PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Predictors of response and disease course in patients with inflammatory bowel disease treated with biological therapy – The Danish IBD Biobank Project: Protocol for a multicenter prospective cohort study
AUTHORS	Zhao, Mirabella; Bendtsen, Flemming; Petersen, A; Larsen, Lone; Dige, Anders; Hvas, Christian; Seidelin, Jakob; Burisch, Johan

VERSION 1 – REVIEW

REVIEWER	Miles Parkes Cambridge University Hospital
REVIEW RETURNED	17-Dec-2019

GENERAL COMMENTS	This is an admirable and important study. Others are undertaking similar work. My main concerns relate to statistical power and heterogeneity in the cohort. There is an assumption that the mechanisms that underlie PNR and 2ry LOR are the same; and that these will be shared between CD and UC. This significantly undermines the power calculation - which also fails to account for the 'multi-omic' approach which is otherwise a potential strength of the study. In addition, to my mind there are too many primary end points. The investigators should focus on ONE primary end point. The rest should be secondary end points. I wonder if it might be possible to expand the sample size? this would strengthen the study considerably.

REVIEWER	Nuru Noor Addenbrooke's Hospital, Cambridge, United Kingdom
REVIEW RETURNED	20-Dec-2019

GENERAL COMMENTS	Overall this is a interesting paper describing an important project with significant potential to uncover novel biological findings around response and non-response in IBD (which is currently the key clinical question in the field). There are a few points highlighted below, either for clarification or for addition, to help further understanding around the methodology of the project for potential readers and to help understand some of the practicalities of the project also: 1. Given this is a Danish Biobank project, perhaps in the introduction the authors could provide specific details on the epidemiological incidence and prevalence of IBD in Denmark. Likewise with regards biologic medications, an indication of how many of these patients

	are currently on biologics and costs in Denmark would be useful to help set the scene and explain why a Danish Biobank specifically is necessary. 2. Please can you provide rationale for why the number of 840 participants was selected for this Biobank cohort? Are there preliminary or other studies from the various -omics fields that you can cite that would indicate that this number would be sufficient to be able to develop potential predictive markers mentioned in the text? 3. The recruitment of 840 participants based on calculations of 5-6 patients starting on biologic therapy at each centre, each month, supposes that almost all patients due to start biologic therapy will enrol into this Biobank. Given that recruitment is already ongoing, could you include how many participants have been recruited to date and over what time period? Is the current recruitment timeline feasible based on recruitment so far? Could you also elaborate on how many participants would be expected to be recruited each year (given likely initial lag to get sites up and running) - as this may mean a greater proportion of patients with more limited follow-up time. 4. Follow-up is currently scheduled until May 2023. Can you provide an indication of whether longer term follow-up is planned to take place thereafter and if for example this were to be through the use of electronic health records, a brief explanation of how this would be done in Denmark and whether patient consent for longer term follow-up has been or will be considered. 5. In the study population paragraph, can you give an indication of proportion of UC and CD patients you would expect to be on these respective biologic medications in Denmark. For example if only very limited numbers on tofacitinib then already gives indication that likely to have low predictive power to develop robust and clinically useful predictor for this treatment. I would certainly still include all such groups, but providing proportions for treatments would give readers a greater understanding on likelihood for development of treatment predictors from this cohort. 6. Regarding choice of primary non-response. Please can you explain reason for selecting 12 week timepoint for ustekinumab and tofacitinib. Whilst further work required, some may feel that this is too early a timepoint for both medications, particularly for ustekinumab. 7. Regarding endoscopic remission for Crohn's disease and SES-CD score of <4, should this be stratified with different scores for patients with ileal disease vs. those with ileocolonic and colonic disease. 8. Regarding clinical data collection, smoking status is not mentioned and would be helpful to collect. 9. Regarding data management, the data collection technique is described. However please can you elaborate on plans for subsequent data curation/cleaning and management, including infrastructure that will be used, coding system that will be used and the personnel that will perform this.

