

User Manual

ExSCEA v1.0.0



< 목 차 >

1. ExSCEA 개요	3
2. Category 사용 방법	5
2.1. HVG	5
2.2. Cell Type	5
2.3. GO	6
3. DE Analysis 사용 방법	8
3.1. Significant Selection	8
3.2. Analysis Graph	18
3.2.1. Volcano Plot	26
3.2.2. Venn Diagram	26
3.3. Third Part Support	18
3.4. Target Gene Plot View & Gene Search	18

1. ExSCEA 개요

(주)이바이오젠은 Single cell RNA-seq data 를 엑셀 기반에서 쉽게 분석할 수 있도록 분석 결과 보고 시 ExSCEA (Excel-based Single Cell Expression Analysis) tool 을 함께 제공한다. ExSCEA tool 은 연구자들이 Single cell RNA-seq 데이터를 보다 쉽게 다루고 원하는 데이터를 쉽게 얻을 수 있도록 사용자 편의를 최대한 반영한 프로그램이며, 엑셀 프로그램 안에서 다양한 분석을 직관적으로 수행할 수 있도록 개발되었다. ExSCEA tool 은 사용자들의 요구사항을 반영하여 데이터 분석과 엑셀 사용에 익숙하지 못한 연구자들도 쉽게 사용이 가능하도록 지속적으로 업데이트 될 예정이다.

(주)이바이오젠에서 제공하는 Single cell RNA-seq (엑셀 데이터)를 열기 전에 함께 제공한 ExSCEA_v1.0.0_Installer.zip 파일의 압축을 풀고 setup.exe 를 실행하면 ExSCEA tool 이 설치된다 (그림 1-1). 설치가 완료되고 ExSCEA format 의 엑셀 데이터를 열면 자동으로 프로그램이 구동된다.



그림 1-1. ExSCEA 설치

ExSCEA 형태의 엑셀 파일은 왼편에 Category 창 (파란색 상자)과 가운데 Expression data (빨간색 상자), 오른편에 DE Analysis 창 (초록색 상자)으로 구성된다. (그림 1-2).

그림 1-2. ExSCEA format

Expression data 는 ID, Gene symbol, Feature ID, Fold change, P-value, Average of Normalized data (log 2)로 구성된다. 각 항목에 대한 설명은 표 1 과 같다.

표 1. Expression data 구성 항목

항목	설명
ID	ExSCEA 내의 유전자 고유 식별 번호
Gene symbol	유전자 이름
Feature ID	Ensembl 데이터베이스 식별자
Fold change	전체 Cluster 대비 특정 Cluster의 비율
P-value	유전자에 대한 Cluster 발현 수준의 유의성
Average of Normalized data (log 2)	해당 Cluster에 대한 유전자의 평균 발현 수준

Category 부분에는 기본 설정된 HVG, Cell Type, GO 가 있고 사용자가 원하는 대로 항목을 구성하여 분석할 수 있다. 자세한 내용은 3. Category 사용 방법에서 설명한다. DE Analysis 창에서는 Fold change, p-value 를 이용하여 군집 간의 발현 차이가 나타나는 유전자를 쉽게 선별할 수 있고 Category 항목 별로 그래프를 작성할 수 있다. 뿐만 아니라, DE Analysis 창에서 Volcano Plot, Venn Diagram 을 직접 그릴 수 있고 선별된 유전자들을 대상으로 Clustering Heatmap 과 KEGG, DAVID 분석을 수행하기 위한 입력 데이터를 자동으로 만들 수 있다. Target Gene Plot View, Gene Search 기능도 이용할 수 있어 연구자가 Single cell RNA-seq data 를 쉽게 활용할 수 있다. 자세한 내용은 4. DE Analysis 사용 방법에서 설명한다.

2. Category 사용 방법

Category 부분에서는 HVG, Cell Type, GO Category 를 이용하여 유전자를 필터링하고 해당 유전자의 데이터 값을 확인한다. 해당 기능을 사용하는 도중 View All Data 를 클릭하면 필터링이 모두 해제되어 전체 결과로 되돌릴 수 있다.

2.1. HVG

HVG 는 Highly Variable Genes 의 약자로, 자신을 제외한 모든 군집과 비교하여 해당 군집에서 유의한 유전자를 필터링 하고 데이터 값을 확인할 수 있는 기능이다. ExSCEA tool 에서는 기본적으로 각 군집에서 P-value 가 0.05 보다 같거나 작은 유전자들이 지정되어 있다. HVG (P-Value) 영역에서 각 군집을 클릭하면 해당 군집에서 유의한 유전자들만 필터링 된다. HVG 에서 여러 군집을 선택할 수 있으며, AND 와 OR 기능을 선택하면 여러 군집에서 동시에 만족하는 유전자 또는 적어도 하나의 군집에 포함된 유전자를 확인할 수 있다.

2.2. Cell Type

Cell Type 은 알려진 세포 타입과 이들의 마커 유전자를 이용하여 군집에 포함된 유전자를 필터링하고 데이터 값을 확인할 수 있는 기능이다 (그림 2-1). ExSCEA tool 에서는 Single cell RNA-seq 에서 자주 사용되는 세포 타입이 지정되어 있다. Cell Type 영역에서 원하는 세포 타입을 선택하면 해당 세포 타입을 구성하는 마커 유전자들만 필터링 된다. Cell Type 에서 여러 세포 타입을 선택할 수 있으며, AND 와 OR 기능을 선택하면 여러 세포 타입에서 동시에 만족하는 유전자 또는 적어도 하나의 세포 타입에 포함된 유전자를 확인할 수 있다. 하단의 Cell Type Setting 을 클릭하면 세포 타입 리스트를 추가, 제거, 수정할 수 있다. 세포 타입은 수동으로 지정할 수 있지만, PanglaoDB 를 이용하여 추가할 수 있다. PanglaoDB 는 현재까지 알려진 세포 타입과 이들의 마커 유전자를 제공하는 데이터베이스이다. 연구자가 원하는 세포 타입을 검색하여 유전자를 필터링 할 수 있다.

