Research Repository UCD

Title	HiQuant: Rapid postquantification analysis of large-scale MS-generated proteomics data
Authors(s)	Bryan, Kenneth, Jarbou, Mohamed-Ali, Raso, Cinzia, McCann, Brendan, Rauch, Jens, et al.
Publication date	2016-04-18
Publication information	Bryan, Kenneth, Mohamed-Ali Jarbou, Cinzia Raso, Brendan McCann, Jens Rauch, and et al. "HiQuant: Rapid Postquantification Analysis of Large-Scale MS-Generated Proteomics Data." American Chemical Society, April 18, 2016. https://doi.org/10.1021/acs.jproteome.5b01008.
Publisher	American Chemical Society
Item record/more information	http://hdl.handle.net/10197/9763
Publisher's statement	This document is the Accepted Manuscript version of a Published Work that appeared in final form in Journal of Proteome Research, copyright © [2016] American Chemical Society after peer review and technical editing by the publisher. To access the final edited and published work see http://pubs.acs.org/doi/abs/10.1021/acs.jproteome.5b01008
Publisher's version (DOI)	10.1021/acs.jproteome.5b01008

Downloaded 2026-05-01 23:44:07

The UCD community has made this article openly available. Please share how this access benefits you. Your story matters! (@ucd_oa)

© Some rights reserved. For more information

HiQuant: Rapid post-quantification analysis of large-scale MS-generated proteomics data

Kenneth Bryan,[*,†] Mohamed-Ali Jarboui,[‡] Cinzia Raso,[¶] Manuel Bernal-Llinares,[†] Brendan McCann,[¶] Jens Rauch,[¶] Karsten Boldt,[‡] and David J. Lynn[†,§]

EMBL Australia Biomedical Informatics Group, Infection & Immunity Theme, South Australian Health and Medical Research Institute, North Terrace, Adelaide, SA 5000, Australia., Medical Proteome Center Institute for Ophthalmic Research, Eberhard-Karls Universität Tübingen, Nägelestrasse 5, D-72074 Tübingen, Germany., Systems Biology Ireland, University College Dublin, Belfield, Dublin 4, Ireland., and School of Medicine, Flinders University, Bedford Park, SA 5042, Australia.

E-mail: kenneth.bryan@sahmri.com

Phone: +61(0)881284056

[*]To whom correspondence should be addressed
[†]South Australian Health and Medical Research Institute (SAHMRI)
[‡]Eberhard-Karls Universität Tübingen
[¶]University College Dublin
[§]Flinders University

Abstract

Recent advances in Mass Spectrometry (MS)-based proteomics are now facilitating ambitious large-scale investigations of the spatial and temporal dynamics of the proteome. However, the increasing size and complexity of these datasets is overwhelming current downstream computational methods, specifically those that support the post-quantification analysis pipeline. Here we present *HiQuant*, a novel application that enables the design and execution of a post-quantification workflow, including common data processing steps, such as assay normalization and grouping, replicate quality control and statistical analysis. *HiQuant* also enables the interpretation of results generated from large-scale datasets by supporting interactive heatmap analysis and also the direct export to Cytoscape and Gephi, two leading network analysis platforms. *HiQuant* may be run via a user-friendly graphical interface and also supports complete one-touch automation via a command-line mode. We evaluate *HiQuant*'s performance by analyzing a large-scale, complex interactome mapping dataset and demonstrate a 200-fold improvement in the execution time over current methods. We also demonstrate *HiQuant*'s general utility by analyzing proteome-wide quantification data generated from both a large-scale public tyrosine kinase siRNA knock-down study and an in-house investigation into the temporal dynamics of the KSR1 and KSR2 interactomes. Download *HiQuant*, sample datasets and supporting documentation at `http://hiquant.primesdb.eu`

Keywords

Bioinformatics, Proteomics, Software, Mass Spectrometry, Data Analysis, Protein Quantification, Visualization, Network Analysis, High-dimensional Data

Introduction

Recent improvements in the resolution, accuracy and speed of Mass Spectrometry (MS) are enabling a revolution in the field of high-throughput proteomics[1]. Accurate, simultaneous quantification of thousands of proteins within multiple biological samples is now routine. This has enabled researchers to characterise the proteome at an unprecedented scale[2,3], and to monitor dynamic changes in the proteome under different experimental conditions[4] . Furthermore, coupling MS with affinity purification (AP-MS) creates a high-throughput method to identify the protein-protein interactions of selected 'bait' proteins[5] and to investigate how the interactome is dynamically 're-wired' over time, under different conditions, or in disease. Recently, for example, this approach has been used to characterise the interactomes of 2,594 human proteins in HEK-293T cells[6] and to map the human Hippo signaling pathway interactome[7,8].

As with preceding advances in genomics[9], the increasing size and complexity of proteomics datasets is pushing the bottleneck from the experimental stage downstream to the data analysis. Software to complete the initial task of transforming the raw MS spectral data into peptide abundances is well developed and includes the popular MaxQuant software[10]. However, the datasets output by MaxQuant can contain tens of thousands of peptide groups, quantified over hundreds of experimental assays and may frequently contain added complexity such as multiple replicates and reverse labeling. As a result, such datasets must undergo an additional series of processing steps (often 10-15 or more) before the data can be interpreted, see Figure 1.

