GeneCards 3.0: An Object-Oriented Approach

Safran, M.², Shen-Orr, S.³, Solomon, I.², Lapidot, M.¹, Shmueli, O.¹, Rosen, N.¹, Adato, A.¹, Ben-Dor, U., Esterman, N., Chalifa-Caspi, V.², and Lancet, D.¹
¹ Dept. of Molecular Genetics, ² Bioinformatics and Biological Computing Unit, Biological Services, ³ Dept. of Molecular Cell Biology, Weizmann Institute of Science

GeneCards is a database of human genes, maps, proteins, and diseases, with associated software that retrieves, integrates, and displays gene centered human genome information [1,2]. Versions 2.xx have stressed features and usability, including query reformulation and grappling with comprehensiveness versus compactness, in order to present just the right mix of detail and hyperlinks. GeneCards has gained widespread popularity, as evidenced by over two million hits at the home site, mirroring by 25 academic institutions around the world and ever-growing commercial interest. Version 3.0 strives to maintain its successful look and feel, data-mining heuristics, feature enhancements and data upgrades, while strengthening the infrastructure, and standardizing data formats using object-oriented and XML (Extensible Markup Language[3] technologies.

We present the pros and cons of using objectoriented Perl[4] and our hybrid approach of implementing an object-oriented skeleton with some non-object-oriented internals to enhance the system's efficiency.

XML[3] is a meta-language that supports customized tags for describing and providing semantic meaning to structured data. This open and selfdescribing format can be easily parsed by other applications and its typed elements can be arranged within others to form a nested hierarchy. We present two XML schemas for representing the GeneCards data, *GeneCardByResource* and *GeneCardByFunction* and their impact on the GeneCards display software. Moving to XML will also facilitate the implementation of an expanded search engine beyond the current text-based capability to include context-specific searches.

References:

1. Rebhan M, Chalifa-Caspi V, Prilusky J and Lancet D, *"GeneCards: a novel functional genomics compendium with automated data mining and query reformulation support"* Bioinformatics, 14(8):656-664 (1998).

2. Safran M *et al, "GeneCards 2002: towards a complete, object-oriented human gene compendium"*, Bioinformatics, **in press.**

3. http://www.w3.org/XML

4. Conway Damian - *Object Oriented Perl*, Manning Publications Co. 1st edition, 2000

e-mail: cards@bioinformatics.weizmann.ac.il website: http://bioinformatics.weizmann.ac.il/ cards/