Check for updates

Received: 2 March 2022 | Revised: 3 August 2022 | Accepted: 5 August 2022

DOI: 10.1002/ajh.26684

CORRESPONDENCE

Single-molecule real-time sequencing of the M protein: Toward personalized medicine in monoclonal gammopathies

To the Editor:

Each patient with a monoclonal gammopathy has a unique monoclonal (M) protein, whose sequence can be used as a tumoral fingerprint to track the presence of the B cell or plasma cell (PC) clone itself. Moreover, the M protein can directly cause potentially life-threatening organ damage, which is dictated by the specific, patient's unique clonal light and/or heavy chain amino acid sequence, as in patients affected by immunoglobulin light chain (AL) amyloidosis.[1] However, patients' specific M protein sequences remain mostly undefined and molecular mechanisms underlying M protein-related clinical manifestations are largely obscure.

We combined the unbiased amplification of expressed immunoglobulin genes through inverse PCR from circularized, double-stranded cDNA using primers annealing to the constant regions of immunoglobulin genes, with single-molecule, real-time, long-read DNA sequencing and bioinformatics and immunogenetic analyses[2–4] (Online Methods, Figures S1, S2, Table S1). The resulting methodology, termed Single-Molecule Real-Time Sequencing of the M protein (SMaRT M-Seq), identifies the full-length sequence of the variable region of expressed immunoglobulin genes and ranks the obtained sequences based on their relative abundance, thus enabling the identification of the full-length variable sequence of light and/or heavy chains from a high number of patients analyzed in parallel.

SMaRT M-Seq has undergone appropriate technical validation (Table S2). Sequencing of contrived bone marrow (BM) samples generated through serial dilutions of κ- or λ-expressing PC lines into control BM, as well as sequencing of replicate, bona fide BM samples from AL patients and comparison with gold-standard techniques of immunoglobulin gene cloning and sequencing, showed: (i) 100% sequence-accuracy at the individual base-pair level; (ii) high repeatability (coefficient of variation <0.8% for sequencing of pentaplicate BM samples) in defining the molecular clonal size (i.e., the fraction of total immunoglobulin sequences coinciding with the clonal sequence); (iii) a high sensitivity in identifying clonal immunoglobulin sequences (10^{-2}–10^{-3} when employing low-coverage sequencing on multiple, pooled samples) (Appendix S1, Figures S3–S5).

To further extend the technical validation of the methodology and assess its throughout, we employed SMaRT M-Seq for the identification of clonal immunoglobulin sequences from BM mononuclear cells of a cohort of 89 consecutive patients with a diagnosis or a suspicion of systemic AL amyloidosis analyzed in parallel in one sequencing round (Figure S6). In 6 of these patients, comparison with standard cloning and sequencing approaches confirmed 100% identity with respect to the sequence obtained by SMaRT M-Seq (Figure S7). In addition, 3 of these patients were analyzed in duplicate with SMaRT M-Seq, and the sequence-based molecular clonal sizes of the two technical replicates were highly comparable (Figure 1). These results further confirm the accuracy and repeatability of this method also when the assay is employed to analyze a higher number of samples in parallel.

Of the 89 sequenced patients, a final diagnosis of systemic AL amyloidosis could be established in 84 patients, including 5 cases with undetectable M protein by means of conventional M protein studies (Figure S8, Table S3).

Of note, SMaRT M-Seq identified a dominant immunoglobulin LC sequence in all 84 patients (but not in patients analyzed in parallel where a monoclonal gammopathy was eventually excluded, Figure S9). The median molecular clonal size was 88.3% (IQR: 70.7%–93%) (Figure 1) and showed a significant correlation with the percentage of BM-PC infiltrate and with serum free LC levels ($p < 0.0001$ in each case) (Figure S10). Patients' clonal sequences proved to be unique (Figure S11). Germline gene usage was in agreement with the expectations for a population of patients with AL amyloidosis (Figure S12) and correlated with selected clinical features (Figure S13).

As an additional way to verify the accuracy of the methodology in identifying the clonal, expressed LC, we compared the sequencing results obtained with SMaRT M-Seq on BM samples with proteomics data from matched, amyloid-containing fat tissues for 4 patients. In all cases, the expected clonal LC variable sequence as assessed by SMaRT M-Seq was the potentially amyloidogenic protein with the highest sequence coverage and was by far the first immunoglobulin LC sequence in terms of unique peptides identified compared to other, published immunoglobulin LCs (Figure S14). Collectively, these data show that SMaRT M-Seq performed on a high number of BM samples from patients with monoclonal gammopathies analyzed in parallel can accurately and reproducibly identify a clonal immunoglobulin LC sequence in all instances, even in cases with low BM-B cell/PC clonal burden and with undetectable M protein by means of conventional diagnostic techniques.

We then investigated whether the full-length variable sequence information attainable at diagnosis using SMaRT M-Seq and the use of inverse PCR coupled to short-read sequencing might enable the detection of low-level, residual clonotypic sequences, as in the context of minimal residual disease (MRD) assessment. Using contrived BM samples mimicking progressively smaller plasma cell clones, a

 © 2022 Wiley Periodicals LLC.

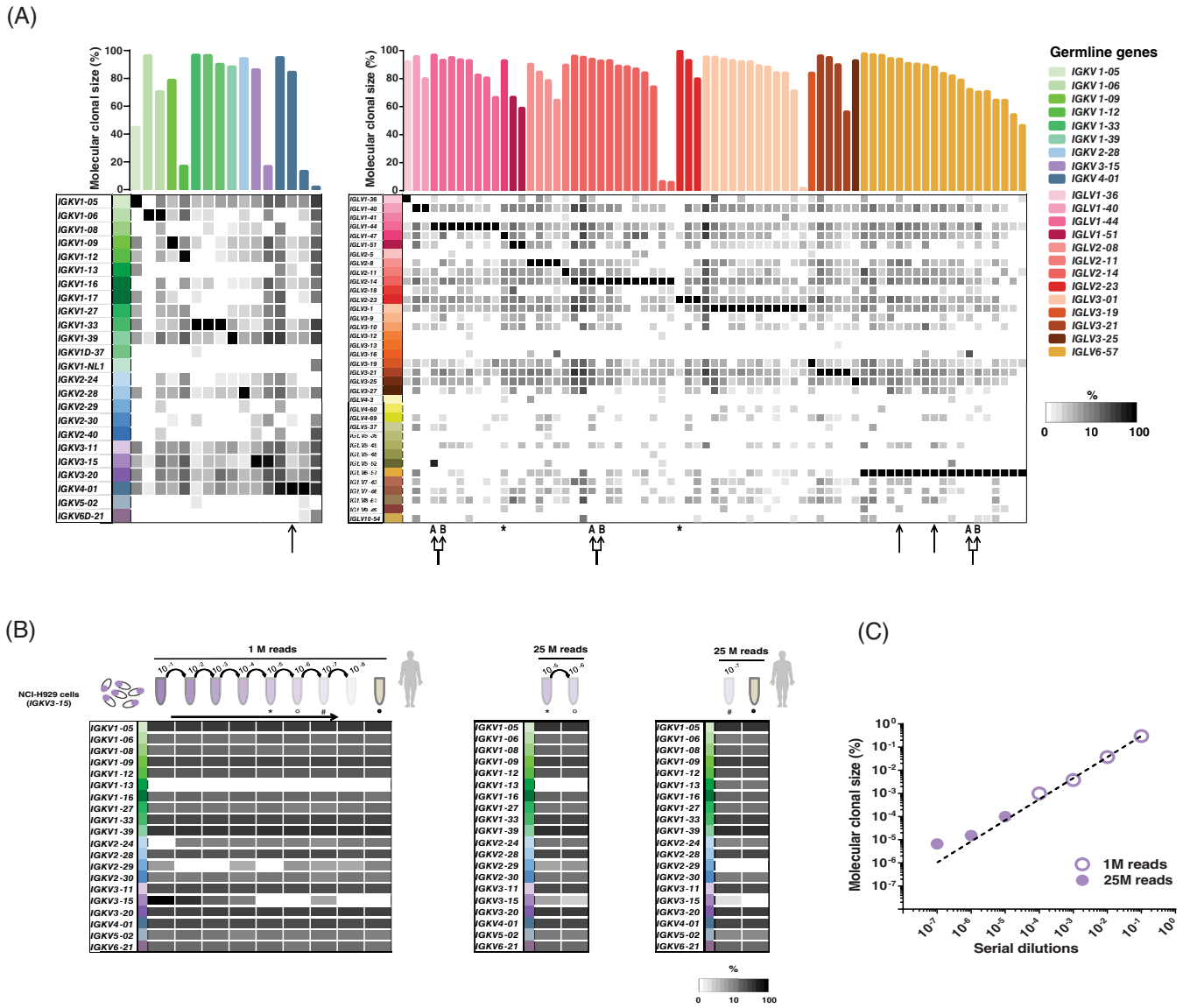

FIGURE 1 Single Molecule Real-Time Sequencing of the M protein (SMaRT M-Seq). (A) Sensitivity, accuracy, repeatability and throughput of SMaRT M-Seq in a cohort of 84 AL patients sequenced in parallel in one sequencing round. Expression levels of different *IGKV* (left) and *IGLV* (right) germline genes (each denoted by a distinctive color) as assessed by SMaRT M-Seq starting from the bone marrow. Bar graphs indicate the molecular clonal size of the dominant clone identified by Vidjil analysis in each tested sample (the corresponding germline gene is indicated with the bar color). In two patients (*) the dominant clone was identified by IMGT/HighV-Quest. Six patients (arrows) were analyzed also through standard cloning and sequencing. Of these, three patients were analyzed in duplicates through SMaRT M-Seq (two-headed arrows). Further analyses are available in Figures S7–14. (B),(C) Sensitive, clonotypic reads detection exploiting full-length variable light chain sequence knowledge. (B) Expression levels (in shades of gray) of different *IGKV* germline genes (each denoted by a distinctive color), starting from serial Log_{10} dilutions of RNA from *IGKV3-15*-secreting NCI-H929 cells into RNA of the bone marrow from a control subject (unspiked control sample in gray). Left: samples sequenced in one run on a 1 M cartridge. Middle and right: samples sequenced in one run on a 25 M cartridge, each. Numbers below each column denote percentages of clonotypic sequences identified in each tested sample. Symbols under tubes (*, °, #, ●) indicate that the same samples are used to generate multiple libraries. (C) Linear correlation plot between serial Log_{10} dilutions of RNA input level and observed molecular clonal sizes

clonotypic sequence was identified up to the 10^{-7} dilution, with a progressively decreasing molecular clonal size, indicating linearity of the amplification and sequencing approach (Figure 1B,C, Appendix S1). Overall, these data suggest that the knowledge of patients' specific, expressed full-length clonal immunoglobulin variable sequence may be exploited to facilitate MRD assessment.

We have established SMaRT M-Seq as a novel, validated assay to reliably identify the full-length variable sequence of M proteins. The assay relies on the unbiased amplification of expressed immunoglobulin gene(s) through an inverse PCR, coupled with single-molecule real-time DNA sequencing. Within a complex biological sample like BM or peripheral blood, SMaRT M-Seq enables ranking the obtained reads

based on their relative abundance, which is a measure of the relative abundance of mRNA molecules within the sample. As such, this method can be employed to infer the full-length variable sequence of an expressed clonal immunoglobulin as the dominant sequence identified within the sample under examination, thus enabling the use of clonal, expressed immunoglobulin sequence information for basic research studies or potential diagnostic applications.

The accuracy of SMaRT M-Seq at identifying the full-length variable sequence of the clonal immunoglobulin gene at the single nucleotide level has been demonstrated through different approaches. First, through the comparison with sequencing results obtained by means of conventional sequencing methods, both in contrived samples spiked with PC lines secreting a known κ- or λ-LC (thus serving as reference material) and in bona fide BM samples. Second, through the investigation of germline gene usage in our analyzed cohort of consecutive AL patients, which closely reflects the expected distribution of immunoglobulin LC genes based on previous studies. Third, through the identification—within amyloid-laden fat tissues—of high amounts of tryptic digestion peptides aligning to the clonal LC sequence obtained through SMaRT M-Seq.

The analysis of multiple BM samples from AL patients as technical replicates demonstrated the repeatability of SMaRT M-Seq for identifying both the full-length variable sequence of clonal immunoglobulin genes at the single nucleotide level, and the molecular clonal size. The latter is a measure of the relative abundance of clonal sequences in a given sample. Differently from genomic DNA-based sequencing methods, where reads are considered to be directly proportionate to the amount of tumoral cells within the sample, in mRNA/cDNA-based methods, including SMaRT M-Seq, reads reflect both the amount of tumoral cells within the sample and the average number of immunoglobulin transcripts per cell, which may differ among different patients. This may reduce linearity between the true clonal burden within the biological sample and the obtained molecular clonal size based on the frequency of clonal sequences and could be regarded as a limitation of mRNA/cDNA-based methods. On the other hand, the excess of light- and/or heavy-chain mRNA molecules compared to a single rearranged genomic DNA molecule in each tumoral cell may favor sensitivity of mRNA/cDNA-based sequencing methods for clonal detection.[5]

Besides the relative abundance of clonal cells within the sample under exam, sensitivity of SMaRT M-Seq is determined also by the number of reads analyzed per sample. This is in turn dictated by the sequencing output of the employed sequencing platform, and by the number of pooled samples analyzed in a given sequencing round, thus proving to be scalable. Even when analyzing multiple samples on a sequencing platform with low sequencing output, the achieved sensitivity of SMaRT M-Seq significantly exceeds the requirements for the identification of clonal B cells/plasma cells in patients with AL amyloidosis at diagnosis.

Besides AL amyloidosis, clinical manifestations causally linked to the presence of specific M proteins can also develop in the context of multiple myeloma, Waldenström macroglobulinemia, and monoclonal gammopathies of clinical significance.[1] The molecular mechanisms underlying these conditions are poorly understood, partly because of the limited number of clinically annotated, sequenced M proteins. Therefore, the possibility of reliably determining the entire variable region of expressed, disease-related immunoglobulin gene sequences from a high number of affected patients has the potential of elucidating molecular mechanisms of pathogenicity and enabling sequence-based predictive models.[6] The identification of pathogenic LCs may improve our capability to make an early diagnosis that is of vital importance. Moreover, at the individual patient level, the identification of expressed clonotypic immunoglobulin sequences could enable approaches of personalized medicine for the sensitive detection of patients' specific M proteins at diagnosis and after anti-clonal therapy.

FUNDING INFORMATION

This work was supported by grants from Amyloidosis Foundation (MN), Italian Ministry of Health (Ricerca Finalizzata, grant #GR-2018-12 368 387, Ricerca Corrente) (EMil, MN), CARIPLO Foundation (grant #2018–0257) (EMil, MN), Cancer Research UK [C355/A26819], FC AECC and AIRC under the Accelerator Award Program (GM, GP, MN), the Italian Ministry of Research and Education (PRIN 20207XLJB2) (SR, GP) and the Fondazione ARISLA (project TDP- 43-STRUCT) (SR).

CONFLICT OF INTEREST

Pasquale Cascino, Giovanni Palladini, and Mario Nuvolone are inventors on a patent application related to this work. EMih owns shares in the company aiNET GmbH.

DATA AVAILABILITY STATEMENT

The data generated in this study are available within the article and its supplementary data files. LC sequences have been deposited to GenBank (MZ595009-MZ595094) and proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (PXD028093 and PXD03587).

