Analysis of five years of controlled access and data sharing compliance at the International Cancer Genome Consortium

To the Editor:

The broad sharing of data from large-scale genomics research and related initiatives has contributed significantly to advances in human genetic research. However, questions have been raised about whether these data sharing practices carry potential privacy risks for research participants[1]. These concerns will likely become more acute in the coming months given the European Court of Justice's recent decision that invalidated data sharing between the European Union and the United States under the Safe Harbour framework[2], although it is still being debated whether this is applicable to academic research. One approach proposed to enhance data privacy requires that users complete an application form, sign a data sharing agreement and receive authorization from a Data Access Committee (DAC) to use the data[3]. However, some members of the scientific community have criticized this 'controlled' approach as overly bureaucratic and time-consuming. To shed some light on this important debate, we have gathered data from 840 access requests processed over the last five years of operation at the International Cancer Genome Consortium (ICGC) Data Access Compliance Office (DACO).

Since the launch of DACO's website in July 2010, the number of access requests has progressively risen from an average of fewer than 2 applications a month (1.42) in its first complete year of operations in 2011 to an average of approximately 31.3 applications per month in 2014. Despite this increase, by streamlining its procedures, DACO was able to reduce the average processing time for new applications from ten days in 2011 to seven days in 2014. Once an applicant is approved, it then takes approximately 24 hours to gain access to the controlled data through the ICGC Data Coordination Centre (DCC) portal at the Ontario Institute for Cancer Research (OICR) and through the European Genome-phenome Archive (EGA). Access to raw sequence reads (consisting of BAM files) held at EGA can take up to three business days. Applications with minor content errors are conditionally approved, which can result in short additional delays until the missing information is supplied by the applicant via the online system. Infrequently, applications with substantial problems are ruled incomplete and must be resubmitted entirely with new signatures and online validation, which can result in a significant delay for an applicant. Nevertheless, it should be noted that resubmitted applications are speedily processed, in an average of four days. As of 31 October 2015, DACO had 975 approved users belonging to 175 different projects. The users' research laboratories were mostly located in North America (49%), Europe (32%) and Asia (14%). Over the years, there have also been a small number of applications turned down, most frequently because the main applicant's professional credentials were not those of a scientific researcher.

Complaints about delays associated with DACO have been very rare, with fewer than six cases identified during the review period. No incidents involving the reidentification of ICGC participants, misuse of data or breaches following approval of other terms of the DACO agreement have been reported. A limited number of minor internal security concerns have been identified by members of the ICGC community and swiftly addressed by DACO, the DCC and the ICGC executive committee.

The DACO application process, including the application form and online submission system, has continually evolved, and it is now more accessible and relatively convenient to complete[3]. The most recent change, still ongoing, regards the adaptation of the application form to enable cloud computing access to the controlled data. On the basis of compiled statistics or revised applications and users' queries, the most onerous components of the application process are the need to obtain a valid OpenID to log into the controlled-access data portals and the requirement to obtain a signature from an official representative of the user's institution.

With the consent of its users, DACO currently posts a list of all approved projects on its website. We suggest that, by openly publishing the number of non-compliance incidents involving controlled data and the manner in which each of these incidents was addressed, research projects would have an opportunity to demonstrate to the community the security, professionalism and reliability of the controlled-access process. DACO's experience demonstrates that 'controlled access' can provide added protection to sensitive data without unduly burdening the research community. The costs and delays associated with this method are modest, and the application requirements have been generally well received by the applicants. Outstanding issues that will need to be addressed over the next few years include the harmonization of DAC's controlled-access requirements across omics research projects[4]; the development of efficient and transparent compliance and accountability frameworks; and the integration of novel bioinformatics tools developed to provide more secure environments for data sharing and analysis. In addition, the Global Alliance for Genomics and Health is currently developing a registered-access model that would fall between the open-access and controlled-access tiers. This new model appears well suited to data sharing contexts where the risk of reidentifying participants or the sensitivity of the data is low.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper (doi:10.1038/ng.3499).

ACKNOWLEDGMENTS
The authors would like to acknowledge the valuable contribution of the following individuals: (DACO consultants) M.H. Zawati, E. Kleiderman, S.O.M. Dyke, D. So and E. Dove, (IDAC members) A. Cambon-Thomsen, R. Yoshida, T. Kaan, P. Nicolas, M. Bobrow and L. Lyman Rodriguez, (OICR/ICGC Web Development team) M. Fukuma, K. Wu, D. Gross and S. Badr, (ICGC DCC) L. Stein, H. Nahaal and V. Ferretti, (ICGC Secretariat) J. Jennings and T.J. Hudson, (research assistance) S. Birko and P3G, and (financial support) FRSQ 30719 and OICR. [**AU: Text ok here at the end of the sentence?**]

COMPETING FINANCIAL INTERESTS
The authors declare competing financial interests: details are available in the online version of the paper.

Yann Joly[1], Emilie de Vries-Seguin[2], Don Chalmers[3], B F Francis Ouellette[4], Joseph Yamada[4], Martin Bobrow[5] & Bartha M Knoppers[1] for the ICGC Data Access Compliance Office and the ICGC International Data Access Committee[6]

*[1]Centre of Genomics and Policy, Faculty of Medicine, Department of Human Genetics, McGill University, Montreal, Quebec, Canada. [2]Public Population Project in Genomics and Society (P3G), Montreal, Quebec, Canada. [3]Faculty of Law, University of Tasmania, Tasmania, Australia. [**AU: Please give the city for affiliation 3.**] [4]Ontario Institute for Cancer Research, Toronto, Ontario, Canada. [5]Department of Medical Genetics, University of Cambridge, Cambridge, UK. [6]Full lists of members and affiliations appear in the* **Supplementary Note**.
e-mail: yann.joly@mcgill.ca
[AU: Per the affiliation note, we have moved the lists of consortium members to a separate supplementary file. Correct that this was intended? If members were meant to be listed in the main text (and as collaborators in PubMed), then numbered affiliations would need to be provided.]

1. Shabani, M., Knoppers, B.M. & Borry, P. *EMBO Mol. Med.* **7**, 507–509 (2015).
2. Rothstein, M.A. *Hastings Cent. Rep.* doi:10.1002/hast.539 (14 December 2015).
3. Joly, Y., Dove, E.S., Knoppers, B.M., Bobrow, M. & Chalmers, D. *PLOS Comput. Biol.* **8**, e1002549 (2012).
4. Kaye, J. & Hawkins, N. *Genome Med.* **6**, 4 (2014).