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.

Fixed-panel tissue RNA prioritization in extraskeletal myxoid chondrosarcoma: CSPG4 evidence across cohorts with comparator and sequencing-year limits

Tristan McRae

trimcrae@gmail.com

Unaffiliated https://orcid.org/0000-0002-1823-1451

Research Article

Keywords: extraskeletal myxoid chondrosarcoma, CSPG4, CHRNA6, tissue RNA, comparator sensitivity, reproducibility

Posted Date: September 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10959636/v1

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

Additional Declarations: The authors declare no competing interests.

Fixed-panel tissue RNA prioritization in extraskeletal myxoid chondrosarcoma: CSPG4 evidence across cohorts with comparator and sequencing-year limits

Version of 2026-09-07 · preprint · Research Square submission copy

Author. Tristan D. McRae

Independent researcher, unaffiliated. Correspondence: trimcrae@gmail.com ORCID: 0000-0002-1823-1451

Running title. Tissue RNA priorities in EMC

Abstract

Background: Tissue RNA enrichment can prioritize experimental assessment of candidate therapeutic addresses in extraskeletal myxoid chondrosarcoma (EMC), but the answer depends on comparator histology, specimen provenance and assay. We evaluated a fixed 11-gene panel and a separate CHRNA6 context control without treating bulk RNA as evidence of accessible protein or normal-tissue sparing.

Methods: We analyzed publicly released gene TPM from a 704-patient soft-tissue-tumor study. Nine primary EMC remained after excluding three explicitly previously reported EMC cases and one recurrence. Primary comparators were myxoid liposarcoma (n=14), low-grade fibromyxoid sarcoma (LGFMS; n=13) and synovial sarcoma (n=18). We calculated per-gene probability of superiority, A=P(EMC>comparator)+0.5P(tie), with equal histology weights, separately for marginal and sequencing-year-matched contrasts. Each matched comparison contained three EMC; their union comprised four patients. A prioritization rule and sensitivity analyses were frozen before target values were inspected. Original GSE24369 arrays provided a separate LGFMS replication anchor (6 EMC biopsies; 17 LGFMS) and three secondary shared histologies.

Results: CSPG4 alone met the 11-gene prioritization rule: marginal A=0.895 and matched A=0.811. CSPG4's LGFMS-specific A was 1.000 in GSE24369 and 0.966 marginal/0.933 matched in the RNA-sequencing cohort. Single-EMC and single-histology deletions preserved the positive composite direction. However, removing sequencing-year 2019 reversed the matched composite to 0.433; the marginal DFSP context comparison was also nonpositive (A=0.467). CHRNA6, the separate control, had A=1 in all primary comparisons. Other candidates displayed weak, reversed or comparator-dependent effects. Normal-expression sources and earlier 3SEQ observations did not support normal-tissue restriction.

Conclusions: The fixed-panel analysis supports a qualified CSPG4 tissue-validation rationale and identifies consequential negative directions for other candidates. The year-sensitive matched result, normal-expression overlap and unresolved malignant-cell localization preclude a broadly robust EMC-selectivity or therapeutic-validation claim.

Keywords

extraskeletal myxoid chondrosarcoma; CSPG4; CHRNA6; tissue RNA; comparator sensitivity; reproducibility

Introduction

An RNA signal can motivate tissue validation of a therapeutic address, but it cannot establish the protein's location, accessibility or therapeutic window. For rare tumors, the available expression cohorts also differ in assay, comparator mixture and provenance. A candidate that ranks highly against one histology may not distinguish the same tumor from another. A useful computational assessment should therefore retain individual comparators and contrary data rather than collapse them into a single favorable tumor-versus-other estimate.

EMC illustrates this distinction. Dulken and colleagues identified CHRNA6 through expression-array analysis of GSE24369 and subsequently demonstrated strong, diffuse CHRNA6 **RNA chromogenic in situ hybridization (CISH)** in 25 EMC cases. Their mimic series contained no threshold-level overexpression, although limited below-threshold signal occurred in 69 of 685 mimics. This was an RNA tissue assay, not protein immunohistochemistry or a therapeutic-accessibility experiment.[1] CHRNA6 is consequently a prior-supported context control here, not a newly discovered address; reanalysis of its discovery array is not an independent validation of that discovery.

