COMPUTATIONAL CANCER RESEARCH / PUBLIC REPORT

Technical-repeat sensitivity of fixed-reference PAM50 subtyping

A public-data computational study across RNA-seq and microarray data

We tested whether a frozen fixed-reference PAM50 method returned the same subtype label for public tumor-expression records linked to the same reported breast-tumor RNA source. In GSE96058, labels agreed for 131 of 136 pairs (96.32%; 95% Wilson interval 91.68%-98.42%) and for 83 of 87 pairs in the prespecified HR-positive/HER2-negative subgroup (95.40%; 88.77%-98.20%). A disclosed post-result TCGA-BRCA check retained labels in all five registered pairs (100%; 56.55%-100%). In the preregistered E-TABM-576 cross-platform sensitivity, 13 of 16 primary pairs retained labels (81.25%; 56.99%-93.41%). The sources were not pooled or tested for a difference. Results support source-bound repeatability evidence, not biological truth, a population agreement rate, or clinical validity.

DATE / VERSION / AUTHOR

11 September 2026 / Public report v1.0.0 / Dalron J. Robertson / NaS Research

INTERNALLY REVIEWED WITH AI ASSISTANCE / NOT PEER REVIEWED

Fully computational and public-data-only. No specimens, outcomes, controlled records, clinical threshold, or patient-facing use. Public posting begins open post-publication review. This report is not independently reviewed, peer reviewed, or intended for clinical use.

Question and design

When two public RNA measurements come from the same reported breast-cancer RNA source, does one frozen PAM50 computer method give them the same subtype label? The study measures repeatability, not whether either label is biologically correct.

Analysis	Attempted pairs	Role	Timing
GSE96058 primary	136	Primary	Preregistered
GSE96058 HR+/HER2-	87	Secondary	Prespecified
TCGA-BRCA	5	External check	Post-result preregistered
E-TABM-576	16	Cross-platform sensitivity	Preregistered

Frozen method

GSE96058 and TCGA-BRCA profiles used the fixed 50-gene panel, immutable GSE81538 reference, five fixed PAM50 centroids, and Spearman correlation, with no cohort centering. E-TABM-576 used the preregistered 59-reporter collapse and gene-wise centering against 112 singleton profiles before the same fixed-centroid correlation. Every linked pair remained in its attempted denominator. No imputation, outcome-guided adaptation, or post-result reliability threshold was allowed. Complete executions had to be byte-identical.

Data boundary

Only public processed expression and necessary public metadata were used. NaS handled no physical specimen and accessed no clinical outcome or controlled record.

Repeat-agreement results

Cohort	Matching / attempted	Agreement	95% Wilson interval
GSE96058 primary	131/136	96.32%	91.68%-98.42%
GSE96058 HR+/HER2-	83/87	95.40%	88.77%-98.20%
TCGA-BRCA external	5/5	100.00%	56.55%-100.00%
E-TABM-576 sensitivity	13/16	81.25%	56.99%-93.41%

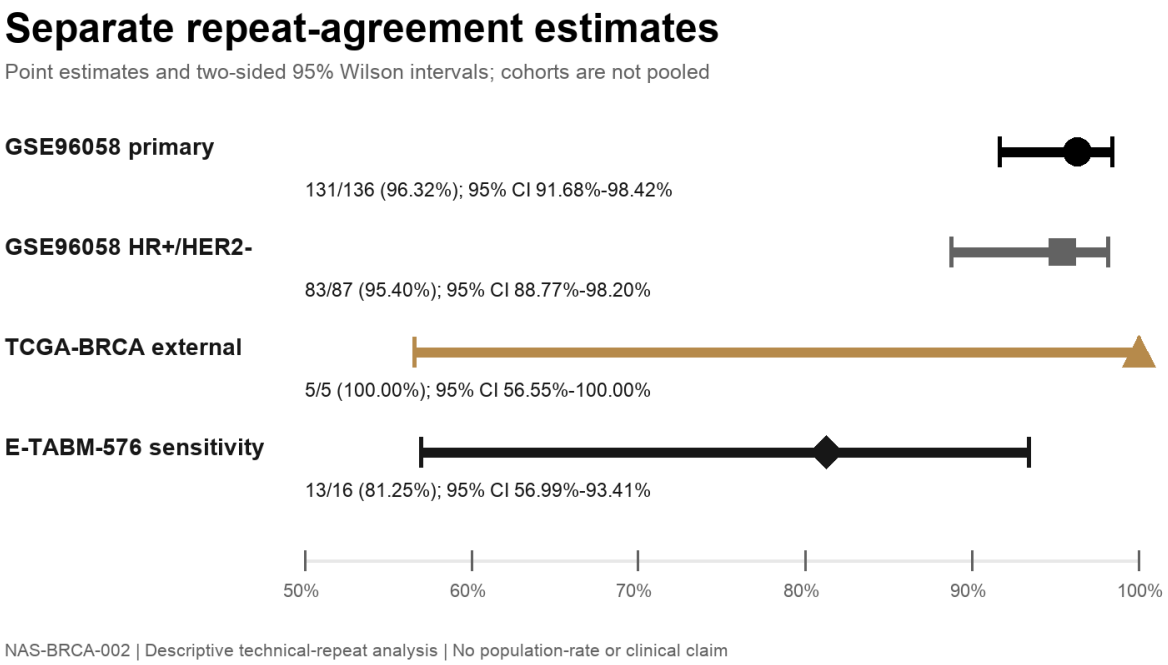

Figure 1. Separate attempted-pair agreement estimates and two-sided 95% Wilson intervals. The five-pair TCGA and 16-pair E-TABM-576 estimates are imprecise. Cohorts were not pooled or tested for a difference.

