geneXplainR: An R interface for the geneXplain platform

Philip Stegmaier[1], Alexander Kel[1], and Edgar Wingender[1,2]

[1]geneXplain GmbH, Am Exer 10b, D-38304 Wolfenbuettel, Germany
[2]Institute of Bioinformatics, University Medical Center Göttingen, D-37077 Göttingen, Germany

12 July 2017

Paper DOI: http://dx.doi.org/10.21105/joss.00412
Software Repository: https://github.com/genexplain/geneXplainR
Software Archive: http://dx.doi.org/10.5281/zenodo.1013318

Summary

The geneXplain platform (A. Kel et al. 2011) is an online toolbox and workflow management system for a broad range of bioinformatic and systems biology applications. The platform is well-known for its upstream analysis (Koschmann et al. 2015), that has been developed to identify causal signalling molecules on the basis of experimental data like expression measurements. Methods integrated into the system include

- molecular network analysis such as pathway enrichment, identification of network clusters, common signaling regulators or effectors,
- analysis of transcription factor binding sites like prediction of binding sites using positional weight matrices, testing for enrichment of binding sites in regulatory sequences, or identification of composite modules (combinations of binding sites), as well as motif finding
- methods to test for enrichment of functional groups or categories, e.g. from the Gene Ontology (Ashburner et al., n.d.), using the Fisher test or Gene Set Enrichment Analysis (GSEA) (Subramanian et al. 2005)
- Flux Balance Analysis (Duarte et al. 2007) to analyze metabolic networks
- methods for processing and statistical analysis of high-throughput data, e.g. Limma (Ritchie et al. 2015) or DESeq2 (Anders and Huber 2010)
- as well as simulation of computational models, e.g. as collected in the BioModels database (Le Novère et al. 2006).

An important feature of the platform is the possibility to define and execute workflows that can implement sequential and parallel multi-step analysis processes. Workflows can be created and edited using a graphical editor. They are an effective tool to define complex analysis pipelines and to document, reuse and to reproduce analysis procedures. Figure 1 shows the graphical user interface of the platform with an example workflow for Flux Balance Analysis.

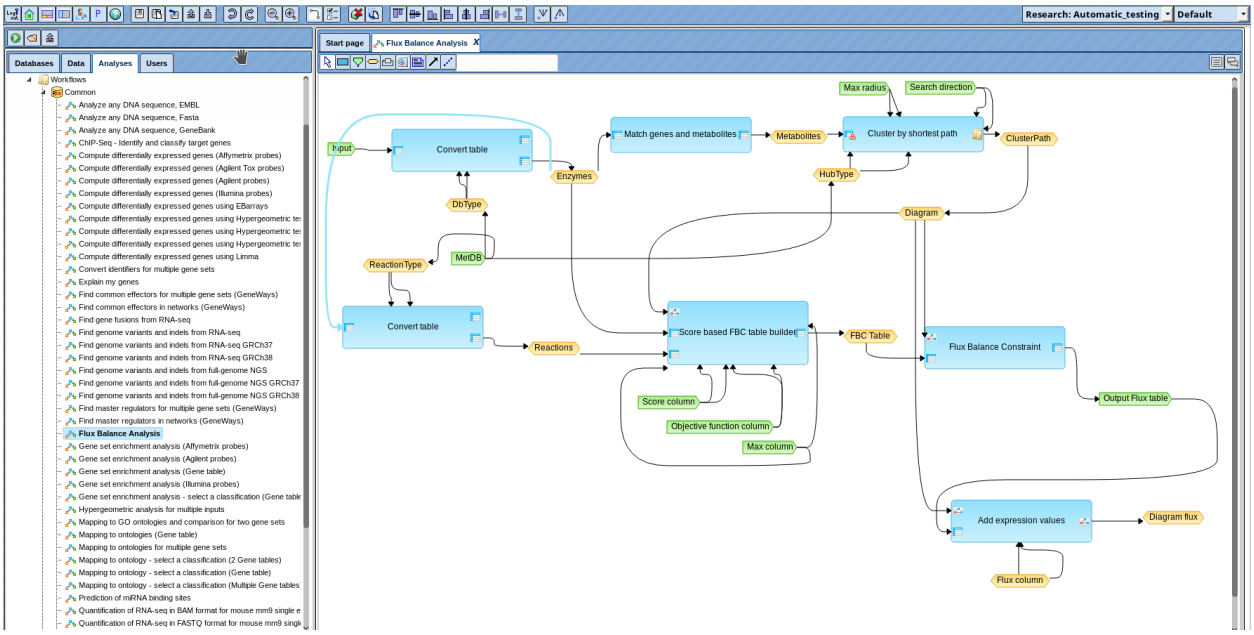

Figure 1. Graphical user interface of the geneXplain platform showing the Flux Balance Analysis workflow.

