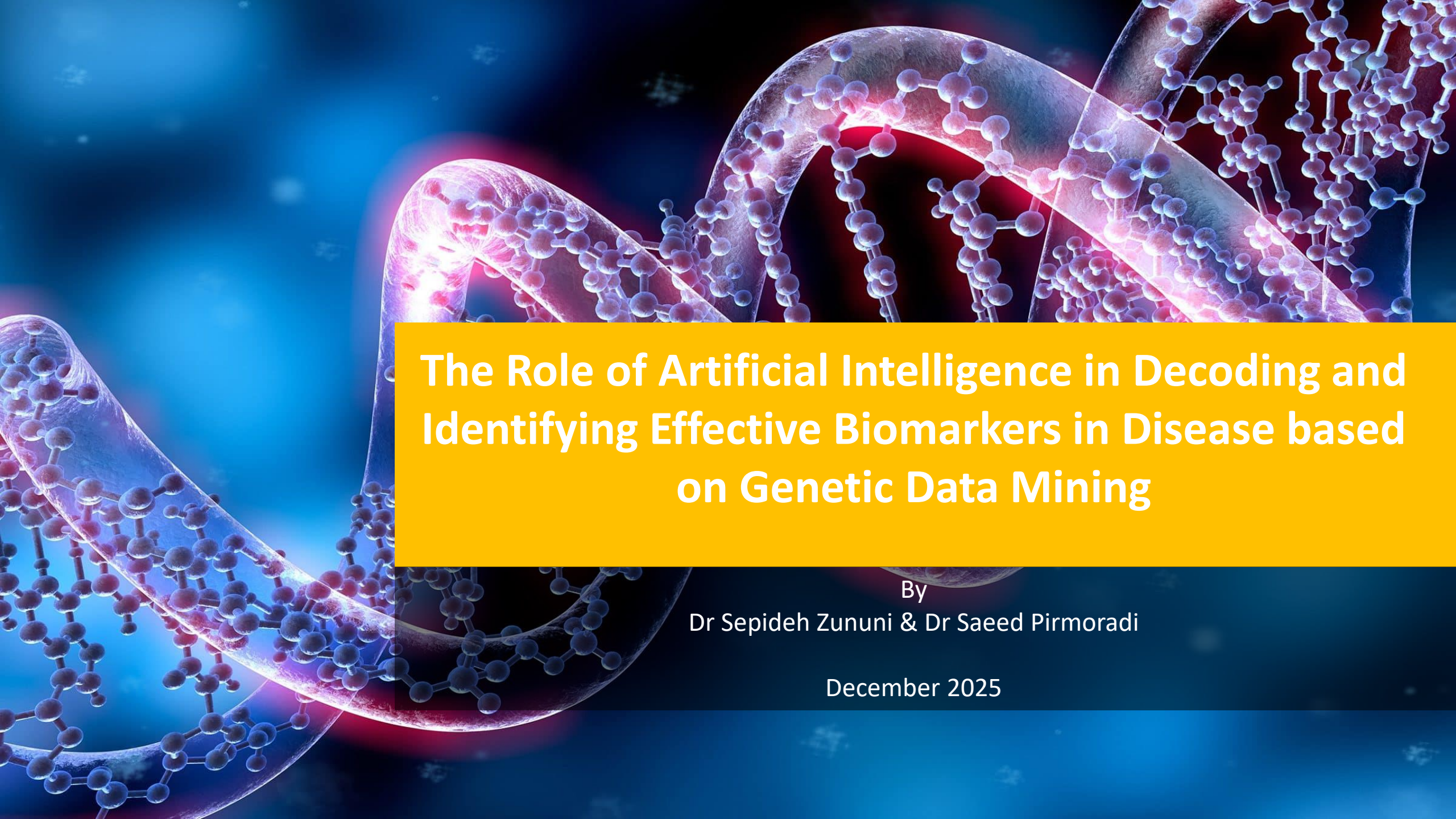


The Role of Artificial Intelligence in Decoding and Identifying Effective Biomarkers in Disease based on Genetic Data Mining

By
Dr Sepideh Zununi & Dr Saeed Pirmoradi

December 2025

Today's Presenters

Dr Sepideh Zununi

- Ph.D. in Medical Biotechnology
- Associate Professor
- Kidney Research Center
- Clinical Research Institute
- Tabriz University of Medical Sciences

Dr Saeed Pirmoradi

- Ph.D. in Artificial Intelligence (AI)
- Post-doc in AI in Medicine
- Certificate of Master of Business Administration (MBA)
- Founder of Khayyam AI Innovation Lab
- Data Scientist in Clinical Research Development Unit of Tabriz Valiasr Hospital

Machine Learning Section

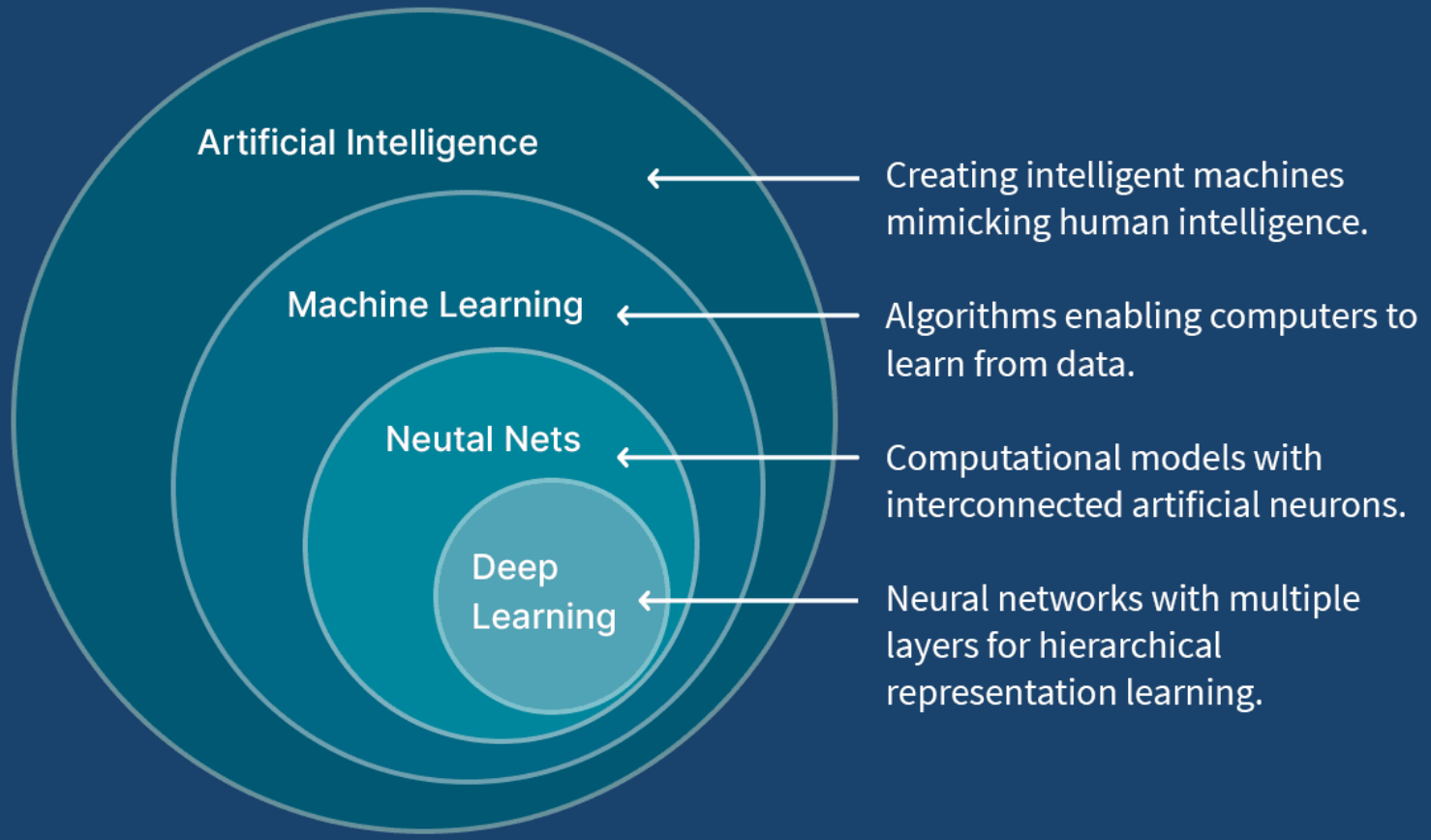
By
Dr Saeed Pirmoradi

December 2025

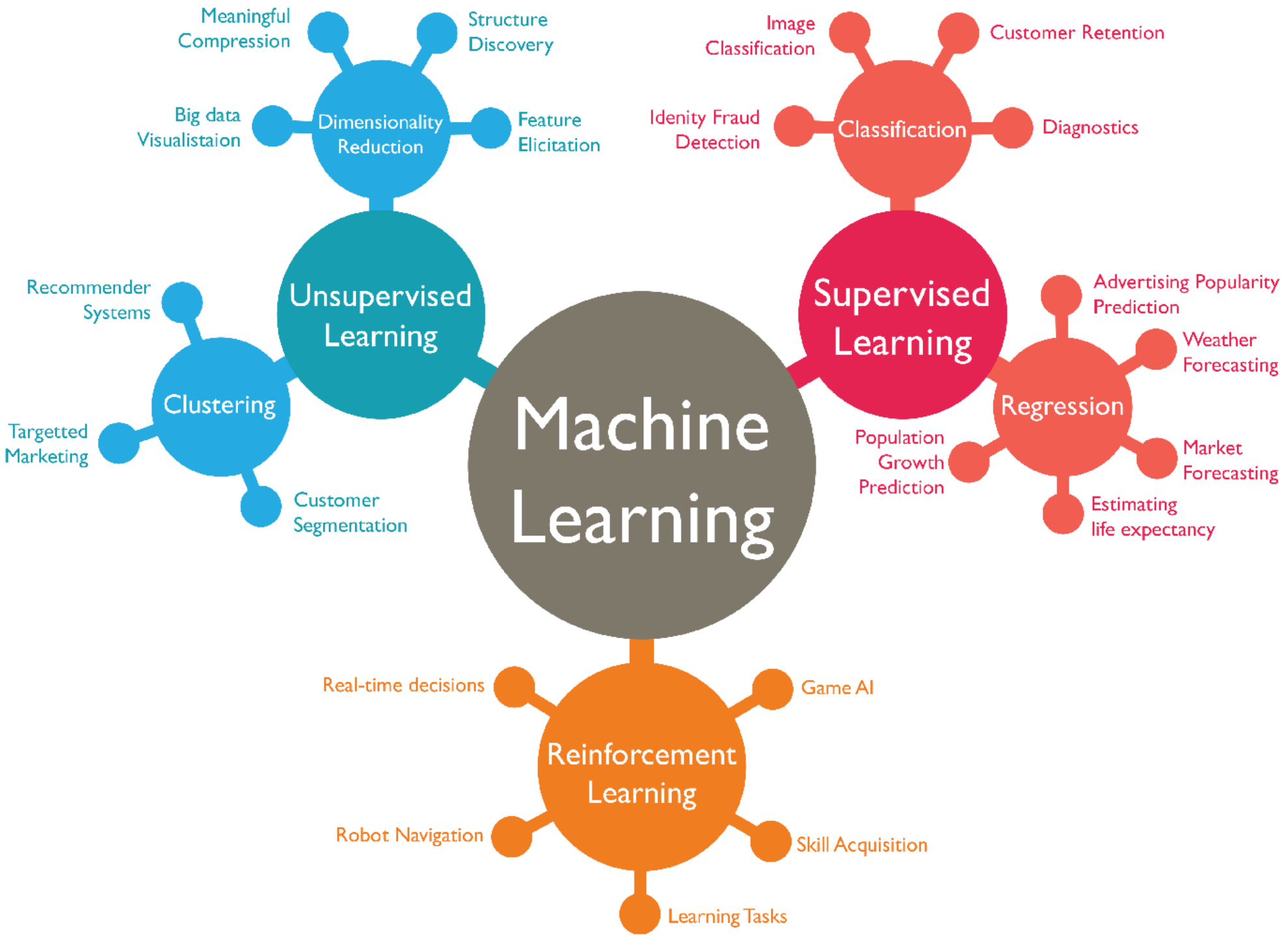
Outline

- Introduction
- Machine Learning roadmap
- Reading Step
- Preprocessing
- Feature Selection Step
- Classification Step
- Explainable AI
- Practical example

What is Machine Learning?



Machine Learning



Machine Learning Process

Datasets

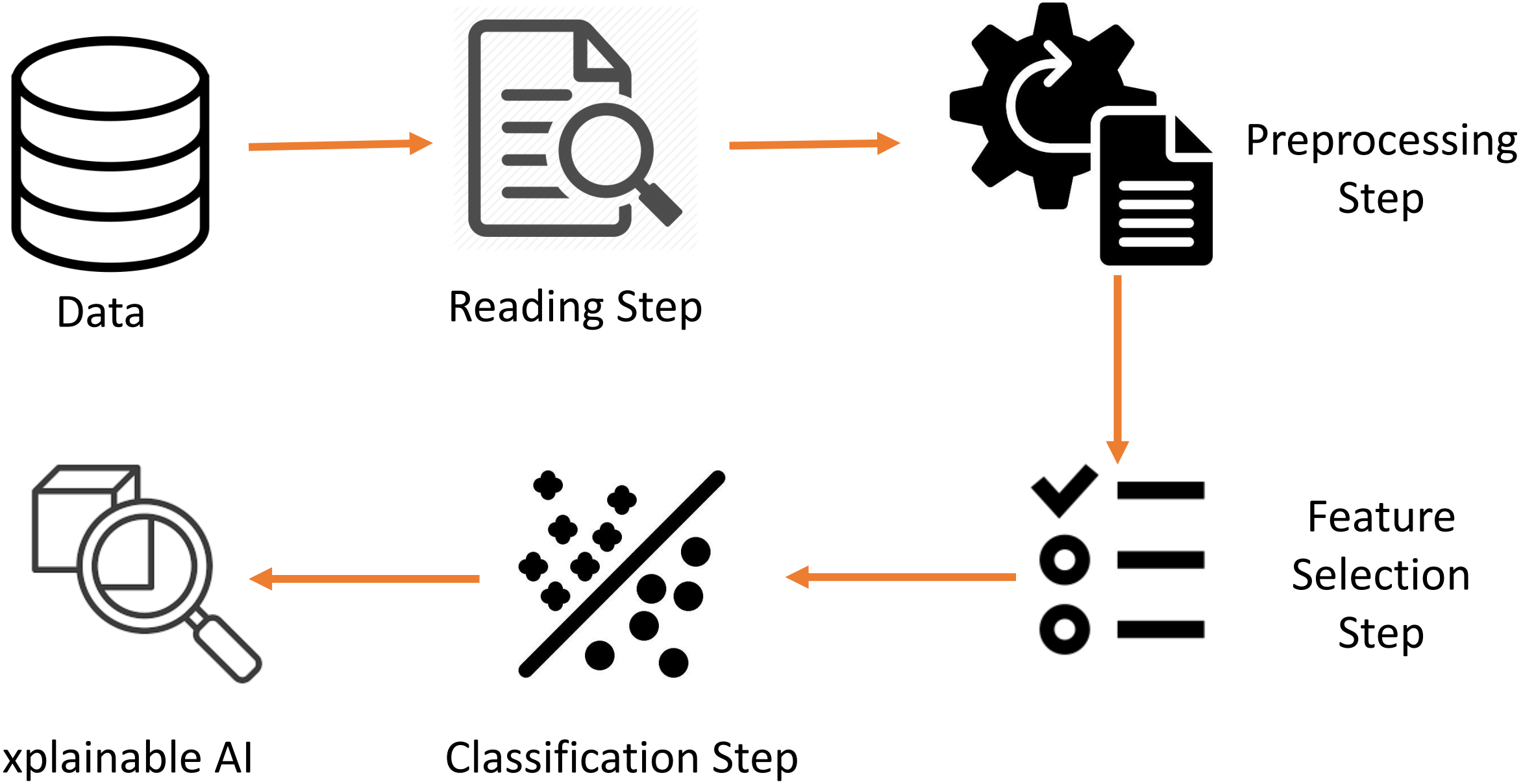
https://portal.gdc.cancer.gov/

https://www.linkedomics.org/login.php

Datasets

Gene Expression Omnibus Platforms

https://www.ncbi.nlm.nih.gov/geo/


Data
Reading Step
Preprocessing Step
Feature Selection Step
Classification Step
Explainable AI

