Published in final edited form as:

J Proteome Res. 2016 December 02; 15(12): 4747–4754. doi:10.1021/acs.jproteome.6b00660.

APOSTL: An Interactive Galaxy Pipeline for Reproducible Analysis of Affinity Proteomics Data

Brent M. Kuenzi^{1,2,‡}, Adam L. Borne^{3,‡}, Jiannong Li⁴, Eric B. Haura³, Steven A. Eschrich⁴, John M. Koomen⁵, Uwe Rix^{1,*}, and Paul A. Stewart^{3,*}

¹Department of Drug Discovery, H. Lee Moffitt Cancer Center & Research Institute, Tampa, Florida 33612-9497, United States

²Cancer Biology Ph.D. Program, University of South Florida, Tampa, Florida 33620, United States

³Department of Thoracic Oncology, H. Lee Moffitt Cancer Center & Research Institute, Tampa, Florida 33612-9497, United States

⁴Department of Biostatistics and Bioinformatics, H. Lee Moffitt Cancer Center & Research Institute, Tampa, Florida 33612-9497, United States

⁵Molecular Oncology, H. Lee Moffitt Cancer Center & Research Institute, Tampa, Florida 33612-9497, United States

Abstract

With continuously increasing scale and depth of coverage in affinity proteomics (AP-MS) data, the analysis and visualization is becoming more challenging. A number of tools have been developed to identify high confidence interactions; however, a cohesive and intuitive pipeline for analysis and visualization is still needed. Here, we present Automated Processing of SAINT Templated Layouts (APOSTL), a freely available, Galaxy-integrated software suite and analysis pipeline for reproducible, interactive analysis of AP-MS data. APOSTL contains a number of tools woven together using Galaxy workflows, which are intuitive for the user to move from raw data to publication-quality figures within a single interface. APOSTL is an evolving software project with

Correspondence should be addressed to Paul Stewart (Paul.Stewart@moffitt.org) and Uwe Rix (Uwe.Rix@moffitt.org).

[‡]These authors contributed equally

Availability

A test server running APOSTL can be accessed at <http://apostl.moffitt.org>. APOSTL is freely available from several sources. The APOSTL tools without the interactive visualization tools can be installed directly to any Galaxy instance from the Galaxy ToolShed (https://toolshed.g2.bx.psu.edu/repository?repository_id=28a2ef7daee9943d&changeset_revision=6f6e2eb3e81b). Users will also find default analysis workflows for a standard SAINTexpress analysis in the Galaxy ToolShed. APOSTL with interactive visualization can be found at the project repository (<https://github.com/bornea/APOSTL>) or can be installed to a Docker instance from the Docker Hub (https://hub.docker.com/r/bornea/apostl_shiny/). APOSTL is distributed under an Open Source License: GNU General Public License (<https://github.com/bornea/APOSTL/blob/master/LICENSE>).

Author Contributions

BMK, ALB, JL, EBH, SAE, JMK, PAS, and UR conceived and designed the project. BMK, ALB and PAS wrote the code and implemented the software. JL performed all proteomics experiments. BMK and JL analyzed the data. BMK, ALB, PAS and UR wrote the manuscript. All authors read and approved the final manuscript.

SUPPORTING INFORMATION

The following files are available free of charge at the ACS website <http://pubs.acs.org>:

Supplementary_figures.pdf: Figures S1–S7 with legends

Supplementary_methods.pdf: Experimental methods for TAP experiments

the potential to customize individual analyses with additional Galaxy tools and widgets using the R web application framework, Shiny. The source code, data and documentation are freely available from GitHub (<https://github.com/bornea/APOSTL>) and other sources.

Graphical Abstract

Keywords

Affinity proteomics; SAINT; SAINTexpress; Galaxy; data analysis; data visualization; AP-MS; pipeline; workflow

Introduction

Detailed knowledge of the interactions engaged by a protein of interest is of critical importance to understand both its molecular function and its role in biological systems. Affinity purification-mass spectrometry includes immunoprecipitation, tandem affinity purification, and drug affinity chromatography; these methods have gained significant momentum over the last decade and have been used to study numerous types of protein interactions including protein-protein interactions (PPI) and protein-ligand interactions¹. These technologies can provide key insights into large protein complexes, signaling pathways as well as novel drug targets^{2,3}.

