v.1.6.0

ExDEGA Manual



(주)이바이오젠

서울특별시 영등포구 선유로13길 25 (문래동6가), 에이스하이테크시티2, 305호 Tel. 02-3141-0791 <u>service@e-biogen.com</u> 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 ExDEGA를 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.

파일 홈 삽입 페이지 레	1이아들	는 수식	데이터	검토 보기	'l ExDEG	A Acrob	at Qo	1면 작업을	원하시나요?	1								1 m	유 공유
Calibri		- 11	- 7î 7î		<i>∾</i> • ∎	P텍스트 줄	바꿈	숫자		-	1	표준 2	표준				∑ 자동 합계 ·	3 7 🔎	
분여넣기 · · · · · · · · · · · · · · · · · · ·	•	• <u>å</u> •	가 ~ 병원 ~	= = =	• •	방합하고 :	가운데 맞렴	• 🖬 •	% * 觉	· 조건부	표니	나뽕	보통	Ŧ	산입 삭제	서식	제우기 -	정렬 및 찾기 및 필터 + 선택 +	
클립보드 5	÷	12	15		맞클	5		5 3	표시 형식	15		스타일			생		Ę	편집	~
P6 * ! × ~	fx.	0																	۷
	1.1	A	В	С	D	E	F	G	н	1	J	K	L	М	N	0	P		
* X	1	Filter: 23420			Fold d	hange			p-v	alue			Average of	Normalized R	C (log2)		-		* X
View All Data	2	ID J	Gene Symbo	A/Conti	B/Conti 👙	B/A 📮	D/C 🛫	A/Contr 👙	B/Contr 🔆	B/A 🛫	D/C 📮	Control 🛫	A 📮	B 🖵	C 🖵	D 📮	Control:	DEG Ana	lysis
Gono Catogoni	3	1	0610005C13Rik	0.830	5 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		alaction
Gene Category	4	2	0510007N19Rik	0.83	7 0.964	1.152	4.518	0.082	0.149	0.115	0.085	10.370	10.113	10.317	4.420	6.596	10.332	G agninicant Gene a	election
Aging	5	3	0610007P14Rik	0.92	0.818	0.886	1.016	0.447	0.356	0.029	0.045	10.298	10.183	10.009	11.666	11.689	10.458	Analysis Graph	
Angiogenesis	7	4	0610008F07RIK	0.4/	0.705	1.030	2.050	0.219	0.008	0.044	0.000	6 349	6.109	5,996	0.702	0.338	6.000	Churterian Maater	an funnad
Apoptotic process	8	6	0610009B22Rik	0.800	5 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	Constanting freating	ab pabbour
Cell evela	9	7	0610009D07Rik	0.980	1.086	1.108	1.391	0.011	0.092	0.050	0.255	10.063	10.034	10.182	10.829	11.305	10.114	Selected Gene Plo	ot (ID input)
Cell cycle	10	8	0610009L18Rik	0.85	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	0	
Cell death	11	9	0610009O20Rik	0.993	L 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	Gene Search	
Cell differentiation	12	10	0610010B08Rik	0.965	0.935	0.965	0.982	0.025	0.008	0.026	0.031	11.892	11.847	11.796	13.680	13.654	12.155		
Coll migration	13	11	0610010F05Rik	1.075	1.235	1.145	1.039	0.395	0.181	0.325	0.510	7.894	8.004	8.199	8.501	8.556	7.772		
Cell migration	14	12	0610010814818	0.90	1.140	1.200	1.560	0.377	0.020	0.465	0.029	10.740	10 295	14.120	9 241	10.025	13.775		
Cell proliferation	16	14	0610012G03Rik	0.835	1.063	1.313	0.757	0.012	0.801	0.269	0.072	9.865	9,611	9,953	10.506	10.104	9.881		
DNA repair	17	15	0610012H03Rik	1.000	21/790	21.794	0.802	1.000	0.022	0.050	0.000	0.000	0.000	4.446	4.686	4.367	0.000		
Enterseally day matrix	18	16	0610030E20Rik	1.200	1.013	0.841	0.798	0.094	0.983	0.264	0.239	10.438	10.706	10.457	10.331	10.006	10.569		
	19	17	0610031J06Rik	0.900	5 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		
Immune response	20	18	0610031016Rik	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		
Inflammatory response	21	19	0610037L13Rik	0.86	7 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		
Nourogaporir	22	20	0610038821Rik	0.81	1 0.745	0.920	1.164	0.235	0.331	0.680	0.300	0.312	6.010	5.890	5,454	5.675	6.557		
iveurogenesis	24	21	0610039K10Rik	0.82	0 710	0.863	0.459	0.354	0.003	0.757	0.400	12 194	11 911	0.200	9.216	8 093	12 132		
RNA splicing	25	23	0610040B10Rik	0.64	1.141	1.764	0,100	0.373	0.841	0.047	0.148	3.857	3.228	4.047	4,940	1.895	3.878		
Secretion	26	24	0610040F04Rik	1.63	1.121	0.686	0.231	0.097	0.892	0.275	0.077	8.192	8.900	8.357	10.137	8.025	8.475		
coll arouth	27	25	0610040J01Rik	0.795	0.847	1.066	0.504	0.022	0.213	0.023	0.434	10.076	9.744	9.836	5.313	4.324	10.168		
Ceir growth	28	26	0610043K17Rik	1.04	1.169	1.114	0.656	0.780	0.050	0.015	0.000	2.374	2.444	2.600	4.005	3.397	2.972		
AND O OR	29	27	1100001G20Rik	1.35	3,445	2,550	1.000	0.455	0.000	0.284	1.000	4.669	5.103	6.454	0.000	0.000	4.830		
· AND O OK	30	28	1110001A16Rik	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	1110001J03Rik	0.843	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		
Gene Category Settings	32	30	1110002L01Rik	1.00	0.969	0.960	1.209	0.040	0.047	0.063	0.189	7.980	7.993	7.934	8.611	8.885	8.080	-	
Gene category settings		C P.	Data inf	0 (+)		1.447	1.314	0.045	u.giu	0.021	0.117	4	4.173	4.001	2.005	9.459	9.451 F		

