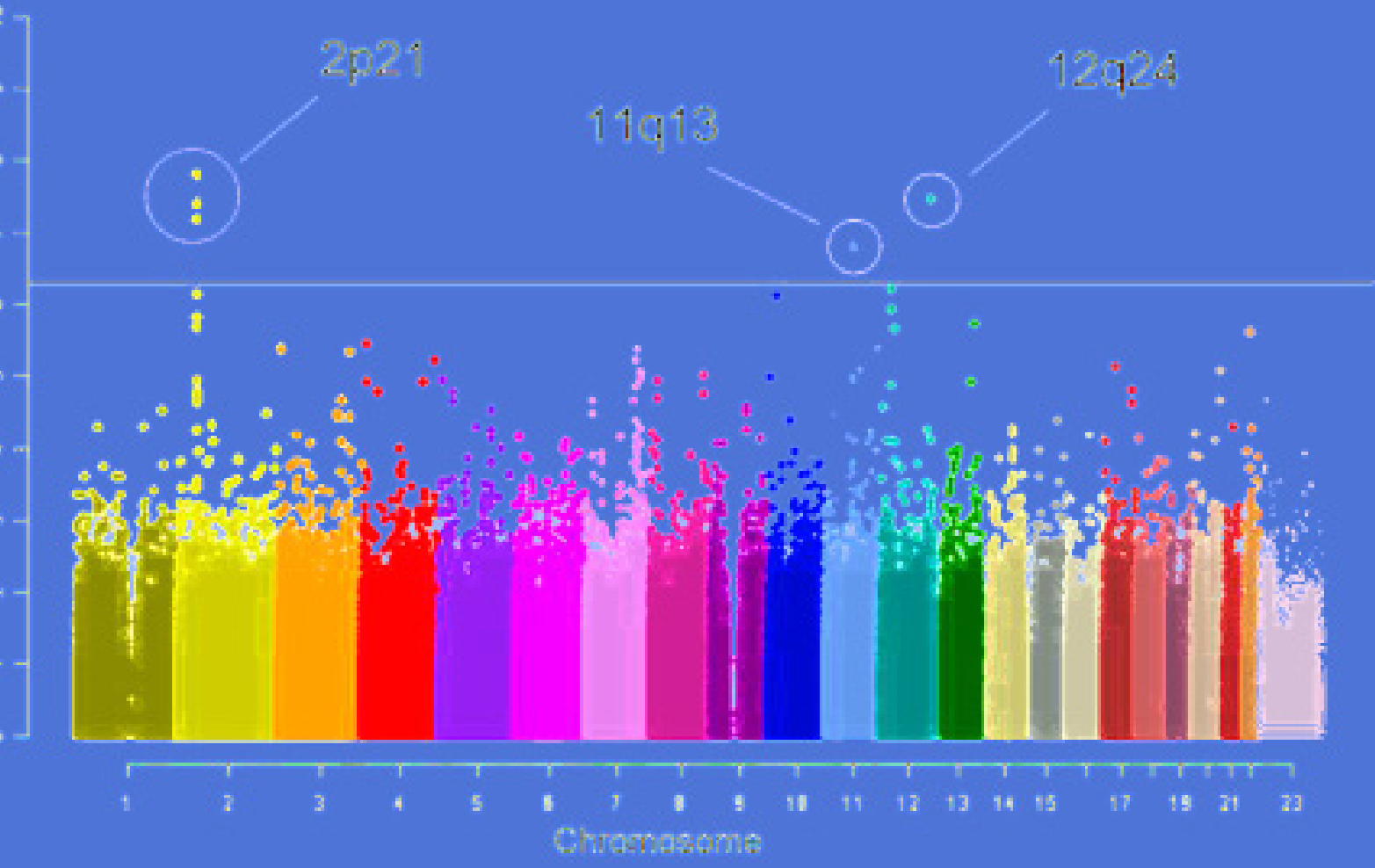
Section of Genetics

Section head
Dr Paul Brennan

Identifying specific genes and gene variants that contribute to the development of cancer will offer a greater understanding of biological pathways that lead to cancer, help elucidate how environmental factors may exert their effects in combination with genes, and aid in identifying individuals who are at high enough risk to benefit from existing risk reduction strategies. The Section of Genetics (GEN) comprises the Genetic Epidemiology Group (GEP), the Genetic Cancer Susceptibility Group (GCS) and the Biostatistics Group (BST), all with the overall mission of identifying genes involved in cancer, characterizing the spectrum of pathogenic sequence variants that they harbor, and understanding how they interact with non-genetic factors.

GEN projects usually involve extensive field work in collaboration with external investigators, so that large-scale epidemiological studies with appropriate clinical and biosample collections can be developed. The primary interest of GEN is the analysis and identification of common genetic susceptibility variants and their interaction with non-genetic risk factors. Genetic analysis comprises either candidate gene type studies (conducted in-house) or genome-wide association studies (GWAS) (currently conducted in collaboration with outside partners, although now also feasible in-house). GEP studies also assess non-genetic exposures, partly in recognition of the importance of non-genetic factors in driving cancer incidence, and also in order to facilitate accurate assessment of gene-environment interactions. By contrast, GCS has a focus on identification of uncommon or rare genetic variants that may have a larger effect than common single nucleotide polymorphisms (SNPs), but that are not sufficiently frequent to be captured by current GWAS genotyping arrays. Such DNA variants are identifiable through a case-control mutation screening approach. To this end, GCS has developed an in silico-driven strategy to stratify variants according to their probability of being pathogenic. Thus, the GCS research programme complements that of GEP, and also provides a facility for bioinformatics prediction on functionality of genetic variants, as well as a capacity to conduct in vitro experiments to validate functional hypotheses on variants of interest that are identified in both groups. In parallel to its own research activities, GCS also maintains and develops the genetic services platform (GSP) and related Laboratory Information Management System (LIMS) to support GEN large-scale molecular epidemiology projects and other IARC genomics projects. BST interacts at all stages to provide overall statistical support.


2p21
11q13
12q24
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
17
19
21
23
Chromosome

Genetic Epidemiology Group (GEP)

Group head
Dr Paul Brennan

Scientists
Dr James McKay
(until September 2010)
Dr Ghislaine Scélo
Dr Mattias Johansson
(from November 2010)
Dr Behnoush Abedi Ardekani

Technical assistants
Ms Valérie Gaborieau
Ms Hélène Renard

Laboratory technician
Ms Amélie Chabrier
(until June 2011)

Secretary
Ms Yvette Granjard

Postdoctoral fellows
Dr Devasena Anantharaman
Dr Dan Chen (until April 2011)
Dr Jajah Fachiroh
(until september 2011)
Dr Mattias Johansson
(until November 2010)
Dr Hooman Khademi Kohnehshahri
Dr Keitaro Matsuo
(until March 2010)
Dr Maria Timofeeva
Dr Kevin Urayama
(until August 2011)
Dr Magdalena Wozniak

Students
Ms Hana Atayallah
(until April 2010)
Mr Darren Brenner
(until September 2011)
Ms Manon Delahaye
(until December 2011)
Ms Erin Jacob
(until September 2010)
Ms Rossana Mendoza Lopez
(until July 2011)
Ms Rajini Nagrani
(until November 2010)

Visiting scientists
Dr David Conway
(until August 2011)
Dr Vinotsole Khamo
(until October 2010)
Professor John McLaughlin
(until October 2011)

The overall goal of the Genetic Epidemiology Group (GEP) is to identify genetic susceptibility variants of various cancer sites and study their interaction with environmental factors. An additional aim is to develop accurate risk prediction models that take both demographic information (e.g. age and sex) and biomarkers (genetic and non-genetic) into account. GEP focuses specifically on cancers related to tobacco and alcohol consumption, as well as rare cancers (e.g. nasopharyngeal cancer (NPC)). Many activities typically involve field work in order to recruit large numbers of cases and controls that have extensive questionnaire information and biological samples. For young onset cancers (such as NPC and childhood cancers) a trio design is also employed. Genetic analyses usually include a genome-wide approach initially, with subsequent large-scale coordinated replication studies in diverse populations. The latter is achieved by the development of international consortia in which GEP takes a leading role. Confirmed susceptibility loci are investigated in more detail with a variety of techniques, including *in silico*, expression and sequencing studies, often conducted in collaboration with other IARC groups. In addition to studies of genetic factors, GEP is conducting a wide range of investigations involving non-genetic factors, including evaluations of circulating biomarkers, such as human papillomavirus (HPV) antibodies, for head and neck cancers; cotinine for lung cancer; and dietary biomarkers for multiple cancers. GEP also performs extensive evaluations of questionnaire data, particularly data that have been collected during field work.

Genetic susceptibility of renal cell carcinoma

The primary GWAS conducted in 2010-2011 was a joint collaboration between IARC and the Centre National de Génotypage (CNG) and focused on renal cell carcinoma (RCC). Based on a series of studies, including a large case-control study from central Europe coordinated by GEP scientists (1400 cases and 2500 controls), the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort study (300 cases and 400 controls), and an additional three studies, the final dataset comprised approximately 2500 cases and 5000 controls. This data was combined with a parallel study coordinated by scientists at the US National Cancer Institute (NCI) (principally Stephen Chanock and Nat Rothman) with approximately 1300 cases and 3400 controls. A combined dataset of approximately 3800 cases and 7400 controls using standardized quality control thresholds and common variable definitions was developed, with coordinated analysis at both IARC and NCI. Further IARC samples for replication, which included 3000 cases and 5000 controls, were also incorporated. This study has resulted in the identification of three novel susceptibility loci for RCC, one of which, *EPAS1* on 2p21, encodes hypoxia-inducible- factor-2 alpha, a transcription factor previously implicated in RCC development (Purdue *et al*., 2011). This finding was notable in former and current smokers but not in non-smokers, suggesting an interaction with smoking (P heterogeneity = 0.004) (Figure 1). This observation raises the possibility that the effect of EPAS1 is dependent on tobacco smoking.

