Cell Systems, Volume *12*

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

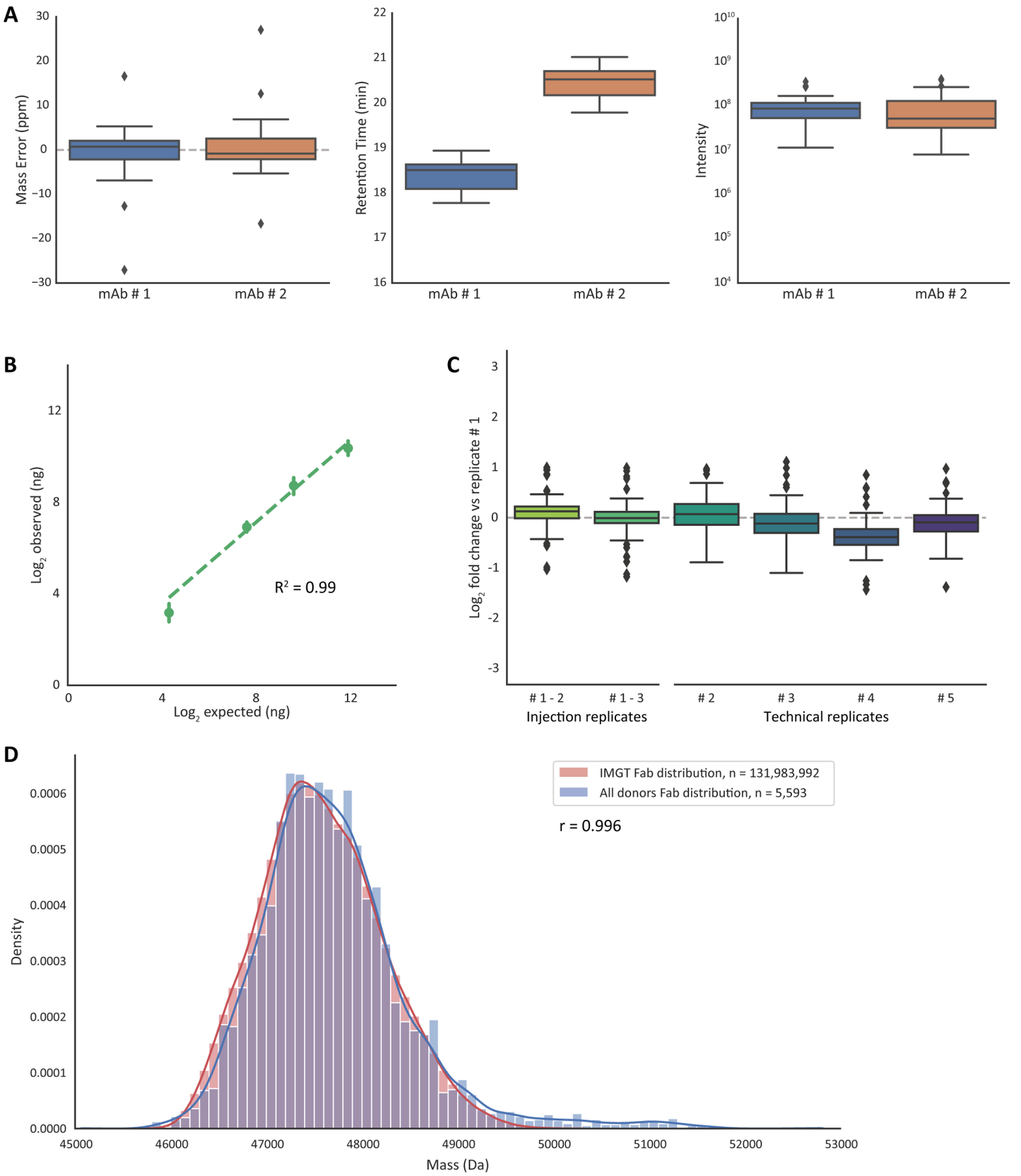
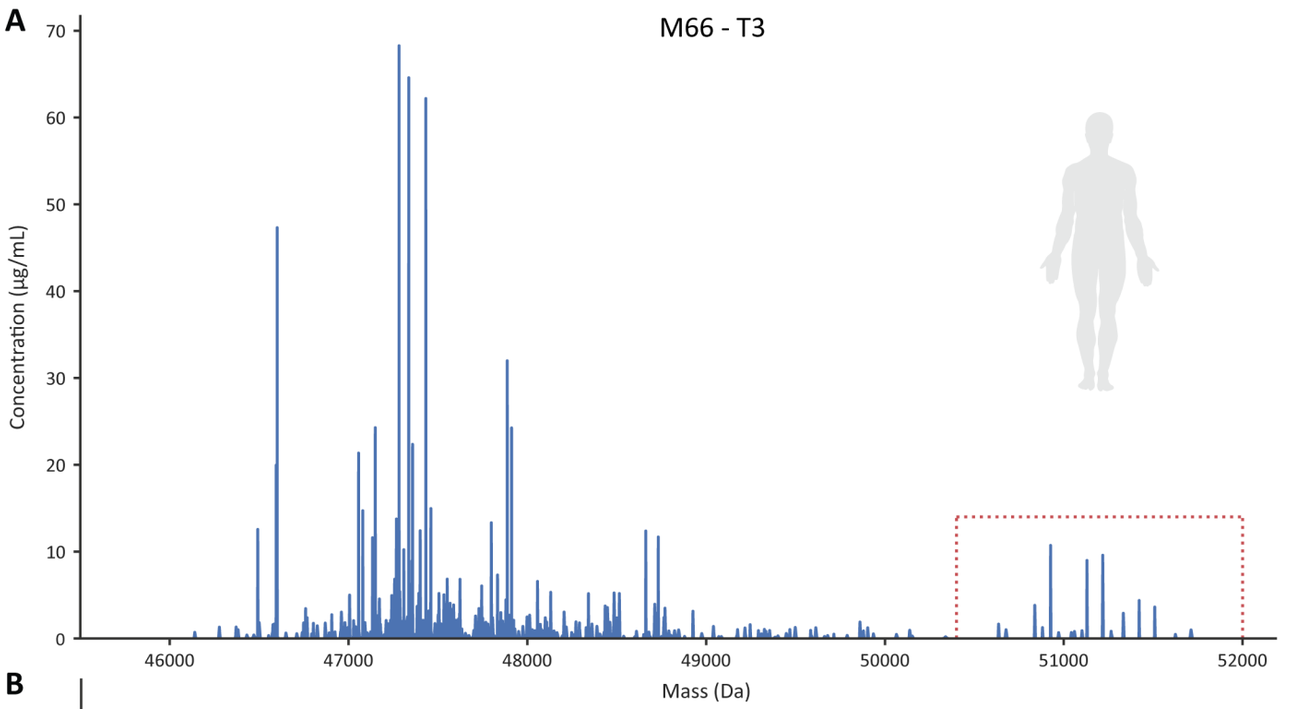


