

**COMMITTEE ON CARCINOGENICITY OF CHEMICALS IN FOOD, CONSUMER PRODUCTS AND THE ENVIRONMENT**

**INTERACTION BETWEEN GENOTYPE AND CHEMICALS IN THE ENVIRONMENT ON THE INDUCTION OF CANCER IN RISK ASSESSMENT – SCOPING PAPER FOR UPDATE REVIEW**

**Introduction**

1. Preceded by a scoping paper in 2000 (CC/2000/07), in 2001 the COC reviewed in detail a number of papers prepared by the DH Toxicology Unit at Imperial College on the subject of interaction between genotype and exposure to chemicals in the environment, and the induction of cancer. The following papers were presented and evaluated, and subsequently a statement was published in 2002. The papers reviewed were:

[CC/01/3](#) - Criteria for the design of gene-environment epidemiology studies

[CC/01/4](#) - A review of potential target genes for susceptibility to carcinogenesis

[CC/01/5](#) - A review of how gene-environment studies should be used in risk assessment process

2. In the statement, one of the recommendations from the Committee was to keep the subject under review, particularly in light of the expected developments from the Environmental Genome Project and other initiatives. At the Horizon Scanning meeting in November 2009, members asked to be provided with a copy of the previous statement as a starting point to determine whether the conclusions could be extended in light of scientific advances over the last decade (Annex 1).

3. Since the publication of the previous COC statement, there have been substantial advances in the human genomic field. Publication of the results from the Human Genome Project (HGP), the International Hapmap Project and the Environmental Genome Project (EGP) have lead to the development of more efficient methods for the discovery of human genetic variants associated with common disease such as cancer. New methods to measure genome variation on a large scale have advanced a new type of genome association study, a Genome-Wide Association Study (GWAS).

4. Population genomics recognizes that, in order to completely understand common disease risk and human health, basic data are needed on genomic variation and on lifestyle behaviours and environmental factors. Hence, there is a need to carry out studies of normal genomic variation across whole populations (Khoury 2004). In recent years, there has been an emergence of population-based Bio-banks in many countries with the objective of quantifying longitudinally the joint influences of genetic and environmental factors on the occurrence of common diseases (Knoppers, 2005).

5. We outline below details of the major projects that have been initiated over the past decade in both the genetic field and exposure assessment and their contribution to the field of gene-environment interaction and cancer risk.

#### **Overview of major projects in the genetic field**

6. The Human Genome Project (HGP) was launched in 1990 with the goal of obtaining a highly accurate sequence of the euchromatic portion of the human genome. In February 2001, the International Genome Sequencing Consortium (IHGSC) reported a draft sequence of the human genome. The 2001 published draft sequence has a number of shortcomings:

- 1) omission of ~10% of the euchromatic genome
- 2) interruption of sequence by ~150,000 gaps
- 3) order and orientation of many segments within local regions were not established

The completed project, with a goal of publishing a complete human sequence, was reported in 2003. The number of gaps was reduced by 400-fold to only 341 compared to the draft sequence, most of which were associated with segmental duplications<sup>1</sup>. The assembled near-completed genome has an error rate of only ~1 per 100,000 bases, measured using error probability assessments generated

---

<sup>1</sup> Segmental duplications are duplicated blocks of genomic DNA, typically ranging in size from 1–200 kb (IHGSC, 2001). They often contain sequence features such as high-copy repeats and gene sequences with intron–exon structure. Thus, being composed of apparently normal genomic DNA, segmental duplications cannot be detected *a priori*; rather, most segmental duplications have, to date, been discovered based on experimental analyses (Bailey et al., 2001).

by DNA base-calling software (Ewing et al., 1998; Gordon et al., 1998; Ewing and Green 1998); it contains 2.85 billion nucleotides and covers 99% of the euchromatic genome. On the basis of the available evidence, the HGP estimated that the total number of protein coding genes were in the region of 20,000 - 25,000.

7. As part of the HGP, the National Human Genome Research Institute and the National Centre for Biotechnology Information established a dbSNP, a public database for worldwide SNP data. DbSNP currently catalogs approximately 7.2 million SNPs with submission obtained from four different categories 1) mined from EST databases, 2) private investigator/corporate experimental results, 3) results of the National Human Genome Research Institute (NHGRI) SNP discovery and 4) SNP mining from the Human Genome Project sequences. Approximately 3.3 million SNPs are considered validated by multiple sequence analyses.