The analysis of quantitative affinity proteomics data, however, is becoming progressively more challenging and computationally demanding due to increased throughput of baits needed to fully map signaling pathways for systems biology and the vast quantities of data generated by highly sensitive mass spectrometers. A number of computational tools have been developed to overcome these challenges⁴⁻⁹. These tools use various methods to assign probabilities to bait and prey interactions, such as by comparing spectral counts of test purifications directly to negative control experiments. Some of the most widely used tools for this purpose are the Significance Analysis of INteractome (SAINT) algorithm and the latest implementation, SAINTexpress^{7,8}. While these algorithms are robust and provide highly accurate predictions of true interactions, there remains a large barrier to the accessibility of these tools as they are operated through a command line interface and require a large amount of data reformatting. In our experience, manually formatting a single data table for a SAINTexpress analysis can take up to an hour and has many opportunities

for user error. Lastly, no visualizations are included in the stand-alone SAINTExpress software, making the analysis and visualization of large datasets including multiple baits increasingly difficult for end-users.

These difficulties constituted a need to develop a solution that could move data from unprocessed AP-MS results from various sources to publication-quality figures in a single workflow. Here, we present Automated Processing of SAINT Templated Layouts (APOSTL), which provides an intuitive user interface to identify novel interactions, interpret data, and visualize information. To our knowledge, APOSTL is the first tool available for affinity proteomics data analysis in Galaxy^{10–12}, the scientific workflow and data analysis platform. We enhance this approach by allowing for either spectral counting (Scaffold, mzIdentML and PeptideShaker¹³) or MS1 intensity data (MaxQuant)¹⁴ to be analyzed using SAINTExpress and additional downstream analysis tools whereas previous tools focused on spectral counting data alone. With APOSTL, we are able to assess various aspects of interaction proteomics at the level of individual interactions, multi-protein complexes, and pathway level enrichments. In fact, one of the most significant improvements over existing methods is the integration of all these analysis steps into a pipeline using Galaxy^{10–12}. All of these aspects are illustrated here in the context of EGFR interactome alterations of EGFR mutant non-small cell lung cancer (NSCLC) cells upon the emergence of acquired resistance to the EGFR inhibitor, erlotinib.

Materials and Methods

Software implementation and overview

APOSTL is a freely available, Galaxy-integrated software suite and analysis pipeline for reproducible, interactive analysis of AP-MS data. The software is distributed as a repository containing individual tools in the Galaxy ToolShed, which allows for customizable workflows. APOSTL consists of a workflow integrated across two analysis environments: a static Galaxy environment and an interactive data analysis environment rendered either within Galaxy or in a separate browser tab (Figure 1). In addition, we have successfully integrated SAINTExpress into Galaxy. Within the static environment, users can process their data for SAINTExpress analysis, query known PPI and generate clustered dot plots¹⁵. Within the interactive data analysis environment, APOSTL uses the R package, Shiny, to provide a user friendly GUI and provides interactive visualization of both R and JavaScript in order to analyze SAINTExpress outputs.

Pre-processing

APOSTL supports the analysis of both spectral counting and MS1 based quantification. For spectral counting-based quantification, users may import their “Samples Report” from the free Scaffold viewer (<http://www.proteomesoftware.com>, commercial license is required to import data into Scaffold) into the Galaxy environment, or generate an “Experiment report” from the Scaffold Export tool in Galaxy for pre-processing. APOSTL also supports the use of mzIdentML files generated from various search engines directly. In addition, users may use the popular proteomics/multi-omic extension of the Galaxy framework, Galaxy-P¹⁶ (<http://usegalaxy.org>) which contains the peptide/protein identification algorithms

‘SearchGUI’ and ‘Peptide Shaker’ (available free of charge in the Galaxy ToolShed)^{13,17,18}. APOSTL is able to recognize either the mzIdentML files directly or the “Protein Report” files exported from PeptideShaker. Alternatively, users may import a “peptides.txt” file from MaxQuant for pre-processing using MS1-based quantification. Using the SAINT Pre-processing tool, spectral counting or MS1 intensity data will then be reformatted into the mandatory input files for SAINTexpress: bait, interaction and prey. The user-defined bait file describes the experimental design by defining sample names, sample groups and test/control enrichments. An interaction file contains the sample names and groups as specified in the bait file as well as the accessions and spectral counts or intensity values. Lastly, the prey file contains the accession, protein length (number of amino acids) and the associated gene name. When analyzing MaxQuant data, a peptide can be mapped to multiple proteins. Therefore we first assign peptides to all possible proteins and then take the Tukey’s biweight, as described previously¹⁹, to obtain a more robust intensity estimation for each individual protein. Peptides are then combined into individual proteins for SAINT analysis. APOSTL utilizes a FASTA database during prey file generation in order to query gene names and calculate protein amino acid lengths. APOSTL expects accessions to be provided as a UniProt ID (e.g. “P00533” or “EGFR_HUMAN”).

