frontiers in Oncology

REVIEW
published: 19 September 2017
doi: 10.3389/fonc.2017.00214

Resources for Interpreting Variants in Precision Genomic Oncology Applications

*Hsinyi Tsang[1,2], KanakaDurga Addepalli[1,2] and Sean R. Davis[3]**

[1] Center for Biomedical Informatics and Information Technology, National Cancer Institute, National Institutes of Health, Gaithersburg, MD, United States, [2] Attain, LLC, McLean, VA, United States, [3] Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD, United States

Precision genomic oncology—applying high throughput sequencing (HTS) at the point-of-care to inform clinical decisions—is a developing precision medicine paradigm that is seeing increasing adoption. Simultaneously, new developments in targeted agents and immunotherapy, when informed by rich genomic characterization, offer potential benefit to a growing subset of patients. Multiple previous studies have commented on methods for identifying both germline and somatic variants. However, interpreting individual variants remains a significant challenge, relying in large part on the integration of observed variants with biological knowledge. A number of data and software resources have been developed to assist in interpreting observed variants, determining their potential clinical actionability, and augmenting them with ancillary information that can inform clinical decisions and even generate new hypotheses for exploration in the laboratory. Here, we review available variant catalogs, variant and functional annotation software and tools, and databases of clinically actionable variants that can be used in an *ad hoc* approach with research samples or incorporated into a data platform for interpreting and formally reporting clinical results.

Keywords: precision oncology, high throughput sequencing, genomic variation, cancer variants, precision medicine, databases, genetic

OPEN ACCESS

Edited by:
Angela Re,
University of Trento, Italy

Reviewed by:
Lawrence Schook,
University of Illinois at Chicago, United States
Myriam Alcalay,
Istituto Europeo di Oncologia, Italy

****Correspondence:***
Sean R. Davis
seandavi@gmail.com

Specialty section:
This article was submitted to Molecular and Cellular Oncology, a section of the journal Frontiers in Oncology

Received: *30 June 2017*
Accepted: *29 August 2017*
Published: *19 September 2017*

Citation:
Tsang H, Addepalli K and Davis SR (2017) Resources for Interpreting Variants in Precision Genomic Oncology Applications. Front. Oncol. 7:214. doi: 10.3389/fonc.2017.00214

1. INTRODUCTION

Genomic technologies and approaches have transformed cancer research and have led to the production of large-scale cancer genomics compendia (1, 2). The resulting molecular characterization and categorization of individual samples from such compendia has driven development of molecular subtypes cancers as well as enhanced understanding of the molecular etiologies of carcinogenesis (3–5). The development of novel and effective targeted therapies has proceeded in parallel with and been accelerated by deeper, faster, and broader genomic characterization (6), enabling early application of molecular characterization at the point of care to inform clinical decision-making (7–10) and to address resistance to primary therapy (11). Genomic characterization also has applications in immune approaches to cancer. For example, chimeric antigen receptor T-cell (CARt) therapy has shown great success in diseases with well-characterized antigens that are relatively tumor-specific (12) as identified by genomic profiling. Variously referred to as precision oncology (13), genomics-driven oncology (14), genomic oncology, and even simply as precision medicine, the paradigm

of applying high-throughput genomic approaches to patient samples is rapidly changing the landscape of oncology care and clinical oncology research.

Conventional approaches to clinical trials design may be inadequate due to molecular heterogeneity of tumors derived from a single primary tissue (15), leading to the adoption of basket, umbrella, and hybrid trials designs. A number of studies are ongoing to determine feasibility and potential impact of precision genomic oncology at the point-of-care (16–18). In addition to studies focused on identifying targetable mutations, immune-based therapeutic approaches are also being informed by HTS applied to patient samples (19–21).

One of the most recent developments in the field of precision oncology is the approval of Pembrolizumab (Keytruda), an anti-PD-1 antibody that functions as a checkpoint inhibitor, by the US Food and Drug Administration for treatment of solid tumors that show genetic evidence of mismatch repair and, therefore, carry very high mutational burdens (22). Pembrolizumab was previously approved for use in melanoma, but the most recent approval is the first that is targeting allows a drug to be used in a non-tissue-specific context in patients showing a specific genomic marker in any solid tumor (23).

As with any clinical testing modality, whether in a research setting or at the point-of-care, a clear understanding of the goals of applying the test is necessary when first designing the test and its validation. However, the flexibility and number of potential data items that arise from even a limited application of HTS has lead the US Food and Drug Administration (FDA) to begin to define its regulatory role (24) and, critically, how existing knowledge bases can be applied in real time to address findings from clinical HTS testing (25).

This review aims to provide an organized set of biological knowledge bases with relevance to the interpretation of small variants, defined as single nucleotide variants or short (on the order of 20 base pairs or fewer) insertions and deletions. The catalogs of observed variants section list large-scale catalogs of variants, useful for filtering known common polymorphisms and identifying previously identified cancer variants. When a variant observed in a clinical sample has not been seen but appears to affect the protein coding sequence, the functional annotation resources section presents a sampling of some of the most common software and databases for predicting the impact on protein function. Finally, we catalog several data products and knowledgebases have been developed to provide decision support (with strong disclaimers and caveats) directly linking observed variants to clinical intervention in point-of-care HTS applications. Integrating the various data sources described in this review with variants observed in individual patients can be accomplished with combinations of software tools for the manipulation of variant datasets.

1.1. Catalogs of Observed Germline and Somatic Variants

Databases of observed variation in normal populations, diseased individuals, and cancer compendia form the map onto which observed variants in patients are projected. Because of the vast quantities of genomic data and, specifically, DNA variants, there is a tension between providing rich, highly curated information about individual variants and producing the largest possible catalog of variants with manageable levels of curation. This section reviews some of the available catalogs (**Table 1**) of genomic variation observed in the germline as well as those that appear in tumors as somatic mutations. Note that many of the databases mentioned below overlap in data sources (some nearly completely), but they may differ in the amount and depth of curation, additional metadata added to each variant, speed of updates, and methods or formats for access.

1.2. Germline

Comprehensive catalogs of germline variants inform decisions about the frequency of variants as seen in the general population as well as to identify variants that are annotated as cancer associated. In the context of tumor sequencing, common variants are unlikely to be genomic drivers of carcinogenesis and are often filtered from a report of potential somatic variants. This filtering process is particularly important when tumor sequencing is not accompanied by matched normal sequencing. Additional germline databases that catalog disease-associated variants can be useful to begin to address familial risk and potentially pharmacogenomic loci (38, 39).

