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 Phitohemoagglutinin 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[1]
[1] *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[1]
[1]*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 reliable 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[1]
[1]*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[1]
[1]*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=9x10-10). This locus was replicated in 5 separate lung cancer studies comprising an additional 2,518 lung cancer cases and 4,752 controls

(p=5x10-20 overall), and it was found to account for 14% of lung cancer cases (Hung et al, Nature 2008). The association region contains several genes, including three that encode nicotinic acetylcholine receptor subunits (CHRNA5, CHRNA3, and CHRNB4).

This effect has been identified in two other studies (Thorgeirsson et al, Nature 2008; Amos et al, Nat Gen 2008), one of which concluded that the primary pathway is via addiction and a greater propensity to be exposed to tobacco products. The contracting arguments for the different interpretations will be discussed.

Finally, we have since extended our genome-wide study of lung cancer by including two further studies comprising an additional 750 cases and 800 controls. We subsequently replicated the top 31 independent findings in a further 4 studies comprising 3339 lung cancer cases and 6064 controls (total 5911 cases and 9416 controls). After pooling the genome-wide and replication phase results, two additional variants were strongly associated with lung cancer, suggesting new susceptibility loci.

10-IS
Statistical Approaches for Genome-Wide Association Studies

D. Balding[1]
[1]*Department of Epidemiology and Public Health, Imperial College London, United Kingdom*

Genome-wide association studies (GWAS) have in the past one or two years been highly successful at identifying genetic variants associated with complex diseases. Yet in a sense they have also been unsuccessful, because the variants discovered show only modest effect on disease risk, and explain little of the phenotypic variation that has been attributed to genetic effects in heritability studies. Diagnostic prediction of cases remains poor for most complex diseases, even in samples enriched for cases, so that for population samples it remains infeasible in practice. There are many plausible hypotheses concerning where the “missing” genetic information lies, including the possibility that structural variants and epigenetic features will be more highly predictive of disease than are SNPs. However there may be also be more predictive power from SNPs than is currently being realized. I will review the statistical methods that have been used to identify causal variants from GWAS results. These tend to be simple one-SNP-at-time analyses, perhaps with some adjustment for population structure and with cryptic kins removed. I will review some alternatives that could improve power, by analyzing multiple SNPs simultaneously and by more sophisticated adjustments for population structure and cryptic kinship. Gene-gene (epistatic) and gene-environment interactions have not yet been widely reported, in the latter case because the case-control study design does not facilitate this. I will review possibilities for incorporating them in future studies, particularly as attention moves towards GWAS of prospective cohorts that often are rich repositories of phenotype and environmental covariates. Epistatic interactions are beginning to be tackled via network and pathway-based approaches that I will also briefly review.

11-IS
Copy number variants: a common mechanism in complex diseases

X. Estivill[1]
[1]*Genes and Disease Program, Centre for Genomic Regulation (CRG), Barcelona, Spain*

Copy number variants (CNVs) may have important implications in complex diseases. Susceptibility to HIV-1 infection, glomerulonephritis associated to Lupus, Crohn’s disease and psoriasis have been found to be associated with common CNVs involving the CCL3L1, FCGR3B and DEFB4 genes, respectively (Gonzalez et al., 2005; Aitman et al., 2006; Fellermann et al., 2006; Fanciulli et al., 2007; Hollox et al., 2008). Rare cases of common disorders (pancreatitis, Alzheimer’s disease and Parkinson’s disease) are also associated with rare CNVs (Le Marechal et al., 2006; Rovelet-Lecrux et al., 2006; Singleton et al., 2003). Since several CNVs are present in each individual and several CNVs contain genes with roles in response to the environment and adaptation, it is likely that regions containing CNVs with a wide range in copy number variation might have important roles in drug-response, susceptibility to infection, inflammation and cancer, among other common traits.