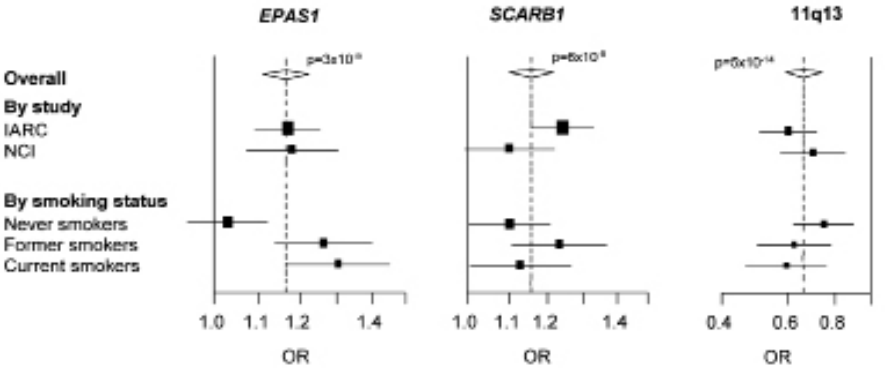


Figure 1. Forest plot shows odds ratios (OR) for three SNPs that were identified as susceptibility variants of renal cell carcinoma, overall and stratified by study and smoking status.

Genetic susceptibility for tobacco and alcohol related cancers

In 2010, we completed a GWAS of head and neck cancers, again in collaboration with the CNG. This involved an initial genome-wide analysis of over 2000 cases and a combined group of 8000 controls from IARC studies. Replication of 20 variants was conducted in a series of 13 independent studies participating in the International Head and Neck Cancer Epidemiology (INHANCE) consortium. Initial results further confirmed evidence for an important role of *ADH* genes in these cancers, as well as evidence for two additional susceptibility loci in regions 4p21 and 12q24 (McKay *et al*., 2011)

Genetic susceptibility to lymphomas

GEP is additionally investigating genetic susceptibility to lymphomas. Within our ongoing collaboration with the CNG, GEP is performing a GWAS of 1200 Hodgkin's lymphoma case-control pairs from eight European countries. Genome-wide genotyping and initial statistical analysis has been completed and multiple independent variants across 6p21 have been identified, two of which were specific for EBV status. An additional novel locus located within the IL13 gene was also identified, suggesting common genetic pathways with other immune-related phenotypes (Urayama *et al*., in press)

Lung cancer and non-genetic risk factors

Numerous studies have reported that fruits and vegetables are protective against lung cancer, and that the one-carbon metabolism pathway (i.e. folate pathway) has been suggested as a mechanism responsible for this protective effect. In order to investigate this hypothesis we analysed serum samples from 900 cases of lung cancer and 1800 matched controls that were collected prospectively within the EPIC cohort of 500 000 European subjects. Serum samples for all subjects, taken on average five years prior to diagnosis among the cases, were analysed for four B vitamins (B2, B6, folate and B12), methionine and homocysteine. After accounting for smoking, a substantial lower risk for lung cancer was seen for

elevated serum levels of B6, as well as for serum methionine. These associations were of sufficient statistical strength to exclude chance as explanation. Similar and consistent decreases in risk were observed in never, former and current smokers, indicating that results were not due to confounding by smoking. A lower risk was also seen for serum folate, although this was only apparent in former and current smokers. When participants were classified by median levels of serum methionine and B6, subjects with above median levels of both had an almost 60% lower lung cancer risk overall. Cumulative lung cancer risk calculations by smoking status and stratified by B6 and methionine levels, indicated important differences in risk associated with having above or below median levels of both vitamin B6 and methionine (Figure 2).

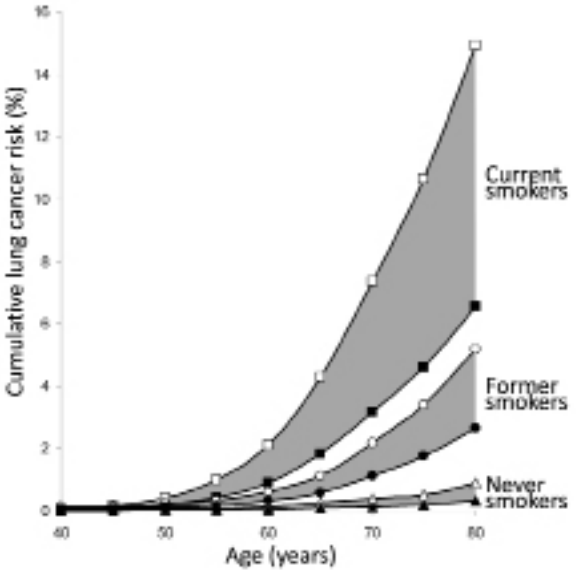


Figure 2. Cumulative risks of lung cancer up to age 79 among never, former and current smokers, stratified for men, and for having above (high/high) and below (low/low) median levels of both vitamin B6 and methionine, respectively.

International study of rare childhood embryonal tumours

Following a meeting held in IARC in 2006 of investigators with an interest in childhood cancers, a large-scale etiological study of rare childhood cancers was piloted in eight countries in 2008-2010 (France, UK, Canada, Australia, Canada, USA, Serbia, Macedonia and Czech Republic). A case-control trio design has been shown to be feasible in all countries participating in the pilot, and five additional countries (Brazil, Japan, India, the Netherlands and El Salvador) may start piloting the protocol in 2011. The aim of the International Study of Rare Childhood Embryonal Tumours (ISET) is to investigate the role of exposure to suspected factors at different key periods at the beginning of life (preconceptional, prenatal and postnatal), as well as genetic susceptibility factors, gene-environment interactions and novel molecular markers. Focusing on Wilms tumour and neuroblastoma, already more than 250 case trios and 1750 unrelated controls have been recruited, with approximately 650 and 450 case trios expected each year, respectively. The full scale study will expand to retinoblastoma, rhabdomyosarcoma and hepatoblastoma.

Priorities for 2012-2013

Genetics and genomics of renal cell carcinoma

GEP will continue to have a central role in the CAGEKID study, which is funded by the European Commission to conduct whole genome sequencing on tumour/non-tumour DNA pairs from at least 100 individuals with kidney cancer. On completion of sequencing, we will join other CAGEKID partners in a comprehensive evaluation of the germline and somatic variation that contributes to risk of RCC. This will be complemented by an examination of gene expression and epigenetic changes. GEP will ensure the availability of at least 100 RCC cases that correspond to the International Cancer Genome Consortium's criteria for full participation in the CAGEKID study, comprising whole genome sequencing and expression analysis of tumour/non-tumour pairs. Further, and in collaboration with colleagues at NCI, we will expand our GWAS of RCC to approximately 8000 cases and 16 000 controls. Funds for this expanded analysis have been provided by NCI.

Lung cancer genetics

Genome-wide data is available on over 15 000 lung cancer cases and 25 000 controls from eight different study groups (with IARC studies contributing about 25% of data). Initial meta-analysis of the majority of these studies has not identified any additional susceptibility loci beyond the three previously reported (5p15, 6p21 and 15q25), although limited subgroup analyses from individual studies has detected heterogeneity of effects for all three loci (by histology for 5p15, by smoking status for 15q25, and by geographic region for 6p21). It is therefore reasonable to expect that a coordinated analysis by subgroup may identify additional susceptibility loci in addition to those previously detected. In collaboration with others, we will carry out a stratified meta-analysis of all studies with a particular focus on identifying subgroup effects by histology, smoking status, family history, age, sex and stage of disease.

Head and neck cancers and human papillomavirus

Further analyses of the role of human papillomavirus (HPV) in head and neck cancers, and potential genetic modifiers, will include a coordinated analysis of the presence of HPV antibodies in pre-diagnostic blood samples within the EPIC cohort, with this being potentially expanded to other cohorts through the cohort consortium. This initiative is being led by GEP scientists along with scientists at NCI (Dr A Kraemer in the

Division of Cancer Epidemiology and Genetics). We will also evaluate the strong geographic differences in HPV prevalence in head and neck cancers by coordinating an analysis of HPV in a large series of samples from Europe, the US and Brazil.

A comprehensive evaluation of the one-carbon metabolism pathway in tobacco-related cancers

We plan to further evaluate the associations of B vitamins with lung cancer risk and risk prediction by initiating a consortium of prospective cohort studies within different populations from Europe, the US, Australia and Asia. The overall aims are to identify the consistency of these associations in different populations, the extent to which they are modified by measurement errors and day-to-day variations in vitamin status, and determine whether analyses of genes found to be associated with these biomarkers provide evidence of causality. In addition, we will assess the potential of using circulating biomarkers in lung cancer risk prediction models. These goals will be achieved by developing a Lung Cancer Cohort Consortium which will include biochemical analysis of at least 5000 lung cancer cases and comparable controls selected from over 20 participating cohorts, with equal proportions of never, former and current smokers. Funds to develop this consortium and to perform these analyses have been provided by the National Cancer Institute of the United States.

We will also extend this analysis to other types of cancer related to tobacco in order to test whether these associations are specific to lung cancer. Our immediate aim is to examine the role of B vitamins in head and neck cancers and RCC. This will initially be assessed within the EPIC cohort (approximately 500 case-control pairs for each cancer type) and may be extended to the cohort consortium based on initial results. These analyses within EPIC will be coordinated with our ongoing GWAS for both cancer types, thereby resulting in a large series of subjects with both genome-wide data and B vitamin measurements. Funds for this analysis have been provided by the World Cancer Research Fund.

Genetic epidemiology of nasopharyngeal cancer

In the short-term, we plan to complete the biorepository of 2000 nasopharyngeal cancer (NPC) cases and controls from the ongoing studies in Singapore, Thailand and Malaysia. Subsequently, we will initiate the evaluation of susceptibility loci that have been identified in Chinese GWAS in order to determine their repeatability in other Asian populations.

Large population cohorts

GEP will continue to coordinate two large population cohorts that were initiated by GEP and other non-IARC scientists. These include a prospective cohort of 200 000 adults from three cities of Siberia Russia (being conducted with the Cancer Research Centre, Moscow; and the Clinical Trials Services Unit, Oxford, UK). Analysis will focus on the role of alcohol on all causes of mortality. Also being investigated is the Golestan Cohort study of 50 000 individuals from Northeastern Iran, being conducted with colleagues from Tehran and NCI. GEP's analysis will focus on the effects of opium, obesity and hypertension on all causes and cause-specific mortality.

