Folklore

PUBLIC SCIENTIFIC REPORT

Cohort 2: Multi-Source Performance Characterization

Three-layer real-world performance study with methodological characterization of all non-full outcomes

Document	HEL-RWP-C2-2026-001-PUBLIC
Version	1.0, August 2026
Study period	March to April 2026
Classifier	Folklore v3.28.0 at study closure
Author	Vladimir Mitev, Helena Bioinformatics Ltd.
ORCID	0009-0001-8013-6850
Report DOI	10.17605/OSF.IO/RTBK8
Public source	https://folklore.helena.bio/studies/cohort-2

This report contains aggregate, non-identifiable results only. It excludes patient registries, VCF files, case manifests, individual clinical reports, and per-case concordance records.

Suggested citation

Mitev V. Cohort 2: Multi-Source Performance Characterization. Helena Bioinformatics public scientific report, 1.0, August 2026. doi:10.17605/OSF.IO/RTBK8.

Abstract

This real-world performance study characterized Folklore v3.28.0 across 20 cases spanning 13 clinical domains, seven variant types, and eight inheritance patterns. The design used a three-layer model covering analytical, classification, and clinical workflow performance and interpreted the peer comparator alongside multi-source evidence. Variant detection was 19 of 20; the single non-detection was attributed by audit to the upstream variant-calling pipeline, with Folklore input-to-database integrity verified. Classification yielded 13 FULL components, six PARTIAL components across four cases, no DISCORDANT outcomes, and one NOT EVALUABLE case. The partial outcomes were assigned to validated design behaviour, feature-gap accumulation, or manual-evidence limitations. This report contains aggregate, non-identifiable results only.

Objective

To characterize Folklore performance across analytical, classification, and clinical workflow layers and to explain every non-full outcome through an explicit methodological-disposition framework.

Scope

This is a Phase 1 scientific performance report. It describes a bounded observational study and does not constitute regulatory clearance, a diagnostic release, a prospective multi-site trial, or authorization for unsupervised clinical use.

Terminology

Variant classes follow the five-tier ACMG/AMP framework: Pathogenic, Likely Pathogenic, Variant of Uncertain Significance, Likely Benign, and Benign [1]. FULL denotes identical class. CLINICAL denotes a one-tier difference within the Pathogenic and Likely Pathogenic boundary. PARTIAL denotes an adjacent boundary difference such as Likely Pathogenic versus VUS. DISCORDANT is reserved for more substantial disagreement. NOT EVALUABLE is reported separately from concordance denominators.

1. Background

Clinical variant classification combines population, computational, functional, segregation, allelic, and disease-specific evidence. The ACMG/AMP framework provides a common vocabulary and evidence structure, but qualified laboratories may still reach different classifications because of evidence availability, rule specification, transcript choice, disease mechanism, and the date of evaluation [1,3,4].

Published inter-laboratory studies show that structured review and data sharing can substantially improve concordance [3,4,6]. These findings support transparent characterization of differences rather than treating any single comparator as unquestioned ground truth. Folklore studies therefore distinguish analytical performance, classification performance, and clinical workflow performance, and they use multi-source evidence where the study design permits [7].

2. Methods

Study design

- Twenty performance-blind selected cases from an established clinical comparator source.
- Coverage of 13 clinical domains, seven variant types, eight inheritance patterns, and both diagnostic and carrier-screen contexts.
- Three reference classes: ClinGen expert-panel curations where available, ClinVar aggregated submissions with review status preserved, and a peer reference comparator.
- Three performance layers: analytical processing, ACMG classification, and clinical workflow output.
- Every PARTIAL, DISCORDANT, or NOT EVALUABLE outcome required a documented methodological disposition.

Comparator and ground truth

The peer reference laboratory served as a comparator. For real-world performance studies, comparator findings were interpreted alongside higher-weight public evidence, including ClinGen expert-panel curations and ClinVar records with review status preserved. The study did not assume that a single source was definitive in every disagreement.

Privacy-preserving analysis

All calculations reported here are cohort-level aggregates. No patient-level phenotype, genotype, sample identifier, referral date, clinical report, or rare case combination is reproduced. Small-cell narratives and individual variants are intentionally suppressed.

3. Results

Metric	Result	Interpretation
Variant detection	19/20 (95%)	One upstream non-detection
Gene assignment	19/19 (100%)	All evaluable targets
FULL components	13	Identical ACMG class
PARTIAL components	6	Across four cases
DISCORDANT outcomes	0	No opposite-direction result
NOT EVALUABLE	1	Upstream pipeline attribution
AI faithfulness	14/20 (70%)	Seven consecutive passes at closure
Change requests validated	10	Trigger-configuration evidence

Selected performance metrics

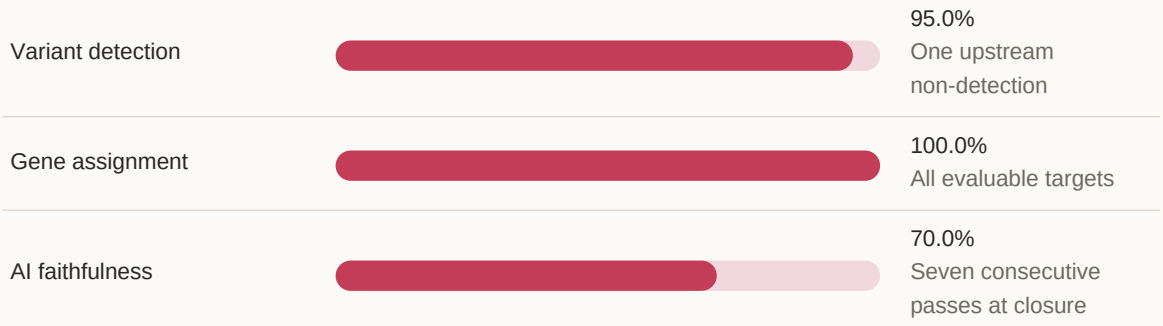

Figure 1. Selected aggregate percentages. Denominators are stated in Table 1.

