MetNet DB: a comprehensive metabolic and regulatory network database

Hailong Zhang

Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd

Recommended Citation
https://lib.dr.iastate.edu/rtd/21374

This Thesis is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
MetNet DB: A comprehensive metabolic and regulatory network database

by

Hailong Zhang

A thesis submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Bioinformatics and Computational Biology

Program of Study Committee:
Eve Syrkin Wurtele, Co-major Professor
Julie Dickerson, Co-major Professor
Les Miller

Iowa State University
Ames, Iowa

2002

Copyright © Hailong Zhang, 2002. All rights reserved.
Graduate College
Iowa State University

This is to certify that the master’s thesis of

Hailong Zhang

has met the thesis requirements of Iowa State University

Signatures have been redacted for privacy
# TABLE OF CONTENTS

**CHAPTER 1. GENERAL INTRODUCTION**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>THESIS ORGANIZATION</td>
<td>2</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>3</td>
</tr>
</tbody>
</table>

**CHAPTER 2. REPRESENTING METABOLIC AND REGULATORY NETWORKS IN A RELATIONAL DATABASE: METNET DB**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>5</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>5</td>
</tr>
<tr>
<td>A GENERAL DATA MODEL</td>
<td>6</td>
</tr>
<tr>
<td>SOME EXAMPLES FOR THE DATA MODEL</td>
<td>8</td>
</tr>
<tr>
<td>METNET DB</td>
<td>9</td>
</tr>
<tr>
<td>METNET DB AND FCMODELER</td>
<td>10</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENT</td>
<td>11</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>11</td>
</tr>
</tbody>
</table>

**CHAPTER 3. INTEGRATED FUNCTIONAL ANNOTATIONS FOR ARABIDOPSIS MICROARRAY PROBES: PROBE DATABASE**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>14</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>14</td>
</tr>
<tr>
<td>PROBE DATABASE</td>
<td>16</td>
</tr>
<tr>
<td>WEB INTERFACE UTILITIES</td>
<td>18</td>
</tr>
<tr>
<td>INTEGRATION OF PROBE DATABASE WITH GENESPRING</td>
<td>20</td>
</tr>
<tr>
<td>SUMMARY</td>
<td>20</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>20</td>
</tr>
</tbody>
</table>
CHAPTER 4. GENERAL CONCLUSION

SUMMARY

FUTURE WORK
CHAPTER 1. GENERAL INTRODUCTION

INTRODUCTION

One of the major tasks and challenges in the post-genomics era is to determine the cellular functions of genes and their products, to understand how the interactions among the entities in cellular contexts could yield a living cell. Microarray, proteomics and other new techniques such as metabolomics are transforming the studies of gene expression. These techniques make it possible to monitor global changes in cellular RNAs, proteins and small molecules simultaneously. With the help of these new techniques, biologists could obtain more vast amounts of valuable data on gene expression than ever [1-8].

All of these point to the need of developing new methodologies to carry out the data mining efficiently in order to uncover the mysterious complex biological networks. A major and exciting challenge is how to extract the useful information from the data sets and combine it with what we already know about pathways and the regulation to achieve a better understanding of how metabolism is regulated in a living cell [1]. To attack this problem, Gene Expression Tool kit (GET) project was launched in Iowa State University. The overall goal of GET project is to develop novel tools that will provide a framework for the formulation of testable hypotheses regarding the function of specific genes, and in the long term provide the basis for identification of genetic regulatory networks that control plant composition and development. These tools will include capabilities for: 1) mapping metabolic and regulatory networks; 2) integrating and visualizing data and information from the literatures; 3) modeling the metabolic or regulatory flow in the network using fuzzy cognitive maps.
The global architecture of GET project is illustrated in Figure 1. PathBinder is a literature-mining tool to pull out the possible interaction and pathway information from PubMed database. ChipViewer is a statistical tool to cluster and process RNA, protein and metabolites profiling data. FCModeler is the visualization and modeling tool to model the metabolic or regulatory flow in the network [9]. MetNet DB serves as an information hub in the GET project.