8. The National Institute of Environmental Health Sciences (NIEHS) initiated the Environmental Genome Project (EGP) in 1997, as a comprehensive systematic approach to understanding how human genetic polymorphism influences individual susceptibility to environmentally induced disease (Kaiser, 1997 & Olden and Wilson, 2000). It is a multi-component effort composed primarily of investigator-initiated grants in five research areas: discovery of human genetic polymorphism; identification of functional genetic polymorphism; population-based epidemiological and clinical studies; technology development; and the ethical, legal, and social implications of genomics-based research.

9. One approach to identifying SNPs is to analyze DNA sequences already deposited in databases (i.e., data mining). The EGP adopted a "Candidate Gene Approach", using a carefully selected set of DNA samples from ninety five unrelated males and females including Americans of European, African, Mexican or Asian descent, and Native Americans. Mammals have evolved sophisticated responses to changing environmental conditions by regulating the expression of specific genes to minimize the biological consequences of the environmental stresses. Genes in these pathways are called environmentally responsive genes

(ERGs) and can be classified into the following 8 categories: cell cycle, DNA repair, cell division, cell signalling, cell structure, gene expression, apoptosis, and metabolism. Since the project started, the EGP has sequenced over 607 environmentally responsive genes in the panel of 95 human DNA samples. To date, mostly metabolism, DNA repair and cell cycle genes have been re-sequenced by the EGP. Another accomplishment was the compilation of the publicly accessible GeneSNP database, which lists the 1000s of new SNPs now identified, for research use. The GeneSNPs database was developed primarily as a web-based, publicly-available graphical display of the newly identified SNPs discovered through re-sequencing ERGs. In 2006, dbSNP/HapMap updates were added to the site to capture and incorporate the major releases of genome-wide SNP data (including validated genotypes) from the HapMap efforts.

10. The international Hapmap Project is an international consortium whose goal was to determine the common patterns of DNA sequence in the human genome and to make this information freely available in the public domain. They developed a map of these patterns across a genome by characterizing sequence variants, their frequencies and correlation between them in DNA samples from populations with ancestry from four geographically diverse populations. The samples obtained for inclusion in the HapMap were 30 adult-and-both-parents trios from Ibadan, Nigeria, 30 trios of U.S. residents of northern and western European ancestry, 44 unrelated individuals from Tokyo, Japan and 45 unrelated Han Chinese individuals from Beijing, China. Phase I of the Hapmap project produced data on 1 million SNPs in the Hapmap samples, evenly spread across the genome. In phase II of the Hapmap project, a further 2.1 million SNPs were successfully genotyped on the same individuals. The resulting Hapmap has a density of approx one SNP per kilobase and is estimated to contain approximately 25-35% of all the 9-10 million SNPs (minor allele frequency (MAF)  $\geq 0.05$ ) in the assembled human genome.

11. The SNP Consortium Ltd was initiated in 1999 as a non profit foundation whose goal was to create a public SNP database. The project expanded to determine allele frequencies and genotypes (Sachidanandam et al., 2001).

12. The 1000 Genome Project was launched in 2008 and is an international research effort to produce an extensive catalog of human genetic variation that will support future medical research studies. The 1000 Genome project builds on the Hapmap project and aims not only to map the SNPs, but also plans to produce a high resolution map of larger differences in the genome structure called structural variants. Structural variants are rearrangements, deletions, or duplications of segments of the human genome. The 1000 samples for the project will come from those used in the extended Hapmap set.

13. Genome-wide association studies (GWAS), also known as whole genome association studies (WGA studies) are an examination of genetic variation across a given genome, designed to identify genetic associations with observable traits. Completion of the HGP in 2003 has made it possible to assess the genetic contributions to common diseases and analyse whole genome samples for genetic variation that contribute to their onset. GWAS are genome-wide, non-hypothesis driven in nature. GWAS use high throughput genotyping technologies to assay hundreds of thousands of SNPs and relate these variants to disease. GWA studies build on lessons learned from candidate gene and family linkage studies, as well as the expanding knowledge of relationships among SNP variants generated by the International Hapmap Project, to work out the majority of common genetic differences among individuals and relate them to health and disease.

14. A typical GWAS has 4 parts; 1) Selection of a large number of individuals with the disease and a suitable comparison group, 2) DNA isolation, genotyping and data review to ensure high quality genotyping, 3) Statistical test for associations between the SNPs and the disease, 4) Replication of identified associations in an independent population sample (Pearson and Manolio, 2008).

