



APOSTL: An interactive Galaxy Pipeline For Reproducible Analysis of Affinity Proteomics Data

Kuenzi, B.M.^{1,2,‡}; Borne, A.L.^{3,‡}; Li, J.⁴; Haura, E.B.³; Eschrich, S.A.⁴; Koomen, J.M.⁵; Rix, U.^{1,§}; Stewart, P.A.^{3,§}

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

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

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

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

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

[‡] These authors contributed equally
[§] Corresponding authors



H. LEE MOFFITT CANCER CENTER & RESEARCH INSTITUTE,
AN NCI COMPREHENSIVE CANCER CENTER – Tampa, FL
1-888-MOFFITT (1-888-663-3488) www.MOFFITT.org

© 2010 H. Lee Moffitt Cancer Center and Research Institute, Inc.

Abstract

BACKGROUND:

- AP-MS data analysis is becoming progressively more challenging
- Formatting large data tables for SAINT analysis is laborious without programming knowledge
- Visualizing large datasets including multiple baits becomes increasingly difficult
- Automated Processing of SAINT Templated Layouts (APOSTL)**
 - freely available, Galaxy-integrated software suite and analysis pipeline for reproducible, interactive analysis of AP-MS data.

INPUTS / PRE-PROCESSING:

- Scaffold “Samples Report”
 - Spectral Counting
- Maxquant “peptides.txt”
 - MS1 Intensity → Tukey’s biweight (peptides) → Summed protein intensity
- Reformats into *inter*, *prey* and *bait* files

SAINTexpress:

- Latest version included with installation

CALCULATIONS:

$$NSAF\ Score_T = \ln \left(\frac{\left(\frac{SpC_T}{L_T} + \epsilon \right)}{\sum_{i=1}^N \left(\frac{SpC_i}{L_i} + \epsilon \right)} \right) = \ln \left(\frac{NSAF_T + \epsilon}{\sum_{i=1}^N \left(\frac{SpC_i}{L_i} + \epsilon \right)} \right)$$

$$\epsilon = 1/SAF_c = 1/mean\left(\frac{SpC}{L}\right)_c$$

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

AVAILABILITY:

- Test server: <http://apostl.moffitt.org>
- Galaxy Toolshed
- Github: <https://github.com/bornea/APOSTL>
- Docker Hub: https://hub.docker.com/r/bornea/apostl_shiny/

VISUALIZATION:

- Replicate correlations
- Protein boxplots
- Density plots
- Bubble graphs
- Bait – Prey Network
- Data table
- KEGG pathway analysis
- Gene Ontology analysis
- PPI network
- Prohits Dotplots

Conclusions

- First software suite for AP-MS analysis in Galaxy
- Increased AP-MS analysis efficiency and reproducibility
 - Hours → Minutes
- Highly adaptable
 - Spectral counting & MS1 compatible
 - Customizable workflows
 - Numerous applications: TAP, BioID, drug affinity chromatography
- Scalable analysis solution
 - Cluster integration, parallel analyses

References

- Zybailov B. et al. *Anal. Chem.* 2005
- Zybailov B. et al. *J. Proteome Res.* 2006
- Choi H. et al. *Nat. Methods* 2011
- Teo G. et al. *J. Proteomics* 2014
- Knight JDR. et al. *Proteomics* 2015

