COMMENTARY

CLiViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer

Malachi Griffith1–4,11, Nicholas C Spies1,11, Kilannin Krysiak1,4,11, Joshua F McMichael1, Adam C Coffman1, Arpad M Danos1, Benjamin J Ainscough1–3, Cody A Ramirez1, Damian T Rieke5, Lynzey Kujan1, Erica K Barnell1, Alex H Wagner1,2, Zachary L Skidmore1, Amber Wollam1, Connor J Liu1, Martin R Jones6, Rachel L Bilski1, Robert Lesur1, Yan-Yang Feng1, Nakul M Shah1, Melika Bonakdar6, Lee Trani1, Matthew Matlock1, Avinash Ramu1, Katie M Campbell1, Gregory C Spies1, Aaron P Graubert1, Karthik Gangavarapu7, James M Eldred1, David E Larson1,3, Jason R Walker1, Benjamin M Good7, Chunlei Wu7, Andrew I Su7, Rodrigo Dienstmann8, Adam A Margolin9, David Tamborero10, Nuria Lopez-Bigas10, Steven J M Jones6, Ron Bose4, David H Spencer1,4, Lukas D Wartman1,2,4, Richard K Wilson1–4, Elaine R Mardis1–4 & Obi L Griffith1–4

CLiViC is an expert-crowdsourced knowledgebase for Clinical Interpretation of Variants in Cancer describing the therapeutic, prognostic, diagnostic and predisposing relevance of inherited and somatic variants of all types. CIViC is committed to open-source code, open-access content, public application programming interfaces (APIs) and provenance of supporting evidence to allow for the transparent creation of current and accurate variant interpretations for use in cancer precision medicine.

Understanding of the genetic heterogeneity and mutational landscape underlying cancer has seen incredible advances in recent years. This has accelerated the implementation of precision medicine strategies in which clinicians and researchers target specific molecular variants with treatments tailored to the individual and their disease1. The biomedical literature describing such associations is large and growing rapidly. As a result, the interpretation of individual variants observed in patients has become a bottleneck in clinical sequencing workflows2. Many cancer hospitals and research centers are engaged in separate efforts to interpret cancer-driving variants and genes in the context of clinical relevance. These efforts are largely occurring within independent ‘information silos’, producing interpretations that require constant updates, lack community consensus and involve intense manual input.

Estimates of the proportion of patients with cancer who would benefit from comprehensive molecular profiling vary substantially3, in part because of the lack of both a community consensus definition of actionability and a comprehensive catalog of specific clinical variant interpretations. Achieving the goals of precision medicine will require this information to be centralized, freely accessible, openly debated and accurately interpreted for application in the clinic. Existing efforts to facilitate clinical interpretation of variants include the Gene Drug Knowledge Database4, the Database of Curated Mutations5, ClinVar6, ClinGen7, PharmGKB8, Cancer Driver Log9, My Cancer Genome10, Jax-Clinical Knowledgebase11, the Personalized Cancer Therapy Knowledgebase, the Precision Medicine Knowledgebase, the Cancer Genome Interpreter, OncoKB and others (Supplementary Table 1). These resources often have barriers to widespread adoption, including some combination of (i) no public access to content, (ii) restrictive content licenses, (iii) no public API, (iv) no bulk data download capabilities and (v) no mechanism for rapid improvement of the content (see Supplementary Table 1 for detailed feature comparison). To address these limitations, we present CLiViC, an open-access, open-source knowledgebase for expert crowdsourcing of Clinical Interpretation of Variants in Cancer (http://civicdb.org/; Fig. 1).

The critical distinguishing features of the CLiViC initiative, in comparison to several of...
the resources cited above, stem from its strong commitment to openness and transparency. We believe that these principles (Box 1) are necessary for widespread adoption of such a resource. The target audience of CIViC is deliberately broad, encompassing researchers, clinicians and patient advocates. CIViC is designed to encourage development of community consensus by leveraging an interdisciplinary, international team of experts collaborating remotely within a centralized curation interface. Variant interpretations are created with a high degree of transparency and detailed provenance. The interface is designed to help keep interpretations current and comprehensive, and to acknowledge the efforts of content creators (Supplementary Fig. 1). CIViC accepts public knowledge contributions but requires that experts review these submissions.

The manner in which the clinical relevance of variants in cancer is presented in the published literature is highly heterogeneous. To represent these data in a more easily interpretable and consistent fashion, the CIViC data model is highly structured and ontology driven (Supplementary Fig. 2). Clinical interpretations are captured and displayed as evidence records consisting of a freeform ‘evidence statement’ and several structured attributes. Each evidence record is associated with a specific gene, variant, disease and clinical action. Evidence records belong to one of four evidence types indicating whether a variant is predictive of response to therapy, prognostic, diagnostic and/or predisposing for cancer. Evidence records are also assigned a level (A to E) to inferential significance (for example, imatinib resistance). The manner in which the clinical relevance of a variant in context of a patient is predictive of response to therapy, prognostic, diagnostic and/or predisposing for cancer. Evidence records are also assigned a level (A to E) to inferential significance (for example, imatinib resistance).
Box 1 CIViC principles

1. **Interdisciplinary.** An interdisciplinary approach is needed to combine the expertise of genome scientists, healthcare providers, patient advocates and others.