Data

- Low sample size
- High Dimension
- Complex patterns

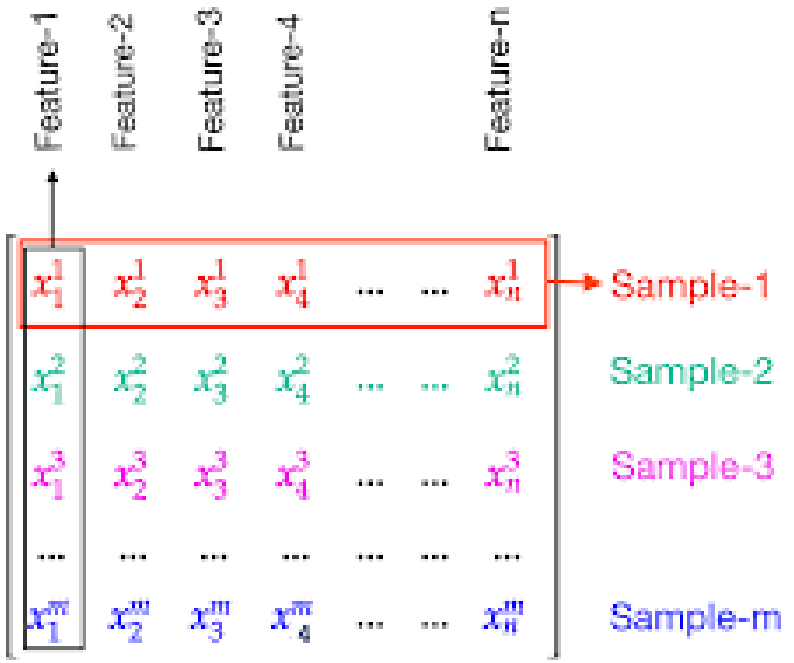
Reading Step

- Reading data (text, excel, …)
- Create data matrix and target vector

$$
\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ \vdots \\ m \end{array}
\begin{array}{c}
\begin{array}{cccc} 1 & 2 & \cdots & n \end{array} \\
\begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
a_{21} & a_{22} & \cdots & a_{2n} \\
a_{31} & a_{32} & \cdots & a_{3n} \\
\vdots & \vdots & \vdots & \vdots \\
a_{m1} & a_{m2} & \cdots & a_{mn}
\end{bmatrix}
\end{array}
$$

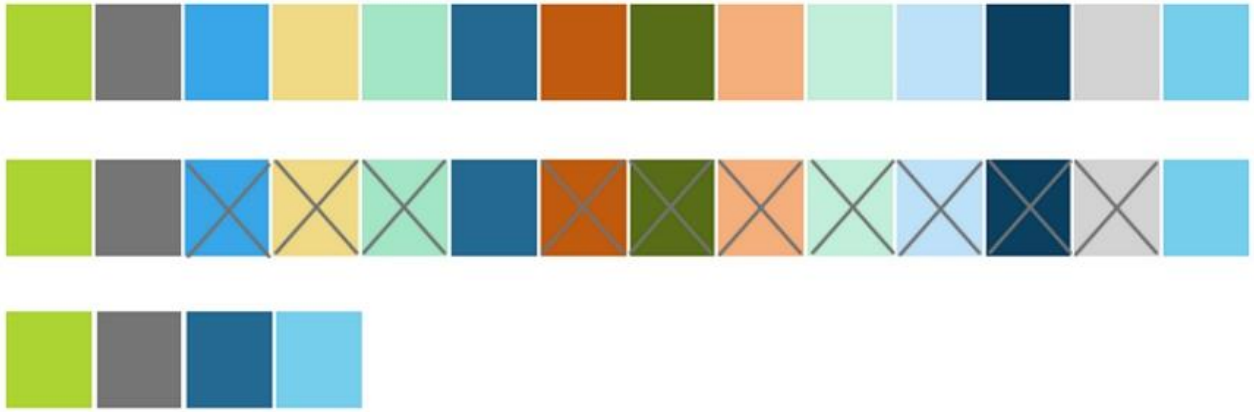
Preprocessing Step

- Cross-Validation (Hold-Out, K-Fold, …)
- Missing values
- Scaling (Min-Max, Standard, …)
- Visualization (Principal Component Analysis, …)



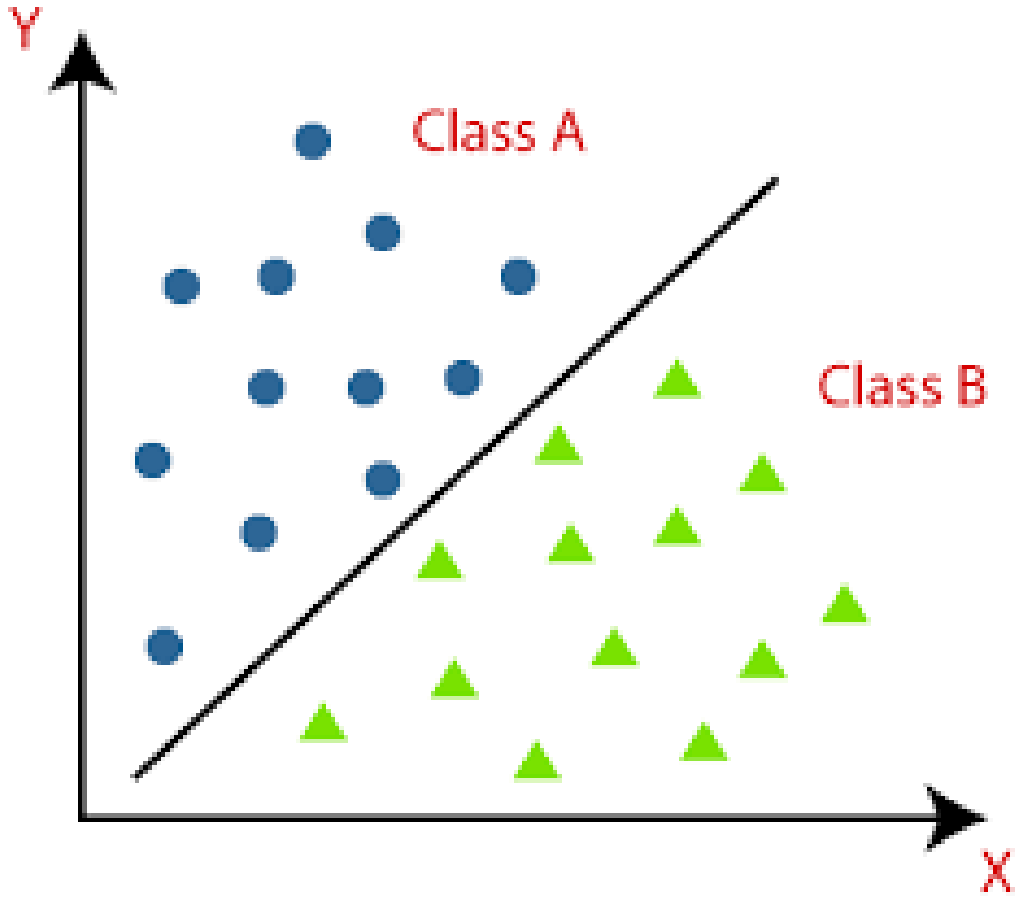
Feature Selection Step

- Filter
- Wrapper
- Embedded
- Hybrid
- Metaheuristic



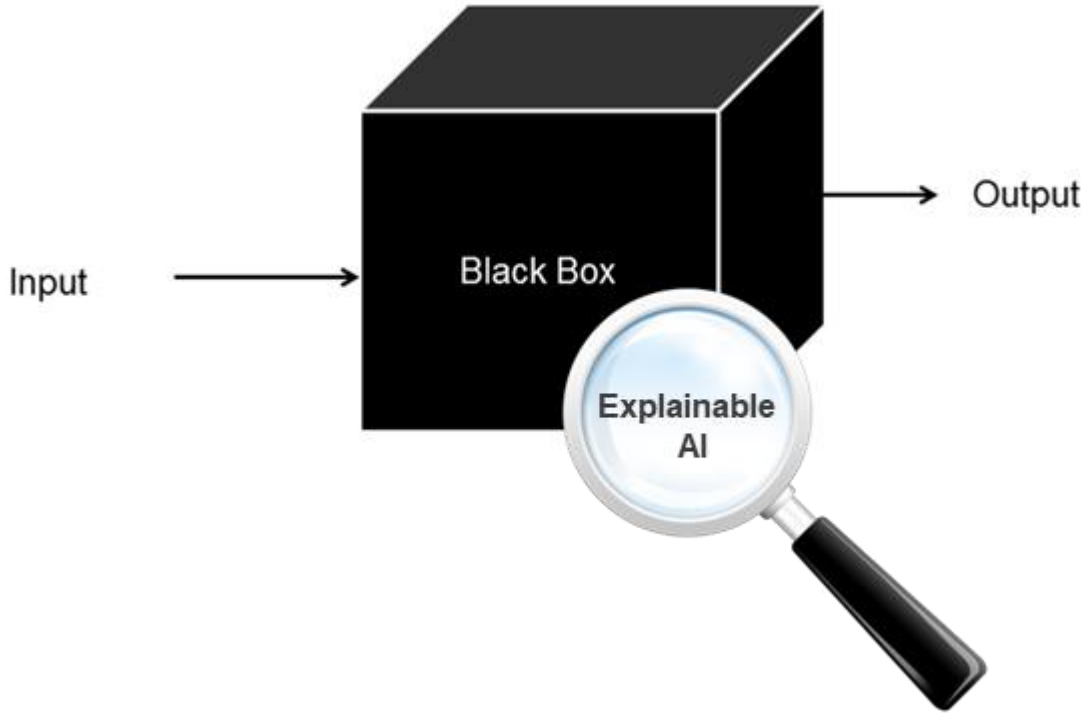
Classification Step

- Naïve Bayes
- LDA or QDA
- Support Vector Machine (SVM
- K-Nearest Neighbors (KNN)
- Artificial Neural Network (ANN)
- Decision Tree
- Random Forest
- …



Explainable AI

- SHAP values
- Association Rule Mining
- …



nature

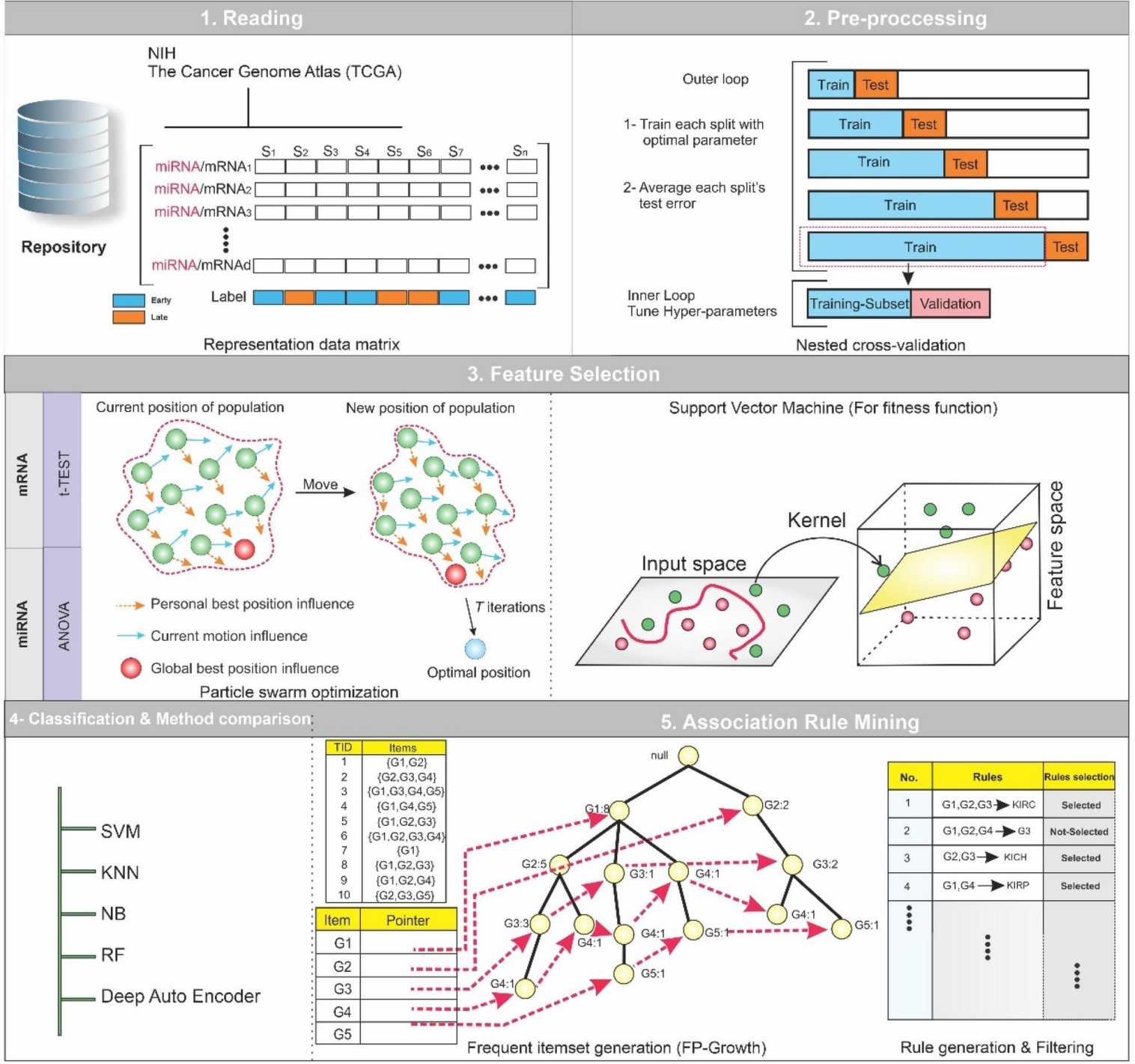
scientific reports

Key therapeutic targets implicated at the early stage of hepatocellular carcinoma identified through machine-learning approaches

Abstract:

Hepatocellular carcinoma (HCC) is the most frequent type of primary liver cancer. Early-stage detection plays an essential role in making treatment decisions and identifying dominant molecular mechanisms. We utilized machine learning algorithms to find significant mRNAs and microRNAs (miRNAs) at the early and late stages of HCC. First, pre-processing approaches, including organization, nested cross-validation, cleaning, and normalization were applied. Next, the t-test/ANOVA methods and binary particle swarm optimization were used as a filter and wrapper method in the feature selection step, respectively. Then, classifiers, based on machine learning and deep learning algorithms were utilized to evaluate the discrimination power of selected features (mRNAs and miRNAs) in the classification step. Finally, the association rule mining algorithm was applied to selected features for identifying key mRNAs and miRNAs that can help decode dominant molecular mechanisms in HCC stages. The applied methods could identify key genes associated with the early (e.g., Vitronectin, thrombin-activatable fibrinolysis inhibitor, lactate dehydrogenase D (LDHD), miR-590) and late-stage (e.g., SPRY domain containing 4, regucalcin, miR-3199-1, miR-194-2, miR-4999) of HCC. This research could establish a clear picture of putative candidate genes, which could be the main actors at the early and late stages of HCC.

