Variability in Somatic Variant Interpretation: Can Software Help?

By Sheryl Elkin, Ph.D., Chief Scientific Officer, QIAGEN Digital Insights, Oncology

Next-generation sequencing (NGS) has facilitated a revolution in cancer research and therapy over the last decade. In the early 2000s, only single gene tests were available, and these tests interrogated a handful of specific mutations within those genes. The mutations were well-established and could be used to direct the application of only a few targeted therapies, most notably in breast and lung cancer. In contrast, NGS is used today to interrogate hundreds to thousands of genes, reporting known and unknown variants across the entire coding regions of genes, even expanding to intronic and intergenic regions of the genome. Hundreds of targeted therapies and, more recently, immunotherapies have been approved for use across over thirty different types of cancer, including solid tumors, hematologic malignancies, and cancers of the central nervous system.[1] Indications for these breakthrough therapies are guided by the results of NGS testing.

Beyond the selection of therapy, sequencing data is now being broadly applied to predict prognosis and refine diagnosis, particularly in hematologic malignancies.

For example, the World Health Organization (WHO), European Leukemia Network (ELN), and International Consensus Classification (ICC) guidance continues to expand, identifying increasing numbers of genes that can be used to derive more information about a patient's illness.[2-4]

With these dramatic improvements in sequencing power has come a now well-known proliferation of data. Laboratories are sequencing large panels of genes, exomes, and whole genomes that yield tremendous numbers of variants, many of which are novel. These laboratories face a bottleneck at the interpretation step: as research continues and science advances, the amount of information that these labs need to process is overwhelming. Furthermore, processed results cannot be reported to clinicians without helping them understand what the results mean for their patients. In addition to managing the crush of data, which requires additional labor for interpretation, laboratory directors also face increasing pressure to reduce test turnaround time.

Compounding this problem is a lack of consistency in variant interpretation, across both individual analysts and institutions. A recent study found that the variation in somatic variant classification among expert curators was as high as 28%, in a study set that included over 70 Variant Call Format (VCF) files spanning hematologic malignancies and solid tumors.[5] This reported level of discrepancy is consistent with the perception in the field that variant interpretation and classification is not a "solved" problem.

Laboratories strive for consistency and accuracy because test interpretation can directly impact patient care through identification of potential targeted therapies as well as prognostic and diagnostic implications. At the extremes, variants with significant evidence and variants with no evidence at all are typically straightforward to classify and exhibit the greatest amount of concordance. However, a large and ever-growing number of variants fall in the middle of the spectrum with varying levels of evidence for their importance; the evaluation and application of this evidence remains variable. Further, for somatic variants, both the biological and clinical impact of the variant must be considered. A variant may

be clearly associated with the development of cancer while experts may debate its level of clinical relevance with respect to therapy, prognosis, or diagnosis.

The question becomes two-fold: what factors contribute to differences in somatic variant interpretation and how can they be mitigated?

The role of disease specificity

Somatic variant interpretation includes two components: oncogenicity and clinical actionability. These components can be separated for the purposes of analysis, as a variant can be functionally oncogenic without providing any insight into clinical management. In the case of an oncogenic variant, the level of clinical actionability for the variant is also crucial for understanding the impact of the variant in the patient's treatment pathway. The stated diagnosis is a critical piece in the consistency of the actionability classification.

In 2017, a set of guidelines for the reporting of somatic variants was published by a collaborative group composed of the Association for Molecular Pathology (AMP), American Society of Clinical Oncology (ASCO) and College of American Pathologists (CAP).[6] Li et al define a set of "tiers" into which variants could be sorted for reporting, where Tier 1 comprises variants of "strong clinical significance;" Tier 2 comprises variants of "potential clinical significance;" Tier 3 comprises variants of "unknown clinical significance;" and Tier 4 comprises "benign" and "likely benign" variants. Li et al further lay out the criteria for fulfilling four levels of evidence, A through D, where levels A and B correspond to Tier 1 and levels C and D correspond to Tier 2. The assessment of the Tier depends on the oncogenicity classification of each variant (not described in Li et al), as well as the level of evidence supporting the therapeutic, prognostic, or diagnostic significance of the variant in a specified disease context.

