August 2021

QCI™ Translational

Release Notes

Sample to Insight

QIAGEN

Contents

Release Highlights

Improved handling of CNV and SV variants

QCI now handles copy number variants (CNV) and structural variants (SV) more comprehensively than before. QCI has been improved to better consume, process, and visualize these variant types at the level of exon and breakpoints. QCI allows you to better assess the physical structure, functional impact, and clinical relevance of copy number alterations and gene fusions detected in NGS secondary analysis pipelines. By utilizing the Test Product Profile (TPP) capability copy number variants in genes of interest are now more accurately described.

Bibliography support for Copy Number Variants

With the addition of breakpoint-level and feature-level matching for copy number variants (CNVs), you can now confidently assess bibliography content for your CNVs of interest. Explore clinically relevant CNV findings from HGMD and expertly curated content from the QIAGEN knowledgebase. New bibliography features allow to you quickly identify and capture relevant references to determine the significance of the variant of interest and to include in your clinical reports.

Additional interpretation support for CNVs

Access more resources to support the interpretation of your CNV variants with active links to public databases including GnomAD, DECIPHER, DGV and dbVar.

CNV classification calculator for hereditary workflows

In the hereditary workflow experience exon-level classification of constitutional copy-number variants using our new CNV classification feature. This is based on the technical standards published by ACMG and ClinGen*.

* Riggs et al, 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 22, 245-257 (2020). https://doi.org/10.1038/s41436-019-0686-8

Improved support for METex14 and EGFRvIII splice variants

With improved structural variant handling, two prominent exon skipping mutations: METex14 and EGFRvIII, are classified more accurately and reveal relevant treatments and clinical trials. This ultimately drives better reporting with treatment options for exon variants detected from RNA.

Integration with DNAnexus through the QIAGEN QCI Connector

In the DNAnexus ecosystem, QIAGEN has published a tool to help DNAnexus pipeline owners terminate the output of bioinformatics pipelines in QCI for tertiary interpretation.

Platform performance improvements

Behind the scenes, QCI is undergoing numerous back-end improvements to make these calculation-intensive operations more efficient.

New and Improved Features

CNV content

With the latest QCIIT release, support for CNV content has been added throughout the platform. Please see an updated list of content sources listed at the end of this document. Content matching is based on the similarity between the uploaded CNV and the CNV described in the matching content. Matching criteria are dependent on specific features of the CNV and the relevant content sources, described in more detail below.

Curated CNV content

CNV content in the QIAGEN knowledgebase fall into three broad categories: precise CNVs, imprecise CNVs and ambiguous CNVs. Precise CNVs are those that can be described at the breakpoint level based on the available data in the source content. For example, if a CNV is described using HGVS format then sufficient detail is present to convert the variant into a precise genomic range based on the genomic coordinates and the reference genome. Imprecise CNVs are those where the reference is described at the feature level and there is sufficient detail to be able to validate the variant based on the gene model and cited or default transcript. For example, a reference that describes a deletion of exon 3 in BRCA1and provides detail about the transcript can be used to calculate a minimum CNV range (in this case the genomic region that describes exon 3 in BRCA1), but the precise breakpoint positions of the CNV are not captured so are therefore considered to be imprecise. Ambiguous CNVs are those where the reference has been curated but there is insufficient detail to accurately resolve the CNV as described in the source. Ambiguous CNV content is not currently used for CNV matching.

Number of (CNV) Variant Findings

For CNVs the number of variant findings is based on the level of similarity match between the variants described in the references and the variant uploaded to QCIIT. The number displayed in the Variant Findings column refers to the number of references that contain at least one variant with an Equivalent match (ie. a similarity score >90%). If there are additional references that contains Relevant (>80% similarity) or Uncertain (Between 5 – 79% similarity) matches a blue cross icon will be displayed. Hovering over the icon will bring up details including the number of additional CNV findings present in the bibliography.

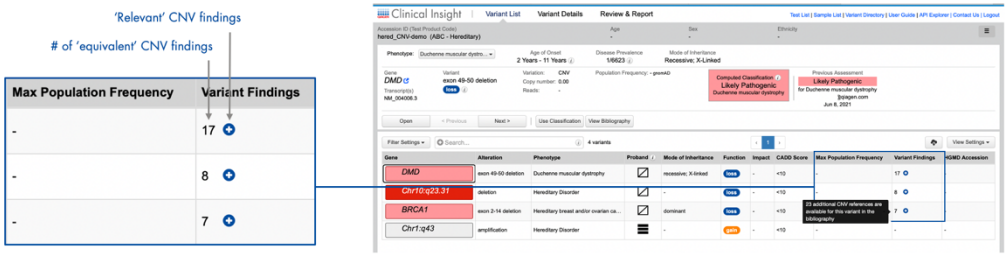


Figure 1: Variant findings for CNV content

Display of CNV references

Details of how the CNV uploaded into QCIIT is related to the CNV described in matching references present in the bibliography can be accessed by clicking the red arrow to the left of each reference. The overall CNV match classification is displayed as a colored tile between each variant described in the reference. Additional detail about the similarity score, genomic regions of the CNVs are also provided.

References

Kwong A et al. (2015) ▶ Comprehensive spectrum of BRCA1 and BRCA2 deleterious mutations in breast cancer in Asian countries. *J Med Genet* 53(1):15-23. Epub 2015 Jul 17 (PMID: 26187060)

- Mutant **human** BRCA1 gene (deletion [germline] with its exon 1 to exon 14 deleted) **increases** the relative risk of **Breast cancer in** human (HGMD: DM).
  - **CNV Match** Equivalent
    - Similarity Score: 100%
    - Test variant region: [17:41226348 - 41277499] (51151 bp) IMPRECISE
    - Reference variant region: [17:41226348 - 41277381] (51033 bp) IMPRECISE
    - Test / Reference variant overlap: 100% / 100%
- Mutant **human** BRCA1 gene (deletion [germline] with its exon 1 to exon 9 deleted) **increases** the relative risk of **Breast cancer in** human (HGMD: DM).
  - **CNV Match** Uncertain
    - Similarity Score: 58%
    - Test variant **COVERS** the reference variant
    - Test variant region: [17:41226348 - 41277499] (51151 bp) IMPRECISE
    - Reference variant region: [17:41247863 - 41277381] (29518 bp) IMPRECISE
    - Test / Reference variant overlap: 58% / 100%

Figure 2: Example reference showing an Equivalent and Uncertain CNV match between a partial gene deletion in BRCA1, and variants referenced in a published study. The first CNV match displayed has the highest significance score (100%), variant regions are identical. The second CNV match displayed is 'Uncertain' with a significance score of 58%. From the description you can see that the uploaded variant completely overlaps (Covers) the variant described in the reference. The similarity score of 58% is based on the Jaccard Index between the two variants as outlined in the final text line in the reference.

