Section of Genetics (GEN)

Identifying specific genes and gene variants that contribute to the development of cancer is important for a number of reasons. These include understanding in greater depth the biological pathways that are involved in cancer, elucidating how environmental factors may exert their effects in combination with genes, and identifying individuals who are at high enough risk that they are likely to benefit from existing risk reduction strategies.

The Genetics Section comprises two Groups with the overall mission of identifying genes involved in cancer, characterising the spectrum of pathogenic sequence variants that they harbour, and understanding how they interact with non-genetic factors. These are the Genetic Epidemiology Group (GEP) and the Genetic Cancer Susceptibility Group (GCS). GEP is mainly involved in coordinating large population-based epidemiological studies and analysis of multiple common genetic variants in order to identify new susceptibility loci. Cancers of primary interest include those of the lung and upper aerodigestive tract (including the nasopharynx) as well as kidney cancer and rarer childhood cancers. GCS is mainly involved in identification of rare variants or mutations in known or strong candidate cancer loci that result in a substantial cancer risk. The main focus is on breast cancer, in particular basal-type breast tumours, with a growing interest in melanoma. Findings from the GCS Group may have direct prevention implications by resulting in more accurate analysis of clinical mutation screening data from high-risk susceptibility genes such as BRCA1, BRCA2, MLH1 and MSH2. GCS also provides a genotyping platform service for both Groups.
GENETIC EPIDEMIOLOGY GROUP (gep)

Group Head
Dr Paul Brennan

Scientists
Dr James McKay
Dr Ghislaine Scelo

Staff
Ms Amelie Chabrier
Ms Valerie Gaborieau
Ms Yvette Granjard
Ms Helene Renard

Visiting scientists
Dr A. Boroda (May 2009–June 2009)

Postdoctoral fellows
Dr D. Anantharaman (from January 2009)
Dr K. Braga Ribeiro (from October 2007)
Dr D. Chen
Dr J. Fachiroh
Dr M. Johansson (from September 2008)
Dr E. Lips (February 2008–January 2009)
Dr K. Szymanska (until April 2008)
Dr M. Timofeeva (from July 2009)
Dr T. Truong (from December 2007)
Dr K. Urayama

Students
M. Delahaye
Mr S. Oh (July 2008–October 2008)
J. Matyjasik (February 2008–July 2008)
M. Lener (February 2008–July 2008)
GEP is currently undertaking large multi-partner genetic epidemiology studies of cancers strongly related to tobacco and alcohol—principally lung and aerodigestive cancers, but also kidney cancers. These include candidate gene studies, and increasingly genomewide association studies.

A series of large multicentre case-control studies of lung, upper aerodigestive and kidney cancers has been completed in Europe and Latin America, comprising over 15,000 subjects. Genomewide association studies are currently underway in collaboration with the Centre National de Genotypage (Evry, France) to help identify new genes for these cancers, and the first results for lung cancer have been published (Hung et al., Nature 2008; McKay et al., Nature Genetics, 2008).

The Group is also working with the International Lung Cancer Consortium (ILCCO) and the International Head and Neck Cancer Epidemiology (INHANCE) Consortium, with the aim of pooling information and results from all large studies of lung and aero-digestive cancers.

A large genomewide study of kidney cancer is also underway in collaboration with the Centre National de Genotypage and the US National Cancer Institute. Complete results are expected before the end of 2009. Plans have also been developed in collaboration with the Centre International de Genotypage for a large-scale tumour sequencing project of kidney tumours (the CAGEKID project).

