Section of Genetics (GEN)

Section head
Dr Paul Brennan

Genetic Epidemiology Group (GEP)

Group head
Dr Paul Brennan

Scientists
Dr Devasena Anantharaman (until July 2016)
Dr Estelle Chanudet-van den Brink
Dr Mattias Johansson
Dr Ghislaine Scélo

Technical assistants
Ms Valérie Gaborieau
Ms Hélène Renard

Laboratory technician
Ms Priscilia Chopard

Project assistant
Ms Laurène Bouvard

Secretariat
Ms Charlotte Volatier

Visiting scientists
Dr Dana Hashim (until February 2017)
Dr Hooman Khademi Kohnehshahri (until September 2016)
Dr Peng Li (until August 2016)
Dr Brent Richards

Postdoctoral fellows
Dr Renata Abrahão
Dr Alice Billot-Grasset (until March 2016)
Dr Robert Carreras Torres
Dr Szilvia Ecsedi (until December 2017)
Dr Anouar Fanidi (until April 2016)
Dr Aida Ferreiro-Iglesias
Dr Florence Guida
Dr Tricia Larose
Dr Ruhina S. Laskar
Dr Corina Lesseur Perez (until December 2016)
Dr Dariush Nasrollahzadeh Nesheli
Dr Sandra Perdomo Velasquez (until February 2016)
Dr Carolina Santamaria Ulloa (until March 2016)
Dr Mahdi Sheikh
Dr Karl Smith Byrne
Dr Chanida Vinayanuwattikun (until May 2016)

Master's student
Ms Sandrine Magat (until August 2017)

Trainee
Ms Linda Kachuri (until September 2016)

Genetic Cancer Susceptibility Group (GCS)

Group head
Dr James McKay

Scientists
Dr Behnoush Abedi-Ardekani
Dr Lynnette Fernandez-Cuesta
Dr Matthieu Foll
Dr Florence Le Calvez-Kelm
Dr Maria Zvereva (until August 2017)

Laboratory technicians
Ms Amélie Chabrier
Mr Geoffroy Durand
Ms Nathalie Forey

Bioinformatician
Ms Catherine Voegele

Secretariat
Ms Isabelle Rondy
Ms Andreea Spanu (until July 2017)

Postdoctoral fellows
Dr Nicolas Alcala
Dr Patrice Avogbe
Dr Md Ismail Hosen
Dr Rim Khlifi
Dr Dariush Nasrollahzadeh Nesheli

Students
Mr Salia Bamba (until September 2017)
Ms Tiffany Delhomme
Mr Jules Derks (until December 2016)
Ms Pauline François (until September 2016)
Ms Aurélie Gabriel
Mr Théo Giffon (until October 2017)
Ms Imen Hemissi
Ms Noemie Leblay
Ms Olesia Lole (until May 2017)
Mr Gabriel Roberti de Oliveira (until February 2017)

The Section of Genetics (GEN) includes the Genetic Epidemiology Group (GEP) and the Genetic Cancer Susceptibility Group (GCS). The work of the Section combines large population-based studies with laboratory and bioinformatics expertise to identify specific genes and genetic profiles that contribute to the development of cancer and elucidate how they exert their effect along with environmental factors. GEN also tries to identify individuals who are at high enough risk that they are likely to benefit from potential screening strategies.

The Section's projects usually involve extensive fieldwork in collaboration with external investigators in order to develop large-scale epidemiological studies with appropriate clinical and exposure data, as well as biosample collection. This typically occurs within GEP. Genetic analysis comprises either candidate gene or genome-wide genotyping studies, as well as extensive sequencing work. GEP studies also assess non-genetic exposures, partly in recognition of the importance of non-genetic factors in driving cancer incidence, and also to facilitate accurate assessment of gene–environment interactions. In contrast, GCS places more focus on identification of uncommon or rare genetic variants that may have a larger effect than common single nucleotide polymorphisms but that are not sufficiently frequent to be captured by current genome-wide association genotyping arrays. The approach of GCS has been to use genomic and bioinformatic techniques to complement more traditional approaches for the study of rare genetic variants. GCS also uses genomics to explore how the variants may be conferring genetic susceptibility to cancer. Thus, the research programme of GCS complements that of GEP, and also provides a facility for high-throughput genomic techniques and the related bioinformatics to support GEN's large-scale molecular epidemiology projects and other IARC genomics projects.