Figure S1 | Performance evaluation of plasma Fab profiling approach using various experimental controls. A) Accuracy and precision in mass, retention time and abundance of spiked-in monoclonal antibody controls. The boxplots show aggregated data from the mAb controls over all plasma measurements. The 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.



C

Estimated Fab glycosylation (%)

	F59	F76	M56	M66	M67	M69	M75	M77
T1	0.51	0.18	0.61	1.23	0.29	0.97	0.16	2.23
T2	0.12	0.24	0.53	3.65	0.25	0.00	0.18	3.10
T3	0.00	0.29	0.68	6.37	0.47	0.85	0.35	3.01
T4	0.04	0.37	0.30	2.91	0.31	0.27	1.86	2.03

	F66H	M57H
M1	3.30	0.52
M2	3.36	0.50
M3	2.92	0.48

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.

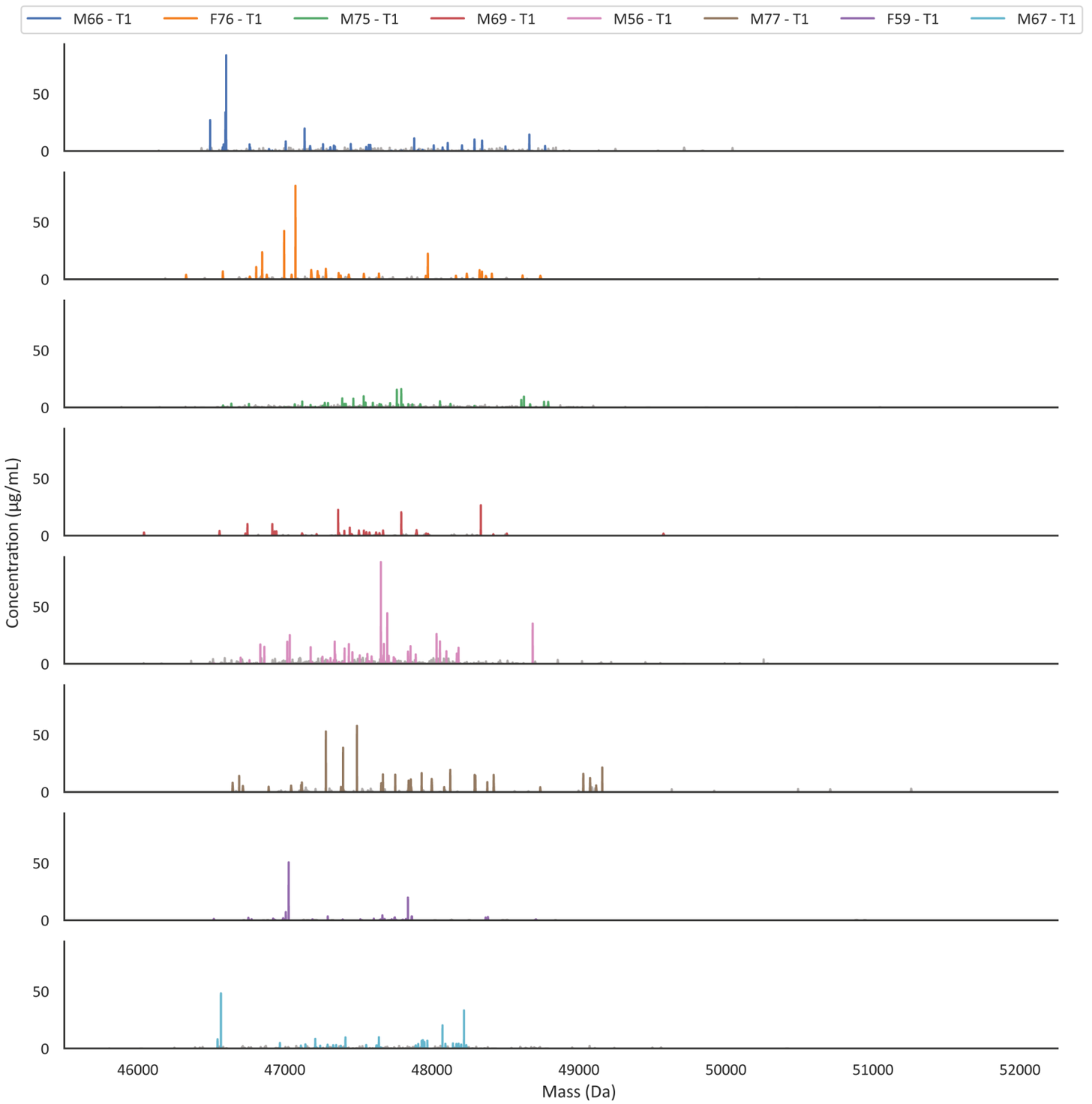


Figure S3 | Fab mass profiles are simple and uniquely individual. The by LC-MS obtained Fab mass profiles are shown for plasma taken from each patient at time point 1 (post-operative). The Fab mass profiles are plotted along the full mass range. In each profile the top 30 most intense clones are colored, 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.

