Quality of Life through Quality of Information
J. Mantas et al. (Eds.)
IOS Press, 2012

doi:10.3233/978-1-61499-101-4-1156

Translational Meta-Analysis Tool for Temporal Gene Expression Profiles

Guenter TUSCH[a,1] and Olvi TOLE[a]
[a] *Medical and Bioinformatics Program, School of Computing and Information Systems, Grand Valley State University, Allendale, MI, USA*

Abstract. Widespread use of microarray technology that led to highly complex datasets often is addressing similar or related biological questions. In translational medicine research is often based on measurements that have been obtained at different points in time. However, the researcher looks at them as a progression over time. If a biological stimulus shows an effect on a particular gene that is reversed over time, this would show, for instance, as a peak in the gene's temporal expression profile. Our program SPOT helps researchers find these patterns in large sets of microarray data. We created the software tool using open-source platforms and the Semantic Web tool Protégé-OWL.

Keywords. Research and education, epidemiological research and clinical trials, bioinformatics, translational research**.**

Introduction

Recent use of microarray technology has led to highly complex datasets often addressing similar biological questions. The statistical methodology of meta-analysis aims to combine results from independent but related studies and is used to potentially extend confirmed results to other species or organs [1]. If the researcher is not interested in confirmatory analysis but in exploring how temporal patterns discovered in a reference study translate into similar studies, he can select interesting studies in an analogous way [2].

Typically, for stimulus response studies a researcher obtains a fold change profile and tries to retrieve similar profiles in microarray databases or clinical databases (that more frequently include microarray data). Peaks in genes in temporal microarray studies represent a biological effect that is reversed after some time. Assume the researcher conducted a temporal microarray study where he/she discovered peaks in certain genes. He/she now wants to see if finding the same effect in related studies can extend his/her hypothesis. This is an example for the kind of research our tool can support. We describe a web interface that allows the researcher to access time-series micro-array datasets from the US National Center for Biotechnology Information Gene Expression Omnibus (NCBI GEO). He/she then can search for the same discovered genes in related temporal data sets and see if the same genes exhibit peaks in, for instance, other tissues, drugs, or even species.

[1] Corresponding Author. Guenter Tusch, email: tuschg@gvsu.edu.

1. Methodology

For stimulus response time-series microarray studies a researcher obtains a fold change profile from the experiment. To retrieve similar profiles in microarray databases would be easiest if the pattern of time points is identical or at least very similar to the original experimental design, for example 0h, 2h, 4h, etc. In general, the pattern of time points depends on the experiment itself and the particular hypothesis the experiment tries to address. Some meta-analyses have used only the intersection of all time points or a selection with only the data available at that particular time. This is most likely only feasible if the studies are closely related.

For our exploratory approach, it is more realistic to look, for instance, for a peak in the profile instead of correlating the entire profile. This can be accomplished by using knowledge-based temporal abstraction [3], dynamic time warping or spline interpolation approaches [4], where time-stamped data points are transformed into an interval-based representation, or some statistical approaches [5].

We implemented these ideas by creating a platform SPOT (**S** - **P**rotégé – **O**WL/SWRL – **T**emporal Abstraction) [3] based on open-source software. It supports the R statistical package (an open source implementation of the S statistical scripting language) and knowledge representation standards (OWL, SWRL) using the Semantic Web tool Protégé-OWL connecting to the user through a web interface.

