

GELIM: An Integrated System with Genetic Network Analyzer and LIMS

GELIM: 유전자 네트워크 분석과 데이터 관리를 위한 통합 시스템

Kyung-Shin Lee^{1*}, Hye-Jung Kim¹, Hwan-Gue Cho¹, Seon-Hee Park², Miyoung Shin², Ho-Youl Jung²

¹ Department of Computer Engineering Pusan National University, Korea

² Bioinformatics team, Electronics and Telecommunications Research Institute, Korea

*To whom correspondence should be addressed. *E-mail: kslee@pearl.cs.pusan.ac.kr

Abstract

생물학적으로 의미 있는 결과를 도출하기 위해서는 많은 실험 데이터가 필요하다. 최근에는 마이크로 어레이 실험 기술이 발달함에 따라 대량의 데이터를 얻을 수 있게 되었고, 이로 인해서 데이터를 체계적으로 관리하고 필요한 정보를 습득할 수 있는 시스템이 필요하게 되었다. LIMS(Laboratory Information Management System)는 이러한 요구 조건을 충족시키기 위한 시스템으로 기존의 파일 시스템에 의존해서 비효율적으로 실험 데이터를 관리해 오던 것을 체계적이고 효율적으로 관리해 주기 위한 시스템이다. 대량의 유전자 발현 데이터의 생산은 유전자의 조절 네트워크 예측을 가능하게 하였다. 유전자간의 상호 작용을 분석하는 것은 세포의 활동을 이해하는데 매우 중요한 요소라고 할 수 있다. 본 논문에서는 기존의 LIMS 기능과 유전자 조절 네트워크 분석 시스템을 통합하여 사용자가 쉽게 데이터를 공유 및 습득할 수 있으며 편리한 사용자 인터페이스를 이용하여 컴퓨터에 익숙하지 않은 실험들도 쉽게 사용할 수 있는 GELIM(an Integrated system with GENetic network analyzer and LIMS)을 소개한다.

서론

생물 정보학의 최종 목표는 생물 체계를 이해하고 질병의 원인을 파악하는 것이며,

더 나아가서는 신약 개발 등 치료법을 연구하는 것이다. 이러한 목적을 이루기 위해서는 실험을 하고, 데이터를 획득하고, 획득한 데이터를 축적하는 과정이 필요하다. 축적된 데이터를 기반으로 분석하여 생물학적인 의미를 도출하는 단계로 연구는 진행된다.

This work is supported by ETRI(Electronics and Telecommunications Research Institute)

생물학적으로 의미 있는 결과를 도출하기 위해서는 많은 실험 데이터가 필요하다. 분석할 수 있는 데이터의 양이 적으면, 실험 오류와 같은 정보에 따라 결과가 달라질 수 있기 때문에, 다양한 조건을 기반으로 실험을 하고 그 실험 결과를 토대로 분석을 시행하여야 한다. 실험 결과가 축적될수록 실험 정보 및 데이터를 관리하는 것이 어렵게 되고, 기존의 파일 시스템에 의해서 관리하면 필요한 정보를 검색하거나 정보를 공유하는데 많은 어려움이 있다. 따라서, 실험에서부터 분석까지의 다양하고 많은 데이터들을 체계적으로 관리, 저장할 수 있는 LIMS(Laboratory Information Management System)를 사용하여야 한다.

축적된 데이터를 이용하여 생물학적으로 의미 있는 결론을 도출하기 위해서는 데이터를 분석해야 한다. 예전에는 생산되는 실험 데이터의 양이 많지 않았으므로 수동으로 진행되었으나, 현재는 실험 기술이 발달하여 수동으로 진행할 수 없는 상태이다. 그러므로 대량의 데이터를 정확하고 빠르게 분석할 수 있는 분석 시스템이 필요하며 이러한 시스템을 이용하면 개발 시간과 개발 비용이 감소한다.

유전자의 발현은 유전자의 상호작용에 의해서 발생하기 때문에 발현 데이터를 이용하여 유전자들의 상호작용을 예측할 수 있다. 유전자들의 상호 작용을 분석하는 것은 유전자 레벨에서의 세포 활동에 대한 정보를 제공하고, 질병의 원인을 예측하는 데 많은 도움을 줄 수 있다.

