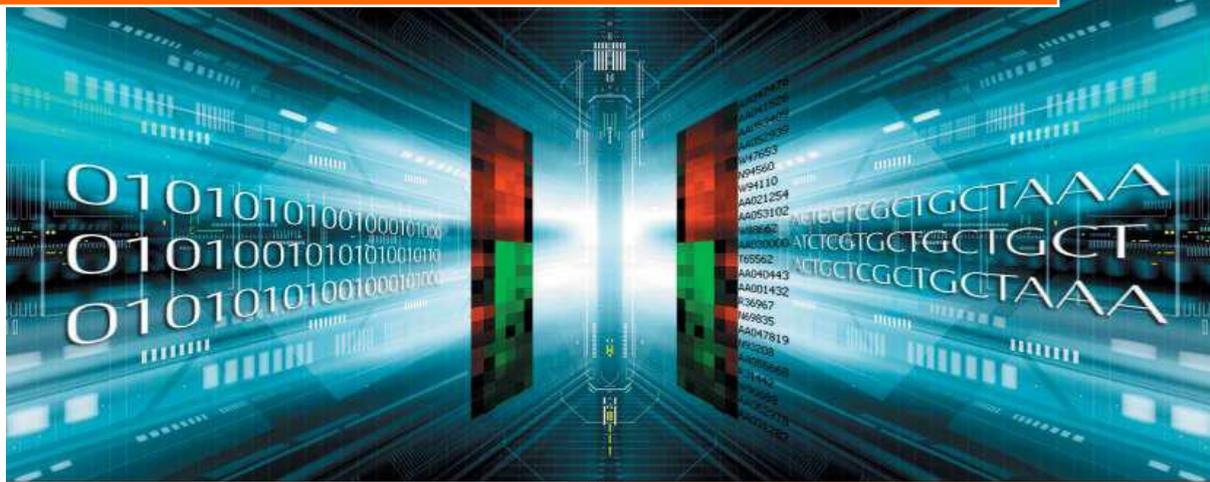


v.1.6.0

# ExDEGA Manual



(주)이바이오젠

서울특별시 영등포구 선유로13길 25  
(문래동6가), 에이스하이테크시티2, 305호  
Tel. 02-3141-0791

[service@e-biogen.com](mailto:service@e-biogen.com)

<http://www.e-biogen.com>

## < 목 차 >

1. ExDEGA Setup

2. ExDEGA Layout

3. Gene Category 사용방법

4. Significant Gene Selection 사용방법

5. Analysis Graph 사용방법

6. Clustering Heatmap Support 사용방법

7. Selected Gene Plot & Gene Search 사용방법

# 1. ExDEGA Setup

㈜이바이오젠은 QuanSeq, mRNA-Seq, Total RNA-Seq 과 Micorarray data 를 엑셀 기반에서 DEG 를 쉽게 분석할 수 있도록 분석보고 시 ExDEGA (Excel based Differentially Expressed Gene Analysis) tool 을 함께 제공한다. ExDEGA 분석툴은 ㈜이바이오젠이 연구자들이 Microarray 및 RNA-Seq 데이터를 보다 쉽게 다루고 원하는 데이터를 쉽게 얻을 수 있도록 사용자 편의를 최대한 반영한 분석툴이고 엑셀 프로그램 안에서 다양한 분석을 직관적으로 수행할 수 있도록 개발되었다. ExDEGA 분석툴은 사용자들의 요구사항을 지속적으로 반영하여 데이터분석과 엑셀사용에 익숙하지 못한 연구자들도 쉽게 사용이 가능하도록 계속 업데이트 될 예정이다.

이바이오젠에서 제공하는 Microarray data 와 RNA-Seq data (엑셀 데이터)를 열기 전에 함께 제공한 ExDEGA(버전).zip 파일의 압축을 풀고 setup 을 실행하면 분석툴이 설치된다(그림 1-1). 설치가 완료되면 보고된 엑셀데이터를 열면 자동으로 ExDEGA 분석툴이 엑셀에 반영된 것을 확인할 수 있다. 참고로 ExDEGA 설치 전에 실행 중인 엑셀 파일이 있으면 종료시킨 후 다시 실행해야 ExDEGA 를 사용할 수 있다.

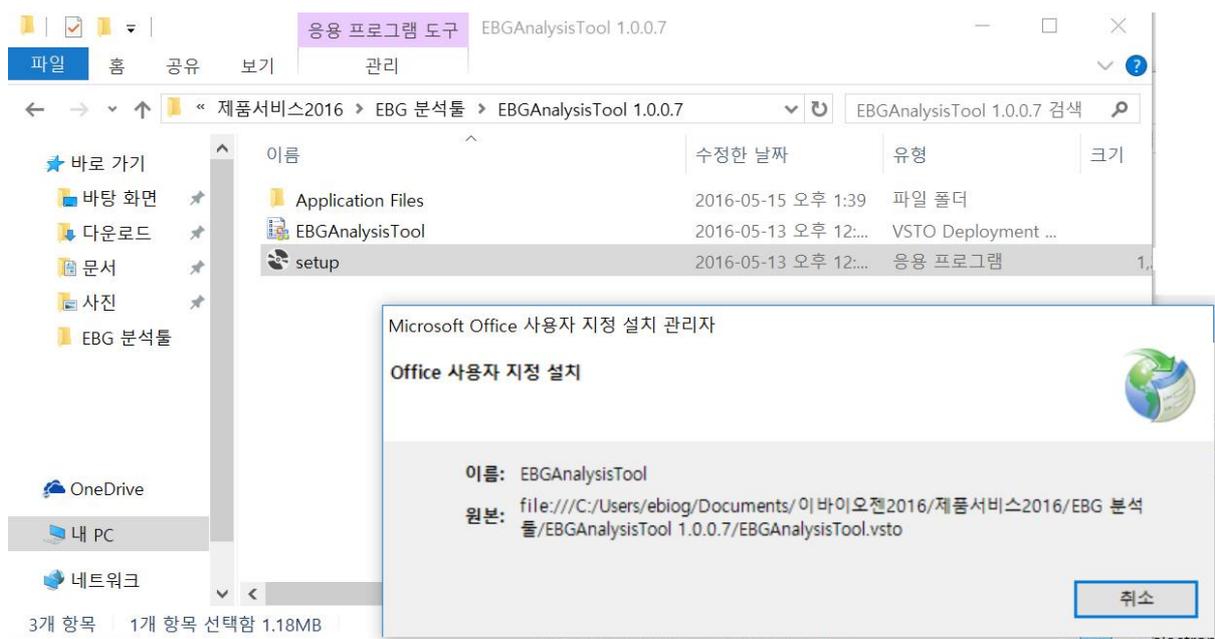


그림 1-1. ExDEGA set up

## 2. ExDEGA Layout

??? ExDEGA Report.xls 파일을 열면 왼쪽에 Gene Ontology (GO) 분석 창과 가운데에 mRNA expression data, 오른쪽에 DEG 분석 창이 나온다(그림 1-2).