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.

	1	A	В	C	D	E	F	G	Н	1	J	К	L	M	N	0
* X	1	Filter: 259			Fold ch	ange			p-val	ue			Average of	Normalized R	C (log2)	
View All Data	2	ID 🛒	Gene Symbo*	A/Conti 🔆	B/Conti 🔅	B/A 🖕	D/C 📮	A/Contr 📜	B/Contr 📜	B/A	D/C 🖕	Control 🖕	A 🗸	в 🛫	C 🖵	D
Gene Category	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
✓ Aging	2094	2092	Adm	0.587	0.722	1.230	0.232	0.057	0.262	0.049	0.061	8.505	7.737	8.035	5.110	3.000
Angiogenesis	2108	2106	Adra1a	1.097	0.846	0.771	0.873	0.286	0.023	0.040	0.011	11.544	11.678	11.304	3.899	3.702
	2116	2114	Adrb3	0.935	0.989	1.058		0.819	0.944	0.042	0.040	9.378	9.281	9.363	11.878	9.298
Apoptotic process	2181	2179	Agt	1.169	0.891	0.762	0.222	0.497	0.044	0.324	0.111	5.895	6.120	5.728	2.672	0.498
Cell cycle	2183	2181	Agtr1a	0.923	0.557	0.604	5.280	0.432	0.173	0.219	0.074	10.984	10.868	10.139	4.796	7.190
	2308	2306	Akt1	0.988	1.040	1.053	0.987	0.802	0.000	0.044	0.048	14.471	14.452	14.527	12.348	12.330
Cell death	2328	2326	Aldh3a1	0.632	0.552	0.874	0.260	0.110	0.009	0.045	0.107	12.316	11.653	11.460	8.772	6.82
Cell differentiation	2367	2365	Alox12	0.682	0.737	1.081	0.852	0.032	0.140	0.712	0.293	9.203	8.652	8.764	8.198	7.96
	2402	2400	Amfr	1.000	0.977	0.977	0.469	0.000	0.088	0.001	0.037	13.080	13.080	13.046	11.417	10.324
Cell migration	2403	2401	Amh	0.681	2.794	4.101	1.050	0.523	0.219	0.146	0.017	3.548	2.995	5.031	5.470	5.541
Cell proliferation	2469	2467	Ankle1	1.760	2.715	1.543	0.840	0.217	0.006	0.032	0.035	3.529	4.345	4.970	8.828	8.576
	2592	2590	Apaf1	1.239	1.083	0.875	0.304	0.187	0.040	0.440	0.008	10.751	11.060	10.867	13.604	11.88
DNA repair	2605	2603	Apex1	0.965	1.292	1.339	1.203	0.336	0.210	0.173	0.189	9.431	9.379	9.801	13.370	13.63
Extracellular matrix	2634	2632	Apod	0.640	0.932	1.456	0.294	0.037	0.049	0.029	0.022	9.585	8.941	9.483	10.581	8.81
	2698	2696	Arg1	1.384	0.774	0.560	3.447	0.234	0.478	0.205	0.500	5.052	5.520	4.683	0.000	1.785
Immune response	2957	2955	Atg7	1.067	1.040	0.975	1.024	0.157	0.796	0.026	0.025	12.788	12.882	12.845	12.001	12.036
Inflammation, response	2965	2963	Atm	1.150	0.942	0.819	1.038	0.206	0.041	0.268	0.476	9.501	9.703	9.415	11.610	11.66
	2967	2965	Atn1	1.124	0.866	0.771	1.320	0.168	0.037	0.324	0.179	13.817	13.985	13.609	14.134	14.53
Neurogenesis	2995	2993	Atp2b1	1.013	0.824	0.814	0.959	0.045	0.000	0.037	0.033	12.416	12.434	12.137	11.775	11.715
PNA colicing	3049	3047	Atp8a2	1.283	1.022	0.796	1.182	0.194	0.856	0.044	0.042	5.129	5.488	5.160	5.078	5.319
