Quickomics: exploring omics data in an intuitive, interactive and informative manner

Benbo Gao[1#], Jing Zhu[1#], Soumya Negi[1], Xinmin Zhang[2], Stefka Gyoneva[1], Fergal Casey[1] and Ru Wei[1*] and Baohong Zhang[1*]

[1]Biogen Inc., Cambridge, Massachusetts, USA

[2]BioInfoRx, Inc., Madison, Wisconsin, USA

[#]Contributed equally to the work

*To whom correspondence should be addressed

Abstract

Summary: We developed Quickomics, a feature-rich R Shiny-powered tool to enable biologists to fully explore complex omics data and perform advanced analysis in an easy-to-use interactive interface. It covers a broad range of secondary and tertiary analytical tasks after primary analysis of omics data is completed. Each functional module is equipped with customized configurations and generates both interactive and publication-ready high-resolution plots to uncover biological insights from data. The modular design makes the tool extensible with ease.

Availability: Researchers can experience the functionalities with their own data or demo RNA-Seq and proteomics data sets by using the app hosted at http://quickomics.bxgenomics.com and following the tutorial, https://bit.ly/3rXIyhL. The source code under GPLv3 license is provided at https://github.com/interactivereport/Quickomics.

Contact: benbo.gao@biogen.com, baohong.zhang@biogen.com

Supplementary information: Supplementary materials are available at https://bit.ly/37HP17g.

1 Introduction

Over the last decade, proteomics and RNA-Seq have become the standard experimental approaches for confidently identifying and accurately quantifying thousands of proteins and genes in complex biological systems. Typically, data generated from those high-throughput experiments are analyzed by data analysts and primary results are provided to end-users as genes by samples matrices, tables of summary statistics derived from differential expression analysis, and often a small set of static figures showing high-level results. The large scale of such datasets demands a more effective delivery method beyond the classic tabulated results handout to enable full utilization of a given dataset. Several tools allowing interactive exploration

and visualization of those complex omics data for end-users without programming skills, such as START (Nelson *et al.*, 2017), PIVOT (Zhu *et al.*, 2018), PaintOmics 3 (Hernández-De-Diego *et al.*, 2018), iSEE (Lun *et al.*, 2018), iDEP (Ge *et al.*, 2018), WIlsON (Schultheis *et al.*, 2019), IRIS-EDA (Monier *et al.*, 2019), DEBrowser (Kucukural *et al.*, 2019), Ideal (Marini *et al.*, 2020), and BEAVR (Perampalam and Dick, 2020), have recently been developed. However, these tools come with limitations regarding input formats, the ability to easily adjust plotting parameters, narrow focus on RNA-Seq data analysis, or lack of comprehensive functionalities covering major secondary and tertiary analytical tasks such as gene set enrichment, co-expression network analysis and comparative pathway analysis (detailed feature comparison is outlined in supplementary table 1).

To address these gaps, we developed Quickomics, an easy-to-use tool for the visualization of both RNA-Seq and proteomics datasets by leveraging newly developed R packages and modern JavaScript plotting libraries to enhance the usability from data quality control to generation of publication-ready figures.

2 Methods

Architecture. Quickomics is built on R Shiny, a web application framework provided as an R package (http://shiny.rstudio.com/). R Shiny allows applications to be deployed on a personal computer using RStudio or hosted on a local or cloud-based Shiny Server. Because it is based on R, Quickomics natively integrates with many R packages, such as ComplexHeatmap (Gu *et al.*, 2016). Thus, all analytical tasks are performed in R in the background while the results are presented in interactive web pages by utilizing web techniques. In addition, Quickomics is based on modular design to ensure extensibility.

Highly configurable visualization is one of major strengths in Quickomics. Often, users have diverse requirements on how to display analyzed data to facilitate interpretation and/or to identify data subsets for further analysis. Here, we take advantage of the dynamic nature of R Shiny where a user can easily query data on-the-fly to only display selected groups, samples or proteins/genes of interest.

Integrated Gene Set Query. We developed xGenesets to provide API for Quickomics to select a subset of genes/proteins when it is needed in the tool. Detailed implementation is described in supplementary tutorial section 7. An example of freely available xGenesets API can be found at https://bxaf.net/genesets/.

