Assessing sub-cellular resolution in spatial proteomics experiments

Laurent Gatto* and Lisa M. Breckels

Computational Proteomics Unit
University of Cambridge, UK

September 8, 2016

Abstract

A meta-analysis assessing and comparing the sub-cellular resolution of spatial proteomics experiments.

1 Introduction

In biology, the localisation of a protein to its intended sub-cellular niche is an necessary condition for it to assume its biological function. Indeed, the localisation of a protein will determine its specific biochemical environment and its unique set of interaction partners. As a result, the same protein can assume different functions in different biological contexts and its mis-localisation can lead to adverse effects.

Spatial proteomics is the systematic and high-throughput study of protein sub-cellular localisation. A wide range of techniques (reviewed in [9]) and computational methods [10] to confidently infer the localisation of thousands of proteomics have been documented. Most techniques rely on some form of sub-cellular fractionation using differential centrifugation or separation

*lg390@cam.ac.uk

along density gradients and the subsequent quantitative assessment of relative protein occupancy profiles in these sub-cellular fractions. Reciprocally, a wide ranging of computational methods have been applied, ranging from unsupervised learning e.g.clustering, dimensionality reduction [19], supervised learning e.g. classification (reviewed in [10], semi-supervised learning and novelty detection [2] and, more recently, transfer learning [3].

Despite these advances, there is surprisingly little agreement in the community as to what constitutes a reliable spatial proteomics experiment, i.e a dataset that generates confident protein assignment results. Is is however implicit that reliability and trust in the results is dependent on adequate sub-cellular resolution, i.e. *enough* separation between the different sub-cellular niches under study to be able to confidently discern protein profiles originating from different sub-cellular niches. And yet, every spatial proteomics publication will somehow arbitrarily claim to have obtained satisfactory resolution.

The importance of adequate sub-cellular resolution reaches beyond the generation of reliable static spatial maps. It is a necessary property of the data to consider tackling more subtle sub-cellular patterns such as multi- and trans-localisation, i.e. the localisation of proteins in multiple sub-cellular niches and the relocation of proteins upon perturbation [10].

In this work, we describe how to understand and interpret widely used dimensionality reduction methods and visualise spatial proteomics data to critically assess their resolution and propose a simple, yet effective method to quantitatively measure resolution and compare it across different experiments. Our recommendations should be useful to spatial proteomics practitioners, to assess the sub-cellular resolution of their experiments and compare it to similar studies while setting up and optimising their experiments, as well biologists interested in critically assessing spatial proteomics studies and their claims.

All the data and software used in this work is available in the `pRoloc` and `pRolocdata` packages [11]. In the interest of readability, we do not display all details in this rendering of the manuscript. Nevertheless, the code to reproduce all results and figures is available in the source of this document in the public manuscript repository [8].

2 Spatial proteomics datasets

For this meta-analysis, we make use of 15 spatial proteomics datasets, summarised in table 1. These data represent a diverse range of species, instruments and methodologies.

We have applied minimal post-processing to the data and have used, as far as possible, the data and annotation provided by the original authors. The data from Foster et al. [7] have been annotated using the curated marker list from Christoforou et al. [5], as only a limited number of markers was provided by the authors[1]. We have also only considered clusters that were defined by at least 7 markers[2]. Where a combined dataset of multiple replicated experiments are provided by the original authors, we have used the combined dataset, rather than individual replicates, as this generally refers to the dataset with best sub-cellular resolution [20]. If there is no combined dataset, or only a single replicate is reported we have taken the (only or) first replicate experiment. In addition, for dimensionality reduction and visualisation, we have systematically replaced missing values by zeros. When calculating distances between protein profiles (see section 3.3), however, missing values were retained.

It is important to highlight that not all experiments used in this study have as main goal the generation of a global sub-cellular map. While the works of Dunkley et al. [6] (*Mapping the Arabidopsis organelle proteome*), Hall et al. [13] (*Mapping organelle proteins and protein complexes in Drosophila melanogaster*) and more recently Christoforou et al. [5] (*A draft map of the mouse pluripotent stem cell spatial proteome*) and Itzhak et al. [14] (*Global, quantitative and dynamic mapping of protein subcellular localization*) explicitly state such goal, other experiments such as Groen et al. [12] (*Identification of trans-golgi network proteins in Arabidopsis thaliana root tissue*) or Nikolovski et al. [16] (*Label free protein quantification for plant Golgi protein localisation and abundance*) have a much more targeted goal (identification of trans-Golgi -network and Golgi apparatus proteins, respectively). When an experiment is targeted at resolving a particular niche, it is often the case that other sub-compartments are less well-resolved as a sacrifice and hence,

[1] This results from the fact that they used a simple distance measurement, termed χ^2, against very few markers to base their sub-cellular localisation prediction

[2] This number is relatively low, and we would typically recommend at least 13 markers per class to perform cross-validation when optimising classifier parameters. See Gatto et al. [10] and the main `pRoloc` tutorial for details.

Data	Proteins	Fractions	Clusters	PC var (%)	Title
hyperLOPIT2015	5032	20	14	72.26	hyperLOPIT experiment on Mouse E14TG2a embryonic stem cells from Christoforou et al. (2016) [5]
hyperLOPIT2015ms2	7114	10	14	74.72	hyperLOPIT experiment on Mouse E14TG2a embryonic stem cells from Christoforou et al. (2016) [5]
andy2011	1371	8	12	65.04	LOPIT experiment on Human Embryonic Kidney fibroblast cells from Breckels et al. (2013) [2]
itzhak2016stcSILAC	5265	30	12	70.61	Data from Itzhak et al. (2016) [14]
rodriguez2012r1	2215	11	12	38.95	Spatial proteomics of human inducible goblet-like LS174T cells from Rodriguez-Pineiro et al. (2012) [17]
tan2009r1	888	4	11	88.49	LOPIT data from Tan et al. (2009) [18]
E14TG2aS1	1109	8	10	65.98	LOPIT experiment on Mouse E14TG2a Embryonic Stem Cells from Breckels et al. (2016) [3]
trotter2010	347	16	10	81.14	LOPIT data sets used in Trotter et al. (2010) [20]
dunkley2006	689	16	9	86.70	LOPIT data from Dunkley et al. (2006) [6]
foster2006	1555	26	8	53.13	PCP data from Foster et al. (2006) [7]
nikolovski2014	1385	20	8	67.97	LOPIMS data from Nikolovski et al. (2014) [16]
groen2014cmb	424	18	7	64.30	LOPIT experiments on Arabidopsis thaliana roots, from Groen et al. (2014) [12]
nikolovski2012imp	1385	32	7	77.82	Meta-analysis from Nikolovski et al. (2012) [15]
andreyev2009rest	2642	36	6	25.39	Six sub-cellular fraction data from mouse macrophage-like RAW264.7 cells from Andreyev et al. (2009) [1]
hall2009	1090	16	5	63.45	LOPIT data from Hall et al. (2009) [13]

Table 1: Summary of the datasets used in this study. The percentage of variance along the principal components (PC) is related to the PCA plots on figure 8. All datasets are available in the `pRolocdata` package.

it is important to keep the overall aim of the studies in mind when assessing their resolution.

