

GENEDECKS: A SYSTEMS BIOLOGY FACILITATOR WITH COMBINATORIAL GENECARDS OUTLOOK

Shany Ron*, Liora Strichman-Almashanu*, Michael Shmoish*, Asaf Madi*, Alexandra Sirota*,
Karin Noy*, Naomi Rosen*, Orit Shmueli*, Marilyn Safran** and Doron Lancet*

*Departments of Molecular Genetics and **Biological Services (Bioinformatics Unit), The Weizmann Institute of
Science, Rehovot, Israel

In the post genomic era, most protein-coding human genes have already been identified. Bioinformatics research is beginning to focus on Systems Biology, incorporating whole cell / whole genome views, focusing on biological networks and pathways. Relevant enrichment of gene and protein databases thus constitutes a major challenge. We introduce GeneDecks, a new research tool aimed at exploiting the rich and varied data found in the GeneCards family of databases. By uncovering relationships within assortments of genes, GeneDecks improves one's ability to reach better systems understanding.

GeneCards is an inclusive compendium of annotative information about human genes. Along with its auxiliary databases GeneLoc, GeneNote, GeneAnnot and GeneTide, it strives to provide a comprehensive resource of gene-centered information. The GeneCards database is organized as a compilation of "cards", each corresponding to a unique human gene, and containing information relevant to the gene from a large variety of databases in different realms. The information provided includes location, expression, pathways, disease associations, gene-related chemical compounds, and carefully selected literature. The gene-centric view afforded by GeneCards is an excellent starting point for browsing a new gene of interest, as it provides rich and relevant annotations at a glance.

The GeneDecks project was initiated in order to provide a structural foundation for supporting more sophisticated and flexible analyses of gene sets in relation to biological characteristics. GeneDecks goals include:

1. For a given gene, provide sets of associated genes through similarity in one or more selected annotation categories.
2. For a set of genes, provide significant shared annotations for this group, based on the large number of available GeneCards categories.
3. Create gene clusters from the GeneCards dataset of genes, using a variety of annotation categories for both one-dimensional and multi-dimensional clustering.
4. Provide access to these annotations via both single gene and gene set queries.

An example to illustrate one-dimensional GeneDecks functionality is as follows: the gene WNT1 has orthologues in 7 species as shown in GeneCards, with protein sequence similarity ranging from 100% in chimpanzee to 31% in nematode. GeneDecks will facilitate finding genes with similar ranges of similarities to orthologues in these species. These genes can be inspected from the angles of their expression pattern in normal tissue, participation in specific pathways, association with diseases, and so on. In this example, clustering involves creating a biologically meaningful distance metric, based on orthologue similarity in available species, and using this function to cluster the genes.

From the infrastructure standpoint, care will be taken to define and classify the variety of GeneCards data-types (including textual strings, numeric information, statistical data and multi-attributed complex objects) to enable generic query and classification mechanisms as well as incorporate machine learning-based components. We plan to support the extension of this method to query relationship sets of 2nd degree or more to a gene of interest, allowing in depth research of gene sets characteristics.

Examples of non-trivial assortments of gene annotations for the aforementioned data-types include protein-protein interactions, relations to diseases and the characterization of SNPs. Enabling meaningful discoveries involves defining new distance metrics for these annotation categories, as well as statistical methods for evaluating the significance of the results of clustering processes. Such discoveries are only possible when applying a systemic view to attributes such as structure, function and location, a view which GeneDecks facilitates.

Genecards and GeneDecks can be found at: <http://genecards.weizmann.ac.il/>