Aus dem Zentrum für Innere Medizin der Universität Heidelberg
(Zentrumssprecher: Prof. Dr. med. Dirk Jäger)
Klinik für Hämatologie, Onkologie und Rheumatologie
(Ärztlicher Direktor: Prof. Dr. med. Carsten Müller-Tidow)

Biological and clinical impact of the immunoglobulin light chain sequence diversity in patients with dominant heart AL amyloidosis

Inauguraldissertation
zur Erlangung des Doctor scientiarum humanarum (Dr. sc. hum.)
an der
Medizinischen Fakultät Heidelberg
der
Ruprecht-Karls-Universität

vorgelegt von
Natalie Berghaus

aus
Wiesbaden

2023

Dekan: Herr Prof. Dr. Dr. h.c. Hans-Georg Kräusslich
Doktorvater: Herr Prof. Dr. med. Stefan O. Schönland

Table of Contents

List of Abbreviations

AA	Amino acid
AApoAI	Apolipoprotein AI amyloidosis
AFib	Fibrinogen α chain amyloidosis
AGG	Beta sheet aggregation tendency value
AL	AL amyloidosis
AL < dFLC	A subgroup of light chain sequences corresponding to AL amyloidosis patients with a dFLC value <180 mg/L
AL > dFLC	A subgroup of light chain sequences corresponding to AL amyloidosis patients with a dFLC value >180 mg/L
AL HC	A subgroup of light chain sequences corresponding to AL amyloidosis patients with detectable clonal heavy chain in serum
AL no HC	A subgroup of light chain sequences corresponding to AL amyloidosis patients without detectable clonal heavy chain in serum
AL_D	AL amyloidosis patients with no dominant heart, kidney and heart, and kidney involvement
AL_H	AL amyloidosis patients with dominant heart involvement
AL_HK	AL amyloidosis patients with dominant heart and kidney involvement
AL_HTX	AL amyloidosis patients who received a heart transplant
AL_K	AL amyloidosis patients with dominant kidney involvement
AL+MM	AL amyloidosis patients who were also diagnosed with multiple myeloma
ALECT2	Leukocyte chemotactic factor 2 amyloidosis
ALZ	Zentrallabor-Analysezentrum University Hospital Heidelberg
Amp	Amplification
ANS	Autonomic nerve system
AP	Alkaline phosphatase
ATTRwt	Wild-type transthyretin amyloidosis
Aβ2m	β2-microglobulin dialysis-related amyloidosis
BM	Bone marrow
BNP	B natriuretic peptide
C	Constant segment
CA	Chromosomal aberrations
CCND1	Cyclin D1
CCND3	Cyclin D3
cDNA	Complementary deoxyribonucleic acid
CDR	Complementary-determining regions
CH	Constant domain heavy chain
Chi	Significance level was calculated using a two-tailored Pearson-Chi-Quadrat-test
CL	Constant domain light chain
CLIA	Chemiluminescence-immunoassay
CT	Computed tomography
cTNT	Cardiac troponin T

D	Diversity segment
Del	Deletion
dFLC	Difference between disease-associated and uninvolved circulation-free light chains
DNA	Deoxyribonucleic acid
DPD	99mTc-labelled 3,3-diphosphono-1,2-propanodicarboxylic acid
ECLIA	Electrochemiluminescence-immunoassay
eGFR	Estimated glomerular filtration rate
Fab	Antigen-binding fragments of an antibody
Fac	Crystallizable fragment of an antibody
FGFR3	Fibroblast growth factor 3
Fischer	Significance level was calculated using a Fischer test
FISH	Fluorescence in situ hybridization
FL	Focal lesion
FR	Framework region
GC	Germinal center
GFR CKD-EPI	Calculation of the glomerular filtration rate using the formula established by the Chronic Kidney Disease Epidemiology Collaboration
GMMG	German-speaking Myeloma Multicenter Group
GRAVY	Grand average of hydropathicity value
GTC	Guanidinium thiocyanate
HC	Heavy chain
hsTNT	Heart-associated troponin T
HY	Hyperidploid
IFE	Immunofixation electrophoresis in serum
Ig	Immunoglobulin
IGH	Immunoglobulin heavy chain
IGK	Immunoglobulin kappa light chain
IGL	Immunoglobulin lambda light chain
J	Joining segment
L1	Lumbar spine
LC	Light chain
LOH	Loss of heterozygosity
MCL	Molekular-cytogenetisches Labor University Hospital Heidelberg
MDE	Myeloma defining events
Median	Significance level was calculated using a median test
MGUS	Monoclonal gammopathy of undermined significance
MM	Multiple myeloma
MM < dFLC	A subgroup of light chain sequences corresponding to multiple myeloma patients with a dFLC value <180 mg/L
MM > dFLC	A subgroup of light chain sequences corresponding to multiple myeloma patients with a dFLC value >180 mg/L
MM HC	A subgroup of light chain sequences corresponding to multiple myeloma patients with detectable clonal heavy chain in serum

MM no HC	A subgroup of light chain sequences corresponding to multiple myeloma patients without detectable clonal heavy chain in serum
MMSET	Multiple myeloma SET domain
$MP_{\_NGS\_PCR}$	Next-generation sequencing approach addressing AL amyloidosis IGLV3-21 samples based on polymerase-chain-reaction products generated using the VLKL12a_Huhn, VLKL3c_Huhn, VLKL4a_Huhn, VLKL7a_Huhn and JLHD123_rv oligonucleotides
M-protein	Monoclonal protein
MRI	Magnetic resonance imaging
Mw	Molecular weight
$N_{\_NGS\_PCR}$	Next-generation sequencing approach addressing AL amyloidosis IGLV3-21 samples based on polymerase-chain-reaction products generated using the VLKL3_A_fw_NB and JLHD123_rv oligonucleotides
NA	Not available
NGS	Next-generation sequencing
NMR	Nuclear magnetic resonance
NT-proBNP	N-terminal provisional brain natriuretic peptide
PC	Plasma cells
PCL	Plasma cell leukemia
PCR	Polymerase-chain-reaction
pI	Isoelectric point
PIC	Physician in charge
PNS	Peripheral nerve system
PYP	99mtc-labeled pyrophosphate
RNA	Ribonucleic acid
RR	Relative risk
SBP	Systolic blood pressure
Sequence_A	Sequences with the highest percentage ≥1 % in the next-generation sequencing approaches, sequences are named in descending alphabetic order
SMM	Smoldering multiple myeloma
SNC	Single nucleotide variant
SOP	Standard operation procedure
STAT3	Signal transducer and activator of transcription 3
t-test	Significance level was calculated using a two-tailored t-test
U-test	Significance level was calculated using a two-tailored Mann-Whitney U-test
V	Variable segment
VH	Variable domain heavy chain
VL	Variable domain light chain
X	Not clearly determined amino acid
ΔpI	Difference between the isoelectric point of the patient-derived light chain and a respective reference sequence

List of Figures

List of Tables

1 Introduction

1.1 Immunoglobulins

Concerning the human immune system, a distinction must be made between the inborn and the acquired, or so-called adaptive, immune system. The adaptive immune system is characterized by a rapidly increased immune response which is triggered as soon as renewed contact with the antigen occurs. This effect can be described as immunogenetic memory and depends on antigen recognition and the specificity of the immunological response (Marshall et al. 2018). The cell types most responsible for the adaptive immune system are lymphocytes, particularly B and T cells with their effector molecules (Kamradt and Ferrari-Kühne 2011). A single B cell that differentiates into an Ig-producing plasma cell produces each immunoglobulin (Ig) molecule. Ig recognize antigens in their native state on the surface of foreign bodies and trigger elimination processes like cell lysis or phagocytosis. T cells recognize processed antigens which are presented as peptides by the highly polymorphic major histocompatibility proteins (Lefranc 2014; Lefranc and Lefranc 2020). These biological functions of B and T cells are accompanied by the need for a large diversity of their antigen receptors. A healthy individuum's potential antigen receptor repertoire is about 2×10^{12} different Ig and T cell receptors. Here, the number of B and T cells an organism is genetically limited to produce is the primary limiting factor (Lefranc and Lefranc 2020).

1.1.1 Structure of Human Immunoglobulins

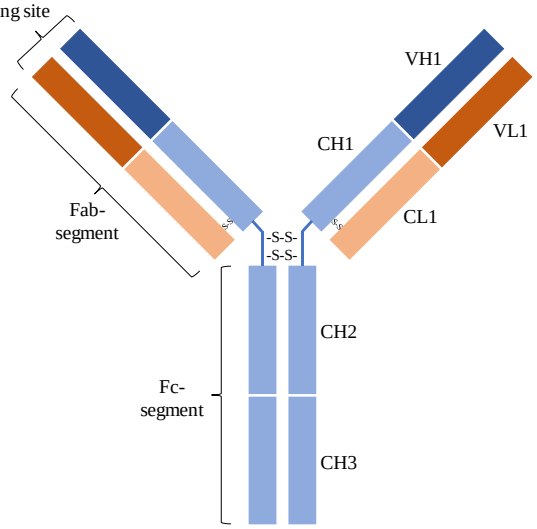
In 1972, Gerald Edelman and Rodney Porter received the Nobel Prize in physiology and medicine for their investigations of the Ig structure (PressRelease_NobelPrize.org 1972). They were the first who were able to resolve the different fragments (Fab, Fc), the characteristic Y-like shape, and the two-chain structure (**Figure 1**) (Edelman 1959; Porter 1959; Wilson and Stanfield 2021).

Addressing the two-chain structure, the Ig consists of two identical light chains (LC) with about 25 kDa and two identical heavy chains (HC) with 50 – 70 kDa. The HCs can be classified into H-mu, H-delta, H-gamma, H-alpha, and H-epsilon, and the LCs into kappa and lambda. (Lefranc and Lefranc 2020; Schroeder and Cavacini 2010). Structurally, LCs can be separated further into two domains, one variable domain (VL) comprising a variable (V) region and a joining (J) region, and a constant domain (CL). A HC also consists of a variable domain (VH) comprising a V, a J, and an additional diversity (D) region, as well as three or four constant domains (CH) (**Figure 1**) (Lefranc and Lefranc 2020; Schroeder and Cavacini 2010).

Further, the V regions comprise three complementary-determining regions (CDR) which are solvent-exposed and form the antigen-binding site (Kamradt and Ferrari-Kühne 2011; Lefranc and Lefranc 2020; Schroeder and Cavacini 2010). This goes together with the fact that CDRs are characterized as highly variable sequence regions. Especially the CDR3 region is defined as highly variable and extends over the C-terminal end of the V region, the complete D region, and the N-terminal end of the J region. Overall, the CDR regions are framed by four stable ″frame″ regions (FR) (Schroeder and Cavacini 2010).

Another way to describe an Ig, is to divide it into two antigen-binding fragments (Fab) and one crystallizable fragment (Fc). The two Fab fragments are characterized by their antigen recognition and binding function and include a complete LC, the VH, and one part of the CH (Schroeder and Cavacini 2010). The Fc fragment was also characterized by Porter, who digested a rabbit antibody with papain and was able to crystalize one of the three fragments (Porter 1959).

A



B

Structure definition			Sequence definition
Chain	Domain type	Domain	Region
H-gamma	V	VH	V-D-J-region
H-delta	C	CH1	C-region
H-alpha	C	CH2	
	C	CH3	
H-mu	V	VH	V-D-J-region
H-epsilon	C	CH1	C-region
	C	CH2	
	C	CH3	
	C	CH4	
kappa	V	V-kappa	V-J-region
	C	C1-kappa	C-region
lambda	V	V-lambda	V-J-region
	C	C1-lambda	C-region

Figure 1. A) Schematic structure of an IgG antibody. B) Human immunoglobulin chain, domain structure, structural labels as well as sequence definition for the heavy chain and light chain classes. blue = heavy chain, orange = light chain, blue line = hinge region, S-S = disulfide bridge, VH = variable domain heavy chain, VL = variable domain light chain, CH = constant domain heavy chain, CL = constant domain light chain, H = heavy, V = variable, C = constant, J = joining, D = diversity. Adapted from (Lefranc and Lefranc 2020). The original article was published under the Creative Commons Attribution 4.0 International License (CC BY 4.0, https://creativecommons.org/licenses/by/4.0/).

Ig are expressed in different forms: as membrane proteins, on the surface of mature and memory B cells, as part of the B cell receptor, or as Ig secreted by plasma cells (Lefranc and Lefranc 2020). The secreted Ig can be divided into five different classes based on their distinct HC: IgM, IgD, IgG, IgA, and IgE. They differ in their physicochemical properties, antigenic

determinants, biological functions, and effector properties as well as their abundance in the serum of a heathy individuum (**Table 1**) (Lefranc and Lefranc 2020).

Table 1. Immunoglobulin classes and their approximate percentage of the total immunoglobulins in the adult serum. Percentage of the Immunoglobulin subclasses in the class: IgG1: 70 – 80 %, IgG2: 18 – 23 %, IgG3: 6 – 8 %, IgG4: 2 – 6 %, IgA1: 90 %, IgA2: 10 %. Ig = immunoglobulin. Adapted from (Lefranc and Lefranc 2020). The original article was published under the Creative Commons Attribution 4.0 International License (CC BY 4.0, https://creativecommons.org/licenses/by/4.0/).

Ig class	IgM	IgD	IgG				IgA		IgE
Ig subclasses			IgG1	IgG2	IgG3	IgG4	IgA1	IgA2	
Approximate % total Ig in adult serum	10	0.2	45-53	11-15	3-6	1-4	11-14	1-4	0.004

1.1.2 B Cell Development

In humans, B cell development starts from hematopoietic stem cells in the bone marrow (**Figure 2**). Here, the rearrangement of the immunoglobulin heavy chain (*IGH*) *D* and *J* locus of the Pro-B cell takes place. Afterward, the rearrangement with the *IGHV* locus allows the Pre-B cell to produce a H-mu chain in the cytoplasm. This H-mu chain is expressed on the surface of the cell with a lambda-like chain. This lambda-like chain forms together with a V-pre-B chain the Pre-B cell receptor (Hieter et al. 1981; Korsmeyer et al. 1981; Lefranc and Lefranc 2020).

For the transition from the Pre-B cell receptor towards an immature B cell, the productive rearrangement of the Ig-kappa (*IGK*) or Ig-lambda (*IGL*) locus is required. In this context, the Pre-B cell receptor induces a signal which inhibits further *IGH-V-D* rearrangements and emits a signal to start the LC *V-J* rearrangement. In a first attempt, a productive rearrangement of one chromosome 2 with the kappa locus is strived – if successful, IgM is produced. However, if a productive rearrangement cannot be obtained, a second attempt with the other chromosome 2 and a kappa LC is sought first. If a kappa LC could not be successfully produced with the second chromosome 2 either, a rearrangement on chromosome 22 and the lambda locus is targeted. The kappa/lambda ratio of antibodies in the serum of a healthy individual is approximately 2:1 (Hollis et al. 1989; Lefranc and Lefranc 2020; Schiff et al. 1990).

After a productive rearrangement of the HC and LC loci, the plasma cells expressing functional IgM on the cell surface are called immature B cells. In the next step, B cells with highly autoreactive B cell receptors are sorted out and the remaining cells (transitional B cells) can migrate to the spleen (Eibel et al. 2014). In the spleen, these cells undergo three different transition states, which can be distinguished by the amount of expressed IgM or other surface markers (Chung et al. 2003; Lefranc and Lefranc 2020).

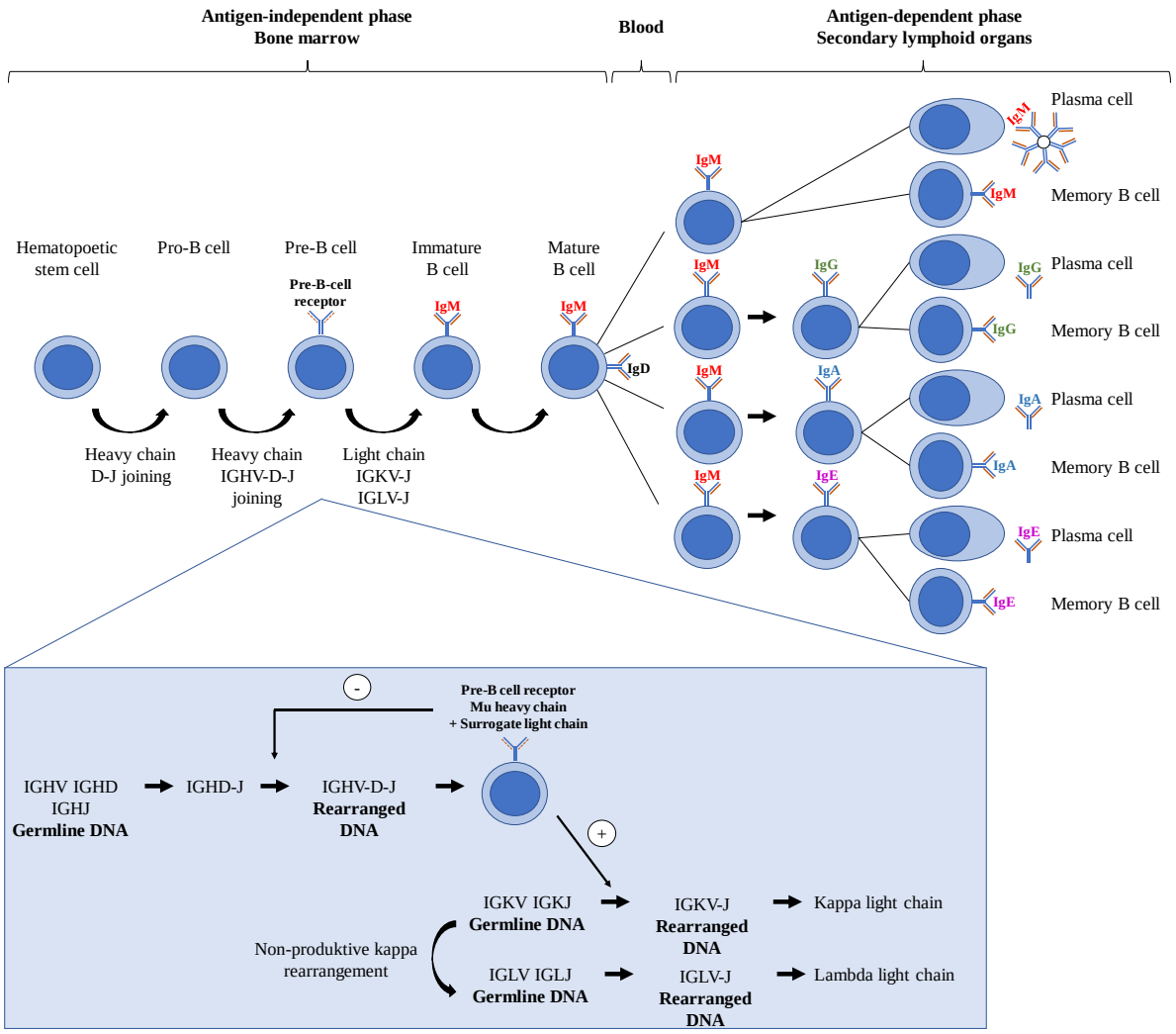


Figure 2. Schema of the B cell differentiation and the regulation of the immunoglobulin gene rearrangements in the pre-B cell. Adapted from (Lefranc and Lefranc 2020). V = variable region, D = diversity region, J = joining region, IGK = immunoglobulin kappa light chain, IGL = immunoglobulin lamba light chain, IGH = immunoglobulin hevay chain, The original article was published under the Creative Commons Attribution 4.0 International License (CC BY 4.0, https://creativecommons.org/licenses/by/4.0/).

1.1.3 Lambda Light Chain Locus

The lambda locus is located on the long arm of chromosome 22q11.2 and spans 1050 kb. It consists of 73 – 74 *IGLV* genes, of which 30 – 36 are potentially functional. Overall, the *IGLV* genes are sorted into 10 IGLV families. Additionally, seven to eleven *IGLJ* and *IGLC* genes, of which four are classified as functional, are located on this chromosome. Each *IGLC* gene is preceded by one corresponding *IGLJ* gene (**Table 2**) (Kawasaki et al. 1997; Lefranc and Lefranc 2020; Schroeder and Cavacini 2010).

Table 2. Potential Homo sapiens IGLV genomic repertoire. Adapted from (Lefranc and Lefranc 2020). IGLV = immunoglobulin lambda light chain variable region. The original article was published under the Creative Commons Attribution 4.0 International License (CC BY 4.0, https://creativecommons.org/licenses/by/4.0/).

IMGT IGLV family	IMGT gene name	Number of functional alleles	Number of pseudogene alleles	Gene order
IGLV1	IGLV1-36	1	-	40
	IGLV1-40	3	-	36
	IGLV1-44	1	-	31
	IGLV1-47	2	-	28
	IGLV1-51	2	-	24
IGKV2	IGLV2-8	3	-	77
	IGLV2-11	3	-	74
	IGLV2-14	4	-	71
	IGLV2-18	4	-	67
	IGLV2-23	3	-	61
IGLV3	IGLV3-1	1	-	84
	IGLV3-9	2	1	76
	IGLV3-10	2	-	75
	IGLV3-12	2	-	73
	IGLV3-16	1	-	69
	IGLV3-19	1	-	66
	IGLV3-21	3	-	64
	IGLV3-22	1	1	63
	IGLV3-25	3	-	59
	IGLV3-27	1	-	56
IGLV4	IGLV4-3	1	-	82
	IGLV4-60	3	-	13
	IGLV4-69	2	-	3
IGLV5	IGLV5-37	1	-	39
	IGLV5-39	2	-	37
	IGLV5-45	4	-	30
	IGLV5-52	1	-	23
IGLV6	IGLV6-57	2	-	17
IGLV7	IGLV7-43	1	-	32
	IGLV7-46	2	1	29
IGLV8	IGLV8-61	3	-	12
IGLV9	IGLV9-49	3	-	26
IGLV10	IGLV10-54	2	1	20

1.1.4 Diversity of Immunoglobulins

1.1.4.1 Combinatorial Diversity and Junctional Diversity

As already mentioned, a high degree of diversity is necessary to ensure the biological function of the Ig. Several mechanisms intertwine for this reason, of which one is the possibility to connect several different IGLVJ(D)C segments (combinatory diversity). This allows the pre-immune repertoire to already achieve a high degree of diversity. This diversity is further enhanced by the enormous variety provided by the *V(D)J* joining (junctional diversity). For example, each *D* gene has the opportunity to encode six different peptide fragments. These different fragments are created through splicing, translation into different reading frames, and

the possibility to rearrange via inversion or deletion (Schroeder and Cavacini 2010). During this rearrangement process, a hairpin structure between the 5' and 3' ends of the *V* and *J* or *D* segments is created. When this hairpin structure is opened dis-symmetrically, a 3' overhang is created and forms a palindromic extension – the P junction (Schroeder and Cavacini 2010). Moreover, the rearranged segments can also lose one or several nucleotides. In contrast, several nucleotides can also be added or replaced in the original germline sequence by the terminal deoxynucleotidyl transferase, which is template-independent and mostly involves 'G' nucleotides (Lefranc and Lefranc 2020; Schroeder and Cavacini 2010).

These two mechanisms alone allow a possible number of 10^{10} different options, for example, for the HC CDR3 region. These CDR3 regions all differ in length and structure (Schroeder and Cavacini 2010).

1.1.5.2 Somatic Hypermutations

Another mechanism for increasing the Ig diversity are somatic hypermutations which are introduced during B cell maturations in the secondary lymphoid organs (Lefranc and Lefranc 2020). This process takes place during the antigen-dependent stages of differentiation and concerns only the rearranged VJ and D segments. Somatic hypermutations in VJ(D) segments are about 10^6 times more frequent than spontaneous mutations in other cells (Lefranc and Lefranc 2020). The introduction of somatic hypermutations is initiated by the activation-induced cytidine deaminase, which induces the deamination from cytosine towards uracil. This uracil is then removed through the uracil deoxyribonucleic acid (DNA) glycosylase and further processed by the DNA base excision repair or DNA mismatch repair system. In the next step, the DNA lesions are repaired, which leads to a transition from cytosine towards thymidine and guanine towards adenine on the opposite strand (Lefranc and Lefranc 2020).

In total, a pre-immune antibody repertoire of 10^{16} different Ig is possible. This is achieved through the connection of the different gene segments, the combinatory effect of the LCs and HCs itself, and the somatic variations in the CDR3 region (Lefranc and Lefranc 2020; Schroeder and Cavacini 2010).

1.2 Monoclonal Gammopathies

The diseases classified as monoclonal gammopathies are all characterized by the monoclonal proliferation of at least one single plasma cell. These plasma cells secrete monoclonal proteins (M protein) that present immunologically and electrophoretically homogenous (Group 2003). Based on this protein, other common names are known for this group of diseases: paraproteinaemias, dysproteinaemias, or immunoglobulinopathies (Kyle and Rajkumar 2007).

In a review by the Mayo Clinic in 2006, several associated plasma cell disorders were listed (Kyle and Rajkumar 2007). Here, patients were diagnosed with multiple myeloma (MM) (16.5 %), AL amyloidosis (11.5 %), lymphoproliferative disorders (4 %), smoldering multiple myeloma (SMM) (3 %), Waldenström's macroglobulinemia (2 %) and solitary or extramedullary plasmacytoma (2 %) (others: 6 %) (Kyle and Rajkumar 2007). However, in more than half of the cases, a monoclonal gammopathy of undermined significance (MGUS) (55 %) was diagnosed.

For a MGUS diagnosis, besides other factors, a serum M protein concentration of >30 g/L, fewer than 10 % plasma cells in the bone marrow, no end organ damage, and no evidence of a related plasma cell proliferative disorder must be considered in the clinical context (Group 2003; Kyle and Rajkumar 2007). From a genetic point of view, primary events like an *IgH* translocation or hyperdiploidy are important. They can promote the transformation of a post-germinal center B cell towards increased monoclonal expansion/proliferation. This proliferation ultimately results in a MGUS diagnosis and thus forms the basis for the development of symptomatic MM (Pawlyn and Morgan 2017) (**Figure 3**).

In this development, the evolutionary intermediate stage of SMM was defined. Here, a malignant transformation can lead to the progression from MGUS towards SMM (Salomon-Perzyński et al. 2021). SMM is further defined by a higher clonal burden than MGUS and a variable risk of progression towards MM (Pawlyn and Morgan 2017; Rajkumar et al. 2015). In contrast, the annual risk from a progression from MGUS to symptomatic MM is approximately only 1 % (Kyle et al. 2018). The progression towards MM is characterized by an expansion of plasma cells outside the bone marrow microenvironment for example in the lymph nodes or soft tissue. The MM underlying biology and clinical presentation are addressed in more detail in chapter 1.4. In a final step, an evolution from MM towards plasma cell leukemia is possible; this stage is defined by peripheral blood involvement (Bhutani et al. 2020; Salomon-Perzyński et al. 2021).

The disease's progression and evolution from MGUS towards plasma cell leukemia is driven by secondary events, like secondary genetic events, epigenetic changes, subclonal competition, subclonal cooperation, or the remodeling of the tumor microenvironment (McGranahan and Swanton 2015; Salomon-Perzyński et al. 2021) (**Figure 3**). Taking these factors together, an increased fitness and adaptive potential of random subclone populations under selective pressure can lead to the evolution of the disease (Nowell 1976; Salomon-Perzyński et al. 2021).

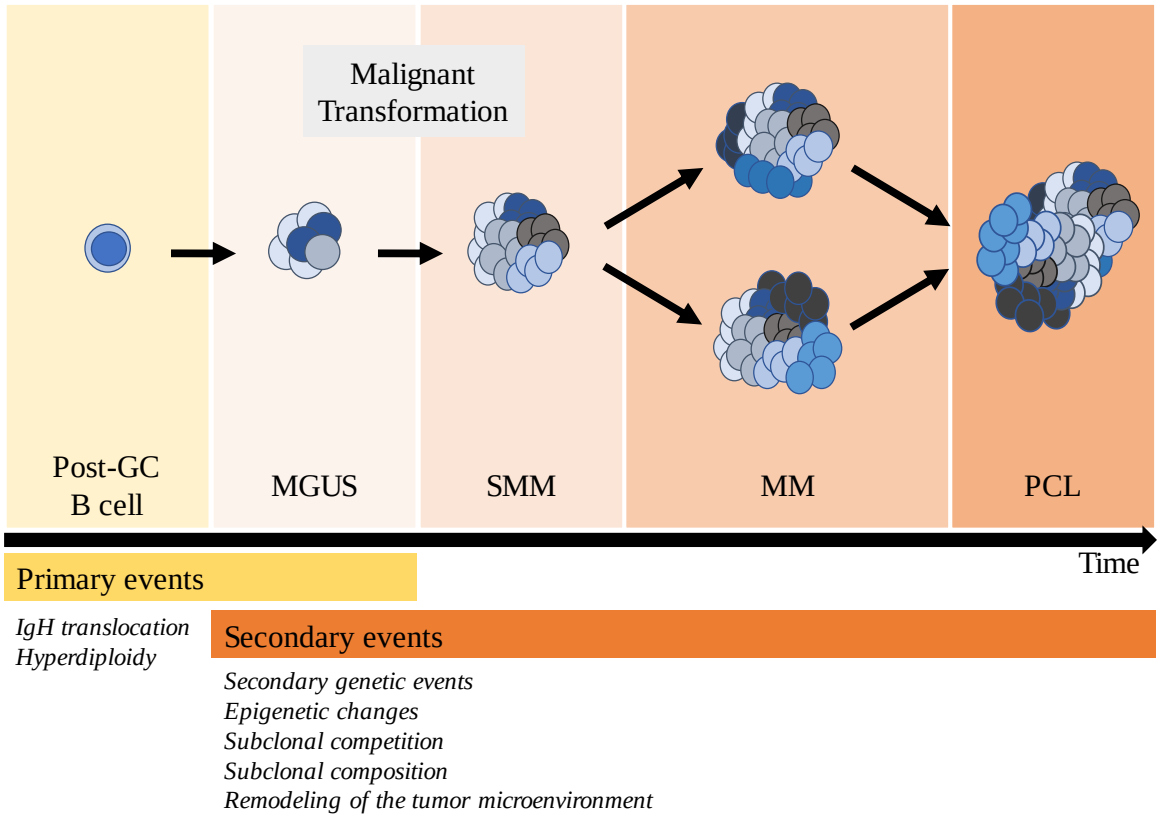


Figure 3. Evolution of plasma cell dyscrasias. Both primary and secondary genetic events are required for the malignant transformation of a post-GC B cell to MM. New genetic events accumulate in malignant plasmocytes over time affecting the fitness of subclones and leading to the further expansion of some and the extinction of others. Progression from MGUS/SMM to MM is driven by subclonal competition (and likely by its cooperation) and by simultaneous changes in the tumor microenvironment. (1) Branching clonal evolution – the dominant evolutionary pattern in the progression of SMM to MM, with some subclones disappearing (pink, light blue) and others appearing (red, purple, brown) over time; (2) static clonal progression – evolutionary pattern in the progression of SMM to MM, in which the tumor's subclonal architecture does not change significantly over time. Abbreviations: GC, germinal center; IgH, immunoglobulin heavy chain; MGUS, monoclonal gammopathy of undetermined significance; MM, multiple myeloma; SMM, smoldering multiple myeloma; PCL, plasma cell leukemia. Image adapted and legend retrieved from a publication by Salomon-Perzyński (Salomon-Perzyński et al. 2021), the original article was published under the Creative Commons Attribution 4.0 International License (CC BY 4.0, https://creativecommons.org/licenses/by/4.0/).

Of note, the development of light chain amyloidosis (AL) is possible at all dressed stages from MGUS towards plasma cell leukemia. A development from MGUS towards AL was observed in 12 % of the patients with a median disease interval of nine years (Kyle and Rajkumar 2007; Kyle et al. 2004a). Besides the fact that co-occurrence of AL and MM was detected in 10 % of the patients, a progression from AL towards MM is extremely rare (Rajkumar et al. 1998).

1.3 Multiple Myeloma

1.3.1 Biology of Multiple Myeloma

As mentioned before, MM is a malignant plasma cell disorder that evolves from the premalignant MGUS and SMM stages. Further, it is defined by an abnormal clonal plasma cell

infiltration in the bone marrow (Greipp et al. 2005; Kyle and Rajkumar 2004; Kyle and Rajkumar 2008). In an epidemiological context, it accounts for approximately 10 % of all hematologic cancers, and the annual age-adjusted incidence in the US is approximately 4.3 per 100000 persons (Kyle and Rajkumar 2008; Kyle et al. 2004b). Additionally, MM is more frequently diagnosed in Black than White persons and slightly more common in males. The overall median age at diagnosis is about 65 years (Firth 2019; Kyle and Rajkumar 2008).

As mentioned in chapter 1.2, translocations, especially of the *IgH* locus, as well as hypodiploidy are described as primary events and occur in almost 100 % of the cases (Walker et al. 2014) (**Table 3**). In this context, hyperdiploid MM is characterized by an overall increased number of 48 – 74 chromosomes, which is based on trisomies of chromosomes 3, 5, 7, 9, 11, 15, 19, and 21 (Chng et al. 2007). However, translocations and hyperdiploidy can be observed for MM as well as MGUS patients and are therefore not sufficient to explain disease progression (Fonseca et al. 2002; Walker et al. 2014).

One possible translocation for the HC locus, t(11;14), is described in the AL chapter 1.4 in more detail. In MM patients, three translocation partners for the HC locus can be frequently observed: a) 20 – 25 % cyclin D1, D3, and possibly D2 (*CCND1, CCND2 and CCND3*); b) 15 % *MMSET* and *FGFR-3*; c) 10 % *c-MAF* and *MAFB* (Chesi et al. 1998; Hanamura et al. 2001; Kuehl and Bergsagel 2002). Besides these primary translocation events, the secondary translocation of *c-MYC* can also frequently be detected in MM patients. This is of special interest due to the fact, that this translocating is associated with an enhanced proliferation (Kuehl and Bergsagel 2002).

Comparable to the results concerning acquired mutations, an increased number of specific genetic aberrations can also be observed during disease progression from MGUS towards MM (López-Corral et al. 2011). MM patients display more often a gain of 1q and a deletion of 13q or 17p as well as a higher amount of the translocations t(11;14) and t(4;14) than MGUS and SMM patients (López-Corral et al. 2011). In general, a gain of 1q, 3q, 9q, 11q, and 15q can be observed in more than 30 % of the MM patients. The most frequent chromosome loss concerns 13q (Kuehl and Bergsagel 2002). However, not only the presence but more specifically the number of affected plasma cells seem to have a prognostic impact and might also influence the disease progression (Avet-Loiseau et al. 2009; López-Corral et al. 2011).

The assumption that additional genetic events are necessary for expansion from MGUS towards the invasive states is further sustained by a higher number of acquired mutations in the context of disease progression (Walker et al. 2014). However, the sides and targeted molecular mechanism did not differ between the different disease stages and are therefore also not fully capable to explain disease progression (Walker et al. 2014).

Table 3. Primary molecular cytogenetic classification of multiple myeloma. IgH = Immunoglobulin heavy chain, MM = multiple myeloma. Retrieved and adapted from (Rajan and Rajkumar 2015). The original article was published under the Creative Commons Attribution 4.0 International License (CC BY 4.0, https://creativecommons.org/licenses/by/4.0/).

Subtype	Gene(s) chromosomes affected	Percentage of MM patients
Trisomic – recurrent trisomies involving odd-numbered chromosomes (except for chromosomes 1, 13, and 21)		42
IgH-translocated		30
t(11;14) (q13;q32)	*CCND1* (cyclin D1)	15
t(4;14) (p16;q32)	*FGFR-3* and *MMSET*	6
t(14;16) (q32;q23)	*C-MAF*	4
t(14;20) (q32;q11)	*MAFB*	<1
Other *IgH* translocations	*CCND3* (cyclin D3) in t(6;14) MM	5
Combined *IgH*-translocated/trisomic – the presence of trisomies and any one of the recurrent *IgH* translocations in the same patient		15
Isolated monosomy 14 – few cases may represent 14q32 translocations involving unknown partner chromosomes		4.5
Other cytogenetic abnormalities in the absence of *IgH* translocations or trisomy or monosomy 14		5.5
Normal/undetected		3

[a] Includes the t(6;14) (p21;q32) translocation and, rarely, other *IgH* translocations involving uncommon partner chromosomes.

Overall, it must be assumed that different genetic events in different pathways lead to the modification of the original non-malignant plasma cell in such a way that the typical MM characteristics are present (Morgan et al. 2012). These characteristics include the already addressed data regarding non-synonymous mutations and copy number abnormalities. In consequence, during the disease progression towards MM, an increase in genetic complexity can be assumed (López-Corral et al. 2011; Walker et al. 2014). This genetic complexity was analyzed using data from more than 800 newly diagnosed MM patients. Here, the natural aging process, defects in the DNA repair mechanism, and mutagenic activity of enzymes, including the activation-induced cytidine deaminase, were identified as important influencing factors for the mutagenesis in MM. (Hoang et al. 2019; Salomon-Perzyński et al. 2021).

Of note, the concept of intra-clonal heterogeneity is also important for MM. In this context, a definition based on different molecular subgroups which are characterized by specific chromosomal aberrations must be given (Manier et al. 2017; Morgan et al. 2012; Rasche et al. 2017). This is necessary since about 80 % of the MM patients display focal lesions (FL). FLs are characterized as an accumulation of malignant plasma cells in restricted regions of the bone marrow (Rasche et al. 2017; Walker et al. 2007). Through a genomic analysis of different FL within the same patients, it was possible to identify spatial heterogeneity. Therefore, the concept of ongoing clonal competition and adaption to the microenvironment can be assigned to MM (Rasche et al. 2017). This is supported by different findings, such as the observation of site-

specific driver aberrations or a high subclonal percentage of heterogenous mutations. Further, the size of the FL correlated with the number of unshared mutations and chromosomal aberrations (Rasche et al. 2017).

To go more into detail, in the already mentioned multi-region genomic analysis, Rasche et al. analyzed four MM cases with available computer tomography (CT)-guided sample sets. These sample sets were used to analyze the polygenetic relationship of clones from different regions (Rasche et al. 2017). In one of these cases, four samples from FLs at the left chest wall, the lumbar spine 1, the sacrum, and the iliac crest were analyzed. In this analysis, it was possible to identify different genomic profiles and unique driver mutations for each FL (**Figure 4**). Since all analyzed positions share a translocation of *MYC* and an amplification of 1q, a shared ancestor can be assumed. This shared ancestor evolved further in the process of multi-focal progression (Rasche et al. 2017). However, this was not noted for all patients, for example, in another case study a common high-risk ancestor with a distribution in a metastatic fashion was observed (Rasche et al. 2017).

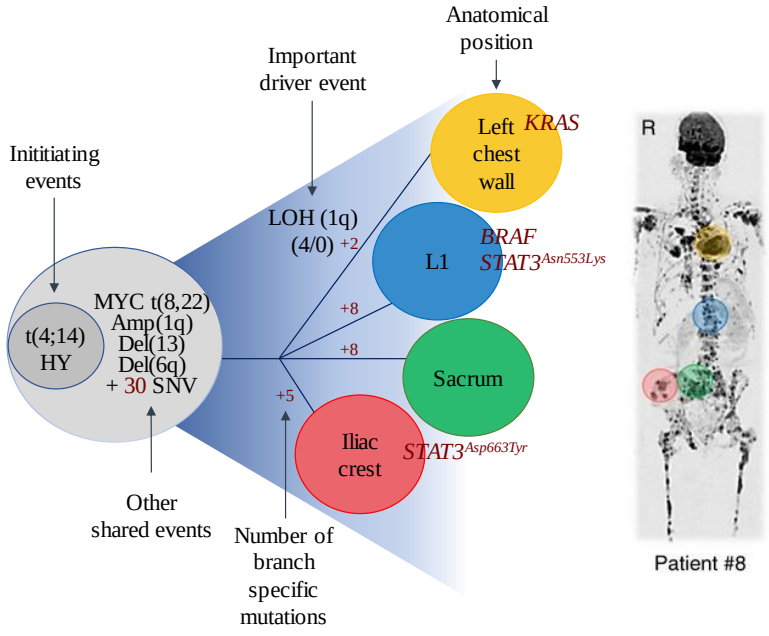


Figure 4. Multi-regional evolution in a multiple myeloma patient. Patient with available multiple CT-guided samples was used to analyze the phylogenetic relationship of clones from different regions. The location of samples is marked in the medical images in the right panel using the color code that was assigned to clones (left panel). The letter R indicates the right side of the body. Focal lesions displayed different genomic profiles with each of them containing unique driver mutations. A mutation of the *BRAF* gene affects the serine/threonine-kinase B-Raf, which is known as a proto-oncogene. KRAS is a monomeric G-protein and is known to play an important role in the proliferation of tumor cells. A mutation in *STAT3* was associated with spatial heterogeneity in the work of Rasche et al. and is defined as a promising candidate as a novel MM driver (Rasche et al. 2017). HY = hyperdiploid, Amp = amplification, Del = deletion, SNV = single nucleotide variant, LOH = loss of heterozygosity, L1 = lumbar spine 1. Figure and legend retrieved and adapted from (Rasche et al. 2017). The original article was published under the creative common CC BY 4.0 license (https://creativecommons.org/licenses/by/4.0/, 05.10.2022).

1.3.2 Clinical Presentation and Diagnosis of Multiple Myeloma

The International Myeloma Working group defined several criteria for a MM diagnosis. In more detail, in addition to more than 10 % of plasma cells in the bone marrow (**Figure 5**, A) or biopsy-proven plasmacytoma, four other myeloma-defining events (MDE) must be present (Rajkumar 2022; Rajkumar et al. 2014). These MDEs are defined by the established CRAB criteria and include hypercalcemia, renal failure, anemia, or lytic bone lesions (**Figure 5**, B). In more detail: serum calcium >0.25 mmol/L needs to be higher than the upper limit of normal, creatinine clearance <40 mL per minute, hemoglobin value of >2 g/dL below the lower limit of normal and at least one lytic lesion on skeletal radiography, CT, or positron emission tomography-CT (Rajkumar et al. 2014). These criteria were updated and refined in 2022 and established as the current standard SLiM-CRAB criteria for treatment indication of SMM. These SLiM-CRAB criteria are defined by ≥60 % clonal bone marrow plasma cells, involved serum free light chain ratio ≥100 mg/L, and more than one FL (≥5 mm in size) on magnetic resonance imaging (Rajkumar 2022). It was possible to associate these additional biomarkers each with an approximately 80 % risk of progression to symptomatic end-organ damage (Rajkumar 2022).

The clinical presentation typically includes bone pain and pathological fractures, recurring infections, and abnormal bleeding (Firth 2019). Rare clinical manifestations may include ischemia, heart failure, neurological impairments, and amyloid diseases (Firth 2019).

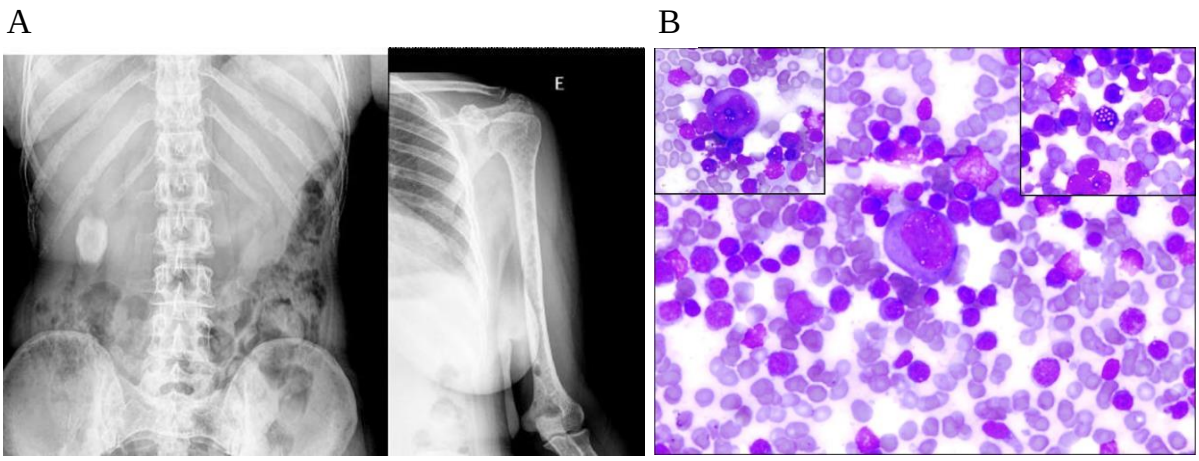


Figure 5. Characteristic osteolysis and plasma cell increase in multiple myeloma patients.
(A) Skeleton X-ray of a multiple myeloma patient showing multiple diffuse osteolytic lesions. Figure and legend were adapted and retrieved from (Caldas et al. 2011). The original article was published under the creative common CC BY-NC 4.0 license (https://creativecommons.org/licenses/by-nc/4.0/, 20.12.2022) (B) Bone marrow aspirate of a multiple myeloma patient showing numerous plasma cells (WG; ×400). Figure and legend were adapted and retrieved from (Khurana et al. 2016). The original article was published under the Creative Commons Attribution 4.0 International License (CC BY 4.0, https://creativecommons.org/licenses/by/4.0/).

In addition to the SLiM-CRAB criteria, the patients should be tested for the presence of a M protein using serum protein electrophoresis and serum immunofixation. Based on these assays, a diagnosis will be established successfully in about 98 % of the MM patients (Chawla et al. 2015; Katzmann et al. 2006; Kyle et al. 2003; Rajkumar 2022). In general, if it is possible to detect more than 5 % of circulating plasma cells in the peripheral blood smear, plasma cell leukemia should be considered (Fernández de Larrea et al. 2021). In addition to the tests necessary for clinical characterization and diagnosis, a fluorescence *in situ* hybridization (FISH) analysis should also be performed for risk stratification. In this context, patients with a deletion of 17p or 1p as well as a t(14;16) or t(14;20) are considered to be at high risk for a lower overall survival (Rajan and Rajkumar 2015).

The estimated prognosis not only considers the CA but also more generally the tumor burden, indicated by staging systems, and the therapy response (Rajkumar 2022; Russell and Rajkumar 2011; Vu et al. 2015). The revised international staging system combines aspects of the tumor burden as well as CA (Palumbo et al. 2015). Stage I is defined by a serum albumin >3.5 gm/dL, a serum beta-2-microglobulin <3.5 mg/L, no high-risk cytogenetics, and a normal serum lactate dehydrogenase level. Stage III applies to patients with a serum beta-2-microglobulin >5.5 mg/L and high-risk cytogenetics like t(4;14), t(14;16), del(17p) or elevated serum lactate dehydrogenase level (Palumbo et al. 2015). Stage II patients are classified as all patients which are not suitable for stage I or III (Palumbo et al. 2015). It was shown that a high level of lactate dehydrogenase can also be associated with lower overall survival (Terpos et al. 2010).

1.4 AL Amyloidosis

1.4.1 Forms of Amyloidosis

Amyloidosis is characterized as a group of rare diseases, which result from the extracellular deposition of amyloid fibrils. These fibrils are defined by their β-sheet confirmation and can be visualized by a congo-red staining and resulting green birefringence (Jin et al. 2003; Westermark 2005).

Until now, more than 30 proteins have been identified to form fibrils in the human body but it is believed that there are even more proteins with amyloidogenic potential (Sipe et al. 2014; Wechalekar et al. 2016). Based on the underlying precursor protein different forms of amyloidosis can be defined (**Table 4**). Acquired systemic AL and hereditary amyloidosis are associated with the presence of an abnormal protein in the serum. The β2-microglobulin (Aβ2m) dialysis-related amyloidosis or reactive systemic amyloidosis are defined by the

excessive occurrence of a normal protein. Wild-type transthyretin (ATTRwt) amyloidosis and atrial natriuretic peptide amyloidosis are forms that are associated with the aging process (Merlini et al. 2018; Wechalekar et al. 2016).

In Western countries, the most frequent form of amyloidosis is systemic AL, next to ATTRwt (Obi et al. 2022; Wechalekar et al. 2016). In general, AL can be defined as an acquired form of amyloidosis (Wechalekar et al. 2016) – only one case study found a hereditary mutation in the CL of a kappa LC (Benson et al. 2015). Besides systemic AL, a localized form is also known. In this case, the LCs deposit at one single anatomic site and form amyloid lesions. However, local AL account only for about 10 % of all amyloidosis cases but no precise population data is available (Basset et al. 2020; Hamidi Asl et al. 1999). In contrast, several studies have been performed to investigate the epidemiology of systemic AL (Merlini et al. 2018). When comparing the results, a range of an annual incidence of 3 to 12.7 per million persons with a median of 10.7 can be calculated (Duhamel et al. 2017; Hemminki et al. 2012; Kyle et al. 1992; Kyle et al. 2004b; Pinney et al. 2013; Quock et al. 2018). Another important epidemiological aspect is that more men than women are affected (Kyle et al. 1992). More precisely, in an analysis of 701 patients, 64 % were male (Kourelis et al. 2017).

Table 4. Most common forms of systemic amyloidosis. Reprinted and adapted by permission from Copyright Clearance Center: Springer Nature, Nature Reviews Disease Oligonucleotides, Systemic immunoglobulin light chain amyloidosis, Merlini et al. (2018) (Merlini et al. 2018) (License number 5398791052309). Legend retrieved from (Merlini et al. 2018).

Designation[a]	Parent protein	Systemic and/or localized	Acquired or hereditary	Organs involved
AL	Immunoglobulin light chain[b]	Systemic or localized	Acquired (hereditary[c])	Heart, kidney, liver, soft tissues, peripheral nervous system (including the autonomic nervous system) and gastrointestinal tract
ATTR	Transthyretin	Systemic	Hereditary	Peripheral nervous system (including the autonomic nervous system), heart, eye, kidney and leptomeninges
		Systemic	Acquired	Heart and ligaments
AA	Serum amyloid A protein	Systemic	Acquired	Predominantly kidney, but may involve liver, gastrointestinal tract and occasionally heart, thyroid and autonomic nervous system
ALECT2	Leukocyte chemotactic factor 2	Systemic	Acquired	Kidney, liver, spleen, adrenals and lungs

Designation[a]	Parent protein	Systemic and/or localized	Acquired or hereditary	Organs involved
AApoAI	Apolipoprotein AI	Systemic	Hereditary	Heart, liver, kidney, peripheral nervous system, testis, larynx and skin
AFib	Fibrinogen α chain	Systemic	Hereditary	Kidney, primarily, with obliterative glomerular involvement
$A\beta_2m$	β_2-microglobulin, wild type	Systemic	Acquired (haemodialysis related)	Musculoskeletal system
	β_2-microglobulin	Systemic	Hereditary	Autonomic nervous system

[a]The amyloid fibril protein is designated protein A and is followed by a suffix that is an abbreviated form of the precursor protein name. For example, when amyloid (A) fibrils are derived from immunoglobulin light (L) chains, the amyloid fibril protein is AL. [b] Rare cases of amyloidosis formed by immunoglobulin heavy chains (AH) and by heavy and light chains (AHL) have been reported. [c] One family with a mutation in the constant region of the κ light chain, with cysteine replacing serine at amino acid residue 131, has been reported (Benson et al. 2015).

In general, AL is caused by at least one malignant plasma cell clone which infiltrates the bone marrow. This infiltration is less than 10 % in half of the AL patients. Therefore, the AL clone is relatively small compared to the one in symptomatic MM (Merlini 2017; Milani et al. 2018). But, only the AL clone produces an abnormal number of free LCs which misfold, form prefibrillar aggregates, and finally form fibrils in the extracellular space. These fibrils cause cellular stress and disturb the normal tissue structure and function which can cause lethal consequences (Merlini 2017). In addition to the differences concerning the clone size and the LC behavior, also the kappa/lambda ratio is shifted in AL. For the B cell repertoire of healthy individuals as well as for the M protein of MM patients, more kappa LC (2:1; kappa/lambda) can be detected. In contrast, AL patients display more frequently a lambda LC (1:3; kappa/lambda) (Kyle and Gertz 1995).

1.4.2 Clinical Presentation and Diagnosis of AL Amyloidosis

In most cases, the AL diagnosis is delayed, and drastic organ damage has already taken place. In addition, more than 69 % of the patients display more than one organ involvement at diagnosis (Obici et al. 2005). Unfortunately, it has been reported that survival significantly decreased the more organs were affected (Muchtar et al. 2019). One possible factor for a delayed diagnosis might be that AL-related symptoms are commonly found in elderly people (Milani et al. 2018), which is of importance since the average age at first diagnosis is 65 years (Schönland et al. 2012). Besides the number of affected organs, the patient's age was also linked to a poorer prognosis. A study conducted by the Mayo Clinic showed that patients ≥65 years had lower survival than younger patients (Muchtar et al. 2019).

Overall, a delayed diagnosis led to higher mortality rates, especially by cardiac involvement or a progression to end-stage renal disease in the first month after diagnosis (Merlini et al. 2018; Muchtar et al. 2017b; Palladini et al. 2014). One study showed that in 40 % of the patients, the diagnosis was confirmed >1 year after symptoms occurred (Lousada et al. 2015). This can especially be observed for MGUS patients – therefore, screening for AL should already be performed at this stage (Al Hamed et al. 2021; Kourelis et al. 2014; Merlini et al. 2018).

Several symptoms can point to a potential AL diagnosis. This includes heart failure with preserved ejection fraction, nephrotic range proteinuria, a mixed axonal demyelinating peripheral neuropathy with autonomic features, or hepatomegaly without imaging abnormalities (Merlini 2017; Merlini et al. 2018; Wechalekar et al. 2016). To secure a diagnosis, it is suggested to perform immunofixation electrophoresis of the serum and urine as well as an immunoglobulin-free LC assay. Further, a tissue biopsy and a histopathological analysis should be conducted (Merlini et al. 2018). However, the underlying biopsy must not be performed on the affected organ. It was shown that a less invasive method like the aspiration of subcutaneous fat or a bone marrow biopsy can lead to a diagnosis in 50 – 80 % of the cases (Merlini et al. 2018; Muchtar et al. 2017a; Quarta et al. 2017). After the detection of amyloid deposits in the biopsies, an accurate typing of the precursor protein is necessary. In this context, a mass spectrometry-based analysis is considered the most successful approach with a reported sensitivity of 88 % and specificity of 96 % (Merlini et al. 2018; Vrana et al. 2014). Besides the already addressed factors, there are also biomarkers available that can detect heart or renal failure with 100 % sensitivity (Merlini et al. 2018). An increased level of the N-terminal provisional brain natriuretic peptide (NT-proBNP) in serum indicates a cardiac involvement (Palladini et al. 2003; Wechalekar et al. 2011), whereas for example, albuminuria can be used for determining renal involvement (Palladini et al. 2014).

AL can affect all organs, except the brain, but in most of the cases, a distinct organ tropism can be observed (Merlini et al. 2018). This includes heart involvement in more than 80 % and a kidney involvement in more than 65 %. Also, the involvement of soft tissue can be observed in over 15 %, followed in descending order by liver involvement, involvement of the peripheral nervous system, and the autonomic nervous system (Merlini et al. 2018). Based on these organ involvements, different symptoms can be noted (**Figure 6**).

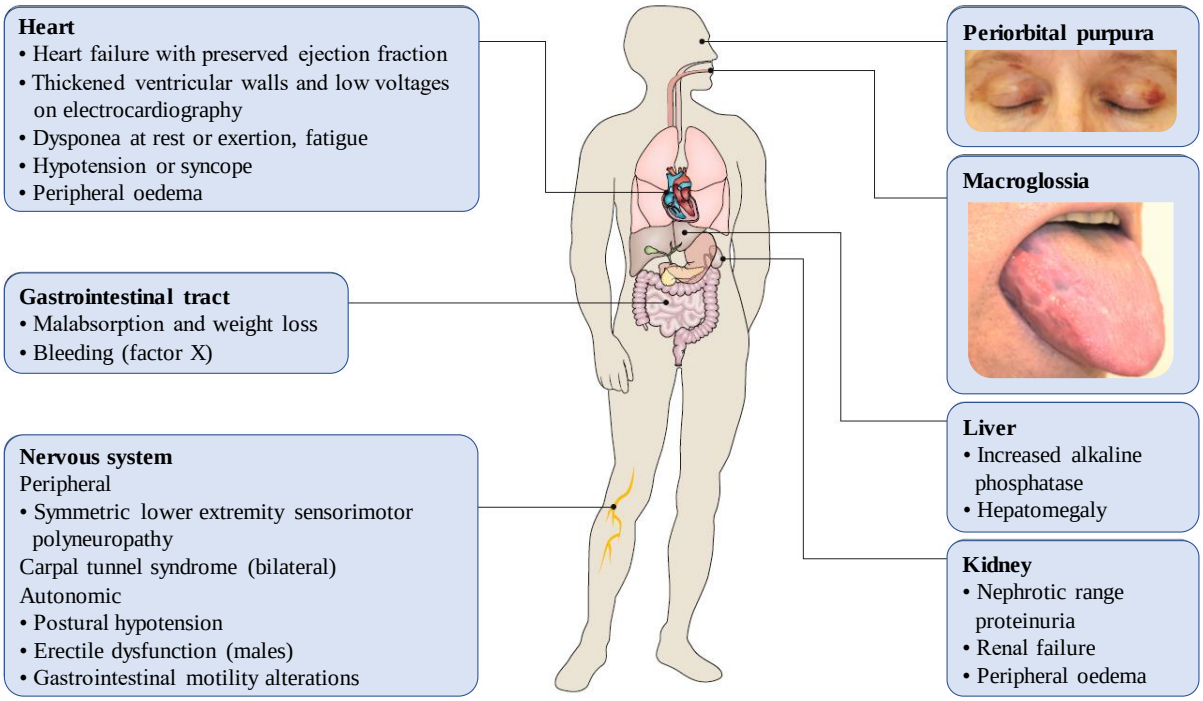


Figure 6. Organ involvement in systemic AL amyloidosis. The symptoms of monoclonal immunoglobulin light chain (AL) amyloidosis are variable and mimic symptoms observed in common conditions of elderly individuals, such as heart failure (fatigue) and diabetes mellitus (proteinuria and peripheral neuropathy), therefore, contributing to late diagnosis. The presence of heart failure with preserved ejection fraction and thickened ventricular walls with low voltages identified using electrocardiography should raise the suspicion of cardiac amyloidosis. Kidney involvement is characterized by proteinuria and progressive renal failure and manifests as peripheral oedema. The involvement of the gastrointestinal tract results in malabsorption and weight loss that can be prominent in some patients, whereas involvement of the autonomic nervous system can cause invalidating postural hypotension. The presence of prototypic signs such as macroglossia (enlargement of the tongue) and periorbital purpura can immediately lead to the right diagnosis. However, such signs are uncommon and, more importantly , appear late in the course of the disease, frequently appearing when the organ damage caused by amyloid is already irreversible. Reprinted and adapted by permission from Copyright Clearance Center: Springer Nature, Nature Reviews Disease Oligonucleotides, Systemic immunoglobulin light chain amyloidosis, Merlini et al. (2018) (Merlini et al. 2018) (License number 5398791052309). Legend retrieved from (Merlini et al. 2018).

As already addressed, the organ involvement influences the survival of AL patients. Patients with irreversible heart damage, due to a late diagnosis, have a median survival of three to six months (Wechalekar et al. 2013). In contrast, patients without heart failure can survive several years even when not responding to first-line therapy (Merlini et al. 2018). Furthermore, the early diagnosis of patients with renal involvement can prevent the progression towards end-stage kidney disease and dialysis. In contrast, late detection of renal disease is associated with an increased risk of progression (Palladini et al. 2014).

In the context of organ involvement, several staging systems have been established (**Table 5**). The most commonly used cardiac staging system in study design and clinical trials was established by the Mayo Clinic and European investigators (Kumar et al. 2012). It is based on the levels of NT-proBNP, cardiac troponin, and the difference between disease-associated and

uninvolved circulating free light chains (dFLC) (Dispenzieri et al. 2004; Kristen et al. 2010; Palladini et al. 2010; Wechalekar et al. 2013). Additionally, a renal staging system predicting the progression towards dialysis has been established with two biomarkers: eGFR and proteinuria. Applying this system, it was not possible to demonstrate a linkage between the staging and the survival of the patients. However, an impact on kidney survival and quality of life was noted, which also influence the efficiency of the treatment (Kastritis et al. 2017; Merlini et al. 2018; Palladini et al. 2014).

In addition to these staging systems, several other factors can affect patient survival. For example, it was shown that patients with a bone marrow plasma cell infiltration of >10 % have poorer survival rates (Kourelis et al. 2013). Another study found that patients with a dFLC <50 mg/L have a significantly better outcome, independent of the cardiac stage (Dittrich et al. 2017; Milani et al. 2017; Sidana et al. 2018).

Table 5. Overview of the most commonly used staging systems in AL amyloidosis. The staging parameters were extracted from the respective publications. (Kumar et al. 2012; Palladini et al. 2014; Wechalekar et al. 2013). dFLC = difference between disease-associated und uninvolved circulation free light chains, cTNT = cardiac troponin T, eGFR = estimated glomerular filtration rate, NT-proBNP = N-terminal provisional brain natriuretic peptide, SBP = systolic blood pressure.

Publication	Alias	Organ	Parameters/staging
Kumar et al. 2012 (J Clin Oncol.)	Mayo	Heart	1 point for each criterion (stage I – IV) o dFLC ≥18 mg/dL o cTNT ≥0.025 ng/mL o NT-proBNP ≥1800 pg/mL
Wechalekar et al. 2013 (Blood)	Euro	Heart	Stratification of advanced cardiac disease with poor prognosis (stage III Mayo) o IIIb: NT-proBNP >8500 ng/L or SBP <100 mm Hg
Palladini et al. 2014 (Blood)	Renal	Kidney	o I: proteinuria <5 g/d and eGFR >50 mL/min/1.73 m^2 o II: proteinuria >5 g/d or eGFR <50 mL/min/1.73 m^2 o III: proteinuria >5 g/d and eGFR <50 mL/min/1.73 m^2

Besides these clinical parameters, chromosomal aberrations (CA) also play an important role in the characterization and risk assessment of AL patients. For this purpose, the FISH analysis has been established as routine clinical testing. Based on this analysis, it was demonstrated that AL patients without underlying MM show a CA in 95 % of the cases (Bochtler et al. 2008). Here, the most prominent difference between AL and MM was found concerning t(11;14). This CA was detected in 61 % of the AL patients and only 15 % of MM, MGUS, or SMM patients (Bochtler et al. 2018).

In general, t(11;14) describes the translocation of the HC locus 14q32 with chromosome 11. Moreover, t(11;14) was found to be associated with the lack of intact Ig in the immunofixation in 82 % of the AL cases (Bochtler et al. 2008). Besides the lack of a HC, t(11;14) was also

found to be associated with AL patients with a lower level of plasma cell infiltration (Bochtler et al. 2018). Overall, it was postulated that the lack of a clonal HC binding partner might be part of the pathologic mechanism of AL (Bochtler et al. 2008). Other HC translocations like t(4;14) and t(14;16) were rarely found in AL but, for example, t(4;14) was detected in 26 % of MM patients (AL = 3 %) (Bochtler et al. 2008).

1.4.3 Pathological Effect of Light Chains and Fibrils

The distinct pathological mechanisms in AL which ultimately lead to impaired organ function were and are a major research focus but are not fully understood yet. Further, it is suggested that a difference between soluble proteins and amyloid aggregates must be made. An analysis in AC16 cardiomyocytes showed that LC amyloid fibrils inhibit cell growth and division. This inhibition was not detected for LC soluble proteins but it was noted that they cause cell death dysfunctions and apoptosis (Marin-Argany et al. 2016). Also, a difference between fresh and aged fibrils was found. Aged fibrils displayed an increase in cytotoxicity, which is potentially based on the fact that fibrils fragment into smaller clusters over time. It was stated, that these smaller clusters might be easier to engulf in the cardiomyocytes (Marin-Argany et al. 2016). Another aspect in the context of cytotoxicity is that the cell-cell contact is disturbed through the attachment of fibrils to the plasma membrane, which leads to a decrease in cell viability (Marin-Argany et al. 2016).

Over the years, several groups have established different animal model systems like *Caenorhabditis elegans* (Diomede et al. 2014), zebrafish (Mishra et al. 2013) or mice (Ayala et al. 2021) to investigate the effect of the disease. Researchers have especially tested the effect of cardiotropic LCs due to the devastating effect on patients with heart involvement (Marin-Argany et al. 2016; Merlini 2017). Here, it was noted that in comparison to LC from MM patients, the amyloidogenic LCs displayed a pathological effect. In more detail, the treated worms showed a pharyngeal dysfunction and a reduced life span. This effect was traced back to structural mitochondrial damage (Diomede et al. 2014; Merlini 2017). Similar damage was observed in amyloid-affected hearts from humans (Merlini 2017). It was also shown that amyloid LCs associated with heart involvement induced the p38 mitogen-activated protein kinase signaling (Shi et al. 2010). Since this pathway also mediate the type B natriuretic peptide (BNP) transcription, a linkage between the increased value of BNP/NT-proBNP and the cardiac pathology of the LC can be made (Merlini 2017).

In addition to increased NT-proBNP values, an enlarged septal thickness can also be detected in AL amyloidosis patients with cardiac involvement. This increased stiffness and enlargement

of the heart can be attributed to the amount of deposited fibrils and lead to an overall reduced functionality (**Figure 7**).

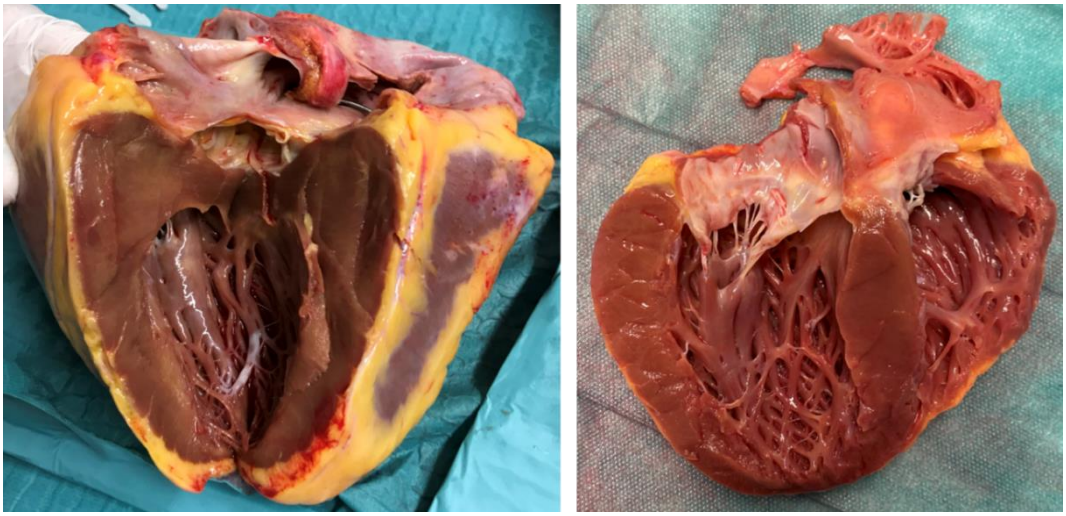


Figure 7. Photos of two heart transplants from AL amyloidosis patients. The heart transplants were donated for research by two patients treated at the amyloidosis center at the Heidelberg University Hospital. The photos were kindly provided by Prof. Dr. Ute Hegenbart (University Hospital Heidelberg).

1.4.4 Pathway of Fibril Formation

It can be assumed that the amyloidogenic protein exists in two stable forms: the native status and the fibril status which are separated by an energy barrier (Absmeier et al. 2022). Kinetically the process of fibril formation can be divided into three phases (**Figure 8**). The first phase, lag phase, is thermodynamically unfavorable and is characterized by completely or partially disordered monomers that convert into nuclei. After the lag phase, an exponential growth phase can be observed, where the nuclei grow through the addition of monomers into fibrils. This formation and extension of fibrils is then followed by an equilibrium phase, which is also called plateau or saturation phase (Chiti and Dobson 2017).

In this process, the native state needs to unfold and adapt to the energetically more favorable amyloid stage (Hartl and Hayer-Hartl 2009). How this stage transition takes place is not fully understood yet. It is assumed that an additional trigger combined with protein unfolding causes the supersaturation barrier to be disrupted. Through this disruption, the formation of amyloid fibril structures can occur (Absmeier et al. 2022; Adachi et al. 2015).

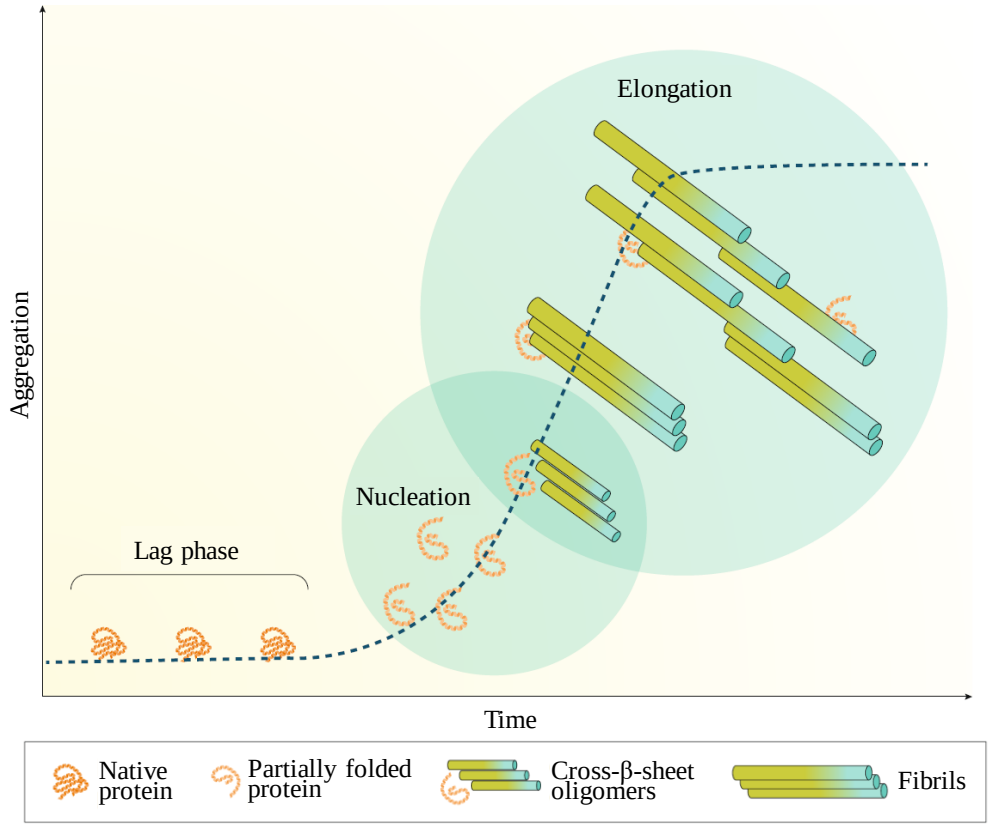


Figure 8. Kinetics of fibril formation in vitro. A specific local concentration of partially folded proteins is necessary for the formation of a critical fibrillar nucleus. The critical concentration for nucleation varies depending on the stability of the light chains. During the lag phase, the conditions do not favour protein aggregation, and fibrils are not formed. However, after the formation of the fibrillar nuclei, protein aggregation into cross-β-sheet oligomers occurs, leading to fibril formation and elongation. The concentration of partially folded proteins necessary for fibril elongation is substantially lower than the concentration required for forming the first nuclei. Reprinted and adapted by permission from Copyright Clearance Center: Springer Nature, Nature Reviews Disease Oligonucleotides, Systemic immunoglobulin light chain amyloidosis, Merlini et al. (2018) (Merlini et al. 2018) (License number 5398791052309). Legend retrieved from (Merlini et al. 2018). The original figure was created by V. Bellotti (University of Pavia, Italy) and the usage within this work was also permitted.

Additionally, several factors can influence this transition from LCs to amyloid fibrils, such as the concentration of the precursor protein, the pH, proteolysis, extracellular chaperones, or lipids (Gottwald and Röcken 2021; Merlini 2017). Also, several non-amyloid factors are known to be present in the amyloid deposits, like the serum amyloid P component. This factor binds to amyloid fibrils and may protect them from recognition by the immune system (Gottwald and Röcken 2021; Pepys 2018). Overall, in a study conducted by Kourelis et al., as much as 161 proteins were detected which seem to be amyloid-associated in AL deposits from patients with heart involvement (Kourelis et al. 2020).

Another aspect that needs to be considered in the process of fibril formation is proteolytic activity (Morgan et al. 2017). This is based on the suggestion that amyloidogenic LC dimers are kinetically unstable and that through endoproteolysis amyloidogenic fragments might be generated (Morgan and Kelly 2016). This is of importance as it was shown that only the variable

domains form fibrils, whereas no amyloidogenity was noted for full-length LCs (Morgan and Kelly 2016). Full-length LCs from patients with amyloidosis unfold more easily than full-length LCs from healthy patients. In consequence, they are assumed to be more prone to proteolytic cleavage (Morgan and Kelly 2016). In this context, several regions seem to be more susceptible to proteolysis than others, which might also be influenced by mutations (Absmeier et al. 2022; Morgan and Kelly 2016).

Kind of the last final step in the process of fibril formation is to bypass the quality control of the endoplasmic reticulum. This control system in the cell monitors proteins to ensure correct folding and assembly and can lead to the degradation of deviant chains. So, only assembled antibodies should be secreted in a healthy individual. (Ellgaard and Helenius 2003; Lee et al. 1999). Nevertheless, also healthy individuals produce LCs in excess of HCs. So, although a small number of free LCs occur among healthy people, it is not clear, how the multitude of LCs in AL patients can bypass this system (Absmeier et al. 2022; Waldmann et al. 1972).

1.4.5 Fibril Structure

Amyloid fibrils can be defined as a linear self-assembly state of polypeptide chains (Chiti and Dobson 2006). Further, they contain spines consisting of many-stranded β-sheets (Eisenberg and Jucker 2012). More precisely, amyloid fibrils are biophysical based on a cross-β fiber diffraction pattern (Eisenberg and Jucker 2012).

In 2019, it was possible to resolve the first two cryo-EM structures of an IGLV1-44 (FOR006) (Radamaker et al. 2019) and IGLV6-57 (AL55) AL fibril (Swuec et al. 2019). In the following years, two additional fibril structures of an IGLV3-19 (FOR005) (Radamaker et al. 2021a) and an IGLV1-51 (FOR001) (Radamaker et al. 2021b) AL fibril were resolved (**Figure 9**). These analyses showed that about 77 % of the residues form the ordered part of the fibrils. These ordered parts also contained 79 % of the detected mutations (Radamaker et al. 2021b). However, a central disordered segment was detected in three out of four fibril structures. This led to the assumption that the mutations are not preferentially located in the fibril core (Radamaker et al. 2021b). In addition, the stable fibril cores of these four cases showed a higher aggregation score than the disordered segments (Radamaker et al. 2021b).

Besides structural aspects, the four fibrils share a mutation at the third position of the CDR2 region (Radamaker et al. 2021b). Besides mutations, post-transcriptional modifications were investigated as well. In this context, FOR001 stood out with three post-transcriptional mutations: a disulfide bond, a pyroglutamylation, and a N-glycosylation (Radamaker et al. 2021b). Even though it is not known how glycosylation can attribute to the fibril formation,

they are overrepresented in AL fibrils – especially for kappa LCs (Dispenzieri et al. 2020; Kumar et al. 2019; Radamaker et al. 2021b).

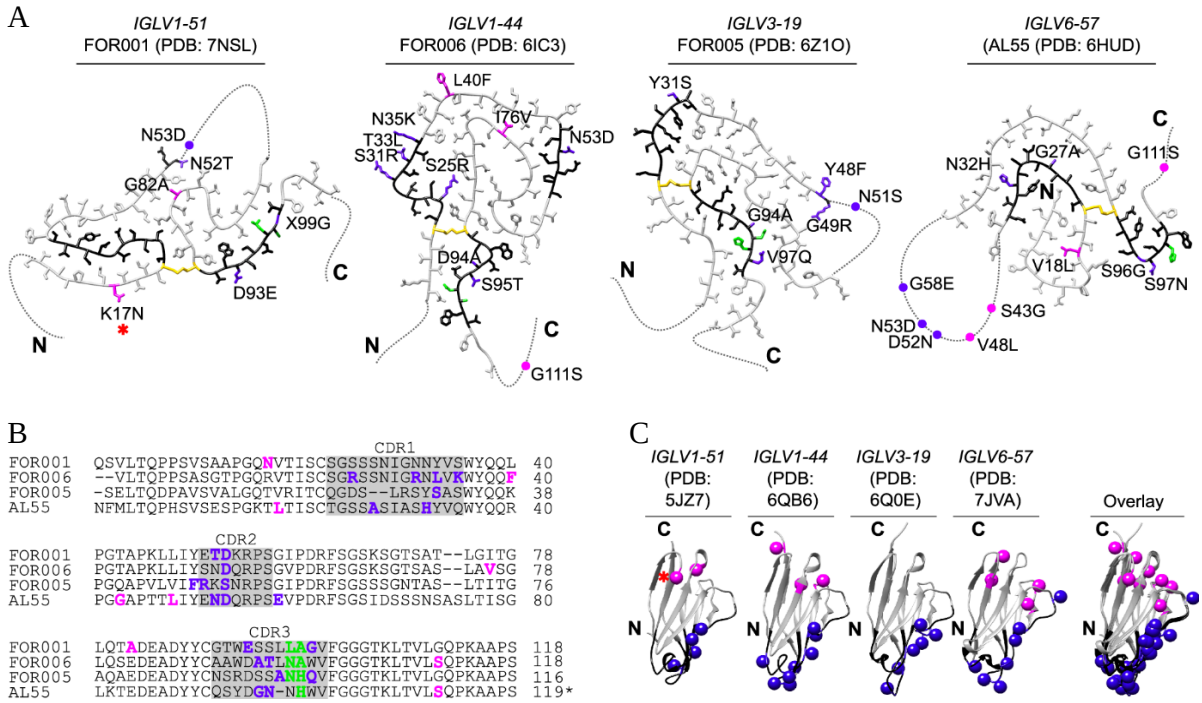


Figure 9. Location of the mutations in known AL amyloid fibrils and natively folded VL domains. (A) Location of the mutations in known AL amyloid fibril structures. The fibrils are derived from the GL segments IGLV1-51*02 λ1 (FOR001, PDB 7NSL) (Radamaker et al. 2021b), IGLV1-44*01 λ1 (FOR006, PDB 6IC3 10.2210/pdb6IC3/pdb) (Radamaker et al. 2019), IGLV3-19*01 λ3 (FOR005, PDB 6Z1O 10.2210/pdb6Z1O/pdb) (Radamaker et al. 2021a), and IGLV6-57*02 λ6 (AL55, PDB 6HUD 10.2210/pdb6HUD/pdb) (Swuec et al. 2019). Disordered parts of the fibril proteins are indicated by dotted lines. Black: CDRs; yellow: Cys; Indigo: mutations in the CDRs and one residue before or after a CDR; magenta: mutations in framework regions; green: residues in the junctional region at the V/J interface. Red star: location of the glycosylation. (B) Sequence alignment of the four fibril proteins. CDRs are marked with gray boxes. Color coding as in (A). (C) Location of mutations in the corresponding, natively folded LC VL domains that are based on the GL segments IGLV1-51*01 (PDB 5JZ7 10.2210/pdb5JZ7/pdb), IGLV1-44*01 (PDB 6QB6 10.2210/pdb6QB6/pdb) IGLV3-19*01 (PDB 6Q0E 10.2210/pdb6Q0E/pdb) and IGLV6-57*02 (PDB 7JVA 10.2210/pdb7JVA/pdb) CDRs are marked in black. Color coding as in (a). This image and legend were retrieved from a publication by Radamaker et al. (Radamaker et al. 2021b). The original article was published under the Creative Commons Attribution 4.0 International License (CC BY 4.0, https://creativecommons. org/licenses/ by/4.0/).

Overall, the analyzed AL fibrils share the typical cross-β structure. This structure is stabilized, for example, through backbone hydrogen bonds or interactions between side chains (Absmeier et al. 2022; Radamaker et al. 2021a; Schmidt et al. 2016; Swuec et al. 2019). Besides these aspects, it is noticeable that different morphologies, post-transcriptional modifications, and mutations can be found (Radamaker et al. 2021b).

1.5 Aim of the Study

What are the main questions addressed in this work?

- Can the difference in the amyloidogenic potential between LCs from AL and MM patients be attributed towards specific sequence characteristics?
- Can the difference in the organ involvement of AL patients be associated with specific sequence characteristics?
- Does the presence of a potential HC binding partner or a high dFLC in the serum of AL patients reflect in specific sequence characteristics of the corresponding LCs?

What is the underlying relevance of these questions?

- AL amyloidosis is a rare disease that is often diagnosed at a late stage with potentially lethal consequences for the patients. This condition would be significantly improved, by screening options for MGUS and SMM patients, based on biomarkers. A promising approach provide the disease's underlying LCs themselves. This work focuses on lambda LCs, as AL patients are more likely to display lambda than kappa LCs (3:1). Nevertheless, these AL LCs should be compared with non-amyloidogenic LCs to identify potential crucial characteristics. For this comparison, LCs from the related plasma cell disorder MM, which do not form fibrils, offer the best option. The gained results can then potentially serve as a starting point for a risk stratification model for the disease progression towards AL.
- The specific organ tropism in AL is a characteristic that is not yet fully understood. To further investigate the pathogenic mechanism of the disease, LCs and particularly their mutations offer a promising approach. The underlying hypothesis of this approach is that the composition and mutations of the LCs first influence the misfolding and then the fibril deposition in different organs.
- One hypothesis in AL addresses the influence of a potential HC binding partner. More specific, the presence or absence of a HC might influence the pathological mechanisms of the disease. In this context, there might be a positive selection of certain LCs which exhibit favorable characteristics for one of the two opportunities. Vice versa, when analyzing these LCs, it might be possible to conclude on the pathological mechanisms. The same approach can be applied to a high dFLC. A high dFLC is, besides other aspects, associated with a higher tumor burden and a more aggressive tumor clone. This might also correlate with specific sequence features of the LCs.

2 Material and Methods

2.1 Material

2.1.1 Equipment and Consumables

The equipment and consumables used in this study are listed in **Table 6** and **Table 7**.

Table 6. Equipment.

Name	Manufacturer	Object
µCuvette G1.0	Eppendorf	RNA and DNA concentration measurement
4200 TapeStation system	Agilent	Electrophoresis and concentration measurements
7-2020	NeoLab	Vortex mixer
Accu-jet® pro	Brand	Pipette
B3 professional series microscope	Motic®	Microscope
Biometra® TProfessional thermocycler	Analytik Jena	Thermocycler
BioSpectrometer® basic	Eppendorf	Spectrometer
Centrifuge 5810	Eppendorf	Centrifuge
Centrifuge 5810 R	Eppendorf	Centrifuge
Countess™ II	Thermo Fischer Scientific	Cell counter
Färbekasten nach Hellendahl mit Erweiterung	Roth	Cell fixation
HTMR-133	HLC	Heating thermomixer
Lightning Volt OSP-250L Power Supply	Owl Scientific Inc.	Voltage supply
MC-24 touch microcentrifuge	Benchmark	Centrifuge
MICROMAT 135	AEG	Microwave
Molecular Imager® Gel DOC™ XR+ with image Lab™Software	Bio RAD	Agarose gel-documentation
multiSUB® MSChoice	Cleaver Scientific	Gel chamber
NextSeq 550 System	Ilumina	Next-generation-sequencer
PerfectSpin 24 plus	VWR Peqlab™	Centrifuge
Pipette “Discovery comfort” 100 – 1000 mL, 20 – 200 mL, 2 – 20 mL, 0.5 – 10 mL	HT lab solutions	Pipette
Pipette “Reference” 10 – 100 mL, 0.5 – 10 mL	Eppendorf	Pipette
Pipette “Research” 2 – 200 mL, 0.5 – 10 mL	Eppendorf	Pipette
PLJ 3500-2NM	Kern	Scale
Quantus™ Fluorometer	Promega	DNA quantification
REAX 2000	Heidolph	Vortex mixer
RoboSep™-S	Stemcell™ Technologies	$CD138^+$ cell sorting
Schott CLS-1172 100 mL	SCHOTT	Flask
Sprout® Plus	Heathrow Scientific®	Centrifuge
Thermo Scientific™ Cytospin 4 centrifuge	Thermo Fischer Scientific	FISH slide centrifuge

Name	Manufacturer	Object
Thermo Scientific™ PICO™ 17 microcentrifuge	Thermo Fischer Scientific	Centrifuge
Tischabzug Typ 1200	Waldner Scala	Fume hood
Vortex-Genie® 2 – G560E	Scientific Industries	Vortex mixer

Table 7. Consumables.

Name	Manufacturer
1.5 mL Eppendorf safe-lock® tubes	Eppendorf
15 mL polystyrene round-bottom tube	Falcon®
2 mL Eppendorf safe-lock® tubes	Eppendorf
5 mL Eppendorf tubes®	Eppendorf
Aqua Spüllösung	B. Braun
Bemis parafilm® M	Heathrow Scientific
C-Chip DHC-N01	NanoEnTek
Corning®Costar® Stripette®, serolog. Pipetten, 5 mL, 10 mL, 25 mL, 50 mL	Merck/Sigma-Aldrich®
Countess™ cell counting chamber slides	Thermo Fischer Scientific/Invitrogen™
Countess™ test beads	Thermo Fischer Scientific/Invitrogen™
DNA LoBind tubes 1.5 mL	neoLab Migge GmbH
Double Cytoslides™ microscope slides	Epredia™
Erlenmeyerkolben 500 mL	NeoLab
EZ double cytofunnel	Epredia™
H_2O Ampuwa Spüllösung	Fresenius Kabi
KIMTECH science precision wipes	Kimberly Clark Professionals™
PCR SingleCap 8er-SoftStripes 0.2 mL; DNA, DNase, and RNase free	Biozym Scientific
pluriStrainer® 100 µm	PluriSelect
RoboSep™ Filter Tip Racks	Stemcell™ Technologies
SafeSeal tips professional 10 µL, 20 µL, 100 µL, 200 µL, 1250 µL	Biozym Scientific
Semper Care latex powder free	Semperit Technische Produkte
Spitzen SafeSeal SurPhob 10 µL, 20 µL, 200 µL, 300 µL, 1250 µL	Biozym Scientific
Sterile Pasteur Pipette	LP Italiana
Vernichtungsbeutel	NeoLab
Water, DNase, RNase-free	MP Biomedicals. LLC
Zentrifugenröhrchen PP 50 mL	Greiner Bio-One

2.1.2 Chemicals and Enzymes

The chemicals and enzymes used in this study are listed in **Table 8**.

Table 8. Chemicals and enzymes.

Application	Name	Manufacturer
Bone marrow preparation, cell storage, FISH-slide preparation	Dulbecco's phosphate-buffered solution (DPBS) 10x	Thermo Fischer Scientific/Gibco™
Bone marrow preparation	EDTA – solution pH 8.0 (0.5 M) for molecular biology	PanReac AppliChem
	Fetal bovine serum (FBS), qualified, HI	Thermo Fischer Scientific/Gibco™

Application	Name	Manufacturer
	Histopaque® 1077	Merck/Sigma-Aldrich®
	UltraPure™ 0.5 M ethylenediaminetetraacetic acid (EDTA), pH 8.0	Thermo Fischer Scientific
	Ammonium chloride solution (NH_4Cl)	Stemcell™ Technologies
Cell counting	Trypan blue stain 0.4 %	Thermo Fischer Scientific/Gibco™
	Tuerk's solution	Merck
Cell storage	Buffer RLT	Qiagen
	2-Mercaptoethanol	Thermo Fischer Scientific/Gibco™
$CD138^+$ cell sorting	RoboSep™ buffer (Ready-to-use)	Stemcell™ Technologies
DNA/RNA Isolation	Ethanol absolute ≥99.8 %	VWR Chemicals
FISH-slide preparation	Acetic acid 100 %	Thermo Fischer Scientific/Carl Roth™
	Methanol 99.9 %, p.a.	Thermo Fischer Scientific/Carl Roth™
Gel electrophoresis	Take5™ 6x loading dye solution	highQu
	Agarose basic for biochemistry	neoFroxx
	SYBR® safe DNA gel stain	Thermo Fischer Scientific/Invitrogen™
	Tris acetate-EDTA (TAE) buffer 10x	Merck/Sigma-Aldrich
	Take5™ 50 bp DNA ladder	highQu
NGS	D1000 reagents	Agilent
	D1000 ScreenTape	Agilent
	NaOH 10 M BioUltra	Merck/Sigma-Aldrich
	Nuclease free water	Ambion®
	Tris-hydrochloride, 1 M-solution (pH 7.0/Mol. Biol.)	Thermo Fischer Scientific
PCR component	AmpliTaq Gold™ DNA polymerase with buffer II and $MgCl_2$	Thermo Fischer Scientific/Applied Biosystems™
	dNTP Set 100 mM solutions	Thermo Fischer Scientific
Reverse transcriptase reaction	Oligo(dT)18 oligonucleotide, 0.5 µg/µL	Thermo Fischer Scientific

2.1.3 Buffer

The buffers used in this study are listed in **Table 9**.

Table 9. Buffer. PBS = phosphate buffered saline, DPBS = dulbecco's phosphate-buffered saline, FBS = fetal bovine serum, EDTA = ethylenediaminetetraacetic acid, TAE = Tris acetate-EDTA.

Buffer	Composition/Manufacturer	Application
GTC buffer	25 mL RLT buffer 250 µL 2-mercaptoethanol	Cell storage

Buffer	Composition/Manufacturer	Application
Methanol – glacial acetic acid	60 mL methanol 20 mL acetic acid	Cell fixation
NH_4Cl – lysis	500 mL ammonium chloride solution 100 mL H_2O distilled	Erythrocyte cell lysis
PBS-buffer	900 mL H_2O distilled 100 mL 10x DPBS	Cytospin, cell storage “dry pellets”
PBS-EDTA-FCS	900 mL H_2O distilled 100 mL 10x DPBS 1 mL FBS 4 mL EDTA 0.5 M	Bone marrow preparation
TAE-buffer 1x	500 mL TAE 10x 4500 mL H_2O	Gel electrophoresis

2.1.4 Oligonucleotides

The oligonucleotides used in this study are listed in **Table 10**.

