nature portfolio

Peer Review File

Proteomic profiling of neonatal herpes simplex virus infection on dried blood spots

Corresponding Author: Dr Kia Dungu

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
Kia Hee Schultz Dungu et al. evaluated the use of proteomic profiling on dried blood spots for screening neonatal herpes simplex virus infection. The identified distinct protein profiles in neonates with disseminated HSV-1 disease compared to controls.
The strategy is interesting but unfortunately the study is difficult to evaluate as presented. Among other issues, the description of the methods is insufficient.

Specific comments:

1. The proposed diagnostic test is indirect as protein levels are measured and concluded to be caused by the HSV infection. What would the protein profile look like in another active virus infection?

2. The Method section has to be improved concerning several different points:
-How many proteins were measured with the Olink Explore 3072?
-In what volume of PBS-T are the DBS resuspended?
-Is there a control that similar amounts of protein are extracted or is there a normalisation of the protein content of each sample? If so how was this implemented?
-How is the relative quantification calculated?
-Is there a reference or a random plate design?
-How many outliners were removed upon consultation of the PCA?
-QC of the data should be presented in supplementary Figures (PCA, NPX-Distribution plots, ...).
-What are the criteria for significant different expression (maximal FDR, minimal NPX difference) for Figure 1 and 2?
-What does the statement There is no significant association between the variables used for case-control matching and NPX levels. Thus, they were excluded from downstream analysis. mean?

3. The number of control samples is about twice that of the disease samples. When comparing the three forms of HSV disease, the ratio of controls and disease is even more extreme. Different sample sizes could influence the performance of ANOVA.

4. The scales in Figure 1 A to C are not the same. Comparison of the data would be easier if the same scale was applied. The position of the dashed lines should be indicated in the legend. Also, what was the basis for the positioning of the dashed line?

5. 23 significantly different proteins are described in the text but only 20 proteins are shown in the figures. Is it because IL6 has been counted 4 times (Suppl. Fig. 1)?

6. Boxplots or other visualisation of the NPX values would be more informative than the heat map presented in Figure 3. What is the scale shown on the right? Also, as shown in the heat map, the control actually clusters to SEM, the control and SEM cluster to the CNS, and these three conditions cluster to DIS. This finding should be explained.

7. Some of the points in Figure 4 and Supplement Figure 2 are almost invisible. According to the legend, some of the pathways are presented by only four proteins, and in the Supplement by only a single protein. Are these pathways indeed significantly enriched?

8. What database (GO terms, Reactome, KEGG, ...) for the pathway analysis in pathfindeR was used?

9. The Discussion is lengthy and might be shortened. The point that the DBS samples are 'mixed of lysed cells and plasma' is important and should be addressed by an appropriate statistic to show that this is not a major influencing factor.

10. Please specify HSV-1 infection or HSV-2 infection or both

Reviewer #2

(Remarks to the Author)
The manuscript from Dungu and colleagues, entitled Proteomic profiling of neonatal herpes simplex virus infection on dried blood spots exploring screening perspectives, employs protein biomarker panels from Olink containing 3072 proteins to screen and perform quantitative analysis of these proteins extracted from dried blood spots. Using statistical analyses, the authors identify 23 proteins that are differentially associated with disseminated HSV disease compared to controls. This appeared fairly specific to this condition as only 3 were shared with skin-eye-mouth disease cases and none with CNS disease cases. Overall the authors goal to identify biomarkers of neonatal HSV is an unmet need and has notable clinical relevance. However, I question whether the manuscript represents a significant advancement. In part, this may be because the novelty of the study has not been fully communicated. Also, the manuscript lacks some technical details, making it hard to evaluate the experimental merits. Please see my comments below for more details.
Primary points
1) The Introduction mentioned omics technologies, but this section could be expanded to put the current study in appropriate context. For example, a representative study from host proteomes from DBS samples (refs 18-21) could be selected to write a few additional sentences that explain the main technique used and the main finding. Also, any transcriptomics studies relevant to neonatal HSV should be included in this section, such as work from Cohen et al. 2021 (https://doi.org/10.1093/ofid/ofab466.082).
2) I did not realize when I read through the manuscript that the approach used was OLink. It is important to describe in the Introduction or Results that the multiplexed proteomic method used was the OLink 3072 panel, including a a few sentences about how this technology works. It will be useful to state explicitly if the whole panel of 3072 proteins was used. And if so, state in the results how many of the proteins were quantified from DBS. I would also suggest including an analysis workflow that would include OLink.
3) Related to OLink, a spreadsheet formatted table should be reported that includes the individual quantitative values for each subject and proteins quantified.
4) I am confused how to evaluate the results the authors reported that were significant for the ANOVA test, for example n=134 for DIS vs Control, versus the subset significant only after the post-hoc test. I understand that the post-hoc significant are more confidence, but is the confidence of ANOVA only low enough that they should not be considered? Some commentary, perhaps in the Methods, on the rationale for reporting both initial set and post-hoc set would be useful.
5) The proteins analyzed by the hierarchical clustering (Figure 3) were already pre-selected as significant for the DIS vs Control, and from Fig 2, these were largely not significant in the other comparisons. Therefore, the conclusion that they clustered by condition is not surprising. Could the authors provide any additional interpretation of this result?
6) The OLink panels contain a fraction of the annotated human proteome. While this may not be an issue for the goal of finding biologically meaningful signatures, it does introduce a bias in the data. Specifically, did the authors take this into consideration when performing their Pathway enrichment? Many enrichment tools use the whole genomes as the background comparison be default. For this study, using the OLink panel, the whole genome background would not be appropriate. Moreover, generally, pathway enrichment has poor sensitivity when used the input contains a low # of genes/proteins. The authors may want to consider a simpler approach that shows assignment of proteins to their annotated pathways (independent of their enrichment).
7) The known physical and/or functional relationships between the differential proteins is not clear. I would suggest the authors perform a network analysis, e.g., using STRING or similar approach.

