

RESEARCH

Open Access



Somatic cancer variant curation and harmonization through consensus minimum variant level data

Deborah I. Ritter^{1†}, Sameek Roychowdhury^{2†}, Angshumoy Roy¹, Shruti Rao³, Melissa J. Landrum⁴, Dmitriy Sonkin⁵, Mamatha Shekar⁶, Caleb F. Davis⁷, Reece K. Hart⁸, Christine Micheel⁹, Meredith Weaver¹⁰, Eliezer M. Van Allen¹¹, Donald W. Parsons¹, Howard L. McLeod¹², Michael S. Watson¹⁰, Sharon E. Plon¹, Shashikant Kulkarni¹, Subha Madhavan^{3*} and on behalf of the ClinGen Somatic Cancer Working Group

Abstract

Background: To truly achieve personalized medicine in oncology, it is critical to catalog and curate cancer sequence variants for their clinical relevance. The Somatic Working Group (WG) of the Clinical Genome Resource (ClinGen), in cooperation with ClinVar and multiple cancer variant curation stakeholders, has developed a consensus set of minimal variant level data (MVLN). MVLN is a framework of standardized data elements to curate cancer variants for clinical utility. With implementation of MVLN standards, and in a working partnership with ClinVar, we aim to streamline the somatic variant curation efforts in the community and reduce redundancy and time burden for the interpretation of cancer variants in clinical practice.

Methods: We developed MVLN through a consensus approach by i) reviewing clinical actionability interpretations from institutions participating in the WG, ii) conducting extensive literature search of clinical somatic interpretation schemas, and iii) survey of cancer variant web portals. A forthcoming guideline on cancer variant interpretation, from the Association of Molecular Pathology (AMP), can be incorporated into MVLN.

Results: Along with harmonizing standardized terminology for allele interpretive and descriptive fields that are collected by many databases, the MVLN includes unique fields for cancer variants such as Biomarker Class, Therapeutic Context and Effect. In addition, MVLN includes recommendations for controlled semantics and ontologies. The Somatic WG is collaborating with ClinVar to evaluate MVLN use for somatic variant submissions. ClinVar is an open and centralized repository where sequencing laboratories can report summary-level variant data with clinical significance, and ClinVar accepts cancer variant data.

Conclusions: We expect the use of the MVLN to streamline clinical interpretation of cancer variants, enhance interoperability among multiple redundant curation efforts, and increase submission of somatic variants to ClinVar, all of which will enhance translation to clinical oncology practice.

Keywords: Cancer genomics, Somatic variant interpretation, Data standard, Somatic variant curation

* Correspondence: sm696@georgetown.edu

†Equal contributors

³Innovation Center for Biomedical Informatics and Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, Washington, DC, USA
Full list of author information is available at the end of the article

Background

To achieve personalized medicine for oncology, it is critical to catalog and curate cancer genomic sequence alterations in order to improve biomarker applications, drug development, and treatment selection and ultimately reduce morbidity and mortality from cancer. Curation and interpretation efforts are underway in multiple centers, often through local internal repositories, web-based platforms curated by laboratory efforts [1], crowd-sourced cancer curation [2], or multi-institutional efforts such as Genomics, Evidence, Neoplasia, Information, Exchange (GENIE) [3]. However, there is no broadly adopted standardized framework to capture clinically relevant data on somatic variants. In order to streamline curation efforts of somatic alterations in the community and increase the clinical utility of cancer variant curation, the Somatic Working Group (WG) of ClinGen, with representation from groups including Clinical Sequencing Exploratory Research (CSER), Association of Medical Pathologists (AMP), American Society of Clinical Oncology (ASCO), and Global Alliance For Genomics and Health (GA4GH), in cooperation with ClinVar and multiple cancer variant curation stakeholders [1, 2, 4–6]

has developed a consensus set of minimal variant level data (MVLDD) that we propose for broad adoption as a standard framework to create a common language for curation and clinical interpretation of somatic alterations.

The Clinical Genome Resource (ClinGen) is a National Institutes of Health (NIH) initiative representing over 75 institutions [6] and is a natural platform for supporting centralized curation of somatic variants and their interpretation. The ClinGen network focuses on clinical and public use of genetic information, with a special emphasis on curation of and education on gene-disease associations and variant interpretations. ClinGen has developed cancer-relevant Clinical Domain Working Groups, including Hereditary Cancer, Pharmacogenomics, and Somatic Cancer. Major challenges for current somatic variant curation include similar but inconsistent terminology, costly and redundant manual curation efforts, and lack of a central framework for housing and coordinating expertise. An example of the necessity of a common framework to describe somatic variants is shown in Fig. 1, where separate databases convey partial information about a somatic variant. Relevant elements are missing from all three; such