Pasquale Cascino[1,2], Alice Nevone[1,2], Maggie Piscitelli[1,2], Claudia Scopelliti[1,2], Maria Girelli[1,2], Giulia Mazzini[1,2], Serena Caminito[1,2], Giancarlo Russo[3], Paolo Milani[2], Marco Basset[1,2], Andrea Foli[1,2], Francesca Fazio[4], Simona Casarini[1,2], Margherita Massa[2], Margherita Bozzola[1,2], Jessica Ripepi[1,2], Melania Antonietta Sesta[1,2], Gloria Acquafredda[5,6], Marica De Cicco[5,6], Antonia Moretta[5,6], Paola Rognoni[1,2], Enrico Milan[7], Stefano Ricagno[8,9], Francesca Lavatelli[1,2], Maria Teresa Petrucci[4], Enkelejda Miho[10,11,12], Catherine Klersy[13], Giampaolo Merlini[1,2], Giovanni Palladini[1,2], Mario Nuvolone[1,2]

[1]*Department of Molecular Medicine, University of Pavia, Pavia, Italy*
[2]*Amyloidosis Research and Treatment Center, Fondazione IRCCS Policlinico San Matteo, Pavia, Italy*
[3]*EMBL partner institute for genome editing, Life Science Center, Vilnius University, Vilnius, Lithuania*

[4]Hematology, Department of Translational and Precision Medicine, Azienda Ospedaliera Policlinico Umberto I, Sapienza University of Rome, Rome, Italy
[5]Pediatric Hematology Oncology Unit, Department of Maternal and Children's Health, Fondazione IRCCS Policlinico San Matteo, Pavia, Italy
[6]Cell Factory and Center for Advanced Cellular Therapies, Department of Maternal and Children's Health, Fondazione IRCCS Policlinico San Matteo, Pavia, Italy
[7]Age related Diseases Unit, Division of Genetics and Cell Biology, San Raffaele Scientific Institute and University Vita-Salute San Raffaele, Milan, Italy
[8]Department of Biosciences, Università degli Studi di Milano, Milan, Italy
[9]Institute of Molecular and Translational Cardiology, IRCCS Policlinico San Donato, Milan, Italy
[10]Institute of Medical Engineering and Medical Informatics, School of Life Sciences, University of Applied Sciences and Arts Northwestern Switzerland FHNW, Muttenz, Switzerland
[11]SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland
[12]aiNET GmbH, Basel, Switzerland
[13]Clinical Epidemiology and Biometry Service, Fondazione IRCCS Policlinico San Matteo, Pavia, Italy

Correspondence
Mario Nuvolone and Giovanni Palladini, Amyloidosis Research and Treatment Center, Fondazione IRCCS Policlinico San Matteo, Viale Golgi 19, Pavia 27100, Italy.
Email: mario.nuvolone@unipv.it and giovanni.palladini@unipv.it

Pasquale Cascino and Alice Nevone contributed equally to this study.

ORCID

Pasquale Cascino https://orcid.org/0000-0002-0032-8466
Alice Nevone https://orcid.org/0000-0002-7724-1043
Maggie Piscitelli https://orcid.org/0000-0003-2115-3439
Giulia Mazzini https://orcid.org/0000-0001-7695-2938
Serena Caminito https://orcid.org/0000-0003-2381-1829
Giancarlo Russo https://orcid.org/0000-0003-0707-2640
Paolo Milani https://orcid.org/0000-0002-2268-9422
Marco Basset https://orcid.org/0000-0003-3829-0499
Andrea Foli https://orcid.org/0000-0002-7489-492X
Simona Casarini https://orcid.org/0000-0003-4730-7467
Margherita Massa https://orcid.org/0000-0001-8037-0439
Margherita Bozzola https://orcid.org/0000-0002-0167-3691
Gloria Acquafredda https://orcid.org/0000-0001-5621-2748
Antonia Moretta https://orcid.org/0000-0001-7040-9715
Paola Rognoni https://orcid.org/0000-0001-6350-1919
Enrico Milan https://orcid.org/0000-0003-3094-4275
Stefano Ricagno https://orcid.org/0000-0001-6678-5873
Francesca Lavatelli https://orcid.org/0000-0002-7693-4825
Maria Teresa Petrucci https://orcid.org/0000-0003-0688-4882
Enkelejda Miho https://orcid.org/0000-0001-6461-0519
Catherine Klersy https://orcid.org/0000-0003-0314-8548
Giampaolo Merlini https://orcid.org/0000-0001-7680-3254
Giovanni Palladini https://orcid.org/0000-0001-5994-5138
Mario Nuvolone https://orcid.org/0000-0001-8334-1684

REFERENCES

1. Fermand JP, Bridoux F, Dispenzieri A, et al. Monoclonal gammopathy of clinical significance: a novel concept with therapeutic implications. *Blood*. 2018;132(14):1478-1485.
2. Perfetti V, Sassano M, Ubbiali P, et al. Inverse polymerase chain reaction for cloning complete human immunoglobulin variable regions and leaders conserving the original sequence. *Anal Biochem*. 1996;239(1):107-109.
3. Eid J, Fehr A, Gray J, et al. Real-time DNA sequencing from single polymerase molecules. *Science*. 2009;323(5910):133-138.
4. Duez M, Giraud M, Herbert R, Rocher T, Salson M, Thonier F. Vidjil: a web platform for analysis of high-throughput repertoire sequencing. *PLoS One*. 2016;11(11):e0166126.
5. Bender S, Javaugue V, Saintamand A, et al. Immunoglobulin variable domain high-throughput sequencing reveals specific novel mutational patterns in POEMS syndrome. *Blood*. 2020;135(20):1750-1758.
6. Garofalo M, Piccoli L, Romeo M, et al. Machine learning analyses of antibody somatic mutations predict immunoglobulin light chain toxicity. *Nat Commun*. 2021;12(1):3532.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Supplementary information

Supplement to

Single-Molecule Real-Time Sequencing of the M protein: toward personalized medicine in monoclonal gammopathies

Pasquale Cascino[1,2]*, Alice Nevone[1,2]*, Maggie Piscitelli[1,2], Claudia Scopelliti[1,2], Maria Girelli[1,2], Giulia Mazzini[1,2], Serena Caminito[1,2], Giancarlo Russo[3], Paolo Milani[1,2], Marco Basset[1,2], Andrea Foli[1,2], Francesca Fazio[4], Simona Casarini[1,2], Margherita Massa[2], Margherita Bozzola[1,2], Jessica Ripepi[1,2], Melania Antonietta Sesta[1,2], Gloria Acquafredda[5,6], Marica De Cicco[5,6], Antonia Moretta[5,6], Paola Rognoni[1,2], Enrico Milan[7], Stefano Ricagno[8,9], Francesca Lavatelli[1,2], Maria Teresa Petrucci[4], Enkelejda Miho[10,11,12], Catherine Klersy[13], Giampaolo Merlini[1,2], Giovanni Palladini[1,2#], Mario Nuvolone[1,2#]

[1] Department of Molecular Medicine, University of Pavia; Pavia, 27100; Italy.
[2] Amyloidosis Research and Treatment Center, Fondazione IRCCS Policlinico San Matteo; Pavia, 27100; Italy.
[3] EMBL partner institute for genome editing, Life Science Center, Vilnius University, Vilnius, Lithuania.
[4] Hematology, Department of Translational and Precision Medicine, Azienda Ospedaliera Policlinico Umberto I, Sapienza University of Rome; Rome, Italy.
[5] Pediatric Hematology Oncology Unit, Department of Maternal and Children's Health, Fondazione IRCCS Policlinico San Matteo; Pavia, Italy.
[6] Cell Factory and Center for Advanced Cellular Therapies, Department of Maternal and Children's Health, Fondazione IRCCS Policlinico San Matteo; Pavia, Italy.
[7] Age related Diseases Unit, Division of Genetics and Cell Biology, San Raffaele Scientific Institute and University Vita-Salute San Raffaele; Milano, Italy.
[8] Department of Biosciences, Università degli Studi di Milano; Milan, Italy.
[9] Institute of Molecular and Translational Cardiology, IRCCS Policlinico San Donato, San Donato Milanese, Milano, Italy.
[10] Institute of Medical Engineering and Medical Informatics, School of Life Sciences, University of Applied Sciences and Arts Northwestern Switzerland FHNW, Muttenz, Switzerland.
[11] SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland.
[12] aiNET GmbH, Basel, Switzerland.
[13] Clinical Epidemiology and Biometry Service, Foundation IRCCS Policlinico San Matteo; Pavia, Italy.
* Equally contributing authors
[#] Corresponding authors

Online Methods

Ethical statements

Clinical records and biological samples were from subjects referred to the Italian Amyloid Center – Fondazione IRCCS Policlinico San Matteo, Pavia, Italy, for diagnostic workout in the suspicion of systemic AL amyloidosis. Per the Declaration of Helsinki, all patients gave their written informed consent for the use of their clinical data and biological samples for research purposes, in agreement with the Institutional Review Board guidelines.

SMaRT M-Seq

Bone marrow (BM) mononuclear cells (MNCs) were obtained from diagnostic leftovers of BM aspirates of therapy-naïve patients with suspected AL amyloidosis through density gradient centrifugation. An aliquot of 10^7 BM-MNCs for immunoglobulin gene sequencing was pelleted, lysed in 1 mL of TRIzol Reagent (Ambion, Invitrogen), and stored at -80 °C until further processing.

Reverse transcription was performed starting from 500-1000 ng of RNA, using an Anchored Oligo $(dT)_{20}$ Primer (Invitrogen) and the SuperScript Double-Stranded cDNA synthesis kit (Invitrogen) according to the manufacturer's instructions. The resulting double-strand cDNA was then resuspended in 10 µL of RNase-free water. The circularization of double-strand cDNA was performed with 1 µL of T4 DNA ligase (1 U/µL) (Invitrogen), 2 µL of 5X Ligase Reaction Buffer (Invitrogen), 3 µL of double-strand cDNA and RNase-free water up to 10 µL of total reaction volume. The reaction was incubated overnight at 14 °C.

The ligated double-strand cDNA was used as input in an inverse PCR with primers targeting a conserved part of the constant region of the *IGKV*/*IGLV* gene. The inverse PCR was performed with 2 µL of the ligated double-strand cDNA, 12.5 µL of NEBNext Ultra II Q5® Master Mix DNA Polymerase (New England Biolabs) with 0.5 µM of forward and reverse primers (Universal κ-C_{LA} or Universal κ-

C_{LB} and Universal λ-C_{LA} and Universal λ-C_{LB} for *IGKV*/*IGLV*, respectively), in a final volume of 25 µL, using a Thermal Cycler C1000, CFX96 Real-Time System (Bio-Rad), at the following conditions: 30 sec at 98 °C; 15 cycles of 10 sec at 98 °C, 30 sec at 72 °C and 30 sec at 72 °C; 1 min at 72 °C (primer sequences in Fig. S2 and Table S1). To reduce the potential impact of base misincorporations during the first amplification cycles, each sample was amplified twice, in parallel, and the two PCR reaction products were pooled before proceeding with the second PCR.

Then, we performed a second PCR to incorporate sample-specific barcodes. This reaction was performed with 2 µl of pooled PCR product obtained from the first PCR, 12.5 µL of NEBNext Ultra II Q5® Master Mix DNA Polymerase (New England Biolabs) with 0.2 µM of Barcoded Universal Forward and Reverse primers from the 96 well- barcode plate (Pacific Biosciences), using a Thermal Cycler C1000, CFX96 Real-Time System (Bio-Rad), at the following conditions: 30 sec at 98 °C; 15 cycles of 10 sec at 98 °C, 30 sec at 72 °C and 30 sec at 72 °C; 1 min at 72 °C.

PCR products were then resolved through electrophoresis on 1% agarose gel, DNA bands of the expected size (700-800 bp) were cut and amplicons were purified using the Mini Elute Gel Extraction Kit (QIAGEN). Purified amplicons were then resubjected to electrophoresis on 1% agarose gel and DNA content was quantified using Qubit (Invitrogen). For each library preparation, identical amounts of barcoded amplicons from different samples were pooled to obtain 1 µg of pooled DNA. Sequencing libraries were prepared using the SMRTbell template prep kit (Pacific Biosciences) and sequenced on a PacBio RSII or on a PacBio Sequel platform at the Functional Genomic Center of Zurich (Switzerland) according to the manufacturer's instructions. Pooled samples were demultiplexed and deprived of SMRT adapter sequences using lima and high-accuracy, consensus reads were subsequently obtained applying Circular Consensus Sequencing (CCS), with minimum 5 passes and 99% accuracy. CCS reads were analyzed using Vidjil tool (http://www.vidjil.org/) with default parameters to identify a consensus clonal sequence for each sample.

Patients' enrollment

Ninety-one consecutive, therapy-naïve patients with suspected AL amyloidosis evaluated between March 2017 and July 2019, for which a BM aspiration was performed as part of the diagnostic workout and with diagnostic leftover of BM aspiration available for analysis, were included in this study.

The diagnosis of AL amyloidosis and the definition of organ involvement were based on clinical examination, echocardiographic, electrocardiographic, serological, cytological and histological analysis according to consensus criteria of the International Society of Amyloidosis[1,2].

The presence of an M protein was assessed by: 1) capillary electrophoresis of serum proteins, which was performed with a commercial CAPI 3 PROTEIN(E) 6 kit on a CAPILLARYS 3 TERA apparatus (Sebia); 2) a semiautomated serum and urine immunofixation electrophoresis, which was performed with a commercial Hydragel 2IF/BJ(HR) kit on a Hydrasys apparatus (Sebia); 3) the κ/λ ratio of serum free light chain (LC) concentrations, assessed by latex-enhanced nephelometry (Freelite assay™, The Binding Site), on a Behring BNII Nephelometer (Dade Behring)[3]. Reference limits for serum FLC concentrations are, respectively, 3.3-19.4 mg/L for κ and 5.7-26.3 mg/L for λ (reference κ/λ ratio interval 0.26–1.65)[4].

In selected cases where an M protein could not be identified with conventional methods, the presence of a monoclonal component was assessed also by high-resolution agarose gel electrophoresis/immunofixation on serum and urine samples obtained at the time of diagnosis. This procedure was performed as previously described[5,6].

To quantify the BM plasma cell (PC) compartment, BM (first aspirate) was collected in the presence of sodium citrate, smeared onto a glass slide and stained with May-Grünwald Giemsa according to standard techniques.

In selected cases, the presence of clonally restricted, immunophenotypically abnormal BMPC was investigated with standard or next generation flow cytometry on BM aspirates (second tube, collected in the presence of heparin) obtained at the time of diagnosis. For next generation flow cytometry, red blood cell lysis, staining, acquisition and analysis were performed using two-tube eight-colour panels following the EuroFlow guidelines[7], achieving a sensitivity of 10^{-6}.

To identify amyloid deposit, periumbilical fat tissue was aspirated using a fine needle and smeared onto glass slides, stained with Congo red as previously described[8] and examined with an Eclipse e600 microscope (Nikon) under polarized light. Amyloid deposits were identified based on the presence of the characteristic birefringence by an experienced operator. To type amyloid fibrils, immuno-electron microscopy was performed as previously described[9].

Cell culture

The human MM cell line NCI-H929 (RRID:CVCL_1600)[10] was a gift from Dr. Tonon, while the amyloidogenic PC line ALMC-2 (RRID:CVCL_M526)[11] was provided by Dr. Jelinek. These cell lines were authenticated through microsatellite analysis using PowerPlex 16 HS System (Promega) or through sequencing of the expressed LC gene, respectively. Negativity to *Mycoplasma spp*. contamination was verified with the EZ-PCR Mycoplasma test Kit (Resnova).

