

International network of cancer genome projects

The International Cancer Genome Consortium*

The International Cancer Genome Consortium (ICGC) was launched to coordinate large-scale cancer genome studies in tumours from 50 different cancer types and/or subtypes that are of clinical and societal importance across the globe. Systematic studies of more than 25,000 cancer genomes at the genomic, epigenomic and transcriptomic levels will reveal the repertoire of oncogenic mutations, uncover traces of the mutagenic influences, define clinically relevant subtypes for prognosis and therapeutic management, and enable the development of new cancer therapies.

The genomes of all cancers accumulate somatic mutations¹. These include nucleotide substitutions, small insertions and deletions, chromosomal rearrangements and copy number changes that can affect protein-coding or regulatory components of genes. In addition, cancer genomes usually acquire somatic epigenetic 'marks' compared to non-neoplastic tissues from the same organ, notably changes in the methylation status of cytosines at CpG dinucleotides.

A subset of the somatic mutations in cancer cells confers oncogenic properties such as growth advantage, tissue invasion and metastasis, angiogenesis, and evasion of apoptosis². These are termed 'driver' mutations. The identification of driver mutations will provide insights into cancer biology and highlight new drug targets and diagnostic tests. Knowledge of cancer mutations has already led to the development of specific therapies, such as trastuzumab for *HER2* (also known as *NEU* or *ERBB2*)-positive breast cancers³ and imatinib, which targets BCR-ABL tyrosine kinase for the treatment of chronic myeloid leukaemia^{4,5}. The remaining somatic mutations in cancer genomes that do not contribute to cancer development are called 'passengers'. These mutations provide insights into the DNA damage and repair processes that have been operative during cancer development, including exogenous environmental exposures^{6,7}. In most cancer genomes, it is anticipated that passenger mutations, as well as germline variants not yet catalogued in polymorphism databases, will substantially outnumber drivers.

Large-scale analyses of genes in tumours have shown that the mutation load in cancer is abundant and heterogeneous^{8–13}. Preliminary surveys of cancer genomes have already demonstrated their relevance in identifying new cancer genes that constitute potential therapeutic targets for several types of cancer, including *PIK3CA*¹⁴, *BRAF*¹⁵, *NF1* (ref. 10), *KDR*¹⁰, *PIK3R1* (ref. 9), and histone methyltransferases and demethylases^{16,17}. These projects have also yielded correlations between cancer mutations and prognosis, such as *IDH1* and *IDH2* mutations in several types of gliomas^{13,18}. Advances in massively parallel sequencing technology have enabled sequencing of entire cancer genomes^{19–22}.

Following the launch of comprehensive cancer genome projects in the United Kingdom (Cancer Genome Project)²³ and the United States (The Cancer Genome Atlas)²⁴, cancer genome scientists and funding agencies met in Toronto (Canada) in October 2007 to discuss the opportunity to launch an international consortium. Key reasons for its formation were: (1) the scope is huge; (2) independent cancer genome initiatives could lead to duplication of effort or

incomplete studies; (3) lack of standardization across studies could diminish the opportunities to merge and compare data sets; (4) the spectrum of many cancers is known to vary across the world; and (5) an international consortium will accelerate the dissemination of data sets and analytical methods into the user community.

Working groups were created to develop strategies and policies that would form the basis for participation in the ICGC. The goals of the consortium (Box 1) were released in April 2008 (http://www.icgc.org/files/ICGC_April_29_2008.pdf). Since then, working groups and initial member projects have further refined the policies and plans for international collaboration.

Bioethical framework

ICGC members agreed to a core set of bioethical elements for consent as a precondition of membership (Box 2). The Ethics and Policy

Box 1 | Goals of the ICGC

The goals of the ICGC are:

- To coordinate the generation of comprehensive catalogues of genomic abnormalities (somatic mutations) in tumours in 50 different cancer types and/or subtypes that are of clinical and societal importance across the globe.
- To ensure high quality by defining the catalogue for each tumour type or subtype to include the full range of somatic mutations, such as single-nucleotide variants, insertions, deletions, copy number changes, translocations and other chromosomal rearrangements, and to have the following features. (1) Comprehensiveness, such that most cancer genes with somatic abnormalities occurring at a frequency of greater than 3% are discovered. (2) High resolution, ideally at a single nucleotide level. (3) High quality, using common standards for pathology and technology. (4) Data from matched non-tumour tissue, to distinguish somatic from inherited sequence variants and aberrations. (5) Generate complementary catalogues of transcriptomic and epigenomic data sets from the same tumours.
- Make the data available to the entire research community as rapidly as possible, and with minimal restrictions, to accelerate research into the causes and control of cancer.
- Coordinate research efforts so that the interests and priorities of individual participants, self-organizing consortia, funding agencies and nations are addressed, including use of the burden of disease and the minimization of unnecessary redundancy in tumour analysis efforts.
- Support the dissemination of knowledge and standards related to new technologies, software, and methods to facilitate data integration and sharing with cancer researchers around the globe.

*A list of participants and their affiliations appears at the end of the paper.

Box 2 | Core bioethical elements

For prospective research, ICGC members should convey to potential participants, that:

- The ICGC is a coordinated effort among related scientific research projects being carried on around the world.
- Participation in the ICGC and its component projects is voluntary.
- Samples and data collected will be used for cancer research, which may include whole-genome sequencing.
- The patient's care will not be affected by their decision about participation.
- The samples collected will be in limited quantities; access to them will be tightly controlled and will depend on the policy and practices of the ICGC-member project. At least a small percentage of the samples may be shared with laboratories in other countries for the purposes of performing quality control studies.
- Data derived from the samples collected and data generated by the ICGC members will be made accessible to ICGC members and other international researchers through either an open or a controlled access database under terms and conditions that will maximize participant confidentiality.
- The researchers accessing data and samples will be required to affirm that they will not attempt to re-identify participants.
- There is a remote risk of being identified from data available on the databases.
- Once data are placed in open databases, those data cannot be withdrawn later.
- In controlled access databases the links to (local) data that can identify an individual will be destroyed after withdrawal. Data previously distributed will continue to be used.
- ICGC members agree not to make claims to possible intellectual property on primary data.
- No profit from eventual commercial products will be returned to subjects donating samples.

For retrospective research, the above guidelines remain the same, with the exception that where the individual is no longer a patient, there will not be a concern that their care could be affected by participation.

For research involving samples and data from deceased individuals:

- Where required by law or ethics, consent should always be obtained from the families of a deceased individual if their samples and data are to be used; if re-consent is not required, however, ethics review is sufficient.
- Ethics committee review should be sought for all research proposing the use of existing sample and data collections.
- Existing collections are a limited and valuable resource; access to them will be tightly controlled.

For research using anonymized samples, ethics review may be required in some jurisdictions.

Committee has created patient consent templates for both prospective collection and retrospective use of samples and data for ICGC projects. Differences in project-specific requirements and national legal frameworks may require some local amendments, while still reflecting the core principles of ICGC.