3 Assessment

3.1 Sub-cellular diversity

A first assessment that provides an important indication of the resolution of the data concerns the number and diversity of sub-cellular niches that are annotated. In the 15 datasets used in this study, this number ranged from 5 (dataset *hall2009*) to 14 (dataset *hyperLOPIT2015*). These numbers should be assessed in the light of about 25 different sub-cellular niches that are documented in all 15 datasets, which are still underestimating the biological sub-cellular diversity.

3.2 Dimensionality reduction and visualisation

Principal component analysis (PCA) is a widely used dimensionality reduction technique in spatial proteomics. It projects the protein occupancy profiles into a new space in such a way as to maximise the spread of all points (i.e. labelled and unlabelled proteins) along the first new dimension (principal component, PC). The second PC is then chosen to be perpendicular to the first one while still maximising the overall variability. Each PC (there are as many as there are fractions) accounts for a percentage of the total variability and it is not uncommon, in well executed experiments, that the two first PCs summarise over 70% of the total variance in the data, confirming that the resulting visualisation remains a reliable and useful simplification of the original multidimensional data.

By firstly summarising the occupancy profiles along PC1 and PC2 (and, possibly, other PCs of interest if necessary, it becomes possible to visualise the complete dataset in a single figure (as opposed to individual sets of profiles - see for example figure 5 in Gatto et al. [9]). In a first instance, it is advised to visualise the data without annotation to confirm the presence of discrete clusters, i.e. dense clouds of points that are well separated from the rest of the data (see for example data from Christoforou et al. [5] on figure 1, left). Such patterns can further be emphasised by using transparency (figure 1, centre) or binned hexagon plots (figure 1, right) to highlight density.

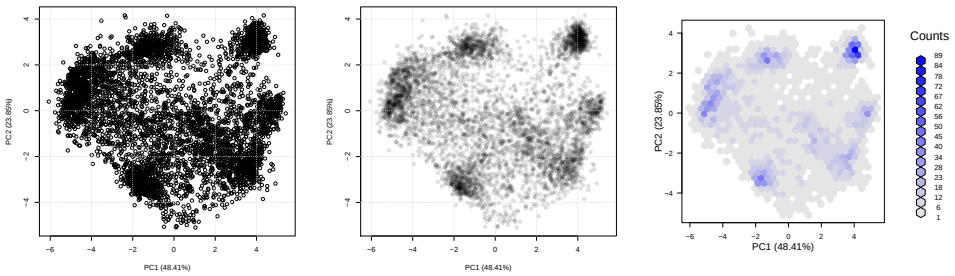

Figure 1: Unsupervised visualisation of spatial resolution using the `plot2D` function from the `pRoloc` package.

In figure 2 we compare three datasets to illustrate different levels of cluster density and separation. We see areas of high density (many proteins per hexagon) are highlighted by dark blue, and less dense areas are white/grey, as indicated by the count key on the right-hand side of each plot. The figure on the left is the hyperLOPIT data from Christoforou et al. [5] (as on figure 1) that used synchronous precursor selection (SPS) MS^3 on an Orbitrap Fusion. The middle figure represents the same experiment and same proteins, analysed using conventional MS^2, illustrating the effect of reduced quantitation accuracy. Finally, on the right, an experiment with considerable less resolution [13].

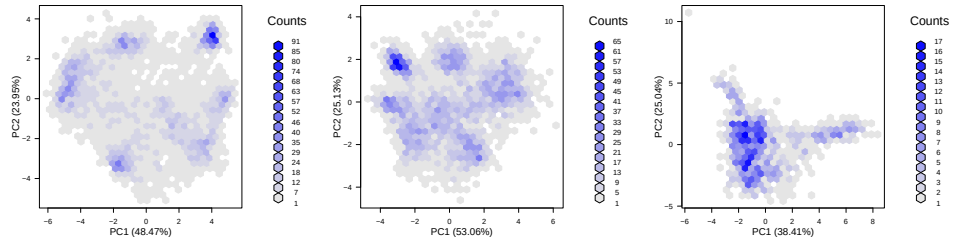

Figure 2: Comparing the cluster density and separation of experiments with excellent (left), intermediate (centre) and poor (right) resolution.

Considering that the aim of sub-cellular fractions is to maximise separation of most sub-cellular niches, one would expect these sub-cellular clusters to be separated optimally in a successful spatial proteomics experiment. In

PCA space, this would equate to the location of the annotated spatial clusters along the periphery of the data points. In other words, the maximum variability of a successful spatial proteomics experiments should be reflected by the separation of the expected/annotated spatial clusters.

Another dimensionality reduction method that is worth mentioning here is linear discriminant analysis. LDA will project the protein occupancy profiles in a new set of dimensions using as criterion the separation of marker classes by maximising the between class variance to the within class variance ratio. As opposed to the *unsupervised* PCA, the *supervised* LDA should not be used as an experiment quality control, but can be useful to assess if one or more organelles have been preferentially separated.

It is important to highlight that these representations, while reflecting a major fraction of the variability in the data, are only a summary of the total variability. Some sub-cellular niches that overlap in 2 dimensions can be separated along further components. It is sometimes useful to visualise data in three dimensions (using for example the `plot3D` function in the `pRoloc` package), which still, however, only reflect part of the total variability. When assessing the resolution of some organelles of interest, one should compare the full occupancy profiles of the marker proteins (`pRoloc`'s `plotDist` function can be used for this) or visualise a dendrogram representing the average distance between cluster profiles (the `mrkHClust` function from `pRoloc` offers this functionality). While detailed exploration of a dataset using these and other visualisation (we also recommend the interactive visualisation apps from the `pRolocGUI` package Breckels et al. [4]) is crucial before analysing and interpreting a new spatial proteomics experiment, a detailed exploration of each of the 15 datasets used in this meta-analysis is out of the scope of this analysis.

3.3 Quantifying resolution

While visualisation of spatial proteomics data remains essential to assess the resolution, and hence the success, of a spatial proteomics experiment, it is useful to be able to objectively quantify the resolution and directly compare different experiments. Here, we present a new infrastructure, termed `QSep`, available in the `pRoloc` package Gatto et al. [11], to quantify the separation of clusters in spatial proteomics experiments. It relies on the comparison of the average euclidean distance *within* and *between* sub-cellular clusters. As

illustrated on the heatmaps in figure 3 for the *hyperLOPIT2015* data, these distances always refer to one reference marker cluster.