SAINTexpress

The latest version of SAINTexpress is included in the installation of APOSTL and will be installed if SAINTexpress is not currently available on the machine operating Galaxy. Following installation, SAINTexpress is then run using a SAINTexpress wrapper within Python that can be accessed using Galaxy. Outputs from the Pre-processing tool can be directly passed to SAINTexpress using Galaxy workflows. SAINTexpress supports both spectral counting and intensity based quantification⁸, which perform similarly with a large overlap in identified interactions (Figure S1A) and can be conducted in parallel with APOSTL to further increase the confidence of identified interactions.

Data transformation

Upon importing data into the APOSTL Shiny server, APOSTL calculates both the normalized spectral abundance factor, NSAF, as described previously⁵ and the NSAF score. The NSAF score is the ratio of the NSAF for test preys versus control preys normalized by the number of controls

$$NSAF\ Score_T = \ln \left(\frac{\left(\frac{(\frac{SpC}{L})_T}{\sum_{i=1}^N (\frac{SpC}{L})_i} + \mathcal{E} \right)}{\left(\frac{(\frac{SpC}{L})_C}{\frac{1}{n} \sum_{i=1}^N (\frac{SpC}{L})_i} + \mathcal{E} \right)} \right) = \ln \left(\frac{NSAF_T + \mathcal{E}}{\frac{NSAF_C}{n} + \mathcal{E}} \right)$$

where SpC is the spectral count (or intensity) of each test T or control C prey, L is the protein length, n is the number of control purifications and the constant $\mathcal{E}$ is added to prevent division by 0. We define $\mathcal{E}$ as the inverse of the average spectral abundance factor (SAF) across all control purifications.

$$\mathcal{E} = 1 / \overline{SAF}_c = 1 / \text{mean}(\frac{SpC}{L})_c$$

The NSAF score performs similarly to the SaintScore in spectral counting data following a sigmoidal relationship (Figure S1B) whereas with MS1 intensity data, NSAF scoring tracks sigmoidally to the log2 fold-change (Figure S1C). The NSAF score is designed as a less stringent alternative to the SaintScore to complement the analysis, functioning as an empirical fold-change.

To conveniently flag frequent background proteins, APOSTL has full integration of the CRAPome²⁰ database (<http://www.crapome.org>), which is a repository for common contaminant proteins identified in AP-MS experiments. Users can query the CRAPome directly from the Galaxy environment using their prey file or a file with a single column of accessions. Additionally, if the user specifies a CRAPome file, APOSTL will calculate the probability of a true interaction based on the abundance of each protein in the CRAPome:

$$CRAPomePCT = 100 \left(1 - \frac{f_i}{N} \right)$$

where f is the frequency of identifications of protein i in the CRAPome and N is the total number of experiments annotated in the CRAPome database. We set a cutoff for *CRAPomePCT* of 80% for visualization where all proteins below 80% probability are shaded with a fixed color.

Data visualization

Following SAINTexpress analysis, users can specify SaintScore cutoffs and query known PPI as annotated in ConsensusPathDB (CPDB). Since PPI annotated in CPDB come from various sources, individual interactions in the database contain varying levels of confidence. Therefore, users can designate an interaction confidence filter when querying PPIs on a scale from 0 to 1. APOSTL will generate a simple interaction file (SIF) of these annotated interactions for import into Cytoscape (<http://www.cytoscape.org>)²¹. We have also integrated the ProHits-Viz dot plot tool¹⁵ into Galaxy (using a customized Python wrapper) for users to analyze spectral counting SAINTexpress outputs (Figure S2). As some dependencies were deprecated and not available for installation, only biclustering is supported in the Galaxy version.