APOSTL Features

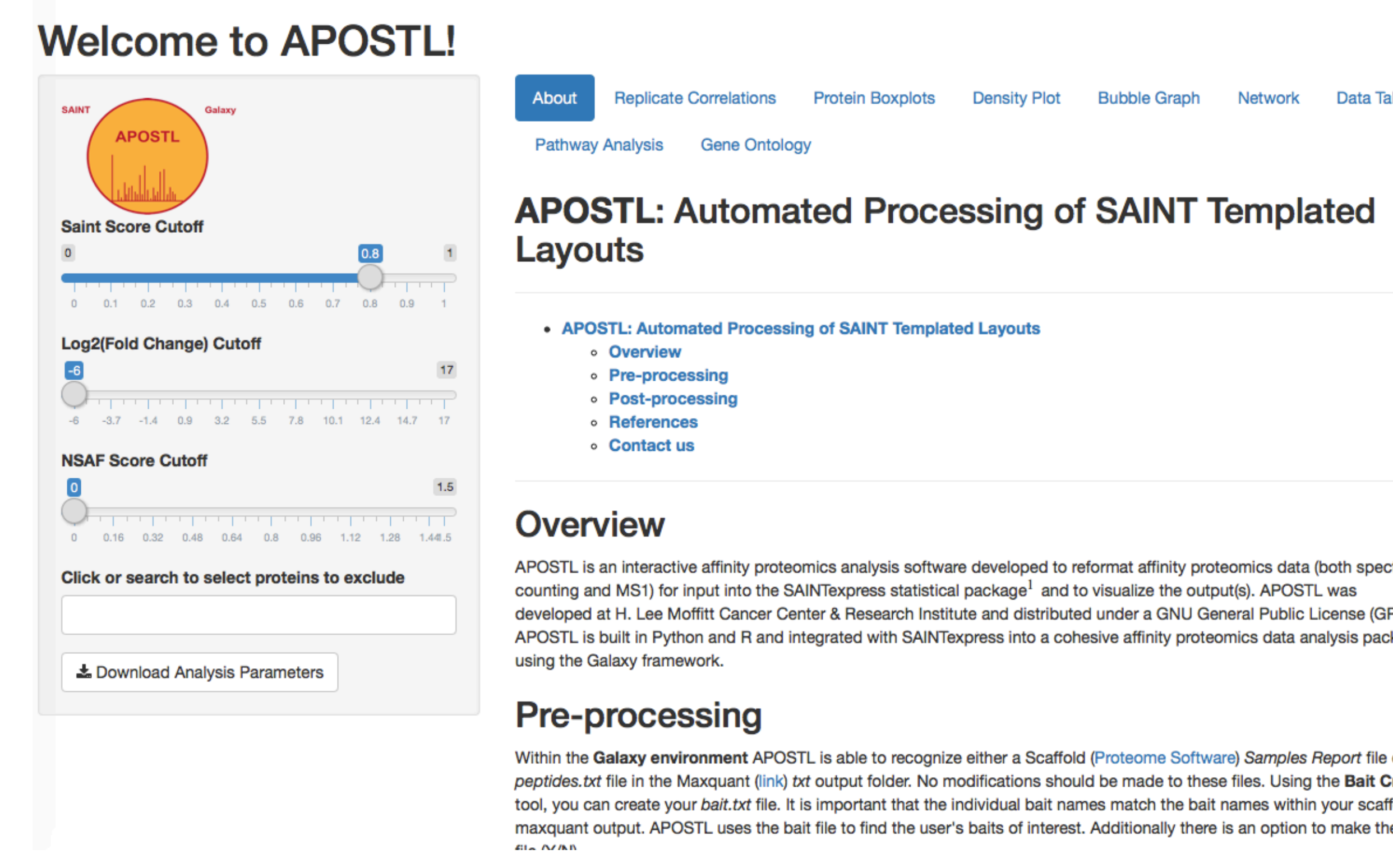