Conventional immunoglobulin gene sequencing with cloning and Sanger sequencing

To sequence the entire variable region of clonal immunoglobulin LCs using conventional Sanger sequencing, we used a previously described method based on double-stranded cDNA synthesis, circularization, inverse PCR, amplicon TA cloning, bacterial transformation and colony Sanger sequencing reported by Perfetti et al[12].

Library preparation and single molecule real-time DNA sequencing

The SMRT bell was produced using the SMRTbell Express Template Prep Kit 1.0 (for RS II sequencing) or 2.0 (for Sequel sequencing, all from Pacific Biosciences). The input amplicon pool concentration was measured using a Qubit Fluorometer dsDNA High Sensitivity assay (Life Technologies). A Bioanalyzer 12Kb assay (Agilent) was used to assess the amplicon pool size distribution and 400-800 ng of amplicon pool was DNA damage repaired and end-repaired using polishing enzymes. A blunt-end ligation was performed to create the SMRT bell template, according to the manufacturer's instructions. The SMRT bell library was quality inspected and quantified using a Bioanalyzer 12Kb assay (Agilent) and on a Qubit Fluorimeter (Life technologies) respectively. A ready to sequence SMRT bell-Polymerase Complex was created using the P6 DNA/Polymerase binding kit 2.0 (for RSII) or the Sequel Binding Kit or 3.0 (for Sequel) according to the manufacturer instructions. The Pacific Biosciences instrument was programmed to sequence the library on 1 SMRT cells v3 with 6-hour movie time (for RSII) or a on 1 Sequel SMRT Cell 1M v3 (PacBio), with 10-hour movie time (for Sequel), using the Sequel Sequencing Kit 3.0 (Pacific Biosciences). After the run, a sequencing report was generated via the SMRT portal, in order to assess the adapter dimer contamination, the sample loading efficiency, the obtained average read-length and the number of filtered sub-reads.

Amplicon short-read sequencing

RNA extraction and synthesis of ligated, double-strand cDNA was performed as for SMaRT M-Seq. The variable region of expressed κ light chains was amplified through an inverse PCR, using 2 µL of the ligated double-strand cDNA, 12.5 µL of NEBNext Ultra II Q5® Master Mix DNA Polymerase (New England Biolabs) with 0.5 µM of forward and reverse primers (κ-CLA and κ-CLB), in a final volume of 25 µL, using a Thermal Cycler C1000, CFX96 Real-Time System (Bio-Rad), at the following conditions: 30 sec at 98 °C; 35 cycles of 10 sec at 98 °C, 30 sec at 67 °C and 30 sec at 72 °C; 1 min at 72 °C. Primer

sequences are reported in Table S1. PCR products were then resolved through electrophoresis on 1% agarose gel, DNA bands of the expected size (700-800 bp) were cut and amplicons were purified using the Mini Elute Gel Extraction Kit (QIAGEN) following the manufacturer's protocol and were quantified using Qubit instrument (Invitrogen). An amount of 100 ng of each amplicon was used as starting material for NGS library construction with Nextera DNA Flex Library Prep (Illumina), according to the manufacturer's instructions, followed by library amplification using the DNA-BLT complex and IDT for Illumina DNA/RNA UD Indexes Set B (IDT-Illumina).

The concentration of eluted libraries and the library sizes were assessed using Qubit High Sensitivity dsDNA kit (Thermo Fisher Scientific) and the High Sensitivity Bioanalyzer chip (Agilent), respectively, and equal amounts of each library were pooled. Paired-end sequencing was performed on a MiSeq system (Illumina) with 502 cycles (251 bp PE sequencing) using a MiSeq Reagent Kit 34.

Immunoinformatics analyses

Initial analysis of the PacBio Continuous Long Reads (CLR) was performed using SMRT Tools v8.0 [https://www.pacb.com/wp-content/uploads/SMRT-Tools-Reference-Guide-v8.0.pdf]. Pooled samples were demultiplexed and deprived of SMRT adapter sequences using lima and high-accuracy, consensus reads were subsequently obtained applying Circular Consensus Sequencing (CCS), with minimum 5 passes and 99% accuracy. CCS reads were analyzed using Vidjil tool (http://www.vidjil.org/), an open-source platform for high-throughput analysis of V(D)J immune repertoire data from next-generation sequencing experiments. Through a seed-based method, the Vidjil algorithm detects sequences with V(D)J recombinations, clusters them into clonotypes based on a 50 bp nucleotide sequences at the V(D)J junction, and then performs a detailed V(D)J assignation after the clustering[13]. The web-based tool of Vidjil was employed selecting the Multi+inc+xxx configuration with default parameters for the visualization, inspection and analysis

of clones and their ranking based on clonotype size for each sample, in order to identify the dominant clone. The obtained clones were then visually inspected and the nucleotide sequence of the dominant clone for each sample was retrieved for further analyses. As reported in the original description of algorithms underlying Vidjil, its limit of clonal detection depends directly on the number of reads actually sequenced and analyzed[14]. In cases where Vidjil analysis performed on a sample from a patient with a known monoclonal gammopathy fails (alert: "few reads analyzed" or "very few reads analyzed"), CCS reads were analyzed with IMGT/HighV-QUEST[15,16] using the following parameters: "species" = "Homo sapiens (human)"; "Receptor type or locus" to be selected based on the sequenced isotype(s). IMGT/V-QUEST [17] was used to identify the V, D and J genes and alleles by alignment with germline gene and allele sequences of the IMGT reference directory, to verify functionality (lack of stop codons, frameshifts, alignment to pseudogenes) and to localize the somatic mutations of the rearranged sequences. Multiple sequence alignment through CLUSTALW was used to verify sequence unicity and homology across clonal sequences from different samples/patients[18]. In order to identify clonotypic sequences within the dilution series subjected to inverse PCR-based amplification and short read sequencing, the clone tracking function of Vidjil was employed and linearity was assessed based on linear regression analysis on the clonotype sizes obtained at the different dilutions[13].

Nano liquid chromatography – tandem mass spectrometry

Subcutaneous fat aspirates were acquired during the routine diagnostic procedures and stored unfixed at -80 °C until use. Samples were extensively washed in cold isotonic saline to remove blood contaminants, followed by manual homogenization in 50 µl of a buffer containing Urea 8 M / RapiGest^{TM} SF 0.5% (Waters, Milford, MA, USA)/ dithiothreitol 0.1 M (Acros Organics). The protein extract found below the floating lipid layer after centrifugation (13.200 rpm, 30 minutes, room

temperature) was quantified using Bio-Rad Protein Assay (Bio-Rad). Protein samples were alkylated with iodoacetamide 0.15 M (Millipore) for 1 hour at room temperature in the dark, and then diluted with five volumes of NH_4HCO_3 0.1 M pH 8.8 / 5% acetonitrile UHPLC-MS grade (VWR International), followed by digestion with sequencing grade modified trypsin (trypsin to protein ratio 1:20 w/w) (37 °C, overnight). Digestion was stopped by adding TFA 1%, followed by centrifugation (13200 rpm, 10 minutes, room temperature). The supernatant was carefully transferred to a new tube, followed by peptide purification on Tip C18 (Pierce) before liquid chromatography-mass spectrometry (LC-MS) analysis.

Nano flow liquid chromatography and tandem mass spectrometry (nLC-MS/MS) was performed on a Dionex Ultimate 3000 UHPLC system coupled to a Q Exactive mass spectrometer (Thermo Fisher Scientific) equipped with an EASY-spray ion source. Peptides were washed on a trap column (C18, 100 Å pore size, 3 um particle size, 0.3 x 5 mm) and then separated on the analytical column (Easy-Spray C18, 100 Å pore size, 3 um particle size, 0.075 x 500 mm) at a flow rate of 250 nl/min, using 0.1% formic acid as eluent A and 0.1% FA in acetonitrile as eluent B. Peptides were eluted by a gradient from 2 to 40% B in 70 minutes and from 40 to 95% B in 10 minutes, followed by a wash step in 95% B for 15 minutes. Full mass spectra were acquired in positive ionization mode at a resolution of 70000 full width at half maximum (FWHM), in scan range from 400 to 1500 m/z. Higher energy collisional induced fragmentation (HCD) of the ten most intense precursor ions with a charge state from 2^+ to 4^+ was performed using following parameters: resolution of 17500 FWHM, maximum injection time of 250 ms, automatic gain control of 50000 ions, 2 m/z isolation window, normalized collision energy of 27, dynamic exclusion of 30 s.

The obtained MS/MS spectra were processed using the Sequest algorithm included in the Proteome Discoverer software, version 2.0 (Thermo Fisher Scientific). The search was performed against the human proteome database (UniProt, August 2019, 74190 entries) supplemented with

immunoglobulin LC sequences from IMGT[19], AL-Base[20] and abYsis[21] databases, plus the LC sequences of interest. Enzymatic cleavage was set as semi tryptic for subcutaneous fat aspirates, with maximum 2 missing cleavage sites. Carbamidomethylation of cysteines as static modification and cyclization of N-terminal glutamine to form pyroglutamate as dynamic modification were specified. Precursor mass tolerance was 10 ppm and fragment mass tolerance was 0.03 Da. Peptide-spectrum matches (PSM) were validated using Percolator node, false discovery rate was 1%. Only high confidence peptides and master proteins were included in the results. Definition of the amyloidosis type in subcutaneous fat aspirates was performed as previously described[22].

Supplementary Data

Initial assessment of SMaRT M-Seq sensitivity, accuracy and repeatability

To initially assess SMaRT M-Seq sensitivity, accuracy and repeatability, we first used the human multiple myeloma (MM) cell line NCI-H929[ref.10], and the human amyloidogenic PC line ALMC-2[ref.11], after having determined or verified the sequence of the expressed κ or λ immunoglobulin LC, respectively (Fig. S3).

We spiked 1 volume of total RNA from either NCI-H929 or ALMC-2 human PCs in 9 volumes of total RNA from the BM of a subject lacking detectable B cell/PC clones, and further prepared 6 additional 1 to 10 serial dilutions into total RNA from the BM of the control subject (from 10^{-1} to 10^{-7}), so as to mimic BM samples containing a progressively smaller PC clone. This procedure resulted in 16 samples (8 for κ and 8 for λ LC sequencing, including dilutions from 10^{-1} to 10^{-7} and the unspiked healthy donor, each), which were subjected to amplification, barcoding, pooling (along with 10 additional samples, as specified below) and single-molecule real-time sequencing on the Pacific Biosciences RSII platform. Without *a priori* knowledge of the expected clonal sequence, Vidjil clonal analysis[13] could identify a dominant clonal sequence up to the 10^{-2} dilution for BM samples spiked with the NCI-H929 cells and subjected to κ LC sequencing, and up to the 10^{-3} dilution for BM samples spiked with the ALMC-2 cells and subjected to λ LC sequencing (Fig. S4).

In the 10^{-3} dilution subjected to κ LC sequencing, 3 nucleotide reads matching the expected clonal sequence were identified (with 100% identity), even if these reads were not sufficient to qualify as a clonal sequence in the Vidjil analysis (Fig. S4). Of note, the dominant clonal sequence identified by the Vidjil analysis showed, in all cases, 100% identity with the expected nucleotide sequence of the immunoglobulin LC genes expressed by NCI-H929 and ALMC-2 cells, respectively, demonstrating the accuracy of clonal sequence determination. Additionally, the obtained molecular clonal size, which

is a measure of the relative frequency of clonal reads in a given sample, progressively decreased in samples with decreasing concentration of the spiked RNA from clonal cells, as expected (Fig. S4).

In parallel, to test the repeatability of SMaRT M-Seq for the detection of clonal immunoglobulin genes in *bona fide* BM samples from patients with a PC dyscrasia, we selected two patients with AL amyloidosis at diagnosis (patient 01 and 02). The nucleotide sequence of the variable region of their clonal κ or λ LC, respectively, was obtained through a standard approach consisting of inverse PCR amplification, TOPO-TA cloning, *E. coli* transformation and multiple colony sequencing[12] (Fig. S5). Next, we split these 2 BM RNA samples into 5 separate tubes each, and we processed the resulting 10 samples independently. Pooling and sequencing for these 10 samples was performed together with the 16 samples originating from the above-mentioned serial dilution experiment. Of note, for both the κ- and the λ-expressing AL patient, the same dominant sequence was obtained in all five cases, with 100% identity with the expected clonal κ and λ sequence as assessed by conventional methods, and a similar molecular clonal size (89.0%, with coefficient of variation, CV, of 0.5% for patient 01 and 92.9%, with CV of 0.7% for patient 02) (Fig. S4).

Collectively, these results indicated that SMaRT M-Seq can accurately and reproducibly identify the entire sequence of the variable region of dominant clonal immunoglobulin genes starting from bulk BM-MNCs.

Sensitive, clonotypic reads detection exploiting full-length variable light chain sequence knowledge

Using NCI-H929 cells and a control BM, we prepared 9 BM samples (dilutions from 10^{-1} to 10^{-8} and the unspiked control sample) which were subjected to inverse PCR-based amplification, barcoding, pooling and paired-end short-read sequencing on an Illumina MiSeq platform with a 1 M cartridge.

Samples were analyzed using the clone tracking function of Vidjil[13]. Within this experiment, a clonotypic sequence was identified up to the 10^{-4} dilution, with a progressively decreasing molecular clonal size (Fig. 1B). Worthy of note, the identified clonotypic sequence (spanning from the 5' terminal part of the complementarity-determining region, CDR1, through the framework region 4, FR4) showed 100% sequence identity with respect to the corresponding region of the κ LC sequence of NCI-H929 cells. To verify the possibility of detecting the clonotypic sequence at higher dilutions, we then subjected selected samples from this dilution series to two novel rounds (10^{-5} and 10^{-6} in the second round and 10^{-7} and the unspiked control sample in the third round) of inverse PCR-based amplification, barcoding, pooling and paired-end short-read sequencing on an Illumina MiSeq platform, this time employing a 25 M cartridge per run. Within these two novel rounds of sequencing, a clonotypic sequence was identified up to the 10^{-7} dilution, with a molecular clonal size which was progressively smaller and in line with the decreasing clonal size detected in the 10^{-1}-10^{-4} dilution series from the first sequencing round (Fig. 1C). Again, the identified clonotypic sequence showed 100% sequence identity with respect to the corresponding region of the κ LC sequence of NCI-H929 cells.

Supplementary Tables

Table S1. Primer sequences

Primer	Sequence primer (5' → 3')
K-C_{LA}	TGCTCATCAGATGGCGGGAA
K-C_{LB}	AAGAGCTTCAACAGGGGAGA
λ-C_{LA}	AGTGTGGCCTTGTTGGCTTG
λ -C_{LB}	GTCACGCATGAAGGGAGCAC
Universal κ-C_{LA}	/5AmMC6/GCAGTCGAACATGTAGCTGACTCAGGTCACTGCTCATCAGATGGCGGGAA
Universal κ-C_{LB}	/5AmMC6/TGGATCACTTGTGCAAGCATCACATCGTAGAAGAGCTTCAACAGGGGAGA
Universal λ-C_{LA}	/5AmMC6/GCAGTCGAACATGTAGCTGACTCAGGTCACAGTGTGGCCTTGTTGGCTTG
Universal λ-C_{LB}	/5AmMC6/TGGATCACTTGTGCAAGCATCACATCGTAGGTCACGCATGAAGGGAGCAC

/5AmMC6/: 5' Amino Modifier C6; K-C_{LA}, K-C_{LB}, λ-C_{LA} and λ-C_{LB} are from ref. [12].