The raw distance matrix (figure 3, top-left) is symmetrical (i.e. the distance bee tween cluster 1 and 2 is the same as between cluster 2 and 1). Within distances are generally the smallest ones, except when two clusters overlap, as the lysosome and endosome in our example. To enable the comparison of these distances within and between experiments (see section 4 for the latter), we further divide each distance by the reference within cluster average distance (figure 3, top-right). This thus informs us as how much the average distance between cluster 1 and 2 is greater than the average distance within cluster 1 (i.e. the tightness of that cluster). At this stage, the distance matrix is not symmetrical anymore. To facilitate the comparison of distances between organelles, the distance distributions can also be visualised as boxplots (figure 3, bottom).

The rational behind these measures is as follows. Intuitively, we assess resolution by contrasting the separation between clusters (formalised by the average distance between two clusters) and the tightness of single clusters (formalised by the average within cluster distance). Ideal sub-cellular fractionation would yield tight and distant clusters, represented by a large normalised between cluster distances on figure 3.

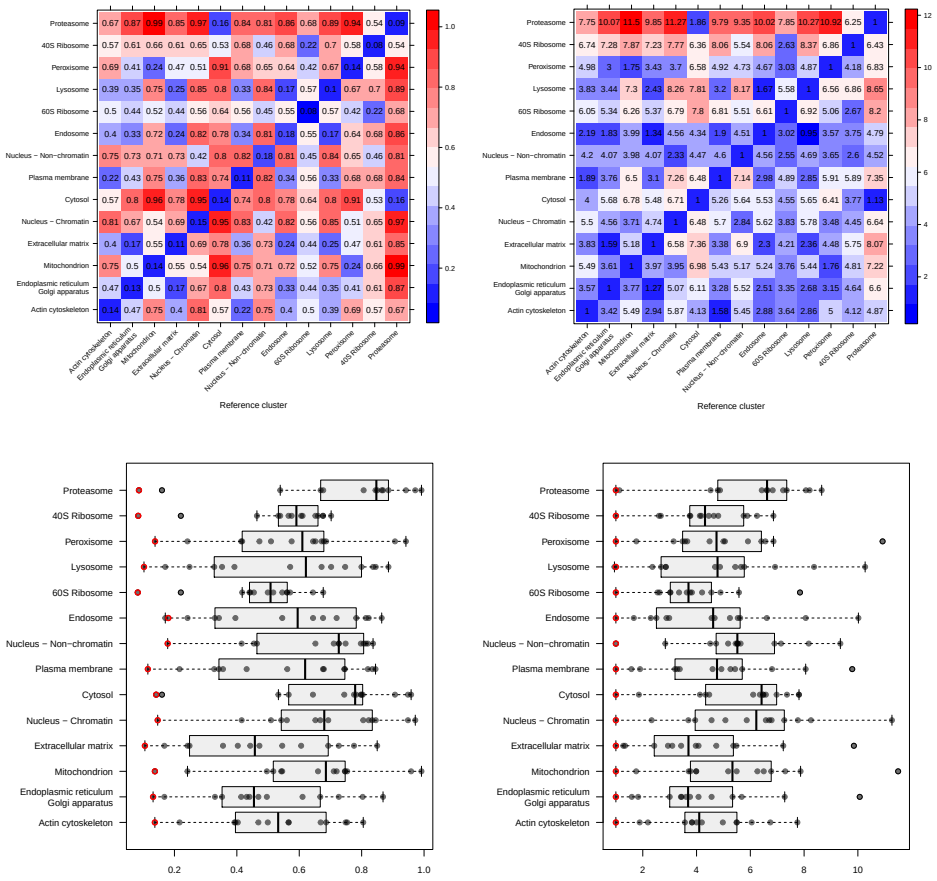

Figure 3: Quantifying resolution of the *hyperLOPIT2015* data Christoforou et al. [5]. The heatmaps at the top illustrate the raw (left) and average normalised (right) within (along the diagonal) and between euclidean cluster distances. The boxplots at the bottom summarise these same values (raw on the left, normalised on the right) to enable easier comparison between clusters, where the within distances are highlighted in red.

3.4 Application of the assessment criteria

To further demonstrate the interpretation of these resolution metrics, we directly compare the two recent global cell maps from [5] (dataset *hyperLOPIT2015*) and [14] (dataset *itzhak2016stcSILAC*). Both feature high protein coverage (7114 and 5265 proteins respectively) and good sub-cellular diversity (14 and 12 annotated clusters respectively). The former contains duplicated experiments, each made of 10 fractions and the latter contains 6 replicates with 5 fractions each. Figure 4 shows the PCA plots applying transparency to identify the underlying structure in the quantitative data and the annotated versions using the markers provided by the respective authors.

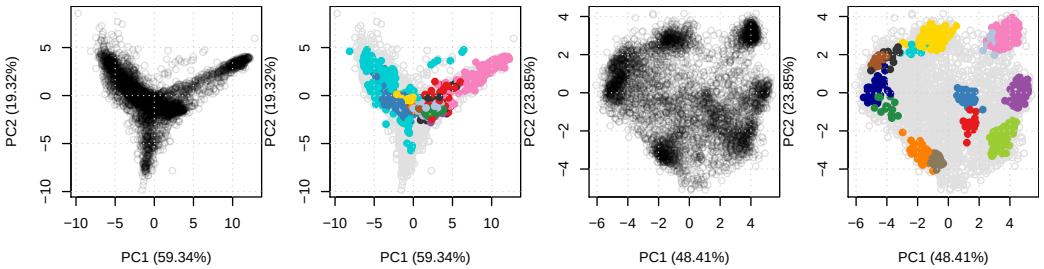

Figure 4: PCA plots for *itzhak2016stcSILAC* (left) and *hyperLOPIT2015* (right). Here, we display PC 1 and 2 for both datasets for comparability. The original authors displayed PC 1 and 3 for the *itzhak2016stcSILAC* data (see figures 7 and 8 below).

Figure 5 illustrates the normalised distance heatmaps and boxplots for the two datasets (*itzhak2016stcSILAC* at the top and *hyperLOPIT2015* at the bottom). The two heatmaps display strikingly different colour patterns. The top heatmap shows a majority of small normalised distances (blue cells) and with only a limited number of large distances (red cells), along the mitochondrial reference cluster. Conversely, the bottom heatmap displays a majority of average (white cells) and large distances (red cells) across all sub-cellular clusters. The boxplots allow to more directly compare the distances across the two datasets. On the top boxplot, we detect relatively short distances for most clusters, with most large distances stemming from the mitochondrion, leading to a median distance of 2.48. The distributions on the bottom boxplot show larger distances, equally spread among all clusters, with an media distance of 4.91.

