Article
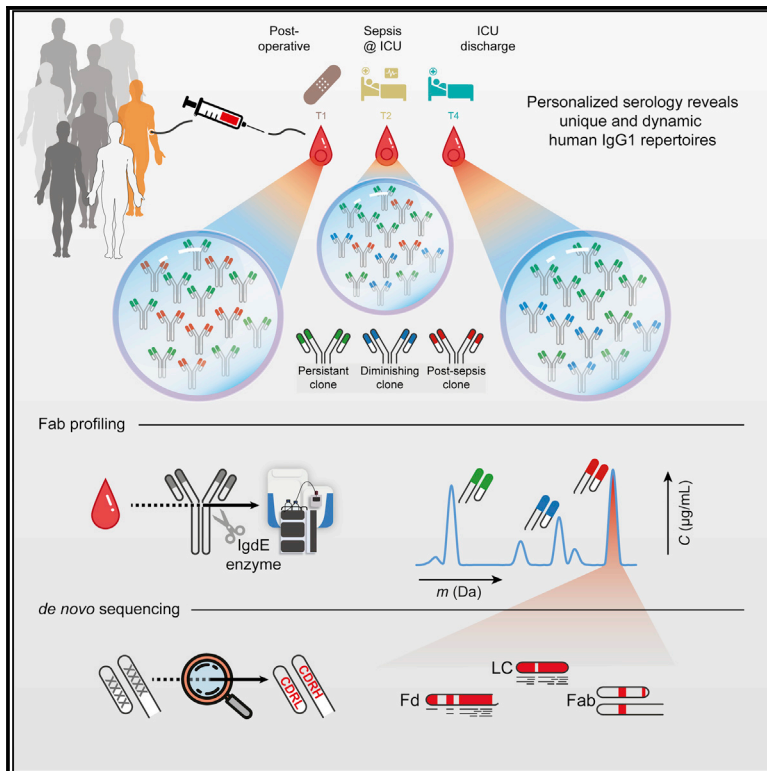
Human plasma IgG1 repertoires are simple, unique, and dynamic

Graphical abstract

Highlights

- Novel LC-MS-based methods enable personalized IgG1 profiling in plasma

- Each donor exhibits a simple but unique serological IgG1 repertoire

- This repertoire adapts to changes in physiology, e.g., sepsis

- Individual plasma IgG1 clones can be identified by combining top-down and bottom-up proteomics

Authors

Albert Bondt, Max Hoek,
Sem Tamara, ..., Olaf L. Cremer,
Marc J.M. Bonten, Albert J.R. Heck

Correspondence

a.j.r.heck@uu.nl

In brief

The human body produces immunoglobulins (Igs) to help combat pathogens. The number of distinct IgG molecules our body can produce exceeds several billions. In contrast to this near-infinite theoretical number, we reveal here, as monitored directly by LC-MS, that only a few dozen distinct clones dominate in abundance the total spectrum of plasma IgG1s of both healthy and diseased donors. Our data indicate that each donor's IgG1 repertoire is distinctively unique. We longitudinally profile an individual's IgG1 repertoire and observe the occurrence or disappearance of specific IgGs over time in response to changes in physiology, e.g., during a septic episode. As a proof of concept, we show that individual plasma IgG1 clones can be quantified and fully sequenced and identified by using a combination of top-down and bottom-up proteomics.

CelPress

Cell Systems

Article

Human plasma IgG1 repertoires are simple, unique, and dynamic

Albert Bondt,[1,2,6] Max Hoek,[1,2,6] Sem Tamara,[1,2,6] Bastiaan de Graaf,[1,2] Weiwei Peng,[1,2] Douwe Schulte,[1,2] Danique M.H. van Rijswijck,[1,2] Maurits A. den Boer,[1,2] Jean-François Greisch,[1,2] Meri R.J. Varkila,[3,4] Joost Snijder,[1,2] Olaf L. Cremer,[4] Marc J.M. Bonten,[3,5] and Albert J.R. Heck[1,2,7,*]

[1]Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, University of Utrecht, Padualaan 8, Utrecht 3584 CH, the Netherlands
[2]Netherlands Proteomics Center, Padualaan 8, Utrecht 3584 CH, the Netherlands
[3]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands
[4]Department of Intensive Care Medicine, University Medical Center Utrecht, Utrecht, the Netherlands
[5]Department of Medical Microbiology, University Medical Center Utrecht, Utrecht, the Netherlands
[6]These authors contributed equally
[7]Lead contact
*Correspondence: a.j.r.heck@uu.nl
https://doi.org/10.1016/j.cels.2021.08.008

SUMMARY

Although humans can produce billions of IgG1 variants through recombination and hypermutation, the diversity of IgG1 clones circulating in human blood plasma has largely eluded direct characterization. Here, we combined several mass-spectrometry-based approaches to reveal that the circulating IgG1 repertoire in human plasma is dominated by a limited number of clones in healthy donors and septic patients. We observe that each individual donor exhibits a unique serological IgG1 repertoire, which remains stable over time but can adapt rapidly to changes in physiology. We introduce an integrative protein- and peptide-centric approach to obtain and validate a full sequence of an individual plasma IgG1 clone *de novo*. This IgG1 clone emerged at the onset of a septic episode and exhibited a high mutation rate (13%) compared with the closest matching germline DNA sequence, highlighting the importance of *de novo* sequencing at the protein level. A record of this paper's transparent peer review process is included in the supplemental information.

INTRODUCTION

The human immune system protects us not only from threats posed by pathogens but also cancer and various other diseases. The immune response in health and disease is crucially dependent on each person's repertoire of immune cells, antibodies, and other circulating plasma proteins. A detailed molecular view of these plasma components is crucial to understanding how they affect each individual's immune response. Immunoglobulins (Igs) represent some of the most important molecules in the human immune system. Ig molecules consist of two identical heavy chains and two identical light chains, held together by a network of disulfide bridges. The heavy chains possess three (IgG, IgA, and IgD) to four (IgM and IgE) immunoglobulin domains with large, conserved regions, which play a role in receptor binding and complement activation. Similar to the heavy chain, the C-terminal domain of the light chain is constant. On the other hand, for both heavy and light chain, the sequence of the N-terminal Ig domains is hypervariable and contains the recognition-determining parts, better known as complementarity-determining regions (CDRs), of the molecule. They are enclosed in the two fragment antigen-binding (Fab) arms of the antibody, consisting of the light chain and the N-terminal parts of the heavy chain (Fd).

The variable regions of the antibody, in particular the CDRs, are optimized to recognize antigens by a process known as affinity maturation. The best antigen binders, modified through somatic recombination and hypermutation of numerous coding gene segment variants, give rise to the mature IgG secreting plasma B cells that produce the antibodies that end up in our circulation. The circulating antibodies, thus, consist of the fully matured heavy- and light-chain variable domain sequences that harbor the CDRs, joined by generally less sequence-variable framework regions (FR). Each unique combination of mature chains is called an Ig clone.

Considering the genes encoding the variable domain sections and the known genomic rearrangements, somatic hypermutations, and post-transcriptional processes that join these sections—resulting into the ultimate protein products—it has been estimated that in humans the theoretical molecular Ig diversity may extend beyond 10^{15} (Schroeder, 2006). Not all theoretically possible Ig clones will be expressed in the human body, since the number of B cells in a human body is several orders of magnitude lower (1–2×10^{11}) (Apostoaei and Trabalka, 2012). Nevertheless,

it has been assumed that the actual repertoire of circulating Igs is extremely large and diverse (Briney et al., 2019; Soto et al., 2019).

Recombinantly expressed clones (mainly IgG) have become a major class of therapeutics, used to fight multiple types of pathologies such as cancers and various infectious diseases. Recent developments have moved the field toward using therapeutic monoclonal antibody (mAb) sequences derived from human subjects instead of laboratory animals; this trend is exemplified by successful new treatments for Ebola (Corti et al., 2016; Dyer, 2019; Mulangu et al., 2019) and COVID-19 (Jones et al., 2021). These therapeutic antibody sequences are inferred from genetic material recovered from patients that successfully overcame the disease. The ability to detect and identify individual mature IgG protein sequences directly from donor specimens would aid such efforts.

To experimentally determine the Ig repertoire, attempts have been made to sequence Ig nucleic acids from bulk B cell populations or B cell subsets from single donors. These Igs are analyzed with high-throughput sequencing (Ig-seq or Rep-seq) at the DNA or RNA level (Georgiou et al., 2014), resulting in datasets of tens to hundreds of thousands of unique reads of variable abundance (Vollmers et al., 2013). Unfortunately, these analyses at the level of DNA and RNA do not measure the actual antibodies of interest, and the presence of a cognate BCR sequence in the B cell population provides no information regarding abundance levels of the antibodies that end up in circulation.

Alternatively, the challenge could be approached from the protein level, analyzing the Ig repertoire present in circulation. The most abundant Ig in human plasma is IgG, at a concentration of approximately 10 mg/mL during health (Cassidy and Nordby, 1975; Gonzalez-Quintela et al., 2008). Of the four IgG subclasses, IgG1 is the most abundant, accounting for more than 50% of all IgGs (van der Giessen et al., 1975) in most people. Given the extremely high theoretical limits on Ig diversity and the large number of experimentally determined variants, most researchers have refrained from analyzing plasma antibodies directly at the protein level. It has been mostly assumed that it is impossible to detect any single Ig clone against the expected background of thousands to millions of other clones.

However, in recent years, several attempts have been made to identify plasma Igs directly. G. Georgiou and colleagues should be considered pioneers. In their method (also referred to as Ig-seq, but protein-based instead of solely gene-based), IgGs are purified from plasma, and tryptic Ig peptides are characterized using liquid chromatography coupled online to tandem mass spectrometry (LC-MS/MS), focusing on the detection of IgG heavy-chain CDR3 peptides (Chen et al., 2017; Lavinder et al., 2014; Lee et al., 2016, 2019; Wine et al., 2013). Another recently developed proteomics approach uses LC-MS to profile intact light chains purified from plasma (Barnidge et al., 2014a, 2014b; He et al., 2019, 2017; Mills et al., 2015; Sharpley et al., 2019). In another report, profiling and sequencing of intact Fabs was attempted, but the sensitivity was too low, and only a few sequence tags could be derived from separated light chains (Wang et al., 2019).

In these approaches, information about the heavy- and light-chain pairing is often lost, which is unfortunate as only the combined CDRs from both heavy and light chain provide the full complementarity against an antigen. Of note, in nearly all these studies, only a subset of plasma immunoglobulins recognizing and binding a specific antigen was targeted, or a mAb spiked into plasma was used as a model. Nevertheless, the remarkable observation was made that a person's plasma Ig repertoire could be several orders less variable than assumed based on the available B cell repertoire and likely dominated by only a limited number of clones. Whether this observed number of IgGs is a consequence of the targeted analysis of antigen-specific IgGs or diseases with monoclonal IgG overexpression (gammopathy) remained unclear.

Here, we introduce a sensitive and efficient approach for quantitative plasma IgG1 clone profiling. The method was applied to a sample set of two healthy control donors, as well as eight critically ill patients, from which sequential plasma samples were retrieved while developing nosocomial sepsis, experiencing a dramatic immunological change in a relatively short time span (Figure 1A). The application of our method revealed several important properties of the human plasma IgG1 repertoire: (1) the total IgG1 repertoire is dominated by a few dozen clones, but (2) is unique for each individual, and (3) single clones are differentially affected by physiological changes. Next, as proof of concept, we *de novo* sequenced a single plasma clone—that appeared at sepsis onset—directly from human plasma, only made possible by using iteratively a combination of protein-centric and peptide-centric proteomics. To validate the correctness of the derived *de novo* IgG1 sequence, we produced its recombinant mAb equivalent and compared its key structural features with the plasma clone. We foresee that this approach will unlock the potential of mass spectrometric analysis and identification of disease-responsive IgGs that may be directly evaluated and used as therapeutic agents, as they do already represent fully matured, fully human Abs.

RESULTS

Mass-spectrometry-based Fab profiling of the human plasma repertoire

To chart and monitor the nature of the plasma IgG1 repertoire, we started our analysis with 10 μL of plasma, derived from a single donor. From such a sample, we first captured all intact IgGs using affinity beads. Subsequently, the captured IgG molecules were digested using the highly specific Ig degrading enzyme (IgdE), cleaving specifically IgG1's at a defined site in the upper hinge region, resulting in the segregated Fc (that remains bound to the affinity beads) and two identical Fabs (Figure 1B) (Spoerry et al., 2016). We focused on the Fab fragments derived from the intact IgGs because this (1) concentrates the clonal signal since each IgG1 provides two identical Fab molecules, (2) results in more homogeneous mass profiles by removing the Fc portions that harbor two heterogeneous *N*-glycosylation sites, and (3) retains all hypervariable CDRs, which define the unique identity and antigen recognition of each clone.

Following elution of the IgG1 Fabs, all these intact 45–53-kDa Fab molecules were subjected to reversed-phase LC-MS. All individual Fab fragments were subsequently characterized by their distinctive mass and chromatography retention time. In our analyses, we spiked-in two monoclonal IgG1 antibodies (Table S1,
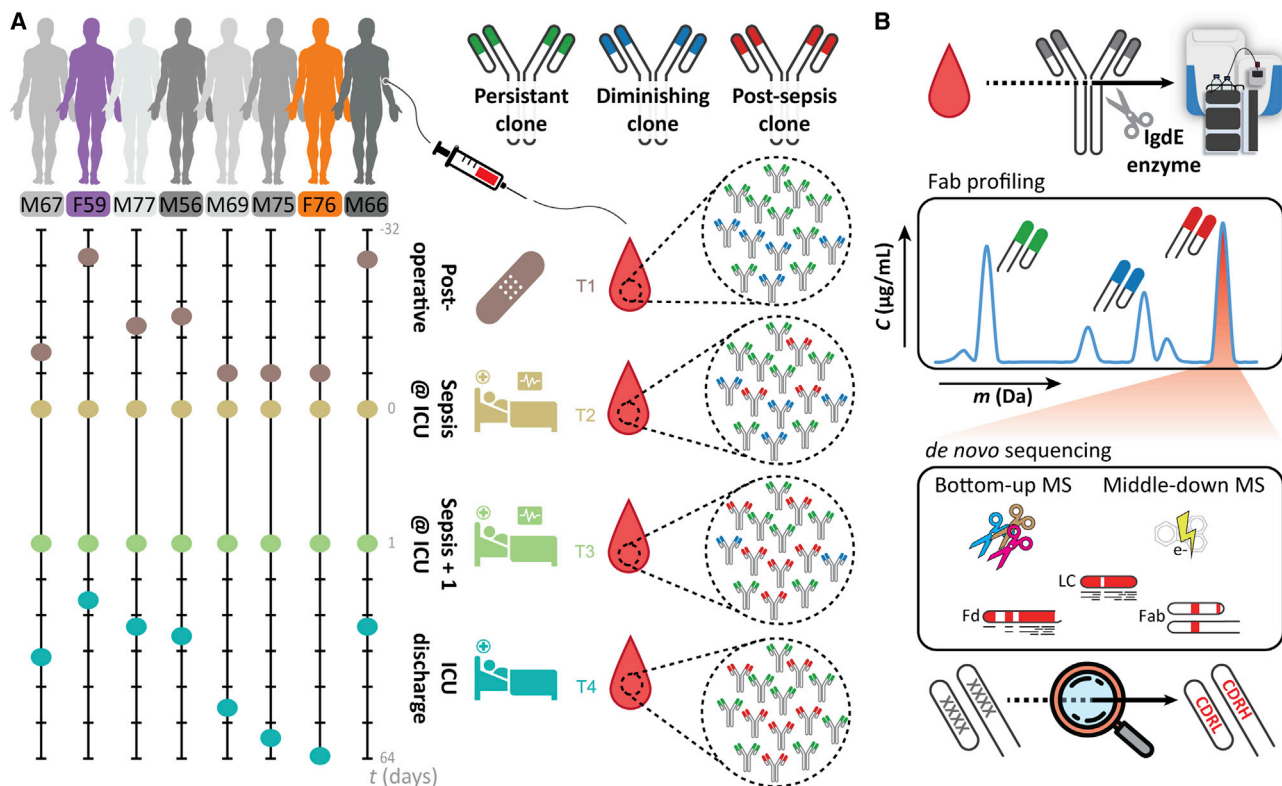
Figure 1. Monitoring individual plasma IgG1 profiles

(A) Longitudinal analysis of the IgG1 repertoire from sepsis patient plasma obtained at four time points, reveals its simplicity and clonal dynamics: some clones are fairly constant (green), some disappear (blue), whereas others appear over time (red).

(B) The experimental approach taken involves IgG capturing from 10–100 μL of serum, followed by the specific enzymatic digestion of the IgG1 molecules in their hinge region, generating two identical Fab portions. All generated Fabs are collected and subsequently subjected to LC-MS analysis. The clonal repertoire is profiled, whereby each identified clone is characterized by its unique mass and retention time. A single post-sepsis clone from one of the patients (F59) was selected for *de novo* sequencing, combining protein- and peptide-centric mass-spectrometry-based sequencing. The extracted full sequence of the plasma IgG1 was validated by analyzing, in a similar manner, a recombinant IgG1 analog of the plasma clone.

mAbs #1 and #2) of known sequence at a defined concentration. These mAbs were used as internal standards for mass and retention time calibration and quantification of all the other distinctive plasma IgG1 clones. This also allowed us to calculate the precision and accuracy of retention time, mass, and quantification in our measurements as illustrated in Figure S1A.

Using a mixture of six monoclonal antibodies (Table S1) spiked into a single-donor plasma background, we furthermore observed linear relationship between (1) the quantity of mAbs that were spiked into the plasma sample, and (2) the quantity that is observed ($R^2 = 0.99$, Figure S1B), using a dilution/titration with 4,000, 800, 200, and 20 ng per mAb. Of note, no discrepancy was observed for the Fab glycosylated mAb as compared to the other mAbs. To evaluate the repeatability (technical replicate and sample preparation replicate), the 100 most abundant plasma-derived clones in this sample were quantified in multiple replicates (Figure S1C). Finally, one of the samples was injected three times to serve as injection replicates (Figure S1C, #1). From all these validation experiments, we concluded that, by using the approach depicted here (Figure 1), the repertoire of Fab clones could be accurately and reproducibly determined from as little as 10 μL of plasma obtained from a single donor at a single time point.

