CLL (Fig.9) (Landau et al., 2015). Among these, approximately 80% of patients carry at least one of four common chromosomal alterations: deletions at 13q14, 11q, or 17p, and trisomy 12 (Döhner et al., 2000). These recurrent aberrations are critical for prognostication and are typically detected using fluorescence in situ hybridisation (FISH) or next-generation sequencing (NGS).

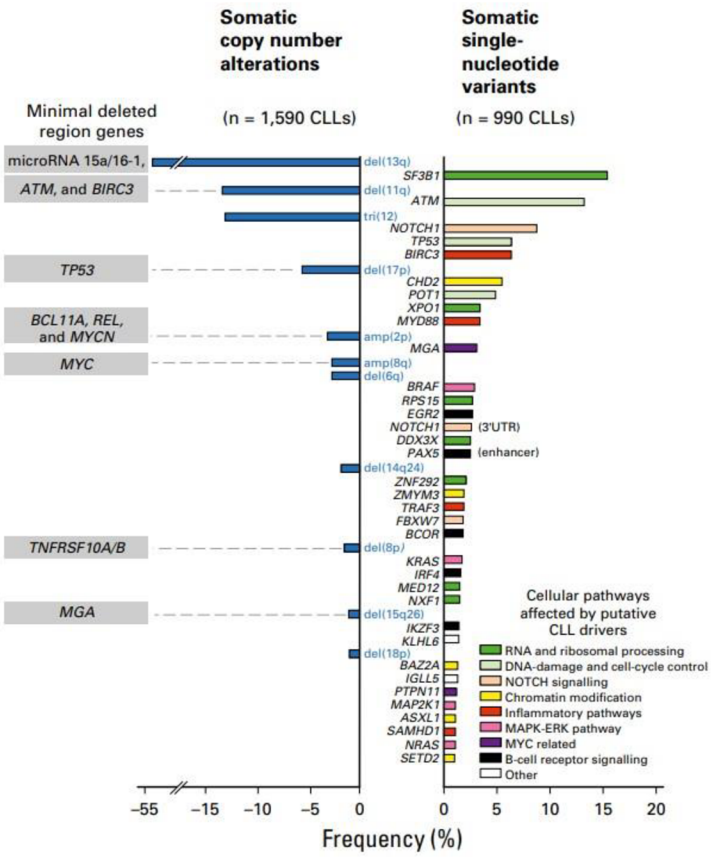


Fig. 9: Gene mutations and recurrent somatic copy number variations in CLL.
The most frequent somatic copy number alterations in CLL include deletions at 13q, 11q, and 17p, as well as trisomy 12. The most observed somatic single-nucleotide variants affect the genes *SF3B1*, *ATM*, *NOTCH1*, and *TP53* (Lazarian et al., 2017; edited).

In addition to the extensive chromosomal abnormalities described above, advances in NGS technologies have led to the identification of a range of recurrently mutated genes in CLL (Landau et al., 2015). The mutations generally affect eight key cellular signalling pathways, whose disruption provides leukemic cells with a selective advantage for survival and proliferation (Lazarian et al., 2017). The affected pathways include BCR signalling, DNA damage response and cell cycle control, Notch signalling pathway, chromatin remodelling, RNA metabolism and ribosomal processing, inflammatory signalling, the MAPK/ERK pathway, and MYC-related signalling (Fig. 9).

1.2.4 Prognosis

CLL is characterised by a highly variable clinical course, with overall survival ranging from a few months to several decades (Binet et al., 1981). While some patients remain

asymptomatic and do not require treatment for many years (Hallek and German CLL Study Group, 2005), others experience disease progression from an initially indolent phase to a more aggressive clinical phenotype. In a subset of patients, the disease presents with aggressive features requiring immediate therapeutic intervention (Chiorazzi et al., 2021).

To assess prognosis, two widely accepted clinical staging systems are commonly used: the Rai classification (Rai et al., 1975; Rai and Montserrat, 1987) and the Binet classification (Binet et al., 1981). Both systems stratify patients into three prognostic groups: low, intermediate, and high risk, each associated with distinct clinical outcomes (Hallek, 2025). These staging systems are simple, cost-effective, and rely on physical examination and basic laboratory tests. They are based on factors such as lymphocytosis, lymphadenopathy, splenomegaly and/or hepatomegaly, as well as the presence of anaemia or thrombocytopenia.

However, in recent years, numerous biological markers have been identified that provide additional prognostic information independent of clinical stage (Cramer and Hallek, 2011; Furman, 2010). Notably, many of the previously described genetic mutations and chromosomal aberrations significantly contribute to disease prognosis. Patients with unmutated IGHV are more frequently associated with high-risk genetic alterations, particularly mutations or deletions in the *TP53* gene, which further worsen prognosis (Döhner et al., 1995; Gonzalez et al., 2011; Zenz et al., 2008). The presence of a complex karyotype, defined as the presence of three or more numerical and/or structural chromosomal abnormalities, is also generally associated with a more aggressive clinical course and significantly shorter overall survival (Baliakas et al., 2014; Haferlach et al., 2007).

As a result, several prognostic scoring systems have been developed that integrate clinical parameters with biological and genetic data to improve risk assessment (Pflug et al., 2014; Shanafelt et al., 2009). Among the most widely adopted is the CLL International Prognostic Index (CLL-IPI), which incorporates a weighted score of five independent prognostic factors: clinical stage according to Rai, patient age, IGHV mutational status, serum β_2-microglobulin concentration, and the presence of del17p and/or *TP53* mutations (International CLL IPI Working Group, 2016).

1.3 Multiclonal chronic lymphocytic leukemia

In a subset of CLL patients, multiple (most commonly double) productive IGH gene rearrangements can be detected. This phenomenon has been independently reported by several research groups and is estimated to occur in approximately 2 – 3% of all CLL cases (Kriangkum et al., 2015; Langerak et al., 2011; Plevova et al., 2014; Visco et al., 2013). Two main mechanisms have been proposed to explain the presence of double productive IGH rearrangements. One possibility is the lack of allelic exclusion at the IGH loci, resulting in a single clonal population expressing two distinct productive IGH gene rearrangements (Rassenti and Kipps, 1997). Alternatively, this finding may reflect the coexistence of two separate clonal B-cell populations within the same patient (Chang and Cerny, 2006; González-Campos et al., 1997; Hsi et al., 2000; Plevova et al., 2014; Sanchez et al., 2003).

In some CLL cases, there is evidence supporting the existence of a single clonal B-cell population expressing two distinct productive IGH gene rearrangements. The underlying molecular mechanisms that could lead to such dual expression involve a lack of allelic exclusion at the IGH loci (Rassenti and Kipps, 1997) or IGHV gene replacement by another IGHV segment (Stamatopoulos et al., 1996). These phenomena have been interpreted as indicative of receptor editing, a process also observed in normal B-cell development. In physiological settings, receptor editing allows B-cells expressing potentially autoreactive BCRs to undergo secondary rearrangements and escape clonal deletion (Nemazee, 2006). A similar mechanism in CLL may provide leukemic clones with a selective survival advantage. Nevertheless, considering current data and recent studies, both IGHV replacement and lack of allelic exclusion appear to be very rare events in CLL (Brazdilova et al., 2018; Kriangkum et al., 2015; Stamatopoulos et al., 2017).

The presence of two distinct clones represents a more probable explanation for the occurrence of double productive IGH rearrangements in CLL. These cases may reflect the coexistence of two clonal B-cell populations, each expressing a distinct BCR, in a single patient. Several studies have documented the presence of concurrent CLL clones within a single patient, challenging the traditional view of CLL as a monoclonal disease (Brazdilova et al., 2018; Chang and Cerny, 2006; González-Campos et al., 1997; Hsi et al., 2000; Plevova et al., 2014; Sanchez et al., 2003). Alternatively, the detection of double productive IGH

gene rearrangements might reflect the coexistence of CLL with another B-cell lymphoproliferative disorder (Sanchez et al., 2003).

The coexistence of multiple productive IGH rearrangements in CLL makes a significant challenge for prognostication, particularly when the co-detected rearrangements display discordant IGHV mutational status (one rearrangement is mutated while the other is unmutated) (Langerak et al., 2011; Plevova et al., 2014; Rosenquist et al., 2017). These discordant cases complicate risk stratification, as IGHV mutational status is a key determinant of clinical outcome.

Studies have shown that approximately two-thirds of multiple productive IGH rearrangements in CLL cases exhibit concordant IGHV mutational status (either both mutated or both unmutated), with a frequency higher than expected by chance (Brazdilova et al., 2018; B. Stamatopoulos et al., 2017; Stranska et al., 2020; Visco et al., 2013). These concordant cases generally follow a clinical course similar to that of patients with a single corresponding IGHV status. In contrast, the remaining discordant cases are associated with more aggressive clinical behaviour, adverse prognostic features, earlier need for treatment, and reduced survival, closely resembling the high-risk profile typical of unmutated IGHV CLL (Visco et al., 2013).

In multiclonal CLL, clonal drift has been observed as a dynamic process characterised by shifts in the proportions of malignant B-cell clones over time, likely driven by overall disease progression or response to treatment. In most multiclonal CLL cases, the dominant clone progressively displaced minor subclones or clonal ratios remained stable over time (Kriangkum et al., 2015; Plevova et al., 2014; B. Stamatopoulos et al., 2017). However, in some cases, a previously minor or even initially undetectable clone, defined by a distinct productive IGH rearrangement, has overgrown the major one and become the dominant (Plevova et al., 2014; Rose-Zerilli et al., 2016).

The disappearance of a clone tends to occur more frequently when it coexists with a more aggressive counterpart. Plevova et al. (2014) reported that clones persisting or emerging over time more often exhibited higher IGHV identity and/or longer IGH CDR3 regions, suggesting a selective advantage. Genomic abnormalities may further influence clonal dynamics. For example, the presence of a *TP53* defect in a clone is associated with

increased clonal fitness and may contribute to clonal drift, underscoring the aggressive nature of p53-defective subpopulations. The impact of other genetic alterations, such as del11q, also warrants consideration. Notably, in discordant multiclonal CLL cases, clonal drift frequently results in the eventual loss of the mutated IGH clone, leading to a shift in the IGHV status classification toward the unmutated (high-risk) profile.

Patients with monoclonal and multiclonal CLL generally exhibit similar clinical and biological characteristics. However, some differences have been reported in specific studies. According to Sanchez et al. (2003), patients with multiclonal CLL had a significantly higher incidence of splenomegaly, lower white blood cell, absolute lymphocyte counts in peripheral blood, and a higher proportion of cases requiring early treatment compared to monoclonal CLL patients. While the incidence of death was slightly higher in the multiclonal group, this difference was not statistically significant.

In contrast, Stranska et al. (2020) did not observe significant differences in clinical or laboratory parameters between oligoclonal and monoclonal CLL cases, except for a significantly higher age at diagnosis in the oligoclonal group. Interestingly, patients with oligoclonal CLL more frequently reported comorbid conditions such as hypercholesterolemia, hypertension, and herpes simplex infection, as well as a higher incidence of tumours (including CLL) in first-degree relatives, suggesting a potential genetic or familial predisposition.

1.4 Methods of immunoglobulin gene rearrangements detection

Detecting IG gene rearrangements is essential in the diagnosis and characterization of B-cell lymphoproliferative disorders, as each B-cell clone creates a unique V(D)J recombination signature that serves as a molecular marker. Several methodologies have been developed to identify these rearrangements, each offering distinct advantages and limitations in terms of sensitivity, specificity, throughput, and applicability to various clinical and research settings.

Sanger sequencing remains the standard clinical method for assessing IG gene rearrangements in CLL due to its simplicity, speed, cost-effectiveness, and high accuracy, with a very low base-calling error rate and straightforward data interpretation that is highly standardized across laboratories (Davi et al., 2020). Commonly used in combination with

polymerase chain reaction (PCR), it enables the precise sequencing of IGHV-IGHD-IGHJ rearrangements, which is essential for determining SHM status (Hotinski et al., 2021). As a monoplex technique, Sanger sequencing captures the most prevalent rearrangement within a sample, providing accurate and unambiguous results in the majority of CLL cases (Davi et al., 2020).

Even though Sanger sequencing has been optimized and standardized to a high level, it may occasionally provide suboptimal and even misleading results (Langerak et al., 2017). In approximately 3 – 4% of patients, Sanger-based analysis may fail entirely or produce uninterpretable results, such as when only a single unproductive rearrangement or dual productive rearrangements are detected (Langerak et al., 2011). Moreover, its low-throughput nature limits its utility in detecting minor clonal populations at low frequencies. The inability of Sanger sequencing to fully resolve the subclonal architecture and intraclonal diversity, particularly in cases of multiple productive IGH rearrangements, poses challenges for comprehensive prognostication and risk stratification in CLL (Davi et al., 2020).

In contrast to Sanger sequencing, NGS enables deep profiling of the B-cell repertoire with high sensitivity and throughput (Davi et al., 2020). It allows the detection of both major and minor clonotypes, facilitating the detailed visualization of subclonal architecture, clonal diversity, and temporal clonal dynamics such as clonal drift and evolution. This enables the identification of independent, unrelated clonotypes that would be missed by conventional methods due to their low abundance or inefficient amplification. Furthermore, NGS allows quantitative assessment of clone frequencies, providing insight into the biological heterogeneity of CLL, which may improve prognosis assessment and treatment strategies (Hotinski et al., 2021).

However, the effective implementation of NGS for immunogenetic analysis requires substantial technical expertise, particularly in assay design and data interpretation (Davi et al., 2020). The process critically depends on robust bioinformatics pipelines capable of distinguishing true clonal rearrangements from polyclonal background noise and managing the large volume of sequencing data. Interpretation remains complex and is not yet fully

standardized across laboratories. There are currently no universally accepted definitions regarding clonal size, numerical cut-off values, or limits of detection (Langerak et al., 2017).

Amplicon-based NGS-based clonality assessment still typically relies on an initial multiplex PCR step, which can be skewed by the presence of polymorphisms and, in particular, SHM that impairs proper primer annealing (Langerak et al., 2017). Moreover, variations in amplification efficiency across different rearrangements can lead to unequal representation, thereby compromising accurate quantification. Amplification biases, resulting from differential primer performance or misannealing, can introduce artificial distortions in IGH gene frequencies, potentially misrepresenting repertoire or skewing clonality status. Consequently, determining clone representation and extrapolating sequencing read frequencies to actual cellular clone sizes requires caution. Nevertheless, the multiplexing capability of NGS, sequencing millions of fragments across multiple targets in parallel, makes it an effective alternative to traditional methods. Its capacity to provide both biological insights and clinically relevant information underscores its potential to improve the understanding and management of CLL (Hotinski et al., 2021).

An alternative approach is to use a hybridization capture-based NGS, which enriches IG loci DNA using probes, thereby enabling a potentially less biased and more comprehensive sequencing of all target regions (Langerak et al., 2017). It can enable simultaneous assessment of IGH gene rearrangements, IGHV SHM status, and mutational profiling of a panel of clinically relevant genes within a single sequencing run (Davi et al., 2020). This integrated approach is particularly advantageous for improving diagnostic precision and is highly meaningful for prognostication and treatment decision-making in CLL. The Lymphoid NeXt-Generation Sequencing (LYNX) panel used in this thesis facilitates broad testing for genetic alterations across common lymphoid neoplasms (Navrkalova et al., 2021). It allows the identification of diverse prognostic and predictive biomarkers within a single assay, an essential advancement toward the effective and personalized management of patients with hematologic malignancies.

2 Aims of the thesis

In a subset of CLL patients, multiple productive IGH gene rearrangements can be detected, reflecting the coexistence of distinct clonal B-cell populations, each expressing a different IG molecule on its surface, within a single patient. The presence of multiple leukemic clones poses a risk for adverse clonal evolution, as each clone may harbour distinct genomic aberrations with potentially divergent biological behaviour. To further investigate this phenomenon, we analysed samples from patients with a history of multiple productive IGH rearrangements. These samples were analysed using three distinct methodologies: Sanger sequencing, capture-based NGS using the LYNX panel, and amplicon-based NGS of IGH gene rearrangements (IG aNGS).

Thus, in the selected cohort, our aims were:

1. To determine the prevalence and diversity of IGH genes in patients with multiclonal CLL and compare it with a group of patients with monoclonal CLL.
2. To attempt the quantitative comparison of the repertoire of IGH rearrangements and resulting clonality depending on the detection method (Sanger sequencing, LYNX panel, and IG aNGS).
3. To describe changes in the IGH repertoire in serial samples over time, assessing clonal dynamics and stability.

3 Material and methods

3.1 Material

Peripheral blood samples from patients diagnosed with CLL, who are monitored at the Department of Internal Medicine, Hematology and Oncology at the University Hospital Brno, were used as the primary material for all experiments. As part of routine diagnostics at the Centre of Molecular Biology and Genetics, University Hospital Brno, the number and type of IGH rearrangements, along with the IGHV mutational status, were determined for all samples via Sanger sequencing (acknowledging the key contribution of Mgr. Hana Skuhrová Francová). For this master's thesis, 76 CLL patients were selected, all of whom had multiple productive IGH rearrangements at a certain disease stage identified between the years 2007 – 2023. All patients provided informed consent for the use of their biological material in the research, under protocols approved by the Ethical Committee of the University Hospital Brno.

In the selected group of patients, we analysed 98 CLL samples (one sample from 58 patients, two serial samples from 18 patients, and three serial samples from two patients). The mean time between the first and last examination for patients with serial samples was 5.3 years, with a median of 4.5 years. The first sample from each patient always corresponded to the time before the patient's treatment began. Of the 76 patients, 30 patients (38%) were already in disease progression at the time of the first sample collection.

3.2 Methods

3.2.1 Separation of B-lymphocytes from peripheral blood

Peripheral blood samples were collected in tubes containing ethylenediaminetetraacetic acid (EDTA) as an anticoagulant. The samples were processed immediately upon arrival at the Centre of Molecular Biology and Genetics, University Hospital Brno, usually within 4 hours, to prevent degradation of biological material. B-lymphocytes were separated from peripheral blood using Ficoll-Paque™ PLUS (GE HealthCare, USA) density gradient centrifugation. The blood was mixed with a cocktail of specific antibodies – 50 µl of

RosetteSep™ Human B Cell Enrichment Cocktail and 25 µl of RosetteSep™ Human CD3+ Depletion Cocktail (STEMCELL Technologies, Canada) per 1 ml of blood, to deplete non-B cells. Non-B cells were crosslinked to red blood cells, forming dense aggregates called immunorosettes, which settled into the lower layer of the density gradient after centrifugation. Purified B-lymphocytes were located at the interface between the blood plasma and the density gradient medium as a buffy coat (Fig. 10), from where they could be easily removed for future processing.

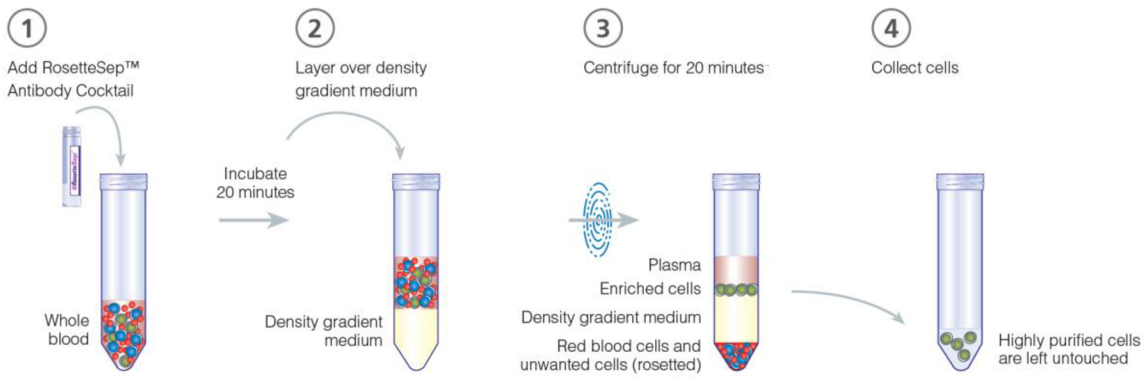


Fig. 10: The Typical RosetteSep™ Cell Enrichment Protocol.
With this immunodensity procedure, it is possible to isolate untouched cells directly from whole blood. By crosslinking unwanted cells to red blood cells present in the sample using specific antibodies, the target cells are purified during standard density gradient centrifugation (https://www.stemcell.com/products/brands/rosettesep-immunodensity-cell-separation.html#protocol).

The suspension of purified B-lymphocytes was dissolved in 2% FBS/PBS (fetal bovine serum/ phosphate-buffered saline) in a 1:1 ratio. After centrifugation, the pellet was resuspended in cold 1x lysis buffer (155 mM NH_4Cl; 10 mM $KHCO_3$; 0.1 mM EDTA; H_2O) (Hospital Pharmacy, University Hospital Brno, Czech Republic). Following another round of centrifugation, the pellet was resuspended in 2% FBS/PBS to create a cell suspension. An appropriate amount of suspension was taken for RNA and DNA isolation (see subsequent chapters). The remaining suspension was centrifuged, and the pellet was resuspended in fresh 10% dimethyl sulfoxide (DMSO) in FBS. The separated cells were stored in a freezing container filled with isopropyl alcohol and initially frozen at -80 °C. The cells were cooled at a rate close to -1 °C per minute. Later, the cells were transferred to liquid nitrogen for long-term storage.

3.2.2 RNA isolation and Sanger sequencing

RNA was isolated from approximately 10×10^6 cells for the determination of IGH rearrangements. The cell pellet was resuspended in 600 µl of RLT buffer (Qiagen, Germany) containing beta-mercaptoethanol. RNA was then purified using the RNeasy® Kit (Qiagen) with the QIAcube® Connect instrument (Qiagen), following the instructions by the manufacturer. The RNA transcript was reverse transcribed into complementary DNA (cDNA), amplified by PCR, and sequenced via Sanger sequencing.

3.2.3 DNA isolation and quality control

Genomic DNA (gDNA) was isolated from approximately 5×10^6 cells resuspended in 400 µl PBS at room temperature. For isolation, the DNeasy® Blood & Tissue Kit (Qiagen) with the QIAcube® Connect, or the MagCore® Genomic DNA Whole Blood Kit (RBC Bioscience, Taiwan) with either the MagCore® HF16 Automated Nucleic Acid Extractor (RBC Bioscience) or the MagCore® Plus II Automated Nucleic Acid Extractor (RBC Bioscience) were used. Both kits use proteinase K for cell lysis and protein degradation. DNA is then eluted in low salt elution buffer or water after the removal of contaminants. The isolated gDNA in our experiments was solubilized in 200 µl 1x Low Tris-EDTA buffer (TE buffer) (10 mM Tris-HCl, 0.1mM EDTA, pH 8.0,) and stored at -20 °C. gDNA isolation followed the instructions by the manufacturer.

The quality and quantity of the isolated gDNA were measured using the NanoDrop ND-1000 Spectrophotometer (Thermo Fisher Scientific, USA). For more accurate values, DNA concentration was also measured on the Qubit® 2.0 Fluorometer (Invitrogen, USA) with the Qubit™ dsDNA BR (Broad Range) Assay Kit (Thermo Fisher Scientific), according to the manufacturer's instructions.

3.2.4 LYmphoid NeXt-Generation Sequencing Panel

The LYNX panel is an integrative, comprehensive, capture-based NGS gene panel designed for the analysis of standard prognostic and predictive markers in common lymphoid neoplasms, such as CLL, acute lymphoblastic leukemia (ALL), diffuse large B-cell lymphoma (DLBCL), follicular lymphoma (FL), mantle cell lymphoma (MCL), Burkitt

lymphoma (BL), Waldenström's macroglobulinemia (WM), and multiple myeloma (MM). The current list of analysed markers is based on a later, unpublished revision of Navrkalova et al. (2021) and can be seen in Fig. 11. It is applicable to standard biological materials, including peripheral blood, bone marrow, fresh frozen tissue, and formalin-fixed, paraffin-embedded blocks (FFPE). The LYNX panel provides accurate detection of mutations in all coding exons and splice sites of 72 lymphoma-related genes, reliable identification of large genome-wide (6 Mb) and chromosomal aberrations (300 kb) recurrent in lymphoid malignancies, the assessment of immunoglobulin and T-cell receptor gene rearrangements, and detection of lymphoma-specific translocations. It represents a versatile sequencing tool for routine testing of patients with lymphoproliferative diseases and can be used in clinical settings with research overlap.

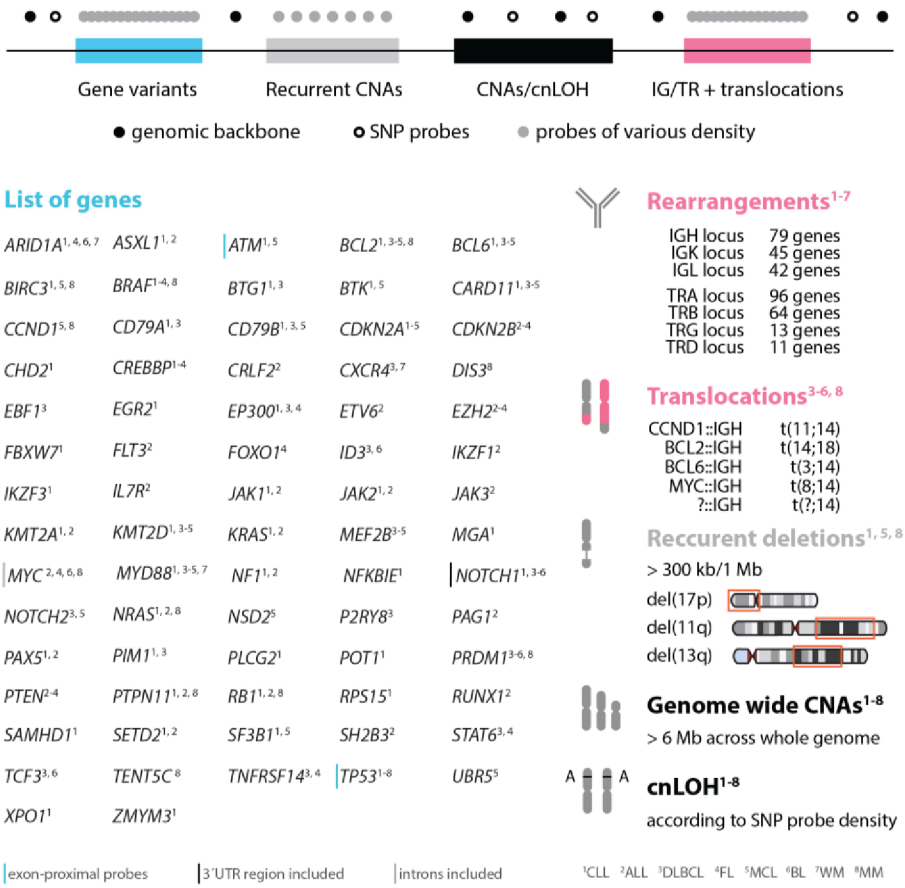


Fig. 11: Illustration of genomic targets and molecular markers in the LYNX panel.
The LYNX panel design for the most common lymphoid malignancies ([1]CLL – chronic lymphocytic leukemia, [2]ALL – acute lymphoblastic leukemia, [3]DLBCL – diffuse large B-cell lymphoma, [4]FL – follicular lymphoma,

[5]MCL – mantle cell lymphoma, [6]BL – Burkitt lymphoma, [7]WM – Waldenström's macroglobulinemia, [8]MM – multiple myeloma). cnLOH, copy neutral loss of heterozygosity; CNV, copy number variant; SNP, single-nucleotide polymorphism (Navrkalova et al., 2021; updated).

Library preparation was performed according to the SureSelect XT HS2 DNA System protocol (Agilent Technologies, USA). Briefly, 200 ng of gDNA was mechanically fragmented by ultrasonication with the Covaris S220 Focused-ultrasonicator (Covaris, UK) to fragments 180 – 250 bp in length. The SureSelect XT HS2 Library Preparation Kit for ILM (Pre PCR) (Agilent Technologies) was used for DNA end repair and ligation of molecular-barcoded adaptors. These steps were performed on the Biometra T Professional Trio Thermocycler (Analytik Jena, Germany). After purification with 80 µl SPRIselect Beads (Beckman Coulter, USA), the adaptor-ligated library was amplified with sample-specific indexes using the SureSelect XT HS2 Index Primer Pairs 97-192 for ILM (Pre PCR) (Agilent Technologies). Pre-capture PCR amplification was performed on the Bio-Rad C1000 Touch™ Thermal Cycler (Bio-Rad, Czech Republic). After purifying the amplified library with 50 µl SPRIselect Beads, the quality of the pre-capture library was assessed using the 2200 TapeStation (Agilent Technologies) with D1000 Reagents and ScreenTape (Agilent Technologies). The quantity of the amplified library was measured with the Qubit™ dsDNA BR Assay Kit on the Quantus™ Fluorometer (Promega, USA).

Subsequently, the DNA libraries from four samples were pooled in equimolar amounts to create the pre-capture pool, with the final amount reaching 1500 ng of DNA in nuclease-free water, resulting in a total volume of 12 µl. If the pre-capture pool had a volume greater than 12 µl, it was concentrated using the Concentrator Plus (Eppendorf, Germany). The prepared gDNA libraries were hybridized to a target-specific probes using the SureSelect XT HS2 Target Enrichment Kit ILM Hyb Module, Box 2 (Post PCR), and the SureSelect Custom Tier2 0.5Mb-2.9Mb Probe, Design ID: 3465851 (both Agilent Technologies) during probe-specific hybridization conditions on the Bio-Rad C1000 Touch™ Thermal Cycler. After hybridization, the targeted molecules were captured on 50 µl Dynabeads™ MyOne™ Streptavidin T1 (Invitrogen), streptavidin-coated magnetic beads. The DNA libraries were mixed vigorously at 1650 rpm for 30 minutes at 23 °C using the ThermoMixer® C (Eppendorf) and then washed with buffers from the SureSelect Target Enrichment Kit ILM Hyb Module, Box 1 (Post PCR) (Agilent Technologies). The SureSelect-enriched DNA

libraries, retained on the streptavidin beads, were amplified during the post-capture PCR amplification on the Bio-Rad C1000 Touch™ Thermal Cycler. Instead of the SureSelect Post-Capture Primer Mix, the AC-PCR Primer from the MGIEasy Universal Library Conversion Kit (App-A) (MGI, China) was used. After purifying the amplified captured libraries with 50 µl SPRIselect Beads, the quality of the post-capture libraries was assessed using the 2200 TapeStation with High Sensitivity D1000 Reagents and ScreenTape (Agilent Technologies). The quantity of the sequencing libraries was measured with the Qubit™ dsDNA HS (High Sensitivity) Assay Kit (Thermo Fisher Scientific) on the Quantus™ Fluorometer.

3.2.5 MGIEasy Universal Library Conversion

The MGIEasy Universal Library Conversion is designed to convert a linear dsDNA library into a circular ssDNA library and is optimized for MGI sequencing platforms. The MGIEasy Universal Library Conversion Kit (App-A) is used to convert dsDNA linear libraries prepared with Next Generation Sequencing Library Preparation Kits from providers other than MGI. The total DNA input should be 1 pmol according to insert size, with a total volume ≤ 48 µl.

The linear dsDNA libraries of index-tagged samples, each 300 bp in size, were pooled together in equimolar amounts to create the post-capture pool, with the final amount reaching 198 ng of DNA in Low TE buffer, resulting in a final volume of 48 µl. If the post-capture pool had a volume greater than 48 µl, it was concentrated using SPRIselect Beads at a 1.2:1 bead-to-sample volume ratio. The quantity of the pool was measured with the Qubit™ dsDNA HS Assay Kit on the Quantus™ Fluorometer. The post-capture pool was then denatured, circularized, and enzymatically digested according to the MGIEasy Universal Library Conversion Kit (App-A) User Manual (MGI). The enzymatic digestion product was purified using SPRIselect Beads and its quality was assessed using the QuantiFluor® ssDNA System (Promega) on the Quantus™ Fluorometer. The final yield should be at least 60 fmol, corresponding to 5.94 ng of ssDNA with a size of 300 bp.

3.2.6 Next-Generation Sequencing

Sequencing of the prepared libraries was performed on the DNBSEQ-G400 machine (MGI). The LYNX libraries were sequenced using either the MGI 1000016982 DNBSEQ-G400RS High-throughput Rapid Sequencing Set (FCS PE150) RUO or the MGI 1000016952 DNBSEQ-G400RS High-throughput Rapid Sequencing Set (FCL PE150) RUO, with a read length of 2×150 bp. Libraries were loaded onto either the FCS (small) flow cell, with approximately 500 million effective reads, or the FCL (large) flow cell, with approximately 1700 million effective reads. Each sample in our dataset was sequenced with approximately 12.5 million reads. The IG aNGS libraries were sequenced using the MGI 940-000152-00 DNBSEQ-G400RS High-throughput Rapid Sequencing Set (FCS PE300) RUO, together with the MGI 1000027472 High-throughput Sequencing Primer Kit (App-C), with a read length 2×300 bp. Sequencing was conducted by the Core Facility Genomics at the Central European Institute of Technology, Masaryk University, Brno.

MGI's DNBSEQ™ technology uses ssDNA circularization and rolling circle replication (RCR) to generate DNA nanoballs (DNBs) directly from the prepared library. The selected dsDNA library, with adapter sequences at the terminal ends, is heated to denature and generate ssDNA. A splint oligonucleotide is hybridized to both ssDNA ends, forming a nicked circle. The nick is then repaired using DNA ligase to form a stable, single-stranded circular template. Subsequently, the circularized ssDNA library is amplified through RCR. RCR produces a long strand of amplified DNA (concatemer), containing 100 to 1000 identical copies of DNA fragments. The concatemer folds into compact, spherical DNBs approximately 200 nm in diameter. The DNBs are accurately loaded onto Patterned Array flow cells, where each binding site is large enough to bind only one single DNB (Fig. 12). The distance between active spots on the chip surface is uniform, ensuring no interference between the fluorescence signals from neighbouring DNBs during laser excitation and subsequent imaging with advanced cameras.

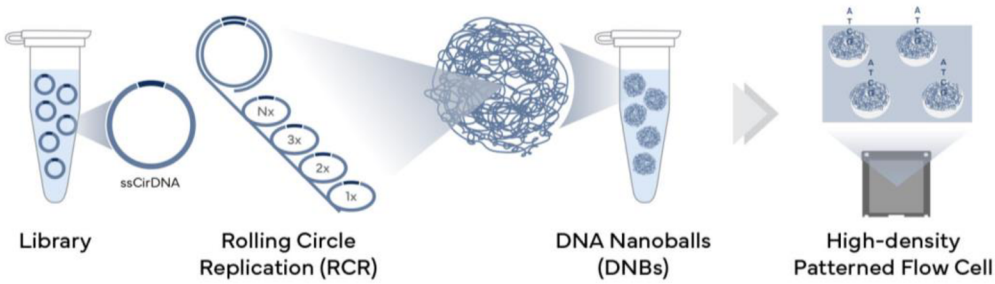


Fig. 12: The DNBSEQ™ technology developed by MGI.

This technology uses ssDNA circularization of dsDNA libraries, followed by their amplification through rolling circle replication (RCR) to form spherical DNA nanoballs (DNBs). The compact DNBs are then loaded onto Patterned Array flow cells, and the sequence is read using four fluorescent probes, each recognizing a specific DNA base (https://www.completegenomics.com/dnbseq-technology/).

3.2.7 Bioinformatic analysis of immunoglobulin rearrangements

The resulting sequences from the LYNX panel were analysed using the LYNX analytical software (https://lynx.ceitec.muni.cz/). These sequences were compared to known germline sequences collected in the immunogenetic database IMGT® (https://www.imgt.org/). A detailed molecular characterization of the detected IG rearrangements was performed with IMGT/V-QUEST (https://www.imgt.org/IMGT_vquest/input). We analysed the productive rearrangements of IGH (VJ:Vh-(Dh)-Jh) and the IGHV mutational status was determined only when a complete, productive IGHV sequence was available. The cut-off value for IGH rearrangement detection was set to three sequence fragments assigned to an event and 1% event frequency. Finally, we calculated the percentage proportions of different IGH rearrangements, representing different cell clones, above the cut-off.

