Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

KilnBio connects assay processing, multiomics integration and biological validation

Varun B. Dwaraka

varun.dwaraka@gmail.com

Michael J. Corley

University of California San Diego Medical Center

Method Article

Keywords: multiomics, biomarkers, batch effects, reference materials, reproducible workflows

Posted Date: November 17th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11749512/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: Competing interest reported. V.B.D. is an employee of TruDiagnostic. M.J.C. serves as a scientific advisor to TruDiagnostic. This work was conducted independently of TruDiagnostic, in personal time and using separate resources.

KilnBio connects assay processing, multiomics integration and biological validation

Varun B. Dwaraka[1]* and Michael J. Corley[2]
[1] Independent researcher, Brooklyn, New York, USA
[2] Department of Medicine, University of California San Diego, La Jolla, California, USA
Varun B. Dwaraka ORCID 0000-0001-9033-3092
Michael J. Corley ORCID 0000-0001-8957-7153
* Corresponding author varun.dwaraka@gmail.com

Acknowledgments

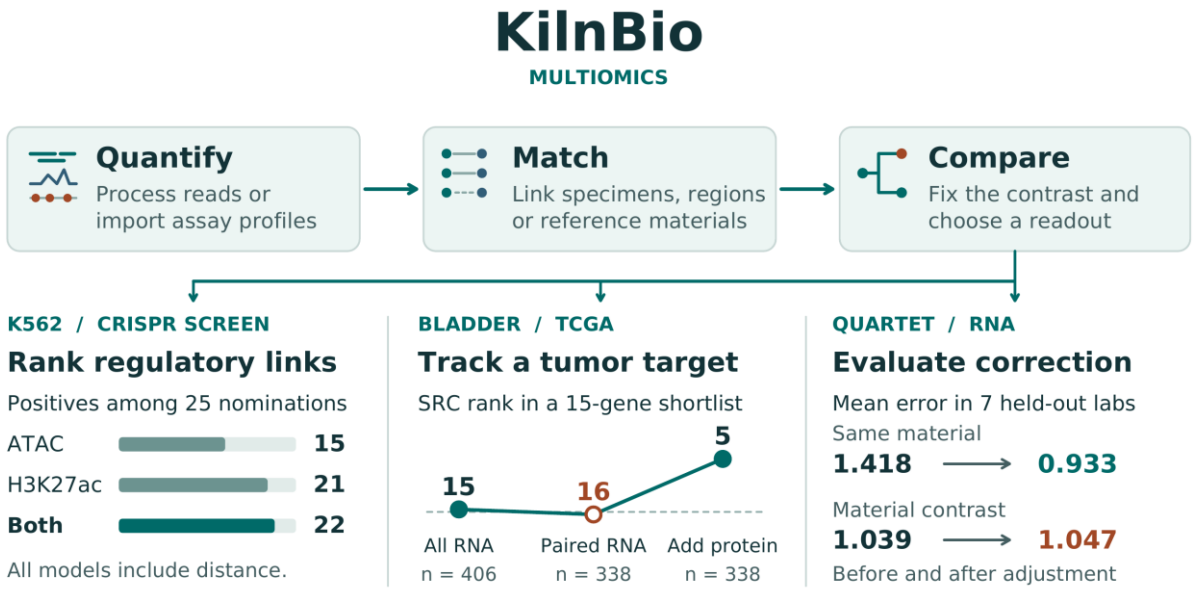
The authors thank the investigators and study participants whose contributions made the datasets available.

Abstract

A laboratory integrating genomic, epigenomic, transcriptomic and proteomic measurements must decide which specimens can be compared and which targets merit experimental follow-up. KilnBio links assay-specific processing to reproducible comparisons that separate the effects of sample selection, measurement availability and analysis choices. Researchers can run the supplied recipes from the command line or adapt individual stages to a new study. In bladder tumors, restricting the gene expression analysis from 406 patients to the 338 with protein measurements moved the SRC proto-oncogene from 15th to 16th, outside a 15-gene shortlist. Adding protein in those same patients moved SRC to fifth. A separate implementation in limma reproduced the rankings. In K562 leukemia cells, accessibility and histone acetylation together prioritized 22 perturbation-supported enhancer–gene links in 25 slots, compared with 15 using accessibility and 21 using histone acetylation. Each ranking included genomic distance. Requiring sufficient methylation coverage excluded 98 of 1,917 links, including six supported by the screen. Across seven laboratories measuring gene expression, a correction improved agreement between repeated measurements while slightly increasing error in biological contrasts. These comparisons let researchers trace why an integration decision changes the experiments they would pursue and evaluate what the added assay contributes.

Keywords multiomics, biomarkers, batch effects, reference materials, reproducible workflows

Graphical abstract



The three branches show screen-supported links, changes in a tumor target's rank and mean root mean squared error before and after laboratory correction. The dashed line marks the 15-gene cutoff. The diagram was drawn with ReportLab.

1 Introduction

Public resources such as ENCODE and The Cancer Genome Atlas have made it possible to investigate the same biological question with several kinds of molecular evidence (Cancer Genome Atlas Network 2012; ENCODE Project Consortium et al. 2020). Consider a laboratory choosing which candidate enhancers to perturb. Chromatin accessibility identifies regions that regulatory proteins might reach, acetylation of histone H3 at lysine 27 (H3K27ac) provides evidence of active chromatin, and bisulfite sequencing measures methylation at covered cytosine–guanine (CpG) sites. The laboratory wants to know which region regulates which gene. That question requires linking measurements made at different genomic scales to a candidate regulatory interaction, then asking whether perturbing the region changes expression of the proposed target (Gschwind et al. 2026).

Much of the analysis can already be done with established software. The nf-core workflows standardize processing within individual assays (Ewels et al. 2020). MultiAssayExperiment links assay records to biological units (Ramos et al. 2017). The Analyst suite and Omics Playground support integration and exploration (Akhmedov et al. 2020; Ewald et al. 2024), and MOFA models

shared and assay-specific variation (Argelaguet et al. 2018). BILBO benchmarks RNA processing against reference measurements (da Silva et al. 2026), while consensusDE and Confidence compare results across expression methods (Waardenberg and Field 2019; Shastry et al. 2026). Directional P-value merging also combines concordant evidence across molecular layers (Slobodyanyuk et al. 2024). For a particular study, the investigator must still specify how the measurements correspond and which observation would support the biological claim. An accessibility peak spans an interval, a methylation call describes a cytosine, and a perturbation screen tests a link between a region and a gene. Getting from those files to a testable list requires decisions and often additional code.

Kabartan Çökeli and Cerrahoğlu (2026) proposed evaluating whether an integrated analysis yields a testable mechanism, makes its failures visible and informs a decision. For a laboratory with a fixed number of experiments, one way to test that proposition is to count how many nominated targets have independent experimental support and to identify which targets disappear when an assay is required.

KilnBio was developed to make those steps reusable. A lab can begin with raw reads or with assay measurements it already has. The software supplies the processing commands, the code for linking observations and worked examples of how to evaluate a biological claim. In the regulatory example, a recipe maps accessibility peaks and methylation sites to candidate enhancers, links each enhancer to a proposed target gene and checks the ranking against a held-out perturbation screen. The recipe applies explicit CpG coverage criteria and records which enhancer–gene links qualify for a methylation estimate. It then compares models on their eligible links and identifies perturbation-supported candidates excluded by the coverage requirement.

Other questions need different links between measurements. RNA and protein from a tumor must refer to the same specimen if the aim is to nominate targets from their joint response. Repeated measurements of reference materials can reveal whether correcting laboratory effects preserves a biological difference. In single-cell RNA data, cells from one donor must retain that shared identity

when responses are compared across donors. The examples below show how the framework handles each of these designs. Table 1 lists the supported entry points. In paired methylation and RNA profiles, a further choice is how individual CpG measurements are summarized before genes are ranked.

2 Materials and methods

2.1 Software and units of analysis

KilnBio accepts raw reads, count matrices, molecular measurements or published feature statistics (Fig. 1). A common launcher invokes versioned Nextflow workflows for bulk and single-cell RNA sequencing, ATAC-seq, ChIP-seq, bisulfite sequencing and autosomal exome capture. Direct R, Bioconductor and Python modules handle statistical analysis and comparison. Supplied analysis recipes execute code directly and require no language-model account, prompt or generated code at runtime. The study command executes explicitly configured stages in dependency order, checks declared output files and supports resumption when the plan, listed inputs and declared outputs remain unchanged. A supplied recipe serves users who do not write analysis code; intermediate users can configure contrasts, covariates and feature maps; advanced users can call individual modules in larger workflows. Users specify the study design, feature mappings and transitions between stages. Each route records the identifiers, assay units, feature coverage and contrast direction needed downstream. A patient identifies a clinical history, a specimen identifies the material assayed, and an enhancer–gene pair identifies a regulatory hypothesis. The applications described here address molecular research questions.