Plasma IgG1 repertoires are dominated by a few clones

Next, we analyzed in parallel a set of 32 plasma samples obtained from eight patients of the Dutch molecular diagnosis and risk stratification of sepsis (MARS) cohort (Figure 1A; Table S2). All these patients underwent major gastrointestinal surgery and subsequently developed an infectious complication (i.e., anastomotic leakage or pneumonia) resulting in sepsis. Plasma samples were obtained from the patients at four different stages: within 24 h of surgery when no signs of sepsis were present (sample T1; between $t = -19$ and $t = -2$ days), on two consecutive days after onset of sepsis (samples T2 and T3; $t = 0$ and $t = 1$ day), and upon intensive care unit (ICU) discharge, when the sepsis had been resolved (sample T4; between $t = 3$ and $t = 61$ days) (Figure 1A). In addition, to monitor the nature of the plasma IgG1 repertoire in healthy donors, we performed an identical analysis on two sets of three sequential healthy donor plasma samples, each collected roughly 1 month apart.

In marked contrast to expectations of extensive IgG1 diversity, we observed that all the LC-MS profiles of IgG1 Fab molecules were dominated by just a few dozen peaks, both in the 32 sepsis plasma samples as well as in the six plasma samples of the healthy donors (Data S1 and S2). In each of the LC-MS runs, we could pick up distinctive IgG1 signals of between 35 and
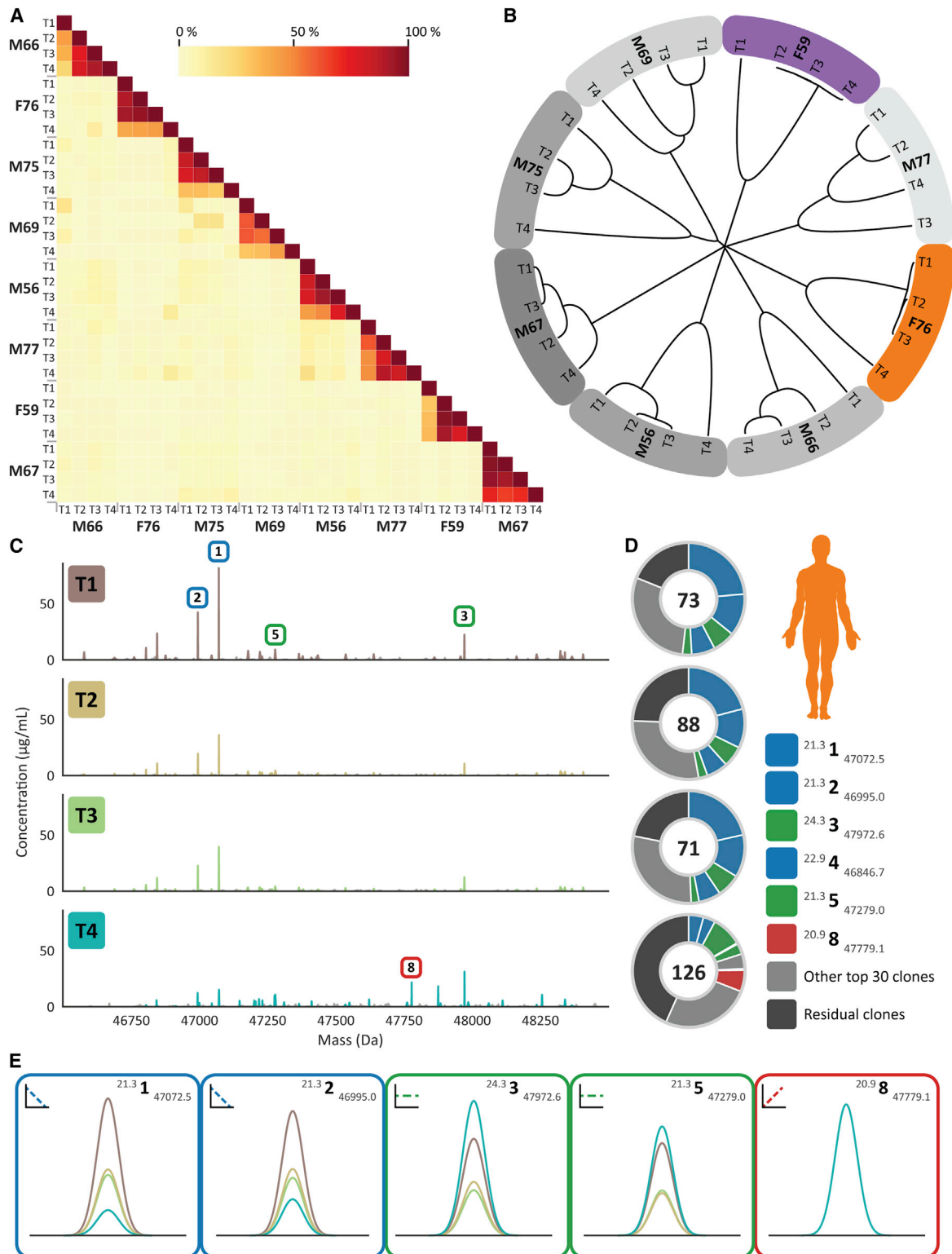
Figure 2. Monitoring personalized plasma Fab repertoires reveals not only their simplicity and extreme donor uniqueness but also longitudinal clonal variations within a single donor

(A) Heatmap illustrating the degree of overlap between the detected IgG1 repertoires in all analyzed sepsis patient plasma samples. For each of the eight donors, four sampling times were available, and Fab profiles were measured by LC-MS analysis. Each LC-MS peak, exhibiting a unique mass and retention time pair, was considered a unique clone and annotated as $^{RT}\#_{mass}$. The amount of Fab molecules, based on intensity that is persistent, is quantified and shown as a

(legend continued on next page)

543 in abundance dominant clones (median 196; Data S1) that we distinguished by their masses in Dalton and retention times (RT) in minutes. Each detected clone was given a unique identifier: $^{RT}\#_{mass}$.

We found that the summed concentrations of the 30 most abundant IgG1 clones account for more than two-thirds of all IgG1 molecules detected from plasma (median 71.8%, range 47.3%–98.3%, Data S1). The full lists of detected clones are provided in Data S1. In addition, the deconvoluted mass plots (similar to Figure 2C) and the raw chromatograms supported by extracted chromatograms for all identified Fabs obtained in the LC-MS measurements are provided in Data S2, and one example of the actual raw mass spectrometry data of Fabs that we have analyzed is provided in Data S3.

Next, we looked at the cumulative mass distribution of all detected IgG1 Fabs in the plasma samples from all donors and at all time points. This cumulative mass distribution—representing more than 5,500 clones experimentally identified—resembled the expected mass distribution of over 130 million IgG1 Fabs constructed from the sequences in the ImMunoGeneTics information system (IMGT) (Lefranc and Lefranc, 2001) database (Figure S1D), revealing that we profiled a representative IgG1 repertoire.

As can be seen in Figure S1D, most Fab fragments exhibit masses between 46 and 49.5 kDa. However, we also did detect some higher Fab masses, which may be indicative of Fab glycosylation. The average mass of Fab glycans is estimated to be around 2,300 Da (Bondt et al., 2016; Hafkenscheid et al., 2017). In two of our donors, annotated M66 and M77, we did detect relatively high levels of Fab glycosylation as shown for M66, time point 3, in Figure S2A, with the annotation of the putative Fab glycosylation annotated in Figure S2B. Still the Fab glycosylated clones represented just a few percent of the total abundance (2%–6% for donor M66 and M77). The fractional abundance of glycosylated Fabs in the other patients was between 0% and 1.86% (with a median of 0.295%) (Figure S2C). Also, in the two healthy donors, one displayed a fractional abundance of glycosylated Fabs of about 3% (F66H), whereas in the other donor this remained around 0.5% at all sampling time points. This fractional abundance is substantially lower than would be predicted based on the IMGT database (∼11% of these 130 million sequences carry at least one consensus N-glycosylation site) and lower than the ∼17% described in literature (Bondt et al., 2016; Hafkenscheid et al., 2017). However, our data reveal that the fractional abundance of glycosylated Fabs is also donor-dependent.

Plasma IgG1 repertoires are unique for each donor

Next, we compared the IgG1 Fab profiles between time points not only within a single donor but also between different donors. Interindividual analyses showed that virtually none of the Fab IDs overlapped between individuals (Figures 2A, S3, S4A, and S4B). Also, hierarchical clustering based on clone IDs clusters each donor distinctively (Figure 2B). Thus, each donor has its own simple albeit unique IgG1 repertoire. However, within each individual, overlap between the measured IgG1 repertoires measured across time was found to be very high, even when the time span largely exceeded the average half-life of IgG1s (Figures 2A–2D and S4). A large portion of the most abundant IgG1s remains present throughout the sampling window of up to 2 months, although we also observe a response in the IgG1 profile due to changes in the patient's physiology (discussed below). To exclude whether these findings were due to the severe physiological state of the eight septic patient donors, we performed a similar analysis on plasma of two healthy donors. In the absence of a dramatic immunological challenge, the IgG1 profiles, as obtained from the two healthy donors, show (1) a very high stability over time within individuals and (2) an interindividual overlap in uniquely RT- and mass-identified IgG1 clones near to zero (Figures S4A and S4B).

Longitudinal quantitative monitoring of single IgG1 clones

By spiking in two recombinant IgG1 mAbs at a known concentration to the plasma samples prior to sample preparation, we could provide additional absolute quantification for the abundance of the detected IgG1s. The concentrations of the LC-MS-detected endogenous IgG1 clones present in plasma ranged from less than 0.05 up to >400 μg/mL (<300 pM up to >2.5 μM, median ∼6.25 nM; Data S1). Monitoring serological IgG1 repertoires over time in patients who had undergone a septic episode, we observed several distinct quantitative patterns. The most recognizable patterns are highlighted in Figures 2C–2E. There are IgG1 clones that become lower in concentration over time (Figures 2C–2E, blue boxes). Another category of IgG1 clones was undetectable in the plasma until post-sepsis but became abundantly present at T4 (Figures 2C–2E, red boxes). Yet, another group of IgG1 clones was found to be rather persistent in concentration over all sampling moments (Figures 2C–2E, green boxes). In healthy donors, the majority of clones were more persistent in concentration, although some subtle changes could be observed for some clones.

---

percentage as indicated by the color bar. In between donors, the overlap is found to be on average 3%, whereas within a single donor at different time points the overlap was found to be in between 26% and 98%.

(B) Hierarchical clustering of the Fab clonal repertoires based on correlation distance. The branch lengths depict the distance between the repertoires. Donors are colored as in Figure 1A.

(C) Longitudinal deconvoluted Fab mass profiles of donor F76 at each of the four time points. Each peak represents a unique Fab at its detected mass and plasma concentration. The top 30 most intense Fab clones in each sample are colored reflecting the time points, the other clones are colored gray. Five peaks are highlighted with a box that is colored based on the longitudinal behavior of the Fab concentration in plasma (blue, diminishing clone; green, persistent clone; and red, post-sepsis clone), a magnified version of each of these Fab signals is shown in (E).

(D) Pie charts portraying the total number and distribution of clones in donor F76 for each time point. The value within the chart displays the number of identified unique Fab molecules. The five most intense Fabs are colored based on longitudinal behavior, and their mass and retention time are depicted in the legend in order of abundance.

(E) Magnified mass plots for each of the highlighted clones. The peaks are colored according to the time points, the surrounding border and sign indicate the longitudinal behavior and the top right shows the annotated clone ID.

Full *de novo* sequencing of an individual plasma IgG1 clone

From the data presented earlier, we can conclude that the plasma IgG1 repertoire of individual healthy and diseased donors is unique and dominated by a few dozen abundant clones. Next, we sought to identify the exact sequences of these clones to obtain further insight into their function and origin. Complete *de novo* sequencing of serological IgGs is notably difficult for several reasons. First, the inherent sequence hypervariability has so far proven to be highly challenging even when (personalized) genome-based sequence templates are available. Second, *de novo* sequencing of antibodies at the protein level by MS is hampered by the complex nature of IgG molecules, stemming from their multichain structure and the intricate network of disulfide bridges. Finally, although shotgun proteomics can be used to obtain (partial) sequences of purified mAbs (Guthals et al., 2017; Sen et al., 2017; Tran et al., 2016), this becomes several orders of magnitude more difficult in a plasma background containing many IgG molecules of closely homologous sequences.

To tackle this challenge, we explored a hybrid and iterative approach combining state-of-the-art peptide-centric (i.e., bottom-up) and protein-centric (i.e., middle-down) mass-spectrometric sequencing methods, using dedicated algorithms to mix-and-match the extracted proteomics-based sequencing data. As proof of concept, we attempted to fully sequence the light and heavy chain of a Fab derived from a single highly abundant IgG1 clone observed in donor F59. This donor showed a plasma IgG1 repertoire dominated by two clones in particular: $^{24.4}1_{47,359.4}$ (average mass 47,359.4 Da, retention time 24.4 min) and $^{20.6}2_{47,025.7}$ (Figure 3A). We focused on the $^{24.4}1_{47,359.4}$ clone, as this clone appeared exclusively after the onset of sepsis.

Following fractionation and selection of the $^{24.4}1_{47,359.4}$ clone, we subjected this Fab to mass-spectrometry-based *de novo* sequencing, combining data from middle-down and bottom-up proteomics (Figure 1B). The *de novo* sequence information from both approaches was used to first select several closely matching light- and heavy-chain templates from the publicly available IMGT database of IgG germline sequences (Figure S5; Tables S3 and S4). Subsequently, the bottom-up and middle-down sequencing data and the measured intact accurate masses of the Fab, light chain, and Fd were used to refine the selected template sequences and ultimately determine the mature sequence present in the donor, revealing discrepancies between the germline and mature sequences.

In the protein-centric approach, we performed electron transfer dissociation (ETD) on the intact Fab, as well as the light chain and Fd separately, obtained by reduction of the Fab molecule (Figure 3A, bottom trace). Several fragmentation scans obtained for the intact Fab, the separated light chain, and Fd were grouped and combined based on their unique precursor mass and retention time (Figure 3B). The ETD mass spectra of the intact Fab yielded accurate masses of the light chain and Fd by cleavage of the interchain disulfide bond, thus providing direct information about the light-chain-heavy-chain pairing (Figure 3C, top spectra). In addition, these ETD spectra yielded extended sequence tags, covering informative parts of the CDR3 and framework (FR) 4 regions of both Fab chains, in a similar manner as previously reported by performing ETD or

ECD of intact IgG molecules (den Boer et al., 2020; Fornelli et al., 2017; Shaw et al., 2020). Complementary, ETD spectra of the separated Fab chains yielded partial sequence information for the FR1, CDR1, FR2, CDR2, and the constant region of the selected clone (Figure 3C, bottom spectra and Figure S6). Although our middle-down MS data provided valuable information about the clone of interest, they did not fully cover the sequence, primarily due to incomplete fragment formation and ambiguous sequence information obtained from the larger fragments.

To further extend our sequencing attempt, we subjected the fraction containing the targeted IgG1 clone to enzymatic digestion, using in parallel four proteases: trypsin, chymotrypsin, thermolysin, and pepsin. The resulting peptides were analyzed by a bottom-up approach, using *de novo* sequencing algorithms for sequence annotation (Peng et al., 2021). Although fractionated and enriched for the desired $^{24.4}1_{47,359.4}$ clone, the bottom-up MS data also contained numerous peptides originating from co-isolated plasma clones (Figure S7), which made it impossible to determine the correct sequence by solely using the bottom-up MS data. Nevertheless, by iteratively extending the sequence information from the middle-down MS approach with the *de novo* peptides from the bottom-up MS approach, we ultimately were able to extract the most likely germline precursor of the targeted clone and, notably, its mature sequence by implementing various single-amino-acid mutations not present in the IMGT database (Figure 4).

In more detail, by using the IMGT database of germline sequences as input, the cumulative MS evidence revealed that the analyzed IgG1 Fab carried a lambda light chain. This light chain originated from a combination of the immunoglobulin lambda (IGL) variable (V) 2-14*01 (IMGT/LIGM-DB: Z73664), IGL joining (J) 2*01 (IMGT/LIGM-DB: M15641), and IGL constant (C) 2*01 (IMGT/LIGM-DB: J00253) alleles. For the heavy-chain Fd portion, we determined that it was constructed from the immunoglobulin heavy (IGH) V3-9*01 (IMGT/LIGM-DB: M99651), IGHJ5*01 (IMGT/LIGM-DB: J00256), and IGHG1*03 (IMGT/LIGM-DB: Y14737) alleles and a diversity (D)-region, which substantially deviated from any reported germline D-region. Although initial identification resulted in just a partial sequence coverage, we could fill the gaps in the germline sequences using sequence tags from the middle-down MS and the *de novo* peptides from the bottom-up MS (Figures S8 and S9). Eventually, our approach resulted in a complete and exact precursor mass match for the light and heavy chains, 100% sequence coverage in bottom-up MS, and near-complete annotation of all available fragments in the middle-down MS data. In this process, numerous mutations had to be incorporated when comparing our data with the germline template sequences (Figures 4B and 4C, in red letters), revealing somatic hypermutation (SHM) of around 13% and 16% for the V gene of the light chain and the heavy chain, respectively. The level of confidence in each identified mutation site is based on several criteria, including support of a mutation by consecutive mass peaks in the middle-down MS retrieved sequence tags, the peptide scores and coverage depth in the bottom-up MS data as well as the frequency of amino acid occurrence at a given position in a pool of experimental and the germline IgG1 sequences (Table S5; Figure S10). Together this provides
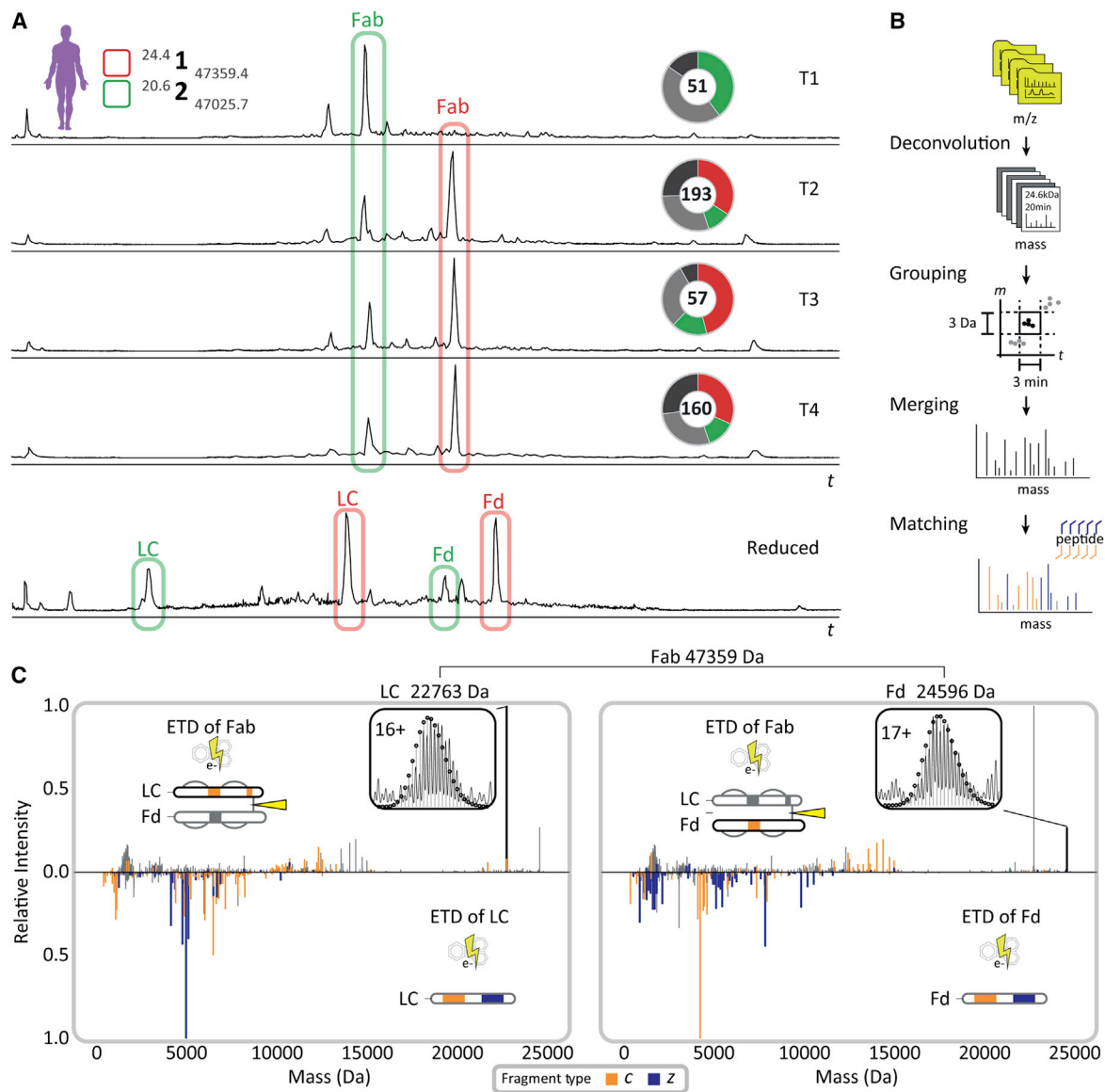
Figure 3. Middle-down sequence characterization of the Fab clone $^{24.4}$ 1 $_{47,359.4}$ that becomes dominant in the repertoire after the onset of sepsis—under reducing and non-reducing conditions

