



## Explanatory notes

### 1. Code of practice

By submitting a completed abstract form, the named investigators are committing to the Code of Practice found on the concluding 2 pages of this document. Note, in particular, that there will be an embargo on publishing results of pan-cancer analyses until at least May 2015, allowing groups involved in single tumour-type sequencing to complete their analyses.

### 2. Filling in the abstract form

Do not exceed the two pages allocated for the abstract. Please outline in the 'background and preliminary data' section the question(s) that you wish to address with the analysis, how the pan-cancer data set will enable these analyses and what steps you might already have taken to address these questions in smaller data sets. The 'timelines and resources' section should outline any key intermediate milestones you will reach during the analysis, and if there are any analyses or data sets from the pan-cancer studies that you would be particularly dependent on. The 'research proposal' section should broadly outline the plan of investigation you will undertake (we accept that this will potentially change as the project develops). The 'legacy plans' section should address the plans you have to make any software, algorithms, visualization tools, statistical analyses available to the research community.

### 3. Data sets

What will be available for scientific working groups:

- a. ~2,000 whole genome sequence pairs, matched tumour and normal, quality filtered, aligned to hg19 and run through three core variant-calling pipelines for base substitutions, indels, copy number, structural variants and retrotranspositions (available as VCF files). Preliminary breakdown of case numbers is appended.
- b. ~200 of the 2,000 WGS cases will have additional cancer genomes sequenced (eg primary & metastasis; diagnosis and progression etc).
- c. RNA-sequencing or cDNA microarray data matched to WGS for ~1,500 cancer samples.
- d. Methylation studies matched to WGS, performed by a variety of different assays, for ~1200-1400 cancer samples.
- e. Additional matched tumour and normal exome pairs, aligned to hg19 (but not run through central variant calling). TCGA exome pairs available for this project include only samples matched to WGS.
- f. Clinical data.

### 4. Cloud computing

Data will be available in a series of cloud compute centres, and we expect most analytic groups to perform their work on cloud-based virtual machines within these centres.

### 5. Deadlines

Abstract submission = Wednesday 27<sup>th</sup> November 2013, 5pm your local time.  
Submit to Jennifer Jennings: [Jennifer.Jennings@oicr.on.ca](mailto:Jennifer.Jennings@oicr.on.ca)

### 6. Personnel

Principal investigators must be affiliated with an ICGC or TCGA tumour type-specific working group or a bioinformatics working group (such as BAWG for ICGC). Junior investigators would typically be graduate or post-doctoral scientists working in the PI's lab. PIs have to apply for junior investigators to access the data or their applications will be rejected. DACO does not accept application from students or post-doc. Investigators not affiliated with ICGC or TCGA can be named as collaborators where their expertise is considered important to the delivery of the research proposal. No more than two PIs and

two collaborators can be named.

**7. Process for choosing scientific working groups**

The scientific working groups will be chosen through an abstract competition judged by a panel of reviewers comprising the WGS pan-cancer steering committee (5 members) and 3 ex officio nominees from the pan-cancer oversight committee. All panel members will rank abstracts for scientific merit and these will be used to guide selection of scientific working groups. The panel reserves the right to re-engineer scientific projects based on complementarity of approaches and broad thematic research questions.

**8. Essential vs exploratory projects**

Approved projects will be divided into essential or exploratory categories. Essential projects represent those which are central to the mission of the WGS pan-cancer project and those which generate data that other projects will depend on. Exploratory projects represent those that are more speculative and/or high risk.

**9. Broad themes of essential projects**

The primary focus of the pan-cancer analysis is to understand the differences and similarities across the different tumour types in cancer genome landscapes. There are of course a number of scientific questions that would be interesting to address with this data set. While by no means exhaustive, we believe the following themes will underpin some of the essential projects:

- a. *Variant calling*
- b. *Functional consequences of non-coding mutation*
- c. *Integration of genome and transcriptome*
- d. *Integration of genome and epigenome*
- e. *Pathway analysis*
- f. *Genomic rearrangement architecture*
- g. *Mutation signatures*
- h. *Interface between germline and somatic genetics*
- i. *Landscape of driver mutations*
- j. *Clinical correlations*
- k. *Temporal evolution of cancer genomes*

**10. Expectations of scientific working groups**

The PIs of the scientific working groups will be expected to attend joint hour-long teleconference calls every 2-3 weeks and present verbal progress reports at these. In addition, the PIs (and/or junior investigators) will be expected to attend the ICGC meetings (China, May 2014; Italy, Feb 2015) and two pan-cancer retreats (March, 2014; Nov, 2014). Written progress reports and oral presentations will be required for these four face-to-face meetings.

