International Cancer Genome Consortium

2nd Scientific Workshop

June 22–24, 2009

Wellcome Trust Conference Center
Hinxton, United Kingdom

TABLE OF CONTENTS

I. Overview

On June 22–24, 2009, the International Cancer Genome Consortium held its second scientific workshop at the Wellcome Trust Conference Centre. With its goal of analyzing the genomes of at least 500 tumors and matched normal tissue for each of 50 specific types or subtypes of cancer, the ICGC may be the most ambitious project stemming from the Human Genome Project.

Goals and challenges of the ICGC

- Obtaining and analyzing the sequence of 50,000 human genomes (25,000 pairs of diseased and normal tissue);
- Obtaining quality biospecimens, with quality annotations;
- Sorting through the huge numbers of mutations in cancer to identify the mutations that drive cancer;
- Maintaining the long-term commitment and support of funders.

Motivation

- Genome sequencing technology is progressing rapidly. At present the cost of sequencing a human genome is approximately US$100,000 and decreasing rapidly. Therefore, the goals of the ICGC are increasingly approachable;
- Cancer is a particularly promising area for personalized medicine;
- ICGC research will lead to greater understanding of the etiology disease, leading to new classifications of disease and fundamental biological insights and new therapies.

History and Progress of the Consortium

The ICGC was launched in Toronto in October 2007. Its goals, structure, policies and guidelines were published in April 2008. Participants representing 22 countries decided to:

- Share data;
- Set the goal of analyzing 50 cancer types or subtypes with at least 500 tumor specimens per cancer type;
- Perform genomic, transcriptomic and epigenomic analysis, including whole-genome sequencing and characterization with interim technologies;
- Set a commitment threshold for funding agencies of US$20 million per tumor type.

In November 2008, at the first scientific workshop in Bethesda MD, the ICGC established scientific committees and working groups.

Currently, several countries are close to making commitments to ICGC projects. Many are considering taking on second or third projects.

Evolving policies

- A working group for data analysis: It was decided not to establish such a group; the decision may be reconsidered;
- Requirement to analyze 500 tumor specimens per tumor type: This expectation has become more flexible because:

    - For rare tumor types fewer specimens may be available;
    - For highly heterogeneous tumors, more than 500 samples may be necessary to capture the genomic diversity.
- *International collaboration on individual tumor types:* For rare tumor types, e.g., pediatric tumors, the ICGC is now encouraging research teams to consider international collaboration to pool resources.

Major Results and Decisions from Working Groups Since November 2008

I. Consent and Data Access

- To oversee access to human subject data, the ICGC established the International Data Access Committee (IDAC). The Data Access Compliance Office (DACO) led by Yann Joly, a lawyer at McGill University, will administer access;
- IDAC and DACO will be independent of the Data Coordination Center (DCC) but will work closely with it;
- To obtain access to controlled data:
    - Researchers will fill out a form requesting access;
    - DACO will review requests;
    - DACO will notify IDAC and DCC, which will enable users to access the controlled database.
- The working group has created a template form for prospective patient informed consent and is planning to develop a template form for retrospective patient informed consent.

II. Tissue and Clinical Annotation

- This group consists of two subcommittees: One on clinical annotations is addressing clinical data elements and quality control. One on tissue sample quality is addressing specimen annotation and quality control.

III. Technologies

- Each ICGC project will need to devote resources to *validating identified changes, including mutations;*
- *Samples* are available to member groups for quality control and to compare technologies:
    - COLO-829, a malignant melanoma cell line, which has been sequenced and has a known catalogue of somatic mutations, is available as a reference genome.
- *Epigenetics*: ICGC is in discussions with Illumina about designing a 500K cytosine phosphate guanine (CpG) array that includes CpG sites with known functional significance.

IV. Data Coordination and Management

- Subcommittees have issued a number of recommendations;
- *Design document* addresses issues such as: How users will query database, data elements, input data format.

Goals of This Meeting

- To obtain progress reports from projects that are already underway;
- To facilitate coordination and sharing of protocols among projects;
- To identify opportunities for new projects that can be shared across countries;
- To define quality assurance (QA)/quality control (QC) procedures; set thresholds;
- To discuss whether to initiate a bioinformatics analysis working group.

II. Status of ICGC Projects

Pancreatic Adenocarcinoma (Canada)

Design

- Chose pancreatic cancer because survival rates remain very low;
- Tumor cells tend to be contaminated with normal cells so decided to use xenograft implants (followed by cell culturing) to obtain more homogeneous specimens and grow more tissue;
- Samples are being collected from the Toronto Hospital, Mayo Clinic and Harvard Hospitals, all with pipelines for creating xenografts;
- Plan to sequence exomes, analyze structural variations and copy number variations, sequence whole transcriptomes, and eventually sequence whole genomes.

Progress and data collected

- Currently doing genome analysis and sequencing small RNAs;
- About 10 tumor specimens have been converted to xenografts with a high (>90 percent) success rate.

Challenges

- Transcriptome analysis is challenging because of short sequence reads;
- Primary tumor is usually less than 50 percent tumor cells;
- Mouse contamination is higher than expected in xenografts (35–50 percent of deoxyribonucleic acid (DNA) is mouse). Therefore, plan to isolate tumor cells.

Pancreatic Adenocarcinoma (Australia)

Status of projects

- Expect funding from Australian government to begin in July 2009. Funding will be from federal and state governments;
- In addition to pancreatic cancer, Australian researchers plan to study ovarian cancer. Both are five-year projects;
- The Queensland Centre for Medical Genomics has been newly constructed and should be operational within weeks.

Design

- Specimens will be from surgically resected tumors;
- Plan to generate xenografts but not on a large scale as in the Canadian project;
- For both pancreatic and ovarian cancer, plan to sequence *transcriptomes* of samples, with 1.5 x 10^8 reads and 10^7 micro ribonucleic acid (RNA) (mRNAs) per tumor. Automated methods will be used for sequencing both large and small RNAs;
- For genome sequencing, plan on 25X coverage for both tumor and matched normal samples using map paired-tag approach. Interested in paired-end read approach in future;
- Expect to be able to do methylome sequencing (20 million reads per tumor).

Viral Hepatitis-Associated Hepatocellular Carcinoma (Japan)

Design

- Tumor under study: Hepatitis virus-associated liver cancers;
- Structure of project:
    - Two research groups supported by two funding agencies;
    - Four collaborating hospitals provide samples;
    - Two collaborating research groups for technology development and bioinformatics and two expert researchers for genome ethics and database access policies.

Progress and lessons learned

- Have obtained 170 paired samples of liver cells (and blood samples);
- Have started whole-genome shotgun sequencing, with paired-end sequencing of short (~300 base pairs) insert and large-insert libraries;
- Copy number analysis and single-nucleotide polymorphism (SNP) genotyping with SNP arrays and oligonucleotide arrays are being evaluated to complement sequencing;
- Compared pipeline of v1.3 and v1.4 Gall sequencer and found v1.4 increased rate of data generated with lower error rate;
- Have used a small number of reads on liver cancer cell lines to detect structural variations, such as deletions, tandem duplications. Structural variations have been validated;
- Shotgun sequencing has been used to detect sites of hepatitis virus integration. In regions of repeat sequences, additional analysis is required;
- Using exon capturing to sequence exons, detected many loss-of-heterozygosity SNPs and candidate mutations found only in tumors. Among the latter, found a missense mutation in tumor protein 53 (TP53).

Challenges

- Detected normal cell contamination in liver tumor samples;
- To deal with normal cell contamination and intra-tumor heterogeneity, deep sequencing can be used and compared with paired samples across techniques.

Alcohol-Related Hepatocellular Carcinoma (HCC) (France)

Design

- Tumor under study: alcohol-associated liver cancer. Liver cancer is associated with several risk factors; no gene mutations were known to be linked to early steps of HCC carcinogenesis;
- Whole-genome sequencing and a range of analytical techniques are being applied to 14 HCC tissues and paired normal specimens. Mutations and rearrangements will be validated on 500 HCC and benign tumors.

