

SOFTWARE

Open Access



# CACSV: a computational web-server that provides classification for cancer somatic genetic variants from different tissues

Nahla AlKurabi<sup>1†</sup>, Ahad AlGahtani<sup>1†</sup> and Turki M. Sobahy<sup>2\*</sup> 

<sup>†</sup>Nahla AlKurabi and Ahad AlGahtani are joint first authors

\*Correspondence: tsobahy@kfshrc.edu.sa

<sup>1</sup> Information Systems Department, King Abdulaziz University (KAU), 7393 Jeddah, Kingdom of Saudi Arabia  
<sup>2</sup> Research Center of King Faisal Specialist Hospital and Research Center-Jeddah (KFSHRC-J), 21499 Jeddah, Kingdom of Saudi Arabia

## Abstract

**Background:** Understanding the role and function of genetic variants is extremely important when analyzing and interpreting a myriad of human disease processes. For cancer in general, cell-specific genetic variants are ubiquitous and distinct tissues have significantly heterogenic genetic profiles. In clinical practice, only a few genetic variants have identifiable clinical utility. Finding clinically relevant genetic variants constitute a challenging process. In addition, there had been no reference protocol to provide guidance for cancer somatic genetic variants classification and interpretation. In 2017, the first version of a reference protocol was published by the Association for Molecular Pathology, the American Society of Clinical Oncology, and the College of American Pathologists. Previously, we incorporated the reference protocol into a computational method to expedite the process of identification of clinically relevant genetic variants. In this work, we developed a computational web-server to increase the accessibility and availability of clinically relevant genetic variants.

**Results:** Our work provides the clinical classification for ~ 3 million cancer genetic variants that are now publicly available in a shareable database on GitHub. We have developed a graphical user interface for the database to enhance the accessibility and ease-of-use.

**Conclusion:** CACSV provides an open-source for about 3 million cancer tissue-specific genetic variants with their assigned clinical annotations.

**Keywords:** Cancer, Somatic, Genetic variants, Classification

## Background

Hundreds of thousands of genetic variants have been associated with single gene disorders, multi-factorial diseases, and cancers. Knowing the clinical annotations and the classification of common and rare genetic variants is important to carry-through precision medicine [1]. Cancer is a heterogeneous disease that manifests distinct genetic and molecular characteristics in different tissues. Cell-specific genetic variants (somatic) could elucidate the molecular functions for cancer driver genes [2]. However, the clinical classification of somatic genetic variants is further affected by



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

the heterogeneous process of creating and using many different classification systems from well-established laboratories [3]. In an effort to manage the complexity of the analysis, the Association for Molecular Pathology (AMP), the American Society of Clinical Oncology (ASCO), and the College of American Pathologists (CAP) developed the first comprehensive algorithm to classify cancer somatic genetic variants (in 2017). The somatic genetic variants are assigned to one of four tiers; variants with strong clinical significance, potential clinical significance, uncertain significance, or benign/likely benign. The classification mostly relies on the availability of clinical and functional evidence in medical professional guidelines and literature, and on the detection of the mutations in small or large cancer screening studies [3].

We incorporated the clinical recommendations into a new computational method as described previously [4]. The National Comprehensive Cancer Network Clinical Practice Guidelines in Oncology (NCCN Guidelines<sup>®</sup>) were used as the source of the practical clinical evidence. Data about actionable genetic variants was extracted from the guidelines and reviewed for somatic genetics variants. In some part of the NCCN guidelines, the genetic variants description was not specific. For example, MET exon 14 skipping mutations in non-small-cell lung carcinoma (NSCLC) are therapeutics biomarkers for Crizotinib. Such “generally” described mutations were curated by acquiring information about their experimental validation available in literature, expert-reviewed databases, or oncogenicity predictions scores (like intOgen & CScape). The Precision Oncology Knowledge Base (OncoKB) was used to collect information about the drug-ability of the genetic variants. The evidence from the literature was measured by available information in COSMIC and cBioPortal databases. The level of a gene-in-tissue involvement was calculated through the data availability in the Cancer Gene Census (CGC) and the Candidate Cancer Gene Database (CCGD).