Table S2. Overview of SMaRT M-Seq validation

	Evidence	Figure
Accuracy	100% sequence identity with respect to clonal sequence identified through standard cloning and sequencing approaches in:	
	• 10^{-1}-10^{-3} points of serial dilution of NCI-H929 or ALMC-2 cell line spiked in control BM (6 samples)	• Figure S4A
	• 5 replicates from Pt. 01 and 02 (10 samples)	• Figure S4B
	• Duplicates from Pt. 22, 37 and 38 (6 samples)	• Figure S7
	• Pt. 39, 40 and 73 (3 samples)	• Figure S7
	Mapping of proteomics data on matched amyloid-positive fat tissues analyzed with mass spectrometry for Pt. 01, 02, 38 and 44 (4 samples)	• Figure S14
	Clonal germline gene usage in line with clonal germline gene usage as assessed by mass spectrometry on the largest series of AL amyloidosis patients described in the literature[#]	• Figure S12
Sensitivity	Investigation of serial dilution (10^{-1}-10^{-7}) of NCI-H929 or ALMC-2 cell line spiked in control BM (16 samples) with a clonal sequence identified up to the 10^{-3} dilution (on a RSII platform with 26 samples sequenced in parallel)	• Figure S4A
	Investigation of a cohort of 86 patients with AL amyloidosis (including 5 cases with negative M protein studies) with a clonal sequence identified in all cases (on a Sequel platform with 92 samples sequenced in parallel)	• Figure 1A
Repeatability	Repeatability of sequence determination in:	
	• 10^{-1}-10^{-3} points of serial dilution of NCI-H929 or ALMC-2 cell line spiked in control BM (6 samples)	• Figure S4A
	Repeatability of sequence and molecular clonal size determination in:	
	• 5 replicates from Pt. 01 and 02 (10 samples)	• Figure S4B
	• Duplicates from Pt. 22, 37 and 38 (6 samples)	• Figure S7
Throughput	Analysis of 26 samples in parallel (RS II platform)	• Figure S4
	Analysis of 92 samples in parallel (Sequel platform)	• Figure 1A

Kourelis TV, Dasari S, Theis JD, et al. Clarifying immunoglobulin gene usage in systemic and localized immunoglobulin light-chain amyloidosis by mass spectrometry. Blood. 2017;129(3):299-306.

Table S3. Clinical characteristics of the 86 patients with AL amyloidosis included in this study

	N (%) / median (IQR) (N=86)°
Male sex	53 (62)
Age, yrs	68 (60-74)
% BMPC	9 (6-13)
dFLC, mg/L	176 (75-370)
Organ	
Heart	62 (72)
Kidney	46 (53)
Soft tissues	34 (40)
Liver	8 (9)
Cardiac stage	
I	12 (15)
II	38 (46)
IIIa	17 (21)
IIIb	15 (18)
Renal stage	
I	49 (57)
II	25 (29)
III	12 (14)
M protein type	
IgAκ / IgAλ	2 / 14

IgDκ / IgDλ	1 / 0
IgGκ / IgGλ	4 / 21
IgMκ / IgMλ	1 / 1
Biclonal (IgAλ + IgGλ)	1
LC only κ / λ	9 / 32

BMPC%: bone marrow plasma cell percentage; dFLC: difference between involved and uninvolved serum free light chain concentration; IQR: interquartile range; yrs: years.
° The 86 AL patients reported here include AL patient 1 and 2 reported in Fig. S4 and the 84 AL patients reported in Fig. 1 (and related supplementary figures)

Table S4. Accession codes of immunoglobulin light chains identified through proteomic analysis

Accession code	Figure code	Patient
ABA70804.1	*IGKV1-05 #1*	Pt. 01
CAA79289.1	*IGKV1-05 #2*	Pt. 01
047269	*IGKV1-05 #3*	Pt. 01
CAA09183.1	*IGKV1-05 #4*	Pt. 01
0709220A	*IGKV1-05 #5*	Pt. 44
AAZ09110.1	*IGKV1-05 #6*	Pt. 44
CAR62893.1	*IGKV1-08 #1*	Pt. 01
A0A0C4DH73	*IGKV1-12 #1*	Pt. 01
ABU90716.1	*IGKV1-16 #1*	Pt. 01
CAJ13417.1	*IGKV1-27 #1*	Pt. 01
P01593	*IGKV1-33 #1*	Pt. 01
AAD09367.1	*IGKV1-33 #2*	Pt. 01
CAA67400.1	*IGKV1-33 #3*	Pt. 01
AAA70229.1	*IGKV1-33 #4*	Pt. 01
751419A	*IGKV1-33 #5*	Pt. 44
AAO11846.2	*IGKV1-39 #1*	Pt. 01
ABB55183.1	*IGKV1-39 #2*	Pt. 01
ABU90636.1	*IGKV1-39 #3*	Pt. 01
ABA71426.1	*IGKV1-39 #4*	Pt. 38
1HEZ_A	*IGKV1-39 #5*	Pt. 44
A0A087WW87	*IGKV2-40 #1*	Pt. 01
AOO95245.1	*IGKV3-20 #1*	Pt. 01

AAZ09095.1	*IGKV3-20 #2*	Pt. 01
AGG22556.1	*IGKV3-20 #3*	Pt. 01
004899	*IGKV3-20 #4*	Pt. 01
040103	*IGKV3-20 #5*	Pt. 01
P01619	*IGKV3-20 #6*	Pt. 01, Pt. 02, Pt. 44
A0A0C4DH25	*IGKV3D-20 #1*	Pt. 01
035923	*IGKV4-01 #1*	Pt. 01
P06312	*IGKV4-01 #2*	Pt. 01
CAI99826.1	*IGKV4-01 #3*	Pt. 44
AAA19493.1	*IGKV4-01 #4*	Pt. 44
751423A	*IGKV4-01 #5*	Pt. 44
ABU90616.1	*IGLV1-36 #1*	Pt. 38
1BJM_A	*IGLV1-36 #2*	Pt. 44
038783	*IGLV1-40 #1*	Pt. 01
AAZ32305.1	*IGLV1-40 #2*	Pt. 01
046681	*IGLV1-40 #3*	Pt. 02
AAZ13626.1	*IGLV1-40 #4*	Pt. 02
CAI99647.1	*IGLV1-40 #5*	Pt. 38
CAE18194.1	*IGLV1-40 #6*	Pt. 38
AKF02486.1	*IGLV1-40 #7*	Pt. 38
CAI99733.1	*IGLV1-40 #8*	Pt. 38
040329	*IGLV1-40 #9*	Pt. 38
AAD16697.1	*IGLV1-40 #10*	Pt. 38
044132	*IGLV1-40 #11*	Pt. 44

A42193	*IGLV1-44 #1*	Pt. 01
ABA71654.1	*IGLV1-44 #2*	Pt. 02
CAE18197.1	*IGLV1-44 #3*	Pt. 02, Pt. 38
040322	*IGLV1-44 #4*	Pt. 02
043768	*IGLV1-44 #5*	Pt. 02
044092	*IGLV1-44 #6*	Pt. 38
ABA70907.1	*IGLV1-44 #7*	Pt. 38
CCN25611.1	*IGLV1-44 #8*	Pt. 38
CAC06662.1	*IGLV1-44 #9*	Pt. 38
CAI99645.1	*IGLV1-44 #10*	Pt. 38
CCN25613.1	*IGLV1-44 #11*	Pt. 38
ABA71522.1	*IGLV1-44 #12*	Pt. 38
ABU90619.1	*IGLV1-44 #13*	Pt. 38
AEM45945.1	*IGLV1-44 #14*	Pt. 38
BAH04817.1	*IGLV1-44 #15*	Pt. 44
AAD16714.2	*IGLV1-47 #1*	Pt. 38
AAD16781.1	*IGLV1-47 #2*	Pt. 38
CAI99658.1	*IGLV1-47 #3*	Pt. 38
ABA71601.1	*IGLV1-47 #4*	Pt. 38
007540	*IGLV1-47 #5*	Pt. 38
P01701	*IGLV1-51 #1*	Pt. 01
007529	*IGLV2-08 #1*	Pt. 02
AAF20404.1	*IGLV2-08 #2*	Pt. 44
CAQ15793.1	*IGLV2-08 #3*	Pt. 44

ABA71589.1	*IGLV2-08 #4*	Pt. 44
AAF20427.1	*IGLV2-08 #5*	Pt. 44
BAC85194.1	*IGLV2-08 #6*	Pt. 44
025363	*IGLV2-08 #7*	Pt. 44
AAF20744.1	*IGLV2-08 #8*	Pt. 44
AAZ13713.1	*IGLV2-11 #1*	Pt. 44
CAC29409.1	*IGLV2-11 #2*	Pt. 44
044053	*IGLV2-14 #1*	Pt. 02
AAD16770.1	*IGLV2-14 #2*	Pt. 02
AAF20466.1	*IGLV2-14 #3*	Pt. 02
ABU90704.1	*IGLV2-14 #4*	Pt. 02, Pt. 44
AAA59006.1	*IGLV2-14 #5*	Pt. 02
007455	*IGLV2-14 #6*	Pt. 02
007526	*IGLV2-14 #7*	Pt. 02
044051	*IGLV2-14 #8*	Pt. 02
CAI99678.1	*IGLV2-14 #9*	Pt. 44
AAF20685.1	*IGLV2-14 #10*	Pt. 44
031833	*IGLV2-14 #11*	Pt. 44
AAF20757.1	*IGLV2-14 #12*	Pt. 44
043965	*IGLV2-14 #13*	Pt. 44
AEM45978.1	*IGLV2-14 #14*	Pt. 44
AAF20451.1	*IGLV2-14 #15*	Pt. 44
044049	*IGLV2-14 #16*	Pt. 44
ABU90729.1	*IGLV2-14 #17*	Pt. 44

BAC01788.1	*IGLV2-14 #18*	Pt. 44
AAO11880.1	*IGLV2-14 #19*	Pt. 44
CAB56050.1	*IGLV2-14 #20*	Pt. 44
ABU90725.1	*IGLV2-14 #21*	Pt. 44
ABU90593.1	*IGLV2-14 #22*	Pt. 44
CAA40960.1	*IGLV2-14 #23*	Pt. 44
AAF20468.1	*IGLV2-18 #1*	Pt. 02
CAI99688.1	*IGLV2-18 #2*	Pt. 02
AAF20752.1	*IGLV2-18 #3*	Pt. 38
AAZ13728.1	*IGLV2-23 #1*	Pt. 02
AAO11875.1	*IGLV2-23 #2*	Pt. 02
043696	*IGLV2-23 #3*	Pt. 44
AAD16661.1	*IGLV2-23 #4*	Pt. 44
1LGV_A	*IGLV2-23 #5*	Pt. 44
AAD16651.1	*IGLV3-01 #1*	Pt. 02
AAC16876.1	*IGLV3-01 #2*	Pt. 02
044125	*IGLV3-19 #1*	Pt. 01
AAB36577.1	*IGLV3-19 #1*	Pt. 02
BAH04820.1	*IGLV3-19 #2*	Pt. 02
AAN15191.1	*IGLV3-19 #3*	Pt. 02
ABU90721.1	*IGLV3-19 #4*	Pt. 02
S02083	*IGLV3-19 #5*	Pt. 44
AAC16828.1	*IGLV3-21 #1*	Pt. 02
043643	*IGLV3-21 #2*	Pt. 02

CAJ75541.1	*IGLV3-21 #3*	Pt. 44
CAE18255.1	*IGLV10-54 #1*	Pt. 02
CAE18258.1	*IGLV10-54 #2*	Pt. 02

Accession code according to NCBI, UniProt or abYsis. Pt: patient.

Supplementary Figures

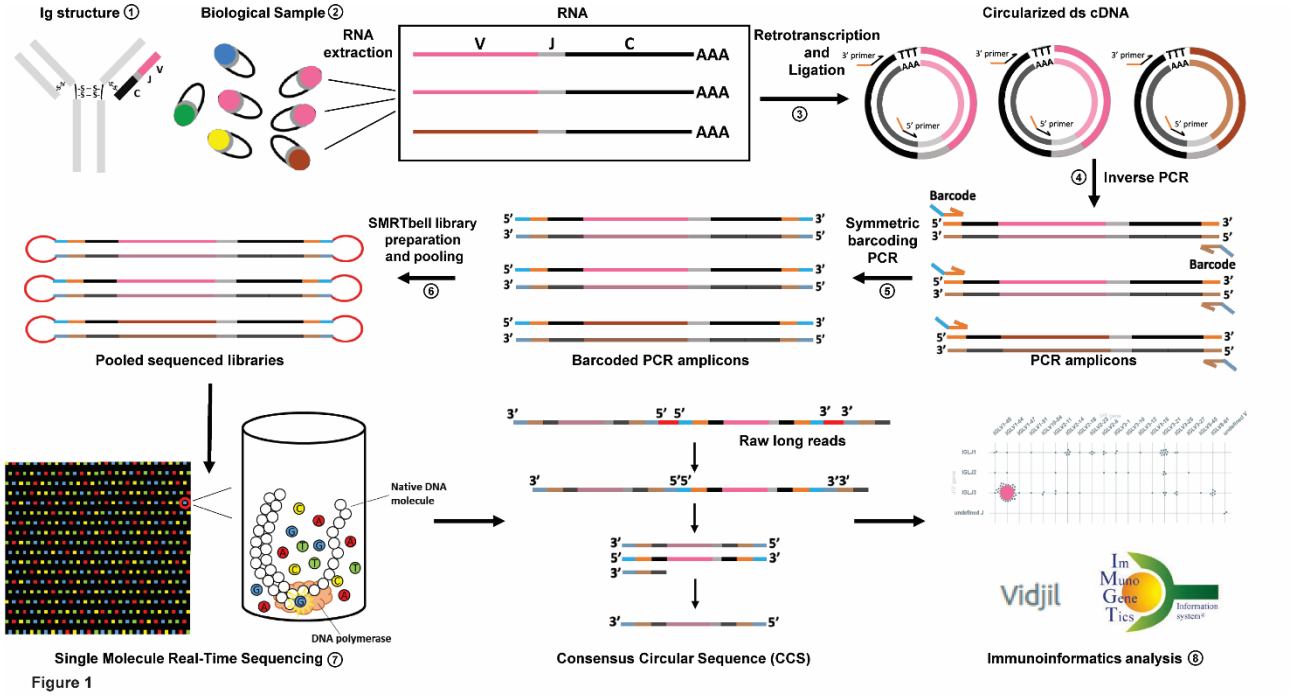

Figure S1. Single Molecule Real-Time Sequencing of the M protein (SMaRT M-Seq)

Schematic representation of the workflow underlying SMaRT M-Seq. The purpose of the assay is to obtain the full-length sequence of the variable region of immunoglobulin (Ig) light (or heavy) chains (①) and to rank the obtained sequences based on their relative abundance, so as to identify potential dominant (clonal) sequences, if a B cell or plasma cell clone is present in the biological sample (e.g. bone marrow- or peripheral blood-derived mononuclear cells) (②). Total RNA is extracted from the biological sample, mRNA is retrotranscribed using an anchored oligo-dT and double stranded complementary DNA (ds cDNA) is synthetized and circularized (③). Two primers (in black) annealing to the constant region of the isotype of interest and containing an adaptor sequence (in orange) are used in the context of an inverse PCR using a high-fidelity DNA polymerase to obtain an amplicon comprising the entire variable region (④). Two primers (in orange) annealing to the adaptor sequence and containing sample ID barcodes (cyan) are used to generate barcoded amplicons (⑤) which are used for library preparation with bell-adaptors (⑥) and subjected to single-molecule real-time sequencing (⑦). Computational methods are employed to analyze raw long reads and to extract circular consensus sequences (CCS). Immunoinformatics analyses, including Vidjil and IMGT/HighV-QUEST, are used to scrutinize repertoires and identify dominant clones (⑧).

