User Manual

ExSCEA v1.0.0



90.2

<목 차>

1.	ExSCE	EA 개요3
2.	Categ	jory 사용 방법5
	2.1.	HVG5
	2.2.	Cell Type5
	2.3.	GO6
3.	DE Ar	nalysis 사용 방법8
	3.1.	Significant Selection8
	3.2.	Analysis Graph18
		3.2.1. Volcano Plot26
		3.2.2. Venn Diagram26
	3.3.	Third Part Support18
	3.4.	Target Gene Plot View & Gene Search18

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.

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

표 1. Expression data 구성 항목

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 해체되어 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 RNAseq 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).

🔒 🗟 🗳 🔯 하· 근·	۵	ExSCEA_Sample.xlsx -	Excel	m - a x
파일 홈 삽입 페이지리	레이아웃 수식 데이터 검토 보기	개발 도구 도움말 EXSCEA 🗘 어떤 작업을 원하시나요?		
응 적 사 · · · · · · · · · · · · · · · · · ·	<u> → 11 → オ オ</u> = = = ユ - 田 - <u>ユ - </u> 増 - 増 - 雪雪 _ □	····································	나동 보통 좋음 경고문 메모 설명 텍스트 젤 확인 연결된 생 스타일 스타일	다. 단, 한,
B331 ▼ I × ✓	∫r Rufy4			
v X View All Data HVG Cell Type CO	A B C 1 Filter: 19312 Feature 1 2 ID v symbor 311 309 lkzf2 ENSMU 6G0 312 310 Spag16 ENSMU 5G0	Gene Category Setting	C P Q R If Cluster 12 Cluster 13 Cluster 14 Cluster - / Other -	X DE Analysis Operations Septement selection De Analysis De An
GO Category Extracellular matrix Immune response Inflammatory response	313 311 Bard1 ENSTUSG0 314 312 Abca12 ENMUSG0 315 313 Atic EISMUSG0 316 314 Fn1 INSMUSG0	Extracellular matrix Immune response	New 56 0.952 3.964 1.081 0. 213 6.326 3.2402 3.550 1.1 263 6.326 0.346 0.225 0.346 0.225 298 4.423 8.573 2.485 1.2	103 Third Part Support 000 Third Part Support 000 Target Grane Floct View (ID input) 000 Grane Search
Neurogenesis RNA splicing Secretion	317 315 Gm4319 ENSMUS60 318 316 Mreg ENSMUS60 319 317 D23001741 ENSMUS60 320 318 Perr ENSMUS60 321 319 Temperation ENSMUS60	Inflammatory response Neurogenesis	Delete 378 44,255 537,795 24,853 1.1 65 5,257 20,252 10,404 0.0 10,404 0.0 Edit 1731 17,918 55,916 49,703 1.4 507 1,074 1,082 1,534 0.0 101 17,918 50,916 49,703 1.4	000
	322 320 Xrec ENSMUSGO 323 321 04-Mar ENSMUSGO 324 322 Starcell ENSMUSGO 325 323 pj37a ENSMUSGO	RNA splicing Secretion	Jab 0.347 0.706 2.445 0.0 283 6.326 12.327 29.313 1.1 1.2 22.3 0.791 1.460 0.1 Import 124 0.300 0.795 1.442 0.0	001 000 717 000
	326 327 (gfbp2 ENSMUSG0 327 316 (gfbp5 ENSMUSG0 328 256 (S30043A11 ENSMUSG0 329 327 6030407001 ENSMUSG0 320 328 Trs1 ENSMUSG0		Web ? 283 6.326 12.227 136.693 1.1 929 1.925 3.306 1.081 1.1 1.1 1.11 1.	300 000 000 000 899
	331 329 Rufy4 ENSMUSG0 332 330 Cxer2 ENSMUSG0 333 331 Cxer1 ENSMUSG0 334 332 Arpc2 ENSMUSG0 333 333 Capber1 ENSMUSG0		View list / Import 10 22/142 113/719 22/22 11 View list / Import 88 0.777 139738 0.880 0. 22/202 50577 1.872 0.880 0. 22/202 50577 1.872 0.880 0. 22/202 50577 1.872 0.880 0. 200 26/1 1.777 1.872 0.489 0. 200 26/1 1.975 1.952 1.22 1.	300 146 000 000
● AND ○ OR	33 334 Aamp EN5MUSG0 7 335 Pnkd EN5MUSG0 38 336 Tmbim1 EN5MUSG0 39 337 Catip ENSMUSG0 340 338 SIC11a1 ENSMUSG0		92 1.198 0.377 1.677 0.0 55 2.059 0.055 1.066 0. 621 0.760 0.355 1.067 0.0 621 0.760 0.355 1.067 0.0 63 0.760 0.355 1.067 0.0 64 0.760 0.355 1.067 0.944 0. 61 0.477 0.574 0.354 0.0 0.0	137 227 067 000 299
Gene Category Setting	341 339 Ctdsp1 ENSMUSG0 342 340 Vil1 ENSMUSG0		701 2.459 0.958 0.878 0.1 664 2.548 0.376 0.506 1.4	338 000 ** •

