Software Capstone Design
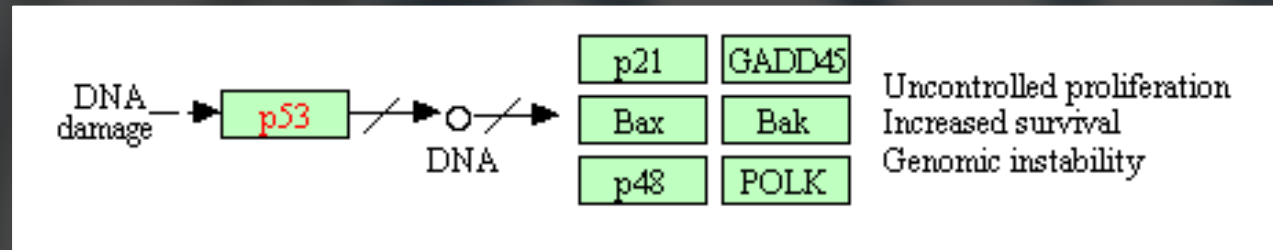
Time-Series Gene Expression Data Analysis
to Identify Drug-Response Specific Gene Networks

Melon Iceflake
(Melon Bingsu)

Finding the gene mixtures that have similarity

How?

Matrix Factorization

2.Method

Latent Space

2.Method

| Matrix Factorization

Latent Space

User들의 영화들에 대한 평점

V

	User 1	User 2	User 3
영화 1	2	3	4
영화 2	5	1	2
영화 3	1	2	3
영화 4	2	4	3

4x3

=

Latent Space

W

	Feature x1	Feature x2
영화 1		
영화 2		
영화 3		
영화 4		

4xN

●

H

	User 1	User 2	User 3
Feature x1			
Feature x2			

Nx3
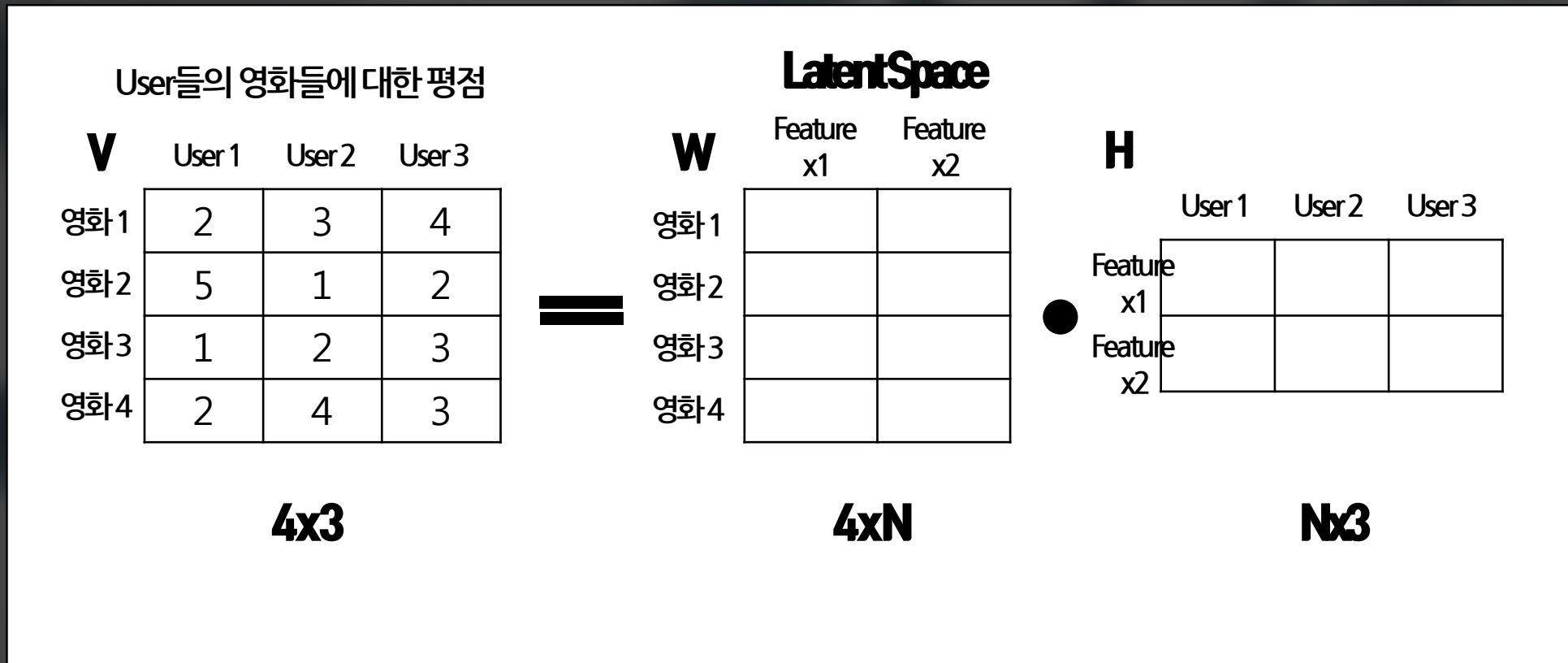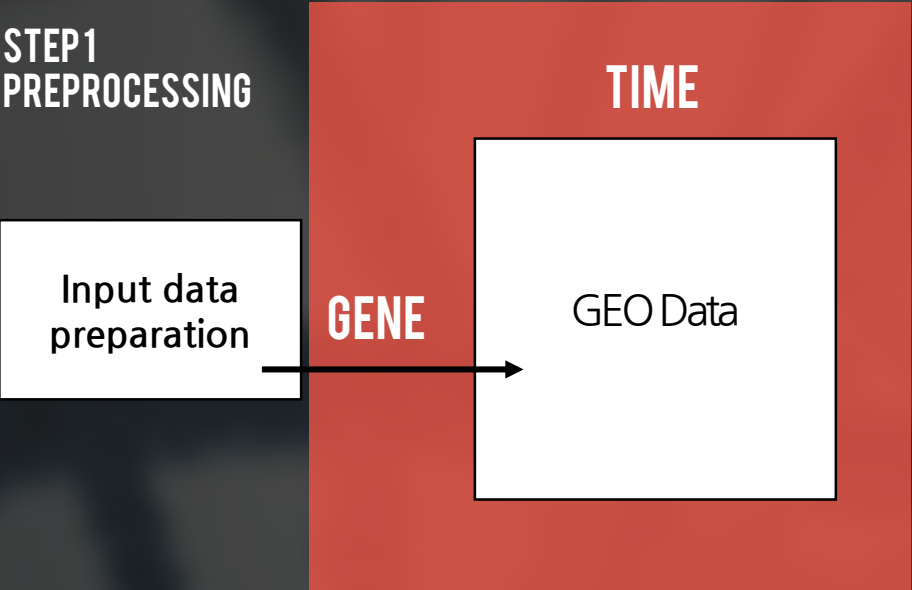
2.Method

