



Genomic & Proteomic Technological Advances in Cancer Research

Genomic & Proteomic Technological Advances in Cancer Research
Co-sponsored by IBC Life Sciences and Center for Cancer Research, NCI, USA.
February 10–11, 2003, Bethesda, MD, USA.

Javed Khan

Principal Investigator, Oncogenomics Section, Pediatric Oncology Branch, Center for Cancer Research, National Cancer Institute, Advanced Technology Center, Room 134E, 8717 Grovemont Circle, Bethesda, MD 20892-4605, USA.

Tel: +1 301 435 2937; Fax: +1 301 480 0314; E-mail: kxanjav@mail.nih.gov

The initial sequencing of the Human Genome Project (HGP) is essentially complete and we have identified most of the letters and 'words' of the human genome in what Francis Collins refers to as the 'Book of Life' [1,2]. Now we have entered into the more complex post-genomic era and the challenge now remains to decipher the biological meaning of these words contained within this complex book, and eventually to 'translate' this to benefit mankind in a tangible way. The purpose of this conference was to summarize the technological advances in the fields of genomics, microarrays, mass spectrometry, genotyping and single nucleotide polymorphism (SNP) analysis for the purpose of preventing, treating, diagnosing and predicting the prognosis of cancer. There were five distinct but inter-related sessions. Following the opening and keynote address, the first group of delegates focused on the central role of bioinformatics to analyze high dimensional genomic-proteomic data. This was followed up by a section on proteomics applied to identifying biomarkers and surrogate end point markers. The second day focused on translational genomics, i.e., how we are beginning to take these new technologies to clinical practice, while the focus of the third and last day was on pharmacogenomics and the application of SNP detection for cancer research.

Introduction

Dr J Carl Barrett (Center for Cancer Research, NCI, NIH, Bethesda, MD, USA) gave the opening address and set the theme for the conference. He outlined several directions of the Center for Cancer Research (CCR), the intramural component of the National Cancer Institute (NCI), to maximally reduce the impact of cancer on the health of the nation. He reasoned that cancer prevention and early detection should be primary goals in cancer research because early detection results in patients presenting with low stage disease at which time the cancer is likely to be curable. In addition, the vast majority of patients present with advanced or currently incurable disease, and in order to improve the survival rates of these patients, new treatment strategies need to be developed, and the promise of a new generation of rationally designed molecularly targeted drugs is looming on the horizon. Early detection and identification of molecular targets may be achieved by both proteomic and genomic-based technologies.

Although the phrase 'molecularly targeted therapy' is currently in vogue, several chemotherapeutic agents are already in use that target specific cancer proteins and pathways, including:

- methotrexate (targeting DHFR)
- 5-fluorouracil (thymidylate synthetase)
- etoposide (topoisomerase II)

However, many of these targets are also expressed in normal cells and thereby inhibit all dividing cells causing severe toxicity. Hence the requirement for drugs that inhibit proteins that are only expressed in diseased tissues is paramount. In addition, for most currently utilized chemotherapeutic drugs we have reached the maximally tolerated doses. Due to the complexity and redundancy of the human genome, different cancers of the same type may mutate different genes and genomic-proteomic research should be geared towards identifying not just the genes but also the pathways that are disrupted. Combined therapy targeting several proteins along one or multiple pathways may not only effectively eliminate the cancer but may also prevent the emergence of resistant clones. One possible achievable goal is to turn cancer into a chronic disease whereby patients live with their tumor, which has been rendered unable to grow.