THESIS ORGANIZATION

Chapter 1 is the introduction to this thesis and provides a summary of the background of the thesis. The global architecture of the GET package is presented. The thesis organization is also provided.

Chapter 2 describes a general data model for representing Metabolic and Regulatory Networks. Several examples are given to explain how the data model could be used to represent the biological networks. The data model is implemented in a relational database: MetNet DB.

Chapter 3 provides a detailed illustration for Probe database, which is one of the practical applications of MetNet DB. Probe database provides the integrated annotations for Arabidopsis microarray probes. The Probe database web application utilities and the integration of Probe database with the other microarray data analysis tools, such as GeneSpring are discussed in this chapter.

Chapter 4 states the conclusion drawn from the work in this thesis and suggests possible directions for the future research.
REFERENCES


Figure 1: Global architecture of the Gene Expression Tool kit (GET) package
CHAPTER 2. REPRESENTING METABOLIC AND REGULATORY NETWORKS IN A RELATIONAL DATABASE: METNET DB

A paper to be submitted to Bioinformatics

Hailong Zhang, Julie A. Dickerson and Eve Syrkin Wurtele

ABSTRACT

Summary: In this paper we will describe a general data model for representing metabolic and regulatory networks. The data model is implemented in a relational database: MetNet DB. MetNet DB serves as an information hub in Gene Expression Tool kit (GET) software package. It provides an integrated and flexible view for metabolic and regulatory networks. With the combination of the other modules of the GET package, sophisticated metabolic network modeling approaches could be achieved. Additionally, the MetNet DB web interface has the functionality of building new interactions and pathways based on user input. This provides a flexible pathway-building tool for the biological research community.

Availability: MetNet DB was implemented in MySQL and the web interface was written in PHP. It is freely available to noncommercial users. MetNet DB is available over the web at
http://get.botany.iastate.edu/metnet/

Contact: Dr. Eve Wurtele mash@iastate.edu

INTRODUCTION

One of the major challenges in the post-genome era is to determine the cellular functions of genes and their products, to understand how the interactions among the entities in cellular contexts could yield a living cell. Recent technological advances enable the high-
throughput detection and measurement of changes in the accumulation of cellular components [1]. This requires us to develop efficient software tools to process tons of valuable data generated every day. In order to achieve an understanding of how living cells function, we must be able to represent the complex biological networks in a way that enables us to manipulate and analyze it by using the full power of computers. The success in deciphering biological functions in a living cell will largely depend on our ability to represent the biological networks in a sophisticated way [2,7]. In eukaryotic cells, the subcellular compartments in which metabolism and regulations occur must be considered in the data model [1].

We have developed a general data model to address these problems. In what follows we will describe how to use the model to represent the metabolic and regulatory biological networks in a rigorous yet flexible manner.

A GENERAL DATA MODEL

The Entity-Relationship models of MetNet DB data model are illustrated in Figures 1 and 2. In the data model, there are two major classes: BlockUnit and Entity.

BlockUnit class is the basic unit to build the biological networks. It includes three sub-classes: Pathway, Interaction and EntityWithContext. The EntityWithContext sub-class represents the physical entities within the cellular context, such as the compound ATP in the chloroplast. The Interaction sub-class represents the physical or indirect interactions among the physical entities within the cellular context, such as a biochemical reaction among several compounds in the mitochondrion. Each Interaction is characterized by its particular list of inputs (EntityWithContext or Interaction) and outputs (EntityWithContext or Interaction).
For example, a biochemical reaction has a set of substrates as inputs and has a set of products as the outputs. The Pathway sub-class represents the pathways in cell. It is a collection of several Interactions. If we use a graph to represent the whole biological network, then the Pathway sub-class is a sub-graph. The Interaction sub-class is an edge and the EntityWithContext sub-class is a vertex. One of the special features of MetNet DB data model is that the Interaction sub-class could have other Interactions or even itself as outputs rather than only EntityWithContexts. This applies to enzyme-catalyzed reactions. In this case, the Interaction (enzyme-involved catalytical reaction) would use another Interaction (biochemical reaction) as the output. The Interactions having Interactions as output are denoted as Control Interactions in MetNet DB data model. Control Interactions can regulate other Interactions.