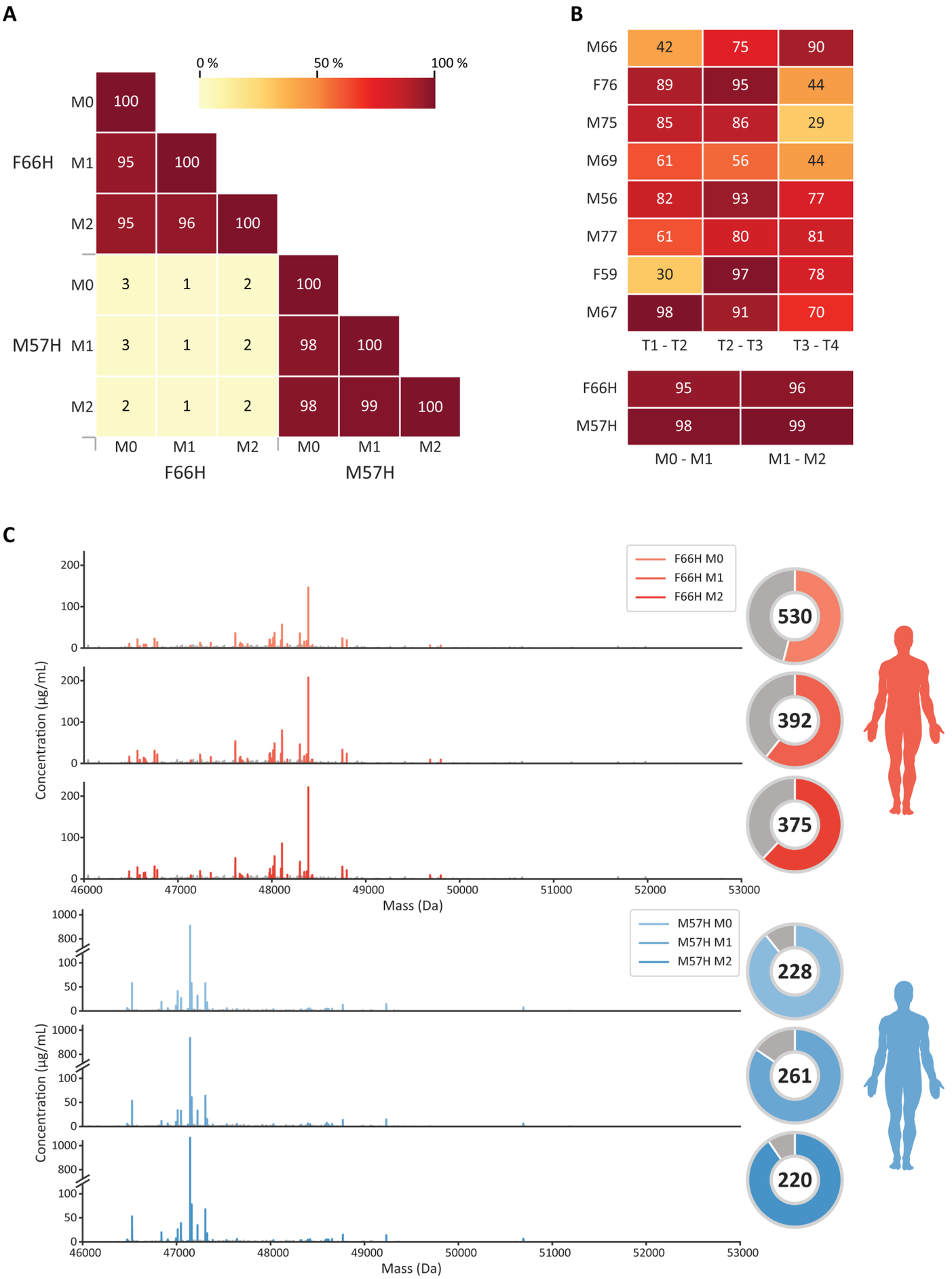


Figure S4 | Longitudinal plasma Fab profiles obtained for two healthy donors. A) Heatmap of healthy donors F66H and M57H constructed using the same method as used in **Figure 2A**. Time points are marked M0, 1, and 2, representing month 0, month 1 and month 2, to clearly distinguish these from the sepsis 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.