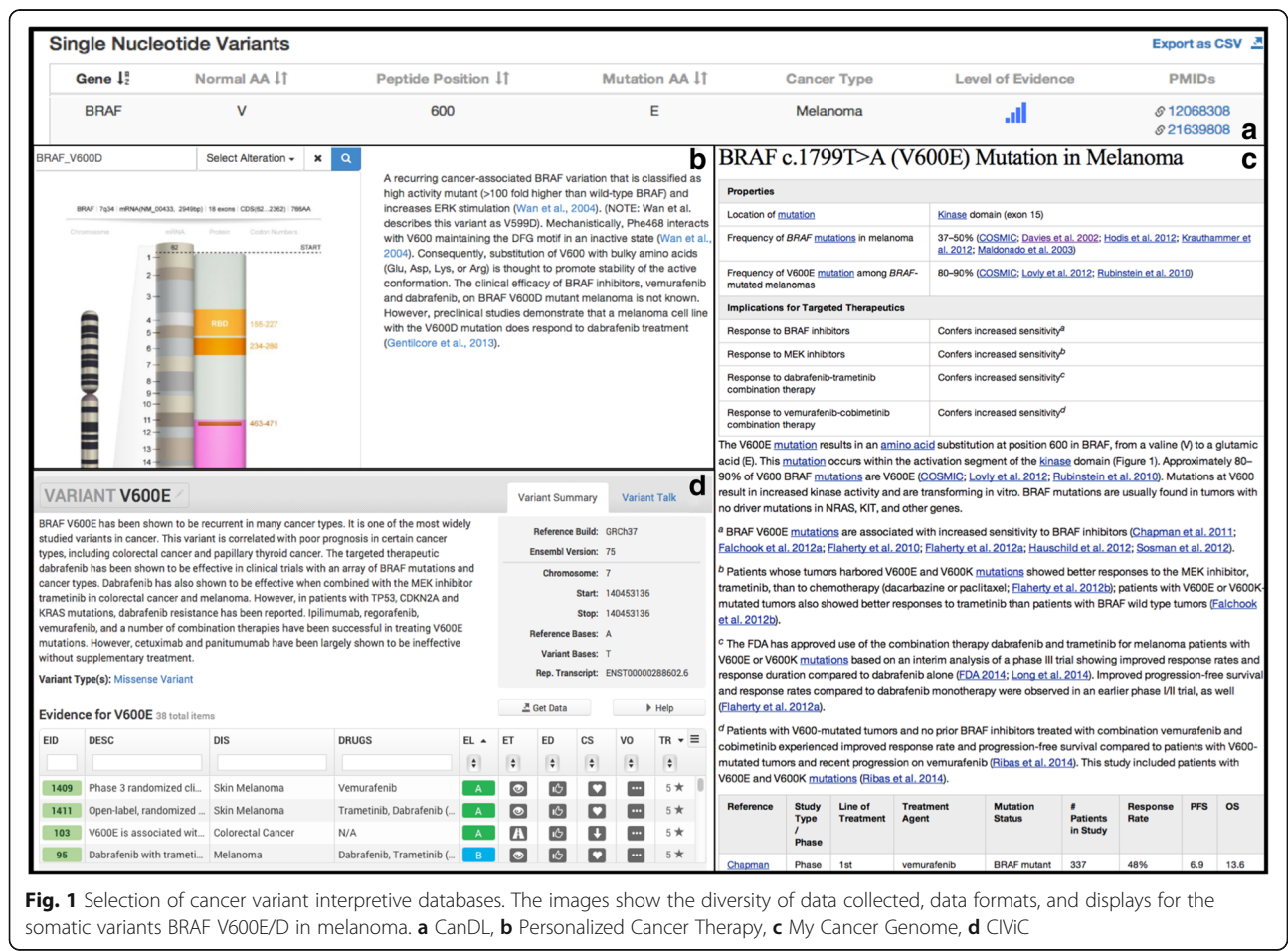


Fig. 1 Selection of cancer variant interpretive databases. The images show the diversity of data collected, data formats, and displays for the somatic variants BRAF V600E/D in melanoma. **a** CanDL, **b** Personalized Cancer Therapy, **c** My Cancer Genome, **d** CIVIC

discrepancies create information gaps and perceived inconsistencies and are a significant communication challenge for clinical or research use of variants.

Methods

Literature, institutional, and metadata reviews inform MVLD development

To develop MVLD, our team first reviewed clinical actionability interpretations from ten institutions participating in ClinGen Somatic WG, seeking information on guidelines or standards utilized to create their actionability frameworks and associated information. We also incorporated input from groups presenting frameworks on ClinGen Somatic WG conference calls [2, 7]. As guidance, we used prior effective efforts to define minimal and structured data, such as minimum information about a microarray experiment (MIAME) adopted by the Gene Expression Omnibus (GEO) database [8, 9]. In addition, we reviewed publications on clinical somatic variant interpretation [10–14]. We uncovered a wide variety of data elements but one distinct common theme emerged: many institutions (half from the survey and multiple others from literature review and group presentations) consistently promoted a somatic variant to the highest actionability if it had a corresponding Food and Drug Association (FDA)-approved drug therapy. The same institutions, as well as many others reviewed in the literature, had varying levels of lower priority tiers (between three to five tier schemes), with a general schema of incorporating lower priority levels of evidence such as guidelines from a professional society, on-going clinical trials, preclinical trials, case series, case reports, in vitro studies, or pathway-based evidence. Each group collected some part of the aforementioned information, and in a variety of formats and data collection systems. Using this as a rough guideline, we next reviewed fields that are currently in use from multiple somatic variant databases.

Current systems capturing somatic variant data inform MVLD development

Databases with variant-level information relevant to cancer can be roughly divided into two types: variant catalogs and variant interpretive databases. Similar to the responding institutions and groups, each database captured somewhat analogous information, but often in distinct formats (Fig. 1). Variant catalogs include sites such as: Catalog of Somatic Mutations in Cancer (COSMIC) and International Cancer Genome Consortium (ICGC) data portal [15, 16]. Variant interpretive datasets include sites such as ClinVar, My Cancer Genome, Clinical Interpretations of Variations in Cancer (CIViC), the Cancer Driver Log (CanDL); multi-omics integration and analysis platforms include Georgetown Database

of Cancer (G-DOC), cBIOPortal, and Personalized Cancer Therapy [1, 4, 5, 17–21]. Here, we briefly describe a selection of somatic variant interpretive sites reviewed in developing MVLD.

ClinVar

Of the cancer variant interpretive sites, ClinVar is the only database with a broader scope that extends beyond cancer to germline variants. ClinVar is an open, centralized variant repository of clinically relevant, interpreted variants and offers a summation of collected data as well as expert panel reviews. Currently, somatic alterations make up a minority usage of ClinVar, roughly 2 % of the total dataset. However, the infrastructure of ClinVar has the potential to be leveraged into somatic variant curation. For germline variant interpretation, there is a well-established approach [22]; a similar, though distinct, language is often applied for interpreting somatic variants (Fig. 2). For example, germline variants may be categorized as pathogenic, while somatic variants are often categorized as diagnostic, prognostic, or predictive biomarkers. Similarly, supporting evidence for a “very strong” pathogenic germline variant may come from published studies reporting high penetrance with segregation data, while supporting evidence for a predictive somatic variant could range from large randomized clinical trials to pre-clinical laboratory data. The Somatic WG worked closely with ClinVar in the development of MVLD, gaining insight into unique considerations in curating somatic variants from the perspective of experienced germline curators.

My Cancer Genome

My Cancer Genome was the first public somatic variant interpretation resource, launched in 2011, and includes information on the effect of tumor variants on sensitivity to targeted therapeutics, as well as a listing of cancer clinical trials that include biomarker information. The information is provided by expert contributors and edited by knowledge resource staff.

