International Agency for Research on Cancer

World Health Organization

Scientific Council **SC/48/8**
Forty-eighth Session 04/11/2011

Lyon, 1–3 February 2012
Auditorium

OPEN SESSION ON SCIENTIFIC TOPICS OF IMPORTANCE TO IARC

The advice from the Scientific Council on the two scientific topics selected by the Senior Leadership Team (SLT) for open discussion will be appreciated. Below is some background regarding the list of points for discussion on Next Generation Sequencing (Topic 1) and on Rare and emerging cancers (Topic 2).

Topic 1: Next Generation Sequencing (prepared by Dr James McKay, Head, GCS)

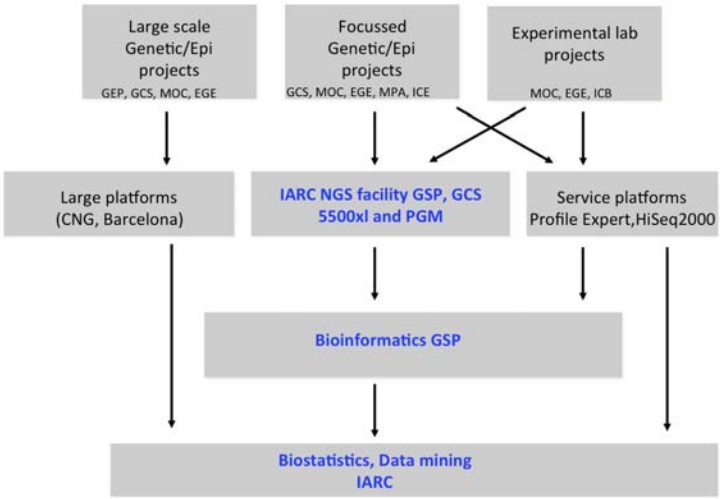
Next generation sequencing (NGS) (or massive parallel sequencing) allows the molecular events involved in carcinogenesis to be described in unprecedented detail. As such, NGS techniques are well-aligned with the research goals of many IARC Groups. The large and diverse biological resources stored within the IARC bio-repositories offer enormous potential for the application of NGS technology. However, this tremendous potential comes with associated challenges. Scientifically, NGS-based experimental designs and analysis strategies need to be sufficiently robust to obtain meaningful results from the large volume of data generated by these platforms. Appropriate methods are needed to subsequently validate findings (both in terms of infrastructure and sample series). Logistically, the running costs remain important and while the adoption of current "best practice" NGS bioinformatics analysis pipelines has facilitated data analysis, the bioinformatic and storage issues are not trivial.

Similar issues arise with the use of data made available through the open data access policies of initiatives such as the International Cancer Genome consortium (ICGC) and The Cancer Genome Atlas (TCGA). These online resources have widespread utility for many IARC Groups, but the network bandwidth needed to obtain these large data series, as well as data storage and analysis, must be considered when planning to use this "free" information.

NGS at IARC

As the scientific activities of IARC research Groups are relatively varied, the nature of IARC's NGS needs are diverse in terms of the type of applications employed and the origin and quality of the biological samples studied. In general the demand for NGS by IARC Groups is likely to be for intermittent, but focused, studies, as opposed to very high throughput studies. The strategy for NGS at IARC has therefore been to implement a sequencer that is as flexible as possible in terms of its function. Large-scale projects will continue to be outsourced to cost-effective

external sequencing facilities (Figure 1).



Flow chart of Next Generation Sequencing at IARC
(acronyms are for IARC Research Groups)

Through consultation with involved Groups and the Laboratory Steering Committee (LSC), a 5500xl sequencer (LifeTechnologies) was chosen based on its flexibility and appropriate cost. Importantly, in 2011 a service provider in the immediate vicinity of IARC has installed the alternative and complementary Illumina HiSeq technology, providing access to this technology to IARC researchers. The 5500xl sequencer was installed through the second part of 2011 as part of the Genetic Services Platform (GSP) within the Genetic Cancer Susceptibility Group (GCS). A smaller capacity Ion Torrent personal genome machine was additionally obtained, and is likely to prove useful in the context of validation studies (both technical confirmation of variants identified by the large scale sequencer, and replication studies of findings arising from the deep sequencing i.e. focused study of a few genes but in the large numbers of samples stored within the large IARC bio-repositories). The GSP is currently incorporating the NGS protocols into the automated workflows of GCS, under the umbrella of the Laboratory Information Management System (LIMS). A medium-size high-performance computing cluster for data analysis, as well as storage and long-term backup, have been installed within the dedicated server room of the Information Technology Services (ITS). The computing cluster is currently being managed by GCS bioinformaticians in collaboration with the ITS. The IARC network capacity has been expanded to facilitate data transfer.