The 15q25 Lung cancer susceptibility locus identified by the IARC lung cancer genome-wide association study. This locus contains three nicotinic acetylcholine receptor genes, CHRNA5, CHRNA3 and CHRNA4. (a) P-values for SNPs genotyped in the 15q25 region (76.4-76.8 mB). The blue line indicates the threshold of p<5x10^-7 at which results were considered genome-wide significant. Points labeled with rs numbers have a p<1x10^-9. Points in red are genotyped in the 317K Illumina panel; points in blue indicate additional genotyped SNPs (Taqman). (b),(c) The high LD genomic region approximately delineated by rs4887053 (76.49 mB) and rs12594247 (76.73 mB) containing the SNPs strongly associated with lung cancer risk. (b) The positions of the 6 known genes. (c) The pairwise r2 estimates for the 46 common SNPs from 76.49 mB and 76.73 mB in controls of the central Europe IARC study, with increasing shades of grey indicating higher degree of r2 values. The majority of pairwise D' estimates for these SNPs exceed 0.8.
RUSSIAN COHORT STUDY

We are coordinating a large cohort study in Russia, along with colleagues in the Cancer Research Centre of Moscow and the Clinical Trials Service Unit of the University of Oxford. Over 200,000 adults have already been recruited from 3 cities in Western Siberia (Barnaul, Biysk and Tomsk) with collection of extensive questionnaire information and DNA. Follow-up is underway to identify cancer and other chronic disease outcomes, and future analyses will focus on understanding the causes of the extremely high mortality rates among adults in middle age in this region. Initial analysis of over 50,000 people from these regions who died of various causes has provided strong evidence that alcohol is the cause of more than half of all Russian deaths at ages 15–54, and accounts for most of the recent large fluctuations in Russian mortality (Zaridze et al., Lancet 2009).

NASOPHARYNGEAL CARCINOMA

Nasopharyngeal carcinoma (NPC) is a malignancy with a wide range of incidence rates across the world. In most areas, it is rare (e.g. 0.5 cases per 100,000 per year in the UK), but in certain regions it occurs in an endemic form with an incidence 10- to 40-fold higher. Endemic regions include the southern parts of China, other parts of Southeast Asia, and the Maghreb (Morocco, Algeria and Tunisia). Along with partners in Malaysia and Thailand, we are conducting studies on the role of genes and environmental factors in the etiology of NPC in Southeast Asia. This study aims to be one of the world’s largest studies of NPC with at least 1000 case-control pairs as well as multi-case families. Currently the study sites consist of nationwide efforts in Thailand coordinated by the National Cancer Institute in Bangkok, and in the Sarawak region of Malaysia coordinated by the Kuching General Hospital. Upon completion of recruitment, we aim to conduct genome-wide studies of NPC to investigate genes associated with onset and survival.

CANCER IN CHILDREN AND YOUNG ADULTS

We are helping to initiate pilot studies of non-central nervous system embryonal cancers that occur in childhood and young adulthood. Apart from most common cancers at these ages (leukemia and central nervous system tumours), there is a lack of large-scale etiological studies in all types of childhood cancers, and data on causes and mechanisms are very limited. With a large international study, we aim to investigate the role of exposure to suspected risk factors at different key periods (preconceptional, prenatal, and postnatal), genetic susceptibility factors and gene-environment interactions, as well as novel molecular markers (e.g. DNA methylation and repair capacity). The study will include retinoblastoma, Wilms’ tumour, rhabdomyosarcoma, neuroblastoma, and hepatoblastoma.
Cancers: A pooled analysis of 13 cancer registries. 


During the 2008 and 2009, the GCS Group has been active in four areas: analysis of unclassified variants in high-risk cancer susceptibility genes, case-control mutation screening of intermediate-risk breast cancer susceptibility genes, the genetics of melanoma susceptibility, and development of an array services platform to support multi-group collaborative projects.

Analysis of unclassified variants. In North America, Europe, Australia and Japan, genetic testing of high-risk cancer susceptibility genes is becoming an increasingly important component of the clinical management of at-risk patients and their close relatives. The vast majority of genetic testing of cancer susceptibility genes is directed towards the established high-risk breast cancer and colon cancer susceptibility genes, especially BRCA1, BRCA2, MLH1 and MSH2. De novo testing of an at-risk patient usually involves a mutation screen of the coding exons and proximal splice junction regions of the underlying susceptibility gene(s), often augmented with a screen for duplications or deletions of individual exons (*Tavtigian and Le Calvez-Kelm, 2007); consequently, the tests are technologically demanding and relatively expensive.