Another major class in the data model is Entity. Entity is the basic unit from which the EntityWithContext sub-class is derived. It represents the physical entities in cell. The Entity class cannot be used to build networks directly, because the cellular context needs to be specified. One good example to distinguish the concept of an Entity from an EntityWithContext is that the compound ATP is an Entity, but the compound ATP in the chloroplast is an EntityWithContext, because the later one includes the specific cellular context. The Entity class includes seven sub-classes: RNA, Gene, Polypeptide, Protein Complex, Compound, Cis-Element and Environmental Effectors. The information contained in most biological databases is for the properties of Entity sub-classes, such as gene sequences, molecular weights and formulas of compounds, etc.
SOME EXAMPLES FOR THE DATA MODEL

By using the data model, the complex cellular interaction processes could be represented, including post-translation modification, protein modification on the different sites and the translocations etc. The following examples are used to illustrate how the goal is achieved.

Translocation

Example case: the protein factor X transports from chloroplast to mitochondrion.

Solution: To represent this process, two EntityWithContexts are needed: X(c) and X(m). Both of the EntityWithContexts are derived from Entity polypeptide X. A translocation Interaction K is formed by using X(c) as input and using X(m) as output.

Enzyme Reaction

Example case: A+B→C, and E is the enzyme that catalyzes the reaction. The reaction occurs in the chloroplast.

Solution: We need two Interactions to represent the above reaction. One is biochemical reaction Interaction P, its inputs are A(c) and B(c) and its output is C(c). A(c), B(c) and C(c) are derived from Entity A, B and C, respectively. The second Interaction Q is a Control Interaction, its input is E(c), which is derived from Entity E, and its output is Interaction P.

Protein Complex Formation

Example case: X, Y, Z are the subunits of the protein M. These three subunits form the mature protein M. This complex formation reaction occurs in mitochondrion.
Solution: One Interaction is needed to represent the above reaction: Complex formation Interaction H. Its inputs are X(m), Y(m) and Z(m) and its output is M(m). X(m), Y(m), Z(m) and M(m) are derived from Entity polypeptide X, Y, Z and Entity Protein Complex M, respectively.

METNET DB

Database Schema

The MetNet DB schema is illustrated in Figure 3. The detailed explanation can be found at the MetNet DB web site http://get.botany.iastate.edu/metnet/new-schema/. There are 23 tables in MetNet DB. The schema provides the functionality to store the very detailed pathway and interaction information including literature references and external database records. The tables are grouped into BlockUnit and Entity two classes, which are mapping with the ER model of MetNet DB.

Data Preparation and Information Source

The portion of MetNet DB data is obtained from the other databases. Entity information is from the KEGG ligand database, TAIR database and AraCyc database, respectively. Some of the Interaction and Pathway information is from the AraCyc database. The KEGG ligand database provides a reasonable comprehensive collection for the compounds in cell [3]. The AraCyc database is an effort of TAIR to annotate the biological pathways of Arabidopsis. It contains Arabidopsis metabolic pathway information [4]. The TAIR Arabidopsis protein/gene collection is the most complete and non-redundant available protein/gene catalog of Arabidopsis [5]. Python codes are used to process the raw data downloaded from
the external databases. Microsoft Access is also used to perform the data preparation before the data is loaded into MetNet DB.

**Web Interface**

MetNet DB web interface is developed in PHP. It serves as a flexible pathway-building tool and a pathway information management system for the research community. Each biological expert user could build the pathways and interactions he/she knows well. This will help the community to establish an integrated network, rather than partial pieces of networks. A Bookmark folder feature has been implemented in MetNet DB web application. By using this feature, one could create personalized Bookmark folder to save all interesting records for later usage. The records in Bookmark folder could be used to build new Interaction and Pathway records. To ensure the quality of user-submitted records and for the security reason, MetNet DB web interface requires user registration and logon to access the MetNet DB information. The users of MetNet DB web interface are in three different privilege levels: Normal User, Expert User and Curator. The Normal User level has the lowest level privilege. The Normal user can only browse and query MetNet DB. The Expert User has the additional privilege to build the new EntityWithContexts, Interactions and Pathways records. The Curator has the highest privilege, including the privilege of deleting records and user account management.