The model for the functioning of the NGS platform within IARC is that the individual Groups take responsibility for the production of the project specific sequencing libraries and the GSP subsequently undertakes the completion of the sequencing itself. Initial data analysis (quality control and mapping of fragments to the given reference sequence) is undertaken by GSP bioinformaticians, with data transmitted back to the Groups for tertiary analysis, with support from the GSP and the Biostatistics Group (BST).

Subsequent to the installation of NGS technology, there has been considerable interest across the Agency, including GCS, EGE, ICB, GEP, MPA, MOC and ICE[1] Groups. Pilot projects (with additional financial support allocated by the Director) have been initiated with the aim of developing the different workflows and providing IARC Groups with access to NGS data.

Projects:

GCS – Exome (coding regions) sequencing to investigate germ line genetic susceptibility, Differential allele expression (RNAseq);

EGE – Targeted (CHIP-seq and MeDIP-seq) as well as genome wide methylation studies (liver cancers);

ICB – Transcriptome analysis for differential analysis of virus infected cells; Detection of novel infections involved in cancer susceptibility;

GEP – Expression and sequenced based analysis of kidney cancer;

MPA – Exome sequencing in the context of familial brain tumours (Li-Fraumeni like syndromes);

MOC – Rescreening for somatic mutations involved in candidate genes;

ICE – Rescreening for somatic mutations involved in candidate genes involved in cervical cancers.

Thanks to this development, it is now possible to undertake mutation analysis in coding regions (exomes) or whole genomes, epigenomes (methylomes), and transcriptomes at IARC. While GCS will coordinate these efforts, individual Groups will play a leading role in the establishment of different assays on this platform and ensure that genome-wide analysis reach a smooth production level in the coming few months.

We would appreciate Scientific Council members' input regarding the NGS model put in place and applications of NGS technologies at IARC. Potential discussion points include:

- Use of NGS technology to identify somatic profiles or signature mutations linked to environmental exposures (description of potential sequenced based biomarkers, exposure classification).
- Use of NGS in the investigation of the contribution to carcinogenesis of infectious agents (known and novel), particular focus on low- and medium-resource countries.
- IARC's contribution to ICGC initiatives. Potential for focus of IARC on cancer types specific to low- and medium-resource countries, and/or determining if genes identified in ICGC in high-resource situations are applicable to lower/medium resource countries with their etiological differences.
- The use of NGS in the development of translational studies through the discovery and validation of genetic and epigenetic biomarkers of tumorigenesis and environmental exposures.
- Application of IARC's bio-repositories using NGS, in particular focused studies (few genes, many samples).
- Ethical issues regarding use of NGS data (appropriate consent, data sharing policies).

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

Topic 2: Rare and Emerging Cancers (prepared by Dr Paul Brennan, Head GEN)

Introduction and Rationale

This briefing note introduces the rationale and implications for an IARC programme that focuses on rare and emerging cancers, which builds on IARC's distinctive expertise and experience.

Research initiatives that target rare and emerging cancers can be of scientific importance, as demonstrated by the following:

- Studies of individuals with certain syndromes associated with increased cancer risks have resulted in discoveries of causal mutations (e.g. APC, BRCA1, BRCA2, MMR genes), which permits screening or prevention interventions to target known carriers, and benefits for the individual.
- The large variation between nations with high versus low cancer incidence rates (i.e. what is rare in one population may be common in another), indicates the presence of strong carcinogenic agents. International studies that incorporate both high and low incidence areas may provide a powerful study design for detection of novel carcinogens.
- Studies of rare syndromes in which disease-causing genetic factors are first detected, may provide evidence of mechanisms that have relevance to a wider range of both rare and common cancers (e.g. p53 mutations in Li-Fraumeni syndrome).
- Studies of rare cancers for which rates were seen to rise in certain populations, have resulted in discoveries of causal factors, and ultimately to new preventive interventions (e.g. HPV, HIV, melanoma and UV). Cancers that have risen in the recent decades for which an underlying causal mechanism is still unclear include testicular cancers and non-Hodgkins lymphomas.
- Studies of rare cancers for which there were concerns about specific exposures have resulted in clarification of some of the possible risk factors (e.g. studies exploring risk factors for childhood cancers).
- Studies of rare subtypes of common tumours, defined by histology or tumour markers, can lead to the identification of distinct etiologies and novel carcinogens (triple negative breast cancers, lung adenocarcinomas in never-smokers).

