GeneNetwork: framework for web-based genetics

Zachary Sloan[1], Danny Arends[2], Karl W. Broman[3], Arthur Centeno[4], Nicholas Furlotte[5], Harm Nijveen[5], Lei Yan[6], Xiang Zhou[7], Robert W. Williams[8], and Pjotr Prins[9]

[1]University of Tennessee Health Science Center, USA
[2]Humboldt University, Berlin, Germany
[3]University of Wisconsin, USA
[4]University of Tennessee Health Science Center, USA
[5]Wageningen University, The Netherlands
[6]University of Tennessee Health Science Center, USA
[7]University of Michigan
[8]University of Tennessee Health Science Center, USA
[9]University Medical Center Utrecht, The Netherlands, University of Tennessee Health Science Center, USA
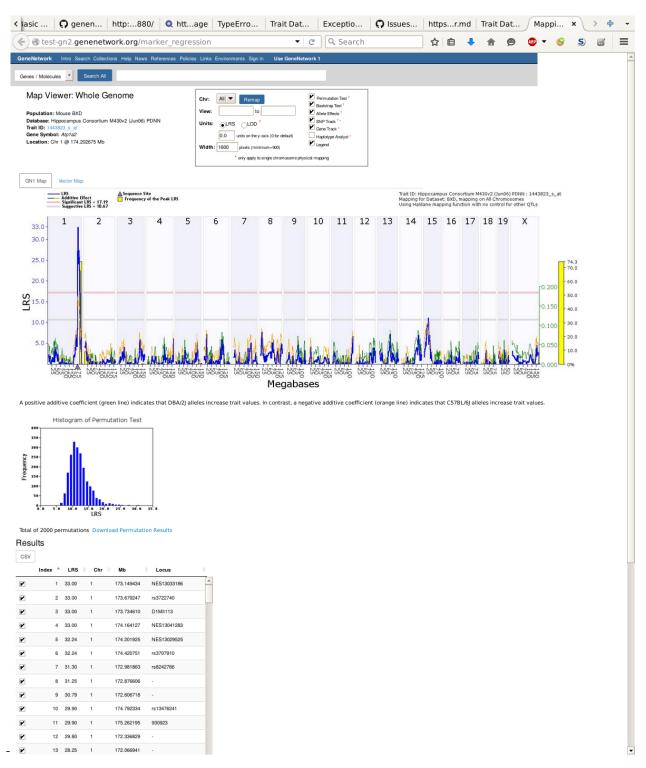
29 May 2016

Summary

GeneNetwork (GN) is a free and open source (FOSS) framework for web-based genetics that can be deployed anywhere. GN allows biologists to upload high-throughput experimental data, such as expression data from microarrays and RNA-seq, and also 'classic' phenotypes, such as disease phenotypes. These phenotypes can be mapped interactively against genotypes using embedded tools, such as R/QTL (Arends et al. 2010) mapping, interval mapping for model organisms and pylmm; an implementation of FaST-LMM (Lippert et al. 2011) which is more suitable for human populations and outbred crosses, such as the mouse diversity outcross. Interactive D3 graphics are included from R/qtlcharts and presentation-ready figures can be generated. Recently we have added functionality for phenotype correlation (Wang et al. 2016) and network analysis (Langfelder and Horvath 2008).

GN is written in python and javascript and contains a rich set of tools and libraries that can be written in any computer language. A full list of included software can be found in the package named 'genenetwork2' and defined in guix-bioinformatics. To make it easy to install GN locally in a byte reproducible way, including all dependencies and a 2GB MySQL test database (the full database is 160GB and growing), GN is packaged with GNU Guix, as described here. GNU Guix deployment makes it feasible to deploy and rebrand GN anywhere.

Future work

More mapping tools will be added, including support for Genome-wide Efficient Mixed Model Association (GEMMA). The Biodiallance genome browser is being added as a Google Summer of Code project with special tracks related to QTL mapping and network analysis. Faster LMM solutions are being worked on, including GPU support.

A REST interface is being added so that data can be uploaded to a server, analysis run remotely on high performance hardware, and results downloaded and used for further analysis. This feature will allow biologist-programmers to use R and Python on their computer and execute computations on GN enabled servers.

References

Arends, D., P. Prins, R. C. Jansen, and K. W. Broman. 2010. "R/qtl: high-throughput multiple QTL mapping." *Bioinformatics* 26 (23): 2990–2. doi:10.1093/bioinformatics/btq565.

Langfelder, P., and S. Horvath. 2008. "WGCNA: an R package for weighted correlation network analysis." *BMC Bioinformatics* 9: 559. doi:10.1186/1471-2105-9-559.

Lippert, C., J. Listgarten, Y. Liu, C. M. Kadie, R. I. Davidson, and D. Heckerman. 2011. "FaST linear mixed models for genome-wide association studies." *Nat Methods* 8 (10): 833–35. doi:10.1038/nmeth.1681.

Wang, X., A. K. Pandey, M. K. Mulligan, E. G. Williams, K. Mozhui, Z. Li, V. Jovaisaite, et al. 2016. "Joint mouse-human phenome-wide association to test gene function and disease risk." *Nat Commun* 7: 10464. doi:10.1038/ncomms10464.