Dr Joseph Schlessinger (Yale University School of Medicine, Newhaven, CT, USA) gave the keynote speech. He spoke on signal transduction molecules, which are perhaps the most ideal targets for cancer therapeutics, particularly the protein tyrosine kinases (PTK). Of the estimated 32,000 genes in the human genome, 6400, or 20%, are currently known to be involved in the signal transduction or growth factor pathway. There are 518 protein kinases (PK) and 180 protein phosphatases (see [3] for an excellent review). There are many examples of both gain and loss of functional mutations of PTK as a cause for human disease, including EGFR and ERBB2-4 (both of which have been found to be amplified and mutated in many cancers). In addition, mutations of FGFR1-4 have been found in bone diseases as well as leukemia, lymphoma

and carcinoma. Several kinases are also involved in chromosomal translocations in cancers, for example, ABL1 in chronic myelogenous leukemia (CML). All these activated PTKs represent potentially 'druggable' targets. Two molecularly targeted drugs against PTK that are currently in clinical trials include Herceptin™ (Genentech, South San Francisco, CA, USA), a humanized monoclonal antibody (mAb) against ERBB2, and STI-571, or Gleevec™ (Novartis, East Hanover, NJ, USA), a drug that inhibits the kinase domains of KIT, platelet derived growth factor receptor, and ABL. In order to identify new drugs for the remaining PK we must first establish their 3D protein structures. Additionally, it is important to identify all of the proteins involved in this signal transduction cascade as well as parallel redundant pathways, since the PKs do not work alone. Dr Schlessinger suggested that we should think of the signal transduction pathway as an electronic circuit and, in order to model these PKs, biologists will need to work with systems engineers.

The complexity of the human genome was reiterated by Dr John Weinstein (CCR, NCI, Bethesda, MD, USA) as he laid out the significant challenge for the scientist: to identify the key players in cancer biology in a sea of 30,000–60,000 genes, 100,000–150,000 splice variants, and 500,000–2,000,000 protein states. He described several tools that are available for the exploration of genomic–proteomic data [101].

Use of transcriptional profiling for compound selection and drug development

Paul Kayne (Bristol-Myers Squibb, Princeton, NJ, USA) spoke on the challenge of identifying good candidate drugs for therapy. In this post-genomic era there is a need for a paradigm shift from the traditional use of toxic, non-specific empirical compounds to non-toxic, specific, molecularly targeted ones. Compound selection is a long and arduous process involving chemistry, computational analysis, structural modeling, assay design, screening for hits, secondary assays, and

toxicity profiles. New compounds are synthesized on the basis of these searches and are fed back into the loop and the process is repeated until a decision is made as to which drug to take to the clinic and at which time point. Currently, only about 10% of compounds make it to clinical trials and therefore we need to develop new *in vivo* and *in vitro* tests to rapidly identify the best drugs and avoid expensive failures. There is also a need for early warning tests that can detect serious and potentially lethal side effects prior to their occurrence. Finally, it is important to predict idiosyncratic toxicity (incidence ~1 in 10,000), which is probably related to the genotype of the patient. In his talk, Paul Kayne asked the question: Can we predict that a given drug will demonstrate toxicity by *in vitro* experimentation prior to introducing it to the clinic? The model his group utilized was liver toxicity caused by PPAR- γ agonists. He utilized microarray expression data generated from hepatocytes exposed to known hepatotoxic reagents. His group found that liver toxicity did correlate with the expression level of a set of genes and the total number of genes that changed correlated with the degree of hepatotoxicity. To identify potential surrogate markers for predicting and monitoring hepatotoxicity, they are planning to perform proteomic analysis using liquid chromatography/mass spectroscopy (LC/MS) on the media in which the cell cultures are growing in the presence of these drugs to identify secreted proteins. However, the data presented was preliminary and several questions remain, such as what is the optimal time point to monitor the gene expression, what is the best cell line to test, how many lines should be tested and will it actually translate to the clinic? Nevertheless, these methods show promise for the development of biomarkers and surrogate tests for predicting the toxicity profile of a compound as well as patient monitoring.

Proteomic analysis of biomarkers and surrogate end points

The ability to profile proteins in human serum reliably and compare these pro-

files in normal and diseased populations offers a path to biomarker discovery. Dr Lance Liotta (CCR, NCI, Bethesda, MD, USA) presented his vision of the future of personalized molecular medicine using proteomic approaches. He focused on early cancer detection through serum proteomics. He suggested that cancer therapeutics would be revolutionized by proteomics such that we will soon be able to deliver combinatory treatments of cocktails of drugs tailored to individual molecular profiles. He proposed that proteomic approaches would allow us to identify and target multiple proteins along the length of the key signal transduction pathways.