	10. Regarding the e-CRF. Can you briefly summarise how many items of data would be collected? How long this takes to complete and which members of staff will be expected or have been completing this to date? Are you able to provide an indication of completeness of data for completed e-CRFs, based on the recruitment to date? 11. Can you elaborate on data sharing plans and whether a data sharing protocol has been developed or due to be developed? How will potential researchers apply to access the data and the biosamples? Will this be an open access Biobank or controlled/closed access resource? 12. I note that funding support has been received from Takeda. Again related to data sharing point above, are there already plans in place for commercial partners to collaborate/ work with or have access to data from this Biobank? 13. Could you elaborate on how any developed predictive marker would be validated in an external cohort? Is collaboration with other Biobanks or Bioresources either in Europe or globally envisaged? 14. The authors have clearly described this Biobank project. Could you clarify from the ethics application/consent form whether there is capacity for patients to be recalled following completion of the study more akin to a Bioresource? Either based on genotype or phenotype e.g. to take part in future Danish IBD research studies? 15. The authors report a lack of patient and public involvement in development of the research question or design of the study. I would encourage adoption of patient and public involvement to help advise ongoing conduct of this study and importantly for future work when results become available. This section in protocol could accordingly be edited to report future plans from incorporation of PPI.

REVIEWER	Ho-Su LEE KU Leuven, Belgium
REVIEW RETURNED	03-Jan-2020

GENERAL COMMENTS	This is a nice study design by Zhao et al describing a plan for biobank with sequential samples from patients with IBD. This study will add to the growing body of literature on molecular pathways. I believe that it is certainly a valuable study. I have the following comments in an attempt to strengthen the study design: 1. There is a need for exclusion criteria such as pregnancy or a very old population. 2. Please clarify the Statistical method for multi-omics analyses.

VERSION 1 – AUTHOR RESPONSE

Reviewer 1 comments:
- This is an admirable and important study. Others are undertaking similar work. My main concerns relate to statistical power and heterogeneity in the cohort. There is an assumption that the

mechanisms that underlie PNR and 2ry LOR are the same; and that these will be shared between CD and UC. This significantly undermines the power calculation - which also fails to account for the 'multi-omic' approach which is otherwise a potential strength of the study.
In addition, to my mind there are too many primary end points. The investigators should focus on ONE primary end point. The rest should be secondary end points.
I wonder if it might be possible to expand the sample size? this would strengthen the study considerably.

Thank you for providing your valuable comments on the manuscript. We fully agree that it is important to take account of heterogeneity in the cohort and we will perform separate analyses for the primary outcomes PNR and LOR and stratify for IBD subtype (CD, UC, IBDU) within the limits of the sample size of each subgroup. Our sample size for primary analysis is calculated based on the assumption that primary non-response will occur in 25% of patients (which is a conservative estimate), accordingly, a sample size consisting of a minimum of 197 patients with primary non-response and 66 responders should be able to detect a 1.3-fold upregulation of relevant biomarkers. The number of 840 participants is a minimum number as the study aims to expand uptake area in the future, this number is estimated based on the average number of biological-naïve IBD patients expected to initiate biological treatment at each study center during the inclusion period, thus, as a result of the study's observational design, the number of participants will also be limited by the rate of initiation of biological treatment in the clinical setting.
We agree with the reviewer that expansion of the sample size would strengthen the power of subgroup analyses, and we do intend to expand the number of participants beyond 840 (which is the minimum enrolment target). In fact, two additional study centers (Odense University Hospital and Hospital Soenderjylland) which are expected to recruit 5 participants per month have joined the study since the manuscript was drafted. Two other study centers have expressed their interest in study participation, and we will continue to work on expanding the study in the future by collaborating with other hospitals. Following the increase in the number of candidate patients available for study enrolment, we expect the final enrolment and sample size to exceed the current minimum estimates. We have addressed the potential limitation of enrolment rate in lines 67-69 and 345-348 and added information on inclusion of the two new study centers to lines 135-136 and lines 151-159.

