Cell Systems Transparent Peer Review Record

Pre-infection antiviral innate immunity contributes to sex differences in SARS-CoV-2 infection

Natalie Sauerwald, Zijun Zhang, Irene Ramos, Venugopalan D. Nair, Alessandra Soares Schanoski, Yongchao Ge, Weiguang Mao, Hala Alshammary, Ana S. Gonzalez-Reiche, Adriana van de Guchte, Carl W. Goforth, Rhonda A. Lizewski, Stephen E. Lizewski , Mary Anne S Amper, Mital Vasoya, Nitish Seenarine, Kristy Guevara, Nada Marjanovic, Clare M. Miller, German Nudelman, Megan A. Schilling, Rachel S. G. Sealfon, Michael S. Termini, Sindhu Vangeti, Dawn L. Weir, Elena Zaslavsky, Maria Chikina, Ying Nian Wu, Harm Van Bakel, Andrew G. Letizia, Stuart C. Sealfon, and Olga G. Troyanskaya

Summary

Initial Submission: Received April 11, 2022

Scientific editor: Bernadett Gaal, DPhil

First round of review: Number of reviewers: 2
2 confidential, 0 signed
Revision invited June 23, 2022
Minor changes anticipated
Revision received July 21, 2022

Second round of review: Number of reviewers: 2
2 original, 0 new
2 confidential, 0 signed
Accepted Oct 18, 2022

Cell Systems Transparent Peer Review Record

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

Dear Dr. Sauerwald,

I hope this email finds you well. The reviews are back on your manuscript and I've appended them below. You'll see that the reviewers find the manuscript compelling and their comments are intended to strengthen an already strong piece of work. We're happy to invite a revision.

In this case addressing Reviewer 2's comments with respect to the analysis seems to warrant special attention. If you have any questions or concerns about the revision, I'd be happy to talk about them, either over email or over Zoom. 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,

Bernadett

Bernadett Gaal, DPhil
Editor-in-Chief, *Cell Systems*

Reviewers' comments:

Reviewer #1: In this manuscript, Sauerwald and colleagues examine sex differences in the immune response to SARS-CoV-2, by studying a young adult cohort of marine recruits using longitudinal analysis before, during and after SARS-CoV2 infection (CHARM study). The analysis showed significant sex differences in several key metrics of infection, including symptoms, initial viral load as well as RNA and protein molecular signatures. Notably, the female patients had a higher baseline expression of anti-viral interferon stimulated genes (ISG), suggesting that increased anti-viral innate immunity may contribute to the sex bias. Very little is known regarding the causal factors driving the established sex bias in SARS-CoV2 infection and COVID-19 disease, and this report sheds new light on this phenomenon which is significant. The manuscript is well written and organized. The use of longitudinal analysis in the patient cohort is a strength. Some weaknesses exist in the study, including lack of some key experimental details and protein validation data for differences in ISG and CD45 in females and males. Overall, this is an interesting study.

Major points
What was the vaccination status of the participants in the study, and how was this status factored in as a covariate in the analysis? Similarly, were antivirals or additional drug treatments administered to the study participants?

Were sex specific changes in CD45 RA/CD45RO accordingly observed in the T or B cell populations?

Similarly, are corresponding changes in ISG proteins observed in male and female samples?

How were the 92 proteins selected for the O-link analysis in Figure 2? Did any of these demonstrate a sex bias for the corresponding mRNA in the RNAseq analysis?

Details regarding how the RNAseq analysis performed are scant. Was RNA isolated from nasal swabs, or from total PBMC? What is the justification for the choice between circulating and nasal molecular signatures?

One of the main weaknesses of this study is the lack of association of male or female immune bias with long term outcomes, largely due to the fact that most of these young (and potentially vaccinated) adult patients showed mild or asymptomatic disease. Can the authors comment on any sort of outcome metrics, in terms of time to total resolution of disease and other outcomes (long covid)?

Reviewer #2: General comments: Sauerwald and coauthors presented an interesting study of comprehensive molecular profiling and systems biology analysis to dissect the sex differences in immune responses to SARS-CoV-2 infection. This study has several considerable strengths, including longitudinal monitoring of (1) SARS-CoV-2 infection by PCR testing, (2) symptoms by questionnaires, and (3) serums and molecular profiling (RNAseq and protein markers) at different stages of infection, as well as a large, relatively homogeneous young healthy individuals of both males and females. This work would be a timely contribution the literature and our understanding of sex difference in immune response to SARS-CoV-2 infection. I have some specific comments as follows that might help the authors clarify some key points.

