Y network and genes of these diseases to form a ranked prior vector and Complex Elucidation) (Vanunu et al., 2006). Recently, we introduced a novel method for prioritizing disease, PRINCE: (i) identifies a set of phenotypically similar diseases (van Driel et al., 2006); (ii) retrieves the known causal genes of these diseases to form a ranked prior vector \( Y \) based on their similarity to the query disease and (ii) propagates the scores of the prior set of genes over a human PPI network to provide association scores for all genes. The final score assigned to each protein in the network combines the prior information with a network-based component. The latter ensures that the resulting scores are smooth over the network nodes. In the input files section, the query disease is selected from a sorted list of OMIM diseases (either by name or MIM code). A textual search for the query disease is also available. Three additional inputs are required: (i) OMIM phenotype–disease associations, extracted from GeneCards (Rebhan et al., 2005); (ii) a disease names file corresponding to the similarity file entries (supplied by default with the plug-in); and (iii) a default set of disease–gene associations, extracted from GeneCards (Rebhan et al., 1998), used also by Vanunu et al. (2010).

The PRINCIPLE plug-in provides three tunable parameters: (i) the weighting parameter \( \alpha \in [0,1] \) (see Formula 1, with a default value of \( \alpha = 0.9 \)); (ii) \( k \in (0, 100) \), the number of top ranked genes to return (default 10); and (iii) \( \epsilon \in (0, 20) \), the number of iterations performed by the algorithm. The score \( F(v) \) can be analytically

\[ F(v) = \alpha \sum_{w \in N(v)} F(w) \cdot w(v, u) + (1 - \alpha)Y(v) \]

Where \( w \) is a normalized matrix representing the weighted PPI network and \( Y(v) \) is the prior weight of the node. Here \( \alpha \) is parameter weighting the relative importance of the prior-based versus the network-based components of the score.

PRINCIPLE leverages on a comprehensive set of weighted PPIs compiled from multiple sources (Vanunu et al., 2010), the disease–disease similarity measures computed by van Driel et al. (2006), and on the disease–gene associations presented in the Online Mendelian Inheritance in Man (OMIM) knowledgebase (Hamosh et al., 2002).

Here we introduce PRINCIPLE (PRINCe InPLementation)—a Cytoscape plug-in (Shannon et al., 2003) implementation of the PRINCE algorithm. Given a query disease, it provides a list of top ranking genes associated with it and an additional visualization of the subnetworks formed by these top ranking genes and their direct interacting neighbors.
An optional output file can be specified, listing the gene scores.

Top 10 associated genes and possible references to Diabetes mellitus. Right clicking on a gene enables retrieving additional information on the protein from multiple data sources.

ACKNOWLEDGEMENTS

We would like to thank Oron Vanunu and Tomer Shlomi for their help in the original PRINCE implementation. We would like to thank Nir Atias for helping with the plugin quality assurance.

Funding: Edmond J. Safra bioinformatics program (to A.G.); Bikura grant from the Israel Science Foundation (to E.R. and R.S.).

Conflict of Interest: none declared.

REFERENCES


Table 1. Top 10 associated genes and possible references to Diabetes mellitus

<table>
<thead>
<tr>
<th>Rank</th>
<th>Gene</th>
<th>Supporting reference</th>
<th>Rank</th>
<th>Gene</th>
<th>Supporting reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PDX1</td>
<td>OMIM</td>
<td>6</td>
<td>PLN</td>
<td>Bergh A et al. (2006)</td>
</tr>
<tr>
<td>2</td>
<td>MAPK8IP1</td>
<td>OMIM</td>
<td>7</td>
<td>BSCL2</td>
<td>Chen et al. (2009)</td>
</tr>
<tr>
<td>3</td>
<td>PPP1R3A</td>
<td>OMIM</td>
<td>8</td>
<td>HNF4A</td>
<td>Möller et al. (1997)</td>
</tr>
<tr>
<td>4</td>
<td>HNF1A</td>
<td>Winckler et al. (2005)</td>
<td>9</td>
<td>NEUROD1</td>
<td>Liu et al. (2007)</td>
</tr>
<tr>
<td>5</td>
<td>MAFA</td>
<td>Kaneto et al. (2008)</td>
<td>10</td>
<td>PCIF1</td>
<td>Claiborn et al.</td>
</tr>
</tbody>
</table>

Fig. 1. An example of the PRINCE output subnetwork for NIDDM, displaying an extract of the top 10 scoring genes and their immediate neighbors. Nodes are colored according to their association scores, with darker colors denoting higher scores.

3 USAGE EXAMPLE

Figure 1 shows a typical output for querying Diabetes mellitus, solving, but for efficiency we compute it using an iterative procedure (Zhu et al., 2004). Typically, the algorithm shows fast convergence, achieving optimal results after 10 iterations (Vanunu et al., 2010).

The results are displayed as the k top priority genes and their direct PPI neighbors, using a color scale signifying relative scores. An optional output file can be specified, listing the gene scores.

A.Gottlieb et al.