Our method was evaluated on a subset of manually reviewed variants and showed a balanced performance on a significantly imbalanced subset [4]. We collected cell-specific mutations from oncology genetic hubs, then filtered germline genetic mutations using the gnomAD database returning around ~3 million cancer mutations. The genetic mutations were then analyzed by the new algorithm producing AMP-ASCO-CAP categorized mutations. The classified data-sets were made available in JSON file format.

To facilitate the use of the classified database for non-bioinformatics users, we created CACSV as an online tool that shows the clinical annotations and classification for cancer mutations along with the source of evidence and any associated approved or investigational drugs.

## Implementation

### Design and execution

The graphical interface and computational engine were developed by using PHP, HTML, and JavaScript. The implementation was carried out by the cross-platform web server application *XAMPP*. The online tool was designed to accept queries by either gene name, the human transcript identification number, or the amino acid mutation description along with selecting a single type of cancer tissue or tissue of origin (11 types altogether).

**Data-sets structure**

In our original work, we covered ten types of tumors including: breast, central nervous system (cns), colon, esophageal, gastric, melanoma, non-small cell lung cancer (nslc), pancreas, rectal, and small cell lung cancer (sclc) [4]. About three million cancer mutations that are available on COSMIC, intOgen, or the bulk dataset of cBioPortal were collected and classified as previously described. Each tissue type has a table size of 21 × 2,952,170, and each table has a MySQL table. (Table 1).

**Classification algorithm**

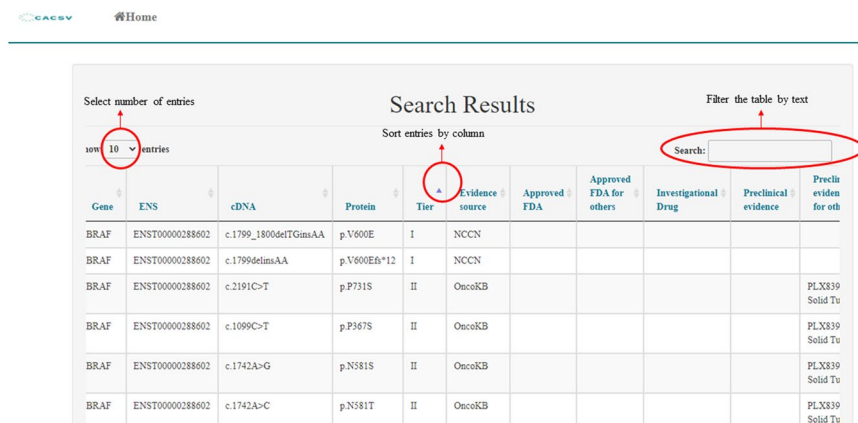
Cell-specific genetic variants are classified based on the availability and on the level of knowledge in medical professional guidelines, literature, mutation prevalence in the genetics databases, and gene-tissue association. Tier I mutations may be delineated in the medical guidelines, they may have fully approved drug(s), or they may be found in PanCan studies in the specified type of tissue. Tier II mutations could have an approved drug(s) for different types of cancer tissues or they could have investigational drugs or preclinical evidence. These mutations may have also been mentioned in small studies. Tier III genetic variants are outlined in a few studies in a different type of tumor tissue and have no reliable experimental functional evaluation.

**Results**

The web-server can take different types of queries from users (gene name, human transcript identification number, or amino acid mutation) with one type of tissue and will return the clinical classification in a formatted table. The table presents different types

**Table 1** Database columns information

Gene	Gene name
ENS	Ensembl transcript identification number
cDNA	Coding DNA variant nomenclature
AA	Protein variant nomenclature
Tier	AMP-ASCO-CAP assigned class
Level	Level of evidence (A, B, C, or D)
Source	Source of the evidence
Bio type	Biomarker type (diagnostic, prognostic, or therapeutic)
Submitted type	Tumor site
NCCN drug	Described drug(s) in the NCCN guidelines
NCCN response	Drug response (resistant or sensitive) in the NCCN guidelines
NCCN condition	Drug related conditions in the NCCN guidelines
NCCN panel	NCCN panel recommendations
FDA for same	FDA-approved drug(s) for the submitted tissue
FDA for other	FDA-approved drug(s) for other tissues
Investigational for same	“Experimental” drugs (in humans) for the same tumor tissue
CT title	Available clinical trials
Status	Clinical trials status (recruiting, suspended, etc.)
Location	Clinical trial location
Preclinical for same	Compelling biological evidence in the same tumor site
Preclinical generic	FDA-approved drug for more than one tumor site
AC	Allele Count in COSMIC