The ICGC recognizes a delicate balance between protecting participants' personal data and sharing these data to accelerate cancer research. Data access policies have been drawn up that are respectful of the rights of the donors, while allowing ICGC data derived from samples to be shared ethically among a wide research community. Two levels of access have been implemented. For data that cannot be used to identify individuals, 'open access' data sets are publicly available. These include data such as gender, age range, histology, normalized gene expression values, epigenetic data sets, somatic mutations, summaries of germline data, and study protocols. 'Controlled access' data sets contain germline genomic data and detailed clinical information that are associated to a unique individual whose personal identifiers have been removed. To access controlled data sets researchers must seek authorizations by contacting the Data Access Compliance Office (DACO) (<http://www.icgc.org/daco>). An independent International Data Access Committee (IDAC) oversees the work of the DACO and provides assistance with resolving issues that arise.

Pathology and clinical annotation

Large-scale genomic studies of human tumours rely on the availability of freshly frozen tumour tissue. To address the paucity of samples that meet ICGC standards, many projects have initiated prospective collections of high-quality source material. Accordingly, the ICGC recommended procedures to promote consistency of sample processing throughout the consortium and ensure a series of quality features such as high tissue integrity and tumour cell content. Each project will need to include diverse data types, such as environmental exposures, clinical history of participants, tumour histopathology, and clinical outcomes.

Tumours show considerable clinical and biological heterogeneity that has resulted in a variety of tumour classifications. Within the ICGC, special measures are taken to promote the consistency of diagnosis. These include the coordination of diagnostic criteria among groups investigating tumours that are related, and policies that all samples will be reviewed by at least two independent reference pathologists. Furthermore, images of the stained tumour sections (or blood smear or cytopins for haematological neoplasias) from which diagnoses were made, will be stored and made available to the community.

Although different tumour types may require specific procedures for tumour acquisition or compilation of clinical and environmental data, the ICGC has set guidelines about the use of common definitions and data standards. This will allow ICGC data users to identify correlations between tumour-specific molecular changes with clinical and histopathological data including prognosis, prediction of therapy response and tumour classification schemes for diagnosis.

Study design and statistical issues

To identify cancer-related genes, one needs to detect genes that are mutated at a higher frequency than the background mutation rate. Given that several driver genes have been found to be mutated at low frequencies, the ICGC will identify somatic mutations observed in at least 3% of tumours of a given subtype. The ICGC determined that 500 samples would be needed per tumour type (although for rare tumour types, a smaller sample size may be justified). In practice, the degree of heterogeneity of a given tumour type is difficult to know in advance, such that some particularly heterogeneous tumour types may require larger sample collections.

Cancer genome analyses

High-quality catalogues of somatic mutations from whole cancer genomes will ultimately be the ICGC standard. Shotgun sequencing using second generation technologies can detect all classes of somatic mutation implicated in cancer. Moreover, if the level of coverage is sufficient, comprehensive high-quality catalogues of somatic mutations from individual cancer genomes can be acquired with >90% sensitivity and >95% specificity. To achieve this, it will be necessary to sequence the genome of both the cancer and a normal tissue from the same individual to distinguish germline variants. Although a few genomes of this standard have already been generated, the cost and the continuing technology development will mean that interim analyses of particularly informative sectors of the genome will be carried out, for example of all coding exons and microRNAs.

For each individual cancer genome, the catalogue of somatic mutations will be supplemented by genome-wide information on the state of methylation of CpG dinucleotides. The optimal strategies and technologies to achieve this are not yet clear. Moreover, the genomes of individual cancers will be accompanied, where possible, by analyses of the transcriptome. Although conventional array-based approaches predominate at present, it is preferable that RNA sequencing becomes the standard as sequencing has a greater dynamic range²⁵ and provides further information including new transcripts and sequence variants²⁶.

ICGC data sets

The distributed nature of the consortium coupled with the large size of the data sets makes it cumbersome to store all data in a single

centralized repository. For this reason, the ICGC has adopted a 'franchise' database model for integrating the information and making it available to the public. Under this model, each member project releases tumour information by copying it into its local franchise database after it has been quality checked. Each franchise database shares a common schema to describe the specimens, the associated clinical information, and their genome characterization data. ICGC primary data files, are sent to the National Center for Biotechnology Information (NCBI) and/or the European Bioinformatics Institute (EBI) for archiving, while interpreted data sets, such as somatic mutation calls, are stored in franchise databases. The ICGC franchise databases and web portal use BioMart²⁷, a data federation technology originally developed for use in Ensembl²⁸, and since adopted for use by several model organism and genome databases. The management of the ICGC data flow is the responsibility of the ICGC Data Coordination Center (DCC) located at the Ontario Institute for Cancer Research.

The DCC also operates the ICGC data portal that allows researchers to access both open and controlled access portions of the ICGC data. The portal provides a variety of user interfaces that range from simple gene-oriented queries ('show me all the non-silent coding mutations identified in *PIK3R1* for all cancers') to queries that integrate genomic, clinical and functional information ('show me all members of the Toll-receptor pathway having deletions in stage III breast cancer'). These queries will be distributed across the franchise databases in a manner that is invisible to the user. The portal will also provide links to the primary files at the NCBI and EBI, interfaces for generating tabular reports, data dumps in common bioinformatics formats, and other visualizations including genome browser tracks, pathway diagrams and survival curves. The portal is available via a link at <http://www.icgc.org>.

At the time of this publication, the following cancer and reference data sets will be available through the ICGC web portal: (1) initial data releases from ICGC members for breast cancer (UK), liver cancer (Japan), and pancreatic cancer (Australia and Canada); (2) a whole genome data set of a metastatic melanoma cell line (COLO829)⁶; (3) open data sets from The Cancer Genome Atlas (TCGA) for glioblastoma multiforme (GBM) and serous cystadenocarcinoma of the ovary (see later); (4) whole exome somatic mutation data from 68 individuals with breast, colorectal, pancreatic cancer and GBM^{11–13}; (5) links to the human reference genome (<http://www.genomereference.org/>) and gene annotations from the GENCODE project (<http://www.sanger.ac.uk/gencode/>) that includes the CCDS gene set²⁹; (6) links to the single nucleotide polymorphism database (dbSNP)³⁰ and the HapMap³¹ databases, providing access to common patterns of variation in reference population samples; (7) links to Reactome³², a curated database of biological pathways in human; and (8) a set of reference gene models, mirrored from ENSEMBL²⁸.

The current version of the web portal provides an entry point to the open access data tier by interactive query as well as bulk download of data files. We expect that in mid-2010 both open access and controlled data will be available.

The ICGC recently established a bioinformatics analysis working group to compare pipelines, analytic methods, consistency within and among algorithms, and establish guidelines or best practices for the consortium. Over time, significant resources will be deployed to develop strategies to analyse the large complex data sets generated by ICGC member projects, and provide value-added views of cancer genomic data by integrating them with other biological and epidemiological data sets.

Data release and intellectual property policies

The data release policies of the ICGC are intended to maximize public benefit while, at the same time, protecting the interests and rights of sample donors and their relatives. Members of the ICGC are committed to the principles of rapid data release (with appropriate controlled access mechanisms), in concordance with the Toronto

statement³³. ICGC members encourage the scientific community to use any data that targets specific genes and mutations, without any restrictions. To allow ICGC members the opportunity to be the first to publish global analyses from data sets they generate, the consortium has also agreed that member projects may specify conditions that include a time limit during which other data users are asked to refrain from publishing global analyses (defined by several ICGC member projects as 100 tumours and matched controls), a provision referred to as a 'publication moratorium'. To allow time for a data set to be analysed and submitted for publication, ICGC members will have at most one year after released data sets reach the specified threshold before third parties are permitted to submit manuscripts describing global analyses. Further details on data release guidelines for data producers, users and reviewers are available <http://www.icgc.org>. Users of ICGC data are expected to respect these terms and to cite this manuscript and the source of pre-publication data, including the version of the data set. In cases of uncertainty, scientists using ICGC data are encouraged to contact the member projects to discuss publication plans.