CNV bibliography

To facilitate the review of potentially relevant CNV references several new features have been added to the bibliography for this release. Firstly, for references reporting CNVs that overlap with the uploaded CNV of interest details of the type of CNV match, the similarity score and genomic ranges are displayed below the specific variant mentioned in the reference. In each reference the CNV matches are sorted in order of highest similarity. Additionally, you can sort the bibliography by the highest similarity score present for each reference. When selecting the 'Show Variant-Specific References only' the bibliography will be filtered to only the CNV matches that are equivalent (i.e., >90% overlap).

To provide further customization of your reference search, a 'CNV Criteria' section is displayed in the top panel. This section allows the bibliography to be filtered based on the type of overlap between the reference and the CNV (within, covers, overlaps 5', overlaps 3'), the breakpoint confidence of the variants in the reference (precise/imprecise), and a specified similarity score cut off (below, above, equal to).

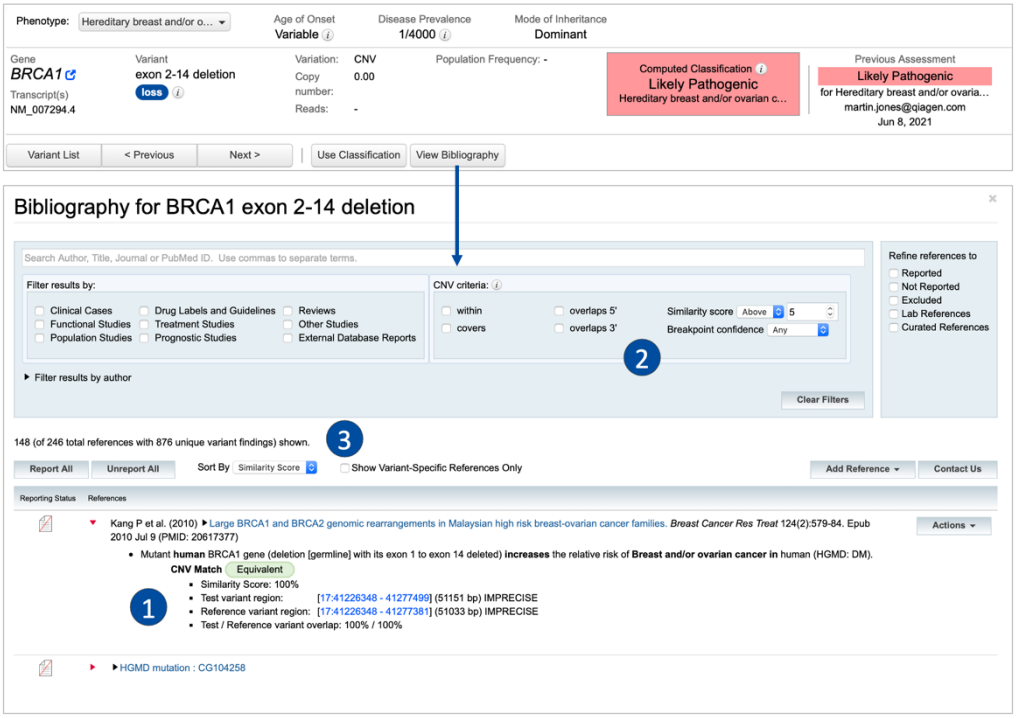


Figure 3: Bibliography features for CNV variants. (1) Matched CNV references contain information about the match criteria for each CNV variant mentioned in the reference including the similarity score, genomic region of the sample CNV, genomic region of the reference (matched) CNV, the breakpoint confidence for each (PRECISE/IMPRECISE) and the region of reciprocal overlap between the sample and reference CNVs. (2) The CNV criteria section allow further filtering of the bibliography based on the CNV matching criteria. (3) The bibliography can

be sorted based on the Similarity score (high to low) and selecting the 'Show Variant-Specific References Only' checkbox will limit reference to only those with a similarity score above 90%.

Additional data support for CNVs

External links to public CNV databases

To support interpretation of CNVs with this release, in addition to matched CNV content in the bibliography, QCIIT now includes link out the external databases that include relevant CNV data. Links are displayed in the 'Effect of Protein' section on the Variant Details page. Supported resources include GnomAD, Decipher (ref), DGV (ref) and dbVar (ref). CNV links are data-driven and will only display if the external resource reports variants that overlap the CNV of interest.

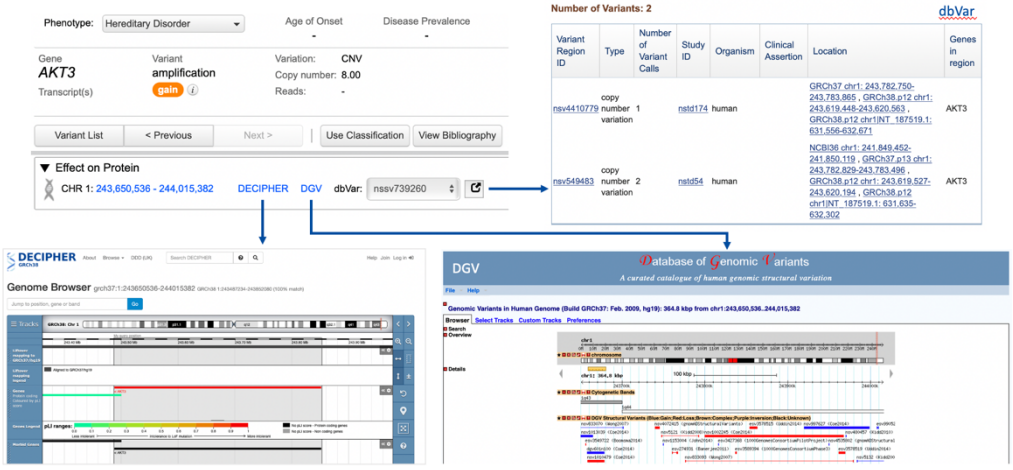


Figure 4: External links to public CNV databases.

Population Frequency for CNV

With this release population frequency data from GnomAD has been added for CNVs. When available, frequency data is displayed in the variant summary panel, the variant table, and in the Rarity in general population section on the Variant Details page (see image below). A link to GnomAD is available in the Rarity in general population section that allows you to access the GnomAD browser directly and review the available data for the CNV of interest. A similarity score of greater than 90 between the uploaded CNV and CNVs present in GnomAD is considered a match. If multiple variants

with the same similarity score are detected, then a population frequency range will be reported for the variant of interest.

Note: A link to the GnomAD browser will be displayed for all uploaded CNVs, even if no population frequency is reported. You can use this link to independently verify if any potentially relevant CNVs below the similarity score are present in the CNV region.