KNA splicing	3061	3059	Atr	0.978	0.898	0.918	0.942	0.047	0.133	0.181	0.319	9.364	9.332	9.208	10.770	10.684
Secretion	3100	3098	Aurkb	1.071	1.368	1.277	1.131	0.026	0.167	0.058	0.152	7.273	7.373	7.725	13.114	13.29
and another	3238	3236	Bak1	1.046	1.070	1.023	1.311	0.586	0.914	0.048	0.068	11.775	11.839	11.872	14.094	14.485
	3267	3265	Bbc3	0.994	1.528	1.538	0.235	0.002	0.022	0.000	0.025	9.330	9.321	9.942	12.002	9.914
	3401	3399	Bcl2	1.185	0.871	0.735	0.900	0.187	0.082	0.044	0.044	10.477	10.722	10.277	6.259	6.10
• AND O OR	3436	3434	Becn1	0.991	1.083	1.093	0.744	0.773	0.026	0.046	0.169	10.938	10.925	11.053	10.723	10.29
	3559	3557	Brca2	1.166	0.934	0.801	1.092	0.368	0.624	0.326	0.054	8.479	8.700	8.380	14.476	14.60
	3759	15929	Pck1	0.391	1.248	3.194	3,466	0.286	0.042	0.046	0.214	9.897	8.542	10.217	0.602	2.39
Gene Category Settings	3803	3801	Cacybo	0.810	1.226	1.513	1.309	0.017	0.044	0.201	0.124	8.967	8.664	9.261	11.525	11.917

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 GO를 추가할 수 있다(그림 1-4). '?' 버튼을 누르면 GO 추가하는 방법이 자세히 설명되어 있다.

• x		Gene Category Settings		· · · · · · · · · · · · · · · · · · ·	н	1	J	К	L	N
	1				alized R			Normalized	IRC (log2)	
View All Data	2	Aging	^	Custom Category	c	• A1 •	A2 💌	B1 💌	B2 💌	c
Cana Catagony	178			New	12.7	57 12.349	12.710	12.717	12.668	12
Gene Category	202				3.1	4.350	4.125	3.672	4.273	2
AND OR	242	Apoptotic process		Delete	11.5	20 11.088	11.273	11.088	11.046	11
	252			E alta	8.2	50 8.538	8.433	7.950	8.399	8
^	270	Cell cycle		Edit	12.1	12.571	13.050	12.468	13.196	11
Aging	350				13.7	13.802	14.558	12.890	13.904	13
Angiogenesis	355	Cell death		QuickGO	14.6	79 14.230	14.218	14.706	14.284	14
	362	Call differentiation		Import	11.8	70 11.971	12.371	11.585	12.212	11
Apoptotic process	401				14.3	39 13.792	13.545	14.170	13.556	14
	458	Cell migration		Web ?	13.2	75 13.218	13.493	13.070	13.379	13
Cell cycle	473				10.1	9.983	8.603	10.979	10.112	10
	585	Cell proliferation			3.9	4.582	4.359	3.261	2.836	4
Cell death	598				9.1	9.191	9.354	8.615	9.688	8
Cell differentiation		DNA repair			1.0	3.902	3.626	5.252	5.417	1
		Extracellular matrix			10.5	37 10.404	10.806	10.684	11.137	10
Cell migration	173				7.0	09 7.001	7.006	6.358	6.137	6
	864	Immune response			11.6	54 12.093	12.001	11.653	12.085	11
Cell proliferation	884				9.1	45 9.193	7.948	9.114	8.363	8
	904	Inflammatory response			10.6	75 10.229	9.676	10.548	10.039	10
	1003				14.1	32 15.003	15.164	14.601	15.237	13
Gene Category Settings	1066	Neurogenesis	\sim		10.2	10.468	10.405	10.700	10.397	10
	1108				2.0	0.000	2.025	0.028	0.009	2

