International Agency for Research on Cancer

World Health Organization

**Scientific Council** **SC/48/6**
**Forty-eighth Session** 18/11/2011

*Lyon, 1–3 February 2012*
*Auditorium*

## UPDATE ON RESEARCH PROGRAMME OF GENETIC CANCER SUSCEPTIBILITY (GCS)

The Genetic Cancer Susceptibility Group (GCS) was reviewed in 2010 as part of the Section of Genetics (GEN). As the GCS Group Head had been appointed immediately prior to the review, the Review Panel recommended that the GCS Group Head provide the Scientific Council with an update of the GCS scientific programme in 2011.

The GCS Group investigates the extent to which genetic variation contributes to cancer etiology. The general strategy of GCS is to take an integrative approach within our studies, including re-sequencing, genotyping, copy number variation and expression analysis, assayed within our laboratory and coupled with bioinformatics approaches, to identify and describe genes involved in cancer susceptibility. GCS's approach is particularly suited to the study of genetic variants with lower population frequencies.

In keeping with the goals of IARC, we aim to investigate genetic susceptibility of cancer sites that are generally less commonly studied by national centres, and with emphasis on cancers important in low- and middle- income countries (for example nasopharyngeal cancer [NPC] in south-east Asia). We also intend to take advantage of the exceptional bio-repositories that have been built up within the GEN Section and IARC. These incorporate extensive collections of biological material and clinical data relevant to lung cancer, kidney cancer and head and neck cancers.

The medium-term future (i.e. next one to four years) of the GCS Group is focused on opportunities stemming from the installation of next generation sequencing (NGS) at IARC. GCS is initially exploiting exome sequencing to identify germ-line susceptibility variants. One of our key challenges is how to distinguish the few variants that contribute to cancer susceptibility from the large number of unrelated variants identified by exome sequencing. Some of our strategies to permit this differentiation include:

- The use of information derived from the somatic events occurring during tumorigenesis to inform our analysis of germ-line events (for example using the fact that many hereditary tumours develop according to Knudson's two hit hypothesis).
- The use of genetics based studies, for example segregation evidence within pedigrees (extended families or trios) to detect disease-related variants (for example a large NPC family from Malaysia).
- The use of *in silico* approaches to enable the triage of variants into those that have a potentially functional effect and those that do not.

- The use of association-based tests to detect disease-related genes when aggregating putative functional variants at a gene level.
- The application of multi-stage experimental designs to allow cost-effective detection and validation of cancer susceptibility genes.
- The adaption of NGS technologies (for example, sequencing a few candidate genes but in many samples) to perform sequencing based replication studies within large IARC based bio-repositories.

We have initiated several new projects that illustrate the general approach that we are taking within GCS.

*Genetic susceptibility to nasopharyngeal carcinoma (NPC)*

NPC has a familial component, but apart from some loci identified by genome-wide association studies (GWAS) and studies of the HLA locus, the variants that explain this genetic risk remain unclear. Isolated populations offer rare opportunities to investigate the genetic cause of human disease. The Bidayuh ethnic subgroup of Sarawak, Malaysia has one of the highest incidence rates of NPC reported (Devi *et al.*, 2004). As part of an ongoing GEN study, we have identified one very large multiplex pedigree originating from an isolated region near Kuching, Sarawak (see Figure 1 below). Blood samples have been collected for 11 of the 26 NPC patients in this pedigree, as well as first-degree relatives for an additional five deceased cases. The samples have been shipped to IARC and DNA extracted.