(A) Reversed-phase LC-MS base peak profiles of the Fab repertoire detected in samples T1-4 from donor F59 (top 4 profiles) reveal the dominance of a small number of clones, whereby especially clone $^{24.4}$ 1 $_{47,359.4}$ becomes dominant in abundance after the onset of sepsis. The LC-MS chromatogram of the reduced and denatured Fab repertoire from donor F59 at T3 is depicted in the bottom panel. The light chains (LCs) and the N-terminal portions of the heavy chains (Fd) of the two dominant clones are annotated with corresponding colors and chain names. All species highlighted in red were subjected to middle-down LC-MS/MS using ETD.

(B) Data processing workflow to prepare middle-down ETD-MS/MS spectra for fragment matching and sequence-tag detection (see STAR Methods section for details).

(C) Deconvoluted ETD-MS/MS spectra of the intact Fab (top spectra) and reduced LC and Fd fragments thereof (mirrored spectra) with the c/z-fragment ions annotated for the LC (left) and Fd (right). The isotopic envelopes of the most abundant charge states of the LC and Fd fragments released from the Fab upon ETD are depicted in the insets with theoretical isotope distributions of the corresponding chain sequences overlaid as black circles. Masses of the LC and Fd fragments and the cumulative mass of the Fab are indicated above the spectra. See also Figure S6 for more detail on the fragment ions identified in these middle-down MS spectra.

proof of concept that it is possible to *de novo* sequence IgG1s present in plasma.

The definitive sequence assignment benefited largely from gathering multiple pieces of experimental evidence, notably (1) the accurate mass of the Fab, (2) the highly accurate masses of the two

individual chains comprising the Fab, (3) the *de novo* identified amino acid sequence reads, retrieved from the middle-down fragmentation of intact chains and intact Fab molecule, and (4) the *de novo* identified amino acid reads from the—multiple proteases-based—peptide-centric bottom-up approach.
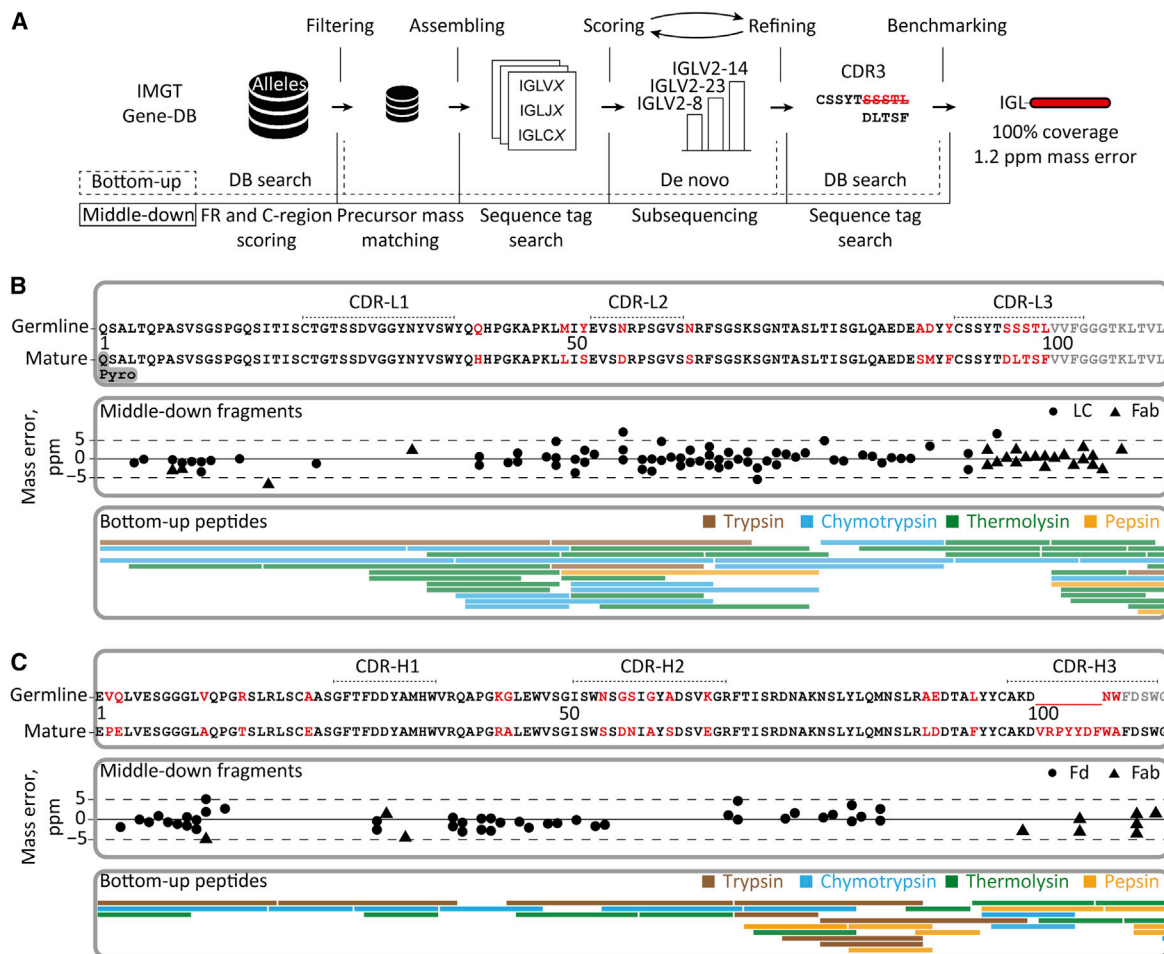
Figure 4. Integrative *de novo* sequencing of the Fab clone [24.4] 1 [47,359.4] from donor F59 combining middle-down and bottom-up MS data

(A) Data analysis pipeline displaying the key steps in the *de novo* sequencing, namely, filtering of the germline database of light- and heavy-chain sequences, assembling of selected allelic variants with mass constraints, scoring of the assembled sequences by using middle-down MS data, iterative refining of the best scoring templates by using peptides in bottom-up MS, and benchmarking of the optimized mature sequences using data from both middle-down and bottom-up MS analysis.

(B) Alignment of the best matching germline IGLV amino acid sequence from the IMGT database (IGLV2-14*01) with the mature sequence that was determined for the light chain of the dominant clone (top box), the fragments from middle-down MS (middle box), and the peptides from bottom-up MS.

(C) Alignment of the best matching germline IGHV amino acid sequence from the IMGT database (IGHV3-9*01) with the mature sequence that was determined for the Fd of donor F59's clone [24.4] 1 [47,359.4] (top box), the fragments from middle-down MS (middle box), and the peptides from bottom-up MS. CDR regions in top panels of (B) and (C) were annotated with reference to the closest matching IMGT sequence. Amino acids that were determined to be different in the mature [24.4] 1 [47,359.4] sequence are highlighted in red.

Validation of the *de novo* sequencing-derived sequence

To validate the accuracy of the full *de novo* sequence of the [24.4] 1 [47,359.4] clone from donor F59, we generated a synthetic recombinant IgG1 clone based on the experimentally determined sequence. We used exactly the same procedures to sequence the recombinant mAb as applied to the plasma-obtained clone, including all the peptide- and protein-centric approaches. Since CDRs are the most critical and hypervariable regions of the antibody, we set out to find peptides in the two datasets covering these regions, so that we could directly compare their fragmentation spectra. A direct comparison of tandem MS spectra of the CDR-spanning peptides from the donor clone and the recombinant mAb are presented in Figures 5A and 5B, covering parts of the light chain and Fd portion, respectively. Above the graphs,

the *de-novo*-obtained sequence is shown with the annotated CDRs, whereby the purple lines indicate the selected peptides. In each panel, the MS/MS spectra obtained from peptides derived from the plasma clone of donor F59 and the recombinant mAb are shown, with the donor spectrum on top and the recombinant mAb spectrum mirrored below. Through visual comparison and as evidenced by the high correlation scores ranging between 0.91 and 0.98, the spectra obtained for peptides originating from the recombinant clone were highly similar to the MS/MS spectra from the peptides derived from the plasma clone [24.4] 1 [47,359.4]. The observed high similarity was not restricted to the *m/z* positions but was found to be also reflected into fragment ion intensities, which are quite sequence specific, thus presenting an additional layer of confidence.
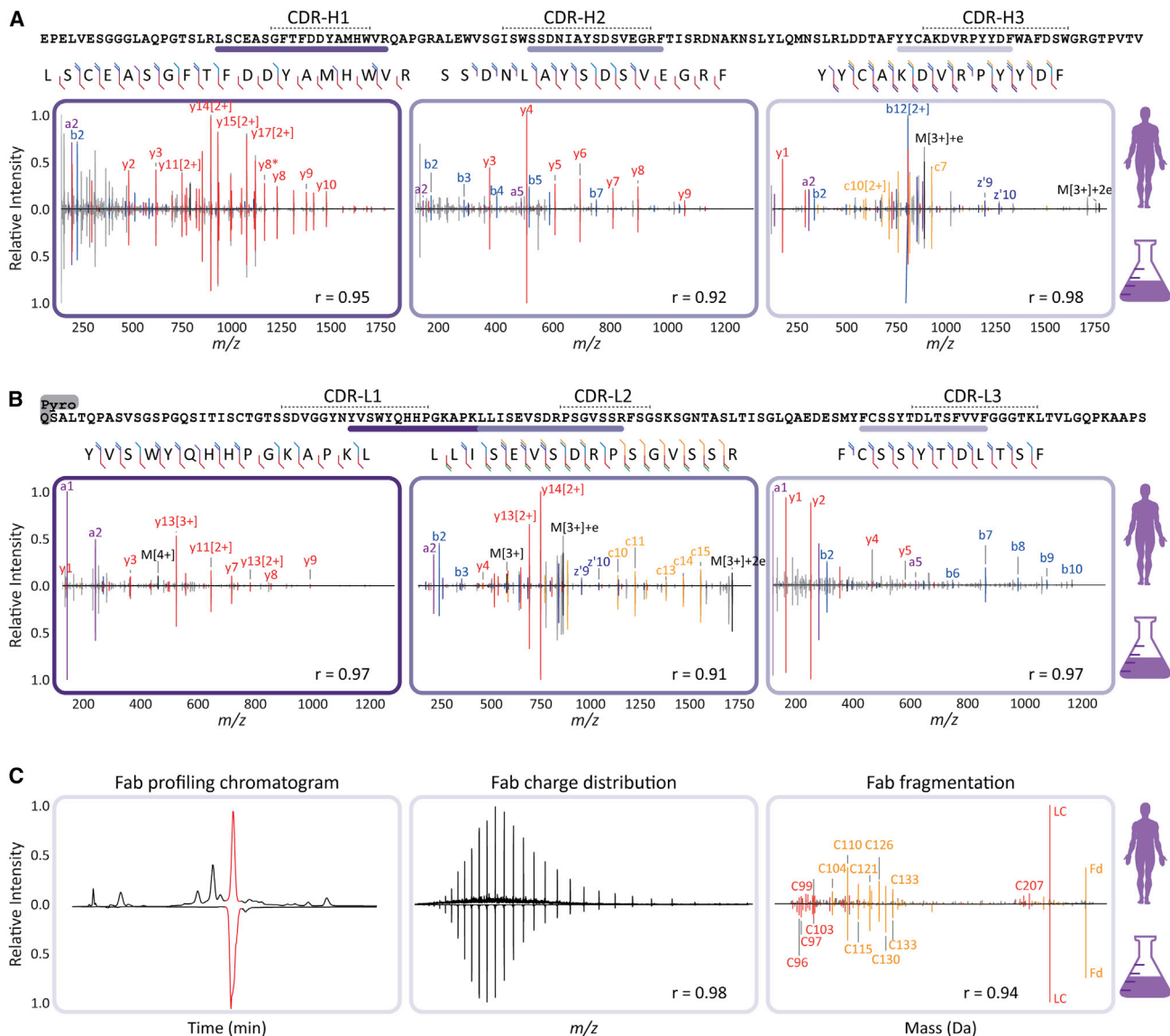
Figure 5. Comparison of sequencing data for clone $^{24.4}1_{47,359.4}$ of the donor F59, and the corresponding recombinant mAb validates the correctness of the *de novo* sequencing approach

(A) Peptide fragmentation spectra of CDR-spanning peptides from the HC of the dominant $^{24.4}1_{47,359.4}$ clone with, mirrored to each other, annotated spectra from the donor (top) and the recombinant IgG1 (bottom).

(B) Peptide fragmentation spectra of CDR-spanning peptides from the LC of the dominant $^{24.4}1_{47,359.4}$ clone with, mirrored to each other, annotated spectra from the donor (top) and the recombinant IgG1 (bottom). Spectra in (A) and (B) are annotated with *a*-ions in purple, *b*-ions in blue, *y*-ions in red, *c*-ions in orange, and *z*-ions in dark blue. Corresponding fragmentation maps are displayed above each spectral pair.

(C) Comparison of the middle-down LC-MS/MS analysis of the $^{24.4}1_{47,359.4}$ clone and the recombinant IgG1. Shown are the base peak chromatograms (left panel), the charge-state distributions detected in MS1 of the Fab (middle panel), and the deconvoluted ETD fragmentation spectra for the donor (top) and recombinant (bottom) IgG (right panel). The Pearson correlation coefficients (r) calculated for all demonstrated spectral pairs in (A), (B), and (C) are indicated at the bottom of each spectrum.

Such a direct comparison of spectral features was extended to the middle-down analysis, used for obtaining sequence tags of the intact Fab (Figure 5C). The intact recombinant Fab displayed a nearly identical retention time profile when compared with the plasma clone $^{24.4}1_{47,359.4}$ Fab (Figure 5C, left panel). Furthermore, both clone $^{24.4}1_{47,359.4}$ and the recombinant mAb emerged with nearly identical charge distributions (Figure 5C, middle panel), whereby the slight dif-

ferences in the distribution are likely due to the underlying background of co-eluting Fabs in the plasma-derived sample. Nevertheless, the masses detected for the two Fabs were identical, i.e., within a 20-ppm mass error. Moreover, when the intact Fabs were subjected to ETD, alike fragment masses and retention times for both the light chain and Fd were observed, comparing the recombinant mAb with the plasma-derived clone. Finally, the generated lower mass fragment

ions used for sequence-tag generation were also very similar (Figure 5C, right panel). Likewise, the in-solution reduction of the Fabs revealed that there were no mass differences between the donor clone and recombinant mAb (light chain and Fd mass within 10 ppm).

Based on all these data, we can conclusively state that the sequence of the plasma-derived clone $^{24.4}$ 1 $_{47,359.4}$ is identical to that of the recombinant mAb, with practically identical data observed at every step of our integrative *de novo* sequencing approach. This not only validates the accuracy of the IgG1 sequence that we obtained for clone $^{24.4}$ 1 $_{47,359.4}$ but also reinforces that the methodology presented here can be used to derive the correct full sequences from individual clones even when they are in a background of other plasma (highly sequence-homologous) IgG1 clones. Although this whole analysis pipeline is still quite arduous, requiring manual validation throughout the process, we consider this proof of concept a major step forward and expect that further fine-tuning of the algorithms will enhance the throughput in the future.

DISCUSSION

The human body can make billions of different antibodies, stemming from the versatile and complex recombination process, accompanied by additional somatic hypermutations, helping us to adapt to a life-long exposure to various pathogens. Here, we demonstrate that it is feasible to profile the IgG1 repertoire of individual donors qualitatively and quantitatively by LC-MS, following the capturing of IgGs from plasma and analyzing the generated IgG1 Fab fragments. From this technical advance, one of the key observations we make is that in all studied donors at all time points, only a limited number of IgG1 clones dominate an individual's repertoire. In all donors, the 30 most abundant clones make up two-thirds of all detected circulating IgG1 molecules; in one donor just two clones contributed ∼50% to the detected serum population of IgG1 molecules. The IgG1 clonal profiles are found to be unique for each donor. Within a donor, the profiles are highly similar across time, but they also adapt to physiological changes (e.g., sepsis). The mass-spectrometry-based approach requires only minute amounts of plasma (∼10 μL) and does not involve labor-intensive enrichment protocols. We further show that specific IgG clones can be extracted from the plasma and analyzed in depth, ultimately leading to the mass-spectrometry-based *de novo* sequencing of the whole Fab molecule. Therefore, one of the holy grails in proteomics, *de novo* sequencing of antibodies directly extracted from plasma, seems to be within reach.