Genetic Epidemiology Group (GEP)

The overall goal for the Genetic Epidemiology Group (GEP) is to contribute to understanding the causes of cancer through the study of genetic susceptibility variants of various cancer sites, and also patterns of genetic mutations that are observed in tumours. An additional goal is to develop accurate risk prediction models that take into account both demographic information (e.g. age and sex) and biomarkers (genetic and non-genetic). The work of GEP includes studies of cancers related to tobacco use and alcohol consumption (lung and aerodigestive tract cancers) and cancers related to obesity (such as kidney, pancreatic, and colorectal cancers). GEP devotes substantial resources to extensive fieldwork, with the goal of recruiting large series of cases and controls, comprising extensive questionnaire information and biological samples. Genetic analyses of inherited susceptibility usually comprise a genome-wide approach initially, with subsequent large-scale coordinated replication studies in diverse populations. This latter aspect is aided 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, which are often conducted in collaboration with other IARC Groups. Analysis of these large genome-wide studies also includes a Mendelian randomization approach that aims to understand how lifestyle factors influence cancer onset.

GEP is also undertaking a large international study of the causes of cancer by analysis of mutation patterns (or mutation signatures) in cancer genomes. Most of the Group's efforts in this domain are included in the Mutographs project, which aims to understand the causes of five different cancers across five continents (see text box).

In addition to studies of genetic factors, GEP is conducting a wide range of studies involving non-genetic factors, including evaluations of circulating biomarkers such as human papillomavirus (HPV) antibodies for head and neck cancers, and a wide range of protein and other biomarkers for lung cancer. The overall goal of these studies is to identify individuals at sufficiently high risk to justify screening and early detection.

Some prominent examples of the Group's work over the 2016–2017 biennium are described here.

Elucidating the etiological role of obesity and related risk factors in multiple cancers – a Mendelian randomization approach

Elevated body mass index (BMI) and obesity-related risk factors have been associated with multiple cancers studied by GEP. Because these risk factors are inherently interrelated, traditional epidemiological studies have not been able to untangle which specific factors exert a causal influence and which are merely correlated with the underlying causal factor.

