# Oncogenicity prediction with Illumina Connected Insights

## Highly accurate, automated variant interpretation using the latest advances in oncology guidelines and proprietary Illumina AI algorithms for clinical research

### Accurate oncogenicity prediction

Achieves greater variant classification accuracy than other classifiers and manual workflows

### Trusted variant interpretation

Maintains high concordance with established clinical tiering and expert-assigned oncogenicity criteria

### Best-in-class *in silico* predictors

Identifies disease-causing variants using breakthrough proprietary algorithms

# Introduction

Interpretation of somatic variants from next-generation sequencing (NGS) methods like comprehensive genomic profiling (CGP) and somatic oncology research studies can be challenging. Tiering guidelines issued by the Association of Molecular Pathology (AMP), the American Society of Clinical Oncology (ASCO), and the College of American Pathologists (CAP), as well as oncogenicity standards established by the Clinical Genome Resource (ClinGen), the Cancer Genomics Consortium (CGC), and the Variant Interpretation for Cancer Consortium (VICC) provide a framework to classify somatic variants by clinical actionability and oncogenicity (their capacity to contribute to tumor formation), respectively. However, current manual interpretation processes are error-prone, time-consuming, and often inconsistent in their results.[1] This challenge creates a need for research labs to employ automated solutions that can apply these standards more efficiently and consistently.

Illumina Connected Insights automatically calculates oncogenicity prediction and leverages PrimateAI-3D and SpliceAI as part of its scoring framework. These breakthrough proprietary algorithms empower researchers to achieve greater variant classification accuracy and confidence in their research insights. PrimateAI-3D and SpliceAI enable even further variant classification accuracy by identifying disease-causing, pathogenic missense variants with superior accuracy as compared to all other classifiers tested.[2,3]

To demonstrate the accuracy of Connected Insights somatic oncology variant interpretaton, it was evaluated for its oncogenicity prediction capability against expert-assigned classifications and established truth sets.

# Methods

The concordance of automated oncogenicity classification in Connected Insights was evaluated against an expert-curated truth set of 94 variants validated under ClinGen/CGC/VICC guidelines. High concordance indicates alignment with expert judgement. Classification accuracy within Connected Insights was also assessed by testing the precision of PrimateAI-3D pathogenicity prediction relative to existing algorithms. Pathogenicity prediction accuracy of PrimateAI-3D was measured for 40 missense variants, benchmarking against 19 other *in silico* predictors (eg, SIFT, PolyPhen-2). Concordance was calculated as the percentage of exact matches to reference classifications. Strong agreement across these evaluations demonstrates the ability of Connected Insights to emulate manual interpretation with high accuracy.

# Results

## Automated variant interpretation with Connected Insights is concordant with expert classifications for oncogenicity

In the guideline validation set of 94 tumor variants, Connected Insights reproduced 96% (90/94) of the expert-assigned oncogenicity labels. The software identified the majority of the same variant labels: oncogenic, likely oncogenic, variant of unknown significance, likely benign, or benign (Table 1). Notably, Connected Insights was able to incorporate up-to-date evidence for certain variants that had evolved since the truth set's publication. In a few cases, Connected Insights found new supporting data from its knowledge base feeds that led to reclassification of variants as oncogenic before post-publication reclassification by external sources (Table 2). This approach highlights how automation in oncology research can match and potentially enhance interpretations by rapidly integrating new evidence. Overall, the high concordance indicates that Connected Insights can achieve expert-level accuracy in applying ClinGen/CGC/VICC oncogenicity criteria.

## Highly accurate prediction of variant oncogenicity with PrimateAI-3D

PrimateAI-3D achieved 100% accuracy for oncogenicity prediction against 40 missense variants, outperforming 19 other *in silico* prediction tools that were evaluated and used as benchmarks (Table 3). In fact, PrimateAI-3D was unique in identifying every known pathogenic missense variant in the benchmark while correctly assigning every benign variant as such (Table 3). These results demonstrate Illumina as a leader in variant pathogenicity prediction, with AI algorithms embedded in Connected Insights that demonstrate higher concordance than other widely-used methods. Of note, the use of primate evolutionary data and 3D protein context by PrimateAI-3D may underlie its superior performance, an advantage that translated into measurably better predictions than traditional algorithms on this data set. The integration of PrimateAI-3D and SpliceAI into Connected Insights ensures that researchers benefit from this top-tier accuracy automatically, improving confidence in variant interpretations.