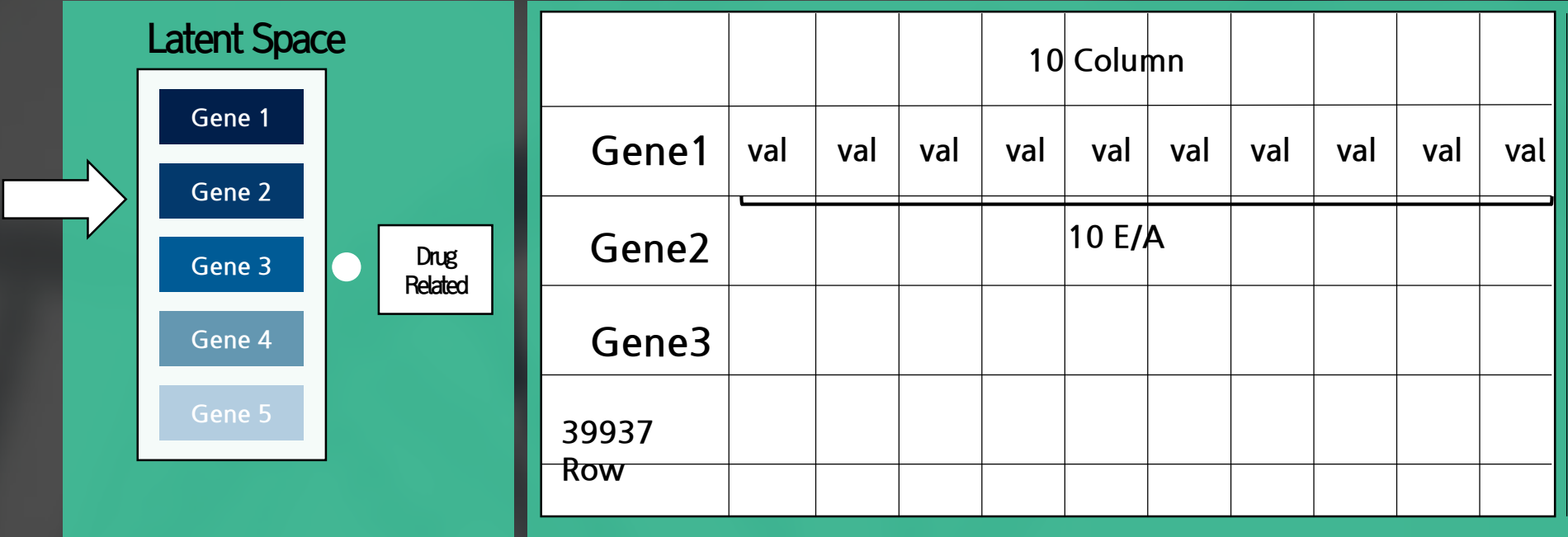## Step 1. Preprocessing

```
      Gene Id  TP      Value
1     16657436 1  -0.063517793
2     16657436 2  -0.072739783
3     16657436 4  -0.046710424
4     16657436 6  -0.017605944
5     16657436 24 -0.055617004
6     16657436 48 0.010705803
7     16657440 1  0.036559191
8     16657440 2  -0.073902344
9     16657440 4  0.018246203
10    16657440 6  0.010045348
11    16657440 24 -0.047593041
12    16657440 48 -0.040935001
```

