

and effects and offers great promise for leading to the future development of treatments tailored according to persons' unique genetic characteristics. We focused on identification of expression profiles potentially predictive of response to treatment using an accessible source such as blood. We defined gene expression profiles of PBLs, in particular of non quiescent T-lymphocytes such as those stimulated with Phytohemagglutinin and interleukin-2 (PHA-PBLs) and of the derived immortalized LCLs, from healthy individuals. These data will be useful to determine the feasibility of developing directly on lymphocytes clinical tests for the prediction of response to treatment, avoiding the need of cell immortalization. The results have been applied to the study of toxicity from ionizing radiation (IR) therapy in cancer, based on the hypothesis that some cases of toxicity may be associated with abnormal transcriptional response to radiation.

Our goal is to establish the basis for a practical clinical test that works directly on peripheral blood and will be used to predict response to treatment, therefore enabling to match the appropriate cure to the appropriate patient. The proposed approach could be extended for the identification of genes signatures associated to molecules that damage DNA in a manner similar to ionizing radiation (genotoxic or radiomimetic agents).

## 6-IS

### Application of proteomics to cancer epidemiology

S. Hanash<sup>1</sup>

<sup>1</sup> Molecular Diagnostics Program, Fred Hutchinson Cancer Research Center, Seattle, USA

Although our understanding of the molecular pathogenesis of common types of cancer has improved considerably, the development of effective strategies for cancer prevention, risk assessment and early detection have lagged behind. Current proteomic strategies allow quantitative profiling of cells, tissues and biological fluids and identification of proteins changes resulting from altered levels, post-translational modifications and amino acid substitutions. A major application of proteomics is assessment of health related changes in the plasma proteome. However the vast dynamic range of protein abundance in plasma and the likely occurrence of biomarkers in the lower range of protein abundance represent a major challenge in applying a proteomics to cancer epidemiology. A combination of innovative strategies promises to overcome these challenges. Recent experience in comprehensive profiling of plasma proteins indicates that low abundance proteins may be identified and quantified with high confidence following extensive plasma fractionation and with the use of protein tagging procedures and high-resolution mass spectrometry. From an experimental design point of view, most cancer biomarker studies, including those aimed at identifying markers for early detection, are initiated with analysis of specimens from newly diagnosed subjects. Specimens obtained from large cohorts are becoming available to allow large-scale investigations of risk factors related to common cancers. The current status of the field and emerging findings will be presented.

## 7-IS

### Proteomics; An epidemiological view

R. Vermeulen<sup>1</sup>

<sup>1</sup> Institute for Risk Assessment Sciences, Utrecht University, The Netherlands

In general the function of a cell can be described by the proteins that are present in the cell and the abundance of these proteins. Although all proteins are based on mRNA precursors, post translational modifications and environmental interactions make it impossible to predict abundance of specific proteins based on gene expression analysis. As such proteomics holds great theoretical promise. However, major challenges still need to be over won both technologically (ability to reliably detect a wide range of proteins), as epidemiologically (study design, sample collection, information on inter- and intra-individual variability). In light of these challenges it might not be surprising that studies until now have only sporadically identified the same proteins associated with disease state. Furthermore, identified proteins have in general been high abundance (housekeeping) proteins and it could be questioned whether these markers provide any new biological insights in the mechanism of disease. However, new techniques like LC/MSMS and protocols like high abundance protein depletion might improve the ability of proteomic techniques to reliably measure low abundance proteins.

Targeted protein screening based on carefully selected protein markers might (at least for the immediate future) be a more promising way forward. In these techniques ligand-binding reagents, which are usually antibodies but may also be alternative protein scaffolds, peptides or nucleic acid aptamers, coated to beads or solid carriers are used to measure multiple protein markers in a small volume of biological materials. Given that a predetermined panel of protein markers is measured basic information on

assay performance (inter- and intra-batch variance), protein stability and inter- and intra-individual variance in the markers of interest can be established a-priori greatly facilitating the correct interpretation of study results. However, these techniques are currently only able to measure up to a few hundred proteins. Furthermore, a number of important technical challenges and bottlenecks in protein array technologies, some of which are unique to proteins while others are common to high throughput methods in general, will need to be solved in order to achieve the maximum capability.