Minor points
1) SEM was not mentioned in the Introduction and only appeared in Results as “SEM”. This should be defined and explained in the Introduction.
2) Generally, the size of the text for the axis labels in Figure 1 and 4, and similar graphs in the Supplement, could be made larger for improved readability.
3) Please include the average NPX, associated errors, and p-values in the Supplemental Tables 2-4.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The manuscript by Kia Hee Schultz Dungu et al. is describing the use of proteomic profiling on dried blood spots for screening neonatal herpes simplex virus infection. The manuscript was altered according to the reviewers comments. In particular, the method part of the manuscript is improved. Furthermore, the heat-map was replaced by a box-plot and the biological interpretation of the differently expressed proteins is explored by using DAVID instead of Pathway enrichment analysis. Some additional supplementary figures are included to support the authors findings.

However, some minor points should be addressed:

- DAVID assesses the overrepresentation of differently expressed proteins (or genes) in defined sets of proteins (or genes). These sets could be GO-terms (molecular function MF, biological pathway BP), KEGG-pathways, Reactome-pathways and UniProt KeyWord sets. Two overrepresented sets found by the analyses were UP_KW-0202~Cytokine and UP_KW-0802~TPR. These are both not per se pathways but rather Key Word sets. This should be marked in the text and the wording changed accordingly.

- The PCAs are shown in Supp. Fig. 1. The presentation of these analyses is appreciated. However, two questions arise: What are the red points on the plots? Are these controls? Please explain in figure caption. It seems as the PCA of "Inflammation II" has another distribution of the red points and PC1 explains 83.2% of the variance, while PC1 of the others only ~20%. Please explain.

Reviewer #2

(Remarks to the Author)
The authors have acknowledged and addressed all my points of concern. The authors have now included sufficient detail to fully evaluate the study. While the technique of O-Link and the sample population size introduces certain caveats in the study interpretation, the authors revised manuscript is sufficiently transparent about these points and uses appropriate language to prevent misinterpretation of the study's relevance and broad impact. Also, I recognize that the nature of the samples being collected and analyzed represents challenges. Therefore, I feel the revised manuscript is in an appropriate state for publication.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.
In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.
The images or other third party material in this Peer Review File are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.
To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/

COPENHAGEN UNIVERSITY HOSPITAL RIGSHOSPITALET

Department of Pediatrics and Adolescent Medicine

Blegdamsvej 9
Copenhagen, 2100 Ø
Denmark

+ 45 35 45 9761
Date: September 1, 2024

REBUTTAL LETTER TO THE REVIEWERS

Manuscript reference number: COMMSMED-23-0926-A

Title: Proteomic profiling of neonatal herpes simplex virus infection on dried blood spots: exploring screening perspectives.

Reviewer 1:
Kia Hee Schultz Dungu et al. evaluated the use of proteomic profiling on dried blood spots for screening neonatal herpes simplex virus infection. The identified distinct protein profiles in neonates with disseminated HSV-1 disease compared to controls. The strategy is interesting but unfortunately the study is difficult to evaluate as presented. Among other issues, the description of the methods is insufficient.

Specific comments:

1. **The proposed diagnostic test is indirect as protein levels are measured and concluded to be caused by the HSV infection. What would the protein profile look like in another active virus infection?**

Response: Thank you for the opportunity to address this important issue. We acknowledge that the protein profile may not be specific to HSV infection. Validation and testing in larger cohorts are indeed necessary before concluding this profile as a diagnostic test. To emphasize this limitation, we have revised the discussion.

Discussion:

reliable protein levels from DBS samples using the Olink Assays (19). Additionally, we acknowledge that the protein profile identified may not be specific to disseminated HSV infection. Validation and testing in larger cohorts are necessary to confirm the diagnostic utility of the identified proteins. Last, our protein

2. **The Method section has to be improved concerning several different points:**

Response: Thank you for the comments. We appreciate the feedback and have revised accordingly.

a) **How many proteins were measured with the Olink Explore 3072?**

Response: The number of proteins measured has been added to the manuscript.

Methods:

with 0.05% tween and protease inhibitor. Protein levels were estimated using the Olink Explore 3072 panels (Olink Proteomics Assays, Uppsala, Sweden), which utilise proximity extension assay (PEA) technology to measure 2941 proteins. In PEA, oligonucleotide-labelled antibodies bind to target proteins,

Results:

We successfully quantified all 2941 proteins included in the Olink Explore 3072 panels from the DBS samples. QC of the variables used for matching and principal component analysis (PCA) analyses

b) **In what volume of PBS-T are the DBS resuspended?**

Methods:

Proteins were extracted from DBS samples according to Olink's recommendations using 40 µl PBS buffer with 0.05% tween and protease inhibitor. Protein levels were estimated using the Olink Explore 3072

c) Is there a control that similar amounts of protein are extracted or is there a normalisation of the protein content of each sample? If so how was this implemented?

Methods:

quantified by sequencing (30). The DBS samples were analysed undiluted and non-normalized, following the Olink protocol. Each run included three external controls: the plate control for data normalization, the sample control to assess potential variation between runs and plates, and the negative control to determine the limit of detection for each assay and to assess potential contamination of assays (31). We employed a

Reference added:

- **Reference number 31**: Olink. Olink data normalization and standardization [Internet]. Available from: https://www.olink.com/content/uploads/2021/09/olink-data-normalization-white-paper-v2.0.pdf

c) **How is the relative quantification calculated?**

Response: The relative quantification calculation is elucidated in the reference "Olink data normalization and standardization white paper (reference 31), which we have now included in the manuscript.

e) Is there a reference or a random plate design?

