Cell Systems

Transparent Peer Review Record

# 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, Albert J.R. Heck

*These authors contributed equally*

## Summary

Initial Submission: Received Nov. 25, 2020
Preprint: https://dx.doi.org/10.2139/ssrn.3749694
Scientific editor: Quincey Justman, Ph.D.

First round of review: Number of reviewers: Three
*Three confidential, zero signed*
Revision invited Feb. 7. 2021
*Major changes anticipated*
Revision received Jun. 29, 2021

Second round of review: Number of reviewers: Two
*Two original, zero new*
*Two confidential, zero signed*
Accepted Aug. 19, 2021



## Editorial decision letter with reviewers' comments, first round of review

Dear Albert,

I'm enclosing the comments that reviewers made on your paper, which I hope you will find useful and constructive. As you'll see, they express interest in the study, but they also have a number of criticisms and suggestions. Based on these comments, it seems premature to proceed with the paper in its current form; however, if it's possible to address the concerns raised with additional experiments and/or analysis, we'd be interested in considering a revised version of the manuscript.

To help guide your revision, I've highlighted portions of the reviews that strike me as particularly critical and made some notes inline. More generally, I appreciate that earlier versions of this paper were shorter-format and encourage you to "hydrate" your text and figures, explaining what you've done in a fulsome way. Don't worry about manuscript length limits for now. Likewise, please keep in mind that you have up to 7 full-page, multi-panel figures at your disposal and that an interested, trained scientist should be able to analyze your findings independently, based on the figures presented. Note that at Cell Systems, we believe that the figures are the scientific backbone of the paper. Currently, it's not possible to understand key aspects of your approach from your figures.

I'd also like to be explicitly clear about an almost philosophical stance that we take at Cell Systems. We believe that understanding how approaches fail is fundamentally interesting: it provides critical insight into understanding how they work. We also believe that all approaches do fail and that it's unreasonable, even misleading, to expect otherwise. Accordingly, when papers are transparent and forthright about the limitations and crucial contingencies of their approaches, we consider that to be a great strength, not a weakness. Please keep this in mind when addressing the reviewers' concerns.

As you work on your revision, it's important that you and I stay on the same page. I'm always happy to talk, either over email or by phone, if you'd like feedback about whether your efforts are moving the manuscript in a productive direction. Do note that we generally consider papers through only one major round of revision, so the revised manuscript would be either accepted or rejected based on the next round of comments we receive from the reviewers. If you have any questions or concerns, please let me know. More technical information and advice about resubmission can be found below my signature. Please read it carefully, as it can save substantial time and effort later.

I look forward to seeing your revised manuscript.

All the best,
Quincey

Quincey Justman, Ph.D.
Editor-in-Chief, Cell Systems

**Reviewers' comments:**

Reviewer #1: In this very interesting and important but also somewhat disorganized paper by Heck and coworkers they present methods for analyzing the overall composition of the antibody repertoire in

plasma at low resolution and separately de novo sequencing of an abundant IgG identified in plasma. A key biological finding from this work is that the serum repertoire appears to be dominated by a few IgG clones, as measured from the MS1 of Fab fragments whereby as few a 30 Fabs with distinct average masses seem to account t for >70% of the Serum IgG. The authors show that the dominant IgG clones are different among individuals but for any particular individual they show significant resistance longitudinally although with the relative amounts of different peaks changing over time. In the second part of this paper the authors succeeded in partial de novo sequencing of an abundant plasma antibody, a significant feat. While the paper is certainly informative and appropriate for publication, it will first require major revision to address the following:

Major:
1) The idea of profiling plasma IgG by first producing Fabs by IgdE digestion and then analyzing and quantifying IgG species from the MS spectra after LC/MS is simple and could be very important. Having said that, there is a lack of detail and controls that are concerning:
2) The authors state that the focused on analysis of Fabs because they lack the Fc glycan which contributes to mass heterogeneity. However in humans 25-30% of Fabs are in fact glycosylated. I am not sure why this is not addressed. I suspect the authors must have observed some clones with a broader mass distribution suggestive of glycosylation heterogeneity.
3) Data showing detection and importantly relative quantification of mock samples comprising of say 20-30 Mabs of known sequence and MW and mixed at different concentrations should be shown to allow the reader to evaluate the sensitivity and quantitation in the method.
4) How was the absolute quantitation of the Fabs validated? I presume quantitation was based on peak area but how exactly was Was it with just the two spiked Man standards. More details are needed.
5) It is quite conceivable that each mass peak is in fact comprised of multiple antibodies having very similar MWs. In fact this is not uncommon, and probably should be expected that each "average mass peak" includes multiple mabs, for example having the same gremlin usage and similar length CDR3s giving comparable masses. This needs to be addressed in detail.
6) Reproducibility among technical replicates should be shown. Also for biological replicates is starting with the same plasma and performing separate digestions and LC/MS then evaluating variability.
7) While impressive the de novo sequencing opens several questions: Was the approach validated e.g. by using middle down and bottom analysis of a known Fab?
8) tHe strategy for filling the gaps in the VH is not presented clearly enough in the main text and SI Fig 7-9 are not quite clear either. How did the authors distinguish Leu from Ile?
9) Validation of the de novo sequencing by synthesizing the Mab and then comparing the middle down and top down spectra with those from the plasma Fab would be helpful -but not required,

Minor but still important
1) The two parts of the paper are disjointed. Given that Fig 1-3 address a very different question than Fig 4 and that extensive additional data is required to firm up the first part of the manuscript maybe the authors should consider splitting it into two parts.
2) The writing needs to be improved. In some parts it reads like a graduate thesis. Detailed information on important controls and quantitation should be provided.
3) Some rationale as to why the authors elected to study samples from septic patients should be provided.