We asked whether a fixed panel of 11 previously nominated address transcripts shows consistent EMC-high ordering across relevant sarcoma histologies in a newly accessible tissue RNA-sequencing cohort, and whether the same-histology direction agrees with original public arrays. The contribution is a comparator-explicit prioritization analysis with falsifiable sensitivity checks, not identification of a universally specific marker. We separate same-histology replication from different comparator composites and retain normal-expression evidence as interpretation context.

Methods

Cohorts and fixed panel

Hofvander et al. released a 19,116-gene by 704-patient TPM matrix with sample metadata and supplementary clinicopathological information.[2] The study primarily recruited patients diagnosed and treated in Lund or Stockholm during 1988–2020, with some specimens from earlier collaborations. Gene quantification used STAR 2.7.10b/GRCh38 and RSEM 1.3.3/Ensembl 104. We retained original Table S1 diagnoses; transcriptomic reclassification was examined separately.

Of 13 EMC specimens, three explicitly previously reported cases were excluded using the source's case references (104-92, 168-97, 536-00), as was one local recurrence (5081-14). Nine primary specimens remained. All nonblank recurrence/metastasis specimen flags were excluded symmetrically in comparator histologies. Positive overlap evidence was removed; absence of a publication reference was not interpreted as proof of universal independence. These exclusions support an overlap-reduced comparison, with probable rather than completely verified independence from all historical datasets.

The prespecified address panel was CD276, SSTR2, PRAME, FAP, CD248, CSPG4, MSLN, L1CAM, GPC3, ALPP and CDH17. CHRNA6 was analyzed separately. PRAME represents an intracellular peptide-HLA address concept rather than an ordinary intact surface protein; no endogenous peptide presentation was measured. Each symbol had one exact gene row in the TPM matrix. No expression-dependent sample or gene filtering was applied.

The primary same-cohort comparators were myxoid liposarcoma (MLPS), LGFMS and synovial sarcoma, with equal weights. Myxofibrosarcoma (MFS) and dermatofibrosarcoma protuberans (DFSP) were separate context comparisons. The original GSE24369 GPL6244 dataset comprised 42 samples: 6 EMC biopsies, 17 LGFMS, 6 MFS, 6 desmoids, 5 solitary fibrous tumors (SFT) and 2 pooled skeletal-muscle RNAs.[3] All original biopsies were retained because lesion stage was not established; they were not assumed to be primary lesions. The two normal

pools were descriptive context, not two individual healthy controls. One uniquely assigned transcript cluster per gene was selected from original platform annotation before inspecting target values. Released RMA log2 signals were used without across-gene standardization or rounding.

Estimands and prioritization

For each gene and histology we calculated A=P(EMC>comparator)+0.5P(tie) across all available specimen pairs. A=0.5 denotes neutral pair ordering; A=0.7 corresponds to 70% superiority after assigning half of tied pairs. This is neither a fold change nor a diagnostic classification accuracy established in an external test set.

We reported marginal A using all nine EMC and, separately, exact-sequencing-year contrasts. Years with at least one EMC and one relevant comparator were retained, including singleton strata. Within each histology, supported-year effects were weighted by the number of EMC in those years. MLPS/LGFMS supported 2019 and 2021; synovial sarcoma supported 2019 and 2020. Thus each matched contrast used three EMC, with four unique patients across the three contrasts. The marginal and matched summaries answer different questions and were not substituted for one another.

A gene met the prospective-to-us tissue-prioritization rule only if its marginal equal-histology A was ≥0.70, each primary histology had A>0.50 under both definitions, and both composite directions remained >0.50 after every single-EMC and single-histology deletion. The 0.70 benchmark was frozen as a tangible 20-percentage-point departure from neutral ordering, not selected from new values and not a clinical cutoff. Sequencing-year, comparator-patient and revised-diagnosis sensitivities were prespecified interpretation qualifiers, not additional pass criteria. No rule was applied to CHRNA6.

The primary cross-cohort anchor was LGFMS-specific A, calculated separately in each cohort. MFS, SFT and desmoid were separate secondary shared-histology comparisons. We compared effect size, sign and single-biopsy deletion sensitivity, without pooling platform values, defining another success cutoff or substituting a different comparator composite. GSE4303 was excluded from abundance replication because the composition of its two-color reference was unresolved.