Perhaps the oldest of the variant catalogs, dbSNP contains 325,658,303 individual variant records (build 150, accessed

TABLE 1 | Catalogs of germline and somatic variants.

Resource	Variant Type	URL	Reference
dbSNP[a]	Germline and somatic	https://www.ncbi.nlm.nih.gov/projects/SNP/	(26)
COSMIC[a]	Somatic	http://cancer.sanger.ac.uk/cosmic	(27)
ClinVar[a]	Germline predisposition and somatic	https://www.ncbi.nlm.nih.gov/clinvar/intro/	(28)
gnomAD[b]	Germline	http://gnomad.broadinstitute.org/	(29)
69 genomes from CGI[c]	Germline	http://www.completegenomics.com/public-data/69-genomes/	(30)
Personalized Genome Project	Germline	http://www.personalgenomes.org/	(31)
NCI Genomic Data Commons	Germline and somatic	https://portal.gdc.cancer.gov/	(32)
cBioPortal	Somatic	http://www.cbioportal.org	(33, 34)
Intogen (Partial TCGA dataset)	Somatic	https://www.intogen.org/search	(35, 36)
Pediatric Cancer Genome Project	Somatic	http://explorepcgp.org	(37)

The most commonly used catalogs include dbSNP, COSMIC, ClinVar, and gnomAD.

[a]*Primary resources useful for all studies.*

[b]*Particularly useful for exome sequencing projects.*

[c]*Useful if the Complete Genomics platform was used.*

May 30, 2017) and is available in multiple formats, searchable, and linked to records in literature and other data resources and databases. While the vast majority of variants in dbSNP have been observed in individuals without cancer, somatic variants are included and annotated in the database. Because dbSNP is driven by community submission of variants, levels of evidence vary among individual variants. The genome Aggregation Database, or gnomAD (29, 40), contains information from 123,136 exomes and 15,496 whole-genomes from unrelated individuals sequenced as part of various disease-specific and population genetic studies (accessed May 30, 2017). These data were collected by numerous collaborations, underwent standard processing, and unified quality control and results area accessible as a searchable online database and as a downloadable VCF-format text file. ClinVar (28), maintained by the NIH National Center for Biotechnology Information (NCBI), is a freely available archive for interpretations of clinical significance of variants for reported conditions. Entries in ClinVar are taken directly from submitters and represent the relationship between variants and clinical significance. When multiple submissions concerning a single variant are available, ClinVar supplies high-level summaries of agreement or disagreement across submitters. Importantly, though, clinical significance in ClinVar is reported as supplied by the submitter. The Personalized Genome Project (31) provides a limited number of fully open-access genome sequencing results provided by volunteers with trait surveys and even some microbiome surveys of participants. A catalog of germline variants derived from 69 genomes sequenced using the Complete Genomics sequencing platform (30) may be useful for groups who have data generated from the same platform, particularly for identifying sequencing-platform-specific false positive results.

1.3. Somatic

Whereas databases of germline variants are useful to filter out variants unlikely to be directly involved in carcinogenesis, databases of somatic variants are useful to identify variants and their frequencies as observed in tumors. In some cases, identified variants may be associated with specific tumor types, offering mechanistic clues, particularly in the rare cancer setting where biological understanding may be limited.

Several catalogs of somatic variants have, at their core, variants derived from The Cancer Genome Atlas (TCGA). These databases vary in the pipelines used to define the variants, the level of annotation associated with individual variants, the proportion of TCGA included, and methods for accessing or querying. Recently, National Cancer Institute (NCI) has established the Genomic Data Commons (GDC) to harmonize clinical information and genomic results across enterprise cancer datasets (32), particularly those funded by NCI, such as TCGA. In addition to the adult tumors profiled as part of the TCGA, the NCI GDC also contains data from several pediatric tumors profiled as part of the Therapeutically Applicable Research To Generate Effective Treatments (TARGET) project (41). Cancer cell line data from the Cancer Cell Line Encyclopedia (CCLE) are also included (42) in the GDC data collection. The GDC is a modern data platform that provides multiple access methods, including a programmatic application programming interface (API), data file download, and web browser-based text and graphical queries and visualization. The International Cancer Genome Consortium (ICGC) is a large, international collaboration with a collection of 76 studies (including TCGA studies) encompassing 21 tissue primary sites. Like the NCI GDC, the ICGC data portal provides modern data platform approaches to data access, visualization, and query (43). The Catalog of Somatic Mutations in Cancer (COSMIC) database is perhaps the largest and best-known cancer variant database. It presents a unified dataset consisting of curated cancer variants for specific genes as well as genomic screens from projects, such as TCGA. Several other cancer variant data resources are listed in **Table 1**.

2. FUNCTIONAL ANNOTATION RESOURCES

When faced with variants with little or no literature or database support, differentiating those that variants that are likely to be deleterious, perhaps contributing to carcinogenesis, versus those that likely are tolerated by the cell is a critical task, particularly in the setting of clinical precision genomic oncology. Note that determing that a variant is deleterious is not likely to result in a change in diagnosis, prognosis, or therapy. However, prioritizing variants for further study, research interest, and for discussion in forums such as a molecular tumor board is a valuable and necessary aspect of applying genomic technologies in the clinical arena.

A number of algorithms and methods have been developed to predict the effect of observed variants on protein structure and function as well as the potential for clinical impact. These prediction methods utilize features of the variant and its context, such as sequence identity, sequence conservation, evolutionary relationship, protein primary and secondary structure, entropy-based protein stability, and approaches such as clustering based on sequence alignments and machine learning. Some of them are specific to the type of variant or mutation, some to a disease type, and some more general. Therefore, applying these functional annotational tools and interpreting the results in a clinical or research setting may require significant human curation before being recognized as clinically actionable. Here, we present a review of a representative set of approaches for predicting pathogenicity of different variants. For a comprehensive list of prediction tools and their details, see **Table 2**. For more detailed scientific and technical explanations of these methods, we refer the reader to a comprehensive review (44).

2.1. SIFT

Sorting Intolerant From Tolerant, or SIFT, that predicts functional impacts of amino acid substitutions (48) is one of the earliest variant effect prediction tools and represents the class of prediction algorithms that utilizes protein conservation. It has since been updated and an online version of the tool is available (67). SIFT uses sequence homology, as measured by protein-level conservation, to classify variants based as tolerated or deleterious based on the associated protein coding changes. SIFT has served as a benchmark against which other methods are compared because

TABLE 2 | Tools, software, and databases for functional prediction and annotation of variant impact.