The publication of these guidelines was widely applauded and broadly adopted, as the field had craved a common set of standards for reporting variants. Despite the acceptance of a common standard, classification of somatic variants according to the guidelines still varies widely from laboratory to laboratory. What causes this variability?

Certainly, guidelines are critical to improving consistency across variant interpreters; however, the assessment still requires analysis of data to determine whether it fulfils criteria. Differences among judgments applied to the evidence can lead to divergent classifications. Classifications of solid tumors tend to be more consistent overall, as the criteria are focused on the availability and evidence for targeted therapies that can be used to treat genetic alterations. The assignment of a Level A (Tier 1) classification requires an available and approved targeted therapy directed to the alteration in the patient's tumor type; this information is straightforward to ascertain and is not subject to judgment as to the level of the evidence.

Beyond Level A, the Tier assignment becomes more complex. Level B (also Tier 1) requires guideline level recommendations and/or "well-powered studies" supporting the association between a marker and a drug, while Level C (Tier 2) includes the qualification that the alteration be a criterion for a clinical trial; Level D (Tier 2) is met if there is preclinical data that the variant could predict response to a drug.

Differences between Level B and Level C (and therefore between Tier 1 and Tier 2) have more subjective elements. Assignment of Level B evidence can be clear if there are guidelines. But in the absence of guidelines, it requires evaluation of the studies, and the number and strength of studies required to attain Level B is not clearly defined. Indeed, the variation in clinical trial design, size, and endpoints makes a simple set of rules difficult to define.

> *"For hematologic malignancies, there are fewer approved targeted therapies than for solid tumors but many more mutations that are associated with prognostic and diagnostic significance."*

For example, an *FGFR2::BICC1* fusion in urothelial carcinoma may clearly be classified as a Tier 1 alteration with Level A evidence for erdafitinib. An *ERBB2* amplification in a salivary gland carcinoma, however, requires closer analysis: the NCCN guidelines state that HER2-targeted therapy, such as trastuzumab, may be applicable in certain circumstances, but it is not supported by "well-powered studies" and is not approved by the FDA for use in this context. Different analysts may choose to classify *ERBB2* amplification in salivary gland carcinoma as Tier 1B or Tier 2C.

For hematologic malignancies, there are fewer approved targeted therapies than for solid tumors but many more mutations that are associated with prognostic and diagnostic significance. As in solid tumors, the assignment of Level A prognostic and diagnostic significance is relatively straightforward; the primary criterion for Level A is the identification of the marker in guidelines, but the distinction between Level B and C (Tier 1 vs Tier 2) is more subjective and results in variability among analysts.

The level of consistency and specificity of the diagnosis itself also contributes to variability in variant Tier classification. For example, *NPM1* mutations are classified as Tier 1 (Level A diagnostic and prognostic) in acute myeloid leukemia (AML), but they do not fulfil the criteria for any level of evidence in other hematologic malignancies, such as myelodysplastic syndrome (MDS) or myeloproliferative neoplasms (MPN). Similarly, a *STAG2* mutation may be classified as Tier 1 with Level A prognostic and diagnostic significance in AML, but in MDS, the level of prognostic evidence may be lLevel B and diagnostic evidence may be Level C.

The Level B and Level C assessments may vary across laboratories. Further, many patients present to physicians with symptoms that are not associated with a clear diagnosis and thus, the diagnosis available at the time of variant assessment can be vague (e.g., myeloid neoplasm). If the diagnosis is unknown, what is the most appropriate classification for a *NPM1* or *STAG2* mutation? Laboratories differ widely in their approach to classifications in the case of a broad diagnosis, with some laboratories choosing the highest Tier available for a diagnosis under that umbrella and others selecting a lower level, reflecting the uncertainty of the diagnosis.

Diagnostic specificity can also affect classifications in solid tumors. The approval of erdafitinib for certain *FGFR2* and *FGFR3* alterations is specific for "urothelial carcinoma," which most commonly occurs in the bladder. Not all "bladder carcinoma" is urothelial; for example, should an *FGFR2* fusion in "bladder carcinoma" be considered a Tier 1 variant? Many laboratories would classify it as Tier 1, because most bladder carcinomas are urothelial. Others might classify it as Tier 2, because it could be a bladder adenocarcinoma or squamous cell carcinoma. Even more challenging are cases of cancer of unknown primary (CUP).