4. Principal findings

Analytical attribution

The single target non-detection was absent from the supplied variant input. A documented audit verified Folklore record correspondence for the supplied inputs and cleared the Folklore processing layer for that event.

Methodological dispositions

Two partial components represented validated conservative guards, three reflected documented evidence-coverage gaps at the LP/VUS boundary, and one depended on manual evidence not available from VCF input alone.

No wider discordance

No DISCORDANT outcome or opposite-direction classification was observed. The results emphasize that an adjacent LP/VUS boundary difference is not equivalent to a Pathogenic versus Benign opposition.

Clinical workflow

Diagnostic cases showed correct tier placement for detected phenotype-matched anchors. AI faithfulness improved across the cohort and reached seven consecutive passes at cohort closure.

5. Discussion

Cohort 2 expands beyond binary concordance by separating platform behaviour, evidence coverage, manual-curation limits, and upstream data quality. This is necessary because a peer laboratory is informative but is not automatically definitive for every variant.

The result supports the use of an explicit disposition taxonomy. A conservative guard operating according to specification should not be counted as a classifier defect, while a repeatable specification breach should. Feature gaps remain scientifically relevant even when they are not defects and should be tracked as roadmap items.

The AI workflow result is reported separately from deterministic classification. This separation prevents narrative quality from being mistaken for evidence used in the ACMG class calculation.

6. Limitations

- Small sample size of 20 cases with wide uncertainty for percentage estimates.
- Single peer reference comparator.
- Convenience sampling within pre-defined diversity constraints rather than random population sampling.
- Some evidence types require manual literature, segregation, or functional review and cannot be derived from VCF input alone.
- The AI faithfulness percentage is observational and not a standalone clinical-safety claim.
- No prospective multi-site or independent external adjudication panel was used.

7. Data availability and privacy

The public data package contains this report, an OSF study summary, and links to the public Folklore evidence page. Patient registries, selection workbooks, VCF or alignment files, individual clinical reports, case manifests, and per-case concordance records are not publicly deposited. Genomic data is inherently identifying, and removal of names alone does not make a clinical genomic record anonymous.

Restricted material may be considered for controlled review only when a lawful basis, purpose limitation, confidentiality terms, an appropriate data-sharing agreement, and a secure transfer mechanism are in place. Access is assessed case by case and is not guaranteed.

8. Conflict of interest and provenance

- The author is founder, managing director, and shareholder of Helena Bioinformatics, which develops Folklore.
- The study used a peer laboratory with which Helena Bioinformatics has a professional collaboration. The comparator was treated as one evidence source, not as unquestioned ground truth.
- No external adjudication panel was engaged at this Phase 1 stage.

9. Ethics and responsible interpretation

This public report is a secondary aggregate account of work performed under collaboration and confidentiality arrangements. It includes no identifiable participant material and introduces no new participant intervention. The report must be read together with its stated limitations and must not be used as an individual diagnostic result.

10. References

1. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants. Genet Med. 2015;17(5):405-424. doi:10.1038/gim.2015.30.

2. Roy S, Coldren C, Karunamurthy A, et al. Standards and guidelines for validating next-generation sequencing bioinformatics pipelines. J Mol Diagn. 2018;20(1):4-27. doi:10.1016/j.jmoldx.2017.11.003.

3. Amendola LM, Jarvik GP, Leo MC, et al. Performance of ACMG-AMP variant-interpretation guidelines among nine laboratories. Am J Hum Genet. 2016;98(6):1067-1076. PMID:27181684.

4. Amendola LM, Muenzen K, Biesecker LG, et al. Variant classification concordance using the ACMG-AMP guidelines across nine genomic implementation research studies. Am J Hum Genet. 2020;107(5):932-941. doi:10.1016/j.ajhg.2020.09.011.

5. Pejaver V, Byrne AB, Feng BJ, et al. Calibration of computational tools for missense variant pathogenicity classification and ClinGen recommendations for PP3/BP4 criteria. Am J Hum Genet. 2022;109(12):2163-2177. doi:10.1016/j.ajhg.2022.10.013.

6. Mighton C, Smith AC, Mayers J, et al. Data sharing to improve concordance in variant interpretation across laboratories. J Med Genet. 2022. PMID:33875564.

7. Mitev V. A methodological framework for real-world performance studies of clinical variant classification platforms at early organizational stages. Zenodo. 2026. doi:10.5281/zenodo.21105027.

Public research links

Study page: https://folklore.helena.bio/studies/cohort-2

OSF validation collection: https://osf.io/xnfzd/

Report DOI: https://doi.org/10.17605/OSF.IO/RTBK8

Parent project DOI: https://doi.org/10.17605/OSF.IO/8QBMH

Contact: https://www.helena.bio/contact