CIViC

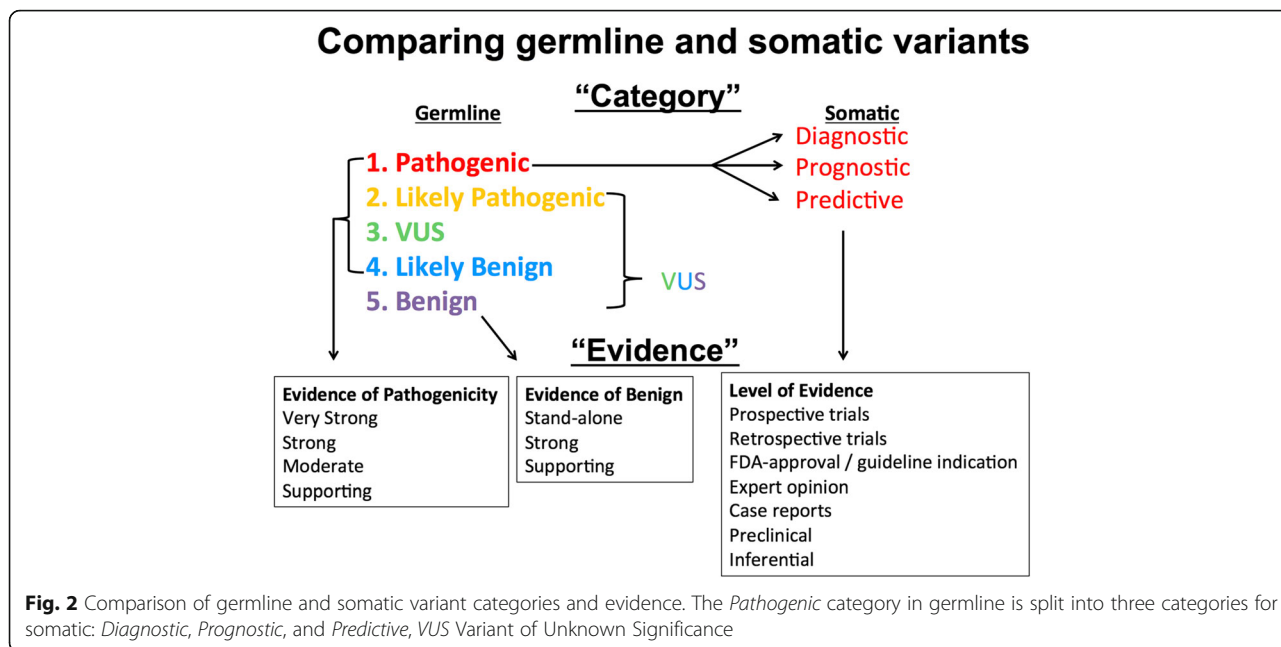
CIViC provides descriptions and evidence levels for publications addressing tumor variants in cancer. The information is crowd-sourced and expert-moderated.

CanDL

CanDL is a curated list of cancer variants that have literature-defined levels of evidence for predicting response to therapy (predictive biomarkers only).

Personalized Cancer Therapy

Personalized Cancer Therapy is a password-protected knowledge resource with free and fee-based levels of



access to the content containing information on the therapeutic implications of 27 genes in cancer, categorized by level of evidence.

Results

MVLD to describe cancer variants

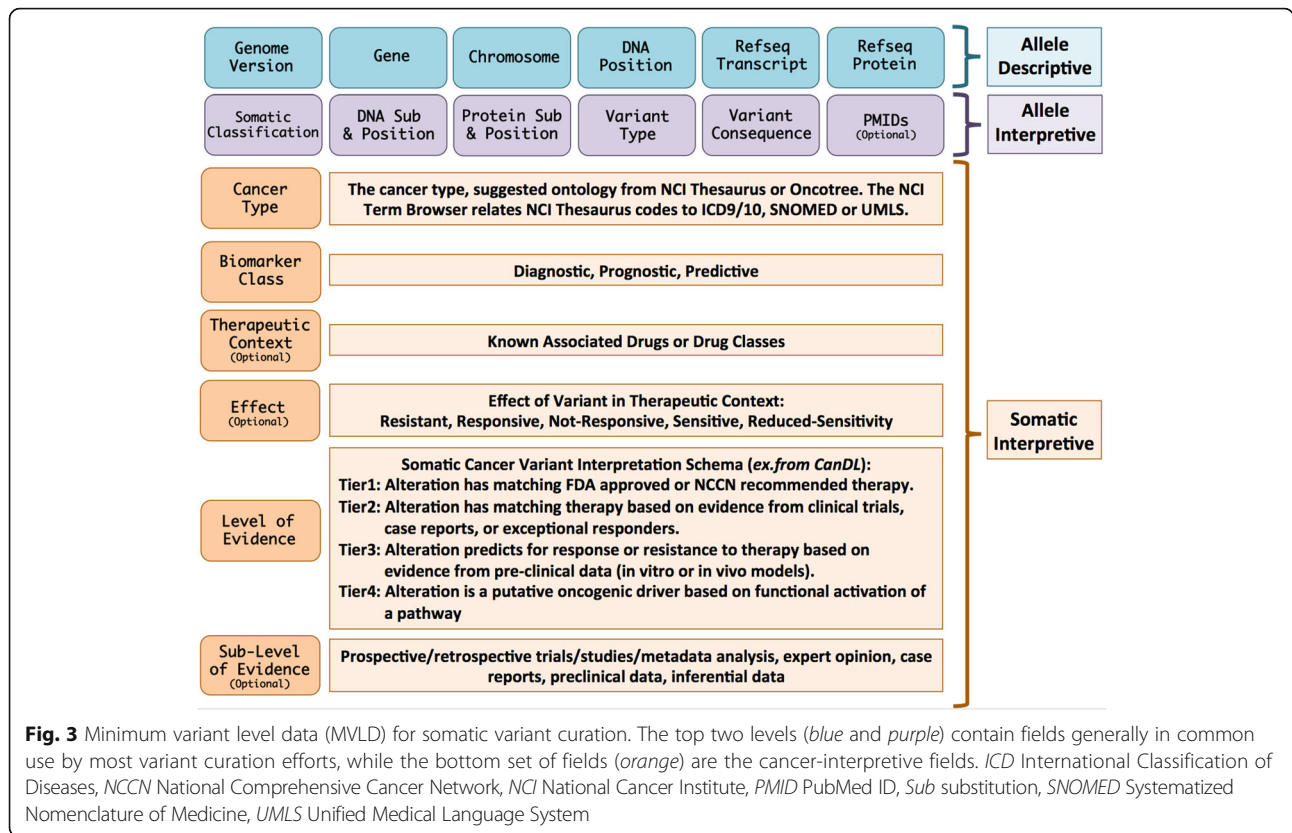
Taking into account institutional surveys, group presentations and discussion as well as prior literature and current websites, we formed the consensus data elements of MVLD, as shown in Fig. 3 and described in further detail below. Throughout the descriptions, we suggest how these elements can be incorporated into the existing structure of ClinVar, although the intent of MVLD is to be a standard data structure not tailored to a specific database.

