



Software/web server article

## GENI: A web server to identify gene set enrichments in tumor samples

Arata Hayashi<sup>a</sup>, Shmuel Ruppó<sup>b</sup>, Elisheva E. Heilbrun<sup>c</sup>, Chiara Mazzoni<sup>c</sup>, Sheera Adar<sup>c</sup>, Moran Yassour<sup>c,d</sup>, Areej Abu Rmaileh<sup>a</sup>, Yoav D. Shaul<sup>a,\*</sup>

<sup>a</sup> Department of Biochemistry and Molecular Biology, The Institute for Medical Research Israel-Canada, Faculty of Medicine, Hebrew University of Jerusalem, Jerusalem 9112001, Israel

<sup>b</sup> Info-CORE, Bioinformatics Unit of the I-CORE at the Hebrew University of Jerusalem, Jerusalem, Israel

<sup>c</sup> Department of Microbiology and Molecular Genetics, The Institute for Medical Research Israel-Canada, Faculty of Medicine, Hebrew University of Jerusalem, Jerusalem 9112001, Israel

<sup>d</sup> School of Computer Science & Engineering, The Hebrew University of Jerusalem, Jerusalem, Israel



### ARTICLE INFO

#### Keywords:

Bioinformatics  
Cancer biology  
TCGA  
Gene Set Enrichment Analysis  
Web-based tools  
Clinical data  
Tumor samples  
Cancer-associated molecular mechanisms  
Multi-Gene Analysis

### ABSTRACT

The Cancer Genome Atlas (TCGA) and analogous projects have yielded invaluable tumor-associated genomic data. Despite several web-based platforms designed to enhance accessibility, certain analyses require prior bioinformatic expertise. To address this need, we developed Gene ENrichment Identifier (GENI, <https://www.shaulab.com/geni>), which is designed to promptly compute correlations for genes of interest against the entire transcriptome and rank them against well-established biological gene sets. Additionally, it generates comprehensive tables containing genes of interest and their corresponding correlation coefficients, presented in publication-quality graphs. Furthermore, GENI has the capability to analyze multiple genes simultaneously within a given gene set, elucidating their significance within a specific biological context. Overall, GENI's user-friendly interface simplifies the biological interpretation and analysis of cancer patient-associated data, advancing the understanding of cancer biology and accelerating scientific discoveries.

## 1. Introduction

The Cancer Genome Atlas (TCGA) [1] and similar initiatives have revolutionized cancer research by establishing extensive repositories of genomic, transcriptomic, and clinical data spanning a diverse array of cancer types [2]. These collaborative databases are pivotal resources for researchers, serving as gateways to untangle the intricate landscape of cancer and identify potential therapeutic targets. They are readily accessible through web-based platforms such as the Genomic Data Commons (GDC) (<https://portal.gdc.cancer.gov/>) [3] and cBioPortal (<http://www.cbioportal.org/>) [4], enabling investigators to examine individual genes and dissect their functions. While these web servers are of substantial value for examining user-selected genes and conducting basic analyses, a serious challenge persists for those without a strong computational background.

Here, we introduce GENI, a user-friendly, web-based platform that facilitates intuitive examination of TCGA data. GENI is designed to promptly compute correlations for genes of interest from the entire transcriptome and rank them against well-established biological gene

sets, by utilizing Gene Set Enrichment Analysis (GSEA) [5]. This analytical technique assesses the behavior of entire gene sets in a certain biological context, by comparing it to predefined gene sets. The streamlined interface of GENI ensures that all users, regardless of their expertise level, can effortlessly access and analyze TCGA data, and obtain publication-quality graphs using a single tool. Furthermore, GENI possesses an additional feature, as it integrates a multigene analysis capability, enabling comprehensive investigations into cancer-associated molecular mechanisms. Together, GENI enhances the accessibility of cancer genomics databases to the diverse community of researchers.

## 2. Material & methods

GENI is a freely available web-based platform developed using the R programming language. It uses various packages from the open-source Bioconductor and CRAN to enable its diverse range of features. Additionally, it includes the Shiny package for efficient data analysis and visualization and the clusterProfiler for GSEA calculation [6].

\* Corresponding author.

E-mail address: [yoav.shaul@mail.huji.ac.il](mailto:yoav.shaul@mail.huji.ac.il) (Y.D. Shaul).

<https://doi.org/10.1016/j.csbj.2023.10.053>

Furthermore, the program calculation of Spearman's and Pearson's correlation coefficients is facilitated by R's built-in functions. The expression data used in GENI were obtained from multiple sources, including the TCGA website (<https://dcc.icgc.org/releases/PCAWG/>), NCBI (<https://ftp.ncbi.nlm.nih.gov/gene/DATA/>), and from the cBioPortal for Cancer Genomics which was accessed using the cBioPortalData R package [4,7]. These diverse datasets were then combined and stored on the Shinyapps.io cloud platform (<https://www.shinyapps.io>). Fig. 1 provides a graphical abstract, illustrating the implementation of GENI. The complete script for GENI is accessible at <https://github.com/ArataHayashi/GENI-Gene-ENrichment-Identifier>.