Reviewer #2: The authors demonstrate antibody tracking and profiling of patient sera using a combination of mass spectrometry technologies. The authors present a clear and meticulously described methodology, which holds great promise for the further integration of MS based analyses into a therapeutic setting. The most exciting piece of the study is the unique sample set the authors were able to obtain.

Although the goals listed in the Summary are phrased broadly, the stated intentions were largely met. My outstanding question after reading the paper is whether the detected, unique antibodies are indeed indicative of sepsis or recovery, and furthermore whether a healthy individual's antibody repertoire would behave similarly. I believe it is crucial to add a longitudinal analysis from a healthy individual to give weight to the idea that the similarities and differences found over time in these patient are meaningful. ***[From QJ: If adding at least on longitudinal analysis from a healthy individual is not possible, please let me know.]*** Without antigen-specific enrichment of the repertoire it is possible the methodology is simply tracking changes in the most abundant circulating antibodies. It may be possible that a healthy individual could deliver a similar result to the patients reported herein in that: a) each individual is different, and b) levels of specific Abs change over time. Were any assays performed on the fractionated patient Ab that was sequenced to show antibody relevance?

"…seemingly sepsis-responsive, single IgG1 clone…" is a quote from page 4. The authors never come back to this nor does there seem to be any evidence to uphold the claim. This statement could be misleading to readers as there is not any demonstrated correlation between the detected Abs and the patient condition.

Was there a rationale for choosing plasma instead of serum for analysis?

Fig S4: It appears there is enhanced CDR-L3 sequence coverage from ETD in the Fab as compared to the reduced LC (also looks true for CDR-H3 in Fab versus Fd). Does this trend hold true with ETD of other Fabs? This would be an interesting observation.
The authors note that a combination of middle down and bottom up data were used to obtain 100% sequence coverage, and although the process of how middle down and bottom up data were combined, it appears that middle down search results were used as a template for bottom up identified peptides to fill in. What benefit did ETD provide? Meaning, if one were to send the bottom up data to PEAKS for de novo assembly, would the correct sequence have been called? How (besides match to intact MW) was the sequence validated?

Finally, it seems the real 'meat' of the manuscript is held within Supplemental. There is a great deal of information represented in the supplemental figures. I suggest to move a representative ETD spectrum from Fig S4 to main since the de novo sequencing aspect of the study is deeply discussed, perhaps a figure to accompany would be in order.

Reviewer #3: Bondt et al provided a manuscript on "Human plasma IgG1 repertoires are simple, unique, and dynamic" in which they describe a mass spectrometry based approach to study the immunoglobulin type IgG1 in human plasma. They sequence the immuno-enriched immunoglobulins on the level of peptides and larger protein fragments and interpret the obtained data to reveal (1) unique IgG1 fingerprints in the tested subjects, (2) changes of these fingerprints in relation to sepsis. While the

strength of this work is clearly grounded in the mass spectrometric expertise of the group, there is lack of evidence for the critical aspects and claims.

(1) From my perspective, a systems biology study of this kind should include other data types to corroborate the observations. ***[From QJ: I appreciate this reveiwers' point of view but respectfully disagree, given the sophistication of the proteomics and the interesting clinical cohort. If you have concerns about how to address this point, please let me know.]*** This can include the integration of RNAseq data from B cells that disentangle the clonal diversity as well as studies of "antibodies in action" by methods capturing the binding pattern. The latter would allow to investigate if any common antigens are being targeted by the immune system. Given the hypothesis that the authors' detect the more abundant IgGs in their study, it should be conceivable to detect these also on protein microarrays and enrich these on their respective antigens. Such analyses could then provide further targeted insights in the clonal aspects and specificity, which makes such a study more attractive to a wider audience.

(2) The data is built on a multi-step workflow, and even though I acknowledge the efforts of establishing such a pipeline, I miss key figures about precision, accuracy and reproducibility. Without such measures it remains difficult to judge the results. Here, true biological replicates are needed that process samples through the whole pipeline. For a layperson, the obvious differences in the attached chromatograms provoke further uncertainty about the reproducibility, even though the matching of the peptides in the main figures tried to contradict such suspicions. I do acknowledge the authors' transparency by providing such figures, but bringing the work beyond a proteomics community would require to explain these.