Table 10. Oligonucleotides. The oligonucleotides were produced and the Tm-calculation was performed by Eurofins. Devel. = developer, SH = Dr. sc. hum. Stefanie Huhn, NB = Natalie Berghaus, NA = not available.

Name	Devel.	Sequence [5'-3']	Purpose (alias)	Tm [C°]
VLKL12a_Huhn[a]	SH	GGTCCTGGGCTCAGTCTG	Amplification of IGLV1 and IGLV2, (1_Huhn_fw)	60.5
VLKL3c_Huhn[a]	SH	TGGTACCAGCAGAAGCCAGG	Amplification of IGLV3, (2_Huhn_fw)	61.4
VLKL4a_Huhn[a]	SH	CCAGCCTGTGCTGACTCA	Amplification of IGLV4, (3_Huhn_fw)	58.2
VLKL7a_Huhn[a]	SH	CAGACTGTGGTGACTCAGGAG	Amplification of IGLV7, (4_Huhn_fw)	61.8
JLHD123_rv[a]	SH	CTAGGACGGTGAGCTTGGTCCC	Amplification of IGLJ1, IGLJ2 and IGLJ3, (5_Huhn_rv)	65.8
JLKL4_rv[a]	SH	GACTCATCTAAAATGATCAGCTGGG	Amplification of IGLJ4, (6_Huhn_rv)	61.3
JLKL5_rv[a]	SH	AGACTCATCTAGGACGGTCAG	Amplification of IGLJ5, (7_Huhn_rv)	59.8
ß2-MRG_3	NA	TGTCGGATTGATGAAACCCAG	Establishment and verification of the full-length λ light chain sequencing method	57.9
ß2-MRG_5	NA	CTCGCGCTACTCTCTCTTTCT	Establishment and verification of the full-length λ light chain sequencing method	59.8
CLKL_A_rv_NB	NB	CACTGTCTTCTCCACGGTG	Amplification of IGLC, (13_NB_rv)	58.8
CLKL_B_rv_NB	NB	CCCTTCATGCGTGACCTG	Amplification of IGLC, (14_NB_rv), discarded in favor of CLKL_A_rv_NB	58.2
VLKL3_A_fw_NB	NB	CCTATGAGCTGACACAGCC	Amplification of IGLV3, N-terminal elongated, (15_NB_rv)	58.2
VLKL6_A_fw_NB	NB	CAGCCCCACTCTGTGTCG	Amplification of IGV6-57, (22_NB_rv)	60.5

Name	Devel.	Sequence [5'-3']	Purpose (alias)	Tm [C°]
VLKL3_H_fw_NB	NB	GCTGACTCAGGACCCTGC	Amplification of IGLV3-19, N-terminal elongated, (29_NB_rv)	60.5
VLKL3_I_fw_NB	NB	CCCTCAGTGTCCGTGTCC	Amplification of IGLV3-1, N-terminal elongated, (36_NB_rv)	65.0
VLKL3_B_fw_NB	NB	GCCAGGATTACCTGTGGG	Amplification of IGLV3-21, N-terminal elongated, discarded, (23_NB_fw)	58.2
VLKL3_E_fw_NB	NB	GTGTCAGTGGCCCCAGG	Amplification of IGLV3-21, N-terminal elongated, discarded, (26_NB_fw)	60.0
VLKL3_G_fw_NB	NB	GGAAGTAAAAGTGTGCACTGG	Amplification of IGLV3-21, N-terminal elongated, discarded, (28_NB_fw)	57.9
MM142_fw	NB	CCTATGAGGTGACTCAGCC	Sequence verification of MM142	64.7
MM120_fw	NB	CCTTTGCGCTGACTCAGC	Sequence verification of MM120	65.5

[a](Huhn 2018)

2.1.5 Kits and Standards

The kits and standards used in this study are listed in **Table 11**.

Table 11. Kits and standards.

Name	Protocol	Manufacturer
AllPrep DNA/RNA/Protein Mini Kit and QIAshredder Kit	″Simultaneous purification of genomic DNA, total RNA and total protein from animal and human Cells″	Qiagen
EasySep™ Human CD138 Positive Selection Kit II	″Human CD138 WB and BM positive selection II 17887″	Stemcell™ Technologies
High Pure PCR-Product Purification Kit	″Purification of PCR products in solution after amplification″	Roche
High-Capacity cDNA Reverse Transcription Kit with RNase Inhibitor	″High capacity cDNA reverse transcriptase kit″	Thermo Fischer Scientific/Applied Biosystems™

2.1.6 Next-Generation Sequencing

The next-generation sequencing kits used in this study are listed in **Table 12**.

Table 12. Next-generation sequencing reagents and kits.

Name	Manufacturer
IDT® for Illumina® DNA/RNA UD Indexes Set B, tagmentation (96 Indexes)	Illumina
Illumina® DNA Prep, (M) tagmentation (24 samples)	Illumina
NextSeq 500/550 Mid Output Kit v2.5 (150 cycles)	Illumina
NextSeq PhiX Control Kit	Illumina
QuantiFluor® ONE dsDNA System	PROMEGA

2.1.7 Counterprograms and Websites

The computer programs and websites used in this study are listed in **Table 13**.

Table 13. Computer programs and Websites.

Name	Application
abYsis (Swindells et al. 2017)	Sequence analysis tool http://www.abysis.org/abysis/ (05.07.2022)
AL-Base (Bodi et al. 2009)	Database https://wwwapp.bumc.bu.edu/BEDAC_ALBase/ (05.07.2022)
ChemSketch (ACD)	Chemical structure drawing software https://www.acdlabs.com/resources/free-chemistry-software-apps/chemsketch-freeware/ (27.10.2022)
Clustal Omega (Madeira et al. 2019)	Sequence alignment tool https://www.ebi.ac.uk/Tools/msa/clustalo/ (05.07.2022)
Ensembl (Howe et al. 2021)	Database https://www.ensembl.org/index.html (05.07.2022)
Ensembl Blast (Howe et al. 2021)	Sequence BLAST tool https://www.ensembl.org/Multi/Tools/Blast?db=core (05.07.2022)
ExPasy – Compute pI/MW (Duvaud et al. 2021)	Sequence analysis tool https://web.expasy.org/compute_pi/ (05.07.2022)
ExPasy – ProtParam (Duvaud et al. 2021)	Sequence analysis tool https://web.expasy.org/protparam/ (05.07.2022)
ExPasy – Translate (Duvaud et al. 2021)	Translation tool https://web.expasy.org/translate/ (05.07.2022)
Genbank (Benson et al. 2013)	Database https://www.ncbi.nlm.nih.gov/genbank/ (05.07.2022)
IBM SPSS® statistics	Statistical analysis program https://www.ibm.com/de-de/spss (12.10.2022)
Image lab	Gel electrophoresis documentation program by BIO RAD
IMGT (Lefranc 2014)	Database https://www.imgt.org/ (05.07.2022)
IMGT/DomainGapAlign (Ehrenmann et al. 2010)	Sequence analysis tool https://www.imgt.org/3Dstructure-DB/cgi/DomainGapAlign.Cgi (08.11.2022)
Local Run Manager 2.2.1.1645	Next-generation-sequencer firmware https://support.illumina.com/sequencing/sequencing_software/local-run-manager/downloads.html (18.10.2022)
LICTOR	Online tool for amyloidogenity prediction http://lictor.irb.usi.ch/lictor/ (07.12.2022)
Local Run Manager Generate FASTQ Analysis Module 2.0.1	Module for generating FASTQ files with the next-generation-sequencer https://support.illumina.com/downloads/local-run-manager-generate-fastq-module-v2.html (18.10.2022)
MedCalc Software Ltd. Relative risk calculator	Web-based tool for relative risk calculations https://www.medcalc.org/calc/relative_risk.php Version 20.2 (15.12.2022)
MEGA – Molecular Evolutionary Genetic Analysis (Tamura et al. 2021)	Sequence alignments and analysis program https://megasoftware.net/ (05.07.2022)
MiXCR (Bolotin et al. 2015)	Analysis program of next-generation sequencing data https://mixcr.readthedocs.io/en/master/ (05.07.2022)

Name	Application
Multiple Oligonucleotide Analyzer (Breslauer et al. 1986)	Oligonucleotide design tool https://www.thermofisher.com/de/de/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/multiple-oligonucleotide-analyzer.html (05.07.2022)
R	Data visualization, analysis, and statistical analysis program https://www.r-project.org/ (05.07.2022) packages: ggplot2 (Wickham 2016), ggsignif (Ahlmann-Eltze and Patil 2021)
R-studio	User interface R https://www.rstudio.com/ (05.07.2022)
Sequence massager	Sequence modification tool https://biomodel.uah.es/en/lab/cybertory/analysis/massager.htm (05.07.2022)
Tango (Rousseau et al. 2006)	Sequence analysis tool http://tango.crg.es/ (05.07.2022)
VBase2 (Retter et al. 2005)	Sequence BLAST tool http://www.vbase2.org/ (05.07.2022)
WebLogo	Web-based application to design sequence logos http://weblogo.berkeley.edu/logo.cgi (19.10.2022)

2.2 AL Amyloidosis Patients and Samples

Samples and data from AL patients which displayed a disease associated lambda LC and had visited the Amyloidosis Centre of the University Hospital Heidelberg between January 2019 and November 2021 were included in this work. The study was approved by the Ethics Committee of the University of Heidelberg (S-123/2006, renewed 07.12.2021) and followed the Helsinki guidelines for research of human subjects. All patients provided written informed consent allowing research on their sample material and clinical data. All patients received therapy according to local standards – additionally, eight patients included in this analysis received a heart transplant (FOR117, FOR130, FOR140, FOR161, FOR196, FOR197, FOR214, and FOR218).

Patients underwent bone marrow aspiration as part of clinical diagnostics workups. The aspirates were forwarded to the GMMG central laboratory and biobank multiple myeloma which is under the lead of Dr. sc. hum. Stefanie Huhn (University Hospital Heidelberg). Patients who underwent bone marrow aspiration before the start of this work (n = 22) and had leftover sample material available in the biobank were also included in this study. Fresh bone marrow aspirate samples were processed according to the following protocols in chapter 2.5. In the framework of this study, biobank-derived derived samples were processed and analyzed as described starting from chapter 2.6.

2.3 Multiple Myeloma Patients and Samples

This study included samples and data from 52 newly diagnosed MM patients which displayed a dominant lambda LC and were treated in GMMG-HD6 clinical trial (EudraCT No.: 2014-003079-40) (Salwender et al. 2019). The trial was performed in accordance with the declaration of Helsinki and the European Clinical Trial Directive (2005) and was approved by the local ethics committees of all participating institutions. All patients provided written informed consent before participating in the study and allowed research on their sample material and clinical data. The GMMG study group provided samples as well as clinical data and the board of GMMG e.V. approved the usage within this work.

Bone marrow sample preparation was not performed within the framework of this work but followed the protocols applied for AL samples (chapters 3.2 – 3.3.2). In addition, LC sequences were generated using a bulk ribonucleic acid (RNA) sequencing approach (chapters 2.7.2 and 2.8.2) which was not performed as part of this work.

In this work, several MM samples were used for sequence validation. In these cases, the GMMG study group provided RNA which was processed and analyzed following the protocols in chapters 2.6.3 – 2.9.

2.4 Clinical Characteristics of the Cohorts

Clinical parameters of the AL patients were collected and analyzed on basis of the clinical patient reports. The definition of organ involvement was performed based on established criteria (Gertz et al. 2005) by Prof. Dr. med Ute Hegenbart (University Hospital Heidelberg) and patients were indicated having “dominant heart involvement” or “dominant heart and kidney involvement” if no other clinically relevant organ manifestation was noted. AL patients with dominant kidney involvement and all patients who could not be classified into one of these three groups were analyzed by Sarah Schreiner (University Hospital Heidelberg).

The clinical parameters of the MM patients were retrieved from the dataset provided by the GMMG study group. The patients enrolled in the GMMG-HD6 trial, must be newly diagnosed with MM, must not be treated beforehand and a systemic AL amyloidosis was excluded. The more detailed inclusion and exclusion criteria for this study have been published by Salwender et al. and are listed in the **Supplementary Information Table 1** (Salwender et al. 2019).

Within the framework of this work, several clinical characteristics were evaluated. The responsible investigator or institution, the starting material, the applied test, and the reference areas are listed in **Table 14**.

Table 14. Overview of the clinical characteristics evaluated in this study, who collected them, the reference areas, and the analysis methods used. No information was available for the MM cohort for the following parameters: bone marrow aspirate volume, percentage of $CD138^+$ plasma cells in the sample, number of cell sorts with no harvested $CD138^+$ plasma cells, GFR CKD-EPI, NT-proBNP, TNT. The bone marrow aspirates of all AL amyloidosis patients, regardless of the organ involvement and a kappa or lambda light chain, performed in the course of clinical diagnostics were processed in equal parts by Sarah Schreiner and myself. In this context, the volume of samples was recorded. The highest measured percentage of a genetic aberration in the FISH result was defined as the proportion of the main clone (Bochtler et al. 2018). Inv. = Investigator, ALZ = Zentrallabor-Analysezentrum University Hospital Heidelberg, NT-proBNP = N-terminal provisional brain natriuretic peptide, PC = plasma cells, LC = light chain, FISH = fluorescence *in situ* hybridization, HC = heavy chain, GFR CKD-EPI = calculation of the glomerular filtration rate using the formula established by the Chronic Kidney Disease Epidemiology Collaboration, AP = alkaline phosphatase, hsTNT = heart associated troponin T. IFE = IFE = immunofixation electrophoresis in serum, MCL = Molekular-cytogenetisches labor University Hospital Heidelberg, CLIA = chemiluminescence-immunoassay, ECLIA = electrochemiluminescence-immunoassay, BM = bone marrow, NB = Natalie Berghaus, SS = Sarah Schreiner, PIC = physician in charge.

Clinical characteristics	Inv.	Starting material and test	Normal range
BM aspirate volume [mL]	NB, SS	Measurement	-
$CD138^+$ PC in the BM sample [%]	NB	Division of the CD138+ PC count by the PC cell count of the complete sample	-
Cell sorts with no harvested $CD138^+$ PC [n]	NB	Counting	-
Percentage of the main clone in the FISH analysis	MCL	FISH analysis (Bochtler et al. 2018)	0 %
Age [y]	PIC	Admission interview	-
Sex [female/male]	PIC	Admission interview	51/49[1]
Timepoint of BM aspiration	PIC	Admission interview	-
PC infiltration in the BM [%]	PIC	Cytological evaluation of BM smear	<5[2]
dFLC at diagnosis [mg/L]	ALZ, NB	Serum, immuneturbimetric, followed by the calculation of the difference between disease-associated and uninvolved circulation free LCs	free lambda LC: 5.7-26.3[3] free kappa LC: 3.3-19.4[3]
Proteinuria [g/d] (range)3	ALZ	Urine, electrophoresis	<0.25[3]
GFR CKD-EPI [mL/min/1.73 m²]	ALZ	Heparin plasma, calculation	>60[3]
NT-proBNP [ng/L]	ALZ	Heparin plasma, CLIA	<450[3]
hsTNT [µg/L]	ALZ	Heparin plasma, ECLIA	<14[3]
AP [U/L]	ALZ	Heparin plasma, photometry	55-105[3]
Detectable M-gradient [n]	ALZ	IFE	0
Serum HC present [n]	ALZ	IFE	0
HC type	ALZ	IFE	0
Gain 1q21 [n]	MCL	FISH analysis	10 %[4]
t(11;14) [n]	MCL	FISH analysis	10 %[4]

[1](Genesis), [2](Kumar et al. 2016), [3]reference information retrieved from the datasheet provided by the Zentrallabor-Analysezentrum University Hospital Heidelberg, [4]reference information retrieved from the datasheet provided by the Molekular-cytogenetisches Labor University Hospital Heidelberg

2.5 Bone Marrow Sample Preparation

2.5.1 Isolation of Mononuclear Cells

To isolate the mononuclear cells, the bone marrow samples were processed following an internal revised quality management document and a standard operating procedure document (SOP) (″SOP für Dichtegradienten-Zentrifugation von PB und KM″, step: ″Ficoll vor Sort″,

last saved 15.08.2022, revised by Stefanie Huhn) (**Supplementary Information Figure 2**). In this context, a density gradient centrifugation followed by an erythrocytes cell lysis was performed.

2.5.2 CD138-Cell Sorting

The $CD138^+$ positive selection was performed as part of the clinical routine diagnostic and led to samples with a high purity of tumor cells. CD138 is an antigen, which is highly expressed on malignant plasma cells in peripheral blood and bone marrow of patients with monoclonal gammopathies – including AL and MM (Chilosi et al. 1999; Kriegsmann et al. 2018; Ridley et al. 1993; Wijdenes et al. 1996). Therefore, analysis and sequencing of cDNA obtained from $CD138^+$ sorted cells should significantly reduce background signals and allow LC analysis of the malignant plasma cells with fewer disturbance of germline signals.

The isolation of $CD138^+$ cells was performed according to an internal revised quality management document and SOP (″SOP zur Roboterunterstützten CD138+ Sortierung von Probenmaterial nach dem StemCell Protokoll″, last saved 15.08.2022, revised by Stefanie Huhn) (**Supplementary Information Figure 3**). The isolated mononuclear cells were used for a robotry (RoboSep™-S – The Fully Automated Cell Separator; Stemcell™ Technologies) supported magnetic-bead-based $CD\text{-}138^+$ cell sorting (EasySep™; Stemcell™ Technologies). For this purpose, the manufacturer provided buffers, magnetic beads, and antibodies. For cell sorting, the protocol ″Human CD138 WB and BM positive Selection II 17887″ was used.

The cell counts before as well as after cell sorting for the $CD138^-$ fraction were determined with 10 µL of the sample and 90 µL of Trypan blue using a cell counter (Countess™ II; Thermo Fischer Scientific). This was performed following an internal revised quality management document and SOP (″SOP zum Zählen der Zellzahl von Proben mittels Countess″, last saved 15.08.2022, revised by Stefanie Huhn) (**Supplementary Information Figure 4**). The cell count of the positive fraction was always determined according to the $CD138^+$ cell sorting protocol using a microscope (10µl Trypan Blue Stain 0.4 % + 10µl cell suspension; C-Chip Neubauer Improved DHC-N01) (**Supplementary Information Figure 5**) (B3 professional series microscope; Motic®).

The purity of the $CD138^+$ samples were not determined due to the rarity of $CD138^+$ cells and their value for research. Nevertheless, FISH analyses based on the $CD138^+$ samples suggest a median purity of 95 %.

2.5.3 Cytospins

To gain information about the genetic aberrations of the patients it is necessary to generate cytospin slides for the FISH analysis. The FISH analysis is a clinical standard diagnostic workup and was carried out in the molecular-cytogenetic laboratory of the University Hospital Heidelberg under the leadership of Prof. Dr. sc. hum. Anna Jauch. The slides were prepared following an internal revised quality management document and SOP (″SOP zur Erstellung von Präparaten für die Zytogenetische Analyse von Myelomzellen″, last saved 15.08.2022, revised by Stefanie Huhn) (**Supplementary Information Figure 5**). Data was – in return – available via the clinical patient reports.

2.5.4 Cell-Pellets

The $CD138^+$ sorted cells were divided up and stored according to an internal revised quality management document and SOP (″Amyloidose – SOP Produkte und Materialaufteilung″, approved: 07.12.20, approved by Stefanie Huhn) (**Supplementary Information Figure 6**). Providing slides for the FISH analysis (3.2.3) has the highest priority; as second priority GTC pellets and as third priority ″dry-pellets″ should be provided. This accounts for the $CD138^-$ fraction as well. Within the framework of this work, GTC pellets were used for DNA and RNA extraction. GTC, or more precisely guanidinium thiocyanate, and β-mercapthoethanol are in this workup commonly used as nucleic acid protectors.

2.6 Molecular Biological Methods

2.6.1 Isolation of Genomic DNA and RNA

The isolation of genomic DNA and RNA was performed using the kit ″AllPrep DNA/RNA/Protein Mini Kit″ (Quiagen), following the protocol ″Simultaneous Purification of Genomic DNA, Total RNA and Total Protein from Animal and Human Cells″. No additional RLT buffer with β-mercaptoethanol was added, due to sample storage in GTC buffer. The lysate was directly used for homogenization using a QIAshredder. For elution of total RNA, 15 µL elution buffer was added and the samples were centrifuged for 1 min at 13000g. This was repeated a second time, to obtain a final volume of 30 µL. A second centrifugation step was also performed for the DNA elution (2x 50 µL) – no protein was extracted. In the framework of this study, only the extracted RNA was used for further experiments.

2.6.2 DNA and RNA Concentration Measurement

The total DNA and RNA concentration was determined using a spectrometer (BioSpectrometer® basic; Eppendorf) and 2 µL of the sample. For DNA measurement 2 µL elution buffer and for RNA measurement 2 µL water were used as a blank.

2.6.3 Reverse Transcriptase Reaction

The reverse transcriptase reaction was performed using the ″High capacity cDNA reverse transcriptase Kit″ (Thermo Fischer Scientific/Applied Biosystems^{TM}) kit following the manufacturer protocol.

2.6.4 Polymerase-Chain Reaction

The polymerase-chain-reaction (PCR) for lambda LC amplification was performed following a published protocol (Huhn 2018) (AmpliTaq Gold^{TM} DNA polymerase with buffer II and $MgCl_2$; Thermo Fischer Scientific/Applied Biosystems^{TM} and dNTP Set 100 mM solutions; Thermo Fischer Scientific). In addition, several oligonucleotides for full-length lambda LC sequencing were adapted from this protocol (**Table 10**). In a first PCR, a multiplex oligonucleotide set consisting of already developed forward oligonucleotides (VLKL12a_Huhn, VLKL3c_Huhn, VLKL4a_Huhn, VLKL7a_Huhn) was used. These oligonucleotides have been shown to bind on the common IGLV families. For sequencing of the full-length lambda LCs, a reverse oligonucleotide located in the IGLC segment was newly established (CLKL_A_rv_NB). This first PCR procedure was developed using provided RNA (by Dr. sc. hum Mohammed H.S. Awwad, University Hospital Heidelberg) from two cell lines (OPM2, RPM1). The RNA was processed following the protocol in chapter 2.6.3. It was shown, that the oligonucleotide combination successfully amplified the complete lambda LC and that the subsequent Sanger sequencing was of good quality (**Supplementary Information Figure 1**). Initially, FOR101 and FOR102 were sequenced with four different oligonucleotide combinations and it was possible to detect the same underlying sequence in all sequencing reactions. The forward oligonucleotides were in all cases a multiplex consisting of VLKL12a_Huhn, VLKL3c_Huhn, VLKL4a_Huhn, VLKL7a_Huhn as reverse oligonucleotide a) JLHD123_rva and JLKL4_rva b) CLKL_A_rv_NB and CLKL_B_rv_NB c) CLKL_A_rv_NB or d) CLKL_B_rv_NB were used. Thus, the protocol (VLKL12a_Huhn, VLKL3c_Huhn, VLKL4a_Huhn, VLKL7a_Huhn, and CLKL_A_rv_NB) was applied to the remaining patient samples. In this process, depending on the sequence quality and the detected IGLV family, additional PCR steps with newly established oligonucleotides were performed (**Figure 10**). When a sequence displayed multiple nucleotide signal overlaps, a PCR with VLKL6_A_fw_NB was performed to confirm or exclude association with the IGLV6-57

family. For LC sequences with poor quality, a single plex PCR using only the respective forward oligonucleotide was performed. Due to the fact, that the VLKL3c_Huhn oligonucleotide is located in the FR2 of the IGLV3 reference sequences an additional PCR (VLKL3_A_fw_NB) was performed to cover the complete IGLV3-segment. In addition, for sequencing of the IGLV3-19 (VLKL3_H_fw_NB) as well as IGLV3-1 (VLKL3_I_fw_NB) family an additional PCR was performed. For sequencing full-length IGLV3-21 LCs, it was possible to generate two N-terminal resolved sequences using the VLKL3_A_fw_NB oligonucleotide. Three additional oligonucleotides (VLKL3_B_fw_NB, VLKL3_E_fw_NB, VLKL3_G_fw_NB) were tested using three samples (FOR154, FOR162, and FOR169) but it was not possible to generate an evaluable N-terminal elongated sequence using the same oligonucleotide in all three cases.

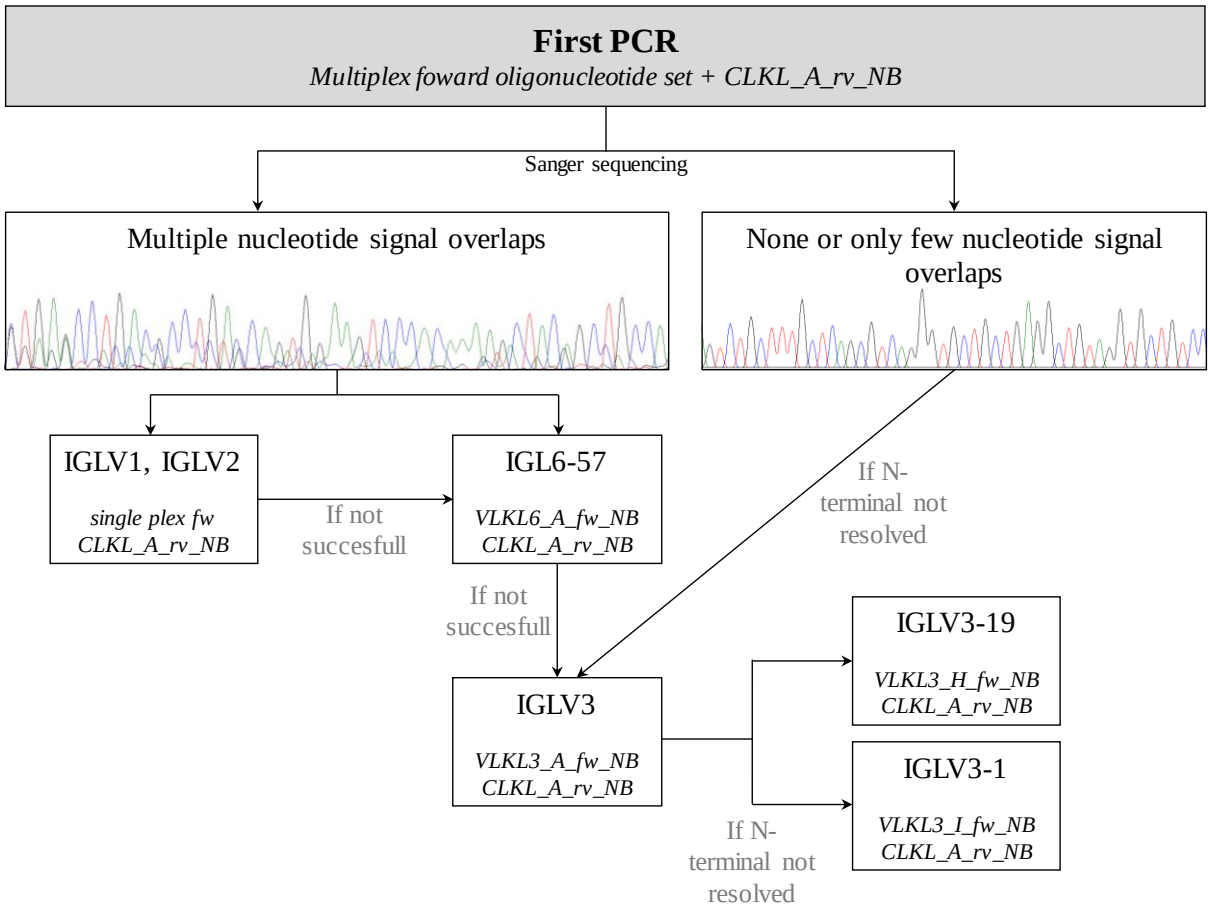


Figure 10. Polymerase chain reaction scheme for sequencing full-length lambda light chains with associated oligonucleotides. Multiplex forward oligonucleotide set: VLKL12a_Huhn, VLKL3c_Huhn, VLKL4a_Huhn, VLKL7a_Huhn.

If it was possible to generate different sequences with the same sequence quality for one sample – indicated by the number of not unambiguously determined AA positions based on nucleotide signal overlaps – these samples were classified as not evaluable.

The verification of the MM sequences was performed using a single-plex PCR with the respective forward oligonucleotide and CLKL_A_rv_NB. For the IGLV3-1 assigned MM sequences MM142 (MM142_fw) and MM120 (MM120_fw) a patient-specific oligonucleotide was used.

2.6.5 Analytic Gel-Electrophoresis

The analytic gel-electrophoresis was performed in horizontal chambers (multiSUB® MSChoice; Cleaver Scientific) filled with TAE-buffer (120V, 15 min). For agarose gels 1.5 % agarose and 1 % Syber Safe (SYBR® Safe DNA gel stain; Thermo Fischer Scientific/Invitrogen™) solved in TAE-buffer was used. For analysis, 5 µL of the sample and 1 µL loading dye (Take5™ 6x loading dye solution; highQu) were mixed and 3 µL gen ruler (Take5™ 50 bp DNA Ladder) were used as reference. The visualization was performed using a gel documentation system with corresponding software (Molecular Imager® Gel DOC™ XR+ with Image Lab™Software; Bio RAD).

2.6.6 PCR-Product Purification

For purification of the PCR-products the ″High Pure PCR-Product Purification Kit″ (Roche) following the protocol ″Purification of PCR-Products in Solution after Amplification″ was used. The elution step was performed with each 15 µL and two centrifugation steps (13000g for 1 min) instead of 50 – 100 µL and a one-minute centrifugation step (13000g).

2.7 DNA Sanger Sequencing

For complementary DNA (cDNA) Sanger sequencing, samples were adjusted to a final concentration according to the manufacturer's instructions. Together with the corresponding oligonucleotides (10 pmol/µL), the samples were sent to Eurofins/GATC (SupremeRun Tube) for sequencing.

2.8 Next-Generation Sequencing

2.8.1 AL Amyloidosis Patients

To further characterize the sequence composition of IGLV3-21 samples a next-generation sequencing (NGS) analysis was performed for seven patients/samples. The Illumina DNA Prep protocol and a Next Seq 550 (Illumina) were used for this purpose. These experiments were

carried out together with Sarah Schreiner (University Hospital Heidelberg) and Dr. rer. nat. Philipp Reichert (University Hospital Heidelberg).

Two different approaches were chosen per patient/sample. In the first approach, PCR samples generated via a multiplex forward oligonucleotide set (VLKL12a_Huhn, VLKL3c_Huhn, VLKL4a_Huhn, VLKL7a_Huhn) and the reverse oligonucleotide (JLHD123_rv) were analyzed ($MP_{\_NGS\_PCR}$). The multiplex oligonucleotide set was shown to successfully amplify the commonly used IGLV segments. The second approach concerns PCR samples that were generated using the oligonucleotides VLKL3_A_fw_NB and JLHD123_rv ($N_{\_NGS\_PCR}$). The PCRs followed the protocol in chapter 2.6.4.

For each approach, two PCRs were performed and pooled before tagmentation. For sample processing, the Illumina document # 1000000025416 v09 (June 2020) ″Illumina DNA Prep Reference Guide″ was used as a guideline as well as the corresponding kits (**Table 12**).

The working steps and aberrations are listed below:

- Tagmentation of Genomic DNA
  500 ng per sample was used
- Post Tagmentation Cleanup
- Amplify Tagmented DNA
- Clean Up Libraries
- Pool Libraries
- Check Library Quality

  The samples were checked for quality using both the Quantus (Quantus™ Fluorometer; Promega) and the Tapestation (4200 TapeStation System; Agilent) systems.

  The quality check using the Tapestation system was performed following an internal revised quality management document and SOP (″SOP zur Automatischen Elektrophorese am TapeStation System″, step: ″DNA, PCR Produkte, NGS Library″, approved 08.06.2021 by Stefanie Huhn). In addition, the Quantus™ fluorometer and the corresponding QuantiFluor® ONE dsDNA System were used for quality check and concentration measurements following the manufacturer protocol (″QuantiFluor ONE dsDNA System Quick Protocol FB200″).

- Dilute Libraries to the Starting Concentration
  In a first step, the molarity of the individual libraries was calculated using the following formula:

$$X\ nM = 10^6 \times \frac{library\ concentration\ (ng/\mu L)}{660\ g/mol \times average\ library\ size\ in\ bp}$$

In the next step, the individual libraries were diluted to achieve a concentration of 2 nM in 15 µL RBS buffer. This was possible for all samples except FOR169 in the $N_{\_NGS\_PCR}$ approach. So, only 15.0 µL instead of 16.7 µL of the sample was used. Subsequently, 5 µL of each library was pooled.

The Illumina protocol ″NextSeq System Denature and Dilute Libraries Guide″ (Document # 15048776 v09 December 2018) was used to dilute the libraries and pHiX to the start concentration of 1.4 pM.

The run was started according to the guidelines of the ″NextSeq 500 System Guide″ (Document # 15069765v06 June 2019) document. In addition, the ″Local Run Manager Generate FASTQ Analysis Module Workflow Guide″ (Document # 1000000003344 July 2018) was used.

2.8.2 Multiple Myeloma Patients

The bulk RNA NGS sequencing of MM samples was not performed within the framework of this work. Sequencing libraries were prepared using the Illumina TruSeq stranded mRNA kit. Sequencing was performed on the Illumina NovaSeq 6000 PE 100 S1 platform.

2.9 Analysis

2.9.1 Sanger Sequencing Analysis

For analysis of the Sanger sequencing results the ab1 files (provided by Eurofins) of the forward and reverse sequencing reaction were analyzed using MEGA (Tamura et al. 2021). To ensure a high sequence quality the sequences were trimmed at the N- und C-terminus and each position was analyzed concerning nucleotide overlaps. If present, the IUPAC-IUB code was used to specify these positions. Afterwards, the forward and reverse sequences were combined into one full-length LC sequence. In the next step, the ExPASy Translate tool (Duvaud et al. 2021) was used for translation into an AA sequence. Each nucleotide triplet containing an uncertain position was analyzed and if possible, the respective AA was complemented.

2.9.2 Next-Generation Sequencing Analysis

The bioinformatic analysis of the AL and MM NGS experiments was not performed within the framework of this work.

The bioinformatical data processing concerning the AL IGLV3-21 samples, was performed by Sarah Schreiner (University Hospital Heidelberg) using MIXCR (Bolotin et al. 2015) and the

following commands: Java –jar mixcr.jar analyze shotgun --verbose -- species hs -- starting-material dna -- receptor-type bcr --contig-assembly sum.fastq analysis.

The bioinformatical data processing concerning the MM samples, was performed by Dr. re. nat. Alexandra M. Poos (University Hospital Heidelberg) also using MIXCR to extract the full sequences of the IG lambda receptor (option: --contig-assembly) (Bolotin et al. 2015)

Within the framework of this work, the cDNA sequences were translated using the ExPASy ″Translate″ tool (Duvaud et al. 2021). In the following, the AL and MM sequences, regardless of Sanger sequencing or NGS, were analyzed equally.

2.9.3 IGL Family Assignment

The IGLV family assignment was performed using the cDNA sequence and VBase2 (Retter et al. 2005) as well as the AA sequence and Ensembl Blast (Howe et al. 2021). If a deviation between the two analysis platforms was indicated, the sequence was stated as non-evaluable. The IGLJ segment was identified using the cDNA sequence and VBase2 and the IGLC segment using the AA sequence and Ensembl Blast. The CDR regions were determined using abYsis (Swindells et al. 2017).

2.9.4 Mutation Analysis

For sequence analysis, the VBase2 IGLV reference sequence was used as standard. However, in some cases, a deviation between the IGLV reference deposited in Vbase2 and Ensembl can be noted and a mutation was only indicated if a deviation of the patient-derived LC towards both reference sequences was noted. This did not concern the IGLV-IGLJ-linker region and in this context, the VBase2 reference was again used as standard. For the IGLJ segment, the reference sequences deposited in GenBank (Benson et al. 2013) were used. IGLJ2*01 and IGLJ3*01 (Gene ID: 28832) show the same cDNA and AA sequence and are therefore defined as IGLJ2. IGLJ3*02 (Gene ID: 28831) is defined as IGLJ3. For the IGLC segment, the Ensembl reference sequences were used but due to the fact, that the AA sequences contain an undefined AA at the first position, the corresponding GenBank reference sequences were considered at this position. Sequence alignments were performed using Clustal omega (Madeira et al. 2019). For the calculation of the median mutation count and frequency, only unambiguously determined mutations were used. In addition, the AL and MM sequences that were to be under comparison with each other were trimmed to the same length. This concerns the N- as well as the C-terminus of the sequences. The sequences were also N-terminally constrained by analyzing only the region covered by the Vbase2 reference. To calculate the median mutation count, only the mutated segments were considered and the median was calculated only from

them. The MM120 sequence displayed an insertion of four AAs, which was in this analysis defined as one mutation. A deletion was also indicated as a mutation.

Besides the mutation distribution, individual positions were also analyzed. If not otherwise specified, only positions that displayed a mutation in ≥50 % were defined as mutation hotspots. In this context, the total number was determined by only counting the ambiguously determined AA positions – in contrast to the calculation of the median mutation count and mutation frequency. Besides the AA exchange, the underlying cDNA nucleotide triplet was also investigated.

The IGLV2-14 MM and AL sequences were analyzed using the web-based tool LICTOR. LICTOR is an application for the prediction of amyloidogenity of lambda LCs based on mutations (Garofalo et al. 2021).

2.9.5 Amino Acid Composition Analysis

The sequences were trimmed as described in chapter 2.9.4. For analysis of the AA composition, the positions that could not be clearly determined were changed to the respective reference AA. Subsequently, an analysis was performed using the ExPASy ″ProtParam″ tool (Duvaud et al. 2021). For the purpose of conciseness, only differences ≥0.5 % were mentioned in the text and differences >0.5 % were numerically specified.

2.9.6 Biophysical Parameter Analysis

The sequences were trimmed as described in chapter 2.9.4 and for AA positions that could not be clearly determined the respective reference AA was used. The beta sheet aggregation tendency (AGG) value was calculated using TANGO (Rousseau et al. 2006). The grand average of hydropathicity (GRAVY) value was calculated using the ExPASy ″ProtParam″ tool (Duvaud et al. 2021). The molecular weight (Mw), as well as the isoelectric point (pI), were determined using the ExPaSy ″Compute pI/Mw″ tool (Duvaud et al. 2021). Not only the pI but also the difference between the patient-derived LC and a reference LC was calculated (ΔpI). In this context, for each patient-derived LC, a reference with the respective IGLV, IGLJ, and IGLC segment as well as the patient-specific linker region was individually built. A not unambiguously determined AA in the linker region was neglected. For several patient-derived LC sequences, no clear assignment towards IGLJ2 and IGLJ3 was possible but the two respective reference sequences did not differ in pI. This also applies to sequences, which were not clearly assigned to IGLC2 or IGLC3. In the analyzed IGLC sections no difference between the two reference sequences can be defined on AA level.

2.9.7 Statistical Analysis and Data Visualization

The statistical analysis was performed using the IBM SPSS statistics program. If not otherwise indicated, values are presented as median.

The exact Fischer test (Fischer), two-tailored, or Chi-Quadrat test (Chi) were used to determine the significance levels of comparison for nominal data. For metric data normal distribution was tested using the Shapiro-Wilk normality test. If normal distribution was indicated in both independent subgroups, the significance level was calculated accordingly by using a two-tailored t-test (t-test). In this context, the variance equality was tested using the Levene-Test. Significance levels between independent subgroups and non-parametric data were calculated using a median test (Median) or two tailored Mann-Whitney U test (U-test). Significance values were adjusted by the Bonferroni correction for multiple tests per test-set. A p-value ≤ 0.05 was considered statistically significant and was numerically specified in the text. The differences concerning the mutation count and frequency were tested using the complete dataset without further separation concerning the frequency and mutation count. In addition, analysis and data visualization was performed using Excel, ChemSketch, WebLogo, R and R-studio using the "ggplot2" (Wickham 2016) and "ggsignif" (Ahlmann-Eltze and Patil 2021) packages.

The relative risk was calculated using the web-based MedCalc relative risk tool.

3 Results

Work program:

Before analyzing the LC sequences themselves, the AL and MM cohorts are clinically characterized. In the next chapter, the composition of the MM samples is analyzed as the LC sequences of the MM patients were obtained using a different method. In the first chapter of the actual LC sequence analysis, the IGLV family and IGLV subfamily distribution are examined. From this point onwards, the analyses are performed not only for AL, MM, and the AL organ tropism but also for two additional parameters: the presence or absence of a clonal HC in the patient or a dFLC > or < 180 mg/L. Next, the linkage between the IGLJ and IGLC segments are analyzed first, and the composition of the complete LCs afterwards. Detailed sequence analyses are performed for IGLV subfamilies that are most frequently associated with AL and to which more than 10 AL sequences can be assigned: IGLV2-14, IGLV3-1, IGLV3-21, and IGLV6-57. Each IGLV subfamily is examined separately concerning patient characteristics, general mutation number and distribution within the LCs, specific mutations detected in the sequence alignment, the AA composition of the LC, and finally, various biophysical parameters. The analysis of IGLV3-21 deviates from this scheme; here, next-generation sequencing experiments were performed to investigate the sequence composition within the samples in more detail. The IGLV6-57 analysis has the particularity that IGLV2-23 sequences are used for comparison. Finally, the results are compared with each other, and the less frequently detected IGLV subfamilies are addressed less extensively. In the last step, the results from these rare IGLV subfamilies are compared with the previous ones.

3.1 Clinical Characteristics of the Cohorts

In the context of this study, 92 bone marrow samples of AL patients were processed. This concerns only samples from AL patients with a dominant heart (AL_H) or dominant heart and kidney involvement (AL_HK). This stratification was performed based on established criteria (chapter 2.4), to further investigate the AL organ tropism. However, the final AL cohort consisted of only 82 AL samples: 61 AL_H and 21 AL_HK. This loss of samples is because after the bone marrow $CD138^+$ cell sorting, these cells need to be used for an obligatory FISH analysis first. Afterwards, two samples corresponding to AL_HK and nine samples corresponding to AL_H did not contain enough $CD138^+$ cells for further analysis (**Table 15**).

In the final AL cohort, the bone marrow samples of the AL_H and AL_HK patients had a comparable median volume of 52 – 60 mL and contained an average of 0.5 % $CD138^+$ plasma

cells. In seven AL_H cases, it was not possible to detect any $CD138^+$ plasma cells. However, it was assumed that enough cells and Ig LC RNA were available for further analysis due to a detection limit of 2.5e+04 cells in the $CD138^+$ cell sorting. For the 52 MM patients included in this study, no information about the initial volume of the bone marrow aspirate and $CD138^+$ cell ratio was available. Further information on the clonality of the $CD138^+$ samples was provided by the FISH analysis. In this context, the highest percentage measured was compared between cohorts. Here, the MM samples (98 %, range: 87 % – 100 %) displayed a significantly higher value than the AL_H (90 %; p = 1.13e-08 (Median)) and AL_HK samples (86 %; p = 0.001 (Median)) – implying a larger clonal size.

Table 15. Comparison of $CD138^+$ cell-sorting and the main clone in the FISH analysis of AL amyloidosis patients. The highest measured percentage of a genetic aberration in the FISH result was defined as the proportion of the main clone (Bochtler et al. 2018). FISH = fluorescence *in situ* hybridization, PC = plasma cells, AL_H = AL amyloidosis patients with dominant heart involvement, AL_HK = AL amyloidosis patients with dominant heart and kidney involvement. More detailed information about the investigator, the sample material, applied tests, and the reference areas are listed in chapter 2.4 and Table 14.

	AL_H n = 61	AL_HK n = 21
Median bone marrow aspirate volume [mL] (range)	52 (8 – 95)	60 (10 – 83)
Median percentage of $CD138^+$ PC in the sample (range)	0.5 % (0 – 7)	0.5 % (0 – 12)
Number of cell sorts with no harvested $CD138^+$ PC n (%)	7 (4 %)	0
Median percentage of the main clone in the FISH analysis (range)	90 % (43 – 98 %)	86 % (10 – 97 %)

When analyzing the clinical characteristics, only a few similarities can be noted between the cohorts (**Table 16**). AL_H as well as AL_HK and MM patients displayed a median age at diagnosis of about 64 – 65 years and most of the samples were obtained at diagnosis (AL_H = 89 %, AL_HK = 86 %, MM = 100 %). As a side note, five AL patients (3x AL_H, 2x AL_HK) also had a MM diagnosis (AL+MM) at the time point of bone marrow aspiration.

Overall, the AL_H and MM subgroups included more male than female patients (AL_HK = male: 11, female: 10) which corresponds with the general epidemiology of the diseases. The AL_H and MM cohorts also displayed a comparable median dFLC value, twice as high as the median dFLC value of the AL_HK subgroup (AL_H = 406 mg/L, AL_HK = 169 mg/L, MM = 411 mg/L). While the AL_H and AL_HK subgroups displayed a comparable median plasma cell infiltration in the bone marrow of about 12 %, the MM subgroup showed a significantly higher median value of about 60 % (AL_H vs. MM p = 3.00e-12 (Median), AL_HK vs. MM p = 0.000042 (Median)). Regarding proteinuria, the

AL_HK patients (3 g/d) displayed a significantly higher median value than the AL_H (0.3 g/d; p = 2.00e-06 (Median)) and MM cohort (1 g/d; p = <0.01 (U-test)). In contrast, the AL_H patients (91 U/L) displayed a higher AP value than the AL_HK (68 U/L) and MM patients (70 U/L). No information about the GFR CKD-EPI, NT-proBNP, and hsTNT values of the MM patients were available. However, the AL_H patients displayed a higher median NT-proBNP (7202 mg/L vs. 3361 mg/L) and GFR CKD-EPI value (71 mL/min/1.73 m² vs. 62 mL/min/1.73 m²) but a lower hsTNT value (70 µg/L vs. 76 µg/L) than AL_HK patients.

A difference between the two diseases was also noted regarding the presence of a M-gradient, which was detected significantly more often in the MM than in AL patients (AL_H = 42 % p = <0.001 (Fischer), AL_HK = 52 % p = <0.001 (Fischer), MM = 87 %) (**Table 16**). Additionally, the AL_HK and MM samples corresponded more often to a patient with detectable clonal HC in serum (AL_H = 44 %; AL_H vs. MM p = <0.001 (Fischer), AL_HK = 67 %, MM = 79 %). These HCs were identified as IgG in more than half of the cases for all subgroups (AL_H = 22/27, AL_HK = 8/14, MM = 33/41). The genetic aberration t(11;14), which is associated with a loss of the HC, was detected in 38 % of AL_H patients, 48 % of AL_HK patients, and 28 % of MM patients. A gain of 1q21 was detected less frequently in the MM than in the AL_H subgroup (AL_H = 37 %, AL_HK = 33 %, MM = 29 %).

Table 16. Overview of selected clinical characteristics of 82 AL amyloidosis and 52 multiple myeloma patients. AL_H = AL amyloidosis patients with dominant heart involvement, AL_HK = AL amyloidosis patients with dominant heart and kidney involvement. AL = AL amyloidosis patients, MM = multiple myeloma patients, LC = light chain, HC = heavy chain, IFE = immunofixation electrophoresis in serum, NA = not available, AL+MM = AL amyloidosis patients who were also diagnosed with multiple myeloma, GFR CKD-EPI = calculation of the glomerular filtration rate using the formula established by the Chronic Kidney Disease Epidemiology Collaboration, AP = alkaline phosphatase, dFLC = difference between disease-associated und uninvolved circulation free light chains, hsTNT = heart associated troponin T. More detailed information about the investigators, the sample material, applied tests, and reference area are listed in chapter 2.4 and Table 14.

	AL n = 82		MM n = 52
	AL_H n = 61	AL_HK n = 21	
Median age, [y] (range)	65 (34 – 86)	64 (51 – 83)	61 (37 – 70)
Sex female/male, n	21/40	10/11	22/30
New-diagnosis, n	54	18	52
Relapse/late relapse/progress, n	7	3	0
AL+MM, n (NA)	3 (6)	2 (1)	0 (0)
Median plasma cell infiltration [%] (range)[1]	12 (1 – 56)	12 (3 – 55)	60 (5 – 100)
Median dFLC at diagnosis [mg/L] (range)	406 (33 – 5322)	169 (48 – 1065)	411 (2 – 16344)
Median proteinuria [g/d] (range)[2]	0.3 (0.1 – 5)	3 (1 – 14)	1 (0 – 13)

	AL n = 82		MM n = 52
	AL_H n = 61	AL_HK n = 21	
Median GFR CKD-EPI [mL/min/1.73 m²] (range)	71 (24 – 106)	62 (17 – 101)	NA
Median NT-proBNP, serum [ng/L], (range)	7202 (288 – 31246)	3361 (854 – 21626)	NA
Median hsTNT [µg/L] (range)	70 (11 – 1227)	76 (11 – 415)	NA
Median AP [U/L] (range)	91 (38 – 233)	68 (41 – 173)	70 (32 – 247)
M-gradient, n, yes (NA)[3]	26 (5)	11 (2)	45 (0)
Serum HC present in IFE, n (%)[4]	27 (44 %)	14 (67 %)	41 (79 %)
Serum HC present in IFE NA, n	3	0	0
IgG, n	22	8	33
IgA, n	4	5	7
IgM, n	0	0	1
IgD, n	0	1	0
IgG and IgD, n	1	0	0
Gain 1q21, n (NA)	23 (13)	7 (4)	10 (0)
t(11;14) n, (NA)	23 (13)	10 (3)	15 (0)

[1]AL_H vs. MM p = 3.00e-12, AL_HK vs. MM p = 0.00004; [2]AL_HK vs. AL_H p = 2.00e-06, AL_HK vs. MM p = <0.01; [3]AL_H vs. MM p = <0.001, AL_HK vs. MM p = <0.001; [4]AL_H vs. MM p = <0.001

3.2 Multiple Myeloma Sample Composition

Since the MM data was generated via a bulk RNA sequencing approach, the sequence composition of these samples was also analyzed (**Table 17**). This analysis was not performed for the AL sequences because Sanger sequencing is not suitable for unambiguously differentiating subsequences and especially not for providing an associated percentage. Concerning the MM data, only additional sequences that were detected in ≥1 % were considered. It should be mentioned that in some cases several sequence sections were given due to the bioinformatic analysis. It is assumed that these sections result from variations at single positions and therefore lead to different sequence sections. This phenomenon will be addressed for each IGLV subfamily separately.

The sequence with the highest percentage (sequence_A) showed a median percentage of 98.85 % (52.46 % – 99.98 %) (**Supplementary Information Table 2**). However, in 16 cases it was possible to detect a second abundant sequence (sequence_B) with a median percentage of 23.54 % (1.23 % – 47.38 %). When a sequence_B was detected, the corresponding sequence_A showed a lower median percentage of 74.86 % (52.46 % – 98.13 %) (p = <0.001 (Median)).

When analyzing the sequences concerning the IGLV family usage, the samples with detectable sequence_B cluster to the IGLV1 family, especially to IGLV1-44 (9/9) (**Table 17**). In general, it was possible to detect a sequence_B in 64 % of the IGLV1 cases (9/14), 24 % of the IGLV2 cases (4/17), and 6 % of the IGLV3 cases (1/17) (**Table 17**, **Supplementary Information**

Table 2). Besides the higher prevalence, the IGLV1 family displayed also a higher percentage of the corresponding sequence_B (IGLV1 = 25.98 %, IGLV2 = 4.28 %, IGLV3 = 1.23 %). This reflects in a lower median percentage of the respective sequence_A (IGLV1 = 78.05 %, IGLV2 = 98.85 %, IGLV3 = 99.41 %). Two sequences were not assigned unambiguously to an IGLV family (sequence_A = 74.57 % and 75.16 %) and displayed a sequence_B with an especially high percentage (25.3 % and 21.79 %).

Table 17. Sequence composition of multiple myeloma samples based on the bulk RNA sequencing results. The sequence composition was analyzed concerning the detected IGLV families in the multiple myeloma cohort (n = 52). sequence_A = sequence with the highest percentage >1 %, sequence_B = sequence with the second highest percentage >1 %.

	n	Median percentage sequence_A (range)	Percentage of sequences with a sequence_B (n)	Median percentage of sequence_A with sequence_B (range)	Median percentage sequence_B (range)
IGLV1	14	78.05 (52.46 – 99.95)	64 (9/14)	73.61 (52.46 – 97.32)	25.98 (2.52 – 57.38)
IGLV2	17	98.85 (93.68 – 99.87)	24 (4/17)	94.61 (93.68 – 95.67)	4.28 (3.56 – 6.02)
IGLV3	17	99.41 (98.08 – 99.98)	6 (1/17)	98.13	1.23
IGLV6	1	99.83	0	0	0
IGLV7	1	99.78	0	0	0
NA	2	74.57, 75.16	100 (2/2)	74.57, 75.16	25.30, 21.79

When analyzing sequence_B in more detail, eight out of nine IGLV1-44 sequences displayed the same IGLV, IGLJ, and IGLC segment composition as the corresponding sequence_A. One sequence_B was not clearly assigned to IGLV1-47 or IGLV1-44 but showed the same IGLJ and IGLC subfamilies as the corresponding sequence_A. Regarding the IGLV2 family, one IGLV2-14 and one IGLV2-23 sequence displayed the same composition for sequence_A and sequence_B. However, it was not possible to assign any IGL segment due to several disconnected sequence sections two times (sequence_A = IGLV2-23). The IGLV3 sequence showed an IGLV3-21/IGLJ2/IGLC2 linkage for both sequences.

Only sequence_A was used in the following detailed IGLV subfamily analysis and comparison with the AL cohort.

3.3 IGL Family Distribution

3.3.1 IGLV Family and Subfamily Distribution

To address the question, if some IGLV families display a higher amyloidogenic potential than others, a comparison between the AL and MM cohort was performed. In general, the IGLV3 family was identified as the most common AL IGLV subfamily (35 %) and one of the two most common MM IGLV families (33 %). IGLV2 was the second most common AL IGLV subfamily (AL = 22 %) and the other most common MM IGLV subfamily (MM = 33 %) (**Figure 11** A). A difference was noted regarding the third most common IGLV family: the MM cohort displayed IGLV1 (MM = 27 %, AL = 11 %; p = <0.001 (Fischer)), and the AL cohort IGLV6. Of note, IGLV6 was detected in 21 % of the AL cases and only once in the MM cohort (2 %; p = <0.001 (Fischer)).

Overall, two rare IGLV families were exclusively detected in one of the cohorts: the IGLV7 family was assigned only once in the MM cohort (2 %), and the IGLV8 family once in the AL cohort (1 %; AL_HK). Moreover, the AL IGLV8 sequence corresponded to one of the five AL patients who were also diagnosed with MM (FOR229_HK). In addition, one AL+MM sequence each was assigned towards IGLV2 (IGLV2-14, FOR159), IGLV1 (IGLV1-44, FOR147), IGLV6 (FOR188_HK) and IGLV3 (IGLV3-1, FOR123). Due to this equal distribution, the AL+MM subgroup was not investigated further (

Supplementary Information Table 3).

However, the IGLV family distribution was not only analyzed between AL and MM but also concerning the AL organ tropism. In this context, a difference was also noted regarding the IGLV6 family (**Figure 11** B). IGLV6 was assigned in 38 % of the AL_HK cases and significantly less often in AL_H cases (15 %; p = 0.032 (Fischer)). Contrastingly, IGLV3 was detected more often in the AL_H than AL_HK subgroup (43 % vs. 14 %; p = 0.032 (Fischer)). IGLV2 (AL_H = 21 %, AL_HK = 24 %) and IGLV1 (AL_H = 11 %, AL_HK = 10 %) were detected in both AL subgroups in a comparable frequency.

Since not only IGLV families, but also IGLV subfamilies can be differentiated from each other, this was investigated in the next step. Here, the AL cohort displayed IGLV6-57 and the MM cohort IGLV3-21 (21 %, AL = 13 %) as the most common IGLV subfamily (**Figure 11** C). The difference concerning the IGLV6 family has already been addressed since IGLV6-57 is the only detected IGLV6 subfamily (AL = 21 %, MM = 2 %; p = <0.001 (Fischer)). IGLV2-14 was detected as one of the most common IGLV subfamilies in both diseases: AL second 20 %, MM third 15 %. For the IGLV1-44 subfamily large deviations were observed between AL and MM.

This family was detected in 6 % of the AL cohort and as the second most common MM IGLV subfamily (17 %). This was reversed in terms of IGLV3-1, which was detected as the third most common AL subfamily (18 %) and in only 6 % of the MM cohort. Furthermore, the IGLV2-23 subfamily was detected only in the MM cohort (p = <0.001 (Fischer)) and – shared with IGLV2-14 – as the third most common IGLV subfamily (15 %).

Regarding the AL organ tropism, a difference was noted regarding IGLV3-21, which was identified as the third most common AL_H IGLV subfamily with an overall three times higher percentage than in the AL_HK subgroup (15 % vs. 5 %) (**Figure 11** D). Besides the already addressed significantly higher IGLV6-57 AL_HK than AL_H association, it is noticeable that four out of eight AL_HTX sequences were assigned to this subfamily. However, the AL_HTX category does not describe an organ involvement, but a certain treatment option for a severe heart manifestation. Further, IGLV3-1 can be described as the most detected AL_H (21 %) and third most detected AL_HK IGLV subfamily (10 %). Both subgroups share IGLV2-14 as the second most used IGLV subfamily (AL_H = 20 %, AL_HK = 19 %).

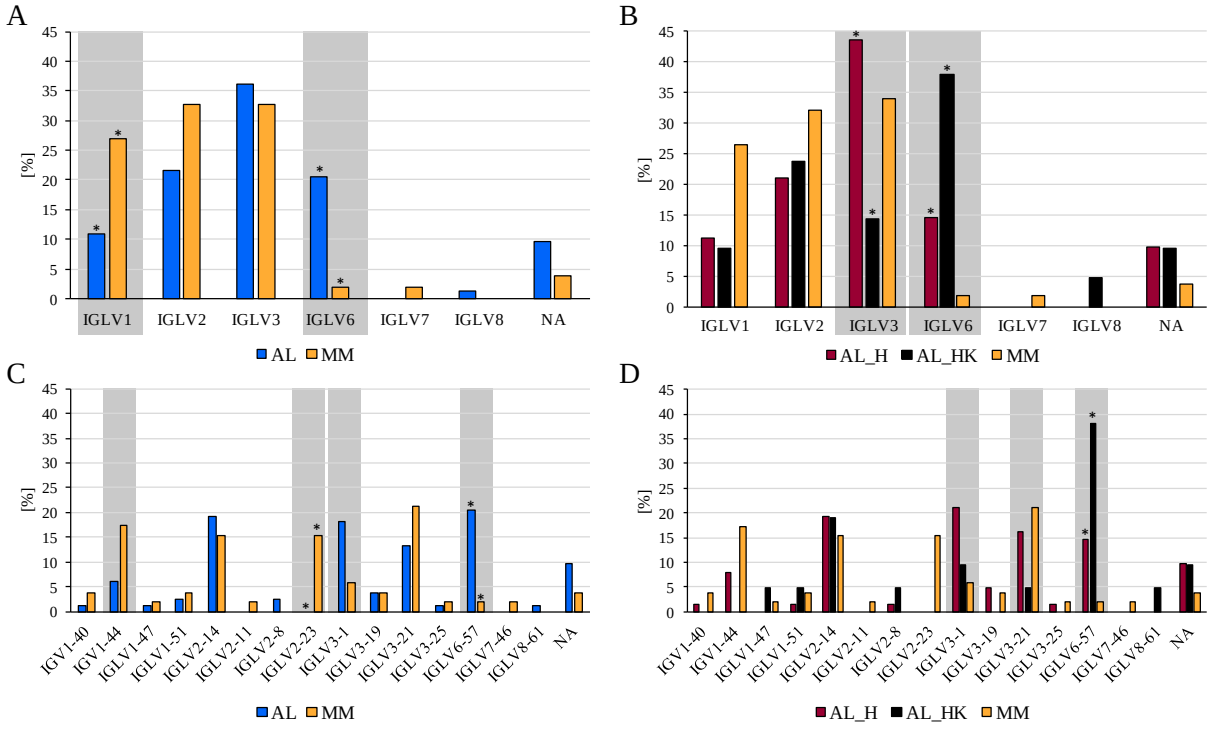


Figure 11. Comparison of the IGLV family and subfamily distribution between the AL amyloidosis and the multiple myeloma cohort as well as between the heart and heart and kidney organ tropism in AL amyloidosis. A) Comparison of the IGLV family distribution between the AL amyloidosis and the multiple myeloma cohort. B) Comparison of the IGLV family distribution between the AL amyloidosis heart and heart and kidney organ tropism as well as the multiple myeloma cohort. C) Comparison of the IGLV subfamily distribution between the AL amyloidosis and the multiple myeloma cohort. D) Comparison of the IGLV subfamily distribution between the AL amyloidosis heart and heart and kidney organ tropism as well as the multiple myeloma cohort. AL = AL amyloidosis patients, MM = multiple myeloma patients, AL_H = AL amyloidosis patients with dominant heart involvement, AL_HK = AL amyloidosis patients with dominant heart

and kidney involvement, NA = not clearly assigned IGLV family, * = p ≤0.05; AL n = 82, MM n = 52, AL_H n = 61, AL_HK n = 21.

To summarize, in the analysis concerning the IGLV family distribution, a prominent difference was noted regarding IGLV6-57. This IGLV family was detected more frequently in AL than MM and AL_HK than AL_H. In contrast, IGLV1 was detected more frequently in the MM than AL cohort, and IGLV3-1 was more frequently in the AL_H than AL_HK cohort. Finally, IGLV2-23 was only detected in the MM cohort.

Parenthesis: An additional analysis was performed to evaluate the effect of a potential HC binding partner and the size and malignancy of the B cell clone. In this context, the sequences were stratified concerning a present clonal heavy chain in serum and a dFLC >180 mg/L in the respective patients (Figure 12). This stratification was performed for IGLV subfamilies that were most frequently associated with AL and to which more than 10 AL sequences were assigned. This concern: IGLV2-14 (AL n = 16, MM n = 8), IGLV3-1 (AL n = 15, MM n = 3), IGLV3-21 (AL n = 10, MM n = 11) and IGLV6-57 (AL n = 17, MM n = 1). In general, the sub-analyses were performed, when more than one sequence can be assigned to one of the respective subgroups.

When stratifying the AL cases for the presence (HC) or absence of a clonal HC (no HC), the IGLV2-14 and IGLV6-57 subfamily displayed an association with AL HC cases (IGLV2-14: 29 % vs. 11 %; IGLV6-57: 27 % vs. 13 %) (Figure 12 A). Interestingly, a certain organ tropism was noted – for AL_H IGLV2-14 sequences an HC association was detected with a six times higher percentage than for no HC cases (37 % vs. 6 %; p = 0.051 (Fischer)). In contrast, AL_HK IGLV2-14 sequences were associated towards no HC cases twice as often than towards HC cases (29 % vs. 14 %). Regarding the IGLV6-57 family, an AL_HK HC association was detected in 50 % of the cases, and only in 14 % of AL_HK no HC cases. However, a comparable percentage was noted in the AL_H subgroup (AL_H HC = 15 %, AL_H no HC = 13 %) (Figure 12 B). Concerning the MM cohort, the MM IGLV3-21 and IGLV2-14 sequences displayed an association with HC cases (IGLV3-21: 24 % vs. 9 %; IGLV2-14: 17 % vs. 9 %), which was similar to the AL cohort.

When analyzing the sequences with respect to a dFLC >180mg/L, no similarities between the two diseases were noted (Figure 12 C). The AL IGLV3-1 and IGLV3-21 sequences were more frequently detected for LCs corresponding to patients with a dFLC >180 mg/L (> dFLC) (IGLV3-1: 27 % vs. 4 %; p = 0.014 (Fischer); IGLV3-21: 14 % vs. 7 %). Contrastingly, AL

IGLV6-57 sequences displayed an association with a dFLC <180 mg/L (< dFLC) (14 % vs. 33 %). Again, a difference regarding the organ tropism was noted – the AL_H IGLV6-57 sequences showed an < dFLC association (33 % vs. 7 %; p = 0.024 (Fischer)), and the AL_H IGLV2-14 and IGLV3-1 sequences towards > dFLC (IGLV2-14 = 24 % vs. 13 %; IGLV3-1 = 29 % vs. 7 %) (Figure 12 D). In contrast, AL_HK IGLV2-14 sequences showed a < dFLC association (11 % vs. 25 %).

Addressing the MM cohort, IGLV1-44 sequences showed an association towards > dFLC cases (26 % vs. 0 %). This was also the case for MM IGLV2-14 sequences (20 % vs. 6 %), similar to the AL_H subgroup. For MM IGLV3-1 and IGLV3-21 sequences an association towards < dFLC cases was noted (IGLV3-1: 3 % vs. 13 %; IGLV3-21: 11 % vs. 38 %).

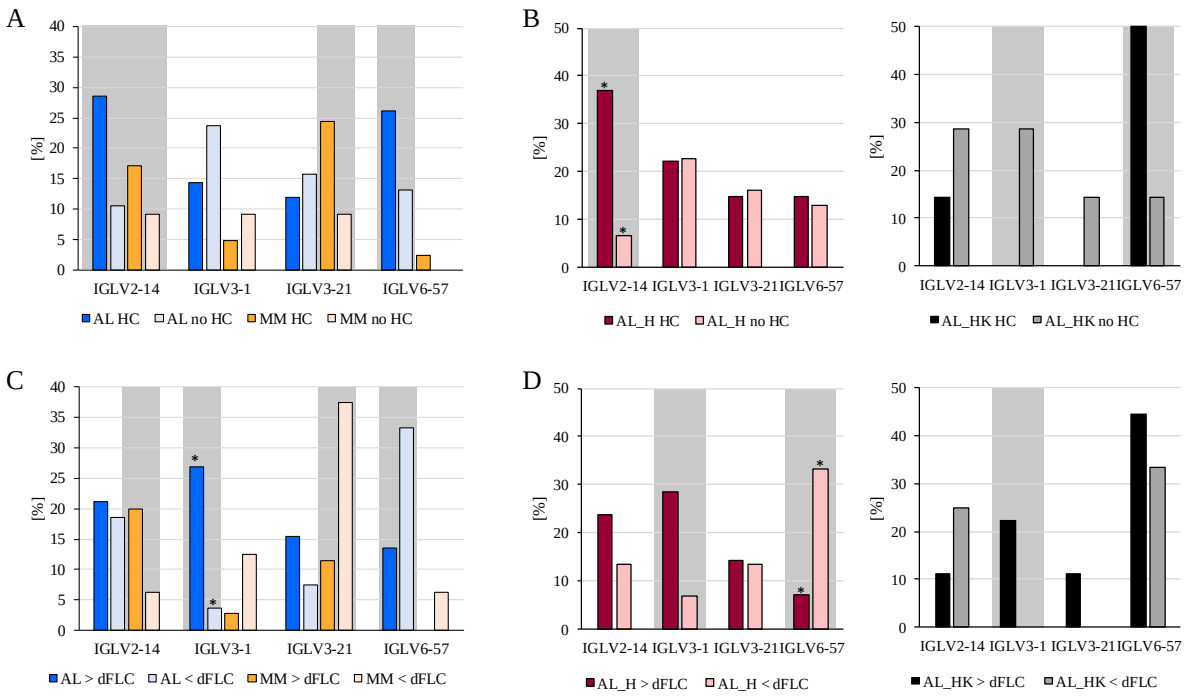


Figure 12. Comparison of the IGLV subfamily distribution between the AL amyloidosis and the multiple myeloma cohort and the presence or absence of a clonal heavy chain or a dFLC >180 mg/L. Analysis of the four IGLV subfamilies with at least ten assigned AL amyloidosis light chain sequences. A) Comparison of the IGLV subfamily distribution between the AL amyloidosis and the multiple myeloma cohort with respect to a present heavy chain. B) Comparison of the IGLV subfamily distribution between AL amyloidosis patients with dominant heart or dominant heart and kidney involvement with respect to a present heavy chain. C) Comparison of the IGLV subfamily distribution between the AL amyloidosis and the multiple myeloma cohort with respect to a dFLC >180 mg/L. B) Comparison of the IGLV subfamily distribution between AL amyloidosis patients with dominant heart or dominant heart and kidney involvement with respect to a dFLC >180 mg/L. AL = AL amyloidosis patients, MM = multiple myeloma patients, HC = detectable clonal heavy chain in the immunofixation electrophoresis in serum, > dFLC = dFLC >180 mg/L, < dFLC = dFLC <180 mg/L; AL_H = AL amyloidosis patients with dominant heart involvement, AL_HK = AL amyloidosis patients with dominant heart and kidney involvement, * = p ≤0.05, AL HC n = 41, AL no HC n = 38, AL > dFLC n = 51, AL < dFLC n = 27, MM HC n = 41, MM no HC n = 11, MM > dFLC n = 35, MM < dFLC n = 16, AL_H HC n = 27, AL_H no HC n = 31, AL_H > dFLC n = 42, AL_H < dFLC n = 15, AL_HK HC n = 14, AL_HK no HC n = 7, AL_HK > dFLC n = 9, AL_HK < dFLC n = 12.

To summarize, IGLV2-14 was significantly more frequently detected in AL_H HC than in AL_H no HC sequences. Concerning the dFLC, IGLV3-1 was significantly more frequent detected in AL > dFLC than AL < dFLC sequences, and IGLV6-57 was significantly more frequent in AL_H < dFLC than AL_H > dFLC sequences.

3.3.2 IGLJ and IGLC Family Distribution

Not only the IGLV distribution but also the composition of the complete LC was evaluated and, in this context, also the IGLJ/IGLC linkage. When considering only the unambiguously determined IGLJ and IGLC segments for both diseases, IGLJ1, IGLJ2, and IGLJ7 were only detected in connection with the corresponding IGLC segment (**Table 18**). In contrast, IGLJ3 was detected in linkage with both IGLC2 and IGLC3. As a side note, when additionally analyzing the AA sequences with the IMGT/DomainGapAlign tool IGLC2*01, IGLC2*02 and IGLC3*04 are all indicated with a 100 % identity.

In general, both diseases displayed an IGLV-IGLJ2/IGLC2 linkage in most of the cases (AL = 38 %; MM = 42 %), second most common towards IGLJ3/IGLC3 (AL = 22 %, MM = 13 %) and third most common towards IGLJ1/IGLC1 (AL = 13 %; MM = 10 %) (**Table 18**). This order was also detected when stratifying the AL sequences with respect to the organ involvement or a present potential HC binding partner, as well as for the MM HC and MM dFLC subgroups. A difference was noted in the AL < dFLC sequences, which presented an IGLJ2/IGLC2 and IGLJ3/IGLC3 linkage with the same frequencies (27 %). In contrast, for AL > dFLC sequences an IGLJ2/IGLC2 linkage was observed in 45 % of the cases and IGLJ3/IGLC3 in only 18 %.

Table 18. IGLJ and IGLC family linkage and distribution between the AL amyloidosis and multiple myeloma cohort. Only detected linkages are shown. AL = AL amyloidosis patients, MM = multiple myeloma patients, AL_H = AL amyloidosis patients with dominant heart involvement, AL_HK = AL amyloidosis patients with dominant heart and kidney involvement, HC = detectable clonal heavy chain in the immunofixation electrophoresis in serum, > dFLC = dFLC >180 mg/L, < dFLC = dFLC <180 mg/L, J = IGLJ, C = IGLC, NA = not clearly assigned, bold = most common linkage. For three AL amyloidosis patients no information about the heavy chain was available, as well as for five AL amyloidosis and one multiple myeloma patients about the dFLC.

IGLJ	1	2		3			2/3			7	7/3	NA
IGLC	1	2	2_3	2	3	2_3	2	3	2_3	7	2	
AL n = 82 [%]	13	**38**	0	7	22	0	2	5	0	1	1	10
AL_H n = 61 [%]	16	**34**	0	8	18	0	3	7	0	2	2	10
AL_HK n = 21 [%]	5	**48**	0	5	33	0	0	0	0	0	0	10
MM n = 52 [%]	10	**42**	8	2	13	8	6	4	2	2	0	4
AL HC n = 41 [%]	15	**37**	0	5	24	0	2	7	0	2	2	5
AL no HC n = 38 [%]	13	**37**	0	11	16	0	3	3	0	0	0	16

IGLJ	1	2		3			2/3			7	7/3	NA
IGLC	1	2	2_3	2	3	2_3	2	3	2_3	7	2	
MM HC n = 41 [%]	10	**41**	7	2	15	7	7	5	0	0	0	5
MM no HC n = 11 [%]	9	**46**	9	0	9	9	9	0	9	9	0	0
AL > dFLC n = 51 [%]	14	**45**	0	6	18	0	2	4	0	2	2	8
AL < dFLC n = 26 [%]	15	**26**	0	11	**26**	0	0	7	0	0	0	15
MM > dFLC n = 36 [%]	8	**36**	8	3	14	8	8	3	3	3	0	6
MM < dFLC n = 16 [%]	13	**50**	6	0	13	6	6	6	0	0	0	0

To summarize, an IGLJ2/IGLC2 connection was detected as the most common linkage in almost all cohorts and subgroups. The only exception was noted for the AL > dFLC subgroup which displayed an IGLJ2/IGLC2 and IGLJ3/IGLC3 linkage equally frequent.

3.3.3 Composition of the IGLV, IGLJ, and IGLC Segments

In the next step, not only the IGLJ/IGLC linkage but also the connection with an IGLV segment was investigated. Again, the five most common AL IGLV subfamilies were analyzed.

Regarding the IGLV2-14 subfamily, the AL (6/16) and MM (5/8) sequences displayed a linkage with IGLJ2/IGLC2 most often (**Table 19**). In addition, the AL_H sequences showed an IGLV2-14 IGLJ1/IGLC1 linkage in five out of twelve cases and none of the AL_HK sequences (n = 4). In contrast, the AL_HK sequences presented more often an IGLJ2/IGLC2 linkage than the AL_H sequences (3/4 vs. 3/12). When stratifying for a present clonal HC or dFLC, the IGLV2-14 AL HC and > dFLC sequences showed an IGLJ1/IGLC1 linkage more frequently than the respective other subgroup (both = 5/12 vs. 1/4). Regarding the MM dFLC subgroups, no pattern was observed and only one MM sequence corresponded to the MM no HC subgroup.

In the IGLV3-1 subfamily, an IGLJ2/IGLC2 linkage was also overserved with the highest frequency in the AL cohort (9/15) with no deviation when stratifying for dominant organ involvement. However, the MM sequences displayed an IGLJ1/IGLC1 linkage in two out of three cases (**Table 19**). In the analysis regarding a clonal HC, the AL no HC sequences showed an IGLJ2/IGLC2 linkage more often than the AL HC subgroup (6/8 vs. 2/6). An analysis regarding the dFLC was not performed since only one AL sequence corresponded to the < dFLC subgroup.

In comparison to the other subfamilies, a difference regarding the most common IGLJ/IGLC linkage was noted in the IGLV3-21 subfamily (**Table 19**). The corresponding AL (n = 10, AL_HK n = 1) sequences displayed a linkage to IGLJ2/IGLC2 and IGLJ3/IGLC3 in four cases each, while the MM IGLV3-21 sequences showed a prominent linkage to IGLJ2/IGLC2 (7/11). While the AL HC sequences more often showed an IGLJ2/IGLC2 linkage (3/5 vs. 1/6), an IGLJ3/IGLC3 linkage was detected more frequently in the AL no HC sequences (0/4 vs. 4/6).

Only two AL sequences corresponded to the < dFLC subgroup but both showed an IGLJ2/IGLC2 linkage. Regarding the MM dFLC subgroups, no pattern was observed, and only one MM sequence corresponded to the MM no HC subgroup.

A generally preferred IGLJ2/IGLC2 linkage was also noted in the AL IGLV6-57 sequences (8/17; MM n = 1 IGLJ3/IGC3) (**Table 19**) and no difference was noted regarding the organ tropism. However, a higher frequency of IGLJ3/IGLC3 was observed for the AL HC compared to the AL no HC subgroup (4/11 vs. 0/6) as well as a more preferred IGLJ2/IGLC2 linkage in the AL > dFLC than the AL < dFLC subgroup (5/9 vs. 3/9).

Table 19. Overview of the five most common IGLV segments and the IGLJ/IGLC linkages between the AL amyloidosis and multiple myeloma cohort. Only detected linkages are shown. AL = AL amyloidosis patients, MM = multiple myeloma patients, NA = not clearly assigned, bold = most common linkage.

IGLJ	1	2		3			2_3			7	7_3	NA
IGLC	1	2	2_3	2	3	2_3	2	3	2_3	7	2	
AL n = 82, n	11	**31**	0	6	18	0	2	4	0	1	1	8
MM n = 52, n	5	**22**	4	1	7	4	4	2	1	1	0	2
IGLV2-14												
AL n = 16, n	5	**6**	0	0	3	0	0	2	0	0	0	
MM n = 8, n	1	**5**	0	0	0	2	0	0	0	0	0	-
IGLV3-1												
AL n = 15, n	3	**9**	0	1	1	0	0	0	0	1	0	-
MM n = 3, n	**2**	1	0	0	0	0	0	0	0	0	0	-
IGLV3-21												
AL n = 10, n	0	**4**	0	0	**4**	0	1	1	0	0	0	-
MM n = 11, n	1	**7**	0	0	1	1	1	0	0	0	0	-
IGLV6-57												
AL n = 17, n	0	**8**	0	3	5	0	0	1	0	0	0	-
MM n = 1, n	0	0	0	0	1	0	0	0	0	0	0	-

When analyzing the AL IGLJ1/IGLC1 (AL n = 11, MM n = 5) frequency separately, a linkage was most noted with IGLV2-14 (5/11) and IGLV3-1 (3/11; MM = 2/3). An AL IGLJ2/IGLC2 linkage was most detected with IGLV3-1 (9/31) and IGLV6-57 (8/31). The MM IGLJ2/IGLC2 sequences displayed most frequently a linkage with IGLV3-21 (7/22) and IGLV2-14 (5/22). The AL IGLJ3/IGLC3 linkage was assigned in 5/18 cases towards IGLV6-57 and 4/10 cases towards IGLV3-21. Regarding the IGLJ3/IGLC3 MM sequences, no favored linkage was noted.

To shortly summarize, for IGLV2-14 the MM and AL sequences presented most frequently and IGLJ2/IGLC2 linkage. This was also the case for the IGLV6-57 AL, IGLV3-1 AL and IGLV3-21 MM sequences. In contrast, the IGLV3-21 AL sequences presented equally frequent an IGLJ2/IGLC2 or an IGLJ3/IGLC3 linkage.

Parenthesis: In the following, the four IGLV subfamilies with at least ten assigned AL amyloidosis sequences were analyzed in a more detailed manner (IGLV2-14, IGLV3-1, IGLV3-21, IGLV6-57). In this context, the overall mutation frequency and count were addressed, and detailed analyses of sequence alignments were performed. This sequence analysis especially concerned positions that were mutated in at least 50 % of the cases. The impact of the detected mutations was then studied in terms of the overall amino acid composition and various biophysical parameters. In addition, subgroup analyses regarding the influence of a potential clonal heavy chain binding partner and the size and malignancy of the B cell clone, in the context of a dFLC >180 mg/L, as well as the AL organ involvement were performed.

3.4 IGLV2-14

IGLV2-14 was detected as the second most common AL subfamily and one of two third most common MM IGLV subfamilies. In general, this subfamily comprised twelve AL_H sequences (1x AL_HTX), four AL_HK sequences, and eight MM sequences (**Table 20**). The AL sequences FOR171 and FOR173 were not used for further analysis due to more than ten AAs that could not be clearly determined. These ambiguously defined AAs results from positions that showed a signal for two or more nucleotides in the Sanger sequencing; consequently, a blur was indicated at these positions.

An additional verification step was included since the AL and MM sequences were generated through different sequencing approaches. To guarantee comparability of the sequences four MM sequences (MM103, MM136, MM129, MM130) were additionally Sanger sequenced and the dominant signals were consistent with the respective sequences in all cases (**Supplementary Information Figure 8**, **Supplementary Information Figure 9**, **Supplementary Information Figure 10**, **Supplementary Information Figure 11**).

3.4.1 Subgroup Analysis

Most AL IGLV2-14 sequences corresponded to patients with a detectable clonal HC in serum (12/16, 11x IgG, 1x IgA) and/or with a dFLC >180 mg/L (11/16) (**Table 20**). Also, seven out of eight MM patients displayed a detectable clonal HC (5x IgG, 1x IgA, 1x IgM) and/or a dFLC >180 mg/L. In context of the presented HC classes, the distribution is well in line with the approximate percentage of the total Ig in the adult serum (Lefranc and Lefranc 2020). Due to the uniform clinical presentation of MM patients with an IGLV2-14 LC, subgroup analyses were only performed for the AL cohort.

Table 20. Overview of selected characteristics of 16 AL amyloidosis and 8 multiple myeloma patients with IGLV2-14 assigned light chain sequences. The column "X AA [n]" defines how many amino acids were not determined unambiguously. Organ inv. = dominant organ involvement of AL amyloidosis patients, AL_H = AL amyloidosis patients with dominant heart involvement, AL_HK = AL amyloidosis patients with dominant heart and kidney involvement, AL_HTX = AL amyloidosis patients who received a heart transplant, AA = amino acid, AL = AL amyloidosis, MM = multiple myeloma, HC in IFE = detectable clonal heavy chain in the immunofixation electrophoresis in serum.

Patient	Disease	Organ inv.	IGLJ	IGLC	X AA [n]	HC in IFE	dFLC [mg/L]
FOR101	AL	AL_H	IGLJ1	IGLC1	1	G	**996.7**
FOR122	AL	AL_H	IGLJ2	IGLC2	1	G	**993.6**
FOR124	AL	AL_H	IGLJ3	IGLC3	0	-	**510.4**
FOR155	AL	AL_H	IGLJ2/3	IGLC3	0	G	52.7
FOR157	AL	AL_H	IGLJ1	IGLC1	1	G	**303.6**
FOR159	AL	AL_H	IGLJ3	IGLC3	2	G	**196.4**
FOR171	AL	AL_H	IGLJ1	IGLC1	17	G	35.7
FOR173	AL	AL_H	IGLJ1	IGLC1	16	-	**405.7**
FOR196	AL	AL_HTX	IGLJ1	IGLC1	1	A	**240.3**
FOR201	AL	AL_H	IGLJ1	IGLC1	0	G	**5322.2**
FOR202	AL	AL_H	IGLJ2/3	IGLC3	0	G	**351.8**
FOR204	AL	AL_H	IGLJ2	IGLC2	5	G	**358.5**
FOR190	AL	AL_HK	IGLJ2	IGLC2	6	-	67.4
FOR220	AL	AL_HK	IGLJ2	IGLC2	8	G	**389.0**
FOR225	AL	AL_HK	IGLJ2	IGLC3	1	-	169.0
FOR230	AL	AL_HK	IGLJ2	IGLC2	0	G	157.1
P006	MM	-	IGLJ1	IGLC1	0	G	**635.9**
P009	MM	-	IGLJ3	IGLC3/2	0	M	**480.2**
P284	MM	-	IGLJ2	IGLC2	0	G	**1103.3**
P312	MM	-	IGLJ2	IGLC2	0	-	**1640.8**
P319	MM	-	IGLJ2	IGLC2	0	G	160.6
P361	MM	-	IGLJ2	IGLC2	0	G	**514.0**
P505	MM	-	IGLJ2	IGLC2	0	A	**350.7**
P563	MM	-	IGLJ3	IGLC3/2	0	G	**313.4**

3.4.2 Mutation Frequency and Count

To examine whether the difference in LC behavior is based on a difference in the general mutation distribution and frequency, these aspects were analyzed within the IGL segments.

The MM sequences presented a significantly higher mutational load in the IGLV segment than the AL sequences (11.5 vs. 8.5, p = 0.026 (Median), reference = 97 AA), but no prominent difference was observed regarding the IGLJ and IGLC segment (**Table 21**).

When analyzing the AL sequences for the organ tropism, a higher mutational load in the IGLV (9.5 vs. 7.5, p = 0.048 (Median)) and IGLJ segment (2.0 vs. 1.0, reference = 12 AA) was detected in the AL_HK sequences than the AL_H sequences (**Table 21**). This difference was especially prominent in FR1 (20 % vs. 75 %) and FR3 (3.0 vs. 1.0, reference = 32 AA). Interestingly, FR1 was detected as the most conserved region in the AL and MM cohort.

When stratifying the AL sequences for the presence or absence of a clonal HC binding partner, the IGLV (9.0 vs. 7.0) and IGLJ segments of AL HC sequences showed a mutation more often (82 % vs. 33 %) – this was particularly observed in the CDR regions (CDR1, CDR2, and CDR3 = 91 % vs. 67 %) (**Table 21**).

In the analysis regarding the dFLC, AL > dFLC sequences presented a higher mutational load in the IGLV segment (9.0 vs. 8.0), and all sequences with a mutation in the IGLC segment were assigned to this subgroup (18 % vs. 0 %). In addition, a higher mutation frequency in CDR1 was detected in the AL > dFLC sequences (91 % vs. 67 %) (**Table 21**).

Table 21. Comparison of the percentage of mutated IGL segments and average mutation count between AL amyloidosis and multiple myeloma IGLV2-14 assigned light chain sequences. The analysis was also performed with respect to different subgroups. AL = AL amyloidosis patients, MM = multiple myeloma patients, AL_H = AL amyloidosis patients with dominant heart involvement, AL_HK = AL amyloidosis patients with dominant heart and kidney involvement, CDR = complementary determining region, FR = framework region, AA = amino acid, HC = detectable clonal heavy chain in the immunofixation electrophoresis in serum, > dFLC = dFLC >180 mg/L, < dFLC = dFLC <180 mg/L, * = $p \leq 0.05$, differences greater than 30 % and one (segments) or two (regions) additional mutations are marked in bold. FOR171, FOR173, and FOR201 were excluded, due to several ambiguously defined amino acids. The median mutation values were calculated using only the mutated segments/regions. The significance level was determined using the complete dataset. The Vbase2 IGLV2-14 reference sequence was used for calculations and the IGLC reference was trimmed at the C-terminus. The CDR3 region includes also the patient-specific linker region and the first two amino acids of the IGLJ segment and therefore spans 9-12 amino acids.

			mutated segments [%] median mutation count [n]								
		n	IGLV 97 AA							IGLJ 12 AA	IGLC 77 AA
				FR1 22 AA	CDR1 14 AA	FR2 15 AA	CDR2 7 AA	FR3 32 AA	CDR3		
AL	[%]	14	100	36	86	86	86	79	86	71	14
	[n]		**8.5***	1.0	1.0	2.0	1.5	1.0	2.5	1.0	1.0
AL_H	[%]	10	100	**20**	90	80	90	80	80	**80**	10
	[n]		**7.5***	1.0	1.0	2.0	1.0	**1.0**	3.0	**1.0**	1.0
AL_HK	[%]	4	100	**75**	75	100	75	75	100	**50**	25
	[n]		**9.5***	1.0	1.0	2.0	2.0	**3.0**	2.0	**2.0**	1.0
MM	[%]	8	100	25	88	88	88	100	100	75	25
	[n]		**11.5***	1.0	2.0	2.0	1.0	2.0	3.5	1.0	1.0
AL HC	[%]	11	100	36	**91**	82	91	82	91	**82**	18
	[n]		**9.0**	1.0	**1.0**	2.0	1.5	1.0	**3.0**	1.0	**1.0**
AL no HC	[%]	3	100	33	**67**	100	67	67	67	**33**	0
	[n]		**7.0**	1.0	2.0	2.0	1.5	2.0	**1.5**	1.0	**0**
MM HC	[%]	7	100	29	86	86	86	100	100	71	14
	[n]		**11.0**	1.0	**2.5**	2.5	1.0	2.0	3.0	1.0	1.0
AL > dFLC	[%]	11	100	36	91	82	82	82	82	73	18
	[n]		**9.0**	1.0	1.5	2.0	1.0	1.0	3.0	1.0	**1.0**
AL < dFLC	[%]	3	100	33	67	100	100	67	67	67	0
	[n]		**8.0**	1.0	1.0	2.0	2.0	1.5	2.0	1.5	**0**
MM > dFLC	[%]	7	100	14	86	86	86	100	100	71	29
	[n]		**11.0**	1.0	2.5	2.0	1.0	2.0	3.0	1.0	1.0

3.4.3 Sequence Alignment

A more detailed analysis of the LC sequences was performed on AA level to identify potential unique mutations or mutation patterns that could influence the biochemical properties of the LC itself or discriminate the subgroups. This was carried out in a step-wise procedure. First, general sequence characteristics such as mutation hotspots (mutations >50 %) and their location were identified separately for both diseases. Then the hotspots were analyzed in terms of an exchange towards a specific AA or AAs with the same side chain properties and the respective cDNA. In this analysis, the LC sequences with more than ten unambiguously determined AAs (marked as X) were included.

The higher overall mutation count in the MM sequences did also reflect in more mutation hotspots in the IGLV segment (9 vs. 3) (**Figure 13** A). The AL sequences only presented two hotspots, which were restricted to a CDR region and shared by the MM sequences: 54S (CDR2, AL = AL = 11/16, MM = 6/8) and 96S (CDR3, AL = 8/16, MM = 5/8). Additional MM hotspots were defined at the following positions: 29V (CDR1, 4/8), 31G (CDR1, 4/8), 49M (FR2, 6/8), 62N (FR3, 6/8), 94T (CDR3, 4/8), 95S (CDR3, 5/8), and 97S (CDR3, 6/8).

The shared mutation hotspot 54S (3rd position CDR2) is the most frequently mutated position (AL = 11/16, MM = 6/8) and both diseases showed an exchange towards threonine in most of the cases (AL = 7/11; MM = 4/6; cDNA = AGT→ACT) (**Figure 13** A). Additionally, a preferential exchange towards threonine was also noted for the second shared hotspots in the CDR3 (S96). In general, the 96S mutation was detected more frequently in AL_HK (3/4) than in AL_H sequences (5/12). Interestingly, the AL_HK sequences presented an S96T mutation in three out of four cases and the same sequences presented a mutation at the eighth (K) or ninth (L or V) position of the IGLJ segment towards methionine (1x X with methionine or leucine as possibilities). This was not detected in the AL_H or MM sequences.

