

# “Honorary Biomedical Sciences & Health Information Lecture Series”



## **STRATEGIES FOR PROTEOMIC PROFILING OF CANCER SPECIMENS AND PLASMA IN MOUSE MODELS OF HUMAN CANCERS: IDENTIFICATION OF NOVEL ALTERNATIVE SPLICE ISOFORMS AS A NEW CLASS OF BIOMARKER CANDIDATES**

### *A Distinguished Lecture*

By

**Dr. Gilbert S. Omenn**

*Professor, Internal Medicine, Human Genetics, and Public Health  
Director, UM Center for Computational Medicine and Biology  
University of Michigan, Ann Arbor, MI, 48109, USA*

**Abstract:** Alternative splicing generates protein diversity without significantly increasing genome size. Most genes have alternative splice variants. Many alternative splice variants are associated with or cause diseases. Alternative splice databases are publicly available. Extensive proteomic analyses have been performed on specimens from genetically-designed mouse models of specific human cancers. We have developed an analytical pipeline to identify and quantify known and novel alternative isoforms from peptide sequences determined by mass spectrometry-based proteomic informatics. Tandem MS/MS spectra were interrogated against a non-redundant database containing an exhaustive 3-frame translation of Ensembl transcripts and gene models from ECgene using X!Tandem software [Fermin et al. Genome Biology 2006;7(4):R35]. The search results were then processed for peptide to protein integration by Trans Proteomic Pipeline (TPP) analysis and our own Michigan Peptide to Protein Integration (MPPI) method [Menon et al Cancer Res 2009;69:300-309]. All peptides matching to our final integrated protein list were searched against the NCBI-nr database using blastp and against the mouse genome using UCSC BLAT browser. Proteins whose peptide sequences aligned perfectly to an existing canonical mouse coding sequence were removed from the list of alternatively spliced protein identifications. We applied the database and the methods to a very large proteomics dataset from plasma of mice with activation of K-Ras (G12D) and deletion of Ink4a/Arf, a genetically-designed model of pancreatic ductal adenocarcinoma in humans. We identified 328 known and 92 novel splice variants from multiple high-scoring spectra (X!Tandem expect value <0.001). We validated all of seven of those novel variants by reverse transcription polymerase chain reaction (RT-PCR). Some of the 92 novel variants were differentially expressed between wild-type and tumor-bearing mutant mice [Menon et al, 2009]. We present similar work in progress on the Her2/neu breast cancer model, with 551 known and 33 novel alternative splice variants in the tumor tissue, compared with wild-type. The exact splicing findings and the biological annotations for these cancers are quite interesting. Alternative splice variants may occur also with the corresponding human cancer and be useful biomarkers for diagnosis and prognosis in a high-throughput mode. This biomarker discovery strategy complements primary MS/MS findings in tumors and in plasma, as well as finding auto-antibodies that amplify tumor protein signals.

**Acknowledgements:** Supported by NCI/SAIC 23XS110A Mouse Models of Human Cancers, MTTC GR 687 Proteomics Alliance for Cancer Research, and U54 DA021519 National Center for Integrative Biomedical Informatics.