APOSTL offers a variety of interactive graphs using the R Shiny framework. Users can perform quality control analyses such as correlations between replicates as well as boxplots of a user-specified protein across all enrichments. Density plots can be generated for all baits to analyze the distributions of log2 fold-changes, SaintScores, ln(NSAF), logOddsScores and NSAF scores. Users can specify cutoffs for SaintScore, log2 fold-change, and NSAF score to further filter their data. Additionally, a protein exclusion list can be appended to remove common contaminant proteins from the analysis that often have high interaction probabilities due to their large amount of identified spectra. While useful for visualizing and

removing confounders from the analysis, excluded proteins should always be reported when publishing data from this tool by exporting the analysis parameters within APOSTL. Following filtering, individual bubble graphs can be generated for all baits where each protein is represented by a circle. Axes and bubble scaling are customizable with the same options as specified in the density plots. Optionally, users can provide a CRAPome file from the CRAPome query tool in APOSTL or from workflow 1 of the CRAPome database²⁰ to assist with data filtering by allowing for a scaled bubble color based on the CRAPome probability described above. Additional visualization options (e.g. color patterns or plotting themes) are fully customizable.

Identified interactions passing SaintScore, log2 fold-change and NSAF score cutoffs can also be visualized in a network diagram directly within APOSTL using the R package visNetwork. These networks can be exported either as an image, SIF file or as a JSON network for import into Cytoscape.js. Finally, pathway analysis (KEGG) and Gene Ontology analysis can be performed using the ClusterProfiler Bioconductor package²² and visualized as bar graphs within APOSTL. Since ClusterProfiler requires Entrez gene IDs, APOSTL will convert all gene symbols into Entrez gene IDs using the mygene R package²³, selecting the hit with the highest score by default.

Results

In order to demonstrate the utility of APOSTL in analyzing affinity proteomics data, we used the well characterized rewiring of the EGFR interactome following erlotinib resistance as a benchmark. We chose to use the erlotinib resistant clone of the EGFR-mutant HCC827 cells, HCC827ER, as these cells have a known amplification and activation of the receptor tyrosine kinase, MET. MET amplification is a validated phenomenon that contributes to erlotinib resistance through activation of compensatory ERBB3 and PI3K signaling^{24–26}. To investigate this shift in signaling dependence, we performed tandem affinity purification (TAP) of EGFR, GRB2, P85B, and ERBB3 in HCC827 cells and erlotinib resistant HCC827ER cells (Figure 2). We used GFP enrichments in HCC827 as a negative control. TAP tagged proteins were expressed at similar levels across HCC827 and HCC827ER (Figure S3) with only P85B being less expressed in HCC827ER than in HCC827. Please see Supplementary Methods for experimental details.

Following LC-MS/MS analysis, data was analyzed using MaxQuant and the resulting peptides.txt file was imported into Galaxy. One of the main advantages of using APOSTL and Galaxy is the ability to implement custom workflows (Figure S4). To demonstrate this, we performed Iterative Rank Order Normalization (IRON)²⁷ and passed the normalized data to the Pre-processing tool of APOSTL. We then performed SAINTexpress analysis in two steps. First, we analyzed both HCC827 and HCC827ER TAP enrichments using GFP as a negative control to gain a global view of the interactome. Second, we analyzed the HCC827ER TAP enrichments using HCC827 as a negative control. This was done to specifically look at the prey interacting proteins that most significantly contribute to rewiring following acquired inhibitor resistance. These datasets were then passed to APOSTL's interactive environment for analysis and visualization (Figure S4).

During the first phase of the analysis, we analyzed the quality of our datasets, first by looking at the correlations between our biological and technical replicates (Figure 3A) and by looking at protein abundances across all replicates (Figure 3B). In order to better assign our cutoffs for analysis, we analyzed density plots of the log2 fold-change, SaintScore, and logOddsScores across all replicates. The majority of proteins had a logOddsScore of 0 (Figure 3C); therefore, we applied relaxed filtering criteria (SaintScore >0.5, log2 fold-change >10) to gain a broad view of the interactome. Using APOSTL's network functionality, we analyzed the shifting interactome upon onset of EGFR inhibitor resistance. There is a large amount of overlap between the EGFR interactomes of HCC827 and HCC827ER with shared nodes including known adaptor proteins SHC1/2 and IRS1/2 (Figure 4A, S5). The HCC827ER interactome had very few nodes that were specific to HCC827ER. However, one specific node, MET, has been described to contribute to erlotinib resistance in this model²⁴.