### 3. Results

#### 3.1. Workflow overview

##### 3.1.1. Single GENI

The GENI platform provides a streamlined search function that identifies correlations between a specific gene and the entire transcriptome using GSEA. The search begins by entering an NCBI gene ID or gene symbol in the "1. Enter your gene of interest" field (Fig. 2A). The user then selects the desired tissue and study in the "2. Search tissue" and "3. Select your study" fields, followed by choosing a gene set library obtained from the Molecular Signatures Database (MSigDB, [8]), using the droplist of "4. Select gene set" (Fig. 2A). Finally, by clicking "Apply GENI," correlation coefficients are calculated (rounded to 6 digits) and organized based on correlation values. Moreover, the "Advanced settings" button expands the search options, including the order of the summary table, correlation method, permutations, the maximum and minimum number of genes in the gene set, the p-value adjustment, the exponent, and the p-value cut-off (Fig. 2B).

The results are displayed as a table, summary dot plots, and a network plot in a new window. The user can further analyze specific gene sets of interest by clicking the indicated row in the table. This results in the appearance of a GSEA plot, enriched plot values, and a summary table of all the correlations. All of the results above can be downloaded in publication-quality PDF format and Excel format through the "Download" button.

##### 3.1.2. Multi GENI

GENI has the ability to analyze several genes simultaneously. The search begins by entering an NCBI gene ID or gene symbol in the "1. Enter your genes (up to 10 genes)" (Fig. 2C). The user selects the desired

tissue and study in the "2. Search tissue" and "3. Select your study" fields. Then, the user selects a specific library and a gene set using the droplist and table found in "4. Select gene set library". Finally, by clicking "Apply GENI," Spearman's correlation coefficients are calculated (rounded to 6 digits) and organized based on correlation values. Moreover, as for a Single GENI, the "Advanced settings" button also expands the search options (Fig. 2B). The results are displayed as a table and a dot plot in a new window. The results above are presented in a high resolution and can be downloaded in PDF and Excel format through the "Download" button.

##### 3.1.3. Apply GENI to your own data

With this feature, users can upload their own data and perform GENI analysis. To ensure optimal utilization of this tool, we strongly recommend that users download the provided example file and upload their data accordingly. Once the data are uploaded, users can select the gene of interest and conduct the same analysis as in the 'Single GENI' tab. Notably, GENI can analyze multiple file formats, including txt, xlsx, tsv, and csv. Detailed instructions for data submission can be found on the main page.

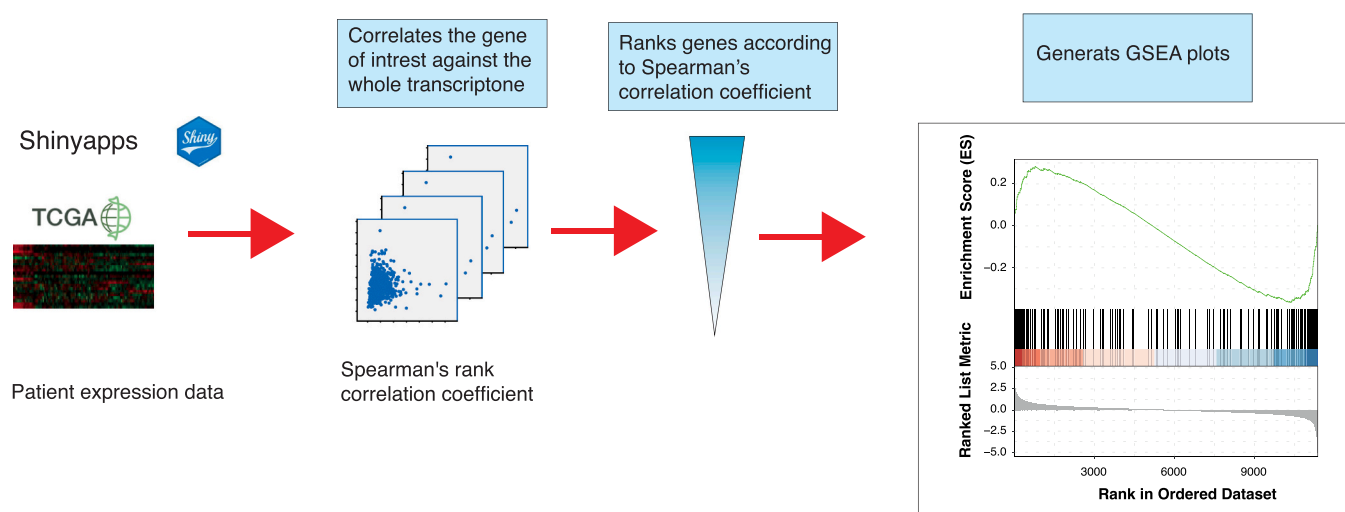
##### 3.1.4. Apply GSEA

GENI provides users with the ability to upload ranked data and perform GSEA. For this functionality, users must upload a file with two columns: one containing the gene list and the other indicating the respective ranks assigned to the genes. To facilitate the analysis process, it also offers a downloadable example table in several formats and a template file for user convenience. One of GENI's advantages is its versatility in analyzing multiple file formats, including txt, xlsx, and csv.