The ultimate mature sequence of $^{24.4}$ 1 $_{47,359.4}$ clone we sequenced here revealed that around 13% and 16% of the amino acids of the V-regions of light and heavy chains were different when compared with the closest germline sequence match within the IMGT database. This number of mutations is higher than the reported average (7%) for IgG1 heavy-chain variable regions, as determined from RNA sequences (Kitaura et al., 2017). This suggests that DNA/RNA templates of the IgG sequences can be helpful, but for obtaining the correct sequence of the circulating clone, analyzing sequences at the level of the proteins will be essential.

The ability to *de novo* sequence the whole Fab molecule is the result of combining, iteratively, middle-down and bottom-up proteomics data. An alternative strategy employed is to combine bottom-up proteomics data with BCR sequences using RNA sequencing of one donor to generate a database to match this donor's Ig bottom-up proteomics data against (Lee et al., 2016). Although also very powerful, a recent application of this approach highlighted further the relevance of antibody sequencing at the level of proteins, when it was shown that for the six potent anti-HIV1 antibodies found by antigen-specific single-B-cell sequencing, only three could also be detected in circulation as IgG protein products (Williams et al., 2017). All these issues highlight the necessity of direct analysis of the serum Ig repertoire at the protein level, as we now demonstrated here to be feasible.

Longitudinal quantitative clonal profiling, as presented here, opens a myriad of future prospects, both fundamental and applied. It allows to advance our understanding of B cell biology and antibody dynamics. Historically, general observations have been made about antibody half-lives using a single dose of labeled antibodies (Morell et al., 1970) or by determining the restoration of normal IgG levels following high-dose administrations of intravenous IgG (Melamed et al., 2018). Through the method presented here, we can monitor the longitudinal abundance of each single clone in the circulation and monitor how it responds to changes in the donor's physiology. Given the approximately 20-day IgG antibody half-life, it is expected that in the time span studied here (10–63 days), a decay in the concentration of clones would be detected. Indeed, we do observe several diminishing clones as depicted in Figures 2 and S4. However, other patterns are also observed, indicative of continuous production of the clone, even over a time window of two months. This is in line with previous reports on the presence of and production of antibodies by long-lived plasma cells, both in mice and humans (Bernasconi et al., 2002; Manz et al., 1997; Slifka et al., 1998). In addition, persistence of autoreactivity has been reported before (Tebani et al., 2020), as well as persistence of antibody clonotypes detected by CDR-H3 proteomics (Lee et al., 2019). From our data, we cannot derive information regarding (auto)reactivity, but the data do suggest that long-term stability is not exclusive to autoantibodies and is instead a common phenomenon.

In summary, the *de novo* sequencing at the protein level of IgG clones circulating in plasma is shown here to be feasible. We demonstrate the synergistic power of combining iteratively peptide- and protein-centric-based sequencing, which is capable of not only identifying the distinct alleles from which a clone originates but also its entire mature sequence. The work presented here is still quite a laborious proof of concept. Our aim for the future is quantitative monitoring and mass-spectrometry-based sequencing of multiple serum immunoglobulins directly at the protein level, i.e., as they appear in circulation and function in the human immune response.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY

Cell Systems
Article

AUTHOR CONTRIBUTIONS

A.J.R.H. conceived the idea for this study. A.B., S.T., and A.J.R.H. designed the research, planned the experiments, and supervised the project. A.B., M.H., M.A.d.B., and D.M.H.v.R. performed the IgG capture from serum and subsequent generation of Fabs, subsequently analyzed by intact LC-MS. M.H. developed the repertoire profiling bioinformatics workflow. S.T. performed the middle-down MS/MS analysis of the clones and generated the bioinformatics workflow to analyze the data. W.P., M.H., D.S., and J.S. performed all bottom-up proteomics experiments and subsequent data analysis. J.-F.G. and B.d.G. contributed to study design and data analysis. M.V., M.J.M.B., and O.L.C. provided the MARS cohort samples and selected the patient population used in this study. A.B., M.H., S.T., and A.J.R.H. wrote the original draft, which was read, improved, and approved by all co-authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Amez Martín, M., Wuhrer, M., and Falck, D. (2021). Serum and plasma immunoglobulin G Fc N-glycosylation is stable during storage. J. Proteome Res. *20*, 2935–2941. https://doi.org/10.1021/acs.jproteome.1c00148.

Apostoaei, A., and Trabalka, J.R. (2012). Review, synthesis, and application of information on the human lymphatic system to radiation dosimetry for chronic lymphocytic leukemia. SENES Oak Ridge, Inc. - Center for Risk Analysis. March 2012. https://www.cdc.gov/niosh/OCAS/pdfs/irep/raddoscll.pdf.

Barnidge, D.R., Dasari, S., Botz, C.M., Murray, D.H., Snyder, M.R., Katzmann, J.A., Dispenzieri, A., and Murray, D.L. (2014a). Using mass spectrometry to monitor monoclonal immunoglobulins in patients with a monoclonal gammopathy. J. Proteome Res. *13*, 1419–1427. https://doi.org/10.1021/pr400985k.

Barnidge, D.R., Dasari, S., Ramirez-Alvarado, M., Fontan, A., Willrich, M.A., Tschumper, R.C., Jelinek, D.F., Snyder, M.R., Dispenzieri, A., Katzmann, J.A., and Murray, D.L. (2014b). Phenotyping polyclonal kappa and lambda light chain molecular mass distributions in patient serum using mass spectrometry. J. Proteome Res. *13*, 5198–5205. https://doi.org/10.1021/pr5005967.

Bernasconi, N.L., Traggiai, E., and Lanzavecchia, A. (2002). Maintenance of serological memory by polyclonal activation of human memory B cells. Science *298*, 2199–2202. https://doi.org/10.1126/science.1076071.

Bondt, A., Rombouts, Y., Selman, M.H.J., Hensbergen, P.J., Reiding, K.R., Hazes, J.M.W., Dolhain, R.J.E.M., and Wuhrer, M. (2014). Immunoglobulin G (IgG) fab glycosylation analysis using a new mass spectrometric high-throughput profiling method reveals pregnancy-associated changes. Mol. Cell. Proteomics *13*, 3029–3039. https://doi.org/10.1074/mcp.M114.039537.

Bondt, A., Wuhrer, M., Kuijper, T.M., Hazes, J.M., and Dolhain, R.J. (2016). Fab glycosylation of immunoglobulin G does not associate with improvement of rheumatoid arthritis during pregnancy. Arthritis Res. Ther. *18*, 274. https://doi.org/10.1186/s13075-016-1172-1.

Briney, B., Inderbitzin, A., Joyce, C., and Burton, D.R. (2019). Commonality despite exceptional diversity in the baseline human antibody repertoire. Nature *566*, 393–397. https://doi.org/10.1038/s41586-019-0879-y.

Cassidy, J.T., and Nordby, G.L. (1975). Human serum immunoglobulin concentrations: prevalence of immunoglobulin deficiencies. J. Allergy Clin. Immunol. *55*, 35–48. https://doi.org/10.1016/s0091-6749(75)80006-6.

Chen, J., Zheng, Q., Hammers, C.M., Ellebrecht, C.T., Mukherjee, E.M., Tang, H.Y., Lin, C., Yuan, H., Pan, M., Langenhan, J., et al. (2017). Proteomic analysis of pemphigus autoantibodies indicates a larger, more diverse, and more dynamic repertoire than determined by B cell genetics. Cell Rep *18*, 237–247. https://doi.org/10.1016/j.celrep.2016.12.013.

Corti, D., Misasi, J., Mulangu, S., Stanley, D.A., Kanekiyo, M., Wollen, S., Ploquin, A., Doria-Rose, N.A., Staupe, R.P., Bailey, M., et al. (2016). Protective monotherapy against lethal Ebola virus infection by a potently neutralizing antibody. Science *351*, 1339–1342. https://doi.org/10.1126/science.aad5224.

den Boer, M.A., Greisch, J.F., Tamara, S., Bondt, A., and Heck, A.J.R. (2020). Selectivity over coverage in de novo sequencing of IgGs. Chem. Sci. *11*, 11886–11896. https://doi.org/10.1039/D0SC03438J.

Dyer, O. (2019). Two Ebola treatments halve deaths in trial in DRC outbreak. BMJ *366*, l5140. https://doi.org/10.1136/bmj.l5140.

Fornelli, L., Ayoub, D., Aizikov, K., Liu, X., Damoc, E., Pevzner, P.A., Makarov, A., Beck, A., and Tsybin, Y.O. (2017). Top-down analysis of immunoglobulin G isotypes 1 and 2 with electron transfer dissociation on a high-field Orbitrap mass spectrometer. J. Proteomics *159*, 67–76. https://doi.org/10.1016/j.jprot.2017.02.013.

Georgiou, G., Ippolito, G.C., Beausang, J., Busse, C.E., Wardemann, H., and Quake, S.R. (2014). The promise and challenge of high-throughput sequencing of the antibody repertoire. Nat. Biotechnol. *32*, 158–168. https://doi.org/10.1038/nbt.2782.

Gonzalez-Quintela, A., Alende, R., Gude, F., Campos, J., Rey, J., Meijide, L.M., Fernandez-Merino, C., and Vidal, C. (2008). Serum levels of immunoglobulins (IgG, IgA, IgM) in a general adult population and their relationship with alcohol consumption, smoking and common metabolic abnormalities. Clin. Exp. Immunol. *151*, 42–50. https://doi.org/10.1111/j.1365-2249.2007.03545.x.

Guthals, A., Gan, Y., Murray, L., Chen, Y., Stinson, J., Nakamura, G., Lill, J.R., Sandoval, W., and Bandeira, N. (2017). De novo MS/MS sequencing of native human antibodies. J. Proteome Res. *16*, 45–54. https://doi.org/10.1021/acs.jproteome.6b00608.

Hafkenscheid, L., Bondt, A., Scherer, H.U., Huizinga, T.W., Wuhrer, M., Toes, R.E., and Rombouts, Y. (2017). Structural analysis of variable domain glycosylation of anti-citrullinated protein antibodies in rheumatoid arthritis reveals the presence of highly sialylated glycans. Mol. Cell. Proteomics 16, 278–287. https://doi.org/10.1074/mcp.M116.062919.

He, L., Anderson, L.C., Barnidge, D.R., Murray, D.L., Dasari, S., Dispenzieri, A., Hendrickson, C.L., and Marshall, A.G. (2019). Classification of plasma cell disorders by 21 tesla Fourier transform ion cyclotron resonance top-down and middle-down MS/MS analysis of monoclonal immunoglobulin light chains in human serum. Anal. Chem. 91, 3263–3269. https://doi.org/10.1021/acs.anal-chem.8b03294.

He, L., Anderson, L.C., Barnidge, D.R., Murray, D.L., Hendrickson, C.L., and Marshall, A.G. (2017). Analysis of monoclonal antibodies in human serum as a model for clinical monoclonal gammopathy by use of 21 tesla FT-ICR top-down and middle-down MS/MS. J. Am. Soc. Mass Spectrom. 28, 827–838. https://doi.org/10.1007/s13361-017-1602-6.

Hunter, J.D. (2007). Matplotlib: A 2D graphics environment. Comput. Sci. Eng. 9, 90–95. https://doi.org/10.1109/MCSE.2007.55.

Jones, B.E., Brown-Augsburger, P.L., Corbett, K.S., Westendorf, K., Davies, J., Cujec, T.P., Wiethoff, C.M., Blackbourne, J.L., Heinz, B.A., Foster, D., et al. (2021). The neutralizing antibody, LY-CoV555, protects against SARS-CoV-2 infection in nonhuman primates. Sci. Transl. Med. 13, eabf1906. https://doi.org/10.1126/scitranslmed.abf1906.

Kitaura, K., Yamashita, H., Ayabe, H., Shini, T., Matsutani, T., and Suzuki, R. (2017). Different somatic hypermutation levels among antibody subclasses disclosed by a new next-generation sequencing-based antibody repertoire analysis. Front. Immunol. 8, 389. https://doi.org/10.3389/fimmu.2017.00389.

Klein Klouwenberg, P.M., Ong, D.S., Bos, L.D., de Beer, F.M., van Hooijdonk, R.T., Huson, M.A., Straat, M., van Vught, L.A., Wieske, L., Horn, J., et al. (2013). Interobserver agreement of Centers for Disease Control and Prevention criteria for classifying infections in critically ill patients. Crit. Care Med. 41, 2373–2378. https://doi.org/10.1097/CCM.0b013e3182923712.

Lavinder, J.J., Wine, Y., Giesecke, C., Ippolito, G.C., Horton, A.P., Lungu, O.I., Hoi, K.H., DeKosky, B.J., Murrin, E.M., Wirth, M.M., et al. (2014). Identification and characterization of the constituent human serum antibodies elicited by vaccination. Proc. Natl. Acad. Sci. USA 111, 2259–2264. https://doi.org/10.1073/pnas.1317793111.

Lee, J., Boutz, D.R., Chromikova, V., Joyce, M.G., Vollmers, C., Leung, K., Horton, A.P., DeKosky, B.J., Lee, C.H., Lavinder, J.J., et al. (2016). Molecular-level analysis of the serum antibody repertoire in young adults before and after seasonal influenza vaccination. Nat. Med. 22, 1456–1464. https://doi.org/10.1038/nm.4224.

Lee, J., Paparoditis, P., Horton, A.P., Frühwirth, A., McDaniel, J.R., Jung, J., Boutz, D.R., Hussein, D.A., Tanno, Y., Pappas, L., et al. (2019). Persistent antibody clonotypes dominate the serum response to influenza over multiple years and repeated vaccinations. Cell Host Microbe 25, 367–376.e5. https://doi.org/10.1016/j.chom.2019.01.010.

Lefranc, M.P., Giudicelli, V., Ginestoux, C., Bodmer, J., Müller, W., Bontrop, R., Lemaitre, M., Malik, A., Barbié, V., and Chaume, D. (1999). IMGT, the international ImMunoGeneTics database. Nucleic Acids Res 27, 209–212. https://doi.org/10.1093/nar/27.1.209.

Lefranc, M.-P., and Lefranc, G.e. (2001). The Immunoglobulin Factsbook (Academic Press).

Manz, R.A., Thiel, A., and Radbruch, A. (1997). Lifetime of plasma cells in the bone marrow. Nature 388, 133–134. https://doi.org/10.1038/40540.

McKinney, W. (2010). Data structures for statistical computing in python. Proc. of the 9th python in science conf. 56–61.

Melamed, I.R., Borte, M., Trawnicek, L., Kobayashi, A.L., Kobayashi, R.H., Knutsen, A., Gupta, S., Smits, W., Pituch-Noworolska, A., Strach, M., et al. (2018). Pharmacokinetics of a novel human intravenous immunoglobulin 10% in patients with primary immunodeficiency diseases: analysis of a phase III, multicentre, prospective, open-label study. Eur. J. Pharm. Sci. 118, 80–86. https://doi.org/10.1016/j.ejps.2018.03.007.

Mills, J.R., Barnidge, D.R., and Murray, D.L. (2015). Detecting monoclonal immunoglobulins in human serum using mass spectrometry. Methods 81, 56–65. https://doi.org/10.1016/j.ymeth.2015.04.020.

Morell, A., Terry, W.D., and Waldmann, T.A. (1970). Metabolic properties of IgG subclasses in man. J. Clin. Invest. 49, 673–680. https://doi.org/10.1172/JCI106279.

Mulangu, S., Dodd, L.E., Davey, R.T., Jr., Tshiani Mbaya, O., Proschan, M., Mukadi, D., Lusakibanza Manzo, M., Nzolo, D., Tshomba Oloma, A., Ibanda, A., et al. (2019). A randomized, controlled trial of Ebola virus disease therapeutics. N. Engl. J. Med. 381, 2293–2303. https://doi.org/10.1056/NEJMoa1910993.

Peng, W., Pronker, M.F., and Snijder, J. (2021). Mass spectrometry-based de novo sequencing of monoclonal antibodies using multiple proteases and a dual fragmentation scheme. J. Proteome Res. 20, 3559–3566. https://doi.org/10.1021/acs.jproteome.1c00169.

Schroeder, H.W., Jr. (2006). Similarity and divergence in the development and expression of the mouse and human antibody repertoires. Dev. Comp. Immunol. 30, 119–135. https://doi.org/10.1016/j.dci.2005.06.006.

Sen, K.I., Tang, W.H., Nayak, S., Kil, Y.J., Bern, M., Ozoglu, B., Ueberheide, B., Davis, D., and Becker, C. (2017). Automated antibody de novo sequencing and its utility in biopharmaceutical discovery. J. Am. Soc. Mass Spectrom. 28, 803–810. https://doi.org/10.1007/s13361-016-1580-0.

Sharpley, F.A., Manwani, R., Mahmood, S., Sachchithanantham, S., Lachmann, H.J., Gillmore, J.D., Whelan, C.J., Fontana, M., Hawkins, P.N., and Wechalekar, A.D. (2019). A novel mass spectrometry method to identify the serum monoclonal light chain component in systemic light chain amyloidosis. Blood Cancer J 9 (16). https://doi.org/10.1038/s41408-019-0180-1.

Shaw, J.B., Liu, W., Vasil Ev, Y.V., Bracken, C.C., Malhan, N., Guthals, A., Beckman, J.S., and Voinov, V.G. (2020). Direct determination of antibody chain pairing by top-down and middle-down mass spectrometry using electron capture dissociation and ultraviolet photodissociation. Anal. Chem. 92, 766–773. https://doi.org/10.1021/acs.analchem.9b03129.

Singer, M., Deutschman, C.S., Seymour, C.W., Shankar-Hari, M., Annane, D., Bauer, M., Bellomo, R., Bernard, G.R., Chiche, J.D., Coopersmith, C.M., et al. (2016). The third international consensus definitions for sepsis and septic shock (Sepsis-3). JAMA 315, 801–810. https://doi.org/10.1001/jama.2016.0287.

Slifka, M.K., Antia, R., Whitmire, J.K., and Ahmed, R. (1998). Humoral immunity due to long-lived plasma cells. Immunity 8, 363–372. https://doi.org/10.1016/S1074-7613(00)80541-5.

Smith, T.F., and Waterman, M.S. (1981). Identification of common molecular subsequences. J. Mol. Biol. 147, 195–197. https://doi.org/10.1016/0022-2836(81)90087-5.

Soto, C., Bombardi, R.G., Branchizio, A., Kose, N., Matta, P., Sevy, A.M., Sinkovits, R.S., Gilchuk, P., Finn, J.A., and Crowe, J.E., Jr. (2019). High frequency of shared clonotypes in human B cell receptor repertoires. Nature 566, 398–402. https://doi.org/10.1038/s41586-019-0934-8.

Spoerry, C., Hessle, P., Lewis, M.J., Paton, L., Woof, J.M., and von Pawel-Rammingen, U. (2016). Novel IgG-degrading enzymes of the IgdE protease family link substrate specificity to host tropism of Streptococcus species. PLoS One 11, e0164809. https://doi.org/10.1371/journal.pone.0164809.

