



## Abstracts: Session I

been reported to range from very low to levels equal to those in Caucasians. We report the results of completed *BRCA1* and *BRCA2* analyses in 20 African American families at risk for breast or ovarian cancer. Families were selected on the basis of a history of breast cancer or breast and ovarian cancer and further subdivided into the following categories: high-risk (three affected first-degree relatives; ten families), moderate-risk (two affected first-degree relatives; seven families) and undetermined-risk (single affected first-degree relative, with medical information being updated). *BRCA1* and *BRCA2* germline alterations were first detected using a series of exon-specific polymerase chain reaction primers for single-strand conformation polymorphism analysis; this was followed by DNA sequencing of polymorphism variants. A limited number of *BRCA1* polymorphic intronic variants detected as a result of these studies were also analyzed for their effect on *BRCA1* mRNA splicing using an assay developed by Myriad Genetics. In this cohort we detected one protein-truncating mutation in either *BRCA1* or *BRCA2* (1/20; 5%). However, we detected splice mutations, missense mutations and several polymorphic variants in both *BRCA1* and *BRCA2*, with a much higher frequency in *BRCA2*. Many of these variants were both new and specific to African American patients; they also occurred in the absence of another disease-causing mutation. Moreover, a new *BRCA1* missense mutation in one of the high-risk families, exon 19(W1718C), seems to co-segregate with breast cancer. We will report the relative frequencies of these *BRCA1* and *BRCA2* variants in patient and control populations. These results agree with previous observations that deleterious mutations in *BRCA1* or *BRCA2* are uncommon in at-risk African American patients. They indicate that more benign variants should be further evaluated for their potential role in the disease process in these patients and that additional, as yet unidentified, genetic factors may contribute to breast cancer risk in African American families.

*Baxevanis, Andreas D.*

[19]

### An integrated approach to the extraction, storage, processing and analysis of microarray gene expression data

Andreas D. Baxevanis<sup>1</sup>, Izabela Makalowska<sup>1</sup>, Kenneth Trout<sup>1</sup>, Qien Zhou<sup>2</sup>, Zheng-Zheng Zhou<sup>2</sup>, Jaime Stein<sup>2</sup>, Edward R. Dougherty<sup>3</sup>, Paul S. Meltzer<sup>4</sup>, Yidong Chen<sup>4</sup>, Michael L. Bittner<sup>4</sup> & Jeffrey M. Trent<sup>4</sup>

<sup>1</sup>Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland

<sup>2</sup>NuTec Services, Stafford, Texas, USA

<sup>3</sup>M.D. Anderson Cancer Center, University of Texas, Houston, Texas

<sup>4</sup>Cancer Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland

A single microarray experiment provides thousands of individual pieces of data; it is therefore essential to focus on effective informatics methods in order to draw new biological conclusions efficiently and confidently. We have developed a unified software package that meets several important needs in the area of microarray data analysis. The software achieves several goals: (1) Integrating algorithms developed at the National Human Genome Research Institute into NuTec's GLEAMS software system, creating a single, integrated core platform that is capable of handling the computational aspects of complementary DNA and oligonucleotide array analysis, including imaging, signal processing, data extraction, database management and higher-order data analysis. (2) Developing new bioinformatics strategies to deduce the sequence- and structure-based relationships that are currently hidden within cDNA and oligonucleotide array data, including clustering methods and connections to gene ontologies. (3) Linking microarray data to data being amassed in clinical laboratories, to combine pedigree, phenotypic and genotypic data with gene expression data. (4) Using parallel computing to facilitate the analysis of microarray data. An overriding goal of this project is that the software be as user-friendly as possible, keeping in mind the level of com-

puter training of most of the biologists who will be using it as well as the hardware that is typically available in most laboratories.