Resource	URL	Reference	Notes
Integrated predictive methods and aggregated databases			
dbNSFP[a,b,c,d]	https://sites.google.com/site/jpopgen/dbNSFP	(45)	Aggregated database of variant information
myvariant.info[a]	http://myvariant.info/	(46)	Aggregated database of variant information
Functional effect prediction software and algorithms			
PolyPhen-2[b]	http://genetics.bwh.harvard.edu/pph2	(47)	Bayesian classification
SIFT[b]	http://sift.jcvi.org	(48)	Alignment scores
MutationAssessor	http://mutationassessor.org	(27)	Conservation, naive Bayes classifier
MutationTaster	http://www.mutationtaster.org	(49)	
PROVEAN	http://provean.jcvi.org/index.php	(50)	
CADD[b,c]	http://cadd.gs.washington.edu	(51)	
GERP++[c]	http://mendel.stanford.edu/SidowLab/downloads/gerp/index.html	(52)	
PhyloP and PhastCons	http://compgen.cshl.edu/phast/index.php	(53, 54)	
nsSNPAnalyzer	http://snpanalyzer.uthsc.edu/	(55)	Random Forest
SNPs&GO	http://snps-and-go.biocomp.unibo.it/snps-and-go/	(56)	SVM
SNAP2	https://rostlab.org/services/snap2web/	(57)	Neural Networks
SNPs3D	http://www.snps3d.org/	(58)	Structure and sequence analysis
MutPred2	http://mutpred.mutdb.org/	(59)	Random Forest
AUTO-MUTE	http://binf2.gmu.edu/automute/	(60)	Topology and statistical contact potential
Panther	http://www.pantherdb.org/tools/csnpScoreForm.jsp	(61)	Hidden Markov Model
stSNP	http://ilyinlab.org/StSNP/	(62)	Comparative modeling of protein structure
Condel[b]	http://bg.upf.edu/fannsdb/	(63)	A weighted average of multiple methods
CoVEC	https://sourceforge.net/projects/covec/files		
CAROL[b]	http://www.sanger.ac.uk/science/tools/carol	(64)	Combines PolyPhen-2 and SIFT
Cancer-specific prediction tools			
CHASM	http://wiki.chasmsoftware.org/index.php/Main_Page	(65)	Random Forest
CanDrA	http://bioinformatics.mdanderson.org/main/CanDrA#CanDrA	(66)	96 structural, evolutionary and gene features

[a]*Aggregated databases combine outputs of other databases and algorithms are, therefore, efficient resources to use in annotation pipelines. Adding these resources to observed variants is supported software in* ***Table 4*** *including Ensembl VEP software (noted[b] in this table), Annovar (noted[c]), and snpEff (noted[a]).*

of its relative simplicity. SIFT considers the type of amino acid change induced by a genomic variant and the position at which the change/mutation occurs. SIFT relies on the presence of sequences from which conservation can be determined; variants for which such databases are limited will potentially lack robust predictions.

2.2. PolyPhen-2

Polymorphism Phenotyping v2, or PolyPhen2, predicts the effecting of coding non-synonymous SNPs on protein structure and function and annotates them (47). This algorithm uses a naive Bayes approach to combine information across a panel of 3D structural, sequence-based, and conservation-based features. Trained on two datasets, HumDiv and HumVar, and associated non-deleterious controls, the PolyPhen2 algorithm represents a class of multivariate prediction algorithms that employ machine learning and multiple features of variant impact.

2.3. Mutation Assessor

Mutation Assessor is an algorithm and tool that, such as SIFT, uses a conservation-based approach. However, Mutation Assessor also incorporates evolutionary information in an attempt to account for shifts in function between subfamilies of proteins (27), potentially extending the functional annotation of variants to "switch of function" as well as loss or gain of function. By quantifying the impact to conserved residues both globally and within subfamilies (residues that distinguish subfamilies from each other are thought to be less tolerant to change), Mutation Assessor defines a functional impact score to predict which variants are likely to be deleterious.

2.4. CONDEL

The CONsensus DELeteriousness, or CONDEL score, is an integrated prediction method for missense mutations that is relatively easy to extend with additional prediction resources (63). Originally implemented as a weighted average of the normalized scores from the output of two computational tools, Mutation Assessor and FATHMM, CONDEL can be extended or adapted to data at hand and represents an "aggregator" approach to variant effect prediction. Condel scores can be derived for a limited set of specified mutations via an online web application. The Ensembl database provides a variation of position-specific CONDEL predictions that combine SIFT and Polyphen-2 for every possible amino acid substitution in all human proteins.

2.5. CHASM

Cancer-specific High-throughput Annotation of Somatic Mutations, or CHASM, is a computational method that identifies and prioritizes the missense mutations likely to enhance tumor cell proliferation (65). CHASM uses machine learning to classify putative "driver" cancer mutations as distinct from "passenger" mutations. Training the CHASM model employed *in silico* simulation to generate realistic "passenger" mutations, specifically modeled to represent variant context and genes that are observed in cancer settings. Multiple features of the variants, including their DNA and protein contexts, were then used to build a machine

learning approach that attempted to maximize the specificity of separating driver mutations from passenger mutations. CHASM represents a relatively specific algorithm focused not on "deleteriousness" but, rather, on the likelihood that an observed variant is a cancer "driver."

2.6. dbNSFP

Recognizing that applying all of the effect prediction tools available is potentially challenging (45), developed a database that aggregates predictions for *all* possible SNVs associated with coding changes (in Gencode gene models). With more than ten different prediction algorithms and extensive additional annotation, this database can be a useful one-stop-shop for adding annotations to variant datasets. The snpEff suite (described below) can be used in conjunction with dbNSFP to efficiently annotate SNPs with the potential to effect coding genes.

3. CLINICAL ACTIONABILITY

The ultimate goal for many of the abovementioned resources is to develop an individualized approach to the diagnosis, prevention, and treatment of cancer, or precision oncology. However, despite recent advances in HTS, determining the clinical relevance of experimentally observed cancer variants remains a challenge in the application of HTS in clinical practice. Difficulties in differentiating driver and passenger mutations, lack of standards and guidelines in reporting and interpretation of genomic variants, lack of clinical evidence in associating genomic variants to clinical outcome, lack of resources to disseminate clinical knowledge to the cancer community, and the precise definition of actionability have been reported to contribute to the bottleneck (68–71). Comprehensive resources linking experimentally determined cancer variants and clinical actionability have been developed to address some of these challenges and address various aspects of translating research results into clinical valuable information to support clinical decisions in precision oncology (see **Table 3**). In recognition of the fact that central curation of information regarding actionability is extremely challenging, several of the resources below use crowdsourcing as a means of gathering updates and enhancing curation efforts. In addition to a web interface, some tools provide additional access via API, mobile app, and/or social media tagging to facilitate dissemination of information and enhance accessibility. While some of these tools share similar functions, in the section below, we highlight distinct features and capabilities for a representative set of resources that might be used as a "starter" set for clinical annotation of variants.