ICGC members believe that maximum public benefit will be achieved if the data remain publicly accessible without patent restrictions, hence no claims to possible intellectual property derived from primary data (including somatic mutations) will be made. Users of ICGC data (including ICGC members) may elect to perform further research and to exercise their intellectual property rights on these downstream discoveries. If this occurs, users are expected to implement licensing policies that do not obstruct further research.

Initial ICGC projects

At present, ten countries and two European consortia have initiated cancer genome projects under the umbrella of the ICGC. The initial projects, listed in Supplementary Table 1, will analyse tumour types found around the globe and throughout the human body affecting a diversity of organs, including blood, brain, breast, kidney, liver, pancreas, stomach, oral cavity and ovary. Over time, the ICGC will investigate 50 or more types and subtypes of cancer in adults and children. In the case of tumours with several subtypes, analyses should be focused on subtypes that may be defined on pathological, molecular, aetiological or geographical differences. It is expected that some cancer types will be studied in parallel in different parts of the world, as the mutation profiles may differ among populations. The consortium has enabled the coordination of initial projects analysing similar cancers in different countries, and in some cases, the redirection of resources to launch new projects.

The Cancer Genome Atlas

TCGA is a comprehensive program in cancer genomics that is jointly supported and managed by the National Cancer Institute and the National Human Genome Research Institute of the US National Institutes of Health. TCGA began in 2006 as a pilot focused on three projects, glioblastoma multiforme (GBM), serous cystadenocarcinoma of the ovary, and lung squamous carcinoma, and has recently expanded to produce comprehensive genomic data sets for at least ten other cancers in the next two years. Given TCGA's contributions in launching the ICGC and cooperation to ensure that its policies (posted at <http://cancergenome.nih.gov>) are coordinated with those of the ICGC, TCGA's participation in the ICGC is considered to be equivalent to that of a full member. TCGA, however, is not able to join the ICGC formally at this time, because of technical and legal issues in the US related to the mechanisms of the distribution of controlled-access data, although such data are directly available to investigators at <http://cancergenome.nih.gov/dataportal>. The National Institutes of Health (NIH) policies relating to distribution of controlled-access data sets are being reviewed with the intent of enabling researchers to integrate and analyse across databases, for example, using the franchise model adopted by the ICGC. Meanwhile, TCGA

is ensuring that projects are coordinated and data sets are compatible with those of the consortium.

ICGC in the next decade

A large proportion of common cancers affecting patients around the world have been or will soon be selected for comprehensive cancer genome studies. Further efforts will be needed to leverage support and expertise to tackle the remaining tumour types, including rare cancers. The challenges of the ICGC are daunting owing to the scope of the initiative, the complexity that is inherent to the heterogeneity of cancer, and the limitations of current technologies to provide accurate long-range assemblies of highly rearranged chromosomes found in tumour cells. These challenges underscore the importance of continued international coordination and further engagement of the scientific community in the next decade.

Moving towards clinical applications

ICGC catalogues, which are expected to grow exponentially, will have immediate relevance in the cancer research community. Early insight into the biology of somatic mutations will come from functional studies in cell-based and animal models of tumours. Mutation screens in retrospective tumour banks linked to registries or clinical trials having significant clinical data will inform on the potential clinical utility of somatic mutations as biomarkers for prognosis or drug-response. Germline variants identified by ICGC projects may allow the discovery of genes predisposing to familial malignancies, such as *PALB2* and pancreatic cancer^{12,34}. High throughput screens of RNA interference or small molecule libraries, and the adaptation of existing model systems, will have a major role in refining potential therapeutic candidates for further study³⁵.

Translating these discoveries into clinical practice will require more sophisticated clinical trials that take into account the increases in phenotypic subdivisions, further coordination to identify subjects having tumours with similar profiles, and increased use of biomarkers, genomic analyses, informatics and other technologies in the clinical development of new therapeutics. Given the tremendous potential for relatively low-cost genomic sequencing to reveal clinically useful information, we anticipate that in the not so distant future, partial or full cancer genomes will routinely be sequenced as part of the clinical evaluation of cancer patients and as part of their continuing clinical management. The successful and appropriate translation of cancer genome research into clinical practice will raise important social and ethical questions. It will be essential to combine the expertise of oncologists, biostatisticians, pathologists, geneticists, policy-makers and members of the biopharmaceutical industry to meet this challenge by developing new policies and clinical models that enable rapid translation of many new biomarkers and cancer targets into new clinical tests and therapeutic interventions that will benefit cancer patients.

- Stratton, M. R., Campbell, P. J. & Futreal, P. A. The cancer genome. *Nature* **458**, 719–724 (2009).
- Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100**, 57–70 (2000).
- Slamon, D. J. *et al.* Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N. Engl. J. Med.* **344**, 783–792 (2001).
- Druker, B. J. *et al.* Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *N. Engl. J. Med.* **344**, 1031–1037 (2001).
- Druker, B. J. *et al.* Activity of a specific inhibitor of the BCR-ABL tyrosine kinase in the blast crisis of chronic myeloid leukemia and acute lymphoblastic leukemia with the Philadelphia chromosome. *N. Engl. J. Med.* **344**, 1038–1042 (2001).
- Pleasance, E. D. *et al.* A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**, 184–190 (2010).
- Pleasance, E. D. *et al.* A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* **463**, 191–196 (2010).
- Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153–158 (2007).
- Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061–1068 (2008).
- Ding, L. *et al.* Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* **455**, 1069–1075 (2008).
- Wood, L. D. *et al.* The genomic landscapes of human breast and colorectal cancers. *Science* **318**, 1108–1113 (2007).
- Jones, S. *et al.* Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* **321**, 1801–1806 (2008).
- Parsons, D. W. *et al.* An integrated genomic analysis of human glioblastoma multiforme. *Science* **321**, 1807–1812 (2008).
- Samuels, Y. *et al.* High frequency of mutations of the *PIK3CA* gene in human cancers. *Science* **304**, 554 (2004).
- Davies, H. *et al.* Mutations of the *BRAF* gene in human cancer. *Nature* **417**, 949–954 (2002).
- van Haften, G. *et al.* Somatic mutations of the histone H3K27 demethylase gene *UTX* in human cancer. *Nature Genet.* **41**, 521–523 (2009).
- Dalgliesh, G. L. *et al.* Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature* **463**, 360–363 (2010).
- Yan, H. *et al.* *IDH1* and *IDH2* mutations in gliomas. *N. Engl. J. Med.* **360**, 765–773 (2009).
- Ley, T. J. *et al.* DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* **456**, 66–72 (2008).
- Mardis, E. R. *et al.* Recurring mutations found by sequencing an acute myeloid leukemia genome. *N. Engl. J. Med.* **361**, 1058–1066 (2009).
- Shah, S. P. *et al.* Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* **461**, 809–813 (2009).
- Stephens, P. J. *et al.* Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature* **462**, 1005–1010 (2009).
- Dickson, D. Wellcome funds cancer database. *Nature* **401**, 729 (1999).
- Collins, F. S. & Barker, A. D. Mapping the cancer genome. Pinpointing the genes involved in cancer will help chart a new course across the complex landscape of human malignancies. *Sci. Am.* **296**, 50–57 (2007).
- Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* **5**, 621–628 (2008).
- Shah, S. P. *et al.* Mutation of *FOXL2* in granulosa-cell tumors of the ovary. *N. Engl. J. Med.* **360**, 2719–2729 (2009).
- Haider, S. *et al.* BioMart Central Portal—unified access to biological data. *Nucleic Acids Res.* **37**, W23–W27 (2009).
- Hubbard, T. J. *et al.* Ensembl 2009. *Nucleic Acids Res.* **37**, D690–D697 (2009).
- Pruitt, K. D. *et al.* The consensus coding sequence (CCDS) project: identifying a common protein-coding gene set for the human and mouse genomes. *Genome Res.* **19**, 1316–1323 (2009).
- Sherry, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29**, 308–311 (2001).
- International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
- Matthews, L. *et al.* Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res.* **37**, D619–D622 (2009).
- Toronto International Data Release Workshop Authors. Prepublication data sharing. *Nature* **461**, 168–170 (2009).
- Jones, S. *et al.* Exomic sequencing identifies *PALB2* as a pancreatic cancer susceptibility gene. *Science* **324**, 217 (2009).
- Chin, L. & Gray, J. W. Translating insights from the cancer genome into clinical practice. *Nature* **452**, 553–563 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank research participants who are generously donating samples and data, as well as physicians and clinical staff contributing to sample annotation and collection. A complete list of organizations that support ICGC projects is in Supplementary Table 1.