Major comments:
1. Based on Figure 4, it's not clear as to whether sex or sex-associated pre-infection immune profile (latent factor) was the independent variable in mediation analysis. Under "Mediation analysis" in Methods, it seems that sex was the independent variable. Please clarify and improve the presentation of Figure 4.
2. Was the nominal or FDR-adjusted ACME p-value used for statistical significance in the mediation analysis? Given a number of mediation analysis were performed, multiple hypothesis testing should be taken into account.
3. The authors should discuss the caveats of causal mediation analysis in order to make "causal" inference. See, for example, VanderWeele (2016) PMID: 26653405, for discussion on assumptions made in causal mediation analysis, in particular, regarding various types of measured and unmeasured confounders. Along this line, would self-reported race be a confounder in the current context? If so, it needs to be properly adjusted.
4. The authors seemed to perform mediation analysis for one mediator at a time. Given that many of the mediators under consideration could be correlated, it might be worth considering multiple-mediator analysis to adjust for each other and quantify the total mediation effect.

Minor comments:
1. References 8 and 9 under "Differential gene expression analysis" in Methods seem mis-aligned with the reference list.

Cell Systems Transparent Peer Review Record

Authors' response to the reviewers' first round comments

Attached.

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

Dear Dr. Sauerwald,

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. 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.

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,

Bernadett

Bernadett Gaal, DPhil
Editor-in-Chief, Cell Systems

Editorial Notes

Transparent Peer Review: Thank you for electing to make your manuscript's peer review process transparent. As part of our approach to Transparent Peer Review, we ask that you add the following sentence to the end of your abstract: "A record of this paper's Transparent Peer Review process is included in the Supplemental Information." Note that this ***doesn't*** count towards your 150 word total!

Also, if you've deposited your work on a preprint server, that's great! Please drop me a quick email with your preprint's DOI and I'll make sure it's properly credited within your Transparent Peer Review record.

Figures and Legends:

Please look over your figures keeping the following in mind:

- When data visualization tools are used (e.g. UMAP, tSNE, etc.), please ensure that the dataset being visualized is named in the figure legend and, when applicable, its accession number is included.
- When color scales are used, please define them, noting units or indicating "arbitrary units," and specify whether the scale is linear or log.
- Bar graphs are not acceptable because they obscure important information about the distributions of the underlying data. Please display individual points within your graphs unless their large number obscures the graph's interpretation. In that case, box-and-whisker plots are a good alternative.
- Please ensure that every time you have used a graph, you have defined "n's" specifically and listed statistical tests within your figure legend.
- Please ensure that all figures included in your point-by-point response to the reviewers' comments are present within the final version of the paper, either within the main text or within the Supplemental Information.

STAR Methods: Note that Cell Press has recently changed the way it approaches "availability" statements for the sake of ease and clarity. Please revise the first section of your STAR Methods as follows, noting that the particular examples used might not pertain to your study. Please consult the STAR Methods guidelines for additional information.

> RESOURCE AVAILABILITY
>
> **Lead Contact:** Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Jane Doe (janedoe@qwerty.com).
>
> **Materials Availability:** This study did not generate new materials. *-OR-* Plasmids generated in this study have been deposited at [Addgene, name and catalog number]. *-OR-* etc.
>
> **Data and Code Availability:**
>
> - **Source data statement** (described below)
> - **Code statement** (described below)
> - Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Data and Code Availability statements **have three parts and each part must be present. Each part should be listed as a bullet point, as indicated above.**

Instructions for section 1: Data. The statements below may be used in any number or combination, but at least one must be present. They can be edited to suit your circumstance. Please ensure that all datatypes reported in your paper are represented in section 1. For more information, please consult this list of standardized datatypes and repositories recommended by Cell Press.

- [Standardized datatype] data have been deposited at [datatype-specific repository] and are publicly available as of the date of publication. Accession numbers are listed in the key resources table.

- [Adjective] data have been deposited at [general-purpose repository] and are publicly available as of the date of publication. DOIs are listed in the key resources table.

- [De-identified human/patient standardized datatype] data have been deposited at [datatype-specific repository]. They are publicly available as of the date of publication until [date or delete "until"]. Accession numbers are listed in the key resources table.