The myvariant.info database is one of the newest and attempts to provide a "one-stop-shop" for variants. It is included in this section because it has recently incorporated the CIViC and Cancer Genome Interpreter databases. In addition, it provides annotations for SNVs from multiple other data sources (a growing list, so see the site for updates) and aggregates functional annotations for variants present in its database, making it a good all-around tool for cancer variant annotation. It is available as a performant web API only at this time.

Clinical Interpretation of Variants in Cancer (CIViC) is an open access and open source platform for community-driven curation and interpretation of cancer variants. It is based on a crowdsourcing model where individuals in the community can contribute to produce a centralized knowledge base with the goal of disseminating knowledge and encouraging active discussion. Users, including patients, patient advocates, clinicians, and researchers, can participate, along with community editors, in various stages of interpreting the clinical significance of cancer variants using standards and guidelines developed by community experts (68, 72).

The Drug Gene Interaction Database (DGIdb) is an open source and open access platform for gene and drug annotation for known interaction and potential druggability. Users can cross-reference genes of interest and drugs against up to 15 sources and in functionally classified gene categories (73, 74). Cancer Genome Interpreter (CGI) identifies mutational events that are biomarkers of drug response or interact with known chemical compounds (75). PharmGKB is a pharmacogenomic resource for building clinical implementation and interpretation based on annotating, integrating, and aggregating knowledge extracted from research-level publications. It provides scored clinical annotation, prescription annotation (drug dosing, prescribing information), as well as pharmacokinetics/pharmacodynamics (PK/PD) annotation, with primary literature reference.

TABLE 3 | In a clinical setting, these databases are the most relevant, as they are maintained to provide clinically actionable and curated content.

Resource	URL	Reference	Crowd-sourcing used	Bulk access
myvariant.info[a]	http://myvariant.info/	(46)	Yes	API[a]
CIViC[a]	https://civic.genome.wustl.edu/home	(72)	Yes	API, Download
DGIdb[a]	http://dgidb.genome.wustl.edu/	(73, 74)	Yes	API, Download
Cancer Genome Interpreter[a]	https://www.cancergenomeinterpreter.org/home	(75)	Yes	API
OncoKb[a]	http://oncokb.org/	(76)		API
Cancer Driver Log	https://candl.osu.edu/	(77)	Yes	Download
Clinical Knowledge Base	https://www.jax.org/clinical-genomics/clinical-offerings/ckb			
My Cancer Genome	http://www.mycancergenome.org	(78)	Yes	(licensed) API
Personalized Cancer Therapy	https://pct.mdanderson.org		Account required	
PharmGKB	https://www.pharmgkb.org/	(79)	Yes	Download
Precision Medicine Knowledge Base (Beta)	https://pmkb.weill.cornell.edu/	(80)	Yes	

While evalutation of each database by both clinical and informatics team members, databases marked with "[a]" are maintained, recently (or continuously) updated, and curated. The myvariant.info database includes both CiVIC and Cancer Genome Interpreter data. The last column in the table notes bulk access approaches as these are relevant when including databases in an annotation pipeline or automated report.

OncoKb contains information on the clinical implication of specific genetic alterations in cancer. Each variant is annotation from multiple sources and scored using Levels of Evidence ranging from Level 1, which includes FDA-approved biomarker predictive of response to an FDA-approved drug, to Level 2, which includes variants for which an FDA-approved or standard of care treatment is available, Level 3 and Level 4 contain variants with investigational and hypothetical therapeutic implications, respectively. A similarly structured scoring system is available for indicating therapeutic implications for variants associated with resistance (76). Cancer Driver Log (CanDL), an expert-curated database for potential driver mutations in cancer, employs a similar four-level scoring system based on FDA approval, clinical, pre-clinical, and experimental functional evidence (77).

MyCancerGenome (MCG) is a knowledge resource highlighting the implication of tumor mutation on cancer care. It allows users to access its content via a mobile app and provide patient-focused information. Patients can access a database entitled DNA-mutation Inventory to Refine and Enhance Cancer Treatment (DIRECT) for Epidermal Growth Factor Receptor (EGFR) mutation for non-small cell lung cancer (NSCLC). Personalized Cancer Therapy (PCT) at the MD Anderson Cancer Center is a resource for clinical response associated with cancer variants and aims to facilitate patient involvement in biomarker-related clinical trials. Drug effectiveness is associated with a specific biomarker and scored based on prospective clinical study as well as Food and Drug Administration (FDA) approval.

4. TOOLS FOR MANIPULATING VARIANT DATASETS

Processing sequence data with the goal of determining variants (somatic or germline) often end with a file in Variant Call Format (VCF format), a loose, self-describing data standard describing variants along a genome, associated statistical and numeric metrics for each variant, and information integrated from data resources such as those described in the preceding sections (81). An ecosystem of tools, listed in **Table 4**, has been developed for basic transformations, manipulations, merge operations, and for adding transcript, protein, and higher-level functional annotations to variants in a VCF file. The vt and bcftools software suites perform operations such as slicing by genomic coordinate, data compression, and, importantly variant normalization, rendering variants more readily comparable across resources. Annovar (82, 83) and the SnpEff suite (84) add annotations relative to gene annotations, including information about transcript and protein-coding changes. The Ensembl Variant Effect Predictor (VEP) utilizes Ensembl gene models to annotate variants in gene context and offers an interesting plugin architecture that supports adding variant information from resources in (**Table 1**) (85). Recently, several software developers of variant annotation tools have developed a standard for reporting gene-centric annotations that has simplified post-processing of variants after annotation. Finally, tools such as Vcfanno (86) have been developed that can flexibly add fields to variants in a VCF file based on relatively sophisticated logic and data transformations, reducing the number of tools required to bring a new data resource into the annotation pipeline.

TABLE 4 | Software tools for manipulating and adding annotations to variant datasets.