GO 분석 창에서는 기본 설정된 GO 와 사용자가 원하는 대로 GO 를 구성하여 분석할 수 있고 DEG 분석과 함께 연동하여 데이터를 쉽게 얻을 수 있다. DEG 분석 창에서는 Fold change, Normalized RC, p-value 등을 선택하여 원하는 데이터를 쉽게 얻을 수 있고 GO graph 를 통해 전체적인 발현패턴을 확인할 수 있다. 뿐만 아니라, DEG 분석 창에서 Scatter Plot, Volcano Plot, Venn Diagram 을 직접 그릴 수 있고 필터링된 유전자들을 대상으로 Clustering heatmap 을 작성하기 위한 MeV 프로그램 input file 을 자동으로 만들 수 있고 Gene expression graph, Gene search 기능도 이용할 수 있어 연구자가 RNA-Seq data 를 쉽게 활용할 수 있다.

Filter: 23420		fold change				p-value				Average of Normalized RC (log2)					
ID	Gene Symbol	A/Cont	B/Cont	B/A	D/C	A/Cont	B/Cont	B/A	D/C	Control	A	B	C	D	Control
3	1.0610005C13Rik	0.836	0.822	0.983	2.133	0.041	0.888	0.039	0.144	7.925	7.666	7.642	9.418	10.511	8.032
4	2.0610007N19Rik	0.837	0.964	1.152	2.466	0.052	0.149	0.115	0.085	10.370	10.113	10.317	4.420	6.596	10.332
5	3.0610007P14Rik	0.925	0.818	0.886	1.016	0.447	0.356	0.029	0.045	10.258	10.183	10.009	11.666	11.689	10.458
6	4.0610008F07Rik	0.876	0.769	1.438	0.810	0.000	0.000	0.000	0.000	1.787	0.697	1.409	6.702	6.398	0.000
7	5.0610009B14Rik	0.847	0.728	0.857	2.000	0.000	0.000	0.055	0.175	6.349	6.108	5.886	2.783	3.818	6.817
8	6.0610009B22Rik	0.806	1.145	1.420	0.961	0.126	0.031	0.161	0.005	8.353	8.042	8.548	10.237	10.180	8.575
9	7.0610009D07Rik	0.980	1.086	1.108	1.391	0.011	0.092	0.050	0.255	10.963	10.034	10.182	10.829	11.805	10.114
10	8.0610009L18Rik	0.851	1.503	1.767	0.567	0.447	0.604	0.023	0.502	7.172	6.940	7.761	5.285	4.466	7.479
11	9.0610009Q20Rik	0.991	0.961	0.970	1.096	0.046	0.913	0.030	0.275	12.033	12.019	11.976	11.851	11.983	11.897
12	10.0610010B08Rik	0.969	0.935	0.965	0.982	0.025	0.008	0.028	0.031	11.892	11.847	11.796	13.680	13.604	12.155
13	11.0610010F05Rik	1.079	1.235	1.145	1.039	0.395	0.181	0.325	0.510	7.894	8.004	8.199	8.501	8.550	7.772
14	12.0610010L14Rik	0.901	1.140	1.266	1.560	0.377	0.020	0.465	0.029	13.911	11.780	14.130	14.384	14.915	13.775
15	13.0610011F06Rik	0.788	1.034	1.313	1.722	0.044	0.008	0.053	0.072	10.740	10.356	10.789	9.241	10.025	10.710
16	14.0610012G68Rik	0.839	1.063	1.268	0.757	0.022	0.801	0.269	0.177	9.865	9.611	9.953	10.506	10.104	9.881
17	15.0610012H03Rik	1.000	1.000	1.000	1.000	1.000	1.000	0.022	0.050	0.000	0.000	0.000	4.446	4.886	4.367
18	16.0610013E20Rik	1.205	1.013	0.841	0.790	0.024	0.903	0.264	0.239	10.438	10.706	10.457	10.331	10.006	10.569
19	17.0610013J06Rik	0.906	0.918	1.013	0.694	0.287	0.000	0.000	0.083	12.628	12.485	12.504	13.066	12.539	12.568
20	18.0610013L16Rik	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000
21	19.0610017L13Rik	0.867	1.140	1.315	0.786	0.028	0.478	0.190	0.099	10.036	9.830	10.225	9.969	9.621	9.974
22	20.0610018B21Rik	0.811	0.746	0.920	1.194	0.235	0.331	0.480	0.300	6.312	6.030	5.890	5.454	5.673	6.557
23	21.0610019L08Rik	1.119	1.050	0.947	1.132	0.394	0.885	0.737	0.466	8.124	8.287	8.208	8.537	8.716	8.303
24	22.0610019K40Rik	0.822	0.710	0.863	0.459	0.155	0.161	0.470	0.097	12.194	11.911	11.699	9.216	8.093	12.132
25	23.0610040B10Rik	0.847	1.141	1.364	0.811	0.373	0.841	0.047	0.148	3.857	3.228	4.047	4.940	1.895	3.878
26	24.0610040F04Rik	1.633	1.121	0.686	0.711	0.097	0.892	0.275	0.077	8.192	8.900	8.357	10.137	8.025	8.475
27	25.0610040D10Rik	0.795	0.847	1.066	0.966	0.022	0.213	0.023	0.434	10.076	9.744	9.936	5.313	4.324	10.168
28	26.0610043K17Rik	1.049	1.169	1.114	0.696	0.000	0.050	0.015	0.000	2.174	2.444	2.800	4.005	3.937	2.972
29	27.110001G20Rik	1.351	1.440	2.550	1.000	0.455	0.000	0.284	1.000	4.669	5.103	6.454	0.000	0.000	4.830
30	28.110001A18Rik	1.005	1.766	1.756	0.816	0.032	0.011	0.037	0.154	8.064	8.072	8.884	9.878	9.585	8.243
31	29.110001J03Rik	0.842	1.944	2.311	0.962	0.198	0.014	0.304	0.014	8.103	7.855	9.063	8.872	8.816	7.920
32	30.110002L16Rik	1.009	0.969	0.960	1.209	0.040	0.047	0.063	0.189	7.980	7.993	7.934	8.621	8.845	8.080
33	31.110004F99Rik	0.774	1.115	1.442	1.314	0.045	0.910	0.021	0.117	9.494	9.123	9.451	9.065	9.458	9.451

그림 1-2. mRNA expression data format made in E-Biogen

## 3. Gene Category 사용 방법

mRNA expression data 는 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. GO 는 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.