Author Contributions See list of consortium authors below.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.J.H. (tom.hudson@oicr.on.ca).

International Cancer Genome Consortium

Executive committee Thomas J. Hudson^{1,2} (Chairperson), Warwick Anderson³, Axel Aretz⁴, Anna D. Barker⁵, Cindy Bell⁶, Rosa R. Bernabe⁷, M. K. Bhan⁸, Fabien Calvo⁹, Iiro Eerola¹⁰, Daniela S. Gerhard⁵, Alan Gutmacher¹¹, Mark Guyer¹², Fiona M. Hemsley¹³, Jennifer L. Jennings¹, David Kerr^{14,15}, Peter Klatt⁷, Patrik Kolar¹⁰, Jun Kusuda¹⁶, David P. Lane¹³, Frank Laplace¹⁷, Youyong Lu¹⁸, Gerd Nettekoven¹⁹, Brad Ozenberger¹², Jane Peterson¹², T.S. Rao⁸, Jacques Remacle¹⁰, Alan J. Schafer²⁰, Tatsuhiro Shibata²¹, Michael R. Stratton²², Joseph G. Vockley⁵, Koichi Watanabe²³, Huanming Yang²⁴, Matthew M. F. Yuen²⁵

Ethics and policy committee Bartha M. Knoppers²⁶ (Leader), Martin Bobrow²⁷, Anne Cambon-Thomsen²⁸, Lynn G. Dressler²⁹, Stephanie O. M. Dyke²², Yann Joly²⁶, Kazuto Kato³⁰, Karen L. Kennedy²², Pilar Nicolás³¹, Michael J. Parker³², Emmanuelle

Rial-Sebbag²⁸, Carlos M. Romeo-Casabona³¹, Kenna M. Shaw⁵, Susan Wallace²⁶, Georgia L. Wiesner^{33,34}, Nikolajs Zeps^{35,36}

Tissue and clinical annotation working group Peter Lichter³⁷ (Leader), Andrew V. Biankin^{38,39}, Christian Chabannon^{9,40}, Lynda Chin^{41,42}, Bruno Clément⁴³, Enrique de Alava⁴⁴, Françoise Degos⁴⁵, Martin L. Ferguson⁴⁶, Peter Geary⁴⁷, D. Neil Hayes⁴⁸, Thomas J. Hudson¹², Amber L. Johns³⁸, Arek Kasprzyk¹, Hidewaki Nakagawa⁴⁹, Robert Penny⁵⁰, Miguel A. Piris⁵¹, Rajiv Sarin⁵², Aldo Scarpa^{53,54}, Tatsuhiro Shibata²¹, Marc van de Vijver^{55,56}

Technologies working group P. Andrew Futreal²² (Leader), Hiroyuki Aburatani⁵⁷, Mónica Bayés^{58,59}, David D.L. Bowtell^{60,61}, Peter J. Campbell^{22,62}, Xavier Estivill^{58,59}, Daniela S. Gerhard⁵, Sean M. Grimmond⁶³, Ivo Gut⁶⁴, Martin Hirst⁶⁵, Carlos López-Otin⁶⁶, Partha Majumder⁶⁷, Marco Marra⁶⁵, John D. McPherson¹⁶⁸, Hidewaki Nakagawa⁴⁹, Zemin Ning²², Xose S. Puente⁶⁶, Yijun Ruan⁶⁹, Tatsuhiro Shibata²¹, Michael R. Stratton²², Hendrik G. Stunnenberg⁷⁰, Harold Swerdlow²², Victor E. Velculescu⁷¹, Richard K. Wilson^{72,73}, Hong H. Xue^{74,75}, Liu Yang⁷⁶

Bioinformatics analyses working group Paul T. Spellman⁷⁷ (Leader), Gary D. Bader^{78,79}, Paul C. Boutros¹, Peter J. Campbell^{22,62}, Paul Flicek⁸⁰, Gad Getz⁸¹, Roderic Guigó⁸², Guangwu Guo²⁴, David Haussler⁸³, Simon Heath⁶⁴, Tim J. Hubbard²², Tao Jiang²⁴, Steven M. Jones⁶⁵, Qibin Li²⁴, Nuria López-Bigas⁸⁴, Ruiyang Luo²⁴, Lakshmi Muthuswamy¹, B. F. Francis Ouellette¹, John V. Pearson⁶³, Xose S. Puente⁶⁶, Victor Quesada⁶⁶, Benjamin J. Raphael⁸⁵, Chris Sander⁸⁶, Tatsuhiro Shibata²¹, Terence P. Speed^{87,88}, Lincoln D. Stein¹, Joshua M. Stuart⁸⁹, Jon W. Teague²², Yasushi Totoki²¹, Tatsuhiro Tsunoda⁴⁹, Alfonso Valencia⁹⁰, David A. Wheeler⁹¹, Honglong Wu²⁴, Shancen Zhao²⁴, Guangyu Zhou²⁴

Data coordination and management working group Lincoln D. Stein¹ (Leader), Roderic Guigó⁸², Tim J. Hubbard²², Yann Joly²⁶, Steven M. Jones⁶⁵, Arek Kasprzyk¹, Mark Lathrop^{64,92}, Nuria López-Bigas⁸⁴, B. F. Francis Ouellette¹, Paul T. Spellman⁷⁷, Jon W. Teague²², Gilles Thomas^{93,94}, Alfonso Valencia⁹⁰, Teruhiko Yoshida²¹

Data release, data tiers and publications working group Karen L. Kennedy²² (Leader), Myles Axton⁹⁵, Stephanie O. M. Dyke²², P. Andrew Futreal²², Daniela S. Gerhard⁵, Chris Gunter⁹⁶, Mark Guyer¹², Thomas J. Hudson¹², John D. McPherson¹⁶⁸, Linda J. Miller⁹⁷, Brad Ozenberger¹², Kenna M. Shaw⁵

