BIOINFORMATICS TOOLKIT FOR CELLULAR ACTIVITIES IN BIOLOGICAL SYSTEMS

Jayanthi Manicassamy & P. Dhavachelvan

Cells are complex systems which are networked together, residing in biological systems. A number of metabolic reactivity that occurs within the cells, which in turns the cells response. Improper cells signaling are the main causes for disease and disorders in biological systems so, it is necessary to understand the systematic organization of cells and its activities. There is lack of formal guidelines and constructive opposition to cell research which helps researchers for further disease identification and in drug discovery. Since, scientists have not yet been able to fully map cells and its functionalities, which are very sensitive and under research this made systems biology researchers in cells to lack behind. In this perspective, understanding and representing the entire activities of the cells at system level involving, cells networking pathways and its signaling is a novel way that have been incorporated in this paper. Since today there is a necessity for development of high throughput tools for effective and efficient representation of system activities. CABS aims at representing cellular activities in biological systems for cancer has been made. Network connectivity of the cells by means of pathway and signaling are done mainly taking disease reactivity or symptoms into consideration in the body. Internally certain metabolic behaviors have also been consideration for representing cells connectivity and communications. Cellular activities including pathway construction and cells signaling in reactive to the environment to the immune system have also been represented. Utilization of this tool will help the researchers for carrying out further research in this area of system biology.

Keywords: Bioinformatics, Cell Biology, Databases, Information Retrieval, Molecular Biology, Performance Evaluation, Systems Biology, Web-Based Tool.

1. INTRODUCTION

Exploiting the potential of computers to model and understand complex biological systems for which its phenomenon have been tremendously increasing day to day. Initially, the approach used by most bioinformatics tools was more analytical and engineering-like. Recently only system-level analysis has grounded its discoveries at molecular-level with the progress of genome sequence project to accumulate in-depth knowledge of molecular nature of biological system. Today more approach are being under development that accounts intra and inter molecular interactions at the cells, tissues, organs and organism level. Systems biology focuses on systems that are composed of molecular components where, both structure of the system and components plays indispensable role forming symbolic state of the system as a whole. There are numbers of exciting and profound issues that are actively investigated, such as robustness of biological systems, network structures and dynamics, and applications to drug discovery. At present understanding the processing activities in the biological systems is primarily towards reductionism from which discovering of new emergent properties of systemic view could occur.

Diseases and disorders in biological systems are more common around the worlds that are caused due to fault signaling at cells level. Cells that reside in the system are the building blocks of life that not identical for all species which are organized as groups for carrying out specialized common functionality. This specialization occurs because different cell types read out different parts of the DNA blueprint and therefore make different proteins [2]. When hundreds of proteins become organized to form macromolecular it is capable of carrying out the cells functions. Cell signaling which is a part of complex system for communication that governs basic cellular activities and coordinates cell actions. Indeed, cell communication is necessary for the ability of cells to perceive and correctly respond to their environment which is the base for immune responses and as well as normal tissue homeostasis. Errors in cellular information processing are responsible for diseases such as cancer, autoimmunity, and diabetes. There is a need to understand cell signaling enough to treat these diseases effectively and efficiently. Traditional work in biology has focused on studying individual parts of cell signaling pathways. However, there is a need to understand the underlying structure of signaling networks and how changes in these networks can affect the transmission of information.

Biological pathways provide significant insights on the interaction mechanisms of molecules. In the course at

^{*} Department of Computer Science, Pondicherry University, Pondicherry, India. *E-mail: jmanic2@yahoo.com, E-mail: dhavachelvan@gmail.com*

systems level various things should be taken into consideration are extent of genomic pathways conserved among different species, requirement of minimum set of pathways required for all the organisms and the distance between the linking pathways [7]. Constraint-based analysis of metabolic pathway by minimization of metabolic adjustment [4] used in system biology tools to investigate dynamical properties of a biological system in a more quantitative and rational way by gives only the basic level view of an organism's functionality. Due to the complex and incomplete nature of biological data, at the present time, fully automated computational is excessively ambitious. A visual approach can complement the biological analysis which would reduce the time spent by specialists for interpreting result [3] [5] [6] or carrying out further researches. This leads us to develop this bioinformatics toolkit for representing the activities of the biological systems at cellular level in graphical manner which has been represented in this paper. This provides a means for community for visualized representation of internal activities of the cells in accordance to the disease and its reactivity to immune system. The novel identification found from the research made on cell functionalities could be made globally available for the community for their specific utilization purpose. Standards procedures for evaluation [1] have been carried out for representing effectiveness and efficiency of CABS.