Progress and data collected

- For several years, a network of biobanks in France has been amassing a liver cancer databank, which now contains more than 1,000 tumors categorized by etiology (alcohol, hepatitis virus C or B, etc.);
- Frequently detected mutations:
    - TP53 is mutated in 20 percent of tumors tested in France; 50 percent in other studies;
    - b-catenin activating mutations.
- Have found several genes that are occasionally mutated in HCC;
- Tumors with stable vs. unstable chromosomes are genetically and clinically distinct. For example, chromosome-stable tumors are not HBV-infected;
- Transcriptome classifications of tumors are correlated with their genomic classifications, including particular mutations and methylation status, and can be related to biological pathways. MicroRNA profiles also relate closely to transcriptomic/genomic classification.

Centre National de Génotypage (CNG)

- Traditionally a genotyping center, CNG has recently established sequencing capability and is participating in the French liver cancer study;
- Used Illumina 610 quad arrays to type 600 kSNPs;
- Completed genome sequencing for 3 pairs of tumors and controls, with 30X coverage;
- Procedures include acoustic shearing, using variable insert-size libraries, and adjusting number of sequencing cycles depending on intensities of reads;
- Modified database (Operon) used previously for genotype data. Gives QC overview of sequencing data and is a good tool for analysis and for transferring data to collaborators;
- Sequences are stored as coordinates, with only differences stored rather than whole sequences. A minimum of 6X coverage allows for SNP calling.

Her2 Positive Breast Cancer (France)

Design

- Studying human epidermal growth factor receptor 2 (Her2) positive breast cancer in international consortium on breast cancer chaired by Mike Stratton;
- To obtain specimens, initiated collaboration with four leading French cancer centers. In autumn 2009, plan to add five additional cancer centers and university hospitals throughout France to this network;
- Necessary criteria in tumors for sequencing: full clinical information available, validation by teams of pathologists, high proportion of tumor cells, tumors larger than 100 mg in size, patients alive in order to provide informed consent and blood.

Progress and data collected

- Developed informed consent procedures for both prospective and retrospective collection of tumor material. Patients have been told that their DNA may be extensively sequenced and transferred abroad;
- Developed common procedures for the selection and processing of samples. In processing, those procedures include extraction of nucleic acid, quantification, QC, comparative genomic hybridization (CGH) array, transcriptome analysis, SNP analysis;
- Have selected first set of tumors as candidates for sequencing. Are analyzing Her2 positive specimens further before committing to sequencing particular tumors;
- French and international pathology groups have met and evaluated 338 tumors. Most limiting criterion has been proportion of tumor cells in samples; samples with more than 70 percent tumor cells are considered best candidates. Have now narrowed down to 12 tumors that are Her2 positive;
- In July and August, planned to obtain informed consent and extract blood DNAs;
- By end of the year, expect sequencing of three pairs of Her2 positive tumors and matched germline DNA.

Challenges

- In contrast to the case for HCC, the National Institute of Cancer of France had not already developed a collection of breast cancer tumors.

Multiple Subtypes of Breast Cancer (United Kingdom)

Design

- The ICGC breast cancer working group contains researchers throughout the world;
- Classifying breast cancers into four categories, based largely on estrogen receptor (ER), progesterone receptor (PR), and Her2 type: (1) Triple negative, (2) Her2 positive, (3) ER positive and Her2 negative (ductal type), (4) lobular and rare carcinomas;
- Expect to sequence 1,500 to 2,000 samples distributed among the four categories.

Progress and data collected

- International working group met at Sanger in June and plans to meet again in January 2010;
- Ethics and consent: Each group that is supplying samples has acquired permission from local IRBs. Re: coding of samples, a proposal to break the connection between patient IDs and samples was unanimously rejected because doing so would prevent subsequent addition of clinical information. However, it is unlikely that information derived from samples will be returned to patients;
- Pathology reviews are being conducted by one group across all ICGC breast cancer projects;
- Have access to 13,500 retrospectively collected primary breast cancer specimens (with normal controls), but many will be unsuitable for extensive analysis;
- Of 314 specimens that were considered promising, only half had at least 70 percent tumor nuclei, making them suitable for sequencing;
- Using paired-end read sequencing to detect rearrangements; 50 million reads per cancer (not high coverage);
- Genomic rearrangements: See different patterns in different tumors (and cell lines derived from tumors). Mostly see intrachromosomal rearrangements—an unexpected finding. Tandem duplications are very common in some types of breast tumors;
- In next 12 months, plan to sequence at least 10 breast cancer whole genomes, at least 30 whole exomes, and perform at least 30 rearrangement screens.

Challenges

- Not many specimens have enough tumor nuclei to be suitable for sequencing. However, pathologists warn that selecting specimens that are not infiltrated by normal cells will lead to the selection of distinct and not necessarily representative biological subsets.

Result

- An individual cancer genome is less likely to reveal new cancer genes as information about mutational processes, e.g., types of mutations that keep recurring. In COLO-829, C:G to T:A mutations were found to be common. This type of mutation is caused by ultraviolet (UV) light, a clue to the origin and etiology of the cancer.

Gastric Adenocarcinoma (China)

Design

- Have chosen to study gastric cancers because of their high incidence in China. Residents of some counties have particularly high risk. Seven hospitals are contributing samples;
- Gastric cancer project has secured funding; several other projects are under consideration in China;
- Have formed seven working groups (e.g., clinical and pathology issues, informed consent and study design) in parallel to ICGC working groups.

Progress and data collected

- Have looked at variations in protein expression patterns among specimens;
- Are focusing on tissue quality for RNA preparation;
- Have analyzed DNA copy number variations, correlating copy number changes with clinical characteristics, microRNA analysis.

Challenges and lessons learned

- Obtaining tumor samples is a rate-limiting factor. Obtaining uniform diagnosis is key;
- Obtaining good pathological and clinical data is key.

Chronic Lymphocytic Leukemia (CLL) (Spain)

Design

- Studying the most frequent leukemia in Western countries, a heterogeneous disease, with some known genetic subtypes (specifically, mutations in immunoglobulin gene lead to clinically distinct path, longer survival).

Progress and data collected

- *Specimens and consent*: Cryopreserved cells from more than 250 patients are available in a national CLL tumor biobank. ICGC researchers are asking patients to give additional consent because of sequencing and transport of samples. So far additional consent has been obtained for 20 specimens that fit ICGC criteria; those patients also donated more blood. Focus is on patients with advanced, active, untreated disease;
- Pathology criteria, SOPs, and research methods have been defined;
- Found misdiagnoses in past specimens based on gene expression profiling: recategorized immunoglobulin mutation status. With SNP arrays, checking for expected deletions;
- Performing DNA structural variation analysis with paired-end sequencing of long insert libraries;
- Performing exome sequencing with ICGC-defined exome;
- Have started whole-genome sequencing using two patients (one with immunoglobulin mutation and one not);
- Have established bioinformatics structure for sharing data within Spain and with Sanger.

Squamous Cell Carcinoma (SCC) of Oral Cavity (India)

Design

- Cancer of oral cavity is most common cancer in Indian males, varying widely in frequency in different cities. Gingivo-buccal cancer is common in India and more likely to metastasize than cancers of tongue and floor of mouth, which are more common in the West. Environmental risk factors for gingivo-buccal cancer are complex: areca nut, tobacco, lime. Some patients have no known environmental risk factors;
- Expect to be able to track genomic changes as lesions develop from pre-malignant conditions. Have a protocol to prospectively follow individual patients;
- Two to four pathologists will examine each sample.

Progress and data collected

- From August 2009 to July 2010, expect to collect 100 tumor samples and control tissues, and 400 samples in all to obtain 500 samples fulfilling all ICGC criteria;
- Collection, genomics, and scientific administration will be done at separate institutions in India. A governing body is in place and working groups have formed;
- *Specimen quality*: From 130 samples collected in a biorepository over last two years, randomly reexamined 32 cryopreserved samples and reconfirmed diagnosis of SCC. Half involved bone. Amount and quality of DNA was adequate. Specimens contain predominantly normal cells, but 4 of 32 specimens were not usable for various reasons (such as necrosis, no tumor present).

Challenges

- Tumors are more heterogeneous and variable in pathology than expected.

The Cancer Genome Atlas (TCGA) - Ovarian Cancer (USA)

Design

- Studying serous ovarian adenocarcinoma, which can be divided into two independent diseases: type I is low grade; type II is high grade and almost always fatal. TCGA is focusing on type II in which TP53 mutation rate is much higher. BRCA1 and BRCA2 are also more likely to be altered.