Some mutations are highly characteristic of a particular cancer type; a CUP with an *EGFR* mutation is almost certainly a non-small cell lung cancer. However, a CUP with a *PIK3CA* mutation could have originated from many different organs; *PIK3CA* mutations would be classified as Tier 1 in breast cancer based on the approval of alpelisib, but would be classified as Tier 2 in all other cancer types. Laboratories need to define rules to handle these cases, but even these rules may depend on the frequency with which the mutation is seen in one specific diagnosis compared to others. Accordingly, an *EGFR* mutation in a CUP case could be a Tier 1 alteration whereas a *PIK3CA* mutation could be a Tier 2 alteration.

Geographical differences can also complicate the consistency of Tier classification. Drug approval agencies vary from country to country, and while

some drugs are approved across multiple agencies, regional differences in drug approvals can cause discrepancies. Coming back to the *FGFR2* fusion discussed above, erdafitinib has been approved by the US FDA to target certain *FGFR2* fusions in urothelial carcinoma, but it has not been approved by the EMA. Accordingly, *FGFR2* fusions can be classified as Tier 2 in Europe while they are classified as Tier 1 in the US.

The role of guidelines

In general, the establishment of guidelines is expected to reduce variability among gene variant reviewers by providing a consistent framework for the evaluation and application of evidence. The American College of Medical Genetics and Genomics (ACMG) and Association for Molecular Pathology (AMP) released a set of guidelines for the interpretation of sequence variants in 2015.[7] Known in the industry as the "ACMG guidelines", they set out criteria for the classification of variants into five categories: pathogenic, likely pathogenic, unknown significance, likely benign, and benign. The ACMG guidelines are widely accepted as the industry standard for the classification of hereditary variants, and they have provided a reliable structure for consistent classification of variants.

Even with clear guidelines, there is room for interpretation and variation. For variants with strong evidence, the guidelines can be clearly applied, and there is little difference in classification from laboratory to laboratory. When the evidence is less clear or less abundant, scientists must apply their own judgment, and may reach different conclusions regarding a particular study. As a result, even with accepted guidelines, there is still disagreement in classification, particularly among variants that are in the gray area between "unknown significance" and "likely pathogenic" or "likely benign."

For example, *FGFR1* p.Arg78Cys and *AKT1* p.Gln79Arg both appear in ClinVar with "Conflicting interpretations of pathogenicity."[8] Both were assessed by established, well-respected laboratories (GeneDx and Invitae), and both were evaluated according to the ACMG guidelines. In both cases, however, the evaluators reached different conclusions, presenting assessments of both "uncertain significance" and "likely pathogenic" for these variants. Similarly, *BRCA1* p.Leu52Phe appears in ClinVar with 12 submissions: six classify the variant as "uncertain significance;" four classify it as "likely benign;" and two classify it as "benign." In all three cases, the variant analysts reached different conclusions using the same guidelines and the same body of evidence.

The ACMG guidelines can be applied to the somatic setting but require adaptation, and as of 2022, a single approach for the functional classification of somatic variants has not been firmly established. The absence of an accepted standard has allowed the development of multiple independent approaches, in some cases adapting the ACMG guidelines to the somatic case, and in other cases adopting a separate functional approach, classifying variants according to loss or gain of function. As can be imagined, independent approaches to functional classification permit significant variation among analysts. Additionally, in the absence of a clear standard, laboratories may apply different weight to different elements of evidence. For example, one laboratory may determine that a single somatic report in the disease of interest is sufficient to classify a variant as biologically relevant, whereas another laboratory may require three somatic reports or functional data demonstrating impact.