#### 3.2. Example

##### 3.2.1. Single gene analysis

To demonstrate GENI's ability to identify the biological context for specific genes, we used markers for the epithelial-mesenchymal transition (EMT) program as an example. This conserved cellular mechanism plays a significant role in cancer progression, contributing to stem-cell properties, drug resistance, immune evasion, and metastasis [9]. The induction of the EMT program is orchestrated by signaling pathways in response to extracellular cues such as the transforming growth factor- $\beta$  (TGF $\beta$ ) [10]. These changes include loss of cell polarity and cell-to-cell adhesion, along with alterations in the expression levels of cell surface receptors and cytoskeletal reorganization [11]. Additionally, this



**Fig. 1. An illustration of the GENI workflow.** Gene expression data are imported from the Shinyapps.io cloud and used to calculate Spearman's or Pearson's correlation for the gene or genes of interest against all protein-coding genes in the transcriptome. The resulting correlation values are then ranked against known gene sets. Finally, GENI produces a table of significantly correlated genes and multiple pathway enrichment graphs of publication quality.

**A**

1. Enter your gene of interest

Please enter your gene of interest. Example: EXT1, IL6, 2131, 1806

2. Search tissue

All

3. Select your study

Show 10 entries

name	Number of patients
Acute Myeloid Leukemia (TCGA, Firehose Legacy)	173
Acute Myeloid Leukemia (TCGA, NEJM 2013)	173
Acute Myeloid Leukemia (TCGA, PanCancer Atlas)	173
Adrenocortical Carcinoma (TCGA, Firehose Legacy)	79
Adrenocortical Carcinoma (TCGA, PanCancer Atlas)	78
Adult Soft Tissue Sarcomas (TCGA, Cell 2017)	206
Bladder Cancer (MSK/TCGA, 2020)	296
Bladder Cancer (TCGA, Cell 2017)	408
Bladder Urothelial Carcinoma (TCGA, Firehose Legacy)	408
Bladder Urothelial Carcinoma (TCGA, Nature 2014)	129

Showing 1 to 10 of 80 entries

4. Select gene set

hallmark gene sets (H)

Advanced settings -

Apply GENI

**B**

Advance settings

Select first order of the sheet

NES\_UP

Select correlation method

spearman

Change permutations

1000

Change minimum number of genes in set

15

Change maximum number of genes in set

500

Select normalization option

BH

Select exponent option

weighted

Change p value cut off

1

Reset

**C**

1. Enter your genes (up to 10 genes)

2. Search tissue

All

3. Select your study

Show 5 entries

name	Number of patients
Acute Myeloid Leukemia (TCGA, Firehose Legacy)	173
Acute Myeloid Leukemia (TCGA, NEJM 2013)	173
Acute Myeloid Leukemia (TCGA, PanCancer Atlas)	173
Adrenocortical Carcinoma (TCGA, Firehose Legacy)	79
Adrenocortical Carcinoma (TCGA, PanCancer Atlas)	78

Showing 1 to 5 of 80 entries

4. Select gene set library

hallmark gene sets (H)

You have selected hallmark gene sets (H).

Show 5 entries

Gene set name
1 HALLMARK_ADIPOGENESIS
2 HALLMARK_ALLOGRAFT_REJECTION
3 HALLMARK_ANDROGEN_RESPONSE
4 HALLMARK_ANGIOGENESIS
5 HALLMARK_APICAL_JUNCTION

Showing 1 to 5 of 50 entries

Advanced settings -

Apply GENI

Calculations may take several minutes.

**Fig. 2. Search panel.** (A) A screenshot of the search panel for Single GENI. (B) A screenshot of the advanced search option. (C) A screenshot of the search panel for Multi GENI.

program induces significant changes in the cell's transcriptomic profile, as the genes associated with the mesenchymal phenotype, such as N-cadherin (CDH2), are upregulated [12]. Hence, identifying genes correlating with the EMT hallmark emerged as a promising strategy to identify unknown factors that potentially function in cancer cell aggressiveness.