Filter: 259		Fold change				p-value				Average of Normalized RC (log2)				
ID	Gene Symbol	A/Contr	B/Contr	B/A	D/C	A/Contr	B/Contr	B/A	D/C	Control	A	B	C	D
1775	1773 Abat	0.793	0.890	1.122	1.898	0.205	0.027	0.723	0.147	9.232	8.899	9.064	9.310	10.234
1989	1987 Ada	0.828	0.857	1.035	0.426	0.410	0.022	0.040	0.028	9.367	9.095	9.145	12.675	11.444
2094	2092 Adm	0.587	0.722	1.230	0.231	0.057	0.262	0.049	0.061	8.505	7.737	8.035	5.110	3.000

그림 1-3. Gene ontology (Aging) selection

'View All Data' 버튼을 누르면 필터를 해제하여 다시 전체 결과를 볼 수 있고 15개의 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.

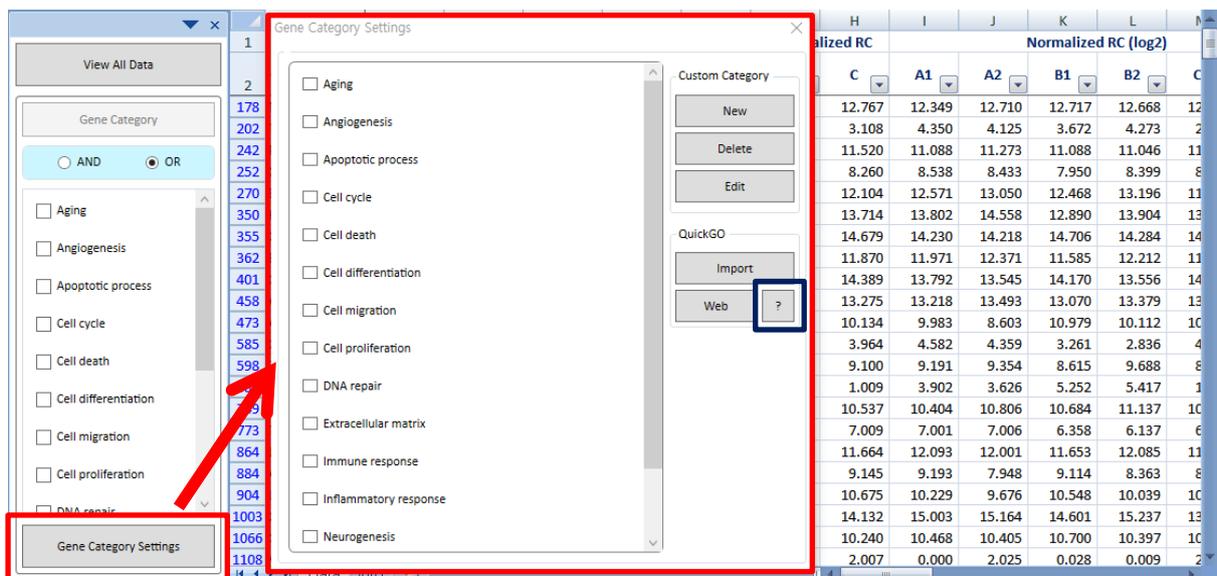


그림 1-4. Gene category settings

만약 원하는 유전자 그룹 목록이 있다면, 직접 입력하여 새로운 Gene Category 를 추가할 수도 있다. Gene Category Settings 버튼을 누른 후 New 를 선택하고 원하는 gene list 입력(or 복사-

붙여넣기) 한 뒤, Gene category 이름 설정 후 저장하면 새로운 GO category 를 확인 할 수 있다(그림 1-5-a,b).

