Apoptosis Gene Information System – AGIS

Kishore R. Sakharkar [1], Marie V. Clement [2], Vincent T.K. Chow [1], and Shazib Pervaiz [3]

[1] Human Genome Laboratory, Department of Microbiology, [2] Department of Biochemistry, [3] Department of Physiology and Oncology Research Institute, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

TABLE OF CONTENTS

1. ABSTRACT

Genes implicated in apoptosis have great relevance to biology, medicine and oncology. Here, we describe a unique resource, **A**poptosis **G**ene **I**nformation **S**ystem (AGIS) that provides data for over 2400 genes involved directly or indirectly, in apoptotic pathways of more than 350 different organisms. The organization of this information system is based on the principle of one-gene, one record. AGIS will be updated on a six monthly basis as new information becomes available. AGIS can be accessed at: http://www.cellfate.org/AGIS/.

2. INTRODUCTION

During development, normal homeostasis involves a tight balance between rates of cell proliferation and cell death, and by implication any digression from the inherent controls on these processes manifests as pathological states. Programmed cell death or apoptosis is a highly conserved process, vital for normal development of organisms as small and simple as *Caenorhabditis elegans* to highly complex multi-cellular mammals. For example, competent apoptotic machinery during embryogenesis is imperative for the morphological and functional maturation of biological systems, negative selection of self-reactive clones of thymocytes, and pruning of over-exuberant neurons exhibiting aberrant circuitry (1-2). However, the same process in a fully developed organism heralds the existence of a pathological process as evidenced by a host of disease states where an abnormally increased or decreased rate of cell death contributes to selective loss or overgrowth of the cell population, respectively. The link between apoptosis and pathological states has been demonstrated in autoimmune disorders, cancer, eradication of neoplastic cells by activated lymphocytes, neurodegenerative disorders, and viral infections. These observations have not only increased our understanding of underlying etiological processes resulting in disease states, but have also provided insights into the intricate cross-talk between proteins involved in diverse cellular pathways during apoptotic signaling. Thus, recognition of the regulated mechanism of cell death marks a new era in many fields of biomedical science including developmental

biology, immunology and oncology (3). Therefore, a greater understanding of the control of apoptosis will contribute towards prophylactic or therapeutic modulation of the process (4), and prevention or protection from disorders in which aberrant death signaling is a central problem (5-6).

The past few years have seen an explosion in the volume of bioinformatics data that is available to researchers. However, as the rate of discovery continues apace, it is becoming ever more difficult to make sense of these data due to problems with data extraction on a specific topic of interest and data inter-operability. The implications of a particular dataset often span the divide between multiple data sources yet existing tools and techniques rarely achieve the same. In order to exploit the information that is available, and to cope with the ever-increasing volume of new data that is now being generated, it is indeed desirable to develop a robust and maintainable mechanism for integrating data resources on apoptosis from different domains. The Apoptosis Gene Information System (AGIS) aims to serve as an online integrated resource of information about proteins involved in apoptosis. AGIS is currently the only resource that compiles the information on apoptotic genes and their products from existing online resources in one place, has a biologist friendly interface and links to bibliography. This is crucial for better understanding of sequence→ structure → function → pathway → phenotype relationships. The information system stores information gleaned from the internet on over 2400 genes/proteins involved in extrinsic and intrinsic apoptotic pathways, thereby serving as a comprehensive educational and research resource on apoptosis. Not only is AGIS useful as a reference for researchers, but it may serve as a basis for experimental work. Through AGIS, researchers may focus on specific genes/proteins and find novel links between the players involved or even construct new testable hypotheses. AGIS is an important step in understanding and integrating the process of apoptosis, and provides a framework upon which a systems biology approach to apoptosis-related genes/proteins and related diseases could be developed. In order to keep pace with the rapidity with which new observations are changing the face of apoptosis biology, the main goal would be to keep the databases up-to-date and continue to enhance the AGIS system so as to provide the optimal starting point for investigating apoptosis.

3. MATERIALS AND METHODS

3.1. Dataset creation

The primary resource in AGIS is a dataset of genes in human and mouse involved in extrinsic and intrinsic apoptosis pathways that were downloaded from ATCC (http://www.atcc.org). This gene list was supplemented with genes involved in apoptosis across other organisms from the SwissProt database and with data gleaned from the literature. The genes known not to be directly involved in apoptosis were clustered under a separate category manually. The dataset currently contains more than 2400 genes.

3.2. Gene information

The information for each gene name was gathered from over 50 online resources and compiled in

one page report. Information (where available) for each gene name provides links to nucleotide and amino acid sequence data, open reading frame (ORF), protein–protein interactions, tissue expression profiles, functional domains, splice variants, disease involvement and many more links to pre-computed data from renowned resources. AGIS thus comprises primary and derived data on apoptotic genes and is designed to serve as a single-stop solution for accession, retrieval and analyses of apoptotic genes and their protein products.

3.3. Ortholog mapping and clustering

Each entry has been mapped to its orthologs in other organisms based on gene name and synonyms. This allows researchers working on human apoptosis to have an understanding of their findings in relation to other model organisms and vice-versa.

