

Poster presentation

Open Access

Pride Wizard: generation of standards compliant quantitative proteomics data

Neil Swainston*¹, Jennifer A Siepen², Andrew R Jones^{2,3}, Sarah R Hart⁴, Henning Hermjakob⁵, Phil Jones⁵ and Simon J Hubbard²

Address: ¹Manchester Centre for Integrative Systems Biology, Manchester Interdisciplinary Biocentre, University of Manchester, Manchester M1 7ND, UK, ²Faculty of Life Sciences, Michael Smith Building, University of Manchester, Manchester M13 9PL, UK, ³School of Computer Science, Kilburn Building, University of Manchester, Manchester M13 9PL, UK, ⁴Michael Barber Centre for Mass Spectrometry, School of Chemistry, Manchester Interdisciplinary Biocentre, University of Manchester, Manchester M1 7ND, UK and ⁵EMBL Outstation, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK

Email: Neil Swainston* - neil.swainston@manchester.ac.uk

* Corresponding author

from BioSysBio 2007: Systems Biology, Bioinformatics and Synthetic Biology
Manchester, UK. 11–13 January 2007

Published: 8 May 2007

BMC Systems Biology 2007, 1(Suppl 1):P27 doi:10.1186/1752-0509-1-S1-P27

This abstract is available from: <http://www.biomedcentral.com/1752-0509/1?issue=S1>

© 2007 Swainston et al; licensee BioMed Central Ltd.

The introduction of the Proteomics Identifications Database (PRIDE) provided the proteomics community with a standards compliant data repository for proteomics data. PRIDE implements standards put forward by the Proteome Standards Initiative (PSI) including mzData.

Many commonly used proteomics software packages do not currently support these standards. As such, formatting data to adhere to the PRIDE schema requires the writing of data parsers to perform the conversion. To address this, the Pride Wizard has been introduced to perform the transformation steps required to convert more commonly used file formats into documents that adhere to the PRIDE schema. Examples of these include .mgf files for peak lists and Mascot .dat files for peptide and protein identifications.

The existing PRIDE schema has no provision for quantitative proteomics data. A new controlled vocabulary is introduced here to allow storage of quantitative I-TRAQ data. It is envisaged that this can be extended to allow submission of quantitative data from other labelling techniques.

The tool has been used to populate a PRIDE database with mass spectra and associated protein identifications and quantifications from a comprehensive set of proteomics experiments from a systems biology study of growth-rate in the eukaryotic cell.