

User Manual

ExPADA v.1.0.3



< 목 차 >

1.	Peak Annotation and Differential enrichment Analysis (ExPADA)-----	3
2.	Functional Annotation Analysis (DAVID) -----	16
3.	Clustering Heatmap Analysis (Clustering) -----	25
4.	Principal Component Analysis (PCA) -----	27
5.	Pathway Analysis (KEGG mapper) -----	34
6.	Gene Set Enrichment Analysis (GSEA) -----	37
7.	Protein-Protein Network Analysis (Cytoscape STRING) -----	42

1. Peak Annotation and Differential enrichment Analysis (ExPADA)

(주)이바이오젠은 DNA-Seq (ChIP-Seq, MBD-Seq)을 엑셀 기반에서 쉽게 분석할 수 있도록 분석 결과 보고 시 ExPADA (Excel based Peak Annotation and Differential enrichment Analysis) tool 을 함께 제공한다. ExPADA 분석 툴은 (주)이바이오젠이 연구자들이 Microarray 및 DNA-Seq 데이터를 보다 쉽게 다루고 원하는 데이터를 쉽게 얻을 수 있도록 사용자 편의를 최대한 반영한 분석 툴이고 엑셀 프로그램 안에서 다양한 분석을 직관적으로 수행할 수 있도록 개발되었다. ExPADA 분석 툴은 사용자들의 요구사항을 지속적으로 반영하여 데이터 분석과 엑셀 사용에 익숙하지 못한 연구자들도 쉽게 사용이 가능하도록 계속 업데이트 될 예정이다.

이바이오젠에서 제공하는 DNA-Seq data (엑셀 데이터)를 열기 전에 함께 제공한 ExPADA(버전).zip 파일의 압축을 풀고 setup 을 실행하면 분석 툴이 설치된다(그림 1-1).

설치가 완료되고 ExPADA format 의 엑셀 데이터를 열면 자동으로 ExPADA 분석 툴이 구동된다. 참고로 ExPADA 설치 전에 실행 중인 엑셀 파일이 있으면 종료시킨 후 다시 실행해야 ExPADA 를 사용할 수 있다.

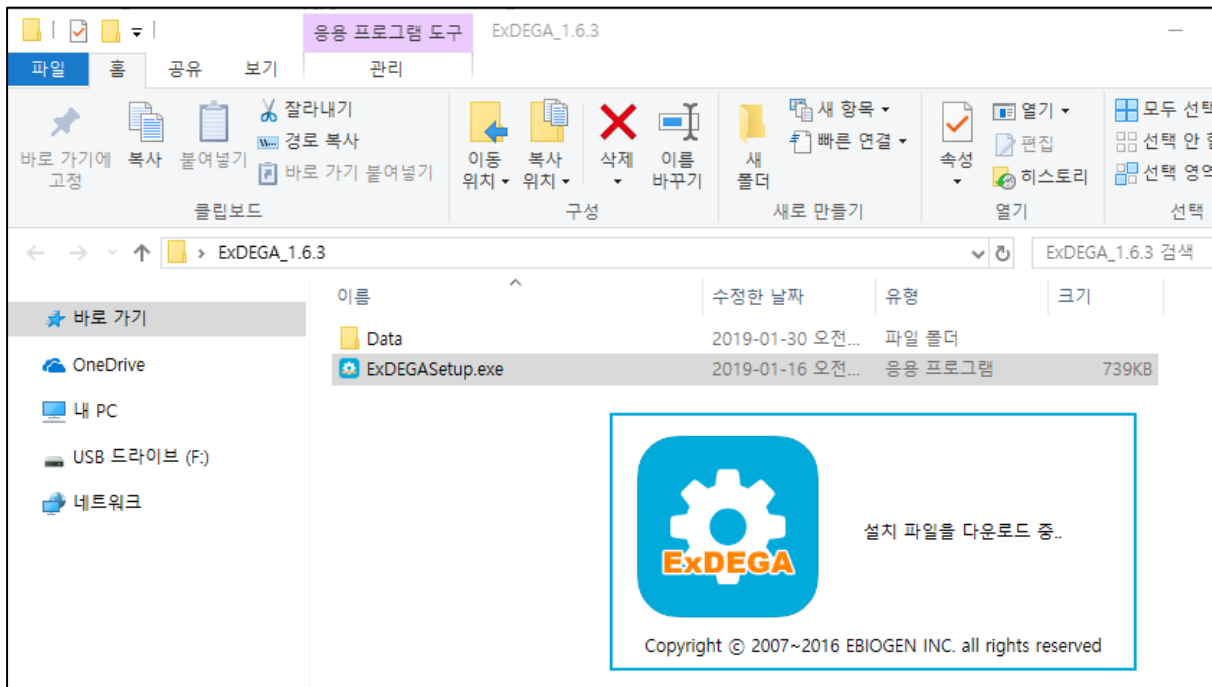


그림 1-1. ExPADA set up

ExpADA format의 엑셀 파일을 열면, 왼쪽에 Gene Category 창과 가운데에 gene expression data, 오른쪽에 DEG Analysis 창이 실행된다(그림 1-2). Gene Category 분석 창에서는 기본 설정된 Gene ontology (GO)가 있고 사용자가 원하는 대로 gene category를 구성하여 분석할 수 있다. Gene category 창과 DEG Analysis 창은 함께 연동하여 데이터를 쉽게 얻을 수 있다. DEG Analysis 창에서는 Fold change, Normalized Data (log2), p-value 등을 선택하여 DEG 선별을 쉽게 할 수 있고 DEGs로 gene category 별로 그래프를 작성할 수 있다. 뿐만 아니라, DEG 분석 창에서 Scatter Plot, Volcano Plot, Venn Diagram을 직접 그릴 수 있고 선별된 유전자들을 대상으로 Clustering heatmap, KEGG 분석, DAVID 분석을 수행하기 위한 input file을 자동으로 만들 수 있다. Gene expression graph, Gene search 기능도 이용할 수 있어 연구자가 DNA-Seq 결과를 쉽게 활용할 수 있다.

ID	Peak Area	Nearest Gene ID	Fold change	Normalized data (log2)	Raw data	Nearest Gene
1	chr1:564453-564853	ORAF29	4.501	0.000	109	0 ORAF29 729759 chr1 564453 564853
2	chr1:567886-568286	ORAF29	0.716	2.811	0.000	47 0 ORAF29 729759 chr1 567886 568286
3	chr1:723982-724382	LOC100288069	0.406	2.396	0.194	33 1 LOC100288069 100288069 chr1 723982 724382
4	chr1:725115-725515	LOC100288069	0.536	1.832	0.899	20 6 LOC100288069 100288069 chr1 725115 725515
5	chr1:726142-726542	LOC100288069	0.808	1.414	1.106	13 8 LOC100288069 100288069 chr1 726142 726542
6	chr1:726770-727170	LOC100288069	0.908	1.932	0.194	22 1 LOC100288069 100288069 chr1 726770 727170
7	chr1:824879-825279	FAM64C	0.456	1.628	0.000	11 0 FAM64C 264559 chr1 824879 825279
8	chr1:951843-951843	HSGL5	0.616	1.482	0.783	14 5 HSGL5 951843 951843
9	chr1:1530486-1530886	FNDC10	0.900	1.932	0.194	22 1 FNDC10 643988 chr1 1530486 1530886
10	chr1:4363056-4363456	LINC01777	0.436	1.980	0.783	23 5 LINC01777 284661 chr1 4363056 4363456
11	chr1:5056170-5056570	AJAP1	1.403	1.414	1.903	13 19 AJAP1 559666 chr1 5056170 5056570
12	chr1:12894311-12894911	A0ACAC4	1.835	1.017	1.729	8 16 A0ACAC4 843066 chr1 12894311 12894911
13	chr1:22699252-22699652	ZBT840	0.342	1.546	0.000	15 0 ZBT840 9523 chr1 22699252 22699652
14	chr1:31760619-31761019	SNRNP40	1.626	1.546	2.247	15 26 SNRNP40 31760619 31761019
15	chr1:35240998-35241398	GJB3	2.895	0.714	2.247	5 26 GJB3 2707 chr1 35240998 35241398
16	chr1:43945341-43945941	SZT2	0.862	2.275	2.060	30 22 SZT2 23334 chr1 43945341 43945941
17	chr1:44789998-44790398	ERL3	0.916	1.189	0.899	10 6 ERL3 79033 chr1 44789998 44790398
18	chr1:87051193-87051593	CLCA4	0.476	1.608	0.518	16 3 CLCA4 22802 chr1 87051193 87051593
19	chr1:88143054-88143454	LINC01364	1.268	1.106	1.449	9 12 LINC01364 100505768 chr1 88143054 88143454
20	chr1:91852704-91853104	HFM1	0.632	7.002	6.340	993 555 HFM1 164045 chr1 91852704 91853104
21	chr1:94139707-94140107	BCAR3	0.606	1.268	0.000	11 0 BCAR3 8412 chr1 94139707 94140107
22	chr1:102291389-102291789	OLFM3	0.822	1.017	1.006	8 7 OLFM3 118427 chr1 102291389 102291789
23	chr1:11162476-11162876	KCNA2	0.496	1.017	0.000	8 0 KCNA2 3737 chr1 11162476 11162876
24	chr1:111628780-111629180	DRAM2	1.288	0.923	1.288	7 10 DRAM2 128338 chr1 111628780 111629180
25	chr1:112713218-112713618	LINC01790	0.732	1.106	0.657	9 4 LINC01790 643355 chr1 112713218 112713618
26	chr1:11354765-11354165	SLC16A1	1.214	0.822	1.106	6 8 SLC16A1 6566 chr1 11354765 11354165
27	chr1:121352060-121352800	EMBP1	0.669	3.272	2.692	125 70 EMBP1 647121 chr1 121352060 121352800
28	chr1:121353043-121353443	EMBP1	0.677	2.350	1.787	32 17 EMBP1 647121 chr1 121353043 121353443
29	chr1:121353640-121354516	EMBP1	0.375	4.732	3.317	437 138 EMBP1 647121 chr1 121353640 121354516
30	chr1:121354608-121355356	EMBP1	0.408	2.700	0.000	55 0 EMBP1 647121 chr1 121354608 121355356
31	chr1:121355448-121356302	EMBP1	0.452	5.192	4.014	523 156 EMBP1 647121 chr1 121355448 121356302
32	chr1:121356765-121357165	EMBP1	0.504	4.854	2.876	218 44 EMBP1 647121 chr1 121356765 121357165
33	chr1:121357819-121358219	EMBP1	0.420	4.964	2.818	236 42 EMBP1 647121 chr1 121357819 121358219
34	chr1:121358818-121359218	EMBP1	0.818	2.644	1.371	41 11 EMBP1 647121 chr1 121358818 121359218
35	chr1:121360559-121360459	EMBP1	1.402	1.106	1.594	9 14 EMBP1 647121 chr1 121360559 121360459

그림 1-2. DNA-seq or Microarray data in ExpADA format

1-1. Gene Category 사용 방법

DNA-Seq 은 수 만개의 유전자를 포함하기 때문에 유전자를 한 개씩 분석하기 보다 기능별로 그룹을 지어 분석을 하는 것이 용이하다. 이를 위해 많은 연구자들이 gene ontology (GO)를 활용한다. GO는 비슷한 기능의 유전자들을 묶어 놓은 그룹이라고 생각하면 이해하기 쉽다.

Gene Category 창은 수많은 GO 중 임의로 15 개를 선택하여 관련 유전자를 필터링 할 수 있도록 만들어 놓은 것이다. 예를 들어, Aging 관련 유전자만 분석을 원할 경우, Gene Category 창에서 Aging 을 선택하면 해당 유전자 리스트만 필터링 된다(그림 1-3). 그리고 Gene Category 의 여러 항목들을 동시에 만족하는 유전자를 필터링할 수 있고 적어도 한 항목만이라도 포함하는 유전자를 보고자 하는 경우도 필터링이 가능하도록 "AND"와 "OR" 기능을 갖추고 있다.

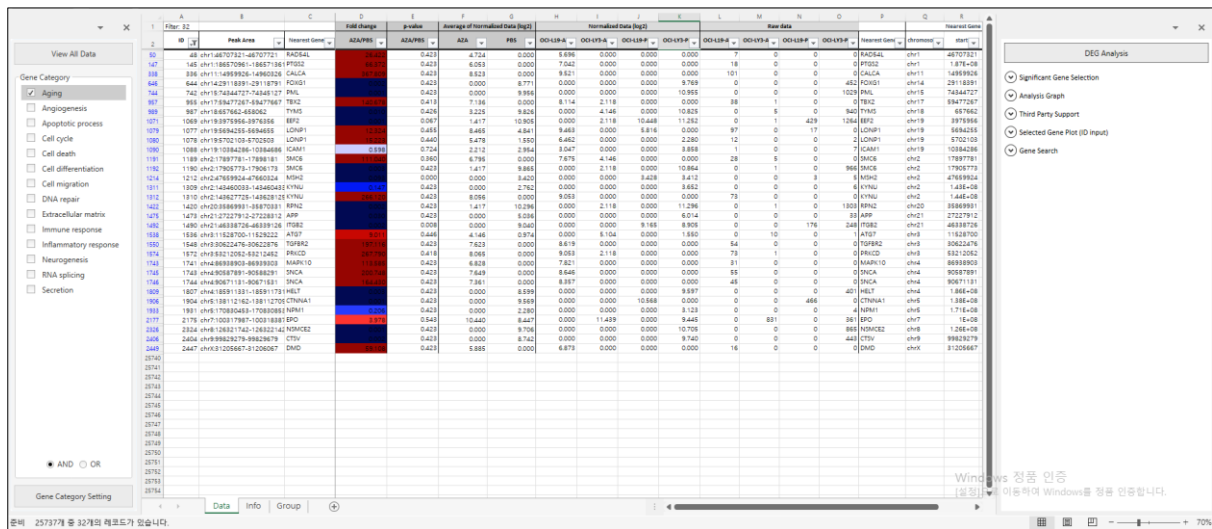


그림 1-3. Gene ontology (Aging) selection

가장 왼쪽 상단에 'View All Data' 버튼을 누르면 필터가 모두 해제되어 다시 전체 결과를 볼 수 있고 15 개의 GO 중 관심 기능이 없다면 'Gene Category Settings' 버튼을 이용하여 Quick GO site 에서 다른 GO 를 추가할 수 있다(그림 1-4). '?' 버튼을 누르면 GO 추가하는 방법이 자세히 설명되어 있다.

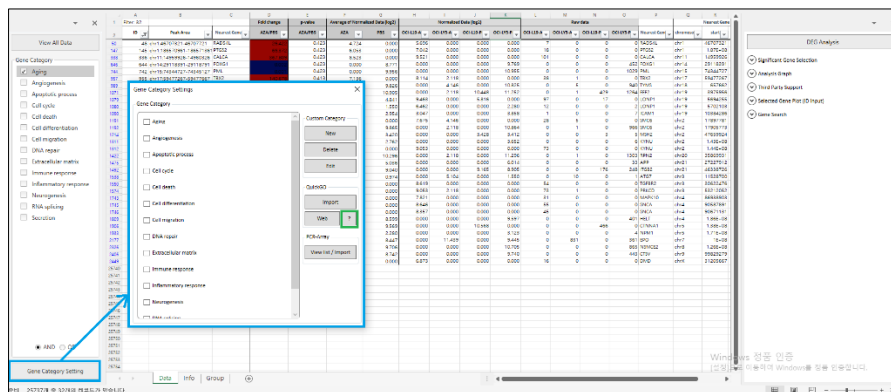


그림 1-4. Gene category settings

만약 원하는 유전자 그룹 목록을 알고 있다면, 직접 입력하여 새로운 Gene Category 를 추가할 수도 있다. Gene Category Settings 버튼을 누른 후 New 를 선택하고 원하는 Nearest Gene ID list 입력(or 복사&붙여넣기) 한 뒤, Gene category 이름 설정 후 저장(OK), Gene category Settings 창을 종료하면 새로운 Gene category 를 확인 할 수 있다(그림 1-5).