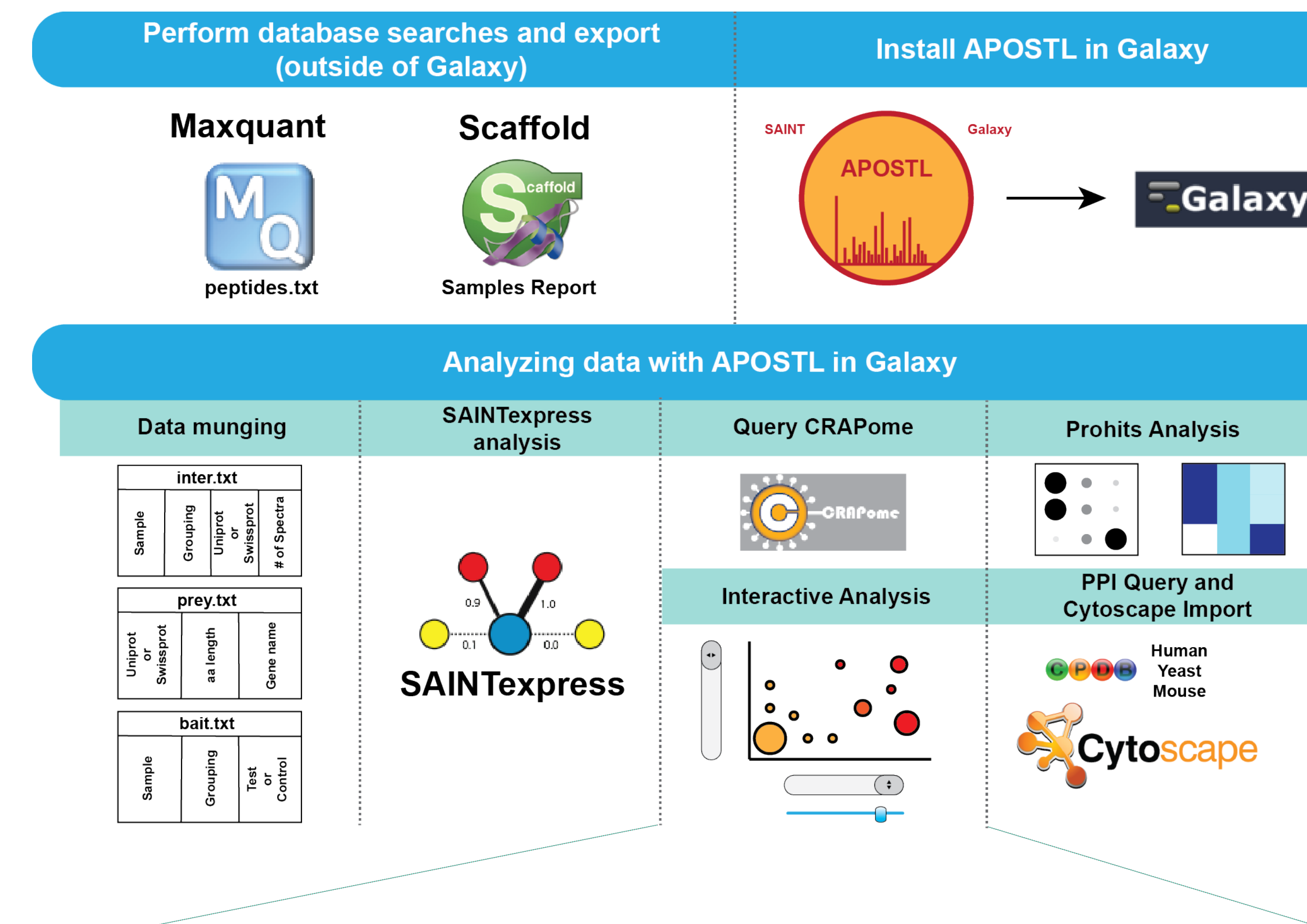


