

Additional file 1: Supplementary Figures

Journal: Genomics & Informatics (Korea Genome Organization)

Manuscript: Single-cell mapping of the endocannabinoid system in breast cancer reveals distributed expression and ECS–NK correlations that track immune composition

Authors: Ekaterina Fudina¹, Sungwon Jeon², Hyoungjin Choi¹, Juok Cho¹

Affiliations: ¹ GenomeLab, Biomedical Engineering, UNIST, Ulsan 44919, South Korea; ² AgingLab, Inc., Ulsan 44919, South Korea

Corresponding authors: Ekaterina Fudina (fudinaekaterina@unist.ac.kr), Juok Cho (juokcho2025@unist.ac.kr)

Date: 2026-09-16

This file (Additional file 1) contains Supplementary Figures S1 and S2. Supplementary Tables S0 to S15 are provided separately as Additional file 2 (Excel workbook); the REMARK checklist is Additional file 3.

Figure S1

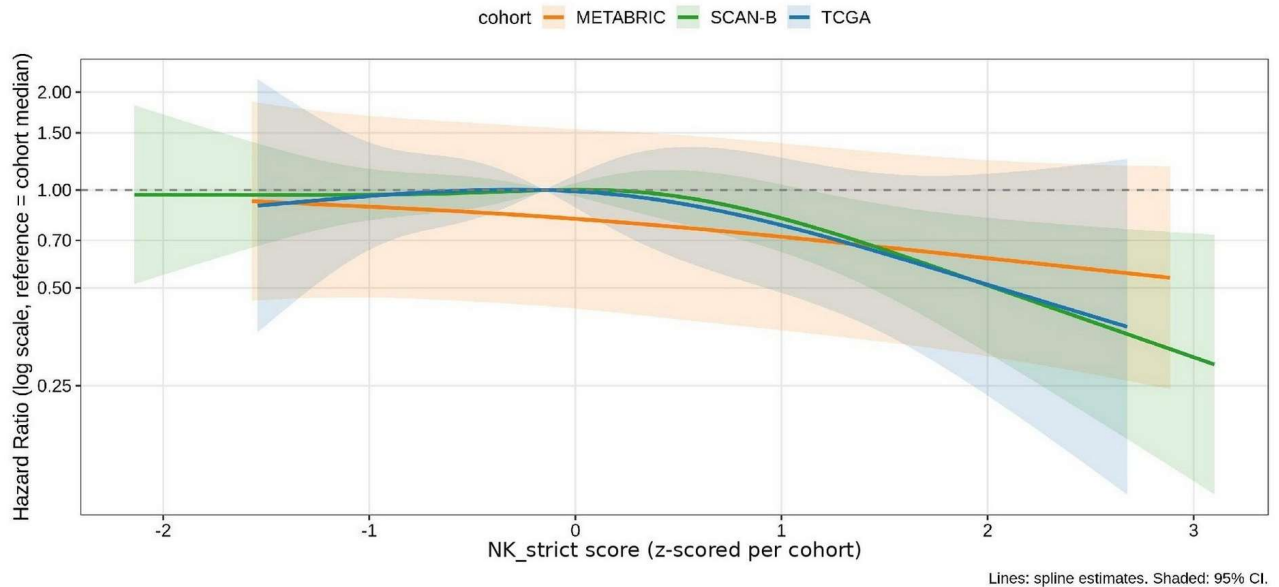


Figure S1. Restricted cubic spline functional form for the NK_strict signature to survival relationship across three cohorts. Restricted cubic splines with 4 knots at Harrell-recommended percentiles (5%, 35%, 65%, 95%) were fitted for the continuous NK_strict score against overall survival in Model 3 (TCGA: age, PAM50, AJCC stage; METABRIC: age, PAM50, stage, grade; SCAN-B: age, PAM50, tumor size, lymph-node status, Nottingham grade). Analysis samples are those of the main Cox models (TCGA $n = 752/88$ events; METABRIC $n = 1,305/735$; SCAN-B $n = 2,818/291$), and the linear-term hazard ratios reproduce Table S6b exactly (0.873, 0.865, 0.845). Lines show spline estimates on a log hazard-ratio scale (reference = cohort mean score); shaded regions represent 95% pointwise confidence intervals. The null hypothesis of linearity was tested by a Wald test on the non-linear terms of the spline. Non-linearity Wald p-values by cohort: TCGA $p = 0.602$; METABRIC $p = 0.473$; SCAN-B $p = 0.146$. None of the three cohorts shows statistical evidence for departure from linearity. The functional form is monotonically decreasing across the observed score range with confidence intervals expanding at the tails due to sparse data. In all three cohorts, hazard ratios below 1 are observed for above-mean NK_strict scores, consistent with the protective association reported in the main Cox analysis. This figure supports the linearity assumption underlying the multivariate Cox proportional hazards analysis. Combined with the Schoenfeld residual test results (Table S8), the linear z-scored parameterization is justified.