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

Gene Category Settings	×	▼ ×		А	В	С	D
			1	Filter: 457	F	old change	
Cell cycle	Custom Category	View All Data	2	Gene symbol	A/C 🔽	B/C 🔽	B/A
Cell death	New		645	COL9A2	0.288	0.300	0.28
		Gene Category	661	EDN2	0.377	0.152	0.37
Cell differentiation	Delete		723	PTCH2	0.494	0.572	0.49
	Edit	C AND C OK	966	NEXN-AS1	0.435	1.108	0.43
Cell migration	Luit		974	IFI44	2.001	5.216	2.00
Cell proliferation	이바이오젠	DNA ranair	975	ADGRL4	2.490	1.986	2.49
			1237	PTPN22	23.460	17.738	23.46
DNA repair	유전자 그룹 설정이 변경 되었습니다. 엑셀 파일을 즉시 저장 하시겠습니까?	Extracellular matrix	1331	TXNIP	2.506	3.261	2.50
			1370	LOC101060524	6.699	2.311	6.69
Extracellular matrix		Immune response	1371	DRD5P2	6.699	2.311	6.69
	예(Y) 아니요(N)		1436	PSMD4	2.018	0.991	2.01
Immune response		Inflammatory response	1721	LOC101928372	0.470	0.338	0.47
Inflammatory recoonse		Neuropeania	1999	LHX9	0.355	1.667	0.35
		Neurogenesis	2037	PTPN7	0.478	0.568	0.47
Neurogenesis		RNA splicing	2107	SLC26A9	0.468	0.764	0.46
			2271	WNT3A	0.268	0.518	0.26
RNA splicing	A 4	Secretion	2410	NLRP3	2.629	1.474	2.62
			2486	LINC00704	2.777	3.115	2.77
Secretion		New category	2490	AKR1C1	2.498	0.923	2.49
		*	2506	IL15RA	2.450	1.717	2.45
I New category	\checkmark	Gene Category Settings	2562	OLAH	2.323	1.012	2.32
		Gene Category Settings	2717	CXCL12	0.498	4.046	0.49

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 95개의 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.