During phase two of the analysis, we used the TAP enrichments in HCC827 as the negative control for the HCC827ER experiments and reanalyzed the data in order to look for significant differences in enrichment between HCC827 and HCC827ER. Here, we applied more stringent filtering (SaintScore > 0.75) to assign only the highest confidence interactions that are more significantly enriched upon erlotinib resistance. We observed a significant enrichment of MET protein with the GRB2 TAP experiments as we had seen in phase one of the analysis (Figure 4B/C). RTK:GRB2 association is an important predictor of RTK activation²⁸. Increased MET:GRB2 association could be indicative of increased signaling through MET, which is the known mechanism of resistance in this model²⁴. Interestingly, while both the HCC827 and HCC827ER TAP enrichments share connections with PIK3R1 (P85A) (Figure 4A, S5), it is much more significantly enriched in the HCC827ER experiments. This suggests that HCC827ER may have increased PI3K signaling, consistent with our previous data showing that HCC827ER cells are exquisitely sensitive to siRNA-mediated knockdown of PI3K signaling components such as P85B and PK3CA²⁹. We also observed an increased association of ERBB4 with GRB2, P85B and ERBB3 (Figure 4B/C). Interestingly ERBB4 is known to associate with GRB2 and P85B, as annotated in CPDB (Figure S6). ERBB4 contributes to acquired resistance to ERBB2/HER2 inhibitors¹ and this interaction could suggest that ERBB4 contributes to erlotinib resistance in this model. Consistently, KEGG pathway analysis and GeneGO analysis (molecular function) significantly enriched for ERBB signaling (Figure 4D) as well as for protein kinase binding proteins (Figure 4E). Together, these data demonstrate the well described role of compensatory activation of MET, PI3K and ERBB3/4 signaling in EGFR TKI resistance^{24–26}.

Discussion

We have developed APOSTL, a Galaxy integrated software suite and analysis pipeline aimed to make the analysis of AP-MS data more accessible and reproducible. APOSTL supports both spectral counting and MS1 intensity datasets for SAINTexpress analysis and interactive visualization. One of the main advantages of APOSTL over other analysis methods is the ability to analyze AP-MS data at the peptide level using MS1 intensity quantification as well as spectral counting. MS1-based measurements are typically more

accurate with a better linear dynamic range. Thus, MS1-based quantification can provide greater accuracy in the low abundance range, which is usually lost with spectral counting³⁰. Additionally, while APOSTL was developed for the analysis of label free affinity proteomics, it could be extended to AP-MS experiments coupled with stable isotope labeling^{31–33} using customized Galaxy workflows to extract data at the peptide level through the development of additional tools.

APOSTL is suitable for the analysis of a wide variety of affinity proteomics applications, including tandem affinity purification mass spectrometry, drug affinity chromatography, and proximity-dependent biotin identification (BioID)³⁴ across human, murine or yeast datasets. All of these applications benefit from more reproducible research and easier replication of results using APOSTL's ability to export analysis parameters and Galaxy's built-in workflow features. An additional benefit of APOSTL is a large increase in analysis efficiency. Individual steps in a typical AP-MS analysis pipeline have been described to take anywhere from 1 day to 1 week³⁵. APOSTL accomplishes much of this within a few minutes by automating pre-processing steps and by providing an intuitive data analysis environment. Since individual APOSTL tools are integrated using Galaxy workflows, APOSTL is also highly adaptable with the potential to incorporate custom workflows, additional analysis tools, and new statistical algorithms.

Alternatively, data can be exported at each stage in the pipeline for offline examination or for analysis using other bioinformatics tools. The scalability of the APOSTL Galaxy platform also offers an institutional-level solution for an affinity proteomics data analysis platform that is accessible to multiple researchers. This allows for parallel analyses to occur across an institution, between institutions, and also enables for group training workshops. Galaxy-based tools and workflows, such as APOSTL, are easily distributed through the Galaxy ToolShed enabling any institution to set up local instances of the framework.