> *"Even with clear guidelines, there is room for interpretation and variation."*

To address this variant interpreter variability conundrum, a joint recommendation from the Clinical Genome Resource (ClinGen), Cancer Genomics Consortium (CGC) and Variant Interpretation for Cancer Consortium (VICC) published earlier this year provides a clear set of guidelines for determining the "oncogenicity" of a somatic variant.[9] The guidelines include a list of criteria, each assigned a numerical weight. The addition of the weights for each criterion satisfied by a variant would result in a number that can be assigned to a point on the oncogenicity spectrum, which corresponds to the pathogenicity classifications outlined by ACMG for hereditary variants. If this set of guidelines achieves broad adoption in the field, the variability among variant interpreters is expected to be significantly reduced.

As demonstrated with the continuing variability in variant interpretation using the 2015 ACMG guidelines, it is extremely difficult, and perhaps impossible, to design the "perfect" set of guidelines. The evaluation of scientific literature to determine whether a study meets one of the criteria still requires judgment which may vary among assessors. Additionally, the criteria themselves must be accepted by the community.

A recent study compared oncogenicity classifications performed according to the ClinGen/CGC/VICC guidelines to classifications performed using a clinical decision support software system, QIAGEN Clinical Insight (QCI®) Interpret One[10] (manuscript in preparation). QCI Interpret One utilizes a classification framework based on the ACMG variant classification guidelines, with adaptations for the somatic use case. The study compared the classification frameworks and compared the specific classifications returned for the variants in the original paper, as well as a separate study set of variants derived from clinical cases. The authors found that the classifications were approximately 90% concordant for the variant set described in the publication and 80% concordant for the experimental set of variants. The differences were primarily identified between adjacent classification levels, and in nearly all cases where there was disagreement, the QCI Interpret One software trended toward calling the variants likely pathogenic as opposed to VUS, or VUS as opposed to likely benign.

This data has implications for laboratory reporting. Some laboratories adopt a more conservative approach and do not want to "over-report" a variant, which could raise the possibility that a physician could make a change to disease management based on a variant that may not be clinically significant. In other laboratories, directors take a more liberal approach and include more variants of debatable significance, because they do not want to miss reporting on a variant that could be helpful for decision-making.

The assessment methodology is important to consider. If the variant classification is fully manual, common for small panels that result in small numbers of variants, the analyst can review all the data for each variant, and if the approach is conservative, they can decide to exclude a variant if the sum of the evidence places a variant in the VUS or likely benign category.

However, automated classification software like QCI Interpret One tends to handle large numbers of variants from larger sequencing panels. The software provides a computed classification for each variant and presents the variants to the user for review.

In a typical case, a comparatively small list of pathogenic and likely pathogenic variants appears at the top of a list that includes a lengthy list of VUS, likely benign, and benign variants. The intended use of the system requires a manual review and approval of the classifications by a variant scientist or laboratory director. The scientist has the opportunity to review the pathogenic and likely pathogenic variants and can downgrade any variants that they feel have been overclassified. It is much easier to identify a variant that needs to be downgraded from a short list than to identify a variant that needs to be upgraded from a long list of VUS and likely benign variants. The chance of error in this scenario is significantly reduced, and the manual portion of the interpretation process is streamlined.

The role of clinical decision support software

As panels grow and the number of variants to interpret increases, laboratories face numerous challenges. In addition to the need for consistent and accurate interpretation, laboratories are under intense pressure to reduce turnaround time. Many laboratories are turning to clinical decision support software systems and/or a curated knowledgebase. These systems can help to present information to the analyst, and many will provide a precomputed classification for the analyst to review. Some of these systems are publicly available, while others require a paid subscription. These systems and knowledgebases, including JAX Clinical Knowledgebase, OncoKB, CIViC, Navify, PierianDx, and QCI Interpret One,[5,11-15] employ varying degrees of manual curation and automation. Some support look-ups of individual variants, while others provide full reports of all variants within a case.

A recent study by Fairley et al[5] undertook an assessment of QCI Interpret One, comparing the performance in variant classification by the software system to that of eight individual laboratories using their internal methods, which could include manual analysis or analysis using an alternate software system. QCI Interpret One provides a unique combination of automation and manual assessment. Described in Fairley et al, the QCI Interpret One application leverages a large database of manually curated literature. Each publication is assessed by curators who extract the relevant information and convert it into computable units. In the context of each case, the software executes a set of rules, based on the AMP criteria for actionability and a modified version of the ACMG criteria for pathogenicity (accounting for the differences in somatic interpretation). The software then generates a computed classification for each variant. The variants are subsequently submitted for expert interpretation, whereby a team of scientists assesses each variant using a topic-based approach, collecting all the data for each variant and providing a holistic assessment and classification. The user is presented with both computed and expert classification and has the ability to review the data and make a final determination.

