

Online Resource 1

Threshold Sensitivity Analysis: Jaccard Similarity Matrix of Conserved Lost Interaction Sets

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 threshold sensitivity analysis for the double-threshold co-expression network filtering strategy described in the main manuscript (Section 3.1.4 and Results Section 4.3). Pairwise Jaccard similarity was computed between the conserved lost interaction sets produced by each of the five threshold configurations evaluated, using fold change 2 (FC2) results. The Jaccard index is defined as:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

where A and B are the conserved interaction sets for two threshold configurations. A value of 1 indicates identical sets and a value of 0 indicates no shared interactions.

Table S1. Pairwise Jaccard Similarity Matrix

Pairwise Jaccard similarity ($|A \cap B| / |A \cup B|$) between conserved lost interaction sets across five threshold configurations (FC2). The selected configuration (control $r \geq 0.7$, tumor $r \leq 0.3$, $n = 18$) is highlighted in green. Color coding: ■ $J \geq 0.50$ (high overlap); ■ $0.25 \leq J < 0.50$ (moderate); ■ $J < 0.25$ (low overlap); ■ diagonal (self-comparison).

Configuration	n	0.6/0.2	0.6/0.3	0.7/0.2	0.7/0.3 [†]	0.8/0.3
0.6 / 0.2	28	1.000	0.429	0.357	0.278	0.036
0.6 / 0.3	79*	0.429	1.000	0.182	0.250	0.024
0.7 / 0.2	10	0.357	0.182	1.000	0.556	0.100
0.7 / 0.3[†]	18	0.278	0.250	0.556	—	0.056
0.8 / 0.3	1	0.036	0.024	0.100	0.056	1.000

[†] Selected configuration (see main text, Section 4.3).

* $n = 79$ is a lower bound: only the top-15 interactions were available for the LUAD-STAD (47 total) and HNSC-STAD (20 total) comparisons at this threshold; the true Jaccard with 0.7/0.3 may be slightly higher than reported.

Table S2. Overlap with the Selected Configuration (0.7/0.3)

Number of interactions shared with the selected configuration (0.7/0.3, $n = 18$) for each alternative threshold (FC2).

Configuration	n (total)	Shared with 0.7/0.3	Union	Jaccard
0.6 / 0.2	28	10	36	0.278
0.6 / 0.3	79*	12	48	0.250*
0.7 / 0.2	10	10	18	0.556
0.7 / 0.3[†]	18	18	18	<i>Reference</i>
0.8 / 0.3	1	1	18	0.056
Top 1% (rank-based, Δr)	3	0	21	0.000
Top 5% (rank-based, Δr)	100	10	108	0.093
Top 10% (rank-based, Δr) [‡]	750	18	750	0.024

[†] Selected configuration.

* Lower bound; see note above.

[‡] All 18 reference interactions (0.7/0.3) are recovered under this alternative rank-based criterion; the reference set is fully nested within the top-10% consensus set (Union = n_{total}).

Table S2b. Per-Dataset Δr Threshold at the Top-1% Rank Cutoff

Dataset	Δr at top 1%
BRCA LumA	0.513
LUAD Upper lobe	0.572
BRCA LumB	0.686
BRCA Basal	0.690
LUAD Lower lobe	0.631
BRCA Her2	0.742
HNSC Oral cavity	0.774
HNSC Oropharynx	0.789
HNSC Larynx	0.801
STAD CIN	0.837
STAD MSI	0.861

The correlation magnitude corresponding to the top 1% of Δr values varies substantially across datasets (range 0.513–0.861), illustrating that rank-based thresholds are subject to the same cross-cohort comparability concern raised for fixed thresholds.

Table S3. Stable Core of Interactions Recovered Across All Configurations from 0.6/0.2 to 0.7/0.2

The following 10 gene pairs were consistently identified as conserved lost interactions in *all* threshold configurations from 0.6/0.2 to 0.7/0.2 (FC2), constituting the most robust subset of the primary 18-interaction set identified at 0.7/0.3.

Stable core of 10 interactions robust to threshold variation (FC2, configurations 0.6/0.2 through 0.7/0.2). Prognostic significance based on Kaplan–Meier log-rank test ($p < 0.05$) reported in the main manuscript (Table 6).

Gene 1	Gene 2	Biological function	Cancer pair	Δr	Prognostic
CST1	MMP11	ECM remodeling / proteolytic balance	BRCA–LUAD	0.799	Yes (HR = 2.01)
COL10A1	CST1	EMT / stromal architecture	BRCA–LUAD	0.762	Yes (HR = 1.93)
GRIN2D	ZIC2	Glutamate signaling / epithelial identity	HNSC–STAD	0.768	Yes (HR = 1.36)
AC010789.1	AL162413.1	lncRNA regulatory buffering	HNSC–STAD	0.739	Yes (HR = 1.34)
HOXC11	ZIC2	Developmental transcription factors	HNSC–STAD	0.927	No
TNXB	WNT2	Wnt signaling / connective tissue	HNSC–STAD	0.853	No
IGHV5-78	TNFSF11	Immunomodulation / RANKL signaling	LUAD–STAD	0.786	No
FCRL4	TNFSF11	B cell regulation / immune microenvironment	LUAD–STAD	0.753	No
IGHV1OR15-2	IGHV4-4	Immunoglobulin repertoire diversity	LUAD–STAD	0.577	No
IGKV3-11	JSRP1	Humoral response / calcium signaling	LUAD–STAD	0.640	No

Rows highlighted in green indicate interactions with significant prognostic associations in the corresponding cancer cohort (log-rank $p < 0.05$). HR: hazard ratio for the high co-expression tertile relative to the low co-expression reference group. Δr : mean delta-correlation across all cancer subtypes at the primary threshold (0.7/0.3).

Interpretation

The results of the sensitivity analysis can be summarized as follows. First, the number of conserved interactions decreases monotonically from 79 (threshold 0.6/0.3) to 1 (threshold 0.8/0.3), with no erratic jumps under parameter variation. Second, 10 of the 18 interactions selected at the primary threshold (0.7/0.3) are consistently recovered across all configurations from 0.6/0.2 to 0.7/0.2, including the four pairs with significant prognostic associations (CST1–MMP11, COL10A1–CST1, GRIN2D–ZIC2, and AC010789.1–AL162413.1). Third, the Jaccard similarity of 0.556 between the selected configuration and the adjacent stricter threshold (0.7/0.2) indicates moderate overlap; the lower values observed with more permissive thresholds (0.250–0.278) are attributable to the inclusion of lower-confidence interactions at reduced stringency.