In addition to insertion-deletion mutations and other protein truncating sequence variants that are highly likely to damage protein function and are consequently generally classified as pathogenic *a priori*, mutation screening often reveals the presence of single nucleotide substitutions and other variants whose effects on gene function and disease risk are not immediately predictable. As a group, these are often referred to as unclassified variants (UVs). Over the last several years, we have contributed to a consortium focusing on the analysis of UVs in BRCA1 and BRCA2. Three notable achievements of our consortium have been: (1) to create a Bayesian method for assessing UVs that combines data across several independent data types in order to calculate an integrated posterior probability that a sequence variant is pathogenic (*Goldgar et al., 2004, Easton et al., 2007; Goldgar et al., 2008; Tavtigian et al., 2008); (2) to convene in February 2008 an IARC Working Group on Unclassified Genetic Variants that resulted in clinically applicable guidelines for UV classification (*Plon et al., 2008) (Tables 1 and 2) and began the diffusion of our Bayesian integrated evaluation beyond the breast cancer genetics community; and (3) to convene in February 2009 an IARC Working Group on Unclassified Genetic Variants in the mismatch repair genes, with the specific intent of adapting the Bayesian integrated evaluation to the colon cancer susceptibility genes.
### Table 1. Proposed Classification System for Sequence Variants Identified by Genetic Testing

<table>
<thead>
<tr>
<th>Class</th>
<th>Description</th>
<th>Probability of being pathogenic</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Definitely pathogenic</td>
<td>&gt;0.99</td>
</tr>
<tr>
<td>4</td>
<td>Likely pathogenic</td>
<td>0.95–0.99</td>
</tr>
<tr>
<td>3</td>
<td>Uncertain</td>
<td>0.05–0.949</td>
</tr>
<tr>
<td>2</td>
<td>Likely not pathogenic or of little clinical signif</td>
<td>0.001–0.049</td>
</tr>
<tr>
<td>1</td>
<td>Not pathogenic or of no clinical significance</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

### Table 2. Testing Recommendations Associated with Each Class of Variant

<table>
<thead>
<tr>
<th>Class</th>
<th>Clinical Testing</th>
<th>Surveillance recommendations if at-risk relative is positive</th>
<th>Research testing of family members</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Test at-risk relatives for variant</td>
<td>Full high-risk surveillance guidelines</td>
<td>Not indicated</td>
</tr>
<tr>
<td>4</td>
<td>Test at-risk relatives for variant*</td>
<td>Full high-risk surveillance guidelines</td>
<td>May be helpful to further classify variant</td>
</tr>
<tr>
<td>3</td>
<td>Do not use for predictive testing in at-risk relatives*</td>
<td>Based on family history (and other risk factors)</td>
<td>May be helpful to further classify variant</td>
</tr>
<tr>
<td>2</td>
<td>Do not use for predictive testing in at-risk relatives*</td>
<td>Treat as “no mutation detected” for this disorder</td>
<td>May be helpful to further classify variant</td>
</tr>
<tr>
<td>1</td>
<td>Do not use for predictive testing in at-risk relatives*</td>
<td>Treat as “no mutation detected” for this disorder</td>
<td>Not indicated</td>
</tr>
</tbody>
</table>

*Recommend continuing to test proband for any additional testing modalities available for the disorder in question: e.g., rearrangement testing.

**Case-control mutation screening of intermediate-risk breast cancer susceptibility genes.** The known high-risk breast cancer susceptibility genes explain about 25% of the familial risk of breast cancer, and the common risk-SNPs detected by recent GWAS studies are not responsible for more than about 10% of the familial relative risk. Thus in breast cancer (as well as colon and prostate cancer) genetics, there is an emerging problem of «missing heritability» (Maher, 2008; Easton and Eeles, 2008). One strong possibility is that uncommon-to-rare variants in intermediate-risk susceptibility genes, typified by ATM and CHEK2, are responsible for an important component of the missing heritability.