In the following, the additional hotspots are analyzed chronologically (**Figure 13** A). At position 29V, the MM sequences presented exclusively a mutation towards isoleucine (MM = 4/8, AL = 3/16, 1x X; cDNA = GTT→ATT). Position 31G was described as a MM mutation hotspot (4/8) but was also found frequently mutated in the AL sequences (7/16). An exchange towards the negatively charged AA aspartic acid was detected twice in the MM sequences and three times in the AL sequences (cDNA = GGT→GAT). The 49M MM mutation hotspot (6/8) was also found frequently mutated in the AL sequences (7/16, 6x AL_H, 1x AL_HK), with a prominent exchange for both diseases towards isoleucine (AL = 5/7, MM = 4/6; cDNA = ATG→ATX). At position 62N, the MM sequences presented a mutation in 6/8 cases but with no directed exchange pattern. Interestingly, the AL sequences showed at this position 5/16 times an exchange towards the negatively charged AA aspartic acid

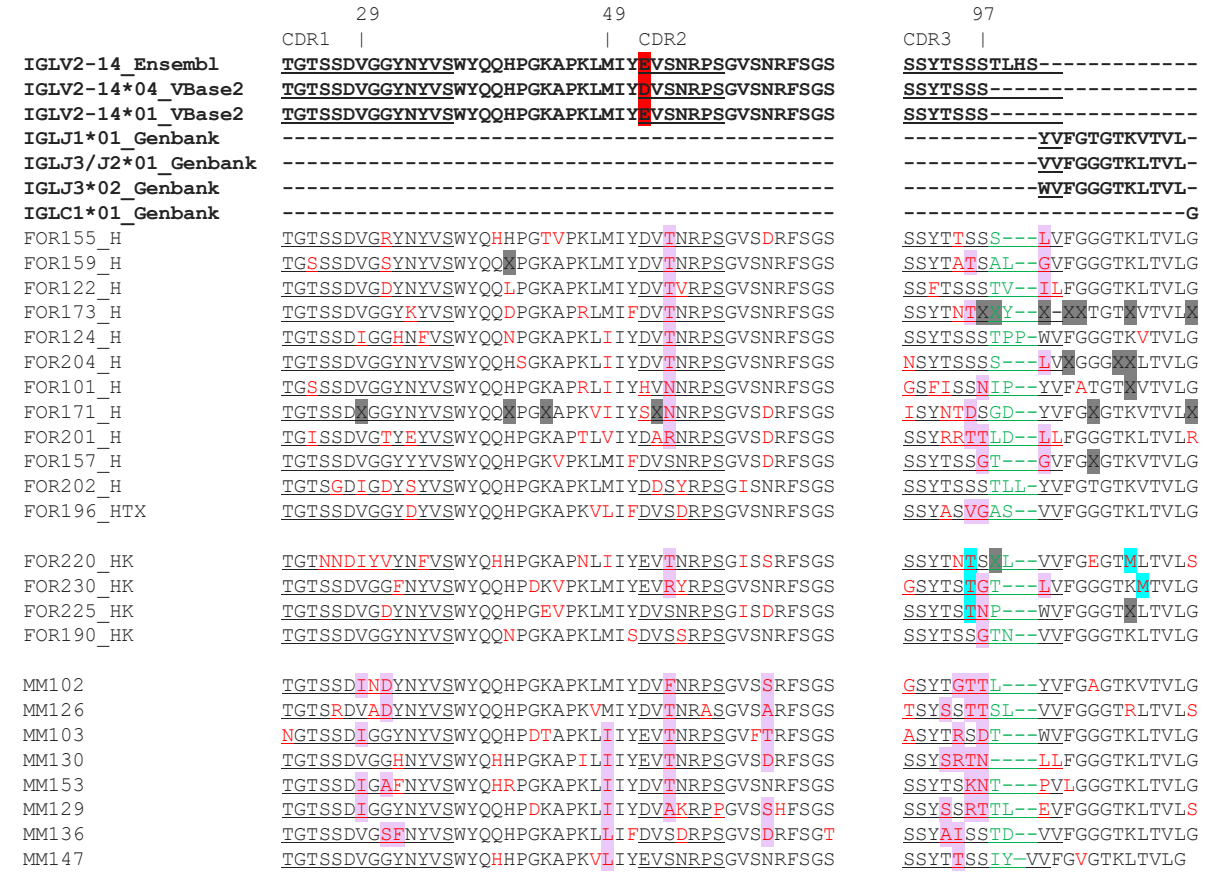
(MM = 2/6; cDNA = AAT→GAT). Regarding the CDR3 MM hotspot 94T, a most prominent exchange towards serine was noted (MM = 3/4, AL = 0/4; cDNA = exchange at the first nucleotide or the second nucleotide in the triplet). No exchange pattern was detected regarding the 95S MM mutation hotspot, but at position 97S (6/8, AL = 7/16 1x X) an exchange towards threonine was detected in three out of six cases.

For the sake of completeness, it should be mentioned that the MM sequences showed a L-13F (cDNA = CTC→TTC) mutation in the leader region in five cases, but due to the oligonucleotide binding site, this region was not covered by the AL sequences.

In general, no consistent IGLV-IGLJ linker region was noted despite an overlapping Ensembl IGLV2-14 reference (**Figure 13** A). Interestingly, the AL sequences presented a mutation at the first position of the IGLJ segment in seven out of sixteen cases (AL_H = 6/12, AL_HK = 1/4) and the MM sequences in three out of eight cases. For the AL sequences, a mutation towards leucine was detected four times and once in the MM sequences (**Figure 13** A).

When stratifying the sequences for a potential clonal HC binding partner, only one IGLV position is remarkable. Two out of three AL no HC sequences displayed a H41N mutation, which was not detected in any other sequence (**Figure 13** B). In addition, two AL no HC sequences presented additional prolines in the linker region (FOR124_H = TPP, FOR225_HK = P) – this was only detected once in the AL HC sequences (FOR101_H = IP). No obvious mutation pattern was detected when stratifying for a dFLC >180 mg/L.

A



B

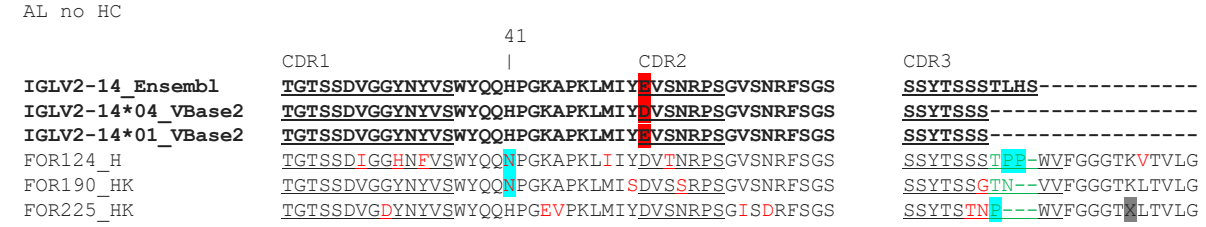


Figure 13. Sequence sections of IGLV2-14 assigned AL amyloidosis and multiple myeloma light chain sequences. A) Sequence sections of IGLV2-14 assigned AL amyloidosis and multiple myeloma light chain sequences B) Sequence sections of IGLV2-14 assigned AL amyloidosis light chain sequences corresponding to patients without detectable clonal heavy chain in serum. Bold = reference sequences, underlined = CDR regions, red highlight = discrepancy between the VBase2 and Ensembl IGLV2-14 reference, red letter = mutation, purple highlight = mutation hotspot, blue highlight = interesting mutation, X and grey highlight = not unambiguously determined amino acid, green letter = linker region, MM = multiple myeloma patients, AL = AL amyloidosis patients, _H = AL amyloidosis patient with dominant heart involvement, _HK = AL amyloidosis patient with dominant heart and kidney involvement, _HTX = AL amyloidosis patient who received a heart transplant. Only one IGLC reference is shown since the first amino acid can be defined as glycine in all cases. Amino acids were numbered according to the VBase2 reference. The complete amino acid sequence alignment is shown in Supplementary Information Figure 12. The complete cDNA sequence alignment is shown in Supplementary Information Figure 13.

Overall, it is noticeable that several regions showed specific mutation patterns (**Figure 13** A). The CDR1 region of AL sequences presented an additional charge in half of the cases (MM = 3/8) and in the remaining AL sequences an accumulation of glycine (4/8) or asparagine

(4/8, MM = 1x) in the CDR3. Only one AL and MM sequence showed an additional charge in the CDR1 region and an additional glycine in the CDR3 (Asparagine = 2x AL, 1x MM). It is also worth mentioning that more than half of the MM sequences presented an insertion of at least one additional charge in the CDR3 region (MM = 5/8; AL = 2/16, p = 0.021 (Fischer)).

Interestingly, these MM sequences corresponded in four out of five cases to a sequence, which displayed a M49 as well as a 54S mutation (AL n = 1). In contrast, a higher content of small AAs like glycine, alanine, or proline was noted in the CDR3 region of AL than MM sequences (**Figure 14**).

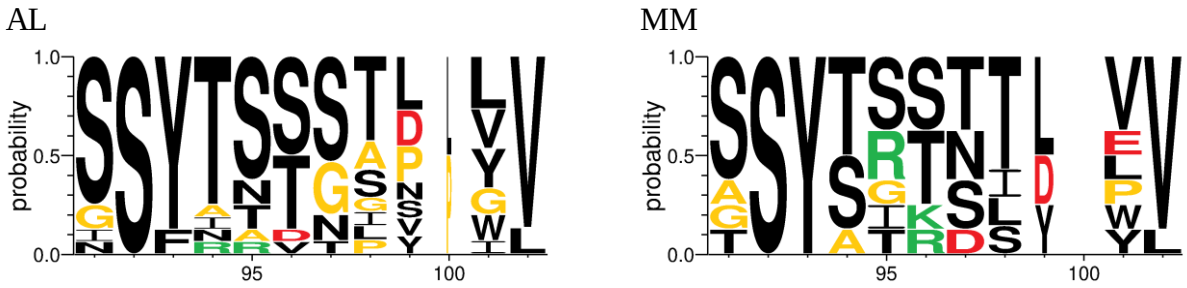


Figure 14. Comparison between the amino acid probability in the CDR3 region of IGLV2-14 assigned AL amyloidosis and multiple myeloma light chain sequences. AL = AL amyloidosis patients, MM = multiple myeloma patients, red letter = negatively charged amino acid, green letter = positively charged amino acid, yellow letter = small amino acids glycine, proline, and alanine. Amino acid numbers were assigned according to the VBase2 IGLV2-14 reference. The linker region spans positions 98-100, the IGLJ segments start at position 101. If only one or two amino acids were detected in the linker region, this is indicated by the width of the letter.

3.4.4 Amino Acid Composition

The overall percentage of the individual AAs was calculated to evaluate the effect of the detected mutations and mutation patterns and to get an overview of the overall LC composition. Only differences ≥0.5 % are mentioned in the following, differences >0.5 % are numerically specified.

The detected differences in the CDR3 region of MM and AL sequences did also reflect in the overall AA composition. The AL sequences displayed a higher amount of glycine, the MM sequences a higher amount of the positively charged AA arginine. In addition, a median percentage of methionine was detected in the AL sequences, which was not the case in the MM sequences (**Supplementary Information Table 4**).

When analyzing the sequences concerning the AL organ tropism, several differences were noticed (**Figure 15**). The AL_H sequences displayed a higher amount of serine (17.0 % vs. 15.7 %) and threonine (9.9 % vs. 9.0 %) than the AL_HK sequences. In addition, a higher

amount of aspartic acid, leucine, and a lower amount of asparagine (3.2 % vs. 4.0 %) and glutamic acid was detected in AL_H than in AL_HK sequences.

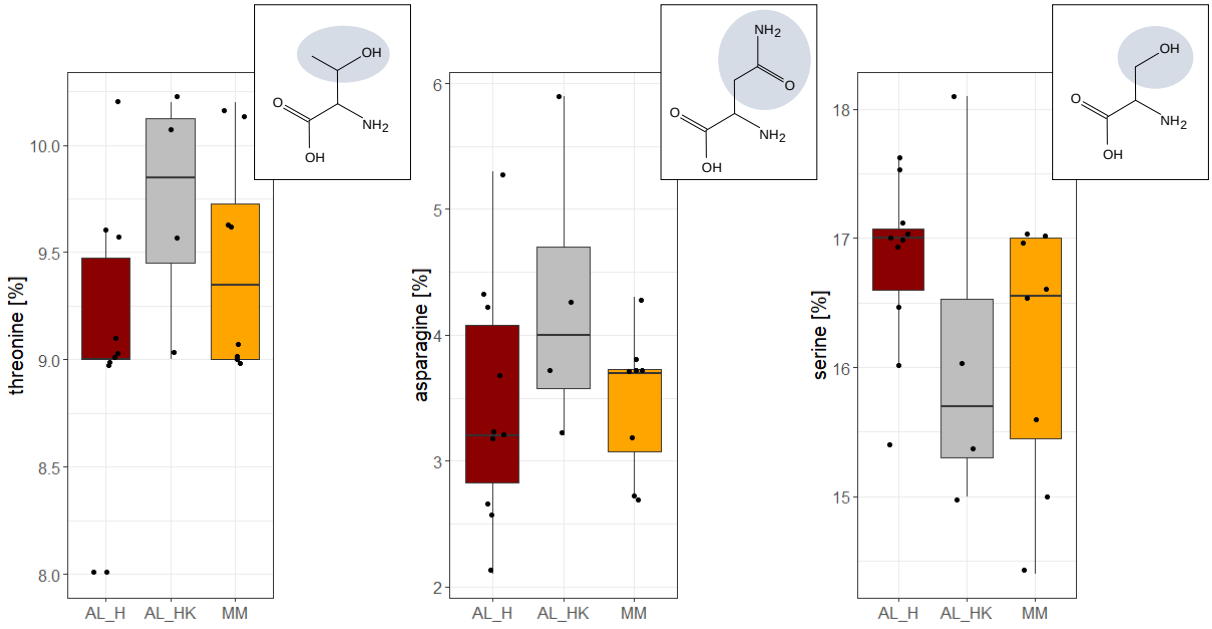


Figure 15. Comparison of the overall amino acid percentage of threonine, asparagine, and serine in the IGLV2-14 light chain sequences between the AL amyloidosis organ tropism and the multiple myeloma cohort. The amino acid percentage was calculated using the complete trimmed light chain sequence. MM = multiple myeloma patients, AL_H = AL amyloidosis patients with dominant heart involvement, AL_HK = AL amyloidosis patients with dominant heart and kidney involvement.

Stratifying the sequences for the presence or absence of a clonal HC, a higher amount of leucine, tyrosine, and valine was detected in the AL HC sequences than in the AL no HC sequences. In contrast, a higher amount of serine (17.5 % vs. 16.9 %) and especially asparagine (4.2 % vs. 3.2 %) was noted in the AL no HC sequences (**Figure 16**).

The AL sequences corresponding to patients with a dFLC >180mg/L showed a higher amount of alanine, asparagine, aspartic acid, isoleucine, tyrosine, and leucine (6.9 % vs. 6.4 %, p = 0.051 (U-test)). The AL < dFLC sequences showed a higher percentage of arginine, glycine, valine, and especially threonine (9.0 % vs. 10.2 %) (**Figure 16**).

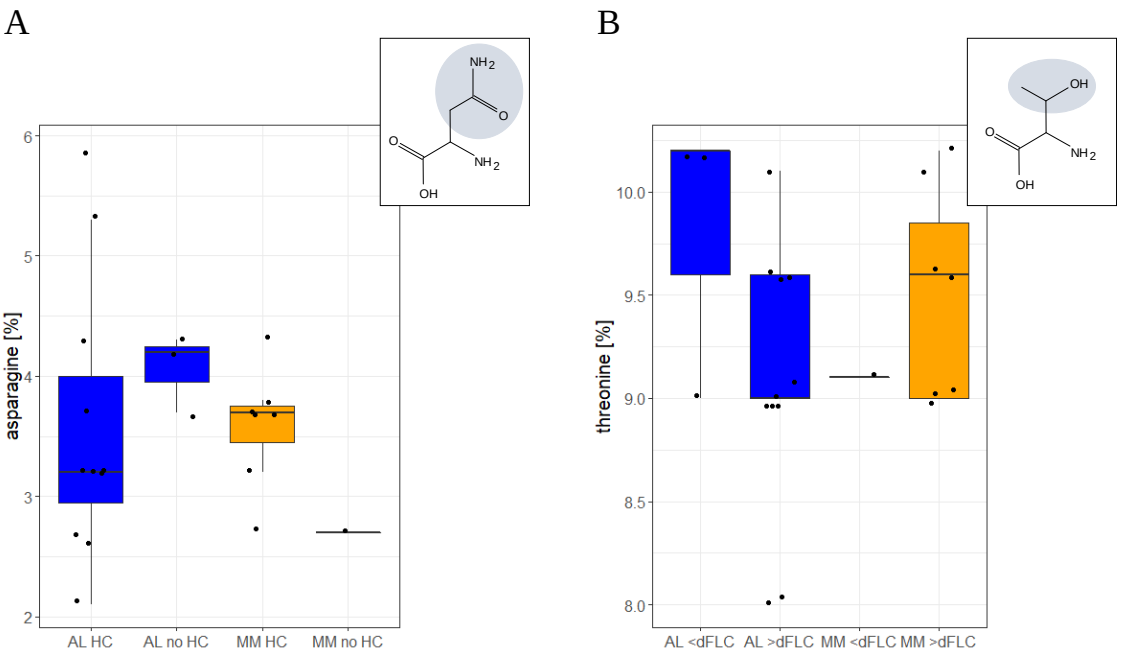


Figure 16. Comparison of the overall amino acid percentage of asparagine and threonine in IGLV2-14 assigned AL amyloidosis and multiple myeloma light chain sequences with respect to the presence or absence of a clonal heavy chain or a dFLC >180 mg/L. (A) Comparison of the overall amino acid percentage of asparagine and threonine in IGLV2-14 assigned AL amyloidosis and multiple myeloma light chain sequences, when stratifying for the presence of a clonal heavy chain in serum (B) Comparison of the overall amino acid percentage of asparagine and threonine in IGLV2-14 assigned AL amyloidosis and multiple myeloma light chain sequences when stratifying for a dFLC >180 mg/L. The amino acid percentage was calculated using the complete trimmed light chain sequence. MM = multiple myeloma patients, AL = AL amyloidosis patients, HC = detectable clonal heavy chain in the immunofixation electrophoresis in serum, > dFLC = dFLC >180 mg/L, < dFLC = dFLC <180 mg/L.

3.4.5 Biophysical Parameters

Several parameters were calculated to evaluate the effect of the mutations and amino acid composition on the biophysical properties of the LCs. The AGG parameter was included to determine the tendency for a beta sheet aggregation. The grand average of hydropathicity (GRAVY) value is defined by the sum of hydropathy values of all AA divided by the protein length. A negative GRAVY-value indicates a hydrophilic nature and a possible interaction of the linear protein with water molecules. The pI was also calculated to examine the theoretical stability of the LCs. In this context, the pI of the patient-derived LCs itself and the difference in comparison to a reference sequence (ΔpI) was calculated. This was performed for the complete LC (IGLVJC) and the IGLV and IGLJ segments (IGLVJ). In addition, the average molecular weight (Mw) of the LCs was specified.

Regarding the AGG value, the MM sequences presented a higher value than the AL sequences (1012.4 vs. 803.8) but no difference was noted when stratifying the AL sequences for dominant organ involvement (Table 22). In contrast, the AL HC sequences (830.1 vs. 753.0, p = 0.018 (U-test)) as well as the AL > dFLC sequences (828.1 vs. 741.1) showed a higher value than the respective other subgroup.

In the analysis of the GRAVY score, the AL sequences showed a lower score than the MM sequences (-0.284 vs. -0.264) – this was especially evident in the AL_HK sequences (-0.292). No differences were noted when stratifying for the presence of a clonal HC binding partner or dFLC (Table 22).

When analyzing the pI and ΔpI of the complete LC and the IGLVJ-segments, a higher value of the MM than AL sequences was noted in all cases (pI: IGLVJC = 6.77 vs. 6.59, IGLVJ = 6.41 vs. 5.92; ΔpI: IGLVJC = -0.32 vs. -0.50, IGLVJ = -0.34 vs. -0.70) (Table 22). Analyzing the AL sequences with respect to the organ involvement, the AL_HK sequences presented a higher pI IGLVJ (6.03 vs. 5.85; ΔpI IGLVJ -0.72 vs. -0.61) but a lower pI IGLVJC (6.15 vs. 6.76; ΔpI IGLVJ -0.70 vs. -0.33) than AL_H sequences. When stratifying the sequences concerning a potential HC binding partner, the AL HC subgroup presented a higher pI than the AL no HC subgroup (pI IGLVJC 6.76 vs. 6.42, pI IGLVJ 6.04 vs. 5.61). AL no HC sequences showed a lower median ΔpI in the IGLVJ segments (-1.13 vs. -0.70) but a higher ΔpI IGLVJC (-0.34 vs. -0.65) than the AL HC sequences. The AL > dFLC sequences displayed a higher pI IGLVJC (6.76 vs. 6.42) and lower pI IGLVJ (5.64 vs. 6.26). In both cases, the ΔpI was lower than in the AL < dFLC sequences (ΔpI IGLVJC: -0.88 vs. -0.34, ΔpI IGLVJ -0.94 vs. -0.49). In summary, a stringent tendency towards a higher pI was only noted in the MM than AL and AL HC than AL no HC cohorts.

The MM sequences also presented a higher Mw than the AL sequences (19655 vs. 19580), which was also noted in the AL_HK sequences in comparison to the AL_H sequences (19751 vs. 19542) (Table 22). No difference was detected when considering a clonal HC binding partner but the AL > dFLC sequences presented a higher Mw than the AL < dFLC sequences (19630 vs. 19539).

Table 22. Overview and comparison between different biophysical parameters of IGLV2-14 assigned AL amyloidosis and multiple myeloma light chain sequences and between different subgroups. AL = AL amyloidosis patients, MM = multiple myeloma patients, AL_H = AL amyloidosis patients with dominant heart involvement, AL_HK = AL amyloidosis patients with dominant heart and kidney involvement, HC = detectable clonal heavy chain in the immunofixation electrophoresis in serum, > dFLC = dFLC >180 mg/L, < dFLC = dFLC <180 mg/L, AGG = β-sheet aggregation tendency, GRAVY = grand average of hydropathicity, pI = isoelectric point, ΔpI = difference between light chain sequence and respective reference sequence, IGLVJC = trimmed full-length light chain, IGLVJ = connected IGLV and IGLJ segments, Mw = average molecular weight, * = p ≤0.05.

	median AGG	median GRAVY	median pI IGLVJC	median ΔpI IGLVJC	median pI IGLVJ	median ΔpI IGLVJ	median Mw (average)
AL	803.8	-0.284	6.59	-0.50	5.92	-0.70	19580
AL_H	803.8	-0.284	6.76	-0.33	5.85	-0.61	19542
AL_HK	791.6	-0.292	6.15	-0.70	6.03	-0.72	19751
MM	1012.4	-0.264	6.77	-0.32	6.41	-0.34	19655

	median AGG	median GRAVY	median pI IGLVJC	median ΔpI IGLVJC	median pI IGLVJ	median ΔpI IGLVJ	median Mw (average)
AL HC	830.1*	-0.284	6.76	-0.65	6.04	-0.70	19595
AL no HC	753.0*	-0.298	6.42	-0.34	5.61	-1.13	19615
MM HC	1131.0	-0.246	6.77	0.01	6.75	0.01	19625
AL > dFLC	828.1	-0.291	6.76	-0.88	5.65	-0.94	19630
AL < dFLC	741.1	-0.284	6.42	-0.34	6.25	-0.49	19539
MM > dFLC	893.8	-0.282	6.76	-0.64	6.06	-0.68	19685

3.4.6 Summary IGLV2-14

IGLV2-14 was one of the most frequently detected IGLV subfamilies in the AL (second, 20 %, n = 16) and MM cohort (third, 15 %, n = 8). Both diseases displayed an IGLJ2/IGLC2 linkage in most of the cases, but the AL sequences presented an IGLJ1/IGLC1 linkage more often than the MM sequences (5/16 vs. 1/8). In addition, an association with the presence of a HC was noted for both diseases and for the MM cohort also towards a dFLC >180 mg/L.

Of note, MM sequences presented a significantly higher median mutation count in the IGLV segment than AL sequences (11.5 vs. 8.5, p = 0.026) which further resulted in more mutational hotspots (9 vs. 2). The two AL-associated hotspots were restricted to a CDR region and shared by the MM cohort. In more detail, position S54 (third position CDR2) was one of the most frequently mutated MM positions and the most frequently mutated AL position. In contrast, for the hotspot position 62N, a difference was detected between the two diseases. While six out of eight MM sequences presented a mutation to various AAs, five out of sixteen AL sequences displayed a mutation towards the negatively charged AA aspartic acid (MM = 2x). Besides the mutation hotspots, several regions presented variations. For example, the CDR1 region of AL sequences showed an additional charge in half of the cases (MM = 3/8), and the other half showed equally distributed an additional glycine or asparagine in the CDR3 (MM = 1x N). More generally, the CDR3 region of AL sequences displayed a tendency towards small AAs like glycine, alanine, or proline. In contrast, the CDR3 region of MM sequences presented an insertion of at least one additional charge in five out of eight cases (AL = 2/16, p = 0.039). This concerned especially sequences with a mutation at position 49M and 54S (4/5). These findings correlate with a higher percentage of glycine in AL and aspartic acid in the full-length MM sequences. Besides the IGLV segment, the first AA of the IGLJ segment also presented a high mutation frequency and exchange tendency. The AL sequences showed a mutation in seven out of sixteen cases (4x leucine) and the MM sequences in three out of eight cases (1x leucine). In the analysis of several biochemical parameters, a higher β-sheet aggregation tendency, Mw, and pI were detected for the MM than AL sequences.

In the analysis regarding the AL organ tropism, both subgroups displayed IGLV2-14 as the second most common IGLV subfamily. However, the AL_H (n = 12) sequences presented more frequently an IGLJ1/IGLC1 linkage, and the AL_HK subgroup (n = 4) an IGLJ2/IGLC2 linkage. Overall, a higher median mutation count was detected in the IGLV (9.5 vs. 7.5) and IGLJ segment of AL_HK than AL_H sequences. Interestingly, this difference was especially prominent in the more stable FR1 and FR3 regions. Besides these overall differences, it was possible to detect two potential associated mutations in three out of four AL_HK sequences: at position S96T and at the seventh or eighth position in the IGLJ segment towards methionine. This might influence the overall higher percentage of serine in AL_H sequences and threonine in AL_HK sequences. In general, the AL_HK sequences showed a higher Mw than the AL_H sequences. Regarding the pI, different tendencies were found depending on considering the complete LC or only the IGLVJ segments.

The IGLV2-14-IGLJ1-IGLC1 linkage was not only found to be associated with AL and AL_H sequences but also with a present clonal HC in serum (AL_H HC: 37 % vs. 6 %; p = 0.034; MM no HC n = 1). Furthermore, AL HC sequences (n = 12) showed a higher median mutation count in the IGLV and IGLJ segments, with the most prominent difference in the CDR regions. Two out of four AL no HC sequences showed a H41N mutation and one or two additional prolines in the linker region. In contrast, only one AL HC sequence showed an additional proline in the linker region. Additionally, the AL HC sequences presented a significantly higher β-sheet aggregation tendency (830.1 vs. 753.0, p = 0.018), a higher pI, and a lower average Mw than the AL no HC sequences.

As already described for the HC analysis, the AL > dFLC (n = 11) subgroup also showed an association with IGLJ1/IGLC1 (MM < dFLC n = 1). Additionally, a higher mutation load in the IGLV and IGLC segment was detected in the AL > dFLC subgroup but no specific sequence feature or mutation was noted. However, the AL > dFLC sequences contained a higher percentage of several AAs including leucine (6.9 % vs. 6.4 %, p = 0.051) and a specifically lower percentage of threonine. In addition, a higher β-sheet aggregation tendency and a higher Mw were detected for the AL > dFLC sequences.

3.5 IGLV3-1

IGLV3-1 was detected as the third most commonly used IGLV subfamily in AL and the fourth most common IGLV subfamily in MM – comprising 13 AL_H (2x AL_HTX), two AL_HK, and three MM sequences (**Table 23**).

Due to the VLKL3c_Huhn oligonucleotide binding site, only the area starting from the FR2 region was amplified and sequenced in the first PCR step. By further sequencing reactions (**Figure 10**), a N-terminal elongated sequence was generated for all AL sequences except FOR132. Therefore, this sequence was excluded in the analysis.

Also, the bulk RNA seq data of the IGLV3-1 MM sequences presented a variation compared to the other IGLV subfamilies. The generated LC sequences did not cover the complete leader region and comprised only the last five AAs. To test the comparability of the sequences, one MM sequences (MM12) was Sanger sequenced and the dominant signals were in all cases consistent with the respective sequences (**Supplementary Information Figure 14**). In addition, the bioinformatic analysis identified several sequence sections in the bulk RNA sequencing data of MM120, leading to sequence gaps and additional sequence variations in the CDR3. These gaps were resolved with the Sanger sequencing results.
Thus, with additional sequencing reactions for AL and MM samples, the LC sequence could be verified and complemented to enable detailed analysis.

3.5.1 Subgroup Analysis

Six out of fifteen AL IGLV3-1 patients displayed a detectable clonal HC in serum (4x IgG, 1x IgA, 1x IgG+IgD) and two out of three MM patients (2x IgG) (**Table 23**). The distribution of the HC classes is in line with the approximate percentage of the total Ig in the adult serum (Lefranc and Lefranc 2020). The third MM sequence corresponded to a patient with a dFLC >180mg/L as well as fourteen out of fifteen AL sequences. Due to the low < dFLC sample number a subgroup analysis regarding a dFLC >180 mg/L was not performed.

Table 23. Overview of selected characteristics of 15 AL amyloidosis and 3 multiple myeloma patients with IGLV3-1 assigned light chain sequences. The column "X AA [n]" defines how many amino acids were not determined unambiguously. Organ inv. = organ involvement, AL_H = AL amyloidosis patients with dominant heart involvement, AL_HK = AL amyloidosis patients with dominant heart and kidney involvement, AL_HTX = AL amyloidosis patients who received a heart transplant, AL = AL amyloidosis, MM = multiple myeloma, HC in IFE = detectable clonal heavy chain in the immunofixation electrophoresis in serum. NA = not available, AA = amino acid, short = no N-terminal longer sequence with additional oligonucleotides could be generated.

Patient	Disease	Organ inv.	IGLJ	IGLC	X AA [n]	HC in IFE	dFLC [mg/L]
FOR123	AL	AL_H	IGLJ2	IGLC2	0	G+D	**1878.0**
FOR179	AL	AL_H	IGLJ2	IGLC2	0	-	**496.8**
FOR130	AL	AL_HTX	IGLJ2	IGLC2	0	-	**1380.2**
FOR161	AL	AL_HTX	IGLJ1	IGLC1	0	-	**198.4**
FOR120	AL	AL_H	IGLJ3	IGLC2	1	-	**863.3**
FOR206	AL	AL_H	IGLJ2	IGLC2	0	G	**2685.3**
FOR215	AL	AL_H	IGLJ7	IGLC7	0	G	**1477.7**
FOR118	AL	AL_H	IGLJ1	IGLC1	1	A	32.5
FOR134	AL	AL_H	IGLJ1	IGLC1	0	G	**694.4**

Patient	Disease	Organ inv.	IGLJ	IGLC	X AA [n]	HC in IFE	dFLC [mg/L]
FOR132	AL	AL_H	IGLJ2	IGLC2	0 - short	-	**296.0**
FOR131	AL	AL_H	IGLJ3	IGLC3	0	-	**381.7**
FOR116	AL	AL_H	IGLJ2	IGLC2	0	G	**1342.8**
FOR195	AL	AL_H	IGLJ2	IGLC2	0	-	**1543.5**
FOR191	AL	AL_HK	IGLJ2	IGLC2	0	-	**467.0**
FOR189	AL	AL_HK	IGLJ1	IGLC1	0	-	**474.0**
MM120	MM	-	IGLJ2	IGLC2	0	G	54.0
MM139	MM	-	IGLJ1	IGLC1	0	G	1.7
MM146	MM	-	IGLJ2	IGLC2	0	-	**4572.4**

3.5.2 Mutation Frequency and Count

To examine whether the difference in LC behavior is based on a difference in the general mutation distribution and frequency, these aspects were analyzed within the IGL segments.

In the analysis regarding the mutational count and frequency, the IGLV3-1 MM sequences presented a higher IGLV median mutational count (12.0 vs. 7.0, p = 0.012 (U-test), reference trimmed = 83 AA) and more often a mutation in the IGLC segment (33 % vs. 7 %, reference trimmed = 77 AA) than the AL sequences (**Table 24**). In more detail, a significantly higher mutational load was detected in the CDR1 (4.0 vs. 2.5, p = 0.032 (U-test), reference = 12 AA) and a higher mutation frequency in the more conserved FR regions of MM sequences. As in IGLV2-14, FR1 was the least often mutated region.

In the analysis regarding the AL organ tropism, the AL_H sequences comprised all sequences with a FR1 (50 % vs. 0 %) and IGLC mutation (8 % vs. 0 %) and showed more frequently a mutation in IGLJ, CDR2, CDR3, and less in FR2 (**Table 24**). However, it should be noted that only two sequences were assigned to the AL_HK subgroup.

When stratifying the sequences concerning a present clonal HC, a higher mutational load was detected in the IGLV segment of AL HC (8.0 vs. 7.0) and MM HC (13.0 vs. 10.0) sequences than in the respective other subgroup (**Table 24**). While only the MM sequences with a mutation in the IGLC segment corresponded to a patient without detectable HC, all IGLC mutated AL sequences corresponded to the AL HC subgroup. In addition, AL HC sequences displayed a higher mutation frequency in FR3 than the AL no HC subgroup (100 % vs. 38 %, 2.0 vs. 2.0, reference = 32 AA). In contrast, the AL no HC sequences presented a higher mutation frequency in CDR1 (100 % vs. 67 %).

Table 24. Comparison of the percentage of mutated IGL segments and average mutation count between AL amyloidosis and multiple myeloma IGLV3-1 assigned light chain sequences. The analysis was also performed with respect to different subgroups. AL = AL amyloidosis patients, MM = multiple myeloma patients, CDR = complementary determining region, FR = framework region, AA = amino acid, AL_H = AL amyloidosis patients with dominant heart involvement, AL_HK = AL amyloidosis patients with dominant heart and kidney involvement. HC = detectable clonal heavy chain in the immunofixation electrophoresis in serum, * = p ≤0.05, Differences greater than 30 % and one (segments) or two (regions) additional mutations are marked in bold. The median mutation values were calculated using only the mutated segments/regions. The significance level was determined using the complete dataset. FOR134 was excluded due to N-terminal shortening. The Vbase2 IGLV3-1 reference sequence was used for calculations. The IGLV and the IGLC references were trimmed at the N- or C-terminus, therefore the FR1 comprised only 10 amino acids. The CDR3 region includes also the patient-specific linker region and the first two amino acids of the IGLJ segment and therefore spans 9-11 amino acids. The MM no HC subgroup comprised only one patient and was therefore neglected.

			mutated segments [%] median mutation count [n]								
		n	IGLV 83 AA							IGLJ 12 AA	IGLC 77 AA
				FR1 10 AA	CDR1 12 AA	FR2 15 AA	CDR2 7 AA	FR3 32 AA	CDR3		
AL	[%]	14	100	43	86	**64**	86	**64**	93	79	**7**
	[n]		**7.0***	1.0	**2.5***	2.0	1.0	2.0	2.0	1.0	1.0
AL_H	[%]	12	100	**50**	83	**58**	**92**	67	**100**	**83**	8
	[n]		7.0	1.0	2.5	2.0	1.0	2.0	2.0	1.5	**1.0**
AL_HK	[%]	2	100	0	100	**100**	**50**	50	**50**	**50**	0
	[n]		6.5	0	2.5	1.5	2.0	2.0	2.0	1.0	**0**
MM	[%]	3	100	67	100	**100**	100	**100**	100	100	**33**
	[n]		**12.0***	1.0	**4.0***	1.0	2.0	1.0	2.0	1.0	1.0
AL HC	[%]	6	100	50	**67**	**50**	100	**100**	100	83	17
	[n]		**8.0**	1.0	**2.5**	2.0	1.0	2.0	2.0	1.0	**1.0**
AL no HC	[%]	8	100	38	**100**	75	75	**38**	88	75	**0**
	[n]		**7.0**	1.0	**2.5**	2.0	1.5	2.0	2.0	1.5	**0**
MM HC	[%]	2	100	**50**	**100**	**100**	100	100	100	100	**0**
	[n]		**13.0**	1.0	**4.0**	2.0	**2.0**	2.5	2.5	1.5	**0**

3.5.3 Sequence Alignment

A more detailed analysis of the LC sequences was performed on AA level to identify potential unique mutations or mutation patterns that could influence the biochemical properties of the LC itself or discriminate the subgroups. This was carried out in a step-wise procedure. First, general sequence characteristics such as mutation hotspots (mutations >50 %) and their location were identified separately for both diseases. Then the hotspots were analyzed in terms of an exchange towards a specific AA or AAs with the same side chain properties and the respective cDNA. In this analysis, the LC sequences with more than ten unambiguously determined AAs (marked as X) were included.

Since only three MM sequences were assigned to the IGLV3-1 family, only positions that were mutated in all sequences were considered mutation hotspots. The AL sequences were still analyzed for a ≥50 % cutoff.

In a more detailed sequence analysis, it was possible to identify three AL hotspots: 32A (CDR1, 7/14), 51S (CDR2, 11/14), and 93S (CDR3, 7/14), which were all located in a CDR region and, except of 51S, shared by the MM sequences (32A = 3/3, 93S = 3/3; 51S = 2/3). The MM sequences further presented the 94S (CDR3, 3/3) mutation hotspot (**Figure 17**).

At the shared hotspot 32A, an exchange toward valine was detected in the MM sequences two times (cDNA valine = GCT→GTT), while the AL sequences presented an exchange towards valine or threonine equally frequent (3/7 each¸ cDNA threonine = GCT→ACT) (**Figure 17**). Position 51S (third position CDR2) can be described as the most frequently mutated AL hotspot (11/14) and in most of the cases, an exchange towards threonine was noted (6/11, cDNA = AGC→ACC). However, the MM sequences displayed a S51N exchange two times (AL = 3x). At the third shared mutation hotspot, 93S, a mutation towards asparagine was detected in three out of seven AL cases and again two times in the MM sequences. The additional MM hotspot 94S did not present a distinct exchange pattern.

Despite not reaching the mutation cutoff for both diseases, position C33 should be mentioned. If mutated, both diseases displayed exclusively an exchange towards serine (MM = 2/3, AL = 5/14, cDNA = TGC→TCC).

It is worth mentioning that the AL sequences displayed the replacement of a positive charge at position 26K (CDR1) in six out of fourteen sequences. In contrast, for two out of three MM sequences, an additional charge in the CDR2 was detected. This concerned especially position 49Q (MM120: Q49R, MM139: Q49E; AL = 1x Q49E).

Overall, the AL and MM sequences did not show a specific linker region, even though a present overlapping Ensembl reference. However, at the first AA of the IGLJ segment, a mutation was detected in more than half of the AL (8/14) and all MM sequences (3/3) (**Figure 17**). The AL sequences equally frequently showed a mutation towards glycine or alanine (3x each), which was also detected one time each in the MM sequences.

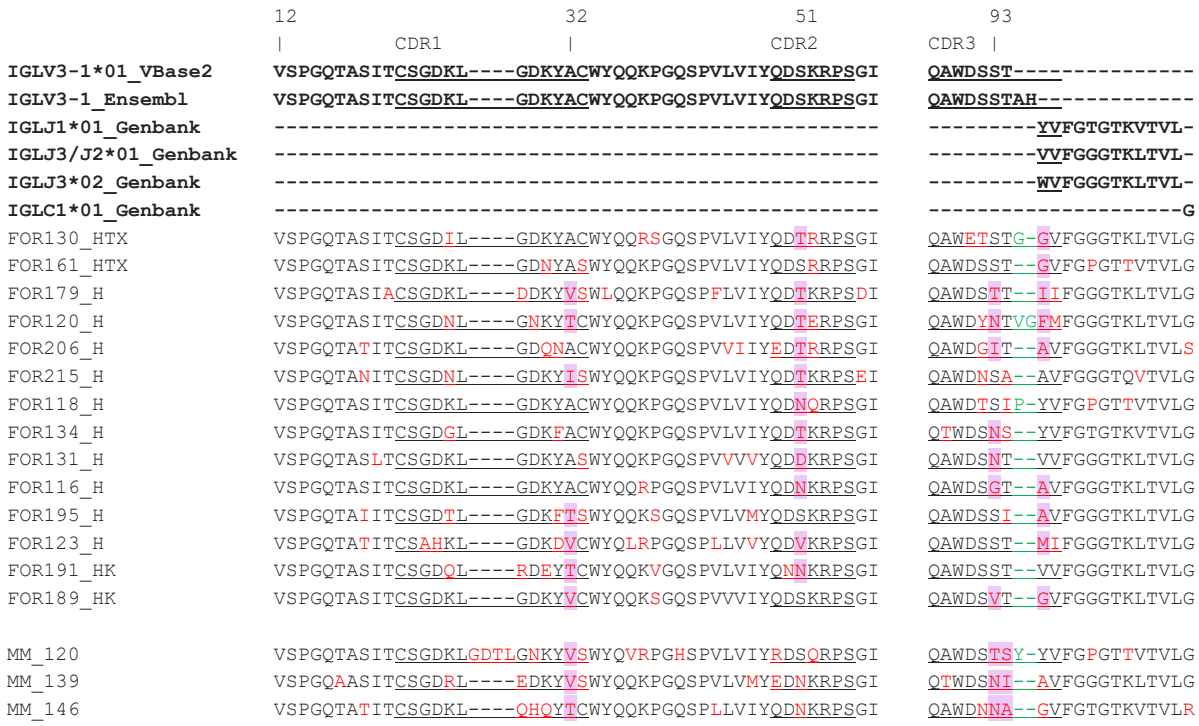


Figure 17. Sequence sections of IGLV3-1 assigned AL amyloidosis and multiple myeloma light chain sequences. Bold = reference sequences, underlined = CDR regions, red letter = mutation, purple highlight = mutation hotspot, X and grey highlight = not unambiguously determined amino acid, green letter = linker region, MM = multiple myeloma patient, _H = AL amyloidosis patient with dominant heart involvement, _HK = AL amyloidosis patient with dominant heart and kidney involvement, HTX = AL amyloidosis patient who which received a heart transplant. Only one IGLC reference is shown, due to the fact, that the first amino acid can be in all cases defined as glycine. Amino acids were numbered according to the VBase2 reference. The complete amino acid sequence alignment is shown in Supplementary Information Figure 15. The complete cDNA sequence alignment is shown in Supplementary Information Figure 16.

No distinct mutation pattern was noted when analyzing the sequences concerning the AL organ tropism. In contrast, in the subgroup analysis regarding the presence of a clonal HC binding partner, several differences were detected. The 32A mutation hotspot seemed to cluster to AL no HC sequences (5/8 vs. 2/6) as well as the C33S mutation (4/8 vs. 1/6). Contrastingly, both MM HC sequences showed the C33S mutation. Both, the CDR3 region of AL HC and AL no HC sequences presented a difference in the alanine and glycine content (**Figure 18**). At position 92S and 93S, two AL HC sequences displayed a mutation towards glycine, at position 94S one time towards alanine and as already described three times at the first AA of the IGLJ segment. The AL no HC sequences showed an additional glycine in the linker region two times as well as three times at the first AA of the IGLJ segment (1x alanine).

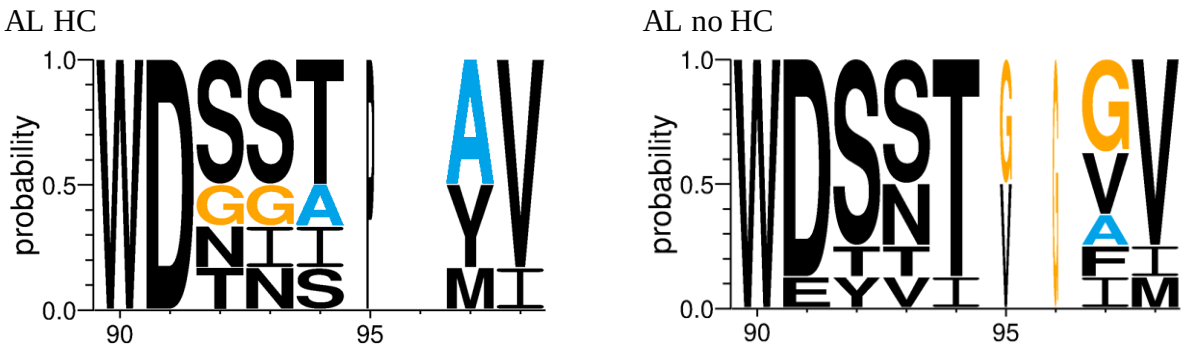


Figure 18. Comparison between the amino acid probability in the CDR3 region of IGLV3-1 assigned AL amyloidosis light chain sequences in the context of the absence or presence of a clonal heavy chain in serum. AL = AL amyloidosis patients, HC = detectable clonal heavy chain in the immunofixation electrophoresis in serum, blue = alanine, yellow = glycine. Amino acid numbers were assigned according to the VBase2 IGLV3-1 reference. The linker region spans positions 95 and 96, the IGLJ segments start at position 97. If only one or two amino acids were detected in the linker region, this is indicated by the width of the letter.

3.5.4 Amino Acid Composition

> *The overall percentage of the individual AAs was calculated to evaluate the effect of the detected mutations and mutation patterns and to get an overview of the overall LC composition. Only differences ≥0.5 % are mentioned in the following, differences >0.5 % are numerically specified.*

Comparing the overall AA composition of the AL and MM sequences, the only difference ≥0.5 % was detected for the positively charged AA arginine – with a higher percentage in MM than AL sequences (1.7 % vs. 1.2 %) (**Figure 19**, **Supplementary Information Table 5**).

When analyzing the AL sequences with regard to the organ involvement, the AL_H sequences presented a higher content of asparagine as well as of the small AAs alanine and proline. In contrast, the AL_HK sequences showed an especially high percentage of serine (13.4 % vs. 12.2 %) and valine (7.9 % vs. 6.7 %) (**Figure 19**).

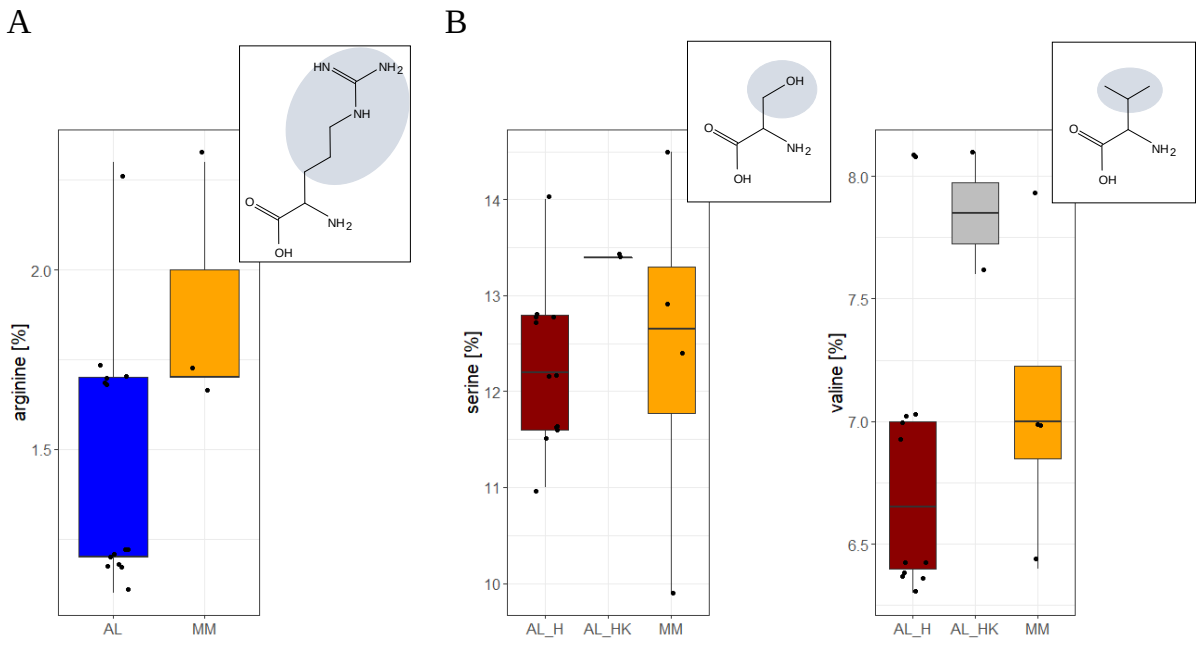


Figure 19. Comparison of the overall amino acid percentage of arginine, serine, and valine in IGLV3-1 assigned AL amyloidosis and multiple myeloma light chain sequences. (A) Comparison of the arginine percentage between IGLV3-1 assigned AL amyloidosis and multiple myeloma light chain sequences (B) Comparison of the serine and valine percentage in IGLV3-1 assigned light chain sequences between the AL organ tropism and multiple myeloma cohort. The amino acid percentage was calculated using the complete trimmed light chain sequence. AL = AL amyloidosis patients, MM = multiple myeloma patients, AL_H = AL amyloidosis patients with dominant heart involvement, AL_HK = AL amyloidosis patients with dominant heart and kidney involvement.

The detected potential alanine enrichment in the CDR3 region of AL HC sequences did also reflect in a higher overall alanine percentage in the complete LC (9.3 % vs. 8.7 %). In addition, the AL HC sequences presented a higher percentage of asparagine (3.8 % vs. 2.9 %) and a lower percentage of serine (11.6 % vs. 12.8 %, p = 0.004 (t-test)) than the AL no HC sequences (**Figure 20**).

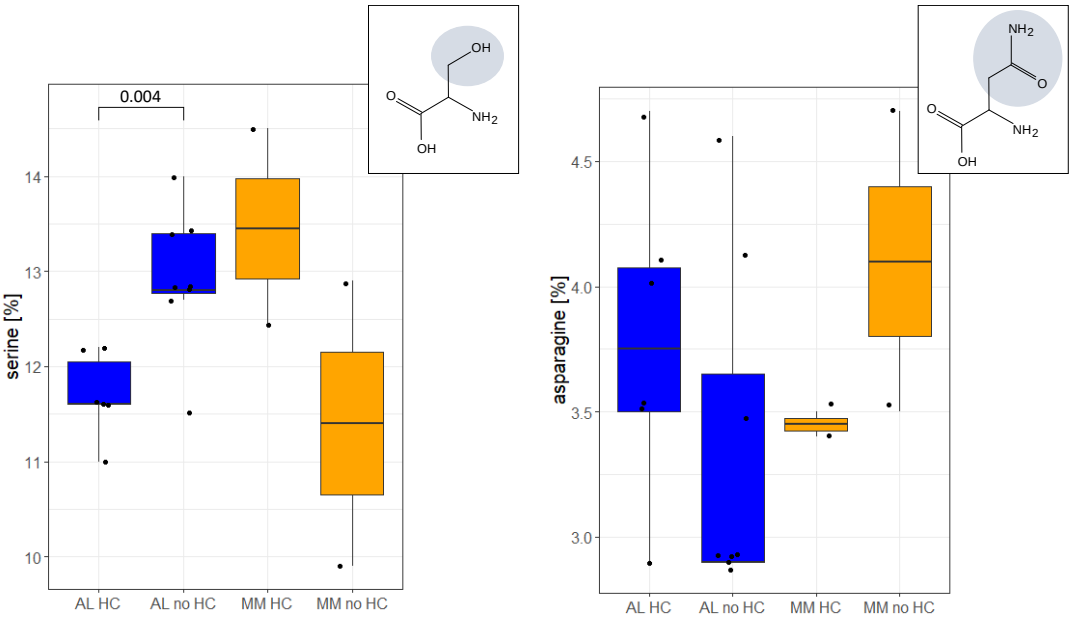


Figure 20. Comparison of the overall amino acid percentage of serine and asparagine between IGLV3-1 assigned AL amyloidosis and multiple myeloma sequences with respect to the presence or absence of a clonal heavy chain. The amino acid percentage was calculated using the complete trimmed light chain sequence. MM = multiple myeloma patients, AL = AL amyloidosis patients, HC = detectable clonal heavy chain in the immunofixation electrophoresis in serum.

3.5.5 Biophysical Parameters

Several parameters were calculated to evaluate the effect of the mutations and amino acid composition on the biophysical properties of the LCs. The AGG parameter was included to determine the tendency for a beta sheet aggregation. The grand average of hydropathicity (GRAVY) value is defined by the sum of hydropathy values of all AA divided by the protein length. A negative GRAVY-value indicates a hydrophilic nature and a possible interaction of the linear protein with water molecules. The pI was also calculated to examine the theoretical stability of the LCs. In this context, the pI of the patient-derived LCs itself and the difference in comparison to a reference sequence (ΔpI) was calculated. This was performed for the complete LC (IGLVJC) and the IGLV and IGLJ segments (IGLVJ). In addition, the average molecular weight (Mw) of the LCs was specified.

The AL sequences presented a higher median AGG score than the MM sequences (1146.7 vs. 1004.0), with an especially high score in the AL_HK sequences (1272.2; AL_H = 1146.7) (Table 25). In addition, the AL HC sequences presented a higher score than the AL no HC subgroup (1222.2 vs. 1146.7).

No difference was noted between the AL and MM cohort with respect to the GRAVY score (Table 25). However, the AL_H sequences presented a lower value than the AL_HK sequences (-0.348 vs. -0.276). In the AL HC sequences, a slightly lower value than in the AL no HC sequences was detected.

Regarding the theoretical stability and solubility, the MM sequences displayed a slightly higher pI IGLVJC value (5.32 vs. 5.04) and comparable IGLVJ value (4.70 vs. 4.60) than the AL sequences; this also reflected in the ΔpI values (IGLVJC = MM: -0.46, AL: -0.51; IGLVJ = MM: -0.15, AL: -0.38) (Table 25). Analyzing the AL organ tropism, the AL_H sequences showed a more pronounced decrease in ΔpI IGLVJC than the AL_HK sequences (-0.51 vs. -0.26). When stratifying for the presence of a clonal HC, no difference was found.

The MM sequences showed a higher Mw than the AL sequences (18446 vs. 18193), with a more pronounced difference compared to the AL_HK sequences (18156, AL_H = 18193) (Table 25). In addition, AL HC sequences displayed a higher value than the respective other subgroup (18205 vs. 18152).

Table 25. Overview and comparison between different biophysical parameters of IGLV3-1 assigned AL amyloidosis and multiple myeloma light chain sequences and between different subgroups. AL = AL amyloidosis patients, MM = multiple myeloma patients, AL_H = AL amyloidosis patients with dominant heart involvement, AL_HK = AL amyloidosis patients with dominant heart and kidney involvement, HC = detectable clonal heavy chain in the immunofixation electrophoresis in serum, AGG = β-sheet aggregation tendency, GRAVY = grand average of hydropathicity, pI = isoelectric point, ΔpI = difference between the light chain sequence and the respective reference sequence, IGLVJC = trimmed full-length light chain, IGLVJ = connected IGLV and IGLJ segments, Mw = average molecular weight. The MM no HC subgroup comprised only one patient and was therefore neglected.

	median AGG	median GRAVY	median pI IGLVJC	median ΔpI IGLVJC	median pI IGLVJ	median ΔpI IGLVJ	median Mw (average)
AL	1146.7	-0.337	5.04	-0.51	4.60	-0.38	18193
AL_H	1146.7	-0.348	5.04	-0.51	4.60	-0.38	18193
AL_HK	1272.2	-0.276	5.13	-0.26	4.78	-0.26	18156
MM	1004.0	-0.378	5.32	-0.46	4.70	-0.34	18446
AL HC	1222.2	-0.348	5.21	-0.51	4.59	-0.37	18205
AL no HC	1146.7	-0.309	5.04	-0.43	4.60	-0.45	18152
MM HC	1180.6	-0.384	5.12	-0.78	4.66	-0.39	18518

3.5.6 Summary IGLV3-1

IGLV3-1 seemed to be more associated with AL (third most common, 18 %, n = 15) than MM (6 %, n = 3), and also the IGLJ/IGLC linkage presented variations. While most AL cases displayed an IGLJ2/IGLC2 linkage, two out of three MM sequences showed an IGLJ1/IGLC1 linkage. Further, resolving the IGLV segment N-terminal was difficult for both diseases and in both sequencing approaches. An analysis regarding the dFLC was not performed, since IGLV3-1 was associated with AL > dFLC sequences.

In a comparison between AL and MM, the MM sequences presented more frequently mutated in the IGLV (12.0 vs. 7.0, p = 0.012) and IGLC segment – including all FR regions and especially CDR1 (4.0 vs. 2.5, p = 0.032). Both cohorts displayed three mutation hotspots, but

the MM cohort comprised only three sequences. So, only when all three sequences displayed a mutation a hotspot was defined. The MM and AL hotspots were all located in a CDR, and only one was exclusive to the AL (MM: 2/3) or MM (AL: 4/15) cohort. Despite sharing mutation hotspots, a different AA exchange tendency was noted. At position 32A, the MM sequences presented an exchange towards valine two times and the AL sequences towards valine or threonine equally. A more pronounced difference was detected for the not shared hotspot at position 51S. At this position, more than half of the AL sequences presented a S51T exchange, and two out of three MM sequences showed a S51N exchange. In addition, the CDR1 of AL sequences frequently presented the loss of a positively charged AA. In contrast, the CDR2 of MM sequences showed an additional charge. Addressing the IGLJ segment, the first AA was mutated in ≥50 % of the sequences regardless of the disease.
Comparing the overall AAs composition, the only difference >0.5 % was noted for the positively charged AA arginine which was detected with a higher median percentage in the MM sequences. Regarding the pI and ΔpI, no major differences between the two diseases were noted but the AL sequences presented a higher aggregation tendency and a lower Mw than the MM sequences.

In general, IGLV3-1 was the most common AL_H (n = 13) IGLV subfamily. In more detail, the AL_H subgroup comprised all sequences with a mutation in FR1 and the IGLC segment as well as a higher mutation frequency in the IGLJ segment, CDR2, and CDR3 regions. No organ-specific mutation pattern was observed but the AL_HK (n = 2) sequences contained a notably higher percentage of serine and valine. Further, the AL_HK sequences showed a higher aggregation tendency (AGG score) and a lower GRAVY score.

In the analysis regarding the presence of a clonal HC, the AL no HC (n = 9) subgroup displayed more frequently an IGLJ2/IGLC2 linkage than the AL HC subgroup (n = 6). However, the HC subgroups showed a higher median mutation count in the IGLV segment for both diseases. While both MM sequences with a mutation in the IGLC segment corresponded to no HC patients, all IGLC mutated AL sequences corresponded to HC patients. Despite not reaching the hotspot cutoff, the C33S mutation seemed to be important. This mutation clustered to AL no HC (4/8 vs. 1/6) and MM HC sequences (2/2 vs. 0) and might be reflected in a higher overall percentage of serine in AL no HC sequences (12.8 % vs. 11.6 %, p = 0.004). Further, AL HC sequences showed an alanine enrichment in the CDR3 region and a higher overall percentage in the complete LC. In contrast, the AL subgroups displayed no major differences in the analysis of several biophysical parameters.

3.6 IGLV3-21

The IGLV3-21 subfamily was the most commonly used MM (n = 11) and the fourth most commonly used AL IGLV subfamily (n = 10). A difference was particularly noted regarding the AL organ tropism. While nine AL_H sequences were assigned to the IGLV3-21 subfamily this was only the case for one AL_HK sequence.

As already described for the IGLV3-1 subfamily, due to the VLKL3c_Huhn oligonucleotide binding site only the area starting from FR2 was amplified and sequenced in the first PCR step. By further sequencing reactions, a N-terminal elongated IGLV3-21 sequence was generated in only two cases (FOR104_H, FOR187_HK). FOR177 exhibited multiple nucleotide signal overlaps in the FR1 and CDR1, resulting in 10 ambiguous AAs in these regions and none in the remaining sequence – therefore, it was not defined as N-terminally resolved. In addition to the difficulty resolving the N-terminus, the sequences of the IGLV3-21 subfamily showed the most abundant nucleotide overlaps in the Sanger sequencing (**Figure 21**), resulting in high numbers of ambiguous identifiable AAs in six out of ten AL cases.

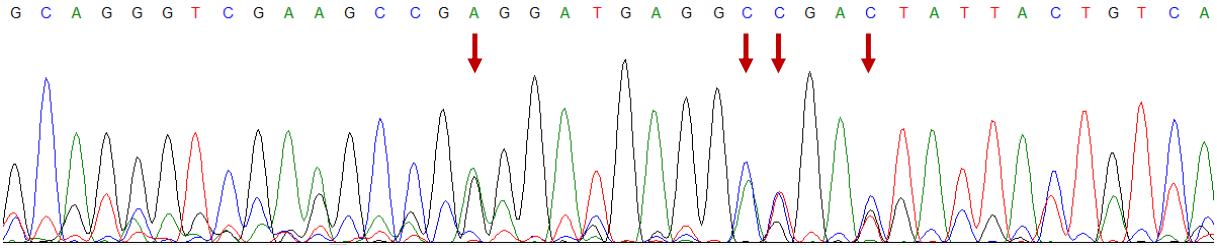


Figure 21. Section of an electropherogram of a sequencing reaction leading to one main sequence and several subsequences. Signal overlaps of several nucleotides can be seen at several positions, some of which are marked with arrows as examples.

Contrastingly, only one MM sequence showed an additional sequence section in the bioinformatic analysis (MM138) and another one an additional sequence_B (MM122). The additional sequence section of MM138 (sequence_A) did not match any sequence region and was therefore neglected. To test the comparability of the sequences, three MM sequences (MM111, MM106, MM119) were also Sanger sequenced. In two cases (MM111, MM106) it was possible to generate a N-terminal elongated sequence without ambiguous AAs (**Supplementary Information Figure 17**, **Supplementary Information Figure 18**). Sanger sequencing of MM119 resulted in an N-terminal extended sequence but some signal overlaps were detected, resulting in not all AA positions being clearly verified (**Supplementary Information Figure 19**). However, based on the results concerning LC verification of the other IGLV subfamilies a reliable sequence can be assumed.

3.6.1 Subgroup Analysis

Four out of ten AL IGLV3-21 patients showed a detectable clonal HC in serum which was in all cases identified as IgG. In the MM cohort, nine out of eleven patients displayed a detectable clonal HC in serum which was identified as IgG in eight cases and once as IgA. This distribution of HC classes is in line with the approximate percentage of the total Ig in the adult serum (Lefranc and Lefranc 2020). The > dFLC subgroups comprised eight out of ten AL and only four out of eleven MM sequences (**Table 26**).

Table 26. Overview of selected characteristics of 10 AL amyloidosis and 11 multiple myeloma patients with IGLV3-21 assigned light chain sequences. The column "X AA [n]" defines how many amino acids were not determined unambiguously. Organ inv. = organ involvement, AL_H = AL amyloidosis patients with dominant heart involvement, AL_HK = AL amyloidosis patients with dominant heart and kidney involvement. AL = AL amyloidosis, MM = multiple myeloma, HC in IFE = detectable clonal heavy chain in the immunofixation electrophoresis in serum, AA = amino acid, NA = not available, short = no N-terminally longer sequence with additional oligonucleotides could be generated, NGS = patient samples analyzed through a next-generation sequencing approach.

Patient	Disease	Organ inv.	IGLJ	IGLC	NGS	X AA [n]	HC in IFE	dFLC [mg/L]
FOR104	AL	AL_H	IGLJ2/3	IGLC3	yes	0	G	**693.3**
FOR105	AL	AL_H	IGLJ2	IGLC2	-	7 - short	G	135.8
FOR127	AL	AL_H	IGLJ2	IGLC2	-	30 - short	G	**332.9**
FOR136	AL	AL_H	IGLJ3	IGLC3	yes	0 - short	-	**1166.8**
FOR162	AL	AL_H	IGLJ2/3	IGLC2	-	16 - short	-	**513.0**
FOR163	AL	AL_H	IGLJ2	IGLC2	yes	19 - short	G	125.4
FOR169	AL	AL_H	IGLJ3	IGLC3	yes	0 - short	-	**708.7**
FOR176	AL	AL_H	IGLJ3	IGLC3	yes	0 - short	-	**314.1**
FOR177	AL	AL_H	IGLJ3	IGLC3	yes	10 - short	-	**252.9**
FOR187	AL	AL_HK	IGLJ2	IGLC2	-	6	-	**1065.0**
MM106	MM	-	IGLJ2	IGLC2	-	0	G	46.3
MM108	MM	-	IGLJ2	IGLC2	-	0	-	**5132.1**
MM111	MM	-	IGLJ3	IGLC3	-	0	G	**3382.4**
MM119	MM	-	IGLJ2	IGLC2	-	0	A	NA
MM122	MM	-	IGLJ2	IGLC2	-	0	G	67.2
MM123	MM	-	IGLJ1	IGLC1	-	0	G	45.6
MM134	MM	-	IGLJ2/3	IGLC2	-	0	G	10.0
MM138	MM	-	IGLJ2	IGLC2	-	0	G	90.0
MM143	MM	-	IGLJ2	IGLC2	-	0	G	**1569.2**
MM144	MM	-	IGLJ2	IGLC2	-	0	G	**288.2**
MM150	MM	-	IGLJ3	IGLC2/3	-	0	G	12.4

Parenthesis: Due to the majority of AL amyloidosis sequences not being N-terminal resolved and containing a large number of ambiguously defined AAs, the sequences were not analyzed concerning mutation distribution, amino acid composition, and biophysical parameters.

Based on the results that two AL amyloidosis sequences, as well as the multiple myeloma sequences, were successfully N-terminal resolved with an additional oligonucleotide, it was assumed that the used oligonucleotide set is capable of binding at the chosen positions in FR1 and IGLC. However, most IGLV3-21 AL amyloidosis light chain sequences showed multiple nucleotide signal overlaps for a large proportion of cDNA positions – so, it can be assumed that additional subsequences are potentially present. To verify the IGLV3-21 subfamily assignment and further investigate the composition of these sequences, next-generation sequencing experiments were performed with seven AL_H sequences (FOR117, FOR136, FOR104, FOR163, FOR169, FOR176, FOR113, **Table 26**). To be able to draw a comparison, the clearly evaluable and N-terminal resolved AL sequence was used as a positive control (FOR104_pC). In contrast, a sequence that had been defined as non-evaluable in the overall analysis was used as negative control (FOR113_nC). For this non-evaluable sequence, a sequence of the IGLV3-21 subfamily with 10 non-uniquely identifiable amino acids was detected in the first Sanger sequencing reaction and a sequence of the IGLV3-25 subfamily with 10 non-evaluable amino acids in the second Sanger sequencing reaction.

3.6.2 Verification of the AL IGLV3-21 Sequences and Subfamily Assignment through a Next-Generation Sequencing Approach

3.6.2.1 Verification of AL IGLV3-21 Subfamily Assignment

In the first NGS assay, a PCR approach based on the multiplex forward oligonucleotide set ($MP_{\_NGS\_PCR}$) was used to verify the IGLV3-21 family assignment and further analyze the sample composition.

In the analysis of the IGLV3-21 assigned sequences (n = 6), all samples displayed an IGLV3-21 sequence_A and in more than half of the cases this sequence was detected in over 90 % (4/6): FOR177 = 99.5 %, FOR169 = 93.4 % and FOR176 = 93.9 % (**Table 27**). This also included the positive control sequence, which showed the highest detected percentage: FOR104_pC = 99.9 %. In contrast, the negative control displayed an IGLV3-9 sequence_A (3.3 %). In total, a median percentage of 93.4 % was detected in the respective sequence_A of all seven samples.

In general, two out of six IGLV3-21 assigned samples displayed four sequences >1 % (FOR163, FOR169) (**Table 27**). In all cases, these additional sequences were assigned to the IGLV3 family – with IGLV3-21 as the most prominent subfamily. In the analysis of FOR113_nC, a total of 16 sequences >1 % were detected, which were assigned twice to IGLV5-45, three times to IGLV3-25, and ten times to IGLV3-21. This led to the detection of IGLV3-21 as the most prominent family for this sample as well.

In conclusion, the IGLV3-21 family assignment determined in Sanger sequencing could be confirmed for all samples by the $MP_{\_NGS\_PCR}$ experiment. The findings also correlate with the Sanger sequencing results. Three out of four sequences without ambiguous AAs in the Sanger sequencing showed only one IGLV3-21 sequence_A in the NGS analysis (**Table 27**). In the fourth sample FOR169, an IGLV3-21 sequence_A was detected in 93.4 %.

Table 27. Evaluation of the sample composition of the IGLV3-21 next-generation sequencing experiment based on multiplex oligonucleotide PCR samples. ($MP_{\_NGS\_PCR}$) Seven AL amyloidosis samples from patients with dominant heart disease and IGLV3-21 assigned light chain sequences were examined. pC = positive control, nC = negative control, short = no N-terminally longer sequence with additional oligonucleotides could be generated, sequence_A = sequence with the highest percentage in the next-generation sequencing analysis. The column "X AA [n]" defines how many amino acids were not determined unambiguously.

Sample	X AA [n]	Number of detected sequences >1 %	IGLV subfamily sequence_A (%)	all sequences >1 % IGLV3-21 [%]	remaining sequences >1 % IGLV3 [%]	remaining sequences >1 % not IGLV3 [%]
FOR104_pC	0	1	IGLV3-21 (99.9 %)	99.9	0	0
FOR136	0 - short	1	IGLV3-21 (52.6 %)	52.6	0	0
FOR163	19 - short	4	IGLV3-21 (1.4 %)	2.7	2.3	0
FOR169	0 - short	4	IGLV3-21 (93.4 %)	93.4	4.0	0
FOR176	0 - short	1	IGLV3-21 (93.9 %)	93.9	0	0
FOR177	10 - short	1	IGLV3-21 (99.5 %)	99.5	0	0
FOR113_nC	-	16	IGLV3-9 (3.3 %)	16.4	9.3	3.6

3.6.2.2 Verification of AL amyloidosis IGLV3-21 Light Chain Sequences

Additionally, it was examined whether the sequences generated by the $MP_{\_NGS\_PCR}$ approach were equivalent to the sequences generated by Sanger sequencing. This was demonstrated for the four sequences that showed only one sequence >1 % in the $MP_{\_NGS\_PCR}$ analysis (FOR136, FOR177, FOR177) – including the FOR104_pC sample (**Figure 22**, **Supplementary**

Information Figure 20). Mismatches were only detected in and after the reverse oligonucleotide binding site. Therefore, the IGLV segments starting from CDR2 and IGLJ onwards were almost or completely covered. In conclusion, not only the family assignment but also the sequence was verified in the MP$_{_NGS_PCR}$ approach.

Verification of the Sanger sequence was also possible for FOR169 (**Supplementary Information Figure 21**). Although this sample showed in the MP$_{_NGS_PCR}$ analysis a total of four sequences >1 %, the corresponding sequence_A was detected with 93.4 % and assigned to IGLV3-21.

Figure 22. Sequence alignment of the Sanger sequence and the sequence with the highest percentage in the IGLV3-21 next-generation sequencing experiment based on multiplex oligonucleotide PCR samples for FOR104_pC and FOR176. (MP_NGS_PCR) (A) sequence alignment positive control FOR104_pC (B) sequence alignment FOR176. CDR region = underlined, IGLJ segment = blue highlight, linker region = green, sequence deviation = red, * = matching position, VLKL3c_huhn = IGLV3 specific forward oligonucleotide, JLHD123_rv = IGLJ segment specific reverse oligonucleotide, Sanger = sequence generated by Sanger sequencing, NGS = sequence with the highest percentage in the next-generation sequencing experiment.

For FOR163 several positions in the cDNA and AA sequence were not clearly determined by Sanger sequencing. (**Figure 23**) and in the MP$_{_NGS_PCR}$ analysis, two IGLV3-21 sequences with a comparable percentage were detected (sequence_A = 1.43 %, sequence_B = 1.31 %). Taken together, a deviation of the Sanger sequence to both MP$_{_NGS_PCR}$ sequences was noted at eight positions, mainly restricted to the CDR3 region (6/8).

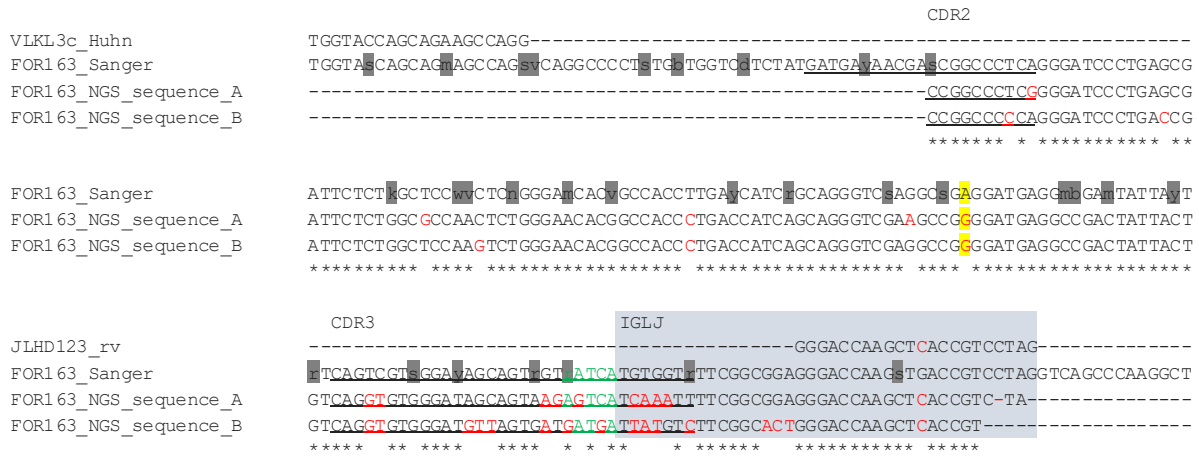


Figure 23. FOR163 sequence alignment of the sequence generated by Sanger sequencing and the sequence with the highest percentage in the IGLV3-21 next-generation sequencing experiment based on multiplex oligonucleotide PCR samples. (MP_NGS_PCR) Next-generation sequencing sequences were designated in descending alphabetical order based on the percentage. CDR region = underlined, IGLJ segment = blue highlight, linker region = green, sequence deviation = red, * = matching position, VLKL3c_huhn = IGLV3 specific forward oligonucleotide, JLHD123_rv = IGLJ segment specific reverse oligonucleotide, Sanger = sequence generated by Sanger sequencing, NGS = sequence generated through next-generation sequencing, yellow highlight = divergent position in all three analyses, grey highlight = IUPAC code for positions with nucleotide overlaps in the sanger sequencing.

In the analysis of the FOR113_nC sequence, ten IGLV3-21 sequences were detected in the MP$_{_NGS_PCR}$ approach. The IGLV3-21 sequence with the highest percentage (sequence_B, 2.44 %) showed a deviation from the Sanger sequence at eleven positions. Taking all ten MP$_{_NGS_PCR}$ IGLV3-21 sequences together, a deviation to the Sanger sequence was noted at four positions. One position in the IGLJ segment can be explained by the different reverse oligonucleotides in the Sanger and NGS sequencing (**Figure 24**)

It is of note that these two samples shared one remarkable position in FR3: at this position, the Sanger sequences of FOR113_nC and FOR163 displayed an adenine, while all detected NGS sequences showed a guanine, which is in accordance with the IGLV3-21 reference (**Figure 23**, **Figure 24**). An adenine would lead to an exchange on AA level at position 80 from glycine towards the negatively charged AA glutamic acid.

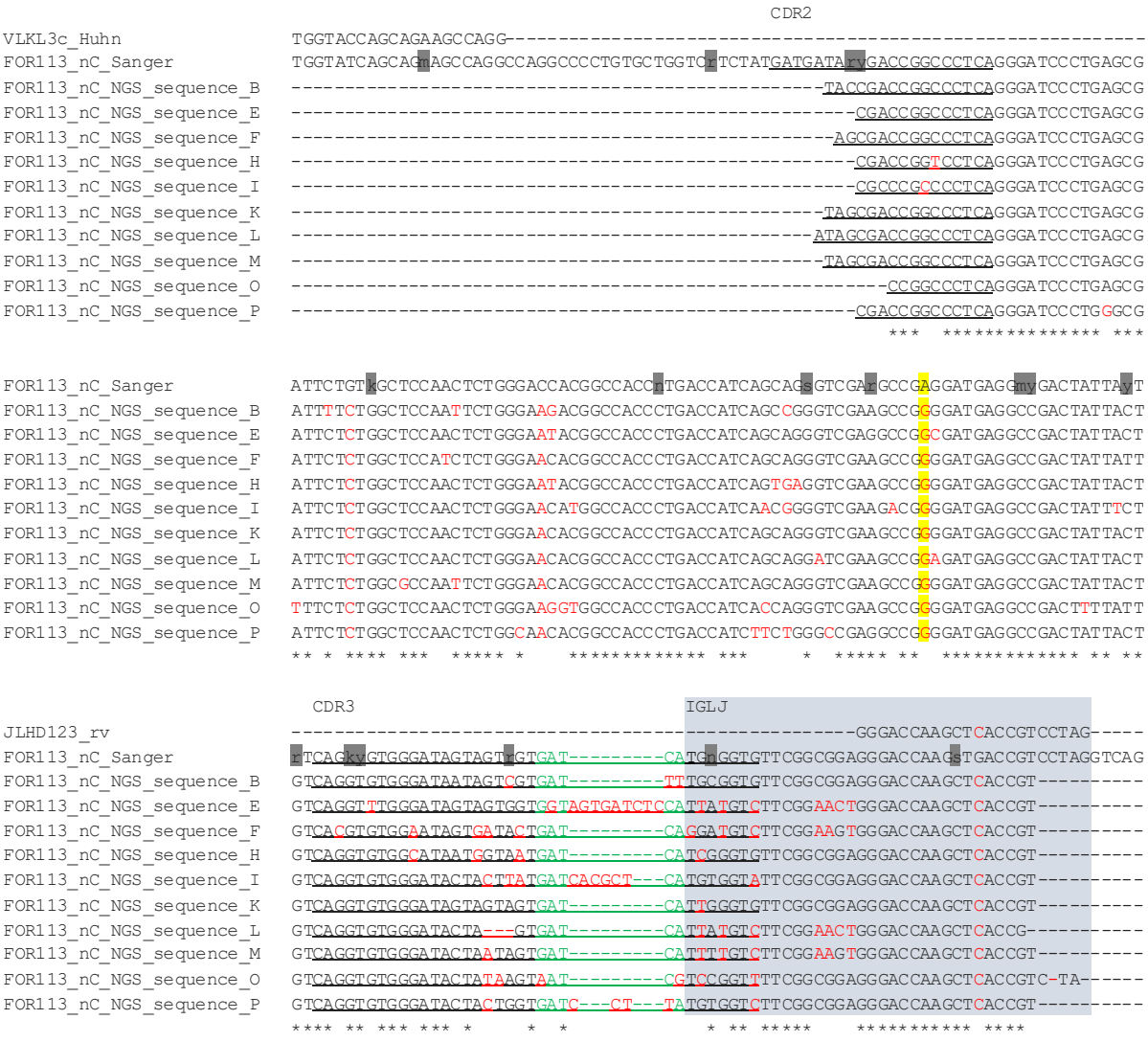


Figure 24. FOR113 sequence alignment of the sequence generated by Sanger sequencing and the sequence with the highest percentage in the IGLV3-21 next-generation sequencing experiment based on multiplex oligonucleotide PCR samples. (MP_NGS_PCR) Next-generation sequencing sequences were designated in descending alphabetical order based on the percentage. CDR region = underlined, IGLJ segment = blue highlight, linker region = green, sequence deviation = red, * = matching position, VLKL3c_huhn = IGLV3 specific forward oligonucleotide, JLHD123_rv = IGLJ segment specific reverse oligonucleotide, Sanger = sequence generated by Sanger sequencing, NGS = sequence generated through next-generation sequencing, yellow highlight = divergent position in all three analyses, grey highlight = IUPAC code for positions with nucleotide overlaps in the sanger sequencing.

Taken together, a clear verification of the Sanger sequence was possible in five cases, including the positive control. The two remaining sequences share a deviation in the FR3 between the Sanger and MP$_{NGS_PCR}$ approach. Additionally, FOR163 showed a total of eight and the negative control sequence four divergences between all IGLV3-21-assigned MP$_{NGS_PCR}$ sequences and the respective Sanger sequence.

3.6.3 Analysis for Lack of N-terminal Coverage of IGLV3-21 AL Sequences through a Next-Generation Sequencing Approach

3.6.3.1 Analysis of the IGLV3-21 Sample Composition and Comparison to the $MP_{\_NGS\_PCR}$ Approach

In a second approach, PCR samples that should generate IGLV3 N-terminally longer sequences were analyzed using NGS ($N_{\_NGS\_PCR}$).

Interestingly, the $N_{\_NGS\_PCR}$ samples contained a higher median number of sequences >1 % than the $MP_{\_NGS\_PCR}$ samples (14 vs. 1; p = 0.038 (U-test), **Table 28**). While in the $MP_{\_NGS\_PCR}$ approach four sequences displayed only one sequence >1 %, this was exclusivity detected for the $N_{\_NGS\_PCR}$ FOR104_pC sequence. However, the FOR113_nC sample displayed the second least number of sequences ≥1 % (n = 5). Furthermore, a lower overall median percentage of all sequence_A compared to the $MP_{\_NGS\_PCR}$ samples was detected (18.4 % vs. 93.4 %). Additionally, while in the $MP_{\_NGS\_PCR}$ approach, all AL samples displayed an IGLV3-21 sequence_A, this was only the case for two samples in the $N_{\_NGS\_PCR}$ analysis (FOR177, FOR104_pC).

However, the positive control sequence FOR104_pC showed an IGLV3-21 sequence_A with an abundance of 99.91 %. The negative control sequence FOR113_nC showed an IGLV3-25 sequence_A with 55.15 % as well as two additional IGLV3-25 sequences (sequence_B = 17.47 %, sequence_C = 6.42 %). Thus, for these two samples, the IGLV family assignment based on the Sanger sequencing and the $N_{\_NGS\_PCR}$ approach was consistent.

FOR177 displayed 20 different sequences with an abundance >1 % but only sequence_A was assigned to the IGLV3-21 subfamily (16.51 %). Based on these 20 sequences, IGLV3-27 was the most prominent IGLV3 subfamily (6x, 28.99 %). IGLV3-27 was also the most prominent IGLV3 subfamily in FOR176 (8x, 49.61 %, sequence_A = IGLV3-10 17.77 %), with a total of 14 sequences >1 % including one IGLV3-21 sequence (sequence_F = 6.07 %). The remaining samples FOR136 (32.5 %), FOR163 (50.04 %), and FOR169 (66.24 %) displayed IGLV3-25 as the most abundant IGLV3 subfamily. None of these three samples contained an IGLV3-21 sequence in a total of 23 (FOR136), 17 (FOR163), or nine (FOR169) sequences ≥1 %. FOR163 and FOR169 displayed an IGLV3-25 sequence_A (FOR163 = 18.36 %, FOR169 = 33.26 %) and FOR136 a sequence which was not ambiguously assigned to IGLV3-25 or IGLV3-16 (13.20 %).

Based on the detected IGLV subfamily distribution, it should be mentioned that the standard multiplex approach has been shown to bind on IGLV3-25 (FOR111). In general, both the

VLKL3c_huhn (20 nucleotides) and VLKL3_A_fw_NB (19 nucleotides) oligonucleotides show no mismatch to the IGLV2-25*03 and IGLV3-25*03 reference sequences. Regarding IGLV3-27: the VLKL3c_Huhn oligonucleotide shows only one (3rd position: A vs. T) and the VLKL3_A_fw_NB oligonucleotide none mismatch to the IGLV3-27 reference sequence. Therefore, the amplification of these IGLV subfamilies should also be ensured in the first MP$_{_NGS_PCR}$ approach.

Table 28. Evaluation of the sample composition of the IGLV3-21 next-generation sequencing experiment based on an IGLV3-specific forward oligonucleotide. (N_NGS_PCR) Seven AL amyloidosis samples from patients with dominant heart disease and IGLV3-21 assigned light chain sequences were examined. pC = positive control, nC = negative control, short = no N-terminally longer sequence with additional oligonucleotides could be generated, sequence_A = sequence with the highest percentage in the next-generation sequencing analysis. The column "X AA [n]" defines how many amino acids were not determined unambiguously.

Sample	Number of detected sequences >1 %	IGLV subfamily sequence_A (%)	Number of IGLV3-21 sequences > 1 %	all sequences >1 % IGLV3-21 [%]	Most prominent IGLV3 subfamily (%)
FOR104_pC	1	IGLV3-21 (99.91)	1	99.9	IGLV3-21 (99.91 %)
FOR136	23	IGLV3-16/IGLV3-25 (13.20)	0	0	IGLV3-25 (32.25 %)
FOR163	17	IGLV3-25 (18.36)	0	0	IGLV3-25 (50.04 %)
FOR169	9	IGLV3-25 (33.26)	0	0	IGLV3-25 (66.24 %)
FOR176	14	IGLV3-10 (17.77)	1	6.1	IGLV3-27 (49.61 %)
FOR177	20	IGLV3-21 (16.51)	1	16.5	IGLV3-27 (28.99 %)
FOR113_nC	6	IGLV3-25 (55.15)	0	0	IGLV3-25 (79.04 %)

3.6.3.2 Sequence Analysis and Comparison to the MP$_{_NGS_PCR}$ Approach

In the analysis of the three samples comprising an IGLV3-21 sequence in both the MP$_{_NGS_PCR}$ and N$_{_NGS_PCR}$ approach, a sequence match was found – including the positive control sample FOR104_pC (**Figure 25**).

Figure 25. Sequence alignment of the IGLV3-21 sequence with the highest percentage in the MP$_{NGS_PCR}$ approach and the IGLV3-21 sequences in the N$_{NGS_PCR}$ approach. (A) sequence alignment positive control FOR104_pC (B) sequence alignment FOR176 (C) sequence alignment FOR136 (D) sequence alignment FOR177. CDR region = underlined, IGLJ segment = blue highlight, linker region = green, sequence deviation = red, * = matching position, VLKL3c_huhn = IGLV3 specific forward oligonucleotide, JLHD123_rv = IGLJ segment specific reverse oligonucleotide, Sanger = sequence generated by Sanger sequencing, NGS = sequence with the highest percentage in the next-generation sequencing experiment, next-generation sequencing sequences were designated in descending alphabetical order based on the percentage.

In the remaining four cases (including FOR113_nC), no IGLV3-21 sequence was detected in the N$_{NGS_PCR}$ approach. To exclude the possibility of erroneous assignment, the sequence_A was compared between the two approaches. In the analysis of FOR136, 43 deviations in a total of 171 overlapping nucleotides were detected between the respective sequence_A sequences. Similar ratios were detected in FOR163 (43 deviations, 168 overlapping nucleotides) and FOR169 (47 deviations, 178 overlapping nucleotides) (**Supplementary Information Figure 22**).

In the next step, it was evaluated if the sequence_A of the N_NGS_PCR approach was also present in the MP_NGS_PCR approach (**Supplementary Information Figure 23**). This analysis could not be performed for FOR136, because only one sequence was detected in the MP_NGS_PCR approach (IGLV3-21). In the analysis of the N_NGS_PCR FOR163 sample, an IGLV3-25 sequence_A was detected and, in the MP_NGS_PCR approach sequence_D was assigned to this subfamily. However, the sequences showed several deviations (22 in 178 overlapping nucleotides). This was also observed for FOR169; the N_NGS_PCR sequence A was assigned to the IGLV3-25 subfamily as well as the MP_NGS_PCR sequence_B, sequence_C, and sequence_D. Sequence_D showed the highest sequence similarity with nine deviations (172 overlapping nucleotides).

In the analysis of the negative control sample FOR113_nC, N_NGS_PCR sequence_A, _B, and _C, as well as the MP_NGS_PCR sequence_C, _D, and _N, were assigned to the IGLV3-25 subfamily. No concordance could be found between the sequences of both approaches. When comparing the N_NGS_PCR IGLV3-25 sequence with the Sanger sequencing result, no overall matching sequence was found. However, taking all three sequences into account, no position showed a deviation in all three sequences – except one, which can be explained by the reverse oligonucleotide binding site.

Taken together, the results concerning the FOR104_pC as well as the FOR113_nC N_NGS_PCR samples match the Sanger sequencing results. In two additional samples, the IGLV3-21 sequence_A of the MP_NGS_PCR approach is in accordance with a detected N_NGS_PCR IGLV3-21 sequence. In the remaining three cases no IGLV3-21 sequence was detected in the N_NGS_PCR approach and in two cases the respective sequence_A showed several deviations from sequences of the same IGLV subfamily in the MP_NGS_PCR approach.

3.6.4 Sequence Alignment

A more detailed analysis of the LC sequences was performed on AA level to identify potential unique mutations or mutation patterns that could influence the biochemical properties of the LC itself or discriminate the subgroups. This was carried out in a step-wise procedure. First, general sequence characteristics such as mutation hotspots (mutations >50 %) and their location were identified separately for both diseases. Then the hotspots were analyzed in terms of an exchange towards a specific AA or AAs with the same side chain properties and the respective cDNA. In this analysis, the LC sequences with more than ten unambiguously determined AAs (marked as X) were included.

When analyzing the sequences in more detail, it was noticed that there seem to be specific positions – especially in the FR regions – that are particularly characterized by nucleotide signal overlaps: 38K (FR2, 3/10 X), 44V (FR2, 3/10 X), 63G (FR3, 3/10 X), 68N (FR3, 3/10 X), 72L

(FR3, 3/10 X), 83A (FR3, 3/10 X), 87C (FR3, 3/10 X) and 89V (CDR3, 3/10 X) and most prominent 84D (FR3, 4/10 X) (**Figure 26** A).

The AL and MM sequences each showed only one mutation hotspot in the IGLV segment, which was also exclusive to the disease (**Figure 26** A). For the AL hotspot 92S (6/10, 1x X, CDR3), a most prominent exchange from serine towards threonine (3/6, cDNA = AGT→ACT) was noted. The MM sequences presented a 92S mutation only twice (1x G, 1x F). The MM hotspot was also located in the CDR3 region (93S) and a mutation was detected in seven out of eleven cases (AL = 4/10) with a most prominent exchange towards a charged AA (2x R, 1x D, AL = 0).