Figure S2

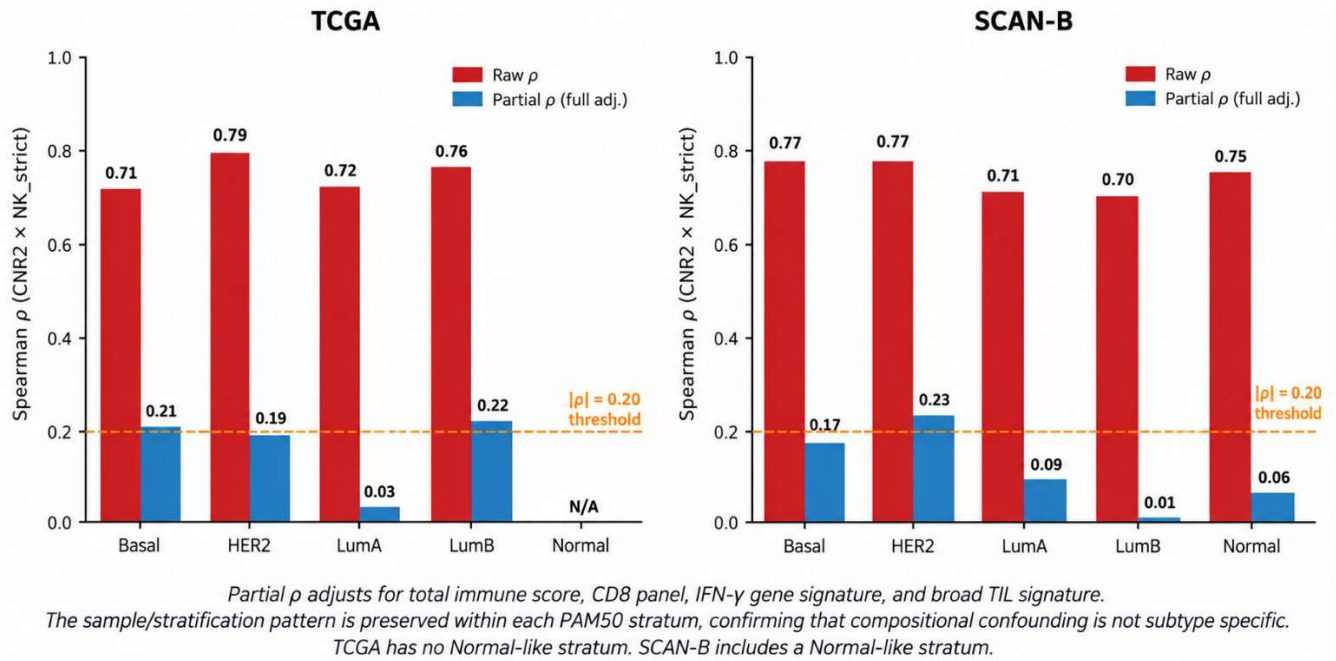


Figure S2. Per-PAM50 intrinsic-subtype stratified raw versus partial Spearman correlation of CNR2 with NK_strict signature score. Grouped bar charts showing raw Spearman ρ (red) and full-adjustment partial Spearman ρ (blue) for $\text{CNR2} \times \text{NK_strict}$ within each available PAM50 intrinsic subtype, computed separately in TCGA-BRCA (left; Basal, HER2-enriched, LumA, LumB; no Normal-like stratum is available in this cohort) and SCAN-B (right; Basal, HER2-enriched, LumA, LumB, Normal-like). Partial correlation adjusts for the 130-gene curated Immune Score (124 to 130 genes available per cohort; Table S15), CD8 T-cell panel (21 genes), IFN- γ 6-gene signature (Ayers), and broad tumor-infiltrating lymphocyte (TIL) signature (50 genes). METABRIC is omitted from this figure because the raw $\text{CNR2} \times \text{NK_strict}$ correlation in that cohort is small in absolute value ($\rho = 0.12$), plausibly reflecting reduced dynamic range of CNR2 detection on the Illumina HT-12 microarray platform relative to RNA-seq (Discussion). The raw to partial ρ collapse pattern observed at the whole-cohort level (Table S7b) is preserved within every PAM50 stratum. In TCGA, raw ρ ranged 0.710 to 0.793 across the four available subtypes while partial ρ ranged 0.033 to 0.216. In SCAN-B, raw ρ ranged 0.702 to 0.772 while partial ρ ranged 0.006 to 0.231. No PAM50 subtype in either cohort retains partial $\rho \geq 0.20$ as a consistent finding across cohorts; the three strata with partial $\rho \geq 0.20$ (Basal in TCGA at 0.205, LumB in TCGA at 0.216, and HER2-enriched in SCAN-B at 0.231) are isolated observations within a broader pattern of collapse from raw values above 0.70. This subtype-stratified analysis rules out the alternative interpretation that the compositional confounding observed at the whole-cohort level is driven by aggregation across heterogeneous PAM50 subtypes. If the raw ECS to NK correlation were driven by subtype-specific biology (for example, if Basal tumors both preferentially expressed CB2 in NK cells and had preferentially high NK infiltration), then within-Basal partial correlation should be preserved. The observation that partial ρ collapses across every available PAM50 stratum, four in TCGA and five in SCAN-B, indicates that the attenuation is a general property of immune-rich tumor biology, not a subtype-driven artifact. The distributed cell-type expression pattern of ECS components (Table 1) is consistent with this interpretation: because CB2 expression in bulk is dominated by co-infiltrating plasma cells and multiple lymphoid populations, any tumor stratum with high general immune infiltration will show elevated bulk CB2 signal irrespective of NK-specific CB2 biology.