Data coordination centre Arek Kasprzyk¹ (Leader), Lincoln D. Stein¹ (Leader), Junjun Zhang¹, Syed A. Haider⁹⁸, Jianxin Wang¹, Christina K. Yung¹, Anthony Cross¹, Yong Liang¹, Saravanamuttu Gnaneshan¹, Jonathan Guberman¹, Jack Hsu¹

International data access committee Martin Bobrow²⁷ (Leader), Don R. C. Chalmers⁹⁹, Karl W. Hasel⁹, Yann Joly²⁶, Terry S. H. Kaan¹⁰⁰, Karen L. Kennedy²², Bartha M. Knoppers²⁶, William W. Lowrance¹⁰¹, Tohru Masui¹⁶, Pilar Nicolás³¹, Emmanuelle Rial-Sebbag²⁸, Laura Lyman Rodriguez¹², Catherine Vergely¹⁰², Teruhiko Yoshida²¹

Cancer genome projects: Pancreatic cancer (ductal adenocarcinoma) and ovarian cancer (serous adenocarcinoma) (Australia) Sean M. Grimmond⁶³ (Leader), Andrew V. Biankin^{38,39}, David D. L. Bowtell^{60,61}, Nicole Cloonan⁶³, Anna deFazio^{103,104}, James R. Eshleman¹⁰⁵, Dariush Etemadmoghadam^{60,61}, Brooke A. Gardiner⁶³, James G. Kench^{38,106}, Aldo Scarpa^{53,54}, Robert L. Sutherland³⁸, Margaret A. Tempero¹⁰⁷, Nicola J. Waddell⁶³, Peter J. Wilson⁶³, **Pancreatic cancer (ductal adenocarcinoma) (Canada)** John D. McPherson¹⁶⁸ (Leader), Steve Gallinger^{108,109}, Ming-Sound Tsao^{110,111}, Patricia A. Shaw¹¹², Gloria M. Petersen¹¹³, Debabrata Mukhopadhyay¹¹⁴, Lynda Chin^{41,42}, Ronald A. DePinho^{41,115}, Sarah Thayer¹¹⁶, Lakshmi Muthuswamy¹, Kamran Shazand¹, Timothy Beck¹, Michelle Sam¹, Lee Timms¹, Vanessa Ballin¹; **Gastric cancer (intestinal- and diffuse-type) (China)** Youyong Lu¹⁸ (Leader), Jiafu Ji¹⁸, Xiuqing Zhang²⁴, Feng Chen¹⁸, Xueda Hu²⁴, Guangyu Zhou²⁴, Qi Yang²⁴, Geng Tian²⁴, Lianhai Zhang¹⁸, Xiaofang Xing¹⁸, Xianghong Li¹⁸, Zhenggang Zhu¹¹⁷, Yingyan Yu¹¹⁷, Jun Yu¹¹⁸, Huanming Yang²⁴; **Renal cancer (renal cell carcinoma; focus on but not limited to clear cell subtype) (European Union/France)** Mark Lathrop^{64,92} (Leader), Jörg Tost^{64,92}, Paul Brennan¹¹⁹, Ivana Holcatova¹²⁰, David Zaridze¹²¹, Alvis Brazma⁸⁰, Lars Egevad¹²², Egor Prokhorchouk¹²³, Rosamonde Elizabeth Banks¹²⁴, Mathias Uhlen¹²⁵, Anne Cambon-Thomsen²⁸, Juris Viksna¹²⁶, Fredrik Ponten¹²⁷, Konstantin Skryabin¹²⁸; **Breast cancer (subtypes defined by an amplification of ER+ HER2 ductal-type) (European Union/United Kingdom)** Michael R. Stratton²² (Leader), P. Andrew Futreal²², Ewan Birney⁸⁰, Ake Borg¹²⁹, Anne-Lise Børresen-Dale^{130,131}, Carlos Caldas¹³², John A. Foekens¹³³, Sancha Martin²², Jorge S. Reis-Filho¹³⁴, Andrea L. Richardson^{135,136}, Christos Sotiropoulos¹³⁷, Hendrik G. Stunnenberg⁷⁰, Gilles Thomas^{93,94}, Marc van de Vijver^{55,56}, Laura van't Veer⁵⁵; **Breast cancer (subtype defined by an amplification of the HER2 gene) (France)** Fabien Calvo⁹ (Leader), Daniel Birnbaum⁴⁰, Hélène Blanche⁹², Pascal Boucher⁹, Sandrine Boyault¹³⁸, Christian Chabannon^{9,40}, Ivo Gut⁶⁴, Jocelyne D. Masson-Jacquemier⁴⁰, Mark Lathrop^{64,92}, Iris Pauporte⁹, Xavier Pivot¹³⁹, Anne Vincent-Salomon¹⁴⁰, Eric Tabone¹³⁸, Charles Theillet¹⁴¹, Gilles Thomas^{93,94}, Jörg Tost^{64,92}, Isabelle Treilleux¹³⁸; **Liver cancer (hepatocellular carcinoma; secondary to alcohol and adiposity) (France)** Fabien Calvo⁹ (Leader), Paulette Bioulac-Sage¹⁴², Bruno Clément⁴³, Thomas Decaens^{43,144}, Françoise Degos⁴⁵, Dominique Franco¹⁴⁵, Ivo Gut⁶⁴, Marta Gut⁹², Simon Heath⁶⁴, Mark

