WebPARE: web-computing for inferring genetic or transcriptional interactions

Cheng-Long Chuang1,2,†, Jia-Hong Wu1,†, Chi-Sheng Cheng1 and Grace S. Shieh1,2,∗

1Institute of Statistical Science, Academia Sinica, Taipei 115, Taiwan and 2Institute of Biomedical Engineering, National Taiwan University, Taipei 106, Taiwan

Received on October 5, 2009; revised on December 4, 2009; accepted on December 7, 2009
Advance Access publication December 10, 2009
Associate Editor: Trey Ideker

ABSTRACT
Summary: Inferring genetic or transcriptional interactions, when done successfully, may provide insights into biological processes or biochemical pathways of interest. Unfortunately, most computational algorithms require a certain level of programming expertise. To provide a simple web interface for users to infer interactions from time course gene expression data, we present WebPARE, which is based on the pattern recognition algorithm (PARE). For expression data, in which each type of interaction (e.g. activator target) and the corresponding paired gene expression pattern are significantly associated, PARE uses a non-linear score to classify gene pairs of interest into a few subclasses of various time lags. In each subclass, PARE learns the parameters in the decision score using known interactions from biological experiments or published literature. Subsequently, the trained algorithm predicts interactions of a similar nature. Previously, PARE was shown to infer two sets of interactions in yeast successfully. Moreover, several predicted genetic interactions coincided with existing pathways; this indicates the potential of PARE in predicting partial pathway components. Given a list of gene pairs or genes of interest and expression data, WebPARE invokes PARE and outputs interactions and their networks in a directed graph.

Availability: A web-computing service WebPARE is publicly available at: http://www.stat.sinica.edu.tw/WebPARE
Contact: gshieh@stat.sinica.edu.tw
Supplementary Information: Supplementary data are available at Bioinformatics online.

1 INTRODUCTION
Genetic interaction (GI) networks may reveal how a group of genes function together to carry out a biological process and unravel cellular buffering mechanisms (Boone et al., 2007), while predicting transcriptional regulatory interactions (TIs) may reveal the regulatory mechanisms in organisms (Wang et al., 2007). Henceforth, we use interactions to denote GIs or TIs. Recently, there have been a few studies on GIs (Wong and Roth, 2005). Paralogs or redundant genes are called SSL gene pairs if the combination of two mutants, neither by itself lethal, causes the organism to die or malfunction. Other types of GIs of interest are transcriptional compensatory and transcriptional diminishment interactions from SSL gene pairs (Chuang et al., 2008). Following a gene’s loss, the expression level of its compensatory gene increases (decreases), and this phenomenon is called transcriptional compensatory (transcriptional diminishment). With the emergence of modern biotechnologies, various computational methods have been proposed to predict interactions using gene expression data and/or other experimental data. Inferring these interactions, when done successfully, can provide insights into biological processes or biochemical pathways of interest. Unfortunately, most computational algorithms require a certain level of programming expertise. A web-computing implementation of such an algorithm provides easy access to predicting interactions that are not annotated in any databases or literature.

We have previously published the pattern recognition algorithm PARE (Chuang et al., 2008), which can infer interactions from time course expression data, provided that each type of interaction, e.g. AT or RT, and the corresponding paired gene expression pattern are significantly associated. PARE uses a non-linear score to classify gene pairs of interest into subclasses of various time lags. In each subclass, PARE learns the parameters in the decision score using known interactions from biological experiments or published literature. Subsequently, the trained algorithm predicts interactions of a similar nature. PARE was shown to infer two sets of interactions in yeast successfully using expression data and existing knowledge such as 112 pairs of qRT-PCR validated GIs. Moreover, several of the predicted GIs coincided with existing pathways in yeast. This indicates that PARE has the potential to predict biochemical pathways, while altered pathways are likely to play key roles in cancers and other human complex diseases (Ding et al., 2008). Recently, we applied PARE to infer TIs involved in human adipogenesis, and preliminary results identified some promising transcription factors for further biological experiments (J.-D. Zucker and K. Clement, unpublished data). Furthermore, a web-computing of PARE will be quite useful to predict GIs for recent large-scale SGA results in yeast.