15. Most GWA studies yield at least one highly statistically significant association but almost all the associations are weak [small odds ratios (ORs), in the order of 1.1–1.6]. According to Vineis et al. (2008), almost all results from cancer GWA studies pointed to novel loci, previously unidentified. Vineis et al. (2008) provide a list of robust cancer GWA studies with relative risks in the order of 1.06–2.23, but

mostly ~1.2. These included studies on breast, prostate and colorectum cancers (Easton et al., 2007; Stacey et al., Hunter et al., 2007; Yeager et al., 2007; Gudmundsson et al., 2007; Tenesa et al., 2008 and Tomlison et al., 2008).

16. Rapid progress in our understanding of genetic factors underlying susceptibility to major common human diseases, such as cancer, is now emerging through the application of GWAS (Wang et al., 2010; Landi et al., 2009; Landi et al., 2009; Song et al., 2009; Thomas et al., 2009; The Wellcome Trust Consortium, 2007; Gold et al., 2007). This approach offers great promise for providing new insights into gene-environment interactions through activities such as the NIH Gene and Environment Initiative (<http://www.gei.nih.gov>). However, the ability of current genome-wide approaches, as implemented on commercially available genotyping products ('chips'), to target the full genetic diversity for broad classes of genes may be limited. The re-sequencing data generated by the EGP provides a useful benchmark for determining the sensitivity of GWAS platforms in detecting common SNPs and genetic associations for this important class of candidate genes. Also the HapMap project provided the necessary background for the conduct of GWA studies by identifying sets of SNPs that cover the entire variability at the genomic level, or most of it.

17. Genes and Environment Initiative (GEI) is a four-year NIH-wide program to identify major genetic susceptibility factors for disease and to develop technologies for reliable and reproducible measurement of potential causative environmental exposures. The GEI program has two main components:

- 1) The Genetics Program, a pipeline for analyzing genetic variation in groups of patients with specific illnesses, involving GWAS. The genetics program includes a consortium for GWAS, as well as replication and fine-mapping studies, sequencing studies, functional studies, development of analytical methods and databases, and pilot clinical translation studies. The GWAS component, named the Gene, Environment Association Studies (GENEVA) consortium, was initiated in 2006 and the goals of the consortium are to (i) identify genetic variants associated with complex diseases and traits in initial genome-wide discovery studies; (ii) identify variations in gene-trait associations related to environmental

exposures; and (iii) ensure the rapid sharing of data to the general scientific community. Curated data are made available in a central, controlled-access database, the database of Genotype and Phenotype (dbGaP), established by the National Centre for Biotechnology Information (NCBI).

2) The Exposure Biology Program (EBP), an environmental technology development program to produce and validate new methods for monitoring environmental exposures that interact with a genetic variation to result in human diseases. The EBP is developing a set of assessment tools for measuring individual exposures to environmental stresses including airborne toxic chemicals, psychosocial stress, addictive substances, diet and physical activity and measures of the biological response to those stressors. These tools will provide the improved accuracy and precision needed to determine how environmental and lifestyle factors interact with genetic factors to determine the risk of developing disease. The program is divided into four major areas: sensors for assessment of chemical exposures (SACE), diet and physical activity (DPA), network on psychosocial stress and addictive substances (NEPSAS), and biological response indicators to environmental agents (BRI).

18. In 1998, the National Office of Public Health Genomics (NOPHG) in the US established the Human Genome Epidemiology Network (HuGENet). It is an informal collaboration of individuals and organisations interested in accelerating the development of the knowledge base on genetic variation and human health. Human Genome Epidemiology (HuGE) uses population-based epidemiologic methods to examine the relationship of genetic variation with health and disease. The HuGE reviews have been a fundamental part of the HuGENet to develop an online resource to house the cumulative information on epidemiological aspects of human genes. A HuGE review is a peer-reviewed, systematic review of gene-disease associations. It collates evidence on population frequencies of genetic variants, genotype-phenotype associations, interactions among genes and environmental exposures. HuGE reviews typically point to gaps in existing epidemiologic and clinical knowledge, thus stimulating further research in these areas. More than 70 HuGE reviews have been completed under the guidance of

HuGENet, with more than 25 reviews discussing the association of a particular gene variant and cancer.