본 논문에서는 LIMS와 유전자 조절 네트워크 분석 시스템을 통합함으로써 데이터를 관리함과 동시에 축적된 데이터를 이용하여 조절 관계를 분석할 수 있는 시스템

GELIM(an integrated system with GENetic network analyzer and LIMs)을 소개한다.

GELIM 소개

GELIM은 LIMS와 유전자 조절 네트워크 분석 시스템이 통합된 시스템이다. 사용 권한이 있는 사용자는 웹으로 접속하여 데이터를 저장, 관리할 수 있으며 저장된 데이터를 이용하여 유전자 조절 네트워크를 분석할 수 있다. 유전자 조절 네트워크 분석 시스템은 자바(JAVA)를 기반으로 하여 독립적으로 개발되었으나 웹에 접속하여 사용할 수 있다. Java Web Start를 이용하여서 연결되었기 때문에 별도의 설치과정을 거치지 않고 사용할 수 있고, 버전이 업그레이드 되면 사용자는 추가 설치 없이 항상 최신 버전의 유전자 조절 네트워크 시스템을 사용할 수 있도록 하였다.

데이터 저장 및 관리 기능

GELIM은 실험 정보 및 데이터를 저장 관리하는 기본적인 LIMS 기능뿐만 아니라 가시화 기능과 간단한 분석 기능도 제공한다. 제공되는 기능은 다음과 같다.

■ 데이터 단위

GELIM은 기본적으로 웹 기반의 생물학 실험에서 나온 데이터의 저장 기능을 지원한다. 데이터를 효율적으로 관리하기 위해 Project, Experiment, Work, Shot의 4단계로 나누어서 관리를 한다.

그림 1에서 보는 바와 같이 Project는 생물학자들이 의미 있는 결과를 얻기 위하여 행하는 일련의 실험 과정으로 하나 또는 그 이상의 Experiment의 집합

이다. 즉, 다양한 실험 조건으로 실험한 Experiment 집합을 의미한다. Experiment는 마이크로 어레이 디자인의 구조는 같고 실험 조건을 다르게 한 Work의 집합이다. 즉 똑같은 어레이 디자인에 같은 유전자를 심어놓고 열, 온도 등과 같은 조건을 달리하여 실험한 Work의 집합이다. Work는 실험 조건이 같고 마이크로 어레이 실험에서 사용한 어레이 디자인이 같은 이미지들의 집합이다. Shot은 마이크로 어레이 데이터의 가장 작은 단위로 물리적으로 마이크로 어레이 실험에서 얻어낸 적색 및 녹색을 가진 이미지를 나타낸다. 만약 벼의 생장에 관한 실험을 한다고 했을 때 Project는 벼의 생장에 관한 실험을 계획하는 단계에 필요한 정보를 저장한다. Experiment는 각각의 실험 조건에 따라 나뉘지게 되는데 열, 온도, 물의 양에 따라 벼의 생장이 어떻게 변화하는가에 따라 나눌 수 있다. 만일 물의 양에 대한 실험을 한다면 Experiment에 이에 대한 정보가 저장된다. Work의 경우 만일 Work가 물의 양에 따른 Experiment에 속하는 경우 10cc의 물, 20cc의 물, 30cc의 물과 같이 각각 물의 양을 달리하여 실험한 결과를 저장한다. 10cc의 물에 따라 실험한 Work에 속하는 Shot의 경우 이 실험에서 얻어진 이미지를 저장하게 되는 것이다.

마이크로 어레이에서 생물학자들이 의미 있는 결과를 도출하기 위해서 마이크로 어레이 실험, 이미지 분석, 통계 분석, 생물학적 의미 도출의 단계를 거친다. 이러한 작업은 서로 다른 연구자들 간의 의사소통과 긴밀한 협력을 바

탕으로 이루어져야 한다.