3.4. Database creation

Primary as well as derived data in AGIS are organized as a dynamic database with a user-friendly front-end and software tools for data analysis and mining. The organization of the information system is a hierarchical report based on the principle of one-gene, one record.

4. RESULTS AND DISCUSSSION

4.1. Data integration and AGIS

The process of building a new database relevant to some field of study in biology involves transforming, integrating, and cleansing multiple external data sources, as well as adding new material and annotations. The number of secondary, value-added databases being developed within the biomedical community is increasing rapidly. Many questions that a biologist is interested in cannot be answered using any single data source. However, some of these queries can be satisfactorily solved by using information from several sources. Hence, it is imperative to build systems that can provide an integrated view on genes and proteins. In this report, we present AGIS – the Apoptosis Gene Information System. AGIS is a broad scale data integration system that allows several databases to be viewed as if they reside in a single united nested system. It automatically handles heterogeneity so that the user can formulate his/her queries in a way that is independent of geographic location and independent of access protocols to the system through a single search page. In AGIS, we consider two factors - data management and knowledge discovery. Data management involves tasks such as integration of relevant data from various evolving sources, transformation of the integrated data into more suitable forms for analysis, cleansing of data to avoid errors in analysis, etc. Knowledge discovery involves the construction and curation of databases to extract various types of information such as splice variants, protein family, gene structure, protein structure and domain shuffling information for a given gene.

4.2. Comparative genomics and apoptosis using AGIS

The morphological and biochemical changes in apoptosis are orchestrated and regulated by intracellular
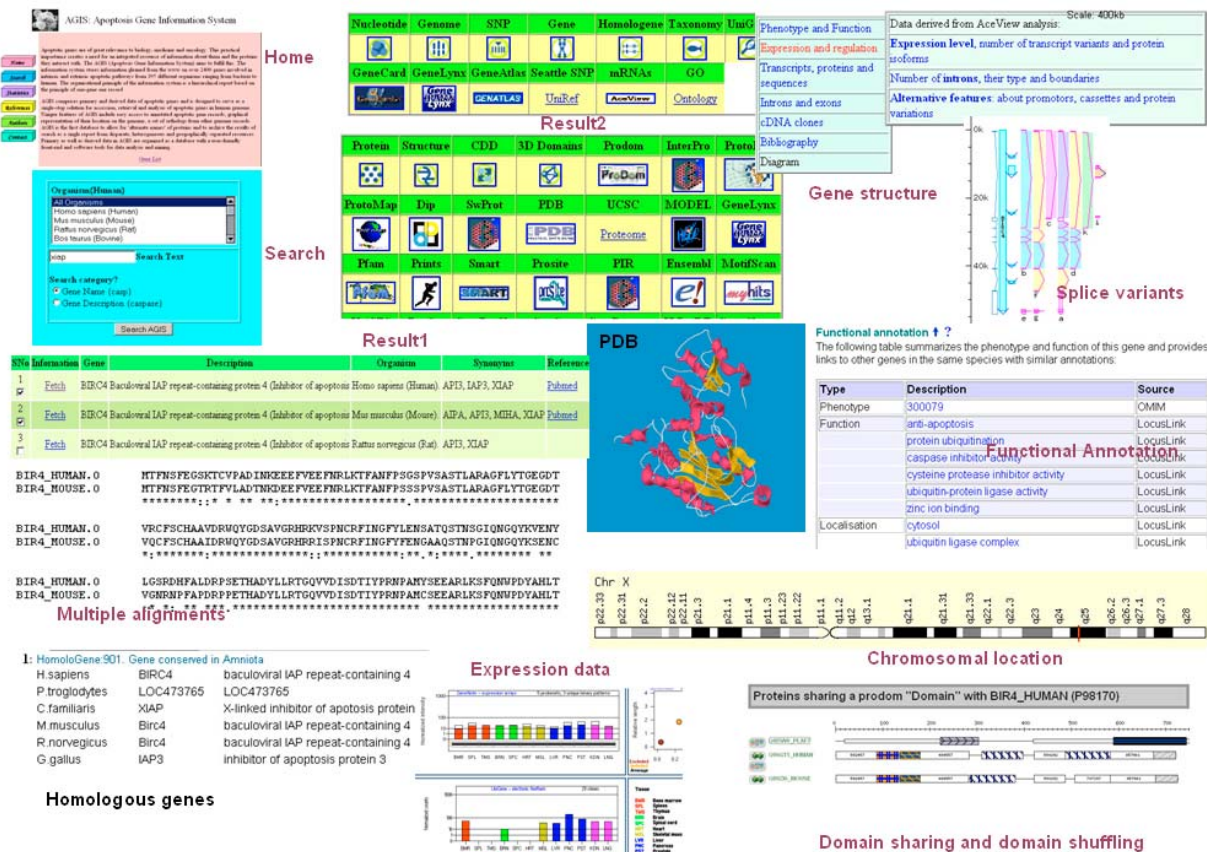
Figure 1. Example of a search on AGIS for the XIAP gene.