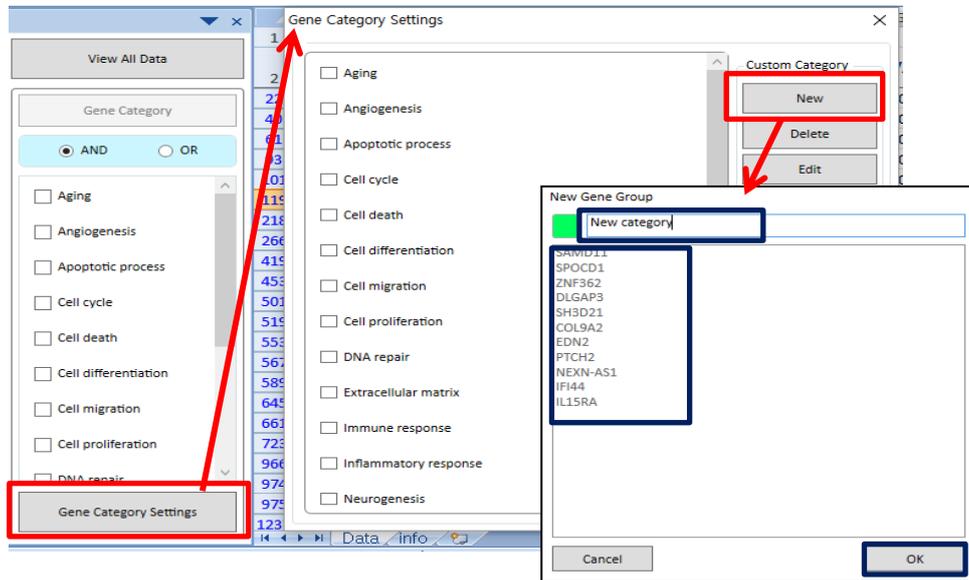


그림 1-5-a. Adding Genes to make a new gene category

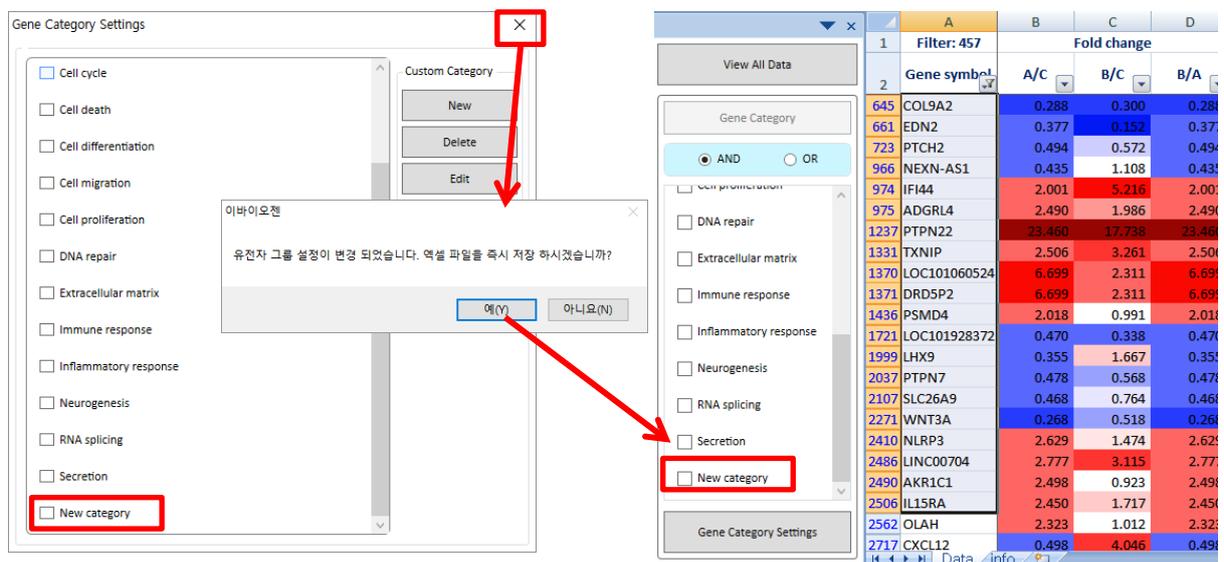


그림 1-5-b. Adding Genes to make a new gene category

#### 4. Significant Gene Selection 사용 방법

오른편의 DEG Analysis 부분에서 "Significant Gene Selection" 창은 전체 결과 중 control 과 test 를 비교한 결과에서 유의하게 발현 차이가 나는 유전자를 필터링 할 수 있도록 만들어 놓은 것이다. 예를 들어, control 기준으로 A 에서 발현이 2 배 이상 증가 또는 감소하고, normalized RC(log)값이

4 이상이고, t-test 결과 p-value 값이 0.05 이하인 유전자(반복 실험한 데이터의 경우)를 선택하면 95 개의 유전자가 필터링 된다(그림 1-6).

그리고 여러 개의 비교그룹에서 동시에 Significant gene 을 선별하고자 할 경우와 적어도 한 비교그룹에서 Significant gene 을 선별하고자 할 경우에는 "AND"와 "OR" 기능을 사용하면 된다.

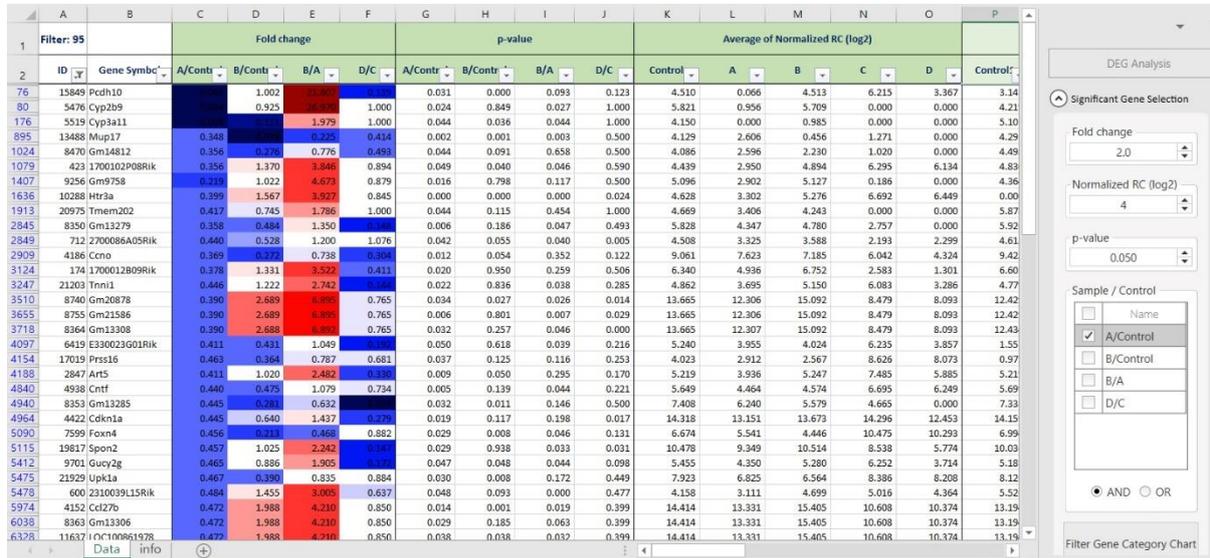


그림 1-6. Significant gene selection

Gene Category 와 Significant gene selection 은 연동 가능하다. 그림 1-7 에서 처럼 Gene Category 의 Cell differentiation 을 선택하면 10 개의 유전자가 필터링 된다(그림 1-7). 10 개의 유전자는 본 데이터에서 Cell differentiation 관련 유전자들 중 A/Control 비교그룹에서 유의하게 발현이 증가 또는 감소한 유전자를 의미한다.

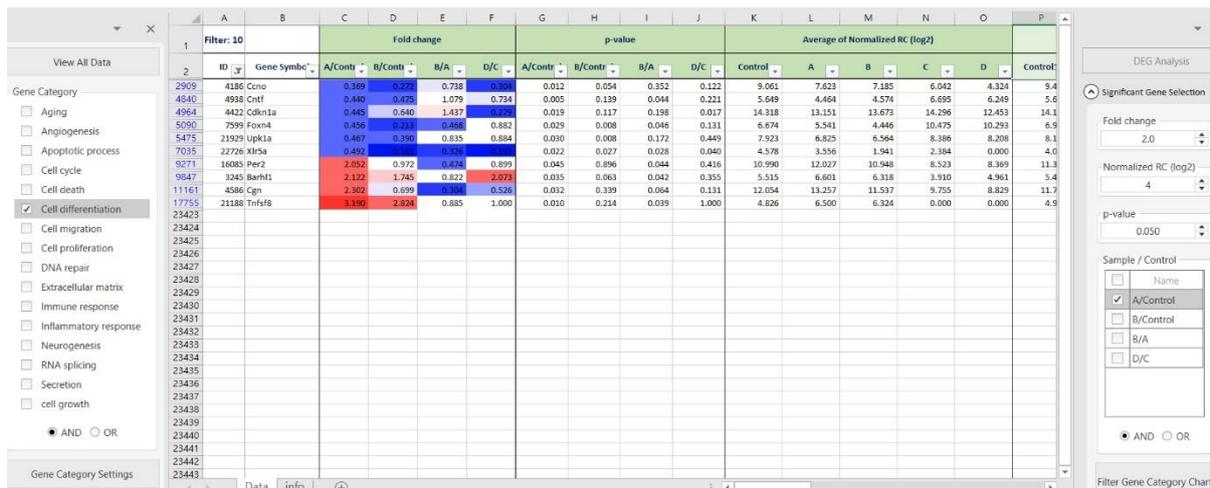


그림 1-7. Significant genes related to Cell cycle

실험 결과에 따라 발현 변화값 (fold change), p-value, normalized RC(log2) 기준을 조정할 수 있고 반복 실험인 경우만 p-value 를 선택할 수 있다.