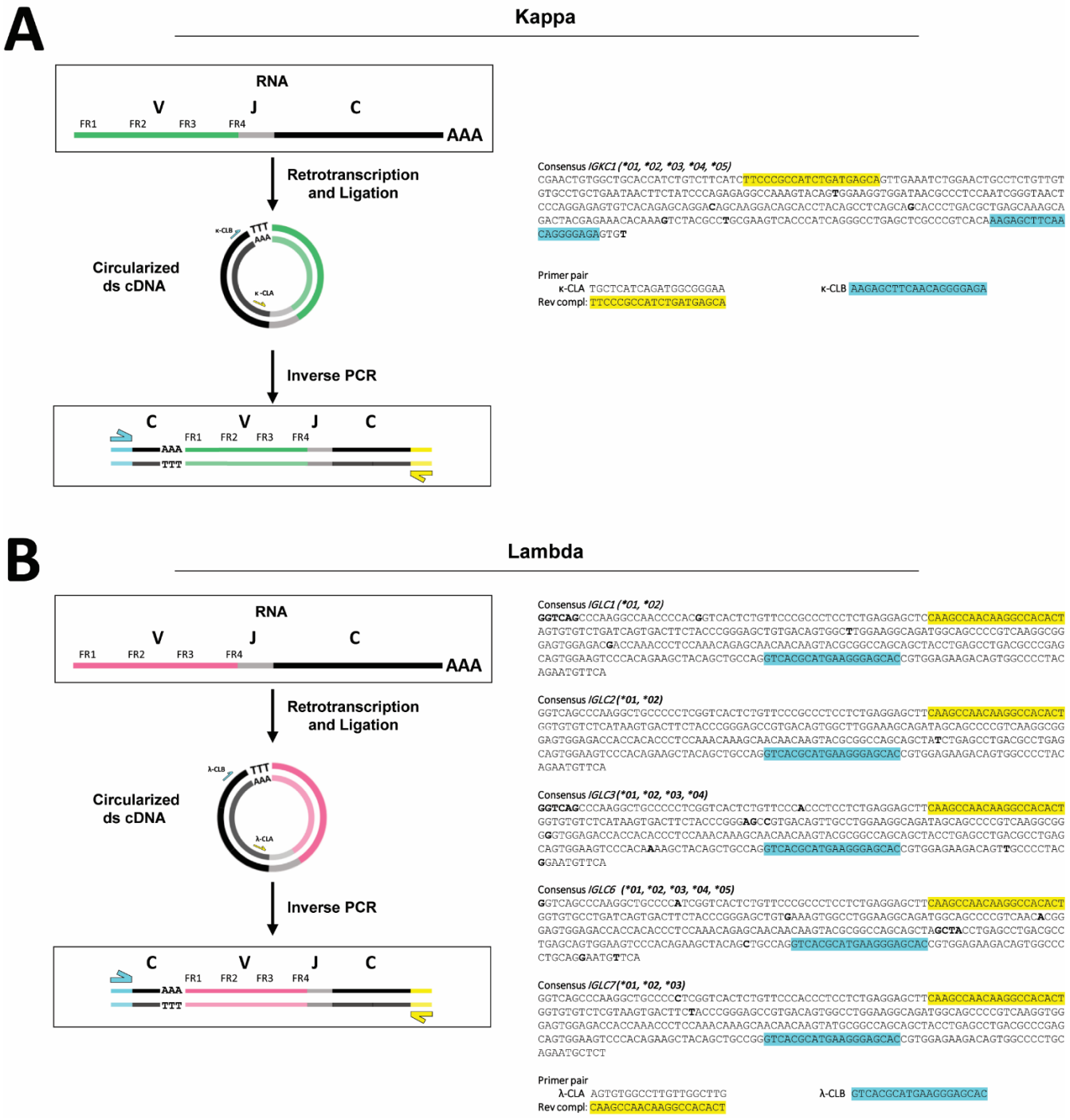

Figure S2. Insight of primers and primer binding regions of the inverse PCR

On the left: Scheme of the employed inverse PCR. Messenger RNA is retrotranscribed using an anchored oligo-dT and double stranded complementary DNA (ds cDNA) is synthesized and circularized. To obtain an amplicon comprising the entire variable region without *a priori* knowledge of the rearranged germline V genes, amplification is carried out on ds cDNA in the context of an inverse PCR, using two primers (in cyan or yellow) that anneal to a conserved, homologous part of the constant regions of the isotype of interest (κ (**A**) or λ (**B**) isotype). On the right: consensus sequences of κ (**A**) and λ (**B**) constant germline reference genes are shown. Nucleotide positions where at least one gene/allele shows a different base with respect to the consensus sequence are

depicted in bold. All nucleotide sequences are reported as 5'→3'. The universal adaptor sequence incorporated in the 5' region of each primer in the context of SMaRT M-Seq is not represented. V: variable region; J: joining region; C: constant region; FR: framework region. Rev. compl.: reverse complement.

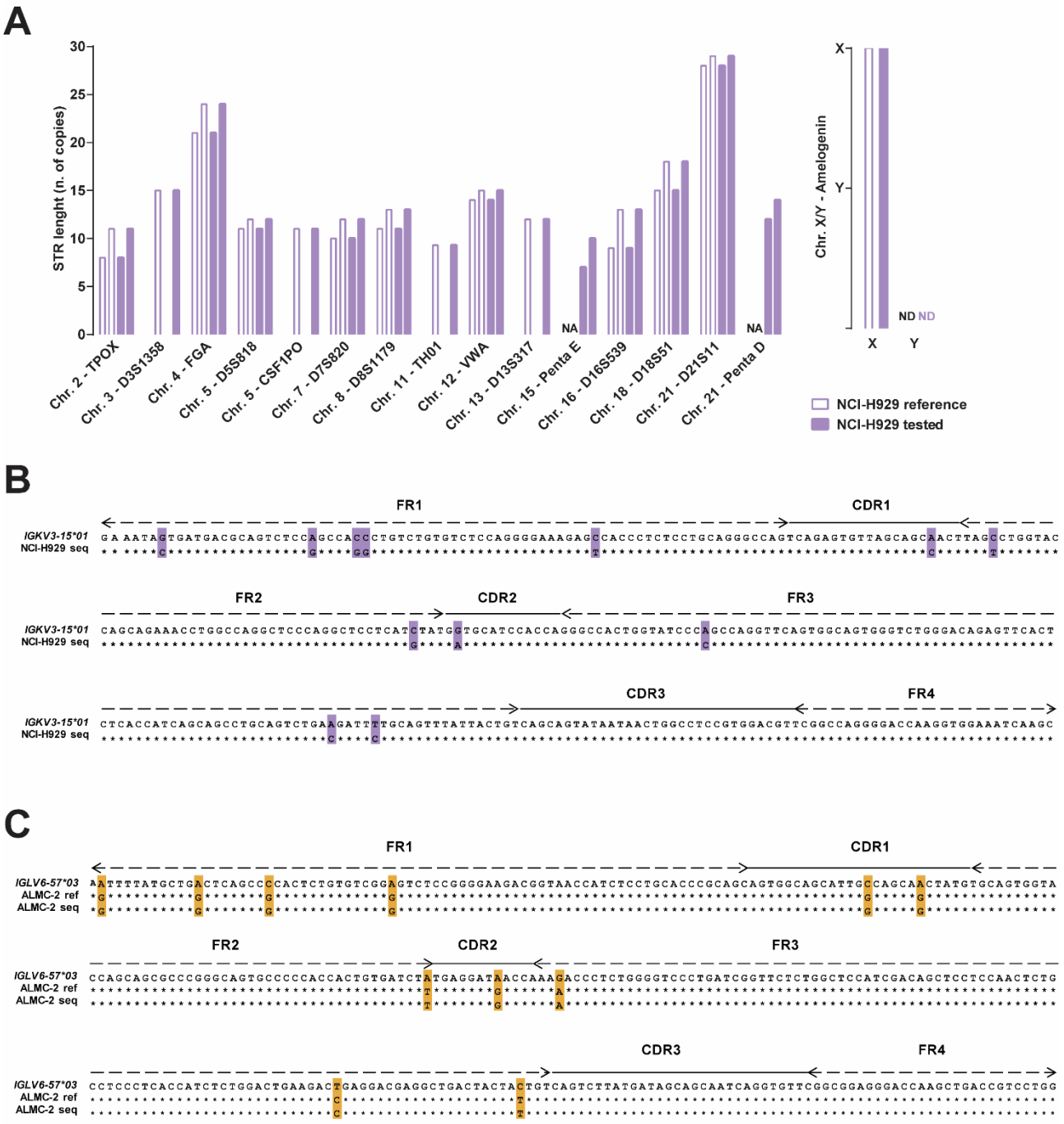

Figure S3. NCI-H929 and ALMC-2 authentication and sequencing of the expressed immunoglobulin light chain

(**A**) Short tandem repeat (STR) profiling of NCI-929 used in this study (NCI-H929 tested), in comparison with reference STR profile for this cell line (according to ATCC, CCRID, Cosmic-CLP, DSMZ, ECACC) (NCI-H929 reference). Chr.: chromosome; NA: not available; ND: not detected. (**B**) Sequence alignment of the variable region of the κ light chain sequence expressed by NCI-H929 cells sequenced in this study (NCI-H929 seq) with the corresponding germline gene (*IGKV3-15*01*). Nucleotide identity is indicated with an asterisk while nucleotide mismatches are highlighted. (**C**) Sequence alignment of the variable region of the λ light chain sequence expressed by ALMC-2 cells sequenced in this study (ALMC-2 seq) with the expected sequence of this cell line according to the original description of this cell line[11] and with the corresponding germline gene (*IGLV6-57*03*).

Nucleotide identity is indicated with an asterisk while nucleotide mismatches are highlighted. FR: framework region; CDR: complementarity determining region.

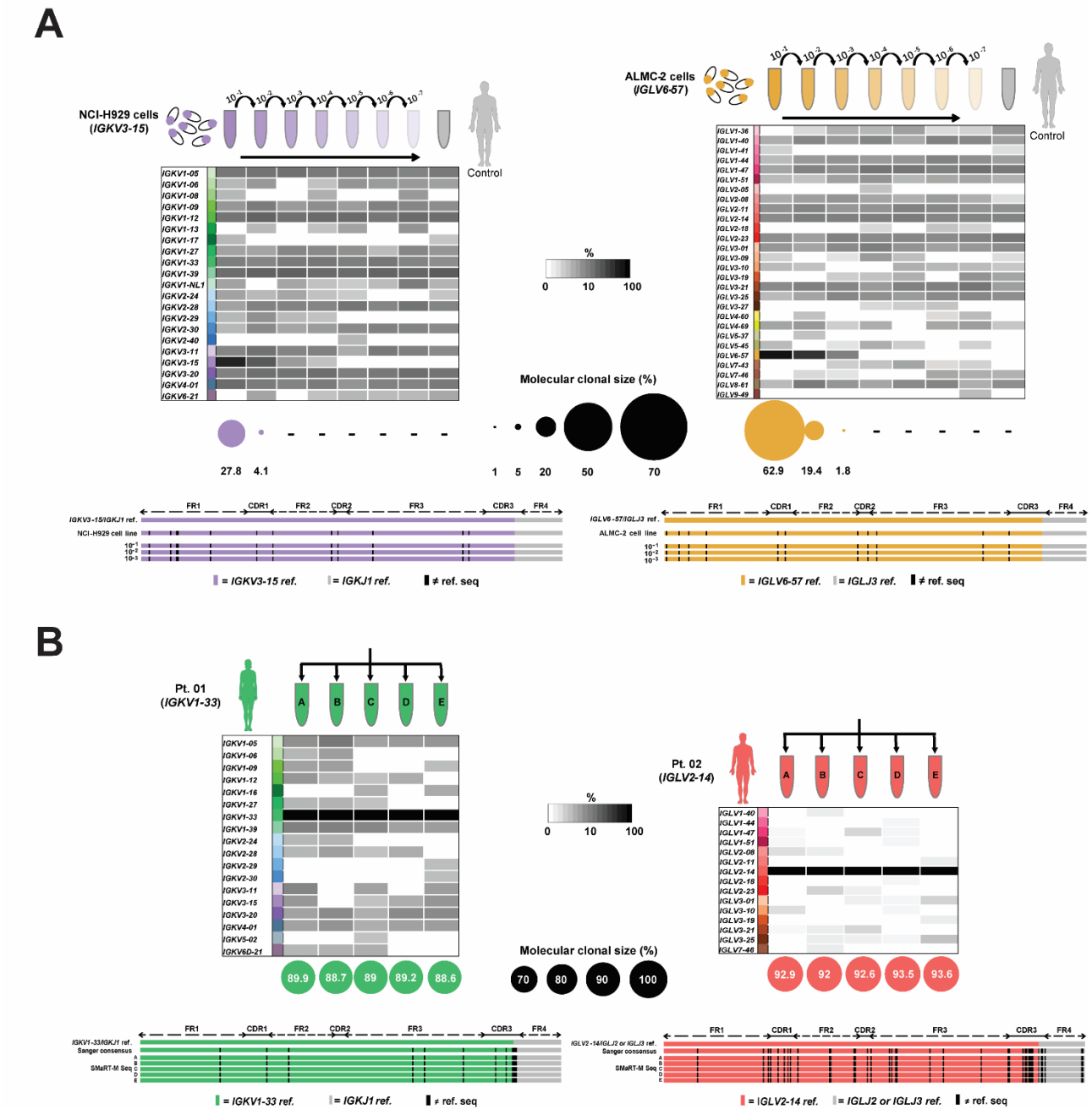

Figure S4. Initial assessment of SMaRT M-Seq sensitivity, accuracy and repeatability

(**A**) Expression levels (in shades of grey) of different *IGKV* (left) and *IGLV* (right) germline genes (each denoted by a distinctive colour) as assessed by SMaRT M-Seq, starting from serial Log_{10} dilutions of RNA from *IGKV3-15*-secreting NCI-H929 cells (left) or *IGLV6-57*-secreting ALMC-2 cells (right) into RNA of the bone marrow from a control subject (unspiked control sample in grey). (**B**) Expression levels of different *IGKV* (left) and *IGLV* (right) germline genes as assessed by SMaRT M-Seq, starting from five replicate bone marrow samples (A to E) from two patients (Pt. 01 and 02) affected by AL amyloidosis, with a plasma cell clone secreting an *IGKV1-33* (left) or an *IGLV2-14* (right) clonal light chain. In both **A** and **B**, scaled pie charts denote the molecular clonal size of the dominant clone identified in each tested sample (the corresponding *IGKV* or *IGLV* germline gene is indicated with the pie chart colour). Minus sign (-) indicates samples where no dominant clone could be identified

with Vidjil. At the bottom, sequence alignments of the clonotypic variable region of the light chain secreted by NCI-H929 and ALMC-2 cells (**A**) or of the clonal light chain from patient 01 and patient 02 (**B**), as assessed by cloning and Sanger sequencing (Sanger) or by SMaRT M-Seq (denoted by the corresponding dilution, **A**, or replicate label, **B**), with the corresponding *IGKV*/*IGKJ* or *IGLV*/*IGLJ* germline genes (ref., with the J gene in grey and the V gene denoted by the corresponding colour). Black tick denotes sequence mismatches (≠ ref. seq.) in the clonotypic light chain with respect to the corresponding germline genes. FR: frame work region; CDR: complementarity determining region; Pt.: patient.

