

# A standardized computational framework for the analysis of mass spectrometry-based single-cell proteomics data

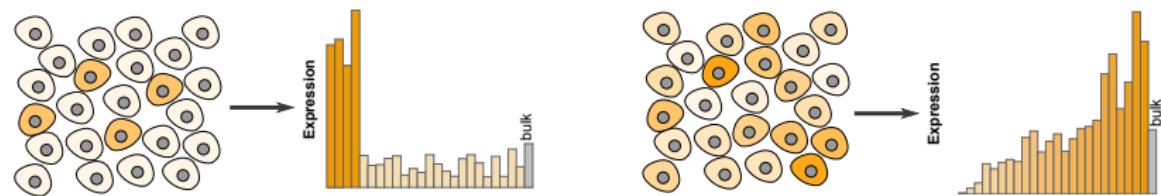
Christophe Vanderaa, Laurent Gatto

December 18, 2020



# Single-cell technology

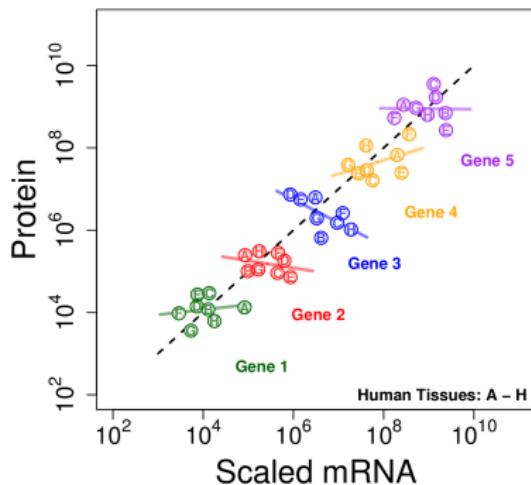
**Single-cell omics** generate measurements for one cell



scRNA-Seq is maturing very quickly with several standardized pipelines, e.g. OSCA book (Amezquita et al. (2019))

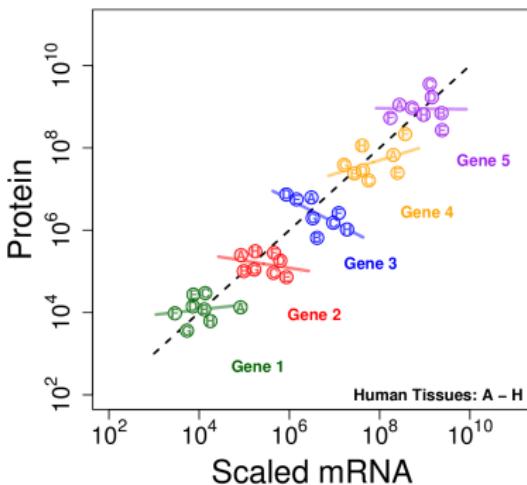
**BUT** cellular function is driven by proteins

# Single-cell proteomics



Source: Specht and Slavov (2018)

# Single-cell proteomics



Source: Specht and Slavov (2018)