Progress and data collected

- Sample acquisition: collecting tumors for last two years and now have more than 300 confirmed high-quality specimens. By end of summer should have >600 specimens;
- Tothill data set (Clin Cancer Res, 2008) found six major classifications for all types of ovarian tumors. TCGA data supports the presence of three major subtypes of serous grade ovarian carcinoma;
- TCGA data will help validate whether the many genes previously found to be amplified in ovarian cancer play functional roles. For example, in certain amplicons, multiple genes are present, and only one may be driving cancer;
- In a pilot sequencing of 600 genes, every one of the first 25 tumors had a mutation in TP53;
- Preliminary data using sequencing and hybrid capture, see about one mutation per megabase. Cover about 80 percent of genes at 10X coverage with three lanes of sequencing on Illumina platform;
- Loss of heterozygosity (LOH) of BRCA1 and BRCA2 are common.

The Cancer Genome Atlas (TCGA) - Glioblastoma Multiforme (USA)

Design

- Three sequencing centers, seven characterization centers, one center for sample processing and management;
- Can use low coverage genome characterization to detect structural variations and then validate the importance of observed mutations by screening a larger cohort of patients through more intensive analysis.

Progress and data collected

- Have 657 matched samples in hand; 309 qualified by pathology and other metrics and sent for sequencing and other analysis;
- In phase 1, published in Nature in November 2008, characterized ~200 samples for transcriptome, genomic and epigenomic changes. Sequenced 91 cases at ~600 candidate genes. Found that certain pathways are very likely to have hits;
- Now addressing what was missed outside the list of candidate genes and outside exons, using hybrid capture and whole-genome sequencing;
- Each of three sequencing centers has taken on a tumor and normal specimen pair for sequencing at 30X coverage. Have so far found 40 mutations (37 previously not known). Mutations in genes from known core pathways. Very low coverage reveals structural variations;
- As a metric for coverage, can use heterozygous SNPs from arrays, validating that found known mutations;
- Co-amplification of chromosome 1 and 7 segments recurs;
- Continuing to perform targeted sequencing for up to 300 additional GBM cases and limited transcriptome sequencing and analysis on a relatively small number of cases (12 to start);
- Rick Wilson indicated that in addition to the TCGA studies described above, Washington University researchers plan to sequence 150 tumor and normal pairs in next year for other types of cancer—AML, breast, lung, ovarian, etc.—and five malignant glioma cases. This work is expected to develop methods, pipeline and analysis tools.

Pediatric Cancers (Germany)

Design

- Funding has come from an alliance between the federal ministry and German Cancer Aid, the largest charity in the country;
- Will define which pediatric tumors to study in coming weeks;
- Pediatric brain tumors are under consideration:
    - With the successful treatment of leukemias, brain tumors are the most common cause of childhood cancer death;
    - Among gliomas, most common are astrocytomas. Among nonglial tumors, most common are embryonal tumors (medulloblastoma and ependymomas);

    - Rationale for studying low-grade astroctyomas: Not many prognostic markers, highly variable rate of relapse, limited response to radio- and chemotherapy, unknown underlying genetics, except for BRAF gene. BRAF inhibition has helped treat one patient;
    - Rationale for medulloblastoma: heterogeneous disease, overtreatment is a common problem but not clear at outset who needs greater treatment; except for cases of genetic syndromes associated with medulloblastoma, no initiating mutations are known;
- Have both medulloblastoma and astrocytoma samples with associated clinical and molecular data. Expect to have 250 cases of each within a few years;
- Plan to perform high resolution, massive paired-end mapping;
- Plan collaboration among: Clinic and pathology, sequencing and analysis, complementary studies such as gene expression and DNA methylation, bioinformatics.

Discussion

Under- or unrepresented cancer types

- Incidence of different cancer varies widely among continents. Some cancers are very common and perhaps deserve study by the ICGC. Among those are cervical cancer and cancers of the gastrointestinal tract (GI);
- Some tumor types are difficult to study because little tissue is available, e.g., prostate cancer. May need improved technology;
- Researchers are reluctant to study tumor types that have good survival rates (such as testicular or bladder cancer), but studying them could shed light on biology. These may be tumors to study later when technology is more developed and cheaper;
- While the ICGC includes GI cancers, additional groups should feel free to pursue GI cancer. There can be more than one project per cancer type;
- No project is focusing on metastatic tumors exclusively. They are relatively easy to obtain compared with primary tumors; analysis of paired primary and metastatic tumors from the same patient provide an ideal source for identification of metastasis-specific genomic alterations.

International collaboration

- There are, so far, few collaborative international projects, except for the breast cancer project. There is interest in such collaboration and a need for funds;
- In study of pancreatic cancer, Canadian and Australian researchers have been discussing how to cooperate and ensure that data sets are comparable. Plan to visit each other's sites every six months, discuss QC and sharing data;
- Groups studying liver cancer are now discussing collaboration;
- Executive Committee (EXEC) will discuss how funding agencies can promote collaboration.

Technology development

- Consortium will share sequencing technology, but development of pathology technology may also be important.

Consistency across ICGC

- A small number of samples (two to three percent) should be sent among centers for validation;
- Technology working group should be able to impose a common suite of analytic steps for all data to ensure consistency in reporting mutations and their consequences; Analyses must be standardized to make data compatible among projects;
- Any recommendations should be flexible because groups are constantly innovating. For example, recommendations for one cell line may not apply to other cancer types.

Data access and consent

- No institutions appear to have had problems with Internal Review Boards (IRBs) restricting the sending of data out of country;
- Use of template consent form? Spanish group used most of the ICGC-developed template but removed questions that seemed tailored to American law.

III. Future Analysis Challenges

The Importance of Analysis Integration

The importance of informatics

- Studying cancer genomes will be at least as complex as other large-scale human genome projects, such as 1000 Genomes, HapMap and ENCODE. Involving people who can analyze the data in various ways will be important;
- Large-scale projects typically follow a technology shift, followed by successful pilot studies. Soon after projects achieve large-scale production scales and international coordination, need integration and analysis;
- *Production informatics* allow for the stable generation and aggregation of data. Analysis informatics is a separate process that allows for integration of data for biological interpretation;
- *Analysis informatics* includes assemblies, gene sets, and linking to biology (making interesting conclusions, e.g., in ENCODE project, that most of the genome is transcribed and that there are many more promoters than previously known).

How to use analysts

- For each of three production groups in ENCODE, there was an analyst. A separate group performed integration analysis;
- Recent genome projects (ENCODE, HapMap) have included explicit funding for analysis. It is important to recognize that analyzing data is not trivial. Analysts need a mixture of skills and connections to production groups. Need programming specialists across disciplines: Math/statistics, engineering and biology;
- Analysis and production need to be both coupled and independent. Analysts closely allied with production groups understand artifacts and are motivated to solve problems but feel attached to their data and can get overly focused on data rather than biology. Independent analysis groups tend to be embedded in a broader context and can make

more evenhanded assessments of datasets and platforms, but need to be regularly immersed in biology; they may be distracted by unnecessary method invention;

- Organizationally, there is a tension between anarchy and control. Leaving analysts independent may lead to trumpeting of artifacts. Controlling them can help in getting low-hanging fruit but analysts may miss important features.

Concluding advice

- The genome community can give concrete recommendations for how to build an analysis bioinformatics program;
- Analysis bioinformatics differs from production bioinformatics and data federation but requires those things;
- ICGC can leverage the technical efforts of ENCODE, 1000 Genomes, probably by cross-pollination of people;
- ICGC should consider funding 15 to 20 analysts.

Discussion

Is ICGC ready to set up an analysis group?

- Nothing this complex has been done before. It is difficult to estimate how hard it will be and how much it will cost;
- The "audience" for this project is a very large cancer research community and the goal is to enable their research. How can this project be as useful as possible for the community of cancer researchers?
    - There have to be layers of access;
    - Cancer researchers should be encouraged to have mathematicians on hand as part of their research projects;
- Should analysis focus on biology or simply on generating a high-quality dataset? There was argument for both;
- Cancer research community will not be able to interpret the data, so it is ICGC's responsibility to analyze it;
- ICGC includes people with medical and biological expertise. There is an opportunity to integrate with analysts from the outset rather than to count on analysis to happen *ad hoc*.

