Curigin

BIOINFORMATICS

logFC 기반 pathway 분석 자동화 시스템

프로젝트 담당자: 염윤석

CONTENTS

1. Program introduction & purpose

- After DEG analysis – For Gene Set Enrichment Analysis
- Gene pathway analysis

2. Method

- DEG Raw data filtering(log FC) - **input**
- "ENRICHR" database pathway searching - **output**
- Intersection analysis of gene pathway result - **ouput**

3. Discussion

1. Program introduction & purpose

Cell line : 253J, FaDu, C4-2B

253J : Human Bladder Carcinoma cell line

FaDu : Hypopharyngeal squamous cell carcinoma cell line

C4-2B : Human prostate cancer cell line

[Target Gene]

mTOR & STAT3

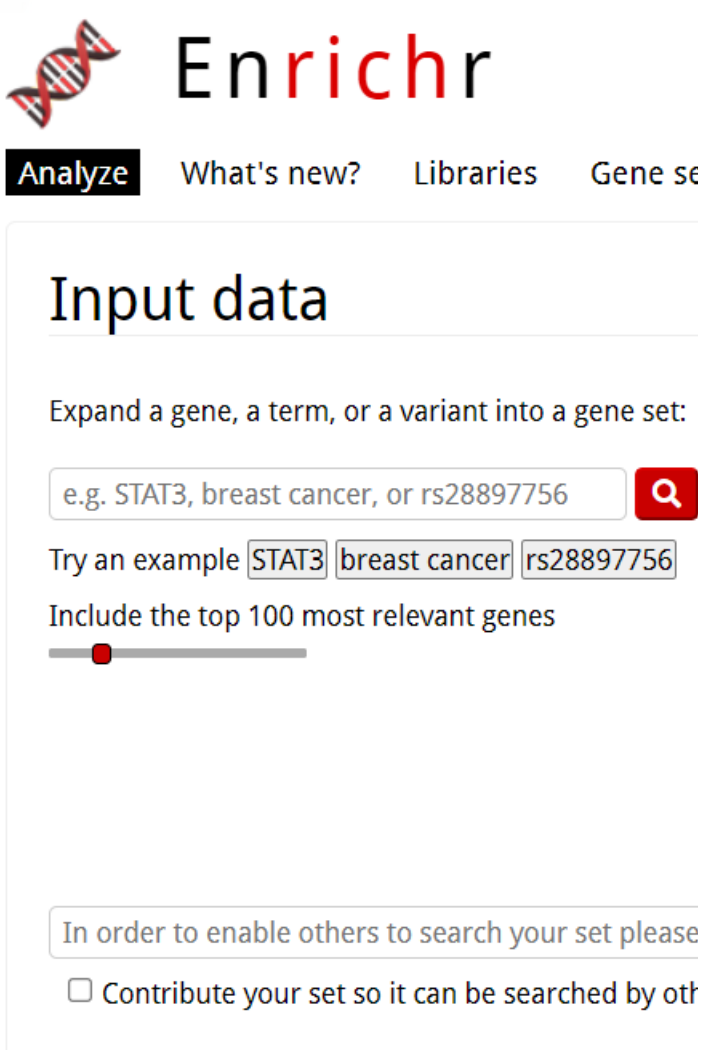
LogFC value

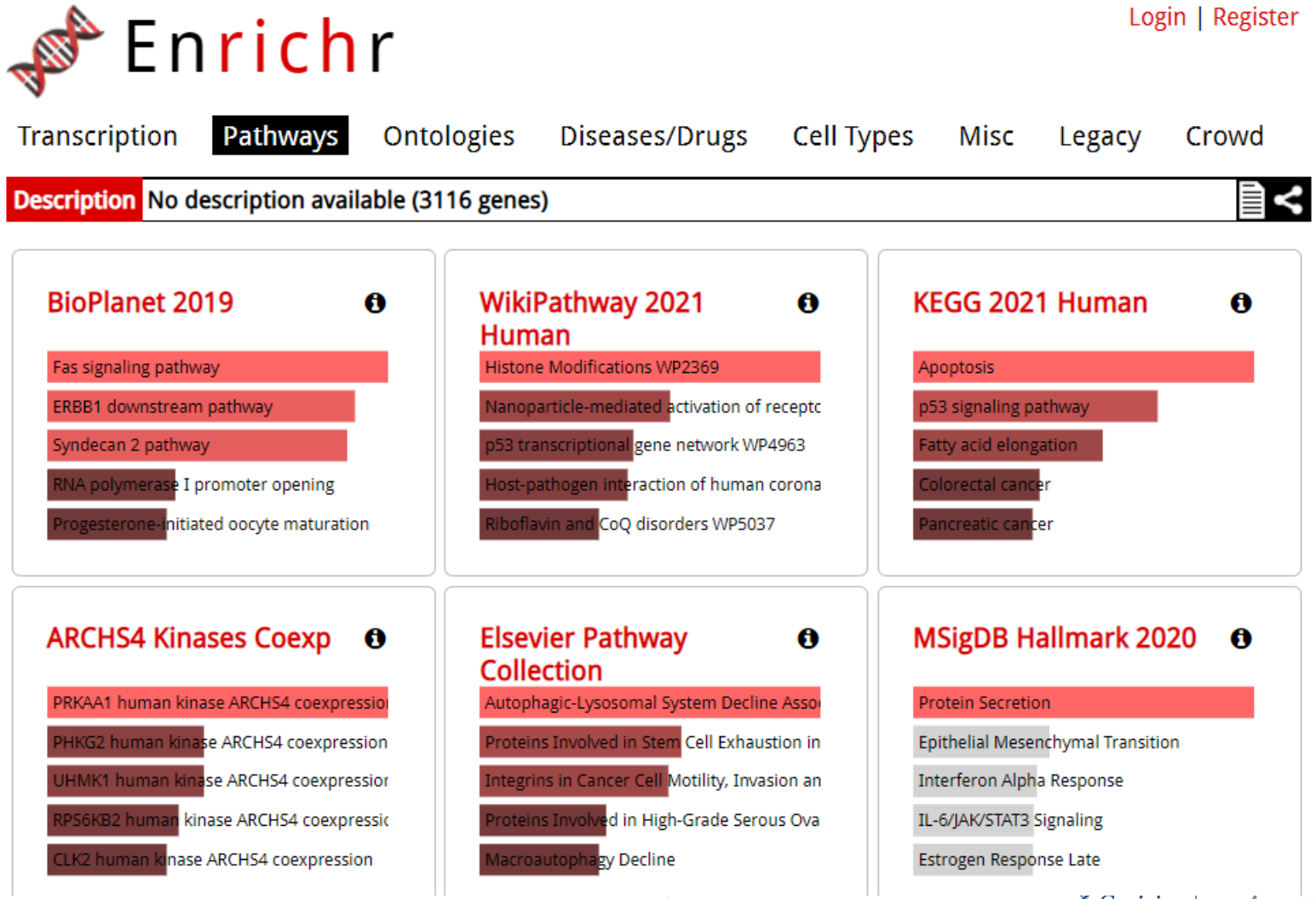
[DEG Analysis]