(1) In the reading step, each dataset was downloaded from the TCGA repository.
(2) The preprocessing step includes two sub-steps, nested cross-validation, and normalization.
(3) The feature selection step contains two sub-steps: the filter method based on t-test for mRNA data and ANOVA for miRNA data, and the wrapper method based on binary particle swarm optimization (PSO) for both mRNA and miRNA data, in which candidate miRNAs/mRNAs with more relevance to early-stage and late-stage Hepatocellular Carcinoma (HCC) were selected.
(4) Mult classifier models were utilized to evaluate the discrimination power of selected miRNAs/mRNAs.
(5) The Association Rule Mining method discovered the hidden relationship between selected miRNAs/mRNAs at the early-stage and late-stage of HCC in the first level and the complex relationship among selected miRNAs/mRNAs in the second level.



The mRNA expression was reported in terms of FPKM values for 60,483 RNA transcripts. In the miRNA profile, 1881 miRNA expression values were recorded using the Illumina HiSeq 2000 platform. The HCC stage system was defined based on the TNM system; (T) the size of the primary tumor, (M) the distant metastasis, and (N) the spread of cancer to lymph nodes. In this study, we considered stage I as an early-stage class and stages II, III, and IV as a late-stage class.

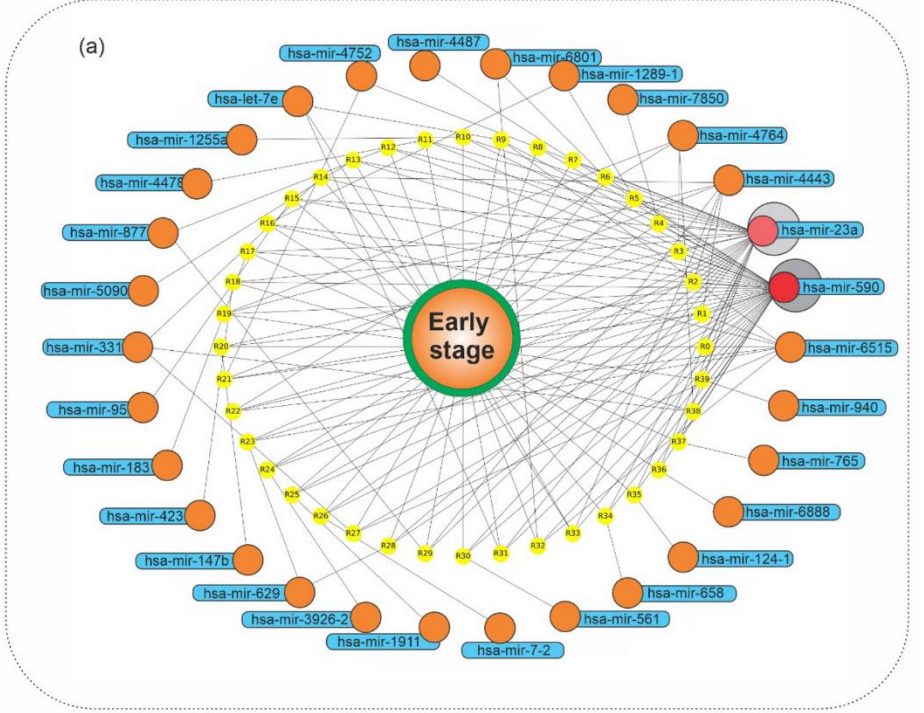
mRNA data		miRNA data	
Early-stage	**Late-stage**	**Early-stage**	**Late-stage**
189	192	190	192

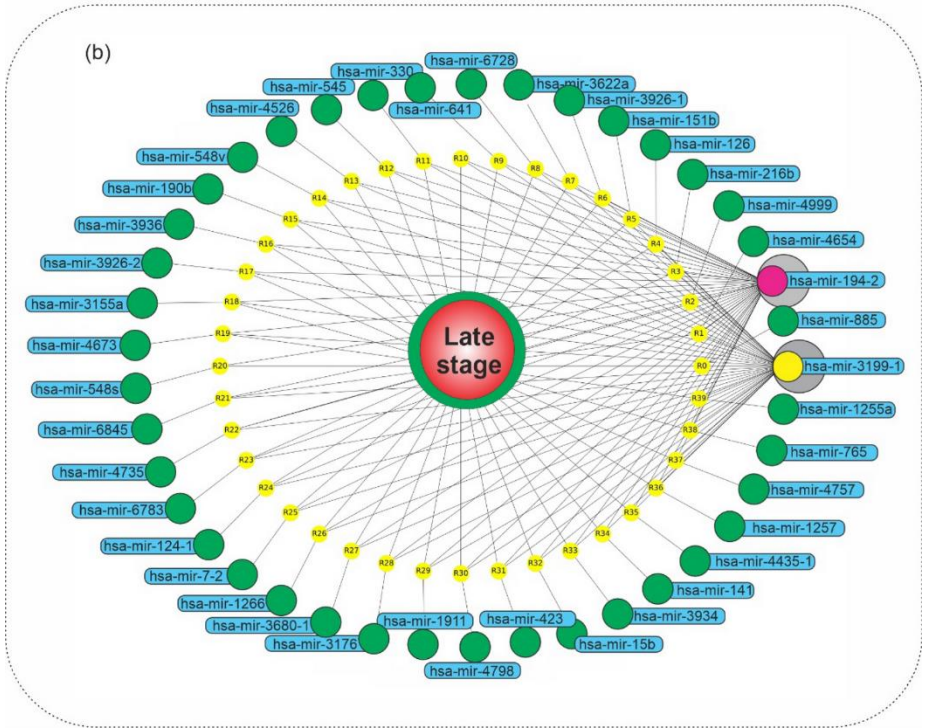
The performance of classifiers based on 123 selected mRNAs.

Classifier	Folds	Accuracy	AUC-ROC	F1-score (Early stage)	F1-score (Late stage)	MCC	Sn	Sp
SVM	Train	96.1	0.96	0.96	0.96	0.92	0.94	0.97
	Validation	75.4	0.75	0.76	0.74	0.51	0.7	0.8
	Test	**76.9**	**0.77**	**0.78**	**0.74**	**0.54**	**0.7**	**0.83**

The performance of classifiers based on 77 selected miRNAs.

Classifier	Folds	Accuracy	AUC-ROC	F1-score (Early stage)	F1-score (Late stage)	MCC	Sn	Sp
SVM	Train	88.8	0.88	0.89	0.88	0.78	0.81	0.95
	Validation	71.3	0.71	0.72	0.7	0.43	0.67	0.75
	Test	**70**	**0.7**	**0.7**	**0.68**	**0.4**	**0.68**	**0.73**

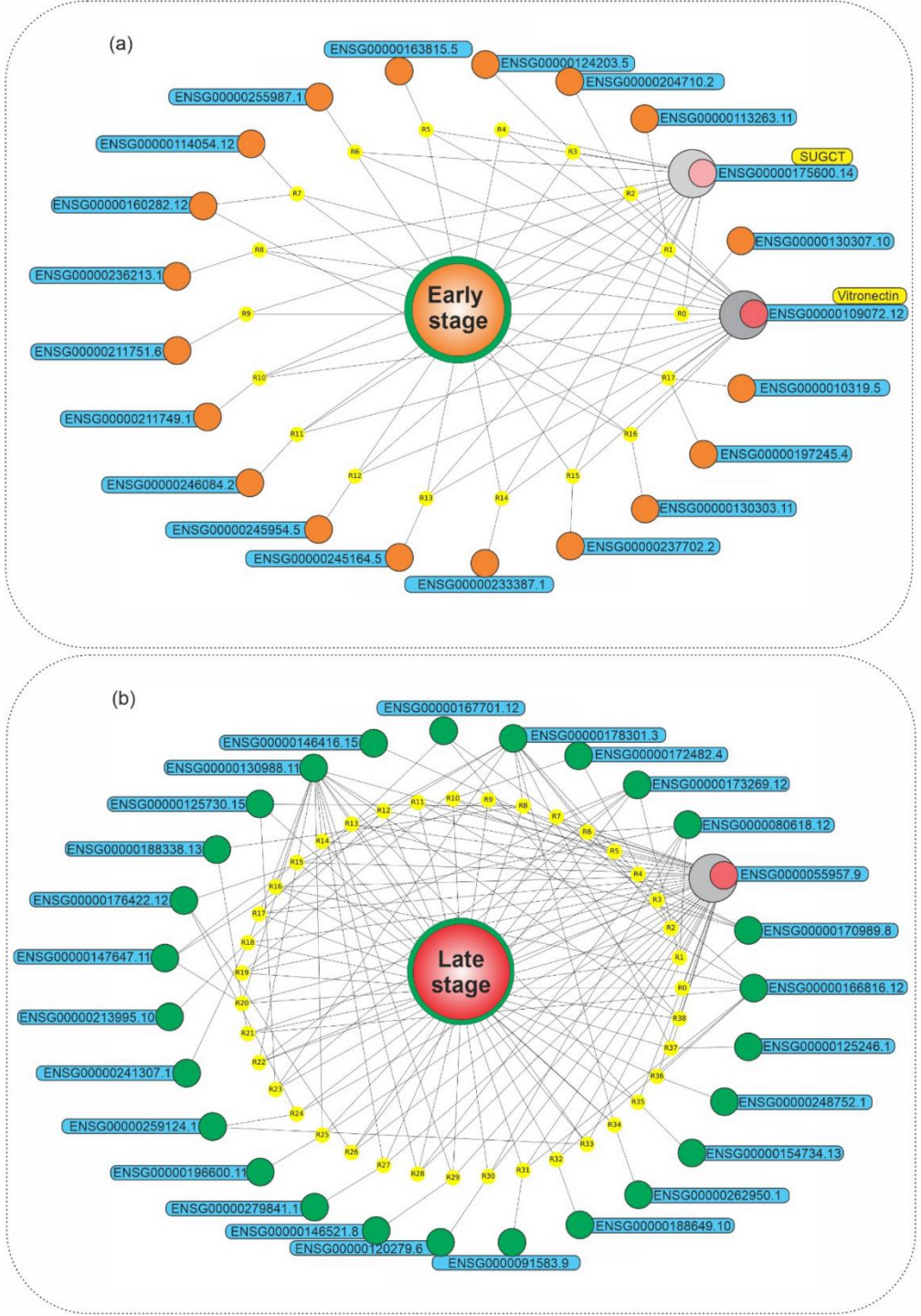




Graph network of miRNAs at early- and late-stage of HCC. Graph network of

a) early-stage related association rules (with lift > 1.16)

b) late-stage related association rules (with lift > 1.2), in which the early-stage phenotype, its rules, and related miRNAs were presented, by orange, yellow, and blue colors, respectively.



Graph network of mRNAs at early- and late-stages of HCC. Graph network of

(a) early-stage related association rules (with lift > 1.21)

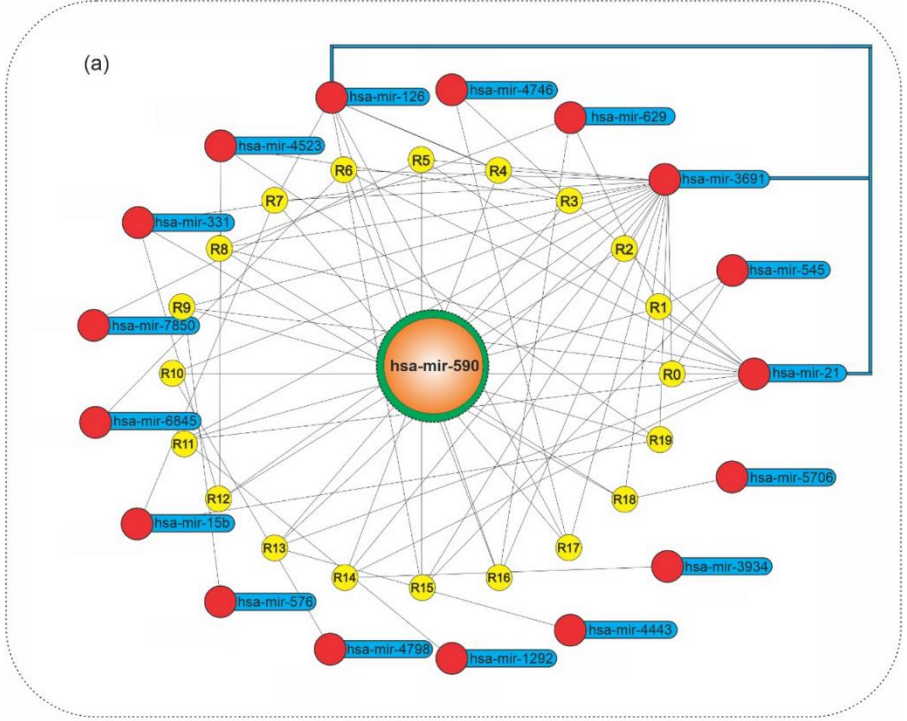
(b) late-stage related association rules (with lift > 1.38), in which the early-stage/late-stage phenotype, its rules, and related mRNAs were presented, by orange/red, yellow, and blue colors, respectively.

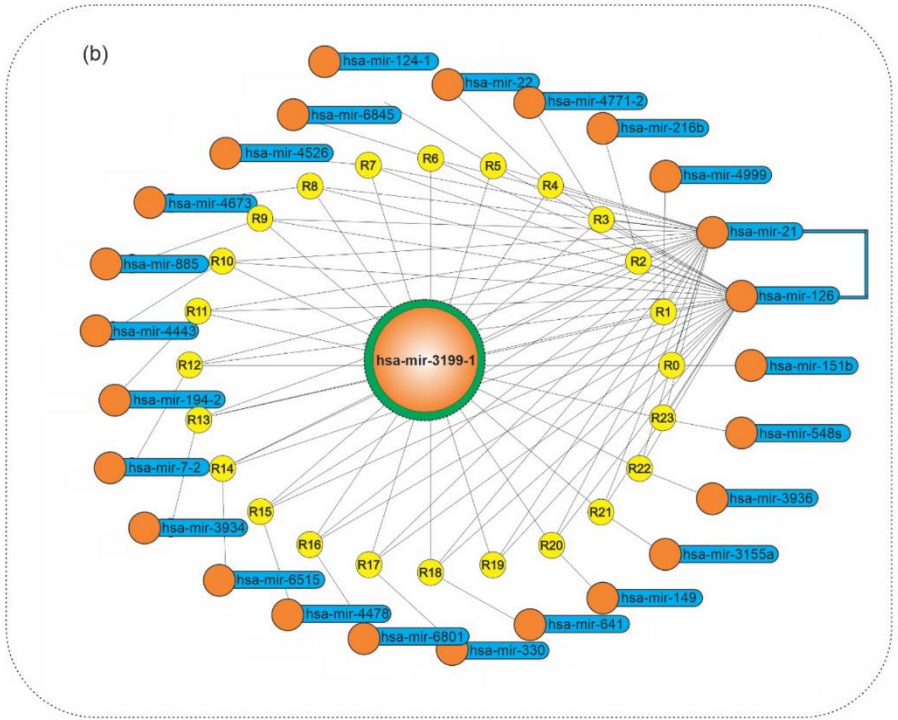
Graph network of has-mir-590 and has-mir-3199-1 in HCC. Graph network of

(a) has-mir-590 (with lift > 1.14)

(b) has-mir-3199-1 (with lift > 1.126)

related association rules, in which the has-mir-590 and has-mir-3199-1, their rules, and their related miRNAs were presented, by orange, yellow, and blue colors, respectively.

