

Cancer Genomes

Moderator: Eleftherios P. Diamandis^{1,2,3*}

Experts: Thomas Hudson,^{4,5,6} Olli Kallioniemi,⁷ Edison T. Liu,^{8,9,10} and Carlos López-Otín¹¹

Cancer has long been regarded as a disease of genes. There are excellent examples of genomic alterations that lead to familial or sporadic cancer; however, despite the spectacular advances in genomics over the last 20 years, the genetic basis of the vast majority of sporadic cancers remains obscure.

New and low-cost sequencing technologies now allow complete DNA sequencing for large numbers of tumors. Recently, the International Cancer Genome Consortium (ICGC)¹² was created to coordinate large-scale cancer genome studies of tumors from 50 different cancer types. It is hoped that systematic analyses of >25 000 cancer genomes at the genomic, epigenomic, and transcriptomic level will reveal previously unidentified oncogenic mutations, which will help us understand cancer pathogenesis and facilitate the development of new cancer therapies. We asked 4 experts to explain more about this organization and its goals.

What are the major goals of the ICGC, and how does your country play a role?



Thomas Hudson: The major goals of the ICGC are to coordinate efforts in many countries to generate and rapidly disseminate extensive catalogues of somatic mutations and other genomic abnormalities present in tumors from at least 50 different cancer types and/or subtypes. With the exception of rare tumors, each project catalogue will include genomic, epigenomic, and transcriptomic data sets from at least 500 pairs of tumor and normal samples.

The Ontario Institute for Cancer Research (OICR)

has already launched a project to study pancreatic ductal adenocarcinoma, a highly fatal form of cancer. The OICR is currently at the planning stage of a prostate cancer genome project, in partnership with Prostate Cancer Canada.

OICR also hosts coordinating bodies of the ICGC: the ICGC Secretariat, which supports communication and workshops among ICGC members, the Data Coordination Center, which manages data exchanges between members, and the ICGC Data Portal (at <http://www.icgc.org>), which provides the scientific community with access to the data and a suite of query tools.



Olli Kallioniemi: The full genomic, transcriptomic, and epigenomic characterization of all the major cancer types and relevant subtypes.



Edison Liu: Singapore is an observer country and not directly participating. Nevertheless, we are fully supportive of the ICGC's goal of identifying the most common mutations in human cancers with the greatest health burdens worldwide.

¹ Department of Pathology and Laboratory Medicine, Mount Sinai Hospital; ² Department of Clinical Biochemistry, University Health Network; and Departments of ³ Laboratory Medicine and Pathobiology, ⁴ Molecular Genetics, and ⁵ Medical Biophysics, University of Toronto; ⁶ Ontario Institute for Cancer Research (OICR), Toronto, Ontario, Canada; ⁷ Institute for Molecular Medicine (FIMM), University of Helsinki, Helsinki, Finland; ⁸ Genome Institute of Singapore (Biomedical Sciences Institutes); ⁹ National University of Singapore; ¹⁰ Health Sciences Authority of Singapore, Singapore; ¹¹ Departamento de

Bioquímica y Biología Molecular, Facultad de Medicina, Universidad de Oviedo, Oviedo, Spain.

* Address correspondence to this author at: 60 Murray St., 6th Floor, Toronto, Ontario, M5G 1X5, Canada. Fax 416-586-8628; e-mail ediamandis@mtnsinai.on.ca.

Received August 25, 2010; accepted August 25, 2010.

¹² Nonstandard abbreviations: ICGC, International Cancer Genome Consortium; OICR, Ontario Institute for Cancer Research.



Carlos López-Otín: The identification of the main genetic alterations in the 50 most prevalent cancers worldwide. To do that, the genome of tumor cells and nontumor cells from the same patient will be completely

sequenced with next-generation sequencing technologies. This will be done in 500 patients per cancer type to identify recurrently mutated genes in at least 3% of tumors. All types of tumor mutations will be analyzed, including point mutations, translocations, and large deletions or insertions. Additionally, epigenomic and transcriptomic data on these tumors will be generated. Our country is leading the sequencing and analysis of chronic lymphocytic leukemia, the most common leukemia in Western countries.

This is an international collaboration. Who is going to fund it?

Thomas Hudson: Each ICGC member project is funded by one or more agencies that provide a minimum of \$20 million US over 5 years. The funds cover a range of activities that include tissue collections, genome analyses, and bioinformatics. In addition, many countries have invested substantially in infrastructure and next-generation sequencing instruments.