LYNX results were compared to the results of Sanger sequencing done at the Centre of Molecular Biology and Genetics, University Hospital Brno, and the results of IG aNGS done by Mgr. Kristýna Taušová, Institute of Medical Genetics and Genomics, Faculty of Medicine, Masaryk University, Brno, for all the samples. The sequences of IG aNGS were analysed using the ARResT/Interrogate Software (https://arrest.tools/interrogate/). We evaluated only productive IGH rearrangements (VJ:Vh-(Dh)-Jh), with a cut-off of 1% of total reads per IGH rearrangement. Finally, we also calculated the percentage proportions of different IGH rearrangements above the cut-off and compared them with the results from the LYNX analysis.

4 Results

In this master's thesis, we analysed 98 CLL samples from 76 patients, each of whom had a history of multiple productive IGH rearrangements detected in diagnostic work-up using Sanger sequencing. These samples were additionally analysed using two distinct methods: capture-based NGS using the LYNX panel and amplicon-based IG aNGS. The results obtained from each method were evaluated separately and subsequently compared with those from a cohort of monoclonal CLL cases (as assessed using Sanger sequencing) monitored at the Department of Internal Medicine, Hematology and Oncology at the University Hospital Brno, as well as compared across all three mentioned methods. A comprehensive overview of the IGHV, IGHD, and IGHJ families and their corresponding genes detected by all methods is provided in Appendices A, B, and C. A detailed list of all detected productive IGH rearrangements is provided in Appendix D.

4.1 Sanger sequencing

Using Sanger sequencing, we analysed 94 out of 98 samples. Results for the remaining 4 samples were not available, however, data for these samples were available from earlier time points, collected between 3 and 11 years prior. In this cohort, 163 unique clonotypes were observed out of a total of 182 detected. The remaining 19 clonotypes were repeated in other samples from the same patients. Among the 94 analysed samples, the presence of multiple productive IGH rearrangements was confirmed in 78 cases (83%), while 16 samples were monoclonal (18%). All monoclonal samples originated from patients who had multiple productive IGH rearrangements. These were detected either in earlier samples or, more commonly, in later ones. Among the multiclonal samples, biclonal cases were the most frequent (69/78; 88.5%), followed by triclonal (8/78; 10.3%) and a single tetraclonal sample (1/78; 1.3%). The clonal composition of all samples is shown in Fig. 13.

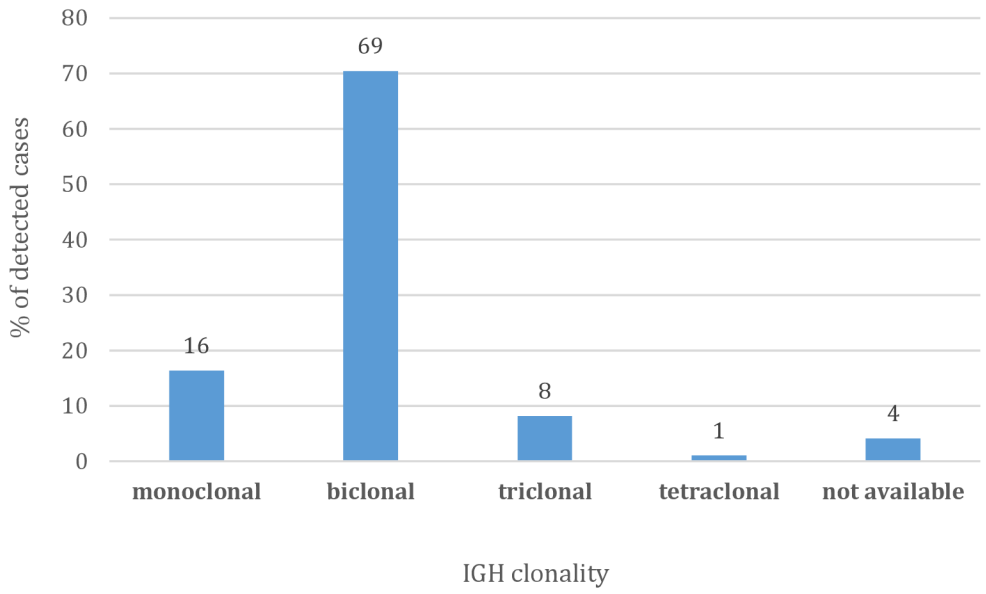


Fig. 13: Clonal composition of IGH rearrangements analysed by Sanger sequencing.
The number above the column represents the total number of samples corresponding the selected IGH clonality.

For further analysis, we selected one multiclonal sample from each patient and examined the IGH repertoire among 163 productive IGH rearrangements. Most of the IGHV genes originated from the IGHV3 family, observed in 59 of 162 cases (36.4%). This was followed by genes from the IGHV1 family (43 cases; 26.5%) and the IGHV4 family (35 cases; 21.6%), as shown in Fig. 14. The least frequently observed were genes from the IGHV7 family (1 case; 0.6%), the IGHV6 family (5 cases; 3.1%), and the IGHV2 family (6 cases; 3.7%).

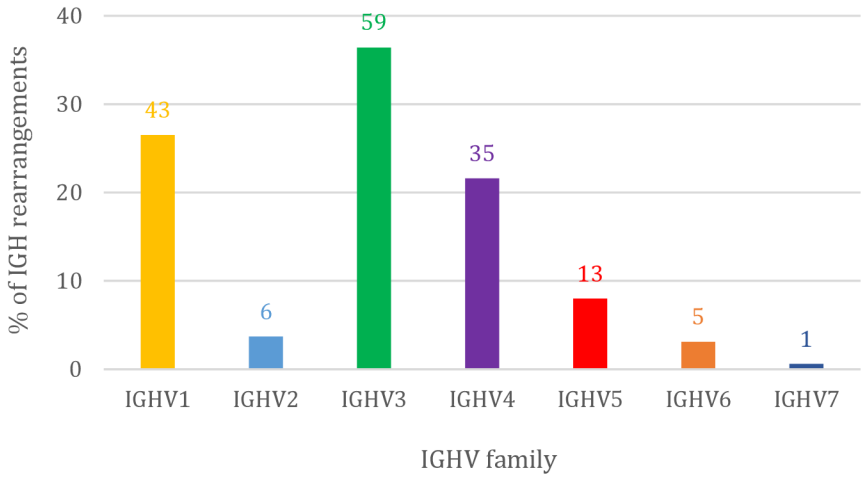


Fig. 14: Distribution of IGHV families analysed by Sanger sequencing.
The number above the column represents the total number of detected IGHV genes in the selected IGHV family.

The most frequent individual IGHV gene overall was IGHV1-69 from the IGHV1 family (21/162; 13%). The second most frequent was IGHV3-23 (15/162; 9.3%), followed by IGHV4-34 (12/162; 7.4%) and IGHV3-7 (10/162; 6.2%). The distribution of all individual IGHV genes among productive IGH rearrangements is illustrated in Fig. 15.

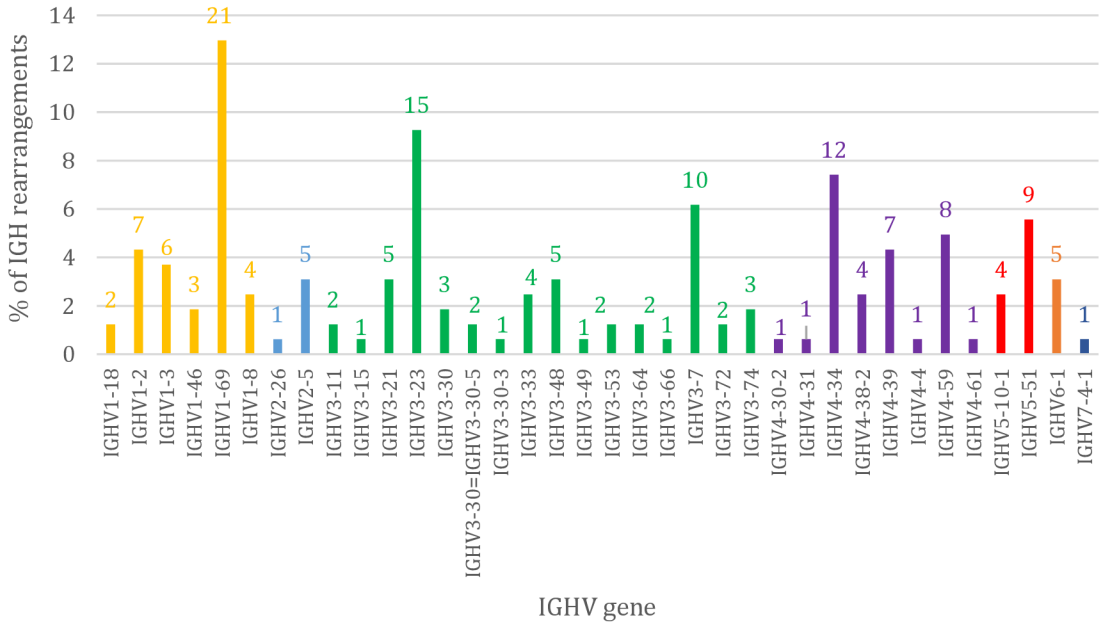


Fig. 15: Distribution of IGHV genes analysed by Sanger sequencing.
The number above the column represents the total number of detected IGHV genes in the selected type of IGHV gene.

In this cohort of multiclonal CLL patients, we also analysed IGHV mutational status using the 98% germline identity cut-off. The distribution of mutational status was balanced, with a slightly higher proportion of unmutated rearrangements (82/162; 50.6%) compared to mutated rearrangements (80/162; 49.4%). The majority of IGH rearrangements showed concordant IGHV mutational status (either mutated or unmutated) with other IGH rearrangement(s) within the same patient (53 out of 76 patients; 69.7%). Discordant cases with coexisting mutated and unmutated IGH rearrangements were observed in 23 out of 76 patients (30.3%). A schematic overview of the IGHV mutational status distribution is shown in Fig.16.

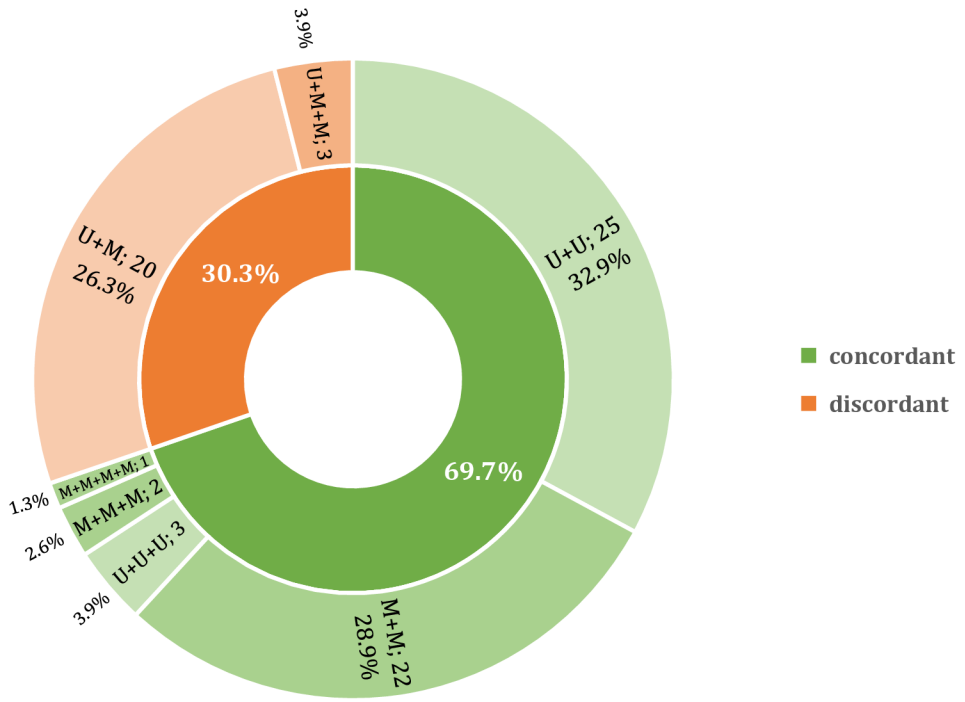


Fig. 16: Distribution of IGHV mutational status and concordance analysed by Sanger sequencing.

The majority of IGHD genes originated from the IGHD3 family, observed in 61 of 162 cases (37.7%). The next most common genes were from the IGHD6 family (35 cases; 21.6%) and the IGHD2 family (32 cases; 19.8%), as shown in Fig. 17. The least frequently observed were genes from the IGHD7 family (1 case; 0.6%).

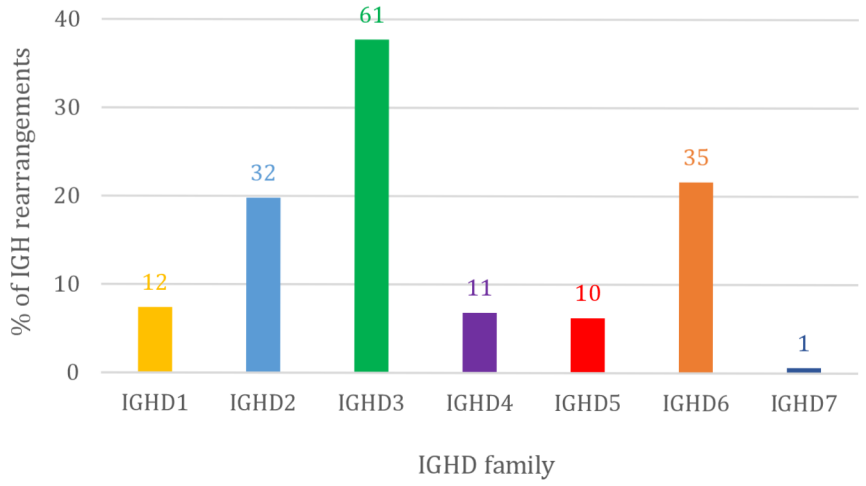


Fig. 17: Distribution of IGHD families analysed by Sanger sequencing.
The number above the column represents the total number of detected IGHD genes in the selected IGHD family.

The most frequent IGHD gene was IGHD3-22 (20/162; 12.4%) followed by IGHD3-3 (19/162; 11.7%), IGHD6-19 (17/162; 10.5%), IGHD6-13 (15/162; 9.3%), and IGHD2-2 (14/162; 8.6%). The distribution of all individual IGHD genes among productive IGH rearrangements is shown in Fig. 18.

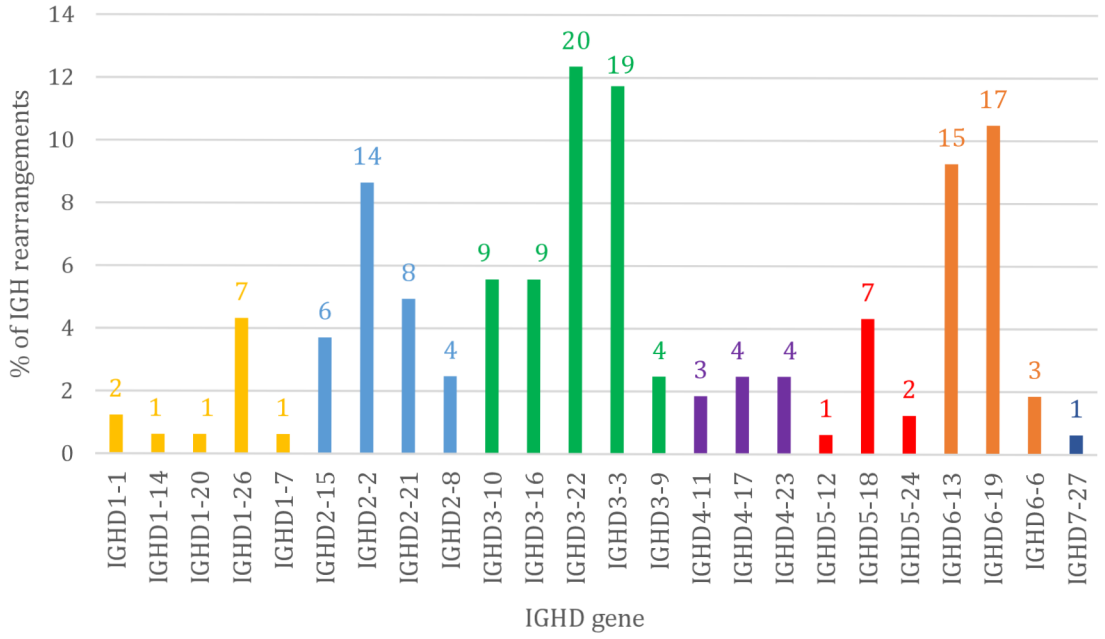


Fig. 18: Distribution of IGHD genes analysed by Sanger sequencing.
The number above the column represents the total number of detected IGHD genes in the selected type of IGHD gene.

Most IGHJ genes originated from the IGHJ4 family, observed in 83 out of 162 cases (51.2%). The second most frequent were genes from the IGHJ6 family (47 cases; 29%), followed by the IGHJ5 family (18 cases; 11.1%). The least frequently observed were genes from the IGHJ1 family (1 case; 0.6%) and the IGHJ2 family (3 cases; 1.9%). The distribution of IGHJ families among productive IGH rearrangements is shown in Fig. 19.

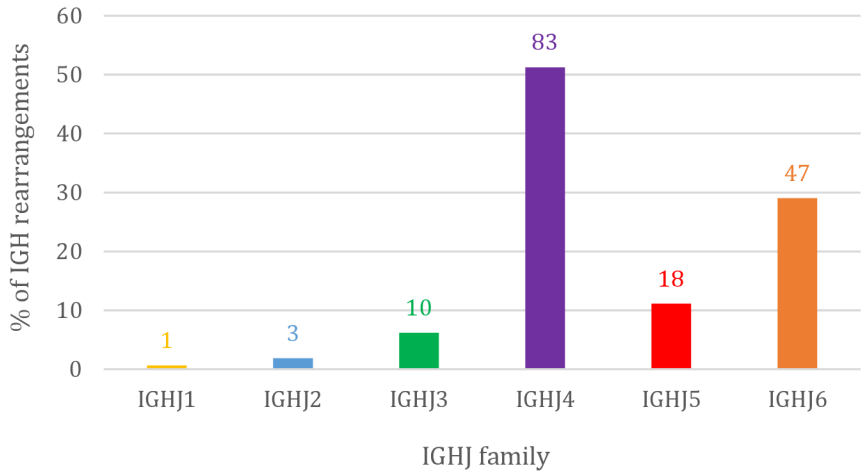


Fig. 19: Distribution of IGHJ families analysed by Sanger sequencing.
The number above the column represents the total number of detected IGHJ genes in the selected IGHJ family.

4.2 Capture-based next-generation sequencing – LYNX panel

Using capture-based NGS via the LYNX panel, we analysed all 98 samples. The sequencing depth varied across the samples, with the median coverage ranging from 454.5 to 2,086.0 reads per sample. The overall median coverage across all samples was 978.5 reads. A graphical representation of the distribution of median coverage across all samples is shown in Fig. 20.

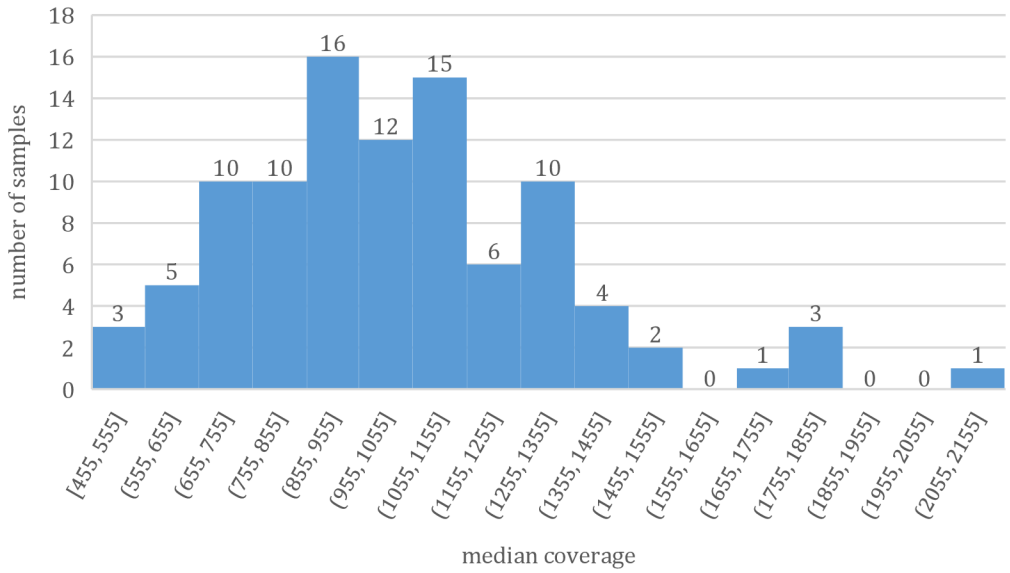


Fig. 20: Distribution of median sequencing coverage across samples analysed by the LYNX panel.

The number of detected fragments per productive IGH rearrangement had an interquartile range (IQR) of 27 to 225 fragments, with a median of 98 fragments. Observed values ranged from 3 to 510 fragments. Six cases exceeded this maximum value. A visual representation of the number of detected fragments per IGH rearrangement is shown in Fig. 21.

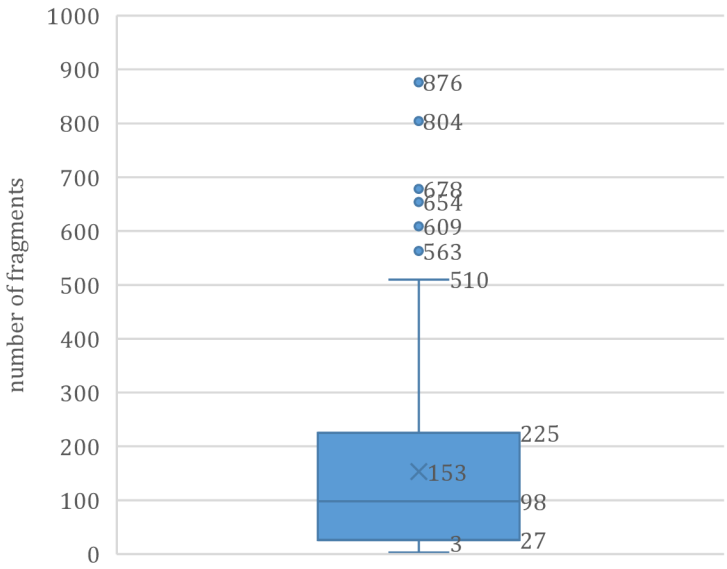


Fig. 21: Variation in the number of detected productive IGH rearrangement fragments in the LYNX panel.

In this cohort, 137 unique clonotypes were identified out of a total of 165 detected. The remaining 28 clonotypes were repeated in other samples from the same patients. In 119 cases (119/165; 72.1%), a complete IGHV sequence was available, allowing for the determination of IGHV mutational status. The presence of multiple productive IGH rearrangements was confirmed in 51 samples (52%) from 44 CLL patients following the set thresholds (i.e., 3 fragments and 1% frequency), while 47 samples from 32 CLL patients were monoclonal (48%). Among the multiclonal samples, biclonal cases were the most frequent (37/51; 72.5%), followed by triclonal (12/51; 23.5%) and tetraclonal cases (2/51; 3.9%). The clonal composition of all samples is shown in Fig. 22.

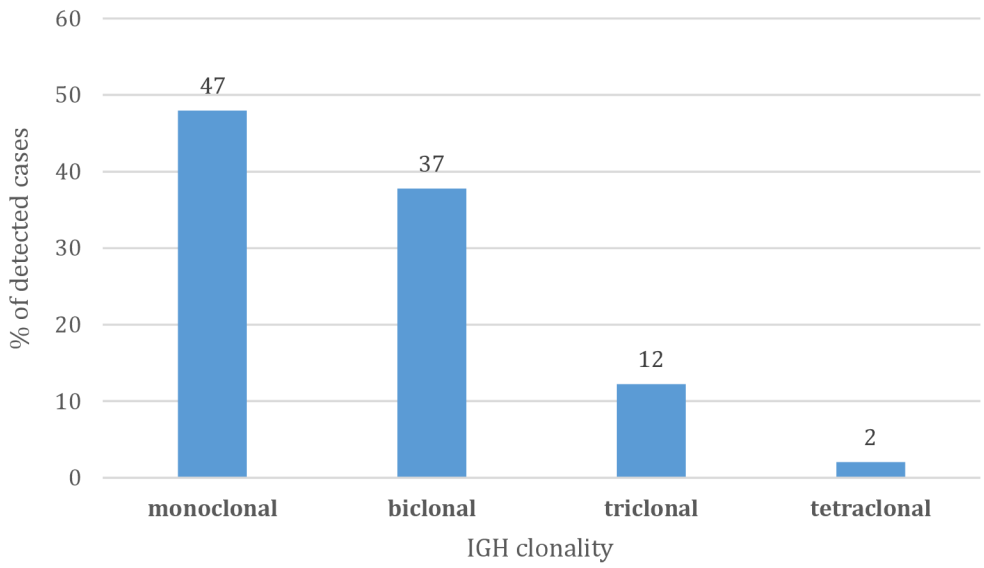


Fig. 22: Clonal composition of IGH rearrangements analysed by the LYNX panel.
The number above the column represents the total number of samples corresponding the selected IGH clonality.

For further analysis, we selected one multiclonal sample from each of the 44 patients in whom multiple productive IGH rearrangements were detected and examined the IGH repertoire among 101 IGH rearrangements. Most IGHV genes originated from the IGHV3 family, observed in 36 of 101 cases (35.6%). This was followed by genes from the IGHV1 family (30 cases; 29.7%) and the IGHV4 family (23 cases; 22.8%), as shown in Fig. 23. The least frequently observed were genes from the IGHV2 family (1 case; 1%) and the IGHV7 family (1 case; 1%).

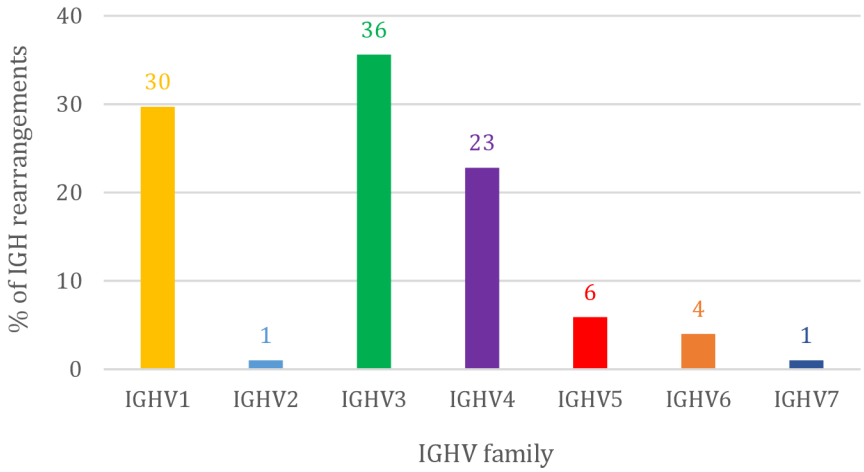


Fig. 23: Distribution of IGHV families analysed by the LYNX panel.
The number above the column represents the total number of detected IGHV genes in the selected IGHV family.

The most frequent individual IGHV gene overall was IGHV1-69 from the IGHV1 family (14/101; 13.9%), followed by IGHV3-23 (8/101; 7.9%), IGHV1-2 (7/101; 6.9%) and IGHV4-39 (7/101; 6.9%). The distribution of all individual IGHV genes among productive IGH rearrangements is illustrated in Fig. 24.

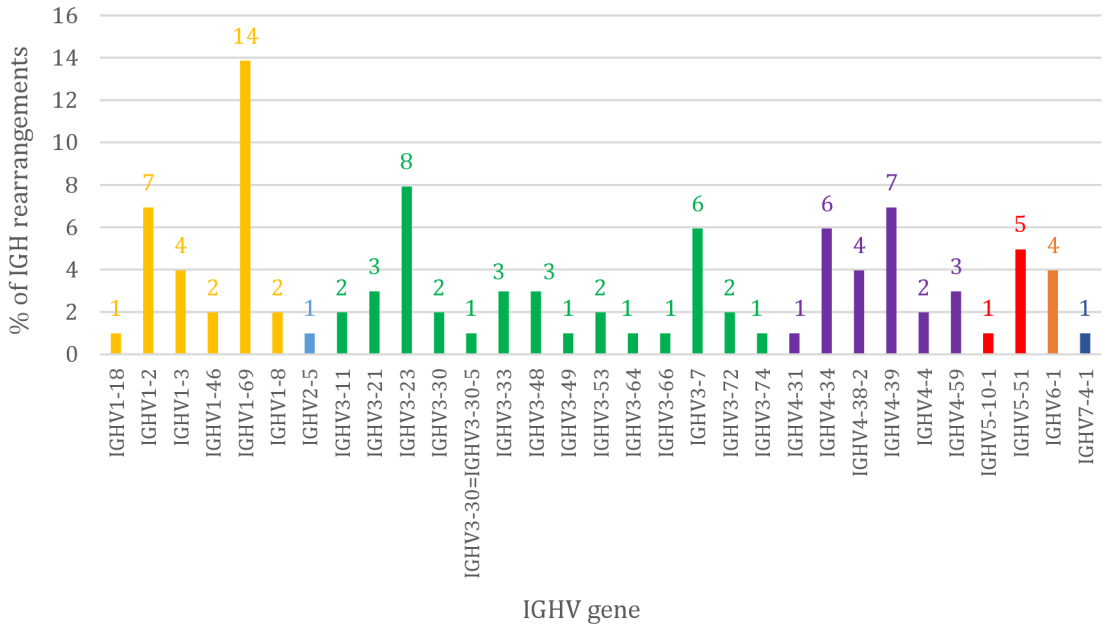


Fig. 24: Distribution of IGHV genes analysed by the LYNX panel.
The number above the column represents the total number of detected IGHV genes in the selected type of IGHV gene.

The majority of IGHD genes originated from the IGHD3 family, observed in 41 of 101 cases (40.6%). The next most common genes were from the IGHD6 family (24 cases; 23.8%) and the IGHD2 family (18 cases; 17.8%), as shown in Fig. 25. The least frequently observed were genes from the IGHD5 family (5 cases; 5%). No genes from the IGHD7 family were observed.

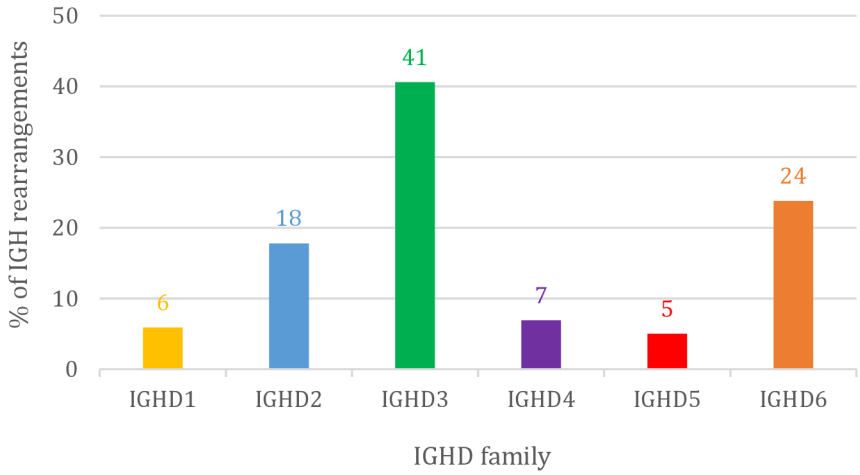


Fig. 25: Distribution of IGHD families analysed by the LYNX panel.
The number above the column represents the total number of detected IGHD genes in the selected IGHD family.

The most frequent individual IGHD gene was IGHD3-3 (14/101; 13.9%). The second most common was IGHD6-19 (12/101; 11.9%), followed by IGHD3-22 (10/101; 9.9%) and IGHD6-13 (10/101; 9.9%). The distribution of all individual IGHD genes among productive IGH rearrangements is shown in Fig. 26.

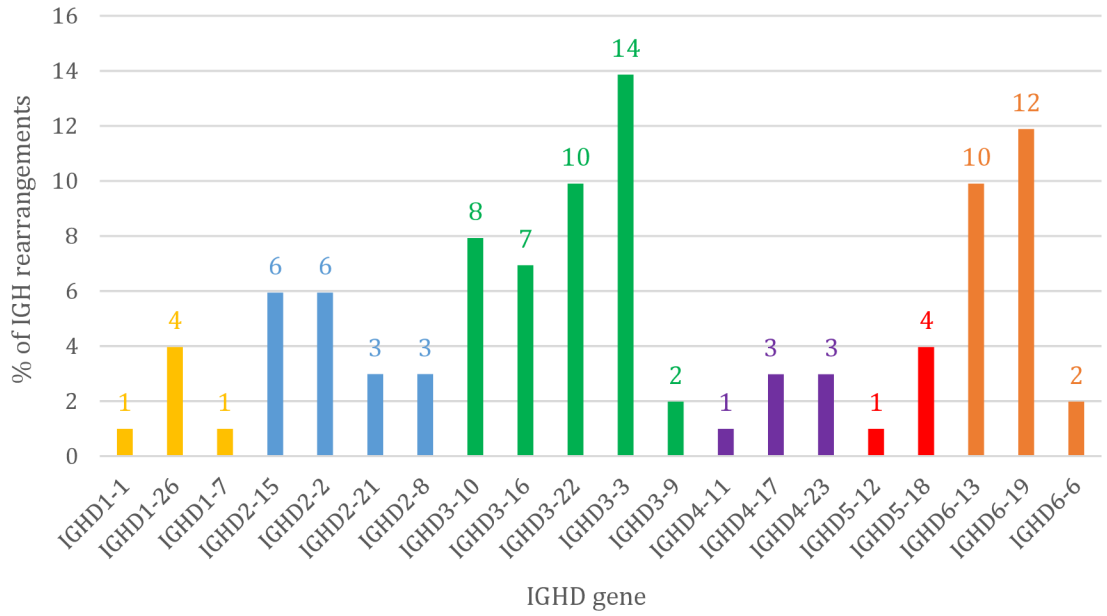


Fig. 26: Distribution of IGHD genes analysed by the LYNX panel.
The number above the column represents the total number of detected IGHD genes in the selected type of IGHD gene.

Most IGHJ genes originated from the IGHJ4 family, observed in 42 out of 101 cases (41.6%). The second most frequent were genes from the IGHJ6 family (36 cases; 35.6%). The least frequently observed were genes from the IGHJ2 family (2 cases; 2%). No genes from the IGHJ1 family were observed. The distribution of IGHJ families among productive IGH rearrangements is shown in Fig. 27.

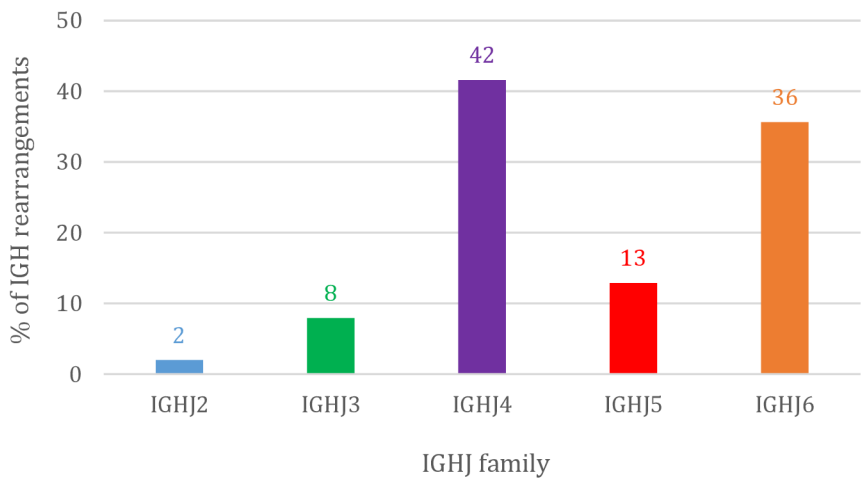


Fig. 27: Distribution of IGHJ families analysed by the LYNX panel.
The number above the column represents the total number of detected IGHJ genes in the selected IGHJ family.

4.3 Amplicon-based next-generation sequencing – IG aNGS

Using amplicon-based IG aNGS, we analysed all 98 samples. The total number of sequencing reads obtained per sample showed considerable variation, ranging from 167,174 to 3,222,645 reads. The median total read count across all samples was 1,719,720. A graphical representation of the distribution of total read counts across all analysed samples is presented in Fig. 28.