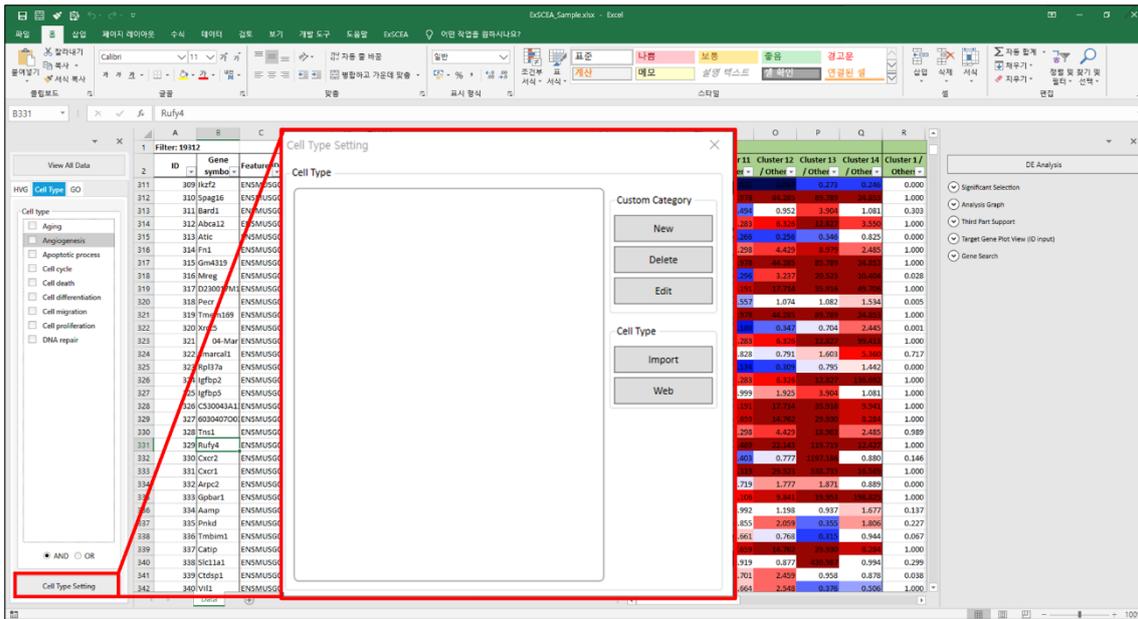


그림 2-1. Cell Type 기능

2.3. GO

GO 는 알려진 Gene Ontology 와 이를 구성하는 유전자를 이용하여 군집에 포함된 유전자를 필터링하고 데이터 값을 확인할 수 있는 기능이다 (그림 2-2). ExSCEA tool 에서는 자주 사용되는 GO 카테고리가 지정되어 있다. GO Category 영역에서 원하는 GO 를 선택하면 해당 GO 를 구성하는 유전자들만 필터링 된다. GO Category 에서 여러 GO 를 선택할 수 있으며, AND 와 OR 기능을 선택하면 여러 GO 에서 동시에 만족하는 유전자 또는 적어도 하나의 GO 에 포함된 유전자를 확인할 수 있다. 하단의 Gene Category Setting 을 클릭하면 GO 리스트를 추가, 제거, 수정할 수 있다. 또한 QuickGO 를 이용하는 새로운 GO 를 추가할 수 있다. QuickGO 를 이용하여 GO 를 추가하는 방법은 물음표 버튼을 클릭하여 확인할 수 있다. PCR-Array 항목의 View list/Import 를 이용하여 Pathway 별 유전자 목록을 추가할 수 있다. 새 창에서 Species (Human, Mouse, Rat)를 선택하고 Keyword 에 추가하고자 하는 Pathway 이름이나 유전자 이름을 검색하고 추가하려는 항목에 Check 를 클릭한 후 Import 버튼을 누르면 자동으로 추가된다 (그림 2-3).