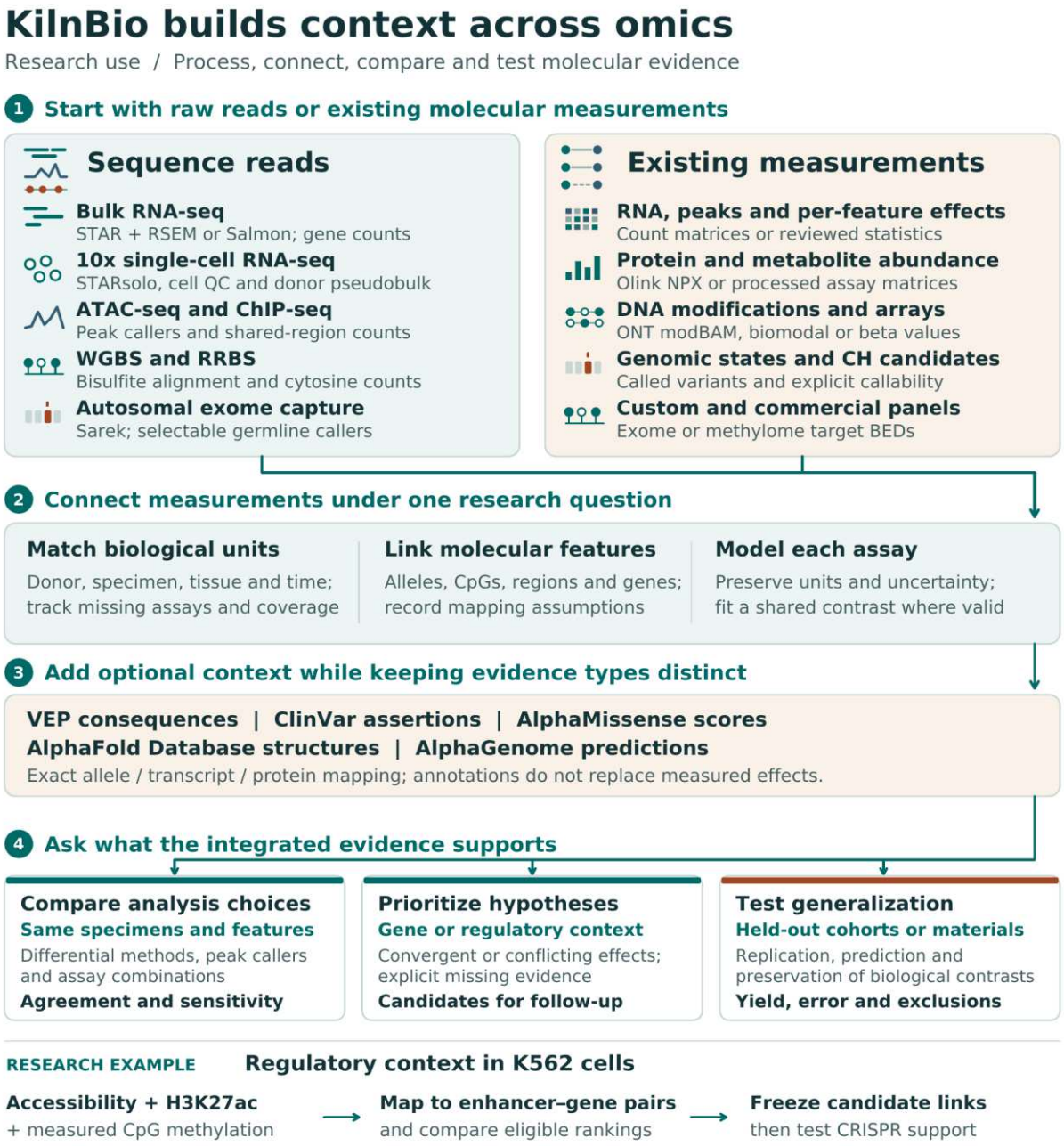


Fig. 1 KilnBio connects assay processing to multiomics research. Reads or deposited measurements enter assay-specific workflows. Biological units and molecular features are linked explicitly before statistical comparison. Optional variant and structure annotations retain their source and evidence type alongside measured effects. The K562 recipe illustrates a path from chromatin and CpG measurements to enhancer–gene hypotheses evaluated against a perturbation screen. Table 1 distinguishes supported interfaces and validation levels. The diagram was drawn in SVG

Table 1 describes what a researcher can do at each supported entry point and the checks performed for that capability. The established tools perform assay-specific processing and statistical inference. KilnBio supplies the interfaces, comparisons and records that connect those outputs to a study

question. The worked examples then evaluate how those choices change target selection, biological agreement or the population an analysis can reach.

Table 1 Assays, upstream methods and KilnBio contributions

Assay and input	Supported analysis	Established tools	What KilnBio adds and how it was checked
Bulk RNA-seq · FASTQ or counts	Quantification and differential expression	nf-core/rnaseq; STAR, RSEM or Salmon; featureCounts; tximport; DESeq2, edgeR, limma-voom	KilnBio checks count inputs and compares expression results on the same genes and donors. Human airway FASTQs supplied the quantification and differential-expression comparisons.
10x RNA-seq · FASTQ or gene-count matrix	Cell QC, clustering and donor pseudobulk	nf-core/scrnaseq; STARsolo; Scanpy; optional Scrublet; DESeq2/edgeR	KilnBio checks chemistry settings and aggregates cells by donor for statistical testing. Three methods were compared on human monocyte pseudobulk counts.
ATAC-seq · FASTQ or region counts	Peak calls and differential accessibility	nf-core/atacseq; MACS2/3, HMMRATAC; DESeq2/edgeR	KilnBio compares callers on matched fragments and tests whether read depth explains a biological contrast. Public ATAC alignments were evaluated against independent DNase profiles.
ChIP-seq · FASTQ or region counts	Narrow/broad peaks; differential signal	nf-core/chipseq; MACS3; Omnipeak comparison; DESeq2/edgeR	KilnBio checks sample and control pairing and compares normalization choices. A human perturbation experiment demonstrated how spike-in scaling changes the estimated response.
WGBS and RRBS · FASTQ or CpG counts	Bisulfite mapping; locus and region differences	nf-core/methylseq; Bismark; optional bwa-meth for WGBS; DSS; dmrseq	KilnBio links covered CpGs to regions for comparison across assays. Public WGBS measurements entered the K562 analysis; sampled human RRBS reads were checked for expected MspI prefixes.

Assay and input	Supported analysis	Established tools	What KilnBio adds and how it was checked
ONT modified DNA · POD5 or modBAM	Basecalling plan; tag QC and CpG counts	Dorado; SAMtools; modkit	KilnBio supplies a basecalling plan and imports modified-base calls. Control reads and public haplotagged human alignments tested count import and recovery of a known methylation difference.
biomodal duet +modC / evoC · vendor reports	Import cytosine counts by chemistry	Vendor decoding upstream; DSS/dmrseq for eligible counts	KilnBio converts chemistry-specific reports into cytosine counts with explicit denominators. Deposited five- and six-letter reports tested conversion; synthetic fixtures tested current vendor-format imports.
DNAm arrays and aging · beta or feature matrix	Selected clock, biomarker and cell-state scores	Published weights; Biolearn resources; PyAging; EpiDiff	KilnBio reports feature eligibility and coverage for model-specific scoring. This entry point uses processed methylation measurements and the requirements of each supported published model.
Proteins and metabolites · measured matrices	Differential proteins, reference comparisons and selected protein clock	limma-trend, t and rank tests; Quartet inputs; PyAging	KilnBio compares statistical methods and laboratory corrections on processed abundance matrices. Quartet proteins and metabolites supplied the method and reference-material comparisons.
Across assays · matrices, predictions or effects	Prediction, replication, shared contrasts and study-design audits	R/Bioconductor models; rank tests; BH; partial conjunction and BY	KilnBio links biological units and features, tracks missing measurements and compares analysis choices. The worked studies test how these choices change nominations, confirmation or prediction coverage.

Assay and input	Supported analysis	Established tools	What KilnBio adds and how it was checked
Genomics · called germline VCF or curated tumor states	Allele dosage; mutation-to-molecular-trait association	VCF genotypes; rank-sum test; BH	KilnBio links supplied genomic calls to other molecular measurements while retaining unknown status. Released leukemia labels were linked to RNA and methylation for TP53-associated comparisons.
Targeted exome or methylome · target BED plus BAM or calls	Target depth and variant or CpG eligibility	SAMtools depth	KilnBio distinguishes features outside the capture design from targets with inadequate depth. Public Twist libraries and synthetic boundary tests checked target eligibility and interval calculations.
Olink · delivered NPX	QC import and shared contrasts with other assays	Vendor-normalized NPX; limma	KilnBio imports delivered NPX values and links them to shared study contrasts. A public synthetic fixture and a fictional three-assay study tested import, modeling and resumption.
Autosomal exome · paired FASTQ	Germline calls; separate somatic route	nf-core/sarek; BWA-MEM/MEM2 or DRAGMAP; HaplotypeCaller, DeepVariant, Strelka or Mutect2	KilnBio configures a bounded autosomal capture workflow with selectable callers. A public empty-callset fixture completed after a documented retry of an upstream filtering failure.
CH · curated calls	Candidate screening and explicit unknown states	Public driver criteria; depth and allele-fraction filters	KilnBio screens supplied calls using depth, allele fraction, population frequency and driver criteria. The interface retains qualifying, excluded and unknown states for research review.
Variant context · normalized alleles and reviewed protein mappings	Consequence, assertion, score and structure annotation	VEP; ClinVar; AlphaMissense; AlphaFold Database; optional AlphaGenome	KilnBio attaches distinct consequence, assertion, prediction and structure fields to a candidate. Public lookups and exact joins were checked; VEP configuration and AlphaGenome requests had interface-level tests.

2.2 Independent testing of genes selected by RNA, protein and methylation evidence

The discovery study supplied directional statistics for RNA, protein and promoter methylation in glioma (Slobodyanyuk et al. 2024). We used the discovery authors' directional merged P values and assay-specific effects. Among 9,209 genes shared by the three assays, 31 met the 5% Benjamini–Hochberg (BH) threshold (Benjamini and Hochberg 1995) for the directional joint statistic and exceeded that threshold in each individual assay. Two comparator lists contained the 31 genes with the smallest unadjusted RNA or protein P values, respectively. Gene symbols broke exact ties. Nominations, effect directions and validation rules were fixed before candidate outcomes were examined. The joint-only list used the discovery direction for the assay being validated. Each single-assay list retained the direction available at selection.

The primary cohort comprised 33 adult historical grade IV tumors with an isocitrate dehydrogenase 1 (IDH1) R132 mutation and 93 with no R132 mutation detected, after excluding known repeat specimens from Erasmus GSE16011 (Gravendeel et al. 2009). The secondary cohort comprised six IDH-mutant and 63 IDH-wild-type patients with tumor proteomics in Huashan PXD056311 (Zhang et al. 2024). Technical measurements were pooled within each patient. A one-sided Mann–Whitney test assessed each predicted direction. At least five patients and measurements in at least half of each group were required. Each endpoint used one BH family comprising all 76 distinct gene–direction hypotheses. Untestable nominations received $P = 1$ and occupied a follow-up slot. Confirmation required the predicted direction and adjusted $P \leq 0.05$. Stratified patient resampling repeated the tests and adjustment 2,000 times with the lists fixed. Reported stability ranges are the central 95% of paired differences in confirmation counts.

2.3 Linking chromatin and methylation measurements to enhancer perturbations

We used the Gschwind et al. K562 CRISPR training and external screen tables on GRCh38 (Gschwind et al. 2026). The authors supplied the outcomes, distances to target transcription start sites and H3K27ac ChIP-seq signals. Released perturbation labels defined screen support; the source

held-out benchmark focused on enhancer-associated effects and filtered positives lacking H3K27ac signal. KilnBio took maximum overlapping ATAC peak signal from the conservative ENCODE IDR peak set and combined CpG measurements from two WGBS replicates (ENCODE Project Consortium et al. 2020). We required at least three common CpG dyads with at least five reads in each replicate. We averaged the two depth-weighted methylation fractions per element. An absent ATAC peak received zero peak signal. A missing methylation value remained unmeasured. For each test chromosome, logistic regression used only training pairs on other chromosomes. All ablations used identical eligible links, split and 25 follow-up slots. We compared average precision across all ranks and counted screen positives among the top 25. The primary comparison added methylation fraction after accounting for common CpG count and depth. Paired resampling of target genes used 2,000 draws with predictions fixed. This extension was designed after earlier analyses of the benchmark. Online Resource 3 contains the complete protocol and both entry points.