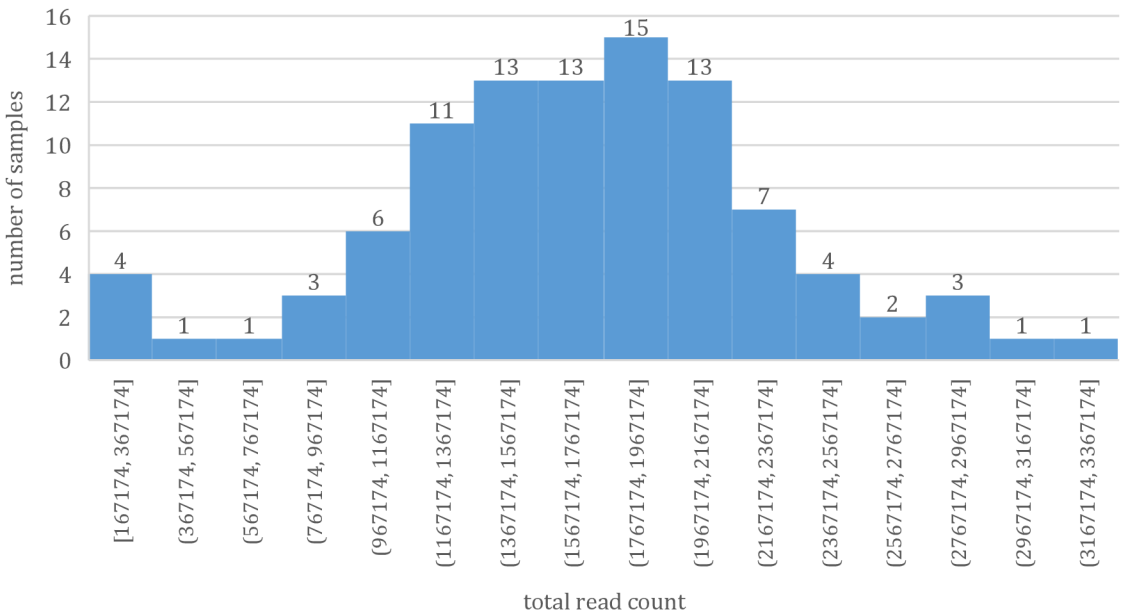


Fig. 28: Distribution of total read counts across samples obtained from IG aNGS analysis.

The number of detected reads per productive IGH rearrangement had an IQR of 58,681 to 1,241,292 reads, with a median of 397,308 reads. Observed values ranged from 3,019 to 2,866,972 reads, which falls within the expected normal range. A visual representation of the number of detected reads per IGH rearrangement is shown in Fig. 29.

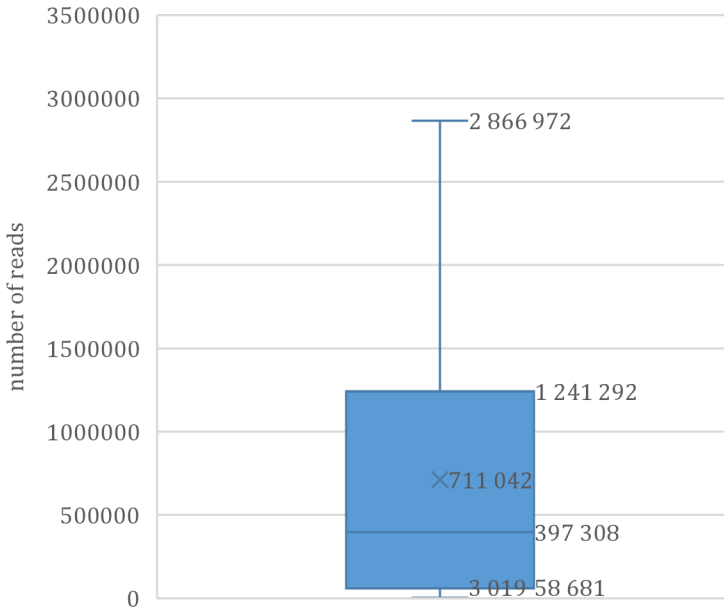


Fig. 29: Variation in detected reads across productive IGH rearrangements in IG aNGS.

In this cohort, 167 unique clonotypes were identified out of a total of 187 detected. The remaining 20 clonotypes were repeated in other samples from the same patients. The presence of multiple productive IGH rearrangements was confirmed in 60 samples (61.2%), while 38 samples were monoclonal (38.8%). Among the multiclonal samples, biclonal samples were the most frequent (37/60; 61.7%), followed by triclonal (16/60; 26.7%), tetraclonal (5/60; 8.3%), and samples with five or more clones (2/60; 3.3%). The clonal composition of all samples is shown in Fig. 30.

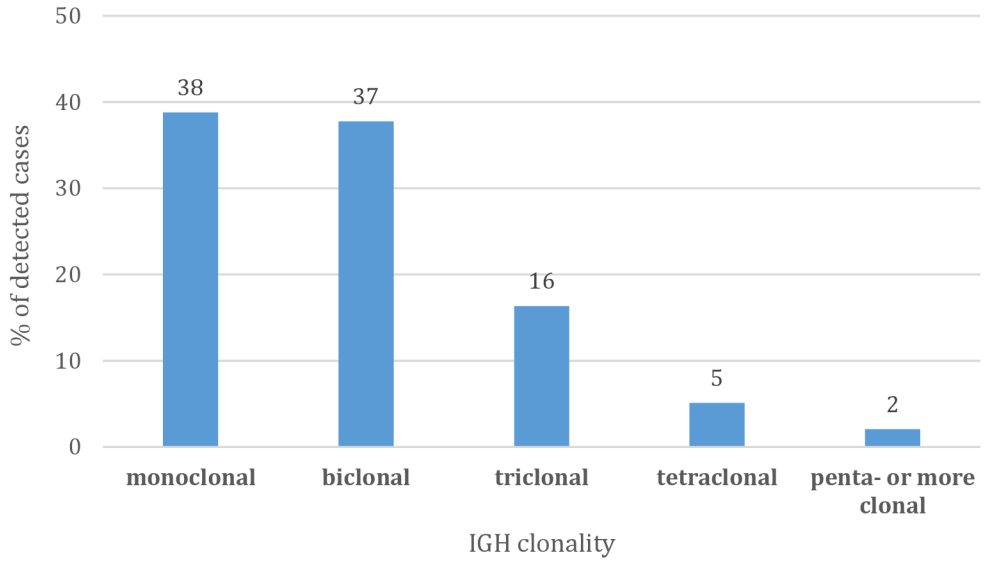


Fig. 30: Clonal composition of IGH rearrangements analysed by IG aNGS.
The number above the column represents the total number of samples corresponding the selected IGH clonality.

For further analysis, we selected one multiclonal sample from each of the 52 patients in whom multiple productive IGH rearrangements were detected, examining the IGH repertoire across 130 IGH rearrangements. In the case 3304, polyclonal composition of the respective sample was detected, and thus only the two major IGH rearrangements, also identified by other methods, were included in this analysis. Most IGHV genes originated from the IGHV3 family, observed in 53 of 130 cases (40.8%), followed by the IGHV1 family (37 cases; 28.5%) and the IGHV4 family (28 cases; 21.5%), as shown in Fig. 31. The least frequently observed were genes from the IGHV7 family (1 case; 0.8%).

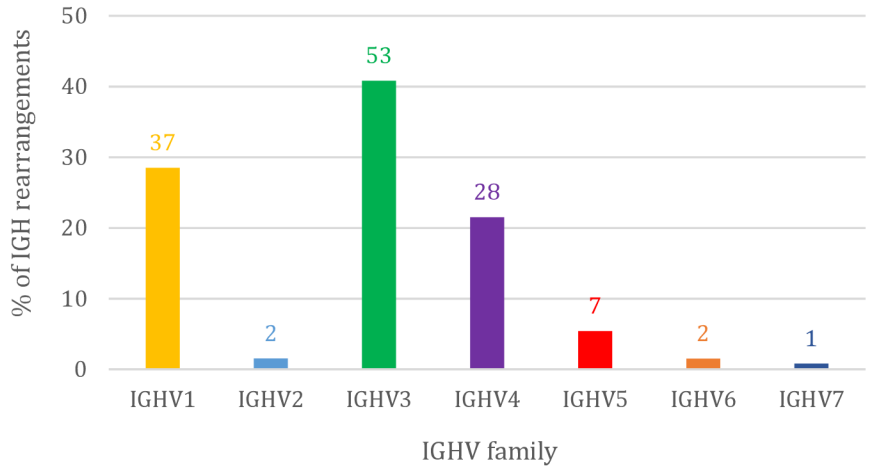


Fig. 31: Distribution of IGHV families analysed by IG aNGS.

The number above the column represents the total number of detected IGHV genes in the selected IGHV family.

The most frequent individual IGHV gene overall was IGHV1-69 from the IGHV1 family (17/130; 13.1%), followed by IGHV3-23 (9/130; 6.9%), IGHV3-30 = IGHV3-30-5 (8/130; 6.2%), and IGHV4-34 (8/130; 6.2%). The distribution of all individual IGHV genes among productive IGH rearrangements is illustrated in Fig. 32.

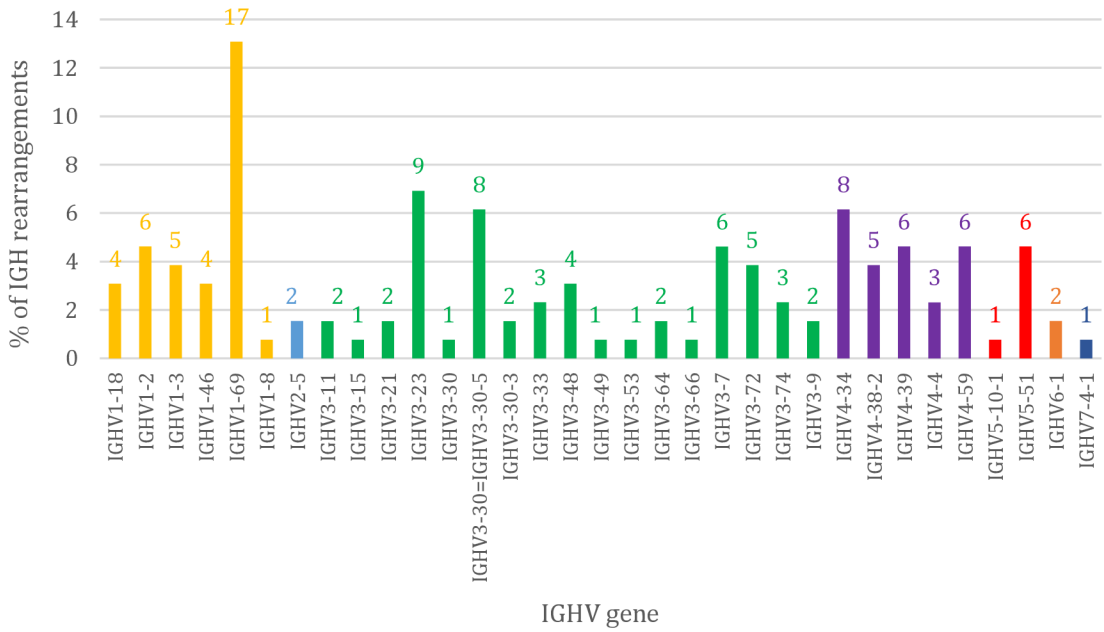


Fig. 32: Distribution of IGHV genes analysed by IG aNGS.

The number above the column represents the total number of detected IGHV genes in the selected type of IGHV gene.

In this cohort of multiclonal CLL patients, we also analysed IGHV mutational status using the 98% germline identity cut-off. The distribution of mutational status was balanced, with a slightly higher proportion of unmutated rearrangements (66/130; 50.8%) compared to mutated rearrangements (64/130; 49.2%). The vast majority of IGH rearrangements showed concordant IGHV mutational status (either mutated or unmutated) with other IGH rearrangement(s) within the same patient (34 out of 52 patients; 65.4%). Discordant cases, in which both mutated and unmutated IGHV rearrangements coexisted, were observed in 18 out of 52 patients (34.6%). A schematic overview of the IGHV mutational status distribution in multiclonal CLL patients is shown in Fig. 33.

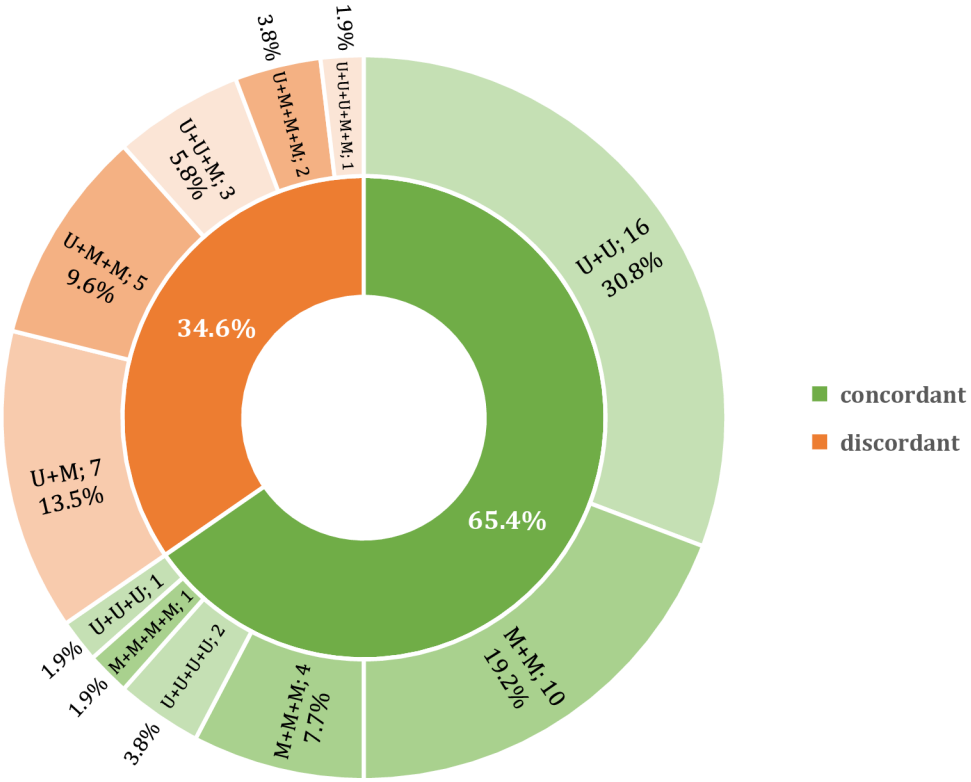


Fig. 33: Distribution of IGHV mutational status and concordance analysed by IG aNGS.

The majority of IGHD genes originated from the IGHD3 family, observed in 53 of 130 cases (40.8%). The next most common genes were from the IGHD6 family (31 cases; 23.8%) and the IGHD2 family (21 cases; 16.2%), as shown in Fig. 34. The least frequently observed were genes from the IGHD1 family (1 case; 0.8%).

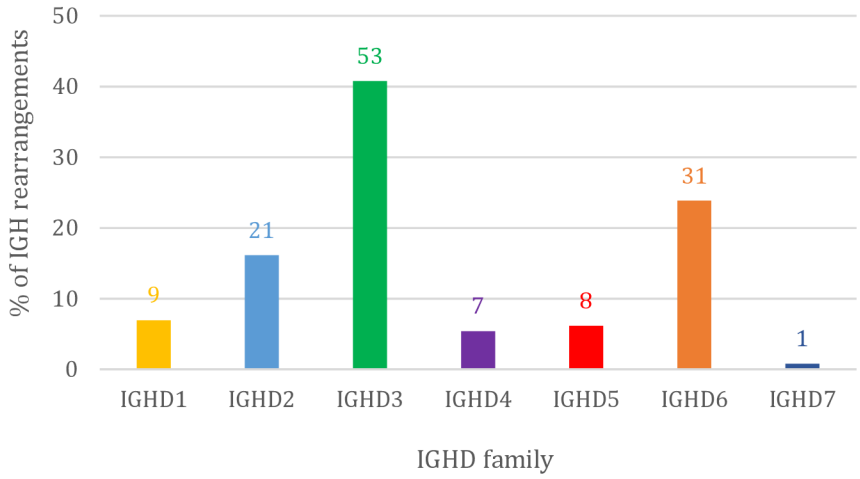


Fig. 34: Distribution of IGHD families analysed by IG aNGS.
The number above the column represents the total number of detected IGHD genes in the selected IGHD family.

The most frequent individual IGHD gene was IGHD6-19 (21/130; 16.2%). The second most common was IGHD3-3 (19/130; 14.6%), followed by IGHD3-10 (13/130; 10%), IGHD2-2 (11/130; 8.5%), and IGHD3-22 (10/130; 7.7%). The distribution of all IGHD genes among productive IGH rearrangements is shown in Fig. 35.

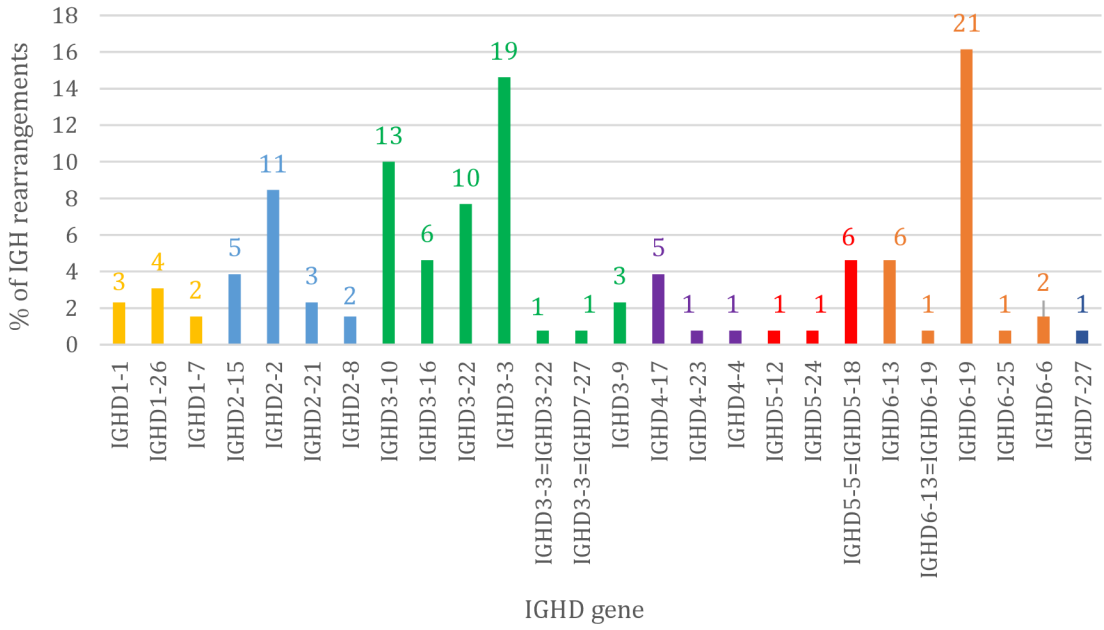


Fig. 35: Distribution of IGHD genes analysed by IG aNGS.
The number above the column represents the total number of detected IGHD genes in the selected type of IGHD gene.

Most IGHJ genes originated from the IGHJ4 family, observed in 62 out of 130 cases (47.7%). The second most frequent were genes from the IGHJ6 family (46 cases; 35.4%). The least frequently observed were genes from the IGHJ1 family (1 case; 0.8%) and the IGHJ2 family (2 cases; 1.5%). The distribution of IGHJ families among productive IGH rearrangements is shown in Fig. 36.

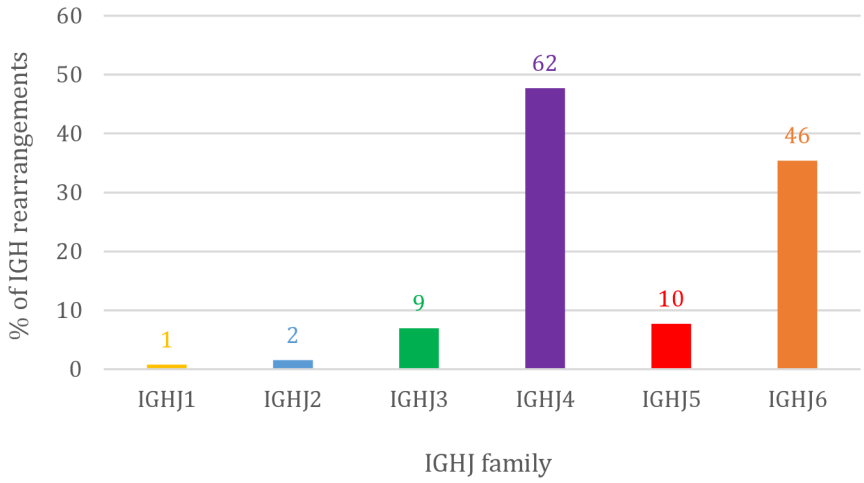


Fig. 36: Distribution of IGHJ families analysed by IG aNGS.
The number above the column represents the total number of detected IGHJ genes in the selected IGHJ family.

4.4 Differences between monoclonal and multiclonal CLL

IGH rearrangements were analysed in a cohort of 2,703 patients with monoclonal CLL as assessed using Sanger sequencing. A comparison between the monoclonal and multiclonal CLL cohort, revealed no significant differences in the overall IGH family usage. The only notable variation was a slightly higher abundance of the IGHD6 family in the multiclonal CLL cohort (21.6 – 23.8%) compared to the monoclonal group (16.8%), surpassing the usage of the IGHD2 family in this group. The distribution of IGH families among productive IGH rearrangements in patients with monoclonal CLL is shown in Fig. 37.

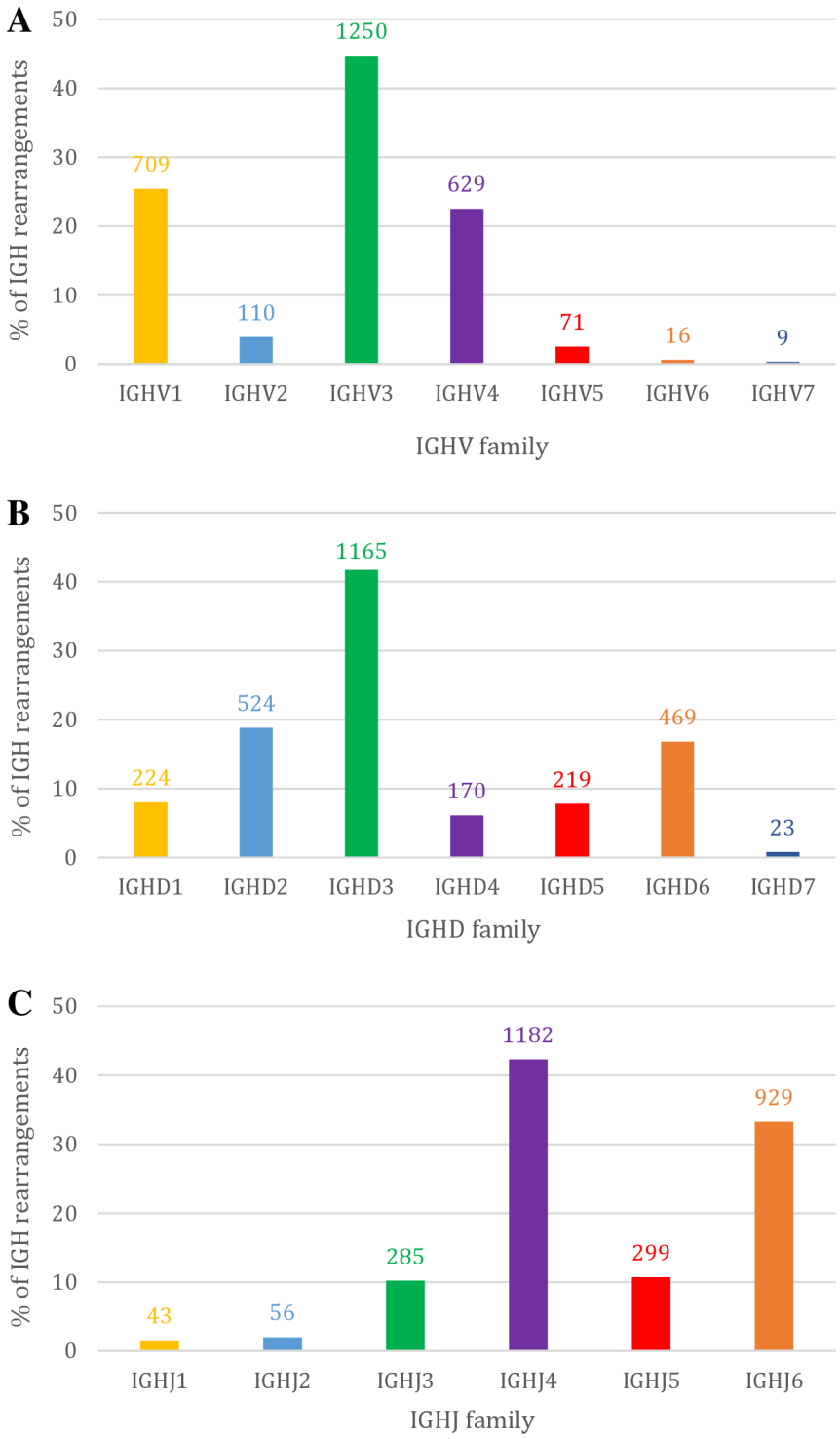
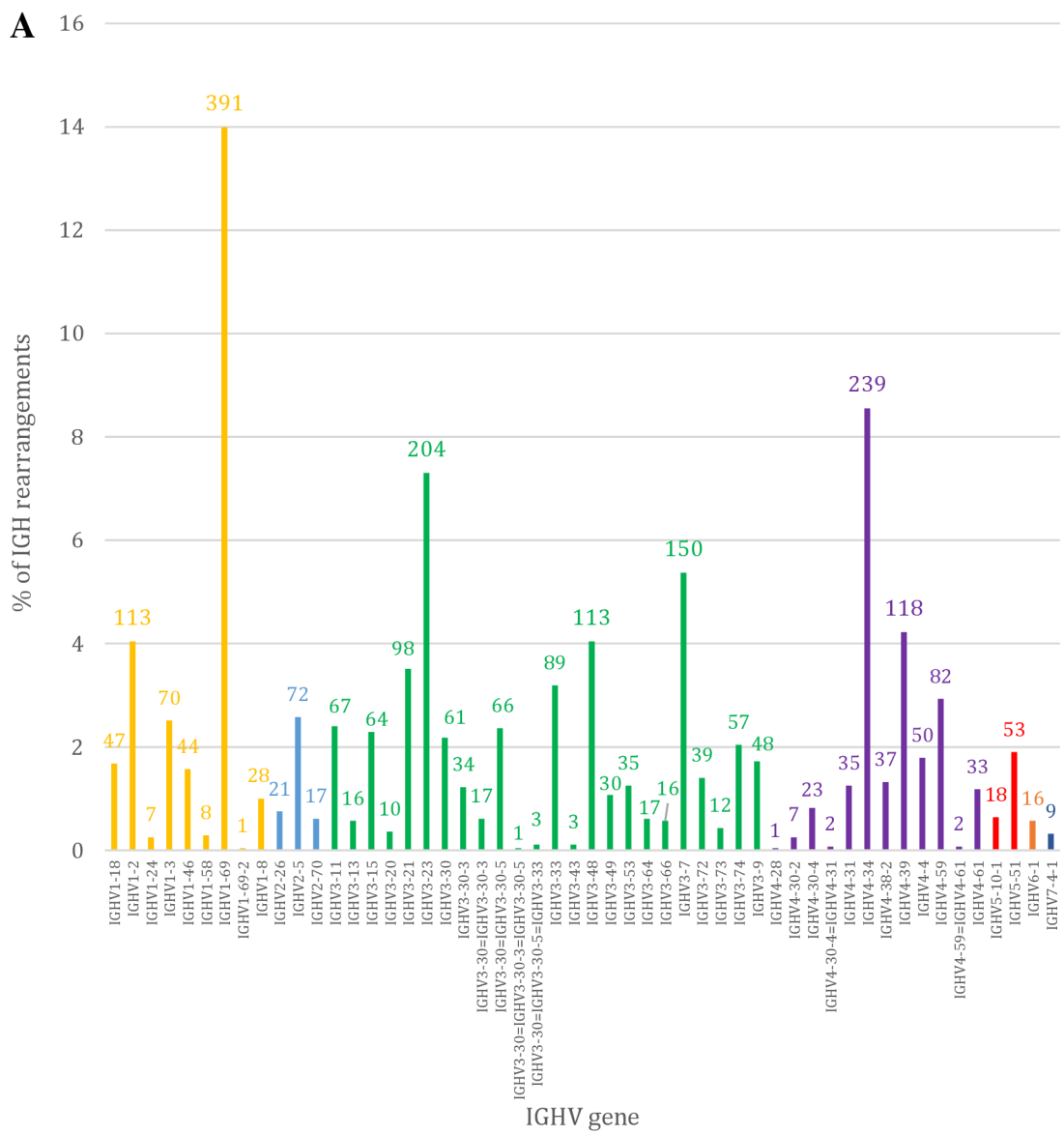


Fig. 37: Distribution of IGH families among monoclonal CLL patients.

The figure illustrates the distribution of IGH rearrangements across the IGHV (A), IGHD (B), and IGHJ (C) families. The number above each column indicates the total number of detected IGH genes within the corresponding IGH family.

A comparison of individual IGHV, IGHD and IGHJ genes revealed no significant differences between the studied cohorts. We also analysed the IGHV mutational status using the 98% germline identity cut-off in monoclonal CLL cohort. The distribution of mutational status was also balanced as in the multiclonal CLL cohort, with a slightly higher proportion of mutated rearrangements (1,426/2,794; 51%) compared to unmutated rearrangements (1,368/2,794; 49%). The distribution of all individual IGHV, IGHD and IGHJ genes among productive IGH rearrangements in patients with monoclonal CLL is illustrated in Fig. 38.



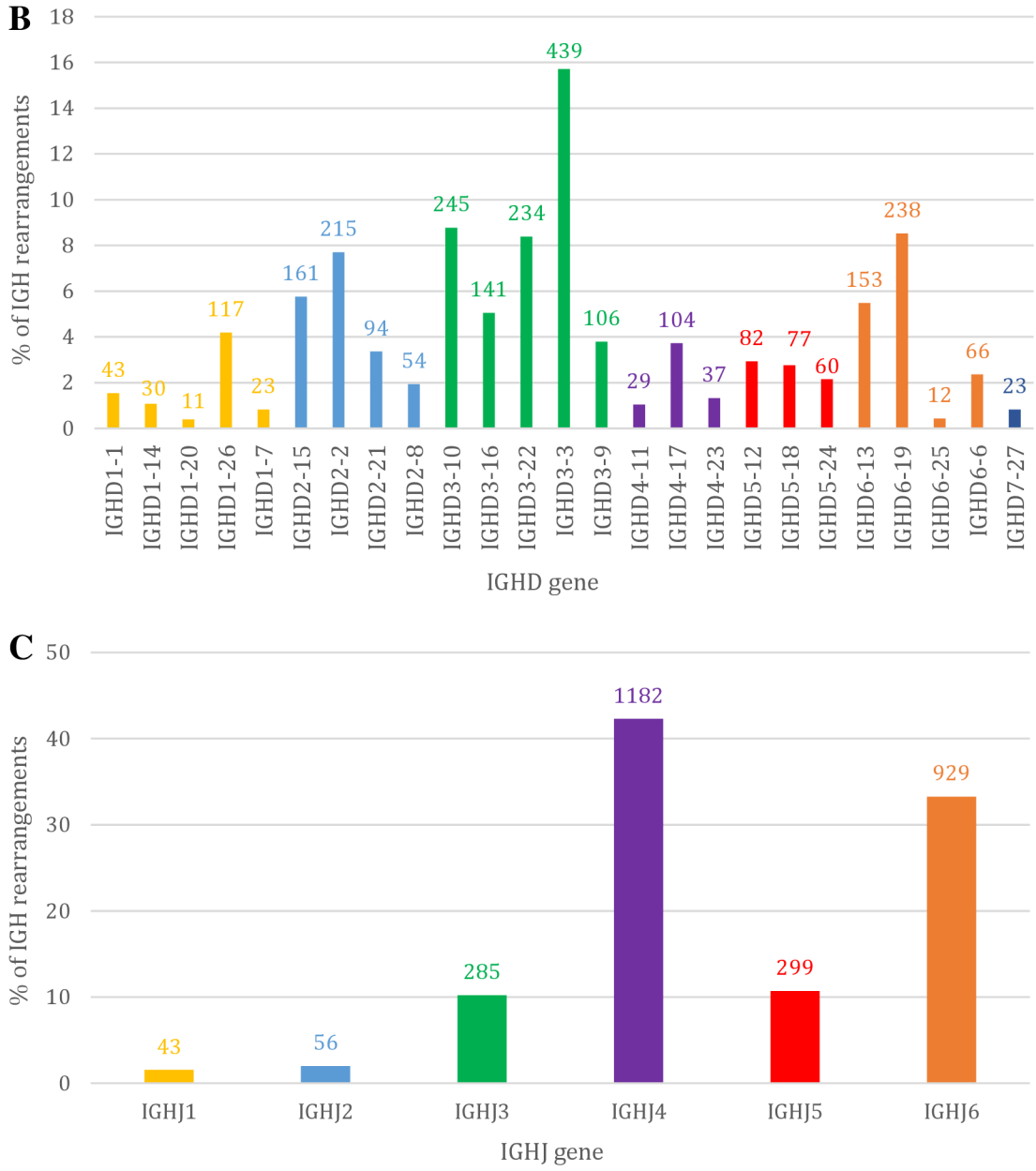


Fig. 38: Distribution of IGH genes among monoclonal CLL patients.

The figure illustrates the distribution of IGH rearrangements across the IGHV (A), IGHD (B), and IGHJ (C) genes. The number above each column indicates the total number of detected IGH genes in the selected type of IGH gene.

4.5 Comparison of methods for IGH rearrangement detection

Of the 98 samples in the cohort, 94 were successfully analysed using all three methods: Sanger sequencing, capture-based NGS with the LYNX panel, and amplicon-based IG

aNGS. This comprehensive approach enabled a direct comparison of detected productive IGH rearrangements and IGH clonality results across all methodologies. Of the 238 productive IGH rearrangements detected across all three methods, 182 (76.5%) were identified by Sanger sequencing, 165 (69.3%) by the LYNX panel, and 187 (78.6%) by IG aNGS.

Of all IGH rearrangements detected by Sanger sequencing (n = 182), 140 (76.9%) were also detected above the cut-offs using the LYNX panel. An additional 14 rearrangements (14/182; 7.7%) were technically identified by the LYNX panel but their abundance did not meet the predefined threshold of 3 fragments for confident detection. The remaining 26 rearrangements (26/182; 14.3%) were detected exclusively by Sanger sequencing and not observed in the LYNX panel results. Furthermore, the LYNX panel identified 17 novel IGH rearrangements that were not reported by Sanger sequencing (excluding 4 samples for which Sanger data were unavailable).

Among the 182 IGH rearrangements identified through Sanger sequencing, 139 (76.4%) were also reliably detected using the IG aNGS platform. In the case of one highly polyclonal sample (3304) from IG aNGS, only the two major IGH rearrangements, also identified by Sanger sequencing, were included in the analysis. In addition to these concordant cases, 39 rearrangements (39/182; 21.4%) were technically detected by IG aNGS but did not surpass the established threshold of 1% relative abundance required for confident reporting. An analysis of these sub-threshold detections showed that 15 rearrangements (15/39; 38.5%) had a relative abundance between 0.5% and 1%, 11 rearrangements (11/39; 28.2%) were present at levels between 0.1% and 0.5%, and 13 rearrangements (13/39; 33.3%) were detected at less than 0.1%. The remaining 4 IGH rearrangements (4/182; 2.2%) were identified exclusively by Sanger sequencing and were not detected by IG aNGS. In addition, IG aNGS revealed 44 novel IGH rearrangements that had not been previously reported by Sanger sequencing (excluding 4 samples for which Sanger data were unavailable).