Figure 1. APOSTL overview and interactive environment graphical user-interface

Experimental Design

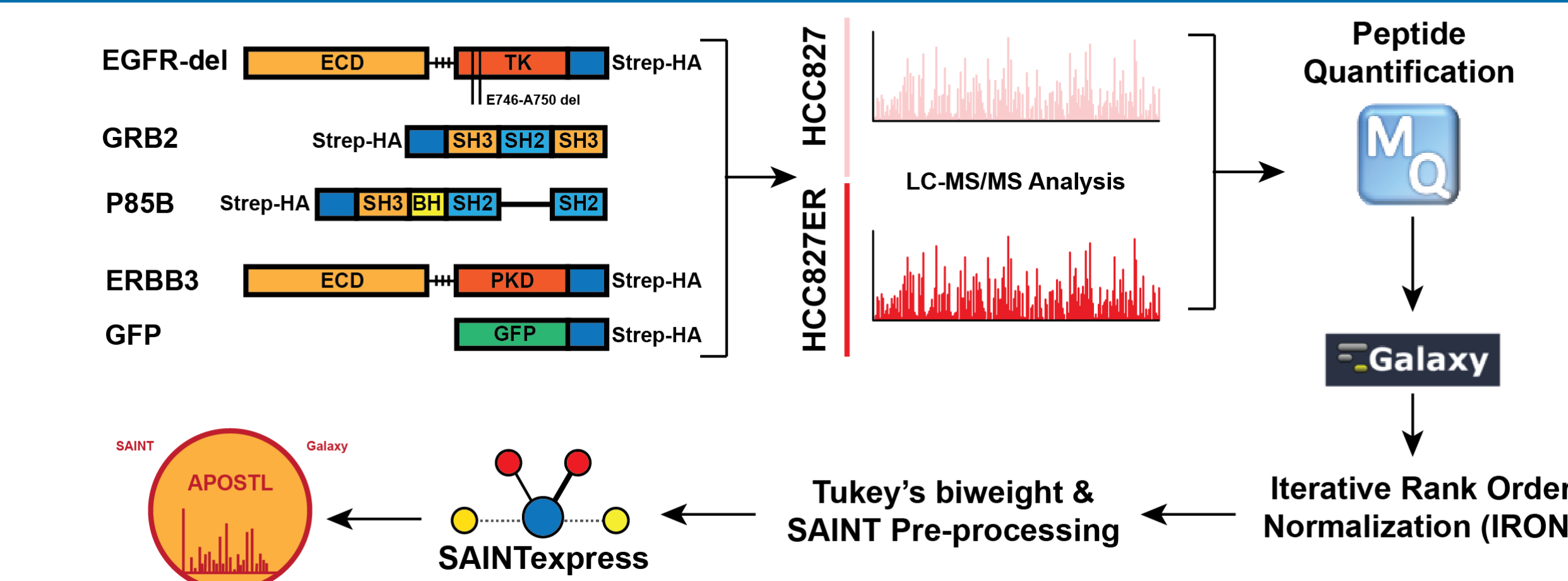


Figure 2. TAP experimental design for APOSTL benchmarking

EGFR Interactome