Besides these two hotspots, two additional positions with a directed exchange pattern should be mentioned. A G80E (FR3) mutation was detected in four AL sequences (MM = 0x) but this position could not be verified in the comparison between the Sanger sequencing and $MP_{\_NGS\_PCR}$ approach. In contrast, the MM sequences showed a Y86F mutation in four cases (1x H, FR3) and none of the AL sequences.

Overall, an additional charge in the CDR3 region was detected in five MM cases and only once in the AL cohort (**Figure 26** B). In this context, the IGLV-IGLJ linker region was not considered, based on the fact that a similar linker region was detected for both diseases. In general, the Ensembl IGLV3-21 reference showed four additional AAs (DHPT) at the C-terminus than the Vbase2 reference sequences. However, most of the AL and MM sequences showed only two AAs in the linker region, and in more than half of the cases, they were identified as aspartic acid and histidine.

A mutation hotspot for both diseases was also defined at the first position of the IGLJ segment (AL = 5/10, 2x X, MM = 7/11). The MM sequences showed a mutation towards the small AAs proline or glycine in two cases each – this was not detected in the AL cohort.

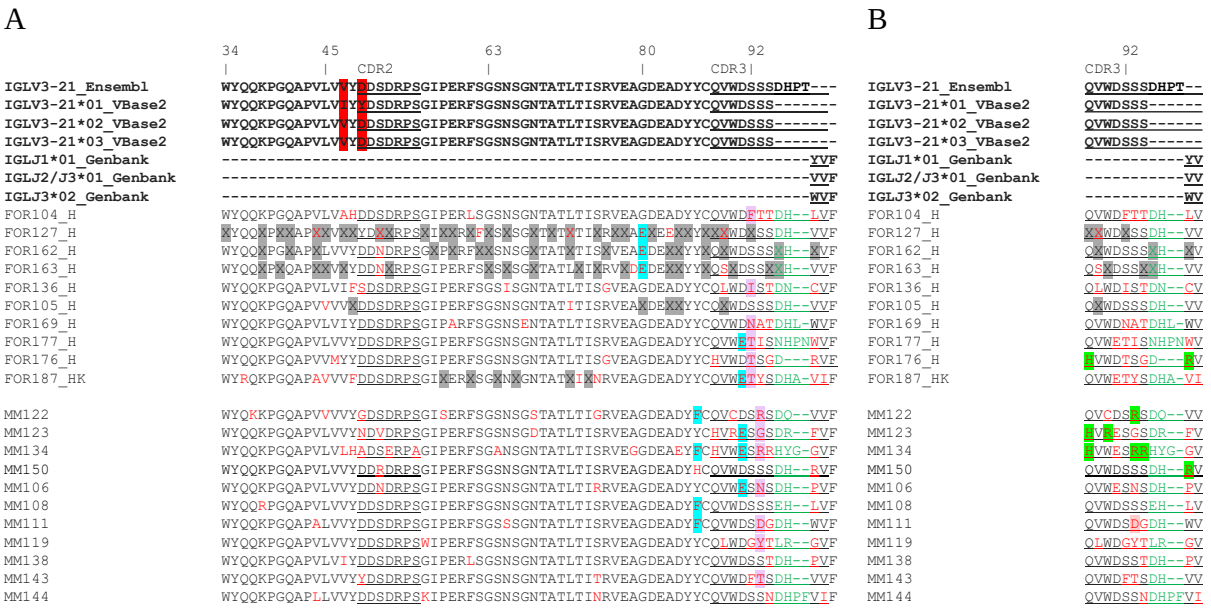


Figure 26. Sequence sections of IGLV3-21 assigned AL amyloidosis and multiple myeloma light chain sequences. A) Sequence sections of IGLV3-21 assigned AL amyloidosis and multiple myeloma light chain sequences B) CDR3 region of IGLV3-21 AL amyloidosis and multiple myeloma sequences with respect to additional charges C) Sequence sections of IGLV3-21 AL amyloidosis sequences stratified for the presence or absence of a clonal heavy chain in serum. Bold = reference sequences, underlined = CDR regions, red highlight = discrepancy between the VBase2 and Ensembl IGLV3-21 reference, red letter = mutation, purple highlight = mutation hotspot, blue highlight = interesting mutation, green highlight = additional positive charge, light red highlight = additional negative charge, X and grey highlight = not unambiguously determined amino acid, green letter = linker region, MM = multiple myeloma patients, AL = AL amyloidosis patients, _H = AL amyloidosis patient with dominant heart involvement, _HK = AL amyloidosis patient with dominant heart and kidney involvement, HC = detectable clonal heavy chain in the immunofixation electrophoresis in serum. Only one IGLC reference is shown because the first amino acid can be defined as glycine in all cases. Amino acids were numbered according to the VBase2 reference. The complete amino acid sequence alignment is shown in Supplementary Information Figure 24. The complete cDNA sequence alignment is shown in Supplementary Information Figure 25.

Interestingly, the sequences with ambiguous AA positions were mostly assigned to the AL HC group (3/4 vs. 2/6). In addition, the AL hotspot at position 92S was enriched in AL no HC sequences (5/6 vs. 1/4) as well as a mutation at the first position of the IGLJ segment (4/6 vs. 1/4, 1x X AL no HC) (**Figure 27**). No specific mutation pattern or sequence characteristics were noted when stratifying for a dFLC >180 mg/L.

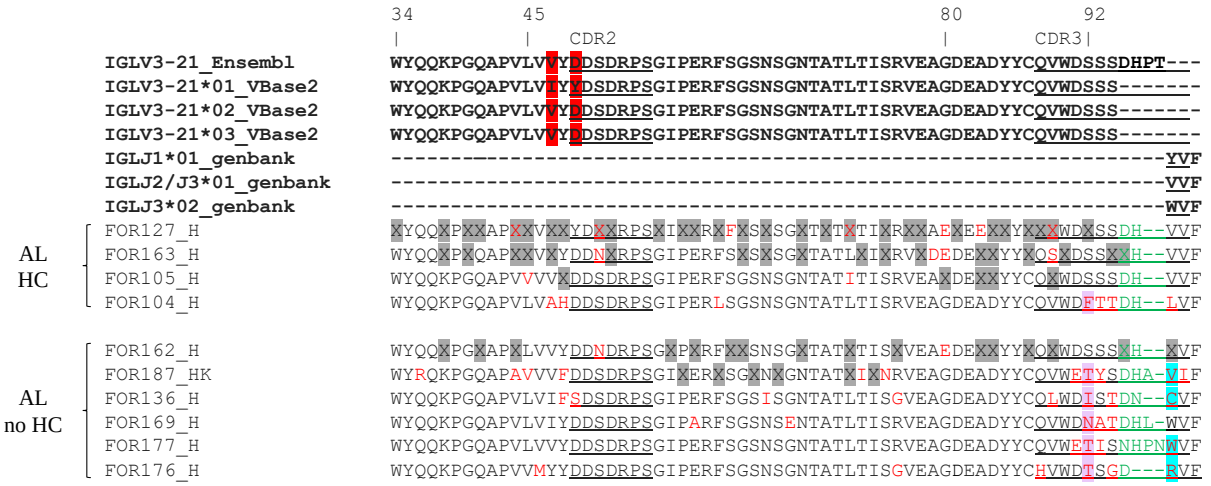


Figure 27. Sequence sections of IGLV3-21 AL amyloidosis sequences stratified for the presence or absence of a clonal heavy chain in serum. Bold = reference sequences, underlined = CDR regions, red highlight = discrepancy between the VBase2 and Ensembl IGLV3-21 reference, red letter = mutation, light blue highlight = mutation hotspot, dark blue highlight = interesting mutation, X and grey highlight = not unambiguously determined amino acid, green letter = linker region, MM = multiple myeloma patients, AL = AL amyloidosis patients, _H = AL amyloidosis patient with dominant heart involvement, _HK = AL amyloidosis patient with dominant heart and kidney involvement, HC = detectable clonal heavy chain in the immunofixation electrophoresis in serum. Amino acids were numbered according to the VBase2 reference. The complete amino acid sequence alignment is shown in Supplementary Information Figure 24. The complete cDNA sequence alignment is shown in Supplementary Information Figure 25.

3.6.5 Summary IGLV3-21

IGLV3-21 was the most common MM IGLV subfamily (21 %, n = 11; AL = 13 %, n = 10) and was found more frequently in AL_H than AL_HK sequences (15 % vs. 5 %). In addition, an association with MM_HC and MM < dFLC was detected. Also, the IGLV-IGLJ-IGLC connection differed between the two diseases. While most of the MM sequences displayed an IGLJ2/IGL2 linkage, the AL sequences showed an IGLJ2/IGLC2 or IGLJ3/IGLC3 linkage equally frequently. In addition, AL HC sequences showed mostly an IGLJ2/IGLC2 connection, and AL no HC sequences an IGLJ3/IGLC3 connection.

Several difficulties occurred in the sequencing of the IGLV segment of the AL sequences: In the majority of the cases, it was not possible to resolve the N-terminus (FR1-CDR1) of the IGLV segment. Further, IGLV3-21 presented as the IGLV subfamily with the most abundant and highest number of ambiguous identifiable AAs. Only four of eleven AL sequences did not show nucleotide signal overlaps resulting in undeterminable AAs. Based on these characteristics, an evaluation regarding the mutation distribution, AA composition, and biophysical parameters was assumed to not be reasonable.

In the next step, NGS experiments were conducted with seven AL sequences to verify the IGLV3-21 subfamily assignment and further investigate the composition of these sequences/samples. These seven samples included an evaluable N-terminal resolved sequence

as the positive control and as the negative control a sequence that was defined as non-evaluable (IGLV3-21 in the first PCR reaction and IGLV3-25 in the second PCR reaction).

At first, NGS was used to analyze a PCR approach based on a multiplex oligonucleotide set ($MP_{\_NGS\_PCR}$). It was possible to detect an IGLV3-21 sequence with the highest percentage for each IGLV3-21 assigned sample (median sequence A: 93.4 %). The negative control sample showed an IGLV3-9 sequence with the highest percentage (3.3 %), but considering all sequences ≥1 %, IGLV3-21 was the most prominent IGLV subfamily. Four samples, including the positive control, showed only one sequence with a percentage ≥1 % and in a comparison with the Sanger sequencing results, the sequences matched (CDR2-IGLJ). Two samples displayed four sequences ≥1 %; in one sample the NGS and Sanger sequencing matched (sequence_A: 93.4 % IGLV3-21). In the second case seven mismatches, mainly restricted to the CDR3 region, were noted (sequence_A: 1.4 % IGLV3-21). The negative control showed 16 sequences with a percentage ≥1 %. Evaluating these 16 sequences together, four positions showed a deviation from the Sanger sequence. Thus, the IGLV subfamily assignment was verified in all cases and in the majority of the cases also the sequence itself. In the remaining samples, the sequences were verified for over 90 % of the positions.

In the next step, NGS was used to analyze PCR samples that should generate IGLV3 N-terminally elongated sequences. Interestingly, these $N_{\_NGS\_PCR}$ samples contained a higher amount of sequence ≥1 % (14 vs. 1; $p = 0.038$) and a lower percentage of the respective sequence_A (18.4 % vs. 93.4 %) compared to the $MP_{\_NGS\_PCR}$ approach. However, only three samples showed an IGLV3-21 sequence ≥1 %. Here, the IGLV3-21 sequence was detected with the highest percentage in two sequences, including the positive control. In all three cases, the IGLV3-21 $N_{\_NGS\_PCR}$ sequence matched the $MP_{\_NGS\_PCR}$ IGLV3-21 sequence_A. It is worth mentioning that for the negative control, the sequence with the highest percentage was identified as IGLV3-25, which matched the Sanger sequencing assignment. Taking all three $N_{\_NGS\_PCR}$ IGLV3-25 sequences into account, no position showed a deviation in all three sequences. So, for the positive and negative control it was possible to generate expected results but beyond that no new information could be gained.

In a next step, the complete AL and MM cohorts were analyzed concerning their mutation hotspots. Here, it was noticeable that the AL (92S) and MM (93S) hotspots were both located in the CDR3 but did not concern the same position. In addition, two positions showed a direct exchange pattern: G80E (AL) and Y86F (MM), but the NGS experiments could not verify the G80E mutation. As already described for the other IGLV subfamilies, it was possible to identify different patterns in the context of charged and small AAs. While the CDR3 region of MM

sequences displayed an additional charge in five out of eleven cases, this was only noted once in the AL cohort. Even though the first AA of the IGLJ segment presented as a mutation hotspot in both cohorts, an exchange towards a small AA was detected in more than half of the cases in the MM sequences and none of the AL sequences. It should also be mentioned that the AL sequences with many ambiguous AA positions belonged to the AL_HC group (4/5 vs. 2/6).

3.7 IGLV6-57 and IGLV2-23

Concerning the IGLV subfamily usage, a significant difference between the two diseases was especially detected for IGLV6-57 and IGLV2-23. IGLV6-57 was found to be the most used AL IGLV subfamily (n = 18) and was only detected once in the MM cohort. In addition, IGLV6-57 presented as the most used AL_HK subfamily and was detected with a more than two times higher percentage than in the AL_H subgroup. However, half of the AL_HTX sequences corresponded to this subfamily (4/8). In contrast, eight MM samples and no AL samples were assigned to the IGLV2-23 subfamily. This is of particular interest due to IGLV2-23 being detected as one of the third most commonly used MM IGLV subfamilies (together with IGLV2-14). Based on this prominent difference and the assumed divergent amyloidogenic potential, a comparison between these two IGLV subfamilies was performed.

The AL IGLV6-57 sequence FOR133 was not used for further analysis due to fourteen not clearly determined AAs. In addition, since only one MM sequence was assigned to the IGLV6-57 subfamily this sequence was also neglected in the evaluation (MM112) – however, it was possible to verify the sequence by Sanger sequencing (**Supplementary Information Figure 26**).

In the bioinformatic analysis of the IGLV2-23 MM sequences MM107 and MM132, it was possible to detect an additional sequence segment. In the case of MM107, this sequence section was not clearly assigned to a specific region but showed partial matches with the IGLC segment (**Supplementary Information Figure 27**). Regarding MM132, the additional sequence section was an exact copy of an IGLC region and was therefore not investigated further (**Supplementary Information Figure 28**).

3.7.1 Subgroup Analysis

More than half of the AL IGLV6-57 sequences corresponded to patients with detectable HC (11/17, 6x IgG, 4x IgA, 1x IgD, 1x NA) and nine to patients with a dFLC >180 mg/L (**Table 29**). Since most IGLV6-57 AL_HK sequences corresponded to a patient with detectable HC (7/8), the respective analysis was performed using only the IGLV6-57 AL_H sequences.

Seven out of eight IGLV2-23 MM patients showed a detectable clonal HC (5x IgG, 2x IgA), and six patients a dFLC >180 mg/L. Therefore, a comparison between these subgroups was not performed.

As a side note, the distribution of HC classes is in line with the approximate percentage of the total Ig in the adult serum for both diseases (Lefranc and Lefranc 2020).

Table 29. Overview of selected characteristics of 18 AL amyloidosis patients and one multiple myeloma patient with IGLV6-57 assigned light chain sequences and 8 multiple myeloma patients with IGLV2-23 assigned light chain sequences. The column "X AA [n]" defines how many amino acids were not determined unambiguously. Organ inv. = organ involvement, AL_H = AL amyloidosis patients with dominant heart involvement, AL_HK = AL amyloidosis patients with dominant heart and kidney involvement, AL_HTX = AL amyloidosis patients who received a heart transplant, AA = amino acid, AL = AL amyloidosis, MM = multiple myeloma, HC in IFE = detectable clonal heavy chain in the immunofixation electrophoresis in serum, NA = not available.

Patient	Disease	Organ inv.	IGLJ	IGLC	X AA [n]	HC in IFE	dFLC [mg/L]
	IGLV6-57						
FOR117	AL	AL_HTX	IGLJ2	IGLC2	0	-	80.9
FOR126	AL	AL_H	IGLJ3	IGLC2	0	G	36.8
FOR133	AL	AL_H	IGLJ2/3	IGLC3	14	-	**633.6**
FOR140	AL	AL_HTX	IGLJ3	IGLC2	0	-	**895.7**
FOR144	AL	AL_H	IGLJ2	IGLC2	0	-	**224.5**
FOR192	AL	AL_H	IGLJ2	IGLC2	0	G	155.3
FOR194	AL	AL_H	IGLJ3	IGLC3	0	A	45.5
FOR197	AL	AL_HTX	IGLJ2	IGLC2	3	G	**442.4**
FOR205	AL	AL_H	IGLJ2/3	IGLC3	0	-	117.2
FOR214	AL	AL_HTX	IGLJ3	IGLC3	0	NA	**484.0**
FOR150	AL	AL_HK	IGLJ3	IGLC2	0	A	86.2
FOR152	AL	AL_HK	IGLJ3	IGLC3	0	G	**211.5**
FOR153	AL	AL_HK	IGLJ3	IGLC3	0	G	103.2
FOR154	AL	AL_HK	IGLJ2	IGLC2	0	G	**202.3**
FOR188	AL	AL_HK	IGLJ3	IGLC3	0	A	85.9
FOR185	AL	AL_HK	IGLJ2	IGLC2	0	A	**754.8**
FOR222	AL	AL_HK	IGLJ2	IGLC2	0	D	168.2
FOR228	AL	AL_HK	IGLJ2	IGLC2	0	-	**218.5**
MM112	MM	-	IGLJ3	IGLC3	0	G	153.8
	IGLV2-23						
MM107	MM	-	IGLJ2/3	IGLC2	0	A	**2396.0**
MM114	MM	-	IGLJ3	IGLC3	0	G	**933.8**
MM117	MM	-	IGLJ2	IGLC2	0	-	**3172.1**
MM125	MM	-	IGLJ3	IGLC3	0	G	**496.0**
MM132	MM	-	IGLJ2/3	IGLC2	0	G	**1496.9**
MM133	MM	-	IGLJ3	IGLC3	0	A	**288.5**
MM135	MM	-	IGLJ2	IGLC2	0	G	28.3
MM149	MM	-	IGLJ2/3	IGLC3	0	G	117.9

Starting with the overall LC composition, the AL IGLV6-57-IGLJ-IGLC linkage has already been described in chapter 3.3.2 and a most prominent linkage with IGLJ2/IGLC2 was detected

(**Table 19**). Three IGLV2-23 MM sequences displayed an IGLJ3/IGLC3 linkage, two an IGLJ2/IGLC2 linkage, and in three cases the IGLJ segment was not determined unambiguously between IGLJ2 and IGLJ3.

3.7.2 Mutation Frequency and Count

To examine whether the difference in LC behavior is based on a difference in the general mutation distribution and frequency, these aspects were analyzed within the IGL segments.

Comparing the AL IGLV6-57 and IGLV2-23 MM sequences, it must be considered that the IGLV segment of the IGLV6-57 sequences was trimmed and the FR1, as well as the first AA of the CDR1 region, were missing. In the following, only the adjusted IGLV2-23 sequences were considered. However, with respect to the analyzed section, the IGLV6-57 reference comprises one additional AA (CDR1-CDR3, IGLV6-57 = 75 AA, IGLV2-23 = 74 AA). Besides, the FR1 of IGLV2-23 MM sequences showed a mutation in 38 % of the cases and, if mutated, a median mutation count of 1.0.

Overall, the IGLV2-23 MM sequences displayed a mutational load almost twice as high as the IGLV6-57 AL sequences (11.5 vs. 6.0, p = 1.60e-06 (t-test)) (**Table 30**). However, not only a higher overall mutation count but also a higher mutation frequency and mutation count was detected in all regions, except CDR3. The most prominent differences were noted in CDR1 (3.5 vs. 1.0, p = 0.003 (t-test), IGLV6-57 = 12 AA, IGLV2-23 = 13 AA), CDR2 (88 % vs. 53 %, p = 0.027 (U-test), both references = 7 AA) and FR3 (100 % vs. 65 %, p = 0.023 (U-test), IGLV6-57 = 34 AA, IGLV2-23 = 32 AA). Additionally, a more frequently mutated IGLJ segment was detected in the MM IGLV2-23 sequences (75 % vs. 53 %, reference = 12 AA).

When stratifying the IGLV6-57 sequences for the AL organ tropism, the AL_H sequences (IGLV = 7.0, IGLJ = 78 %, IGLC = 22 %, reference = 74 AA) and especially the AL_HTX sequences (IGLV = 7.5, IGLJ = 75 %, IGLC = 25 %) showed a higher mutational load or frequency than the AL_HK sequences in all segments (IGLV = 5.0, IGLJ = 25 %, IGLC = 0 %)(**Table 30**). In addition, a higher mutation frequency was detected in the FR3 of AL_H sequences compared to AL_HK sequences (100 % vs. 25 %, p = 0.007 (t-test), reference = 34 AA).

In the analysis of IGLV6-57 AL_H sequences regarding a potential HC binding partner, the AL_H HC sequences presented less frequently mutated than the AL_H no HC sequences in the IGLV (6.5 vs. 7.5), IGLJ (50 % vs. 100 %) and IGLC segment (0 % vs. 50 %) (**Table 30**). This difference was also noted for CDR2 (25 % vs. 75 %).

When stratifying for a dFLC >180 mg/L, the IGLV6-57 AL < dFLC sequences showed, if mutated, more mutations in the IGLJ segment (2.0 vs. 1.0). Despite no overall detected difference in the IGLV segment, the FR3 showed more mutations in the AL > dFLC subgroup (1.0 vs. 2.5) (**Table 30**).

Table 30. Comparison of the percentage of mutated IGL segments and the average mutation count between the IGLV6-57 assigned AL amyloidosis and IGLV2-23 assigned multiple myeloma light chain sequences. The analysis was also performed with respect to different subgroups. The IGLV6-57 MM light chain sequence is not shown. AA = amino acid, CDR = complementary determining region, FR = framework region, AL = AL amyloidosis patients, MM = multiple myeloma patients, AL_H = AL amyloidosis patients with dominant heart involvement, AA = amino acid, AL_HK = AL amyloidosis patients with dominant heart and kidney involvement. HC = detectable clonal heavy chain in the immunofixation electrophoresis in serum, > dFLC = dFLC >180 mg/L, < dFLC = dFLC <180 mg/L, * = p ≤0.05, differences greater than 30 % and one (segments) or two (regions) additional mutations are marked in bold. The median mutation values were calculated using only the mutated segments/regions. The significance level was determined using the complete dataset. FOR133 was excluded due to several not unambiguously defined amino acids. The median mutation values were calculated using only the mutated segments/regions. The Vbase2 reference sequences were used for calculations, the IGLC reference was trimmed at the C-terminus. The CDR3 region also includes the patient-specific linker region and the first two amino acids of the IGLJ segment. It, therefore, spans 9-11 amino acids in the IGLV6-57 subgroup and 8-11 amino acids in IGLV2-23.

			mutated segments [%] median mutation count [n]							
		n	IGLV						IGLJ 12 AA	IGLC 74 AA
IGLV6-57 / IGLV2-23 AA			75/71	CDR1 12/13	FR2 15/15	CDR2 7/7	FR3 34/32	CDR3		
AL	[%]	17	100	82	59	**53**	**65**	100	53	12
– IGLV6-57	[n]		**6.0***	**1.0***	1.0	2.0*	2.0*	2.0	1.0	1.0
AL_H	[%]	9	100	78	67	44	**100**	100	**78**	22
– IGLV6-57	[n]		**7.0**	1.0	2.0	1.0	2.0*	2.0	**1.0**	**1.0**
AL HTX	[%]	4	100	100	75	50	100	100	**75**	25
– IGLV6-57	[n]		**7.5**	1.0	2.0	1.5	2.5	1.5	**1.0**	**1.0**
AL_HK	[%]	8	100	88	50	63	**25**	100	**25**	0
– IGLV6-57	[n]		**5.0**	1.0	1.0	2.0	1.5*	2.0	**2.0**	**0**
MM	[%]	8	100	100	88	**88**	**100**	100	75	13
– IGLV2-23	[n]		**11.5***	**3.5***	1.0	2.0*	2.5*	2.5	1.0	1.0
AL_H HC	[%]	4	100	75	50	**25**	100	100	**50**	**0**
– IGLV6-57	[n]		**6.5**	1.0	1.5	1.0	2.0	2.0	1.5	**0**
AL_H no HC	[%]	4	100	75	75	**75**	100	100	**100**	**50**
– IGLV6-57	[n]		**7.5**	2.0	2.0	1.0	2.5	1.5	1.0	**1.0**
AL > dFLC	[%]	8	100	88	**75**	38	50	100	50	13
– IGLV6-57	[n]		6.0	1.0	1.5	2.0	**2.5**	2.0	**1.0**	1.0
AL < dFLC	[%]	9	100	78	**44**	67	78	100	56	11
– IGLV6-57	[n]		6.0	2.0	1.0	1.0	**1.0**	2.0	**2.0**	1.0

3.7.3 Sequence Alignment

A more detailed analysis of the LC sequences was performed on AA level to identify potential unique mutations or mutation patterns that could influence the biochemical properties of the LC itself or discriminate the subgroups. This was carried out in a step-wise procedure. First, general sequence characteristics such as mutation hotspots (mutations >50 %) and their location were identified separately for both diseases. Then the hotspots were analyzed in terms of an exchange towards a specific AA or AAs with the same side chain properties and the respective cDNA. In this analysis, the LC sequences with more than ten unambiguously determined AAs (marked as X) were included.

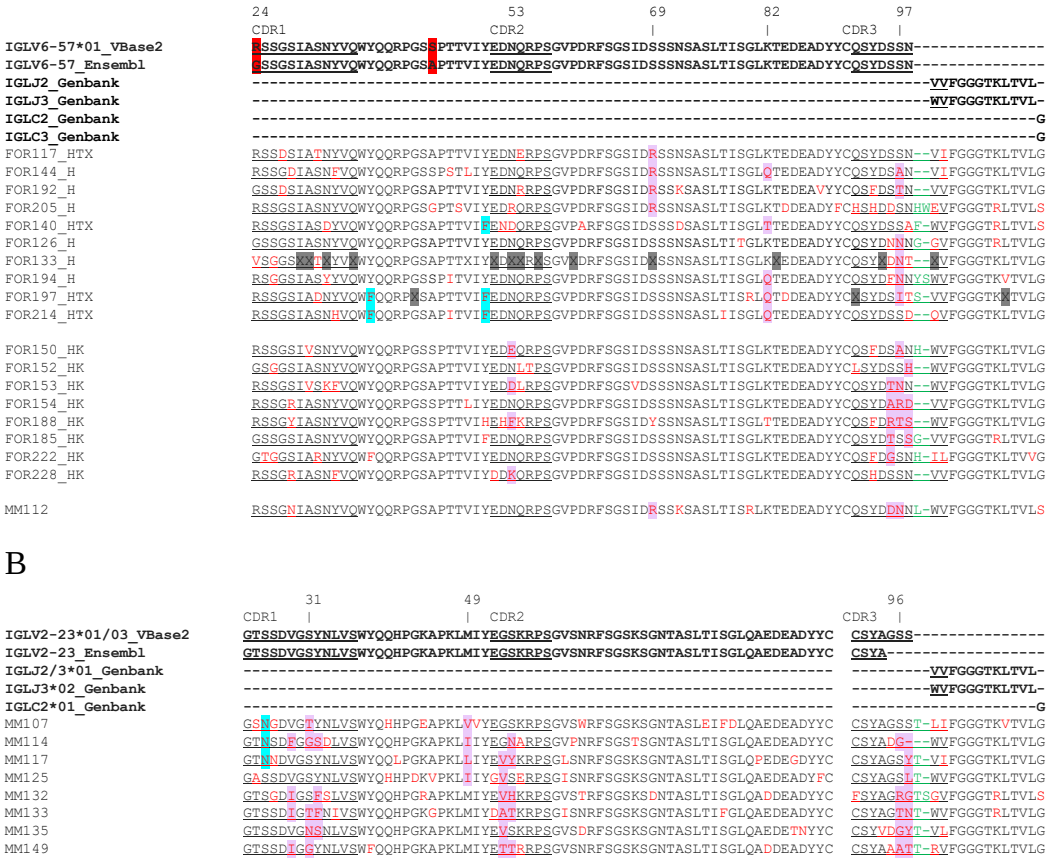
When analyzing the IGLV6-57 sequences in more detail, it is noticeable that only one position in the CDR3 region is mutated in over 50 % of the AL sequences (97S), including four out of eight AL_HK and six out of nine AL_H sequences (**Figure 28** A). When considering the organ involvement separately, three hotspots were additionally defined for the AL_HK sequences: 53N (3rd position CDR2, 4/8), 96S (CDR3, 5/8), and 98N (CDR3, 4/8). While at position 53N an exchange towards a charged AA was noted in three out of four AL_HK cases, no directed exchange pattern was observed for the other two hotspots. In contrast, two additional AL_H hotspots were identified, and both showed a specific exchange pattern. At position 69S (FR3), four out of nine AL_H sequences displayed an exchange towards the positively charged AA arginine (1x X with R as a possibility, cDNA = AGC→AGG or AGA). At position 82 (FR3), an exchange from the positively charged AA lysine towards glutamine was detected four times and one time towards threonine (cDNA glutamine = AAG→CAG).

Overall, no consistent pattern was noted in the AL IGLV6-57 IGLV-IGLJ linker region.

When stratifying the AL_H sequences for the absence or presence of a clonal HC, it is noticeable that an exchange towards phenylalanine was detected in three out of four AL_H HC sequences at position 50Y and at position 37Y twice. Remarkably, the two Y50F and Y37F mutated sequences corresponded to AL_HTX patients. All Y50F mutated sequences also cluster to the > dFLC subgroup (cDNA Y50F = TAT→TTT).

Regarding the IGLV2-23 MM sequences, eight positions – mainly restricted to the CDR regions – were identified as mutation hotspots: 29V (CDR1, 4/8), 31S (CDR1, 5/8), 32Y (CDR1, 4/8), 49M (FR2, 4/8), 53G (CDR2, 6/8), 54S (3rd position CDR2, 5/8), 96S (CDR3, 5/8) and 97S (CDR3, 7/8) (**Figure 28** B). However, only at two positions, a distinct exchange was noted: V29I = 3/4 (cDNA = GTT→ATT) and G53V = 4/6 (cDNA = GGC→GTC). Also, in three out of eight cases, an exclusive exchange towards asparagine was noted at position 26S. Although there is no overlapping reference, the linker region contained only a threonine in six cases (1x TS, 1x nothing).

A



B

Figure 28. Sequence sections of IGLV6-57 assigned AL amyloidosis and multiple myeloma light chain sequences and IGLV2-23 assigned multiple myeloma light chain sequences. (A) Sequence section of IGLV6-57 assigned AL amyloidosis and multiple myeloma light chain sequences. (B) Sequence section of IGLV2-23 assigned multiple myeloma light chain sequences. Bold = reference sequences, underlined = CDR regions, red letter = mutation, purple highlight = mutation hotspot, blue highlight = interesting mutation, X and grey highlight = not unambiguously determined amino acid, green letter = linker region, MM = multiple myeloma, _H = AL amyloidosis patient with dominant heart involvement, _HK = AL amyloidosis patient with dominant heart and kidney involvement, _HTX = AL amyloidosis patient who received a heart transplant. Only one IGLC reference is shown because the first amino acid can be in all cases defined as glycine. Amino acids were numbered according to the VBase2 reference. The complete IGLV6-57 amino acid sequence alignment is shown in Supplementary Information Figure 29. The complete IGLV6-57 cDNA sequence alignment is shown in Supplementary Information Figure 30. The complete IGLV2-23 sequence amino acid alignment is shown in Supplementary Information Figure 31. The complete IGLV2-23 sequence cDNA alignment is shown in Supplementary Information Figure 32.

In a comparison between AL IGLV6-57 and MM IGLV2-23, the MM IGLV2-23 sequences presented more mutation hotspots (1 vs. 8) and the two cohorts share mutation hotspots at the third position of the CDR2 (AL_HK) and at the two last positions in the IGLV segment (one only AL_HK). Moreover, mutations in the IGLJ segment as well as the first AA of the IGLC segment were detected in both subgroups with the same AA exchanges.

As already observed for the other IGLV subfamilies, the IGLV6-57 sequences presented charge changes in specific regions as well (**Figure 29**). More than half of the AL IGLV6-57 sequences presented an additional charge in the CDR1 (10/18) and interestingly, a certain organ tropism

was noted. While five AL_H sequences presented an exchange towards the negatively charged AA aspartic acid (1x histidine), four AL_HK sequences presented an additional positive charge (3x arginine, 1x lysine, p = 0.036 (Chi)). Also, the CDR2 region showed an insertion of an additional charge in eight AL IGLV6-57 sequences, which clustered to the < dFLC subgroup (6/9 vs. 2/9). In contrast, the CDR3 region displayed an enrichment of charged AAs especially prominent in AL_HK sequences (6/8, AL_H n = 1; p = 0.012 (Fischer)). Regarding the IGLV2-23 MM sequences, an accumulation of additional charges was only noted in the CDR3 region (4/8, CDR1 1/8, CDR2 1/8).

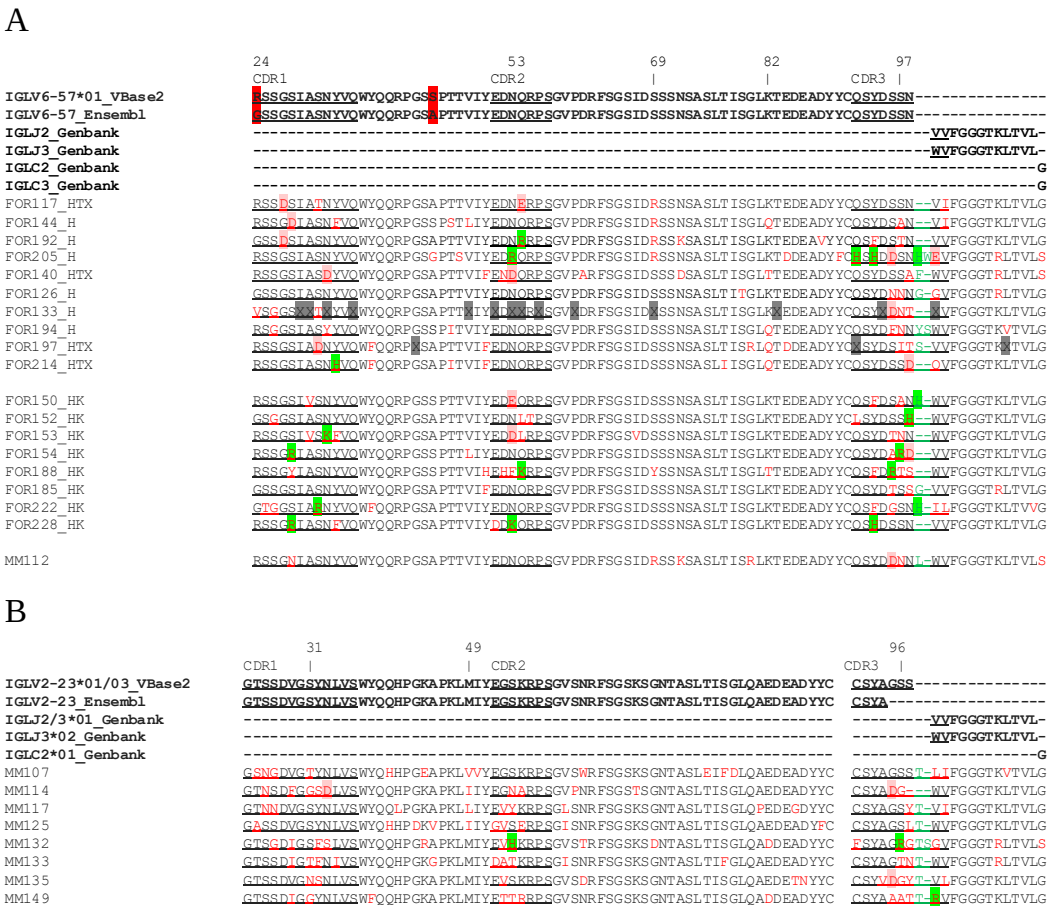


Figure 29. Sequence section of IGLV6-57 assigned AL amyloidosis and multiple myeloma light chain sequences and IGLV2-23 assigned multiple myeloma light chain sequences with respect to additional charges. (A) Sequence section of IGLV6-57 assigned AL amyloidosis and multiple myeloma light chain sequences. (B) Sequence section of IGLV2-23 assigned multiple myeloma light chain sequences. Bold = reference sequences, underlined = CDR regions, red letter = mutation, light red highlight = additional negative charge in a CDR region, green highlight = additional positive charge in a CDR region, X and grey highlight = not unambiguously determined amino acid, green letter = linker region, MM = multiple myeloma, _H = AL amyloidosis patient with dominant heart involvement, _HK = AL amyloidosis patient with dominant heart and kidney involvement, _HTX = AL amyloidosis patient who received a heart transplant. Only one IGLC reference is shown because the first AA can in all cases be defined as glycine. Amino acids were numbered according to the VBase2 reference. The complete IGLV6-57 amino acid sequence alignment is shown in Supplementary Information Figure 29. The complete IGLV6-57 cDNA sequence alignment is shown in Supplementary Information Figure 30.The complete IGLV2-23 sequence amino acid alignment is shown in Supplementary Information Figure 31. The complete IGLV2-23 sequence cDNA alignment is shown in Supplementary Information Figure 32.

3.7.4 Amino Acid Composition

The overall percentage of the individual AAs was calculated to evaluate the effect of the detected mutations and mutation patterns and to get an overview of the overall LC composition. Only differences ≥0.5 % are mentioned in the following, differences >0.5 % are numerically specified.

To achieve comparability between IGLV6-57 and IGLV2-23, the median AA percentage of the respective full-length reference sequences, without patient-specific linker region, was also calculated. Considering the reference VBase2 IGLV segment individually, it is striking that there is almost no commonality between these subfamilies. The IGLV6-57 reference contains an especially lower amount of alanine (3.1 % vs. 7.2 %) and glycine (6.1 % vs. 11.3 %) as well as a higher amount of aspartic acid (6.1 % vs. 3.1 %) than the IGLV2-23 reference. Overall, the IGLV6-57 reference comprises a higher number of negatively charged AAs (10 vs. 6). Also, differences regarding the positively charged AAs were noted; the IGLV2-23 reference sequence contained a lower amount of arginine (2.1 % vs. 4.0 %) and a higher amount of lysine (4.1 vs. 2.0).

In the IGLV6-57 AL as well as IGLV2-23 MM LC sequences no difference (≥ 0.5 %) was detected for the AAs cysteine, glutamine, glutamic acid, glycine, leucine, methionine, proline, and tryptophan for the full-length reference sequences (**Supplementary Information Table 6**, **Supplementary Information Table 7**). In contrast, in both the IGLV2-23 MM sequences and the IGLV6-57 AL, an increase in aspartic acid (AL = 4.8 % vs. 5.5 %, MM = 2.7 % vs. 3.3 %), isoleucine (AL = 3.0 % vs. 3.6 %, MM = 2.7 % vs. 3.3 %), phenylalanine (AL = 2.4 % vs. 3.0 %, MM = 2.2 % vs. 2.7 %) and threonine (AL = 7.9 % vs. 8.4 %, MM = 7.7 % vs. 8.2 %), as well as a decrease in serine (AL = 17.6 % vs. 16.3 %, MM = 17.5 % vs. 15.8 %), lysine (AL = 4.8 % vs. 4.2 %, MM = 6.0 % vs. 5.4 %) and tyrosine (AL = 5.5 vs. 4.8, MM = 4.9 % vs. 4.4 %) was noted in the complete LC.

Taken IGLV6-57 separately, an increase in alanine (7.9 % vs. 8.5 %), arginine (2.4 % vs. 3.0 %), and a decrease for asparagine (4.2 % vs. 3.6 %) in comparison to the reference was noted. In the IGLV2-23 sequences, the only exclusive difference was noted concerning valine which displayed a decrease compared to the reference (6.9 % vs. 6.3 %).

Considering the AL IGLV6-57 organ tropism separately, the AL_H sequences presented a higher amount of arginine, aspartic acid, and isoleucine (3.6 % vs. 3.0 %, p = 0.047 (t-test)) as well as a lower amount of histidine, and lysine (4.2 % vs. 4.8 %, p = 0.031 (t-test)) than the AL_HK sequences (**Figure 30**). In the AL_HTX sequences, a lower amount of glycine and a

higher amount of serine and phenylalanine than in the AL_H and AL_HK sequences were detected.

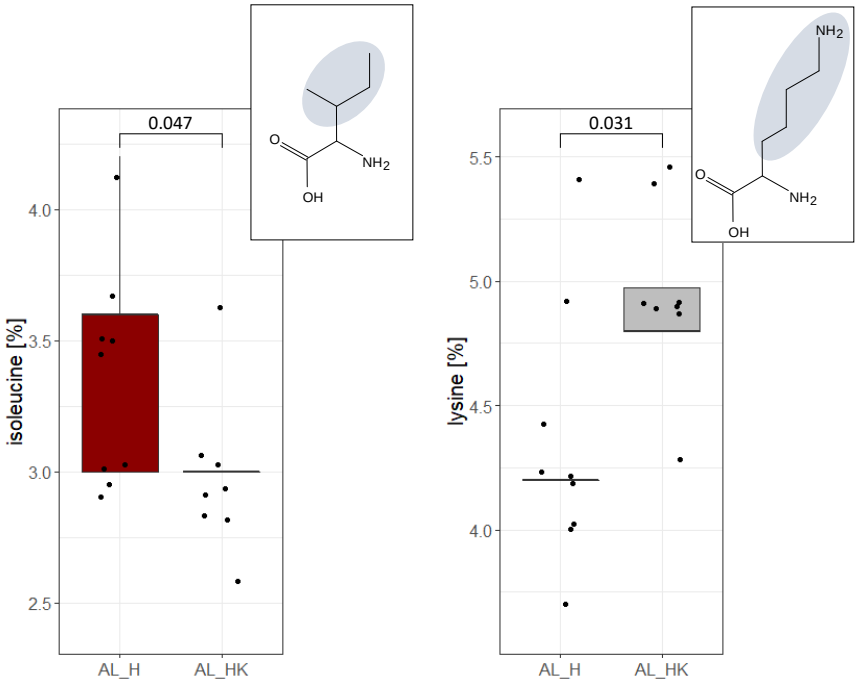


Figure 30. Comparison of the overall isoleucine and lysine percentage between IGLV6-57 assigned AL amyloidosis light chain sequences corresponding to patients with dominant heart or dominant heart and kidney involvement. The amino acid percentage was calculated using the complete trimmed light chain sequence. AL_H = AL amyloidosis patients with dominant heart involvement, AL_HK = AL amyloidosis patients with dominant heart and kidney involvement.

For the AAs aspartic acid and serine, a difference was observed between the AL subgroups in both the stratification for the presence of a HC and a dFLC >180 mg/L. While the AL_H HC sequences presented a lower percentage (aspartic acid = 4.8 % vs. 5.5 %, serine = 16.0 % vs. 16.8 %), a higher percentage was noted in the AL > dFLC sequences (aspartic acid = 5.5 % vs. 4.8 %, serine = 16.7 % vs. 15.8 % p = 0.012 (t-test)) (**Figure 31**). In addition, AL_H HC sequences showed a higher proportion of valine than AL_H no HC sequences. In AL > dFLC sequences, a higher percentage of alanine (8.5 % vs. 7.9 %; p = 0.015 (U-test)), arginine, glutamine, and a lower percentage of lysine, and tryptophan than in the respective other subgroup was noted.

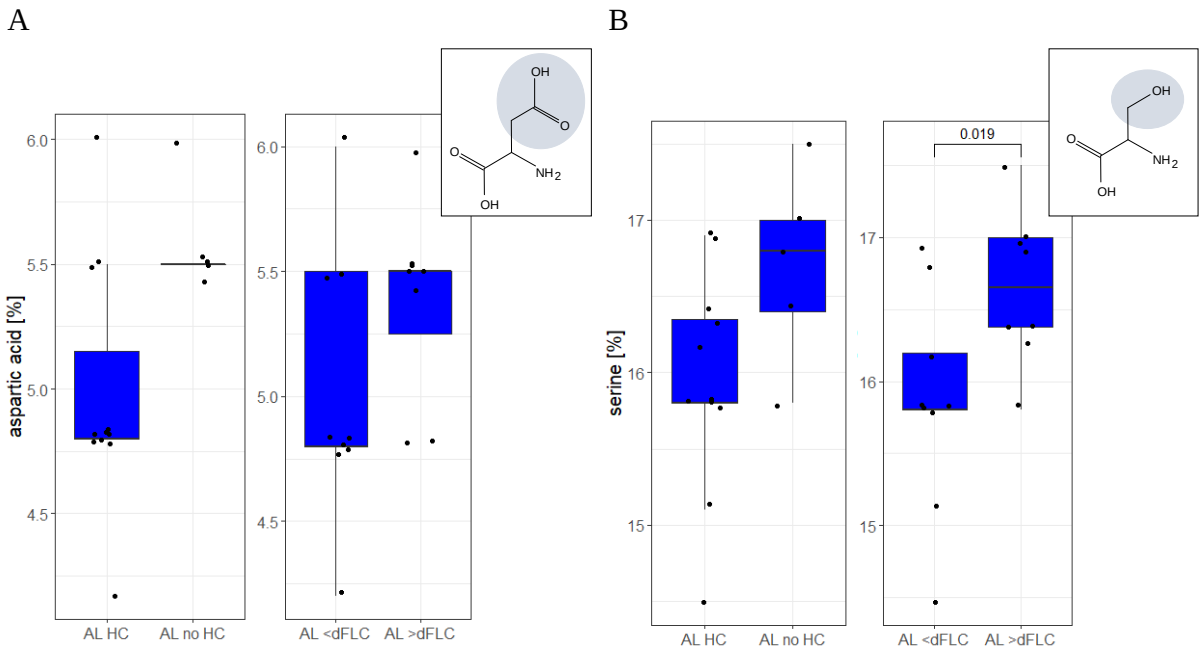


Figure 31. Comparison of the overall amino acid percentage of serine and aspartic acid between IGLV6-57 assigned AL amyloidosis light chain sequences with respect to the presence or absence of a clonal HC or a dFLC >180 mg/L. The amino acid percentage was calculated using the complete trimmed light chain sequence. MM = multiple myeloma patients, AL = AL amyloidosis patients, HC = heavy chain, > dFLC = dFLC >180 mg/L, < dFLC = dFLC <180 mg/L.

3.7.5 Biophysical Parameters

Several parameters were calculated to evaluate the effect of the mutations and amino acid composition on the biophysical properties of the LCs. The AGG parameter was included to determine the tendency for a beta sheet aggregation. The grand average of hydropathicity (GRAVY) value is defined by the sum of hydropathy values of all AA divided by the protein length. A negative GRAVY-value indicates a hydrophilic nature and a possible interaction of the linear protein with water molecules. The pI was also calculated to examine the theoretical stability of the LCs. In this context, the pI of the patient-derived LCs itself and the difference in comparison to a reference sequence (ΔpI) was calculated. This was performed for the complete LC (IGLVJC) and the IGLV and IGLJ segments (IGLVJ). In addition, the average molecular weight (Mw) of the LCs was specified.

In the IGLV2-23 reference sequence, a higher aggregation tendency (AGG score; 126 vs. 55) as well as a lower Mw than for the IGLV6-57 reference can be detected (10038 vs. 10643) (**Table 31**). Due to sequence trimming, a comparison between the LCs of the two diseases was not reasonable for the AGG value and the Mw. Since the GRAVY score is calculated by the sum of hydropathy values of all AAs, divided by the number of residues, a comparison between the two IGLV subfamilies can be performed. Here, a lower value was detected for the AL IGLV6-57 sequences (-0.434 vs. -0.223, reference IGLV = -0.751 vs. -0.452). Interestingly, the IGLV segment also shows a lower pI for the IGLV6-57 reference sequence than for the IGLV2-

23 reference (4.60 vs. 6.74). However, the MM IGLV2-23 full-length LC sequences displayed a more pronounced decrease in ΔpI in the IGLVJ (-0.55) and IGLVJC segments (-1.36) than the AL IGLV6-57 sequences (IGLVJC = -0.14, IGLVJ = 0.05).

A higher beta sheet aggregation tendency was detected in the IGLV6-57 AL_H (814) and particularly the AL_HTX (867) sequences compared to AL_HK sequences (655) (**Table 31**). In addition, the AL_H HC sequences presented a higher AGG value than the AL_H no HC sequences (852 vs. 762). No difference was found when stratifying for a dFLC >180 mg/L.

In the analysis of the GRAVY score, no major differences were found concerning the AL organ tropism, the presence of a HC, or a dFLC value >180 mg/L (**Table 31**).

Regarding the pI, the AL_HK sequences displayed almost no changes in pI and ΔpI in the full-length LC, while in the AL_H sequences a decrease was noted (**Table 31**). This difference was especially prominent when considering the AL_HTX sequences (pI IGLVJC 4.99 vs. 4.72, p = 0.043 (Median); pI IGLVJ 4.36 vs. 4.68, p = 0.043 (Median)). When stratifying for the presence or absence of a HC binding partner or a > dFLC, no difference for the pI or ΔpI was found in the AL sequences.

In the context of the Mw, no difference was detected when stratifying the AL sequence for a dominant organ involvement or the presence of a clonal HC. However, the AL > dFLC sequences displayed a higher average Mw than the AL < dFLC sequences (17712 vs. 17562) (**Table 31**).

Table 31. Overview and comparison between different biophysical parameters of IGLV6-57 assigned AL amyloidosis and IGLV2-23 assigned multiple myeloma light chain sequences and between different subgroups. AL = AL amyloidosis patients, MM = multiple myeloma patients, AL_H = AL amyloidosis patients with dominant heart involvement, AL_HK = AL amyloidosis patients with dominant heart and kidney involvement, AL_HTX = AL amyloidosis patients who received a heart transplant, HC = detectable clonal heavy chain in the immunofixation electrophoresis in serum, AGG = β-sheet aggregation tendency, GRAVY = grand average of hydropathicity, pI = isoelectric point, ΔpI = difference between the light chain sequence and the respective reference sequence, IGLVJC = trimmed full-length light chain, IGLVJ = connected IGLV and IGLJ segments, Mw = average molecular weight.

	median AGG	median GRAVY	pI IGLVJC	ΔpI IGLVJC	pI IGLVJ	ΔpI IGLVJ	Mw (average)
AL – IGLV6-57	664.6	-0.434	4.78	-0.14	4.47	0.05	17648
AL_H – IGLV6-57	813.9	-0.445	4.73	-0.14	4.37	0.02	17668
AL_HTX – IGLV6-57	866.6	-0.392	4.72	-0.21	4.36	-0.05	17668
AL_HK – IGLV6-57	655.4	-0.426	4.99	0.02	4.68	0.20	17593
MM – IGLV2-23	870.3	-0.223	7.78	-0.55	6.53	-1.36	19176

	median AGG	median GRAVY	pI IGLVJC	ΔpI IGLVJC	pI IGLVJ	ΔpI IGLVJ	Mw (average)
AL_H HC – IGLV6-57	852.4	-0.423	4.76	-0.12	4.43	0.05	17689
AL_H no HC – IGLV6-57	762.3	-0.487	4.71	-0.16	4.33	0	17668
AL > dFLC – IGLV6-57	761.3	-0.392	4.72	-0.15	4.36	0.01	17562
AL < dFLC – IGLV6-57	719.7	-0.445	4.91	0	4.60	0.19	17712

3.7.6 Summary IGLV6-57 and IGLV2-23

In the analysis regarding the IGLV subfamily usage between AL and MM, the most prominent differences were found concerning IGLV6-57 and IGLV2-23. IGLV6-57 was detected almost exclusively in AL and was even the most commonly used IGLV subfamily (21 % vs. 8 %; p = 0.001). In contrast, IGLV2-23 was detected exclusively in the MM cohort and as one of the most used IGLV subfamilies (0 % vs. 15 %; p = <0.001). Both IGLV subfamilies share a preferred linkage with IGLJ2/IGLC2. However, while the IGLV2-23 MM sequences were associated with a detectable clonal HC in serum, the AL IGLV6-57 sequences seem to relate to the no HC, < dFLC (AL_H p = 0.038), and AL_HK subgroups.

In a more detailed analysis, the IGLV2-23 MM (n = 8) sequences presented a significantly higher median mutation count in the IGLV segment than the AL IGLV6-57 sequences (n = 18) (trimmed = 11.5 vs. 6.0; p = 1.60e-06) – including the CDR1 (3.5 vs. 1.0; p = 0.003), CDR2 (88 % vs. 53 %; p = 0.027), and FR3 region (100 % vs. 65 %; p = 0.023). This finding was further reflected in more mutation hotspots (8 vs. 1), especially in the CDR regions of MM sequences. Despite three frequently mutated positions in both cohorts, charged amino acids differed between the IGLV subfamilies. The AL IGLV6-57 sequences showed an additional charge in the CDR1 (10/18), CDR2 (8/18), and CDR3 (9/18) region, whereas, in the MM IGLV2-23 sequences, only the CDR3 region presented an additional charge. Moreover, the IGLV6-57 reference sequence generally contains a higher proportion of the negatively charged AA aspartic acid (6.1 % vs. 3.1 %) and the positively charged AA arginine (4.0 % vs. 2.1 %). However, in the overall amino acid composition of the LCs, both cohorts share several changes compared to the respective reference sequence. Taken IGLV6-57 separately, an increase in alanine and arginine as well as a decrease in asparagine was noted compared to the reference. In the IGLV2-23 sequences, the only exclusive difference was detected for valine, with a decrease compared to the reference (6.9 % vs. 6.3 %). Differences between the reference sequences were not only detected in the overall AA composition but also in the biophysical parameters. Considering only the IGLV segment, the IGLV2-23 reference shows a higher beta

sheet aggregation tendency as well as a higher pI than the IGLV6-57 reference. In contrast, a higher Mw was detected in the IGLV6-57 reference. In the analysis of the patient-derived LCs regarding the ΔpI value, it is noticeable that while the MM IGLV2-23 sequences displayed almost no change compared to the reference sequence, a decrease was found in the IGLV6-57 AL LCs. This was not only observed in the complete LC (IGLV6-57 AL = -1.36, IGLV2-23 MM = -0.14) but also for the IGLV-IGLJ segments (IGLV6-57 AL = -0.55, IGLV2-23 MM = 0.05). In addition, a higher GRAVY score was detected for the IGLV2-23 MM sequences than for the IGLV6-57 sequences.

Not only is the strong association of AL and IGLV6-57 prominent but a specific organ tropism was further detected. The AL_HK sequences (n = 8) were assigned to IGLV6-57 twice as often as the AL_H sequences (n = 10) (38 % vs. 15 %; p = 0.032). However, this was reversed in the small AL_HTX group, which showed an IGLV6-57 sequence in half of the cases (4/8). Interestingly, AL_H sequences and especially the AL_HTX sequences (IGLV = 7.5, IGLJ = 75 %, IGLC = 25 %) presented a higher mutational load or frequency than the AL_HK sequences in all segments (IGLV = 5.0, IGLJ = 25 %, IGLC = 0 %). Besides the fact that several discriminating hotspots were identified, the insertion of charge can be defined as the biggest difference between these subgroups. The CDR1 of AL_H sequences frequently presented the insertion of negatively charged AA, while in AL_HK sequences additional positive charges were noted (p = 0.036). In contrast, only the CDR3 of AL_HK sequences presented an additional charge in most of the cases (6/8 vs. 3/9; p = 0.02). These sequence findings also reflected in differences regarding the overall percentage of charged AAs. The AL_H sequences presented a higher percentage of arginine and aspartic acid as well as a lower amount of the two positively charged AAs histidine and lysine (4.2 % vs. 4.8 %, p = 0.031). These differences naturally affected the calculated biophysical parameters, especially the pI and ΔpI. While the AL_HK full-length sequences showed almost no changes in pI and ΔpI, the AL_H, especially AL_HTX sequences, showed a decrease. This led to a significantly lower pI for the full-length LCs (pI IGLVJC 4.99 vs. 4.72, p = 0.043) and IGLVJ segments (pI IGLVJ 4.36 vs. 4.68, p = 0.043) of AL_HTX sequences compared to AL_HK sequences. In addition, the AL_H (814) and especially the AL_HTX sequences (867) presented a higher beta sheet aggregation tendency compared to AL_HK sequences (655).

Only the AL_H sequences were analyzed concerning a detectable clonal HC since an AL_HK HC association was noted. The AL_H no HC (n = 4) sequences showed an especially high mutation frequency in the IGLJ (100 % vs. 50 %) and IGLC segment (50 % vs. 0 %) as well as a higher mutation count in the IGLV segment (7.5 vs. 6.5). In a more detailed sequence

analysis, the AL_H HC (n = 4) sequences presented a Y50F mutation in three out of four cases. Furthermore, two AL_HTX HC cases showed a Y50F mutation and a similar exchange at position 37 from tyrosine towards phenylalanine. Overall, the AL_H HC sequences showed a higher aggregation tendency as well as a lower percentage of aspartic acid and serine and a higher percentage of valine than AL_H no HC sequences.

Comparing the AL dFLC subgroups, the > dFLC subgroup (n = 8) showed a higher median mutation count in the IGLJ-segment (2.0 vs. 1.0) and a lower median mutation count in the FR3 (1.0 vs. 2.5). Interestingly, the Y50F mutation was associated with AL_H HC and AL_HTX HC and further detected exclusively in the < dFLC subgroup (n = 9). In general, for the AL > dFLC subgroup, a higher percentage of serine (16.7 % vs. 15.8 %; p = 0.012) and alanine (8.5 % vs. 7.9 %; p = 0.015) was noted. Concerning the biophysical parameters, only a difference in molecular weight was detected – which was higher in the > dFLC subgroup.

3.8 Summary of Common IGLV Subfamilies

Each frequently detected IGLV subfamily was associated with either a specific subgroup like a present HC, a dFLC >180 mg/L, or a distinct organ tropism in AL. However, no common pattern or discrimination between AL and MM was found (**Table 32**). In the following summary, IGLV3-21 will be neglected, since the majority of sequence features could not be evaluated.

To examine whether the difference in the LC behavior is based on a difference in the general mutation distribution and frequency, these aspects were analyzed within the IGL segments. Regarding the IGLV mutation count, it is noticeable that the MM sequences presented a median mutation count of 11.5-12.0 in all three analyzed IGLV subfamilies. In contrast, the AL IGLV6-57 subfamily displayed the lowest value with 6.0, followed by AL IGLV3-1 sequences with 7.0, and AL IGLV2-14 sequences with 8.5. If evaluable, the FR1 presented as the least frequently mutated region (IGLV3-1, IGLV2-14, IGLV2-23).

Analyzing the IGLJ segment mutation frequency between AL and MM, no difference >30 % was detected. Overall, the AL IGLV6-57 sequences presented the lowest IGLJ mutation frequency compared to the other AL subgroups (53 % vs. 71 – 79 %). Notably, the AL_H or AL_HTX sequences showed an increased IGLJ mutation count/frequency compared to the AL_HK sequences in all IGLV subfamilies.

Regarding the IGLC segment, several common observations were made. The MM IGLV2-14 and IGLV3-1 sequences presented a higher IGLC mutation frequency than the respective AL

subgroup. In addition, the AL HC sequences presented an increased mutation frequency in both IGLV subfamilies. In contrast, no difference between AL IGLV6-57 and MM IGLV2-23 sequences was noted regarding mutation frequency and IGLV6-57 AL no HC sequences showed an increased mutation frequency compared to AL HC sequences.

Table 32. Comparison of analyzed AL amyloidosis and multiple myeloma IGLV subfamilies in the context of common associations and mutation frequency and count in the IGL segments. AL = AL amyloidosis patients, MM = multiple myeloma patients, AL_H = AL amyloidosis patients with dominant heart involvement, AL_HK = AL amyloidosis patients with dominant heart and kidney involvement, AL_HTX = AL amyloidosis patients who received a heart transplant, HC = detectable clonal heavy chain in the immunofixation electrophoresis in serum, > dFLC = dFLC >180 mg/L, < dFLC = dFLC <180 mg/L, AA = amino acid, FR = framework region, CDR = complementary determining region, ↑ = increase, ↓ = decrease, * = p ≤0.05

IGLV subfamily	IGLV2-14		IGLV3-1		IGLV6-57	IGLV2-23
	AL n = 16	MM n = 8	AL n = 14	MM n = 3	AL n = 17	MM n = 8
Associations	AL_H HC*; IGLJ1/ IGLC1	> dFLC	AL_H, > dFLC*,	no HC, < dFLC, IGLJ1/ IGLC1	AL* AL_HK*, AL_HTX*, AL_HK HC, < dFLC (AL_H*)	only MM*, MM HC, MM > dFLC
Median mutation count/ frequency IGLV	8.5* ↑AL_HK ↑HC (97 AA)	11.5* (97 AA)	7.0* FR1 only in AL_H (83 AA)	12.0* ↑CDR1* (83 AA)	6.0* ↑AL_HTX ↓AL_HK (75 AA)	11.5* ↑CDR1* ↑CDR2* ↑FR3* (75 AA)
Median mutation count/ frequency IGLJ (12 AA)	1.0 – 71 % ↑AL_H	1.0 – 75 %	1.0 – 79 % ↑AL_H	1.0 – 100 %	1.0 – 53 % ↑AL_HTX ↓AL_HK ↑no HC	1.0 – 75 %
Median mutation count/frequency IGLC	1.0 – 14 % ↑HC ↑> dFLC (77 AA)	1.0 – 25 % (77 AA)	1.0 – 7 % AL_H only; ↑AL_HC (77 AA)	1.0 – 33 % no HC (77 AA)	1.0 – 12 % ↑AL_HTX ↓AL_HK ↑no HC (78 AA)	1.0 – 13 % (74 AA)

The increased IGLV mutation count in MM sequences was also reflected in more mutation hotspots in almost all analyzed IGLV subfamilies (**Table 33**). For example, while the IGLV2-14 AL sequences displayed only two hotspots, the MM sequences displayed nine – which also comprised the AL hotspots. In a comparison between AL IGLV6-57 and MM IGLV2-23 sequences, this was even more pronounced with one AL hotspot and eight MM hotspots.

In a more detailed sequence analysis, it was not possible to find one discriminating hotspot that was present in all or most of the AL sequences and none or only a few of the respective MM sequences. It should be highlighted that the third position of the CDR2 and the last two AAs of

the IGLV segment appear to be the most frequently mutated in all analyzed IGLV subfamilies in both diseases.

Interestingly, in the IGLV2-14 as well as IGLV3-1 sequences, it was possible to detect differences regarding the insertion or replacement of charged AAs as well as small AAs between the two diseases. The IGLV2-14 AL sequences presented additional charges in the CDR1 region and enrichment of small AAs in the CDR3 – which reflected in an overall higher percentage of glycine than in the MM sequences. In contrast, the IGLV2-14 MM sequences presented additional charges in the CDR3 and overall a higher percentage of the negatively charged AA aspartic acid. A similar phenomenon was observed in the IGLV3-1 subfamily: The AL sequences showed a loss of charge in the CDR1, the MM sequences an additional one in the CDR2 as well as a higher overall percentage of the positively charged AA arginine. For both of the diseases, a mutation towards alanine or glycine was detected at the first AA of the IGLJ segment. Interestingly, the IGLV3-1 AL HC sequences showed more often an additional alanine, and the no HC sequences a glycine. Remarkably, the IGLV2-23 MM sequences, like the IGLV2-14 MM sequences, showed additional charges in the CDR3. As a side note, it was also possible to detect an additional charge in the CDR3 of IGLV3-21 MM sequences more frequently than in the respective AL sequences (5/11 vs. 1/10).

In the analysis of IGLV6-57 AL sequences, not only the insertion of charged AAs in all CDR regions but also a specific organ tropism was noted. The IGLV6-57 AL_H sequences presented the insertion of negatively charged AA and the AL_HK positive ones in the CDR1. Moreover, the CDR3 region also displayed an enrichment of charged AAs which was especially prominent in AL_HK sequences. Additionally, the AL IGLV6-57 sequences presented an increased percentage of glycine compared to the respective reference.

Table 33. Comparison of analyzed AL amyloidosis and multiple myeloma IGLV subfamilies in the context of mutation hotspots, sequence characteristics, and the overall amino acid composition. AL = AL amyloidosis patients, MM = multiple myeloma patients, AL_H = AL amyloidosis patients with dominant heart involvement, AL_HK = AL amyloidosis patients with dominant heart and kidney involvement, HC = detectable clonal heavy chain in the immunofixation electrophoresis in serum, > dFLC = dFLC >180 mg/L, < dFLC = dFLC <180 mg/L, AA = amino acid, AL_HTX = AL amyloidosis patients who received a heart transplant, FR = framework region, CDR = complementary determining region, ↑ = increase, ↓ = decrease, * = p ≤0.05.

IGLV subfamily	IGLV2-14		IGLV3-1		IGLV6-57	IGLV2-23
	AL n = 16	MM n = 8	AL n = 14	MM n = 3	AL n = 17	MM n = 8
Mutation hotspots [n]	2 not exclusive; N62D, first AA IGLJ →L	9 N62	3 HC: 32A, C33S;	3	1 3 AL_HK hotspots, 53N 2 AL_H hotspots, S69R, K82Q; HC: Y50F and Y37F > dFLC: Y50F, < dFLC: additional charge CDR2	8 CDR regions, V29I, G53V
Sequence features	additional charge CDR1; small AA CDR3;	additional charge CDR3*; especially in 49M and 54S mutated sequences	HC CDR3: ↑A no HC CDR3: ↑G; loss of charge CDR1	additional charge CDR2	additional charge CDR1*, AL_H: - AL_HK: +; additional charge CDR2; additional charge CDR3, AL_HK*	additional charge CDR3
AA composition	↑G and M ↑S and T AL_H ↑N AL_HK ↑S and N HC ↑L* > dFLC ↑T < dFLC	↑ D	↑ S and V AL_HK ↑ A and N HC ↑S* no HC	↑R	↑D and T ↓S and Y ↑I* AL_H ↑L* AL_HK ↓D and S HC ↑D and S* > dFLC ↑A* > dFLC	↑D and T ↓S and Y

Several parameters were calculated to evaluate the effect of the mutations and amino acid composition on the biophysical properties of the LCs (**Table 34**). In the context of the beta-sheet aggregation tendency, the IGLV2-14 and IGLV3-1 AL HC sequences presented a higher aggregation tendency than the AL no HC sequences.

The GRAVY value is defined by the sum of hydropathy values of all AA divided by the protein length and can therefore be compared between all subgroups. The AL IGLV6-57 sequences presented the lowest value (-0.434), which indicates the most hydrophilic nature and a possible interaction of the linear protein with water molecules. In contrast, the IGLV2-23 MM sequences presented an almost twice as high value (-0.223) and therefore the highest of all analyzed subgroups. Comparable values were observed for the AL (-0.284) and MM (-0.264) IGLV2-14 subgroups and lower values for the IGLV3-1 AL sequences (AL = -0.337, MM = -0.397).

Regarding the pI and ΔpI, the IGLV2-14 full-length LCs sequences presented higher values than the AL subgroups. The IGLV2-14 and IGLV3-1 AL sequences showed a comparable ΔpI value of -0.5. Overall, the IGLV2-23 MM sequences presented the lowest ΔpI value of all analyzed IGLV subgroups, and the IGLV6-57 AL sequences the highest (-0.55 vs. -0.14).

In addition, the Mw of the LCs was calculated. Here, the MM IGLV2-14 and IGLV3-1 sequences presented a higher value compared to the respective AL sequences. Remarkably, a higher value was further observed for the > dFLC subgroup in the AL IGLV2-14 and IGLV6-57 sequences.

Table 34. Comparison of analyzed AL amyloidosis and multiple myeloma IGLV subfamilies in the context of biophysical parameters. AL = AL amyloidosis patients, MM = multiple myeloma patients, AL_H = AL amyloidosis patients with dominant heart involvement, AL_HK = AL amyloidosis patients with dominant heart and kidney involvement, HC = detectable clonal heavy chain in the immunofixation electrophoresis in serum, > dFLC = dFLC >180 mg/L, < dFLC = dFLC <180 mg/L, AA = amino acid, AL_HTX = AL amyloidosis patient who received a heart transplant, FR = framework region, CDR = complementary determining region, ↑ = increase, ↓ = decrease, * = p ≤0.05.

IGLV subfamily	IGLV2-14		IGLV3-1		IGLV6-57	IGLV2-23
	AL n = 16	MM n = 8	AL n = 14	MM n = 3	AL n = 17	MM n = 8
Median AGG	803.8 ↑HC* ↑> dFLC	1012.4	1146.7 ↑AL_HK ↑HC	1004.0	664.6 ↓AL_HK ↑AL_HTX	870.3
Median GRAVY	-0.284	-0.264	-0.337 ↓AL_H	-0.397	-0.434	-0.223
Median pI IGLVJC	6.59 ↓AL_HK ↑HC	6.77	5.04	5.32	4.78 ↑AL_HK	7.78
Median ΔpI IGLVJC	-0.50 ↓AL_HK ↓HC	-0.32	-0.51 ↓ AL_H	-0.46	-0.14 ↑AL_HK	-0.55
Median Mw	19580 ↑AL_HK ↑> dFLC	19655	18193 ↓AL_HK ↑HC	18446	17648 ↑> dFLC	19176

3.9 Rare IGLV Subfamilies

3.9.1 IGLV1-40

It was possible to assign one AL_H (AL = 1 %, AL_H = 2 % FOR138) and two MM sequences (4 %; MM104, MM140) to the IGLV1-40 subfamily (**Supplementary Information Figure 33**, **Supplementary Information Figure 34**). In the bioinformatic analysis of the MM sequence MM104, it was not possible to generate a continuous LC sequence – two sequence segments were indicated. However, this does not affect the region covered by the VBase2 reference, which was verified by Sanger sequencing (**Supplementary Information Figure 35**). The AL_H sequence showed a total of nine mutations in the trimmed IGLV segment and one each in the IGLJ and IGLC segments. In the MM sequences, a lower number of five (MM104) and eight (MM140) mutations in the IGLV segment were detected. In addition, both MM sequences displayed a mutation from valine towards isoleucine at the second position of the IGLJ segment. A second, disease-spanning hotspot was defined at position 54S (3rd position CDR2, AL = N, MM = T).

3.9.2 IGLV1-44

Despite not reaching the cut-off, IGLV1-44 was one of the frequently detected IGLV subfamilies. Five AL_H sequences (AL = 6 %, AL_H = 8 %) and nine MM sequences (17 %) were assigned to this subfamily.

As discussed in section 4.2, all associated MM samples showed an additional sequence_B in the bioinformatics analysis. In sequence_A, an additional sequence segment – leading to sequence gaps – was identified in the bioinformatic analysis in seven cases. Also, in sequence_B several sequence sections were detected but in eight out of nine cases, only the leader region was affected. For the two sequences without an additional sequence segment in sequence_A, no sequence that covers the complete IGLV segment N-terminal was generated (**Supplementary Information Figure 36**). Four MM samples were additionally Sanger sequenced (MM116 (**Supplementary Information Figure 37**), MM119 (**Supplementary Information Figure 38**), MM137 (**Supplementary Information Figure 39**), MM141 (**Supplementary Information Figure 40**)) and in three cases a sequence coverage with the respective sequence_B was detected. The Sanger sequencing of MM116 led to a sequence that showed several overlapping nucleotide signals but matched the respective sequence_B – in the majority of nucleotides. Based on these results, the corresponding MM sequence_B was used for the analysis.

The AL and MM sequences presented a comparable mutation count and frequency of the IGLJ (AL = 4/5, range 0 – 2; MM = 6/8, range 0 – 2) and IGLC segment (AL = 2/5, range 0 – 1;

MM = 3/8, range 0 – 1). However, a large range in the IGLV mutation count was observed in the MM compared to AL sequences (5 – 15, AL = 7 – 11) (**Supplementary Information Table 8**).

In addition, the MM sequences presented four mutation hotspots 33T (7/9, CDR1), 39Q (5/9, FR2), 51S (6/9, CDR2), 54Q (5/9, CDR2), the AL sequences only one 35N (4/5, CDR1) (Figure 32) – which was not shared with the MM cohort. For the AL hotspot 35N (4/5, 1xX), a mutation towards serine or threonine was detected twice each. In contrast, the MM sequences displayed a mutation at this position in only four out of nine cases (2x threonine, 1x serine). Position 33T was defined as the most frequently mutated position in the MM sequences (7/9) and a mutation towards proline was detected in a majority of the cases (4/7). Only one AL sequence displayed a mutation at this position (serine). At the 39Q MM hotspot, a mutation towards histidine was detected in four out of nine MM and two out of five AL cases. At position 51S, no directed exchange pattern was detected but at position 54Q a mutation towards a positively charged AA was detected in four out of nine MM cases (2x histidine, 2x arginine, AL n = 0).

Besides these IGLV mutation hotspots, the first AA of the IGLJ segment was also defined as a mutation hotspot for both diseases, and a preferred exchange towards proline was noted (AL = 4/5, MM = 5/9).

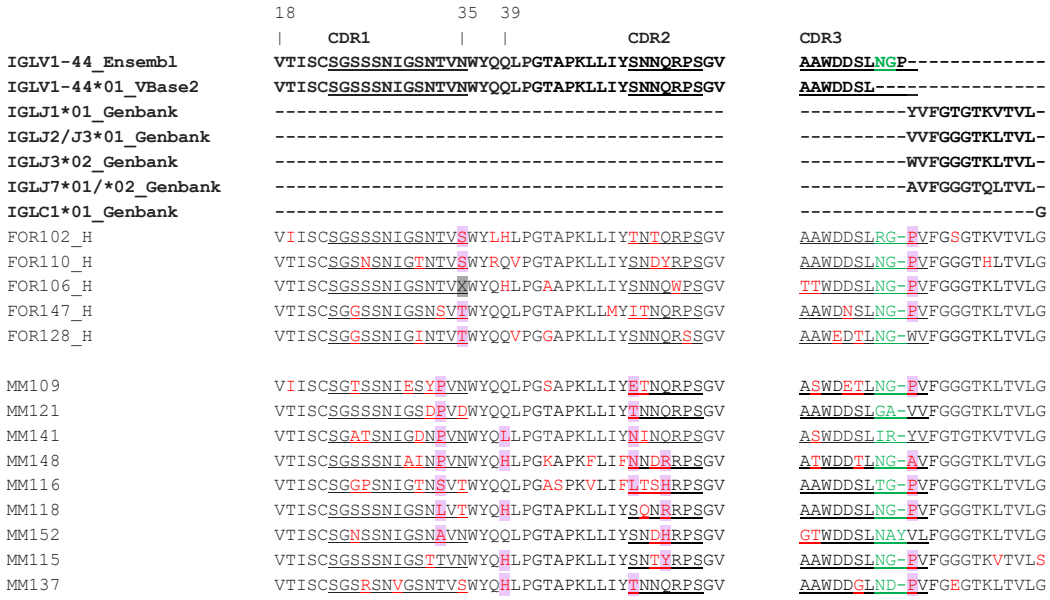


Figure 32. Sequence sections of IGLV1-44 assigned AL amyloidosis and multiple myeloma light chain sequences. Bold = reference sequences, underlined = CDR regions, red letter = mutation, purple highlight = mutation hotspot, X and grey highlight = not unambiguously determined amino acid, green letter = linker region, MM = multiple myeloma patient, _H = AL amyloidosis patient with dominant heart involvement. Only one IGLC reference is shown because the first amino acid can in all cases be defined as glycine. Amino acids were numbered according to the VBase2 reference. The complete amino acid sequence alignment is shown in Supplementary Information Figure 41. The complete cDNA sequence alignment is shown in Supplementary Information Figure 42.

3.9.3 IGLV1-47

One AL_HK (AL = 1 %, AL_HK = 5 %, FOR168) and one MM sequence (2 %, MM151) were assigned to the IGLV1-47 subfamily (**Supplementary Information Figure 43**, **Supplementary Information Figure 44**). In the bioinformatic analysis of the MM sequence, it was not possible to generate a continuous LC sequence and two sequence segments were indicated. However, this does not affect the region covered by the VBase2 reference. The AL_HK sequence displayed three mutations in the IGLV segment, two in the IGLJ and one in the IGLC segment. In contrast, the MM sequence presented six mutations in the IGLV segment (including 3rd position CDR2). No position showed a mutation in both sequences.

3.9.4 IGLV1-51

Two AL (AL = 2 %, AL_H 2 % FOR193, AL_HK 5 % FOR149) and two MM sequences (4 %, MM110, MM131) of the IGLV1-51 subfamily were identified (**Supplementary Information Figure 45**, **Supplementary Information Figure 46**). The AL sequences displayed eleven (FOR139) and six (FOR149) mutations in the IGLV segment and one in the IGLJ segment. In MM110 six mutations and in MM131 five mutations were detected in the IGLV segment and two (MM110) or four mutations (MM131) in the IGLJ segment. A mutation in the IGLC segment was not detected in all cases. In a detailed sequence analysis, two positions are noticeable: the exchange N53D (3rd position CDR2, AL = 2/2, MM = 1/2) and an undirected exchange at the first position of the IGLJ segment.

3.9.5 IGLV2-8

One AL_H (2 %, FOR182) and one AL_HK (4 %, FOR164) sequence were assigned to IGLV2-8 (AL = 2 %) and no MM sequence (**Supplementary Information Figure 47**, **Supplementary Information Figure 48**). The AL_H sequence displayed ten mutations in the IGLV segment and one in the IGLJ segment. In contrast, the AL_HK sequence contained only six mutations in the IGLV segment. The sequences shared three mutated positions (27S, 49M, and 54S (3rd position CDR2)). Interestingly, both sequences displayed an exchange towards isoleucine at position 49M.

3.9.6 IGLV2-11

Only one MM sequence (2 %, MM127) was assigned to the IGLV2-11 subfamily (**Supplementary Information Figure 49**, **Supplementary Information Figure 50**). In general, four mutations in the IGLV segment, including the third position in the CDR2 (S54T) as well as one mutation in the IGLJ segment, were detected. Additional Sanger sequencing

verified most of the cDNA positions but some nucleotide signal overlaps were noted (**Supplementary Information Figure 51**).

3.9.7 IGLV3-19

Three AL_H (AL = 4 %, AL_H = 5 %; FOR103, FOR148, FOR216) and two MM sequences (4 %; MM128, MM145) were assigned to the IGLV3-19 subfamily (**Supplementary Information Figure 52**, **Supplementary Information Figure 53**). The MM128 sequence was verified through additional Sanger sequencing (**Supplementary Information Figure 54**). The AL sequences were missing seven (2/3) or eight (1/3) AAs at the N-terminus of the IGLV segment due to the oligonucleotide binding site. FOR103 displayed four mutations and FOR216 eight mutations in the IGLV segment as well as one mutation in the IGLJ segment. In FOR148 no mutation was detected at all, however, several AAs were not determined unambiguously. Both MM sequences showed seven mutations in the IGLV segment and shared mutations at positions 29S, 65S, and 92S. Interestingly, the MM sequences displayed a S65T mutation, while two out of three AL sequences displayed a S65N mutation. In addition, MM145 displayed three mutations in the IGLJ segment and one in the IGLC segment.

3.9.8 IGLV3-25

One AL_H (AL = 1 %, AL_H = 2 %, FOR111) and one MM (2 %, MM105) sequence were assigned to the IGLV3-25 subfamily (**Supplementary Information Figure 55**, **Supplementary Information Figure 56**). However, it was not possible to generate a N-terminal elongated AL sequence and therefore no detailed sequence analysis was performed. The MM sequence displayed eleven mutations in the IGLV segment (including the 3rd CDR2 position: S51H) and none in the IGLJ and IGLC segment. It should be noted that through additional Sanger sequencing, it was not possible to generate an evaluable MM sequence.

3.9.9 IGLV7-46

One MM (2 %, MM113) sequence was assigned to the IGLV7-46 subfamily (**Supplementary Information Figure 57**, **Supplementary Information Figure 58**). In the bioinformatic analysis, it was not possible to generate a continuous LC sequence and two sequence segments were indicated. However, only partial sequence matches were detected. It was possible to detect 15 mutations in the IGLV segment (including the 3rd CDR2 position 54S).

3.9.10 IGLV8-61

Only one AL_HK sequence (FOR229) was assigned to the IGLV8-61 subfamily (**Supplementary Information Figure 59**, **Supplementary Information Figure 60**). Compared to the VBase2 reference, four AAs were missing at the N-terminus due to the

oligonucleotide binding site. The sequence displayed 14 mutations in the IGLV segment as well as one in the IGLJ segment. Moreover, the sequence corresponded to an AL patient who was also diagnosed with MM.

3.10 Summary of Rare IGLV Subfamilies

In the analysis of the most common AL-associated IGLV subfamilies, several differences, as well as commonalities, between AL and MM were noted.

Regarding the IGLV segment, the MM sequences presented a significantly higher mutation count than the respective AL subgroups. Interestingly, all analyzed IGLV subfamilies showed a comparable median mutation count of 11.5-12.0 in the MM sequences, which did not reflect in the other, rarer, IGLV subfamilies (**Table 35**). Furthermore, the higher mutation frequency of MM sequences in the IGLC segment, could not be unambiguously noted for the remaining IGLV subfamilies. In contrast, AL and MM sequences of less common IGLV subfamilies also frequently presented the mutation hotspot at the third position of the CDR2 region.

AL and MM sequences frequently showed differences in charge and small AAs which was also noted for the IGLV1-44 subfamily. Despite not reaching the cut-off, IGLV1-44 was one of the frequently detected IGLV subfamilies, comprising five AL and nine MM sequences. While four MM cases presented a mutation towards a positively charged AA at position 54Q, this was not observed in the AL cohort. Interestingly, besides the shared mutation hotspot at the first AA of the IGLJ segment and a prominent exchange towards proline, the MM sequences further presented a T33P mutation in more than half of the cases.

Table 35. Comparison of relative rare AL amyloidosis and multiple myeloma associated IGLV subfamilies in context of mutations, and sequence features. In the case of IGLV1-44, a mutation range was specified. AL = AL amyloidosis patients, MM = multiple myeloma patients, + = positively charged amino acid, , NA= not available, CDR = complementary determining region, → = mutation towards.

	n		Mutation count [n]						Mutation 3rd CDR2 position		private sequence features	
			IGLV		IGLJ		IGLC					
	AL	MM	AL	MM	AL	MM	AL	MM	AL	MM	AL	MM
IGLV1-40	1	2	9	5, 8	1	1, 1	1	0, 0	S54N	S54T		IGLJ 3rd →I/V
IGLV1-44	5	9	7 – 11	5 – 15	0 – 2	0 – 2	0 – 1	0 – 1	no	no	35N	T33P, Q54+,
IGLV1-47	1	1	3	6	2	0	1	0	no	N53D		
IGLV1-51	2	2	6, 11	5, 6	1	2, 4	0	0	N53D	N53D (1/2)	no	no

	n		Mutation count [n]						Mutation 3rd CDR2 position		private sequence features	
			IGLV		IGLJ		IGLC					
	AL	MM	AL	MM	AL	MM	AL	MM	AL	MM	AL	MM
IGLV2-8	2	0	6, 10		0, 1		0, 0		54S		M49I	
IGLV2-11	0	1		4		1		0		S54T		
IGLV3-19	3	2	0, 8	7, 7	0, 1	0, 3	0	0, 1	no	no	S65N	S65T
IGLV3-25	1	1	NA	11	NA	0	NA	0	NA	S51H	NA	
IGLV7-46	0	1		15		0		0		S54D		
IGLV8-61	1	0	14		1		0		no			

Based on the analysis of the relatively rare AL as well as MM-associated IGLV subfamilies, no new insights were gained beyond those already obtained above. However, the results concerning IGLV1-44 supported the findings regarding a difference in charge and small AAs between AL and MM.

4 Discussion

This study aimed to investigate AL-specific lambda LC sequence features to I) draw conclusions on the pathogenic mechanism and II) establish a risk stratification for disease progression from MGUS/ SMM/MM towards AL. For these analyses, MM sequences were used as reference set since AL and MM patients usually display an increased level of free LCs but those of MM patients do not deposit as fibrils.

Summary Key-Note Results

Cohorts:
- 52 MM patients and 82 AL patients, comprising 61 AL_H and 21 AL_HK patients
- stratification for a present clonal HC and dFLC > or < 180 mg/L
- more detailed analyses: IGLV2-14, IGLV3-1, IGLV3-21, IGLV6-57 and IGLV2-23

IGLV subfamily distribution:
- IGLV6-57 more frequent in: AL than MM*, AL_HK than AL_H*, and AL_H < dFLC than AL_H > dFLC*
- IGLV2-23 more frequent in: MM than AL*
- IGLV2-14 more frequent in: AL_H HC than AL_H no HC*
- IGLV3-1 more frequent in: AL > dFLC than AL < dFLC*

IGLV-IGLJ-IGLC composition:
- IGLJ2/IGLC2 most frequent linkage for AL and MM
- IGLV3-21 AL: IGLJ2/IGLC2 and IGLJ3/IGLC3 equally frequent

IGLV3-21 subfamily (AL):
- large number of ambiguously identified AA and LCs N-terminal only amplified starting from FR2 → additional NGS experiments
- $MP_{\_NGS\_PCR}$: verification of the IGLV3-21 assignment and sequences itself
- $N_{\_NGS\_PCR}$: no additional insights could be gained

Mutation count and frequency:
- IGLV: higher in MM than AL LCs*
- IGLJ: no major differences between AL and MM LCs; higher in AL_H than AL_HK LCs
- IGLC: IGLV2-14: higher in MM than AL LCs;
  IGLV2-14 and IGLV3-1: higher in AL HC than AL no HC LCs

Mutation hotspots and AA composition:
- more mutation hotpots in MM than AL LCs
- higher percentage of charged AA in IGLV2-14 (D) and IGLV3-1 (R) MM than AL LCs
- differences concerning charged and small AAs between AL and MM LCs: IGLV2-14*, IGLV3-1, IGLV3-21, and IGLV1-44
- differences concerning charged AAs between IGLV6-57 AL_H and AL_HK*

Biophysical parameters:
- AGG: higher for IGLV2-14 and IGLV3-1 AL HC than AL no HC LCs
- GRAVY score: no significant differences between AL and MM LCs;
  IGLV6-57 LCs the lowest value and IGLV2-23 LCs the highest value
- pI and ΔpI: IGLV2-14 MM higher pI than AL full-length LCs;
  IGLV2-23 MM LCs the lowest ΔpI value and IGLV6-57 AL LCs the highest
- Mw: IGLV2-14 and IGLV3-1 MM higher than AL LCs;
  higher for IGLV2-14 and IGLV6-57 AL > dFLC than AL < dFLC LCs

4.1 How to be Sure to Sequence the Amyloidogenic Clone?

4.1.1 Verification Steps

Before evaluating and discussing more detailed sequence characteristics, a more fundamental question shall be considered first: how to be sure to sequence the amyloidogenic clone? In this context, a profound advantage of this study was sequencing cDNA from $CD138^+$ sorted plasma cells. Using this starting material, it is assumed that the analyzed samples contain a high proportion of clonal cells and can therefore in general reflect the disease-associated clone (Chilosi et al. 1999; Kriegsmann et al. 2018; Ridley et al. 1993; Wijdenes et al. 1996)).

Additionally, within the framework of this study, four different approaches were used to verify the sequencing results: (A) The initial development of a full-length lambda LC sequencing protocol included the first verification step. Here, four different oligonucleotide combinations were used to sequence FOR101 and FOR102 and it was possible to detect the same underlying LC sequence in all cases (chapter 2.6.4). (B) An additional verification step was necessary since the MM sequences were obtained through a different sequencing approach. Sanger sequencing of several MM LCs verified not only the MM sequences but further the AL method and oligonucleotides. (C) IGLV3-21 proved to be the most difficult AL lambda IGLV subfamily to sequence, and additional NGS experiments were performed. Here, the $MP_{\_NGS\_PCR}$ approach verified both the family assignment and the sequences. (D) It was possible to demonstrate that the generated LC sequences match the amyloidogenic protein in the patients, which will be explained in the following paragraph in more detail.

This fourth verification step was feasible due to a collaboration with a research group at Ulm University headed by Prof. Dr. rer. nat. Markus Fändrich and Dr. rer. nat. Christian Haupt. Determining the amyloid load in abdominal fat aspirates is a standard tool for supporting an AL diagnosis and further evaluating the prognosis (van Gameren et al. 2010). These fat samples were obtained from all patients treated at the amyloidosis center in Heidelberg and analyzed for their amyloid load. If a high amyloid load in the fat aspirates and an analyzable LC sequence were present, these fat samples were sent to Ulm. In the next step, Julian Baur (Ulm University) extracted the underlying amyloidogenic proteins from these fat samples and additionally from heart tissues. These heart tissues were obtained from heart transplants donated for research. In total, 54 lambda sequences and corresponding samples were sent to Ulm. This sample set comprised 34 AL_H samples, eight AL_HK samples, nine samples from AL patients with dominant kidney involvement (AL_K), and three samples from AL patients with no dominant heart, kidney, or heart and kidney manifestation (AL_D). Based on this dataset, ten sequences and corresponding proteins were analyzed in more detail (Baur et al. 2022). This concerned

eight sequences that were generated within the framework of this work (6x AL_H, 2x AL_K) and in collaboration with Sarah Schreiner (University Hospital Heidelberg, AL_K and AL_D sequences). The two remaining sequences and samples were already analyzed through cryo-EM and corresponded to AL_HTX patients (FOR005 (Radamaker et al. 2021a), FOR006 (Radamaker et al. 2019)).

These eight extracted proteins were analyzed by mass spectrometry and the provided sequences corresponded well to the actual amyloidogenic protein. Further, the proteins comprised the IGLV segment, IGLJ segment, and up to 14 AAs of the IGLC segment (Baur et al. 2022). It is worth mentioning that it was possible to verify sequences from three different IGLV families and four IGLV subfamilies: IGLV1-40, IGLV1-44, IGLV2-14, and IGLV3-19. As a side note, the AA sequence of three cases (FOR101, FOR142, FOR159) contained ambiguous AA positions. These positions were resolved by mass spectrometry data and identified within one of the two AA possibilities that resulted from the cDNA sequencing (Baur et al. 2022). However, the cDNA sequence was still crucial since it was not always possible to extract enough protein from the fat samples to generate a complete sequence coverage by the mass spectrometry results. Another problem of mass spectrometry is that leucine and isoleucine, due to the same Mw, cannot be distinguished from each other.