*Beaulieu, Martin*

[20]

### High-throughput genotyping using candidate region mismatch scanning (PCR-CRMS)

Martin Beaulieu<sup>1</sup>, Ching Ouyang<sup>1</sup>, Garry Larson<sup>1</sup>, Louis Geller<sup>1</sup>, Steven Flanagan<sup>2</sup> & Theodore Krontiris<sup>1</sup>

<sup>1</sup>Division of Molecular Medicine, City of Hope National Medical Center, Duarte, California, USA

<sup>2</sup>Division of Neurosciences, City of Hope National Medical Center, Duarte, California, USA

To identify loci affecting disease susceptibility, linkage and association analyses are performed by scoring previously characterized sequence variation, such as microsatellites and single-nucleotide polymorphisms. However, such analyses can be expensive and time-consuming. Moreover, examination of particular candidate gene regions for potential new linkages and associations may be limited by the infrequent occurrence of known markers. The typing by currently available methods of a sufficiently large number of markers to achieve a robust analysis is often beyond the reach of the typical research laboratory. Various techniques that score either unknown or known sequence variation have been developed, yet none of these strategies is equally well suited for mutation detection and the scoring of genotypes. To perform both tasks, a technique should be both sensitive enough to detect all mutation types and sufficiently quantitative that the translation of the data into allele-sharing status is feasible. We investigated the possibility of replacing conventional genotyping with a mutation-detection approach. Genome mismatch scanning is a hybridization-based mutation-detection technique employing the *Escherichia coli* mismatch-detection enzymes MutHLS. It was originally developed to enrich for identical-by-descent regions between two whole genomic DNA samples. To accommodate polymerase chain reaction products of candidate regions, we have conducted a comprehensive biochemical optimization of the technique for target sizes ranging from 260 to 1,250 base pairs. Our modifications, which we now collectively designate polymerase chain reaction-candidate region mismatch scanning, have simplified the assay, rendered it quantitative and demonstrated its potential for cost-effective, high-throughput genotyping. Strategies for exploiting the method in the study of candidate regions or in genome-wide studies will be discussed.

*Belbin, Thomas*

[21]

### Molecular classification of head and neck squamous cell carcinoma using cDNA microarrays

Thomas Belbin<sup>1,2</sup>, Bhuvanesh Singh<sup>3</sup>, Ilana Barber<sup>2</sup>, Bruce Wenig<sup>2</sup>, Richard Smith<sup>2</sup>, Mike Prystowsky<sup>2</sup> & Geoffrey Childs<sup>1</sup>

<sup>1</sup>Department of Molecular Genetics, Albert Einstein College of Medicine, New York, New York, USA

<sup>2</sup>Department of Pathology, Albert Einstein College of Medicine, New York, New York, USA

<sup>3</sup>Head and Neck Service, Memorial Sloan-Kettering Cancer Center, New York, New York, USA

Head and neck squamous cell carcinoma (HNSCC) is the fifth most common cancer worldwide and accounts for 95% of the head and neck cancer cases in the Western world. Over the past three decades advancements in management have



produced equivocal survival benefits. We have used complementary DNA microarrays containing 9,216 genes to monitor gene expression in HNSCC tissue samples, adjacent normal margins and lymph node metastases. We extracted RNA from 46 surgically resected HNSCC tumor samples and compared it with cultured normal human adult keratinocytes. We are also comparing at least ten additional matched samples of primary and metastatic tumors from these same patients. Hierarchical clustering methodology has allowed us to classify these tumors on the basis of their gene expression patterns, and we are currently correlating these results with the clinical data collected for each patient. We have identified a group of genes that are generally overexpressed in these tumors, as well as genes that are diagnostic of specific tumor subtypes. Our results demonstrate that it is possible to classify HNSCC tumors based solely on global patterns of gene expression and to identify candidate genes that are the best predictors of tumor subtypes. We are studying these genes using other methodologies, such as tissue microarrays, to examine further their suitability as possible targets for drug treatment.

Bertoni, Francesco

[22]

### Construction of a chromosome 11q22–23 BAC/PAC contig to characterize deletions in lymphoproliferative disorders

Francesco Bertoni, Rebecca Auer, Roman Müllenbach, Chris Jones & Finbarr Cotter

*Department of Experimental Haematology, The Royal London and St. Bartholomew's School of Medicine and Dentistry, The Royal London Hospital, London, UK*