19. The task of identifying genetic determinants for complex, multigenic diseases is hampered by small studies, publication and reporting biases, and lack of common standards worldwide. The creation of a network of networks that include groups of investigators collecting data for human genome epidemiology research was proposed at a HuGENet meeting in 2005. The Network of Networks currently has 35 consortia that are represented by hundreds of collaborators interested in sharing knowledge, experience and resources in the conduct, analysis and dissemination of results of human genome epidemiology investigations. In 2006, the Network of Networks published a “Roadmap” for using consortia-driven pooled meta-analyses to accelerate the knowledge base of gene-disease association.

The Roadmap involved a number of different components

- 1) working with genetic epidemiology study platforms to improve the performance and productivity of the consortia and groups
- 2) promoting the publication of methodologically sound genetic association studies with transparent reporting of their methods and avoidance of selective reporting
- 3) developing methods of synthesis and meta-analysis of the literature on genetic associations
- 4) developing field synopses with an online encyclopedia summarizing the known information about genetic associations through a systematic assessment of their cumulative evidence

20. In 2006, the HuGENet implemented the field synopsis concept. A field synopsis is a regularly updated snapshot of the current state of knowledge about genetic associations in a particular field of research defined by a disease, phenotype, or family of genes. A workshop in Venice generated interim guidelines for grading the cumulative evidence in genetic associations based on three criteria 1) the amount of evidence; 2) the extent of replication; and 3) protection from bias. The proposed scheme allows for three categories of

descending credibility (A,B,C) for each of these criteria and also for a composite assessment of “strong,” “moderate,” or “weak” credibility (Ioannidis et al., 2007). Briefly, an overall strong rating is reserved for an AAA rating, while an overall “weak” rating is reserved for associations with one or more C ratings. The rest are labeled as “moderate”. Pilot studies in the areas of Alzheimer disease, bladder cancer, schizophrenia, preterm birth, coronary heart disease as well as DNA repair genes and cancer phenotypes have been undertaken.

21. Genetic Susceptibility to Environmental Carcinogenesis Study (GSEC) is an international collaborative study that began in 1999 to study the genetic susceptibility to environmental carcinogens. The study involved information on polymorphism in the genes that metabolise environmental carcinogens using original data from both published and unpublished studies. The genes studied to date include CYP1A1, GSTT1, GSTM1, NAT2, CYP2E1, and CYP2D6.

22. The UK Biobank is a long-term initiative, which began in 2007. It is investigating the respective contribution of genetic predisposition and environmental exposure (nutrition, lifestyle and medication) to the development of disease. 400,000 participants had been recruited by November 2009. The study hopes to follow 500,000 volunteers in the UK aged 40-69 years for 25 years thereafter. The participants taking part complete an automated questionnaire, an interview about lifestyle, medical history and nutritional habits. Blood and urine samples are taken and preserved for DNA extraction and will be used to measure other important substances in the future.

23. EPIC is a European Wide prospective cohort study of the relationship between diet and cancer. EPIC started in 1993 with data and blood collections in 10 European countries; Germany, France, Denmark, Spain, Greece, Holland, Italy, Norway, UK and Sweden, coordinated by the WHO's IARC. The full cohort consists of 521,468 individuals (of whom 366,521 are women), mostly between the ages of 34-69 years. Up to 2004, there were over 24,000 new cases of cancer recorded among participants with the majority of cancers being those of the breast, colorectum, prostate, and lung. The study and analysis is ongoing but the key results for cancer retrieved in 2008 are

- 1) Increased fat intake increases risk of breast cancer
- 2) High levels of sex hormones increase risk of breast cancer
- 3) Obesity increases a number of cancer risks
- 4) High dietary fibre protects against bowel cancer

24. The GenAir study, which ran from 2002 to 2004, investigated the effects of air pollution and Environmental Tobacco Smoke (ETS) on cancers of the lung, bladder, pharynx and larynx in non-smokers in 9 European Countries. The study was conducted using a nested case-control design in the EPIC investigation, by means of biomarkers of exposure and susceptibility (n=1100 cases and 3000 controls). The key findings were:

- 1) Higher exposures to (traffic related) air pollutants can increase the risk of lung cancer among never smokers and ex smokers of at least 10 years.
- 2) ETS exposure during childhood showed an association with lung cancer, particularly in never smokers (Hazard Ratio = 3.63, 95% CI 1.19 – 11.11).
- 3) The study found that the presence of DNA adducts was associated with a subsequent risk of lung cancer
- 4) The amount of plasma DNA was found to be associated with overall cancer onset after adjustment for age and centre. They found that the presence of gene mutations in plasma DNA was predictive for development of bladder cancer, but not of other types included in the GenAir study.