Several pieces of software have been developed over the last few years to support post-quantification analysis[11–14] (see Supplementary Table S2 for a comparative overview of current tools). Perseus is the most widely used, and implements a range of data manipulation, analytical and visualization functions (see www.perseus-framework.org). There are, however, several key limitations of Perseus and other related software including an excess of manual operations (i.e. 'point-and-click'), the lack of a command-line interface for high-throughput

analysis and the inability to define a workflow that can be automatically applied to related datasets (e.g. the other bait proteins in an interactome mapping experiment). As a result, current software does not scale well for large, complex proteomics datasets, which are becoming increasingly prevalent.

We found these limitations particularly evident through our involvement in the PRIMES consortium (http://www.primes-fp7.eu/), an international effort to map the dynamic interactomes of more than 90 proteins in the epidermal growth factor receptor (EGFR) pathway in an oncogenic and non-oncogenic cell-line. The PRIMES experimental design involves triple SILAC[15], forward and reverse labelling and biological and technical replication. Protein identification and quantification by MaxQuant[10] yields a dataset containing approximately 10,000 protein groups measured over more than 3,000 assays. Using Perseus, it takes an expert user at least 20-30 mins to manually apply a typical post-quantification analysis pipeline to the 36 assays generated by a *single* bait protein experiment (see Supplementary Video S1). As this analysis must be repeated for each bait, analysis of all 93 baits in the PRIMES dataset with Perseus would involve more than 40 hours of manual processing.

To address these limitations, we have developed the high-throughput protein quantification analysis tool (*HiQuant*). *HiQuant* implements a customizable post-quantification data analysis pipeline including several data processing, quality control, normalization and statistical analysis steps which can be applied simultaneously to hundreds of assays within a MS-based proteomics experiment.

In the following *Materials and Methods* section we provide an overview of the principal features of *HiQuant* and outline our experimental methods. In *Results and Discussion*, we evaluate the performance and scope of *HiQuant* by analyzing three proteome-wide quantification datasets generated across a range of experimental designs.

Materials and Methods

HiQuant Overview

Graphical and Command-Line Modes

HiQuant can be run either via a graphical user interface (GUI) or the command line, see Figure 2. The 'wizard-like' GUI, see Supplementary Figure S1, enables the user to develop and execute a custom post-quantification analysis workflow. This workflow's parameter settings can also be saved (as a plain text '.config' file), and used to execute the analysis solely via the command line interface, thus facilitating complete automation. Importantly, the plain text '.config' file can also be quickly modified to enable the analysis of future proteomic datasets. The command line interface also enables *HiQuant* to interact with other programs and to facilitate analyses that would otherwise be prohibitively time consuming. For example, one can perform the extensive iterations required to investigate the sensitivity of results to different parameter settings or assess the expected false discovery rate (FDR) for a given study, see Supplementary Data for more detail.

HiQuant Input Formats and Analysis Automation

HiQuant supports the analysis of both labelled and label-free quantification (LFQ) data and is compatible with the MaxQuant *proteinGroups.txt* file (or any other plain text table). The software also supports conversion from the HUPO proteomics standards initiative (PSI) *.mzTab* standard exchange format[16]. A protein quantification input file for a given study can contain many different experimental units (e.g. each bait protein investigated). Within each experimental unit there may be different conditions (e.g. bait A in disease; bait A in normal and bait A in empty vector control) and within each condition there is usually biological and/or technical replication (including forward and reverse labeling). Selecting and defining each column manually (e.g. as performed in Perseus) is extremely tedious and time-consuming for large datasets. Such manual processing is, additionally, a

potential source of human error[17]. *HiQuant* enables the user to define pattern matching criteria through which data columns can be automatically assigned to the various sub-groups within each experimental unit (see *HiQuant* user documentation). The ability of *HiQuant* to support the generic description of sample groups (for example when applying label-swap reversal or specifying the members of experimental groupings) is a vital part of facilitating the level of pipeline automation required to deal with large-scale proteomics datasets with complex experimental designs. *HiQuant* then supports several common data pre-processing steps including the removal of flagged contaminants, label-swap inversion, data normalization, evaluation of replicate consistency, filtering of protein groups based on the number of supporting replicates/spectra and also various replicate merging options, see Figure 1 and Supplementary Figure S2(a).

Statistical Analysis and Output Visualization

HiQuant also implements several statistical tests applicable to labelled (e.g. one-sample t-test) or label-free (e.g. two-sample t-test) data as well as the Q-function 'tail probability test' (*Significance A test*)[10] often used in proteomics. *HiQuant* outputs results (i.e. one or more lists of statistically significant protein abundances) as annotated plain text tables. Importantly, and unlike existing methods, *HiQuant* also supports the visualization of all results (e.g. all detected proteins vs. all bait experiments) via a custom, PDF exportable, heatmap, see Figure 2 and Supplementary Figure S2(b). These output data may also be exported as two attribute-rich graph formats (namely '.gexf' and '.xgmml'), see *HiQuant* user documentation. This latter feature facilitates further analysis and visualization of result data via the leading network analysis applications in biology (*Cytoscape*)[18] and machine learning (*Gephi*)[19], see Figures 2 and 3 and Supplementary Video S6.

