

# Project report & Task Assignment

# Table of Contents

1. Xác định mối liên hệ giữa miRNA và targeted genes
2. cBioportal
3. Mối liên hệ giữa DNA methylation vs genes.
4. Others
5. Phân công tasks

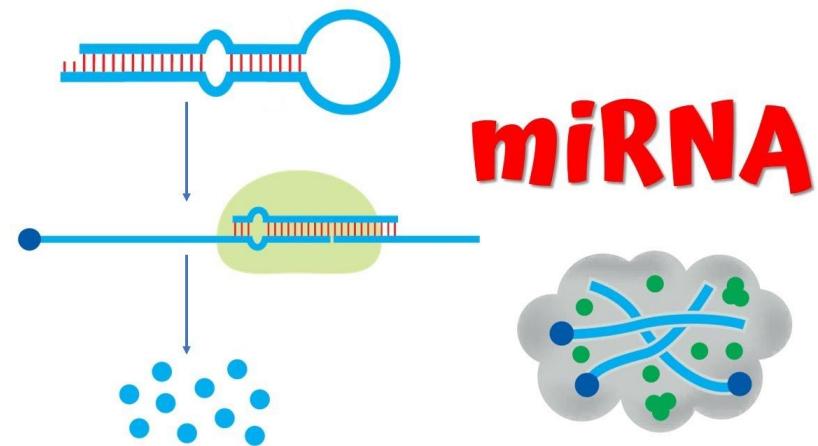
# 1. Xác định mối liên hệ giữa miRNA và targeted genes

## miRNA

- Là các phân tử RNA ngắn (khoảng 21-23 nucleotide)
- Không mã hóa protein
- Điều hòa biểu hiện gen ở mức sau phiên mã

## Target genes

- Là các gene mà miRNA gắn vào và điều hòa
- Một miRNA có thể điều hòa nhiều gen đích
- Một gene có thể bị điều hòa bởi nhiều miRNA khác nhau



# 1. Xác định mối liên hệ giữa miRNA và targeted genes

Cơ chế tác động

Quá trình nhận diện gene đích:

1. miRNA gắn vào vùng 3'UTR của mRNA đích (thường gấp nhất)
2. Sự gắn kết dựa trên nguyên tắc bổ sung base (không hoàn toàn)
3. Vùng "seed region" (nucleotide 2-8) của miRNA đóng vai trò quan trọng trong nhận diện

Kết quả:

- Ức chế dịch mã: miRNA ngăn ribosome dịch mã mRNA thành protein
- Phân hủy mRNA: miRNA gây phân hủy mRNA đích
- Kết quả: Giảm hoặc ngừng sản xuất protein từ gene đích

# 1. Xác định mối liên hệ giữa miRNA và targeted genes

Dữ liệu hiện tại:

Dữ liệu giữa mirna\_id với gene symbol (Có thể chuyển sang Ensembl Gene sau)

Tuy nhiên mirna id ở đây khác với dữ liệu ở TCGA data là “hsa-miR-122-5p” với “hsa-mir-122”.

“5p” là miRNA phát triển từ nhánh 5’

→ Cần tìm hiểu ý nghĩa, có thể tính weight giữa các connection này.

mirna_id, gene_id
hsa-miR-122-5p, SLC7A1
hsa-miR-122-5p, ADAM17
hsa-miR-122-5p, NUMBL
hsa-miR-122-5p, EGLN3
hsa-miR-122-5p, TRIB1
hsa-miR-122-5p, SLC7A11
hsa-miR-122-5p, FOXJ3
hsa-miR-122-5p, XP06
hsa-miR-122-5p, AP3M2
hsa-miR-122-5p, G6PC3
hsa-miR-122-5p, GALNT10
hsa-miR-122-5p, DSTYK
hsa-miR-122-5p, TPD52L2

# 2. cBioportal

Lung Adenocarcinoma (TCGA, GDC)

TCGA Lung Adenocarcinoma. Source data from GDC and generated in Jul 2025 using Cancer Data Aggregator.