2.4 Separating the effect of patient selection from the contribution of protein measurements

Using the expanded TCGA-LUAD PanCancer Atlas data release, with the original TCGA study as biological context (Cancer Genome Atlas Research Network 2014), we compared a standard exact sample-ID inner join and limma (Ritchie et al. 2015) with KilnBio pairing and nomination audits. Before downloading molecular measurements, we fixed the study, EGFR contrast and budget of 15 nominations. We retained 128 genes with unambiguous total-protein measurements and adequate coverage in the discovery groups. EGFR itself was excluded. We fitted limma to log-transformed RNA from all eligible primary tumors, to RNA in the subset with protein, and to protein in that same subset. The joint ranking averaged signed percentile ranks of absolute moderated t statistics. Nominations and directions were frozen before downloading the independent stage I–II lung tumor series GSE31210 (Okayama et al. 2012). External confirmation required agreement in sign and a two-sided BH-adjusted P value at or below 0.05 across all 126 externally measurable genes. We resampled external patients 1,000 times within EGFR groups, refitting limma with nominations

fixed. A separate positive control exchanged two protein records' donor labels while preserving their released specimen identifiers. Online Resource 4 contains the protocol, code and target lists.

We applied the same 15-gene follow-up budget to a second cohort, the TCGA bladder PanCancer Atlas release, with the original study as context (Robertson et al. 2017). Before inspecting its rankings, we specified a contrast between primary tumors with a reported protein-altering FGFR3 mutation and mutation-profiled tumors without a qualifying call. We excluded FGFR3 from the candidate genes and matched RNA and reverse-phase protein array (RPPA) profiles by specimen identifier, with one specimen per patient. We retained the same genes across rankings, requiring a unique RNA symbol, a single eligible total-protein antibody and measurements in at least 80% of each group in each analysis. We ranked genes by RNA in all eligible tumors, by RNA in the matched tumors, and by the mean of the signed RNA and protein ranks in those same tumors. The endpoint was a change in follow-up nominations.

A retrospective implementation comparison then used the same seven checksum-verified bladder inputs, transformations, shared 129-gene panel, limma model (Ritchie et al. 2015) and 15-gene budget. The earlier bladder results were already known. A standalone R script read the original files and fitted matched RNA and protein with group-specific means and an explicit contrast, rather than the recipe's intercept and binary-group design. It combined the same signed ranks and reported the complete-case estimates, sample identifiers and nominations. The unchanged KilnBio recipe additionally supplied the all-RNA comparison and specimen, feature and nomination-change records. An extended version of the direct script added those comparisons as a positive control. Numerical equivalence required identical ranks and directions, moderated-t differences no greater than 1e-8 and score differences no greater than 1e-10. The endpoints were numerical agreement and the analysis context returned by each implementation. Online Resource 9 contains the protocol and executable comparison.

2.5 Laboratory agreement and preservation of differences between reference materials

The Quartet studies released measurements of related cell-line reference materials and compared methods for batch correction (Yu et al. 2023; Zheng et al. 2024). Published $\log_2$ matrices covered 15 batches per assay, distributed across seven RNA, ten protein and five metabolite laboratories. One entire laboratory was held out per fold. Only the first replicate of materials D5, F7 and M8 from the remaining laboratories could enter control selection or fitting. Candidate controls had the lowest training ratios of between-material to residual variance. The selection took up to 20 features, capped at one fifth of the measured features with a minimum of three. A linear model included material identity and the first principal component of standardized controls, using a custom control-feature projection motivated by the principle of estimating unwanted variation from control features (Risso et al. 2014). Each held-out sample was adjusted using the training centers, scales, loadings and coefficients, without its material label.

The comparison included the published scale and subtraction of the concurrent D6 reference on the $\log_2$ scale. Replicates two and three were reserved for evaluation. For same-material agreement, the reference for a held-out measurement was the same material measured in the other laboratories on the same processing scale. For biological agreement, each within-batch difference between two materials was compared with the unadjusted difference measured by the other laboratories. Each laboratory contributed equally to the reference. Selected controls were excluded from both scores. Root mean squared error (RMSE) was averaged within laboratory before comparison across laboratories. A smaller value indicates closer agreement with the relevant reference.

2.6 Assay controls and computational verification

Four paired donors from the airway smooth-muscle dexamethasone experiment (GSE52778) supplied a known RNA response (Himes et al. 2014). Five million complete read pairs per library were sampled with seed 74129. Gene-level effects were compared across RSEM, featureCounts, DESeq2 and edgeR on a common set of 11,854 genes (Li and Dewey 2011; Liao et al. 2014; Love et

al. 2014; Chen et al. 2025). Two PC9 chromatin immunoprecipitation sequencing campaigns tested the effect of 1 μM GSK126 for five days against vehicle, using five million reads per library (GSE64243). Joint alignment to human and fly references supplied a spike-in denominator for histone H3 lysine 27 trimethylation (H3K27me3) (Egan et al. 2016; Patel et al. 2024). The Sequencing Quality Control (SEQC) study provided published featureCounts matrices from six sites (SEQC/MAQC-III Consortium 2014). Lanes were pooled within each of four local preparations per material and site. Three distinct balanced two-versus-two partitions per material tested technical nulls. Independent quantitative polymerase chain reaction measurements supplied a comparator for the C-to-D mixture contrast.

To test whether sequencing depth explained a cell-line contrast, we used released ENCODE ATAC alignments from two GM12878 and two K562 cultures (ENCODE Project Consortium et al. 2020). We counted uniquely assigned paired fragments in 10 kb autosomal tiles after masking assembly gaps and blacklisted regions. We sampled 21,788,309 assigned fragments per culture without replacement, the smallest library total, and fitted the same edgeR contrast at both depths. For each cell line, we assessed the direction of the ATAC effect in tiles supported by both DNase replicates from that line and neither replicate from the other. We retained all tiles that passed the count filter, irrespective of ATAC significance.

Online Resource 1 gives the protocol and observations for two additional library checks. These cover three peak callers against independent DNase profiles on a three-chromosome HepG2 panel and the expected MspI read prefix in one reduced representation bisulfite sequencing (RRBS) preparation from treated YB5 cells (Zhang et al. 2018). We also checked haplotype-specific 5mC using public HG002 reads, imported published human five- and six-letter cytosine counts, and assessed target coverage using public Twist exome and methylome libraries (Online Resource 5).

Outcome counts and numerical contrasts were checked against saved analysis tables. The glioma rank tests and multiplicity adjustments were checked against SciPy and R implementations. GTEx

HC3 regression and false-discovery-rate calculations were checked against statsmodels; 126 pairwise TMM normalization checks agreed with edgeR. These checks assess computational consistency.

2.7 Assay availability and specimen matching

Processed chronic lymphocytic leukemia profiles from the MOFA release were used to predict the ex vivo response D_001_3, the middle tested concentration of the first coded drug (Argelaguet et al. 2018; Dietrich et al. 2018). Of 200 donors, 184 had an outcome and recorded gender; 121 had both RNA and DNA methylation. Four ridge-model arms shared 90 development donors, five inner folds and 31 complete test donors. Scaling, feature selection and penalty selection used only the training rows of each fit. A subsequent assay-availability analysis reused these frozen fits. It retained the 31 complete test donors and selected one quarter of each incomplete availability stratum, rounded up, by a fixed identifier hash. The two-assay policy used methylation alone when RNA was absent and abstained without methylation. Mean absolute error (MAE) was compared with methylation alone on the same donors; 2,000 paired resamples held predictions fixed.

For TCGA breast cancer (Cancer Genome Atlas Network 2012), GDC release 46.0 supplied all project cases and open STAR gene-count and 450K methylation beta file records. Eligibility was evaluated at three levels: any samples from one case, primary tumor for both assays, and the same primary-tumor sample-vial barcode. The report returned a unique file pair only when exactly one vial and one file per assay qualified. Pairing was evaluated at the deposited source-vial identifier level.

2.8 Method agreement and genomic linkage

The three-method comparison used integer counts from four paired airway donors and, separately, CD14-positive monocyte pseudobulk counts from eight paired donors in the interferon experiment (Kang et al. 2018). DESeq2, edgeR quasi-likelihood and limma-voom (Law et al. 2014; Love et al. 2014; Chen et al. 2025) tested the same features, retaining counts of at least ten in at least the smaller group size. For each method, Benjamini–Hochberg adjustment used the full retained family. The support fraction counted significant calls in the prevailing direction divided by the number of

planned methods; opposite directions and missing tests were reported separately. For protein abundance, the published Quartet $\log_2$ matrix provided F7 and M8 in ten laboratories. One replicate from the first batch in each laboratory supplied a paired material contrast for limma-trend, paired t and signed-rank tests. Moderated inference for quantitative proteomics has prior support (Kammers et al. 2015). The ten laboratories provided paired measurements of the two reference materials.

To test genomic linkage, the publicly released CLL TP53 alteration indicator was matched by donor to the released RNA and methylation matrices (Argelaguet et al. 2018; Dietrich et al. 2018). Recorded positives and negatives defined the groups; unknown calls were excluded from testing and counted. Two-sided rank-sum tests used a separate BH family per assay. This exploratory comparison used preselected molecular features without adjustment for disease subtype, ancestry, purity or other alterations.

2.9 Promoter definitions and target priorities in matched GTEx profiles

We linked published GTEx EPIC methylation measurements (GSE213478) (Oliva et al. 2023; Jain et al. 2024) to GTEx v8 gene read counts using donor, tissue and specimen-prefix identifiers. The three primary tissues were lung, ovary and transverse colon. Donor hashes assigned all tissues from a donor to discovery or validation; singleton methylation plates within a partition were excluded because the fixed-plate HC3 model would otherwise have unit leverage. The original discovery/validation groups contained 89/62, 76/47 and 47/36 profiles, respectively. Promoter features averaged M-values across at least two autosomal CpGs annotated to TSS200 or TSS1500. Genes required an unambiguous RNA symbol and at least one count per million in 20% of discovery samples. Separate RNA and methylation regressions estimated age effects per decade, adjusting for sex where variable, RNA integrity, ischemic time and methylation plate. RNA-only, methylation-only and combined rankings allocated 25 follow-up slots; the combined score averaged percentile ranks of absolute age t statistics.

Original nominations and both assay directions were fixed before validation. Directional confirmation used the maximum of the RNA and methylation one-sided P values, followed by BH adjustment over the nomination union for each source/validation-tissue pair. We retained the original lists and tested TMM normalization (Chen et al. 2025), voom inference (Law et al. 2014), alternative promoter summaries, additional donor covariates, available EpiSCORE cell-composition estimates (Zhu et al. 2022) and a three-component nuisance sensitivity. Controlled mean-versus-median comparisons held RNA normalization, donors, eligible genes and the ranking rule fixed. CpG checks assessed effect-direction coherence and removal of each probe in turn. Twenty further donor partitions tested nomination stability. These sensitivity analyses were specified after the original validation outcomes had been inspected, and overlapping partitions are not independent replications. Online Resources 6 and 7 give the methods, results and analysis code.

