Online Resource 5

Computational protocol for the secondary analysis

Acute kidney injury is associated with less reversal of urinary metabolomic changes after coronary artery bypass grafting

Journal: Metabolomics

Donglin Xu[†], Zengkun Liu[†], Haixuan Zhao, Ang Li, Qiduo Chen, Jing Liu and Hourong Sun[*]

Hourong Sun; Department of Cardiovascular Surgery, Qilu Hospital of Shandong University, 107 Wenhua Xi Road, Jinan, Shandong 250012, China; Email: sunhourong2016@126.com; Tel.: +86-185-6008-6790

Donglin Xu and Zengkun Liu contributed equally.

Protocol status and scope

This protocol documents the completed secondary analysis retrospectively; it is not a prospective registration. It implements comparisons of postoperative urine profiles with each patient's preoperative profile and with a cohort trajectory, followed by separate urine and kidney-tissue analyses. It does not include new specimen collection, instrument acquisition or clinical diagnostic adjudication. No Protocols.io DOI has been assigned to this document.

1. Required environment and input files

Use R 4.4.1 with readxl, openxlsx and Rcpp. On Windows, provide an Rtools toolchain compatible with R 4.4.1 and a UTF-8 locale. The script checks its R version and required packages and compiles the embedded C++ routines.

The required longitudinal files are pr5c00113_si_002.xlsx (AKI Table S1) and pr5c00113_si_003.xlsx (non-AKI Table S2), associated with https://doi.org/10.1021/acs.jproteome.5c00113. These labels identify clinical groups, not positive/negative ionization modes.

The independent urine file is 11306_2026_2423_MOESM2_ESM.xlsx, worksheet fill0.5min, from https://doi.org/10.1007/s11306-026-02423-7. The script also accepts the specified local alias after content checks.

The kidney metadata input is Stage13_GSE183276/GSE183276_Kidney_Healthy-Injury_Cell_Atlas_scCv3_Metadata_03282022.txt.gz below the input directory, or a path supplied through CSA_AKI_KPMP_FILE. The source record is https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE183276.

2. Configure and invoke the analysis

Extract Analysis_Code.R from Online Resource 2. Set CSA_AKI_DATA_DIR to the directory containing the downloaded public inputs and CSA_AKI_OUTPUT_DIR to a writable output directory. The optional CSA_AKI_KPMP_FILE overrides the default kidney metadata path. Then run Rscript --vanilla Analysis_Code.R.

The script creates its output directory and cpp_cache subdirectory, uses the Mersenne-Twister generator with Inversion and Rejection settings, and sets calculation-specific seeds internally. Do not replace the fixed seeds or permutation counts when reproducing the reported estimates.

3. Check the computational kernels and assemble the longitudinal matrix

Before analysis, the script compares optimized assignment, rank-sum, reference-distance, fully refitted permutation and region-statistic kernels with direct R implementations. Tolerances are specified in the code. A failed comparison stops execution.

Intersect the two source tables by feature names and require 4,048 shared features. Parse sample identifiers of the form A or NA, patient number, underscore and timepoint 1–4. Retain patients with exactly one sample at every timepoint; require 145 complete patients, comprising 50 AKI and 95 non-AKI. This yields 580 included patient-timepoint profiles from the 622 available source samples.

Preserve the source AKI labels. Do not infer clinical onset, KDIGO stage or recovery from the molecular sampling labels. Patient-level clinical confounder adjustment cannot be reconstructed from these input matrices.

4. Calculate the within-patient return measures

Use the full 622-sample, group-blind feature matrix for this part. For each feature containing zeros, add half its minimum positive value to all observations of that feature; leave features without zeros without a pseudocount. Apply log2, feature-median centering and robust scaling by MAD multiplied by 1.4826. Fall back to IQR divided by 1.349, sample standard deviation and 1 as needed. No non-finite values were present in the analyzed source matrix.

For each complete patient, form the T3-minus-T1 earlier-change vector, T4-minus-T3 later-change vector and T4-minus-T1 residual vector. Residual ratio is the residual-vector norm divided by the earlier-change norm. Reversal fidelity is the cosine between the later-change vector and the negative earlier-change vector. Numerical denominators include 10^-12 as in the source script.

Compare AKI with non-AKI using median differences and Cliff's delta. Use 100,000 rank-sum label permutations (upper tail for residual ratio, lower tail for reversal fidelity), 20,000 stratified patient bootstrap draws and Benjamini–Hochberg correction across these two tests. Omit each patient once to assess effect direction.