HiQuant Implementation

HiQuant is implemented in Java and is therefore platform independent. *HiQuant* utilizes the *JFreeChart*, *JHeatChart*, *JmzTab* and *ItextPDF* java libraries. In line with recent bioinformatics software standard recommendations[20] it should be noted that *HiQuant* is currently in beta version and care must be taken by the user when entering/editing parameter values in the GUI or the plain text '.config' file. Furthermore, for large input files (e.g. 100's of Megabytes) additional RAM may need to be assigned, which requires the use of the Java runnable JAR file version of *HiQuant*, see *HiQuant* user documentation for further details. Benchmarking and analysis was performed on a standard desktop machine (MacOS 10.9.5;Intel Core i5, 3.2GHz;16Gb/1600HHz RAM). *HiQuant* is distributed under GNU general public license version 2 and may be downloaded, along with the supporting user manual, example datasets, configuration files and step-by-step tutorials at `http://hiquant.primesdb.eu`.

Quantification of the Dynamic Interactomes of KSR1 and KSR2

Cell lines

HEK-293 (human embryonic kidney) cells were cultured in standard Dulbecco's modified Eagle's medium containing 10% fetal calf serum and 2mM L-glutamine.

Immunoprecipitation & Mass spectrometry

All samples were prepared as 3 biological replicates. 24 hours post transfection, cells were serum-starved for 12-18 hours before stimulation with 10 nM epidermal growth factor (EGF) (Roche) for 0, 5, 15, 30 & 120 min. MS-analysis was performed as previously reported[21]. The complete immunoprecipitation and mass spectrometry protocol is detailed in Supplementary Data.

Generation of Protein Quantification Data

MaxQuant (Version 1.3.0.5) was used to analyse raw mass spectrometric data files from LC-MS/MS for LFQ. Default settings were used unless stated otherwise, including the following parameters: trypsin/P digest; variable modifications included oxidation of methionine and acetyl (Protein N-term); fixed modification included carbamidomethylation of cysteine only; multiplicity = 1; first search at 20 ppm; main search at 6 ppm mass accuracy (MS) and 20 mass deviation for the fragment ions; data searched against a human database (UniProt release 2014_09); minimum peptide length of 6; unfiltered for labelled amino acids; false discovery rate (FDR) of 0.01 selected for peptides and proteins; results refined through re-quantify option; match between runs selected with 1 min. time window; LFQ minimum ratio count set at 1. The full list of parameters used in MaxQuant analysis are detailed in Supplementary Data.

Results and Discussion

Case Study 1: Performance Benchmark using a Large-scale Interactome Dataset

A key feature of *HiQuant* is its ability to analyze large proteomics datasets much more efficiently in comparison to existing, more manually intensive, methods. As discussed in the introduction, using the leading post-quantification analysis method, Perseus, it would take at least 40 hours to analyze the PRIMES dynamic interactome dataset. In comparison, it takes approximately 1 min. to define the workflow using *HiQuant* GUI interface and only seconds to apply it to each bait experiment, see Supplementary Videos S2 and S3. As a result, analysis of the entire PRIMES dataset of 93 baits measured over $\approx$ 3500 assays takes just 12 mins. using *HiQuant*, see Supplementary Video S4. This represents a 200-fold improvement in straight-line performance compared to Perseus.

Case Study 2: Analysis of a Large-scale Labeled Quantification Dataset

The development of *HiQuant* also now makes it considerably more feasible to re-interrogate large-scale publicly available proteomics datasets. To demonstrate this we have also used *HiQuant* to analyze a publicly available labeled quantification dataset generated from an investigation into the impact of the siRNA knock-down of 66 tyrosine kinases (TK) on the proteome of the MCF-7 breast cancer cell-line (PRIDE project ID: PXD002065)[22]. After configuring workflow parameters, analysis of the full dataset took just 34 seconds. Here, we also made use of *HiQuant*'s ability to export the results as a bipartite 'experiment vs. protein' network file, see Figure 2(e) and Supplementary Figure S3. Network visualization greatly aids the interpretation of results, and one can immediately appreciate that the siRNA knockdown of receptor tyrosine kinases EGFR and NTRK3 (*neurotrophic tyrosine kinase receptor type 3*) have both the greatest and also a similar (32 shared 'targets') impact on the proteome of MCF-7. Interestingly, ETV6-NTRK3 is a recurrent gene fusion in Secretory Breast Carcinomas[23]. Furthermore, the antagonism of EGFR along with ERBB3 (HER3) has also been shown to enhance the response to PI3K-Akt pathway inhibitors in triple negative breast cancer[24]. One can also quickly identify the proteins most altered in their expression in response to multiple different TK knock-downs (e.g. FTHL3, CA12 and CD44). The top non-TK hubs (SYPL1, SYT7, CA12, APOB, FTHL3, CD44, TADA2B, CAV1, FTLP2, VDAC2, VDAC1P1), that are effected by the knockdown of at least 10 TK in MCF-7 are enriched for both Hexokinase (HK) binding (VDAC1P1, VDAC2) and Ferritin binding (FTLP2, FTHL3), which have been shown to be related to tumor progression[25,26] (see Supplementary Data for further discussion). This case study demonstrates the added insight that network based analysis can provide for large-scale proteomics datasets.

