Single cell proteomics with Mascot

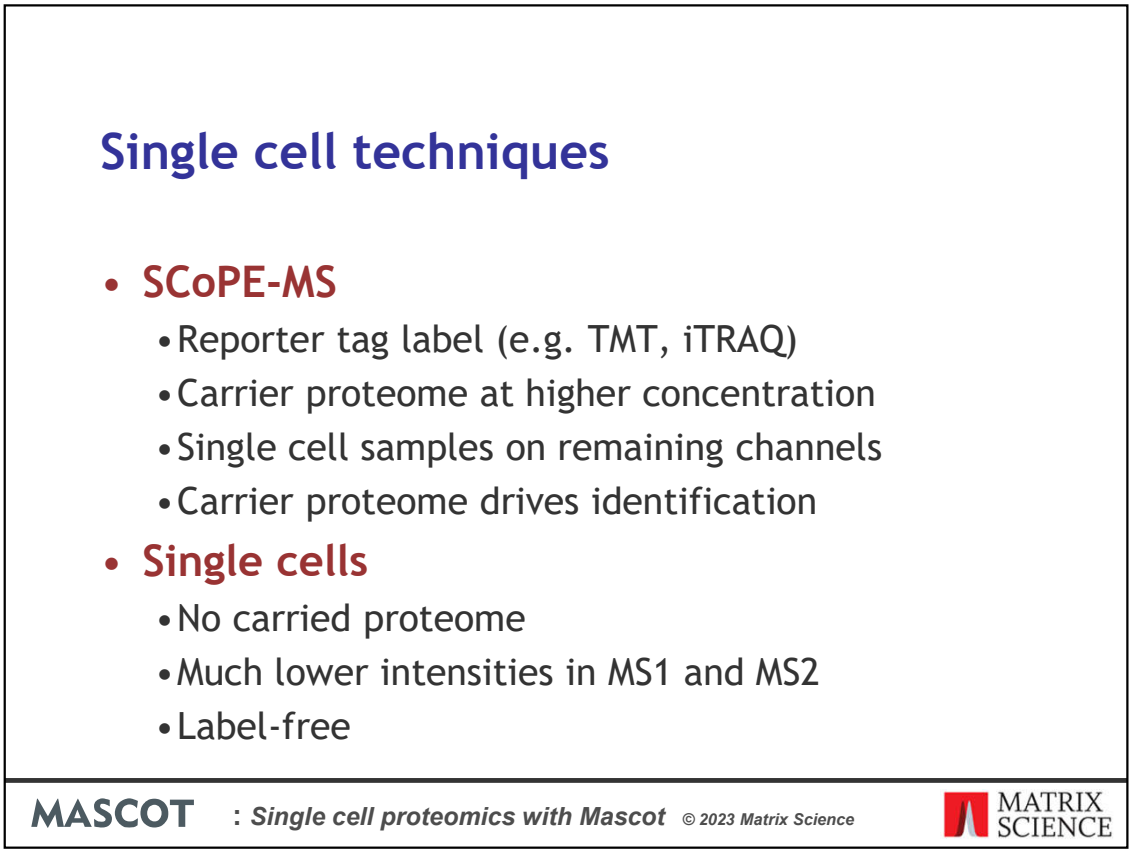
Patrick Emery
Matrix Science Ltd

Bulk vs. Single Cell studies

- **Bulk tissue analysis**
  - Averages differences between cell diversity
- **Single-cell analysis**
  - Characterisation of individual cells

Traditionally, all Mass-spec based proteomics experiments have been based on bulk tissue/sample analysis, meaning the data are derived from multiple source cells, and the results are the averages between the included cells