Response: We employed a triplet design (case + control + control) on the same plate but randomized positions with 8 positions apart from each other.

Methods:

the limit of detection for each assay and to assess potential contamination of assays (31). We employed a triplet design (case + control + control) on the same plate, with samples randomized to positions 8 wells

f) How many outliners were removed upon consultation of the PCA?

Response: Sixteen outliers were removed after reviewing the PCA.

Methods:

Analyze R package (version 3.4.1). Samples with quality control (QC) and Assay Warnings were excluded. Samples within +/- 3 standard deviations (SD) in Principal Components 1 and 2 calculated from Panel-Assay wise NPX values were considered valid, while outliers (n=16) were removed. ~~There~~

g) QC of the data should be presented in supplementary Figures (PCA, NPX-Distribution plots, ...)

Response: PCA plots have been added as a supplemental Figure 1.

Results:

samples. QC of the variables used for matching and principal component analysis (PCA) analyses revealed no differences between cases and controls (Supplemental Figure 1). The permutation test showed

h) What are the criteria for significant different expression (maximal FDR, minimal NPX difference) for Figure 1 and 2?

Response: Since the study was exploratory in nature, we did not predefine a specific mean NPX difference threshold for determining significance. Instead, we employed rigorous statistical methods to identify differentially expressed proteins, including the use of FDR-adjusted p-values and post hoc testing. The threshold for significance was set at an FDR-adjusted p-value of <0.05. We have revised the methods and former Figure 1+2 legends to clarify these points. Due to revisions in the number of figures, previous Figures 1+2 are now Figures 2+3.

Methods:

Neonatal Screening, Department of Congenital Disorders, Statens Serum Institut. Since the study was exploratory in nature, a specific mean NPX difference threshold was not predefined. The threshold for significance was set at an FDR-adjusted p-value of <0.05. Considering the limited sample size, we opted

Figure 2 (former Figure 1), legend:

Volcano plots depicting protein differences in neonates with the neonatal HSV phenotypes compared with controls; A) disseminated disease, B) CNS disease, and C) Skin-eye-mouth disease. The x-axis represents the normalized protein expression (NPX) mean difference, and the y-axis represents the log10 p-value. The horizontal dashed line indicates the significance threshold (false discovery rate adjusted p-value <0.05). The vertical dashed lines indicate the maximum (controls) and minimum (cases) NPX mean difference at significant false discovery rate adjusted p-values. The labelled proteins were significant in the post-hoc analyses when comparing individual HSV phenotypes with controls.

Figure 3 (former Figure 2), legend:

Venn diagram illustrating the significantly different proteins for the neonatal HSV infection phenotypes compared to controls. The labelled proteins were significant in the post-hoc analyses when comparing individual HSV phenotypes with controls. IFIT1, IFIT3 and CXCL10 were significant for both disseminated and skin-eye-mouth disease. There were no post-hoc significant proteins for CNS disease versus controls. The threshold for significance was set at a false discovery rate adjusted p-value of <0.05.

i) What does the statement There is no significant association between the variables used for case-control matching and NPX levels. Thus, they were excluded from downstream analysis. mean?

Response: Thank you. We apologize for any confusion and acknowledge that we have not effectively communicated this aspect of our methodology. Due to the few individuals in our study, we were restricted in our choice of statistical approach and consequently employed ANOVA without adjusting for the matching variables. Despite this limitation, we tested for associations between these matching variables and NPX values, ultimately deciding to exclude them in our analysis.

Methods:

significance was set at an FDR-adjusted p-value of <0.05. Considering the limited sample size, we opted for Analysis of Variance (ANOVA) as statistical model, which precluded the adjustment for matching variables. Since no statistically significant associations were found between the matching variables and NPX levels, we excluded them in our analysis. To identify differentially expressed proteins between cases

3. **The number of control samples is about twice that of the disease samples. When comparing the three forms of HSV disease, the ratio of controls and disease is even more extreme. Different sample sizes could influence the performance of ANOVA.**

Response: Thank you for emphasizing this important point. We acknowledge that the unequal distribution of samples, i.e., the higher number of controls compared to cases and the varying ratios across the three HSV phenotypes, could potentially influence the performance of ANOVA. Our rationale for including more controls than cases was to obtain a more robust mean NPX protein levels for the non-HSV infected population. Given that the study is exploratory, we accepted some level of statistical uncertainty. The results should be viewed primarily as indications of proteins that may be relevant for future biomarker development. We have revised the discussion to reflect these considerations.

Discussion:

assumptions. Second, the unequal distribution of samples, with more controls than cases and varying ratios across the HSV phenotypes, could influence the ANOVA performance. However, to increase the robustness of the controls' mean NPX levels, we decided to include more controls than cases. Since the study is exploratory, we accepted some level of statistical uncertainty, and the results should be viewed as preliminary indications of proteins that may be relevant for future biomarker development. Another

4. **The scales in Figure 1 A to C are not the same. Comparison of the data would be easier if the same scale was applied. The position of the dashed lines should be indicated in the legend. Also, what was the basis for the positioning of the dashed line?**

Response: Thank you for the comment. The scales in Figure 1 A-C have been aligned and the figure legend has been revised. Due to revisions in the number of figures, the numbering has been updated to Figure 2.