In this example, we utilized GENI to search for genes co-expressed with N-cadherin in the breast invasive carcinoma (TCGA, Firehose legacy) dataset. Upon clicking the "Apply GENI" button, the intermediate table (Fig. 3A), summary plots (Fig. 3B), and network plot (Fig. 3C) of the GSEA result are displayed in the main panel. From the intermediate table, we selected the "hallmark of epithelial-mesenchymal transition" gene set as it demonstrated the highest normalized enrichment score (NES) value (Fig. 3B). Upon selecting this gene set, a GSEA plot (Fig. 3D), a detailed result table (Fig. 3E), and a gene list for the selected gene set were displayed (Fig. 3F). For comparison to the hallmark of the EMT, we selected the "hallmark of oxidative phosphorylation," as it was the gene set demonstrating the lowest NES value (Fig. 3G). Our analysis identified that N-cadherin co-expressed genes were strongly correlated with EMT, demonstrating the usefulness of GENI in identifying potential factors that function in cancer cell aggressiveness.

In a previous study utilizing GENI, we identified that the expression level of the adaptor protein dihydropyrimidines like-2 (DPYSL2) in breast cancer patients correlated with the EMT markers [13]. Next, we validated that this adaptor protein interacts with the signaling molecule Janus kinase 1 (JAK1) by applying biochemical and mice-based experimental settings. Moreover, this interaction is essential for activating the signal transducer and activator of transcription 3 (STAT3), a

downstream transcription factor that regulates cancer cell aggressiveness. Hence, this study validated GENI as a tool to predict the biological function of a given gene.

### 3.2.2. Multigene analysis

Recent studies from our lab have highlighted the critical role of metabolic rewiring in the proper execution of the EMT program [14]. To further investigate this phenomenon, we developed a web-based tool called the Metabolic gENE Rapid Visualizer (MERAV), <http://merav.wi.mit.edu> [15] to systematically identify metabolic genes that are exclusively expressed in particular tumor subtypes. This analysis identified 44 metabolic genes upregulated in high-grade tumors bearing mesenchymal markers, which we designated the "mesenchymal metabolic signature" (MMS) [16]. In addition, our group confirmed three MMS genes, dihydropyrimidine dehydrogenase (DPYD), exostosin glycosyltransferase 1 (EXT1) [16], and glutathione peroxidase 8 (GPX8) [17], as EMT-promoting enzymes. In this example, we utilized Multi GENI to search for these two genes expressed in the breast invasive carcinoma (TCGA, Firehose legacy) dataset. We selected the Hallmark gene sets library and the "Hallmark Epithelial Mesenchymal Transition" gene set. Upon clicking the "Apply GENI" button, the summary table (Fig. 4A) and a plot (Fig. 4B) of the GSEA result were displayed in the main panel. As predicted, both genes demonstrated a significant correlation with the EMT hallmarks. In contrast, the epithelial marker occludin (OCLN) [18] presented a significant and negative NES (Fig. 4B). In addition, the expression profile of the known epithelial marker, E-cadherin (CDH1), was in contrast to the mesenchymal marker, N-cadherin (CDH2), demonstrating a clear "cadherin switch" phenomenon [9]. Therefore,

A

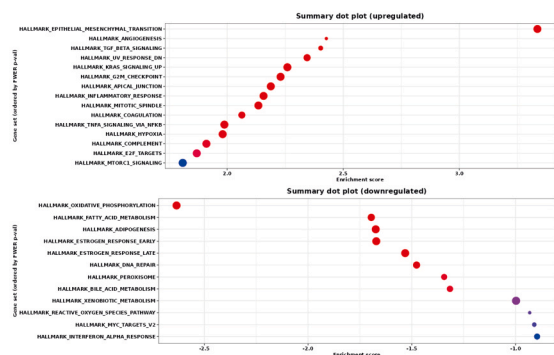
Select the Gene Set for further analysis