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, 356–364. https://doi.org/10.1016/j.jmb.2016.08.019.

Tabb, D.L., Ma, Z.Q., Martin, D.B., Ham, A.J., and Chambers, M.C. (2008). DirecTag: accurate sequence tags from peptide MS/MS through statistical scoring. J. Proteome Res. 7, 3838–3846. https://doi.org/10.1021/pr800154p.

Tebani, A., Gummesson, A., Zhong, W., Koistinen, I.S., Lakshmikanth, T., Olsson, L.M., Boulund, F., Neiman, M., Stenlund, H., Hellström, C., et al. (2020). Integration of molecular profiles in a longitudinal wellness profiling cohort. Nat. Commun. 11, 4487. https://doi.org/10.1038/s41467-020-18148-7.

Tran, N.H., Rahman, M.Z., He, L., Xin, L., Shan, B., and Li, M. (2016). Complete de novo assembly of monoclonal antibody sequences. Sci. Rep. *6*, 31730. https://doi.org/10.1038/srep31730.

van de Waterbeemd, M., Tamara, S., Fort, K.L., Damoc, E., Franc, V., Bieri, P., Itten, M., Makarov, A., Ban, N., and Heck, A.J.R. (2018). Dissecting ribosomal particles throughout the kingdoms of life using advanced hybrid mass spectrometry methods. Nat. Commun. *9*, 2493. https://doi.org/10.1038/s41467-018-04853-x.

van der Giessen, M., Rossouw, E., van Veen, T.A., van Loghem, E., Zegers, B.J., and Sander, P.C. (1975). Quantification of IgG subclasses in sera of normal adults and healthy children between 4 and 12 years of age. Clin. Exp. Immunol. *21*, 501–509.

van der Walt, S., Colbert, S.C., and Varoquaux, G. (2011). The NumPy array: A structure for efficient numerical computation. Comput. Sci. Eng. *13*, 22–30. https://doi.org/10.1109/MCSE.2011.37.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2.

Vollmers, C., Sit, R.V., Weinstein, J.A., Dekker, C.L., and Quake, S.R. (2013). Genetic measurement of memory B-cell recall using antibody repertoire sequencing. Proc. Natl. Acad. Sci. USA *110*, 13463–13468. https://doi.org/10.1073/pnas.1312146110.

Wang, Z., Liu, X., Muther, J., James, J.A., Smith, K., and Wu, S. (2019). Top-down mass spectrometry analysis of human serum autoantibody antigen-binding fragments. Sci. Rep. *9*, 2345. https://doi.org/10.1038/s41598-018-38380-y.

Williams, L.D., Ofek, G., Schätzle, S., McDaniel, J.R., Lu, X., Nicely, N.I., Wu, L., Lougheed, C.S., Bradley, T., Louder, M.K., et al. (2017). Potent and broad HIV-neutralizing antibodies in memory B cells and plasma. Sci. Immunol. *2*, eaal2200. https://doi.org/10.1126/sciimmunol.aal2200.

Wine, Y., Boutz, D.R., Lavinder, J.J., Miklos, A.E., Hughes, R.A., Hoi, K.H., Jung, S.T., Horton, A.P., Murrin, E.M., Ellington, A.D., et al. (2013). Molecular deconvolution of the monoclonal antibodies that comprise the polyclonal serum response. Proc. Natl. Acad. Sci. USA *110*, 2993–2998. https://doi.org/10.1073/pnas.1213737110.

Zabrouskov, V., Senko, M.W., Du, Y., Leduc, R.D., and Kelleher, N.L. (2005). New and automated MSn approaches for top-down identification of modified proteins. J. Am. Soc. Mass Spectrom. *16*, 2027–2038. https://doi.org/10.1016/j.jasms.2005.08.004.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Longitudinal plasma samples	Molecular Diagnosis and Risk Stratification of Sepsis (MARS) biorepository	ClinicalTrials.gov: NCT01905033
Longitudinal EDTA plasma samples from two healthy Caucasian donors	Precision Med (Solana Beach, CA, US)	7005-8200
Chemicals, peptides, and recombinant proteins		
CaptureSelect™ FcXL affinity matrix	Thermo Fisher	194328010
FabALACTICA	Genovis	A0-AG1-020
Ni-NTA beads	Thermo Fisher	R90101
Trastuzumab	Roche, Penzberg, Germany	N/A
Alemtuzumab	Genmab, Utrecht, The Netherlands	N/A
Cetuximab, Rituximab, Bevacizumab, and Infliximab	Evidentic, Berlin, Germany	N/A
Oasis HLB	Waters Corporation, Milford, MA, USA	WAT058951
Deposited data		
Bottom-up dataset – sequencing	This paper	https://doi.org/10.6084/m9.figshare.13194005
Middle-down dataset – sequencing	This paper	https://doi.org/10.6084/m9.figshare.13193996
Profiling LC-MS dataset	This paper	https://doi.org/10.6084/m9.figshare.13182743
Software and algorithms		
BioPharmaFinder 3.2	Thermo Fisher	OPTON-30985
Pandas 1.2.4, Numpy 1.20.2, Scipy 1.6.2, Matplotlib 3.3.4, and Seaborn 0.11.1	McKinney, 2010; van der Walt et al., 2011; Virtanen et al., 2020; Hunter, 2007	N/A
PEAKS X	Bioinformatics Solutions Inc., Waterloo, ON, Canada	https://www.bioinfor.com/peaks-studio/
Xtract	Zabrouskov et al., 2005	N/A
ReSpect	Thermo Fisher Scientific, Bremen, Germany	N/A
DirecTag algorithm	Tabb et al., 2008	N/A
Thermo Proteome Discoverer (version 2.3.0.523)	Thermo Fisher Scientific, Bremen, Germany	OPTON-30945
Data analysis code	This paper	https://doi.org/10.6084/m9.figshare.14794173
Scripts used to generate the figures	This paper	https://doi.org/10.6084/m9.figshare.14794173
Other		
MAbPac™ Reversed Phase HPLC Columns	Thermo Fisher	303184

RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Albert Heck (a.j.r.heck@uu.nl)

Materials availability
This study did not generate new unique reagents.

Data and code availability

- Raw-files source data have been deposited at figshare and are publicly available at
- Bottom-up dataset – sequencing: https://doi.org/10.6084/m9.figshare.13194005
- Middle-down dataset – sequencing: https://doi.org/10.6084/m9.figshare.13193996
- Profiling dataset: https://doi.org/10.6084/m9.figshare.13182743

- Data analysis code has been deposited at figshare and is publicly available at https://doi.org/10.6084/m9.figshare.14794173.
- The scripts used to generate the figures reported in this paper have been deposited at figshare and are publicly available at https://doi.org/10.6084/m9.figshare.14794173

Any additional information required to reproduce this work is available from the Lead Contact.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human subjects

We obtained longitudinal plasma samples from the Molecular Diagnosis and Risk Stratification of Sepsis (MARS) biorepository (ClinicalTrials.gov identifier NCT01905033), for which subjects had been included in the mixed ICU of a tertiary teaching hospital in the Netherlands (University Medical Centre Utrecht, Utrecht) since January 2011. Donors were enrolled via an opt-out consent method approved by the institutional review board of the UMC Utrecht (IRB no. 10-1056C). Daily leftover EDTA plasma (obtained from blood drawn during routine care) was stored at -80 °C until use.

For the current study, we included eight patients with esophageal or gastroesophageal junction cancer who underwent an elective esophagectomy and gastric pull-through procedure and had subsequently developed an infectious complication (i.e., anastomotic leakage or pneumonia). These patients were all admitted to the Intensive Care Unit on two occasions. The first admission concerned routine observation after elective resection followed by an uncomplicated ICU stay of fewer than 2 days. All patients were subsequently readmitted to the ICU due to sepsis. For all episodes of sepsis, microbiological cultures were obtained either before or during ICU readmission, and clinical infection was adjudicated highly plausible according to pre-defined criteria (Klein Klouwenberg et al., 2013). Furthermore, all infectious episodes met SIRS criteria and had a Sequential Organ Failure Assessment (SOFA) score ≥ 2, thus fulfilling current Sepsis-3 definitions (Singer et al., 2016). All patients were ultimately discharged from ICU in a clinically stable condition. We analyzed plasma samples obtained at four well-defined time points: within 24 hours of surgery when no signs of sepsis were present (sample T1), on two consecutive days after onset of sepsis (samples T2 and T3), and upon ICU discharge following resolution of sepsis (sample T4). Patient characteristics are summarized in Table S2. All patients had been treated with neoadjuvant chemoradiotherapy prior to surgery. None of the patients received treatment with immunoglobulins or monoclonal antibodies either prior to or during ICU admission.

In addition, longitudinal EDTA plasma samples from two healthy Caucasian donors were purchased from Precision Med (Solana Beach, CA, US). The samples were part of the 'Normal Control Collections', protocol number 7005-8200. These donors were selected having similar characteristics as the sepsis donors regarding age and gender.

METHOD DETAILS

Plasma IgG purification and Fab generation

The IgG purification and generation of Fabs was adapted from an earlier published protocol (Bondt et al., 2014). The FcXL affinity matrix used in the workflow, which binds to the CH3 domain of the IgG constant region, has recently been shown not to provide a bias regarding analysis of the Fc glycosylation residing in the CH2 domain (Amez Martín et al., 2021). Mobicol spin filters were assembled according to the manufacturer's instructions and placed in 2 mL Eppendorf tubes. Then 20 μL FcXL affinity matrix slurry was added to the spin filter, followed by three washing steps with 150 μL PBS, in which the liquid was removed by centrifugation for 1 min at 1000 × g. Two additional washing steps with 150 μL were performed. After washing, the 2 mL tube was replaced by a 1.5 mL tube. The affinity matrix was resuspended in 150 μL PBS, and 10 μL plasma was added. Furthermore, 1 μL of a solution containing two known mAbs at 200 μg/mL each was added, corresponding to 20 μg/mL when calculated to the plasma concentration. The samples were then incubated, under shaking conditions for one hour at room temperature. After the incubation, the flow-through was collected, and the affinity matrix with bound IgGs was washed four times with 150 μL PBS. Finally, 50 μL PBS containing 100 U of the IgdE (FabALACTICA; Genovis AB, Lund, Sweden) protease was added before incubating on a thermal shaker at 37 °C for >16 hours. After the incubation, 10 μL of Ni-NTA beads were added to bind and remove the His-tagged protease, whereafter the samples were incubated for an additional 30 minutes. The flow through after centrifugation contained the Fab fragments generated from IgG1.

Method optimization and validation using a mixture of recombinant mAbs

In an array of experiments, we optimized and validated the robustness and accuracy of our IgG1 capturing approach, the generation of the Fab fragments and the analysis of the Fab LC-MS profiles. Therefore, we prepared a mixture of six IgG1 monoclonal antibodies, including the two spiked into every plasma sample subsequently analyzed. For the method optimization the 6 recombinant mAbs

were added to the plasma of a single donor in different quantities: 4,000, 800, 200, 20, 2, or 0.5 ng per mAb. The mAbs used were trastuzumab (Roche, Penzberg, Germany), alemtuzumab (Genmab, Utrecht, The Netherlands), the Fab glycosylated cetuximab, rituximab, bevacizumab, and infliximab (Evidentic, Berlin, Germany). Fab sequences and theoretical masses of these mAbs, including the most abundant cetuximab glycoforms, are shown in Table S1. All these samples were injected once, except for the 200 ng sample which was injected three times to provide additional injection replicates.

LC-MS(/MS)

Reversed-phase liquid chromatography was performed by using a Thermo Scientific Vanquish Flex UHPLC instrument, equipped with a 1 mm x 150 mm MAbPac RP analytical column, directly coupled to an Orbitrap Fusion Lumos Tribrid (Thermo Scientific, San Jose, CA, USA) or Q Exactive HF-X mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). The column preheater, as well as the analytical column chamber, were heated to 80 °C during chromatographic separation. Both samples, either containing intact Fabs or separate Fab chains, were separated over 62 min at a 150 μL/min flow rate. Gradient elution was achieved by using two mobile phases A (0.1% HCOOH in Milli-Q) and B (0.1% HCOOH in CH_3CN) and ramping up B from 10 to 25% over one minute, from 25 to 40% over 55 min, and from 40 to 95% over one minute. MS data were collected with the instrument operating in Intact Protein and Low Pressure mode. Spray voltage was set at 3.3 kV, capillary temperature 350 °C, probe heater temperature 100 °C, sheath gas flow 35, auxiliary gas flow 10, and source-induced dissociation was set at 15 V. The electrospray voltage was applied after 2 min to prevent the salts in the sample from entering the MS. Intact Fabs were recorded with a resolution setting of 7,500 (@ *m/z* 200) in MS1, which allows for better detection of charge distributions of the large proteins (> 30 kDa) (van de Waterbeemd et al., 2018). Separate Fab chains were analyzed with a resolution setting of 120,000 (@ *m/z* 200) in MS1, which allows for more accurate mass detection of smaller proteins (< 30 kDa). MS1 scans were acquired in a range of 500-3,000 Th with the 250% AGC target and a maximum injection time set to either 50 ms for the 7,500 resolution or 250 ms for the 120,000 resolution. In MS1, 2 μscans were recorded for the 7,500 resolution and 5 μscans for the 120,000 resolution per scan. Data-dependent mode was defined by the number of scans: single scan for intact Fabs and two scans for separate Fab chains. In both cases, MS/MS scans were acquired with a resolution of 120,000, a maximum injection time of 500 ms, a 1,000% AGC target, and 5 μscans averaged and recorded per scan. The ions of interest were mass-selected by quadrupole in a 4 Th isolation window and accumulated to the AGC target prior to fragmentation. The electron-transfer dissociation (ETD) was performed using the following settings: 16 ms reaction time, a maximum injection time of 200 ms, and the AGC target of 1e6 for the ETD reagent. For the data-dependent MS/MS acquisition strategy, the intensity threshold was set to 2e5 of minimum precursor intensity. MS/MS scans were recorded in the range of *m/z* = 350-5,000 Th using high mass range quadrupole isolation.

Clonal profiling data analysis

Masses were retrieved from the generated RAW files using BioPharmaFinder 3.2 (Thermo Scientific). Deconvolution was performed using the ReSpect algorithm between 5 and 57 min using 0.1 min sliding window with 50% offset and a merge tolerance of 50 ppm, with noise rejection set at 95%. The output mass range was set at 10,000 to 100,000 with a target mass of 48,000 and mass tolerance of 30 ppm. Charge states between 10 and 60 were included, and the Intact Protein peak model was selected. Further data analysis was performed using Python 3.8.10 (with libraries: Pandas 1.2.4 (McKinney, 2010), Numpy 1.20.2 (van der Walt et al., 2011), Scipy 1.6.2 (Virtanen et al., 2020), Matplotlib 3.3.4 (Hunter, 2007) and Seaborn 0.11.1). Masses of the BioPharmaFinder identifications (components) were recalculated using an intensity weighted mean considering only the most intense peaks comprising 90% of the total intensity. Furthermore, using the data of two spiked-in mAbs (trastuzumab and alemtuzumab) a mass correction was applied based on the difference between the calculated and observed mAb masses, and similarly, a retention time alignment was applied to minimize deviation between runs.

Components between 45,000 and 53,000 kDa with the most intense charge state above *m/z* 1,000 and score ≥ 40 were considered Fab portions of IgG clones. The clones were matched between runs using average linkage (unweighted pair group method with arithmetic mean UPGMA) L^∞ distance hierarchical clustering. Flat clusters were formed based on a cophenetic distance constraint derived from the mass and retention time tolerance. These tolerances were defined as three times the standard deviation of the mAb standards, which were 1.4 Da and 0.8 min, respectively. Clones within a flat cluster were considered identical between runs.

Peptide-centric (bottom-up) *de novo* sequencing

Clones of interest were captured through fraction collection using the same chromatography setup used for LC-MS profiling. Samples were dried under vacuum and resuspended in a 50 mM ammonium bicarbonate buffer. To boost signal intensity, the fractions were pooled across the time points. Samples were equally split for digestion with four proteases.

For digestion with trypsin, chymotrypsin and thermolysin, a sodium deoxycholate (SDC) buffer was added to a total volume of 80 μL, 200 mM Tris pH 8.5, 10 mM tris(2-carboxyethyl)phosphine (TCEP), 2% (w/v) SDC final concentration. For pepsin digestion, a urea buffer was added to a total volume of 80 μL, 2M Urea, 10 mM TCEP. Samples were denatured for 10 min at 95 °C followed by reduction for 20 min at 37 °C. Next, iodoacetic acid was added to a final concentration of 40 mM and incubated in the dark for 45 min at room temperature for alkylation of free cysteines. Then for trypsin, chymotrypsin and thermolysin 50 mM ammonium bicarbonate buffer was added to a total volume of 100 μL. For pepsin 1 M HCl was added to a final concentration of 0.04 M. 0.1 μg of each protease was added and incubated for 4 hours at 37 °C. After digestion 2 μL HCOOH was added to precipitate

the SDC. SDC was removed by centrifugation for 20 min at max speed (20,817 × *g*) after which the supernatant was moved to a new tube.

Samples were desalted by Oasis HLB (Waters Corporation, Millford, MA, USA) following a 5-step protocol. 1) Sorbent was wetted using 2x 200 μL CH₃CN, 2) followed by equilibration with 2x 200 μL water/10% HCOOH. 3) The sample was loaded and 4) washed with 2x 200 μL water/10% HCOOH. 5) Finally, the sample was eluted using 2x 50 μL water/50% CH₃CN /10% HCOOH and dried down by vacuum centrifuge. Prior to MS analysis samples were reconstituted in 2% HCOOH.

LC-MS/MS

Data acquisition was performed on the Orbitrap Fusion Tribrid Mass Spectrometer (Thermo Scientific, San Jose, CA, USA) coupled to UHPLC 1290 system (Agilent Technologies, Santa Clara, CA, USA). Peptides were trapped (Dr. Maisch Reprosil C18, 3 μm, 2 cm × 100 μm) prior to separation (Agilent Poroshell EC-C18, 2.7 μm, 500 mm × 75 μm). Trapping was performed for 10 min in solvent A (0.1% HCOOH in Milli-Q), and the gradient was as follows: 0 – 13% solvent B (0.1% HCOOH in 80% CH3CN) over 5 min, 13 – 44% solvent B over 65 min, 44 – 100% solvent B over 4 min, and 100% B for 4 min (flow was split to achieve the final flowrate of approximately 200 nL/min). Mass spectrometry data was collected in a data-dependent fashion with survey scans ranging from 350-2,000 Th (resolution of 60,000 @ *m/z* 200), and up to 3 sec for precursor selection and fragmentation with either stepped higher-energy collisional dissociation (HCD) set to [25%, 35%, 50%] or electron transfer dissociation (ETD), used with charge-normalized settings and supplemental activation of 27%. The MS2 spectra were recorded at a resolution of 30,000 (@ *m/z* 200). The AGC targets for both MS and MS2 scans were set to standard within a maximum injection time of 50 and 250 ms, respectively.