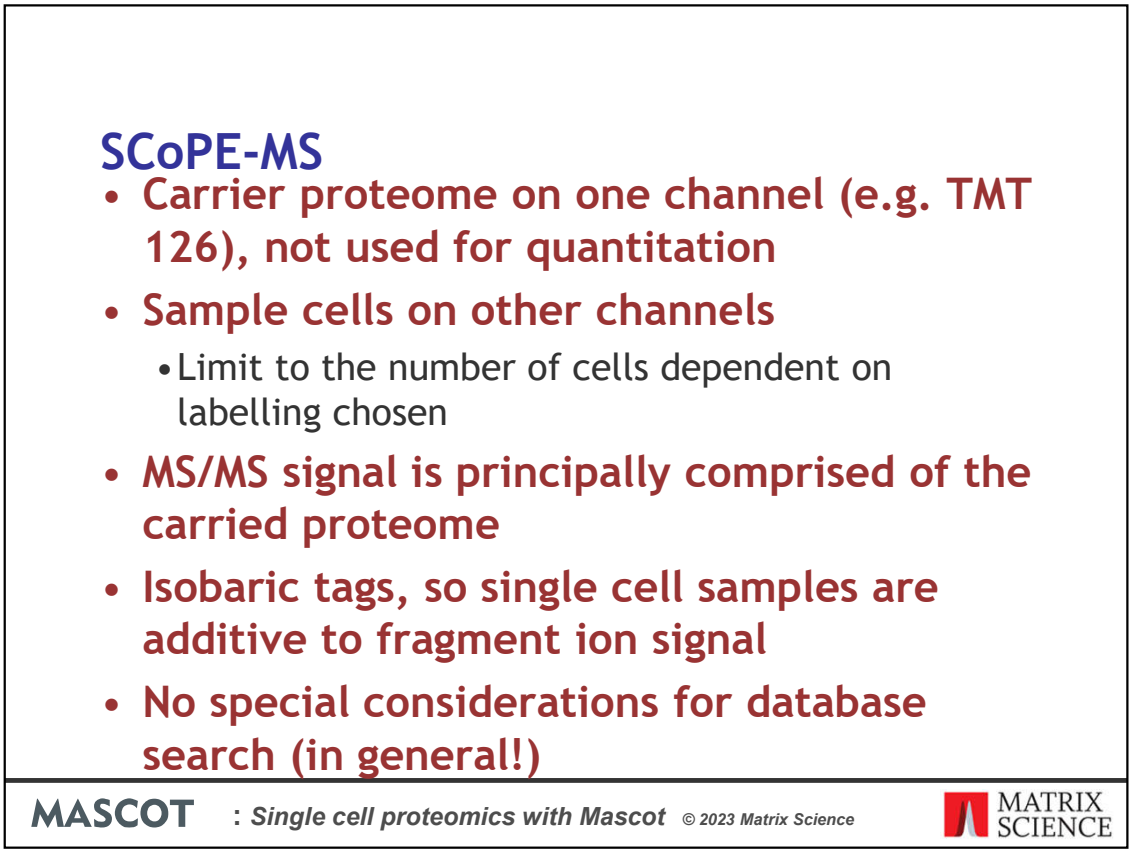
Recent improvements in instrument sensitivity, often with additional ion mobility separation, combined with accurate cell sorting has enabled a rapid growth in the field of single-cell proteomics, allowing for the characterisation of individual cells.



The two main approaches to single cell proteomics analyses that you are likely to encounter are:

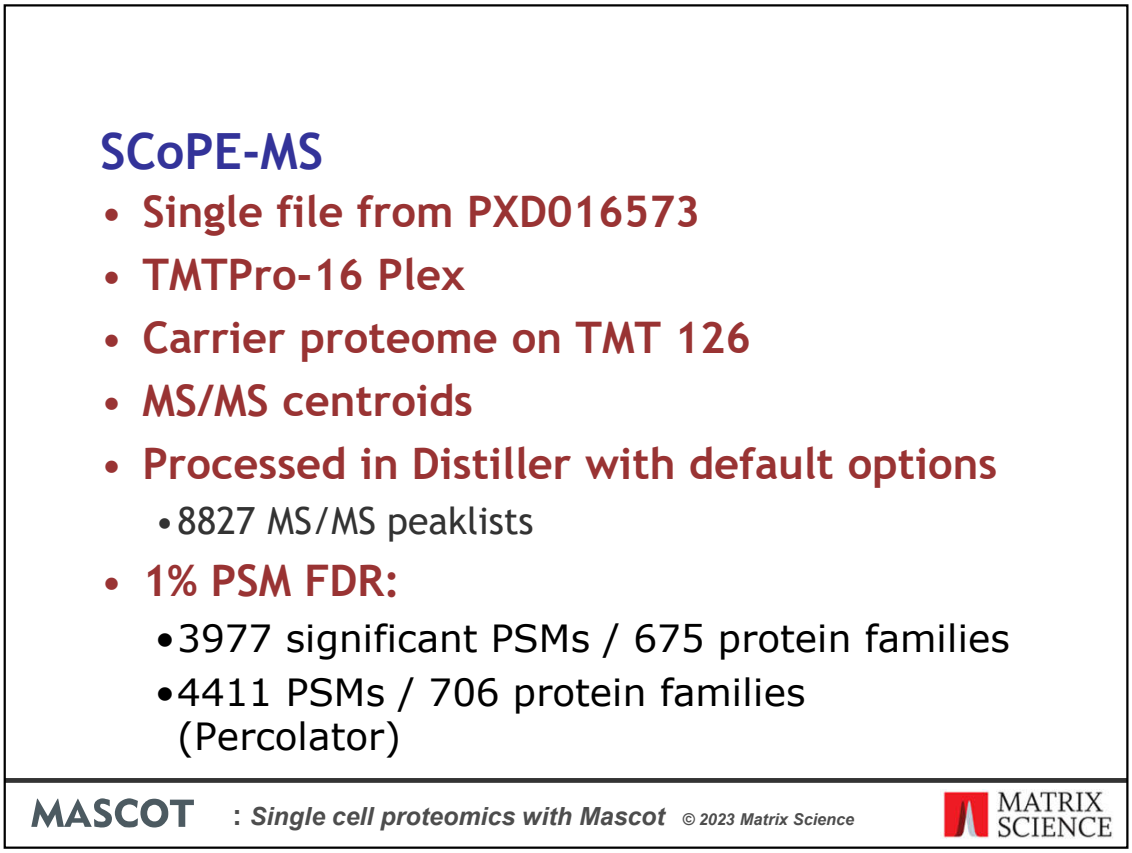
Single Cell ProtEomics by Mass Spectrometry (SCoPE-MS). With this approach, an isobaric label such as iTRAQ or TMT is used. A carried proteome is labelled with one channel. This carrier proteome is derived from multiple cells and is therefore at a much higher protein concentration than the individual cells which are labelled with the remaining channels. The carrier proteome is used to give more typical fragmentation spectra, avoiding any issues associated with low intensity spectra from the single cells and drive identification – it is generally not used for quantitation, which is confined to the single-cell channels.

