

LARGE SCALE QUANTITATION OF STABLE ISOTOPE LABELLED PROTEOMES USING RETENTION AND DRIFT TIME PROFILING

Andrew P Collins¹; Antony McCabe¹; Ian Morns²; Johannes PC Vissers³; Andrew R Jones¹

¹Institute of Integrative Biology, University of Liverpool, United Kingdom; ²Nonlinear Dynamics Limited, Newcastle upon Tyne, United Kingdom; ³Waters Corporation, Wilmslow, United Kingdom

INTRODUCTION

The use of stable isotope labelling is routinely applied in LC-MS proteomics based quantitation. Labels are incorporated metabolically, enzymatically, chemically or by stable isotope labelling. The comparison of multiple samples necessitates a higher degree of label multiplexing, which in turn requires higher separation power LC systems, more sensitive, faster and advanced MS to acquire more and better quality product ion spectra, and informatics tools that can take advantage of the high resolution and mobility separation afforded by modern MS. Here, the principle and application of novel informatics is described to enable stable isotope labelling quantitation using commercial software originally developed for large scale quantitation of label-free LC-MS data, incorporating novel retention and ion mobility drift time profiling based scoring algorithms.

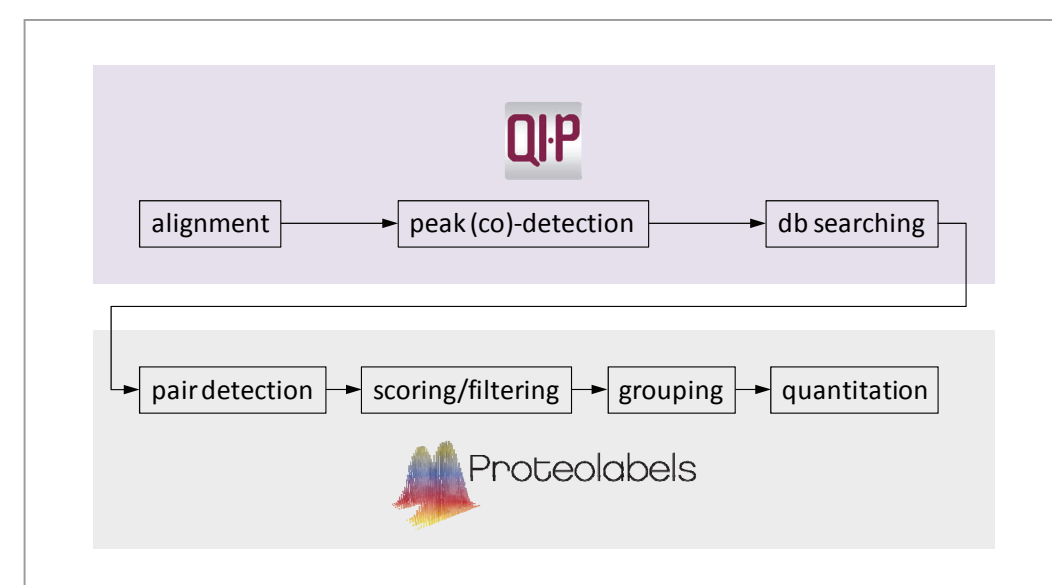


Figure 1. Progenesis QI for proteomics/Proteolabels workflow.

METHODS

Sample preparation

Two sample types were investigated following reduction, alkylation and trypsin digestion: 1) Dimethyl (CH₃ 'light' and CD₃ 'heavy') labelled HL60 human B cells and 2) Non-labelled and ¹³C₆-lysine labelled BW5147-derived mouse T-cell hybridoma cells.

LC-MS conditions

Nanoscale LC separation of tryptic peptides was conducted with a trap column configuration using a M-class system and a 90 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 nano-scale 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 and reviewed UniProt protein sequence databases. Quantitative analysis of the peptides and protein grouping was conducted with Proteolabels — Figure 1.