Interactive and High-Resolution Plots. R Shiny controlled widgets are used to enable users to configure many graphic parameters like font size, color palette, width and height, etc. Moreover, svglite package was deployed to allow exporting of interactively generated plots in high-resolution PDF and SVG format.

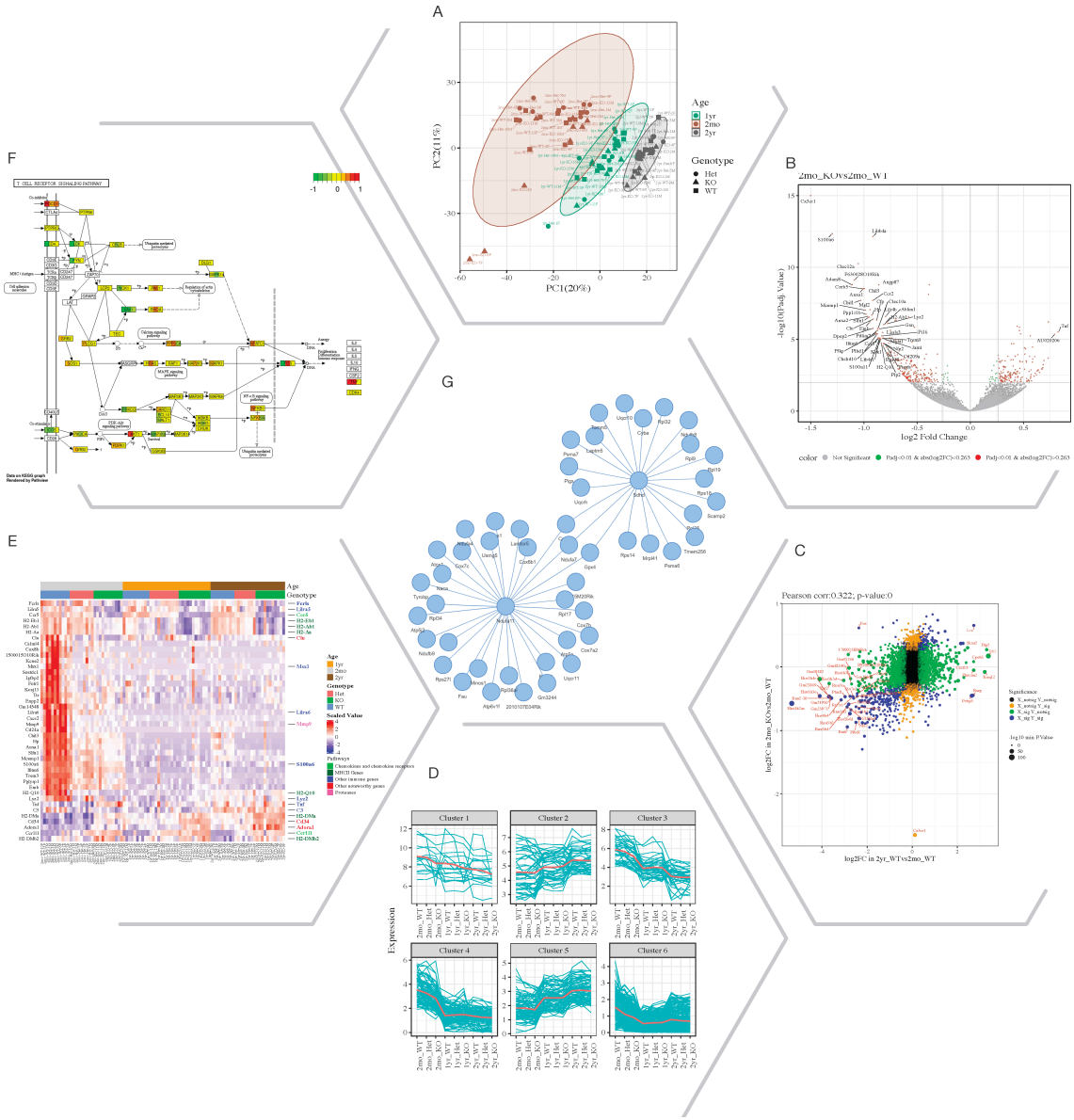

Fig. 1. Selected Quickomics functions applied to a dataset of microglial RNA-seq gene expression from three mouse genotypes over time. A) PCA based on full dataset highlights primary sample separation by mouse age, at which the