GEP is grateful to the following persons for their collaborations:

Marta Vilensky, Buenos Aires, Argentina; Victor Wünsch-Filho, Sao Paulo, Brazil; Rayjean Hung, Toronto, Canada; Lenka Foretova, Brno, Vladimir Janout and Eleonora Fabianova, Olomouc, Vladimir Bencko, Ivana Holcatova, Prague, Czech Republic; Mark Lathrop, Paris, France; Rajesh Dikshit, Mumbai, India; Reza Malekzadeh, Tehran, Iran; Beena Devi, Kuching, Malaysia; Per Magne Ueland, Bergen, Norway; Jolanta Lissowska, Warsaw, Poland; Dana Mates and Jinga Viorel, Bucharest, Romania; David Zaridze, Moscow, Russia ; Miodrag and Simona Ognjanovic, Belgrade, Serbia; Tam Ha, Singapore; Eleonora Fabianova, Banska Bystrica, Slovakia; Sulee Sangrajrang, Bangkok, Thailand; Paolo Vineis, London, Caroline Relton, Newcastle, Angus Roberts, Sheffield, UK; Nathaniel Rothman, Stephen Chanock, Bethesda, Chris Amos, Houston, Aimee Kreemer, Rockville, Jon Wakefield, Washington, USA.

Financial support from the following bodies is gratefully acknowledged:

European Commission, Brussels, Belgium
Institut National du Cancer, Paris, France
National Institutes of Health, USA
World Cancer Research Fund, London, UK

Publications

Anantharaman D, Marron M, Lagiou P *et al.* (2011). Population attributable risk of tobacco and alcohol for upper aerodigestive tract cancer. *Oral Oncol,* 47:725–731.doi:10.1016/j.oraloncology.2011.05.004 PMID:21684805

Andreotti G, Boffetta P, Rosenberg PS *et al.* (2010). Variants in blood pressure genes and the risk of renal cell carcinoma. *Carcinogenesis*, 31:614–620. doi:10.1093/carcin/bgp321 PMID:20047954

Baltar VT, Xun WW, Chuang SC *et al.* (2011). Smoking, secondhand smoke, and cotinine levels in a subset of EPIC cohort. *Cancer Epidemiol Biomarkers Prev,* 20:869–875.doi:10.1158/1055-9965.EPI-10-1235 PMID:21357382

Bosetti C, Scelo G, Chuang SC *et al.* (2011). High constant incidence rates of second primary cancers of the head and neck: a pooled analysis of 13 cancer registries. *Int J Cancer,* 129:173–179.doi:10.1002/ijc.25652 PMID:20824702

Brennan P, Hainaut P, Boffetta P (2011). Genetics of lung-cancer susceptibility. *Lancet Oncol,* 12:399–408.doi:10.1016/S1470-2045(10)70126-1 PMID:20951091

Broeks A, Schmidt MK, Sherman ME *et al.*; Genica Network; kConFab; AOCS (2011). Low penetrance breast cancer susceptibility loci are associated with specific breast tumor subtypes: findings from the Breast Cancer Association Consortium. *Hum Mol Genet,* 20:3289–3303.doi:10.1093/hmg/ddr228 PMID:21596841

Butterbach K, Beckmann L, de Sanjosé S *et al.* (2011). Association of JAK-STAT pathway related genes with lymphoma risk: results of a European case-control study (EpiLymph). *Br J Haematol,* 153:318–333.doi:10.1111/j.1365-2141.2011.08632.x PMID:21418178

Campa D, Hüsing A, Chang-Claude J *et al.* (2011). Genetic variability of the fatty acid synthase pathway is not associated with prostate cancer risk in the European Prospective Investigation on Cancer (EPIC). *Eur J Cancer,* 47:420–427.doi:10.1016/j.ejca.2010.09.029 PMID:20965718

Campa D, Hüsing A, Dostal L *et al.* (2011). Genetic variability of the forkhead box O3 and prostate cancer risk in the European Prospective Investigation on Cancer. *Oncol Rep,* 26:979–986. PMID:21725602

Campa D, Hüsing A, Stein A *et al.* (2011). Genetic variability of the mTOR pathway and prostate cancer risk in the European Prospective Investigation on Cancer (EPIC). *PLoS One,* 6:e16914.doi:10.1371/journal.pone.0016914 PMID:21373201

Campa D, Kaaks R, Le Marchand L *et al.* (2011). Interactions between genetic variants and breast cancer risk factors in the breast and prostate cancer cohort consortium. *J Natl Cancer Inst,* 103:1252–1263.doi:10.1093/jnci/djr265 PMID:21791674

Chang JS, Wiemels JL, Chokkalingam AP *et al.* (2010). Genetic polymorphisms in adaptive immunity genes and childhood acute lymphoblastic leukemia. *Cancer Epidemiol Biomarkers Prev,* 19:2152–2163.doi:10.1158/1055-9965.EPI-10-0389 PMID:20716621

Chen D, McKay JD, Clifford G *et al.* (2011). Genome-wide association study of HPV seropositivity. *Hum Mol Genet.* doi:10.1093/hmg/ddr383 PMID:21896673

Chen D, Truong T, Gaborieau V *et al.* (2011). A sex-specific association between a 15q25 variant and upper aerodigestive tract cancers. *Cancer Epidemiol Biomarkers Prev,* 20:658–664.doi:10.1158/1055-9965.EPI-10-1008 PMID:21335511

Chu A, Heck JE, Ribeiro KB *et al.* (2010). Wilms' tumour: a systematic review of risk factors and meta-analysis. *Paediatr Perinat Epidemiol,* 24:449–469.doi:10.1111/j.1365-3016.2010.01133.x PMID:20670226

Chuang SC, Scélo G, Lee YC *et al.* (2010). Risks of second primary cancer among patients with major histological types of lung cancers in both men and women. *Br J Cancer,* 102:1190–1195.doi:10.1038/sj.bjc.6605616 PMID:20354532

Cocco P, t'Mannetje A, Fadda D *et al.* (2010). Occupational exposure to solvents and risk of lymphoma subtypes: results from the Epilymph case-control study. *Occup Environ Med*, 67:341–347.doi:10.1136/oem.2009.046839 PMID:20447988

Conde L, Halperin E, Akers NK *et al.* (2010). Genome-wide association study of follicular lymphoma identifies a risk locus at 6p21.32. *Nat Genet,* 42:661–664.doi:10.1038/ng.626 PMID:20639881

Conway DI, McKinney PA, McMahon AD *et al.* (2010). Socioeconomic factors associated with risk of upper aerodigestive tract cancer in Europe. *Eur J Cancer,* 46:588–598.doi:10.1016/j.ejca.2009.09.028 PMID:19857956

De Stefani E, Ronco AL, Deneo-Pellegrini H *et al.* (2010). Dietary patterns and risk of advanced prostate cancer: a principal component analysis in Uruguay. *Cancer Causes Control,* 21:1009–1016. doi:10.1007/s10552-010-9527-4 PMID:20198507

Does M, Scélo G, Metayer C *et al.* (2011). Exposure to electrical contact currents and the risk of childhood leukemia. *Radiat Res,* 175:390–396. doi:10.1667/RR2357.1 PMID:21388283

Ekburanawat W, Ekpanyaskul C, Brennan P *et al.* (2010). Evaluation of non-viral risk factors for nasopharyngeal carcinoma in Thailand: results from a case-control study. *Asian Pac J Cancer Prev,* 11:929–932. PMID:21133603

Elliott KS, Zeggini E, McCarthy MI *et al.*; Australian Melanoma Family Study Investigators; PanScan Consortium (2010). Evaluation of association of HNF1B variants with diverse cancers: collaborative analysis of data from 19 genome-wide association studies. *PLoS One,* 5:e10858.doi:10.1371/journal.pone.0010858 PMID:20526366

Furberg H, Kim YJ, Dackor J *et al.*; Tobacco and Genetics Consortium (2010). Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet,* 42:441–447. doi:10.1038/ng.571 PMID:20418890

Gaudet MM, Olshan AF, Chuang SC *et al.* (2010). Body mass index and risk of head and neck cancer in a pooled analysis of case-control studies in the International Head and Neck Cancer Epidemiology (INHANCE) Consortium. *Int J Epidemiol,* 39:1091–1102.doi:10.1093/ije/dyp380 PMID:20123951

Grulich AE, Vajdic CM, Falster MO *et al.* (2010). Birth order and risk of non-hodgkin lymphoma–true association or bias? *Am J Epidemiol,* 172:621–630. doi:10.1093/aje/kwq167 PMID:20720098

Gu F, Schumacher FR, Canzian F *et al.* (2010). Eighteen insulin-like growth factor pathway genes, circulating levels of IGF-I and its binding protein, and risk of prostate and breast cancer. *Cancer Epidemiol Biomarkers Prev,* 19:2877–2887.doi:10.1158/1055-9965.EPI-10-0507 PMID:20810604

Heck JE, Berthiller J, Vaccarella S *et al.* (2010). Sexual behaviours and the risk of head and neck cancers: a pooled analysis in the International Head and Neck Cancer Epidemiology (INHANCE) consortium. *Int J Epidemiol,* 39:166–181.doi:10.1093/ije/dyp350 PMID:20022926

Heck JE, Charbotel B, Moore LE *et al.* (2010). Occupation and renal cell cancer in Central and Eastern Europe. *Occup Environ Med,* 67:47–53. doi:10.1136/oem.2009.046250 PMID:19737732

Herceg Z, Vaissière T (2011). Epigenetic mechanisms and cancer: an interface between the environment and the genome. *Epigenetics*, 6:804–819.doi:10.4161/epi.6.7.16262 PMID:21758002

Hermann S, Rohrmann S, Linseisen J *et al.* (2010). Level of education and the risk of lymphoma in the European prospective investigation into cancer and nutrition. *J Cancer Res Clin Oncol,* 136:71–77. doi:10.1007/s00432-009-0638-9 PMID:19582474