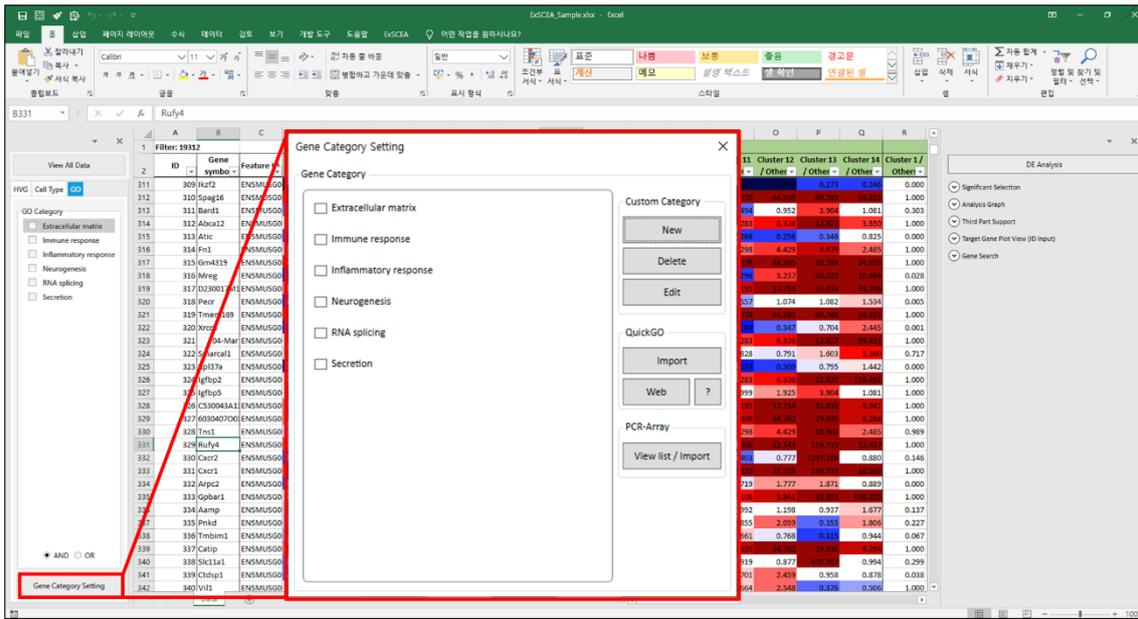


그림 2-2. GO Category 기능

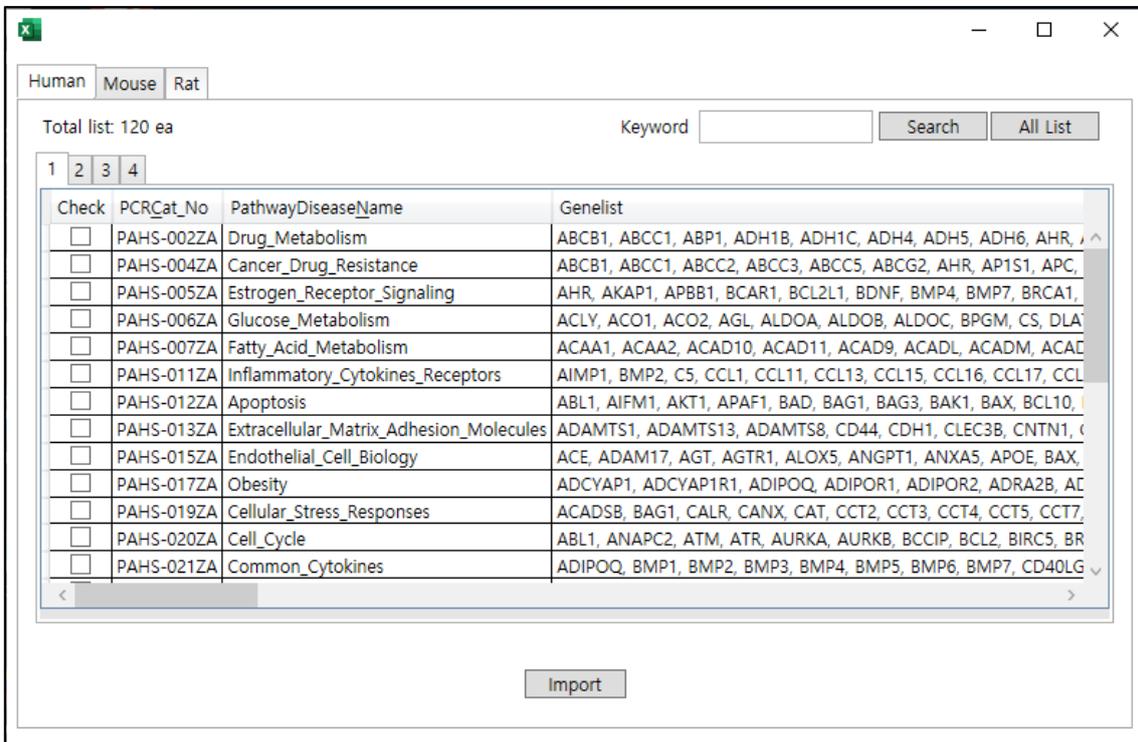


그림 2-3. PCR-Array Pathway

3. DE Analysis 사용 방법

3.1. Significant Selection

DE Analysis 부분에서 Significant Selection 창은 전체 결과 중 전체 군집 대비 특정 군집에서 유의한 유전자를 필터링 할 수 있는 기능이다. 예를 들어, Cluster 1/Others 비교조건을 선택하고 Fold change 를 2, p-value 를 0.05 로 선택하면 Cluster 1 을 제외한 나머지 군집 대비 Cluster 1 에서 2 배 이상 발현이 증가 또는 감소하면서 p-value 값이 0.05 이하인 유전자가 필터링 된다 (그림 3-1).

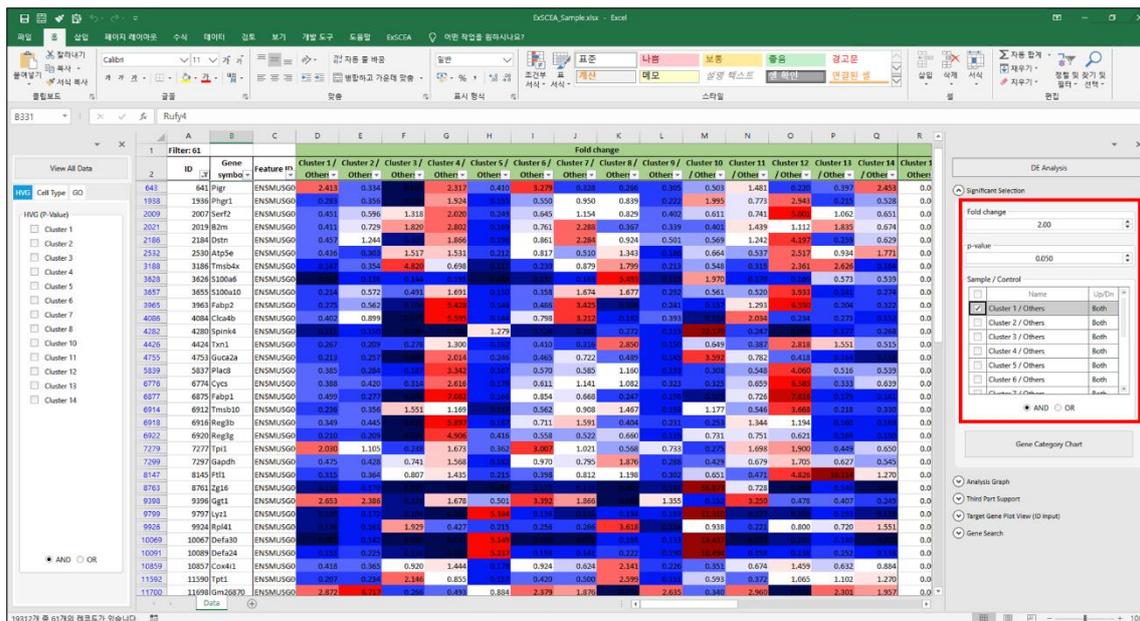


그림 3-1. Significant Selection

Significant Selection 에서 증가 또는 감소한 유전자를 각각 보고 싶다면 Up/Dn 항목에서 선택할 수 있다. Both 는 증가 또는 감소된 모든 유전자를 필터링, Up 은 증가된 유전자를 필터링, Dn 은 감소된 유전자를 필터링한다 (그림 3-2).