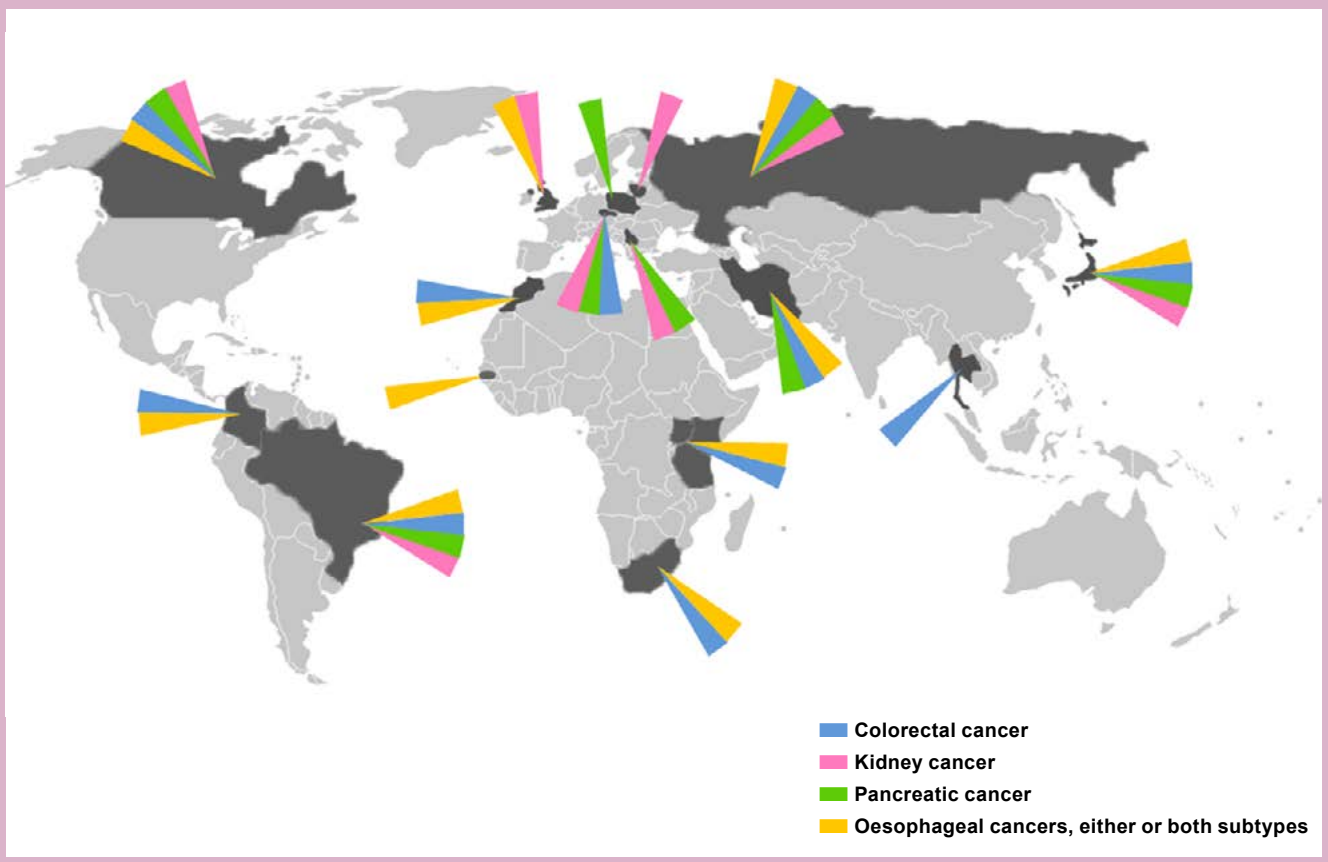
By leveraging data from genome-wide association studies of tens of thousands of cancer cases and controls that GEP has led or contributed to, we have conducted a series of studies where we have interrogated the causal relevance for several obesity-related risk factors for various cancers. Because these analyses were based on genetic instruments, they are not influenced by reverse causation and are less sensitive

Mutographs of cancer

A new and major initiative of the Section is an effort to understand the causes of cancer by generating mutation signature profiles based on whole-genome sequence data. The study results from a major Cancer Research UK (CRUK) Grand Challenge grant – one of the world's most ambitious cancer research awards – and is co-led by Dr Paul Brennan together with overall principal investigator (PI) Professor Sir Mike Stratton from the Sanger Institute (Cambridge, United Kingdom) and four other co-PIs. The overall name of the project is Understanding of the Causes of Cancer through Studies of Mutational Signatures – Mutographs.

Different patterns of somatic mutation are generated by the different environmental, lifestyle, and genetic factors that cause cancer; many of them are still unknown. Within the Mutographs project, GEP is coordinating the recruitment of 5000 individuals with cancer (colorectal cancer, kidney cancer, pancreatic cancer, oesophageal adenocarcinoma, or oesophageal squamous cancer) across five continents to explore whether different mutational signatures explain the marked variation in incidence. Through an international network of collaborators, biological materials are collected, along with demographic, histological, clinical, and questionnaire data. Whole-genome sequences of tumour–germline DNA pairs are generated at the Sanger Institute. Extracted somatic mutational signatures are then correlated with data on risk factors.

