A Bioconductor workflow for processing and analysing spatial proteomics data
Reply to reviewers

Lisa Breckels and Laurent Gatto

March 26, 2018

Reviewer 1 - Daniel J. Stekhoven

We thank reviewer 1 for their comments. Please find our responses to these inset below.

> Next to reducing the dimensions of data for visualisation, PCA also offers a way to understand how the variability is distributed across the multidimensional data by providing linear combinations of the variables which then constitute the actual PCs. On that note it would be nice to mention this in Visualising markers section on page 16, where PC7 explains not much variability but due to the correct weighing of the variables we do get a separation between mitochondrial and peroxisome. This then can be further motivated with Figure 9 - where we probably can see that the weights for the fractions where the two localisations differ are larger than otherwise.

We have added a paragraph reminding users what PCA does in the Visualising Markers section to motivate the choice of looking at PC's 1 and 7. Figure 9 (now Figure 8), the corresponding code and an explanation of the plotDist function has been moved to this section to lead on from the reference to marker profiles and separation.

> I was unable to reproduce Figure 13 comparing the two MSnSets. While I was able to look at each set separately using pRolocVis(hllst@x[[I]]), where i is 1 or 2, I only got an error using the code from the manuscript. When using ?remap=FALSE? it actually works, but since this makes barely sense it is of no use - but just as a hint at debugging it.

This should be fixed in the latest version of `pRolocGUI`.

> You really need to make the results from the phenoDisco classification available too. It is super disappointing that one cannot continue reproducing the code from page 23 on, because it takes 24 hours to compute it using 40 cores?

The results are available and stored in 'pRolocdata' for users. This is what is called in the manuscript under the hood:

```
package = "pRolocdata") csvfile <- dir(extdatadir, full.names =
TRUE, pattern = "hyperLOPIT-SIData-ms3-rep12-intersect.csv") f0 <-
dir(extdatadir, full.names = TRUE, pattern = "bpw-pdres.rds") pdres
<- readRDS(f0) hl <- addMarkers(hl, pdres, mcol = "pd", verbose =
FALSE)
```

Please note that we already say in the workflow: "Note: We do not evaluate this code chunk in this document as the algorithm is computationally intensive and best parallelised over multiple workers. This phenoDisco analysis took 24 hours to complete when parallelised over 40 workers."

However, we have moved this statement to the end of the section for clarity. We state the output of these analyses is readily available for readers and stored in the `pd` column of the `MSnSet` object called `hl`. Readers can examine

these results and get an idea of the type of output one can expect from this a `phenoDisco` analysis.

Specifically we have added the following to the end of the section to make this clear.

"Please note, in this document we can not evaluate the call to `phenoDisco` in the code chunk above as the algorithm is computationally intensive and best parallelised over multiple workers. This phenoDisco analysis took 24 hours to complete when parallelised over 40 workers. The ouput of running the `phenoDisco` algorithm is an `MSnSet` containing the new data clusters, appended to the `featureData` under the name `pd`. We have made the results readily available for users to interrogate and get an idea of the type of output one may gain from using this function. The results can be displayed by using the `getMarkers` function. We see that 5 new phenotype data clusters were found."

> The above comment is of course also true for the KNN TL Optimisation on page 33 - this needs to be downloadable, since not everyone has access to Cambridge?s HPC and probably even less have 76 hours to spare.

The same as for the phenoDisco analysis and svm, the TL results are stored as a RDS in pRolocdata and loaded in the backgroun
We thank the reviewer for their constructive and positive comments.

Reviewer 2: Leonard J. Foster