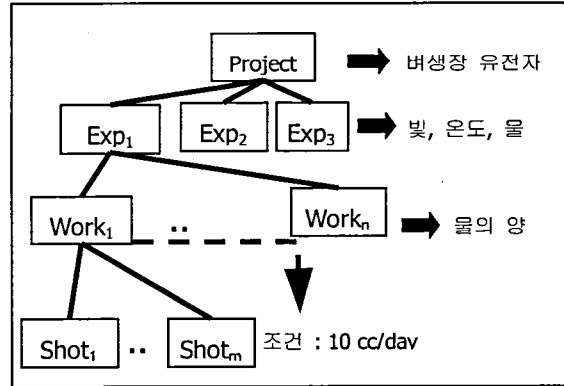


그림1 계층적 데이터 단위

이는 서로 다른 작업 환경과 배경을 가진 연구자들 사이에서 혼란이 야기될 수 밖에 없다. 이러한 문제를 해결하기 위해서 하나의 의사소통 공간을 제공한다. 이를 위해 각각 이미지 분석, 통계 분석, 생물학적 분석을 통한 결과 파일을 업로드 할 수 있는 기능을 제공하고 있다. 이러한 기능은 각 분야의 연구자들이 작업의 진척도와 일정들을 일일이 연락해서 확인해야 하는 불편함과 시간적 낭비들을 해결해 줄 수 있다.

■ 마이크로 어레이 이미지 가시화

마이크로 어레이 실험으로 나오는 이미지 데이터와 이미지 데이터 분석 파일을 통해 마이크로 어레이 실험에 대한 결과를 도출하게 되므로 이미지 데이터에 대한 가시화 기능이 요구된다. GELIM에서는 이를 위해 이미지를 보여주는 기능뿐만 아니라 한 이미지에 구성된 유전자들의 각 채널의 발현 정도에 따른 Histogram 및 spot plot을 제공한다. 또한 이러한 발현 정도에 대한 변화를 비교하기 위해 시간에 따른 유전자 발현 정보의 변화와 반복 실험에서 시

간에 따른 유전자 발현 정보의 변화에 대한 그래프를 보여준다.

■ 메타 파일 생성

마이크로 어레이 실험과 분석을 통해 여러 개의 데이터가 만들어지면 생물학자들은 이러한 데이터의 조합을 통해 최종적으로 생물학적인 의미를 도출하게 된다. 예를 들어 특정한 유전자의 시간대별 발현 양상을 관찰하고자 하는 경우 각 시간대별 파일을 모두 확인해야 하는 번거로움이 있다. 그리고 실험의 정확성을 위해서 반복 실험을 한 경우 반복 실험 당 만들어지는 분석 파일들에서 붉은색과 녹색의 비율을 모두 살펴보고자 하는 경우에도 여러 파일을 모두 찾아서 보아야 한다. 본 시스템에서는 위와 같은 문제를 극복하고자 사용자가 원하는 데이터의 내용을 조합하여 메타 데이터를 생성하고 사용자가 쉽게 데이터를 분석할 수 있도록 하였다.

■ 데이터 검색

마이크로 어레이 데이터는 엄청나게 많은 이미지와 분석 데이터로 이루어져 있으므로 여기서 의미 있는 결과를 도출하기 위해서는 데이터 필터링과 검색이 필수적이다. GELIM에서는 데이터 관리를 위한 계층(Project, Experiment, Work)에 따라 데이터를 검색할 수 있도록 할 뿐만 아니라 유전자의 이름, 이미지 분석 결과 파일에서 특정 발현 강도 이상의 스팟들을 검색하여 보여줌으로써 사용자가 편리하게 생물학적인 의미를 도출할 수 있도록 도와준다.

■ 데이터 별 백업 & 복구

마이크로 어레이는 데이터의 양이 많

을 뿐만 아니라 한 번의 실험에 많은 돈이 들기 때문에 이 데이터를 잃을 경우 손실이 크다. 그러므로 이러한 데이터를 Project단위로 백업 및 복구할 수 있는 기능을 제공하고 있다.