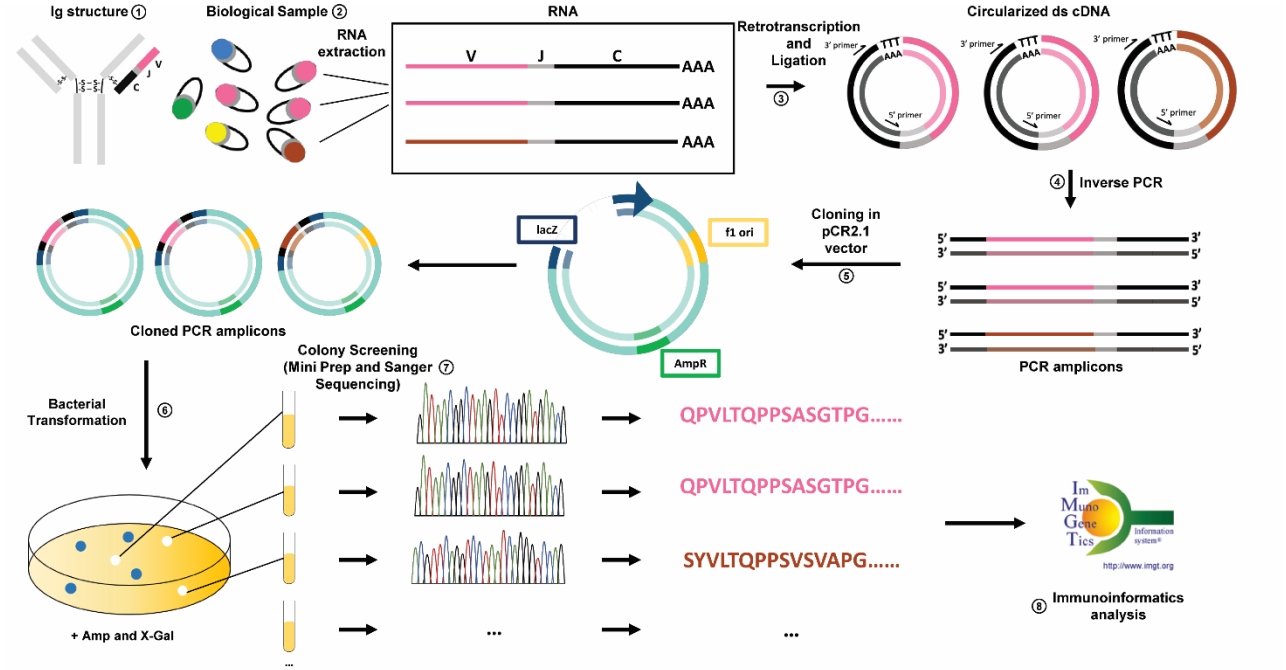

Figure S5. Identification of clonal immunoglobulin variable sequence based on cloning and Sanger sequencing

Schematic representation of the workflow enabling the identification of clonal immunoglobulin (Ig) variable sequence based on cloning and Sanger sequencing. The purpose of the assay is to identify the full-length sequence of the variable region of immunoglobulin light (or heavy) chains (①), so as to identify potential dominant (clonal) sequences, if a B cell or plasma cell clone is present (②). Total RNA is extracted from the biological sample (e.g., bone marrow derived mononuclear cells), mRNA is retrotranscribed using a standard oligo-dT and double stranded complementary DNA (ds cDNA) is synthetized and circularized (③). Two primers (in black) annealing to the constant region of the isotype of interest are used in the context of an inverse PCR using a conventional Taq polymerase to obtain an amplicon comprising the entire variable region (④). The resulting amplicons are cloned in a pCR2.1 vector (⑤), followed by *E.coli* transformation (⑥) and plating in the presence of ampicillin (Amp) and X-Gal for white-blue colony screening. White colonies are then picked, grown, and extracted plasmid DNA is subjected to Sanger sequencing (⑦). The resulting sequences are analyzed with IMGT/V-QUEST to identify a consensus dominant sequence (⑧).

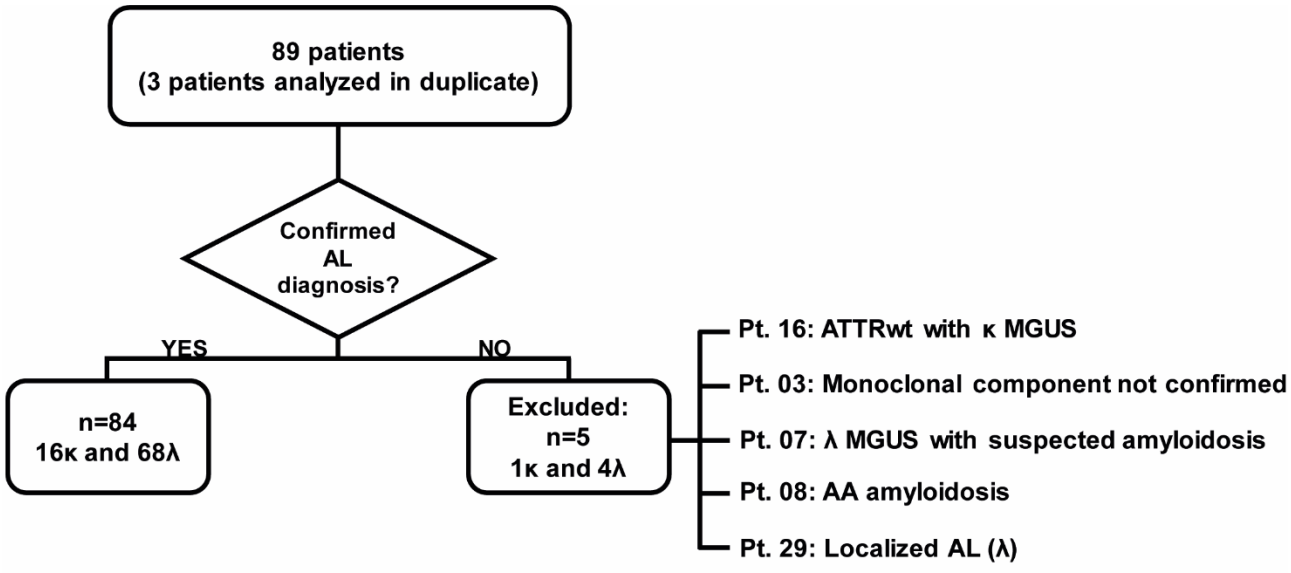

Figure S6. Patients' enrollment

Flow chart depicting the 89 patients subjected to SMaRT M-Seq in parallel in one sequencing round. Three cases were analyzed in duplicates (for a total of 92 samples). A confirmed, biopsy-proven final diagnosis of systemic AL amyloidosis was used to include (n=84) or exclude (n=5) patients. Final diagnosis of excluded patients (if available) is indicated. Pt.: patient; AA: amyloid A amyloidosis; ATTRwt: wild-type transthyretin-related amyloidosis (formerly senile cardiac amyloidosis or senile systemic amyloidosis); MGUS: monoclonal gammopathy of undetermined significance.

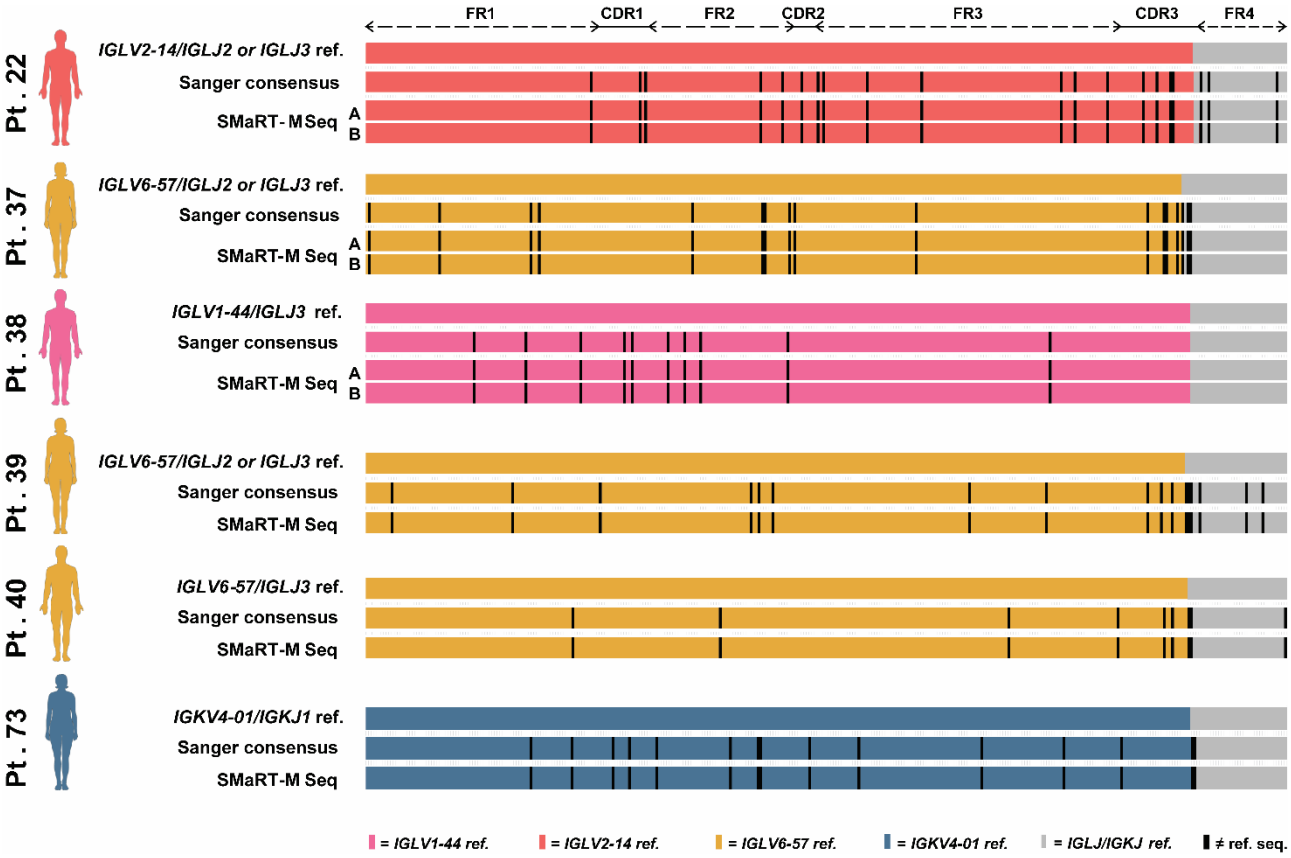

Figure S7. Accuracy and repeatability of SMaRT M-Seq performed on a cohort of 89 patients analyzed in parallel

Sequence alignments of the clonal light chain from six patients (from the cohort of 89 patients analyzed in parallel) as assessed by cloning and Sanger sequencing (Sanger) or by SMaRT M-Seq (A-B indicate technical duplicates), with the corresponding *IGKV*/*IGKJ* or *IGLV*/*IGLJ* germline genes (ref., with the J gene in grey and the V gene denoted by the corresponding colour). Black tick denotes sequence mismatches (≠ ref. seq.) in the clonotypic light chain with respect to the corresponding germline genes. FR: frame work region; CDR: complementarity determining region; Pt.: patient.

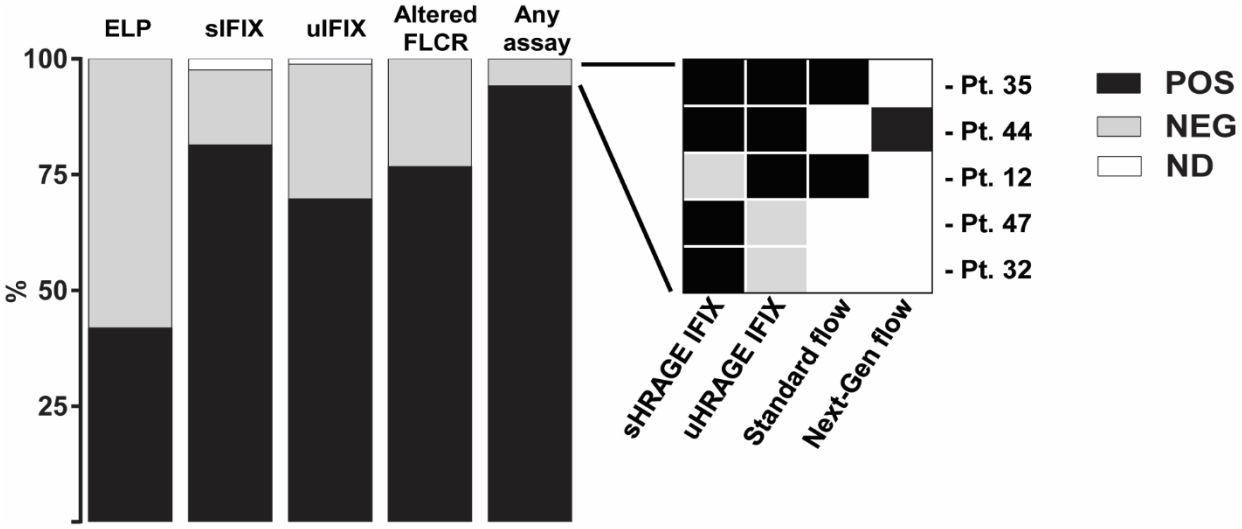

Figure S8. Results of M protein studies in the cohort of AL patients

The bar graph indicates results of individual, standard M protein studies in the cohort of AL patients examined in this study or of the combination thereof (Any assay). Inset: For the five patients with negative standard M protein studies, the grid shows results of second-level M protein studies leading to the identification of a monoclonal gammopathy. Results are denoted as follows: POS: positive (black); NEG: negative (grey); ND: not done (white). ELP: serum protein electrophoresis; sIFIX: serum immunofixation; uIFIX: urine immunofixation; FLCR: serum free light chain ratio; sHRAGE IFIX: serum high-resolution agarose gel electrophoresis with immunofixation; uHRAGE IFIX: urine high-resolution agarose gel electrophoresis with immunofixation; standard flow: standard flow cytometry on bone marrow; Next-Gen flow: next-generation flow cytometry on bone marrow (EuroFlow protocol, 10^{-6} sensitivity).