### **Developments in the field of Exposure Assessment**

25. The precise contribution of individual environmental factors and their interaction with genes continues to be difficult to elucidate. The lack of accurate, quantitative measures of exposure is a source of uncertainty in epidemiological studies, limiting the power of such studies to enable definitive conclusions about the association between exposure and disease.

26. In recent years, progress has been made in the development of new technologies and methods for assessing human exposure to chemicals, dietary and lifestyle factors, infectious agents and other stressors. A meeting of the US

committee of Environmental Exposure Technology Development in 2004 identified a “toolbox” of methods that can be used alone or in combination to provide information about individual and population exposure scenarios. The methods identified included environmental sensors and Geographical Information Systems (GIS) to derive information about external environmental exposure and biologic sensors, toxicogenomics, toxicoproteomics and body burden assays to derive measurements of internal biological exposure.

27. Environmental sensors are devices capable of providing information from the macro environment level such as industrial effluents, to microscale exposure levels such as household, workplace and personal environment to nanoscale exposure levels at the point of human contact. There have been significant developments in the field of environmental sensors in the past decade. Microscale sensors, such as tiny, inexpensive, personal exposure monitors that use fluorescence or cell function to detect chemicals in nanoscale sample volumes, and macroscale sensors, such as infrared radiation–based monitors that detect sulphur and nitrogen oxides in industrial-stack effluents, have been developed making it easier and less expensive to measure actual exposures as well as environmental concentrations (Weis et al. 2005).

28. GIS technology is an increasingly popular tool for determining exposure in epidemiological studies. The GIS Spatiotemporal visualization tools can be used to produce smooth, continuous space–time maps that assign exposure estimates through a spatial query procedure, as well as smooth, continuous temporally variant histograms and scatter plots that allow for examination of relationships between exposure-related variables at any moment in time (Meliker et al. 2005).

29. Toxicogenomics is a field that identifies the activities, regulation and interaction of genes, mRNA transcripts (transcriptomics), proteins (proteomics), adducts (adductomics) and metabolites (metabolomics) in a biological sample. These technologies can each provide a “molecular signature” or “fingerprint” of exposure or early effect which may enhance our understanding of the mechanisms by which chemicals cause toxicity and contribute to disease, at a range of environmentally relevant exposure levels. These signatures and

ultimately, risk, induced by exposure, are determined by the unique genomic composition of each individual and biomarkers of susceptibility can be determined through genomic analyses. Toxicogenomic-based methods have been used to classify exposure to chemicals and environmental toxicants (examples include cadmium (Griffin et al., 2001), cigarette smoke (Spira et al., 2004) and arsenate (Fry et al., 2007)) and for classifying health outcomes such as prostate cancer (Posadas et al., 2004 and Troyer et al., 2004) and acute lymphoblastic leukemia (Holleman et al., 2004). Toxicogenomics provides an opportunity to obtain a more realistic view of exposure involving multiple chemicals at potentially low environmental concentrations and multiple biological response pathways (Weis et al., 2005). Limitations exist for toxicogenomic-based methods in exposure assessment including the lack of information on background levels of expression of various mRNA transcripts, proteins and metabolites. A useful database on toxicogenomics, started in 2004, is the Comparative Toxicogenomics Database (CTD; <http://ctd.mdibl.org>), which provides insights into the mechanisms of chemical actions, disease susceptibility, toxicity, and therapeutic drug interactions by curating and integrating data describing relationships between chemicals, genes/proteins, and human diseases (Mattingly et al., 2004).

30. Most human biomonitoring assays provide a quantitative, reliable measurement of chemicals, their metabolites or reaction products in biological samples but some assays lack the specificity and sensitivity needed to detect the broad range of compounds present in biological samples, have not been validated at background levels in the environment and many assays can only accommodate moderate sample throughput (Weis et al., 2005). Hair, pulmonary air, nails and saliva can also be used as matrices for biomonitoring purposes (Angerer et al., 2007). Biomonitoring measurements can be performed on chemical substances such as PAHs, phthalates, dioxins, aromatic amines, environmental tobacco smoke and volatile organic compounds and Angerers et al. (2007) provides a substantial list of specific biomarkers of internal exposure that can be detected in biological samples. New biomonitoring methods for detecting DNA and protein adducts, based on established technologies, have been published in the literature (Whyatt et al., 2001; Perera et al., 2004; Barr et al., 2003 and Koc and Swenberg, 2002).

