CDDpress

www.nature.com/cdd

COMMENT

The *TP53* Database: transition from the International Agency for Research on Cancer to the US National Cancer Institute

Kelvin César de Andrade[1], Elaine E. Lee[2], Elise M. Tookmanian[3], Chimene A. Kesserwan[4], James J. Manfredi[5], Jessica N. Hatton[1], Jennifer K. Loukissas[3], Jiri Zavadil[6], Lei Zhou[7], Magali Olivier[6], Megan N. Frone[1], Owais Shahzada[8], William J. R. Longabaugh[2], Christian P. Kratz[9], David Malkin[10], Pierre Hainaut[11] and Sharon A. Savage[1]✉

Cell Death & Differentiation (2022) 29:1071–1073; https://doi.org/10.1038/s41418-022-00976-3

From 1994 through 2021, The International Agency for Research on Cancer (IARC) and the World Health Organization maintained a comprehensive database on variations in the tumor protein p53 gene (*TP53*), one of the most frequently mutated genes in human cancer. *TP53* plays crucial roles in cell signaling, apoptosis, metabolism, DNA repair and transcription, earning it the moniker "guardian of the genome" (reviewed in [1]). Germline genetic variants in *TP53* are the primary cause of Li-Fraumeni syndrome (LFS, OMIM 151623), a hereditary cancer predisposition disorder [2] associated with an approximately 24 times higher lifetime incidence of any cancer compared with the general population [3]. The database was initiated by Hollstein et al. [4] and, posteriorly, developed and curated at IARC by Pierre Hainaut and Magali Olivier from 1995 until 2021. During this period, the dataset grew from 2500 to over 50,000 annotated variations in the current database release [5], making it the largest single-locus cancer database. The database has served as an important resource for numerous *TP53*- and LFS-associated studies. Since 1997, the key publications describing and referencing the database have accumulated over 9000 citations in the scientific and medical literature (source: Google Scholar; selected significant papers include [5–8]). Data from the IARC *TP53* database have been widely mined and analyzed to systematically explore functional and structural properties of p53 variants [9–11], genotype–phenotype associations [12], temporal patterns of cancer penetrance [13], carcinogen-induced mutation signatures [14–16], and cancer prognosis and outcomes [17]. The most recent publication using this database, as of the writing of this commentary, focused on the germline dataset and investigated differences in variant distribution and cancer patterns to better refine the variable LFS-associated phenotypic spectrum [18].

On October 25th, 2021, the IARC-sponsored *TP53* database was fully transferred to the US National Cancer Institute (NCI) to host and facilitate important upgrades to its infrastructure (https://tp53.isb-cgc.org). The original *TP53* Database was run on an on-premises server at IARC, using a Microsoft platform. The NCI-sponsored *TP53* Database is hosted on the Google Cloud Platform, primarily using its App Engine, BigQuery, and Cloud Storage services. The high-level architecture, along with additional specifications, are illustrated in Fig. 1. The web application was rewritten using the Python-based Flask framework, and now runs on App Engine, which provides automatic load balancing to ensure scalability and high availability. Data files available for download and files used to support the web application are kept in Cloud Storage. There are 49 tables and 21 views stored in BigQuery dataset, *isb-cgc-tp53.P53_data*, that serves as the read-only database for the application. This dataset is publicly readable and can thus be used by researchers directly for cloud-based analyses using the Google BigQuery API. For the initial rollout, the existing database and files were copied directly from the IARC system. In order to obtain input from the *TP53*-associated scientific and clinical community, the database content was divided into three main subgroups according to the data types available: germline, tumor (somatic), and mouse and other experimental models. We invited 380 individuals previously registered for two conferences (the 17th and 18th International p53 Workshops) to participate in the working groups convened to oversee the transition. Seventy individuals, from 11 countries, expressed interest in being included in at least one of the working groups. The "Germline Variants" working group is made up of 48 members, the "Mouse and Other Experimental models" working group has 25 members, and the "Somatic Variants" working group has 23 members (Supplementary Table 1).

Based on insights from the working groups and digital media/user-experience experts, several important updates were made. The web interface was redesigned to implement user-centered design principles and modern aesthetics, optimized for search, and programmed to function responsively across device platforms. The language throughout the site, including the user manual and database descriptions, was updated to improve clarity and usability. One of the major enhancements was allowing users to easily preview the downloadable data, and filter certain rows by the column values. This was designed to locate and download data of interest more efficiently. Whenever possible, the database will follow the "Findable, Accessible, Interoperable, Reusable" principles (FAIR) to guide data