Case Study 3: Quantification of the Dynamic Interactomes of Scaffolding Proteins KSR1 and KSR2

To further explore the benefits of network visualization, and to demonstrate the application of *HiQuant* to LFQ data, we analyzed a previously unpublished dataset in which we investigated the dynamic interactome of two homologous scaffolding proteins, KSR1 (kinase suppressor of Ras 1) and KSR2. The primary function of KSR1 is the regulation of ERK signalling via its interactions ERK pathway kinases Raf, MEK (MAP2K1/2) and ERK (MAPK1/3). While MEK is constitutively bound, Raf (Raf-1 or B-Raf) and ERK are recruited to KSR1 upon mitogen stimulation. KSR2, a KSR1 homolog (60% identical at the amino acid level), also interacts with the core MAPK signalling machinery, but, in addition, is a key regulator of cellular energy homeostasis and is associated with obesity and insulin resistance[27]. We used an AP-MS approach to identify the interacting proteins of KSR1 and KSR2 in HEK293 cells at 0, 5, 15, 30 and 120 mins. post EGF (*epidermal growth factor*) stimulation, see *Materials and Methods*. MaxQuant was used to analyse raw mass spectrometry data files from LC-MS/MS for LFQ and *HiQuant* was used for post-quantification analysis (see Supplementary Data for parameter settings). *HiQuant* was then used to export these results as a rich graph file ('.gexf') which was then visualized using Gephi, see Supplementary Videos S5 and S6.

After EGF stimulation, a total of 362 KSR1 interactors were identified (*fold change* $\geq$ *2.0 and two-sample one-sided t-test P-value* $<$ *0.05*) in one or more of the five time-points (see Supplementary Table S1). For KSR2, 139 interactors were detected at one or more time-points. Following EGF stimulation there is a substantial re-wiring of the KSR1 and KSR2 interactomes, see Figure 3. Interestingly, 20 KSR1 interactors were also found to interact with KSR2, including MAPK signalling proteins (MAP2K1, MAP2K2 and MAPK1), supporting their known functional redundancy. However, MAPK1 appears to transiently interact KSR1 at an earlier post-stimulation time-point (t=0 min.) compared to KSR2 (t=5 min. and t=15 min.). Other novel findings include the interaction of angiomotin (AMOT) with KSR1 across all time-points, suggesting a possible point of crosstalk between MEK-ERK

and MST-LATS pathways, both of which are regulated by Raf1. Additionally, unlike KSR2, KSR1 appears to interact with both the mini-chromosome maintenance (MCM) complex (members 2-7) and the mediator (MED) complex (7 members), perhaps suggesting a greater role in DNA replication, transcriptional regulation and oncogenesis, see Supplementary Data for further discussion. This case study demonstrates the ability of *HiQuant* to support the further analysis of LFQ data and also to bridge the gap between protein quantification and more advanced interpretation of the data from a network biology perspective.

Conclusion

MS-based quantitative proteomics is now facilitating the large-scale analysis of the dynamics of both the proteome and the interactome and leading to the generation of significantly larger and more complex datasets. This in turn requires the development of new software capable of analyzing such datasets efficiently. *HiQuant* is a novel application enabling the rapid post-quantification analysis of large-scale MS-based proteomics datasets, up to 200 times faster than existing methods, and supports the downstream visualization of high-dimensional experimental results. Such datasets may also include applications other than those presented here, such as those generated from phosphoproteomics experiments, for example. *HiQuant* also supports the visualization and interpretation of results generated by large-scale investigations by coupling the post-quantification pipeline to powerful network analysis platforms. In this way *HiQuant* will provide invaluable support as researchers attempt to further understand the dynamic behaviour of the proteome and begin to elucidate the role of the interactome in complex biological systems.

Acknowledgements

Funding: The research leading to these results received funding from the European Union Seventh Framework Programme (FP7/2007-2013) PRIMES project under grant agreement

number FP7-HEALTH-2011-278568. The Lynn Group is also supported by EMBL Australia. Science Foundation Ireland under grant No. 14/IA/2395. Science Foundation Ireland under grant No. 06/CE/B1129. Many thanks to Professor Matthias Wilm for helpful input and comment.

Author Contributions

K.B. conceived of the idea, wrote the paper, designed and developed the software. M.J. conceived of the idea, and helped to design and test the data analysis workflow. C.R. carried out the experiments and designed and tested data analysis workflow. M.B.L. developed the software and supporting website. B.M. performed the scaffold protein experiments. J.R. conceived of the scaffold protein research idea and contributed to writing the paper. K.B. conceived of the idea, and helped to design and test the data analysis workflow. D.L. helped to develop and design the software and wrote the paper.

Competing Interests

The authors declare that they have no competing financial interests.

Supporting Information

Supplementary Data: Main supplementary data document.