Subsequently, we compared the two NGS methods: the LYNX panel and IG aNGS. Together, these methods identified 207 distinct IGH rearrangements. In the case of one highly polyclonal sample (3304) from IG aNGS, only the two major IGH rearrangements,

also identified by the LYNX panel, were included in the analysis. Of all IGH rearrangements detected by the LYNX panel (n = 165), 145 (87.9%) were also reliably detected using IG aNGS, indicating strong concordance between the two methods. An additional 20 rearrangements (20/165; 12.1%) were identified by the LYNX panel, however, their abundance in IG aNGS did not exceed the 1% threshold required for confident reporting. An analysis of these sub-threshold detections showed that 10 rearrangements (10/20; 50%) had a relative abundance between 0.5% and 1%, 6 rearrangements (6/20; 30%) were present at levels between 0.1% and 0.5%, and 4 rearrangements (4/20; 20%) were detected at less than 0.1%. Furthermore, IG aNGS identified 42 novel IGH rearrangements that were not reported by the LYNX panel. Of these, 9 rearrangements (9/42; 21.4%) were technically detected by LYNX, but their abundance did not meet the predefined threshold of 3 fragments required for confident detection by that method.

Of the total 238 IGH rearrangements, 130 (130/238; 54.6%) were identified by all three methods. When including rearrangements detected by Sanger sequencing but found below the predefined threshold for confident detection in the NGS methods, this number increased to 156 (156/238; 65.5%). Among the 182 rearrangements detected by Sanger sequencing, 21 (21/182; 11.5%) were identified by only one of the two NGS methods, and 31 (31/182; 17%) were either not detected at all or were found below the detection threshold in one or both NGS platforms. Of the 50 IGH rearrangements not detected by Sanger sequencing, 9 (9/50; 18%) were confirmed by both NGS methods, while 41 (41/50; 82%) were detected by one or the other. In 4 cases where Sanger sequencing data were not available, both NGS methods identified the same IGH rearrangements.

In 123 instances of all identified IGH rearrangements by NGS methods (123/207; 59.4%), both NGS methods detected IGH rearrangements in the same relative order of abundance. The absolute difference in the proportional representation of individual rearrangements between LYNX and IG aNGS, expressed in percentage, had an IQR from 0.36% to 9.88%, with a median difference of 3.89%. Changes in the abundance of minor IGH rearrangements, defined as those that were not the most abundant in a sample, occurred in 68 cases (68/207; 32.9%), with the majority of these (60/68; 88.2%) detected by only one of the two NGS methods. In these cases, the absolute difference in the proportional

representation of individual rearrangements between LYNX and IG aNGS were similar to those seen in rearrangements with concordant order of abundance, with an IQR from 1.59% to 5.51%, and a median difference of 3.26%. In 16 cases (16/207; 7.7%), a change in the major IGH rearrangement, defined as the most abundant in a sample, was observed. Here, the absolute differences in proportional representation of individual rearrangements between LYNX and IG aNGS were substantially larger, with an IQR ranging from 24.92% to 55.51% and a median difference of 30.87%. The numerical variation in the proportional representation of individual IGH rearrangements between NGS methods is illustrated in Fig. 39.

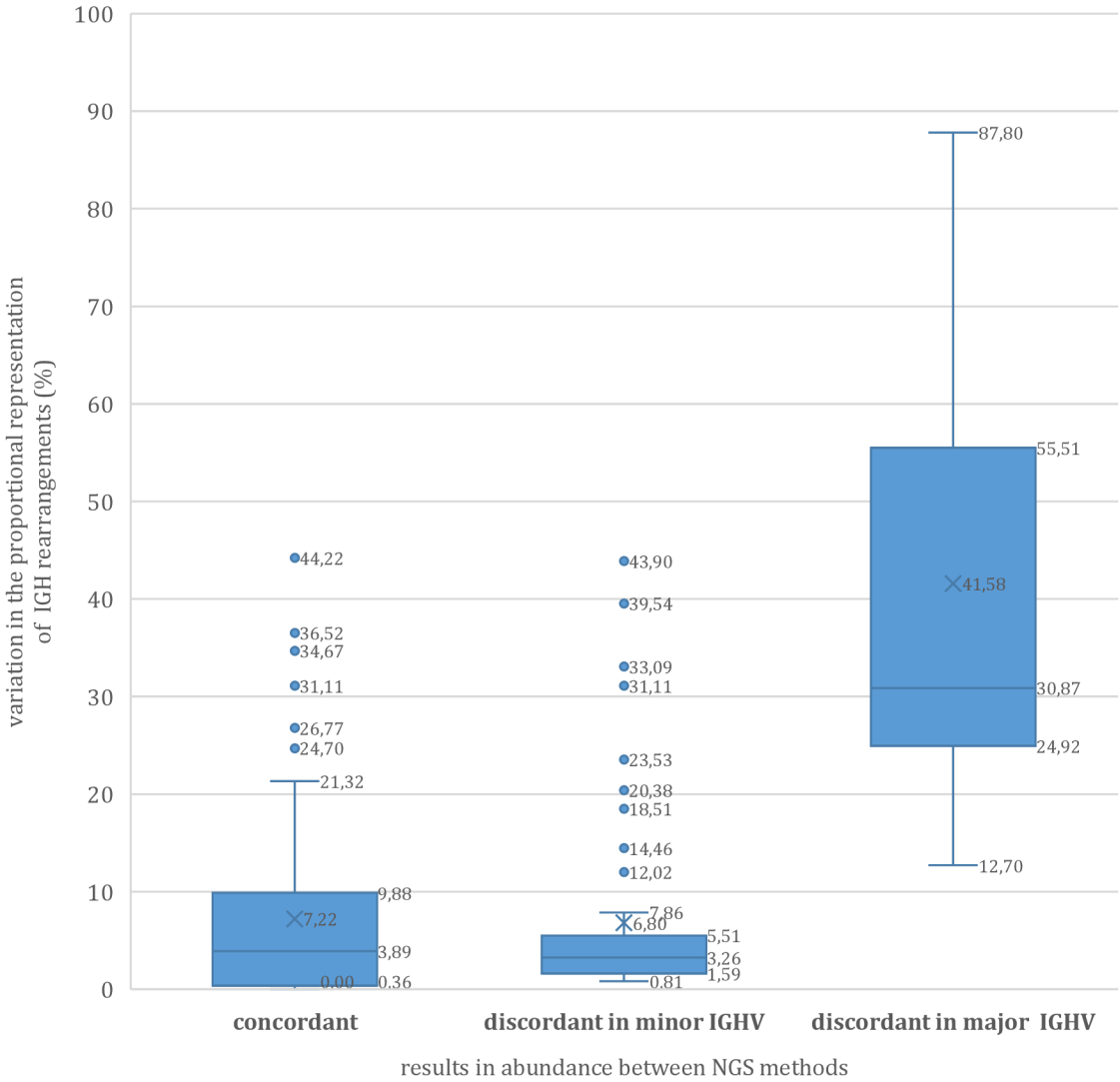


Fig. 39: Relative abundance differences of IGH rearrangements detected by NGS methods.

Subsequently, we compared IGH clonality results across all methodologies. As illustrated in Fig. 40, the results of IGH clonality obtained using the LYNX panel and IG aNGS were highly concordant, indicating strong agreement between these two NGS-based approaches. In contrast, Sanger sequencing showed greater variability across the different IGH clonality groups.

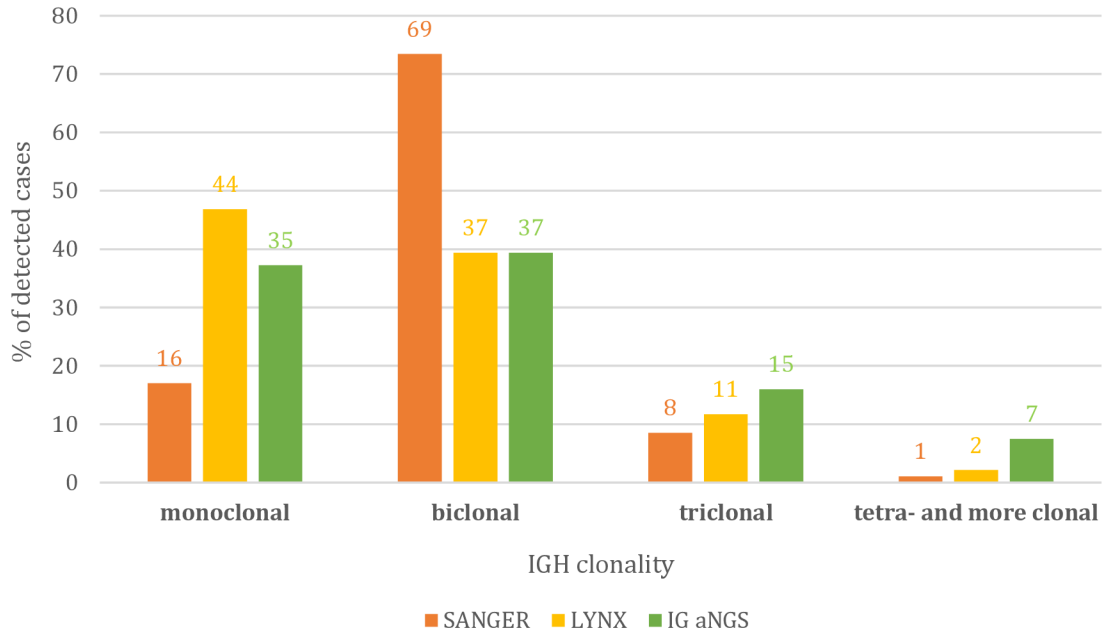


Fig. 40: Comparative analysis of three IGH clonality detection methods in multiclonal CLL. The number above the column represents the total number of detected cases with the selected IGH clonality.

Biclonal samples represented the most frequently observed clonality type across most of the methods implemented. Sanger sequencing reported the highest number of biclonal cases, almost twice as many as detected by the LYNX panel or IG aNGS. In contrast, tri-, tetra-, and higher-order clonal samples were less common overall, but their detection rates increased with the application of NGS technologies. IG aNGS demonstrated enhanced sensitivity for identifying these lower-frequency clonal populations, detecting clonotypes that were not captured by Sanger sequencing.

Notably, the detection of monoclonal samples varied substantially across the three methods. Sanger sequencing identified the fewest monoclonal cases. In contrast, both NGS-based approaches detected more monoclonal cases. Specifically, IG aNGS identified approximately 2.5 times more monoclonal cases than Sanger sequencing, while the LYNX

panel detected nearly three times as many. A comprehensive overview of all samples, including their IGH clonality results as determined by each of the three methods, is provided in Appendix E.

Among the 98 samples analysed, 47 (48%) showed identical IGH clonality results between Sanger sequencing and the LYNX panel. In another 47 samples (48%), discrepancies were observed between these two methods. In 34 of these discrepant cases (72.3%), the LYNX panel detected fewer IGH rearrangements than Sanger sequencing, indicating a lower number of detected IGH clonotypes. However, in 10 of these 34 cases (29.4%), the IGH rearrangements identified by Sanger sequencing were detected by the LYNX panel but fell below the predefined threshold of 3 fragments. Conversely, in 13 of the 47 discrepant cases (27.7%), the LYNX panel identified a higher number of detected IGH clonotypes than Sanger sequencing. For 4 of the 98 samples (4%), a direct comparison was not possible due to the unavailability of Sanger sequencing results from the same sampling.

Of the 98 samples analysed, 48 (49%) showed concordant IGH clonality between Sanger sequencing and IG aNGS. In 46 samples (47%), discrepancies were observed between these two methods. In 26 of these cases (56.5%), IG aNGS detected fewer IGH rearrangements, indicating lower IGH clonality compared to Sanger sequencing. Notably, in 20 of these 26 cases (76.9%), the IGH rearrangements identified by Sanger sequencing were also detected by IG aNGS but fell below the established threshold of 1% relative abundance. In contrast, IG aNGS identified higher IGH clonality in 20 of the 46 discrepant cases (43.5%). For 4 samples (4/98; 4%), comparison was not possible due to the absence of Sanger sequencing data from the same sampling.

Among the 98 samples analysed, 58 (59.2%) showed identical IGH clonality between the LYNX panel and IG aNGS. In 40 samples (40.8%), discrepancies were observed. In 14 of these cases (35%), IG aNGS detected fewer IGH rearrangements, indicating a lower number of detected IGH clonotypes. However, in all 14 cases, the additional rearrangements identified by LYNX were also present in IG aNGS but fell below the 1% relative abundance threshold. Conversely, in 26 cases (65%), IG aNGS reported higher clonality than the LYNX

panel. In 7 of these cases (26.9%), the additional IGH rearrangements were present in the LYNX panel but below its predefined threshold of 3 fragments.

Excluding the 4 samples without available Sanger sequencing results from the same time point, all three methods yielded concordant IGH clonality in 32 of 94 cases (34%). Among the 62 discordant cases, 31 (50%) showed agreement between one of the NGS methods and Sanger sequencing: 15 cases (48.4%) with the LYNX panel and 16 cases (51.6%) with IG aNGS. The remaining 31 discordant cases (50%) showed no concordance between Sanger and either NGS method. Of these, 22 samples (71%) had concordant results between the LYNX panel and IG aNGS. In all 9 of the remaining samples (29%), IG aNGS detected more IGH rearrangements than the LYNX panel. In 2 of those 9 cases (22.2%), the additional rearrangements detected by IG aNGS were also present in the LYNX panel but fell below its predefined threshold of 3 fragments.

Finally, we performed a technical comparison of the two NGS methods by analysing the relative abundance of fragments (LYNX) or reads (IG aNGS) corresponding to mutated and unmutated IGHV sequences. As illustrated in Fig. 41A, the LYNX panel captured mutated IGHV sequences at a lower fragment count compared to unmutated ones. In contrast, IG aNGS (Fig. 41B) did not show this discrepancy, with both mutated and unmutated IGHV sequences being detected at comparable read counts.

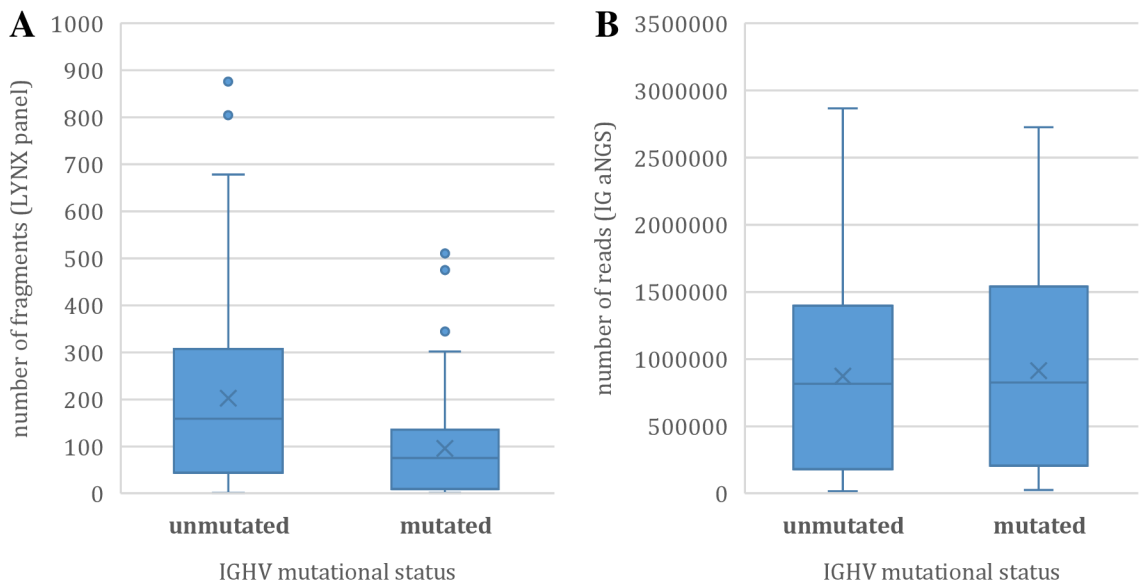


Fig. 41: Comparative detection efficiency of IGHV mutational status in LYNX (A) and IG aNGS (B).

Due to the previously observed variability between the two NGS methodologies, we further normalized the detection results to account for overall sequencing output. For the LYNX panel, we related the number of detected fragments per productive IGH rearrangement to the mean sample coverage. For IG aNGS, we calculated the ratio of reads per productive IGH rearrangement to the total read count per sample. As shown in Fig. 42A, the LYNX panel exhibited a lower ratio with decreasing IGHV somatic hypermutation (SHM) load, suggesting reduced detection efficiency for highly mutated IGHV sequences. In contrast, IG aNGS (Fig. 42B) displayed no clear dependency between the read ratios and IGHV SHM load, indicating more consistent performance regardless of mutational status.

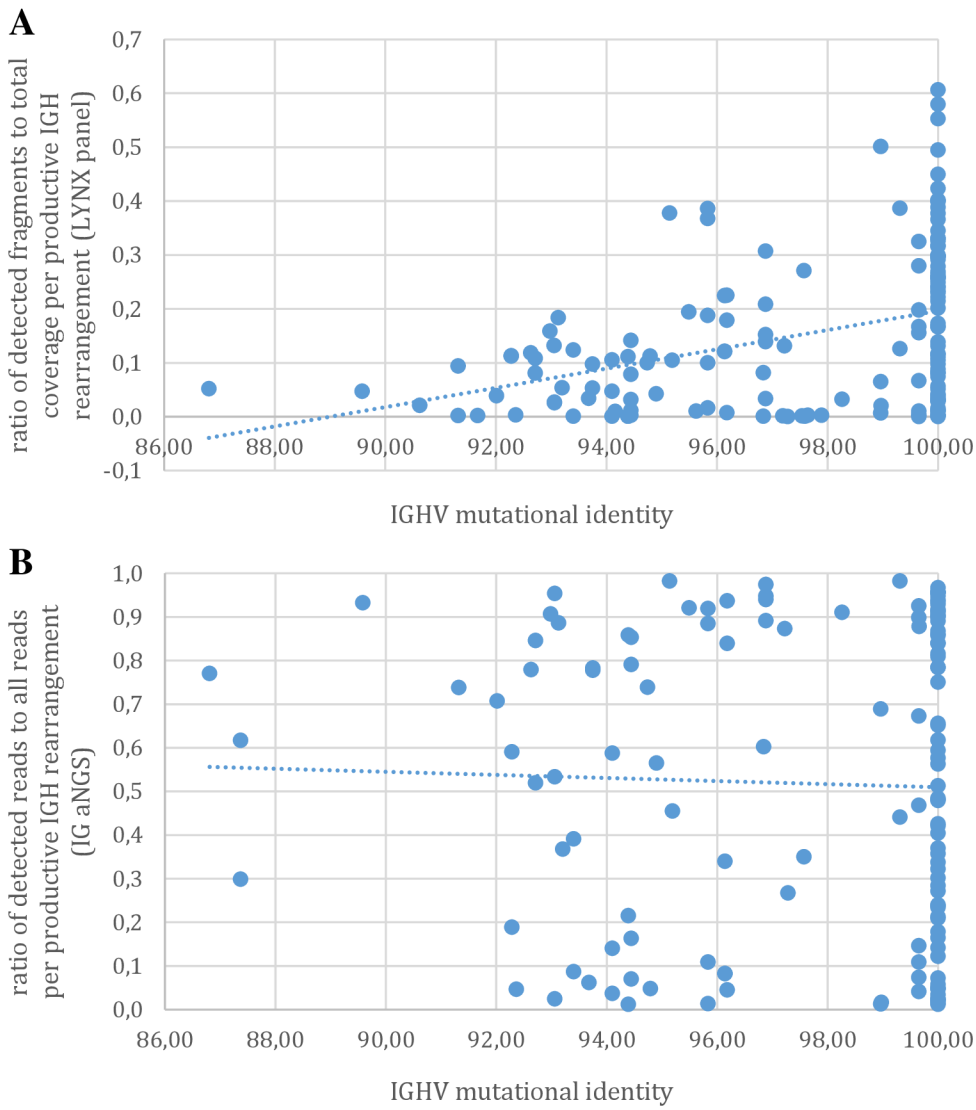


Fig. 42: Impact of IGHV mutational identity on detection efficiency in LYNX (A) and IG aNGS (B).

Then we proceeded to compare the individual ratios between the two NGS methods. Initially, this comparison was limited to monoclonal samples detected by both methods, to determine whether technical differences in capture efficiency existed between mutated and unmutated IGHV sequences in LYNX and IG aNGS. As shown in Fig. 43, unmutated clones exhibited, on average, approximately twice the fragment-to-coverage ratio in the LYNX panel compared to mutated clones and also displayed greater variability. In contrast, mutated clones tended to cluster at higher read ratios in the IG aNGS data, while unmutated clones were more heterogeneous, with some showing markedly low read support. Upon further inspection, we found that all five unmutated cases with low read ratios (red arrows) were affected by the presence of an unproductive IGH rearrangement in the same sample, which had a significantly higher read count than the productive rearrangement.

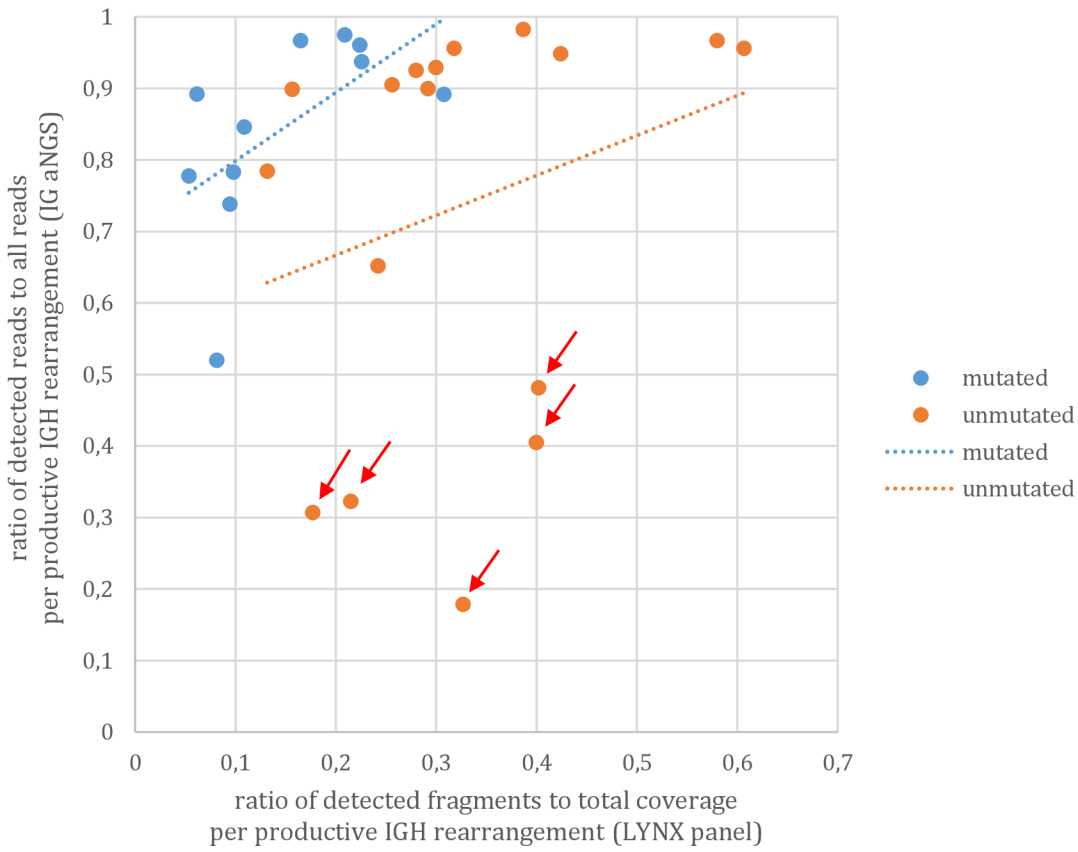


Fig. 43: Comparison of capture efficiency for mutated and unmutated IGHV sequences in monoclonal CLL across NGS platforms.

Red arrows in the chart represent cases affected by the presence of an unproductive IGH rearrangement in the same sample, which had a significantly higher read count than the productive IGH rearrangement.

To further assess the relationship between the two NGS methods, we compared the individual ratios of all detected productive IGH rearrangements between LYNX and IG aNGS. This analysis aimed to evaluate how efficiently fragment counts (normalized by coverage in LYNX) corresponded to sequencing reads (normalized by total read count in IG aNGS), and how this relationship differs depending on IGHV mutational status.

In both mutated and unmutated IGHV sequences, an increase in the fragment-to-coverage ratio in LYNX generally correlated with a higher read ratio in IG aNGS, reflecting increasing clonal size (Fig. 44). However, the distribution of these ratios varied markedly between the groups. Mutated clones were compressed into lower fragment-to-coverage ratio values, while unmutated clones reached considerably higher ratios. In contrast, read ratios in IG aNGS appeared more diffusely distributed across the entire range in both mutated and unmutated IGHV sequences.

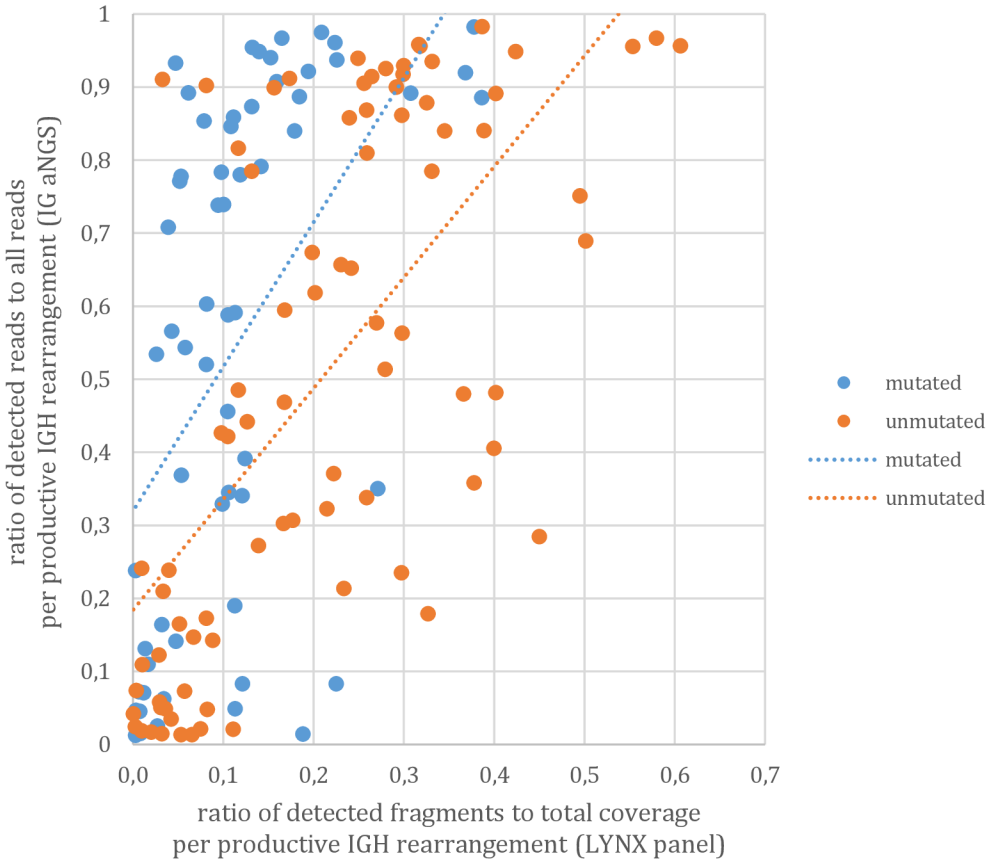


Fig. 44: Comparison of capture efficiency for mutated and unmutated IGHV sequences in all detected productive IGH rearrangements across NGS platforms.

Notably, 25 out of 145 sequences (17.2%) from IG aNGS, derived from 17 patients, were influenced by the presence of unproductive IGH rearrangements in the same sample. In these cases, the unproductive rearrangement exhibited either comparable or substantially higher read counts than the productive one. These skewed cases were predominantly linked to unmutated IGHV sequences (18/25; 72%), while 7 cases (28%) were associated with mutated IGHV sequences. Fig. 45 illustrates the relationship between fragment-to-coverage and read ratios after excluding these sequences influenced by unproductive IGH rearrangements.

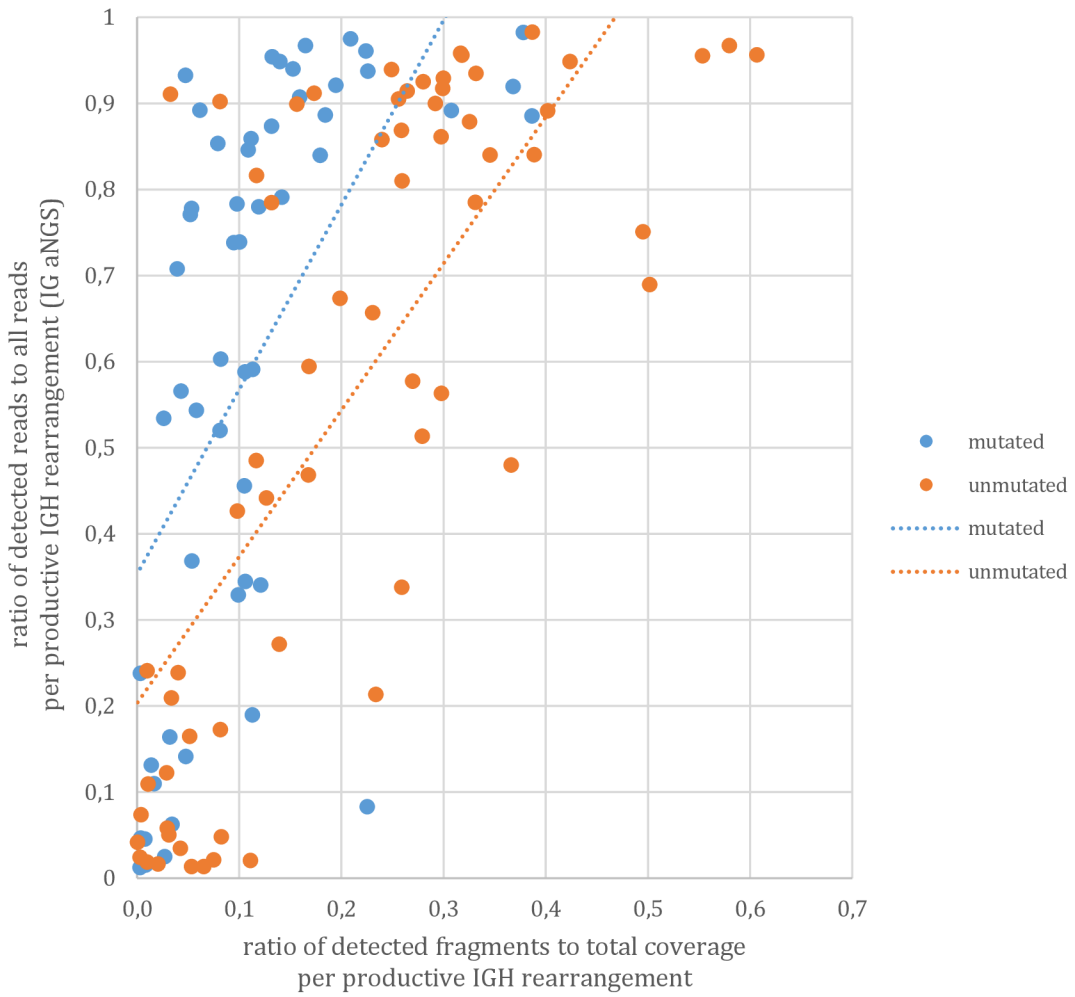


Fig. 45: Comparison of capture efficiency for mutated and unmutated IGHV sequences in all detected productive IGH rearrangements unaffected by unproductive IGH rearrangements across NGS platforms.

Analysis of the IGH gene usage in these unproductive rearrangements revealed that IGHV genes most frequently originated from the IGHV3 family (9/17; 52.9%), followed by IGHV2 and IGHV4 (3/17 each; 17.6%), and IGHV1 (2/17; 11.8%). For IGHD segments, the most represented was IGHD3 (8/17; 47.1%), followed by IGHD2 (6/17; 35.3%), IGHD1 (2/17; 11.8%), and IGHD4 (1/17; 5.9%). Regarding IGHJ gene usage, IGHJ4 was the most common (7/17; 41.2%), followed by IGHJ6 (6/17; 35.3%), IGHJ3 (2/17; 11.8%), and IGHJ5 (2/17; 11.8%).

4.6 Clonal evolution of IGH rearrangements in multiclonal CLL

In the cohort of 76 multiclonal CLL patients, we conducted an analysis of 20 patients for whom multiple serial samples were available. Specifically, two samples were analysed from 18 patients, and three samples from two patients. This analysis focused on the dynamics of IGH clonality, assessing how the number and relative proportions of detected IGH rearrangements changed over time.

Out of the 20 patients analysed, 15 (75%) exhibited changes in the total number of detected IGH rearrangements between their first and last available samples, indicating a shift in clonality. Among the 15 patients with clonality changes, 13 (86.7%) showed a decrease in the number of detected IGH rearrangements. In 10 of these 13 patients (76.9%), there was a reduction by one IGH rearrangement, most commonly a shift from biclonality to monoclonality. In 2 of the 13 patients (15.4%), a decrease of two IGH rearrangements was observed, representing a transition from triclonality to monoclonality. In one patient, the number of decreasing IGH rearrangements varied depending on the method used. In contrast, 2 of the 15 patients (13.3%) displayed the emergence of new IGH rearrangements in later samples. In 11 of these 15 patients (73.3%), the major clone, defined by the IGH clonotype with the highest representation, remained the same across all analysed time points.

In 13 of 15 patients (86.7%), the overall IGHV mutational status of the patient remained unchanged, despite a detectable change in the number of IGH rearrangements. In 9 of these cases (9/13; 69.2%), there was a reduction in the number of unmutated IGHV rearrangements; in 3 cases (3/13; 23.1%), there was a decrease in mutated IGHV rearrangements; and in 1 case (1/13; 7.7%), we observed an increase in unmutated IGHV

rearrangements alongside another unmutated one. In the remaining 2 of 15 patients (13.3%), the overall IGHV mutational status changed. In one patient (BRNO2543), a previously present mutated IGHV rearrangement was no longer detectable, which led to a transition from a discordant IGHV mutational profile to the unmutated status. In the other patient (BRNO1841), a previously absent unmutated IGHV rearrangement was detected, changing the IGHV mutational status from mutated to discordant, with two IGH rearrangements showing different mutational statuses.

Among the 13 patients with a declining number of clones, 9 (69.2%) showed an increased dominance of the major clone, while minor clones either decreased in abundance (3/9; 33.3%) or disappeared completely (6/9; 66.7%) over time. Moreover, in 4 of the 13 patients (30.8%) with decreasing clonal diversity, we observed more complex dynamics. In these cases, the major IGH rearrangement was eventually replaced by another IGH rearrangement in a later sample, indicating that not only are minor clones capable of persisting, but they may also expand to overtake the initially dominant clone. The dynamics of IGH clonotypes in these four patients, identified through the LYNX panel, is illustrated below.

In patient BRNO0319, we detected four distinct clonotypes across three collected samples. In the second of the three serial samples, a notable shift in the dominant IGH rearrangement was observed. The previously minor clonotype, represented by the CDR3 sequence CARGQQGDYVYGMDVW, initially present at 30.4% in the first sample, expanded to 85%, becoming the new dominant IGH rearrangement. Concurrently, the formerly dominant clonotype, CAKAREQQLPPEDYW, decreased markedly from 59% to 11.6%. By the third sample, CAKAREQQLPPEDYW was no longer detectable, while CARGQQGDYVYGMDVW further expanded to comprise 99.2% of the total clonal population. The remaining two IGH clonotypes, each accounting for less than 10% in the first sample, gradually declined over time. One of these disappeared entirely by the final sample, while the other fell below 1%. A visual representation of the dynamics of IGH rearrangements in patient BRNO0319 is shown in Fig. 46.

