

Online Resource 3

Per-Dataset Correlation Value Distributions and Rank-Based Threshold Comparability Analysis

A multi-cohort computational framework for detection and prognostic validation of conserved gene co-expression network dissolution in solid tumors

Marc Ríos-Cadenas¹, Iván Segura-Carmona¹, Aurelio López-Fernández¹, Francisco A. Gómez-Vela^{1,*}

¹SynergIA Laboratory (SIALAB), Universidad Pablo de Olavide, Ctra. Utrera km. 1, ES-41013 Seville, Spain.

*Corresponding author: fgomez@upo.es

Network Modeling Analysis in Health Informatics and Bioinformatics

Overview

This Online Resource reports the full per-dataset distribution of pairwise Pearson correlation values (r_N : normal/control tissue; r_T : tumor tissue) underlying the double-threshold filtering strategy described in the main manuscript (Section 3.1.3). Correlation matrices were computed at the FC2 threshold from DESeq2 size-factor (median-of-ratios) normalized counts; the variance-stabilizing transformation (VST) described in the main manuscript was used exclusively for sample-level quality control (outlier detection) and was not used as input to correlation network inference. Values were extracted from the upper triangle of each gene×gene correlation matrix for all eleven subdatasets (1.7–6.3 million gene pairs per dataset, depending on DEG-set size). Table S1 and Table S2 report summary statistics and percentiles of r_N and r_T , respectively. Table S3 quantifies the percentile rank of the fixed control threshold ($r_N \geq 0.7$) within each cohort's own distribution, directly addressing whether an identical absolute threshold corresponds to an equivalent proportion of gene pairs across cohorts. Figure S1 shows the corresponding density plots.

Table S1. Distribution of Pairwise Correlation Values — Normal (Control) Tissue

Summary statistics and percentiles of pairwise Pearson correlation values (r_N) in normal/control tissue, FC2, all DEG–DEG gene pairs.

Dataset	n pairs	Mean	SD	P25	P50	P75	P90	P95	P99	P99.9
LUAD Upper lobe	4,054,128	0.022	0.200	-0.112	-0.003	0.130	0.279	0.386	0.633	0.907
LUAD Lower lobe	2,745,996	0.021	0.219	-0.127	-0.009	0.138	0.303	0.422	0.709	0.948
HNSC Larynx	1,740,045	0.057	0.283	-0.130	-0.020	0.181	0.458	0.682	0.931	0.978
HNSC Oral cavity	3,483,480	0.049	0.271	-0.145	0.003	0.211	0.433	0.570	0.800	0.953
HNSC Oropharynx	3,407,355	0.045	0.277	-0.153	0.003	0.210	0.429	0.571	0.819	0.969
STAD CIN	6,253,416	0.051	0.338	-0.187	-0.036	0.247	0.553	0.757	0.961	0.993
STAD MSI	3,396,921	0.034	0.393	-0.257	-0.090	0.301	0.679	0.813	0.926	0.973
BRCA Basal	3,483,480	0.012	0.262	-0.144	-0.013	0.147	0.351	0.505	0.772	0.950
BRCA Her2	3,420,420	0.010	0.285	-0.177	-0.022	0.168	0.398	0.550	0.790	0.938
BRCA LumA	2,049,300	0.012	0.202	-0.113	-0.017	0.107	0.263	0.389	0.679	0.896
BRCA LumB	3,890,655	0.009	0.283	-0.170	-0.020	0.165	0.391	0.546	0.785	0.936

Table S2. Distribution of Pairwise Correlation Values — Tumor Tissue

Summary statistics and percentiles of pairwise Pearson correlation values (r_T) in tumor tissue, FC2, all DEG–DEG gene pairs.