Supplementary Figure S1: Overview of *HiQuant*'s graphical user interface.

Supplementary Figure S2: *HiQuant* graphical user interface.

Supplementary Figure S3. Visualization of the tyrosine kinase knockdown network.

Supplementary Table S1: The list of prey proteins detected in the KSR1 and KSR2 dynamic interactome (Case Study 3).

Supplementary Table S2: A comparison of the current tools used for the post-quantitative analysis of protein quantification data and HiQuant.

Supplementary Video S1: Analysis of the quantification data generated from a single bait (SRC protein) experiment (Case Study 1) using the Perseus application.

Supplementary Video S2: *HiQuant* analysis of the quantification data generated from a single bait (SRC protein) experiment (Case Study 1).

Supplementary Video S3: *HiQuant* analysis of the quantification data generated from a single bait (SRC protein) experiment (Case Study 1) using a configuration file.

Supplementary Video S4: Demo of *HiQuant*'s command line version to analyze a large-scale bait-prey dataset 93 (3500 assays) using the Runnable Jar File mode.

Supplementary Video S5: Analysing the label-free quantification data generated in the KSR1/2 study (Case Study 3) using HiQuant.

Supplementary Video S6: Step-by-step guide to generating a network visualization from *HiQuant*'s output file using Gephi.

HiQuant Website: *HiQuant* software download and running information, example datasets, configuration files and user documentation and video tutorials.

References

(1) Leitner, A.; Aebersold, R. SnapShot: Mass spectrometry for protein and proteome analyses. *Cell* **2013**, *154*, 252–252.

(2) Nagaraj, N.; Kulak, N. A.; Cox, J.; Neuhauser, N.; Mayr, K.; Hoerning, O.; Vorm, O.; Mann, M. System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra HPLC runs on a bench top Orbitrap. *Molecular & Cellular Proteomics* **2012**, *11*, M111–013722.

(3) Kim, M.-S.; Pinto, S. M.; Getnet, D.; Nirujogi, R. S.; Manda, S. S.; Chaerkady, R.; Madugundu, A. K.; Kelkar, D. S.; Isserlin, R.; Jain, S. A draft map of the human proteome. *Nature* **2014**, *509*, 575–581.

(4) Dephoure, N.; Gygi, S. P. Hyperplexing: a method for higher-order multiplexed quantitative proteomics provides a map of the dynamic response to rapamycin in yeast. *Science signaling* **2012**, *5*, rs2–rs2.

(5) Ewing, R. M.; Chu, P.; Elisma, F.; Li, H.; Taylor, P.; Climie, S.; McBroom-Cerajewski, L.; Robinson, M. D.; O'Connor, L.; Li, M. Large-scale mapping of human protein–protein interactions by mass spectrometry. *Mol. Sys. Biol.* **2007**, *3*.

(6) Huttlin, E. L. et al. The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell* **2015**, *162*, 425–440.

(7) Kwon, Y.; Vinayagam, A.; Sun, X.; Dephoure, N.; Gygi, S. P.; Hong, P.; Perrimon, N. The Hippo signaling pathway interactome. *Science* **2013**, *342*, 737–740.

(8) Couzens, A. L.; Knight, J. D.; Kean, M. J.; Teo, G.; Weiss, A.; Dunham, W. H.; Lin, Z.-Y.; Bagshaw, R. D.; Sicheri, F.; Pawson, T.; Wrana, J. L.; Choi, H.; Gingras, A.-C. Protein interaction network of the mammalian Hippo pathway reveals mechanisms of kinase-phosphatase interactions. *Science Signaling* **2013**, *6*, rs15–rs15.

(9) McPherson, J. D. Next-generation gap. *Nature methods* **2009**, *6*, S2–S5.

(10) Cox, J.; Mann, M. MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnol.* **2008**, *26*, 1367–1372.

(11) Kuzniar, A.; Kanaar, R. PIQMIe: a web server for semi-quantitative proteomics data management and analysis. *Nucleic acids research* **2014**, *42*, W100–W106.

(12) Colaert, N.; Helsens, K.; Impens, F.; Vandekerckhove, J.; Gevaert, K. Rover: a tool to visualize and validate quantitative proteomics data from different sources. *Proteomics* **2010**, *10*, 1226–1229.

(13) Taverner, T.; Karpievitch, Y. V.; Polpitiya, A. D.; Brown, J. N.; Dabney, A. R.; Anderson, G. A.; Smith, R. D. DanteR: an extensible R-based tool for quantitative analysis of-omics data. *Bioinformatics* **2012**, *28*, 2404–2406.

(14) Rigbolt, K. T.; Vanselow, J. T.; Blagoev, B. GProX, a user-friendly platform for bioinformatics analysis and visualization of quantitative proteomics data. *Molecular & Cellular Proteomics* **2011**, *10*, O110–007450.

(15) Ong, S.-E.; Mann, M. Mass spectrometry–based proteomics turns quantitative. *Nat. Chem. Biol.* **2005**, *1*, 252–262.