Software	URL	Reference
vt	http://genome.sph.umich.edu/wiki/Vt	(87)
bcftools	http://www.htslib.org/download/	(88)
ANNOVAR	http://annovar.openbioinformatics.org/en/latest/	(83)
Ensembl Variant Effect Predictor (VEP)	http://www.ensembl.org/vep	(85)
SnpEff	http://snpeff.sourceforge.net/	(84)
Oncotator	https://portals.broadinstitute.org/oncotator/	(89)
vcfanno	https://github.com/brentp/vcfanno	(86)

Variant calling produces a list of observed variants. The tools in this table are useful for adding biological interpretation and for annotating the variants with information from resources in ***Tables 1–3****.*

5. DISCUSSION

5.1. Pragmatic Details

Despite advanced toolsets for manipulating variant files and increasing adoption available standard formats, practical pitfalls and challenges remain to the basic manipulation of variant datasets. Some data resources are available in multiple formats and not all formats contain identical information. Matching variants between resources and observed variants can be challenging, as some variants can be represented validly in multiple forms. Ideally, variants are cataloged with clarity with respect to a reference genome and, whenever possible, using HGVS nomenclature (90). In spite of increasing awareness and uptake of HGVS standard nomenclature, the critical step of matching variants across tools and databases in assessing clinical significance is still hampered by inconsistencies across tools and databases (91). Particularly, when handling clinical samples, an information system that provides results from multiple resources when assessing novel variants, incorporates *in silico* controls when adding or updating data resources (to avoid introducing errors), and adheres to HGVS nomenclature wherever possible in data processing pipelines can increase the likelihood of discovering potentially relevant variants.

5.2. Where to Start?

This review is meant to be comprehensive, so the reader might wonder "Where do we start?." While it is difficult to make hard-and-fast recommendations about what resources, tools, and databases are "the best" given the lack of gold-standard datasets on which to base such evalutations, annotations in **Tables 1–3** are meant to provide context for prioritization. The context for sequencing (clinical or not, targeted mutations, trial setting, or novel variant and biomarker discovery) will also drive annotation pipeline development. Not all data resources need to be added simultaneously if developing a pipeline for annotating cancer variants for precision oncology applications. In a clinical setting, targeting the reporting workflow and working with clinicians to understand the most relevant annotations is the most efficient

approach to determining relevant resources for annotation. Developing a modular informatics pipeline, perhaps using a computational workflow framework (https://github.com/pditommaso/awesome-pipeline) that can be easily extended and re-run on previously annotated data is helpful to keep pace with the rapidly changing and growing collection of annotation resources. Newer aggregation resources such as myvariant.info offer a wholistic solution (annotation, catalog, and clinical actionability), but with some risk of "lossiness" with respect to the primary resources contained within.

Finally, given the rapid pace of new development in this space, we have established a crowd-sourced list of cancer variant resources for precision medicine available at https://github.com/seandavi/awesome-cancer-variant-databases.

5.3. Conclusion

Robust sequencing technologies and increasingly reliable bioinformatics pipelines, combined with parallel development of therapeutics and diagnostics has bolstered the field of precision genomic oncology. However, the sheer number of resources available that can inform the interpretation of small variants is staggering, except for the very few variants with well-established clinical relevance or an associated targeted therapy. This review has highlighted a number of important data resources individually. For other variants, data integration remains a significant hurdle to the rapid turnaround required to apply HTS in a clinical context. Expert panel review (the molecular tumor board) has been effective for some groups (13, 92, 93) while other groups have adopted a protocol-based approach (94). Even when molecularly targetable lesions are identified, barriers to delivering therapy have been observed, limiting the impact of precision genomic oncology in some settings (95). Not covered in this review is the increasing utility of HTS in the burgeoning field of immunotherapy, where early efforts to predict response based on HTS results have been promising (19, 96, 97).

Some interesting trends are evident in the databases and resources presented in this review that highlight the overarching trends in delivering precision medicine. First is the sheer volume and rapid growth of numbers of observations to learn about the spectrum of variation cancer and normal genomes. Projects such as GnomAD, COSMIC, and other data sharing efforts enhance precision by cataloging rare variants as well as precise estimates of the frequencies of common variants. Second is the use of crowd-sourcing to produce rich clinical annotation (e.g., CiVIC) in response to the need for intensive human interaction to interpret the clinical impact of a variant or its relationship to potential medical intervention. On the other hand, with volumes of data ever-increasing, machine learning techniques drive many of the most commonly used approaches for assigning scores for impact of observed variants. As well-annotated datasets and variant catalogs grow, application of machine learning will become both more common and more powerful.

While significant progress has been made in applying technology to precision oncology, cancer arises in an individual after a typically complex and incompletely understood set of oncogenic events that are increasingly observable at the molecular level. Progress in cancer prevention, early detection, diagnosis, prognosis, and treatment is increasingly driven by insight gained through the analysis and interpretation of large genomic, proteomic, and pharmacological knowledge bases. Reductionist approaches to cancer biology can achieve only limited success in understanding cancer biology and improving therapy. Cancer is a disease associated with disruption of normal cellular circuitry and processes that leads to abnormal or uncontrolled proliferative growth, characterized by a complex spectrum of biochemical alterations that affects biological processes at multiple scales from the molecular activity and cellular homeostasis to intercellular and inter-tissue signaling. The cancer research community has made great strides in measuring the oncogenic events that lead to the development of cancer and therapy resistance. Because of the complexity inherent in protein networks, intercellular signaling, cellular heterogeneity, and the dynamic nature of cancer, future progress will require a more wholistic approach to precision oncology, including multiscale systems and modeling approaches that address the interrelatedness of the biological processes underlying cancer.

AUTHOR CONTRIBUTIONS

SD initiated the manuscript. SD, KA, and HT all contributed to the writing and editing of the manuscript.

FUNDING

This work was supported by the National Cancer Institute Center for Biomedical Informatics and Information Technology and the National Cancer Institute Center for Cancer Research in the Intramural Research Program at the National Institutes of Health.