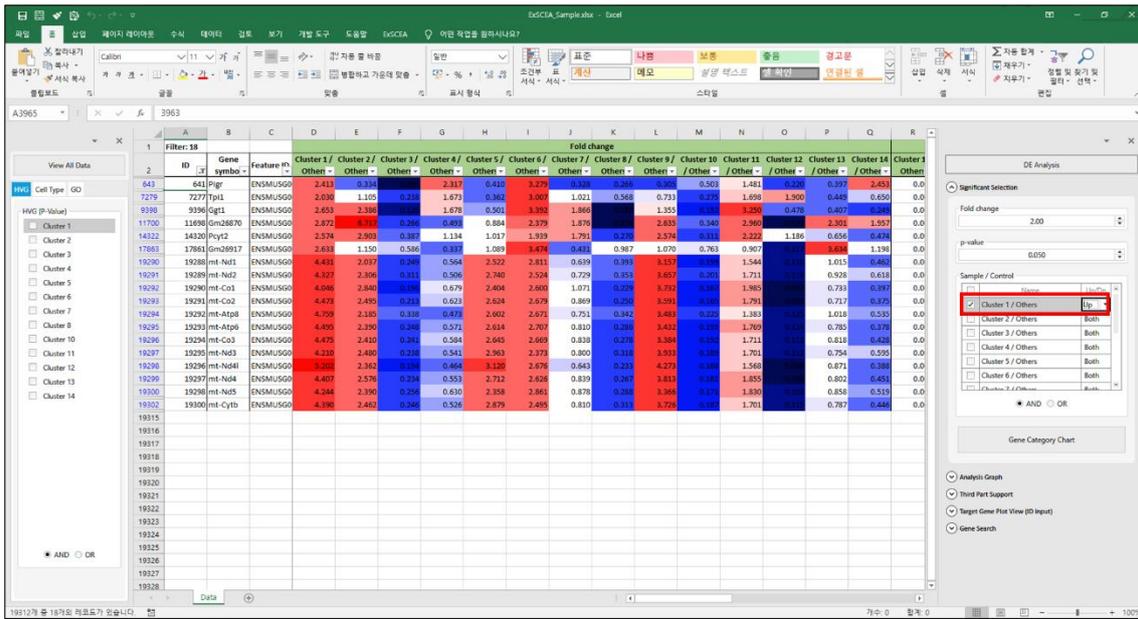


그림 3-2. Significant Selection (Up/Dn 기능)

Significant Selection 부분은 Category 창의 HVG, Cell Type, GO 와 연동이 가능하다. 그림 3-3 과 같이 Significant Selection 에서 조건을 입력하고 Category 창에서 원하는 항목을 선택하면 유전자가 추가적으로 필터링 된다.

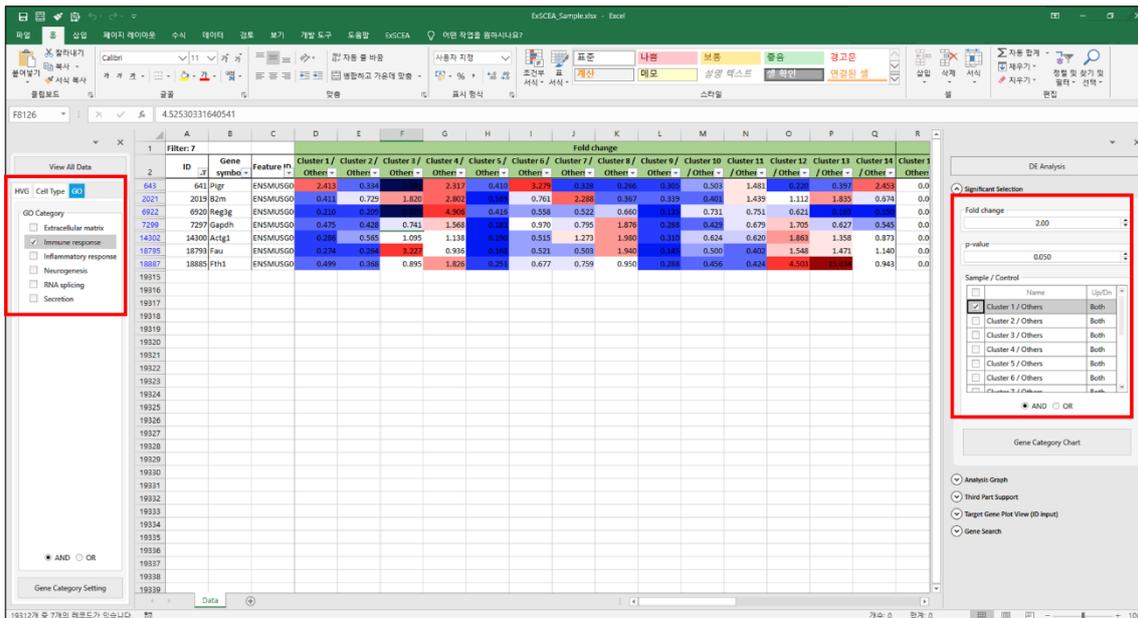


그림 3-3. Significant Selection (Category 창과 연동)

Gene Category Chart 는 HVG, Cell Type, GO 와 관련된 유전자 중 유의한 유전자의 비율과 수를 나타낸 그래프이다 (그림 3-4). 전체 데이터 상태에서 Significant Selection 의 비교 그룹을 선택하고 Gene Category Chart 를 클릭하면 증가/감소한 유전자를 대상으로 HVG, Cell Type, GO 에 대한 그래프가 생성된다. Pie chart 의 특정 영역을 클릭하면 해당 HVG, Cell Type, GO 에 대하여

증가/감소된 유전자가 필터링 된다. Bar plot 에서 상단의 숫자는 해당 유전자 수이며, 해당 영역을 클릭하면 유전자가 필터링 된다 (그림 3-5).