Data analysis

Raw LC-MS/MS data were processed using PEAKS X software (Bioinformatics Solutions Inc., Waterloo, ON, Canada) for *de novo* sequencing of peptides. The following parameters were used for *de novo* sequencing: parent mass error tolerance – 12 ppm, fragment mass error tolerance – 0.02 Da, max number of variable PTMs per peptide – 3. Fixed modifications: Carboxymethyl; variable modifications: Oxidation (HW), Oxidation (M), Pyro-glu from E, Pyro-glu from Q, Carboxymethyl (KW, X@N-term), and Carbamylation. Resulting *de novo* peptide tables were exported as.csv files and used for filtering of the IMGT database and determination of the mature [24.4] 1 [47359.4] clone sequences.

Protein-centric (middle-down) de novo sequencing

Fab samples were prepared without treatment as well as under denaturing and reducing conditions for analysis of intact Fab and separate Fab chains, respectively. The latter were denatured and reduced in 10 mM TCEP at 60 °C for 30 min prior to LC-MS/MS analysis. Approximately 2-5 μg of each sample was injected for a single middle-down LC-MS(/MS) experiment using the parameters described above.

Data analysis

Full middle-down MS spectra were deconvoluted with either Xtract (Zabrouskov et al., 2005) or ReSpect (Thermo Fisher Scientific, Bremen, Germany) for isotopically-resolved (separate Fab chains) or unresolved (intact Fabs) data, respectively. Middle-down LC-MS/MS data were charge-deconvoluted and deisotoped into singly-charged mass spectra using the 'Parallel Xtract' node and converted to mascot generic format (mgf) files in Thermo Proteome Discoverer (version 2.3.0.523; Thermo Fisher Scientific, Bremen, Germany). Deconvolution parameters were set as follows: ReSpect: precursor *m/z* tolerance – 0.2 Th; relative abundance threshold – 0 %; precursor mass range from 3 to 100 kDa; precursor mass tolerance of 30 ppm; charge states between 3 and 100. Xtract: signal/noise threshold of 3; *m/z* range – 500-3,000 Th.

For the analysis of the final mature Fab chains of the most abundant clone in the plasma of patient F59, we used an integrative approach that utilizes bottom-up and middle-down data and the international ImMunoGeneTics information system (IMGT) database (Figure 3B). First, the replicate middle-down MS/MS spectra were grouped per deconvoluted mass feature in the LC-MS-only runs by using a 3 Da mass window and a 3 min retention time window. The resulting grouped spectra were merged into a single spectrum, whereby peaks' intensities were combined when they coincided within a 2 ppm window. The identity of the constant domain (C-) gene was determined by matching the fragments in these combined spectra to a database of all functional, open reading frame, and in-frame pseudogene alleles for C-genes retrieved from IMGT/Gene-DB (Lefranc et al., 1999). Next, bottom-up LC-MS/MS spectra of the fractionated [24.4] 1 [47359.4] clone were screened against a database of all functional, open reading frame, and in-frame pseudogene alleles for the variable domain (V-) genes retrieved from the IMGT/Gene-DB (Lefranc et al., 1999), using local Smith-Waterman alignment with the BLOSUM62 matrix in which the common *de novo* sequencing errors I/L, Q/E and N/D were modified to neutral substitutions (Smith and Waterman, 1981). From any gene regions with confident peptide matches (FDR < 1%), we then took the FR1, 2 and 3 regions and subjected them to an in-house scoring algorithm to score their agreement with our middle-down data (Figure S5). In short: the algorithm searches for peak patterns that would occur as a result of fragmentation of the provided sequence regions, disregarding preceding and succeeding parts of the initial sequence. Ranking the gene regions by a resulting composite score enabled us to select top-scoring templates as a starting point for our sequencing efforts as well as discard low-scoring regions from further analyses. The remaining gene regions were then used to *in silico* generate a database of germline light and heavy chains.

Then, using a custom implementation of the DirecTag algorithm (Tabb et al., 2008), all possible sequence tags were detected and annotated in the combined middle-down spectra. These sequence tags were used to search the filtered germline light and heavy IgG chains. For the best scoring germline sequences, consistent sequence tags with a length of more than 4 amino acids were searched against *de novo* predicted peptides originating from the bottom-up peptide-centric MS data. In an iterative manner, the matching

peptides were used to modify the best scoring selected germline sequences until the mass of the final sequence matched the precursor masses determined by middle-down MS (Figures S8 and S9). In more detail, the gaps between the consecutive sequence tags extracted from the middle-down MS data were first filled with amino acids from the best matching germline sequence. Then, the filled gaps were compared to the highest scoring peptides retrieved from the bottom-up MS data, aligned to the region of interest using Clustal Omega algorithm. When aligned peptides showed discrepancies from the germline sequence the amino acid residues in the gaps were altered and the theoretical mass of the gap was compared to the experimental mass, defined by the mass difference between consecutive sequence tags. Finally, the modified sequences were rescored by spectral alignment, sequence-tag detection, and a bottom-up database search, providing the final mature Fab sequences. The final predicted sequences – and more specifically the identified mutations when compared to the most closely related gene regions – were additionally compared to the frequency of amino acid occurrence at their specific positions (as numbered by IMGT) in both the AbYsis database (Swindells et al., 2017) and the recombined full IMGT database (Lefranc et al., 1999). This screening yielded an estimate of how likely the mutations were to occur. While some of the predictions are rather rare, none of them are impossible as reported by AbYsis (Table S5).

QUANTIFICATION AND STATISTICAL ANALYSIS

For quantification of LC-MS profiling data, the intensity values of the two mAb standards (trastuzumab and alemtuzumab) were averaged in each run and set to 20 μg/mL. The intensity values of all other detected Fabs were normalized to these values in order to determine the concentration of each individual clone. For the quantification of mAbs in the validation experiment, a slightly different normalization was used. The intensity values of the detected mAbs in all runs were normalized to the intensity values of trastuzumab and alemtuzumab as measured in the first 200 ng replicate.

Statistical values in figures depicted as lower-case letter r indicate Pearson correlation coefficients. Distances between samples as shown in Figure 2B were determined by distance correlation. Linear regression for validation of quantification Figure S1B was determined by ordinary least squares regression with the coefficient of determination given as uncentered R^2. The error-bars in the figure represent the standard error of the mean (SEM).
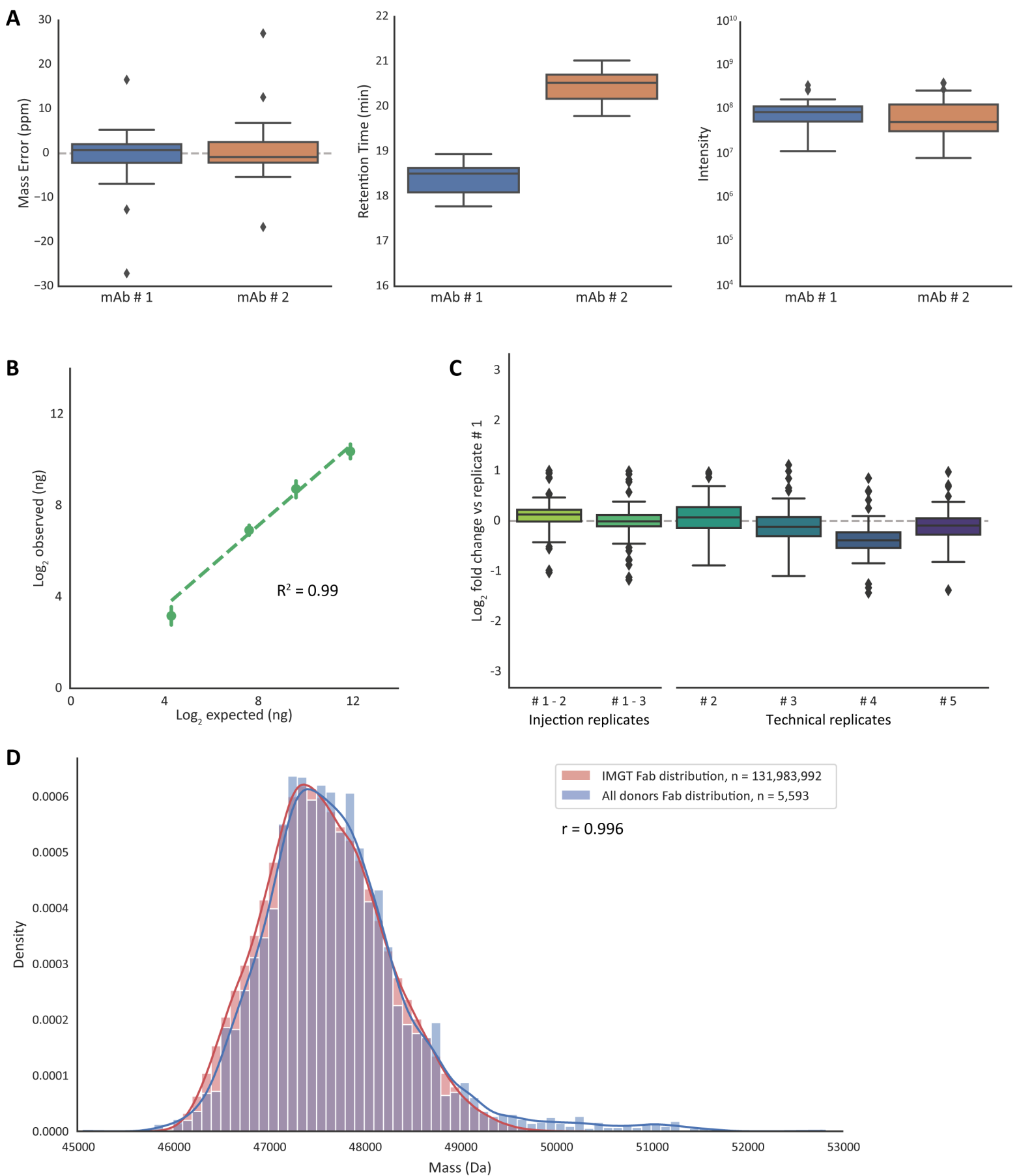
Supplemental information

Human plasma IgG1 repertoires

are simple, unique, and dynamic

Albert Bondt, Max Hoek, Sem Tamara, Bastiaan de Graaf, Weiwei Peng, Douwe Schulte, Danique M.H. van Rijswijck, Maurits A. den Boer, Jean-François Greisch, Meri R.J. Varkila, Joost Snijder, Olaf L. Cremer, Marc J.M. Bonten, and Albert J.R. Heck