■ 해외 데이터베이스 결과 참고

마이크로 어레이 데이터로부터 의미 있는 결과를 도출하기 위해 마이크로 어레이 이미지를 분석해서 나온 유전자에 대해 이미 밝혀져 있거나 연구되고 있는 내용을 참고하면 분석에 도움이 될 수 있다. 이를 위해 NCBI 에서 제공하는 Entrez[17], UniGene[19], PubMed[18]와 같은 외부 데이터베이스에 연결하여 본 시스템에 등록된 특정 유전자에 대하여 이미 공개된 데이터 정보를 얻을 수 있도록 하였다.

■ Quality Control 의 결과 저장

마이크로 어레이 실험의 정확도를 높이기 위해 같은 내용의 마이크로 어레이 실험을 중복하여 그 결과들을 비교하게 된다. 똑같은 마이크로 어레이 디자인으로 반복 실험한 결과들을 실험시 제작한 마이크로 어레이 디자인 정보를 이용하여 평가하면 실제 결과를 내는데 도움을 줄 수 있다. 따라서 각 실험에 대한 마이크로 어레이 Quality Control 의 결과를 저장하여 도움을 줄 수 있도록 하였다.

■ 데이터 공유를 위한 표준 포맷 사용

실험자간에 데이터를 공유하기 위해서는 표준화된 데이터 포맷이 필요하다. MGED[22]에서는 마이크로 어레이 실험 데이터를 공유하기 위해 필요한 최소한의 정보인 MIAME

(Minimum Information About Microarray Experiment)[16]를 정의하여 이러한 문제를 극복하고자 노력하고 있다. 본 시스템은 MIAME 를 이용하여 실험자간 데이터 공유를 쉽게 할 수 있도록 하였다.

유전자 조절 네트워크 예측 기능

유전자 조절 네트워크란

세포의 활동은 유전자 하나의 독립적인 활동에 의해서 일어나는 것이 아니라 유전자 또는 유전자의 활동에 의한 산물인 단백질 또는 물질 대사간의 상호작용으로 일어난다. 그러므로 생명현상을 이해하기 위해서는 이들간의 조절 체계를 이해하는 것이 매우 중요한 문제이다. 세포활동에 관여하는 상호작용은 세가지 레벨로 나누어진다. 첫 번째는 유전자 레벨(mRNA 레벨)에서 발생하는 유전자 네트워크(Genetic Network)이고, 두 번째는 단백질 레벨에서 발생하는 단백질 상호작용(Protein-Protein Interaction)이며 세 번째 레벨은 물질대사간의 상호작용인 대사 경로(Metabolic Pathway)이다. 그림 2는 세포내의 전체 상호작용을 나타내고 있다. 그림 2에서 알 수 있듯이 유전자간에는 직접적인 상호작용은 존재하지 않는다. 그러나 유전자 활동의 산물인 단백질이나 대사 산물에 의해서 직접적으로 영향을 받는다. 그러므로 유전자간에는 간접적인 상호작용이 존재하는 것으로 볼 수 있다. 그림 2의 실선은 직접적인 상호작용을 나타내며 그 영향력이 활성화 시키는 것인지, 비활성화 시키는 것인지에 대한

표시는 나타나지 않고 있다. 점선은 유전자 레벨에서 간접적으로 나타나는 상호작용을 보여주고 있으며, 이것이 유전자 네트워크이다[1].

유전자 네트워크를 알게 되면 세포내의 전반적인 조절 메커니즘을 이해할 수 있게 되고 이를 기반으로 유전자, 단백질의 기능을 예측할 수 있다. 더 나아가서는 질병의 원인을 파악하고 그에 따른 치료법이나 신약을 개발할 수 있기 때문에 연구 가치가 매우 높다.

