Folklore

PUBLIC SCIENTIFIC REPORT

Cohort 1: Pathogenic Variant Concordance Study

Pre-registered evaluation of aggregate concordance against an established clinical reference comparator

Document	HEL-VAL-C1-2026-001-PUBLIC
Version	1.0, August 2026
Study period	February to March 2026
Classifier	Folklore v3.6.4 at study closure
Author	Vladimir Mitev, Helena Bioinformatics Ltd.
ORCID	0009-0001-8013-6850
Report DOI	10.17605/OSF.IO/KVJ8E
Public source	https://folklore.helena.bio/studies/cohort-1

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 1: Pathogenic Variant Concordance Study. Helena Bioinformatics public scientific report, 1.0, August 2026. doi:10.17605/OSF.IO/KVJ8E.

Abstract

This pre-registered validation study evaluated aggregate concordance between Folklore and an established clinical reference laboratory in 20 cases with known Pathogenic or Likely Pathogenic findings. The cohort covered 11 clinical domains, six variant types, and three zygosity categories. All 20 cases met the three primary criteria for variant detection, gene assignment, and ACMG class concordance. Fourteen cases were FULL matches and six were CLINICAL one-tier differences within the Pathogenic and Likely Pathogenic boundary. The pre-defined threshold of at least 18 of 20 cases was exceeded. No PARTIAL or DISCORDANT outcome was observed. The report presents only aggregate, non-identifiable results.

Objective

To test whether Folklore met a pre-defined 90 percent acceptance threshold for detection, gene assignment, and clinically equivalent ACMG classification in a performance-blind selected cohort.

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 cases selected before Folklore processing from an established reference laboratory registry.
- Coverage of 11 clinical domains, six variant types, and heterozygous, homozygous, and compound-heterozygous states.
- Primary criteria: P1 variant detection, P2 gene assignment, and P3 ACMG class concordance.
- Secondary criteria: HGVS notation, molecular consequence, and zygosity agreement.
- GO threshold defined before analysis as at least 18 of 20 cases meeting all primary criteria.

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	20/20 (100%)	All target findings detected
Gene assignment	20/20 (100%)	Correct gene assignment
FULL ACMG concordance	14/20 (70%)	Identical five-tier class
CLINICAL P/LP concordance	6/20 (30%)	One-tier P/LP difference
Overall FULL + CLINICAL	20/20 (100%)	All primary criteria met
PARTIAL or DISCORDANT	0/20 (0%)	No clinically wider difference
HGVS, consequence, zygosity	20/20 (100%)	All secondary criteria met

Selected performance metrics

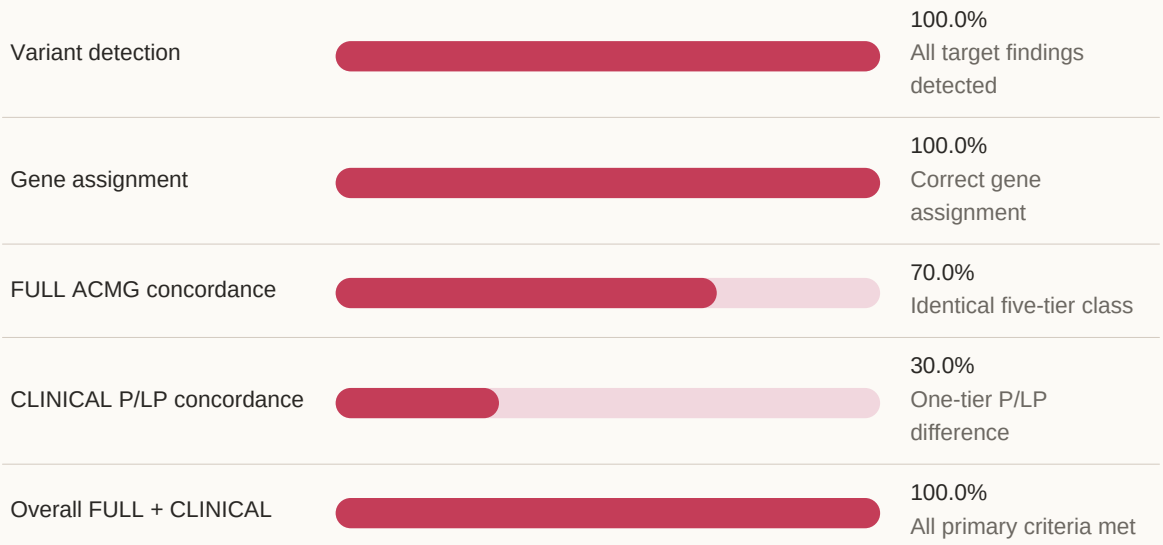

Figure 1. Selected aggregate percentages. Denominators are stated in Table 1.

4. Principal findings

Acceptance decision

All 20 cases met all primary criteria, exceeding the pre-registered threshold of 18 cases. The study therefore reached its pre-defined GO decision for this bounded cohort.

Classification differences

The six non-identical classifications remained within the Pathogenic and Likely Pathogenic boundary. Such differences are compatible with known inter-laboratory variability and did not create a Pathogenic or Likely Pathogenic versus Benign or Likely Benign opposition.

Iterative improvement

The study identified four classifier defects during development. Each was corrected and the affected cases were reprocessed before the final v3.6.4 aggregate analysis.

5. Discussion

Cohort 1 demonstrates complete performance against the study's bounded acceptance criteria. The result is strongest as evidence for the selected variant spectrum and weakest where the cohort had no representation, including population controls, benign specificity, complex structural variation, and novel evidence-poor variants.

The distinction between FULL and CLINICAL concordance is important. A one-tier P/LP difference reflects evidence weighting within the actionable boundary and should not be interpreted as identical methodology. Transparent reporting preserves that distinction while recognizing that no wider disagreement occurred.

The corrections made during the study are part of the scientific result. They show that concordance analysis was used as an engineering feedback mechanism and that the published aggregate reflects the final locked version, not the initial development state.

6. Limitations

- Small sample size of 20 cases.
- Enrichment for known Pathogenic or Likely Pathogenic findings.
- Single peer reference comparator.
- No population controls or direct specificity estimate.
- Limited representation of complex structural and non-coding variation.
- The study is descriptive and cannot establish regulatory-grade generalizability.

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.
- A scientific-advisory family relationship exists and is disclosed. The pre-defined acceptance threshold, independent operator role, and separately generated comparator reports were used as bias mitigations.
- The comparator laboratory produced its clinical reports independently of this public report.

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-1

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

Report DOI: https://doi.org/10.17605/OSF.IO/KVJ8E

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

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