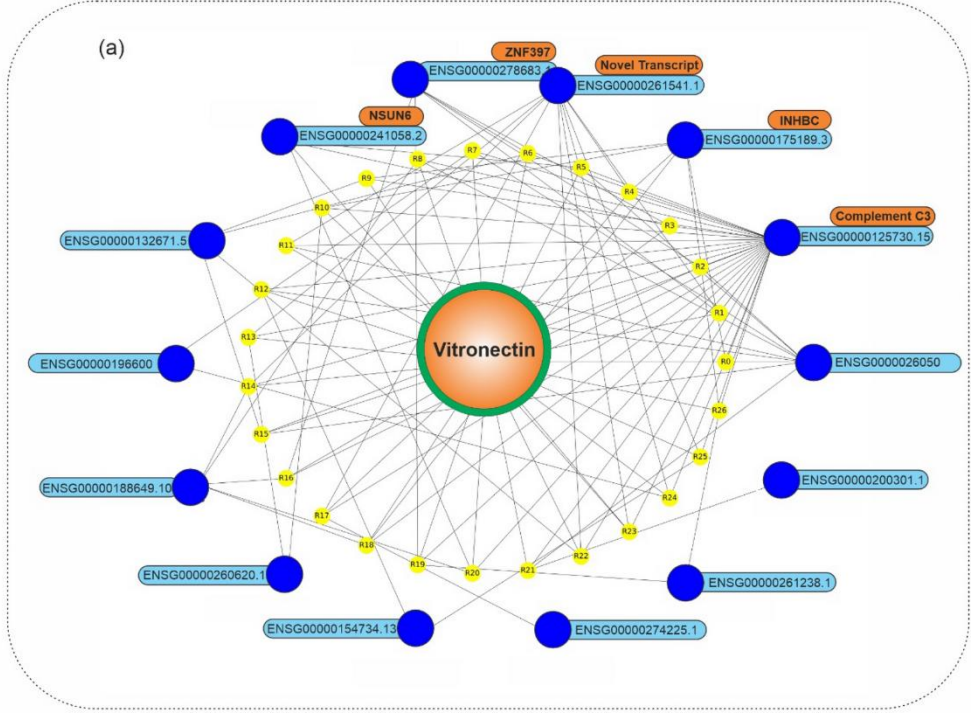


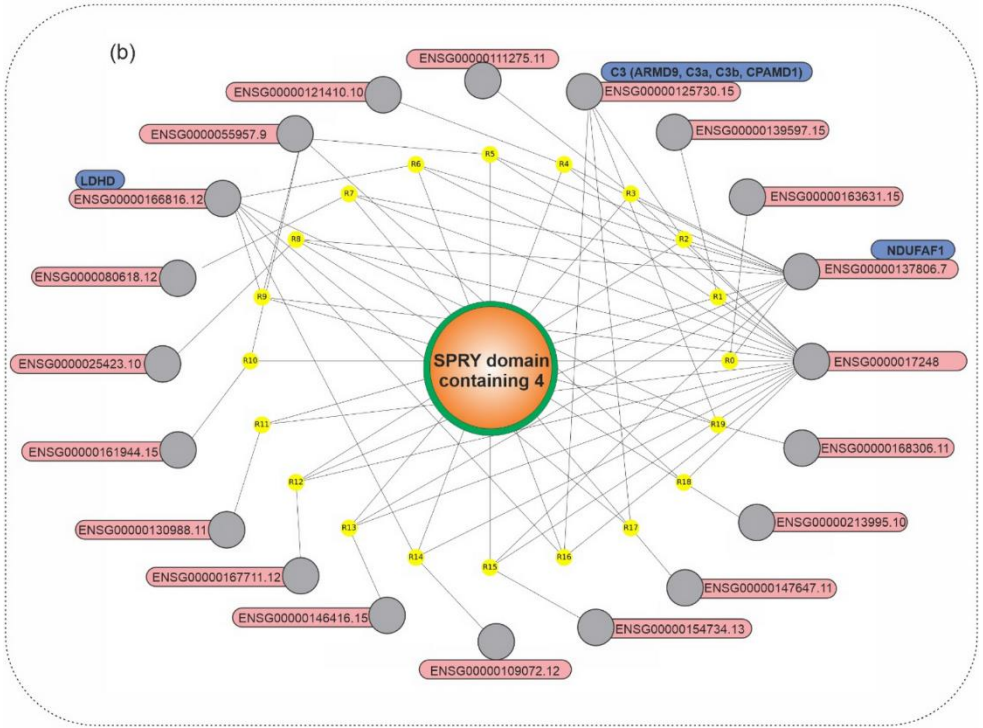


Graph network of Vitronectin and SPRY domain containing 4 in HCC.

(a) Graph network of Vitronectin (with lift > 1.42) related association rules, in which the Vitronectin, its rules, and related mRNAs were presented, by orange, yellow, and blue colors, respectively. Vitronectin, the most frequent mRNA in the early-stage association rules, has a high dependency on ENSG00000125730 (Complement C3).

(b) Graph network of the SPRY domain containing 4 (with lift > 1.46) related association rules, in which the SPRY domain containing 4, its rules, and related mRNAs were presented, by orange, yellow, and blue colors, respectively.





Challenges

- Lack of Iranian population genetic data
- Lack of communication between genetic research and clinical professionals
- Lack of a wet lab experiment in the follow-up of the Machine Learning results

Part # 2
Interpreting candidate RNAs highlighted by AI:

From model importance to biological insight

Sepideh Zununi Vahed
Associate Professor of Medical Biotechnology
Tabriz University of Medical Sciences

December 2025

How to think when AI gives you a list of RNAs and asks you to interpret them:

- Biologically
- Pathologically
- clinically.

By the end of this lecture, you should clearly know:

- What these RNA lists really mean
- Interpret AI–identified RNAs
- A step-by-step biological interpretation workflow
- Handle known, partially known, and novel RNAs
- And what you are *obligated* to do before making biological or clinical claims.

Why Interpretation Matters

The real danger is not wrong prediction,
the real danger is **wrong interpretation**.

- AI frequently outputs lists of 'top RNAs'
- RNA lists are cheap, but interpretation is expensive.
  - Misinterpretation can lead to false biological claims, waste years of experimental work, mislead clinicians, conclusions that do not survive validation
  - Clinical translation depends on correct interpretation

So interpretation is not a luxury.
It is a **scientific responsibility**.

nature communications

Article

https://doi.org/10.1038/s41467-024-53851-9

Deep generative AI models analyzing circulating orphan non-coding RNAs enable detection of early-stage lung cancer

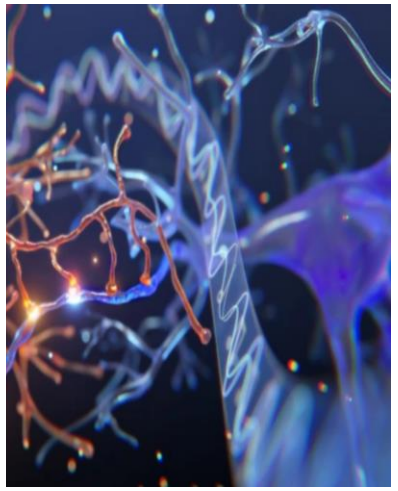
Received: 21 May 2024

Accepted: 22 October 2024

Published online: 21 November 2024

Check for updates

Mehran Karimzadeh[1,7], Amir Momen-Roknabadi[1,7], Taylor B. Cavazos[1,7], Yuqi Fang[1], Nae-Chyun Chen[1], Michael Multhaup[1], Jennifer Yen[1], Jeremy Ku[1], Jieyang Wang[1], Xuan Zhao[1], Philip Murzynowski[1], Kathleen Wang[1], Rose Hanna[1], Alice Huang[1], Diana Corti[1], Dang Nguyen[1], Ti Lam[1], Seda Kilinc[1], Patrick Arensdorf[1], Kimberly H. Chau[1], Anna Hartwig[1], Lisa Fish[1], Helen Li[1], Babak Behsaz[1], Olivier Elemento[2], James Zou[3], Fereydoun Hormozdiari[1,4] ✉, Babak Alipanahi[1] ✉ & Hani Goodarzi[5,6] ✉



Are these the most important RNAs in cancer biology?

When AI methods analyze RNA-sequence data, they typically output:
Ranked candidate RNAs (mRNA, lncRNA, circRNA, miRNA)

AI does not do

- ✘ AI does **not** discover biologically important RNAs.
- ✘ AI does **not** understand cancer.
- ✘ AI does **not** know pathways, oncogenes, or patients.

✓ It optimizes a **mathematical objective function**.

- Minimize prediction error
- Maximize classification accuracy
- Reduce uncertainty

✘ These RNAs drive cancer.

✓ **These RNAs helped *me*, as a model, perform a specific task.**

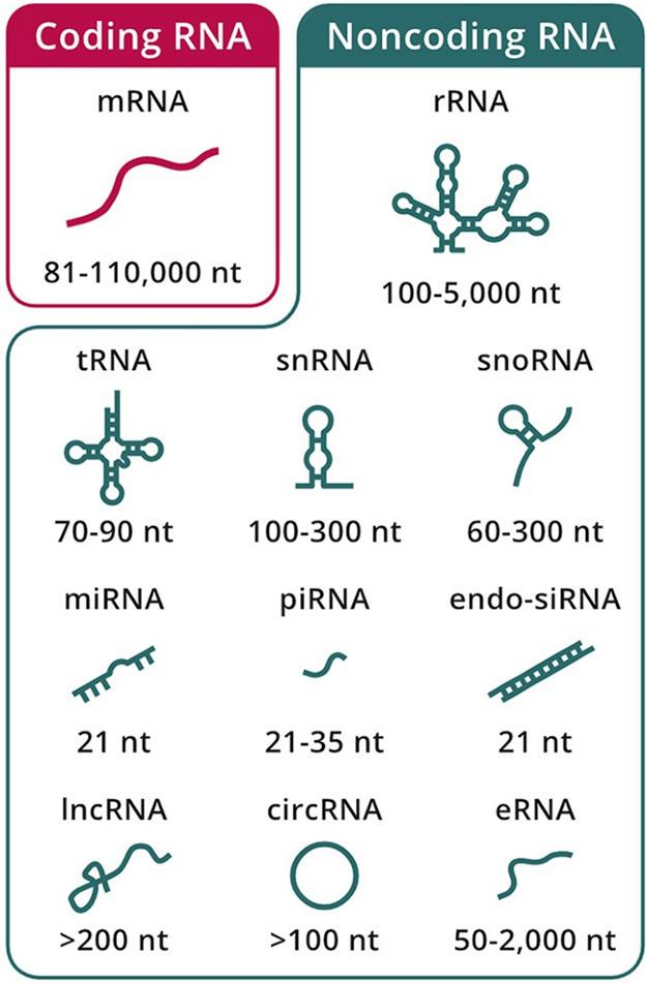
what *does* AI actually identify?

AI identifies RNAs that are **informative** for:

- ✓ A specific task
- ✓ On a specific dataset
- ✓ Under specific model assumptions

An **RNA** can be:

- •Highly informative
- •Highly predictive
- •And completely non-causal



AI **learns** from public cancer databases like:
TCGA, GEO, ArrayExpress
That contain real patient samples:
different hospitals, protocols, and sequencing platforms

A **candidate RNA** is:
- A model-derived signal
- A hypothesis generator
- Not a biological conclusion

AI gives you *questions*, not *answers*.
Biology begins **after** the model outputs the RNA list.

Your job as a biologist–oncologist is to **translate algorithmic importance into biological significance**.

20 top RNAs identified by AL-based methods on early stage of HCC:

	mRNA	#	mRNA ID	Repeat Count
0	M9	9	ENSG00000109072.12	1553
1	M6	6	ENSG00000080618.12	1398
2	M48	48	ENSG00000166816.12	1039
3	M28	28	ENSG00000137806.7	931
4	M32	32	ENSG00000146416.15	254
5	L24	24	ENSG00000130307.10	254
6	L103	103	ENSG00000255987.1	250
7	L99	99	ENSG00000245954.5	250
8	L98	98	ENSG00000245164.5	250
9	L91	91	ENSG00000236213.1	250
10	L90	90	ENSG00000233387.1	250
11	L77	77	ENSG00000211751.6	250
12	L76	76	ENSG00000211749.1	250
13	L100	100	ENSG00000246084.2	250
14	L46	46	ENSG00000163815.5	250
15	L92	92	ENSG00000237702.2	250
16	L17	17	ENSG00000124203.5	247
17	L11	11	ENSG00000113263.11	247
18	L113	113	ENSG00000264468.1	247
19	L0	0	ENSG00000010319.5	237
20	L63	63	ENSG00000178343.4	233

	miRNA	#	miRNA ID	Repeat Count
0	L59	59	hsa-mir-590	1330
1	L22	22	hsa-mir-23a	827
2	L37	37	hsa-mir-4443	662
3	L30	30	hsa-mir-3691	448
4	L73	73	hsa-mir-877	447
5	L27	27	hsa-mir-331	427
6	L63	63	hsa-mir-6515	396
7	L60	60	hsa-mir-629	376
8	L48	48	hsa-mir-4764	355
9	L72	72	hsa-mir-7850	273
10	L0	0	hsa-let-7e	256
11	L40	40	hsa-mir-4523	238
12	L8	8	hsa-mir-1289-1	213
13	L4	4	hsa-mir-1255a	212
14	L69	69	hsa-mir-6888	211
15	L67	67	hsa-mir-6801	206
16	L46	46	hsa-mir-4752	206
17	L39	39	hsa-mir-4487	192
18	L57	57	hsa-mir-5706	179
19	L76	76	hsa-mir-95	171
20	L35	35	hsa-mir-423	166

Step-by-step interpretation workflow

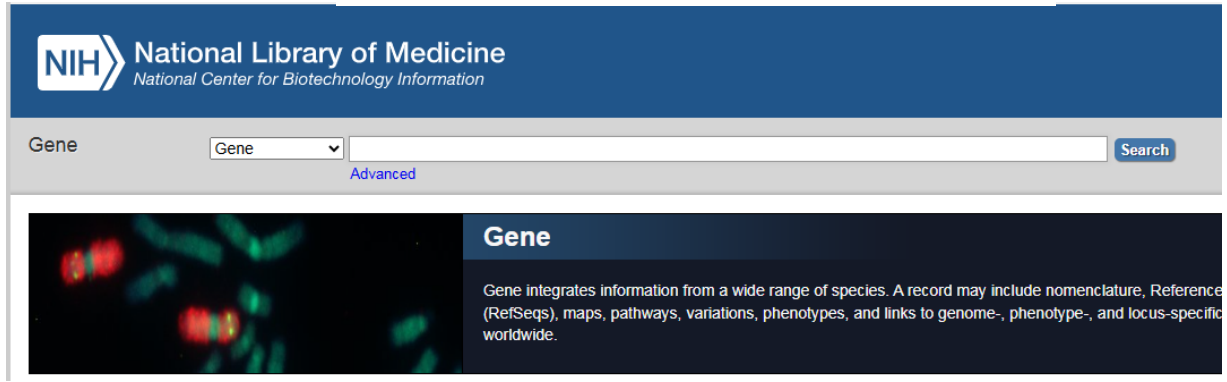
Answer why RNA discovery in a disease is important?

1. Identify RNA and gene annotation
2. Coding *vs* non-coding classification
3. Expression patterns in a disease (HCC)
4. Functional and pathway context
5. Clinical relevance

Step 1: RNA Identification & Annotation

- Gene symbol
- Ensembl ID
- Chromosomal location
- Transcript variants

- **Databases:**
- Ensembl
- NCBI Gene



Never proceed without correct annotation.

Step 2: Coding *vs* Non-Coding

- Protein-coding?
- lncRNA?
- miRNA?
- Pseudogene?

Databases:

- NCBI RefSeq
- UniProt
- Ensembl biotype

This determines downstream interpretation entirely.

Step 3: Expression analysis

- Tumor *vs* normal liver
- Early *vs* late-stage HCC
- Etiology-specific expression (HBV, NASH)
- Cell-type specificity

Databases:

- GDC / TCGA-LIHC
- GTEx
- GEO

Contextual expression matters more than fold-change alone.

Step 4: Functional & Pathway Context

- Known molecular functions
- Pathway involvement
- Protein–protein interactions
- Regulatory roles

Tools:

- KEGG
- Reactome
- STRING

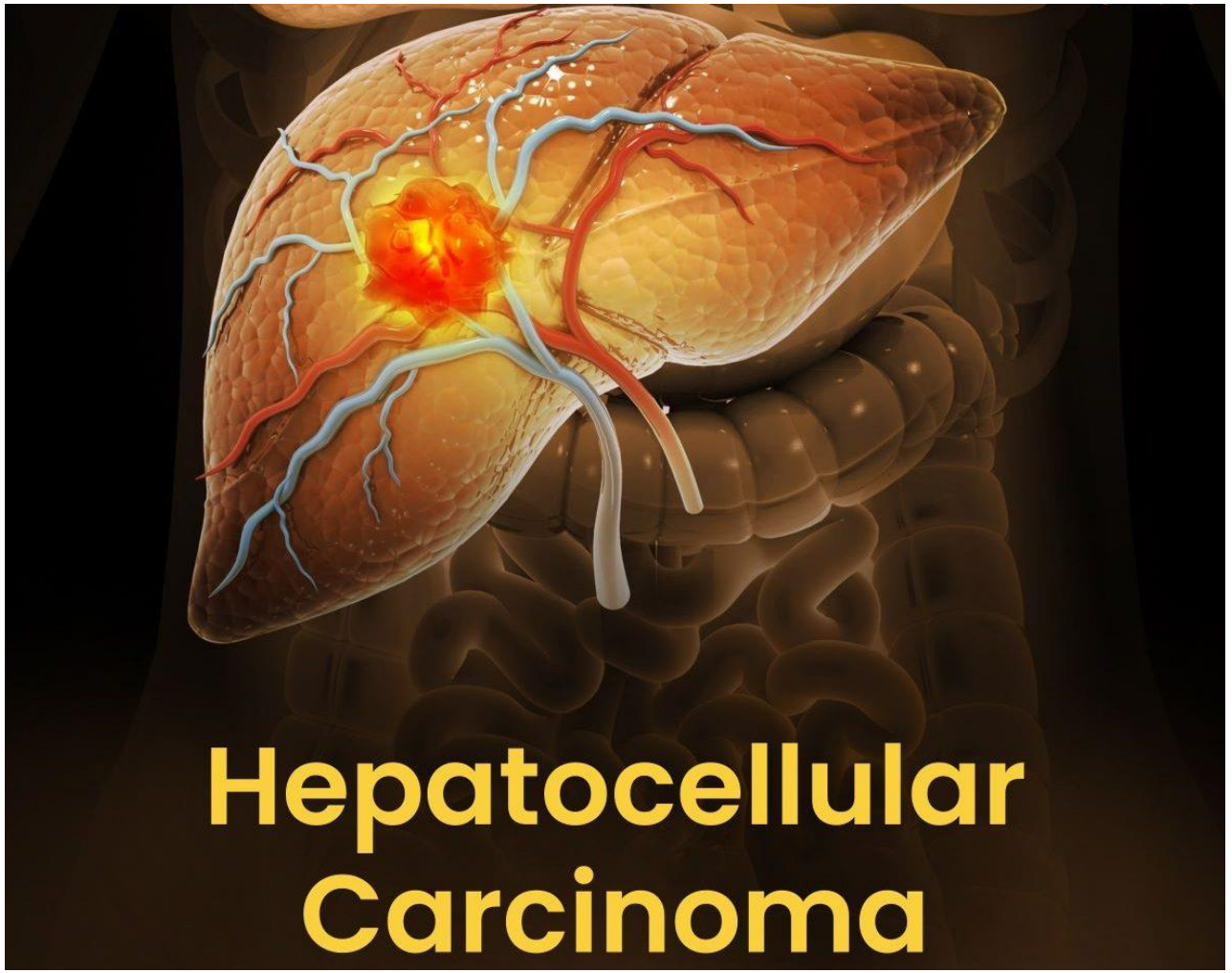
Map the RNA into known biological systems.

Step 5: Clinical Interpretation

- Biomarker potential
- Prognostic vs diagnostic
- Drug targetability
- Circulating RNA/protein detectability

Clinical relevance is the final filter.

An example: HCC



Why HCC Needs AI-Based RNA Discovery

- 6th most common cancer worldwide
- 3rd leading cause of cancer-related death
- Late diagnosis is common
- High molecular heterogeneity

HCC is often diagnosed late. Early molecular biomarkers (especially RNA-based) are urgently needed, which is why AI-based discovery is attractive.

Essential Background: Pathology

Before interpreting any RNA, you must understand the disease context.

- Originates from hepatocytes
- Most common primary liver cancer

- HCC typically arises in chronically injured livers:

✓Chronic hepatitis B or C
✓Alcoholic liver disease
✓Non-alcoholic steatohepatitis (NASH)

- Progression: Inflammation → Fibrosis → Cirrhosis → HCC

- Microenvironment: Fibrosis, immune dysregulation, angiogenesis

This chronic inflammatory environment shapes the transcriptome.

HCC Pathophysiology

- Chronic injury → inflammation → fibrosis → cirrhosis
- Repeated hepatocyte regeneration
- Accumulation of genetic and epigenetic alterations
- Clonal selection of malignant cells

RNA expression profiles capture these progressive molecular adaptations.

Essential Background:

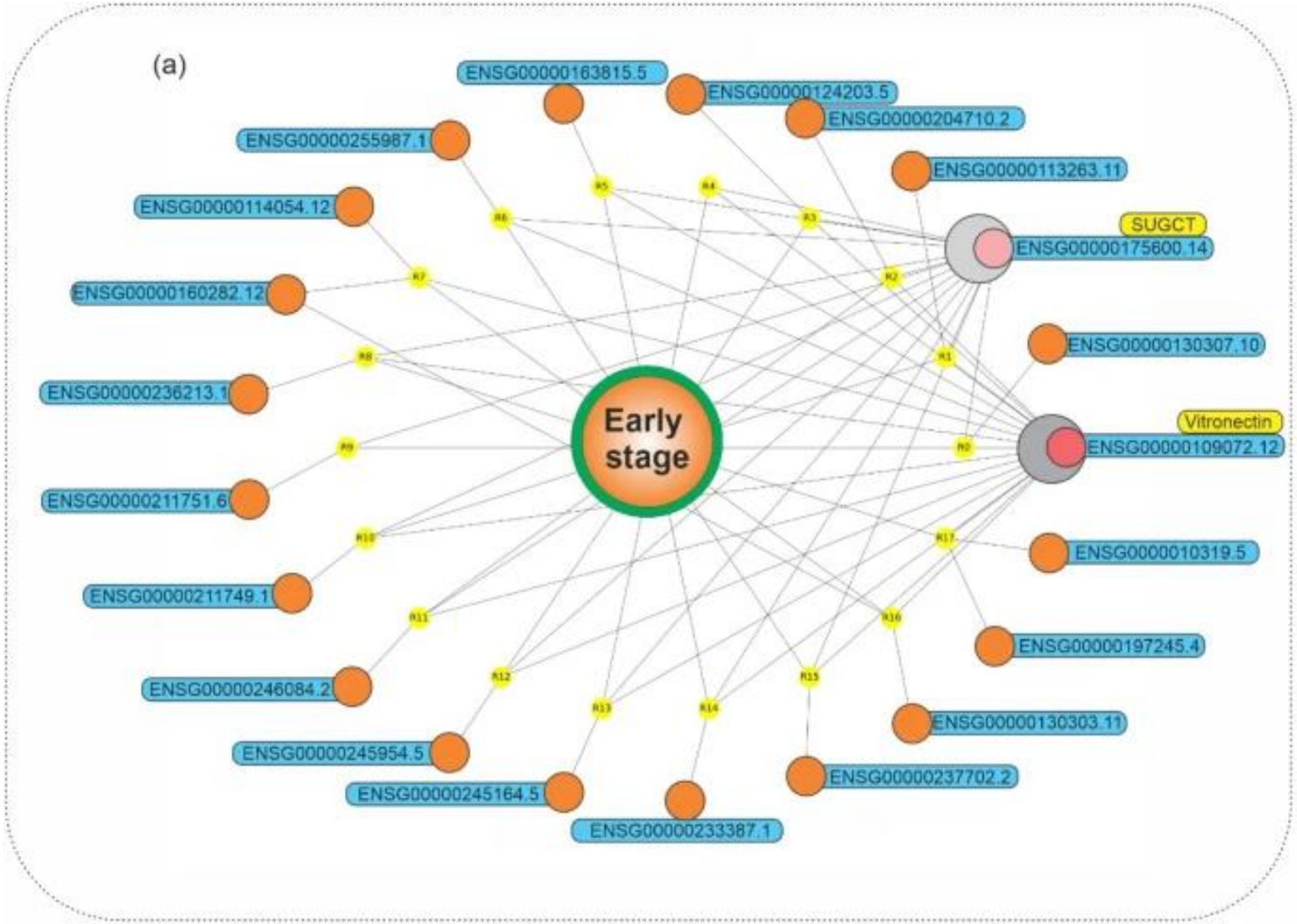
Key Molecular Signaling Pathways in HCC

Pathway	Importance in HCC
WNT/β-catenin	Tumor initiation, stemness
PI3K–AKT–mTOR	Growth, metabolism
MAPK/ERK	Proliferation
TGF-β	EMT, fibrosis, immune evasion
Hippo/YAP-TAZ	Liver size control, oncogenesis
p53	Genomic instability & inflammation
NF-κB	Inflammation-driven tumorigenesis

Any RNA candidate must be evaluated in the context of these pathways.

Example: HCC

- ✓ENSG00000109072
- ✓ENSG00000175600



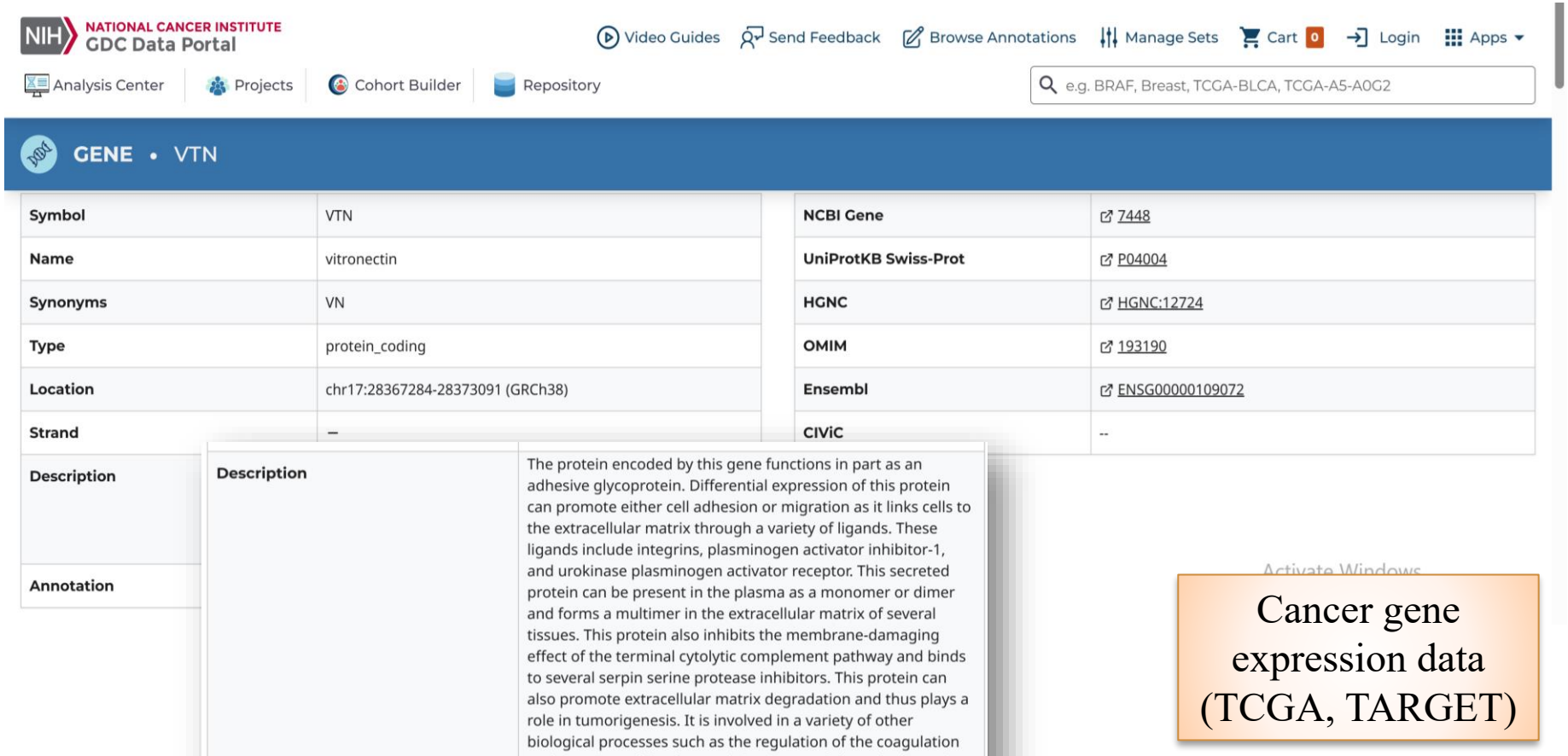
Example: HCC


Google
ENSG00000109072
Ensembl
https://www.ensembl.org › Homo_sapiens › Summary
Gene: VTN (ENSG00000109072) - Summary
Gene: VTN ENSG00000109072 ... Chromosome 17: 28,367,265-28,373,091 reverse stran
gene has 14 transcripts (splice variants), 261 orthologues and 1 ... Read more
Ensembl
http://www.ensembl.org › Homo_sapiens › Summary
Gene: VTN (ENSG00000109072) - Summary
NIH National Library of Medicine
National Center for Biotechnology Information
GEO Profiles
GEO Profiles
ENSG00000109072
Create alert Advanced
Summary
VTN - Intraportal tacrolimus flushing effect on liver allograft
Annotation: VTN, vitronectin
Organism: Homo sapiens
Reporter: GPL11209, ENSG00000109072 (ID_REF), GDS4459, NM_000638
DataSet type: Expression profiling by array, transformed count, 17 samples
ID: 112435600
GEO DataSets Gene Profile neighbors Chromosome neighbors Homologene neighbors
NIH National Library of Medicine
National Center for Biotechnology Information
Gene
Gene
ENSG00000109072
Create RSS Save search Advanced
Full Report
Showing Current items.
VTN vitronectin [Homo sapiens (human)]
Gene ID: 7448, updated on 25-Nov-2025
Summary
Official Symbol VTN provided by HGNC
Official Full Name vitronectin provided by HGNC
Primary source HGNC:HGNC:12724
See related Ensembl:ENSG00000109072 MIM:193190; AllianceGenome:HGNC:12724
Gene type protein coding
RefSeq status REVIEWED
Organism Homo sapiens
Lineage Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Hominidae; Homo
Also known as VN; V75; VNT
Summary The protein encoded by this gene functions in part as an adhesive glycoprotein. Differential expression of this protein
adhesion or migration as it links cells to the extracellular matrix through a variety of ligands. These ligands include int
inhibitor-1, and urokinase plasminogen activator receptor. This secreted protein can be present in the plasma as a mo
multimer in the extracellular matrix of several tissues. This protein also inhibits the membrane-damaging effect of the
pathway and binds to several serpin serine protease inhibitors. This protein can also promote extracellular matrix deg
role in tumorigenesis. It is involved in a variety of other biological processes such as the regulation of the coagulation
tissue remodeling. The heparin-binding domain of this protein give it anti-microbial properties. It is also a lipid binding
component of high density lipoprotein. [provided by RefSeq, Aug 2020]

Example: HCC

- The Genomic Data Commons (GDC)

- Secreted extracellular matrix glycoprotein
- Roles in cell adhesion, migration, coagulation, complement system



Cancer gene expression data (TCGA, TARGET)

Example: HCC

Step 1: Annotation of VTN
- Gene: **VTN (Vitronectin)**
- Ensembl ID: ENSG00000109072
- Chromosome 17
- Secreted extracellular protein

VTN is well-annotated, reducing obscurity.

