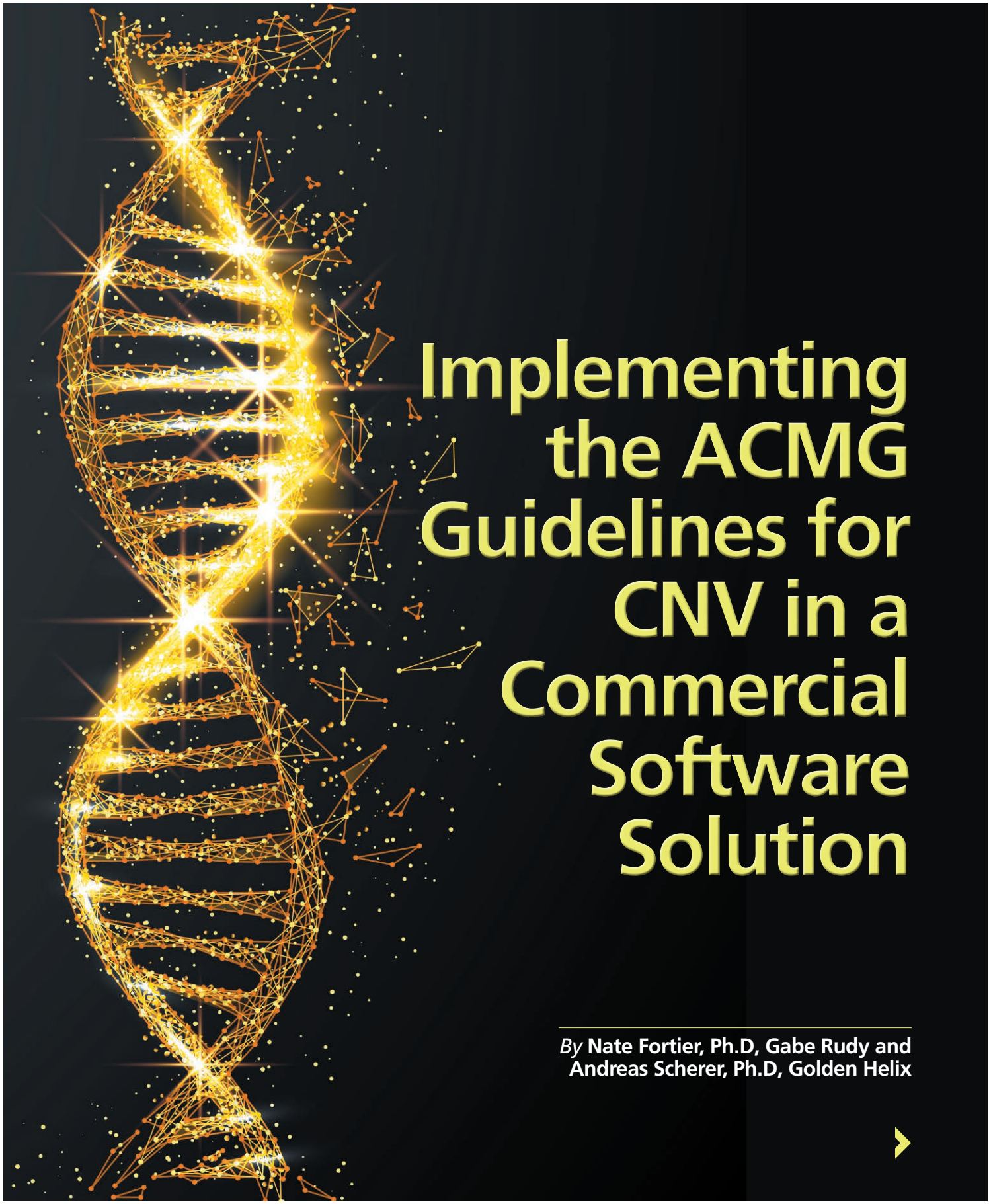

Implementing the ACMG Guidelines for CNV in a Commercial Software Solution

By **Nate Fortier, Ph.D, Gabe Rudy and Andreas Scherer, Ph.D, Golden Helix**

COPY NUMBER VARIATIONS (CNVs) are an important source of variation in the human genome and have been associated with numerous disorders. Technologies for detecting CNVs have been utilized in the clinical setting for well over a decade to identify disease-causing deletions and duplications across the genome. To assist clinical labs in the interpretation and reporting of CNVs, the American College of Medical Genetics and Genomics (ACMG) has developed a new set of professional guidelines in collaboration with the Clinical Genome Resource (ClinGen) project.

While these new guidelines provide a robust set of criteria for interpreting and classifying CNVs, the complexity involved in implementing these guidelines can be daunting for many clinical labs. To address this, Golden Helix has developed a new guided clinical workflow tool for the interpretation of CNVs based on these new guidelines. This software automates much of the classification process and guides the user through the interpretation of CNVs and related publications. By automating much of the analysis and hiding irrelevant information, this tool helps to reduce decision fatigue and ensure the repeatability of clinical workflows.

Chapter 1: Introduction

In 2020, the American College of Medical Genetics and Genomics (ACMG) developed a new set of clinical guidelines for the interpretation of CNVs in collaboration with the Clinical Genome Resource (ClinGen) (Riggs, *et al*., 2020). Golden Helix was the first commercial company to implement these guidelines and has translated the criteria into a user-friendly workflow that guides the user through every step of the CNV interpretation process.