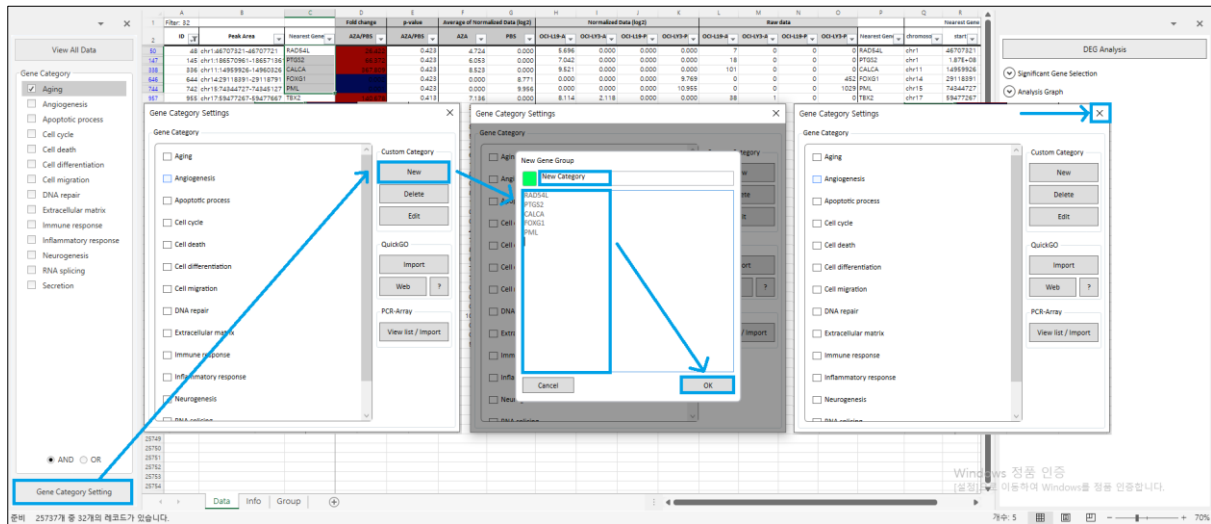


그림 1-5. Adding Genes to make a new gene category

PCR-Array 항목의 View list / Import 를 이용하여 Pathway 별 Gene list 를 추가 할 수 있다. Gene Category Settings 버튼을 누른 후 View list / Import 버튼을 누른다. Sub Window 창에서 Human/Mouse/Rat 등 species 를 선택하고, Keyword 에 추가하고자 하는 Pathway 이름이나 유전자 이름을 검색하고(A) Check box 에 체크한 뒤(B) Import 버튼을 누르면 자동으로 추가된다(그림 1-6).

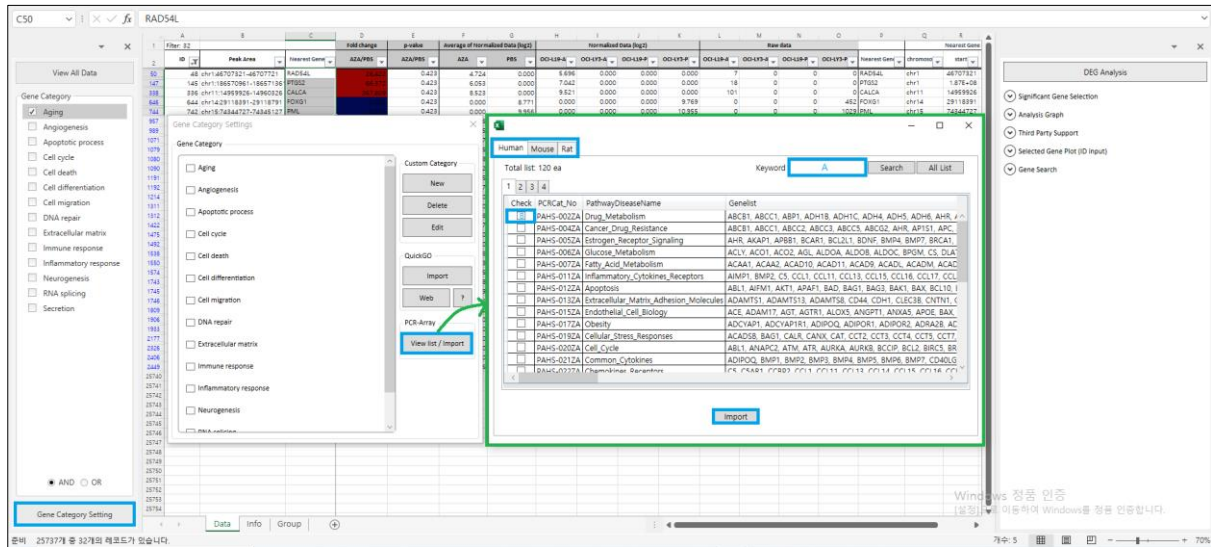


그림 1-6. PCR-Array Pathway settings

1-2. Significant Gene Selection 사용 방법

오른편의 DEG Analysis 부분에서 "Significant Gene Selection" 창은 전체 결과 중 대조군과 실험군을 비교한 결과에서 유의하게 발현 차이가 나는 유전자를 필터링 할 수 있도록 만들어 놓은 것이다. 예를 들어, B/A 비교조건을 선택하고 fold change:2, Normalized Data (log2):4, p-value:0.05 를 선택하면, A 대비 B 에서 2 배 이상 발현이 증가 또는 감소하고, Normalized Data (log2)값이 4 이상이고, p-value 값이 0.05 이하인 유전자가 필터링 된다(그림 1-7). P-value 는 반복 실험한 데이터(N>=2)의 경우만 제공된다. 비교그룹을 다중 선택할 수 있다. "AND"나 "OR"를 기능을 이용하면 선택한 비교 그룹들에서 공통적인 DEGs (교집합) 또는 하나의 비교그룹 이상 DEGs (합집합)을 선별할 수 있다.

참고로 유전자 선별 시 보통 fold change 2 이상, p-value 0.05 이하를 기준으로 선별한다. Normalized data (log2)는 정해져 있는 기준은 없으나 FPKM 의 경우엔 1 이상, RC 의 경우엔 4~6 이상을 기준으로 선별하면 발현값이 낮은 유전자들을 제외할 수 있다. 위 기준을 꼭 따라야 하는 것은 아니며, 유전자 선별 기준은 연구자의 데이터에 맞게 조정하여 사용할 수 있다.

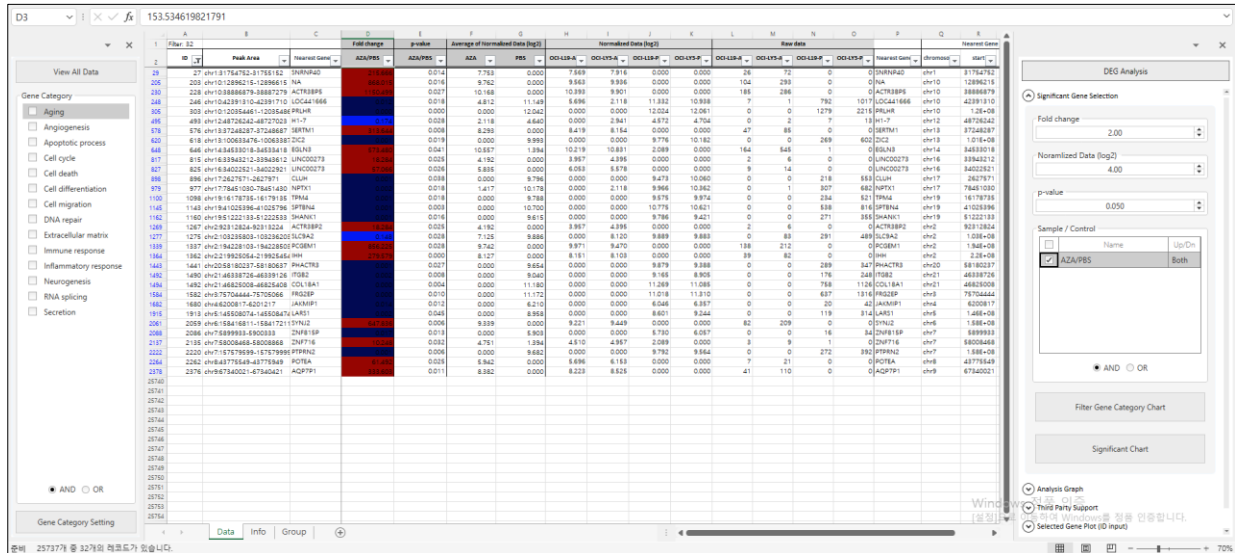


그림 1-7. Significant gene selection

Significant gene selection 에서 증가 또는 감소한 유전자를 각각 보고 싶다면 Up/Dn 의 selection box 에서 선택할 수 있다. Both 는 증가, 감소 유전자가 모두 필터링 되고 Up 은 증가한 유전자만, Dn 은 감소한 유전자만 따로 필터링 할 수 있다(그림 1-8)

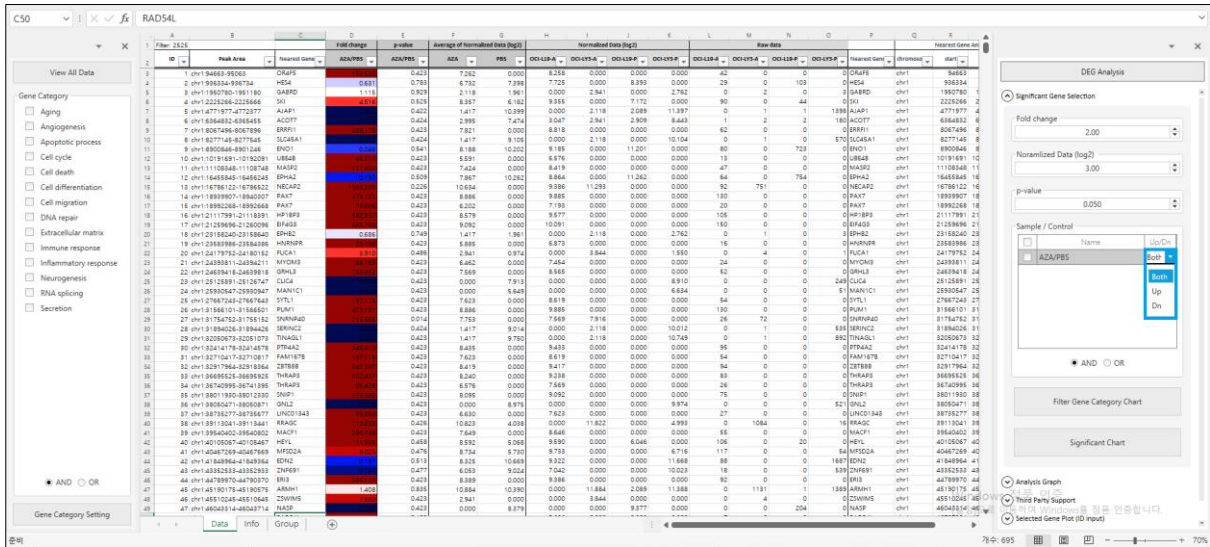


그림 1-8. Significant genes (separately up and down)

Gene Category와 Significant gene selection은 연동 가능하다. 그림 1-8에서처럼 Significant Gene Selection에서 DEGs를 선택하고 Gene Category의 Cell differentiation을 선택하면 9개의 유전자가 필터링 된다(그림 1-9). 9개의 유전자는 본 데이터에서 Cell differentiation 관련 유전자들 중 B/A 비교그룹에서 유의하게 발현이 증가 또는 감소한 유전자를 의미한다.

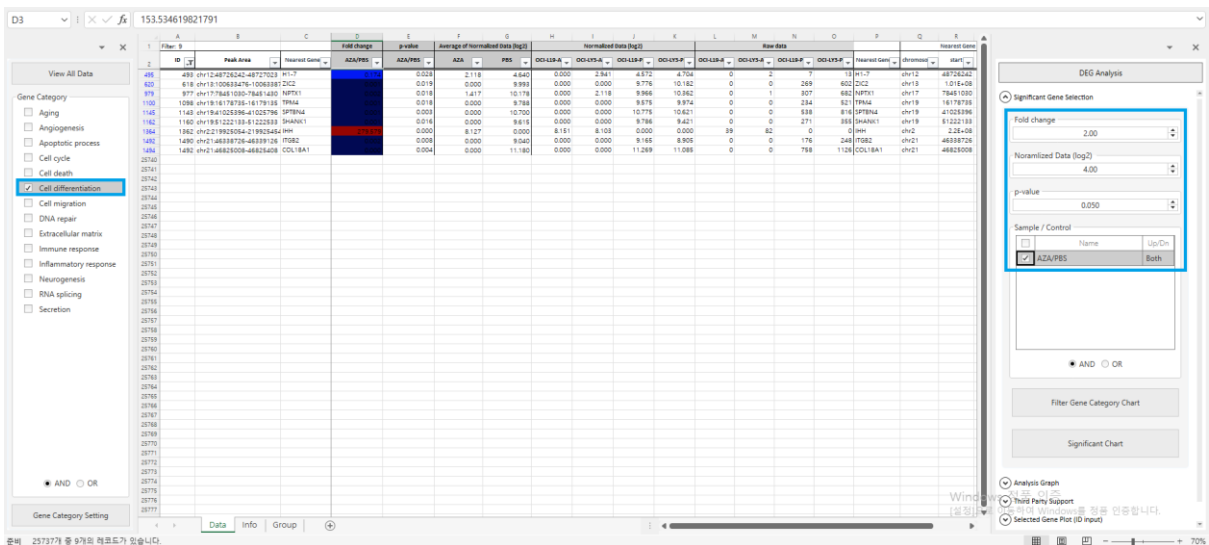


그림 1-9. Significant genes related to Cell differentiation

Gene Category Chart는 각 GO 관련 유전자 중 발현이 유의하게 차이 나는 유전자의 %와 수를 나타낸 그래프이다. 본 분석을 통해 어떤 GO의 유전자들이 상대적으로 많은 발현 변화가 있었는지를 확인할 수 있다. 전체 데이터 상태에서 Significant Gene Selection의 비교 그룹을 선택하고 "View Gene Category Chart"를 클릭하면 증가/감소한 유전자들 대상으로 GO Chart가 생성된다. 그래프의 각 영역을 클릭하면 해당 유전자들이 필터링 된다. 예를 들어 왼쪽의 Pie chart의 특정영역을 클릭하면 해당 GO의 증가/감소된 유전자가 함께 필터링 된다. 오른쪽의

증가/감소된 bar chart 에서 bar 상단의 숫자는 해당 유전자 수이며 bar 를 클릭했을 때 해당 유전자가 필터링 된다(그림 1-10).

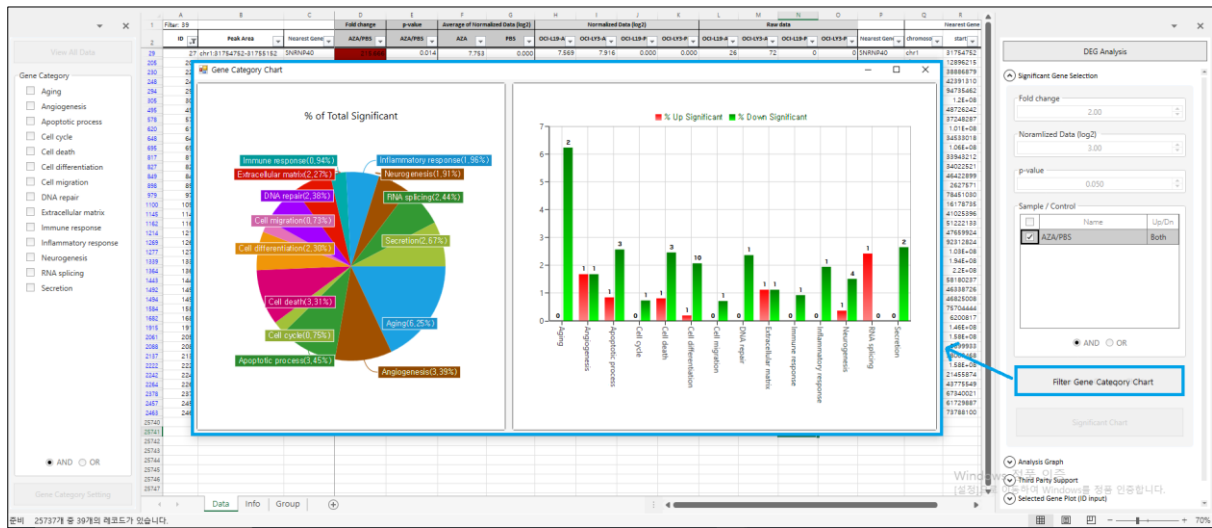


그림 1-10. View Gene Category Chart

1-3. Analysis Graph 사용 방법

DEG Analysis 부분에서 “Analysis Graph” 창을 펼치면 그림 1-11 와 같이 Scatter Plot, Volcano Plot, Venn Diagram 을 엑셀에서 쉽게 그릴 수 있다.

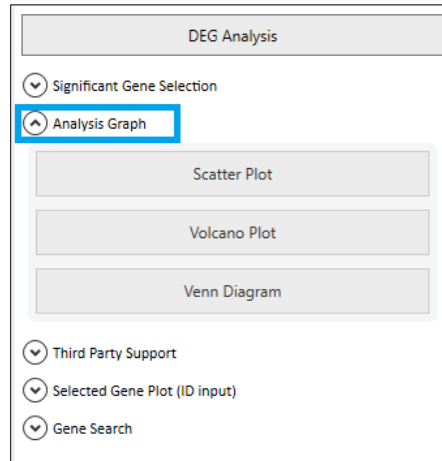


그림 1-11. Analysis Graph Tool

1-3-1. Scatter plot

Scatter Plot 은 대조군과 실험군의 발현양상을 확인할 수 있는 이미지이다. 오른쪽에 샘플 비교 그룹과 Fold threshold line (예시: 2fold)을 선택하고 “Graph View”를 클릭하면 왼쪽에 선택한 비교 그룹을 대상으로 Scatter Plot 이 자동 생성된다. x 축은 대조군의 normalized data (log2), y 축은 실험군의 normalized data (log2)이다. 초록색 사선 아래는 2fold 이상 감소한 유전자들, 빨간색 사선 위 2fold 이상 증가한 유전자들이다. 유의한 유전자를 식별하기 위한 색은 빨강, 파랑, 초록 중 사용자가 선택할 수 있다. Plot 에서 특정 spot 을 클릭하면 해당 유전자명이 표시되고 마우스 오른쪽을 클릭하여 지울 수도 있다. 그리고 여러 개의 유전자를 동시에 표시하고 싶다면 “Gene

Select(ID Input)" 창에 해당 유전자 ID 를 복사하여 입력하고 "Add"를 클릭하면 Nearest Gene ID 이 자동 생성된다(그림 1-12).



그림 1-12. Analysis Graph Tool – Scatter Plot

1-3-2. Volcano plot

Volcano Plot 은 반복 실험($N \geq 2$)이 된 경우에만 분석 가능하다. Volcano Plot 은 Scatter Plot 의 기능과 거의 동일한데 오른쪽에 샘플 비교 그룹과 Fold threshold line (예시: 2fold), p-value (예시: p-value 0.05)를 선택하고 "Graph View"를 클릭하면 왼쪽에 선택한 비교 그룹을 대상으로 Plot 이 자동 생성된다. 초록색 세로선 왼쪽은 2fold 이상 감소한 유전자들, 빨간색 세로선 오른쪽은 2fold 이상 증가한 유전자들, 검은색 가로선 위는 p-value 0.05 이하인 유전자들이다. 유의한 유전자를 식별하기 위한 색은 빨강, 파랑, 초록 중 사용자가 선택할 수 있다. Plot 에서 특정 spot 을 클릭하면 해당 유전자명이 표시되고 마우스 오른쪽쪽을 클릭하여 표시를 지울 수도 있다. 그리고 여러 개의 유전자를 동시에 표시하고 싶다면 "Gene Select(ID Input)" 창에 해당 유전자 ID 를 복사하여 입력하고 "Add"를 클릭하면 Nearest Gene ID 가 자동 생성된다(그림 1-13).

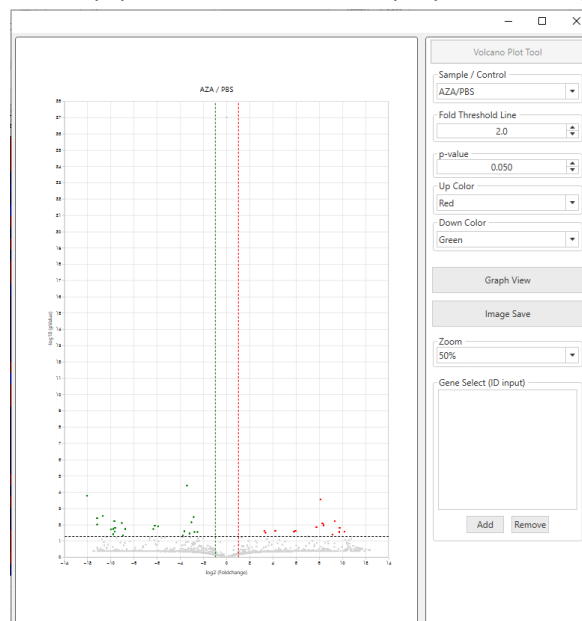


그림 1-13. Analysis Graph Tool – Volcano Plot

1-3-3. Venn diagram

Venn Diagram 을 통해 4 개 이하의 비교그룹을 대상으로 Venn Diagram 을 작성할 수 있다. Venn Diagram 을 그릴 샘플 비교그룹과 Fold Change, Normalized data (log2), p-value 을 선택 후, Diagram View 를 클릭하면 결과를 확인할 수 있으며 그룹은 최대 4 그룹까지 선택 가능하다. 아래의 그림은 B/A, C/A, C/B 결과 중, 2fold, Normalized data (log2) 4 이상, p-value 0.05 이하인 유전자 list 를 가지고 Venn Diagram 을 작성한 결과이다(그림 1-14).

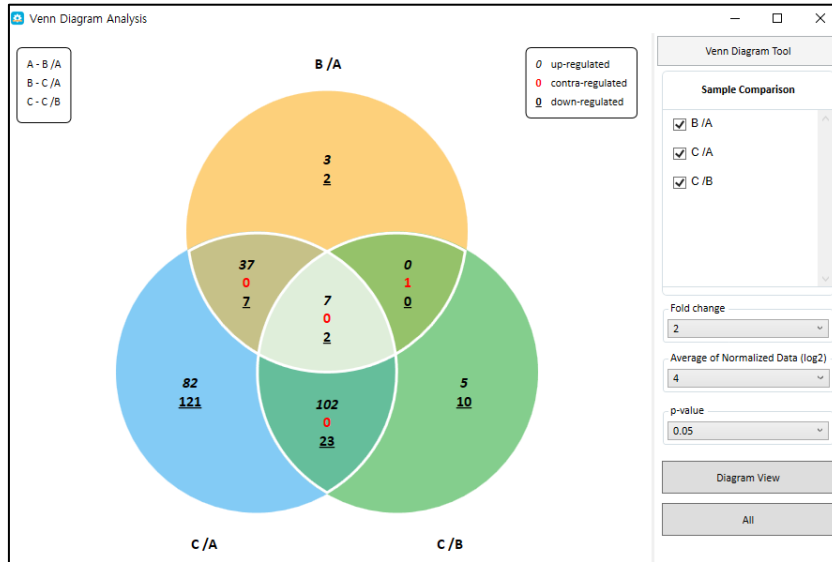


그림 1-14. Analysis Graph Tool – Venn Diagram

Venn Diagram 결과에서 표시되는 형식은 다음과 같다(그림 1-15).

1. **기울어진 숫자** : up-regulated 된 gene 수
2. **빨간색 숫자** : regulation 이 대조되는 gene 수
(예: B/A 에서는 up 되고 C/A 에서는 down 되는 gene 수)
3. **밑줄 친 숫자** : down-regulated 된 gene 수

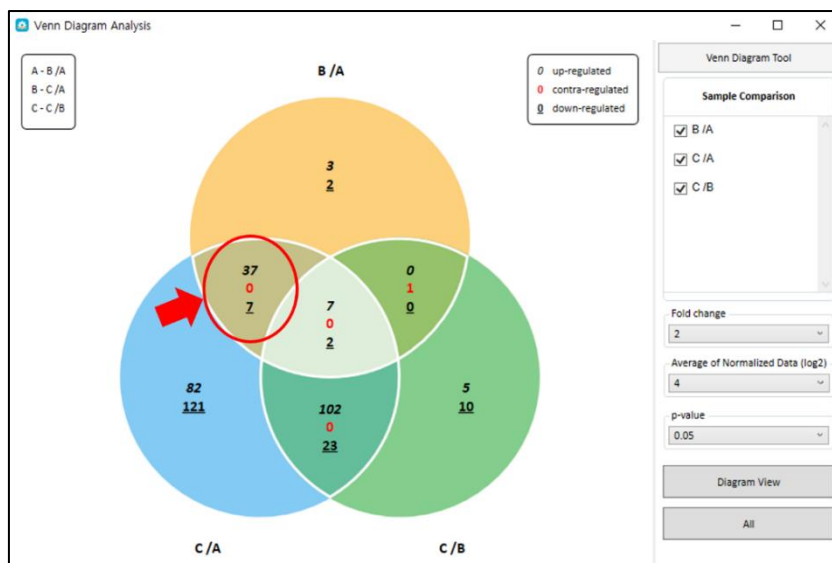


그림 1-15. For example of up, down, contra-regulated in Venn Diagram

Venn Diagram 각 영역에 어떤 유전자들이 있는지 확인할 수도 있다. 예를 들어, B/A 에서만 up 이 되는 유전자를 보고 싶으면, Venn Diagram 에서 B/A 에서만 해당되는 영역을 찾아 마우스 오른쪽 클릭하고 up-regulated 를 선택하면 유전자 list 3 개가 엑셀 data sheet 에 filter 된다(그림 1-16).

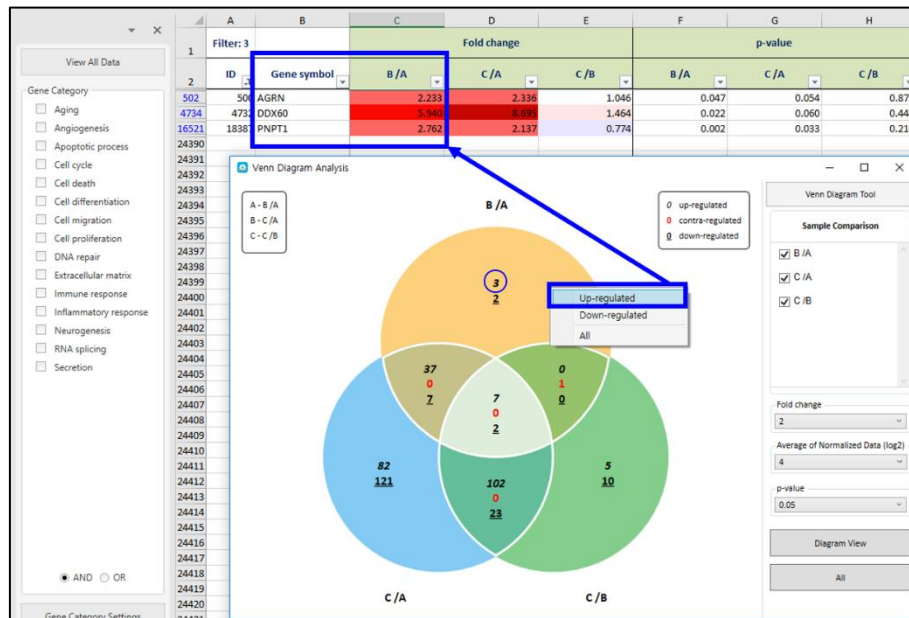


그림 1-16. Filtering 2fold up-regulated gene list in Venn Diagram

ExpADA 에서 제공되는 모든 이미지는 오른쪽마우스를 눌러 'Save image' 버튼을 통해 저장 가능하다(그림 1-17).



그림 1-17. Save image

1-4. Third Party Support 사용 방법

Third Party Support 는 연구자가 선택한 유전자를 기반으로 Clustering heatmap 과 KEGG 분석, DAVID 분석을 수행하기 위한 입력 데이터를 제공한다. 먼저, Input File 제작에 앞서 유전자를 선별하는 것이 필요하다. 유전자 선별은 유의성 있는 DEG 분석, Gene Ontology 분석 등으로 선별할 수 있다 (그림 1-18).

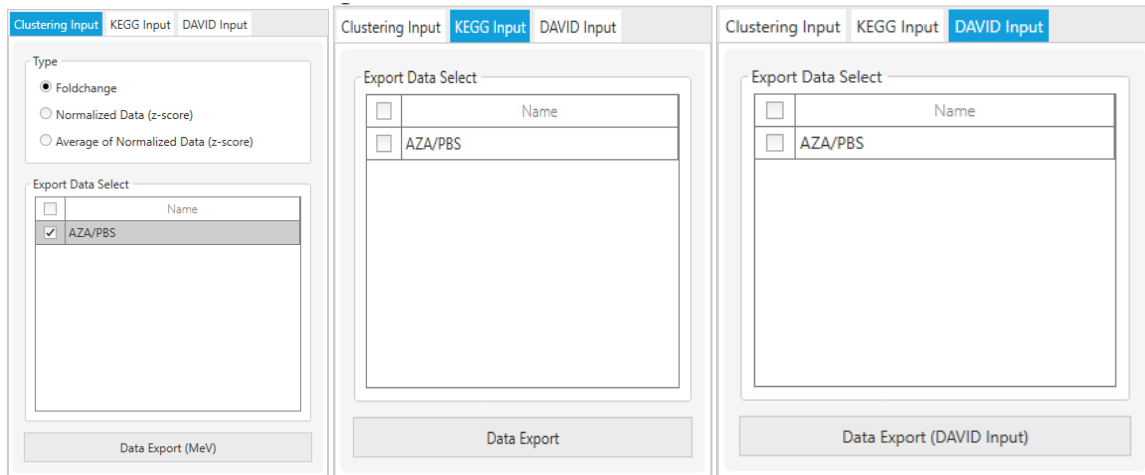


그림 1-18. Third Party Support (왼편부터 Clustering Input, KEGG Input, DAVID Input)

필터링 된 유전자 리스트를 대상으로 Clustering Heatmap 을 작성하려면 크게 두 종류의 데이터를 이용할 수 있다. Clustering Input 은 비교 조합의 Fold change 를 이용하여 제작된다. 원하는 비교 조합을 선택한 후 Data Export 버튼을 누르면 파일을 저장할 수 있다. 파일명에는 띄어쓰기가 들어가지 않도록 주의한다. Clustering Input 파일은 유전자 이름과 선택한 비교 조합, Z-score 로 구성된다. Clustering Input 파일은 **MeV** 프로그램을 이용하여 Clustering heatmap 을 작성할 수 있다.