5. Construct the leave-one-patient-out trajectory

Hold out one complete patient, including all four samples, in each of 145 folds. Estimate feature pseudocounts, medians and scales from the other 144 complete patients. In this trajectory analysis, unlike the preceding calculation, add the training half-minimum-positive pseudocount to every value before log2 transformation.

Express each training patient's values as changes from their own T1. Average these vectors at 0, 6, 20 and 44 h and join the four means with three straight segments. Subtract the held-out patient's own T1, project onto every segment with the projection coefficient restricted to [0,1], and retain the nearest point in the full feature space.

Record the nearest-point Euclidean distance, interpolated time, chosen segment and T4-to-T1 norm. Assigned time is bounded by 0 and 44 h and is exactly zero at baseline. Do not use AKI labels to construct the reference path. Repeat using centered unit-length within-sample ranks and sample-median-centered log profiles.

6. Compare trajectory distance and evaluate sensitivity

At T4, report the AKI-minus-non-AKI mean difference, Hedges g, 100,000 upper-tail patient-label permutations and 10,000 stratified patient bootstrap intervals. Apply the same comparisons to each representation and omit one patient score at a time. These delete-one checks do not rebuild every trajectory.

Define lag as 44 h minus assigned molecular time. For magnitude sensitivity, use the already calculated primary T4 scores: regress log1p(off-trajectory distance) on log1p(T4-to-T1 norm) using the other 144 scores, obtain the held-out residual, and compare residuals between groups. This is not adjustment for clinical covariates and not fully nested re-estimation of the original trajectories.

For the exploratory two-coordinate display, median-center and robustly scale residual ratio and T4 distance. Split at the two medians for descriptive regions, compare the 2×4 table by label permutation, and fit the joint logistic association. These display thresholds are not clinical cutoffs and do not establish patient subtypes.

7. Analyze the separate early postoperative urine cohort

Use the source fill0.5min sheet and verify 55 patient identifiers, 29 AKI and 26 non-AKI labels, and 1,384 features. Exclude features with more than 20% non-finite values or zero variance; replace remaining non-finite entries by feature medians and remove post-imputation constants. In the analyzed input, all 76,120 values were finite and all features were retained.

Construct primary robustly scaled, sample-median-centered and rank/unit representations. Score each AKI patient against the 26 non-AKI profiles and each non-AKI patient against the other 25. The score is the root-mean-square standardized difference from the reference median. Follow the representation-specific reference-scale fallback rules in the script; rank/unit scores do not use a second scale division.

For the primary AKI contrast, rebuild the reference and scores for every one of 100,000 label permutations. The 10,000 bootstrap intervals resample fixed patient scores and do not rebuild the reference. Compare the 13 severe and

16 mild source-labeled AKI patients as a secondary analysis using 100,000 permutations and 10,000 stratified bootstraps. The single sample per patient cannot test return toward baseline.

8. Aggregate and compare kidney-tissue cell states

Use GSE183276 source annotations rather than reclustering cells. Require the patient, condition.l1, state.l2, class and structure fields listed in the script. Select source-defined tubular epithelial cells in proximal, distal, collecting and intermediate tubules and aggregate counts within each patient.

The source contains 45 patients and 79,336 eligible tubular cells. Compare 12 AKI with 18 reference kidneys; the 15 CKD patients are excluded from the primary contrast. The compared 30 patients contain 39,756 cells, and each patient has at least 50 eligible cells. Compute fractions for degenerative, adaptive epithelial, transitioning, cycling and reference states.

Use patients as the resampling unit. Compare the degenerative fraction with 100,000 one-sided permutations and 10,000 stratified bootstraps. The other four state comparisons are two-sided and Benjamini–Hochberg adjusted together. The tissue data are not paired with the urine data.

9. Inspect the output and retain interpretation limits

The script writes CSA_AKI_manuscript_reproduction.xlsx with quality-control summaries, kernel checks, patient scores, trajectory scores, effect estimates, sensitivity results, runtime warnings and session information. Inspect warnings and cohort counts before interpreting the estimates. Individual bootstrap/permutation draws are not the reported study observations.

The principal reported residual-ratio difference is 0.0548 and reversal-fidelity difference is −0.0927. The T4 distance contrast is 24.980; the magnitude-adjusted and time-lag results remain inconclusive. Use the manuscript and Online Resource 1 for exact estimates, confidence intervals and rounding; do not interpret urinary profile return as observed kidney-function recovery.

This written protocol and journal-formatting revision do not constitute a new execution or an independent numerical reproduction. The original supplied analysis code is unchanged; its SHA-256 is e78969ac3a1a3936a287317acfd96bdfb8fd743b5ac8a98d4c3c948b8dd65bb0.