

## An integrated system for mining relations among microRNAs, drugs and phenotypes

A. Pulvirenti<sup>\*1</sup>, R. Giugno<sup>\*1</sup>, S. Di Bella<sup>2</sup>, G. Nigita<sup>2</sup>, V. Macca<sup>2</sup>, A. Giummarra<sup>2</sup>, D. Garofalo<sup>2</sup>, G. Caruso<sup>2</sup>, V. Bonnici<sup>3</sup>, A. Ferro<sup>\*1</sup>✉

\*these authors contributed equally

<sup>1</sup>Dipartimento di Biomedicina Clinica e Molecolare, Università di Catania, Catania, Italy

<sup>2</sup>Dipartimento di Matematica e Informatica, Università di Catania, Italy

<sup>3</sup>Dipartimento di Informatica, Università di Verona, Verona, Italy

### Motivations

Interactions among genes together with their expression level determine a phenotype in a given organism. Genetic origin of a disease is often discovered once its phenotype has been clearly defined. Human phenotypes can be easily described by their observable features. Most human phenotypes can be found in the database OMIM (Online Mendelian Inheritance in Man). Van Driel et al. proposed a text mining technique to be applied to OMIM in order to find phenotypes similarity. On the other hand modern studies have shown that genetic variations may affect

drug response. Moreover, drugs similarity may be used to predict drug-diseases relations. The aim of such research area is to computationally predict drugs response and their side effects. Indeed this information may be used to reduce new drugs development cost. It is well known that microRNAs (miRNAs) expressions strongly affect phenotypes sometimes causing diseases. Consequently, miRNAs may provide insights on drug response, for example, by analysing their regulating interaction with drug targets. Simple associations miRNA-disease through their target genes can be obtained querying a knowledge

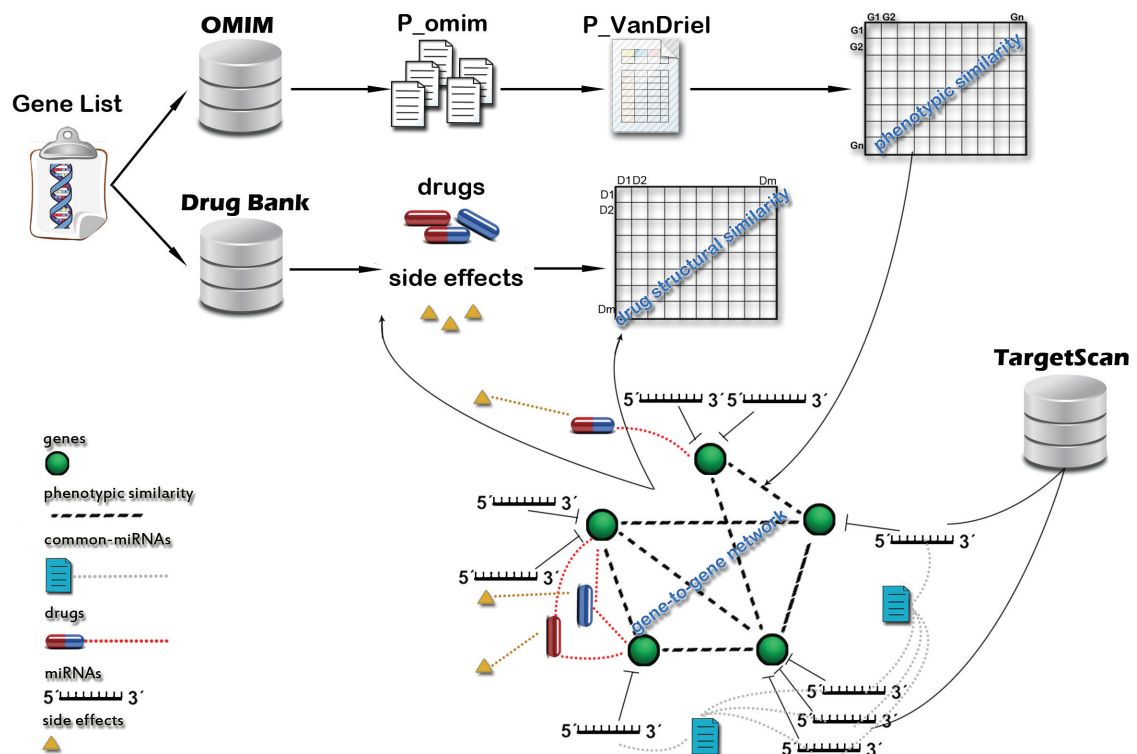


Figure 1. Data mining workflow.

base such as miRò. The subject of the present study is the design of an environment in which microRNAs-drugs-phenotypes relationships can be deduced using more advanced network-based techniques.

#### Methods

Starting from matrices of phenotypic and drugs similarities, the tool provides users an interface to establish those miRNAs, known to be responsible for the regulation of genes involved in a disease, associated with other diseases showing similar phenotypes. Given a list of genes, e.g. those composing a metabolic pathway, we lead a phenotypic and miRNAs-based analysis by calculating distances among gene pairs. The phenotypic analysis consists of two steps. In the first step, we extract for each pairs of genes the phenotypes from OMIM. Then, using Van Driel et al., for each phenotype, we select the one hundred most similar to it. This allows the construction of a

complete list of phenotypes associated to each genes pair. In the second step, we calculate the phenotypic distance matrix between any two genes according to the their common phenotypes. This can be viewed also as a weighted gene network. Applying TargetScan (or any other targeting program) to the endpoints of each edge in the network, we build a list of miRNAs. This yields a list of miRNA-phenotype associations. Genes in the networks are also correlated with drugs targeting them together with their side effects through drug-bank repositories. Drugs structural similarity is computed by Tanimoto technique.

#### Results

We propose a software tool that using gene-phenotypes association techniques in connection with gene network-based correlation algorithms allows the identification of statistically significant microRNA-phenotype-drug associations.