Conclusions

To our knowledge, APOSTL is the first software suite for the analysis of affinity proteomics data available within Galaxy. The possibility of analyzing these data within a single tool represents a promising approach to improving the efficiency and reproducibility of affinity proteomics data analysis. At the same time, APOSTL's adaptability provides a powerful way to customize individual analyses and support for further APOSTL development.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We wish to acknowledge the Moffitt Lung Cancer Center of Excellence and the Moffitt Proteomics, Bioinformatics and Biostatistics Core Facilities. Moffitt Core Facilities are supported by the National Cancer Institute (Award No. P30-CA076292) as a Cancer Center Support Grant. Proteomics is also supported by the Moffitt Foundation. We would also like to thank Guillermo Gonzalez-Calderon for critical assistance hosting the APOSTL test server.

ABBREVIATIONS

AP-MS	Affinity Proteomics
APOSTL	Automated Processing of SAINT Templated Layouts
SAINT	Significance Analysis of Interactomes
PPI	Protein-protein interaction
NSAF	Normalized Spectral Abundance Factor
TAP	Tandem Affinity Purification

References

1. Canfield K, Li J, Wilkins OM, Morrison MM, Ung M, Wells W, Williams CR, Liby KT, Vullhorst D, Buonanno A, et al. Receptor Tyrosine Kinase ERBB4 Mediates Acquired Resistance to ERBB2 Inhibitors in Breast Cancer Cells. *Cell Cycle*. 2015; 14:648–655. [PubMed: 25590338]
2. Woods NT, Mesquita RD, Sweet M, Carvalho MA, Li X, Liu Y, Nguyen H, Thomas CE, Iversen ES, Marsillac S, et al. Charting the Landscape of Tandem BRCT Domain-Mediated Protein Interactions. *Sci Signal*. 2012; 5:rs6. [PubMed: 22990118]
3. Gupta GD, Coyaude É, Gonçalves J, Mojarad BA, Liu Y, Wu Q, Gheiratmand L, Comartin D, Tkach JM, Cheung SWT, et al. *Cell*. 2015; 163:1484–1499. [PubMed: 26638075]
4. Zybailov B, Coleman MK, Florens L, Washburn MP. Correlation of Relative Abundance Ratios Derived from Peptide Ion Chromatograms and Spectrum Counting for Quantitative Proteomic Analysis Using Stable Isotope Labeling. *Anal Chem*. 2005; 77:6218–6224. [PubMed: 16194081]
5. Zybailov B, Mosley AL, Sardi ME, Coleman MK, Florens L, Washburn MP. Statistical Analysis of Membrane Proteome Expression Changes in *Saccharomyces cerevisiae*. *J Proteome Res*. 2006; 5:2339–2347. [PubMed: 16944946]
6. Zhang Y, Wen Z, Washburn MP, Florens L. Refinements to Label Free Proteome Quantitation: How to Deal with Peptides Shared by Multiple Proteins. *Anal Chem*. 2010; 82:2272–2281. [PubMed: 20166708]
7. Choi H, Larsen B, Lin ZY, Breitkreutz A, Mellacheruvu D, Fermin D, Qin ZS, Tyers M, Gingras AC, Nesvizhskii AI. SAINT: Probabilistic Scoring of Affinity Purification-Mass Spectrometry Data. *Nat Methods*. 2011; 8:70–73. [PubMed: 21131968]
8. Teo G, Liu G, Zhang J, Nesvizhskii AI, Gingras AC, Choi H. SAINTexpress: Improvements and Additional Features in Significance Analysis of INTERactome Software. *J Proteomics*. 2014; 100:37–43. [PubMed: 24513533]
9. Wenger CD, Phanstiel DH, Lee MV, Bailey DJ, Coon JJ. COMPASS: A Suite of Pre- and Post-Search Proteomics Software Tools for OMSSA. *Proteomics*. 2011; 11:1064–1074. [PubMed: 21298793]
10. Goecks J, Nekrutenko A, Taylor J. Galaxy Team. Galaxy: A Comprehensive Approach for Supporting Accessible, Reproducible, and Transparent Computational Research in the Life Sciences. *Genome Biol*. 2010; 11:R86. [PubMed: 20738864]
11. Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A, Taylor J. Galaxy: A Web-Based Genome Analysis Tool for Experimentalists. *Curr Protoc Mol Biol*. 2010; Chapter 19(Unit 19.10.1–21)
12. Taylor J, Schenck I, Blankenberg D, Nekrutenko A. Using Galaxy to Perform Large-Scale Interactive Data Analyses. *Curr Protoc Bioinformatics*. 2007; Chapter 10(Unit 10.5)
13. Vaudel M, Burkhardt JM, Zahedi RP, Oveland E, Berven FS, Sickmann A, Martens L, Barsnes H. PeptideShaker Enables Reanalysis of MS-Derived Proteomics Data Sets. *Nat Biotech*. 2015; 33:22–24.