The Fairley study assessed 77 VCF files, across both solid tumors and hematological malignancies. Each VCF was assessed using QCI Interpret One and separately by an independent laboratory, and the results were compared to the classifications in the original reports from the submitting laboratory. Any discrepant results were reviewed by a panel of experts. After all review was complete, the classifications from QCI Interpret One and the expert panel were concordant in 91% of variants, with only 9% discordance. However, discrepancies were apparent between manual reviewers in 28% of variants, resulting in only 72% concordance.

Consider one example: a *BRAF* p.D594N alteration was detected in a melanoma case. While a *BRAF* p.V600E mutation in melanoma is clearly a Tier 1A alteration, QCI Interpret One computes a Tier 2 classification for p.D594N, a "type III" *BRAF* alteration that results in reduced B-Raf protein kinase activity. The two reviewing laboratories disagreed on the classification, with one laboratory reporting the variant as Tier 2 and the other reporting the variant as Tier 3. The expert panel confirmed the QCI Interpret One computed classification of Tier 2, based primarily on the availability of clinical trials which could include this mutation as a criterion.

The results reported by Fairley et al suggest strongly that the use of a clinical decision support system, particularly a system with considerable manual review and human judgment built into its components, could significantly improve the consistency and accuracy of variant interpretation.

Where does this leave us?

Variant interpretation is not a solved problem, and there is no single answer that will provide the solution. Accurate diagnoses can help to provide more reliable classifications. Guidelines are critical, but analysis is still required to evaluate the criteria. Clinical decision support systems that integrate the guidelines can apply evidence in a standardized and consistent manner, and they are most effective when they include a manual component of analysis and oversight. Integration of all these components will bring the field closer to the ideal. JoPM

Sheryl Krevsky Elkin, PhD

Sheryl is the Chief Scientific Officer for QIAGEN Digital Insights, Oncology. Dr. Elkin joined QIAGEN in 2019 as part of a merger with N-of-One, a clinical interpretation company. An early member of the N-of-One team, Dr. Elkin played a lead role in the development of the N-of-One clinical interpretation methodology, establishing a rigorous process for the analysis of scientific and clinical evidence and presentation of the evidence to physicians to help guide their therapeutic decisions. At QIAGEN Digital Insights, Dr. Elkin helped to integrate expert clinical interpretation into QIAGEN's clinical decision support software and is responsible for scientific oversight of the oncology products. Dr. Elkin completed her postdoctoral fellowship at the Massachusetts Institute of Technology's Center for Cancer Research, where she earned a fellowship from the Leukemia and Lymphoma Society. She earned her doctorate in Biological and Biomedical Sciences from Harvard Medical School and an A.B. in Biology and Music from Amherst College, graduating Phi Beta Kappa and summa cum laude, with High Distinction in Biology.