2.10 Additional interfaces for integrated research

The Olink interface imports delivered NPX measurements with sample and assay identifiers, product/panel information and explicit measurement-QC exclusions. The input retains the vendor-provided normalization, detection-limit information where available and product identity. Shared R models retain assay-specific units: eligible RNA counts use voom, continuous measurements use limma and array methylation beta values are transformed to M-values (Law et al. 2014; Ritchie et al. 2015). Covariates, contrasts, feature-to-gene maps and directions must be specified before integration. Sequencing methylation counts require their separate depth-aware analysis rather than this array transformation.

A separate evidence module implements directional Bonferroni partial-conjunction tests across prespecified assay layers (Benjamini and Heller 2008), with Benjamini–Yekutieli adjustment across the declared gene family (Benjamini and Yekutieli 2001). The claim tested is association in at least a specified number of layers, subject to valid directional input tests and fixed mappings and

orientations. This implementation applies established statistical ideas to the specified assay layers. The glioma benchmark retained its original directional merged statistics.

The exome adapter configures nf-core/sarek for paired-end autosomal capture data (Hanssen et al. 2024), with selectable established germline callers and a separate somatic route. Clonal hematopoiesis screening retains call depth, alternate-read counts, allele fraction, population-frequency evidence and curated driver criteria. Longitudinal clonal hematopoiesis research motivates this use case (Uddin et al. 2024). Unknown or inadequately covered states are retained separately from qualifying negatives.

Optional annotation commands preserve VEP transcript consequences (McLaren et al. 2016), ClinVar assertions and review status (Landrum et al. 2018), AlphaMissense scores (Cheng et al. 2023) and AlphaFold Database structure metadata (Bertoni et al. 2026). Allele joins require the same declared assembly and normalized biallelic representations. Protein substitutions require an explicit accession and reference-residue check; structure lookup retrieves existing AlphaFold Database models. An AlphaGenome adapter can request regulatory predictions from the provider when explicitly enabled. The original AlphaGenome paper supports the model's scientific rationale (Avsec et al. 2026); the version 2 interface is documented separately by its pinned SDK (Google DeepMind 2026). Predictions and database assertions are labeled separately from measured associations. Standard pipeline execution invokes no general-purpose language model. Online Resource 8 reports the verification procedures and supported inputs for these interfaces.

2.11 AI use

The authors wrote the initial analysis scripts and used Codex to review and integrate them. They wrote the first draft of the manuscript and used GPT-6 to review and edit it.

3 Results

The results are organized around the decisions a laboratory must make when combining molecular data. Sections 3.1–3.3 ask which genes or regulatory links would be selected for follow-up, using external measurements or perturbation screens where available. Sections 3.4–3.8 examine why those conclusions can change: laboratory correction, processing methods, missing assays, specimen matching and methylation feature definitions. Section 3.9 reports execution checks for additional interfaces. Each example identifies its starting point—raw reads, deposited measurements or published statistics—so that the added analysis can be distinguished from the work supplied by the source investigators.

3.1 Joint evidence expanded gene selection, while RNA ranking gave more expression confirmations

Slobodyanyuk et al. developed directional P-value merging and applied it to RNA, protein and promoter methylation differences between IDH groups in glioma. Their study identified genes and pathways with concordant molecular changes and evaluated pathway-level consistency in additional datasets (Slobodyanyuk et al. 2024). KilnBio began with their published gene-level statistics. We used those statistics to ask a different experimental question: if a laboratory could follow up 31 genes, how many would reproduce the expected difference in independent patients? We fixed a list of 31 genes significant only after joint integration and compared it with equally sized lists ranked by RNA or protein alone. In the Erasmus RNA cohort, 11 joint-only genes confirmed, compared with 27 RNA-selected and 18 protein-selected genes (Table 2 and Fig. 2). In paired patient resamples, the joint-only list yielded 16 fewer confirmations than RNA selection (central 95% range −18 to −3) and seven fewer than protein selection (−11 to 1). Joint integration therefore identified replicating expression associations that fell below each individual-assay significance threshold, while selecting the strongest RNA signals gave the higher expression-confirmation yield at the same follow-up budget.

The protein endpoint showed why assay coverage must remain part of that comparison. In Huashan, two joint-only, two RNA-only and five protein-only candidates confirmed. Only seven of the 31 RNA-selected genes met the protein measurement requirements; retaining all 31 allocated slots records the yield a laboratory would obtain from the original list. The joint-minus-RNA difference was zero (range −3 to 3), and no joint-only candidate confirmed in both external assays. After inspecting both cohorts, we ranked all 9,209 discovery genes by the published joint P value, allowing strong individual-assay signals to remain eligible. With BH adjustment over the revised families of 66 signed hypotheses for RNA and 67 for protein, this joint-overall list confirmed 24 of 31 genes in RNA, versus 27 for RNA selection (paired difference −3, range −6 to 1). It confirmed 3 of 31 in protein, versus 5 for protein selection (difference −2, range −5 to 0). KilnBio thus makes two distinct uses of integration testable: finding additional candidates beyond individual-assay thresholds, and ranking all candidates for the highest yield at a defined endpoint. The selection rule, available measurements and follow-up assay determined which strategy performed better here.

Table 2 Prespecified confirmations per 31 nominations (76 signed hypotheses per cohort)

Discovery rule	Erasmus tumor RNA	Huashan tumor protein
Joint-only	11/31 (29 testable)	2/31 (17 testable)
RNA-only	27/31 (31 testable)	2/31 (7 testable)
Protein-only	18/31 (29 testable)	5/31 (15 testable)

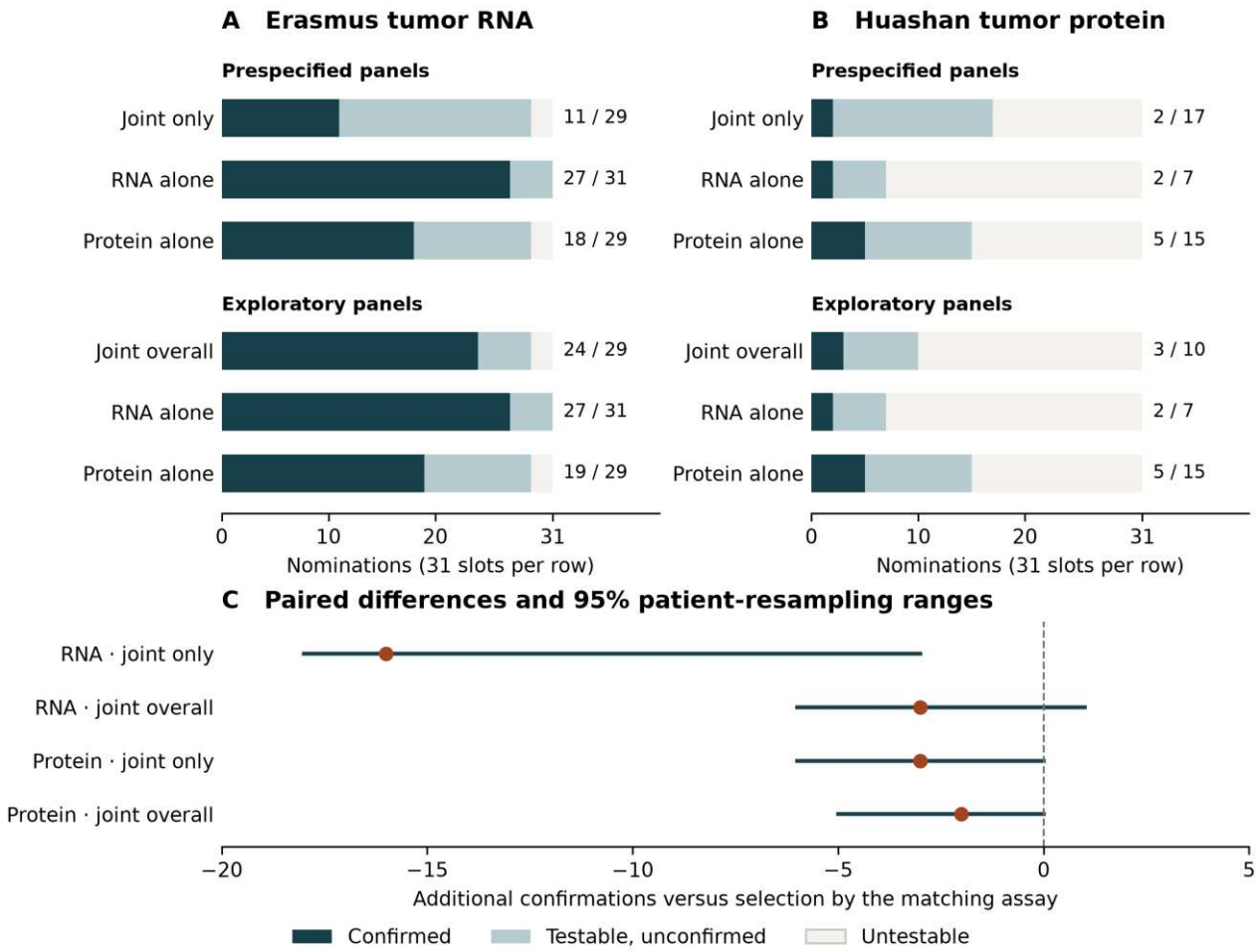


Fig. 2 External confirmations under a fixed budget of 31 nominations. In panels A and B the right-hand labels give confirmed and testable counts. Pale segments show measurable genes that fell short of confirmation. Each analysis retained all 31 slots. The joint-only rule was fixed before validation; the joint-overall ranking was examined after both outcomes had been seen. Changing the lists also changed the family of hypotheses used for adjustment. Panel C compares each joint list with selection by the assay measured in the external cohort. Bars span the central 95% of paired differences across 2,000 genotype-stratified patient resamples. RNA and protein were measured in different patients. The figure was drawn with Matplotlib

3.2 Chromatin measurements prioritized screen-supported regulatory links and exposed the cost of methylation coverage

Gschwind et al. developed enhancer–gene predictors and a benchmark assembled from genetic perturbation experiments (Gschwind et al. 2026). KilnBio reused their released K562 training and held-out screen tables, including the tested links, outcome labels, genomic distances and H3K27ac measurements, and added deposited ENCODE ATAC peaks and WGBS CpG calls (ENCODE Project Consortium et al. 2020). The analysis therefore started from processed measurements. We

asked which assay combination would most effectively narrow a laboratory's follow-up list to 25 regulatory links. Requiring sufficient methylation coverage reduced the external candidate set from 1,917 to 1,819 links, excluding 98 links, six of which were supported by the perturbation screen. The common set contained 661 regulatory elements, 414 genes and 111 screen-supported links. On those same candidates, distance plus accessibility selected 15 supported links in 25 slots, distance plus histone acetylation selected 21, and distance plus both chromatin measurements selected 22 (Table 3). Thus, 88% of the combined chromatin shortlist had screen support, compared with 60% and 84% for the respective single-chromatin-assay models. These are links between regions and genes, with support supplied by the published perturbation outcomes. The combined model concentrated 22 of the 111 supported links into 25 nominations from 1,819 candidates, increasing the observed yield by seven links over accessibility and by one over histone acetylation.