Through an enhanced understanding of cancer etiology, the unprecedented effort within the Mutographs project is anticipated to outline modifiable risk factors, lead to new approaches to prevent cancer, and provide opportunities to empower early detection, refine high-risk groups, and contribute to therapeutic development.



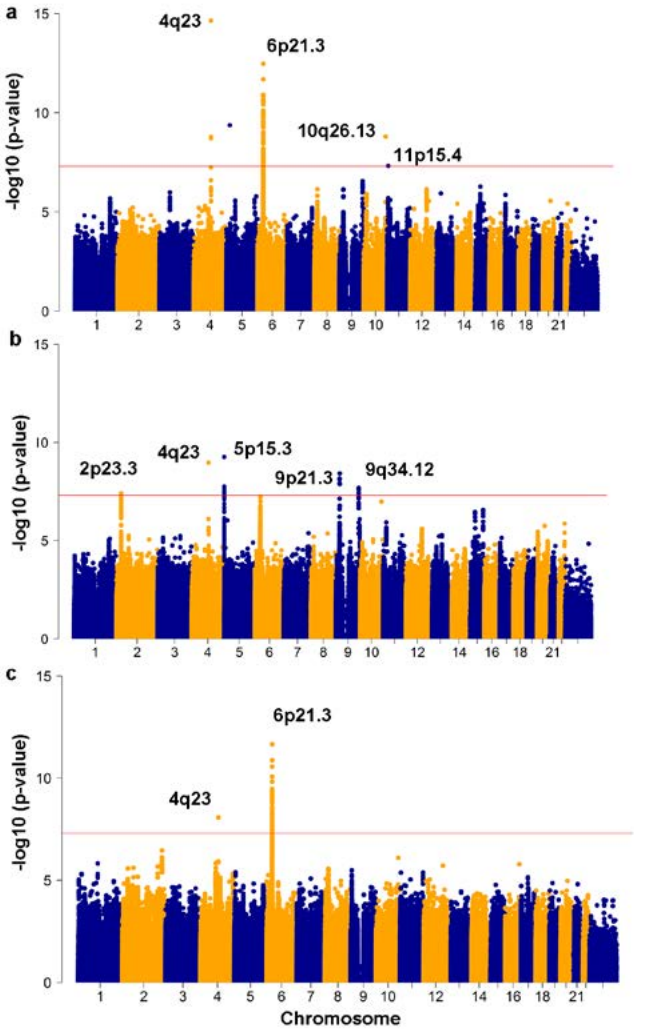
Countries contributing to Mutographs recruitment.

to confounding than those using direct exposure measures. The results were illuminating for both kidney cancer and pancreatic cancer, not only because they confirmed elevated BMI as an important cause of these cancers, but in particular because they highlighted insulin as an important mediator of the risk increase caused by higher BMI. These findings significantly improved the understanding of the importance of obesity in the etiology of kidney cancer and pancreatic cancer. They have also highlighted a potentially important role for obesity and insulin resistance in lung cancer (Carreras-Torres et al., 2017a, 2017b).

Genome-wide analysis of tobacco-related cancers

GEP has coordinated a large OncoArray analysis of more than 7000 cancers of the oral cavity or oropharynx, along with a similar number of controls. A prominent finding from this study is the important role of the HLA region for oropharyngeal cancer. Further analysis of this locus identified a HLA haplotype that was also associated with cervical cancer, suggesting an important interaction with this specific haplotype and papillomavirus (Lesseur et al., 2016) (Figure 1). GEP also led a large genome-wide association analysis of more than 10 000 renal cancer cases and 20 000 controls, in partnership with the United States National Cancer Institute, and identified an additional seven susceptibility loci for renal cancer in addition to six that had been previously discovered (Scelo et al., 2017).