2. Methodology

Today diseases and disorders in biological system has become an inevitable one in and around the world. Functionalities of biological systems could be identified from the disease chromosome of the genome and its level of infection of the particular species. Based on the different symptoms along with functionalities that takes place in the biological systems from level of disease contamination cells are modeled. For each symptom pathway is constructed and signaling is made along with functional details that are represented in graphics mode which could be further visualized as movie. The tools processing is primarily developed for making a visualization of human biological system for cancer syndrome which could provides information about the cells functionalities in depth.

Methodologies for carrying out representation for cellular activities in biological system for a particular syndrome through cellular pathway network, signaling and other activities have been represented here. Due to the complexity that avail in the biological systems it is difficult to represent the whole biological system accurately. The development made falls as an appropriate visualized representation of system activity at the level of cells which provides a base for carrying out further researches.

2.1 Genome Processing

Species affected by *Cancer* with the specific type of cancer, its sample genome will be extracted along with related and required information like *causes for cancer*, *symptoms*, *activities*, *cells related information's etc...* from databases like *NCBI*, *PubMed*, *KEGG and Metacyc*. This information extraction is based only the infection level of the species which are used in representing the activities in biological systems at cellular level.

2.2 Graph Construction

Based on the affect species sample genome, *Circular Genome Graph* is constructed through *GenomeVx* and pathway chart of the specific disease for the species is extracted from KEGG database for visualization for researchers. Taking into consideration of the disease level (Low, Medium and High) changes in functionalities or symptoms in the system occurs. Flexibility of changing the disease level can be made according to the user's requirement.

2.3 Cellular Representation

Disease and its symptoms along with functionalities in the biological systems cells level are taken into consideration for constructing Biological cellular network and signaling. Cellular networking construction currently involves disease level along with symptoms and functionalities in the biological system. Here nodes are cells, edges are pathways and signaling is carried out between edges between nodes. Some pathways in the network might or might exist for signaling passing. Specific types of cells in the specific subsystems along with the required cells that perform relevant functionalities are grouped together as a single sub cellular network. Cells status in human system has been represented in Figure 1 in symbolic form which denotes a particular nodes (Cells) status. A change in the status of the cells depends upon disease level or in reactive to effectors like external environment, molecule interaction, etc..



Figure 1: Cells Status Representation

Normal cells status represents whatever the level of disease, the cells properly activates. Infected refers to cells that are practically infected where few improper activity take place that are hard to identify. Defective cells that don't perform the activity that should be actually perform. Defective cells could also be represented as fully infected cells. In the construction of cellular network two major things that have been considered. The first things considered are the sub-system cells in which priority is given to immune system cells. The second is based on symptoms along with functionalities in the system which is also considered for cells signaling. Apart from this if the level of disease found to be low then infection in cells found to be low and defective cells found to be null or approximately 1%-2% of the infected cells. 25%-40% of the cells found be infected of which partially found to be defective for disease level found to be medium. While defective cells found to be more of which few are found to be infected for high disease level. The representation of cellular network and signaling is only a blue print of the cells activates in the human system.

Immune system is responsible for protecting its host from pathogens and other foreign substances which also involved in responding to the antigens that are injected to the system. This was the reason behind for giving first priority to the immune systems cells and its activities. Cells of the immune system are always in active statues. System has 30% of which 25% RBC and 3% WBC of the 100 Trillions cells in the whole adult human body. Here we have considered three adaptive immune system cells B-cells, T-cells and NK cells of which B-cells and T-cells are considered to be main. Normal human immune system without and infection consists of 30% of B-cells and 70% of T-cells based on this, disease infection level, cellular network representing immune system cells will be constructed. Based on IMC infected immune system, cells in the whole cellular network have been represented in which *n* is the total number of immune cells which is evaluated as

$$n = [(N \times 1000) \times (\text{Total\% of cells of the} \\ \text{Immune System})]$$
(1)