**Fig. 1** The search page is marked to show the multiple options to filter and sort the output table

of information that include the clinical classification (tier), available drugs or clinical trials, and source of evidence. The user may search the full content by string-based search through a filtering box and sort the table by any column (Fig. 1). The users can also choose the number of displayed entries per page: 10, 25, 50, or 100. Data per gene can be exported as well.

**Discussion**

The number of genetics and genomics applications in oncology are immerse and the associated clinical yield is improving. Multiple research and clinical groups work on the genetic information curation and revision process. In the realm of clinical practice, the decisions of the cancer genetics variants “serviceability” to the patients are made through meetings and discussions by the tumor molecular board (TMB). Computational methods providing clinical annotations for thousands of cancer somatic genetic variants can assist in the aforementioned process. Our web-application would provide an open-source to clinical geneticists and medical oncologists during the board meetings.

Future improvements for the software are warranted. The new released clinical standards by the Clinical Genome Resource (ClinGen), Cancer Genomics Consortium (CGC), and Variant Interpretation for Cancer Consortium (VICC) may further improve the clinical annotations [5]. From technical prospective, the software could be also made easy-to-integrate with the existing sequencing analysis workflows and systems.

**Conclusion**

CACSV is an easy-to-use computational method that produces clinical annotations for about 3 million cancer tissue-specific genetic variants.

**Abbreviations**

- GUI Graphical user interface
- AMP The Association for Molecular Pathology
- ASCO American Society of Clinical Oncology
- CAP College of American Pathologists
- CACSV Clinically actionable cancer somatic variants
- COSMIC Catalogue of somatic mutations in cancer
- CNS Central nervous system
- ACMG American College of Medical Genetics

**Acknowledgements**

The authors would like to thank Emily L. Heaphy for manuscript English review.

**Author contributions**

T.M.S designed the web-application interface and deployed the application on AWS Cloud. NA and AA developed all the different aspects of the web-application. All authors read and approved the final manuscript.

**Funding**

T.M.S is employed by King Faisal Specialist Hospital & Research Center-Jeddah (KFSHRC-J). The funding body played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

**Availability of data and materials**

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study. Project name: Clinically Actionable Cancer Somatic Variants (CACSV). Project home page: <http://44.203.161.191/cacsv/>. Operating system(s): Platform independent. Programming language: PHP, HTML, MySQL, and JavaScript. Other requirements: update to date web-browser. License: MIT. Any restrictions to use by non-academics: license needed.

**Declarations****Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare no conflict of interest.

Received: 27 October 2022 Accepted: 23 February 2023

Published online: 15 March 2023

**References**

1. Sahni N, et al. Widespread macromolecular interaction perturbations in human genetic disorders. *Cell*. 2015;161:647–60.
2. Tamborero D, et al. Cancer genome Interpreter annotates the biological and clinical relevance of tumor alterations. *Genome Med*. 2018;10:25.
3. Li MM, et al. Standards and guidelines for the interpretation and reporting of sequence variants in cancer: a joint consensus recommendation of the association for molecular pathology, American Society of clinical oncology, and college of American pathologists. *JMD*. 2017;19:4–23.
4. Sobahy TM, et al. Clinically actionable cancer somatic variants (CACSV): a tumor interpreted dataset for analytical workflows. *BMC*. 2022;15:95.
5. Horak P, et al. Standards for the classification of pathogenicity of somatic variants in cancer (oncogenicity): joint recommendations of clinical genome Resource (ClinGen), cancer genomics consortium (CGC), and variant interpretation for cancer consortium (VICC). *Genet Med: Off J Am Coll Med Genet*. 2022;24:5.

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