Lathrop^{64,92}, Didier Samuel^{146,147}, Gilles Thomas^{93,94}, Jessica Zucman-Rossi¹⁴⁸, **Paediatric brain tumours (medulloblastoma, paediatric pilocytic astrocytoma) (Germany)** Peter Lichter³⁷ (Leader), Roland Eils^{37,149} (Leader), Benedikt Brors³⁷, Jan O. Korbe^{80,150}, Andrey Korshunov¹⁵¹, Pablo Landgraf¹⁵², Hans Lehrach¹⁵³, Stefan Pfister^{37,154}, Bernhard Radlwimmer³⁷, Guido Reifenberger¹⁵⁵, Michael D. Taylor^{156,157}, Christof von Kalle^{158,159}; **Oral cancer (gingivobuccal) (India)** Partha P. Majumder⁶⁷ (Leader), Rajiv Sarin⁵², T. S. Rao⁶, M. K. Bhan⁸; **Rare pancreatic tumours (enteropancreatic endocrine tumours and rare pancreatic exocrine tumours; intraductal papillary mucinous neoplasms, solid pseudopapillary tumours, mucinous cystic neoplasms and other rarer tumours) (Italy)** Aldo Scarpa^{53,54} (Leader), Paolo Pederzoli¹⁶⁰, Rita T. Lawlor⁵⁴, Massimo Delledonne¹⁶¹, Alberto Bardelli^{162,163}, Andrew V. Biankin^{38,39}, Sean M. Grimmond⁶³, Thomas Gress¹⁶⁴, David Klimstra¹⁶⁵, Giuseppe Zamboni⁵³; **Liver cancer (hepatocellular carcinoma; virus associated) (Japan)** Tatsuhiro Shibata²¹ (Leader), Yusuke Nakamura^{49,166}, Hidewaki Nakagawa⁴⁹, Jun Kusuda¹⁶, Tatsuhiro Tsunoda⁴⁹, Satoru Miyano¹⁶⁶, Hiroyuki Aburatani⁵⁷, Kazuto Kato³⁰, Akihiro Fujimoto⁴⁹, Teruhiko Yoshida²¹; **Chronic lymphocytic leukaemia (with mutated and unmutated IgVH) (Spain)** Elias Campo¹⁶⁷ (Leader), Carlos López-Otin⁶⁶, Xavier Estivill^{58,59}, Roderic Guigó⁸², Silvia de Sanjose¹⁶⁸, Miguel A. Piris⁵¹, Emili Montserrat¹⁶⁷, Marcos González-Díaz²⁴, Xose S. Puente⁶⁶, Pedro Jares¹⁶⁷, Alfonso Valencia⁹⁰, Heinz Himmelbaue⁵⁸, Victor Quesada⁶⁶, Silvia Bea¹⁶⁷; **Breast cancer (triple negative/lobular/other) (United Kingdom)** Michael R. Stratton²² (Leader), P. Andrew Futreal²², Peter J. Campbell^{22,62}, Anne Vincent-Salomon¹⁴⁰, Andrea L. Richardson^{135,136}, Jorge S. Reis-Filho¹³⁴, Marc van de Vijver^{55,56}, Gilles Thomas^{93,94}, Jocelyne D. Masson-Jacquemier⁴⁰, Samuel Aparicio¹⁶⁹, Ake Borg¹²⁹, Anne-Lise Børresen-Dale^{130,131}, Carlos Caldas¹³², John A. Foekens¹³³, Hendrik G. Stunnenberg⁷⁰, Laura van't Veer⁵⁵, Douglas F. Easton¹⁷⁰, Paul T. Spellman⁷⁷, Sancha Martin²²; **The Cancer Genome Atlas (United States)** Anna D. Barker⁵, Lynda Chin^{41,42}, Francis S. Collins¹⁷¹, Carolyn C. Compton⁵, Martin L. Ferguson⁴⁶, Daniela S. Gerhard⁵, Gad Getz⁸¹, Chris Gunter⁹⁶, Alan Guttmacher¹¹, Mark Guyer¹², D. Neil Hayes⁴⁸, Eric S. Lander⁸¹, Brad Ozenberger¹², Robert Penny⁵⁰, Jane Peterson¹², Chris Sander⁸⁶, Kenna M. Shaw⁵, Terence P. Speed^{87,88}, Paul T. Spellman⁷⁷, Joseph G. Vockley⁵, David A. Wheeler⁹¹, Richard K. Wilson^{72,73}

Initial scientific planning committee Thomas J. Hudson^{1,2} (Chairperson), Lynda Chin^{41,42}, Bartha M. Knoppers²⁶, Eric S. Lander⁸¹, Peter Lichter³⁷, Lincoln D. Stein¹, Michael R. Stratton²², Warwick Anderson³, Anna D. Barker⁵, Cindy Bell⁶, Martin Bobrow²⁷, Wylie Burke¹⁷², Francis S. Collins¹⁷¹, Carolyn C. Compton⁵, Ronald A. DePinho^{41,115}, Douglas F. Easton¹⁷⁰, P. Andrew Futreal²², Daniela S. Gerhard⁵, Anthony R. Green¹⁷³, Mark Guyer¹², Stanley R. Hamilton¹⁷⁴, Tim J. Hubbard²², Olli P. Kallioniemi¹⁷⁵, Karen L. Kennedy²², Timothy J. Ley^{72,176}, Edison T. Liu⁶⁹, Youyong Lu¹⁸, Partha Majumder⁶⁷, Marco Marra⁶⁵, Brad Ozenberger¹², Jane Peterson¹², Alan J. Schafer²⁰, Paul T. Spellman⁷⁷, Hendrik G. Stunnenberg⁷⁰, Brandon J. Wainwright¹⁷⁷, Richard K. Wilson^{72,73}, Huanming Yang²⁴

¹Ontario Institute for Cancer Research, Toronto, Ontario M5G 0A3, Canada.

²Departments of Medical Biophysics and Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A1, Canada. ³National Health and Medical Research Council, Canberra, Australian Capital Territory 2601, Australia. ⁴Project Management Agency, German Aerospace Center (DLR), 53175 Bonn, Germany. ⁵National Cancer Institute, US National Institutes of Health, Bethesda, Maryland 20892, USA. ⁶Genome Canada, Ottawa, Ontario K2P 1P1, Canada. ⁷Secretariat of State for Research, Ministry of Science and Innovation, 28027 Madrid, Spain. ⁸Department of Biotechnology, Ministry of Science & Technology, Government of India, New Delhi, Delhi 110003, India. ⁹Institut National du Cancer, 92513 Boulogne-Billancourt, France. ¹⁰Genomics and Systems Biology Unit, Health Research Directorate, European Commission, B-1049 Brussels, Belgium. ¹¹Eunice Kennedy Shriver National Institute of Child Health and Human Development, US National Institutes of Health, Bethesda, Maryland 20892, USA. ¹²National Human Genome Research Institute, US National Institutes of Health, Bethesda, Maryland 20892, USA. ¹³Cancer Research UK, London WC2A 3PX, UK. ¹⁴Sidra Medical and Research Center, Qatar Foundation, Doha, Qatar. ¹⁵Department of Clinical Pharmacology, University of Oxford, Oxford OX2 6HE, UK. ¹⁶National Institute of Biomedical Innovation, Ibaraki, Osaka 567-0085, Japan. ¹⁷Division of Molecular Life Sciences, Federal Ministry of Education and Research, 11055 Berlin, Germany. ¹⁸Beijing Cancer Institute and Hospital, Peking University School of Oncology, 100142 Beijing, China. ¹⁹German Cancer Aid, 53113 Bonn, Germany. ²⁰Wellcome Trust, London NW1 2BE, UK. ²¹National Cancer Center Research Institute, Chuo-ku, Tokyo 104-0045, Japan. ²²Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK. ²³Yokohama Institute, RIKEN, Yokohama, Kanagawa 230-0045, Japan. ²⁴BGI-Shenzhen, Shenzhen, 518083 Guangdong, China. ²⁵The Hong Kong University of Science and Technology, Hong Kong, China. ²⁶Centre of Genomics and Policy, McGill University and Genome Québec Innovation Centre, Montreal, Québec H3A 1A4, Canada. ²⁷Department of Medical Genetics, Cambridge Institute for Medical Research, University of Cambridge, Cambridge CB2 0XY, UK. ²⁸U558, INSERM, 31073 Toulouse, France. ²⁹University of North Carolina School of Pharmacy, Division of Pharmaceutical Outcomes and Policy, Institute for Pharmacogenomics and Individualized Therapy, Chapel Hill, North Carolina 27599, USA. ³⁰Institute for Research in Humanities, Graduate School of Biostudies, Institute for Integrated Cell-Material Sciences, Kyoto University, Kyoto, Kyoto 606-8501, Japan. ³¹Inter-University Chair in Law and the Human Genome, University of Deusto, Bilbao, 48007 Bizkaia, Spain. ³²The Ethox Centre, University of Oxford, Oxford OX3 7LF, UK. ³³Department of Genetics, Case Western Reserve University, Cleveland, Ohio 44106, USA. ³⁴Center for Human Genetics, University Hospitals Case Medical Center, Cleveland, Ohio 44106, USA. ³⁵St John of God Pathology, Subiaco, Western

Australia 6008, Australia.³⁶Schools of Surgery and Pathology and Laboratory Medicine, The University of Western Australia, Nedlands, Western Australia 6009, Australia.

