Doron Lancet

Jacques S. Beckmann

Nili Avidan
Edna Ben-Asher
Tsviya Olender
Marilyn Safran
Tamar Koch
Muriel Shemla
Michael Shmoish
Liora Almashanu
Orit Shmueli
Clara Singer
Itai Yanai
Daniela Amann
Iris Grossman
Hila Benjamin
Inbar C. Gihon
Yehudit Hasin
Naama Kopelman
Naomi Rosen
Maxim Shklar

Harvesting the genome - from sequence to function

Department of Molecular Genetics

Tel. 972 8 934 **4121** **Fax.** 972 8 934 **4487**
E-mail: doron.lancet@weizmann.ac.il
Web page: http://bioinfo.weizmann.ac.il/genome_center

A. Identification of monogenic diseases genes

The completion of the human genome project allowed us to perform the steps beyond linkage analysis to a genomic interval, to enabling effective discovery of monogenic disease genes and their mutations. Six such projects have been successfully completed, as shown in the table, and Fig. 1.

Disease	Disease description	Population	Gene	ref
Mucolipidosis type 4 ML4	Developmental and mental retardation	Ashkenazi Jews	MCOLN1-(novel gene)	1*
Hered. Inclusion body myopathy HIBM	Muscular dystrophy	Persian Jews	GNE	2
Polym. Ventricular tachycardia PVT	Heart failure	Galilee Bedouins	CASQ2	3
Usher syndrome type 3 USH3	Progressive deafness and blindness	Diverse	clarin-1 (novel gene)	4
Congenital diserythropoyetic anemia CDAI	Anemia	Negev Bedouins	codanin-1 (novel gene)	5
Asthenoterato-zoospermia	Male Infertility	French	CATSPER2	8

* Numbers indicate position in list of references

B. Single Nucleotide Polymorphism (SNP) Technologies for polygenic traits

The most prevalent genomic inter-individual variations are Single Nucleotide Polymorphisms (SNPs). The Genome Center has established the only high through-put SNP scoring facility in Israel, based on a Sequenom mass spectrometry instrument. This enables the statistical association between SNPs and polygenic disease phenotypes. Ongoing projects include a search for schizophrenia-associated genes. osteoporosis associated genes, genes linked with smoking craving and a study of polymorphisms in olfactory receptor genes. In parallel, Quantitative Traits Loci in tomato, chicken and cattle are studied. SNPs associated with variations in drug response (pharmacogenetics) are studied in multiple sclerosis (with Teva and the Technion) and schizophrenia (with HUJ). Recent technology upgrades include effort-saving SNP scoring in DNA pools, mass-spectrometry-based RNA quantification and an Affymetrix 10,000-SNP array for whole genome mapping.

C. Functional and Evolutionary genomics

A crucial Genome Center activity is computational whole genome analyses. A centerpiece of this endeavor is GeneCards (http://bioinfo.weizmann.ac.il/genecards), a public integrated compendium of human genes (also commercially available through Xennex Inc). GeneCards uses advanced data mining to provide users with annotated records, with links to 47 world-wide resources. It is equipped with global query capabilities and a semi-automated quality assurance tool (GeneQArds). In the past year GeneCards has been augmented to encompass a suite of databases, including GeneLoc, which integrates gene location

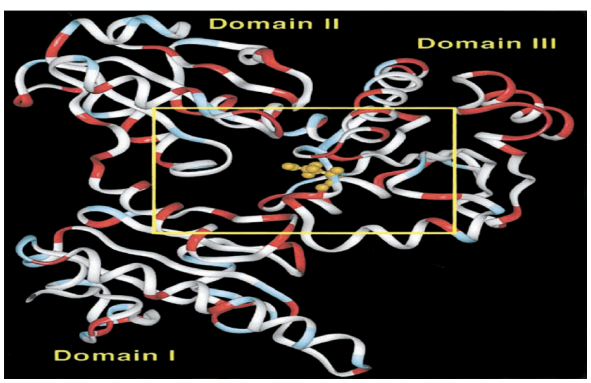

Fig. 1 *Homology model of human CASQ2. The mutated amino-acid (A_{5p}307His) leadind to heart failure is shown in yellow. The basic and the acidic amino-acids are colored in blue and red, respectively.*