We have developed geneXplainR, an R (R Core Team 2016) interface for the geneXplain platform, that makes it possible to define analysis pipelines in the R language using tools, workflows and other resources integrated in the platform. The package is based on and extends the rbiouml package (Yevshin and Valeev 2013). The geneXplainR adds basic functionality not covered by rbiouml such as creation of projects, folders or deletion of items from the workspace as well as functions that provide direct access to certain tools or workflows. Another purpose of geneXplainR is to offer a suite of examples scripts in the *example* branch that help users to get started with the software. We have also developed a similar project denoted as the *genexplain-api* (P. Stegmaier 2017) that addresses Java developers and shall be described elsewhere. With geneXplainR, developers can easily take advantage of other bioinformatics software and resources available in R, e.g. through the popular Bioconductor project (Gentleman et al. 2004).

Acknowledgements

The development of geneXplainR has been supported by the MyPathSem, a collaborative project funded by the German Federal Ministry of Education and Research (BMBF) in the funding program "i:DSem – Integrative Datensemantik in der Systemmedizin", as well as by MIMOmics, a collaborative project funded by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 305280, research area FP7-HEALTH-2012-INNOVATION-1, topic HEALTH.2012.2.1.1-3: Statistical methods for collection and analysis of –omics data.

Support

Issue reports and support requests are welcome either by writing an e-mail to info@genexplain.com or through the GitHub issue system (https://github.com/genexplain/geneXplainR/issues).

References

Anders, S., and W. Huber. 2010. "Differential Expression Analysis for Sequence Count Data." *Genome Biology* 11 (10): R106. doi:10.1186/gb-2010-11-10-r106.

Ashburner, M., C.A. Ball, J.A. Blake, D. Botstein, and H. Butler. n.d. "Gene Ontology: Tool for the Unification of Biology. the Gene Ontology Consortium." *Nature Genetics* 25 (1): 25–29. doi:10.1038/75556.

Duarte, N., S.A. Becker, N. Jamshidi, I. Thiele, M.L. Mo, and others. 2007. "Global Reconstruction of the Human Metabolic Network Based on Genomic and Bibliomic Data." *Proc Natl Acad Sci U S A* 104 (6): 1777–82. doi:10.1073/pnas.0610772104.

Gentleman, R.C., V.J. Carey, D.M. Bates, Bolstadm B., M. Dettling, and others. 2004. "Bioconductor: Open Software Development for Computational Biology and Bioinformatics." *Genome Biology* 5 (10): R80. doi:10.1186/gb-2004-5-10-r80.

Kel, A., F. Kolpakov, V. Poroikov, and G. Selivanova. 2011. "GeneXplain — Identification of Causal Biomarkers and Drug Targets in Personalized Cancer Pathways." *J. Biomol. Tech.*, no. 22: S1.

Koschmann, J., A. Bhar, P. Stegmaier, A. E. Kel, and E. Wingender. 2015. "'Upstream Analysis': An Integrated Promoter-Pathway Analysis Approach to Causal Interpretation of Microarray Data." *Microarrays*, no. 2: 270–86. doi:10.3390/microarrays4020270.

Le Novère, N., B. Bornstein, A. Broicher, M. Courtot, M. Donizelli, and others. 2006. "BioModels Database: A Free, Centralized Database of Curated, Published, Quantitative Kinetic Models of Biochemical and Cellular Systems." *Nucleic Acids Research* 34 (Database issue): D689–691. doi:10.1093/nar/gkj092.

R Core Team. 2016. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Ritchie, M.E., B. Phipson, D. Wu, Y. Hu, C.W. Law, and others. 2015. "Limma Powers Differential Expression Analyses for Rna-Sequencing and Microarray Studies." *Nucleic Acids Research* 43 (7): e47. doi:10.1093/nar/gkv007.

Stegmaier, Philip. 2017. "Genexplain-Api - the geneXplain Platform Java Api." *GitHub Repository*. https://github.com/genexplain/genexplain-api; GitHub.

Subramanian, A., P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, and others. 2005. "Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles." *Proc Natl Acad Sci U S A* 102 (43): 15545–50. doi:10.1073/pnas.0506580102.

Yevshin, Ivan, and Tagir Valeev. 2013. *Rbiouml: Interact with Biouml Server*.