IV. Potential Collaborative Projects on Pediatric Rare Tumors

Confounding issues in pediatrics

- Children typically receive chemotherapy before surgery;
- Biopsies are typically very small;
- Consent is through parents; researchers may need to obtain consent again as children reach age of majority (retroactive consent); assent needs to be obtained from child;
- Survival rates can vary widely in what seem like the same disease (e.g., for patients with acute lymphoblastic leukemia (ALL), outcomes tend to be good but are much worse if Philadelphia chromosome is involved);

- Expression patterns of different genes may vary with normal stages of infant-child development;
- Gene expression varies naturally as children progress toward adulthood.

Potential collaborative projects include:

Neuroblastoma (USA)

The disease

- Causes about 15 percent of childhood cancer mortality. Highest rate of spontaneous remission, but about half of cases present with explosive metastatic disease;
- Cure rate over last two decades has been stagnant.

Infrastructure

- In North America, the Children's Oncology Group (COG) has performed centralized collection of biospecimens (about 650 cases per year), collecting rich phenotype information;
- Also: NBL-GWAS, a genome-wide association study at the Children's Hospital of Philadelphia, and NBL-TARGET, a collaborative genomic project to discover genes that could be targeted clinically. These are NCI-funded studies. NBL-TARGET future plans include epigenetic studies (methylation and miRNA);
- There is an existing history of international collaboration among neuroblastoma researchers, with regular meetings, including a biennial meeting on translational genomics in neuroblastoma and extensive biobanking and genomics efforts.

Future steps

- Sample sets are in hand for complete resequencing of neuroblastoma genome;
- Focus on refractory neuroblastoma genome: prevalence of mutations is higher in cases where there is relapse;
- Focus on translating discoveries into therapies (anaplastic lymphoma kinase (ALK) gene an example).

Pediatric Lymphoma and Leukemia (USA)

Infrastructure

- NCI launched TARGET (Therapeutically Applicable Research to Generate Effective Treatments) projects after TCGA. TARGET collects data in a centralized database with access at two levels: public and protected;
- TARGET has ongoing programs for ALL and neuroblastoma;
- TARGET is being expanded under the 2009 economic stimulus plan ARRA (American Recovery and Reinvestment Act of) funding to cover new tumors.

Findings

- For ALL, alterations in gene IKZF1 (or IKAROS) predict a new class of childhood ALL with high risk for relapse. Small molecules are being developed as drugs.
- Mutations in janus family of kinases and IKZF1 alternations result in disease recurrence in 84 percent of the cases

Osteosarcoma (UK)

The disease

- Most common primary malignant bone tumor; tends to be a disease of teenagers and young adults;
- Stagnant survival rates over last two decades;
- There are many available specimens with normal paired tissue; samples tend to be large;
- However, most samples are obtained post-chemotherapy. The bias is toward specimens from patients who responded poorly to chemotherapy. To obtain untreated specimens, may be able to use very small samples (200 ng DNA) from needle biopsies;
- Tumor specimens tend to have little normal cell contamination;
- Metastases (particulary lung) lend themselves to analysis as 'cherry-picking' of lung mets (even in the setting of recurrence) is an established therapeutic approach.

Infrastructure

- In Europe, EuroBoNeT is banking tumors from a number of countries. About 20 percent of samples include pre-chemo needle biopsies;
- Many countries have sample collections;
- Adrienne Flanagan's group in the UK has started to generate xenografts;
- Osteosarcoma is rare, so it is difficult for any one group to obtain large amounts of funding. Collaboration would be valuable;
- COG is banking osteosarcomas from all centres (approximately 600/year) along with paired blood specimens.

Soft Tissue Sarcoma (France)

The disease

- Heterogeneous, occurring in any site of the body. Often large tumors;
- Some subtypes respond to targeted treatment;
- Connective tissue tumors have molecular subtypes associated with specific translocations;
- Treatment has changed dramatically over the last decade: Many more treatment agents are available. Where agents were developed with knowledge of the molecular basis of the cancer, median survival has improved from one year to more than four years.

Infrastructure

- A network in Europe for connective tissue cancer, CONTICANET, has nearly 5,000 patients and tumors collected in a virtual tumor bank (www.conticabase.org), all reviewed by expert pathologists;
- Active participants include 20 centers in France as well as centers in Germany, Italy, and Belgium;
- There are sarcoma groups around the world. There are plans to develop a World Sarcoma Network, bringing the groups together;
- Informed consent is done locally, not centralized.

Wilms' Tumor (UK)

The disease

- Rare childhood cancer of the kidney; very good survival rates;
- Treatment varies: In the USA, nephrectomy precedes chemotherapy; in Europe, the reverse occurs;
- Genetics: Prognostic markers include genomic events (1q and 16q loss). Known germline susceptibility through WT1 gene. There are known somatic changes in genes. Cases of Wilms' tumor are not entirely homogeneous;
- In tissue specimens, there are usually a high proportion of tumor cells.

Infrastructure

- Most patients with Wilms' Tumor are involved in clinical trials;
- A number of groups have collected frozen samples over the years;
- There is a tradition of collaboration among researchers, as with other childhood cancers.

Future steps

- Studying about 200 cases of Wilms' tumor and its variants may be sufficient.

Discussion

- Communities appear to be organized around each of these tumor types. There is no need to start from scratch;
- Key question is how much to do prospectively, vs. staying in existing collaborations;
- How many whole genomes are targeted for each disease in TARGET?
    - In case of neuroblastoma, would like to do 100 whole genomes, but amount of future funding is not clear.
- Could two or three groups analyze 100 or more genomes each in collaboration? Funders could promote integration upfront;
- Specimens collected immediately post-chemotherapy are often necrotic;
- Cancer stem cells may be a future target, if researchers can isolate and analyze a very small number of these cells;

- Relapse tumor specimens have separate issues. Cancer may have acquired new mutations. This is a complication, but does not necessarily exclude analysis. Researchers can still attempt to isolate driver mutations.

Working Group Reports

I. Consent and Data Access Policies Working Group

Participation in working group

- This working group would like to have representatives from China and India;
- Participants can be scientists rather than specialists in ethical and legal issues.

Motivation

- Consent procedures should look ahead to when data will be shared internationally. To prevent future delays in research, basic elements for international collaboration should be present in consent forms.

Status of model forms

- *Model consent form*: Working group created a sample prospective consent form, which it attempted to make simpler and clearer in its breakout session. All ICGC participants should receive a committee-approved model form about a month after the workshop. Researchers in the field can add or subtract elements. Using the form is not required as long as research procedures are in concordance with ICGC policies;
- *Retrospective consent form*: For existing cohorts, committee is preparing a retrospective form for obtaining reconsent from living individuals. This form should be ready by the March meeting;
- *Data access form*: Awaiting feedback from the Data Coordination and Management committee.

Role of data access compliance office and working groups

- The ICGC Data Access Compliance Office (DACO) will be situated in McGill University;
- DACO will not second-guess university ethics review. It will ask for proof that an official at each institution has considered or performed an ethics review;
- Three levels of review:
    - Researcher states that no ethics review is required and why;
    - Institution waives ethics review;
    - Ethics review occurs.
- The International Data Access Committee (IDAC) will address situations where clarification is required.

Future directions

- Issues group plans to study in the coming year;
    - Return of results: When legal and ethical constraints may require return of results to patients;

    - Issues in longitudinal studies of pediatric cohorts;
    - Timing of consent: Is consent authentic when it is given during trying times of illness? Is consent authentic when given ahead of time?
- *Future of ICGC*
    - ICGC should consider a communication strategy, creating a portal not just for researchers to share results and collaborate but also for the public and the media, saying what the ICGC is about. Researchers should give 150-word lay summaries of their research for posting.

Discussion

- More countries should be represented on the consent and data access policies committee. It is important to take into account local as well as international context. Language used should be sufficiently broad to speak to a range of cultures;
- Draft data release policies are not yet written though the draft data access application form says that signers are aware of the policies:
    - Data release policies should be communicated about a month after this meeting and resolved about two to three months after this meeting;
    - Data release policies should promote early data release while protecting data producers so that they can be the first to write global analyses. Studies on mutations by outsiders may be allowed before global analyses are done;
    - By early fall, data should start being released to the public. Before that, data will be available among ICGC members. By the fall, policies should be in place.
- The icgc.org web site is currently underpopulated. ICGC members should suggest content to Jennifer Jennings jennifer.jennings[at]icgc.org to build up this portal.