Summarizing, these four approaches verified the IGL family assignment and the cDNA sequence. Further, the sequences correspond well to the deposited AL protein in the patients, so it can be considered that they reflect the amyloidogenic clone.

4.1.2 Challenges

One of the hallmarks of LCs is their high variability and several mechanisms are intertwined to ensure this. Not only does this concern the junctional diversity and somatic hypermutations, but also the combinatorial diversity. Up to 36 different IGLV, four IGLJ, and four IGLC genes, each can be potentially combined to finally lead to a functional lambda LC (Kawasaki et al. 1997; Lefranc and Lefranc 2020; Schroeder and Cavacini 2010). Additionally, although a dominant sequence in AL can be assumed, there is nevertheless a certain degree of sequence and respective clonal heterogeneity (Bochtler et al. 2018).

In this context, PCR-based Sanger sequencing is expected to map a certain degree of sequence heterogeneity. However, applying this method may also mean that some LC sequences will be difficult or even impossible to resolve. In the scope of this work, only 10 % (8/82) of the

generated LC sequences were not assigned to an IGLV segment, which not always concerned a not evaluable sequence. Here, cases in which specific mutations do not allow a clear distinction between two reference IGLV segments were also included.

Further, only 15 % (12/82) of the generated LC sequences showed at least ten not clearly identified AAs. In this context, in particular, the IGLV3-21 subfamily was prominent. IGLV3-21 was detected frequently in both AL and MM but AL full-length LC sequencing proved difficult. The AL sequences showed an above-average number of sequence blurs and several oligonucleotides failed to bind in FR1. In contrast, it was possible to resolve three MM sequences N-terminally and to generate a sequence without uncertain nucleotide positions twice. So, it can be postulated that the sequencing difficulties probably only relate to AL sequences. These difficulties could be caused by a mutation, insertion, or deletion that could be common or even exclusively characteristic of AL and prevent oligonucleotide binding. Interestingly, IGLV3-21 is the only detected IGLV gene that comprises three instead of only two exons – however, the first exon is described as noncoding (Ensembl). In this context, amplification could further be prevented by different splicing variants since the oligonucleotide binding site is located at the beginning of the third exon.

The IGLV3-21 next-generation sequencing experiments performed within this work demonstrated that the sequences can be assigned unambiguously to IGLV3-21 and the $MP_{\_NGS\_PCR}$ approach even verified the Sanger sequences. In contrast, the $N_{\_NGS\_PCR}$ approach only conditionally detected IGLV3-21 sequences. However, the positive control showed that the PCR protocol works in principle.

Nevertheless, the phenomenon of prominent sequence blurs in certain IGLV subfamilies and lacking oligonucleotide binding, should be studied in more detail. Bulk RNA sequencing could be helpful to investigate not only the sequences but also the composition of the expressed LCs in more detail. This approach would also shed light on the intra-clonal heterogeneity and provide more information about the actual diversity of the expressed LCs in a single AL patient.

Summarizing, within this work, 90 % of the AL LCs were assigned to a specific IGLV segment and 85 % of the sequence showed only a few uncertain AA positions.

4.2 Discrimination between Amyloidogenic and Non-Amyloidogenic Light Chains

This work primary aimed to investigate and define AL-specific lambda LC sequence features – mostly in comparison to MM lambda LCs. In the first step, both cohorts were clinically

characterized and met the expectations. Therefore, it is assumed that these are representative groups, with one limitation: The AL cohort only comprised AL_H and AL_HK patients, to perform an analysis regarding the characteristic organ tropism. In the following chapter, the identified communalities and differences between the two diseases will be discussed and put into context, to potentially infer on pathogenic mechanisms.

4.2.1 IGLV Subfamily Usage

After the clinical characterization, the IGLV subfamily usage was investigated to examine if the IGLV families present a difference in their amyloidogenic potential. Several studies have been published to address this aspect (Abraham et al. 2003; Comenzo et al. 2001; Kourelis et al. 2020; Perfetti et al. 2002; Perfetti et al. 2012; Sidana et al. 2021). The comparison of five of them showed that IGLV6-57 and IGLV3-1 are the most frequently AL-associated IGLV subfamilies Additionally, IGLV2-14 was among the commonly AL-associated IGLV subfamilies in three out of five studies (Kourelis et al. 2017; Perfetti et al. 2012; Sidana et al. 2021). The data set presented in this work is well in line with previously published preliminary data, which also included AL_K patients (Berghaus et al. 2022; AL_K patients were analyzed by Sarah Schreiner (University Hospital Heidelberg)).

Despite the comprehensive data about IGL families in AL, the analysis of exclusively AL sequences is insufficient to conclude on associations and potential crucial sequence features. Past work usually used healthy donors or LC sequences stored in the ″AL-Base″ database as a reference set. The advantage of this study is the comparison of LC sequences from clinically well-characterized AL and MM cohorts. The MM patients were treated within the GMMG-HD6 clinical trial and screened for amyloid deposits as potential exclusion criteria. Later progression to AL is a very rare phenomenon but cannot be precluded. Nevertheless, the median follow-up of the GMMG-HD6 trial was 49.8 month (Goldschmidt et al. 2021) and in total, one out of 564 patients was excluded due to developing an amyloidosis during this trial (unpublished data). This excluded MM patient presented a kappa LC and was therefore not part of the cohort analyzed within this work. So, an actual comparison of amyloidogenic and non-amyloidogenic LCs is assumed.

Two remarkable results from the comparison of AL and MM LC sequences, are a significant association of IGLV6-57 with AL LCs and the exclusive detection of IGLV2-23 for MM LCs. To compare the results of this study with a larger data set, LCs deposited in AL-Base were utilized. AL-Base is an internet LC database, established and hosted by the Boston University (the Amyloidosis Center at the Boston University School of Medicine) (Bodi et al. 2009). The stored LC sequences are classified into three major clinical groups ″AL – plasma cell disorder″,

″other – plasma cell disorders″ and ″ non – plasma cell disorders″, which can be further specified in e.g. ″AL″, ″MM″ or ″healthy individual″. Additionally, the LC sequences can be sorted by kappa and lambda and the corresponding IGV subfamilies (Bodi et al. 2009). Analyzing the deposited lambda AL (n = 416) and MM sequences (n = 91), an AL IGLV6-57 association was most frequently detected, namely in 24 % of the AL cases and only in 2 % of the MM cases. In contrast, AL sequences showed an IGLV2-23 assignment in only 3 % and MM sequences in 10 %. Thus, the data deposited in AL-Base also showed a striking difference for these two IGLV subfamilies. However, AL-Base is not a completely curated database. The stored LC sequences cannot always be assigned to a publication, which means that the methodology used is also unknown. Further, the length of the sequences varies and information about clinical parameters are usually not available.

Even though this work focused on only AL_H and AL_HK patients, an association of certain IGLV subfamilies with a specific type of organ involvement should be discussed in more detail. All studies cited above, which investigated the organ tropism, coincided with the enrichment of IGLV2-14, IGLV1-40, and especially IGLV6-57 for AL_K patients. The analysis regarding AL_H patients presented more varieties. Here, most prominent LCs of the IGLV2-14 and IGLV1-44 subfamilies were detected (followed by IGLV3-1 and IGLV3-21) (Abraham et al. 2003; Comenzo et al. 2001; Kourelis et al. 2020; Perfetti et al. 2002; Perfetti et al. 2012; Sidana et al. 2021).Within this work, a prominent association for IGLV3-1 and IGLV3-21 with AL_H as well as for IGLV6-57 with AL_HK was noted. Interestingly, IGLV6-57 was also detected for four out of eight AL_HTX LCs. The previously published own preliminary dataset is well in line with these observations (Berghaus et al. 2022; AL_K patients were analyzed by Sarah Schreiner (University Hospital Heidelberg)).

Certain trends can be noted when comparing the AL_H and AL_HK data of this work with own previously published AL_K data (Berghaus et al. 2022; AL_K patients were analyzed by Sarah Schreiner (University Hospital Heidelberg)). IGLV1 showed a more pronounced association with AL_K sequences. The IGLV2 subfamilies were found commonly in AL_H and AL_K and less in AL_HK sequences; the IGLV3 subfamilies seemed to be more associated with AL_H sequences. In contrast, the IGLV6-57 subfamily was detected equally frequently in AL_H and AL_K sequences and noticeable more often in the AL_HK sequence (Berghaus et al. 2022; AL_K patients were analyzed by Sarah Schreiner (University Hospital Heidelberg)). Accordingly, this strengthens the hypothesis that a divergent amyloidogenic potential in general but even an association towards an organ involvement can be assumed.

In addition to the organ tropism, the sequences were further stratified for a potential HC binding partner and, concerning tumor size and malignancy, for a dFLC >180 mg/L. Interestingly, IGLV2-14 was detected significant more often in AL_H HC cases, and IGLV6-57 in AL_H < dFLC cases. More general, IGLV3-1 was more frequently detected in the AL > dFLC subgroup. These aspects might influence a risk stratification based on the respective IGLV subfamily. The analyses concerning the stratifications for a present clonal HC and the dFLC will be addressed in a more detailed manner in chapter 4.3.

To conclude, the results of this study support the assumption that IGLV subfamilies present a difference in their amyloidogenic potential and are well in line with previously published studies and the data deposited in AL-Base. Additionally, certain associations were not only detected for AL and MM but also for AL_H, AL_HK, when stratifying for the presence or absence of a clonal HC, and a dFLC >180 mg/L. Therefore, it can be postulated that certain IGLV subfamilies may display favorable characteristics for different pathogenic mechanism.

4.2.2 Mutation Frequency and Count

The already addressed IGLV subfamily usage is part of the combinatory diversity, a mechanism to increase the LC diversity. This LC diversity is even further increased through junctional diversity and somatic hypermutations. It is assumed, that the introduced mutations and changes can increase the LC capability/propensity to aggregate and form fibrils (Blancas-Mejía and Ramirez-Alvarado 2013). To address this assumption, the AL and MM LC sequences were analyzed for their median mutation count and frequency.

It was striking at this point that all MM LCs displayed a median of 11.5 – 12.0 mutations in the IGLV segment. Even though the sequences were trimmed differently, the affected FR1 region was defined as the least mutated region in several IGLV subfamilies. Thus, the results can be placed in a comparative framework. In this analysis, the AL sequences showed significantly fewer mutations than the MM sequences. Further, the IGLV2-14 and IGLV3-1 MM sequences also presented a higher mutation frequency in the IGLC segment than the respective AL sequences. These findings contradict the hypothesis that an excessive accumulation of mutations leads to destabilization and misfolding of the AL-associated LCs.

An explanation for this difference between AL and MM might be that an excessive mutation rate due to genomic instability can be hypothesized for MM. In a study published by Walker et al., 37/40 MM patients showed an indicator for DNA double-strand breaks in plasma cells

(Walters et al. 2011). DNA double-strand breaks can lead to CA and thus potentially to genomic instability and, ultimately, to cancer (Walters et al. 2011). This could serve as a preliminary explanation for the higher mutational load compared to AL LCs. However, no statement about the actual germline of the patients can be made, since cDNA was sequenced within this work. Another hypothesis might be that MM is associated with a larger and therefore also even further evolved tumor clone than the ones observed in AL. In this context, it was already shown that during the tumor evolution from MGUS towards MM mutations are acquired (Walker et al. 2014), which could than serve as additional or alternative explanation for the differences between AL and MM.

Apart from the differences between AL and MM, the AL IGLV subfamilies also differed in their median mutation count. In general, a range of four up to fifteen mutations compared to a respective reference was reported for AL-associated LCs, more precisely the VL domain (Absmeier et al. 2022). In this study, IGL subfamily-dependent median mutation counts, within the range of four up to fifteen mutations, were detected.

IGLV2-14 LC sequences displayed the highest value, followed by IGLV3-1, and IGLV6-57 LCs the lowest. Here, it must be considered that IGLV2-14 is one of the most frequently detected AL and MM IGLV subfamilies, IGLV3-1 has been associated with AL more frequently, and most importantly that IGLV6-57 was detected almost exclusively in the AL cohort. Based on these results, several alternative hypotheses can be postulated which complement the assumption that certain IGLV families display already primary favorable properties for the amyloid formation: a) fewer mutations are sufficient to increase the aggregation capability of inherent aggregation prone IGLV subfamilies b) the increased mutation count in MM LCs might lower their aggregation potential c) it can be assumed that a certain number of mutations are necessary to form a functional LC.

To confirm the hypothesis of divergent primarily favorable amyloidogenic properties of IGLV subfamilies, the ‘normal’ IGLV repertoire of healthy individuals or other plasma cell disorders need to be analyzed. The data deposited in AL-Base concerning plasma cells and categorized as “normal repertoire” (n = 102) displays IGLV6-57 in 14 % of the cases and most frequently IGLV1-44 or IGLV3-19 (both 16 %). At first sight, this seems to contradict the hypothesis but again, it must be considered that the source of the deposited sequences is not validated and is only partially reproducible. This is also reflected in the median mutation counts of the IGLV6-57 LCs classified as “normal repertoire”. Here, the sequences showed a range from zero (2x) up to 19 mutations (in the analyzed AA range of this work). Nevertheless, the median mutation count was with 8.5 higher than the median mutation count of the AL IGLV6-57

sequences in this work. This might support the hypothesis, that a certain number of mutations must be acquired to form a functional LC. Ultimately, the analysis of a larger study with a well-characterized cohort is needed to confirm or reject the hypotheses.

Sequencing not only the IGLV segment but the complete LC was an important step forward of this study. In this analysis, the IGLV segment and also the IGLC segment were mutated more frequently in MM than in AL sequences. This finding might also correlate with an assumed increased genomic instability in MM patients and a possible subordinate role of the IGLC segment in fibril formation. Regarding the IGLJ segment, AL and MM sequences showed no significant differences. Here, IGLV6-57-associated AL LCs displayed a notable division. They presented approximately 20 % less frequently mutated in the IGLJ segment than AL LCs from other IGLV subfamilies. Despite the difference concerning IGLV6-57, a certain organ tropism was also noted. The AL_H sequences always displayed a higher mutation frequency than the AL_HK sequences of the corresponding IGLV subfamily (IGLV6-57: AL_HTX). A reason for this phenomenon is not yet apparent, and further studies concerning the AL organ tropism and comparisons towards AL_K and AL_D sequences are needed.

To conclude, the results of this study contradict the assumption that an accumulation of mutations could serve as an explanation for amyloidogenity. An alternative hypothesis can be postulated based on the mutation counts, shortly summarized as “quality before quantity”.

4.2.3 Mutation Hotspots and Sequence Characteristics

A more detailed analysis of the mutation hotspots and AAs exchanges was needed in the context of the postulated hypotheses. Here, the detected higher median mutation count in MM sequences was also reflected in more mutation hotspots than in the AL sequences. The low AL IGLV6-57 mutation count is also shown in this analysis since only one hotspot was identified. However, it was not possible to detect common mutation hotspots that were present exclusively in the AL sequences and would therefore discriminate between AL and MM. As side note, AL and MM sequences frequently presented the mutation hotspot at the third position of the CDR2 region, which was already described as AL hotspots in other studies (Poshusta et al. 2009; Radamaker et al. 2021b). This missing discrimination potential also concerns the IGLV-IGLJ linker region, which is part of the junctional diversity. In this context, it was also not possible to identify a disease-specific mutation pattern.

So, a more holistic approach was chosen, since the lack of clear discrimination potential by the mutation hotspots and linker region. In general, various effects of mutations that subsequently

favor the formation of fibrils can be assumed. Here, two main aspects need to be noted: a destabilizing effect that leads to the misfolding of the soluble precursor protein or a stabilizing effect of the final amyloid structure (Absmeier et al. 2022). The stabilizing effect can, for example, be achieved by interactions between stored monomers of the protofilament or within the steric zipper and cross-motifs inside the fibril core (Absmeier et al. 2022; Pradhan et al. 2020; Radamaker et al. 2019). Therefore, the insertion or deletion of charge in specific regions might induce or enhance interactions. The importance of charge was also represented in the cryo-EM structure of an IGLV1-44 fibril. Here, three mutations added charge to the surface of the fibril, one removed charge from a non-polar cavity, and one additional contributed to compensate native conserved charges. Therefore, in this case study, five out of eight mutations induced or removed charged AAs in the IGLV segment (Radamaker et al. 2019).

Strikingly, it was possible to detect differences concerning charge and small AAs between AL and MM for each IGLV subfamily analyzed within this work. Further, such differences were detected in the frequently AL-associated IGLV subfamilies (IGLV2-14, IGLV3-1, IGLV3-21) and even for IGLV1-44. These differences translated into a lower overall percentage of aspartic acid for IGLV2-14 AL and arginine for IGLV3-1 AL sequences compared to the respective MM LCs. So, individual pathogenic mechanisms for each IGLV subfamily can be hypothesized. This assumption is supported by the fact that all published fibril structures of different IGLV subfamilies showed a different morphology (Radamaker et al. 2021a; Radamaker et al. 2021b; Radamaker et al. 2019; Swuec et al. 2019).

Not only differences between AL and MM and the IGLV subfamilies but also concerning the AL organ tropism were detected. In this context, IGLV6-57 can again serve as a model to differentiate the AL_H and AL_HK sequences. Next to exclusive AL_H or AL_HK hotspots, these subgroups showed interesting differences regarding charged AAs. While both subgroups displayed an additional charge in CDR1, AL_H sequences were significantly associated with an additional negative one and AL_HK sequences with a positive one. Additionally, both subgroups displayed the introduction of the charge in the CDR2, and AL_HK sequences exclusively showed a third additional charge in the CDR3. That these deviations are limited to the CDR regions is surprising, considering its natural function to enhance the diversity of LCs. However, others demonstrated that not only mutations in the more conserved FR regions contribute to destabilization (Hurle et al. 1994). Further, mutations in CDR regions can potentially induce structural changes or enhance backbone dynamics (Absmeier et al. 2022; Blancas-Mejía et al. 2014; Rottenaicher et al. 2021).

In conclusion, the hypothesis that IGLV families probably have an inherently divergent amyloidogenic potential can be complemented by the fact that charge might also be an important factor – not only in the context of amyloidogenity but also for the AL organ tropism. Based on the results that each IGLV family presented a different mutation pattern, it can also be postulated that individual pathological sequence-based mechanisms may exist for each IGLV subfamily separately.

4.2.4 Biophysical Parameters

Various biophysical parameters were investigated to detect an effect of the mutation patterns on the LCs. This is of interest in the context of the postulated hypothesis that different IGLV families have different inherent amyloidogenic potentials. However, a direct comparison between the patient-derived LCs and their IGLV subfamilies is only possible for the GRAVY score and ΔpI value. This is due to the fact that the GRAVY value is calculated by the sum of hydropathy values of all AAs divided by the protein length. The ΔpI value was determined using the difference between the patient-derived LC and a reference sequence.

The results of this study suggest a relationship between a low GRAVY score and an increased amyloidogenic potential. However, this seems to contradict the phenomena of insoluble amyloid fibrils. The MM IGLV2-23 LC sequences showed the highest value, and the AL IGLV6-57 the lowest. At first glance, this is surprising, as it would suggest that IGLV6-57 LCs are most soluble in water in the linear state. Nevertheless, more generally, a low GRAVY score indicates that more hydrophilic AAs are present. When comparing the VBase2 IGLV reference sequences, a correlation between AL with IGLV references exhibiting especially low values was observed (**Figure 33** A). More precisely, the IGLV6-57 reference (AL enriched) shows the lowest value of -0.751. The IGLV3-1 reference (AL enriched) shows a slightly higher value of -0.635, and the IGLV2-14 reference (AL and MM) has a value of -0.512. The IGLV1-44 reference sequence showed a comparable value of -0.492 and was detected to be the third most common IGLV subfamily in MM and only in 6 % of the AL cohort. The IGLV2-23 reference (MM only) shows the highest value (-0.452), followed by IGLV3-21 (-0.490), the most frequently detected IGLV subfamily in MM. However, further studies are needed to characterize an association in more detail.

Even though different IGLV segment reference sequences exhibit distinct pI values, no direct correlation was observed (**Figure 33** A). However, the IGLV2-23 (MM only) reference shows the highest value with 6.74, followed by IGLV3-21 (5.60; MM enriched) and IGLV2-14 (5.55; AL and MM). The lowest value was noted for IGLV3-1 (4.52; AL enriched), IGLV6-57 (4.60;

AL enriched), and IGLV1-44 (4.60; MM enriched). Nevertheless, comparing the full-length IGLV2-14 and IGLV3-1 patient-derived AL and MM LC sequences, higher pI and ΔpI and values were observed for the MM LC sequences. An analysis of urinary monoclonal LCs from patients with and without AL revealed a lower mean pI of 4.8 for AL-associated LCs compared to a pI of 6.2 for non-AL LCs – which is in concordance with the results of this study (Bellotti et al. 1990). AL IGLV6-57 sequences displayed the lowest value for the AL cohort and it was noticeable that the AL_HK subgroup showed almost no deviation from the reference sequences.

The pI, in combination with the pH of the environment, can have a strong influence on the net charge of a protein, in consequence, affects its solubility and aggregation propensity (Schein 1990; Schmittschmitt and Scholtz 2003). When neglecting factors such as protein folding and modifications, proteins are most stable and insoluble if the pH of the buffer solution is within 1.0 pH unit of the protein's pI. In a previously published own study (n = 10, mixed IGLV subfamilies, AL_H, AL_K, AL_HTX), the analyzed AL LCs (IGLVJ) displayed a significant increase in pI compared to a reference sequence (Baur et al. 2022). However, this is not in concordance with the results of this larger study (n = 82, AL_H, AL_HK).

Overall, within this work different pI trends as well as differences in charged AAs were noted between AL and MM. This was also the case in the analysis regarding the AL organ tropism. In this context, it must be considered that in the case of AL, both, blood and tissue pH must be taken into account. So, the hypothesis can be postulated that not only mutations can increase the amyloidogenic potential but that these LCs must also be in the ″right″ environment to initially deposit as fibrils. These first fibrils could then, through a seeding effect, lead to the additional deposition of fibrils in other organs. However, further studies are needed to support these assumptions.

Since amyloid fibrils consist of many-stranded β-sheets, the AGG parameter, which indicates the tendency towards a β-sheet aggregation, was calculated. While the GRAVY score matches the IGLV family and disease association, this was not the case for the AGG parameter (**Figure 33** C). When comparing the IGLV reference sequences, IGLV1-44 (20.9; MM enriched) and IGLV6-57 (55.4; AL enriched) show by far the lowest value, followed by IGLV2-23 (132.5; MM only) and IGLV2-14 (145.7; AL and MM). The IGLV3-21 (566.0; MM enriched) and IGLV3-1 (524.6; AL enriched) reference sequences displayed the highest values. When comparing actual patient-derived AL and MM IGLV2-14 and IGLV3-1 LCs, different trends were observed. Additionally, divergences regarding the AL organ tropism were noted: AL_HK IGLV6-57 sequences displayed a lower value, and IGLV3-1 sequences a higher value than the respective AL_H sequences. Thus, based on the results regarding the AGG parameter, no clear

trend or correlation was identified, and the results neither support nor reject the postulated hypotheses.

In the already addressed work by Belotti et al., a significant difference in the Mw of LCs was detected. They found LC fragments with a low Mw in the urine of 30 out of 34 AL patients and only 15 out of 51 non-AL patients (Bellotti et al. 1990). In this study, no significant difference was detected but the MM sequences also showed a slightly higher Mw compared to AL sequences of the corresponding IGLV subfamily. Interestingly, this contrasts with the Mw of the IGLV reference sequences. The IGLV1-44 (10016; MM enriched) reference displays the lowest Mw, followed by IGLV2-23 (10038, MM only) and IGLV3-21 (10073, MM enriched) (**Figure 33** D). These three IGLV families showed an increased MM association. Further, IGLV3-1 (10093, AL enriched) and IGLV2-14 (10116, AL and MM) reference sequences show a higher Mw, but by far the highest value is detected for the IGLV6-57 reference (10643, AL enriched).

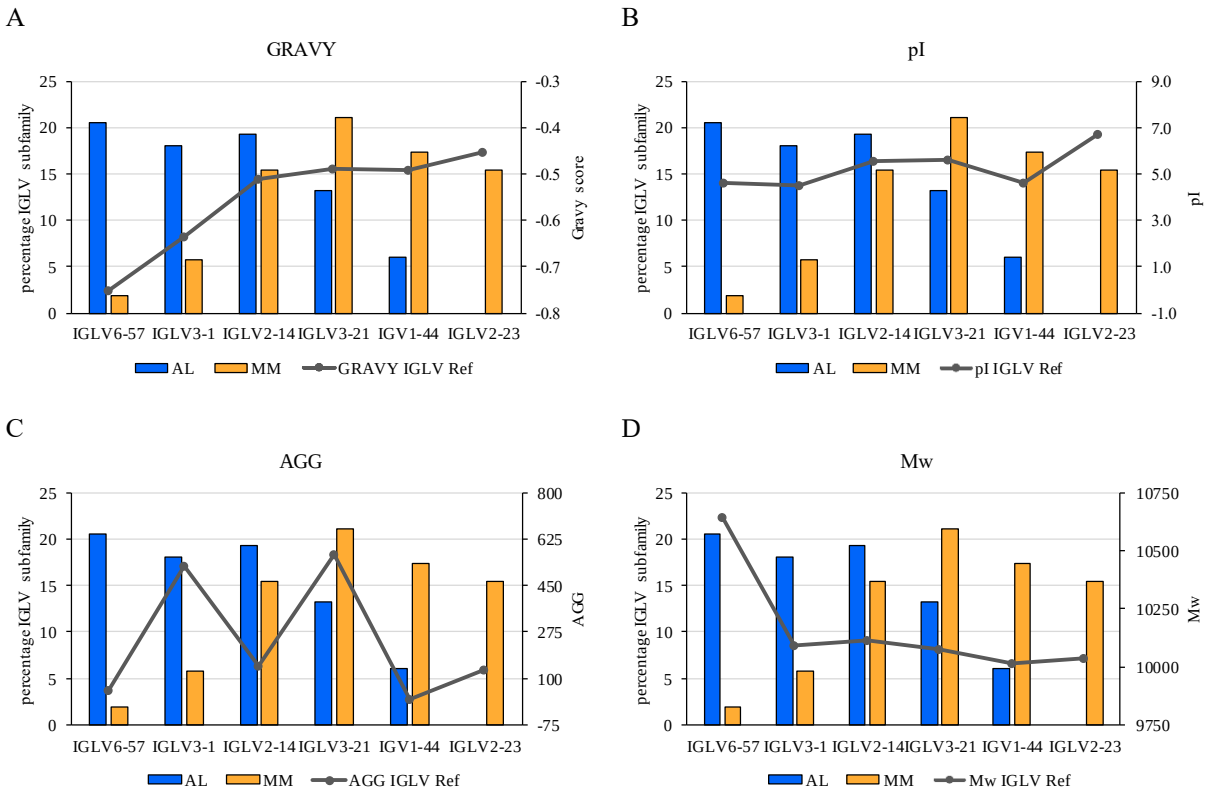


Figure 33. Correlation of the percentage of the detected IGLV subfamilies in the AL amyloidosis and multiple myeloma cohort, with the (A) grand average of hydropathicity score, (B) Isoelectric point, (C) AGG parameter and (D) the molecular weight of the respective IGLV reference sequence. AL = AL amyloidosis, MM = multiple myeloma, IGLV Ref = IGLV subfamily reference sequence, GRAVY = grand average of hydropathicity score, pI = isoelectric point, Mw = molecular weight

If postulating that different IGLV subfamilies differ in their inherent amyloidogenic potential, a possible relation between a low GRAVY score and a higher amyloidogenic potential can be assumed. As already mentioned, at first glance, an increased hydrophobicity contradicts the definition of insoluble amyloid fibrils. However, the GRAVY score only concerns the linear protein and neglects other factors influencing protein stability. Additionally, the results of this study are in concordance with the observations made by Belotti et al.: it was possible to detect a higher pI and molecular weight in MM than in AL patients. However, while the pI values partially match the IGLV subfamily/disease association, this was almost completely reversed in terms of the Mw. However, it also applies to the pI that only the linear protein is considered.

In conclusion, no additional general mechanisms were found that explain the divergent behavior of AL and MM LC based on the calculated biophysical parameters. Based on the results concerning the AL organ tropism and the pI of the LCs, it can be postulated that amyloidogenity does not only emerge from the LCs itself but that the tissue environment could also be a decisive factor.

4.3 Influence of a Potential Heavy Chain Binding Partner or dFLC >180 mg/L

In order to obtain even deeper insights into the potential pathogenic mechanisms in AL, the sequences were stratified with respect to two additional parameters. In the following chapter, the results of this work are placed in the context of the underlying hypotheses and discussed in more detail.

4.3.1 Potential Heavy Chain Binding Partner

One potential hypothesis for AL concerns the presence or absence of a potential clonal HC-binding partner (Bochtler et al. 2008). This is further supported by the fact that the CA t(11;14) is not only associated with a loss of the HC in 80 % of the cases but also more frequently associated with AL than MM (Bochtler et al. 2008; Bochtler et al. 2018). The results of this study suggest that charges and small AAs have a decisive role in the differentiation between AL and MM LCs. Further, it is assumed that each IGLV subfamily should be considered individually and may have its own pathological mechanisms. This is particularly evident when stratifying for the absence of a clonal HC.

To bring IGLV2-14 into focus, the AL HC subgroup showed a higher IGLV mutation count and comprised all LCs with a mutation in the IGLC segment. Further, differences in charge and small AAs were noted between the AL subgroups. Only the AL HC sequences showed an H41N

mutation and additional prolines in the linker region. However, the biophysical parameters showed a striking difference. Here, the AL HC sequences showed a significantly higher AGG parameter than the AL no HC sequences. This is complemented by the fact that the AL HC sequences further showed a higher pI than the AL no HC subgroup

These phenomena were also partially noted for the IGLV3-1 subfamily. Although the differences were not as prominent, all IGLV3-1 AL sequences with a mutation in the IGLC segment belonged to the AL HC subgroup. Interestingly, the IGLV2-14 and IGLV3-1 AL HC sequences presented a higher mutation frequency or count in the FR3 than the respective other subgroup. It was also noticeable that the AL IGLV3-1 CDR3 region showed a difference between these subgroups: the AL HC sequences showed a higher proportion of alanine and the AL no HC sequences of glycine. Both AAs are small and no obvious discriminating aspect can be mentioned – thus, the possible effects or consequences need to be further investigated in protein analyses.

An additional hypothesis can be formed based on these results: In the presence of a potential clonal HC binding partner, a misfit between HC and LCs might influence the excess of free LCs. This can be supported by an increased mutation count in the IGLV segment, especially in FR3, and the IGLC segment as well as certain sequence characteristics. In the context of AL, there might be a positive selection for LCs that exhibit a higher aggregation tendency or a pI close to the physiological pH of the blood and might therefore be insoluble within the patient's bloodstream.

4.3.2 Influence of a dFLC > or < 180mg/L

As expected, the detected differences and resulting hypothesis do not apply to all IGLV subfamilies. For IGLV6-57, AL_H no HC sequences showed a higher IGLJ mutation frequency, and all IGLC-mutated sequences belonged to this subgroup. However, IGLV6-57 AL sequences showed the most prominent difference when stratifying for the dFLC. This stratification approach was chosen to investigate if the size and malignancy of the B cell tumor correlate with the characteristics of the respective LCs. Although clinical risk stratifications apply different thresholds for AL and MM patients, this study cut-off was set at >180 mg/L. Here, the IGLV6-57 subfamily stood out, despite mutation number and frequency showing no significant differences between the AL dFLC subgroups.

IGLV6-57 AL < dFLC sequences frequently showed an insertion of charged AAs in the CDR2. Further, the Mw of the LCs showed a notable difference. In this context, the AL > dFLC subgroups displayed a higher Mw – not only in the IGLV6-57 but also in the IGLV2-14 subfamily. This is of special interest when considering that a larger Mw was also detected for MM compared to AL sequences. Most AL patients present with an underlying MGUS with low tumor burden and less frequently with underlying MM and a respective higher tumor load. If postulating that a higher tumor burden is associated with a higher Mw of the disease-associated LCs, this also seems reasonable for AL and a dFLC >180 mg/L. However, the underlying mechanism and consequence remain unclear.

4.4 Prediction of Amyloidogenity

The detected differences between AL and MM, as well as the subgroup analyses, can not only be used to further investigate the pathogenic mechanisms but also be brought into a more clinical context: toxicity prediction and risk stratification. A clinical risk stratification tool that could predict a patient's risk for developing AL from MGUS, SMM, or MM is highly sought-after. Identifying high-risk patients would allow monitoring strategies, early diagnosis, and early treatment and ultimately, prevent end-organ damage.

In 2021, a web-based tool for toxicity prediction of lambda LCs based on somatic mutations was established: LICTOR. The underlying machine learning analysis is said to have an 80 % accuracy in LC classification (Garofalo et al. 2021). When applying this tool to the IGLV2-14 subfamily, seven out of eight MM sequences and nine out of fourteen AL sequences were indicated as toxic. In two AL cases, the tool failed and provided no evaluation. So, when considering only the IGLV2-14 AL sequences, the accuracy is only 64 %. The overall accuracy is down to 41 % when also including the IGLV2-14 MM sequences. However, it is a newly established application, and the prediction will be improving over time, so it will potentially become a helpful tool.

Within this work, a different approach for risk stratification is needed since no mutation or mutations allowed a distinct differentiation of AL and MM LCs. In this context, the IGLV subfamily usage may play a decisive role (**Figure 34**). The findings of this study are well in line with these of previously published ones and suggest a higher amyloidogenic potential for some IGLV subfamilies. Further, the IGLV subfamilies appear to be associated with distinct organ involvement, high dFLC, or the presence of a potential HC-binding partner.

So, based on the results of this study, a new model for risk stratification can be suggested. In this context, IGLV6-57 and IGLV2-23 can serve as a model for further explanation. MGUS

patients with IGLV6-57-assigned LCs have an increased risk for disease progression towards AL: relative risk of 11.63 (p = 0.015). This is hypothesized since IGLV6-57 was significantly associated with AL. Vice versa, MGUS patients with an IGLV2-23 assigned LC should have a much lower risk for progression towards AL: relative risk 0.04 (p = 0.023). This is stated because IGLV2-23 was found exclusively in the MM cohort.

Further, not only the IGLV subfamilies presented variation but also an influence of a potential HC binding partner or the dFLC can be detected. For example, IGLV3-1 was frequently detected in the MM and AL cohort but also a significant AL > dFLC association was noted. So, IGLV subfamilies frequently detected in both cohorts must be considered in more detailed sequence analyses. More generally, an increased risk for the devolvement of AL may be indicated by a low mutation count in the IGLV and IGLC segment, as well as fewer charged AAs.

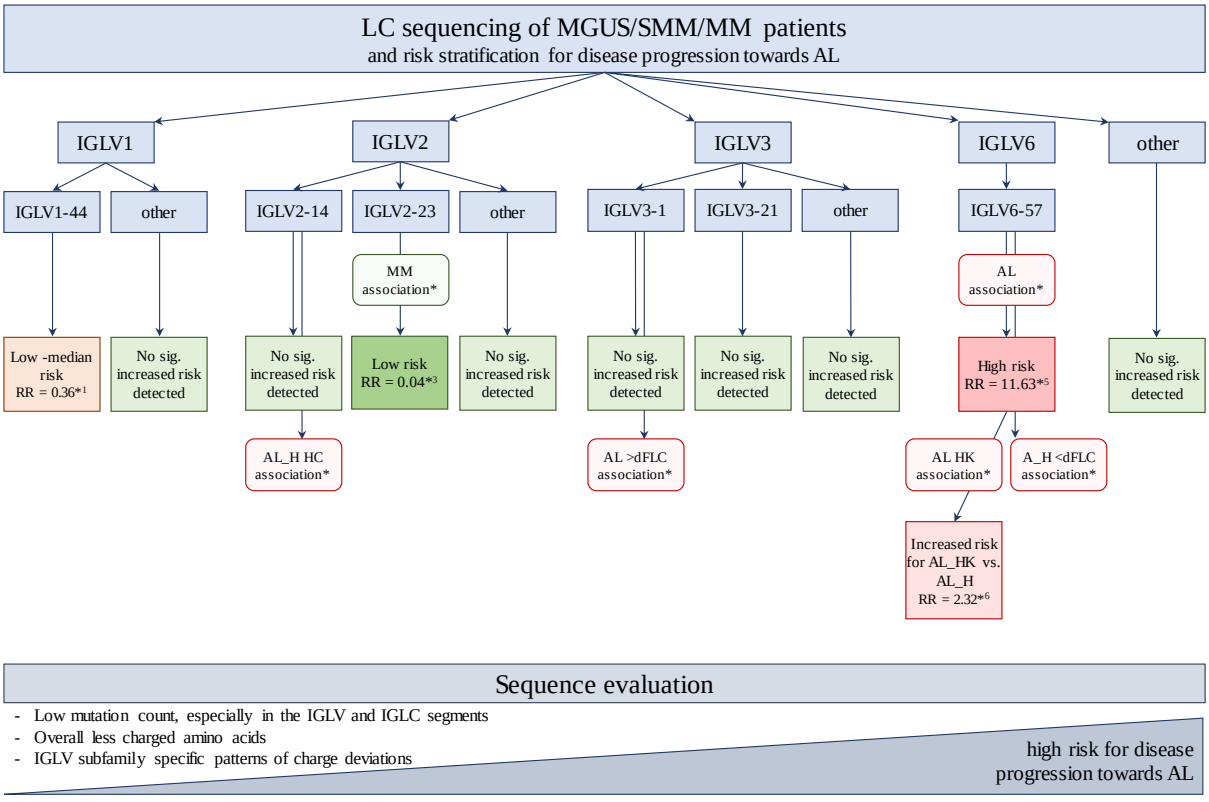


Figure 34. Risk stratification of monoclonal gammopathy of undefined significance/ smoldering multiple myeloma/multiple myeloma patients for the progression towards AL amyloidosis based on light chain sequencing results. The scheme was developed based on the data of this work and specifically refers to the detected differences in the IGLV subfamily distribution between AL amyloidosis patients with dominant heart involvement, AL amyloidosis patients with dominant heart and kidney involvement, and multiple myeloma patients. MGUS = monoclonal gammopathy of undefined significance, SMM = smoldering mutliple myeloma, MM = multiple myeloma AL = AL amyloidosis, LC = light chain, HC = heavy chain; RR = relative risk. * = p ≤0.05. The relative risk was calculated using the web-based MedCalc relative risk tool. P-values of the IGLV subfamily associations are mentioned in chapter 3.3. [1]p = 0.053, [2]p = 0.054, [3]p = 0.023, [4]p = 0.047, [5]p = 0.015, [6]p = 0.036, [7]p = 0.008.

4.5 Outlook

Among the aspects that should be studied in more detail, two thematic areas can be distinguished. The first area concerns the biology of AL and the disease underlying pathological mechanisms. Here, it is still not evident why some IGLV subfamilies are more associated with AL and others are not. To address this question, a more holistic view of the sequence and protein level should be approached. In this context, AL, MM, and even reference sequences should be compared systematically. These comparisons should provide necessary information to understand especially the IGLV subfamily distribution and the hypothesized divergent inherent amyloidogenic potential. In more detail, which biophysical parameters and sequence characteristics or modifications are ultimately decisive for LCs to form fibrils or which other factors need to be considered more intensively.

The landmark for this work has already been established based on an AL LC of the IGLV2-14 subfamily. This work examined each individual mutation site regarding its influence on fibril formation in vitro, and indeed, one mutation was crucial for fibril formation (Kazman et al. 2020).

Besides the IGLV subfamily usage, the hypothesis that a misfit between HC and LC influences the excess of free LCs still needs to be analyzed in more detail (Bochtler et al. 2008). This analysis should concern the sequence and protein level. Addressing the protein level, nuclear magnetic resonance experiments could provide information about the interaction between these two proteins in more detail. On top of this, bioinformatics models/predictions could be useful to compare references as well as MM and AL HC and LC pairs.

This study was limited by analyzing a not completely representative AL cohort; nevertheless, comparing AL_H and AL_HK sequences provided new insights. To further investigate the organ tropism in AL, the results of this work should be merged with the results of the work of Sarah Schreiner, which analyzes AL_K and AL_D sequences. In the context of the postulated hypothesis that the tissue environment may play a crucial role, further bioinformatic or protein analyses are necessary. Here, investigations of the different tissues themselves and different amyloid-associated proteins are needed (Gottwald and Röcken 2021). Beyond the LCs, it should be analyzed which pathways or even known plasma cell factor-associated factors/proteins show a deviation between AL and MM and between the AL organ involvements.

The second thematic aspect concerns risk stratification for a progression towards AL. In the context of this study, an approach based on the IGLV subfamily distribution and clinical parameters was chosen. However, not only sample sets covering a disease progression but also larger cohorts are necessary. These additional samples are also necessary to further investigate the detected sequence characteristics and differences between AL and MM and to establish them as potential biomarkers. Through this work, a preliminary model could be postulated, which, however, still needs to be expanded and verified through additional analyses.

5 Summary

Immunoglobulins and their component light chains are essential parts of the human immune system. Even though, in the context of an AL amyloidosis disease, light chains can be pathogenic. This pathogenicity arises from at least one malignant plasma cell, which infiltrates the bone marrow and produces an excess of free light chains – which misfold and ultimately deposit as fibrils in the extracellular space. Why and how these naturally soluble light chains transit into fibrils is still unknown.

One research approach relies on the natural function of the light chains themselves: antigen recognition and underlying sequence diversity. The analysis of this sequence diversity and corresponding mutations may allow making inferences about potential pathogenic mechanisms in AL amyloidosis. Nevertheless, a control group is needed to draw distinct conclusions. Here, light chains from the related plasma cell disorder multiple myeloma, which do not form fibrils, offer a suitable option. This work compared full-length lambda light chain sequences from 82 AL amyloidosis and 52 multiple myeloma patients and revealed several striking aspects – such as, a significant association between IGLV6-57 and AL amyloidosis as well as between IGLV2-23 and multiple myeloma. Overall, the results of this work are well in line with previously published studies and strengthen the assumption that IGLV subfamilies present a difference in their amyloidogenic potential. Intuitively, mutations may alter this inherent amyloidogenic potential. In the context of this work, mutation analyses showed that not the quantity but rather the distinct exchange pattern and location might be crucial. This hypothesis rests upon the findings that the multiple myeloma sequences always displayed a significantly higher mutation count than the respective AL amyloidosis sequences and the detection of different exchange patterns. In more detail, these patterns mostly concerned charged amino acids and even differed between the analyzed IGLV subfamilies. Therefore, individual pathological sequence-based mechanisms for each IGLV subfamily are hypothesized.

In addition to the characteristic of light chain fibril formation, AL amyloidosis patients often show a distinct organ tropism. To further analyze this aspect, the AL amyloidosis patients were stratified accordingly beforehand. In the end, this study included 61 patients with a dominant heart and 21 patients with a heart and kidney involvement. When comparing these subgroups, it was possible to detect differences concerning the IGLV subfamily assignment, mutation counts, and again most prominently, charged amino acids. In the context of charged amino acids, the IGLV6-57 assigned sequences stood out. Here, it was possible to detect diverging patterns between the two AL amyloidosis subgroups in several regions. Taken together, these results suggest that not exclusively the light chains themselves, but also the tissue environment

might influence their amyloidogenic potential which are mirrored in certain sequence characteristics.

Besides the already mentioned high levels of free light chains, AL amyloidosis patients frequently display a loss of the clonal heavy chain. To further investigate these parameters and to find out if they reflect in certain light chain sequence characteristics, the AL amyloidosis and multiple myeloma cohort were accordingly stratified. Addressing the level of free light chains, the most prominent difference concerned the molecular weight of the light chains. IGLV6-57 and IGLV2-14 AL amyloidosis sequences from patients with a high level of free light chains displayed a higher molecular weight than the other subgroup. The second analysis concerning the presence or absence of a clonal heavy chain revealed an increased mutation count in the IGLV and the IGLC segment of light chains with potential heavy chain binding partners. These results led to the hypothesis that mismatched heavy chains and light chains might influence the excess of free light chains. In this context, there might also be a positive selection for light chains that exhibit a higher aggregation tendency and isoelectric point close to the physiological pH of the blood – which would render them insoluble within the patient's bloodstream.

In conclusion, the results of this study lead to several intertwining hypotheses concerning the pathogenic mechanisms in AL amyloidosis: *Certain IGLV subfamilies innately present with favorable properties for amyloid formation. Here, individual pathological sequence-based mechanisms, which frequently involve alterations of charged amino acids, can be assumed. Further, not only the light chains themselves but also external factors, like the tissue environment and a potential clonal heavy chain binding partner, might influence their amyloidogenity. In this context, there might also be a positive selection for light chains which exhibit favorable characteristics like an increased aggregation tendency or a favorable isoelectric point.*

Besides the biological questions concerning the pathogenic mechanisms, there is an urgent need for a screening option for patients with a monoclonal gammopathy of undefined significance on the clinical/medical side. Identifying patients at high risk for disease progression towards AL amyloidosis would possibly prevent end-organ damage due to earlier diagnosis. Based on the results of this work, a risk stratification mainly based on the IGLV subfamily usage was established. Keeping the previously mentioned IGLV associations in mind, it is reasonable that the highest significant relative risk, 11.63, was calculated for IGLV6-57. In contrast, the lowest significant relative risk, 0.04, was detected for IGLV2-23.

In summary, this work contributes to resolving the pathogenic mechanisms of AL amyloidosis and provides a starting point for clinical risk assessment based on light chain sequencing.

6 Zusammenfassung

Immunglobuline und als ihr Bestandteil somit auch Leichtketten, sind ein wichtiger Teil des menschlichen Immunsystems. Im Rahmen einer AL Amyloidose Erkrankung können Leichtketten jedoch auch einen pathogenen Effekt haben. Dieser resultiert daraus, dass mindestens eine maligne Plasmazelle das Knochenmark infiltriert und einen Überschuss an Leichtketten produziert. Anschließend kommt es zur Fehlfaltung dieser Leichtketten, die sich schlussendlich in Form von Fibrillen im extrazellulären Raum ablagern. Warum und wie genau diese Leichtkettentransition stattfindet, ist jedoch noch nicht vollständig verstanden.

Ein Ansatz, um diesen Mechanismus näher zu untersuchen, basiert auf der natürlichen Funktion der Leichtketten: der Erkennung von Antigenen und der damit einhergehenden Sequenz Diversität. Durch eine Analyse dieser Diversität und den zugrunde liegenden Mutationen sollten Rückschlüsse auf die pathogenen Mechanismen gezogen werden können. Hierbei ist eine Kontrollgruppe essenziell. Die beste Option bieten Patienten und Patientinnen mit Multiplem Myelom – hier werden Leichtketten zwar übermäßig produziert, bilden aber keine Fibrillen.

Im Rahmen dieser Arbeit wurden die kompletten lambda Leichtketten Sequenzen von 82 AL Amyloidose und 52 Multiplem Myelom Patienten und Patientinnen miteinander verglichen. Hierbei konnte unter anderem eine signifikante Assoziation zwischen IGLV6-57 und der AL Amyloidose sowie IGLV2-23 und dem Multiplem Myelom detektiert werden. In Einklang mit bereits publizierten Daten kann somit postuliert werden, dass die IGLV Subfamilien ein unterschiedliches amyloidogenes Potenzial aufweisen. Es ist naheliegend, dass dieses amyloidogene Potenzial durch Mutationen beeinflusst werden kann. In den Mutationsanalysen zeigte sich jedoch, dass nicht die Anzahl, sondern eher der spezifische Austausch und die Region ausschlaggebend sein könnten. Diese Annahme basiert darauf, dass die Multiple Myelom Sequenzen eine signifikant höhere Anzahl von Mutationen aufwiesen und spezifische Mutationsmuster detektiert werden konnten. Diese Mutationsmuster betrafen im Besonderen geladene Aminosäuren und unterschieden sich zwischen den analysierten IGLV Subfamilien. Daher kann postuliert werden, dass es zusätzlich individuelle sequenzbasierte pathogene Mechanismen für die einzelnen IGLV Subfamilien gibt.

Ein weiteres charakteristisches Merkmal der AL Amyloidose ist der spezifische Organtropismus. Um dieses Phänomen näher zu untersuchen, wurden die Patienten und Patientinnen vor dem Beginn dieser Studie entsprechend stratifiziert, sodass die AL Amyloidose Kohorte 61 Fälle mit dominantem Herzbefall und 21 Fälle mit Herz- und Nierenbefall umfasste. Auch in diesem Vergleich konnten Unterschiede in Bezug auf die IGLV

Subfamilien und Mutationen detektiert werden. Herausstechend waren erneut Unterschiede in Bezug auf geladene Aminosäuren und insbesondere die IGLV6-57 Subfamilie. Diese Ergebnisse führten zu der Hypothese, dass nicht nur die Leichtketten selbst, sondern auch externe Faktoren wie das umliegende Gewebe einen Einfluss auf die Amyloidogenität haben und sich in spezifischen Sequenzcharakteristika wiederfinden.

Neben dem erhöhten Wert freier Leichtketten zeigen AL Amyloidose Patienten und Patientinnen ebenfalls häufig einen Verlust der klonalen schweren Kette. Um zu untersuchen, ob diese beiden Parameter sich auch in den Eigenschaften der Leichtketten widerspiegeln, wurden die Kohorten entsprechend stratifiziert. Auffallend war, dass IGLV2-14 und IGLV6-57 AL Amyloidose Sequenzen, die mit einem hohen Wert an freien Leichtketten assoziiert waren, auch ein höheres Molekulargewicht zeigten. In der zweiten Analyse bezüglich der An- oder Abwesenheit einer schweren Kette zeigten die Leichtketten Sequenzen mit potenziellem Bindungspartner mehr Mutationen im IGLV und IGLC Segment. Diese Ergebnisse führten insgesamt zu der Hypothese, dass der Überfluss an freien Leichtketten potenziell durch nicht passende schwere Ketten beeinflusst wird. Zusätzlich kann eine positive Selektion für Leichtketten, die eine erhöhte Aggregationstendenz und einen isoelektrischen Punkt nahe dem des Blutes aufweisen, angenommen werden.

Die auf dieser Arbeit basierenden Hypothesen werden im Folgenden kurz zusammengefasst: *Es gibt IGLV Subfamilien, die ein erhöhtes primäres amyloidogenes Potenzial aufweisen. In diesem Kontext können individuelle sequenzbasierte pathogene Mechanismen angenommen werden, die im Besonderen geladene Aminosäuren involvieren. Es kann ebenfalls davon ausgegangen werden, dass externe Faktoren wie das umliegende Gewebe oder auch ein potenzieller Bindungspartner einen Einfluss auf die Amyloidogenität haben. Hierbei ist eine zugrunde liegend positive Selektion für Leichtketten mit vorteilhaften Eigenschaften denkbar.*

Aus klinischer Perspektive besteht die dringende Notwendigkeit für eine Screening-Option für Patienten und Patientinnen mit einer monoklonalen Gammopathie unklarer Signifikanz hinsichtlich ihres Risikos, eine AL Amyloidose zu entwickeln. Dadurch könnten drastische Organschäden durch eine frühere Diagnose verhindert werden. Diese Arbeit nutzte im Besonderen die IGLV Subfamilien Verteilung für eine Risikoeinschätzung. Darauf basierend konnte für IGLV6-57 das höchste signifikante relative Risiko mit 11,63 und für IGLV2-23 das niedrigste signifikante relative Risiko mit 0,04 berechnet werden.

Zusammenfassend trägt diese Arbeit dazu bei, die pathogenen Mechanismen in der AL Amyloidose weiter zu charakterisieren und liefert einen Ansatzpunkt für eine Risikobewertung im klinischen Kontext.

7 Literature

Abraham, R. S., Geyer, S. M., Price-Troska, T. L., Allmer, C., Kyle, R. A., Gertz, M. A. and Fonseca, R. (2003). **Immunoglobulin light chain variable (V) region genes influence clinical presentation and outcome in light chain–associated amyloidosis (AL)**. Blood *101 (10)*, 3801-3807, doi: 10.1182/blood-2002-09-2707.

Absmeier, R. M., Rottenaicher, G. J., Svilenov, H. L., Kazman, P. and Buchner, J. (2022). **Antibodies gone bad - The molecular mechanism of light chain amyloidosis**. Febs J *290 (6)*, 1398-1419, doi: 10.1111/febs.16390.

Adachi, M., So, M., Sakurai, K., Kardos, J. and Goto, Y. (2015). **Supersaturation-limited and Unlimited Phase Transitions Compete to Produce the Pathway Complexity in Amyloid Fibrillation**. J Biol Chem *290 (29)*, 18134-18145, doi: 10.1074/jbc.M115.648139.

Ahlmann-Eltze, C. and Patil, I. (2021). **ggsignif: R Package for Displaying Significance Brackets for 'ggplot2'**. PsyArXiv, doi: 10.31234/osf.io/7awm6.

Al Hamed, R., Bazarbachi, A. H., Bazarbachi, A., Malard, F., Harousseau, J. L. and Mohty, M. (2021). **Comprehensive Review of AL amyloidosis: some practical recommendations**. Blood Cancer J *11 (5)*, 97, doi: 10.1038/s41408-021-00486-4.

Avet-Loiseau, H., Leleu, X., Roussel, M., Mathiot, C., Caillot, D., Hulin, C., Marit, G., Facon, T., Attal, M., Harousseau, J.-L., Minvielle, S. p., Campion, L. and Moreau, P. (2009). **Deletion of the 17p Chromosomal Region Is Associated with a Very Poor Outcome in Multiple Myeloma Independently of the Type of Treatment**. Blood *114 (22)*, 1817-1817, doi: 10.1182/blood.V114.22.1817.1817.

Ayala, M. V., Bender, S., Anegon, I., Menoret, S., Bridoux, F., Jaccard, A. and Sirac, C. (2021). **A rat model expressing a human amyloidogenic kappa light chain**. Amyloid *28 (3)*, 209-210, doi: 10.1080/13506129.2021.1877651.

Basset, M., Hummedah, K., Kimmich, C., Veelken, K., Dittrich, T., Brandelik, S., Kreuter, M., Hassel, J., Bosch, N., Stuhlmann-Laeisz, C., Blank, N., Müller-Tidow, C., Röcken, C., Hegenbart, U. and Schönland, S. (2020). **Localized immunoglobulin light chain amyloidosis: Novel insights including prognostic factors for local progression**. Am J Hematol *95 (10)*, 1158-1169, doi: 10.1002/ajh.25915.

Baur, J., Berghaus, N., Schreiner, S., Hegenbart, U., Schönland, S. O., Wiese, S., Huhn, S. and Haupt, C. (2022). **Identification of AL proteins from 10 λ-AL amyloidosis patients by mass spectrometry extracted from abdominal fat and heart tissue**. Amyloid *30 (1)*, 27-37, doi: 10.1080/13506129.2022.2095618.

Bellotti, V., Merlini, G., Bucciarelli, E., Perfetti, V., Quaglini, S. and Ascari, E. (1990). **Relevance of class, molecular weight and isoelectric point in predicting human light chain amyloidogenicity**. Br J Haematol *74 (1)*, 65-69, doi: 10.1111/j.1365-2141.1990.tb02539.x.

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. and Sayers, E. W. (2013). **GenBank**. Nucleic Acids Res *41 (Database issue)*, D36-42, doi: 10.1093/nar/gks1195.

Benson, M. D., Liepnieks, J. J. and Kluve-Beckerman, B. (2015). **Hereditary systemic immunoglobulin light-chain amyloidosis**. Blood *125 (21)*, 3281-3286, doi: 10.1182/blood-2014-12-618108.

Berghaus, N., Schreiner, S., Granzow, M., Müller-Tidow, C., Hegenbart, U., Schönland, S. O. and Huhn, S. (2022). **Analysis of the complete lambda light chain germline usage in patients with AL amyloidosis and dominant heart or kidney involvement**. PLoS One *17 (2)*, e0264407, doi: 10.1371/journal.pone.0264407.

Bhutani, M., Foureau, D. M., Atrash, S., Voorhees, P. M. and Usmani, S. Z. (2020). **Extramedullary multiple myeloma**. Leukemia *34 (1)*, 1-20, doi: 10.1038/s41375-019-0660-0.

Blancas-Mejía, L. M. and Ramirez-Alvarado, M. (2013). **Systemic amyloidoses**. Annu Rev Biochem *82*, 745-774, doi: 10.1146/annurev-biochem-072611-130030.

Blancas-Mejía, L. M., Tischer, A., Thompson, J. R., Tai, J., Wang, L., Auton, M. and Ramirez-Alvarado, M. (2014). **Kinetic control in protein folding for light chain amyloidosis and the differential effects of somatic mutations**. J Mol Biol *426 (2)*, 347-361, doi: 10.1016/j.jmb.2013.10.016.

Bochtler, T., Hegenbart, U., Cremer, F. W., Heiss, C., Benner, A., Hose, D., Moos, M., Bila, J., Bartram, C. R., Ho, A. D., Goldschmidt, H., Jauch, A. and Schonland, S. O. (2008). **Evaluation of the cytogenetic aberration pattern in amyloid light chain amyloidosis as compared with monoclonal gammopathy of undetermined significance reveals common pathways of karyotypic instability**. Blood *111 (9)*, 4700-4705, doi: 10.1182/blood-2007-11-122101.

Bochtler, T., Merz, M., Hielscher, T., Granzow, M., Hoffmann, K., Krämer, A., Raab, M. S., Hillengass, J., Seckinger, A., Kimmich, C., Dittrich, T., Müller-Tidow, C., Hose, D., Goldschmidt, H., Hegenbart, U., Jauch, A. and Schönland, S. O. (2018). **Cytogenetic intraclonal heterogeneity of plasma cell dyscrasia in AL amyloidosis as compared with multiple myeloma**. Blood Adv *2 (20)*, 2607-2618, doi: 10.1182/bloodadvances.2018023200.

Bodi, K., Prokaeva, T., Spencer, B., Eberhard, M., Connors, L. H. and Seldin, D. C. (2009). **AL-Base: a visual platform analysis tool for the study of amyloidogenic immunoglobulin light chain sequences**. Amyloid *16 (1)*, 1-8, doi: 10.1080/13506120802676781.

Bolotin, D. A., Poslavsky, S., Mitrophanov, I., Shugay, M., Mamedov, I. Z., Putintseva, E. V. and Chudakov, D. M. (2015). **MiXCR: software for comprehensive adaptive immunity profiling**. Nat Methods *12 (5)*, 380-381, doi: 10.1038/nmeth.3364.

Breslauer, K. J., Frank, R., Blöcker, H. and Marky, L. A. (1986). **Predicting DNA duplex stability from the base sequence**. Proc Natl Acad Sci USA *83 (11)*, 3746-3750, doi: 10.1073/pnas.83.11.3746.

Caldas, A. R., Brandao, M. and Marinho, A. (2011). **Non-Secretory Myeloma or Light Chain-Producing Multiple Myeloma: A Case Report**. J Med Cases *2 (3)*, 97-100.

Chawla, S. S., Kumar, S. K., Dispenzieri, A., Greenberg, A. J., Larson, D. R., Kyle, R. A., Lacy, M. Q., Gertz, M. A. and Rajkumar, S. V. (2015). **Clinical course and prognosis of non-secretory multiple myeloma**. Eur J Haematol *95 (1)*, 57-64, doi: 10.1111/ejh.12478.

Chesi, M., Bergsagel, P. L., Shonukan, O. O., Martelli, M. L., Brents, L. A., Chen, T., Schröck, E., Ried, T. and Kuehl, W. M. (1998). **Frequent dysregulation of the c-maf proto-oncogene at 16q23 by translocation to an Ig locus in multiple myeloma**. Blood *91 (12)*, 4457-4463.

Chilosi, M., Adami, F., Lestani, M., Montagna, L., Cimarosto, L., Semenzato, G., Pizzolo, G. and Menestrina, F. (1999). **CD138/syndecan-1: a useful immunohistochemical marker of normal and neoplastic plasma cells on routine trephine bone marrow biopsies**. Mod Pathol *12 (12)*, 1101-1106.

Chiti, F. and Dobson, C. M. (2006). **Protein misfolding, functional amyloid, and human disease**. Annu Rev Biochem *75*, 333-366, doi: 10.1146/annurev.biochem.75.101304.123901.

Chiti, F. and Dobson, C. M. (2017). **Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade**. Annu Rev Biochem *86 (1)*, 27-68, doi: 10.1146/annurev-biochem-061516-045115.

Chng, W. J., Kumar, S., VanWier, S., Ahmann, G., Price-Troska, T., Henderson, K., Chung, T.-H., Kim, S., Mulligan, G., Bryant, B., Carpten, J., Gertz, M., Rajkumar, S. V., Lacy, M., Dispenzieri, A., Kyle, R., Greipp, P., Bergsagel, P. L. and Fonseca, R. (2007). **Molecular Dissection of Hyperdiploid Multiple Myeloma by Gene Expression Profiling**. Cancer Res *67 (7)*, 2982-2989, doi: 10.1158/0008-5472.

Chung, J. B., Silverman, M. and Monroe, J. G. (2003). **Transitional B cells: step by step towards immune competence**. Trends in Immunol *24 (6)*, 342-348, doi: 10.1016/s1471-4906(03)00119-4.

Comenzo, R. L., Zhang, Y., Martinez, C., Osman, K. and Herrera, G. A. (2001). **The tropism of organ involvement in primary systemic amyloidosis: contributions of Ig VL germ line gene use and clonal plasma cell burden**. Blood *98 (3)*, 714-720, doi: 10.1182/blood.V98.3.714.

Diomede, L., Rognoni, P., Lavatelli, F., Romeo, M., del Favero, E., Cantù, L., Ghibaudi, E., di Fonzo, A., Corbelli, A., Fiordaliso, F., Palladini, G., Valentini, V., Perfetti, V., Salmona, M. and Merlini, G. (2014). **A Caenorhabditis elegans-based assay recognizes immunoglobulin light chains causing heart amyloidosis**. Blood *123 (23)*, 3543-3552, doi: 10.1182/blood-2013-10-525634.

Dispenzieri, A., Gertz, M. A., Kyle, R. A., Lacy, M. Q., Burritt, M. F., Therneau, T. M., Greipp, P. R., Witzig, T. E., Lust, J. A., Rajkumar, S. V., Fonseca, R., Zeldenrust, S. R., McGregor, C. G. and Jaffe, A. S. (2004). **Serum cardiac troponins and N-terminal pro-brain natriuretic peptide: a staging system for primary systemic amyloidosis**. J Clin Oncol *22 (18)*, 3751-3757, doi: 10.1200/jco.2004.03.029.

Dispenzieri, A., Larson, D. R., Rajkumar, S. V., Kyle, R. A., Kumar, S. K., Kourelis, T., Arendt, B., Willrcih, M., Dasari, S. and Murray, D. (2020). **N-glycosylation of monoclonal light chains on routine MASS-FIX testing is a risk factor for MGUS progression**. Leukemia *34 (10)*, 2749-2753, doi: 10.1038/s41375-020-0940-8.

Dittrich, T., Bochtler, T., Kimmich, C., Becker, N., Jauch, A., Goldschmidt, H., Ho, A. D., Hegenbart, U. and Schönland, S. O. (2017). **AL amyloidosis patients with low amyloidogenic free light chain levels at first diagnosis have an excellent prognosis**. Blood *130 (5)*, 632-642, doi: 10.1182/blood-2017-02-767475.

Duhamel, S., Mohty, D., Magne, J., Lavergne, D., Bordessoule, D., Aboyans, V. and Jaccard, A. (2017). **Incidence and Prevalence of Light Chain Amyloidosis: A Population-Based Study**. Blood *130 (Supplement 1)*, 5577-5577, doi: 10.1182/blood.V130.Suppl_1.5577.5577.

Duvaud, S., Gabella, C., Lisacek, F., Stockinger, H., Ioannidis, V. and Durinx, C. (2021). **Expasy, the Swiss Bioinformatics Resource Portal, as designed by its users**. Nucleic Acids Res *49 (W1)*, W216-w227, doi: 10.1093/nar/gkab225.

Edelman, G. M. (1959). **DISSOCIATION OF γ-GLOBULIN**. J Am Chem Soc *81 (12)*, 3155-3156, doi: 10.1021/ja01521a071.

Ehrenmann, F., Kaas, Q. and Lefranc, M. P. (2010). **IMGT/3Dstructure-DB and IMGT/DomainGapAlign: a database and a tool for immunoglobulins or antibodies, T cell receptors, MHC, IgSF and MhcSF**. Nucleic Acids Res *38 (Database issue)*, D301-307, doi: 10.1093/nar/gkp946.

Eibel, H., Kraus, H., Sic, H., Kienzler, A. K. and Rizzi, M. (2014). **B cell biology: an overview**. Curr Allergy Asthma Rep *14 (5)*, 434, doi: 10.1007/s11882-014-0434-8.

Eisenberg, D. and Jucker, M. (2012). **The Amyloid State of Proteins in Human Diseases**. Cell *148 (6)*, 1188-1203, doi: 10.1016/j.cell.2012.02.022.

Ellgaard, L. and Helenius, A. (2003). **Quality control in the endoplasmic reticulum**. Nat Rev Mol Cell Biol *4 (3)*, 181-191, doi: 10.1038/nrm1052.

Ensembl **IGLV3-21-201 Transcript**. URL: https://www.ensembl.org/Homo_sapiens/ Transcript/ Summary?db=core;g=ENSG00000211662;r=22:22711689-22713203;t=ENST00000390308 [as of 16.11.2022].

Fernández de Larrea, C., Kyle, R., Rosiñol, L., Paiva, B., Engelhardt, M., Usmani, S., Caers, J., Gonsalves, W., Schjesvold, F., Merlini, G., Lentzch, S., Ocio, E., Garderet, L., Moreau, P., Sonneveld, P., Badros, A., Gahrton, G., Goldschmidt, H., Tuchman, S., Einsele, H., Durie, B., Wirk, B., Musto, P., Hayden, P., Kaiser, M., Miguel, J. S., Bladé, J., Rajkumar, S. V. and Mateos, M. V. (2021). **Primary plasma cell leukemia: consensus definition by the International Myeloma Working Group according to peripheral blood plasma cell percentage**. Blood Cancer J *11 (12)*, 192, doi: 10.1038/s41408-021-00587-0.

Firth, J. (2019). **Haematology: multiple myeloma**. Clin Med (Lond) *19 (1)*, 58-60, doi: 10.7861/clinmedicine.19-1-58.

Fonseca, R., Bailey, R. J., Ahmann, G. J., Rajkumar, S. V., Hoyer, J. D., Lust, J. A., Kyle, R. A., Gertz, M. A., Greipp, P. R. and Dewald, G. W. (2002). **Genomic abnormalities in monoclonal gammopathy of undetermined significance**. Blood *100 (4)*, 1417-1424.

Garofalo, M., Piccoli, L., Romeo, M., Barzago, M. M., Ravasio, S., Foglierini, M., Matkovic, M., Sgrignani, J., De Gasparo, R., Prunotto, M., Varani, L., Diomede, L., Michielin, O., Lanzavecchia, A. and Cavalli, A. (2021). **Machine learning analyses of antibody somatic mutations predict immunoglobulin light chain toxicity**. Nature Commun *12 (1)*, 3532-3532, doi: 10.1038/s41467-021-23880-9.

Genesis, S. B. D. URL: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Bevoelkerungsstand/Tabellen/deutsche-nichtdeutsche-bevoelkerung-nach-geschlecht-deutschland.html#fussnote-1-249820 [as of 11.04.2023].

Gertz, M. A., Comenzo, R., Falk, R. H., Fermand, J. P., Hazenberg, B. P., Hawkins, P. N., Merlini, G., Moreau, P., Ronco, P., Sanchorawala, V., Sezer, O., Solomon, A. and Grateau, G. (2005). **Definition of organ involvement and treatment response in immunoglobulin light chain amyloidosis (AL): a consensus opinion from the 10th International Symposium on Amyloid and Amyloidosis, Tours, France, 18-22 April 2004**. Am J Hematol *79 (4)*, 319-328, doi: 10.1002/ajh.20381.

Goldschmidt, H., Mai, E. K., Bertsch, U., Besemer, B., Haenel, M., Miah, K., Fenk, R., Schlenzka, J., Munder, M., Dürig, J., Blau, I. W., Huhn, S., Hose, D., Jauch, A., Kunz, C., Neubauer, A., Scheid, C., Schroers, R., Metzler, I. v., Schieferdecker, A., Thomalla, J., Reimer, P., Mahlberg, R., Graeven, U., Kremers, S., Martens, U. M., Kunz, C., Hensel, M., Seidel-Glaetzer, A., Weisel, K., Raab, M.-S. and Salwender, H. (2021). **Elotuzumab in Combination with Lenalidomide, Bortezomib, Dexamethasone and Autologous Transplantation for Newly-Diagnosed Multiple Myeloma: Results from the Randomized Phase III GMMG-HD6 Trial**. Blood *138 (Supplement 1)*, 486-486, doi: 10.1182/blood-2021-147323.

Gottwald, J. and Röcken, C. (2021). **The amyloid proteome: a systematic review and proposal of a protein classification system**. Crit Rev Biochem Mol Biol *56 (5)*, 526-542, doi: 10.1080/10409238.2021.1937926.

Greipp, P. R., San Miguel, J., Durie, B. G., Crowley, J. J., Barlogie, B., Bladé, J., Boccadoro, M., Child, J. A., Avet-Loiseau, H., Kyle, R. A., Lahuerta, J. J., Ludwig, H., Morgan, G., Powles, R., Shimizu, K., Shustik, C., Sonneveld, P., Tosi, P., Turesson, I. and Westin, J. (2005). **International staging system for multiple myeloma**. J Clin Oncol *23 (15)*, 3412-3420, doi: 10.1200/jco.2005.04.242.

Group, I. M. W. (2003). **Criteria for the classification of monoclonal gammopathies, multiple myeloma and related disorders: a report of the International Myeloma Working Group**. Br J Haematol *121 (5)*, 749-757.

Hamidi Asl, K., Liepnieks, J. J., Nakamura, M. and Benson, M. D. (1999). **Organ-specific (localized) synthesis of Ig light chain amyloid**. J Immunol *162 (9)*, 5556-5560.

Hanamura, I., Iida, S., Akano, Y., Hayami, Y., Kato, M., Miura, K., Harada, S., Banno, S., Wakita, A., Kiyoi, H., Naoe, T., Shimizu, S., Sonta, S. I., Nitta, M., Taniwaki, M. and Ueda, R. (2001). **Ectopic expression of MAFB gene in human myeloma cells carrying (14;20)(q32;q11) chromosomal translocations**. Jpn J Cancer Res *92 (6)*, 638-644, doi: 10.1111/j.1349-7006.2001.tb01142.x.

Hartl, F. U. and Hayer-Hartl, M. (2009). **Converging concepts of protein folding in vitro and in vivo**. Nat Struct Mol Biol *16 (6)*, 574-581, doi: 10.1038/nsmb.1591.

Hemminki, K., Li, X., Försti, A., Sundquist, J. and Sundquist, K. (2012). **Incidence and survival in non-hereditary amyloidosis in Sweden**. BMC Public Health *12*, 974, doi: 10.1186/1471-2458-12-974.

Hieter, P. A., Korsmeyer, S. J., Waldmann, T. A. and Leder, P. (1981). **Human immunoglobulin κ light-chain genes are deleted or rearranged in λ-producing B cells**. Nature *290 (5805)*, 368-372, doi: 10.1038/290368a0.

Hoang, P. H., Cornish, A. J., Dobbins, S. E., Kaiser, M. and Houlston, R. S. (2019). **Mutational processes contributing to the development of multiple myeloma**. Blood Cancer J *9 (8)*, 60, doi: 10.1038/s41408-019-0221-9.

Hollis, G. F., Evans, R. J., Stafford-Hollis, J. M., Korsmeyer, S. J. and Mckearn, J. P. J. P. o. t. N. A. o. S. o. t. U. S. o. A. (1989). **Immunoglobulin lambda light-chain-related genes 14.1 and 16.1 are expressed in pre-B cells and may encode the human immunoglobulin omega light-chain protein**. Proc Natl Acad Sci USA *86 (14)*, 5552-5556, doi: 10.1073/pnas.86.14.5552.

Howe, K. L., Achuthan, P., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M., Azov, A. G., Bennett, R., Bhai, J., Billis, K., Boddu, S., Charkhchi, M., Cummins, C., Da Rin Fioretto, L., Davidson, C., Dodiya, K., El Houdaigui, B., Fatima, R., Gall, A., Garcia Giron, C., Grego, T., Guijarro-Clarke, C., Haggerty, L., Hemrom, A., Hourlier, T., Izuogu, O. G., Juettemann, T., Kaikala, V., Kay, M., Lavidas, I., Le, T., Lemos, D., Gonzalez Martinez, J., Marugán, J. C., Maurel, T., McMahon, A. C., Mohanan, S., Moore, B., Muffato, M., Oheh, D. N., Paraschas, D., Parker, A., Parton, A., Prosovetskaia, I., Sakthivel, M. P., Salam, Ahamed I A., Schmitt, B. M., Schuilenburg, H., Sheppard, D., Steed, E., Szpak, M., Szuba, M., Taylor, K., Thormann, A., Threadgold, G., Walts, B., Winterbottom, A., Chakiachvili, M., Chaubal, A., De Silva, N., Flint, B., Frankish, A., Hunt, S. E., Iisley, G. R., Langridge, N., Loveland, J. E., Martin, F. J., Mudge, J. M., Morales, J., Perry, E., Ruffier, M., Tate, J., Thybert, D., Trevanion, S. J., Cunningham, F., Yates, A. D., Zerbino, D. R. and Flicek, P. (2021). **Ensembl 2021**. Nucleic Acids Res *49 (D1)*, D884-D891, doi: 10.1093/nar/gkaa942.

Huhn, S. (2018). **ELDA qASO-PCR for High Sensitivity Detection of Tumor Cells in Bone Marrow and Peripheral Blood**. In: Multiple Myeloma: Methods and Protocols, eds. Heuck, C. and Weinhold, N., Springer New York, New York, NY, pp. 1-14.

Hurle, M. R., Helms, L. R., Li, L., Chan, W. and Wetzel, R. (1994). **A role for destabilizing amino acid replacements in light-chain amyloidosis**. Proc Natl Acad Sci USA *91 (12)*, 5446-5450, doi: 10.1073/pnas.91.12.5446.

Jin, L.-W., Claborn, K. A., Kurimoto, M., Geday, M. A., Maezawa, I., Sohraby, F., Estrada, M., Kaminksy, W. and Kahr, B. (2003). **Imaging linear birefringence and dichroism in cerebral amyloid pathologies**. Proc Natl Acad Sci USA *100 (26)*, 15294-15298, doi: 10.1073/pnas.2534647100.

Kamradt, T. and Ferrari-Kühne, K. (2011). **Adaptive immunity**. Dtsch Med Wochenschr *136 (33)*, 1678-1683, doi: 10.1055/s-0031-1281577.

Kastritis, E., Gavriatopoulou, M., Roussou, M., Migkou, M., Fotiou, D., Ziogas, D. C., Kanellias, N., Eleutherakis-Papaiakovou, E., Panagiotidis, I., Giannouli, S., Psimenou, E., Marinaki, S., Apostolou, T., Gakiopoulou, H., Tasidou, A., Papassotiriou, I., Terpos, E. and Dimopoulos, M. A. (2017). **Renal outcomes in patients with AL amyloidosis: Prognostic factors, renal response and the impact of therapy**. Am J Hematol *92 (7)*, 632-639, doi: 10.1002/ajh.24738.

Katzmann, J. A., Dispenzieri, A., Kyle, R. A., Snyder, M. R., Plevak, M. F., Larson, D. R., Abraham, R. S., Lust, J. A., Melton, L. J., 3rd and Rajkumar, S. V. (2006). **Elimination of the need for urine studies in the screening algorithm for monoclonal gammopathies by using serum immunofixation and free light chain assays**. Mayo Clin Proc *81 (12)*, 1575-1578, doi: 10.4065/81.12.1575.

Kawasaki, K., Minoshima, S., Nakato, E., Shibuya, K., Shintani, A., Schmeits, J. L., Wang, J. and Shimizu, N. (1997). **One-megabase sequence analysis of the human immunoglobulin lambda gene locus**. Genome Res *7 (3)*, 250-261, doi: 10.1101/gr.7.3.250.

Kazman, P., Vielberg, M. T., Pulido Cendales, M. D., Hunziger, L., Weber, B., Hegenbart, U., Zacharias, M., Köhler, R., Schönland, S., Groll, M. and Buchner, J. (2020). **Fatal amyloid formation in a patient's antibody light chain is caused by a single point mutation**. Elife *9*, doi: 10.7554/eLife.52300.

Khurana, U., Joshi, D., Santoshi, J. A., Sharma, T. and Kapoor, N. (2016). **Oligosecretory multiple myeloma: a case report**. Blood Res. *51 (1)*, 63-65, doi: 10.5045/br.2016.51.1.63.

Korsmeyer, S. J., Hieter, P. A., Ravetch, J. V., Poplack, D. G., Waldmann, T. A. and Leder, P. (1981). **Developmental hierarchy of immunoglobulin gene rearrangements in human leukemic pre-B-cells**. Proc Natl Acad Sci USA *78 (11)*, 7096-7100, doi: 10.1073/pnas.78.11.7096.

Kourelis, T. V., Dasari, S., Theis, J. D., Ramirez-Alvarado, M., Kurtin, P. J., Gertz, M. A., Zeldenrust, S. R., Zenka, R. M., Dogan, A. and Dispenzieri, A. (2017). **Clarifying immunoglobulin gene usage in systemic and localized immunoglobulin light-chain amyloidosis by mass spectrometry**. Blood *129 (3)*, 299-306, doi: 10.1182/blood-2016-10-743997.

Kourelis, T. V., Dasari, S. S., Dispenzieri, A., Maleszewski, J. J., Redfield, M. M., Fayyaz, A. U., Grogan, M., Ramirez-Alvarado, M., Abou Ezzeddine, O. F. and McPhail, E. D. (2020). **A Proteomic Atlas of Cardiac Amyloid Plaques**. JACC CardioOncol *2 (4)*, 632-643, doi: 10.1016/j.jaccao.2020.08.013.

Kourelis, T. V., Kumar, S. K., Gertz, M. A., Lacy, M. Q., Buadi, F. K., Hayman, S. R., Zeldenrust, S., Leung, N., Kyle, R. A., Russell, S., Dingli, D., Lust, J. A., Lin, Y., Kapoor, P., Rajkumar, S. V., McCurdy, A. and Dispenzieri, A. (2013). **Coexistent multiple myeloma or increased bone marrow plasma cells define equally high-risk populations in patients with immunoglobulin light chain amyloidosis**. J Clin Oncol *31 (34)*, 4319-4324, doi: 10.1200/jco.2013.50.8499.

Kourelis, T. V., Kumar, S. K., Go, R. S., Kapoor, P., Kyle, R. A., Buadi, F. K., Gertz, M. A., Lacy, M. Q., Hayman, S. R., Leung, N., Dingli, D., Lust, J. A., Lin, Y., Zeldenrust, S. R., Rajkumar, S. V. and Dispenzieri, A. (2014). **Immunoglobulin light chain amyloidosis is diagnosed late in patients with preexisting plasma cell dyscrasias**. Am J Hematol *89 (11)*, 1051-1054, doi: 10.1002/ajh.23827.

Kriegsmann, K., Dittrich, T., Neuber, B., Awwad, M. H. S., Hegenbart, U., Goldschmidt, H., Hillengass, J., Hose, D., Seckinger, A., Müller-Tidow, C., Ho, A. D., Schönland, S. and Hundemer, M. (2018). **Quantification of number of CD38 sites on bone marrow plasma cells in patients with light chain amyloidosis and smoldering multiple myeloma**. Cytometry B Clin Cytom *94 (5)*, 611-620, doi: 10.1002/cyto.b.21636.

Kristen, A. V., Giannitsis, E., Lehrke, S., Hegenbart, U., Konstandin, M., Lindenmaier, D., Merkle, C., Hardt, S., Schnabel, P. A., Röcken, C., Schonland, S. O., Ho, A. D., Dengler, T. J. and Katus, H. A. (2010). **Assessment of disease severity and outcome in patients with systemic light-chain amyloidosis by the high-sensitivity troponin T assay**. Blood *116 (14)*, 2455-2461, doi: 10.1182/blood-2010-02-267708.

Kuehl, W. M. and Bergsagel, P. L. (2002). **Multiple myeloma: evolving genetic events and host interactions**. Nat Rev Cancer *2 (3)*, 175-187, doi: 10.1038/nrc746.

Kumar, S., Dispenzieri, A., Lacy, M. Q., Hayman, S. R., Buadi, F. K., Colby, C., Laumann, K., Zeldenrust, S. R., Leung, N., Dingli, D., Greipp, P. R., Lust, J. A., Russell, S. J., Kyle, R. A., Rajkumar, S. V. and Gertz, M. A. (2012). **Revised prognostic staging system for light chain amyloidosis incorporating cardiac biomarkers and serum free light chain measurements**. J Clin Oncol *30 (9)*, 989-995, doi: 10.1200/jco.2011.38.5724.

Kumar, S., Murray, D., Dasari, S., Milani, P., Barnidge, D., Madden, B., Kourelis, T., Arendt, B., Merlini, G., Ramirez-Alvarado, M. and Dispenzieri, A. (2019). **Assay to rapidly screen for immunoglobulin light chain glycosylation: a potential path to earlier AL diagnosis for a subset of patients**. Leukemia *33 (1)*, 254-257, doi: 10.1038/s41375-018-0194-x.

Kumar, S., Paiva, B., Anderson, K. C., Durie, B., Landgren, O., Moreau, P., Munshi, N., Lonial, S., Bladé, J., Mateos, M.-V., Dimopoulos, M., Kastritis, E., Boccadoro, M., Orlowski, R., Goldschmidt, H., Spencer, A., Hou, J., Chng, W. J., Usmani, S. Z., Zamagni, E., Shimizu, K., Jagannath, S., Johnsen, H. E., Terpos, E., Reiman, A., Kyle, R. A., Sonneveld, P., Richardson, P. G., McCarthy, P., Ludwig, H., Chen, W., Cavo, M., Harousseau, J.-L., Lentzsch, S., Hillengass, J., Palumbo, A., Orfao, A., Rajkumar, S. V., Miguel, J. S. and Avet-Loiseau, H. (2016). **International Myeloma Working Group consensus criteria for response and minimal residual disease assessment in multiple myeloma**. Lancet Oncol *17 (8)*, e328-e346, doi: 10.1016/S1470-2045(16)30206-6.

Kyle, R. A. and Gertz, M. A. (1995). **Primary systemic amyloidosis: clinical and laboratory features in 474 cases**. Semin Hematol *32 (1)*, 45-59.

Kyle, R. A., Gertz, M. A., Witzig, T. E., Lust, J. A., Lacy, M. Q., Dispenzieri, A., Fonseca, R., Rajkumar, S. V., Offord, J. R., Larson, D. R., Plevak, M. E., Therneau, T. M. and Greipp, P. R. (2003). **Review of 1027 patients with newly diagnosed multiple myeloma**. Mayo Clin Proc *78 (1)*, 21-33, doi: 10.4065/78.1.21.

Kyle, R. A., Larson, D. R., Therneau, T. M., Dispenzieri, A., Kumar, S., Cerhan, J. R. and Rajkumar, S. V. (2018). **Long-Term Follow-up of Monoclonal Gammopathy of Undetermined Significance**. N Engl J Med *378 (3)*, 241-249, doi: 10.1056/NEJMoa1709974.

Kyle, R. A., Linos, A., Beard, C. M., Linke, R. P., Gertz, M. A., O'Fallon, W. M. and Kurland, L. T. (1992). **Incidence and natural history of primary systemic amyloidosis in Olmsted County, Minnesota, 1950 through 1989**. Blood *79 (7)*, 1817-1822.

Kyle, R. A. and Rajkumar, S. V. (2004). **Multiple Myeloma**. N Engl J Med *351 (18)*, 1860-1873, doi: 10.1056/NEJMra041875.

Kyle, R. A. and Rajkumar, S. V. (2007). **Epidemiology of the plasma-cell disorders**. Best Pract Res Clin Haematol *20 (4)*, 637-664, doi: 10.1016/j.beha.2007.08.001.

Kyle, R. A. and Rajkumar, S. V. (2008). **Multiple myeloma**. Blood *111 (6)*, 2962-2972, doi: 10.1182/blood-2007-10-078022 .

Kyle, R. A., Therneau, T. M., Rajkumar, S. V., Larson, D. R., Plevak, M. F. and Melton, L. J. (2004a). **Long-term Follow-up of 241 Patients With Monoclonal Gammopathy of Undetermined Significance: The Original Mayo Clinic Series 25 Years Later**. Mayo Clin Proc *79 (7)*, 859-866, doi: 10.4065/79.7.859.

Kyle, R. A., Therneau, T. M., Rajkumar, S. V., Larson, D. R., Plevak, M. F. and Melton, L. J. I. J. C. (2004b). **Incidence of multiple myeloma in Olmsted County, Minnesota: Trend over 6 decades**. Cancer *101 (11)*, 2667-2674, doi: 10.1002/cncr.20652.

Lee, Y. K., Brewer, J. W., Hellman, R. and Hendershot, L. M. (1999). **BiP and immunoglobulin light chain cooperate to control the folding of heavy chain and ensure the fidelity of immunoglobulin assembly**. Mol Biol Cell *10 (7)*, 2209-2219, doi: 10.1091/mbc.10.7.2209.

Lefranc, M.-P. (2014). **Immunoglobulin and T Cell Receptor Genes: IMGT® and the Birth and Rise of Immunoinformatics**. Front Immunol *5 (22)*, doi: 10.3389/fimmu.2014.00022.

Lefranc, M. P. and Lefranc, G. (2020). **Immunoglobulins or Antibodies: IMGT(®) Bridging Genes, Structures and Functions**. Biomedicines *8 (9)*, doi: 10.3390/biomedicines8090319.

López-Corral, L., Gutiérrez, N. C., Vidriales, M. B., Mateos, M. V., Rasillo, A., García-Sanz, R., Paiva, B. and San Miguel, J. F. (2011). **The progression from MGUS to smoldering myeloma and eventually to multiple myeloma involves a clonal expansion of genetically abnormal plasma cells**. Clin Cancer Res *17 (7)*, 1692-1700, doi: 10.1158/1078-0432.Ccr-10-1066.

Lousada, I., Comenzo, R. L., Landau, H., Guthrie, S. and Merlini, G. (2015). **Light Chain Amyloidosis: Patient Experience Survey from the Amyloidosis Research Consortium**. Adv Ther *32 (10)*, 920-928, doi: 10.1007/s12325-015-0250-0.

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A. R. N., Potter, S. C., Finn, R. D. and Lopez, R. (2019). **The EMBL-EBI search and sequence analysis tools APIs in 2019**. Nucleic Acids Res *47 (W1)*, W636-w641, doi: 10.1093/nar/gkz268.

Manier, S., Salem, K. Z., Park, J., Landau, D. A., Getz, G. and Ghobrial, I. M. (2017). **Genomic complexity of multiple myeloma and its clinical implications**. Nat Rev Clin Oncol *14 (2)*, 100-113, doi: 10.1038/nrclinonc.2016.122.

Marin-Argany, M., Lin, Y., Misra, P., Williams, A., Wall, J. S., Howell, K. G., Elsbernd, L. R., McClure, M. and Ramirez-Alvarado, M. (2016). **Cell Damage in Light Chain Amyloidosis: FIBRIL INTERNALIZATION, TOXICITY AND CELL-MEDIATED SEEDING**. J Biol Chem *291 (38)*, 19813-19825, doi: 10.1074/jbc.M116.736736.

Marshall, J. S., Warrington, R., Watson, W. and Kim, H. L. (2018). **An introduction to immunology and immunopathology**. Allergy Asthma Clin Immunol *14 (Suppl 2)*, 49, doi: 10.1186/s13223-018-0278-1.

McGranahan, N. and Swanton, C. (2015). **Biological and Therapeutic Impact of Intratumor Heterogeneity in Cancer Evolution**. Cancer Cell *27 (1)*, 15-26, doi: 10.1016/j.ccell.2014.12.001.

Merlini, G. (2017). **AL amyloidosis: from molecular mechanisms to targeted therapies**. Hematology *2017 (1)*, 1-12, doi: 10.1182/asheducation-2017.1.1.

Merlini, G., Dispenzieri, A., Sanchorawala, V., Schönland, S. O., Palladini, G., Hawkins, P. N. and Gertz, M. A. (2018). **Systemic immunoglobulin light chain amyloidosis**. Nat Rev Dis Primers *4 (1)*, 38, doi: 10.1038/s41572-018-0034-3.

Milani, P., Basset, M., Russo, F., Foli, A., Merlini, G. and Palladini, G. (2017). **Patients with light-chain amyloidosis and low free light-chain burden have distinct clinical features and outcome**. Blood *130 (5)*, 625-631, doi: 10.1182/blood-2017-02-767467.

Milani, P., Merlini, G. and Palladini, G. (2018). **Light Chain Amyloidosis**. Mediterr J Hematol Infect Dis *10 (1)*, e2018022, doi: 10.4084/mjhid.2018.022.

Mishra, S., Guan, J., Plovie, E., Seldin, D. C., Connors, L. H., Merlini, G., Falk, R. H., MacRae, C. A. and Liao, R. (2013). **Human amyloidogenic light chain proteins result in cardiac dysfunction, cell death, and early mortality in zebrafish**. Am J Physiol Heart Circ Physiol *305 (1)*, H95-103, doi: 10.1152/ajpheart.00186.2013.

Morgan, G. J. and Kelly, J. W. (2016). **The Kinetic Stability of a Full-Length Antibody Light Chain Dimer Determines whether Endoproteolysis Can Release Amyloidogenic Variable Domains**. J Mol Biol *428 (21)*, 4280-4297, doi: 10.1016/j.jmb.2016.08.021.