**METNET DB AND FCMODELER**

FCModeler is a visualization and network-modeling tool. By using FCModeler, one could perform sophisticated network modeling and analysis computations, such as to identify the
alternative routes, to perform simulations, etc. FCModeler models the metabolic and regulatory flow in the network using fuzzy cognitive maps [6].

ACKNOWLEDGEMENT

The authors are grateful to Dr. Lucas Mueller for providing AraCyc database bulk data files, and Dr. Carol Foster and Beth Fatland for helpful biological input.

REFERENCES


---

**Figure 1:** Entity-Relationship model of BlockUnit class

---

**Figure 2:** Entity-Relationship model of Entity class
Figure 3: MetNet DB Schema
CHAPTER 3. INTEGRATED FUNCTIONAL ANNOTATIONS FOR ARABIDOPSIS MICROARRAY PROBES: PROBE DATABASE

A paper to be submitted to Plant Molecular Biology

Hailong Zhang, Julie A. Dickerson and Eve S. Wurtele

ABSTRACT

Summary: Probe database is a web database to provide the integrated functional annotations for Arabidopsis microarray probes. Presently Affymetrix Arabidopsis GeneChip AtGenome1, ATH1 and AFGC EST microarray, three large datasets are supported. The web interface of Probe database also provides useful utilities for microarray data processing. Probe database could be seamlessly integrated to other microarray data analysis tools, such as GeneSpring. This provides an efficient annotation for mining Arabidopsis RNA profiling data.

Availability: Probe database was implemented in MySQL and the web interface was written in PHP. It is freely available to noncommercial users. Probe database is available over the web at http://get.botany.iastate.edu/probe/

Contact: Dr. Eve Wurtele mash@iastate.edu

INTRODUCTION

The Arabidopsis genome sequencing has been completed [1]. This genome-scale and cDNA sequencing efforts provide the essential information required for the construction of entire genome microarray chips for gene expression studies. Affymetrix GeneChip Arabidopsis Genome Array and Arabidopsis Functional Genomics Consortium (AFGC) EST
Array are two most common-used microarray data sets in *Arabidopsis* research community. Affymetrix *Arabidopsis* GeneChip AtGenome1 Array contains probe sets interrogating more than 8,200 genes and 100 EST clusters of *Arabidopsis*. Eighty percent of the genes represented on the array are predicted coding sequences from genomic BAC entries. Twenty percent of the probe sets represent high quality cDNA sequences [2]. ATH1 Array contains probe sets interrogating more than 22,000 genes [2]. AFGC EST array probes represent more than 14,000 high quality cDNA sequences [3].

The efficient mining of the valuable microarray gene expression data largely depends on the accuracy and integration level of the functional annotations for the microarray probes, because the major and interesting challenge during the data mining process is how to extract the useful information from the expression data and combine it together with what we already know about the genes functions in the different cellular contexts [4]. Unfortunately, the *Arabidopsis* annotation files provided by Affymetrix and AFGC only contain the GenBank identifiers with the very brief GenBank descriptions. It could be extremely tedious and time consuming for biological researchers to search the major databases for the functional annotation for each gene individually only based on the GenBank identifiers. Especially in Affymetrix data set, around eighty percent probes are from predicted genomic BAC sequences. To search GenBank directly by using the GenBank identifiers of these probes, will return the whole BAC records instead of the individual coding regions (CDS).

To address these problems, we are making Probe database available to the *Arabidopsis* research community. Probe database provides an integrated view of the functional annotations for *Arabidopsis* microarray probes: It integrates the associated information from Affymetrix, AFGC, GenBank, TAIR, Swiss-Prot, AraCyc and MetNet DB.
Probe database web interface provides links to these major databases for *Arabidopsis* probes. The *Arabidopsis* Information Resource (TAIR) provides a comprehensive resource for the scientific community working with *Arabidopsis* [5]. Swiss-Prot database is a database of protein sequences produced by the EBI. It contains high-quality manually checked annotation [6]. The AraCyc database is an effort of TAIR to annotate the biological pathways of *Arabidopsis* [7]. MetNet DB is a relational database developed in Iowa State University, for representing and analyzing metabolic and regulatory networks. MetNet DB is a module of Gene Expression Tool kit (GET) software package [8].