STEP1
PREPROCESSING

Input data preparation → GENE → **TIME** GEO Data

	Time Point 0	Time Point 1	Time Point 2	
Gene1	Test - Control	Test - Control	Test - Control	
Gene2	Test - Control	Test - Control	Test - Control	●●●
Gene3	Test - Control	Test - Control	Test - Control	
		●		
		●		

2.Method

Step 2. MF Experiment

LATENT SPACE

Latent Space

Gene 1
Gene 2
Gene 3
Gene 4
Gene 5

Drug Related

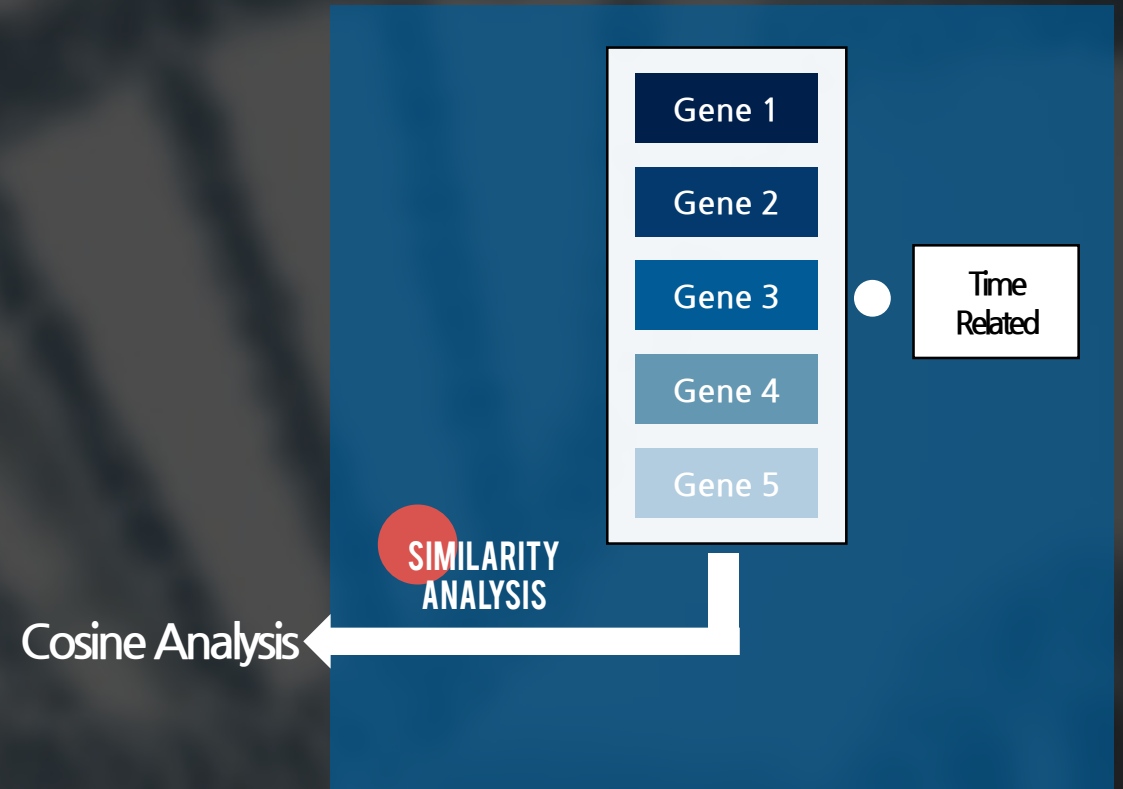|										
	10 Column									
Gene1	val	val	val	val	val	val	val	val	val	val
Gene2					10 E/A					
Gene3										
39937 Row										

```
 1  Dimension: 39937 x 10
    [0.002539091    0.007624203    0.0066787438    -0.006508924    0.008181381  0.008740933 -0.0123922   0.008969718 -0.013079663    -0.010385518]
    [0.00379i466    -3.3209336E-4    -0.026933813    -0.03870391 -0.008306302    0.025093056 -0.004232476    0.026975768 -0.01220879 -0.02693758]
    [-0.008702863    -0.004907232    -0.0044590733    -0.011056392    0.0013728082    -4.3674436E-4   -0.003486521    -0.013119135    -0.004690911
    -0.0065090163]
 5  [0.001436993    0.008147015 -0.004651517    -0.007582642    0.0071523045    -0.0018798077    0.005026365 0.01146251   -0.005152842    -0.012261407]