그림 2-2. GO Category 기능

	louse Rat							
Fotal list	: 120 ea		Keyword Search All List					
1 2 3	4							
Check	PCRCat_No	PathwayDisease <u>N</u> ame	Genelist					
	PAHS-002ZA	Drug_Metabolism	ABCB1, ABCC1, ABP1, ADH1B, ADH1C, ADH4, ADH5, ADH6, AHR,					
	PAHS-004ZA	Cancer_Drug_Resistance	ABCB1, ABCC1, ABCC2, ABCC3, ABCC5, ABCG2, AHR, AP1S1, APC,					
	PAHS-005ZA	Estrogen_Receptor_Signaling	AHR, AKAP1, APBB1, BCAR1, BCL2L1, BDNF, BMP4, BMP7, BRCA1,					
	PAHS-006ZA	Glucose_Metabolism	ACLY, ACO1, ACO2, AGL, ALDOA, ALDOB, ALDOC, BPGM, CS, DLA					
	PAHS-007ZA	Fatty_Acid_Metabolism	ACAA1, ACAA2, ACAD10, ACAD11, ACAD9, ACADL, ACADM, ACAD					
	PAHS-011ZA	Inflammatory_Cytokines_Receptors	AIMP1, BMP2, C5, CCL1, CCL11, CCL13, CCL15, CCL16, CCL17, CCL					
	PAHS-012ZA	Apoptosis	ABL1, AIFM1, AKT1, APAF1, BAD, BAG1, BAG3, BAK1, BAX, BCL10,					
	PAHS-013ZA	Extracellular_Matrix_Adhesion_Molecules	ADAMTS1, ADAMTS13, ADAMTS8, CD44, CDH1, CLEC3B, CNTN1, (
	PAHS-015ZA	Endothelial_Cell_Biology	ACE, ADAM17, AGT, AGTR1, ALOX5, ANGPT1, ANXA5, APOE, BAX,					
	PAHS-017ZA	Obesity	ADCYAP1, ADCYAP1R1, ADIPOQ, ADIPOR1, ADIPOR2, ADRA2B, AD					
	PAHS-019ZA	Cellular_Stress_Responses	ACADSB, BAG1, CALR, CANX, CAT, CCT2, CCT3, CCT4, CCT5, CCT7,					
	PAHS-020ZA	Cell_Cycle	ABL1, ANAPC2, ATM, ATR, AURKA, AURKB, BCCIP, BCL2, BIRC5, BR					
	PAHS-021ZA	Common_Cytokines	ADIPOQ, BMP1, BMP2, BMP3, BMP4, BMP5, BMP6, BMP7, CD40LG					
<			>					

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 1을 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).