- *Return of results to patients*
    - If returning relevant information to patients, what responsibility will researchers have to look for relevant information? That would be an extra load;
    - Researchers cannot act as if they have a therapeutic relationship with patients and cannot foster “the therapeutic misconception” in patients who are actively being treated for cancer. To induce patients to participate with the promise of therapeutic benefit risks dishonesty;
    - Returning data to patients may be especially difficult if data are encrypted;
    - Public bulletins could be used, giving researchers an open-ended duty to reveal results over time;
    - If there is a misdiagnosis or other finding of clinical significance, there may be a responsibility to give feedback to patients, perhaps in a time-limited way;
    - It is impractical to expect researchers to look for and evaluate problems unrelated to their own expertise;
    - ICGC should have a clear position on this issue, but it can be very helpful to leave door open for possibly contacting patients;

    - A committee participant would like to gather data about outcomes in countries, such as Spain, Japan, and France, that have laws and policies on return of results: How often does situation arise? What procedures have been used? What problems have occurred?
    - While ICGC participants should not be responsible for scouring genomes of patients for any possible insight, it is a different matter if researchers come across unexpected, clinically important information that they cannot ignore.
- Data access process is a barrier. Each project should enable outside researchers to be able to do as much as possible without having to access controlled data, e.g., by providing somatic mutation tables, high-level analyses, good clinical information about, e.g., survival, age;
- Consent and data access forms are meant to protect patients, but policies must be simple, understandable, and workable;
- The consent and data access working group welcomes researchers' questions and suggestions.

II. Tissue and Clinical Annotation Working Group

This working group has two subgroups.

Tasks, progress, recommendations

- Clinical annotation subgroup:
    - Task: To develop unified guidelines for clinical annotation and for the use of software applications for annotation and metadata analysis;
    - In establishing data categories for clinical history:
        - Precise definitions are needed for accurate submission of data;
        - Fewer fields should be required than proposed so far;
        - Diagnosis-specific data should be separated from common data;
        - Additional committee review is required before finalizing a system of annotation.
    - In data definition process:
        - ICGC should harmonize with CTRNet using recognized sources, such as caBIG, WHO, etc;
        - The working group will need the contribution of ICGC subject-matter experts;
    - Software recommendation:
        - There are several good alternatives for local software solutions;
        - CTRNet has offered use of its tumor-banking suite (ATiM), and initial view of working group is that ATiM will meet needs. CTRNet will schedule a webinar to demonstrate ATiM. Others will be invited to do same.
- Tissue sample quality subgroup:
    - Task: To develop unified guidelines for specimen annotation and quality assurance/quality control;

    - Have proposed to modify the third of the four policies on quality standards of specimens listed in the April 2008 ICGC guidelines document:
        - Existing policy: All samples should be reviewed by two or more reference pathologists using sections from the same piece of tissue;
        - Proposed change: Histological examination must be documented and optical images must be stored and made available to those studying results from the sample. Virtual slides are recommended.
- *Proposed change*: Tumor and control (germline) samples do not always match. Genotyping of both should be mandatory for all cases preceding extensive sequencing. (Using a genotyping chip for this is also a good way to check on the accuracy of future sequencing).
- *Enforcement:*
    - About one to five percent of tumors should be sent to an independent site, such as another ICGC project, for the isolation of DNA and RNA;
    - Isolated analytes, especially DNA, should be send to an independent site for quality assurance;
    - Tissue specimens should be examined by an independent reference pathologist at a different location, if possible.
- The group communicates with other working groups: technology, data coordination and management;
- DCC catalogue data feedback: Have reached consensus on data elements and which should be mandatory or optional.

To be done

- Clinical annotation subgroup:
    - Complete clinical annotation guidelines;
    - Make recommendation for use of software for annotations and metadata analysis.
- Tissue sample quality subgroup:
    - Establish further guidelines for QA/QC, including QC issues for xenografts;
    - Review existing QA/QC procedures among ICGC projects for insights.

Discussion

- Are any of the recommendations controversial or difficult?
    - For sample processing, no;
    - For clinical data, there is already an installed base of definitions and changing definitions people use will be challenging;
- There is an opportunity to change the consent forms biobanks use as their standard to add a statement that in the future, specimens may be subject to genome sequencing;
- DCM working group: How much clinical history should be captured;
- There should not be many categories of clinical information but they should be precise, e.g., survival period and tumor-specific;

- Should patient treatment course be recorded? Since many projects hope to find patients who have been as homogeneously treated as possible, information is helpful, but perhaps should not be required;
- Required information should be minimal;
- Re: proposed modification to 3rd policy on specimen quality: Should digitized images be treated as controlled data?
    - Slides will not be individually identifiable and files will be large;
    - Consensus to keep images public.

III. Technologies Working Group

- At breakout session, working group revisited issues discussed in recent weeks to confirm agreement;
- ICGC members are welcome to see copies of minutes or to participate in calls of subgroups;
- Libraries and variant calling subgroups in particular invite additional participants.

Progress and recommendations

I. Validation subgroup
- A high quality threshold should be set for the catalogue of somatic variants from each cancer genome;
- Some in group say standard should be higher than the one currently set of detecting 90 percent of somatic variants, with no more than five percent of those being false positives;
- Every cancer genome will need validation in the short and medium term;
- As the cost of sequencing drops, validation may take an increasingly large proportion of effort. It may be possible to do less validation or apply it to only to a subset of cancers.

II. Common sample subgroup
- It is useful for individual centers to be able to evaluate their protocols and to analyze their pipelines before generating data in large scales by using one or more well-studied, common samples with inexhaustible amounts of DNA and RNA;
- COLO-829, a malignant melanoma cell line sequenced at Sanger is available. A good amount of data is available about this cell line;
- Sanger researchers validated efficacy of Illumina for finding known mutations in this cell line;
- Cell line has more than 30,000 somatic mutations (5,000 is considered average), including 272 mutations in protein-coding regions, and 171 nonsynonymous mutations.

III. Libraries subgroup: This group will continue to meet.
- Most group members prefer libraries with a mixture of short (<500 bp) and long (1–3 kb or greater) insert sizes. Preferences for different insert sizes vary; it will be useful for this subgroup to continue meeting;
- Agree on at least 30X sequence coverage of both normal and tumor;

- No decision on optimal read lengths. Agreement to wait for data;
- Agree that need to periodically reevaluate recommendations based on data and polls of what research groups are doing;
- Rather than mapping to a reference genome, is preassembly of a cancer genome possible? Group has not yet reached a conclusion on the value of preassembly.

IV. Variant calling subgroup: This group will also continue to meet.
- Each class of variants needs the development of its own algorithm;
- Many algorithms are in development;
- Many problems and artifacts occur, some of which depend on platform;
- Variant calling will have to be assessed for each individual cancer;
- Calling algorithms will have to be assessed as they are developed;
- Standards for deposition of data into ICGC databases need to be established: Should all data be entered, or do data have to meet minimum standards, as in the Human Genome Project?

V. Exome subgroup
- An exome has been designed and submitted to providers;
- This needs to be checked and coordinated with other exome designs; designs should match;
- Further interaction with providers can lead to benefits from scale, reducing costs.

VI. Epigenomics subgroup
- Assessment of MeDIP, MeCAP and bisulphite sequencing approaches is in progress;
- Design of a CpG chip for ICGC groups: Illumina has expressed interest in designing such a chip. ICGC has to decide if it wants to commission one. The chip would need to be used by groups outside ICGC to justify the cost of generating it;
- Monitoring epigenomics approaches.

VII. Transcriptome technology subgroup
- Currently scoping key issues likely to affect interpretation and integration of NGS (Next Generation Sequencing) transcriptome data;
- Also considering issues related to sample preparation, methodologies, and analysis.