References

1. Targeted Therapy Drug List by Cancer Type – NCI. Available online: https://www.cancer.gov/about-cancer/treatment/types/targeted-therapies/approved-drug-list (accessed on 31 October 2022).
2. Arber, D.A.; Orazi, A.; Hasserjian, R.P.; Borowitz, M.J.; Calvo, K.R.; Kvasnicka, H.-M.; Wang, S.A.; Bagg, A.; Barbui, T.; Branford, S.; et al. International Consensus Classification of Myeloid Neoplasms and Acute Leukemias: integrating morphologic, clinical, and genomic data. Blood 2022, 140, 1200-1228. https://doi.org/10.1182/blood.2022015850.
3. Khoury, J.D.; Solary, E.; Abla, O.; Akkari, Y.; Alaggio, R.; Apperley, J.F.; Bejar, R.; Berti, E.; Busque, L.; Chan, J.K.C.; et al. The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Myeloid and Histiocytic/Dendritic Neoplasms. Leukemia 2022, 36, 1703-1719. https://doi.org/10.1038/s41375-022-01613-1.
4. Dohner, H.; Wei, A.H.; Appelbaum, F.R.; Craddock, C.; DiNardo, C.D.; Dombret, H.; Ebert, B.L.; Fenaux, P.; Godley, L.A.; Hasserjian, R.P.; et al. Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN. Blood 2022, 140, 1345-1377. https://doi.org/10.1182/blood.2022016867.
5. Fairley, J.A.; Deans, Z.C.; Treacy, R.J.L.; Grieg, E.; Bungartz, K.; Burton, R.; Hayes, J.; Elkin, S.K. An Independent Assessment of a Commercial Clinical Interpretation Software Indicates that Software Can Mitigate Variation in Human Assessment. J. Mol. Pathol. 2022, 3, 125-139. https://doi.org/10.3390/jmp3030012.
6. Li, M.M.; Datto, M.; Duncavage, E.J.; Kulkarni, S.; Lindeman, N.I.; Roy, S.; Tsimberidou, A.M.; Vnencak-Jones, C.L.; Wolff, D.J.; Younes, A.; et al. Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer. J. Mol. Diagn. 2017, 19, 4-23. https://doi.org/10.1016/j.jmoldx.2016.10.002.
7. Richards, S.; Aziz, N.; Bale, S.; Bick, D.; Das, S.; Gastier-Foster, J.; Grody, W.W.; Hegde, M.; Lyon, E.; Spector, E. et al. Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet. Med. 2015, 17, 405-424. https://doi.org/10.1038/gim.2015.30.
8. ClinVar. Available online: https://www.ncbi.nlm.nih.gov/clinvar/ (accessed on 14 Nov 2015).
9. Horak, P.; Griffith, M.; Danos, A.M.; Pitel, B.; Madhavan, S.; Liu, X.; Chow, C.; Williams, H.; Carmody, L.; Barrow-Liang, L.; et al. Standards for the classification of pathogenicity of somatic variants in cancer (oncogenicity): Joint recommendations of Clinical Genome Resource (ClinGen), Cancer Genomics Consortium (CGC), and Variant Interpretation for Cancer Consortium (VICC). Genet. Med. 2022, 24, 986-988. https://doi.org/10.1016/j.gim.2022.01.001.
10. Goverdhan, A.; Mullineaux, L.; Pryzbylski, A.; Teigen, C.; Elkin, S.K.; Pitel, B. A Comparative Analysis of Somatic Cancer Variant Interpretation Using the ClinGen/CGC/VICC Framework and QIAGEN Clinical Insight Decision Support Software. In Proceedings of the Association for Molecular Pathology, Phoenix, Arizona, USA, 5 November 2022.
11. Patterson, S.E.; Statz, C.M.; Yin, T.; Mockus, S.M. Utility of the JAX Clinical Knowledgebase in capture and assessment of complex genomic cancer data. NPJ Precis. Oncol. 2019, 3, 2. https://doi.org/10.1038/s41698-018-0073-y.
12. Chakravarty, D.; Gao, J.; Phillip, S.; Kundra, R.; Zhang, H.; Wang, J.; Rudolph, J.E.; Yaeger, R.; Soumerai, T.; Nissan, M.H.; et al. OncoKB: A Precision Oncology Knowledge Base. JCO Precis. Oncol. 2017, 1, 1–16. https://doi.org/10.1200/PO.17/00011.
13. Griffith, M.; Spies, N.C.; Krysiak, K.; McMichael, J.F.; Coffman, A.C.; Danos, A.M.; Ainscough, B.J.; Ramirez, C.A.; Rieke, D.T.; Kujan, L.; et al. CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. Nat. Genet. 2017, 49, 170–174. https://doi.org/10.1038/ng.3774.
14. Yaung, S.J.; Pek, A. From Information Overload to Actionable Insights: Digital Solutions for Interpreting Cancer Variants from Genomic Testing. J. Mol. Pathol. 2021, 2, 312–318. https://doi.org/10.3390/jmp2040027.
15. PierianDx—Home. Available online: https://www.pieriandx.com (accessed on 28 April 2022).