14. Cox J, Mann M. MaxQuant Enables High Peptide Identification Rates, Individualized P.p.b.-Range Mass Accuracies and Proteome-Wide Protein Quantification. *Nat Biotechnol.* 2008; 26:1367–1372. [PubMed: 19029910]
15. Knight JDR, Liu G, Zhang JP, Pasculescu A, Choi H, Gingras AC. A Web-Tool for Visualizing Quantitative Protein-Protein Interaction Data. *Proteomics.* 2015; 15:1432–1436. [PubMed: 25422071]
16. Jagtap PD, Johnson JE, Onsongo G, Sadler FW, Murray K, Wang Y, Shenykman GM, Bandhakavi S, Smith LM, Griffin TJ. Flexible and Accessible Workflows for Improved Proteogenomic Analysis Using the Galaxy Framework. *J Proteome Res.* 2014; 13:5898–5908. [PubMed: 25301683]
17. Barsnes H, Vaudel M, Colaert N, Helsens K, Sickmann A, Berven FS, Martens L. Compomics-Utilities: An Open-Source Java Library for Computational Proteomics. *BMC Bioinformatics.* 2011; 12:1–6. [PubMed: 21199577]
18. Vaudel M, Barsnes H, Berven FS, Sickmann A, Martens L. SearchGUI: An Open-Source Graphical User Interface for Simultaneous OMSSA and X!Tandem Searches. *Proteomics.* 2011; 11:996–999. [PubMed: 21337703]
19. Stewart PA, Parapatics K, Welsh EA, Müller AC, Cao H, Fang B, Koomen JM, Eschrich SA, Bennett KL, Haura EB. A Pilot Proteogenomic Study with Data Integration Identifies MCT1 and GLUT1 as Prognostic Markers in Lung Adenocarcinoma. *PLoS One.* 2015; 10:e0142162. [PubMed: 26539827]
20. Mellacheruvu D, Wright Z, Couzens AL, Lambert JP, St-Denis NA, Li T, Miteva YV, Hauri S, Sardi ME, Low TY, et al. The CRAPome: A Contaminant Repository for Affinity Purification-Mass Spectrometry Data. *Nat Methods.* 2013; 10:730–736. [PubMed: 23921808]
21. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* 2003; 13:2498–2504. [PubMed: 14597658]
22. Yu G, Wang LG, Han Y, He QY. clusterProfiler: An R Package for Comparing Biological Themes Among Gene Clusters. *Omi A J Integr Biol.* 2012; 16:284–287.
23. Mark, A.; Thompson, R.; Wu, C. R Packag. version 1.6.0. 2014. Mygene: Access MyGene Info Services.
24. Suda K, Murakami I, Katayama T, Tomizawa K, Osada H, Sekido Y, Maehara Y, Yatabe Y, Mitsudomi T. Reciprocal and Complementary Role of MET Amplification and EGFR T790M Mutation in Acquired Resistance to Kinase Inhibitors in Lung Cancer. *Clin Cancer Res.* 2010; 16:5489–5498. [PubMed: 21062933]
25. Engelman JA, Zejnullahu K, Mitsudomi T, Song Y, Hyland C, Park JO, Lindeman N, Gale CM, Zhao X, Christensen J, et al. MET Amplification Leads to Gefitinib Resistance in Lung Cancer by Activating ERBB3 Signaling. *Science (80-).* 2007; 316:1039–1043.
26. Gazdar AF. Activating and Resistance Mutations of EGFR in Non-Small-Cell Lung Cancer: Role in Clinical Response to EGFR Tyrosine Kinase Inhibitors. *Oncogene.* 2009; 28:S24–S31. [PubMed: 19680293]
27. Welsh EA, Eschrich SA, Berglund AE, Fenstermacher DA. Iterative Rank-Order Normalization of Gene Expression Microarray Data. *BMC Bioinformatics.* 2013; 14:153. [PubMed: 23647742]
28. Lowenstein EJ, Daly RJ, Batzer AG, Li W, Margolis B, Lammers R, Ullrich A, Skolnik EY, Bar-Sagi D, Schlessinger J. The SH2 and SH3 Domain-Containing Protein GRB2 Links Receptor Tyrosine Kinases to Ras Signaling. *Cell.* 1992; 70:431–442. [PubMed: 1322798]
29. Li J, Bennett K, Stukalov A, Fang B, Zhang G, Yoshida T, Okamoto I, Kim JY, Song L, Bai Y, et al. Perturbation of the Mutated EGFR Interactome Identifies Vulnerabilities and Resistance Mechanisms. *Mol Syst Biol.* 2013; 9:705. [PubMed: 24189400]
30. Choi H, Glatter T, Gstaiger M, Nesvizhskii AI. SAINT-MS1: Protein-Protein Interaction Scoring Using Label-Free Intensity Data in Affinity Purification-Mass Spectrometry Experiments. *J Proteome Res.* 2012; 11:2619–2624. [PubMed: 22352807]
31. Blagoev B, Kratchmarova I, Ong SE, Nielsen M, Foster LJ, Mann M. A Proteomics Strategy to Elucidate Functional Protein-Protein Interactions Applied to EGF Signaling. *Nat Biotechnol.* 2003; 21:315–318. [PubMed: 12577067]