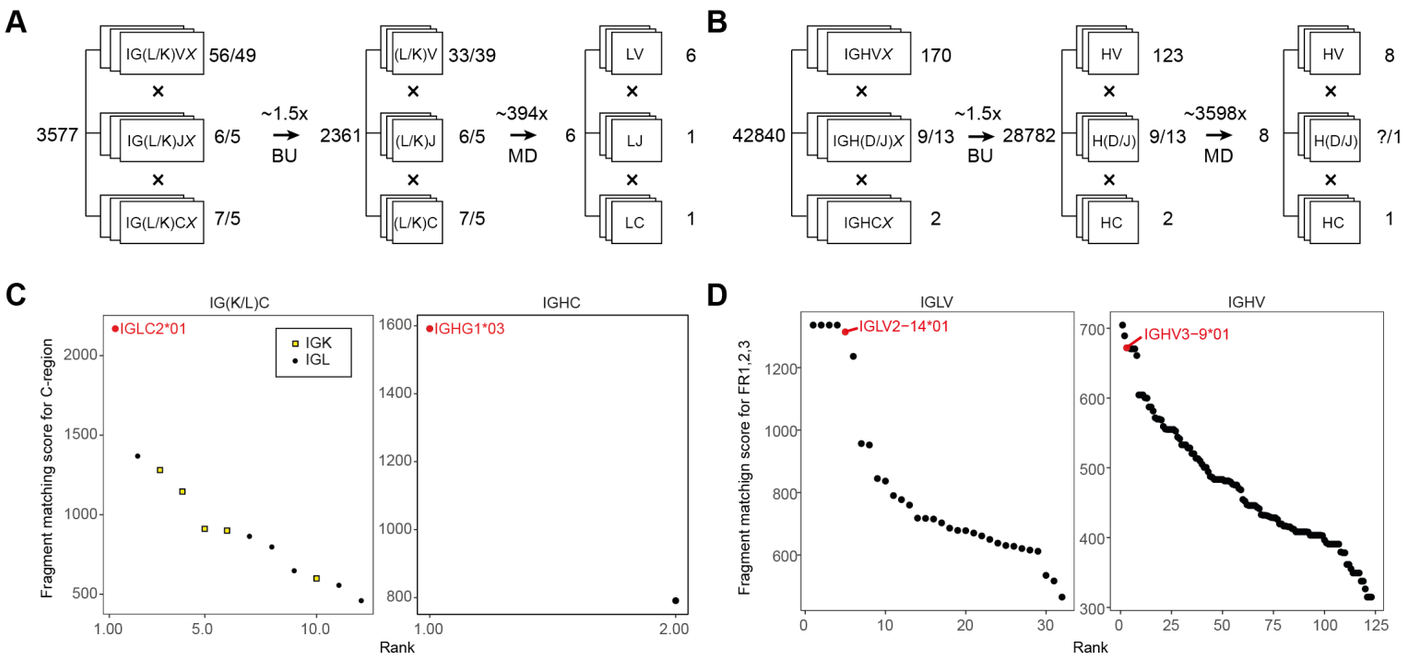


Figure S5 | Template matching of the obtained sequencing data for the Fab clone $^{24.4}1_{47359.4}$ versus the IMGT database. The filtering of IMGT database and scoring of the germline IGXV and IGXC-alleles was performed by using iteratively bottom-up (BU) and middle-down (MD) proteomics data. **A)** 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.