Carlos López-Otín: This is an international collaboration in which each country proposes a cancer type to be studied and provides funding to perform the studies for that particular tumor type. The role of the ICGC is to coordinate the different types of cancers studied to prevent duplication of efforts and provide a similar methodology, quality standards, and objectives, so that after the completion of the project, the results generated by different members of the consortium will be comparable and accessible to other scientists.

Over the last few years a few cancer genomes have been published. What did we learn from these initial studies?

Thomas Hudson: Although many cancer genomes will need to be studied to help distinguish driver from passenger mutations, I am amazed that the first cancer genome publications describing one or a handful of tumors have provided considerable insight about cancer processes. For example, studies of melanoma and lung cancer cell lines have shown that the specific patterns of mutations observed in tumors match the patients' exposures to carcinogens such as ultraviolet

light and tobacco. Other studies comparing primary tumors, metastases, and xenografts provide evidence of selection processes in tumors and suggest genes that may be implicated in metastatic and engraftment processes. The high number of somatic mutations, often exceeding 10 000 per tumor, offers some tantalizing evidence of new cancer genes, but the validation of these needs further work.

Olli Kallioniemi: A deeper understanding that cancers are very heterogeneous in terms of their mutational spectrum and the genes involved. Only a few common genetic changes, such as *BRAF* (v-raf murine sarcoma viral oncogene homolog B1) mutations, have been found so far. However, despite an apparent lack of mutational hot spots, many mutations target different members along key cancer pathways. Integrated molecular profiles (according to the ICGC protocol) from large numbers of tumors, with full transcriptomic and epigenomic profiling, do not yet exist.

Edison Liu: First, the heterogeneity and the singularity of most of these mutations suggest that only a few common mutations will be found. The majority appear to be private mutations thus far. However, I do believe that when we have the patient numbers, we will find higher-order "structure" to these mutations relative to gene clusters and pathways.

Carlos López-Otín: The first initiatives in this regard have revealed that the number of mutated genes in a tumor is higher than previously thought. From these studies, some genes have already been identified as frequently mutated in a certain type of tumor, and subsequent studies have confirmed the importance of these genes in tumor development. Additionally, the genomic studies of melanoma or lung carcinoma cells have revealed the presence of specific patterns of mutations that inform us about the exposure to the specific carcinogens that have been at the origin of the development of these particular tumors. Another lesson we have learned from these initial studies is the heterogeneity of tumor mutations, which makes it very difficult to identify the driver mutations and to distinguish them from passenger mutations, thus reinforcing the need to study a large number of samples to understand the mechanisms of cancer development and progression.

Do you believe that sequencing whole genomes, not just exons, will provide additional valuable information?

Thomas Hudson: Absolutely. We cannot ignore the evidence that there is extensive regulatory activity, noncoding RNAs, and other functional units outside of

the protein-coding genes. While the tools to study these features of the genome are less developed and the ability to predict the consequences of the thousands of somatic mutations outside exons is relatively limited, I anticipate that cancer cells will have as many driver mutations in noncoding regions of the genome as they do in exons. A consequence of analyzing a large number of tumor genomes will be the observation of recurrent mutations outside of exons, which will fascinate molecular biologists and spur research into this class of driver mutations.

Olli Kallioniemi: As long as sequencing capacity is not limiting, sequencing whole genomes is important to gain a completely unbiased understanding of mutational events across the genome. At the moment, our ability to interpret all the changes in the nonexonic sites is limited, but this is specifically an area where the ICGC data will be very helpful.

Edison Liu: Yes I do. There is evidence that structural mutations or variations may have a larger role in cancer progression than previously thought. Such rearrangements do not fall in exons, and the functional rearrangements occur in introns and nongenic regions. The only way to ascertain these is by whole genome sequencing.

Carlos López-Otín: Exome sequencing is a very useful tool to identify point mutations or small insertions/deletions rapidly and cost-effectively and will probably be very useful to identify the vast majority of mutations implicated in tumor development. Nevertheless, exome sequencing cannot identify some of the genetic alterations that are important for tumor growth, including translocations, large insertions or deletions, inversions, or changes in copy number. Moreover, there are many alterations important for cancer development and progression that lie outside of the protein-coding regions. Therefore, whole genome sequencing will be essential to get the complete view of the genetic landscape of cancer.

What do you expect the translational outcome of this initiative could be?