[1]Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA. [2]Institute for Systems Biology, Seattle, WA, USA. [3]Office of the Director, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, MD, USA. [4]Genetics Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA. [5]Department of Oncological Sciences and Tisch Cancer Center, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [6]Epigenomics and Mechanisms Branch, International Agency for Research on Cancer, WHO, Lyon, France. [7]Department of Molecular Genetics & Microbiology, College of Medicine, University of Florida, Gainesville, FL, USA. [8]General Dynamics Information Technology, Rockville, MD, USA. [9]Pediatric Hematology and Oncology, Hannover Medical School, Hannover, Germany. [10]Division of Hematology/Oncology, The Hospital for Sick Children, Department of Pediatrics, University of Toronto, Toronto, ON, Canada. [11]Institute for Advanced Biosciences, Institut National de la Santé et de la Recherche Médicale 1209 Centre National de la Recherche Scientifique, 5309, Universitè Grenoble Alpes, Grenoble, France. ✉email: savagesh@mail.nih.gov

Received: 15 December 2021 Revised: 28 February 2022 Accepted: 4 March 2022
Published online: 29 March 2022

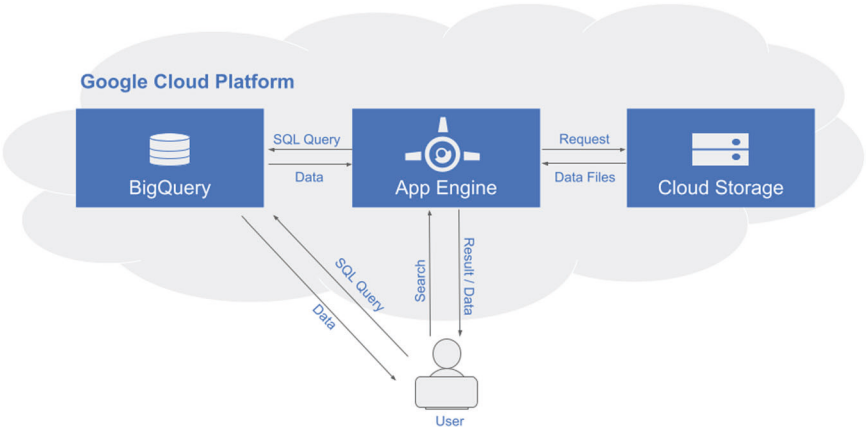

Fig. 1 A high-level system architecture of the redesigned NCI-sponsored *TP53* database (https://tp53.isb-cgc.org). Application Environment: Google Cloud platform-Google App Engine Flexible Environment. Application Framework: Flask. Programming Language: Python 3.7. Object Data Storage: Google Cloud Storage. Database Platform: Google BigQuery. For questions and requests: tp53-info@isb-cgc.org. The application source code can be found in GitHub: https://github.com/isb-cgc/TP53.

management [19], and efforts will be made to remove and replace tools available behind a paywall with publicly available resources. Links to external databases, additional tools, and publications will be added on an ongoing basis to ensure the database remains an effective and up to date resource to facilitate variant curation, achieved with close collaboration with ClinGen's *TP53* Variant Curation Expert Panel (https://clinicalgenome.org/affiliation/50013/). We also intend to add resources to advance studies on emerging topics, such as the variable LFS phenotypic spectrum and potential role of genetic modifiers, abnormal *TP53* variant allele frequency, and *TP53*-related clonal hematopoiesis. Variant annotation will focus on pertinent new *in silico* prediction tools and functional assays, mutational signatures and hotspots, investigating variant-specific DNA-binding affinity, characterizing variants-associated neoantigens, haplotype associations, and other mechanisms of p53 impairment. We also seek to include data on additional model organisms (such as *Drosophila melanogaster* [fruit fly], *Dario rerio* [zebrafish], *Caenorhabditis elegans*, among others) to enable comparative genomics studies, promote collaborative research, and maximize the use of reagents and strains of animal models.

The mission of the NCI-Sponsored *TP53* Database is to serve as a publicly available resource by providing data to better understand existing and new aspects related to the *TP53* gene, its pathways, and the phenotypic manifestations caused by changes in its structure. The NCI *TP53* Database team will coordinate data inclusion and curation requests. Prioritization of database updates will be made in consultation with the working groups, and with research consortia and other collaborative efforts such as the LiFT UP study (ClinicalTrials.gov Identifier: NCT04541654) and the ClinGen's *TP53* Variant Curation Expert Panel. Future endeavors will be geared towards curation of new literature, fostering research, collecting and linking resources to new types of data, and integrating collaborative efforts among clinicians, scientists, and commercial laboratories to expand the characterization of both *TP53* and LFS.