 6  [0.00349877777    0.00216807 -0.007908275    -0.0144901285    -0.009042019    0.016016616 0.0073771575    -0.0035307198    0.008654798 -9.2725165E-4]
 7  [-0.0012364885    -0.010056057    0.0066342247    0.0025381336    0.0065873987    -0.0072787483    0.004108262 -0.0034971633    -0.0054715485
    0.0022469114]
 8  [-0.008680478    -4.5615196E-4    -0.024883041    0.0026553508    0.004092916 -0.0011068434    -0.0010314971    0.00813664   -0.005329893    -0.0042766808]
 9  [-0.0043327087 0.0026543085    0.0015715753    -0.0018582086    0.0018010591    0.0025897103    -0.009036142    0.0030566095    0.0026155796
    3.34628E-4]
```
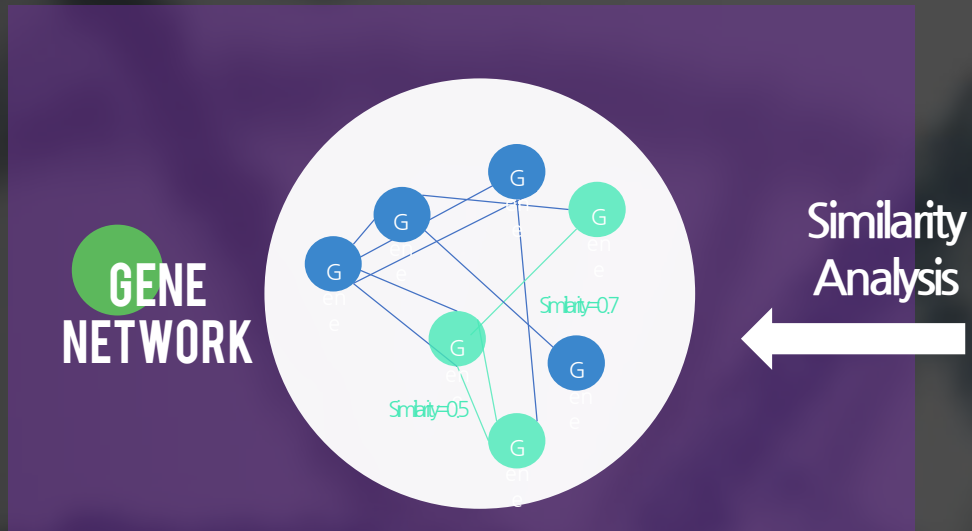
Gene 1
Gene 2
Gene 3

2.Method

Network will be made
with results of Similarity analysis test

| Gene 1 |
| Gene 2 |
| Gene 3 |
| Gene 4 |
| Gene 5 |

Time
Related

SIMILARITY
ANALYSIS

Cosine Analysis ◀

2.Method

Step 4. Network

GENE NETWORK

Similarity=0.7

Similarity=0.5

Similarity Analysis

If value x from Similarity Analysis is in this ranges,
We can infer that Genes have " ～ " relation in between.