32. Ranish JA, Yi EC, Leslie DM, Purvine SO, Goodlett DR, Eng J, Aebersold R. The Study of Macromolecular Complexes by Quantitative Proteomics. *Nat Genet.* 2003; 33:349–355. [PubMed: 12590263]
33. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M. Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Mol Cell Proteomics.* 2002; 1:376–386. [PubMed: 12118079]
34. Kim DI, Birendra KC, Zhu W, Motamedchaboki K, Doye V, Roux KJ. Probing Nuclear Pore Complex Architecture with Proximity-Dependent Biotinylation. *Proc Natl Acad Sci U S A.* 2014; 111:E2453–E2461. [PubMed: 24927568]
35. Morris JH, Knudsen GM, Verschueren E, Johnson JR, Cimerancic P, Greninger AL, Pico AR. Affinity Purification–mass Spectrometry and Network Analysis to Understand Protein-Protein Interactions. *Nat Protoc.* 2014; 9:2539–2554. [PubMed: 25275790]

Figure 1. APOSTL overview
A pipeline for identifying and visualizing high confidence protein interactions.

Figure 2. EGFR interactome analysis workflow

Analysis workflow of EGFR interactome in HCC827 and erlotinib resistant HCC827ER cells. ECD = extracellular domain, TK = tyrosine kinase domain, PKD = pseudo-kinase domain, SH: Src homology domain. Box indicates processes performed within Galaxy.

Figure 3. Quality control plots in APOSTL

A. Replicate correlations between technical replicates of EGFR TAP experiments in HCC827 cells. **B.** Representative boxplot for ABCC5 protein present in all datasets. Biological and technical replicates are collapsed into experiments and rendered as a boxplot. **C.** Density plot of the log odds score across all replicates for HCC827 and HCC827ER. HCC827 ERBB3 experiments were excluded for visualization of the density plot. GFP enrichments are not displayed as they are used as negative controls and are not influenced by data filtering.

Figure 4. Analysis of EGFR inhibitor resistance interactome alterations

A. EGFR interactome before and after onset of erlotinib resistance. Data was filtered following SAINT analysis (SaintScore >0.5, log₂ fold-change >10) using GFP as negative control. Data was analyzed in cytoscape following SIF export in APOSTL. Parental HCC827 specific nodes are highlighted in light blue, HCC827ER specific nodes are highlighted in dark blue and overlapping interactions are highlighted in orange. **B.** Bubble plot of HCC827ER enrichments following SAINT analysis using parental HCC827 enrichments as the control. Saint score of 0.75 was used for filtering. CrapomePCT is defined as the probability of a true interaction based on the abundance of each protein in the CRAPome. Proteins with CrapomePCT >80% are labeled. The bait proteins GRB2 and PIK3R2 were excluded from the GRB2 and P85B TAP enrichment visualization,

respectively **C.** EGFR interactome following erlotinib resistance. **D.** KEGG pathway analysis of HCC827ER enrichments following SAINT analysis and filtering (SaintScore=0.5). **E.** Gene ontology (molecular function) analysis of HCC827ER enrichments following SAINT analysis and filtering (SaintScore=0.5).