20 20 40 400 100<	日日 🗸 🖞 ち・さ																						
Norm	파업 홈 삼업 페이기	지 레이아운		1이터 검						업을 입하시니													
	🚔 👗 잡락내기 🛛 Calibr	ri	V 11	 ✓ 1 n[*] 	==-	4. 3	^낮 자동 물 바중		일반	~		2 H C		나방	보통		좋음	경고문	K		🔉 😭 Σমন্ত হব	· 7.	0
SER. SE <	물여넣기 해 석사 부사 가 기	<u>1</u> .	- 0-2	- 11	= = =	-] 병합하고 개	୫ଗ ହଳ •	or . %	, 12.2	조건부	표계산		메모	48	텍스트	셀 확인	연결된	18	실업	4체 서식 · 지유기 ·	정말맞찾	기및
ANDE I A C D E D <thd< th=""> <thd< th=""></thd<></thd<>	열립보드 5	i	22	5		24	8		표시	형식	5 Add - V	191 -			스타일					* ÷	g (월드 11 년 문집	a
No No No P Q R Verter Image: State of the sta	A3965 * 1 × 4	1 50	3963																				
Very 1 Prince: III Very Outer I Outer I <th< td=""><td></td><td></td><td>A</td><td>в</td><td>с</td><td>D</td><td>E</td><td>F</td><td>G</td><td>н</td><td>1</td><td>J</td><td>к</td><td>L</td><td>м</td><td>N</td><td>0</td><td>P</td><td>Q</td><td>R</td><td></td><td></td><td></td></th<>			A	в	с	D	E	F	G	н	1	J	к	L	м	N	0	P	Q	R			
Ver NDAC 2 0 3 open Filter 10 Obter 20	*)	K 1	Filter: 18									Fold cl	hange										* X
With Call yee 0.3 0.4 0.9 0.5 0.50	View All Data	2	ID T	Gene	Feature ID	Cluster 1/	Cluster 2/	Cluster 3/	Cluster 4/	Cluster 5/	Cluster 6/	Cluster 7/	Cluster 8 /	Cluster 9/	Cluster 10	Cluster 11	Cluster 12	Cluster 13	Cluster 14	Cluster 1	DE A	nalysis	
Main Type	MVG CellTime (C)	643	641	Pigr	ENSMUSGO	2.413	0.334	- and	2.317	0.410	3.279	0.328	0.266	0.305	0.503	1.481	0.220	0.397	2,453	0.0	Significant Selection		
Diff Diff Unit 999 999 999 999 2400 ± 10 5500 ± 200 1.100 5500 ± 200 2.500 ± 200 2	tito certype do	7279	7277	Tpi1	ENSMUSGO	2.030	1.105	0.238	1.673	0.362	3.007	1.021	0.568	0.733		1.698	1.900	0.449	0.650	0.0	0		
■ Outer 1 1100 1100 0.2007 0.5400000 2.27 3.28 0.88 2.300 1.27 0.200 2.200 1.20 0.000 ○ Outer 2 1100 1120 0.2007 0.5400000 2.27 1.38 0.280	HVG (P-Value)	9398	9396	Ggt1	ENSMUSGO	2.653	2.386	0.126	1.678	0.501	3.392	1.866	0.000	1.355	0.152	3.250	0.478	0.407	0.249	0.0	Fold change		
□ Cutor 2 0 MMUSO 2.374 288 0.87 1.10 0.28 0.87 0.07 0.97 0.88 0.05 0.88 0.07 0.07 0.97 0.88 0.05 0.88 0.07 0.07 0.97 0.88 0.07 0.07 0.87 0.07 0.87 0.07 0.87 0.07 0.87 0.07 0.87 0.07 0.87 0.07 0.87 0.07 0.87 0.07 0.87 0.07 0.87 0.07 0.87 0.07 0.87 0.07 0.88 0.07 0.07 0.87 0.07 0.87 0.07 0.87 0.07 0.87 0.07 0.87 0.07 0.87 0.07 0.87 0.07 0.87 0.07 0.87 0.07 0.87 0.07 0.87 0.07 0.87 0.07 0	Cluster 1	11700	11698	Gm26870	ENSMUSGO	2.872	6.717	0.266	0.493	0.884	2.379	1.876		2.635	0.340	2.960	0220	2.301	1.957	0.0	2.0	D	0