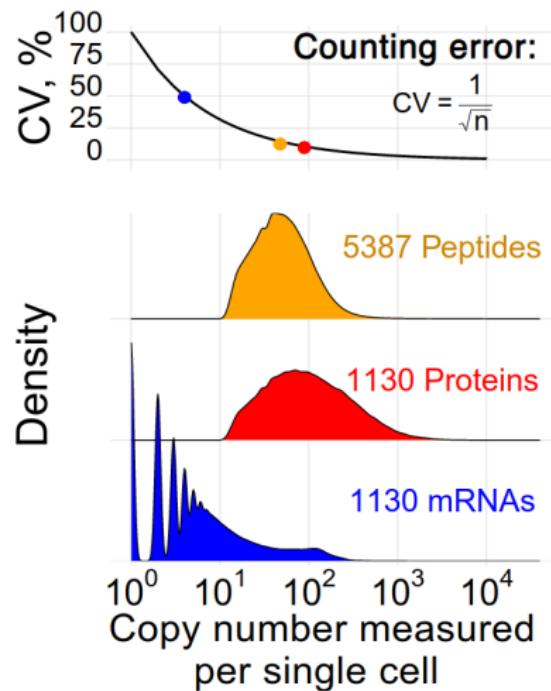
## Flow cytometry or CITE-Seq

- ▶ High throughput
- ▶ High sensitivity

## BUT antibody based:

- ▶ Targeted approach
- ▶ Specificity issues

# MS-based single-cell proteomics I

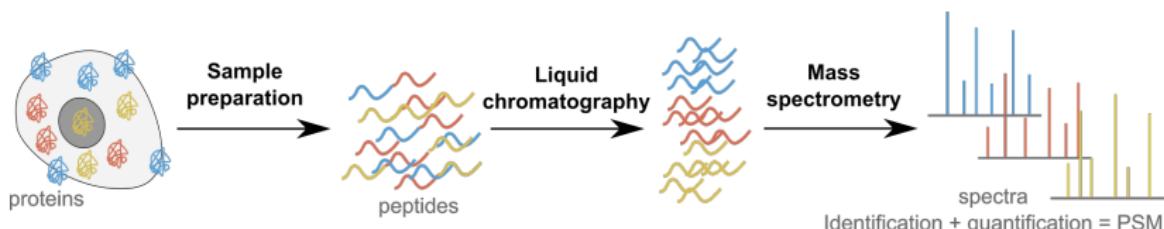


Source: Specht et al. (2020)

# MS-based single-cell proteomics II

Recent milestone : SCoPE2 protocol (Specht et al. (2020))

- ▶ ~ 1500 single cells
- ▶ ~ 3000 quantified proteins
- ▶ ~ 600 proteins expressed per cell



Between 1 (label-free) and 10 (TMT multiplexing) single cells per run

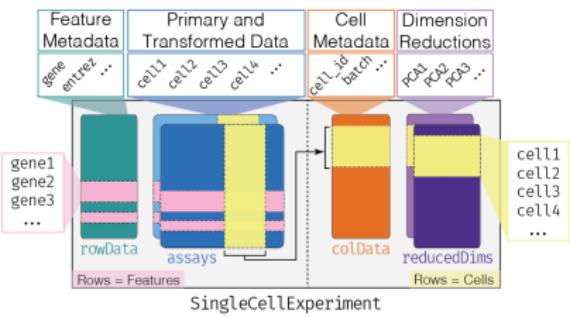
# Our contribution

We offer a solution to the **lack of good computational tools** for handling SCP data.

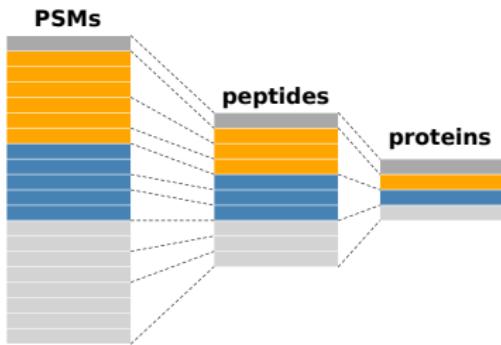
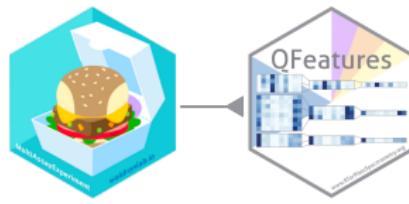
- ▶ SCP data framework combines existing Bioconductor classes
- ▶ `scpdata` disseminates curated SCP data sets for method development and benchmarking (soon on Bioconductor)
- ▶ `scp` implements functions to streamline the analysis of SCP data (available on Bioconductor)

# Data framework I

## Single-cell

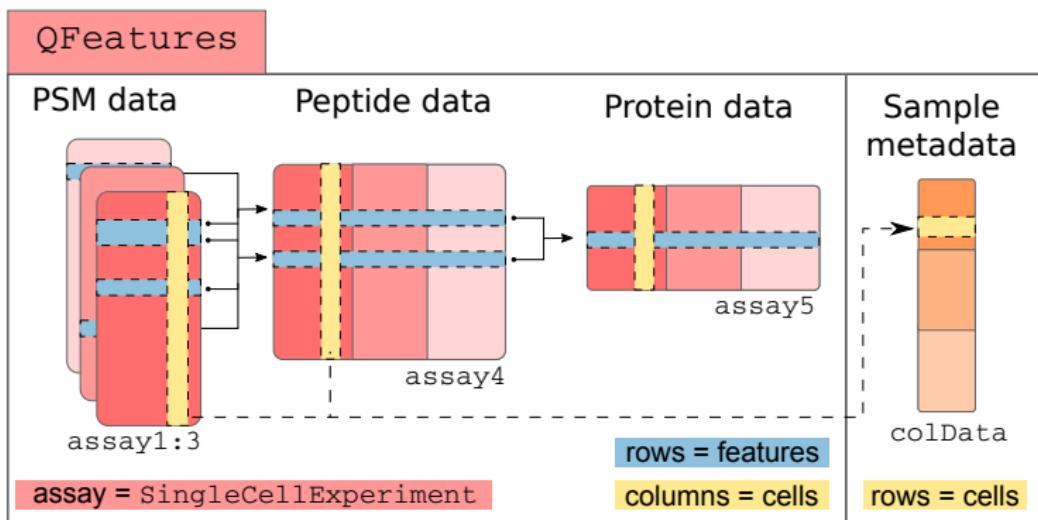
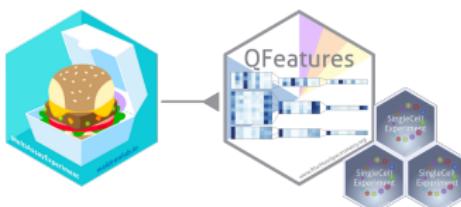