Threshold

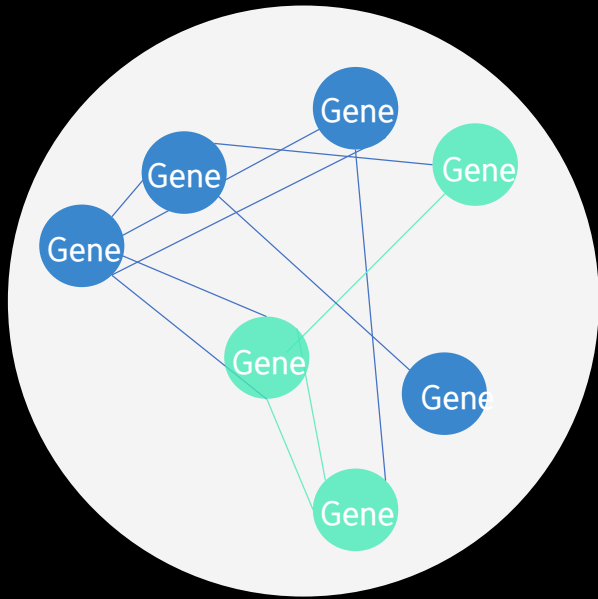$0.3 < x < 1.0$ 1(Positively Related)

$-0.3 <= x <= 0.3$ ➡ 0(x)

$-1.0 < x < -0.3$ -1(Negatively Related)

16657436 : (16657436, 1) (16657440, 1) (16657447, 1) (16657476, 1) (16657492, -1) (16657502, -1) (16657509, -1) (16657527, 1) (16657529, 1) (16657534,
1) (16657554, 1) (16657572, 1) (16657652, -1) (16657654, 1) (16657680, 1) (16657730, -1) (16657737, 1) (16657758, -1) (16657764, 1) (16657767, 1)
(16657793, 1) (16657917, 1) (16657923, 1) (16657948, -1) (16657993, 1) (16658005, 1) (16658023, -1) (16658064, -1) (16658119, -1) (16658122, 1)
(16658125, -1) (16658129, 1) (16658135, -1) (16658184, -1) (16658192, -1) (16658226, -1) (16658231, 1) (16658248, -1) (16658254, -1) (16658268, 1)
(16658278, -1) (16658289, 1) (16658300, -1) (16658428, 1) (16658449, -1) (16658457, 1) (16658471, -1) (16658479, -1) (16658523, 1) (16658528, 1)
(16658536, 1) (16658589, -1) (16658594, 1) (16658622, 1) (16658641, -1) (16658644, 1) (16658655, -1) (16658674, 1) (16658690, -1) (16658800, 1)
(16658864, -1) (16658904, 1) (16658909, -1) (16658926, 1) (16658947, 1) (16658950, -1) (16658987, -1) (16659036, -1) (16659133, 1) (16659138, 1)
(16659140, 1) (16659200, -1) (16659238, -1) (16659338, -1) (16659340, -1) (16659343, 1) (16659349, -1) (16659357, -1) (16659366, 1) (16659380, -1)
(16659395, -1) (16659407, 1) (16659413, -1) (16659434, -1) (16659440, -1) (16659443, -1) (16659455, -1) (16659475, -1) (16659545, 1) (16659637, -1)
(16659686, -1) (16659727, -1) (16659746, -1) (16659794, 1) (16659825, -1) (16659856, -1) (16659904, 1) (16659907, 1) (16659911, 1) (16659916, 1)

3.Validation

VALIDATION OF MODEL

Our model

Ex)

```
found value : (16896049, 16913761), num_line : 20896
찾은 값 : 1
found value : (16821365, 16896049), num_line : 14239
찾은 값 : 1
found value : (16821330, 16896049), num_line : 14238
찾은 값 : 1
```

```
found value : (16879408, 16948209), num_line : 19441
찾은 값 : -1
```

```
found value : (16833839, 16898788), num_line : 15339
찾은 값 : 1
```

```
found value : (16848807, 17046135), num_line : 16600
cannot found (0)

found value : (16833839, 16913806), num_line : 15339
cannot found (0)
```