RESULTS

Co-detection

The workflow builds upon the Progenesis QI for proteomics co-detection feature, shown in Figure 2, in which features are aligned across LC-MS runs. This affords gains in sensitivity and reproducibility. Sensitivity is improved since identifications can be propagated across runs, for example enabling low abundance features to be analysed, even if they have not been identified in a given run. Since the same basis for quantification is used across all runs, highly reproducible peptide ratios across replicates are achieved.

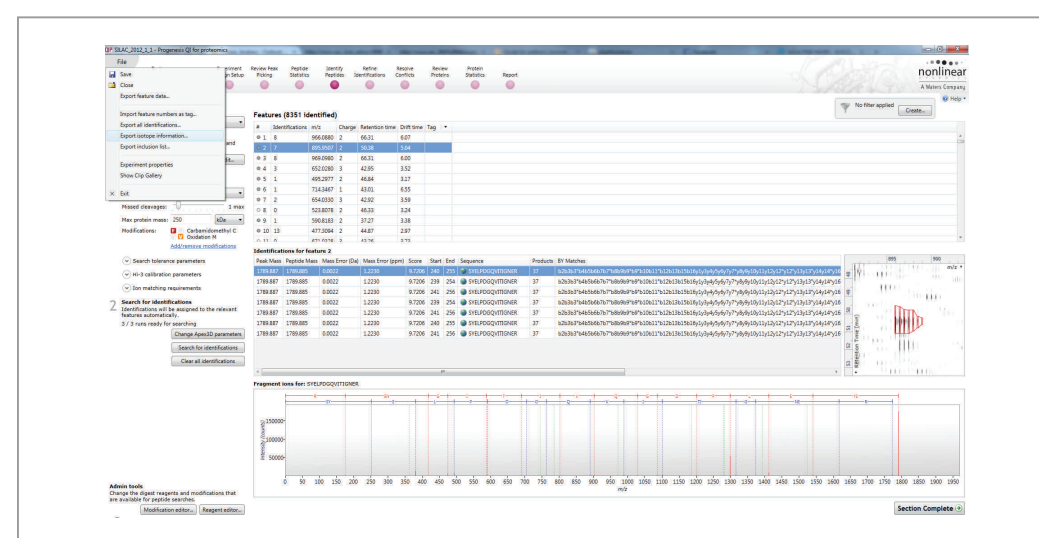


Figure 2. Export of the isotopic information, of the co-detected features and identified peptides, including chromatographic and drift time profiles.

Auto-detect

Proteolabels has been engineered to make a first pass at profiling the peptide pairs present in the data, to detect automatically the experiment design, i.e. the quantification label used, and the optimal settings for mass/charge and retention time tolerance. The "Auto-Detect" mode gives users confidence that the software is functioning optimally, removing a common source of frustration that expert use and understanding is required for parameters setting — Figure 3.