Uncertainty, normal context and reproducibility

For primary Hofvander estimates, 2000 bootstrap resamples (seed 20260906) were stratified by histology and sequencing year, sharing each EMC resample across comparisons. Reported 2.5th–97.5th percentiles are pointwise intervals conditional on observed strata. Singleton strata remain fixed and can produce overly narrow or degenerate intervals. We therefore also report raw denominators, individual placements and all single-patient, histology and year deletions. Removing a year or last specimen can change support; weights were recalculated and unsupported contrasts remained undefined. No P values or multiplicity-selected discovery set was produced.

HPA gene XML/JSON records supplied separate normal-expression and localization context.[4, 5] Version 25 XML entries and current 25.1 documentation were distinguished. Normal IHC reliability, absent records, compartment discordance and antibody ambiguity were retained. No HPA nTPM-to-tumor-TPM safety ratio was calculated. Historical GSE28866 findings were considered under their distinct depth-scaled, square-root-compressed 3SEQ peak estimand, without pooling values or recounting four EMC library records as four established independent patients.[6]

The sample/probe contracts and prioritization rule were frozen before new target-expression values were inspected. Source files, metadata and code are hash-pinned. An independent implementation reparsed the original arrays, platform annotation and RNA-sequencing matrix and reproduced the estimates and deletion calculations. The supplement gives the exact files and remaining provenance limitations.

Use of artificial intelligence. OpenAI Codex assisted with source recovery and organization, analysis-code implementation, execution, verification, and manuscript drafting. Independent implementations checked original source values and the reported arithmetic. AI assistance does not establish biological validity; responsibility for the final interpretation and released article remains with the author.

Results

Broad prioritization differs from a favorable single comparator

CSPG4 alone met the address-panel rule, with marginal composite A=0.89454 (pointwise conditional interval 0.86614–0.91806) and matched A=0.81111 (0.75278–0.86111). Its marginal A values were 0.78571 versus MLPS, 0.96581 versus LGFMS and 0.93210 versus synovial sarcoma; corresponding matched values were 0.66667, 0.93333 and 0.83333. CHRNA6 had A=1 across these comparisons but remained outside the 11-gene pass count. Figure 1 and Table S1 display every gene, including reversed and neutral directions.