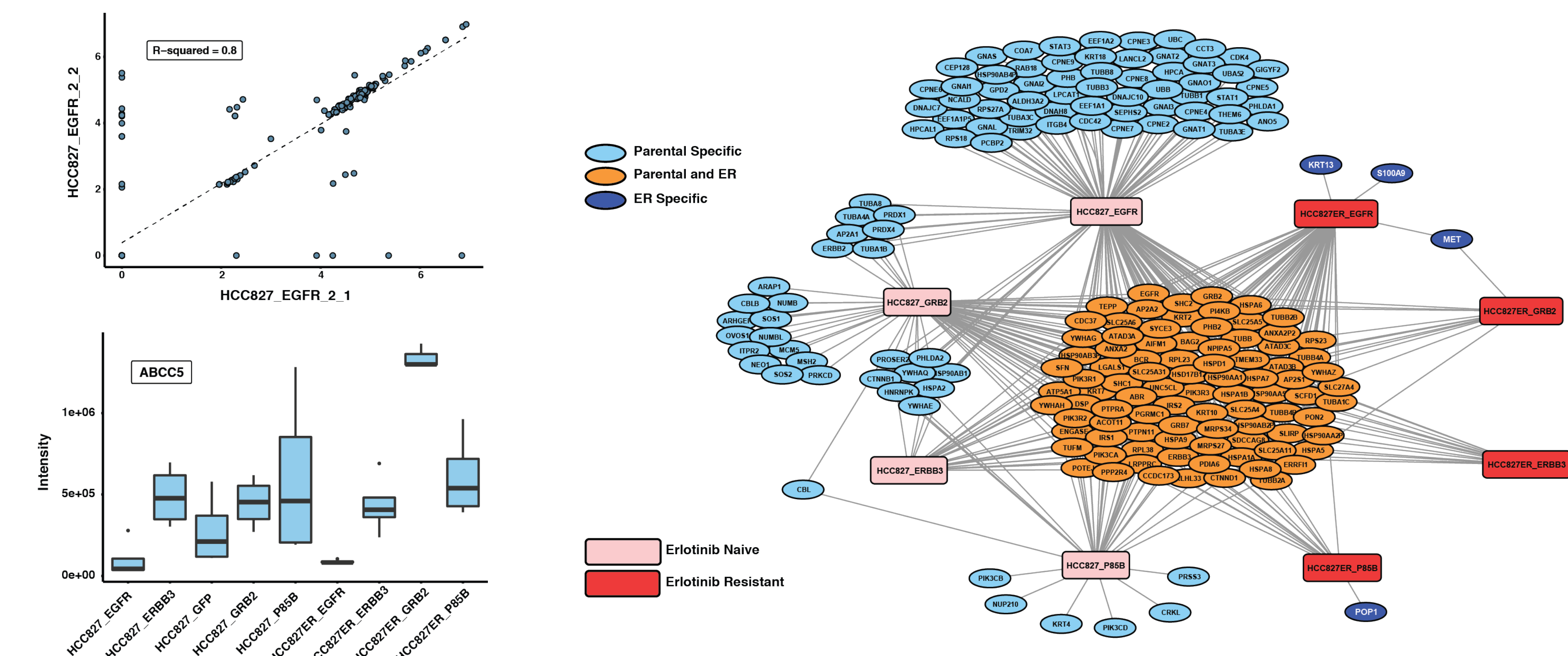


Figure 3. Example QC analysis using APOSTL's interactive analysis tool

Figure 4. EGFR interactome before and after onset of erlotinib resistance

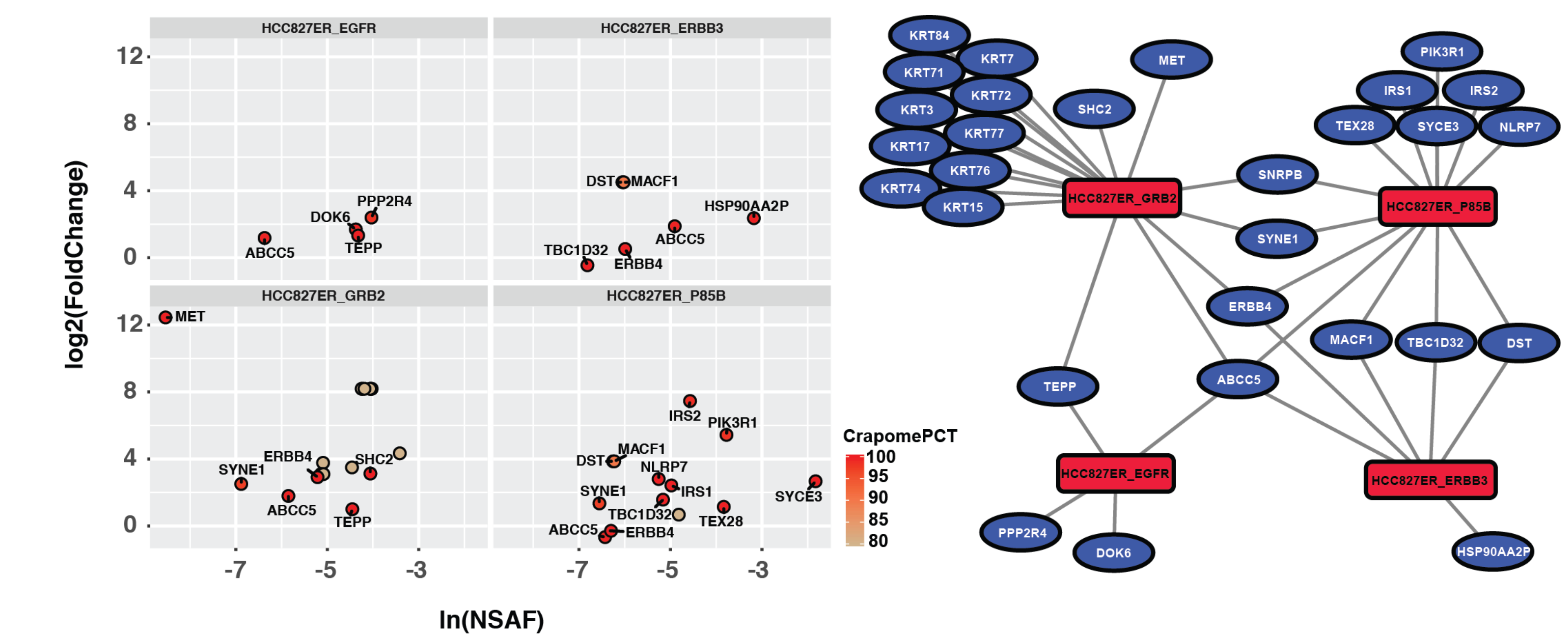