Allele descriptive fields

The first section of Fig. 3 (blue) incorporates overall standard fields already in common use to describe and characterize a genomic variant. *Genome Build* should be in GRCh37/GRCh38 format and should, if possible, use the actual version of the reference genome used to call the variant. *Gene Name* should be the Human Genome Organization (HUGO)-approved gene name. *Chromosome* and *DNA Position* should be the number or letter representation of the chromosome on which the variant is found and the corresponding genomic coordinate in HGVS format. For *Refseq Transcript* and *Refseq Protein* RefSeq transcripts and protein identifiers should be used. Since the transcript is often not known in sequencing data, all applicable transcripts may be used, or the most commonly accepted transcript may be used.

Allele interpretive fields

The second section of Fig. 3 (purple) pertains to the allele interpretive fields. Other than *Somatic Classification*, this section also contains generally standardized fields used by most curation efforts. The *Somatic Classification* field is necessary for cancer variant curation. Many centers do not yet require paired or matched normal sample sequencing, which is needed to distinguish cancer-specific variants from individual or rare germline variants. Suggested terms for the somatic classification are “Confirmed somatic”, “Confirmed germline”, or “Unknown.” The “Unknown” term could be a placeholder for submitters lacking matched normal sequencing or entities submitting data on behalf of literature or websites where paired sample information is not available. *DNA Substitution and Position* and *Protein Substitution and Position* should be written in HGVS format for both DNA and protein positions (if applicable). If there is a noncoding variant, the DNA position only may be supplied. For all other variant types, it is strongly suggested to include both DNA and protein annotations, but only one is required. *Variant Type* should represent the type of variant, such as single nucleotide variant (SNV), multi-nucleotide variant (MNV), insertion (INS), or deletion (DEL). Complex variants, such as deletion plus substitution, should be described as MNV. *Variant Consequence* should be the “molecular consequence” of a variant and rendered in the suggested terms “Nonsense”, “Missense”, “Silent”, “Frame shift”, “In-frame”, “3UTR”, “5UTR”, “Splice”, “Splice-region”, “Intronic”, “Upstream”, or “Downstream”. For *Variant Type* and *Variant Consequence*, these terms are all available in



MISO Sequence Ontology codes as well [23]. *PubMed IDs* (PMIDs) are strings that reference supporting publications for the variant deposited. It is strongly suggested to use PMIDs to support variant evidence, but PMIDs are optional in this iteration of MVL.

Cancer interpretive fields

The additional cancer-relevant fields are the main consensus data fields developed for the curation and dissemination of clinically relevant cancer variants. For *Cancer Type*, incorporating a standardized terminology for cancers when reporting variants is critically important to sharing datasets. Thus, a versatile ontology should be used to describe the cancer type. In clinical practice, sequencing requisition forms often require International Classification of Diseases (ICD) codes. However, this terminology set was not expressly created for describing cancers [24]. Several cancer-focused ontologies are available, such as National Cancer Institute (NCI) Thesaurus [25], a set of encoded terms to describe cancers and tissue pathology, and Oncotree, a set of 519 tumor types in 32 tissues [26]. Oncotree includes the NCI Thesaurus code for each cancer type. Through the NCI Term Browser, the NCI Thesaurus codes can also be related to other ontologies, such as the aforementioned ICD, or others such as SNOMED and UMLS. We suggest use of either NCI Thesaurus or Oncotree. The NCI Thesaurus

can additionally describe histopathological tissue changes associated with cancers, while Oncotree, though limited to cancers, is very useful as a short, readily interpretable initialism of the cancer type (e.g. RGNT is rosette-forming glioneuronal tumor). The *Biomarker Class* field describes the clinical utility of the variant, and we suggest three standardized terms: “Diagnostic”, “Prognostic”, or “Predictive”. These are terms already in common use and are drawn from concepts proposed by the Institute of Medicine 2012 Translational Omics Report [27]. The *Therapeutic Context* field includes drugs that are specific to the variant reported. This field should first be populated with any FDA or National Comprehensive Cancer Network (NCCN) recommended treatment, followed by relevant drugs from commonly used drug databanks, including The DrugBank [28]. This is an optional field, though clinical relevance of a variant is greatly enhanced by including information of the relevant therapeutic context. The *Effect* field will hold keywords describing the effect of the variant in the therapeutic context. Deinstmann et al. [10] proposed a five-term vocabulary that we reason describes most cases: resistant, responsive, not-responsive, sensitive, reduced sensitivity. We have adopted this and added one field, “other”, to allow a free-text field descriptor only if none of the five descriptors apply. As this field is dependent upon the *Therapeutic Context*, it is also an

optional field in this iteration of MVLD. While the *Level of Evidence* field can hold any variant-scoring framework, we suggest users adopt the forthcoming somatic variant interpretation guidelines issued by AMP. However, adopting a new framework may not be feasible for current projects or prior publications. Thus, any well-described interpretive or scoring framework may substitute here. The ClinVar field “Review Status (Assertion Method)” takes a similar approach. The user can submit a variant interpreted with a described published schema and the variant receives expert review with consideration of the referenced schema. One simple stratified somatic interpretation framework already in use is that of CanDL, the Cancer Driver Log [1]. In this somatic variant interpretation system, the level of evidence can be broken into four tiers, each tier using a short structured sentence to describe the utility of a cancer variant in clinical practice as follows: tier 1, “Alteration has matching FDA approved or NCCN recommended therapy”; tier 2, “Alteration has matching therapy based on evidence from clinical trials, case reports, or exceptional responders”; tier 3, “Alteration predicts for response or resistance to therapy based on evidence from pre-clinical data (in vitro or in vivo models)”; tier 4, “Alteration is a putative oncogenic driver based on functional activation of a pathway”. The *Sub-Level of Evidence* field further substantiates the *Level of Evidence* assertion and is generally composed of six possible subsections: (1) prospective trials/studies, (2) retrospective trials/studies and metadata analysis, (3) expert opinion, (4) case reports, (5) published preclinical data, and (6) inferential data or publications. Standardizing language and terminology in this field is important and will be an ongoing effort as MVLD further develops. In this iteration of MVLD, this field is optional. For prospective and retrospective trials and studies and metadata analysis, we recommend supplying the clinical trial number (NTC) for any clinical trial, which can be found on websites such as the NIH repository for clinical trials (ClinicalTrials.gov), MolecularMatch, or the International Clinical Trials Registry. For case reports or published pre-clinical data, always cite at least the PMID. Case reports are a single, unique observation in an individual (for example, [29]). Pre-clinical data are often functional data that have not been tested in a clinical trial but have strong implications for clinical utility (for example, [30]). For expert opinion, cite the date, name, and academic or medical affiliation of the expert or members of an expert panel. For published inferential data (for example, [31]), cite the PMID, or for in silico predictions, cite the name of the programs used.