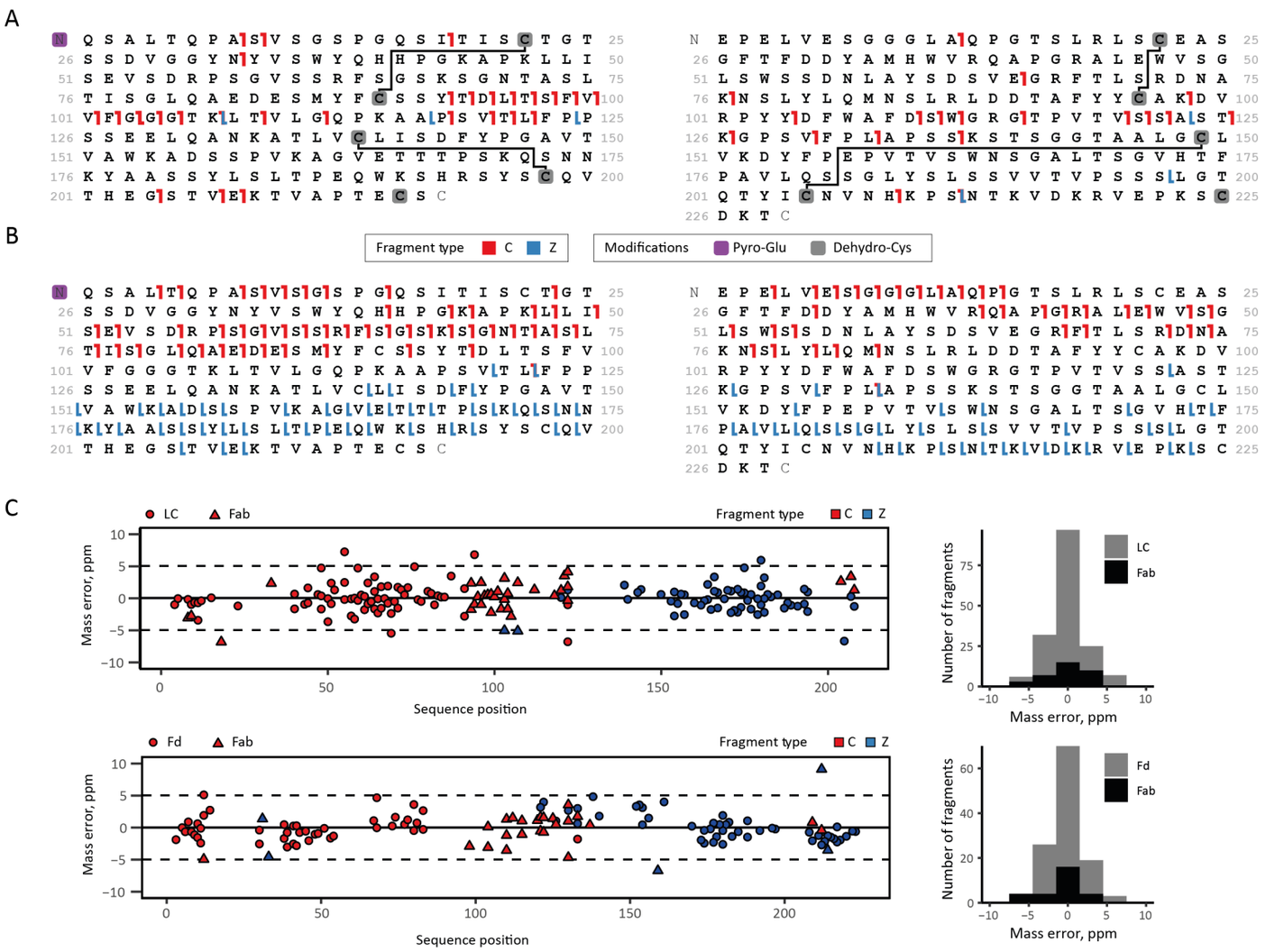


Figure S6 | Middle-down ETD analysis and sequence annotation of the light chain and the N-terminal portion of the heavy chain from clone $^{24.4}1_{47359.4}$ from donor F59. **A)** Fragmentation maps of the light chain (left) and Fd (right) when subjected to ETD within the intact Fab molecule. **B)** 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.

