

## Harvesting the human genome - from sequence to function

**Doron Lancet**

**Edna Ben-Asher**

**Tsviya Olender**

**Liora Strichman-Almashanu**

**Irina Dalah**

**Tsippi Iny-Stein**

**Asaf Madi**

**Naomi Rosen**

**Marilyn Safran**

**Michael Shmoish**

**Alexandra Sirota**

**Miriam Khen**

**Tamar Koch**

**Ester Feldmesser**

**Yehudit Hassin**

**Idan Menashe**

**Shany Ron**

**Daniela Zalcenstein-Amann**

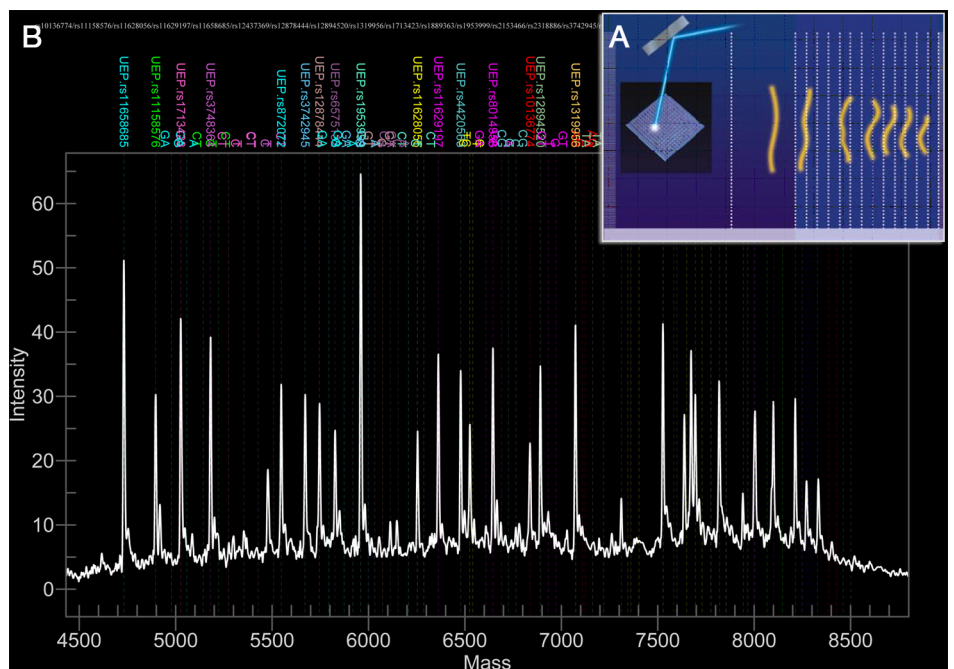
### A. Genomic analyses of hereditary diseases

#### i. Common diseases and traits

Common diseases, such as asthma, diabetes, heart disease, etc. are multi-factorial, involving both genetic and environmental factors and affecting large fractions of the population. Genetic studies of complex traits, in which combinations of several genes are involved, rely heavily on Single Nucleotide Polymorphisms (SNP) scoring in large groups of affected and unaffected individuals. The availability of large sets of SNPs facilitate the definition and characterization of genomic regions manifesting genetic linkage with defined phenotypes, and thus play a central role in the elucidation of the genetics of multigenic inherited traits. [The Crown Human Genome Center](#) has established the only [high throughput SNP](#) scoring facility in Israel, based on a Sequenom mass spectrometry instrument (Figure 1). Ongoing projects, mostly collaborations around Israel, include the identification of candidate schizophrenia-associated genes, polymorphic olfactory receptor genes underlying odor blindness, as well as a search for genes associated with smoking craving. In parallel, Quantitative Traits Loci in tomato and cattle are being studied. SNPs associated with variations in drug response (pharmacogenetics) are analyzed with respect to multiple sclerosis and schizophrenia. Recent technology upgrades include a new Sequenom method, iPLEX, that enables very high multiplexing of genotyping reactions, SNP scoring in DNA pools, and mass-spectrometry-based Gene Expression analyses.

#### ii. Monogenic diseases

In a search for the gene responsible for Medullary Cystic Kidney disease 1 (MCKD1), a linkage disequilibrium study, followed by additional mapping procedures, has indicated a section of 600kb on chromosome 1 as carrying the disease-causing mutation in an Israeli family of Ashkenazi origin. Careful positional cloning analysis of the region has not yet revealed the disease-causing mutation, and therefore the linkage region may be



**Fig. 1** Analyzing Single Nucleotide Polymorphisms (SNPs) using Sequenom MassArray System. A. DNA fragments which carry SNP have different mass and can be separated by a MALDI-TOF (Matrix Assisted Laser Desorption time of flight) mass spectrometer. B. The new iPLEX technology allows up to 25 SNPs' scoring in a single assay.