Objectives

Potential areas where IARC fostered studies of rare and emerging cancers may be particularly valuable include:

- Tracking of cancer trends to provide an early warning of rare cancers that are potentially increasing in incidence in more than one population;
- Development of causal hypotheses by bringing together working groups and initiation of subsequent studies of the role of exogenous and endogenous factors in the development of rare and emerging cancers;

- To foster, lead and participate in international consortia and projects that are implemented in multiple centres to cover large enough populations so that rare events and complex etiological relationships can be investigated;
- To share expertise, resources and guidance that permit studies of rare and emerging cancers in populations where there is particular need, including studies of rarer tumours that have an important burden in low- and middle-income countries; and
- To contribute new evidence that supports improvements in cancer prevention and control.

Definitions and Criteria

Definitions for rare cancers have been considered for various purposes, and accordingly, a wide range of definitions have arisen. Definitions based on an annual incidence rate of 15 per 100 000 population or greater have been used in some settings, but these include a large fraction of all cancers. Clinical research consortia (e.g. Rare Cancers Europe – www.rarecancers.eu) use an annual incidence rate of < 6 per 100 000, which in an analysis of European cancer registries classified 194 specific tumour types as rare, accounting for 19% of all cancers.

Using 6 per 100 000 to classify worldwide cancer incidence rates, approximately 20% of all cancers would be deemed rare. International differences include cancers of the liver and oesophagus that are relatively more frequent in some lower- and middle-income countries but rare in high-income countries, while in contrast, several cancers are relatively frequent in high-income countries but rare in low/middle income countries.

More restrictive definitions have also been considered, such as for 'ultra-rare' cancer based on an incidence rate of < 1 per 100 000 annually. Studies of cancers with this degree of rarity must refer to very large populations to achieve reasonable levels of scientific quality and logistical efficiency, usually involving international consortia for collaborative research, such as those established by IARC.

While any definition of rarity requires a decision about the point at which rates indicate that an event is appropriately infrequent, consideration must also be given to the definition of the event itself. Given the clinical or scientific implications of the heterogeneity that exists between cancers, and the fact that it is increasingly feasible to identify more homogeneous subgroups, such as by molecular markers, these subgroups may then fit the definition chosen for rarity. Criteria used to identify rare subsets of cancers at specific sites and for specific purposes include:

- ***Age***, such as in the recognition that cancer in children is usually distinct from cancer in adults.
- ***Genetic predisposition,*** such as for an inherited cancer or syndrome that includes the occurrence of one or more types of cancer (e.g. Familial adenomatous polyposis, Hereditary non-polyposis colon cancer, retinoblastoma…).

- ***Molecular or cytogenetic features*** that characterize variants of common cancers that may be acquired during tumourigenesis (e.g. breast cancers that are negative for estrogen, progesterone and HER2 receptors).
- ***Clinical or biological features*** of common cancers such that the nature of the disease is different from that cancer's typical behaviour (e.g. early stage lung cancer or a precursor phase such as myelodyplastic syndrome that may lead to leukemia).
- ***Other epidemiological features,*** such as: a cancer that appears to have a variant form in a specific setting, ethnic or demographic group; where trends indicate that a cancer is emerging in a certain population; or where a cancer arises from a specific cause or exposure (e.g. Kaposi's sarcoma, Burkitt's lymphoma, mesothelioma, melanoma, distinguishing cholangiocarcinomas from more common liver cancers, clear cell adenocarcinoma of the vagina among young women after diethylstilbestrol).

