KilnBio connects assay processing, multiomics integration and biological validation

Functional & Integrative Genomics
Varun B. Dwaraka and Michael J. Corley
Corresponding author: Varun B. Dwaraka, Independent researcher, Brooklyn, New York, USA
varun.dwaraka@gmail.com

Online Resource 8

Software and public-input checks for Olink, exome and variant annotation interfaces.

Online Resource 8 Integration interfaces and verification

These interfaces connect delivered measurements, variant calls and optional scientific-model annotations to an explicit research design. This supplement reports the inputs, execution checks and observed outputs for each addition, complementing the biological reanalyses in the main article.

Reproducible study execution

The study command reads a JSON plan containing explicit inputs, configuration templates, allowed KilnBio stages, dependencies and expected output files. It executes prerequisites first, records logs and hashes, stops when a stage fails and verifies declared completed outputs on resume. Failed attempts are preserved. The investigator specifies the study design, transitions between stages and output files to verify. The development checkout ran 105 tests and the research checkout ran 93. Each suite skipped two optional-dependency tests. After the study runner was updated to track changes to its source code, all nine runner tests passed.

Olink and cross-assay evidence

A public synthetic Olink fixture contained 52 baseline fictional donors and 184 assays, giving 9,568 NPX values. The independent matrix comparison agreed exactly. An additional fictional study comprised 60 donors and 200 genes across RNA, protein and methylation, with eight prespecified planted signals. One command completed NPX import, shared-contrast modeling and directional evidence aggregation; resume skipped the completed stages. Eight planted genes and none of 192 other genes were selected at the configured threshold. This seeded example verified import, stage orchestration and recovery of the planted signals under the configured decision rule. Partial-conjunction and BY procedures supplied the established statistical methods [1, 2].

Exome and clonal hematopoiesis

The exome interface uses nf-core/sarek [3] for a bounded paired-end autosomal capture route. In a 40,001-base chromosome 22 public fixture, HaplotypeCaller returned no variants. The first execution failed in upstream CNN tranche filtering of that empty callset. A separate retry explicitly omitted that filter, completed 26 tasks and still returned no variants. The successful retry used the same inputs with skip_tools set to haplotypecaller_filter in nf-core/sarek 3.12.0. Public clonal hematopoiesis work [4] provides the biological context for research screening. The exome execution check used the public fixture.

Allele and protein annotation

The public software check used the first 30 records from the retrieved GRCh38 ClinVar VCF and AlphaMissense table, preserving source headers, plus one deliberately unmatched query. The union contained 61 distinct alleles. Independent exact joins matched 30/61 queries to each respective subset, with every query retained. Unmatched alleles retained their unannotated state. VEP consequences,

ClinVar assertions and AlphaMissense predictions remained separate evidence fields [5–7]. The check verified exact allele matching and preservation of both annotated and unmatched queries.

Structure and regulatory predictions

A live AlphaFold Database query for UniProt P04637 returned one version 6 model. Two protein substitutions were checked: R175H matched the deposited reference residue and the deliberately inconsistent A175H did not. The lookup retrieved existing structure metadata [8] and verified residue correspondence. The VEP 116.3/cache 116 route was checked through configuration, version and failure-path tests. AlphaGenome 2 SDK contracts were tested offline with mocked transport. The original AlphaGenome paper [9] describes the preceding model, while version 2 request behavior is defined by the pinned SDK documentation [10]. These checks verified the respective lookup and request interfaces used to attach structural and regulatory annotations.

Contribution to reusable integration

The additions provide a configured route from delivered measurements and variant calls to shared study contrasts and annotated research candidates. The outputs preserve the distinction between measured associations, database assertions and model predictions. Assay dependence and the declared feature mappings determine how evidence is combined. The study runner records inputs, stage completion and recoverable failures, allowing the same configured analysis to be resumed and inspected.

Public sources and acquisition scope

KilnBio research source, OlinkAnalyze fixture, exome fixture. The annotation inputs were retrieved on November 13, 2026 from ClinVar GRCh38, the AlphaMissense hg38 table, and the AlphaFold Database record for P04637. Resource licenses and model-service terms apply to the respective data and model outputs.

References

[1] Benjamini Y, Heller R (2008) Screening for partial conjunction hypotheses. Biometrics 64(4):1215-22. https://doi.org/10.1111/j.1541-0420.2007.00984.x

[2] Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple testing under dependency. Ann Stat 29:1165–1188. https://doi.org/10.1214/aos/1013699998

[3] Hanssen F, Garcia MU, Folkersen L et al (2024) Scalable and efficient DNA sequencing analysis on different compute infrastructures aiding variant discovery. NAR Genom Bioinform 6(2):lqae031. https://doi.org/10.1093/nargab/lqae031

[4] Uddin MM, Saadatagah S, Niroula A et al (2024) Long-term longitudinal analysis of 4,187 participants reveals insights into determinants of clonal hematopoiesis. Nat Commun 15(1):7858. https://doi.org/10.1038/s41467-024-52302-9

[5] McLaren W, Gil L, Hunt SE et al (2016) The Ensembl Variant Effect Predictor. Genome Biol 17(1):122. https://doi.org/10.1186/s13059-016-0974-4

[6] Landrum MJ, Lee JM, Benson M et al (2018) ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res 46(D1):D1062-D1067. https://doi.org/10.1093/nar/gkx1153

[7] Cheng J, Novati G, Pan J et al (2023) Accurate proteome-wide missense variant effect prediction with AlphaMissense. Science 381(6664):eadg7492. https://doi.org/10.1126/science.adg7492

[8] Bertoni D, Tsenkov M, Magana P et al (2026) AlphaFold Protein Structure Database 2025: a redesigned interface and updated structural coverage. Nucleic Acids Res 54(D1):D358-D362. https://doi.org/10.1093/nar/gkaf1226

[9] Avsec Ž, Latysheva N, Cheng J et al (2026) Advancing regulatory variant effect prediction with AlphaGenome. Nature 649(8099):1206-1218. https://doi.org/10.1038/s41586-025-10014-0

[10] Google DeepMind (2026) AlphaGenome Python SDK, version 2.0.0. Software documentation. https://github.com/google-deepmind/alphagenome/tree/v2.0.0. Accessed 12 November 2026