## **Conclusions**

31. There has been substantial activity since the COC last reviewed this topic in 2001-2 on all aspects of GEI: technical, statistical, databases, consortia and other multi-group work, and a great deal has been published on the number and nature of associations. It will not possible to review all the work comprehensively. However, it might be useful for the Committee to review selected projects, as suggested below, and to consider to what extent the limitations in gene-environment studies highlighted in 2002 still exist, i.e.:

- i) Are study designs now adequate for purpose, in terms of power and statistical strategy?
- ii) Are hypotheses being lodged *a priori* with a third party and are findings being replicated?
- iii) Is the rationale for gene selection being stated; is this becoming redundant with genome-wide studies?
- iv) Are the functional consequences of SNPs identified in association studies being investigated, as proposed by COC, in a tiered approach?
- v) What is the status of polymorphisms of xenobiotic metabolising enzymes and cancer risk? Have any consistent and/or strong associations been identified?

## **Questions for the committee:**

- 1) Would the Committee like to review any specific project in detail or are there any other areas that the COC considers it would be useful to review?
- 2) Would the Committee be interested in reviewing the cancer relevant HuGENet reviews and meta-analysis studies available to date?
- 3) For Gene Wide Association Studies (GWAS), would the Committee agree that one approach that might be helpful to this discussion would be to look at GWAS and one particular endpoint such as bladder, breast, prostate or colorectum cancers?

4) Would the committee be interested in reviewing the area of field synopsis studies for bladder cancer and/or DNA repair genes and cancer phenotypes?

Secretariat

June 2010

## References

Angerer, J, Ewers, U, Wilhelm, M. Human biomonitoring: State of the art. International Journal of Hygiene and Environmental Health Volume 210, Issues 3-4, 22 May 2007, Pages 201-228.

Bailey, JA., Yavor, AM., Massa, HJ., Trask, BJ and Eichler, EE (2001) Segmental Duplications: Organization and Impact Within the Current Human Genome Project Assembly. *Genome Res.* 11: 1005-1017

Barr JR, Maggio VL, Barr DB, Turner WE, Sjodin A, Sandau CD, et al. New high-resolution mass spectrometric approach for the measurement of polychlorinated biphenyls and organochlorine pesticides in human serum. *J Chromatog B.* 2003;794(1):137–148.

Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* (2007) 447:1087–1093.

Ewing, B. & Green, P. (1998) Base-calling of automated sequencing traces using Phred. II. Error probabilities. *Genome Res.* 8, 186–194.

Ewing, B., Hiller, L. Wendl, M. & Green, P (1998). Base-calling of automated sequence traces using Phred. I. Accuracy assessment. *Genome Res.* 8, 175–185.

Fry RC, Navasumrit P, Valiathan C, Svensson JP, Hogan BJ, Luo M, Bhattacharya S, Kandjanapa K, Soontararuks S, Nookabkaew S, Mahidol C, Ruchirawat M, Samson (2007) Activation of inflammation/NF-kappaB signaling in infants born to arsenic-exposed mothers. *Genet.* 2007 Nov;3(11):e207.

Gold B, Kirchhoff T, Stefanov S, Lautenberger J, Viale A, Garber J, Friedman E, Narod S, Olshen AB, Gregersen P, Kosarin K, Olsh A, Bergeron J, Ellis NA, Klein RJ, Clark AG, Norton L, Dean M, Boyd J, Offit K.(2008) Genome-wide

association study provides evidence for a breast cancer risk locus at 6q22.33. *Proc Natl Acad Sci U S A.* 105(11):4340-5.

Gordon, D., Abajian, C. & Green, P (1998). Consed: A graphical tool for sequence finishing. *Genome Res.* 8, 195–202 (1998).

Griffin JL, Walker LA, Shore RF, Nicholson JK. Metabolic profiling of chronic cadmium exposure in the rat. *Chem Res Toxicol.* 2001;14(10):1428–1434.

Gudmundsson J, Sulem P, Manolescu A, et al. (2007). Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat. Genet.* (2007) 39:631–637.

Holleman A, Cheek MH, den Boer ML, Yang W, Veerman AJ, Kazemie PD, et al. Gene-expression patterns in drug-resistant acute lymphoblastic leukemia cells and response to treatment. *N Engl J Med.* 2004;351:533–542.