Figure S1 | Performance evaluation of plasma Fab profiling approach using various experimental controls. A) Accuracy and precision in mass, retention time and abundance of spiked-in monoclonal antibody controls. The boxplots show aggregated data from the mAb controls over all plasma measurements. The box indicates median and inter quartile ranges (IQRs), and the whiskers span 1.5 times the IQR. Values outside this range (fliers) are marked with diamonds. From left to right, the panels show observed mass error of these mAbs, observed retention time, and detected intensity. **B)** Linearity of detection. For these experiments six monoclonal antibodies (Trastuzumab, Cetuximab, Rituximab, Campath, Bevacizumab and Infliximab) were added at 20, 200, 800 and 4000 ng in a plasma background. The detected response of all of these mAbs was compared to the expected response visualized as scatterplot. The error bars depict the standard error, and the dotted line shows an ordinary least squares (OLS) linear regression accompanied by a R^2. **C)** Reproducibility of quantitation. The reproducibility of the top 100 most intense clones in a plasma were measured over several replicates and visualized as boxplots. The values are shown as fold change of the concentration compared to the first replicate measurement. The first two boxplots depict injection replicates, i.e. replicates from multiple injections of the same sample. The other boxplots show technical replicates, which constitute the entire sample preparation procedure starting from the plasma. The boxes are constructed using the same method as the boxplots in panel **(A)**. **D)** Distributions of detected Fab masses compared to the expected mass distribution. Kernel density estimation of all Fabs detected in all sepsis donors, at all analyzed time points, compared against an *in silico* generated distribution of Fabs from the IMGT database. The number of Fabs used to generate each distribution is shown in the figure legend. Both distribution histograms use a bin size of 100 Da. The Pearson correlation coefficient (r) was calculated between both kernel density estimations.
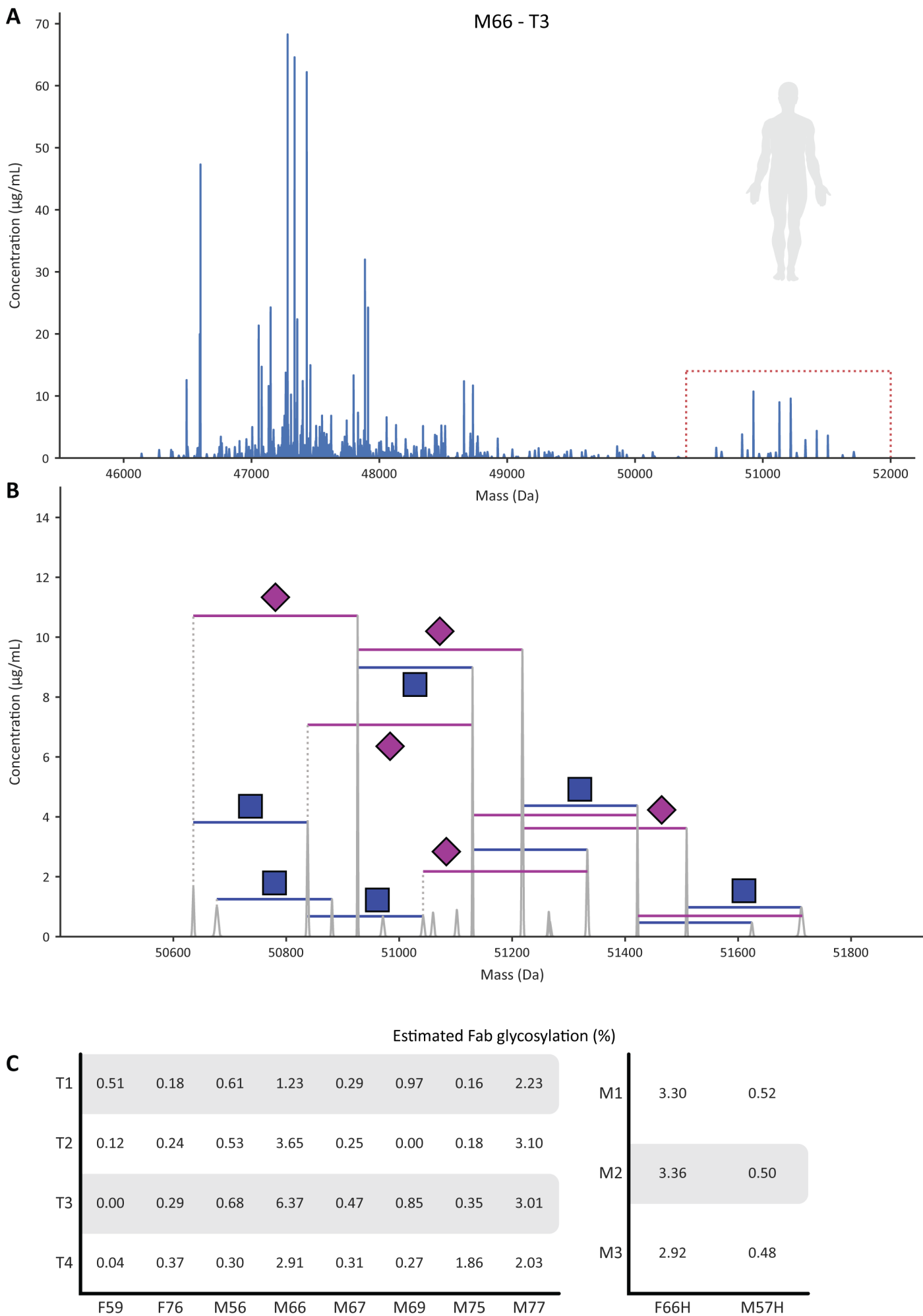
Figure S2 | Extent of Fab glycosylation in the plasma repertoire. A) Fab mass profile of donor M66, taken from the plasma sample at time point 3. The mass range between 50,400 Da and 52,000 Da is boxed in red and shown magnified in panel **(B)**. **B)** Zoomed-in mass profile with annotation of glycan-related masses. Monosaccharides mass differences between peaks are annotated as follows: blue square = GlcNAc (203 Da), magenta diamond = sialic acid (291 Da). For annotation of the glycosylation a mass tolerance of 1 Da and a retention time tolerance of 0.6 min was used. **C)** Estimated percentages of plasma Fab molecules being glycosylated in all samples measured. For this, Fab clones with a mass >49,500 Da were assumed to carry one or more Fab glycans. This value was chosen because the *in silico* Fab distribution generated from the IMGT database (shown in **Figure S1D**) extends up to 49,500 Da, the majority of Fabs has a mass between 47,000 and 48,000, and the average literature described Fab glycan has a mass of approximately 2,300 Da. The validity of this assumption is illustrated for M66 – T3 in panels **(A)** with the glycosylated Fabs being in mass quite separated from the other clones. The percentage of plasma Fab molecules being glycosylated was calculated by taking the sum of Fab concentrations above 49,500 Da and dividing these by the total detected concentration in each sample. On the left in C) are shown the % Fab glycosylation in the plasmas of the septic patients, on the right the % observed in two healthy donors. In general, we observe that the % Fab glycosylation is < 1%, although in some donors it is substantially higher, i.e. M66. M77 and F66H.
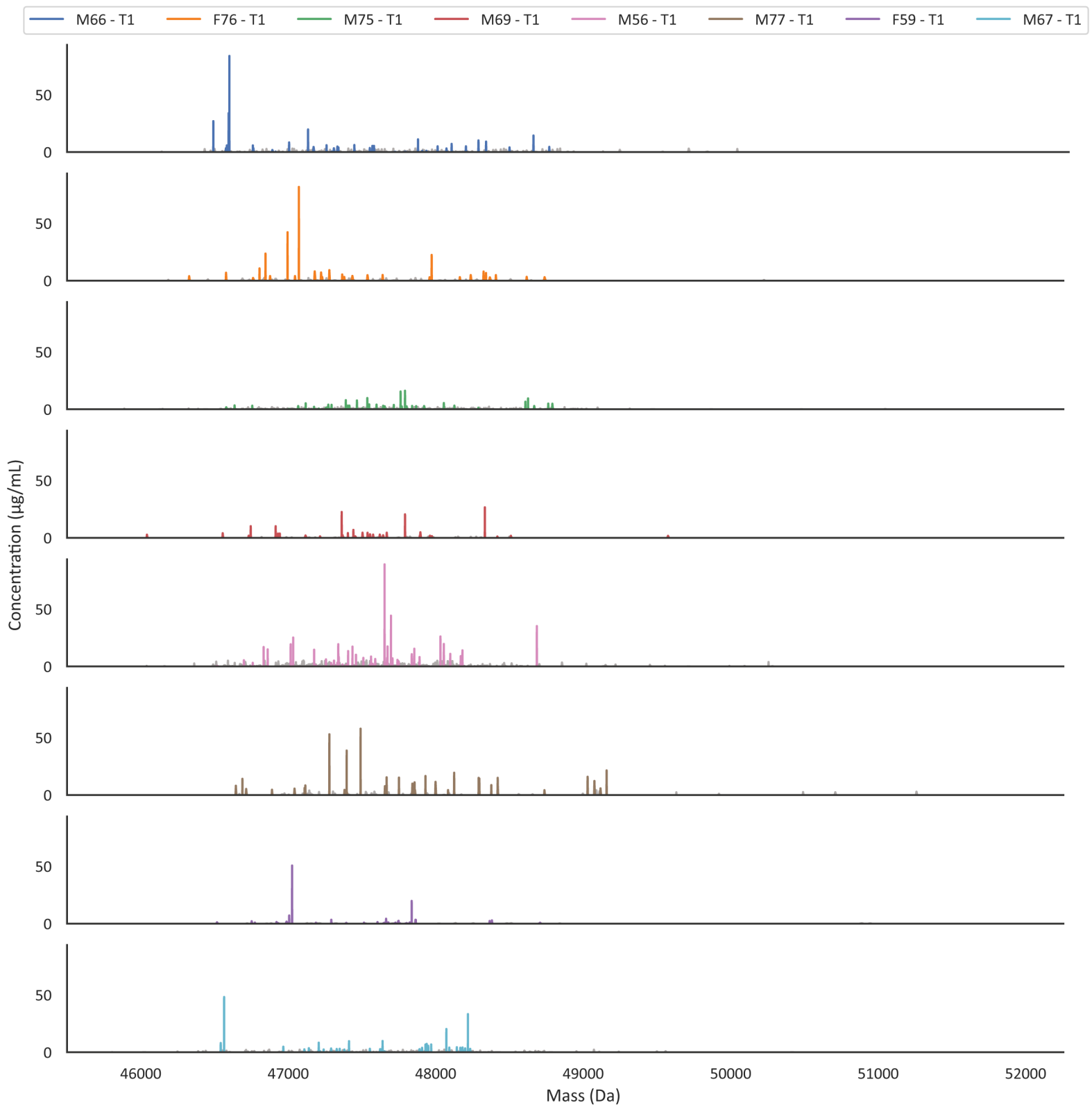
Figure S3 | Fab mass profiles are simple and uniquely individual. The by LC-MS obtained Fab mass profiles are shown for plasma taken from each patient at time point 1 (post-operative). The Fab mass profiles are plotted along the full mass range. In each profile the top 30 most intense clones are colored, with a separate color for each donor. The remaining clones are shown in grey. The concentrations were determined from the LC-MS intensities, normalized against two spiked-in recombinant mAbs.
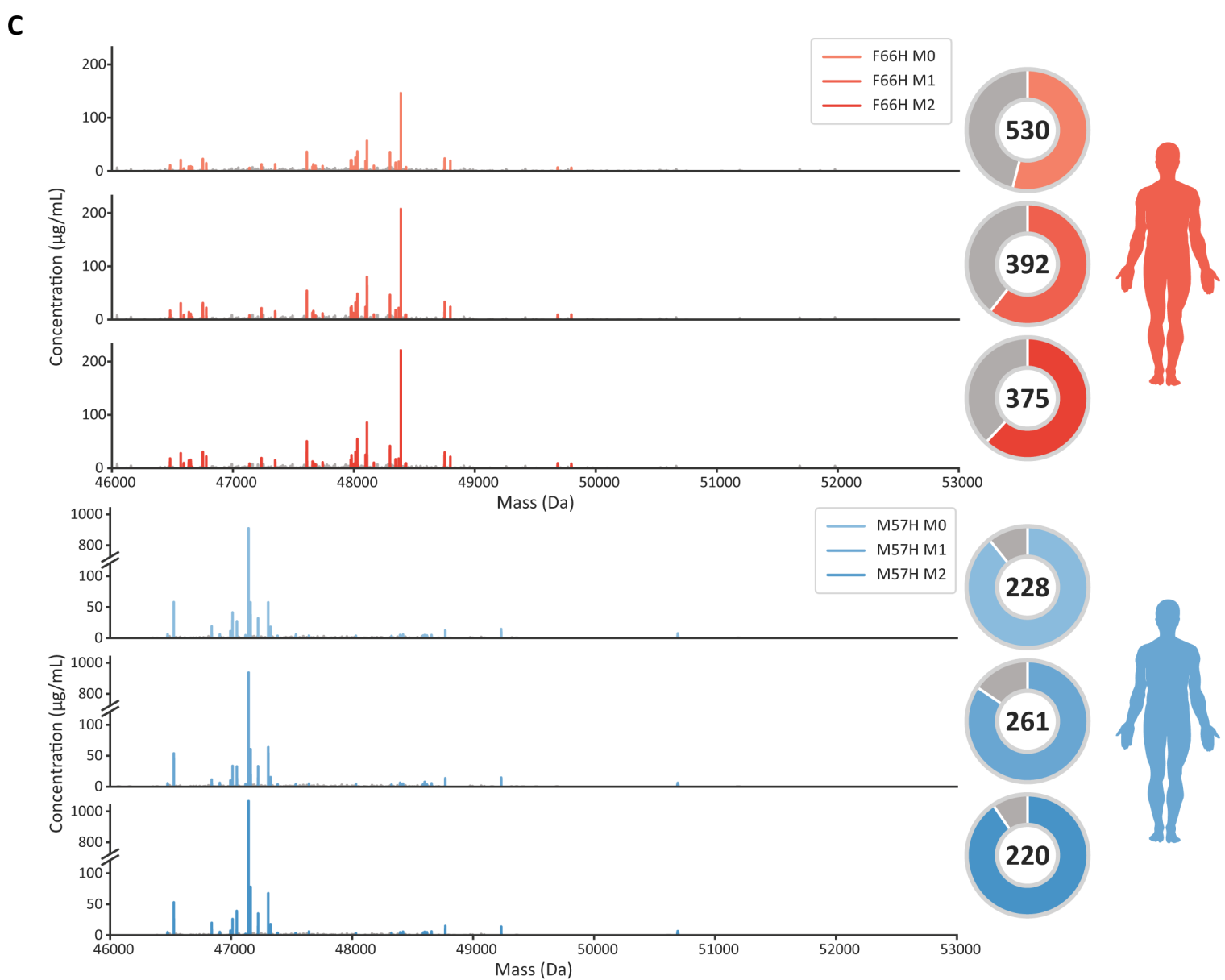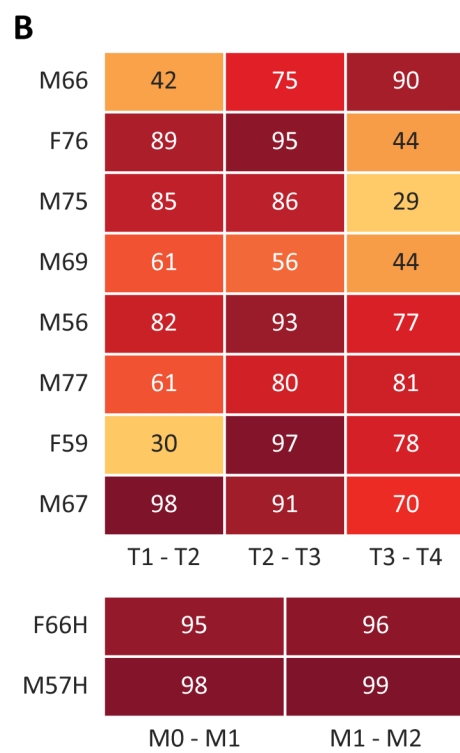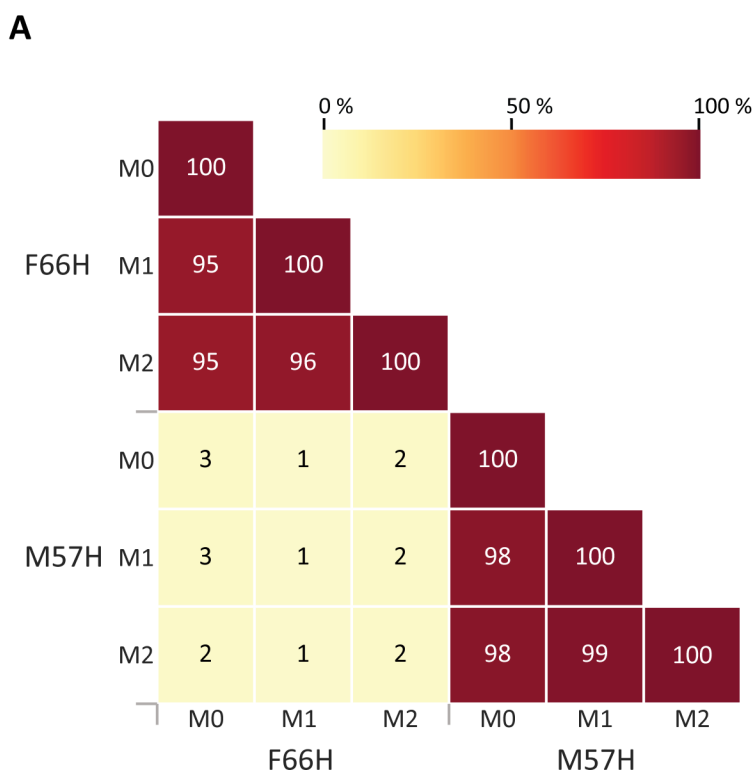
Figure S4 | Longitudinal plasma Fab profiles obtained for two healthy donors. A) Heatmap of healthy donors F66H and M57H constructed using the same method as used in **Figure 2A**. Time points are marked M0, 1, and 2, representing month 0, month 1 and month 2, to clearly distinguish these from the sepsis donor time points. Inside each cell of the heatmap a percentage value shows the degree of overlap between samples, which is also represented by the color bar. **B)** Heatmap showing the Fab overlap in consecutive time points of all healthy and sepsis affected donors, showing only the degree of overlap for consecutive time points within each donor. The colors match those of the color bar from panel **A**. **C)** Mass profiles of healthy donors with donut charts. For each mass profile the top 30 most intense clones are colored, and the remaining clones are colored grey. In the donut charts the colored slice displays the distribution of the top 30 most intense clones compared to the other clones. The value inside the donut shows the total number of detected clones.
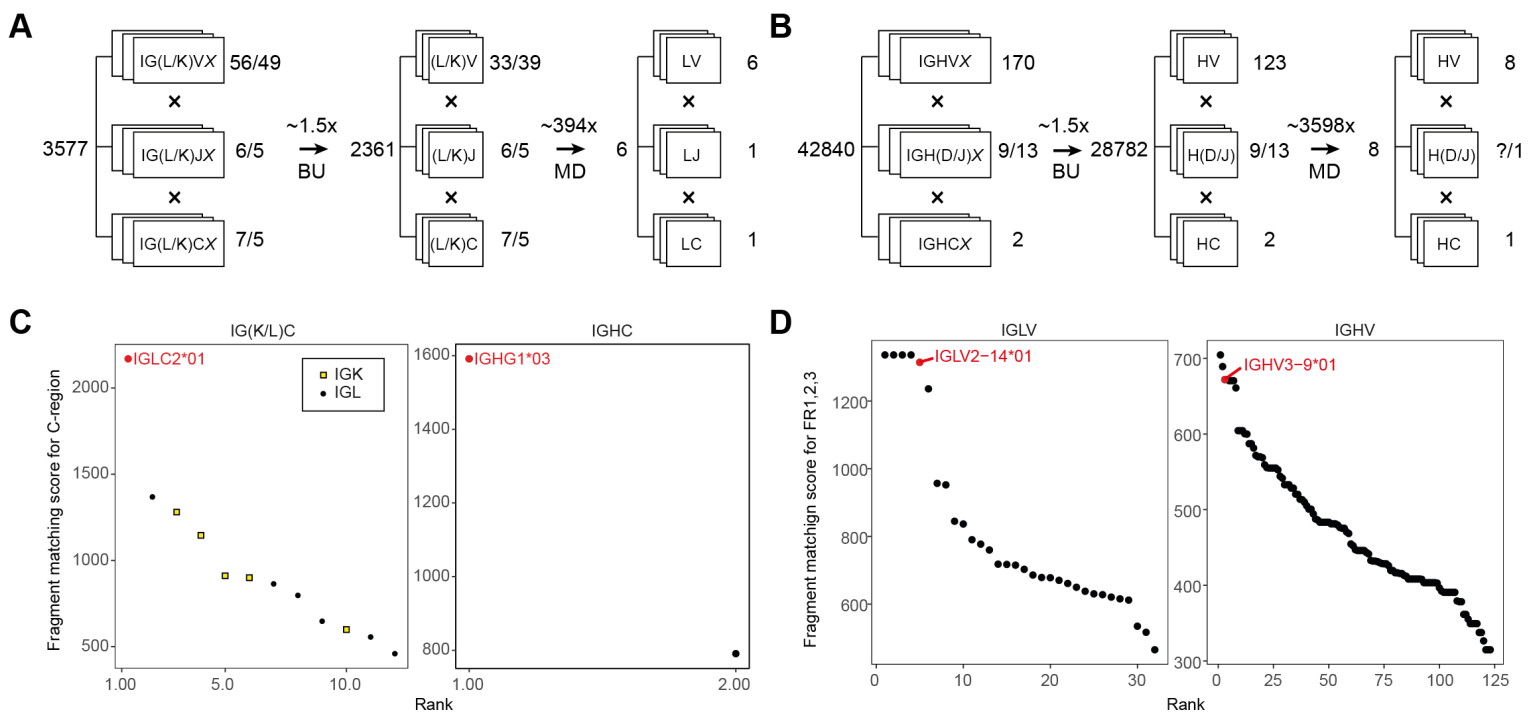
Figure S5 | Template matching of the obtained sequencing data for the Fab clone [24.4] 1 [47359.4] versus the IMGT database.
