

## **Genomics and proteomics offers new hopes towards a personalized approach to lung cancer prevention and treatment**

**Lionel Gil\***

Programa de Biología Molecular y Celular  
Facultad de Medicina  
Universidad de Chile  
Santiago, Chile  
Tel: 56 2 6786068  
Fax: 56 2 7376320  
E-mail: lgil@machi.med.uchile.cl

**Marta Adonis**

Programa de Biología Molecular y Celular  
Facultad de Medicina  
Universidad de Chile  
Santiago, Chile  
Tel: 56 2 6786068  
Fax: 56 2 7356373  
E-mail: madonis@canela.med.uchile.cl

Cancer involves a pathological breakdown in the cellular process that control proliferation, differentiation and death of particular cells. World wide approximately 10 million people are diagnosed with cancer annually and more than 6 million die of the disease every year, over 22 million people in the world are cancer patients. In the period 1990-2000 an increase of around 19% in incidence and 18% in mortality has been observed. It is expected by 2020 that the number of cancer cases will double to 20 millions with an annual death of 12 millions. According to the World Health Organization, lung cancer is the most common malignant disease worldwide particularly among men, representing 12.3% of all cancers. It is the major cause of death from cancer, accounts for 1.1 million deaths a year and 17.8% of all cancer deaths, thus, represents the type of cancer with worse prognosis. No effective treatment is available for lung cancer; the five year survival rate for lung cancer patients is less than 15%.

### **RISK FACTORS**

It is estimated that at least 75% of the cancers are caused by chemical compounds in a process call chemical carcinogenesis. Lung cancer is generally a consequence of chronic exposure over long period of time to environmental carcinogen mixtures as well as other environmental, life style, diet and host factors. Environmental factors include air, water and soil air pollution, occupational exposures. Host factor include: genetic pattern, carcinogen exposure, carcinogen metabolism, DNA repair activity, oncogene and tumour suppressor expression and nutritional status.

Arsenic in drinking water in several areas of the world, as

in the II Region of Chile, has been related to increased risk of lung cancer and other cancers such as bladder and skin. Tobacco smoking is the main known cause of cancer related death worldwide. In the USA, Japan and Europe smoking accounts by 83-92% in men and 57-80% in women of the lung cancer deaths. The risk of lung cancer among smokers relative to the risk of non smokers is in the order of 8-15 in men and 2-10 in women. Many occupations and some specific chemicals encountered at work are associated with increase of cancer risk especially in newly- industrialized countries where most industrial activity take place in multiple small scale operations.

Outdoor and indoor pollution are also risk factors for lung cancer. Very high lung cancer rates occur in some regions of China and other Asian countries among non smoking women suggesting that indoor pollution as a result of combustions sources for heating and cooking is also a risk factor. Lung cancers is also attributable to outdoor air pollution, several studies have compared residence in urban vs. rural areas as a risk factor. In general lung cancer rates were higher in urban areas and correlate well with levels of respirable particulate matter as well as with specific pollutants such as polycyclic aromatic hydrocarbons (PAHs) or with mutagenic extracts in bacterial assay systems (Adonis et al. 1993).

A recent study done in 500.000 people in 116 cities in USA indicates that after eliminating different confounding factors, fine particulate (PM<sub>2.5</sub>) air pollution exposures were associated with significant increases in lung cancer mortality and each 10 ug /m<sup>3</sup> elevation in fine particulate air pollution was associated with 8% in cancer mortality

---

\*Corresponding author

(Pope et al. 2002). But, this situation might be even worse in developing countries where regulations to control vehicle and industrial emissions are poorly accomplished (Adonis and Gil, 2000; Gil et al. 2000).

It is important to point out that many of the carcinogenic pollutants present in the smog of big cities are the same that are present in cigarette smoke. One characteristic of air environmental pollutants is that the individuals lack control over their level of exposure. For instance in winter in downtown Santiago of Chile we have detected 5 ng/m<sup>3</sup> of benzo(a)pyrene (BaP), the most studied environmental carcinogen in respirable particles. Thus, a non smoking inhabitant breath benzo(a)pyrene BaP levels equivalent to smoke 10 cigarette a day.