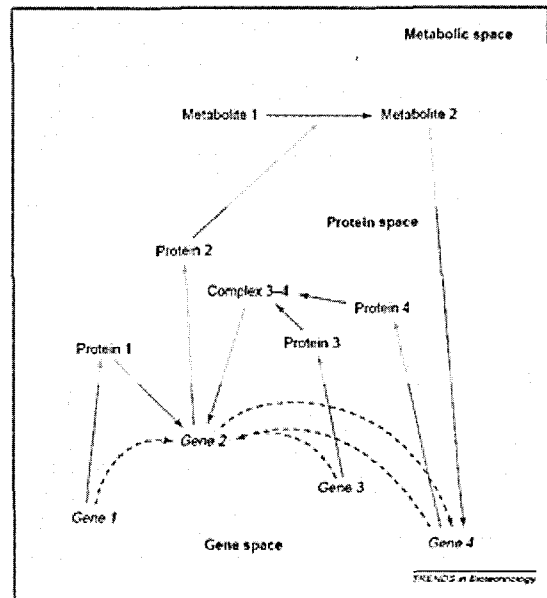


그림 2 세포내의 상호 작용

전처리 단계

전처리 단계는 정확도 높은 유전자 네트워크를 예측하기 위해서 필요한 여러 가지 정보 및 분석을 제공하는 단계로써 세부적인 기능은 다음과 같다.

■ 입력 데이터 보정

마이크로어레이 실험 데이터의 경우, 실험 시 발생한 노이즈나 이미지 분석

시 발생한 오류로 인한 미성값을 보정해준다. 보정하기 위한 방법으로는 KNN(K-Nearest Neighbor)알고리즘[3]을 기반으로 하여 가장 가까운 값, 가장 가까운 K 개 값의 평균, 우선 순위를 적용한 K 개의 평균을 이용하여 보정하는 방법을 제공한다

■ 데이터 도메인(domain)변환

이미지 분석 결과로 나오는 데이터는 실수의 형태로 출력된다. 실수 형태의 데이터를 이용해서도 분석할 수 있지만 데이터로부터 의미 있는 정보만을 추출하여 분석하면 데이터를 단순화 함으로써 노이즈에 강한 장점이 있다. 그러므로 데이터를 이진화[4], 삼진화로 변형하여 데이터를 단순화 시키는 기능을 제공한다.

■ 패턴 가시화 및 쿼리 처리

텍스트 형식의 데이터는 직관적으로 인식하기 어려우므로 특징을 한눈에 파악하기 어려운 단점이 있다. 그래서 사용자가 관심 있는 유전자를 선택하여 패턴을 다양한 색상으로 볼 수 있으며 유사도 분석을 통해서 관심이 있는 유전자와 유사도가 높은 유전자 등을 검색할 수 있다. 유사도를 측정하기 위한 방법으로는 City-Block distance, Maximum distance, Euclidean distance 를 제공한다.

$$\text{City-Block distance} : d = \sum_{i=1}^n |x_i - y_i|$$

$$\text{Maximum distance} : d = \max |x_i - y_i|$$

$$\text{Euclidean distance} : d = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

■ 데이터 추출 기능

사용자는 특정 유전자 리스트만

연구하거나, 특징이 있는 소수의 유전자들만 골라서 분석할 수 있다. 이러한 기능을 편리하게 제공하기 위해서 사용자의 리스트로부터 자동으로 유전자의 발현 데이터를 추출하여 입력 데이터를 생성해 주는 기능과 입력 데이터에 있는 유전자를 체크하여 새로운 입력 파일을 생성해 주는 기능을 제공하고 있다.

■ 클러스터링 분석

클러스터링이란 특징이 유사한 개체들을 그룹화 하는 것을 말한다. 여기서 특징이란 유전자의 발현 양상을 말하는데, 이것은 발현 양상이 유사할수록 유사한 기능에 속할 것이라는 가정을 기반으로 한다. 제공하는 클러스터링 방법으로는 계층적 클러스터링, K-Means 클러스터링, Graph Based 클러스터링, SOM(Self-Organizing Map) 클러스터링이 있다.

네트워크 생성 단계

유전자 조절 네트워크를 예측 하기 위해서는 전처리 단계를 거친 데이터를 입력으로 하여 조절 관계를 나타내는 그래프 형태의 결과를 도출한다. 분석을 위해서 사용하는 모델은 베이지안 네트워크 모델이다. 베이지안 모델은 부분적으로 직접적인 연관성이 있거나 의존성이 있는 데이터들의 관계를 잘 나타내며, 통계적인 이론을 토대로 모델의 계산 알고리즘 등이 많이 연구되어 있기 때문에 유전자들의 조절 관계를 잘 나타내어 줄 것으로 기대되고 있다.