첫 번째, Fold change 값을 이용할 시 Type 부분에 Fold change 를 체크하고 Export Data Select 에서 Heatmap 에 표현할 비교그룹을 체크한다. "Data Export"를 클릭한 후 "(input 명).txt"로 저장한다. Input file 명에는 띄어쓰기가 들어가지 않도록 주의한다.

두 번째, 개별 샘플의 발현 값인 Normalized Data 로 표현하고자 할 때 Z-Score 를 체크하고 확인하고자 하는 샘플을 체크한다. "Data Export"를 클릭한 후 "(input 명).txt"로 저장한다. 단, Z-score 로 그릴 때는 샘플 3 개 이상에서만 가능하다.

* Z-score 는 일반적으로 평균으로부터 얼마만큼 떨어져 있느냐를 판단하는 지표이다. 계산방식은 Normalized data 를 log10 으로 변환 후 평균값을 뺀 후 표준편차로 나누어 계산한다.

$$Z\text{-score} = \frac{\{\text{Normalized data} (\log_{10}) - \text{average of Normalized data} (\log_{10})\}}{\text{standard deviation of Normalized data}(\log_{10})}$$

Clustering heatmap input file 은 Nearest Gene ID 과 fold change (log2) 또는 z-score 로 구성된다(그림 1-19). Clustering heatmap 작성법은 본 매뉴얼의 3. Clustering heatmap analysis 부분에 설명되어 있다.

Nearest Gene ID	AZA/PBS
OR4F5	7.26242000252593
HES4	-0.665359419780513
GABRD	0.156928548725499
SKI	2.17492823841489
AJAP1	-8.98256701264007
ACOT7	-4.479341166372
ERRF1	7.82126504330931
SLC45A1	-7.68804942783477
ENO1	-2.01408459953478
UBE4B	5.59135328983294
MASP2	7.42369161236887
EPHA2	-2.39529403012508
NECAP2	10.6341694072588
PAX7	8.8860958804934

그림 1-19. Clustering heatmap input file

KEGG input 은 분석 결과에서 Up-/Down-regulated 된 유전자들이 어떤 Pathway 에 속하는지 확인하고자 할 때 **KEGG Mapper** 를 이용하기 위한 입력 데이터를 제공한다. KEGG Input 에서는 하나의 비교 조합만 선택 가능하다. 비교 조합 선택 후 Data Export 를 선택하면 그림 1-20 과 같이 유전자 이름과 비교 조합, 발현 수준에 따른 색 코드로 구성된다. KEGG Input 파일은 **KEGG Mapper** 의 입력 데이터로 사용하여 Pathway 상에 속하는 유전자와 이들의 발현 수준을 확인할 수 있다.

Nearest Gene ID	AZA/PBS
OR4F5	#FF0000,black
HES4	#B0E0E6,black
GABRD	.
SKI	#FF4500,black
AJAP1	#1E90FF,black
ACOT7	#1E90FF,black
ERRF1	#FF0000,black
SLC45A1	#1E90FF,black
ENO1	#00BFFF,black
UBE4B	#FF0000,black
MASP2	#FF0000,black
EPHA2	#00BFFF,black
NECAP2	#FF0000,black
PAX7	#FF0000,black
PAX7	#FF0000,black
HP1BP3	#FF0000,black
EIF4G3	#FF0000,black

그림 1-20. KEGG input file

DAVID 는 다양한 데이터베이스를 기반으로 유전자의 상관관계를 통계적으로 분석하여 유전자의 주요 기능을 예측하는 Analysis tool 이다. DAVID 는 3 천 개 이상의 유전자는 분석할 수 없으므로 3 천 개 이하로 유전자를 선별해야 한다. Data Export (DAVID Input)는 DAVID 에서 유전자 이름을 입력하는 부분에 사용되는 파일이다. 이 파일은 유전자 이름으로 구성된다 (그림 1-21). 유전자 이름을 DAVID 의 입력 데이터로 사용하여 분석하려는 GO 또는 Pathway 에 대한 데이터를 다운로드 받는다.



그림 1-21. DAVID input file

1-5. Selected Gene Plot & Gene Search 사용 방법

ExpADA의 기능 중에 선별한 유전자 또는 연구자가 관심있는 유전자들을 대상으로 발현 패턴을 그래프로 표현하고자 할 때는 "Selected Gene Plot" 기능을 사용할 수 있다. 선별한 유전자의 ID를 복사하여 Selected Gene Plot 창에 붙여 넣고 "Expression Plot View"를 누르면 Normalized data (log2) 값, Fold change (log2) 값으로 line graph가 그려진다(그림 1-23).

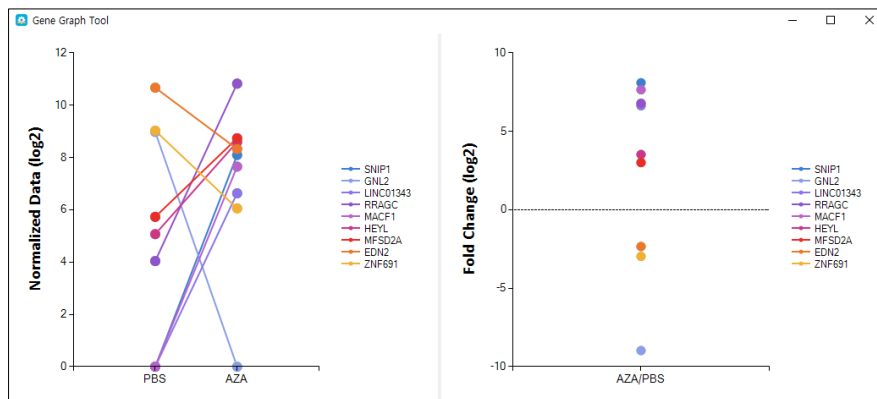


그림 1-23. Gene graph

특정 keyword 관련 유전자를 검색하고 싶을 때는 gene search 창을 이용하면 된다. 예를 들어 'PAX'를 검색하면 엑셀 Data Sheet에 'PAX' keyword를 포함하는 행만 필터링하여 확인할 수 있다(그림 1-24).

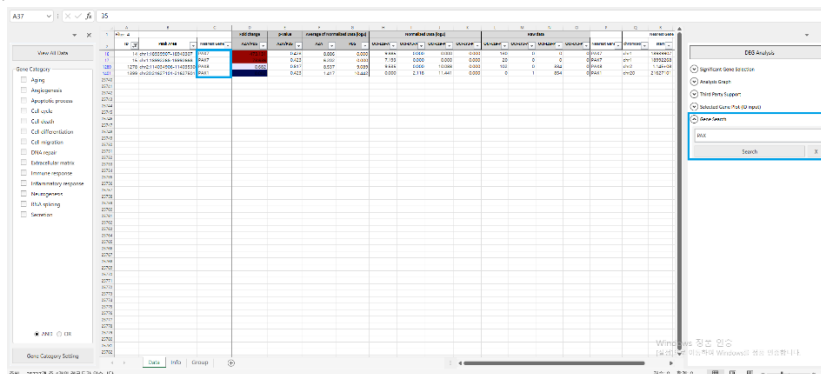


그림 1-24. Genes related to PAX

2. Functional Annotation Analysis (DAVID)

2-1. DAVID 분석 틀을 이용한 Functional Annotation 분석

DAVID 는 다양한 데이터 베이스를 기반으로 유전자의 상관관계를 통계적으로 분석하여 유전자의 주요 기능을 예측하는 analysis tool 이다. 분석 과정은 그림 2-1 과 같다.

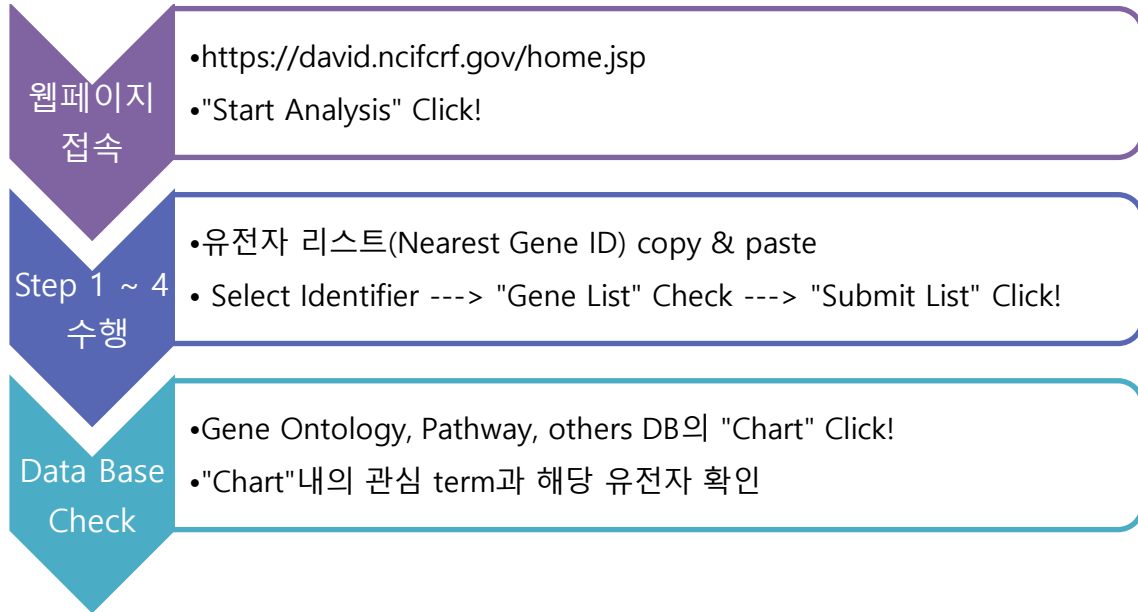


그림 2-1. DAVID tool analysis process

DAVID 에서는 3 천 개 이상의 유전자는 분석할 수 없으므로 3 천 개 이하로 유전자를 선별해야 한다. DNA-Seq 결과에서 significant gene 을 선별하여 DAVID 분석을 한다. DAVID 홈페이지 (<https://david.ncicrf.gov/home.jsp>)에 접속하여 “Start Analysis”을 클릭한다 (그림 2-2).

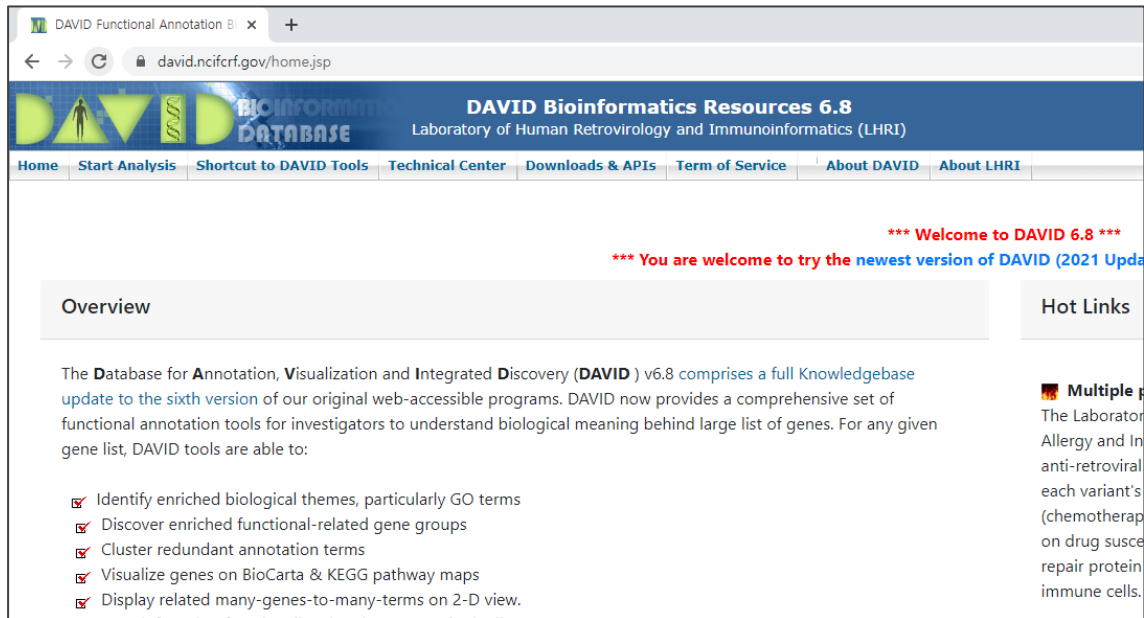


그림 2-2. DAVID tool webpage

“Upload” 탭에서 Step 1 에서 Step 4 까지 수행한다(그림 2-3). Step 1 에서 선별한 유전자의 Gene Symbol 을 복사하고 “A: Paste a list” 창에 붙여 넣는다. Step 2 에서 “OFFICIAL_GENE_SYMBOL”를 선택한다. 만약 step 1 에서 Gene Bank No.를 넣었다면 “GENEBANK_ACCESSION”을 선택한다. Step 2a 에서 분석하는 종의 학명을 입력한다. Step 3 에서 “Gene List”를 체크하고 Step 4 에서 “Submit List”를 누른다.

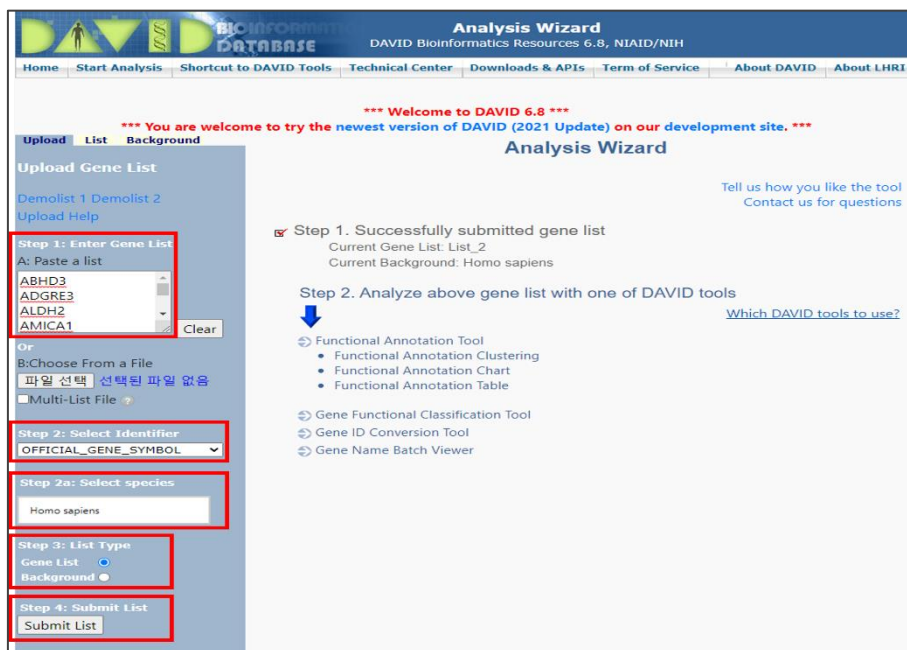


그림 2-3. DAVID tool : Step 1 ~ Step 4

List sheet 에서 분석하고자 하는 종을 선택한다(그림 2-4. a). "List" Sheet 에서 해당 종(숫자)로 표기되어 있고 가로 안의 숫자가 분석에 적용된 유전자의 개수이다. 예시에서는 269 개의 유전자 리스트를 넣었고 데이터베이스에서 기능이 밝혀진 267 개만이 Functional Annotation 분석에 이용되었다는 의미이다. Current Background 에 분석하고자 하는 종이 아닌 다른 종이 나왔다면 좌측 "Background" Sheet 에서 알맞은 종을 선택하여 "Use"를 클릭한다(그림 2-4. b).

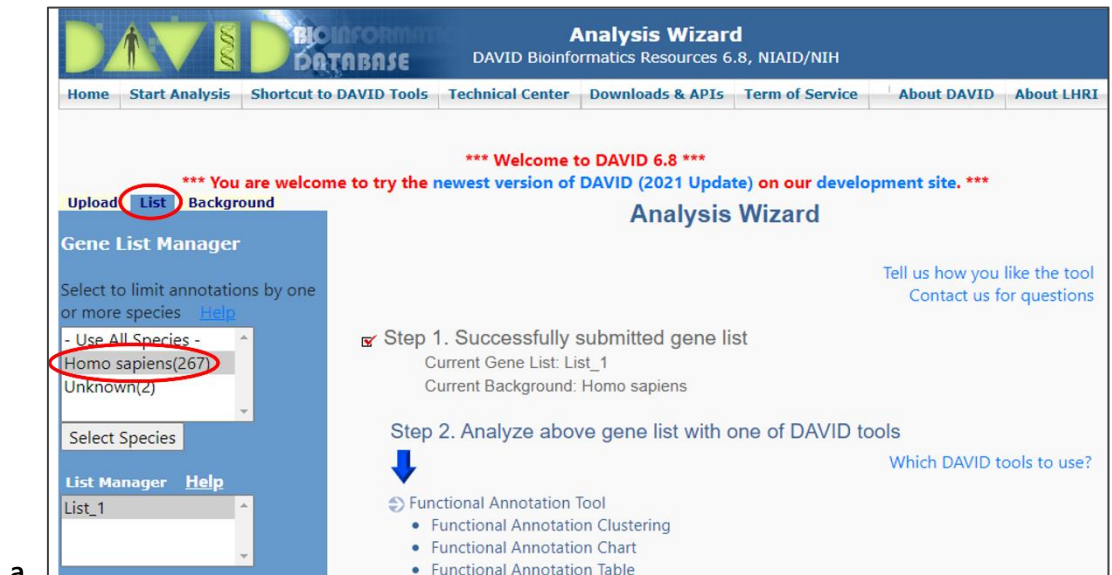


그림 2-4. DAVID tool : Select Species

DAVID 분석 결과 중 Gene Ontology Biological Process 결과를 확인하려면 "Gene_Ontology"의 "+" 표시를 클릭하여 결과 창을 열고 "GOTERM_BP_Direct"의 "Chart"를 누른다(그림 2-5). Input 한 유전자들이 유의하게 관여하는 GO list가 나온다. 관심 GO를 클릭하면 QuickGO 데이터베이스로 연결되어 각 GO의 정보를 확인할 수 있다. GO의 Gene 막대를 클릭하면 해당 GO 관련 유전자들을 확인할 수 있다.

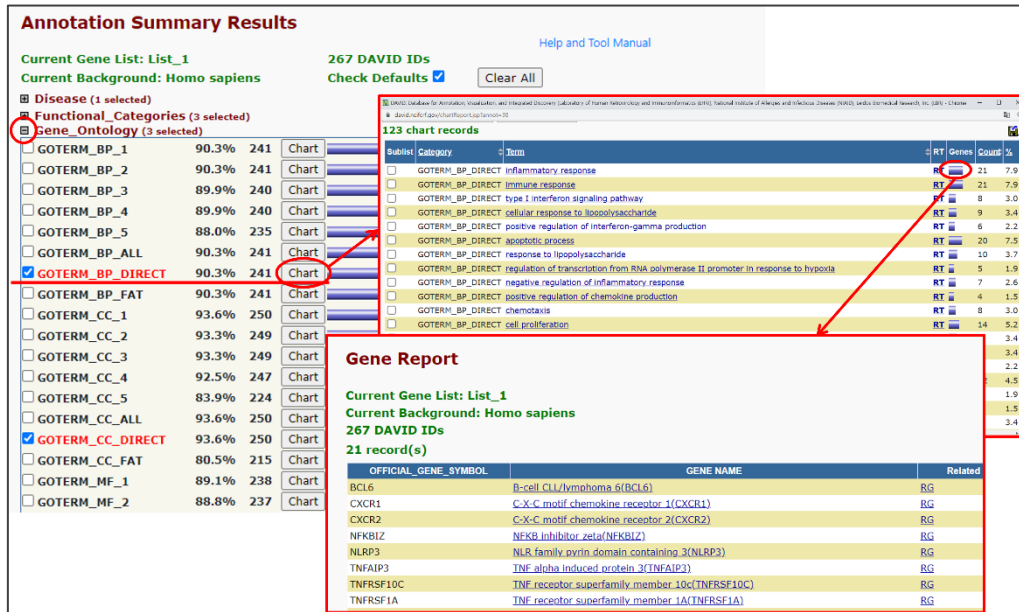


그림 2-5. DAVID tool : exploring Gene Ontology analysis result

이와 같은 방법으로 Pathway 결과를 확인해 보면 KEGG_PATHWAY database에서 주요 Pathway가 나온다(그림 2-6). 각 pathway를 누르면 pathway 그림을 확인할 수 있다. pathway 그림에서 별 표시가 되어 있는 유전자가 input 유전자 중 해당 pathway에 관여하는 유전자이다. 유전자를 클릭하면 유전자 정보도 자세히 알 수 있다.

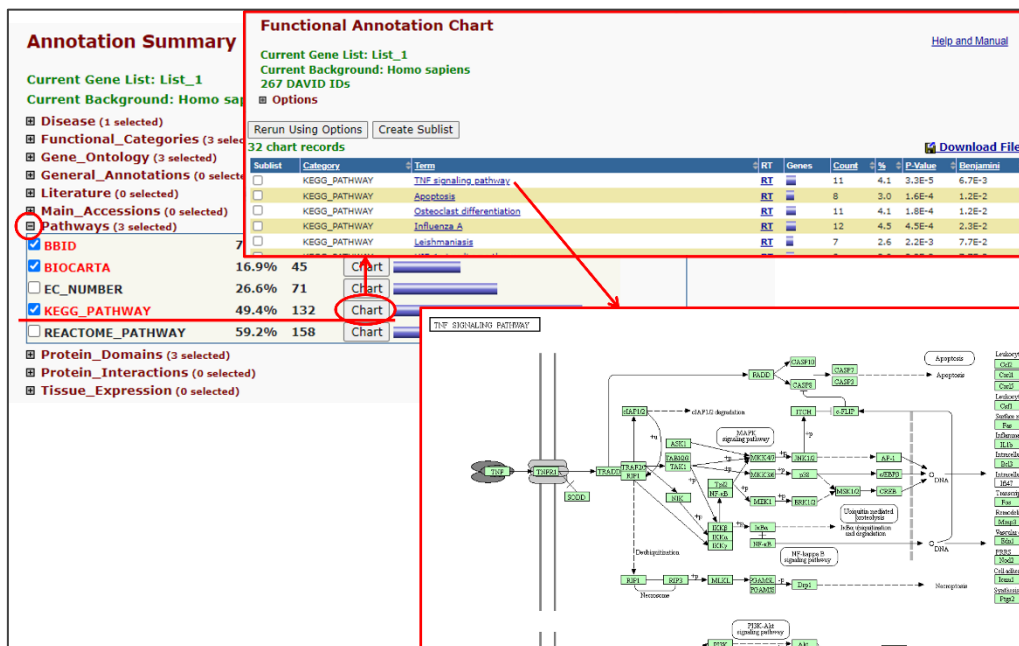


그림 2-6. DAVID tool : exploring Pathway analysis result

DAVID 분석은 input 한 유전자들이 유의하게 관련되는 GO, pathway 등을 분석하기에 유용한 tool 이다. 즉, input 한 유전자에서 많은 유전자들이 관련되는 GO, pathway 만 결과로 나오기 때문에 input 유전자 중 적은 수가 관련되는 GO, pathway 는 결과에 나오지 않는다. 또한 input 유전자의 수가 적으면 분석 결과가 없을 수도 있다. DAVID 에서는 유전자 2 개 이상, EASE score 0.1 이하를 default 로 분석하여 이 기준에 적합한 결과를 보여준다. option 에서 이 기준을 조정하여 리스트를 더 볼 수 있다. David 분석 결과의 각 항목은 DAVID 홈페이지의 Help and Tool Manual 에 자세히 설명되어 있다(그림 2-7).

Annotation Summary Results

Current Gene List: List_1 94 DAVID IDs
Current Background: Homo sapiens Check Defaults Clear All

- Disease (1 selected)
- Functional_Categories (3 selected)
- Gene_Ontology (3 selected)
- General_Annotations (0 selected)
- Literature (0 selected)

[Help and Tool Manual](#)

Functional Annotation Chart

Current Gene List: demolist1
Current Background: Homo sapiens
171 DAVID IDs

Options

Count Threshold EASE Threshold # of Records Displayed

Rerun Using Options Create Sublist [Download File](#)

Sublist	Category	Term	RT	Genes	Count	%	P-Value
<input type="checkbox"/>	SP_PIR_KEYWORDS	signal	RT	<div style="width: 47%; height: 10px; background-color: blue;"></div>	47	27.5%	3.0E-10
<input type="checkbox"/>	SP_PIR_KEYWORDS	glycoprotein	RT	<div style="width: 51%; height: 10px; background-color: blue;"></div>	51	29.8%	4.9E-8
<input type="checkbox"/>	GOTERM_CC_ALL	extracellular region	RT	<div style="width: 32%; height: 10px; background-color: blue;"></div>	32	18.7%	1.1E-7
<input type="checkbox"/>	SP_PIR_KEYWORDS	alternative splicing	RT	<div style="width: 49%; height: 10px; background-color: blue;"></div>	49	28.7%	6.4E-6
<input type="checkbox"/>	SP_PIR_KEYWORDS	chromoprotein	RT	<div style="width: 7%; height: 10px; background-color: blue;"></div>	7	4.1%	1.1E-5
<input type="checkbox"/>	SP_PIR_KEYWORDS	direct protein sequencing	RT	<div style="width: 33%; height: 10px; background-color: blue;"></div>	33	19.3%	1.2E-5
<input type="checkbox"/>	SP_PIR_KEYWORDS	phosphorylation	RT	<div style="width: 31%; height: 10px; background-color: blue;"></div>	31	18.1%	1.6E-5
<input type="checkbox"/>	UP_SEQ_FEATURE	signal peptide	RT	<div style="width: 47%; height: 10px; background-color: blue;"></div>	47	27.5%	3.7E-5
<input type="checkbox"/>	SP_PIR_KEYWORDS	metalloprotein	RT	<div style="width: 8%; height: 10px; background-color: blue;"></div>	8	4.7%	4.7E-5
<input type="checkbox"/>	GOTERM_BP_ALL	response to chemical stimulus	RT	<div style="width: 14%; height: 10px; background-color: blue;"></div>	14	8.2%	6.1E-5

Percentage, e.g. 14/171=8.2% (involved genes/total genes)

그림 2-7. DAVID Help and Tool Manual

DAVID 분석 결과를 내 컴퓨터에 저장하려면, 초록색으로 표시된 Download File 링크를 마우스 오른쪽 버튼을 클릭한 후 다른 이름으로 저장을 선택하면 파일을 다운로드 받을 수 있다(그림 2-8).

* 주의사항

: internet explorer 를 이용할 경우 다른 이름으로 저장 버튼이 보이지 않기 때문에, Chrome 을 이용하여 분석하기를 권장한다.

Functional Annotation Chart
 Current Gene List: demolist1
 Current Background: Homo sapiens
 171 DAVID IDs

Options
 Count Threshold: 2 EASE Threshold: 0.1 # of Records Displayed: 1000

Download File (highlighted with a green arrow)

Sublist	Category	Term	RT	Genes	Count	%	P-Value
<input type="checkbox"/>	SP_PIR_KEYWORDS	signal	RT		>47	27.5%	>3.0E-10
<input type="checkbox"/>	SP_PIR_KEYWORDS	glycoprotein	RT		51	29.8%	4.9E-8
<input type="checkbox"/>	GOTERM_CC_ALL	extracellular region	RT		32	18.7%	1.1E-7
<input type="checkbox"/>	SP_PIR_KEYWORDS	alternative splicing	RT		49	28.7%	6.4E-6
<input type="checkbox"/>	SP_PIR_KEYWORDS	chromoprotein	RT		7	4.1%	1.1E-5
<input type="checkbox"/>	SP_PIR_KEYWORDS	direct protein sequencing	RT		33	19.3%	1.2E-5
<input type="checkbox"/>	SP_PIR_KEYWORDS	phosphorylation	RT		31	18.1%	1.6E-5
<input type="checkbox"/>	UP_SEQ_FEATURE	signal peptide	RT		47	27.5%	3.7E-5
<input type="checkbox"/>	SP_PIR_KEYWORDS	metalloprotein	RT		8	4.7%	4.7E-5
<input type="checkbox"/>	GOTERM_BP_ALL	response to chemical stimulus	RT		14	8.2%	6.1E-5

Annotations:
 - Gene list and population background being analyzed (points to 'Current Gene List' and 'Current Background')
 - Minimum number of genes for the corresponding term (points to 'Count Threshold')
 - Maximum EASE Score/P-Value (points to 'EASE Threshold')
 - Maximum number of record per page (points to '# of Records Displayed')
 - Original database/resource where the terms orient (points to 'GOTERM_BP_ALL')
 - Enriched terms associated with your gene list (points to 'response to chemical stimulus')
 - Related Term Search (points to 'RT' column)
 - Genes involved in the term (points to 'Genes' column)
 - Percentage, e.g. 14/171=8.2% (involved genes/total genes) (points to '%' column)
 - Modified Fisher Exact P-Value, EASE Score. The smaller, the more enriched. (points to 'P-Value' column)

그림 2-8. DAVID data download

3. Pathway analysis (KEGG mapper)

DNA-Seq 분석 결과에서 up/down-regulated genes들이 어떤 Pathway에 속하는지 확인하고자 한다면 KEGG에서 제공하는 KEGG Mapper를 이용한다. 사용방법은 그림 5-1과 같은 순서로 진행된다.

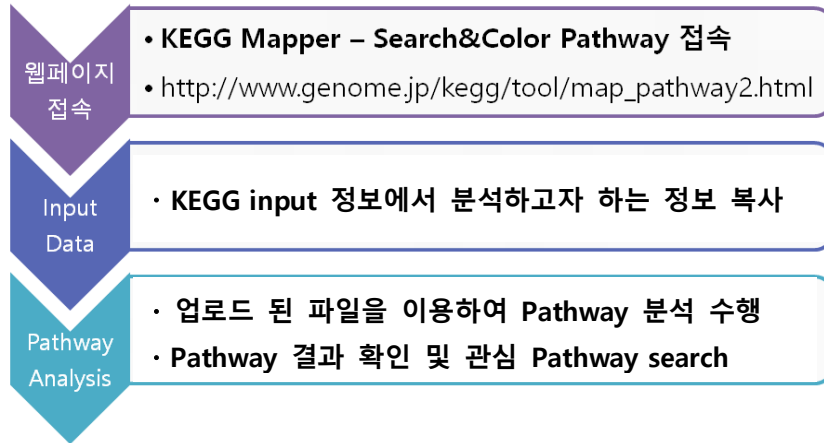


그림 5-1. KEGG Mapper tool analysis process

그림 5-2는 mRNA-Seq report에서 2fold, normalized data(log2)>4, p-value<0.05을 기준으로 선별한 유전자를 KEGG 분석하는 과정이다.

Significant gene selection에서 Fold change, Normalized Data(log2), p-value (반복실험의 경우) 값을 지정하고, 확인하고자 하는 Fold change 조합을 선택하여 필터를 적용한다.

필터를 적용하여 선별된 유전자를 대상으로 KEGG input [Entrez ID, FC Color(#숫자,black)]를 복사하여, KEGG 분석에 사용한다. 원하는 비교조합의 컬러 정보가 반드시 Entrez ID열 바로 옆에 나란히 놓이게 한 후(열 숨기기 기능 이용), 진행해야 한다.

그림 5-2. KEGG Mapper input file generation process

그림 5-3과 같이 KEGG Mapper 웹페이지(<https://www.genome.jp/kegg/mapper/>)에 접속하고 Color 항목에 들어가면 아래와 같은 화면을 볼 수 있다.

- (1) 분석하고자 하는 유전자의 species를 선택 (Human이면 hsa 선택)
- (2) KEGG input 정보에서 분석하고자 하는 정보를 복사하여 빈칸에 입력
- (3) "Use uncolored diagram"과 "Include aliases" 항목에 체크를 한 후
- (4) Exec 버튼을 누른다.

KEGG Mapper – Color

KEGG2 About Reconstruct Search Color Join Convert ID Assign KO Taxonomy

Color tool

The Color tool searches various KEGG objects, including genes, KOs, EC numbers, metabolites and drugs, against KEGG pathway maps. Found objects may be marked in any combination of background and foreground colors.

Search mode: Reference hsa other org

Enter KEGG identifiers followed by color specification

81671	#FF6347,black
8875	#FF6347,black
84272	#FF6347,black
80149	#00BFFF,black
84243	#FF6347,black
253832	#87CEEB,black
55625	#87CEEB,black
7538	#87CEEB,black
677	#FF6347,black
25946	#87CEEB,black

Examples:
Select

Or upload file:

Default bgcolor:

Use uncolored diagrams

Include aliases (for hsa and other org modes)

그림 5-3. KEGG Mapper tool analysis process

분석결과, 입력한 유전자들이 관여하는 pathway list 가 나온다(그림 5-4). pathway 이름 옆에 있는 괄호 안 숫자는 입력한 유전자 중 각 pathway 에 관여하는 유전자의 수이다. 괄호 안 숫자를 클릭하면 해당 유전자 목록을 볼 수 있다. pathway 이름을 클릭하면 해당 pathway chart 가 열리고 입력한 유전자의 발현 up/down (red/blue)이 색으로 표시되어 있다. Pathway 이미지는 "다른 이름으로 저장"이 가능하고 "html"으로 저장하면 이미지에 링크된 항목을 그대로 유지해서 저장이 가능하다.

***참고사항**

만약 오른쪽마우스 버튼을 클릭했을 때 다른 이름으로 저장이 보이지 않을 경우, Internet explorer 대신 chrome 창을 이용하면 확인할 수 있다.

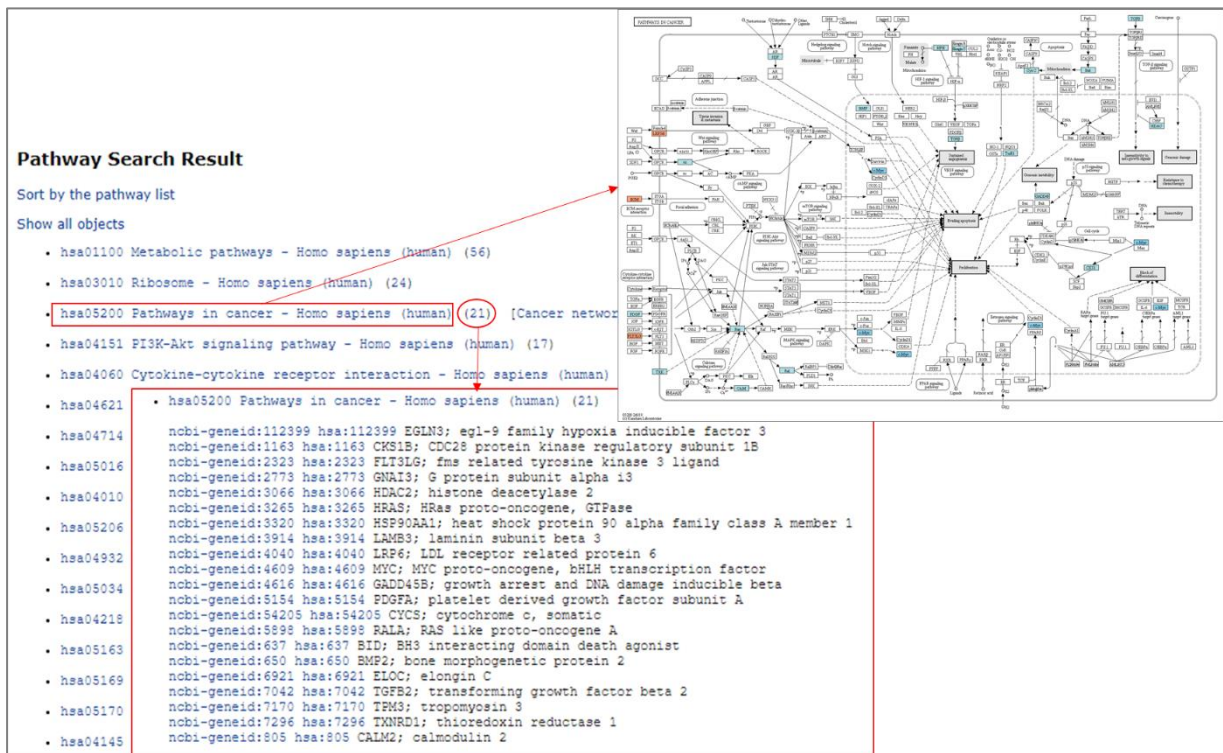


그림 5-4. KEGG Mapper tool analysis result

4. Gene set enrichment analysis (GSEA)

Gene set enrichment analysis (GSEA)는 Microarray 또는 DNA-Seq data 를 넣어 대조군, 실험군에서 유의한 gene set 을 분석하는 프로그램이다. GSEA 는 human, mouse, rat 만 분석 가능하다. MSigDB 에 있는 gene set (GO, pathway 등)을 기반으로 분석한다.

분석 과정은 그림 6-1 과 같다.

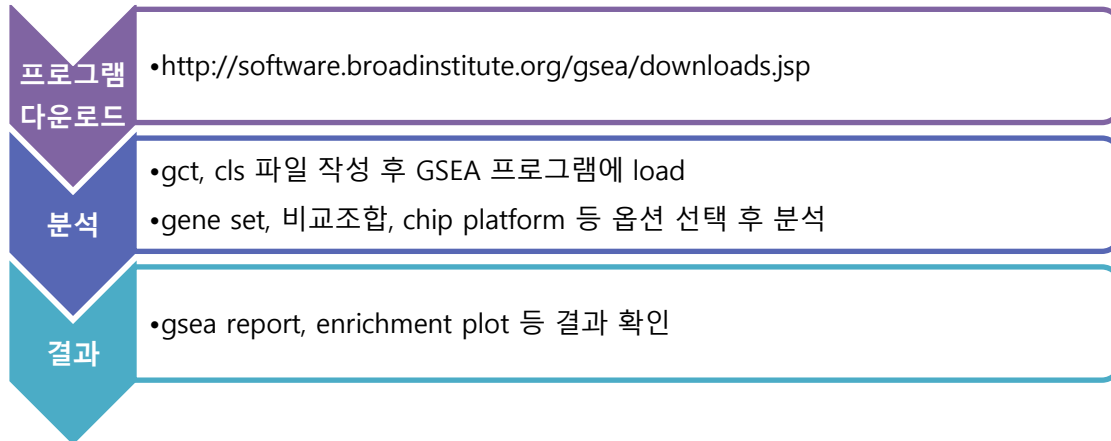


그림 6-1. GSEA tool analysis process

GSEA 홈페이지(<http://software.broadinstitute.org/gsea/downloads.jsp>)에 들어가 회원가입 후 로그인 하여 GSEA 프로그램을 다운로드 받는다 (그림 6-2).

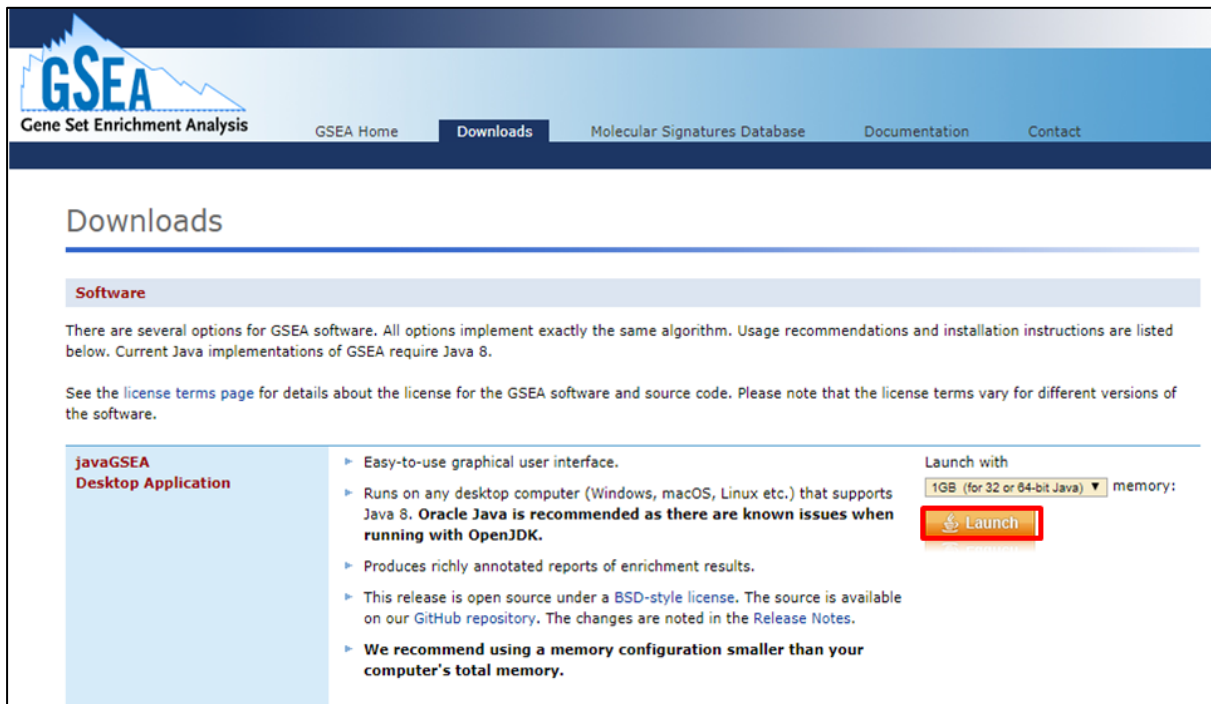


그림 6-2. GSEA program download

GSEA 분석을 위해서는 유전자 발현값 정보가 포함되어 있는 gct 파일과 샘플 정보가 포함되어 있는 cls 파일이 필요하다. Gct 파일은 그림 6-3과 같은 형식으로 만든다. A1칸에는 항상 "#1.2", A2칸에 유전자 수, B2칸에 샘플 수를 기입한다. DNA-Seq data는 A열에 Nearest Gene ID, B열에 Gene title (gene description), C열부터 각 샘플의 normalized data (log2 변환하지 않은 값)을 기입한다. Microarray data는 A열에 probe ID, B열에 Nearest Gene ID, C열부터 각 샘플의 normalized data (log2 변환하지 않은 값)을 기입한다. 파일 저장할 때는 파일명 뒤에 ".gct"를 붙이고 파일 형식은 "텍스트 (탭으로 분리) 파일"로 저장한다.

	A	B	C	D	E	F	G	H	I	J	K
1	#1.2										
2	24424	9									
3	Gene sym	Gene Title	A1	A2	A3	B1	B2	B3	C1	C2	C3
4	A1BG	NA	1.544002	1.369196	1.864898	1.630973	1.149183	1.289642	0.842837	0.742837	0.942837
5	A1BG-AS1	NA	0.629423	0.166494	0.220651	0.514093	0.562111	0.158501	0.943667	0.843667	1.043667
6	A1CF	NA	4.39E-05	5.74E-05	0.050181	1.67E-05	2.14E-05	0	0.1	0	0.2
7	A2M	NA	0.230622	0.330027	0.100307	0.641994	0.090967	0.071063	0.328091	0.228091	0.428091
8	A2M-AS1	NA	0.303073	0.798804	0.111377	1.771126	0.91748	0.29448	0.922872	0.822872	1.022872

파일 이름(N): gsea_input.gct
 파일 형식(T): 텍스트 (탭으로 분리)

그림 6-3. gct file

cls 파일은 그림 6-4와 같은 형식으로 만든다. A1칸에는 "샘플수(띄어쓰기)그룹수(띄어쓰기)1", A2칸에는 "#그룹이름", A3칸에는 gct파일에 기입한 샘플의 순서대로 각 샘플이 어떤 그룹에 속하는지 그룹이름을 기입한다. A2, A3칸에서 띄어쓰기로 그룹을 구분한다. 파일 저장할 때는 파일명 뒤에 ".cls"를 붙이고 파일형식은 "텍스트 (탭으로 분리) 파일"로 저장한다.

	A	B	C
1	9 3 1		
2	#A B C		
3	A A A B B B C C C		
4			

파일 이름(N): gsea_input.cls
 파일 형식(T): 텍스트 (탭으로 분리)

그림 6-4. cls file

GSEA 프로그램을 열어 Load data 버튼을 누르고 Browse for files 버튼을 누른 후 gct, cls 파일을 연다(그림 6-5). gct, cls 파일은 파일의 경로가 길면 input 파일을 잘 인식하지 못하여 되도록 바탕화면에 두고 수행한다.