While previous CNV interpretation guidelines were primarily focused on the interpretation of large cytogenetic events, the ACMG guidelines provide robust standards for the interpretation of small intragenic CNVs in addition to large multi-gene events. The new guidelines classify variants according to the same 5-tier classification system that is used for the interpretation of small sequence variants (Richards, *et al*., 2015) and uncouples the classification of a CNV's pathogenicity from its implications for any particular individual.

The CNV guidelines specify a list of over 80 distinct criteria, which are applied based on information about the CNV's gene impact, the literature surrounding the affected genes, and information about the current patient. This complexity is further compounded by many important caveats, exceptions, and considerations which are known to many in the CNV clinical workspace but are not explicitly mentioned in the published guidelines.

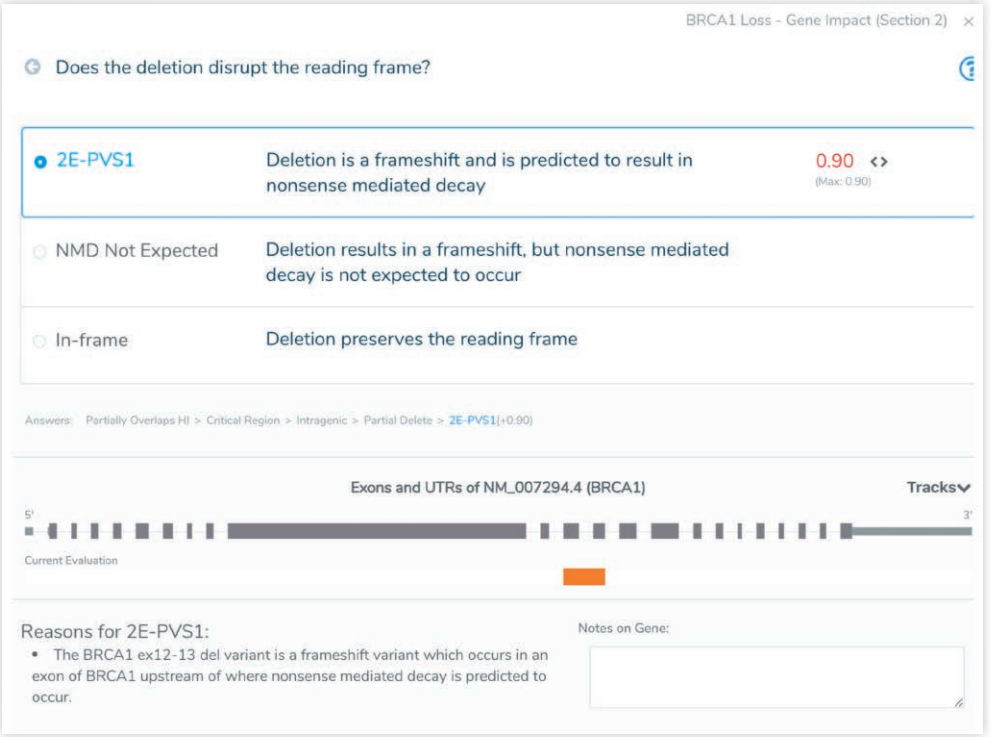

Figure 1: VSClinical Workflow Section 2E

"One of the most challenging aspects of this development process was the translation of the individual criteria into a set of guided workflows."

Over the past year, we have thoroughly explored these new guidelines along with the plethora of available supplementary material in order to develop a comprehensive workflow to streamline the process of interpreting CNVs in accordance with the guidelines. This workflow has been integrated into our existing clinical variant interpretation software VSClinical, allowing CNVs to be seamlessly interpreted alongside small variants in a single workflow tool. We have completely automated the process of scoring many of the criteria specified in the guidelines and, where such automation is not possible, we have developed tools to support and guide the user during the manual interpretation process.

One of the most challenging aspects of this development process was the translation of the individual criteria into a set of guided workflows. For each major section described in the guidelines, there are many individual criteria that may apply, but the guidelines provide no structure to sift through them. To solve this, we constructed a decision tree around a given section to guide the user to the relevant criteria. Each workflow asks

Figure 2: VSClinical Sections 1A, 3A-C, and 4O

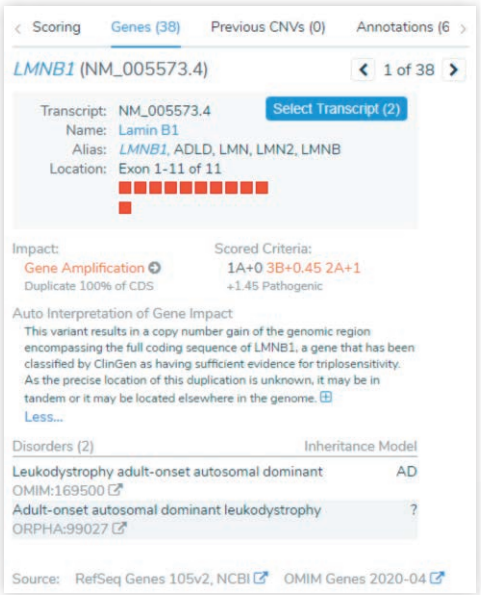

Figure 3: VSClinical Gene Details

the user a series of simple questions about the CNV or related publication under consideration. An example of the sort of questions asked is shown in **Figure 1**. This series of questions guides the user along a path through the decision tree while hiding irrelevant information. Once the user reaches the end of the decision tree, the relevant criteria is recommended to the user. By only showing the user the most relevant information at each step of the interpretation process, we can reduce decision fatigue and ensure the repeatability of a given clinical workflow.