The alternative approach is to carry out analysis directly on single cell samples, with no carried proteome. This results in much lower intensity spectra – both in the survey and MS/MS scans – but it does allow the use of label-free quantitation, avoiding the requirement for reporter ion labels.



As mentioned, for a SCoPE-MS experiment, the carrier proteome is present on one of the reporter ion tags and provides the bulk of the fragment ion signal in the MS/MS scans. The single cells, labelled on the remaining channels, are additive to this signal because the tags are isobaric of course. The result is that the MS/MS spectra generated in a SCoPE-MS experiment look very similar to equivalent spectra from a typical bulk analysis – and in general we can treat these data in much the same way.

One limitation is of course that the chosen labelling method will therefore also put a limit on the number of cells which can be studied in the run.



As an example, we took a single file from a SCoPE-MS dataset in the PRIDE public repository. This is a TMTPro-16 Plex dataset with the carrier proteome on the 126 channel.

The MS/MS spectra have been saved as centroids, so we carried out peak picking in Mascot Distiller using the default options – which simply take the centroids from the MS/MS scans.

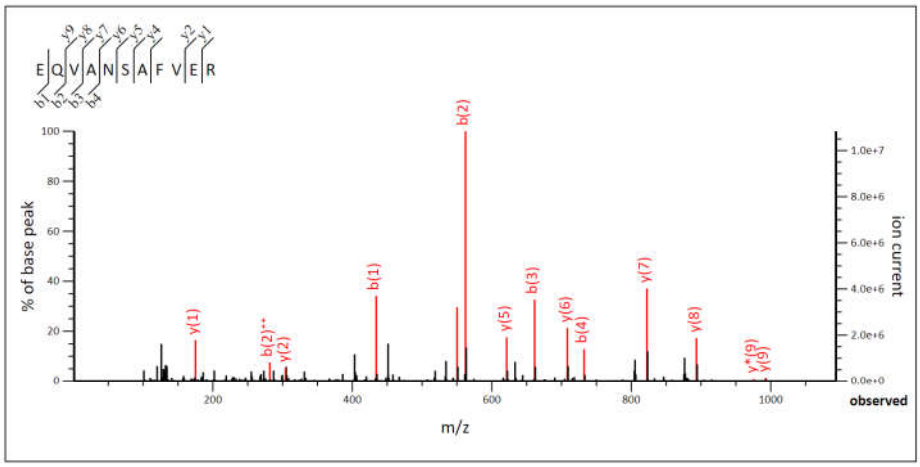
This resulted in 8827 peaklists. These were searched with Mascot Server 2.8 using the search settings described in the project. Results were filtered for 1% FDR at the Peptide Spectrum Match level. This yielded 3977 significant PSMs with scores above the homology threshold across 675 protein families.

Enabling Percolator rescoring further improved the results, giving us 4411 significant PSMs and 706 protein families.

So, we're confidently identifying close to 50% of the peaklists, which is a good yield and we've not had to do anything "special" to process the data.

MS/MS Fragmentation of **EQVANSAFVER**
Found in **HS90B_HUMAN** in **SwissProt**, Heat shock protein HSP 90-beta OS=Homo sapiens OX=9606 GN=HSP90AB1 PE=1 SV=4

Match to Query 4875: 1552.833072 from(518.618300,3+) intensity(76402696.0000) scans(4374) rawscans(sn4374) rtinseconds(1072.087) index(2588)
Title: 2589: Scan 4374 (rt=1072.09) [E:\PXD016573\8-plex-16x-2e5-500ms_Fig3f.raw]
Data file 8-plex-16x-2e5-500ms_Fig3f.temp.mgf



1.07 to 1093.53

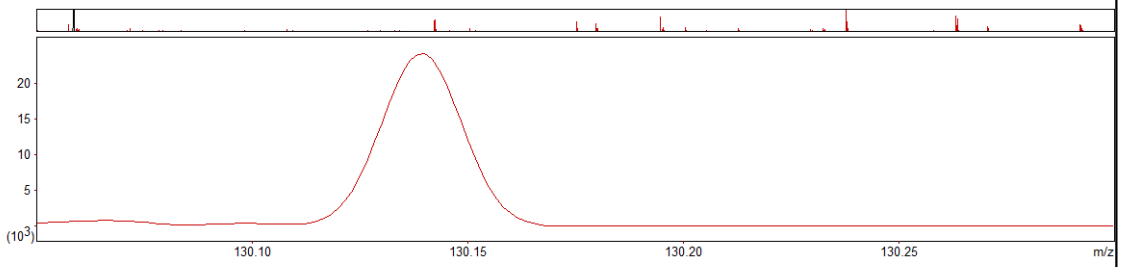
Label all possible matches ○ Label matches used for scoring ◉

Monoisotopic mass of neutral peptide Mr(calc): 1552.8170
Fixed modifications: Carbamidomethyl (C),TMTpro (N-term),TMTpro (K) (apply to specified residues or termini only)
Ions Score: 67 **Expect:** 2.7e-06
Matches : 14/116 fragment ions using 18 most intense peaks (help)