Table 1: Highly concordant variant interpretation with Connected Insights

| | | Connected Insights | | |
|---|---|---|---|---|
| | | O + LO | VUS | LB + B |
| ClinGen | O + LO | 43 | 3 | 0 |
| | VUS | 1 | 32 | 0 |
| | LB + B | 0 | 0 | 15 |

O, oncogenic; LO, likely oncogenic; VUS, variant of unknown significance; LB, likely benign; B, benign. Concordance of variant interpretation is represented by color, where green indicates variant classifications were in alignment between ClinGen and Connected Insights, white indicates no variants were given conflicting classifications between ClinGen and Connected Insights, and blue indicates variant classifications were not in alignment between ClinGen and Connected Insights.

Table 2: Reclassification of variants as oncogenic with Connected Insights

| Variant | Description | Link |
|---|---|---|
| *PIK3CA* N1068Kfs*5 | Results in increased cell proliferation in the absence of growth factors | OncoKB |
| | Results in increased cell proliferation, migration, and transformation ability | CKB |
| *RB1* S567L | Results in decreased binding to E2F, failure to induce G1 arrest, and failure to promote differentiation | OncoKB |
| *TP53* G334R | Results in impaired suppression ability in a colony reduction assay, impaired activation of a subset of *TP53* target genes, and reduced thermal stability | OncoKB |

Variants included in this table were reclassified by Connected Insights as oncogenic before expert knowledge bases. Variants were classified as oncogenic according to oncogenic strong-2 (OS2) guidelines, with well-established *in vitro* or *in vivo* functional studies supportive of an oncogenic effect of the variant.

Table 3: Highly accurate oncogenicity prediction with PrimateAI-3D

| | | PrimateAI-3D | |
|---|---|---|---|
| | | O + LO | LB + B |
| ClinGen | O + LO | 33 | 0 |
| | LB + B | 0 | 7 |

O, oncogenic; LO, likely oncogenic; LB, likely benign; B, benign. Concordance of variant interpretation is represented by color, where green indicates variant classifications were in alignment between ClinGen and Connected Insights and white indicates no variants were given conflicting classifications between ClinGen and Connected Insights.

## Summary

Collectively, these results demonstrate that Illumina Connected Insights delivers high analytical accuracy in automated variant interpretation. The tool reproduces classifications with high fidelity, using best-in-class AI predictors and the latest evidence to update interpretations with the latest data that manual methods would otherwise miss. The consistency in applying stringent oncology guidelines and the demonstrated alignment with external benchmarks underline the reliability of Connected Insights for clinical research use.

**Learn more →**

Request a free trial of Illumina Connected Insights


For Research Use Only. Not for use in diagnostic procedures.

## References


1. Austin-Tse CA, Jobanputra V, Perry DL, et al. Best practices for the interpretation and reporting of clinical whole genome sequencing. *NPJ Genom Med*. 2022;7:27 doi.org/10.1038/s41525-022-00295-z
2. Gao H, Hamp T, Ede J, et al. The landscape of tolerated genetic variation in humans and primates. *Science*. 2023;380(6648):eabn8153. doi:10.1126/science.abn8197
3. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, et al. Predicting Splicing from Primary Sequence with Deep Learning. *Cell*. 2019;176(3):535-548.e24. doi:10.1016/j.cell.2018.12.015


illumina®

1.800.809.4566 toll-free (US) | +1.858.202.4566 tel
techsupport@illumina.com | www.illumina.com

© 2026 Illumina, Inc. All rights reserved. All trademarks are the property of Illumina, Inc. or their respective owners. For specific trademark information, see www.illumina.com/company/legal.html.
M-GL-04087 v1.0