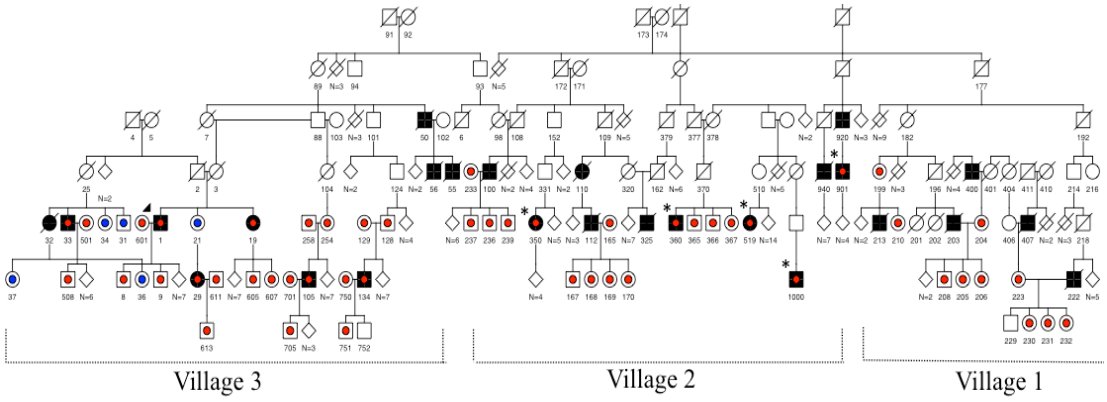


Figure 1. Large extended Bidayuh pedigree from a village located in a mountainous region located approximately 100 kilometers north east of Sarawak, Borneo, Malaysia. Males are squares, females circles, diamond indicate multiple individuals. Triangle indicates the proband. Internal circles mark individuals for which a blood sample has been collected. Oral history obtained at interview suggests that Village 1 is founder of Villages 2 and 3.

This project aims to perform comprehensive genetic analysis of this pedigree using whole exome DNA sequencing. Six individuals have been exome sequenced at the time of writing. After variant filtering to include only novel putative functional variants via bioinformatics analysis, we will assess for segregation of alleles within the pedigree. Variants of interest will be investigated further in additional NPC cases from Malaysia (from four pedigrees and 92 trios), as well as a broader case control series from Malaysia, Thailand and Singapore.

Additionally, we have developed a collaboration with Dr Allan Hildesheim of the division of Cancer Epidemiology and Genetics at the US NCI who is conducting a parallel study in

Taiwanese NPC pedigrees. With Dr Hildesheim we plan to compare, contrast and cross validate results from Bidayuh and Taiwanese pedigrees.

We have applied for extra-budgetary funding to support this project (NCI R03CA165037: Genetic analysis of a large multiplex nasopharyngeal carcinoma family, Impact/priority score 22).

*Integration of somatic information into genetic studies: Using the two-hit model to assist in lung cancer susceptibility gene discovery*

Many tumours resulting from a hereditary predisposition appear to act along the lines of Knudson's hypothesis: one mutant allele is inherited and the other (wild-type) allele is altered by a somatic mutation during tumorigenesis. The co-occurrence of (different) germ-line and sporadic mutations in the same gene in a given individual is a relatively rare event by chance alone. Unusual clustering of dual gene mutation events may potentially be used to map genes involved in cancer susceptibility. *PALB2* was recently identified as a familial pancreatic cancer susceptibility gene using such an approach (Jones *et al.,* 2009), a finding that was independently replicated (Casadei *et al.,* 2011).

We are exploring this method as a gene discovery tool. We have initially focussed on lung cancer within the exceptional bio-repository developed within the GEN Section. This now stores DNA from approximately 6000 cases and controls, and of these for more than 325 lung cancer patients there is a blood sample and a snap frozen resection of the corresponding tumour sample (oversampled for early stage tumours). In an initial pilot project, we have selected 12 patients with potentially *a priori* high genetic risk (selecting patients with two first degree affected relatives also with lung cancer, or an early age of onset ≤ 40). We will use exome re-sequencing (germ-line and tumour) and copy number analysis (tumour) to assay the putative functional variants carried by each patient. Subsequently we will triage variants to identify genes potentially acting along the lines of the two-hit model, i.e. a germline mutation and a different somatic mutation in the same gene. We will also use lung cancer sequencing data from The Cancer Genome Anatomy (TCGA) project, particularly those with an early age of onset, to complement the IARC based sequencing.

Noteworthy genes will then be validated through targeted re-sequencing in the remaining 313 paired normal and lung tumour samples in the IARC lung cancer bio-repository, as well as cases and controls sourced from the wider lung cancer bio-repository (focusing on high genetic risk).

*Investigation of common and rare genetic variants in melanoma cancer 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 goal of this study is the identification of new melanoma susceptibility genes and the characterization of the pathogenic sequence variants associated with increased risk of developing melanoma. Bioinformatics tools can assess the functional consequence of the variants *in silico* using the degree of evolutionary conservation observed at each position on the protein sequence (SIFT [Ng *et al*., 2002;

Adzhubei, *et al.*, 2010], Align-GVGD [Tavtigian *et al.*, 2008]). We have applied a case-control mutation screening approach to demonstrate the efficiency of ranking rare missense substitutions using *in silico* prediction programs before comparing the distribution and frequencies of the different classes of variants in a series of early-onset breast cancer cases and controls (Tavtigian *et al.*, 2009; Le Calvez-Kelm *et al.*, 2011). We have set up a large-scale case-control mutation screening study nested in the EPIC cohort and intend to investigate candidate genes involved in the skin pigmentation pathway through targeted sequencing using the Ion Torrent instrument.