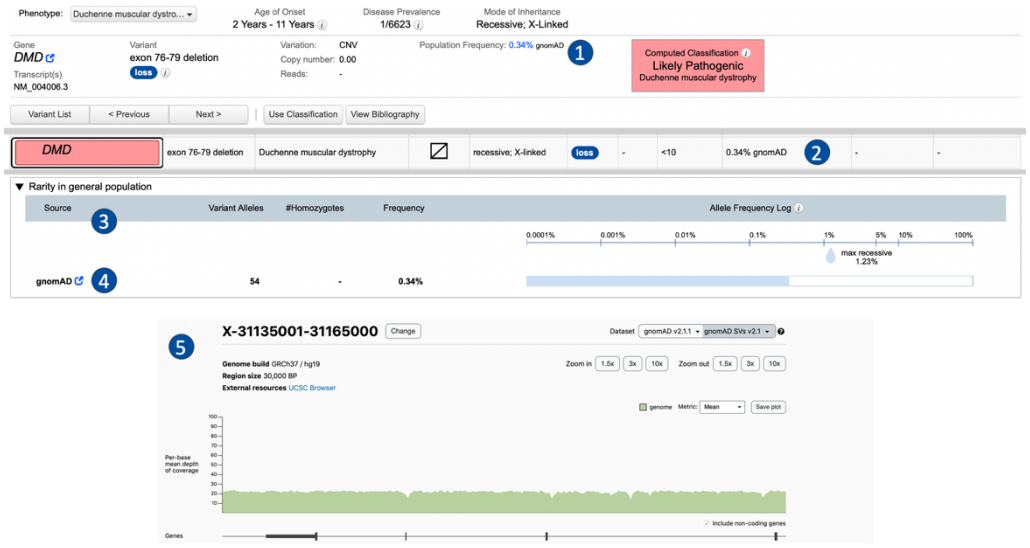


Figure 5: GnomAD frequency data for CNVs. The GnomAD population frequency for the selected CNV is displayed in the Variant summary (1), the Variant Table (2) and the Rarity in general population section (3). A link to GnomAD for the variant is available in the Rarity in general population section (4). This link will take you to the GnomAD browser view for the genomic region of the CNV (5).

CNV classification calculator

A major new feature for this release is the introduction of a variant classification system specific to CNV variants for the Hereditary workflow. This CNV classification calculator is based on the joint consensus recommendation from ACMG and ClinGen published in 2020 (Ref) and uses a points-based system to stratify CNV variants incorporating genomic location, variant structure and matched content from literature and public sources. This classification calculator is also implemented as a web app by ClinGen here: http://cnvcalc.clinicalgenome.org/cnvcalc/.

A key feature of this classification tool is its application to 'established haploinsufficient genes' (HI genes) based on the consensus review from ClinGen documented here https://dosage.clinicalgenome.org/acmg.shtml. We have built on this implementation to expand its application to both HI genes reported by ClinGen and a list of QIAGEN HI genes curated from the literature. Furthermore, to support carrier-testing and hereditary cancer workflows the CNV classifier has been expanded to apply to genes where loss of function is a known mechanism of disease (LOF genes). This follows the current logic for triggering PVS1 for CNVs in QCIIT in previous releases and maintains expected pathogenicity classification between releases. For each CNV, including multigenic CNVs, the classification rules clearly state in which context (either HI vs LOF) the rule is being triggered and for which specific gene in the CNV region.

CNV classifier design & implementation

The new CNV classifier replaces the existing computed classification that is based primarily on the ACMG PVS1 rule. Equivalent PVS1criteria are built into section 2 of the CNV classifier as detailed below. When evaluating CNV variants in the Hereditary workflow the classification will be based solely on the new classification system. With this release the new CNV classification rules cannot be manually applied or modified. However, you can manually add a rationale for each of the rules triggered if desired and adjust the strength of the final pathogenicity if additional data is available to you. You also have the option to set the reportability for each variant as is standard in QCIIT.

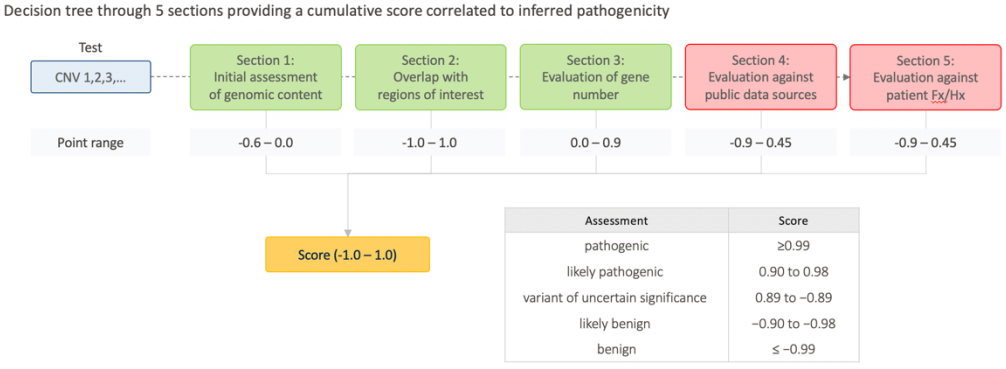


Figure 6: Schematic showing the 5 different ClinGen classification sections. The schematic demonstrates how CNVs uploaded through an analysis (blue) are classified in the ClinGen decision tree. Sections 1 to 3 are implemented for this release (green). Sections 4 and 5 are not currently available (red). The points applied to the variant are summed into a final score (yellow) that correlated to the calculated pathogenicity classification (table inset).

CNV classification rules that have triggered for a particular variant will be displayed in the assessment section of the variant details page (Number 1 in the image below) and are also summarized in the pop-up accessed by clicking on the info icon in the Computed Classification tile (Number 2 in the image below). To review how the points system is correlated to a pathogenicity classification you can click on the info icon in the Points column to reveal a descriptive pop-up (Number 3 in the image below). The table below summarizes the section 1, 2 and 3 rules that are now used in QCIIT for CNV classification.

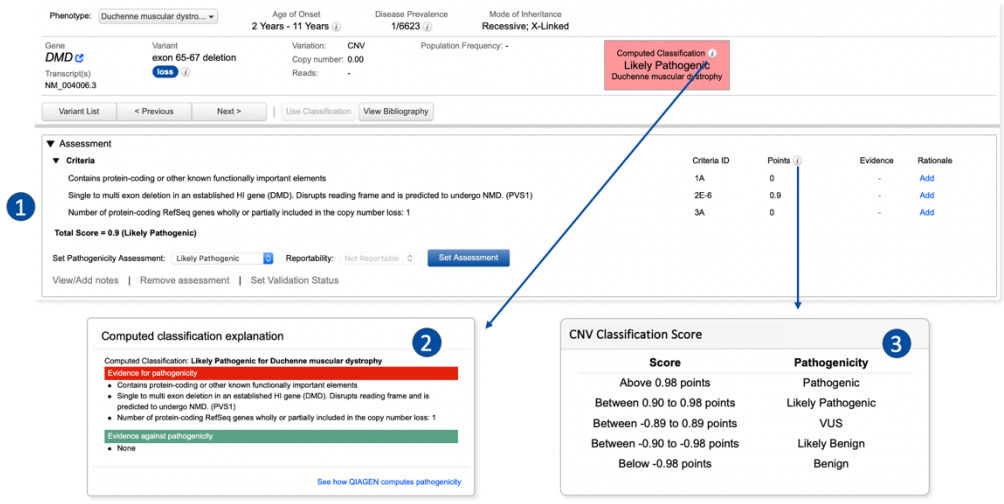


Figure 7: New UI features for the CNV classification calculator. The assessment section (1). The Computed classification explanation (2). CNV classification score (3).