He presented his previously published work on serum protein pattern diagnostics, which was done in collaboration with the US Food and Drug Administration (FDA) [4]. He analyzed proteomic spectra generated by mass spectroscopy (surface-enhanced laser desorption ionization or SELDI) with genetic algorithms and self-organizing cluster analysis, and was able to accurately diagnose ovarian cancer.

He and others have performed similar studies on adenocarcinoma of the colon, lung and breast. An improvement over the original SELDI method has been the use of newer MS technologies such as the Applied Biosystems (CA, USA) QSTAR™ Pulsar Q TOF MS system which has a higher resolution and is also able to directly identify differentially expressed proteins. This may lead to the development of an ELISA test for these proteins. The NCI is currently developing a reference laboratory to use these methods for diagnostic purposes. Other investigators have used proteomic profiles from non-serum body fluids for early diagnosis of cancers. Dr Tatyana Zhukov (University of South Florida, Tampa, FL, USA) presented her study on performing SELDI on preserved sputum to diagnose lung cancer. She found the specificity was 100% but sensitivity was only 75%, which is too low to be used as a diagnostic test. However, when these results were combined with sputum cytology and immunohistochemistry

there was an improvement in the sensitivity to detect lung cancer.

Dr Sushmita Mimi Roy (SurroMed, Mountain View, CA, USA) presented a very nice example of using proteomics-based approaches on serum samples to identify biomarkers of disease using asthma as a disease mode. The SurroMed proteomic platform is based on taking a very small amount of body fluid (3–5 µl), which is fractionated according to size by a variety of methods followed by gas chromatography (GC)-or Electro-spray Ionization-Time of Flight (ESI-TOF) performed on the different eluted fractions. After sophisticated data analysis, comparing normal and patients with asthma, they are able to identify the differentially expressed proteins by examining the different peaks in the spectral data. With this method they have confirmed the increased levels of haptoglobin in patients with asthma. These types of proteomic analysis on body fluids, on a host of diseases including cancer, are likely to detect novel biomarkers as well as surrogate markers that correlate with diagnosis, prognosis, response to treatment, and early detection of toxicities of therapy.

Translation of genomics–proteomics into clinical practice

Dr Karol Sikora (AstraZeneca, Mereside, Cheshire, UK) was the chair and first presenter of the session on integrating genomic–proteomics to clinical practice. It is clear that there will be a dramatic increase in the incidence of cancer worldwide over the next 20 years due to the aging population. Considerable advances have been made in local therapy for minimally invasive cancers as well as improvements in the delivery of radiotherapy using sophisticated computer imaging and treatment planning. However, the most promising advances have been the increasing understanding of the molecular genetics of cancer fuelled by advances in our knowledge of the human genome. This will have a significant impact on prevention, screening, diagnosis and treatment of cancer and herald a ‘golden age’ of drug discovery.

Over the past 5 years, there has been an amassing of information from sequence data to biological processes in oncogenic and tumor suppressor pathways. This has been the fertile ground for the hunt for rationally based anti-cancer drugs. There is now a record number of novel compounds currently in clinical trials and most of the Phase I drugs are molecularly targeted. Dr Sikora’s speculation is that over the next few years precise molecular targets will be identified in specific cancers and consequently, traditional empiric chemotherapy will give way to tumor specific protein-targeted therapy, thereby increasing effectiveness and reducing toxicity through increased specificity and selectivity.

The dream outcome for patients and the pharmaceutical industry is that these drugs will make cancer a chronic illness and that we will have a society that is willing to pay for these. Dr Sikora predicted that there will be several new molecularly targeted drugs by the year 2010 for the four main cancers (breast, colon, lung and prostate), that this will become a US\$64 billion market and that within 25 years cancer will indeed become a chronic controllable disease.