- [De-identified human/patient standardized datatype] data have been deposited at [datatype-specific repository], and accession numbers are listed in the key resources table. They are available upon request until [date or delete "until"] if access is granted. To request access, contact [insert name of governing body and instructions for requesting access]. [Insert the following when applicable] In addition, [summary statistics describing these data/processed datasets derived from these data] have been deposited at [datatype-specific repository] and are publicly available as of the date of publication. These accession numbers are also listed in the key resources table.

- Raw [standardized datatype] data derived from human samples have been deposited at [datatype-specific repository], and accession numbers are listed in the key resources table. Local law prohibits depositing raw [standardized datatype] datasets derived from human samples outside of the country of origin. Prior to publication, the authors officially requested that the raw [adjective] datasets reported in this paper be made publicly accessible. To request access, contact [insert name of governing body and instructions for requesting access]. [Insert the following when applicable] In addition, [summary statistics describing these data/processed datasets derived from these data] have been deposited at [datatype-specific repository] and are publicly available as of the date of publication. These accession numbers are also listed in the key resources table.

- The [adjective] data reported in this study cannot be deposited in a public repository because [reason]. To request access, contact [insert name of governing body and instructions for requesting access]. [Insert the following when applicable] In addition, [summary statistics describing these data/processed datasets derived from these data] have been deposited at [datatype-specific or general-purpose repository] and are publicly available as of the date of publication. [Accession numbers or DOIs] are listed in the key resources table.

• This paper analyzes existing, publicly available data. These accession numbers for the datasets are listed in the key resources table.

• [Adjective or all] data reported in this paper will be shared by the lead contact upon request.

Instructions for section 2: Code. The statements below may be used in any number or combination, but at least one must be present. They can be edited to suit your circumstance. ***If you are using GitHub, please follow the instructions here to archive a "version of record" of your GitHub repo at Zenodo, then report the resulting DOI. Additionally, please note that the Cell Systems strongly recommends that you also include an explicit reference to any scripts you may have used throughout your analysis or to generate your figures within section 2.***

• All original code has been deposited at [repository] and is publicly available as of the date of publication. DOIs are listed in the key resources table.

• All original code is available in this paper's supplemental information.

• This paper does not report original code.

Instructions for section 3. Section 3 consists of the following statement: Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

In addition,

STAR Methods follows a standardized structure and must be included within the main manuscript. Please reorganize your experimental procedures to include these specific headings in the following order: LEAD CONTACT AND MATERIALS AVAILABILITY (including the three statements detailed above); EXPERIMENTAL MODEL AND SUBJECT DETAILS (when appropriate); METHOD DETAILS (required); QUANTIFICATION AND STATISTICAL ANALYSIS (when appropriate); ADDITIONAL RESOURCES (when appropriate). We're happy to be flexible about how each section is organized and encourage useful subheadings, but the required sections need to be there, with their headings. They should also be in the order listed. Please see the STAR Methods guide for more information or contact me for help.

Please ensure that original code has been archived in a general purpose repository recommended by Cell Press and that its DOI is provided in the Software and Algorithms section of the Key Resources Table. If you've chosen to use GitHub, please follow the instructions here to archive a "version of record" of your GitHub repo at Zenodo, complete with a DOI. Thank you!

Currently, you don't have a **Key Resources Table** (KRT). Note that the key resources table is required for manuscripts with an experimental component, and if a purely computational manuscript links to any external datasets (previously published or new), code-containing websites (e.g. a GitHub repo, noting that DOIs are strongly preferred), or uses non-standard software, it needs to include a key resources table that details these aspects of the paper. Purely computational or theoretical papers that don't contain any external links and use standard software don't require a key resources table,

although you're welcome to include one if you like. For details, please refer to the Table Template or feel free to ask me for help.

Thank you!

Reviewer comments:

Reviewer #1: In the revised manuscript submitted by Sauerwald and colleagues, the authors have taken pains to address the comments and concerns raised during the prior review. This includes clarifying the SARS-CoV-2 vaccination status of the patient cohort, adding additional experimental details for the RNASeq analysis, detailing the variables in the statistical analysis (mediation analysis), and clarifying the FDR correction used to assess ISG in the study. Mainly, they have provided a coherent rebuttal to some of the main limitations of the study, which satisfies this Reviewer. The central observation is rigorous and sound.

Reviewer #2: This reviewer thanks the authors for addressing all the comments.