Hosgood HD 3rd, Boffetta P, Greenland S *et al.* (2010). In-home coal and wood use and lung cancer risk: a pooled analysis of the International Lung Cancer Consortium. *Environ Health Perspect,* 118:1743–1747.doi:10.1289/ehp.1002217 PMID:20846923

Hsiung CA, Lan Q, Hong YC *et al.* (2010). The 5p15.33 locus is associated with risk of lung adenocarcinoma in never-smoking females in Asia. *PLoS Genet,* 6:e1001051.doi:10.1371/journal.pgen.1001051 PMID:20700438

Hudson TJ, Anderson W, Artez A *et al.*; International Cancer Genome Consortium (2010). International network of cancer genome projects. *Nature*, 464:993–998.doi:10.1038/nature08987 PMID:20393554

Johansson I, Van Guelpen B, Hultdin J *et al.* (2010). Validity of food frequency questionnaire estimated intakes of folate and other B vitamins in a region without folic acid fortification. *Eur J Clin Nutr,* 64:905–913.doi:10.1038/ejcn.2010.80 PMID:20502473

Johansson M, Relton C, Ueland PM *et al.* (2010). Serum B vitamin levels and risk of lung cancer. *JAMA*, 303:2377–2385.doi:10.1001/jama.2010.808 PMID:20551408

Karami S, Boffetta P, Brennan P *et al.* (2011). Renal cancer risk and occupational exposure to polycyclic aromatic hydrocarbons and plastics. *J Occup Environ Med,* 53:218–223.doi:10.1097/JOM.0b013e31820a40a3 PMID:21270648

Karami S, Boffetta P, Stewart P *et al.* (2010). Occupational sunlight exposure and risk of renal cell carcinoma. *Cancer,* 116:2001–2010.doi:10.1002/cncr.24939 PMID:20213683

Karami S, Boffetta P, Stewart PS *et al.* (2011). Occupational exposure to dusts and risk of renal cell carcinoma. *Br J Cancer,* 104:1797–1803. doi:10.1038/bjc.2011.148 PMID:21540858

Karami S, Brennan P, Navratilova M *et al.* (2010). Vitamin d pathway genes, diet, and risk of renal cell carcinoma. *Int J Endocrinol,* 2010:879362. doi:10.1155/2010/879362 PMID:20049159

Khademi H, Etemadi A, Kamangar F *et al.* (2010). Verbal autopsy: reliability and validity estimates for causes of death in the Golestan Cohort Study in Iran. *PLoS One,* 5:e11183.doi:10.1371/journal.pone.0011183 PMID:20567597

Kiran S, Cocco P, Mannetje A *et al.* (2010). Occupational exposure to ethylene oxide and risk of lymphoma. *Epidemiology*, 21:905–910.doi:10.1097/EDE.0b013e3181f4cc0f PMID:20811284

Krais AM, Hautefeuille AH, Cros MP *et al.* (2011). CHRNA5 as negative regulator of nicotine signaling in normal and cancer bronchial cells: effects on motility, migration and p63 expression. *Carcinogenesis*, 32:1388–1395.doi:10.1093/carcin/bgr090 PMID:21586512

Laantri N, Jalbout M, Khyatti M *et al.* (2011). XRCC1 and hOGG1 genes and risk of nasopharyngeal carcinoma in North African countries. *Mol Carcinog,* 50:732–737.doi:10.1002/mc.20754 PMID:21520294

Landi MT, Chatterjee N, Caporaso NE *et al.* (2010). GPC5 rs2352028 variant and risk of lung cancer in never smokers. *Lancet Oncol,* 11:714–716, author reply 716.doi:10.1016/S1470-2045(10)70096-6 PMID:20688270

Lindstrom S, Schumacher F, Siddiq A *et al.* (2011). Characterizing associations and SNP-environment interactions for GWAS-identified prostate cancer risk markers–results from BPC3. *PLoS One,* 6:e17142. doi:10.1371/journal.pone.0017142 PMID:21390317

Linseisen J, Rohrmann S, Bueno-de-Mesquita B *et al.* (2011). Consumption of meat and fish and risk of lung cancer: results from the European Prospective Investigation into Cancer and Nutrition. *Cancer Causes Control,* 22:909–918.doi:10.1007/s10552-011-9764-1 PMID:21479828

Lips EH, Gaborieau V, McKay JD *et al.* (2010). Association between a 15q25 gene variant, smoking quantity and tobacco-related cancers among 17 000 individuals. *Int J Epidemiol,* 39:563–577. doi:10.1093/ije/dyp288 PMID:19776245

Lissowska J, Foretova L, Dabek J *et al.* (2010). Family history and lung cancer risk: international multicentre case-control study in Eastern and Central Europe and meta-analyses. *Cancer Causes Control,* 21:1091–1104.doi:10.1007/s10552-010-9537-2 PMID:20306329

Lubin JH, Gaudet MM, Olshan AF *et al.* (2010). Body mass index, cigarette smoking, and alcohol consumption and cancers of the oral cavity, pharynx, and larynx: modeling odds ratios in pooled case-control data. *Am J Epidemiol,* 171:1250–1261. doi:10.1093/aje/kwq088 PMID:20494999

Lubin JH, Muscat J, Gaudet MM *et al.* (2011). An examination of male and female odds ratios by BMI, cigarette smoking, and alcohol consumption for cancers of the oral cavity, pharynx, and larynx in pooled data from 15 case-control studies. *Cancer Causes Control,* 22:1217–1231.doi:10.1007/s10552-011-9792-x PMID:21744095

Macfarlane TV, Macfarlane GJ, Oliver RJ *et al.* (2010). The aetiology of upper aerodigestive tract cancers among young adults in Europe: the ARCAGE study. *Cancer Causes Control,* 21:2213–2221.doi:10.1007/s10552-010-9641-3 PMID:20835759

Macfarlane TV, Macfarlane GJ, Thakker NS *et al.* (2011). Role of medical history and medication use in the aetiology of upper aerodigestive tract cancers in Europe: the ARCAGE study. *Ann Oncol.* doi:10.1093/annonc/mdr335 PMID:21828376

Marron M, Boffetta P, Zhang ZF *et al.* (2010). Cessation of alcohol drinking, tobacco smoking and the reversal of head and neck cancer risk. *Int J Epidemiol,* 39:182–196.doi:10.1093/ije/dyp291 PMID:19805488

Maule M, Scélo G, Pastore G *et al.* (2011). Second malignancies after childhood noncentral nervous system solid cancer: Results from 13 cancer registries. *Int J Cancer,* n/a.doi:10.1002/ijc.26135 PMID:21520035

McKay JD, Truong T, Gaborieau V *et al.* (2011). A genome-wide association study of upper aerodigestive tract cancers conducted within the INHANCE consortium. *PLoS Genet*, 7:e1001333. doi:10.1371/journal.pgen.1001333 PMID:21437268

Metayer C, Scélo G, Chokkalingam AP *et al.* (2011). Genetic variants in the folate pathway and risk of childhood acute lymphoblastic leukemia. *Cancer Causes Control,* 22:1243–1258.doi:10.1007/s10552-011-9795-7 PMID:21748308

Moore LE, Boffetta P, Karami S *et al.* (2010). Occupational trichloroethylene exposure and renal carcinoma risk: evidence of genetic susceptibility by reductive metabolism gene variants. *Cancer Res,* 70:6527–6536.doi:10.1158/0008-5472.CAN-09-4167 PMID:20663906

Nieters A, Bracci PM, de Sanjosé S *et al.* (2011). A functional TNFRSF5 polymorphism and risk of non-Hodgkin lymphoma, a pooled analysis. *Int J Cancer,* 128:1481–1485.doi:10.1002/ijc.25420 PMID:20473910

Olsson AC, Fevotte J, Fletcher T *et al.* (2010). Occupational exposure to polycyclic aromatic hydrocarbons and lung cancer risk: a multicenter study in Europe. *Occup Environ Med,* 67:98–103.doi:10.1136/oem.2009.046680 PMID:19773276

Olsson AC, Gustavsson P, Kromhout H *et al.* (2011). Exposure to diesel motor exhaust and lung cancer risk in a pooled analysis from case-control studies in Europe and Canada. *Am J Respir Crit Care Med,* 183:941–948. doi:10.1164/rccm.201006-0940OC PMID:21037020

Paliwal A, Vaissière T, Krais A *et al.* (2010). Aberrant DNA methylation links cancer susceptibility locus 15q25.1 to apoptotic regulation and lung cancer. *Cancer Res,* 70:2779–2788.doi:10.1158/0008-5472.CAN-09-4550 PMID:20332232

Park SL, Lee YC, Marron M *et al.* (2011). The association between change in body mass index and upper aerodigestive tract cancers in the ARCAGE project: multicenter case-control study. *Int J Cancer,* 128:1449–1461.doi:10.1002/ijc.25468 PMID:20506380

Purdue MP, Johansson M, Zelenika D *et al.* (2011). Genome-wide association study of renal cell carcinoma identifies two susceptibility loci on 2p21 and 11q13.3. *Nat Genet,* 43:60–65.doi:10.1038/ng.723 PMID:21131975

Ribeiro KB, Levi JE, Pawlita M *et al.* (2011). Low human papillomavirus prevalence in head and neck cancer: results from two large case-control studies in high-incidence regions. *Int J Epidemiol*, 40:489–502. doi:10.1093/ije/dyq249 PMID:21224273

Rothman N, Garcia-Closas M, Chatterjee N *et al.* (2010). A multi-stage genome-wide association study of bladder cancer identifies multiple susceptibility loci. *Nat Genet,* 42:978–984.doi:10. 1038/ng.687 PMID:20972438

Samoli E, Lagiou A, Nikolopoulos E *et al.* (2010). Mediterranean diet and upper aerodigestive tract cancer: the Greek segment of the Alcohol-Related Cancers and Genetic Susceptibility in Europe study. *Br J Nutr,* 104:1369–1374.doi:10.1017/S0007114510002205 PMID:20500931