The IARC (International Agency for Research in Cancer) has a program to evaluate the data of carcinogenic hazards to humans as consequence of exposure to particular chemical, physical and biological agents and mixtures. IARC has classified a great number of chemicals as: human carcinogens, probably carcinogens to humans, possible carcinogenic to humans and non carcinogens.

## UNDERSTANDING THE MECHANISMS OF CHEMICAL CARCINOGENESIS

Molecular cancer epidemiology used advanced laboratory methods in combination with analytical epidemiology tools to identify at the molecular level specific exogenous or/and host factors that play a role in human cancer causation. Biomarkers of internal dose, of biologically effective dose, of response and susceptibility give important information for prevention and treatment of the disease. Molecular epidemiology of lung cancer has focused on environmental causes because it is believed that one of the main factors of incidence is exogenous and hence preventable. The potential contribution of molecular epidemiology includes: proving evidence that environmental agents pose carcinogenic risk, helping establish the causal of environmental factors in cancer, identifying environmental susceptibility interactions and populations at higher risks. Thus, molecular epidemiology has identified a number of carcinogenic hazards, in some cases providing definitive etiologic data furthering our understanding of individual genetic and acquired susceptibility to environmental carcinogens.

Molecular endpoints used in studies of molecular carcinogenesis include metabolites in body fluids, DNA and protein adducts, mutation in reporter genes, in oncogenes and in suppressor genes, genomic instability, aberrant gene expression and altered cell culture, which give key information about both genetic and environmental susceptibility factors.

For any given environmental exposure, individual differences in susceptibility might have a genetic basis. Genetic differences in metabolic activation and

detoxification of environmental carcinogens may partially explain host susceptibility to chemically induced cancers. Some PAHs are the most ubiquitous carcinogenic pollutants. They are present in incomplete combustion process including vehicles exhaust, chimney emissions, smoked foods, cigarette smoke and indoor heating and cooking systems. [Figure 1](#) shows that PAHs are metabolized to reactive DNA binding diols epoxides by phase I (CYP1A1, CYP1B1) enzymes and detoxified by phase II enzymes (GSTs) before reaching their target. But, during these processes BaP is also activated to ultimate carcinogens that reacts covalently with DNA, generating DNA adducts which are very difficult to be repaired by DNA repairing enzymes and that are responsible for the initiation step in chemical carcinogenesis, leading to tumour development. Thus, individual variations in metabolic activities in each phase or in the coordination of these two phases regulate the clearance of DNA toxic metabolites and might be partially responsible for individual host susceptibility to PAHs. Several polymorphisms have been described in CYP1A1 but one in the 3'-non coding region (CYP1A1\*2A) has been the most studied. Polymorphisms in CYP1A1 as well as in GSTM1 have been associated with different types of cancer risk including lung cancer. In a case-control study in the Chilean population we have found that the estimated relative risk (Odds Ratio=OR) for lung cancer associated to a single mutated allele in CYP1A1\*2A was 2.41 and in GSTM1 null OR was 2.46, whereas for individuals carrying combined CYP1A1\*2A and GSTM1 mutated alleles OR was 3.56, meaning that individuals carrying mutations in both genes have a risk 3.56 times higher than individuals carrying the wild type polymorphisms. These results suggest that Chilean people carrying single or combined GSTM1 and CYP1A1 polymorphisms could be more susceptible to lung cancer induced by environmental pollutants, (Quiñones et al. 2001). In addition in a recent study done in diesel exposed workers we have found that workers carrying the CYP1A1\*2A alleles showed significantly higher 1-hydroxypyrene urinary levels than non exposed workers from rural areas carrying the same genotype, showing the importance of the relationship between exposure and genetic risk polymorphisms. Furthermore the highest 1-hydroxypyrene pyrene urinary levels was found in those workers carrying the combined CYP1A1\*2A and GSTM1 null genotypes (Adonis et al. 2003).

Knowledge for specific genetic polymorphisms conferring susceptibility to lung cancer should provide more power for the detection and characterization of the environmental factors that increase the risk in an individual. Thus, individuals carrying risk polymorphisms might be more susceptible to exposure to environmental pollutants.

