Ageing Proteomics

Ashfaq Ali (ashfaq.ali@nbis.se)

6/15/2020

Contact: valentina.siino@immun.lth.se

PI: Fredrik Levander

NBIS Staff

Ashfaq Ali

Dag Ahrén

Support #5219 Plasma proteome changes during ageing

Agreed analyses:

- Calculations of association of proteins with ageing using correlation and group comparisons. Inclusion of covariates.
- Clustering of protein profiles as function of age
- Comparison of protein levels between individuals living at different locations

If time permits and depending on discussion of first two points:
- Comparison with published data.
- Functional enrichment of proteins from previous analyses

Short Summary

Project analyses was carried out in several iterations where analyses was adjusted/added as per user request. Feedback on results was provided through meetings or email communication. Results were delivered using Lund University box.

Following is the list of analyses carried out.

Data preparation

Loess normalized data provided by Valentina was used in the analyses in the project. In the data preparation and QC, following steps were used.

- Prepare ´SummarisedExperiment´object with sample meta data and protein data provided
- Analyze missingness and filter out poor quality samples and proteins

Analyze missingness

Data missingnes was analyzed using "mice" package in R. We considered proteins present in at least 50% of the samples for further analyses.

Exploratory analyses

PCA analyses

Principle component analyses was used to perform exploratory analyses of the proteomics data. Explained variance by each principle component was calculated using the output from prcomp´ function in R. Correlation between principle components and phenotype (clinical) variables was calculated using ´lm´ function in R and "adj.r.squared" values were extracted from the summary statistics produced by the ´lm´ function.

Plot sample overview using PCA

Cluster samples

Explained variance by PCs

This step calculates explained variance by each principle component of the protein data.

Correlation of principle component with Phenotype data

To identify the most relevant clinical variables associated with protein data, we calculate correlation between the principle components and the phenotype variables.

Differential expression analyses

For differential expression analyses we used "limma" package in R. Following models were used to analyze deferentially abundant proteins.

- *Basic model* protein ~ Age
- *Gender and BMI adjusted model* protein ~ gender + BMI + Age
- *Interaction with environmental variables* protein ~ gender + bmi + Age + ENV + Age*ENV

Age in gender and BMI adjusted model

Test for Age, location and Age*location interaction (proteins with missingness removed)

Volcano plots

Volcano plots to visualize the differntially abundant proteins modified versions of EnhancedVolcano plots was used.

Test for interaction with location with full protein data

Heatmap for significantly proteins with age

Heatmap for proteins significantly associated with age

Plot interaction terms using ggplots2

An example for interaction plot

Code for creating interaction plots for all significantly regulated proteins

Extract and format data

Function for plotting interaction

Plot interactions

Differential expression analyses of proteins related to the UA2 variable

Differential expression analyses using following models.

1. protein ~ Gender + UA2
2. protein ~ Gender + UA2 + Age + UA2*AGE

Basic model

Model for interaction

Volcano plots for UA2

Interaction plot for UA2*Age interactions.

Test for handgrip associated proteins

Handgrip*Age interaction

Cluster analyses to identify Age related clusters

Age related protein clustering

To identify patterns associated with age we performed cluster analyses of on proteins associated with age. To visualize trends we create age related intervals as following.

1. Plot for groups between 0-40, 61-60, 61-80 and 81-112.
2. Plots for groups for age span of 10 years each

Mean protein intensity was calculated for proteins in each cluster followed by average over each age intervals.

For clustering, protein data was scaled followed by calculating dsitance between protein observation using euclidean distance. Clustering was then performed by using complete linkage between observations.

Extract age associated proteins

Summarize proteins for interval mentioned in 1.

Summarize proteins for interval mentioned in 2.

Cluster analyses

To identify patterns in protein data clustering using complete linkage was performed to identify cluster of proteins associated with Age.

Export results

Plot clusters

Export for 10 year interval

Plot for 10 year interval

SWANDe analyses

Differential expression Sliding window analysis distinguishes (DEswan) was used to identify waves of aging plasma proteins. Age span between 20 years to 120 years with interval size of 10 years was slected for sliding window analyses in the ´DEswan´ function implemented i ´DEswan´ package. Sex was used as covariate in the DEswan analyses.

Reshape and export resulst from DEswan

Plot number of siginifcant protein in each interval

Comparison with Nature Benoit Lehallier et al doi

https://doi.org/10.1038/s41591-019-0673-2

Death related proteins protein

Gene set enrichment analyses using MSigDb

Load gene sets

Gene Set Enrichment analyses

Plot GSEA results

GSEA with background

Gene annotations

MMP correlation

References

Acknowledgments

Data responsibility

NBIS & Uppnex Unfortunately, we do not have resources to keep any files associated with the support request. We suggest that you safely store the results delivered by us. In addition, we ask that you remove the files from UPPMAX/UPPNEX after analysis is completed. The main storage at UPPNEX is optimized for high-speed and parallel access, which makes it expensive and not the right place for long time archiving.

Long-term backup The responsibility for data archiving lies with universities and we recommend asking your local IT for support with long-term data archiving. Also a newly established Data Office at SciLifeLab may be of help to discuss other options.

Acknowledgments

If you are presenting the results in a paper, at a workshop or conference, we kindly ask you to acknowledge us.

```
NBIS Staff are encouraged to be co-authors when this is merited in accordance to the ethical recommenda
```

NGI In publications based on data from NGI Sweden, the authors must acknowledge SciLifeLab, NGI and UPPMAX: "The authors would like to acknowledge support from Science for Life Laboratory, the National Genomics Infrastructure, NGI, and Uppmax for providing assistance in massive parallel sequencing and computational infrastructure."

Support project closing procedures

Once the final report is delivered, you should be contacted by one of our managers, Jessica Lindvall jessica.lindvall@nbis.se or Henrik Lantz henrik.lantz@nbis.se, with a request to close down the project in our internal system and for invoicing matters. If we do not hear from you within 30 days the project will be automatically closed and invoice sent. Again, we would like to remind you about data responsibility and acknowledgements, see Data Responsibility and Acknowledgments sections.

You are naturally more than welcome to come back to us with further data analysis request at any time via http://nbis.se/support/support.html.

Thank you for using NBIS and all the best for future research.