If it is =>75% then the disease level is considered to be high. If it is =50% < 75% then the disease level is considered to be medium. Level of disease below 50% is considered to be low

$$IMC = \left[\left(\left(\frac{n}{\% \text{ of } B \text{ - Cells}} \right) \times \% \text{ of disease level} \right) + \left(\left(\frac{n}{\% \text{ of } T \text{ - Cells}} \right) \times \% \text{ of disease level} \right) \right]$$
(2)

The other systems considered are the Nervous system which has 1% of the total cells. Respiratory system has 0.00015% of the total cells, Digestive system has 40% of the cells and Integumentary system has 4% of cells. While rest of the 25% contained by Endocrine system, Circulatory system, Urinary and Reproductive system. Based on this percentage level, nodes are been represented and connected together in a network for each system. Since in the cellular activities the Nervous system plays the major role all the systems network are connected to the centralized Nervous system which form cellular network of the biological system as a whole. In the representation of the nodes along with the immune system, symptoms and functionalities of the cells are taken into consideration.

Since, construction of biological system cellular network is a complex task in representing in a visualized form and it is very hard to understand along with cellular communication and representing its reactivity. For this reason number of nodes representation and its evaluation falls on the following consequences based on the percentage level of the cells of the sub-systems of the whole biological system. If the cells level of the sub-system < 3% then the total number of nodes (N_1) is $N_1 = AN/100$. If the cells level of the sub-system is from 4% to 10% then the total number of nodes (N_2) is $N_2 = AN/(100 \times 4)$. If the cells level of the sub-system is > 10% then the total number of nodes (AN) that should exist for the particular sub-system is evaluated as

$$AN = [(N \times 1000) \times \% \text{ of cells of the particular}$$

sub-system] (3)

N is the approximate total number of cells exist in the whole biological system. If odd number of nodes exist than, [No. of Nodes = (No. of Nodes + 1)]. Nodes interconnectivity is based on complete graph theory where the total number of links is

No. of links =
$$[No. of Nodes \times (No. of Nodes-1)]/2$$

If number of Nodes in a subsystem is greater than 50 then No. of Nodes = (No. of Nodes)/2. If Nodes found to odd than No. of Nodes = No. of Nodes + 1. Some links (paths) are virtual through which cells signaling doesn't take place or signaling takes place through the path due to infection of defective of the cells. The evaluation carried out for is mainly for reducing the actual number of node and its interconnectivity for clear representation and visualization of cells signaling and its reactivity. Figure 2 represents adult human system type and its mode of representation of cells. Figure 3 represents a complex fully interconnected sample cellular network without any infection or disease.





Signaling represents communication between cells or communication from the environment to the cells. Here three signaling modes are considered Autocrine *signaling*, Endocrine signaling and Paracrine signaling. Identification of cancer mainly found through the overproduction of certain paracrine growth factors. Signaling within the network and the reactivity of the nodes depends upon the functionality or sign in the biological system. Cells reactivity and functionalities of the whole system are to be represented in graphical manner. Based on the complexity found from Figure 3 a reduce representation is made in Figure 4 for constructing the biological network to better understand the cellular activities in the system based on the disease. Interlinking between the sub-systems are represented on two bases static and dynamic. Static linking is a static linkage to integrate the cellular network sub-systems with is integrated with the centre point of each sub-cellular network. Dynamic linking is the actual pathway representation depends on the activities of cells within or with other subsystems.



Figure 3: Complex Interconnected Cellular Network



Figure 4: Biological System Cellular Network (a) Sample Cellular Network for System without Disease (b) Sample Cellular Network for System with Disease

2.4 Movie Construction

The graphical display for visualizing signals within cells and it reactivity or functionality of a particular disease as a whole is converted as a movie with users specificity which is of customizable. The whole community could visualize cells activities in accordance to the disease and its immune system or antiviral reactivity as movie through means of various searches. It also provides a way for further researches in the area of drug discovery and also researches in disease identification. Figure 5 represents the overall functionality of the tool.



Figure 5: Overall Functionality of the Tool

3. RESULT ANALYSIS AND EVALUATION

Cancer is one of the imperative diseases that most people get affected all around the world that leads most deaths. Cancer is group of many different diseases with different phenotypic characteristics in higher multi-cellular organisms. Though cancer cells have some commonly increased metabolic pathways, such as those involved in nucleic acid synthesis, there is tremendous biochemical heterogeneity among malignant neoplasms which differentiate cancers based on enzymatic alterations.