## Proteomics



# Data framework II

## SCP data framework



# scpdata

The package currently contains 11 datasets accessible through `ExperimentHub`



Example: the SCoPE2 dataset

```
1 library(scpdata)
2 scpd <- specht2019v3()
```

Overview of the dataset:

```
1 show(scpd)
2 An instance of class QFeatures containing 179 assays:
3 [1] 190222S_LCA9_X_FP94AA: SingleCellExperiment with 2823 rows and 11 columns
4 [2] 190222S_LCA9_X_FP94AB: SingleCellExperiment with 4297 rows and 11 columns
5 [3] 190222S_LCA9_X_FP94AC: SingleCellExperiment with 4956 rows and 11 columns
6 ...
7 [177] 191110S_LCB7_X_APNOV16plex2_Set_9: SingleCellExperiment with 4626 rows and 16
8 [178] peptides: SingleCellExperiment with 9208 rows and 1018 columns
9 [179] proteins: SingleCellExperiment with 2772 rows and 1018 columns
```

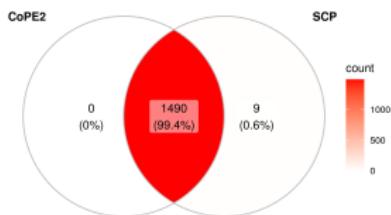
## SCoPE2 pipeline:

1. Input data
2. QC on features
3. Peptide aggregation
4. QC on samples
5. Log-normalization
6. Feature selection
7. Protein aggregation
8. Imputation

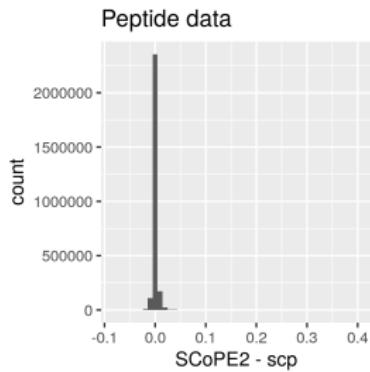
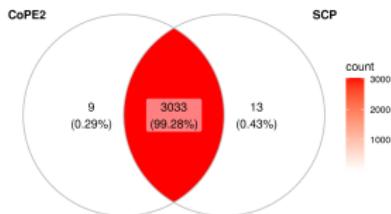
```
1 readSCP(quantTable = quantData ,  
2         metaTable = metaData ,  
3         channelCol = "Channel" ,  
4         batchCol = "Set") %>%  
5 zeroIsNA(i = 1:4) %>%  
6 computeSCR(i = 1:4 ,  
7             colDataCol = "SampleType" ,  
8             carrierPattern = "Carrier" ,  
9             samplePattern = "Monocyte") %>%  
10 filterFeatures(~ Potential.contaminant != "+" &  
11                 .meanSCR < 0.1) %>%  
12 subsetByAssay(dims(.)[1, ] > 150) %>%  
13 joinAssays(i = 4:6, name = "peptides") %>%  
14 computeMedianCV(i = 1:3 ,  
15                 proteinCol = "protein" ,  
16                 peptideCol = "peptide") %>%  
17 subsetByColumn(.\$MedianCV < 0.4) %>%  
18 normalize(i = "peptides" ,  
19             method = "median" , na.rm = TRUE) %>%  
20 logTransform(i = "normAssay" ,  
21                 base = 2) %>%  
22 aggregateFeatures(i = "logAssay" ,  
23                     name = "proteins" ,  
24                     fcol = "protein") %>%  
25 impute(i = "normAssay" ,  
26                 method = "knn") ->  
27 scp
```