"By only showing the user the most relevant information at each step of the interpretation process, we can reduce decision fatigue and ensure the repeatability of a given clinical workflow."

In the next chapter we provide an overview of the five major sections of the CNV guidelines. This is followed by a series of examples showcasing how the guidelines are applied in the context of several specific CNVs using our guided workflow system.

Chapter 2: ACMG Classification Criteria for CNVs

The CNV guidelines specify a list of criteria, which are applied based on information about the CNV, current patient, and related literature. Each criterion in the CNV guidelines has an associated score and the scores for all applicable criteria are added up to obtain the final score for a given CNV. The CNV is then classified using the simple thresholding system described below:

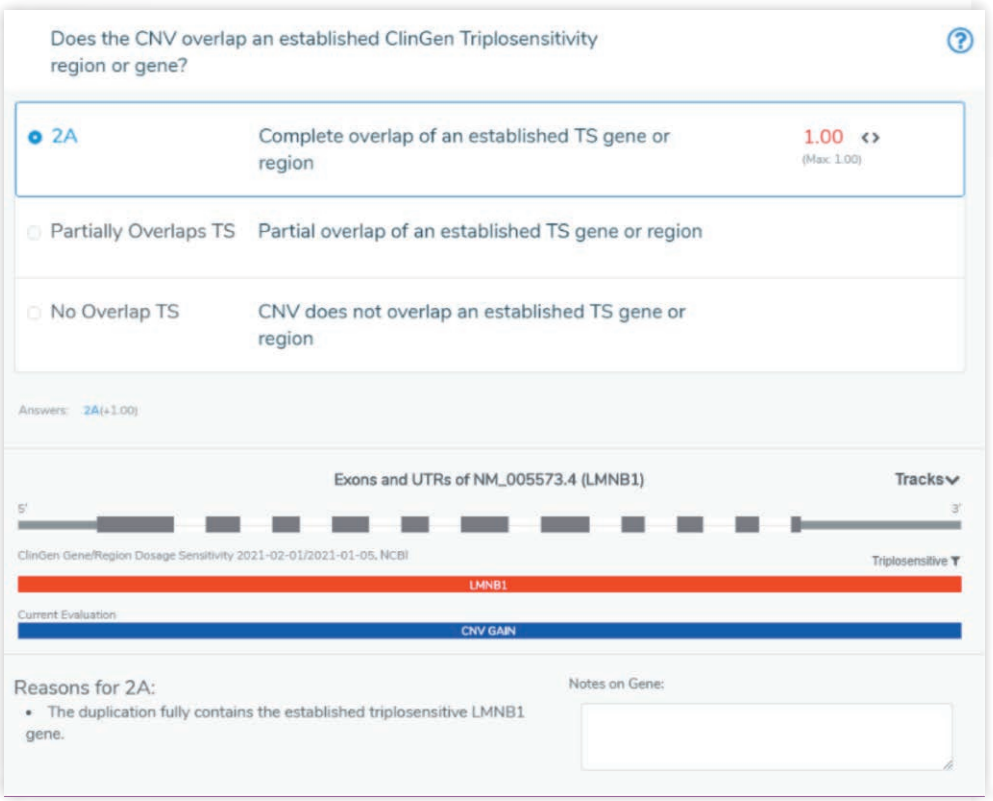

Figure 4: VSClinical Section 2A

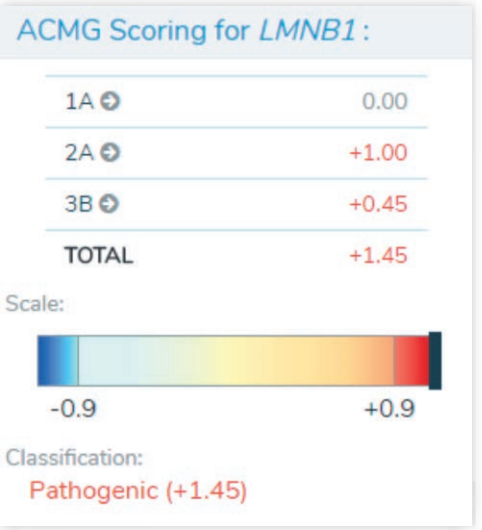

Figure 5: Pathogenic Classification

- Pathogenic: score ≥ 0.99
- Likely Pathogenic: 0.90 ≤ score < 0.99
- Uncertain Significance: -0.90 < score < 0.90
- Likely Benign: -0.99 < score ≤ 0.90
- Benign: score ≤ -0.99

The criteria used to score a CNV are divided into five sections. Sections 1-3 assess the impact of the CNV on the gene or gene product for clinically relevant overlapping genes, while sections 4-5 assess whether any overlapping genes or genomic regions are likely to be either haploinsufficient (in case of deletions and intragenic duplications) or triplosensitive (in the case of whole-gene duplications). To aid clinicians in the identification of known haploinsufficient and triplosensitive genes, ClinGen has developed the Dosage Sensitivity Map, a publicly available resource cataloging the evidence for the dosage sensitivity of genes and genomic regions.

