

Supplementary information: fixed-panel tissue RNA prioritization in extraskeletal myxoid chondrosarcoma

Supplementary Information to *Fixed-panel tissue RNA prioritization in extraskeletal myxoid chondrosarcoma: CSPG4 evidence across cohorts with comparator and sequencing-year limits* — accompanies [emc-tissue-rna-prioritization-preprint.pdf](#).

Research manuscript; not peer reviewed. · Version of 2026-09-06 · supplementary information · built from 8bb8c9d29

Source inventory and specimen scope

Source	Units and usable specimens	Assay / release	Role
Hofvander 2026	704 source patients; 9 primary EMC after fixed exclusions	19,116 gene symbols, gene TPM; author release v1.0.1	New-to-this-analysis tissue cohort
GSE24369	42 original arrays: 6 EMC, 17 LGFMS, 6 MFS, 6 desmoid, 5 SFT, 2 muscle pools	GPL6244 transcript-cluster RMA log2 signal	Retrospective same-histology anchor; also CHRNA6 discovery data
GSE28866	Four EMC library/STT records; individual patient count unverified	Selected 3SEQ peaks, supplied compressed normalized scale	Historical, different-estimand sample/organ context only
HPA	12 gene XML/JSON pairs; tissue/cell records are not independent subjects	XML version 25; current methods/download 25.1, Ensembl 109	Descriptive normal RNA, IHC and ICC/IF context
Cammareri 2023	Five EMC rows in original supplementary workbook	Whole-section PRAME QR005 IHC; both reader columns negative	External negative protein evidence, not paired with present RNA

Nine retained Hofvander EMC IDs: 11881-19, 3371-22, 3372-22, 4716-22, 4840-13, 5149-18, 5241-06, 7931-19, 8102-22. Excluded previously reported cases 104-92,168-97 and 536-00 map to source MDB9736 cases 3, 4 and 7; 5081-14 is a local recurrence. Original Table S1 diagnosis and nonblank specimen exception are preserved for every 704-source sample. No patient date is inferred from an accession suffix. Blank publication references do not prove independence. Gene-fusion positive in this source can mean current RNA-sequencing or an earlier assay, not necessarily a known partner in every case.

Original-diagnosis primary comparators contain 14 MLPS, 13 LGFMS, 18 synovial sarcomas; separate contexts contain 60 MFS and 10 DFSP. Shared-histology secondary contexts add 11 primary SFT and 11 desmoids. Source metadata sometimes uses revised or transcriptomic labels; original labels are primary and the separately reported revision sensitivity changes all relevant rows symmetrically. Purity and grade are unavailable for EMC adjustment.

Sequencing years for the nine EMC are 2019(n=2), 2020(n=1), 2021(n=1), 2022(n=2), 2023(n=3). MLPS has 2 comparators in 2019 and 5 in 2021; LGFMS has 7 and 5; synovial sarcoma has 10 in 2019 and 8 in 2020. Consequently, each primary matched contrast contains 3 EMC and the union is 4 EMC. Context DFSP has only 1 matched EMC and 3 comparators in 2021. Unsupported year cells remain null. All counts refer to the exact specimen policy, not the full published histology totals.

Exact measurement contract

TPM source SHA256: b0d665d1bd1d96ace1faf66cc5a4d7ab7e41cb487c8f0f61734f102a1f9a7af3. Original GSE24369 family gzip SHA256: 98c83c8ca23b7052cf0d4d0099a7bf1af6c3c972276038c3a633e2a5349b3c37. Full source digests and byte sizes are in the two frozen metadata manifests. Analysis uses released numeric precision, not the earlier rounded cache.

