CAT PETR: A Graphical User Interface for Differential Analysis of ProteinPhosphorylation Data

Keegan Flanagan[1,2], Steven Pelech[3,4], Yossef Av-Gay[1*] and Khanh Dao Duc[2*]

[1]Department of Microbiology and Immunology, University of British Columbia, Vancouver, BC V6T 1Z3, Canada
[2]Department of Mathematics, University of British Columbia, Vancouver, BC V6T 1Z4, Canada
[3]Faculty of Medicine, University of British Columbia, Vancouver, BC V6H 3Z6, Canada
[4]Kinexus Bioinformatics Corporation, Suite 1 – 8755 Ash Street, Vancouver, V6P 6T3, Canada

[*]To whom correspondence should be addressed: yossi@mail.ubc.ca, kdd@math.ubc.ca

Abstract

Analysis of signal transduction pathways relies on monitoring global changes in cellular response to perturbation or genetic manipulation. Antibody microarray data provides an increasingly viable tool for the monitoring and analyzing high throughput expression and protein phosphorylation data. However, computational tools designed specifically for the analysis of antibody microarray data are few and far between; forcing researchers to use applications primarily designed for other data types. Here we present CAT PETR (Convenient Analysis Tool for Phosphorylation and Expression Testing), a user friendly web application that provides interactive and publication ready plots for the differential analysis of expression and phosphorylation data collected via antibody microarrays. Our tool is made available as a web application at https://av-gay-ubc.shinyapps.io/CAT-PETR/. The code is publicly available on GitHub (https://github.com/Raif-fl/CAT-PETR) with documentation containing tutorials and datasets.

Introduction

Antibody microarrays use immobilized antibody spots to capture specific targets from a given sample, allowing for the binding of many different targets to be quantified [1]. The ever growing number of array validated antibodies have made antibody microarrays an increasingly powerful tool for biomarker and drug target discovery [2], protein modification analysis [3–5], and protein-protein interaction detection [6]. Yet, computational tools are still lacking to efficiently analyze the data. While a variety of analytical tools are available for RNA-seq and DNA microarray data [7–9], the inputs to these applications are often data type specific, making the use of antibody microarray data difficult. As a result, the only current way to analyze antibody microarray data is through a combination of packages, coding languages, and web applications, increasing the chance for inappropriate or unoptimized analysis methods to be used.

Here we present CAT PETR, an application that combines differential analysis tools with flexible and interactive visualizations to create a streamlined, all in one tool for exploratory expression and phosphorylation analysis of antibody microarray data, or data collected via RNA sequencing, DNA microarray, and mass spectrometry. CAT PETR is available as a web application or can be easily downloaded and run as a desktop application. A user manual including a tutorial and instructions for different use cases and example file formats is also available in the user manual SI file and as part of the application.