3.Validation

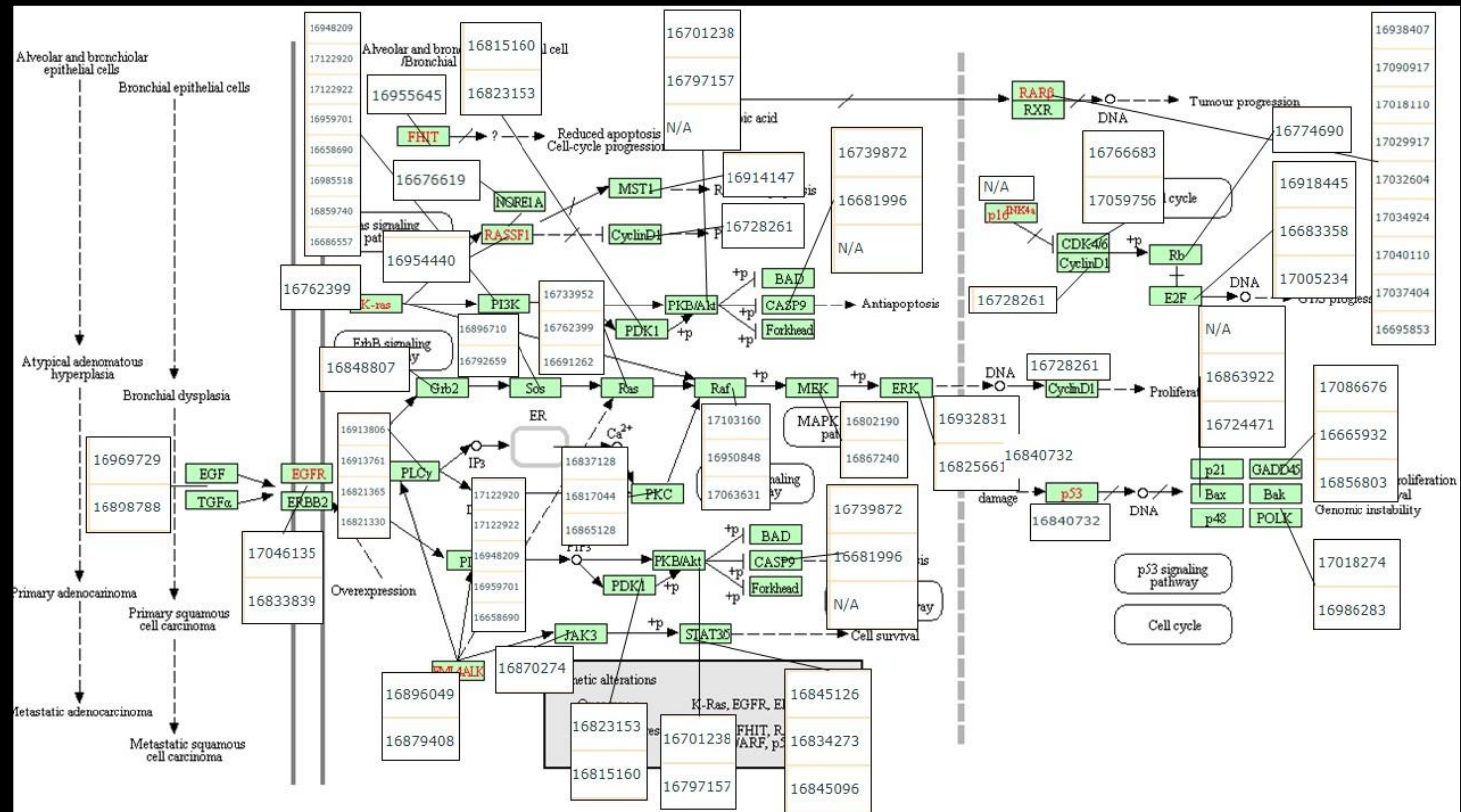## VALIDATION OF MODEL

Our model

Gene pathway


```
found_value : (16896049, 16913761), num_line : 20896
찾은 값 : 1
found_value : (16821365, 16896049), num_line : 14239
찾은 값 : 1
found_value : (16821330, 16896049), num_line : 14238
찾은 값 : 1

found_value : (16879408, 16948209), num_line : 19441
찾은 값 : -1

found_value : (16833839, 16898788), num_line : 15339
찾은 값 : 1

found_value : (16848807, 17046135), num_line : 16600
cannot found (0)

found_value : (16833839, 16913806), num_line : 15339
cannot found (0)
```

Compare
our model with
gene pathway

Validation ratio : 0.337423

4. VideoClip

Thank you