We have now limited primary outcomes to only include primary non-response and secondary loss-of-response (lines 194-211):

Primary outcomes in this study are as follows:
1. PNR to treatment: defined as lack of clinical response with induction therapy defined as a decrease in SCCAI of ≥2 points from baseline in UC patients; or a decrease in HBI of >3 points from baseline in CD patients, as well as patients who undergo intestinal resection or colectomy due to IBD, or as fistula revision in patients with perianal CD during the period of induction therapy.
2. LOR to treatment: defined as patients achieving clinical response (as measured by clinical activity indices) during the period of induction therapy, but who later suffer from clinical relapse during maintenance therapy, including the need for rescue therapy with corticosteroids or an alternative biological therapy, or surgery for IBD.

Secondary outcomes in this study are as follows:
1. Clinical remission to treatment: defined as a SCCAI of ≤2 in UC patients; or a HBI of ≤4 in CD patients.
2. Endoscopic remission: defined as an UCEIS of ≤ 1 in UC patients; or a SES-CD of <4 in CD patients.
3. Surgery: defined as intestinal resection or colectomy due to disease activity of IBD not responding to medical therapy; or as fistula revision or drainage of abscesses after initiation of biological therapy in patients with perianal CD.

Response to Reviewer 2
Reviewer Name: Nuru Noor
Institution and Country: Addenbrooke's Hospital, Cambridge, United Kingdom

Reviewer 2 comments:
- Overall this is a interesting paper describing an important project with significant potential to uncover novel biological findings around response and non-response in IBD (which is currently the key clinical question in the field).
There are a few points highlighted below, either for clarification or for addition, to help further understanding around the methodology of the project for potential readers and to help understand some of the practicalities of the project also:

1. Given this is a Danish Biobank project, perhaps in the introduction the authors could provide specific details on the epidemiological incidence and prevalence of IBD in Denmark. Likewise with regards biologic medications, an indication of how many of these patients are currently on biologics and costs in Denmark would be useful to help set the scene and explain why a Danish Biobank specifically is necessary.

This is a very good suggestion, we have now added details on the incidence and prevalence of IBD in Denmark based on results from a nationwide cohort study from 2017 in lines 73-76:
"Between 1980 and 2013, the prevalence of IBD in Denmark is estimated to be 52.730, and the overall incidence of IBD was found to be steadily increasing and approached 25.9 per 100.000 person years."
We have also added information on the proportion of IBD patients treated with biologicals and the estimated annual costs in Denmark based on results from a nationwide cohort study from 2019 in lines 90-94:
"In a nationwide prospective cohort of IBD patients diagnosed in 2003 to 2004, the proportion of CD and UC patients exposed to biological therapy was 28% and 9%, respectively, and the annual cost of biologic therapy in these patients was estimated to constitute 1.9 million euros."

2. Please can you provide rationale for why the number of 840 participants was selected for this Biobank cohort? Are there preliminary or other studies from the various -omics fields that you can cite that would indicate that this number would be sufficient to be able to develop potential predictive markers mentioned in the text?

The number of 840 participants was selected based on an estimate of incident biological use in biological-naïve IBD patients at the four study centers, each study center initiates biological therapy in 5-6 biological-naive patients per month yielding a total of 840 patients during the three-year inclusion period, however, we plan to expand the study to include other study centers located in Denmark and the number of 840 patients represents a minimum enrolment target. The current estimate on sample size for primary analysis is calculated using the validated freeware system G*Power developed by Düsseldorf University, assuming that primary non-response will occur in 25% of patients which is a conservative approach, a sample size consisting of 197 patients with primary non-response and 66 responders will be able to detect a 1.3 fold upregulation (more correctly: a 0.3 fold up- or downregulation compared to responders) of relevant biomarkers. Since the manuscript was drafted, we have already included two more study centers (Odense University Hospital and Hospital Soenderjylland located in Region of Southern Denmark) where patient recruitment is expected to initiate in medio 2020, these centers are expected to contribute with 70 participants per recruitment year. Furthermore, two other centers have expressed interest in joining the project. Thus, we expect enrolment to exceed the minimum target of 840, although enrolment rate will depend on the actual

number of incident biological users at the study centers due as a result of the observational design of the study. We have addressed the potential limitation of enrolment rate in lines 67-69 and lines 345-348 and information on inclusion of the two new study centers has now been added to lines 135-136 and lines 151-159.