(3) My main concern though lies in the effect of the enrichment of IgGs from plasma. There is no data on the completeness of the enrichment. More importantly, there is no information to rule out that the enrichment matrix favours certain IgG clones over others. This would introduce a bias into which IgG clones are being analyzed instead of reflecting the original clonal distribution in plasma. Analysis of IgG from neat plasma or other enrichment matrices are needed. Also, the known and mentioned diversity in IgG modifications of both the Fc and Fab do contribute. These modifications have been previously studied by the first author (PMID: 25004930) and I would expect that these have an effect on the enrichment and peptide detection. Please comment and expand on this matter.

(4) It is known that antibody levels provide a unique fingerprint and remain extremely stable over time in clinically healthy subject in comparison to other omics data types (see PMID: 32900998). The authors chose a small cohort of septic patients to demonstrate disease perturbations. They investigated the changes over time and in relation to pre- and post-diagnosis samplings. While this is certainly a study worth pursuing, I doubt that the differences in sampling timepoints provide sufficient similarity to allow a conclusion that can be meaningful. Sepsis has a very strong impact on the immune system and these individuals have most certainly received treatments to their condition to counteract symptoms and causes. These variables make it very difficult to collect sufficient common evidence for the given statements. There is growing acceptance that inter-individual diversity in the 'omes, heterogeneity in treatment response and health trajectories further complicate the efforts in finding commonalities in the data between patients. It would therefore be preferable to first study individual health profiles in consecutively sampled asymptomatic donors (eg 3 subjects and 10 samples) that do not have to be reduced to one common clinical phenotype or that are being treated .

(4) The section about "mass spectrometry-based de novo sequencing of an individual plasma IgG1 clone"

confuses the flow of the manuscript. I understand the value of these proof-of-concept investigations but they would be better suited as a main content of a separate submission. Otherwise, I would expect that these investigations would include the clinical samples as well.

(5) The discussion did not provide the expected critical reflections on the own data and limitations that would put this study into a context. It rather presented as an expanded outline of the topics already covered by the introduction.

In summary, there is no doubt that this work is a fine proteomics study and suitable for an audience that acknowledges these analytical challenges. There are a few critical aspects that require more data to provide convincing arguments for a broader interpretation of the observations. This includes to anchor the investigation in other data dimensions.

---

## Authors' response to the reviewers' first round comments

Attached.

---

## Editorial decision letter with reviewers' comments, second round of review

< Dear Albert,

Thank you very much for your patience, and I hope you've enjoyed your holiday. I'm very pleased to let you know that the reviews of your revised manuscript are back, the peer-review process is complete, and only a few minor, editorially-guided changes are needed to move forward towards publication.

In addition to the final comments from the reviewers, I've made some suggestions about your manuscript within the "Editorial Notes" section, below. I'll also send you a lightly line-edited version of your manuscript directly over email. Please consider my editorial suggestions carefully, ask any questions of me that you need, make all warranted changes, and then upload your final files into Editorial Manager.

As you look forward to acceptance, please do consider submitting one of the protocols you've developed in this paper to STAR Protocols, or extending this offer to one of your trainees. STAR Protocols is geared towards trainees and its key purpose is to provide complete and consistent instructions for how to conduct reproducible experiments. If you have any questions, please email starprotocols@cell.com.

I'm looking forward to going through these last steps with you. Although we ask that our editorially-guided changes be your primary focus for the moment, you may wish to consult our FAQ (final formatting checks tab) to make the final steps to publication go more smoothly. More technical information can be found below my signature, and please let me know if you have any questions.

All the best,

Quincey

Quincey Justman, Ph.D.
Editor-in-Chief, Cell Systems

---

**Editorial Notes**

*Title:* Your title is excellent and declarative, but I think you should consider highlighting the proteomic advance in the title itself. For example:

**Direct proteomic measurement reveals that human plasma IgG1 repertoires are simple, unique, and dynamic**

As you re-consider your title, note that an effective title is easily found on Pubmed and Google. A trick for thinking about titles is this: ask yourself, "How would I structure a Pubmed search to find this paper?" Put that search together and see whether it comes up is good "sister literature" for this work. If it does, feature the search terms in your title.

*Abstract:* I've gone over your abstract with the goals of increasing accessibility to a broad audience and making it more concrete. See what you think. Please feel free to revert anything that you don't like or that you feel distorts your meaning! I apologize if there are instances of the latter.

> *Although somatic recombination and hypermutation humans can theoretically produce billions of IgG1 variants, the diversity of IgG1 clones circulating in human blood plasma has largely eluded direct proteomic characterization. Here we use an LC-MS based proteomics approach to reveal that the circulating IgG1 repertoire in human plasma is dominated by a limited number of clones in healthy 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-centric 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 to 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.*

*Manuscript Text:* As mentioned above, I'll send you a very lightly line-edited version of your paper. Note that I've also checked your manuscript's length and approved it -- please do not cut your text to meet any stated word/character limits. Also please note that house style disallows editorializing within the text (e.g. strikingly, surprisingly, importantly, etc.), especially the Results section. These terms are a distraction and they aren't needed—your excellent observations are certainly impactful enough to stand on their

own. Please remove these words and others like them. “Notably” is suitably neutral to use once or twice if absolutely necessary.