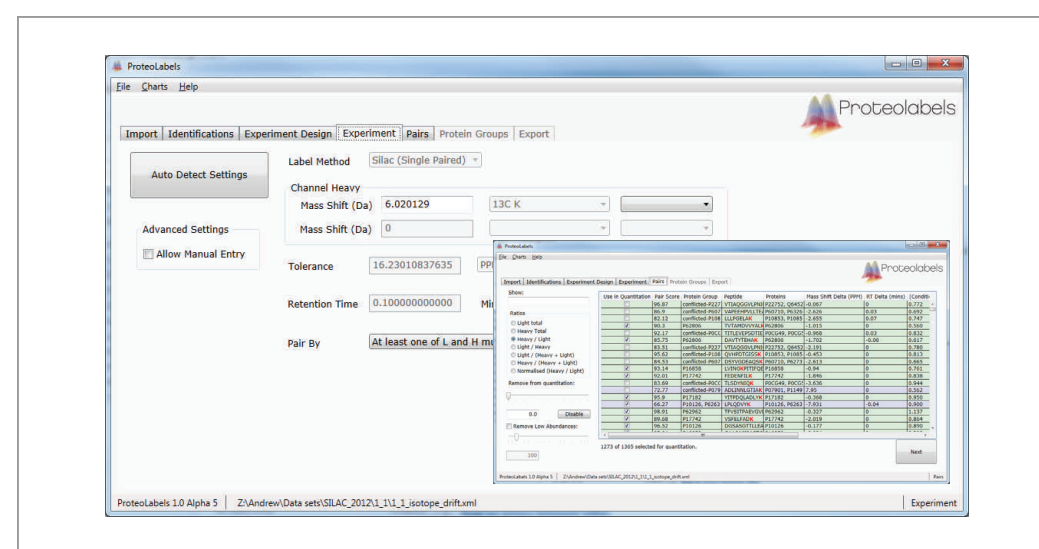


Figure 3. Auto-detect feature of Proteolabels.

Pair scoring

Proteolabels contains novel metrics for scoring the quality of peptide pairs, enabling users to measure the reliability of peptide-level quantitation, remove those pairs that are suspect, giving higher quality protein-level quantification. The pair scoring is based upon profiling m/z, retention time and chromatogram matching between light and heavy pairs across all identified ions — Figure 4.