3. The recruitment of 840 participants based on calculations of 5-6 patients starting on biologic therapy at each centre, each month, supposes that almost all patients due to start biologic therapy will enrol into this Biobank. Given that recruitment is already ongoing, could you include how many participants have been recruited to date and over what time period? Is the current recruitment timeline feasible based on recruitment so far? Could you also elaborate on how many participants would be expected to be recruited each year (given likely initial lag to get sites up and running) - as this may mean a greater proportion of patients with more limited follow-up time.

Thank you for your suggestion. We have included 160 patients to date, over a period of 7 months, these participants are recruited at three of the four active study centers, as recruitment at one of the study sites (Aarhus University Hospital) was postponed due to long processing time for the establishment and approval of a collaboration contract and was just initiated in December 2019. According to our estimate of 5-6 participants per month at each study center, the current recruitment has exceeded the anticipated number of 105-126 participants (at three study centers in 7 months). We believe that the current recruitment timeline is feasible, based on the current recruitment rate we expect approximately 300 patients to be recruited each year. However, due to the observational study design, enrolment rate will depend on the actual number of bio-naïve IBD patients initiating biological treatment in the clinical setting. Furthermore, the annual number of recruited participants may be higher following the inclusion of the two new study centers in 2020. This information has now been added to lines 316-320.

4. Follow-up is currently scheduled until May 2023. Can you provide an indication of whether longer term follow-up is planned to take place thereafter and if for example this were to be through the use of electronic health records, a brief explanation of how this would be done in Denmark and whether patient consent for longer term follow-up has been or will be considered.

This is an interesting and important point considering that future research questions may arise based on results from the current study. We do intend to extend the follow-up of the study cohort in the future, however, the current approval of the study protocol only covers study conduct (including data collection from health records) till May 2023. It will be possible to extend the follow-up period by seeking approval from the Danish Ethics Committee and Data Regulatory Agency, in that case, a separate consent will be sought from study participants. The staffs conducting the study at each local center will then contact the participants in advance to study termination and ask for their consent for an extended follow-up.

5. In the study population paragraph, can you give an indication of proportion of UC and CD patients you would expect to be on these respective biologic medications in Denmark. For example if only very limited numbers on tofacitinib then already gives indication that likely to have low predictive power to develop robust and clinically useful predictor for this treatment. I would certainly still include all such groups, but providing proportions for treatments would give readers a greater understanding on likelihood for development of treatment predictors from this cohort.

This is a very good suggestion. All Danish hospitals are advised to follow the National Treatment Guidelines for biological treatment of IBD patients issued by the Medicine Council, ”National guidelines and recommendations from the National Board for the Use of Expensive Hospital Medication” (reference number seven in the manuscript), according to these guidelines, 80% of CD patients initiating biological treatment due to luminal activity are expected to receive either 1)

infliximab 2) adalimumab 3) vedolizumab as 1st or 2nd line treatment, all three drugs and ustekinumab may also be used as 3rd or 4th line treatment. These recommendations also apply to fistulizing CD patients except for the use of vedolizumab which is only approved as 3rd or 4th line treatment. In acutely severe UC, patients in need of 'rescue'-treatment with biologicals will receive infliximab. In chronic active UC who will initiate biological therapy, 80% are expected to receive 1) infliximab, 2) vedolizumab or 3) golimumab as 1st or 2nd line treatment, furthermore, tofacitinib may be used as 2nd line treatment, all above-mentioned drugs and adalimumab may also be used as 3rd line treatment. This information has now been elaborated in lines 165-176.