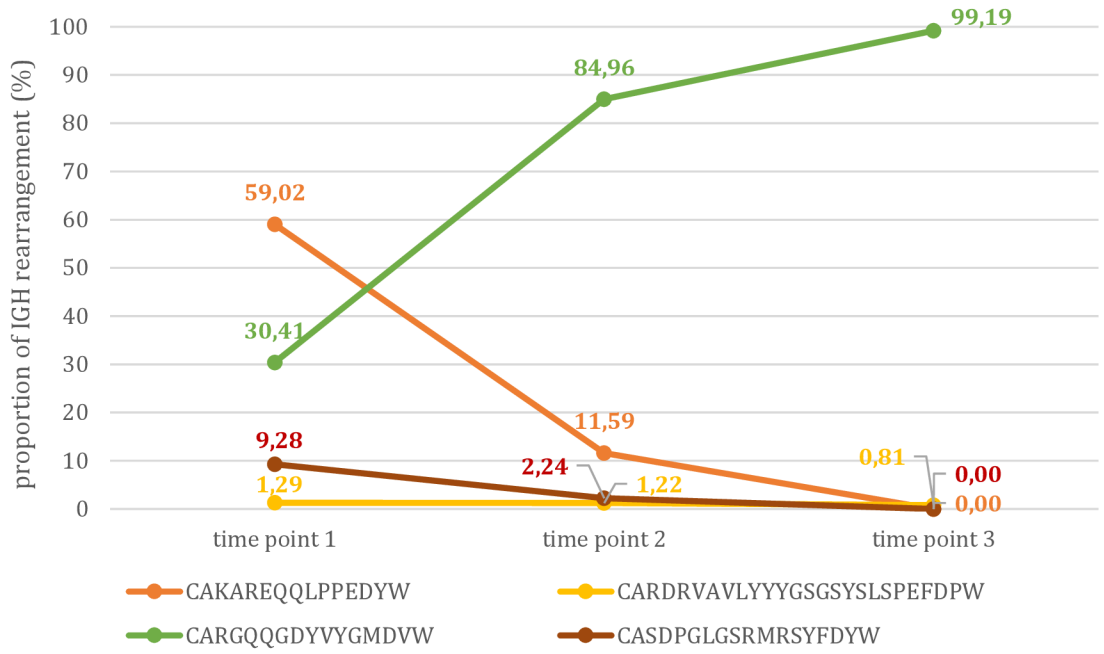


Fig. 46: Clonal dynamics of IGH rearrangements in patient BRNO0319 over time.

In patient BRNO1048, we detected two distinct clonotypes across three collected samples. In the second of the three serial samples, we observed a significant shift in the dominant IGH rearrangement. The previously minor clonotype, represented by the CDR3 sequence CARGAAATGDFDYW, initially present at 10.5% in the first sample, expanded to 83.8%, becoming the new dominant IGH rearrangement. Concurrently, the formerly dominant clonotype, CARGYTESGEGSGYYFNWFDPW, decreased markedly from 88.5% to 16.2%. By the third sample, CARGYTESGEGSGYYFNWFDPW was no longer detectable, while CARGAAATGDFDYW had expanded to fully dominate the clonal population. A visual representation of the dynamics of IGH rearrangements in patient BRNO1048 is shown in Fig. 47.

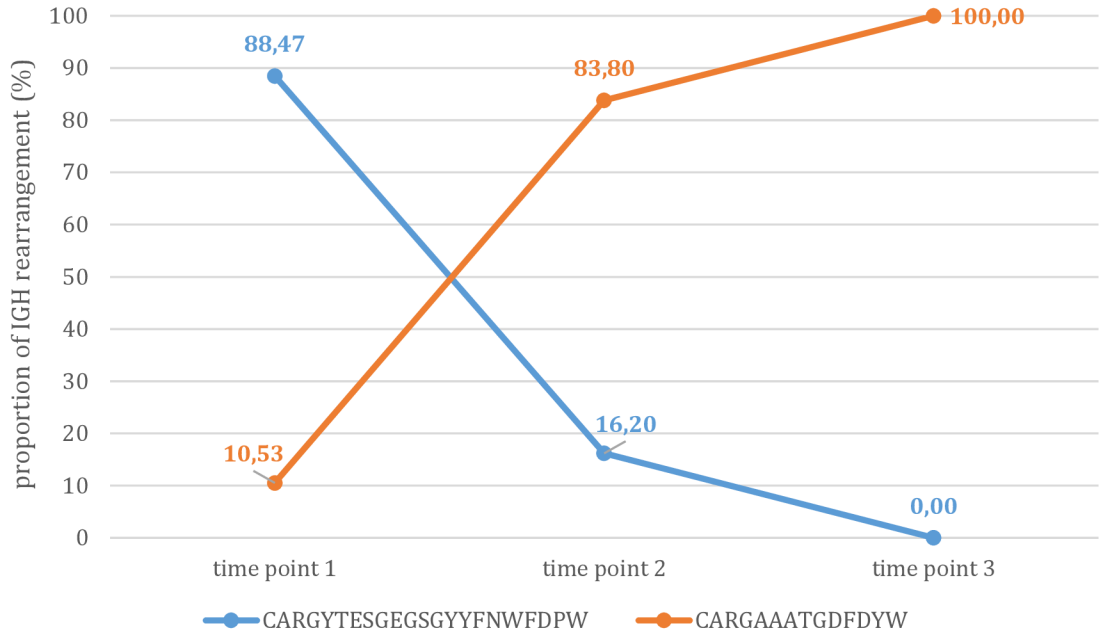


Fig. 47: Clonal dynamics of IGH rearrangements in patient BRNO1048 over time.

In patient BRNO1502, we detected three distinct clonotypes across two collected samples. In the first sample, two major IGH rearrangements were relatively balanced in proportion. By the second timepoint sample, we observed a notable shift in the proportions of these rearrangements. The previously minor clonotype, represented by the CDR3 sequence CTRDANYFDGSAYYDAFDIW, initially present at 43.9% in the first sample, expanded to 97%, becoming the new dominant IGH rearrangement. Conversely, the formerly dominant clonotype, CARSPSLYGDWYFDLW, declined from 53.1% to 3%. The third IGH rearrangement, which accounted for 3.1% in the first sample, disappeared entirely by the final timepoint. A visual representation of the dynamics of IGH rearrangements in patient BRNO1502 is shown in Fig. 48.

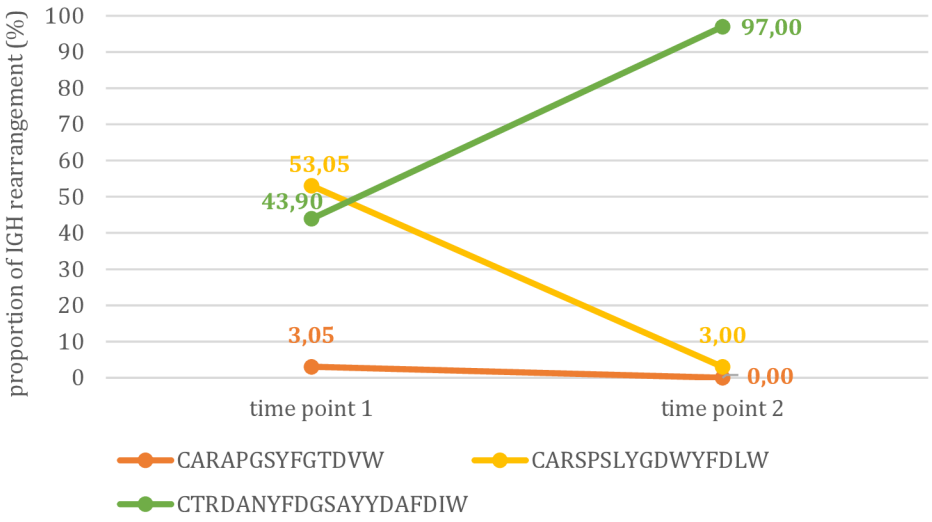


Fig. 48: Clonal dynamics of IGH rearrangements in patient BRNO1502 over time.

In patient BRNO1718, we detected two distinct clonotypes across two collected samples. In the second timepoint sample, a significant shift in the dominant IGH rearrangement was observed. The previously minor clonotype, represented by the CDR3 sequence CARGQSSIAAWYYFDYW, initially present at 39.1% in the first sample, expanded to 100%, fully dominating the clonal population. Concurrently, the formerly dominant clonotype, CARSSGEPYDFWTNYYGMDVW, which had accounted for 60.9% in the first sample, decreased markedly and was no longer detectable. A visual representation of the dynamics of IGH rearrangements in patient BRNO1718 is shown in Fig. 49.

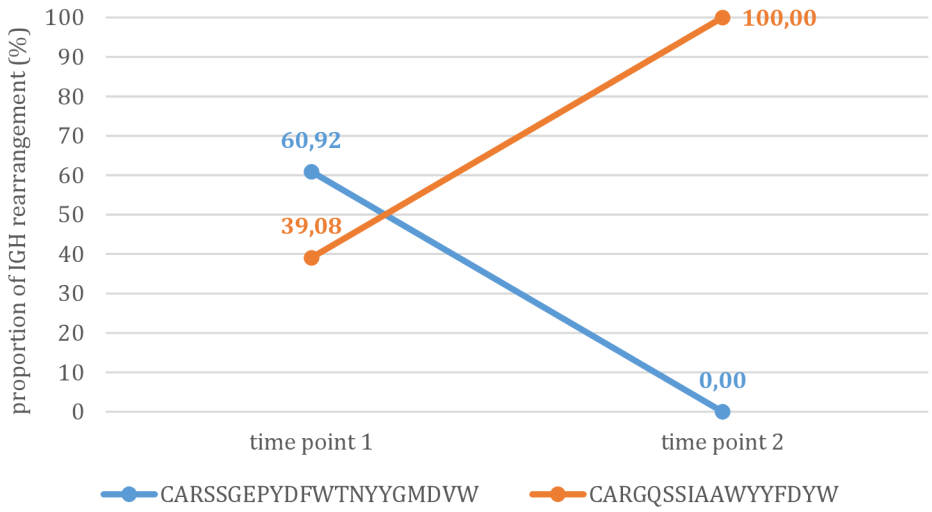


Fig. 49: Clonal dynamics of IGH rearrangements in patient BRNO1718 over time.

5 Discussion

CLL is typically characterised by the monoclonal expansion of a single mature B-lymphocyte, defined by a unique IGHV-IGHD-IGHJ rearrangement within the IGH gene. Additionally, the SHM status of the clonotypic IGHV gene serves as a critical biomarker for prognostic stratification and predicting therapeutic responses (Agathangelidis et al., 2022). In this master's thesis, we performed a comprehensive analysis of the IGH gene repertoire in the cohort of CLL samples collected in the Centre of Molecular Biology and Genetics, University Hospital Brno. The results of this analysis, detailed in Chapter 4.4, revealed patterns of IGH gene usage consistent with established findings reported by Karan-Djurasevic et al. (2017).

Importantly, in a subset of CLL patients, multiple (most commonly two) productive IGH gene rearrangements can be detected. The most likely explanation is the presence of two or more distinct leukemic B-cell clones within the patient, each defined by a different IGH rearrangement (Brazdilova et al., 2018; Chang and Cerny, 2006; González-Campos et al., 1997; Hsi et al., 2000; Plevova et al., 2014; Sanchez et al., 2003). The coexistence of multiple productive IGH rearrangements in CLL poses a significant challenge for prognostication, particularly when the co-detected rearrangements exhibit discordant IGHV mutational statuses (Langerak et al., 2011; Plevova et al., 2014; Rosenquist et al., 2017). These discordant cases complicate risk stratification, as IGHV mutational status remains a key determinant of clinical outcome.

In this master's thesis, we focused on the analysis of IGH rearrangements in a cohort of 76 CLL patients with a history of multiple productive IGH rearrangements detected by Sanger sequencing. These samples were further analysed using two different NGS methods: capture-based NGS with the LYNX panel, and amplicon-based IG aNGS. In the multiclonal CLL cohort, the distribution of IGH gene usage was consistent across all tested methods. A comparison between the monoclonal and multiclonal CLL cohorts, revealed no significant differences in the overall IGH gene usage. The only notable variation was a slightly higher abundance of IGHD6 family over IGHD2 in the multiclonal CLL cohort compared to the monoclonal CLL group. This difference may result from distinct combinations of rearranged

IGHV, IGHD, and IGHJ genes or from potential variations in antigenic stimulation during the patient's lifetime (Chiorazzi et al., 2021). Additionally, the distribution of IGHV mutational status was generally balanced between the monoclonal and multiclonal CLL cohorts.

After comparing IGH gene usage between the monoclonal and multiclonal CLL cohorts, we analysed the number of detected productive IGH rearrangements per sample in the multiclonal CLL cohort. The number of detected IGH clonotypes varied depending on the detection method employed in this study. Conventional Sanger sequencing, which is widely used in CLL diagnostics, identified a greater number of multiclonal cases overall compared to NGS methods. However, NGS technologies, i.e., capture-based (LYNX panel) and amplicon-based (IG aNGS) approaches, although identifying fewer multiclonal cases, often detected additional productive IGH rearrangements within multiclonal samples that may be missed by Sanger sequencing. This methodological disparity can significantly impact the assessment of clonality in CLL, particularly in cases with low-frequency or multiple rearrangements.

As our study was based on the selection of multiclonal CLL cases identified by Sanger sequencing at the Centre of Molecular Biology and Genetics, University Hospital Brno, this method detected the highest number of multiclonal samples. Unlike NGS, which was performed on DNA, Sanger sequencing is conducted on RNA as an input material and does not apply a threshold for reliable IGH detection, so every detected IGH rearrangement is interpreted as clonal. This presents a limitation of this method, because the absence of a quantitative threshold means that the detected IGH rearrangements cannot be accurately quantified. Consequently, the size of the detected clones remains unclear, making their clinical interpretation ambiguous. In cases where Sanger sequencing indicated multiclonality, but NGS revealed only a single dominant clone, the additional IGH rearrangements were either absent from the NGS data or fell below the predefined thresholds for confident detection. This discrepancy suggests that some IGH rearrangements detected by Sanger sequencing may represent low-frequency subclones with relatively high IGH RNA expression levels, which is why they are detected only by Sanger sequencing. These

differences may also stem from technical factors such as differences in primer binding affinity or preferential amplification of certain sequences over others.

Among multiclonal CLL cases, biclonal samples were the most frequently observed across all three methods. Although tri-, tetra-, and more complex clonal architectures are relatively uncommon, their detection increased notably with the use of NGS. The capture-based LYNX panel and the amplicon-based IG aNGS approach showed the highest level of concordance in results and offered complementary advantages in resolving clonal architecture, particularly in identifying subclonal rearrangements that may go undetected with low-throughput methods. These observations highlight the superior resolution and sensitivity of NGS for IGH clonality assessment. Overall, our findings underscore the value of high-throughput sequencing in providing a more comprehensive and precise characterization of the IGH repertoire in CLL, especially in cases involving low abundance clones or multiclonality.

However, more precise characterization of the IGH repertoire introduces another challenge: there is currently no strict threshold for defining what constitutes a malignant CLL clone, or at minimum, a MBL clone, for reporting purposes. Many of the identified minor clones may represent non-malignant, highly reactive B-cell populations that developed in response to antigenic stimulation over the patient's lifetime. Moreover, the literature (Langerak et al., 2017) and published guidelines on IGH analysis (Langerak et al., 2011; Rosenquist et al., 2017) emphasize the need for careful interpretation of multiple clone findings, without clearly defining which criteria should be used to reliably identify true CLL clones. Thus, finding multiple expanded clones within the CD5+ B-cells with high-throughput methods is very often in patients with CLL, indicating that leukemogenesis is a multiclonal process that likely involves competition among B-cells with special BCR features (Bagnara et al., 2025).

One of the NGS approaches discussed in this thesis is the capture-based method using the LYNX panel. Our study demonstrates that this approach is reliable in detecting the majority of productive IGH rearrangements, reinforcing its utility for capturing clinically relevant clonotypes. However, this method may be limited by the efficiency of probes hybridization with IGH genes. Since IGHV genes undergo somatic hypermutation (SHM),

the probes may not hybridize effectively with mutated IGHV regions. As a result, fewer mutated IGHV rearrangements may be detected, or they may be detected with lower-than-true abundance. Our present findings strongly suggest this limitation. Therefore, interpreting the clonal abundances of IGH clones in samples can be ambiguous, particularly when mutated IGHV rearrangements are present. This may also be one of the reasons for the differing IGH clonality results observed when comparing the capture-based LYNX approach with the amplicon-based IG-aNGS method, in addition to technical factors such as assay sensitivity and lower coverage rates.

The other NGS approach discussed in this thesis is the amplicon-based IG-aNGS method. IG aNGS consistently detected a greater number of productive IGH rearrangements, resulting in the classification of more samples as multiclonal compared to the LYNX panel. This difference underscores the superior sensitivity of the IG aNGS approach in identifying low abundance clonotypes that may remain undetected by other methods. Therefore, it provides a substantial advantage in capturing minor subclonal populations. However, the size of IGH clonotypes may be skewed due to the amplification bias, where certain sequences are preferentially amplified over others. As we have shown, the abundance of clonotypes can also be skewed by the excessive amplification of certain unproductive IGH rearrangement. Moreover, variations in amplification efficiency across different rearrangements can lead to unequal representation, thereby compromising accurate quantification (Langerak et al., 2017). The differing IGH clonality results observed when comparing the capture-based LYNX panel may also be influenced by other technical factors, such as assay sensitivity and varying coverage rates.

To sum up, both capture-based and amplicon-based NGS approaches are reliable for detecting productive IGH rearrangements and capturing clinically relevant clonotypes. These methods provide highly overlapping results with enhanced resolution and sensitivity, enabling detailed characterization of the IG clonal architecture in CLL. This advantage is particularly important in clinical and research settings, where the detection of minor subclones may impact disease monitoring, prognosis, and treatment decision-making. However, the lack of a strict thresholds for defining what constitutes a malignant CLL clone complicates result interpretation. Therefore, analysis by Sanger sequencing simplify the

diagnostic process and data interpretation. Findings from our study highlight the importance of considering methodological limitations when interpreting clonality results. They also support the value of validating multiclonal findings with orthogonal techniques capable of detecting minor clonal populations with greater confidence.

Besides comparing individual methods for detecting IGH rearrangements, we also had the opportunity to investigate longitudinal evolution in clonal architecture within a subset of multiclonal CLL patients with available serial samples over time. Among these patients, 15/20 (75%) exhibited alterations in the total number of detected productive IGH rearrangements between their earliest and most recent samples, underscoring the dynamic nature of clonal composition in CLL. Notably, in 13 of the 15 patients (86.7%), a reduction in the number of detected productive IGH rearrangements was observed, suggesting a contraction of clonal diversity. This decline likely reflects the clonal dominance of one or a few leukemic B-cell populations, potentially driven by selective pressures such as therapeutic intervention or microenvironmental factors that may favour the expansion of more aggressive or treatment-resistant clones (Landau et al., 2017). Such patterns are consistent with the established concept of clonal evolution and selection in CLL, wherein fitter clones outcompete others over time.

Moreover, in 4 of the 13 patients (30.8%) with decreasing clonal diversity, we observed more complex dynamics. In these cases, the major IGH rearrangement was eventually replaced by another IGH rearrangement at a later sample, indicating that not only are minor clones capable of persisting, but they may also expand to overtake the initially dominant clone. This shift could be driven by selection advantages of the minor IGH rearrangement, potentially due to different genomic aberrations or gene mutations. Alternatively, the BCR of that clone may exhibit more distinctive responses to antigenic stimuli.

Conversely, in 2 of the 15 patients (13.3%), new productive IGH rearrangements emerged at later time points. This finding suggests the potential development of novel B-cell clones or the expansion of previously undetectable minor subclones, possibly as a result of disease progression or treatment-related selective escape. Together, these observations highlight the importance of longitudinal monitoring using sensitive, high-resolution

methods such as NGS to capture the evolving clonal landscape in CLL and its potential clinical implications.

In 13 of the 15 cases (86.7%), the overall IGHV mutational status of the patient remained stable over time, even when a change in the number of detectable productive IGH rearrangements was observed. This stability is likely due to the fact that, in patients with concordant IGHV mutational status, the loss of one IGH rearrangement does not alter the overall mutational classification, as both rearrangements share the same mutational status. However, we identified cases with a change in IGHV mutational status over time. This shift typically involved either the disappearance of a previously detected mutated rearrangement or the appearance of a new unmutated rearrangement, leading to an overall transition toward an unmutated or discordant IGHV profile. This change implies a possible shift toward a more aggressive disease phenotype and worse prognosis. These findings are consistent with observations reported by Plevova et al. (2014), who also documented longitudinal shifts toward unmutated IGHV status, further highlighting the potential clinical relevance of clonal dynamics and the necessity for ongoing molecular monitoring in CLL.

6 Conclusion

Genomic composition of leukemic populations in patients with multiclonal chronic lymphocytic leukemia

In 2–3% of all chronic lymphocytic leukemia (CLL) patients, clonal expansion and coexistence of distinct clonal B-lymphocyte populations occur, characterized by the expression of multiple, most commonly double, productive immunoglobulin heavy chain (IGH) gene rearrangements. In this master's thesis, we analysed a cohort of 76 CLL patients, each with a history of multiple productive IGH rearrangements detected during diagnostic work-up using Sanger sequencing. All patients were subsequently analysed using next-generation sequencing (NGS) with two orthogonal methods: capture-based NGS using the LYNX panel and amplicon-based NGS of IGH rearrangements (IG aNGS).

The experimental part of this thesis presents a quantitative comparison of IGH rearrangement repertoires in monoclonal and multiclonal CLL, along with the comparison of IGH clonotypes between individual detection methods. The results show no significant differences in overall IGH gene usage between monoclonal and multiclonal CLL, and the distribution of IGH gene usage was consistent across all tested methods within the multiclonal CLL cohort. However, the number of detected IGH clonotypes varied notably depending on the detection method employed. While Sanger sequencing identified multiple CLL clones in a greater number of samples, NGS-based approaches demonstrated significantly higher sensitivity in detecting low-frequency minor clones.

A key challenge highlighted by these findings is the absence of a standardized and reliable threshold across methods for defining what constitutes a malignant CLL clone for reporting purposes. Although the literature emphasizes the need for careful interpretation of multiple clone findings, clear and universally accepted criteria for reliably identifying a clinically relevant CLL clone remain undefined. Therefore, this master's thesis contributes important insights into the detection and characterization of multiclonal CLL, supporting the development of more accurate interpretive guidelines in future diagnostic and clinical practice.

Bibliography

Agathangelidis, A., Chatzidimitriou, A., Chatzikonstantinou, T., Tresoldi, C., Davis, Z., Giudicelli, V., Kossida, S., Belessi, C., Rosenquist, R., Ghia, P., Langerak, A.W., Davi, F., Stamatopoulos, K., ERIC, the European Research Initiative on CLL, 2022. Immunoglobulin gene sequence analysis in chronic lymphocytic leukemia: the 2022 update of the recommendations by ERIC, the European Research Initiative on CLL. Leukemia 36, 1961–1968. https://doi.org/10.1038/s41375-022-01604-2

Agathangelidis, A., Chatzidimitriou, A., Gemenetzi, K., Giudicelli, V., Karypidou, M., Plevova, K., Davis, Z., Yan, X.-J., Jeromin, S., Schneider, C., Pedersen, L.B., Tschumper, R.C., Sutton, L.-A., Baliakas, P., Scarfò, L., van Gastel, E.J., Armand, M., Tausch, E., Biderman, B., Baer, C., Bagnara, D., Navarro, A., Langlois de Septenville, A., Guido, V., Mitterbauer-Hohendanner, G., Dimovski, A., Brieghel, C., Lawless, S., Meggendorfer, M., Brazdilova, K., Ritgen, M., Facco, M., Tresoldi, C., Visentin, A., Patriarca, A., Catherwood, M., Bonello, L., Sudarikov, A., Vanura, K., Roumelioti, M., Skuhrova Francova, H., Moysiadis, T., Veronese, S., Giannopoulos, K., Mansouri, L., Karan-Djurasevic, T., Sandaltzopoulos, R., Bödör, C., Fais, F., Kater, A.P., Panovska, I., Rossi, D., Alshemmari, S., Panagiotidis, P., Costeas, P., Espinet, B., Antic, D., Foroni, L., Montillo, M., Trentin, L., Stavroyianni, N., Gaidano, G., Francia di Celle, P., Niemann, C., Campo, E., Anagnostopoulos, A., Pott, C., Fischer, K., Hallek, M., Oscier, D., Stilgenbauer, S., Haferlach, C., Jelinek, D., Chiorazzi, N., Pospisilova, S., Lefranc, M.-P., Kossida, S., Langerak, A.W., Belessi, C., Davi, F., Rosenquist, R., Ghia, P., Stamatopoulos, K., 2021. Higher-order connections between stereotyped subsets: implications for improved patient classification in CLL. Blood 137, 1365–1376. https://doi.org/10.1182/blood.2020007039

Agathangelidis, A., Darzentas, N., Hadzidimitriou, A., Brochet, X., Murray, F., Yan, X.-J., Davis, Z., van Gastel-Mol, E.J., Tresoldi, C., Chu, C.C., Cahill, N., Giudicelli, V., Tichy, B., Pedersen, L.B., Foroni, L., Bonello, L., Janus, A., Smedby, K., Anagnostopoulos, A., Merle-Beral, H., Laoutaris, N., Juliusson, G., di Celle, P.F., Pospisilova, S., Jurlander, J., Geisler, C., Tsaftaris, A., Lefranc, M.-P., Langerak, A.W., Oscier, D.G., Chiorazzi, N., Belessi, C., Davi, F., Rosenquist, R., Ghia, P., Stamatopoulos, K., 2012. Stereotyped B-cell receptors in one-third of chronic lymphocytic leukemia: a molecular classification with implications for targeted therapies. Blood 119, 4467–4475. https://doi.org/10.1182/blood-2011-11-393694

Allen, D., Cumano, A., Dildrop, R., Kocks, C., Rajewsky, K., Rajewsky, N., Roes, J., Sablitzky, F., Siekevitz, M., 1987. Timing, genetic requirements and functional

consequences of somatic hypermutation during B-cell development. Immunol. Rev. 96, 5–22. https://doi.org/10.1111/j.1600-065x.1987.tb00506.x

Alt, F.W., Yancopoulos, G.D., Blackwell, T.K., Wood, C., Thomas, E., Boss, M., Coffman, R., Rosenberg, N., Tonegawa, S., Baltimore, D., 1984. Ordered rearrangement of immunoglobulin heavy chain variable region segments. EMBO J. 3, 1209–1219. https://doi.org/10.1002/j.1460-2075.1984.tb01955.x

Bagnara, D., Mazzarello, A.N., Cardente, N., Vergani, S., Kasar, S., Fernandes, S., Ferrer, G., Ghiotto, F., Barrientos, J.C., Kolitz, J.E., Rai, K.R., Alen, S.L., Colombo, M., Fais, F., Brown, J.R., Ferrarini, M., Chiorazzi, N., 2025. Chronic lymphocytic leukemia often arises by a multiclonal selection process. Haematologica. https://doi.org/10.3324/haematol.2024.286380

Baliakas, P., Iskas, M., Gardiner, A., Davis, Z., Plevova, K., Nguyen-Khac, F., Malcikova, J., Anagnostopoulos, A., Glide, S., Mould, S., Stepanovska, K., Brejcha, M., Belessi, C., Davi, F., Pospisilova, S., Athanasiadou, A., Stamatopoulos, K., Oscier, D., 2014. Chromosomal translocations and karyotype complexity in chronic lymphocytic leukemia: a systematic reappraisal of classic cytogenetic data. Am. J. Hematol. 89, 249–255. https://doi.org/10.1002/ajh.23618

Bassig, B.A., Au, W.-Y., Mang, O., Ngan, R., Morton, L.M., Ip, D.K.M., Hu, W., Zheng, T., Seow, W.J., Xu, J., Lan, Q., Rothman, N., 2016. Subtype-specific incidence rates of lymphoid malignancies in Hong Kong compared to the United States, 2001-2010. Cancer Epidemiol. 42, 15–23. https://doi.org/10.1016/j.canep.2016.02.007

Bassing, C.H., Swat, W., Alt, F.W., 2002. The mechanism and regulation of chromosomal V(D)J recombination. Cell 109 Suppl, S45-55. https://doi.org/10.1016/s0092-8674(02)00675-x

Berek, C., Berger, A., Apel, M., 1991. Maturation of the immune response in germinal centers. Cell 67, 1121–1129. https://doi.org/10.1016/0092-8674(91)90289-b

Berek, C., Milstein, C., 1987. Mutation Drift and Repertoire Shift in the Maturation of the Immune Response. Immunol. Rev. 96, 23–41. https://doi.org/10.1111/j.1600-065X.1987.tb00507.x

Binet, J.L., Auquier, A., Dighiero, G., Chastang, C., Piguet, H., Goasguen, J., Vaugier, G., Potron, G., Colona, P., Oberling, F., Thomas, M., Tchernia, G., Jacquillat, C., Boivin, P., Lesty, C., Duault, M.T., Monconduit, M., Belabbes, S., Gremy, F., 1981. A new prognostic classification of chronic lymphocytic leukemia derived from a multivariate

survival analysis. Cancer 48, 198–206. https://doi.org/10.1002/1097-0142(19810701)48:1<198::aid-cncr2820480131>3.0.co;2-v

Bomben, R., Dal Bo, M., Capello, D., Forconi, F., Maffei, R., Laurenti, L., Rossi, D., Del Principe, M.I., Zucchetto, A., Bertoni, F., Rossi, F.M., Bulian, P., Cattarossi, I., Ilariucci, F., Sozzi, E., Spina, V., Zucca, E., Degan, M., Lauria, F., Del Poeta, G., Efremov, D.G., Marasca, R., Gaidano, G., Gattei, V., 2009. Molecular and clinical features of chronic lymphocytic leukaemia with stereotyped B cell receptors: results from an Italian multicentre study. Br. J. Haematol. 144, 492–506. https://doi.org/10.1111/j.1365-2141.2008.07469.x

Brazdilova, K., Plevova, K., Skuhrova Francova, H., Kockova, H., Borsky, M., Bikos, V., Malcikova, J., Oltova, A., Kotaskova, J., Tichy, B., Brychtova, Y., Mayer, J., Doubek, M., Pospisilova, S., 2018. Multiple productive IGH rearrangements denote oligoclonality even in immunophenotypically monoclonal CLL. Leukemia 32, 234–236. https://doi.org/10.1038/leu.2017.274

Burrows, P.D., Lebien, T., Zhang, Z., Davis, R.S., Cooper, M.D., 2004. The Development of Human B Lymphocytes, in: Molecular Biology of B Cells. Elsevier Inc., pp. 141–154. https://doi.org/10.1016/B978-012053641-2/50010-1

Cabaniols, J.P., Fazilleau, N., Casrouge, A., Kourilsky, P., Kanellopoulos, J.M., 2001. Most alpha/beta T cell receptor diversity is due to terminal deoxynucleotidyl transferase. J. Exp. Med. 194, 1385–1390. https://doi.org/10.1084/jem.194.9.1385

Campbell, K.S., Cambier, J.C., 1990. B lymphocyte antigen receptors (mIg) are non-covalently associated with a disulfide linked, inducibly phosphorylated glycoprotein complex. EMBO J. 9, 441–448. https://doi.org/10.1002/j.1460-2075.1990.tb08129.x

Campbell, K.S., Hager, E.J., Friedrich, R.J., Cambier, J.C., 1991. IgM antigen receptor complex contains phosphoprotein products of B29 and mb-1 genes. Proc. Natl. Acad. Sci. U. S. A. 88, 3982–3986. https://doi.org/10.1073/pnas.88.9.3982

Chang, H., Cerny, J., 2006. Molecular characterization of chronic lymphocytic leukemia with two distinct cell populations: Evidence for separate clonal origins. Am. J. Clin. Pathol. 126, 23–28. https://doi.org/10.1309/0YYF-17GF-KFJF-NP5G

Chaudhuri, J., Alt, F.W., 2004. Class-switch recombination: interplay of transcription, DNA deamination and DNA repair. Nat. Rev. Immunol. 4, 541–552. https://doi.org/10.1038/nri1395

Chaudhuri, J., Tian, M., Khuong, C., Chua, K., Pinaud, E., Alt, F.W., 2003. Transcription-targeted DNA deamination by the AID antibody diversification enzyme. Nature 422, 726–730. https://doi.org/10.1038/nature01574

Chen, J., Trounstine, M., Kurahara, C., Young, F., Kuo, C.C., Xu, Y., Loring, J.F., Alt, F.W., Huszar, D., 1993. B cell development in mice that lack one or both immunoglobulin kappa light chain genes. EMBO J. 12, 821–830. https://doi.org/10.1002/j.1460-2075.1993.tb05722.x

Chi, X., Li, Y., Qiu, X., 2020. V(D)J recombination, somatic hypermutation and class switch recombination of immunoglobulins: mechanism and regulation. Immunology 160, 233–247. https://doi.org/10.1111/imm.13176

Chiorazzi, N., Chen, S.-S., Rai, K.R., 2021. Chronic Lymphocytic Leukemia. Cold Spring Harb. Perspect. Med. 11, a035220. https://doi.org/10.1101/cshperspect.a035220

Cook, G.P., Tomlinson, I.M., 1995. The human immunoglobulin VH repertoire. Immunol. Today 16, 237–242. https://doi.org/10.1016/0167-5699(95)80166-9

Corfe, S.A., Paige, C.J., 2009. Development of B Lymphocytes, in: Wickrema, A., Kee, B. (Eds.), Molecular Basis of Hematopoiesis. Springer, New York, NY, pp. 173–199. https://doi.org/10.1007/978-0-387-85816-6_8

Cramer, P., Hallek, M., 2011. Prognostic factors in chronic lymphocytic leukemia-what do we need to know? Nat. Rev. Clin. Oncol. 8, 38–47. https://doi.org/10.1038/nrclinonc.2010.167

Critchlow, S.E., Bowater, R.P., Jackson, S.P., 1997. Mammalian DNA double-strand break repair protein XRCC4 interacts with DNA ligase IV. Curr. Biol. CB 7, 588–598. https://doi.org/10.1016/s0960-9822(06)00258-2

Damle, R.N., Wasil, T., Fais, F., Ghiotto, F., Valetto, A., Allen, S.L., Buchbinder, A., Budman, D., Dittmar, K., Kolitz, J., Lichtman, S.M., Schulman, P., Vinciguerra, V.P., Rai, K.R., Ferrarini, M., Chiorazzi, N., 1999. Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. Blood 94, 1840–1847.

Davi, F., Langerak, A.W., de Septenville, A.L., Kolijn, P.M., Hengeveld, P.J., Chatzidimitriou, A., Bonfiglio, S., Sutton, L.-A., Rosenquist, R., Ghia, P., Stamatopoulos, K., 2020. Immunoglobulin gene analysis in chronic lymphocytic leukemia

in the era of next generation sequencing. Leukemia 34, 2545–2551. https://doi.org/10.1038/s41375-020-0923-9

Di Noia, J.M., Neuberger, M.S., 2007. Molecular mechanisms of antibody somatic hypermutation. Annu. Rev. Biochem. 76, 1–22. https://doi.org/10.1146/annurev.biochem.76.061705.090740

Döhner, H., Fischer, K., Bentz, M., Hansen, K., Benner, A., Cabot, G., Diehl, D., Schlenk, R., Coy, J., Stilgenbauer, S., Volkmann, M., Galle, P.R., Poustka, A., Hunstein, W., Lichter, P., 1995. p53 gene deletion predicts for poor survival and non-response to therapy with purine analogs in chronic B-cell leukemias. Blood 85, 1580–1589.