Dataset	n pairs	Mean	SD	P25	P50	P75	P90	P95	P99	P99.9
LUAD Upper lobe	4,054,128	0.013	0.130	-0.064	-0.013	0.066	0.169	0.249	0.476	0.818
LUAD Lower lobe	2,745,996	0.017	0.165	-0.087	-0.018	0.089	0.225	0.328	0.603	0.864
HNSC Larynx	1,740,045	0.022	0.174	-0.086	-0.020	0.088	0.229	0.343	0.719	0.968
HNSC Oral cavity	3,483,480	0.025	0.132	-0.058	-0.001	0.085	0.193	0.273	0.465	0.753
HNSC Oropharynx	3,407,355	0.023	0.193	-0.106	-0.019	0.118	0.283	0.400	0.646	0.873
STAD CIN	6,253,416	0.051	0.175	-0.056	-0.001	0.103	0.266	0.409	0.749	0.937
STAD MSI	3,396,921	0.068	0.211	-0.080	0.024	0.179	0.363	0.484	0.703	0.878
BRCA Basal	3,483,480	0.013	0.123	-0.059	-0.012	0.061	0.162	0.243	0.448	0.755
BRCA Her2	3,420,420	0.016	0.146	-0.077	-0.013	0.083	0.204	0.295	0.501	0.750
BRCA LumA	2,049,300	0.026	0.132	-0.049	-0.008	0.065	0.190	0.296	0.523	0.760
BRCA LumB	3,890,655	0.026	0.124	-0.047	-0.003	0.069	0.174	0.265	0.489	0.749

Table S3. Cross-Cohort Comparability of the Fixed Correlation Threshold

For each dataset, the percentile rank of $r_N = 0.7$ (the fixed control-tissue threshold used throughout the main pipeline) was computed within that dataset’s own r_N distribution (Table S1), by linear interpolation between adjacent percentile points. The ratio SD_{r_T}/SD_{r_N} quantifies the relative dispersion of the tumor-tissue distribution compared to the normal-tissue distribution within the same dataset.

Percentile rank of the fixed threshold ($r_N \geq 0.7$) within each dataset’s control-tissue correlation distribution, and the tumor-to-normal standard deviation ratio.

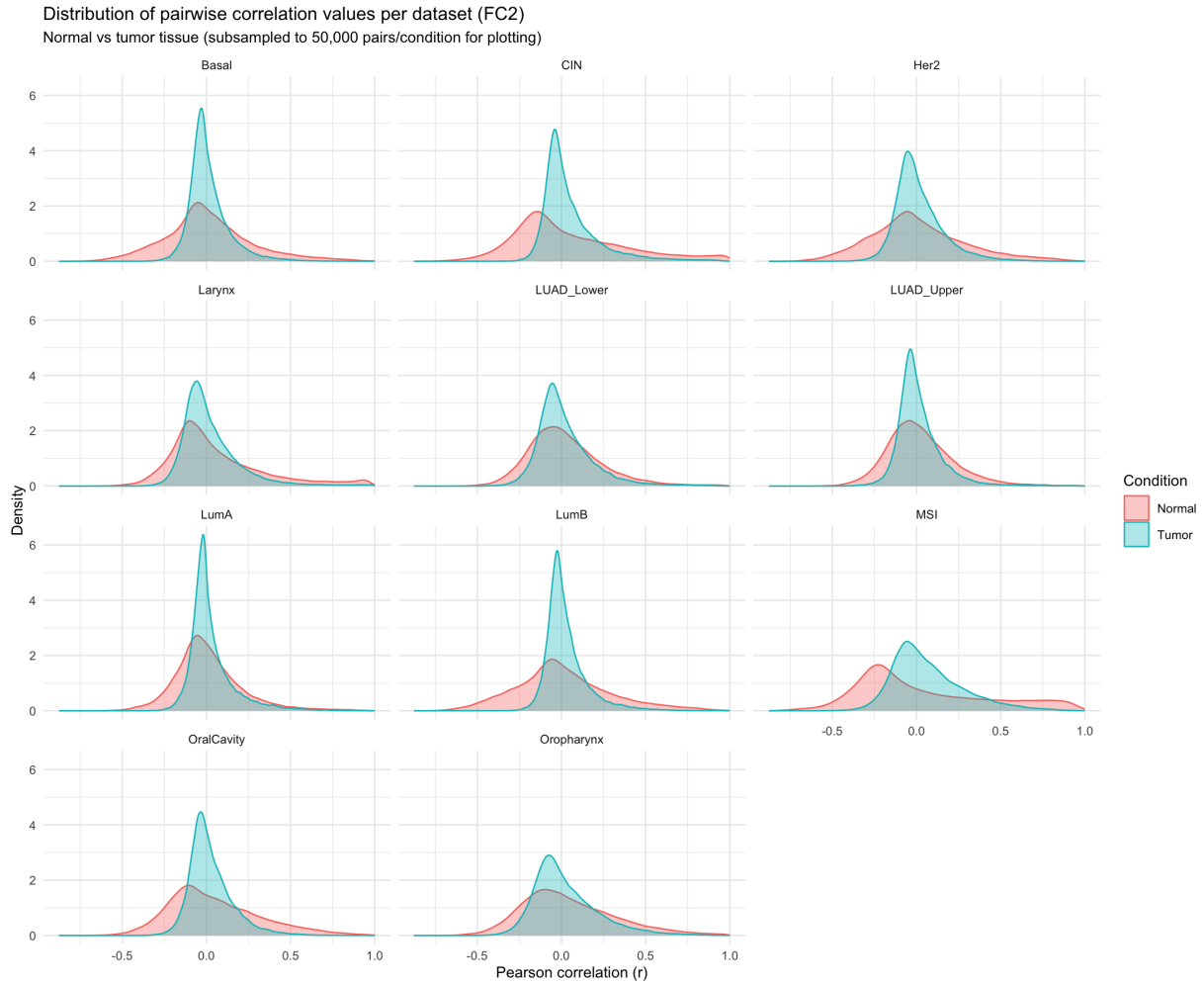
Dataset	Percentile of $r_N=0.7$	SD r_N	SD r_T	Ratio SD_{r_T}/SD_{r_N}
LUAD Upper lobe	99.2	0.200	0.130	0.651
LUAD Lower lobe	98.9	0.219	0.165	0.757
HNSC Larynx	95.3	0.283	0.174	0.613
HNSC Oral cavity	97.3	0.271	0.132	0.486
HNSC Oropharynx	97.1	0.277	0.193	0.696
STAD CIN	93.6	0.338	0.175	0.517
STAD MSI	90.8	0.393	0.211	0.536
BRCA Basal	97.9	0.262	0.123	0.470
BRCA Her2	97.5	0.285	0.146	0.514
BRCA LumA	99.1	0.202	0.132	0.656
BRCA LumB	97.6	0.283	0.124	0.437