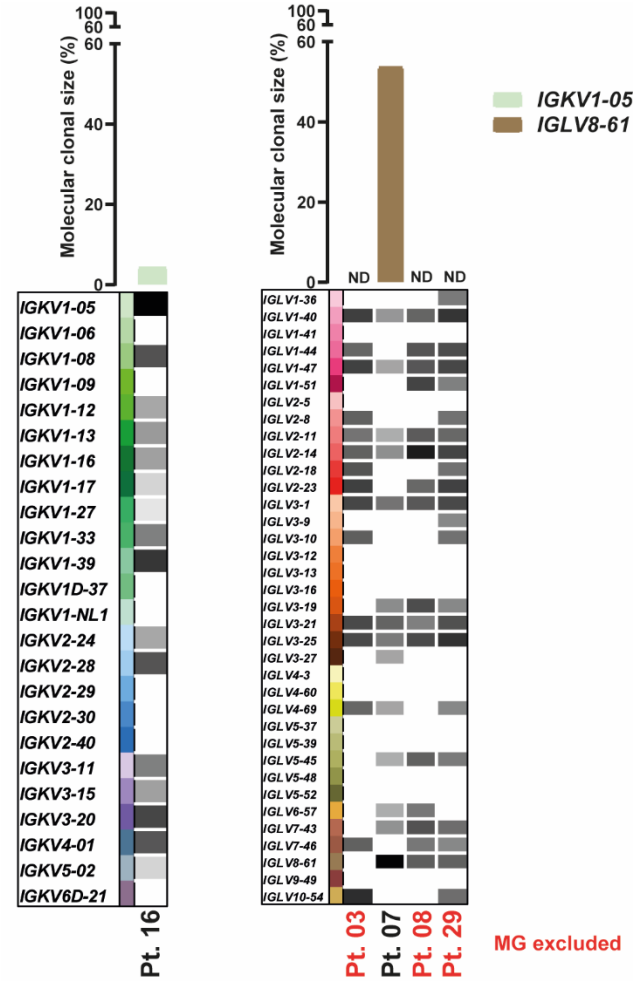

Figure S9. SMaRT M-Seq results in excluded patients

Expression levels of different *IGKV* (left) and *IGLV* (right) germline genes (each denoted by a distinctive color) as assessed by SMaRT M-Seq, starting from the bone marrow of the 5 patients for whom a final diagnosis of systemic AL amyloidosis could not be confirmed or was excluded. Bar graphs indicate the molecular clonal size of the dominant clone identified by Vidjil analysis in each tested sample (the corresponding germline gene is indicated with the bar colour). For patients in red a monoclonal gammopathy was eventually excluded after diagnostic workup for a suspected AL amyloidosis. ND: not detectable; Pt.: patient; MG: monoclonal gammopathy.

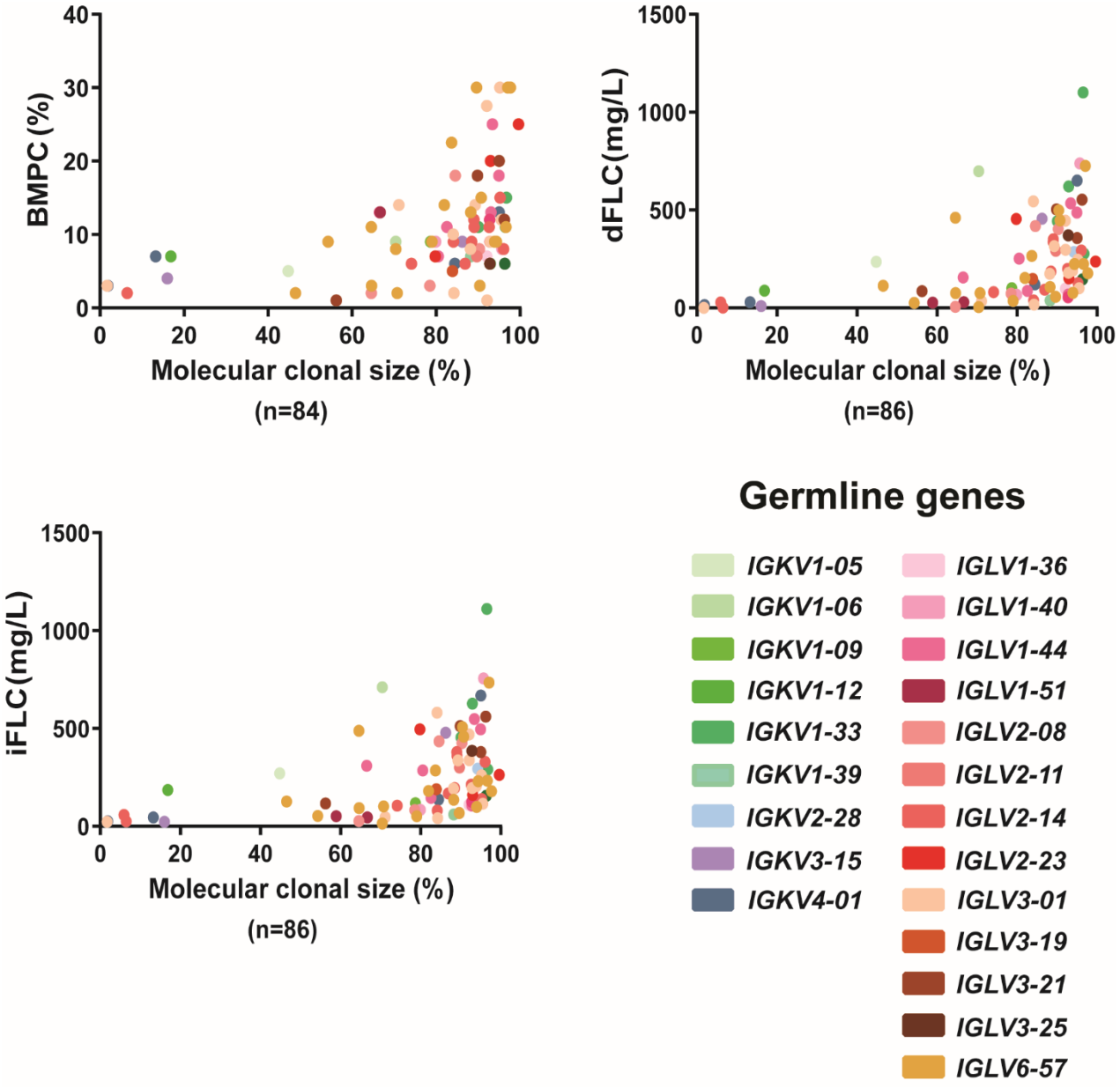

Figure S10. Correlation of molecular clonal size and plasma cell clonal burden
Correlation of molecular clonal size as determined by SMaRT M-Seq and bone marrow plasma cell (BMPC) percentage (top left), differential serum free light chain (dFLC) concentration (top right) and involved free light chain (iFLC) concentration (bottom left). Each dot denotes one patient. The number of patients included in each analysis is indicated at the bottom of each graph (n). Colors refer to the corresponding germline gene of the clonal immunoglobulin light chain.

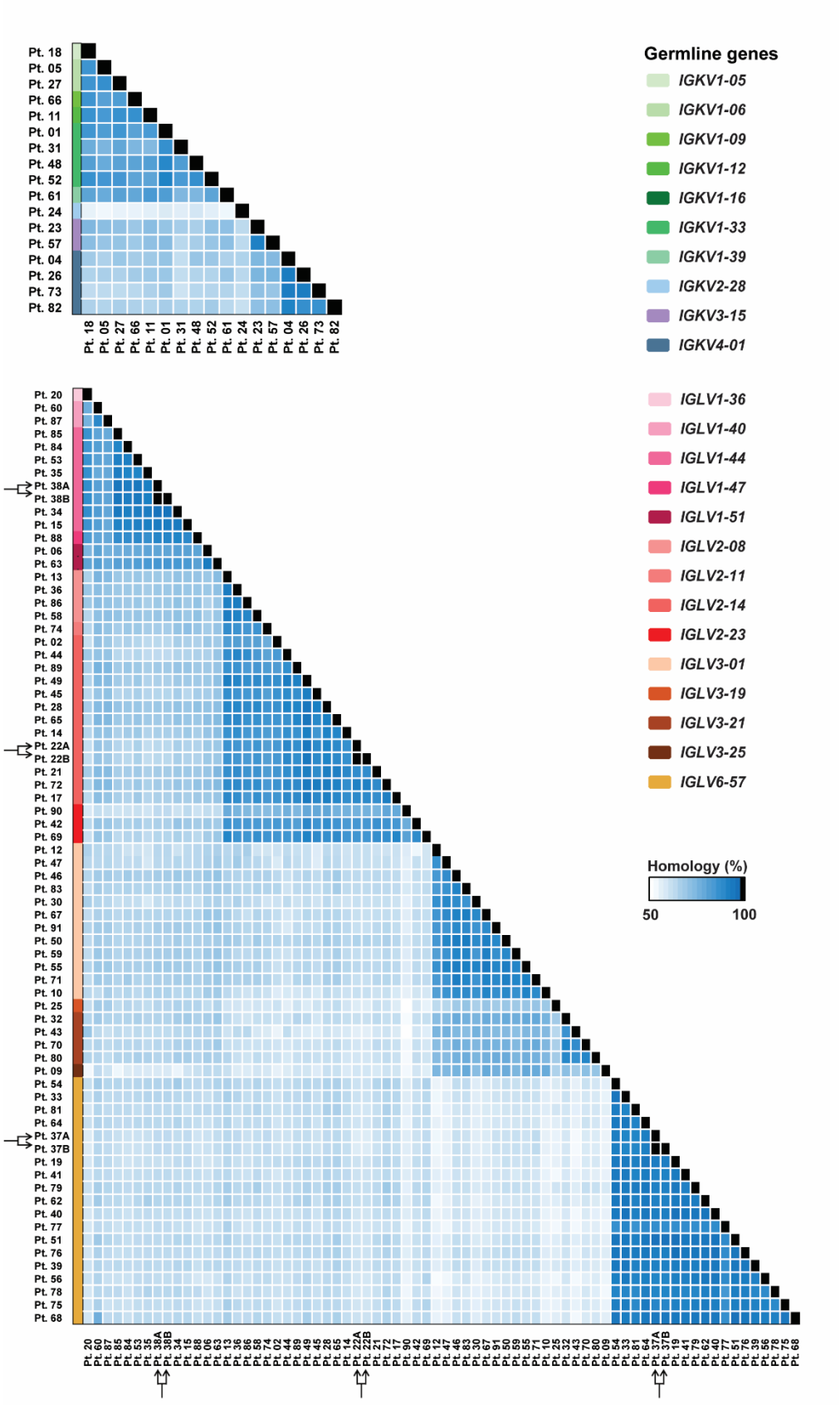

Figure S11. Clonal immunoglobulin sequence homology in the cohort of AL patients
Heatmaps show immunoglobulin sequence homology of 17 κ (top) and 69 λ (bottom) AL patients ranked by *IGKV* (top) and *IGLV* (bottom) germline genes (each denoted by a distinctive color). Three λ patients were analyzed in duplicates (arrows). Pt.: patient.

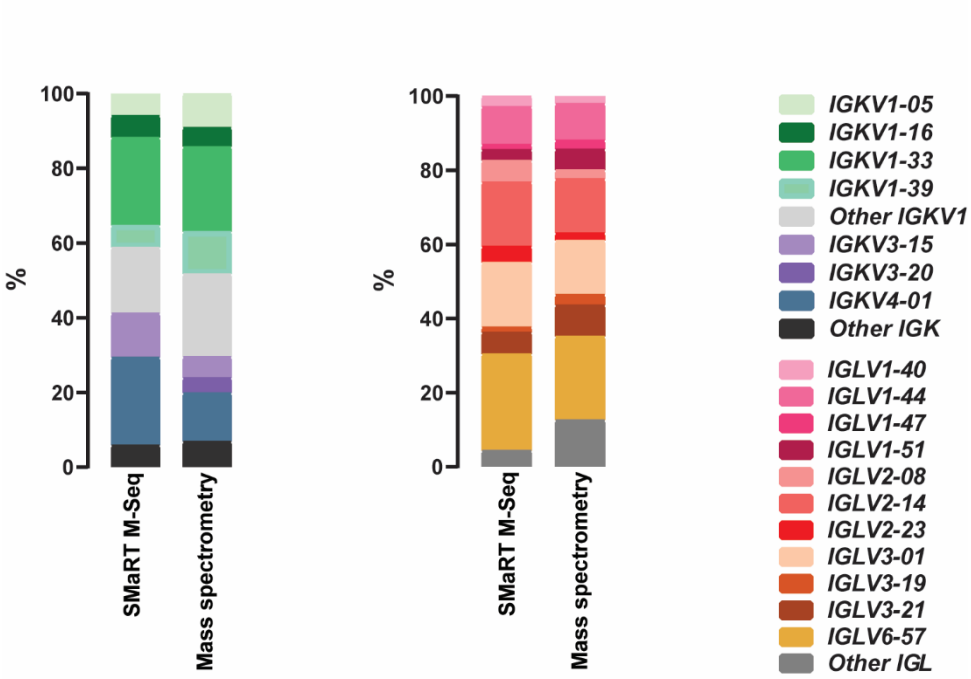

Figure S12. Germline gene usage in AL amyloidosis

Relative frequencies of different *IGKV* (left) and *IGLV* (right) germline genes (each denoted by a distinctive color) in patients with systemic AL amyloidosis as assessed by SMaRT M-Seq in this study (n=17 κ sequences, n=69 λ sequences, including patients 01 and 02 reported in Fig. S4) and by laser-capture and laser microdissection and mass spectrometry (Mass spectrometry) [n=176 κ sequences, n=450 λ sequences, reported in ref. [23]].

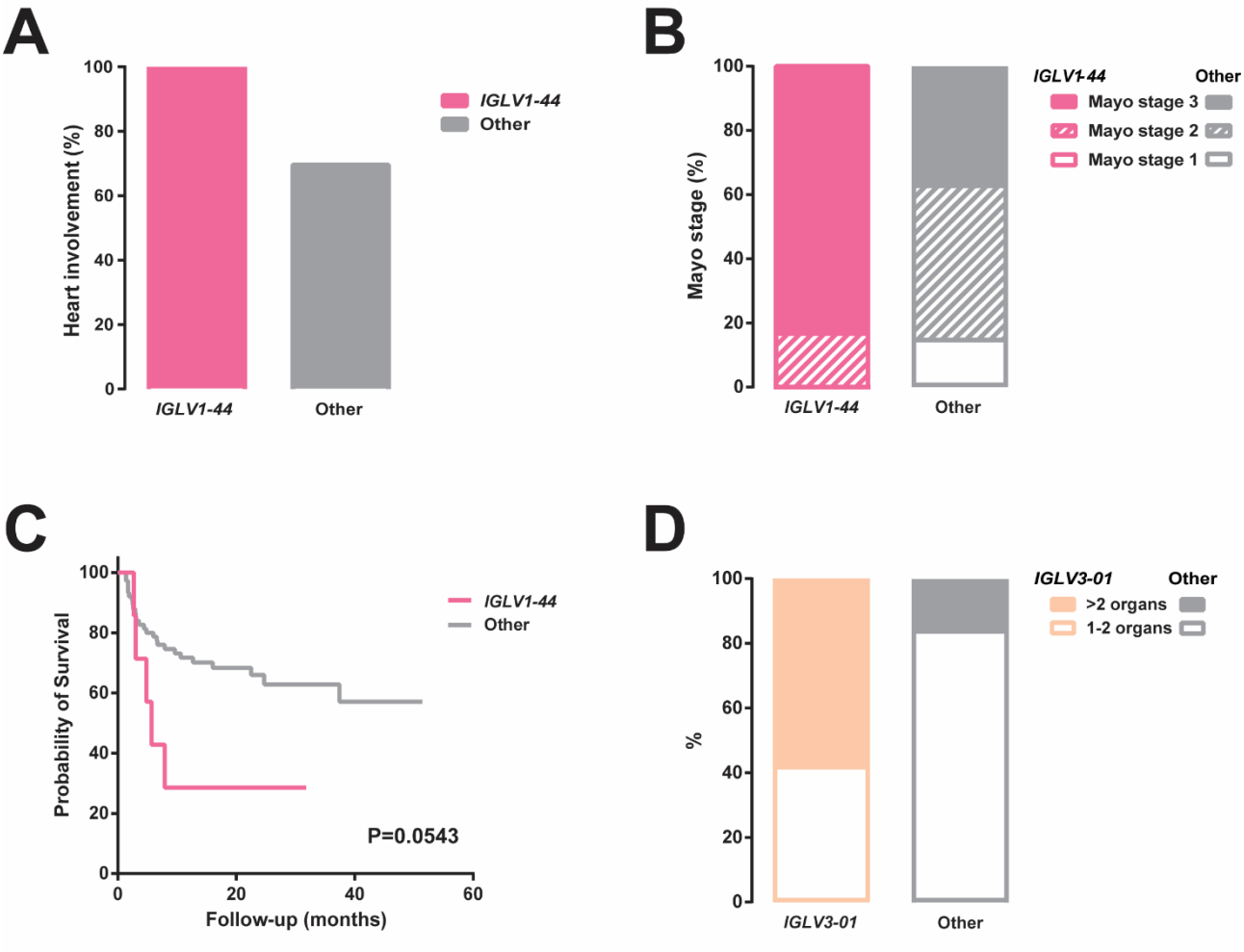

Figure S13. Association of germline gene usage and clinical characteristics
(**A-C**) Percentage of heart involvement (**A**), distribution of Mayo stage (**B**) and Kaplan Meier survival estimate curve (**C**) in AL patient with a clonal *IGLV1-44* gene (pink) or a non-*IGLV1-44* gene (grey). (**D**) Percentage of patients with 1-2 or >2 involved organs in AL patient with a clonal *IGLV3-01* gene (light pink) or a non-*IGLV3-01* gene (grey).