# SCoPE2 replication results

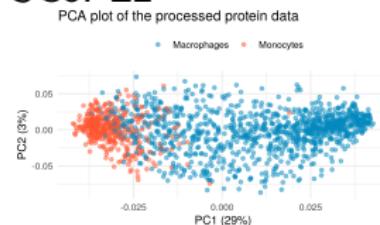
## Cell filter



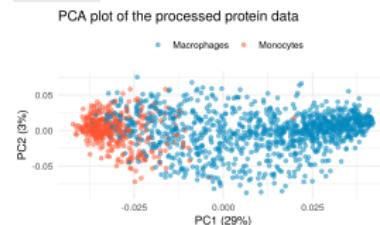
## Protein filter



## ScCoPE2



## scp



# Replication progress

We used `scp` to reproduce 2 SCP analyses:

1. Successful replication of SCoPE2 analysis (multiplexed) + identification of **issues and improvements** to the analysis
2. Failed replication of Zhu et al. (2019) (label-free): lack of good documentation

This demonstrate the successful application of our software to various SCP datasets.



Preprint is being written

# Single-cell challenges



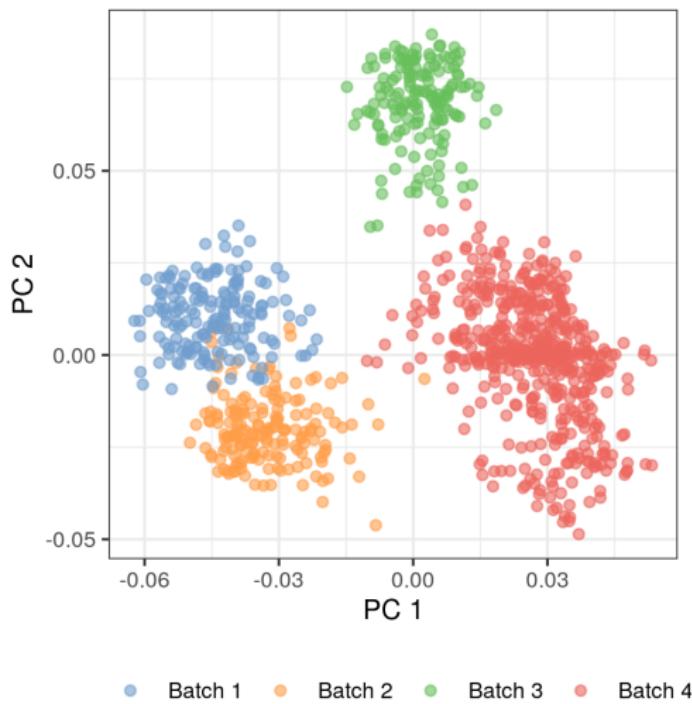
# Single-cell challenges



- ▶ Technical challenges: automation, minute sample amount, cost per cell
- ▶ Computational challenges: big data, missingness, complex batch effects
- ▶ Conceptual challenges: what is a cell type? what is biologically relevant?

# SCP challenges: batch effect

Batch effect on quantification

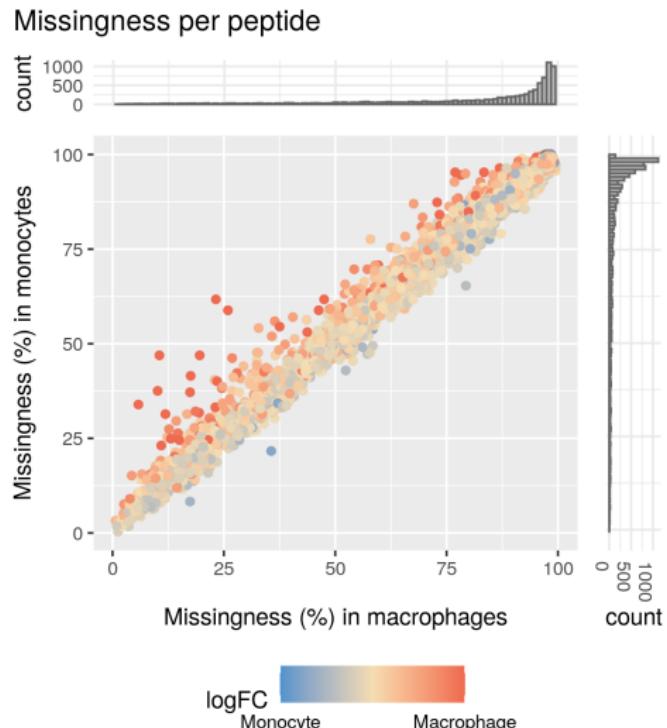


- ▶ Sample collection batch
- ▶ Chromatographic batch
- ▶ Mass spectrometer maintenance