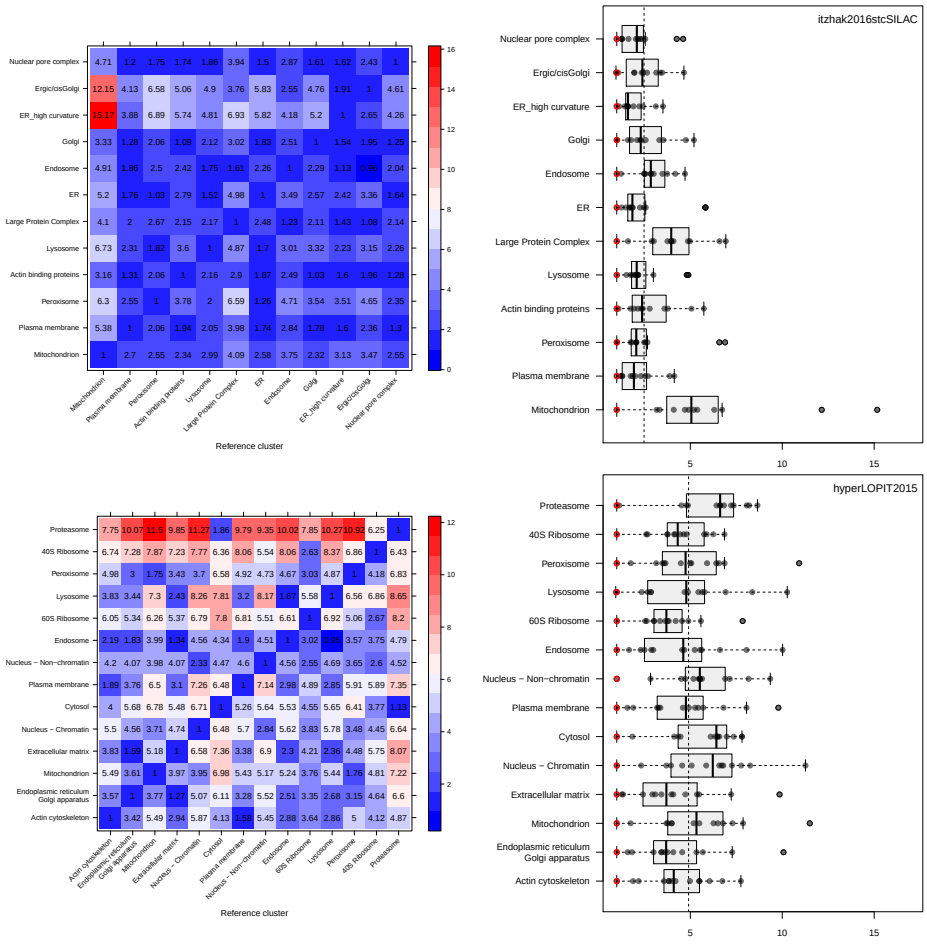

Figure 5: Contrasting quantitative separation assessment between the *itzhak2016stcSILAC* [14] (top) and *hyperLOPIT2015* [5] (bottom) datasets. The dashed vertical lines on the boxplots represent the overall media between cluster distance, 2.48 and 4.91 for *itzhak2016stcSILAC* and *hyperLOPIT2015* respectively.

4 Comparative study

We now apply the quantitative assessment of spatial resolution described in section 3.3 to compare the 15 experiments presented in section 2. On figure 6, we show, for each dataset, a boxplot illustrating the distribution of the global average normalised distances for all spatial clusters. The datasets have been ordered using the experiment-wide median between distance. It is important to always refer back to the original data when considering summarising metrics like these, to put the resolution into context; the density and annotated PCA plots discussed in section 3.2 are provided in figures 7 and 8 and the quantitative assessment boxplots and heatmaps are shown in figures 9 and 10.

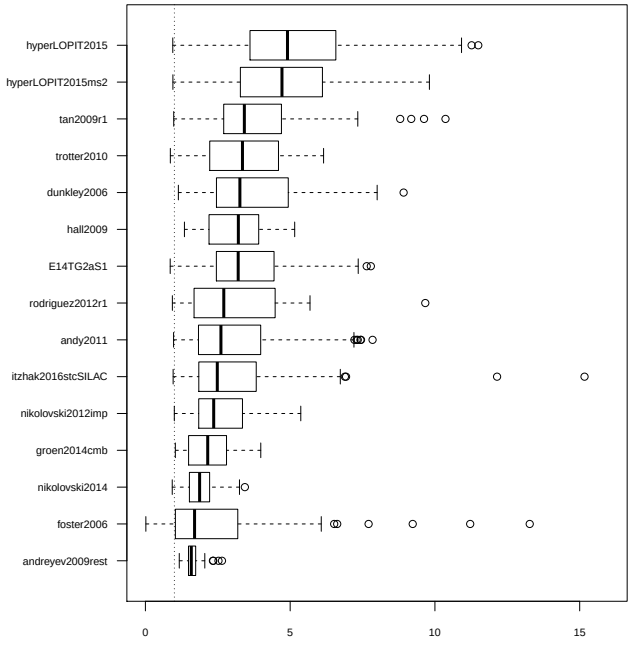

Figure 6: Quantitative separation assessment using experiment-wide normalised between cluster distances.

The hyperLOPIT experiments *hyperLOPIT2015* and *hyperLOPIT2015ms2* [5] using SPS MS^3 and conventional MS^2 show the best global, experiment-wide resolution. As documented by the authors and illustrated in section 3.2, the increased quantitation accuracy of the former result in better sub-cellular

resolution.

The next set of experiment are *tan2009r1*, *trotter2010*, *dunkley2006*, *hall2009* and *E14TG2aS1*. It is important to highlight that most of datasets (as well as *andy2011*, discussed later) have either been directly re-analysed using a semi-supervised novelty detection algorithm *phenoDisco* [2] (the only exception here being *hall2009*), or, in the case of *trotter2010*, have been annotated using markers based on the *phenoDisco* re-analysis. The novelty detection algorithm, *phenoDisco*, searches for new clusters of unlabelled proteins, using the marker proteins to guide the clustering of unlabelled features. These new clusters, termed *phenotypes*, are then validated by the user for coherence with known sub-cellular niches. This re-analysis has proven quite successful [2] and has identified previously undetected sub-cellular niches that form tight and well-resolved clusters (see for example ribosomial and trans-golgi network (TGN) in *dunkley2006*, or proteasome and nucleus in *hall2009r1* to cite only a few), which in turn favour good resolution scores.