Polymorphisms in genes coding for enzymes involved in the metabolism and detoxification of carcinogens including many cytochromes P-450 that codify for Phase I enzymes or genes that codify for conjugation Phase II enzymes, or

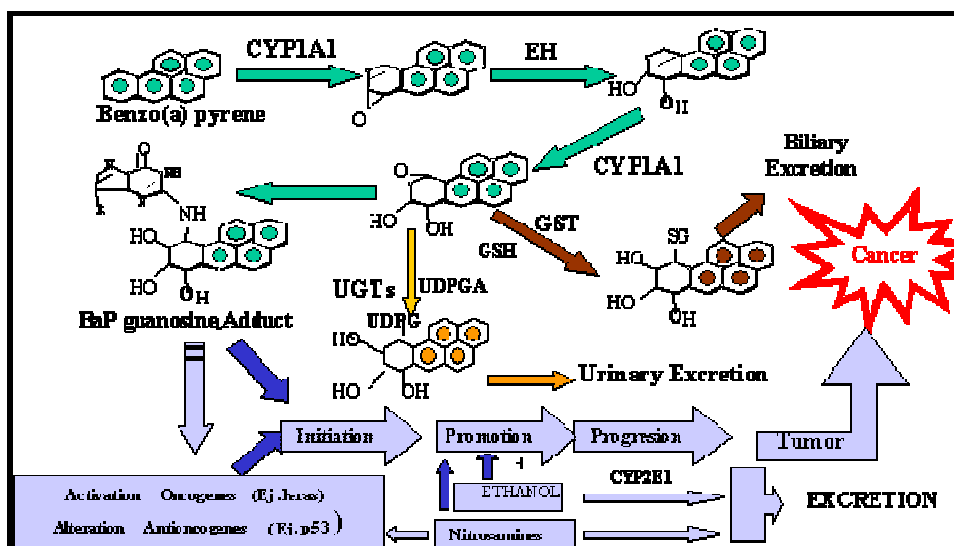


Figure 1. Relationship between PAH metabolism and chemical carcinogenesis.

EH: Epoxyde Hydrase;  
 GST: Glutathion Transferases;  
 GSH: Glutathion;  
 UGTs: Uridinglucoronyl transferases;  
 UDPGA: Uridin diphospho-glucuronic acid;  
 CYP1A1: Cytochrome P-450 1A1

for DNA repair enzymes and for tumour suppressor proteins might be determinant in the risk to lung cancer by environmental exposure. [Table 1](#) shows some genetic polymorphisms that have been associated with lung cancer. The frequency of many of these polymorphisms varies within different ethnicities suggesting that some populations around the world might be more susceptible to environmental carcinogens pollutants.

It is hoped that a more unified approach to cancer epidemiology and genetics will identify those combinations of genetic susceptibility and environmental exposures that lead to significant increases in risk at the individual and population level. This could lead changes in lifestyle and avoidance of specific exposures in genetically susceptible individuals.

A positive familial history of lung cancer has been identified as a risk factor associated with certain polymorphisms of the cytochrome P-450, phase II genes, and DNA repair genes. Activation point mutations in the KRAS oncogene occur in adenocarcinoma with a prevalence ranging from 15-60% which is higher in smokers than in non smokers and this mutations might be a relatively early event in lung carcinogenesis.

Furthermore mutations in the p53 are also frequent events in lung cancer. Among lung cancers the proportion of p53 mutations increases with duration and amount of tobacco smoking. Polymorphisms of the p53 gene related to single base changes could cause aminoacid replacements which

could modify the functionality of the protein. The polymorphism at codon 72 that presents the arginine (Arg) or proline (Pro) genotype has been associated with genetically determined susceptibility to smoking related cancers. In a recent study done in Chileans we have found a relationship between the presence of the Pro allele and lung cancer risk in male smokers. Thus, O.R was 2.47 for one single nucleotide polymorphic allele Pro and O.R= 3.88 for the Pro: Pro genotype was observed (Irrázabal et al. 2003).

## ADVANCES IN GENOMICS AND PROTEOMICS

The marked increase in the amount of information available on genetic sequences and gene expressed by humans offers great opportunities for improving our understanding of lung cancer. We should expect to witness a fast increase in the rate of discovery of genes involved in lung cancer pathogenesis and we should be able to develop reliable molecular criteria for classifying lung cancers and predicting biological properties of individual tumours.