DNA microarrays for cancer diagnosis and prediction of prognosis

DNA microarrays are being increasingly used for diagnostic classification of cancers [5] as well as for predicting prognosis [6,7]. Drs Louis Staudt (CCR, NCI, NIH, Bethesda, MD, USA) and William Evans (Memphis, TN, USA) presented their ground breaking studies on the applications of microarrays for these purposes for diffuse large B cell lymphoma [6] and childhood acute lymphoblastic leukemia (ALL) [8], respectively.

Over the past 20 years, event free survival of childhood ALL has improved from 50–80% simply by using old drugs in more optimal ways without the discovery of new antileukemia agents. However, there remains 20% of patients that do not respond to treatment that may benefit from the field of genomics

in a broader sense. Dr Evans’ group used data from DNA gene expression studies to develop a molecular classifier and were able to predict relapse and the risk of developing secondary acute myelogenous leukemia (AML). Treatment of pediatric acute lymphoblastic/lymphocytic leukemia (ALL) is currently based on the concept of tailoring the intensity of therapy to a patient’s risk of relapse using clinical, pathological and cytogenetics parameters. He showed, however, that using the gene expression levels of as few as seven genes he was able to predict the risk of relapse in T-ALL with > 97% accuracy, and was able to predict the risk of secondary AML in TEL-AML1 positive ALL with 100% accuracy. However, the caveats of these and other prognosis studies are that the numbers of patients analyzed are relatively small and the predictions have not been verified with independent test sets or in prospective trials. One gene, FLT3, was identified as predicting relapse in hyperdiploid ALL. Inhibitors for FLT3 now exist, which opens up the prospect of using these agents for this subgroup of patients.

Application of proteomics for diagnostics

Jorge Leon (Aventis Pharma, Cambridge, MA, USA) addressed the financial implications of using proteomics to discover diagnostic markers. The US diagnostic market is worth about US\$1.5 billion, and will grow to a US\$3 billion market by the year 2007. Current homebrew tests, in which the consumer purchases their own in house self-diagnostic test, have produced a US\$10 billion market. Return on investment in a diagnostic program is greater than a successful drug development program with significantly lower risk. He estimated that it costs US\$300 million and 10–15 years to develop a particular drug with an investment return of US\$300 million per year, and a good drug company produces a new drug every 2–3 years. In contrast, a diagnostic test costs US\$10 million to develop and takes 3 years, but the return is US\$50 million per year in revenue.

This area of research is important because there are several problems with the current diagnostic screening tests. For example, Pap smear misses 10–20% of the time in the most sophisticated labs. The prostate specific antigen (PSA) test is a very lucrative test but still has a significant false positive and negative detection rate and still misses some prostate cancer (PCA). For many cancers there is currently no surrogate marker for diagnosis or to predict response or toxicity. For diagnostic markers to be useful they need to be able to both diagnose a condition and offer options to tailor therapy. A good example of this are the kits available that confirm high expression of ERBB2 in breast cancer to the guide physicians as to which patients will benefit from Herceptin treatment. Finally, for a test to be useful it has to have > 90% clinical sensitivity.

Mouse models of cancers

The mouse is currently the pre-eminent experimental mammalian model for the study of cancer biology, aided by the recent announcement detailing the analysis of the draft mouse genome [9]. Some of the advantages of using mice as models include their uniform genetic background, the ability to study the effects of knockouts of single genes, e.g., tumor suppressor genes, and the ability to make transgenics or knock-ins for the study of specific oncogenes and growth factors. Many mice are available from the NCI repository [102] and from the Jackson Laboratory [103]. Several are relevant to human cancer including *erbB2*, *myc*, *RAS*, *PyMT* (MMTV-polyoma middle-T), *Notch4*, and *cyclinD1* transgenic mouse models as well as knockouts of *p53* and *RB*.