cells were isolated. (B) Volcano plot visualizes differentially expressed genes, most of which show reduced expression in 2mo KO compared to 2mo_WT microglia. For spacing purpose, absolute log2FC (Fold Change) and negative log10 adjusted p-value are capped at 1.5 and 15, respectively. (C) Correlation analysis between two comparisons shows that aging and Cx3cr1-KO have a similar effect on gene expression. (D) Pattern clustering identifies subsets of genes with similar expression over the samples. The clustering is mostly driven by age, with the KO genotype having a similar, but smaller effect. (E) Heatmap of all samples allows the identification of gene clusters with expression regulated by age and/or genotype. Key genes and the pathways they belong to are highlighted on the right. (F) After pathway enrichment analysis, KEGG pathways (Kanehisa and Goto, 2000) of interest can be displayed in a cellular context. The color bars with each stripe representing one comparison show log2 fold changes in various comparisons, allowing project-wide insights for patterns of expression. (G) Correlation network shows potential links between genes of interest.

3 Results

Quickomics provides a comprehensive visualization workflow for major secondary and tertiary analytical tasks of high dimensional data, which is composed of nine main modules for result visualization and/or analysis and one output module. The detailed guidance about how to format, upload and explore a dataset is provided in supplementary tutorial. The functionalities of Quickomics are exemplified by applying the tool to a time course study of mouse microglia (Gyoneva *et al.*, 2019) as shown in Figure 1**.**

4 Conclusions

In summary, Quickomics is a powerful tool to help biologists explore complex omics data and interpret results through a user-friendly interface. Firstly, it provides comprehensive quality control analysis with adjustable options for visualization. Secondly, it supports most of the major secondary and tertiary analytical tasks including volcano plot, heatmap, expression plot, gene set enrichment, pattern clustering, correlation network and Venn diagram. Finally, it is released as open source to entice suggestions and contributions from the bioinformatics community to further enhance this versatile tool.

Conflict of Interest: B.G., J.Z., S.N., S.G., F.C, R.W. and B.Z are employees of Biogen and hold stocks from the company. X.Z is an employee of BioInfoRx, Inc.

References

Ge,S.X. *et al.* (2018) iDEP: An integrated web application for differential expression and pathway analysis of RNA-Seq data. *BMC Bioinformatics*, **19**, 1–24.

Gu,Z. *et al.* (2016) Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, **32**, 2847–2849.
Gyoneva,S. *et al.* (2019) Cx3cr1-deficient microglia exhibit a premature aging transcriptome. *Life Sci. alliance*, **2**.
Hernández-De-Diego,R. *et al.* (2018) PaintOmics 3: A web resource for the pathway analysis and visualization of multi-omics data. *Nucleic Acids Res.*, **46**, W503–W509.
Kanehisa,M. and Goto,S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.*, **28**, 27–30.
Kucukural,A. *et al.* (2019) DEBrowser: interactive differential expression analysis and visualization tool for count data. *BMC Genomics*, **20**, 6.
Lun,A.T.L. *et al.* (2018) iSEE: Interactive SummarizedExperiment Explorer [version 1; referees: 2 approved]. *F1000Research*, **7**.
Marini,F. *et al.* (2020) ideal: an R/Bioconductor package for Interactive Differential Expression Analysis. *bioRxiv*, 1–17.
Monier,B. *et al.* (2019) IRIS-EDA: An integrated RNA-seq interpretation system for gene expression data analysis. *PLoS Comput. Biol.*, **15**, 1–15.
Nelson,J.W. *et al.* (2017) The START App: a web-based RNAseq analysis and visualization resource. *Bioinformatics*, **33**, 447–449.
Perampalam,P. and Dick,F.A. (2020) BEAVR: a browser-based tool for the exploration and visualization of RNA-seq data. *BMC Bioinformatics*, **21**, 221.
Schultheis,H. *et al.* (2019) WIlsON: Web-based Interactive Omics VisualizatioN. *Bioinformatics*, **35**, 1055–1057.
Zhu,Q. *et al.* (2018) PIVOT: Platform for interactive analysis and visualization of transcriptomics data. *BMC Bioinformatics*, **19**, 4–11.