Show 10 entries

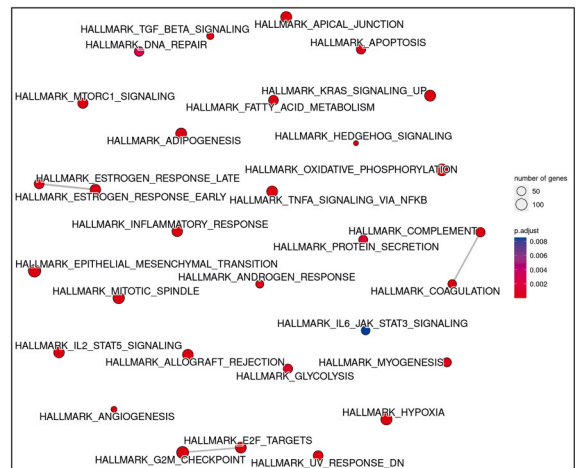
Download	GeneSet	Size	NES	NOM p-val	FWER p-val	FDR q-val
<input type="checkbox"/>	HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION	197	3.33745429587854	5.39573827169327e-63	2.69786913584664e-61	7.38364184547501e-62
<input type="checkbox"/>	HALLMARK_ANGIOGENESIS	36	2.41448546709555	1.68679245424121e-9	8.43396227120603e-9	2.30824230580376e-9
<input type="checkbox"/>	HALLMARK_TGF_BETA_SIGNALING	54	2.39118558608873	4.58052701235216e-10	2.54473722908453e-9	6.96454399538924e-10
<input type="checkbox"/>	HALLMARK_UV_RESPONSE_DN	144	2.33541496899198	1.58982872409456e-13	1.58982872409456e-12	4.3511101922588e-13
<input type="checkbox"/>	HALLMARK_KRAS_SIGNALING_UP	200	2.25627156084416	1.06075268392705e-14	1.76792113987842e-13	4.83852101440411e-14
<input type="checkbox"/>	HALLMARK_G2M_CHECKPOINT	200	2.22691523783589	4.1811414071306e-14	5.22642675891325e-13	1.43039048138678e-13
<input type="checkbox"/>	HALLMARK_APICAL_JUNCTION	200	2.18492259375488	3.85857916325812e-13	3.21548263640484e-12	8.80026826707993e-13
<input type="checkbox"/>	HALLMARK_INFLAMMATORY_RESPONSE	200	2.15371526856084	1.5416777294489e-12	1.10119837817778e-11	3.01380608764446e-12
<input type="checkbox"/>	HALLMARK_MITOTIC_SPINDLE	199	2.13509172879265	1.29954991491485e-11	8.12218696821782e-11	2.22291432814383e-11
<input type="checkbox"/>	HALLMARK_COAGULATION	138	2.06188089283665	2.10226307551784e-8	8.08562721353017e-8	2.2129085005451e-8

Showing 1 to 10 of 50 entries

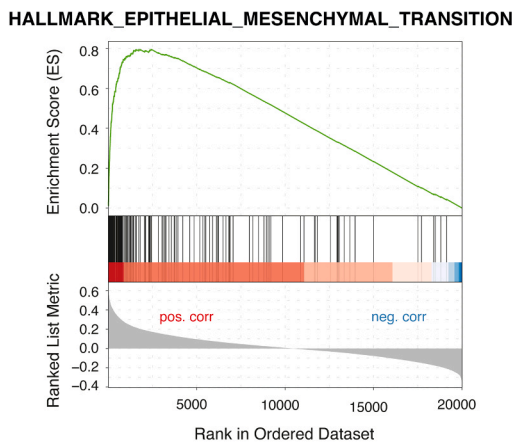
B



C



D



E

Name	Value
Enrichment Score (ES)	0.7954
Normalized Enrichment Score (NES)	3.3364
Nominal p-value	3.370e-63
FWER p-value	1.685e-61
FDR q-value	4.612e-62

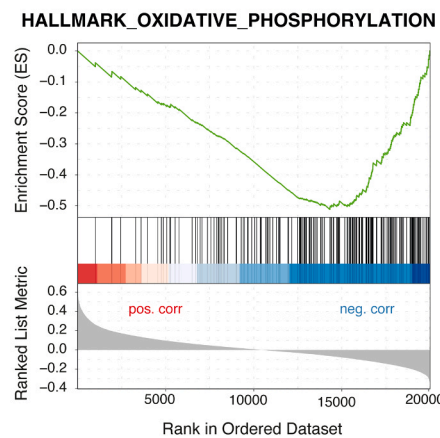
F

Show 10 entries

Symbol	Entrez gene ID	Rank Metric Score	Core Enrichment
FN1	2335	0.640229654533688	Yes
COL11A1	1301	0.587180934380594	Yes
INHBA	3624	0.583802665615582	Yes
THBS2	7058	0.578346150199702	Yes
CALD1	800	0.561598211452162	Yes
LOXL2	4017	0.558487256206945	Yes
LOX	4015	0.555544482668819	Yes
GREM1	26585	0.55504449753124	Yes
CDH11	1009	0.549207823759469	Yes
COL5A2	1290	0.547781985126619	Yes