For the developed tool various result analyses have been made and evaluated. Here in this section illustration for Lung Cancer with result analyses and evaluation have been narrated. The disease is due to uncontrolled cell growth in tissues of the lung. This is the most common cause of cancerrelated death in men. Lung cancer has four main histological categories based on tumor cells morphology they are squamous cell carcinomas, small cell lung carcinomas, adenocarcinomas and large cell carcinomas. The etiology of lung cancer is mostly due to cigarette smoking which have more effect over the cardiovascular system where the cancer cells react to hypoxic conditions by upregulating expression of HIF-1. Other proposed causes include environmental pollution, passive smoke inhalation, and radon exposure. These are all minor players and, in fact, the role of passive smoke and radon in the home as causes are insignificant, in spite of environmental. Mutations in p53 are common and to be observed in early tumors which may be mutated in up to 50% of lung cancer cases caused by tobacco carcinogens have high frequency of $G \rightarrow T$ transversions that may be induced by benzo (pyrene). For lung cancers caused by radon exposure p53mutations comprise $C \rightarrow A$ transversions, $G \rightarrow A$ transitions and the hot spot mutation AGG \rightarrow ATG in codon 249.

Lung cancer can spread to any organ in the body particularly adrenal glands, lymph nodes, liver, brain, kidney, thyroid, spleen, pleura diaphragm and bones are the most common sites for lung-cancer metastasis. The major symptoms of lung cancer are.

Coughing:

Physiological action of coughing is responsible in part to the vagus nerve, which runs through the lungs to the brain. Coughs are prolonged and forceful, and caused by decreased blood flow to the brain, secondary to raised intrathoracic pressure due to cough. Other acute complications include insomnia, cough-induced vomiting, chest pain due to muscular strain of incessant coughing, rupture of bulla causing pneumothorax. Pain exists due to pressure that resides from the lungs.

Dyspnea (Shortness of Breath):

Lungs become stiff and require increased effort to expand

during inhalation. Severe curvature of the spine (scoliosis) also restricts breathing by reducing the movement of the rib cage. The heart pumps blood through the lungs so heart function must be proper for the lungs to function normally. If the heart is pumping inadequately fluid will accumulate in the lung which is pulmonary edema. This condition causes dyspnea that is often accompanied by a feeling of smothering or heaviness in the chest. The fluid accumulation in the lungs may also lead to airway narrowing and wheezing.

Hemoptysis (Coughing Up of Blood from the Respiratory System):

The cause is blood from the nose that has traveled down to the airways and then is coughed up. Blockage of an artery by a blood clot causes hemoptysis due to death of lung tissue. Other causes of hemoptysis are high blood pressure in the pulmonary veins, lung circulation problems, inflammatory conditions of the pulmonary blood vessels.

Pain in the Chest and Shoulder:

The cause is due to destruction of tissue by the tumor, infection, stretching of internal organs and obstruction. Pain that results from tissue destruction is the bone pain from invasive that cause periosteal irritation and bone fractures. Infection pain that often occurs in cancer system with decreased immunity is herpes zoster, which produces a nerve pain or causalgia due to distribution of the affected nerve.

Loss of Appetite:

Lung tumor tissue wasting predominantly muscle and fat, probably involves all organs with the exception of the heart, liver, and brain. Direct involvement of the liver, bile duct, or pancreas can reduce the production of digestive enzymes that are needed for digestion of foodstuffs with decreased in nitrogen content. Due to decreased digestive functions, metabolic demands of the tumor of the host's defense systems causes for decrease of appetite in the system.

Fatigue:

The most common hematologic effect of cancer is anemia. Anemia resulting from decreased levels of folic acid and iron-deficiency as a result of malabsorption or hemolysis. Anemia or blood loss caused due to decreased number of red blood cells, which carry oxygen to the tissues.

Bone Fractures:

In severe cases, prolonged coughing can cause fatigue fractures of lower ribs and costochondritis. Tissues inflammation exists that is connected between the breastbone and the ribs. Tumors commonly have bony metastases and produce hypercalcemia. This leads to direct destruction of bone, with release of calcium into the bloodstream.