Schmeisser N, Conway DI, McKinney PA *et al.* (2010). Life course social mobility and risk of upper aerodigestive tract cancer in men. *Eur J Epidemiol,* 25:173–182.doi:10.1007/s10654-010-9429-5 PMID:20143252

Skibola CF, Bracci PM, Nieters A *et al.* (2010). Tumor necrosis factor (TNF) and lymphotoxin-alpha (LTA) polymorphisms and risk of non-Hodgkin lymphoma in the InterLymph Consortium. *Am J Epidemiol,* 171:267–276. doi:10.1093/aje/kwp383 PMID:20047977

Szymańska K, Hung RJ, Wünsch-Filho V *et al.* (2011). Alcohol and tobacco, and the risk of cancers of the upper aerodigestive tract in Latin America: a case-control study. *Cancer Causes Control,* 22:1037–1046.doi:10.1007/s10552-011-9779-7 PMID:21607590

Szymańska K, Levi JE, Menezes A *et al.* (2010). TP53 and EGFR mutations in combination with lifestyle risk factors in tumours of the upper aerodigestive tract from South America. *Carcinogenesis*, 31:1054–1059.doi:10.1093/carcin/bgp212 PMID:19955396

Szymańska K, Matos E, Hung RJ *et al.* (2010). Drinking of maté and the risk of cancers of the upper aerodigestive tract in Latin America: a case-control study. *Cancer Causes Control,* 21:1799–1806.doi:10.1007/s10552-010-9606-6 PMID:20623173

Szymańska K, Moore LE, Rothman N *et al.* (2010). TP53, EGFR, and KRAS mutations in relation to VHL inactivation and lifestyle risk factors in renal-cell carcinoma from central and eastern Europe. *Cancer Lett,* 293:92–98. doi:10.1016/j.canlet.2009.11.024 PMID:20137853

Taverne CW, Verheus M, McKay JD *et al.* (2010). Common genetic variation of insulin-like growth factor-binding protein 1 (IGFBP-1), IGFBP-3, and acid labile subunit in relation to serum IGF-I levels and mammographic density. *Breast Cancer Res Treat,* 123:843–855.doi:10.1007/s10549-010-0778-2 PMID:20155489

Timofeeva MN, McKay JD, Smith GD *et al.* (2011). Genetic Polymorphisms in 15q25 and 19q13 Loci, Cotinine Levels, and Risk of Lung Cancer in EPIC. *Cancer Epidemiol Biomarkers Prev.* PMID:21862624

Timpson NJ, Brennan P, Gaborieau V *et al.* (2010). Can lactase persistence genotype be used to reassess the relationship between renal cell carcinoma and milk drinking? Potentials and problems in the application of Mendelian randomization. *Cancer Epidemiol Biomarkers Prev,* 19:1341–1348.doi:10.1158/1055-9965.EPI-09-1019 PMID:20447925

Truong T, Hung RJ, Amos CI *et al.* (2010). Replication of lung cancer susceptibility loci at chromosomes 15q25, 5p15, and 6p21: a pooled analysis from the International Lung Cancer Consortium. *J Natl Cancer Inst,* 102:959–971.doi:10.1093/jnci/djq178 PMID:20548021

Truong T, Sauter W, McKay JD *et al.*; EPIC-lung (2010). International Lung Cancer Consortium: coordinated association study of 10 potential lung cancer susceptibility variants. *Carcinogenesis*, 31:625–633. doi:10.1093/carcin/bgq001 PMID:20106900

Urayama KY, Holcatova I, Janout V *et al.* (2011). Body mass index and body size in early adulthood and risk of pancreatic cancer in a central European multicenter case-control study. *Int J Cancer.* doi:10.1002/ijc.25959 PMID:21520034

van Bemmel DM, Boffetta P, Liao LM *et al.* (2011). Comprehensive analysis of 5-aminolevulinic acid dehydrogenase (ALAD) variants and renal cell carcinoma risk among individuals exposed to lead. *PLoS One,* 6:e20432.doi:10.1371/journal.pone.0020432 PMID:21799727

Vineis P, Chuang SC, Vaissière T *et al.*; Genair-EPIC Collaborators (2011). DNA methylation changes associated with cancer risk factors and blood levels of vitamin metabolites in a prospective study. *Epigenetics,* 6:195–201.doi:10.4161/epi.6.2.13573 PMID:20978370

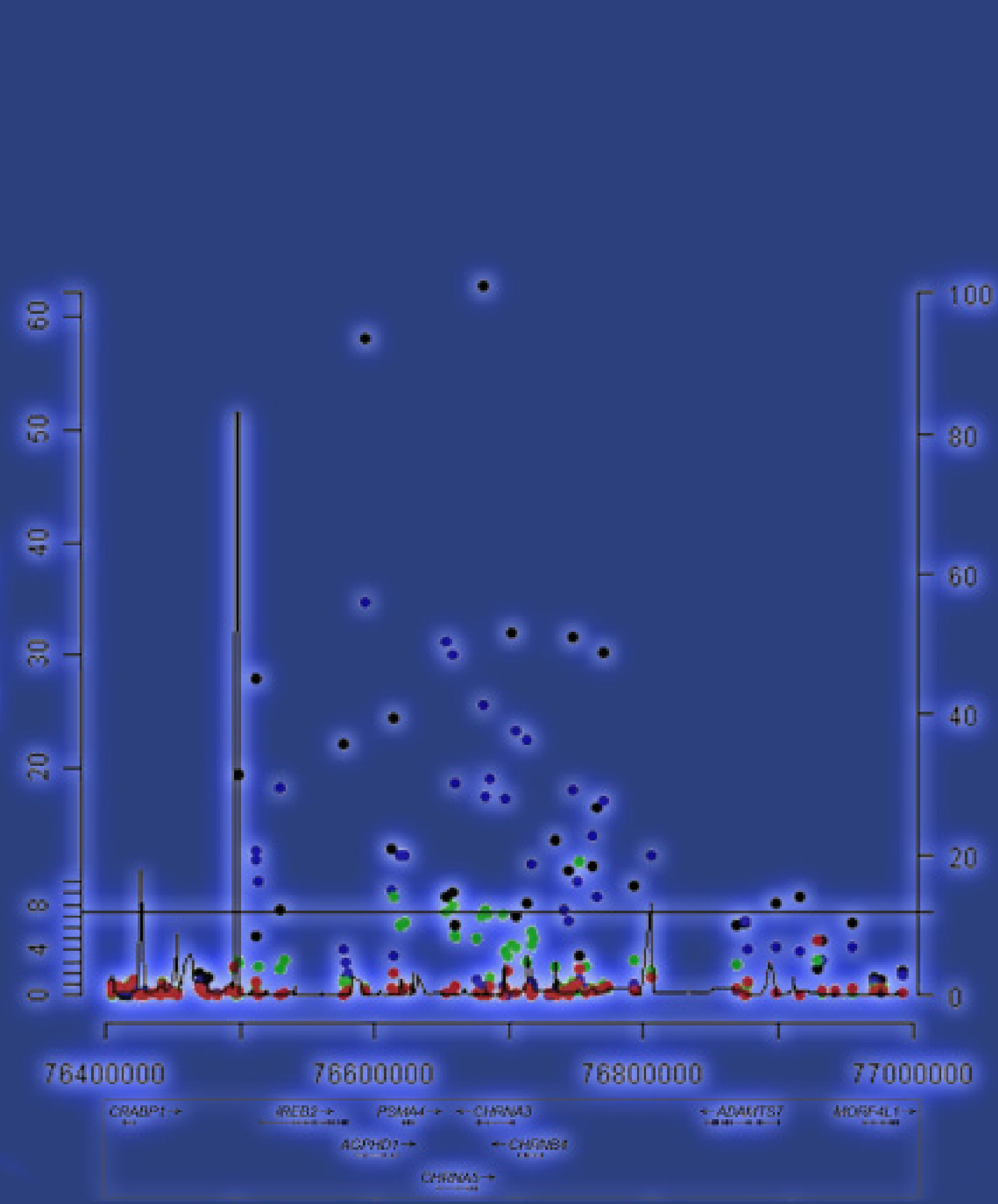
Wei S, Niu J, Zhao H *et al.* (2011). Association of a novel functional promoter variant (rs2075533 C>T) in the apoptosis gene TNFSF8 with risk of lung cancer–a finding from Texas lung cancer genome-wide association study. *Carcinogenesis*, 32:507–515. doi:10.1093/carcin/bgr014 PMID:21292647

Wiemels JL, Kang M, Chang JS *et al.* (2010). Backtracking RAS mutations in high hyperdiploid childhood acute lymphoblastic leukemia. *Blood Cells Mol Dis,* 45:186–191.doi:10.1016/j.bcmd.2010.07.007 PMID:20688547

Xun WW, Brennan P, Tjonneland A *et al.* (2011). Single-nucleotide polymorphisms (5p15.33, 15q25.1, 6p22.1, 6q27 and 7p15.3) and lung cancer survival in the European Prospective Investigation into Cancer and Nutrition (EPIC). *Mutagenesis*, 26:657–666. doi:10.1093/mutage/ger030 PMID:21750227

Yang XR, Chang-Claude J, Goode EL *et al.* (2011). Associations of breast cancer risk factors with tumor subtypes: a pooled analysis from the Breast Cancer Association Consortium studies. *J Natl Cancer Inst,* 103:250–263.doi:10.1093/jnci/djq526 PMID:21191117

Yu H, Zhao H, Wang LE *et al.* (2011). An analysis of single nucleotide polymorphisms of 125 DNA repair genes in the Texas genome-wide association study of lung cancer with a replication for the XRCC4 SNPs. *DNA Repair (Amst),* 10:398–407.doi:10.1016/j.dnarep.2011.01.005 PMID:21296624


0
4
8
20
30
40
50
60
0
20
40
60
80
100
76400000
76600000
76800000
77000000
GRABP1
IRF3
PSMA4
CHRNA3
ADAMTS7
MORF4L1
ACP1
CHRNB4
CHRNA5

Genetic Cancer Susceptibility Group (GCS)

Group head
Dr James McKay
(from September 2010)