*Figures and Legends:* Your figures are excellent. Note that you do have up to 7 full-page, multi-panel main text figures, so if you would like to move any figures or panels from the supplement into the main text, please feel free to do that. Also, a minor suggestion: please consider expanding Figure 1 to include a graphic depiction of IgG structure, as described in the first paragraph of your introduction. Regarding your figure legends, please double-check that they include all of the information necessary to interpret the figures they describe, without help from the main text. Note that we encourage defining n each time it appears: in the STAR Methods, main text, and Figure Legends.

**Thank you!**

---

**Reviewer comments:**

Reviewer #1: The authors addressed all my criticisms and provided extensive additional data on reproducibility and validation. The data on the stability of the serum repertoire in healthy individuals is of particular interest.

A side comment regarding comment #4 by reviewer #3: "It is known that antibody levels provide a unique fingerprint and remain extremely
stable over time in clinically healthy subject in comparison to other omics data types
(see PMID: 32900998)." As far as I could tell the cited paper did not present data on the Fab composition of blood. I believe there is literature on antibody titer stability over time (see work by Slifka and coworkers) but not on Fab composition in blood which is a separate issue

Reviewer #2: The authors have applied considerable and noticeable effort to revise the manuscript. I believe they have addressed most of the reviewers comments and concerns, and they have added experiments and clarity which have enhanced the overall story. In general the manuscript feels more cohesive and reads well. I also think with the additional experimental detail included this would make it possible for another group to perform similar studies. Aside from overall revision of the message and detail, the two most important updates of the revised version to me were a) taking a known mAb through the same process, b) construct generation of the F59 clone and comparison of MS and MSMS spectra (Fig 5) to recombinant, and c) the longitudinal analysis of healthy patient sera.

I think it's a very significant (controversial?) claim in Discussion (second to last paragraph) that challenges the dogma of the long-accepted ~20 day Ab half-life. The authors claim that "...continuous production seems to be the normal behavior of clones...", and that this phenomenon dominated in both sepsis patients and healthy. The authors reference observation of autoantibodies exhibiting the same behavior. Can the authors find other substiantiations of the claim? Or speak to why their results are contrary to multiple reports claiming otherwise (https://pubmed.ncbi.nlm.nih.gov/3183495/, just as one

example)? Perhaps there is a technical reason for the Abs that do not change levels longitudinally. For me I think this is extremely interesting and would be an amazing find if true, but I don't think there is enough evidence from this study to substiantiate the claim, so it should either be softened or discussed with a stronger argument. ***[From QJ: Please do carefully consider this comment and edit your discussion -- beyond my line edits -- to be appropriately circumspect about this point.]***

Although the clarity of the writing has improved from the first submission, I still stumble in areas in the manuscript where there are unusual grammar or phrasing choices that distract from the story. I would encourage a final revision with a critical eye for phrasing and flow.

---

Reviewers' comments:
Reviewer #1: In this very interesting and important but also somewhat disorganized paper by Heck and coworkers they present methods for analyzing the overall composition of the antibody repertoire in plasma at low resolution and separately de novo sequencing of an abundant IgG identified in plasma. A key biological finding from this work is that the serum repertoire appears to be dominated by a few IgG clones, as measured from the MS1 of Fab fragments whereby as few a 30 Fabs with distinct average masses seem to account t for >70% of the Serum IgG. The authors show that the dominant IgG clones are different among individuals but for any particular individual they show significant resistance longitudinally although with the relative amounts of different peaks changing over time. In the second part of this paper the authors succeeded in partial de novo sequencing of an abundant plasma antibody, a significant feat. While the paper is certainly informative and appropriate for publication, it will first require major revision to address the following:

We are pleased to note the appreciation of this reviewer for the key findings of our work both on the "simplicity and uniqueness of the IgG1 profiles" and the "de novo sequencing of a IgG1 by combining middle-down and bottom-up proteomics". We agree that the write-up of the paper could be improved, and followed the suggestion, also of the editor, to extend/hydrate the paper with more descriptions and new experiments as outlined further below.

Major:
1) The idea of profiling plasma IgG by first producing Fabs by IgdE digestion and then analyzing and quantifying IgG species from the MS spectra after LC/MS is simple and could be very important. Having said that, there is a lack of detail and controls that are concerning:
2) The authors state that the focused on analysis of Fabs because they lack the Fc glycan which contributes to mass heterogeneity. However in humans 25-30% of Fabs are in fact glycosylated. I am not sure why this is not addressed. I suspect the authors must have observed some clones with a broader mass distribution suggestive of glycosylation heterogeneity.

Concerning 1, we have now extended and revised the manuscript and included several control experiments with spiked in mAbs, as also described further below.

Concerning 2, indeed part of the circulating IgG1 may be glycosylated on their Fab, and we do see evidence for that as now described in the revised manuscript. However, from our quantitative assessment of their abundance we observe much less Fab glycosylated molecules than the quoted 25-30%. We find it is patient specific and ranges in between 0.5 and 7% in the donors we studied (Suppl. Figure S3). There may be several reasons for this seeming discrepancy, notably that a) the signal intensity for glycosylated Fabs in our profiles is spread over the different glycoforms and therefore less well detected, or that b) Fab glycosylation is less abundant than previously proposed. In response to the reviewer, we now discuss the presence of Fab-glycosylated IgG1 in the paragraph 'Plasma IgG1 repertoires are dominated by a few clones' of the Results section. We also show evidence of them in the data of both the healthy and sepsis donors, showing that the Fab glycosylation levels are really donor specific.