Showing 1 to 10 of 197 entries

G

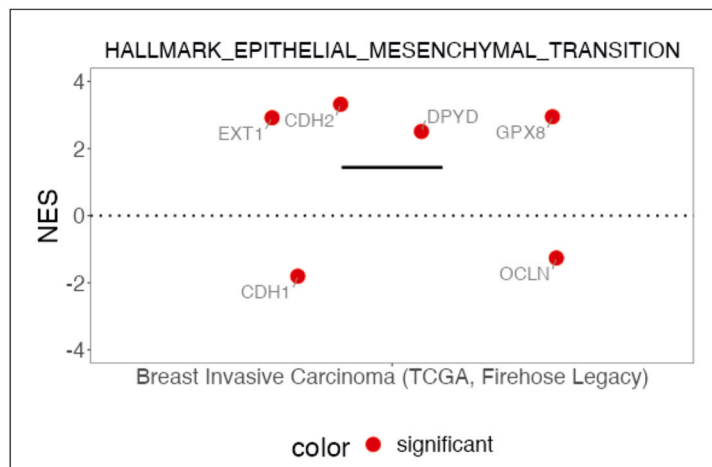


**Fig. 3. CDH2 expression correlates with EMT markers in breast cancer patients.** (A) A screenshot of the summary table. Size: number of genes in the gene set. NES: Normalized enrichment score. NOM p-value: nominal p-value. FWER p-value: Family-wise error rate. FDRq-value: False discovery rate. (B) A screenshot of a summary dot plot of upregulated and downregulated gene sets of CDH2 correlated genes. The results are colored based on the FDR values, and the ball size is set by the number of genes in the gene set. The gene sets are sorted ascendingly according to their FWER p-values. (C) A screenshot of the network plot of CDH2 correlated gene sets. The results are colored by FDR, and the ball size is set by the gene set size. (D) GSEA plot of the Hallmark of Epithelial-Mesenchymal Transition (EMT) shows a positive and significant correlation of CDH2 with the EMT markers in breast cancer patients. (E) Screenshot of the detailed result of the hallmark of the EMT. (F) A screenshot of the EMT gene list and Spearman's rank correlation coefficient with CDH2 expression. (G) GSEA plot of Hallmark of Oxidative Phosphorylation (OXPHOS) shows a negative correlation with CDH2 expression in breast cancer patients.

A

GeneSymbol	GeneID	enrichmentScore	NES	NOM p-val	FWER p-val	FDR q-val
CDH1	999	-0.397671702170594	-1.81663503694078	0.0000022287606680756	0.00000742920222691865	0.00000234606386113221
CDH2	1000	0.79543816420859	3.31609346261411	1.04614941397714e-63	5.23074706988569e-62	1.4315728822845e-62
DPYD	1806	0.65294885396334	2.50465639913162	7.14181025131778e-30	5.10129303665556e-29	1.61093464315439e-29
OCLN	100506658	-0.2596126970949	-1.25179166063184	0.0548446069469835	0.0979367981196135	0.0474220285631813
EXT1	2131	0.603816032861264	2.92426576421957	4.98791714941629e-28	2.49395857470814e-26	6.82557083604334e-27
GPX8	493869	0.794655737913531	2.95488407832997	8.25457191634919e-60	4.1272859581746e-58	1.65091438326984e-58

B



**Fig. 4. Comparison of multi-genes by Multi GENI.** (A) A screenshot of the summary table. NES: Normalized enrichment score. NOM p-val: nominal p-value. FWER p-val: Family-wise error rate. FDR q-val: False discovery rate. (B) Dot plot of the Hallmark of Epithelial Mesenchymal Transition (EMT) on Breast invasive carcinoma (TCGA, Firehose Legacy). Colored by FDR q-value (significant: value < 0.05) NES: Normalized enrichment score.

GENI can distinguish between mesenchymal and epithelial genes based on the hallmark of EMT, further validating the accuracy and biological relevance of the methodology implemented in this tool.

4. Discussion

Various publicly available web tools are commonly utilized to ascertain the biological context of provided expression samples. Several web tools such as DAVID (<https://david.ncifcrf.gov/home.jsp>), Metascape (<https://metascape.org>), and Enricher (<https://maayanlab.cloud/Enrichr/>) (Table 1) analyze the expression pattern of a given and limited set of genes. Moreover, certain web tools specialize in the analysis of specific gene subsets, such as the Gene Regulatory Network Database (GRAND) (<https://grand.networkmedicine.org>), focusing on transcription factors, and EMTome (<http://www.emtome.org>) on EMT genes.

Web-based platforms such as the GDC and cBioPortal have significantly impacted cancer research by providing user-friendly interfaces for researchers to access and analyze large-scale databases. Additionally, specialized bioinformatic tools such as Xena (<https://xena.ucsc.edu>) and Gene Expression Profiling Interactive Analysis (GEPIA) (<http://gepia.cancer-pku.cn/index.html>) have been developed to analyze TCGA data effectively. Specifically, these web-based tools allow the researchers to perform Kaplan-Meier survival analysis, compare tumor vs. normal tissues within or across tumors, determine the association between increased gene expression and the promoter epigenetic landscape, and create subgroups. Another notable tool is LinkedOmics (<https://>

**Table 1**

**Summary table of commonly used web tools.** *Gene Input:* Indicates the number of genes submitted for analysis in a single run. *Global or Selective Analysis:* Denotes the type of analysis conducted by the web tool. "Global" refers to comparing the gene of interest against the entire genome, while "Selective" indicates analysis performed exclusively on a specific set of genes. *Gene Enrichment:* Represents the type of analysis conducted by the web tool, focusing on gene enrichment. TF-transcription factors. EMT-Epithelial mesenchymal transition genes. Ref-references.

