

**Abstracts: Session III**

Reboul, J.

[13]

**The gene number dilemma: Direct evidence for at least 19,000 protein-encoding genes in *C. elegans* and implications for the human genome**J. Reboul<sup>1</sup>, P. Vaglio<sup>1</sup>, N. Tzellas<sup>1</sup>, C. Jackson<sup>2</sup>, T. Moore<sup>2</sup>, Y. Kohara<sup>3</sup>, J. Thierry-Mieg<sup>4</sup>, D. Thierry-Mieg<sup>4</sup>, J. Hitti<sup>5</sup>, L. Doucette-Stamm<sup>5</sup>, J. Hartley<sup>6</sup>, G. Temple<sup>6</sup>, M. Brasch<sup>6</sup>, D.E. Hill<sup>1</sup> & M. Vidal<sup>1</sup><sup>1</sup>Dana-Farber Cancer Institute and Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA<sup>2</sup>Research Genetics, Huntsville, Alabama, USA<sup>3</sup>Center for Genetic Resource Information, National Institute of Genetics, Shizuoka, Japan<sup>4</sup>National Center for Biotechnology Information, National Institutes of Health, Bethesda, Maryland, USA<sup>5</sup>Genome Therapeutics Corporation, Waltham, Massachusetts, USA<sup>6</sup>Life Technologies Inc., Rockville, Maryland, USA

Estimates of the number of genes of organisms with a wide range of biological complexity will emerge from genome sequencing projects, and this information will be crucial in correlating gene number with biological complexity. Computer-based annotations of the genome sequences of *Saccharomyces cerevisiae*, *Caenorhabditis elegans* and *Drosophila melanogaster* predict approximately 6,000, 19,000 and 13,600 genes, respectively. It is paradoxical that *C. elegans* contains 50% more genes than *D. melanogaster* and that the latter contains only twice as many genes as *S. cerevisiae*. However, these numbers should be considered with caution until gene predictions are verified using experimental approaches. To verify gene predictions experimentally, we amplified predicted open reading frames from a *C. elegans* complementary DNA library using the polymerase chain reaction, cloned them and generated an open reading frame sequence tag (OST) for each of them. In a random sample ( $n=1,200$ ) of the approximately 10,000 genes predicted from the entire *C. elegans* genome but experimentally unverified so far, 90% of these predicted genes could be unambiguously verified by OST sequencing. When added to the list of approximately 9,500 genes previously verified by individual laboratories, by expressed sequence tag sequencing or both, our data confirm the existence of at least 19,000 protein-encoding genes in *C. elegans*. We suggest that ORFeome cloning coupled with OST analysis could be used to verify gene number predictions for other organisms. Furthermore, the demonstration and identification of 10,000 new genes provides a resource for comparative genomics and gene expression studies. We also discuss our findings in the context of the current debate on the human gene number.

Reeve, Anthony

[14]

**Microarray analysis of colorectal cancer cell lines**Aaron Jeffs<sup>1</sup>, Sharon Pattison<sup>2</sup>, Junjie Xu<sup>2</sup> & Anthony Reeve<sup>2</sup><sup>1</sup>Otago Genomics Facility, Department of Biochemistry, University of Otago, Dunedin, New Zealand<sup>2</sup>Cancer Genetics Laboratory, Department of Biochemistry, University of Otago, Dunedin, New Zealand

The incidence of colorectal cancer (CRC) in New Zealand non-Maori women is recognized as the highest in the world, and the highest global rates of CRC occur in New Zealand and Australian populations. We are profiling the gene expression of

CRC cell lines using complementary DNA microarrays to provide baseline data for a clinical investigation of the genetic events involved in metastasis of this disease. The genetic mechanisms underlying metastasis of CRC are at present poorly understood. The development of metastasis significantly reduces five-year survival of CRC patients from at least 70% in those with no metastatic involvement to 35–65% in those with lymph node involvement and 5% in those with distant metastasis. We have generated cell lines from patient tumor samples with varying Dukes classifications of disease pathology. Of particular interest are two cell lines from the same individual, obtained one year apart, that are derived from non-metastatic CRC at the Dukes B stage and metastatic CRC at the Dukes C stage. Gene expression data comparing the metastatic and nonmetastatic cell lines will be presented. Greater understanding of the molecular processes involved in the initiation and progression of CRC may lead to new clinical options for CRC screening and treatment, and for prediction and prevention of metastatic events in CRC.

Riggins, Gregory J.

[15]

**Comprehensive gene expression analysis as a tool for hunting cancer genes**Gregory J. Riggins<sup>1</sup>, Anita Lal<sup>1</sup>, W. Troy Loging<sup>1</sup> & Robert L. Strausberg<sup>2</sup><sup>1</sup>Duke University Medical Center, Durham, North Carolina, USA<sup>2</sup>Cancer Genome Anatomy Project, Office of the Director, National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA

Nearly three million messenger RNA transcripts have been catalogued by tissue origin to create one of the largest free databases for cancer gene expression analysis. Seventy different cell populations were studied by serial analysis of gene expression to provide precise gene expression levels in brain, breast, colon, ovarian and prostate cancers and corresponding normal tissues. Additionally, several model systems have been analyzed, providing insight into the genes transcriptionally regulated by fundamental oncogenic processes, such as angiogenesis, oncogene amplification and telomere maintenance. These data, digital northern displays and comparison tools for virtual experiments are made available by the Cancer Genome Anatomy Project at <http://www.ncbi.nlm.nih.gov/SAGE/>. Despite the many challenges inherent in applying large-scale gene expression data, we focus on two examples in which application of project data has yielded progress. First, mining this data readily provides candidate tumor markers or tumor antigens for the major tumors represented in the database. Using fluorescent real-time polymerase chain reaction and western blotting we showed that the mined tumor markers are expressed in an independent set of tumors and not in other normal tissues. Large-scale gene expression analysis may therefore provide a means for identifying therapeutic or prognostic targets. Second, analysis of a cell culture model for telomere maintenance (an important function of transformed cells) has provided insight into genes not previously implicated in this process. This example helps demonstrate that functional genomics on the basis of comprehensive gene expression analysis is a useful tool for hunting cancer-related genes.