pathways involving a host of proteins that are activated in a sequential manner to trigger or control processes of death commitment, execution, and clearance. The heightened interest in studying the signaling networks involved during apoptotic commitment, execution, and clearance stems from the highly preserved nature of this pathway in cnidaria (7), nematodes (8), plants (9), mammals (10), insects (11) and yeasts (12). The assessment of diversity and distribution of these proteins is possible with the availability of genome data. AGIS currently contains more than 2400 genes from 395 different organisms ranging from bacteria to human. This allows for comparisons between human apoptosis and that of other organisms, as well as to understand the evolutionary forces shaping apoptosis. Comparison of apoptotic proteins across organisms may shed light on alternate pathways in different organisms and also on the evolution of genes involved in apoptosis. AGIS is the first database to allow searching '*alternate names*' of proteins, and to archive the results of search as a single report from disparate, heterogeneous and geographically distinct resources.

4.3. AGIS interface and availability

AGIS is available in the internet at: http://www.cellfate.org/AGIS/. Search by a gene name retrieves the gene of interest and all the entries that have the gene of interest as a keyword. The information system allows for search using gene name (e.g. casp), gene description (e.g. caspase), organism (e.g. *Homo sapiens*) and accession number. The retrieved results can be subjected to multiple sequence alignment using ClustalW, thereby allowing viewing of conserved regions and analysis of aspects of protein evolution. Links to more than 50 heterogeneous databases provide information on various aspects ranging from splice variants (via Aceview server that may provide information on the reasons for the variations in the protein sequence seen via ClustalW), domain shuffling (via Prodom server), protein structure (via PDB server, if available), literature (via PubMed), etc. AGIS thus provides a unified view about a specific apoptotic gene from diverse sources, that when integrated may provide important clues about the biological functions and evolution of the gene.

4.4. AGIS goal and update

Our goal is to provide the maximum available information for each gene/protein involved in apoptosis and to facilitate researchers in their quest for information directly or indirectly related to apoptotic signal transduction pathways. We propose to update AGIS regularly every six months. In future, AGIS aims to become a major resource in understanding the genetics of human apoptosis.

5. Search example

Figure 1 shows an example of a search using AGIS.

6. REFERENCES

1. Danial N.N. and S.J. Korsmeyer, Cell death: critical control points. *Cell*, 116, 205-219 (2004)

2. Armstrong S.A. and S.J. Korsmeyer, Comparison of human genomics and genetic models of cancer to identify novel therapeutic targets. *Cell Cycle*, 2, 408-409 (2003)

3. James E.R. and D.R. Green, Manipulation of apoptosis in the host-parasite interaction. *Trends Parasitol.*, 20, 280-287 (2004)

4. Pervaiz S. and M.V. Clement, Tumor intracellular redox status and drug resistance - serendipity or a causal relationship? *Curr Pharm Des.*, 10, 1969-1977 (2004)

5. Akram S., H.F. Teong, L. Fliegel, S. Pervaiz and M.V. Clement, Reactive oxygen species-mediated regulation of the Na(+)-H(+) exchanger 1 gene expression connects intracellular redox status with cells' sensitivity to death triggers. *Cell Death Differ.*, Sep 23 (in press) (2005)

6. Clement M.V., J.L. Hirpara and S. Pervaiz, Decrease in intracellular superoxide sensitizes Bcl-2-overexpressing tumor cells to receptor and drug-induced apoptosis independent of the mitochondria. *Cell Death Differ.*, 10, 1273-1285 (2003)

7. Cikala M., B. Wilm, E. Hobmayer, A. Bottger and C.N. David, Identification of caspases and apoptosis in the simple metazoan Hydra. *Curr Biol.*, 9, 959-962 (1999)

8. Kinchen J.M. and M.O. Hengartner, Tales of cannibalism, suicide, and murder: Programmed cell death in C. elegans. *Curr Top Dev Biol.*, 65, 1-45 (2005)

9. Solomon M., B. Belenghi, M. Delledonne, E. Menachem and A. Levine, The involvement of cysteine proteases and protease inhibitor genes in the regulation of programmed cell death in plants. *Plant Cell*, 11, 431-444 (1999)

10. Tittel J.N. and H. Steller, A comparison of programmed cell death between species. *Genome Biol.,* 1, Reviews 0003 (2000)

11. Richardson H. and S. Kumar, Death to flies: Drosophila as a model system to study programmed cell death. *J Immunol Methods*, 265, 21-38 (2002)

12. Skulachev V.P., Programmed death in yeast as adaptation? *FEBS Lett.*, 528, 23-26 (2002).

Send Correspondence to: Shazib Pervaiz, MD, PhD, Department of Physiology and Oncology Research Institute, Yong Loo Lin School of Medicine, National University of Singapore,2 Medical Drive, MD9, #02-05 Singapore 117597, Tel: 65-96614732, Fax: 65-67735461, E-mail: phssp@nus.edu.sg