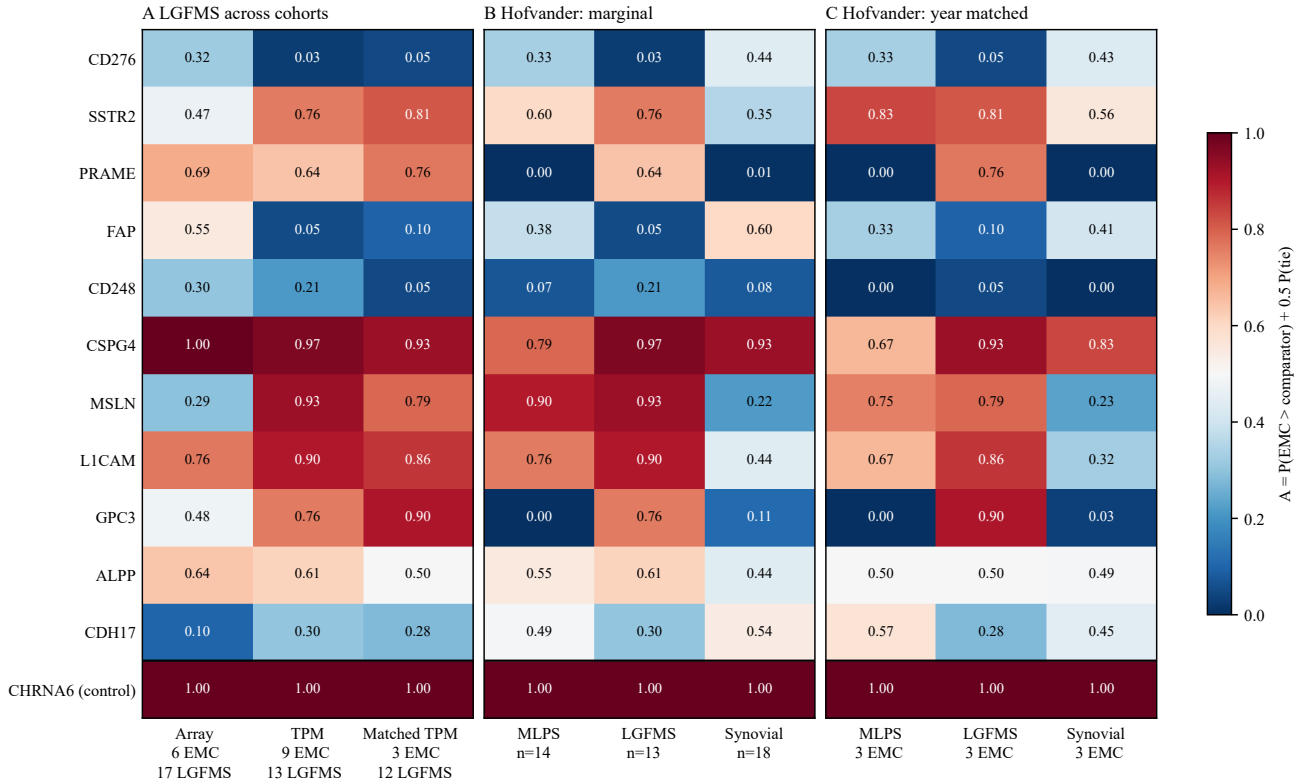


Figure 1. Fixed-panel probability of superiority by comparator and cohort. Rows retain all 11 address genes plus the separately marked CHRNA6 control. Panel A shows the same LGFMS comparison in original arrays (6 EMC, 17 LGFMS), Hofvander marginal (9 EMC, 13 LGFMS), and Hofvander year-matched (3 EMC, 12 LGFMS across 2019/2021). Panels B and C show all three primary Hofvander histologies separately, marginally and within supported years. Color and printed values encode A on 0–1 with a neutral midpoint 0.5; there are no significance stars. The matched columns do not represent nine independent matched patients.

PRAME and L1CAM had positive LGFMS directions in both datasets but failed broad prioritization because other Hofvander histologies had lower or reversed contrasts. For PRAME, marginal A was 0 versus MLPS and 0.00617 versus synovial sarcoma. L1CAM's marginal composite was 0.69931, near but below 0.70; importantly, its synovial comparison was independently nonpositive (0.43827 marginal; 0.32500 matched), so its failure did not depend solely on a close numerical threshold.

MSLN, SSTR2, GPC3 and FAP had opposed GSE24369-versus-Hofvander LGFMS directions. CD276, CD248 and CDH17 were below 0.5 against LGFMS in both cohorts. ALPP had positive marginal LGFMS ordering but matched

A=0.5. These contrasts are retained rather than interpreted through the most favorable histology.

CSPG4 is consistent across named cohorts but sensitive to sequencing year

Against LGFMS, CSPG4 A was 1.000 in the arrays and 0.96581 marginal/0.93333 matched in Hofvander. Every single array EMC or LGFMS biopsy deletion retained A=1. Secondary shared contexts also had positive CSPG4 directions: array A=1 versus each of MFS, SFT and desmoid; Hofvander marginal/matched values were 0.85370/0.85714, 0.87879/0.62500 and 1/1 respectively. These comparisons support consistent ordering against those particular histologies, not universal specificity.

Within Hofvander, single-EMC deletion composite ranges were 0.88136–0.94035 marginal and 0.71667–0.94444 matched. Single-histology deletion ranges were 0.85891–0.94896 and 0.75000–0.88333. However, deleting 2019 lowered the matched composite to 0.43333 while the marginal value remained 0.83333 (Figure 2). The apparently positive matched summary therefore depends on a small set of supported year cells; it cannot be called batch-robust. The conditional bootstrap does not resolve that limitation. Revised diagnoses gave similar composites (0.89475 marginal; 0.81111 matched), but partly expression-informed revisions are not independent confirmation.