PROBE DATABASE

Database Schema

The schema of Probe database is illustrated in Figure 1. There are totally eight tables in Probe database. The detailed explanation could be found on Probe database web site at http://get.botany.iastate.edu/probe/schema/.

Data Processing and Information Sources

The statistical data of Probe Database is summarized in Table 1. The data processing procedures and the information sources for each individual table are described in details as following.

**Affy_AC table:**

It is a table to save GenBank annotation for Affymetrix AtGenome1 data set. The information in this table is based on the accession numbers in Affymetrix annotation file. A Python program is developed to retrieve and process GenBank records for Affymetrix data
set. BioPython GenBank module is used in this program [9]. The data processing steps are all performed on the fly. The program takes the Affymetrix annotation file as input. For each individual probe record, firstly, the program retrieves and parses the whole corresponding GenBank record based on the GenBank identifier in Affymetrix annotation file; secondly, based on the parsed result, the program determines if the probe is from a BAC or EST sequence record. If the probe is from a BAC record, further processing is performed: the locus ID in the Affymetrix annotation file is used to identify the appropriate coding region (CDS). Finally all functional related information, including sequence similarity notes in GenBank is output. The final output result is saved in a tab-delimited text format.

**AffyATH1_AC table:**

It is a table to save the Affymetrix ATH1 Array annotation. The Probe Description is from Affymetrix NetAffy website.

**Affy_Tair, AffyATH1_Tair and AFGC_Tair tables:**

They are tables to save the cross-references of Affymetrix AtGenome1, ATH1 and AFGC probes with TAIR protein data set. The information in the tables is based on the results of sequence similarity comparison. The translated Blast (blastx) is performed. The sequences which Affymetrix AtGenome1, ATH1 and AFGC probes are designed from are used to search against TAIR protein data set (the file ATH1.pep.08102001.total, which is downloaded from TAIR FTP site). NCBI standalone Blast program is used. The threshold of Blast expected value is e-5. The Blast result is processed with a Python program.

**Swiss_Tair table:**

It is a table to save the cross-reference of Swiss-Prot Arabidopsis protein data set with TAIR protein data set. The information in the tables is based on the result of sequence
similarity comparison. The standard protein Blast (blastp) is performed. *Arabidopsis* Swiss-Prot sequences (31,308 sequences, retrieved by using SRS6 system at http://srs.ebi.ac.uk/) are used to search against TAIR protein data set (the file ATH1.pep.08102001.total, which is downloaded from TAIR FTP site). NCBI standalone Blast program is used. The threshold of Blast expected value is e-5. The Blast result is processed with a Python program.

**TairAnnotation table:**

It is a table to save the TAIR annotations for TAIR protein data set. It is based on the file ATH1.pep.08102001.total, which is downloaded from TAIR FTP site.

**TairProperty table:**

It is a table to save the computed properties of TAIR protein data set. The properties include molecular weight, PI, predicted sub-cellular location and the number of trans-membrane domains. The file containing the computed properties is downloaded from TAIR FTP site.

**Pathway_Tair and Pathway tables:**

They are tables to save the cross-references of AraCyc pathways with TAIR protein data set. These two tables are derived from the results of SQL queries to MetNet DB. MetNet DB contains 163 AraCyc pathways details.

**WEB INTERFACE UTILITIES**

The web interface of Probe database provides useful utilities for microarray data procession. One query example is illustrated in Figure 2. The special features of Probe database web interface are summarized as following:
Utility to generate the user-defined Probe list with annotation links in Excel spreadsheet format

Currently lots of biological researchers are working with the microarray experiment data in Excel spreadsheet format. It could be time-consuming to query each individual probe through Probe database web interface one by one. Given a user-defined probe list, Probe database web utility could generate an Excel file for the probe list with the hypertext links to the functional annotation pages. In this way, we could access Probe annotation pages directly from Excel files. This utility usage details could be found on Probe database web site.