Humanizing Genomics macrogen

Drug Sample group	Test vs Control groups
253J-BV-102	253J-BV-102 vs. 253J-BV-control
253J-BV-10G	253J-BV-10G vs. 253J-BV-control
253J-BV-control	253J-BV-10G vs. 253J-BV-102

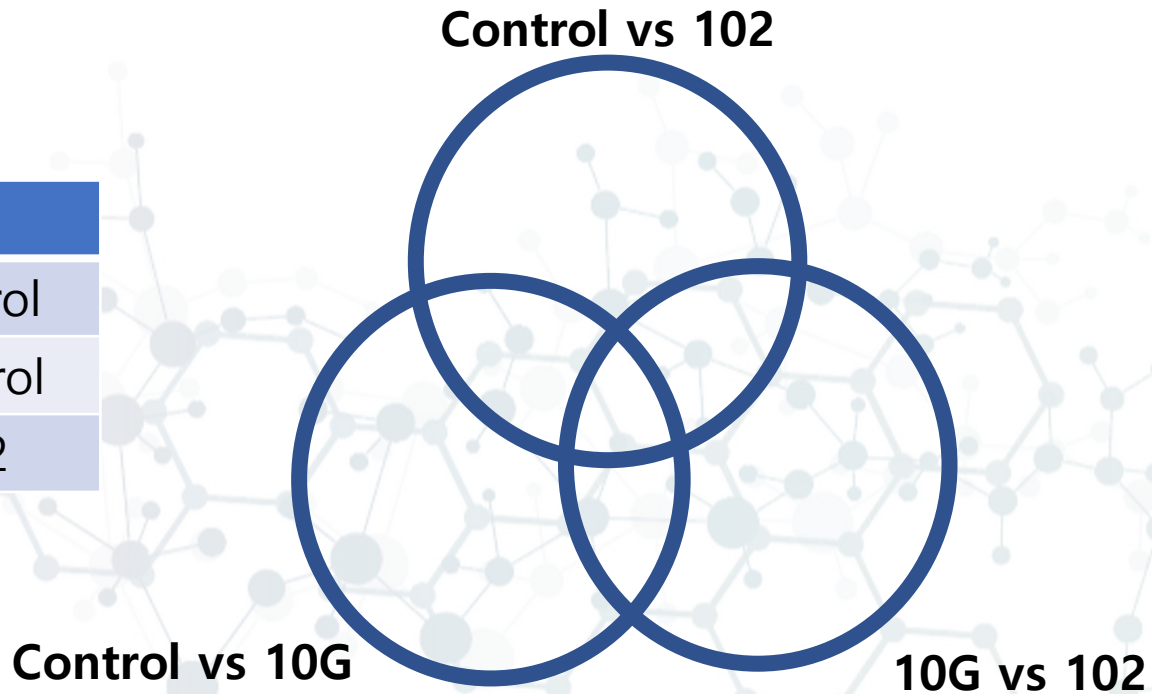
1. Program introduction & purpose





1. Program introduction & purpose

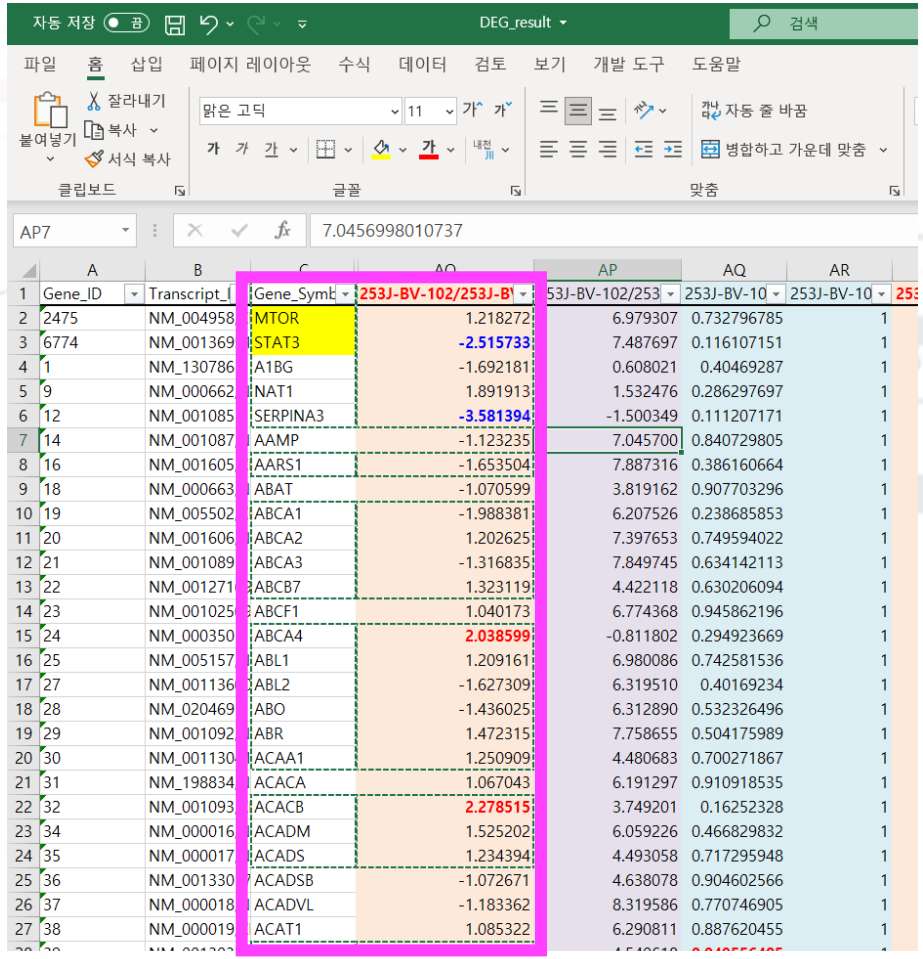
Test vs Control groups
253J-BV-102 vs. 253J-BV-control
253J-BV-10G vs. 253J-BV-control
253J-BV-10G vs. 253J-BV-102



Analyzing Significant target gene pathway
Down or up - regulated by drug samples!