Gene	GPL6244 probe	Role
CD276	7984743	Address
SSTR2	8009526	Address
PRAME	8074856	Address
FAP	8056257	Address
CD248	7949588	Address
CSPG4	7990545	Address
MSLN	7992071	Address
L1CAM	8175871	Address
GPC3	8175234	Address
ALPP	8049123	Address
CDH17	8151795	Address
CHRNA6	8150550	Separate context control

All transcript-cluster assignments for each selected probe resolve to the same gene. No probe mean, best probe, alias substitution or expression-dependent feature filter is used. Array finite signed log signals are valid; TPM must be finite and nonnegative. Zero observations remain ties. Missing/duplicate observations cause a recorded technical failure rather than deletion or imputation. Actual runs completed without these failures.

For EMC values x_i and comparator values y_j , $A = (1/nm) \sum_i \sum_j [I(x_i > y_j) + 0.5I(x_i = y_j)]$. Pair counts are not independent-patient counts. The marginal summary is the mean of the 3 histology effects. Matched effects first average within exact year, weighting by supported EMC counts, then average equally across histologies. A common EMC specimen can contribute to multiple histology contrasts; bootstrap sampling preserves that dependence by sharing its resample. No cross-platform scale or patient pooling is performed.

Bootstrap seed 20260906, 2000 replicates, stratified by histology/year. The conditional 95% percentile intervals are pointwise, not familywise and not used to select genes. For CSPG4, a degenerate MLPS matched interval [0.66667, 0.66667] reflects observed sparse strata and the resampling design, not absence of biological uncertainty. All singleton flags and patient placements are supplied. Independent 5000-draw resampling reproduced the main matched interval endpoints; this numerical check does not remove small-sample limitations.

The allocation rule requires marginal composite ≥ 0.70 , each primary marginal/matched histology > 0.5 , and both summaries > 0.5 after each single EMC and histology deletion. Year/comparator deletion and revised diagnoses are qualifiers. Deletions renormalize retained supported-year weights; a lost contrast remains undefined. Complete-case selection or imputation is not performed. No new replication success cutoff was added. The separate cross-cohort output compares signs and magnitudes against LGFMS primarily, MFS/SFT/desmoid secondarily.

Tables and evidence files

Table S1, `../../../../autonomy/atlas-hofvander-validation-2026-09-06/all12-gene-effects.csv`, includes every gene, both primary composites, conditional intervals and same-LGFMS effects. Table S2, `../../../../autonomy/atlas-hofvander-validation-2026-09-06/all-hofvander-contrasts.csv`, contains 60 gene $\times$ histology rows; `../../../../autonomy/atlas-hofvander-validation-2026-09-06/all-year-cells.csv` provides all supported/unsupported cells. Table S3, `../../../../autonomy/atlas-hofvander-validation-2026-09-06/all-shared-histology-replication.csv`, contains 48 gene $\times$ shared-histology comparisons and individual-deletion ranges. `../../../../autonomy/atlas-hofvander-validation-2026-09-06/all-primary-deletions.csv` retains every summary deletion. Full JSON adds every specimen value, placement, individual contrast and revised-label mapping.

Table S4, `../autonomy/atlas-hofvander-validation-2026-09-06/draft/normal-context.csv`, contains 12 source-linked HPA records with normal IHC reliability/missingness and assay-location context. It is a descriptive extract, not a normal-sparing assessment. Source data include 1092 RNA tissue/block records and 1054 categorical normal-IHC cell records; neither total is a count of newly measured independent donors. HPA consensus RNA summarizes maxima across HPA/GTEX sources/subtissues. GPC3/CHRNA6 lack tissue-IHC rows in this recovery; absence of a row is not a negative stain. SSTR2/FAP uncertain IHC and ALPP multi-gene antibody warning remain explicit. PRAME's localization tags are not evidence of endogenous EMC peptide-HLA presentation.

PRAME supplementary workbook SHA256:

```
20b4bd1243be22e4eee2da7470da986091f91cc4959fa416708748ba99187ca7
```

Sheet 1 rows 128–132 contain five EMC, negative in both reader columns; blank percentage fields were not converted to measurements. Case 2 has EWSR1::NR4A3 RNA-sequencing support, cases 3–5 NR4A3-positive FISH, case 1 no recorded ancillary method. This supports a negative result in those 5 specimens, not universal PRAME absence or absent peptide-HLA presentation.