분석 방법은 데이터의 양과 사용자가 지정한 분석 형태에 따라서 다른 방법으로 수행된다. 유전자 조절 네트워크를 예측하는 것은 계산 복잡도가 높은 방법이고, 분석할 데이터가 아주 많은 경우에는 오랜 시간이 필요하다. 그러므로 탐색할 양이 많은 경우에는 휴리스틱한 유전자 알고리즘을 이용하여 최적에 가까운 결과를 보여주며, 탐색할 양이 적은 경우에는 모든 경우를 탐색하여 최적의 해를 찾아준다.

가시화 단계

가시화 기능은 분석 데이터의 양이 많을수록 그 중요성이 높아지는 기능이다. 데이터의 형태로는 예측된 조절 네트워크의 특징을 잘 파악할 수 없기 때문이다. 유전자 조절 네트워크는 그래프와 동일한 형태를 가지고 있기 때문에 각 유전자는 그래프의 노드로, 조절 관계는 에지로 표현된다. 이때, 발현을 활성화 시키는 조절은 붉은색의 실선, 발현을 억제시키는 조절은 파란색의 점선으로 표시하여 쉽게 구분되도록 하였으며 이러한 속성은 편집 가능하다.

가시화 단계는 yWorks (<http://www.yworks.com>)에서 개발한 자바 기반의 라이브러리인 yFiles 를 사용하여 개발하였다. 가시화 단계에서 제공하는 기능은 다음과 같다.

■ 조절 네트워크 가시화 및 편집

입력 데이터로부터 유전자 조절 네트워크를 가시화 해 주며 유전자와 조절 관계를 나타내는 색상, 모양 등의 편집 기능과 네트워크의 확대, 축소

기능을 제공한다. 분석할 유전자의 수가 많을수록 가시화에서 중요한 기능이 레이아웃이며, 레이아웃이 잘 되지 않은 경우에는 네트워크로부터 정보를 쉽게 습득할 수 없는 문제점이 발생한다. 본 시스템에서는 Circular, Orthogonal, Hierarchical 등의 다양한 레이아웃을 제공하고, 이렇게 얻어진 이미지는 JPEG, GIF 등의 포맷으로 변환이 가능하다.

■ 외부 생물 데이터 베이스와의 연결

분석하는 유전자에 대한 생물학적인 정보를 제공하기 위해서 생물 데이터 베이스와 연결하였다. 관심 있는 유전자를 선택하고 데이터 베이스를 선택하면 자동으로 연결되어 해당 유전자의 정보를 찾아준다. 연결 가능한 데이터 베이스로는 GeneBank, Entrez, PubMed 가 있다.

■ 조절 네트워크 확장 기능

여러 개의 결과 파일로부터 하나의 조절 네트워크를 생성하는 네트워크 확장 기능을 제공한다. 이 기능은 각각 분석한 데이터에서 공통된 유전자를 이용하여 통합하는 것으로 전체 네트워크를 구성할 때 유용한 기능이다.

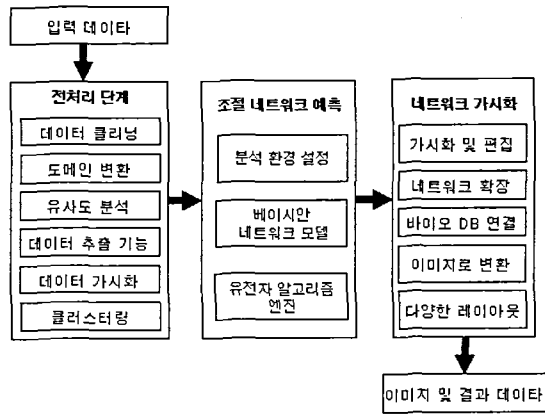


그림3 유전자 네트워크 분석 시스템의 구조

타 시스템과의 비교

마이크로어레이 실험에서 나오는 데이터를 효율적으로 관리하기 위해 여러 데이터베이스 시스템이 존재하고 있다. 이러한 시스템들은 기본적으로 데이터의 저장, 관리, 검색하는 기능을 제공하며 부가적으로 데이터를 분석할 수 있는 기능들을 제공하고 있다.