The Somatic WG is implementing MVLD with ClinVar on a dataset from Baylor College of Medicine Advancing Sequencing Into Childhood Cancer Care

(BASIC³) through the NIH-funded CSER. This set of somatic variants derives from pediatric solid tumor sequencing [7]. The published dataset was first transformed into a standard ClinVar submission to understand how current fields in ClinVar may be adapted to refine somatic variant handling. Then, the data were transformed to MVLD fields and categories such as “Biomarker Class”, which were not initially included in the dataset reviewed by CSER working group members. Currently we are seeking other datasets and groups who would like to upload somatic variants to ClinVar or test the MVLD format. We provide an example of MVLD in using the BASIC³ dataset in Additional file 1 (“Example of MVLD data format”). The completed, published BASIC³ dataset in MVLD format, with relation of the MVLD fields to those in ClinVar, will be made available as an example file for groups interested in testing MVLD formatting on their data.

Discussion and conclusion

In its current state, the relevant information for the clinical use of a cancer variant is often dispersed, with different formats for similar information or relevant information missing. Inconsistency in cancer variant data creates knowledge gaps, complicates the exchange of cancer variant data, and uses considerable resources for repeated data transformations. In creating MVLD through a consensus approach, and promoting the adoption of a standardized framework across stakeholders in cancer variant curation, the Somatic WG of ClinGen aims to minimize redundant data handling and create a consistent set of elements for the clinical utility of cancer variants.

For a set of data elements such as MVLD to become a standard relies heavily on user/community adoption, uptake, and continued usage. The membership in ClinGen Somatic WG spans multiple institutions and includes at least five major cancer variant curation knowledge bases as well as multiple representatives from industry. In addition, members of the ClinGen Somatic WG are collectively members of many current major efforts to curate the cancer genome, including CSER, GA4GH, GENIE, Oncology Research Information Exchange Network (ORIEN), and The Cancer Genome Atlas (TCGA). MVLD was developed through a consensus approach with input from multiple groups that agree on the necessity of a practical and useful data standard for cancer variant curation. We are also working closely in partnership with ClinVar both to understand somatic variant handling and to test migration from MVLD to ClinVar submissions. MVLD will likely evolve during a testing and adoption phase, and we are actively reaching out to groups interested in submitting somatic data to ClinVar to assist with data handling,

increase somatic submissions to ClinVar, and solicit MVLD testing and feedback. In subsequent iterations of MVLD, we plan to extend the reach to international databases such as COSMIC. Also, creating a parallel somatic rating schema to that of the current germline variant “star” system is currently being discussed, as are the creation of expert panels for somatic variant reviews.

It is important to note that in its current format, MVLD mainly applies to aggregated variant-level data and not necessarily to case-level data. For example, a recent study describes an individual with four somatic variants in a rosette-forming glioneuronal tumor (RGNT) and a pathogenic germline variant [32]. At the variant level, MVLD does not capture information relating multiple variants in one sample. This is important to consider because individual samples can often have multiple somatic variants and possible drug contraindications [14]. Currently, initiatives such as GA4GH, GENIE, and ORIEN are focusing efforts on modeling case-level data. In addition, the MVLD currently requires Refseq identifiers, as well as select ontologies that can be mapped to NCI Thesaurus. Although total transcript content may be comparatively increased in sets such as Gencode, UCSC, or Ensembl, the true impact on limitation for variants in MVLD format is currently unknown. We suggest that using ClinVar to track bulk somatic variant submissions lacking mappable Refseq identifiers could help gauge estimations. Also, additional ontologies may be rapidly added. For example, Disease Ontology also maps to NCI codes and thus further iterations may offer a “local ontology” optional field alongside a more circumspect selection of required standardized terms [33]. While standardization often necessitates limitation, current tools in development to convert between ontologies and annotation sets will greatly enhance the perspective usage scope. In addition, MVLD captures only DNA sequencing data and does not in the current format capture RNA data, structural variation data, “outcomes” level data, or other cancer-relevant test data. However, additional types of data beyond next-generation sequencing variants will build a more personalized approach to cancer care, and handling the complex challenge of standardizing diverse data is made simpler by having first a basic set of common data elements for cancer variant curation.

Standardizing the data elements that represent a somatic variant is critically important to enhancing communication and utility of genetic data for clinicians, researchers, and the public. The FDA has recently proposed creating an FDA recommendation for variant databases, and standardization of data across databases is a core feature of their proposed initiative. A standard set of key data elements specified using controlled

vocabularies to describe somatic variants from clinical tests will enable large-scale analysis of molecular diagnostic and theranostic data from multiple sources and drive forward cancer precision medicine, as well as ensure continued and broad use of clinical and research data.

Additional file

Additional file 1: Example of the MVLD data format. (XLSX 32 kb)

Abbreviations

AMP: Association for Molecular Pathology; BASIC³: Baylor College of Medicine Advancing Sequencing Into Childhood Cancer Care; CanDL: Cancer Driver Log; CIVIC: Clinical Interpretations of Variations in Cancer; ClinGen: Clinical Genome Resource; COSMIC: Catalog of Somatic Mutations in Cancer; CSER: Clinical Sequencing Exploratory Research; FDA: Food and Drug Association; GA4GH: Global Alliance For Genomics and Health; GENIE: Genomics, Evidence, Neoplasia, Information, Exchange; ICD: International Classification of Diseases; MNV: Multi-nucleotide variant; MVLD: Minimum Variant Level Data; NCCN: National Comprehensive Cancer Network; NCI: National Cancer Institute; NIH: National Institutes of Health; ORIEN: Oncology Research Information Exchange Network; PMID: PubMed ID; WG: Working Group

Funding

ClinGen is funded by the National Human Genome Research Institute, with additional funding from the Eunice Kennedy Shriver National Institute of Child Health and Human Development and the National Cancer Institute (U41 HG006834, U01 HG007436, U01 HG007437, HHSN261200800001E). ClinVar is supported by the Intramural Research Program of the NIH, National Library of Medicine. Additional funding: 01HG008390 (NHGRI, BD2K Program) to SM; U01HG006492 and K08CA188615 to EMV. The BASIC³ study is a Clinical Sequencing Exploratory Research (CSER) program project supported by NHGRI/NCI 1U01HG006485 to AR, SEP, and DWP.