Here's a typical match from the search – as you can see, it's a pretty standard looking match with a good run of y-ions

SCoPE-MS

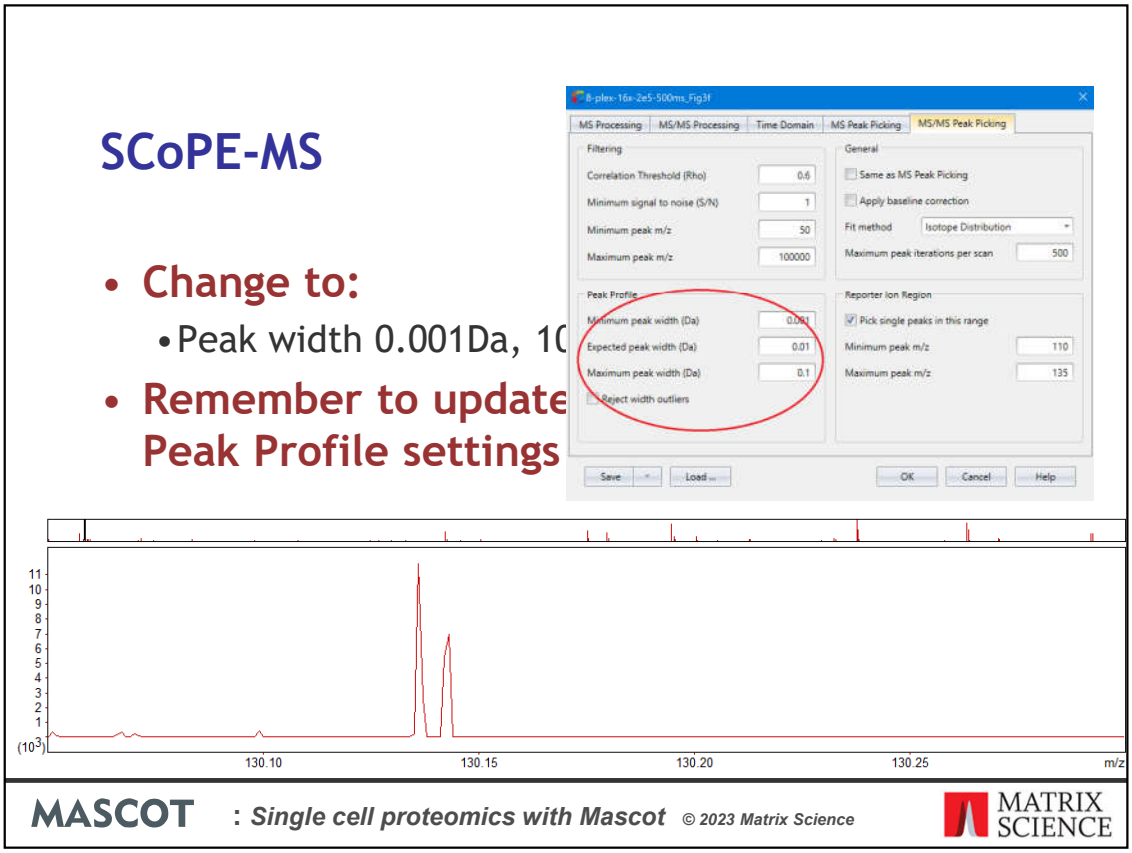
- **Uncentroiding TMTPro data**
  - Default peak width in Distiller is too large
    - Peak width 0.02Da, 600 points per Da
  - N and C peaks end up merged



MASCOT : *Single cell proteomics with Mascot* MATRIX SCIENCE

One word of caution. If your data are saved as centroids, you can sometimes get better results if you uncentroid the data in Distiller and allow Mascot Distiller to carry out it's own peak detection on the MS/MS scans – particularly if you have higher charge state precursors.

If you're using TMTPro tags, then the N and C tags are so close in mass, that the default uncentroiding options will cause the peaks to be merged together, as you can see in the example below.



So, for TMTPro data, we'd need to reduce the uncentroiding peak width to 0.001 Da – now you can see we're retaining the N and C peaks. If you have to do this, remember to adjust the peak profile settings under MS/MS Peak Picking in Distiller

Single cells

- **Much lower signal intensity than either bulk or SCoPE-MS data**
  - MS and MS/MS
- **Often combined with ion mobility (e.g. PASEF, FAIMS) to improve signal to noise**

If you're analysing single cells without a carrier proteome or any tag multiplexing, then you'll find your MS and MS/MS signal is much lower intensity than either bulk or SCoPE-MS analyses. Additional ion mobility separation can help so you can find a number of studies in the PRIDE dataset which have used PASEF, FAIMS etc

Single cells


RETURN TO ISSUE ‹ PREV ARTICLE NEXT ›

Features of Peptide Fragmentation Spectra in Single-Cell Proteomics

Hannah Boekweg, Daisha Van Der Watt, Thy Truong, S. Madisyn Johnston, Amanda J. Guise, Edward D. Plowey, Ryan T. Kelly, and Samuel H. Payne*

Cite this: *J. Proteome Res.* 2022, 21, 1, 182−188
Publication Date: December 17, 2021
https://doi.org/10.1021/acs.jproteome.1c00670
Copyright © 2021 The Authors. Published by American Chemical Society
RIGHTS & PERMISSIONS