Acuity[21]는 Axon사에서 만든 상용 LIMS로 Windows에서 실행할 수 있는 응용 프로그램이다. 데이터의 저장과 관리뿐만 아니라 데이터의 보안 기능과 리포트 기능, 다양한 입출력 포맷을 지원한다. 또한 clustering, ANOVA, PCA(principle component analysis)와 같은 다양한 분석 기능을 제공하고 있어 저장되어 있는 데이터에 대해 많은 분석을 해볼 수 있다.

BASE[13]는 Lund대학에서 개발한 웹 기반의 시스템으로 각종 데이터, 이미지, 유전자 발현 데이터의 관리 및 plug-in을 통한 정규화 기능, 데이터 가시화, 분석 기능이 통합되어 제공된다.

이 두 시스템과 GELIM의 기능을 비교해

보면 표1 과 같다.

기능	BASE	Acuity	GELIM
계층적 관리	○	○	○
백업 및 복구기능	X	○	○
메타파일 생성	○	X	○
반복실험 처리	○	○	○
정규화 기능	○	○	X
정보검색 기능	○	○	○
3D 가시화	○	○	X
외부 DB 연결	○	○	○
이미지 분석	○	○	X
조절네트워크분석	X	X	○
레포팅	X	○	○
사용자관리, 보안	○	○	○
웹 기반	○	X	○

표 1 타 시스템과 GELIM과의 기능 비교

BASE와 Acuity, GELIM 모두 데이터를 효율적으로 관리하기 위해 계층적으로 분류해서 데이터를 관리한다. 모두 기본적인 검색 기능과 반복 실험에 대한 데이터 관리기능을 제공하고, 현재 등록되어 있는 데이터에 대해서 NCBI의 데이터를 검색한 페이지를 링크해두었다. 또한 사용자의 관리를 위해 그룹으로 나뉘어서 관리하거나 사용자의 권한에 따라 계층으로 나누어 관리하고 보안을 위해서 각 데이터에 대한 접근 권한을 설정할 수 있도록 되어 있다. BASE는 데이터의 백업 및 복구 기능을 자체 프로그램에서 제공하는 것이 아니라 BASE에서 사용하고 있는 데이터베이스인 MySQL의 기능에만 의존하고 있어 이 부분에서 취약점으로 보이고 있다. 또한 현재 등록된 데이터에 대한 등록 상태를 볼 수 있는 레포팅 기능을 제공하지 않는다. Acuity는 여러 데이터들을 한 눈에 비교해볼 수 있도록 분석 데이터를

조합해서 만드는 메타 파일 생성 기능을 제공하지 않고 있어 불편하며, 웹 기반의 프로그램이 아니라 Windows에서만 동작하는 응용 프로그램으로서 Linux 사용자가 사용하는데 불편한 단점이 있다. GELIM은 데이터의 저장, 관리 및 검색 기능을 제공하고 있지만 정규화나 이미지 분석과 같은 기본적인 분석 기능을 제공하고 있지 않다. 하지만 유전자 조절 네트워크 분석을 할 수 있는 프로그램을 제공하여 유전자 조절 네트워크를 구성하여 유전자의 기능을 예측할 수 있는 기능을 제공하고 있다.

결론

실험 데이터의 양이 늘어남에 따라서 이를 체계적으로 관리하기 위한 LIMS의 필요성이 더욱 확대되었다. GELIM은 대량의 데이터와 실험 정보를 체계적으로 관리하는 웹 기반의 시스템으로써 여러 사용자들이 정보를 공유할 수 있고, 이로 인해서 불필요한 중복 실험을 방지할 수 있다는 장점이 있다. 또한 편리한 인터페이스를 제공함으로써 컴퓨터에 익숙하지 않은 사용자들도 쉽게 사용할 수 있다. 축적된 데이터는 유전자 조절 네트워크 분석 시스템을 이용해서 유사도 측정, 클러스터링, 조절 네트워크 예측 등의 다양한 분석을 할 수 있으므로 실험실에서 유용하게 사용할 수 있다. GELIM에 대한 정보는 <http://neobio.cs.pusan.ac.kr:8080/gelim> 에서 볼 수 있다.