The complete sequencing of the human genome is allowing the identification of expressed sequences and polymorphic sequences providing information that is increasing available to medical researchers that make possible to visualize in the near future a more personalized protocols for the disease treatment as well as to identify persons with higher risk. Thus, the discovery genes involved in the pathogenesis may lead to new targets for diagnosis and treatment. A knowledge in the polymorphisms that make each of us unique individuals

**Table 1. Some genetic polymorphisms related to lung cancer.**

Gene	Polymorphism	Codon (SNP)	Protein Function
<b>CYP1A1</b>	Msp1 (CYP1A1*2A)	T6235C	Phase I: PAHs Activation
<b>CYP1A2</b>	CYP1A2*F	C734A	Phase I: HAPs, Nitrosamines and arylamines metabolism
<b>CYP2E1</b>	DraI  PstI	T7668A  C1091T	Phase I: Procarcinogens Activation (4- methylnitrosamine, aromatic amines
<b>CYP3A4</b>	CYP3A4*3	Met445Treo	Phase I: Drug metabolism. Procarcinogens Activation (PAHs)
<b>mEH (Epoxide hydrolase)</b>	EPHX1*3	Tir113His	Phase I: Procarcinogens Activation (HAPs)
<b>GSTs (Glutathion transferases)</b>	GSTM1  GSTT1  GSTP1	mu  teta deletion  pi deletion, Ile105Val	Phase II: Glutathion conjugation of hydrophobic and electrophilic compounds
<b>UGTs (UDP- glucuronosyltransferases)</b>	UGT1A1*28	Insertion/deletion of a repetead sequence TA, in the promotor region	Phase II: Glucuronic acid Conjugation of hydrophobic and electrophilic compounds
<b>NATs (N-acetyltransferases)</b>	NAT1*10	T1088A and C1095A	Phase II: Acetylation of aromatic amines and heterocyclic amines activation
<b>ERCC2</b>	XPD (NER)	Lis751Gln Asp312Asn	DNA repair
<b>XRCC1</b>	BER	Arg280Gln	DNA repair
<b>p53</b>	Codon 72	Arg72Pro	Cellular cycle regulation Tumor supressor
<b>MDR</b>	MDR1	C3435T	Drug Resistance

could be the key in the future for predicting individual risks for developing the disease as well as the individual responses to antineoplastic drugs. However the impact of genomics is still unknown, of the 30.000 human genes only few thousands have known functions or even names. Thus, the cancer researchers are presently searching in genome databases trying to identify candidates for genes important in lung cancer pathogenesis. Chromosomal aberrations and loci of chromosomal deletions have already been defined in lung cancer and with the increasing availability of gene maps we are now seeing an acceleration in our recognition of new genes involved in lung cancer pathogenesis.

Some of the most promising applications of genomics to lung cancer research come from measurements of the gene expression of cancer cells. Thus, serial analysis of gene expression, oligonucleotide arrays and cDNA arrays are now tools that allow investigators to measure the expression of thousands of gene in a single experiment. This allows to learn about selected clones, genes over expressed in certain types of lung cancers and to identify potential therapeutic targets. Gene arrays have offered the opportunity to test large numbers of gene as potential predictive markers of clinical applications. However, because the number variables measured generally exceeds the number of samples in such studies some skeptics have argued that any association between a single marker and outcome would be meaningless by traditional criteria, but statisticians have developed methods to focus on overall patterns of gene expression rather in individual genes.

Clinical lung cancer researchers recognizes that the standard morphological classification of lung cancers is unable to provide critical information on the aggressiveness of a particular cancer and how the cancer will respond to radiation or chemotherapy. Developing a well defined taxonomy for lung cancer is important both for clinical management of the disease and for research. Because of implications for treatment and prognosis no one will question the significance of differentiating for instance lymphoma from carcinoma or small cell carcinoma from non- small cell carcinoma in the evaluation of lung tumour and in their applications for treatment and prognosis.

New genomic techniques will facilitate changes in tumour classification more useful for therapy. Gene expression profiles also might help to distinguish between different phenotypes of lung cancer. In this case is very important first to purify line cancer cells from tissue samples from other cells, techniques like laser capture micro-dissections and scraping nearly pure clusters of neoplastic cells from tumour tissues are very useful. The development and utilization of tissue-processing methods will, be essential for the successful execution of molecular phenotyping projects. Another key element for cancer classification projects will be the development of reliable and reproducible techniques for measuring the genes that are more relevant for lung cancer. For instances there is big hopes in the development of custom neumochips arrays that

represents the genes expressed in respiratory epithelium in lung cancer.