1	A	В	С	D	E	F	G	н	I	J	К	L	M	N	0	P 🔺				
Fi	lter: 95			Fold cf	ange			p-val	ue			Average of	Normalized RC	(log2)						*
,	ID J	Gene Symbo'	A/Conti 🔆	B/Conti 🔅	B/A	D/C 🚽	A/Contr 👙	B/Contr 📜	B/A	D/C	Control	A	в	c 📮	D 🖵	Control:		D	DEG Analy	rsis
6	15849	Pcdh10		1.002	21.807	0.139	0.031	0.000	0.093	0.123	4.510	0.066	4.513	6.215	3.367	3.14	0.			
)	5476	Cyp2b9	i ilitik	0.925	26-970	1.000	0.024	0.849	0.027	1.000	5.821	0.956	5.709	0.000	0.000	4.21	Sig	nifican	nt Gene Se	lection
5	5519	Cyp3a11	91556	0.121.	1.979	1.000	0.044	0.036	0.044	1.000	4.150	0.000	0.985	0.000	0.000	5.10				
5	13488	Mup17	0.348	and the second	0,225	0.414	0.002	0.001	0.003	0.500	4.129	2.606	0.456	1.271	0.000	4.29	Fo	ld cha	inge	
1	8470	Gm14812	0.356	0.276	0.776	0.493	0.044	0.091	0.658	0.500	4.086	2.596	2.230	1.020	0.000	4.49			2.0	
1	423	1700102P08Rik	0.356	1.370	3.846	0.894	0.049	0.040	0.046	0.590	4.439	2.950	4.894	6.295	6.134	4.83				
	9256	Gm9758	0.219	1.022	4.673	0.879	0.016	0.798	0.117	0.500	5.096	2.902	5.127	0.186	0.000	4.36	-Nr	ormali	zed RC (l	0021-
1	10288	Htr3a	0.399	1.567	3.927	0.845	0.000	0.000	0.000	0.024	4.628	3.302	5.276	6.692	6.449	0.00	145	////una	rea ne ti	'gu/
1	20975	Tmem202	0.417	0.745	1.786	1.000	0.044	0.115	0.454	1.000	4.669	3.406	4.243	0.000	0.000	5.87			4	
	8350	Gm13279	0.358	0.484	1.350		0.006	0.186	0.047	0.493	5.828	4.347	4.780	2.757	0.000	5.92				
	712	2700086A05Rik	0.440	0.528	1.200	1.076	0.042	0.055	0.040	0.005	4.508	3.325	3.588	2.193	2.299	4.61	p-1	value		
	4186	Ccno	0.369	0.272	0.738	0.304	0.012	0.054	0.352	0.122	9.061	7.623	7.185	6.042	4.324	9.42		1	0.050	
2	174	1700012B09Rik	0.378	1.331	3,522	0.411	0.020	0.950	0.259	0.506	6.340	4.936	6.752	2.583	1.301	6.60				
1	21203	Tnni1	0.446	1.222	2.742		0.022	0.836	0.038	0.285	4.862	3.695	5.150	6.083	3.286	4.77	. 52	mole	/ Control	
	8740	Gm20878	0.390	2.689	6.895	0.765	0.034	0.027	0.026	0.014	13.665	12.306	15.092	8.479	8.093	12.42	30	mpie /	/ control	
	8755	Gm21586	0.390	2.689	6.895	0.765	0.006	0.801	0.007	0.029	13.665	12.306	15.092	8.479	8.093	12.42			Name	3
	8364	Gm13308	0.390	2.688	6.892	0.765	0.032	0.257	0.046	0.000	13.665	12.307	15.092	8.479	8.093	12.43		7 .	10 mm -	
	6419	E330023G01Rik	0.411	0.431	1.049	0.192	0.050	0.618	0.039	0.216	5.240	3.955	4.024	6.235	3.857	1.55		Y A	/Control	
	17019	Prss16	0.463	0.364	0.787	0.681	0.037	0.125	0.116	0.253	4.023	2.912	2.567	8.626	8.073	0.97		B	/Control	
	2847	Art5	0.411	1.020	2.482	0.330	0.009	0.050	0.295	0.170	5.219	3.936	5.247	7.485	5.885	5.21	1	-		
1	4938	Cntf	0.440	0.475	1.079	0.734	0.005	0.139	0.044	0.221	5.649	4.464	4.574	6.695	6.249	5.69		В,	/A	
T.	8353	Gm13285	0.445	0.281	0.632		0.032	0.011	0.146	0.500	7.408	6.240	5.579	4.665	0.000	7.33	Ī	D	/C	
1	4422	Cdkn1a	0.445	0.640	1.437	0.279	0.019	0.117	0.198	0.017	14.318	13.151	13.673	14.296	12.453	14.15				
)	7599	Foxn4	0.456	0.213	0.468	0.882	0.029	0.008	0.046	0.131	6.674	5.541	4.446	10.475	10.293	6.99				
	19817	Spon2	0,457	1.025	2.242	0.347	0.029	0.938	0.033	0.031	10.478	9.349	10.514	8.538	5.774	10.03				
	9701	Gucy2g	0.465	0.886	1.905		0.047	0.048	0.044	0.098	5.455	4.350	5.280	6.252	3.714	5.18				
	21929	Upk1a	0.467	0.390	0.835	0.884	0.030	0.008	0.172	0.449	7.923	6.825	6.564	8.386	8.208	8.12				
	600	2310039L15Rik	0.484	1.455	3.005	0.637	0.048	0.093	0.000	0.477	4.158	3.111	4.699	5.016	4.364	5.52		• A	ND O	OR
1	4152	Ccl27b	0.472	1.988	4.210	0.850	0.014	0.001	0.019	0.399	14.414	13.331	15.405	10.608	10.374	13.19				
ĩ	8363	Gm13306	0.472	1.988	4.210	0.850	0.029	0.185	0.063	0.399	14.414	13.331	15.405	10.608	10.374	13.19				
3	11637	I OC100861978	0.472	1.988	4,210	0.850	0.038	0.038	0.032	0.399	14.414	13.331	15.405	10.608	10.374	13.19 *		-		
	-	Data info	(+)								4					Þ	Filte	er Gen	le catego	ry Cr