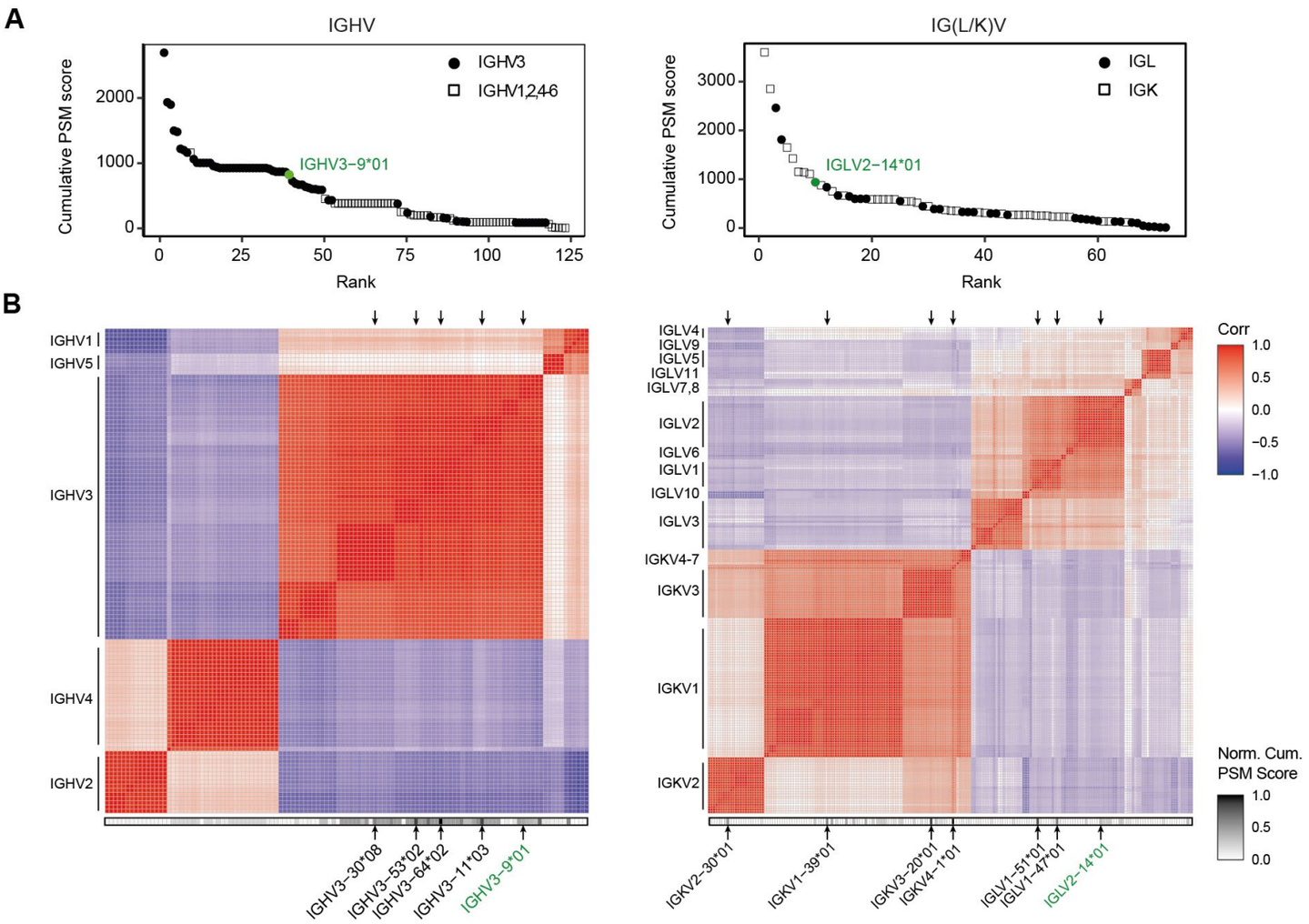


Figure S7 | Large homologous families of Ig V-gene alleles (e.g. IGHV3) are observed among the top-scoring identifications as extracted from the bottom-up proteomics data. **A)** Cumulative PSM scores determined for the germline V-gene alleles of the Fab heavy (left) and light chains (right). On the 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.