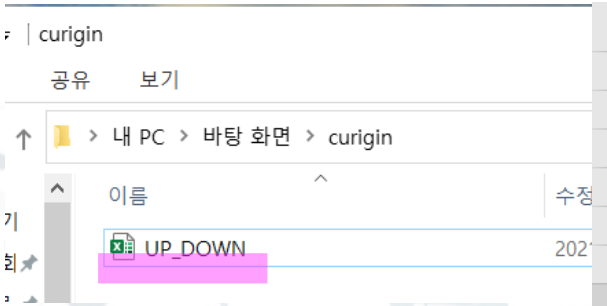
2. Method – Input file [UP_DONW]

Raw Data : DEG analysis result



	A	B	C	AO	AP	AQ	AR
1	Gene_ID	Transcript_I	Gene_Symb	253J-BV-102/253J-B	53J-BV-102/253	253J-BV-10	253J-BV-10
2	2475	NM_004958	MTOR	1.218272	6.979307	0.732796785	1
3	6774	NM_001369	STAT3	-2.515733	7.487697	0.116107151	1
4	1	NM_130786	A1BG	-1.692181	0.608021	0.40469287	1
5	9	NM_000662	NAT1	1.891913	1.532476	0.286297697	1
6	12	NM_001085	SERPINA3	-3.581394	-1.500349	0.111207171	1
7	14	NM_001087	AAMP	-1.123235	7.045700	0.840729805	1
8	16	NM_001605	AARS1	-1.653504	7.887316	0.386160664	1
9	18	NM_000663	ABAT	-1.070599	3.819162	0.907703296	1
10	19	NM_005502	ABCA1	-1.988381	6.207526	0.238685853	1
11	20	NM_001606	ABCA2	1.202625	7.397653	0.749594022	1
12	21	NM_001089	ABCA3	-1.316835	7.849745	0.634142113	1
13	22	NM_001271	ABCB7	1.323119	4.422118	0.630206094	1
14	23	NM_001025	ABCF1	1.040173	6.774368	0.945862196	1
15	24	NM_000350	ABCA4	2.038599	-0.811802	0.294923669	1
16	25	NM_005157	ABL1	1.209161	6.980086	0.742581536	1
17	27	NM_001136	ABL2	-1.627309	6.319510	0.40169234	1
18	28	NM_020469	ABO	-1.436025	6.312890	0.532326496	1
19	29	NM_001092	ABR	1.472315	7.758655	0.504175989	1
20	30	NM_001130	ACAA1	1.250909	4.480683	0.700271867	1
21	31	NM_198834	ACACA	1.067043	6.191297	0.910918535	1
22	32	NM_001093	ACACB	2.278515	3.749201	0.16252328	1
23	34	NM_000016	ACADM	1.525202	6.059226	0.466829832	1
24	35	NM_000017	ACADS	1.234394	4.493058	0.717295948	1
25	36	NM_001330	ACADSB	-1.072671	4.638078	0.904602566	1
26	37	NM_000018	ACADVL	-1.183362	8.319586	0.770746905	1
27	38	NM_000019	ACAT1	1.085322	6.290811	0.887620455	1

1. Input data : "UP_DONW.xlsx"



	A	B	C	D	E
1	Gene	LogFC			
2	MTOR	0.912989			
3	STAT3	0.370732			
4	SNORD35	1907.727			
5	SNORA44	24.51009			
6	SNORD97	16.16849			
7	SNORD15	12.8939			
8	SNORA6	12.64529			
9	RNVU1-20	8.408803			
10	BCYRN1	5.977254			
11	CKB	5.709024			
12	SNORD10	5.513206			
13	NUDT4P1	5.161954			
14	DEFB109P	4.850933			
15	BMP8B	4.833422			
16	ACTA2	4.764732			
17	TNFRSF10	4.666227			
18	FDXR	4.54199			
19	GP1BB	4.419471			
20	HES4	4.278054			
21	DPYSL4	4.272934			
22	CKLF-CMT	4.195083			
23	CRABP2	4.189488			
24	CELSR3-AS	4.073673			

10G_102 CON_10G CON_102

준비

2. Method – Input file [Chrome webdriver]

"WebCrolling module" – Chrome Webdriver

Current Releases

- If you are using Chrome version 95, please download ChromeDriver 95.0.4638.17
- If you are using Chrome version 94, please download ChromeDriver 94.0.4606.61
- If you are using Chrome version 93, please download ChromeDriver 93.0.4577.63
- For older version of Chrome, please see below for the version of ChromeDriver that supports it.

If you are using Chrome from Dev or Canary channel, please following inst

For more information on selecting the right version of ChromeDriver, plea

ChromeDriver 95.0.4638.17

Supports Chrome version 95

- Resolved issue 3857: Host header or origin header is specified and is
- Added `--allowed-origins` argument

For more details, please see the release notes.