“View Gene Category Chart” 버튼을 누르면 각 GO 관련 유전자 중 발현이 유의하게 차이 나는 유전자의 %와 수가 그래프로 그려진다. 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 유전자 들 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 되고 오른쪽의 증가/감소된 bar chart에서 bar 상단의 숫자는 해당 유전자 수이고 bar를 클릭하면 해당 유전자가 필터링 된다(그림 1-8).



그림 1-8. View Gene Category Chart

## 5. Analysis Graph 사용 방법

DEG Analysis 부분에서 “Analysis Graph” 창을 펼치면 아래 그림 1-9와 같이 Scatter Plot, Volcano Plot, Venn Diagram을 엑셀에서 쉽게 그릴 수 있다.

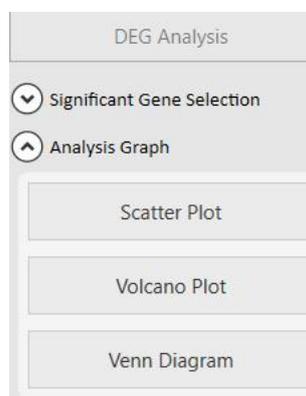


그림 1-9. Analysis Graph Tool

첫번째 Scatter Plot 은 오른쪽에 샘플 비교그룹을 선택하고 Fold threshold line 을 선택하고 "Graph View"를 클릭하면 왼쪽에 선택한 비교그룹을 대상으로 Scatter Plot 이 자동 생성된다. Plot 에서 특정 spot 을 클릭하면 해당 유전자가 표시되고 마우스 오른쪽을 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 Gene Symbol 이 자동 생성된다(그림 1-10).

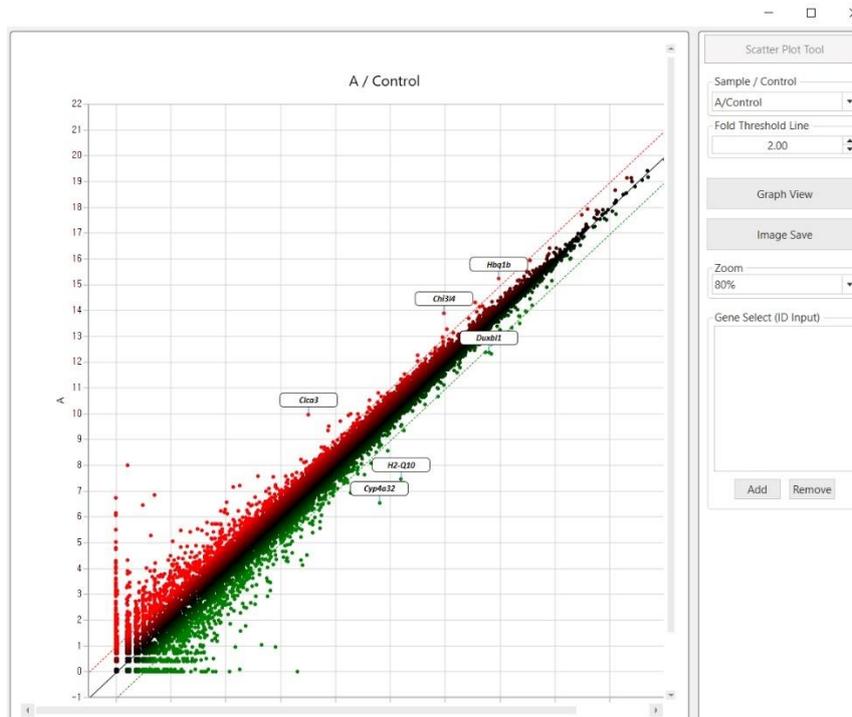


그림 1-10. Analysis Graph Tool – Scatter Plot

두번째 Volcano Plot 은 Scatter Plot 의 기능과 거의 동일한데 오른쪽에 샘플 비교그룹을 선택하고 Fold threshold line 과 p-value 를 선택하고 "Graph View"를 클릭하면 왼쪽에 선택한 비교그룹을 대상으로 Scatter Plot 이 자동 생성된다. Plot 에서 특정 spot 을 클릭하면 해당 유전자가 표시되고 마우스 오른쪽을 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 Gene Symbol 이 자동 생성된다(그림 1-10).

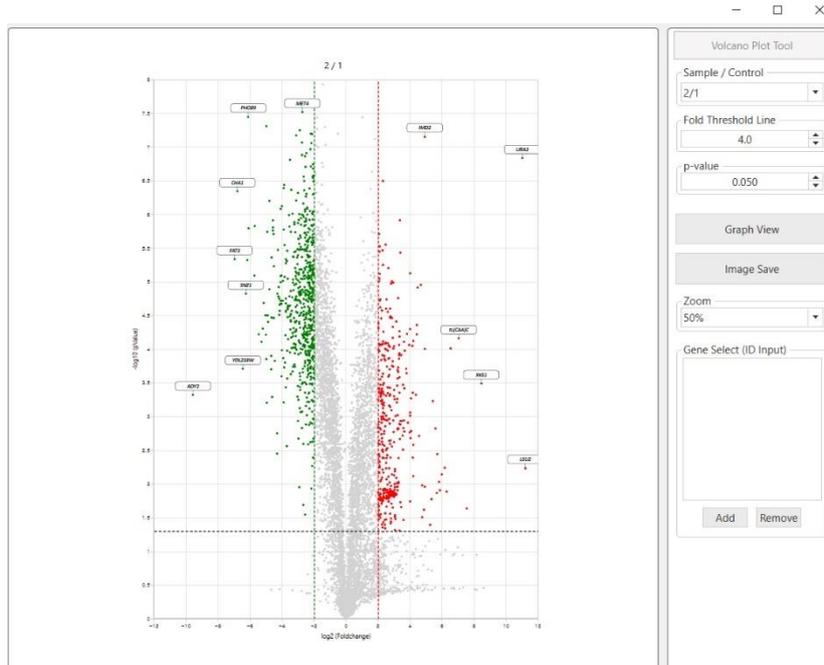


그림 1-11. Analysis Graph Tool – Volcano Plot

세번째 Venn Diagram 을 통해 2 개, 3 개 또는 4 개 까지의 비교그룹을 대상으로 Venn Diagram 을 작성할 수 있다. Venn Diagram 을 그릴 샘플 비교그룹과 Fold Change, p-value(반복실험시)을 선택 후, Diagram View 를 클릭하면 결과를 확인할 수 있으며 그룹은 최대 4 그룹까지 선택 가능하다. 아래의 그림은 A/C 와 B/C, B/A 결과 중, 2fc 이상 up, down 된 list 를 가지고 Venn Diagram 을 작성한 결과이다(그림 1-12).