The *hall2009* dataset is relatively poorly annotated (only 5 sub-cellular clusters, which is the lowest in all test datasets). As long as these few clusters are well separated, poor annotation can positively influence the resolution scoring. Hence, it is important to visualise the data as described in section 3.2 and consider the sub-cellular diversity exposed in the analysed data (see section 3.1).

The next set of experiments that show comparable resolution profiles are *andy2011*, *itzhak2016stcSILAC* and *nikolovski2012imp*. Note that the quantitative separation measurement is robust to questionable marker annotation. For example, the *Large Protein Complex* class defined by the original authors in the *itzhak2016stcSILAC* data could be dropped as it loosely defines many niches and thus lacks resolution. This would only marginally influence the overall assessment metrics as only the distances to/from that class are affected (i.e. 23 out of 144 distances) and as such it would not change its rank among the test datasets.

As mentioned earlier, the *groen2014cmb* and *nikolovski2014* are targeted experiment, focusing on the trans-Golgi-network and Golgi niches respectively. Such experiments do not aim for best global resolution, which is reflected by relatively low resolution.

The *foster2006* experiment displays relatively poor separation. This might be due to the relatively high number of missing values (42.4 %). Finally, the *andreyev2009rest* dataset suffers from very broad sub-cellular clusters (compared to separation between clusters).

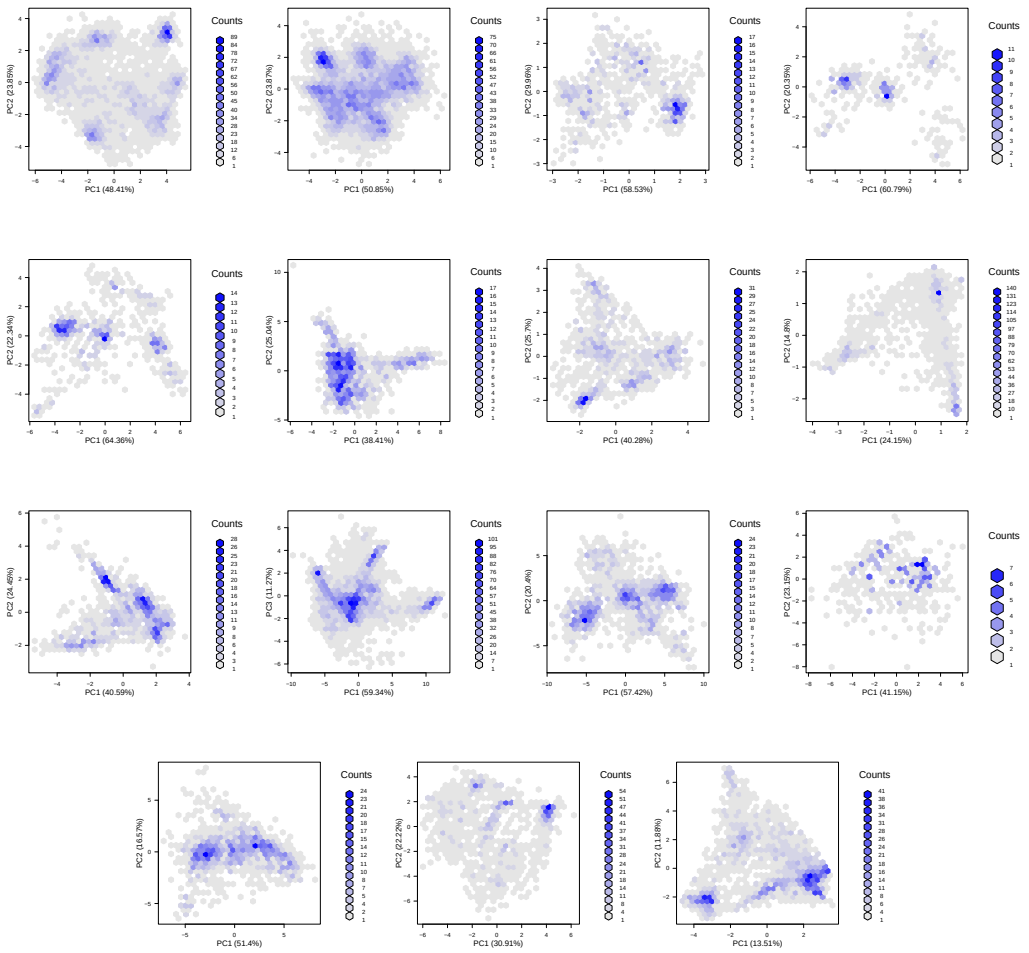

Figure 7: Density PCA plots for the 15 experiments used in this study. The experiments are ordered according to the median average between cluster distance (see figure 6). Figures have been generated using the `plot2D` function from the `pRoloc` package.