- **Section 1: Initial assessment of genomic content:** This section of the guidelines begins by assessing whether the CNV overlaps any known protein-coding genes or other functionally important genomic material. If the CNV does not overlap any functionally important material, then it is classified as a variant of uncertain significance and the evaluation is concluded. However, if the CNV does overlap a known protein-coding gene, then further evaluation is required.
- **Section 2: Overlap with Established Dosage Sensitive Genes:** This is applied if the CNV overlaps an established haploinsufficient, triplosensitive, or benign gene/region. A gene is considered haploinsufficient or triplosensitive if it has been classified as having "sufficient

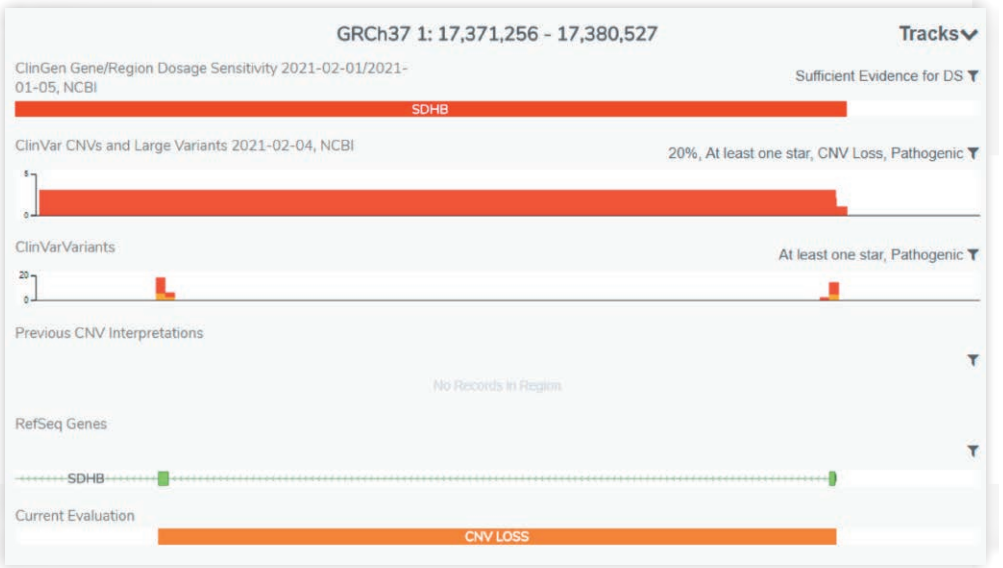

Figure 6: SDHB Gene Loss

evidence for dosage pathogenicity" either in the ClinGen Dosage Sensitivity Map or in the clinical lab's internal database. Established benign genes are those that have been shown to have a variable copy number in the general population (classified as "dosage sensitivity unlikely" by ClinGen).

While CNVs completely containing an established dosage sensitive gene or region can be immediately classified as pathogenic, CNVs partially overlapping such genes require further evaluation to determine the likely functional effect. Similar logic is applied to CNVs overlapping benign genes or genomic regions. Generally, when a CNV is completely contained by a benign gene/region, it can be safely classified as benign. The exception to this rule is when both breakpoints of an intragenic duplication are within a triplobenign gene that has not also been classified as haplobenign. Such intragenic duplications may result in an in-frame insertion or may cause a frameshift resulting in nonsense mediated decay. These duplications should therefore be evaluated in accordance with the updated PVS1 guidelines as potential loss of function variants. Additionally, when a CNV overlapping an established benign region includes additional genomic material, further evaluation is required.

- **Section 3: Evaluation of Gene Number:** While the size of a CNV is often used as a proxy for pathogenicity, there are notable exceptions. Sometimes large cytogenetically visible CNVs can be benign, while small intragenic CNVs can be pathogenic due to the disruption of important protein-coding genes.

"When a CNV overlapping an established benign region includes additional genomic material, further evaluation is required."

Thus, the ACMG guidelines only recommend a classification of likely pathogenic based on gene content alone when the CNV overlaps an extremely large number of genes (35 in the case of deletions and 50 in the case of duplications).

- **Section 4: Detailed Evaluation of Genomic Content from Published Literature, Public Databases, and Internal Lab Data:** Many CNVs overlap genes which do not have sufficient evidence to establish dosage sensitivity. When interpreting such CNVs, other data sources such as peer-reviewed medical literature and public databases must be explored to identify evidence that supports or refutes the clinical significance of the gene or region being evaluated. The criteria described in this section can be broken up into three broad categories: Individual Case Evidence, Case-Control Studies, and Population Evidence

Individual case evidence includes other reported probands from the literature, public databases, or internal lab data with CNVs or small variants similar in genomic content to the CNV being evaluated. When evaluating individual case evidence, one must consider the specificity of the observed phenotypic data. Generally, probands with highly specific well-defined phenotypes provide stronger evidence than probands with non-specific phenotypes. Inheritance information can also affect the strength of individual case evidence. For instance, a de novo occurrence in an affected individual with no family history of disease provides evidence supporting the variants clinical significance.