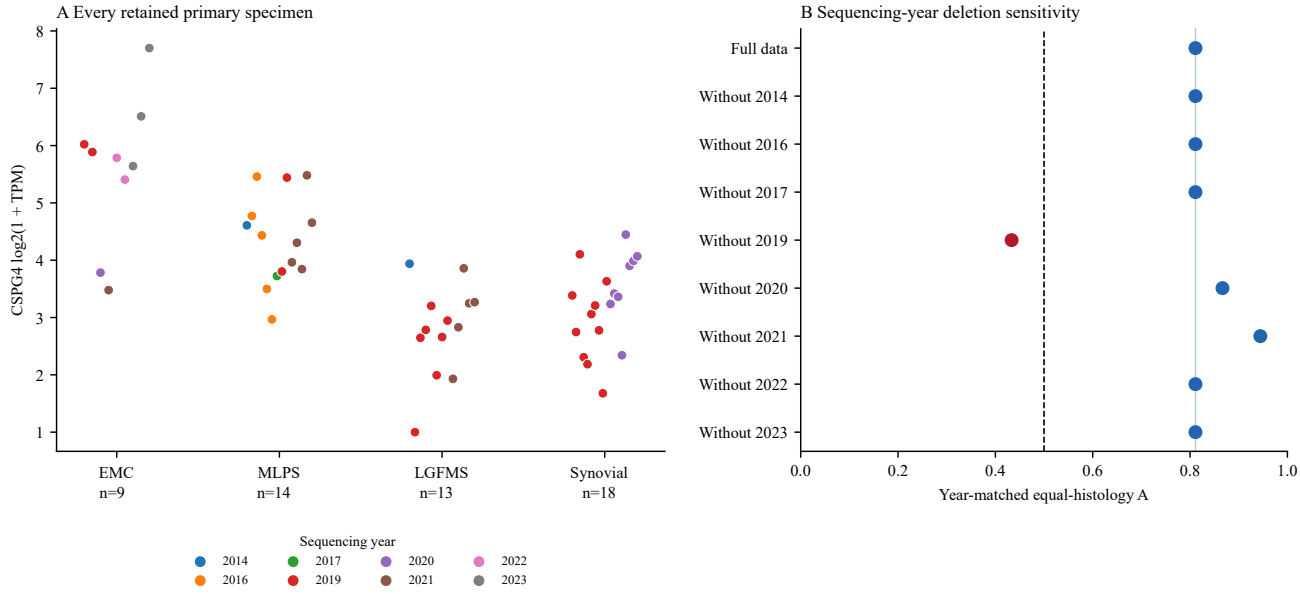


Figure 2. CSPG4 specimens and sequencing-year sensitivity. Panel A shows each retained primary specimen's log2(1+TPM), grouped by EMC and the three primary comparator histologies; color denotes sequencing year. The display transformation does not enter rank calculations. Panel B shows the matched equal-histology summary with each year deleted, beside the full-data value and neutral 0.5 reference. The 2019 deletion reverses the summary. Points are descriptive observed/deletion values, not confidence limits; no normal-tissue or protein measurements are plotted.

CSPG4 was not higher than DFSP marginally (A=0.46667). The matched DFSP comparison was especially sparse, with one EMC and three DFSP specimens in 2021. This contrary histology was not included in the primary success rule and remains consequential context.

RNA priority does not establish protein availability or normal sparing

The retrieved HPA CSPG4 record describes broad normal cytoplasmic IHC, whereas ICC/IF includes membrane localization. Different assay compartments cannot establish EMC cell-surface density or normal sparing. Historical GSE28866 CSPG4 peak medians were broadly positive, yet a normal-colon record exceeded the lowest EMC library value. That observation uses a different estimand, but directly argues against treating positive group medians as universal individual separation. Historical CHRNA6 peak signal was nonuniform; it neither invalidates the published RNA CISH assay nor implies absence of the whole transcript.

Independent protein evidence further limits RNA-only extrapolation. Cammareri et al. assessed PRAME clone QR005 on whole sections from 350 soft-tissue tumors and mimics.[7] Their original supplementary rows 128–132 contain five EMC cases, all negative in both reader columns; four have recorded fusion/rearrangement support and one has no recorded ancillary test. Negative meant 0% positive cells. PRAME expression was common in MLPS and synovial sarcoma in that study. Positive PRAME ranks relative to LGFMS therefore do not establish PRAME protein positivity or an EMC therapeutic address. These protein observations come from a separate cohort and assay, not a paired RNA–protein experiment.

Discussion

The result identifies a qualified next tissue-validation question: whether CSPG4 protein is present in the malignant compartment and accessible in EMC specimens spanning relevant technical and biological variation. This question follows from positive same-histology RNA ordering in both cohorts, but the matched 2019-deletion reversal and DFSP direction prevent a general EMC-selectivity claim. Validation would need to resolve cellular source, localization and normal-context overlap, rather than merely reproduce a pooled RNA rank.

