ANALYSIS WORKFLOW FOR QUANTITATIVE PROTEOMICS, EMPLOYING TRIPLEX DIMETHYL LABELLING AND ION MOBILITY ASSISTED DATA INDEPENDENT ACQUISITION

Andrew Collins[1,2], Antony McCabe[1,2], Christopher J Hughes[3], Ian Morns[4], Johannes PC Vissers[3], Andrew R Jones[1,2] , Jan Claereboudt[5],
[1] Institute of Integrative Biology, University of Liverpool, UK; [2] OMIC ANALYTICS LTD, Liverpool, UK; [3] Waters Corporation, Wilmslow, UK; [4] Nonlinear Dynamics Limited, Newcastle upon Tyne, UK; [5] Waters Corporation, Zellik, Belgium

INTRODUCTION

Stable isotope labelling can be routinely applied in LC-MS proteomics, for accurate and reproducible quantitative profiling on a large scale. Labels are incorporated metabolically, enzymatically, chemically or by stable isotope labelling. Here we demonstrate a novel informatics processing pipeline for data sets generated using dimethyl chemical labelling, applied in triplex. This method has the advantage of being broadly applicable to any sample type, and has quantitative reproducibility close to that achievable with metabolic labelling. We demonstrate the benefits of using software able to incorporate retention time (t_r) alignment and profiling, as well as profiling of ion mobility (IM) drift times (t_d) to increase confidence in peptide quantification and sensitivity.

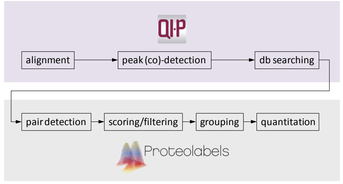

Figure 1. Progenesis QI for proteomics/Proteolabels workflow.

METHODS

Sample preparation

Triple dimethyl samples were prepared as previously described[1] and the proteomes (HeLa, yeast and *E.coli*) pooled and labeled as shown in Figure 2.

LC-MS conditions

Nanoscale LC separation of tryptic peptides was conducted with a trap column configuration using a M-class system and a 90 or 120 min gradient from 5-40% ACN (0.1% FA) at 300 nl/min using a BEH 1.7 µm C18 reversed phase 75 µm x 20 cm nanoscale LC column. MS data were acquired in triplicate in data independent analysis mode in ion mobility enabled data independent analysis mode (LC-IM-DIA-MS) using a Synapt G2 -*Si* instrument.

Informatics

The LC-MS peptide data were aligned, peak detected and searched with Progenesis QI for proteomics using a reviewed UniProt protein sequence databases. Quantitative analysis of the peptides and protein grouping was conducted with Proteolabels.

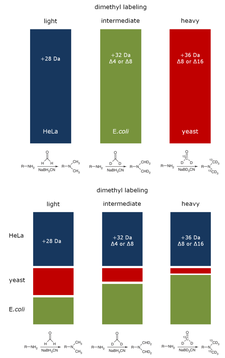

Figure 2. Experimental designs (top: proteome unique; bottom: HeLa 100:100:100, yeast 50:25:10, and E. coli 50:75:90, 'light' : 'intermediate' : 'heavy', respectively).

RESULTS

Multidimensional data acquisition and analysis

Example data are shown in Figure 3, illustrating a two-dimensional representation of the data (intensity *vs.* m/z) and inset two three dimensional counterparts of the same set of triplets, showing intensity as a function of t_r and *m/z*, and intensity as a function of t_d and *m/z* . Progenesis QI for proteomics utilizes drift time to confirm ion alignment across multiple runs and to correlate DIA precursor and product ions, whereas Proteolabels uses the same information for pair/triplet detection and scoring.

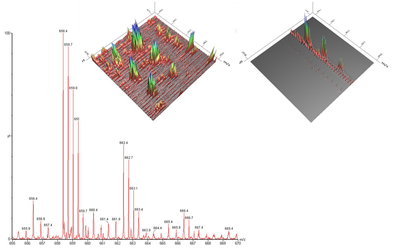

Figure 3. Dimethyl 50:25:10 (yeast) triplet separated by m/z, t_r, t_r and intensity.