General Description

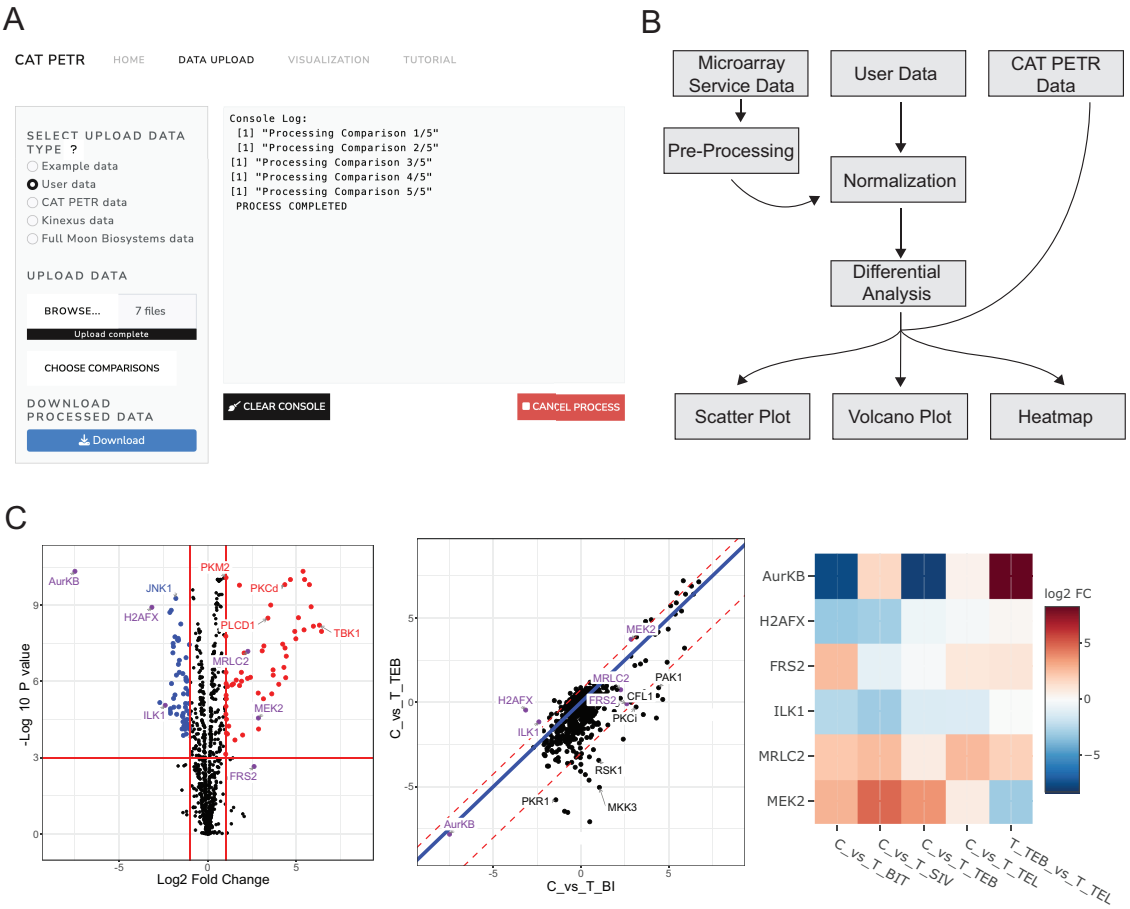

Figure 1: Overview of CAT PETR analysis workflow with interface and visualizations. (**A**). Data upload interface. (**B**) Network diagram showing the analysis pathways for the three acceptable input formats. (**C**) Example of interactive plots provided by CAT PETR including (left to right) volcano plot, scatter plot of fold changes, and heatmap.

Data Upload

The data upload tab (Fig. 1A) allows users to upload, normalize, and perform differential analysis on their data. CAT PETR accepts three different data formats that lead to three analysis pathways (Fig. 1B).

Microarray Service Data: Data provided by antibody microarray assay services such as those offered by Kinexus Bioinformatics and Full Moon Biosystems can be directly uploaded. Both companies' microarrays have recently been used to quantify protein expression and covalent modification, such as in studies of the central nervous system [10], cancer [11], and infectious diseases [12]. Data provided by similar microarray services must be converted to the general user data format, described next.

User Data: The user can input data as tables of gene/protein names and associated expression/phosphorylation metrics, with the option, if appropriate, to specify phosphorylation site information. It is expected that pre-processing steps, such as background correction for microarray data or the calculation of TPM (transcript per million) values for RNA-seq data, have been completed beforehand (see user manual SI file for more details).

CAT PETR Data: Data that have already been analyzed by CAT PETR can be downloaded and re-uploaded to skip the analysis step.

Data Processing

The user can choose to normalize the uploaded data using a logarithmic transformation or variance stabilization normalization [13]. After normalization, differential analysis is completed via traditional pairwise t-tests or Cyber-T t-tests, [14], which have been widely applied to studies using antibody microarrays, DNA microarrays, mass spectrometry, and next generation sequencing [15]. More details on the methods are available in the user manual SI file.

Visualizations

The visualization tab allows for the creation of 3 separate and interactive plot types: volcano plots, scatter plots, and heatmaps (Figure 1C), described below. These plots commonly use a universal gene/protein search bar to decide which genes/proteins will be labeled or displayed; allowing for the same genes/proteins to be visualized across different plots. More details on the interactive plotting tools are available in the user manual SI file.