Reproduction and figure status

Empirical programs are `../autonomy/atlas-hofvander-validation-2026-09-06/analyze.py` and `../autonomy/atlas-hofvander-validation-2026-09-06/replication.py`; all 8 pre-outcome approved files are listed in `../autonomy/atlas-hofvander-validation-2026-09-06/coordinator-authorization.json`. Original scientific protocol bytes are preserved alongside frontmatter-corrected versions. Each analysis ran once: command elapsed 46.38 and 12.72 seconds, exit 0. Runtime was Python 3.12.14, NumPy 2.3.5, openpyxl 3.1.5; exact executable/version receipt is in `../autonomy/atlas-hofvander-validation-2026-09-06/draft/runtime.json`. Independent original-source arithmetic reproduced 8,448 TPM values, 504 array values, 2,436 estimate/deletion blocks and 266,772 scalar comparisons without discrepancy. These checks verify computation, not causal or clinical validity.

Source locations in the original manifests remain historical workstation paths. The portable offline wrapper accepts an explicit bundle root, stages exact archive/gzip members and changes only `source_location` in derived manifests. One fresh-directory replay on the same Windows host completed with 829,482 exact scalar comparisons and zero discrepancies across 29 scientific JSON files and five CSV exports; two execution records separately matched status and final stage. This is not a second-machine or second-OS claim. Python 3.12.14, NumPy 2.3.5, openpyxl 3.1.5 and et-xmlfile 2.0.0 are pinned. Historical approval is preserved, while the compatibility record explicitly identifies the mechanical replay. Installation instructions suffice; no bundled runtime or wheels are claimed.

The canonical `plot_emc_tissue_rna.py` adapter preserves the frozen draft plotting statistics and geometry, adds SVG output for the manuscript renderer alongside vector PDF and 300-dpi PNG, and rejects rendering until exact source and plotting-script bytes are committed. Figures remain descriptive; no uncertainty bars or new statistical tests are introduced. Normal-expression context remains Table S4. Figure rendering and visual inspection are recorded separately from numerical replay.

Author metadata were transcribed from `../aso/fusion-junction-aso-journal-article.md`: the Author, affiliation/correspondence and ORCID block. No study-specific funding, conflicts or ethics statements were inferred from that separate article. OpenAI Codex assisted with implementation, source organization, verification and drafting; this assistance is disclosed in the main article.

The historical analysis, draft and replay packets are preserved unchanged. Canonical manuscript routing and the SVG plotting adapter are separate presentation files; no scientific decision or source membership changed.

Table S1. All fixed-panel effects

A=0.5 is neutral ordering; the control is separate from the address pass count. Matched columns use the supported-year specimens, not all nine EMC. Intervals and complete sensitivities are retained in the machine-readable tables described above.

Gene	Hofvander marginal A	Hofvander matched A	LGFMS array A	LGFMS Hofvander A	Rule
CD276	0.26575	0.27143	0.32353	0.02564	Fail
SSTR2	0.56926	0.73373	0.47059	0.76068	Fail
PRAME	0.21573	0.25397	0.68627	0.64103	Fail
FAP	0.34367	0.27897	0.54902	0.05128	Fail
CD248	0.12178	0.01587	0.30392	0.21368	Fail
CSPG4	0.89454	0.81111	1.00000	0.96581	Pass
MSLN	0.68008	0.58968	0.29412	0.92735	Fail
L1CAM	0.69931	0.61627	0.76471	0.90171	Fail
GPC3	0.29060	0.31270	0.48039	0.76068	Fail
ALPP	0.53336	0.49722	0.63725	0.61111	Fail
CDH17	0.44520	0.43115	0.09804	0.30342	Fail
CHRNA6	1.00000	1.00000	1.00000	1.00000	Control only