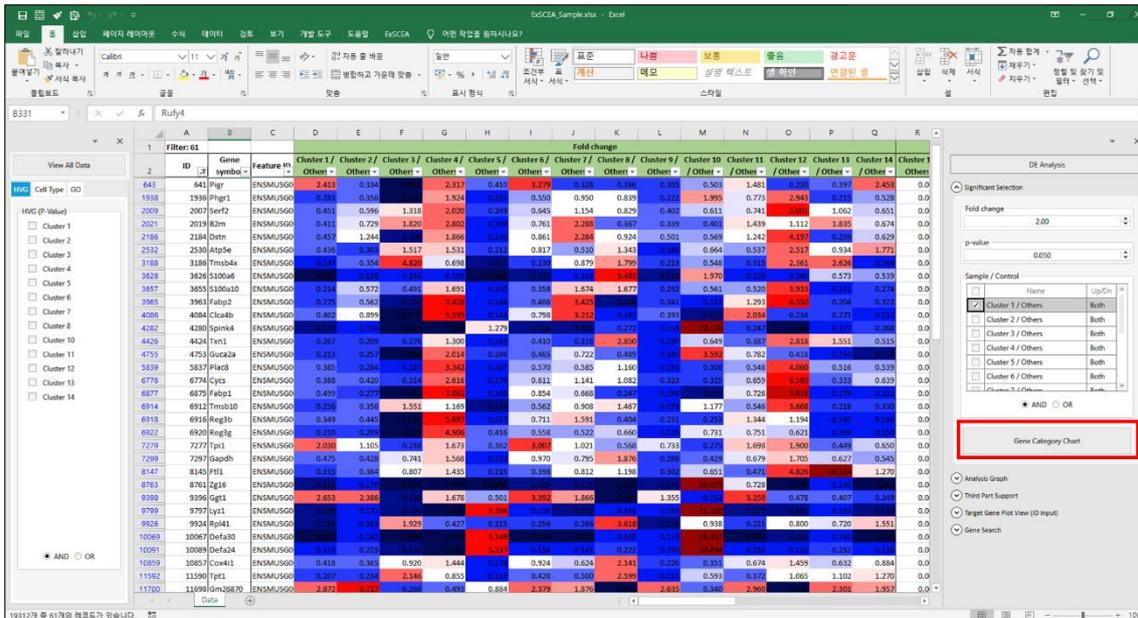


그림 3-4. Gene Category Chart

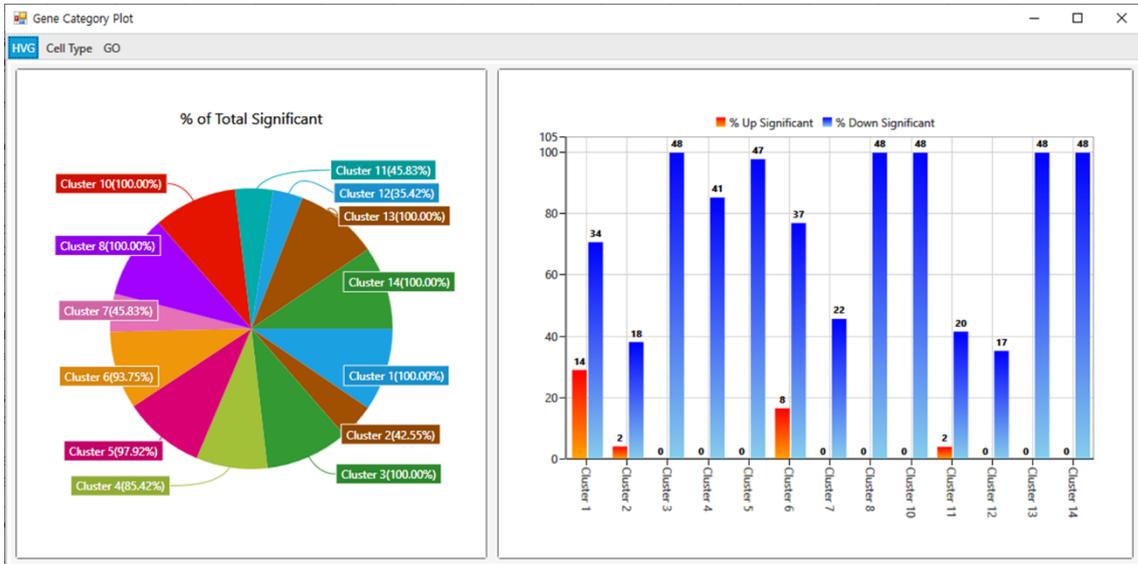


그림 3-5. Gene Category Chart (HVG 그래프)

3.2. Analysis Graph

DE Analysis 에서 Analysis Graph 를 클릭하면 Volcano Plot 과 Venn Diagram 이미지를 생성할 수 있다 (그림 3-6).