Utility to generate the probe list for all genes involved in a particular pathway

If the Probe database query result shows the queried probe is involved in one pathway, the web utility provides the option to generate the probe list for all genes involved in the particular pathway. The generated probe list could be easily loaded into microarray analysis tools, such as GeneSpring. In this way, we could monitor the gene expression level changes for all genes involved in one particular pathway easily. By using this utility, more sophisticated machine learning approaches could be developed for the novel pathway inference from RNA profiling data.

Utilities for further Data Processing

Probe database provides links to the other databases utilities pages, such as the Arabidopsis annotation evidence pages on TAIR, Nice-Prot pages for Swiss-Prot records, graphical illustrations of AraCyc Metabolic pathways etc. This provides the possibility of
further data processing. One could perform the motif finding, sequence similarity searching and 2D gel region prediction etc further data processing by using these utilities.

INTEGRATION OF PROBE DATABASE WITH GENESPRING

Probe database could be integrated to other microarray analysis tools, such as GeneSpring. GeneSpring supports the customized external web database links in version 4.2.1 and later versions. By successful configuration, one could access Probe database Arabidopsis probe functional annotation directly from GeneSpring Gene Inspector. Figure 3 is the screen shot of GeneSpring with the integration of Probe database. The detailed instruction of GeneSpring configuration could be found at Probe database web site.

SUMMARY

Probe database provides the integrated functional annotations for Arabidopsis microarray probes. With the combination of the other microarray analysis softwares, it provides an efficient tool for mining Arabidopsis RNA profiling data.

REFERENCE


http://www.affymetrix.com/products/arrays/specific/arab.affx


[8] Hailong Zhang, Julie A. Dickerson and Eve S. Wurtele, Unpublished Result


### Table 1: Statistical Data of Probe Database

<table>
<thead>
<tr>
<th>Database/ Data Set</th>
<th>Numbers of Processed Records in Probe Database</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affymetrix AtGenome1</td>
<td>8,297 Probes</td>
</tr>
<tr>
<td>Affymetrix ATH1</td>
<td>22,814 Probes</td>
</tr>
<tr>
<td>AFGC</td>
<td>14,393 Probes</td>
</tr>
<tr>
<td>TAIR</td>
<td>26,156 Gene/Proteins</td>
</tr>
<tr>
<td>Swiss-Prot</td>
<td>31,308 Proteins</td>
</tr>
<tr>
<td>AraCyc</td>
<td>163 Pathways</td>
</tr>
</tbody>
</table>
Figure 1: Probe Database Schema
Query Result:

<table>
<thead>
<tr>
<th>Affymetrix ID</th>
<th>Probe Type</th>
<th>Gene Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>203&gt;35_at</td>
<td>Probe from mBTA</td>
<td>Arabidopsis AT31 gene</td>
<td>for glycerol-3-phosphate acyltransferase</td>
</tr>
</tbody>
</table>

**Figure 2:** Screen Shot of one Affymetrix Probe Query Example

**Figure 3** Screen Shot of GeneSpring Integrated with Probe Database
CHAPTER 4. GENERAL CONCLUSION

SUMMARY

In this thesis, we describe a general data model for representing the metabolic and regulatory networks in a living cell. The key contribution of the data model is the general rules that it provides for associating the individual biological entities and interactions into large complex networks of cellular processes. MetNet DB, which implements this model, currently handles information on metabolic pathway and gene regulation. The web interface of MetNet DB provides a flexible pathway-building tool for the biological research community. Based on the information derived from MetNet DB, Probe database is developed to provide the functional annotations for *Arabidopsis* microarray probes.

FUTURE WORK

MetNet DB data model is still evolving as our experience grows and the types of data we handle expands. With the accumulation of pathway and interaction data in MetNet DB, the advanced query functionalities are needed for efficient processing pathway data. As a modeling tool, FCModeler needs to deal with the more complex cases. Additionally, Probe database web interface provides the utility to generate the probe list for all genes involved in one particular pathway. By using this utility, more sophisticated machine learning approaches could be developed for the novel pathway inference from RNA profiling data.