Step 2: Coding Status
- Protein-coding gene
- UniProt-confirmed glycoprotein
- Abundant in plasma and ECM

This immediately raises biomarker potential.

Step 3: Expression in HCC
- TCGA-LIHC: elevated in tumors
- Particularly increased in early-stage HCC
- Associated with fibrotic microenvironment

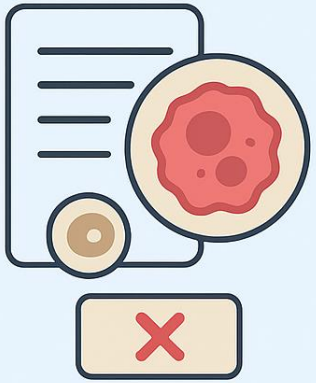
Early-stage enrichment is key here.

Scenario-based interpretation

Published Articles on RNA in Hepatocellular Carcinoma

1. Review the literature
2. Evaluate the findings

RNA in Other Cancers or Diseases

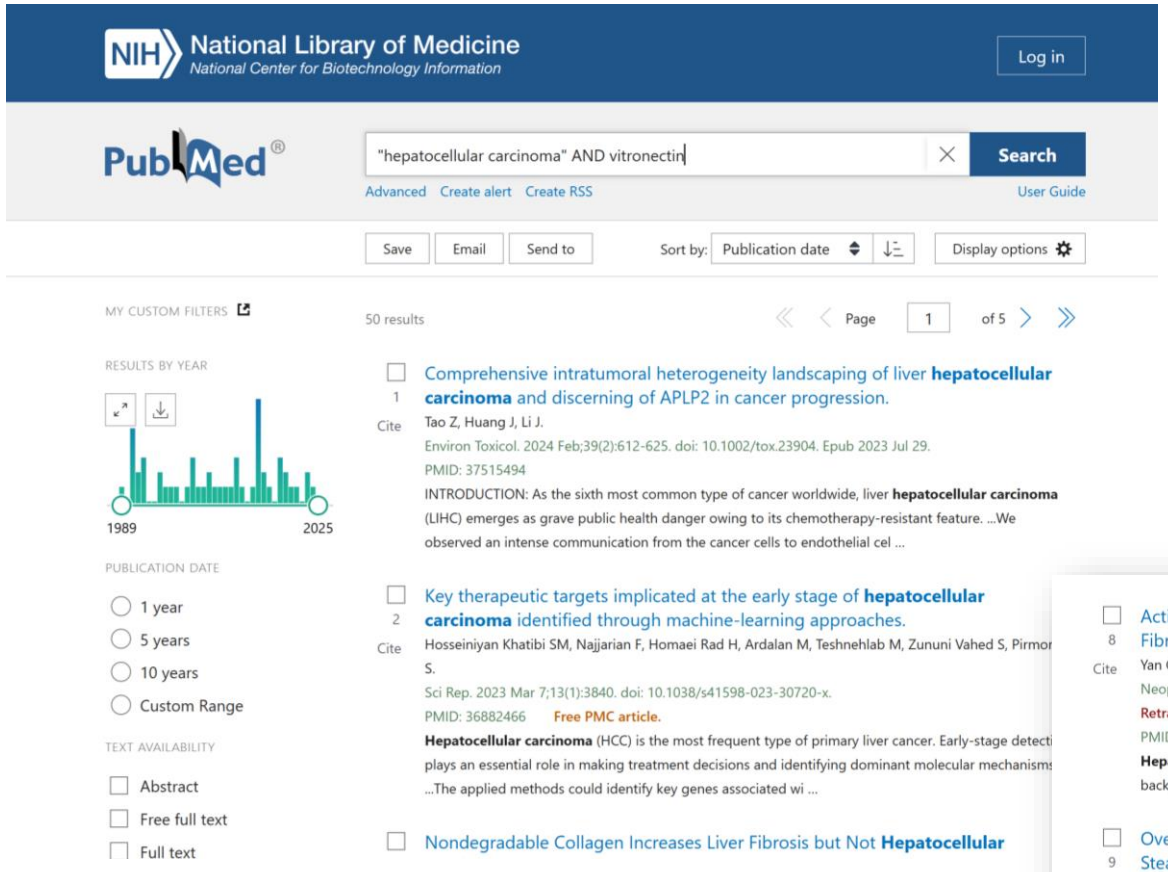


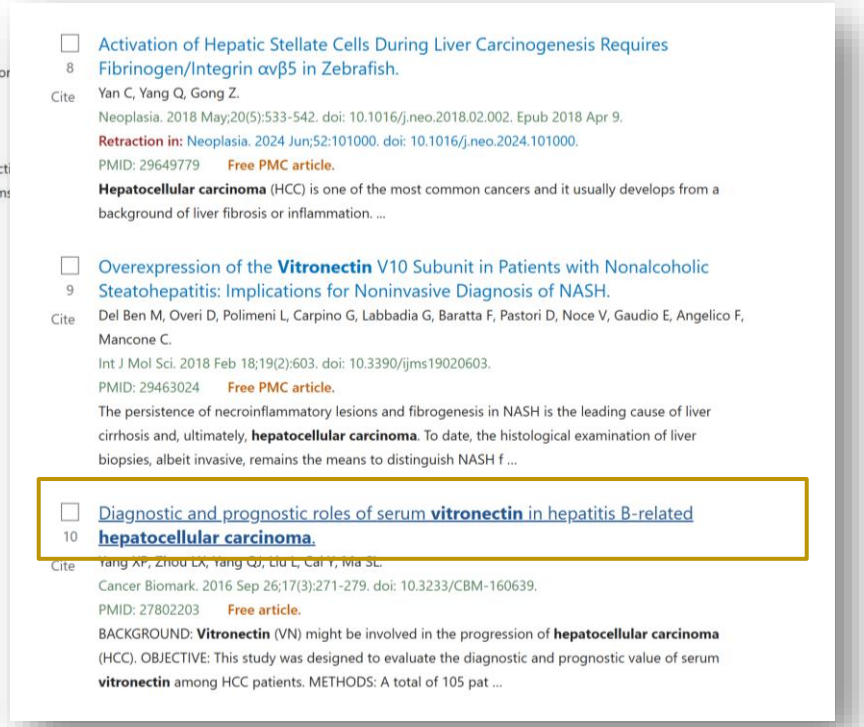
1. Assess relevance
2. Consider hypothesis

Lack of Published Articles

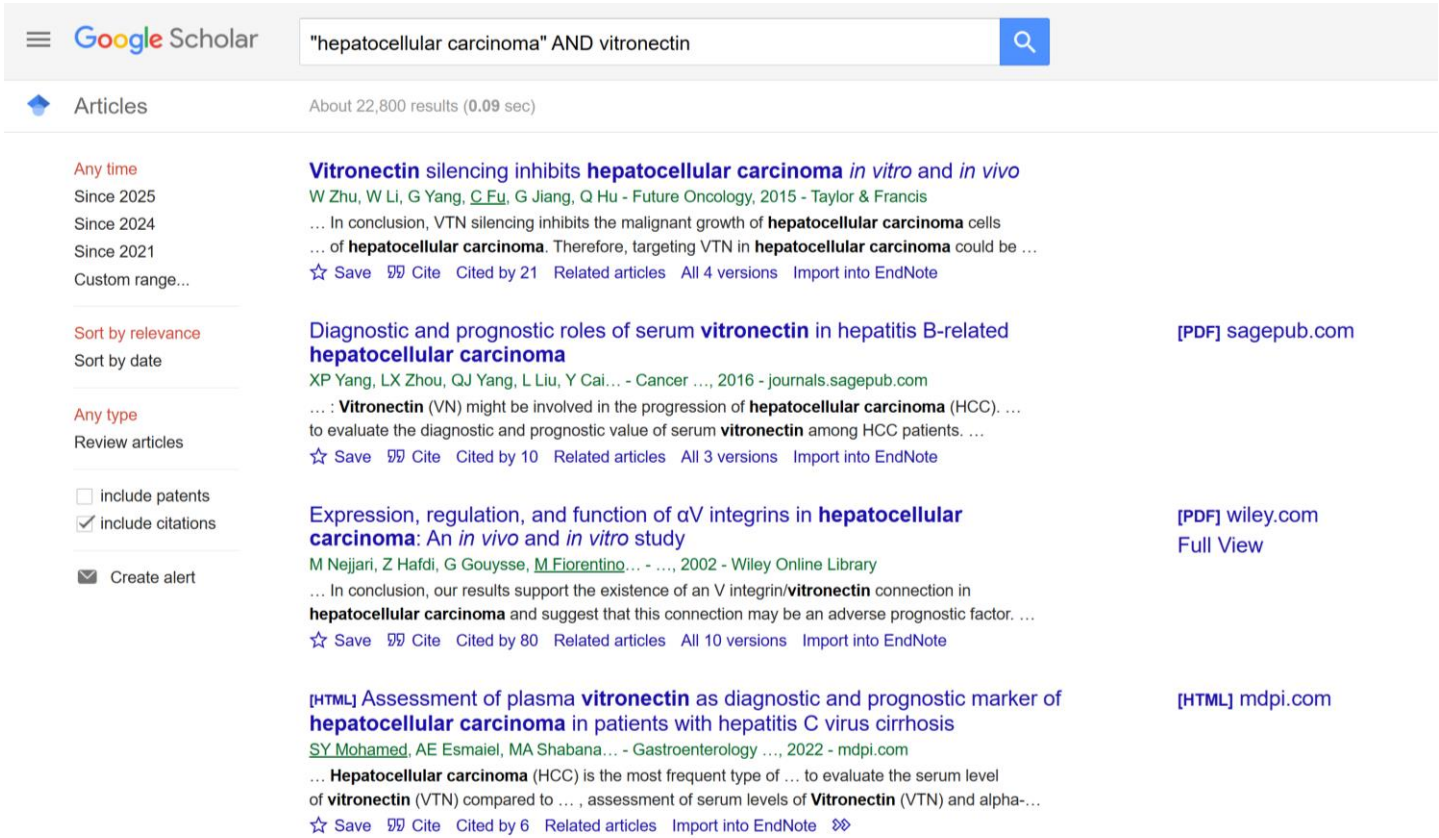
1. Explore basic functions
2. Plan novel studies

Scenario 1-HCC





Scenario 1-HCC



Published articles already report a role of Vitronectin in HCC

- The most common mistake is assuming:

  “Because it’s published, it’s true and complete.”

- Your first action is critical appraisal, not acceptance.

Scenario 1: RNA Published in HCC

1.Read the literature critically

- Mechanism
- Experimental models
- Clinical cohort size

2.Compare:

- Reported direction of expression
- Your dataset's findings

3.Validate consistency:

- Prognostic relevance?
- Functional assays?

Note: **Your model did not "discover" the RNA—it rediscovered it with computational evidence.**

- Subtype specificity
- Interaction networks
- Prediction of therapeutic response
- Early vs late-stage relevance

Before interpreting biology, categorize the literature:

A. Nature of the studies:

- Bioinformatic only?
- Expression-based?
- Functional?
- Clinical cohort analyses?

B. Level of evidence

- Multi-cohort patient validation
- Functional perturbation with phenotypic effects
- Correlative expression studies
- Single-dataset mining

Map agreement and tension in the literature.

- Do all studies report: The same direction of Vitronectin expression?
- The same prognostic implication?

Are findings:

- Consistent?
- Context-dependent?
- Contradictory?

Distinguish correlation from causation explicitly

Even if published work exists, you must ask:
Does Vitronectin drive HCC behavior, or does HCC alter Vitronectin?
state this uncertainty openly.

“While Vitronectin expression correlates with tumor progression, whether it plays
a causal role or reflects adaptive stress responses remains unresolved.”
This protects scientific credibility.

Identify gaps left by existing studies

This is the most important step.

- What did the published studies not address?
  - Mechanism?
  - Tumor microenvironment?
  - Therapy response?
  - Temporal dynamics?

- you do not repeat known findings — you extend them.

Frame conclusions with precision

❌ “Vitronectin plays an important role in HCC”

✅ “Current evidence supports a context-dependent role for Vitronectin at early stage of HCC”

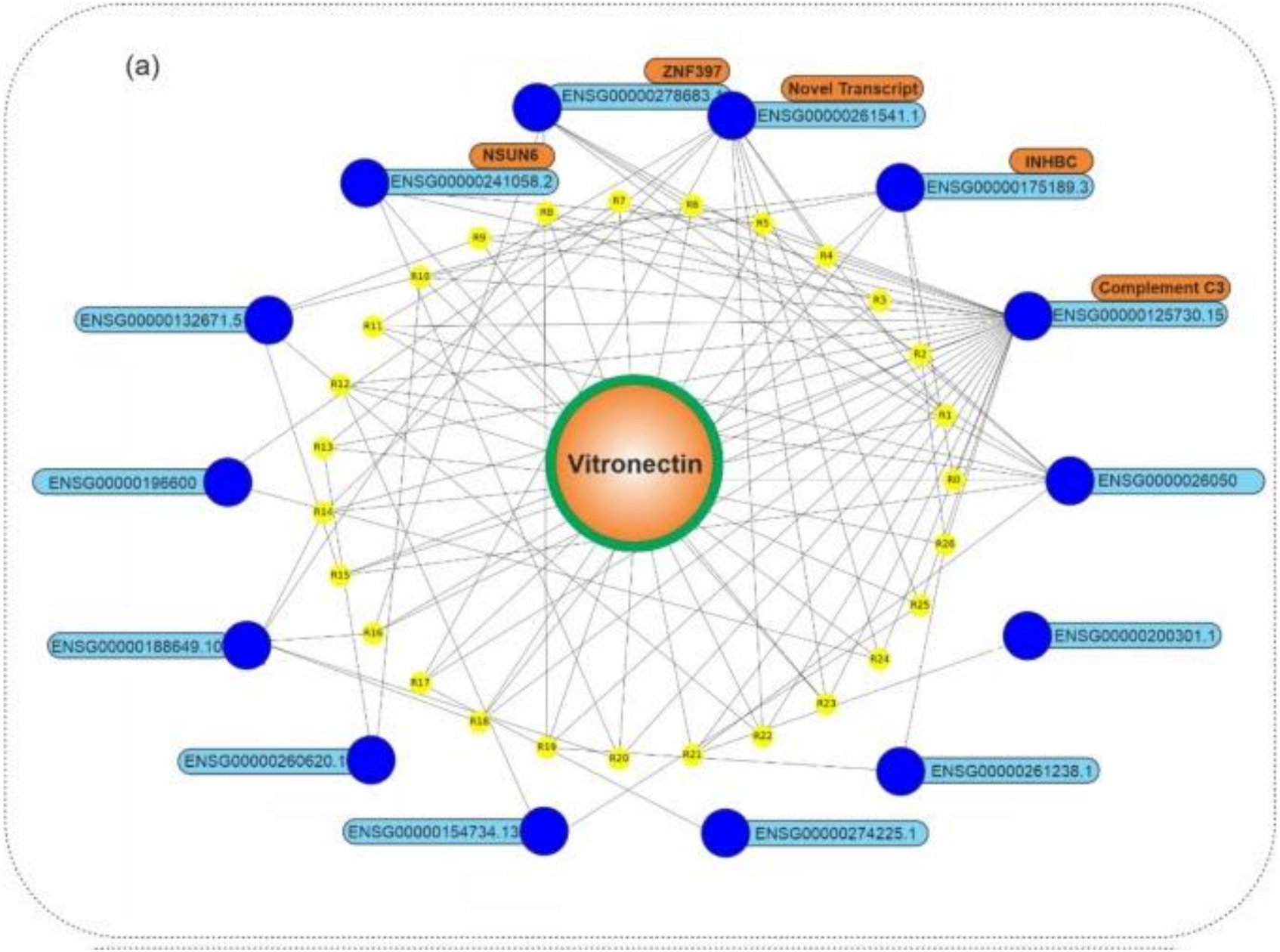
Step 4: Functional Context
•ECM remodeling
•Integrin binding
•Cell adhesion and migration
•Tumor–stroma interaction

These functions are central to HCC progression.