**11. Manuscripts**

As the projects take shape, the direction of potential manuscripts should become clearer, and may necessitate re-engineering of the original plans. We expect open sharing of draft manuscripts as the deadline for co-ordinated submission approaches. There will also be a marker paper that extracts and summarises key metrics of the WGS pan-cancer analysis. About 6 months before submission date, we will collate the extended research abstracts for presubmission enquiries with journals that might be interested.

**12. Data freezes**

Once the 'essential' scientific working groups have been identified, we will undertake a dependency analysis in which we determine which pieces of analysis are dependent on prior analyses. This will be used to assign phased data freeze dates.

**13. Broad timelines**

- a. Pilot analysis in computing centres (core technical group) – November 15<sup>th</sup> 2013
- b. Abstract submission – 27<sup>th</sup> November, 2013
- c. Scientific working groups identified 10<sup>th</sup> December, 2013
- d. Genome uploads – November 2013 to June 2014
- e. Core variant calling – December 2013 to September 2014
- f. Scientific analyses – November 2013 to December 2014
- g. Manuscript preparation – January to February 2015
- h. Manuscript submission – 20<sup>th</sup> March 2015.

**14. Legacy plans**

We expect that the computational steps used to produce publication-ready results on the pan-cancer data set will be embodied in executable code and sufficiently well documented to enable replication by third parties. For example, a documented virtual machine image or RStudio / knitr interwoven code + output + explanatory text.

**15. Prioritisation**

The steering group and core technical group will monitor computer usage of scientific working groups. Where compute capacity becomes limiting, the steering committee may assign the 'essential' scientific working groups a higher priority for this resource than the 'exploratory' groups.

**16. Handling of potential conflicts of interest**

We anticipate that there may arise situations in which a pan-cancer working group makes a discovery about a single tumour type. In such situations, we will expect the scientific working group to report this to the steering committee, and establish a dialogue with the appropriate tumour type working group.

**17. Intellectual property rights**

Participants agree to abide by the IP policies of the ICGC and TCGA. These policies are listed here:

ICGC IP policy:

<http://www.icgc.org/icgc/goals-structure-policies-guidelines/e4-intellectual-property-policy>

TCGA IP policy:

[http://cancergenome.nih.gov/pdfs/Data Use Certv082014](http://cancergenome.nih.gov/pdfs/Data_Use_Certv082014)

**18. Companion methodology papers**

We anticipate that there will be significant methodology development alongside the scientific development. As interwoven code and virtual machines will be generated as part of this project, this provides the opportunity to publish companion methodology papers alongside the scientific papers. We will address whether, for example, Nature Methods or Genome Research might be interested in a co-ordinated release of these alongside the results papers. Provided that there is no large-scale release of data or results, it may be feasible to publish the methodology papers early with agreement from the full scientific working group.

**Estimates of the number and distribution of whole genome cases (estimated in March 2013). Note that these are estimates only, and numbers will change.**

<b>Disease type</b>	<b>Centres</b>	<b>Number of patients</b>
Pancreatic cancer	Australia	50-100 (some triplets)
	Canada	100
	TCGA	50
Prostate cancer	Canada	100-150
	UK	30 (incl some triplets)
	TCGA	50
	Germany	50
Gastric cancer	TCGA	50
	China	50
Liver cancer	Japan	300
	France	?
	TCGA	50
Paediatric brain cancer	Germany	250
Lymphoma	Germany	80
Oral cancer	India	25
Acute myeloid leukaemia	TCGA	50
	South Korea	10
	Spain	100 (incl some triplets)
Chronic lymphocytic leukaemia	UK & France	200 (incl some triplets)
	TCGA	50
	South Korea	100
Chronic myeloid cancers	UK	40-50 (all triplets)
Bone cancer	UK	50
Oesophageal cancer	UK	75-100
Ovarian cancer	TCGA	50
	Australia	50 (incl some triplets)
Glioma	TCGA	100
Endometrial cancer	TCGA	50
Head & neck cancer	TCGA	50
Non-small cell lung cancer	TCGA	100
Melanoma	TCGA	50
Colorectal cancer	TCGA	50
Bladder cancer	TCGA	50
Kidney cancer	TCGA	50
Cervical cancer	TCGA	50