ChromeDriver 95.0.4638.10

Supports Chrome version 95

- Resolved issue 3857: Host header or origin header is specified and is not whitelisted or localhost [Pri-1]
- Added `--allowed-origins` argument

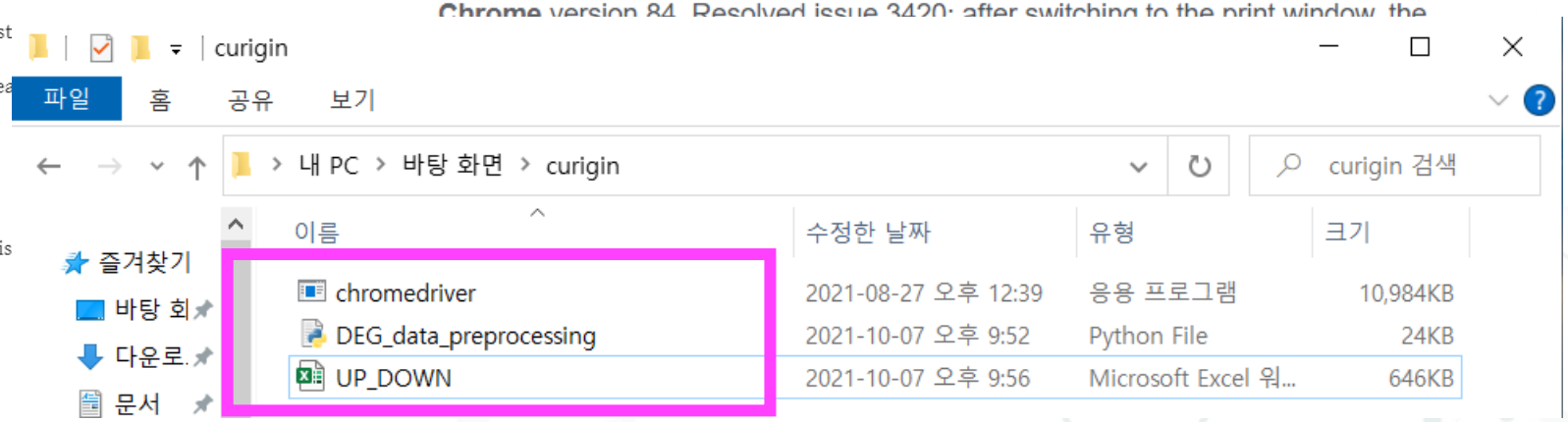
For more details, please see the release notes.

ChromeDriver 94.0.4606.61

romedriver.chromium.org › downloads

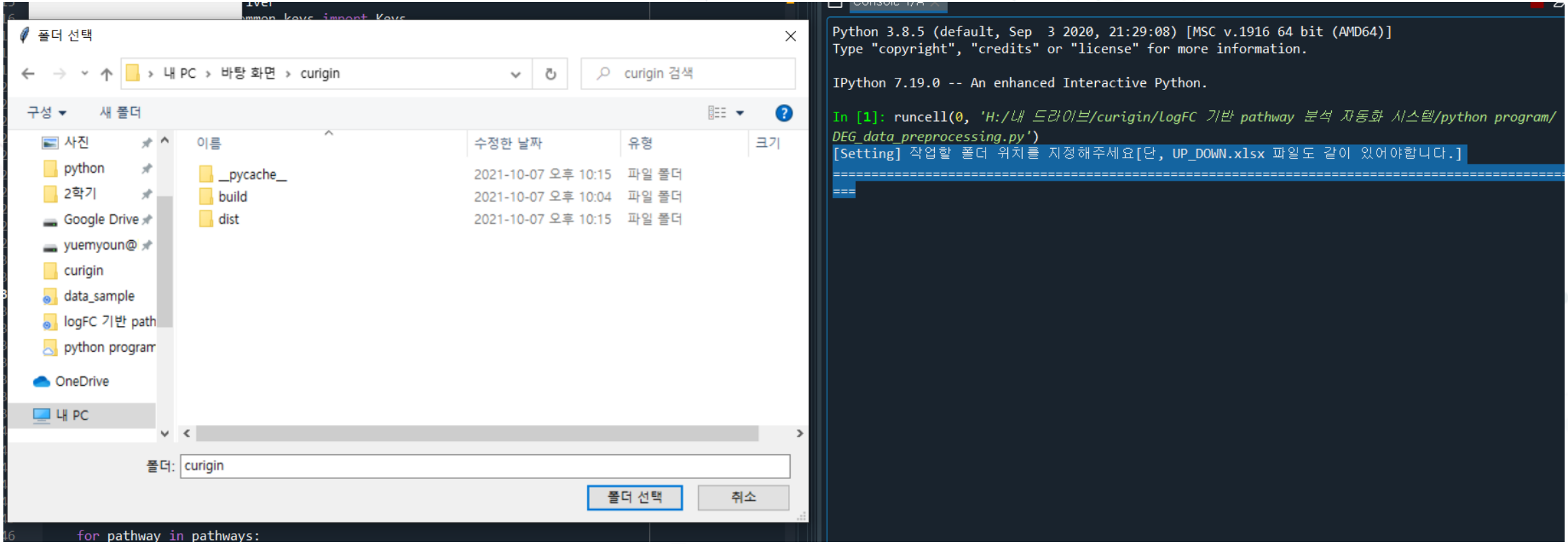
eDriver - WebDriver for Chrome - Downloads

Chrome version 84. Resolved issue 3420: after switching to the print window, the



2. Method – processing [Setting]

[Setting] 작업할 폴더 위치를 지정해주세요 [단, UP_DONW.xlsx 파일도 같이 있어야합니다.]



2. Method – processing [Setting]

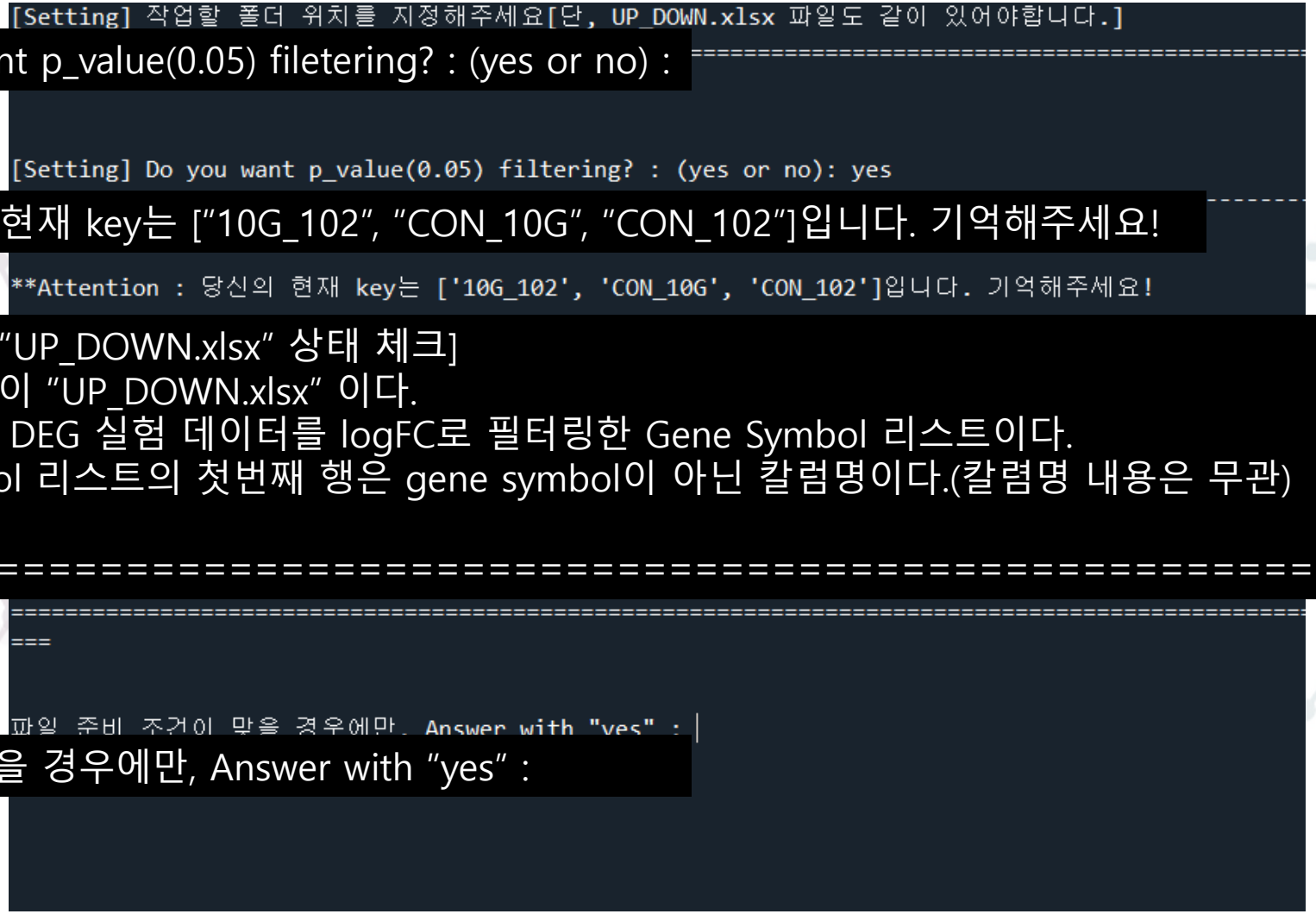
[Setting] Do you want p_value(0.05) filetering? : (yes or no) :

**Attention : 당신의 현재 key는 [“10G_102”, “CON_10G”, “CON_102”]입니다. 기억해주세요!

[현재 준비된 파일 – “UP_DOWN.xlsx” 상태 체크]
조건 1 : 파일의 이름이 “UP_DOWN.xlsx” 이다.
조건 2 : 파일 내용은 DEG 실험 데이터를 logFC로 필터링한 Gene Symbol 리스트이다.
조건 3: Gene Symbol 리스트의 첫번째 행은 gene symbol이 아닌 칼럼명이다.(칼렴명 내용은 무관)

==========

파일 준비 조건이 맞을 경우에만, Answer with “yes” :



2. Method - processing

```
[What do you want me to do?]
-------------------------------------------------------------

Mode : 1. ENRICHR(gene 리스트로 pathway raw 데이터 만들기)
Mode : 2. COMB_COUNT(pathway raw 데이터로 double, triple... updown counting 하기
```

```
-------------------------------------------------------------------------------

Case 1 : Just do the 1. ENRICHR
Case 2: I did 1. ENRICHR I wan you to do 2. COMB_COUNT
Case 3: Do 1. ENRICHR & 2. COMB_COUNT at ONCE!
Case 4: Let me think again, END this program

[YOUR CHOICE?] (1 : Case 1 | 2 : Case 2 | 3 : Case3 | 4 : Case 4)  =
```

2. Method - processing

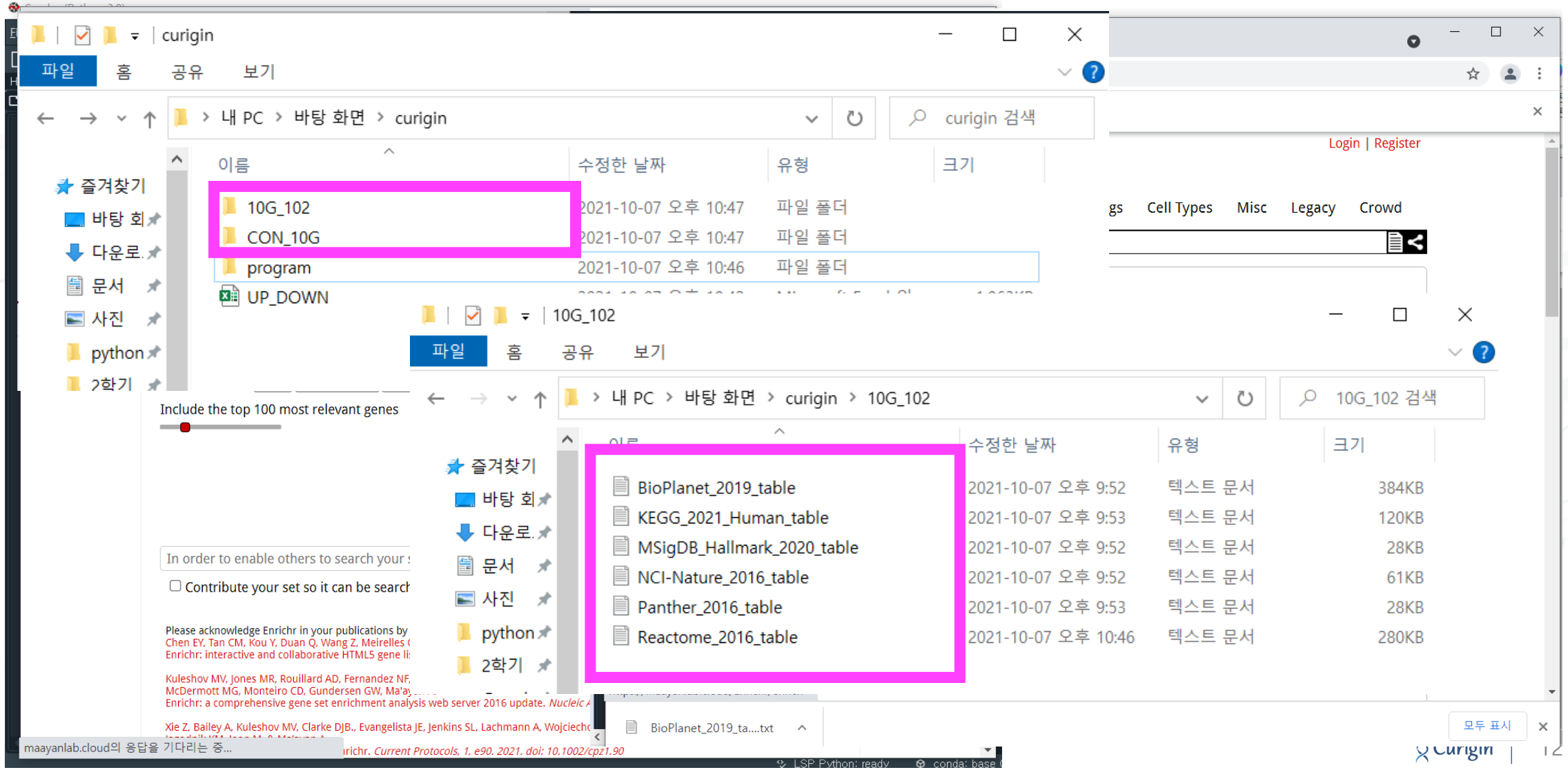
```
[PATHWAY EXAMPLES]
BioPlanet_2019
MSigDB_Hallmark_2020
NCI-Nature_2016
Panther_2016
KEGG_2021_Human
Reactome_2016


모든 pathway에 대해 작업을 원하시면 "all"을, 원하는 pathway를 정하고 싶다면 "no"를 입력하세요 :
```

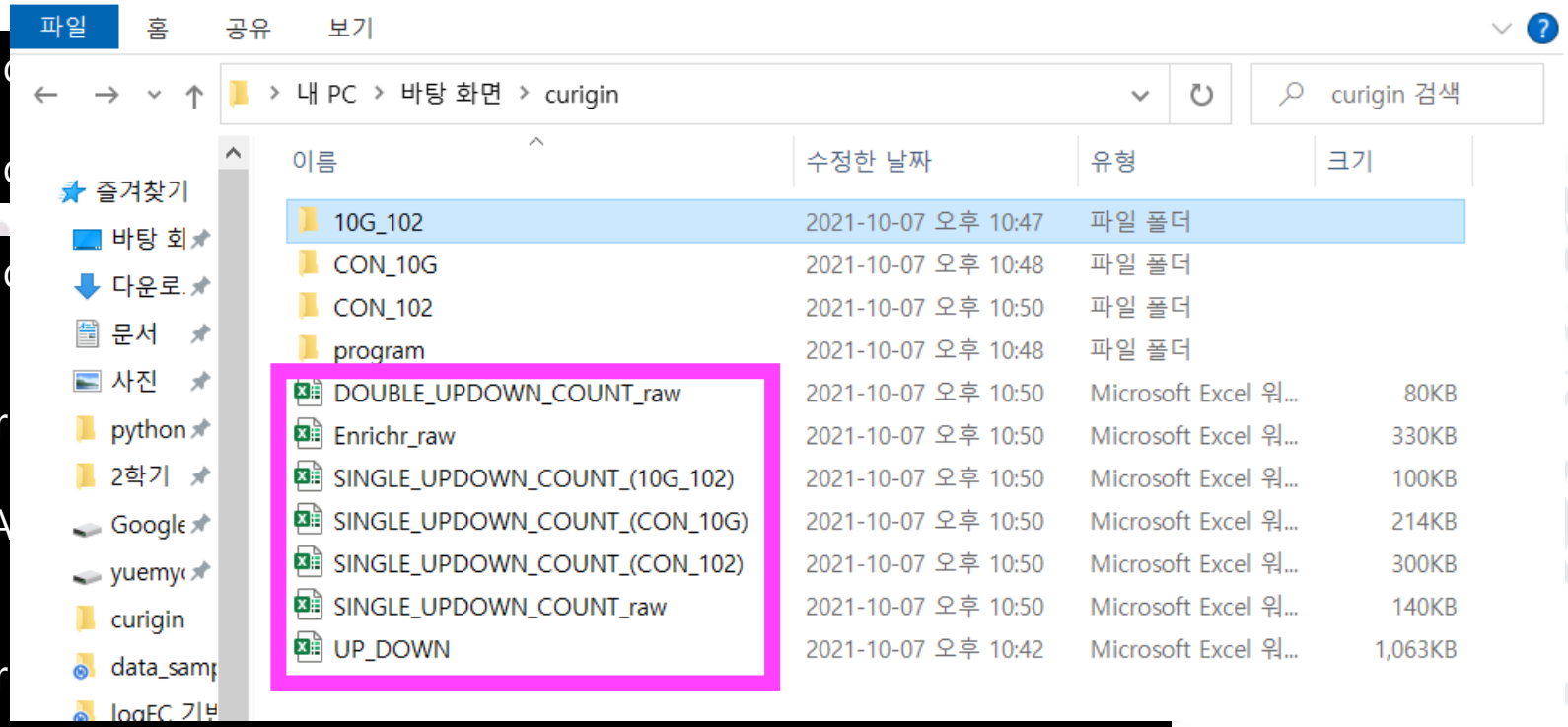
```
[Setting] Chrome webdriver의 위치를 알려주세요!

[Setting] 컴퓨터의 "다운로드" 위치를 알려주세요!
```

2. Method - processing



2. Method - processing

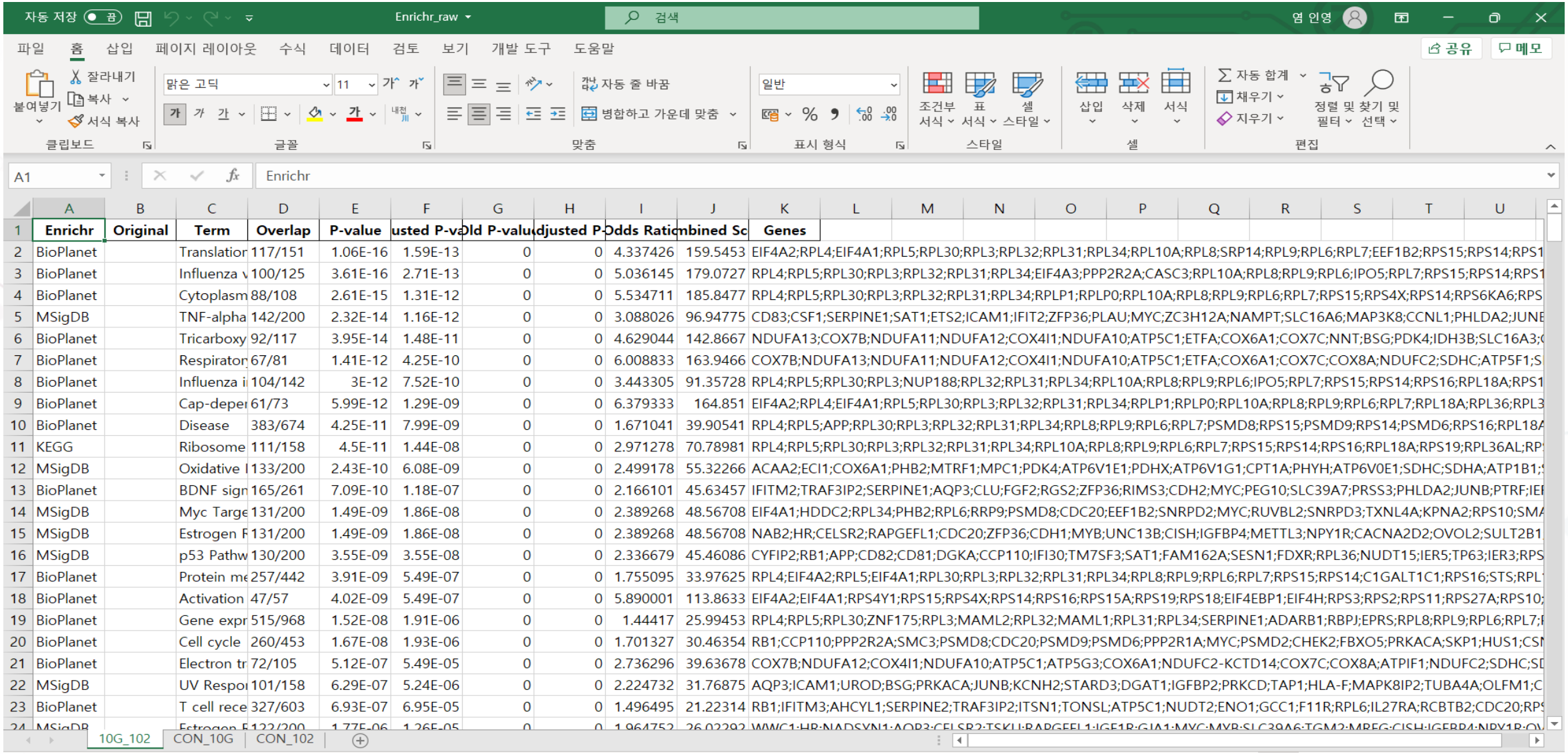



```
[LOADING] I am making DOUBLE combinations of keys...

your current keys : ['10G_102', 'CON_10G', 'CON_102']

[LOADING] I am making TRIPLE combinations of keys...
```

2. Method – Output file [Enrichr]



Enrichr_raw

	A	B	C	D	E	F	G	H	I	J	K
1	Enrichr	Original	Term	Overlap	P-value	usted P-va	ld P-valu	djusted P-	Odds Ratio	mbined Sc	Genes
2	BioPlanet		Translation	117/151	1.06E-16	1.59E-13	0	0	4.337426	159.5453	EIF4A2;RPL4;EIF4A1;RPL5;RPL30;RPL3;RPL32;RPL31;RPL34;RPL10A;RPL8;SRP14;RPL9;RPL6;RPL7;EEF1B2;RPS15;RPS14;RPS1
3	BioPlanet		Influenza v	100/125	3.61E-16	2.71E-13	0	0	5.036145	179.0727	RPL4;RPL5;RPL30;RPL3;RPL32;RPL31;RPL34;EIF4A3;PPP2R2A;CASC3;RPL10A;RPL8;RPL9;RPL6;IPO5;RPL7;RPS15;RPS14;RPS1
4	BioPlanet		Cytoplasm	88/108	2.61E-15	1.31E-12	0	0	5.534711	185.8477	RPL4;RPL5;RPL30;RPL3;RPL32;RPL31;RPL34;RPLP1;RPLP0;RPL10A;RPL8;RPL9;RPL6;RPL7;RPS15;RPS4X;RPS14;RPS6KA6;RPS
5	MSigDB		TNF-alpha	142/200	2.32E-14	1.16E-12	0	0	3.088026	96.94775	CD83;CSF1;SERPINE1;SAT1;ETS2;ICAM1;IFIT2;ZFP36;PLAU;MYC;ZC3H12A;NAMPT;SLC16A6;MAP3K8;CCNL1;PHLDA2;JUNE
6	BioPlanet		Tricarboxy	92/117	3.95E-14	1.48E-11	0	0	4.629044	142.8667	NDUFA13;COX7B;NDUFA11;NDUFA12;COX4I1;NDUFA10;ATP5C1;ETFA;COX6A1;COX7C;NNT;BSG;PDK4;IDH3B;SLC16A3;(
7	BioPlanet		Respirator	67/81	1.41E-12	4.25E-10	0	0	6.008833	163.9466	COX7B;NDUFA13;NDUFA11;NDUFA12;COX4I1;NDUFA10;ATP5C1;ETFA;COX6A1;COX7C;COX8A;NDUFC2;SDHC;ATP5F1;S
8	BioPlanet		Influenza i	104/142	3E-12	7.52E-10	0	0	3.443305	91.35728	RPL4;RPL5;RPL30;RPL3;NUP188;RPL32;RPL31;RPL34;RPL10A;RPL8;RPL9;RPL6;IPO5;RPL7;RPS15;RPS14;RPS16;RPL18A;RPS1
9	BioPlanet		Cap-depe	61/73	5.99E-12	1.29E-09	0	0	6.379333	164.851	EIF4A2;RPL4;EIF4A1;RPL5;RPL30;RPL3;RPL32;RPL31;RPL34;RPLP1;RPLP0;RPL10A;RPL8;RPL9;RPL6;RPL7;RPL18A;RPL36;RPL3
10	BioPlanet		Disease	383/674	4.25E-11	7.99E-09	0	0	1.671041	39.90541	RPL4;RPL5;APP;RPL30;RPL3;RPL32;RPL31;RPL34;RPL8;RPL9;RPL6;RPL7;PSMD8;RPS15;PSMD9;RPS14;PSMD6;RPS16;RPL18A
11	KEGG		Ribosome	111/158	4.5E-11	1.44E-08	0	0	2.971278	70.78981	RPL4;RPL5;RPL30;RPL3;RPL32;RPL31;RPL34;RPL10A;RPL8;RPL9;RPL6;RPL7;RPS15;RPS14;RPS16;RPL18A;RPS19;RPL36AL;RP
12	MSigDB		Oxidative I	133/200	2.43E-10	6.08E-09	0	0	2.499178	55.32266	ACAA2;ECI1;COX6A1;PHB2;MTRF1;MPC1;PDK4;ATP6V1E1;PDHX;ATP6V1G1;CPT1A;PHYH;ATP6V0E1;SDHC;SDHA;ATP1B1;S
13	BioPlanet		BDNF sign	165/261	7.09E-10	1.18E-07	0	0	2.166101	45.63457	IFITM2;TRAF3IP2;SERPINE1;AQP3;CLU;FGF2;RGS2;ZFP36;RIMS3;CDH2;MYC;PEG10;SLC39A7;PRSS3;PHLDA2;JUNB;PTRF;IEF
14	MSigDB		Myc Targe	131/200	1.49E-09	1.86E-08	0	0	2.389268	48.56708	EIF4A1;HDDC2;RPL34;PHB2;RPL6;RRP9;PSMD8;CDC20;EEF1B2;SNRPD2;MYC;RUVBL2;SNRPD3;TXNL4A;KPNA2;RPS10;SMA
15	MSigDB		Estrogen F	131/200	1.49E-09	1.86E-08	0	0	2.389268	48.56708	NAB2;HR;CELSR2;RAPGEFL1;CDC20;ZFP36;CDH1;MYB;UNC13B;CISH;IGFBP4;METTL3;NPY1R;CACNA2D2;OVOL2;SULT2B1
16	MSigDB		p53 Pathw	130/200	3.55E-09	3.55E-08	0	0	2.336679	45.46086	CYFIP2;RB1;APP;CD82;CD81;DGKA;CCP110;IFI30;TM7SF3;SAT1;FAM162A;SESN1;FDXR;RPL36;NUDT15;IER5;TP63;IER3;RPS
17	BioPlanet		Protein me	257/442	3.91E-09	5.49E-07	0	0	1.755095	33.97625	RPL4;EIF4A2;RPL5;EIF4A1;RPL30;RPL3;RPL32;RPL31;RPL34;RPL8;RPL9;RPL6;RPL7;RPS15;RPS14;C1GALT1C1;RPS16;STS;RPL
18	BioPlanet		Activation	47/57	4.02E-09	5.49E-07	0	0	5.890001	113.8633	EIF4A2;EIF4A1;RPS4Y1;RPS15;RPS4X;RPS14;RPS16;RPS15A;RPS19;RPS18;EIF4EBP1;EIF4H;RPS3;RPS2;RPS11;RPS27A;RPS10;
19	BioPlanet		Gene expr	515/968	1.52E-08	1.91E-06	0	0	1.44417	25.99453	RPL4;RPL5;RPL30;ZNF175;RPL3;MAML2;RPL32;MAML1;RPL31;RPL34;SERPINE1;ADARB1;RBPJ;EPRS;RPL8;RPL9;RPL6;RPL7;F
20	BioPlanet		Cell cycle	260/453	1.67E-08	1.93E-06	0	0	1.701327	30.46354	RB1;CCP110;PPP2R2A;SMC3;PSMD8;CDC20;PSMD9;PSMD6;PPP2R1A;MYC;PSMD2;CHEK2;FBXO5;PRKACA;SKP1;HUS1;CSN
21	BioPlanet		Electron tr	72/105	5.12E-07	5.49E-05	0	0	2.736296	39.63678	COX7B;NDUFA12;COX4I1;NDUFA10;ATP5C1;ATP5G3;COX6A1;NDUFC2-KCTD14;COX7C;COX8A;ATPIF1;NDUFC2;SDHC;SD
22	MSigDB		UV Respor	101/158	6.29E-07	5.24E-06	0	0	2.224732	31.76875	AQP3;ICAM1;UROD;BSG;PRKACA;JUNB;KCNH2;STARD3;DGAT1;IGFBP2;PRKCD;TAP1;HLA-F;MAPK8IP2;TUBA4A;OLFM1;C
23	BioPlanet		T cell rece	327/603	6.93E-07	6.95E-05	0	0	1.496495	21.22314	RB1;IFITM3;AHCYL1;SERPINE2;TRAF3IP2;ITSN1;TONSL;ATP5C1;NUDT2;ENO1;GCC1;F11R;RPL6;IL27RA;RCBTB2;CDC20;RPS
24	MSigDB		Estrogen F	122/200	1.77E-06	1.26E-05	0	0	1.964752	26.02292	WWC1;HR;NADSYN1;AQP3;CELSR2;TSKU;RAPGEFL1;IGF1R;GJA1;MYC;MYB;SLC39A6;TGM2;MREG;CISH;IGFBP4;NPY1R;OV

10G_102 | CON_10G | CON_102

2. Method – Output file [UP_DOWN]

3. Discussion

- Webcrolling – webdriver version update
- Visualization
- Application Development – UX_UI

Curigin

THANK YOU

"Speedy Development of Safe Gene Therapy based on Smart Platform"