Hunter DJ, Kraft P, Jacobs KB, et al. (2007). A genome-wide association study identifies alleles in *FGFR2* associated with risk of sporadic postmenopausal breast cancer. *Nat. Genet.* 39:870–874.

IHGSC (International Human Genome Sequencing Consortium) (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860–921.

Kaiser, J. (1997) Environment institute lays plans for gene hunt. *Science* 278, 569–570.

Koc, H and Swenberg, JA. (2002). Applications of mass spectrometry for quantitation of DNA adducts. *Journal of Chromatography B* Volume 778, Issues 1-2, 5, Pages 323-343

Knoppers BM. (2005). *J Law Med Ethics.* 33(1):7-14.

Landi MT, Chatterjee N, Yu K, Goldin LR, Goldstein AM, Rotunno M, Mirabello L, Jacobs K, Wheeler W, Yeager M, Bergen AW, Li Q, Consonni D, Pesatori AC, Wacholder S, Thun M, Diver R, Oken M, Virtamo J, Albanes D, Wang Z, Burdette L, Doheny KF, Pugh EW, Laurie C, Brennan P, Hung R, Gaborieau V, McKay JD, Lathrop M, McLaughlin J, Wang Y, Tsao MS, Spitz MR, Wang Y, Krokan H, Vatten L, Skorpen F, Arnesen E, Benhamou S, Bouchard C, Metsapalu A, Vooder T, Nelis M, Väik K, Field JK, Chen C, Goodman G, Sulem P, Thorleifsson G, Rafnar T, Eisen T, Sauter W, Rosenberger A, Bickeböller H, Risch A, Chang-Claude J, Wichmann HE, Stefansson K, Houlston R, Amos CI, Fraumeni JF Jr, Savage SA, Bertazzi PA, Tucker MA, Chanock S, Caporaso NE. (2009) A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. *Am J Hum Genet.* 85(5):679-91.

Mattingly CJ, Colby GT, Rosenstein MC, Forrest JN Jr, Boyer JL. (2004). Promoting comparative molecular studies in environmental health research: an overview of the comparative toxicogenomics database (CTD). *Pharmacogenomics J.*; 4(1):5-8.

Meliker JR, Slotnick MJ, Avruskin GA, Kaufmann A, Fedewa SA, Goovaerts P, Jacquez GJ, Nriagu JO. (2007). Individual lifetime exposure to inorganic arsenic using a space-time information system. *Int Arch Occup Environ Health.* 80(3):184-97.

Olden, K. and Wilson, S (2000). Environmental health and genomics: visions and implications. *Nat. Rev. Genet.* 1, 149–153.

Pearson TA, Manolio TA. (2008). How to interpret a genome-wide association study. *JAMA.* 299(11):1335-44.

Perera FP, Rauh V, Whyatt RM, Tsai WY, Bernert JT, Tu YH, et al. (2004). Molecular evidence of an interaction between prenatal environmental exposures and birth outcomes in a multi-ethnic population. *Environ Health Perspect.*;112:626–630.

Posadas EM, Davidson B, Kohn EC.(2004). Proteomics and ovarian cancer: implications for diagnosis and treatment: a critical review of the recent literature. *Curr Opin Oncol.*;16(5):478–484

Sachidanandam et al., 2001

Song H, Ramus SJ, Tyrer J, Bolton KL, Gentry-Maharaj A, Wozniak E, Anton-Culver H, Chang-Claude J, Cramer DW, DiCioccio R, Dörk T, Goode EL, Goodman MT, Schildkraut JM, Sellers T, Baglietto L, Beckmann MW, Beesley J, Blaakaer J, Carney ME, Chanock S, Chen Z, Cunningham JM, Dicks E, Doherty JA, Dürst M, Ekici AB, Fenstermacher D, Fridley BL, Giles G, Gore ME, De Vivo I, Hillemanns P, Hogdall C, Hogdall E, Iversen ES, Jacobs IJ, Jakubowska A, Li D, Lissowska J, Lubiński J, Lurie G, McGuire V, McLaughlin J, Medrek K, Moorman PG, Moysich K, Narod S, Phelan C, Pye C, Risch H, Runnebaum IB, Severi G, Southey M, Stram DO, Thiel FC, Terry KL, Tsai YY, Tworoger SS, Van Den Berg DJ, Vierkant RA, Wang-Gohrke S, Webb PM, Wilkens LR, Wu AH, Yang H, Brewster W, Ziogas A; Australian Cancer (Ovarian) Study; Australian Ovarian Cancer Study Group; Ovarian Cancer Association Consortium, Houlston R, Tomlinson I, Whittemore AS, Rossing MA, Ponder BA, Pearce CL, Ness RB, Menon U, Kjaer SK, Gronwald J, Garcia-Closas M, Fasching PA, Easton DF, Chenevix-Trench G, Berchuck A, Pharoah PD, Gayther SA (2009). A genome-wide association study identifies a new ovarian cancer susceptibility locus on 9p22.2. *Nat Genet.* 41(9):996-1000.