Table 1: Summary of the computed classification criteria used in QCI-Interpret for CNV variants in the hereditary workflows.

Rule	Source	Variant matching	Description	Score	Additional detail
1A	ACMG ClinGen	Gain & Loss	Contains protein-coding or other known functionally important elements.	0.00	Includes 5′/3′ UTR
1B	ACMG ClinGen	Gain & Loss	Does NOT contain protein-coding or any known functionally important elements.	-0.60	Intergenic genome regions
2A	ACMG ClinGen	Loss	Complete overlap of an established HI/LOF gene.	1.00	Includes 5′/3′ UTR
2B	ACMG ClinGen	N/A	Partial overlap of an established HI/LOF genomic region.	0.00	Rule not available for this release
2C-1	ACMG ClinGen	Loss	Partial overlap with the 5′ end of an established HI/LOF gene and coding sequence is involved.	0.90	Minimum overlap is 1bp.
2C-2	ACMG ClinGen	Loss	Partial overlap with the 5′ end of an established HI/LOF gene and only the 5′ UTR is involved.	0.00	Minimum overlap is 1bp.
2D-1	ACMG ClinGen	Loss	Partial overlap with the 3′ end of an established HI/LOF gene and only the 3′ UTR is involved.	0.00	Minimum overlap is 1bp.
2D-2	ACMG ClinGen	Loss	Partial overlap with the 3′ end of an established HI/LOF gene and only the last exon is involved. Other established pathogenic[1] variants have been reported in this exon.	0.90	Minimum overlap is 1bp.
2D-3	ACMG ClinGen	Loss	Partial overlap with the 3′ end of an established HI/LOF gene and only the last exon is involved. No other established pathogenic[1] variants have been reported in this exon.	0.30	Minimum overlap is 1bp.
2D-4	ACMG ClinGen	Loss	Partial overlap with the 3′ end of an established HI/LOF gene and it includes other exons in addition to the last exon. NMD[2] is expected to occur.	0.90	Minimum overlap is 1bp.
2E-1	ACMG ClinGen	Loss	Both breakpoints are within an established HI/LOF gene. Deletes entire coding region.	0.90	Equivalent to PVS1
2E-2	ACMG ClinGen	Loss	Single to multi exon deletion in an established HI/LOF gene. Preserves the reading frame. Region overlaps a critical or well-established functional domain[3].	0.90	Equivalent to PVS1
2E-3	ACMG ClinGen	Loss	Single to multi exon deletion in an established HI/LOF gene. Preserves the reading frame. LoF variants in this region are NOT frequent in the general population[4]. Variant removes >10% of protein.	0.45	Equivalent to PVS1_strong
2E-4	ACMG ClinGen	Loss	Single to multi exon deletion in an established HI/LOF gene. Preserves the reading frame. LoF variants in this region are NOT frequent in the general population[4]. Variant removes <10% of protein.	0.00	
2E-5	ACMG ClinGen	Loss	Single to multi exon deletion in an established HI/LOF gene. Preserves the reading frame. LoF variants in this region are frequent in the general population[4].	0.00	

Rule	Source	Variant matching	Description	Score	Additional detail
Rule	Source	Variant Matching	Description	Score	Details
2E-6	ACMG ClinGen	Loss	Single to multi exon deletion in an established HI/LOF gene. Disrupts the reading frame, predicted to undergo NMD[2].	0.45	Equivalent to PVS1_strong
2E-7	ACMG ClinGen	Loss	Single to multi exon deletion in an established HI/LOF gene. Disrupts the reading frame, NOT predicted to undergo NMD[2]. Region overlaps a critical or well-established functional domain[3].		Equivalent to PVS1_moderate
2E-8	ACMG ClinGen	Loss	Single to multi exon deletion in an established HI/LOF gene. Disrupts the reading frame, NOT predicted to undergo NMD[2]. LoF variants in this region are NOT frequent in the general population[4]. Variant removes >10% of protein.	0.45	Equivalent to PVS1_strong
2E-9	ACMG ClinGen	Loss	Single to multi exon deletion in an established HI/LOF gene. Disrupts the reading frame, NOT predicted to undergo NMD[2]. LoF variants in this region are NOT frequent in the general population[4]. Variant removes <10% of protein.	0.30	Equivalent to PVS1_moderate
2E-10	ACMG ClinGen	Loss	Single to multi exon deletion in an established HI/LOF gene. Disrupts the reading frame, NOT predicted to undergo NMD[2]. LoF variants in this region are frequent in the general population[4].	0.45	Equivalent to PVS1_strong
2E-11	QIAGEN	Loss	Intronic deletion in an established HI/LOF gene. Disrupts a canonical splice site position.	0.00	+2/-2bp at the intron/exon boundary
2E-12	QIAGEN	Loss	Intronic deletion in an established HI/LOF gene. Does not disrupt any canonical splice site positions.	0.45	Equivalent to PVS1_strong
2F	ACMG ClinGen	Loss	Completely contained within an established benign CNV region[5].	-1.00	
2G	ACMG ClinGen	Loss	Overlaps an established benign CNV region[5] but includes additional genomic material.	0.00	
2H	ACMG ClinGen	N/A	Two or more HI predictors suggest that AT LEAST ONE gene in the interval is HI.	0.15	Rule not available for this release
2I	QIAGEN	Loss	Does NOT overlap an established HI/LOF gene or benign CNV.	0.00	
3A	ACMG ClinGen	Gain & Loss	Number of protein-coding RefSeq genes wholly or partially included in the copy-number loss = 0 – 34 or more genes.	0.00	
3B	ACMG ClinGen	Gain & Loss	Number of protein-coding RefSeq genes wholly or partially included in the copy-number loss = 35 – 49 or more genes.	0.45	

Rule	Source	Variant matching	Description	Score	Additional detail
3C	ACMG ClinGen	Gain & Loss	Number of protein-coding RefSeq genes wholly or partially included in the copy-number loss = 50 or more genes.	0.90	

HI = haploinsufficient, LOF = loss of function,
1. Established pathogenic variants is based on curated variant information from the QIAGEN knowledgebase.
2. NMD is inferred to be expected only if a variant occurs outside of the last exon or the last 50 bps of the penultimate exon.
3. The Definition of a Critical or well-established functional domain is based on sources of functional domain information from InterPro, PFAM, SMART, CDD.
4. LoF variant frequency in general population is calculated by the presence of putative LoF variants in the genomic region based on GnomAD v2.1.
5. Established benign CNV regions are generated based on GnomAD 2.1 SV data with a maximum population frequency >1%.