Scientists
Dr Fabienne Lesueur
Dr Florence Le Calvez-Kelm

Laboratory technicians
Ms Sandrine Chopin-McKay
(until October 2010)
Ms Amélie Chabrier
(since July 2011)
Mr Geoffroy Durand
Ms Nathalie Forey
Ms Jocelyne Michelon
Ms Nivonirina Robinot

Bioinformatician
Ms Catherine Voegele

Secretary
Ms Antoinette Trochard

Postdoctoral fellows
Dr Jamil Ahmad
(until November 2010)
Dr Francesca Damiola
(until October 2011)
Dr Mona Ellaithi
Dr Javier Oliver
Dr Maroulio Pertesi
(since March 2011)
Dr Dewajani Purnomosari

Students
Ms Manon Delahaye
Ms Aurélie Fillon (until July 2010)
Ms Célia Jolivet (until June 2011)
Ms Bin Thieu Tù Nguyen-Dumont
(until December 2010)
Mr Wee Loon Ong (until August 2010)
Ms Fanny Paquet
Ms Maroulio Pertesi
(until December 2010)
Mr Maxime Vallée

The focus of the Genetic Cancer Susceptibility Group (GCS) is to investigate the contribution of inherited genetic factors to the etiology of cancer. An integrative approach is used, applying a variety of genomics-based techniques, including genotyping, mutation screening and expression studies, coupled with bioinformatics analysis, to both family- and case-control-based study designs. Of particular interest is the contribution of variants with a relatively rare population frequency, although we remain involved with studies of common SNPs through our collaboration with GEP. GCS also maintains and develops the Genetic Services Platform (GSP) and related Laboratory Information Management System (LIMS) to support genomics-based projects within the GEN section, as well as other IARC genomics groups.

Investigation of common and rare genetic variants in breast cancer susceptibility genes

GCS aims to measure the genetic risk attributable to the different types of genetic variation in breast cancer. These projects integrate data from case-control genotyping, case-control mutation screening, bioinformatics and allelic imbalance expression studies to identify dysfunctional variants responsible for cancer susceptibility, with a focus on breast cancer susceptibility genes and the characterization of new moderate- to high-risk susceptibility alleles.

Bioinformatics tools can quantify the functional consequence of the variants in silico using the degree of evolutionary conservation observed at this site. We have applied our case-control mutation screening approach to the breast cancer susceptibility genes *ATM* (Tavtigian *et al.*, 2008) and *CHEK2* (Le Calvez *et al.*, 2011) to demonstrate the efficiency of ranking rare missense substitutions using in silico programs before comparing the distribution and frequencies of the different types of variants in a series of early-onset breast cancer cases and controls. Although loss-of-function mutations in *ATM* and *CHEK2* have been associated with intermediate-risk of breast cancer, our strategy allowed us to demonstrate that a subset of rare missense substitutions make a comparable contribution to disease susceptibility.

Differential allelic expression in CHEK2 alleles

Differential allelic expression (DAE) occurs when two alleles of a particular gene are expressed unequally. DAE can result in quite major expression level differences between alleles, for example when a truncating mutation results in nonsense mediated mRNA decay or when more subtle differences in allele expression levels are due to a sequence variant in a regulatory element. Our initial DAE assay, based on high-resolution melting curve analysis and probes designed for a genetic marker located in the target gene's mRNA, has demonstrated DAE in lymphoblastoid cell line mRNA in heterozygote carriers of the c.1100delC truncating mutation in the *CHEK2* cancer susceptibility gene (Nguyen-Dumont *et al.*, 2010). We are now developing potentially more sensitive and higher throughput DAE tests by taking advantage of possibilities offered by the massive parallel technologies.

In collaboration with GEP, GCS has additionally provided genetic and analytical expertise for the completion of multiple genome-wide association studies (GWAS), notably, HNSCC (McKay *et al.*, 2011) and kidney cancers (Purdue *et al.*, 2011). We are actively pursuing GWAS of Hodgkin's lymphoma and cancers of the oral cavity. The Genetic Services Platform in GCS additionally plays an important role in GWAS by performing genotyping to quickly replicate and validate findings.

The Genetic Services Platform

Nested within the GCS, the Genetic Services Platform (GSP) aims to implement cutting edge genomics techniques and make them available to all IARC groups, along with relevant technical expertise and support. Genomics applications are piloted prior to implementation to ensure their suitability to IARC's large population-based studies. In 2010-2011, the GSP undertook eight pilot and 14 collaborative research projects using the set up platforms in collaboration with GEP, MOC, EGE, PAT, ICB and RAD.

In parallel, GSP has worked closely with ITS and MOC to develop a Sample Management System for IARC biobank to allow the efficient handling of the increasing number of biological samples hosted at IARC (Voegele *et al.*, 2010).

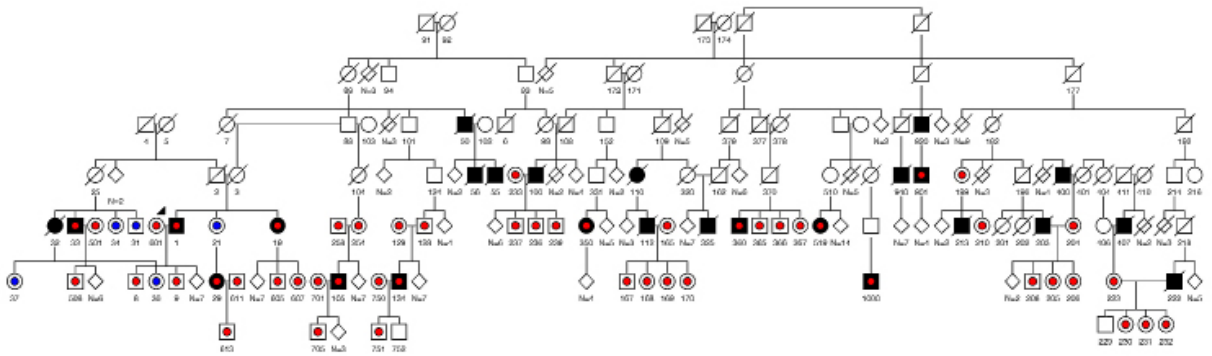
During 2011, the GSP also installed next generation sequencing (NGS) technology at IARC. A Life technologies 5500xl SOLiD next generation sequencer and a medium capacity Ion-Torrent personal genome machine have been installed and the associated workflows are currently being implemented under the umbrella of the GCS LIMS. We are now piloting a variety of NGS applications (particularly exome resequencing, RNAseq and methylation-related sequencing) with IARC scientists and technical staff. We have put in place a high-performance computing cluster and relevant software to enable analysis of large quantities of data produced by a NGS platform, as well as its long-term storage and backup. GCS bioinformaticians are developing bioinformatic workflows for the analysis of NGS data.

Projects from 2012-2013

Over the next few years, GCS will apply NGS to our projects. We will remain focused on describing genetic susceptibility, but we will also endeavor to use knowledge from the somatic events occurring during tumorigenesis, assayed using NGS and other genomics-based techniques, to inform our analysis of the germline events.

Genetic susceptibility to nasopharyngeal carcinoma

Isolated populations offer rare opportunities to investigate the genetic cause of human disease. We have identified one exceptionally large multiplex pedigree from the Bidayuh population where the incidence of nasopharyngeal carcinoma (NPC) is one of the highest in the world. Blood samples have been collected for 11 of the 26 NPC patients in this pedigree. We will use exome DNA sequencing to conduct a comprehensive genetic analysis to identify potential genes segregating in this pedigree. Through our collaboration with Dr Allan Hildesheim of the Division of Cancer Epidemiology and Genetics at NCI, who is conducting parallel efforts in Taiwanese NPC pedigrees, and further



Large extended Bidayuh pedigree from Sarawak, Borneo, Malaysia. Red dots indicate where blood samples have been collected

study in additional NPC patients at IARC, we will compare, contrast and cross validate results from this large pedigree.

Investigation of the Knudson's two-hit model in tumorigenesis

Many tumours resulting from a hereditary predisposition under Knudson's 'two-hit' genetic model contain one inherited mutant allele and the remaining allele is mutated somatically during tumorigenesis. The co-occurrence of germline and somatic mutations in the same gene in a given individual is a relatively rare event by chance alone. The observation of only a few dual mutation events will be sufficient to pinpoint a gene as noteworthy. We intend to perform whole exome sequencing of both the normal and tumour material of lung cancer patients with a positive family history, and use this data to identify genes acting along the lines of the two-hit model. Noteworthy genes can then be further investigated in broader/larger case-control studies to validate and further explore significant findings.

Investigation of common and rare genetic variants in melanoma susceptibility genes

Malignant melanoma is a rare tumour of melanocytes that, because of its aggressive nature, causes the majority of deaths related to skin cancer. The objective of this study is to identify new melanoma susceptibility genes and the characterization of the pathogenic sequence variants associated with increased risk of developing melanoma. We have set up a large-scale case-control mutation screening study nested in the EPIC cohort. Also in collaboration with Dr Françoise Clavel-Chapelon's group (INSERM U1018, Villejuif, France), we are investigating the relationship between genetic factors, pigmentary phenotype (sun exposure) and risk of cutaneous malignant melanoma (CMM) in the E3N prospective cohort. Results emerging from the association study in this cohort will complement findings from the case-control mutation screening study in EPIC, where dysfunctional variants will be sought in genes of the same pathways (pigmentation, nevi development).