We next isolated the contribution of the measured methylation fraction from the availability of CpGs and read depth. A model containing chromatin, CpG count and depth selected 20 supported links in 25 slots; adding methylation also selected 20. The two lists shared 24 nominations: methylation brought in a screen-supported link to CD52 and displaced a screen-supported link to FEN1. Across all ranks, average precision changed by −0.015 (central 95% range −0.033 to 0.004 from paired target-gene resampling). The same predictions yielded nine supported links in ten slots with either model; at 50 slots, yield decreased from 38 to 34 after adding methylation (Online Resource 3). In this example, combined chromatin measurements produced the highest observed top-25 yield among the models in Table 3. Adding methylation changed the regulatory experiment a laboratory would choose without increasing its top-25 yield, while the coverage requirement had already removed six supported candidates. KilnBio exposes both consequences: how additional molecular evidence changes priorities among measurable links, and which supported hypotheses disappear from the search space before ranking.

Table 3 Follow-up yield on 1,819 methylation-covered K562 enhancer–gene links

Predictors	Average precision	Screen-supported links in 25 nominations
Distance	0.337	15
Distance and ATAC	0.364	15
Distance and H3K27ac	0.522	21
Distance, ATAC and H3K27ac	0.526	22
Chromatin model plus CpG count and depth	0.544	20
Coverage model plus methylation fraction	0.529	20

3.3 Why a target leaves or reenters a follow-up list

The TCGA lung study characterized tumors across molecular layers (Cancer Genome Atlas Research Network 2014), and Okayama et al. supplied an independent expression cohort with EGFR status (Okayama et al. 2012). Starting from their deposited molecular measurements and clinical labels, we asked why adding protein changes a target list: does protein supply useful ranking information, or does requiring that assay first change the patients being compared? KilnBio separated those steps. Among 510 RNA- and mutation-profiled primary lung tumors, 65 had an EGFR alteration; 356 had matched total-protein profiles, including 44 altered and 312 comparator tumors. Restricting the RNA analysis to those 356 patients changed three of 15 nominations: CCNB1, MYH11 and RAD51 left, while EIF4EBP1, ITGA2 and YAP1 entered. Both RNA lists yielded 11 confirmations in the independent cohort of 127 EGFR-mutant and 99 other tumors (paired difference 0, central 95% resampling range −2 to 1). Adding protein for the same 356 patients replaced eight more nominations and yielded seven RNA confirmations, four fewer than matched-patient RNA selection (range −5 to 0). The added assay therefore changed the proposed experiments but lowered the observed confirmation yield for this expression endpoint. Reporting the intermediate RNA analysis identifies how much of the target-list change came from excluding patients.

On the released records, an exact sample-ID join and the KilnBio specimen check produced identical joint rankings. We then deliberately exchanged the donor labels on two protein records while

preserving their specimen identifiers. KilnBio stopped at the contradictory donor assignments, and the source records identified both affected specimens. Joining by donor alone replaced KIT with EEF2K on the follow-up list. Withholding the two conflicting records made the same substitution. Joining by the preserved specimen identifiers recovered KIT. Each list still had seven external RNA confirmations. A count of confirmations alone would therefore have concealed the change in target identity.

The original bladder study characterized molecular subtypes (Robertson et al. 2017). We reused the expanded public RNA, RPPA and mutation records to apply the same separation of patient selection from protein addition. SRC illustrates why comparing only the starting and final lists can obscure the explanation. Among 129 eligible genes, SRC ranked 15th by RNA in all 406 patients, 58 with a qualifying FGFR3 mutation. Restricting RNA to the 338 patients with protein measurements, including 50 with a qualifying mutation, moved SRC to 16th and replaced it with STMN1 in the 15-gene shortlist. On those same patients, SRC ranked fourth by protein and fifth in the joint analysis, with the same association sign throughout. Restricting patients changed one nomination; adding protein changed five. KilnBio identifies SRC's return as a consequence of protein evidence in the matched patients and records its earlier exclusion as a consequence of changing the analyzed population. This makes the reason for following up SRC inspectable, even though it appears on both the starting and final lists.

The direct limma comparator and KilnBio agreed on all 387 gene ranks and directions across matched RNA, protein and joint analyses, including all 15 joint nominations. The largest moderated-t difference was 2.0e-13; signed scores were identical. KilnBio's supplied recipe additionally recorded the 68 patients excluded by requiring protein measurements, including eight with a qualifying FGFR3 mutation, and separated the one nomination changed by patient restriction from the five changed by adding protein. Its deposited-identifier audit accepted 338 unambiguous pairs and found no ambiguous donors.

When the conventional script was extended with the all-RNA fit and explicit eligibility and decision records, it reproduced all 516 four-arm ranks, the full specimen roster, all 138 prefilter feature eligibility records and the named nomination changes. The extended comparator therefore reproduced both the estimates and their interpretation. KilnBio supplies the corresponding fits, eligibility records and nomination comparisons in a reusable recipe, allowing the analysis to be repeated from the original files (Online Resource 9).

3.4 Improving repeatability slightly distorted a biological contrast in RNA

The Quartet studies supplied related cell-line reference materials, measurements across laboratories and a comparison of batch-correction methods (Yu et al. 2023; Zheng et al. 2024). Starting from their published $\log_2$ abundance matrices, we asked whether closer agreement between repeated measurements also preserves the biological differences a study aims to detect. KilnBio evaluated these two outcomes separately in laboratories excluded from fitting. The control-feature correction reduced same-material error in seven of seven RNA laboratories, eight of ten protein laboratories and two of five metabolite laboratories. Yet error in the differences between materials increased in every assay-specific fold (Fig. 3). For RNA, mean same-material error fell from 1.418 to 0.933, while mean contrast error rose from 1.039 to 1.047. Mean contrast errors before and after correction were 2.954 and 2.989 for protein, and 1.168 and 1.191 for metabolites. Each comparison uses its assay's measurement scale. The practical finding is that a correction favored by technical repeatability can still shift the biological contrast. A researcher can use both outputs to judge the correction against the purpose of the study.

Subtracting the same D6 value from both materials leaves their within-batch difference unchanged on the log scale. The observed zero change in this comparator checks the calculation. The fitted control projection instead allows a sample-specific adjustment and can alter that difference. For a new laboratory, the relevant choice depends on which contrast the study must preserve and how much distortion the application tolerates.

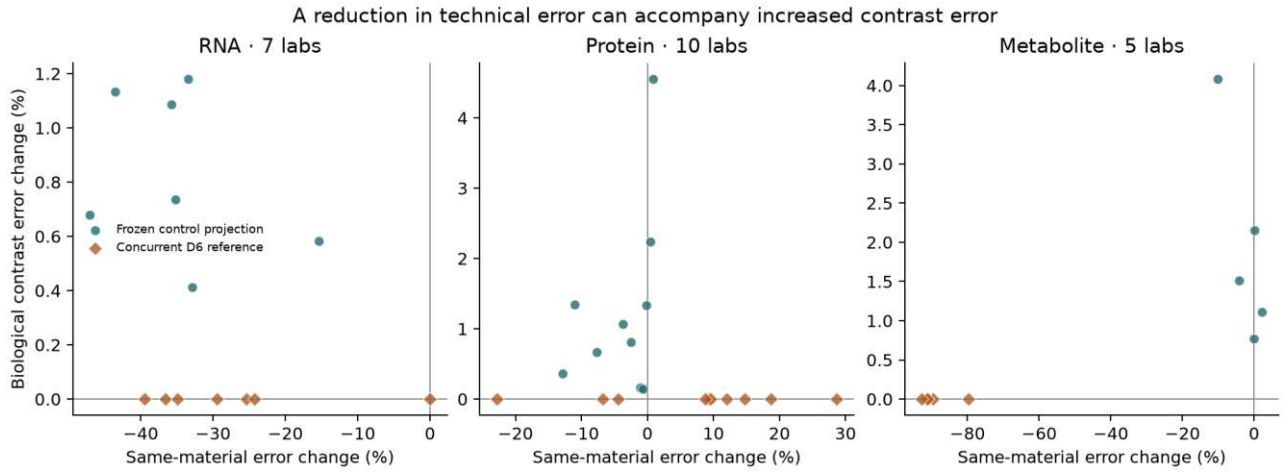


Fig. 3 Changes in same-material and contrast error within each held-out laboratory. Circles show the frozen control-feature projection and diamonds show concurrent D6 subtraction. Values are percentage changes in root mean squared error from each laboratory's published-scale baseline. Points to the left of zero have lower same-material error. Points above zero have higher contrast error. The panels have assay-specific axis ranges. Figure generated with Matplotlib

3.5 Processing choices that recover a known signal can still change its magnitude

The airway and PC9 experiments provide known biological responses against which to evaluate processing decisions (Himes et al. 2014; Egan et al. 2016). Here KilnBio began with sampled raw reads: five million paired reads per airway library and five million reads per PC9 library. All three airway analysis routes recovered glucocorticoid induction of CRISPLD2 in four of four donors. Across the common genes, however, changing quantification had a larger effect than changing the differential-expression method. With DESeq2 fixed, substituting featureCounts for RSEM gave a Spearman correlation of 0.895 between treatment effects; with featureCounts fixed, substituting edgeR for DESeq2 gave 0.998. In PC9 cells, the human-to-fly treated-to-vehicle ratios were 0.228 and 0.195 in the two campaigns. Across 87,103 fixed regions, median $\log_2$ fold change was −0.056 under library-composition scaling and −2.513 under spike-in scaling. KilnBio reproduced the published perturbation direction while showing how the processing reference changes its estimated magnitude. The spike-in comparison depends on the experimental reference and antibody assumptions (Patel et al. 2024).

SEQC supplied processed featureCounts matrices and independent quantitative measurements (SEQC/MAQC-III Consortium 2014). KilnBio used them to compare recovery of a known mixture contrast with calls between preparations of the same material. Across six sites, C-to-D RNA effects correlated with the independent measurements at 0.909–0.924 (Spearman correlation, 880–881 matched genes per site). At nominal BH 5%, DESeq2 called at least one gene in 25 of 72 same-material partitions and edgeR in 24 of 72. Reporting both outcomes shows why a strong correlation with a positive control is only one part of evaluating an analysis. The partition counts summarize reused preparations.