Segregation of a variant among affected family members can also provide evidence for clinical significance. Similarly, cases in which

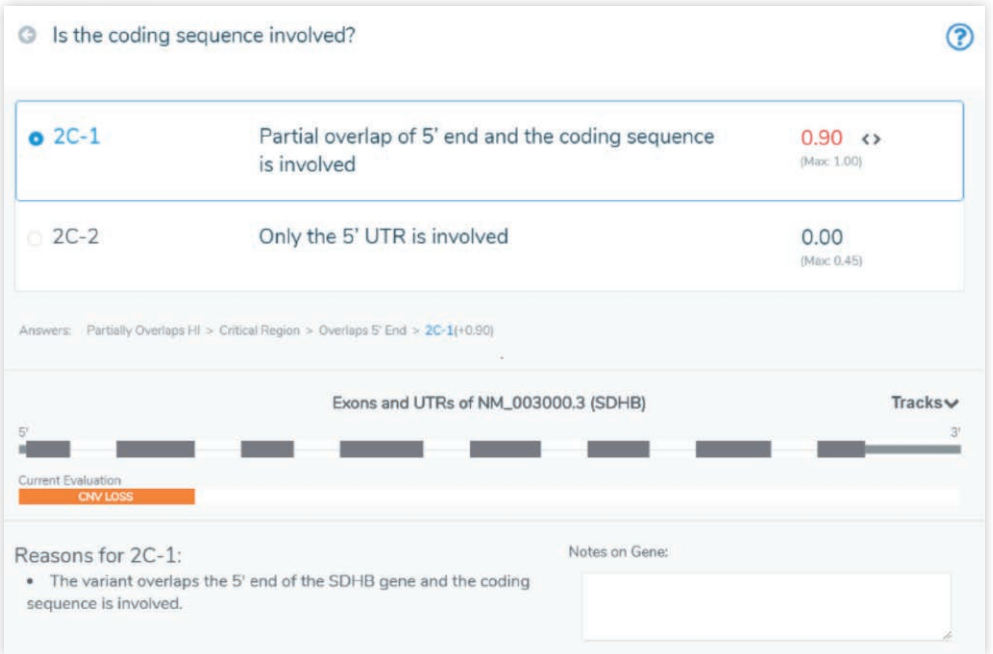

Figure 7: VSClinical Section 2C-1

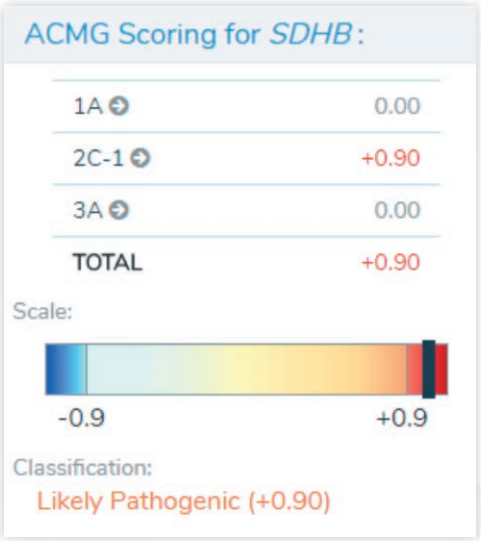

Figure 8: Likely Pathogenic Classification

a variant does not segregate among affected family members can provide evidence against pathogenicity. Additionally, variants that have been studied as part of a well-powered case-control study may provide evidence for or against pathogenicity. Finally, it may be possible to gather information about a variant's presence in the general population using population catalogs such as the Database of Genomic Variants (MacDonald, Ziman, Yuen, Feuk, & Scherer, 2014) and gnomAD (Karczewski, *et al.*, 2020).