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.

- ~	- M	A	В	с	D	E	F	G	н	T	J	к	L	M	N	0	P 🔺	
* ^	1 61	lter: 10			Fold ch	ange			p-val	ue			Average of	f Normalized R	C (log2)			
View All Data	2	ID J	Gene Symbo'	A/Conti 🕽	B/Conti 👙	B/A	D/C 🖕	A/Contr 📜	B/Contr 👙	B/A	D/C	Control 🖕	A	в	c 🖕	D	Control:	DEG Analysis
Gana Category	2909	4186 C	cno	0.369	0.272	0.738	0.304	0.012	0.054	0.352	0.122	9.061	7.623	7.185	6.042	4.324	9.4	
Gene category	4840	4938 C	intf	0,440	0.475	1.079	0.734	0.005	0.139	0.044	0.221	5.649	4.464	4.574	6.695	6.249	5.6	· · · · · · · · · · · · · · · · · · ·
Aging	4964	4422 C	dkn1a	0.445	0.640	1.437	0.279	0.019	0.117	0.198	0.017	14.318	13.151	13.673	14.296	12.453	14.1	Fald shares
Angiogenesis	5090	7599 F	oxn4	0.456	0.213	0.468	0.882	0.029	0.008	0.046	0.131	6.674	5.541	4.446	10.475	10.293	6.9	Polo change
	5475	21929 U	Jpk1a	0.467	0.390	0.835	0.884	0.030	0.008	0.172	0.449	7.923	6.825	6.564	8.386	8.208	8.1	2.0 🗘
Apoptotic process	7035	22726 X	llr5a	0.492	0.161	0.326	0:192	0.022	0.027	0.028	0.040	4.578	3.556	1.941	2.384	0.000	4.0	
Cell cycle	92/1	16085 P	er2	2.052	0.972	0.474	0.899	0.045	0.896	0.044	0.416	10.990	12.027	10.948	8.523	8.369	11.3	Normalized RC (log2)
Call death	9047	3243 8	lamit	2.122	1.745	0.822	2.073	0.035	0.003	0.042	0.355	3.515	12 257	0.318	0.755	4.961	3.4	4 🗘
Cell deall	17755	21189 T	ofefe	2,502	2 924	0.995	1.000	0.032	0.333	0.004	1.000	4 926	6 500	6 324	0.000	0.000	11.7	
 Cell differentiation 	23423		maio	No. AND		0.005	1.000	0.010	0.2.24	0.035	1.000	4.020	0.300	0.524	0.000	0.000	4.2	p-value
Cell migration	23424																	0.050
Cell proliferation	23425																	
DNA renair	23420																	Sample / Control
	23428										-							Name
Extracellular matrix	23429																	
Immune response	23430																	A/Control
laftermeters recently	23431																	B/Control
Innaminatory response	23432																	P/A
Neurogenesis	23433																	0/A
RNA solicing	23434																	D/C
	23435																	
Secretion	23430																	
cell growth	23437										-							
	23439										-							
AND OR	23440																	AND O OP
	23441																	C AND C ON
	23442																	
Gene Category Settings	23443		14														Ψ.	51
	4 1	D	ata info	(+)								(P.	Filter Gene Category Chart

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 Selection 의 비교 그룹을 선택하고 "View Gene Category Chart"를 클릭하면 증가/감소한 유전자 들 대상으로 GO Chart 가 생성된다. 그래프의 각 영역을 클릭하면 해당 유전자들이 필터링 된다. 예를 들어 왼쪽의 Pie chart 의 특정영역을 클릭하면 해당 GO 의 증가/감소된 유전자가 함께 필터링 되고 오른쪽의 증가/감소된 bar chart 에서 bar 상단의 숫자는 해당 유전자 수이고 bar 를 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 1-9 와 같이 Scatter Plot, Volcano Plot, Venn Diagram 을 엑셀에서 쉽게 그릴 수 있다.