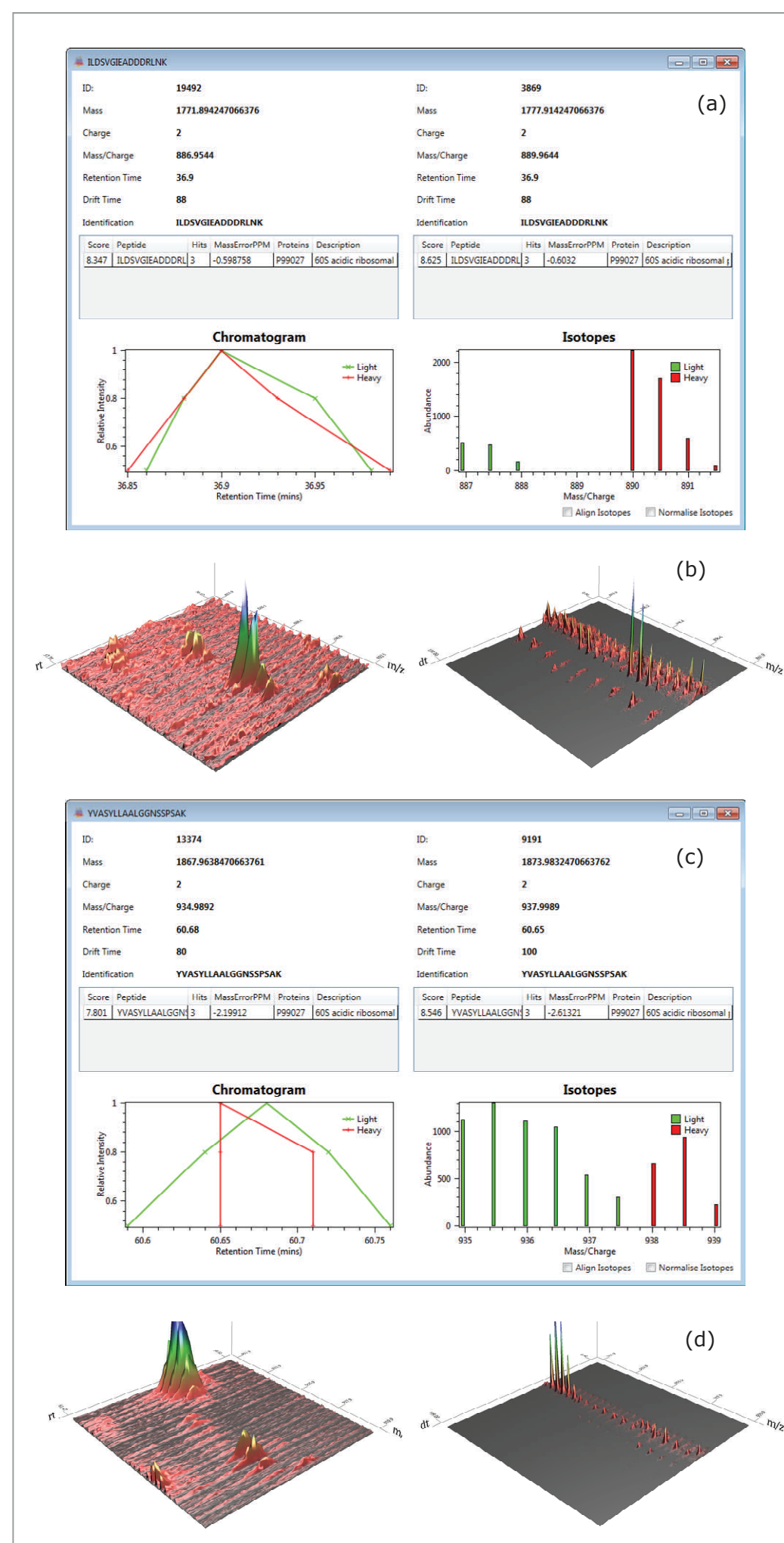


Figure 4. Detection of peptide pair showing good agreement between the chromatographic, isotopic, identification and drift properties, (a) and (b), and a pair with disagreement in chromatographic and drift properties caused by an isobaric co-eluting peptide, (c) and (d), interfering with the heavy channel.

QC Metrics

Proteolabels contains a large number of QC metric and data plotting options, enabling users to explore, interact and understand their data, and, where necessary, make informed decisions as to how arrive at the best protein set — Figure 5. These QC tools complement the raw data QC metrics implemented in Progenesis QI for proteomics.

Intelligent protein inference

A challenge in bottom-up proteomic workflows is how to generate protein-level quantification from peptide signals. We incorporate a protein grouping step, to account for different proteins supported by the same sets of peptides. "Conflicted peptides", those that cannot be uniquely assigned to a protein group, are removed from quantification by default, but can be manually added back into a given group as required.

Protein-level quantification can be performed by a number of different mathematical methods (median, sum, mean, "best scoring pair"), giving users flexibility for special cases. However, by default we recommend the use of Proteolabels "Weighted average" scheme, by which the overall abundance of the light/heavy pair, and the "Pair Score" is taken into account to arrive at the protein-level ratio — Figure 6. Shown as well are the benefits of co-detection across samples and the ability to match pairs, both contributing to increased peptide and protein coverage.