□ Outor) 1796 1796 1796 1796 1797 1798	Cluster 2	14322	14320	Pcyt2	ENSMUSGO	2.574	2.903	0.387	1.134	1.017	1.939	1.791	0.270	2.574	0.311	2.222	1.186	0.656	0.474	0.0	p-value		
□ Cutor 4 1908 1928 mAA MAMAGE 2.22 2.33 0.39 0.39 1.22 0.43 0.12 0.43 0.12 0.43 0.12 0.43 0.12 0.43 0.12 0.43 0.12 0.43 0.12 0.43 0.12 0.43 0.12 0.43 0.12 0.43 0.12 0.43 0.12 0.43 0.12 0.12 0.12 0.12 0.12 0.12 0.12 0.12 0.12 0.12 0.12 0.12 0.12 0.12 0.12 0.12 0.02 <td>Cluster 3</td> <td>17863</td> <td>17861</td> <td>Gm26917</td> <td>ENSMUSGO</td> <td>2.633</td> <td>1.150</td> <td>0.586</td> <td>0.337</td> <td>1.089</td> <td>3,474</td> <td>0.431</td> <td>0.987</td> <td>1.070</td> <td>0.763</td> <td>0.907</td> <td>0.117</td> <td>3.634</td> <td>1.198</td> <td>0.0</td> <td>0.0</td> <td>0</td> <td>0</td>	Cluster 3	17863	17861	Gm26917	ENSMUSGO	2.633	1.150	0.586	0.337	1.089	3,474	0.431	0.987	1.070	0.763	0.907	0.117	3.634	1.198	0.0	0.0	0	0
□ Outor 5 1928 1928 1428 1428 238 0.0 0.00 0.00 128 128 0.00	Cluster 4	19290	19288	mt-Nd1	ENSMUSGO	4,431	2.037	0.249	0.564	2.522	2.811	0.639	0.393	3.157		1.544	0.116	1.015	0.462	0.0			
□ Outor 6 1982 12980 mic.0a 100 0.02 2.02 2.02 100 0.02 100 0.02	Cluster 5	19291	19285	mt-Nd2	ENSMUSGO	4.327	2.306	0.333	0.506	2.740	2.524	0.729		3.657		1.711	0.115	0.928	0.618	0.0	Sample / Control		
□ Outor 7 1929 12921 ACA 2400 1202	Cluster 6	19292	19290	mt-Co1	ENSMUSGO	4,846	2.840	0.196	0.679	2.404	2.600	1.071	0.229	3.732		1.985	0.000	0.733	0.397	0.0	Name		Un/Dn *
Custor 0 1925 1222 min Allas 0 0000000 0.000	Chuster 7	19293	19291	mt-Co2	ENSMUSGO	4,473	2,495	0.213	0.623	2.624	2.679	0.869	0.250	3.591		1.791		0.717	0.375	0.0	Cluster 1 / Others		Up •
□ Cuter 10 1005 1220 M140 100 0230 1200 M140 1200 1200 M140 1200 1200 M140 1200	Chater 8	19294	19292	mt-Atp8	ENSMUSGO	4.759	2.185	0.338	0.473	2.602	2.671	0.751	0.342	3.483		1.383	0.125	1.018	0.535	0.0	Cluster 2 / Others		Both
□ 0.040 1000 1200 0.04	Chuter 10	19295	19293	mt-Atpo	ENSNUSGO	4,430	2.390	0.248	0.5/1	2.014	2.707	0.810	0.230	3,432		1.709	0.10	0.785	0.375	0.0	Cluster 3 / Others		Both
□ 100 1200 100 100 1200 100		19290	19294	mt-Cos	ENSMUSGO	4,475	2,410	0.291	0.564	2,040	2.009	0.838	0.278	3.304		1.711	0.111	0.315	0.425	0.0	Cluster 4 / Others		Both
□ 0.000 1237 m 402 2550 0.00	Cluster II	10200	19290	mt-Nid4l	ENISMUSCO	5 302	2,262	0.256	0.0454	2.505	2.575	0.643	0.310	4 373		1.569		0.971	0.395	0.0	Cluster 5 / Others		Both
□ 0.04 or 13 1000 1200 1400 0.434 2.30 0.04 0.037 0.328 0.30 0.328 0.035 0.0	Cluster 12	10200	10207	mt-Nd4	ENSMUSCO	4 407	2 576	0.724	0.552	2 712	2 626	0.939	0.267	2 612		1 855	-	0.802	0.451	0.0	Cluster 6 / Others		Both