Step 5: Clinical interpretation
•Potential **early diagnostic biomarker**
•Detectable at protein level (serum)
•Reflects tumor microenvironment activity

This makes VTN clinically attractive.

Scenario 1-HCC



Scenario 2: RNA Known in Other Cancers

- Study function in other diseases
  1. Oncogene or tumor suppressor?
  2. Tissue-specific effects?

- Evaluate liver-specific relevance
  1. Expression in normal organ?
  2. Known organr-enriched transcription factors?

- Hypothesize **mechanistic transferability**
  1. Same pathway?
  2. Different microenvironment?

- Propose conserved or context-specific roles

scientific reports

Explore content ▾ About the journal ▾ Publish with us ▾

nature › scientific reports › articles › article

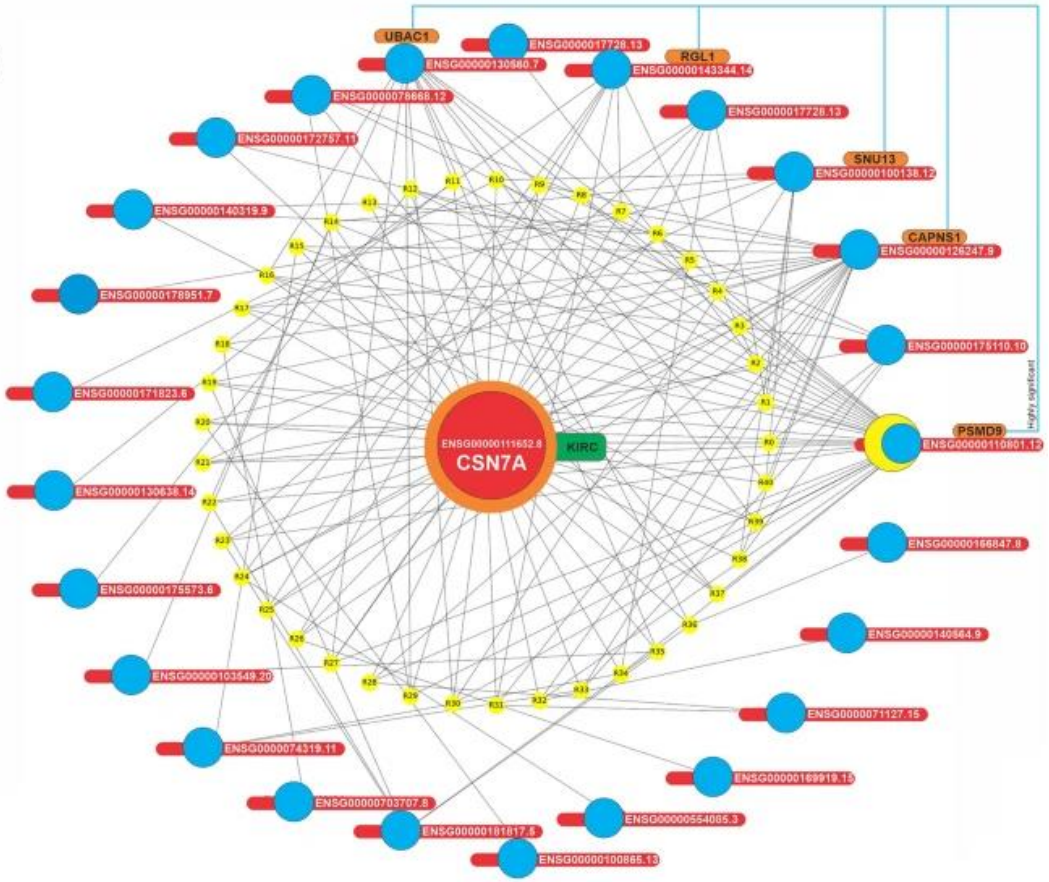
Panels of mRNAs and miRNAs for decoding molecular mechanisms of Renal Cell Carcinoma (RCC) subtypes utilizing Artificial Intelligence approaches

Seyed Mahdi Hosseiniyan Khatibi, Mohammadreza Ardalan, Mohammad Teshnehlab, Sepideh Zununi Vahed & Saeed Pirmoradi

Article | Open access | Published: 30 September 2022

Scientific Reports **12**, Article number: 16393 (2022) | Cite this article

(b)



Scenario 2

UniProt BLAST Align Peptide search ID mapping SPARQL UniProtKB Advanced | List Search Help

- Function
- Names & Taxonomy
- Subcellular Location
- Disease & Variants
- PTM/Processing
- Expression
- Interaction
- Structure
- Family & Domains
- Sequence
- Similar Proteins

Q9UBW8 · CSN7A_HUMAN

Protein	COP9 signalosome complex subunit 7a	Amino acids	275 (go to sequence)
Gene	COPS7A	Protein existence	Evidence at protein level
Status	UniProtKB reviewed (Swiss-Prot)	Annotation score	5/5
Organism	Homo sapiens (Human)		

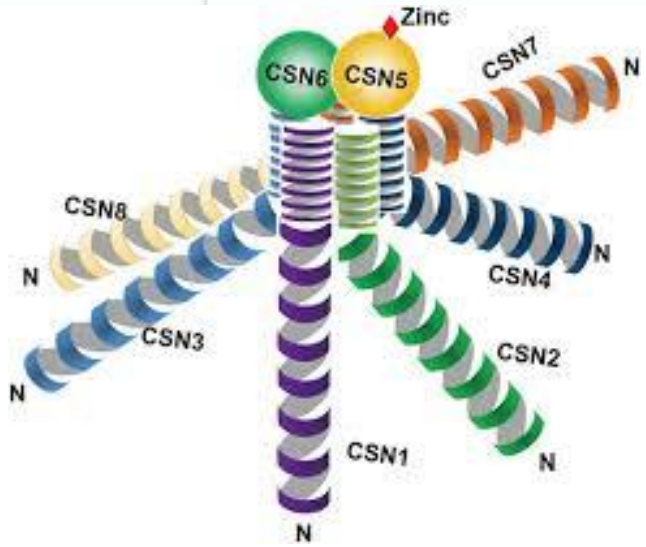
Entry | Variant viewer 270 | Feature viewer | Genomic coordinates | Publications | External links | History

Tools | Download | Add | Add a publication | Entry feedback

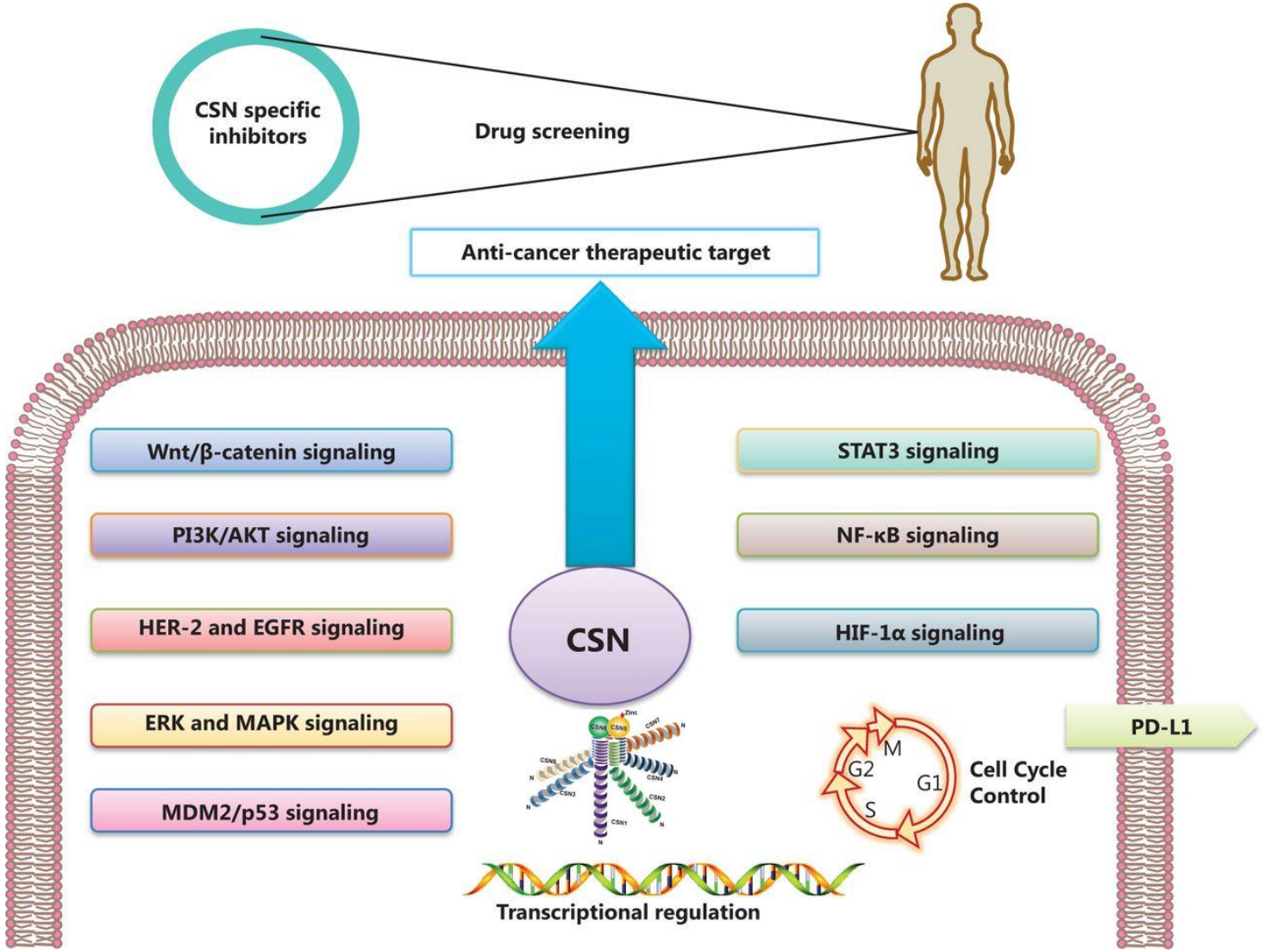
Function

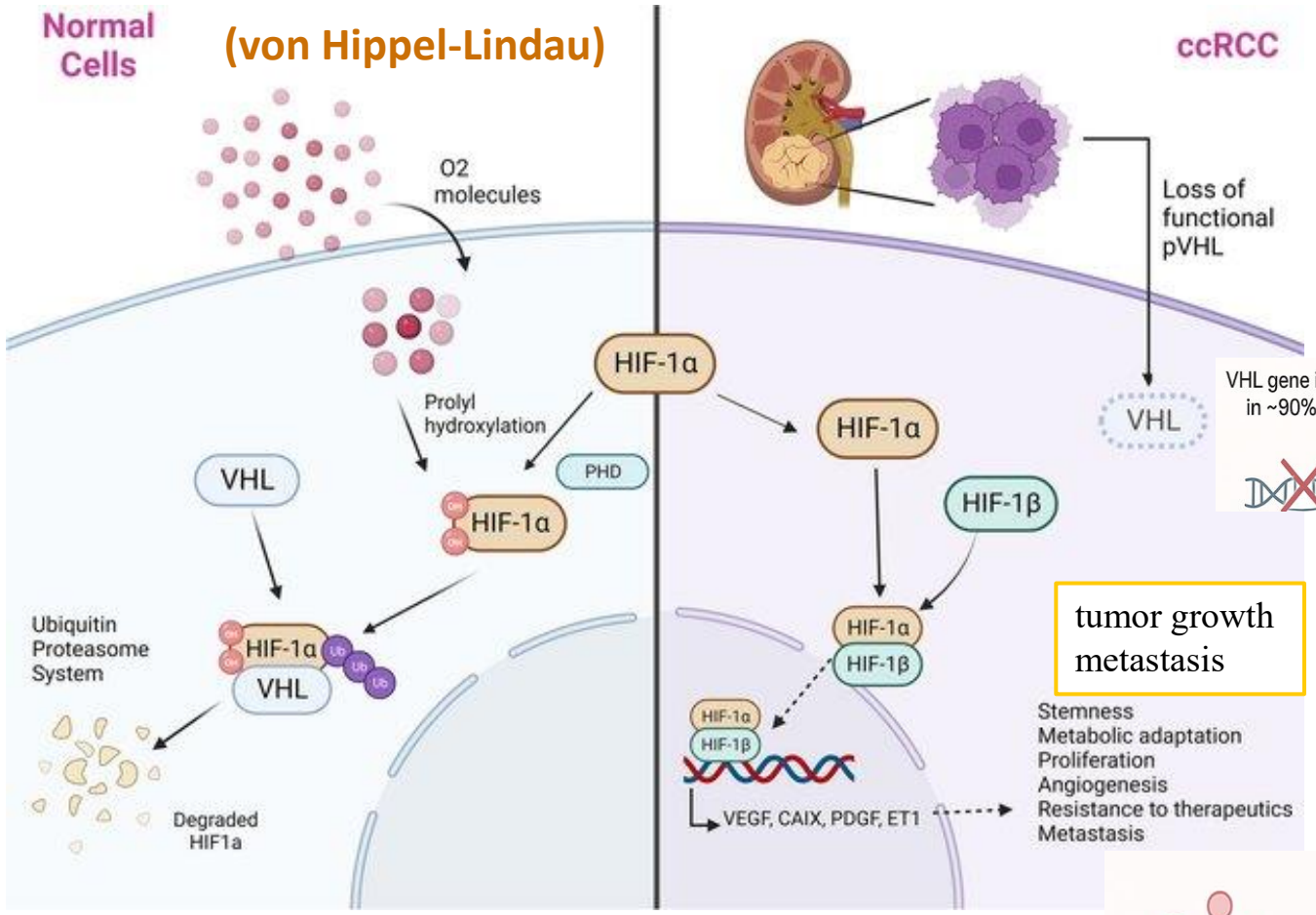
Component of the COP9 signalosome complex (CSN), a complex involved in various cellular and developmental processes. The CSN complex is an essential regulator of the ubiquitin (Ubl) conjugation pathway by mediating the deneddylation of the cullin subunits of SCF-type E3 ligase complexes, leading to decrease the Ubl ligase activity of SCF-type complexes such as SCF, CSA or DDB2. The complex is also involved in phosphorylation of p53/TP53, JUN, I-kappa-B-alpha/NFKBIA, ITPK1 and IRF8/ICSBP, possibly via its association with CK2 and PKD kinases. CSN-dependent phosphorylation of TP53 and JUN promotes and protects degradation by the Ubl system, respectively. 5 Publications

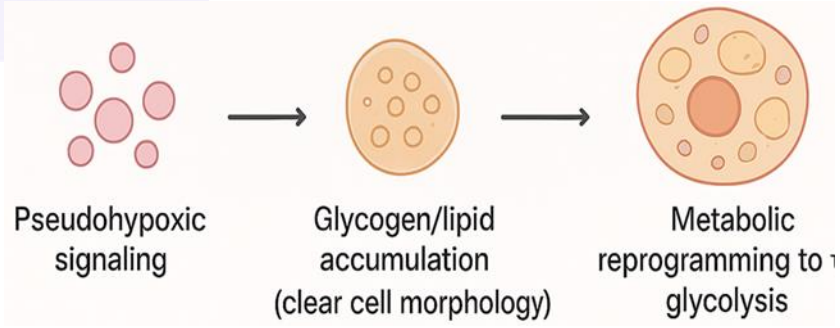
Help

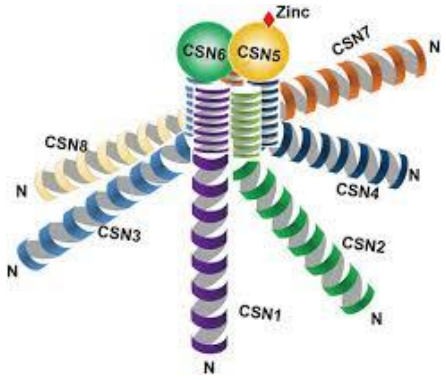


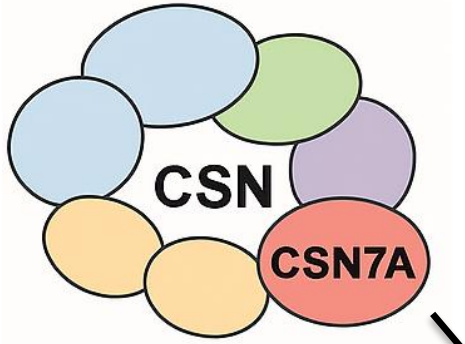
Scenario 2


CSN specific inhibitors
Drug screening
Anti-cancer therapeutic target
Wnt/β-catenin signaling
PI3K/AKT signaling
HER-2 and EGFR signaling
ERK and MAPK signaling
MDM2/p53 signaling
STAT3 signaling
NF-κB signaling
HIF-1α signaling
CSN
PD-L1
M
G2
G1
S
Cell Cycle Control
Transcriptional regulation