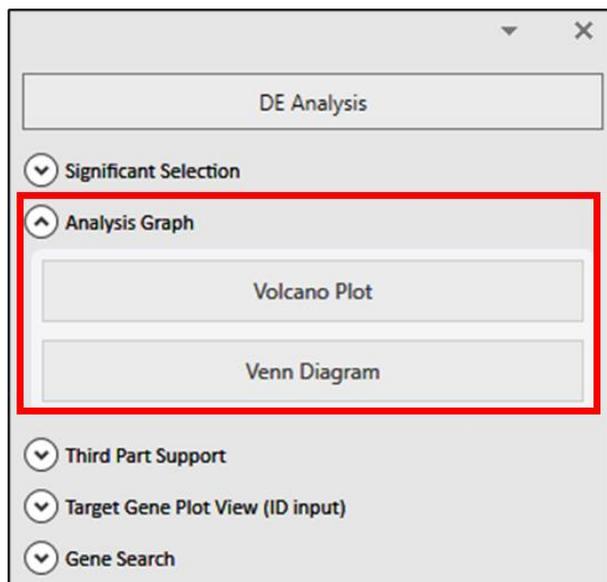


그림 3-6. Analysis Graph

3.2.1. Volcano Plot

Volcano Plot 은 Fold change 와 P-value 를 기반으로 생성되는 이미지이다. Volcano Plot 창에서 비교 조합과 Fold Threshold Line, P-value 를 선택하고 Graph View 버튼을 클릭하면 이미지가 생성된다. 초록색 세로선 왼편은 2 Fold 이상 감소한 유전자, 빨간색 세로선 오른편은 2 Fold 이상 증가한 유전자, 검은색 가로선 위편은 p-value 가 0.05 이하인 유전자이다. 그림 3-7 은 Cluster 1 과 나머지 군집에 대한 비교 조합에 대하여 Fold Threshold Line 은 2, P-value 는 0.05 로 지정한 Volcano Plot 이다.

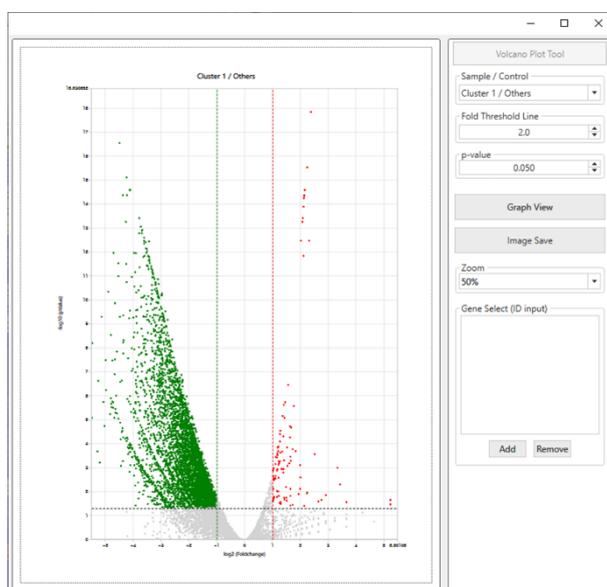


그림 3-7. Volcano Plot

이미지에서 점을 클릭하면 해당 유전자 이름이 표시된다. 유전자 이름을 마우스 오른쪽 버튼을 클릭하면 해당 표시를 지울 수 있다. 여러 개의 유전자를 동시에 표현하고 싶다면 Gene Select (ID input)에 유전자 ID 를 복사한 후 Add 버튼을 클릭하면 해당 유전자들의 이름이 이미지 상에 표시된다. 빈 공간에서 마우스 오른쪽 버튼을 클릭하면 이미지 상에 표시된 모든 유전자 이름을 지울 수 있다. Image Save 버튼을 클릭하면 Volcano Plot 이미지를 저장할 수 있다.

3.2.2. Venn Diagram

Venn Diagram 은 여러 비교 조합 간의 유의한 유전자 개수를 비교하기 위한 이미지이다. Venn Diagram 을 그릴 비교 조합과 Fold change, P-value 를 선택한 후 Diagram View 버튼을 클릭하면 결과를 확인할 수 있으며, 비교 조합은 최대 4 개 그룹까지 선택할 수 있다. 그림 3-8 은 Cluster 1/Others, Cluster 3/Others, Cluster 6/Others 세 그룹에 대하여 Fold change 는 2, P-value 는 0.05 로 지정한 Venn Diagram 이다.

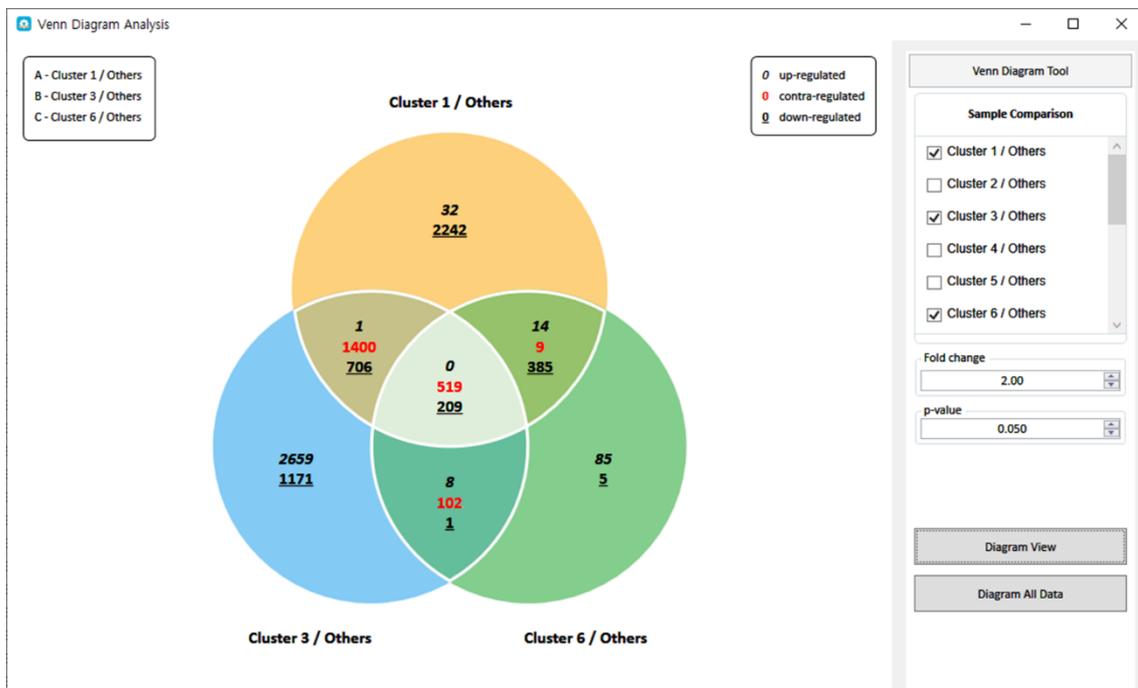


그림 3-8. Venn Diagram

그림 3-8 에서 Venn Diagram 이미지에 표시되는 글자에 대한 범례는 다음과 같다.

- **기울어진 숫자** : Up-regulated 유전자의 수
- **빨간색 숫자** : Regulation 이 대조되는 유전자의 수 (예: Cluster 1/Others 에서는 Up-regulated 이면서 Cluster 2/Others 에서는 Down-regulated 인 유전자)
- **밑줄 친 숫자** : Down-regulated 유전자의 수

Venn Diagram 이미지에서 해당 영역에 마우스 오른쪽 클릭을 하면 Up-regulated, Down-regulated, Region All 버튼을 확인할 수 있다. Up-regulated 는 해당 비교 조합에서 Up-regulated

된 유전자를 필터링, Down-regulated 는 해당 비교 조합에서 Down-regulated 된 유전자를 필터링, Region All 은 해당 비교 조합에서 검출된 Up-/Down-regulated 유전자를 필터링한다.

3.3. Third Part Support

Third Part Support 는 연구자가 선택한 유전자를 기반으로 Clustering heatmap 과 KEGG 분석, DAVID 분석을 수행하기 위한 입력 데이터를 제공한다 (그림 3-9). 입력 데이터 제작에 앞서 유전자를 선별하는 과정이 필요하다. 유전자 선별은 Category 와 DE Analysis 를 이용하여 선별할 수 있다.

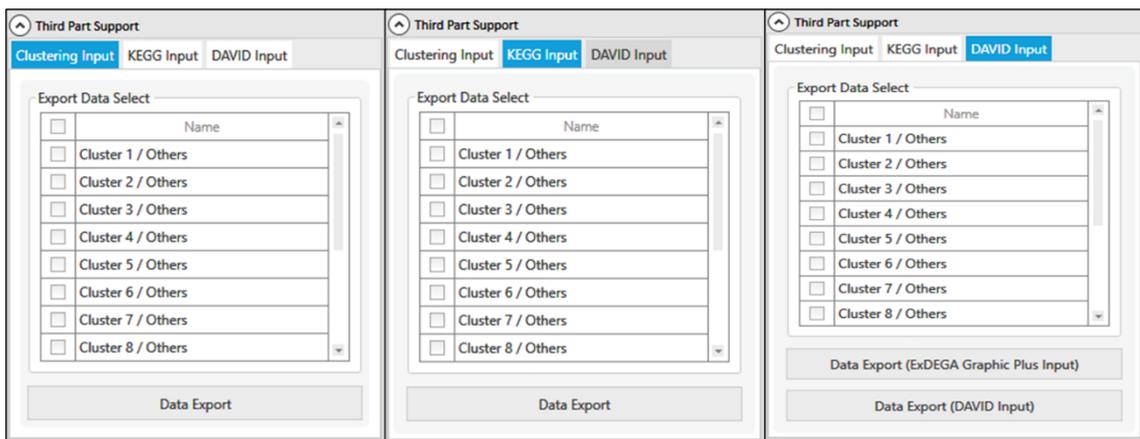


그림 3-9. Third Part Support (왼편부터 Clustering Input, KEGG Input, DAVID Input)

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 그림 3-10 과 같이 유전자 이름과 선택한 비교 조합, Z-score 로 구성된다. Clustering Input 파일은 **MeV** 또는 **ExDEGA Graphic Plus** 프로그램을 이용하여 Clustering heatmap 을 작성할 수 있다.

Gene Symbol	Cluster 1 / Others	Cluster 2 / Others	Cluster 3 / Others
Pigr	2.412656	0.3338233	0.09458916
Phgr1	0.2832479	0.3560095	0.08909768
Serf2	0.4505849	0.5958751	1.317955
B2m	0.4114005	0.7285938	1.820078
Dstn	0.4570626	1.244223	0.1060638
Atp5e	0.4355506	0.3025525	1.517178
Tmsb4x	0.1468177	0.3535689	4.820166
S100a6	0.09239034	0.1359687	0.1436486
S100a10	0.2138821	0.5718173	0.491002
Fabp2	0.274729	0.562046	0.1061139
C1ca4b	0.4021292	0.8994936	0.07165406
Spink4	0.1114207	0.1497428	0.0844293
Txn1	0.2667113	0.2091366	0.2780316
Guca2a	0.2133063	0.2566306	0.06780826
Plac8	0.385346	0.2835954	0.1872116
Cyca	0.3875297	0.4196959	0.3136992
Fabp1	0.4988641	0.277174	0.09452607
Tmsb10	0.2360803	0.3562078	1.550741
Reg3b	0.3487499	0.4446786	0.08138297
Reg3g	0.2104861	0.2088951	0.07882152
Tpi1	2.030389	1.104762	0.2382583
Gapdh	0.4746698	0.4283723	0.7407911
Fil1	0.3149678	0.3639748	0.8072144
Zg16	0.1162529	0.1698134	0.07674919

그림 3-10. Clustering Input 파일

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 3-11 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.

