Simulating Systems Genetics with SysGenSIM

Andrea Pinna[1], Nicola Soranzo[1], and Ina Hoeschele[2,3] and Alberto de la Fuente[1]

[1]CRS4 Bioinformatica, Parco Scientifico e Tecnologico, POLARIS, Edificio 1, Loc. Piscina Manna, 09010, Pula (CA), Italy
[2]Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061
[3]Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0477

SysGenSIM [1] is a software package to simulate Systems Genetics experiments in model organisms, for the purpose of evaluating and comparing statistical and computational methods and their implementations for analyses of Systems Genetics data (e.g. methods for Expression Quantitative Trait Loci (eQTL) mapping and network inference). Users are allowed to select a variety of network topologies, genetic and kinetic parameters to simulate Systems Genetics data (genotyping, gene expression and phenotyping) with genome-wide networks. The software has been recently updated by adding to the above features the newly implemented capability of reproducing experimental perturbations, i.e. single-gene knock-out, knock-down, and over-expression experiments.

Simulated data can be of great use for the evaluation of algorithms for the inference of gene networks, as demonstrated in DREAM4 [2,3] and DREAM5 [4,5] challenges, and recently in a meeting of the STATSEQ network, where our *in-silico* data turned out as valuable benchmarks for testing the developed inference algorithms.

SysGenSIM, released under the GNU General Public License, can be freely downloaded from [6]. Data produced by SysGenSIM are published in [7].

References.
[1] A. Pinna, N. Soranzo, I. Hoeschele, and A. de la Fuente. Simulating systems genetics data with SysGenSIM. Bioinformatics, 27(17):2459–2462, 2011.
[2] A. Pinna, N. Soranzo, and A. de la Fuente. From knockouts to networks: establishing direct cause-effect relationships through graph analysis. PloS ONE, 5(10):e12912, 2010.
[3] A. Pinna, S. Heise, R.J. Flassig, A. de la Fuente, and S. Klamt. Reconstruction of large-scale regulatory networks based on perturbation graphs and transitive reduction: Improved methods and their performance and applications. In preparation.
[4] N. Soranzo, A. Pinna, V. De Leo, and A. de la Fuente. Elucidating transcriptional regulatory networks from heterogeneous gene expression compendia. In preparation.
[5] D. Marbach, J.C. Costello, R. Kueffner, N.M. Vega, R.J. Prill, D.M. Camacho, K.R. Allison, M. Kellis, J.J. Collins, A. Pinna, N. Soranzo, V De Leo, A. de la Fuente, G. Stolovitzky, et al. Wisdom of crowds for robust gene network inference. Nature Methods, (9):796–804, 2012.
[6] SysGenSIM website. [http://sysgensim.sourceforge.net].
[7] SysGenSIM benchmarks. [http://sysgensim.sourceforge.net/datasets.html].