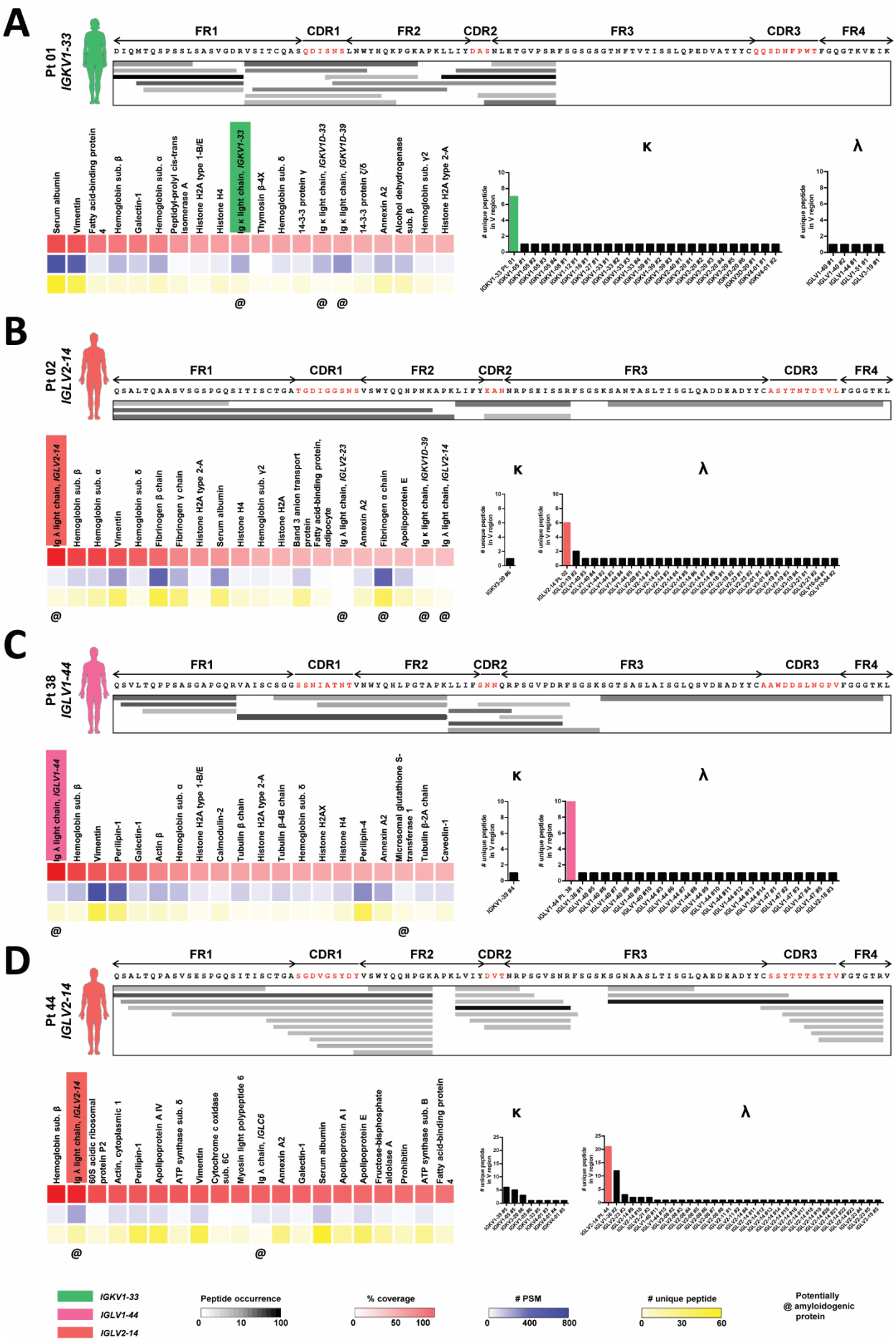
A
Pt 01
IGKV1-33
FR1
CDR1
FR2
CDR2
FR3
CDR3
FR4
DIQMTQSPSSLSASVGDRVSITCQASQDISNSLNWYHQKPGKAPKLLIYDASNLETGVPSRFSGSGSGTNFTVTISSLQPEDVATYYCQQSDNFPWTFGQGTKVEIK
Serum albumin
Vimentin
Fatty acid-binding protein 4
Hemoglobin sub. β
Galectin-1
Hemoglobin sub. α
Peptidyl-prolyl cis-trans isomerase A
Histone H2A type 1-B/E
Histone H4
Ig κ light chain, IGKV1-33
Thymosin β-4X
Hemoglobin sub. δ
14-3-3 protein γ
Ig κ light chain, IGKV1D-33
Ig κ light chain, IGKV1D-39
14-3-3 protein ζ/δ
Annexin A2
Alcohol dehydrogenase sub. β
Hemoglobin sub. γ2
Histone H2A type 2-A
κ
λ
unique peptide in V region
B
Pt 02
IGLV2-14
QSALTQAASVSGSPGQSITISCTGATGDIGGSNSVSWYQQHPNKAPKLIFYEANNRPSEISSRFSGSKSANTASLTISGLQADDEADYYCASYTNTDTVLFGGGTKL
Ig λ light chain, IGLV2-14
Hemoglobin sub. β
Hemoglobin sub. α
Vimentin
Hemoglobin sub. δ
Fibrinogen β chain
Fibrinogen γ chain
Histone H2A type 2-A
Serum albumin
Histone H4
Hemoglobin sub. γ2
Histone H2A
Band 3 anion transport protein
Fatty acid-binding protein, adipocyte
Ig λ light chain, IGLV2-23
Annexin A2
Fibrinogen α chain
Apolipoprotein E
Ig κ light chain, IGKV1D-39
Ig λ light chain, IGLV2-14
C
Pt 38
IGLV1-44
QSVLTQPPSASGAPGQRVAISCSGGSSNIATNTVNWYQHLPGTAPKLLIFSNNQRPSGVPDRFSGSKSGTSASLAISGLQSVDEADYYCAAWDDSLNGPVFGGGTKL
Ig λ light chain, IGLV1-44
Hemoglobin sub. β
Vimentin
Perilipin-1
Galectin-1
Actin β
Hemoglobin sub. α
Histone H2A type 1-B/E
Calmodulin-2
Tubulin β chain
Histone H2A type 2-A
Tubulin β-4B chain
Hemoglobin sub. δ
Histone H2AX
Histone H4
Perilipin-4
Annexin A2
Microsomal glutathione S-transferase 1
Tubulin β-2A chain
Caveolin-1
D
Pt 44
IGLV2-14
QSALTQPASVSESPGQSITISCTGASGDVGSYDYVSWYQQHPGKAPKLVIYDVTNRPSGVSNRFSGSKSGNAASLTISGLQAEDEADYYCSSYTTTSTYVFGTGTRV
Hemoglobin sub. β
Ig λ light chain, IGLV2-14
60S acidic ribosomal protein P2
Actin, cytoplasmic 1
Perilipin-1
Apolipoprotein A IV
ATP synthase sub. δ
Vimentin
Cytochrome c oxidase sub. 6C
Myosin light polypeptide 6
Ig λ chain, IGLC6
Annexin A2
Galectin-1
Serum albumin
Apolipoprotein A I
Apolipoprotein E
Fructose-bisphosphate aldolase A
Prohibitin
ATP synthase sub. B
Fatty acid-binding protein 4
IGKV1-33
IGLV1-44
IGLV2-14
Peptide occurrence
0 10 100
% coverage
0 50 100
PSM
0 400 800
unique peptide
0 30 60
Potentially @ amyloidogenic protein

Figure S14. Clonal light chain sequence verification through mass spectrometry-based amyloid typing

(**A-D**) Top: Liquid chromatography and tandem mass spectrometry in fat pad tissue from four AL patients. Peptide ions mapping is performed against an augmented database including the human proteome, known immunoglobulin sequences from public databases and each patient's clonal light chain sequence (protein sequence predicted based on the nucleotide sequence determined through SMaRT M-Seq). Physical distribution and absolute number (in shades of grey) of peptide ions mapping against each patient's clonal light chain sequence. Bottom, left: Heatmaps show the 20 proteins with the highest sequence coverage identified in each patient's fat pad, with their relative coverage (shades of red), absolute number of peptide-spectrum matches (PSM, shades of blue) and absolute number of unique peptides (shades of yellow). Bottom, right: Bar graphs indicate the number of unique peptides assigned to κ or λ light chains. Patient's clonal light chain is colored according to the corresponding germline gene. Other light chains are designated with their corresponding germline gene and a progressive number (the list of accession codes is available in Table S4). FR: framework region, CDR: complementarity determining region, Pt.: patient.

References

1. Gertz MA, Comenzo R, Falk RH, et al. 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.* 2005;79(4):319-328.
2. Gertz MA, Merlini G. Definition of organ involvement and tratment response in immunoglobulin light chain amyloidosis (AL): a consensus opinion. *Amyloid.* 2010;17(Suppl. 1):48-49.
3. Bradwell AR, Carr-Smith HD, Mead GP, et al. Highly sensitive, automated immunoassay for immunoglobulin free light chains in serum and urine. *Clinical chemistry.* 2001;47(4):673-680.
4. Katzmann JA, Clark RJ, Abraham RS, et al. Serum reference intervals and diagnostic ranges for free kappa and free lambda immunoglobulin light chains: relative sensitivity for detection of monoclonal light chains. *Clinical chemistry.* 2002;48(9):1437-1444.
5. Palladini G, Russo P, Bosoni T, et al. Identification of amyloidogenic light chains requires the combination of serum-free light chain assay with immunofixation of serum and urine. *Clinical chemistry.* 2009;55(3):499-504.
6. Merlini G, Marciano S, Gasparro C, Zorzoli I, Bosoni T, Moratti R. The Pavia approach to clinical protein analysis. *Clinical chemistry and laboratory medicine : CCLM / FESCC.* 2001;39(11):1025-1028.
7. van Dongen JJ, Lhermitte L, Bottcher S, et al. EuroFlow antibody panels for standardized n-dimensional flow cytometric immunophenotyping of normal, reactive and malignant leukocytes. *Leukemia.* 2012;26(9):1908-1975.
8. Westermark P. Subcutaneous adipose tissue biopsy for amyloid protein studies. *Methods in molecular biology.* 2012;849:363-371.
9. Arbustini E, Verga L, Concardi M, Palladini G, Obici L, Merlini G. Electron and immuno-electron microscopy of abdominal fat identifies and characterizes amyloid fibrils in suspected cardiac amyloidosis. *Amyloid.* 2002;9(2):108-114.
10. Gazdar AF, Oie HK, Kirsch IR, Hollis GF. Establishment and characterization of a human plasma cell myeloma culture having a rearranged cellular myc proto-oncogene. *Blood.* 1986;67(6):1542-1549.
11. Arendt BK, Ramirez-Alvarado M, Sikkink LA, et al. Biologic and genetic characterization of the novel amyloidogenic lambda light chain-secreting human cell lines, ALMC-1 and ALMC-2. *Blood.* 2008;112(5):1931-1941.
12. Perfetti V, Sassano M, Ubbiali P, et al. Inverse polymerase chain reaction for cloning complete human immunoglobulin variable regions and leaders conserving the original sequence. *Anal Biochem.* 1996;239(1):107-109.
13. Duez M, Giraud M, Herbert R, Rocher T, Salson M, Thonier F. Vidjil: A Web Platform for Analysis of High-Throughput Repertoire Sequencing. *PLoS One.* 2016;11(11):e0166126.
14. Giraud M, Salson M, Duez M, et al. Fast multiclonal clusterization of V(D)J recombinations from high-throughput sequencing. *BMC Genomics.* 2014;15:409.
15. Brochet X, Lefranc MP, Giudicelli V. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucleic Acids Res.* 2008;36(Web Server issue):W503-508.
16. Alamyar E, Duroux P, Lefranc MP, Giudicelli V. IMGT((R)) tools for the nucleotide analysis of immunoglobulin (IG) and T cell receptor (TR) V-(D)-J repertoires, polymorphisms, and IG mutations: IMGT/V-QUEST and IMGT/HighV-QUEST for NGS. *Methods in molecular biology.* 2012;882:569-604.
17. Giudicelli V, Chaume D, Lefranc MP. IMGT/V-QUEST, an integrated software program for immunoglobulin and T cell receptor V-J and V-D-J rearrangement analysis. *Nucleic Acids Res.* 2004;32(Web Server issue):W435-440.
18. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994;22(22):4673-4680.
19. Lefranc MP, Lefranc G. IMGT((R)) and 30 Years of Immunoinformatics Insight in Antibody V and C Domain Structure and Function. *Antibodies (Basel).* 2019;8(2).

20. Bodi K, Prokaeva T, Spencer B, Eberhard M, Connors LH, Seldin DC. AL-Base: a visual platform analysis tool for the study of amyloidogenic immunoglobulin light chain sequences. *Amyloid.* 2009;16(1):1-8.
21. Swindells MB, Porter CT, Couch M, et al. abYsis: Integrated Antibody Sequence and Structure-Management, Analysis, and Prediction. *J Mol Biol.* 2017;429(3):356-364.
22. Brambilla F, Lavatelli F, Di Silvestre D, et al. Reliable typing of systemic amyloidoses through proteomic analysis of subcutaneous adipose tissue. *Blood.* 2012;119(8):1844-1847.
23. Kourelis TV, Dasari S, Theis JD, et al. Clarifying immunoglobulin gene usage in systemic and localized immunoglobulin light-chain amyloidosis by mass spectrometry. *Blood.* 2017;129(3):299-306.