Availability of data and material

BASIC³ data will be available in ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>) and is taken from a published dataset of Parsons et al. 2016 (doi:10.1001/jamaoncol.2015.5699).

Authors' contributions

SM proposed the concept of minimum variant level data (MVLD). SM and SK co-chaired two in-person conferences and bi-monthly conference calls to discuss and develop MVLD. SR provided extensive background and input of the somatic interpretive fields and made a major contribution to the manuscript in describing MVLD. MS provided extensive edits and revisions to the manuscript as well as development of the somatic interpretive fields. DIR provided literature and data organization, ClinVar submission, and made a major contribution to the manuscript. ML and DS provided ClinVar somatic variant handling. CM made a major contribution to the manuscript. DWP contributed to development of cancer interpretive fields and guidance on the published BASIC³ dataset. SEP is a Principal Investigator of ClinGen and a co-chair of the Clinical Domain Working Group and provided expertise in cancer variant curation. SK and SM are ClinGen Somatic Working Group Chairs. SEP is on the ClinGen Steering Committee. All authors contributed to the development of MVLD and have read and approved the manuscript.

Competing interests

Baylor College of Medicine (BCM) and Miraca Holdings Inc. have formed a joint venture with shared ownership and governance of the Baylor Genetics Laboratory which performs clinical genetic testing. SEP is an employee of BCM and serves on the Scientific Advisory board of Baylor Genetics Laboratory. The remaining authors declare that they have no competing interests.

Author details

¹Baylor College of Medicine and Texas Children's Hospital, Houston, TX, USA. ²Ohio State University, Columbus, OH, USA. ³Innovation Center for Biomedical Informatics and Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, Washington, DC, USA. ⁴National

Center for Biotechnology Information, Bethesda, MD, USA. ⁵National Cancer Institute, Rockville, MD, USA. ⁶Illumina, San Diego, CA, USA. ⁷MolecularMatch, Houston, TX, USA. ⁸Invitae, San Francisco, CA, USA. ⁹Vanderbilt University School of Medicine, Nashville, TN, USA. ¹⁰American College of Medical Genetics and Genomics, Bethesda, MD, USA. ¹¹Dana-Farber Cancer Institute, Boston, MA, USA. ¹²Moffitt Cancer Center, Tampa, FL, USA.