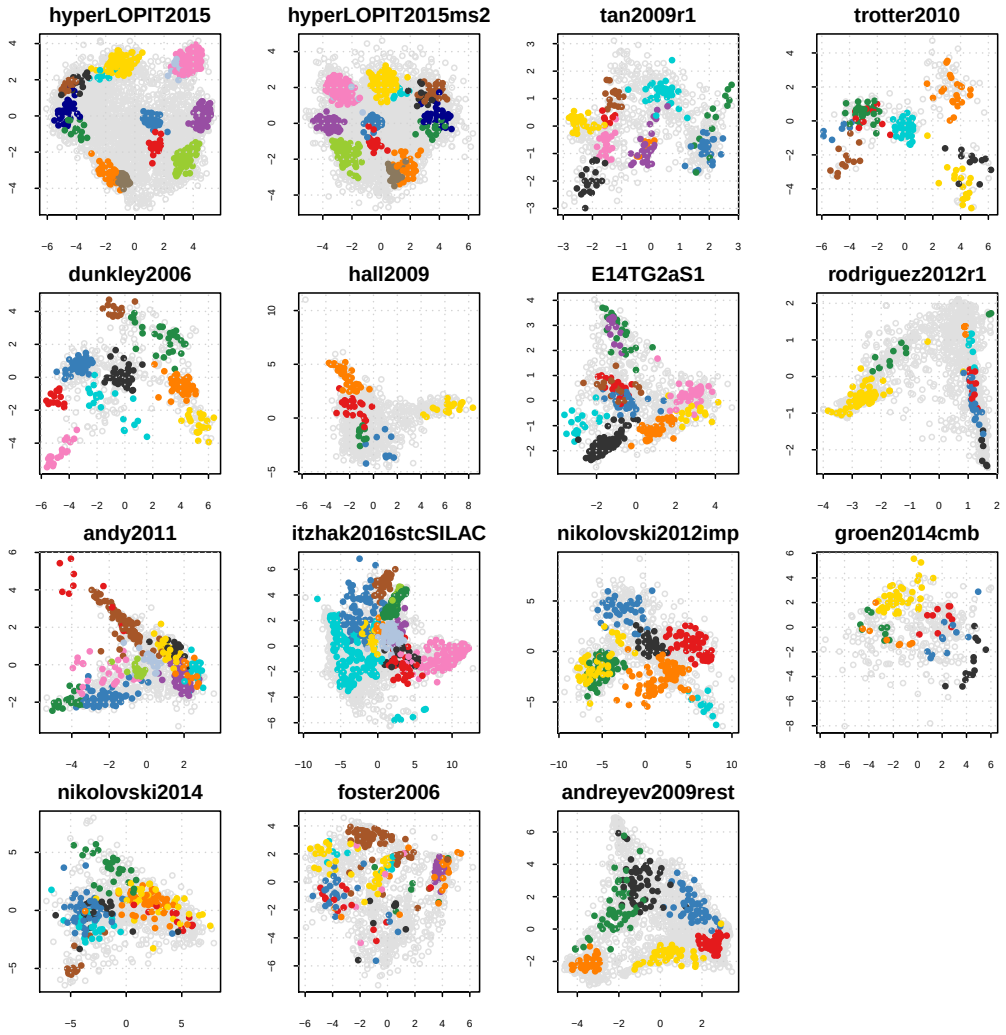

Figure 8: PCA plots for the 15 experiments used in this study. PC 1 and 2 were used except for *itzhak2016stcSILAC*, where PC 1 and 3 were used to conform to the original authors figures. The experiments are ordered according to the median average between cluster distance (see figure 6). The percentage of variance explained along the 2 PCs on the plots can be found in table 1. Figures have been generated using the `plot2D` function from the `pRoloc` package.

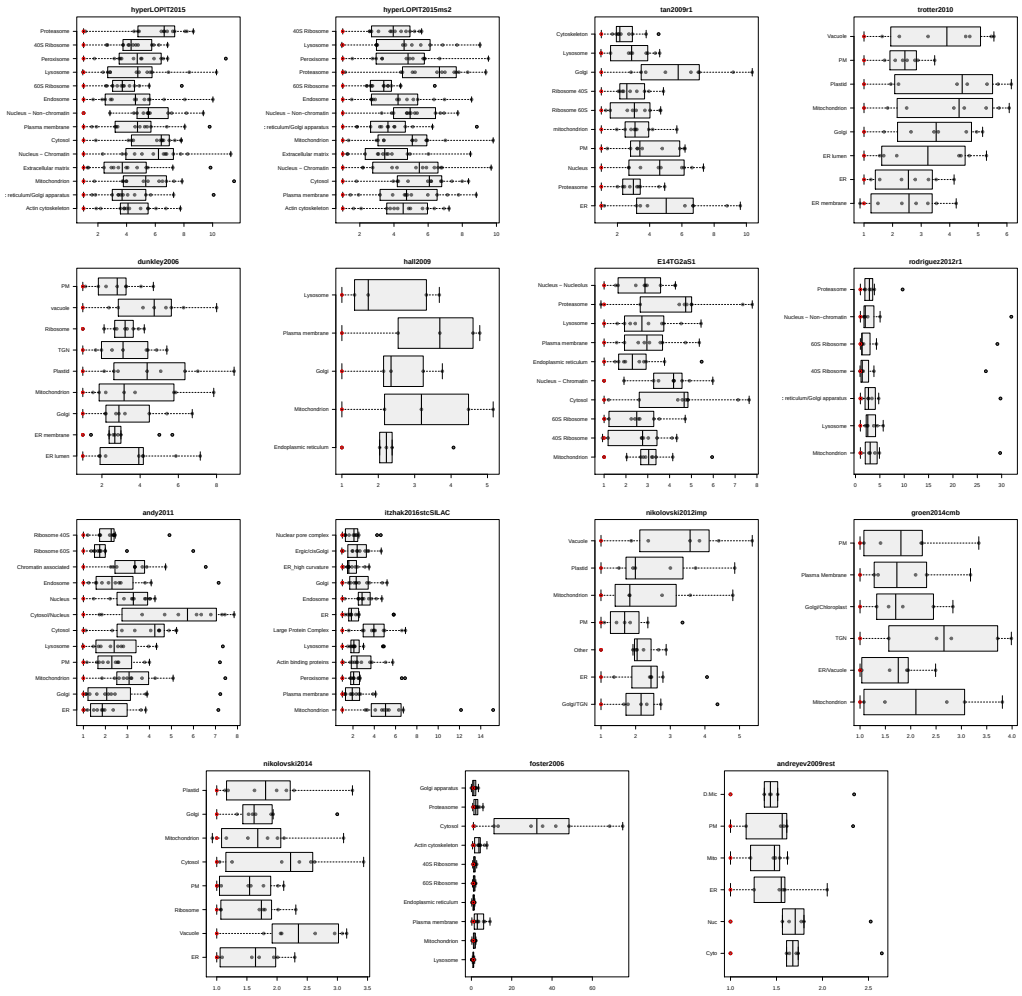

Figure 9: Quantitative separation boxplot for the 15 experiments used in this study. The experiments are ordered according to the median average between cluster distance (see figure 6).

