

GeneDecks: A Systems Biology Facilitator with Combinatorial GeneCards Display

Doron Lancet¹, Shany Ron¹, Michael Shmoish¹, Asaf Madi¹, Alexandra Sirota¹, Karin Noy¹, Naomi Rosen¹, Ohad Greenshpan¹, Orit Shmueli¹, Marilyn Safran², Yonatan Aumann³ and Liora Strichman-Almashanu¹

¹Departments of Molecular Genetics and ²Biological Services (Bioinformatics Unit), The Weizmann Institute of Science, Rehovot 76100, ³Department of Computer Science, Bar-Ilan University, Ramat Gan, Israel.

In the post-genome era, Bioinformatics is beginning to focus on Systems Biology, incorporating whole cell / whole genome views and biological networks and pathways. Relevant enrichment of gene and protein databases thus constitutes a major challenge. GeneCards (<http://www.genecards.org/>), a compendium of annotative information about human genes, has stood up to this challenge for almost a decade. Along with its auxiliary databases GeneLoc, GeneNote, GeneAnnot and GeneTide, it strives to provide a comprehensive resource of gene-centric information. Every GeneCard contains knowledge relevant to a specific human gene, automatically mined from ~50 databases worldwide. GeneCards' gene-centric view is an excellent starting point for genome browsing, providing rich and relevant annotation at a glance. A particular emphasis is placed on gene expression (see <http://genecards.weizmann.ac.il/genenote/>), and pathway information. We are currently introducing batch queries to enhance the power of analyzing gene lists stemming from microarray data as well as other high-throughput experiments.

GeneDecks is a new research tool, aimed at exploiting the rich and varied data found in the GeneCards family of databases for system-level analysis. Whereas GeneCards is *gene-centric*, GeneDecks is *set-centric*, promoting the research and analysis of groups of genes. By uncovering combinatorial relationships within assortments of genes, GeneDecks improves one's ability to reach better systems understanding. The GeneDecks project was initiated in order to provide a foundation for supporting sophisticated and flexible analyses of gene sets in relation to biological characteristics. GeneDecks includes features such as:

1. For a given gene, extract a set of associated genes through similarity in one or more selected GeneCards categories.
2. For a set of genes, provide significant shared annotations for this group, based on the large number of available GeneCards categories.
3. Provide On-Line Analytical Processing (OLAP) functionality for browsing and exploring relationships between GeneCards genes and sets thereof.
4. Find sets of genes with interesting data relationships, using data mining on the GeneCards annotations space

On these relationships, a judicious use of context dependent distance measures will be defined for classifying the variety of GeneCards data-types (including textual strings, numeric information, statistical data and complex objects with multiple attributes), in order to enable the discovery of assorted sets of genes of interest to biological researchers. A generic query mechanism will allow user defined retrieval of relevant genes and information related to any set of GeneCards genes. This query will support retrieving of related gene sets of 2nd degree or more, allowing in-depth research of gene sets characteristics.

Examples of non-trivial assortment categories of gene annotations for the aforementioned data-types include transcription patterns, protein-protein interactions, alternative splicing variants, etc. Such discoveries are only possible when applying a systemic view to attributes such as structure, function and location, transforming *de facto* the data itself into the research arena, an ability which GeneDecks facilitates.