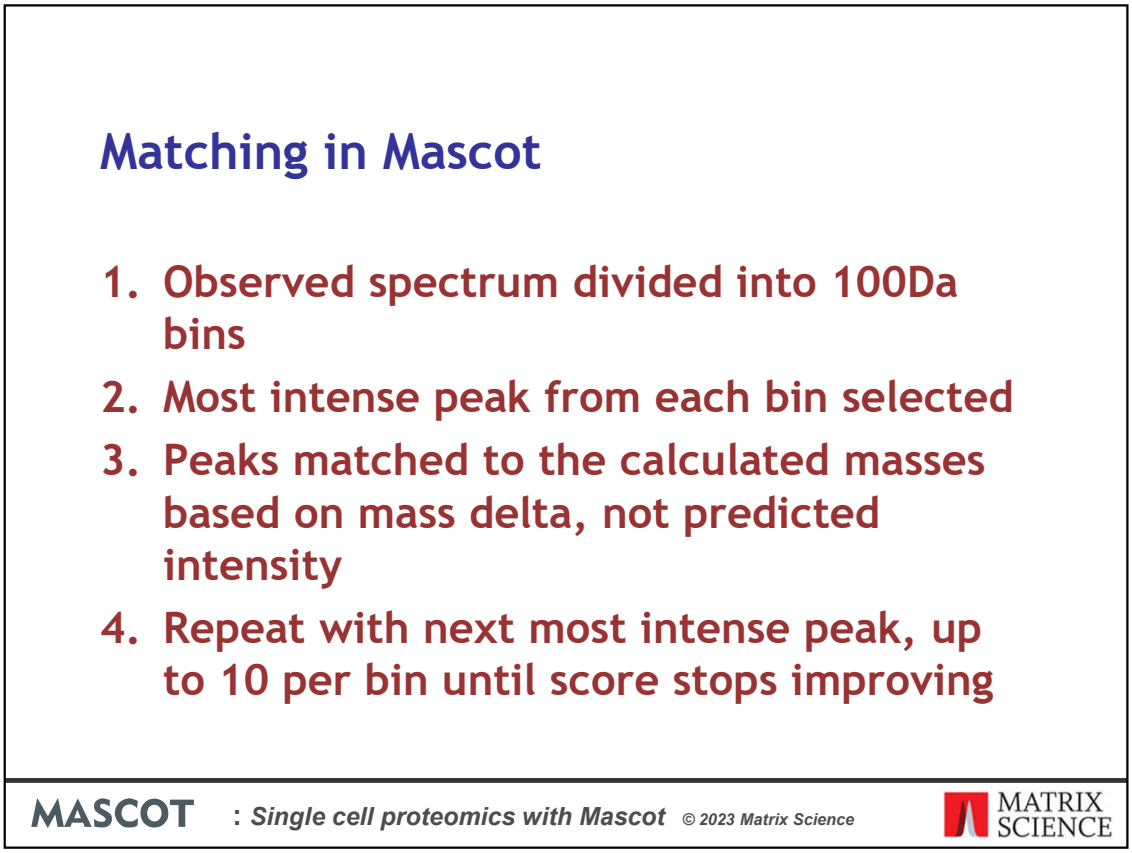
Article Views 2691 | Altmetric 23 | Citations 11
LEARN ABOUT THESE METRICS

Share Add to Export RIS

- **Intensity compression**
- **Fewer y-ion peaks**
  - Different peaks observed
  - Suggest single cell specific spectral libraries
- **Worse signal to noise, fewer peaks**
  - Lower scores for candidate peptides

Boekweg *et al* in a paper published in the Journal of Proteome Research last year looked at the features of spectra derived from single cells. They highlighted that you get intensity compression, blurring the line between signal and noise. The spectra generally have fewer y-ion peaks present in a different pattern to bulk samples, so the authors suggest using single cell specific spectral libraries. It would also be an important consideration when training machine learning models to predict spectra.

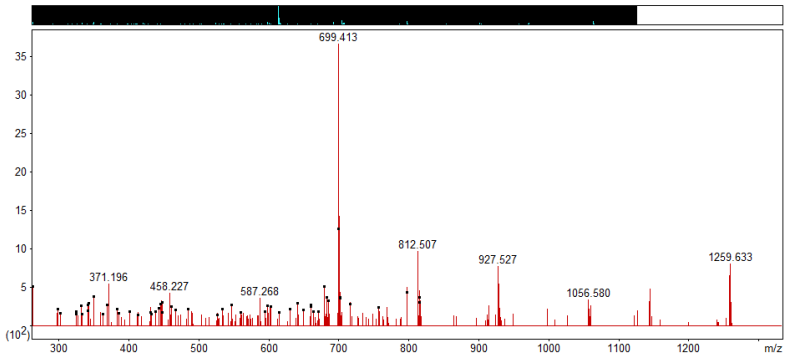
The poorer signal to noise and fewer peaks results in lower scores for candidate peptides.



Does this affect Mascot? The procedure used by Mascot for matching a spectrum to a candidate peptide sequence is outlined here. In the scoring, there is a penalty term that balances the sum of matched peak intensity with the sum intensity of the selected peaks. As long as the noise is random and doesn't completely overwhelm the fragment peaks in every bin, the algorithm is usually able to pick the signal from the noise. Adding noise tends to decrease the match score rather than prevent a match. Importantly, Mascot does not rely on predicted fragment peak intensity, which is important because, as suggested by Boekweg *et al,* this can be very different between bulk and single cell samples.

