

GeneNetwork: framework for web-based genetics

Zachary Sloan¹, Danny Arends², Karl W. Broman³, Arthur Centeno⁴, Nicholas Furlotte⁵, Harm Nijveen⁵, Lei Yan⁶, Xiang Zhou⁷, Robert W. Williams⁸, and Pjotr Prins⁹

¹University of Tennessee Health Science Center, USA

²Humboldt University, Berlin, Germany

³University of Wisconsin, USA

⁴University of Tennessee Health Science Center, USA

⁵Wageningen University, The Netherlands

⁶University of Tennessee Health Science Center, USA

⁷University of Michigan

⁸University of Tennessee Health Science Center, USA

⁹University Medical Center Utrecht, The Netherlands, University of Tennessee Health Science Center, USA

29 May 2016

Paper DOI: <http://dx.doi.org/10.21105/joss.00025>

Software Repository: <https://github.com/genenetwork/genenetwork2>

Software Archive: <http://dx.doi.org/10.5281/zenodo.53740>

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/qtcharts 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).

test-gn2.genenetwork.org/marker_regression

GeneNetwork Intro Search Collections Help News References Policies Links Environments Sign in Use GeneNetwork 1

Genes / Molecules Search All

Map Viewer: Whole Genome

Population: Mouse BXD
 Database: Hippocampus Consortium M430v2 (Jun06) PDNN
 Trait ID: 1443823_s_at
 Gene Symbol: *Alpt2a2*
 Location: Chr 1 @ 174.202675 Mb

Chr: All Remap
 View: to
 Units: LRS LOD
 Width: 1600 pixels (minimum=900)
 * only apply to single chromosome physical mapping

Permutation Test
 Bootstrap Test
 Allele Effects
 SNP Track
 Gene Track
 Haplotype Analyst
 Legend

GN1 Map Vector Map

LRS Additive Effect Significant LRS = 17.19 Suggestive LRS = 10.67 Sequence Site Frequency of the Peak LRS

Trait ID: Hippocampus Consortium M430v2 (Jun06) PDNN : 1443823_s_at
 Mapping for Dataset: BXD, mapping on All Chromosomes
 Using Haldane mapping function with no control for other QTLs

A positive additive coefficient (green line) indicates that *DBA/2j* alleles increase trait values. In contrast, a negative additive coefficient (orange line) indicates that *C57BL/6j* alleles increase trait values.

Histogram of Permutation Test

Total of 2000 permutations [Download Permutation Results](#)

Results

CSV

Index	LRS	Chr	Mb	Locus
<input checked="" type="checkbox"/>	1	33.00	1	173.149434 NES13033186
<input checked="" type="checkbox"/>	2	33.00	1	173.679247 rs37222740
<input checked="" type="checkbox"/>	3	33.00	1	173.734610 D1Mit113
<input checked="" type="checkbox"/>	4	33.00	1	174.164127 NES13041283
<input checked="" type="checkbox"/>	5	32.24	1	174.201925 NES13029525
<input checked="" type="checkbox"/>	6	32.24	1	174.420751 rs3707910
<input checked="" type="checkbox"/>	7	31.30	1	172.981863 rs8242766
<input checked="" type="checkbox"/>	8	31.25	1	172.876606 -
<input checked="" type="checkbox"/>	9	30.79	1	172.606718 -
<input checked="" type="checkbox"/>	10	29.90	1	174.792334 rs13476241
<input checked="" type="checkbox"/>	11	29.90	1	175.262195 930923
<input checked="" type="checkbox"/>	12	29.80	1	172.336829 -
<input checked="" type="checkbox"/>	13	28.25	1	172.066941 -

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/qlt: 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.