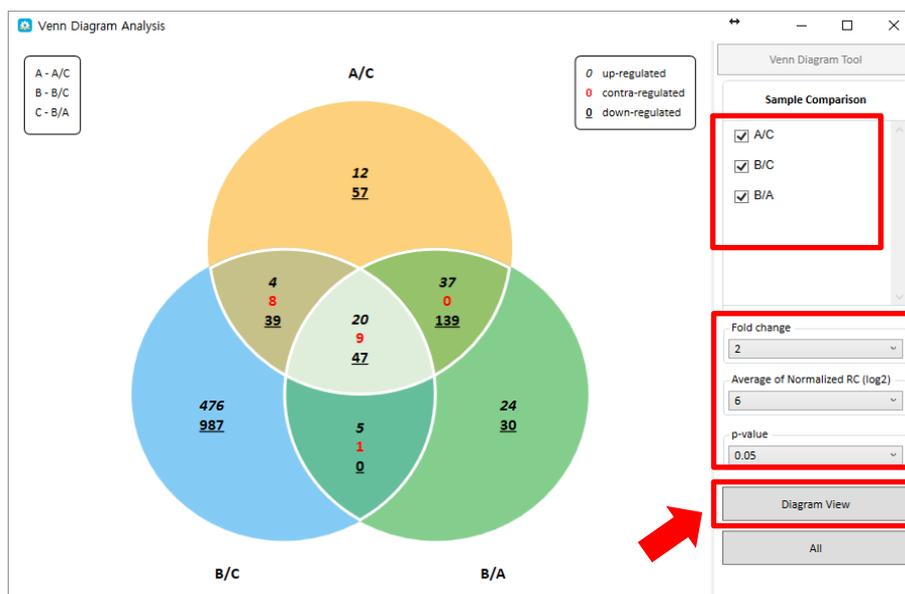


그림 1-12. Analysis Graph Tool – Venn Diagram

Venn Diagram 결과에서 표시되는 형식은 다음과 같다(그림 1-13).

1. **기울어진 숫자** : 2fold 이상 up-regulated 된 gene 수
2. **빨간색 숫자** : regulation 이 대조되는 gene 수
3. **밑줄 친 숫자** : 2fold 이상 down-regulated 된 gene 수

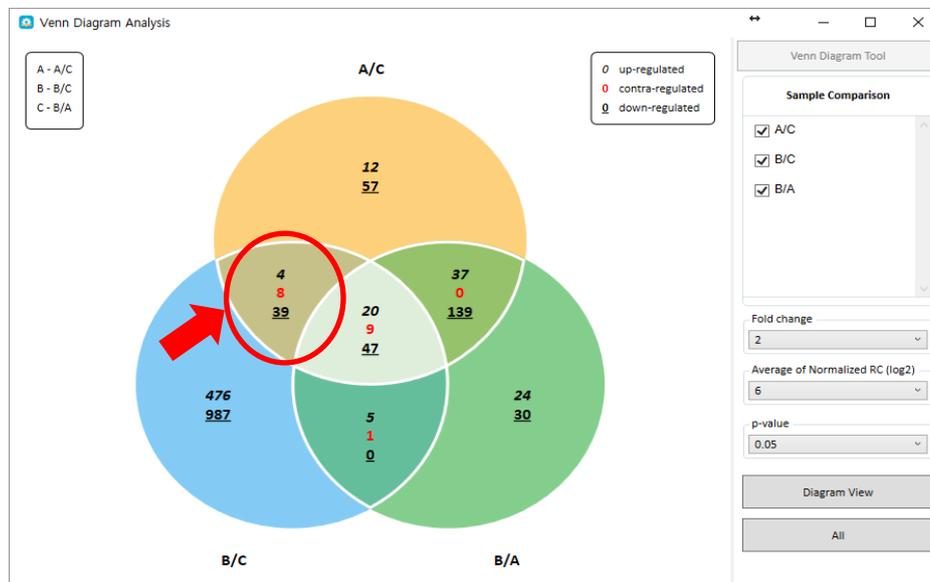


그림 1-13. For example of up ,down, contra-regulated in Venn Diagram

Venn Diagram 이미지를 오른쪽 클릭하면 Venn Diagram 각 영역에 어떤 유전자들이 있는지 확인할 수 있다. 예를 들어, A/C 에서만 2fold up 이 되는 유전자를 보고 싶으면, Venn Diagram 에서 A/C 에서만 해당되는 영역을 찾아 마우스 오른쪽 클릭 하면 2fold up 된 유전자 list 4 개가 엑셀 sheet 에 filter 된다(그림 1-14).

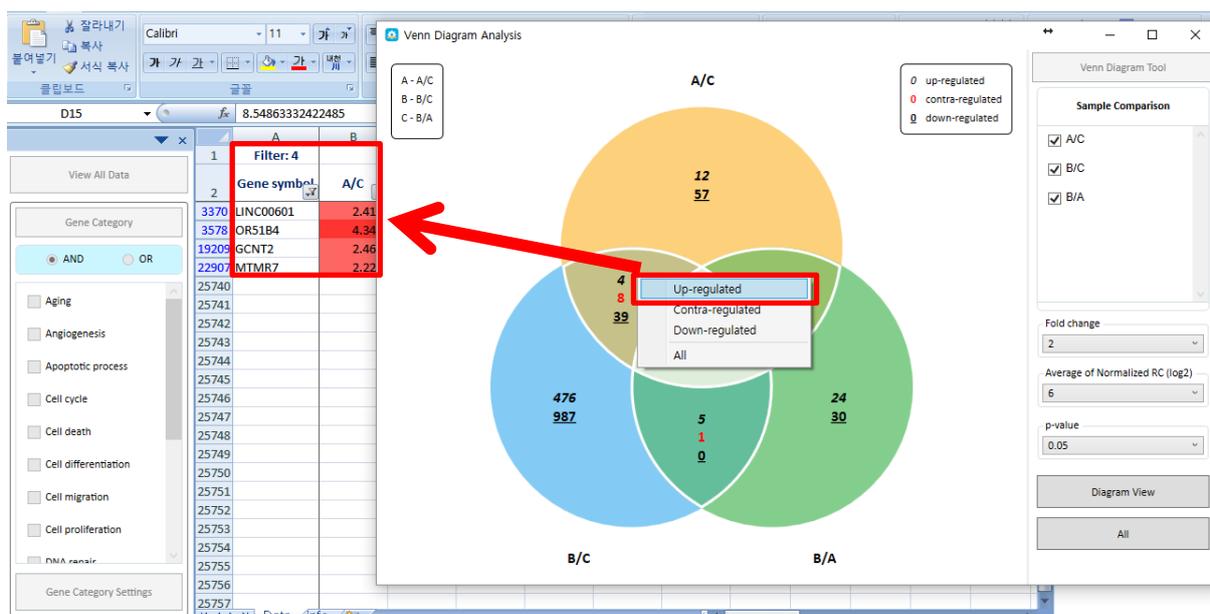


그림 1-14. Filtering 2fold up-regulated gene list in Venn Diagram

ExDEGA 에서 제공되는 모든 이미지는 오른쪽마우스를 눌러 'Save image' 버튼을 통해 저장이 가능하다(그림 1-15).

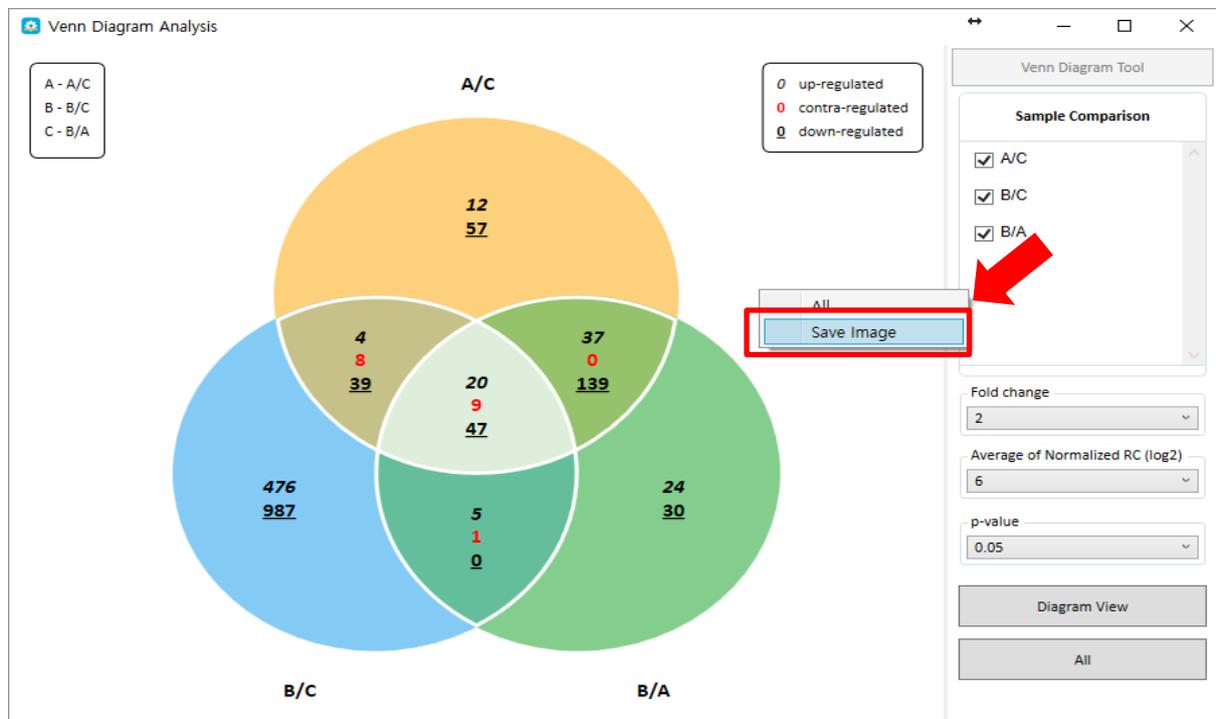


그림 1-15. Save image

## 6. Clustering Heatmap Support 사용 방법

ExDEGA 의 DEG Analysis 에서는 Significant Gene Selection 또는 Venn Diagram 등을 통해 Data Mining 을 수행한 후 정리된 유전자 리스트를 대상으로 Clustering Heatmap 을 쉽게 작성할 수 있도록 지원한다.

당사에서 추천하는 Clustering Heatmap 프로그램은 MeV 인데 ExDEGA 에서 MeV 용 Input file 을 자동 생성해 주고 MeV 에서 해당 파일을 불러오면 된다. 이후의 Clustering 방법 및 이미지 가공 및 저장 방법은 본 매뉴얼 "4. MeV Software 이용 Clustering Heatmap 작성" 부분을 참고하면 된다.

그림 1-16 에서 필터링된 유전자 리스트를 대상으로 Clustering Heatmap 을 작성하려면 크게 2 종류의 데이터를 이용할 수 있는데 첫번째는 Fold change 값을 이용할 시 Type 부분에 Fold change 를 체크하고 Export Data Select 에서 Heatmap 에 표현할 비교그룹을 체크하여 "Data Export"를 클릭한 후 "???.txt"로 저장하면 된다. 두번째는 발현값(Raw Data(RC))으로 표현하고자 할 때 Raw Data 를 체크하고 샘플이 3 개 이상이면 z-score 를 체크하고 샘플이 2 개면 median 을 체크하고 Export Data Select 에서 Heatmap 에 표현할 비교그룹을 체크하여 "Data Export"를 클릭한 후 "???.txt"로 저장하면 된다

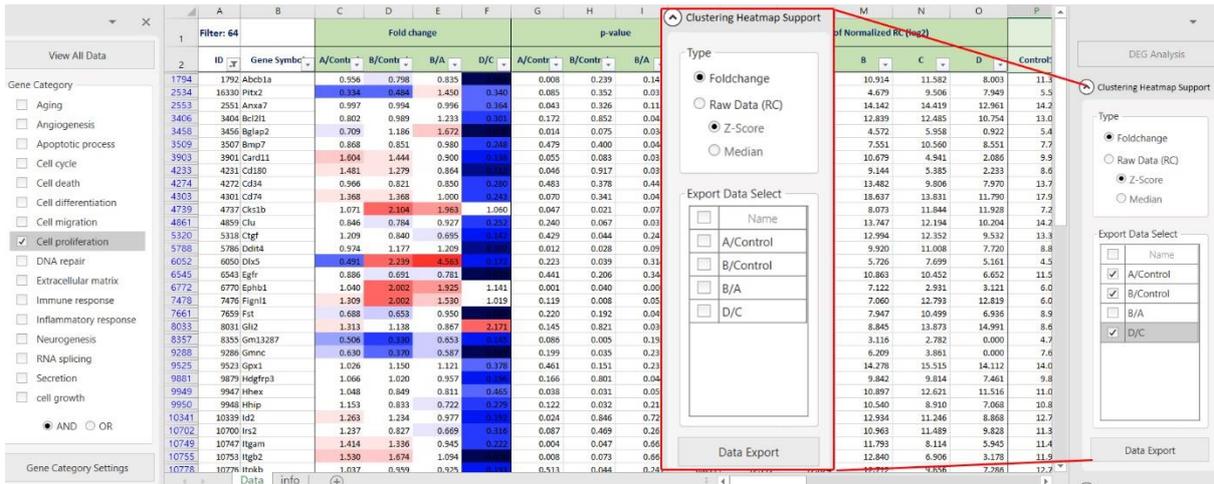


그림 1-16. Clustering Heatmap Support

## 7. Selected Gene Plot & Gene Search 사용 방법

ExDEGA의 기능 중에 선별한 유전자 또는 연구자가 관심있는 유전자들을 대상으로 발현패턴을 그래프로 표현하고자 할 때는 "Selected Gene Plot" 기능을 사용하면 된다.

선별한 유전자의 gene symbol 을 복사하여 Selected Gene Plot 창에 붙여 넣고 "Expression Plot View"를 누르면 normalized RC(log2) 값, fold change 값으로 line graph 가 그려진다(그림 1-17).

그리고 특정 keyword 관련 유전자를 검색하고 싶을 때는 gene search 창을 이용하면 된다. 예를 들어 'insulin'을 검색하면 엑셀 Data Sheet 에 'insulin' keyword 을 포함하는 모든 유전자가 검색되어 필터링 된다(그림 1-18).

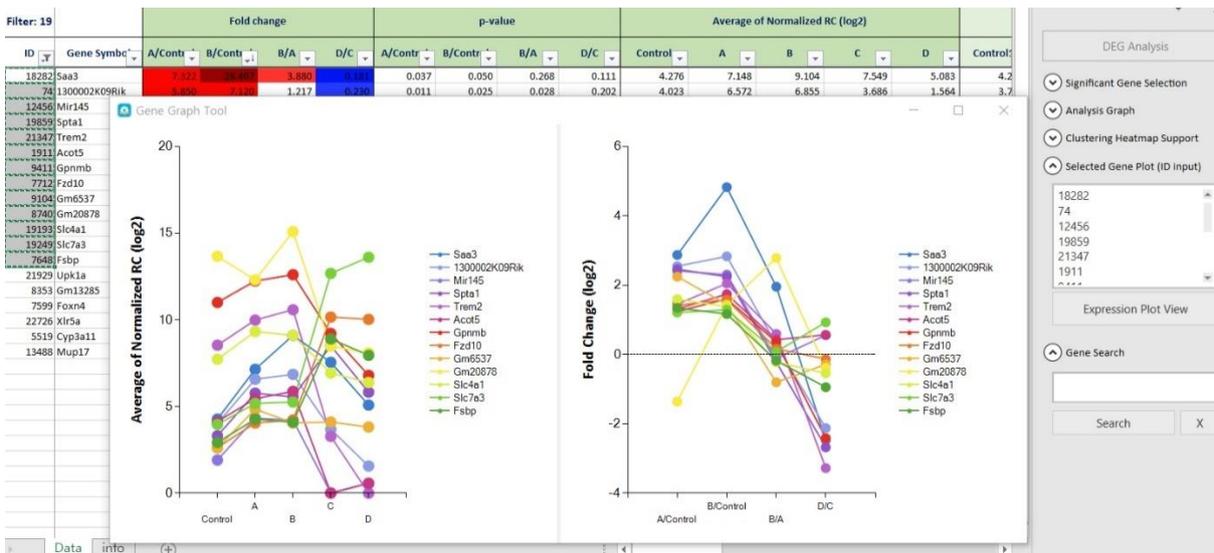


그림 1-17. Gene graph

1	A	B	Fold change				p-value				Average of Normalized RC (log2)					P	Q	R	S
			B/Con	E232-25/Con	E305-100/Con	LPS/Con	B/Con	E232-25/Con	E305-100/Con	LPS/Con	Con	B	E232-25	E305-100	LPS				
36	Filter:																		
2	ID	Gene symb																	
375	373	lgfbp2	0.747	0.385	0.384	0.596	0.570	0.017	0.017	0.222	1.397	0.975	0.019	0.016	0.650	1.510	0.000	0.000	0.000
376	374	lgfbp5	1.337	0.715	0.305	0.427	0.811	0.750	0.409	0.488	1.745	2.164	1.261	0.033	0.517	2.485	2.994	1.902	0.000
458	456	ins1	1.699	0.709	0.623	0.724	0.704	0.756	0.644	0.771	1.929	2.694	1.432	1.247	1.463	2.693	3.578	2.111	1.303
685	683	insig2	0.756	0.983	1.091	0.877	0.661	0.980	0.902	0.817	9.202	8.799	9.178	9.327	9.012	8.483	8.150	8.283	8.403
999	1997	lgf1	0.547	0.460	0.484	0.481	0.451	0.311	0.340	0.380	13.831	12.961	12.711	12.784	12.774	14.318	13.550	12.740	13.073
1484	2482	lgfbp1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1485	2483	lgfbp3	0.674	0.883	1.213	3.131	1.000	0.755	0.750	0.477	0.569	0.000	0.390	0.847	2.215	0.915	0.000	0.658	1.357
1535	3533	lgf2bp1	1.487	0.311	0.705	0.444	0.556	0.351	0.586	0.123	1.838	2.411	5.042	1.334	5.729	1.450	1.576	5.840	1.974
1635	3633	lgfbp4	3.897	0.311	0.311	0.446	0.357	0.270	0.530	0.486	9.947	11.932	5.293	8.977	8.778	10.658	12.617	4.215	9.315
1861	4359	insm2	0.816	0.628	0.828	1.553	0.862	0.694	0.845	0.749	2.093	1.801	1.422	1.822	2.728	2.882	2.577	2.098	1.296
1417	7415	lgfbp6	1.422	4.056	0.980	1.822	0.455	0.249	0.664	0.428	0.053	0.561	2.087	0.024	0.918	0.000	0.000	2.643	0.000
7702	7700	lgf2bp2	0.626	2.391	1.330	1.217	0.332	0.253	0.319	0.611	9.974	9.298	11.231	10.385	10.257	10.274	9.668	11.668	10.487
1264	8262	lgf2r	1.190	0.763	0.946	0.776	0.885	0.810	0.958	0.811	10.204	10.455	9.814	10.124	9.838	8.350	8.365	8.406	8.783
1472	8470	lgfals	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0355	10353	ins16	0.558	1.070	0.453	1.998	0.470	0.925	0.341	0.361	8.821	7.979	8.919	7.679	9.820	8.025	7.045	8.008	7.236
0420	10418	ide	1.049	0.963	1.224	1.234	0.834	0.900	0.587	0.370	7.537	7.606	7.483	7.829	7.840	7.804	7.612	7.710	8.130
0590	10588	ins1	0.823	0.417	0.416	0.418	0.825	0.414	0.414	0.413	1.778	0.996	0.016	0.013	0.010	1.929	1.580	0.000	0.000
1094	12092	insm1	0.676	2.366	0.685	1.065	1.000	0.270	0.383	0.905	0.564	0.000	1.806	0.019	0.855	0.928	0.000	2.253	0.000
2930	12928	Rxfp1	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

DEG Analysis

Significant Gene Selection

Analysis Graph

Clustering Heatmap Support

Selected Gene Plot (ID Input)

Gene Search

insulin

Search X

그림 1-18. Genes related to insulin