The fixed panel also produces useful negative evidence. Some transcripts separate EMC from LGFMS while failing against MLPS or synovial sarcoma; others change direction across platforms. Preserving these results avoids presenting the preferred comparator as if it represented sarcoma generally. CHRNA6 recapitulates a previously supported RNA marker and is deliberately not counted as a new target. In particular, neither its established CISH performance nor the present rank separation demonstrates protein immunoreactivity or therapeutic accessibility.

Several limitations remain. The source cohort was a convenience sample, and sequencing year is an incomplete proxy for library chemistry, RNA quality and processing batch. Restricting to supported years substantially reduces the EMC sample. The array biopsy lesion stage is unknown, and exclusion of explicit old cases cannot prove absence of every historical overlap. Bulk tissue admixture can produce differences unrelated to malignant-cell expression; EMC purity and grade were unavailable for adjustment. TPM and RMA signals have different measurement properties despite a common rank estimand. Normal tissues were not matched to tumors, HPA includes incomplete or uncertain protein evidence, and the two muscle pools cannot establish organ safety. Finally, the prioritization threshold is an allocation rule rather than a clinically validated effect size or a statistical discovery procedure.

The present evidence supports testing CSPG4 localization in tissue while retaining an explicitly year-sensitive interpretation. It does not establish a therapeutic target, an independently validated diagnostic test, a normal-sparing intervention or efficacy. The complete fixed-panel result—including comparator failures and normal-context limits—is the scientific output.

Data, code and declarations

Data availability. Original source data are publicly available from the Hofvander author repository and Zenodo record (doi:10.5281/zenodo.17866629), GEO GSE24369/GPL6244, GEO GSE28866 and the versioned HPA entries cited here. This study's supplementary package includes exact source hashes and specimen/probe mappings, the frozen protocols, full-panel effect tables, analysis code and verification receipts. The offline replay accepts an explicit source-bundle directory and reproduces the historical results with the pinned runtime. No new deposit DOI or public availability of an unreleased analysis package is claimed. Source-specific attribution and reuse conditions are retained; HPA database licensing is not assigned to source articles or third-party data.

Author and correspondence. Tristan D. McRae, independent researcher, unaffiliated, is the corresponding author at trimcrae@gmail.com; ORCID 0000-0002-1823-1451. These details match the established author record used in the

repository's ASO manuscript.

Research scope and declarations. This work reanalyzes public data and reports no newly collected participant specimens, intervention, protein experiment or clinical treatment. No new ethics approval, exemption or waiver is asserted.

Funding. This research received no funding.

Competing interests. The author declares no competing interests.

References

1. Dulken BW et al. CHRNA6 RNA In Situ Hybridization Is a Useful Tool for the Diagnosis of Extraskeletal Myxoid Chondrosarcoma. Modern Pathology. 2024; 37:100464. https://doi.org/10.1016/j.modpat.2024.100464
2. Hofvander et al. Transcriptomic Subgroups in Soft Tissue Tumors Correlate with Morphologic Subtype, Genomic Features, and Outcome. Clinical Cancer Research. 2026; 32:1825–1834. https://doi.org/10.1158/1078-0432.CCR-25-3740 ; data https://doi.org/10.5281/zenodo.17866629
3. NCBI GEO. GSE24369 and GPL6244, original series and transcript-cluster annotation. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE24369 ; https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GPL6244
4. Uhlén M et al. Tissue-based map of the human proteome. Science. 2015; 347:1260419. https://doi.org/10.1126/science.1260419 ; HPA versioned CSPG4 entry linked in Table S4.
5. Thul PJ et al. A subcellular map of the human proteome. Science. 2017; 356:eaal3321. https://doi.org/10.1126/science.aal3321
6. GSE28866 original normalized peak matrix and primary 3SEQ study. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE28866 ; https://doi.org/10.1186/gb-2012-13-8-r75
7. Cammareri et al. PRAME immunohistochemistry in soft tissue tumors and mimics: a study of 350 cases highlighting its imperfect specificity but potentially useful diagnostic applications. Virchows Archiv. 2023; 483:145–156. https://doi.org/10.1007/s00428-023-03606-6 ; original supplementary workbook, rows 128–132.

Supplementary Files

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

- emctissuernaprioritizationsupplementaryinformation.pdf
- emctissuernacodedatasupplement.zip
- priorpreprintandreusenote.txt