Discussion

- Are there any issues that the technologies working group should be considering that are not yet being considered? Mike Stratton invited e-mails.
- Proposed CpG chip
  - Will cover CpGs outside islands as well as in them; complementary to enrichment methods;
  - Will ask Illumina to avoid CpGs in repeats and to include variable sites determined by various experts. Want to annotate a table for 20 million CpGs and to select perhaps a million for a chip;

- Assay will be quantitative. Chips would not be complete but a useful tool and would provide a standard for comparison.

IV. Data Coordination and Management Working Group

Progress in developing franchise database

Design

- Each ICGC genome characterization center will have a local database, loaded periodically and integrating each site's own data with several standard data sets;
- DCC portal will allow for queries across all databases. The portal will allow for Google-type searches of genes, patients, SNPs, CNVs, structural variations. A user can choose filters, e.g., chromosome number, type of variation, gene pathway. Results can be shown as text or graphics.

Deployment timeline

- In June, a prototype was populated with synthetic data;
  In the month after meeting, DCC planned to import melanoma data and TCGA data sets;
- By mid to late August, alpha release is planned for testing by interested centers. Following these initial tests, data model might be changed dramatically;
- By end of 2009, intend to have beta release in all centers;
- Hope for Release 1.0 by next workshop in March 2010.

Issues considered in breakout session

- Walked through data model field by field. Noted need for certain fixes;
- How to relate raw data to interpreted data to preserve the chain of evidence.

- Data exists in various levels, each stored in a different place:
  - Level 1: raw data (reads from machines); stored in archival databases, mostly in European Bioinformatics Institute (EBI) genotype archive;
  - Level 2: minimally processed data (alignments); stored in genome centers locally as a short-term solution;
  - Level 3: interpreted data (mutations and their consequences); stored in franchise database;
  - In addition, written protocols will be collected in a Wiki at DCC.
- How to archive sequence reads and alignments for community access:
  - For raw reads, recommend submitting to EGA (at EBI), which has agreed to accept DACO/DCC authorization procedures for access to controlled data; there will be no need for users to obtain an additional account with EGA. May add dbGAP (at NCBI) pending a similar reciprocal agreement;

- For alignments, important to maintain evidence chain. EGA/NCBI will accept alignments as a pilot project starting this summer. Alignments will be stored with these groups if they decide to accept alignments regularly in the future. Otherwise, each genome center will be asked to maintain alignments on a Web-accessible, password-protected server. Because of storage and bandwidth implications, if no other solution emerges, data may move to DCC.

- How to manage genome builds over the lifetime of the project:
  - The most recent build of the human reference genome is H37. ICGC members should align sequences relative to H37;
  - The data management working group feels that alignments and interpreted data need to be updated regularly but not too frequently, perhaps once every three years.

Unresolved issues

- Pathology images:
  - Would like to make accessible to all;
  - Where will the images reside? DCC can ask each center to place images on a community-accessible File Transfer Protocol (FTP) server. This is consistent with a distributed franchise database model, but there are IT implications.
- Clinical and histopathological data model:
  - Need guidance from clinical and tissues working group on fields to use, e.g., level of granularity in clinical history.

Security procedures

- Open vs. protected data: ICGC policy is binary, with data either publicly accessible or controlled. A user with access to controlled data has access to all data;
- Model for authentication for use of controlled data: Plan to use OpenID instead of setting up ICGC-specific accounts. Users can obtain OpenIDs from a large number of providers (e.g., VeriSign is one) and use them to access a range of accounts (e.g.,Gmail);
- Approval procedure: Bundle of material transfer agreements and certifications goes to DACO. DACO approves and sends list of OpenIDs to DCC;
- Authentication process: User sends OpenID to database by logging in. ICGC database interacts behind scenes with user through OpenID provider (such as VeriSign). Any of the affiliated databases in the franchise database can check OpenIDs against authorization database.

Unresolved security issues

- Authorization expires after one year, but students or postdocs who leave lab should be deauthorized more quickly. Should PIs have to periodically confirm that all authorized staff are still present, or should PIs be allowed to remove staff at their own discretion via a web form? To add staff, PI should file a request with the DACO;

    - There are dozens of OpenID providers (e.g., Yahoo, Google, Flicker, VeriSign, AOL and others). Must decide which ones to trust.
- The data management working group encourages participation from others. People can either join as formal members or drop in on teleconferences.

Discussion

Sharing data

- Until DCC is fully operational, how should data be shared among Consortium members?
- The alpha test version will be available by the end of summer; will only include open data, not protected data. Alpha version is intended to be fully functional. If it does not work, an FTP site will be set up in Ontario for sharing data.

Genome versions

- In new versions of human genome, tend to add new areas rather than have considerable realignments of previously aligned sections;
- Is updating every three years too infrequent? Perhaps ICGC data should be updated when there is a new complete build of the human genome. This is likely to be every two or three years.

Archiving sequence reads

- Is it a real risk that short-read archives will run out of space? No, as long as there is agreement on how to condense information on level of intensities into quality assessment scores. Few people use intensity data; the number who use base calls and quality scores is much higher. However, with cancer specimens, samples get used up, so perhaps ICGC participants should keep as much information as possible;
- Storing one called base plus four qualities (one for each base) may be easiest. Done this way, the cost of archiving is now two percent of the cost of sequencing;
- Beginning to archive earlier could save space.

Privacy issues

- Data about germline polymorphisms will be in controlled tier;
- Should assume that information provided on public side can be used to identify individuals more readily than many imagine. What type of investigators will examine public data without going to protected side of data?
- Aggregate somatic genetic data is expected to be open. From aggregate allele frequency data, it is hard to identify individual data, but risk is not zero. However, there is also a risk to the research community if protect more data;
- NIH and Wellcome Trust have moved aggregate germline genetic data to controlled access pending further consideration;
- Perhaps only PI-level people should have access to controlled data. IRBs may be cautious about providing access to students without supervision. Perhaps bar to access should be high. However, PIs will typically immediately pass along task to a student or postdoc. Given that PIs will share passwords, ICGC members had decided that it is better to recognize the real people who will be using the data. This will permit better tracking of any breaches;

- As always a balance must be struck between enabling research and protecting individual privacy. Ultimately, some level of trust is required;
- It would be helpful to have a communication plan to convey how data may be used;
- Need an action plan for what to do if something goes wrong.

V. Moving Forward

Development of ICGC as a Consortium

The ICGC has become a functional consortium for sharing ideas and expertise, such as how to acquire samples, generate data and deal with data. The Consortium is also moving toward coordination and collaboration. Funders are supporting investigators, counting on the idea that the ICGC will allow for coordination.

Funding and Projects

At the launch of the ICGC in 2007, there was one funded project. By the first workshop In November 2008, there were eight projects. Now, there are 10 projects with three or four more in the offing. Together, these do not yet fulfill the ICGC mission of covering a broad range of human cancer types. Discussions at this workshop on rare and childhood cancers were fruitful; pulling sample collections together is a worthwhile approach.

The Work of Generating Cancer Genomes

- *Ethics:* There has been progress;
- *Samples:* Collection is underway. Groups are encountering expected problems; many sample collections are not of adequate quality. Prospective collections are moving forward quickly;
- *Pathology:* QC is in place with at least two pathologists looking at each case introduced into a project;
- *Data management:* Procedures are in development and the Consortium is working toward an acceptable model;
- *Data:* The credibility of the Consortium will depend on the quality of data generated on the scale that has been promised.

The challenge before the next workshop in Madrid is to generate a large amount of data and make it publicly viewable and accessible to cancer researchers.

Data analysis issue remains a challenge and a matter for further discussion. Each team should have informaticists. There should also be an overarching analysis group. Beyond providing coordination and tools, ICGC participants should discuss whether the Consortium should do more to analyze data across cancer types—to analyze all the data collectively. Some form of overarching data analysis should be discussed over the coming months and years as participants generate large amounts of data.

Next Workshop

The third ICGC scientific workshop is scheduled for March 21–23, 2010 in Madrid, Spain.

