Mantra 2.0: An online collaborative resource for drug mode of action and repurposing by network analysis

Diego Carrella[1], Francesco Napolitano[1], Rossella Rispoli[1], Mario Miglietta[2], Annamaria Carissimo[1], Luisa Cutillo[1,4], Francesco Sirci[1], Francesco Gregoretti[3], Diego Di Bernardo[1,5*]

[1]Telethon Institute of Genetics and Medicine, Via P. Castellino, 111, Naples, Italy
[2]Interactive SRL, Via Fratelli Bisogno 5 83100, Avellino, Italy
[3] Institute for High-Performance Computing - ICAR - CNR, Via P. Castellino, 111, Naples, Italy
[4] Dip. di Studi Aziendali e Quantitativi Università degli studi di Napoli "Parthenope", v. Generale Parisi, 13, 80132, Naples, Ialy
[5] Dep. Electrical Engineering and Information Technology, University of Naples "Federico II", v. Claudio 21, 80125, Naples, Italy

ABSTRACT

Summary: Elucidation of molecular targets of a compound (mode of action, MoA) and of its off-targets is a crucial step in drug development. We developed an online collaborative resource (MANTRA 2.0) that supports this process by exploiting similarities between drug-induced transcriptional profiles. Drugs are organised in a network of nodes (drugs) and edges (similarities) highlighting "communities" of drugs sharing a similar MoA. A user can upload gene expression profiles (GEPs) before and after drug treatment in one or multiple cell types. An automated processing pipeline transforms the GEPs into a unique drug "node" embedded in the drug-network. Visual inspection of the neighbouring drugs and communities helps in revealing its MoA, and to suggest new applications of known drugs (drug repurposing). MANTRA 2.0 allows storing and sharing user-generated network nodes, thus making MANTRA 2.0 a collaborative ever-growing resource.

Availability: The web tool is freely available for academic use at http://mantra.tigem.it.
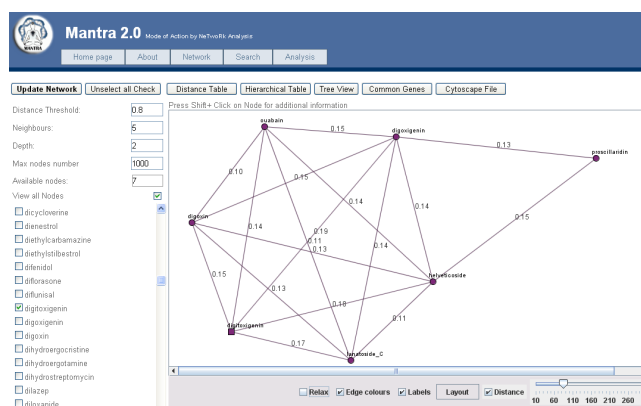
Contact: dibernardo@tigem.it

Fig. 1. MANTRA 2.0 web interface. Drugs are represented as nodes in a network. Edges are labelled with the distance between the two drugs.

1 INTRODUCTION

We developed a computational approach to predict drug mode-of-action and drug repurposing from the analysis of the Connectivity Map (Cmap, Lamb *et al.* 2006), a compendium of gene expression profiles (GEPs) following drug treatment of human cell lines with 1,309 bioactive small molecules (Iorio *et al.*, 2010). The approach was based on generating, for each drug, a single "prototype" ranked list (PRL) of differentially expressed genes following drug treatment across multiple cell lines, or at different dosages (Iorio *et al.*, 2010). We devised a drug similarity measure based on gene set enrichment analysis (GSEA, Subramanian *et al.* 2005) to compute similarity between two PRLs. Transcriptional similarities were represented as a network. The position of a drug within the network was shown to provide insights about its MoA by exploiting previous knowledge on neighboring drugs.

Here we introduce a collaborative online resource, named MANTRA 2.0, implementing our computational approach to drug discovery integrated with a relational database engineered to analyse and store user-provided GEPs and meta-data following small-molecule treatments, single gene perturbations or disease states. An example of a workflow can be found in the help section of the MANTRA 2.0 web site. MANTRA 2.0 guides the user beginning with GEP upload and annotation, which are then automatically transformed into a node in the network by connecting it to similar nodes, according to the our similarity measure. The result is visualised as an interactive network (Figure 1) that can be explored to reveal similarities among drugs, genes and diseases.

---

*to whom correspondence should be addressed

2 METHODS

2.1 Gene expression profile database

We used a relational database (PostgreSQL) for GEPs. The database contains three types of profiles: Drug-induced, Gene Perturbation and Disease. Each experiment is stored together with additional meta-data.