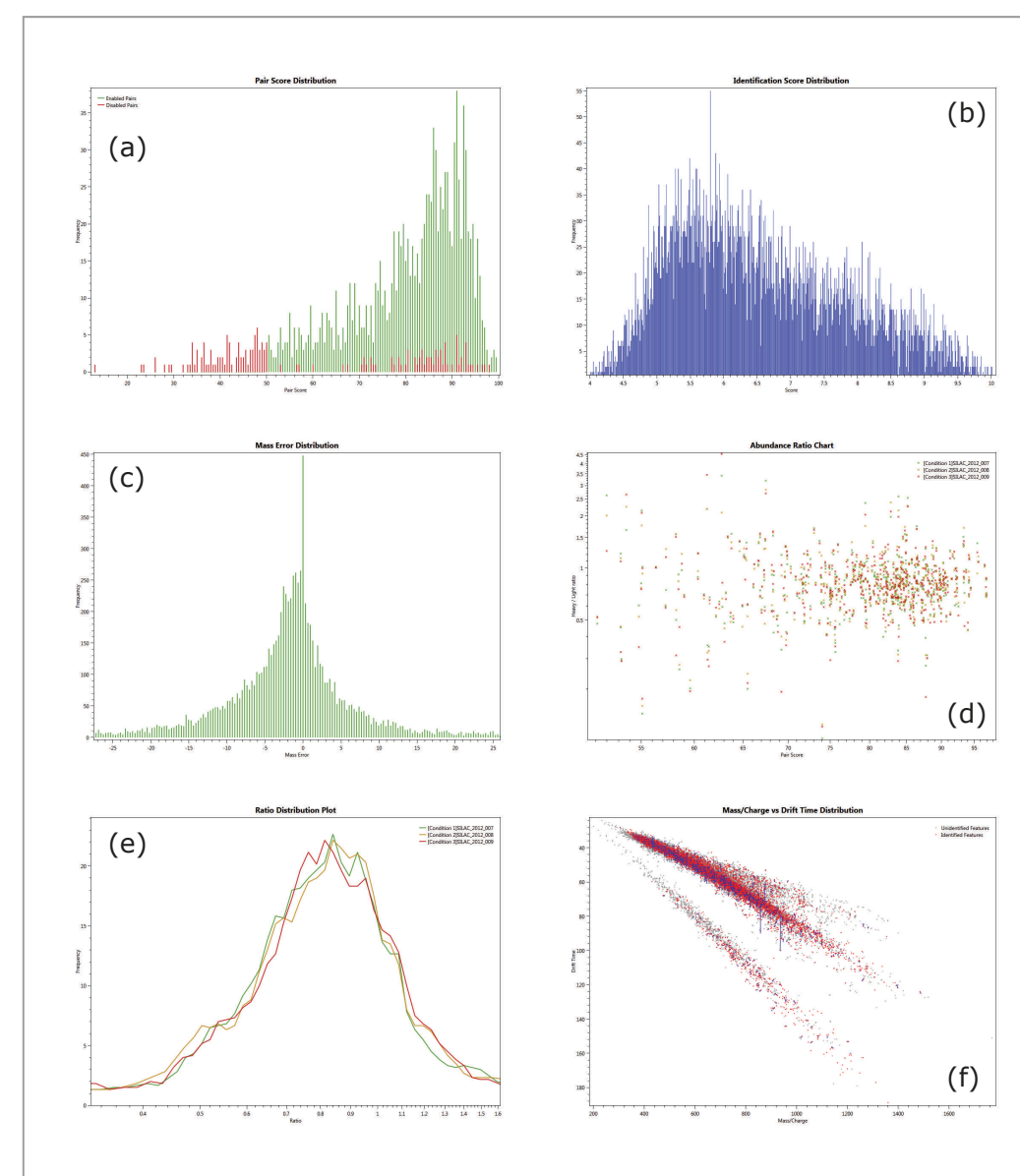


Figure 5. Proteolabels QC metrics and visualisation plots, including pair quantitation score distribution (a), peptide identification score distribution (b), search error distribution (c), peptide ratio abundance ratio (d), protein ratio distribution profiles (e), and m/z vs. drift with identified metabolically labelled peptide pairs (f).

Integrated workflow

Proteolabels has been engineered to work simply alongside Progenesis QI for proteomics, giving users an integrated workflow — from a straightforward install, through data analysis, and to data export in flexible formats.