Rows highlighted in yellow mark the datasets with the most extreme percentile values (STAD MSI, lowest; BRCA LumA, among the highest), illustrating the range (90.8–99.2 percentile) spanned by the same nominal threshold across cohorts.

Figure S1. Density Distributions of Pairwise Correlation Values by Dataset

Interpretation

Three observations follow from this analysis. First, pairwise correlation distributions are unimodal and right-skewed in both tissue states across all eleven subdatasets, with no evidence of pathological multimodality or degenerate behavior attributable to the absence of log-transformation prior to correlation estimation. Second, the correlation magnitude corresponding to the fixed control threshold ($r_N \geq 0.7$) is not an equivalent statistical event across cohorts: it ranges from the 90.8th percentile (STAD MSI) to the 99.2nd percentile (LUAD upper lobe) of the respective control-tissue distributions (Table S3). This confirms that, as with any fixed threshold, an identical absolute correlation value does not select an identical proportion of gene pairs in every cohort — the same limitation raised by the reviewer for fixed thresholds applies, in mirrored form, to rank-based thresholds (Online Resource 1, Table S2b), where the Δr value corresponding to a fixed percentile (e.g., top 1%) itself varies substantially across cohorts (0.513–0.861). Third, the standard deviation of tumor-tissue correlations is consistently smaller than that of normal-tissue correlations in all eleven datasets (ratio range 0.44–0.76), indicating a general compression of pairwise co-expression structure in the malignant state that is not restricted to the 18 conserved lost interactions reported in the main text.



Density of pairwise Pearson correlation values (r) per dataset, normal (red) versus tumor (teal) tissue, FC2. Subsamped to 50,000 pairs per condition per dataset for plotting; summary statistics in Tables S1–S2 are computed from the complete (non-subsamped) distributions.