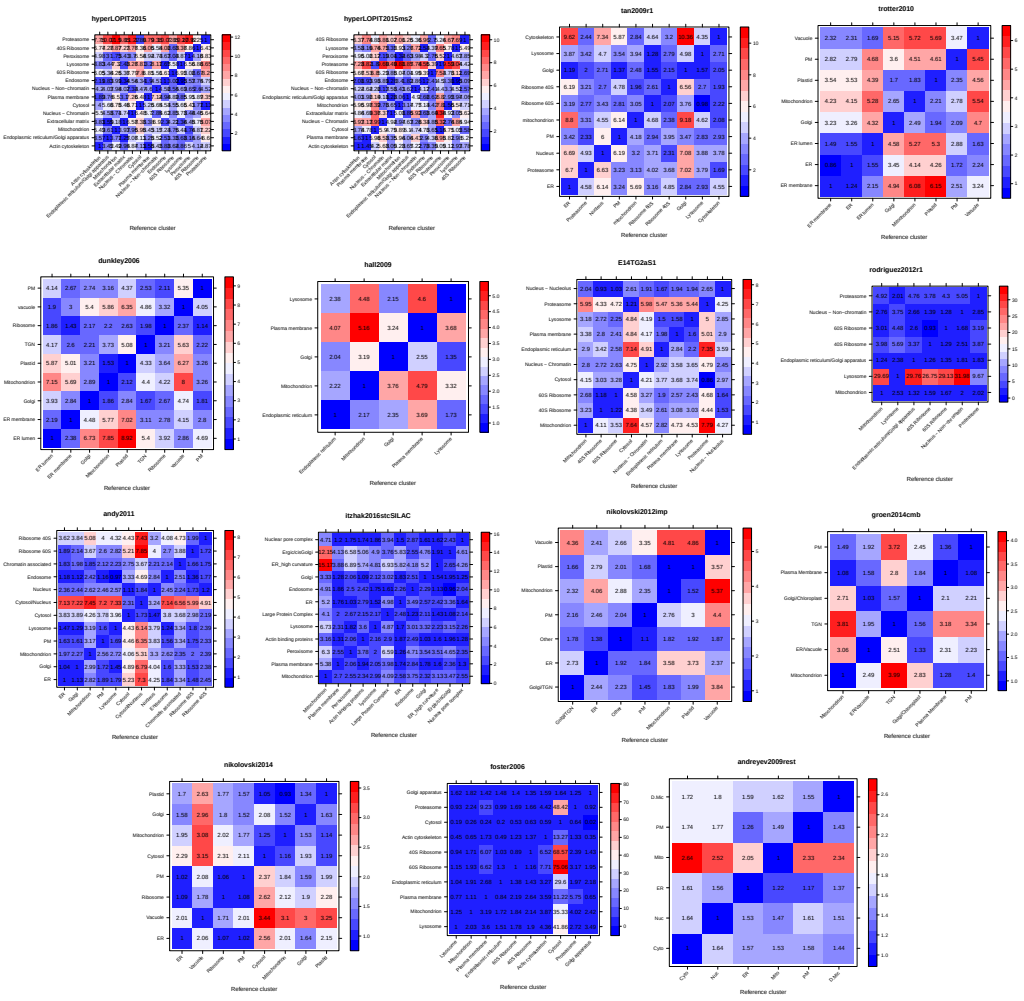

Figure 10: Quantitative separation heatmaps for the 15 experiments used in this study. The experiments are ordered according to the median average between cluster distance (see figure 6).

5 Conclusions

It is necessary to emphasise the importance and effect of marker definition on estimating and assessing the resolution of spatial proteomics experiments and, of course, the assignment of proteins to their most likely sub-cellular compartments. In this work, we have used the markers provided by the original authors (except for [7] to assess the data as originally presented.

The ordering suggested in section 4 should not be taken as absolute. Its main purpose is to provide a guide to compare different experiments. It will be useful for laboratories that do spatial studies on different models and with different fractionation and/or quantitation methods, to assess the impact of these variables (such as, for example hyperLOPIT MS^2 and MS^3). It is also useful to compare separation between different labs, as demonstrated in our comparative study (section 4). We anticipate that it will also prove useful for the researcher wanting to assess the resolution of newly published studies, and put them into a wider context.

Sub-cellular resolution is of course only one aspect of spatial proteomics, albeit an essential one, that critically determines the reliability of protein assignments to spatial niches as well as the identification of multi- and trans-localisation events.

6 Session information

The software and versions used to produce this document are summarised below. In particular, recent versions of `pRoloc` and `pRolocdata` [11], which require versions 1.13.14 and 1.11.7 or later, respectively. The source of this document enabling to reproduce all results and figures is available in the source of this document in the public manuscript repository [8].

- R version 3.3.1 Patched (2016-08-02 r71022), `x86_64-pc-linux-gnu`
- Base packages: base, datasets, graphics, grDevices, methods, parallel, stats, stats4, utils
- Other packages: annotate 1.51.0, AnnotationDbi 1.35.4, Biobase 2.33.3, BiocGenerics 0.19.2, BiocParallel 1.7.8, cluster 2.0.4, hexbin 1.27.1, IRanges 2.7.15, MLInterfaces 1.53.1, MSnbase 1.99.1, mzR 2.7.4, pRoloc 1.13.14, pRolocdata 1.11.7, ProtGenerics 1.5.1, Rcpp 0.12.7, S4Vectors 0.11.13, XML 3.98-1.4, xtable 1.8-2
- Loaded via a namespace (and not attached): affy 1.51.1, affyio 1.43.0, assertthat 0.1, base64enc 0.1-3, BiocInstaller 1.23.9, biomaRt 2.29.2, bitops 1.0-6, car 2.1-3, caret 6.0-71, class 7.3-14, codetools 0.2-14, colorspace 1.2-6, DBI 0.5, dendextend 1.3.0, DEoptimR 1.0-6, digest 0.6.10, diptest 0.75-7, doParallel 1.0.10, dplyr 0.5.0, e1071 1.6-7, evaluate 0.9, flexmix 2.3-13, FNN 1.1, foreach 1.4.3, formatR 1.4, fpc 2.1-10, gbm 2.1.1, gdata 2.17.0, genefilter 1.55.2, ggplot2 2.1.0, ggvis 0.4.3, grid 3.3.1, gtable 0.2.0, gtools 3.5.0, highr 0.6, htmltools 0.3.5, htmlwidgets 0.7, httpuv 1.3.3, hwriter 1.3.2, impute 1.47.0, iterators 1.0.8, jsonlite 1.0, kernlab 0.9-24, knitr 1.14, lattice 0.20-34, limma 3.29.21, lme4 1.1-12, lpSolve 5.6.13, magrittr 1.5, MALDIquant 1.15, MASS 7.3-45, Matrix 1.2-7.1, MatrixModels 0.4-1, mclust 5.2, mgcv 1.8-14, mime 0.5, minqa 1.2.4, mlbench 2.1-1, modeltools 0.2-21, munsell 0.4.3, mvtnorm 1.0-5, mzID 1.11.2, nlme 3.1-128, nloptr 1.0.4, nnet 7.3-12, pbkrtest 0.4-6, pcaMethods 1.65.0, pls 2.5-0, plyr 1.8.4, prabclus 2.2-6, preprocessCore 1.35.0, proxy 0.4-16, quantreg 5.29, R6 2.1.3, randomForest 4.6-12, RColorBrewer 1.1-2, RCurl 1.95-4.8, rda 1.0.2-2, reshape2 1.4.1, robustbase 0.92-6, rpart 4.1-10, RSQLite 1.0.0, sampling 2.7, scales 0.4.0, sfsmisc 1.1-0, shiny 0.13.2,

SparseM 1.72, splines 3.3.1, stringi 1.1.1, stringr 1.1.0, survival 2.39-5, threejs 0.2.2, tibble 1.2, tools 3.3.1, trimcluster 0.1-2, vsn 3.41.0, whisker 0.3-2, zlibbioc 1.19.0

Acknowledgements

L.M.B is supported by a Wellcome Trust Technology Development Grant (Grant number 108467/Z/15/Z). L.G. is supported by the BBSRC Strategic Longer and Larger grant (Award BB/L002817/1). The authors would like to thank Dr Claire M. Mulvey for helpful comments on the quantitative assessment.