To our best knowledge, treatment pattern with the different types of biological drugs except anti-TNF inhibitors in IBD patients in Denmark has not yet been elucidated in the existing literature. Among biological-naïve UC patients who initiated treatment with anti-TNF inhibitors in the period between 2005-2014, 92% received infliximab, while 8% received adalimumab (reference: PMID: 27913244) in biological-naïve CD patients, 84% received infliximab, while 16% received adalimumab.(reference: PMID: 29239001) Unpublished data from a nationwide cohort study showed that among IBD patients in Denmark diagnosed in 2016 who had received biological treatment, the cumulative exposure to biologicals within two years of diagnosis among CD patients was 88% for infliximab, 12% for adalimumab, 5% for vedolizumab and 11% for ustekinumab. Among UC patients, cumulative exposure to biologicals within two years of diagnosis was 80% for infliximab, 26% for adalimumab, 11% for vedolizumab and 6% for golimumab. We would like to ask for permission from the editor and the reviewers to not include these unpublished data on the proportion of IBD patients on different types of biological treatment in the manuscript as these data are currently being prepared for publication in a separate manuscript.

6. Regarding choice of primary non-response. Please can you explain reason for selecting 12 week timepoint for ustekinumab and tofacitinib. Whilst further work required, some may feel that this is too early a timepoint for both medications, particularly for ustekinumab.

This is an interesting and relevant point. There are indeed studies suggesting that a second dose of ustekinumab eight weeks after the induction dose increases response rates assessed at week 16 in comparison to those assessed at week 8 after the induction dose (i.e. the UNIFI induction trial), likewise, some studies suggest that an extended induction therapy with 10mg tofacitinib for 16 weeks improves response rates in non-responders to 8 weeks of tofacitinib treatment (i.e. the OCTAVE Open trial). However, these results have yet to be validated in future studies and 8-12 weeks timepoint is more commonly used in the existing literature. We have selected 12 weeks timepoint in accordance to recommendations in the Danish national treatment guidelines, according to which response to treatment should be assessed after 12 weeks of treatment in Crohn's disease patients and 8 weeks of treatment in ulcerative colitis patients.

7. Regarding endoscopic remission for Crohn's disease and SES-CD score of <4, should this be stratified with different scores for patients with ileal disease vs. those with ileocolonic and colonic disease.

This is a relevant and interesting point, considering that there are studies supporting the theory that ileal disease presents a distinct disease subtype in comparison to colonic Crohn's disease and differs both in terms of disease course, treatment response and biomarker-profile from Crohn's disease otherwise. However, to our best knowledge, no study has yet validated use of stratified scoring with SES-CD dependent on disease location in CD patients, therefore, we have chosen not to use different cut-off values of SES-CD score dependent on disease location. However, we also collect data from imaging (MRI and abdominal CT) exams during the study period whenever available to evaluate mucosal healing using the Lemann Index in patients with ileal disease in addition to endoscopic remission.

8. Regarding clinical data collection, smoking status is not mentioned and would be helpful to collect.

Thank you for making us aware of the lack of information on smoking status. We have collected data on smoking status in the eCRF, information on smoking status is collected at baseline and updated at each study visit, smoking status will be updated every 6 months after the first year. Information on smoking status is specified as current use, former use or never user, for current and former users, information on duration and average number of cigarettes per day is recorded. This has now been added to lines 217-218 in the manuscript.

9. Regarding data management, the data collection technique is described. However please can you elaborate on plans for subsequent data curation/cleaning and management, including infrastructure that will be used, coding system that will be used and the personnel that will perform this.