Döhner, H., Stilgenbauer, S., Benner, A., Leupolt, E., Kröber, A., Bullinger, L., Döhner, K., Bentz, M., Lichter, P., 2000. Genomic aberrations and survival in chronic lymphocytic leukemia. N. Engl. J. Med. 343, 1910–1916. https://doi.org/10.1056/NEJM200012283432602

Dong, Y., Pi, X., Bartels-Burgahn, F., Saltukoglu, D., Liang, Z., Yang, J., Alt, F.W., Reth, M., Wu, H., 2022. Structural principles of B cell antigen receptor assembly. Nature 612, 156–161. https://doi.org/10.1038/s41586-022-05412-7

Dörner, T., Foster, S.J., Farner, N.L., Lipsky, P.E., 1998. Somatic hypermutation of human immunoglobulin heavy chain genes: targeting of RGYW motifs on both DNA strands. Eur. J. Immunol. 28, 3384–3396. https://doi.org/10.1002/(SICI)1521-4141(199810)28:10<3384::AID-IMMU3384>3.0.CO;2-T

Dudley, D.D., Chaudhuri, J., Bassing, C.H., Alt, F.W., 2005. Mechanism and control of V(D)J recombination versus class switch recombination: similarities and differences. Adv. Immunol. 86, 43–112. https://doi.org/10.1016/S0065-2776(04)86002-4

Dunham, I., Shimizu, N., Roe, B.A., Chissoe, S., Hunt, A.R., Collins, J.E., Bruskiewich, R., Beare, D.M., Clamp, M., Smink, L.J., Ainscough, R., Almeida, J.P., Babbage, A., Bagguley, C., Bailey, J., Barlow, K., Bates, K.N., Beasley, O., Bird, C.P., Blakey, S., Bridgeman, A.M., Buck, D., Burgess, J., Burrill, W.D., O'Brien, K.P., 1999. The DNA sequence of human chromosome 22. Nature 402, 489–495. https://doi.org/10.1038/990031

Dunnick, W., Hertz, G.Z., Scappino, L., Gritzmacher, C., 1993. DNA sequences at immunoglobulin switch region recombination sites. Nucleic Acids Res. 21, 365–372. https://doi.org/10.1093/nar/21.3.365

Ehlich, A., Schaal, S., Gu, H., Kitamura, D., Müller, W., Rajewsky, K., 1993. Immunoglobulin heavy and light chain genes rearrange independently at early stages of B cell development. Cell 72, 695–704. https://doi.org/10.1016/0092-8674(93)90398-a

Eichhorst, B., Robak, T., Montserrat, E., Ghia, P., Niemann, C.U., Kater, A.P., Gregor, M., Cymbalista, F., Buske, C., Hillmen, P., Hallek, M., Mey, U., ESMO Guidelines Committee, 2021. Chronic lymphocytic leukaemia: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann. Oncol. Off. J. Eur. Soc. Med. Oncol. 32, 23–33. https://doi.org/10.1016/j.annonc.2020.09.019

Emanuel, B.S., Cannizzaro, L.A., Magrath, I., Tsujimoto, Y., Nowell, P.C., Croce, C.M., 1985. Chromosomal orientation of the lambda light chain locus: V lambda is proximal to C lambda in 22q11. Nucleic Acids Res. 13, 381–387. https://doi.org/10.1093/nar/13.2.381

Ermert, K., Mitlöhner, H., Schempp, W., Zachau, H.G., 1995. The immunoglobulin κ locus of primates. Genomics 25, 623–629. https://doi.org/10.1016/0888-7543(95)80003-5

Flaswinkel, H., Reth, M., 1994. Dual role of the tyrosine activation motif of the Ig-alpha protein during signal transduction via the B cell antigen receptor. EMBO J. 13, 83–89. https://doi.org/10.1002/j.1460-2075.1994.tb06237.x

Frippiat, J.P., Williams, S.C., Tomlinson, I.M., Cook, G.P., Cherif, D., Le Paslier, D., Collins, J.E., Dunham, I., Winter, G., Lefranc, M.P., 1995. Organization of the human immunoglobulin lambda light-chain locus on chromosome 22q11.2. Hum. Mol. Genet. 4, 983–991. https://doi.org/10.1093/hmg/4.6.983

Fugmann, S.D., Lee, A.I., Shockett, P.E., Villey, I.J., Schatz, D.G., 2000. The RAG proteins and V(D)J recombination: complexes, ends, and transposition. Annu. Rev. Immunol. 18, 495–527. https://doi.org/10.1146/annurev.immunol.18.1.495

Furman, R.R., 2010. Prognostic markers and stratification of chronic lymphocytic leukemia. Hematol. Am. Soc. Hematol. Educ. Program 2010, 77–81. https://doi.org/10.1182/asheducation-2010.1.77

Gay, D., Saunders, T., Camper, S., Weigert, M., 1993. Receptor editing: an approach by autoreactive B cells to escape tolerance. J. Exp. Med. 177, 999–1008. https://doi.org/10.1084/jem.177.4.999

Ghia, P., Stamatopoulos, K., Belessi, C., Moreno, C., Stilgenbauer, S., Stevenson, F., Davi, F., Rosenquist, R., European Research Initiative on CLL, 2007. ERIC

recommendations on IGHV gene mutational status analysis in chronic lymphocytic leukemia. Leukemia 21, 1–3. https://doi.org/10.1038/sj.leu.2404457

Gonzalez, D., Martinez, P., Wade, R., Hockley, S., Oscier, D., Matutes, E., Dearden, C.E., Richards, S.M., Catovsky, D., Morgan, G.J., 2011. Mutational status of the TP53 gene as a predictor of response and survival in patients with chronic lymphocytic leukemia: results from the LRF CLL4 trial. J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. 29, 2223–2229. https://doi.org/10.1200/JCO.2010.32.0838

González-Campos, J., Ríos-Herranz, E., De Blas-Orlando, J.M., Martín-Noya, A., Parody-Ruiz-Berdejo, R., Rodríguez-Fernández, J.M., 1997. Chronic lymphocytic leukemia with two cellular populations: a biphenotypic or biclonal disease. Ann. Hematol. 74, 243–246. https://doi.org/10.1007/s002770050292

Goossens, T., Klein, U., Küppers, R., 1998. Frequent occurrence of deletions and duplications during somatic hypermutation: implications for oncogene translocations and heavy chain disease. Proc. Natl. Acad. Sci. U. S. A. 95, 2463–2468. https://doi.org/10.1073/pnas.95.5.2463

Gorman, J.R., Alt, F.W., 1998. Regulation of immunoglobulin light chain isotype expression. Adv. Immunol. 69, 113–181. https://doi.org/10.1016/s0065-2776(08)60607-0

Grawunder, U., Wilm, M., Wu, X., Kulesza, P., Wilson, T.E., Mann, M., Lieber, M.R., 1997. Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. Nature 388, 492–495. https://doi.org/10.1038/41358

Grawunder, U., Zimmer, D., Fugmann, S., Schwarz, K., Lieber, M.R., 1998. DNA ligase IV is essential for V(D)J recombination and DNA double-strand break repair in human precursor lymphocytes. Mol. Cell 2, 477–484. https://doi.org/10.1016/s1097-2765(00)80147-1

Haferlach, C., Dicker, F., Schnittger, S., Kern, W., Haferlach, T., 2007. Comprehensive genetic characterization of CLL: a study on 506 cases analysed with chromosome banding analysis, interphase FISH, IgV(H) status and immunophenotyping. Leukemia 21, 2442–2451. https://doi.org/10.1038/sj.leu.2404935

Hallek, M., 2025. Chronic Lymphocytic Leukemia: 2025 Update on the Epidemiology, Pathogenesis, Diagnosis, and Therapy. Am. J. Hematol. 100, 450–480. https://doi.org/10.1002/ajh.27546

Hallek, M., Cheson, B.D., Catovsky, D., Caligaris-Cappio, F., Dighiero, G., Döhner, H., Hillmen, P., Keating, M., Montserrat, E., Chiorazzi, N., Stilgenbauer, S., Rai, K.R., Byrd, J.C., Eichhorst, B., O'Brien, S., Robak, T., Seymour, J.F., Kipps, T.J., 2018. iwCLL guidelines for diagnosis, indications for treatment, response assessment, and supportive management of CLL. Blood 131, 2745–2760. https://doi.org/10.1182/blood-2017-09-806398

Hallek, M., German CLL Study Group, 2005. Chronic lymphocytic leukemia (CLL): first-line treatment. Hematol. Am. Soc. Hematol. Educ. Program 285–291. https://doi.org/10.1182/asheducation-2005.1.285

Hamblin, T.J., Davis, Z., Gardiner, A., Oscier, D.G., Stevenson, F.K., 1999. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. Blood 94, 1848–1854.

Hermanson, G.G., Eisenberg, D., Kincade, P.W., Wall, R., 1988. B29: a member of the immunoglobulin gene superfamily exclusively expressed on beta-lineage cells. Proc. Natl. Acad. Sci. U. S. A. 85, 6890–6894.

Hieter, P.A., Hollis, G.F., Korsmeyer, S.J., Waldmann, T.A., Leder, P., 1981. Clustered arrangement of immunoglobulin lambda constant region genes in man. Nature 294, 536–540. https://doi.org/10.1038/294536a0

Hiom, K., Gellert, M., 1998. Assembly of a 12/23 paired signal complex: a critical control point in V(D)J recombination. Mol. Cell 1, 1011–1019. https://doi.org/10.1016/s1097-2765(00)80101-x

Hombach, J., Lottspeich, F., Reth, M., 1990a. Identification of the genes encoding the IgM-alpha and Ig-beta components of the IgM antigen receptor complex by amino-terminal sequencing. Eur. J. Immunol. 20, 2795–2799. https://doi.org/10.1002/eji.1830201239

Hombach, J., Tsubata, T., Leclercq, L., Stappert, H., Reth, M., 1990b. Molecular components of the B-cell antigen receptor complex of the IgM class. Nature 343, 760–762. https://doi.org/10.1038/343760a0

Honjo, T., 1983. Immunoglobulin genes. Annu. Rev. Immunol. 1, 499–528. https://doi.org/10.1146/annurev.iy.01.040183.002435

Hotinski, A.K., Best, O.G., Kuss, B.J., 2021. The future of laboratory testing in chronic lymphocytic leukaemia. Pathology (Phila.), Malignant haematology 2021: impact of recent

advances on the diagnostic laboratory 53, 377–384. https://doi.org/10.1016/j.pathol.2021.01.006

Hsi, E.D., Hoeltge, G., Tubbs, R.R., 2000. Biclonal chronic lymphocytic leukemia. Am. J. Clin. Pathol. 113, 798–804. https://doi.org/10.1309/V8AN-A2XP-7TDV-HR0T

Huber, C., Huber, E., Lautner-Rieske, A., Schäble, K.F., Zachau, H.G., 1993. The human immunoglobulin kappa locus. Characterization of the partially duplicated L regions. Eur. J. Immunol. 23, 2860–2867. https://doi.org/10.1002/eji.1830231120

Hulkkonen, J., Vilpo, L., Hurme, M., Vilpo, J., 2002. Surface antigen expression in chronic lymphocytic leukemia: clustering analysis, interrelationships and effects of chromosomal abnormalities. Leukemia 16, 178–185. https://doi.org/10.1038/sj.leu.2402363

International CLL IPI Working Group, 2016. An international prognostic index for patients with chronic lymphocytic leukaemia (CLL-IPI): a meta-analysis of individual patient data. Lancet Oncol. 17, 779–790. https://doi.org/10.1016/S1470-2045(16)30029-8

Jacobs, H., Bross, L., 2001. Towards an understanding of somatic hypermutation. Curr. Opin. Immunol. 13, 208–218. https://doi.org/10.1016/s0952-7915(00)00206-5

Janeway, C., Travers, P., Walport, M., Shlomchik, M.J., 2001. Immunobiology 5: the immune system in health and disease, 5th ed. Garland Pub, New York.

Karan-Djurasevic, T., Pavlovic, S., Karan-Djurasevic, T., Pavlovic, S., 2017. Somatic Hypermutational Status and Gene Repertoire of Immunoglobulin Rearrangements in Chronic Lymphocytic Leukemia, in: Lymphocyte Updates - Cancer, Autoimmunity and Infection. IntechOpen. https://doi.org/10.5772/intechopen.69110

Kawasaki, K., Minoshima, S., Nakato, E., Shibuya, K., Shintani, A., Asakawa, S., Sasaki, T., Klobeck, H.G., Combriato, G., Zachau, H.G., Shimizu, N., 2001. Evolutionary dynamics of the human immunoglobulin kappa locus and the germline repertoire of the Vkappa genes. Eur. J. Immunol. 31, 1017–1028.

Kawasaki, K., Minoshima, S., Nakato, E., Shibuya, K., Shintani, A., Schmeits, J.L., Wang, J., Shimizu, N., 1997. One-megabase sequence analysis of the human immunoglobulin lambda gene locus. Genome Res. 7, 250–261. https://doi.org/10.1101/gr.7.3.250

Kawasaki, K., Minoshima, S., Shimizu, N., 2000. Propagation and maintenance of the 119 human immunoglobulin Vlambda genes and pseudogenes during evolution. J. Exp. Zool.

288, 120–134. https://doi.org/10.1002/1097-010x(20000815)288:2<120::aid-jez4>3.0.co;2-i

Khanna, K.K., Jackson, S.P., 2001. DNA double-strand breaks: signaling, repair and the cancer connection. Nat. Genet. 27, 247–254. https://doi.org/10.1038/85798

Kirsch, I.R., Morton, C.C., Nakahara, K., Leder, P., 1982. Human immunoglobulin heavy chain genes map to a region of translocations in malignant B lymphocytes. Science 216, 301–303. https://doi.org/10.1126/science.6801764

Kocks, C., Rajewsky, K., 1989. Stable Expression and Somatic Hypermutation of Antibody V Regions in B-Cell Developmental Pathways. Annu. Rev. Immunol. 7, 537–559. https://doi.org/10.1146/annurev.iy.07.040189.002541

Komori, T., Okada, A., Stewart, V., Alt, F.W., 1993. Lack of N Regions in Antigen Receptor Variable Region Genes of TdT-Deficient Lymphocytes. Science. https://doi.org/10.1126/science.8356451

Korsmeyer, S.J., Hieter, P.A., Ravetch, J.V., Poplack, D.G., Waldmann, T.A., Leder, P., 1981. Developmental hierarchy of immunoglobulin gene rearrangements in human leukemic pre-B-cells. Proc. Natl. Acad. Sci. U. S. A. 78, 7096–7100. https://doi.org/10.1073/pnas.78.11.7096

Kriangkum, J., Motz, S.N., Mack, T., Beiggi, S., Baigorri, E., Kuppusamy, H., Belch, A.R., Johnston, J.B., Pilarski, L.M., 2015. Single-Cell Analysis and Next-Generation Immuno-Sequencing Show That Multiple Clones Persist in Patients with Chronic Lymphocytic Leukemia. PloS One 10, e0137232. https://doi.org/10.1371/journal.pone.0137232

Kubagawa, H., Cooper, M.D., Carroll, A.J., Burrows, P.D., 1989. Light-chain gene expression before heavy-chain gene rearrangement in pre-B cells transformed by Epstein-Barr virus. Proc. Natl. Acad. Sci. U. S. A. 86, 2356–2360. https://doi.org/10.1073/pnas.86.7.2356

Landau, D.A., Sun, C., Rosebrock, D., Herman, S.E.M., Fein, J., Sivina, M., Underbayev, C., Liu, D., Hoellenriegel, J., Ravichandran, S., Farooqui, M.Z.H., Zhang, W., Cibulskis, C., Zviran, A., Neuberg, D.S., Livitz, D., Bozic, I., Leshchiner, I., Getz, G., Burger, J.A., Wiestner, A., Wu, C.J., 2017. The evolutionary landscape of chronic lymphocytic leukemia treated with ibrutinib targeted therapy. Nat. Commun. 8, 2185. https://doi.org/10.1038/s41467-017-02329-y

Landau, D.A., Tausch, E., Taylor-Weiner, A.N., Stewart, C., Reiter, J.G., Bahlo, J., Kluth, S., Bozic, I., Lawrence, M., Böttcher, S., Carter, S.L., Cibulskis, K., Mertens, D., Sougnez, C.L., Rosenberg, M., Hess, J.M., Edelmann, J., Kless, S., Kneba, M., Ritgen, M., Fink, A., Fischer, K., Gabriel, S., Lander, E.S., Nowak, M.A., Döhner, H., Hallek, M., Neuberg, D., Getz, G., Stilgenbauer, S., Wu, C.J., 2015. Mutations driving CLL and their evolution in progression and relapse. Nature 526, 525–530. https://doi.org/10.1038/nature15395

Langerak, A.W., Brüggemann, M., Davi, F., Darzentas, N., van Dongen, J.J.M., Gonzalez, D., Cazzaniga, G., Giudicelli, V., Lefranc, M.-P., Giraud, M., Macintyre, E.A., Hummel, M., Pott, C., Groenen, P.J.T.A., Stamatopoulos, K., EuroClonality-NGS Consortium, 2017. High-Throughput Immunogenetics for Clinical and Research Applications in Immunohematology: Potential and Challenges. J. Immunol. Baltim. Md 1950 198, 3765–3774. https://doi.org/10.4049/jimmunol.1602050

Langerak, A.W., Davi, F., Ghia, P., Hadzidimitriou, A., Murray, F., Potter, K.N., Rosenquist, R., Stamatopoulos, K., Belessi, C., European Research Initiative on CLL (ERIC), 2011. Immunoglobulin sequence analysis and prognostication in CLL: guidelines from the ERIC review board for reliable interpretation of problematic cases. Leukemia 25, 979–984. https://doi.org/10.1038/leu.2011.49

Lanzavecchia, A., 1985. Antigen-specific interaction between T and B cells. Nature 314, 537–539. https://doi.org/10.1038/314537a0

Lazarian, G., Guièze, R., Wu, C.J., 2017. Clinical Implications of Novel Genomic Discoveries in Chronic Lymphocytic Leukemia. J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol. 35, 984–993. https://doi.org/10.1200/JCO.2016.71.0822

Lefranc, M.P., 2001a. Nomenclature of the human immunoglobulin heavy (IGH) genes. Exp. Clin. Immunogenet. 18, 100–116. https://doi.org/10.1159/000049189

Lefranc, M.P., 2001b. Nomenclature of the human immunoglobulin kappa (IGK) genes. Exp. Clin. Immunogenet. 18, 161–174. https://doi.org/10.1159/000049195

Lefranc, M.P., 2001c. Nomenclature of the human immunoglobulin lambda (IGL) genes. Exp. Clin. Immunogenet. 18, 242–254. https://doi.org/10.1159/000049203

Li, Z., Otevrel, T., Gao, Y., Cheng, H.L., Seed, B., Stamato, T.D., Taccioli, G.E., Alt, F.W., 1995. The XRCC4 gene encodes a novel protein involved in DNA double-strand break repair and V(D)J recombination. Cell 83, 1079–1089. https://doi.org/10.1016/0092-8674(95)90135-3

Liu, J.-C., Zhang, K., Zhang, X., Guan, F., Zeng, H., Kubo, M., Lee, P., Candotti, F., James, L.K., Camara, N.O.S., Benlagha, K., Lei, J.-H., Forsman, H., Yang, L., Xiao, W., Liu, Z., Liu, C.-H., 2024. Immunoglobulin class-switch recombination: Mechanism, regulation, and related diseases. MedComm 5, e662. https://doi.org/10.1002/mco2.662

Löffert, D., Ehlich, A., Müller, W., Rajewsky, K., 1996. Surrogate Light Chain Expression Is Required to Establish Immunoglobulin Heavy Chain Allelic Exclusion during Early B Cell Development. Immunity 4, 133–144. https://doi.org/10.1016/S1074-7613(00)80678-0

Ma, Y., Pannicke, U., Schwarz, K., Lieber, M.R., 2002. Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. Cell 108, 781–794. https://doi.org/10.1016/s0092-8674(02)00671-2

Manis, J.P., Tian, M., Alt, F.W., 2002. Mechanism and control of class-switch recombination. Trends Immunol. 23, 31–39. https://doi.org/10.1016/s1471-4906(01)02111-1

Marti, G.E., Rawstron, A.C., Ghia, P., Hillmen, P., Houlston, R.S., Kay, N., Schleinitz, T.A., Caporaso, N., International Familial CLL Consortium, 2005. Diagnostic criteria for monoclonal B-cell lymphocytosis. Br. J. Haematol. 130, 325–332. https://doi.org/10.1111/j.1365-2141.2005.05550.x

Matsuda, F., Ishii, K., Bourvagnet, P., Kuma, K. i, Hayashida, H., Miyata, T., Honjo, T., 1998. The complete nucleotide sequence of the human immunoglobulin heavy chain variable region locus. J. Exp. Med. 188, 2151–2162. https://doi.org/10.1084/jem.188.11.2151

Matutes, E., Wotherspoon, A., Catovsky, D., 2007. Differential diagnosis in chronic lymphocytic leukaemia. Best Pract. Res. Clin. Haematol. 20, 367–384. https://doi.org/10.1016/j.beha.2007.03.001

McBlane, J.F., van Gent, D.C., Ramsden, D.A., Romeo, C., Cuomo, C.A., Gellert, M., Oettinger, M.A., 1995. Cleavage at a V(D)J recombination signal requires only RAG1 and RAG2 proteins and occurs in two steps. Cell 83, 387–395. https://doi.org/10.1016/0092-8674(95)90116-7

McBride, O.W., Battey, J., Hollis, G.F., Swan, D.C., Siebenlist, U., Leder, P., 1982a. Localization of human variable and constant region immunoglobulin heavy chain genes on subtelomeric band q32 of chromosome 14. Nucleic Acids Res. 10, 8155–8170.

McBride, O.W., Hieter, P.A., Hollis, G.F., Swan, D., Otey, M.C., Leder, P., 1982b. Chromosomal location of human kappa and lambda immunoglobulin light chain constant region genes. J. Exp. Med. 155, 1480–1490. https://doi.org/10.1084/jem.155.5.1480

McConnell, S.A., Casadevall, A., 2025. New insights into antibody structure with implications for specificity, variable region restriction and isotype choice. Nat. Rev. Immunol. https://doi.org/10.1038/s41577-025-01150-9

McHeyzer-Williams, L.J., McHeyzer-Williams, M.G., 2005. Antigen-specific memory B cell development. Annu. Rev. Immunol. 23, 487–513. https://doi.org/10.1146/annurev.immunol.23.021704.115732

McKean, D., Huppi, K., Bell, M., Staudt, L., Gerhard, W., Weigert, M., 1984. Generation of antibody diversity in the immune response of BALB/c mice to influenza virus hemagglutinin. Proc. Natl. Acad. Sci. U. S. A. 81, 3180–3184. https://doi.org/10.1073/pnas.81.10.3180

Melchers, F., ten Boekel, E., Seidl, T., Kong, X.C., Yamagami, T., Onishi, K., Shimizu, T., Rolink, A.G., Andersson, J., 2000. Repertoire selection by pre-B-cell receptors and B-cell receptors, and genetic control of B-cell development from immature to mature B cells. Immunol. Rev. 175, 33–46.

Milstein, C., Neuberger, M.S., Staden, R., 1998. Both DNA strands of antibody genes are hypermutation targets. Proc. Natl. Acad. Sci. U. S. A. 95, 8791–8794. https://doi.org/10.1073/pnas.95.15.8791

Molica, S., 2006. Sex differences in incidence and outcome of chronic lymphocytic leukemia patients. Leuk. Lymphoma 47, 1477–1480. https://doi.org/10.1080/10428190600555819

Moshous, D., Callebaut, I., de Chasseval, R., Corneo, B., Cavazzana-Calvo, M., Le Deist, F., Tezcan, I., Sanal, O., Bertrand, Y., Philippe, N., Fischer, A., de Villartay, J.P., 2001. Artemis, a novel DNA double-strand break repair/V(D)J recombination protein, is mutated in human severe combined immune deficiency. Cell 105, 177–186. https://doi.org/10.1016/s0092-8674(01)00309-9

Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., Honjo, T., 2000. Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell 102, 553–563. https://doi.org/10.1016/s0092-8674(00)00078-7

Muramatsu, M., Nagaoka, H., Shinkura, R., Begum, N.A., Honjo, T., 2007. Discovery of activation-induced cytidine deaminase, the engraver of antibody memory. Adv. Immunol. 94, 1–36. https://doi.org/10.1016/S0065-2776(06)94001-2

Muramatsu, M., Sankaranand, V.S., Anant, S., Sugai, M., Kinoshita, K., Davidson, N.O., Honjo, T., 1999. Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. J. Biol. Chem. 274, 18470–18476. https://doi.org/10.1074/jbc.274.26.18470

Navrkalova, V., Plevova, K., Hynst, J., Pal, K., Mareckova, A., Reigl, T., Jelinkova, H., Vrzalova, Z., Stranska, K., Pavlova, S., Panovska, A., Janikova, A., Doubek, M., Kotaskova, J., Pospisilova, S., 2021. LYmphoid NeXt-Generation Sequencing (LYNX) Panel: A Comprehensive Capture-Based Sequencing Tool for the Analysis of Prognostic and Predictive Markers in Lymphoid Malignancies. J. Mol. Diagn. 23, 959–974. https://doi.org/10.1016/j.jmoldx.2021.05.007

Nemazee, D., 2006. Receptor editing in lymphocyte development and central tolerance. Nat. Rev. Immunol. 6, 728–740. https://doi.org/10.1038/nri1939

Nemazee, D., Weigert, M., 2000. Revising B cell receptors. J. Exp. Med. 191, 1813–1817. https://doi.org/10.1084/jem.191.11.1813

Neuberger, M.S., 2008. Antibody diversification by somatic mutation: from Burnet onwards. Immunol. Cell Biol. 86, 124–132. https://doi.org/10.1038/sj.icb.7100160

Neuberger, M.S., Milstein, C., 1995. Somatic hypermutation. Curr. Opin. Immunol. 7, 248–254. https://doi.org/10.1016/0952-7915(95)80010-7

Novobrantseva, T.I., Martin, V.M., Pelanda, R., Müller, W., Rajewsky, K., Ehlich, A., 1999. Rearrangement and expression of immunoglobulin light chain genes can precede heavy chain expression during normal B cell development in mice. J. Exp. Med. 189, 75–88. https://doi.org/10.1084/jem.189.1.75

Oettinger, M.A., Schatz, D.G., Gorka, C., Baltimore, D., 1990. RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. Science 248, 1517–1523. https://doi.org/10.1126/science.2360047

Osmond, D.G., Rico-Vargas, S., Valenzona, H., Fauteux, L., Liu, L., Janani, R., Lu, L., Jacobsen, K., 1994. Apoptosis and macrophage-mediated cell deletion in the regulation of B lymphopoiesis in mouse bone marrow. Immunol. Rev. 142, 209–230. https://doi.org/10.1111/j.1600-065x.1994.tb00891.x

Panovská, A., Doubek, M., Brychtová, Y., Mayer, J., 2010. Chronic lymphocytic leukemia and focusing on epidemiology and management in everyday hematologic practice: recent data from the Czech Leukemia Study Group for Life (CELL). Clin. Lymphoma Myeloma Leuk. 10, 297–300. https://doi.org/10.3816/CLML.2010.n.061

Pargent, W., Meindl, A., Thiebe, R., Mitzel, S., Zachau, H.G., 1991. The human immunoglobulin kappa locus. Characterization of the duplicated O regions. Eur. J. Immunol. 21, 1821–1827. https://doi.org/10.1002/eji.1830210807

Parikh, S.A., Rabe, K.G., Kay, N.E., Call, T.G., Ding, W., Schwager, S.M., Bowen, D.A., Conte, M., Jelinek, D.F., Slager, S.L., Shanafelt, T.D., 2014. Chronic lymphocytic leukemia in young (≤ 55 years) patients: a comprehensive analysis of prognostic factors and outcomes. Haematologica 99, 140–147. https://doi.org/10.3324/haematol.2013.086066

Pech, M., Höchtl, J., Schnell, H., Zachau, H.G., 1981. Differences between germ-line and rearranged immunoglobulin Vκ coding sequences suggest a localized mutation mechanism. Nature 291, 668–670. https://doi.org/10.1038/291668a0

Pflug, N., Bahlo, J., Shanafelt, T.D., Eichhorst, B.F., Bergmann, M.A., Elter, T., Bauer, K., Malchau, G., Rabe, K.G., Stilgenbauer, S., Döhner, H., Jäger, U., Eckart, M.J., Hopfinger, G., Busch, R., Fink, A.-M., Wendtner, C.-M., Fischer, K., Kay, N.E., Hallek, M., 2014. Development of a comprehensive prognostic index for patients with chronic lymphocytic leukemia. Blood 124, 49–62. https://doi.org/10.1182/blood-2014-02-556399

Pike, K.A., Ratcliffe, M.J.H., 2002. Cell surface immunoglobulin receptors in B cell development. Semin. Immunol. 14, 351–358. https://doi.org/10.1016/s1044-5323(02)00068-4

Plevova, K., Francova, H.S., Burckova, K., Brychtova, Y., Doubek, M., Pavlova, S., Malcikova, J., Mayer, J., Tichy, B., Pospisilova, S., 2014. Multiple productive immunoglobulin heavy chain gene rearrangements in chronic lymphocytic leukemia are mostly derived from independent clones. Haematologica 99, 329–338. https://doi.org/10.3324/haematol.2013.087593

Rada, C., Ehrenstein, M.R., Neuberger, M.S., Milstein, C., 1998. Hot spot focusing of somatic hypermutation in MSH2-deficient mice suggests two stages of mutational targeting. Immunity 9, 135–141. https://doi.org/10.1016/s1074-7613(00)80595-6

Radic, M.Z., Erikson, J., Litwin, S., Weigert, M., 1993. B lymphocytes may escape tolerance by revising their antigen receptors. J. Exp. Med. 177, 1165–1173. https://doi.org/10.1084/jem.177.4.1165

Radic, M.Z., Zouali, M., 1996. Receptor editing, immune diversification, and self-tolerance. Immunity 5, 505–511. https://doi.org/10.1016/s1074-7613(00)80266-6

Rai, K.R., Montserrat, E., 1987. Prognostic factors in chronic lymphocytic leukemia. Semin. Hematol. 24, 252–256.

Rai, K.R., Sawitsky, A., Cronkite, E.P., Chanana, A.D., Levy, R.N., Pasternack, B.S., 1975. Clinical staging of chronic lymphocytic leukemia. Blood 46, 219–234.

Rajewsky, K., 1996. Clonal selection and learning in the antibody system. Nature 381, 751–758. https://doi.org/10.1038/381751a0

Ramsden, D.A., van Gent, D.C., Gellert, M., 1997. Specificity in V(D)J recombination: new lessons from biochemistry and genetics. Curr. Opin. Immunol. 9, 114–120. https://doi.org/10.1016/s0952-7915(97)80167-7

Rassenti, L.Z., Kipps, T.J., 1997. Lack of allelic exclusion in B cell chronic lymphocytic leukemia. J. Exp. Med. 185, 1435–1445. https://doi.org/10.1084/jem.185.8.1435

Rawstron, A.C., Bennett, F.L., O'Connor, S.J.M., Kwok, M., Fenton, J.A.L., Plummer, M., de Tute, R., Owen, R.G., Richards, S.J., Jack, A.S., Hillmen, P., 2008. Monoclonal B-cell lymphocytosis and chronic lymphocytic leukemia. N. Engl. J. Med. 359, 575–583. https://doi.org/10.1056/NEJMoa075290

Reth, M., 1989. Antigen receptor tail clue. Nature 338, 383–384. https://doi.org/10.1038/338383b0

Rogozin, I.B., Kolchanov, N.A., 1992. Somatic hypermutagenesis in immunoglobulin genes. II. Influence of neighbouring base sequences on mutagenesis. Biochim. Biophys. Acta 1171, 11–18. https://doi.org/10.1016/0167-4781(92)90134-l

Rosenquist, R., Ghia, P., Hadzidimitriou, A., Sutton, L.-A., Agathangelidis, A., Baliakas, P., Darzentas, N., Giudicelli, V., Lefranc, M.-P., Langerak, A.W., Belessi, C., Davi, F., Stamatopoulos, K., 2017. Immunoglobulin gene sequence analysis in chronic lymphocytic leukemia: updated ERIC recommendations. Leukemia 31, 1477–1481. https://doi.org/10.1038/leu.2017.125

Rose-Zerilli, M.J.J., Gibson, J., Wang, J., Tapper, W., Davis, Z., Parker, H., Larrayoz, M., McCarthy, H., Walewska, R., Forster, J., Gardiner, A., Steele, A.J., Chelala, C., Ennis, S., Collins, A., Oakes, C.C., Oscier, D.G., Strefford, J.C., 2016. Longitudinal copy number, whole exome and targeted deep sequencing of "good risk" IGHV-mutated CLL patients with progressive disease. Leukemia 30, 1301–1310. https://doi.org/10.1038/leu.2016.10

Sakaguchi, N., Kashiwamura, S., Kimoto, M., Thalmann, P., Melchers, F., 1988. B lymphocyte lineage-restricted expression of mb-1, a gene with CD3-like structural properties. EMBO J. 7, 3457–3464. https://doi.org/10.1002/j.1460-2075.1988.tb03220.x

Sanchez, M., Misulovin, Z., Burkhardt, A.L., Mahajan, S., Costa, T., Franke, R., Bolen, J.B., Nussenzweig, M., 1993. Signal transduction by immunoglobulin is mediated through Ig alpha and Ig beta. J. Exp. Med. 178, 1049–1055. https://doi.org/10.1084/jem.178.3.1049

Sanchez, M.-L., Almeida, J., Gonzalez, D., Gonzalez, M., Garcia-Marcos, M.-A., Balanzategui, A., Lopez-Berges, M.-C., Nomdedeu, J., Vallespi, T., Barbon, M., Martin, A., de la Fuente, P., Martin-Nuñez, G., Fernandez-Calvo, J., Hernandez, J.-M., San Miguel, J.F., Orfao, A., 2003. Incidence and clinicobiologic characteristics of leukemic B-cell chronic lymphoproliferative disorders with more than one B-cell clone. Blood 102, 2994–3002. https://doi.org/10.1182/blood-2003-01-0045

Sant, M., Allemani, C., Tereanu, C., De Angelis, R., Capocaccia, R., Visser, O., Marcos-Gragera, R., Maynadié, M., Simonetti, A., Lutz, J.-M., Berrino, F., HAEMACARE Working Group, 2010. Incidence of hematologic malignancies in Europe by morphologic subtype: results of the HAEMACARE project. Blood 116, 3724–3734. https://doi.org/10.1182/blood-2010-05-282632

Schamel, W.W., Reth, M., 2000. Monomeric and oligomeric complexes of the B cell antigen receptor. Immunity 13, 5–14. https://doi.org/10.1016/s1074-7613(00)00003-0

Schatz, D.G., Oettinger, M.A., Baltimore, D., 1989. The V(D)J recombination activating gene, RAG-1. Cell 59, 1035–1048. https://doi.org/10.1016/0092-8674(89)90760-5

Schroeder, H.W., Cavacini, L., 2010. Structure and function of immunoglobulins. J. Allergy Clin. Immunol. 125, S41-52. https://doi.org/10.1016/j.jaci.2009.09.046

Shanafelt, T.D., Jenkins, G., Call, T.G., Zent, C.S., Slager, S., Bowen, D.A., Schwager, S., Hanson, C.A., Jelinek, D.F., Kay, N.E., 2009. Validation of a New Prognostic Index for Patients With Chronic Lymphocytic Leukemia (CLL). Cancer 115, 363–372. https://doi.org/10.1002/cncr.24004

Shenoy, P.J., Malik, N., Sinha, R., Nooka, A., Nastoupil, L.J., Smith, M., Flowers, C.R., 2011. Racial differences in the presentation and outcomes of chronic lymphocytic leukemia and variants in the United States. Clin. Lymphoma Myeloma Leuk. 11, 498–506. https://doi.org/10.1016/j.clml.2011.07.002

Shi, Z., Zhang, Q., Yan, H., Yang, Y., Wang, P., Zhang, Y., Deng, Z., Yu, M., Zhou, W., Wang, Q., Yang, X., Mo, X., Zhang, C., Huang, J., Dai, H., Sun, B., Zhao, Y., Zhang, L., Yang, Y.-G., Qiu, X., 2019. More than one antibody of individual B cells revealed by single-cell immune profiling. Cell Discov. 5, 1–13. https://doi.org/10.1038/s41421-019-0137-3

Shinkura, R., Tian, M., Smith, M., Chua, K., Fujiwara, Y., Alt, F.W., 2003. The influence of transcriptional orientation on endogenous switch region function. Nat. Immunol. 4, 435–441. https://doi.org/10.1038/ni918

Stamatopoulos, B., Timbs, A., Bruce, D., Smith, T., Clifford, R., Robbe, P., Burns, A., Vavoulis, D.V., Lopez, L., Antoniou, P., Mason, J., Dreau, H., Schuh, A., 2017. Targeted deep sequencing reveals clinically relevant subclonal IgHV rearrangements in chronic lymphocytic leukemia. Leukemia 31, 837–845. https://doi.org/10.1038/leu.2016.307

Stamatopoulos, K., Agathangelidis, A., Rosenquist, R., Ghia, P., 2017. Antigen receptor stereotypy in chronic lymphocytic leukemia. Leukemia 31, 282–291. https://doi.org/10.1038/leu.2016.322

Stamatopoulos, K., Belessi, C., Moreno, C., Boudjograh, M., Guida, G., Smilevska, T., Belhoul, L., Stella, S., Stavroyianni, N., Crespo, M., Hadzidimitriou, A., Sutton, L., Bosch, F., Laoutaris, N., Anagnostopoulos, A., Montserrat, E., Fassas, A., Dighiero, G., Caligaris-Cappio, F., Merle-Béral, H., Ghia, P., Davi, F., 2007. Over 20% of patients with chronic lymphocytic leukemia carry stereotyped receptors: Pathogenetic implications and clinical correlations. Blood 109, 259–270. https://doi.org/10.1182/blood-2006-03-012948

Stamatopoulos, K., Kosmas, C., Stavroyianni, N., Loukopoulos, D., 1996. Evidence for immunoglobulin heavy chain variable region gene replacement in a patient with B cell chronic lymphocytic leukemia. Leukemia 10, 1551–1556.