Received: 16 July 2016 Accepted: 13 October 2016

Published online: 04 November 2016

References

- Damodaran S, Miya J, Kautto E, Zhu E, Samorodnitsky E, Datta J, Reeser JW, Roychowdhury S. Cancer Driver Log (CanDL): catalog of potentially actionable cancer mutations. *J Mol Diagn*. 2015;17(5):554–9.
- Good BM, Ainscough BJ, McMichael JF, Su AI, Griffith OL. Organizing knowledge to enable personalization of medicine in cancer. *Genome Biol*. 2014;15(8):438.
- Rose S. Huge data-sharing project launched. *Cancer Discov*. 2016;6(1):4–5.
- Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Hoover J, Jang W, Katz K, Ovetsky M, Riley G, Sethi A, Tully R, Villamarin-Salomon R, Rubinstein W, Maglott DR. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res*. 2016;44(D1):D862–8.
- Taylor AD, Micheel CM, Anderson IA, Levy MA, Lovly CM. The path(way) less traveled: a pathway-oriented approach to providing information about precision cancer medicine on My Cancer Genome. *Transl Oncol*. 2016;9(2):163–5.
- Rehm HL, Berg JS, Brooks LD, Bustamante CD, Evans JP, Landrum MJ, Ledbetter DH, Maglott DR, Martin CL, Nussbaum RL, Plon SE, Ramos EM, Shery ST, Watson MS, ClinGen. ClinGen—the Clinical Genome Resource. *N Engl J Med*. 2015;372(23):2235–42.
- Parsons DW, Roy A, Yang Y, Wang T, Scollon S, Bergstrom K, Kerstein RA, Gutierrez S, Petersen AK, Bavlle A, Lin FY, Lopez-Terrada DH, Monzon FA, Hicks MJ, Eldin KW, Quintanilla NM, Adesina AM, Mohila CA, Whitehead W, Jea A, Vasudevan SA, Nuchtern JG, Ramamurthy U, McGuire AL, Hilsenbeck SG, Reid JG, Muzny DM, Wheeler DA, Berg SL, Chintagumpala MM, Eng CM, Gibbs RA, Plon SE. Diagnostic yield of clinical tumor and germline whole-exome sequencing for children with solid tumors. *JAMA Oncol*. 2016;2(5):616–624.
- Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FC, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J, Vingron M. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat Genet*. 2001;29(4):365–71.
- Brazma A. Minimum Information About a Microarray Experiment (MIAME)—successes, failures, challenges. *ScientificWorldJournal*. 2009;9:420–3.
- Dienstmann R, Dong F, Borgers D, Dias-Santagata D, Ellisen LW, Le LP, Iafraite AJ. Standardized decision support in next generation sequencing reports of somatic cancer variants. *Mol Oncol*. 2014;8(5):859–73.
- Meric-Bernstam F, Johnson A, Holla V, Bailey AM, Brusco L, Chen K, Routbort M, Patel KP, Zeng J, Kopetz S, Davies MA, Piha-Paul SA, Hong DS, Eterovic AK, Tsimberidou AM, Broaddus R, Bernstam EV, Shaw KR, Mendelsohn J, Mills GB. A decision support framework for genomically informed investigational cancer therapy. *J Natl Cancer Inst*. 2015;107(7). 10.1093/jnci/djv098. Print 2015.
- Sawyers CL. The cancer biomarker problem. *Nature*. 2008;452(7187):548–52.
- Sukhai MA, Craddock KJ, Thomas M, Hansen AR, Zhang T, Siu L, Bedard P, Stockley TL, Kamel-Reid S. A classification system for clinical relevance of somatic variants identified in molecular profiling of cancer. *Genet Med*. 2016;18(2):128–36.
- Van Allen EM, Wagle N, Stojanov P, Perrin DL, Cibulskis K, Marlow S, Jané-Valbuena J, Friedrich DC, Kryukov G, Carter SL, McKenna A, Sivachenko A, Rosenberg M, Kiezun A, Voet D, Lawrence M, Lichtenstein LT, Gentry JG, Huang FW, Fostel J, Farlow D, Barbie D, Gandhi L, Lander ES, Gray SW, Joffe S, Janne P, Garber J, MacConaill L, Lindeman N, Rollins B, Kantoff P, Fisher SA, Gabriel S, Getz G, Garraway LA. Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide precision cancer medicine. *Nat Med*. 2014;20(6):682–8.
- Forbes SA, Beare D, Gunasekaran P, Leung K, Bindal N, Boutselakis H, Ding M, Bamford S, Cole C, Ward S, Kok CY, Jia M, De T, Teague JW, Stratton MR, McDermott U, Campbell PJ. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res*. 2015;43(Database issue):D805–11.
- International Cancer Genome Consortium, Hudson TJ, Anderson W, Artez A, Barker AD, Bell C, Bernabe RR, Bhan MK, Calvo F, Eerola I, Gerhard DS, Guttmacher A, Guyer M, Hemsley FM, Jennings JL, Kerr D, Klatt P, Kolar P, Kusada J, Lane DP, Laplace F, Youyong L, Nettekoven G, Ozenberger B, Peterson J, Rao TS, Remacle J, Schafer AJ, Shibata T, Stratton MR, Vockley JG, Watanabe K, Yang H, Yuen MM, Knoppers BM, Bobrow M, Cambon-Thomsen A, Dressler LG, Dyke SO, Joly Y, Kato K, Kennedy KL, Nicolas P, Parker MJ, Rial-Sebbag E, Romeo-Casabona CM, Shaw KM, Wallace S, Wiesner GL, Zeps N, Lichter P, Biankin AV, Chabannon C, Chin L, Clement B, de Alava E, Degos F, Ferguson ML, Geary P, Hayes DN, Hudson TJ, Johns AL, Kasprzyk A, Nakagawa H, Penny R, Piris MA, Sarin R, Scarpa A, Shibata T, van de Vijver M, Futreal PA, Aburatani H, Bayes M, Botwell DD, Campbell PJ, Estivill X, Gerhard DS, Grimmond SM, Gut I, Hirst M, Lopez-Otin C, Majumder P, Marra M, McPherson JD, Nakagawa H, Ning Z, Puente XS, Ruan Y, Shibata T, Stratton MR, Stunnenberg HG, Swerdlow H, Velculescu VE, Wilson RK, Xue HH, Yang L, Spellman PT, Bader GD, Boutros PC, Campbell PJ, Flicek P, Getz G, Guigo R, Guo G, Haussler D, Heath S, Hubbard TJ, Jiang T, Jones SM, Li Q, Lopez-Bigas N, Luo R, Muthuswamy L, Ouellette BF, Pearson JV, Puente XS, Quesada V, Raphael BJ, Sander C, Shibata T, Speed TP, Stein LD, Stuart JM, Teague JW, Totoki Y, Tsunoda T, Valencia A, Wheeler DA, Wu H, Zhao S, Zhou G, Stein LD, Guigo R, Hubbard TJ, Joly Y, Jones SM, Kasprzyk A, Lathrop M, Lopez-Bigas N, Ouellette BF, Spellman PT, Teague JW, Thomas G, Valencia A, Yoshida T, Kennedy KL, Axton M, Dyke SO, Futreal PA, Gerhard DS, Gunter C, Guyer M, Hudson TJ, McPherson JD, Miller LJ, Ozenberger B, Shaw KM, Kasprzyk A, Stein LD, Zhang J, Haider SA, Wang J, Yung CK, Cros A, Liang Y, Gnaneshan S, Guberman J, Hsu J, Bobrow M, Chalmers DR, Hasel KW, Joly Y, Kaan TS, Kennedy KL, Knoppers BM, Lowrance WW, Masui T, Nicolas P, Rial-Sebbag E, Rodriguez LL, Vergely C, Yoshida T, Grimmond SM, Biankin AV, Bowtell DD, Cloonan N, de Fazio A, Eshleman JR, Etemadmoghadam D, Gardiner BB, Kench JG, Scarpa A, Sutherland RL, Tempero MA, Waddell NJ, Wilson PJ, McPherson JD, Gallinger S, Tsao MS, Shaw PA, Petersen GM, Mukhopadhyay D, Chin L, DePinho RA, Thayer S, Muthuswamy L, Shazand K, Beck T, Sam M, Timms L, Ballin V, Lu Y, Ji J, Zhang X, Chen F, Hu X, Zhou G, Yang Q, Tian G, Zhang L, Xing X, Li X, Zhu Z, Yu Y, Yu J, Yang H, Lathrop M, Tost J, Brennan P, Holcatova I, Zaridze D, Brazma A, Egevard L, Prokhorchouk E, Banks RE, Uhlen M, Cambon-Thomsen A, Viksna J, Ponten F, Skryabin K, Stratton MR, Futreal PA, Birney E, Borg A, Borresen-Dale AL, Caldas C, Foekens JA, Martin S, Reis-Filho JS, Richardson AL, Sotiriou C, Stunnenberg HG, Thoms G, van de Vijver M, van't Veer L, Calvo F, Birnbaum D, Blanche H, Boucher P, Boyault S, Chabannon C, Gut I, Masson-Jacquemier JD, Lathrop M, Pauporte I, Pivot X, Vincent-Salomon A, Tabone E, Theillet C, Thomas G, Tost J, Treilleux I, Calvo F, Bioulac-Sage P, Clement B, Decaens T, Degos F, Franco D, Gut I, Gut M, Heath S, Lathrop M, Samuel D, Thomas G, Zucman-Rossi J, Lichter P, Eils R, Brors B, Korbel JO, Korshunov A, Landgraf P, Lehach H, Pfister S, Radlwimmer B, Reifemberger G, Taylor MD, von Kalle C, Majumder PP, Sarin R, Rao TS, Bhan MK, Scarpa A, Pedersoli P, Lawlor RA, Delledonne M, Bardelli A, Biankin AV, Grimmond SM, Gress T, Klimstra D, Zamboni G, Shibata T, Nakamura Y, Nakagawa H, Kusada J, Tsunoda T, Miyano S, Aburatani H, Kato K, Fujimoto A, Yoshida T, Campo E, Lopez-Otin C, Estivill X, Guigo R, de Sanjose S, Piris MA, Montserrat E, Gonzalez-Diaz M, Puente XS, Jares P, Valencia A, Himmelbauer H, Quesada V, Bea S, Stratton MR, Futreal PA, Campbell PJ, Vincent-Salomon A, Richardson AL, Reis-Filho JS, van de Vijver M, Thomas G, Masson-Jacquemier JD, Aparicio S, Borg A, Borresen-Dale AL, Caldas C, Foekens JA, Stunnenberg HG, van't Veer L, Easton DF, Spellman PT, Martin S, Barker AD, Chin L, Collins FS, Compton CC, Ferguson ML, Gerhard DS, Getz G, Gunter C, Guttmacher A, Guyer M, Hayes DN, Lander ES, Ozenberger B, Penny R, Peterson J, Sander C, Shaw KM, Speed TP, Spellman PT, Vockley JG, Wheeler DA, Wilson RK, Hudson TJ, Chin L, Knoppers BM, Lander ES, Lichter P, Stein LD, Stratton MR, Anderson W, Barker AD, Bell C, Bobrow M, Burke W, Collins FS, Compton CC, DePinho RA, Easton DF, Futreal PA, Gerhard DS, Green AR, Guyer M, Hamilton SR, Hubbard TJ, Kallioniemi OP, Kennedy KL, Ley TJ, Liu ET, Lu Y, Majumder P, Marra M, Ozenberger B, Peterson J, Schafer AJ, Spellman PT, Stunnenberg