Drs Jeffrey Green and Thomas Reid (both CCR, NCI, NIH, Bethesda, MD, USA) presented some data on the characterization of several mouse cancer models, using molecular cytogenetics and genomic approaches, including DNA microarrays, fluorescent *in situ* hybridization (FISH), spectral karyotyping (SKY) and comparative genomic hybridization (CGH), and compared them to their human coun-

terparts. Dr Reid found that mouse models for epithelial cancers (*Brca1*^{-/-}, *MMTV-V-myc*, *erb2*, *SV40Tag*) revealed centrosome amplifications similar to human cancer, however those induced by oncogene overexpression had fewer recurring imbalances and the pattern and distribution of imbalances was most similar to human cancers in the conditional *Brca1*^{-/-} mice. Through these studies, he concluded that mouse models can be used for cloning of cancer-associated genes and the genetic similarities to human cancer suggest that mouse models can be useful preclinical models.

Pharmacogenomics

Roberto Weinmann (Bristol-Myers Squibb, Princeton, NJ, USA) and Dr Stephen Chanock (CCR, NCI, NIH, Bethesda, MD, USA) discussed the impact that the integration of genomics will have on predicting drug response to therapy. A classical example of how genomics has impacted cancer therapy is in leukemia therapy where there has been the discovery of the polymorphism of the thiopurine methyltransferase (TPMT) gene which catalyzes the S-methylation of thiopurine drugs such as 6-mercaptopurine. Individual variation in the toxicity and therapeutic efficacy of thiopurine drugs is associated with a common genetic polymorphism that controls levels of TPMT activity. Genetic polymorphisms in the TPMT gene are such that about 90% of Caucasians have high TPMT activity, 10% have intermediate activity, and 1 in 300 individuals have low activity. The test for this polymorphism is now a polymerase chain reaction (PCR)-based Clinical Laboratory Improvement Act (CLIA)-certified test that allows clinicians to distinguish between the low and high drug metabolizers and reduce dosage accordingly for those who might have severe hematopoietic toxicity due to slow drug elimination, and increase dosage accordingly for those with higher metabolism. Tests for other candidate genes involved in the metabolism of various chemotherapy agents are currently being developed.

There are ~1.5 million known SNPs and although good examples of polymorphisms that can guide therapy exist, e.g., TPMT, for the majority of drugs there are currently no methods to predict their metabolism rates in a cost effective way.

One goal of pharmacogenomics profiling for drug development is to identify surrogate markers that can predict the pharmacodynamic (PD) efficacy of a drug so that it can be rapidly transitioned to the clinic.

Dr Weinmann also raised the question, how does one develop an assay or marker to monitor the effectiveness of novel compounds in preclinical development? A good PD marker for an assay must be expressed in both target and a surrogate tissue, e.g., peripheral blood mononuclear cells (PBMC). It must then be clinically validated and detectable in high-throughput tests.

Currently, transcription profiling by microarrays offers an attractive way of pharmacogenomic profiling compounds and developing PD markers, since each pathway has a series of transcriptional reporters on the microarrays.

The strategy outlined by Bristol-Myers Squibb to develop markers for a molecularly targeted compound involves assaying cell lines, xenografts and surrogate tissues (e.g., PBMC) treated with the drug, antisense, monoclonal antibodies, and RNAis etc. to knock down the gene and look at common genes that change by microarray experiments. They then replicate the results using reverse transcription-polymerase chain reaction (RT-PCR) for these genes and ultimately perform clinical trials to validate these markers

Next, Dr Weinmann presented some preliminary work to attempt to define a set of genes that may predict whether a breast cancer will respond to a drug. For a given drug, the strategy was to determine the expression profile of five sensitive and six resistant cell lines and identify the genes that distinguish these two groups. Following this, they treated all of the cell lines with the same drug, and identified genes that are downregulated in the majority of the sensitive cell

Highlights

- The sequencing of the genomes of several species are either completed or on the way to completion.
- Increasing annotation of genes and their function.
- Expansion of computing power and understanding of high dimensional data has allowed more extensive data mining.
- Several affordable proteomic approaches are emerging to identify differentially expressed proteins in cancer versus normal.
- A compendium of differentially expressed genes/proteins in cancer and normal samples are being developed using genomic and proteomic approaches.

lines. Next they performed expression profiles on a series of breast tumors and used these two discriminatory gene lists to classify the tumors by simple hierarchical clustering. Although the tumors they presented appeared to cluster into two main groups, what remains unknown is whether the separation has any biological or clinical relevance. Regardless of the outcome of this particular study, their proposed strategy has potential application and requires further validation.

