Doron Lancet

Chaim Kahana

Avital Adato
Rivi A.Goell
Nili Avidan
Jacques Beckmann
Edna Ben-Asher
Shifra Ben-Dor
Moshe Binyamin
Vered C. Caspi
Leon Esterman
Rhoda Frydman
Shirley H. Saban
Miriam Khen
Michal Lapidot
Tevie Mehlman
Tsviya Olender
Ron Ophir
Irit Orr
Eitan Rubin
Marilyn Safran
Alla Shainskaya
Orit Shmueli
Irina Solomon
Anna Weisman

Harnessing new DNA and protein technologies to cope with the genome revolution

Depts. of Biological Services & Molecular Genetics & The Crown Human Genome Center

Tel. 972 8 934 3683 / 2745 **Fax.** 972 8 934 4487 / 4113
E-mail: Doron.Lancet@Weizmann.ac.il Chaim.Kahana@Weizmann.ac.il

A. Harvesting the genome

The genome's first draft contains treasures of information, and is advantageous in the context of Israel's unique genetic populations. The Genome Center, together with the Bioinformatics and Biological Computing Unit (BBCU), provide scientists at Weizmann and elsewhere with an array of technologies and knowhow essential for such an endeavor. One important example of such activity is linkage analysis of a genetic disease among members of affected families, which enables defining a refined genomic interval for the disease-causing mutation. The WIS plays a central role in the transit from linkage mapping to disease gene discovery, by providing genome-related practical know-how as well as computer-intensive tools with regards to DNA sequence and variation analyses. Three genes for monogenic diseases have already been successfully identified; they cause mental retardation (Bargal R, et al., 2000, 2001) muscular disctrophy (Eisenberg I., et al., 2001) and heart failure (Lahat H., et al., 2000). As more genomic sequences become assembled and annotated, novel strategies for analysis of large genomic intervals and identification of multigenic diseases-causing polymorphisms are currently being developed.

B. The MassArray SNP Technology - Genetic Variation Association studies

Genetic variation analysis is the most important challenge of the Human Genome Project for pharmacogenomic applications and the elucidation of multigenic diseases. During the last years, various technologies have been developed for the study of Single Nucleotide Polymorphisms (SNPs), the most common type of genetic variation. In order to join the worldwide effort for such studies the WIS has recently acquired the Sequenom MassArray System, the technology of choice for very high throughput SNP analysis. Using Matrix Assisted Laser Desorption/Ionization Time-of-Flight (MALDI-TOF) mass spectrometry, the system measures target DNA associated with SNPs and other forms of genetic variation directly. By combining MassEXTEND primer extension chemistries with high-density SpectroCHIP Arrays, the MassARRAY system is capable of determining more than 4000 SNPs per day. Several Genetic Variation association studies have been started already at the WIS in collaboration with scientists at the institute as well as in other universities and companies.

C. Bioinformatics Systems Development

Among the bioinformatics tools developed is GeneCards (Rebhan, M., et al., 1998), an integrated database of human genes, that offers concise information and links about human genes, including synonyms, chromosomal location, protein and sequence data, orthologs (similar genes in other organisms), genetic disorders, research literature, SNPs, and more. It carries out automated data mining from a variety of sources, and allows sophisticated web-based queries. GeneCards is mirrored freely at 25 academic sites around the world, and is available for commercial use via a partnership with DoubleTwist Inc. The Unified Database (UDB) integrates information by interpolation from disparate maps of the human genome. UDB provides an important tool for coping with the fragmentary data of the First Draft of the Human genome. It is associated with CroW21, a more extensive mapping tool for the fully sequenced human chromosome 21. GESTALT (GEnomic Sequence Total Analysis and Lookup Tool) is a software workbench for genomic sequence analysis, comparison, and annotation, with a strong emphasis on visualisation. HORDE (The Human Olfactory Receptor Data Exploratorium) is a web site that provides information on 906 human OR genes and their associated proteins including structure, function and evolution.