Cross-platform sensitivity

All 16 registered E-TABM-576 primary pairs were valid. Thirteen retained their PAM50 labels and three did not, for 81.25% agreement with a 56.99%-93.41% Wilson interval. The median within-pair centered-profile Spearman correlation was 0.619.

E-TABM-576 view	Matching / attempted	Agreement	95% Wilson interval
Primary independent-source pairs	13/16	81.25%	56.99%-93.41%
Secondary dependent comparisons	31/39	79.49%	64.47%-89.22%

Interpretation boundary

The microarray sensitivity point estimate is numerically lower than the GSE96058 estimate, but no between-source test was planned or performed. Platform, processing, reference, cohort, preservation, and other differences remain entangled. The result does not establish a statistical difference or its cause.

The 39 secondary comparisons reuse profiles and are not independent. Participant-level nonoverlap with GSE96058 was not identifier-verified.

Score margins

For each valid pair, the smaller winning-versus-runner-up score margin was summarized. The few nonmatching GSE96058 pairs had lower medians than matching pairs, but no group test or cutoff was planned or selected.

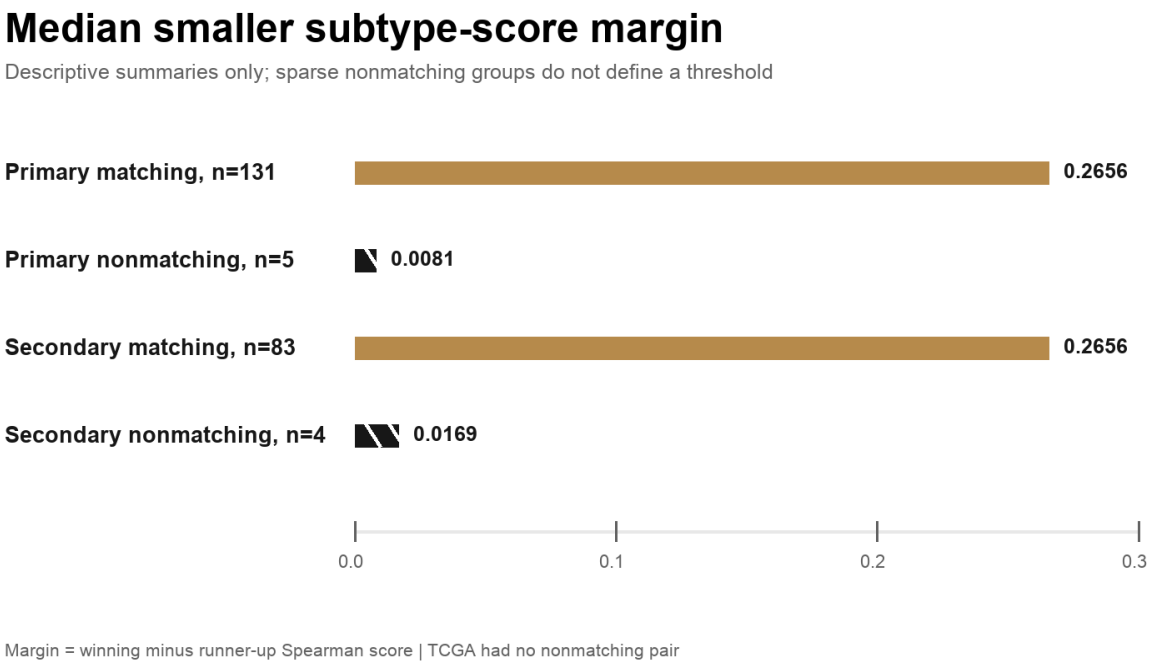

Figure 2. Median smaller PAM50 score margin by repeat-label state. Nonmatching groups contain only five primary and four secondary pairs. This descriptive contrast does not define a reliability threshold.

Numerical incident

The first TCGA implementation attempt rejected two exact log2(0.1) boundary values. That provisional output was rejected and not interpreted. The accepted run applied only the 1e-12 floor tolerance frozen before TCGA molecular access.

Interpretation

The frozen implementation showed high but imperfect repeat-label agreement in GSE96058 and retained labels in all five TCGA-BRCA pairs. The five-pair external result is encouraging but its 56.55% lower confidence limit shows substantial uncertainty.

Limitations

The external candidate was found after the original result. GDC and SCAN-B used different upstream expression workflows, while the fixed reference shares the SCAN-B context. GSE96058 mixes repeat mechanisms. The discordant groups and external cohort are small. Technical-repeat agreement cannot identify biological truth or establish diagnostic, prognostic, predictive, treatment, clinical-utility, or general transportability validity.

Public-source search update

Continued public-source screening admitted E-TABM-576 only as a separate cross-platform sensitivity analysis. No larger eligible unchanged-method repeat dataset has been identified, and search completeness is not claimed.

Conclusion

The evidence supports a narrow source-bound repeatability conclusion. It does not support a population rate or clinical claim. This public release invites open post-publication review; it is not independent review or journal peer review.

Reproducibility and sources

The package accompanying this public report contains the versioned manuscript, checksum receipts, review packet, numerical-incident record, current search refresh, exact figure and paper builder, and the shared NaS renderer and palette. It excludes raw molecular data, identifiers, credentials, outcomes, and controlled records.

Review provenance: the founder directed an AI-assisted internal release audit and authorized website publication after an audit pass. The report has not received independent scientific review or journal peer review and is not for clinical use.

Public sources

GSE96058 - NCBI Gene Expression Omnibus

TCGA-BRCA - NCI Genomic Data Commons

E-TABM-576 - EMBL-EBI BioStudies

Original PAM50 description - PubMed 19204204

Published repeatability context - PubMed 36892725