References

[1] A Y Andreyev, Z Shen, Z Guan, A Ryan, E Fahy, S Subramaniam, C R Raetz, S Briggs, and E A Dennis. Application of proteomic marker ensembles to subcellular organelle identification. *Mol Cell Proteomics*, 9(2):388–402, Feb 2010. doi: 10.1074/mcp.M900432-MCP200.

[2] L M Breckels, L Gatto, A Christoforou, A J Groen, K S Lilley, and M W Trotter. The effect of organelle discovery upon sub-cellular protein localisation. *J Proteomics*, 88:129–40, Aug 2013. doi: 10.1016/j.jprot.2013.02.019.

[3] L M Breckels, S B Holden, D Wojnar, C M Mulvey, A Christoforou, A Groen, M W Trotter, O Kohlbacher, K S Lilley, and L Gatto. Learning from heterogeneous data sources: An application in spatial proteomics. *PLoS Comput Biol*, 12(5):e1004920, May 2016. doi: 10.1371/journal.pcbi.1004920.

[4] Lisa M Breckels, Thomas Naake, and Laurent Gatto. *pRolocGUI: Interactive visualisation of spatial proteomics data*. URL `http://ComputationalProteomicsUnit.github.io/pRolocGUI/`. R package version 1.7.5.

[5] A Christoforou, C M Mulvey, L M Breckels, A Geladaki, T Hurrell, P C Hayward, T Naake, L Gatto, R Viner, A Martinez Arias, and K S Lilley.

A draft map of the mouse pluripotent stem cell spatial proteome. *Nat Commun*, 7:8992, 2016. doi: 10.1038/ncomms9992.

[6] T P. J. Dunkley, S Hester, I P Shadforth, J Runions, T Weimar, S L Hanton, J L Griffin, C Bessant, F Brandizzi, C Hawes, R B Watson, P Dupree, and K S Lilley. Mapping the arabidopsis organelle proteome. *Proc Natl Acad Sci USA*, 103(17):6518–6523, Apr 2006. doi: 10.1073/pnas.0506958103.

[7] L J Foster, C L de Hoog, Y Zhang, Y Zhang, X Xie, V K. Mootha, and M Mann. A mammalian organelle map by protein correlation profiling. *Cell*, 125(1):187–199, Apr 2006. doi: 10.1016/j.cell.2006.03.022.

[8] L Gatto. Assessing sub-cellular resolution in spatial proteomics experiments. `https://github.com/ComputationalProteomicsUnit/QSep-manuscript/`, 2016.

[9] L Gatto, J A Vizcaíno, H Hermjakob, W Huber, and K S Lilley. Organelle proteomics experimental designs and analysis. *Proteomics*, 10 (22):3957–69, Nov 2010. doi: 10.1002/pmic.201000244.

[10] L Gatto, L M Breckels, T Burger, D J Nightingale, A J Groen, C Campbell, N Nikolovski, C M Mulvey, A Christoforou, M Ferro, and K S Lilley. A foundation for reliable spatial proteomics data analysis. *Mol Cell Proteomics*, 13(8):1937–52, Aug 2014. doi: 10.1074/mcp.M113.036350.

[11] L Gatto, L M Breckels, S Wieczorek, T Burger, and K S Lilley. Mass-spectrometry-based spatial proteomics data analysis using pRoloc and pRolocdata. *Bioinformatics*, 30(9):1322–4, May 2014. doi: 10.1093/bioinformatics/btu013.

[12] A J Groen, G Sancho-Andrs, L M Breckels, L Gatto, F Aniento, and K S Lilley. Identification of trans-golgi network proteins in arabidopsis thaliana root tissue. *J Proteome Res*, 13(2):763–76, Feb 2014. doi: 10.1021/pr4008464.

[13] S L Hall, S Hester, J L Griffin, K S Lilley, and A P Jackson. The organelle proteome of the dt40 lymphocyte cell line. *Mol Cell Proteomics*, 8(6):1295–1305, Jun 2009. doi: 10.1074/mcp.M800394-MCP200.

[14] D N Itzhak, S Tyanova, J Cox, and G H Borner. Global, quantitative and dynamic mapping of protein subcellular localization. *Elife*, 5, 2016. doi: 10.7554/eLife.16950.

[15] N Nikolovski, D Rubtsov, M P Segura, G P Miles, T J Stevens, T P Dunkley, S Munro, K S Lilley, and P Dupree. Putative glycosyltransferases and other plant golgi apparatus proteins are revealed by LOPIT proteomics. *Plant Physiol*, 160(2):1037–51, Oct 2012. doi: 10.1104/pp.112.204263.

[16] N Nikolovski, P V Shliaha, L Gatto, P Dupree, and K S Lilley. Label free protein quantification for plant golgi protein localisation and abundance. *Plant Physiol*, Aug 2014. doi: 10.1104/pp.114.245589.

[17] A M Rodrguez-Pieiro, S van der Post, M E Johansson, K A Thomsson, A I Nesvizhskii, and G C Hansson. Proteomic study of the mucin granulae in an intestinal goblet cell model. *J Proteome Res*, 11(3):1879–90, Mar 2012. doi: 10.1021/pr2010988.

[18] DJL Tan, H Dvinge, A Christoforou, P Bertone, A Martinez Arias, and KS Lilley. Mapping organelle proteins and protein complexes in drosophila melanogaster. *J Proteome Res*, 8(6):2667–2678, Jun 2009. doi: 10.1021/pr800866n.

[19] M Tomizioli, C Lazar, S Brugire, T Burger, D Salvi, L Gatto, L Moyet, L M Breckels, A M Hesse, K S Lilley, D Seigneurin-Berny, G Finazzi, N Rolland, and M Ferro. Deciphering thylakoid sub-compartments using a mass spectrometry-based approach. *Mol Cell Proteomics*, 13(8):2147–67, Aug 2014. doi: 10.1074/mcp.M114.040923.

[20] M W B Trotter, P G Sadowski, T P J Dunkley, A J Groen, and K S Lilley. Improved sub-cellular resolution via simultaneous analysis of organelle proteomics data across varied experimental conditions. *PROTEOMICS*, 10(23):4213–4219, 2010. ISSN 1615-9861. doi: 10.1002/pmic.201000359.