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 9

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

Online Resource 9

Direct limma and KilnBio on identical public bladder inputs

Direct limma and KilnBio produced equivalent estimates. KilnBio also reported patient restriction and protein addition separately; an extended conventional script reproduced those records. The endpoints are numerical agreement and the analysis context returned by each implementation.

Design and provenance

The seven public TCGA bladder PanCancer Atlas inputs are pinned to Datahub revision 860b63a7f2d23540c94e25dd35302e133c29f48b. Robertson et al. [1] describe the source study; counts here derive from the expanded release. The FGFR3 contrast, transformations, feature mappings, 80% coverage rule and 15-gene budget were unchanged. Standalone limma [2] read original source files, without importing KilnBio functions or intermediate results. Both implementations used the same feature universe.

The protocol preceded these executions but followed the earlier SRC result. This retrospective comparison used the released specimen labels and the previously specified nomination endpoint. The complete-case workflow wrote matched estimates, fitted sample identifiers and group sizes. The extended comparator added the all-RNA fit, specimen and feature ledgers, and named nomination changes.

Numerical comparison

Endpoint	Direct complete-case	Extended direct
Gene ranks and directions	387/387 agree	516/516 agree
Joint nominations	15/15 agree	15/15 agree
Maximum |t difference|	2.0e-13	2.0e-13
Maximum score difference	0	0
All-RNA comparison	Not requested	Reproduced
Full specimen and feature ledgers	Not requested	Reproduced
Named selection and protein changes	Not requested	Reproduced

Frozen tolerances were 1e-8 for moderated t and 1e-10 for scores, with exact ranks and directions. “Not requested” identifies outputs added in the extended comparator. Both implementations use limma; agreement is computational replication.

Population and interpretation

Of 406 eligible RNA patients, 58 had a qualifying FGFR3 mutation. Complete cases comprised 338 patients, including 50 positive; requiring protein excluded 68 patients, including eight positive and 60 without a qualifying call. The deposited identifiers yielded 338 unambiguous assay pairs and no ambiguous donors. Pairing was assessed using deposited identifiers. The shared universe comprised 129 genes from 138 mapped features before coverage filtering.

Restricting patients changed one of 15 nominations: SRC left and STMN1 entered. Adding protein in the same 338 patients changed five: FOXM1, MSH2, MSH6, RAD51 and TGM2 left; GAPDH, IRS1, PREX1, SRC and XRCC5 entered. Retained targets did not reverse sign. SRC ranked 15th, 16th, fourth and fifth in all RNA, matched RNA, protein and joint analyses respectively, with positive direction throughout. Figure 4 shows the joint-rank equivalence and the SRC trajectory.

Reproduction

The accompanying package includes the unchanged recipe at research commit d604bdace930153a6da7bfb45f1c8b0a86ca05d0, the standalone comparator, source checksums, protocol and scorer. From the extracted package, run:

python run.py --cache PUBLIC_CACHE --out NEW_RESULTS

Requirements are Python 3.9 or later and R with limma. Options: --rscript, --r-library and --offline. The tested environment used R 4.6.2, limma 3.68.5 and statmod 1.5.2. The separate packaged replay reproduced all 23 checked scientific outputs. Commands and session receipts are retained for reproduction.

Contribution to reproducible interpretation

The comparison reproduces the earlier SRC finding and separates patient selection from the contribution of protein measurements. KilnBio supplies these comparisons and eligibility records in a reusable recipe. Agreement with the extended comparator verifies that the additional analysis context can be reconstructed from the original public inputs.

[1] 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

[2] 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