3) Data showing detection and importantly relative quantification of mock samples comprising of say 20-30 Mabs of known sequence and MW and mixed at different concentrations should be shown to allow the reader to evaluate the sensitivity and quantitation in the method.

We have included in the revision several control experiments and summarized the data in **Figure S1**: "Performance evaluation of plasma Fab profiling approach using various experimental controls". We assessed the accuracy and precision in mass, retention time and abundance of spiked-in monoclonal antibody controls. For these experiments six monoclonal antibodies (trastuzumab, cetuximab, rituximab, alemtuzumab, bevacizumab and infliximab) were added at 20, 200, 800 and 4000 ng in a plasma background. These experiments revealed great performance and reproducibility in accuracy and precision of measured mass and retention times. Moreover, these experiments also revealed a linearity in quantitation down to 20 ng of mAbs (translates to the concentration of 2 µg/mL for a clone in plasma) that could also be reproduced in injection replicates and full analysis workflow replicates.

4) How was the absolute quantitation of the Fabs validated? I presume quantitation was based on peak area but how exactly was Was it with just the two spiked Man standards. More details are needed.

All samples were deconvoluted the same way using BioPharma Finder 3.1 software (BPF; Thermo Scientific, San Jose, USA). The intensities of the mAbs reported by BPF were averaged and the average intensity was set at 20 µg/mL (since it was 200 ng in 10 µL plasma). All other intensities were normalized to this value. This is now more extensively described in the Methods section. See also the answer above, describing the control experiments now included and depicted in Suppl. Figure S1.

5) It is quite conceivable that each mass peak is in fact comprised of multiple antibodies having very similar MWs. In fact this is not uncommon, and probably should be expected that each "average mass peak" includes multiple mabs, for example having the same gremlin usage and similar length CDR3s giving comparable masses. This needs to be addressed in detail.

With the theoretical amount of Fabs present being very large (millions to billions) it can certainly not be excluded that two Fabs can have the same "exact mass" and retention time. The chance however that two of these are among the top 30-100 most abundant Fabs is already much lower. Moreover, in the ETD runs for the clone sequencing we also measured the masses of the LC and Fd of fragmented Fabs with high resolution, and of course two Fabs with identical exact mass have likely not the same identical masses for their LC and Fd chains. Therefore, we think that the chance that in our analysis two (highly abundant, i.e. top-100) Fabs exhibit identical mass and retention time is low.

6) Reproducibility among technical replicates should be shown. Also for biological

replicates is starting with the same plasma and performing separate digestions and LC/MS then evaluating variability.

We performed these experiments, and the data is presented in our revised manuscript, in the paragraph 'Mass spectrometry-based Fab profiling of the human plasma repertoire' of the Results section, and Figure S1. See also our answer to question 3) . This point is also addressed by the new experiments including the Fab profiling of samples from healthy donors taken at three successive timepoints

7) While impressive the de novo sequencing opens several questions: Was the approach validated e.g. by using middle down and bottom analysis of a known Fab?

To reinforce the applicability of our integrative *de novo* sequencing approach and address comments of the reviewer we followed this suggestion and had a monoclonal antibody generated with the exact sequence as determined in this study *de novo*. Next, we repeated all the experimental steps, including Fab profiling, middle-down and bottom-up LC-MS/MS for this synthesized monoclonal antibody. Pleasingly, the obtained data is fully identical. Masses of the Fab, light chain, and Fd portion are within 20 ppm from the masses determined for the clone observed in the donor F59. We also could find back all of the key peptides that were used for the determination of the sequences from the endogenous clone, including the important peptides that span the CDR regions. Please see the new Figure 5 for a comparison of the key features that were used for integrative *de novo* sequencing between the donor clone and the newly synthesized recombinant antibody. Also, we extended the text with an additional part on "Validation of *de novo* sequenced Fab from donor F59". The new work described in the revised manuscript supports more in general that our combined bottom-up and middle-down approach enables to *de novo* sequence by mass by MS highly abundant plasma IgGs.

8) tHe strategy for filling the gaps in the VH is not presented clearly enough in the main text and SI Fig 7-9 are not quite clear either. How did the authors distinguish Leu from Ile?

In the revised manuscript we have extended our description on the *de novo* sequencing to make clearer how we performed the gap filling procedure. In particular the paragraph 'Protein-centric (middle-down) de novo sequencing' in the Methods section provides now more detail on the gap filling procedure.

Distinguishing between Leu and Ile is a very challenging task even for *de novo* sequencing of recombinant monoclonal antibodies, whereby very specific fragmentation approaches and formation of secondary fragments are essential. In order to have some degree of support for placing either Leu or Ile at a specific position we used available databases of germline sequences and experimental antibody sequences, from which either Leu or Ile was chosen based on its prevalence of appearance.