GCS is grateful to the following persons for their collaboration:

Georgia Chenevix-Trench, Terry Dwyer, Parkville, Melissa Southey, Melbourne, Amanda Spurdle, Herston, Australia; Maria Isabel Waddington Achatz, Edenir Inez Palmero, São Paulo, Brazil; Irène Andrulis, Toronto, William Foulkes, Jacques Simard, Quebec, Canada; Françoise Clavel-Chapelon, Florent de Vathaire, Villejuif, David Cox, Charles Dumontet, Lars Jordheim, Sylvie Mazoyer, Olga Sinilnikova, Gilles Thomas, Lyon, Janet Hall, Mark Lathrop, Paris, France; Beena Devi, Kuching, Sarawak, Suleeporn Sangrajrang, Bangkok, Thailand; Doug Easton, Cambridge, United Kingdom; Fergus Couch, Rochester, MN, David Goldgar, Salt Lake City, UT, Esther John, Fremont, CA, Sean Tavtigian, Alan Thomas, Salt Lake City, UT, USA and the following Consortia: European Prospective Investigation into Cancer and Nutrition (EPIC), EPIC Skin Cancer Working Group, Melanoma Genetics (GenoMel), International Lung Cancer Consortium (ILCCO), International Head and Neck Cancer Epidemiology (INHANCE)

Financial support from the following bodies is gratefully acknowledged:

Electricité de France
Huntsman Cancer Institute, Salt Lake City, USA
Institut National du Cancer, Paris, France
Mayo Clinic, Rochester, MN, USA
National Institutes of Health, USA
Université Laval, Québec, Canada

Publications

Braschinsky M, Tamm R, Beetz C *et al.* (2010). Unique spectrum of SPAST variants in Estonian HSP patients: presence of benign missense changes but lack of exonic rearrangements. *BMC Neurol,* 10:17. doi:10.1186/1471-2377-10-17 PMID:20214791

Campa D, Hüsing A, McKay JD *et al.* (2010). The INSIG2 rs7566605 polymorphism is not associated with body mass index and breast cancer risk. *BMC Cancer,* 10:563.doi:10.1186/1471-2407-10-563 PMID:20955599

Carvalho LV, Pereira EM, Frappart L *et al.* (2010). Molecular characterization of breast cancer in young Brazilian women. *Rev Assoc Med Bras,* 56:278–287.doi:10.1590/S0104-42302010000300010 PMID:20676533

Chen D, McKay JD, Clifford G *et al.* (2011). Genome-wide association study of HPV seropositivity. *Hum Mol Genet*.doi:10.1093/hmg/ddr383 PMID:21896673

Chen D, Truong T, Gaborieau V *et al.* (2011). A sex-specific association between a 15q25 variant and upper aerodigestive tract cancers. *Cancer Epidemiol Biomarkers Prev,* 20:658–664.doi:10.1158/1055-9965.EPI-10-1008 PMID:21335511

Dumontet C, Landi S, Reiman T *et al.* (2010). Genetic polymorphisms associated with outcome in multiple myeloma patients receiving high-dose melphalan. *Bone Marrow Transplant,* 45:1316–1324.doi:10.1038/bmt.2009.335 PMID:19966851

Figueroa JD, Garcia-Closas M, Humphreys M *et al.*; The GENICA Network; for the kConFab AOCS Management Group (2011). Associations of common variants at 1p11.2 and 14q24.1 (RAD51L1) with breast cancer risk and heterogeneity by tumor subtype: findings from the Breast Cancer Association Consortium. *Hum Mol Genet*. doi:10.1093/hmg/ddr368 PMID:21852249

Galy O, Chemin I, Le Roux E *et al.* (2011). Mutations in TP53 and CTNNB1 in Relation to Hepatitis B and C Infections in Hepatocellular Carcinomas from Thailand. *Hepat Res Treat,* 2011: 1–9.doi:10.1155/2011/697162 PMID:21760996

Garritano S, Gemignani F, Palmero EI *et al.* (2010). Detailed haplotype analysis at the TP53 locus in p.R337H mutation carriers in the population of Southern Brazil: evidence for a founder effect. *Hum Mutat,* 31:143–150.doi:10.1002/humu.21151 PMID:19877175

Hernandez-Vargas H, Lambert MP, Le Calvez-Kelm F *et al.* (2010). Hepatocellular carcinoma displays distinct DNA methylation signatures with potential as clinical predictors. *PLoS One,* 5:e9749. doi:10.1371/journal.pone.0009749 PMID:20305825

Hernández-Vargas H, Ouzounova M, Le Calvez-Kelm F *et al.* (2011). Methylome analysis reveals Jak-STAT pathway deregulation in putative breast cancer stem cells. *Epigenetics,* 6:428–439. doi:10.4161/epi.6.4.14515 PMID:21266853

Le Calvez-Kelm F, Lesueur F, Damiola F *et al.*; Breast Cancer Family Registry (2011). Rare, evolutionarily unlikely missense substitutions in CHEK2 contribute to breast cancer susceptibility: results from a breast cancer family registry case-control mutation-screening study. *Breast Cancer Res*, 13:R6.doi:10.1186/bcr2810 PMID:21244692

McKay JD, Truong T, Gaborieau V *et al.* (2011). A genome-wide association study of upper aerodigestive tract cancers conducted within the INHANCE consortium. *PLoS Genet,* 7:e1001333. doi:10.1371/journal.pgen.1001333 PMID:21437268

Milne RL, Goode EL, García-Closas M *et al.*; for the GENICA Network; for the kConFab Investigators and the AOCS Group (2011). Confirmation of 5p12 as a susceptibility locus for progesterone-receptor-positive, lower grade breast cancer. *Cancer Epidemiol Biomarkers Prev*, 20:2222–2231. PMID:21795498

Mouchawar J, Korch C, Byers T *et al.* (2010). Population-based estimate of the contribution of TP53 mutations to subgroups of early-onset breast cancer: Australian Breast Cancer Family Study. *Cancer Res,* 70:4795–4800.doi:10.1158/0008-5472.CAN-09-0851 PMID:20501846

Nguyen-Dumont T, Jordheim LP, Michelon J *et al.* (2011). Detecting differential allelic expression using high-resolution melting curve analysis: validation on the breast cancer susceptibility gene CHEK2. *BMC Med Genomics,* 4:4. PMID:21223598

Purdue MP, Johansson M, Zelenika D *et al.* (2011). Genome-wide association study of renal cell carcinoma identifies two susceptibility loci on 2p21 and 11q13.3. *Nat Genet,* 43:60–65.doi:10.1038/ng.723 PMID:21131975

Rio Frio T, Bahubeshi A, Kanellopoulou C *et al.* (2011). DICER1 mutations in familial multinodular goiter with and without ovarian Sertoli-Leydig cell tumors. *JAMA*, 305:68–77.doi:10.1001/jama.2010.1910 PMID:21205968

Royce SG, Alsop K, Haydon A *et al.* (2010). The role of SMAD4 in early-onset colorectal cancer. *Colorectal Dis,* 12:213–219.doi:10.1111/j.1463-1318.2009.01779.x PMID:19183329

Szymańska K, Moore LE, Rothman N *et al.* (2010). TP53, EGFR, and KRAS mutations in relation to VHL inactivation and lifestyle risk factors in renal-cell carcinoma from central and eastern Europe. *Cancer Lett,* 293:92–98.doi:10.1016/j.canlet.2009.11.024 PMID:20137853

Timofeeva MN, McKay JD, Smith GD *et al.* (2011). Genetic Polymorphisms in 15q25 and 19q13 Loci, Cotinine Levels, and Risk of Lung Cancer in EPIC. *Cancer Epidemiol Biomarkers Prev.* PMID:21862624

Truong T, Hung RJ, Amos CI *et al.* (2010). Replication of lung cancer susceptibility loci at chromosomes 15q25, 5p15, and 6p21: a pooled analysis from the International Lung Cancer Consortium. *J Natl Cancer Inst,* 102:959–971.doi:10.1093/jnci/djq178 PMID:20548021

Truong T, Sauter W, McKay JD *et al.*; EPIC-lung (2010). International Lung Cancer Consortium: coordinated association study of 10 potential lung cancer susceptibility variants. *Carcinogenesis*, 31:625–633. doi:10.1093/carcin/bgq001 PMID:20106900

Velkova A, Carvalho MA, Johnson JO *et al.* (2010). Identification of Filamin A as a BRCA1-interacting protein required for efficient DNA repair. *Cell Cycle,* 9:1421–1433.doi:10.4161/cc.9.7.11256 PMID:20305393

Voegele C, Alteyrac L, Caboux E *et al.* (2010). A sample storage management system for biobanks. *Bioinformatics*, 26:2798–2800.doi:10.1093/bioinformatics/btq502 PMID:20807837

Wei S, Niu J, Zhao H *et al.* (2011). Association of a novel functional promoter variant (rs2075533 C>T) in the apoptosis gene TNFSF8 with risk of lung cancer–a finding from Texas lung cancer genome-wide association study. *Carcinogenesis*, 32:507–515.doi:10.1093/carcin/bgr014 PMID:21292647

Yang XR, Chang-Claude J, Goode EL *et al.* (2011). Associations of breast cancer risk factors with tumor subtypes: a pooled analysis from the Breast Cancer Association Consortium studies. *J Natl Cancer Inst,* 103:250–263.doi:10.1093/jnci/djq526 PMID:21191117

Yu H, Zhao H, Wang LE *et al.* (2011). An analysis of single nucleotide polymorphisms of 125 DNA repair genes in the Texas genome-wide association study of lung cancer with a replication for the XRCC4 SNPs. *DNA Repair (Amst),* 10:398–407.doi:10.1016/j.dnarep.2011.01.005 PMID:21296624

Biostatistics Group (BST)

Group head
Dr Graham Byrnes

Secretary
Ms Yvette Granjard

Students
Ms Amandine Coquillat
(until July 2010)
Ms Sally Hinchcliffe
(until April 2010)
Ms Marie Paturel (until July 2011)

Visiting scientist
Professor Nanny Wermuth

After moving to GEN in 2010, the Biostatistics Group (BST) has increased collaborations with the other GEN groups, while maintaining former collaborations and cross-agency activities. In particular, BST is responsible for the training and development of IARC's statistical personnel located in the various sections.

In silico functional classification

BST continues to collaborate with GCS on *in silico* functional classification, which has taken on new directions with the advent of next generation sequencing (NGS). While the now traditional genome-wide association studies (GWAS) have been successful in identifying many genetic features which predispose to various cancers, they have small individual effects and jointly explain only a small part of the familial clustering of all common cancers. It is possible that many of the remaining genetic variants are individually very rare, but occur in such variety that together they explain an important fraction of the heredity of cancer. Moreover, since the effects of individual variants are greater, they are more likely to provide insight into mechanisms of oncogenesis. However, their individual rarity precludes the use of standard genetic epidemiological techniques. It is necessary to first use other evidence, such as evolutionary variability between species, to classify these variants into categories of possibly similar effect. These categories can then reach population frequencies that allow them to be investigated by standard epidemiological methods. This approach has helped identify the importance of missense mutations in several genes for predicting the risk of breast cancer (Le Calvez-Kelm *et al.*, 2011; Southey *et al.*, 2010).