³⁷German Cancer Research Center, 69120 Heidelberg, Germany. ³⁸Garvan Institute of Medical Research, University of New South Wales, Darlinghurst, Sydney, New South Wales 2010, Australia. ³⁹Department of Surgery, Bankstown Hospital, Bankstown, Sydney, New South Wales 2200, Australia. ⁴⁰Institut Paoli-Calmettes, 13273 Marseille, France. ⁴¹Belfer Institute for Applied Cancer Science, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. ⁴²Department of Dermatology, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁴³U991, INSERM, 35043 Rennes, France. ⁴⁴Department of Hematology, Centro de Investigación del Cáncer, Hospital Universitario, Universidad de Salamanca, 37007 Salamanca, Spain. ⁴⁵Hôpital Beaujon, 92110 Clichy, France. ⁴⁶MLF Consulting, Arlington, Massachusetts 02474, USA. ⁴⁷Canadian Tumour Repository Network, Winnipeg, Manitoba R3M 0V5, Canada. ⁴⁸Department of Internal Medicine, Division of Medical Oncology, Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, North Carolina 27599, USA. ⁴⁹Center for Genomic Medicine, RIKEN, Yokohama, Kanagawa 230-0045, Japan. ⁵⁰International Genomics Consortium, Phoenix, Arizona 85004, USA. ⁵¹Molecular Pathology Programme, Spanish National Cancer Research Centre (CNIO), 28029 Madrid, Spain. ⁵²Advanced Centre for Treatment, Research and Education in Cancer, Tata Memorial Centre, Kharghar, Navi Mumbai, Maharashtra 410210, India. ⁵³Department of Pathology, University of Verona, 37134 Verona, Italy. ⁵⁴Center for Applied Research on Cancer (ARC-NET), Verona University Hospital, 37134 Verona, Italy. ⁵⁵Netherlands Cancer Institute, 1066 CX Amsterdam, The Netherlands. ⁵⁶Academic Medical Center, 1015 AZ Amsterdam, The Netherlands. ⁵⁷Research Center for Advanced Science and Technology, University of Tokyo, Meguro-ku, Tokyo 153-8904, Japan. ⁵⁸Center for Genomic Regulation, Pompeu Fabra University, 08003 Barcelona, Spain. ⁵⁹Public Health and Epidemiology Network Biomedical Research Center (CIBERESP), Barcelona, 08003 Catalonia, Spain. ⁶⁰Peter MacCallum Cancer Centre, Melbourne, Victoria 3002, Australia. ⁶¹Department of Biochemistry and Molecular Biology, University of Melbourne, Parkville, Victoria 3010, Australia. ⁶²Department of Haematology, University of Cambridge, Cambridge CB2 2XY, UK. ⁶³Queensland Centre for Medical Genomics, Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland 4067, Australia. ⁶⁴CEA/DSV/IG-Centre National de Genotypage, 91057 Evry, France. ⁶⁵Canada's Michael Smith Genome Sciences Centre, BC Cancer Agency, Vancouver, British Columbia V5Z 1L3, Canada. ⁶⁶Departamento de Bioquímica y Biología Molecular, Instituto Universitario de Oncología, Universidad de Oviedo, 33006 Oviedo, Spain. ⁶⁷National Institute of Biomedical Genomics, Kalyani, West Bengal 741251, India. ⁶⁸Department of Medical Biophysics, University of Toronto, Toronto, Ontario M5S 1A1, Canada. ⁶⁹Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore 138672, Singapore. ⁷⁰Nijmegen Centre for Molecular Life Sciences, Radboud University Nijmegen, 6500 HB Nijmegen, The Netherlands. ⁷¹Ludwig Center for Cancer Genetics and Therapeutics, Johns Hopkins Kimmel Cancer Center, Baltimore, Maryland 21231, USA. ⁷²The Genome Center, Washington University School of Medicine, St. Louis, Missouri 63108, USA. ⁷³Siteman Cancer Center, Washington University School of Medicine, St. Louis, Missouri 63108, USA. ⁷⁴Applied Genomics Center, Fok Ying Tung Graduate School, HKUST, Hong Kong, China. ⁷⁵Department of Biochemistry, The Hong Kong University of Science and Technology, Hong Kong, China. ⁷⁶Cancer Institute, Zhejiang University, 310009 Hangzhou, China. ⁷⁷Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94510, USA. ⁷⁸Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario M5S 3E1, Canada. ⁷⁹Banting and Best Department of Medical Research, University of Toronto, Toronto, Ontario M5S 3E1, Canada. ⁸⁰European Molecular Biology Laboratory-European Bioinformatics Institute, Hinxton, Cambridge CB10 1SD, UK. ⁸¹Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA. ⁸²Spanish National Bioinformatics Institute (INB) and Center for Genomic Regulation, Universitat Pompeu Fabra, 08003 Barcelona, Spain. ⁸³Howard Hughes Medical Institute and Center for Biomolecular Science and Engineering, University of California Santa Cruz, Santa Cruz, California 95064, USA. ⁸⁴Research Unit on Biomedical Informatics, Department of Experimental and Health Science, Pompeu Fabra University, 08003 Barcelona, Spain. ⁸⁵Department of Computer Science & Center for Computational Molecular Biology, Brown University, Providence, Rhode Island 02912, USA. ⁸⁶Computational Biology Center, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. ⁸⁷Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria 3052, Australia. ⁸⁸Department of Statistics, University of California Berkeley, Berkeley, California 94720, USA. ⁸⁹Department of Biomolecular Engineering, University of California Santa Cruz, Santa Cruz, California 95064, USA. ⁹⁰Spanish National Bioinformatics Institute (INB) and Structural Biology and Biocomputing Programme, Spanish National Cancer Research Centre (CNIO), 28029 Madrid, Spain. ⁹¹Human Genome Sequencing Center & Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA. ⁹²Fondation Jean Dausset, Centre d'Etude du Polymorphisme Humain, 75010 Paris, France. ⁹³Université Claude Bernard Lyon 1, 69622 Villeurbanne, France. ⁹⁴Fondation Synergie Lyon Cancer, 69008 Lyon, France. ⁹⁵Nature Genetics, New York, New York 10013-1917, USA. ⁹⁶HudsonAlpha Institute for Biotechnology, Huntsville, Alabama 35806, USA. ⁹⁷Nature and the Nature research journals, New York, New York 10013, USA. ⁹⁸Computer Laboratory, University of Cambridge, Cambridge CB3 0FD, UK. ⁹⁹Faculty of Law, University of Tasmania, Hobart, Tasmania 7001, Australia. ¹⁰⁰Faculty of Law, National University of Singapore, Singapore 259776, Singapore. ¹⁰¹Consultant in Health Research Ethics and Policy, 34280 La Grande Motte, France. ¹⁰²ISIS 39 rue Camille Desmoulins, Institut Gustav Roussy, Pediatric Sce, 94805 Villejuif, France. ¹⁰³Department of Gynaecological Oncology, Westmead Hospital, Westmead, Sydney, New South Wales 2145, Australia. ¹⁰⁴Westmead Institute for Cancer Research, University of Sydney at the Westmead Millennium Institute, Westmead, Sydney, New South Wales 2145, Australia. ¹⁰⁵Sol Goldman Pancreatic Cancer Research