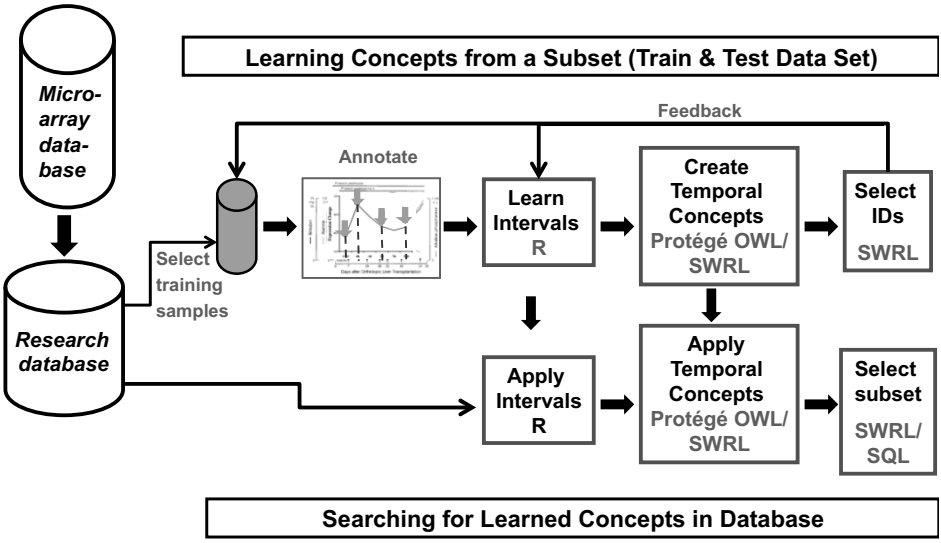

Figure 1. The SPOT Design.

2. System Description and Evaluation

The researcher (user) needs to perform the following steps using the web interface: He/she selects interesting studies, currently only from NCBI GEO. Typically, a

researcher will choose datasets from one microarray platform or related platforms, e.g. Affymetrix. The system is very flexible and gives the user a variety of choices. The system currently uses the temporal abstraction paradigm. This means that time pattern is constructed from generic intervals with specific time properties, for instance, "increasing", "decreasing", or "constant". A peak can be described as an increasing interval immediately followed by a decreasing interval using Allen's interval algebra.

The user has to select training samples (genes and time-series) and within those identify the intervals that allow the system to learn their properties and automatically apply them to the entire database. Thus temporal patterns like, for instance, peaks in fold changes of the expression data of a particular set of genes can be found in the entire database. More complex patterns can be defined using the SWRL language in the Protégé-OWL SWRL built-ins [6]. The program generates an initial set OWL/SWRL rules from templates that can be modified by the user. The user also can create customized rules within the Protégé software environment. The researcher (user) can search for similar profiles in the database of the previously selected interesting studies. He/she can interrupt the process any time and come back to it later.

The system is a multi-user system, where several researchers can work in parallel. The tools we developed to support this process use the Protégé-OWL ontology development toolkit (compare [3,6]). For a more sophisticated analysis the data can be downloaded from the web site as a Protégé OWL file and analyzed locally in Protégé. For testing purposes we used GEO GSE1001 [7] as the reference study. Furthermore, a biologist has extensively evaluated the user interface and usability of the system by systematically selecting temporal microarray studies from NCBI GEO with sufficient number of time points. The necessary changes have been incorporated into the system. The next step is the use of data from a new study from our partner institution Van Andel Research Institute in Grand Rapids, MI, USA.

References

[1] Wren JG, A global meta-analysis of microarray expression data to predict unknown gene functions and estimate the literature-data divide. Bioinformatics 2009;25(13): 1694-1701.
[2] Ramasamy A, Mondry A, Holmes CC, Altman DG, Key Issues in Conducting a Meta-Analysis of Gene Expression Microarray Datasets. PLoS Med 2008;5(9): e184.
[3] Tusch G, Bretl C, O'Connor M, Das A, SPOT -towards temporal data mining in medicine and bioinformatics, AMIA Annu Symp Proc. 2008;1157.
[4] Smith AA, Vollrath A, Bradfield C, Craven M, Similarity queries for temporal toxicogenomic expression profiles, PLoS Computational Biology 2008; e1000116
[5] Sboner A, Romanel A, Malossini A, Ciocchetta F, Demichelis F, Azzini I, Blanzieri E, Dell'Anna R, Simple methods for peak and valley detection in time series microarray data, in P. McConnell, S. Lin, P. Hurban (eds.), Methods of Microarray Data Analysis V, New York: Springer; 2007: 27-44
[6] O'Connor MJ, Shankar RD, Parrish DB, Das AK. Knowledge-Data Integration for Temporal Reasoning in a Clinical Trial System. Int J Med Inform . 2009; 78(1): S77-S85.
[7] Vázquez-Chona F, Song BK, Geisert EE Jr: Temporal changes in gene expression after injury in the rat retina. Invest Ophthalmol Vis Sci 2004; 45(8): 2737-46.