2.2 Microarray Analysis Pipeline

MANTRA 2.0 integrates a microarray data analysis pipeline to upload experiments in the drug network. Raw file formats (CEL) for Affymetrix chips HG-U133A, HT_HG-U133A, HG-U133A_2 and HG-U133_Plus_2 are supported. The pipeline implements the MAS5 and RMA algorithms (Gautier *et al.*, 2004). Differential expression analysis between treated and untreated samples is performed with a Bayesian t-test (Baldi and Long, 2001) and Benjiamini-Hochberg FDR correction (false discovery rate). MANTRA 2.0 also accepts pre-processed ranked lists of probes (HG-U133A) if raw data are not available or a custom data analysis pipeline is preferred.

2.3 Generation of prototype ranked lists and evaluation of similarity

PRL are generated by merging the ranked lists corresponding to the same drug by applying the Borda count method in a hierarchical fashion (Iorio *et al.*, 2010). The resulting PRL represents a node in the network. In order to obtain pairwise distances between the new node and the nodes already present in the network, we implemented a GSEA-based distance (Iorio *et al.*, 2010). Both the merging procedure and the similarity computation algorithms have been implemented in C language and built as a dynamic-link library (DLL).

2.4 Visualisation and interactive Graphical User Interface

The network can then be searched for neighboring nodes or communities. The user interface has been implemented in ASP.NET and VB.NET frameworks. A customized version of the MEDUSA (Hooper and Bork, 2005) visualization system (v.1.03) has been used for the network view.

3 IMPLEMENTATION

MANTRA 2.0 has been designed with attention to usability and reliability. The tool consists of three main workspaces: Analysis, Network and Search.

3.1 The Analysis workspace

The analysis workspace allows the user: (i) to add or remove nodes in the network. A node can be created by uploading microarray-based GEPs with meta-data. The PRL is added to the internal database, similarity distances are computed and the new node becomes part of the network; (ii) to identify which drugs in the network upregulate, or downregulate, a specific set of genes by applying GSEA against each PRL in the network.

3.2 The Network workspace

The "Network" workspace, shown in Figure 1, provides an interactive visualization of the network. The network is made up of nodes representing PRLs and edges connecting significantly similar pairs of PRLs. Nodes corresponding to drug treatments are color coded if they belong to the same drug community (Iorio *et al.*, 2010). The network can be navigated by means of a point and click interface. A number of options are available in order to limit the number of nodes and edges visualized according to both topological and similarity distance thresholds. The visualized subnetwork can be downloaded in the form of tabular data or exported to Cytoscape (Shannon *et al.*, 2003). It is also possible to obtain a summary of common differentially expressed genes for the selected nodes.

3.3 The Search workspace

The "Search" workspace allows searching the network for nodes of interest. Queries can be performed with drug/gene synonyms, by matching text in the node description or by proximity to a given node within a threshold.

4 CONCLUSION

MANTRA 2.0 is a collaborative resource for elucidating drug mode-of-action and for drug repurposing. The network framework makes it easy to discover unexpected relationships among drugs, genes and diseases. The possibility of uploading and sharing user's data allows for the growth of the network, thus making MANTRA 2.0 an ever-growing collaborative resource for drug discovery.

REFERENCES

Baldi, P. and Long, A. D. (2001). A bayesian framework for the analysis of microarray expression data: regularized t-test and statistical inferences of gene changes. *Bioinformatics*, **17**(6), 509–519.

Gautier, L., Cope, L., Bolstad, B. M., and Irizarry, R. A. (2004). affy—analysis of affymetrix genechip data at the probe level. *Bioinformatics*, **20**(3), 307–315.

Hooper, S. D. and Bork, P. (2005). Medusa: a simple tool for interaction graph analysis. *Bioinformatics*, **21**(24), 4432–4433. PMID: 16188923.

Iorio, F., Bosotti, R., Scacheri, E., Belcastro, V., Mithbaokar, P., Ferriero, R., Murino, L., Tagliaferri, R., Brunetti-Pierri, N., Isacchi, A., and Bernardo, D. d. (2010). Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proceedings of the National Academy of Sciences*, **107**(33), 14621–14626. PMID: 20679242.

Lamb, J., Crawford, E. D., Peck, D., Modell, J. W., Blat, I. C., Wrobel, M. J., Lerner, J., Brunet, J.-P., Subramanian, A., Ross, K. N., Reich, M., Hieronymus, H., Wei, G., Armstrong, S. A., Haggarty, S. J., Clemons, P. A., Wei, R., Carr, S. A., Lander, E. S., and Golub, T. R. (2006). The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science*, **313**(5795), 1929–1935.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research*, **13**(11), 2498–2504. PMID: 14597658.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., and Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, **102**(43), 15545–15550. PMID: 16199517.