Harvesting the human genome - from sequence to function

Department of
Molecular Genetics

Prof. Doron Lancet

Edna Ben-Asher, Tsviya Olender, Marilyn Safran, Miriam Khen, Tamar Koch, Irina Dalah, Tsippi Iny-Stein, Naomi Rosen, Michael Shmoish, Elena Matusevich, Yakov Perlman, Justin Alexander, Gil Stelzer, Arik Harel, Daniela Zalcenstein-Amann, Yehudit Hassin, Alexandra Sirota, Asaf Madi, Yaron Sole, Emily Brewster

☎ 972 8 934 3683

FAX 972 8 934 4487

@ doron.lancet@weizmann.ac.il

www.weizmann.ac.il/ molgen/ members/lancet.html

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. 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 Fish and cattle are being studied. SNPs associated with variations in drug response (pharmacogenetics) are analyzed with respect to multiple sclerosis and schizophrenia. The last 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. Next Generation Sequencing Technologies

In the race towards '$1000 per genome sequencing' new sequencing technologies, thousands folds more efficient than the previous ones, have been invented. The Genome Center in collaboration with the Biological Services Unit at the WIS are in the process of introducing the next generation technologies to the WIS. Various applications like gene expression, bacterial sequencing, Chromatin Immunoprecipitation Sequencing and whole genome transcriptome are being tested by WIS scientists with the Solexa Genome Analyzer model of Illumina (Fig 1). In the near future the SOLiD system of Applied Biosystems will also be tested at the WIS.

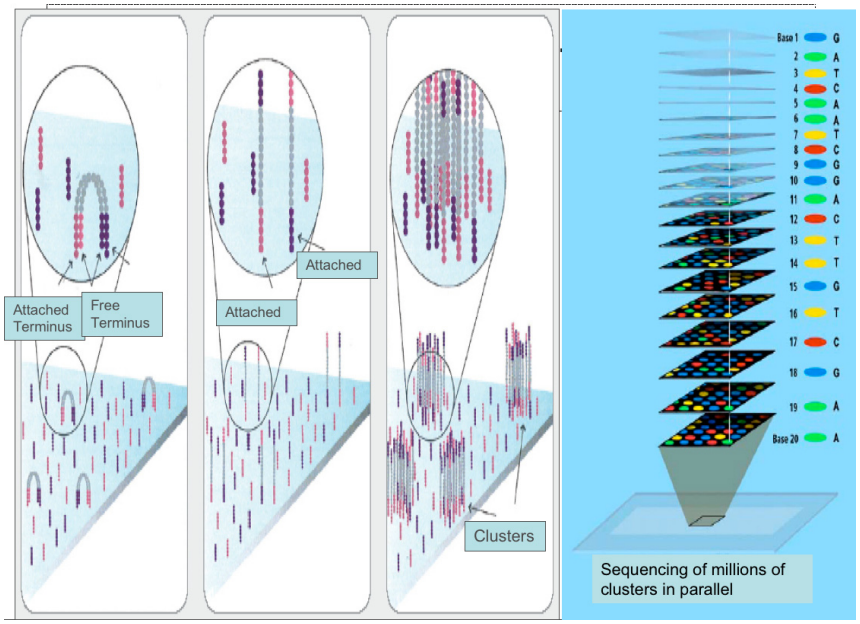

Fig. 1 *Solexa Sequencing Technology. Sequencing templates are immobilized on a flow cell surface designed to present the DNA in a manner that facilitates access to enzymes while ensuring high stability of surface-bound template. Solid phase amplification is employed to create 1000 identical copies of each single molecule in close proximity. This technology can achieve up to ten million single molecule clusters per square centimeter. Sequencing by synthesis uses four fluorescently-labeled modified nucleotides to sequence the millions of clusters in parallel.*

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 70 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