This project has two facets: supporting the adaptation of previously developed tools (AlignGVGD) to study the genetics of melanoma (with Dr Fabienne Lesueur); and investigating methods better adapted to high-throughput data generated by NGS.

Omics integration

Two key reasons BST moved to NGS are because of the extremely large number of genetic variables investigated (potentially every element of the genome), and the ability to obtain several complementary types of data (i.e. germline genotype, tissue genotype, tissue-specific RNA expression and high-density methylation data). The number of measured variables makes traditional case-control comparisons impractical because of the need for multiple comparisons correction. However, the variety of data types creates the possibility of novel approaches, such as using tissue-specific expression to localize areas of the genome likely to be active in that tissue, then examining these areas for GWAS.

Another option is the comparison of genomic and somatic modification. BST is collaborating with Dr James McKay to develop the statistical methodology to use matched comparisons within individuals to identify regions showing modification in both tumour and germline DNA, as would be expected under Knudson's two-hit hypothesis.

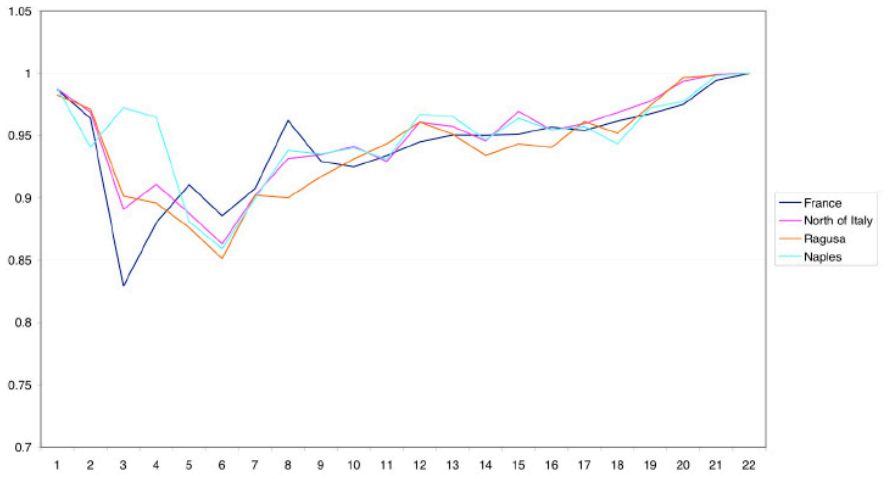


Figure 1. Cumulative explanation of variation in dietary patterns in EPIC by principal components, according to centre of recruitment

General biostatistics support

A common feature of most epidemiological research is that exposure data is approximate. This applies to questionnaires, self-reported smoking behaviour, estimated radiation exposures and imputed genotypes. In all cases, incorporating these noisy data into conventional regression analyses can result in biased estimates and inappropriate inferences. Hence, a core methodological interest is the development, evaluation and use of methods such as Markov Chain maximum likelihood and regression calibration. This underpins continued collaborations with the ENV, as well as within GEN (Timofeeva *et al.*, 2011).

BST also provides guidance on the use of traditional biostatistical tools, both within IARC and in outside collaborations (Hery *et al.*, 2010; Burgess *et al.*, 2011)

Statistical education and training

IARC has adopted a model of disseminated statistical support by dispersing individuals with differing levels of statistical training among the various sections. Encouraging development and sharing of statistical skills is therefore one of the responsibilities of BST.

Training is carried out through BST with the largest block of teaching taking place within the Introduction to Epidemiology Summer School. This course was also repeated in 2010 for IARC staff unable to attend as part of the Summer School and it was fully subscribed. In 2011, BST facilitated a course on the use of the statistical package R in epidemiology (coordinated by Dr Martyn Plummer, INF). Further courses and a seminar series are under development.

In addition, BST funding has been used to provide support for IARC staff to attend meetings specific to statistical methodology and to bring senior statisticians for short visits to give seminars and otherwise discuss advanced methodologies. Visitors in 2011 included Prof Per Kragh Andersen

from the University of Copenhagen and Dr Frank Dudbridge from the London School of Hygiene and Tropical Medicine. Other visits are planned and BST will host Prof Nanny Wermuth, a distinguished statistician, as a Senior Visiting Scientist in 2011-2012.

Priorities for 2012-13

Over the next two years, BST will continue the development and evaluation of methodological approaches both for genetic epidemiology, using the variety of data types available from IARC's recently acquired NGS capabilities, and for more general statistical issues in epidemiology. In parallel to this, it will be crucial to establish collaborations with external statisticians to complement the resources at IARC. Such collaborations will contribute to the development of a culture of cross-institutional cooperation on statistical work.

Publications

Büchner FL, Bueno-de-Mesquita HB, Linseisen J *et al.* (2010). Fruits and vegetables consumption and the risk of histological subtypes of lung cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC). *Cancer Causes Control,* 21:357–371.doi:10.1007/s10552-009-9468-y PMID:19924549

Burgess JA, Matheson MC, Gurrin LC *et al.* (2011). Factors influencing asthma remission: a longitudinal study from childhood to middle age. *Thorax*, 66:508–513.doi:10.1136/thx.2010.146845 PMID:21450787

Chen D, Truong T, Gaborieau V *et al.* (2011). A sex-specific association between a 15q25 variant and upper aerodigestive tract cancers. *Cancer Epidemiol Biomarkers Prev,* 20:658–664.doi:10.1158/1055-9965.EPI-10-1008 PMID:21335511

Gandini S, Boniol M, Haukka J *et al.* (2011). Meta-analysis of observational studies of serum 25-hydroxyvitamin D levels and colorectal, breast and prostate cancer and colorectal adenoma. *Int J Cancer,* 128:1414–1424.doi:10.1002/ijc.25439 PMID:20473927

Héry C, Tryggvadóttir L, Sigurdsson T *et al.* (2010). A melanoma epidemic in Iceland: possible influence of sunbed use. *Am J Epidemiol,* 172:762–767. doi:10.1093/aje/kwq238 PMID:20813801

Jenab M, Bueno-de-Mesquita HB, Ferrari P *et al.* (2010). Association between pre-diagnostic circulating vitamin D concentration and risk of colorectal cancer in European populations: a nested case-control study. *BMJ*, doi:10.1136/bmj.b5500 PMID:20093284

Le Calvez-Kelm F, Lesueur F, Damiola F *et al.*; Breast Cancer Family Registry (2011). Rare, evolutionarily unlikely missense substitutions in CHEK2 contribute to breast cancer susceptibility: results from a breast cancer family registry case-control mutation-screening study. *Breast Cancer Res*, 13:R6.doi:10.1186/bcr2810 PMID:21244692

McKay JD, Truong T, Gaborieau V *et al.* (2011). A genome-wide association study of upper aerodigestive tract cancers conducted within the INHANCE consortium. *PLoS Genet,* 7:e1001333. doi:10.1371/journal.pgen.1001333 PMID:21437268

Odefrey F, Stone J, Gurrin LC *et al.*; Australian Twins and Sisters Mammographic Density Study (2010). Common genetic variants associated with breast cancer and mammographic density measures that predict disease. *Cancer Res,* 70:1449–1458.doi:10.1158/0008-5472.CAN-09-3495 PMID:20145138

Olsson CA, Foley DL, Parkinson-Bates M *et al.* (2010). Prospects for epigenetic research within cohort studies of psychological disorder: a pilot investigation of a peripheral cell marker of epigenetic risk for depression. *Biol Psychol,* 83:159–165.doi:10.1016/j.biopsycho.2009.12.003 PMID:20018225

Russell FM, Carapetis JR, Balloch A *et al.* (2010). Hyporesponsiveness to re-challenge dose following pneumococcal polysaccharide vaccine at 12 months of age, a randomized controlled trial. *Vaccine*, 28:3341–3349.doi:10.1016/j.vaccine.2010.02.087 PMID:20206670

Sighoko D, Bah E, Haukka J *et al.* (2010). Population-based breast (female) and cervix cancer rates in the Gambia: evidence of ethnicity-related variations. *Int J Cancer,* 127:2248–2256.doi:10.1002/ijc.25244 PMID:20162609

Southey MC, Ramus SJ, Dowty JG *et al.* (2011). Morphological predictors of BRCA1 germline mutations in young women with breast cancer. *Br J Cancer,* 104:903–909.doi:10.1038/bjc.2011.41 PMID:21343941

Southey MC, Teo ZL, Dowty JG *et al.*; kConFab for the Beast Cancer Family Registry (2010). A PALB2 mutation associated with high risk of breast cancer. *Breast Cancer Res,* 12:R109.doi:10.1186/bcr2796 PMID:21182766

Stone J, Gurrin LC, Hayes VM *et al.* (2010). Sibship analysis of associations between SNP haplotypes and a continuous trait with application to mammographic density. *Genet Epidemiol,* 34:309–318.doi:10.1002/gepi.20462 PMID:19918759

Timofeeva MN, McKay JD, Smith GD *et al.* (2011). Genetic Polymorphisms in 15q25 and 19q13 Loci, Cotinine Levels, and Risk of Lung Cancer in EPIC. *Cancer Epidemiol Biomarkers Prev.* PMID:21862624

Tsilidis KK, Allen NE, Key TJ *et al.* (2010). Oral contraceptives, reproductive history and risk of colorectal cancer in the European Prospective Investigation into Cancer and Nutrition. *Br J Cancer,* 103:1755–1759.doi:10.1038/sj.bjc.6605965 PMID:21045829

Tsilidis KK, Allen NE, Key TJ *et al.* (2011). Menopausal hormone therapy and risk of colorectal cancer in the European Prospective Investigation into Cancer and Nutrition. *Int J Cancer,* 128:1881–1889.doi:10.1002/ijc.25504 PMID:20533550