Design and implementation of analysis pipeline for single cell type proteomics data

By

Lukas Gamp

in partial fulfilment of the requirement
for the degree of MSc
in Bioinformatics

mm yy

Abstract

(the spacing is set to 1.5)

no more than 250 words for the abstract

- a description of the research question/knowledge gap – what we know and what we don't know
- how your research has attempted to fill this gap
- a brief description of the methods
- brief results
- key conclusions that put the research into a larger context

Contents

Acknowledgements

Thank you for following this tutorial!

I hope you'll find it useful to write a very professional dissertation.

1 Introduction

1.1 Mass Spectrometry

1.1.1 MS-Spectrum

Every peptide is reflected by its‘ indivual fingerprint in the ms-spectrum. The fingerprint is based on the chemical properties and modifications. In order to identify proteins, fingerprints are matched against a sequence database (Cox & Mann 2008).

1.1.2 Interpretation of the data

Since ms data has a high resoultuion, algorithms are used to convert peaks to an interpretable form. These algorithms find local minima of the function to seperate peaks from each other. Peaks include all isotopes of the proteins `containing atoms. MaxQuant \citep{Cox2008} is one of the sofware packages to process the data and provides it for further analysis and statistical testing. To find the istopic distribution of a biomolecue, MaxQuant creates a vertex of every single peak and connects them with their possible isotopic counterparts by finding the proportion of mass of an average aminoacid to its` respective isotope (averagine (Senko et al. 1995)). After this procedure a large amount of noise is reduced and a single peak reflects a small biomolecule. These biomolecules can now be searched in a database in forward and reverse direction. The peptide identification (P-) score indicates the fit of the data to the found sequence in the database according to the length of the peptide and is used to calculate the posterior error probability (FDR). The calculation of the false discovery rate is then calculated by taking FDR into contrast. After these calculation the peptide peaks are joined to a protein and can be quantified.

follow sctranscr. analysis pipe

signatures of cell clustures

monocyte data acvailiabe

tmt label –> noise reduction

tools for clustering

`plst, knm,`

follow sctranscr. analysis pipe

signatures of cell clustures

monocyte data acvailiabe

tmt label –> noise reduction

tools for clustering

`plst, knm,`

2 Methods

3 Materials and Methods

3.1 Materials

For analysis two types of cells were used. One type is the Jurkat-based cell line (J-lat) with integrated HIV.

The other type of cells are macrophages with a sample size of 72 cells. The analysis is done with two groups. A HIV negative (HIV-) control group and a HIV positive (HIV+) group.

3.2 Data

3.2.1 Acquisition

Analysis of the data was done with MaxQuant (Cox & Mann 2008)

4 Results

Some more guidlines from the School of Geosciences.

This section should summarise the findings of the research referring to all figures, tables and statistical results (some of which may be placed in appendices). - include the primary results, ordered logically - it is often useful to follow the same order as presented in the methods. - alternatively, you may find that ordering the results from the most important to the least important works better for your project. - data should only be presented in the main text once, either in tables or figures; if presented in figures, data can be tabulated in appendices and referred to at the appropriate point in the main text.

Often, it is recommended that you write the results section first, so that you can write the methods that are appropriate to describe the results presented. Then you can write the discussion next, then the introduction which includes the relevant literature for the scientific story that you are telling and finally the conclusions and abstract – this approach is called writing backwards.

5 Discussion

the purpose of the discussion is to summarise your major findings and place them in the context of the current state of knowledge in the literature. When you discuss your own work and that of others, back up your statements with evidence and citations. - The first part of the discussion should contain a summary of your major findings (usually 2 – 4 points) and a brief summary of the implications of your findings. Ideally, it should make reference to whether you found support for your hypotheses or answered your questions that were placed at the end of the introduction. - The following paragraphs will then usually describe each of these findings in greater detail, making reference to previous studies. - Often the discussion will include one or a few paragraphs describing the limitations of your study and the potential for future research. - Subheadings within the discussion can be useful for orienting the reader to the major themes that are addressed.

6 Conclusion

The conclusion section should specify the key findings of your study, explain their wider significance in the context of the research field and explain how you have filled the knowledge gap that you have identified in the introduction. This is your chance to present to your reader the major take-home messages of your dissertation research. It should be similar in content to the last sentence of your summary abstract. It should not be a repetition of the first paragraph of the discussion. They can be distinguished in their connection to broader issues. The first paragraph of the discussion will tend to focus on the direct scientific implications of your work (i.e. basic science, fundamental knowledge) while the conclusion will tend to focus more on the implications of the results for society, conservation, etc.

7 Bibliography

Cox, J. & Mann, M. (2008), 'Maxquant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification', *Nature Biotechnology* **26**, 1367–1372.

Senko, M. W., Beu, S. C. & McLaffertycor, F. W. (1995), 'Determination of monoisotopic masses and ion populations for large biomolecules from resolved isotopic distributions', *Journal of the American Society for Mass Spectrometry* **6**, 229–233.

8 Appendix(ces)

8.1 Appendix A: additional tables

Insert content for additional tables here.

8.2 Appendix B: additional figures

Insert content for additional figures here.

8.3 Appendix C: code

Insert code (if any) used during your dissertation work here.