We have applied for extra-budgetary funding to support this project (NCI R03CA156624-01A1: A new approach to identify rare genetic variants influencing melanoma risk, Impact/priority score 22).

*The GCS Group's interaction with the GEN Section*

GCS remains complementary to work carried out throughout the GEN Section. GCS remains involved with the GWAS coordinated by the Genetic Epidemiology Group (GEP), with publication of GWAS on upper aero-digestive tract cancers (McKay *et al.,* 2011; Chen *et al.,* 2011), kidney cancer (Purdue *et al.,* 2011), and human papillomavirus (HPV) serology (Chen *et al.,* 2011) during 2011. We have also finalized scans of Hodgkin lymphoma (Urayama *et al.,* JNCI submitted) and oral-oropharyngeal cancer (Johansson *et al.,* submitted). During 2012, GEP and GCS will undertake an extensive GWAS of kidney cancer (U01CA155309, PI: G. Scelo, GEP; Co-PIs: J. McKay and F. Le Calvez-Kelm, GCS). GCS will be particularly involved in expression analysis, with the incorporation of expression data into GWAS studies (particularly via an eQTL analysis) and using expression profiles as determinants of kidney cancer survival.

The Biostatistics Group (BST) is involved with the majority of GCS projects, both in terms of individual projects and development of the bioinformatics aspects of NGS data.

**Medium-term plans**

Within the medium term, GCS intends to expand several aspects of these projects and launch additional initiatives. For example:

- Using genomics techniques (exome re-sequencing and gene expression profiles) and extensive characterization of HPV status to compare and contrast the molecular profiles of HPV positive (50) and negative (50) oro-pharyngeal tumours collected from Europe and South America.
- Investigate the possibility of the "two-hit" model mapping, as well as other exome sequencing based analysis strategies, in other cancer sites, particularly kidney and head and neck cancers enriched for *a priori* high genetic risk.
- Expand the focus of NPC studies to explore genetic susceptibility to NPC across additional case-control series collected from Thailand, Indonesia and Singapore.

*The Genetic Services Platform (GSP)*

Finally, GCS also remains committed to developing and supporting genomics-based techniques across the Agency through the Genetic Services Platform (GSP). GSP has carried out genomics based studies (principally Illumina beadarrays, High Resolution melting genotyping/mutation screening and Taqman genotyping) in collaborative studies with GEP, MOC, EGE, MPA, ICB, ICE and ENV[1] through 2011.

A major GCS priority during 2011 has been the installation of NGS within the GSP. The strategy for NGS at IARC was to implement a versatile sequencer to meet the relatively diverse sequencing needs of the Agency, with large scale projects outsourced. A LifeTechnologies 5500xl sequencer was installed in July and the first run was initiated in September 2011. An Ion Torrent Personal genome machine (medium-scale next generation sequencer) was installed in late October 2011. We are incorporating the NGS protocols into the automated workflows of GCS, under the umbrella of the Laboratory Information Management System (LIMS). There is considerable interest in this technology across the Agency, with pilot projects from GCS, GEP, ICB, EGE and MPA being initiated in exome sequencing, RNAseq, CHIP-seq and methylation.

GCS has put in place a medium-sized high-performance computing cluster for data analysis, as well as long-term data storage and backup. We are developing bioinformatic pipelines principally orientated around the LifeScope, the ABI's specialized analysis software designed for SOLiD reads, in consultation with the BST Group. A GCS staff bioinformatician (LY5) is managing the high-performance computing cluster (in collaboration with the Information Technology Services Group [ITS]), the LifeScope software, as well as informatics management of the sequencers themselves. We additionally have two very promising bioinformatics orientated PhD students working in the development of bioinformatic pipelines. We plan to grow the bioinformatics team organically, with the retention of one of these students as a bioinformatics post-doctoral fellow during early 2012, with a professional bioinformatician position, included in the 2012–2013 regular budget to be recruited in the later part of the year. In addition the wider GCS Group (staff scientists and post-doctoral fellows) have received initial training in LifeScope. We are additionally working with scientists from other IARC groups regarding bioinformatics analysis aspects relevant to their activities (RNAseq, CHIP-seq).