Bleeding and Blood Clots:

Disseminated intravascular coagulation (DIC), characterized by an extensive activation of the blood coagulation system within the circulatory system and the deposition of fibrin clots in small blood vessels. It has been postulated that DIC is due to the release of thromboplastin-like substances by the tumor.

Neurological Symptoms:

Autoimmune response to neuronal antigens expressed in cancer cells are neurological symptom such as imbalance, memory loss, muscle weakness, or vision loss due to manifestations of hypercalcemia that include central nervous system effects. Most commonly PND's are observed in cancer affected systems.

Based on the specific cancer, symptoms in the biological system along with internal activities of the system are extracted from the databases with sample human genome. Based on the sample genome graph is constructed along with specific disease pathway as represented. Construction of the biological cellular network is done based on the user specified disease level which has been represented for lung cancer with medium disease level with connected nodes. Links within the sub systems are represented in grey with represents pathways that may exist. Based on the symptoms and its functionalities links (pathways) between nodes (cells) will be graphically represented in black within sub systems of with other sub systems with graphical representation of cells activities and its representation in textual form. The tool has provided the facility of converting the whole cellular activities in movie format which could be made available for the whole community based on the user specific search made as shown in Figure 6 for visualizing and understanding cancer cells activities together with the overall system activities. Table 1 represents the list of features and functionalities of the novel tool. For this result tool found to have 0.80 precision which is evaluated as Precision [1] =

Table 1						
CABS	Tool	Features	and	Functionalities		

License Free	Windows Platform	
GUI	Han Estandia	
External Database	User Friendly	
Automatic Detection	Context-Sensitive	
Online Help	Navigate able	
Graph Theory	Fully web-based	
Chart	Browser compactable	
Graphics	Customizable	

(number of named entity a system correctly detected) / (total number of named entity identified by the system), 0.85 recall which is evaluated as Recall [1] = number of named entity a system correctly detected / total number of named entity contained in the input text and 0.82 *F*-score which is calculated as *F*-score [8] = [(2*Precision*Recall) / (Precision + Recall)]. Apart from this various inputs have been given and the tools have been evaluated.



Figure 6: Lung Cancer (a) Search and Cellular Movie Visualization (b) Genome Graph and Pathway (c) Cellular Network and Cells Functionality

In the tool evaluation scheme precision represents performance of the system and the correct information that are been accurately retrieved. The good the precision performance is good with comparison made with high or good recall. The lower the recall the retrieval ability of relevant information is considered to be less then even if precision is high the performance of the tool is considered to be less or better [1]. *F*-score denotes the effectiveness of the tool variation based on the evaluation carried out. Based on the results analyzed, precision found to be good and the retrieval ability also found to be good. The effectiveness of the tool is less in variation, thus the performance of the tool is considered to be good.

5. CONCLUSION AND FUTURE ENHANCEMENT

The main aim of the paper is to represent a visualized novel tool for representing internal activities of the biological systems at cellular level for various species which has been initially stepped for human species. Cancer is the major diseases that have been considered are brain cancer, ovarian cancer and lung cancer. The tool will be further extended for exact representation of cellular activities of the biological systems for all diseases and for all species. The research is carried out apart from this for modeling a generalized structure for various cells based on its functionalities. Apart from the cellular network construction will be considered for utilizing cells activities representation at molecule level also with pathways. This upgradation will further provides an exact visualized representation of cells activities in the biological system.

The tool utilization benefits not only researchers for globally making their novel identification found on cells, but also helps other communities to visualize cells and its functionalities in a graphical manner in the cancer affected biological systems. This tool could provide a base for treating disease and in drug discovery. The experiment has been carried out in such a way by observing and evaluating the tool performance also. From the result observed CABS tool performance found to be good with appropriated representation cellular activities to some extent.