Figure 1. Genome-wide association meta-analysis results. Red lines correspond to $P = 5 \times 10^{-8}$. The vertical axes show $-\log_{10}(P\text{-values})$. (a) Overall oral cavity and pharyngeal cancer analysis with 6034 cases and 6585 controls. (b) Oral cavity cancer analysis with 2990 cases and 6585 controls. (c) Oropharyngeal cancer analysis with 2641 cases and 6585 controls. Loci with technically validated genome-wide significant single nucleotide polymorphisms are tagged by genomic location. Reprinted from Lesseur et al. (2016) by permission from Macmillan Publishers Ltd, copyright 2016.



Genetic Cancer Susceptibility Group (GCS)

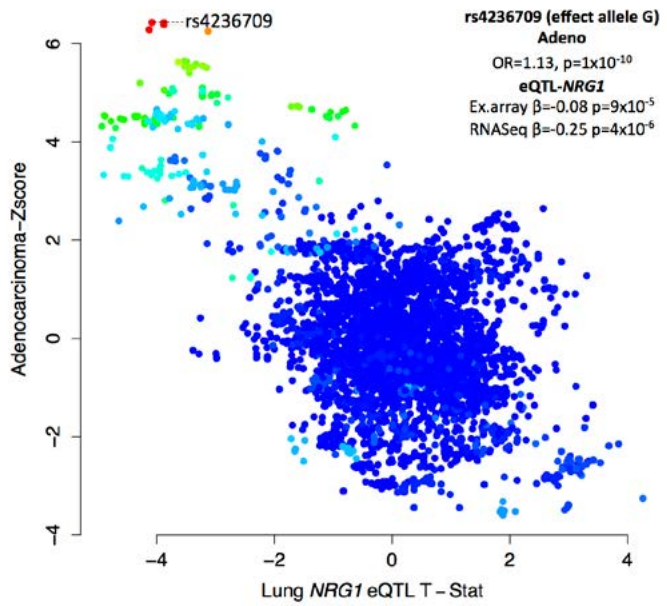
The Genetic Cancer Susceptibility Group (GCS) contains a multidisciplinary scientific team, covering genetics, genomics, bioinformatics, and pathology. These combined skills are used to undertake genetic and genomic research to identify cancer-related genes and explore their mechanisms of action. Through this knowledge, the aim is to gain insights into cancer etiology and apply that to early cancer detection and prevention.

In the context of germline genetics, working within the International Lung Cancer Case–Control Consortium (ILCCO) and the United States National Cancer Institute Genetic Associations and Mechanisms in Oncology (GAME-ON) OncoArray consortium, GCS undertook a genome-wide association study (GWAS) of lung cancer that included nearly 30 000 lung cancer patients and 57 000 controls. This GWAS identified 18 susceptibility loci, including 10 novel loci. These susceptibility alleles have been explored by integrating additional genetic data from more than 250 000 people, including from studies of gene expression in the lung and other tissues,

measures of smoking propensity, lung spirometry (forced vital capacity [FVC]/ forced expiration volume [FEV]), and leukocyte telomere length (McKay et al., 2017a). The analysis of lung epithelial tissue gene expression implicated genes not previously implicated in lung cancer etiology, such as *RNASET2* or *SECISBP2L*, and genes like *NRG1*, a gene infrequently somatically translocated in lung adenocarcinomas (Figure 2). In addition, we identified lung cancer susceptibility variants at 8p12 that influence expression levels of *CHRNA2*, a cholinergic nicotinic receptor. In contrast with our previous observation with genetic variants in cholinergic nicotinic receptor genes and lung cancer, these 8p12 variants were not associated with the amount smoked but with factors such as age at smoking initiation. These variants appeared to be linked with *CHRNA2* expression levels, particularly in the cerebellum, adding to the emerging evidence that this region of the brain may indeed play a role in aspects of addictive behaviour. In addition, GCS has been awarded funding to continue to explore the mechanisms of action of these genetic susceptibility variants (from France Genomique and the Institut national du Cancer, France) and has initiated projects to investigate genomic events in rare thoracic tumours, such as lung carcinoid tumours (supported by La Ligue nationale contre le Cancer Rhône-Alpes, France; the Dutch Cancer Society, The Netherlands; and the National Cancer Institute, USA) and mesothelioma (supported by the Institut national du Cancer, France).