For ATAC-seq, KilnBio began with released ENCODE alignments and asked whether unequal read depth explained a contrast between two cell lines (ENCODE Project Consortium et al. 2020). Equalizing assigned-fragment depth preserved the expected direction in most DNase-supported tiles: 4,186 of 4,220 GM12878 tiles (99.2%) agreed, compared with 4,183 (99.1%) at full depth; 11,567 of 14,528 K562 tiles (79.6%) agreed, compared with 11,779 (81.1%). Across 214,580 tiles tested at both depths, the Spearman correlation of log fold changes was 0.992. Thus, the cell-line pattern persisted after matching depth. KilnBio supplies a way to check this technical explanation before interpreting differential accessibility biologically.

The HepG2 comparison asked whether a caller with greater agreement to independent DNase data should replace the existing default. On matched fragments from two released alignments, HMMRATAC increased mean DNase F1 from 0.338 to 0.415, while replicate Jaccard fell from 0.873 to 0.776. The loss exceeded the prespecified 0.020 switching tolerance. KilnBio made the trade-off explicit: the alternative improved one criterion while failing the frozen reproducibility requirement. Online Resource 1 also reports the expected restriction-site composition in sampled RRBS reads.

Additional examples checked whether the framework preserved biological distinctions when importing other assays. At the previously reported HG002 SNRPN locus, deposited haplotagged

nanopore alignments gave pooled 5mC of 55.4% across 229 CpGs, versus 37.0% and 74.1% on the two haplotypes. KilnBio recovered the published difference that pooling obscures. A separate targeted-exome check began with a 10,000-pair raw-read subset; an independent interval calculation agreed on all 807,103 core target bases with positive depth. These examples connect an assay-specific input to an interpretable quantity: methylation by chromosome copy or measured depth within the intended capture design. Online Resource 5 gives the input stages, denominators and five- and six-letter count-import checks.

3.6 Missing assays limit prediction coverage and specimen pairing

The MOFA release supplied processed leukemia measurements and an ex vivo drug-response endpoint (Argelaguet et al. 2018; Dietrich et al. 2018). KilnBio used these data to ask whether a prediction strategy remains useful for patients who lack an assay. Only 31 of 48 test donors had both RNA and methylation. Allowing a methylation-only prediction when RNA was absent extended coverage to 46 donors, including 15 additional people. On those same 46 donors, methylation alone had mean absolute error 0.11097, compared with 0.11327 for the policy that used RNA when available. The paired difference favoring the combined policy was −0.00230 response units (95% bootstrap interval −0.00607 to 0.00118). The fallback expanded coverage relative to requiring both assays, while methylation alone already reached all 46 and had slightly lower observed error. KilnBio makes coverage and prediction error visible together, allowing a laboratory to judge both the assay requirement and the model’s performance. Using the same public leukemia resource, Kabartan Çökeli and Cerrahoğlu (2026) evaluated how incomplete mRNA measurements affected agreement between inferred clusters and immunoglobulin heavy-chain variable-region mutation status. The drug-response analysis here evaluates a different consequence of assay availability: how many donors can receive a prediction and its error on a common set of donors.

A metadata-only audit of public TCGA breast cancer records asked a preceding question: do files recorded for the same patient represent material suitable for the intended cross-assay comparison

(Cancer Genome Atlas Network 2012)? Among 1,098 cases, 789 had RNA and methylation files somewhere in the record. Requiring primary tumor for both assays left 782; requiring the same primary-tumor vial left 781. Of these, 773 had one eligible file pair, while eight required review. KilnBio turns the choice of matching level into an explicit cohort and review list, showing why 789 patients with both assays on file is different from 773 unambiguous primary-tumor pairs.

3.7 Comparing methods and linking genotype to molecular phenotypes

The airway RNA counts, published monocyte experiment (Kang et al. 2018) and Quartet protein matrix supplied three settings for a common question: which findings persist when the statistical method changes? KilnBio applied the methods to the same eligible features and contrasts. At least one of three RNA methods called 3,082 of 12,060 airway genes, and all three called 1,781 in the same direction. All estimated CRISPLD2 induction at $\log_2$ fold changes of 2.58–2.60. In donor-level monocyte pseudobulk data, 2,340 of 2,544 genes called by any method had support from all three, among 3,703 tested. In the Quartet protein comparison, moderated and paired t tests shared 141 calls among 3,489 proteins, while the approximate signed-rank test yielded zero discoveries after BH adjustment. Requiring all three methods would therefore have removed the entire protein shortlist. KilnBio distinguishes broadly supported findings from method-sensitive findings and shows the consequence of a chosen agreement rule before that rule determines follow-up.

Genomic status can instead define the biological groups whose other molecular measurements are compared. Using the released leukemia TP53 indicator (Argelaguet et al. 2018; Dietrich et al. 2018), KilnBio linked status to RNA in 131 donors and methylation in 190. It separately recorded unknown status for five additional RNA donors and six methylation donors. The unadjusted comparisons identified 337 of 5,000 transcript features and 742 of 4,248 CpGs at BH 5% within each assay. The resulting candidates describe molecular associations with TP53 status, with their assay and donor coverage retained for follow-up. This example starts from curated alteration labels and processed molecular matrices.

3.8 Promoter summaries changed GTEx target priorities

The GTEx source studies mapped methylation relationships with genetic variation, expression and chronological age across tissues (Oliva et al. 2023; Jain et al. 2024). KilnBio began with deposited methylation beta values and gene counts, linked to authorized donor metadata, and asked how the definition of a promoter measurement changes a combined RNA–methylation shortlist. With RNA normalization, donors and eligible genes held fixed, replacing mean promoter M-values with median M-values replaced 11 of 25 combined nominations in lung, 15 in ovary and 14 in transverse colon. Across 20 additional donor partitions, the median numbers replaced were 12, 11.5 and 12, with ranges of 10–18, 6–18 and 7–16, respectively. The reanalysis identifies a source of target instability that arises before integration: different summaries of the same CpGs can send a laboratory toward different genes.

KilnBio then separated stable nomination from statistical confirmation. In the full original validation groups, none of the frozen nominations achieved BH-adjusted two-assay confirmation across 31 same-tissue and 89 source/validation-tissue specifications. Removing individual CpGs preserved the promoter age-effect direction for 21 of 25 lung, 24 of 25 ovarian and 23 of 25 colon combined nominations, but none retained corrected methylation support after every possible single-probe removal. Newly selected ovarian combined lists yielded 0–14 joint confirmations across additional, overlapping donor partitions. Together, these outputs show which feature definitions change target identity and how those nominations depend on the available donors and probes. A laboratory can examine these sensitivities before committing to validation experiments. Online Resource 6 supplies the specification-level results.

3.9 Verification of the additional software interfaces

The additional interfaces address a practical integration task: carrying delivered measurements and variant annotations through a configured study without rebuilding the connections between tools. In a fictional RNA, protein and methylation study, one command completed NPX import, shared-

contrast fitting and evidence aggregation, then resumed without rerunning completed stages. Public resource subsets independently matched 30 ClinVar and 30 AlphaMissense alleles among 61 distinct queries, including a deliberately unmatched query. A live AlphaFold Database request returned a TP53 model and distinguished a matching from an incorrect reference residue. A bounded public exome fixture completed after a separately recorded retry omitted an upstream filter that failed on its empty callset; zero variants were called. These execution checks demonstrate the tested data handoffs, annotation matching and failure handling. Online Resource 8 describes the inputs and verification level for each interface.

4 Discussion

Table 4 What KilnBio reanalyses add to published datasets and how the comparisons can be reused

Research question	Public inputs and KilnBio analysis	What the reanalysis revealed
How can chromatin and methylation guide an enhancer experiment?	Link ENCODE peaks and CpGs to the published K562 perturbation pairs and rank the same covered links.	Chromatin predictors recovered 22 positives in 25 slots. After adjustment for coverage, adding methylation exchanged a FEN1 link for a CD52 link. Both had screen support.
Do genes nominated by joint evidence replicate?	Freeze equal-sized lists from published directional glioma statistics and test independent RNA and protein cohorts.	Eleven joint-only genes replicated in RNA. RNA selection confirmed 27. Ranking all joint signals confirmed 24 in an exploratory analysis.
Why did the tumor shortlist change?	Rank RNA in all patients and again in those with matched protein, then add protein for those same patients. Test lung nominations in GSE31210.	In lung cancer, restriction replaced three of 15 targets; protein replaced eight. In bladder cancer, SRC moved from 15th to 16th to fifth. Comparing only the first and last lists masked its exclusion.
Does batch correction preserve biological differences?	Use Quartet replicates and material identities to score two kinds of error in held-out laboratories.	Repeatability improved in all seven RNA labs while error in the material contrast increased slightly.
Who can receive a prediction when an assay is missing?	Keep donors in the MOFA leukemia test roster when an assay is missing and compare fallback predictions.	A methylation fallback extended coverage from 31 to 46 of 48 donors. Optional RNA did not improve average error on those 46.

Research question	Public inputs and KilnBio analysis	What the reanalysis revealed
Do two assay records represent the same material?	Match TCGA breast cancer records at case, tissue and vial level, flagging ambiguous file pairs.	Of 789 breast cancer cases with both assays on file, 773 had an unambiguous primary-tumor pair.
Which results depend on the analysis method?	Compare statistical methods on the same donor contrasts in airway RNA and on the same Quartet proteins.	Three methods agreed on 1,781 airway genes. Requiring all three protein methods would have left no discoveries.
Which molecular traits vary with a genomic alteration?	Link recorded leukemia TP53 status to donor RNA and methylation, retaining unknown status separately.	The comparison identified transcript and CpG associations with TP53 status for follow-up.
Do target priorities depend on the methylation summary?	Compare mean and median promoter M-values in matched GTEx profiles while fixing donors, RNA and eligible genes.	Changing the summary replaced 11–15 of 25 combined targets. The original frozen lists had no corrected two-assay confirmations under the tested full-validation specifications.

Table 4 brings the examples together by the research decision they address. Their common contribution is a reproducible comparison that links molecular measurements to the experiments a laboratory would choose. In K562, the source investigators supplied the perturbation evidence (Gschwind et al. 2026); KilnBio measured how chromatin and methylation inputs change the yield, identity and eligibility of nominated links. In glioma, the source investigators supplied joint statistics (Slobodyanyuk et al. 2024); KilnBio followed fixed gene lists into additional patients and counted confirmations at the same follow-up budget as single-assay selection. In lung and bladder cancer, it separated changes caused by specimen availability from changes caused by protein evidence. These added comparisons explain the consequences of integration using the source studies' biological observations.

