

Characterizing and Evaluating Cell Specialization Through the Gini Index of Gene Expression: A TCGA Normal Vs. Tumor Case Study



Fabio Cumbo¹ and Daniele Santoni^{2,*}

¹Center for Computational Life Sciences, Lerner Research Institute, Cleveland Clinic, 9500 Euclid Avenue, Cleveland, 44195OH, USA

²Institute for Systems Analysis and Computer Science "Antonio Ruberti", National Research Council of Italy, Via dei Taurini 19, Rome, 00185 RM, Italy

© 2025 The Author(s). Published by Bentham Science.

This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Public License (CC-BY 4.0), a copy of which is available at: <https://creativecommons.org/licenses/by/4.0/legalcode>. This license permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.



* Address correspondence to this author at the Institute for Systems Analysis and Computer Science "Antonio Ruberti", National Research Council of Italy, Via dei Taurini 19, Rome, 00185 RM, Italy; E-mail: daniele.santoni@iasi.cnr.it

Published: May 29, 2025

Cite as: Cumbo F, Santoni D. Characterizing and Evaluating Cell Specialization Through the Gini Index of Gene Expression: A TCGA Normal Vs. Tumor Case Study. Open Bioinform J, 2025; 18: e18750362364938. <http://dx.doi.org/10.2174/0118750362364938250520114456>



Send Orders for Reprints to reprints@benthamscience.net

SUMMARY TABLE OF THE STATISTICAL ANALYSIS OF STANDARD DEVIATIONS

Table 1 below reports the results of the statistical analysis performed on standard deviations and based on z-

scores and Wilcoxon rank-sum test for each of the 17 tumor types considered in this study. Z-scores are computed considering the standard deviation of the actual paired samples and of those in the same 1,000 randomized gene expression profiles used for computing Gini index.

Table 1. The Gini Coefficients and standard deviations for both normal and tumor samples of each patient belonging to TCGA-BLCA are reported.

Patient ID	Gini Coefficient (normal)	Gini Coefficient (tumor)	Standard Deviation (normal)	Standard Deviation (tumoral)
A3BM	0.9264239871	0.9080490119	154.68883	96.43031
A2C5	0.9224576714	0.9282946197	151.95858	222.19769
A20Q	0.9359826374	0.9236195676	133.56378	149.16459
A0YR	0.9180891384	0.9344879804	109.83594	198.76369
A20U	0.9449703359	0.9117340418	195.59044	93.63012
A2LA	0.9336658289	0.8897754671	123.53428	67.89630
A0YN	0.9269895324	0.9240204421	105.14391	163.99434
A13J	0.9308255413	0.9161325754	122.80717	92.53150
A54R	0.9260035463	0.9145792187	135.71554	87.97194
A6I3	0.9281541403	0.9340961722	127.75173	151.45745
A2LB	0.9458646358	0.9209801424	209.29742	134.35776
A3OP	0.9243242865	0.9160922793	143.76300	162.05531
A3WC	0.9187152779	0.9164581422	133.00635	151.24905
A20N	0.9331593033	0.9149552461	112.11224	105.30660
A3OQ	0.9074732666	0.9073534907	124.90734	71.27612
A5RI	0.9186601531	0.9161467308	111.28156	104.39026

Patient ID	Gini Coefficient (normal)	Gini Coefficient (tumor)	Standard Deviation (normal)	Standard Deviation (tumoral)
A3WV	0.9192367999	0.9008157607	150.26802	110.87332
A20W	0.9051581964	0.9194939622	108.33031	189.54448
A20R	0.9401306602	0.9181210557	133.90115	131.66324

Table 2. Summary table of the statistical analysis performed on the number of normal-tumor samples (second column) for each of the involved 17 tumor types (first column) based on the standard deviations. Columns 3-5 report the total number and the percentage of paired samples with a significant and not significant p-value, also discriminated as positive (with z-scores >0) and negative (with z-score <0). Plus and minus symbols near the tumor type under the first column represent the result of the Wilcoxon rank-sum test.

Tumor Type	Samples normal - tumor (paired)	Comparison of standard dev. for each patient through z-score		
		Positive ($p<0.01$)	Not significant	Negative ($p<0.01$)
BLCA	19 - 408 (19)	7 (36.8%)	9 (47.3%)	3 (15.7%)
BRCA +	113 - 1090 (112)	45 (40.1%)	53 (47.3%)	14 (12.5%)
CHOL +	9 - 36 (9)	4 (44.4%)	5 (55.5%)	0 (0%)
COAD +	41 - 456 (41)	13 (31.7%)	25 (60.9%)	3 (7.31%)
ESCA +	11 - 161 (8)	0 (0%)	8 (100%)	0 (0%)
HNSC +	44 - 500 (43)	13 (30.2%)	29 (67.4%)	1 (2.3%)
KICH -	24 - 65 (23)	0 (0%)	5 (21.7%)	18 (78.2%)
KIRC +	72 - 530 (72)	43 (59.7%)	25 (34.7%)	4 (5.5%)
KIRP	32 - 288 (31)	9 (29.0%)	18 (58.0%)	4 (12.9%)
LIHC +	50 - 371 (50)	7 (14.0%)	43 (86.0%)	0 (0.0%)
LUAD	59 - 513 (57)	9 (15.7%)	48 (84.2%)	0 (0.0%)
LUSC +	49 - 501 (49)	18 (36.7%)	28 (57.1%)	3 (6.1%)
PRAD	52 - 495 (52)	4 (7.6%)	36 (69.2%)	12 (23.0%)
READ	10 - 166 (9)	2 (22.2%)	6 (66.6%)	1 (11.1%)
STAD +	32 - 375 (27)	6 (22.2%)	19 (70.3%)	2 (7.4%)
THCA	58 - 502 (58)	7 (12.0%)	35 (60.3%)	16 (27.5%)
UCEC	35 - 543 (23)	5 (21.7%)	10 (43.4%)	8 (34.7%)

DISCLAIMER: The above article has been published, as is, ahead-of-print, to provide early visibility but is not the final version. Major publication processes like copyediting, proofing, typesetting and further review are still to be done and may lead to changes in the final published version, if it is eventually published. All legal disclaimers that apply to the final published article also apply to this ahead-of-print version.