Expert opinion and outlook

As we look toward the future one cannot help but contemplate as to whether we are really heading towards a futuristic world such as the one portrayed by the movie ‘Gattaca’, which, set in the twenty-first century, depicts a society which predetermines and dictates the life of every individual based on whole genome screening. During the conference there was extreme optimism about the integration of genomic–proteomic based approaches to medicine. We heard that very soon, using sophisticated genomic–proteomic profiles, we would be able to predict the probability of developing a certain disease, have a precise molecular diagnosis, deliver individualized therapy based on these profiles, and predict whether you will live or die

from the disease. We heard : “There is a new target.....new drug....but it is still round the corner! There are newer and more innovative technologies emerging. Tomorrow we will have personalized medicine”. Several of the speakers have predicted that cancer will become a chronic illness by 2025.

From this prophetic golden outlook let us step down to earth momentarily and discuss some creases that still need some ironing. Firstly, there is still a lack of standard operating procedures for sample acquisition by most surgeons and pathologists that are in the cutting edge of patient care. A wide variety of artifactual changes can be introduced by differences in sample handling procedures such as delays in freezing to allow samples to thaw during transport. Secondly, too often there is a lack of ‘good’ clinical information accompanying the samples, which is crucial to perform correlative studies. Thirdly, there are multiple platforms in which microarrays are performed (Affymetrix, cDNA, Oligoarrays, Compugen, Operon, Agilent, Amersham/Motorola etc.) and it has been difficult for one investigator to compare his results with another. Fourthly, there are no standard methodologies for microarray experiments, i.e., direct-, indirect labeling, amplification, various controls etc. Lastly, as more and more diagnostic or prognostic signature studies are performed and published with small arrays and relatively small numbers of patients, questions remain as to whether these ‘diagnostic & prognostic’ signature genes truly are the best ones. By the very evolving nature of the data released by the HGP, more and more genes are being discovered as well as their splice variants, so that many of the currently utilized microarrays are incomplete. Therefore, are we ready to take these to the clinic to determine therapy? Should we wait for the best set of genes? How do we define the best? What is the best platform to use? In the realm of therapeutics, our knowledge of gene ontology and pathways is still in its infancy and each microarray may yield tens of potential targets. Which ones should we choose to target?

Although these are significant issues that need addressing, there is nevertheless cause for optimism in the future for cancer therapeutics. In the author’s opinion, the human genome project has indeed heralded the golden age of drug discovery and has impacted the way that medicine is practiced and will be practiced in the near future.

Bibliography

Papers of special note have been highlighted as either of interest (•) or of considerable interest (••) to readers.

1. Human Genome Special. *Science* 291 (2001).
2. The Human Genome. Genomic Special. *Nature* 409 (2001).
3. Blume-Jensen P, Hunter T: Oncogenic kinase signalling. *Nature* 411, 355-365 (2001).
4. Petricoin EF, Ardekani AM, Hitt BA *et al.*: Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* 359, 572-577 (2002).
5. Khan J, Wei JS, Ringner M *et al.*: Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat. Med.* 7, 673-679 (2001).
6. Rosenwald A, Wright G, Chan WC *et al.*: The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N. Engl. J. Med.* 346, 1937-1947, (2002).
7. van 't Veer LJ, Dai H, van de Vijver MJ *et al.*: Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415, 530-536 (2002).
8. Yeoh EJ, Ross ME, Shurtleff SA *et al.*: Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell* 1, 133-143 (2002).
9. Waterston RH, Lindblad-Toh K, Birney E *et al.*: Initial sequencing and comparative analysis of the mouse genome. *Nature* 420, 520-562 (2002).

Websites

101. <http://discover.nci.nih.gov>
102. <http://web.ncifcrf.gov/researchresources/mmhcc/>
103. <http://jaxmice.jax.org/>