- HG, Wainwright BJ, Wilson RK, Yang H. International network of cancer genome projects. *Nature*. 2010;464(7291):993–8
17. Griffith M, Spies NC, Krysiak K, Coffman AC, McMichael JF, Ainscough BJ, Rieke DT, Danos AM, Kujan L, Ramirez CA, Wagner AH, Skidmore ZL, Liu CJ, Jones MR, Bilski RL, Lesurf R, Barnell EK, Shah NM, Bonakdar M, Trani L, Matlock M, Ramu A, Campbell KM, Spies GC, Graubert AP, Gangavarapu K, Eldred JM, Larson DE, Walker JR, Good BM, Wu C, Su AI, Dienstmann R, Jones SJM, Bose R, Spencer DH, Wartman LD, Wilson RK, Mardis ER, Griffith OL. CIVIC: a knowledgebase for expert-crowdsourcing the clinical interpretation of variants in cancer. *bioRxiv*. 2016. <http://dx.doi.org/10.1101/072892>.
 18. Madhavan S, Gauba R, Song L, Bhuvaneshwar K, Gusev Y, Byers S, Juhl H, Weiner L. Platform for Personalized Oncology: integrative analyses reveal novel molecular signatures associated with colorectal cancer relapse. *AMIA Jt Summits Transl Sci Proc*. 2013;2013:118.
 19. Madhavan S, Gusev Y, Harris M, Tanenbaum DM, Gauba R, Bhuvaneshwar K, Shinohara A, Rosso K, Carabet LA, Song L, Riggins RB, Dakshanamurthy S, Wang Y, Byers SW, Clarke R, Weiner LM. G-DOC: a systems medicine platform for personalized oncology. *Neoplasia*. 2011;13(9):771–83.
 20. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, Antipin Y, Reva B, Goldberg AP, Sander C, Schultz N. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012;2(5):401–4.
 21. Johnson A, Zeng J, Bailey AM, Holla V, Litzenburger B, Lara-Guerra H, Mills GB, Mendelsohn J, Shaw KR, Meric-Bernstam F. The right drugs at the right time for the right patient: the MD Anderson precision oncology decision support platform. *Drug Discov Today*. 2015;20(12):1433–8.
 22. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehms HL. ACMG Laboratory Quality Assurance Committee. 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*. 2015;17(5):405–24.
 23. Eilbeck K, Lewis SE, Mungall CJ, Yandell M, Stein L, Durbin R, Ashburner M. The Sequence Ontology: a tool for the unification of genome annotations. *Genome Biol*. 2005;6(5):R44.
 24. Jette N, Quan H, Hemmelgarn B, Drosler S, Maass C, Moskal L, Paoin W, Sundararajan V, Gao S, Jakob R, Ustun B, Ghali WA. IMECCHI Investigators. The development, evolution, and modifications of ICD-10: challenges to the international comparability of morbidity data. *Med Care*. 2010;48(12):1105–10.
 25. Fragoso G, de Coronado S, Haber M, Hartel F, Wright L. Overview and utilization of the NCI thesaurus. *Comp Funct Genomics*. 2004;5(8):648–54.
 26. Oncotree. CMO Tumor Type Tree. <http://oncotree.mskcc.org/oncotree/>. Accessed 15 July 2016.
 27. IOM (Institute of Medicine). *Evolution of Translational Omics: Lessons Learned and the Path Forward*. Washington: The National Academies Press; 2012.
 28. Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, Maciejewski A, Arndt D, Wilson M, Neveu V, Tang A, Gabriel G, Ly C, Adamjee S, Dame ZT, Han B, Zhou Y, Wishart DS. DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res*. 2014;42(Database issue):D1091–7.
 29. Gaudichon J, Jeanne-Pasquier C, Deparis M, Veysiere A, Heyndrickx M, Minckes O, Orbach D. Complete and repeated response of a metastatic ALK-rearranged inflammatory myofibroblastic tumor to crizotinib in a teenage girl. *J Pediatr Hematol Oncol*. 2016;38(4):308–11.
 30. Gannon HS, Kaplan N, Tsherniak A, Vazquez F, Weir BA, Hahn WC, Meyerson M. Identification of an "exceptional responder" cell line to MEK1 inhibition: clinical implications for MEK-targeted therapy. *Mol Cancer Res*. 2016;14(2):207–15.
 31. Kasaian K, Wiseman SM, Walker BA, Schein JE, Hirst M, Moore RA, Mungall AJ, Marra MA, Jones SJ. Putative BRAF activating fusion in a medullary thyroid cancer. *Cold Spring Harb Mol Case Stud*. 2016;2(2):a000729.
 32. Lin FY, Bergstrom K, Person R, Bavle A, Ballester LY, Scollon S, Raesz-Martinez R, Jea A, Birchansky S, Wheeler DA, Berg SL, Chintagumpala MM, Adesina AM, Eng C, Roy A, Plon SE, Parsons DW. Integrated tumor and germline whole exome sequencing identifies mutations in MAPK and PI3K pathway genes in an adolescent with rosette-forming glioneuronal tumor of the fourth ventricle. *Cold Spring Harb Mol Case Stud*. 2016;2(5):a001057. doi:10.1101/mcs.a001057.
 33. Wu TJ, Schriml LM, Chen QR, Colbert M, Crichton DJ, Finney R, Hu Y, Kibbe WA, Kincaid H, Meerzaman D, Mittra E, Pan Y, Smith KM, Srivastava S, Ward S, Yan C, Mazumder R. Generating a focused view of disease ontology cancer terms for pan-cancer data integration and analysis. *Database (Oxford)*. 2015;2015:bav032.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