Figure 2 (former Figure 1), legend:

Volcano plots depicting protein differences in neonates with the neonatal HSV phenotypes compared with controls; A) disseminated disease, B) CNS disease, and C) Skin-eye-mouth disease. The x-axis represents the normalized protein expression (NPX) mean difference, and the y-axis represents the log10 p-value. The horizontal dashed line indicates the significance threshold (false discovery rate adjusted p-value <0.05). The vertical dashed lines indicate the maximum (controls) and minimum (cases) NPX mean difference at significant false discovery rate adjusted p-values. The labelled proteins were significant in the post-hoc analyses when comparing individual HSV phenotypes with controls.

5. **23 significantly different proteins are described in the text but only 20 proteins are shown in the figures. Is it because IL6 has been counted 4 times (Suppl. Fig. 1)**

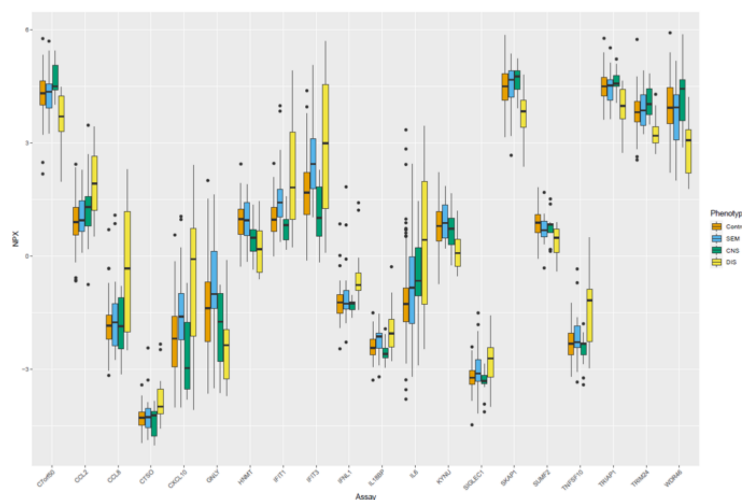
Response: Thank you for the comment. We apologize for the incorrect number of significantly different proteins. The correct number of proteins is indeed 20. We have revised throughout the manuscript.

6. **Boxplots or other visualisation of the NPX values would be more informative than the heat map presented in Figure 3. What is the scale shown on the right? Also, as shown in the heat map, the control actually clusters to SEM, the control and SEM cluster to the CNS, and these three conditions cluster to DIS. This finding should be explained.**

Response: Thank you for the comment. We agree that the heatmap was not providing the necessary clarity and information as intended. To address this, we have decided to present the data using boxplots. Due to revisions in the number of figures, the numbering has been updated to Figure 4.

Figure 4 (former Figure 3):

FIGURE 4~~3~~: Boxplot visualisation of NPX levels for post-hoc significant proteins. ~~Unsupervised hierarchical clustering analysis.~~



Boxplots illustrating the differences in NPX levels for the post-hoc significant proteins (n=20) among the neonatal HSV disease phenotypes and controls (IFIT1 IFIT3 and CXCL10 overlapping for disseminated and skin-eye-mouth disease versus controls). DIS=disseminated disease, CNS=central nervous system disease, SEM=skin-eye-mouth disease.

7. **Some of the points in Figure 4 and Supplement Figure 2 are almost invisible. According to the legend, some of the pathways are presented by only four proteins, and in the Supplement by only a single protein. Are these pathways indeed significantly enriched?**

Response: Thank you for the comment. As suggested by Reviewer 2 (comment number 6), we agree that pathway enrichment analysis may not be the most suitable approach. Therefore, we decided to go with a simpler approach by functional annotation using the Database for Annotation, Visualization, and Integrated Discovery (DAVID: https://david.ncifcrf.gov/home.jsp). The methods, results, and discussion sections along with Figure 4 and Supplemental Table 5 have been revised. Due to revisions in the number of figures, the numbering has been updated to Figure 5. The supplemental Figure 2 and Supplemental Tables 6-7 has been removed and the numbering of supplemental figures has been changed.

Methods:

To identify biological pathways associated with the clinical phenotypes, we performed functional annotation of the differentially expressed proteins using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) (32). We included only proteins that remained significant in the post-hoc analyses when comparing the individual neonatal HSV phenotypes with controls. For the background list in DAVID, we used the full list of proteins in the Olink Explore 3072 panels.~~pathway enrichment analysis using pathfindR in R (31). This approach involved comparing the proteins of interest to a pathway database and through an active subnetwork enrichment analysis. Significantly elevated and reduced proteins from the Olink analyses were included along with their mean NPX differences based on adjusted p-values < 0.05. We considered enriched pathways with a p-value < 0.05 (31).~~

Results:

Biological pathways associated with disease phenotypes

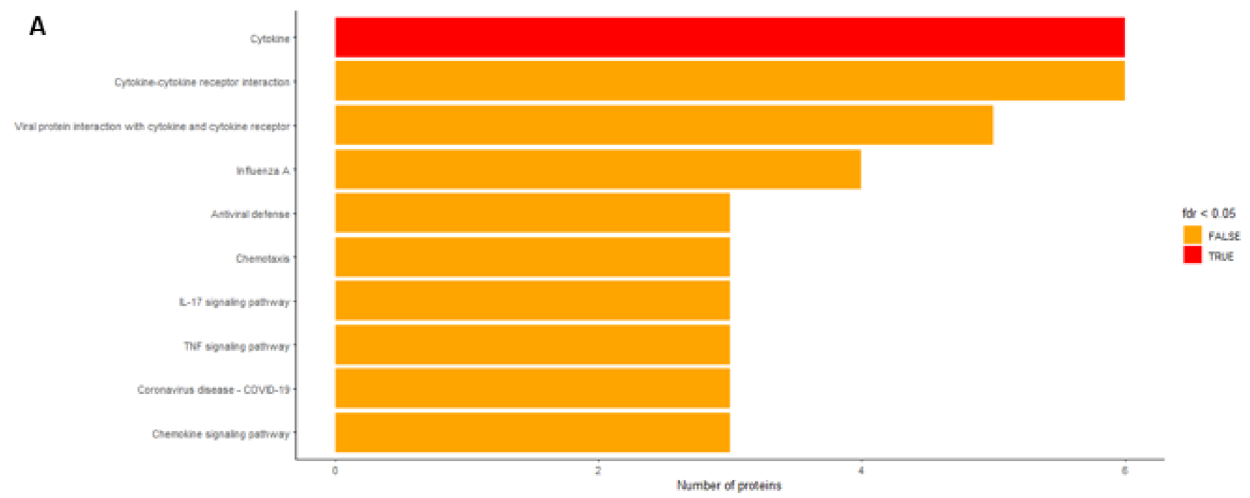
Functional annotation of the post-hoc significant proteins allowed us to identify biological pathways associated with the clinical phenotypes compared to controls (Figure 5A, Supplemental Table 5). For neonates with disseminated disease, we identified 33 pathways, but only the pathway “Cytokine” was statistically significant (FDR adjusted p-value < 0.05). Other relevant pathways, such as “Cytokine-cytokine receptor interaction”, “Viral protein interaction with cytokine and cytokine receptor” and “Chemotaxis”, were identified but did not reach statistical significance. In the comparison of SEM disease versus controls, the pathway “Tetratricopeptide Repeat was significant. Other relevant pathways, such as “Interferon alpha/beta signalling” and “Cytokine signalling in the immune system” were not significant. The “Antiviral defence” pathway appeared in both comparisons: disseminated disease versus controls and SEM disease versus controls. No pathways were identified for CNS disease versus controls, as there were no post-hoc significant proteins in this comparison (Supplemental Table 5).~~Pathway enrichment analyses~~

Discussion:

Our results suggest that the early stages of disseminated HSV disease are~~is~~ characteris~~z~~ed by a pronounced cytokine production, as evidenced by the significantly elevated levels of several~~numerous~~ inflammatory cytokines and ~~the enrichment of~~ annotated pathways associated with cytokine activation. ~~In addition, w~~We found indications of high activity in~~of~~ both the innate and adaptive immune systems, with elevated ~~levels of~~ markers of cytotoxic T cells and macrophages, along with ~~several~~ proteins and ~~enriched~~ pathways linked to antiviral ~~defense~~defence ~~and cell death~~. These findings ~~collectively~~ indicate a highly

Figure 5A:

FIGURE 5~~4~~: ~~Pathway enrichment analysis~~ Functional annotation and protein network in disseminated HSV disease versus controls.



A) Functional annotation of the proteins that remained significant when comparing disseminated HSV disease versus controls in the post-hoc analyses. For these proteins, one pathway “Cytokine” was statistically significant (false discovery rate adjusted p-value < 0.05). B) Protein network visualisation illustrating the relationships between the significant proteins in disseminated HSV disease compared to controls. Nodes represent individual proteins, while edges indicate different types of associations: co-occurrence (blue lines), co-expression (black lines), text mining (yellow lines), neighbourhood (green lines) and known interactions from curated databases (teal lines). The network was created using the STRING database (www.string-db.org).

Reference deleted:

- Ulgen E, Ozisik O, Sezerman OU. pathfindR: An R Package for Comprehensive Identification of Enriched Pathways in Omics Data Through Active Subnetworks. Front Genet. 2019;10:858.

Reference added:

- Huang DW, Sherman BT, Tan Q, Kir J, Liu D, Bryant D, et al. DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. Nucleic Acids Res. 2007 Jul;35(Web Server issue):W169-175.

8. **What database (GO terms, Reactome, KEGG, ...) for the pathway analysis in pathfindeR was used?**

Response: Thank you for the comment. As suggested by reviewer 2 (comment number 6), we agree that pathway enrichment analysis may not be the most suitable approach. Therefore, we decided to go with a simpler approach by functional annotation using the Database for Annotation, Visualization, and Integrated Discovery (DAVID: https://david.ncifcrf.gov/home.jsp). The point has been addressed in response to comment 7. The functional annotation results are provided in Supplemental Table 5 including the databases used.

9. **The Discussion is lengthy and might be shortened. The point that the DBS samples are 'mixed of lysed cells and plasma' is important and should be addressed by an appropriate statistic to show that this is not a major influencing factor.**

Response: Thank you for the comment. We acknowledge the concern regarding the use of DBS samples for proteomics due to the mix of lysed cells and plasma. This is an inherent condition when working with DBS samples, and it is unlikely to change in future studies. While we understand the importance of addressing this issue through appropriate statistical analysis, it is not feasible in our current study due to the nature of the samples. We have taken extensive measures to ensure the reliability and accuracy of our results. Furthermore, various sections of the discussion have been shortened while retaining the main points.

10. **Please specify HSV-1 infection or HSV-2 infection or both**

Response: Thank you for the comment. We have revised the methods to clarify, specifying that we included both HSV-1 and HSV-2 infection.

Methods:

This nationwide matched case-control study included all neonates aged 0-28 days with HSV-1 and HSV-2 infections from all hospitals with neonatal and paediatric departments in Denmark from 2010 to 2019.