• AND • AND <td< td=""><td>Cluster 13</td><td>19300</td><td>19295</td><td>mt-Nd5</td><td>ENSMUSGO</td><td>4.744</td><td>2 390</td><td>0.256</td><td>0.630</td><td>2 358</td><td>2.861</td><td>0.878</td><td>0.288</td><td>3,365</td><td></td><td>1.830</td><td>0.103</td><td>0.858</td><td>0.519</td><td>0.0</td><td>Charles 7.10thore</td><td></td><td>Roth</td></td<>	Cluster 13	19300	19295	mt-Nd5	ENSMUSGO	4.744	2 390	0.256	0.630	2 358	2.861	0.878	0.288	3,365		1.830	0.103	0.858	0.519	0.0	Charles 7.10thore		Roth
# AND O CK Oracle 0	Cluster 14	19302	19300	mt-Cytb	ENSMUSGO	4.390	2.462	0.246	0.526	2.879	2,495	0.810		3.725		1.701	0.115	0.787	0.446	0.0	AND	OOR	
• AND OR <		19315																-					
#1937 Gene Category Chart 1938 Gene Category Chart 1939 Gene Category Chart 1939 Gene Category Chart 1939 Gene Category Chart 1939 Gene Category Chart 1932 Gene Category Chart		19316																					
• AND OR One 0		19317																			Gene Catr	gory Chart	
AND OK Ox 0 </td <td></td> <td>19318</td> <td></td>		19318																					
WAND OR 9920 Other Price Search 1922 9923 Other Price Search Other Price Search 1923 9924 Other Price Search Other Price Search 1924 9924 Other Price Search Other Price Search 1925 9924 Other Price Search Other Price Search 1926 1000 1000 1000 Other Price Search		19319																			0		
# AND O 04 Out = 0		19320																			Analysis Graph		
# AND OR 9922		19321																			 Third Part Support 		
* AND O (A P1922 O Gene Sand *1002 P1925 P1925 *1002 P1025 P1025		19322																			Target Gene Plot View (ID in	put)	
* AND O CR 19324 Columna 19324 Columna Columna <td< td=""><td></td><td>19323</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>(a) Canada and</td><td></td><td></td></td<>		19323																			(a) Canada and		
* AND CM 19933 Image: Control of the state of th		19324																			- vene search		
91/27 # 10/28 41E27 % 444 10 10 10 10 10 10 10 10 10 10 10 10 10	AND O OF	19325																					
1927 1927 1927 1927 1927 1927 1927 1927 1927 1927 1927 1927 192 192 192 192 192 19 1	● AnD ○ OK	19326																					
19228 ■ <td></td> <td>19327</td> <td></td>		19327																					
1 1 1 1912월 1952 월 157 운영나다. 혐의 가수 0 월 편 0 표 100		19328	-	-																	*		
9812개 중 13개의 적료도가 있습니다. 혐 🔟 - 🖡 + 100%			2	uta (÷)								4							E.			
	19312개 중 18개의 레코드가 있습	비다, 혐																	개수: 0	합계:(+ 1009