REFERENCES

1. International cancer genome consortium. (2017). Available from: http://icgc.org
2. Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, et al. The cancer genome atlas pan-cancer analysis project. *Nat Genet* (2013) 45(10):1113–20. doi:10.1038/ng.2764
3. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* (2012) 490(7418):61–70. doi:10.1038/nature11412
4. Cancer Genome Atlas Research Network. The molecular taxonomy of primary prostate cancer. *Cell* (2015) 163(4):1011–25. doi:10.1016/j.cell.2015.10.025
5. Network TCGAR. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* (2008) 455(7216):1061–8. doi:10.1038/nature07385
6. Blumenthal GM, Mansfield E, Pazdur R. Next-generation sequencing in oncology in the era of precision medicine. *JAMA Oncol* (2016) 2(1):13–4. doi:10.1001/jamaoncol.2015.4503
7. Flaherty KT, Infante JR, Daud A, Gonzalez R, Kefford RF, Sosman J, et al. Combined BRAF and MEK inhibition in melanoma with BRAF V600 mutations. *N Engl J Med* (2012) 367(18):1694–703. doi:10.1056/NEJMoa1210093
8. Shaw AT, Kim D-W, Nakagawa K, Seto T, Crinó L, Ahn M-J, et al. Crizotinib versus chemotherapy in advanced ALK-positive lung cancer. *N Engl J Med* (2013) 368(25):2385–94. doi:10.1056/NEJMoa1214886

9. Maemondo M, Inoue A, Kobayashi K, Sugawara S, Oizumi S, Isobe H, et al. Gefitinib or chemotherapy for non-small-cell lung cancer with mutated EGFR. *N Engl J Med* (2010) 362(25):2380–8. doi:10.1056/NEJMoa0909530
10. Mughal TI, Schrieber A. Principal long-term adverse effects of imatinib in patients with chronic myeloid leukemia in chronic phase. *Biologics* (2010) 4:315–23. doi:10.2147/BTT.S5775
11. Ai J, Tiu RV. Practical management of patients with chronic myeloid leukemia who develop tyrosine kinase inhibitor-resistant BCR-ABL1 mutations. *Ther Adv Hematol* (2014) 5(4):107–20. doi:10.1177/2040620714537865
12. Grupp SA, Kalos M, Barrett D, Aplenc R, Porter DL, Rheingold SR, et al. Chimeric antigen receptor-modified T cells for acute lymphoid leukemia. *N Engl J Med* (2013) 368(16):1509–18. doi:10.1056/NEJMoa1215134
13. Sohal DPS, Rini BI, Khorana AA, Dreicer R, Abraham J, Procop GW, et al. Prospective clinical study of precision oncology in solid tumors. *J Natl Cancer Inst* (2016) 108(3):djv332. doi:10.1093/jnci/djv332
14. Garraway LA. Genomics-driven oncology: framework for an emerging paradigm. *J Clin Orthod* (2013) 31(15):1806–14. doi:10.1200/JCO.2012.46.8934
15. Simon R. Genomic alteration-driven clinical trial designs in oncology. *Ann Intern Med* (2016) 165(4):270–8. doi:10.7326/M15-2413
16. Cheng DT, Mitchell TN, Zehir A, Shah RH, Benayed R, Syed A, et al. Memorial sloan kettering-integrated mutation profiling of actionable cancer targets (MSK-IMPACT): a hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. *J Mol Diagn* (2015) 17(3):251–64. doi:10.1016/j.jmoldx.2014.12.006
17. NCI-MATCH trial (molecular analysis for therapy choice). (2017). Available from: https://www.cancer.gov/about-cancer/treatment/clinical-trials/nci-supported/nci-match
18. Lopez-Chavez A, Thomas A, Rajan A, Raffeld M, Morrow B, Kelly R, et al. Molecular profiling and targeted therapy for advanced thoracic malignancies: a biomarker-derived, multiarm, multihistology phase II basket trial. *J Clin Oncol* (2015) 33(9):1000–7. doi:10.1200/JCO.2014.58.2007
19. Bethune MT, Joglekar AV. Personalized T cell-mediated cancer immunotherapy: progress and challenges. *Curr Opin Biotechnol* (2017) 48:142–52. doi:10.1016/j.copbio.2017.03.024
20. Chalmers ZR, Connelly CF, Fabrizio D, Gay L, Ali SM, Ennis R, et al. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Med* (2017) 9(1):34. doi:10.1186/s13073-017-0424-2
21. Faltas B, Bhinder B, Beltran H, Tagawa ST, Molina AM, Nanus DM, et al. Generating a neoantigen map of advanced urothelial carcinoma by whole exome sequencing. *J Clin Oncol* (2016) 34(2_suppl):354. doi:10.1200/jco.2016.34.2_suppl.354
22. Le DT, Durham JN, Smith KN, Wang H, Bartlett BR, Aulakh LK, et al. Mismatch-repair deficiency predicts response of solid tumors to PD-1 blockade. *Science* (2017) 357:409–13. doi:10.1126/science.aan6733
23. Garber K. In a major shift, cancer drugs go 'tissue-agnostic'. *Science* (2017) 356(6343):1111–2. doi:10.1126/science.356.6343.1111
24. FDA. *Optimizing FDA's Regulatory Oversight of Next Generation Sequencing Diagnostic Tests—Preliminary Discussion Paper*. (2015). Available from: https://www.fda.gov/downloads/medicaldevices/newsevents/workshopsconferences/ucm427869.pdf
25. FDA. *Draft Guidance for Stakeholders and Food and Drug Administration Staff*. (2016). Available from: https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM509838.pdf
26. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* (2001) 29(1):308–11. doi:10.1093/nar/29.1.308
27. Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, et al. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res* (2016) 45:D777–83. doi:10.1093/nar/gkw1121
28. Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res* (2016) 44(D1):D862–8. doi:10.1093/nar/gkv1222
29. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* (2016) 536(7616):285–91. doi:10.1038/nature19057
30. Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* (2010) 327(5961):78–81. doi:10.1126/science.1181498
31. Church GM. The personal genome project. *Mol Syst Biol* (2005) 1:2005.0030. doi:10.1038/msb4100040
32. Grossman RL, Heath AP, Ferretti V, Varmus HE, Lowy DR, Kibbe WA, et al. Toward a shared vision for cancer genomic data. *N Engl J Med* (2016) 375(12):1109–12. doi:10.1056/NEJMp1607591
33. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cbio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* (2012) 2(5):401–4. doi:10.1158/2159-8290.CD-12-0095
34. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* (2013) 6(269):l1. doi:10.1126/scisignal.2004088
35. Rubio-Perez C, Tamborero D, Schroeder MP, Antolin AA, Deu-Pons J, Perez-Llamas C, et al. In silico prescription of anticancer drugs to cohorts of 28 tumor types reveals novel targeting opportunities. *Cancer Res* (2015) 75(15 Suppl):2983. doi:10.1158/1538-7445.AM2015-2983
36. Gonzalez-Perez A, Perez-Llamas C, Deu-Pons J, Tamborero D, Schroeder MP, Jene-Sanz A, et al. IntOGen-mutations identifies cancer drivers across tumor types. *Nat Methods* (2013) 10(11):1081–2. doi:10.1038/nmeth.2642
37. Downing JR, Wilson RK, Zhang J, Mardis ER, Pui C-H, Ding L, et al. The pediatric cancer genome project. *Nat Genet* (2012) 44(6):619–22. doi:10.1038/ng.2287
38. Wheeler HE, Maitland ML, Dolan ME, Cox NJ, Ratain MJ. Cancer pharmacogenomics: strategies and challenges. *Nat Rev Genet* (2013) 14(1):23–34. doi:10.1038/nrg3352
39. Relling MV, Evans WE. Pharmacogenomics in the clinic. *Nature* (2015) 526(7573):343–50. doi:10.1038/nature15817
40. gnomAD browser. (2017). Available from: http://gnomad.broadinstitute.org/
41. *Therapeutically Applicable Research to Generate Effective Treatments (TARGET)*. (2017). Available from: https://ocg.cancer.gov/programs/target
42. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* (2012) 483(7391):603–7. doi:10.1038/nature11003
43. Zhang J, Baran J, Cros A, Guberman JM, Haider S, Hsu J, et al. International cancer genome consortium data portal – a one-stop shop for cancer genomics data. *Database (Oxford)* (2011) 2011:bar026. doi:10.1093/database/bar026
44. Addepalli K. *Models Predicting Effects of Missense Mutations in Oncogenesis*. Ph.D. thesis. George Mason University (2014).
45. Liu X, Wu C, Li C, Boerwinkle E. dbNSFP v3.0: a one-stop database of functional predictions and annotations for human nonsynonymous and splice-site SNVs. *Hum Mutat* (2016) 37(3):235–41. doi:10.1002/humu.22932
46. Xin J, Mark A, Afrasiabi C, Tsueng G, Juchler M, Gopal N, et al. High-performance web services for querying gene and variant annotation. *Genome Biol* (2016) 17(1):91. doi:10.1186/s13059-016-0953-9
47. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet* (2013) Chapter 7: Unit 7.20. doi:10.1002/0471142905.hg0720s76
48. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res* (2003) 31(13):3812–4. doi:10.1093/nar/gkg509
49. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods* (2014) 11(4):361–2. doi:10.1038/nmeth.2890
50. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. *PLoS One* (2012) 7(10):e46688. doi:10.1371/journal.pone.0046688
51. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* (2014) 46(3):310–5. doi:10.1038/ng.2892
52. Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S. Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput Biol* (2010) 6(12):e1001025. doi:10.1371/journal.pcbi.1001025
53. Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, et al. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* (2005) 15(8):1034–50. doi:10.1101/gr.3715005

54. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res* (2010) 20(1): 110–21. doi:10.1101/gr.097857.109
55. Bao L, Zhou M, Cui Y. nsSNPAnalyzer: identifying disease-associated nonsynonymous single nucleotide polymorphisms. *Nucleic Acids Res* (2005) 33(Web Server issue):W480–2. doi:10.1093/nar/gki372
56. Calabrese R, Capriotti E, Fariselli P, Martelli PL, Casadio R. Functional annotations improve the predictive score of human disease-related mutations in proteins. *Hum Mutat* (2009) 30(8):1237–44. doi:10.1002/humu.21047
57. Hecht M, Bromberg Y, Rost B. Better prediction of functional effects for sequence variants. *BMC Genomics* (2015) 16(Suppl 8):S1. doi:10.1186/1471-2164-16-S8-S1
58. Yue P, Melamud E, Moult J. SNPs3D: candidate gene and SNP selection for association studies. *BMC Bioinformatics* (2006) 7(1):166. doi:10.1186/1471-2105-7-166
59. Pejaver V, Urresti J, Lugo-Martinez J, Pagel KA, Lin GN, Nam H-J, et al. MutPred2: inferring the molecular and phenotypic impact of amino acid variants. *bioRxiv* (2017) 134981. doi:10.1101/134981
60. Masso M, Vaisman II. AUTO-MUTE: web-based tools for predicting stability changes in proteins due to single amino acid replacements. *Protein Eng Des Sel* (2010) 23(8):683–7. doi:10.1093/protein/gzq042
61. Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, et al. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res* (2003) 13(9):2129–41. doi:10.1101/gr.772403
62. Uzun A, Leslin CM, Abyzov A, Ilyin V. Structure SNP (StSNP): a web server for mapping and modeling nsSNPs on protein structures with linkage to metabolic pathways. *Nucleic Acids Res* (2007) 35(Web Server issue):W384–92. doi:10.1093/nar/gkm232
63. González-Pérez A, López-Bigas N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, condel. *Am J Hum Genet* (2011) 88(4):440–9. doi:10.1016/j.ajhg.2011.03.004
64. Lopes MC, Joyce C, Ritchie GRS, John SL, Cunningham F, Asimit J, et al. A combined functional annotation score for non-synonymous variants. *Hum Hered* (2012) 73(1):47–51. doi:10.1159/000334984
65. Carter H, Chen S, Isik L, Tyekucheva S, Velculescu VE, Kinzler KW, et al. Cancer-specific high-throughput annotation of somatic mutations: computational prediction of driver missense mutations. *Cancer Res* (2009) 69(16): 6660–7. doi:10.1158/0008-5472.CAN-09-1133
66. Mao Y, Chen H, Liang H, Meric-Bernstam F, Mills GB, Chen K. CanDrA: cancer-specific driver missense mutation annotation with optimized features. *PLoS One* (2013) 8(10):e77945. doi:10.1371/journal.pone.0077945
67. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* (2009) 4(7):1073–81. doi:10.1038/nprot.2009.86
68. Li MM, Datto M, Duncavage EJ, Kulkarni S, Lindeman NI, Roy S, et al. Standards and guidelines for the interpretation and reporting of sequence variants in cancer. *J Mol Diagn* (2017) 19(1):4–23. doi:10.1016/j.jmoldx.2016.10.002
69. Prawira A, Pugh TJ, Stockley TL, Siu LL. Data resources for the identification and interpretation of actionable mutations by clinicians. *Ann Oncol* (2017) 28(5):946–57. doi:10.1093/annonc/mdx023
70. Uzilov AV, Ding W, Fink MY, Antipin Y, Brohl AS, Davis C, et al. Development and clinical application of an integrative genomic approach to personalized cancer therapy. *Genome Med* (2016) 8(1):62. doi:10.1186/s13073-016-0313-0
71. Hedley Carr T, McEwen R, Dougherty B, Johnson JH, Dry JR, Lai Z, et al. Defining actionable mutations for oncology therapeutic development. *Nat Rev Cancer* (2016) 16(5):319–29. doi:10.1038/nrc.2016.35
72. Griffith M, Spies NC, Krysiak K, Coffman AC, McMichael JF, Ainscough BJ, et al. CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. *Nat Genet* (2017) 49(2):170–4. doi:10.1038/ng.3774
73. Wagner AH, Coffman AC, Ainscough BJ, Spies NC, Skidmore ZL, Campbell KM, et al. DGIdb 2.0: mining clinically relevant drug-gene interactions. *Nucleic Acids Res* (2016) 44(D1):D1036–44. doi:10.1093/nar/gkv1165
74. Griffith M, Griffith OL, Coffman AC, Weible JV, McMichael JF, Spies NC, et al. DGIdb: mining the druggable genome. *Nat Methods* (2013) 10(12): 1209–10. doi:10.1038/nmeth.2689
75. Tamborero D, Rubio-Perez C, Deu-Pons J, Schroeder MP, Vivancos A, Rovira A, et al. Cancer genome interpreter annotates the biological and clinical relevance of tumor alterations. *bioRxiv* (2017) 140475. doi:10.1101/140475