Volcano Plots: Volcano plots visualize the effect size and significance of each datapoint for each control-treatment comparison. The top hits, as defined by the Euclidean distance to the origin point of the graph, are automatically labeled. and can be added to the search bar.

Scatter Plots: Scatter plots visualize the fold change from two separate control-treatment comparisons along the x and y axis allowing for the comparison of changes in expression/phosphorylation between multiple control and treatment groups. A line is drawn to represent a 1:1 ratio of log fold change between comparisons, with additional lines that can be adjusted to indicate quantile thresholds. Top hits are automatically labelled.

Heatmaps: Heatmaps act as another tool for visualizing multiple control-treatment comparisons at the same time. The heatmap only shows genes/proteins present in the search bar.

Acknowledgments

Funding for this work was provided by CIHR PJT-148646 and PJT-152931. Steven Pelech and his family are the majority shareholders of Kinexus Bioinformatics Corporation.

References

[1] Chen Z, Dodig-Crnković T, Schwenk JM, Tao Sc. Current applications of antibody microarrays. Clinical proteomics. 2018;15(1):1–15.

[2] Wilson JJ, Burgess R, Mao YQ, Luo S, Tang H, Jones VS, et al. Antibody arrays in biomarker discovery. Advances in clinical chemistry. 2015;69:255–324.

[3] Yang Y, Lian S, Meng L, Qu L, Shou C. Antibody array revealed PRL-3 affects protein phosphorylation and cytokine secretion. PLoS One. 2017;12(1):e0169665.

[4] Wang X, Shi Z, Lu HY, Kim JJ, Bu W, Villalobos JA, et al. High-throughput profiling of histone post-translational modifications and chromatin modifying proteins by reverse phase protein array. Journal of Proteomics. 2022;262:104596.

[5] Lambert Yue CS, Arora N, Winkler DF, Pelech S. Antibody microarray and immunoblotting analyses of the EGF signaling phosphorylation network in human A431 epidermoid carcinoma cells. 2017;.

[6] Fasolo J, Im H, Snyder MP. Probing high-density functional protein microarrays to detect protein-protein interactions. JoVE (Journal of Visualized Experiments). 2015;(102):e51872.

[7] Russo F, Angelini C. RNASeqGUI: a GUI for analysing RNA-Seq data. Bioinformatics. 2014;30(17):2514–2516.

[8] Nelson JW, Sklenar J, Barnes AP, Minnier J. The START App: a web-based RNAseq analysis and visualization resource. Bioinformatics. 2017;33(3):447–449.

[9] Choi K, Ratner N. iGEAK: an interactive gene expression analysis kit for seamless workflow using the R/shiny platform. BMC genomics. 2019;20(1):1–7.

[10] Noël G, Tham DKL, Guadagno E, MacVicar B, Moukhles H. The laminin-induced phosphorylation of PKCδ regulates AQP4 distribution and water permeability in rat astrocytes. Cellular and Molecular Neurobiology. 2021;41(8):1743–1757.

[11] Falcomatà C, Bärthel S, Widholz SA, Schneeweis C, Montero JJ, Toska A, et al. Selective multikinase inhibition sensitizes mesenchymal pancreatic cancer to immune checkpoint blockade by remodeling the tumor microenvironment. Nature Cancer. 2022;3(3):318–336.

[12] Peppenelli MA, Miller MJ, Altman AM, Cojohari O, Chan GC. Aberrant regulation of the Akt signaling network by human cytomegalovirus allows for targeting of infected monocytes. Antiviral research. 2018;158:13–24.

[13] Huber W, Von Heydebreck A, Sültmann H, Poustka A, Vingron M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics. 2002;18(suppl_1):S96–S104.

[14] Baldi P, Long AD. A Bayesian framework for the analysis of microarray expression data: regularized t-test and statistical inferences of gene changes. Bioinformatics. 2001;17(6):509–519.

[15] Kayala MA, Baldi P. Cyber-T web server: differential analysis of high-throughput data. Nucleic acids research. 2012;40(W1):W553–W559.