Reviewer 2:

The manuscript from Dungu and colleagues, entitled Proteomic profiling of neonatal herpes simplex virus infection on dried blood spots exploring screening perspectives, employs protein biomarker panels from Olink containing 3072 proteins to screen and perform quantitative analysis of these proteins extracted from dried blood spots. Using statistical analyses, the authors identify 23 proteins that are differentially associated with disseminated HSV disease compared to controls. This appeared fairly specific to this condition as only 3 were shared with skin-eye-mouth disease cases and none with CNS disease cases. Overall the authors goal to identify biomarkers of neonatal HSV is an unmet need and has notable clinical relevance. However, I question whether the manuscript represents a significant advancement. In part, this may be because the novelty of the study has not been fully communicated. Also, the manuscript lacks some technical details, making it hard to evaluate the experimental merits. Please see my comments below for more details.

Primary points

1. The Introduction mentioned omics technologies, but this section could be expanded to put the current study in appropriate context. For example, a representative study from host proteomes from DBS samples (refs 18-21) could be selected to write a few additional sentences that explain the main technique used and the main finding. Also, any transcriptomics studies relevant to neonatal HSV should be included in this section, such as work from Cohen et al. 2021 (https://doi.org/10.1093/ofid/ofab466.082).

Response: Thank you for the comment. We agree and have revised the introduction as suggested. Furthermore, we have highlighted the paper by Cohen et al. in the discussion.

Introduction:

markers and disease pathophysiology (16,17). For decades, DBS samples have been used in omics methods, particularly in metabolomics and genomics (18). However, recent advancements have broadened their feasibility to include proteomics and transcriptomics (19–21). Although these approaches are not widely used in routine clinical practice, the accurate measurement of proteins in stored DBS samples using multiplex proximity extension assays (PEA) has been demonstrated (19). Additionally, host transcriptional profiles in whole blood samples from neonates with HSV infection, identifying an RNA signature comprising 1322 differentially expressed genes (22). ~~Previous studies have demonstrated~~

Discussion:

pathways linked to antiviral ~~defense~~defence ~~and cell death~~. These findings ~~collectively~~ indicate a highly activated ~~and potentially exaggerated~~ immune response in neonates with disseminated disease, aligning with the RNA signatures identified by Cohen et al. (22). Among the antiviral proteins identified, IFIT1

2. I did not realize when I read through the manuscript that the approach used was OLink. It is important to describe in the Introduction or Results that the multiplexed proteomic method used was the OLink 3072 panel, including a a few sentences about how this technology works. It will be useful to state explicitly if the whole panel of 3072 proteins was used. And if so, state in the results how many of the proteins were quantified from DBS. I would also suggest including an analysis workflow that would include OLink.

Response: Thank you for the comments. The abstract, introduction, methods and results have been revised. Furthermore, a study cohort overview and proteomics workflow has been provided in Figure 1.

Abstract:

most infections are recognisedpresent. We investigated high-throughputhigh throughput multiplexed proteomics on DBSdried blood spot samples from a nationwide cohort of neonates with HSV infection, using the Olink Explore 3072 panels and proximity extension assays to gain insight into screening and

Introduction:

This is the first study to investigate of high--throughput, multiplexed proteomic analysis on DBS samples obtained from neonates prior to developingwith HSV infection using the Olink Explore 3072 panels and PEA technology. Our aim was to gain insight into screening potential and disease pathophysiology perspectives.

Methods:

Proteins were extracted from DBS samples according to Olink's recommendations using 40 µl PBS buffer with 0.05% tween and protease inhibitor. Protein levels were estimated using the Olink Explore 3072 panels (Olink Proteomics Assays, Uppsala, Sweden), which utilise proximity extension assay (PEA) technology to measure 2941 proteins. In PEA, oligonucleotide-labelled antibodies bind to target proteins, allowing for the formation of a unique DNA sequence corresponding to the protein, which is then quantified by sequencing (30). The DBS samples were analysed undiluted and non-normalized, following the Olink protocol. Each run included three external controls: the plate control for data normalization, the

Reference added (30):

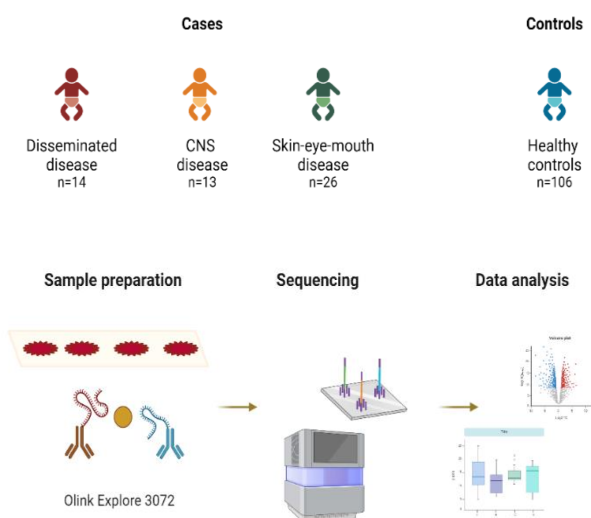
- Olink. Olink Explore Overview [Internet]. Available from: https://7074596.fs1.hubspotusercontent-na1.net/hubfs/7074596/01-User%20Manuals%20for%20website/1187-olink-explore-overview-user-manual.pdf

Results:

Quality control and data structure

We successfully quantified all 2941 proteins included in the Olink Explore 3072 panels from the DBS samples. QC of the variables used for matching and principal component analysis (PCA) analyses

Figure 1:



A) Summary of the study cohort, displaying the distribution of HSV phenotypes among cases versus controls. B) Laboratory and analytical workflow from preparation of the dried blood spot samples, proximity extension assay using Olink Explore 3072 panels and proximity extension assay technology, library preparation, high-throughput sequencing (NovaSeq 6000, Illumina), and data analysis (Olink Analyze R package) including quality control, normalization, statistical analysis, and functional annotation. Created with Biorender.com.