Click gene symbols below or enter here  Query

Summary Clinical Data CN Segments Plots Beta Selected: 585 patients | 587 samples Custom Selection Columns Groups Help

587 results Study Page Help

Patient ID	Sample ID	Number of Samples Per Patient	Mutation Count	Fraction Genome Altered	Diagnosis Age	Overall Survival Status	Sample Type	Sex	Ethnicity Category	Race Category	AJCC Pathologic M-Stage	AJCC Pathologic N-Stage	AJCC Pathologic Stage	AJCC Pathologic T-Stage	American Joint Committee on Cancer Publication Version	Biopsy Site	Birt Init Pat Dia Dat
TCGA-55-6987	TCGA-55-6987-01	1	141	0.0776	77	0:LIVING	Primary Tumor	Male	NOT HISPANIC OR LATINO	WHITE	M0	N0	Stage IA	T1	6th	Lung	-28
TCGA-75-5147	TCGA-75-5147-01	1	34	0.4207		0:LIVING	Primary Tumor	Female			M0	N0	Stage IB	T2	6th	Lung	
TCGA-05-4410	TCGA-05-4410-01	1	866	0.3382	62	0:LIVING	Primary Tumor	Male			M0	N0	Stage IB	T2	6th	Lung	-22
TCGA-78-7153	TCGA-78-7153-01	1	88	0.2226	65	0:LIVING	Primary Tumor	Female		WHITE	M0	N0	Stage IB	T2	6th	Lung	-23

## 2. cBioportal

Patient: TCGA-55-6987, Male, 77 years old, Lung Adenocarcinoma (Lung Adenocarcinoma), LIVING (70 months), DiseaseFree (42 months)  
Samples: TCGA-55-6987-01, Primary

Lung Adenocarcinoma (TCGA, GDC)

Summary Pathways Clinical Data Pathology Report Tissue Image Patient Page Help

**RTK-RAS**

```
graph TD; EGFR[EGFR] --- FGFR1[FGFR1]; EGFR --- ERBB2[ERBB2]; EGFR --- ERBB3[ERBB3]; EGFR --- ERBB4[ERBB4]; FGFR1 --- FGFR2[FGFR2]; FGFR1 --- FGFR3[FGFR3]; FGFR1 --- FGFR4[FGFR4]; FGFR2 --- FGFR3; FGFR2 --- FGFR4; FGFR3 --- FGFR4; RET[RET] --- ROS1[ROS1]; RET --- ALK[ALK]; ROS1 --- ALK; MET[MET] --- KIT[KIT]; MET --- NTRK1[NTRK1]; KIT --- IGF1R[IGF1R]; NTRK1 --- NTRK2[NTRK2]; CBL[CBL] --- EERRF11[ERRF11]; SOS1[SOS1] --- NF1[NF1]; SOS1 --- RASA1[RASA1]; SOS1 --- KRAS[KRAS]; SOS1 --- HRAS[HRAS]; SOS1 --- NRAS[NRAS]; SOS1 --- RIT1[RIT1]; PTPN11[PTPN11] --- SOS1; KRAS --- RAF1[RAF1]; KRAS --- BRAF[BRAF]; KRAS --- ARAF[ARAF]; RAF1 --- MEK1[MEK1]; RAF1 --- MEK2[MEK2]; MEK1 --- MAPK1[MAPK1]; MEK2 --- MAPK1; MAPK1 --- RAC1[RAC1]; MAPK1 --- MAP2K1[MAP2K1]; MAPK1 --- MAP2K2[MAP2K2]
```

Pathway name Pathway altered Genes matched

<input checked="" type="radio"/> RTK-RAS	Yes	RIT1 NTRK1 MET
<input type="radio"/> MYC	Yes	MLXIPL MYCL
<input type="radio"/> PI3K	Yes	RICTOR AKT3
<input type="radio"/> HIPPO	Yes	PTPN14
<input type="radio"/> NRF2	Yes	KEAP1
<input type="radio"/> TP53	No	
<input type="radio"/> WNT	No	
<input type="radio"/> TGF-Beta	No	
<input type="radio"/> NOTCH	No	
<input type="radio"/> Cell Cycle	No	

Showing 1-10 of 10  
 Show TCGA PanCancer Atlas pathways only

### 3. Mối liên hệ giữa cpG (DNA methylation) với genes

Nhóm các cpGs theo exon để giảm kích thước của đồ thị

## 4. Các phương pháp khác

Xây dựng các mô hình nhận đầu vào là dữ liệu đã được xử lý (4 omics data

- đã được chọn lọc dựa trên paper CtAE)

Thực hiện một vài phân tích dữ liệu

Đầu ra là các phân cụm bệnh nhân (clustering).

Đánh giá kết quả phân cụm dựa vào C-index

Các phương pháp có thể tham khảo:

1. PCA + Kmeans
2. AutoEncoder (variants) → Kmeans

## 5. Phân công công việc

1. Mối quan hệ giữa cpgs và genes - Lê Tùng Lâm
2. Mối quan hệ giữa miRNA với genes - Khuất Đức Bảo
3. Nghiên cứu cBioportal và các vấn đề liên quan: Nguyễn Hải Anh
4. Các phương pháp khác: