Systems biology

ViTraM: visualization of transcriptional modules

Hong Sun[1], Karen Lemmens[2], Tim Van den Bulcke[1], Kristof Engelen[2], Bart De Moor[1] and Kathleen Marchal[2,*]

[1]Department of Electrical Engineering, Katholieke Universiteit Leuven, Kasteelpark Arenberg 10 and
[2]Department of Microbial and Molecular Systems, Katholieke Universiteit Leuven, Kasteelpark Arenberg 20, 3001 Heverlee, Belgium

ABSTRACT

Motivation: We developed ViTraM, a tool that allows visualizing overlapping transcriptional modules in an intuitive way. By visualizing not only the genes and the experiments in which the genes are co-expressed, but also additional properties of the modules such as the regulators and regulatory motifs that are responsible for the observed co-expression, ViTraM can assist in the biological analysis and interpretation of the output of module detection tools.

Availability: The ViTraM software is platform-independent. The software and supplementary material are available at: http://homes.esat.kuleuven.be/~kmarchal/ViTraM/Index.html

Contact: kathleen.marchal@biw.kuleuven.be

1 INTRODUCTION

Previous studies have unveiled the modular organization of the transcriptional network (Hartwell *et al.*, 1999). Module detection tools are developed to reveal this modularity by identifying biclusters (or modules), i.e. groups of genes that show a similar expression profile in a subset of experiments (Madeira and Oliveira, 2004). Some transcriptional module detection tools go one step beyond: not only do they search for the modules, but also the regulatory program responsible for the observed co-expression behavior of the genes in the module is identified (Bar-Joseph *et al.*, 2003; Lemmens *et al.*, 2006, 2009; Tanay *et al.*, 2004; Xu *et al.*, 2004).

Usually, many overlapping modules are identified by module detection tools. Having a visual overview of how these modules overlap, gives insight in the structure of the biological system. The problem with visualizing overlapping modules simultaneously, however, is that the overlap in multiple dimensions complicates the choice of an appropriate layout. Only few tools exist that are capable of visualizing modules simultaneously (Grothaus *et al.*, 2006; Santamaria *et al.*, 2008).

In this study, we developed ViTraM ('Visualization of Tran-scriptional Modules'). In comparison with previously developed tools, ViTraM not only allows for a dynamic visualization of overlapping modules, but also allows displaying additional information on the modules. As such, ViTraM can provide more

insight into the modules and makes the biological interpretation of the identified modules more accessible to biologists.

2 METHODS

ViTraM is a java-based tool. The minimal information required by ViTraM to visualize modules are the genes and experiments composing the modules. ViTraM visualizes the modules in a 2D display, called the ModuleDisplay, in which the rows represent the genes and the columns the experiments. Each regulatory module is represented in this display as a transparent colored rectangle. To improve the visualization of overlapping modules, ViTraM includes two different ordering algorithms. In addition, the user can make his own selection of modules or filter a module based on criteria outlined below.

Additional properties of the genes, experiments and modules are optional and can also be visualized by ViTraM:

- A gene property includes membership to a particular gene ontology class or the presence of a transcription factor binding site in or the binding of a regulator to the upstream region of the gene.
- An experiment property includes the membership of an experiment to a particular conditional class, which gives information on the major cue that was measured during the experiment.
- A module property is only available when the module inference tool was capable of identifying the regulatory program of the module. Module properties thus include a list of regulators or motifs that were assigned to the module by the inference algorithm.
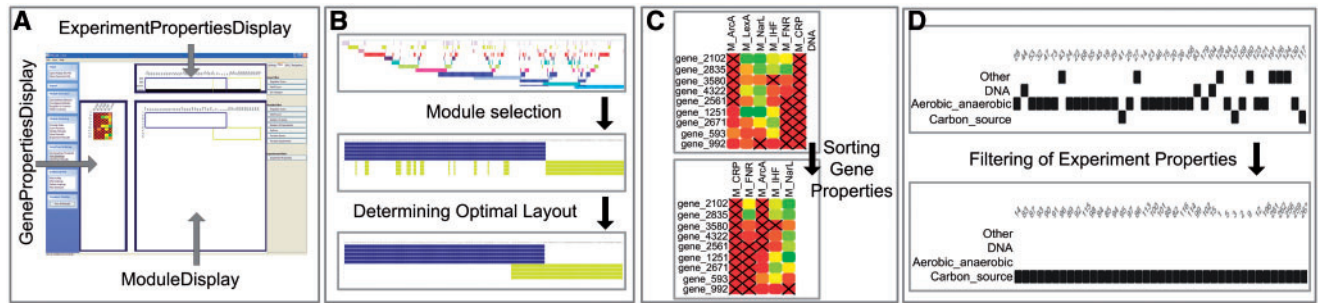
Two displays, the GenePropertiesDisplay and ExperimentPropertiesDisplay, show respectively the gene properties and the experiment properties. Both displays are dynamically linked to the ModuleDisplay, meaning that if the order of genes or experiments changes in the ModuleDisplay, their order will also change in the other two displays. In addition the user can change the properties that are shown and their order.

Finally, the OverviewDisplay, given in a separate window, provides a less detailed but total overview of all currently displayed modules and allows the user to keep track of which part of the module selection is currently displayed in the ModuleDisplay. Interactively navigating through the ModuleDisplay allows to see the rest of the selected modules into detail.

3 RESULTS AND CONCLUSIONS

To demonstrate that ViTraM can assist a user in analyzing the output of a module detection tool, it was applied on the results obtained by the module detection tool DISTILLER (Lemmens *et al.*, 2009). DISTILLER identified overlapping regulatory modules and

---

*To whom correspondence should be addressed.

Fig. 1. Overview on some of the functionalities of ViTraM. (**A**) The ViTraM software consists of several displays. The ModuleDisplay shows the modules in a 2D matrix in which the rows represent the genes and the columns the experiments. The GenePropertiesDisplay shows the genes and their properties, whereas the ExperimentPropertiesDisplay shows the experiments present in the modules and their properties. (**B**) An overview of all modules can be seen in the OverviewDisplay. Visualization of a subset of modules is possible, for instance, selecting all modules to which a module inference tool (here DISTILLER) assigned a particular motif (here CRP). After applying one of ViTraM's ordering methods, all modules can be displayed in a coherent way. (**C**) The GenePropertiesDisplay shows the scores for the binding of a regulator or the presence of a motif by a color gradient in which green is the lowest value and red the highest value. After sorting the gene properties based on these scores, it is clear that the CRP motif assigned by the module inference tool indeed has a high score for both modules. (**D**) The ExperimentPropertiesDisplay displays the conditional categories to which an experiment belongs. A selection of experiments that should be visualized can be made, for instance, only experiments that measure the influence of the carbon source can be visualized.

assigned at least one motif to each separate module. Based on this module property, all modules to which the module detection tool has assigned the regulatory motif for the regulator CRP were selected by using one of ViTraM's module selection techniques. Subsequently, the optimal layout of these modules is determined (Figure 1B).

In addition to the motifs assigned by DISTILLER, we screened all genes for the presence of known regulatory binding sites according to RegulonDB (Gama-Castro *et al.*, 2008). These scores are displayed in the GenePropertiesDisplay. When sorting the transcription factor binding sites according to their scores (Figure 1C) it is clear that the binding sites for CRP indeed showed the highest overrepresentation, confirming the results of DISTILLER that assigned the regulatory motif for regulator CRP to both modules.

The ExperimentPropertiesDisplay shows the experiments present in the modules and information on the conditional classes of which module experiments are part of. Many experiments in which the genes of these CRP modules were co-expressed belong to either the conditional category 'carbon-source' or 'anaerobiosis_aerobiosis'. A selection of experiments that measure for instance the influence of the carbon source was made (Figure 1D). These findings are consistent with the known functions of the assigned regulator CRP. The catabolite repressor is known to be active during glucose starvation and to interact with the regulators ArcA and FNR in response to oxygen (Gama-Castro *et al.*, 2008).

Figure 1 shows how ViTraM allows zooming in on and studying in detail a subset of modules and their properties, while maintaining an overview on how the different modules are related to each other. This interactive exploration of results can help biologists in the interpretation of the many modules that are present in the output of module inference tools.

REFERENCES

Bar-Joseph,Z. *et al.* (2003) Computational discovery of gene modules and regulatory networks. *Nat. Biotechnol.*, **21**, 1337–1342.

Gama-Castro,S. *et al.* (2008) RegulonDB (version 6.0): gene regulation model of Escherichia coli K-12 beyond transcription, active (experimental) annotated promoters and Textpresso navigation. *Nucleic Acids Res.*, **36**, D120–D124.

Grothaus,G.A. *et al.* (2006) Automatic layout and visualization of biclusters. *Algorithms Mol. Biol.*, **1**, 15.

Hartwell,L.H. *et al.* (1999) From molecular to modular cell biology. *Nature*, **402**, C47–C52.

Lemmens,K. *et al.* (2009) DISTILLER: a data integration framework to reveal condition dependency of complex regulons in Escherichia coli. *Genome Biol.*, **10**, R27.

Lemmens,K. *et al.* (2006) Inferring transcriptional modules from ChIP-chip, motif and microarray data. *Genome Biol.*, **7**, R37.

Madeira,S.C. and Oliveira,A.L. (2004) Biclustering algorithms for biological data analysis: a survey. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, **1**, 24–45.

Santamaria,R. *et al.* (2008) BicOverlapper: a tool for bicluster visualization. *Bioinformatics*, **24**, 1212–1213.

Tanay,A. *et al.* (2004) Revealing modularity and organization in the yeast molecular network by integrated analysis of highly heterogeneous genomewide data. *Proc. Natl Acad. Sci. USA*, **101**, 2981–2986.

Xu,X. *et al.* (2004) Learning module networks from genome-wide location and expression data. *FEBS Lett.*, **578**, 297–304.