(16) Griss, J.; Jones, A. R.; Sachsenberg, T.; Walzer, M.; Gatto, L.; Hartler, J.; Thallinger, G. G.; Salek, R. M.; Steinbeck, C.; Neuhauser, N. The mzTab Data Exchange Format: Communicating Mass-spectrometry-based Proteomics and Metabolomics Experimental Results to a Wider Audience. *Molecular & Cellular Proteomics* **2014**, *13*, 2765–2775.

(17) Powell, S. G.; Baker, K. R.; Lawson, B. A critical review of the literature on spreadsheet errors. *Decision Support Systems* **2008**, *46*, 128–138.

(18) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* **2003**, *13*, 2498–2504.

(19) Bastian, M.; Heymann, S.; Jacomy, M. Gephi: An Open Source Software for Exploring and Manipulating Networks. International AAAI Conference on Weblogs and Social Media, 2009; pp 361–362.

(20) Martens, L.; Kohlbacher, O.; Weintraub, S. T. Managing Expectations When Publishing Tools and Methods for Computational Proteomics. *Journal of proteome research* **2015**, *14*, 2002–2004.

(21) Turriziani, B.; Garcia-Munoz, A.; Pilkington, R.; Raso, C.; Kolch, W.; von Kriegsheim, A. On-beads digestion in conjunction with data-dependent mass spectrometry: a shortcut to quantitative and dynamic interaction proteomics. *Biology* **2014**, *3*, 320–332.

(22) Stebbing, J.; Zhang, H.; Xu, Y.; Grothey, A.; Ajuh, P.; Angelopoulos, N.; Giamas, G. Characterization of the Tyrosine Kinase-Regulated Proteome in Breast Cancer by Combined use of RNA interference (RNAi) and Stable Isotope Labeling with Amino Acids in Cell Culture (SILAC) Quantitative Proteomics. *Molecular & Cellular Proteomics* **2015**, *14*, 2479–2492.

(23) Laé, M.; Fréneaux, P.; Sastre-Garau, X.; Chouchane, O.; Sigal-Zafrani, B.; Vincent-Salomon, A. Secretory breast carcinomas with ETV6-NTRK3 fusion gene belong to the basal-like carcinoma spectrum. *Modern Pathology* **2009**, *22*, 291–298.

(24) others,, et al. Antagonism of EGFR and HER3 enhances the response to inhibitors of the PI3K-Akt pathway in triple-negative breast cancer. *Science signaling* **2014**, *7*, ra29–ra29.

(25) others,, et al. Hexokinase 2 is required for tumor initiation and maintenance and its systemic deletion is therapeutic in mouse models of cancer. *Cancer cell* **2013**, *24*, 213–228.

(26) Alkhateeb, A. A.; Han, B.; Connor, J. R. Ferritin stimulates breast cancer cells through an iron-independent mechanism and is localized within tumor-associated macrophages. *Breast cancer research and treatment* **2013**, *137*, 733–744.

(27) Costanzo-Garvey, D. L. et al. KSR2 is an essential regulator of AMP kinase, energy expenditure, and insulin sensitivity. *Cell metabolism* **2009**, *10*, 366–378.