9) Validation of the de novo sequencing by synthesizing the Mab and then comparing

the middle down and top down spectra with those from the plasma Fab would be helpful -but not required,

Please see the answer to comment #7. We indeed generated the monoclonal antibody based on the sequence determined for the clone from donor F59 and repeated all the experimental steps for validation of *de novo* sequencing procedures used in the current approach. This is now all included in the revised manuscript. Pleasingly, the recombinant mAb displayed highly alike features in the LC-MS, bottom-up and middle-down proteomics experiments.

Minor but still important
1) The two parts of the paper are disjointed. Given that Fig 1-3 address a very different question than Fig 4 and that extensive additional data is required to firm up the first part of the manuscript maybe the authors should consider splitting it into two parts.

Discussing this further also with the editor we made a fully revised extended revision and improved the link between the sections. We feel that the combination of the IgG profiling and the proof-of-concept that we can sequence a plasma IgG clone *de novo* make this work more impactful and thus we prefer not to separate these parts.

2) The writing needs to be improved. In some parts it reads like a graduate thesis. Detailed information on important controls and quantitation should be provided.

As above, discussing this further also with the editor we made a fully revised extended revision and improved the link between the sections, improved the figures and the legends. Moreover, a whole new section and figure has been added to the revised manuscript with detailed information on controls and quantitation.

3) Some rationale as to why the authors elected to study samples from septic patients should be provided.

The rationale for taking samples from septic patients is two-fold. First of all, we expected those people to show a rapid and measurable antibody response to the pathogen(s) responsible for the sepsis. Secondly as these people were admitted to the ICU, we could receive longitudinal samples from a biobank administered by our local hospital. The first reason we now mention in the revised manuscript. Of note, we included in our revised manuscript now also data on two healthy donors, whereby per donor 3 serum samples were taken over a time period of 2 months, sampling each consecutive month.

**Reviewer #2**: The authors demonstrate antibody tracking and profiling of patient sera using a combination of mass spectrometry technologies. The authors present a clear and meticulously described methodology, which holds great promise for the further integration of MS based analyses into a therapeutic setting. The most exciting piece of the study is the unique sample set the authors were able to obtain.

Although the goals listed in the Summary are phrased broadly, the stated intentions were largely met. My outstanding question after reading the paper is whether the detected, unique antibodies are indeed indicative of sepsis or recovery, and furthermore whether a healthy individual's antibody repertoire would behave similarly. I believe it is crucial to add a longitudinal analysis from a healthy individual to give weight to the idea that the similarities and differences found over time in these patient are meaningful. ***[From QJ: If adding at least one longitudinal analysis from a healthy individual is not possible, please let me know.]***

Without antigen-specific enrichment of the repertoire it is possible the methodology is simply tracking changes in the most abundant circulating antibodies. It may be possible that a healthy individual could deliver a similar result to the patients reported herein in that: a) each individual is different, and b) levels of specific Abs change over time. Were any assays performed on the fractionated patient Ab that was sequenced to show antibody relevance?

Although it was harder for us to receive donor samples from healthy individuals (with consent) we were able to purchase plasma samples of two healthy donors whereby blood sampling was undertaken at three consecutive time points (each a month apart). The analysis of these samples is now included in the revised manuscript. It revealed that indeed also the human plasma IgG1 repertoires of healthy donors are equally "simple", and unique, albeit that they revealed (as expected) much less dynamics in changes in the repertoires.

"…seemingly sepsis-responsive, single IgG1 clone…" is a quote from page 4. The authors never come back to this nor does there seem to be any evidence to uphold the claim. This statement could be misleading to readers as there is not any demonstrated correlation between the detected Abs and the patient condition.

By including in the revision also the analysis of the two healthy donors at three blood sampling times, it has become apparent that there are more changes in the IgG1 profile over time in the sepsis patients than in the healthy controls, and therefore it seems that the occurrence of the abundant clone in F59 is seemingly sepsis-responsive. Nevertheless, to avoid over-interpretation of our statement, we modified the sentence to "a single plasma clone – that appeared at sepsis onset –".

Was there a rationale for choosing plasma instead of serum for analysis?

We are highly dependent on clinical collaborators for the samples. In this case, for the sepsis cohort it was easiest for the clinicians to provide plasma.

Of note, our method begins with the affinity purification of the IgG, and this process works equally well in serum as in plasma (and the majority of other biofluids).

Fig S4: It appears there is enhanced CDR-L3 sequence coverage from ETD in the Fab as compared to the reduced LC (also looks true for CDR-H3 in Fab versus Fd). Does this trend hold true with ETD of other Fabs? This would be an interesting observation.

The reviewer is right, we do observe better CDR-L3 and CDR-H3 coverage in the non-reduced Fabs when fragmented by ETD. This phenomenon came to us not as a surprise as we and others previously demonstrated that for intact antibodies, Fab2, or Fab molecules when dissociating them using electron-based approaches (see "Selectivity over coverage in de novo sequencing of IgGs." by den Boer MA, et al in Chem Sci. 2020;11(43):11886-11896. doi: 10.1039/d0sc03438j). Such preferential formation of fragment ions was earlier explained by the accessibility of this regions located between two intra-chain disulfide loops.