Thomas Hudson: The translation of newly discovered cancer genes and mutations into the clinic as well-validated tests, services, or products will take years and substantial work. Early (i.e., 5-year) outcomes having clinical benefits would be the identification of mutations or gene expression patterns that classify tumors into new subclasses that are clinically relevant. For example, it would be useful to distinguish indolent vs aggressive forms of prostate cancer and avoid inconti-

nence and impotence associated with surgery in patients believed to have nonthreatening tumors. I suspect that it will take more than 15 years to develop novel interventions for new cancer genes and mutations discovered by ICGC projects, given the time lines to conduct studies spanning target validation, early drug discovery, and preclinical and clinical studies. The pressure to adopt new approaches, particularly for advanced cancers that do not respond to conventional drugs, will be considerable.

Olli Kallioniemi: ICGC will develop the ultimate reference database of genomic changes in cancer and will help to serve as a reference for diagnostic genomic studies. ICGC data will hopefully highlight specific diagnostic, prognostic, and therapeutic targets and will facilitate the ongoing transition toward personalized medicine.

Edison Liu: First will be the conceptual advance in understanding the detailed genetic mutations and mutational load of human cancers. The coalescence of pathways with large mutational loads can direct us to therapeutic options. Second, we may find ready diagnostics in the resulting data. Lastly, the mutational spectrum and the sequence context of these mutations, when examined in such large numbers, will inevitably provide clues as to the genesis of cancer-associated mutations.

Carlos López-Otín: Novel markers for cancer diagnosis and prognosis, better classification of patients for response to therapy, and novel therapeutic targets.

Why aren't some notable institutions and investigators part of the initiative?

Thomas Hudson: It is correct that although over 50 institutions and 200 investigators are involved with the ICGC, there are excellent teams that have not yet been involved. The main reason for this is funding. It takes considerable resources to support an ICGC project, as each project involves large teams of clinicians, pathologists, cancer biologists, genome scientists, bioinformaticians, and others. The ICGC is still open to new members and projects. Several funding agencies are planning to join or support additional projects, particularly for tumor types not selected yet.

Edison Liu: Only large sequencing centers can effectively participate in this initiative because of the sequencing scope and the production requirements. Smaller sequencing centers with a focus on cell biology may have neither the interest nor the infrastructure to do this effectively.

Carlos López-Otín: The sequencing of tumor genomes will contribute to a better understanding of the mechanisms implicated in cancer development but will not provide all the answers to this disease. Some investigators consider that the sequencing of 25 000 genomes will take important financial resources from other areas of cancer research and are skeptical about the future outcomes from this initiative. I remember that similar questions and doubts were raised at the first stages of the Human Genome Project. However, to my view, the sequencing of the human genome has proven an invaluable resource for the advance of science at all levels, and especially for cancer research. That's why I am convinced that the completion of the ICGC project will also be of enormous value for the entire scientific community.

The expected data from this effort would be thousands of genetic alterations (point mutations, insertions/deletions, chromosomal rearrangements, and copy number variations) that may be "passengers" (vast majority, noncausative) or "drivers" (very few, causative). Who will sort out the "drivers" from the "passengers" and how?

Thomas Hudson: Statistical evidence of "drivers" vs "passengers" requires the detection of genes that are mutated at a higher frequency than the background mutation rate. At the design phase of the ICGC, it was deemed important that each project would have the power to detect mutated genes at the 3% level, a level that is higher than the background mutation rate (estimated to be less than 1 mutation per gene for a sample size of 500).

Functional validation of drivers, particularly those that may be clinically relevant, is outside the scope of the ICGC. Basic and clinical researchers will design a wide spectrum of studies that will evaluate cancer genes. It is anticipated that high-throughput methods (such as RNA interference screens) will be important to prioritize targets, given the high number of mutated genes that will need further investigations.

Olli Kallioniemi: Part of the answer will be achieved through bioinformatic analysis of the actual ICGC data and integration of these data with other independent data sets. The public availability of ICGC data allows any biologist to link up their specific findings on genes, proteins, and pathways with ICGC data. Thus, the global cell biology and cancer research community will eventually help to sort out the key driver events. Functional studies in cancer cell lines, such as high-throughput genome-scale RNA interference, will help to sort out the key driver mutations. My personal view is that many of the rare mutations that we currently

consider "passenger mutations" on the basis of theoretical and statistical considerations may eventually turn out to contribute to some aspects of tumor biology, thus being eventually labeled as "drivers."

Edison Liu: The rest of the community will do the sorting. This information, like the HapMap, will provide truly enabling knowledge for the larger scientific community to exploit.