DISCLAIMER

Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy, or views of the International Agency for Research on Cancer/World Health Organization. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

1. Levine AJ. Spontaneous and inherited TP53 genetic alterations. Oncogene. 2021;40:5975–83.
2. Guha T, Malkin D. Inherited TP53 mutations and the Li-Fraumeni syndrome. Cold Spring Harb Perspect Med. 2017;7:a026187.
3. de Andrade KC, Khincha PP, Hatton JN, Frone MN, Wegman-Ostrosky T, Mai PL, et al. Cancer incidence, patterns, and genotype-phenotype associations in individuals with pathogenic or likely pathogenic germline TP53 variants: an observational cohort study. Lancet Oncol. 2021;22:1787–98.
4. Hollstein M, Rice K, Greenblatt MS, Soussi T, Fuchs R, Sorlie T, et al. Database of p53 gene somatic mutations in human tumors and cell lines. Nucleic Acids Res. 1994;22:3551–5.
5. Bouaoun L, Sonkin D, Ardin M, Hollstein M, Byrnes G, Zavadil J, et al. TP53 variations in human cancers: new lessons from the IARC TP53 database and genomics data. Hum Mutat. 2016;37:865–76.
6. Hainaut P, Hollstein M. p53 and human cancer: the first ten thousand mutations. Adv Cancer Res. 2000;77:81–137.
7. Olivier M, Eeles R, Hollstein M, Khan MA, Harris CC, Hainaut P. The IARC *TP53* database: new online mutation analysis and recommendations to users. Hum Mutat. 2002;19:607–14.
8. Petitjean A, Mathe E, Kato S, Ishioka C, Tavtigian SV, Hainaut P, et al. Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database. Hum Mutat. 2007;28:622–9.
9. Kato S, Han SY, Liu W, Otsuka K, Shibata H, Kanamaru R, et al. Understanding the function-structure and function-mutation relationships of p53 tumor suppressor protein by high-resolution missense mutation analysis. Proc Natl Acad Sci USA. 2003;100:8424–9.
10. Giacomelli AO, Yang X, Lintner RE, McFarland JM, Duby M, Kim J, et al. Mutational processes shape the landscape of TP53 mutations in human cancer. Nat Genet. 2018;50:1381–7.
11. Kotler E, Shani O, Goldfeld G, Lotan-Pompan M, Tarcic O, Gershoni A, et al. A systematic p53 mutation library links differential functional impact to cancer mutation pattern and evolutionary conservation. Mol Cell. 2018;71:873.
12. Olivier M, Goldgar DE, Sodha N, Ohgaki H, Kleihues P, Hainaut P, et al. Li-Fraumeni and related syndromes: correlation between tumor type, family structure, and *TP53* genotype. Cancer Res. 2003;63:6643–50.
13. Amadou A, Waddington Achatz MI, Hainaut P. Revisiting tumor patterns and penetrance in germline TP53 mutation carriers: temporal phases of Li-Fraumeni syndrome. Curr Opin Oncol. 2018;30:23–9.
14. Hollstein M, Hergenhahn M, Yang Q, Bartsch H, Wang ZQ, Hainaut P. New approaches to understanding p53 gene tumor mutation spectra. Mutat Res. 1999;431:199–209.
15. Pfeifer GP, Denissenko MF, Olivier M, Tretyakova N, Hecht SS, Hainaut P. Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. Oncogene. 2002;21:7435–51.
16. Olivier M, Hollstein M, Hainaut P. TP53 mutations in human cancers: origins, consequences, and clinical use. Cold Spring Harb Perspect Biol. 2010;2:a001008.
17. Petitjean A, Achatz MI, Borresen-Dale AL, Hainaut P, Olivier M. TP53 mutations in human cancers: functional selection and impact on cancer prognosis and outcomes. Oncogene. 2007;26:2157–65.

18. Kratz CP, Freycon C, Maxwell KN, Nichols KE, Schiffman JD, Evans DG, et al. Analysis of the Li-Fraumeni spectrum based on an international germline TP53 variant data set: an International Agency for Research on Cancer TP53 database analysis. JAMA Oncol. 2021;7:1800–5.
19. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR guiding principles for scientific data management and stewardship. Sci Data. 2016;3:160018.

ACKNOWLEDGEMENTS

We would like to thank the valuable input from members of the three working groups.

AUTHOR CONTRIBUTIONS

KCA and SAS contributed to project conceptualization. KCA contributed to data curation, formal analysis, and project administration. EEL, OS, and WJRL contributed to methodology, software, and visualization. KCA, EEL, EMT, CAK, JJM, JNH, JKL, JZ, LZ, MO, MNF, OS, WJRL, CPK, DM, PH, and SAS contributed to writing and review of the paper. SAS was responsible for funding acquisition and supervision.

FUNDING

Intramural Research Program, NCI, National Institutes of Health. ISB-CGC is a component of the NCI Cancer Research Data Commons and has been funded in whole or in part with Federal funds from the NCI, National Institutes of Health, Department of Health and Human Services, under Contract No. HHSN261201400008C and ID/IQ Agreement No. 17 × 146 under Contract No. HHSN261201500003I.

COMPETING INTERESTS

The authors declare no competing interests. KCA, CAK, JNH, MNF, CPK, and SAS are unpaid members of the ClinGen *TP53* Variant Curation Expert Panel. MNF is co-developer of CancerGene Connect and member of the National Accreditation Program for Breast Centers Board of Directors representing the National Society of Genetic Counselors.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41418-022-00976-3.

Correspondence and requests for materials should be addressed to Sharon A. Savage.