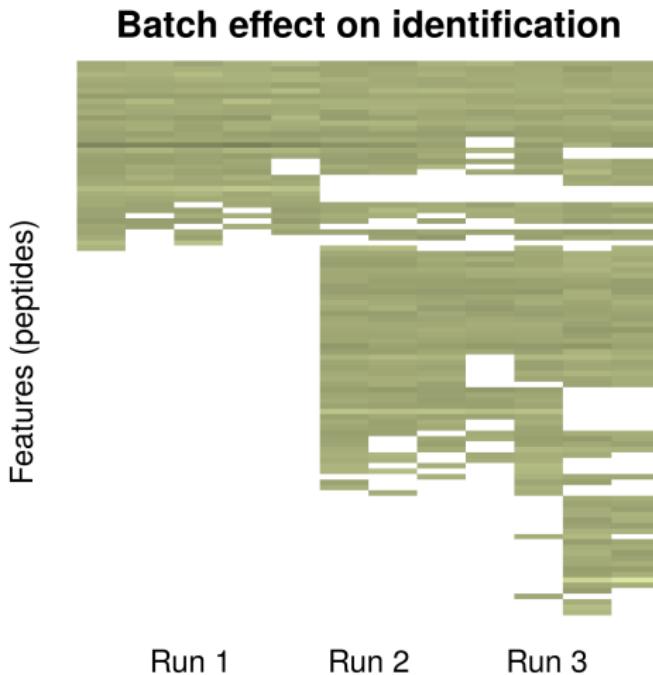
SCoPE2 analysis removed batch effects using

ComBat

# SCP challenges: missingness



# SCP challenges: batch effect + missingness



# Takehome message

- ▶ SCP is an emerging but very promising field!
- ▶ We developed a computational infrastructure to formalize SCP data analyses
- ▶ The infrastructure could be applied to reproduce 2 published analyses
- ▶ Exciting challenges are yet to be solved

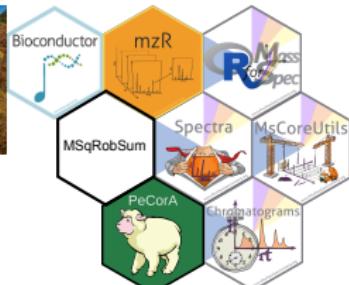
Single-cell



scp framework



Proteomics



# Acknowledgements

- ▶ UCLouvain-CBIO lab: Laurent Gatto
- ▶ Slavov lab: Nikolai Slavov, Harrison Specht
- ▶ Bioconductor team: Lori Shepherd
- ▶ Bioconductor and EuroBioc community

Thank you for your attention!  
I'm happy to take questions now or at the discussion tables



# References I

- Robert A Amezquita, Aaron T L Lun, Etienne Becht, Vince J Carey, Lindsay N Carpp, Ludwig Geistlinger, Federico Martini, Kevin Rue-Albrecht, Davide Risso, Charlotte Soneson, Levi Waldron, Hervé Pagès, Mike L Smith, Wolfgang Huber, Martin Morgan, Raphael Gottardo, and Stephanie C Hicks. Orchestrating single-cell analysis with bioconductor. *Nat. Methods*, pages 1–9, December 2019.
- Harrison Specht and Nikolai Slavov. Transformative opportunities for Single-Cell proteomics. *J. Proteome Res.*, 17(8):2565–2571, August 2018.
- Harrison Specht, Edward Emmott, Aleksandra A Petelski, R Gray Huffman, David H Perlman, Marco Serra, Peter Kharchenko, Antonius Koller, and Nikolai Slavov. Single-cell proteomic and transcriptomic analysis of macrophage heterogeneity. October 2020.
- Ying Zhu, Mirko Scheibinger, Daniel Christian Ellwanger, Jocelyn F Krey, Dongseok Choi, Ryan T Kelly, Stefan Heller, and Peter G Barr-Gillespie. Single-cell proteomics reveals changes in expression during hair-cell development. *Elife*, 8, November 2019.