Staudt, L.M., Lenardo, M.J., 1991. Immunoglobulin Gene Transcription. Annu. Rev. Immunol. 9, 373–398. https://doi.org/10.1146/annurev.iy.09.040191.002105

Stavnezer, J., 1996. Immunoglobulin class switching. Curr. Opin. Immunol. 8, 199–205. https://doi.org/10.1016/s0952-7915(96)80058-6

Steen, S.B., Gomelsky, L., Roth, D.B., 1996. The 12/23 rule is enforced at the cleavage step of V(D)J recombination in vivo. Genes Cells Devoted Mol. Cell. Mech. 1, 543–553. https://doi.org/10.1046/j.1365-2443.1996.d01-259.x

Steiner, L.A., Eisen, H.N., 1967. Sequential changes in the relative affinity of antibodies synthesized during the immune response. J. Exp. Med. 126, 1161–1183. https://doi.org/10.1084/jem.126.6.1161

Stranska, K., Plevova, K., Skuhrova Francova, H., Skabrahova, H., von Jagwitz-Biegnitz, M., Radova, L., Panovska, A., Hrobkova, S., Brychtova, Y., Urbanova, R., Smolej, L., Simkovic, M., Zuchnicka, J., Mohammadova, L., Spacek, M., Mayer, J., Pospisilova, S., Doubek, M., 2020. Profiling of biological and environmental risk factors in immunogenetic subgroups of chronic lymphocytic leukemia - Czech national study. Biomed. Pap. Med. Fac. Univ. Palacky Olomouc Czechoslov. 164, 425–434. https://doi.org/10.5507/bp.2019.046

Sutton, L.-A., Hadzidimitriou, A., Baliakas, P., Agathangelidis, A., Langerak, A.W., Stilgenbauer, S., Pospisilova, S., Davis, Z., Forconi, F., Davi, F., Ghia, P., Rosenquist, R., Stamatopoulos, K., 2017. Immunoglobulin genes in chronic lymphocytic leukemia: key to understanding the disease and improving risk stratification. Haematologica 102, 968–971. https://doi.org/10.3324/haematol.2017.165605

Taccioli, G.E., Rathbun, G., Oltz, E., Stamato, T., Jeggo, P.A., Alt, F.W., 1993. Impairment of V(D)J recombination in double-strand break repair mutants. Science 260, 207–210. https://doi.org/10.1126/science.8469973

Thai, T.-H., Kearney, J.F., 2004. Distinct and opposite activities of human terminal deoxynucleotidyltransferase splice variants. J. Immunol. Baltim. Md 1950 173, 4009–4019. https://doi.org/10.4049/jimmunol.173.6.4009

Thai, T.-H., Purugganan, M.M., Roth, D.B., Kearney, J.F., 2002. Distinct and opposite diversifying activities of terminal transferase splice variants. Nat. Immunol. 3, 457–462. https://doi.org/10.1038/ni788

Tiegs, S.L., Russell, D.M., Nemazee, D., 1993. Receptor editing in self-reactive bone marrow B cells. J. Exp. Med. 177, 1009–1020. https://doi.org/10.1084/jem.177.4.1009

Tobin, G., Thunberg, U., Laurell, A., Karlsson, K., Aleskog, A., Willander, K., Soderberg, O., Merup, M., Vilpo, J., Hultdin, M., Sundstrom, C., Roos, G., Rosenquist, R., 2005. Patients with chronic lymphocytic leukemia with mutated VH genes presenting

with Binet stage B or C form a subgroup with a poor outcome. Haematologica 90, 465–469. https://doi.org/10.3324/%x

Tolar, P., Sohn, H.W., Pierce, S.K., 2005. The initiation of antigen-induced B cell antigen receptor signaling viewed in living cells by fluorescence resonance energy transfer. Nat. Immunol. 6, 1168–1176. https://doi.org/10.1038/ni1262

Tonegawa, S., 1983. Somatic generation of antibody diversity. Nature 302, 575–581. https://doi.org/10.1038/302575a0

Torres, R.M., Imboden, J., Schroeder, H.W., 2008. 4 - Antigen receptor genes, gene products, and co-receptors, in: Rich, R.R., Fleisher, T.A., Shearer, W.T., Schroeder, H.W., Frew, A.J., Weyand, C.M. (Eds.), Clinical Immunology (Third Edition). Mosby, Edinburgh, pp. 53–77. https://doi.org/10.1016/B978-0-323-04404-2.10004-1

Treanor, B., 2012. B-cell receptor: from resting state to activate. Immunology 136, 21–27. https://doi.org/10.1111/j.1365-2567.2012.03564.x

Troshchynsky, A., Dzneladze, I., Chen, L., Sheng, Y., Saridakis, V., Wu, G.E., 2015. Functional analyses of polymorphic variants of human terminal deoxynucleotidyl transferase. Genes Immun. 16, 388–398. https://doi.org/10.1038/gene.2015.19

van Gent, D.C., Ramsden, D.A., Gellert, M., 1996. The RAG1 and RAG2 proteins establish the 12/23 rule in V(D)J recombination. Cell 85, 107–113. https://doi.org/10.1016/s0092-8674(00)81086-7

Vasicek, T.J., Leder, P., 1990. Structure and expression of the human immunoglobulin lambda genes. J. Exp. Med. 172, 609–620.

Venkitaraman, A.R., Williams, G.T., Dariavach, P., Neuberger, M.S., 1991. The B-cell antigen receptor of the five immunoglobulin classes. Nature 352, 777–781. https://doi.org/10.1038/352777a0

Visco, C., Moretta, F., Falisi, E., Facco, M., Maura, F., Novella, E., Nichele, I., Finotto, S., Giaretta, I., Ave, E., Perbellini, O., Guercini, N., Scupoli, M.T., Trentin, L., Trimarco, V., Neri, A., Semenzato, G., Rodeghiero, F., Pizzolo, G., Ambrosetti, A., 2013. Double productive immunoglobulin sequence rearrangements in patients with chronic lymphocytic leukemia. Am. J. Hematol. 88, 277–282. https://doi.org/10.1002/ajh.23396

Waheed, O., 2018. Investigation of the molecular mechanisms that determine signal strength from B-cell antigen receptors (BCRs) of the IgM and IgD isotypes and the affinity

of chronic lymphocytic leukemia derived BCRs. (Master's Thesis). University of Eastern Finland, Kuopio.

West, R.B., Lieber, M.R., 1998. The RAG-HMG1 complex enforces the 12/23 rule of V(D)J recombination specifically at the double-hairpin formation step. Mol. Cell. Biol. 18, 6408–6415. https://doi.org/10.1128/MCB.18.11.6408

Wiesendanger, M., Kneitz, B., Edelmann, W., Scharff, M.D., 2000. Somatic Hypermutation in Muts Homologue (Msh)3-, Msh6-, and Msh3/Msh6-Deficient Mice Reveals a Role for the Msh2–Msh6 Heterodimer in Modulating the Base Substitution Pattern. J. Exp. Med. 191, 579–584. https://doi.org/10.1084/jem.191.3.579

Williams, A.F., Barclay, A.N., 1988. The Immunoglobulin Superfamily—Domains for Cell Surface Recognition. Annu. Rev. Immunol. 6, 381–405. https://doi.org/10.1146/annurev.iy.06.040188.002121

Williams, S.C., Frippiat, J.P., Tomlinson, I.M., Ignatovich, O., Lefranc, M.P., Winter, G., 1996. Sequence and evolution of the human germline V lambda repertoire. J. Mol. Biol. 264, 220–232. https://doi.org/10.1006/jmbi.1996.0636

Wilson, P.C., Wilson, K., Liu, Y.J., Banchereau, J., Pascual, V., Capra, J.D., 2000. Receptor revision of immunoglobulin heavy chain variable region genes in normal human B lymphocytes. J. Exp. Med. 191, 1881–1894. https://doi.org/10.1084/jem.191.11.1881

Wilson, T.M., Vaisman, A., Martomo, S.A., Sullivan, P., Lan, L., Hanaoka, F., Yasui, A., Woodgate, R., Gearhart, P.J., 2005. MSH2–MSH6 stimulates DNA polymerase η, suggesting a role for A:T mutations in antibody genes. J. Exp. Med. 201, 637–645. https://doi.org/10.1084/jem.20042066

Zenz, T., Kröber, A., Scherer, K., Häbe, S., Bühler, A., Benner, A., Denzel, T., Winkler, D., Edelmann, J., Schwänen, C., Döhner, H., Stilgenbauer, S., 2008. Monoallelic TP53 inactivation is associated with poor prognosis in chronic lymphocytic leukemia: results from a detailed genetic characterization with long-term follow-up. Blood 112, 3322–3329. https://doi.org/10.1182/blood-2008-04-154070

Zhu, L., Peng, Q., Wu, Y., Yao, X., 2023. scBCR-seq revealed a special and novel IG H&L V(D)J allelic inclusion rearrangement and the high proportion dual BCR expressing B cells. Cell. Mol. Life Sci. CMLS 80, 319. https://doi.org/10.1007/s00018-023-04973-8

Appendix A Overview of IGHV genes in CLL

Tab. 1: A comprehensive overview of the IGHV families and their corresponding genes.

IGHV family	IGHV gene rearrangement	monoclonal CLL (number of samples)	multiclonal CLL (number of samples)		
			Sanger	LYNX	IG aNGS
IGHV1	IGHV1-18	47 (1.68%)	2 (1.23%)	1 (0.99%)	4 (3.08%)
	IGHV1-2	113 (4.04%)	7 (4.32%)	7 (6.93%)	6 (4.62%)
	IGHV1-24	7 (0.25%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
	IGHV1-3	70 (2.51%)	6 (3.70%)	4 (3.96%)	5 (3.85%)
	IGHV1-46	44 (1.57%)	3 (1.85%)	2 (1.98%)	4 (3.08%)
	IGHV1-58	8 (0.29%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
	IGHV1-69	391 (13.997%)	21 (12.96%)	14 (13.86%)	17 (13.08%)
	IGHV1-69-2	1 (0.04%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
	IGHV1-8	28 (1.00%)	4 (2.47%)	2 (1.98%)	1 (0.77%)
IGHV2	IGHV2-26	21 (0.75%)	1 (0.62%)	0 (0.00%)	0 (0.00%)
	IGHV2-5	72 (2.58%)	5 (3.09%)	1 (0.99%)	2 (1.54%)
	IGHV2-70	17 (0.61%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
IGHV3	IGHV3-11	67 (2.40%)	2 (1.23%)	2 (1.98%)	2 (1.54%)
	IGHV3-13	16 (0.57%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
	IGHV3-15	64 (2.29%)	1 (0.62%)	0 (0.00%)	1 (0.77%)
	IGHV3-20	10 (0.36%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
	IGHV3-21	98 (3.54%)	5 (3.09%)	3 (2.97%)	2 (1.54%)
	IGHV3-23	204 (7.30%)	15 (9.26%)	8 (7.92%)	9 (6.92%)
	IGHV3-30	61 (2.18%)	3 (1.85%)	2 (1.98%)	1 (0.77%)
	IGHV3-30-3	34 (1.22%)	1 (0.62%)	0 (0.00%)	2 (1.54%)
	IGHV3-30 = IGHV3-30-3	17 (0.61%)	0 (0%)	0 (0%)	0 (0%)
	IGHV3-30 = IGHV3-30-5	66 (2.36%)	2 (1.23%)	1 (0.99%)	8 (6.15%)
	IGHV3-30 = IGHV3-30-3 = IGHV3-30-5	1 (0.04%)	0 (0%)	0 (0%)	0 (0%)
	IGHV3-30 = IGHV3-30-5 = IGHV3-33	3 (0.11%)	0 (0%)	0 (0%)	0 (0%)
	IGHV3-33	89 (3.19%)	4 (2.47%)	3 (2.97%)	3 (2.31%)
	IGHV3-43	3 (0.11%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
	IGHV3-48	113 (4.04%)	5 (3.09%)	3 (2.97%)	4 (3.08%)

IGHV family	IGHV gene rearrangement	monoclonal CLL (number of samples)	multiclonal CLL (number of samples)		
			Sanger	LYNX	IG aNGS
IGHV3	IGHV3-49	30 (1.07%)	1 (0.62%)	1 (0.99%)	1 (0.77%)
	IGHV3-53	35 (1.25%)	2 (1.23%)	2 (1.98%)	1 (0.77%)
	IGHV3-64	17 (0.61%)	2 (1.23%)	1 (0.99%)	2 (1.54%)
	IGHV3-66	16 (0.57%)	1 (0.62%)	1 (0.99%)	1 (0.77%)
	IGHV3-7	150 (5.37%)	10 (6.17%)	6 (5.94%)	6 (4.62%)
	IGHV3-72	39 (1.40%)	2 (1.23%)	2 (1.98%)	5 (3.85%)
	IGHV3-73	12 (0.43%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
	IGHV3-74	57 (2.04%)	3 (1.85%)	1 (0.99%)	3 (2.31%)
	IGHV3-9	48 (1.72%)	0 (0.00%)	0 (0.00%)	2 (1.54%)
IGHV4	IGHV4-28	1 (0.04%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
	IGHV4-30-2	7 (0.25%)	1 (0.62%)	0 (0.00%)	0 (0.00%)
	IGHV4-30-4	23 (0.82%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
	IGHV4-30-4 = IGHV4-31	2 (0.07%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
	IGHV4-31	35 (1.25%)	1 (0.62%)	1 (0.99%)	0 (0.00%)
	IGHV4-34	239 (8.55%)	12 (7.41%)	6 (5.94%)	8 (6.15%)
	IGHV4-38-2	37 (1.32%)	4 (2.47%)	4 (3.96%)	5 (3.85%)
	IGHV4-39	118 (4.22%)	7 (4.32%)	7 (6.93%)	6 (4.62%)
	IGHV4-4	50 (1.79%)	1 (0.62%)	2 (1.98%)	3 (2.31%)
	IGHV4-59	82 (2.93%)	8 (4.94%)	3 (2.97%)	6 (4.62%)
	IGHV4-59 = IGHV4-61	2 (0.07%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
	IGHV4-61	33 (1.18%)	1 (0.62%)	0 (0.00%)	0 (0.00%)
IGHV5	IGHV5-10-1	18 (0.64%)	4 (2.47%)	1 (0.99%)	1 (0.77%)
	IGHV5-51	53 (1.90%)	9 (5.56%)	5 (4.95%)	6 (4.62%)
IGHV6	IGHV6-1	16 (0.57%)	5 (3.09%)	4 (3.96%)	2 (1.54%)
IGHV7	IGHV7-4-1	9 (0.32%)	1 (0.62%)	1 (0.99%)	1 (0.77%)

Appendix B Overview of IGHD genes in CLL

Tab. 2: A comprehensive overview of the IGHD families and their corresponding genes.

IGHD family	IGHD gene rearrangement	monoclonal CLL (number of samples)	multiclonal CLL (number of samples)		
			Sanger	LYNX	IG aNGS
IGHD1	IGHD1-1	43 (1.54%)	2 (1.23%)	1 (0.99%)	3 (2.31%)
	IGHD1-14	30 (1.07%)	1 (0.62%)	0 (0.00%)	0 (0.00%)
	IGHD1-20	11 (0.39%)	1 (0.62%)	0 (0.00%)	0 (0.00%)
	IGHD1-26	117 (4.19%)	7 (4.32%)	4 (3.96%)	4 (3.08%)
	IGHD1-7	23 (0.82%)	1 (0.62%)	1 (0.99%)	2 (1.54%)
IGHD2	IGHD2-15	161 (5.76%)	6 (3.70%)	6 (5.94%)	5 (3.85%)
	IGHD2-2	215 (7.70%)	14 (8.64%)	6 (5.94%)	11 (8.46%)
	IGHD2-21	94 (3.36%)	8 (4.94%)	3 (2.97%)	3 (2.31%)
	IGHD2-8	54 (1.93%)	4 (2.47%)	3 (2.97%)	2 (1.54%)
IGHD3	IGHD3-10	245 (8.77%)	9 (5.56%)	8 (7.92%)	13 (10.00%)
	IGHD3-16	141 (5.05%)	9 (5.56%)	7 (6.93%)	6 (4.62%)
	IGHD3-22	234 (8.38%)	20 (12.35%)	10 (9.90%)	10 (7.69%)
	IGHD3-3	439 (15.71%)	19 (11.73%)	14 (13.86%)	19 (14.62%)
	IGHD3-3 = IGHD3-22	0 (0.00%)	0 (0.00%)	0 (0.00%)	1 (0.77%)
	IGHD3-3 = IGHD7-27	0 (0.00%)	0 (0.00%)	0 (0.00%)	1 (0.77%)
	IGHD3-9	106 (3.79%)	4 (2.47%)	2 (1.98%)	3 (2.31%)
IGHD4	IGHD4-11	29 (1.04%)	3 (1.85%)	1 (0.99%)	0 (0.00%)
	IGHD4-17	104 (3.72%)	4 (2.47%)	3 (2.97%)	5 (3.85%)
	IGHD4-23	37 (1.32%)	4 (2.47%)	3 (2.97%)	1 (0.77%)
	IGHD4-4	0 (0.00%)	0 (0.00%)	0 (0.00%)	1 (0.77%)
IGHD5	IGHD5-12	82 (2.93%)	1 (0.62%)	1 (0.99%)	1 (0.77%)
	IGHD5-18	77 (2.76%)	7 (4.32%)	4 (3.96%)	0 (0.00%)
	IGHD5-24	60 (2.15%)	2 (1.23%)	0 (0.00%)	1 (0.77%)
	IGHD5-5 = IGHD5-18	0 (0.00%)	0 (0.00%)	0 (0.00%)	6 (4.62%)
IGHD6	IGHD6-13	153 (5.48%)	15 (9.26%)	10 (9.90%)	6 (4.62%)
	IGHD6-13 = IGHD6-19	0 (0.00%)	0 (0.00%)	0 (0.00%)	1 (0.77%)
	IGHD6-19	238 (8.52%)	17 (10.49%)	12 (11.88%)	21 (16.15%)
	IGHD6-25	12 (0.43%)	0 (0.00%)	0 (0.00%)	1 (0.77%)
	IGHD6-6	66 (2.36%)	3 (1.85%)	2 (1.98%)	2 (1.54%)

IGHD family	IGHD gene rearrangement	monoclonal CLL (number of samples)	multiclonal CLL (number of samples)		
			Sanger	LYNX	IG aNGS
IGHD7	IGHD7-27	23 (0.82%)	1 (0.62%)	0 (0.00%)	1 (0.77%)

Appendix C Overview of IGHJ genes in CLL

Tab. 3: A comprehensive overview of the IGHJ families and their corresponding genes.

IGHJ gene rearrangement	monoclonal CLL (number of samples)	multiclonal CLL (number of samples)		
		Sanger	LYNX	IG aNGS
IGHJ1	43 (1.54%)	1 (0.62%)	0 (0.00%)	1 (0.77%)
IGHJ2	56 (2.00%)	3 (1.85%)	2 (1.98%)	2 (1.54%)
IGHJ3	285 (10.20%)	10 (6.17%)	8 (7.92%)	9 (6.92%)
IGHJ4	1182 (42.30%)	83 (51.23%)	42 (41.58%)	62 (47.69%)
IGHJ5	299 (10.70%)	18 (11.11%)	13 (12.87%)	10 (7.69%)
IGHJ6	929 (33.25%)	47 (29.01%)	36 (35.64%)	46 (35.38%)

Appendix D Overview of IGH rearrangements in multiclonal CLL

Tab. 4: A detailed list of all detected productive IGH rearrangements in multiclonal CLL.

The table presents a comprehensive list of all detected productive IGH rearrangements in the multiclonal CLL cohort, providing detailed information on gene usage, sequence identity, IGHV mutational status, corresponding detection method, and estimated clonal abundance. In sample 3304, due to its highly polyclonal composition, only the two major IGH rearrangements, also confirmed by other methods, were included in the table for clarity and consistency. U, unmutated IGHV; M, mutated IGHV; BDL, below detection limit.

sample ID	IGHV	IGHV identity	IGHV mutational status	IGHD	IGHJ	CDR3	Sanger	LYNX	LYNX clone proportion	IG aNGS	IG aNGS clone proportion
463B	IGHV1-2	100.00	U	IGHD2-8	IGHJ6	CARSFLGYCTNGVCYDPPWVAAAGTGYYYYGMDVW	yes	yes	74.58	yes	67.04
	IGHV1-69	100.00	U	IGHD2-21	IGHJ6	CASYGGDCYNLCMDVW	yes	yes	25.42	yes	32.96
467	IGHV1-3	100.00	U	IGHD6-19	IGHJ4	CAREQWLVTPNFDYW	yes	yes	86.53	yes	82.64
	IGHV4-34	100.00	U	IGHD3-9	IGHJ6	CARLLAGAYYYYYYGMDVW	yes	yes	13.47	yes	17.36
521	IGHV1-69	100.00	U	IGHD2-15	IGHJ6	CAREVGIVATLYYYYGMDVW	yes	yes	76.70	yes	83.03
	IGHV5-51	100.00	U	IGHD6-19	IGHJ4	CARQQWLGNYYFDYW	yes	yes	23.30	yes	2.92
	IGHV3-72	94.65	M	IGHD3-10	IGHJ4	CASFTAKFGSNIRKFDYW	no	no	-	yes	6.39
	IGHV1-18	97.64	M	IGHD2-2	IGHJ6	CARDQGYCSSTSCYEGNYYYYYGMDVW	no	no	-	yes	3.94
	IGHV3-30 = IGHV3-30-5	100.00	U	IGHD2-2	IGHJ6	CAKDLTPTEYCSSTSCYYYYYYYGMDVW	no	no	-	yes	3.72
568	IGHV3-72	94.90	M	IGHD5-18	IGHJ4	CASFTAKFGSNIRKFDYW	yes	yes	26.44	yes	62.08
	IGHV4-39	98.97	U	IGHD2-15	IGHJ4	CASRDWGSGEDYW	yes	yes	12.64	yes	1.83
	IGHV1-8	96.53	M	IGHD4-17	IGHJ4	CARGWAYGDNADYW	yes	no	-	BDL	-
	IGHV1-18	97.64	M	IGHD2-2	IGHJ6	CARDQGYCSSTSCYEGNYYYYYGMDVW	no	yes	60.92	yes	36.09
1123	IGHV4-38-2	93.06	M	IGHD3-22	IGHJ5	CARGYTESGEGSGYYFNWFDPW	yes	yes	88.47	yes	68.33
	IGHV3-11	96.95	M	IGHD6-13	IGHJ4	CARGAAGTGDFDYW	no	yes	10.53	yes	30.42

sample ID	IGHV	IGHV identity	IGHV mutational status	IGHD	IGHJ	CDR3	Sanger	LYNX	LYNX clone proportion	IG aNGS	IG aNGS clone proportion
1123	IGHV3-9	96.89	M	IGHD3-3	IGHJ6	CVKCRGFDFWSDHLHYYYGMDVW	no	BDL	-	yes	1.28
1436	IGHV4-59	94.14	M	IGHD5-18	IGHJ5	CARDIYWIGYMYQGNWFDPW	yes	yes	59.66	yes	36.09
	IGHV3-66	96.84	M	IGHD3-3	IGHJ3	CARDLPGVFEARYGAFDIW	yes	yes	40.34	yes	63.91
1475	IGHV1-69	100.00	U	IGHD3-10	IGHJ6	CARVEVQGVINLYYYYYGMDVW	yes	yes	76.15	yes	67.80
	IGHV3-48	100.00	U	IGHD3-3	IGHJ6	CARDRDFWSGYYGYYYYYGMDVW	yes	yes	10.22	yes	28.73
	IGHV1-69	100.00	U	IGHD3-16	IGHJ3	CARDHITYYDYIWALDAFDIW	no	yes	13.63	yes	1.61
	IGHV3-30-3	98.65	U	IGHD2-2	IGHJ6	CARESLYCSTTICHHGLDVW	no	no	-	yes	1.86
1485	IGHV1-69	99.65	U	IGHD3-3	IGHJ4	CARGGEDYDFWSGYFTDRGSGRGHDVW	yes	yes	94.44	yes	62.87
	IGHV4-39	100.00	U	IGHD6-13	IGHJ5	CARQGGYSSSWYGRINWFDPW	yes	yes	5.56	yes	32.33
	IGHV3-21	99.32	U	IGHD3-3	IGHJ6	CARDIRSVIIWSGQQNYGMDVW	no	no	-	yes	2.44
	IGHV1-2	100.00	U	IGHD6-19	IGHJ4	CAREQWLALSHFDYW	no	no	-	yes	2.36
1531	IGHV3-33	100.00	U	IGHD4-17	IGHJ6	CARGQQGDYVYGMDVW	yes	yes	83.11	yes	81.31
	IGHV4-39	100.00	U	IGHD3-3	IGHJ4	CASDPGLGSRMRSYFDYW	yes	yes	2.18	BDL	-
	IGHV3-30 = IGHV3-30-5	100.00	U	IGHD6-13	IGHJ4	CAKAREQQLPPEDYW	no	yes	13.66	yes	18.69
	IGHV3-21	100.00	U	IGHD3-10	IGHJ5	CARDRVAVLYYYGSGSYSLSPEFDPW	no	yes	1.04	BDL	-
1622	IGHV3-23	94.44	M	IGHD6-13	IGHJ4	CAREGAGTDLLDFDSW	yes	yes	74.80	yes	84.85
	IGHV3-7	94.10	M	IGHD2-21	IGHJ6	CALSGLDLDYYYYYDMDVW	yes	yes	25.20	yes	15.15
1731	IGHV3-23	93.40	M	IGHD6-13	IGHJ6	CAKLGQHLVQIYYYHIDVW	yes	yes	52.33	yes	88.90
	IGHV5-51	94.79	M	IGHD5-18	IGHJ4	CARLRYSYGLSTSPADYW	yes	yes	47.67	yes	11.10
1732	IGHV1-2	100.00	U	IGHD3-3	IGHJ6	CARDTRDDDFWSGYLPEYYYYYYGMDVW	yes	yes	80.30	yes	72.81
	IGHV1-69	100.00	U	IGHD3-3	IGHJ6	CARDLYDFWSGYRPTGAYYYYGMDVW	yes	yes	15.40	yes	22.55
	IGHV1-69	100.00	U	IGHD2-8	IGHJ6	CARAARPYCTNGVCYLEIAAYYYYGMDVW	yes	yes	4.29	BDL	-
	IGHV3-72	94.65	M	IGHD3-10	IGHJ4	CASFTAKFGSNIRKFDYW	no	no	-	yes	4.64

sample ID	IGHV	IGHV identity	IGHV mutational status	IGHD	IGHJ	CDR3	Sanger	LYNX	LYNX clone proportion	IG aNGS	IG aNGS clone proportion
1738	IGHV1-69	100.00	U	IGHD2-15	IGHJ3	CARLFPPYCSGGSCYGWAPSGAFDIW	yes	yes	76.78	yes	71.64
	IGHV1-69	99.65	U	IGHD3-10	IGHJ6	CARAMVQGVISYYYYYYYMDVW	yes	yes	23.22	yes	28.36
1741	IGHV3-30	100.00	U	IGHD6-13	IGHJ4	CAKAREQQLPPEDYW	yes	yes	59.02	yes	65.29
	IGHV3-33	100.00	U	IGHD4-17	IGHJ6	CARGQQGDYVYGMDVW	yes	yes	30.41	yes	30.77
	IGHV4-39	100.00	U	IGHD3-3	IGHJ4	CASDPGLGSRMRSYFDYW	yes	yes	9.28	yes	3.93
	IGHV3-21	100.00	U	IGHD3-10	IGHJ5	CARDRVAVLYYYGSGSYSLSPEFDPW	no	yes	1.29	BDL	-
1750	IGHV3-23	96.88	M	IGHD3-22	IGHJ4	CAKFPRGAGYYYYFDYW	yes	yes	100.00	yes	100.00
	IGHV4-59	96.84	M	IGHD5-24	IGHJ4	CASARDAYTNFDSW	yes	BDL	-	BDL	-
1766	IGHV1-2	100.00	U	IGHD6-19	IGHJ4	CAREQWLALSHFDYW	yes	yes	68.81	yes	53.77
	IGHV3-21	99.31	U	IGHD3-3	IGHJ6	CARDIRSVIIWSGQQNYGMDVW	yes	yes	31.19	yes	46.23
1894	IGHV7-4-1	100.00	U	IGHD3-3	IGHJ6	CARSSGEPYDFWTNYYGMDVW	yes	yes	60.92	yes	97.44
	IGHV1-46	100.00	U	IGHD6-6	IGHJ4	CARGQSSIAAWYYFDYW	yes	yes	39.08	yes	2.56
1933	IGHV3-23	100.00	U	IGHD3-10	IGHJ3	CAKDRGSPRIPLLWFGELGDDAFDIW	yes	yes	95.60	yes	85.72
	IGHV4-38-2	99.65	U	IGHD2-21	IGHJ4	CARDTIHSTLDRLGYRFDYW	yes	yes	4.40	yes	14.28
1934	IGHV3-7	94.10	M	IGHD3-10	IGHJ4	CARDPGNSGSYEFDYW	yes	yes	66.27	yes	61.47
	IGHV3-72	93.20	M	IGHD1-26	IGHJ4	CARVIVGPSAFDYW	yes	yes	33.73	yes	38.53
	IGHV3-23	89.93	M	IGHD2-2	IGHJ4	CAKYDRGTIRSSNGFDHW	yes	no	-	BDL	-
1960	IGHV4-59	96.14	M	IGHD2-2	IGHJ2	CARSPSLYGDWYFDLW	yes	yes	53.05	yes	49.62
	IGHV3-7	95.83	M	IGHD3-22	IGHJ3	CTRDANYFDGSAYYDAFDIW	yes	yes	43.90	BDL	-
	IGHV4-34	94.39	M	IGHD3-10	IGHJ2	CARGRAFYSGSGDHWFFDVW	yes	BDL	-	yes	7.34
	IGHV4-4	87.80	M	IGHD2-15	IGHJ6	CARAPGSYFGTDVW	no	yes	3.05	yes	43.04
2101	IGHV3-23	99.65	U	IGHD3-3	IGHJ6	CAKWRVDDFWSGYDVPYYYYGMDVW	yes	yes	100.00	yes	100.00
2129	IGHV5-10-1	100.00	U	IGHD6-19	IGHJ4	CAREQWLGGDYFDYW	yes	yes	100.00	yes	100.00
	IGHV3-23	98.61	U	IGHD4-11	IGHJ4	CAKEGRDYSTPFDYW	yes	no	-	BDL	-

sample ID	IGHV	IGHV identity	IGHV mutational status	IGHD	IGHJ	CDR3	Sanger	LYNX	LYNX clone proportion	IG aNGS	IG aNGS clone proportion
2606	IGHV5-51	99.65	U	IGHD2-2	IGHJ5	CARIGERLGYCSSTSCYFWFDPW	yes	yes	100.00	yes	100.00
	IGHV2-26	90.38	M	IGHD3-22	IGHJ4	CTRISGLDSYQSGALSTDFW	yes	no	-	no	-
2669	IGHV3-23	93.75	M	IGHD3-10	IGHJ4	CAKDSSGGYSNPYNFDFW	yes	yes	100.00	yes	100.00
	IGHV4-34	97.19	M	IGHD2-2	IGHJ6	CARGRSSCDRSSCSPQNYYYGMDVW	yes	BDL	-	BDL	-
2683	IGHV5-51	95.83	M	IGHD4-23	IGHJ6	CARLREGGTSLDAMDVW	yes	yes	100.00	yes	94.54
	IGHV3-30	99.66	U	IGHD6-19	IGHJ4	CARGSSGWYHDYW	no	no	-	yes	5.46
2748	IGHV4-34	92.28	M	IGHD3-9	IGHJ4	CARSRSPTFDYYFDQW	yes	yes	100.00	yes	21.43
	IGHV3-64	87.37	M	IGHD1-26	IGHJ4	CAREEGGWSHYDKW	yes	no	-	yes	69.78
	IGHV3-7	90.37	M	IGHD6-13	IGHJ4	CARGVIAAGLFDSW	no	no	-	yes	5.52
	IGHV3-23	88.70	M	IGHD3-16	IGHJ4	CARGFTRLAELDSW	no	no	-	yes	3.27
2777	IGHV3-11	96.88	M	IGHD6-13	IGHJ4	CARGAAATGDFDYW	yes	yes	83.80	yes	97.43
	IGHV4-38-2	93.06	M	IGHD3-22	IGHJ5	CARGYTESGEGSGYYFNWFDPW	yes	yes	16.20	yes	2.57
2807	IGHV1-69	100.00	U	IGHD2-15	IGHJ3	CARLFPPYCSGGSCYGWAPSGAFDIW	yes	yes	100.00	yes	98.06
	IGHV1-69	100.00	U	IGHD3-10	IGHJ4	CARDTGSFGVIATSYYFDYW	no	no	-	yes	1.94
2995	IGHV3-7	100.00	U	IGHD5-18	IGHJ6	CARVDSYGPYYYYYYGMDVW	yes	yes	100.00	yes	100.00
	IGHV1-2	91.32	M	IGHD6-13	IGHJ4	CARGIAITGAGEDFDFW	yes	BDL	-	BDL	-
3034	IGHV1-3	100.00	U	IGHD3-10	IGHJ4	CARESWGLVGEFDYW	yes	yes	96.80	yes	98.00
	IGHV1-2	100.00	U	IGHD6-19	IGHJ4	CAREQWLWGNYFDYW	no	yes	3.20	yes	2.00
3066	IGHV4-39	100.00	U	IGHD6-19	IGHJ6	CARLGIAVARGYGMDVW	yes	yes	75.84	yes	49.74
	IGHV3-11	100.00	U	IGHD3-16	IGHJ4	CARDKITYYDYVWGSYRLDYW	yes	yes	24.16	yes	50.26
3190	IGHV4-34	94.39	M	IGHD3-22	IGHJ4	CARGPSLVQQRSGYHYFDSW	yes	yes	87.36	yes	93.40
	IGHV3-48	96.18	M	IGHD3-3	IGHJ4	CAAHYDFWSGYFLGYW	yes	yes	6.02	yes	4.94
	IGHV4-4	94.24	M	IGHD3-22	IGHJ4	CARVYDSSDYTFDYW	no	yes	6.63	yes	1.66
3239	IGHV1-69	100.00	U	IGHD3-10	IGHJ6	CARVEVQGVINLYYYYYGMDVW	yes	yes	95.09	yes	100.00