Morgan, G. J., Usher, G. A. and Kelly, J. W. (2017). **Incomplete Refolding of Antibody Light Chains to Non-Native, Protease-Sensitive Conformations Leads to Aggregation: A Mechanism of Amyloidogenesis in Patients?** Biochemistry *56 (50)*, 6597-6614, doi: 10.1021/acs.biochem.7b00579.

Morgan, G. J., Walker, B. A. and Davies, F. E. (2012). **The genetic architecture of multiple myeloma**. Nat Rev Cancer *12 (5)*, 335-348, doi: 10.1038/nrc3257.

Muchtar, E., Dispenzieri, A., Lacy, M. Q., Buadi, F. K., Kapoor, P., Hayman, S. R., Gonsalves, W., Warsame, R., Kourelis, T. V., Chakraborty, R., Russell, S., Lust, J. A., Lin, Y., Go, R. S., Zeldenrust, S., Rajkumar, S. V., Dingli, D., Leung, N., Kyle, R. A., Kumar, S. K. and Gertz, M. A. (2017a). **Overuse of organ biopsies in immunoglobulin light chain amyloidosis (AL): the consequence of failure of early recognition**. Ann Med *49 (7)*, 545-551, doi: 10.1080/07853890.2017.1304649.

Muchtar, E., Gertz, M. A., Kumar, S. K., Lacy, M. Q., Dingli, D., Buadi, F. K., Grogan, M., Hayman, S. R., Kapoor, P., Leung, N., Fonder, A., Hobbs, M., Hwa, Y. L., Gonsalves, W., Warsame, R., Kourelis, T. V., Russell, S., Lust, J. A., Lin, Y., Go, R. S., Zeldenrust, S., Kyle, R. A., Rajkumar, S. V. and Dispenzieri, A. (2017b). **Improved outcomes for newly diagnosed AL amyloidosis between 2000 and 2014: cracking the glass ceiling of early death**. Blood *129 (15)*, 2111-2119, doi: 10.1182/blood-2016-11-751628.

Muchtar, E., Gertz, M. A., Kyle, R. A., Lacy, M. Q., Dingli, D., Leung, N., Buadi, F. K., Hayman, S. R., Kapoor, P., Hwa, Y. L., Fonder, A., Hobbs, M., Gonsalves, W., Kourelis, T. V., Warsame, R., Russell, S., Lust, J. A., Lin, Y., Go, R. S., Zeldenrust, S., Rajkumar, S. V., Kumar, S. K. and Dispenzieri, A. (2019). **A Modern Primer on Light Chain Amyloidosis in 592 Patients With Mass Spectrometry-Verified Typing**. Mayo Clin Proc *94 (3)*, 472-483, doi: 10.1016/j.mayocp.2018.08.006.

Nowell, P. C. (1976). **The Clonal Evolution of Tumor Cell Populations**. Science *194 (4260)*, 23-28, doi: 10.1126/science.959840.

Obi, C. A., Mostertz, W. C., Griffin, J. M. and Judge, D. P. (2022). **ATTR Epidemiology, Genetics, and Prognostic Factors**. Methodist Debakey Cardiovasc J *18 (2)*, 17-26, doi: 10.14797/mdcvj.1066.

Obici, L., Perfetti, V., Palladini, G., Moratti, R. and Merlini, G. (2005). **Clinical aspects of systemic amyloid diseases**. Biochim Biophys Acta *1753 (1)*, 11-22, doi: 10.1016/j.bbapap.2005.08.014.

Palladini, G., Barassi, A., Klersy, C., Pacciolla, R., Milani, P., Sarais, G., Perlini, S., Albertini, R., Russo, P., Foli, A., Bragotti, L. Z., Obici, L., Moratti, R., Melzi d'Eril, G. V. and Merlini, G. (2010). **The combination of high-sensitivity cardiac troponin T (hs-cTnT) at presentation and changes in N-terminal natriuretic peptide type B (NT-proBNP) after chemotherapy best predicts survival in AL amyloidosis**. Blood *116 (18)*, 3426-3430, doi: 10.1182/blood-2010-05-286567.

Palladini, G., Campana, C., Klersy, C., Balduini, A., Vadacca, G., Perfetti, V., Perlini, S., Obici, L., Ascari, E., d'Eril, G. M., Moratti, R. and Merlini, G. (2003). **Serum N-terminal pro-brain natriuretic peptide is a sensitive marker of myocardial dysfunction in AL amyloidosis**. Circulation *107 (19)*, 2440-2445, doi: 10.1161/01.Cir.0000068314.02595.B2.

Palladini, G., Hegenbart, U., Milani, P., Kimmich, C., Foli, A., Ho, A. D., Vidus Rosin, M., Albertini, R., Moratti, R., Merlini, G. and Schönland, S. (2014). **A staging system for renal outcome and early markers of renal response to chemotherapy in AL amyloidosis**. Blood *124 (15)*, 2325-2332, doi: 10.1182/blood-2014-04-570010.

Palumbo, A., Avet-Loiseau, H., Oliva, S., Lokhorst, H. M., Goldschmidt, H., Rosinol, L., Richardson, P., Caltagirone, S., Lahuerta, J. J., Facon, T., Bringhen, S., Gay, F., Attal, M., Passera, R., Spencer, A., Offidani, M., Kumar, S., Musto, P., Lonial, S., Petrucci, M. T., Orlowski, R. Z., Zamagni, E., Morgan, G., Dimopoulos, M. A., Durie, B. G., Anderson, K. C., Sonneveld, P., San Miguel, J., Cavo, M., Rajkumar, S. V. and Moreau, P. (2015). **Revised International Staging System for Multiple Myeloma: A Report From International Myeloma Working Group**. J Clin Oncol *33 (26)*, 2863-2869, doi: 10.1200/jco.2015.61.2267.

Pawlyn, C. and Morgan, G. J. (2017). **Evolutionary biology of high-risk multiple myeloma**. Nat Rev Cancer *17 (9)*, 543-556, doi: 10.1038/nrc.2017.63.

Pepys, M. B. (2018). **The Pentraxins 1975-2018: Serendipity, Diagnostics and Drugs**. Front Immunol *9*, 2382-2382, doi: 10.3389/fimmu.2018.02382.

Perfetti, V., Casarini, S., Palladini, G., Vignarelli, M. C., Klersy, C., Diegoli, M., Ascari, E. and Merlini, G. (2002). **Analysis of Vλ-Jλ expression in plasma cells from primary (AL) amyloidosis and normal bone marrow identifies 3r(λIII) as a new amyloid-associated germline gene segment**. Blood *100 (3)*, 948-953, doi: 10.1182/blood-2002-01-0114.

Perfetti, V., Palladini, G., Casarini, S., Navazza, V., Rognoni, P., Obici, L., Invernizzi, R., Perlini, S., Klersy, C. and Merlini, G. (2012). **The repertoire of λ light chains causing predominant amyloid heart involvement and identification of a preferentially involved germline gene, IGLV1-44**. Blood *119 (1)*, 144-150, doi: 10.1182/blood-2011-05-355784.

Pinney, J. H., Smith, C. J., Taube, J. B., Lachmann, H. J., Venner, C. P., Gibbs, S. D., Dungu, J., Banypersad, S. M., Wechalekar, A. D., Whelan, C. J., Hawkins, P. N. and Gillmore, J. D. (2013). **Systemic amyloidosis in England: an epidemiological study**. Br J Haematol *161 (4)*, 525-532, doi: 10.1111/bjh.12286.

Porter, R. R. (1959). **The hydrolysis of rabbit γ-globulin and antibodies with crystalline papain**. Biochem J *73 (1)*, 119-127, doi: 10.1042/bj0730119.

Poshusta, T. L., Sikkink, L. A., Leung, N., Clark, R. J., Dispenzieri, A. and Ramirez-Alvarado, M. (2009). **Mutations in specific structural regions of immunoglobulin light chains are associated with free light chain levels in patients with AL amyloidosis**. PLoS One *4 (4)*, e5169, doi: 10.1371/journal.pone.0005169.

Pradhan, T., Annamalai, K., Sarkar, R., Huhn, S., Hegenbart, U., Schönland, S., Fändrich, M. and Reif, B. (2020). **Seeded fibrils of the germline variant of human λ-III immunoglobulin light chain FOR005 have a similar core as patient fibrils with reduced stability**. J Biol Chem *295 (52)*, 18474-18484, doi: 10.1074/jbc.RA120.016006.

PressRelease NobelPrize.org (1972). **Nobel Prize Outreach AB 2022** URL: https://www.nobelprize.org/prizes/medicine/1972/press-release/ [as of 12.10.2022].

Quarta, C. C., Gonzalez-Lopez, E., Gilbertson, J. A., Botcher, N., Rowczenio, D., Petrie, A., Rezk, T., Youngstein, T., Mahmood, S., Sachchithanantham, S., Lachmann, H. J., Fontana, M., Whelan, C. J., Wechalekar, A. D., Hawkins, P. N. and Gillmore, J. D. (2017). **Diagnostic sensitivity of abdominal fat aspiration in cardiac amyloidosis**. Eur Heart J *38 (24)*, 1905-1908, doi: 10.1093/eurheartj/ehx047.

Quock, T. P., Yan, T., Chang, E., Guthrie, S. and Broder, M. S. (2018). **Healthcare resource utilization and costs in amyloid light-chain amyloidosis: a real-world study using US claims data**. J Comp Eff Res *7 (6)*, 549-559, doi: 10.2217/cer-2017-0100.

Radamaker, L., Baur, J., Huhn, S., Haupt, C., Hegenbart, U., Schönland, S., Bansal, A., Schmidt, M. and Fändrich, M. (2021a). **Cryo-EM reveals structural breaks in a patient-derived amyloid fibril from systemic AL amyloidosis**. Nat Commun *12 (1)*, 875, doi: 10.1038/s41467-021-21126-2.

Radamaker, L., Karimi-Farsijani, S., Andreotti, G., Baur, J., Neumann, M., Schreiner, S., Berghaus, N., Motika, R., Haupt, C., Walther, P., Schmidt, V., Huhn, S., Hegenbart, U., Schönland, S. O., Wiese, S., Read, C., Schmidt, M. and Fändrich, M. (2021b). **Role of mutations and post-translational modifications in systemic AL amyloidosis studied by cryo-EM**. Nat Commun *12 (1)*, 6434, doi: 10.1038/s41467-021-26553-9.

Radamaker, L., Lin, Y.-H., Annamalai, K., Huhn, S., Hegenbart, U., Schönland, S. O., Fritz, G., Schmidt, M. and Fändrich, M. (2019). **Cryo-EM structure of a light chain-derived amyloid fibril from a patient with systemic AL amyloidosis**. Nature Commun *10 (1)*, 1103-1103, doi: 10.1038/s41467-019-09032-0.

Rajan, A. M. and Rajkumar, S. V. (2015). **Interpretation of cytogenetic results in multiple myeloma for clinical practice**. Blood Cancer J *5 (10)*, e365-e365, doi: 10.1038/bcj.2015.92.

Rajkumar, S. V. (2022). **Multiple myeloma: 2022 update on diagnosis, risk stratification, and management**. Am J Hematol *97 (8)*, 1086-1107, doi: 10.1002/ajh.26590.

Rajkumar, S. V., Dimopoulos, M. A., Palumbo, A., Blade, J., Merlini, G., Mateos, M. V., Kumar, S., Hillengass, J., Kastritis, E., Richardson, P., Landgren, O., Paiva, B., Dispenzieri, A., Weiss, B., LeLeu, X., Zweegman, S., Lonial, S., Rosinol, L., Zamagni, E., Jagannath, S., Sezer, O., Kristinsson, S. Y., Caers, J., Usmani, S. Z., Lahuerta, J. J., Johnsen, H. E., Beksac, M., Cavo, M., Goldschmidt, H., Terpos, E., Kyle, R. A., Anderson, K. C., Durie, B. G. and Miguel, J. F. (2014). **International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma**. Lancet Oncol *15 (12)*, e538-548, doi: 10.1016/s1470-2045(14)70442-5.

Rajkumar, S. V., Gertz, M. A. and Kyle, R. A. (1998). **Primary systemic amyloidosis with delayed progression to multiple myeloma**. Cancer *82 (8)*, 1501-1505.

Rajkumar, S. V., Landgren, O. and Mateos, M.-V. (2015). **Smoldering multiple myeloma**. Blood *125 (20)*, 3069-3075, doi: 10.1182/blood-2014-09-568899.

Rasche, L., Chavan, S. S., Stephens, O. W., Patel, P. H., Tytarenko, R., Ashby, C., Bauer, M., Stein, C., Deshpande, S., Wardell, C., Buzder, T., Molnar, G., Zangari, M., van Rhee, F., Thanendrarajan, S., Schinke, C., Epstein, J., Davies, F. E., Walker, B. A., Meissner, T., Barlogie, B., Morgan, G. J. and Weinhold, N. (2017). **Spatial genomic heterogeneity in multiple myeloma revealed by multi-region sequencing**. Nat Commun *8 (1)*, 268, doi: 10.1038/s41467-017-00296-y.

Retter, I., Althaus, H. H., Münch, R. and Müller, W. (2005). **VBASE2, an integrative V gene database**. Nucleic Acids Res *33 (suppl_1)*, D671-D674, doi: 10.1093/nar/gki088.

Ridley, R., Xiao, H., Hata, H., Woodliff, J., Epstein, J. and Sanderson, R. (1993). **Expression of syndecan regulates human myeloma plasma cell adhesion to type I collagen**. Blood *81 (3)*, 767-774, doi: 10.1182/blood.V81.3.767.767.

Rottenaicher, G. J., Weber, B., Rührnößl, F., Kazman, P., Absmeier, R. M., Hitzenberger, M., Zacharias, M. and Buchner, J. (2021). **Molecular mechanism of amyloidogenic mutations in hypervariable regions of antibody light chains**. J Biol Chem *296*, 100334, doi: 10.1016/j.jbc.2021.100334.

Rousseau, F., Schymkowitz, J. and Serrano, L. (2006). **Protein aggregation and amyloidosis: confusion of the kinds?** Curr Opin Struct Biol *16 (1)*, 118-126, doi: 10.1016/j.sbi.2006.01.011.

Russell, S. J. and Rajkumar, S. V. (2011). **Multiple myeloma and the road to personalised medicine**. Lancet Oncol *12 (7)*, 617-619, doi: 10.1016/s1470-2045(11)70143-7.

Salomon-Perzyński, A., Jamroziak, K. and Głodkowska-Mrówka, E. (2021). **Clonal Evolution of Multiple Myeloma-Clinical and Diagnostic Implications**. Diagnostics (Basel) *11 (9)*, 1534, doi: 10.3390/diagnostics11091534.

Salwender, H., Bertsch, U., Weisel, K., Duerig, J., Kunz, C., Benner, A., Blau, I. W., Raab, M. S., Hillengass, J., Hose, D., Huhn, S., Hundemer, M., Andrulis, M., Jauch, A., Seidel-Glaetzer, A., Lindemann, H. W., Hensel, M., Fronhoffs, S., Martens, U., Hansen, T., Wattad, M., Graeven, U., Munder, M., Fenk, R., Haenel, M., Scheid, C. and Goldschmidt, H. (2019). **Rationale and design of the German-speaking myeloma multicenter group (GMMG) trial HD6: a randomized phase III trial on the effect of elotuzumab in VRD induction/consolidation and lenalidomide maintenance in patients with newly diagnosed myeloma**. BMC Cancer *19 (1)*, 504, doi: 10.1186/s12885-019-5600-x.

Schein, C. H. (1990). **Solubility as a function of protein structure and solvent components**. Biotechnology (N Y) *8 (4)*, 308-317, doi: 10.1038/nbt0490-308.

Schiff, C., Bensmana, M., Guglielmi, P., Milili, M., Lefranc, M. P. and Fougereau, M. (1990). **The immunoglobulin lambda-like gene cluster (14.1, 16.1 and F lambda 1) contains gene(s) selectively expressed in pre-B cells and is the human counterpart of the mouse lambda 5 gene**. Int Immunol *2 (3)*, 201-207, doi: 10.1093/intimm/2.3.201.

Schmidt, A., Annamalai, K., Schmidt, M., Grigorieff, N. and Fändrich, M. (2016). **Cryo-EM reveals the steric zipper structure of a light chain-derived amyloid fibril**. Proc Natl Acad Sci USA *113 (22)*, 6200-6205, doi: 10.1073/pnas.1522282113.

Schmittschmitt, J. P. and Scholtz, J. M. (2003). **The role of protein stability, solubility, and net charge in amyloid fibril formation**. Protein Sci *12 (10)*, 2374-2378, doi: 10.1110/ps.03152903.

Schönland, S., Blank, N., Kristen, A. V., Beimler, J., Ganten, T. and Hegenbart, U. (2012). **Systemic amyloidoses**. Internist (Berl) *53 (1)*, 51-64, doi: 10.1007/s00108-011-2952-y.

Schroeder, H. W., Jr. and Cavacini, L. (2010). **Structure and function of immunoglobulins**. J Allergy Clin Immunol *125 (2 Suppl 2)*, S41-52, doi: 10.1016/j.jaci.2009.09.046.

Shi, J., Guan, J., Jiang, B., Brenner, D. A., Del Monte, F., Ward, J. E., Connors, L. H., Sawyer, D. B., Semigran, M. J., Macgillivray, T. E., Seldin, D. C., Falk, R. and Liao, R. (2010). **Amyloidogenic light chains induce cardiomyocyte contractile dysfunction and apoptosis via a non-canonical p38alpha MAPK pathway**. Proc Natl Acad Sci USA *107 (9)*, 4188-4193, doi: 10.1073/pnas.0912263107.

Sidana, S., Dasari, S., Kourelis, T. V., Dispenzieri, A., Murray, D. L., King, R. L., McPhail, E. D., Ramirez-Alvarado, M., Kumar, S. K. and Gertz, M. A. (2021). **IGVL gene region usage correlates with distinct clinical presentation in IgM vs non-IgM light chain amyloidosis**. Blood Adv *5 (8)*, 2101-2105, doi: 10.1182/bloodadvances.2020003671.

Sidana, S., Tandon, N., Dispenzieri, A., Gertz, M. A., Buadi, F. K., Lacy, M. Q., Dingli, D., Fonder, A. L., Hayman, S. R., Hobbs, M. A., Gonsalves, W. I., Hwa, Y. L., Kapoor, P., Kyle, R. A., Leung, N., Go, R. S., Lust, J. A., Russell, S. J., Zeldenrust, S. R., Rajkumar, S. V. and Kumar, S. K. (2018). **Clinical presentation and outcomes in light chain amyloidosis patients with non-evaluable serum free light chains**. Leukemia *32 (3)*, 729-735, doi: 10.1038/leu.2017.286.

Sipe, J. D., Benson, M. D., Buxbaum, J. N., Ikeda, S., Merlini, G., Saraiva, M. J. and Westermark, P. (2014). **Nomenclature 2014: Amyloid fibril proteins and clinical classification of the amyloidosis**. Amyloid *21 (4)*, 221-224, doi: 10.3109/13506129.2014.964858.

Swindells, M. B., Porter, C. T., Couch, M., Hurst, J., Abhinandan, K. R., Nielsen, J. H., Macindoe, G., Hetherington, J. and Martin, A. C. (2017). **abYsis: Integrated Antibody Sequence and Structure-Management, Analysis, and Prediction**. J Mol Biol *429 (3)*, 356-364, doi: 10.1016/j.jmb.2016.08.019.

Swuec, P., Lavatelli, F., Tasaki, M., Paissoni, C., Rognoni, P., Maritan, M., Brambilla, F., Milani, P., Mauri, P., Camilloni, C., Palladini, G., Merlini, G., Ricagno, S. and Bolognesi, M. (2019). **Cryo-EM structure of cardiac amyloid fibrils from an immunoglobulin light chain AL amyloidosis patient**. Nature Commun *10 (1)*, 1269-1269, doi: 10.1038/s41467-019-09133-w.

Tamura, K., Stecher, G. and Kumar, S. (2021). **MEGA11: Molecular Evolutionary Genetics Analysis Version 11**. Mol Biol Evol *38 (7)*, 3022-3027, doi: 10.1093/molbev/msab120.

Terpos, E., Katodritou, E., Roussou, M., Pouli, A., Michalis, E., Delimpasi, S., Parcharidou, A., Kartasis, Z., Zomas, A., Symeonidis, A., Viniou, N.-A., Anagnostopoulos, N., Economopoulos, T., Zervas, K., Dimopoulos, M. A. and on behalf of the Greek Myeloma Study Group, G. (2010). **High serum lactate dehydrogenase adds prognostic value to the international myeloma staging system even in the era of novel agents**. Eur J Haematol *85 (2)*, 114-119, doi: 10.1111/j.1600-0609.2010.01466.x.

van Gameren, I. I., Hazenberg, B. P. C., Bijzet, J., Haagsma, E. B., Vellenga, E., Posthumus, M. D., Jager, P. L. and van Rijswijk, M. H. (2010). **Amyloid load in fat tissue reflects disease severity and predicts survival in amyloidosis**. Arthritis Care Res (Hoboken) *62 (3)*, 296-301, doi: 10.1002/acr.20101.

Vrana, J. A., Theis, J. D., Dasari, S., Mereuta, O. M., Dispenzieri, A., Zeldenrust, S. R., Gertz, M. A., Kurtin, P. J., Grogg, K. L. and Dogan, A. (2014). **Clinical diagnosis and typing of systemic amyloidosis in subcutaneous fat aspirates by mass spectrometry-based proteomics**. Haematologica *99 (7)*, 1239-1247, doi: 10.3324/haematol.2013.102764.

Vu, T., Gonsalves, W., Kumar, S., Dispenzieri, A., Lacy, M. Q., Buadi, F., Gertz, M. A. and Rajkumar, S. V. (2015). **Characteristics of exceptional responders to lenalidomide-based therapy in multiple myeloma**. Blood Cancer J *5 (10)*, e363, doi: 10.1038/bcj.2015.91.

Waldmann, T. A., Strober, W. and Mogielnicki, R. P. (1972). **The renal handling of low molecular weight proteins. II. Disorders of serum protein catabolism in patients with tubular proteinuria, the nephrotic syndrome, or uremia**. J Clin Invest *51 (8)*, 2162-2174, doi: 10.1172/jci107023.

Walker, B. A., Wardell, C. P., Melchor, L., Brioli, A., Johnson, D. C., Kaiser, M. F., Mirabella, F., Lopez-Corral, L., Humphray, S., Murray, L., Ross, M., Bentley, D., Gutiérrez, N. C., Garcia-Sanz, R., San Miguel, J., Davies, F. E., Gonzalez, D. and Morgan, G. J. (2014). **Intraclonal heterogeneity is a critical early event in the development of myeloma and precedes the development of clinical symptoms**. Leukemia *28 (2)*, 384-390, doi: 10.1038/leu.2013.199.

Walker, R., Barlogie, B., Haessler, J., Tricot, G., Anaissie, E., Jr, J. D. S., Epstein, J., Hemert, R. v., Erdem, E., Hoering, A., Crowley, J., Ferris, E., Hollmig, K., Rhee, F. v., Zangari, M., Pineda-Roman, M., Mohiuddin, A., Yaccoby, S., Sawyer, J. and Angtuaco, E. J. (2007). **Magnetic Resonance Imaging in Multiple Myeloma: Diagnostic and Clinical Implications**. J Clin Oncol *25 (9)*, 1121-1128, doi: 10.1200/jco.2006.08.5803.

Walters, D. K., Wu, X., Tschumper, R. C., Arendt, B. K., Huddleston, P. M., Henderson, K. J., Dispenzieri, A. and Jelinek, D. F. (2011). **Evidence for ongoing DNA damage in multiple myeloma cells as revealed by constitutive phosphorylation of H2AX**. Leukemia *25 (8)*, 1344-1353, doi: 10.1038/leu.2011.94.

Wechalekar, A. D., Gillmore, J. D. and Hawkins, P. N. (2016). **Systemic amyloidosis**. Lancet *387 (10038)*, 2641-2654, doi: 10.1016/s0140-6736(15)01274-x.

Wechalekar, A. D., Gillmore, J. D., Wassef, N., Lachmann, H. J., Whelan, C. and Hawkins, P. N. (2011). **Abnormal N-terminal fragment of brain natriuretic peptide in patients with light chain amyloidosis without cardiac involvement at presentation is a risk factor for development of cardiac amyloidosis**. Haematologica *96 (7)*, 1079-1080, doi: 10.3324/haematol.2011.040493.

Wechalekar, A. D., Schonland, S. O., Kastritis, E., Gillmore, J. D., Dimopoulos, M. A., Lane, T., Foli, A., Foard, D., Milani, P., Rannigan, L., Hegenbart, U., Hawkins, P. N., Merlini, G. and Palladini, G. (2013). **A European collaborative study of treatment outcomes in 346 patients with cardiac stage III AL amyloidosis**. Blood *121 (17)*, 3420-3427, doi: 10.1182/blood-2012-12-473066.

Westermark, P. (2005). **Aspects on human amyloid forms and their fibril polypeptides**. FEBS J *272 (23)*, 5942-5949, doi: 10.1111/j.1742-4658.2005.05024.x.

Wickham, H. (2016). **ggplot2 Elegant Graphics for Data Analysis**, Springer, Berlin.

Wijdenes, J., Vooijs, W. C., Clément, C., Post, J., Morard, F., Vita, N., Laurent, P., Sun, R.-X., Klein, B. and Dore, J.-M. (1996). **A plasmocyte selective monoclonal antibody (B-B4) recognizes syndecan-1**. Br J Haematol *94 (2)*, 318-323, doi: 10.1046/j.1365-2141.1996.d01-1811.x.

Wilson, I. A. and Stanfield, R. L. (2021). **50 years of structural immunology**. J Biol Chem *296*, 100745, doi: 10.1016/j.jbc.2021.100745.

8 Own Contribution to Data Collection, Analysis and Publications

This work was conducted as part of the research consortium FOR2969 entitled ″Mechanisms of antibody light chain misfolding in systemic AL amyloidosis″.

In the course of clinical diagnostics, bone marrow aspirates of AL amyloidosis patients were obtained by physicians. These bone marrows aspirates were processed in equal parts by Sarah Schreiner and myself – regardless a kappa or lambda LC or the organ involvement. In general, stratification for the organ tropism was carried out by Prof. Dr. Ute Hegenbart.

Starting from GTC138^{+} cell pellets, the sample processing and targeted amplification of the lambda LCs of AL_H and AL_HK patients were performed mainly by myself (Sarah Schreiner: AL_K and AL_D patients). Twenty-two AL_H and AL_HK patients underwent bone marrow aspiration prior to the start of this study. These patients were also included in this study since leftover sample material was available in the biobank.

The underlying PCR method for the amplification of full-length lambda LCs was adapted, extended, and verified entirely by myself. In principle, this method and some forwards oligonucleotides were extracted from a published protocol by Dr. Stefanie Huhn. The cell lines initially used as PCR templates were provided by Dr. Mohammed H.S. Awwad.

The extraction and evaluation of the AL_H and AL_HK sequences and the extraction and analysis of the clinical parameters from the patient reports were also carried out entirely by myself. These two aspects together constitute the main part of this work.

The additional NGS AL IGLV3-21 experiments were planned by myself and subsequently performed together with Sarah Schreiner and Dr. Philipp Reichert. The bioinformatic processing and extraction of the LC sequences were carried out by Sarah Schreiner. The subsequent detailed sequence analysis of the AL_H samples was carried out exclusively by myself and followed the protocols described in chapters 2.9.3 to 2.9.7 (including translation).

The lambda MM LC sequences were generated using a bulk RNA sequencing approach, which was not part of this work. The subsequent bioinformatic analysis and extraction of the LCs were also not carried out as part of this work, but rather by Dr. Alexandra M. Poos. In addition to the sequences, the clinical parameters were also provided by the GMMG study group. In the framework of this work, the translation and analysis of the lambda MM LC sequences, as described in chapters 2.9.3 to 2.9.7, was exclusively performed by myself. Additionally, I extracted and analyzed the parameters necessary for a comparison from the information provided.

In summary, the lambda MM as well as AL_H and AL_HK LC sequences and associated clinical parameters were analyzed in detail and compared exclusively by myself. The subsequent evaluation, visualization, statistical analysis as well as interpretation of the data in the context of the discussion was also carried out entirely by myself.

Partial results of the present work have been pre-published in the following papers:

1. Berghaus, N., Schreiner, S., Granzow, M., Müller-Tidow, C., Hegenbart, U., Schönland, S. O. and Huhn, S. (2022). **Analysis of the complete lambda light chain germline usage in patients with AL amyloidosis and dominant heart or kidney involvement.** PloS One *17(2)*, e0264407.
2. Baur, J.*, Berghaus, N.*, Schreiner, S., Hegenbart, U., Schönland, S. O., Wiese, S., Huhn, S. and Haupt, C. (2022). **Identification of AL proteins from 10 λ-AL amyloidosis patients by mass spectrometry extracted from abdominal fat and heart tissue.** Amyloid *1–11*, doi:10.1080/13506129.2022.2095618.
3. Berghaus, N., Schreiner, S., Poos, A. M., Raab, M. S., Goldschmidt, H., Mai, E. K., Salwender, HJ., Bernhard, H., Thurner, L., Müller-Tidow, C., Weinhold, N., Hegenbart, U., Schönland S. O. and Huhn, S. (2023). **Comparison of IGLV2-14 light chain sequences of patients with AL amyloidosis or multiple myeloma.** (provisional acceptance FEBS J)

*authors contributed equally to this work.

Publication 1 included a preliminary dataset consisting of lambda LC sequences from AL_H, AL_HK, and AL_K patients. The clinical parameters of the AL_H and AL_HK patients are part of the analysis in chapter 3.1. The analysis of the IGLV families and subfamilies is part of chapter 3.3.1, and the data for the analysis of the complete LCs are included in chapters 3.3.2 and 3.3.3. The IGLV subfamily distribution of AL_K LC sequences, analyzed by Sarah Schreiner, was used for comparison in chapter 4.2.1. My contribution to the publication included the generation and analysis of the AL_H and AL_HK sequences, as well as the compilation of the corresponding patient data. The clinical parameter and sequence data concerning AL_K patients were analyzed by Sarah Schreiner. I performed the merger and visualization of the complete dataset. Additionally, the manuscript draft with the introduction, material methods, result, and discussion part were written by myself.

Publication 2 was carried out in cooperation with a research group in Ulm under the leadership of Prof. Markus Fändrich and Dr. Christian Haupt. The involved PhD student Julian Baur and I share the first authorship on this publication. The work includes sequences of AL_HTX

(n = 2), AL_H (n = 6), and AL_K (n = 2) patients, which constituted the essential basis for further conducted investigations. My contribution to this work was the generation and analysis of the AL_H sequences and the collection of clinical data. The previously published sequences of the AL_HTX sequences were also reanalyzed and put into context by myself. The patient data of the AL_H sequences are included in the analysis in chapter 3.1. The sequence analyses of the six AL_H LCs are included in the corresponding chapters for the respective IGLV subfamily (chapters 3.4, 3.9.1, 3.9.2, 3.9.7). The additional mass spectrometry experiments are included in the discussion under chapter 4.1. Additionally, I have written the corresponding material methods part, the results part as well as the part of the discussion of the manuscript that concerns the sequence data.

Publication 3 included a preliminary dataset consisting of IGLV2-14 LC sequences from AL_H, AL_HK, AL_K, and MM patients. The respective AL_H, AL_HK, and AL_K sequences were already part of **publication 1.** My contribution to this work was the generation and analysis of the AL_H and AL_HK sequences and collection of the clinical data as well as the sequence analysis of the MM LCs. The AL_K sequences were generated and analyzed by Sarah Schreiner, who also collected and analyzed the clinical data of the AL_K patients. The merger, visualization, and statistical analysis of the complete dataset was also performed by myself. Additionally, I wrote the manuscript draft including the introduction, material methods, result and discussion part.

Further publications with own contributions:

4. Radamaker, L., Karimi-Farsijani, S., Andreotti, G., Baur, J., Neumann, M., Schreiner, S., Berghaus, N., Motika, R., Haupt, C., Walther, P., Schmidt, V., Huhn, S., Hegenbart, U., Schönland, S. O., Wiese, S., Read, C., Schmidt, M. and Fändrich, M. (2021). **Role of mutations and post-translational modifications in systemic AL amyloidosis studied by cryo-EM.** Nature Commun *12(1)*, 6434, doi:10.1038/s41467-021-26553-9.
5. Schreiner, S., Berghaus, N., Poos, A. M., Raab, M. S., Besemer, B., Fenk, R., Goldschmidt, H., Mai, E. K., Müller-Tidow, C., Weinhold, N., Hegenbart, U., Huhn, S. and Schönland S. O. (2023). **Sequence diversity of kappa light chains from patients with AL amyloidosis and multiple myeloma.** (under Review, Amyloid)

Publication 4 is a cryo-EM analysis of a patient-derived AL amyloid fibril. Within this work, the newly generated fibril and associated LC sequence were compared with other known sequences of cryo-EM resolved fibrils. My part in this work was the detailed sequence analysis

of the individual LC sequences and a consequential comparison. Furthermore, I was also involved in the editing of the manuscript.

Publication 5 includes the analysis of 41 AL amyloidosis and 83 MM kappa LC sequences. The complete cohorts were analyzed regarding the IGKV and IGKJ gene usage and the AL cohort also in context of the organ involvement. Further, the AL and MM LC sequences assigned to IGKV1/D-33 were analyzed in terms of mutations and biochemical properties. I was involved in processing the underlying AL bone marrow samples and editing of the manuscript.

Appendix

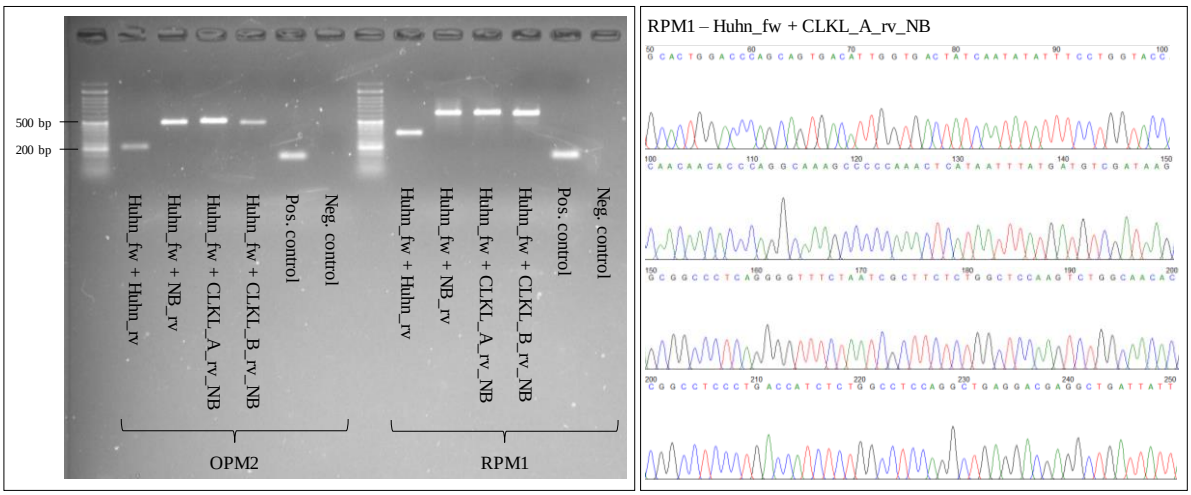
Appendix Material and Methods

Supplementary Information Table 1. Inclusion and exclusion criteria for the GMMG-HD6 trial. Retrieved from (Salwender et al. 2019). The text was extracted from the publication without changes, the layout of the table was modified. The original article was published under the creative common CC BY license (CC BY 4.0; https://creativecommons.org/licenses/by/4.0/) and the Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article.

Inclusion criteria	Exclusion criteria
Patients meeting all of the following criteria will be considered for admission to the trial:	Patients presenting with any of the following criteria will not be included in the trial:
• Confirmed diagnosis of untreated multiple myeloma requiring systemic therapy (diagnostic criteria (IMWG updated criteria (2014). For some patients systemic therapy may be required though these diagnostic criteria are not fulfilled. In this case the GMMG study office has to be consulted prior to inclusion. • Measurable disease, defined as any quantifiable monoclonal protein value, defined by at least one of the following three measurements • Measurable disease, defined as any quantifiable monoclonal protein value, defined by at least one of the following three measurements - Serum M-protein ≥10 g/l (for IgA ≥ 5 g/l) - Urine light-chain (M-protein) of ≥200 mg/24 h - Serum FLC assay: involved FLC level ≥ 10 mg/dl provided sFLC ratio is abnormal • Age 18–70 years inclusive • WHO performance status 0–3 (WHO = 3 is allowed only if caused by MM and not by co-morbid conditions) • Negative pregnancy test at inclusion (women of childbearing potential) • For all men and women of childbearing potential: patients must be willing and capable to use adequate contraception during the complete therapy. Patients must agree on the requirements regarding the lenalidomide pregnancy prevention programme. • All patients must - agree to abstain from donating blood while taking lenalidomide and for 28 days - following discontinuation of lenalidomide therapy - agree not to share study drug lenalidomide with another person and to return all - unused study drug to the investigator or pharmacist • Ability of patient to understand character and individual consequences of the clinical trial	• Patient has known hypersensitivity to any drugs given in the protocol, notably bortezomib, lenalidomide, dexamethasone and elotuzumab or to any of the constituent compounds (incl. Boron and mannitol). • Systemic AL amyloidosis (except for AL amyloidosis of the skin or the bone marrow) • Previous chemotherapy or radiotherapy during the past 5 years except local radiotherapy in case of local myeloma progression. (Note: patients may have received a cumulative dose of up to 160 mg of dexamethasone or equivalent as emergency therapy within 4 weeks prior to study entry.) • Severe cardiac dysfunction (NYHA classification III-IV, see appendix IIIB) • Significant hepatic dysfunction (serum bilirubin ≥1,8 mg/dl and/or ASAT and/or ALAT ≥ 2.5 times normal level), unless related to myeloma. (Note: if the mentioned limits for bilirubin or ASAT/ALAT are exceeded, but there is no significant hepatic dysfunction at investigator's discretion, the GMMG study office has to be consulted prior to inclusion) • Patients with renal insufficiency requiring hemodialysis • HIV positivity • Patients with active or history of hepatitis B or C • Patients with peripheral neuropathy or neuropathic pain, CTC grade 2 or higher (as defined by the NCI Common Terminology Criteria for Adverse Events (NCI CTCAE) Version 4.0, see appendix V) • Patients with a history of active malignancy during the past 5 years with the exception of basal cell carcinoma of the skin or stage 0 cervical carcinoma treated with curative intent • Patients with acute diffuse infiltrative pulmonary and/or pericardial disease • Autoimmune hemolytic anemia with positive Coombs test or immune thrombocytopenia • Platelet count < 75 × 109/l, or, dependent on bone marrow infiltration by plasma cells, platelet count < 30 × 109/l (patients with platelet count < 75 × 109/l, but > 30 × 109/l may be

Inclusion criteria	Exclusion criteria
• Written informed consent (must be available before enrollment in the trial)	eligible if percentage of plasma cells in bone marrow is ≥50%), (transfusion support within 14 days before the test is not allowed) • Haemoglobin < 8.0 g/dl, unless related to myeloma • Absolute neutrophil count (ANC) < 1.0 × 109/l (the use of colony stimulating factors within 14 days before the test is not allowed), unless related to myeloma • Pregnancy and lactation • Participation in other clinical trials. This does not include long-term follow-up periods without active drug treatment of previous studies during the last 6 months. • No patient will be allowed to enrol in this trial more than once.

A B



Supplementary Information Figure 1. Establishment of a PCR for full-length light chain sequencing. The two cell lines OPM2 and RPM1 were used for this purpose. PCR was confirmed by gel electrophoresis (A) and also by Sanger sequencing (B). (A) Huhn_fw= multiplex forward oligonucleotide set, which has been shown to bind on the common IGLV families (VLKL12a_Huhn, VLKL3c_Huhn, VLKL4a_Huhn, VLKL7a_Huhn), Huhn_rv= multiplex reverse oligonucleotide set which has been shown to bind on the common IGLJ-subfamilies (JLHD123_rv, JLKL4_rv, JLKL5_rv), NB_rv= multiplex reverse oligonucleotide set, which bind on the IGLV segment (CLKL_A_rv_NB, CLKL_B_rv_NB), Pos. controle= RPM1 as template and oligonucleotides: ß2-MRG_3 and ß2-MRG_5, Neg. control = H2O as template and oligonucleotides: V266_IgH_fw_a2 and V266_IgH_rv_a. (B) Electropherogram section of the Sanger sequencing of the RPM1 Huhn_fw Multiplex (VLKL12a_Huhn, VLKL3c_Huhn, VLKL4a_Huhn, VLKL7a_Huhn) and CLKL_A_rv_NB sample.

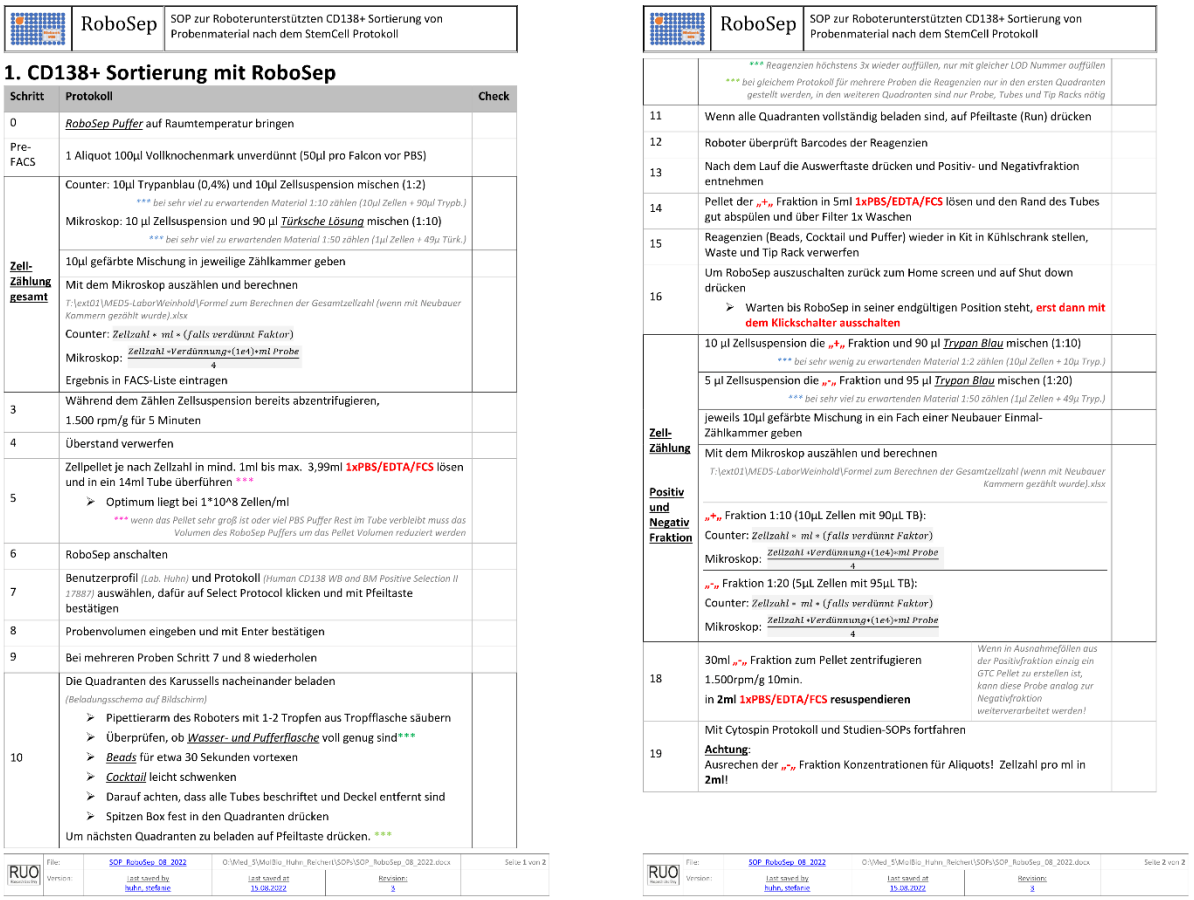
	Ficoll	SOP für Dichtegradienten-Zentrifugation von PB und KM

1. Ficoll PB & KM

Schritt	NUR TROCKENPELLET	Ficoll vor SORT
1	PB/KM - Röhrchen/Spritzen in 50ml Falcon Tube überführen Volumen PB/KM notieren	KM-Spritzen einzeln durch gelben Filter in 50ml Falcon Tube überführen Volumen KM notieren
	100µl KM an Pre-FACS abgeben *Proben >20ml auf 2 Ficols splitten* max. 20ml PB/KM pro Ficoll auf 35ml mit 1xPBS+EDTA+FCS auffüllen	100µl KM an Pre-FACS abgeben *Proben >20ml auf 2 Ficols splitten* dann auf 35ml mit 1xPBS+EDTA+FCS durch den Filter waschen dabei leicht rühren
2	Mischung auf 15ml Ficoll überschichten *Proben <5ml Proben in 15er Falcons (5ml Ficoll + 10ml Probe +Puffer)*	Mischung auf 15ml Ficoll überschichten *Proben <5ml Proben in 15er Falcons (5ml Ficoll + 10ml Probe +Puffer)*
3	30min. bei 2800rpm ohne Bremse zentrifugieren	30min. bei 2800rpm ohne Bremse zentrifugieren
4	Interphase abnehmen und Zellsuspension bis auf 50ml mit 1xPBS+EDTA+FCS auffüllen	Interphase abnehmen und über einen gelben Filter geben und mit 1xPBS+EDTA+FCS durch den Filter waschen (bis auf 50ml)
5	10min bei 1800rpm mit Bremse zentrifugieren	10min bei 1800rpm mit Bremse zentrifugieren
6	Überstand verwerfen *gesplittete Proben **VOR** der Lyse zusammenführen*	Überstand verwerfen *gesplittete Proben **VOR** der Lyse zusammenführen*
7	*** *falls Pellet nach Waschschritt noch sehr Rot:* *in 5ml EL-Puffer suspendieren* *5min einwirken lassen* *5ml 1xPBS+EDTA+FCS zum Stoppen zugeben* *10min bei 1800rpm mit Bremse zentrifugieren* *Überstand verwerfen*	**Ery-Lyse** *evtl. gesplittete Proben **VOR** der Lyse zusammenführen mit Hilfe des Ammonium Chlorid (NH_4Cl) Puffers* ➢ Zellpellet nach Ficol mit 4-6ml *NH_4Cl* 10min inkubieren ➢ auf 50ml mit 1xPBS+EDTA+FCS auffüllen ➢ Abzentrifugieren 5min. 1.500rpm ➢ Überstand verwerfen in 10ml 1xPBS+EDTA+FCS resuspendieren
8	Pellet mit 1ml 1xPBS+EDTA+FCS in 2ml Eppendorf Tube überführen	Falls gefordert (Studien-SOP, Zeitpunkte) 500µl für Trockenpellet abnehmen
9	1 min bei 13300 rpm mit Bremse zentrifugieren Überstand abnehmen mit 200µl 1xPBS **OHNE** EDTA/FCS resuspendieren	1 min bei 13300 rpm mit Bremse zentrifugieren Überstand abnehmen mit 200µl 1xPBS **OHNE** EDTA/FCS resuspendieren
10	Trockenpellet bei -20°C im Tiefkühlschrank lagern.	Trockenpellet bei -20°C im Tiefkühlschrank lagern. Restmaterial: weiter nach SORT SOP

RUO	File:	SOP_Ficoll_08_2022	O:\Med_5\MolBio_Huhn_Reichert\SOPs\SOP_Ficoll_08_2022.docx		Seite 1 von 1
	Version:	Last saved by huhn, stefanie	Last saved at 15.08.2022	Revision: 4	

Supplementary Information Figure 2. Internal revised quality document for the preparation of bone marrow using a density gradient centrifugation. The protocol „Ficoll vor Sort“ was used. SOP: “SOP für Dichtegradienten-Zentrifugation von BP und KM”, last saved by Stefanie Huhn on 15.08.2022. Stefanie Huhn agreed to the usage of this document in this work.



RoboSep | SOP zur Roboterunterstützten CD138+ Sortierung von Probenmaterial nach dem StemCell Protokoll

1. CD138+ Sortierung mit RoboSep

Schritt	Protokoll	Check
0	*RoboSep Puffer* auf Raumtemperatur bringen	
Pre-FACS	1 Aliquot 100µl Vollknochenmark unverdünnt (50µl pro Falcon vor PBS)	
Zell-Zählung gesamt	Counter: 10µl Trypanblau (0,4%) und 10µl Zellsuspension mischen (1:2) **** bei sehr viel zu erwartenden Material 1:10 zählen (10µl Zellen + 90µl Trypb.)* Mikroskop: 10 µl Zellsuspension und 90 µl *Türksche Lösung* mischen (1:10) **** bei sehr viel zu erwartenden Material 1:50 zählen (1µl Zellen + 49µ Türk.)*	
	10µl gefärbte Mischung in jeweilige Zählkammer geben	
	Mit dem Mikroskop auszählen und berechnen *T:\ext01\MED5-LaborWeinhold\Formel zum Berechnen der Gesamtzellzahl (wenn mit Neubauer Kammern gezählt wurde).xlsx* Counter: $Zellzahl * ml * (falls\ verdünnt\ Faktor)$ Mikroskop: $\frac{Zellzahl * Verdünnung * (1e4) * ml\ Probe}{4}$ Ergebnis in FACS-Liste eintragen	
3	Während dem Zählen Zellsuspension bereits abzentrifugieren, 1.500 rpm/g für 5 Minuten	
4	Überstand verwerfen	
5	Zellpellet je nach Zellzahl in mind. 1ml bis max. 3,99ml **1xPBS/EDTA/FCS** lösen und in ein 14ml Tube überführen *** ➢ Optimum liegt bei 1*10^8 Zellen/ml **** wenn das Pellet sehr groß ist oder viel PBS Puffer Rest im Tube verbleibt muss das Volumen des RoboSep Puffers um das Pellet Volumen reduziert werden*	
6	RoboSep anschalten	
7	Benutzerprofil *(Lab. Huhn)* und Protokoll *(Human CD138 WB and BM Positive Selection II 17887)* auswählen, dafür auf Select Protocol klicken und mit Pfeiltaste bestätigen	
8	Probenvolumen eingeben und mit Enter bestätigen	
9	Bei mehreren Proben Schritt 7 und 8 wiederholen	
10	Die Quadranten des Karussells nacheinander beladen *(Beladungsschema auf Bildschirm)* ➢ Pipettierarm des Roboters mit 1-2 Tropfen aus Tropfflasche säubern ➢ Überprüfen, ob *Wasser- und Pufferflasche* voll genug sind*** ➢ *Beads* für etwa 30 Sekunden vortexen ➢ *Cocktail* leicht schwenken ➢ Darauf achten, dass alle Tubes beschriftet und Deckel entfernt sind ➢ Spitzen Box fest in den Quadranten drücken Um nächsten Quadranten zu beladen auf Pfeiltaste drücken. ***	

RUO | File: SOP_RoboSep_08_2022 | Version: | Last saved by huhn, stefanie | Last saved at 15.08.2022 | Revision: 3 | O:\Med_5\MolBio_Huhn_Reichert\SOPs\SOP_RoboSep_08_2022.docx | Seite 1 von 2

RoboSep | SOP zur Roboterunterstützten CD138+ Sortierung von Probenmaterial nach dem StemCell Protokoll

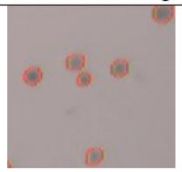
	**** Reagenzien höchstens 3x wieder auffüllen, nur mit gleicher LOD Nummer auffüllen* **** bei gleichem Protokoll für mehrere Proben die Reagenzien nur in den ersten Quadranten gestellt werden, in den weiteren Quadranten sind nur Probe, Tubes und Tip Racks nötig*	
11	Wenn alle Quadranten vollständig beladen sind, auf Pfeiltaste (Run) drücken	
12	Roboter überprüft Barcodes der Reagenzien	
13	Nach dem Lauf die Auswerftaste drücken und Positiv- und Negativfraktion entnehmen	
14	Pellet der „+" Fraktion in 5ml **1xPBS/EDTA/FCS** lösen und den Rand des Tubes gut abspülen und über Filter 1x Waschen	
15	Reagenzien (Beads, Cocktail und Puffer) wieder in Kit in Kühlschrank stellen, Waste und Tip Rack verwerfen	
16	Um RoboSep auszuschalten zurück zum Home screen und auf Shut down drücken ➢ Warten bis RoboSep in seiner endgültigen Position steht, **erst dann mit dem Klickschalter ausschalten**	
Zell-Zählung **Positiv und Negativ Fraktion**	10 µl Zellsuspension die „+" Fraktion und 90 µl *Trypan Blau* mischen (1:10) **** bei sehr wenig zu erwartenden Material 1:2 zählen (10µl Zellen + 10µ Tryp.)* 5 µl Zellsuspension die „-" Fraktion und 95 µl *Trypan Blau* mischen (1:20) **** bei sehr viel zu erwartenden Material 1:50 zählen (1µl Zellen + 49µ Tryp.)* jeweils 10µl gefärbte Mischung in ein Fach einer Neubauer Einmal-Zählkammer geben Mit dem Mikroskop auszählen und berechnen *T:\ext01\MED5-LaborWeinhold\Formel zum Berechnen der Gesamtzellzahl (wenn mit Neubauer Kammern gezählt wurde).xlsx* „+" Fraktion 1:10 (10µL Zellen mit 90µL TB): Counter: $Zellzahl * ml * (falls\ verdünnt\ Faktor)$ Mikroskop: $\frac{Zellzahl * Verdünnung * (1e4) * ml\ Probe}{4}$ „-" Fraktion 1:20 (5µL Zellen mit 95µL TB): Counter: $Zellzahl * ml * (falls\ verdünnt\ Faktor)$ Mikroskop: $\frac{Zellzahl * Verdünnung * (1e4) * ml\ Probe}{4}$	
18	30ml „-" Fraktion zum Pellet zentrifugieren 1.500rpm/g 10min. in **2ml 1xPBS/EDTA/FCS resuspendieren**	*Wenn in Ausnahmefällen aus der Positivfraktion einzig ein GTC Pellet zu erstellen ist, kann diese Probe analog zur Negativfraktion weiterverarbeitet werden!*
19	Mit Cytospin Protokoll und Studien-SOPs fortfahren **Achtung**: Ausrechen der „-" Fraktion Konzentrationen für Aliquots! Zellzahl pro ml in **2ml**!	

RUO | File: SOP_RoboSep_08_2022 | Version: | Last saved by huhn, stefanie | Last saved at 15.08.2022 | Revision: 3 | O:\Med_5\MolBio_Huhn_Reichert\SOPs\SOP_RoboSep_08_2022.docx | Seite 2 von 2

Supplementary Information Figure 3. Internal revised quality document for the CD138$^+$ cell sorting following the Stem cell protocol and using a robotry. SOP: "SOP zur Roboterunterstützten CD138$^+$ Sortierung von Probenmaterial nach dem Stelm Cell Protocoll", last saved by Stefanie Huhn on 15.08.2022. Stefanie Huhn agreed to the usage of this document in this work.

	Countess	SOP zum Zählen der Zellzahl von Proben mittels Countess Cell-Counter

1. Zählen der Zellzahl mittels Countess Cell-Counter

Schritt	Protokoll
0	Counter ist in der Regel eingeschaltet, durch Tippen auf Bildschirm testen; sonst Einschaltknopf am Gerät hinten bedienen
CAVE!	**Nur** Gesamtzellzahl und "-" Fraktion am Counter zählen, "+" Fraktion wird manuell am Mikroskop gezählt *siehe SOP zum Sortieren der KM-Proben "Zellzählung Positiv Fraktion"*
Mo. u.n.B.	Countess test beads 1:2 mit Trypanblau mischen, 10µl in "Countess Cell" Zählkammer pipettieren; Ergebnis sollte sein: $1x10^6$/mL ± 10%
Vorbereiten der Probe	10µl Trypanblau (0,4%) und 10µl Zellsuspension mischen (1:2) *** bei sehr großem Pellet eine höhere Verdünnung wählen (z.B. 1:10 [10µl Zellen + 90µl TB] oder höher); Der optimale Messbereich liegt bei 1x 10^5 - 4x 10^6 Zellen/mL (Anzeige am Counter /mL (Anzeige am Counter)
	10µl Mix in "Countess Cell" Zählkammer geben (ca. 30sek setzen lassen)
	Zählkammer in den Counter schieben; korrekte Position ist erreicht, wenn man einen soften Klick hört
Zellzählung	In der linken oberen Ecke ist das laufende Profil (gängig ist: *"Standard"*) zu sehen; sollte dies nicht der Fall sein, auf "Standard", dann auf **"Adjust"** klicken, dann das notwendige Profil auswählen und laden (**"Load"**)
	Das gängige Profil "Standard" hat die Einstellungen:
	 
	Die Zellen werden automatisch fokussiert, danach müssen Zellen mit hellem Kern sichtbar sein (Capture Bildschirm)
	Einstellungen anpassen: Helligkeit optimieren (**"Light source"** verschieben auf ± 10)
	Evtl. solange manuell nachfokussieren bis Ringe zu sehen sind ("Focus" Button drücken und mit dem Slider verschieben), in der Regel liegt der Fokus bei ca. 2900±; wenn nötig, auch den Zoom hinzuziehen ("Zoom" Button drücken und mit Slider zoomen und nochmals fokussieren)
	"Capture" drücken, das Ergebnis wird angezeigt; Total Concentration beinhaltet schon die 1:2 Verdünnung
	Durch kurzes Drücken des Slides wird er automatisch ausgeworfen
	 korrekter Fokus |  inkorrekter Fokus
Berechnen	Angegebene Zellzahl des Counters berechnen: *T:\ext01\MED5-LaborWeinhold**Formel zum Berechnen der Gesamtzellzahl *Counter:* $Zellzahl * ml * (falls\ verdünnt\ Faktor)$ (bei höherer Verdünnung z.B. 1:10, x 5; bei 1:50 x 25 etc.)
	Danach für RoboSep weiterverfahren nach SOP zum Sortieren der KM-Proben ab Punkt 3
	Neues Profil laden/ändern bzw. Troubleshooting siehe Benutzerhandbuch *T:\ext01\MED5-LaborWeinhold\ACHTUNG_LaborOrganisation\Geräte\Manuals_Geräte\countess_II_man*

RUO	File:	SOP_Counter_08_2022	O:\Med_5\MolBio_Huhn_Reichert\SOPs\SOP_Counter_08_2022.docx	Seite 1 von 1
	Version:	Last saved by huhn, stefanie	Last saved at 15.08.2022	Revision: 2

Supplementary Information Figure 4. Internal revised quality document for the cell counting using the Countess cell counter. SOP: "SOP zum Zählen der Zellzahl von Proben mittels Countess Cell Counter", last saved by Stefanie Huhn on 15.08.2022. Stefanie Huhn agreed to the usage of this document in this work.

	Cytospin	SOP zur Erstellung von Präparaten für die Zytogenetische Analyse von Myelomzellen

1. CYTOSPIN

Schritt	Protokoll	Check
1	Blaue Fertigobjektträger beschriften: Studie, Studien-Nummer / P-Nummer / Datum auf die Hälfte des Beschriftungsfelds	
2	Aliquotieren der entsprechenden Menge Positivfraktion zum Erstellen der entsprechenden Anzahl an Spots mit jeweiligen Zellen pro Spot: **Studien-SOP für die Anzahl der Zellen pro Spot**	
3	zentrifugiere diese 5min bei 1.500**rcf/g** in einem 2ml Eppi.	
4	Überstand abnehmen	
5	Zellpellet in 100µL/Spot kaltem 1xPBS, ohne EDTA lösen	
6	Filter der Doppelfunnels mit 1xPBS, ohne EDTA anfeuchten	
7	Blaue Fertigobjektträger in die Doppelfunnels einspannen und in die Zentrifuge setzen	
8	100µl Zellsuspension pro Funnel-Kammer an die Trennwand pipettieren (2x pro Doppelfunnel)	
9	6min bei 600rpm (Beschleunigung high) zentrifugieren	
10	Objektträger entnehmen und 5-10min im Abzug trocknen lassen	
11	10 Minuten fixieren (**Me**thanol-Eisessig Lösung **4:1** [90ml 100% **Me**thanol + 30ml Eisessig], ABZUG!) **Me**thanol-Eisessig Lösung **4:1** wiederverwendbar, bei -20°C lagern	
12	Objektträger entnehmen und trocknen lassen (ABZUG!)	
13	Qualität der Spots mikroskopisch prüfen	
14	Objektträger bei -20°C für eine Nacht/übers Wochenende lagern	
15	Übergabe an die Humangenetik: 1. Slides 2. Überweisungsschein mit P-Nummer 3. Laufzettel der Studie mit P-Nummer und Reinheit (so bekannt)	

RUO	File:	SOP_CYTOSPIN_06_2021	O:\Med_5\MolBio_Huhn_Reichert\SOPs\SOP_CYTOSPIN_06_2021.docx	Seite 1 von 1
	Version:	Last saved by huhn, stefanie	Last saved at 15.08.2022	Revision: 3

Supplementary Information Figure 5. Internal revised quality document for the preparation of cytospins for the cytogenetic analysis if myeloma cells. SOP: "SOP zur Erstellung von Präperaten für die Zytogenetische Analyse von Myelomzellen", last saved by Stefanie Huhn at 15.08.2022. Stefanie Huhn agreed to the usage of this document in this work.

Amyloidose | SOP Produkte und Materialaufteilung

Achtung!
Ausdruck könnte veraltet sein.
Aktuell gültige Versionen sind laminiert.

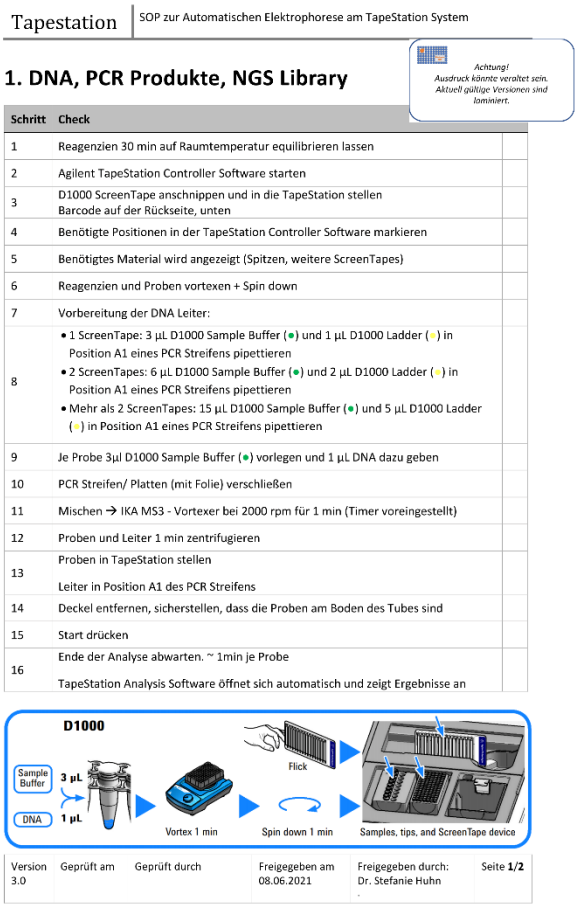
1. Produkte - Amyloidose

Schritt	Menge	Protokoll	Check
FICOL	**80ml**	**Ficoll Protokoll dann weiter nach Sort SOP**	
KM Trocken-Pellet	500µl nach Ery-Lyse	Erstellung siehe Ficoll Protokoll	
Sortierung	**80ml**		
Positiv Fraktion			
1. Zytogenetik	• Minimum 60.000 Zellen (10.000/Spot) • Optimum 150.000 Zellen (25.000/Spot)	siehe Cytospin Protokoll • 6 Spots auf 3 Slides +6 Spots auf 3 Slides wenn GTC/Lebend/Trocken vollständig sind	
2. CD138 + GTC-Pellets (RLT+βME[§])	• Minimum 1 Aliquot je $2x10^5$ • Optimum 2Aliquots > $2x10^7$ *wenn möglich gleichmäßig zw. GTC und Trockenpellet aufteilen bei <$4x10^5$ nach FISH alles in 1x GTC* *Es wird nichts verworfen!*	• Aliquotiere • Zellsuspension zentrifugieren (30sek 13.300 rpm) • Überstand abnehmen • Zellen mit GTC re-suspendiere[1,2] • bei -80°C Lagern. >> RNA Extrahieren	
3. CD 138 + Trocken Pellet	• Minimum 1Aliquot je $2x10^5$ • Optimum 2Aliquots > $2x10^7$	• Aliquotiere • Zellsuspension zentrifugieren (30sek 13.300 rpm) • Überstand abnehmen • bei -20°C Lagern. >> DNA Extrahieren	
Negativ Fraktion			
4. CD138 - GTC-Pellets (RLT+βME[§])	• Minimum 1Aliquot je $2x10^5$ • Optimum 2Aliquots > $2x10^7$	• Aliquotiere • Zellsuspension zentrifugieren (30sek 13.300 rpm) • Überstand abnehmen • Zellen mit GTC re-suspendiere[1,2] • bei -80°C Lagern. >> RNA Extrahieren	
5. CD 138 - Trocken Pellet	• Minimum 1Aliquot je 2x105 • Optimum 2Aliquots > 2x107 *Rest nach TP und GTC verwerfen*	• Aliquotiere • Zellsuspension zentrifugieren (30sek 13.300 rpm) • Überstand abnehmen • bei -20°C Lagern. >> DNA Extrahieren	

1. <$1x10^7$ Zellen in Aliquot = 350µl GTC Puffer;
2. ab $1x10^7$ Zellen 700µl GTC Puffer

Version 3.0	Geprüft am	Geprüft durch	Freigegeben am 07.12.2020	Freigegeben durch: Dr. Stefanie Huhn .pdf zum Nachdruck abgelegt	Seite **1/1**

Supplementary Information Figure 6. Internal revised quality document for the material division of amyloidosis samples. SOP: "Amyloidose – SOP Produkte und Materialaufteilung", approved by Stefanie Huhn 07.12.2020. Stefanie Huhn agreed to the usage of this document in this work.



Tapestation | SOP zur Automatischen Elektrophorese am TapeStation System

1. DNA, PCR Produkte, NGS Library

Achtung! Ausdruck könnte veraltet sein. Aktuell gültige Versionen sind laminiert.

Schritt	Check	
1	Reagenzien 30 min auf Raumtemperatur equilibrieren lassen	
2	Agilent TapeStation Controller Software starten	
3	D1000 ScreenTape anschnippen und in die TapeStation stellen Barcode auf der Rückseite, unten	
4	Benötigte Positionen in der TapeStation Controller Software markieren	
5	Benötigtes Material wird angezeigt (Spitzen, weitere ScreenTapes)	
6	Reagenzien und Proben vortexen + Spin down	
7	Vorbereitung der DNA Leiter:	
8	• 1 ScreenTape: 3 µL D1000 Sample Buffer (●) und 1 µL D1000 Ladder (●) in Position A1 eines PCR Streifens pipettieren • 2 ScreenTapes: 6 µL D1000 Sample Buffer (●) und 2 µL D1000 Ladder (●) in Position A1 eines PCR Streifens pipettieren • Mehr als 2 ScreenTapes: 15 µL D1000 Sample Buffer (●) und 5 µL D1000 Ladder (●) in Position A1 eines PCR Streifens pipettieren	
9	Je Probe 3µl D1000 Sample Buffer (●) vorlegen und 1 µL DNA dazu geben	
10	PCR Streifen/ Platten (mit Folie) verschließen	
11	Mischen → IKA MS3 - Vortexer bei 2000 rpm für 1 min (Timer voreingestellt)	
12	Proben und Leiter 1 min zentrifugieren	
13	Proben in TapeStation stellen Leiter in Position A1 des PCR Streifens	
14	Deckel entfernen, sicherstellen, dass die Proben am Boden des Tubes sind	
15	Start drücken	
16	Ende der Analyse abwarten. ~ 1min je Probe TapeStation Analysis Software öffnet sich automatisch und zeigt Ergebnisse an	

Version 3.0	Geprüft am	Geprüft durch	Freigegeben am 08.06.2021	Freigegeben durch: Dr. Stefanie Huhn	Seite **1/2**

Tapestation | SOP zur Automatischen Elektrophorese am TapeStation System

2. genomische DNA

Schritt	Check	
1	Reagenzien 30 min auf Raumtemperatur equilibrieren lassen	
2	Agilent TapeStation Controller Software starten	
3	Genomic ScreenTape anschnippen und in die TapeStation stellen Barcode auf der Rückseite, unten	
4	Benötigte Positionen in der TapeStation Controller Software markieren	
5	Benötigtes Material wird angezeigt (Spitzen, weitere ScreenTapes)	
6	Reagenzien und Proben vortexen + Spin down	
7	Vorbereitung der DNA Leiter:	
8	• 1 -2 ScreenTapes: 10 µL Genomic DNA Sample Buffer (●) und 1 µL Genomic DNA Ladder (●) in Position A1 eines PCR Streifens pipettieren. • Für mehr als 2 ScreenTapes: 20 µL Genomic DNA Sample Buffer (●) und 2 µL Genomic DNA Ladder (●)in Position A1 eines PCR Streifens pipettieren.	
9	Je Probe 10 µL Genomic DNA Sample Buffer (●) vorlegen und 1 µL DNA dazu geben	
10	PCR Streifen/ Platten (mit Folie) verschließen	
11	Durch Anschnippen mischen (nicht Vortexen → Scherkräfte!)	
12	Proben und Leiter 1 min zentrifugieren	
13	Proben in TapeStation stellen Leiter in Position A1 des PCR Streifens	
14	Deckel entfernen, sicherstellen, dass die Proben am Boden des Tubes sind	
15	Start drücken	
16	Ende der Analyse abwarten. ~ 1min je Probe TapeStation Analysis Software öffnet sich automatisch und zeigt Ergebnisse an	

Version 1.0	Erstellt	Geprüft	Freigegeben	Seite **2/2**

sop_tapestation_06_2021

Supplementary Information Figure 7. Internal revised quality document for concentration determination and gel electrophoresis using the tapestation system. SOP zur Automatischen Elektrophorese am TapeStation System″, step: ″DNA, PCR Produkte, NGS Library″, approved 08.06.2021 by Stefanie Huhn. Stefanie Huhn agreed to the usage of this document in this work.

Appendix Results

Supplementary Information Table 2. Sample composition and light chain composition of the multiple myeloma cohort. Sequence_A = sequence with the highest percentage >1 %, sequence_B = sequence with the second highest percentage >1 %.

sample	sequence_A [%]	IGLV sequence_A	IGLJ sequence_A	IGLC sequence_A	sequence sections	sequence_B [%]	IGLV sequence_B	IGLJ sequence_B	IGLC sequence_B	sequence sections
MM101	74.6	**NA**	**IGLJ3**	**IGLC3**	2	25.3	**IGLV1-44**	**IGLJ3**	**IGLC3**	2
MM102	95.7	**IGLV2-14**	**IGLJ1**	**IGLC1**	1	4.0	**IGLV2-14**	**IGLJ1**	**IGLC1**	1
MM103	99.7	**IGLV2-14**	**IGLJ3**	**IGLC2/3**	1					
MM104	99.7	**IGLV1-40**	**IGLJ2**	**IGLC2**	2					
MM105	98.1	**IGLV3-25**	**IGLJ2**	**IGLC2**	2					
MM106	99.4	**IGLV3-21**	**IGLJ2**	**IGLC2**	1					
MM107	99.8	**IGLV2-23**	**IGLJ2/3**	**IGLC2**	2					
MM108	99.7	**IGLV3-21**	**IGLJ2**	**IGLC2**	1					
MM109	69.6	**IGLV1-44**	**IGLJ2**	**IGLC2/3**	1	26.1	**IGLV1-44**	**IGLJ2**	**IGLC2**	2
MM110	95.2	**IGLV1-51**	**IGLJ7**	**IGLC7**	1					
MM111	100.0	**IGLV3-21**	**IGLJ3**	**IGLC3**	1					
MM112	99.8	**IGLV6-57**	**IGLJ3**	**IGLC3**	1					
MM113	99.8	**IGLV7-46**	**IGLJ2**	**IGLC2**	2					
MM114	94.1	**IGLV2-23**	**IGLJ3**	**IGLC3**	1	4.5	**IGLV2-23**	**IGLJ3**	**IGLC3**	2
MM115	97.3	**IGLV1-44**	**IGLJ3**	**IGLC2/3**	2	2.5	**IGLV1-44**	**IGLJ3**	**IGLC2**	2
MM116	65.0	**IGLV1-44**	**IGLJ3**	**IGLC3**	2	34.8	**IGLV1-44**	**IGLJ3**	**IGLC3**	3
MM117	98.8	**IGLV2-23**	**IGLJ2**	**IGLC2**	1					
MM118	74.4	**IGLV1-44**	**IGLJ3**	**IGLC2**	2	25.4	**IGLV1-44**	**IGLJ3**	**IGLC2**	2
MM119	98.7	**IGLV3-21**	**IGLJ2**	**IGLC2**	1					
MM120	99.7	**IGLV3-1**	**IGLJ1**	**IGLC1**	6					
MM121	52.5	**IGLV1-44**	**IGLJ2**	**IGLC2/3**	2	47.4	**IGLV1-44/47**	**IGLJ2**	**IGLC2/3**	3
MM122	98.1	**IGLV3-21**	**IGLJ2**	**IGLC2**	1	1.2	**IGLV3-21**	**IGLJ2**	**IGLC2**	2
MM123	99.0	**IGLV3-21**	**IGLJ1**	**IGLC1**	1					
MM124	75.2	**NA**	**IGLJ3**	**IGLC2/3**	2	21.8	**IGLV1-44**	**IGLJ3**	**IGLC2**	2
MM125	99.9	**IGLV2-23**	**IGLJ3**	**IGLC3**	1					
MM126	97.7	**IGLV2-14**	**IGLJ2**	**IGLC2**	1					
MM127	99.7	**IGLV2-11**	**IGLJ2**	**IGLC2/3**	1					
MM128	99.3	**IGLV3-19**	**IGLJ2**	**IGLC2**	1					
MM129	99.3	**IGLV2-14**	**IGLJ2**	**IGLC2**	1					
MM130	99.6	**IGLV2-14**	**IGLJ2**	**IGLC2**	1					
MM131	100.0	**IGLV1-51**	**IGLJ2**	**IGLC2/3**	1					
MM132	99.9	**IGLV2-23**	**IGLJ2/3**	**IGLC2**	2					
MM133	93.7	**IGLV2-23**	**IGLJ3**	**IGLC3**	1	6.0	**NA**	**NA**	**NA**	3
MM134	98.9	**IGLV3-21**	**IGLJ2/3**	**IGLC2**	1					
MM135	97.4	**IGLV2-23**	**IGLJ2**	**IGLC2**	1					
MM136	99.4	**IGLV2-14**	**IGLJ2**	**IGLC2**	1					
MM137	81.7	**IGLV1-44**	**IGLJ2/3**	**IGLC2/3**	1	18.0	**IGLV1-44**	**IGLJ2/3**	**IGLC2**	2
MM138	99.9	**IGLV3-21**	**IGLJ2**	**IGLC2**	2					
MM139	98.4	**IGLV3-1**	**IGLJ2**	**IGLC2**	1					
MM140	99.9	**IGLV1-40**	**IGLJ2**	**IGLC2**	1					
MM141	74.1	**IGLV1-44**	**IGLJ1**	**IGLC1**	2	25.7	**IGLV1-44**	**IGLJ1**	**IGLC1**	2

sample	sequence_A [%]	**IGLV sequence_A**	**IGLJ sequence_A**	**IGLC sequence_A**	sequence sections	sequence_B [%]	**IGLV sequence_B**	**IGLJ sequence_B**	**IGLC sequence_B**	sequence sections
MM142	99.4	**IGLV3-1**	**IGLJ2/3**	**IGLC2**	1					
MM143	98.8	**IGLV3-21**	**IGLJ2**	**IGLC2**	1					
MM144	99.9	**IGLV3-21**	**IGLJ2**	**IGLC2**	1					
MM145	99.7	**IGLV3-19**	**IGLJ2**	**IGLC2**	1					
MM146	99.6	**IGLV3-1**	**IGLJ1**	**IGLC1**	1					
MM147	98.5	**IGLV2-14**	**IGLJ2**	**IGLC2**	1					
MM148	73.6	**IGLV1-44**	**IGLJ2/3**	**IGLC3**	2	26.0	**IGLV1-44**	**IGLJ2/3**	**IGLC2**	2
MM149	95.1	**IGLV2-23**	**IGLJ2/3**	**IGLC3**	1	3.6	**NA**	**NA**	**NA**	3
MM150	99.9	**IGLV3-21**	**IGLJ3**	**IGLC2/3**	1					
MM151	97.5	**IGLV1-47**	**IGLJ3**	**IGLC3**	2					
MM152	71.1	**IGLV1-44**	**IGLJ2**	**IGLC2**	2	28.7	**IGLV1-44**	**IGLJ2**	**IGLC2**	2
MM153	98.8	**IGLV2-14**	**IGLJ3**	**IGLC2/3**	1					

Supplementary Information Table 3. Sequence analysis from AL amyloidosis patients who were also diagnosed with multiple myeloma. The column "X AA [n]" defines how many amino acids were not determined unambiguously. Org. inv. = organ involvement, H = AL amyloidosis patients with dominant heart involvement, AA = amino acid, Mut. = mutations, NA = not available.

Patient	Org. inv.	IGLV	IGLJ	IGLC	X AA [n]	Mut. IGLV [n]	Mut. IGLJ [n]	Mut. IGLC [n]
FOR123	H	3-1	2	2	0	12	2	0
FOR159	H	2-14	3	3	2	5	1	0
FOR147	H	1-44	3	2	0	9	1	0
FOR188	HK	6-57	3	3	0	11	0	0
FOR229	HK	8-61	3	3	0	14	1	0

Appendix Results: IGLV2-14

```
MM103_NGS         ATGGCCTGGGCTCTGCTGCTCCTCACCCTCCTCACTCAGGGCACAGGGTCCTGGGCCCAG
MM103_Sanger      ------------------------------------------------------------AG
                                                                              **

MM103_NGS         TCTGCCCTGACTCAGCCTGCCTCCGTGTCTGGGTCTCCTGGACAGTCGATCACCATCTCC
MM103_Sanger      TCTGCCCTGACTCAGCCTGCCTCCGTGTCTGGGTCTCCTGGACAGTCGATCACCATCTCC
                  ************************************************************

MM103_NGS         TGCAATGGAACCAGCAGTGACATTGGTGGTTATAACTATGTCTCCTGGTACCAACAACAC
MM103_Sanger      TGCAATGGAACCAGCAGTGACATTGGTGGTTATAACTATGTCTCCTGGTACCAACAACAC
                  ************************************************************

MM103_NGS         CCGGACACAGCCCCCAAACTCATTATTTATGAGGTCACTAATCGGCCCTCAGGGGTTTTT
MM103_Sanger      CCGGACACAGCCCCCAAACTCATTATTTATGAGGTCACTAATCGGCCCTCAGGGGTTTTT
                  ************************************************************

MM103_NGS         ACTCGCTTCTCTGGCTCCAAGTCTGGCAACACGGCCTCCCTGACCATCTCTGGCCTCCAG
MM103_Sanger      ACTCGCTTCTCTGGCTCCAAGTCTGGCAACACGGCCTCCCTGACCATCTCTGGCCTCCAG
                  ************************************************************

MM103_NGS         GGTGAGGACGAGGCTGATTATTACTGCGCCTCATATACCAGAAGTGACACCTGGGTGTTC
MM103_Sanger      GGTGAGGACGAGGCTGATTATTACTGCGCCTCATATACCAGAAGTGACACCTGGGTGTTC
                  ************************************************************

MM103_NGS         GGCGGAGGGACCAAACTGACCGTCCTAGGTCAGCCCAAGGCTGCCCCCTCGGTCACTCTG
MM103_Sanger      GGCGGAGGGACCAAACTGACCGTCCTAGGTCAGCCCAAGGCTGCCCCCTCGGTCACTCTG
                  ************************************************************

MM103_NGS         TTCCCGCCCTCCTCTGAGGAGCTTCAAGCCAACAAGGCCACACTGGTGTGTCTCATAAGT
MM103_Sanger      TTCCCGCCCTCCTCTGAGGAGCTTCAAGCCAACAAGGCCACACTGGTGTGTCTCATAAGT
                  ************************************************************

MM103_NGS         GACTTCTACCCGGGAGCCGTGACAGTGGCCTGGAAGGCAGATAGCAGCCCCGTCAAGGCG
MM103_Sanger      GACTTCTACCCGGGAGCCGTGACAGTGGCCTGGAAGGCAGATAGCAGCCCCGTCAAGGCG
                  ************************************************************

MM103_NGS         GGAGTGGAGACCACCACACCCTCCAAACAAAGCAACAACAAGTACGCGGCCAGCAGCTAC
MM103_Sanger      GGAGTGGAGACCACCACACCCTCCAAACAAAGCAACAACAAGTACGCGGCCAGCAGCTAC
                  ************************************************************

MM103_NGS         CTGAGCCTGACGCCTGAGCAGTGGAAGT------------------------------
MM103_Sanger      CTGAGCCTGACGCCTGAGCAGTGGAAGTCCCACAAAAGCTACAGCTGCCAGGTCACGCAT
                  ****************************

MM103_NGS         -----------------
MM103_Sanger      GAAGGGAGCACCGTGGA
```

Supplementary Information Figure 8. Sequence comparison between MM103 IGLV2-14 assigned light chain sequences generated through a next-generation sequencing and Sanger sequencing approach. For the sequence generated by the Sanger sequencing approach, only dominant signals were used. * = nucleotide position consistent, MM = multiple myeloma patient, NGS = light chain sequence generated through next-generation sequencing, Sanger = light chain sequence generated through the Sanger sequencing. Sequencing was performed using the VLKL12a_Huhn and CLKL_A_rv_NB oligonucleotides.

```
MM129_NGS         ATGGCCTGGGCTCTGCTATTCCTCACCCTCCTCACTCAGGGCACAGGGTCCTGGGCCCAG
MM129_Sanger      ----------------------------------------------------------CAG
                                                                             ***

MM129_NGS         TCTGCCCTGACTCAGCCTGCCTCCGTGTCTGGGTCTCCTGGACAGTCGATCACCATCTCC
MM129_Sanger      TCTGCCCTGACTCAGCCTGCCTCCGTGTCTGGGTCTCCTGGACAGTCGATCACCATCTCC
                  ************************************************************

MM129_NGS         TGCACTGGAACCAGCAGTGACATTGGTGGTTATAACTATGTCTCCTGGTACCAACAACAC
MM129_Sanger      TGCACTGGAACCAGCAGTGACATTGGTGGTTATAACTATGTCTCCTGGTACCAACAACAC
                  ************************************************************

MM129_NGS         CCAGACAAAGCCCCCAAGCTCATCATTTATGATGTCGCTAAACGGCCCCCAGGGGTTTCT
MM129_Sanger      CCAGACAAAGCCCCCAAGCTCATCATTTATGATGTCGCTAAACGGCCCCCAGGGGTTTCT
                  ************************************************************

MM129_NGS         AGTCACTTCTCTGGCTCCAAGTCTGGCAACACGGCCTCCCTGACCATCTCTGGCCTGCAA
MM129_Sanger      AGTCACTTCTCTGGCTCCAAGTCTGGCAACACGGCCTCCCTGACCATCTCTGGCCTGCAA
                  ************************************************************

MM129_NGS         CCTGAGGACGAAGCTGATTACTTCTGCAGTTCATATTCAAGCCGCACCACTCTCGAAGTC
MM129_Sanger      CCTGAGGACGAAGCTGATTACTTCTGCAGTTCATATTCAAGCCGCACCACTCTCGAAGTC
                  ************************************************************

MM129_NGS         TTCGGCGGAGGGACCAAGCTGACCGTCCTCAGTCAGCCCAAGGCTGCCCCCTCGGTCACT
MM129_Sanger      TTCGGCGGAGGGACCAAGCTGACCGTCCTCAGTCAGCCCAAGGCTGCCCCCTCGGTCACT
                  ************************************************************

MM129_NGS         CTGTTCCCGCCCTCCTCTGAGGAGCTTCAAGCCAACAAGGCCACACTGGTGTGTCTCATA
MM129_Sanger      CTGTTCCCGCCCTCCTCTGAGGAGCTTCAAGCCAACAAGGCCACACTGGTGTGTCTCATA
                  ************************************************************

MM129_NGS         AGTGACTTCTACCCGGGAGCCGTGACAGTGGCCTGGAAGGCAGATAGCAGCCCCGTCAAG
MM129_Sanger      AGTGACTTCTACCCGGGAGCCGTGACAGTGGCCTGGAAGGCAGATAGCAGCCCCGTCAAG
                  ************************************************************

MM129_NGS         GCGGGAGTGGAGACCACCACACCCTCCAAACAAAGCAACAACAAGTACGCGGCCAGCAGC
MM129_Sanger      GCGGGAGTGGAGACCACCACACCCTCCAAACAAAGCAACAACAAGTACGCGGCCAGCAGC
                  ************************************************************

MM129_NGS         TATCTGAGCCTGACGCCTGAG-------------------------------------
MM129_Sanger      TATCTGAGCCTGACGCCTGAGCAGTGGAAGTCCCACAGAAGCTACAGCTGCCAGGTCACG
                  ********************

MM129_NGS         -------------------
MM129_Sanger      CATGAAGGGAGCACCGTGG
```

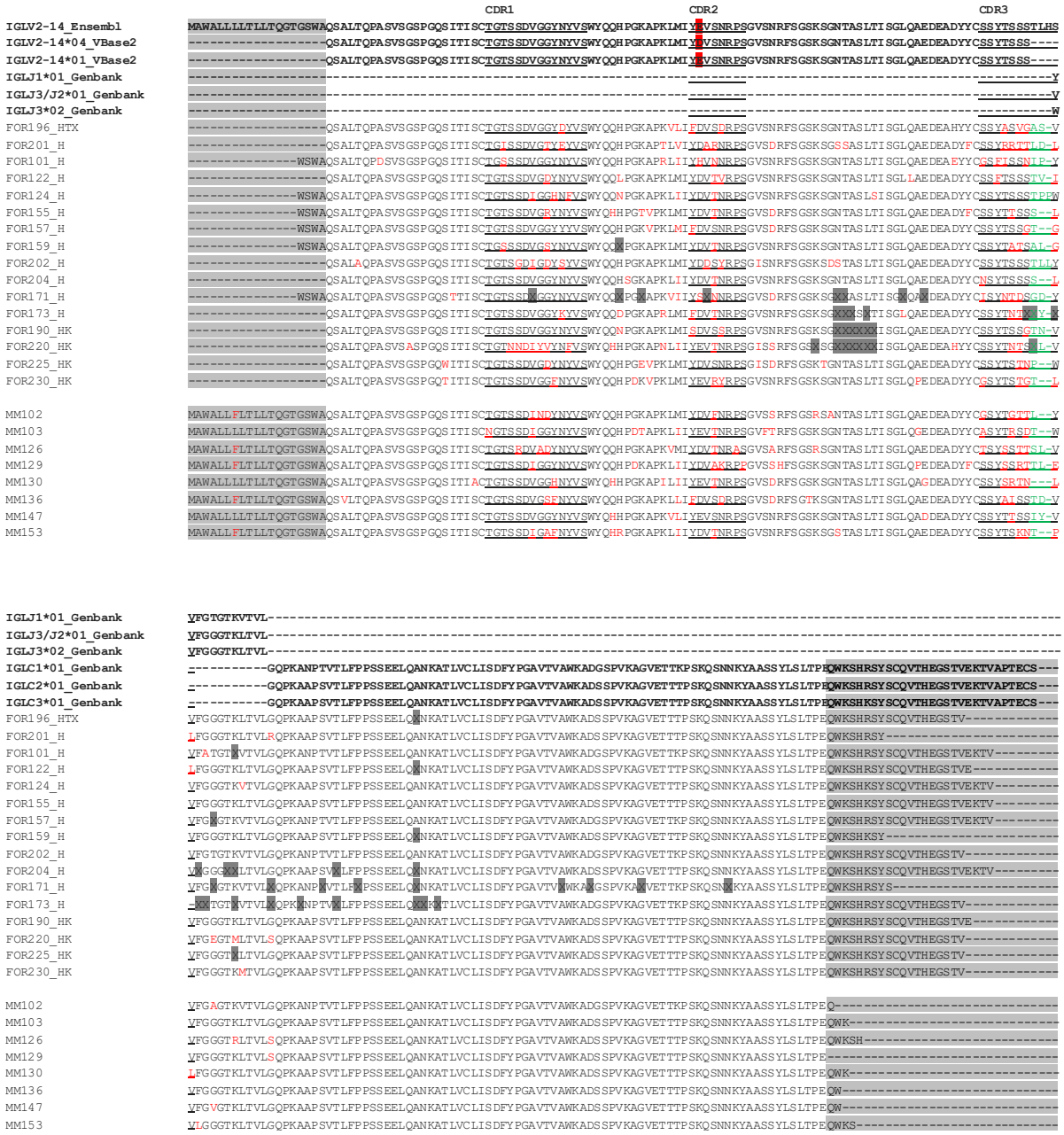
Supplementary Information Figure 9. Sequence comparison between MM129 IGLV2-14 assigned light chain sequences generated through a next-generation sequencing and Sanger sequencing approach. For the sequence generated by the Sanger sequencing approach only dominant signals were used. * = nucleotide position consistent, MM = multiple myeloma patient, NGS = light chain sequence generated through next-generation sequencing, Sanger = light chain sequence generated through Sanger sequencing. Sequencing was performed using the VLKL12a_Huhn and CLKL_A_rv_NB oligonucleotides.

```
MM130_NGS         ATGGCCTGGGCTCTGCTGCTCCTCACCCTCCTCACTCAGGGCACAGGGTCCTGGGCCCAG
MM130_Sanger      ----------------------------------------------------------CAG
                                                                             ***

MM130_NGS         TCTGCCCTGACTCAGCCTGCCTCCGTGTCTGGGTCTCCTGGACAGTCGATCACCATCGCC
MM130_Sanger      TCTGCCCTGACTCAGCCTGCCTCCGTGTCTGGGTCTCCTGGACAGTCGATCACCATCGCC
                  ************************************************************

MM130_NGS         TGCACTGGAACCAGCAGTGACGTTGGTGGTCATAACTATGTCTCCTGGTACCAACATCAC
MM130_Sanger      TGCACTGGAACCAGCAGTGACGTTGGTGGTCATAACTATGTCTCCTGGTACCAACATCAC
                  ************************************************************

MM130_NGS         CCAGGCAAAGCCCCCATACTCATAATTTATGAGGTCACTAATCGGCCCTCAGGGGTTTCT
MM130_Sanger      CCAGGCAAAGCCCCCATACTCATAATTTATGAGGTCACTAATCGGCCCTCAGGGGTTTCT
                  ************************************************************

MM130_NGS         GATCGCTTCTCTGGCTCCAAGTCTGGCAACACGGCCTCCCTGACCATCTCTGGGCTCCAG
MM130_Sanger      GATCGCTTCTCTGGCTCCAAGTCTGGCAACACGGCCTCCCTGACCATCTCTGGGCTCCAG
                  ************************************************************

MM130_NGS         GCTGGCGACGAGGCTGATTATTACTGCAGCTCATACAGCAGAACCAATTTGCTATTCGGC
MM130_Sanger      GCTGGCGACGAGGCTGATTATTACTGCAGCTCATACAGCAGAACCAATTTGCTATTCGGC
                  ************************************************************

MM130_NGS         GGAGGGACCAAATTGACCGTCCTAGGTCAGCCCAAGGCTGCCCCCTCGGTCACTCTGTTC
MM130_Sanger      GGAGGGACCAAATTGACCGTCCTAGGTCAGCCCAAGGCTGCCCCCTCGGTCACTCTGTTC
                  ************************************************************

MM130_NGS         CCGCCCTCCTCTGAGGAGCTTCAAGCCAACAAGGCCACACTGGTGTGTCTCATAAGTGAC
MM130_Sanger      CCGCCCTCCTCTGAGGAGCTTCAAGCCAACAAGGCCACACTGGTGTGTCTCATAAGTGAC
                  ************************************************************

MM130_NGS         TTCTACCCGGGAGCCGTGACAGTGGCCTGGAAGGCAGATAGCAGCCCCGTCAAGGCGGGA
MM130_Sanger      TTCTACCCGGGAGCCGTGACAGTGGCCTGGAAGGCAGATAGCAGCCCCGTCAAGGCGGGA
                  ************************************************************

MM130_NGS         GTGGAGACCACCACACCCTCCAAACAAAGCAACAACAAGTACGCGGCCAGCAGCTATCTG
MM130_Sanger      GTGGAGACCACCACACCCTCCAAACAAAGCAACAACAAGTACGCGGCCAGCAGCTATCTG
                  ************************************************************

MM130_NGS         AGCCTGACGCCTGAGCAGTGGAAGTC--------------------------------
MM130_Sanger      AGCCTGACGCCTGAGCAGTGGAAGTCCCACAGAAGCTACAGCTGCCAGGTCACGCATGAA
                  *************************

MM130_NGS         ------------
MM130_Sanger      GGGAGCACCGTGG
```

Supplementary Information Figure 10. Sequence comparison between MM130 IGLV2-14 assigned light chain sequences generated through a next-generation sequencing and Sanger sequencing approach. For the sequence generated by the Sanger sequencing approach only dominant signals were used. * = nucleotide position consistent, MM = multiple myeloma patient, NGS = light chain sequence generated through next-generation sequencing, Sanger = light chain sequence generated through Sanger sequencing. Sequencing was performed using the VLKL12a_Huhn and CLKL_A_rv_NB oligonucleotides.

```
MM136_NGS         ATGGCCTGGGCTCTGCTATTCCTCACCCTCCTCACTCAGGGCACAGGGTCCTGGGCCCAG
MM136_Sanger      ----------------------------------------------------------AG
                                                                            **

MM136_NGS         TCTGTCCTGACTCAACCTGCCTCCGTGTCTGGGTCTCCTGGACAGTCGATCACCATCTCC
MM136_Sanger      TCTGTCCTGACTCAACCTGCCTCCGTGTCTGGGTCTCCTGGACAGTCGATCACCATCTCC
                  ************************************************************

MM136_NGS         TGCACTGGAACCAGCAGTGACGTTGGTTCTTTTAATTATGTCTCCTGGTACCAACAACAC
MM136_Sanger      TGCACTGGAACCAGCAGTGACGTTGGTTCTTTTAATTATGTCTCCTGGTACCAACAACAC
                  ************************************************************

MM136_NGS         CCAGGCAAAGCCCCCAAACTCTTAATTTTTGATGTCAGTGATAGGCCCTCAGGGGTTTCT
MM136_Sanger      CCAGGCAAAGCCCCCAAACTCTTAATTTTTGATGTCAGTGATAGGCCCTCAGGGGTTTCT
                  ************************************************************

MM136_NGS         GATCGCTTCTCTGGCACCAAGTCTGGCAACACGGCCTCCCTGACCATCTCTGGGCTCCAG
MM136_Sanger      GATCGCTTCTCTGGCACCAAGTCTGGCAACACGGCCTCCCTGACCATCTCTGGGCTCCAG
                  ************************************************************

MM136_NGS         GCTGAGGACGAGGCTGATTATTACTGCAGCTCATATGCAATCAGCAGTACGGATGTGGTT
MM136_Sanger      GCTGAGGACGAGGCTGATTATTACTGCAGCTCATATGCAATCAGCAGTACGGATGTGGTT
                  ************************************************************

MM136_NGS         TTCGGCGGAGGGACAAAGCTGACCGTCCTAGGTCAGCCCAAGGCTGCCCCCTCGGTCACT
MM136_Sanger      TTCGGCGGAGGGACAAAGCTGACCGTCCTAGGTCAGCCCAAGGCTGCCCCCTCGGTCACT
                  ************************************************************

MM136_NGS         CTGTTCCCGCCCTCCTCTGAGGAGCTTCAAGCCAACAAGGCCACACTGGTGTGTCTCATA
MM136_Sanger      CTGTTCCCGCCCTCCTCTGAGGAGCTTCAAGCCAACAAGGCCACACTGGTGTGTCTCATA
                  ************************************************************

MM136_NGS         AGTGACTTCTACCCGGGAGCCGTGACAGTGGCCTGGAAGGCAGATAGCAGCCCCGTCAAG
MM136_Sanger      AGTGACTTCTACCCGGGAGCCGTGACAGTGGCCTGGAAGGCAGATAGCAGCCCCGTCAAG
                  ************************************************************

MM136_NGS         GCGGGAGTGGAGACCACCACACCCTCCAAACAAAGCAACAACAAGTACGCGGCCAGCAGC
MM136_Sanger      GCGGGAGTGGAGACCACCACACCCTCCAAACAAAGCAACAACAAGTACGCGGCCAGCAGC
                  ************************************************************

MM136_NGS         TATCTGAGCCTGACGCCTGAGCAGTGG------------------------------
MM136_Sanger      TATCTGAGCCTGACGCCTGAGCAGTGGAAGTCCCACAGAAGCTACAGCTGCCAGGTCACG
                  ***************************

MM136_NGS         --------------------
MM136_Sanger      CATGAAGGGAGCACCGTGGA
```

Supplementary Information Figure 11. Sequence comparison between MM136 IGLV2-14 assigned light chain sequences generated through a next-generation sequencing and Sanger sequencing approach. For the sequence generated by the Sanger sequencing approach only dominant signals were used. * = nucleotide position consistent, MM = multiple myeloma patient, NGS = light chain sequence generated through next-generation sequencing, Sanger = light chain sequence generated through Sanger sequencing. Sequencing was performed using the VLKL12a_Huhn and CLKL_A_rv_NB oligonucleotides.