Spira A, Schembri F, Beane J, Shah V, Liu G, Brody JS. (2004). Impact of cigarette smoke on the normal airway transcriptome. *Chest.* 125(5 Suppl):115S.

Stacey SN, Manolescu A, Sulem P, et al. (2007). Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat. Genet.* 39:865–869.

Tenesa A, Farrington SM, Prendergast JG, et al. (2008). Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nat. Genet.* 40:631–637.

Thomas G, Jacobs KB, Kraft P, Yeager M, Wacholder S, Cox DG, Hankinson SE, Hutchinson A, Wang Z, Yu K, Chatterjee N, Garcia-Closas M, Gonzalez-Bosquet J, Prokunina-Olsson L, Orr N, Willett WC, Colditz GA, Ziegler RG, Berg CD, Buys SS, McCarty CA, Feigelson HS, Calle EE, Thun MJ, Diver R, Prentice R, Jackson R, Kooperberg C, Chlebowski R, Lissowska J, Peplonska B, Brinton LA, Sigurdson A, Doody M, Bhatti P, Alexander BH, Buring J, Lee IM, Vatten LJ, Hveem K, Kumle M, Hayes RB, Tucker M, Gerhard DS, Fraumeni JF Jr, Hoover RN, Chanock SJ, Hunter DJ. (2009) A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). *Nat Genet*;41(5):579-84.

The Wellcome Trust Case Control Consortium. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447:661–678.

Tomlinson IP, Webb E, Carvajal-Carmona L, et al. (2008). A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nat. Genet.* 40:623–630.

Troyer DA, Mubiru J, Leach RJ, Naylor SL. (2004). Promise and challenge: markers of prostate cancer detection, diagnosis and prognosis. *Dis Markers.* 20(2):117–128.

Vineis P, Brennan P, Canzian F, Ioannidis JP, Matullo G, Ritchie M, Stromberg U, Taioli E, Thompson J. (2008). Expectations and challenges stemming from genome-wide association studies. *Mutagenesis.* 23(6):439-44.

Wang X, Pankratz VS, Fredericksen Z, Tarrell R, Karaus M, McGuffog L, Pharaoh PD, Ponder BA, Dunning AM, Peock S, Cook M, Oliver C, Frost D; EMBRACE, Sinilnikova OM, Stoppa-Lyonnet D, Mazoyer S, Houdayer C; GEMO, Hogervorst FB, Hoening MJ, Ligtenberg MJ; HEBON, Spurdle A,

Chenevix-Trench G; kConFab, Schmutzler RK, Wappenschmidt B, Engel C, Meindl A, Domchek SM, Nathanson KL, Rebbeck TR, Singer CF, Gschwantler-Kaulich D, Dressler C, Fink A, Szabo CI, Zikan M, Foretova L, Claes K, Thomas G, Hoover RN, Hunter DJ, Chanock SJ, Easton DF, Antoniou AC, Couch FJ.(2010) Common variants associated with breast cancer in genome-wide association studies are modifiers of breast cancer risk in BRCA1 and BRCA2 mutation carriers. *Human Mol Genet.* 2010 May 14.

Weis BK, Balshaw D, Barr JR, Brown D, Ellisman M, Liroy P, et al. (2005). Personalized exposure assessment: promising approaches for human health research *Environ Health Perspect* 113: 84048

Whyatt RM, Barr DB. 2001. Measurement of organophosphate metabolites in postpartum meconium as a potential biomarker of prenatal exposure: a validation study. *Environ Health Perspect* 109:417–420.

Yeager M, Orr N, Hayes RB, et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat. Genet.* (2007) 39:645–649.

**Annex 1 to CC/2010/02**