Standardised and reproducible analysis of mass spectrometry-based single-cell proteomics data

Christophe Vanderaa, Laurent Gatto

Computational biology and bioinformatics, de Duve Institute, UCLouvain

christophe.vanderaa@uclouvain.be

Summary

Recent advances in sample preparation, processing and mass spectrometry (MS) have enabled the emergence of MS-based single-cell proteomics (SCP). We have developed a computational framework that facilitates the analysis of such data. We therefore provide two packages, `scpdata` and `scp`, and demonstrate the application of our pipeline by replicating and improving the analyses of two published data sets and show its relevance in the processing and interpretation of MS-based SCP data.

MS-SCP technology

MS-SCP is enabled by recent advances in sample preparation, liquid chromatography and mass spectrometry.

There are 2 main strategies:

- **Label-free acquisition**: **low throughput** (∼ 15 samples/day) and **low identification** rate, but **accurate quantification** (*e.g.* Zhu et al. 2019 [1])
- **Multiplexed acquisition**: **high throughput** (∼ 200 samples/day), but label **cross-contamination** (*e.g.* Specht et al. 2020 [2])

Both methods generate data with high **missingness** and considerable **batch effects**.

SCP data framework

SCP data framework = `QFeatures` + `SingleCellExperiment`

scp

```
readSCP(quantTable = quantData, metaTable = metaData,
        channelCol = "Channel", batchCol = "Set") %>%
    zeroIsNA(i = 1:4) %>%
    filterFeatures(~ Reverse != "+" & Potential.contaminant != "+") %>%
    subsetByAssay(dims(.)[1, ] > 150) %>%
    computeSCR(i = 1:3, colDataCol = "SampleType",
               carrierPattern = "Carrier",
               samplePattern = "Macrophage|Monocyte") %>%
    filterFeatures(~ !is.na(.meanSCR) & .meanSCR < 0.1) %>%
    aggregateFeaturesOverAssays(i = 1:3, fcol = "peptide",
                                name = paste0("peptides_", names(.)),
                                fun = robustSummary) %>%
    joinAssays(i = 4:6, name = "peptides") %>%
    computeMedianCV(i = "peptides_filter1", proteinCol = "protein",
                    peptideCol = "peptide", batchCol = "Set") %>%
    normalize(i = "peptides", name = "peptides_norm",
              method = "median", na.rm = TRUE) %>%
    logTransform(i = "peptides_norm", name = "peptides_log",
                 base = 2) %>%
    aggregateFeatures(i = "peptides_log", name = "proteins",
                      fcol = "protein", robustSummary) %>%
    impute(i = "proteins_norm", method = "knn", k = 3) ->
    scp
```

Conclusion

The amount of available MS-based SCP data is rapidly increasing and so is the need for dedicated analysis software. We hence offer two R/Bioconductor packages. The first package, `scpdata`, disseminates curated SCP data sets for method development and benchmarking. The second package, `scp`, implements functions to streamline the analysis of SCP data. They will facilitate the development of dedicated algorithms, namely for dealing with missing data and batch effects that are characteristic of MS-SCP data.

scpdata

	Title	Description	Species	Date	# Assays
1	specht2019v2	Specht et al. 2019: macrophage...	Homo sapiens	2019-12-05	179
2	specht2019v3	Specht et al. 2019: macrophage...	Homo sapiens	2019-10-04	179
3	dou2019_lysates	Dou et al. 2019: HeLa lysates ...	Homo sapiens	2019-10-15	3
4	dou2019_mouse	Dou et al. 2019: single cells ...	Mus musculus	2019-10-15	13
5	dou2019_boosting	Dou et al. 2019: testing boost...	Mus musculus	2019-10-15	7
6	zhu2018MCP	Zhu et al. 2018 (Mol. Cel. Pro...	Rattus norvegicus	2018-09-01	1
7	zhu2018NC_hela	Zhu et al. 2018 (Nat. Comm.): ...	Homo sapiens	2018-02-28	1
8	zhu2018NC_lysates	Zhu et al. 2018 (Nat. Comm.): ...	Homo sapiens	2018-02-28	1
9	zhu2018NC_islets	Zhu et al. 2018 (Nat. Comm.): ...	Homo sapiens	2018-02-28	1
10	cong2020AC	Cong et al. 2020 (Ana. Chem.):...	Homo sapiens	2020-01-02	9
11	zhu2019EL	Zhu et al. 2019 (eLife): chick...	Gallus gallus	2020-11-04	62

Imputation

Batch correction

This work is funded by an Aspirant FRS-FNRS fellowship awarded to Christophe Vanderaa. The poster is available at .

References

[1] Y. Zhu, M. Scheibinger, D. C. Ellwanger, J. F. Krey, D. Choi, R. T. Kelly, S. Heller, and P. G. Barr-Gillespie, "Single-cell proteomics reveals changes in expression during hair-cell development," *Elife*, vol. 8, Nov. 2019.

[2] H. Specht, E. Emmott, A. A. Petelski, R. Gray Huffman, D. H. Perlman, M. Serra, P. Kharchenko, A. Koller, and N. Slavov, "Single-cell proteomic and transcriptomic analysis of macrophage heterogeneity." Oct. 2020.