Manuscript Number: CELL-SYSTEMS-D-22-00142
"Pre-infection antiviral innate immunity contributes to sex differences in SARS-CoV-2 infection"

Response to Reviewers

Reviewer #1: In this manuscript, Sauerwald and colleagues examine sex differences in the immune response to SARS-CoV-2, by studying a young adult cohort of marine recruits using longitudinal analysis before, during and after SARS-CoV2 infection (CHARM study). The analysis showed significant sex differences in several key metrics of infection, including symptoms, initial viral load as well as RNA and protein molecular signatures. Notably, the female patients had a higher baseline expression of anti-viral interferon stimulated genes (ISG), suggesting that increased anti-viral innate immunity may contribute to the sex bias. Very little is known regarding the causal factors driving the established sex bias in SARS-CoV2 infection and COVID-19 disease, and this report sheds new light on this phenomenon which is significant. The manuscript is well written and organized. The use of longitudinal analysis in the patient cohort is a strength. Some weaknesses exist in the study, including lack of some key experimental details and protein validation data for differences in ISG and CD45 in females and males. Overall, this is an interesting study.

Major points
What was the vaccination status of the participants in the study, and how was this status factored in as a covariate in the analysis? Similarly, were antivirals or additional drug treatments administered to the study participants?

> Thank you for pointing out these important clarifications. This observed study was conducted prior to FDA approved vaccines and treatments specifically directed at SARS-CoV-2 and none of these participants were involved in any clinical trials at the time. All symptomatic participants were treated as outpatients and none received any type of medication beyond symptomatic treatment such as non-steroidal anti-inflammatory drugs or acetaminophen. This has been noted in the text as well.

Were sex specific changes in CD45 RA/CD45RO accordingly observed in the T or B cell populations? Similarly, are corresponding changes in ISG proteins observed in male and female samples?

> Unfortunately we only have bulk data available to us in this study, so we are unable to validate the CD45 isoform changes in specific cell populations. Only six of the ISGs were measured at a protein level in the O-link data (CCL19, CCL4, CX3CL1, CXCL10, CXCL11, CXCL9), and none of these six showed significant sex differences either in gene expression or protein levels.

How were the 92 proteins selected for the O-link analysis in Figure 2? Did any of these demonstrate a sex bias for the corresponding mRNA in the RNAseq analysis?

> The O-link proteins were not individually selected - we used the protein biomarker inflammation panel from the O-link company. Due to the small number of O-link samples from female participants, we did not rely heavily on the sex differences identified from these data but focused on the RNA-seq which contained sufficient samples for statistical analyses.

Details regarding how the RNAseq analysis performed are scant. Was RNA isolated from nasal swabs, or from total PBMC? What is the justification for the choice between circulating and nasal molecular signatures?

> The RNA-seq data was sequenced from whole blood samples; the nasal swabs were used only for the SARS-CoV-2 qPCR testing. Blood was targeted for study because of interest in characterizing the systemic host response to SARS-CoV-2 infection as opposed to the responses to the effects of local infection which would be reflected in the nasal swab.

One of the main weaknesses of this study is the lack of association of male or female immune bias with long term outcomes, largely due to the fact that most of these young (and potentially vaccinated) adult patients showed mild or asymptomatic disease. Can the authors comment on any sort of outcome metrics, in terms of time to total resolution of disease and other outcomes (long covid)?

> Due to the nature of this study, involving very young and very healthy adults with times between consecutive sample collections of about 7 days, we are unable to draw conclusions on longer-term outcomes. The majority of cases resolved within a week, making it impossible to observe any differences in time to disease resolution with samples taken 7 days apart. We did explore this, but unfortunately we do not have the resolution to identify any differences in disease time course, and none of the participants in our cohort experienced long COVID symptoms.

Reviewer #2: General comments: Sauerwald and coauthors presented an interesting study of comprehensive molecular profiling and systems biology analysis to dissect the sex differences in immune responses to SARS-CoV-2 infection. This study has several considerable strengths, including longitudinal monitoring of (1) SARS-CoV-2 infection by PCR testing, (2) symptoms by questionnaires, and (3) serums and molecular profiling (RNAseq and protein markers) at different stages of infection, as well as a large, relatively homogeneous young healthy individuals of both males and females. This work would be a timely contribution the literature and our understanding of sex difference in immune response to SARS-CoV-2 infection. I have some specific comments as follows that might help the authors clarify some key points.

Major comments:
1. Based on Figure 4, it's not clear as to whether sex or sex-associated pre-infection immune profile (latent factor) was the independent variable in mediation analysis. Under "Mediation analysis" in Methods, it seems that sex was the independent variable. Please clarify and improve the presentation of Figure 4.