2. **Community consensus.** The interpretations of clinical actionability required to enable precision medicine should be freely available and openly discussed across a diverse community. To facilitate consensus building, the interface must support direct contribution from members of the community.

3. **Transparency.** Content should be created with transparency, kept current, be comprehensive, track provenance and acknowledge the efforts of its creators.

4. **Computationally accessible.** The interface should be both structured enough to allow computational data mining (via APIs) and agile enough to handle the product of openly debated human interpretation.

5. **Freely accessible.** Curated knowledge will remain free and can be accessed anonymously without login unless the user wishes to contribute to content. No fees will be introduced.

6. **Open license.** CIViC will encourage both academic and commercial engagement through flexible licensing. Access will not be restricted by exclusive licensing.
diseases\textsuperscript{21}, a field where the ACMG–AMP have proposed detailed standards and guidelines for variant classification\textsuperscript{22}. This report identified a low rate of interpretation agreement between laboratories (34% concordance). However, discussion and review of criteria were able to more than double this concordance, demonstrating the need for and success of open discourse in clinical variant interpretation\textsuperscript{21}. Recently, the Somatic Working Group (WG) of the Clinical Genome Resource (ClinGen) has published a consensus set of minimal variant-level data (MVLD) to help standardize data elements needed for curation of the clinical utility of somatic cancer variants\textsuperscript{23}. At present, cancer variant interpretation efforts that nominally have the same goals show a remarkably low overlap in source publications cited for these interpretations (1.6–71.6%, but generally less than 25%; Supplementary Table 2). This suggests that no single effort has comprehensively identified or summarized even the most relevant literature in this area, further illustrating the high curation burden involved. Conversely, these small overlaps emphasize the importance of reducing duplication of effort moving forward, especially considering the vastness of the existing literature and its tremendous growth rate. In CIViC, curation efforts thus far have focused on variants relevant to cancer types of particular interest at our center (for example, acute myeloid leukemia, breast cancer and lung cancer; Supplementary Fig. 8b), on variants identified as high priority by early CIViC partners\textsuperscript{4,19} and on variants targeted by proof-of-principle precision medicine ‘basket’ clinical trials such as NCI-MATCH (also known as EAY131 or NCT02465060). Our ability to provide expertise in these areas is complemented by the expert knowledge of other groups and organizations, making CIViC a more comprehensive resource than would be possible with a ‘siloed data’ approach. To this end, recruitment of external contributors and domain experts from multiple fields is a top priority. This is accomplished in part through planning of CIViC-sponsored events in the cancer research and treatment community. We also allow for different levels of external community involvement, including submission of suggested publications to a queue to guide others to generate new evidence records (Supplementary Fig. 14).

Additional challenges faced by CIViC include long-term sustainability of funding, ensuring broad clinical engagement and maintaining the enthusiasm for the crowdsourcing efforts. We are addressing each of these challenges by engagement with other resources, experts and funding agencies with track records of long-term maintenance of informatics resources (see the Supplementary Note for further discussion). To facilitate such engagement and seek broad input and external guidance for our efforts, we have recently established a Variant Interpretation for Cancer Consortium (VICC) as part of the Global Alliance for Genomics Health (GA4GH). We have also established a panel of clinical domain experts to provide independent guidance on development of the resource and to assess the completeness and accuracy of our variant curation efforts.

CIViC is designed to address many of the challenges of cancer variant interpretation. To our knowledge, CIViC is the only variant interpretation effort currently capable of leveraging community experts and additionally has the most open model (open-access content, open-source code and an open API). We believe that this open strategy represents a sustainable model for achieving current, standardized and comprehensive interpretations of the clinical relevance of cancer variants. As the community of contributors grows, an increased incentive will emerge to help keep CIViC updated with cutting-edge clinical trial and US Food and Drug Administration (FDA) investigational new drug (IND) findings. As we have created a comprehensive and modern API, centers can rapidly integrate CIViC into automated clinical report generation for gene panel, exome, whole-genome and RNA sequencing of tumor samples. While there are many challenges faced by an effort such as this one, we hope that, with input from a critical mass of interested parties, these challenges can be largely overcome. We invite all researchers, healthcare providers and patient advocates engaged in clinical interpretation of variants to join the community at CIViC (http://civicdb.org/).