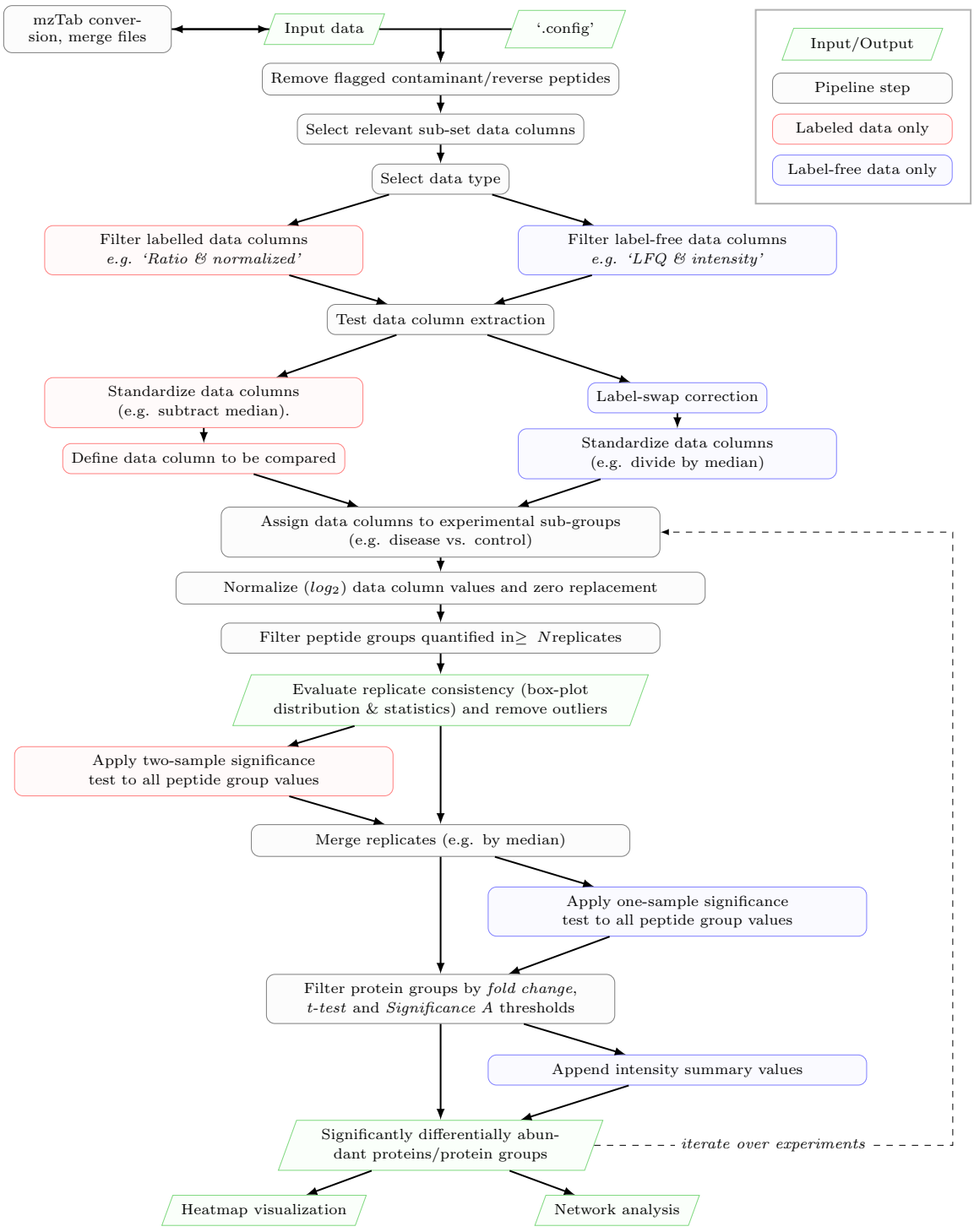


Figure 1: *HiQuant* post-quantification analysis workflow inputs, steps and outputs. The workflow may be iterated (dashed arrow) over multiple related experiments, of the same design, contained in the input file (e.g. bait proteins, siRNA knockdowns, time points etc.)