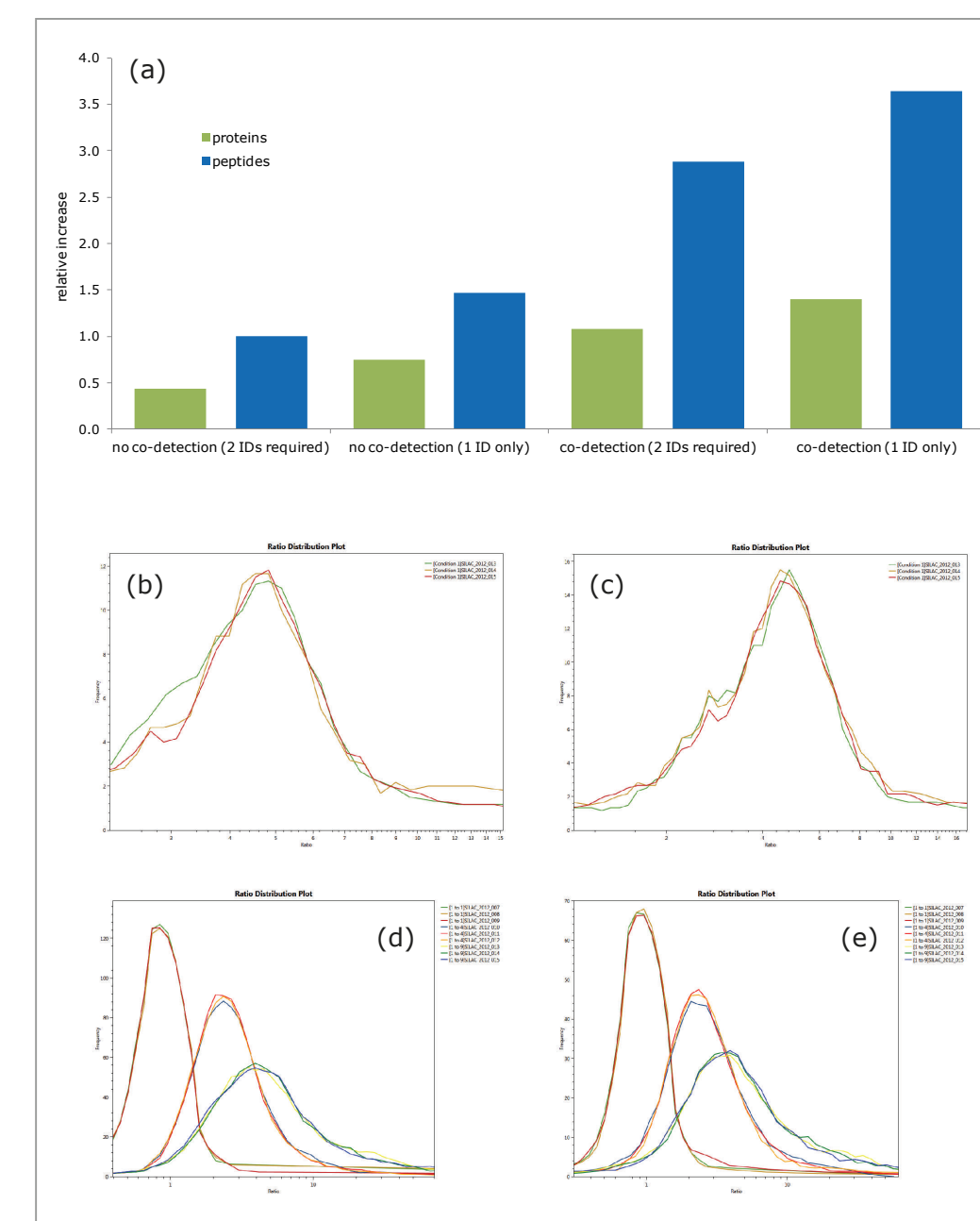


Figure 6. Analysis of a 1:9 SILAC spike quantified in isolation and as part of a wider analysis of a 1:1; 1:4 and 1:9 experiment (a). Both components of a peptide pair identified (2 IDs required; no co-detection (b) and with co-detection (d)) or only the light or heavy equivalent identified (1 ID only; no co-detection (c) and with co-detection (e)). Co-detection provided a ~ 3 fold increase in quantifiable peptides and proteins. The ability to quantify singletons provided a ~ 50% gain. No loss in ratio quality was observed ((b) to (e)) through the increments in numbers of quantifiable peptides and proteins.

Quantitative precision, abundance and dynamic range

Three technical replicates of a biological dimethyl labelled sample were analysed to assess the precision and dynamic range of the complete workflow — Figure 7. Technical replicates showed great similarity in terms of reported ratio across the complete dynamic range of the sample. Moreover, no variance dependency was observed as a function of either ratio or abundance. Moreover, relatively small fold changes could be determined reliably across the complete observed dynamic range.