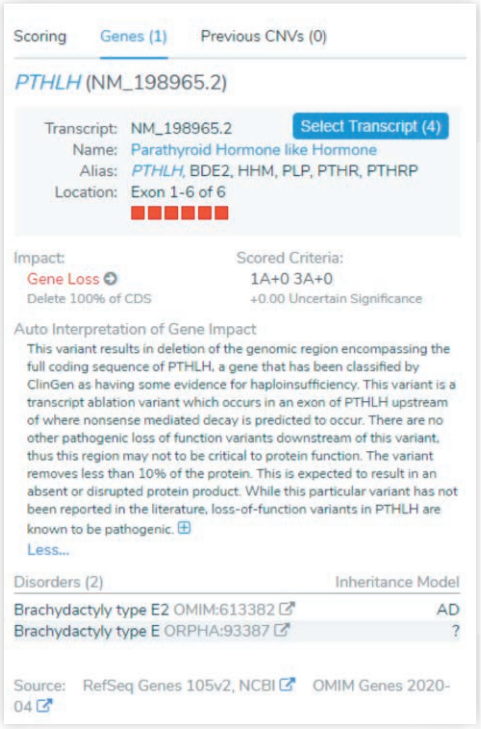

Figure 9: PTHLH Gene Details

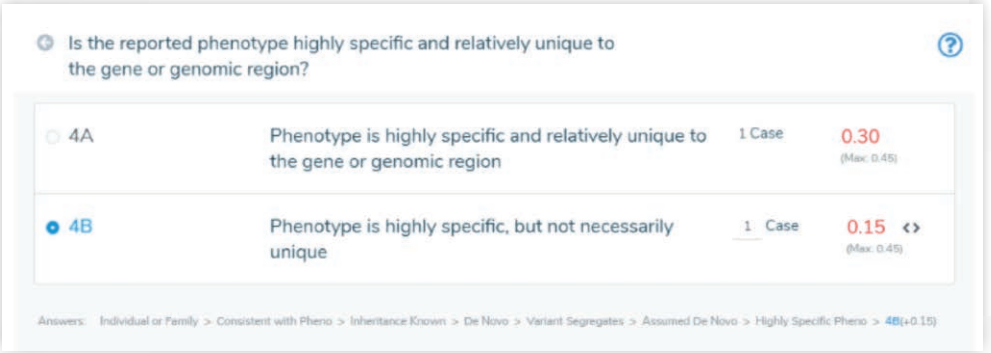

Figure 10: VSClinical Section 4B

■ **Section 5: Evaluation of Inheritance Pattern/ Family History for Patient Being Studied:** The presence of inheritance information and family history for the current patient can provide evidence for the dosage sensitivity of the gene or region under consideration. However, it should be noted that it is difficult to determine dosage sensitivity based on the inheritance pattern of a CNV in a single family. In fact, much of the criteria described in Section 5, simply require the application of the Section 4 criteria to the current proband, essentially treating the current patient as a single piece of individual case-level evidence.

It is only through the accumulation of evidence across many different families that a true measure of clinical significance can be obtained. For this reason, VSClinical stores all criteria applied in Section 5, so that it can be leveraged in future interpretations as Section 4 case-level evidence. As the clinical lab accumulates evidence across different individuals and families, a clearer picture of the gene's dosage sensitivity can emerge, allowing for the classification of variants that were previously of uncertain significance.

"As the clinical lab accumulates evidence across different individuals and families, a clearer picture of the gene's dosage sensitivity can emerge, allowing for the classification of variants that were previously of uncertain significance."

Chapter 3: CNV Interpretation with VSClinical

This chapter goes over a few example CNVs that have been classified using VSClinical's CNV Guidelines workflow. This will give you a better understanding of how to interpret CNVs in accordance with the ACMG Guidelines and demonstrate how the interpretation process is streamlined by having all relevant information embedded in a clinical workflow solution.

Copy Number Gain: dup(5)(q23.1q23.3)

Let's start by looking a large duplication spanning 14.5 million base pairs of chromosome 5 (115,714,503 – 130,229,314). We begin by looking at Sections 1 and 3 of the CNV guidelines. Looking at **Figure 2**, we can see that this duplication overlaps 38 protein-coding genes. Thus, we apply criteria 1A and 3B, assigning 0.45 points.

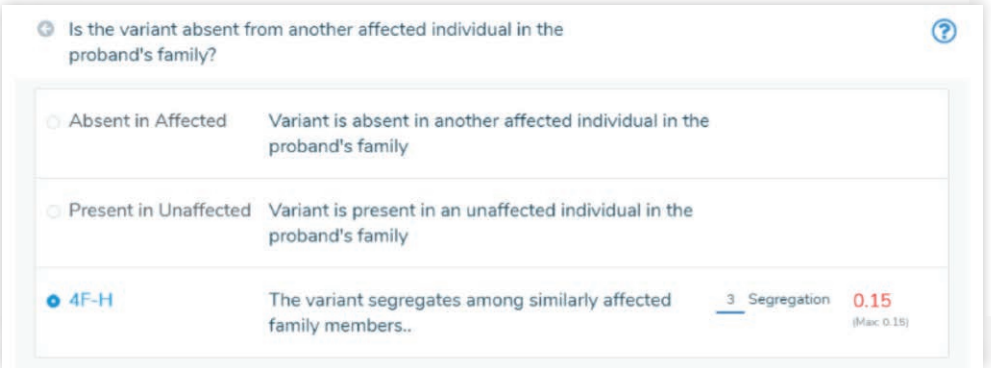

Figure 11: VSClinical Sections 4F-H

Figure 12: Section 4 Evidence Summary

Since this duplication does not meet the threshold of 50 genes required to classify a duplication as likely pathogenic using gene content alone, we must examine the dosage sensitivity of the individual genes within the duplicated region.

By examining the prioritized gene-list in VSClinical (shown in **Figure 3**), we can see that this duplication fully contains the gene LMNB1, which is associated with adult-onset leukodystrophy and has been classified as triplosensitive in the ClinGen Dosage Sensitivity Map.

For CNVs overlapping haploinsufficient or triplosensitive genes, VSClinical automatically assesses the gene impact and recommends the appropriate section 2 criteria. Looking at **Figure 4**, we can see that this duplication fully contains a known triplosensitive gene. Thus, we can apply criterion 2A and assign an additional 1.0 points to the CNV's score.

This results in a total score of 1.45 and a classification of Pathogenic, as shown in **Figure 5**.

Since this duplication fully contains an established triplosensitive gene, sections 4-5 are not applicable, and we can conclude our evaluation of this duplication.

Copy Number Loss: SDHB Exons 1-2

Next, we will look at a deletion of the first two exons of the gene SDHB. This gene has been classified as haploinsufficient by ClinGen and is associated with a number of disorders including Cowden Syndrome and various forms of gastrointestinal cancer. We begin by looking at Section 1 of the guidelines. Given that this deletion overlaps a protein-coding gene which is known to be haploinsufficient (shown in **Figure 6**), we can apply criterion 1A and proceed to Section 2 of the guidelines.