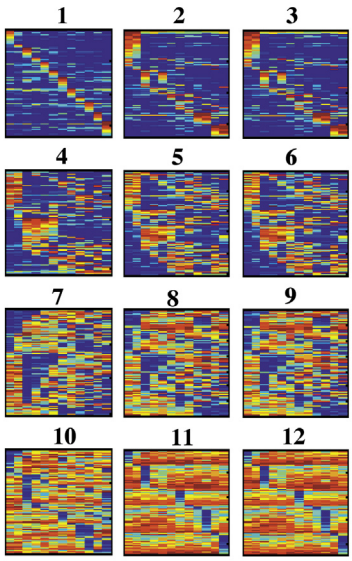

Fig. 2 *Profile clustering according to binary patterns. Each panel shows expression profiles with the indicated number of tissues with high expression. Blue - lowest expression, red - highest expression. Tissue columns from left to right are Brain Spinal cord, Bone marrow, Spleen, Thymus, Lung, Pancreas, Prostate, Heart, Skeletal muscle, Kidney, Liver.*

resources, and GeneAnnot, relating microarray probe sets to genes. GeneCards features include a comprehensive alias list, chromosomal coordinate information; Protein sequence, domains and family data; functional ontology, pathways and disease information; DNA sequence and species orthologs, SNPs and mutations. Comprehensive assignment to the 5 million human expressed sequence tags (ESTs) is done by a new Terra Incognita Discovery Endeavor (TIDE) tool.

A related key activity is the GeneNote project, with a large WIS-generated gene expression dataset. Transcription patterns collected by DNA arrays encompass most human genes, across 12 normal human tissues. The data are used as a powerful annotation tool in GeneCards, and are shown in detail in the GeneNote database. A comparison is provided to electronic Northern (cDNA library statistics) and to Serial Analysis of Gene Expression (SAGE) data. GeneNote further provides a variety of Tissue Specificity Indices (TSIs), continuously spanning the range from housekeeping to tissue-specific, and assisting in the definition of a gene's status as related to tissue differentiation.

Based on GeneNote and related data, we are seeking the genomic properties – including alternative splicing, domain composition, and codon usage – that determine a gene's mode of expression. Evolutionary insight through comparative genomic analyses of human with other species, is proving most useful. For example, a comparison of human and mouse expression profiles suggests widespread neutral expression. Another mode of analysis has revealed a tree of relationships suggesting a timeline for their relative appearance in evolution.

Selected Publications

Bargal, R. et. al. (2000) Identification of the gene causing mucolipidosis IV. Nat Genet, 26, 118- 123.

Eisenberg, I. et. al. (2001) The UDP-N-acetylglucosamine 2-epimerase gene is mutated in hereditary inclusion body myopathy. Nat Genet, 29, 83-87.

Lahat, H. et. al. (2001) A missense mutation in a highly conserved region of CASQ2 is associated with catecholamine-induced polymorphic ventricular tachycardia. Am J Hum Genet, 69, 1378-1384.

Adato, A. et. al. (2002) USH3A transcripts encode clarin-1, a four-transmembrane-domain protein with a possible role in sensory synapses. Eur J Hum Genet, 10, 339-350.

Dgany, O. et. al. (2002) Congenital dyserythropoietic anemia type I is caused by mutations in codanin-1. Am J Hum Genet, 71, 1467-1474.

Safran, M. et. al. (2002) GeneCards 2002: towards a complete, object-oriented, human gene compendium. Bioinformatics, 18, 1542-1543.

Kirstein-Grossman, I. et. al. (2002) Pharmacogenetic Development of Personalized Medicine: Multiple Sclerosis Treatment as a Model. Drug News Perspect, 15, 558-567.

Avidan, N. et. al. (2003) CATSPER2, a human autosomal nonsyndromic male infertility gene. Eur J Hum Genet, 11, 497-502.

Rosen, N. et. al. (2003) GeneLoc: exon-based integration of human genome maps. Bioinformatics, 19 Suppl 1, I222-I224.

Safran, M. et. al. (2003) Human Gene-Centric Databases at the Weizmann Institute of Science. Nucleic Acids Res, 31, 142-146.

Shmueli, O. et. al. (2003) GeneNote: whole genome expression profiles in normal human tissues. Comptes Rendus Biologies, 326, 1067-1072.

Acknowledgements:

German-Israel Foundation

Israel Ministry of Science - National Lab. for Genome Infrastructure (knowledge center)

XenneX Inc. and The Yeda Fund; The Abraham and Judith Goldwasser Fund; The Klutznick and Cleaver Associates Fund; The Crown Human Genome Center