APPENDIX 1

WORKSHOP AGENDA

Second ICGC Scientific Workshop

June 22–24, 2009
Hinxton, UK

Monday, June 22		
6 and 7 p.m.	Transfers from Homerton College to Wellcome Trust Conference Centre	Departs from outside Homerton College on Hills Road
6:30 – 9 p.m.	Registration and Reception Wellcome Trust Conference Centre Wellcome Trust Genome Campus Hinxton, Cambridge Tel 01223 495000	Conference centre foyer
9 p.m.	Transfer from Wellcome Trust Conference Centre to Homerton College	Departs from the conference centre entrance
Tuesday, June 23		
7:30 – 8 a.m.	Breakfast	Homerton College
8:15 a.m.	Transfer from Homerton College to Wellcome Trust Conference Centre	Departs from Outside Homerton College on Hills Rd
8:30 – 9 a.m.	Registration and Poster Set Up	Conference centre foyer
Welcome and Purpose of Meeting		The Francis Crick Auditorium
9– 9:07 a.m.	Welcome from Conveners	Mike Stratton (UK)
9:08 – 9:15 a.m.	Opening Remarks	Mark Walport (UK)
9:15 – 9:40 a.m.	An Overview and Update on the Consortium • Progress of Consortium as a whole and purpose of the meeting	Tom Hudson (Canada)
Status of ICGC Projects		
9:40 a.m. – 2:30 p.m.	Moderator • Participants in this session will have between 5 and 12 minutes to present the status of each cancer genome project: design, progress and data collected.	Fabien Calvo (France)

	Pancreatic adenocarcinoma	John McPherson (Canada)
	Pancreatic adenocarcinoma	Sean Grimmond (Australia)
	Morning coffee	Conference centre foyer
	Comprehensive genomic analysis of viral hepatitis-associated HCC	Tatsuhiro Shibata (Japan)
	Her2 positive breast cancer	Gilles Thomas (France)
Status of ICGC Projects, Continued		
	Hepatocellular carcinoma (HCC) (alcohol-related)	Jessica Zucman-Rossi (France) & Ivo Gut (France)
	Breast (multiple subtypes)	Mike Stratton (UK)
	Gastric adenocarcinoma	Youyong Lu (China)
	Chronic lymphocytic leukemia	Elias Campo (Spain)
	Squamous cell carcinoma (oral cavity)	Partha Majumder (India) & Rajiv Sarin (India)
	Lunch and group photo	Hall restaurant
	TCGA (ovarian - serous cystadenocarcinoma)	Paul Spellman (USA)
	TCGA (glioblastoma multiforme)	Rick Wilson (USA)
	Pediatric (subtypes)	Peter Lichter (Germany)
	Panel Discussion (led by Moderator)	
Future Analysis Challenges		
2:30 – 2:45 p.m.	The importance of analysis integration	Ewan Birney (UK)
2:45 – 3 p.m.	Panel Discussion (led by Moderator)	Alfonso Valencia (Spain)
Collaborative ICGC Projects		
3 – 3:30 p.m.	Collaborative Projects for Pediatric/Rare Tumors Potential collaborations include projects on: • Neuroblastoma • Pediatric lymphoma and leukemia	David Malkin (Canada) John Maris (USA) Daniela Gerhard (USA)

	• Osteosarcoma	Adrienne Flanagan (UK)
	• Soft tissue sarcoma	Jean-Yves Blay (France)
	• Wilms' tumor	Mike Stratton (UK)
3:30 – 4 p.m.	Panel Discussion (led by Moderator)	
4 – 4:30 p.m.	Afternoon tea	Conference centre foyer
Breakout Sessions		
Consent and Data Access Policies Working Group		The Library
4:30 – 7 p.m.	Discussion (led by Working Group Leader)	Bartha Knoppers (Canada)
Tissue and Clinical Annotation Working Group		Loft Room 2
4:30 – 7 p.m.	Discussion (led by Working Group Leader)	Peter Lichter (Germany)
Technologies Working Group		The Pompeiian Room
4:30 – 7 p.m.	Discussion (led by Working Group Leader)	Mike Stratton (UK)
Data Coordination and Management Working Group		The Green Room
4:30 – 7 p.m.	Discussion (led by Working Group Leader)	Lincoln Stein (Canada)
Meeting for Funders/EXEC		The Tennis Court Room
5:30 – 7 p.m.	Discussion (led by Moderator)	Tom Hudson (Canada)
7 – 7:30 p.m.	Drinks reception	Cloisters
7:30 – 9 p.m.	Dinner*	Hall restaurant
9:15 p.m.	Transfer from Wellcome Trust Conference Centre to Homerton College	Departs from the conference centre entrance

*A 7:30 bus will be prearranged for attendees wanting to skip dinner.

7:30 p.m.	Transfer from Wellcome Trust Conference Centre to Homerton College	Departs from the conference centre entrance

Wednesday, June 24		
7 – 7:30 a.m.	Breakfast	Homerton College
7:30 a.m.	Transfer from Homerton College to Wellcome Trust Conference Centre	Departs from Outside Homerton College on Hills Rd
Report Back from Working Group Leaders		The Francis Crick Auditorium
8 – 8:30 a.m.	Consent and Data Access Policies	Bartha Knoppers (Canada)
8:30 – 8:45 a.m.	Panel Discussion (led by Working Group Leader)	
8:45 – 9:15 a.m.	Tissue and Clinical Annotation	Peter Lichter (Germany)
9:15 – 9:30 a.m.	Panel Discussion (led by Working Group Leader)	
9:30 – 10 a.m.	Morning coffee	Conference centre foyer
10 – 10:40 a.m.	Technologies	Mike Stratton (UK)
10:40 – 11 a.m.	Panel Discussion (led by Working Group Leader)	
11 – 11:30 a.m.	Data Coordination and Management	Lincoln Stein (Canada)
11:30 – 11:45 a.m.	Panel Discussion (led by Working Group Leader)	
Wrap-Up Session – Moving Forward		
11:45 a.m. – 12 p.m.	Summary Speaker	Mike Stratton (UK)
12 p.m.	Meeting adjourns	
12 – 1:30 p.m.	Lunch	Hall restaurant
1:30 p.m.	Return transfers to Stansted & Heathrow Airport & Cambridge train station	Departs from the conference centre entrance

APPENDIX II

PARTICIPANT LIST

Second ICGC Scientific Workshop

June 22–24, 2009
Hinxton, UK

First Name	Last Name	Affiliation	Country	Email
Andrew	Biankin	Garvan Institute	Australia	a.biankin@garvan.org.au
David	Bowtell	Peter MacCallum Cancer Centre	Australia	David.Bowtell@petermac.org
Sean	Grimmond	Institute for Molecular Bioscience, University of Queensland	Australia	s.grimmond@imb.uq.edu.au
Terry	Speed	The Walter And Eliza Hall Institute of Medical Research	Australia	terry@wehi.edu.au
Nikolajs	Zeps	Government of Western Australia, Department of Health	Australia	nik.zeps@sjog.org.au
Samuel	Aparicio	BC Cancer Research Centre, BC Cancer Agency	Canada	saparicio@bccrc.ca
Cindy	Bell	Genome Canada	Canada	cbell@genomecanada.ca
Peter	Geary	Canadian Tumour Repository Network (CTRNet)	Canada	Peter.Geary@CTRNet.ca
Martin	Hirst	Genome Sciences Centre, BC Cancer Agency	Canada	mhirst@bcgsc.ca
Tom	Hudson	Ontario Institute for Cancer Research	Canada	tom.hudson@oicr.on.ca
Jennifer	Jennings	Ontario Institute for Cancer Research	Canada	jennifer.jennings@oicr.on.ca
Yann	Joly	Université de Montréal	Canada	yann.joly@umontreal.ca
Arek	Kasprzyk	Ontario Institute for Cancer Research	Canada	arek.kasprzyk@oicr.on.ca
Bartha	Knoppers	Université de Montreal	Canada	bartha.maria.knoppers@umontreal.ca
David	Malkin	The Hospital for Sick Children	Canada	david.malkin@sickkids.ca
John	McPherson	Ontario Institute for Cancer Research	Canada	john.mcpherson@oicr.on.ca
Francis	Ouellette	Ontario Institute for Cancer Research	Canada	francis@oicr.on.ca