The authors note that a combination of middle down and bottom up data were used to obtain 100% sequence coverage, and although the process of how middle down and bottom up data were combined, it appears that middle down search results were used as a template for bottom up identified peptides to fill in. What benefit did ETD provide?

First of all, as mentioned in the previous response, ETD of the Fab provides relatively clean series of N-terminal c-ion fragments in the region between the two disulfide loops. Such specificity of fragmentation enables the detection of sequence tags for consecutive fragment ion peaks. Moreover, ETD uniquely dissociates also inter-chain disulfides in the Fab molecule, therefore providing the masses of the constituting LC and Fd chains. Nevertheless, the presence of intra-chain disulfides hampers retrieving sequence information in most other regions. Therefore, the additional middle-down analysis under reducing conditions turned out to be crucial.

Meaning, if one were to send the bottom up data to PEAKS for de novo assembly, would the correct sequence have been called? How (besides match to intact MW) was the sequence validated?

We attempted this with the Supernovo software from Protein metrics, which should provide the same output as PEAKS-AB, for de novo identification of the dominant clone from donor F59. However, the presence of peptides of co-isolated contaminant Fabs poses a significant hurdle for this kind of analysis. Such software solutions are still heavily optimized for the sequencing of a single purified mAb and fail in our case.

We have revised the manuscript and supplemented it with the data on a recombinant monoclonal antibody, generated based on the sequence determined by us for the dominant clone from donor F59. By performing all experimental steps that were initially used for the analysis of the endogenous clone we have further validated the sequence through comparison of various features, including mass and retention time in the Fab LC-MS, and comparing fragmentation patterns in the middle-down ETD of the Fab, and comparing the same peptides and their respective fragmentation spectra in the bottom-up approach. Additionally, when we submit the sequencing bottom-up MS data generated for the recombinant clone to Supernovo, it yielded the correct sequence for both the light and heavy chains.

Finally, it seems the real 'meat' of the manuscript is held within Supplemental. There is a great deal of information represented in the supplemental figures. I suggest to move a representative ETD spectrum from Fig S4 to main since the de novo sequencing aspect of the study is deeply discussed, perhaps a figure to accompany would be in order.

We have revised our manuscript and included such ETD data in the new main Figures 3 and 5.

**Reviewer #3**: Bondt et al provided a manuscript on "Human plasma IgG1 repertoires are simple, unique, and dynamic" in which they describe a mass spectrometry based approach to study the immunoglobulin type IgG1 in human plasma. They sequence the immuno-enriched immunoglobulins on the level of peptides and larger protein fragments and interpret the obtained data to reveal (1) unique IgG1 fingerprints in the tested subjects, (2) changes of these fingerprints in relation to sepsis. While the strength of this work is clearly grounded in the mass spectrometric expertise of the group, there is lack of evidence for the critical aspects and claims.

(1) From my perspective, a systems biology study of this kind should include other data types to corroborate the observations. ***[From QJ: I appreciate this reveiwers' point of view but respectfully disagree, given the sophistication of the proteomics and the interesting clinical cohort. If you have concerns about how to address this point, please let me know.]*** This can include the integration of RNAseq data from B cells that disentangle the clonal diversity as well as studies of "antibodies in action" by methods capturing the binding pattern. The latter would allow to investigate if any common antigens are being targeted by the immune system. Given the hypothesis that the authors' detect the more abundant IgGs in their study, it should be conceivable to detect these also on protein microarrays and enrich these on their respective antigens. Such analyses could then provide further targeted insights in the clonal aspects and specificity, which makes such a study more attractive to a wider audience.

We find this an interesting perspective, but really a different one than what we have taken here and wish to pursue. When one seeks to find antibodies that target a specific antigen indeed the procedure suggested by reviewer would be required. However, here the question has been more focused on how the general (healthy) human plasma IgG repertoire looks like. The advances we have made that are described in the manuscript are to do this directly at the site where this question should be answered, namely at the protein level and in the plasma. Also, by including now new experiments on healthy controls we observe very simple IgG1 clonal repertoires. For us, an interesting question that still remains to be resolved is; what do these abundant clones in healthy controls target? However, that is outside the scope of the current manuscript, but target of follow up studies.

(2) The data is built on a multi-step workflow, and even though I acknowledge the efforts of establishing such a pipeline, I miss key figures about precision, accuracy and reproducibility. Without such measures it remains difficult to judge the results. Here, true biological replicates are needed that process samples through the whole pipeline. For a layperson, the obvious differences in the attached chromatograms provoke further uncertainty about the reproducibility, even though the matching of the peptides in the main figures tried to contradict such suspicions. I do acknowledge the authors' transparency by providing such figures, but bringing the work beyond a proteomics community would require to explain these.

This is now extensively discussed in the revised manuscript and summarized in supplemental Figure S1. See also our more detailed response to reviewer 1, point 3.