Figure 5. Protein-protein interactions significantly enriched in context of erlotinib resistance

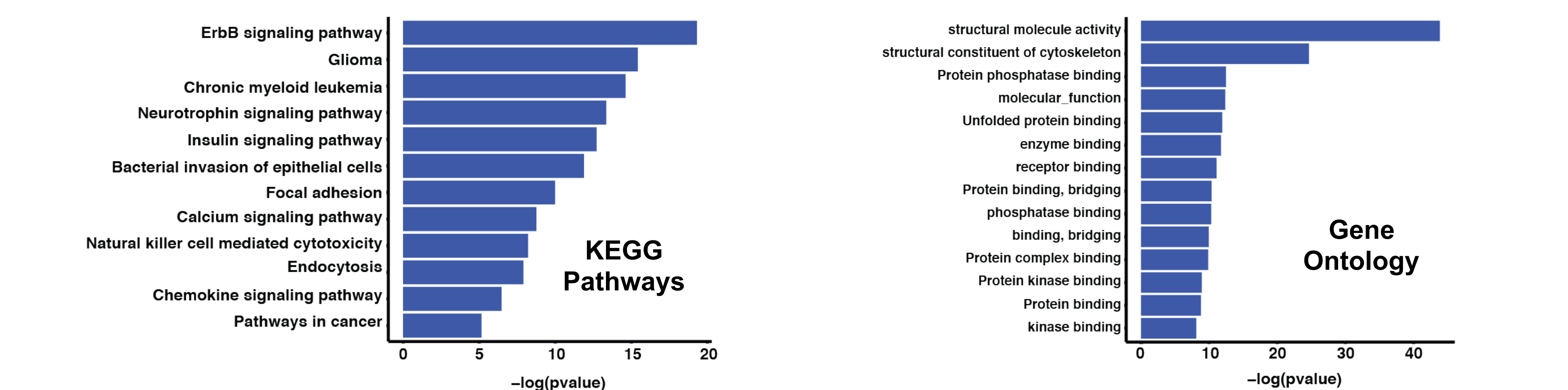


Figure 6. KEGG pathway and Gene Ontology analysis of interactions enriched in erlotinib resistant cells