Supplementary Information Figure 12. Sequence alignment of IGLV2-14 assigned AL amyloidosis and multiple myeloma amino acid light chain sequences. N- and C-terminal sequence regions that were excluded in the analysis are highlighted in grey. Bold = reference sequences, underlined = CDR regions, red highlight = discrepancy between the VBase2 and Ensembl IGLV2-14 reference, red letter = mutation, X and grey highlight = not unambiguously determined amino acid, green letter = linker region, MM = multiple myeloma patient, _H = AL amyloidosis patient with dominant heart involvement, _HK = AL amyloidosis patient with dominant heart and kidney involvement, _HTX = AL amyloidosis patient who received a heart transplant.

Supplementary Information Figure 13. Sequence alignment of IGLV2-14 assigned AL amyloidosis and multiple myeloma cDNA light chain sequences. green letter = linker region, MM = multiple myeloma patient, _H = AL amyloidosis patient with dominant heart involvement, _HK = AL amyloidosis patient. _HTX = AL amyloidosis patient who received a heart transplant. Nucleotide signal overlaps in the Sanger sequencing were specified according to the IUPAC code.

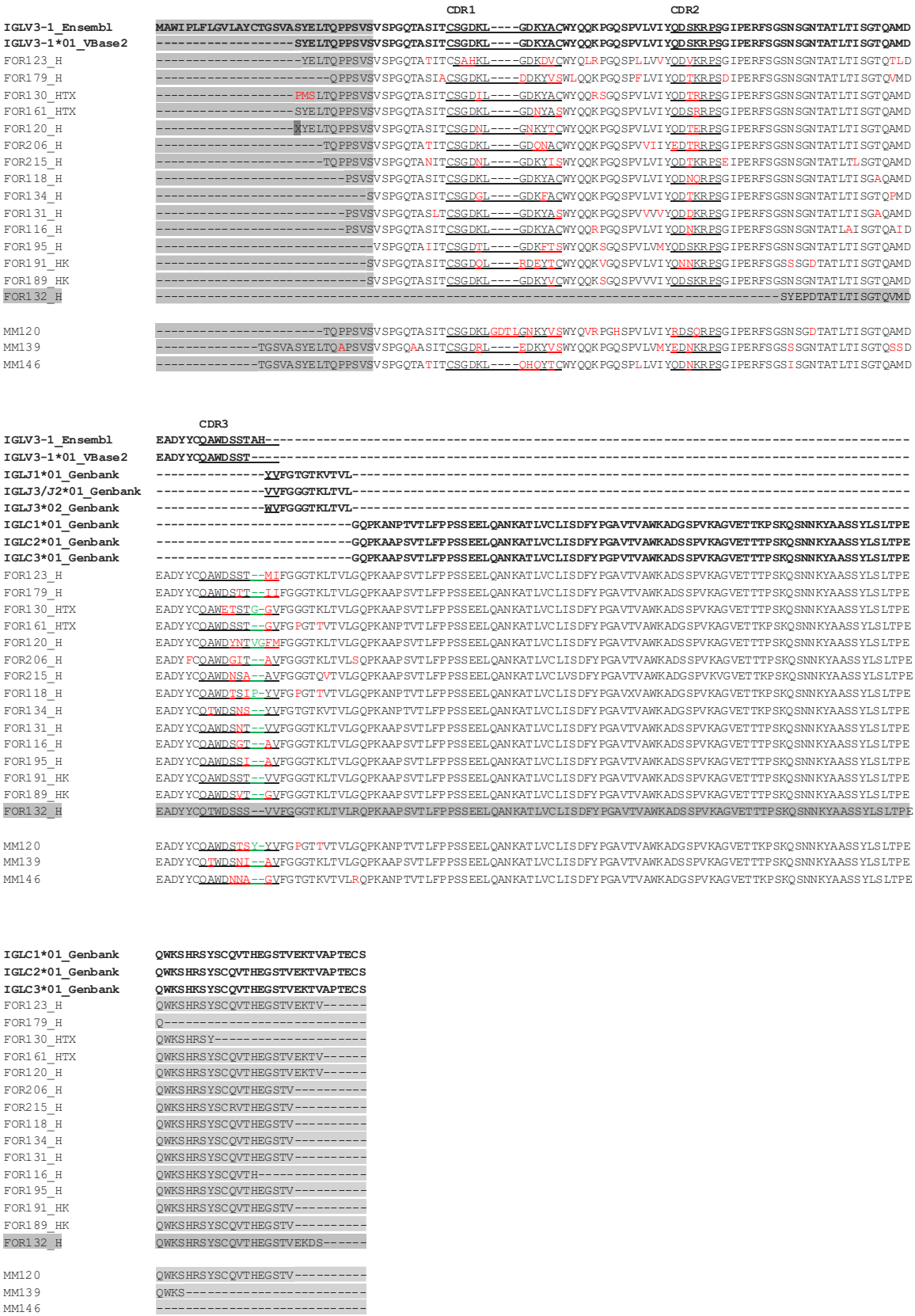
Supplementary Information Table 4. Overview of the amino acid percentage of IGLV2-14 assigned AL amyloidosis and multiple myeloma light chain sequences. MM = multiple myeloma patient, _H = AL amyloidosis patient with dominant heart involvement, _HK = AL amyloidosis patient with dominant heart and kidney involvement, _HTX = AL amyloidosis patient who received a heart transplant.

	Amino acid [%]																			
Sample	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
FOR101_H	8.0	1.6	5.3	2.7	1.6	4.8	3.7	8.5	1.1	4.3	5.9	5.3	0.0	2.7	7.4	15.4	8.0	1.1	5.3	7.4
FOR122_H	8.5	1.1	3.2	3.7	1.6	4.3	3.2	8.0	0.0	3.2	8.0	5.3	0.5	2.7	6.9	17.0	9.6	1.1	4.8	7.4
FOR124_H	8.5	1.1	4.2	3.2	1.6	4.8	3.2	8.5	0.5	3.7	5.8	5.3	0.0	2.6	7.9	17.5	9.0	1.6	4.2	6.9
FOR155_H	8.0	1.6	3.2	3.7	1.6	4.3	3.2	8.0	1.1	2.7	7.0	4.8	0.5	2.7	7.0	17.1	10.2	1.1	4.8	7.5
FOR157_H	7.5	1.1	3.2	3.7	1.6	4.8	3.2	9.6	0.5	2.7	5.9	5.9	0.5	2.7	7.0	16.0	9.6	1.1	5.3	8.0
FOR159_H	9.6	1.1	3.7	3.2	1.6	4.8	3.2	8.5	0.5	2.7	6.9	5.3	0.5	2.1	6.9	17.0	9.0	1.1	5.3	6.9
FOR196_HTX	9.6	1.1	2.7	3.7	1.6	4.8	3.2	9.0	1.1	2.7	6.4	5.3	0.0	2.7	6.9	17.0	8.0	1.1	4.8	8.5
FOR202_H	8.5	1.1	2.6	4.8	1.6	4.8	3.2	7.9	0.5	3.7	6.9	5.8	0.5	2.1	6.9	16.9	9.0	1.1	6.3	5.8
FOR204_H	8.6	1.1	4.3	3.2	1.6	4.8	3.2	8.6	0.5	3.2	7.0	5.3	0.0	2.1	6.4	17.6	9.1	1.1	5.3	7.0
FOR201_H	9.0	3.2	2.1	4.3	1.6	4.8	3.7	7.4	0.5	3.2	8.0	4.8	0.0	2.7	6.9	16.5	9.0	1.1	4.8	6.4
FOR190_HK	8.5	1.1	4.3	3.2	1.6	4.8	3.2	9.0	0.0	2.7	6.4	5.3	0.5	2.1	6.9	18.1	9.0	1.1	4.8	7.4
FOR220_HK	9.0	1.1	5.9	2.1	1.6	4.3	4.3	5.9	1.6	4.3	6.9	4.3	0.5	2.7	6.9	16.0	9.6	1.1	5.3	6.9
FOR225_HK	8.0	1.1	3.7	4.3	1.6	4.8	3.7	8.0	0.5	3.2	6.4	4.8	0.5	2.1	7.4	15.4	10.1	2.1	5.3	6.9
FOR230_HK	7.5	1.6	3.2	3.2	1.6	4.8	3.7	9.1	0.5	2.7	6.4	5.3	1.1	2.7	7.5	15.0	10.2	1.1	5.3	7.5
MM102	9.1	1.6	4.3	3.7	1.6	4.8	3.2	8.0	0.5	3.2	6.4	5.3	0.5	2.7	7.0	14.4	9.6	1.1	5.9	7.0
MM103	8.6	1.6	3.7	3.7	1.6	4.8	3.7	8.6	0.5	3.7	6.4	4.8	0.0	2.7	7.0	15.0	10.2	1.6	5.3	6.4
MM126	10.1	2.7	3.2	3.7	1.6	4.8	3.2	6.9	0.5	2.7	6.4	4.3	0.5	2.1	6.4	16.5	10.1	1.1	5.3	8.0
MM129	8.5	1.1	2.7	3.7	1.6	4.8	3.7	7.4	1.1	3.7	6.9	5.9	0.0	2.7	8.0	17.0	9.0	1.1	4.8	6.4
MM130	9.1	1.6	3.8	3.2	1.6	4.3	3.2	9.1	1.6	3.8	7.5	4.8	0.0	2.2	7.0	15.6	9.1	1.1	4.8	6.5
MM136	8.5	1.1	2.7	4.8	1.6	4.8	3.2	8.0	0.5	3.2	6.9	5.3	0.0	3.2	6.9	17.0	9.0	1.1	4.3	8.0
MM147	8.5	1.1	3.7	3.2	1.6	4.3	3.2	8.0	1.1	3.2	6.4	5.3	0.0	2.1	6.9	17.0	9.0	1.1	5.9	8.5
MM153	9.1	1.6	3.7	3.2	1.6	4.3	3.2	8.0	0.5	3.7	7.0	5.9	0.0	2.1	7.5	16.6	9.6	1.1	4.8	6.4

Appendix Results: IGLV3-1

```
MM120_NGS_A          GATCCATGGCCTCCTTTGCGCTGACTCAGCCACCCTCAGTGTCCGTGTCCCCAGGACAGA
MM120_Sanger         ----------------------ACTCAGCCACCCTCAGTGTCCGTGTCCCCAGGACAGA
                                            ************************************

MM120_NGS_A          CAGCCAGCATCA--------------------------------------------------
MM120_NGS_B          -------------------------------GGAGATACATTGGGGAATAAATATG---
MM120_Sanger         CAGCCAGCATCACCTGCTCTGGAGATAAATTGGGAGATACATTGGGGAATAAATATGTTT
                     ************                    ************************

MM120_NGS_C          -----TATCAGGTGAGGCCAGGCCACTCCCCTGTATTGGTCATCTATCGAGATAGTCAGC
MM120_Sanger         CGTGGTATCAGGTGAGGmCAGGCCACTCCCCTGTrTTGGTCATCTATCGAGATAGTCAsC
                          ******************************************************

MM120_NGS_C          GGCCCTCAGGGATCCCTGAGCGATTCTCTGGCTCCAACTCTGGGGACACAGCCACTCTGA
MM120_Sanger         GGCCCTCAGGGATCCCTGAGCGATTCTCTGGCTCCAACTCTGGGGACACAGCCACTCTGA
                     ************************************************************

MM120_NGS_C          CCATCAGCGGGACCCAGGCTATGGATGAGGCTGACTACTACTGTCAGGCGTGGGACAGCA
MM120_Sanger         CCATCAGCGGGACCCAGGCTATGGATGAGGCTGACTACTACTGTCAGGCGTGGGACAGCA
                     ************************************************************

MM120_NGS_C          CCTCCTATTATGTCTTCGGGCCTGGGACCACGGTCACTGTCCTAGGTCAGCCCAAGGCCA
MM120_Sanger         CCTCCTATTATGTCTTCGGGCCTGGGACCACGGTCACTGTCCTAGGTCAGCCCAAGGCCA
                     ************************************************************

MM120_NGS_C          ACCCCACTGTCACTCTGTTCCCGCCCTCCTCTGAGGAGCTCCAAGCCAACAAGGCCACAC
MM120_Sanger         ACCCCACTGTCACTCTGTTCCCGCCCTCCTCTGAGGAGCTCCAAGCCAACAAGGCCACAC
                     ************************************************************

MM120_NGS_C          TAGTGTGTCTGATCAGTGACTTCTACCCGGGAGCTGTGACAGTGGCCTGGAAGGCAGATG
MM120_Sanger         TAGTGTGTCTGATCArTGACTTCTACCCGGGAGCTGTGACAGTGGCCTGGAAGGCAGATG
                     ************************************************************

MM120_NGS_C          GCAGCCCCGTCAAGGCGGGAGTGGAGACCACCAAACCCTCCAAACAGAGCAACAACAAGT
MM120_Sanger         GCAGCCCyGTCAAGGCGGGAGTGGAGACCACCAAACCCTCCAAACAGAGCAACAACAAGT
                     ************************************************************

MM120_NGS_C          ACGCGGCCAGCAGCTACCTGAGCCTGACGCCCGAGC-----------------------
MM120_Sanger         ACGCGGCCAGCAGCTACCTGAGCCTGACGCCCGAGCAGTGGAAGTCCCACAGAAGCTACA
                     **********************************

MM120_NGS_C          --------------------------------
MM120_Sanger         GCTGCCAGGTCACGCATGAAGGGAGCACCGTGGA
```

Supplementary Information Figure 14. Sequence comparison between MM120 IGLV3-1 assigned light chain sequences generated through a next-generation sequencing and Sanger sequencing approach. For the sequence generated by the Sanger sequencing approach only dominant signals were used. * = nucleotide position consistent, MM = multiple myeloma patient, NGS = light chain sequence generated through next-generation sequencing, Sanger = light chain sequence generated through Sanger sequencing. Sequencing was performed using the VLKL3_I_fw_NB and CLKL_A_rv_NB oligonucleotides. In the bioinformatic analysis of the bulk RNA sequencing, several sequence sections were given – indicated by NGS_A, NGS_B, and NGS_C. Nucleotide signal overlaps in the Sanger sequencing were specified accoding to the IUPAC code. The fourth sequence section is not shown as it shows only partial sequence overlap with the CDR3 region, but cannot be completely aligned.



Supplementary Information Figure 15. Sequence alignment of IGLV3-1 assigned AL amyloidosis and multiple myeloma amino acid light chain sequences. N- and C-terminal sequence regions that were excluded in the analysis are highlightet in grey. Bold = reference sequences, underlined = CDR regions, red letter = mutation, X and grey highlight = not unambiguously determined amino acid, green letter = linker region, MM = multiple myeloma patient, _H = AL amyloidosis patient with dominant heart involvement, _HK = AL amyloidosis patient with dominant heart and kidney involvement, _HTX = AL amyloidosis patient who received a heart transplant, FOR132 was neglected in the analysis.

Supplementary Information Figure 16. Sequence alignment of IGLV3-1 assigned AL amyloidosis and multiple myeloma cDNA light chain sequences. green letter = linker region, MM = multiple myeloma patient, _H = AL amyloidosis patient with dominant heart involvement, _HK = AL amyloidosis patient. _HTX = AL amyloidosis patient who received a heart transplant. Nucleotide signal overlaps in the Sanger sequencing were specified accoding to the IUPAC code.

Supplementary Information Table 5. Overview absolute the amino acid percentage of IGLV3-1 assigned AL amyloidosis and multiple myeloma light chain sequences. MM = multiple myeloma patient, _H = AL amyloidosis patient with dominant heart involvement, _HK = AL amyloidosis patient with dominant heart and kidney involvement, _HTX = AL amyloidosis patient who received a heart transplant.

	Amino acid [%]																			
Sample	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
FOR123_H	8.7	1.7	2.9	4.7	2.3	5.2	3.5	7.6	0.6	2.9	8.1	5.8	0.6	2.3	7.6	12.2	10.5	1.7	4.1	7.0
FOR179_H	8.7	1.2	2.9	5.8	1.7	5.8	3.5	7.0	0.0	4.1	7.0	6.4	0.6	2.9	7.6	12.8	9.9	1.7	4.1	6.4
FOR130_HTX	9.2	2.3	2.9	4.0	2.3	5.8	4.0	9.2	0.0	3.5	6.4	4.6	0.6	2.3	6.9	12.7	10.4	1.7	4.6	6.4
FOR161_HTX	8.7	1.7	4.1	4.7	1.7	5.8	3.5	8.7	0.0	2.9	5.8	5.2	0.6	2.3	8.1	12.8	9.9	1.7	4.7	7.0
FOR120_H	8.6	1.1	4.6	4.0	2.3	5.7	4.0	8.6	0.0	2.9	6.3	5.2	1.1	2.9	7.5	11.5	10.3	1.7	5.2	6.3
FOR206_H	9.9	1.7	3.5	4.7	2.3	5.8	4.1	8.1	0.0	4.1	5.8	5.2	0.6	2.9	7.6	11.6	10.5	1.7	3.5	6.4
FOR215_H	9.3	1.2	4.7	4.7	1.7	6.4	4.1	8.1	0.0	2.3	6.4	5.8	0.6	2.3	7.6	11.6	8.7	1.7	4.7	8.1
FOR118_H	9.2	1.2	4.0	4.6	2.3	6.4	3.5	8.1	0.0	3.5	5.8	5.8	0.6	2.3	8.7	11.0	9.2	1.7	5.2	6.9
FOR134_H	7.6	1.2	4.1	4.7	2.3	5.8	3.5	8.7	0.0	2.9	5.8	6.4	0.6	2.9	8.1	11.6	10.5	1.7	4.7	7.0
FOR132_H	9.9	1.2	3.5	5.2	1.7	5.8	3.5	8.1	0.0	1.7	6.4	6.4	0.6	2.3	7.6	12.8	8.7	1.7	4.7	8.1
FOR131_H	10.5	1.7	3.5	4.7	2.3	5.8	3.5	8.7	0.0	3.5	6.4	5.8	0.0	2.3	7.6	12.2	8.7	1.7	4.7	6.4
FOR116_H	9.3	1.2	2.9	4.7	1.7	5.8	3.5	8.1	0.0	3.5	6.4	5.8	1.2	2.9	7.0	14.0	9.9	1.7	4.1	6.4
FOR195_H	8.7	1.7	2.9	4.7	2.3	6.4	4.1	7.6	0.0	2.9	6.4	5.2	0.6	2.3	7.0	13.4	9.9	1.7	4.7	7.6
FOR191_HK	8.7	1.2	2.9	4.7	2.3	5.8	3.5	8.7	0.0	2.9	5.8	6.4	0.6	2.3	7.0	13.4	9.3	1.7	4.7	8.1
FOR189_HK	8.7	1.7	2.9	4.7	2.3	5.2	3.5	7.6	0.6	2.9	8.1	5.8	0.6	2.3	7.6	12.2	10.5	1.7	4.1	7.0
MM120	7.9	2.3	3.4	5.1	1.7	4.0	3.4	8.5	0.6	2.8	6.2	5.6	0.6	2.3	7.9	12.4	10.2	1.7	5.6	7.9
MM139	8.7	1.7	3.5	4.7	1.7	5.2	4.7	7.6	0.0	2.9	6.4	5.8	0.6	2.3	7.6	14.5	8.7	1.7	4.7	7.0
MM146	8.8	1.8	3.5	3.5	2.3	5.8	4.1	8.2	0.0	3.5	6.4	6.4	0.6	2.3	7.6	12.9	8.8	1.8	4.7	7.0

Appendix Results: IGLV3-21

```
MM106_NGS           ATGGCCTGGACCGTTCTCCTCCTCGGCCTCCTCTCTCACTGCACAGGCTCTGTGACCTCC
MM106_Sanger        ------------------------------------------------------------


MM106_NGS           TATGTGCTGACTCAGCCACCCTCGGTGTCAGTGGCCCCAGGACAGACGGCCAGGATTACC
MM106_Sanger        ------------CAGCCACCCTCGGTGTCAGTGGCCCCAGGACAGACGGCCAGGATTACC
                            ********************************************

MM106_NGS           TGTGGGGGAAACAACATTGGAAGTAAAAGTGTGCACTGGTATCAGCAGAAGCCAGGCCAG
MM106_Sanger        TGTGGGGGAAACAACATTGGAAGTAAAAGTGTGCACTGGTATCAGCAGAAGCCAGGCCAG
                    ************************************************************

MM106_NGS           GCCCCTGTGCTGGTCGTCTATGATGATAACGACCGGCCCTCAGGGATCCCTGAGCGATTC
MM106_Sanger        GCCCCTGTGCTGGTCGTCTATGATGATAACGACCGGCCCTCAGGGATCCCTGAGCGATTC
                    ************************************************************

MM106_NGS           TCTGGCTCCAACTCTGGGAACACGGCCACCCTGACCATCAGGAGGGTCGAAGCCGGGGAT
MM106_Sanger        TCTGGCTCCAACTCTGGGAACACGGCCACCCTGACCATCAGGAGGGTCGAAGCCGGGGAT
                    ************************************************************

MM106_NGS           GAGGCCGACTATTACTGTCAGGTGTGGGAGAGTAATAGTGATCATCCGGTATTTGGCGGA
MM106_Sanger        GAGGCCGACTATTACTGTCAGGTGTGGGAGAGTAATAGTGATCATCCGGTATTTGGCGGA
                    ************************************************************

MM106_NGS           GGGACCAAGCTGACCGCCCTAGGTCAGCCCAAGGCTGCCCCCTCGGTCACTCTGTTCCCG
MM106_Sanger        GGGACCAAGCTGACCGCCCTAGGTCAGCCCAAGGCTGCCCCCTCGGTCACTCTGTTCCCG
                    ************************************************************

MM106_NGS           CCCTCCTCTGAGGAGCTTCAAGCCAACAAGGCCACACTGGTGTGTCTCATAAGTGACTTC
MM106_Sanger        CCCTCCTCTGAGGAGCTTCAAGCCAACAAGGCCACACTGGTGTGTCTCATAAGTGACTTC
                    ************************************************************

MM106_NGS           TACCCGGGAGCCGTGACAGTGGCCTGGAAGGCAGATAGCAGCCCCGTCAAGGCGGGAGTG
MM106_Sanger        TACCCGGGAGCCGTGACAGTGGCCTGGAAGGCAGATAGCAGCCCCGTCAAGGCGGGAGTG
                    ************************************************************

MM106_NGS           GAGACCACCACACCCTCCAAACAAAGCAACAACAAGTACGCGGCCAGCAGCTATCTGAGC
MM106_Sanger        GAGACCACCACACCCTCCAAACAAAGCAACAACAAGTACGCGGCCAGCAGCTATCTGAGC
                    ************************************************************

MM106_NGS           CTGACGCCTGAGCAGTGGAAGTC-----------------------------------
MM106_Sanger        CTGACGCCTGAGCAGTGGAAGTCCCACAGAAGCTACAGCTGCCAGGTCACGCATGAAGGG
                    ***********************

MM106_NGS           ----------
MM106_Sanger        AGCACCGTGG
```

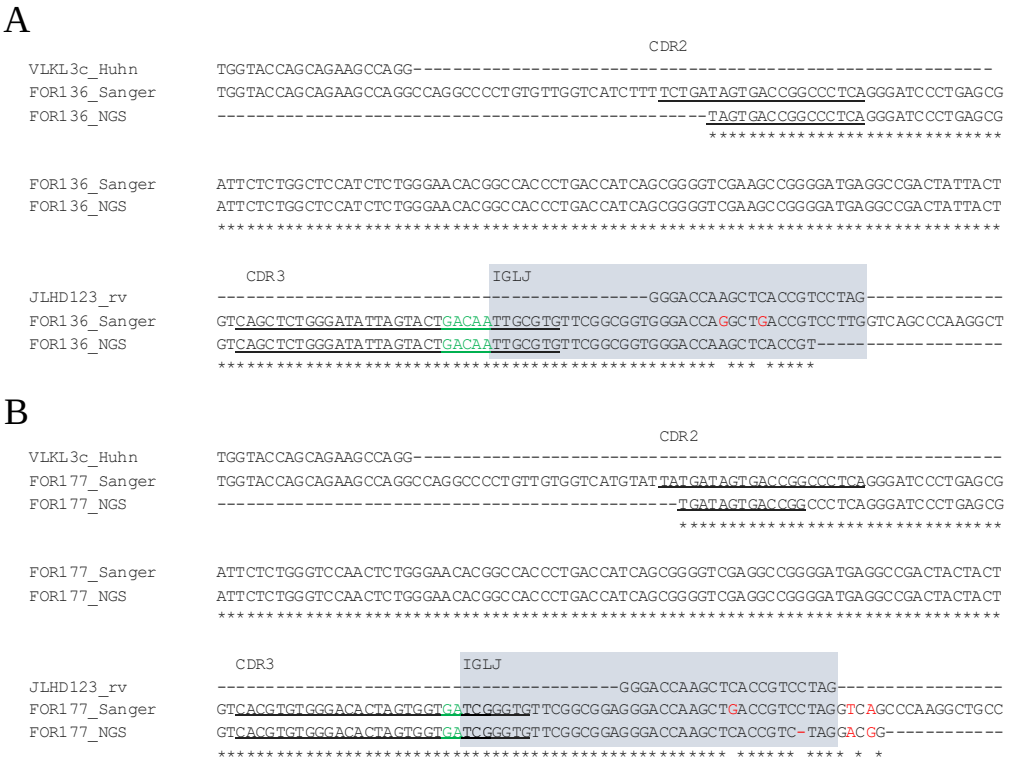
Supplementary Information Figure 17. Sequence comparison between MM106 IGLV3-21 assigned light chain sequences generated through a next-generation sequencing and Sanger sequencing approach. For the sequence generated by the Sanger sequencing approach only dominant signals were used. * = nucleotide position consistent, MM = multiple myeloma patient, NGS = light chain sequence generated through next-generation sequencing, Sanger = light chain sequence generated through Sanger sequencing. Sequencing was performed using the VLKL3_A_fw_NB and CLKL_A_rv_NB oligonucleotides.

```
MM111_NGS           ATGGCCTGGACCGTTCTCCTCCTCGGCCTCCTCTCTCACTGCACAGGGTCTGTGACCTCC
MM111_Sanger        ------------------------------------------------------------


MM111_NGS           TTTGTGCTGACTCAGCCACCCTCGGTGTCAGTGGCCCCAGGACAGACGGCCAGGTTCACC
MM111_Sanger        ------------CAGCCACCCTCGGTGTCAGTGGCCCCAGGACAGACGGCCAGGTTCACC
                                ********************************************


MM111_NGS           TGTGGGGGAAACAACATTGGAAGTGAAAGTGTGCACTGGTACCAGCAGAAGCCAGGCCAG
MM111_Sanger        TGTGGGGGAAACAACATTGGAAGTGAAAGTGTGCACTGGTACCAGCAGAAGCCAGGCCAG
                    ************************************************************


MM111_NGS           GCCCCTGCGTTGGTCGTCTATGATGACAGCGACCGGCCCTCAGGGATCCCTGAGCGATTC
MM111_Sanger        GCCCCTGCGTTGGTCGTCTATGATGACAGCGACCGGCCCTCAGGGATCCCTGAGCGATTC
                    ************************************************************


MM111_NGS           TCTGGCTCCAGCTCTGGGAACACGGCCACCCTGACCATCAGCAGGGTCGAAGCCGGGGAT
MM111_Sanger        TCTGGCTCCAGCTCTGGGAACACGGCCACCCTGACCATCAGCAGGGTCGAAGCCGGGGAT
                    ************************************************************


MM111_NGS           GAGGCCGACTATTTCTGTCAGGTGTGGGATAGTGATGGTGATCATTGGGTGTTCGGCGGA
MM111_Sanger        GAGGCCGACTATTTCTGTCAGGTGTGGGATAGTGATGGTGATCATTGGGTGTTCGGCGGA
                    ************************************************************


MM111_NGS           GGGACCAAGCTGAGCGTCCTAGGTCAGCCCAAGGCTGCCCCCTCGGTCACTCTGTTCCCA
MM111_Sanger        GGGACCAAGCTGAGCGTCCTAGGTCAGCCCAAGGCTGCCCCCTCGGTCACTCTGTTCCCA
                    ************************************************************


MM111_NGS           CCCTCCTCTGAGGAGCTTCAAGCCAACAAGGCCACACTGGTGTGTCTCATAAGTGACTTC
MM111_Sanger        CCCTCCTCTGAGGAGCTTCAAGCCAACAAGGCCACACTGGTGTGTCTCATAAGTGACTTC
                    ************************************************************


MM111_NGS           TACCCGGGAGCCGTGACAGTGGCCTGGAAGGCAGATAGCAGCCCCGTCAAGGCGGGAGTG
MM111_Sanger        TACCCGGGAGCCGTGACAGTGGCCTGGAAGGCAGATAGCAGCCCCGTCAAGGCGGGAGTG
                    ************************************************************


MM111_NGS           GAGACCACCACACCCTCCAAACAAAGCAACAACAAGTACGCGGCCAGCAGCTACCTGAGC
MM111_Sanger        GAGACCACCACACCCTCCAAACAAAGCAACAACAAGTACGCGGCCAGCAGCTACCTGAGC
                    ************************************************************


MM111_NGS           CTGACGCCTGAGCAGTGG------------------------------------------
MM111_Sanger        CTGACGCCTGAGCAGTGGAAGTCCCACAAAAGCTACAGCTGCCAGGTCACGCATGAAGGG
                    ******************


MM111_NGS           ---------
MM111_Sanger        AGCACCGTG
```

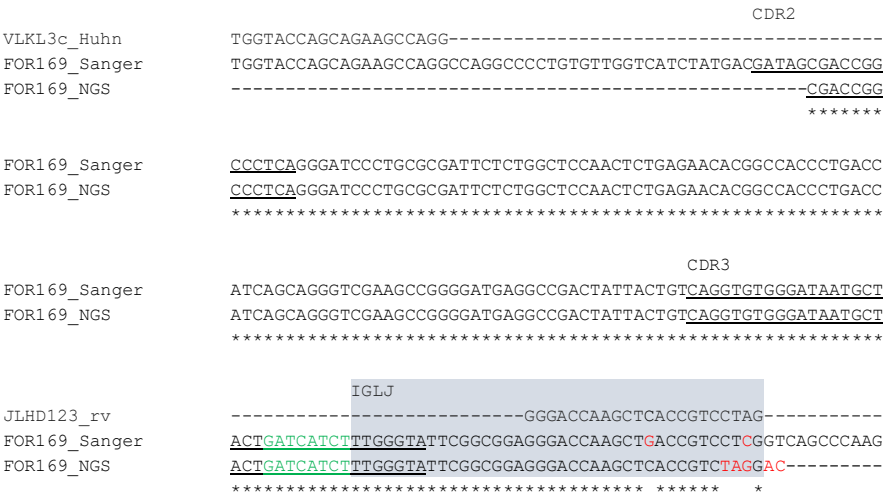
Supplementary Information Figure 18. Sequence comparison between MM111 IGLV3-21 assigned light chain sequences generated through a next-generation sequencing and Sanger sequencing approach. For the sequence generated by the Sanger sequencing approach only dominant signals were used. * = nucleotide position consistent, MM = multiple myeloma patient, NGS = light chain sequence generated through next-generation sequencing, Sanger = light chain sequence generated through Sanger sequencing. Sequencing was performed using the VLKL3_A_fw_NB and CLKL_A_rv_NB oligonucleotides.

```
MM119_NGS       ATGGCCTGGACCGTTCTCCTCCTCGGCCTCCTCTCTCACTGCACAGGGTCTCTGAATTCT
MM119_Sanger    ------------------------------------------------------------


MM119_NGS       TATGTGCTGACTCAGCCACCCTCGGTGTCAGTGGCCCCAGGACAGACGGCCAGGATTACC
MM119_Sanger    ------------CAGCCACCCTCGGTGTCAGTGGCCCCAGGACAGACGGCCAGGATTACC
                            ********************************************


MM119_NGS       TGTGAGGGAAACGACATTGGAGATAAAAGTTTGCAGTGGTACCAGCAGAAGCCAGGCCAG
MM119_Sanger    TGTGAGGGAAACGACATTGGArrTAAAAGTdTGCAGTGGTACCAGCAGAAGCCAGGCCAG
                ************************************************************


MM119_NGS       GCCCCTGTGTTGGTCGTCTATGATGATAGCGACCGGCCCTCATGGATCCCTGAGCGATTC
MM119_Sanger    GCCCCTGTGTTGGTsrTCTATGAyGATAGCGACCGGCCCTCAkGGATCCCTGAGCGATTC
                ************************************************************


MM119_NGS       TCTGGCTCCAACTCTGGGAACACGGCCACCCTGACCATCAGCAGGGTCGAAGCCGGGGAT
MM119_Sanger    TCTGGCTCCArCTCwGGGAmCACGGCCACCyTGACCATCAGCrGGGTCGAAGCCGrGGAT
                ************************************************************


MM119_NGS       GAGGCCGACTATTATTGTCAGTTGTGGGATGGTTATACTCTCAGGGGGGTATTCGGCGGA
MM119_Sanger    GAGGCyGACTATTATTGTCAGTyGTsGrATGGywATACTCTCAkGGGGGTATTCGGCGGA
                ************************************************************


MM119_NGS       GGGACCAAGCTGACCGTCCTAGGTCAGCCCAAGGCTGCCCCCTCGGTCAGTCTGTTCCCG
MM119_Sanger    GGGACCAAGCTGACCGTCCTAGGTCAGCCCAAGGCTGCCCCCTCGGTCAvTCTGTTCCCG
                ************************************************************


MM119_NGS       CCCTCCTCTGAGGAGCTTCAAGCCAACAAGGCCACACTGGTGTGTCTCATAAGTGACTTC
MM119_Sanger    CCCTCCTCTGAGGAGCTTCAAGCCAACAAGGCCACACTGGTGTGTCTCATAAkTGACTTC
                ************************************************************


MM119_NGS       TACCCGGGAGCCGTGACAGTGGCCTGGAAGGCAGATAGCAGCCCCGTCAAGGCGGGAGTG
MM119_Sanger    TACCCGGrAGCCGTGACAGkGGCCTGGAAGGCArATAGCAGCCCCGTCAAGGCGGGAGTG
                ************************************************************


MM119_NGS       GAGACCACCACACCCTCCAAACAAAGCAATAACAAGTACGCGGCCAGCAGCTA-------
MM119_Sanger    GAGACCACCACACCCTCCAAACAAAGCAAyAACAAGTACGCGGCCAGCAGCTAyCTGAGC
                ************************************************


MM119_NGS       --------------------
MM119_Sanger    CTGACGCCTGAGCAGTGGAA
```

Supplementary Information Figure 19. Sequence comparison between MM111 IGLV3-21 assigned light chain sequences generated through a next-generation sequencing and Sanger sequencing approach. For the sequence generated by the Sanger sequencing approach only dominant signals were used. * = nucleotide position consistent, MM = multiple myeloma patient, NGS = light chain sequence generated through next-generation sequencing, Sanger = light chain sequence generated through Sanger sequencing. Sequencing was performed using the VLKL3_A_fw_NB and CLKL_A_rv_NB oligonucleotides. Nucleotide signal overlaps in the Sanger sequencing were specified accoding to the IUPAC code.



Supplementary Information Figure 20. Sequence alignment of the Sanger sequence and the sequence with the highest percentage in the IGLV3-21 next-generation sequencing experiment based on multiplex oligonucleotide PCR samples for FOR136 and FOR177. (MP_NGS_PCR) (C) sequence alignment FOR136 (D) sequence alignment FOR177. CDR region = underlined, IGLJ segment = blue highlight, linker region = green, sequence deviation = red, * = matching position, VLKL3c_huhn = IGLV3 specific forward oligonucleotide, JLHD123_rv = IGLJ segment specific reverse oligonucleotide, Sanger = sequence generated by Sanger sequencing, NGS = sequence with the highest percentage in the next-generation sequencing experiment.



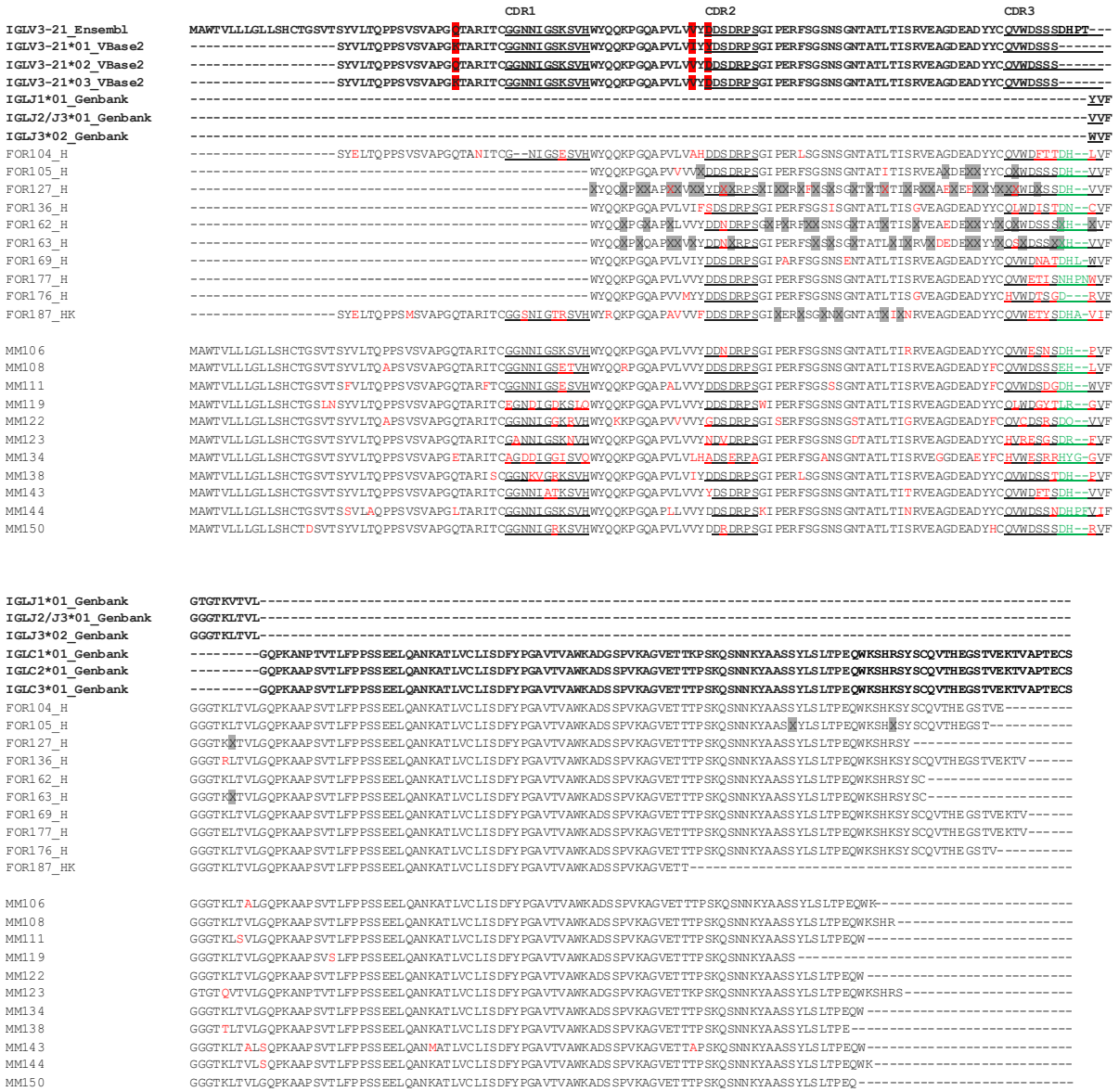
Supplementary Information Figure 21. Sequence alignment of the FOR169 Sanger sequence and the sequence with the highest percentage in the IGLV3-21 next-generation sequencing experiment based on multiplex oligonucleotide PCR samples. (MP_NGS_PCR). CDR region = underlined, IGLJ segment = blue highlight, linker region = green, sequence deviation = red, * = matching position, VLKL3c_huhn = IGLV3 specific forward oligonucleotide, JLHD123_rv = IGLJ segment specific reverse oligonucleotide, Sanger = sequence generated by Sanger sequencing, NGS = sequence with the highest percentage in the next-generation sequencing experiment.

```
FOR113_NGS_PCR_MP_sequenceA      TAACAACCGGCCCTCTGGGATCCCTGAGCGATTCTCTGCCTCCAACTCGGGGAACACGGC
FOR113_NGS_PCR_N_sequenceA       -----AGAGGCCCTCAGGGATCCCTGAGCGATTCTCTGGCTCCAGTTCAGGGACAATAGC
                                      *  ******* ********************** *****  ** ****  *  **

FOR113_NGS_PCR_MP_sequenceA      CACCCTGACCATCAGCAGAGCCCAAGCCGGGGATGAGGCTGACTATTACTGTCAGGTGTG
FOR113_NGS_PCR_N_sequenceA       CACGTTGACCATCAGTGGAGTCCAGGCAGACGACGAGGCTGTCTATTACTGTCAATCAGC
                                 ***  **********  *** *** ** *  ** ******* ************

FOR113_NGS_PCR_MP_sequenceA      GGACATCAGTATT------GGGGTGTTCGGCGGAGGGACCAAGCTCACCGT---
FOR113_NGS_PCR_N_sequenceA       ATACAACAGTGGTAAATCTTTTGTCTTCGGAACTGGGACCAAGCTCACCGTCTA
                                   *** ****  *         ** *****    *****************


FOR136_NGS_PCR_MP_sequence_A      ---TAGTGACCGGCCCTCAGGGATCCCTGAGCGATTCTCTGGCTCCATCTCTGGGAACAC
FOR136_NGS_PCR_N_sequence_A       GACAACTGAGAGGCCCTCAGGGATCCCTGACCGATTCTCTGGCTCCAGCTCAGGGACAAT
                                      * ***  ****************** *************** *** ****  *

FOR136_NGS_PCR_MP_sequence_A      GGCCACCCTGACCATCAGCGGGGTCGAAGCCGGGGATGAGGCCGACTATTACTGTCAGCT
FOR136_NGS_PCR_N_sequence_A       AGTCACGCTGACCATCAGTGGGGTCCAGGCTGAAGACGAGGCTGACTATTATTGTCAATC
                                   * *** *********** ****** * ** *  ** ***** ******** *****

FOR136_NGS_PCR_MP_sequence_A      CTGGGATATTAGTACTGACAATTGCGTGTTCGGCGGTGGGACCAAGCTCACCGT
FOR136_NGS_PCR_N_sequence_A       AGCAGACAGCAGCGGCACTTATCGGGTGTTCGGCGGAGGGACCAAGCTCACCGT
                                      ** *  **        ** * *********** *****************


FOR163_NGS_PCR_MP_sequence_A      -----CCGGCCCTCGGGGATCCCTGAGCGATTCTCTGGCGCCAACTCTGGGAACACGGCC
FOR163_NGS_PCR_N_sequence_A       AATGAGAGGCCCTCAGGGATCCCTGAGCGATTCTCTGGCTCCAGCTCAGGGACTACAGCC
                                         ******* ************************ *** *** ****  ** ***

FOR163_NGS_PCR_MP_sequence_A      ACCCTGACCATCAGCAGGGTCGAAGCCGGGGATGAGGCCGACTATTACTGTCAGGTGTGG
FOR163_NGS_PCR_N_sequence_A       ACGTTGACCATCAGTGCAGTCCAGCCAGAGGACGAGGCTGACTTCTACTGTCAATCACTT
                                  **  **********    *** *  * * *** ***** ****  ********

FOR163_NGS_PCR_MP_sequence_A      GATAGCAGTAAGAGTCATCAAATTTTCGGCGGAGGGACCAAGCTCACCGTCTA
FOR163_NGS_PCR_N_sequence_A       GACACCAGTGGCACTTCTGTGGTTTTCGGCGGAGGGACCAAGCTCACCGTCTA
                                  ** * ****   * *  *    ******************************


FOR169_NGS_PCR_MP_sequence_A      ----------CGACCGGCCCTCAGGGATCCCTGCGCGATTCTCTGGCTCCAACTCTGAGA
FOR169_NGS_PCR_N_sequence_A       ATAAGGACAGTGAGAGGCCCTCAGGGATCCCTGAGCGATTCTCTGGCTCCACCTCAGGGA
                                            **  ****************** ***************** *** * **

FOR169_NGS_PCR_MP_sequence_A      ACACGGCCACCCTGACCATCAGCAGGGTCGAAGCCGGGGATGAGGCCGACTATTACTGTC
FOR169_NGS_PCR_N_sequence_A       CAACAGTCACGCTGACCATCAGTGGAGTCCAGGCAGAAGACGAGGCTGACTATTACTGTC
                                    ** * *** ***********  * *** * ** *  ** ***** *************

FOR169_NGS_PCR_MP_sequence_A      AGGTGTGGGATAATGCTACTGATCATCTTTGGGTATTCGGCGGAGGGACCAAGCTCACCG
FOR169_NGS_PCR_N_sequence_A       AATCAGCAGAC--AGCGGTGG-AACCCATGTGGTTTTCGGCGGTGGGACCAAGCTCACCG
                                  *       **    **    *     * *  *** ******** ****************

FOR169_NGS_PCR_MP_sequence_A      TCTAGGAC--------
FOR169_NGS_PCR_N_sequence_A       TCTAGGACGGTGAGCT
                                  ********
```

Supplementary Information Figure 22. Sequence alignment of the FOR113, FOR136, FOR163 and FOR169 sequences with the highest percentage in the two IGLV3-21 next-generation sequencing experiments. * = matching position, _A = sequence with the highest percentage in the next-generation sequencing experiment; NGS_PCR_MP = sequence generated by the next-generation sequencing approach based on multiple forward oligonucleotides, NGS_PCR_N = sequence generated by the next-generation sequencing approach based on the VLKL3_A_fw_NB forward oligonucleotide.

```
FOR113_NGS_PCR_MP_sequence_C      -------GGCCCTCAGGGATCCCTGAGCGATTCTCTGGCTCCAGCTCAGGGACAACAGTC
FOR113_NGS_PCR_MP_sequence_D      AGTCTGAGGCCCTCAGGGATCCCTGAGCGATTCTCTGGCTCCACCTCAGGGGCAGCAGTC
FOR113_NGS_PCR_MP_sequence_N      ----AGAGGCCCTCAGGGATCCCTGACCGATTCTATGGCTCCAGTTCAGGGACAACAGTC
FOR113_NGS_PCR_N_sequence_A       ----AGAGGCCCTCAGGGATCCCTGAGCGATTCTCTGGCTCCAGTTCAGGGACAATAGCC
FOR113_NGS_PCR_N_sequence_C       --TGAGAGGCCCTCAGGGATCCCTGAGCGATTCTCTGGCTCCAGTTCAGGGACAATAGCC
                                        ****************** ******* ********  ****** **  ** *

FOR113_NGS_PCR_MP_sequence_C      ACATTGACCATCAGTGGAGTCCAGGCAGAGGACGAGGCTGACTATTATTGTCAATCAGCA
FOR113_NGS_PCR_MP_sequence_D      ACGTTGACCATCAGTGGAGTCCAGGCAGAAGACGAGGCTGACTATTACTGTCAATCATCA
FOR113_NGS_PCR_MP_sequence_N      ACATTGACCATCAGTGGAGTCCAGGCAGAAGACGAGGGTGACTACTACTGTCAATCAGCA
FOR113_NGS_PCR_N_sequence_A       ACGTTGACCATCAGTGGAGTCCAGGCAGACGACGAGGCTGTCTATTACTGTCAATCAGCA
FOR113_NGS_PCR_N_sequence_C       ACGTTGACCATCAGTGGAGTCCAGGCAGACGACGAGGCTGTCTATTACTGTCAATCAGCA
                                  ** ************************* ******* ** *** ** ********* **

FOR113_NGS_PCR_MP_sequence_C      GACACCATTGGTACTTATCGGGTGTTCGGCGGAGGGACCAAGCTCACCGT----
FOR113_NGS_PCR_MP_sequence_D      GATAGCAGTGGTACATAT---GTCTTCGGAACTGGGACCAAGCTCACCGT----
FOR113_NGS_PCR_MP_sequence_N      GACGCCAGTGGCAATTATGTGGTCTTCGGTATTGGGACCAAGCTCACCGTCTAG
FOR113_NGS_PCR_N_sequence_A       TACAACAGTGGTAAATCTTTTGTCTTCGGAACTGGGACCAAGCTCACCGTCTA-
FOR113_NGS_PCR_N_sequence_C       TACAACAGTGGTAAATCTTTTGTCTTCGGAACTGGGACCAAGCTCACCGT----
                                   *   ** *** *  * *   ** *****    *****************



FOR163_NGS_PCR_MP_sequence_D      ----AGCGGCCCTCAGGGATCCCTGAGCGATTCTCTGGCTCCAGTTTAGGGACAACAGCC
FOR163_NGS_PCR_N_sequence_A       AATGAGAGGCCCTCAGGGATCCCTGAGCGATTCTCTGGCTCCAGCTCAGGGACTACAGCC
                                      ** ************************************ * ****** ******

FOR163_NGS_PCR_MP_sequence_D      ACGTTGACTATCAGTGAAGTCCAGGCAGAGGACGAGGCTGACTATTACTGTCAATCAGCA
FOR163_NGS_PCR_N_sequence_A       ACGTTGACCATCAGTGCAGTCCAGCCAGAGGACGAGGCTGACTTCTACTGTCAATCACTT
                                  ******** ******* ******* ******************  ************

FOR163_NGS_PCR_MP_sequence_D      GACAACAGTGGTGGTTATGGGGTGGCATTCGGCGGAGGGACCAAGCTCACCGT---
FOR163_NGS_PCR_N_sequence_A       GACACCAGTGGCACTTCT---GTGGTTTTCGGCGGAGGGACCAAGCTCACCGTCTA
                                  **** ******   ** *   ****  *************************



FOR169_NGS_PCR_MP_sequence_B      -----GACAGTGAGAGGCCCTCAGGGATCCCTGAGCGATTCTCTGGCTCCAGCTCAGGGA
FOR169_NGS_PCR_MP_sequence_C      ---AAGACAGTGAGAGGCCCTCAGGGATCCCTGAGCGATTCTCTGGCTCCACCTCAGGGG
FOR169_NGS_PCR_MP_sequence_D      ------ACACTGAGAGGCCCTCAGGGATCCCTGAGCGATTCTCTGGCTCCACCTCAGGGA
FOR169_NGS_PCR_N_sequence_A       ATAAGGACAGTGAGAGGCCCTCAGGGATCCCTGAGCGATTCTCTGGCTCCACCTCAGGGA
                                        *** ************************************** *******

FOR169_NGS_PCR_MP_sequence_B      CAACAGTCACGTTGACCATCAGTGGAGTCCAGGCAGAAGACGAGGCTGACTATCACTGTC
FOR169_NGS_PCR_MP_sequence_C      CAACAGCCACGTTGACCATTAGTGGAGTCCGGACAGAAGACGAGGCTGACTATTACTGTC
FOR169_NGS_PCR_MP_sequence_D      CAACAGTCACGTTGACCATCAGTGGAGTCCAGGCAGAAGACGAGGGTGACTATTACTGTC
FOR169_NGS_PCR_N_sequence_A       CAACAGTCACGCTGACCATCAGTGGAGTCCAGGCAGAAGACGAGGCTGACTATTACTGTC
                                  ****** **** ******* ********** * ************ ******* ******

FOR169_NGS_PCR_MP_sequence_B      AATCAGCAGACAGCAGTGGTA------CTTATGTGGTATTCGGCGGAGGGACCAAGCTCA
FOR169_NGS_PCR_MP_sequence_C      AATCAGCAGACAGCACTAGTC------CTTATG---TCTTCGGAACTGGGACCAAGCTCA
FOR169_NGS_PCR_MP_sequence_D      AATCAACAGACAACAGTGGTCCTTATCCCTTTGTGGTATTCGGCGGAGGGACCAAGCTCA
FOR169_NGS_PCR_N_sequence_A       AATCAGCAGACAGCGGTGGAA------CCCATGTGGTTTTCGGCGGTGGGACCAAGCTCA
                                  ***** ****** *  * *        *   **   * *****    *************

FOR169_NGS_PCR_MP_sequence_B      CCG----------------
FOR169_NGS_PCR_MP_sequence_C      CCGT---------------
FOR169_NGS_PCR_MP_sequence_D      CCGT---------------
FOR169_NGS_PCR_N_sequence_A       CCGTCTAGGACGGTGAGCT
                                  ***
```

Supplementary Information Figure 23. Sequence comparison of the FOR113, FOR136, FOR163 and FOR169 sequences with the highest percentage in the N$_{\text{NGS_PCR}}$ approach and corresponding sequences in the MP$_{\text{NGS_PCR}}$ approach. * = matching position, NGS_PCR_MP = sequence generated by the next-generation sequencing approach based on multipley forward oligonucleotides, NGS_PCR_N = sequence generated by the next-generation sequencing approach based on the VLKL3_A_fw_NB forward oligonucleotide, next-generation sequencing sequences were designated in descending alphabetical order based on the percentage.



Supplementary Information Figure 24. Sequence alignment of IGLV3-21 assigned AL amyloidosis and multiple myeloma amino acid light chain sequences. N- and C-terminal sequence regions that were excluded in the analysis are highlightet in grey. Bold = reference sequences, underlined = CDR regions, red highlight = discrepancy between the VBase2 and Ensembl IGLV3-21 reference, red letter = mutation, X and grey highlight = not unambiguously determined amino acid, green letter = linker region, MM = multiple myeloma patient, _H = AL amyloidosis patient with dominant heart involvement, _HK = AL amyloidosis patient with dominant heart and kidney involvement.

Supplementary Information Figure 25. Sequence alignment of IGLV3-21 assigned AL amyloidosis and multiple myeloma cDNA light chain sequences. green letter = linker region, MM = multiple myeloma patient, _H = AL amyloidosis patient with dominant heart involvement, _HK = AL amyloidosis patient. Nucleotide signal overlaps in the Sanger sequencing were specified accoding to the IUPAC code.

Appendix Results: IGLV6-57 and IGLV2-23

```
MM112_NGS         ATGGCCTGGGCTCCACTACTTCTCACCCTCCTCGCTCACTGCACAGGTTCTTGGGCCAAT
MM112_Sanger      ------------------------------------------------------------


MM112_NGS         TTTATGCTGACTCAGCCGCACTCTGTGTCGGAGTCTCCGGGGAAGACGGTGACCATCTCC
MM112_Sanger      ------------------------GTGTCGGAGTCTCCGGGGAAGACGGTGACCATCTCC
                                          **********************************


MM112_NGS         TGCACCCGCAGCAGTGGCAACATTGCCAGCAACTATGTGCAGTGGTACCAGCAGCGCCCG
MM112_Sanger      TGCACCCGCAGCAGTGGCAACATTGCCAGCAACTATGTGCAGTGGTACCAGCAGCGCCCG
                  ************************************************************


MM112_NGS         GGCAGTGCCCCCACCACTGTGATCTATGAGGATAATCAAAGACCCTCTGGGGTCCCTGAT
MM112_Sanger      GGCAGTGCCCCCACCACTGTGATCTATGAGGATAATCAAAGACCCTCTGGGGTCCCTGAT
                  ************************************************************


MM112_NGS         CGGTTCTCTGGCTCCATCGACAGGTCCTCCAAGTCTGCCTCCCTCACCATCTCTCGACTG
MM112_Sanger      CGGTTCTCTGGCTCCATCGACAGGTCCTCCAAGTCTGCCTCCCTCACCATCTCTCGACTG
                  ************************************************************


MM112_NGS         AAGACTGAGGACGAGGCTGACTACTACTGTCAGTCTTATGATGACAACAATCTTTGGGTG
MM112_Sanger      AAGACTGAGGACGAGGCTGACTACTACTGTCAGTCTTATGATGACAACAATCTTTGGGTG
                  ************************************************************


MM112_NGS         TTCGGCGGAGGGACCAAGCTGACCGTCCTGAGTCAGCCCAAGGCTGCCCCCTCGGTCACT
MM112_Sanger      TTCGGCGGAGGGACCAAGCTGACCGTCCTGAGTCAGCCCAAGGCTGCCCCCTCGGTCACT
                  ************************************************************


MM112_NGS         CTGTTCCCACCCTCCTCTGAGGAGCTTCAAGCCAACAAGGCCACACTGGTGTGTCTCATA
MM112_Sanger      CTGTTCCCACCCTCCTCTGAGGAGCTTCAAGCCAACAAGGCCACACTGGTGTGTCTCATA
                  ************************************************************


MM112_NGS         AGTGACTTCTACCCGGGAGCCGTGACAGTGGCCTGGAAGGCAGATAGCAGCCCCGTCAAG
MM112_Sanger      AGTGACTTCTACCCGGGAGCCGTGACAGTGGCCTGGAAGGCAGATAGCAGCCCCGTCAAG
                  ************************************************************


MM112_NGS         GCGGGAGTGGAGACCACCACACCCTCCAAACAAAGCAACAACAAGTACACGGCCAGCAGC
MM112_Sanger      GCGGGAGTGGAGACCACCACACCCTCCAAACAAAGCAACAACAAGTACACGGCCAGCAGC
                  ************************************************************


MM112_NGS         TACCTGAGCCTGACGCCTGAGCAGT----------------------------------
MM112_Sanger      TACCTGAGCCTGACGCCTGAGCAGTGGAAGTCCCACAAAAGCTACAGCTGCCAGGTCACG
                  ************************


MM112_NGS         --------------------
MM112_Sanger      CATGAAGGGAGCACCGTGG
```

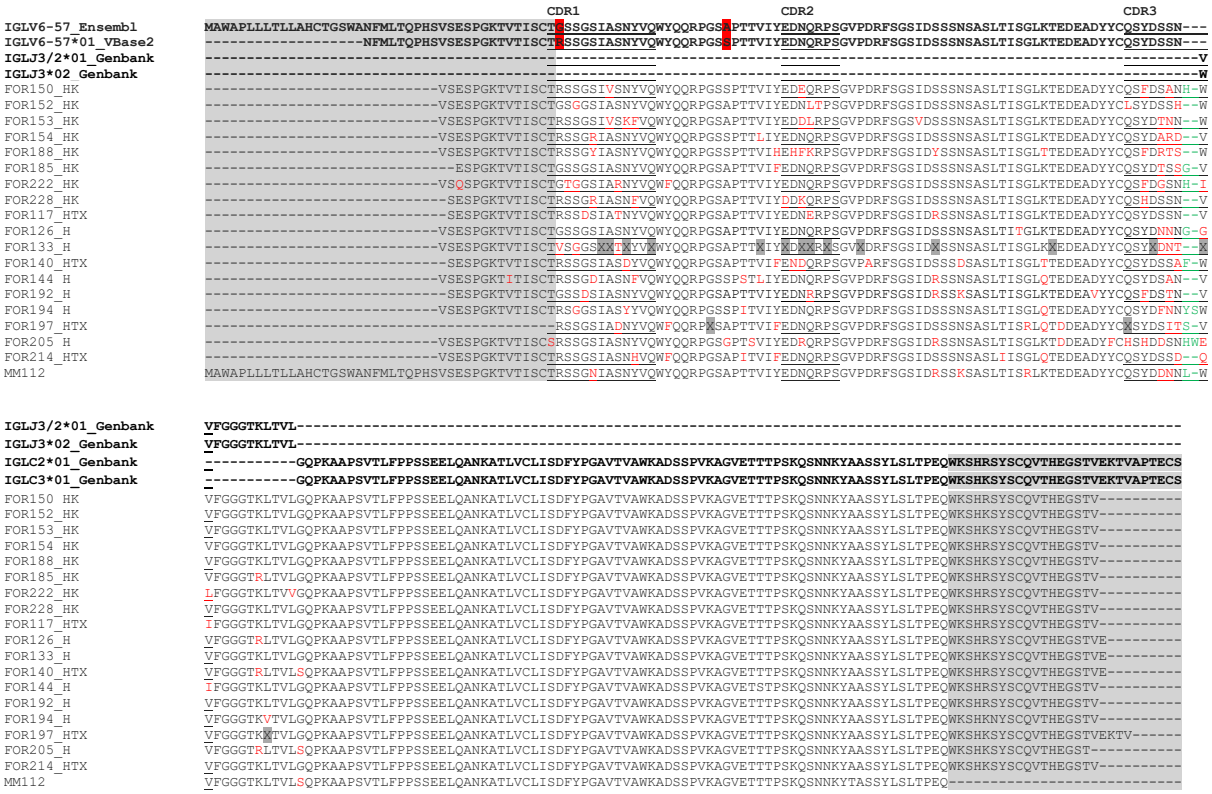
Supplementary Information Figure 26. Sequence comparison between MM112 IGLV6-57 assigned light chain sequences generated through a next-generation sequencing and Sanger sequencing approach. For the sequence generated by the Sanger sequencing approach only dominant signals were used. * = nucleotide position consistent, MM = multiple myeloma patient, NGS = light chain sequence generated through next-generation sequencing, Sanger = light chain sequence generated through Sanger sequencing. Sequencing was performed using the VLKL3_A_fw_NB and CLKL_A_rv_NB oligonucleotides.

```
MM107_A      ------------------------------------------------------------
MM107_B      GGGTCACAAGAGGCAGCGCTCTCGGGACGTCTCCACCATGGCCTGGGCTCTGCTGCTCCT

MM107_A      ------------------------------------------------------------
MM107_B      CACTCTCCTCACTCAGGACACAGGGTCCTGGGCCCAGTCTGCCCTGACTCAGCCTGCCTC

MM107_A      ------------------------------------------------------------
MM107_B      CGTGTCTGGGTCTCTTGGACAGTCGGTCACCATCTCCTGCTCTGGATCCAACGGTGATGT

MM107_A      ------------------------------------------------------------
MM107_B      TGGGACTTATAACCTTGTCTCTTGGTATCAACACCACCCAGGCGAAGCCCCCAAACTCGT

MM107_A      ------------------------------------------------------------
MM107_B      GGTTTATGAGGGCAGTAAGCGGCCCTCCGGGGTTTCTTGGCGCTTCTCTGGCTCCAAGTC

MM107_A      ------------------------------------------------------------
MM107_B      TGGCAACACGGCCTCCCTGGAAATTTTTGACCTCCAGGCTGAGGACGAGGCCGATTATTA

MM107_A      ------------------------------------------------------------
MM107_B      CTGCTGCTCCTATGCAGGTAGTAGTACTTTGATTTTCGGCGGAGGGACCAAAGTGACCGT

MM107_A      ----------------------------------------CCCGCCCTCCTCTGAGGCAAA
MM107_B      CCTAGGTCAGCCCAAGGCTGCCCCCTCGGTCACTCTGTTCCCGCCCTCCTCTGAGGAGCT
                                                      ****************

MM107_A      GGGG----ATAAGA-----CAG---------------------------------------
MM107_B      TCAAGCCAACAAGGCCACACTGGTGTGTCTCATAAGTGACTTCTACCCGGGAGCCGTGAC
                     * ***      * *

MM107_A      ------------------------------------------------------------
MM107_B      AGTGGCCTGGAAGGCAGATAGCAGCCCCGTCAAGGCGGGAGTGGAGACCACCACTCCCTC

MM107_A      ------------------------------------------------------------
MM107_B      CAAACAGAGCAACAACAAGTACGCGGCCAGCAGCTATCTGAGCCTGACGCCTGAGCAGTG

MM107_A      ---
MM107_B      GAA
```

Supplementary Information Figure 27. Alignment of sequence sections given by the bioinformatic analysis of the bulk RNA sequencing approach of M107. * = nucleotide position consistent, MM = multiple myeloma patient. The two different sequence sections are indicated by _A and _B.

```
MM132_A      ------------------------------------------------------------
MM132_B      CTGGGGTCACAAGAGGCAGCGCTCTCGGGACGTCTCCACCATGGCCTGGGCTCTGCTGCT

MM132_A      ------------------------------------------------------------
MM132_B      CCTCACCCTCCTCACTCAGGACACAGGGTCCTGGGCCCAGTCTGCCCTGACTCAGCCTGC

MM132_A      ------------------------------------------------------------
MM132_B      CTCCGTGTCTGCTTCTCCTGGACAGTCGATCACCATCTCCTGCACTGGAACCAGCGGTGA

MM132_A      ------------------------------------------------------------
MM132_B      TATTGGGAGTTTTAGCCTTGTCTCCTGGTACCAACAACACCCAGGCAGAGCCCCCAAACT

MM132_A      ------------------------------------------------------------
MM132_B      CATGATTTACGAGGTCCATAAGCGGCCCTCAGGGGTTTCTACTCGCTTCTCTGGCTCCAA

MM132_A      ------------------------------------------------------------
MM132_B      GTCTGACAACACGGCCTCCCTGACAATCTCTGGGCTCCAGGCTGACGACGAGGCTGATTA

MM132_A      ------------------------------------------------------------
MM132_B      TTACTGCTTCTCATATGCAGGTAGAGGCACTTCGGGAGTCTTCGGCGGAGGGACCAGGCT

MM132_A      -----------------------------------------TGTTCCCGCCCTCCTCTGA
MM132_B      GACCGTCCTAAGTCAGCCCAAGGCTGCCCCCTCGGTCACTCTGTTCCCGCCCTCCTCTGA
                                                      ******************

MM132_A      GG----------------------------------------------------------
MM132_B      GGAGCTTCAAGCCAACAAGGCCACACTGGTGTGTCTCATAAGTGACTTCTACCCGGGAGC
             **

MM132_A      ------------------------------------------------------------
MM132_B      CGTGACAGTGGCCTGGAAGGCAGATAGCAGCCCCGTCAAGGCGGGAGTGGAGACCACCAC

MM132_A      ------------------------------------------------------------
MM132_B      ACCCTCCAAACAAAGCAACAACAAGTACGCGGCCAGCAGCTATCTGAGCCTGACGCCTGA

MM132_A      ---------
MM132_B      GCAGTGGAA
```

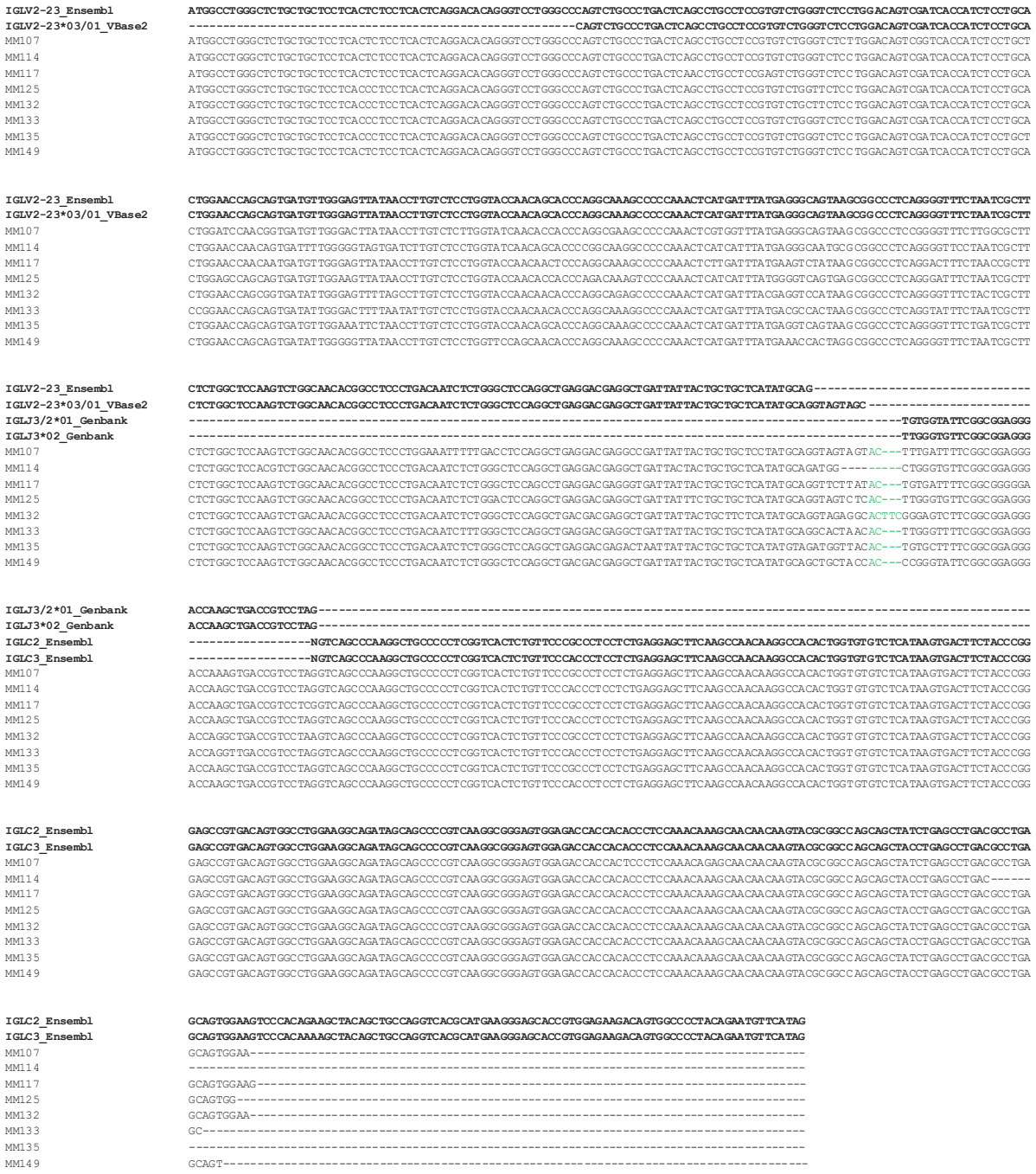
Supplementary Information Figure 28. Alignment of sequence sections given by the bioinformatic analysis of the bulk RNA sequencing approach of M132. * = nucleotide position consistent, MM = multiple myeloma patient. The two different sequence sections are indicated by _A and _B.



Supplementary Information Figure 29. Sequence alignment of IGLV6-57 assigned AL amyloidosis and multiple myeloma amino acid light chain sequences. N- and C-terminal sequence regions that were excluded in the analysis are highlightet in grey. Bold = reference sequences, underlined = CDR regions, red highlight = discrepancy between the VBase2 and Ensembl IGLV6-57 reference, red letter = mutation, X and grey highlight = not unambiguously determined amino acid, green letter = linker region, MM = multiple myeloma patient, _H = AL amyloidosis patient with dominant heart involvement, _HK = AL amyloidosis patient with dominant heart and kidney involvement, _HTX = AL amyloidosis patient who received a heart transplant.

Supplementary Information Figure 30. Sequence alignment of IGLV6-57 assigned AL amyloidosis and multiple myeloma cDNA light chain sequences. N- and C-terminal sequence regions that were excluded in the analysis are highlightet in grey. Bold = reference sequences, underlined = CDR regions, red letter = mutation, X and grey highlight = not unambiguously determined amino acid, green letter = linker region, MM = multiple myeloma patient. Nucleotide signal overlaps in the Sanger sequencing were specified accoding to the IUPAC code.

Supplementary Information Figure 31. Sequence alignment of IGLV2-23 assigned multiple myeloma amino acid light chain sequences. N- and C-terminal sequence regions that were excluded in the analysis are highlightet in grey. Bold = reference sequences, underlined = CDR regions, red letter = mutation, X and grey highlight = not unambiguously determined amino acid, green letter = linker region, MM = multiple myeloma patient.



Supplementary Information Figure 32. Sequence alignment of IGLV2-23 assigned multiple myeloma cDNA light chain sequences. N- and C-terminal sequence regions that were excluded in the analysis are highlightet in grey. Bold = reference sequences, underlined = CDR regions, red letter = mutation, X and grey highlight = not unambiguously determined amino acid, green letter = linker region, MM = multiple myeloma patient. Nucleotide signal overlaps in the Sanger sequencing were specified accoding to the IUPAC code.

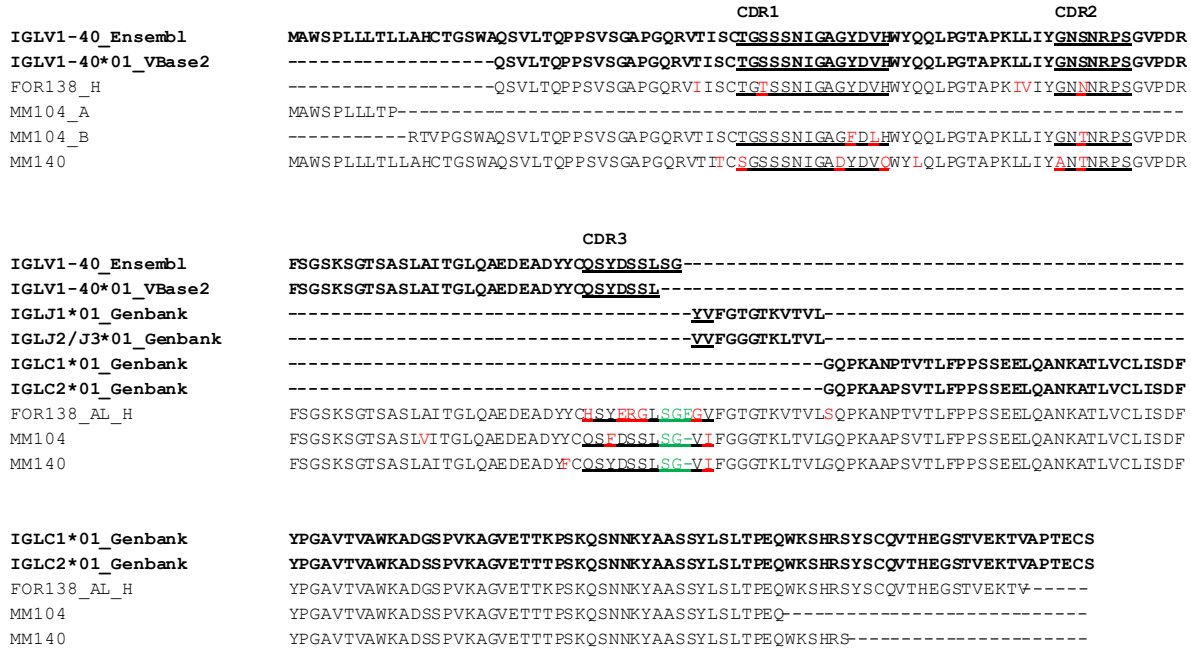
Supplementary Information Table 6. Overview absolute the amino acid percentage of IGLV6-57 assigned AL amyloidosis and multiple myeloma light chain sequences. MM = multiple myeloma patient, _H = AL amyloidosis patient with dominant heart involvement, _HK = AL amyloidosis patient with dominant heart and kidney involvement, _HTX = AL amyloidosis patient who received a heart transplant. For comparison, the median amino acid percentage of all detected IGLJ-IGLC combinations and the VBase2 IGLV6-57 reference sequence was calculated.

	Amino acid [%]																			
Sample	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
FOR150_HK	7.8	2.4	3.6	4.8	1.2	5.4	4.8	6.6	0.6	3.0	6.0	4.8	0.0	3.0	7.2	16.9	7.8	1.8	4.8	7.8
FOR152_HK	8.5	1.2	3.6	4.8	1.2	4.2	4.2	7.9	0.6	3.0	7.3	4.8	0.0	2.4	7.3	16.4	8.5	1.8	5.5	8.5
FOR153_HK	7.9	2.4	3.6	5.5	1.2	4.8	4.2	6.7	0.0	2.4	6.7	5.5	0.0	3.0	7.3	15.8	8.5	1.8	4.8	7.9
FOR154_HK	8.5	3.6	3.6	5.5	1.2	5.5	4.2	6.7	0.0	3.0	6.7	4.8	0.0	2.4	7.3	15.8	7.9	1.2	5.5	8.5
FOR188_HK	7.9	3.0	3.0	4.2	1.2	4.8	4.2	6.7	1.2	3.0	6.1	4.8	0.0	3.6	7.3	15.8	9.1	1.8	5.5	7.9
FOR185_HK	8.4	2.4	3.6	4.8	1.2	5.4	4.2	7.8	0.0	3.0	6.0	4.2	0.0	3.0	7.2	16.9	8.4	1.2	4.8	8.4
FOR222_HK	8.4	2.4	4.2	4.8	1.2	5.4	4.2	8.4	0.6	3.6	6.0	4.8	0.0	3.6	7.2	14.5	8.4	1.2	4.2	8.4
FOR228_HK	8.5	3.0	3.6	5.5	1.2	5.5	3.6	6.7	0.6	3.0	6.1	5.5	0.0	3.0	7.3	16.4	7.9	1.2	4.2	8.5
FOR117_HTX	8.5	3.0	4.2	5.5	1.2	4.8	4.8	6.1	0.0	3.6	6.1	4.8	0.0	2.4	7.3	15.8	8.5	1.2	5.5	8.5
FOR126_H	8.4	2.4	5.4	4.8	1.2	5.4	4.2	8.4	0.0	3.0	6.0	4.2	0.0	2.4	7.2	15.1	8.4	1.2	5.4	8.4
FOR140_HTX	9.6	3.0	2.4	5.4	1.2	5.4	4.2	6.0	0.0	3.0	6.0	3.6	0.0	3.6	7.2	17.5	8.4	1.8	4.8	9.6
FOR144_H	8.5	3.0	4.2	5.5	1.2	6.1	4.2	6.7	0.0	3.6	6.7	4.2	0.0	3.0	7.3	17.0	6.7	1.2	4.8	8.5
FOR192_H	8.5	3.0	3.6	4.8	1.2	4.8	4.2	6.7	0.0	3.0	6.1	5.5	0.0	3.0	7.3	15.8	8.5	1.2	4.8	8.5
FOR194_H	7.8	2.4	4.2	4.8	1.2	6.0	4.2	7.2	0.0	3.6	5.4	4.2	0.0	3.0	7.2	16.2	7.2	1.8	6.6	7.8
FOR197_HTX	8.4	3.0	3.6	6.0	1.2	6.0	3.6	6.0	0.0	3.6	6.0	4.2	0.0	3.6	7.2	16.3	8.4	1.2	4.2	8.4
FOR205_H	7.8	4.2	3.6	6.0	1.2	4.8	4.2	6.6	1.8	3.0	6.0	4.2	0.0	3.0	7.2	16.8	7.2	1.8	4.2	7.8
FOR214_HTX	8.5	2.4	3.6	5.5	1.2	6.7	4.2	6.7	0.6	4.2	6.1	4.2	0.0	3.6	7.3	17.0	6.7	1.2	3.6	8.5
MM112	7.8	3.6	4.8	5.4	1.2	5.4	4.2	5.4	0.0	3.0	6.6	5.4	0.0	2.4	7.2	15.1	8.4	1.8	5.4	7.8
Reference	**7.9**	**2.4**	**4.2**	**4.8**	**1.2**	**5.5**	**4.2**	**6.7**	**0.0**	**3.0**	**6.0**	**4.8**	**0.0**	**2.4**	**7.3**	**17.6**	**7.9**	**1.2**	**5.5**	**7.3**

Supplementary Information Table 7. Overview absolute the amino acid percentage of IGLV2-23 assigned multiple myeloma light chain sequences. MM = multiple myeloma patient. For a comparison, the median amino acid percentage of all detected IGLJ-IGLC combinations and the VBase2 IGLV2-23 reference sequence was calculated.

	Amino acid [%]																			
Sample	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
MM107	9.2	1.1	3.3	3.3	2.2	4.3	4.3	9.2	1.1	2.2	7.6	5.4	0.0	2.7	6.0	16.3	7.1	1.6	4.9	8.2
MM114	9.9	1.1	3.8	3.8	2.2	4.9	3.3	9.9	0.5	3.3	7.1	4.9	0.0	2.7	7.1	14.8	8.2	1.6	4.4	6.0
MM117	8.2	1.1	4.3	2.7	2.2	4.9	3.8	9.2	0.0	3.3	8.7	6.0	0.0	2.2	7.1	15.2	8.2	1.1	6.0	6.0
MM125	9.2	1.1	3.3	3.3	2.2	4.3	3.3	8.7	1.1	3.8	7.6	5.4	0.0	2.7	6.5	16.8	7.6	1.6	4.3	7.1
MM132	9.7	2.7	2.2	3.8	1.6	4.9	2.7	8.6	1.1	3.2	7.0	4.9	0.5	3.2	6.5	16.8	8.6	1.1	4.3	6.5
MM133	9.2	1.6	3.8	3.3	2.2	4.9	2.7	9.2	0.5	4.3	6.5	5.4	0.5	3.3	6.5	14.7	9.8	1.6	4.3	5.4
MM135	8.2	1.1	3.8	3.3	2.2	4.9	3.3	8.7	0.5	2.7	7.6	6.0	0.5	2.2	6.5	16.8	8.2	1.1	4.9	7.6
MM149	10.3	2.2	3.3	3.3	2.2	4.9	2.7	8.7	0.5	3.3	7.1	5.4	0.5	2.7	6.5	15.2	9.8	1.1	4.3	6.0
Reference	**9.3**	**1.1**	**3.3**	**2.7**	**2.2**	**4.9**	**3.3**	**9.3**	**0.5**	**2.7**	**7.1**	**6.0**	**0.5**	**2.2**	**6.6**	**17.5**	**7.7**	**1.4**	**4.9**	**6.9**

Appendix Results: Rare IGLV Subfamilies



Supplementary Information Figure 33. Sequence alignment of IGLV1-40 assigned AL amyloidosis and multiple myeloma amino acid light chain sequences. Bold = reference sequences, underlined = CDR regions, red letter = mutation, green letter = linker region, MM = multiple myeloma patient, _H = AL amyloidosis patient with dominant heart involvement.