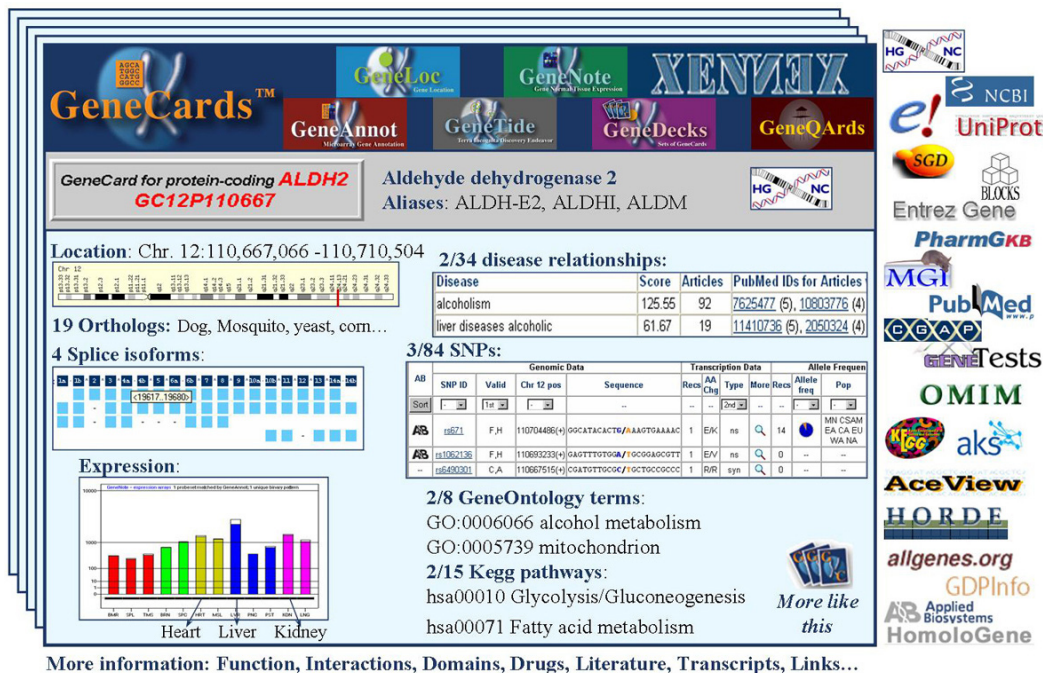
Department of  
molecular genetics

972 8 934 3683

972 8 934 4487

doron.lancet@weizmann.ac.il

[www.weizmann.ac.il/molgen/members/lancet.html](http://www.weizmann.ac.il/molgen/members/lancet.html)



**Fig. 2** A cartoon of the GeneCard for ALDH2. This is a sample of the many topics presented in GeneCards. On the right hand side are logos of our sources for information.

enlarged for further gene searching. This study is carried out in collaboration with Prof. Moti Shohat from the department of Medical Genetics, Rabin Medical Center.

Premature ovarian failure (POF) is another disease which has genetic etiology. Linkage analysis in 11 French families, and one family from East Jerusalem, indicate that the linkage interval is on chromosome 7q22.1. The study, performed in collaboration with Prof. Marc Fellous from Inserm/Hopital Cochin (Paris), is still in progress.

## B. Computational tools for genome-wide bioinformatics

An important Genome Center tool is the [GeneCards database](#), a public integrated database of human genes (also commercially available via licensing from Xennex Inc.), now in the process of expanding to allow for genome-wide computational analyses. GeneCards currently displays information from and links to over 50 different resources worldwide, organized by topic (Figure 2). The types of information in a [GeneCard](#) include names, genomic coordinates, protein features including domains and families, gene functions and pathways, interactions with other proteins and compounds, expression patterns, SNPs, alternative transcripts, orthologues, disorders information, relevant literature, and links to relevant wet-lab services.

The GeneCards suite is a collection of several in-house databases, each specializing in a different topic, with the summarized information also presented in GeneCards: [GeneLoc](#) integrates gene location resources; [GeneAnnot](#) maps microarray probesets to genes; [GeneTide](#) provides a comprehensive and ranked list of EST associations with genes, and helps predict novel genes; [GeneNote](#) provides graphically displayed transcription patterns in human tissues based on WIS-generated gene datasets, and is currently being expanded to include other external datasets.

GeneCards addresses the growing need to understand the complexity of biological systems in the era of systems biology in several different ways: At the single GeneCard level we are adding more information on "systemic issues" such as pathways, molecular interactions, and alternative splicing. We are in the process of organizing information in a function-oriented rather than resource-oriented manner. At the database level we are currently adding a batch query mechanism for retrieval of

selected annotations for a list of genes (GeneAlaCart), as well as planning the development of a search mechanism which will enable elaborate queries for cross sectioning and retrieval of subsets of the data.