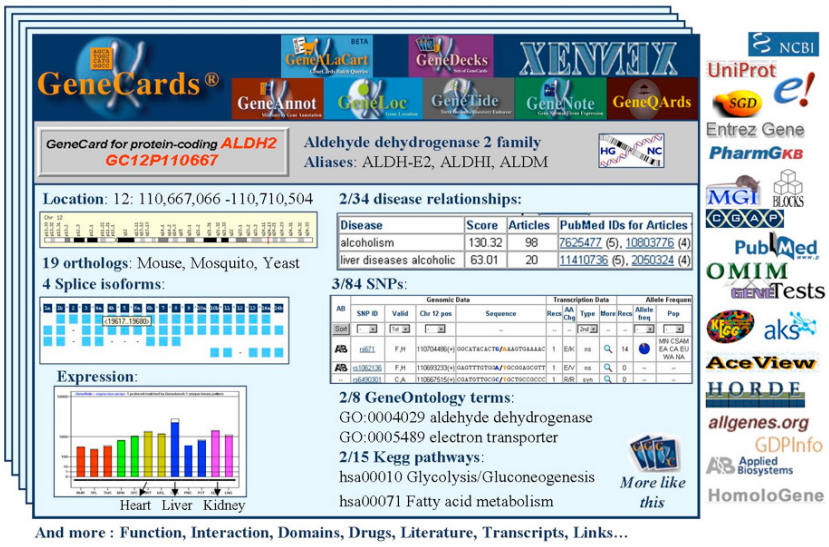

Fig. 2 *A cartoon of the GeneCard for ALDH2. This is a sample of the many topics presented in GeneCards. On the right hand side are logos of our sources for information.*

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 continuing to enhance our 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., Cattan, 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(7):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. 1;56(3):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., Fidder, HH., Zidan, J., Lancet, D. and Friedman, E. (2005)

Genotype phenotype correlations in Israeli colorectal cancer patients. Int J Cancer. 10;114(1):58-73.

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.

Amann-Zalcenstein, D., Avidan, N., Kanyas, K., Ebstein, RP., Kohn, Y., Hamdan, A., Ben-Asher, E., Karni, O., Mujaheed, M., Segman, RH., Maier, W., Macciardi, F., Beckmann, JS., Lancet, D., and Lerer, B. (2006)

AHI1, a pivotal neurodevelopmental gene, and C6orf217 are associated with susceptibility to schizophrenia.

Eur J Hum Genet. Oct;14(10):1111-9. Epub 2006 Jun 14. Erratum in: Eur J Hum Genet. 2007 Mar;15(3):387.

Rigbi, A., Kanyas, K., Yakir, A., Greenbaum, L., Pollak, Y., Ben-Asher, E., Lancet, D., Kertzman, S., Lerer B. (2007) Why do young women smoke? V. Role of direct and interactive effects of nicotinic cholinergic receptor gene variation on neurocognitive function. Genes Brain Behav. Jun 7;

Greenbaum, L., Strous, RD., Kanyas, K., Merbl, Y., Horowitz, A., Karni, O., Katz, E., Kotler, M., Olender, T., Deshpande, SN., Lancet, D., Ben-Asher, E., and Lerer, B.. (2007) Association of the RGS2 gene with extrapyramidal symptoms induced by treatment with antipsychotic medication. Pharmacogenet Genomics. Jul;17(7):519-28.

Grossman, I., Avidan, N., Singer, C., Goldstaub, D., Hayardeny, L., Eyal, E., Ben-Asher, E., Paperna, T., Pe'er, I., Lancet, D., Beckmann, JS., and Miller, A. (2007) Pharmacogenetics of glatiramer acetate therapy for multiple sclerosis reveals drug-response markers. Pharmacogenet Genomics. Aug;17(8):657-66.

Acknowledgements

D. Lancet holds the Ralph and Lois Chair in Human Genetics. Research 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.; The Abraham and Judith Goldwasser Fund; The Gurwin Fund; The Crown Human Genome Center

INTERNAL support

Part of the work is supported by Benoziyo Center for Neurological Disease, the Crown Human Genome Center the Abraham and Judy Goldwasser Fund, and the Gurwin Fund