This means that for searching single-cell spectra with Mascot, we want to maximise the signal to noise, and we should look at using Percolator to see if that can improve scoring.

Single cells



timsTOF dataset (PXD024043)

- Low fragment ion intensity
  - Many peaks barely above noise peaks
- Peaks are narrower than default processing options

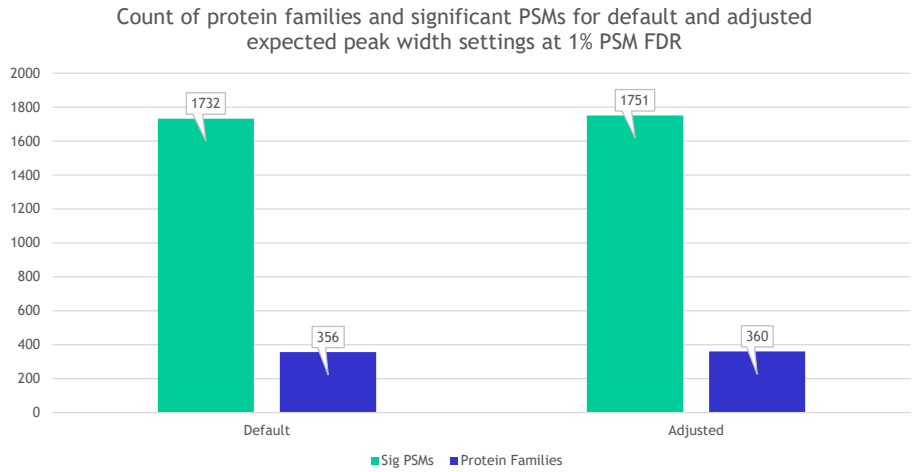
Here we have a typical MS/MS spectrum from a timsTOF instrument – taken from a publically available single-cell dataset in the PRIDE repository. As you can see, the fragment peaks are mainly low intensity. It's the entire peak area which is reduced of course, not just the zenith – so in addition to a lower than normal maximum intensity, the peaks are considerably narrower than assumed by the default timsTOF processing options in Mascot Distiller.

Peak picking with Distiller – peak width

- default.BrukerTIMS.opt expected peak width
  - 0.012 Da
- Actual average peak width for this dataset
  - ~ 0.0045 Da

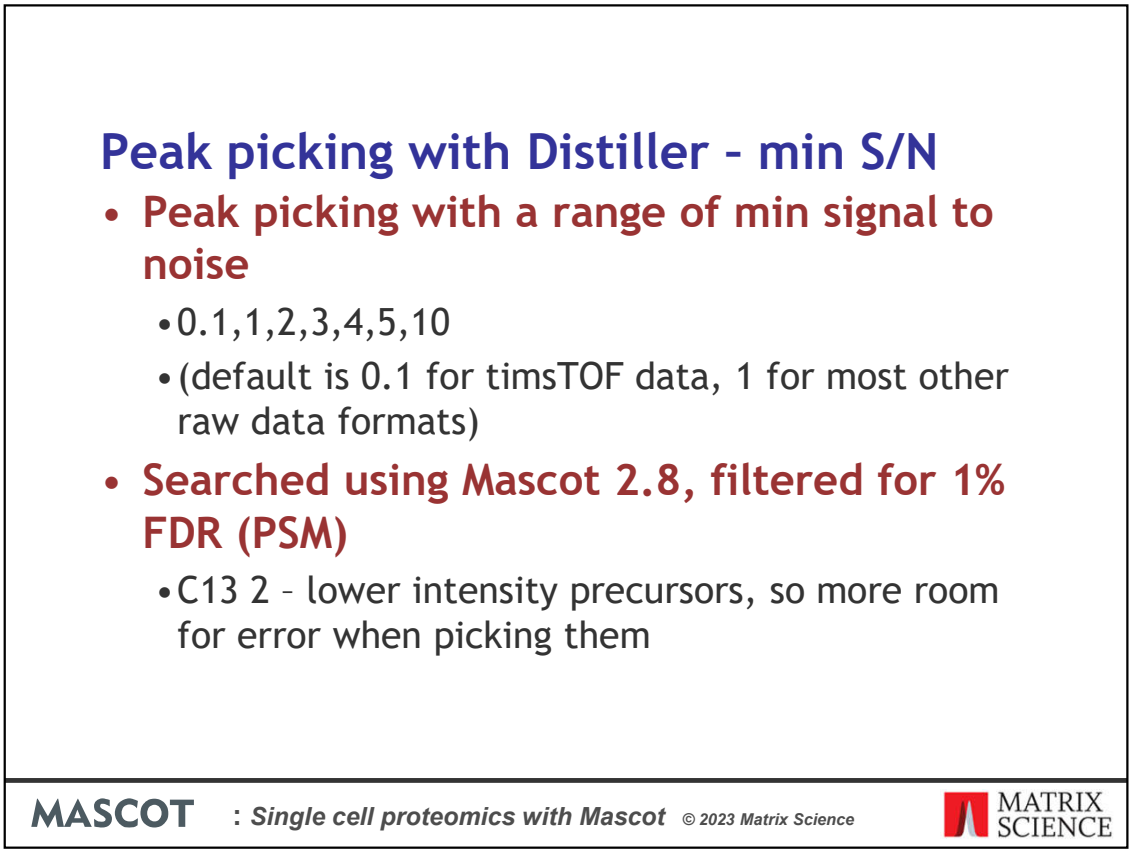
So, to get the best results from these types of data, unlike with SCoPE-MS datasets we need to take a look at our peak-picking. Starting with peak widths in the MS/MS spectra. The default.BrukerTIMS.opt options in Mascot Distiller have an expected peak with of 0.012 Da, but a quick look at this dataset shows that the actual average peak width is closer to 0.0045 Da.