In this case criteria 2C-1 is automatically recommended by VSClinical , as this deletion overlaps the 5' end of the gene and the coding sequence is involved. This is shown in **Figure 7**.

The application of criterion 2C-1 results in a score of 0.90 and a classification of Likely Pathogenic, as shown in **Figure 8**.

Since the deletion overlaps an established haploinsufficient gene, sections 4-5 are not applicable, and we can conclude our evaluation of this deletion.

"By examining the prioritized gene-list in VSClinical, we can see that this duplication fully contains the gene LMNB1"

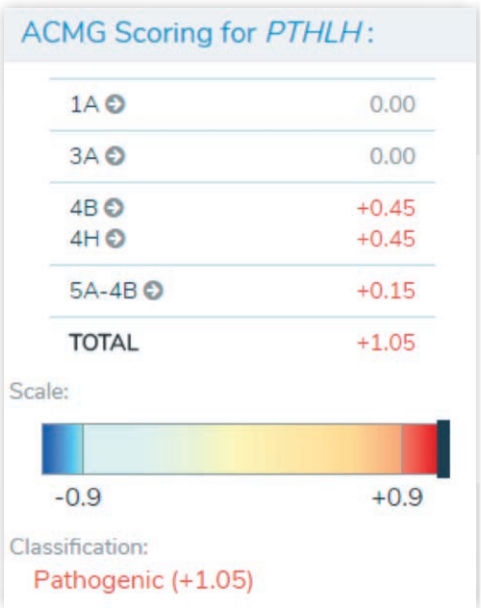

Figure 13: PTHLH Pathogenic Classification

Copy Number Loss: PTHLH Deletion

Finally, we will look at a deletion in chromosome 12 encompassing the gene PTHLH. This deletion is de novo and was identified in a patient with Brachydactyly type E (BDE). As before, we begin by applying criterion 1A, as the CNV overlaps a single protein-coding gene. Looking at **Figure 9**, we see that this gene is known to be associated with BDE and is present in the ClinGen dosage sensitivity map. However, ClinGen has not found sufficient evidence to establish haploinsufficiency.

Therefore, Section 2 cannot be applied for this gene and we must explore the literature to evaluate the gene's clinical significance using the Section 4 Criteria.

While BDE is a relatively rare disorder, is has been associated with variants in several genes, including HOXD13, PPHP, PDE3A, HDAC4, and PTHLH. Given the limited heterogeneity of BDE, this phenotype may be considered highly specific, but not necessarily unique.

De Novo Occurrences

Reviewing the literature for mutations in PTHLH associated with BDE, we find a publication by Thomas-Teinturier *et al.* where the authors describe a female proband with BDE and no reported family history (Thomas-Teinturier, *et al.*, 2016). The authors identified a heterozygous de novo deletion in PTHLH (c.101+3delAAGT). Because the authors did not confirm parental relationships, criterion 4B applies and 0.15 points are assigned, as shown in **Figure 10**.

After further review, we discover a similar variant reported by Pereda *et al.* in a female with BDE and no reported family history (Pereda, *et al.*, 2017). The author's identified a heterozygous stop gain mutation (c.166C>T). The variant was also detected in a mosaic state in the girl's unaffected father, but the ClinGen gene curation scoring guidelines recommend counting variants detected in a mosaic state as de novo. Given that parental relationships were not confirmed, we again apply criterion 4B and assign 0.15 points.

Another heterozygous stop gain mutation was identified by Jamsheer *et al.* in a female with BDE and no reported family history (Jamsheer, *et al.*, 2016). While the variant was said to be de novo, no information regarding confirmation of parental relationships was presented. As a result, we again apply criterion 4B and assign an additional 0.15 points.

Segregation Among Affected Family Members

Reyes *et al.* identified a splice region variant -3

bp upstream from exon 3 of PTHLH in a female proband with BDE (Reyes, Bravenboer, & Juppner, 2019). This mutation resulted in a heterozygous insertion causing a frameshift. This variant was also found in the proband's affected mother, maternal aunt, and monozygous twin sons. Thus, we add 3 segregations to our evaluation, as the monozygous twins are only considered a single segregation (shown in **Figure 11**).

A stop gain mutation (c.169C>T) in PTHLH was reported by Bae *et al.* in a male proband with BDE and his affected mother (Bae, Choi, Park, Lee, & Lee, 2018). Based on this finding, we add single additional segregation to our evaluation.

Another segregation was described by Thomas-Teinturier *et al.* The authors identified a heterozygous deletion, c.47_101+73del128 in PTHLH in a female with BDE and her affected mother (Thomas-Teinturier, *et al.*, 2016). Thus, we add an additional segregation to our evaluation.

Finally, Jamsheer *et al.* describe a frameshift mutation (c.258delC) in female proband with BDE (Jamsheer, *et al.*, 2016). This mutation was also present in her affected father and sister. Based on this information, we add two additional segregations to our evaluation for a total of 7 segregations.

The large number of segregations identified in the literature allow us to apply criteria 4H and assign an additional 0.45 points. VSClinical will tabulate each of these individual pieces of literature evidence in accordance with the guidelines rules and apply the correct code and point level. This concludes our Section 4 evaluation. A summary of the criteria applied based on the relevant literature is presented by VSClinical, as shown in **Figure 12**.