Table 2: How ACMG/ClinGen criteria scores are used to calculate Pathogenicity in the Computed Classification. Note that classification computation for the ACMG/ClinGen model is based on a scoring system. Rules that trigger different scoring criteria are mutually exclusive and as such a 'conflicting criteria' resolution scheme is not required.

CNV pathogenicity Classification	Score
Pathogenic	Above 0.98 points
Likely Pathogenic	0.90 to 0.98 points
VUS	-0.89 to 0.89 points
Likely Benign	-0.90 to -0.98 points
Benign	Below -0.98 points

CNV handling & display

Exon-level display behavior for partial-gene CNVs

To provide additional clarity into the extend of CNV variants in QCIIT, exon-level information for CNV deletion variants was introduced in a previous release. With this release this behavior has now been extended to CNV amplifications (see example for the partial amplification of exon14-15 for the FLT3 gene in the example below). An additional enhancement to this behavior, that applies to all CNV variants, is that when a CNV impacts the entire coding region of a gene the variant will not display the impacted exon information to make it clear that a whole gene loss or gain is being described by the variant. In the example below for TP53 all coding exons 2-11 have been deleted and the variant is now described simply as 'TP53 deletion'.

It is important to note that many analysis pipelines for panel data do not call CNVs on the entire region of the genes of interest. In these cases, the user should be aware when a partial gene amplification or loss is intended to convey the overall copy number state of the gene (See example for ERBB2 in the figure below). For this reason, content matching for partial gene gains/losses will be done at the gene level (e.g., a partial amplification for ERBB2 will match to ERBB2 amplification content).

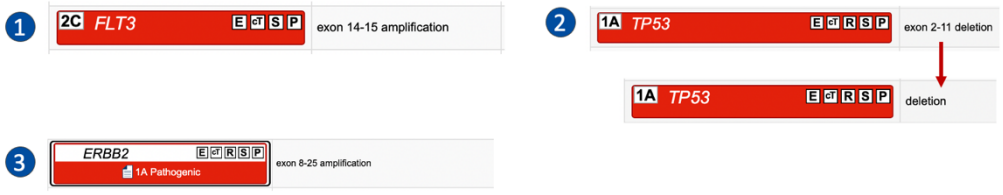


Figure 8: Exon display behavior for CNV variants. 1. A possible ITD of exons 14-15 for FLT3 is displayed as a partial gene amplification. 2. TP53 deletion of all coding exons (exon 2-11) now displayed as a whole gene deletion. 3. Example of panel data displaying the exon 8-25 genomic region to infer gene-level CNV state for the ERBB2 gene.

Note: Impacted exon information can still be found in the Variant Details section

Enhanced Variant details section

For this release the variant details section on the Variant Details page has been enhanced to provide more detailed information for CNV variants. Specifically, to facilitate faster review of the region impacted by the CNV both breakpoints are now displayed as a range. Clicking on the IGV link will show the impacted range in the genome browser. Additionally, information about the impact of each

breakpoint at the cytoband, gene region and exon level is now displayed where applicable (see example for the DMD gene in the image below).

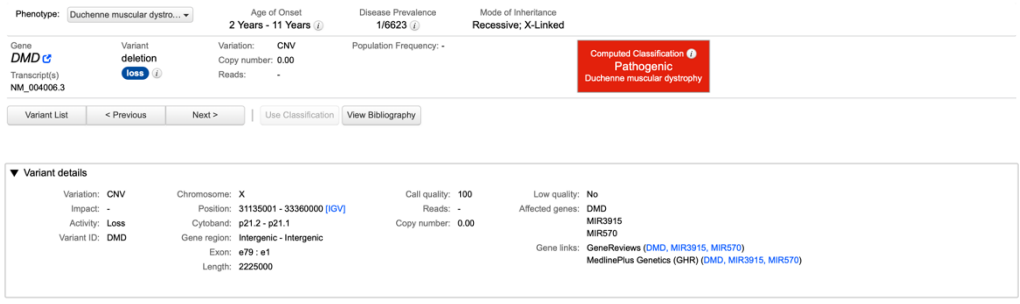


Figure 9: Updated variant details display for a deletion of the DMD gene.

Handling of multiple CNV variants in the same gene in a single test/analysis.

Prior to this release multiple CNV variants in the same gene were merged to accommodate CNV calling pipelines based on a tiling/binning approach (for example, CNV data from microarray). With this release the software behavior has been enhanced to support CNV calling pipelines where merging overlapping or adjacent CNVs is not appropriate (for example Illumina TSO500 splice variant calling data). The new behavior is described in the image below. Briefly, multiple-exon CNVs will be assumed to be independent events and will not be combined, while single-exon events will be assumed to be dependent and will be combined ONLY if they are adjacent in the same transcript and have the same CN value.

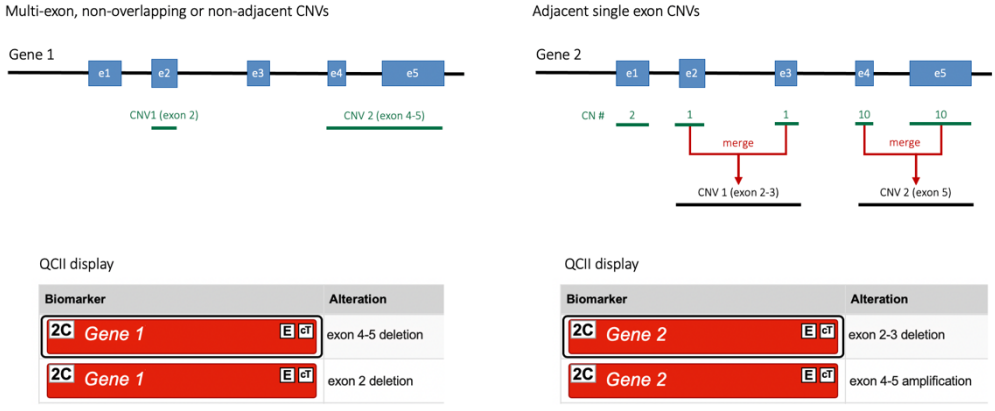


Figure 10: Schematic showing CNV merging behavior. (Left) CNV merging is not done when CNVs are multi-exonic or are non-adjacent single exons. (Right) Adjacent single exons that have the same CN classification will be merged into a single non-overlapping variant.