The examples also show why the useful outcome is sometimes a changed interpretation rather than a longer target list. Adding methylation exchanged two screen-supported K562 hypotheses while preserving top-25 yield. Adding protein produced fewer confirmations at the external lung RNA endpoint. A Quartet correction improved repeatability while slightly distorting the reference contrast.

These results help researchers decide what evidence is relevant to their intended experiment and what information would be lost by an eligibility or analysis rule. Earlier work has evaluated the value of additional data types for survival prediction (Li et al. 2024); KilnBio provides reusable recipes for examining such decisions across the target-selection, assay-processing and study-design settings demonstrated here.

The direct limma comparison connects this scientific interpretation to the software contribution (Fig. 4). Conventional code recovered the same conditional rankings, and its extended version recovered the specimen exclusions and target-list changes. KilnBio supplies the corresponding fits, records and comparisons together so that another investigator can repeat them from the original inputs. For a researcher using a supported recipe, this reduces the code that must be assembled to connect an assay result to a follow-up decision. For a bioinformatician, the same stages provide explicit inputs, outputs and assumptions that can be adapted within a larger workflow.

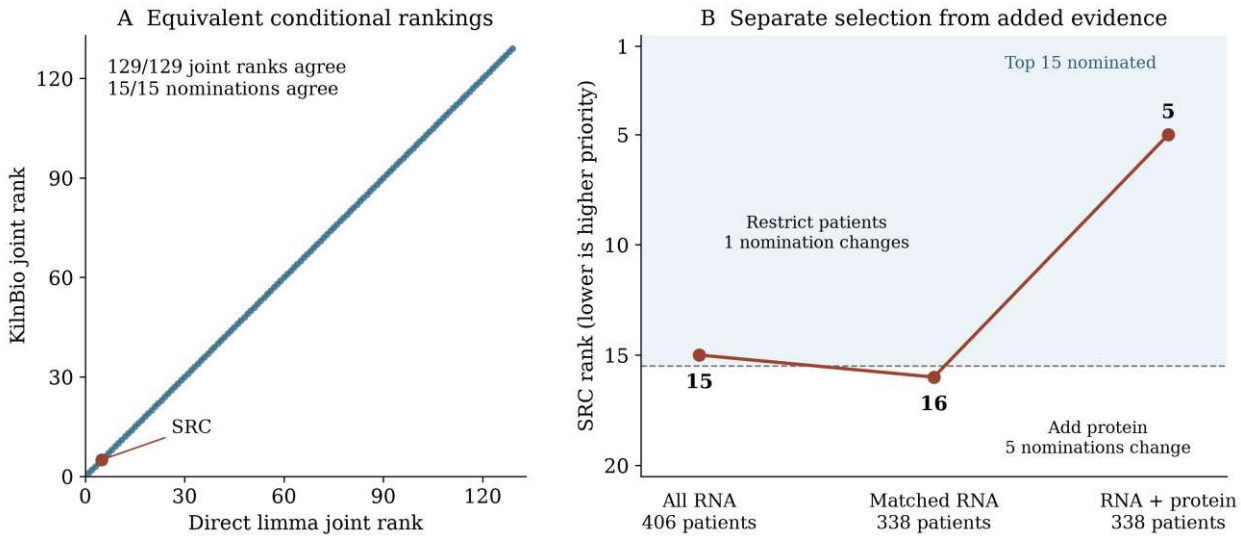


Fig. 4 Direct-workflow equivalence and analysis context. A, all 129 joint ranks and 15 nominations agree; SRC is highlighted. B, SRC crosses the top-15 boundary upon restricting RNA from 406 to 338 patients, then returns when protein is added. Listed changes count shortlist replacements. Extended conventional code reproduces these results. The figure was drawn with Matplotlib

4.1 Future development and evaluation

Further biological evaluation can extend the questions addressed by these reanalyses. Larger glioma cohorts with comprehensive IDH genotyping and adjustment for subtype, age, purity and cell composition would extend the current comparisons, in which the assays measured different patients, historical R132 testing covered part of IDH variation and the secondary protein cohort contained six mutant patients with incomplete feature coverage. Independent cohorts and perturbation experiments could test whether nominated genes and regulatory links predict functional responses. In GTEx, larger validation groups and more complete cell-composition references, including an ovarian reference, would help distinguish promoter summaries whose rankings differ. Matched-culture measurements would extend the K562 example, while broader reference panels would extend Quartet beyond one family with shared upstream processing. Crossed biological groups and platforms, together with controls or replicate measurements, would support evaluation of technical correction while preserving biological contrasts. Where biology and platform are completely confounded, such additional measurements are needed to identify their separate effects. Genetic fingerprints could extend deposited-identifier checks to physical specimen identity across assays and cohorts.

The implementation checks provide concrete starting points for new validation datasets. Empirical Olink cohorts could evaluate the cross-assay workflow demonstrated with fictional data. Whole-exome truth sets spanning coverage and allele fractions, together with low-frequency clonal hematopoiesis controls, could establish calling and screening performance beyond the current bounded execution check. Raw nanopore signal, positive 5hmC controls, full capture libraries and matched cross-chemistry measurements would extend the deposited-read and count-import examples. Full-cache VEP execution and live AlphaGenome requests would extend the current interface checks. Independent analyst studies could measure time, interpretation accuracy and usability relative to conventional workflows. Functional or prospective endpoints could then evaluate

the contribution of the added annotations to target selection. Clinical translation would require separate, task-specific analytical and clinical validation for diagnosis, screening or treatment selection.

Scientific-model annotations provide a further route to generating testable hypotheses: an observed allele can be linked to its predicted regulatory or protein consequence and compared with measured RNA, chromatin or protein changes (McLaren et al. 2016; Landrum et al. 2018; Cheng et al. 2023; Avsec et al. 2026; Bertoni et al. 2026). Evaluation against new experiments would assess the biological support for those predictions. KilnBio's specimen maps, feature definitions and provenance could also underpin future model-development datasets. A dedicated training and export workflow would need to preserve data rights and provider restrictions, separate donors or cohorts between training and evaluation, and compare model annotations with measured-only baselines. Future releases can incorporate additional public datasets, methods and analysis recipes as this evidence develops. Versioned code, input manifests and retained analysis protocols would preserve the reproducibility of the published results while documenting what each subsequent release adds.

5 Conclusions

KilnBio connects assay-specific processing to analyses that draw on several types of molecular evidence. In K562, measurements taken at peaks and individual cytosines could be brought to bear on the same enhancer–gene hypotheses. In lung and bladder tumors, repeating the RNA analysis in the patients with matched protein measurements separated the effect of patient selection from the contribution of protein to a target's rank. In the Quartet study, comparing the same reference materials across laboratories exposed a trade-off between technical repeatability and preservation of an expression difference. In GTEx, changing how promoter CpGs were summarized altered the combined shortlist on the same donors and gene family. The common feature is a recorded link between each measurement, the biological unit it represents and the observation used to assess it.

The direct limma comparison confirms that the supplied procedure preserves the underlying estimates while making the eligibility and decision comparisons reproducible.

A researcher can start with a supported sequencing workflow or enter with existing measurements, then adapt the recipe that matches the experiment. A study of regulatory mechanisms needs a link from region to gene and a perturbation readout. A prediction study needs an eligible patient roster and an outcome, including a plan for patients with a missing assay. A study combining laboratories needs a reference contrast it intends to preserve. KilnBio provides code for those designs and keeps the samples, features and comparisons visible as the analysis proceeds. The investigator can then explain both what the integration revealed and how the data support that interpretation. Supplied commands reduce the need to assemble these software interfaces from scratch. Investigators specify the biological question and study design, then use the supplied analyses to connect molecular observations to evidence for follow-up experiments.

Data and code availability

The KilnBio research code is available at https://github.com/varunorama/kilnbio-research. The public release archives contain the versions used in this study. v0.2.0-research.1 covers glioma, Quartet, K562 and lung verification. v0.2.0-research.2 covers the bladder analysis. v0.2.0-research.3 covers targeted sequencing. Commit d604bda identifies the later snapshot used for the direct-workflow comparison. Research use requires project-specific written approval under the KilnBio Noncommercial Research License.

Glioma discovery and validation data are available from the cited studies, GSE16011 and PXD056311. Quartet measurements use Figshare files 39433822, 39433819, 39433852 and 39433825. Airway and PC9 data are GSE52778 and GSE64243. SEQC matrices and reference measurements are distributed in the Bioconductor seqc package. The leukemia matrices are in bioFAM/MOFAdata revision 5c1e71b; the TCGA-BRCA audit used public GDC release 46.0 metadata. Primary glioma and Quartet analyses used KilnBio 0.2.0-dev.12; the assay and specimen

audits used 0.2.0-dev.13. We verified the glioma, Quartet, K562 and lung analyses against research release v0.2.0-research.1. We ran the bladder analysis from v0.2.0-research.2. Its README and bladder recipe identify the seven public cBioPortal TCGA PanCancer Atlas files (blca_tcga_pan_can_atlas_2018) by revision and checksum. A separate implementation rebuilt the patient roster and reproduced every gene rank and direction from those source files. Online Resource 2 contains the version-pinned analysis code for the glioma and Quartet results, the source URLs and checksums, expected tables, and a command to regenerate Table 2 and Figures 2 and 3 from 12 public source files. The package can be run independently of the development repository. Online Resource 3 reproduces the K562 comparison in Table 3 from measured interval features or from the released ENCODE peak and CpG files. Online Resource 4 gives the pinned cBioPortal TCGA-LUAD data revision, public GSE31210 inputs, signed target lists and replication code for the lung cohort. The ChIP signal and CRISPR outcomes are from the published benchmark. The ATAC regions are ENCFF855PCP and the WGBS files are ENCFF867JRG and ENCFF721JMB. Online Resource 5 identifies the nanopore, cytosine-report and capture-panel inputs. The capture-panel workflow and target definitions are documented in v0.2.0-research.3.

The ATAC experiments were ENCSR095QNB (GM12878), ENCSR868FGK (K562) and ENCSR291GJU (HepG2). Their DNase comparators were ENCSR000EMT, ENCSR000EOT and ENCSR149XIL, respectively. The RRBS reads came from GSM2813724 in GSE104998 (SRR6175605 and SRR6175606).

GTEx methylation beta values are deposited under GSE213478; RNA gene counts are from the GTEx Analysis v8 release. Exact ages and extended donor covariates were obtained from GTEx v9 phenotype metadata supplied with the provider's permission for use in this study. These donor-level inputs are not redistributed here, and full rerunning requires appropriate access to the original metadata. Online Resource 6 documents sample matching, promoter construction and sensitivity results. Online Resource 7 supplies analysis scripts and aggregate gene-level results, with a script

that reconstructs the reported summary counts without donor-level data. This GTEx extension is separate from the software releases cited above.