Evaluation of Inheritance for Current Patient

Once we have concluded our Section 4 analysis, we can move on to Section 5, where we will examine the inheritance pattern and family history for the current patient. Since the variant of interest is de novo and the patient has been diagnosed with BDE, we will apply criterion 5A, which specifies that we use the de novo scoring categories from Section 4 to determine the appropriate score. Because we have not confirmed parental relationships, we should apply criterion 5A-4B and assign an additional 0.15 points, for a final score of 1.05 and a classification of Pathogenic, as shown in **Figure 13**.

Based on the literature review described above, we can now classify the dosage sensitivity for the gene PTHLH as Haploinsufficient and save this classification for use in future evaluations. Once this haploinsufficiency classification has been saved, future CNVs occurring in the PTHLH gene can be evaluated using the criteria in Section 2, thereby streamlining future evaluations.

"Such tools are essential for streamlining the clinical interpretation process, reducing decision fatigue, and ensuring the repeatability of clinical workflows."

Chapter 4: Conclusion

In 2020, ACMG in collaboration with the ClinGen working group developed a new set of guidelines for the clinical interpretation of CNVs. While these guidelines provide a robust set of rules for interpreting intragenic deletions and duplications, the implementation of these guidelines in a clinical setting can be quite complex. To address this complexity, we have developed a comprehensive workflow to guide users through the clinical interpretation of CNVs in accordance with the new guidelines. Golden Helix is the first commercial company to implement these guidelines in a guided clinical workflow and the examples described above illustrate the way in which this tool reduces the complexity of the interpretation process. Such tools are essential for streamlining the clinical interpretation process, reducing decision fatigue, and ensuring the repeatability of clinical workflows. JoPM

Nate Fortier, Ph.D

Nate, Director of Research for Golden Helix, joined the development team in June of 2014. Nathan obtained his Bachelor's degree in Software Engineering from Montana Tech University in May 2011, received a Master's degree in Computer Science from Montana State University in May 2014, and received his Ph.D. in Computer Science from Montana State University in May 2015. Nathan works on data curation, script development, and product code. When not working, Nathan enjoys hiking and playing music.

Gabe Rudy

Gabe is Vice President of Product and Engineering of Golden Helix, and team member since 2002. Gabe thrives in the dynamic and fast-changing field of bioinformatics and genetic analysis. Leading a killer team of Computer Scientists and Statisticians in building powerful products and providing world-class support, Gabe puts his passion into enabling Golden Helix's customers to accelerate their research. When not reading or blogging, Gabe enjoys the outdoor Montana lifestyle. But most importantly, Gabe truly loves spending time with his sons, daughter, and wife.

Andreas Scherer, Ph.D

Andreas has managed global software and services businesses working for publicly traded companies such as Netscape and AOL as well as privately held companies. As part of his academic work, he has developed algorithms to conduct DNA sequence analysis. In the last decade, he has focused on accelerating R&D processes of Fortune 500 pharmaceutical, biotech, and medical device companies. As a result of this work, he is intimately familiar with the domestic and international life sciences market. Dr. Scherer holds a PhD in Computer Science from the University of Hagen, Germany, and a Master of Computer Science from the University of Dortmund, Germany. He is author and co- author of over 20 international publications and has written books on project management, the Internet, and artificial intelligence.

References

Bae, J., Choi, H. S., Park, S. Y., Lee, D.-E., & Lee, S. (2018). Novel mutation in PTHLH related to brachydactyly type E2 initially confused with unclassical pseudopseudohypoparathyroidism. Endocrinology and Metabolism, 252.

Jamsheer, A., Sowinska-Seidler, A., Olech, E. M., Socha, M., Kozlowski, K., Pyrkosz, A., . . . Latos-Bielenska, A. (2016). Variable expressivity of the phenotype in two families with brachydactyly type E, craniofacial dysmorphism, short stature and delayed bone age caused by novel heterozygous mutations in the PTHLH gene. Journal of human genetics, 457-461.

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alfoldi, J., Wang, Q., . . . Birnbaum, D. P. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nature, 434-443.

MacDonald, J. R., Ziman, R., Yuen, R. K., Feuk, L., & Scherer, S. W. (2014). The Database of Genomic Variants: a curated collection of structural variation in the human genome. Nucleic acids research, D986-D992.

Pereda, A., Garzon-Lorenzo, L., Garin, I., Cruz-Rojo, J., Sanchez del Pozo, J., & Perez de Nanclares, G. (2017). The p. R56* mutation in PTHLH causes variable brachydactyly type E. American Journal of Medical Genetics Part A, 816-819.

Reyes, M., Bravenboer, B., & Juppner, H. (2019). A Heterozygous Splice-Site Mutation in PTHLH Causes Autosomal Dominant Shortening of Metacarpals and Metatarsals. Journal of Bone and Mineral Research, 482-489.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., . . . Spector, E. (2015). 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. Genetics in Medicine, 405-423.

Riggs, E. R., Andersen, E. F., Cherry, A. M., Kantarci, S., Kearney, H., Patel, A., . . . Thorland, E. C. (2020). Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). Genetics in Medicine, 245-257.

Thomas-Teinturier, C., Pereda, A., Garin, I., Diez-Lopez, I., Linglart, A., Silve, C., & de Nanclares, G. P. (2016). Report of two novel mutations in PTHLH associated with brachydactyly type E and literature review. American Journal of Medical Genetics Part A, 734-742.