Here, a web-computing implementation of PARE (WebPARE) is presented, which attempts to provide a simple web interface for users to infer interactions from time course gene expression data. In addition, a graphical display of the predicted network is also provided. In the following, we outline the architecture of WebPARE, and conclude...
WebPARE checks whether the uploaded expression data is in
where select the value of and (ii) Fisher's exact test for the training data, in which users can
proceed to classify each gene pair into a particular subclass in
or use the default values (Fisher's test, according to the guidelines (Supplementary Material)
validated yeast GIs and known TIs in
Some integrated existing interactions, e.g. the 112 pairs of qRT-PCR
arrived in the queue, WebPARE distributes it to the computing unit.
formed if a list of genes is uploaded. After a new request has
in the queue, WebPARE distributes it to the computing unit.
Some integrated existing interactions, e.g. the 112 pairs of qRT-PCR
validated yeast GIs and known TIs in Arabidopsis; yeast, mouse and
human, are used to train parameters of PARE or a set of default
parameters can be used. More existing TIs in other species will be
integrated in the near future. All gathered information is then passed
to the computing unit.
The key procedures of the computing unit are outlined as follows; we refer to Chuang et al. (2008) for the details of PARE. First, WebPARE checks whether the uploaded expression data is in
PreCLustering file format; see the website for an example. Next, a
filtering process applied to expression data checks whether uploaded
data satisfies the assumptions of PARE. Namely, (i) whether any
gene expression curve of interest is too ‘flat’ to be predicted [to satisfy Equation (1) in Chuang et al., 2008],
\[
\max \left\{ \frac{G_i(t)}{G_j(t)} \right\} - C \geq 0
\]
where \(G_i(t)\) denotes gene \(i\)’s expression after smoothing at time \(t\), and (ii) Fisher’s exact test for the training data, in which users can
select the value of \(C\) and the percentage of the training passing
Fisher’s test, according to the guidelines (Supplementary Material)
or use the default values (\(C = 1.4\) and 50%). Once the dataset passes
this filtering step, among subclasses with a few time lags, PARE
proceeds to classify each gene pair into a particular subclass in
which an interaction occurs most probably. In each subclass, either
the particle swarm optimization algorithm is used to optimize the
parameters using known interactions or the default values are used.
Finally, all gene pairs are scored by PARE.
After WebPARE finishes a request, an email will notify the user
to download the result, in which the most probable time lag, the
associated PARE score and the predicted interaction type for each
gene pair are outputted. In addition, a directed graph of the predicted
interactions (a Cytoscape session file) is reported (Supplementary Material), in which each node denotes a gene and is labeled with the
gene name, while each edge represents a significant predicted interaction; non-significant interactions, those where the absolute
values of PARE scores are smaller than the threshold, are not plotted.
A solid edge represents an AT (or transcriptional diminishment)
interaction, while a dashed edge denotes a RT (or transcriptional
compensatory) interaction when inferring TIs (or GIs).
The web-interface unit of WebPARE is written in ASP, and runs
on Microsoft internet information services web server, while the
computing unit is written in MATLAB. Currently, WebPARE allows
100 thousands queries/access.

3 AN EXAMPLE
Suppose that a list of 15 gene pairs involved in cell cycle using
expression data from cyclin-mutant yeast cells (Orlando et al.,
2008) were uploaded to WebPARE, and TIs of these gene pairs
were of interest. In the filtering step, since all 15 pairs were to be
predicted, following the guidelines (Supplementary Material)
the user relaxed the value of \(C\) to 1.1 such that all gene pairs
passed the filtering process of Equation (1). Next, the 162 integrated
(prestored) pairs of known TIs in yeast passed Equation (1) with
\(C = 1.4\), and 100% of them passed the Fisher’s exact test. Therefore,
WebPARE was invoked. All integrated yeast TI pairs were classified
into subclasses with distinct time lags based on their PARE scores
with the default weights (1, 1, 3.5). In each subclass, the integrated
known TIs were used to train the parameters of PARE, and the TIs of
interest were predicted. After comparing the predicted results with
published literature, the modified true positive rate (mTPR) was
60% (9/15), where mTPR was defined as the ratio of the number
of correctly predicted interactions to the total known interactions
among all gene pairs. However, if the user preferred more accurate
predictions, following the guidelines (Supplementary Material) the
user would apply larger values of \(C\). Setting \(C\) to 1.4 and 1.5
reduced the number of gene pairs to be predicted to 10 and 10,
respectively, and both their mTPRs were 70%. This echoes the
guideline that a larger value of parameter \(C\) in Equation (1) leads to
more accurate predictions, but has a risk of filtering out gene pairs
of interest. The significant predicted network for the 10 pairs is in
the (Supplementary Material). A pilot study of predicting 99 gene
pairs (Supplementary Material) resulted in mTPRs 72% and 82% for
\(C\) equal to 1.1 and 1.5, respectively; the experiment took \sim16\ min,
which was conducted by PC with Pentium Core 2 1.86 GHz and
1.0GB RAM.

ACKNOWLEDGEMENTS
We wish to thank Dr Chung-Ming Chen for suggesting a web-
computing of PARE in 2008 and Mr Chia-Chang Wang for
participating in collecting TIs.
C.-L. Chuang et al.

Funding: National Science Council, Republic of China (98-2118-M-001-017).

Conflict of Interest: none declared.

REFERENCES

Boone, C. et al. (2007) Exploring genetic interactions and networks with yeast. Nat Rev Genet, 8, 437–449.

Chuang, C.L. et al. (2008) A pattern recognition approach to infer time-lagged genetic interactions. Bioinformatics, 24, 1183–1190.

Ding, L. et al. (2008) Somatic mutations affect key pathways in lung adenocarcinoma. Nature, 455, 1069–1075.

Orlando, D.A. et al. (2008) Global control of cell-cycle transcription by coupled CDK and network oscillators. Nature, 453, 944–947.

Wang, R.S. et al. (2007) Inferring transcriptional regulatory networks from high-throughput data. Bioinformatics, 23, 3056–3064.

Wong, S.L. and Roth, F.P. (2005) Transcriptional compensation for gene loss plays a minor role in maintaining genetic robustness in Saccharomyces cerevisiae. Genetics, 171, 829–833.