We are just finishing Year 2 of a 5-year NIH-funded project to examine this hypothesis. The main approach of the project is full open reading frame mutation screening of carefully selected candidate genes from a series of 1250 breast cancer cases and a similar number of ethnically-matched controls. The candidate genes are selected each year by an advisory committee, and the majority of the cases and controls are from the population centers of the NIH sponsored Breast Cancer Family Registries. Preliminary results have been encouraging. We have published a laboratory methods paper (*Nguyen et al.*, 2009) and an analysis of the intermediate risk susceptibility gene ATM (*Tavtigian et al.*, 2009). In the latter work we demonstrate the effectiveness of our bioinformatic approach to analysis of rare missense substitutions while also demonstrating the importance of rare missense substitutions in ATM to breast cancer susceptibility. Over the next three and a half years, we will be able to analyse a considerable number of candidate genes via this approach, and look forward to further elucidating the genetic basis of breast cancer susceptibility.

**Genetics of melanoma susceptibility.** Mutations in two genes encoding cell cycle regulatory proteins have been shown to cause familial cutaneous malignant melanoma (CMM). About 20% of melanoma-prone families bear a point mutation in the CDKN2A locus at 9p21, which encodes two unrelated proteins, p16 (INK4a) and p14 (ARF). Rare mutations in CDK4 have also been linked to the disease. Although the CDKN2A gene has been shown to be the major melanoma predisposing gene, there remains a significant proportion of melanoma kindreds linked to 9p21 in which germline mutations of CDKN2A have not been identified through direct exon sequencing. To assess the contribution of large rearrangements in CDKN2A to the disease, we performed multiplex ligation-dependent probe amplification (MLPA) in the French melanoma-prone families set. Overall, we showed that genomic deletions represent 2.1% of total mutations in this series (*Lesueur et al.*, 2008). In melanoma-prone families, the effect of CDKN2A is modified by subject-related phenotypes such as skin type, nevus count and sun sensitivity, as well as genetic variants in the highly polymorphic BIennal report 2008/2009
pigmentation gene MC1R. We have investigated the effect of the GST genes, which are involved in detoxification of metabolites after UV exposure, on melanoma risk in multigenerational melanoma-prone families with CDKN2A mutations. We found that the GSTT1 null allele modifies the risk of developing melanoma in carriers of a high-risk CDKN2A mutation, even after adjustment for MC1R genotype and host factors. Thus it is becoming clear that multiple genetic modifiers influence melanoma risk (*Chaudru et al., 2009).

Following a strategy similar to one we have developed to identify and analyse intermediate-risk genes for breast cancer, our next goal is to investigate strong candidate genes of the pigmentation pathway through a case–control mutation screening using subjects from the EPIC cohort.

**Array services.** The GCS Group took delivery of an Illumina BeadArray reader/Goldengate platform in April 2008. Workflows for SNP genotyping, methylation profiling and gene expression profiling have been validated, and GCS staff have been trained accordingly. Several projects have been executed on the Illumina platform. In support of a GCS breast cancer genetics project, we have created and validated a custom 384-SNP worldwide ancestry informative marker panel. In support of an EGE project, we used the Illumina Cancer Panel I methylation kit to profile the promoter methylation of 807 cancer-related genes in a series of hepatocellular, breast and esophageal carcinomas and surrounding tissues. Manuscripts related to the methylation studies are in preparation, and larger sample series will likely be analysed in the near future. In support of a MOC project, the Illumina Platform was used to perform gene expression profiling on a series of breast cancer cell lines to assess how p53 status affects the transcriptional response of these cells to estradiol or to the selective estrogen receptor modulator tamoxifen. Analyses are ongoing and a manuscript should follow.
REFERENCES


PUBLICATIONS


REFERENCES


PUBLICATIONS