Note: edge cases may exist for complex cases where adjacent single-exon splice variants are combined

Nomenclature display choice for multi-genic CNVs (gene vs gene-region)

In QCIIT, CNV variants that span genomic regions with overlapping genes are displayed in the Chr:Cytoband format to indicate that multiple genes are impacted by the variant. This behavior is not ideal for some applications where there is only a single gene of interest in the specified CNV (for example gene panels with a discrete number of relevant genes). To account for this QCIIT now uses the 'Genes Tested' list from the Test Product Profile (TPP) to determine which gene of interest should be displayed in a given CNV (see the example for ERBB2 in the image below). If you need to confirm if your current TPP includes the appropriate gene list for your panel, or you wish to create a TPP with a new gene list please contact your QIAGEN account manager.

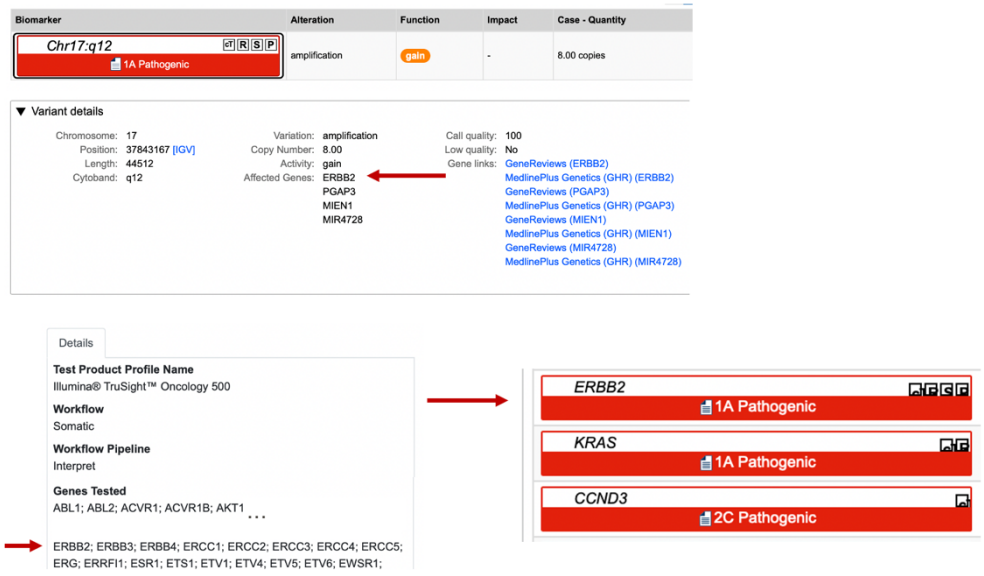


Figure 11: Example of a gene of interest being displayed based on the 'Genes Tested' list in a TPP. Display in the genomic region format for the ERBB2 amplification without an associated list of tested genes is shown in the top left. Using a TPP with a gene list that includes ERBB2 (bottom left) results in the CNV now being recognized as an amplification of a gene of interest and the expected gene name is displayed (bottom right).

Note: If overlapping genes are present in the TPP the display will default back to the existing region display.

Gene fusion handling & display

To support more detailed assessment of structural variants uploaded from VCF files, QCIIT now uses breakpoint and transcript information to calculate and display exon-level information for gene fusions. This information can be useful for determining if a detected fusion conforms to a known functionally relevant structure. The variant details section also now contains additional information including showing both breakpoints for the uploaded structural variants, the genomic regions impacted by the breakpoint, and information about the 5′ and 3′ impacted exons for breakpoints falling within genes.

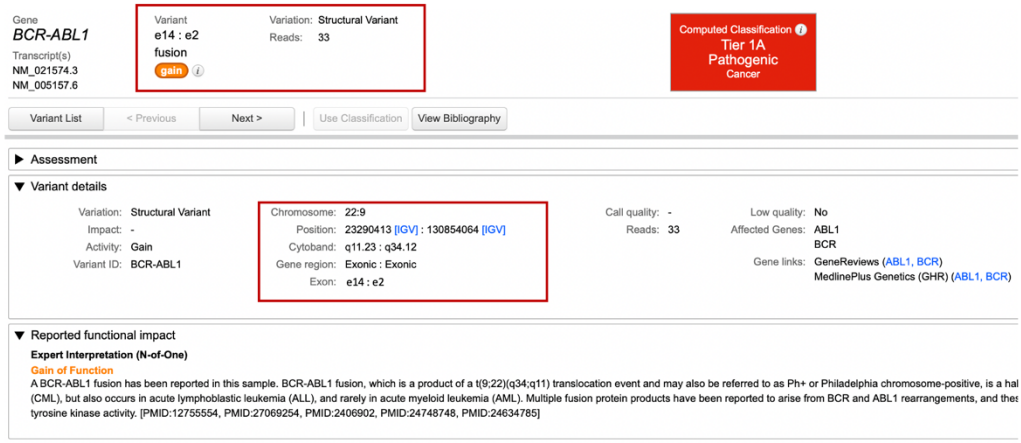


Figure 12: Screenshot showing new variant details for gene fusions. In the example above breakpoint and transcript information is used to determine the exon-level information for the BCR-ABL1. The variant details section now contains information about both fusion breakpoints as well as the gene regions impacted.

Note: For QCII-One customers breakpoint information for gene fusions will also be sent to N-of-One and details will be included as part of the expert interpretation review.

Note: Exon-level information is not used to match gene fusion content. Content matching is based only on gene ID and the 5′– 3′ orientation of the gene pair.

Note: With the inclusion of more detailed breakpoint information QCIIT now supports the display of multiple gene fusions detected with different breakpoints in a single analysis.

Splice variant handling & display for MET and EGFR

In this release we have added support for clinically relevant splice variants in MET and EGFR detected from RNA pipelines such as TSO500 and OncoMine DX. Specifically, when uploaded into QCIIT in VCF format encoded either as rearrangements or deletion variants, METex14 skipping variants and EGFRvIII splice variants now match appropriate treatment and trial content. Schematics of the supported structure of each variant and example display in the QCIIT IU is shown in the images below.

Figure 13: Support for METex14 splice variants detected from RNA. MET exon 14 skipping variants matching the canonical breakpoint structure and encoded in rearrangement (SVTYPE=BND) or deletion (SVTYPE=DEL) format, will match to MET exon14 mutation positive treatment and trial content. The appropriate expert interpretation for functional impact will also be displayed.