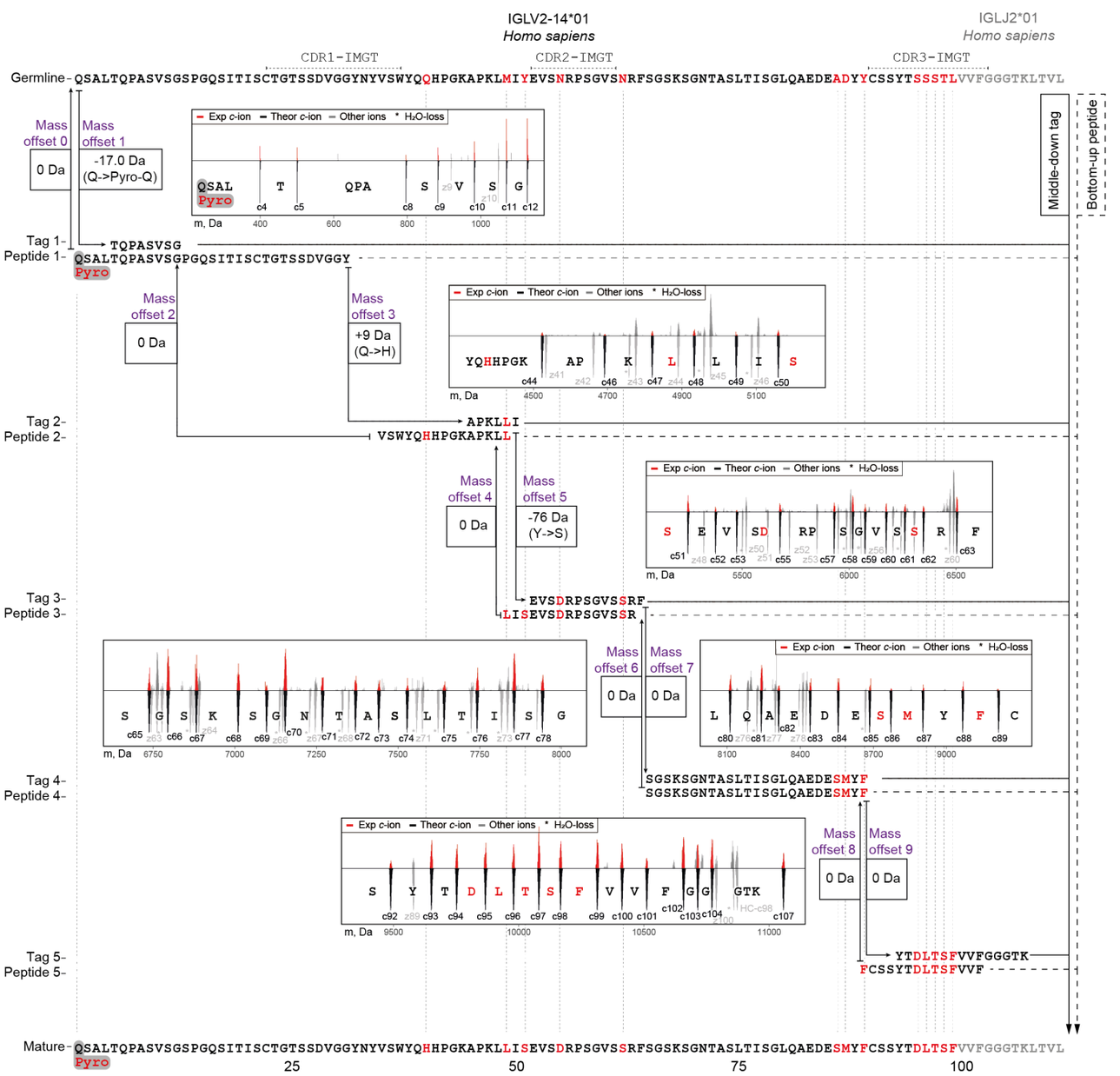


Figure S8 | Refining of the sequence of clone $^{24.4}1_{47359.4}$ light chain germline IGLV2-14*01-IGLJ2*01, based on the iterative integration of middle-down and bottom-up proteomics data. First, the sequence tags detected in the middle-down MS data were used as arrays of consecutive fragment peaks, 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.