3. Related to OLink, a spreadsheet formatted table should be reported that includes the individual quantitative values for each subject and proteins quantified.

Response: Thank you for the comment. The individual quantitative values for each subject and proteins are provided in a separate Excel file (Supplemental Data 1). Additionally, F-statistics are included in a separate file (Supplemental Data 2).

Methods:

Data and code availability

Data are submitted (Supplemental Data 1 and Supplemental Data 2).and cCodes will be made public upon publication.

4. I am confused how to evaluate the results the authors reported that were significant for the ANOVA test, for example n=134 for DIS vs Control, versus the subset significant only after the post-hoc test. I understand that the post-hoc significant are more confidence, but is the confidence of ANOVA only low enough that they should not be considered? Some commentary, perhaps in the Methods, on the rationale for reporting both initial set and post-hoc set would be useful.

Response: Thank you for the comment. In our manuscript, we report the number of significant proteins for each HSV phenotype versus controls as identified by the ANOVA test to provide a comprehensive overview of potential differential expression. However, we place greater emphasis on the subset of proteins with post-hoc statistical significance. The method section has been revised to elaborate on this.

Methods:

NPX levels, we excluded them in our analysis. To identify differentially expressed proteins between cases and controls, ~~Analysis of Variance (~~ANOVA was~~) and unsupervised hierarchical clustering were~~ performed with the Olink Analyze R package. This allowed us to capture a broad set of proteins with potential differential expression. However, recognizing the risk of false positive findings, correction for ~~M~~multiple comparisons~~hypothesis correction~~ was applied using the Benjamini-Hochberg method.~~; and~~ ~~p~~Post-hoc correction (Tukey test) was subsequently done to test the significance of pairwise comparisons of the individual HSV phenotypes versus controls. As the data distribution deviated slightly from a

5. The proteins analyzed by the hierarchical clustering (Figure 3) were already pre-selected as significant for the DIS vs Control, and from Fig 2, these were largely not significant in the other comparisons. Therefore, the conclusion that they clustered by condition is not surprising. Could the authors provide any additional interpretation of this result?

Response: Thank you for the comment. As mentioned in our response to Reviewer 1, comment 6, we recognized that the heatmap did not provide the necessary clarity and information as intended. To address this, we have decided to present the data using boxplots (Figure 4, former Figure 3).

6. The OLink panels contain a fraction of the annotated human proteome. While this may not be an issue for the goal of finding biologically meaningful signatures, it does introduce a bias in the data. Specifically, did the authors take this into consideration when performing their Pathway enrichment? Many enrichment tools use the whole genomes as the background comparison be default. For this study, using the OLink panel, the whole genome background would not be appropriate. Moreover, generally, pathway enrichment has poor sensitivity when used the input contains a low # of genes/proteins. The authors may want to consider a simpler approach that shows assignment of proteins to their annotated pathways (independent of their enrichment).

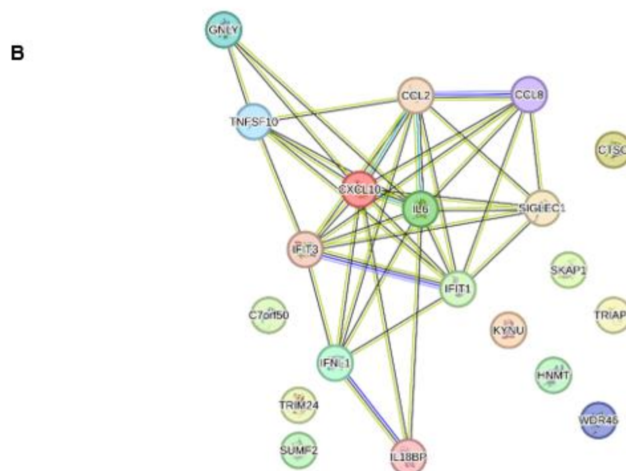
Response: Thank you for the important comment. We agree that pathway enrichment analysis may not be the most suitable approach. Therefore, we decided to go with a simpler approach by functional annotation

using the Database for Annotation, Visualization, and Integrated Discovery (DAVID: https://david.ncifcrf.gov/home.jsp). The point has been addressed in response to Reviewer 1, comment 7.

7. The known physical and/or functional relationships between the differential proteins is not clear. I would suggest the authors perform a network analysis, e.g., using STRING or similar approach.

Response: Thank you for the comment. We agree and have added a protein network plot in Figure 5B.

Figure 5B:



A) Functional annotation of the proteins that remained significant when comparing disseminated HSV disease versus controls in the post-hoc analyses. For these proteins, one pathway "Cytokine" was statistically significant (false discovery rate adjusted p-value < 0.05). B) Protein network visualisation illustrating the relationships between the significant proteins in disseminated HSV disease compared to controls. Nodes represent individual proteins, while edges indicate different types of associations: co-occurrence (blue lines), co-expression (black lines), text mining (yellow lines), neighbourhood (green lines) and known interactions from curated databases (teal lines). The network was created using the STRING database (www.string-db.org).

Minor points

1. SEM was not mentioned in the Introduction and only appeared in Results as "SEM". This should be defined and explained in the Introduction.

Response: Thank you for the comment. The introduction has been revised to include definition of 'SEM'.