Strategic Approach and Logistics

Research on rare and emerging cancers is consistent with IARC's mission, but also builds on the strong foundation established through its leadership of worldwide cancer research. This approach integrates IARC's roles: in leading international, inter-disciplinary collaborations that focus on the causes of cancer with the aim of improving prevention; as the global reference centre for cancer information, through cancer registries and pathology standards; as the provider of high quality data and recommendations to guide research and global cancer control policies; in identifying and addressing knowledge gaps of particular interest to low- and middle-income countries; in capacity building by sharing specialized expertise and resources, and through education and training. A significant feature of IARC is its expertise in coordinating research across countries and organizations, which is facilitated by its independent role as an international organization and its proven record of accomplishments.

Elements of research related to rare and emerging cancers that are consistent with IARC's ongoing activities include:

- Large international consortia that employ common protocols and harmonized procedures would be necessary to ensure data quality and that sufficient numbers of the rare cancers can be included;
- It may be possible to conduct studies of certain rare cancers using existing cohort and case-control consortia, particularly as cohort follow-up extends;
- Expertise in setting standards and in coordinating cancer registration and pathology classification will be important elements of any new projects;
- The setting of priorities for new studies and consortia can be based on the expertise in cancer information and descriptive epidemiology, as rare, very rare or emerging cancers can be identified and monitored;
- New hypotheses and the design of new studies can be motivated by: (a) ongoing monitoring of rare cancer rates that detects a recurrence of trends or patterns; and (b) cancers that are rare and that have increasing rates, as this would suggest a novel pattern of carcinogen exposure;

- Some areas of research were never before feasible due to the extreme rarity of the tumours, yet the successes of other large consortia for environmental studies, genetic studies and cohort pooling, demonstrate that it is now feasible to undertake international, collaborative studies (e.g. for rare childhood cancers such as Wilms tumour, neuroblastoma, brain tumours);
- The selection of locations to launch new studies can take into consideration where the needs and opportunities are greatest, with continuing efforts to enable studies in lower-income countries;
- Existing relationships with national cancer programmes and their directors may provide an appropriate mechanism for establishing studies in new regions.

Potential contribution of IARC Sections and Groups

There is an important leadership role for each IARC Section and Group as worldwide studies of rare cancers are launched and coordinated. IARC Sections can guide the development of new research and set directions for future prevention and control of rare and emerging cancers by:

- **Cancer Information:** collecting, analysing and disseminating information regarding the global cancer burden, patterns and trends, thereby identifying and monitoring rare and emerging cancers that require investigation.
- **IARC Monographs:** identifying gaps in knowledge and opportunities for new research, and by aiding in the interpretation of new findings as part of the ongoing rigorous review and interpretation of contemporary evidence.
- **Mechanisms of Carcinogenesis:** identifying gaps in knowledge and scientific opportunities, and by coordinating research on mechanisms that underlie interactions between the environment, the genome and the epigenome.
- **Molecular Pathology:** maintaining standards for histopathological and genetic diagnostic criteria, which are the basis of worldwide cancer information and research, and by coordinating research and evaluation related to the discovery and characterization of tumour markers.
- **Infections/Environment and Radiation/Nutrition and Metabolism/Genetics:** identifying gaps and opportunities, and by coordinating studies of risk factors.
- **Early Detection and Prevention:** identifying needs, opportunities and priorities for new research, by establishing cancer control policy implications of new findings, and by evaluating resultant primary and secondary prevention strategies.

Funding Mechanisms

On one level this proposal restates IARC's long-standing commitment to all cancers (which has consistently included rare cancers), as well as to worldwide collaborations and impact; however, more emphasis on rare cancers does have resource implications. One advantage of this area of work is that many of the essential resources presently exist throughout IARC Sections, as well as in the far-reaching collaborations that have already been established. Even so, for studies to be done in low- and middle-income nations, funding from other sources will usually be required. Major external funding agencies will continue to be primary sources for IARC's research (e.g. EU, NIH, AICR, etc.), and to succeed it must be clear that the value, quality and cost-effectiveness of the research will not be diminished by the rarity of the cancers studied, or the places where the studies should occur. However, for issues related to scientific importance, health impacts, feasibility, and cost-effectiveness of study logistics, plus IARC's unique role, expertise and record of productivity, a programme that targets rare and emerging cancers, although challenging, can certainly succeed.