**URLs.** The Clinical Interpretation of Variants in Cancer resource described by this work is available at http://www.civicdb.org/. Personalized Cancer Therapy Knowledgebase, https://pct.manderson.org/; Precision Medicine Knowledgebase, https://pmkb.weill.cornell.edu/; Cancer Genome Interpreter, https://www.cancergenomeinterpreter.org/; OncoKB, http://oncokb.org/; Github, https://github.com/; GA4GH Variant Interpretation for Cancer Consortium (VICC), http://ga4gh.org/#/vicc; MIT license, https://opensource.org/licenses/MIT; CC0 license, https://creativecommons.org/publicdomain/zero/1.0/.

**Note:** Any Supplementary Information and Source Data files are available in the online version of the paper (doi:10.1038/ng.3774).

**ACKNOWLEDGMENTS**

First and foremost, we are grateful to the community of curators, editors, domain experts and users who make CIViC possible. CIViC is supported by the National Cancer Institute (NCI) of the National Institutes of Health (NIH) under award number U01CA209936 to O.L.G. (with M.G. and E.R.M. as co-principal investigators). This work was also supported by a grant to R.K.W. from the National Human Genome Research Institute (NHGRI) of the NIH under award number U54HG003079. D.T.R. was supported by the German Federal Ministry of Education and Research under award numbers 031L0030E and 031L0023B. The Institute for Research in Biomedicine is a recipient of a Severo Ocho Centre of Excellence Award from MINECO (Government of Spain) and is supported by CERCA (Generalitat de Catalunya). D.T. is funded by a grant from the Spanish Ministry of Economy and Competitiveness and FEDER/UE (SAF2015-74072-JIN). M.G. is supported by the NHGRI under award number K99HG007940. O.L.G. is supported by the NCI under award number K22CA188163. L.D.W. is supported by the NCI under award number K08CA166229. The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

**AUTHOR CONTRIBUTIONS**

M.G., N.C.S., K.K. and O.L.G. wrote the paper with input from A.M.D., E.K.B., E.R.M., A.H.W., R.L., A.C.C., J.F.M., B.J.A., Z.L.S., C.A.R., C.J.L., D.T.R. and G.C.S. A.C.C led the back-end code development with contributions from N.C.S., J.F.M., M.G., O.L.G., K.K., G.C.S and K.G. J.F.M. led the front-end code development with contributions from A.P.G., A.C.C., K.K., M.G., O.L.G. and N.C.S. B.M.G., A.I.S. A.A.M., D.T., N.L.-B. and S.J.M.J. contributed ideas relating to crowdsourcing functionality and integration with existing community resources. J.M.E., D.E.L., C.W. and J.B.W. contributed software engineering expertise. Substantial curation efforts were contributed by M.G., N.C.S., K.K., A.M.D., B.J.A., C.A.R., D.T.R., I.K., E.K.R., A.H.W., Z.L.S., A.W., C.J.L., M.R.J., R.L., R.L., Y.-Y.F., N.M.S., M.B., L.T., M.M., A.K., K.M.C., R.D., R.B., D.H.S., L.D.W., E.R.M. and O.L.G. Guidance on developing CIViC for clinical applications was provided by R.D., R.B., D.H.S. and L.D.W. Trainee supervision and project leadership were provided by M.G., R.K.W., E.R.M. and O.L.G.

**COMPETING FINANCIAL INTERESTS**

The authors declare no competing financial interests.

1. Collins, F.S. & Varmus, H. A new initiative on precision medicine. *N. Engl. J. Med.* 372, 793–795 (2015).
2. Good, B., Ainscough, B.L., McMichael, D.F., Su, A.I. & Griffith, D.L. Organizing knowledge to enable personalized medicine in cancer. *Genome Biol.* 15, 438 (2014).
3. Marie-Beristain, F. et al. Feasibility of large-scale genomic testing to facilitate enrollment onto genomically matched clinical trials. *J. Clin. Oncol.* 33, 2753–2762 (2015).
4. Dienstmann, R., Jiang, I.S., Bot, B., Friend, S. & Guinney, J. Database of genomic biomarkers for cancer
Commentary

2016 update. Nucleic Acids Res. 44 D1, D717–D725 (2016).
19. Laskin, J. et al. Lessons learned from the application of whole-genome analysis to the treatment of patients with advanced cancers. Cold Spring Harb. Mol. Case Stud. 1, a000570 (2015).
20. Parnell, L.D. et al. BioStar: an online question & answer resource for the bioinformatics community. PLOS Comput. Biol. 7, e1002216 (2011).
21. Amendola, L.M. et al. Performance of ACMG-AMP Variant-Interpretation Guidelines among nine laboratories in the Clinical Sequencing Exploratory Research Consortium. Am. J. Hum. Genet. 98, 1067–1076 (2016).
22. Richards, S. et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet. Med. 17, 405–424 (2015).
23. Ritter, D.I. et al. Somatic cancer variant curation and harmonization through consensus minimum variant level data. Genome Med. 8, 117 (2016).