## Code of practice for members of WGS pan-cancer studies

I/We will adhere to the following principles for the ICGC/TCGA WGS pan-cancer project:

**1. Confidentiality**

We are committed to maintaining confidentiality of all discussions of the scientific working groups for the WGS pan-cancer project. Project proposals and progress reports will be kept confidential in accordance with good scientific practice. We have signed the Confidential Disclosure Agreement for the WGS pan-cancer project.

**2. Ethics**

We will respect the ethical standards applicable for analysis of personal genome data.

**3. Conflict of interest**

We will seek advice from the WGS pan-cancer steering committee on potential conflicts of interest, especially among scientific working groups of the pan-cancer analysis or between a pan-cancer analysis group and a tumour type-specific expert group. We understand these conflicts of interest will be managed on a case-by-case basis.

**4. Intellectual property rights**

Participants agree to abide by the IP policies of the ICGC and TCGA. These policies are listed here:

ICGC IP policy:

<http://www.icgc.org/icgc/goals-structure-policies-guidelines/e4-intellectual-property-policy>

TCGA IP policy: [http://cancergenome.nih.gov/pdfs/Data\\_Use\\_Certv082014](http://cancergenome.nih.gov/pdfs/Data_Use_Certv082014)

**5. Quality assurance**

We maintain quality assurance through constant review of all aims, activities and outcomes of our research. We commit to personally attending (or sending a delegate for) the conference calls, the two face-to-face meetings and the ICGC meetings. We will provide regular progress reports as stipulated by the steering committee.

**6. Data protection**

We will abide by the terms, conditions and spirit of the ICGC/DACO and TCGA/dbGAP data access agreements involving the confidentiality and privacy of data against re-identification by third parties.

**7. Publication and embargo policy**

The WGS pan-cancer project is committed to the principle of rapid distribution of primary data, alignments files, and secondary analysis data (such as mutation calls) to all members of the ICGC/TCGA community that have applied for (and have become approved for) data access. The primary data contributed to the project will fall under any publication/embargo policy of the contributing institute or consortia (see item 8 below). The community resource for the ICGC/TCGA PCAWG project comprises newly generated files produced by this collaboration (unified alignments, mutation calls and any secondary analysis data). As is the case for similar "community resource projects", users of data generated and distributed by the WGS pan-cancer data are asked to respect the desire of the pan-cancer consortium to publish reports on the generation and analysis of their data. Hence, scientific works with a primary focus on pan-cancer (across entity) analyses that made use of the WGS pan cancer community resource will be under publication embargo until the WGS pan-cancer consortium publishes its main paper (anticipated May 2015). Methodology papers may be published prior to this embargo, with agreement from the full scientific working group, as discussed under item 18 above.

8. **Policy for unpublished cancer genome data contributed to the WGS pan-cancer effort**

The original whole genome sequencing data or any other original data that is contributed to this project (i.e. data that is not newly generated as part of the PCAWGs resource described in item 7 above) do not fall under the publication and embargo policy.

Contributing groups can produce analysis and publications using those original files, subject to any publication/embargo policy of the contributing institute or consortia. The contributing groups will have access to variant calls made by the pan-cancer technical group in a tumour type for analysis on a single disease entity (for example, prostate cancer-specific analysis) prior to the date the pan-cancer embargo is lifted. In return, data generators who contribute to this project agree not to demand withdrawal of samples for reasons of publication priority.

9. **Companion papers & authorship**

Associated WGS pan-cancer companion papers may be published back-to-back with the WGS pan-cancer consortium main paper, or at any time after the publication embargo has lifted. WGS pan-cancer companion papers that make use of the pan-cancer resources are asked to mention within their author list that the work has been performed on behalf of the WGS pan-cancer consortium (example: Smith J & Doe J, on behalf of the WGS pan cancer consortium). The individual tumour-specific ICGC/TCGA groups that have contributed data to the pan-cancer project should be mentioned in the acknowledgements or supplementary information of all manuscripts. The main WGS pan-cancer paper will explicitly list pan-cancer and tumour type-specific contributors as authors.