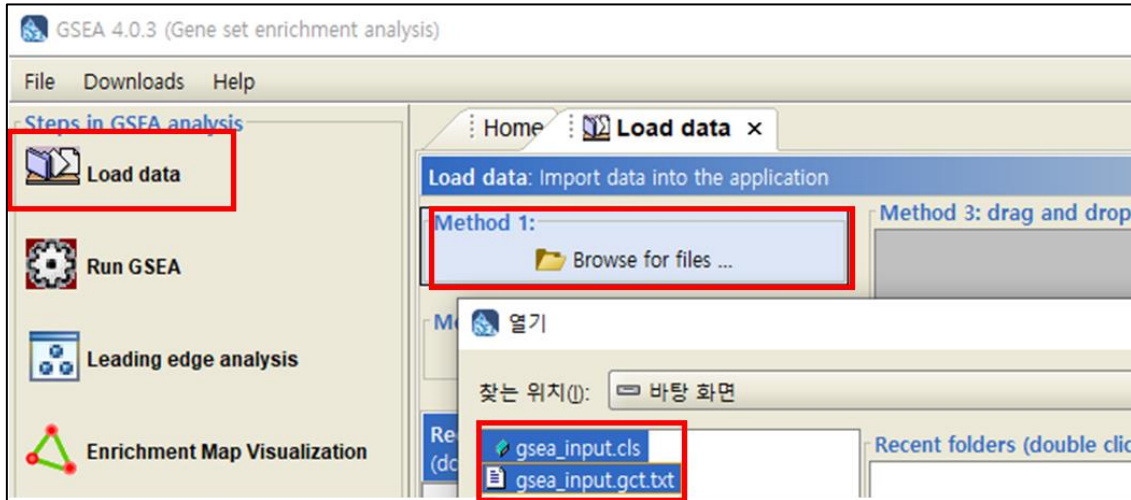


그림 6-5. Load data in GSEA program

Run GSEA 를 누르고 Expression dataset 는 gct 파일명을 선택, gene sets database 는 분석하고자 하는 gene set 을 선택한다(그림 6-6). pathway 분석을 하고자 하면 c2 를 선택, gene ontology 분석을 하고자 하면 c5 를 선택한다. Gene set 에 대한 자세한 설명은 GSEA 홈페이지 (<http://software.broadinstitute.org/gsea/msigdb/collections.jsp>)에 있다. Number of permutations 은 1000 으로 기입하고, Phenotype labels 은 분석하고자 하는 비교조합(test versus control)을 선택한다. Collapse/remap to Nearest Gene IDs 은 Collapse 을 선택하고, permutation type 은 gene_set 을 선택한다. Chip platform 은 DNA-Seq 의 경우엔 Human(or Mouse or Rat)_Symbol_with_Remapping_MSigDB.v7.0.chip 을 선택한다. Microarray 의 경우엔 실험한 chip 을 선택한다. Run 버튼을 누르면 분석이 시작된다.

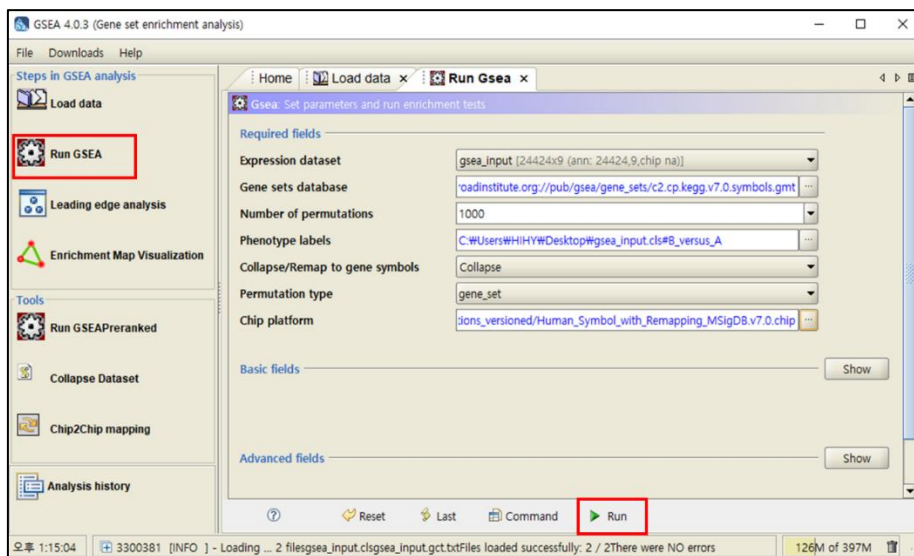


그림 6-6. Run GSEA

분석이 완료되면 GSEA 왼쪽 아래 GSEA reports 창에 status가 Success로 바뀐다. Show results folder를 누르면 GSEA 분석 결과 창이 열린다(그림 6-7).

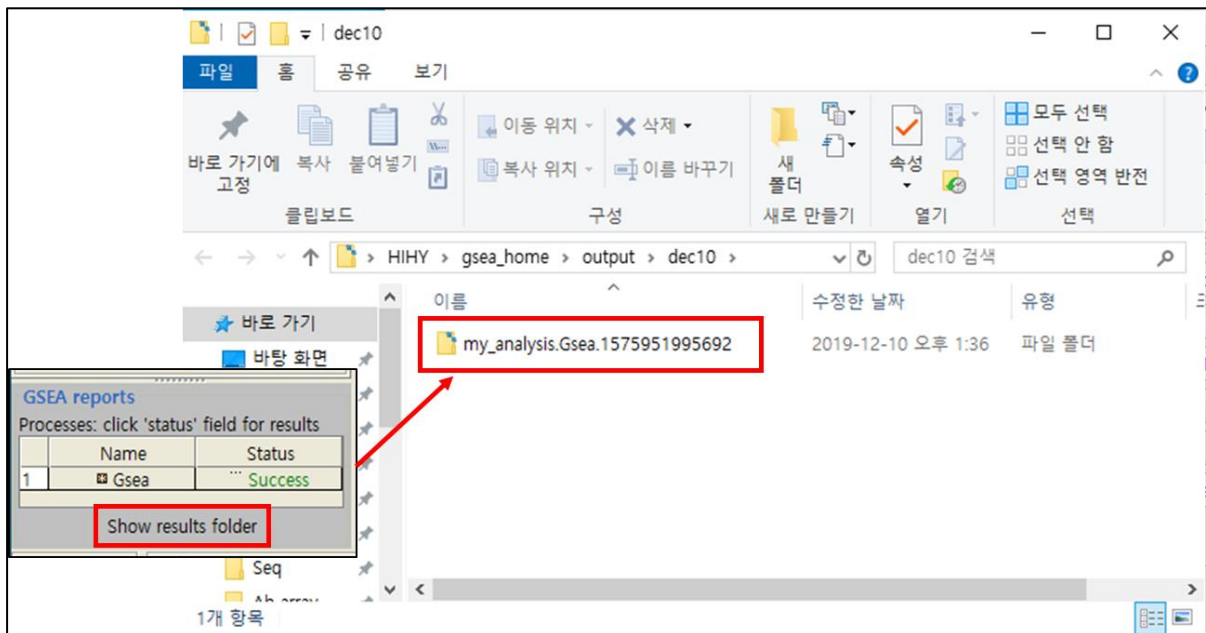


그림 6-7. GSEA results folder

GSEA 결과 중 중요 파일은 'gsea_report_for'로 시작하는 엑셀 파일이다. _for 대조군 파일은 대조군에서 유의한 gene set, _for 실험군 파일은 실험군에서 유의한 gene set이다(그림 6-8).

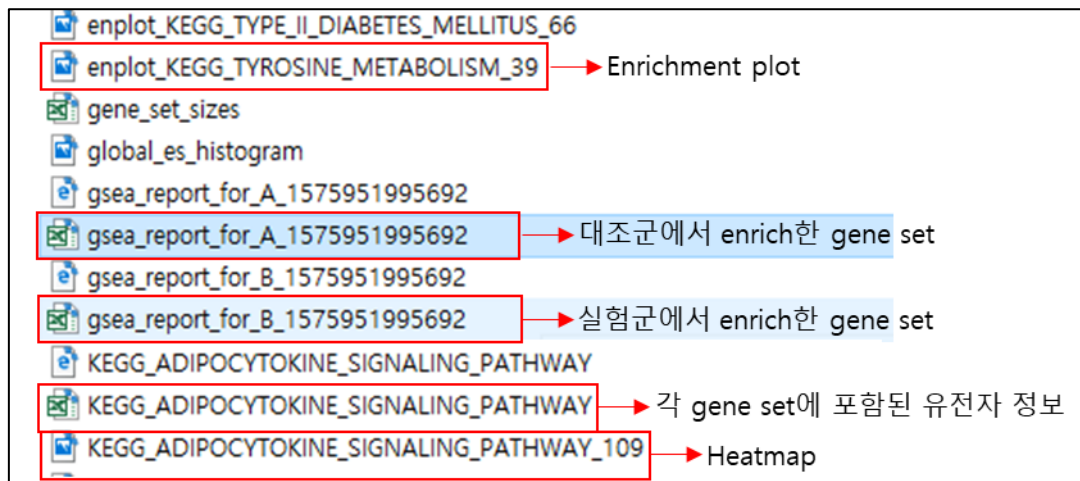
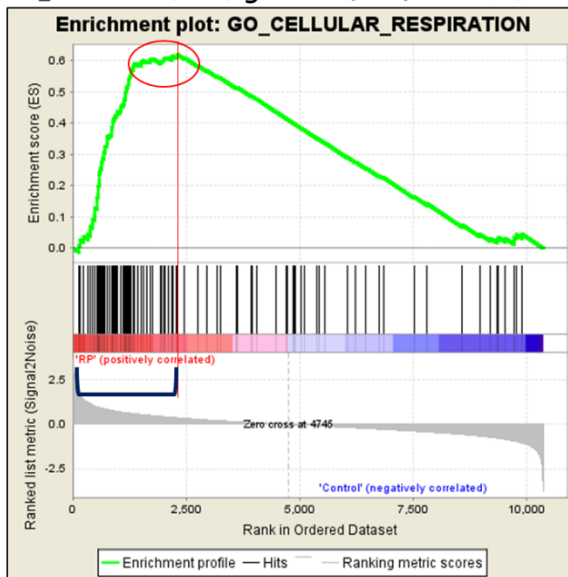


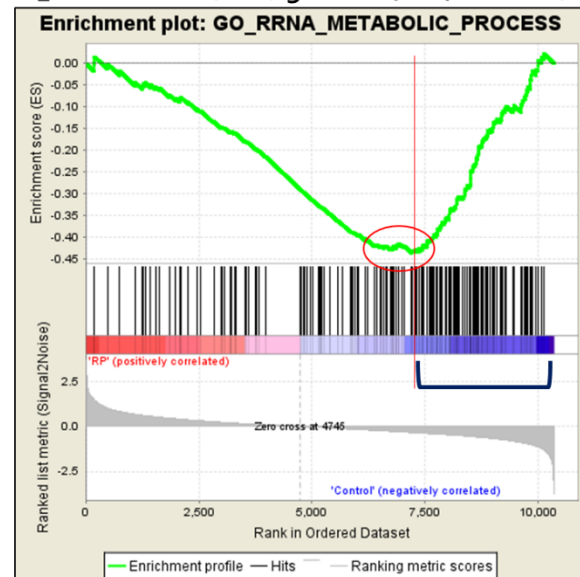
그림 6-8. GSEA result files

_for 대조군 파일에는 enrichment score (ES)와 Normalized enrichment score (NES)가 음수, _for 실험군 파일에는 ES와 NES는 양수다. 음수 양수와 관계없이 NES의 절대값이 큰 순서로 ranking 되어 있다. 음수는 DOWN (ranking 하위)에서 core gene의 밀집도가 있다는 것을, 양수는 UP (ranking 상위)에서 core gene의 밀집도가 있다는 것을 의미한다. NES 절대값이 높을수록 유의한 gene set이다. 상위 20개 gene set은 enrichment plot, heatmap, 각 gene set에 포함된 유전자들의 정보가 담긴 excel file이 있다. GSEA 분석 결과 중 Enrichment plot이 논문에 많이 실린다. Enrichment plot 이미지에서 세로 선이 해당 gene set에 포함된 유전자들이며 fold change 순으로 나열된다(그림 6-9). Peak가 왼쪽에 생기면 대조군 대비 실험군에서 up된 유전자들이 많다는 의미이고, peak가 오른쪽에 생기면 down된 유전자가 많다는 의미이다.

▼_for test 파일내 gene set/ ES, NES 양수



▼_for control 파일내 gene set/ ES, NES 음수



core enrichment(=core gene) 영역, 관련된 유전자영역이 밀집되어 있는 곳

그림 6-9. GSEA enrichment plot

GSEA 분석과정 및 결과에 대한 카테고리의 자세한 의미는 GSEA user guide (<https://software.broadinstitute.org/gsea/doc/GSEAUserGuideFrame.html>)에서 확인할 수 있다.

5. Protein-Protein Network Analysis (Cytoscape STRING)

STRING tool 은 Protein-Protein Interaction 데이터 베이스를 기반으로 유전자의 상관관계를 통계적으로 분석하여 Network 을 작성해주는 분석 툴이다. 분석 과정은 그림 7-1 과 같다.

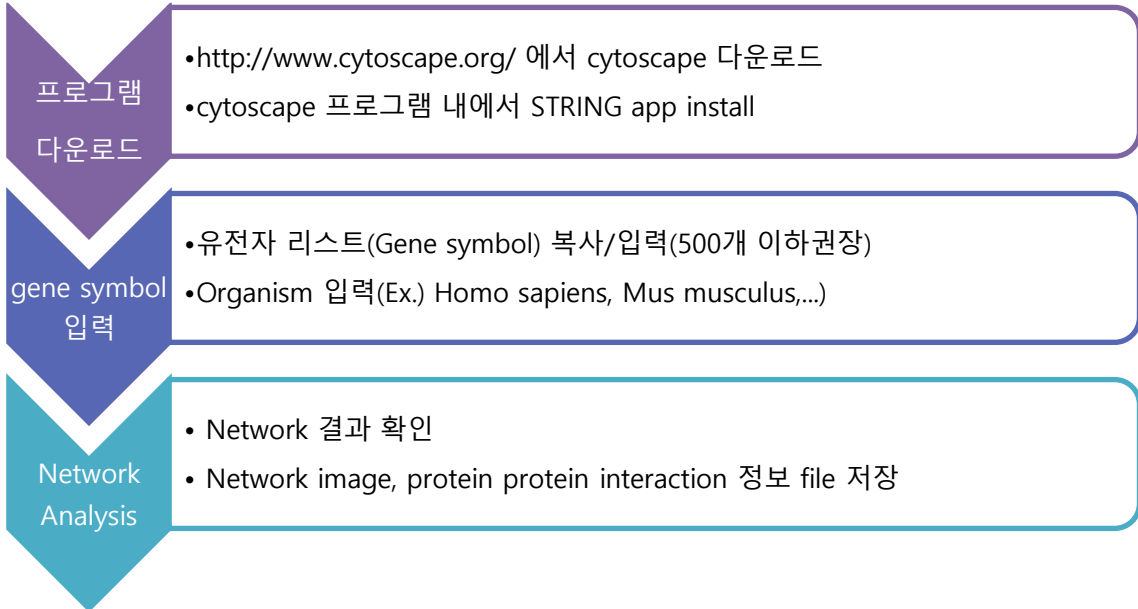


그림 7-1. STRING analysis process

Cytoscape 홈페이지 (<http://www.cytoscape.org/>)에서 cytoscape 프로그램을 다운로드 받아 설치한다(그림 7-2).



그림 7-2. Cytoscape download

Cytoscape 프로그램을 열어 상위에 있는 메뉴 중 [Apps] > [App Manager]로 들어간다(그림 7-3). StringApp 을 선택 후 Install 버튼을 누른다.