Peak picking with Distiller - peak width



MASCOT : *Single cell proteomics with Mascot* *© 2023 Matrix Science* MATRIX SCIENCE

If we adjust the MS/MS peak picking peak profile options to match the expected peak width, we do get slightly improved results from this data file, with an additional 19 significant PSMs at 1% FDR with 4 additional protein families. Nothing too dramatic, but the processing time was roughly the same so it's worth having.



We can actually get further improvements by combining this with adjusting the minimum signal to noise setting in the peak filtering in Distiller. To demonstrate this, we carried out peak picking with a range of minimum S/N values and then took the number of significant PSMs at a 1% PSM FDR.

Peak picking with Distiller – min S/N

Minimum S/N	PSMs above homology (1% FDR)
0.1	1751
1	1900
2	1904
3	1907
4	1934
5	1912
10	1675

By increasing the minimum signal to noise from 0.1 to 1, we get an immediate jump of 149 in the number of significant PSMs at 1% FDR. There is then a steady increase in the number of significant matches up to a signal to noise filter of 4, as the overall signal to noise of the spectrum is being improved by excluding noise peaks. After this the numbers fall – presumably by that stage we're excluding too many real fragment ion peaks.

Peak picking with Distiller – min S/N

Minimum S/N	PSMs above homology (1% FDR)	Percolator (1% FDR)
0.1	1751	2832
1	1900	2857
2	1904	2810
3	1907	2867
4	1934	2914
5	1912	2853
10	1675	2735

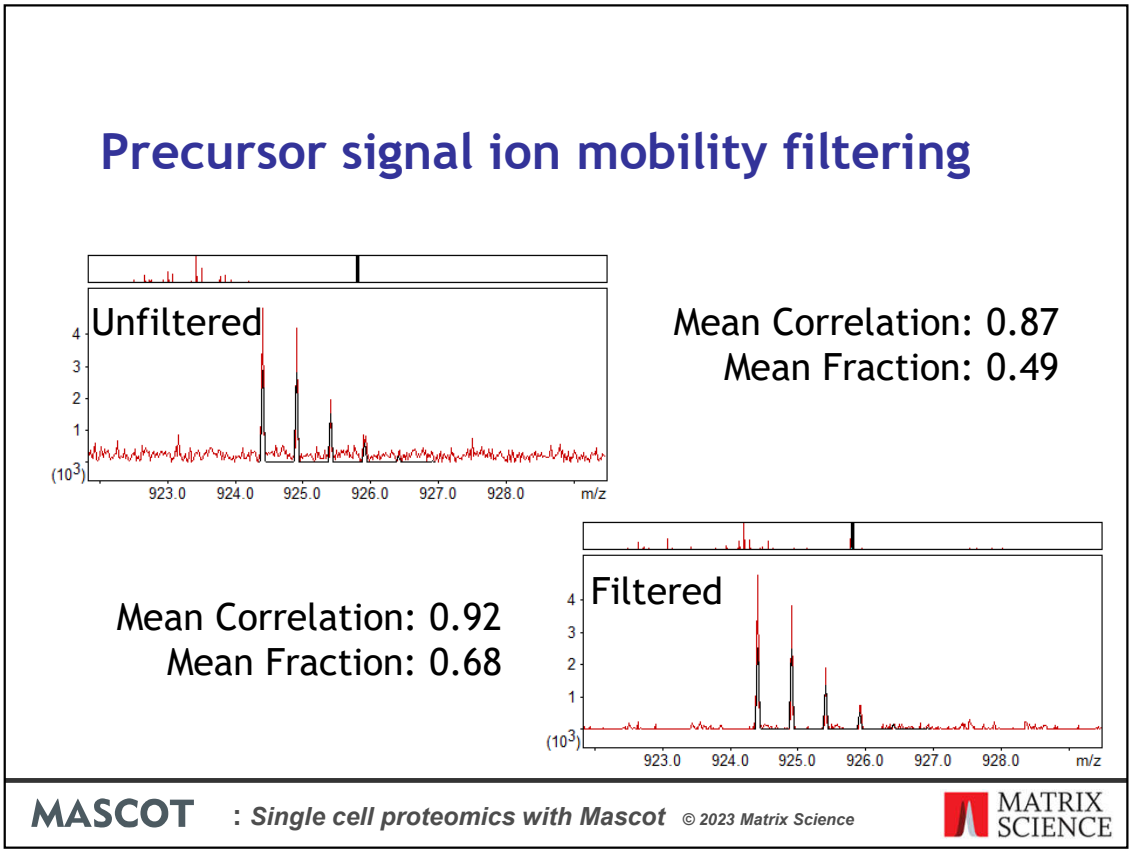
However, just enabling Percolator rescoring gets us over 1000 additional PSMs

Peak picking with Distiller - min S/N

Minimum S/N	PSMs above homology (1% FDR)	Percolator (1% FDR)	Percolator+RT (1% FDR)
0.1	1751	2832	2976
1	1900	2857	3015
2	1904	2810	3084
3	1907	2867	2978
4	1934	2914	3070
5	1912	2853	3038
10	1675	2735	3001

And if we enable the retention time feature in Percolator, we gain another 150 or so PSMs.