Although gene expression patterns are closed linked to cell's function, it is ultimately genetic alterations that are responsible for the cancer phenotype. Thus, for detecting mutations, genomic DNAs techniques such as oligonucleotide arrays can represent a series of different sequences for each gene, including wild type and single-base mismatched sequences.

## **THE FUTURE IN LUNG CANCER RESEARCH AND THEIR IMPACT IN LUNG CANCER DIAGNOSIS, TREATMENT AND PREVENTION**

In the future considerable efforts will be made for classify lung cancers, to discover genes involved in lung cancer pathogenesis and to study the biochemistry and function of those genes in the disease. The main tools for tissues processing, array production and decision-based statistical analysis strategies all see to be in place for these efforts. Promising new proteomics techniques might not only measure protein levels but might also recognize post translational modifications of proteins. Integrating these measurements with those of gene expression could add a whole new dimension to our understanding of lung cancer. Our current focus of genomic on gene expression virtually ignores that 95% of the genome that does not encode for proteins or regulatory information. Although the function of this vast amount of our genome is still unknown, it is often thought to be involved in stabilizing chromosomes and thus should be considered a likely target for lung cancer related aberrations.

Although survival for stage I cancers may reach about 65%, in lung cancer survival rarely exceeds 15%. In the light of poor survival rates, the development of novel techniques for early diagnosis, prevention and therapies is a high priority to diminish lung cancer incidence, prognosis and mortality. Thus, clinical scientists, pathologists, molecular biologists, statisticians, computer specialists working as a team will make lung cancer genomics programs successfully. Collaborative efforts will probably be increasing important in cancer research because sophisticated tools required specialized expertise.

## **ACKNOWLEDGMENTS**

We acknowledge the support from grants : FONDECYT N° 1000628, European Community, INCO N° 962731 and DID , University of Chile ENL 03/12.

## **REFERENCES**

ADONIS, M. and GIL, L. Mutagenicity of organic extracts from Santiago (Chile) airborne particulate matter. *Mutation Research*, 1993, vol. 292, p. 51-61.

ADONIS, M. and GIL, L. Polycyclic aromatic hydrocarbons levels and mutagenicity of inhalable

**Gil, L. and Adonis, M.**

particulate matter in Santiago, Chile. *Inhalation Toxicology*, 2000, vol. 12, no. 12, p. 1173-1183.

ADONIS, M.; MARTÍNEZ, V.; RIQUELME, R.; ANCIC, P.; GONZÁLEZ, G.; TAPIA, R.; CASTRO, M.; LUCAS, D.; BERTHOU, F. and GIL, L. Susceptibility and exposure biomarkers in people exposed to PAH-diesel exhaust. *Toxicology Letters*, 2003, vol. 144, no. 1, p. 3-15.

GIL, L.; KING, L. and ADONIS, M. Trends of polycyclic aromatic hydrocarbons levels and mutagenicity in Santiago's inhalable airborne particles in the period 1992–1996. *Inhalation Toxicology*, 2000, vol. 12, no. 12, p. 1185-1204.

IRARRÁZABAL, C.; ROJAS, C.; ARACENA, R.; MÁRQUEZ, C. and GIL, L. Chilean pilot study on the risk of lung cancer associated with a codon 72 polymorphism in the gene of protein p53. *Toxicology Letters*, 2003, vol. 144, no. 1, p. 69-76.

QUIÑONES, L.; LUCAS, D.; GODOY, J.; CÁCERES, C.; BERTHOU, B.; VARELA, N.; LEE, K.; ACEVEDO, L.; MARTÍNEZ, L.; AGUILERA, A. M. and GIL, L. CYP1A1, CYP2E1 and GSTM1 genetic polymorphisms. The effect of single and combined genotypes on lung cancer susceptibility in Chilean people. *Cancer Letters*, 2001, vol. 174, no. 1, p. 35-44.

POPE, C.A.; BURNETT, R.T.; THUN, M.J.; CALLE, E.E.; KREWSKI, D.; ITO, K. and THURSTON, G.D. Lung cancer, cardiopulmonary mortality and long-term exposure to particulate air pollution. *JAMA*, 2002, vol. 287, no. 9, p. 1132-1141.