Introduction:

INTRODUCTION

Herpes simplex virus (HSV) infection is a life-threatening neonatal infection, with a mortality rate of up to 70-80% when disseminated and neurological damage of up to 50% when affecting the central nervous system (CNS) (1–3). In contrast, skin-eye-mouth (SEM) disease is generally less severe, characterized by vesicular lesions on mucocutaneous surfaces, but can progress to more severe forms if left untreated (2).

2. Generally, the size of the text for the axis labels in Figure 1 and 4, and similar graphs in the Supplement, could be made larger for improved readability.

Response: Thank you for the comment. Figure 2 (former Figure 1) and Figure 5 (former Figure 4) has been changed.

3. Please include the average NPX, associated errors, and p-values in the Supplemental Tables 2-4.

Response: Thank you for the comment. The requested information has been included in Supplemental Tables 2-4 (provided in separate Excel files).

Additional revisions:

We have revisited the language and improved throughout the manuscript.

Rigshospitalet

COPENHAGEN UNIVERSITY HOSPITAL RIGSHOSPITALET

Department of Pediatrics and Adolescent Medicine

Blegdamsvej 9
Copenhagen, 2100 Ø
Denmark

+ 45 35 45 9761
Date: October 21, 2024

REBUTTAL LETTER TO THE REVIEWERS

Manuscript reference number: COMMSMED-23-0926-B

Title: Proteomic profiling of neonatal herpes simplex virus infection on dried blood spots: exploring screening perspectives.

Reviewer #1 (Remarks to the Author):

The manuscript by Kia Hee Schultz Dungu et al. is describing the use of proteomic profiling on dried blood spots for screening neonatal herpes simplex virus infection. The manuscript was altered according to the reviewers comments. In particular, the method part of the manuscript is improved. Furthermore, the heat-map was replaced by a box-plot and the biological interpretation of the differently expressed proteins is explored by using DAVID instead of Pathway enrichment analysis. Some additional supplementary figures are included to support the authors findings.

However, some minor points should be addressed:

- DAVID assesses the overrepresentation of differently expressed proteins (or genes) in defined sets of proteins (or genes). These sets could be GO-terms (molecular function MF, biological pathway BP), KEGG-pathways, Reactome-pathways and UniProt KeyWord sets. Two overrepresented sets found by the analyses were UP_KW-0202~Cytokine and UP_KW-0802~TPR. These are both not per se pathways but rather Key Word sets. This should be marked in the text and the wording changed accordingly.

Response: Thank you for the comment. We appreciate the correction and have revised accordingly.

Methods:

To identify biological pathways and Key Word sets associated with the clinical phenotypes, we performed functional annotation of the differentially expressed proteins using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) (32). We included only proteins that

Results:

Biological pathways associated with disease phenotypes

Functional annotation of the post-hoc significant proteins allowed us to identify biological pathways and Key Word sets associated with the clinical phenotypes compared to controls (Figure 5A, Supplemental Table 5). For neonates with disseminated disease, we identified 33 pathways and Key Word sets, but only the Key Word setpathway "Cytokine" was statistically significant (FDR adjusted p-value < 0.05). Other relevant pathways, such as "Cytokine-cytokine receptor interaction", "Viral protein interaction with cytokine and cytokine receptor" and "Chemotaxis", were identified but did not reach statistical significance. In the comparison of SEM disease versus controls, the Key Word setpathway "Tetratricopeptide Repeat was significant. Other relevant pathways, such as "Interferon alpha/beta signalling" and "Cytokine signalling in the immune system" were not significant. The "Antiviral defence" pathway appeared in both comparisons: disseminated disease versus controls and SEM disease versus controls. No pathways or Key Word sets were identified for CNS disease versus controls, as there were no post-hoc significant proteins in this comparison (Supplemental Table 5).

Figure 5A, legend:

A) Functional annotation of the proteins that remained significant when comparing disseminated HSV disease versus controls in the post-hoc analyses. For these proteins, one Key Word setpathway "Cytokine" was statistically significant (false discovery rate adjusted p-value < 0.05). B) Protein network

- The PCAs are shown in Supp. Fig. 1. The presentation of these analyses is appreciated. However, two questions arise: What are the red points on the plots? Are these controls? Please explain in figure caption. It seems as the PCA of "Inflammation II" has another distribution of the red points and PC1 explains 83.2% of the variance, while PC1 of the others only ~20%. Please explain.

Response: Thank you for the comment. The red points represent outliers, defined as data points that fall more than three standard deviations from either PC1 or PC2. These outliers are not categorised by phenotype, as the definition does not differentiate between them. Specifically, the outliers include 5 control samples and 3 case samples. In the Inflammation_II dataset, where PC1 explains 83.2% of the total variance, this indicates that PC1 captures a larger proportion of the overall variability compared to PC1 in the other panels. This is visually evident by the broader spread along the PC1 axis in the corresponding PCA plot.

Supplemental Figure 1, legend:

PCA plots depicting data quality control for each Olink Explore 3072 panel (Olink Proteomics Assays, Uppsala, Sweden). Red points represent outliers, defined as data points falling more than three standard deviations from either PC1 or PC2.

Reviewer #2 (Remarks to the Author):

The authors have acknowledged and addressed all my points of concern. The authors have now included sufficient detail to fully evaluate the study. While the technique of O-Link and the sample population size introduces certain caveats in the study interpretation, the authors revised manuscript is sufficiently transparent about these points and uses appropriate language to prevent misinterpretation of the study's relevance and broad impact. Also, I recognize that the nature of the samples being collected and analyzed represents challenges. Therefore, I feel the revised manuscript is in an appropriate state for publication.

Response: Thank you very much.