So the take home message is to adjust your peakpicking settings for single cell spectra as the spectra are quite different to bulk peptide spectra, and enable Percolator rescoring to further improve your results.

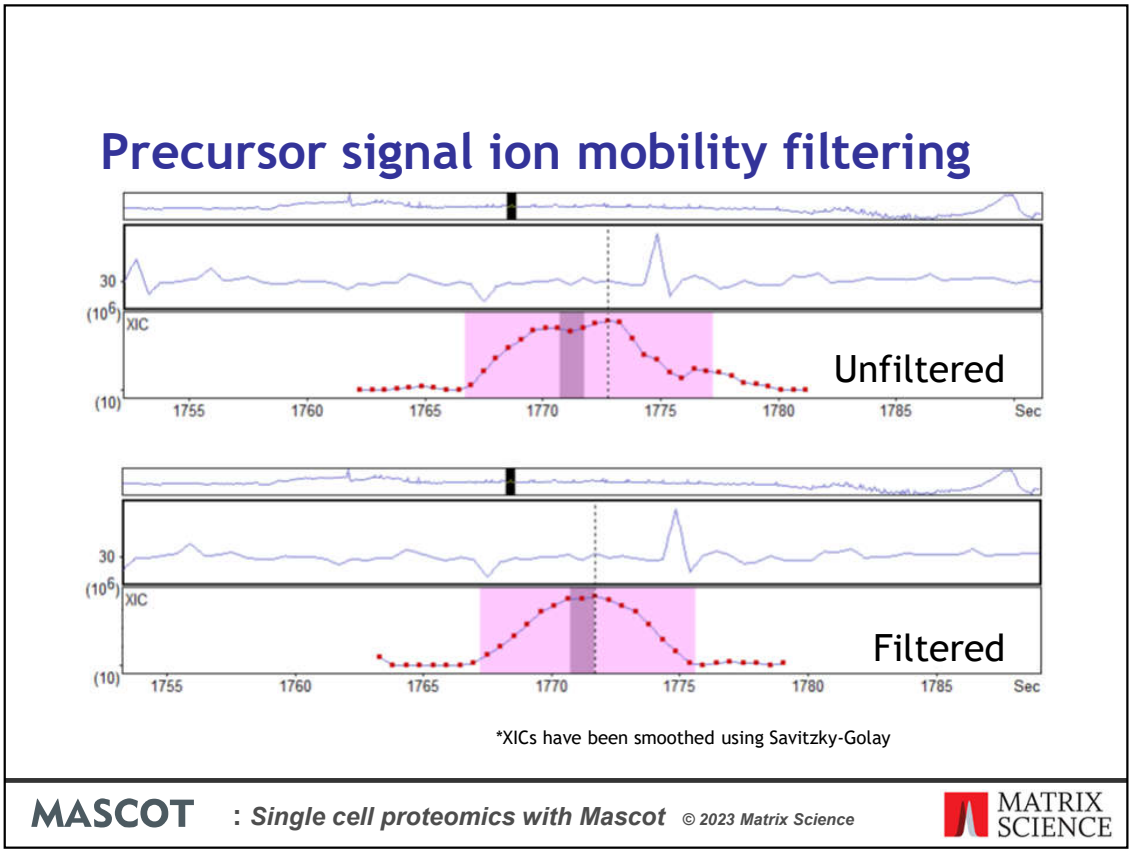


If you have timsTOF data and you include the ion mobility values in the peaklists, then the precursor signal can be filtered by Mascot Distiller during quantitation based on the peptide ion mobility range. This allows us to get a clean signal for some very low intensity precursors, such as this one here. With the filtering enabled, we're getting a clean baseline, with the fourth and fifth peaks of the isotope distribution much clearer from the baseline than without filtering.

This improves the quality of the quantitation results. For this datafile, without ion mobility filtering the average correlation coefficient between the predicted and observed precursor isotope distributions is 0.87, and the average fraction of the peak area in the precursor region is 0.49.

With ion mobility filtering, these improve to 0.92 and 0.68 respectively.

Enabling this requires the search was carried out with Mascot 2.7 or later.



For the very low intensity signal we often see with single cell samples, removing this baseline noise can significantly improve our XIC detection, as is shown here. The unfiltered XIC will probably be over-estimating the peptide abundance due to including more baseline noise in the area.

Conclusions

- **SCoPE-MS spectra are much more similar to standard bulk sample spectra**
  - But be careful if you're uncentroiding TMTPro data!
- **Single cell spectra are quite different to bulk sample spectra**
  - Fewer y-ions
  - Intensity compression
- **Peak detection, Percolator rescoring are important for getting the best results**
  - Precursor Ion mobility filtering of timsTOF data helps with low intensity signal