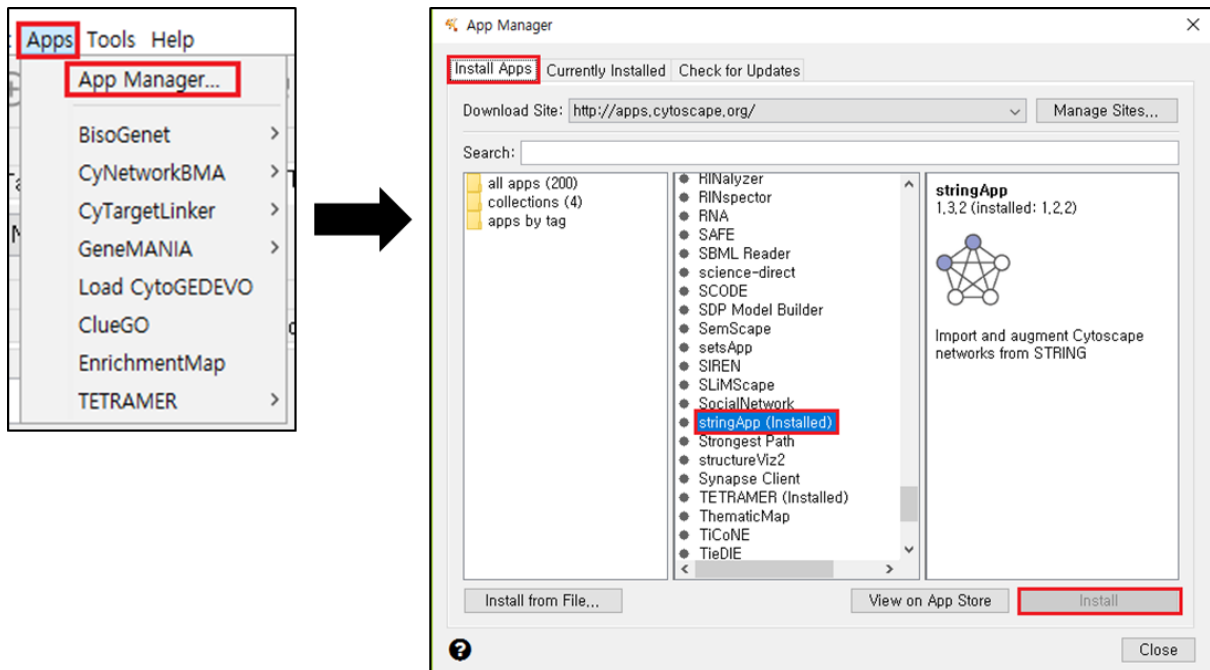


그림 7-3. STRING app installation in Cytoscape

Cytoscape 상위 메뉴 중 [File] > [Import] > [Network from Public Databases]로 들어간다(그림 7-4).

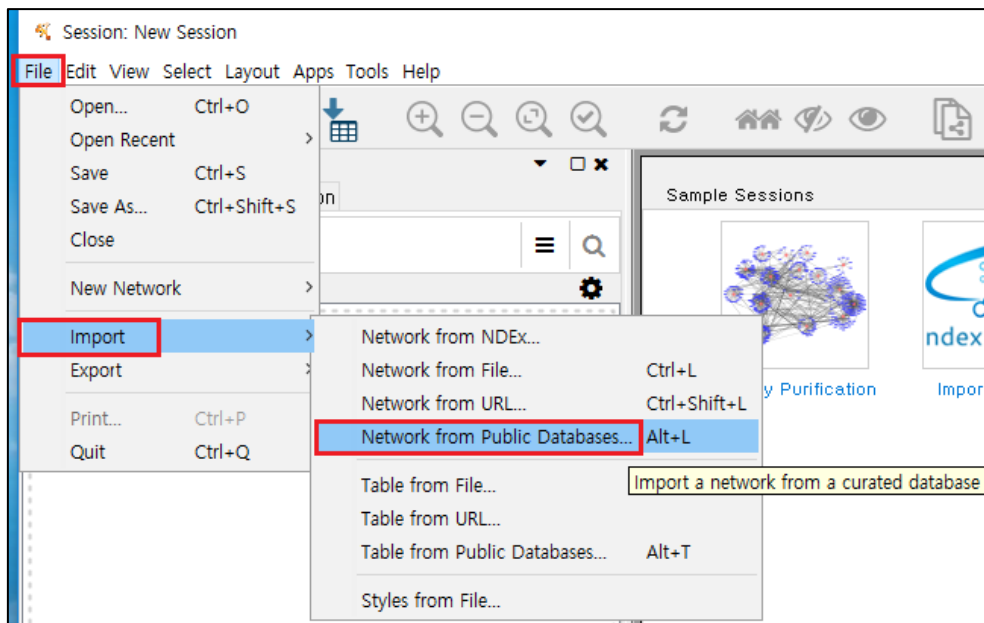


그림 7-4. STRING analysis process 1

Data Source 를 "STRING : protein query" 선택하고 Species 를 선택한다(그림 7-5). 분석하고자 하는 유전자들의 Nearest Gene ID 을 입력한다. Confidence (score)는 Protein-Protein Interaction 강도를 뜻하는 것으로 0 부터 1 까지 이고, 1 로 갈 수록 Interaction 이 강함을 의미한다. Maximum additional interactors 를 0으로 하면 input 한 유전자 안에서만 network 이 그려지고 숫자를 높이면 input 하지 않은 neighborhood protein 까지 network 이 그려진다.

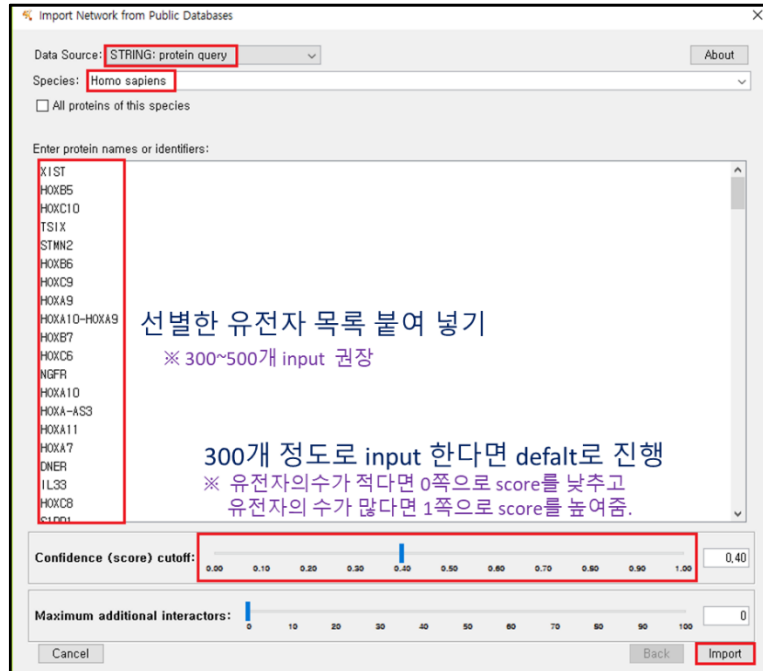


그림 7-5. STRING analysis process 2

Input 한 Nearest Gene ID 과 match 가 되지 않는 protein 이 있으면 그림 7-6 과 같은 화면이 나온다. 두 개 이상의 protein 이름이 나타나는 경우는 유사 protein 을 확인하라고 한다. 연구자의 선택에 따라 모두 check 또는 해지한다. Import 버튼을 누르면 분석이 진행된다.

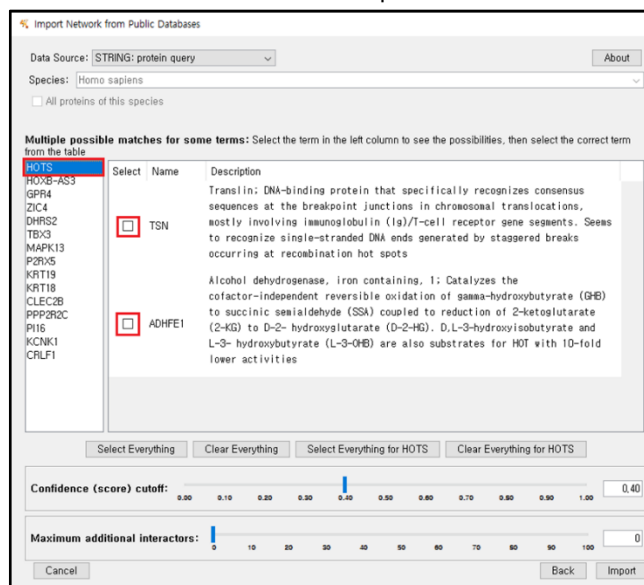


그림 7-6. Not matched proteins in STRING

분석이 완료되면 network image 가 나온다(그림 7-7).

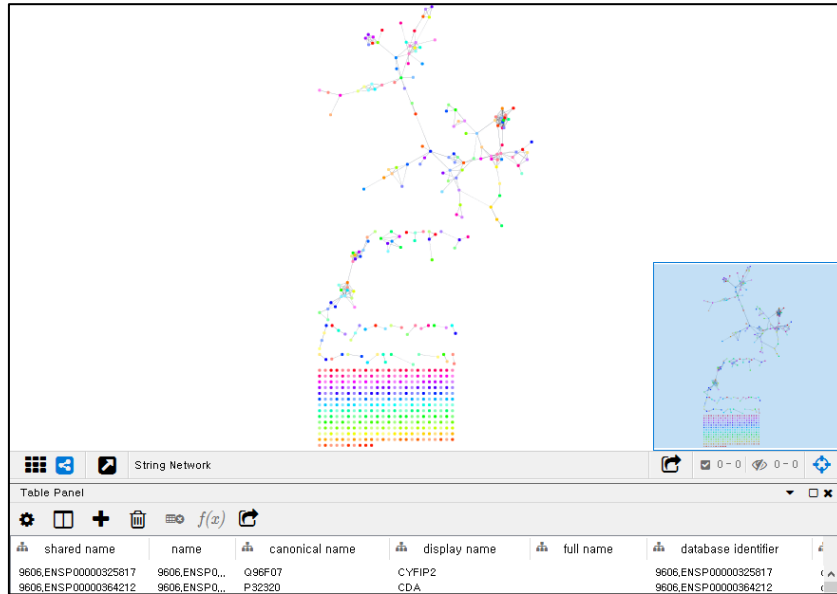


그림 7-7. Network result

[File] > [Export] > [Network to Image]를 눌러 이미지를 저장한다(그림 7-8). PDF 파일형식으로 저장하는 것을 권장한다. Pdf 파일로 저장하면 확대를 하여도 이미지가 깨지지 않는다.

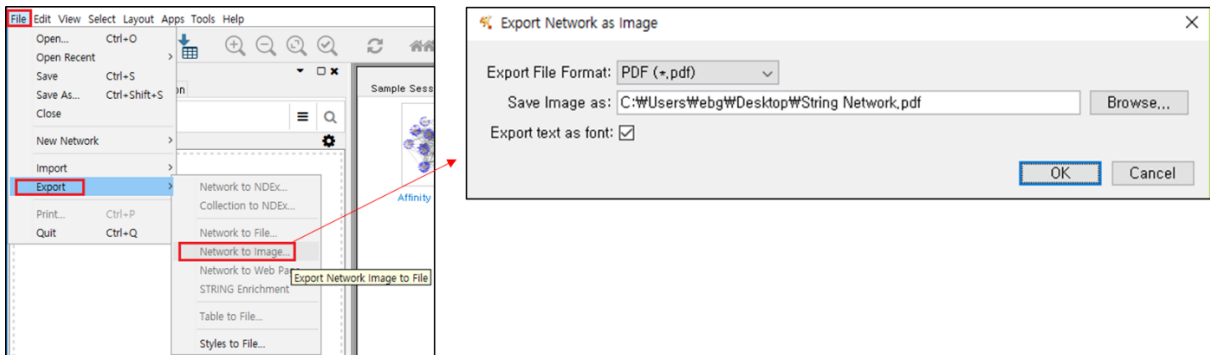


그림 7-8. Save network image

어떤 유전자들이 protein-protein interaction 을 하는지 정보를 저장하고 싶으면 [File] > [Export] > [Table to File...]로 들어가 String Network default edge 파일을 저장한다(그림 7-9).

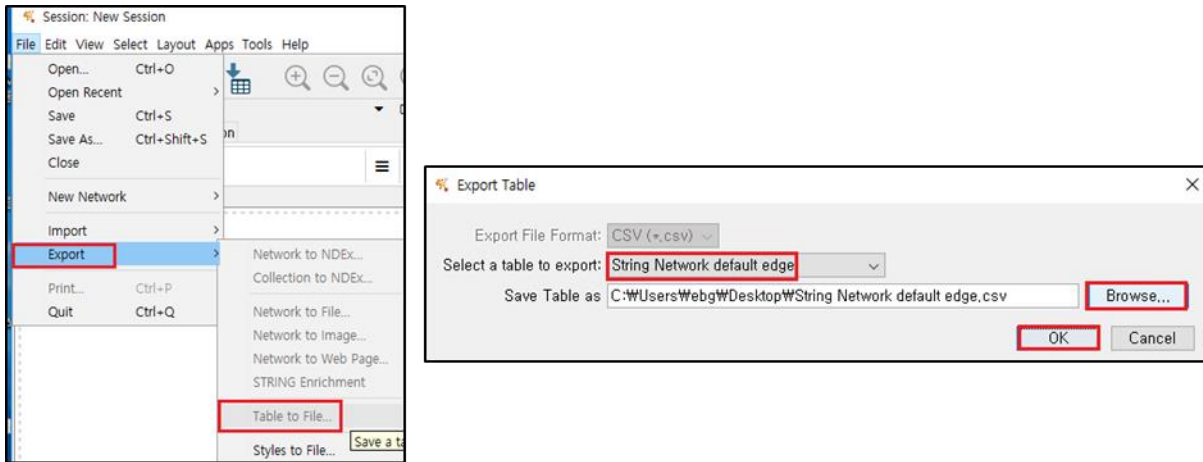


그림 7-9. Save edge table

String Network default edge 파일에서 name 에 interaction 정보, score 에 confidence score 가 나와있다(그림 7-10). Name 에 A (pp) B 라고 적혀있으면 A 유전자와 B 유전자가 Protein-Protein Interaction 한다는 것이고 score 값이 1 에 가까울수록 interaction 이 강한 것이다.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	SUID	coexpri	coocui	databa	experir	fusion	interac	intersp	name	neighb	score	selecte	shared	shared	textmit
2	701	0.397		0.9	0.961		pp		EFNB2 (pp) EPHA4		1.000	FALSE	pp	EFNB2 (pp)	0.926
3	702	0.397		0.9	0.817		pp		EFNB2 (pp) EPHB1		0.999	FALSE	pp	EFNB2 (pp)	0.887
4	961	0.929		0.9			pp		IFIT1 (pp) MX1	0.064	0.998	FALSE	pp	IFIT1 (pp)	0.722
5	700	0.397		0.9	0.76		pp		EFNB2 (pp) EPHA5		0.997	FALSE	pp	EFNB2 (pp)	0.768
6	680	0.878		0.9	0.292		pp		OAS3 (pp) IFIT1		0.996	FALSE	pp	OAS3 (pp)	0.519
7	683	0.888		0.9	0.132		pp		OAS3 (pp) MX1		0.995	FALSE	pp	OAS3 (pp)	0.506
8	935	0.892		0.9			pp		IFI6 (pp) MX1		0.995	FALSE	pp	IFI6 (pp) N	0.503

그림 7-10. Interaction information in edge table

Network image 에서 색이나 모양을 변경하고 싶은 경우에는 STRING Manual ([Download link](#))에서 image 수정 방법을 확인할 수 있다.