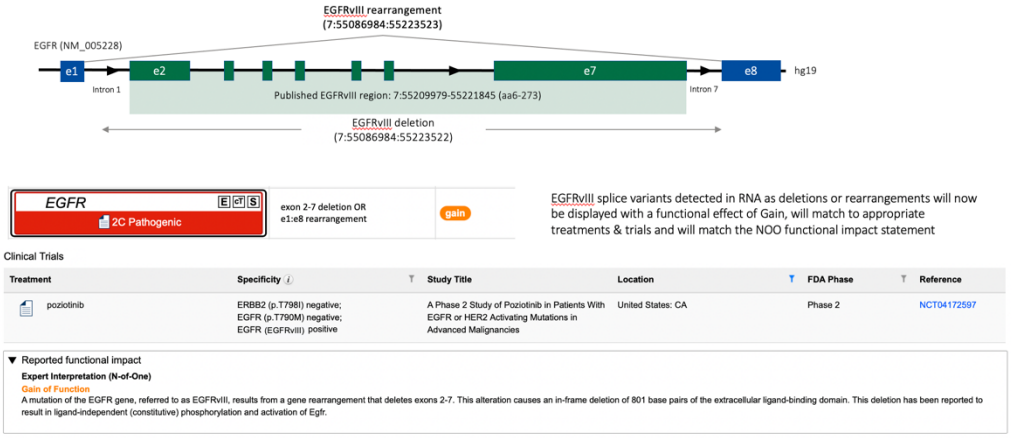


Figure 14: Support for the EGFRvIII splice variant detected from RNA. EGFR splice variants matching the canonical EGFRvIII breakpoint structure and encoded in rearrangement (SVTYPE=BND) or deletion (SVTYPE=DEL) format, will match to EGFRvIII mutation positive content. The appropriate expert interpretation for functional impact will also be displayed.

New structural variant support for inversions and insertions encoded in 'SVTYPE' VCF format

Inversion and insertion variants encoded in VCF as structural variants using the SVTYPE=INV and SVTYPE=INS formats are now recognized and displayed in QCIIT. Examples of how variant will be displayed in the UI are shown below.

Figure 15: Example display in QCIIT of an inversion variant encoded using the SVTYPE=INV VCF format. In the example above an inversion impacted exon 8 of EPCAM to exon 3 of MSH2 encoded using the SVTYPE=INV format is displayed in the QCIIT UI (top). The structure of the variant format is similar to a CNV encoded as SVTYPE=DEL, though no CNV value is encoded (bottom).

Figure 16: Example display in QCIIT of an insertion variant encoded using the SVTYPE=INS VCF format. In the example above an 63bp is inserted at exon 14 of FLT3 (top). Note that even though this insertion might represent

an internal tandem duplication (ITD) the SVTYPE=INS provides no information of the origin of the inserted region unless specifically captured in the ALT field as <INS:Tandem> (bottom).

Note: Variants will be displayed in the QCIIT UI and can be manually assessed and reported out, but no automatic content matching for these variant types will be available for this release.

Locus-region variant support

There are a number of genomic loci that are involved in clinically relevant variants (the immunoglobulin heavy chain locus (IGH) in follicular B-cell lymphomas for example). To better support the ability to display, match and filter on variants that impact these loci QCIIT now recognizes the genomic regions these loci represent as being distinct. The table below provides a list of the loci that are now supported, the genomic regions that are specified for each loci and the corresponding transcripts ID's on which these loci are based. The figure below shows how a BCL2-IGH variant is displayed when uploaded via VCF.

Table 3: Locus regions supported in QCIIT. The genomic regions below define the Loci supported in QCIIT (hg19 coordinates). Variants impacting these regions will be treated in a manner similar to coding genes with respect to UI display, variant filtering and content matching.

Chr	Start Position	End Position	Name	Transcript ID
14	106052774	107288051	IGH	NC_000014.8:107288051_106052774
2	89156874	90274235	IGK	NC_000002.11:89156874_90274235
22	22380474	23265085	IGL	NC_000022.10:22380474_23265085
14	22090057	23021075	TRA	NC_000014.8:22090057_23021075
7	141998851	142510972	TRB	NC_000007.13:141998851_142510972

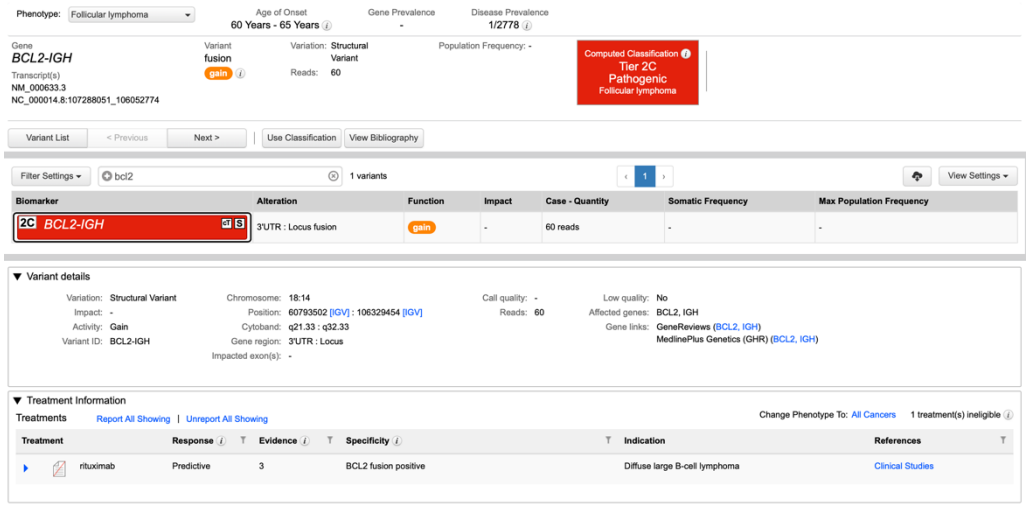


Figure 17: BCL2 gene fusion involving the IGH locus as displayed in QCIIT. The example above shows how a BCL2-IGH gene fusion is displayed in QCIIT. Variant structure describing the fusion of BCL2 at the 3'UTR to the IGH locus is noted in the variant details section. Variant matching to BCL2 fusion positive treatment information is also shown.

ACMG Pathogenicity Improvements

A number of minor improvements have been made to the logic that drives the ACMG computed classification:

- We have improved the implementation for triggering PVS7, PS7 and PM7 such that only primary reference sources are used to derive the independent somatic observation counts. This prevents multiple counting of findings.
- We have made a change to prevent somatic-specific QIAGEN rules PA1 and PA3 from triggering for variants in hereditary workflows.
- We have additionally updated logic to strengthen the somatic-specific QIAGEN PA1 and PA3 rules in the presence of conflicting information as follows: Pathogenic + Benign = VUS (EXCEPT in cases where PA1, PA3 criteria has been invoked).

Integration with DNAnexus through the QIAGEN QCI Connector

Users of the DNAnexus secure cloud infrastructure/bioinformatics ecosystem are now able to integrate upstream bioinformatics pipelines with QCI for interpretation and analysis. The QIAGEN QCI

Connector App is available in the public Tools Library for any DNAnexus account holder. Terminating an upstream bioinformatics pipeline's vcf or vcf.gz output on this app allows a convenient method to push the sample into QCI for interpretation. This application supports both QCII and QCIIT account holders in the US and UK/Europe.