Acquisition and DIA search specificity

The first experimental design shown in Figure 2 was used to access the specificity of the LC-IM-DIA-MS acquisition method and the DIA search algorithm embedded in Progenesis QI for proteomics. A limited amount of 100 ng (all proteomes combined) of labeled samples was analysed using default triple dimethyl search parameters and automatic precursor and product ion search tolerances (5 and 12 ppm, respectively).

The results shown in Figure 4 illustrate how the peptide identification FDR was estimated by expressing the incorrectly labeled peptide, as they are proteome unique, as a function of the number of identified peptides that are correctly labeled. The overall observed peptide FDR equaled 0.5%.

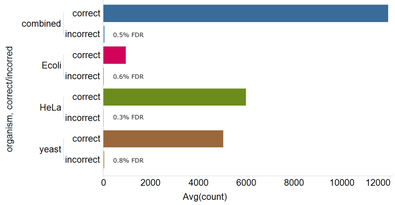

Figure 4. Peptide FDR estimation for the individual proteomes and all peptide searches combined across all species.

Auto-detection and pair scoring

Proteolabels has been designed to make a first pass at profiling the peptide pairs/triplets present in the data, to detect automatically the experiment design, and the optimal settings for *m/z* and t_r tolerance. In addition, novel metrics for scoring the quality of peptide feature groups have been integrated, enabling the measurement of the reliability of peptide-level quantitation. The pair/triplet scoring is based upon profiling *m/z*, retention time and chromatogram matching between light, intermediate and heavy peptides across all identified ions as illustrated in Figure 5.

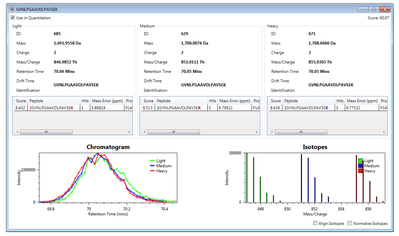

Figure 5. Identification of a dimethyl 100:100:100 (HeLa) peptide triplet showing good agreement between the chromatographic, isotopic, identification and drift properties of the detected and identified peptide.

QC Metrics

Shown in Figure 6 are Proteolabels QC metric and data summary graphics that enable detailed exploration, interaction and interpretation to arrive at the best protein quantitation set. These QC tools complement the raw data QC metrics implemented in Progenesis QI for proteomics (not shown).

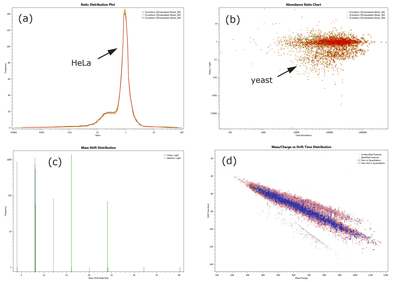

Figure 6. Proteolabels QC metrics and visualization plots, including peptide ratio distribution (a), peptide abundance ratio (b), detected mass shift distributions (c), and m/z vs. drift with identified and quantified chemically labelled peptide feature groups (pairs/triples) (d).

Protein inference and quantitation

A protein grouping step, to account for different proteins supported by the same peptides, is conducted prior to quantitation. "Unique peptides" (assigned to a single protein) and "resolved peptides" (assigned to a same-set group of proteins) are used for quantitation. "Conflicted peptides" (cannot be uniquely assigned to a protein group), are removed from quantification by default, but can be manually added back in.

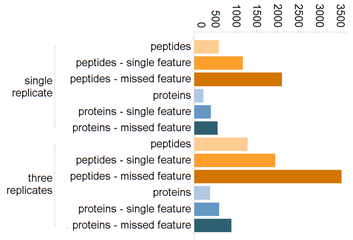

Figure 6. Analysis of data from the second design - quantified in isolation (single replicate) or analyzed in tandem (three technical replicates) showing the increase in peptides and proteins (groups) quantified requiring the peptide to be identified in all channels (light orange/blue), only a single feature required to be identified (medium orange/blue) and allowing features to missed in one of the channels (dark orange/blue).

Protein-level quantification is performed in Proteolabels using a novel "Weighted average" system, by which the overall abundance of a pair/triplet and the "Pair/Triplet Score" is taken into account to arrive at a protein-level ratio. The advantage of this method over a simple median or mean is that more reliable peptide feature groups contribute more heavily to the overall protein-level quantitation value, but while allowing other plausible/reliable peptide feature groups to contribute in a systematic manner.