First Name	Last Name	Affiliation	Country	Email
Lincoln	Stein	Ontario Institute for Cancer Research	Canada	lincoln.stein@gmail.com
Susan	Wallace	Université de Montréal	Canada	susan.elizabeth.wallace@umontreal.ca
Jingxiang	Li	Beijing Genomics Institute, Shenzhen	China	lijx@genomics.org.cn
Youyong	Lu	Peking University	China	yongylu@public.bta.net.cn; youyonglu@sina.com
Hong	Xue	Applied Genomics Center, Hong Kong University of Science and Technology	China	hxue@ust.hk
Xiuqing	Zhang	Beijing Genomics Institute, Shenzhen	China	zhangxq@genomics.org.cn
Iiro	Eerola	European Commission	European Union	iiro.eerola@ec.europa.eu
Jean-Yves	Blay	Hôpital Edouard-Herriot	France	blay@lyon.fnclcc.fr
Fabien	Calvo	Institut National du Cancer (INCa)	France	fcalvo@institutcancer.fr
Christian	Chabannon	Institut Paoli-Calmettes	France	chabannonc@marseille.fnclcc.fr
Bruno	Clément	Inserm U620	France	bruno.clement@rennes.inserm.fr
Olivier	Delattre	Institut Curie	France	olivier.delattre@curie.fr
Ivo	Gut	Centre National de Génotypage	France	ivo.gut@cng.fr
Marta	Gut	Centre National de Génotypage	France	mgut@cng.fr
Simon	Heath	Centre National de Génotypage	France	heath@cng.fr
Emmanuelle	Rial-Sebbag	Institut national de la santé et de la recherche médicale (INSERM)	France	rialseb@cict.fr
Gilles	Thomas	Fondation Synergie	France	gd.thomas@me.com
Jessica	Zucman-Rossi	Fondation Jean Dausset - CEPH	France	zucman@cephb.fr
Roland	Eils	German Cancer Research Center	Germany	r.eils@dkfz-heidelberg.de
Franz	Kohlhuber	German Cancer Aid	Germany	kohlhuber@krebshilfe.de
Hans	Lehrach	Max Planck Institute for Molecular Genetics	Germany	lehrach@molgen.mpg.de
Peter	Lichter	German Cancer Research Centre	Germany	Peter.Lichter@Dkfz-Heidelberg.de

First Name	Last Name	Affiliation	Country	Email
Michal	Schweiger	Max Planck Institute for Molecular Genetics	Germany	mschweig@molgen.mpg.de
Holger	Sültmann	German Cancer Research Center	Germany	h.sueltmann@dkfz.de
Stefan	Wiemann	German Cancer Research Center	Germany	s.wiemann@dkfz.de
Partha	Majumder	Indian Statistical Institute	India	ppm@isical.ac.in
T.S.	Rao	Ministry of Science & Technology, Dept. of Biotechnology	India	tsrao.dbt@nic.in
Rajiv	Sarin	Advanced Centre for Treatment, Research and Education in Cancer (ACTREC)	India	rsarin@actrec.gov.in
Hiroyuki	Aburatani	RCAST, University of Tokyo	Japan	haburata-tky@umin.ac.jp
Kazuto	Kato	Kyoto University	Japan	kato@zinbun.kyoto-u.ac.jp
Jun	Kusuda	NiBio	Japan	jkusuda@nibio.go.jp
Hidewaki	Nakagawa	RIKEN	Japan	hidewaki@ims.u-tokyo.ac.jp
Tatsuhiro	Shibata	NCC, Japan	Japan	tashibat@ncc.go.jp
Koichi	Watanabe	RIKEN	Japan	koichi.watanabe@riken.jp
Teruhiko	Yoshida	NCC, Japan	Japan	tyoshida@ncc.go.jp
Henk	Stunnenberg	Nijmegen Centre for Molecular Life Sciences	Netherlands	H.Stunnenberg@ncmls.ru.nl
Mònica	Bayés	Centre for Genomic Regulation	Spain	monica.bayes@crg.es
Elias	Campo	Hospital Clinic, University of Barcelona	Spain	ecampo@clinic.ub.es
Enrique	de Alava	Centro de Investigacion del Cancer (CIC)	Spain	edealava@usal.es
Xavier	Estivill	Centre for Genomic Regulation	Spain	xavier.estivill@crg.es
Roderic	Guigo	Center for Genomic Regulation	Spain	roderic.guigo@crg.cat
Peter	Klatt Brückl	Spanish Ministry of Science and Innovation	Spain	peter.klatt@micinn.es
Núria	López-Bigas	Pompeu Fabra University	Spain	nuria.lopez@upf.edu
Pilar	Nicolás	Universities of Deusto and of The Basque Country	Spain	pnicolas@genomelaw.deusto.es

First Name	Last Name	Affiliation	Country	Email
Miguel Angel	Piris	Spanish National Cancer Research Centre (CNIO)	Spain	mapiris@cnio.es
Xose	Puente	University of Oviedo	Spain	xspuente@uniovi.es
Carlos	Romeo-Casabona	Universities of Deusto and of The Basque Country	Spain	cromeo@genomelaw.deusto.es
Alfonso	Valencia	Spanish National Cancer Research Center (CNIO)	Spain	valencia@cnio.es
Ewan	Birney	European Bioinformatics Institute	United Kingdom	birney@ebi.ac.uk
Martin	Bobrow	University of Cambridge	United Kingdom	mb238@cam.ac.uk
Susan	Bull	University of Oxford	United Kingdom	susan.bull@well.ox.ac.uk
Peter	Campbell	Wellcome Trust Sanger Institute	United Kingdom	pc8@sanger.ac.uk
Michael	Dunn	Wellcome Trust	United Kingdom	M.DUNN@wellcome.ac.uk
Richard	Durbin	Wellcome Trust Sanger Institute	United Kingdom	rd@sanger.ac.uk
Stephanie	Dyke	Wellcome Trust Sanger Institute	United Kingdom	sd4@sanger.ac.uk
Adrienne	Flanagan	UCL Cancer Institute	United Kingdom	rmkdamf@ucl.ac.uk
Andy	Futreal	Wellcome Trust Sanger Institute	United Kingdom	paf@sanger.ac.uk
Tim	Hubbard	Wellcome Trust Sanger Institute	United Kingdom	th@sanger.ac.uk
Karen	Kennedy	Wellcome Trust Sanger Institute	United Kingdom	kk6@sanger.ac.uk
Sancha	Martin	Wellcome Trust Sanger Institute	United Kingdom	sd1@sanger.ac.uk
Michael	Stratton	Wellcome Trust Sanger Institute	United Kingdom	mrs@sanger.ac.uk
Harold	Swerdlow	Wellcome Trust Sanger Institute	United Kingdom	hps@sanger.ac.uk
Jon	Teague	Wellcome Trust Sanger Institute	United Kingdom	jwt@sanger.ac.uk
Mark	Walport	Wellcome Trust	United Kingdom	m.walport@wellcome.ac.uk
Lynda	Chin	Harvard University	USA	lynda_chin@dfci.harvard.edu
Jim	Downing	St. Jude Children's Research Hospital	USA	jim.downing@stjude.org

First Name	Last Name	Affiliation	Country	Email
Lynn	Dressler	University of North Carolina	USA	dressler@email.unc.edu
Daniela	Gerhard	NCI	USA	gerhardd@mail.nih.gov
Mark	Guyer	NHGRI	USA	mark_guyer@nih.gov
Karin	Jegalian	N/A	USA	karinjeg@verizon.net
John	Maris	Abramson Family Cancer Research Institute	USA	maris@chop.edu
Debabrata	Mukhopadhyay	Mayo Clinic	USA	Mukhopadhyay.Debabrata@mayo.edu
Brad	Ozenberger	NHGRI	USA	bozenberger@mail.nih.gov
Robert	Penny	International Genomics Consortium	USA	rpenny@intgen.org
Jane	Peterson	NHGRI	USA	jane_peterson@nih.gov
Paul	Spellman	LBL	USA	ptspellman@lbl.gov
Sarah	Thayer	Massachusetts General Hospital	USA	sthayer@partners.org
Joe	Vockley	NCI	USA	vockleyj@mail.nih.gov
David	Wheeler	Baylor College of Medicine	USA	wheeler@bcm.tmc.edu
Georgia	Wiesner	Center for Human Genetics, University Hospitals Case Medical Center	USA	Georgia.Wiesner@uhhospitals.org
Richard	Wilson	Washington University	USA	rwilson@watson.wustl.edu