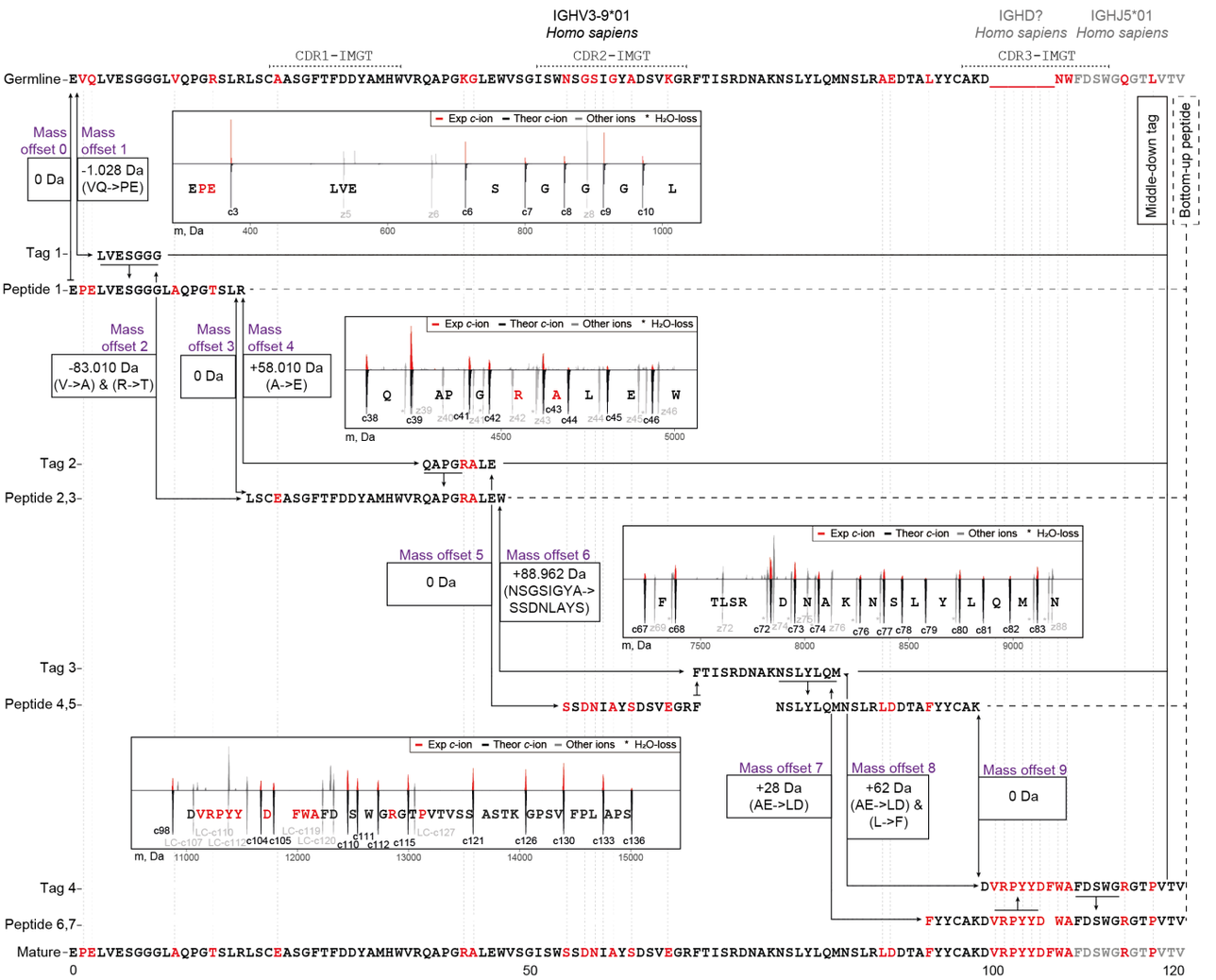


Figure S9 | Refining of the sequence of clone $^{24.4}1_{47359.4}$ heavy chain germline IGHV3-9*01-IGHJ5*01, based on the iterative integration of middle-down and bottom-up proteomics data. First, sequence tags were detected in the middle-down MS data as arrays of consecutive fragment peaks similar to 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**.

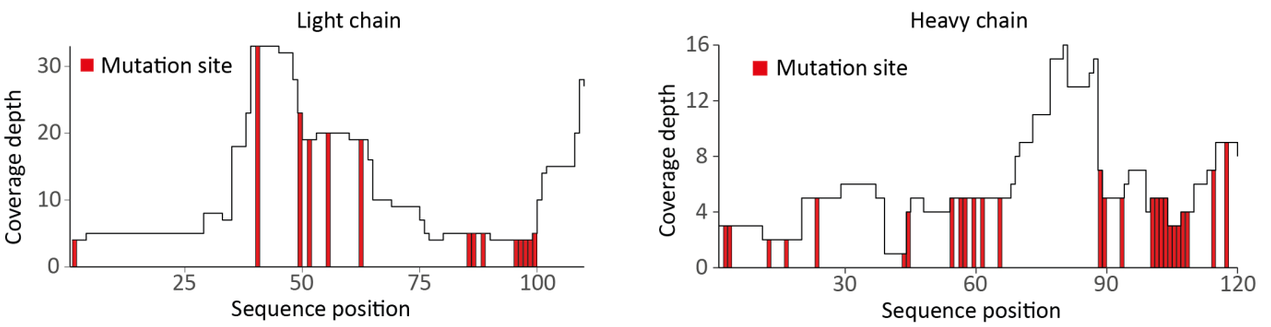


Figure S10 | Coverage depths for *de novo* sequenced light and heavy chains of the clone $^{24.4}1_{47359.4}$ from donor F59. Values at each position represent the number of unique peptides identified in the bottom-up LC-MS/MS data. The determined mutation sites are depicted in red. Only the first 110 and 120 amino acids are shown for the light and heavy chain, respectively.