Functional sequences within CNVs might provide new insights on population differences in pharmacogenomics or disease predisposition. Information about the population distribution of CNV frequencies is crucial for association studies. HapMap DNA samples from Caucasian, Asian and African populations have been used to generate CNV data at the genome level (Redon et al., 2006; Wong et al., 2006). Resequencing of DNA samples is also showing that the human DNA contains many differences at the structural level (Korbel et al., 2007), defining what is perhaps an infinite amount of structural variability of the genome organization. The mechanisms by which CNVs could have functional consequences include a direct gene dosage effect of the gene or genes embedded in the CNV, or a positional effect on genes proximal or distal to the CNV (Estivill and Armengol, 2007). A comprehensive catalogue of the spectrum of alleles at CNV regions of the human genome should provide the appropriate tools to explore the relationship between CNV loci and phenotypes, and will define specific structural changes that modify gene expression and function.

We have studied for CNVs, samples of the HGDP-CEPH Human Genome Diversity Cell Line Panel, a widely used resource for studies of human genetic variation. By comparing samples from twelve population groups we have identified over 200 genomic regions that vary in the genomic structure between groups. Most of these regions coincide with already known CNVs and segmental duplications. These regions contain genes with a role in immune response, adaptation to environment, and metabolic pathways. This data set of CNVs allows the most comprehensive characterization to date of human genetic variation with a potential role in disease susceptibility in different human population groups.

12-IS
Molecular cancer epidemiology and public health

R. Saracci[1]
[1] *IFC- National Research Council, Pisa, Italy*

Molecular epidemiology , i.e. the use in epidemiological studies of techniques of molecular biology, has pervaded in the last two decades all areas of epidemiology, and of cancer epidemiology in particular, ranging from studies of aetiology and pathogenesis to investigations of biomarkers for early detection of minimal disease, potentially applicable to screening programmes . It might be pointed out that enthusiasm for the adoption of molecular epidemiology methods in cancer epidemiology has not been matched by translation into material prevention advances at the population level, as the largest scope for prevention still derives from knowledge on such exposures as tobacco, alcohol, energy balance or occupational and environmental carcinogens acquired by more traditional epidemiological approaches. This view however underrates the time lag often necessarily intervening between novel research approaches and practical applications. These have in fact started to emerge, notably in the area of cancers induced by infectious agents. Cervical cancer is the most relevant example : at the level of exposure biomarkers, the aetiological link with human human papilloma viruses has been established by advanced molecular epidemiology techniques, and at the level of early effect/outcome biomarkers, the detection of the virus in cervical cells has come into discussion as complement to the pap-test for screening purposes. In the area of susceptibility biomarkers expectations are high of major prevention advances , well beyond what is already possible and implemented for the important but comparatively uncommon cancers determined by single genes. Susceptibility biomarkers (in the broad sense) encompass genetic variants affecting not only cancer risk via a multiplicity of direct or mediating paths (e.g. through activation or inactivation of carcinogens) but also conditioning exposure to external agents (e.g. tobacco or alcohol) via behaviour-related effects. The working hypothesis of an “integrative epidemiology” approach is that a combined study of environmental exposures, susceptibility biomarkers with their proteomic and metabolic expressions as well as of biomarkers for molecular sub-typing of cancers may lead to a more and more refined stratification of risk by individual traits and by cancer type, paving the way to individualized prevention. While certainly fertile for research this working hypothesis raises two issues from a public health viewpoint. First it cannot be taken for granted, and needs to be critically examined from different quantitative angles, that the bulk of cases of a common cancer can be prevented by an individualized risk approach, a refined variant of the “ high risk” strategy of prevention. Second environmental factors, often more difficult to measure with accuracy than molecular bodily components, may become secondary elements in the constellation of demonstrable determinants in the carcinogenic process : hence prevention may in practice become restricted to molecular determinants as modifiable by pharmacological means.

13-IS
Abstract not received

14-IS
Epidemiology of childhood leukemia: a transdisciplinary approach

J. Ross[1], K.J. Johnson[1], D.A. Largaespada[1], L.G. Spector[1], S. Ognjanovic[1]
[1]*University of Minnesota Cancer Center, Minneapolis, MN, USA*

Leukemia, including acute lymphoblastic (ALL) and acute myeloid (AML), is the most common childhood cancer in developed countries; over 3000 new cases are diagnosed each year in the United States. Incidence rates have been increasing significantly since the 1970s. Despite several studies, the etiology remains largely unknown. Steadily growing evidence suggests