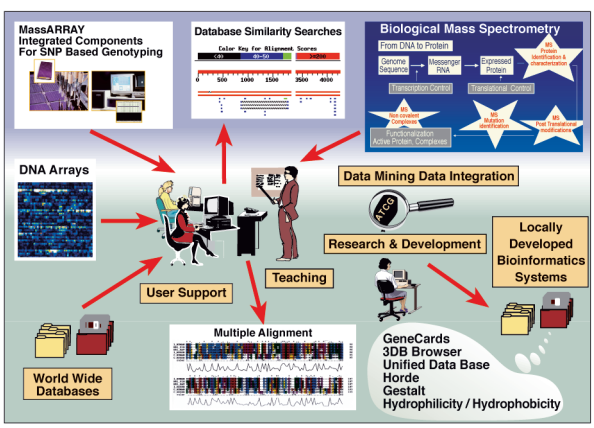

Fig. 1 *Genomic and post-genomic facilities.*

D. Bioinformatics Services

To support the ability of Weizmann scientists to use the rapidly growing wealth of biological data and tools, the Bioinformatics and Biological Computing Unit (BBCU) maintains a large professional team. It serves as a knowledge base, by offering consulting services for bioinformatics users and developers. BBCU organizes many bioinformatics training activities such as courses, workshops and seminars. The BBCU also identifies, purchases, and downloads useful data and software, as well as provides information about the availability and usefulness of these tools. The software includes Celera's CDS, Compugen's Gencarta, Accelerys' GCG, and more. Data sources include Genbank, EMBL, Swissprot, NCBI's NR, Transfac and over 50 additional databases. Finally, the BBCU maintains the infrastructure required to facilitate bioinformatics usage, including computers, software, data and mirror sites. Some of the BBCU services are supported through the Israeli National Node (INN), and are provided to the entire Israeli academic community.

E. DNA Chips

Some of the most exciting developments in genome technologies are in the realm of DNA arrays, known also as DNA chips. These methods are used mainly for profiling global gene expression in specific cells or tissues, but numerous other applications are feasible. For these purposes, the DNA Array unit harbors two sub-technologies: (1) DNA chips which are fabricated in-house by means of attaching hundreds or thousands of DNA fragments of choice onto a glass surface (such as a microscope slide). These arrays are used to query (by hybridization) RNA obtained from the cells tested. Results are monitored by a laser-based scanner, which detects signals of fluorescent probes. (2) the Affymetrix GeneChip system, for large-scale experiments. The pre-made arrays contain 200,000 oligonucleotides representing over 12,000 genes on a 1cm2 surface. Both sub-technologies are located jointly and provided as a service for the benefit of researchers.

Using the latter method, Project 40 has been launched, to elucidate the genome-wide patterns of gene expression in 40 human tissues.

F. Biological Mass Spectrometry

Genome sequence information is becoming available in unprecedented quantities. The absence of a direct functional correlation between gene transcripts and their corresponding proteins, however, represents a significant roadblock for implementation of this information.

Within the last few years, mass spectrometry has revolutionized analysis of proteins by providing rapid and sensitive identification of proteins in 1-D and 2-D gels. It is considered now to be one of the most sensitive and specific techniques for the identification and characterization of biomolecules, especially peptides and proteins.

The strategy of our facility is to provide our community with the equipment, tools, services and knowledge in the field of biological mass spectrometry. Our unit has launched three mass spectrometers: (1) Bruker REFLEX TM MALDI-TOF Mass Spectrometer, (2) API-3000 Triple Quadrupole and (3) the API-QSTAR Pulsar-i Quadrupole TOF with o-MALDI source electrospray mass spectrometers.

We perform mass spectrometric analysis for protein identification services, quality control of synthetic peptides and oligonucleotides, and for analysis of recombinant and mutant proteins. Some on-going projects enable the identification of post-translational modifications and analysis of hydrophobic proteins.

Selected Publications

Rebhan, M., et al., (1998) GeneCards: A novel functional genomics compendium with automated data mining and query reformulation support. Bioinformatics 14, 656-664.

Bargal R, et al., (2000) Identification of the gene causing mucolipidosis type IV. Nat Genet. 26(1), 118-23.

Bargal R., et al., (2001) Mucolipidosis type IV: Novel MCOLN1 mutations in Jewish and non-Jewish patients and the frequency of the disease in the Ashkenazi Jewish population. Human Mutation 17, 397-402.

Eisenberg I. , et al., (2001), The UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase gene is mutated in recessive hereditary inclusion body myopathy. Nat. Genet 29(1), 83-87.

Lahat H., et al., (2001) A Missense Mutation in a Highly Conserved Region of CASQ2 Is Associated With Autosomal Recessive Catecholamine Induced Polymorphic Ventricular Tachycardia in Bedouin Families from Israel. Am J Hum Genet (in press).

Acknowledgements

Supported by a Ministry of Science grants to the National Laboratory for Genome Infrastructure, and for an Indian-Israeli genome collaboration, Yan Hanadiv, The Krupp foundation, DoubleTwist Inc., Weizmann Institute Glasberg, Levy, Nathan Brunschwig and Levine funds, and the Abraham and Judith Goldwasser foundation.