References

- Jayanthi Manicassamy and P. Dhavachelvan, "Metrics based Performance Control Over Text Mining Tools in Bioinformatics", ACM Portal, 171-176, (January 2009).
- [2] Rayavarapu Bhargavi, G. Madhavi Sastry, U. Suryanarayana Murthy and G. Narahari Sastry, "Structural and Active Site Analysis of Plasmepsins of Plasmodium Falciparum: Potential Anti-Malarial Targets", *Elsevier*, (August 2005).
- [3] Rodrigo Santamaria, Roberto Thero'n and Luis Quintales, "BicOverlapper: A Tool for Bicluster Visualization", ACM Portal, 1212–1213, (March 2008).
- [4] Matteo Fattore and Patrizio Arrigo, "Knowledge Discovery and System Biology in Molecular Medicine: An Application on Neurodegenerative Diseases", *ISB*, (December 2004).
- [5] Michael Baitaluk, Mayya Sedova, Animesh Ray and Amarnath Gupta, "BiologicalNetworks: Visualization and Analysis Tool for Systems Biology", *Oxford Journal*, W466-W471, (March 2006).
- [6] Eran Hodis, Gideon Schreiber, Kistian Rother and Joel L. Sussman, "eMovie: A Storyboard-based Tools for Making Molecular Movies", *Elsevier*, (April 2007).
- [7] Ali Cakmak and Gultekin Ozsoyoglu, "Mining Biological Networks for Unknown Pathways", Oxford Journal, 2775– 2783, (August 2007).

- [8] Richard J. Edwards, Norman E. Davey and Denis C. Shields, "CompariMotif: Quick and Easy Comparisons of Sequence Motifs", ACM Portal, 1307–1309, (March 2008).
- [9] Cigdem Demir. B., Chandandeep Nagi, Bulent Yener, "Cell-Graph Mining for Breast Tissue Modeling and Classification", *PubMed*, 5311-4, (2007).
- [10] Marta Sabou, Chris Wroe, Carole Goble, Gilad Mishne, "Learning Domain Ontologies for Web Service Descriptions: An Experiment in Bioinformatics", *Citeseer*, (May 2005).
- [11] Richard Tzong-Han Tsai, Shih-Hung Wu, Wen-Chi Chou, Yu-Chun Lin, Ding He, Jieh Hsiang, Ting-Yi Sung and Wen-Lian Hsu, "Various Criteria in the Evaluation of Biomedical Named Entity Recognition", *PubMed*, 7-92, (February 2006).
- [12] Jung-Jae Kim, Piotr Pezik and Dietrich Rebholz-Schuhmann, "MedEvi: Retrieving Textual Evidence of Relations between Biomedical Concepts from Medline", *ACM Portal*, 1410–1412, (March 2008).
- [13] Arek Gladki, Pawel Siedlecki, Szymon Kaczanowski and Piotr Zielenkiewicz, "e-LiSe—An Online Tool for Finding Needles in the '(Medline) Haystack'", ACM Portal, 1115– 1117, (March 2008).
- [14] Mario Falchi and Christian Fuchsberger, "Jenti: An Efficient Tool for Mining Complex Inbred Genealogies", ACM Portal, 724–726, (January 2008).
- [15] Yang Jin, Ryan T. McDonald, Kevin Lerman, Mark A. Mandel, Steven Carroll, Mark Y. Liberman, Fernando C. Pereira, Raymond S. Winters and Peter S. White, "Automated Recognition of Malignancy Mentions in Biomedical Literature", *BMC Bioinformatics*, 7-492, (November 2006).
- [16] Maryann E. Martone, Joshua Tran, Willy W. Wong, Joy Sarigis, "The Cell Centered Database Project: An Update on Building Community Resources for Managing and Sharing 3D Imaging Data", *Elsevier*, (October 2007).
- [17] Raymond W. Ruddon, "Cancer Biology", Fourth Edition, Oxford University Press, 2007.
- [18] Wolfgang Arthur Schluz, "Molecular Biology of Human Cancers: An Advanced Students Textbook", Springer, (2005).
- [19] Dominik Wodarz, Natalia L. Komarova, "Computational Biology of Cancer: Lecture Notes and Mathematical Modeling", World Scientific Publishing Co. Pte. Ltd., (2005).
- [20] Georg F. Weber, "Molecular Mechanisms of Cancers", *Springer*, (2007).
- [21] James V. Freeman, Frederick E. Dewey, David M. Hadley, Jonathan Myers, and Victor F. Froelicher, "Autonomic Nervous System Interaction with the Cardiovascular System during Exercise", *Progress in Cardiovascular Diseases*, 48, (5) (March/April), (2006) 342-362.