Supplementary Information Figure 34. Sequence alignment of IGLV1-40 assigned AL amyloidosis and multiple myeloma cDNA light chain sequences. Bold = reference sequences, green letter = linker region, MM = multiple myeloma patient. _H = AL amyloidosis patient with dominant heart involvement. In the bioinformatic analysis of the bulk RNA sequencing, several sequence sections were given – indicated by _A and _B. Nucleotide signal overlaps in the Sanger sequencing were specified accoding to the IUPAC code.

```
MM104_NGS_A         ATGGCCTGGTCTCCTCTCCTCCTCACTC------------------------------
MM104_NGS_B         ----------------------------CTCGCACAGTTCCAGGGTCCTGGGCCCAGTCT
MM104_Sanger        ------------------------------------------------------TCAGTCT
                                                                           ******

MM104_NGS_B         GTGCTGACGCAGCCGCCCTCAGTGTCTGGGGCCCCAGGGCAGAGGGTCACCATCTCCTGC
MM104_Sanger        GTGCTGACGCAGCCGCCCTCAGTGTCTGGGGCCCCAGGGCAGAGGGTCACCATCTCCTGC
                    ************************************************************

MM104_NGS_B         ACTGGGAGCAGCTCCAACATCGGGGCAGGTTTTGATTTACACTGGTACCAGCAACTTCCA
MM104_Sanger        ACTGGGAGCAGCTCCAACATCGGGGCAGGTTTTGATTTACACTGGTACCAGCAACTTCCA
                    ************************************************************

MM104_NGS_B         GGAACAGCCCCCAAACTCCTCATCTATGGTAACACCAATCGGCCCTCAGGGGTCCCTGAC
MM104_Sanger        GGAACAGCCCCCAAACTCCTCATCTATGGTAACACCAATCGGCCCTCAGGGGTCCCTGAC
                    ************************************************************

MM104_NGS_B         CGATTCTCTGGCTCCAAGTCTGGCACCTCAGCCTCCCTGGTCATCACTGGGCTCCAGGCT
MM104_Sanger        CGATTCTCTGGCTCCAAGTCTGGCACCTCAGCCTCCCTGGTCATCACTGGGCTCCAGGCT
                    ************************************************************

MM104_NGS_B         GAGGATGAGGCTGATTATTACTGCCAGTCTTTTGACAGCAGCCTGAGTGGTGTGATATTC
MM104_Sanger        GAGGATGAGGCTGATTATTACTGCCAGTCTTTTGACAGCAGCCTGAGTGGTGTGATATTC
                    ************************************************************

MM104_NGS_B         GGCGGAGGGACCAAGCTGACCGTCCTAGGTCAGCCCAAGGCTGCCCCCTCGGTCACTCTG
MM104_Sanger        GGCGGAGGGACCAAGCTGACCGTCCTAGGTCAGCCCAAGGCTGCCCCCTCGGTCACTCTG
                    ************************************************************

MM104_NGS_B         TTCCCGCCCTCCTCTGAGGAGCTTCAAGCCAACAAGGCCACACTGGTGTGTCTCATAAGT
MM104_Sanger        TTCCCGCCCTCCTCTGAGGAGCTTCAAGCCAACAAGGCCACACTGGTGTGTCTCATAAGT
                    ************************************************************

MM104_NGS_B         GACTTCTACCCGGGAGCCGTGACAGTGGCCTGGAAGGCAGATAGCAGCCCCGTCAAGGCG
MM104_Sanger        GACTTCTACCCGGGAGCCGTGACAGTGGCCTGGAAGGCAGATAGCAGCCCCGTCAAGGCG
                    ************************************************************

MM104_NGS_B         GGAGTGGAGACCACCACACCCTCCAAACAAAGCAACAACAAGTACGCGGCCAGCAGCTAT
MM104_Sanger        GGAGTGGAGACCACCACACCCTCCAAACAAAGCAACAACAAGTACGCGGCCAGCAGCTAT
                    ************************************************************

MM104_NGS_B         CTGAGCCTGACGCCTGAGCAGTG-----------------------------------
MM104_Sanger        CTGAGCCTGACGCCTGAGCAGTGGAAGTCCCACAGAAGCTACAGCTGCCAGGTCACGCAT
                    ***********************

MM104_NGS_B         --------------
MM104_Sanger        GAAGGGAGCACCGT
```

Supplementary Information Figure 35. Sequence comparison between MM104 IGLV1-40 assigned light chain sequences generated through a next-generation sequencing and Sanger sequencing approach. For the sequence generated by the Sanger sequencing approach only dominant signals were used. * = nucleotide position consistent, MM = multiple myeloma patient, NGS = light chain sequence generated through next-generation sequencing, Sanger = light chain sequence generated through Sanger sequencing. Sequencing was performed using the VLKL12a_Huhn and CLKL_A_rv_NB oligonucleotides. In the bioinformatic analysis of the bulk RNA sequencing, several sequence sections were given – indicated by _A and _B.

Supplementary Information Figure 36. Sequence Alignment of IGLV1-44 assigned multiple myeloma light chain sequences with an additional sequence in the bioinformatic analysis of the bulk RNA sequencing approach. Sequence_A = sequence with the highest percentage >1 %, sequence_B = sequence with the second highest percentage >1 %. In the bioinformatic analysis of the bulk RNA sequencing, several sequence sections were given – indicated by _A, _B and _C.

```
MM116_NGS_sequenceB      CCTCACCCTCCTCACCCACCGGTCCTGGGCCCAGTCCGTGCTGACTCAGCCACCCTCAGC
MM116_NGS_sequenceA_A    ------------------------------------------------------------
MM116_NGS_sequenceA_B    ------------------------------------------------------------
MM116_Sanger             --------------------TGGTCCTGGGCTCAGTCTGTGCTGACTCAGCCACCCTCAGC
                                             ************************************

MM116_NGS_sequenceB      GTCTGGGACCCCCGGGCAGACGGTCACCATCTCTTGTTCTGGAGGCCCCTCCAACATCGG
MM116_NGS_sequenceA_A    ------------CGGGCAGACGGTCACCATCTCTTGTTCTGGAGGCCCCTCCAACATCGG
MM116_NGS_sequenceA_B    ------------------------------------------------------------
MM116_Sanger             GTCTGGGACCCCCGGGCAGACGGTCACCATCTCTTGTTCTGGAGGCCCCTCCAACATCGG
                                     **********************************************

MM116_NGS_sequenceB      AACTAATTCAGTAACCTGGTACCAGCAACTCCCAGGAGCGTCGCCCAAGGTCCTCATCTT
MM116_NGS_sequenceA_A    AACTAATTCAGTAACCTGGTACCAGCAACTCCCAGGAGCGTCGCCCAAGGTCCTCA----
MM116_NGS_sequenceA_B    ------------------------------------------------------------
MM116_Sanger             AACTAATTCAGTAACCTGGTACCAGCAACTCCCAGGAGCGTCGCCCAAGGTCCTCATCTT
                         ************************************************************

MM116_NGS_sequenceB      TCTCACTAGTCATCGGCCCTCAGGGGTCCCTGACCGATTCTCTGGCTCCAAGTCTGCCAC
MM116_NGS_sequenceA_B    -------------CGGCCCTCAGGGGTCCCTGACCGATTCTCTGGCTCCAAGTCTGCCAC
MM116_Sanger             TCTCAmTAGTCATCGGCCCTCAGGGGTCCCTGACCGATTCTCTGGCTCCAAGTCTGCCAC
                         ************************************************************

MM116_NGS_sequenceB      CTCAGCCTCCCTGGCCATCAGTGGTCTCCAGTCTGAGGATGAGGCTGATTATTACTGTGC
MM116_NGS_sequenceA_B    CTCAGCCTCCCTGGCCATCAGTGGTCTCCAGTCTGAGGATGAGGCTGATTATTACTGTGC
MM116_Sanger             CTCAGCCTCCCTGGCCTGsTCykGksyTCAGTCTGAGGATGAGGCTGATTATTACTGTGC
                         ***************  *  ****** ********************************

MM116_NGS_sequenceB      AGCATGGGATGACAGCCTGACTGGTCCGGTGTTCGGCGGAGGGACCAAGTTGACCGTCCT
MM116_NGS_sequenceA_B    AGCATGGGATGACAGCCTGACTGGTCCGGTGTTCGGCGGAGGGACCAAGTTGACCGTCCT
MM116_Sanger             AGCATGGGATGACAGCCTGACTGGTCCGGTGTTCGGCGGAGGGACCAAGTTGACCGTCCT
                         ************************************************************

MM116_NGS_sequenceB      AGGTCAGCCCAAGGCTGCCCCCTCGGTCACTCTGTTCCCACCCTCCTCTGAGGAGCTTCA
MM116_NGS_sequenceA_B    AGGTCAGCCCAAGGCTGCCCCCTCGGTCACTCTGTTCCCACCCTCCTCTGAGGAGCTTCA
MM116_Sanger             AGGTCAGCCCAAGGCTGCCCCCTCGGTCACTCTGTTCCCACCCTCCTCTGAGGAGCTTCA
                         ************************************************************

MM116_NGS_sequenceB      AGCCAACAAGGCCACACTGGTGTGTCTCATAAGTGACTTCTACCCGGGAGCCGTGACAGT
MM116_NGS_sequenceA_B    AGCCAACAAGGCCACACTGGTGTGTCTCATAAGTGACTTCTACCCGGGAGCCGTGACAGT
MM116_Sanger             AGCCAACAAGGCCACACTGGTGTGTCTCATAAGTGACTTCTACCCGGGAGCCGTGACAGT
                         ************************************************************

MM116_NGS_sequenceB      GGCCTGGAAGGCAGATAGCAGCCCCGTCAAGGCGGGAGTGGAGACCAC------------
MM116_NGS_sequenceA_B    GGCCTGGAAGGCAGATAGCAGCCCCGTCAAGGCGGGAGTGGAGACCACCACACCCTCCAA
MM116_Sanger             GGCCTGGAAGGCAGATAGCAGCCCCGTCAAGGCGGGAGTGGAGACCACCACACCCTCCAA
                         ***********************************************

MM116_NGS_sequenceB      ------------------------------------------------------------
MM116_NGS_sequenceA_B    ACAAAGCAACAACAAGTACGCGGCCAGCAGCTACCTGAGCCTGACGCCTGAGCAGTG---
MM116_Sanger             ACAAAGCAACAACAAGTACGCGGCCAGCAGCTACCTGAGCCTGACGCCTGAGCAGTGGAA

MM116_NGS_sequenceB      ------------------
MM116_NGS_sequenceA_B    ------------------
MM116_Sanger             GTCCCACAAAAGCTACAG
```

Supplementary Information Figure 37. Sequence comparison between MM116 IGLV1-44 assigned light chain sequences generated through a next-generation sequencing and Sanger sequencing approach. For the sequence generated by the Sanger sequencing approach only dominant signals were used. * = nucleotide position consistent, MM = multiple myeloma patient, NGS = light chain sequence generated through next-generation sequencing, Sanger = light chain sequence generated through a Sanger sequencing approach. Sequence_A = sequence with the highest percentage >1 %, sequence_B = sequence with the second highest percentage >1 %. In the bioinformatic analysis of the bulk RNA sequencing, several sequence sections were given – indicated by _A, _B and _C. Sequencing was performed using the VLKL12a_Huhn and CLKL_A_rv_NB oligonucleotides. Nucleotide signal overlaps in the Sanger sequencing were specified accoding to the IUPAC code.

```
MM119_NGS_sequenceB_A     ATGGCCAGCTTCCCTCTCCTCCTCACCCTCCTCAC-------------------------
MM119_NGS_sequenceB_B     ----------------------------------GTCCTGGGCCCAGTCTGTCCTGACT
MM119_NGS_sequenceA       ------------------------------------------------------------
MM119_Sanger              -----------------------------------------CAGTCTGTCCTGACT


MM119_NGS_sequenceB_B     CAGCCACCCTCAGCGTCTGGGACCCCCGGGCAGAGGGTCATCATTTCTTGTTCTGGAACC
MM119_NGS_sequenceA       ------------------------------------------------------------
MM119_Sanger              CAGCCACCCTCAGCGTCTGGGACCCCCGGGCAGAGGGTCATCATTTCTTGTTCTGGAACC


MM119_NGS_sequenceB_B     AGCTCCAACATCGAAAGTTATCCTGTAAATTGGTATCAGCAGCTCCCGGGATCCGCCCCC
MM119_NGS_sequenceA       ------------------------------------------------------------
MM119_Sanger              AGCTCCAACATCGAAAGTTATCCTGTAAATTGGTATCAGCAGCTCCCGGGATCCGCCCCC


MM119_NGS_sequenceB_B     AAACTCCTCATCTATGAAACTAATCAGCGGCCCTCAGGGGTCCCTGACCGATTCTCTGGC
MM119_NGS_sequenceA       ------------------------------CCCTCAGGGGTCCCTGACCGATTCTCTGGC
MM119_Sanger              AAACTCCTCATCTATGAAACTAATCAGCGGCCCTCAGGGGTCCCTGACCGATTCTCTGGC
                                                        ******************************


MM119_NGS_sequenceB_B     TCCAAGTCTGGCAGCTCAGCCTCCCTGGCCATCACTGGTCTCCAGTCTGAAGATGAGTCT
MM119_NGS_sequenceA       TCCAAGTCTGGCAGCTCAGCCTCCCTGGCCATCACTGGTCTCCAGTCTGAAGATGAGTCT
MM119_Sanger              TCCAAGTCTGGCAGCTCAGCCTCCCTGGCCATCACTGGTCTCCAGTCTGAAGATGAGTCT
                          ************************************************************


MM119_NGS_sequenceB_B     GATTATTACTGTGCTTCATGGGATGAAACCCTGAATGGTCCGGTATTCGGCGGAGGGACC
MM119_NGS_sequenceA       GATTATTACTGTGCTTCATGGGATGAAACCCTGAATGGTCCGGTATTCGGCGGAGGGACC
MM119_Sanger              GATTATTACTGTGCTTCATGGGATGAAACCCTGAATGGTCCGGTATTCGGCGGAGGGACC
                          ************************************************************


MM119_NGS_sequenceB_B     AAGCTGACCGTCCTGGGTCAGCCCAAGGCTGCCCCCTCGGTCACTCTGTTCCCGCCCTCC
MM119_NGS_sequenceA       AAGCTGACCGTCCTGGGTCAGCCCAAGGCTGCCCCCTCGGTCACTCTGTTCCCGCCCTCC
MM119_Sanger              AAGCTGACCGTCCTGGGTCAGCCCAAGGCTGCCCCCTCGGTCACTCTGTTCCCGCCCTCC
                          ************************************************************


MM119_NGS_sequenceB_B     TCTGAGGAGCTTCAAGCCAACAAGGCCACACTGGTGTGTCTCATAAGTGACTTCTACCCG
MM119_NGS_sequenceA       TCTGAGGAGCTTCAAGCCAACAAGGCCACACTGGTGTGTCTCATAAGTGACTTCTACCCG
MM119_Sanger              TCTGAGGAGCTTCAAGCCAACAAGGCCACACTGGTGTGTCTCATAAGTGACTTCTACCCG
                          ************************************************************


MM119_NGS_sequenceB_B     GGAGCCGTGACAGTGGCCTGGAAGGCAGATAGCAGCCCCGTCAAGGCGGGAGTGGAGACC
MM119_NGS_sequenceA       GGAGCCGTGACAGTGGCCTGGAAGGCAGATAGCAGCCCCGTCAAGGCGGGAGTGGAGACC
MM119_Sanger              GGAGCCGTGACAGTGGCCTGGAAGGCAGATAGCAGCCCCGTCAAGGCGGGAGTGGAGACC
                          ************************************************************


MM119_NGS_sequenceB_B     ------------------------------------------------------------
MM119_NGS_sequenceA       ACCACACCCTCCAAACAAAGCAACAACAAGTACGCGGCCAGCAGCTACCTGAGCCTGACG
MM119_Sanger              ACCACACCCTCCAAACAAAGCAACAACAAGTACGCGGCCAGCAGCTACCTGAGCCTGACG


MM119_NGS_sequenceB_B     ------------------------------------------------------------
MM119_NGS_sequenceA       CCTGAGCAGTG-------------------------------------------------
MM119_Sanger              CCTGAGCAGTGGAAGTCCCACAGAAGCTACAGCTGCCAGGTCACGCATGAAGGGAGCACC


MM119_NGS_sequenceB_B     ----
MM119_NGS_sequenceA       ----
MM119_Sanger              GTGG
```

Supplementary Information Figure 38. Sequence comparison between MM119 IGLV1-44 assigned light chain sequences generated through a next-generation sequencing and Sanger sequencing approach. For the sequence generated by the Sanger sequencing approach only dominant signals were used. * = nucleotide position consistent, MM = multiple myeloma patient, NGS = light chain sequence generated through next-generation sequencing, Sanger = light chain sequence generated through a Sanger sequencing approach. Sequence_A = sequence with the highest percentage >1 %, sequence_B = sequence with the second highest percentage >1 %. In the bioinformatic analysis of the bulk RNA sequencing, several sequence sections were given – indicated by _A, _B and _C. Sequencing was performed using the VLKL12a_Huhn and CLKL_A_rv_NB oligonucleotides.

```
MM137_NGS_sequenceA       ------------------------------------------------------------
MM137_NGS_sequenceB_A     ATGGCCAGCTTCCCTCTCCTCCTCAC-------------------------------
MM137_NGS_sequenceB_B     -------------------------GTCCTGGGCCCAGTCTGTGCTGACTCAGCCACCT
MM137_Sanger              ------------------------------------AGTCTGTGCTGACTCAGCCACCT
                                                              **********************

MM137_NGS_sequenceA       ------------------------------------------------------------
MM137_NGS_sequenceB_B     TCAGCGTCTGGGACCCCCGGGCAGAGGGTCACCATCTCTTGTTCTGGCAGCAGGTCCAAC
MM137_Sanger              TCAGCGTCTGGGACCCCCGGGCAGAGGGTCACCATCTCTTGTTCTGGCAGCAGGTCCAAC
                          ************************************************************

MM137_NGS_sequenceA       ------------------------------------------------------------
MM137_NGS_sequenceB_B     GTCGGAAGTAACACTGTAAGTTGGTACCAGCACCTCCCGGGAACGGCCCCCAAACTCCTC
MM137_Sanger              GTCGGAAGTAACACTGTAAGTTGGTACCAGCACCTCCCGGGAACGGCCsCCAAACTCCTC
                          ************************************************************

MM137_NGS_sequenceA       -------------------------AGGGGTCCCTGACCGATTCTCCGGCTCCAAGTCT
MM137_NGS_sequenceB_B     ATCTATACTAATAATCAGCGGCCCTCAGGGGTCCCTGACCGATTCTCCGGCTCCAAGTCT
MM137_Sanger              ATCTATACTAATAATCAGCGGCCCTCAGGGGTCCCTGACCGATTCTCCGGCTCCAAGTCT
                          ************************************************************

MM137_NGS_sequenceA       GGCACCTCAGCCTCTCTGGCCATCAGTGGGCTCCGGTCTGAGGATGAGGCTGACTATTAC
MM137_NGS_sequenceB_B     GGCACCTCAGCCTCTCTGGCCATCAGTGGGCTCCGGTCTGAGGATGAGGCTGACTATTAC
MM137_Sanger              GGCACCTCAGCCTCTCTGGCCATCAGTGGGCTCCGGTCTGAGGATGAGGCTGACTATTAC
                          ************************************************************

MM137_NGS_sequenceA       TGTGCAGCATGGGATGACGGCCTGAATGATCCCGTTTTCGGCGAAGGGACCAAGCTGACC
MM137_NGS_sequenceB_B     TGTGCAGCATGGGATGACGGCCTGAATGATCCCGTTTTCGGCGAAGGGACCAAGCTGACC
MM137_Sanger              TGTGCAGCATGGGATGACGGCCTGAATGATCCCGTTTTCGGCGAAGGGACCAAGCTGACC
                          ************************************************************

MM137_NGS_sequenceA       GTCCTAGGTCAGCCCAAGGCTGCCCCCTCGGTCACTCTGTTCCCGCCCTCCTCTGAGGAG
MM137_NGS_sequenceB_B     GTCCTAGGTCAGCCCAAGGCTGCCCCCTCGGTCACTCTGTTCCCGCCCTCCTCTGAGGAG
MM137_Sanger              GTCCTAGGTCAGCCCAAGGCTGCCCCCTCGGTCACTCTGTTCCCGCCCTCCTCTGAGGAG
                          ************************************************************

MM137_NGS_sequenceA       CTTCAAGCCAACAAGGCCACACTGGTGTGTCTCATAAGTGACTTCTACCCGGGAGCCGTG
MM137_NGS_sequenceB_B     CTTCAAGCCAACAAGGCCACACTGGTGTGTCTCATAAGTGACTTCTACCCGGGAGCCGTG
MM137_Sanger              CTTCAAGCCAACAAGGCCACACTGGTGTGTCTCATAAGTGACTTCTACCCGGGAGCCGTG
                          ************************************************************

MM137_NGS_sequenceA       ACAGTGGCCTGGAAGGCAGATAGCAGCCCCGTCAAGGCGGGAGTGGAGACCACCACACCC
MM137_NGS_sequenceB_B     ACAGTGGCCTGGAAGGCAGATAGCAGCCCCGTCAAGGCGGGAGTGGAGACCAC-------
MM137_Sanger              ACAGTGGCCTGGAAGGCAGATAGCAGCCCCGTCAAGGCGGGAkTGGAGACCACCACACCC
                          ************************************************************

MM137_NGS_sequenceA       TCCAAACAAAGCAACAACAAGTACGCGGCCAGCAGCTACCTGAGCCTGACGCCTGAGCAG
MM137_NGS_sequenceB_B     ------------------------------------------------------------
MM137_Sanger              TCCAAACAAAGCAACAACAAGTACGCGGCCAGCAGCTACCTGAGCCTGACGCCTGAGCAG
                          ************************************************************

MM137_NGS_sequenceA       TGGAAGTCC---------------------------------------------
MM137_NGS_sequenceB_B     ----------------------------------------------------
MM137_Sanger              TGGAAGTCCCACAGAAGCTACAGCTGCCAGGTCACGCATGAAGGGAGCACCGTGG
                          *********
```

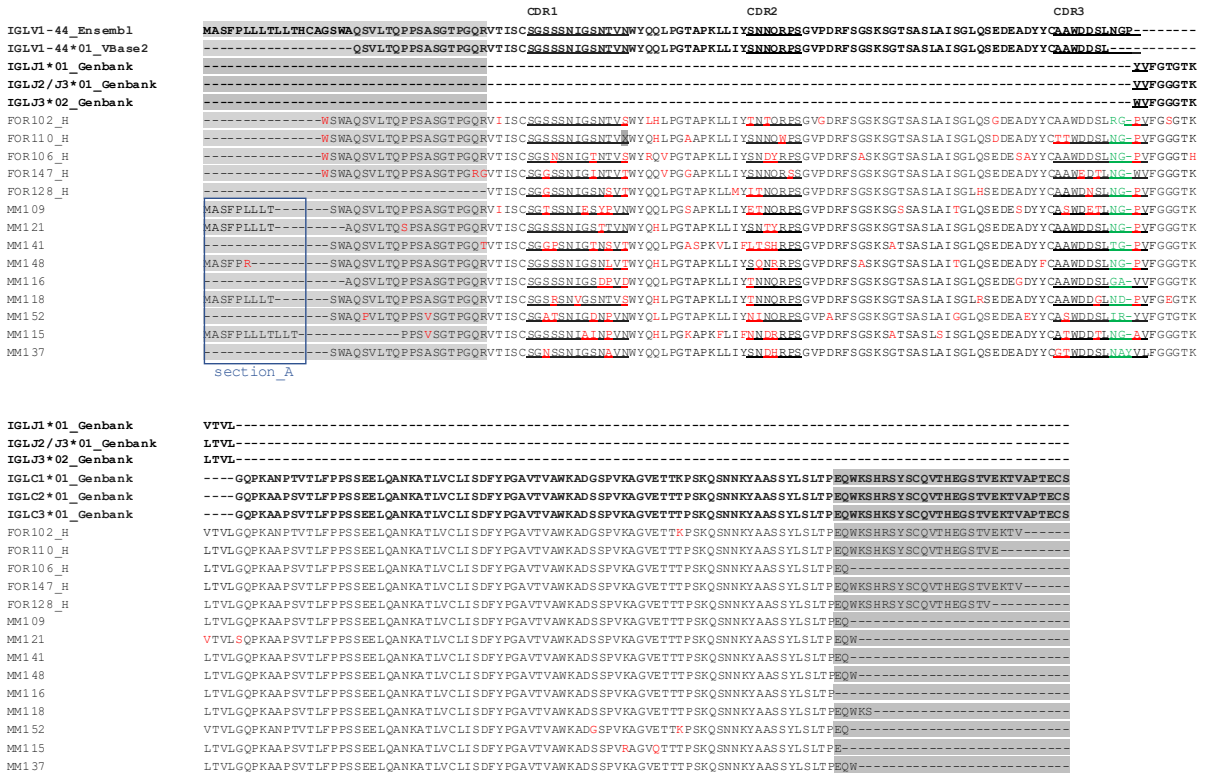
Supplementary Information Figure 39. Sequence comparison between MM137 IGLV1-44 assigned light chain sequences generated through a next-generation sequencing and Sanger sequencing approach. For the sequence generated by the Sanger sequencing approach only dominant signals were used. * = nucleotide position consistent, MM = multiple myeloma patient, NGS = light chain sequence generated through next-generation sequencing, Sanger = light chain sequence generated through a Sanger sequencing approach. Sequence_A = sequence with the highest percentage >1 %, sequence_B = sequence with the second highest percentage >1 %. In the bioinformatic analysis of the bulk RNA sequencing, several sequence sections were given – indicated by _A, _B and _C. Sequencing was performed using the VLKL12a_Huhn and CLKL_A_rv_NB oligonucleotides. Nucleotide signal overlaps in the Sanger sequencing were specified accoding to the IUPAC code.

```
MM141_NGS_sequenceB_A  ATGGCCAGCTTCCCTCTCCTCCTCATGGCCAGCTTCCCTCTCCTCCTC------------
MM141_NGS_sequenceB_B  ----------------------------------------------GGTCCTGGGCCC
MM141_NGS_sequenceA_A  ------------------------------------------------------------
MM141_NGS_sequenceA_B  ------------------------------------------------------------
MM141_Sanger           ------------------------------------------------------------

MM141_NGS_sequenceB_B  AGCCTGTGCTGACTCAGCCACCCTCAGTGTCTGGGACCCCCGGCCAGAGGGTCACCATCT
MM141_NGS_sequenceA_A  ----------------------------------------GCCAGAGGGTCACCATCT
MM141_Sanger           AGTCTGTGCTGACTCAGCCACCCTCAGTGTCTGGGACCCCCGGCCAGAGGGTCACCATCT
                                                              ******************

MM141_NGS_sequenceB_B  CTTGTTCTGGAGCCACTTCCAACATCGGCGACAACCCTGTTAACTGGTACCAGCTGCTCC
MM141_NGS_sequenceA_A  CTTGTTCTGGAGCCACTTCCAACATCGGCGACAACCCTGTTAACTGGTACCAGCTGCTCC
MM141_NGS_sequenceA_B  ------------------------------------------------------------
MM141_Sanger           CTTGTTCTGGAGCCACTTCCAACATCGGCGACAACCCTGTTAACTGGTACCAGCTGCTCC
                       ************************************************************

MM141_NGS_sequenceB_B  CAGGGACGGCCCCCAAGTTACTCATCTATAATATTAATCAGCGGCCCTCAGGGGTCCCTG
MM141_NGS_sequenceA_A  CAGGGAAGGCCCCCAA--------------------------------------------
MM141_NGS_sequenceA_B  ----------------------------------------------CTCAGGGGTCCCTG
MM141_Sanger           CAGGGACGGCCCCCAAGTTACTCATCTATAATATTAATCAGCGGCCCTCAGGGGTCCCTG
                       ****** **************************************************

MM141_NGS_sequenceB_B  CCCGATTCTCTGGCTCCAAGTCTGGCACCTCAGCCTCCCTGGCCATCGGTGGGCTCCAGT
MM141_NGS_sequenceA_B  CCCGATTCTCTGGCTCCAAGTCTGGCACCTCAGCCTCCCTGGCCATCGGTGGGCTCCAGT
MM141_Sanger           CCCGATTCTCTGGCTCCAAGTCTGGCACCTCAGCCTCCCTGGCCATCGGTGGGCTCCAGT
                       ************************************************************

MM141_NGS_sequenceB_B  CTGAGGATGAGGCTGAGTATTACTGTGCATCATGGGATGACAGCCTGATTCGTTATGTCT
MM141_NGS_sequenceA_B  CTGAGGATGAGGCTGAGTATTACTGTGCATCATGGGATGACAGCCTGATTCGTTATGTCT
MM141_Sanger           CTGAGGATGAGGCTGAGTATTACTGTGCATCATGGGATGACAGCCTGATTCGTTATGTCT
                       ************************************************************

MM141_NGS_sequenceB_B  TCGGAACTGGGACCAAGGTCACCGTCCTAGGTCAGCCCAAGGCCAACCCCACTGTCACTC
MM141_NGS_sequenceA_B  TCGGAACTGGGACCAAGGTCACCGTCCTAGGTCAGCCCAAGGCCAACCCCACTGTCACTC
MM141_Sanger           TCGGAACTGGGACCAAGGTCACCGTCCTAGGTCAGCCCAAGGCCAACCCCACTGTCACTC
                       ************************************************************

MM141_NGS_sequenceB_B  TGTTCCCGCCCTCCTCTGAGGAGCTCCAAGCCAACAAGGCCACACTAGTGTGTCTGATCA
MM141_NGS_sequenceA_B  TGTTCCCGCCCTCCTCTGAGGAGCTCCAAGCCAACAAGGCCACACTAGTGTGTCTGATCA
MM141_Sanger           TGTTCCCGCCCTCCTCTGAGGAGCTCCAAGCCAACAAGGCCACACTAGTGTGTCTGATCA
                       ************************************************************

MM141_NGS_sequenceB_B  GTGACTTCTACCCGGGAGCTGTGACAGTGGCCTGGAAGGCAGATGGCAGCCCCGTCAAGG
MM141_NGS_sequenceA_B  GTGACTTCTACCCGGGAGCTGTGACAGTGGCCTGGAAGGCAGATGGCAGCCCCGTCAAGG
MM141_Sanger           GTGACTTCTACCCGGGAGCTGTGACAGTGGCCTGGAAGGCAGATGGCAGCCCCGTCAAGG
                       ************************************************************

MM141_NGS_sequenceB_B  CGGG------------------------------------------------------
MM141_NGS_sequenceA_B  CGGGAGTGGAGACCACCAAACCCTCCAAACAGAGCAACAACAAGTACGCGGCCAGCAGCT
MM141_Sanger           CGGGAGTGGAGACCACCAAACCCTCCAAACAGAGCAACAACAAGTACGCGGCCAGCAGCT
                       ************************************************************

MM141_NGS_sequenceB_B  ------------------------------------------------------------
MM141_NGS_sequenceA_B  ACCTGAGCCTGACGCCCGAGCAGT-----------------------------------
MM141_Sanger           ACCTGAGCCTGACGCCCGAGCAGTGGAAGTCCCACAGAAGCTACAGCTGCCAGGTCACGC
                       ************************

MM141_NGS_sequenceB_B  ----------------
MM141_NGS_sequenceA_B  ----------------
MM141_Sanger           ATGAAGGGAGCACCGT
```

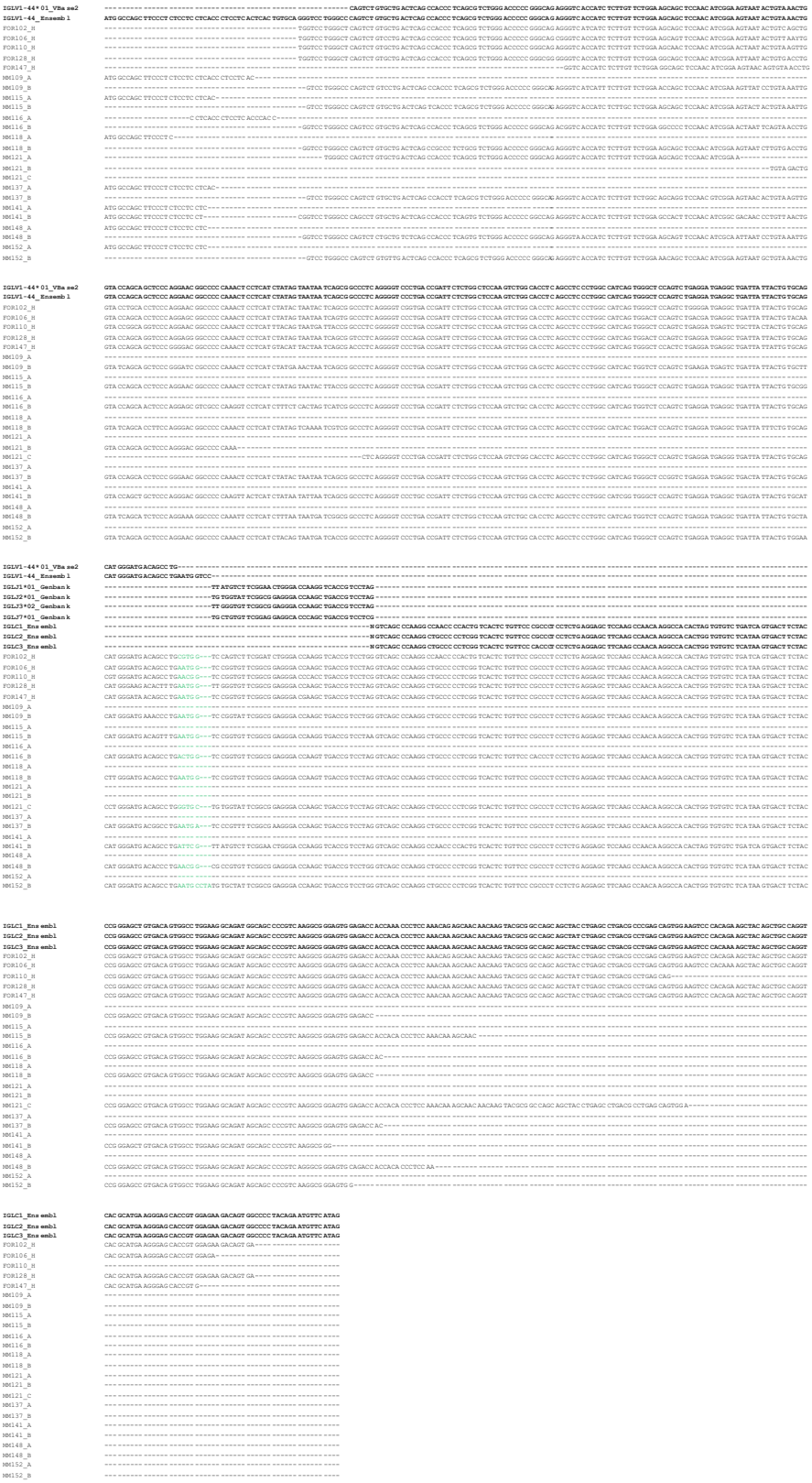
Supplementary Information Figure 40. Sequence comparison between MM141 IGLV1-44 assigned light chain sequences generated through a next-generation sequencing and Sanger sequencing approach. For the sequence generated by the Sanger sequencing approach only dominant signals were used. * = nucleotide position consistent, MM = multiple myeloma patient, NGS = light chain sequence generated through next-generation sequencing, Sanger = light chain sequence generated through a Sanger sequencing approach. Sequence_A = sequence with the highest percentage >1 %, sequence_B = sequence with the second highest percentage >1 %. In the bioinformatic analysis of the bulk RNA sequencing, several sequence sections were given – indicated by _A, _B and _C. Sequencing was performed using the VLKL12a_Huhn and CLKL_A_rv_NB oligonucleotides.

Supplementary Information Table 8. Mutation and light chain composition analysis of IGLV1-44 assigned AL amyloidosis and multiple myeloma light chain sequences. MM = multiple myeloma patient, _H = AL amyloidosis patient with dominant heart involvement.

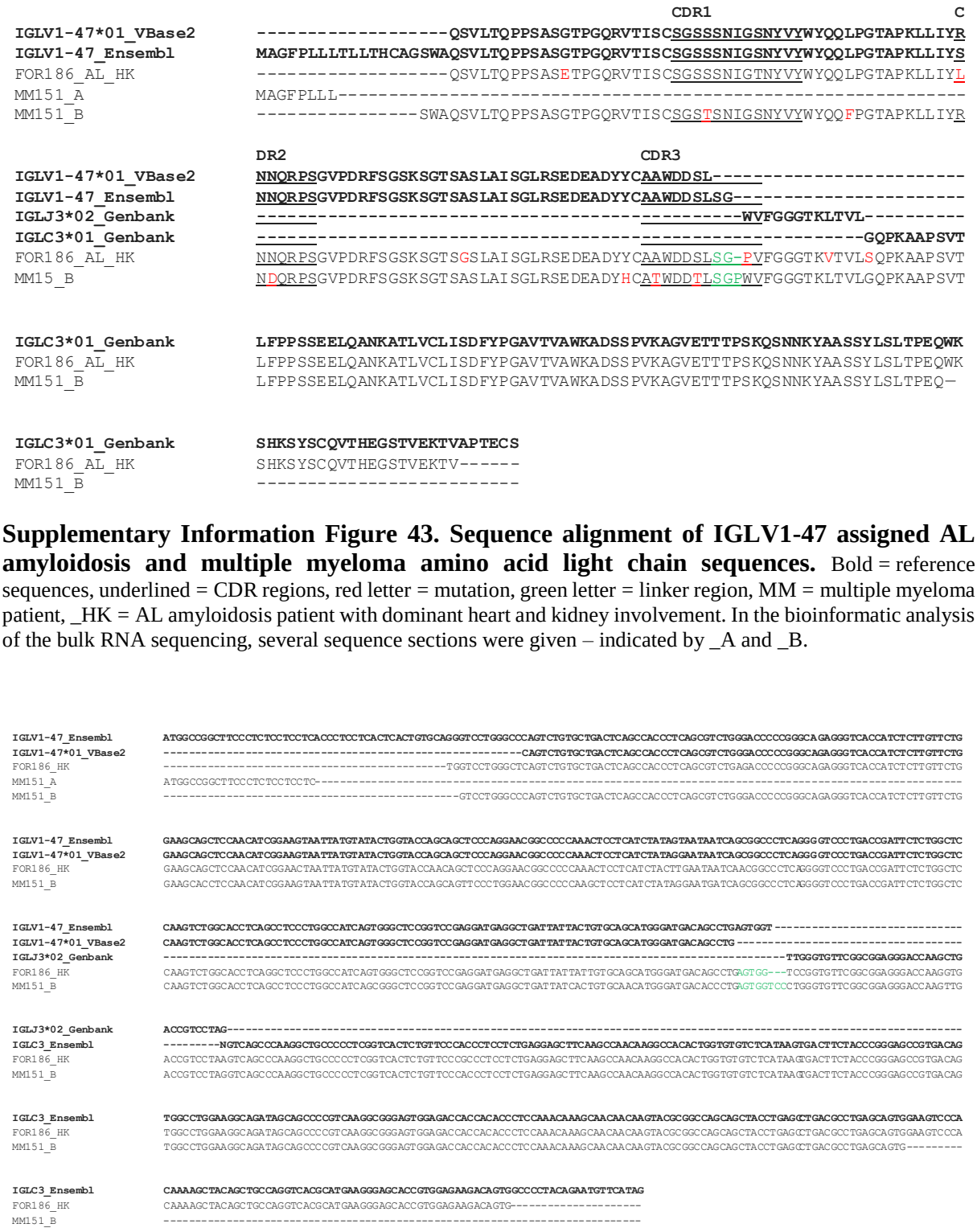
	FOR102_H	FOR106_H	FOR110_H	FOR128_H	FOR147_H	MM109	MM115	MM116	MM118	MM121	MM137	MM141	MM148	MM152
mutations IGLV segment	9	7	11	8	9	15	5	15	9	5	8	11	15	6
mutations IGLJ segment	2	1	2	0	1	1	2	1	1	0	2	0	1	0
mutations IGLC segment	1	1	0	0	0	0	1	0	0	0	0	1	1	0
IGLJ family	1	3	7/3	3	3	2	3	3	3	2	2/3	1	2/3	2
IGLC family	1	3	2	2	2	2/3	2/3	3	2	2/3	2/3	1	3	2



Supplementary Information Figure 41. Sequence alignment of IGLV1-44 assigned AL amyloidosis and multiple myeloma amino acid light chain sequences. Bold = reference sequences, underlined = CDR regions, red letter = mutation, green letter = linker region, MM = multiple myeloma patient, _H = AL amyloidosis patient with dominant heart involvement. In the bioinformatic analysis of the bulk RNA sequencing, several sequence sections were given – section A is framed. The sequence with the second highest percentage in the overall analysis was used for alignment and analysis of the IGLV1-44 multiple myeloma sequences (sequence_B).

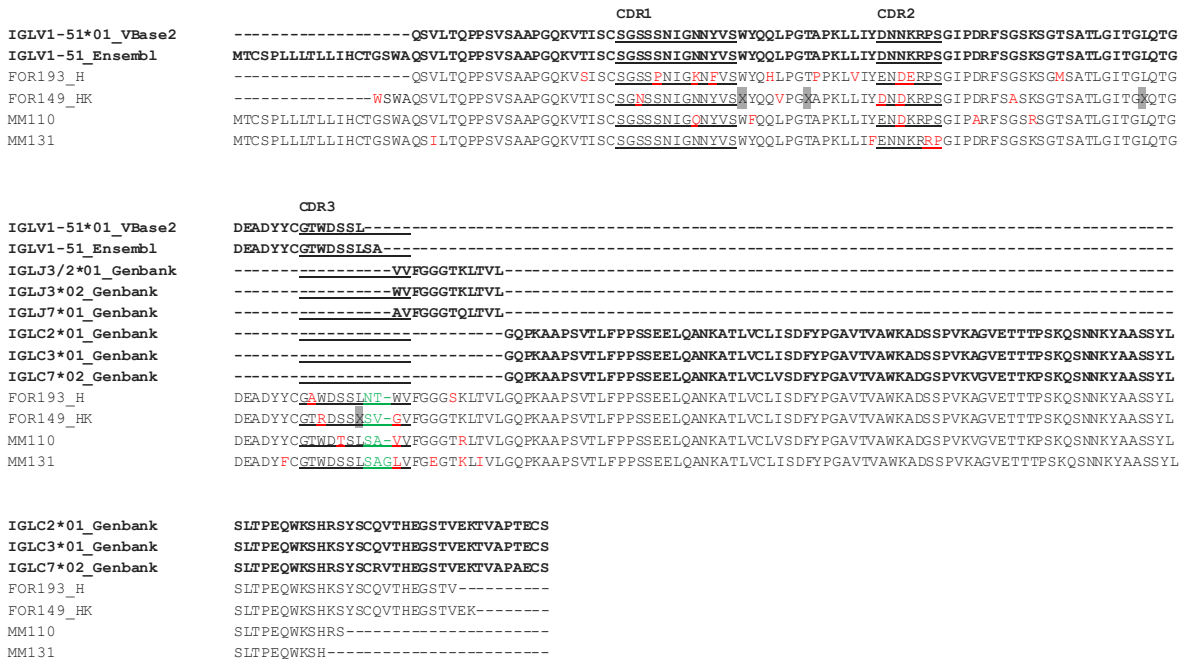


Supplementary Information Figure 42. Sequence alignment of IGLV1-44 assigned AL amyloidosis and multiple myeloma cDNA light chain sequences. Bold = reference sequences, green letter = linker region, MM = multiple myeloma patient, _H = AL amyloidosis patient with dominant heart involvement. In the bioinformatic analysis of the bulk RNA sequencing, several sequence sections were given – indicated by _A, _B, and _C. Nucleotide signal overlaps in the Sanger sequencing were specified according to the IUPAC code. The sequence with the second highest percentage in the overall analysis was used for alignment und analysis of the IGLV1-44 multiple myeloma sequences (sequence_B).

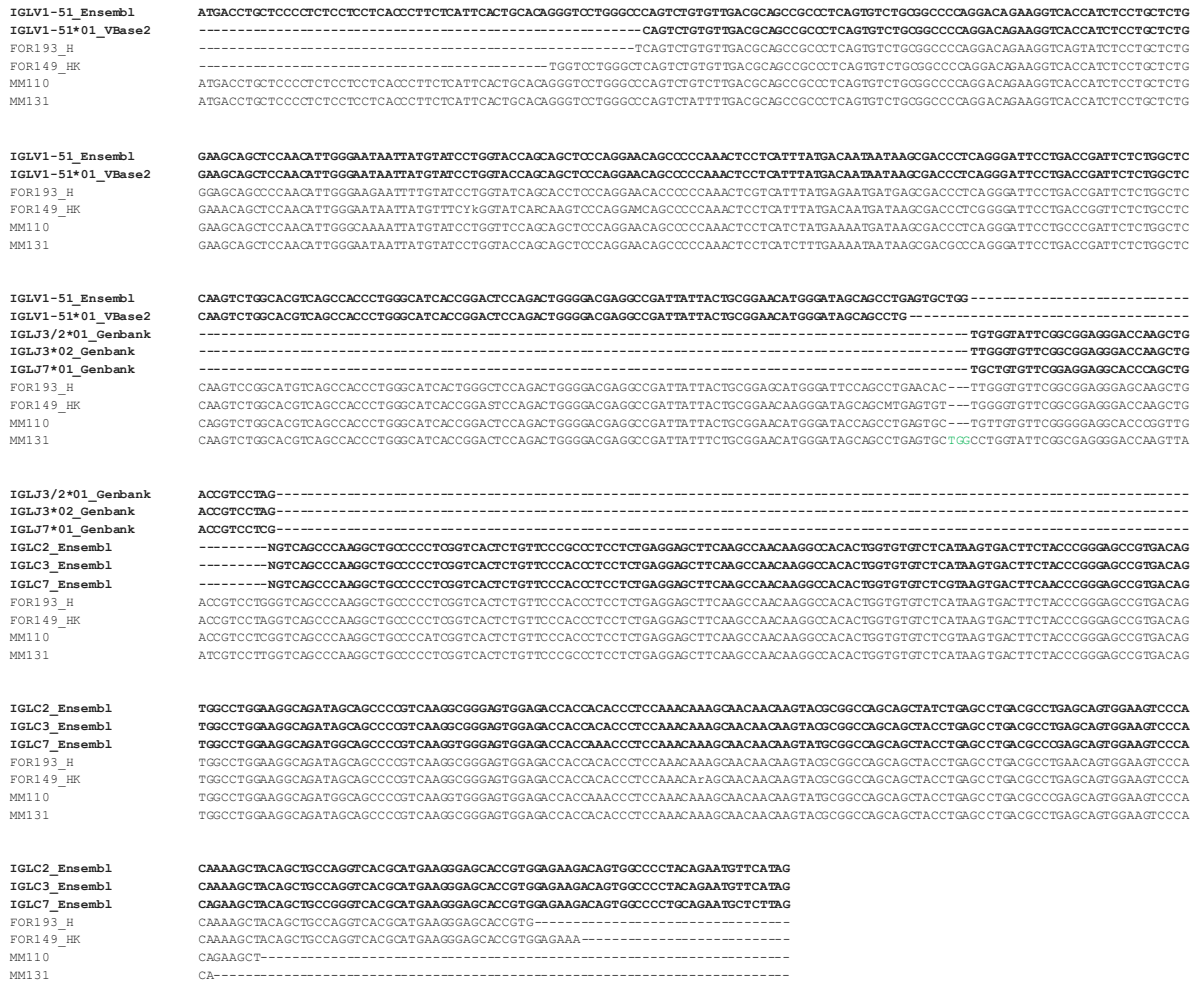


Supplementary Information Figure 43. Sequence alignment of IGLV1-47 assigned AL amyloidosis and multiple myeloma amino acid light chain sequences. Bold = reference sequences, underlined = CDR regions, red letter = mutation, green letter = linker region, MM = multiple myeloma patient, _HK = AL amyloidosis patient with dominant heart and kidney involvement. In the bioinformatic analysis of the bulk RNA sequencing, several sequence sections were given – indicated by _A and _B.

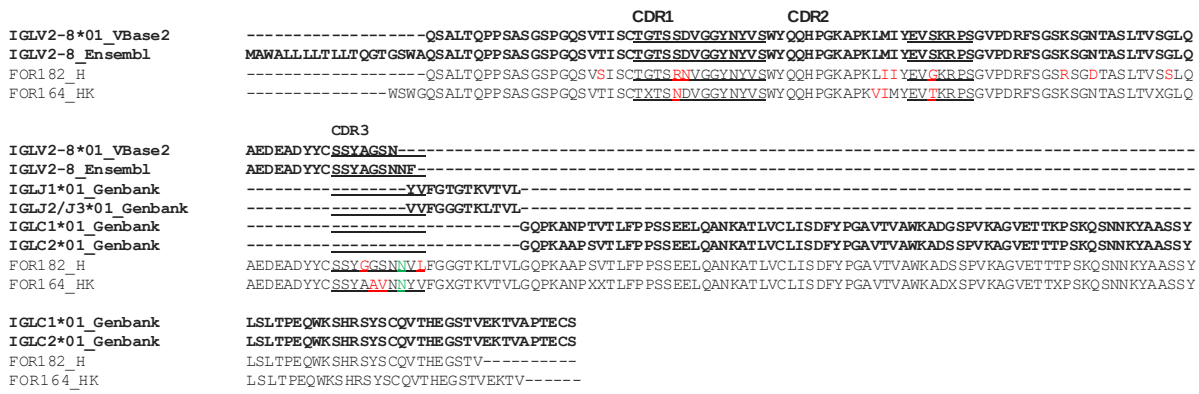
Supplementary Information Figure 44. Sequence alignment of IGLV1-47 assigned AL amyloidosis and multiple myeloma cDNA light chain sequences. Bold = reference sequences, green letter = linker region, MM = multiple myeloma patient, _HK = AL amyloidosis patient with dominant heart and kidney involvement. In the bioinformatic analysis of the bulk RNA sequencing, several sequence sections were given – indicated by _A and _B. Nucleotide signal overlaps in the Sanger sequencing were specified accoding to the IUPAC code.



Supplementary Information Figure 45. Sequence alignment of IGLV1-51 assigned AL amyloidosis and multiple myeloma amino acid light chain sequences. Bold = reference sequences, underlined = CDR regions, red letter = mutation, green letter = linker region, MM = multiple myeloma patient, _HK = AL amyloidosis patient with dominant heart and kidney involvement, _H = AL amyloidosis patient with dominant heart involvement, X and grey highlight = not unambiguously determined amino acid.



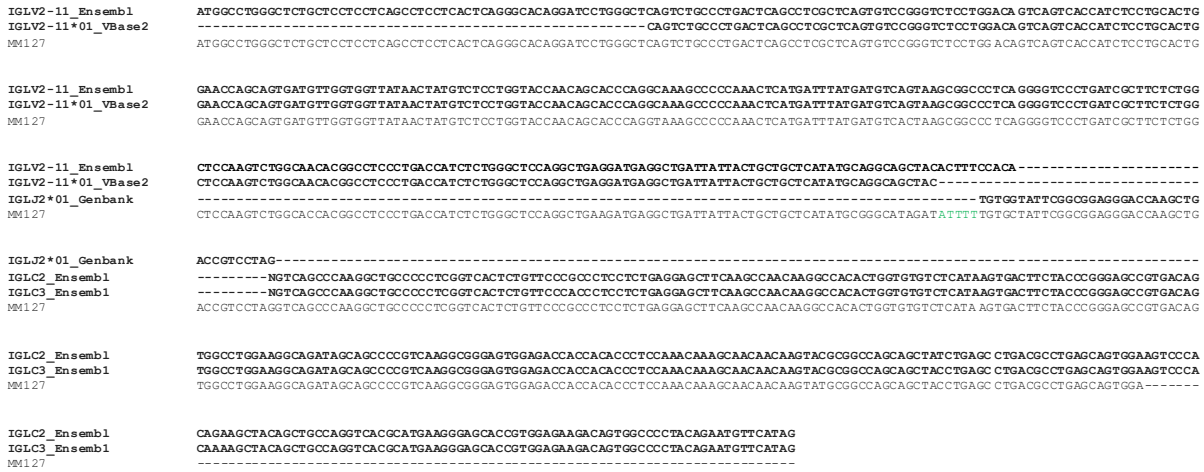
Supplementary Information Figure 46. Sequence alignment of IGLV1-51 assigned AL amyloidosis and multiple myeloma cDNA light chain sequences. Bold = reference sequences, green letter = linker region, MM = multiple myeloma patient, _HK = AL amyloidosis patient with dominant heart and kidney involvement, _H = AL amyloidosis patient with dominant heart involvement Nucleotide signal overlaps in the Sanger sequencing were specified accoding to the IUPAC code.



Supplementary Information Figure 47. Sequence alignment of IGLV2-8 assigned AL amyloidosis amino acid light chain sequences. Bold = reference sequences, underlined = CDR regions, red letter = mutation, green letter = linker region, _HK = AL amyloidosis patient with dominant heart and kidney involvement, _H = AL amyloidosis patient with dominant heart involvement.

Supplementary Information Figure 48. Sequence alignment of IGLV2-8 assigned AL amyloidosis cDNA light chain sequences. Bold = reference sequences, green letter = linker region, _HK = AL amyloidosis patient with dominant heart and kidney involvement, _H = AL amyloidosis patient with dominant heart involvement Nucleotide signal overlaps in the Sanger sequencing were specified accoding to the IUPAC code.

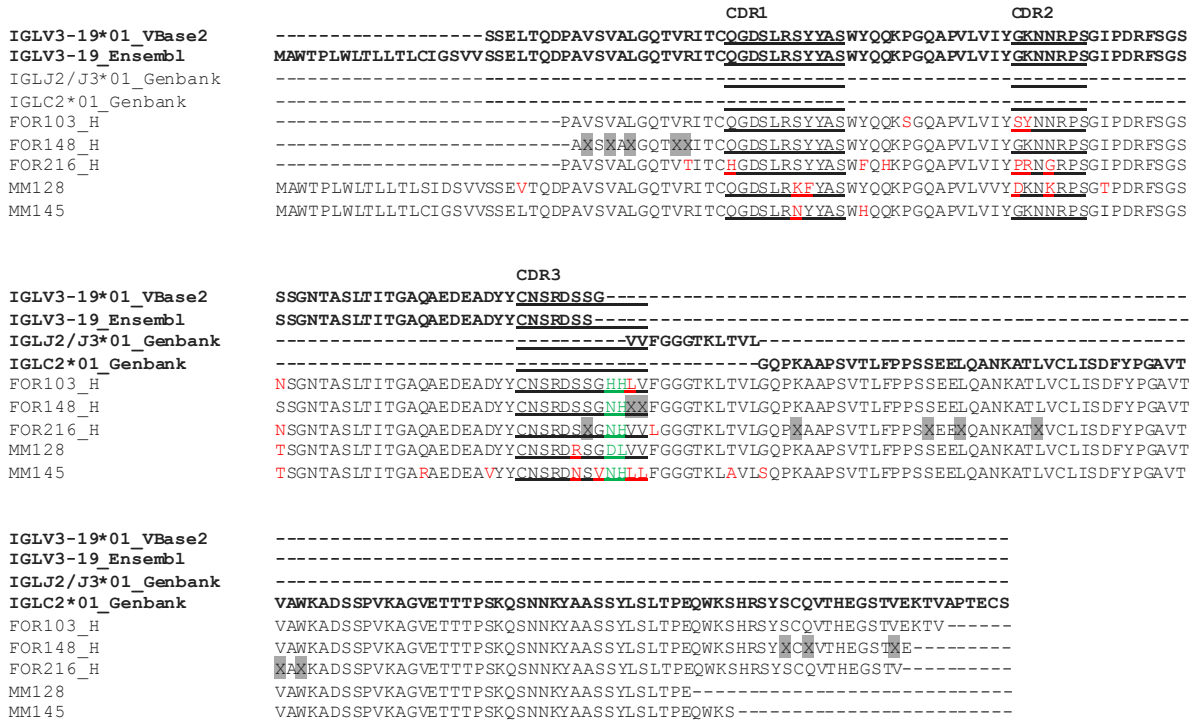
Supplementary Information Figure 49. Sequence alignment of the IGLV2-11 assigned MM127 multiple myeloma amino acid light chain sequence. Bold = reference sequences, underlined = CDR regions, red letter = mutation, MM = multiple myeloma patient, green letter = linker region, X and grey highlight = not unambiguously determined amino acid.



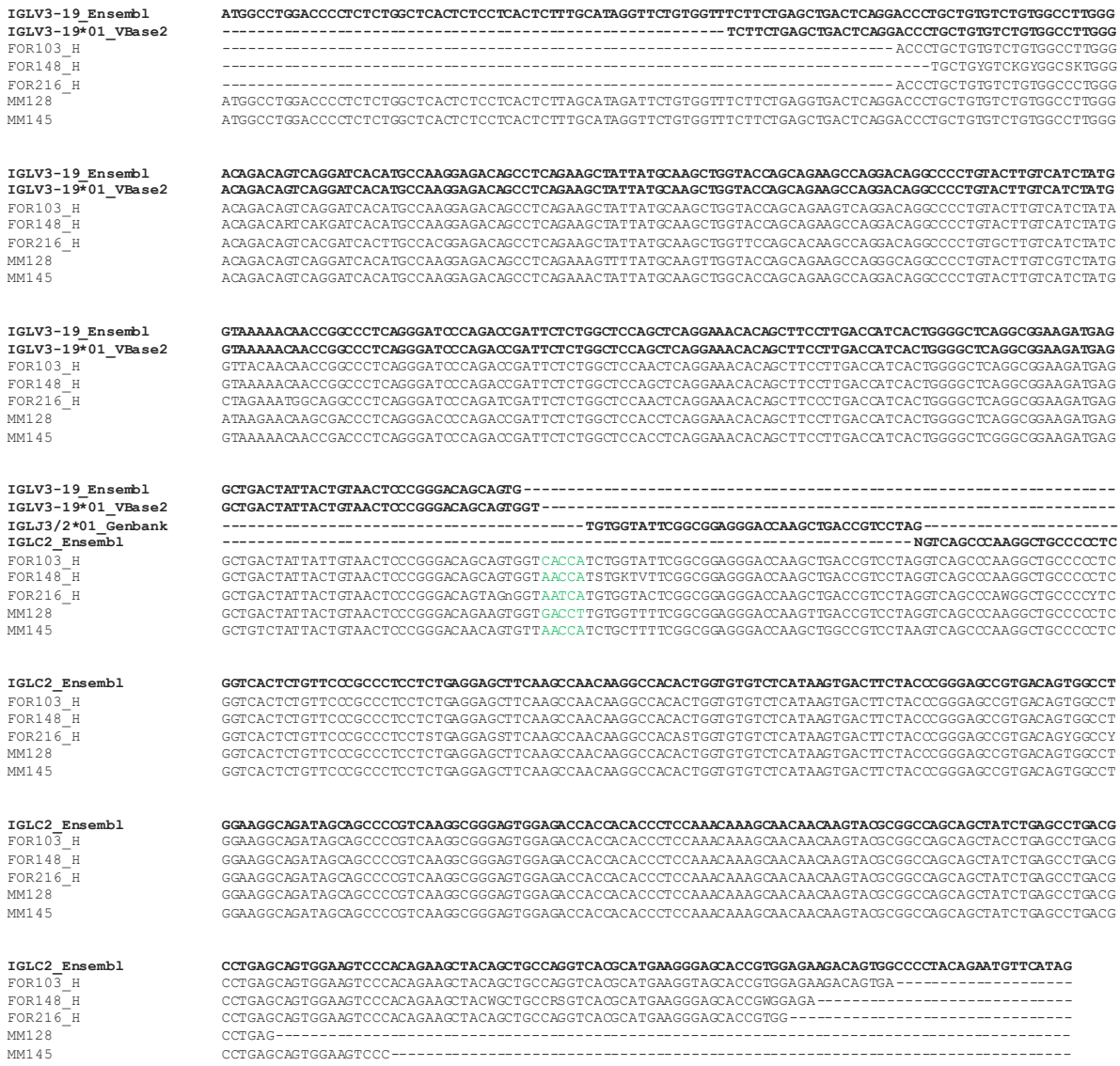
Supplementary Information Figure 50. Sequence alignment of the IGLV2-11 assigned MM127 multiple myeloma cDNA light chain sequence. Bold = reference sequences, MM = multiple myeloma patient, green letter = linker region.

```
MM127_NGS           ATGGCCTGGGCTCTGCTCCTCCTCAGCCTCCTCACTCAGGGCACAGGATCCTGGGCTCAG
MM127_Sanger        ----------------------------------------------------------CAG
                                                                               ***

MM127_NGS           TCTGCCCTGACTCAGCCTCGCTCAGTGTCCGGGTCTCCTGGACAGTCAGTCACCATCTCC
MM127_Sanger        TCTGCCCTGACTCAGCCTCGCTCAGTGTCCGGGTCTCCTGGACAGTCAGTCACCATCTCC
                    ************************************************************

MM127_NGS           TGCACTGGAACCAGCAGTGATGTTGGTGGTTATAACTATGTCTCCTGGTACCAACAGCAC
MM127_Sanger        TGCACTGGAACCAGCAGTGATGTTGGTGGTTATAACTATGTCTCCTGGTACCAACAGCAC
                    ************************************************************

MM127_NGS           CCAGGTAAAGCCCCCAAACTCATGATTTATGATGTCACTAAGCGGCCCTCAGGGGTCCCT
MM127_Sanger        CCAGGTAAAGCCCCCAAACTCATGATTTATGATGTCACTAAGCGGCCCTCAGGGGTCCCT
                    ************************************************************

MM127_NGS           GATCGCTTCTCTGGCTCCAAGTCTGGCACCACGGCCTCCCTGACCATCTCTGGGCTCCAG
MM127_Sanger        GATCGCTTCTCTGGCTCCAAGTCTGGCACCryGGmswmsmyGryCATCTCTGGGCTCCAG
                    ************************************************************

MM127_NGS           GCTGAAGATGAGGCTGATTATTACTGCTGCTCATATGCGGGCATAGATATTTTTGTGCTA
MM127_Sanger        GCTGAAGATGAGGCTGATTATTACTGCTGCTCATATGCGGGCATAGATATTTTTGTGCTA
                    ************************************************************

MM127_NGS           TTCGGCGGAGGGACCAAGCTGACCGTCCTAGGTCAGCCCAAGGCTGCCCCCTCGGTCACT
MM127_Sanger        TTCGGCGGAGGGACCAAGCTGACCGTCCTAGGTCAGCCCAAGGCTGCCCCCTCGGTCACT
                    ************************************************************

MM127_NGS           CTGTTCCCGCCCTCCTCTGAGGAGCTTCAAGCCAACAAGGCCACACTGGTGTGTCTCATA
MM127_Sanger        CTGTTCCCGCCCTCCTCTGAGGAGCTTCAAGCCAACAAGGCCACACTGGTGTGTCTCATA
                    ************************************************************

MM127_NGS           AGTGACTTCTACCCGGGAGCCGTGACAGTGGCCTGGAAGGCAGATAGCAGCCCCGTCAAG
MM127_Sanger        AGTGACTTCTACCCGGGAGCCGTGACAGTGGCCTGGAAGGCAGATAGCAGCCCCGTCAAG
                    ************************************************************

MM127_NGS           GCGGGAGTGGAGACCACCACACCCTCCAAACAAAGCAACAACAAGTATGCGGCCAGCAGC
MM127_Sanger        GCGGGAGTGGAGACCACCACACCCTCCAAACAAAGCAACAACAAGTATGCGGCCAGCAGC
                    ************************************************************

MM127_NGS           TACCTGAGCCTGACGCCTGAGCAGTGGA--------------------------------
MM127_Sanger        TACCTGAGCCTGACGCCTGAGCAGTGGAAGTCCCACAGAAGCTACAGCTGCCAGGTCACG
                    ****************************

MM127_NGS           ------------------
MM127_Sanger        CATGAAGGGAGCACCGTGG
```

Supplementary Information Figure 51. Sequence comparison between MM127 IGLV2-11 assigned light chain sequences generated through a next-generation sequencing and Sanger sequencing approach. For the sequence generated by the Sanger sequencing approach only dominant signals were used. * = nucleotide position consistent, MM = multiple myeloma patient, NGS = light chain sequence generated through next-generation sequencing, Sanger = light chain sequence generated through a Sanger sequencing approach. Sequencing was performed using the VLKL12a_Huhn and CLKL_A_rv_NB oligonucleotides.



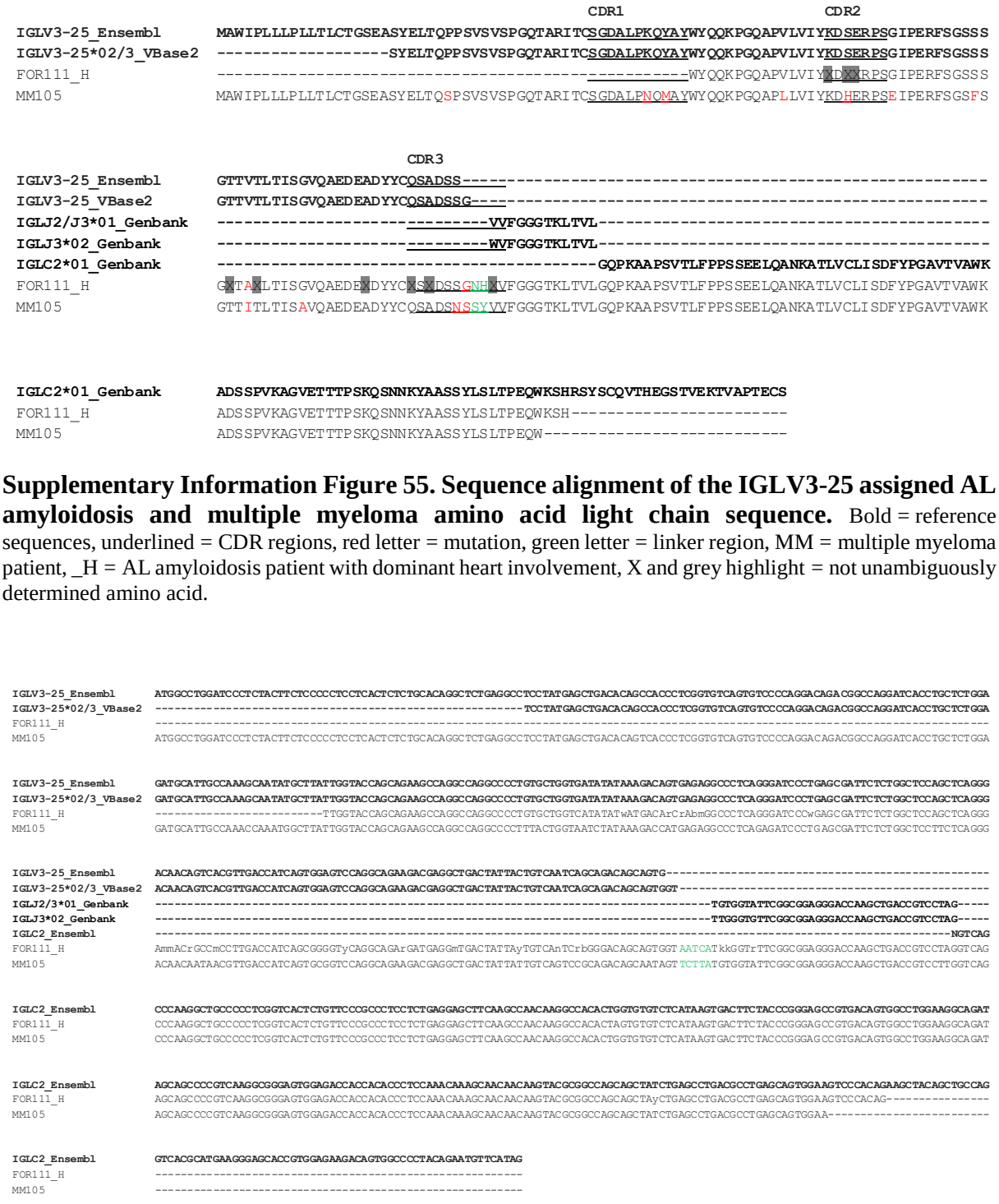
Supplementary Information Figure 52. Sequence alignment of IGLV3-19 assigned AL amyloidosis and multiple myeloma amino acid light chain sequences. Bold = reference sequences, underlined = CDR regions, red letter = mutation, green letter = linker region, MM = multiple myeloma patient, _H = AL amyloidosis patient with dominant heart involvement, X and grey highlight = not unambiguously determined amino acid.



Supplementary Information Figure 53. Sequence alignment of IGLV3-19 assigned AL amyloidosis and multiple myeloma cDNA light chain sequences. green letter = linker region, MM = multiple myeloma patient, _H = AL amyloidosis patient with dominant heart involvement.

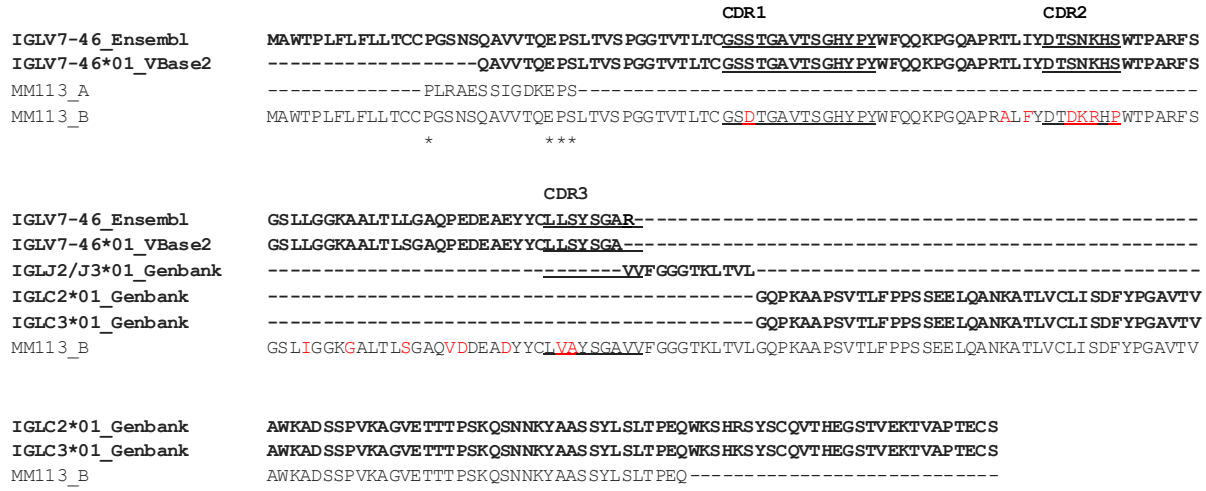
```
MM128_NGS               ATGGCCTGGACCCCTCTCTGGCTCACTCTCCTCACTCTTAGCATAGATTCTGTGGTTTCT
MM128_Sanger            ------------------------------------------------------------


MM128_NGS               TCTGAGGTGACTCAGGACCCTGCTGTGTCTGTGGCCTTGGGACAGACAGTCAGGATCACA
MM128_Sanger            -----------------CCCTGCTGTGTCTGTGGCCTTGGGACAGACAGTCAGGATCACA
                                         ****************************************


MM128_NGS               TGCCAAGGAGACAGCCTCAGAAAGTTTTATGCAAGTTGGTACCAGCAGAAGCCAGGGCAG
MM128_Sanger            TGCCAAGGAGACAGCCTCAGAAAGTTTTATGCAAGTTGGTACCAGCAGAAGCCAGGGCAG
                        ************************************************************


MM128_NGS               GCCCCTGTACTTGTCGTCTATGATAAGAACAAGCGACCCTCAGGGACCCCAGACCGATTC
MM128_Sanger            GCCCCTGTACTTGTCGTCTATGATAAGAACAAGCGACCCTCAGGGACCCCAGACCGATTC
                        ************************************************************


MM128_NGS               TCTGGCTCCACCTCAGGAAACACAGCTTCCTTGACCATCACTGGGGCTCAGGCGGAAGAT
MM128_Sanger            TCTGGCTCCACCTCAGGAAACACAGCTTCCTTGACCATCACTGGGGCTCAGGCGGAAGAT
                        ************************************************************


MM128_NGS               GAGGCTGACTATTACTGTAACTCCCGGGACAGAAGTGGTGACCTTGTGGTTTTCGGCGGA
MM128_Sanger            GAGGCTGACTATTACTGTAACTCCCGGGACAGAAGTGGTGACCTTGTGGTTTTCGGCGGA
                        ************************************************************


MM128_NGS               GGGACCAAGTTGACCGTCCTAGGTCAGCCCAAGGCTGCCCCCTCGGTCACTCTGTTCCCG
MM128_Sanger            GGGACCAAGTTGACCGTCCTAGGTCAGCCCAAGGCTGCCCCCTCGGTCACTCTGTTCCCG
                        ************************************************************


MM128_NGS               CCCTCCTCTGAGGAGCTTCAAGCCAACAAGGCCACACTGGTGTGTCTCATAAGTGACTTC
MM128_Sanger            CCCTCCTCTGAGGAGCTTCAAGCCAACAAGGCCACACTGGTGTGTCTCATAAGTGACTTC
                        ************************************************************


MM128_NGS               TACCCGGGAGCCGTGACAGTGGCCTGGAAGGCAGATAGCAGCCCCGTCAAGGCGGGAGTG
MM128_Sanger            TACCCGGGAGCCGTGACAGTGGCCTGGAAGGCAGATAGCAGCCCCGTCAAGGCGGGAGTG
                        ************************************************************


MM128_NGS               GAGACCACCACACCCTCCAAACAAAGCAACAACAAGTACGCGGCCAGCAGCTATCTGAGC
MM128_Sanger            GAGACCACCACACCCTCCAAACAAAGCAACAACAAGTACGCGGCCAGCAGCTATCTGAGC
                        ************************************************************


MM128_NGS               CTGACGCCTGAG-------------------------------------------------
MM128_Sanger            CTGACGCCTGAGCAGTGGAAGTCCCACAGAAGCTACAGCTGCCAGGTCACGCATGAAGGG
                        ************


MM128_NGS               -----------
MM128_Sanger            AGCACCGTGGA
```

Supplementary Information Figure 54. Sequence comparison between MM128 IGLV3-19 assigned light chain sequences generated through a next-generation sequencing and Sanger sequencing approach. For the sequence generated by the Sanger sequencing approach only dominant signals were used. * = nucleotide position consistent, MM = multiple myeloma patient, NGS = light chain sequence generated through next-generation sequencing, Sanger = light chain sequence generated through a Sanger sequencing approach. Sequencing was performed using the VLKL3_H_fw_NB and CLKL_A_rv_NB oligonucleotides.

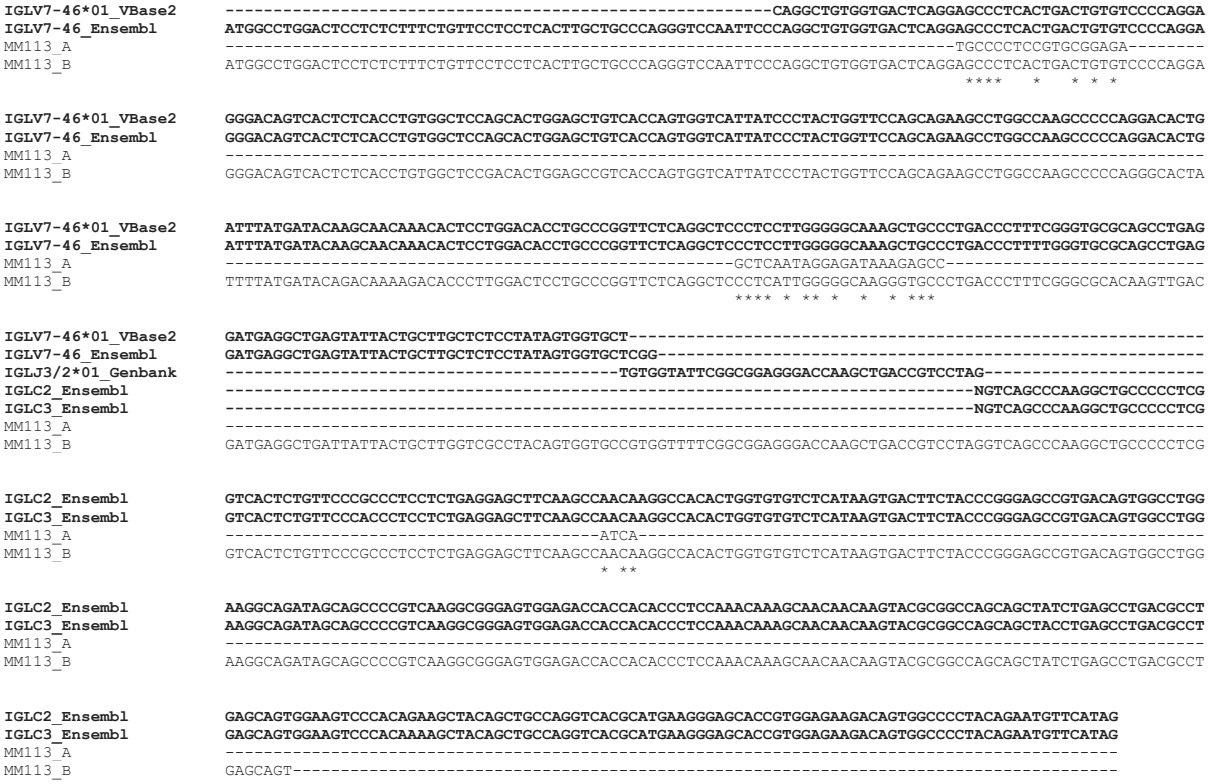


Supplementary Information Figure 55. Sequence alignment of the IGLV3-25 assigned AL amyloidosis and multiple myeloma amino acid light chain sequence. Bold = reference sequences, underlined = CDR regions, red letter = mutation, green letter = linker region, MM = multiple myeloma patient, _H = AL amyloidosis patient with dominant heart involvement, X and grey highlight = not unambiguously determined amino acid.

Supplementary Information Figure 56. Sequence alignment of the IGLV3-25 assigned AL amyloidosis and multiple myeloma cDNA light chain sequence. MM = multiple myeloma patient, _H = AL amyloidosis patient with dominant heart involvement, green letter = linker region. Nucleotide signal overlaps in the Sanger sequencing were specified according to the IUPAC code.



Supplementary Information Figure 57. Sequence alignment of the IGLV7-46 assigned MM113 multiple myeloma amino acid light chain sequence. Bold = reference sequences, underlined = CDR regions, red letter = mutation, MM = multiple myeloma patient, * = nucleotide position consistent. In the bioinformatic analysis of the bulk RNA sequencing, several sequence sections were given – indicated by _A and _B.



Supplementary Information Figure 58. Sequence alignment of the IGLV7-46 assigned MM113 multiple myeloma cDNA light chain sequence. Bold = reference sequences, MM = multiple myeloma patient. * = nucleotide position consistent. In the bioinformatic analysis of the bulk RNA sequencing, several sequence sections were given – indicated by _A and _B. Nucleotide signal overlaps in the Sanger sequencing were specified according to the IUPAC code.

```
                                                                                   CDR1
IGLV8-61_Ensembl         MSVPTMAWMMLLLGLLAYGSGVDSQTVVTQEPSFSVSPGGTVTLTCGLSSGSVSTSYYPSWYQQTPGQAP
IGLV8-61*01_VBase2       ------------------------QTVVTQEPSFSVSPGGTVTLTCGLSSGSVSTSYYPSWYQQTPGQAP
FOR229_HK                ----------------------------TQEPSLSVSPGGTVTLTCGLRSGSVSSNYHPSWHQQIPGQVP

                              CDR2                                       CDR3
IGLV8-61_Ensembl         RTLIYSTNTRSSGVPDRFSGSILGNKAALTITGAQADDESDYYCVLYMGSGI------------------
IGLV8-61*01_VBase2       RTLIYSTNTRSSGVPDRFSGSILGNKAALTITGAQADDESDYYCVLYMGSG-------------------
IGLJ3*02_Genbank         ----------------------------------------------------WVFGGGTKLTVL------
IGLC3*01_Genbank         ----------------------------------------------------------------GQPKAA
FOR229_HK                RTLIYSTNIRSSGVPHRFSGSIVGKKAALTITGAQADDESDYYCMLYLGSGVSVFGGGTKLTVLGQPKAA

IGLC3*01_Genbank         PSVTLFPPSSEELQANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYLSLTP
FOR229_HK                PSVTLFPPSSEELQANKATLVCLISDFYPGAVTVAWKADSSPVKAGVETTTPSKQSNNKYAASSYLSLTP

IGLC3*01_Genbank         EQWKSHKSYSCQVTHEGSTVEKTVAPTECS
FOR229_HK                EQWKSHKSYSCQVTHEGSTV----------
```

Supplementary Information Figure 59. Sequence alignment of the IGLV8-61 assigned FOR229 AL amyloidosis amino acid light chain sequence. Bold = reference sequences, underlined = CDR regions, red letter = mutation, _HK = AL amyloidosis patient with dominant heart and kidney involvement.

```
IGLV8-61_Ensembl      ATGAGTGTCCCCACCATGGCCTGGATGATGCTTCTCCTCGGACTCCTTGCTTATGGATCAGGAGTGGATTCTCAGACTGTGGTGACCCAGG AGCCATCGTT
IGLV8-61*01_VBase2    ------------------------------------------------------------------------CAGACTGTGGTGACCCAGGAGCCATCGTT
FOR229_HK             ----------------------------------------------------------------------------------TGACTCAGGAGCCATCGTT

IGLV8-61_Ensembl      CTCAGTGTCCCCTGGAGGGACAGTCACACTCACTTGTGGCTTGAGCTCTGGCTCAGTCTCTACTAGTTACTACCCCAGCTGGTACCAGCAGACCCCAGGCC
IGLV8-61*01_VBase2    CTCAGTGTCCCCTGGAGGGACAGTCACACTCACTTGTGGCTTGAGCTCTGGCTCAGTCTCTACTAGTTACTACCCCAGCTGGTACCAGCAGACCCCAGGCC
FOR229_HK             GTCAGTGTCCCCTGGAGGGACAGTCACACTCACTTGTGGCTTGAGGTCTGGATCAGTCTCAAGTAATTACCACCCCAGCTGGCACCAGCAGATCCCAGGCC

IGLV8-61_Ensembl      AGGCTCCACGCACGCTCATCTACAGCACAAACACTCGCTCTTCTGGGGTCCCTGATCGCTTCTCTGGCTCCATCCTTGGGAACAAAGCTGCCCTCACCATC
IGLV8-61*01_VBase2    AGGCTCCACGCACGCTCATCTACAGCACAAACACTCGCTCTTCTGGGGTCCCTGATCGCTTCTCTGGCTCCATCCTTGGGAACAAAGCTGCCCTCACCATC
FOR229_HK             AGGTTCCACGCACCCTCATCTACAGCACAAATATTCGCTCTTCTGGGGTCCCTCATCGCTTCTCTGGCTCCATCGTTGGGAAAAAGGCTGCACTCACCATC

IGLV8-61_Ensembl      ACGGGGGCCCAGGCAGATGATGAATCTGATTATTACTGTGTGCTGTATATGGGTAGT-----------------------------------------------
IGLV8-61*01_VBase2    ACGGGGGCCCAGGCAGATGATGAATCTGATTATTACTGTGTGCTGTATATGGGTAGTGGCGGCATTTC------------------------------------
IGLJ3*02_Genbank      -----------------------------------------------------------------TTGGGTGTTCGGCGGAGGGACCAAGCTGACCGTCCTAG-
IGLC3_Ensembl         ------------------------------------------------------------------------------------------------------NG
FOR229_HK             ACGGGGGCCCAGGCAGATGATGAATCTGATTATTATTGTATGCTGTATTTGGGTAGTGGCGTTTCGGTGTTCGGCGGAGGGACCAAGCTGACCGTCCTGGG

IGLC3_Ensembl         TCAGCCCAAGGCTGCCCCCTCGGTCACTCTGTTCCCACCCTCCTCTGAGGAGCTTCAAGCCAACAAGGCCACACTGGTGTGTCTCATAAGTGACTTCTACC
FOR229_HK             TCAGCCCAAGGCTGCCCCCTCGGTCACTCTGTTCCCACCCTCCTCTGAGGAGCTTCAAGCCAACAAGGCCACACTGGTGTGTCTCATAAGTGACTTCTACC

IGLC3_Ensembl         CGGGAGCCGTGACAGTGGCCTGGAAGGCAGATAGCAGCCCCGTCAAGGCGGGAGTGGAGACCACCACACCCTCCAAACAAAGCAACAACAAGTACGCGGCC
FOR229_HK             CGGGAGCCGTGACAGTGGCCTGGAAGGCAGATAGCAGCCCCGTCAAGGCGGGAGTGGAGACCACCACACCCTCCAAACAAAGCAACAACAAGTACGCGGCC

IGLC3_Ensembl         AGCAGCTACCTGAGCCTGACGCCTGAGCAGTGGAAGTCCCACAAAAGCTACAGCTGCCAGGTCACGCATGAAGGGAGCACCGTGGAGAAGACAGTGGCCCC
FOR229_HK             AGCAGCTACCTGAGCCTGACGCCTGAGCAGTGGAAGTCCCACAAAAGCTACAGCTGCCAGGTCACGCATGAAGGGAGCACCGTGG----------------

IGLC3_Ensembl         TACAGAATGTTCATAG
FOR229_HK             ----------------
```

Supplementary Information Figure 60. Sequence alignment of the IGLV8-61 assigned FOR229 AL amyloidosis cDNA light chain sequence. Bold = reference sequences, underlined = CDR regions, red letter = mutation, _HK = AL amyloidosis patient with dominant heart and kidney involvement. Nucleotide signal overlaps in the Sanger sequencing were specified according to the IUPAC code.

Curriculum Vitae

Personal Details

Surname and forename	Berghaus, Natalie
Date of birth	24.11.1994
Place of birth	Wiesbaden, Germany
Nationality	German

School History

Secondary school: Max-Planck-Schule, Rüsselsheim
Secondary school: Fürst-Johann-Ludwig Schule, Hadamar
Abitur: Fürst-Johann-Ludwig Schule, Hadamar (12.06.2013)

Academic History

since 12.2022
Scientific Employee
University Hospital Heidelberg
Department of Neuropathology, Prof. Dr. Andreas von Deimling
Section of Molecular Neuropathology, Prof. Dr. Dr. Felix Sahm

since 09.2019
PhD Position Dr. sc. hum.
University Hospital Heidelberg – supervisor: Prof. Stefan O. Schönland
"Biological and clinical impact of the immunoglobulin light chain sequence diversity in patients with dominant heart AL amyloidosis"

10.2016 – 04.2019
Master Molecular Biology – grade: 1.5
Goethe University Frankfurt am Main – supervisor: Prof. Helge Bode
″Investigation of the Fabclavine-Biosynthesis″

10.2013 – 08.2016
Bachelor Biology – grade: 1.9
Goethe University Frankfurt am Main – supervisor: Prof. Volker Müller
″Investigation on the acetaldehyde-metabolism of the acetogenic bacterium *Acetobacterium woodii*″

Experiences and Awards

2022
- oral presentation at the "International symposium on Amyloidosis"
- DGAK Young Investigator Award

2020
- successful grand application as lead proposer (Stiftungsmittel Universitätsklinikum Heidelberg; 4000 €)

2018
- poster presentation at the VAAM Workshop: "Biology of bacteria producing natural products"

Danksagung

Zu aller erst möchte ich mich bei allen Patienten und Patientinnen bedanken!

Ich möchte mich bei Prof. Dr. Stefan Schönland, Prof. Dr. Ute Hegenbart und Dr. Stefanie Huhn für die Überlassung des Themas bedanken. Ich möchte mich auch bei ihnen dafür bedanken, dass sie immer hilfsbereit waren und mich unterstützt haben – trotz ihrer hohen Belastung in Klinik und Forschung. Ich habe den größten Respekt vor ihnen.

Herrn Prof. Dr. Stefan Schönland möchte ich darüber hinaus dafür danken, dass er sich die Zeit genommen hat, mit mir meine Forschung zu diskutieren und mich immer gefördert und gefordert hat. Sein Bestreben, diese Erkrankung weiter zu erforschen und schlussendlich ihre Ursache aufzuklären, hat mich sehr beeindruckt. Frau Prof. Dr. Ute Hegenbart möchte ich herzlichst für die Beantwortung aller fachlichen Fragen danken und bewundere sie sehr für ihre Hingabe ihren Patienten und Patientinnen gegenüber. Ich möchte mich auch dafür bedanken, dass sie trotz ihrer hohen Arbeitsbelastung immer ein Lächeln oder ein nettes Wort übrighatte. Ebenso möchte ich Frau Dr. Stefanie Huhn dafür danken, dass sie immer hinter mir stand, sich meine Ideen angehört hat und mit mir daran gearbeitete hat, sie umzusetzen. Ich möchte mich auch dafür bedanken, dass sie mir zugehört hat und wir immer eine offene Kommunikation führen konnten.

Natürlich möchte ich mich auch bei dem ganzen Molekularbiologischen Labor bedanken, für das zusammensitzen, lachen und dafür, dass sie immer ein offenes Ohr hatten. Bei Dr. Philipp Reichert möchte ich mich dafür bedanken, dass er immer sofort zur Stelle war und versucht hat bei allen Problemen und Herausforderungen zu helfen.

Im Besonderen möchte ich meiner Kollegin Sarah Schreiner danken, für alle gemeinsam durchlebten Höhen und Tiefen, in denen wir es geschafft haben, immer aufeinander Rücksicht zu nehmen und uns gegenseitig zu unterstützen. Wir waren immer Mitarbeiterinnen des Monats, haben uns mehr gesehen als jede andere Person und alles geteilt von Gefühlen über ein Zimmer bis hin zu Wein. Wir haben das zusammen durchgestanden und das werde ich nie vergessen.

Ich möchte auch all meinen Freunden danken und im Besondern Gina, da sie die Geduld hatte, diese Arbeit zu lesen und alle meine Kommafehler und sprachlichen Irrwege zu korrigieren.

Besonderer Dank gilt meinen Eltern, die mich immer bedingungslos unterstützt haben und mir ermöglicht haben, meiner Leidenschaft für die Biologie nachzugehen. Ich möchte mich auch bei Philippe bedanken, für das Zuhören, den Versuch zu verstehen, an was ich forsche, seine Wertschätzung und Liebe.

Eidesstattliche Versicherung

1. Bei der eingereichten Dissertation zu dem Thema

   „Biological and clinical impact of the immunoglobulin light chain sequence diversity in patients with dominant heart AL amyloidosis“

   handelt es sich um meine eigenständig erbrachte Leistung.

2. Ich habe nur die angegebenen Quellen und Hilfsmittel benutzt und mich keiner unzulässigen Hilfe Dritter bedient. Insbesondere habe ich wörtlich oder sinngemäß aus anderen Werken übernommene Inhalte als solche kenntlich gemacht.

3. Die Arbeit oder Teile davon habe ich bislang nicht an einer Hochschule des In- oder Auslands als Bestandteil einer Prüfungs- oder Qualifikationsleistung vorgelegt.

4. Die Richtigkeit der vorstehenden Erklärungen bestätige ich.

5. Die Bedeutung der eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unrichtigen oder unvollständigen eidesstattlichen Versicherung sind mir bekannt. Ich versichere an Eides statt, dass ich nach bestem Wissen die reine Wahrheit erklärt und nichts verschwiegen habe.

_____ _____

Ort und Datum Unterschrift