The application of technologies such as DNA microarrays and the ongoing human genome sequencing project will help the study of the genes underlying the pathogenesis of cancer. Chromosome 11q22–23 is frequently deleted in human solid and lymphoid neoplasms, such as breast and colorectal cancer, chronic lymphocytic leukemia and mantle cell lymphoma. Thus several cancer genes are likely to map to this region. To analyze the deletions in lymphoproliferative disorders, a PAC/BAC contig of chromosome 11q22–q23.3 was constructed, using as a guide the yeast artificial chromosome contig previously applied to characterize the deletions in patients affected by Jacobsen syndrome. BAC and PAC clones were identified and localized by screening of different human genomic libraries using the polymerase chain reaction, filter hybridization and database searches, using sequence-tagged sequences, expressed sequence tags or gene sequences known to map to the region of interest. The PAC and BAC end sequences were then used as new sequence-tagged sequences for database analysis or to design polymerase chain reaction primers after eliminating the repetitive elements using BLAST software. The assembled PAC/BAC contig stretches from D11S1897 to FRA11B. Construction of a transcript map to identify possible cancer genes is continuing. The ordered genomic clones will be used to analyze DNA samples of lymphoid tumors to determine regions of chromosome loss or amplification within the region of the contig using DNA microarray technology or the Invader assay.

Berx, Geert

[23]

### E-cadherin controlling epithelial differentiation in human carcinomas

Geert Berx, Kristin Strumane, Joke Comijn & Frans Van Roy

*Molecular Cell Biology Unit, Department of Molecular Biology, VIB-Ghent University, Ghent, Belgium*

Acquisition of invasive and metastatic capabilities is frequently associated with loss of cell–cell adhesion. One of the major constituents of the cell–adhesion complex in epithelial cells is E-cadherin, a transmembrane protein involved in homotypic cell–cell adhesion. E-cadherin exerts a potent invasion-suppressing role in experimental tumor cell systems. Partial or total loss of E-cadherin expression apparently occurs in the majority of human carcinomas. Several different mechanisms seem to cause the loss of E-cadherin function, including transcriptional downregulation and mutational inactivation. E-cadherin mutations are already present at the early noninvasive *in situ* stage of lobular breast tumors, and E-cadherin germline mutations in cases of familial diffuse gastric cancer argue strongly in favor of a real tumor suppressor role for E-cadherin, in addition to its role as an invasion suppressor. The variety of effects on E-cadherin expression (such as epithelial cell morphology, cell polarization, and growth inhibitory and tumor suppressive effects) is not solely the result of cell–cell adhesion but implies that E-cadherin expression mediates complex signaling processes and transcriptional regulatory events. Therefore we set up cell models that mimic the E-cadherin deficiency in human carcinomas and analyzed them for differences in messenger RNA expression, using molecular indexing and microarrays. We are currently analyzing genes with differential expression in these matched tumor cell pairs (with and without E-cadherin expression) to determine their specific role in growth control and loss of cell differentiation during epithelial tumor progression.

Bharucha, Vandana

[24]

### Analysis of the CBFA2, CBFA2 and AF-9 genes in familial acute myelogenous leukemia

Vandana Bharucha, Asim Belgaumi, Kishor Bhatia, Hassan El Solh & Colin Hodgkinson

*King Faisal Hospital and Research Centre, Riyadh, Saudi Arabia*

The cloning of chromosomal translocation breakpoints from sporadic leukemias has identified genes that are implicated in the pathogenesis of acute myelogenous leukemia (AML). Although common, such chromosomal translocations are not necessary for the development of leukemias, suggesting that other underlying genetic factors are involved in the early stages of leukemia development. Familial AML is extremely rare, and families with the disease provide a unique resource for the study of the genes involved in the initial stages of cancer development. We investigate a single family in which several members have developed AML. Three members (a mother and two offspring) are affected with AML M4eo, and a third offspring is affected with AML M2. The average age of onset in the children (7 years) is significantly lower than that for the mother (30 years). Transmission of the disease seems to be in an autosomal-dominant manner. The small size of the pedigree reduces the power of genome-wide linkage analysis and therefore we used a candidate gene approach. Initially three genes were selected for analysis based on their reported involvement in familial and sporadic AML: *CBFA2*, *CBFB* and *AF9*. We performed genotyping at these loci to identify shared haplotypes between affected individuals. We will also report the results of direct mutation analysis of the candidate genes.