Websites	Gene Input	Global/Selective	Gene Enrichment	Ref
DAVID	multiple	selective	V	[19]
Metascape	multiple	selective	V	[20]
Enricher	single	selective (up to 500 genes)	V	[21]
GRAND	single/multiple (limited to TF)	selective		[22]
EMTome	single (limited to EMT)	global	V	[23]
cBioPortal	single	global		[4]
GDC	single	global		[3]
Xena	single	global		[24]
GEPIA	single	global		[25]
LinkedOmics	single	global	V	[26]
GENI	single/multiple	global	V	

linkedomics.org), a web-based platform integrating multi-omics data from cancer studies, facilitating the analysis and interpretation of molecular profiles in the context of clinical outcomes. LinkedOmics provides the option to conduct various analyses such as differential expression, survival analysis, pathway enrichment, and network analysis.

Unlike these platforms with various options for analysis, GENI offers a more focused tool that results in an easy-to-use interface. This approach provides a biological context for the analysis of gene expression data and allows for the identification of potential pathways and processes involved in cancer progression. Notably, GENI has no restriction in its input, enabling the analysis of any gene in the genome. In addition, a significant advantage of GENI is its capability to analyze multiple genes simultaneously (Table 1). Moreover, the user-friendly interface of GENI makes it easy for researchers to perform complex analyses without requiring extensive bioinformatics expertise. GENI's unique feature adds another layer of analysis to this wealth of information and provides a valuable resource for the scientific community.

GENI is built on the Shinyapp platform, utilizing its capabilities to provide an intuitive and effortless interface for exploring and analyzing TCGA data through GSEA. However, it is essential to note that Shinyapp's underlying infrastructure influences the stability of the web tool. Shinyapp is a third-party service, so we cannot directly modify its structure to address stability concerns. Nevertheless, recognizing the significance of a stable user experience, we have proactively optimized our script to enhance computational efficiency within this framework. We are committed to continuously monitoring and fine-tuning GENI to ensure the best possible stability and performance, striving to provide an optimal platform for users to seamlessly engage with TCGA data and derive valuable insights into their research questions.

We have outlined forthcoming features to extend the platform's capabilities and offer a more comprehensive analytical experience. One of our main goals for the near future is to incorporate data integration capabilities with the Cancer Cell Line Encyclopedia (CCLE) and the Gene Expression Omnibus (GEO), broadening the scope of GENI's utility. These enhancements aim to enable users to conduct more comprehensive analyses by incorporating a broader range of data sources.

## 5. Conclusion

GENI provides a user-friendly and powerful web-based platform for exploring the TCGA database by allowing researchers to investigate gene expression levels relative to known gene sets. Its features are a result of its ability to conduct precise GSEA, profoundly elevating the analytical depth and biological relevance of gene expression data. Noteworthy is GENI's capability to extend the analysis beyond individual genes, which provides a broader understanding of gene interactions. Thus, GENI can potentially aid in detecting therapeutic targets and bring new perspectives to the study of cancer progression, ultimately benefiting the scientific community. Overall, GENI's importance lies in its simplicity, biological relevance, and accessibility, making it an attractive tool for cancer research.

## Funding

This work was supported by the Israel Science Foundation (Grant 299/21) and the Israel Cancer Research Fund project grant. AH is supported by the Brodie fellowship for breast cancer research, AH and CM by the Hebrew University International Ph.D. Talent Scholarship. MY is the Rosalind, Paul and Robin Berlin Faculty Development Chair in Perinatal Research, also supported by the Azrieli Foundation.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence

the work reported in this paper.

## Acknowledgment

The authors thank the Shaul Lab team for the helpful discussions about user questions and Emilia Malachi for illustrating the GENI logo. We would like to express our gratitude to Dr. Gillian Kay for her assistance with English editing.