For complete instructions that walk you through the 10-minute integration exercise, log into DNAnexus with a valid account and navigate to this page:

https://platform.dnanexus.com/app/qiagen_qci_connector_app

Minor Improvements & Bug Fixes

- A minor update has been made to the QIAGEN default 'Illumina® TruSight™ Oncology 500' Test Product Profile to add a previously unmapped gene.
- Functional impact inference and content matching for predicted METex14 skipping variants has been extended to include the entire poly-pyrimidine tract at the exon 14 canonical acceptor splice-site (HG19 Chr 7:116411884-116411905) and the +10 position from the exon 14 canonical donor splice-site (HG19 7:116412041 -116412054), see figure below.

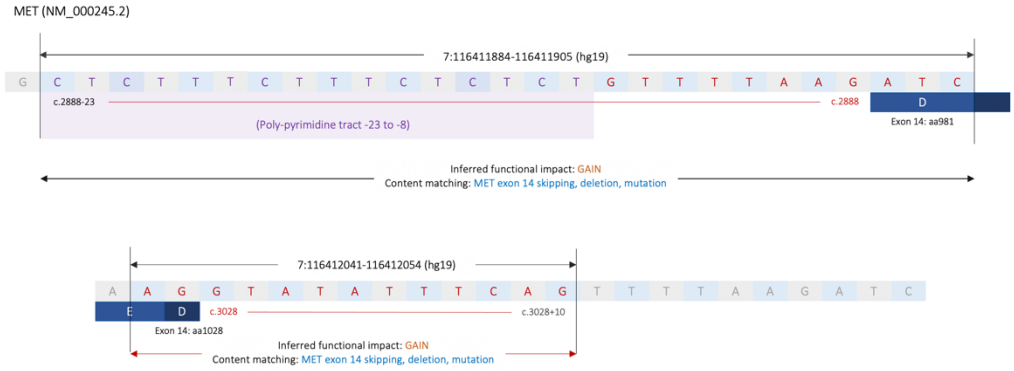


Figure 18: Genomic regions used to determine mutations that may lead to METex14 skipping. Image above shows a schematic representation of the genomic regions around the 5' and 3' donor and acceptor splice sites for MET exon 14. Non-synonymous variants with no existing literature curation falling into the defined regions will be flagged as having a predicted gain of function, will be assessed as likely pathogenic, and will match to MET exon 14 curated literature, including treatments and clinical trials.

- QCIIT now supports gene fusions described using a colon delimiter. This resolves the issue where canonical gene IDs that contain a hyphen (eg. HLA-A) were not properly recognized on import. Colon delimiters are supported for both manual and XML/API import. See the example for the gene fusions HLA-A:NCOA4 below.

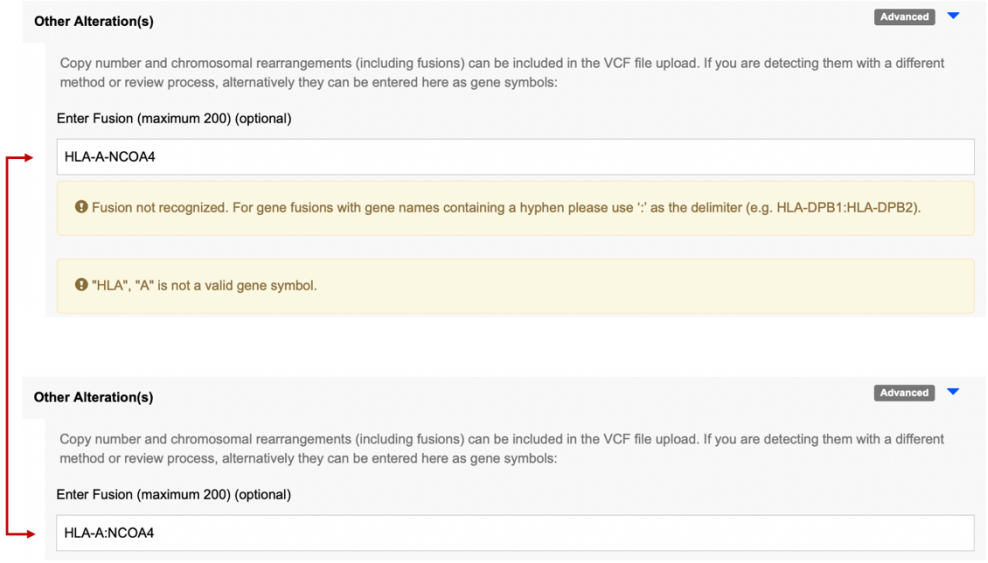
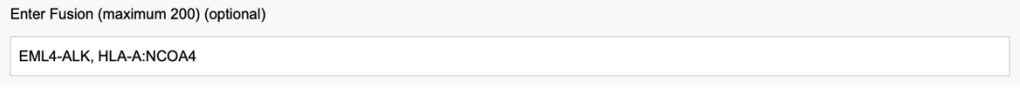


Figure 19: Support for colon delimiter usage for gene fusion import into QCIIT. Top, a gene fusion containing a hyphenated gene name and using a hyphen as a delimiter is not correctly recognized during variant upload. Bottom, replacing the delimiter to use a colon resolves the issue.

Note: QCIIT supports a mix of hyphen and colon delimiters in a list of gene fusions as long as they are correctly separated by a comma (see below). This can be useful if you only need to manually edit the delimiter for a small number of fusion in a long list.



Note: Export of gene fusions from QCIIT via API will continue to use a hyphen as the delimiter i.e. in the example above the input would be EML4-ALK, HLA-A:NCOA4 but output will still be EML4-ALK, HLA-A-NCOA4.

Updated Content Versions

Within the application QCI offers a hyperlink that will present the user with a report of the specific version of every data source that has been integrated into a knowledge base that drives the application. As of this release, the content sources include:

Data Source	Version
CADD	v1.6
NCBI Gene	2021-02-19
Allele Frequency Community	2019-09-25
EVS	ESP6500SI-V2
Refseq Gene Model	2021-02-19
JASPAR	2013-11
Ingenuity Knowledge Base Snapshot Timestamp	2021-07-23 16:14:57.512
Vista Enhancer	2012-07
Clinical Trials	D-release
MITOMAP: A Human Mitochondrial Genome Database. http://www.mitomap.org, 2019	2020-06-19
PolyPhen-2	v2.2.2 (HumVar)
1000 Genome Frequency	phase3v5b
ExAC	0.3.1
TargetScan	7.2
phyloP	NCBI36 (hg18) 2009-11, GRCh37 (hg19) 2014-02, GRCh38 2015-05
GENCODE	Release 37
CentoMD	5.3
Ingenuity Knowledge Base	D-release
dbVar	2021_04
OMIM	May 07, 2021
gnomAD	2.1.1
BSIFT	2016-02-23
TCGA	2013-09-05
Clinvar	2021-04-26
DGV	2016-05-15
COSMIC	v92
HGMD	2021.2
OncoTree	oncotree_2019_03_01
dbSNP	NCBI36 (hg18) 151, GRCh37 (hg19) 154, GRCh38 154
SIFT4G	2016-02-23