GeneDecks, our newest project, aims to provide more elaborate analysis methods related to multiple genes. GeneDecks will include using combinations of biological characteristics to retrieve similar genes to a gene of interest, as well as discovering significant relationships between biological annotations of gene sets. GeneDecks's long-term goal is to supply an interactive interface for complex ad-hoc queries on GeneCards gene subsets that will enable researchers to expand their thinking processes.

## Selected publications

Shmueli, O., Horn-Saban, S., Chalifa-Caspi, V., Shmoish, M., Ophir, R., Benjamin-Rodrig, H., Safran, M., Domany, E., and Lancet, D. (2003) GeneNote: whole genome expression profiles in normal human tissues. *Comptes Rendus Biol.*, 326, 1067-1072.

Rosen, N., Chalifa-Caspi, V., Shmueli, O., Adato, A., Lapidot, M., Stampnitzky, J., Safran, M., and Lancet, D. (2003) GeneLoc: Exon-based integration of human genome maps. *Bioinformatics*, 19 Suppl. 1, i222-224.

Safran, M., Chalifa-Caspi, V., Shmueli, O., Olender, T., Lapidot, M., Rosen, N., Shmoish, M., Peter, Y., Glusman, G., Feldmesser, E., Adato, A., Peter, I., Khen, M., Atarot, T., Groner, Y., and Lancet, D. (2003) Human Gene-Centric Databases at the Weizmann Institute of Science: GeneCards, UDB, CroW 21 and HORDE. *Nucleic Acids Res.*, 31, 142-146.

Avidan, N., Tamary, H., Dgany, O., Cattani, D., Pariente, A., Thulliez, M., Borot, N., Moati, L., Barthelme, A., Shalmon, L., Krasnov, T., Ben-Asher, E., Olender, T., Khen, M., Yaniv, I., Zaizov, R., Shalev, H., Delaunay J, Fellous M, Lancet D, Beckmann JS. (2003) CATSPER2, a human autosomal nonsyndromic male infertility gene. *Eur J Hum Genet.* 11, 497-502.

Chalifa-Caspi, V., Yanai, I., Ophir, R., Rosen, N., Shmoish, M., Benjamin-Rodrig, H., Iny Stein, T., Shmueli, O., Safran, M. and Lancet, D. (2004) GeneAnnot: comprehensive two-way linking between oligonucleotide array probesets and GeneCards genes. *Bioinformatics*, 20, 1457-1458.

Korostishevsky, M., Kaganovich, M., Cholostoy, A., Ashkenazi, M., Ratner, Y., Dahary, D., Bernstein, J., Bening-Abu-Shach,

- U., Ben-Asher, E., Lancet, D., Ritsner, M. and Navon, R. (2004) Is the G72/G30 locus associated with schizophrenia? single nucleotide polymorphisms, haplotypes, and gene expression analysis. *Biol Psychiatry*. 56, 169-76.
- Shklar, M., Strichman-Almashanu, L., Shmueli, O., Shmoish, M., Safran, M., and Lancet, D. (2005) GeneTide--Terra Incognita Discovery Endeavor: a new transcriptome focused member of the GeneCards/GeneNote suite of databases. *Nucleic Acids Res.*, 33, D556 - D561.
- Yanai, I., Benjamin, H., Shmoish, M., Chalifa-Caspi, V., Shklar, M., Ophir, R., Bar-Even, A., Horn-Saban, S., Safran, M., Domany, E., Lancet, D., and Shmueli, O. (2005) Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification *Bioinformatics*, 21, 650-9.
- Starinsky, S., Figer A, Ben-Asher, E., Geva, R., Flex, D., Fidler, HH., Zidan, J., Lancet, D. and Friedman, E. (2005) Genotype phenotype correlations in Israeli colorectal cancer patients. *Int J Cancer*. 10, 58-73.
- Greenbaum, L., Kanyas, K., Karni, O., Merbl, Y., Olender, T., Horowitz, A., Yakir, A., Lancet, D., Ben-Asher, E. and Lerer, B. (2005) Why do young women smoke? I. Direct and interactive effects of environment, psychological characteristics and nicotinic cholinergic receptor genes. *Mol Psychiatry*. 11, 312-322, 223
- Amann, D., Avidan, N., Kanyas, K., Kohn, Y., Hamdan, A., Ben-Asher, E., Macciardi, F., Beckmann, JS., Lancet, D. and Lerer, B. (2006) The trace amine receptor 4 gene is not associated with schizophrenia in a sample linked to chromosome 6q23. *Mol Psychiatry*. 11(2):119-21.

#### **Acknowledgements**

D. Lancet holds the Ralph and Lois Chair in Human Genomics. This research is supported by the Crown Human Genome Center, by an Israel Ministry of Science and Technology grant to the National Knowledge Center in Genomics. XenneX Inc. and The Yeda Fund; The Abraham and Judith Goldwasser Fund; Part of the work is supported by Benozziyo Center for Neurological Disease.