Correlation analysis of gene expression time series at multiple scales: from the entire genome to metabolic and signalling pathways

Remondini D (1,2,3), Neretti N (1,4), Milanesi L (5), Tatar M (6), Sedivy JM (7), Franceschi C (1,8), Castellani GC (1,2,3,4)

(1) Centro Interdipartimentale "L. Galvani", Università di Bologna IT
(2) DIMORFIPA, Università di Bologna IT
(3) INFN sezione di Bologna, IT
(4) Brown University, Providence RI USA
(5) Istituto di tecnologie biomediche (ITB) del CNR, Milano IT
(6) Dept. of Ecology and Evolutionary Biology, Brown University, Providence RI USA
(7) Dept. of Molecular and Cell Biology and Biochemistry, Brown University, Providence RI USA
(8) Dip. di Patologia Sperimentale, Università di Bologna IT

Motivation

High-throughput genomic data (microarray) can be very informative on cell state, but an emerging challenge is to retrieve useful informations about gene-gene interaction network from gene expression dynamics, obtained by array sampling collected over time. We propose a method based on the similarity between gene expression dynamics following a cell perturbation. Three examples are considerd: the characterization of the regulatory cascade of c-myc proto-oncogene following Tamoxifen stimulation in engineered rat fibroblas cells; the same characterization of genomic response in Drosophila after nutrition changes; patterns of gene activity as a consequence of ageing occurring over a life-span time series (25y-90y) sampled from T-cells of human donors.

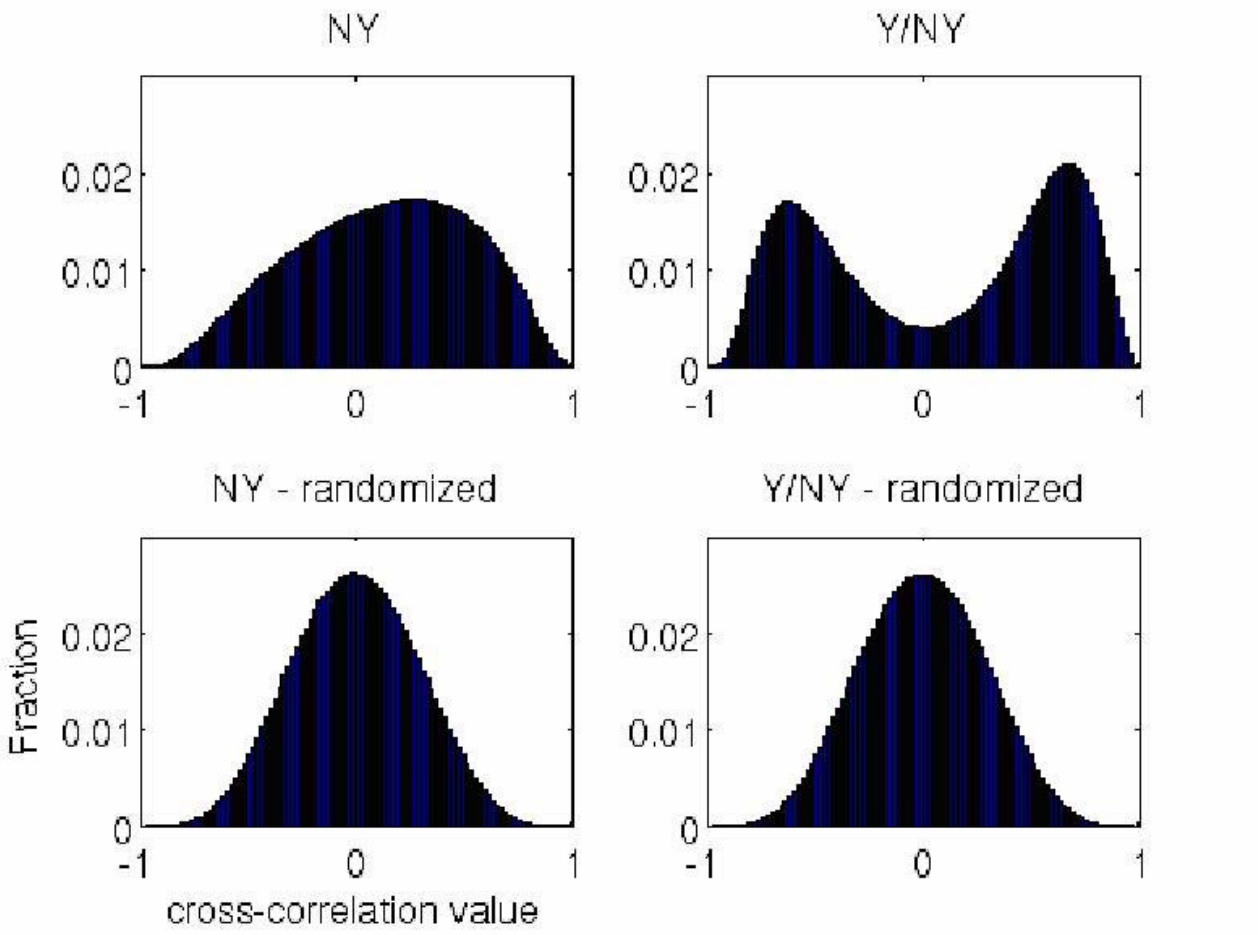
Methods

The key for extracting useful gene network features both at a global level (quantifying the cell response to perturbation) and at a single gene level (gene targeting) is the correlation between expression time series. Thresholding methods are applied for noise removal and for network reconstruction, together with a (optional) processing that allows a preliminar selection of the genes possibly responding to perturbation.

Results

The method is applied to the different datasets: The correlation-based model can establish a clear relationship between network structure and the cascade of activated genes, reflected on multiple scales, from the whole genome down to pathways. The method results very sensitive to the temporal structure of the data, since data shuffling destroys the observed relations.

Contact email: gastone.castellani@unibo.it


NY
Y/NY
NY - randomized
Y/NY - randomized
Fraction
0.02
0.01
0
-1
0
1
cross-correlation value