> Thank you for pointing this out. We have added a diagram in the supplement (Figure S5) as well as additional text to clarify the variables in the mediation analysis. We have also added a panel to Figure 4 (Fig 4A) quantifying the mediation effects of all relationships tested in the causal mediation analysis to aid clarity of this section.

2. Was the nominal or FDR-adjusted ACME p-value used for statistical significance in the mediation analysis? Given a number of mediation analysis were performed, multiple hypothesis testing should be taken into account.

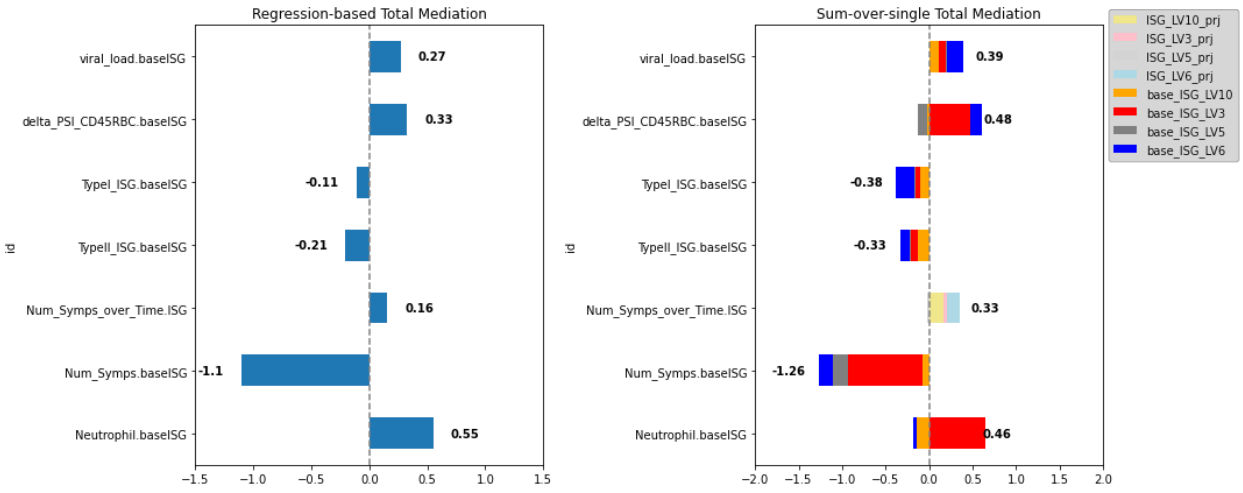
> The mediation analysis was performed using only the 4 latent variables identified from the full set of ISGs, resulting in only 32 total hypotheses tested. Multiple hypothesis correction has been included in the results (Supplementary Table S5) with a column for FDR values. The majority of significant relationships were not changed under the correction, due to the relatively few hypotheses tested.

3. The authors should discuss the caveats of causal mediation analysis in order to make "causal" inference. See, for example, VanderWeele (2016) PMID: 26653405, for discussion on assumptions made in causal mediation analysis, in particular, regarding various types of measured and unmeasured confounders. Along this line, would self-reported race be a confounder in the current context? If so, it needs to be properly adjusted.

> We did not observe clinical differences based on ethnicity or race, but there were some significant differences observed in ISG latent variables across racial and ethnic groups (although the differences were much smaller than those observed across sexes). Both race and ethnicity have been included as covariates in the causal mediation analysis and full results have been updated, but no significant qualitative changes were observed due to inclusion of these confounders. We have also added some text to address the caveats of causal mediation analysis to the discussion, as proposed.

4. The authors seemed to perform mediation analysis for one mediator at a time. Given that many of the mediators under consideration could be correlated, it might be worth considering multiple-mediator analysis to adjust for each other and quantify the total mediation effect.

> We thank the reviewer for this suggestion. We have computed the total mediation effects for the mediators, with results shown below. This analysis supports the findings of significant mediation through neutrophil-associated ISGs in particular. The total mediation effects were largely consistent with those computed by single mediator analysis, so we keep the single mediator analysis in the main text as it allows us to quantify the effects of each individual mediator with corresponding statistical testing.



Minor comments:

1. References 8 and 9 under "Differential gene expression analysis" in Methods seem mis-aligned with the reference list.

> Thank you for pointing this out, it has been fixed.