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 plot에서 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).

	*	×					
DE Analysis							
Significant Selection							
Analysis Graph							
Volcano Plot							
Venn Diagram							
Third Part Support							
Target Gene Plot View (ID input)							
🕑 Gene Search							

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-value가 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 Upregulated 이면서 Cluster 2/Others 에서는 Down-regulated 인 유전자)
- **<u>밑줄 친 숫자</u>** : Down-regulated 유전자의 수

Venn Diagram 이미지에서 해당 영역에 마우스 오른쪽 클릭을 하면 Up-regulated, Downregulated, 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.

Third Part Support	Third Part Support	Third Part Support				
Clustering Input KEGG Input DAVID Input	Clustering Input KEGG Input DAVID Input	Clustering Input KEGG Input DAVID Input				
Export Data Select	Export Data Select	Export Data Select				
Name 🔺	Name A	Name *				
Cluster 1 / Others	Cluster 1 / Others	Cluster 1 / Others				
Cluster 2 / Others	Cluster 2 / Others	Cluster 3 / Others				
Cluster 3 / Others	Cluster 3 / Others	Cluster 4 / Others				
Cluster 4 / Others	Cluster 4 / Others	Cluster 5 / Others				
Cluster 5 / Others	Cluster 5 / Others	Cluster 6 / Others				
Cluster 6 / Others	Cluster 6 / Others	Cluster 7 / Others				
Cluster 7 / Others	Cluster 7 / Others	Cluster 8 / Others				
Cluster 8 / Others	Cluster 8 / Others					
		Data Export (ExDEGA Graphic Plus Input)				
Data Export	Data Export	Data Export (DAVID Input)				

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.

Clusteri	ng_input.txt - Windo	ws 메모장	
파일(F) 편	집(E) 서식(Q) 보기	│(⊻) 도움말	
Gene Syr	mbol Clust	er 1 / Others Cluste	er 2 / Others Cluster 3 / Others
Pigr	2.412656 0.333	38233 0.094	58916
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 0.106	0638
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 0.562	2046 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 0.283	35954 0.187	2116
Cycs	0.3875297	0.4196959	0.3136992
Fabp1	0.4988641	0.277174 0.094	52607
Tmsb10	0.2360803	0.3562078	1.550741
Reg3b	0.3487499	0.4446786	0.08138297
Reg3g	0.2104861	0.2088951	0.07882152
Tpi1	2.030389 1.104	762 0.2382583	
Gapdh	0.4746698	0.4283723	0.7407911
Ftl1	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.

☐ KEGG_input.txt - Windows 메모장					
파일(F) 편	집(E) 서식(O) 보기(V) <u>도</u> 움말				
Gene Syr	nbol 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).

DAV	/ID_input.f	txt - Wind	ows 메모징	ł
파일(F)	편집(E)	서식(0)	보기⊘	<u>도</u> 움말
Pigr				
Phgr1				
Serf2				
B2m				
Dstn				
Atp5e				
Tmsb4	х			
S100a6	5			
S100a1	10			
Fabp2				
Clca4b				
Spink4				
Txn1				
Guca2a	a			
Plac8				

그림 3-12. Data Export (DAVID Input) 파일

🧾 Graphicplus_input.txt - Windows 메모장						
파일(F) 편	집(E) 서식(O) 보기(V)	<u>도</u> 움말				
Gene Syr	mbol Cluster 1	/ Others Clus	ter 2 / Others Cluster 3 / Others			
Pigr	2.412656 0.333823	0.09	458916			
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 0.10	60638			
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	6 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 0.283595	0.18	72116			

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).



그림 3-15. Gene Search