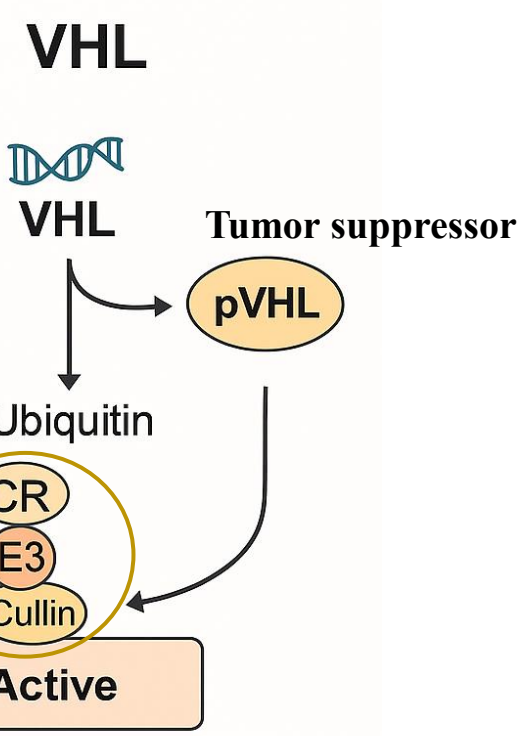

Normal Cells
(von Hippel-Lindau)
ccRCC
O2 molecules
Loss of functional pVHL
HIF-1α
Prolyl hydroxylation
PHD
VHL
HIF-1α
VHL gene inactivation in ~90% of cases
HIF-1α
HIF-1β
Ubiquitin Proteasome System
HIF-1α
VHL
HIF-1α
HIF-1β
tumor growth
metastasis
HIF-1α
HIF-1β
Stemness
Metabolic adaptation
Proliferation
Angiogenesis
Resistance to therapeutics
Metastasis
VEGF, CAIX, PDGF, ET1
Degraded HIF1a


Pseudohypoxic signaling
Glycogen/lipid accumulation (clear cell morphology)
Metabolic reprogramming to
glycolysis





cullin-RING E3 ubiquitin ligases **CRLs**

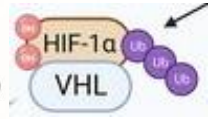


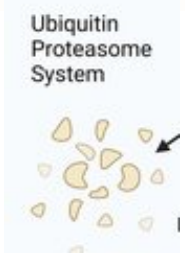
Inactive

p53 Cyclin E HIF-1α

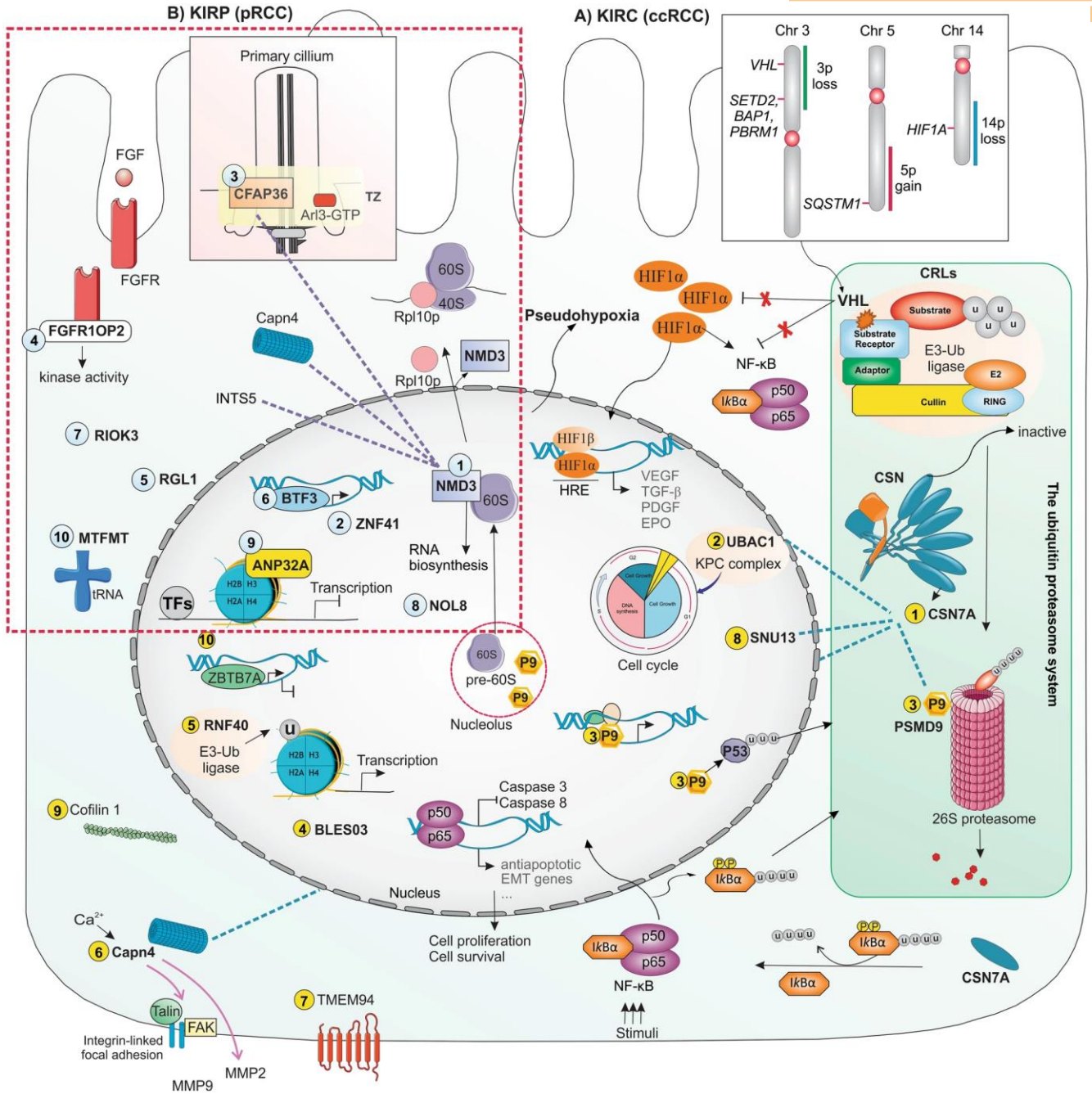
p53 Cyclin E HIF-1α

p53 Cyclin E HIF-1α





Scenario 2


B) KIRP (pRCC)
A) KIRC (ccRCC)
Primary cillium
3 CFAP36
TZ
Arl3-GTP
FGF
FGFR
4 FGFR1OP2
kinase activity
Capn4
60S
40S
Rpl10p
NMD3
INTS5
7 RIOK3
5 RGL1
6 BTF3
2 ZNF41
1 NMD3
60S
10 MTFMT
tRNA
9 ANP32A
TFs
Transcription
RNA biosynthesis
8 NOL8
10 ZBTB7A
60S P9 pre-60S P9
Nucleolus
5 RNF40
E3-Ub ligase
Transcription
9 Cofilin 1
4 BLES03
p50 p65
Caspase 3
Caspase 8
antiapoptotic EMT genes
Nucleus
Cell proliferation
Cell survival
Ca2+
6 Capn4
Talin
FAK
Integrin-linked focal adhesion
MMP9
MMP2
7 TMEM94
Chr 3
Chr 5
Chr 14
VHL
3p loss
SETD2, BAP1, PBRM1
5p gain
HIF1A
14p loss
SQSTM1
Pseudohypoxia
HIF1α
HIF1β
HRE
VEGF
TGF-β
PDGF
EPO
NF-κB
IκBα
p50
p65
CRLs
VHL
Substrate
Substrate Receptor
Adaptor
E3-Ub ligase
E2
Cullin
RING
inactive
CSN
2 UBAC1
KPC complex
Cell cycle
8 SNU13
1 CSN7A
3 P9
PSMD9
26S proteasome
P53
The ubiquitin proteasome system
PP
IκBα
NF-κB
Stimuli
CSN7A

Your interpretation should be conservative and defensible:

“Based on its known roles in protein turnover and cell-cycle regulation in other malignancies, CSN7A may modulate stress adaptation and survival pathways relevant to ccRCC biology.”

“Given its established role in disease X via pathway Y, RNA Z may contribute to disease A through a shared biological process.”

Due to a frequent mutation in protein-coding regions and an elevated burden of unfolded proteins; an elevated protein turnover was necessary for those speedily dividing cancer cells. Hence, the inhibition of the UPS components appeared to be a hopeful strategy for KIRC therapy.

Scenario 3: Completely Novel RNA

- Verify annotation quality
- Check conservation
- Evaluate tissue specificity
- Use computational predictions
- Design validation experiments

Novelty increases uncertainty—but also impact.

Look for indirect cancer logic

- Forget cancer labels and return to fundamentals:
  What does X do in normal cells?

- Many cancer-relevant genes were discovered because they:
- Regulate stress responses
- Maintain protein homeostasis
- Stabilize signaling networks

- In silico analysis
- Bench validation
- Orthogonal data
- Iterative refinement

Be extremely disciplined in interpretation

- Avoid declaring oncogenic or suppressive roles prematurely
- Use conditional language:
- “may”
- “suggests”
- “is consistent with”

- This protects scientific integrity.

Scenario 3

Based on the available literature and bioinformatics analysis.

SPRINGER NATURE Link

Log in

Find a journal | Publish with us | Track your research | Search | Cart

Home > Journal of Cancer Research and Clinical Oncology > Article

A diagnostic miRNA panel to detect recurrence of ovarian cancer through artificial intelligence approaches

Research | Published: 15 November 2022

Volume 149, pages 325–341, (2023) Cite this article

Journal of Cancer Research and Clinical Oncology

Aims and scope →

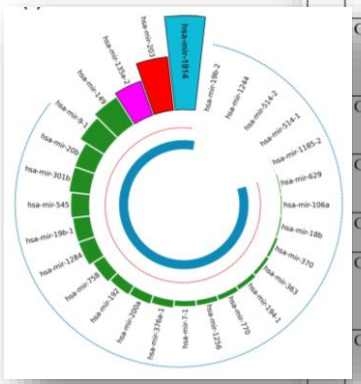
Submit manuscript →

Reyhaneh Aghayousefi, Seyed Mahdi Hosseiniyan Khatibi, Sepideh Zununi Vahed, Milad Bastami, Saeed Pirmoradi & Mohammad Teshnehlab

Part of a collection: Precision Oncology

Table S5. Top ten enriched gene ontology (GO) targeted by miR-1914 in the recurrence of ovary cancer

	miR-1914 Gene Set	Description	Size	Ratio	P-value	FDR	Gene counts	Genes
1	GO:0048870	cell motility	1518	2.3900	0.000022153	0.069538	26	APEX1, ARHGEF2, ARPC3, ARRB2, BARHL1, CALR, CFL1, DMTN, MSN, NAV1, NRG4, NUP93, OGDH, OLR1, PLCG1, PLXND1, PPIA, PPP1R9B, SLC16A1, SP1, SRCIN1, STAT5B, TACR3, VASH1, WASF2, ZNF609
2	GO:0051674	localization of cell	1518	2.3900	0.000022153	0.069538	26	APEX1, ARHGEF2, ARPC3, ARRB2, BARHL1, CALR, CFL1, DMTN, MSN, NAV1, NRG4, NUP93, OGDH, OLR1, PLCG1, PLXND1, PPIA, PPP1R9B, SLC16A1, SP1, SRCIN1, STAT5B, TACR3, VASH1, WASF2, ZNF609
3	GO:0040011	Locomotion	1738	2.3283	0.000010891	0.069538	29	APEX1, ARHGEF2, ARPC3, ARRB2, BARHL1, CALR, CFL1, CREB1, DMTN, MSN, NAV1, NRG4, NUP93, OGDH, OLR1, PLCG1, PLXND1, PPIA, PPP1R9B, RNF165, SHANK3, SLC16A1, SP1, SRCIN1, STAT5B, TACR3, VASH1, WASF2, ZNF609
	GO:0006928	movement of cell or subcellular component	1967	2.1282	0.000041981	0.098835	30	APEX1, ARHGEF2, ARPC3, ARRB2, BARHL1, CALR, CFL1, DMTN, KIF3B, MSN, NAV1, NOS1AP, NRG4, NUP93, OGDH, OLR1, PLCG1, PLXND1, PPIA, PPP1R9B, RNF165, SHANK3, SLC16A1, SP1, SRCIN1, STAT5B, TACR3, VASH1, WASF2, ZNF609
	GO:0040018	positive regulation of multicellular organism growth	33	16.914	0.000087445	0.15043	4	CREB1, IGF2, PPIB, STAT5B
	GO:0016477	cell migration	1368	2.3460	0.000095845	0.15043	23	APEX1, ARHGEF2, ARRB2, BARHL1, CALR, CFL1, DMTN, MSN, NAV1, NUP93, OGDH, OLR1, PLCG1, PLXND1, PPIA, PPP1R9B, SLC16A1, SP1, SRCIN1, STAT5B, VASH1, WASF2, ZNF609
	GO:0048524	positive regulation of viral process	83	8.4059	0.00032338	0.43504	5	CFL1, LARP1, PPIA, PPIB, SP1
	GO:0031328	positive regulation of cellular biosynthetic process	1919	1.9633	0.00042004	0.49444	27	ADIRF, APEX1, ARHGEF2, ARRB2, BARHL1, CAMTA2, CREB1, CTDNEP1, E2F3, EIF5AL1, FOXK1, IGF2, KMT2B, LARP1, MAFG, MTA1, NFIC, NFIX, NOS1AP, PLAGL2, PLXND1, PTGES2, RBPJL, RNF187, SP1, STAT5B, ZNF609
	GO:0009891	positive regulation of biosynthetic process	1949	1.9330	0.00053645	0.56131	27	ADIRF, APEX1, ARHGEF2, ARRB2, BARHL1, CAMTA2, CREB1, CTDNEP1, E2F3, EIF5AL1, FOXK1, IGF2, KMT2B, LARP1, MAFG, MTA1, NFIC, NFIX, NOS1AP, PLAGL2, PLXND1, PTGES2, RBPJL, RNF187, SP1, STAT5B, ZNF609
10	GO:0003062	regulation of heart rate by chemical signal	6	46.512	0.00074982	0.70611	2	NOS1AP, YWHAE



- **one-sentence expert takeaway for each scenario**

- **Scenario 1:**
  *Your job is to **refine truth**, not repeat literature.*

- **Scenario 2:**
  *Your job is to **translate mechanisms responsibly**, not import cancer labels.*

- **Scenario 3:**
  *Your job is to let **normal biology** guide cancer relevance without speculation.*

Final message

AI does not replace biological thinking.
AI reshapes the hypothesis space.

But interpretation
is where biology, pathology, and medicine take control again.

If you remember one thing from today, remember this:
A candidate RNA is a starting point, not a conclusion

Thank you

Email: said.Pirmoradi@gmail.com
Email: sepide.zununi@gmail.com

Phone: 0914-311-6396

Website: https://khayyamlab.ir