All data in the project will be collected in a project database which is hosted centrally by the Capital Region in Denmark using the electronic data capture system REDCap. REDCap is a secure, web-based application designed to support data capture for research studies, providing an intuitive interface for validated data entry and audit trails for tracking data manipulation and export procedures. All study centers will have authorized access to study data. All access to the server and other server maintenances will be logged. Study setup and hosting will be performed by a PhD student at Hvidovre University Hospital. Roles in the system are given according to functions. The authorized staff at the clinical centers can add data to the electronic database and will keep the database current to reflect subject status during the study period. Once the eCRF for a subject is completed, the project personnel at each local center will approve the data using an electronic signature and thereby confirm the accuracy of the data recorded. Upon study termination, electronic data will be stored for an additional ten years before deletion, biological data will be transferred to a separate biobank for future research – this has been approved by the Data Regulatory Agency and consent from the participants is sought upon recruitment. Biological data will be stored at a centrally regulated biobank facility hosted by the Capital Region of Denmark (BIOSEK) with authorized access. Data will be stored for an additional 15 years before destruction. This information has been added to lines 252-267 in the manuscript.

10. Regarding the e-CRF. Can you briefly summarise how many items of data would be collected? How long this takes to complete and which members of staff will be expected or have been completing this to date? Are you able to provide an indication of completeness of data for completed e-CRFs, based on the recruitment to date?

The items of data collected in the e-CRF are specified in lines 213-225. There are 61 items to be collected at inclusion, 98 items at each study visit in addition to 33 items on laboratory test results at each study visit. However, not all items are specified for each patient but depend on disease phenotype and medical history of the given patient. Some items, such as those related to disease phenotype, comorbidities, will only be specified upon changes, otherwise the staff member responsible for data collection will specify the item as “no change since last visit”.
Likewise, data on surgery, medication, hospital admission and IBD-related examination are recorded upon the occurrence of such an event. The e-CRF takes about 10-15 minutes to complete for each study visit. At present, four staff members (one PhD student, one student research assistant, two clinicians) are responsible for eCRF completion which is performed continuously during study period - at each study visit for each participant. To date, we have complete data for e-CRFs (100%). This has now been elaborated in the manuscript in lines 252-258.

11. Can you elaborate on data sharing plans and whether a data sharing protocol has been

developed or due to be developed? How will potential researchers apply to access the data and the biosamples? Will this be an open access Biobank or controlled/closed access resource?

Development of the data sharing protocol for the study is still in process. In order to maintain responsible data sharing and to keep patient data confidentially, data collected in the study will not be shared as an open access resource, however, we intend to collaborate with other research groups and external researchers are welcome to apply for access to the biobank material for future projects by contacting the steering group of the project. In Denmark, collaboration with external research partners requires separate approvals from the Data Regulation Agency and the establishment of a specific data processing agreement, therefore, data sharing will be decided on a case-by-case basis in the steering group. This has now been elaborated in lines 266-273.

12. I note that funding support has been received from Takeda. Again related to data sharing point above, are there already plans in place for commercial partners to collaborate/ work with or have access to data from this Biobank?

The project has received an unrestricted grant from Takeda (Takeda Pharma A/S), Takeda will not be involved in data collection, analysis, interpretation, nor publication of the results, there are no plan in place for commercial partners to collaborate or access the data from the Biobank. This is now more clearly described in lines 370-345.

13. Could you elaborate on how any developed predictive marker would be validated in an external cohort? Is collaboration with other Biobanks or Bioresources either in Europe or globally envisaged?

We plan to establish an independent external cohort to validate the panel of predictive biomarkers detected in this study consisting of both IBD patients who initiate biological treatment and control patients naïve to biological treatment. The panel's performance in terms of accuracy, sensitivity and specificity of the biomarker panel will be evaluated in ROC-analyses, according to which the biomarker panel will be down-selected and refined. The validation cohort would consist of patients from participating centers who have not contributed to the current study, but we are also looking for possibilities to engage in international collaboration to facilitate and strengthen biomarker detection within the field of IBD on the global plan.