Center, Johns Hopkins Medical Institutions, Baltimore, Maryland 21231, USA. ¹⁰⁶Department of Anatomical Pathology, Royal Prince Alfred Hospital, University of Sydney, Camperdown, Sydney, New South Wales 2050, Australia. ¹⁰⁷Helen Diller Family Comprehensive Cancer Center, University of California San Francisco, San Francisco, California 94115, USA. ¹⁰⁸Department of General Surgery, Toronto General Hospital, Toronto, ON M5G 2C4, Canada. ¹⁰⁹Samuel Lunenfeld Research Institute, Toronto, Ontario M5S 1A1, Canada. ¹¹⁰Ontario Cancer Institute, University Health Network, Toronto, Ontario M5G 2M9, Canada. ¹¹¹Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario M5S 1A1, Canada. ¹¹²Department of Pathology, University Health Network, Toronto, Ontario M5G 2C4, Canada. ¹¹³Department of Health Science Research, Mayo Clinic, Rochester, Minnesota 55905, USA. ¹¹⁴Department of Biochemistry and Molecular Biology, Mayo Clinic, Rochester, Minnesota 55905, USA. ¹¹⁵Department of Medicine and Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. ¹¹⁶Department of Surgery, Harvard Medical School, Boston, Massachusetts 02115, USA. ¹¹⁷Shanghai Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China. ¹¹⁸Institute of Digestive Disease, The Chinese University of Hong Kong, Hong Kong, China. ¹¹⁹International Agency for Research on Cancer, 69372 Lyon, France. ¹²⁰Institute of Hygiene and Epidemiology, First Faculty of Medicine, Charles University in Prague, 121 08 Prague, Czech Republic. ¹²¹Department of Epidemiology and Prevention, N. N. Blokhin Russian Cancer Research Centre, Moscow 115478, Russian Federation. ¹²²Karolinska Institutet, Karolinska University Hospital, SE-171 76 Stockholm, Sweden. ¹²³Bioengineering Center, Russian Academy of Sciences, Moscow 117312, Russian Federation. ¹²⁴Cancer Research UK Centre, Leeds Institute for Molecular Medicine, St James's University Hospital, Leeds LS9 7TF, UK. ¹²⁵Science for Life Laboratory, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden. ¹²⁶Institute of Mathematics and Computer Science, University of Latvia, Riga LV-1459, Latvia. ¹²⁷Uppsala University, SE-751 05 Uppsala, Sweden. ¹²⁸Kurchatov Scientific Center, Moscow 123182, Russian Federation. ¹²⁹Department of Oncology, Lund University, SE-221 85 Lund, Sweden. ¹³⁰Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, 0310 Oslo, Norway. ¹³¹Faculty of Medicine, University of Oslo, 0316 Oslo, Norway. ¹³²Department of Oncology, University of Cambridge and Cancer Research UK Cambridge Research Institute, Li Ka Shing Centre, Cambridge CB2 0RE, UK. ¹³³Department of Medical Oncology, Erasmus MC Rotterdam, Josephine Nefkens Institute and Cancer Genomics Centre, 3015 CE Rotterdam, The Netherlands. ¹³⁴Breakthrough Breast Cancer Research Centre, Institute of Cancer Research, London SW3 6JB, UK. ¹³⁵Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. ¹³⁶Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. ¹³⁷Jules Bordet Institute, B-1000 Brussels, Belgium. ¹³⁸Centre Léon Bérard, 69373 Lyon, France. ¹³⁹Hôpital Jean Minjot, 25030 Besançon, France. ¹⁴⁰Institut Curie, 75231 Paris, France. ¹⁴¹Centre Val d'Aurelle Paul-Lamarque, 34298 Montpellier, France. ¹⁴²Hôpital Pellegrin, 33076 Bordeaux, France. ¹⁴³Hôpital Henri Mondor, 94010 Créteil, France. ¹⁴⁴U955, INSERM, 94000 Créteil, France. ¹⁴⁵Hôpital Antoine Bécélère, 92141 Clamart, France. ¹⁴⁶Centre Hepato-Bilaire, AP-HP Hôpital Paul-Brousse, 94800 Villejuif, France. ¹⁴⁷U785, INSERM, 94800 Villejuif, France. ¹⁴⁸U674, INSERM, 75010 Paris, France. ¹⁴⁹BioQuant, Heidelberg University, 69120 Heidelberg, Germany. ¹⁵⁰Genome Biology Unit, European Molecular Biology Laboratory, 69126 Heidelberg, Germany. ¹⁵¹Department of Neuropathology, Heidelberg University Hospital, 69120 Heidelberg, Germany. ¹⁵²Clinic for Pediatric Oncology, Hematology and Immunology, Heinrich-Heine University Hospital, 40225 Düsseldorf, Germany. ¹⁵³Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany. ¹⁵⁴Department of Pediatric Hematology and Oncology, Heidelberg University Hospital, 69120 Heidelberg, Germany. ¹⁵⁵Institute of Neuropathology, Heinrich-Heine University, 40001 Düsseldorf, Germany. ¹⁵⁶Division of Neurosurgery, Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. ¹⁵⁷The Arthur and Sonia Labatt Brain Tumour Research Centre, Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. ¹⁵⁸National Center for Tumor Diseases, 69120 Heidelberg, Germany. ¹⁵⁹Division of Translational Oncology, German Cancer Research Center, 69120 Heidelberg, Germany. ¹⁶⁰Department of Surgery, University Hospital Trust of Verona, 37134 Verona, Italy. ¹⁶¹Functional Genomics Center, Department of Biotechnology, University of Verona, 37134 Verona, Italy. ¹⁶²Laboratory of Molecular Genetics, Institute for Cancer Research and Treatment, University of Torino, 10060 Torino, Italy. ¹⁶³FIRC Institute of Molecular Oncology, 20139 Milan, Italy. ¹⁶⁴Department of Gastroenterology, Endocrinology, Metabolism and Infectiology, University of Marburg, 35043 Marburg, Germany. ¹⁶⁵Department of Pathology, Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. ¹⁶⁶Human Genome Center, Institute of Medical Science, University of Tokyo, Minato-ku, Tokyo 108-8639, Japan. ¹⁶⁷Hospital Clinic, University of Barcelona, 08036 Barcelona, Spain. ¹⁶⁸Unit of Infections and Cancer, Cancer Epidemiology Research Programme, CIBER Epidemiología y Salud Pública, Institut Català d'Oncologia-IDIBELL, 08907 Hospitalet de Llobregat, Spain. ¹⁶⁹BC Cancer Research Centre, BC Cancer Agency, Vancouver, British Columbia V5Z 1L3, Canada. ¹⁷⁰Departments of Public Health and Primary Care and Oncology, University of Cambridge, Cambridge CB1 8RN, UK. ¹⁷¹US National Institutes of Health, Bethesda, Maryland 20892, USA. ¹⁷²Department of Bioethics and Humanities, University of Washington, Seattle, Washington 98195, USA. ¹⁷³Cambridge Institute for Medical Research and Department of Haematology, University of Cambridge, Cambridge CB2 2XY, UK. ¹⁷⁴Pathology and Laboratory Medicine, The University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030, USA. ¹⁷⁵Institute for Molecular Medicine Finland, University of Helsinki, FIN-00290 Helsinki, Finland. ¹⁷⁶Departments of Medicine and Genetics, Washington University School of Medicine, St. Louis, Missouri 63110, USA. ¹⁷⁷Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland 4072, Australia.