GCS has also developed collaborative links with Professor Gilles Thomas, who has recently formed a bioinformatics group (Synergie Cancer) focused on analysis of NGS data at the Centre Léon Bérard (Lyon). Bi-weekly seminars are held between the two groups regarding bioinformatics issues, and in particular quality control, analysis workflows and pipelines and tertiary analysis of NGS data. Professor Thomas has provided GCS (and other GEN researchers) with access to his large high-performance computing cluster, and Professor Thomas's group will have reciprocal access to IARC's NGS platforms.

[1] GEP = Genetic Epidemiology Group; MOC = Molecular Carcinogenesis Group; EGE = Epigenetics Group; MPA = Section of Molecular Pathology; ICB = Infections and Cancer Biology Group; ICE = Infections and Cancer Epidemiology Group; ENV = Section of Environment and Radiation.

*Literature cited*

Devi BC et al. High incidence of nasopharyngeal carcinoma in native people of Sarawak, Borneo Island. *Cancer Epidemiol Biomarkers Prev*. 2004;13(3):482-6

Jones S et al. Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene. Science. 2009;324(5924):217.

Casadei S, et al. Contribution of inherited mutations in the BRCA2-interacting protein PALB2 to familial breast cancer. Cancer Res. 2011 15;71(6):2222-9.

Ng PC, Henikoff S. Accounting for human polymorphisms predicted to affect protein function. Genome Res. 2002 Mar;12(3):436-46.

Adzhubei IA, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010 Apr;7(4):248-9.

Tavtigian SV, et al. Classification of rare missense substitutions, using risk surfaces, with genetic- and molecular-epidemiology applications. Hum Mutat. 2008 Nov;29(11):1342-54.

Tavtigian SV, et al Rare, evolutionarily unlikely missense substitutions in ATM confer increased risk of breast cancer. Am J Hum Genet. 2009 Oct;85(4):427-46.

Le Calvez-Kelm F, et al. 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. 2011 ;13(1):R6.

McKay JD et al. A genome-wide association study of upperaerodigestive tract cancers conducted within the INHANCE consortium. PLoS Genet. 2011 Mar;7(3):e1001333.

Chen D et al. A sex-specific association between a 15q25 variant and upper aerodigestive tract cancers. Cancer EpidemiolBiomarkers Prev. 2011 Apr;20(4):658-64.

Purdue MP, et al. Genome-wide association study of renal cell carcinoma identifiestwo susceptibility loci on 2p21 and 11q13.3. Nat Genet. 2011 Jan;43(1):60-5.

Chen D, et al. Genome-wide association study of HPV seropositivity. Hum Mol Genet. 2011 Sep 10. [Epub ahead of print]

## APPENDIX

## GCS staff and fellows

| | |
|---|---|
| Group Head | Dr James McKay (since 6 September 2010) |
| Scientists | Dr Fabienne Lesueur |
| | Dr Florence Le Calvez-Kelm |
| Laboratory Technicians | Ms Sandrine Chopin-McKay (until 31 October 2010) |
| | Ms Amélie Chabrier (since 1 July 2011) |
| | Mr Geoffroy Durand |
| | Ms Nathalie Forey |
| | Ms Jocelyne Michelon |
| | Ms Nivonirina Robinot |
| Bioinformatician | Ms Catherine Voegele |
| Secretariat | Ms Antoinette Trochard |
| Postdoctoral fellows | Dr Jamil Ahmad (26 November 2007–22 November 2010) |
| | Dr Francesca Damiola (10 November 2008–7 October 2011)) |
| | Dr Mona Ellaithi (since November 2011) |
| | Dr Javier Oliver (since 1 February 2011) |
| | Dr Maroulio Pertesi (since 14 March 2011) |
| | Dr Dewajani Purnomosari (since 2 September 2011) |
| Students | Ms Manon Delahaye (since January 2011) |
| | Ms Aurélie Fillon (12 April 2010–23 July 2010) |
| | Ms Célia Jolivet (11 April 2011–24 June 2011) |
| | Ms Bin Thieu Tù Nguyen-Dumont (26 September 2005–31 December 2010) |
| | Mr Wee Loon Ong (7 June 2010–13 August 2010) |
| | Ms Fanny Paquet (5 September 2011–18 February 2012) |
| | Ms Maroulio Pertesi (1 September 2010–1 December 2010) |
| | Mr Maxime Vallée (since 6 August 2007) |