Shown in Figure 6 are the benefits of co-detection across samples and the ability to create peptide feature groups (using Proteolabels "pair/triplet scoring" for increased confidence), both contributing to increased peptide (non-redundant counts) and protein coverage. Note that the number of quantified peptides approximately scales with the number of identified peptides shown in Figure 4 per number of conditions.

Quantitative precision, abundance and dynamic range

Three technical replicates of the second experimental design were used to estimate quantitative precision. Shown in Figure 7 are the non-normalized $\log_2$ transformed 'light'/'heavy' peptide ratio values/channel. Despite a slight overall ratio offset, good agreement with the expected relative abundance values (dashed lines) was observed. The slight difference between expected and observed is likely due to imperfect mixing/labeling.

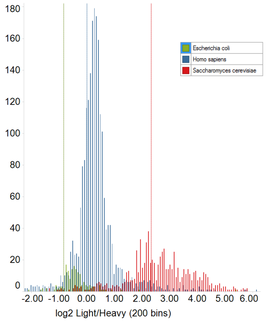

Figure 7. Raw, unfiltered dimethyl peptide fold change illustrating observed, i.e. 'light' vs. 'heavy' condition/channel comparison for expected 1:1 (Homo sapiens - HeLa), 5:1 (Saccharomyces cerevisiae - yeast) and a 1.8:1 (Escherichia coli as represented by the dashed lines.

Weighted average protein centric quantification results in Volcono plot format for the same mixed proteomics dimethyl labeling quantitation experiment are shown in Figure 8, contrasting the 'light' versus the 'intermediate' and 'heavy' channels, respectively. As a reference, dashed two-fold change trend lines have been added as a reference, demonstrating that precise quantitation is feasible. Figure 8 shows clear separation (by fold change and p-value) of the *E. coli* (green) and yeast (red) proteomes (changing in abundance) from the Hela (blue, human) cells (not changing in abundance)

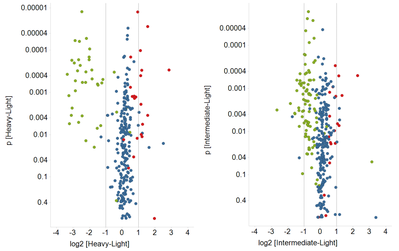

Figure 8. Weighted average normalized protein fold change as a function of regulation probability (Student's T test) for 'light' vs. 'heavy' (left) and 'light' vs 'intermediate' labeled mixed proteomes according to the second experimental design shown in Figure 2. Colors as per Figure 7.

CONCLUSION

- **Quantitative functionality of Progenesis QI for proteomics and Proteolabels has been extended to include the analysis of dimethyl isotopically labelled samples in triplex for the large scale analysis of LC-MS data**
- **High quantitative precision was observed, including reliable quantitation of the expected ratios according to the spike-in design**
- **Co-detection across LC-MS runs and metabolic pair matching afforded ~ 2-fold gain in both peptide and protein detection/quantification**
- **Proteolabels "Weighted averaging" based profiling of peptide feature groups delivers high-level quantitative accuracy at the protein-level**
- **The analysis demonstrated that accurate and reliable differential expression could be detected of yeast and *E. coli* proteins, against a dominant background of unchanging human proteins.**
- **The Proteolabels QC metrics further enhance accurate quantitation by enabling users to explore numerous aspects of the data, including criteria for grouping peptides (pairs/triples), and overall trends for peptide and protein quantitation.**

References

1. Boersema PJ, Raijmakers R, Lemeer S, Mohammed S, Heck AJ. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. Nat Protoc. 2009;4(4):484-94.
2. http://www.nonlinear.com/progenesis/qi-for-proteomics/
3. Qi D, Brownridge P, Xia D, Mackay K, Gonzalez-Galarza FF, Kenyani J, Harman V, Beynon RJ, Jones AR. A software toolkit and interface for performing stable isotope labeling and top3 quantification using Progenesis LC-MS. OMICS. 2012 Sep;16(9):489-95.
4. Geromanos SJ, Hughes C, Ciavarini S, Vissers JP, Langridge JI. Using ion purity scores for enhancing quantitative accuracy and precision in complex proteomics samples. Anal Bioanal Chem. 2012 Sep;404(4):1127-39.

TO DOWNLOAD A COPY OF THIS POSTER, VISIT WWW.WATERS.COM/POSTERS