Furthermore, we have put an effort in making the chromatograms clearer. In addition to the (now smoothed) raw chromatograms we have included an additional line in all chromatograms that shows the proportion of the profile that was identified as IgG1 Fabs. Noisy interference in some of the chromatograms is caused by the (sometimes only partially removed) IgdE enzyme, and by some remaining albumin from the samples. Since these proteins have a clearly different masses (65-70 kDa), they do not interfere with further downstream analysis.

(3) My main concern though lies in the effect of the enrichment of IgGs from plasma. There is no data on the completeness of the enrichment. More importantly, there is no information to rule out that the enrichment matrix favours certain IgG clones over others. This would introduce a bias into which IgG clones are being analyzed instead of reflecting the original clonal distribution in plasma. Analysis of IgG from neat plasma or other enrichment matrices are needed. Also, the known and mentioned diversity in IgG modifications of both the Fc and Fab do contribute. These modifications have been previously studied by the first author (PMID: 25004930) and I would expect that these have an effect on the enrichment and peptide detection. Please comment and expand on this matter.

We have minimized the number of enrichment and depletion steps in the protocol to avoid the introduction of a bias. Furthermore, in our approach we have applied a common IgG enrichment matrix that is designed to recognize all IgGs in the CH3 domain, as far away as possible from the variable domain, the domain where all clones differ and that may introduce preferential enrichment. For this affinity matrix it has been shown that no bias is introduced when using the glycopeptide (CH2 domain) as readout (Amez Martín et al., 2021). It cannot be excluded with absolute certainty that no bias is introduced, but we have chosen the materials and steps in the protocol that should minimize this. Unfortunately, even more commonly used affinity matrices (such as protein A or protein G) have some affinity for certain Fab domains in addition to their preferred binding site on the Fc.

In order to confirm that even glycans on the variable domain do not pose a threat to our approach we have included the Fab-glycosylated monoclonal antibody cetuximab in the validation experiments, also mentioned in response to Reviewer 1 remark #3. Cetuximab performs as well as the other mAbs in the experiment.

This is now also discussed in the revised manuscript.

(4) It is known that antibody levels provide a unique fingerprint and remain extremely stable over time in clinically healthy subject in comparison to other omics data types (see PMID: 32900998). The authors chose a small cohort of septic patients to demonstrate disease perturbations. They investigated the changes over time and in relation to pre- and post-diagnosis samplings. While this is certainly a study worth pursuing, I doubt that the differences in sampling timepoints provide sufficient similarity to allow a conclusion that can be meaningful. Sepsis has a very strong impact on the immune system and these individuals have most certainly received treatments to their condition to counteract symptoms and causes. These variables make it very difficult to collect sufficient common evidence for the given statements. There is growing acceptance that inter-individual diversity in the 'omes, heterogeneity in

treatment response and health trajectories further complicate the efforts in finding commonalities in the data between patients. It would therefore be preferable to first study individual health profiles in consecutively sampled asymptomatic donors (eg 3 subjects and 10 samples) that do not have to be reduced to one common clinical phenotype or that are being treated .

To address the issues indicated by the reviewer regarding the clinically complex sepsis sample set, we now additionally included longitudinal data for two healthy individuals. The sampling covers a similar time span as for the sepsis samples. Also, for these healthy individuals we can conclude that the IgG clonal profile is dominated by a limited number of clones (>50% concentration in only 30 clones), the profiles are unique for each individual, and the repertoire is dynamic. In particular the last statement is a lot less pronounced in the healthy individuals, since they do not go through major immunological challenges.

We are very excited about the remarkable dynamics we can observe in the sepsis samples. Our data does however indeed not allow any clinical conclusions, but the mere visualization of the antibody plasticity was the main target in this study.

In relation to the reference presented by the reviewer, that is regarding autoantibodies. The authors of the particular manuscript describe an extremely stable recognition of certain autoantigens in healthy individuals. However, the used methodologies cannot detect whether this always concerns the same clone, or different clones recognizing the same target.

(4) The section about "mass spectrometry-based de novo sequencing of an individual plasma IgG1 clone" confuses the flow of the manuscript. I understand the value of these proof-of-concept investigations but they would be better suited as a main content of a separate submission. Otherwise, I would expect that these investigations would include the clinical samples as well.

By extending our manuscript and including new experiments and data we have also expanded the explanation on how we performed our analysis and sequencing from one of the clinical samples.

(5) The discussion did not provide the expected critical reflections on the own data and limitations that would put this study into a context. It rather presented as an expanded outline of the topics already covered by the introduction.

We expect that the revised version of the manuscript meets the expectations of the reviewer. Nearly all issues and potential limitations (as addressed by this and the other reviewers) are resolved or addressed in the text.

In summary, there is no doubt that this work is a fine proteomics study and suitable for an audience that acknowledges these analytical challenges. There are a few critical aspects that require more data to provide convincing arguments for a broader interpretation of the observations. This includes to anchor the investigation in other data dimensions.

In summary and in response to all reviewers, we hope that by including all the control experiments on the mAbs, the new data on the healthy controls and by sequencing also a recombinant mAb based on the earlier by us derived *de novo* sequence we have lived up to the expectations of this reviewer and the other reviewers.



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. 10.1021/acs.jproteome.1c00148.