## References

- [1] Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, et al. An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell* 2018;173(400–416):e11. <https://doi.org/10.1016/j.cell.2018.02.052>.
- [2] Network CGAR Weinstein JN, Collisson EA, Shaw KRM Mills GB, Ozenberger BA, et al. The cancer genome atlas pan-cancer analysis project. *Nat Genet* 2013;45:1113–20. <https://doi.org/10.1038/ng.2764>.
- [3] L GR, P HA, Vincent F, E VH, R LD, A KW, et al. Toward a shared vision for cancer genomic data. *N Engl J Med* 2016;375:1109–12. <https://doi.org/10.1056/nejmp1607591>.
- [4] Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *p11–p11 Sci Signal* 2013;6. <https://doi.org/10.1126/scisignal.2004088>.
- [5] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 2005;102:15545–50. <https://doi.org/10.1073/pnas.0506580102>.
- [6] Yu G. clusterProfiler: an universal enrichment tool for functional and comparative study. *Biorxiv* 2018:256784. <https://doi.org/10.1101/256784>.
- [7] Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2012;2:401–4. <https://doi.org/10.1158/2159-8290.cd-12-0095>.
- [8] Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 2011;27:1739–40. <https://doi.org/10.1093/bioinformatics/btr260>.
- [9] Nieto MA, Huang RY-J, Jackson RA, Thiery JP. EMT: 2016. *Cell* 2016;166:21–45. <https://doi.org/10.1016/j.cell.2016.06.028>.
- [10] Katsuno Y, Meyer DS, Zhang Z, Shokat KM, Akhurst RJ, Miyazono K, et al. Chronic TGF- $\beta$  exposure drives stabilized EMT, tumor stemness, and cancer drug resistance with vulnerability to bitopic mTOR inhibition. *Sci Signal* 2019;12:eaau8544. <https://doi.org/10.1126/scisignal.aau8544>.
- [11] Brabletz S, Schuhwerk H, Brabletz T, Stemmler MP. Dynamic EMT: a multi-tool for tumor progression. *Embo J* 2021;40:e108647. <https://doi.org/10.15252/embj.2021108647>.
- [12] Bakir B, Chiarella AM, Pitarresi JR, Rustgi AK. EMT, MET, plasticity, and tumor metastasis. *Trends Cell Biol* 2020;30:764–76. <https://doi.org/10.1016/j.tcb.2020.07.003>.
- [13] Rmaileh AA, Solaimuthu B, Khatib A, Lavi S, Tanna M, Hayashi A, et al. DPYSL2 interacts with JAK1 to mediate breast cancer cell migration. *J Cell Biol* 2022;221:e202106078. <https://doi.org/10.1083/jcb.202106078>.
- [14] Sciacovelli M, Frezza C. Metabolic reprogramming and epithelial-to-mesenchymal transition in cancer. *FEBS J* 2017;284:3132–44. <https://doi.org/10.1111/febs.14090>.
- [15] Shaul YD, Yuan B, Thiru P, Nutter-Upham A, McCallum S, Lanzkron C, et al. MERAV: a tool for comparing gene expression across human tissues and cell types. *Nucleic Acids Res* 2016;44:D560–6. <https://doi.org/10.1093/nar/gkv1337>.
- [16] Shaul YD, Freinkman E, Comb WC, Cantor JR, Tam WL, Thiru P, et al. Dihydropyrimidine accumulation is required for the epithelial-mesenchymal transition. *Cell* 2014;158:1094–109. <https://doi.org/10.1016/j.cell.2014.07.032>.
- [17] Khatib A, Solaimuthu B, Yosef MB, Rmaileh AA, Tanna M, Oren G, et al. The glutathione peroxidase 8 (GPX8)/IL-6/STAT3 axis is essential in maintaining an aggressive breast cancer phenotype. *Proc Natl Acad Sci* 2020;117:21420–31. <https://doi.org/10.1073/pnas.2010275117>.
- [18] Aigner K, Dampier B, Descovich L, Mikula M, Sultan A, Schreiber M, et al. The transcription factor ZEB1 (deltaEF1) promotes tumour cell dedifferentiation by repressing master regulators of epithelial polarity. *Oncogene* 2007;26:6979–88. <https://doi.org/10.1038/sj.onc.1210508>.
- [19] Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, et al. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol* 2003;4:P3. <https://doi.org/10.1186/gb-2003-4-5-p3>.
- [20] Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* 2019;10:1523. <https://doi.org/10.1038/s41467-019-09234-6>.
- [21] Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* 2016;44:W90–7. <https://doi.org/10.1093/nar/gkw377>.
- [22] Ben Guebila M, Lopes-Ramos CM, Weighill D, Sonawane AR, Burkholz R, Shamsaei B, et al. GRAND: a database of gene regulatory network models across human conditions. *Nucleic Acids Res* 2021;50:gkab778. <https://doi.org/10.1093/nar/gkab778>.

- [23] Vasaikar SV, Deshmukh AP, Hollander P, den, Addanki S, Kuburich NA, Kudaravalli S, et al. EMTome: a resource for pan-cancer analysis of epithelial-mesenchymal transition genes and signatures. *Brit J Cancer* 2021;124:259–69. <https://doi.org/10.1038/s41416-020-01178-9>.
- [24] Goldman MJ, Craft B, Hastie M, Repěčka K, McDade F, Kamath A, et al. Visualizing and interpreting cancer genomics data via the Xena platform. *Nat Biotechnol* 2020; 38:675–8. <https://doi.org/10.1038/s41587-020-0546-8>.
- [25] Yang R, Tang Z, Guo J, Li C, Lin Z, Kang B, et al. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res* 2017; 45:W98–102. <https://doi.org/10.1093/nar/gkx247>.
- [26] Vasaikar SV, Straub P, Wang J, Zhang B. LinkedOmics: analyzing multi-omics data within and across 32 cancer types. *Nucleic Acids Res* 2017;46:gkx1090. <https://doi.org/10.1093/nar/gkx1090>.