sample ID	IGHV	IGHV identity	IGHV mutational status	IGHD	IGHJ	CDR3	Sanger	LYNX	LYNX clone proportion	IG aNGS	IG aNGS clone proportion
3239	IGHV1-69	100.00	U	IGHD3-16	IGHJ3	CARDHITYYDYIWALDAFDIW	no	yes	4.91	BDL	-
3289	IGHV1-3	100.00	U	IGHD3-10	IGHJ4	CARESWGLVGEFDYW	yes	yes	92.11	yes	93.54
	IGHV1-2	100.00	U	IGHD6-19	IGHJ4	CAREQWLWGNYFDYW	yes	yes	7.89	yes	6.46
3304	IGHV4-34	96.14	M	IGHD2-15	IGHJ6	CASRLYSSGHYNYYMDVW	yes	yes	100.00	yes	12.02
	IGHV1-3	93.40	M	IGHD1-20	IGHJ3	CARNARIRATHDDSFDMW	yes	BDL	-	yes	12.70
3377	IGHV1-69	100.00	U	IGHD3-16	IGHJ4	CARDTGSFGVIATSYYFDYW	yes	yes	100.00	yes	94.03
	IGHV3-7	100.00	U	IGHD1-26	IGHJ6	CARVIRGELLYYYYYGMDVW	yes	BDL	-	BDL	-
	IGHV3-72	94.65	M	IGHD3-10	IGHJ4	CASFTAKFGSNIRKFDYW	no	no	-	yes	2.71
	IGHV3-7	96.93	M	IGHD3-22	IGHJ4	CASYYCDQKGCPIEKWDW	no	no	-	yes	1.69
	IGHV1-18	97.64	M	IGHD2-2	IGHJ6	CARDQGYCSSTSCYEGNYYYYYGMDVW	no	no	-	yes	1.56
3436	IGHV1-2	100.00	U	IGHD2-8	IGHJ6	CARSFLGYCTNGVCYDPPWVAAAGTGYYYYGMDVW	yes	yes	100.00	yes	100.00
3582	IGHV1-69	100.00	U	IGHD3-16	IGHJ6	CARDPDMITFGGVIVIPGTRGGMDVW	yes	yes	85.54	yes	100.00
	IGHV5-10-1	96.88	M	IGHD2-15	IGHJ4	CARLGLGYCNGGNCPYDYW	yes	yes	14.46	BDL	-
3653	IGHV3-33	100.00	U	IGHD3-9	IGHJ6	CATQYYDILTGYYNYYYYGMDVW	yes	yes	87.50	yes	94.47
	IGHV1-69	100.00	U	IGHD3-10	IGHJ6	CASGMVRGVIEYYYYYYGMDVW	yes	yes	9.12	yes	5.53
	IGHV4-39	100.00	U	IGHD3-16	IGHJ4	CARRAVGDYVLPNYFDYW	yes	yes	3.38	BDL	-
3698	IGHV3-21	100.00	U	IGHD6-19	IGHJ4	CARQLYSSGWYPSRAFDYW	yes	yes	95.90	yes	100.00
	IGHV6-1	95.62	M	IGHD3-22	IGHJ4	CARSGYFREWLDCW	yes	yes	4.10	BDL	-
3868	IGHV1-69	86.81	M	IGHD4-17	IGHJ4	CVRDRGLTVTSGGYFDYW	yes	yes	90.20	yes	93.55
	IGHV3-49	100.00	U	IGHD1-26	IGHJ3	CTRSGSYWGGDAFDIW	yes	yes	4.90	yes	2.94
	IGHV6-1	97.64	M	IGHD2-21	IGHJ4	CARGDSNCDYW	yes	yes	4.90	BDL	-
	IGHV3-30 = IGHV3-30-5	95.61	M	IGHD3-22	IGHJ6	CAKDLGGYYDSSGYLYYFYGMDVW	no	BDL	-	yes	3.50
3952	IGHV3-64	100.00	U	IGHD3-3	IGHJ5	CVKEEYYDFWSGYPPPWFDPW	yes	yes	57.75	yes	2.24

sample ID	IGHV	IGHV identity	IGHV mutational status	IGHD	IGHJ	CDR3	Sanger	LYNX	LYNX clone proportion	IG aNGS	IG aNGS clone proportion
3952	IGHV4-59	100.00	U	IGHD4-11	IGHJ6	CARERQVTTDVYGMDVW	yes	yes	42.25	yes	97.76
4001	IGHV3-23	92.71	M	IGHD2-21	IGHJ6	CARGVATSSFYDYGLDVW	yes	yes	100.00	yes	100.00
	IGHV3-7	92.71	M	IGHD2-21	IGHJ4	CARGPFVVISAPDYW	yes	no	-	BDL	-
4020	IGHV3-7	96.88	M	IGHD3-22	IGHJ4	CASYYCDQKGCPIEKWDW	yes	yes	98.16	yes	100.00
	IGHV4-34	97.89	M	IGHD6-13	IGHJ4	CARGWIAAPGTPNDFDYW	yes	yes	1.84	BDL	-
4180	IGHV3-48	97.57	M	IGHD4-23	IGHJ3	CARDGVGAPIW	yes	yes	75.39	yes	71.09
	IGHV1-69	100.00	U	IGHD1-26	IGHJ6	CARDGPLRIVGATTQYYYYYGMDVW	yes	yes	24.61	yes	28.91
4207	IGHV1-69	98.96	U	IGHD2-8	IGHJ4	CATSYYFDYW	yes	yes	100.00	yes	75.30
	IGHV4-59	94.39	M	IGHD3-22	IGHJ4	CARETYYYETSGYYPYYFDSW	yes	no	-	yes	23.53
	IGHV4-4	99.66	U	IGHD6-19	IGHJ6	CARVGQWLANSYYYYGMDVW	no	no	-	yes	1.18
4211	IGHV4-38-2	100.00	U	IGHD6-13	IGHJ6	CARDDKQQLVQTGPWDYYYGMDVW	yes	yes	93.36	yes	81.09
	IGHV3-33	94.44	M	IGHD4-23	IGHJ4	CARVRGSPISYFDYW	yes	yes	6.64	yes	18.91
4264	IGHV4-59	100.00	U	IGHD2-21	IGHJ3	CARNPYCGGDCYSDAFDIW	yes	yes	100.00	yes	100.00
	IGHV3-21	94.79	M	IGHD3-16	IGHJ4	CARDLRLGDLSLYNEFDYW	yes	no	-	BDL	-
4307	IGHV4-34	92.28	M	IGHD3-9	IGHJ4	CARSRSPTFDYYFDQW	yes	yes	100.00	yes	65.33
	IGHV3-64	87.37	M	IGHD1-26	IGHJ4	CAREEGGWSHYDKW	yes	no	-	yes	33.09
	IGHV3-7	90.37	M	IGHD6-13	IGHJ4	CARGVIAAGLFDSW	no	no	-	yes	1.58
4313	IGHV1-69	99.65	U	IGHD3-3	IGHJ4	CARGGEDYDFWSGYFTDRGSGRGHDVW	yes	yes	100.00	yes	94.23
	IGHV1-69	100.00	U	IGHD3-10	IGHJ4	CARDTGSFGVIATSYYFDYW	no	no	-	yes	3.24
	IGHV3-9	100.00	U	IGHD3-3	IGHJ4	CAKDYYDFWSGYPNASPFDYW	no	no	-	yes	2.53
4338	IGHV5-51	95.83	M	IGHD4-23	IGHJ6	CARLREGGTSLDAMDVW	yes	yes	99.02	yes	90.96
	IGHV3-30	99.65	U	IGHD6-19	IGHJ4	CARGSSGWYHDYW	yes	yes	0.98	yes	7.59
	IGHV3-9	100.00	U	IGHD6-13	IGHJ1	CANTFLPLSSSWPTEYFQHW	no	no	-	yes	1.44
4432	IGHV3-11	96.88	M	IGHD6-13	IGHJ4	CARGAAATGDFDYW	yes	yes	100.00	yes	100.00

sample ID	IGHV	IGHV identity	IGHV mutational status	IGHD	IGHJ	CDR3	Sanger	LYNX	LYNX clone proportion	IG aNGS	IG aNGS clone proportion
4439	IGHV1-8	98.26	U	IGHD2-2	IGHJ6	CARVGGYDSRTSTSNYYNGMDVW	yes	yes	98.40	yes	95.62
	IGHV3-23	99.65	U	IGHD3-3	IGHJ6	CAKWRVDDFWSGYDVPYYYYGMDVW	yes	yes	1.60	yes	4.38
4444	IGHV3-23	100.00	U	IGHD3-10	IGHJ3	CAKDRGSPRIPLLWFGELGDDAFDIW	yes	yes	100.00	yes	100.00
4446	IGHV3-74	95.14	M	IGHD5-18	IGHJ4	CAREGNGYTYGTPDYW	yes	yes	100.00	yes	100.00
	IGHV1-8	99.31	U	IGHD6-13	IGHJ6	CARDTRRIEGNGMDVW	yes	no	-	no	-
4451	IGHV3-7	95.83	M	IGHD3-22	IGHJ3	CTRDANYFDGSAYYDAFDIW	yes	yes	97.00	yes	100.00
	IGHV4-59	91.03	M	IGHD2-2	IGHJ2	CARSPSLYGDWYFDLW	no	yes	3.00	BDL	-
4478	IGHV4-39	100.00	U	IGHD6-19	IGHJ6	CARLGIAVARGYGMDVW	yes	yes	100.00	yes	100.00
4481	IGHV3-53	100.00	U	IGHD3-3	IGHJ5	CARVRSRKILFGKQYNWFDPW	yes	yes	98.900	yes	100.00
	IGHV1-2	100.00	U	IGHD6-13	IGHJ5	CAREVYIAAAVRIKRGWFDPW	yes	yes	1.10	BDL	-
4636	IGHV3-21	91.32	M	IGHD6-13	IGHJ6	CARDKWDSSNWDSYGMDVWGQGGMDVW	yes	yes	100.00	yes	100.00
	IGHV5-51	93.06	M	IGHD3-22	IGHJ4	CAKEVVTRIHFFDNW	yes	no	-	BDL	-
4724	IGHV1-69	100.00	U	IGHD3-16	IGHJ4	CARDTGSFGVIATSYYFDYW	yes	yes	100.00	yes	100.00
4731	IGHV2-5	93.13	M	IGHD2-2	IGHJ4	CARRLLTVVPYYFDFW	yes	yes	100.00	yes	95.97
	IGHV3-74	94.10	M	IGHD6-19	IGHJ4	CAAAFHSGCPGSYW	yes	BDL	-	yes	4.03
4834	IGHV4-61	92.98	M	IGHD2-2	IGHJ6	CARGVARVQLLDVFYYYFAMDVW	yes	yes	100.00	yes	98.53
	IGHV1-8	96.88	M	IGHD2-2	IGHJ5	CARGIVIKPAVYWYWFDPW	yes	no	-	BDL	-
	IGHV4-38-2	96.93	M	IGHD3-3	IGHJ4	CARNYDFWSGHDYW	no	BDL	-	yes	1.47
4854	IGHV2-5	95.19	M	IGHD2-2	IGHJ5	CSHWATSCWSTNCYESGGWENNWFDPW	yes	yes	71.72	yes	83.45
	IGHV4-34	93.68	M	IGHD3-16	IGHJ4	CARGTRQGGAATVDYW	yes	yes	23.45	yes	11.43
	IGHV3-7	95.95	M	IGHD1-26	IGHJ4	CVRDKDGSIVGATLLDYW	no	yes	4.83	BDL	-
	IGHV3-48	99.32	U	IGHD7-27	IGHJ4	CASYQNEGFW	no	no	-	yes	5.12
4904	IGHV1-46	100.00	U	IGHD6-6	IGHJ4	CARGQSSIAAWYYFDYW	yes	yes	100.00	yes	100.00
4905	IGHV3-33	100.00	U	IGHD4-17	IGHJ6	CARGQQGDYVYGMDVW	yes	yes	99.19	yes	100.00

sample ID	IGHV	IGHV identity	IGHV mutational status	IGHD	IGHJ	CDR3	Sanger	LYNX	LYNX clone proportion	IG aNGS	IG aNGS clone proportion
4905	IGHV3-21	100.00	U	IGHD3-10	IGHJ5	CARDRVAVLYYYGSGSYSLSPEFDPW	no	yes	0.81	BDL	-
4912	IGHV4-34	92.63	M	IGHD1-7	IGHJ5	CARGQYNWNYRFDSW	yes	yes	78.86	yes	82.63
	IGHV3-23	94.44	M	IGHD5-12	IGHJ6	CAKDGDDSCMDVW	yes	yes	21.14	yes	17.37
4941	IGHV6-1	100.00	U	IGHD6-19	IGHJ6	CARDLSSGWPRYYYYYMDVW	yes	yes	69.97	yes	22.41
	IGHV4-34	94.74	M	IGHD6-19	IGHJ6	CGAVAGNPDYYGMDVW	yes	yes	30.03	yes	77.59
4947	IGHV3-23	93.75	M	IGHD3-10	IGHJ4	CAKDSSGGYSNPYNFDFW	yes	yes	100.00	yes	100.00
4954	IGHV1-69	100.00	U	IGHD3-3	IGHJ4	CASAHDFWSGLDYW	yes	yes	100.00	yes	100.00
	IGHV4-59	97.54	M	IGHD3-22	IGHJ4	CASGRGSYYFDSGGFDPPDYW	yes	BDL	-	BDL	-
	IGHV3-23	93.75	M	IGHD6-6	IGHJ5	CAKDAQFGIIGRPSW	yes	no	-	BDL	-
4973	IGHV1-2	100.00	U	IGHD6-19	IGHJ4	CARDGTDISSGSQSPFYYFDYW	yes	yes	80.19	yes	98.55
	IGHV5-51	98.96	U	IGHD2-8	IGHJ4	CARHFTGEVDYW	yes	yes	19.81	yes	1.45
5032	IGHV3-48	99.31	U	IGHD7-27	IGHJ4	CASYQNEGFW	yes	yes	100.00	yes	100.00
	IGHV5-10-1	91.67	M	IGHD1-1	IGHJ5	CARRLAVW	yes	BDL	-	BDL	-
5061	IGHV1-46	100.00	U	IGHD3-22	IGHJ5	CASGIGGYYDSSGYYELRWGNWFDPW	yes	yes	97.28	yes	100.00
	IGHV5-51	98.96	U	IGHD2-2	IGHJ6	CARLYCSSTSCFYYYYGMDVW	yes	yes	2.72	BDL	-
5300	IGHV3-7	94.10	M	IGHD3-10	IGHJ4	CARDPGNSGSYEFDYW	NA	yes	100.00	yes	100.00
5413	IGHV1-3	100.00	U	IGHD6-19	IGHJ4	CAREQWLVMITFDYW	yes	yes	93.05	yes	87.28
	IGHV3-53	100.00	U	IGHD5-18	IGHJ6	CASQLGNVDTAMVSGYYYYYGMDVW	yes	yes	6.95	yes	12.72
5417	IGHV1-18	97.57	M	IGHD2-2	IGHJ6	CARDQGYCSSTSCYEGNYYYYYGMDVW	NA	yes	63.47	yes	38.29
	IGHV3-72	94.90	M	IGHD5-18	IGHJ4	CASFTAKFGSNIRKFDYW	NA	yes	34.73	yes	60.35
	IGHV3-7	92.88	M	IGHD2-21	IGHJ4	CARGYCGGDCFYW	NA	yes	1.80	yes	1.37
5434	IGHV3-23	96.18	M	IGHD3-3	IGHJ4	CAKQTYYDFWSGYYRYFDSW	yes	yes	94.58	yes	96.26
	IGHV4-31	94.16	M	IGHD3-22	IGHJ2	CARRLDYYDTGGYHLGATDWYFDLW	yes	yes	5.42	BDL	-
	IGHV1-46	96.23	M	IGHD3-3	IGHJ5	CAKSSEGGGFDPW	no	no	-	yes	2.50

sample ID	IGHV	IGHV identity	IGHV mutational status	IGHD	IGHJ	CDR3	Sanger	LYNX	LYNX clone proportion	IG aNGS	IG aNGS clone proportion
5434	IGHV5-51	100.00	U	IGHD6-19	IGHJ4	CARHQWYLLYYFDYW	no	no	-	yes	1.24
5510	IGHV4-39	100.00	U	IGHD6-13	IGHJ5	CARVWGYSSSWYGGTNWFDPW	yes	yes	84.62	yes	66.96
	IGHV3-23	95.83	M	IGHD3-22	IGHJ4	CAKGGALWFYDSSGLGYFDYW	yes	yes	8.46	yes	12.35
	IGHV3-7	95.56	M	IGHD3-16	IGHJ6	CARKAHYAMDVW	no	yes	6.92	yes	14.78
	IGHV4-59	94.46	M	IGHD3-22	IGHJ4	CAKHFYESSGYPDW	no	BDL	-	yes	5.92
5529	IGHV3-7	92.71	M	IGHD2-2	IGHJ4	CASAYKSRCLFEFW	yes	yes	100.00	yes	100.00
	IGHV5-51	93.75	M	IGHD1-26	IGHJ4	CARLSTGSYSLVDQW	yes	no	-	BDL	-
5549	IGHV3-48	94.44	M	IGHD1-1	IGHJ3	CAREVNWNDLRFEPFDFW	yes	BDL	-	BDL	-
	IGHV4-34	94.39	M	IGHD5-24	IGHJ1	CARQGTIEVATMSAW	yes	BDL	-	BDL	-
	IGHV1-3	96.53	M	IGHD4-11	IGHJ5	CARIPYMSTVTTGWFDPW	yes	no	-	BDL	-
	IGHV3-33	93.75	M	IGHD6-13	IGHJ4	CARDCTIGSSWYCGLDYW	yes	no	-	no	-
	IGHV3-15	89.16	M	IGHD4-17	IGHJ4	CTTAHDYGDYWTRGGLGLRDYW	no	yes	100.00	yes	100.00
5550	IGHV5-10-1	100.00	U	IGHD6-19	IGHJ4	CARQQWLGGEYFDYW	yes	yes	100.00	yes	55.78
	IGHV3-15	97.28	M	IGHD4-17	IGHJ4	CTPDHIVVPPTHPYLLTTVTTVDCW	yes	BDL	-	yes	44.22
5559	IGHV4-38-2	100.00	U	IGHD6-13	IGHJ6	CARDDKQQLVQTGPWDYYYGMDVW	NA	yes	100.00	yes	100.00
5599	IGHV1-3	100.00	U	IGHD3-10	IGHJ4	CARDQWLTKLPALFDYW	yes	yes	73.38	yes	94.70
	IGHV6-1	100.00	U	IGHD6-6	IGHJ4	CSRGPSIGDPFDYW	yes	yes	25.26	yes	5.30
	IGHV1-2	99.66	U	IGHD4-17	IGHJ6	CARDSQKTTVTRRWLYYYYYMDVW	no	yes	1.37	BDL	-
5726	IGHV4-34	100.00	U	IGHD2-2	IGHJ6	CSIVVVPAAMIPDYYYYGMDVW	yes	yes	100.00	yes	100.00
	IGHV2-5	100.00	U	IGHD1-26	IGHJ4	CAHRLSWVGSFYFDYW	yes	no	-	no	-
5733	IGHV5-51	100.00	U	IGHD4-23	IGHJ4	CARHQWYLLYYFDYW	yes	yes	100.00	yes	93.80
	IGHV1-18	100.00	U	IGHD3-3	IGHJ4	CARVTTEIYFIRFLEWSKVRTPFDYW	yes	no	-	BDL	-
	IGHV1-46	96.23	M	IGHD3-3	IGHJ5	CAKSSEGGGFDPW	no	no	-	yes	6.20

sample ID	IGHV	IGHV identity	IGHV mutational status	IGHD	IGHJ	CDR3	Sanger	LYNX	LYNX clone proportion	IG aNGS	IG aNGS clone proportion
5764	IGHV3-30 = IGHV3-30-5	95.49	M	IGHD6-19	IGHJ6	CAKVALAGTFDDYKYYGMDVW	yes	yes	100.00	yes	98.75
	IGHV2-5	95.88	M	IGHD3-22	IGHJ4	CTHSWIWRGVVISTTYFDYW	yes	no	-	no	-
	IGHV3-74	95.27	M	IGHD5-18	IGHJ4	CAREGNGYTYGTPDYW	no	no	-	yes	1.25
5804	IGHV4-39	100.00	U	IGHD2-2	IGHJ6	CARHRLGYCSSTSCYYYYYGMDVW	yes	yes	89.36	yes	88.02
	IGHV1-69	100.00	U	IGHD6-19	IGHJ4	CARGPNSSGWSGFGYYFDYW	yes	yes	8.51	yes	11.98
	IGHV1-8	100.00	U	IGHD3-3	IGHJ4	CARGLGGKDDFWGGNRDSW	no	yes	2.13	BDL	-
5814	IGHV3-23	89.58	M	IGHD6-13	IGHJ5	CAKEGAGEGRPTFDPW	yes	yes	68.89	yes	100.00
	IGHV3-30	90.63	M	IGHD2-15	IGHJ5	CVRDFTLRGGGGFYGVDHW	yes	yes	31.11	BDL	-
5837	IGHV3-23	92.01	M	IGHD6-19	IGHJ4	CANSVGGYSYQWLEGGAYYFDYW	yes	yes	100.00	yes	92.37
	IGHV3-21	100.00	U	IGHD2-15	IGHJ6	CARDPRYCSGGSCYSVWWNVW	yes	no	-	no	-
	IGHV1-69	99.66	U	IGHD2-2	IGHJ6	CASSRDLIVVVPAAIPGDYYYYGMDVW	no	no	-	yes	6.01
	IGHV3-30 = IGHV3-30-5	93.20	M	IGHD3-10	IGHJ4	CAKSGGIFMIYVQSQPLDYW	no	no	-	yes	1.62
5842	IGHV4-38-2	94.44	M	IGHD3-22	IGHJ3	CATARWFLPHGPFDIW	yes	yes	95.83	yes	93.56
	IGHV3-74	92.36	M	IGHD3-16	IGHJ4	CAGLRMGEGLDSW	yes	yes	4.17	yes	5.13
	IGHV3-30 = IGHV3-30-5	90.54	M	IGHD5-18	IGHJ3	CAKEYSHGSWAFDIW	no	no	-	yes	1.31
5851	IGHV3-7	93.06	M	IGHD3-22	IGHJ4	CVRGGYYHDFW	yes	yes	100.00	yes	98.71
	IGHV6-1	94.95	M	IGHD3-16	IGHJ4	CARCGPSTAGGVCAYW	yes	no	-	BDL	-
	IGHV3-30 = IGHV3-30-5	97.30	M	IGHD1-1	IGHJ4	CAKDPHPWKDNNIWYNEGPDYW	no	no	-	yes	1.29
5880	IGHV1-18	100.00	U	IGHD3-9	IGHJ4	CARDLFADQFKAQLRYFDWLQNYFDYW	yes	yes	100.00	yes	98.70
	IGHV4-59	100.00	U	IGHD3-22	IGHJ4	CARGATYYYDSSGYISDLCMDYW	yes	no	-	BDL	-
	IGHV1-46	96.23	M	IGHD3-3	IGHJ5	CAKSSEGGGFDPW	no	no	-	yes	1.30

sample ID	IGHV	IGHV identity	IGHV mutational status	IGHD	IGHJ	CDR3	Sanger	LYNX	LYNX clone proportion	IG aNGS	IG aNGS clone proportion
5912	IGHV1-46	96.18	M	IGHD3-3	IGHJ5	CAKSSEGGGFDPW	yes	yes	100.00	yes	100.00
	IGHV4-4	98.26	U	IGHD5-18	IGHJ4	CARGLVGGYNHGYHTTHSANW	yes	no	-	no	-
5920	IGHV3-74	95.27	M	IGHD5-18	IGHJ4	CAREGNGYTYGTPDYW	NA	yes	100.00	yes	100.00
5931	IGHV1-69	100.00	U	IGHD3-3	IGHJ6	CARVQMDNDDFWSGYYQPLRYYYYGMDVW	yes	yes	100.00	yes	100.00
	IGHV3-7	95.83	M	IGHD1-14	IGHJ4	CVKGGWENPIDYW	yes	no	-	no	-
5932	IGHV1-69	100.00	U	IGHD3-16	IGHJ6	CARDPGDYIWGSLVARTPASPYGMDVW	yes	yes	100.00	yes	100.00
	IGHV2-5	97.59	M	IGHD2-21	IGHJ4	CAHTDRLVSSSYFDYW	yes	BDL	-	BDL	-
5940	IGHV1-69	99.65	U	IGH2-2	IGHJ6	CASSRDLIVVVPAAIPGDYYYYGMDVW	yes	yes	100.00	yes	98.59
	IGHV3-30-3	100.00	U	IGHD3-22	IGHJ4	CARGSRPPIYYDSSGSLDYW	yes	no	-	yes	1.41
6004	IGHV1-69	100.00	U	IGHD3-22	IGHJ4	CASPQLPYYYDSSGYYLPLDYW	yes	yes	100.00	yes	100.00
	IGHV4-30-2	95.19	M	IGHD3-22	IGHJ6	CARDHIRDGMDVW	yes	no	-	BDL	-
6065	IGHV3-30 = IGHV3-30-5	97.22	M	IGHD6-13	IGHJ4	CAKDPHPWKDNNIWYNEGPDYW	yes	yes	100.00	yes	97.09
	IGHV1-69	100.00	U	IGHD3-3	IGHJ6	CARVSIFGVVNIYYYYYYMDVW	yes	BDL	-	yes	2.91

Appendix E Overview of IGH clonality in multiclonal CLL

Tab. 5: A comprehensive overview of IGH clonality results in multiclonal CLL patients.
The table presents a detailed overview of IGH clonality results in multiclonal CLL patients, summarizing the number and characteristics of productive IGH rearrangements identified by each detection method. U, unmutated IGHV; M, mutated IGHV; X, unknown IGHV mutational status.

patient ID	sample ID	sampling date	Sanger clonality	Sanger IGHV mutational status	LYNX clonality	LYNX IGHV mutational status	IG aNGS clonality	IG aNGS IGHV mutational status
BRNO0319	1741	15.04.2011	triclonal	U+U+U	tetraclonal	U+U+X+X	triclonal	U+U+U
	1531	16.04.2014	biclonal	U+U	tetraclonal	U+U+U+X	biclonal	U+U
	4905	01.04.2021	monoclonal	U	biclonal	U+U	monoclonal	U
BRNO0511	1732	14.09.2007	triclonal	U+U+U	triclonal	U+U+X	triclonal	U+U+M
BRNO0604	1738	29.04.2011	biclonal	U+U	biclonal	U+U	biclonal	U+U
	2807	16.01.2017	monoclonal	U	monoclonal	U	biclonal	U+U
BRNO0814	521	15.12.2008	biclonal	U+U	biclonal	X+X	pentaclonal	U+U+U+M+M
BRNO0862	463B	25.05.2010	biclonal	U+U	biclonal	X+X	biclonal	U+U
	3436	05.06.2018	monoclonal	U	monoclonal	U	monoclonal	U
BRNO0923	467	22.06.2009	biclonal	U+U	biclonal	U+X	biclonal	U+U
BRNO1003	2129	06.11.2015	biclonal	U+U	monoclonal	U	monoclonal	U
BRNO1029	1933	17.03.2010	biclonal	U+U	biclonal	U+X	biclonal	U+U
	4444	30.10.2018	monoclonal	U	monoclonal	U	monoclonal	U
BRNO1037	568	04.04.2011	triclonal	U+M+M	triclonal	X+X+X	triclonal	U+M+M
	5417	04.05.2022	NA	-	triclonal	M+M+X	triclonal	M+M+M
BRNO1046	3582	26.10.2018	biclonal	U+M	biclonal	U+X	monoclonal	U
BRNO1048	1123	29.04.2010	monoclonal	M	biclonal	M+X	triclonal	M+M+M
	2777	10.02.2017	biclonal	M+M	biclonal	M+M	biclonal	M+M

patient ID	sample ID	sampling date	Sanger clonality	Sanger IGHV mutational status	LYNX clonality	LYNX IGHV mutational status	IG aNGS clonality	IG aNGS IGHV mutational status
BRNO1048	4432	27.04.2020	monoclonal	M	monoclonal	M	monoclonal	M
BRNO1054	1731	01.06.2011	biclonal	M+M	biclonal	M+M	biclonal	M+M
BRNO1072	1766	12.07.2010	biclonal	U+U	biclonal	U+U	biclonal	U+U
BRNO1132	1622	17.01.2011	biclonal	M+M	biclonal	M+M	biclonal	M+M
BRNO1278	1436	17.09.2012	biclonal	M+M	biclonal	M+M	biclonal	M+M
BRNO1282	2669	14.11.2016	biclonal	M+M	monoclonal	M	monoclonal	M
	4947	30.04.2021	monoclonal	M	monoclonal	M	monoclonal	M
BRNO1502	1960	20.09.2013	triclonal	M+M+M	triclonal	M+M+X	triclonal	M+M+M
	4451	13.05.2020	monoclonal	M	biclonal	M+M	monoclonal	M
BRNO1525	1475	15.11.2013	biclonal	U+U	triclonal	U+U+U	tetraclonal	U+U+U+U
	3239	14.09.2016	monoclonal	U	biclonal	U+U	monoclonal	U
BRNO1565	1485	17.03.2014	biclonal	U+U	biclonal	U+X	tetraclonal	U+U+U+U
	4313	08.03.2017	monoclonal	U	monoclonal	U	triclonal	U+U+U
BRNO1616	3066	31.07.2017	biclonal	U+U	biclonal	U+U	biclonal	U+U
	4478	28.05.2020	monoclonal	U	monoclonal	U	monoclonal	U
BRNO1625	1934	16.09.2014	triclonal	M+M+M	biclonal	M+M	biclonal	M+M
	5300	24.01.2022	NA	-	monoclonal	M	monoclonal	M
BRNO1680	1750	30.01.2015	biclonal	M+M	monoclonal	M	monoclonal	M
BRNO1689	4180	18.11.2019	biclonal	U+M	biclonal	U+X	biclonal	U+M
BRNO1718	1894	11.05.2015	biclonal	U+U	biclonal	U+U	biclonal	U+U
	4904	01.04.2021	monoclonal	U	monoclonal	U	monoclonal	U
BRNO1774	4954	05.05.2021	triclonal	U+M+M	monoclonal	U	monoclonal	U
BRNO1788	2606	21.06.2016	biclonal	U+M	monoclonal	U	monoclonal	U
BRNO1791	3952	09.10.2015	biclonal	U+U	biclonal	U+U	biclonal	U+U

patient ID	sample ID	sampling date	Sanger clonality	Sanger IGHV mutational status	LYNX clonality	LYNX IGHV mutational status	IG aNGS clonality	IG aNGS IGHV mutational status
BRNO1794	2101	12.10.2015	monoclonal	U	monoclonal	U	monoclonal	U
	4439	06.05.2020	biclonal	U+U	biclonal	U+X	biclonal	U+U
BRNO1795	3190	21.11.2017	biclonal	M+M	triclonal	M+M+M	triclonal	M+M+M
BRNO1797	5032	18.06.2021	biclonal	U+M	monoclonal	U	monoclonal	U
BRNO1841	2683	10.102016	monoclonal	M	monoclonal	M	biclonal	U+M
	4338	18.02.2020	biclonal	U+M	biclonal	M+X	triclonal	U+U+M
BRNO1908	5061	26.07.2018	biclonal	U+U	biclonal	U+X	monoclonal	U
BRNO1938	4020	23.08.2019	biclonal	M+M	biclonal	M+X	monoclonal	M
BRNO1999	2748	23.01.2017	biclonal	M+M	monoclonal	M	tetraclonal	M+M+M+M
	4307	11.01.2019	biclonal	M+M	monoclonal	M	triclonal	M+M+M
BRNO2003	3034	17.07.2017	monoclonal	U	biclonal	U+U	biclonal	U+U
	3289	12.02.2018	biclonal	U+U	biclonal	U+U	biclonal	U+U
BRNO2013	2995	10.03.2017	biclonal	U+M	monoclonal	U	monoclonal	U
BRNO2088	3377	06.04.2018	biclonal	U+U	monoclonal	U	tetraclonal	U+M+M+M
	4724	23.11.2020	monoclonal	U	monoclonal	U	monoclonal	U
BRNO2174	3304	16.02.2018	biclonal	M+M	monoclonal	M	polyclonal	-
BRNO2336	3653	16.01.2019	triclonal	U+U+U	triclonal	U+X+X	biclonal	U+U
BRNO2351	3698	19.02.2019	biclonal	U+M	biclonal	U+X	monoclonal	U
BRNO2422	3868	27.05.2019	triclonal	U+M+M	triclonal	M+X+X	triclonal	U+M+M
BRNO2471	4001	09.08.2019	biclonal	M+M	monoclonal	M	monoclonal	M
BRNO2483	4446	12.05.2020	biclonal	U+M	monoclonal	M	monoclonal	M
	5920	21.06.2023	NA	-	monoclonal	X	monoclonal	M
BRNO2538	4207	02.12.2019	biclonal	U+M	monoclonal	U	triclonal	U+U+M
BRNO2543	4211	05.12.2019	biclonal	U+M	biclonal	M+X	biclonal	U+M

patient ID	sample ID	sampling date	Sanger clonality	Sanger IGHV mutational status	LYNX clonality	LYNX IGHV mutational status	IG aNGS clonality	IG aNGS IGHV mutational status
BRNO2543	5559	05.09.2022	NA	-	monoclonal	X	monoclonal	U
BRNO2568	4264	24.01.2020	biclonal	U+M	monoclonal	U	monoclonal	U
BRNO2614	5764	08.02.2023	biclonal	M+M	monoclonal	M	biclonal	M+M
BRNO2628	4481	29.05.2020	biclonal	U+UU	biclonal	U+X	monoclonal	U
BRNO2680	4636	10.09.2020	biclonal	M+M	monoclonal	M	monoclonal	M
BRNO2727	5932	04.07.2023	biclonal	U+M	monoclonal	X	monoclonal	U
BRNO2737	4731	27.11.2020	biclonal	M+M	monoclonal	M	biclonal	M+M
BRNO2792	4834	19.02.2021	biclonal	M+M	monoclonal	M	biclonal	M+M
BRNO2798	4854	08.03.2021	biclonal	M+M	triclonal	M+M+X	triclonal	U+M+M
BENO2831	4912	09.04.2021	biclonal	M+M	biclonal	M+M	biclonal	M+M
BRNO2848	4941	27.04.2021	biclonal	U+M	biclonal	U+M	biclonal	U+M
BRNO2860	4973	18.05.2021	biclonal	U+U	biclonal	U+U	biclonal	U+U
BRNO2959	5837	13.04.2023	biclonal	U+M	monoclonal	M	triclonal	U+M+M
BRNO3082	5413	02.05.2022	biclonal	U+U	biclonal	U+X	biclonal	U+U
BRNO3093	5434	23.05.2022	biclonal	M+M	biclonal	M+X	triclonal	U+M+M
BRNO3142	5510	21.07.2022	biclonal	U+M	triclonal	U+X+X	tetraclonal	U+M+M+M
BRNO3149	5529	18.08.2022	biclonal	M+M	monoclonal	X	monoclonal	M
BRNO3160	5549	26.08.2022	tetraclonal	M+M+M+M	monoclonal	M	monoclonal	M
BRNO3161	5550	26.08.2022	biclonal	U+M	monoclonal	U	biclonal	U+M
BRNO3184	5599	23.09.2022	biclonal	U+U	triclonal	U+U+X	biclonal	U+U
BRNO3246	5726	05.01.2023	biclonal	U+U	monoclonal	U	monoclonal	U
BRNO3252	5733	11.01.2023	biclonal	U+U	monoclonal	U	biclonal	U+M
BRNO3294	5804	10.03.2023	biclonal	U+U	triclonal	U+X+X	biclonal	U+U
BRNO3300	5814	17.03.2023	biclonal	M+M	biclonal	M+X	monoclonal	M

patient ID	sample ID	sampling date	Sanger clonality	Sanger IGHV mutational status	LYNX clonality	LYNX IGHV mutational status	IG aNGS clonality	IG aNGS IGHV mutational status
BRNO3305	5851	28.04.2023	biclonal	M+M	monoclonal	X	biclonal	M+M
BRNO3313	5842	21.04.2023	biclonal	M+M	biclonal	M+X	triclonal	M+M+M
BRNO3330	5880	22.05.2023	biclonal	U+U	monoclonal	X	biclonal	U+M
BRNO3351	5912	19.06.2023	biclonal	U+M	monoclonal	M	monoclonal	M
BRNO3356	5931	03.07.2023	biclonal	U+M	monoclonal	X	monoclonal	U
BRNO3364	5940	19.07.2023	biclonal	U+U	monoclonal	U	biclonal	U+U
BRNO3390	6004	12.09.2023	biclonal	U+M	monoclonal	X	monoclonal	U
BRNO3413	6065	24.10.2023	biclonal	U+M	monoclonal	X	biclonal	U+M