Gene Symbol	Cluster 1 / Others
Pigr	#FF6347,black
Phgr1	#87CEEB,black
Serf2	#87CEEB,black
B2m	#87CEEB,black
Dstn	#87CEEB,black
Atp5e	#87CEEB,black
Tmsb4x	#00BFFF,black
S100a6	#1E90FF,black
S100a10	#00BFFF,black
Fabp2	#87CEEB,black
Clca4b	#87CEEB,black
Spink4	#1E90FF,black
Txn1	#87CEEB,black
Guca2a	#00BFFF,black

그림 3-11. KEGG Input 파일

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. DAVID input 은 크게 Data Export (ExDEGA Graphic Plus Input)와 Data Export (DAVID Input)로 구성된다. Data Export (DAVID Input)는 DAVID 에서 유전자 이름을 입력하는 부분에 사용되는 파일이다. 이 파일은 유전자 이름으로 구성된다 (그림 3-12). 유전자 이름을 DAVID 의 입력 데이터로 사용하여 분석하려는 GO 또는 Pathway 에 대한 데이터를 다운로드 받는다. Data Export (ExDEGA Graphic Plus Input)는 **ExDEGA Graphic Plus** 프로그램에서 DAVID 분석을 수행하기 위하여 사용되는 파일을 구축한다. 이 파일은 유전자 이름과 비교 조합, Fold change 로 구성된다 (그림 3-13).



그림 3-12. Data Export (DAVID Input) 파일

Gene Symbol	Cluster 1 / Others	Cluster 2 / Others	Cluster 3 / Others
Pigr	2.412656	0.3338233	0.09458916
Phgr1	0.2832479	0.3560095	0.08909768
Serf2	0.4505849	0.5958751	1.317955
B2m	0.4114005	0.7285938	1.820078
Dstn	0.4570626	1.244223	0.1060638
Atp5e	0.4355506	0.3025525	1.517178
Tmsb4x	0.1468177	0.3535689	4.820166
S100a6	0.09239034	0.1359687	0.1436486
S100a10	0.2138821	0.5718173	0.491002
Fabp2	0.274729	0.562046	0.1061139
Clca4b	0.4021292	0.8994936	0.07165406
Spink4	0.1114207	0.1497428	0.0844293
Txn1	0.2667113	0.2091366	0.2780316
Guca2a	0.2133063	0.2566306	0.06780826
Plac8	0.385346	0.2835954	0.1872116

그림 3-13. Data Export (ExDEGA Graphic Plus Input) 파일

3.4. Target Gene Plot View & Gene Search

선별된 유전자 또는 연구자가 관심을 가지는 유전자를 대상으로 비교 조합마다 Fold change 의 패턴을 그래프로 표현하고자 할 때 Target Gene Plot View 를 사용한다. 그림 3-14 와 같이 선별한 유전자의 ID 를 복사하여 Target Gene Plot View 창에 붙여 넣고 Expression Plot View 버튼을 클릭하면 Fold change (log 2) 값으로 그래프가 생성된다

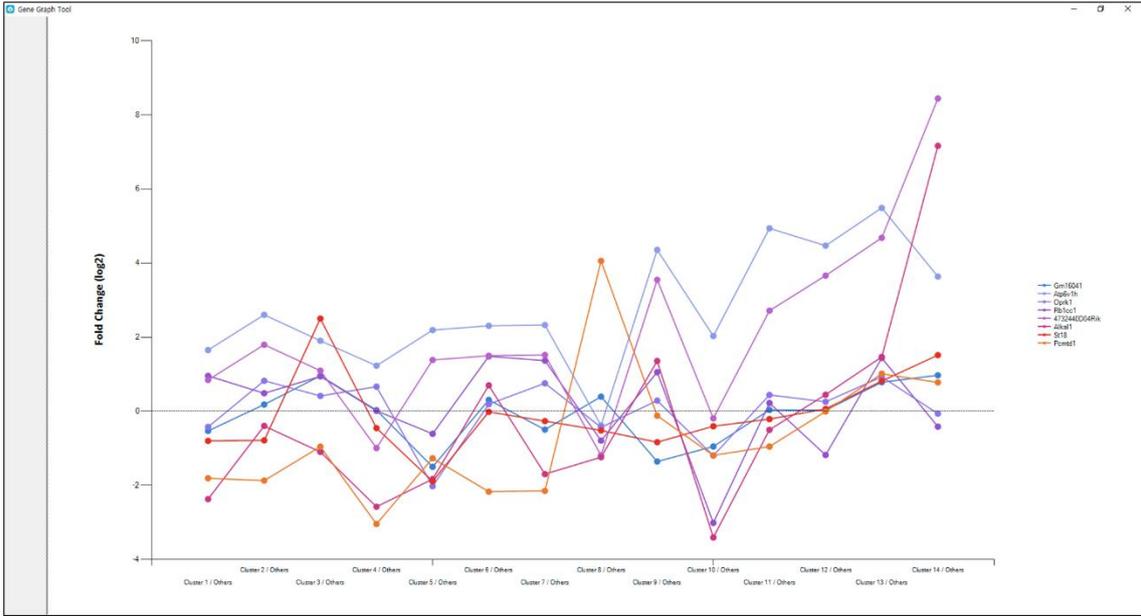


그림 3-14. Target Gene Plot View

특정 유전자를 검색하고 싶을 때는 Gene Search 창을 이용하면 된다. 유전자의 이름을 입력한 후 Search 버튼을 클릭하면 전체 데이터 중에서 해당 유전자에 대한 정보만 확인할 수 있다 (그림 3-15).

ID	Gene symbol	Feature ID	Cluster 1 / Others	Cluster 2 / Others	Cluster 3 / Others	Cluster 4 / Others	Cluster 5 / Others	Cluster 6 / Others	Cluster 7 / Others	Cluster 8 / Others	Cluster 9 / Others	Cluster 10 / Others	Cluster 11 / Others	Cluster 12 / Others	Cluster 13 / Others	Cluster 14 / Others
4084	C1ca4b	ENSMERGCG	0.402	0.899	3.567	0.344	0.798	9.212	0.374	0.393	0.774	2.034	0.234	0.273	0.347	0.0

그림 3-15. Gene Search