GCS has also explored the potential of circulating tumour DNA (ctDNA) as a biomarker for cancer detection. The study of ctDNA presents important technical challenges; both the DNA quality and the allele fractions of tumour-derived genetic alterations in the total cell-free DNA (cfDNA) in plasma are lower than that usually acceptable for next-generation sequencing analysis. Furthermore, to be applicable in early detection, mutations must also be identified without the prior knowledge of the tumour genotype and across any region of the gene. We have combined our laboratory and bioinformatic skills to develop an analysis pipeline specifically tailored to ctDNA, called Needlestack (https://github.com/IARCbioinfo/needlestack; Figure 3). We applied this approach to retrospective case–control studies of lung and pancreatic cancer, demonstrating that ctDNA is found in cases and, importantly, in patients with early-stage disease (Fernandez-Cuesta et al., 2016; Le Calvez-Kelm et al., 2016). Whereas ctDNA is strongly over-represented in patients, tumour-related mutations were also consistently noted in an unexpected proportion of controls (~3–10%). This somewhat surprising observation in controls, with the limitations that it implies, highlights the insights that can be gained when these techniques are applied to molecular epidemiology-based studies at IARC. We are now exploring the application of these methods for ctDNA in additional settings, particularly bladder cancer (supported by the Association pour la Recherche en biologie moléculaire, France, and La Ligue nationale contre le Cancer Rhône-Alpes, France) and oesophageal cancer (supported by the National Institute for Medical Research Development, Islamic Republic of Iran).

Figure 2. Scatter plots comparing variants across the 5417 variants at the 8p12 susceptibility loci and their associated with lung adenocarcinoma (vertical axis) and the lung *cis* expression quantitative trait loci (eQTL) Genotype-Tissue Expression (GTEx) (horizontal axis). Each variant is coloured relative to the degree of linkage disequilibrium (R^2) with a sentinel lung cancer variant (rs4236709 marked) at that locus (red for high, blue for low). Inset table (top right): association between sentinel variant and lung adenocarcinoma as well as the eQTL evidence in lung epithelium in five cohorts, first RNA expression based on microarray and RNASeq technologies. The variants associated with lung adenocarcinoma tend to be those that are lung *cis*-eQTL for *NRG1*. Curiously, while somatic translocations are generally linked with ectopic NRG1 activation and never-smokers, the germline genetic risk correlated with decreased *NRG1* expression and was present in lung adenocarcinomas from ever- and never-smokers. © IARC.



GCS plays an active role in the development of genomics capabilities at IARC. GCS, with important contributions from other Groups, has continued to build links within the genomics community at IARC, through a role in the Bioinformatics Steering Committee and the related Bioinformatics Working Group, and through the Laboratory Steering Committee, as well as providing access to the laboratory techniques, pathology expertise, and computational resources for genomics-related activities at IARC. These developments are also made available to the scientific community via a GitHub site: https://github.com/IARCbioinfo/.

Figure 3. Two examples of variants called using Needlestack's regression model to detect rare allele fraction in outlying individuals (variant carriers). Each dot represents a sequenced individual (two dots per sample) coloured according to its phred-scaled *q*-value. The black regression line shows the estimated sequencing-error rate along with the 99% confidence interval (black dotted lines) containing samples. Coloured-dotted lines correspond to the limits of regions defined for different significance *q*-value thresholds. Both technical duplicates appear as outliers from the regression (in red), and are therefore classified as carrying the given mutation. Reprinted from Fernandez-Cuesta et al. (2016). Copyright 2016, with permission from Elsevier.