Online Resource 8 documents the later integration interfaces, their software tests and public-resource checks. These additions belong to the subsequent development snapshot and are separate from the archived releases used for the original biological benchmarks. Original data-processing and statistical protocols were not changed by adding these interfaces.

Online Resource 9 supplies the conventional-workflow comparison, a standalone limma implementation and the unchanged bladder recipe exported from research commit d604bdace930153a6da7bfb45f1c8b0a86ca05d0. The source manifest fixes all seven public inputs. The reproduction command runs the complete-case, KilnBio and extended conventional analyses and checks numerical and reporting agreement. The comparison-specific protocol was fixed after the earlier bladder results were known; the original scientific design and prior benchmark releases were not changed.

Supplementary information

Online Resource 1. Held-out chromatin-accessibility caller comparison and restriction-site checks in bisulfite sequencing reads.

Online Resource 2. Code, public input manifests and expected results for the glioma target-selection and Quartet laboratory-correction comparisons.

Online Resource 3. K562 enhancer–gene ranking with ATAC, histone acetylation and CpG measurements, including sensitivity to the number of follow-up slots.

Online Resource 4. Lung tumor analysis separating specimen eligibility and protein addition, with independent RNA confirmation and a specimen-label control.

Online Resource 5. Haplotype-specific nanopore methylation, five- and six-letter cytosine imports and targeted sequencing coverage.

Online Resource 6. GTEx methods and results for sensitivity to promoter construction and donor partitioning.

Online Resource 7. GTEx analysis scripts, aggregate result tables and a command to reproduce the sensitivity summaries.

Online Resource 8. Software and public-input checks for Olink, exome and variant annotation interfaces.

Online Resource 9. Bladder tumor comparison of KilnBio and direct limma, including the report, executable recipe, protocol and source manifest.

Statements and Declarations

Funding

This work received no specific funding.

Competing interests

Varun B. Dwaraka is an employee of TruDiagnostic. Michael J. Corley serves as a scientific advisor to TruDiagnostic. This work was conducted independently of TruDiagnostic, in personal time and using separate resources.

Ethics and consent

This study reanalyzed existing molecular measurements and associated metadata; no participants were recruited and no new biospecimens were collected. Donor-level GTEx inputs are not redistributed. No identifiable participant images or individual case descriptions are presented.

Author contributions

Varun B. Dwaraka conceived the study, specified the research questions, developed the analysis code, performed the analyses and prepared the original draft. Michael J. Corley contributed to conceptualization and methodology, advised on the scientific interpretation and supplied GTEx metadata with permission for use in this study.

References

Akhmedov M, Martinelli A, Geiger R et al (2020) Omics Playground: a comprehensive self-service platform for visualization, analytics and exploration of Big Omics Data. NAR Genomics and Bioinformatics 2:lqz019. https://doi.org/10.1093/nargab/lqz019

Argelaguet R, Velten B, Arnol D et al (2018) Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets. Mol Syst Biol 14:e8124. https://doi.org/10.15252/msb.20178124

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

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

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 57:289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

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

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

Cancer Genome Atlas Network (2012) Comprehensive molecular portraits of human breast tumours. Nature 490:61–70. https://doi.org/10.1038/nature11412

Cancer Genome Atlas Research Network (2014) Comprehensive molecular profiling of lung adenocarcinoma. Nature 511:543–550. https://doi.org/10.1038/nature13385

Chen Y, Chen L, Lun ATL et al (2025) edgeR v4: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets. Nucleic Acids Res 53(2):gkaf018. https://doi.org/10.1093/nar/gkaf018

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

da Silva VLC, Alvarenga JVR, Linhares-Neto MV et al (2026) BILBO: a self-hosted platform for large-scale RNA-Seq analysis with a functional annotation clustering framework. Funct Integr Genomics 26:306. https://doi.org/10.1007/s10142-026-02091-9

Dietrich S, Oleś M, Lu J et al (2018) Drug-perturbation-based stratification of blood cancer. J Clin Invest 128:427–445. https://doi.org/10.1172/JCI93801

Egan B, Yuan CC, Craske ML et al (2016) An Alternative Approach to ChIP-Seq Normalization Enables Detection of Genome-Wide Changes in Histone H3 Lysine 27 Trimethylation upon EZH2 Inhibition. PLoS One 11(11):e0166438. https://doi.org/10.1371/journal.pone.0166438

ENCODE Project Consortium, Moore JE, Purcaro MJ et al (2020) Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature 583:699–710. https://doi.org/10.1038/s41586-020-2493-4

Ewald JD, Zhou G, Lu Y et al (2024) Web-based multi-omics integration using the Analyst software suite. Nat Protoc 19(5):1467-1497. https://doi.org/10.1038/s41596-023-00950-4

Ewels PA, Peltzer A, Fillinger S et al (2020) The nf-core framework for community-curated bioinformatics pipelines. Nat Biotechnol 38(3):276-278. https://doi.org/10.1038/s41587-020-0439-x

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

Gravendeel LA, Kouwenhoven MC, Gevaert O et al (2009) Intrinsic gene expression profiles of gliomas are a better predictor of survival than histology. Cancer Res 69(23):9065-72. https://doi.org/10.1158/0008-5472.CAN-09-2307

Gschwind AR, Mualim KS, Karbalayghareh A et al (2026) An encyclopedia of human enhancer–gene regulatory interactions. Nature 657:179–190. https://doi.org/10.1038/s41586-026-10781-4

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

Himes BE, Jiang X, Wagner P et al (2014) RNA-Seq transcriptome profiling identifies CRISPLD2 as a glucocorticoid responsive gene that modulates cytokine function in airway smooth muscle cells. PLoS One 9(6):e99625. https://doi.org/10.1371/journal.pone.0099625

Jain N, Li JL, Tong L et al (2024) DNA methylation correlates of chronological age in diverse human tissue types. Epigenetics Chromatin 17:25. https://doi.org/10.1186/s13072-024-00546-6

Kabartan Çökeli E, Cerrahoğlu F (2026) Multi-omics data integration in bioinformatics: methods, applications, and the path toward biological reasoning. Funct Integr Genomics 26:326. https://doi.org/10.1007/s10142-026-02107-4

Kammers K, Cole RN, Tiengwe C et al (2015) Detecting significant changes in protein abundance. EuPA Open Proteomics 7:11-19. https://doi.org/10.1016/j.euprot.2015.02.002

Kang HM, Subramaniam M, Targ S et al (2018) Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. Nature Biotechnology 36:89-94. https://doi.org/10.1038/nbt.4042

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

Law CW, Chen Y, Shi W et al (2014) voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biology 15:R29. https://doi.org/10.1186/gb-2014-15-2-r29

Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12:323. https://doi.org/10.1186/1471-2105-12-323

Li Y, Herold T, Mansmann U et al (2024) Does combining numerous data types in multi-omics data improve or hinder performance in survival prediction? Insights from a large-scale benchmark study. BMC Med Inform Decis Mak 24:244. https://doi.org/10.1186/s12911-024-02642-9

Liao Y, Smyth GK, Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30(7):923-30. https://doi.org/10.1093/bioinformatics/btt656

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15(12):550. https://doi.org/10.1186/s13059-014-0550-8

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

Okayama H, Kohno T, Ishii Y et al (2012) Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. Cancer Res 72:100–111. https://doi.org/10.1158/0008-5472.CAN-11-1403

Oliva M, Demanelis K, Lu Y et al (2023) DNA methylation QTL mapping across diverse human tissues provides molecular links between genetic variation and complex traits. Nat Genet 55:112–122. https://doi.org/10.1038/s41588-022-01248-z

Patel LA, Cao Y, Mendenhall EM et al (2024) The Wild West of spike-in normalization. Nat Biotechnol 42(9):1343-1349. https://doi.org/10.1038/s41587-024-02377-y

Ramos M, Schiffer L, Re A et al (2017) Software for the Integration of Multiomics Experiments in Bioconductor. Cancer Res 77(21):e39–e42. https://doi.org/10.1158/0008-5472.CAN-17-0344

Risso D, Ngai J, Speed TP et al (2014) Normalization of RNA-seq data using factor analysis of control genes or samples. Nat Biotechnol 32(9):896-902. https://doi.org/10.1038/nbt.2931

Ritchie ME, Phipson B, Wu D et al (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43:e47. https://doi.org/10.1093/nar/gkv007

Robertson AG, Kim J, Al-Ahmadie H et al (2017) Comprehensive molecular characterization of muscle-invasive bladder cancer. Cell 171:540–556.e25. https://doi.org/10.1016/j.cell.2017.09.007

SEQC/MAQC-III Consortium (2014) A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. Nat Biotechnol 32(9):903-14. https://doi.org/10.1038/nbt.2957

Shastry A, Ott BP, Nandani T et al (2026) Confidence: a web app for cross-platform differential gene expression analysis, gene scoring, and enrichment analysis. Scientific Reports 16:18680. https://doi.org/10.1038/s41598-026-50527-w

Slobodyanyuk M, Bahcheli AT, Klein ZP et al (2024) Directional integration and pathway enrichment analysis for multi-omics data. Nat Commun 15(1):5690. https://doi.org/10.1038/s41467-024-49986-4

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

Waardenberg AJ, Field MA (2019) consensusDE: an R package for assessing consensus of multiple RNA-seq algorithms with RUV correction. PeerJ 7:e8206. https://doi.org/10.7717/peerj.8206

Yu Y, Zhang N, Mai Y et al (2023) Correcting batch effects in large-scale multiomics studies using a reference-material-based ratio method. Genome Biol 24(1):201. https://doi.org/10.1186/s13059-023-03047-z

Zhang H, Pandey S, Travers M et al (2018) Targeting CDK9 reactivates epigenetically silenced genes in cancer. Cell 175:1244–1258.e26. https://doi.org/10.1016/j.cell.2018.09.051

Zhang J, Sun R, Lyu Y et al (2024) Proteomic profiling of gliomas unveils immune and metabolism-driven subtypes with implications for anti-nucleotide metabolism therapy. Nat Commun 15(1):10005. https://doi.org/10.1038/s41467-024-54352-5

Zheng Y, Liu Y, Yang J et al (2024) Multi-omics data integration using ratio-based quantitative profiling with Quartet reference materials. Nat Biotechnol 42(7):1133-1149. https://doi.org/10.1038/s41587-023-01934-1

Zhu T, Liu J, Beck S et al (2022) A pan-tissue DNA methylation atlas enables in silico decomposition of human tissue methylomes at cell-type resolution. Nat Methods 19:296–306. https://doi.org/10.1038/s41592-022-01412-7

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ESM1.pdf
- ESM2.zip
- ESM3.zip
- ESM4.zip
- ESM5.pdf
- ESM6.pdf
- ESM7.zip
- ESM8.pdf
- ESM9.zip