BIOINFORMATICS APPLICATIONS NOTE Vol. 25 no. 11 2009, pages 1468–1469
doi:10.1093/bioinformatics/btp246

Systems biology

ViaComplex: software for landscape analysis of gene expression networks in genomic context

Mauro A. A. Castro[1,*], José L. Rybarczyk Filho[2], Rodrigo J. S. Dalmolin[1], Marialva Sinigaglia[3], José C. F. Moreira[1], José C. M. Mombach[4] and Rita M. C. de Almeida[2]

[1]Unidade de Bioinformática, Departamento de Bioquímica, [2]Instituto de Física, [3]Núcleo de Bioinformática, Departamento de Genética, Universidade Federal do Rio Grande do Sul, Rua Ramiro Barcelos 2600-anexo, Porto Alegre 90035-003 and [4]Departamento de Física, Universidade Federal de Santa Maria, Santa Maria 97105-900, Brazil

Received on January 13, 2009; revised on March 22, 2009; accepted on April 6, 2009

Advance Access publication April 15, 2009

Associate editor: Thomas Lengauer

ABSTRACT

ViaComplex is an open-source application that builds landscape maps of gene expression networks. The motivation for this software comes from two previous publications (*Nucleic Acids Res*., **35**, 1859–1867, 2007; *Nucleic Acids Res.*, **36**, 6269–6283, 2008). The first article presents a network-based model of genome stability pathways where we defined a set of genes that characterizes each genetic system. In the second article we analyzed this model by projecting functional information from several experiments onto the gene network topology. In order to systematize the methods developed in these articles, ViaComplex provides tools that may help potential users to assess different high-throughput experiments in the context of six core genome maintenance mechanisms. This model illustrates how different gene networks can be analyzed by the same algorithm.

Availability: http://lief.if.ufrgs.br/pub/biosoftwares/viacomplex

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Contact: mauro@ufrgs.br or rita@if.ufrgs.br

1 INTRODUCTION

Genome maintenance mechanisms (GMM) are critical for cell homeostasis. Evolution has shaped sophisticated repair systems that cover most of the insults that can cause genome damages. Defects in any of these systems can predispose to cancer (Castro *et al.*, 2008). At least four DNA damage repair pathways operate in mammals that, together with apoptosis and chromosome stability pathways, comprise the basis of GMM (Hoeijmakers, 2001; Zhivotovsky and Kroemer, 2004).

We have previously constructed a network-based model of human GMM in which different gene activity data were projected onto the interaction map (Castro *et al.*, 2007, 2008). Here, we extend these previous studies and develop a new software which could serve as a generalized tool to evaluate gene expression networks. With a graphical user interface, ViaComplex can either compute gene activity data for the internal model or import customized models of gene/protein interaction networks. In this case, the GMM network model illustrates the type of problem that can be dealt with the software for different gene networks. It can be used to produce publication quality images where data are visualized as functional landscapes projected onto gene network maps. ViaComplex also provides a statistical module based on the concept of information theory where multiple hypotheses are controlled by the false discovery rate (FDR) approach.

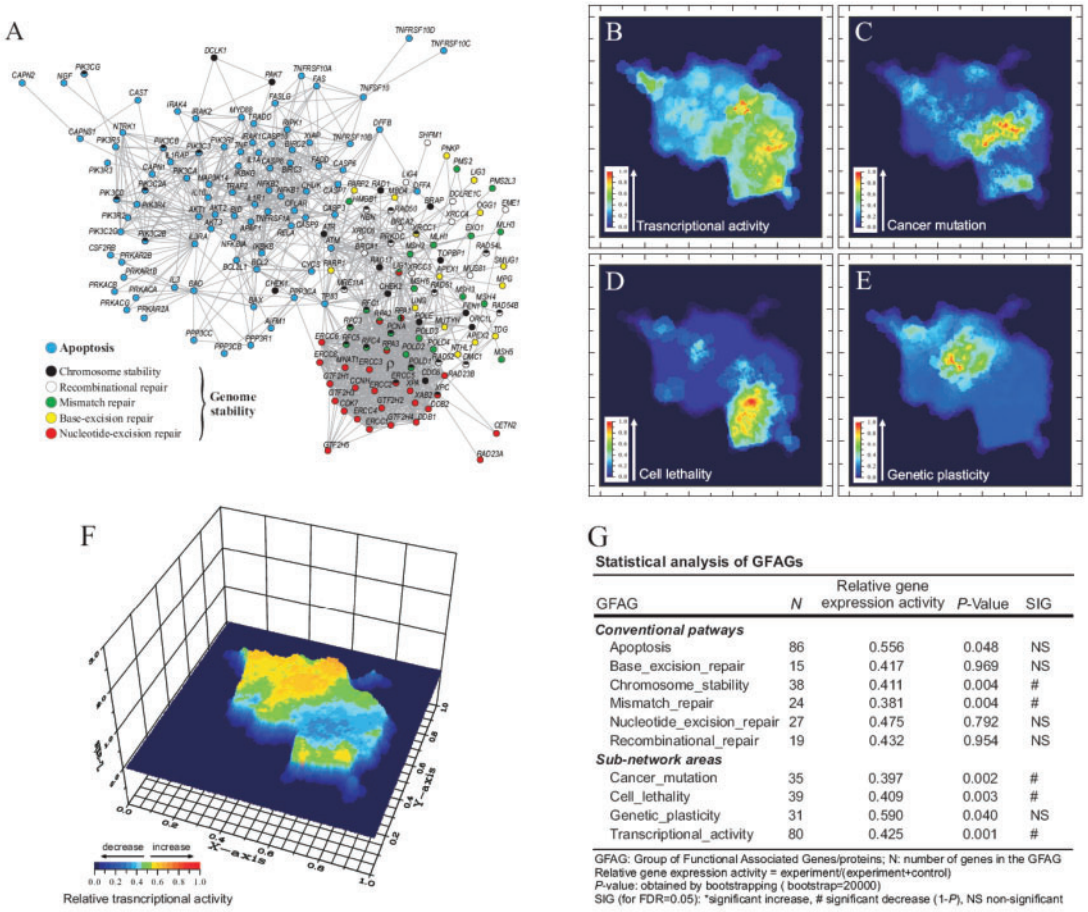
2 IMPLEMENTATION

ViaComplex program code is written for Linux and Windows Intel FORTRAN compilers (version 10.1.025) and is linked with Dislin 9.4, a scientific plotting library (http://www.dislin.de/). The main advantage of this program is that it is able to distribute a given quantity (quantitative or qualitative data) onto gene/protein interaction networks. To do this, ViaComplex overlaps functional information with interaction information (e.g. the network-based model of GMM).

The GMM network model comprises 180 genes that participate in human apoptosis and genome-stability functions as previously described (Castro *et al.*, 2007) and is depicted in Figure 1A. As an example of ViaComplex capabilities, in Figure 1B we show a microarray data analysis processed by the landscape module where gene expression activity is plotted over the network topology. In this figure the software distributed the microarray signal according to the coordinates of the network objects (i.e. nodes and links). By default, ViaComplex will distribute the signal on both nodes and links, but the user can change this option together with other ones available in the console. Also, the same algorithm can map qualitative data, as exemplified with cancer mutations (Fig. 1C), cell lethality (Fig. 1D) and genetic plasticity (Fig. 1E). Alternatively, user can compare two different functional states of the same gene network topology (Fig. 1F, methylated versus nonmethylated states).

The install package includes a comprehensive help file that provides the user all necessary details to prepare the data input and to execute the data analysis. ViaComplex can read the common

*To whom correspondence should be addressed.



G

Statistical analysis of GFAGs

GFAG	*N*	Relative gene expression activity	*P*-Value	SIG
Conventional patways				
Apoptosis	86	0.556	0.048	NS
Base_excision_repair	15	0.417	0.969	NS
Chromosome_stability	38	0.411	0.004	#
Mismatch_repair	24	0.381	0.004	#
Nucleotide_excision_repair	27	0.475	0.792	NS
Recombinational_repair	19	0.432	0.954	NS
Sub-network areas				
Cancer_mutation	35	0.397	0.002	#
Cell_lethality	39	0.409	0.003	#
Genetic_plasticity	31	0.590	0.040	NS
Transcriptional_activity	80	0.425	0.001	#

GFAG: Group of Functional Associated Genes/proteins; N: number of genes in the GFAG
Relative gene expression activity = experiment/(experiment+control)
P-value: obtained by bootstrapping (bootstrap=20000)
SIG (for FDR=0.05): *significant increase, # significant decrease (1-*P*), NS non-significant

Fig. 1. Landscape analysis of GMM. (**A**) Graph of interactions among genes involved GMM, as previously described in Castro *et al.* (2007). (**B**) Example of gene expression data analysis using breast MCF7 cell line (GEO accession no. GSM155194). Color gradient represents the transcriptional activity mapped onto graph. The same algorithm can be used to map different data, e.g. (**C**) genes causally implicated in human cancer, (**D**) yeast lethality data and (**E**) genetic plasticity, as defined in Castro *et al.* (2008). (**F**) Two-state landscape analysis. It compares the transcription profile of MCF7 cells in hypomethylated state (e.g. *state a*) versus hipermethylated state (e.g. *state b*) (GEO accessions GSE5816 versus GSM155194), where the colour gradient $Z = a/(a+b)$. A summary of the statistical analysis of this data is presented in (**G**). Figure 1A reprinted/adapted with permission from *NAR* (Oxford University Press).

gene/protein identifiers (e.g. EMBL, ENTREZ, UniProt, HGNC, RefSeq and UniGene) and the resulting graphs can be previewed as XWIN format or saved as PDF, EPS or PostScript files.

The install package also comprises an extensive library of published studies that exemplifies all procedures by a simple mouse click. In this sense, the GMM network model can be used to observe the functionality of the algorithm, which can analyze different gene networks. Supplementary Figure 1 illustrates this possibility for a large network (with 1892 genes). Such option is available at the 'custom model' module. It is semantically focused in genes, matching gene IDs and node IDs. If there is more than one microarray probe interrogating the expression of a given gene, then the software takes the average of the expression values, which allows the use of different microarray platforms (i.e. it does not involve comparisons between network nodes and probe tag IDs). Other numerical samples of different sizes are available at ViaComplex homepage.

Additional features of the software include a statistical module where two microarray datasets can be compared following the method described in Castro *et al.* (2007), as exemplified in Figure 1G for the data used to build the Figure 1F. We anticipate that ViaComplex will be useful to mine graph patterns from high-throughput experimental data.

Funding: Brazilian Agencies FAPERGS, CAPES and CNPq.

Conflict of Interest: none declared.

REFERENCES

Castro,M.A.A. *et al.* (2007) Impaired expression of NER gene network in sporadic solid tumors. *Nucleic Acids Res.*, **35**, 1859–1867.

Castro,M.A.A. *et al.* (2008) Evolutionary origins of human apoptosis and genome-stability gene networks. *Nucleic Acids Res.*, **36**, 6269–6283.

Hoeijmakers,J.H.J. (2001) Genome maintenance mechanisms for preventing cancer. *Nature*, **411**, 366–374.

Zhivotovsky,B. and Kroemer,G. (2004) Apoptosis and genomic instability. *Nat. Rev. Mol. Cell Biol.*, **5**, 752–762.