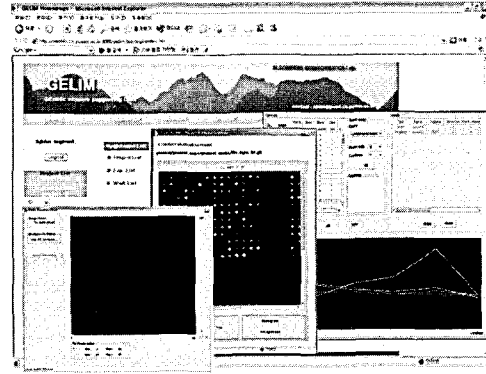


그림 4 GELIM의 실행 화면

Acknowledgements

This work is supported by a grant(2010-2004-0014) from Electronics and Telecommunications Research Institute.

References

- [1] Paul Brazhnik and Alberto de la Fuente and Pedro Mendes, Gene networks:how to put the function in genomics, *TRENDS in Biotechnology*, 2002.
- [2] Jiang Qiu and Jimmy Lin et al., Prediction of regulatory networks : genome-wide identification of transcription factor targets from gene expression data, *Bioinformatics*, 2003.
- [3] Olga Troyanskaya and Michael Cantor et al., Missing value estimation methods for DNA microarrays, *Bioinformatics*, 2001.
- [4] Ilya Shmulevich and Wei Zhang, Binary analysis and optimization-based normalization of gene expression data, *Bioinformatics*, 2002.
- [5] Nir Friedman and Michal Linial et al., Using Bayesian Networks to Analyze Expression

- Data, *Journal of Computational Biology*, 2000.
- [6] Tong I. Lee et al., Transcriptional Regulatory Networks in *Saccharomyces cerevisiae*, *Science*, 2002.
- [7] Hitoshi Iba and Atsushi Mimura, Inference of a gene regulatory network by means of interactive evolutionary computing, *Information Sciences*, 2002.
- [8] Jiang Qian and Marisa Dolled-Filhart et al., Beyond Synexpression Relationships: Local Clustering of Time-shifted and Inverted Gene Expression Profiles Identifies New, Biologically Relevant Interactions, *Journal of Molecular Biology*, 2001.
- [9] Ziv Bar-Joseph and Georg K Gerber et al., Computational discovery of gene modules and regulatory networks, *Nature biotechnology*, 2003.
- [10] Haiyuan Yu and Nicholas M Luscombe et al., Genomic analysis of gene expression relationships in transcriptional regulatory networks, *TRENDS in Genetics*, 2003.
- [11] 이경신, Genetic Algorithm을 이용한 유전자 조절 네트워크 구성하기, *Technical Report*, 2003.
- [12] 이경신, yFiles를 이용한 그래프 가시화, *Technical Report*, 2004.
- [13] Lao H. Saal, Carl Troein, Johan Vallon-Christersson et al., BioArray Software Environment: A Platform for Comprehensive Management and Analysis of Microarray Data, *Genome Biology*, 2002.
- [14] Pascale Anderle and Manuel Duval and et al., Gene Expression Databases and Data Mining, *BioTechniques*, 2003.
- [15] Paul T. Spellman, Design and implementation of microarray gene expression markup language(MAGE-ML), *Genome Biology*, 2002.
- [16] Alvis Brazma, Minimum information about a microarray experiment(MIAME) - toward standards for microarray data, *Nature*, 2001.
- [17] Entrez cross-data search.
<http://www.ncbi.nlm.nih.gov/Entrez/>
- [18] Entrez PubMed.
<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed>
- [19] UniGene.
<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene>
- [20] Microarray Software Comparison,
http://ihome.cuhk.edu.hk/~b400559/arraysoft_database.html
- [21] Acuity Microarray Analysis, Visualization, and Database software,
http://www.axon.com/gn_Acuity.html
- [22] MGED, <http://www.mged.org>