Building systems for interactive data exploration in systems biology

Bjørn Fjukstad[1], Vanessa Dumeaux[2], Karina Standahl Olsen[3], Eiliv Lund[3] and Lars Ailo Bongo[1*]

Abstract

Background: In scientific fields such as systems biology there is a need for data exploration tools that can enable discovery of new insights. Such tools need to integrate advanced statistical analyses with known biology from up-to-date databases to leverage the wealth of existing knowledge in data exploration. However there are few systems that enable the development of new tools that combine these.

Results: We have designed an approach for developing data exploration applications in systems biology, and demonstrated its viability through a web application for exploring and comparing transcriptional profiles from blood and tumor samples. Our approach makes it possible to visualize data from advanced statistical software packages in any modern programming language. We show that it is possible to leverage advanced statistical software packages together with up-to-date online databases to create applications that integrates data with known biology.

Conclusions: Our approach and reference implementation Kvik, enables easy development of data exploration tools that provide reproducible analyses using efficient processing. Kvik is open-sourced at github.com/fjukstad/kvik and the web application for exploring transcriptional profiles, MIxT, is availible at github.com/fjukstad/mixt.

Keywords:

Background

Analysis frameworks: Bioconductor in R, Python. Visualization: Cytoscpape, D3, BioJS. Databases: MSigDB, KEGG, PubMed. Parallel and distributed: Hadoop, Spark, Kubernetes, Azure.

Data wrangling and analysis done in R or other languages. Visualization and presentation of final datasets by external tools. Manual database lookup is tedious and gets out of date.

Related work: OpenCPU, biogo, renjin,

The big issue: There is no link between analysis and final presentation

Methods

Collection of packages in the Go programming language for building data exploration applications. Interfaces with popular online databases such as MSigDB and KEGG. Provides an interface to the R programming language. Typically used to build web apps, but commandline tools are also possible!

Results and Discussion

MIxT. Interactive web application for exploring WGCNA modules from blood and biopsy. Integrate results with known genesets and online databases. Docker containers for reproducibility and portability.

Data analysis: R package + Kvik = R in the browser. Visualization: D3, SigmaJS, R Plots

How can we improve Kvik and its features: R user input sanitation, security,

Conclusions

Sum it all up.

List of abbreviations used
Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Computer Science, UiT – The Arctic University of Tromsø, 9037 Tromsø, NO. [2]University of ...., , . [3]Department of Community Medicine, UiT – The Arctic University of Tromsø, 9037 Tromsø, NO.

References

*Correspondence: larsab@cs.uit.no
[1]Department of Computer Science, UiT – The Arctic University of Tromsø, 9037 Tromsø, NO
Full list of author information is available at the end of the article