Carlos López-Otín: Nowadays, the discrimination between driver and passenger mutations is a difficult task. Functional validation in cell lines or in animal models will definitely provide an answer to this problem. However, those approaches are generally time-consuming and expensive, although necessary. Previous studies on driver genes have shown that most driver genes appear mutated in different patients with the same tumor type. In the ICGC project, a minimum number of 500 patients per tumor type has been selected for sequencing because with this number it will be possible to identify genes that show statistically more mutations than what would be expected by chance on the basis of gene size. It is assumed that within the ICGC each member should be able to identify those genes that are recurrently mutated in at least 3% of patients. In addition, all data generated by the ICGC will be made available to the whole community, and metaanalysis of different tumor types to identify recurrently mutated genes across tumor types will be performed either by members of the ICGC or by any other researcher. With this approach, it is likely that many patients will have mutations in genes that are not recurrently mutated in other tumors but nevertheless might be the driver for that particular tumor. There are 2 initiatives to deal with these genes: The first one is to study signaling pathways and not only individual genes, as the outcome might be similar if 2 genes form part of the same pathway. One group at ICGC will perform that kind of analysis, and data will be also available for the community. The second one will be to perform functional analyses of those genes in cells or in animal models.

Some believe that the best bang for the buck could be achieved by funding individual researchers (such as NIH RO1 grants) rather than big consortia and advise waiting until sequencing costs are further reduced. For example, the human genome could have been completed today, instead of 10 years ago, at <0.1% of its cost, saving billions of dollars. What do you say?

Thomas Hudson: This question was debated extensively by funding agencies and scientists at the planning phase of the ICGC (2007–2008). The drastic reductions in sequencing costs were anticipated, and it was impor-

tant to decide “when” the project should start. Pilot projects in the US and UK clearly showed the importance of getting organized early. It was also recognized that the major bottlenecks and costs would move from genome sequencing to sample collections, data storage, and analysis. Finally, it was recognized that it would be more costly for funding agencies to fund large numbers of laboratories doing their own sequencing and that these projects would generally not meet the low costs and higher standards that can be achieved and monitored in larger centers or result in widespread data dissemination to a large community of users. Choosing the right balance of strategic investments, such as ICGC and investigator-driven research projects, was a matter of debate in every participating country.

The sharing of costs of the project (currently \$400 million in 12 countries; projected costs over one decade of \$1 billion) makes the project more affordable to funding agencies.

From an economic perspective, a mere 1% reduction in mortality from cancer would save nearly \$500 billion to current and future Canadians and Americans (and presumably trillions worldwide). An infusion of \$1 billion to cancer genome research to wage a more effective “war on cancer” would clearly yield an excellent return on investment.

Olli Kallioniemi: Both types of funding schemes are necessary, and they should not be considered as alternatives. Technology drives the progress in genomic science, and scientific needs will make technology change. Discovery cannot wait, and it will always be more expensive in the beginning. Without examples like the Human Genome Project, none of our current sequencing technologies would have evolved as far as they have now done. The sequencing of 25 000 tumors is now appearing very realistic and affordable and will be much cheaper in 3–5 years. Sequencing throughput has undergone an exponential improvement in the past 5 years, with a corresponding decrease in the costs per base, and these trends will most likely continue for some years to come.

Edison Liu: I disagree with this statement. The only reason why sequencing is so much better now than 10

years ago is because there were indeed scientific and monetary drivers to develop the next-generation sequencing technologies in the *absence* of any obvious clinical utility. In addition, a 10-year delay in the disclosure of the human genome would have meant a 10-year delay in our scientific discoveries and breakthroughs. Literally every aspect of human biology is dependent on sequence information. There must be a balance between big science and individual science. One cannot optimally perform without the other.

Carlos López-Otín: The human genome has catalyzed research in many areas and has saved probably billions of dollars in sequencing to individual labs, speeding up the identification of genes implicated in disease. A proper balance is necessary, but this aspect has been extensively discussed during the planning phase of the ICGC initiative.

Author Contributions: *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.*

Authors' Disclosures of Potential Conflicts of Interest: *Upon manuscript submission, all authors completed the Disclosures of Potential Conflict of Interest form. Potential conflicts of interest:*

Employment or Leadership: None declared.

Consultant or Advisory Role: O. Kallioniemi, Medisapiens; E.T. Liu, Lilly Singapore Centre for Drug Discovery and VeracYTE.

Stock Ownership: O. Kallioniemi, Medisapiens.

Honoraria: None declared.

Research Funding: O. Kallioniemi, Bayer Schering Pharma and Hoffman-La Roche.

Expert Testimony: None declared.

Other Remuneration: O. Kallioniemi, intellectual property rights, patent, and licensing relationship with Vysis/Abbott Laboratories on fluorescence in situ hybridization/comparative genomic hybridization.

Role of Sponsor: The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, or preparation or approval of manuscript.

Previously published online at DOI: 10.1373/clinchem.2010.152140