The filtering of IMGT database and scoring of the germline IGXV and IGXC-alleles was performed by using iteratively bottom-up (BU) and middle-down (MD) proteomics data. **A)** Filtering of germline IGL and IGK alleles with BU and MD mass spectrometry (MS) reduces the number of possible germline light chain sequences from 3,577 to 6 candidate sequences (~600-fold reduction). **B)** Filtering of germline IGH alleles with BU MS and MD MS reduces the number of possible germline heavy chain sequences from 42,840 to 8 candidate sequences (~5,000-fold reduction). **C)** Fragment matching scores for the germline C-gene alleles of the light (left) and heavy (right) chain of the Fab clone [24.4] 1 [47359.4] using the middle-down MS data. **D)** Fragment matching scores for the Framework Regions 1, 2, and 3 of the germline V-gene alleles of light (left) and heavy (right) chains of IgG1 determined by using the middle-down MS data.
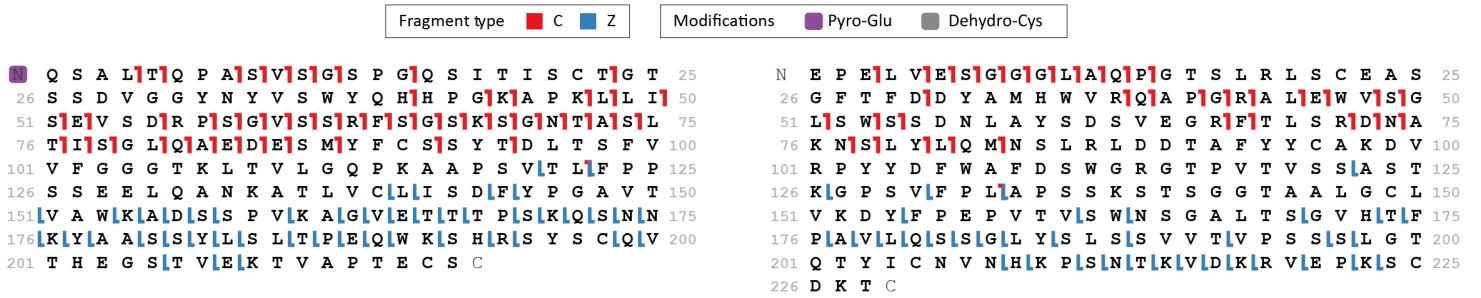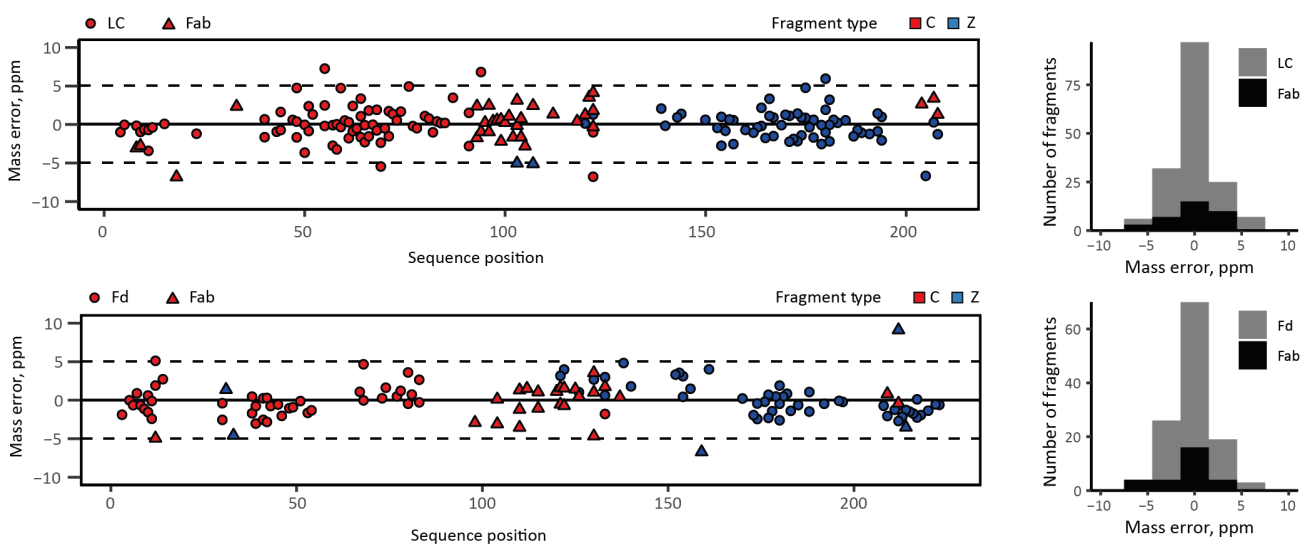
Figure S6 | Middle-down ETD analysis and sequence annotation of the light chain and the N-terminal portion of the heavy chain from clone [24.4]1[47359.4] from donor F59. A) Fragmentation maps of the light chain (left) and Fd (right) when subjected to ETD within the intact Fab molecule. B) Fragmentation maps of the light chain (left) and Fd (right) when subjected to ETD after reduction and denaturation of the precursor Fab. C) Mass errors and their distribution of the light chain fragments observed in ETD of Fab and the light chain alone, and mass errors and distribution thereof for Fd fragments detected in ETD of Fab and Fd alone.
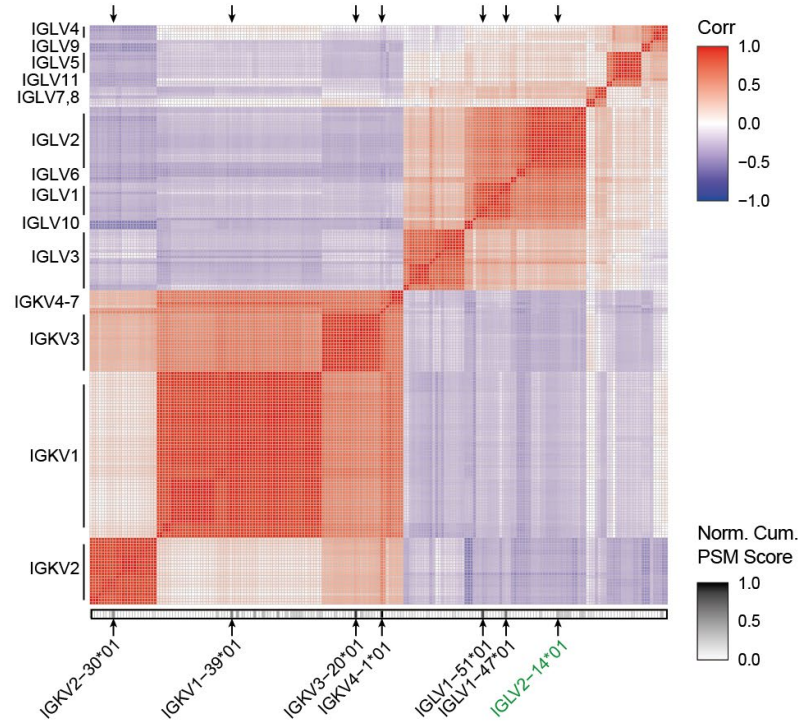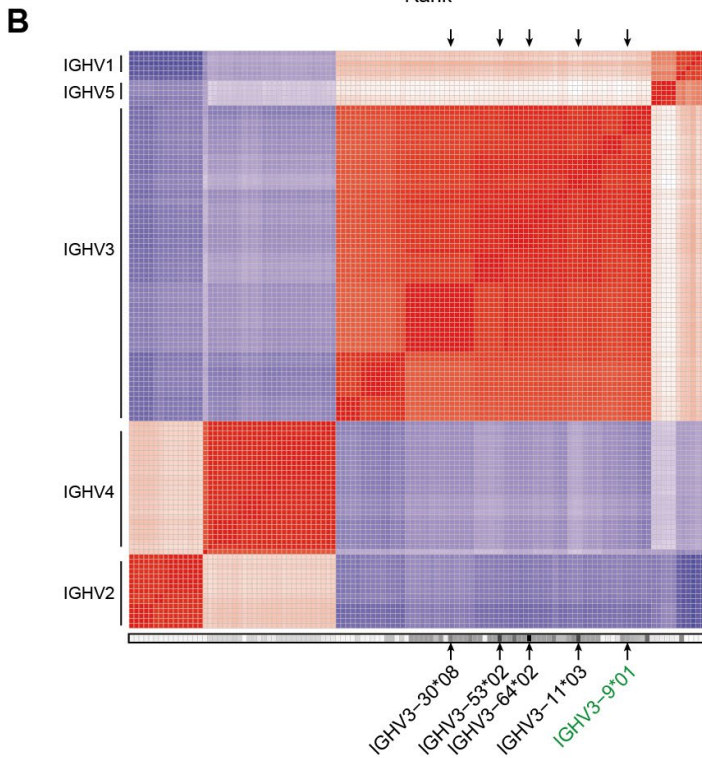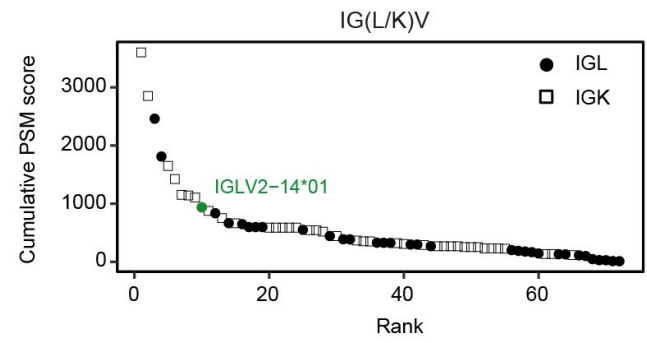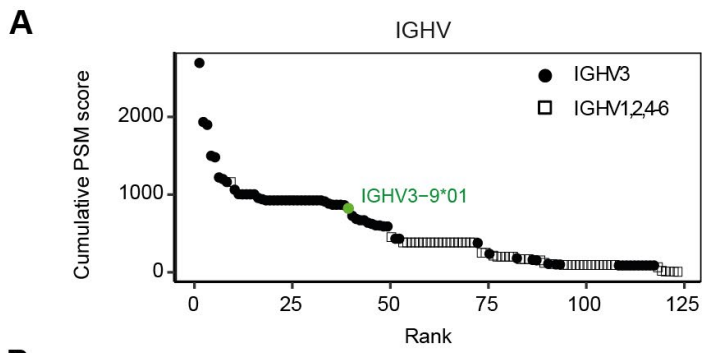
Figure S7 | Large homologous families of Ig V-gene alleles (e.g. IGHV3) are observed among the top-scoring identifications as extracted from the bottom-up proteomics data. **A)** Cumulative PSM scores determined for the germline V-gene alleles of the Fab heavy (left) and light chains (right). On the left, alleles from the largest IGHV3 family are displayed as filled circles; alleles of other IGHV families are shown as empty squares. On the right, alleles from larger and more homologous IGKV families are shown as empty squares, while filled circles display alleles of IGLV families. Germline V-gene sequences were downloaded from IMGT. **B)** Correlation matrix displaying sequence similarity among all germline V-gene sequences of the Fab heavy (left) and light (right) chain. Normalized cumulative PSM scores are shown below the correlation maps. Some of the top-scoring V-gene sequences are indicated with black arrows. The V-genes ultimately determined for clone [24.4] 1 [47359.4] by the integrative *de novo* bottom-up and middle-down sequencing are highlighted in green.
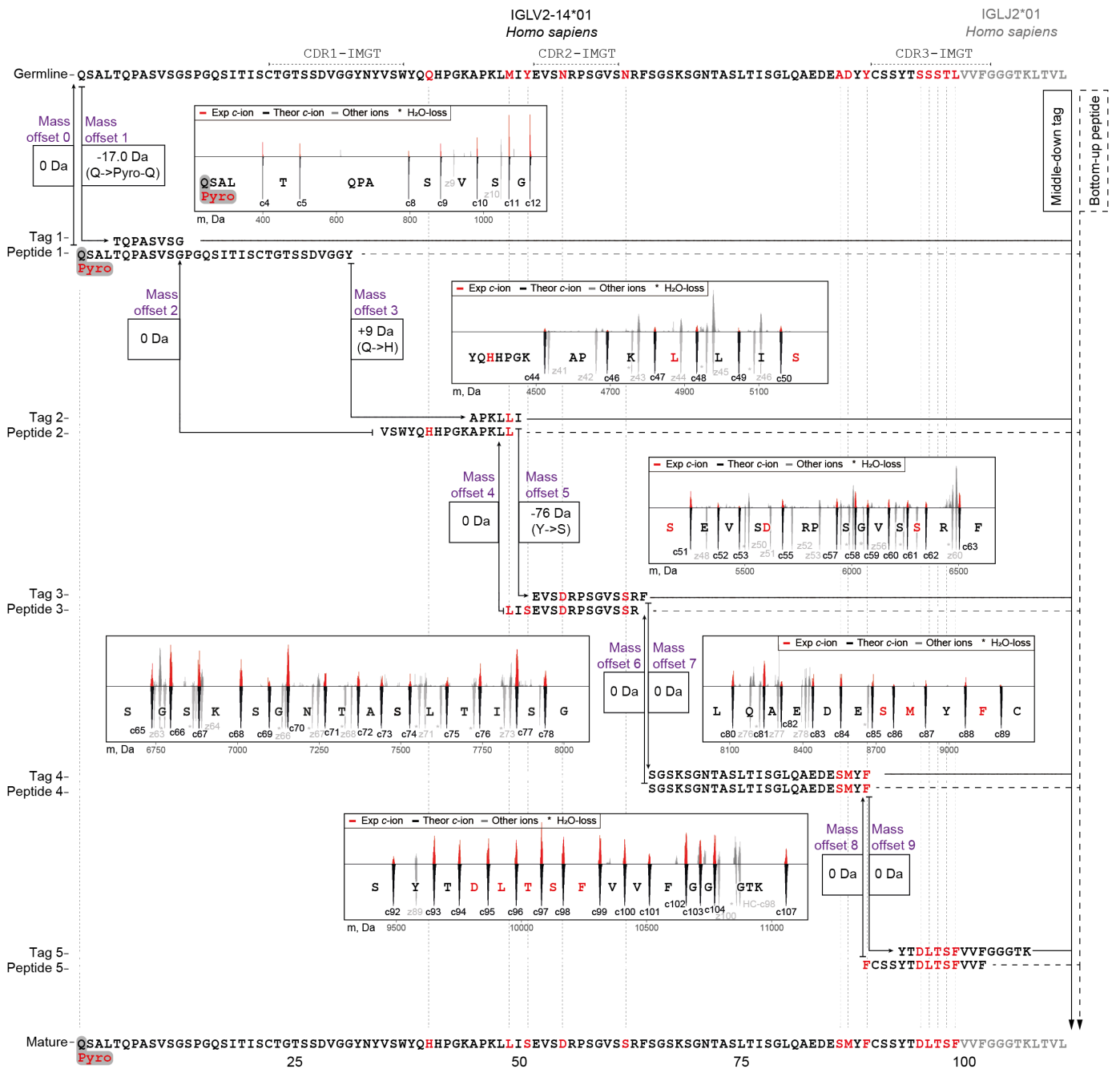
Figure S8 | Refining of the sequence of clone [24.4]1 [47359.4] light chain germline IGLV2-14*01-IGLJ2*01, based on the iterative integration of middle-down and bottom-up proteomics data. First, the sequence tags detected in the middle-down MS data were used as arrays of consecutive fragment peaks, which directly hinted at the presence of 11 mutations (M49L, Y51S, Y51S, N55D, N62S, A85S, D86M, Y88F, S95D, S96L, S97T, T98S, and L99F). Next, these tags were aligned to the *de novo* sequenced peptide sequences obtained by bottom-up MS, revealing 2 additional mutations. The highest-scoring aligned peptides were used to extend the initial sequence tags, and then these steps were iteratively repeated. At each step of tag extension, the mass offsets were calculated by comparing a mass gap between two consecutive tags to the mass of amino acid residues in the corresponding gap in the germline sequence. Iteratively, middle-down tags were extended with aligning peptides until all (if possible) mass offsets become equal to 0 Da. Eventually, 13 mutations and one modified residue (Pyro-Q) were determined for the [24.4]1 [47359.4] light chain sequence. De-charged isotopic distributions of the fragments involved in each sequence tag are displayed as red peaks in the corresponding insets with the theoretical isotopic distributions for these fragments displayed underneath each fragment. Fragmentation spectra of the peptides used in this refining process for the CDRs are shown in **Figure 5**. See also **Supplemental Table 5** for an overview of the evidence supporting each detected amino acid mutation.
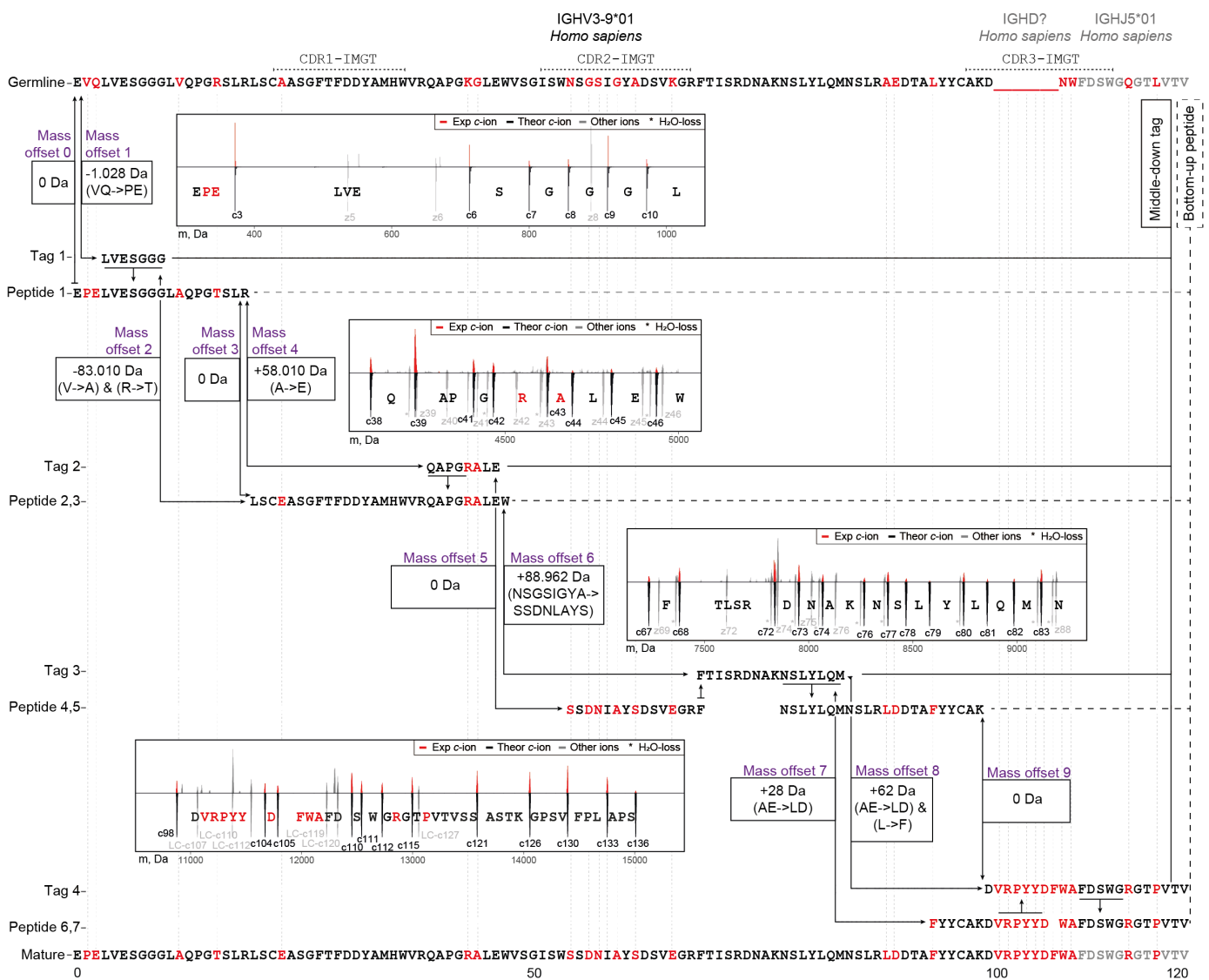
Figure S9 | Refining of the sequence of clone [24.4]1 [47359.4] heavy chain germline IGHV3-9*01-IGHJ5*01, based on the iterative integration of middle-down and bottom-up proteomics data. First, sequence tags were detected in the middle-down MS data as arrays of consecutive fragment peaks similar to refining of the light chain sequence. Next, these tags were aligned to the *de novo* sequenced peptides from bottom-up MS. The highest-scoring aligned peptides were used to extend the initial tags, and then this step was repeated. At each step of tag extension, the mass offsets were calculated by comparing a mass gap between two consecutive tags to the mass of amino acid residues in the corresponding gap in the germline sequence. Iteratively, tags were extended with aligning peptides until all (if possible) mass offsets become equal to 0 Da. Eventually, more than 20 mutations were determined for the N-terminal portion of the heavy chain for clone [24.4]1 [47359.4]. De-charged isotopic distributions of the fragments involved in each sequence tag are displayed as red peaks in the corresponding insets with the theoretical isotopic distributions for these fragments displayed underneath each fragment. Fragmentation spectra of the peptides used in this refining process for the CDRs are shown in **Figure 5**.
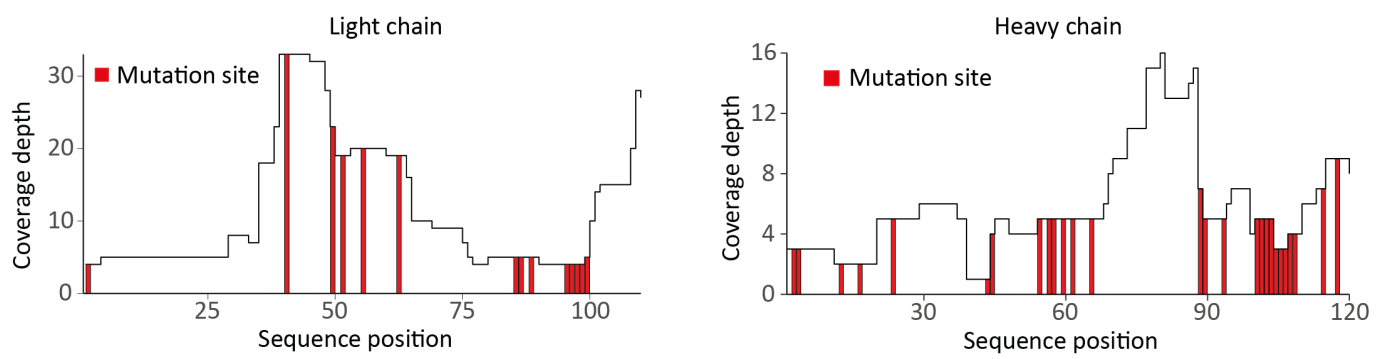
Figure S10 | Coverage depths for *de novo* sequenced light and heavy chains of the clone [24.4]1[47359.4] from donor F59. Values at each position represent the number of unique peptides identified in the bottom-up LC-MS/MS data. The determined mutation sites are depicted in red. Only the first 110 and 120 amino acids are shown for the light and heavy chain, respectively.