A more fundamental issue might be however the highly variable proteome in combination with the often unknown influence of collection media and storage conditions on the stability of proteins. This would undoubtedly lead to an unfavorable ratio between the inter- and intra-individual variability and as such complicating its application in epidemiological research. This becomes even a bigger issue when one moves from studies targeted to disease-recognition to early (prospective) detection of diseases where one would expect relatively small changes in biological parameters. It is therefore imperative that appropriate, well-powered study designs are employed using highly standardized biological collection protocols. At current full scale proteomics, despite of the theoretical advantages over genomics and transcriptomics might not be ready to be used in large scale(prospective) epidemiological research.

## 8-IS

### Case-control mutation screening – insights and lessons from breast cancer susceptibility

N. Rahman<sup>1</sup>

<sup>1</sup> Human Genetics, Section of Cancer Genetics, Institute of Cancer Research, Sutton, United Kingdom

Approximately 1 in 10 women develop breast cancer. Epidemiological studies have demonstrated that first-degree relatives of breast cancer cases are at two-fold risk of developing the disease. Currently, three components of the genetic architecture of breast cancer have been delineated. 1) Rare, high penetrance (>10 fold) autosomal dominant cancer predisposition genes such as BRCA1 and BRCA2. 2) Common, low penetrance (<1.5 fold) susceptibility alleles that have emerged from genome-wide tag-SNP searches in breast cancer. 3) Rare, intermediate penetrance (2-4 fold) susceptibility alleles discovered through large-scale, case-control resequencing analyses. We have identified four DNA repair genes, ATM, CHEK2, BRIP1 and PALB2 which exemplify this final class, through mutation screening of familial breast cancer cases and controls. These susceptibility genes are similar to BRCA1 and BRCA2 as they are characterised by multiple, individually rare, monoallelic, truncating mutations but they are associated with smaller increases in risk, approximately doubling the risk of breast cancer. All four genes encode proteins that function in DNA repair pathways and biallelic mutations in three of them (ATM, BRIP1 and PALB2) cause childhood developmental disorders associated with high risks of cancer, similar to biallelic BRCA2 mutations. Identification of rare, intermediate / low penetrance genes is currently challenging because of the dependence on high-throughput sequencing in large case-control series and correct candidate gene selection. However, this will likely change in the near future with the advent of whole genome sequencing. Moreover, it is highly plausible that this class of susceptibility allele is making a contribution to many diseases.

## 9-IS

### Lessons from genome-scans – the example of lung cancer

P. Brennan<sup>1</sup>

<sup>1</sup> International Agency for Research on Cancer – Genetic Epidemiology Group, Lyon, France

Lung cancer is the most common cause of cancer globally, representing 1 in 8 of all cancer cases and 1 in 6 of all cancer deaths occurring in 2002. The predominant risk factor is tobacco smoking, with the risk of developing lung cancer by age 75 in Europe and North America ranging from approximately 15% in lifelong smokers to less than 1% in never-smokers. In populations where the large majority of smokers have quit smoking, such as men in many parts of Europe and North America, an increasing proportion of lung cancer cases now occur among ex-smokers. This trend is likely to continue and emphasizes the importance of elucidating further the aetiology of lung cancer. While a heritable component for lung cancer has long been recognized, progress in identifying susceptibility genes has been slow. To identify genetic factors that modify disease risk, we conducted a genome-wide association study of lung cancer. The initial phase constituted an analysis of 317,139 SNPs in 1,989 lung cancer cases and 2,625 controls from 6 central European countries. We identified a locus in chromosome region 15q25 that was strongly associated with lung cancer ( $p=9 \times 10^{-10}$ ). This locus was replicated in 5 separate lung cancer studies comprising an additional 2,518 lung cancer cases and 4,752 controls