76. Chakravarty D, Gao J, Phillips S, Kundra R, Zhang H, Wang J, et al. OncoKB: a precision oncology knowledge base. *JCO Precis Oncol* (2017) 1(1):1–16. doi:10.1200/PO.17.00011
77. Damodaran S, Miya J, Kautto E, Zhu E, Samorodnitsky E, Datta J, et al. Cancer driver log (CanDL): catalog of potentially actionable cancer mutations. *J Mol Diagn* (2015) 17(5):554–9. doi:10.1016/j.jmoldx.2015.05.002
78. Micheel CM, Lovly CM, Levy MA. My cancer genome. *Cancer Genet* (2014) 207(6):289. doi:10.1016/j.cancergen.2014.06.016
79. Hewett M, Oliver DE, Rubin DL, Easton KL, Stuart JM, Altman RB, et al. PharmGKB: the pharmacogenetics knowledge base. *Nucleic Acids Res* (2002) 30(1):163–5. doi:10.1093/nar/30.1.163
80. Huang L, Fernandes H, Zia H, Tavassoli P, Rennert H, Pisapia D, et al. The cancer precision medicine knowledge base for structured clinical-grade mutations and interpretations. Journal of the American Medical Informatics Association? *J Am Med Inform Assoc* (2017) 24(3):513–9. doi:10.1093/jamia/ocw148
81. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics* (2011) 27(15):2156–8. doi:10.1093/bioinformatics/btr330
82. Yang H, Wang K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nat Protoc* (2015) 10(10):1556–66. doi:10.1038/nprot.2015.105
83. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* (2010) 38(16):e164. doi:10.1093/nar/gkq603
84. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (2012) 6(2):80–92. doi:10.4161/fly.19695
85. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The ensembl variant effect predictor. *Genome Biol* (2016) 17(1):122. doi:10.1186/s13059-016-0974-4
86. Pedersen BS, Layer RM, Quinlan AR. Vcfanno: fast, flexible annotation of genetic variants. *Genome Biol* (2016) 17(1):118. doi:10.1186/s13059-016-0973-5
87. Tan A, Abecasis GR, Kang HM. Unified representation of genetic variants. *Bioinformatics* (2015) 31(13):2202–4. doi:10.1093/bioinformatics/btv112
88. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics* (2009) 25(16):2078–9. doi:10.1093/bioinformatics/btp352
89. Ramos AH, Lichtenstein L, Gupta M, Lawrence MS, Pugh TJ, Saksena G, et al. Oncotator: cancer variant annotation tool. *Hum Mutat* (2015) 36(4): E2423–9. doi:10.1002/humu.22771
90. den Dunnen JT, Dalgleish R, Maglott DR, Hart RK, Greenblatt MS, McGowan-Jordan J, et al. HGVS recommendations for the description of sequence variants: 2016 update. *Hum Mutat* (2016) 37(6):564–9. doi:10.1002/humu.22981
91. Yen JL, Garcia S, Montana A, Harris J, Chervitz S, Morra M, et al. A variant by any name: quantifying annotation discordance across tools and clinical databases. *Genome Med* (2017) 9(1):7. doi:10.1186/s13073-016-0396-7
92. Knepper TC, Bell GC, Hicks JK, Padron E, Teer JK, Vo TT, et al. Key lessons learned from moffitt's molecular tumor board: the clinical genomics action committee experience. *Oncologist* (2017) 22(2):144–51. doi:10.1634/theoncologist.2016-0195
93. Beltran H, Eng K, Mosquera JM, Sigaras A, Romanel A, Rennert H, et al. Whole-exome sequencing of metastatic cancer and biomarkers of treatment response. *JAMA Oncol* (2015) 1(4):466–74. doi:10.1001/jamaoncol.2015.1313
94. Ghazani AA, Oliver NM, St Pierre JP, Garofalo A, Rainville IR, Hiller E, et al. Assigning clinical meaning to somatic and germ-line whole-exome sequencing data in a prospective cancer precision medicine study. *Genet Med* (2017) 19:787–95. doi:10.1038/gim.2016.191
95. Bryce AH, Egan JB, Borad MJ, Stewart AK, Nowakowski GS, Chanan-Khan A, et al. Experience with precision genomics and tumor board, indicates

frequent target identification, but barriers to delivery. *Oncotarget* (2017) 8(16):27145–54. doi:10.18632/oncotarget.16057

96. Wang R-F, Wang HY. Immune targets and neoantigens for cancer immunotherapy and precision medicine. *Cell Res* (2017) 27(1):11–37. doi:10.1038/cr.2016.155
97. Yarchoan M, Johnson BA III, Lutz ER, Laheru DA, Jaffee EM. Targeting neoantigens to augment antitumour immunity. *Nat Rev Cancer* (2017) 17(4):209–22. doi:10.1038/nrc.2016.154

Conflict of Interest Statement: This work was performed while KA and HT were employed by Attain, LLC, in support of bioinformatics projects at the National Cancer Institute. The authors declare that the work was conducted in the absence of any commercial or financial relationships that constitute a potential conflict of interest.