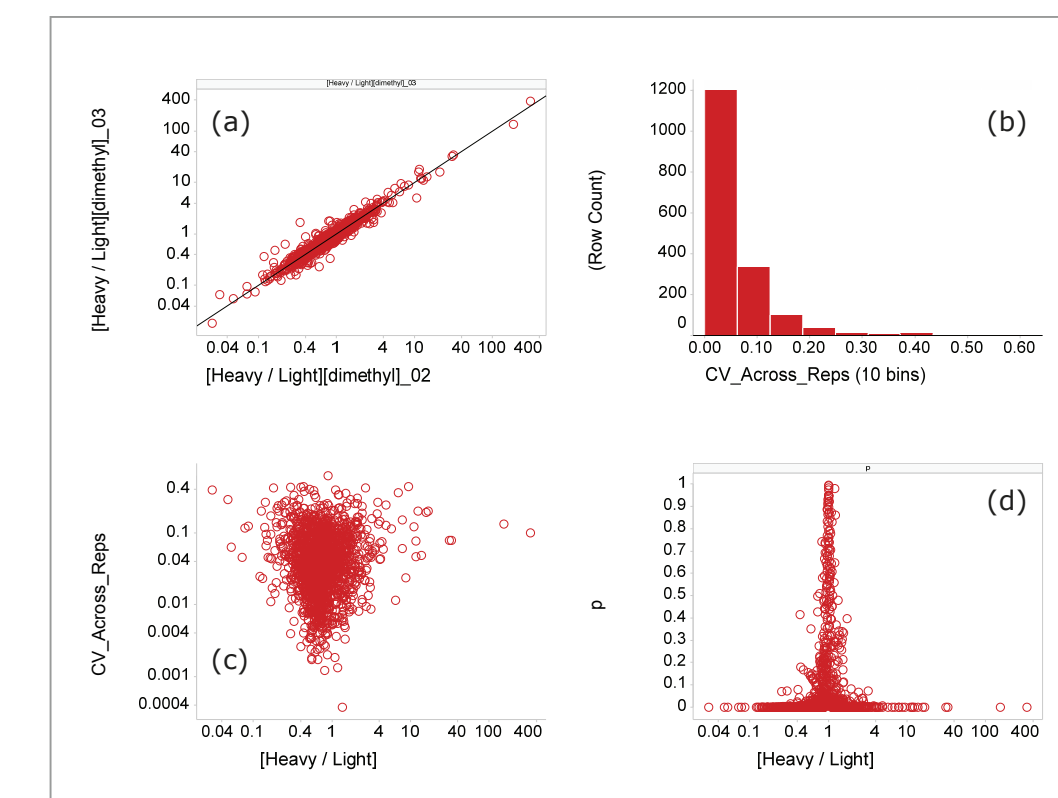


Figure 7. Quantitative precision and dynamic range analysis, of technical replicates, of dimethyl labelled samples, log ratio comparison two technical replicates (a), co-efficient of variance distribution (CV) across replicates (b), CV as a function of average ratio (c), and significant difference (p) as a function of ratio (d).

CONCLUSION

- Quantitative functionality of Progenesis QI for proteomics extended to include the analysis of isotopically labelled samples by integrating Proteolabels software in the workflow for the large scale analysis of LC-MS data
- Metrics introduced to score the quality of a peptide based on m/z and chromatographic matching, including drift time/ion mobility as an additional validation filter, which complement the raw data quality control metrics in Progenesis QI for proteomics
- The accuracy and precision of the workflow were in agreement with expected values and were of high quality
- Co-detection across LC-MS runs and metabolic pair matching afforded ~ 3-fold gain in both peptide and protein detection/quantification

References

- http://www.nonlinear.com/progenesis/qi-for-proteomics/
- 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.
- Geromanos SJ, Hughes C, Clavari S, Vissers JP, Langridge JJ. Using ion purity scores for enhancing quantitative accuracy and precision in complex proteomics samples. Anal Bioanal Chem. 2012 Sep;404(4):1127-39.

Acknowledgement

Andrew Williamson and Anthony Whetton from the School of Cancer and Imaging Sciences, University of Manchester, United Kingdom are kindly acknowledged for providing SILAC labelled BW5147-derived mouse T-cell hybridoma cells. Dimethyl labelled HL60 human B cells were donated by Bobby Florea, Bio-organic Synthesis, Faculty of Science, Leiden Institute of Chemistry, Netherlands.