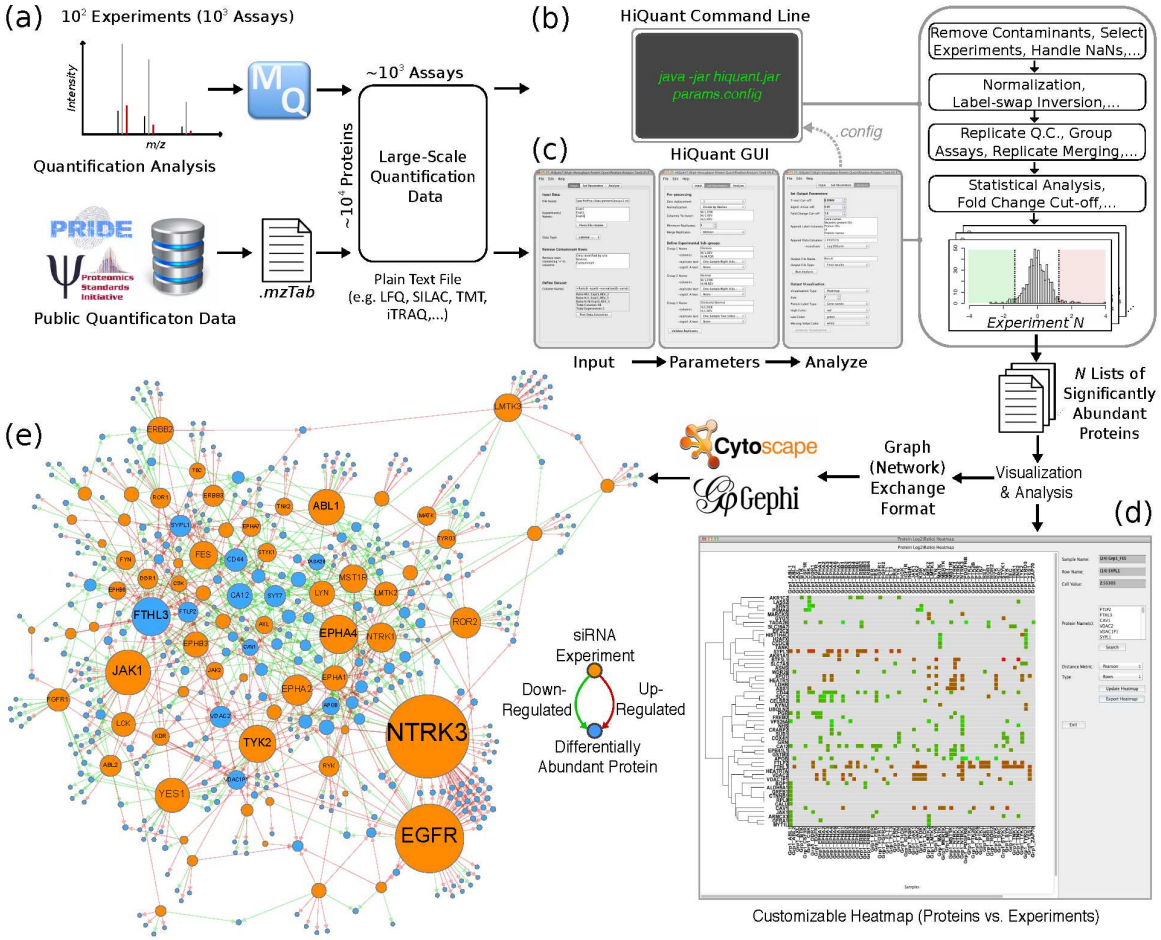
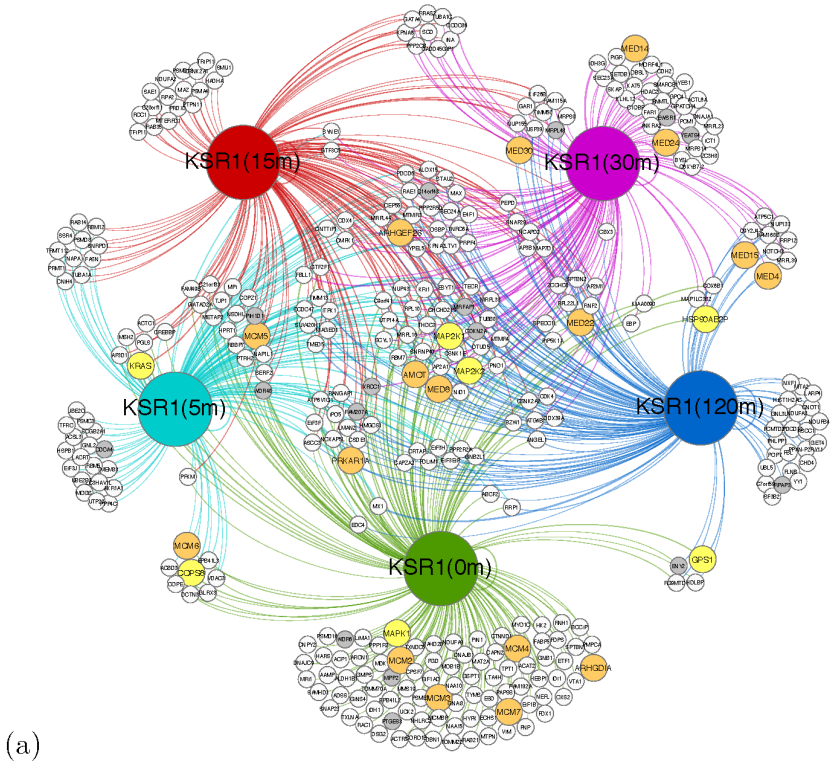
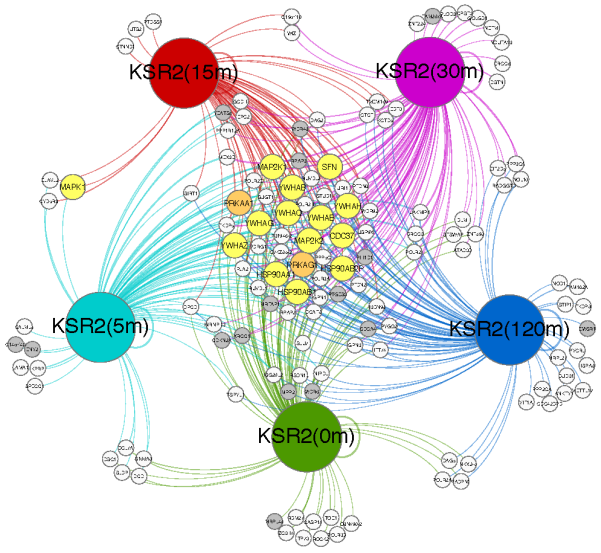


Figure 2: An overview of *HiQuant* Analysis. (a) Large-scale protein quantification data may be generated in-house or retrieved from a public repository, such as the PRIDE database. Compatible data types include those generated via LFQ or label based methods that yield relative quantification values (SILAC, TMT, iTRAQ,.. etc.)(b) *HiQuant* post-quantification analysis pipeline may be implemented via a fully automated Command Line Interface or (c) a graphical user interface (GUI). (d) Results may be exported as plain text, visualized with a customizable heatmap or (e) exported to a rich graph format to facilitate network layout and analysis via Gephi or Cytoscape. The network of 66 tyrosine kinases (TK) knocked-down by siRNA (orange nodes) in the MCF-7 cell-line and the proteins that increase (red edges) or decrease (green edges) in their relative abundance (blues nodes), node size corresponds to degree (number of incident edges), see Case Study 2 and Supplementary Data.



(a)



(b)

Figure 3: The dynamic interactomes of (a) KSR1 and (b) KSR2 at 0, 5, 15, 30, and 120 mins. post-stimulation with EGF (green, turquoise, red, magenta and blue nodes respectively). Known interactors (yellow nodes) and prey common to both KSR1 and KSR2 (grey nodes) are highlighted. See text for further details on potential novel interactors (orange). Network files (.gexf) were generated by HiQuant and laid out (ForceDirected2 algorithm), and edited, using Gephi (see Supplementary Video S6).