14. The authors have clearly described this Biobank project. Could you clarify from the ethics application/consent form whether there is capacity for patients to be recalled following completion of the study more akin to a Bioresource? Either based on genotype or phenotype e.g. to take part in future Danish IBD research studies?

According to rules and regulations set by the Danish Ethics Committee and the Data Regulation Agency, patients recruited in this study will not be able to be recalled for a future study unless the investigators apply for a separate consent from the patients for the specific study. However, by asking for patients' consent for storage and use of biological samples in a future biobank, the biological samples collected in this study will be stored for at least 15 years after study termination and may be used in future projects upon approval from the Data Regulation Agency.

15. The authors report a lack of patient and public involvement in development of the research question or design of the study. I would encourage adoption of patient and public involvement to help advise ongoing conduct of this study and importantly for future work when results become available. This section in protocol could accordingly be edited to report future plans from incorporation of PPI.

Thank you for this suggestion. We agree with the reviewer that involvement of patients in conduct and design of the study is highly valuable, we will investigate the possibility to involve the Danish patient

organization for Crohn's and Colitis patients as well as participants in the study in relation to development of future research questions which may arise and to involve participants in the study in relation to amendments to the current protocol.

Response to Reviewer 3
Reviewer Name: Ho-Su LEE
Institution and Country: KU Leuven, Belgium

Reviewer 3 comments:
- This is a nice study design by Zhao et al describing a plan for biobank with sequential samples from patients with IBD. This study will add to the growing body of literature on molecular pathways. I believe that it is certainly a valuable study.
I have the following comments in an attempt to strengthen the study design:

1. There is a need for exclusion criteria such as pregnancy or a very old population.

Thank you for your constructive suggestion. We have not defined exclusion criteria related to pregnancy or age in this study, because we aim to initiate a cohort which covers, as far as possible, all biological-naive IBD patients in the uptake area similar to the concept of an inception cohort to minimize potential selection bias and to reflect the true population of biological users in the real-life setting. However, we will stratify for potential confounding factors such as pregnancy and age differences in the analyses, in relation to age differences, we will also account for comorbidities which might impact biomarker findings in the elder patient group.

2. Please clarify the Statistical method for multi-omics analyses.

We intend to use a systems-biology based approach by integrating datasets consisting of transcriptomic, proteomic and microbial data with clinical phenotype in a combined analysis either by multivariate analysis using a supervised approach such as the mixOmics R-package which allows for dimension reduction and data integration between different types of omics-datasets using the multivariate framework DIABLO as well as data visualization. Alternatively, the omics-datasets would be integrated using a stepwise approach by first identifying relevant features to our clinical outcome e.g. treatment response within each dataset, followed by an assessment of interrelations between datasets, roughly, this would involve dimensionality reduction for each type of omics-data to reduce the complexity of high-throughput data, e.g. into functional modules, followed by filtering of features within each dataset using the relation of each module with our clinical outcome (e.g. treatment response) to include only features of high relevance to our outcome. At last, the interrelation between features across datasets will be assessed. The importance of each single feature can then be evaluated by performing leave-one-out analyses. This approach has been described by Pedersen et al. in a Nature Protocol article (PMID: 30382244). The final statistical analysis plan (SAP) for this study is still in the planning phase, therefore, methods for these analyses have not been described in detail in the present manuscript. We will seek professional bioinformatical assistance both in the planning phase and the conduct of the analyses in this study and the SAP will be completed prior to completion of data collection.

VERSION 2 – REVIEW

REVIEWER	Nurulamin Noor

	Addenbrooke's Hospital, Cambridge University Hospitals NHS Trust, Cambridge, United Kingdom
REVIEW RETURNED	23-Jan-2020

GENERAL COMMENTS	All queries and methodological points satisfactorily answered in responses and from amendments to the manuscript. I wish the authors the best of luck for this important and interesting project.

REVIEWER	Ho-Su LEE KU Leuven, Belgium
REVIEW RETURNED	31-Jan-2020

GENERAL COMMENTS	The authors have addressed my comments.