	DEG Analysis
Signi	ficant Gene Selection
Anal	ysis Graph
	Scatter Plot
	Volcano Plot
	Venn Diagram

그림 1-9. Analysis Graph Tool

첫번째 Scatter Plot은 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 spot을 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 ID를 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 2개, 3개 또는 4개 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 Heatmap에 표현할 비교그룹을 체크하여 "Data Export"를 클릭한 후 "???.txt"로 저장하면 된다.

		A	В	C	D	E	F	G	н	1	Clustering Heatmap Support	м	N	0	P 🔺	
* X	1	Filter: 64			Fold ch	ange			p-val	ue	O close ing reaching support	of Normalized R	(1082)			*
View All Data	2	ID J	Gene Symbo*	A/Conti	B/Conti 👙	B/A 🕌	D/C	A/Contr 👙	B/Contr 🖕	B/A	Туре	в	c 🗸	D	Control:	DEG Analysis
Sene Category	179	1792	Abcb1a	0.956	0.798	0.835		0.008	0.239	0.14	Foldchange	10.914	11.582	8.003	11.3	
circ category	253-	16330	Pitx2	0.334	0.484	1.450	0.340	0.085	0.352	0.03		4.679	9.506	7.949	5.5	Clustering Heatmap Suppo
Aging	255	2551	Anxa7	0.997	0.994	0.996	0.364	0.043	0.326	0.11	Raw Data (RC)	14.142	14.419	12.961	14.2	
Angiogenesis	340	3404	Bcl2l1	0.802	0.989	1.233	0.301	0.172	0.852	0.04		12.839	12.485	10.754	13.0	Туре
Angiogenesis	345	3456	Bglap2	0.709	1.186	1.672	A.m.	0.014	0.075	0.03	• Z-Score	4.572	5.958	0.922	5.4	Foldshange
Apoptotic process	350	3507	Bmp7	0.868	0.851	0.980	0.248	0.479	0.400	0.04	O Madian	7.551	10.560	8.551	7.7	© Poluciange
Cell cycle	390	3901	Card11	1.604	1.444	0.900		0.055	0.083	0.03	O Wedian	10.679	4.941	2.086	9.9	Raw Data (RC)
	423	4231	Cd180	1.481	1.279	0.864		0.046	0.917	0.03		9.144	5.385	2.233	8.6	
Cell death	427-	4272	Cd34	0.966	0.821	0.850	0.280	0.483	0.378	0.44		13.482	9.806	7.970	13.7	• Z-Score
Call differentiation	430	4301	Cd74	1.368	1.368	1.000	0.243	0.070	0.341	0.04	Export Data Select	18.637	13.831	11.790	17.9	O Median
	473	4737	Cks1b	1.071	2.104	1.963	1.060	0.047	0.021	0.07		8.073	11.844	11.928	7.2	
Cell migration	486	4859	Clu	0.846	0.784	0.927		0.240	0.067	0.03	Name	13.747	12.194	10.204	14.2	
Coll proliferation	532	5318	Ctgf	1.209	0.840	0.695		0.429	0.044	0.24	A/Control	12.994	12.352	9.532	13.3	Export Data Select
Cell promeration	578	5786	Ddit4	0.974	1.177	1.209		0.012	0.028	0.09	Avconaor	9.920	11.008	7.720	8.8	Nama
DNA repair	605	6050	DIx5	0.491	2.239	4.563		0.223	0.039	0.31	B/Control	5.726	7.699	5.161	4.5	LINdille
Extracollular matrix	654	6543	Egfr	0.886	0.691	0.781		0.441	0.206	0.34		10.863	10.452	6.652	11.5	A/Control
	677.	6770	Ephb1	1.040	2.002	1.925	1.141	0.001	0.040	0.00	B/A	7.122	2.931	3.121	6.0	R/Control
Immune response	747	7476	Fign11	1.309	2.002	1.530	1.019	0.119	0.008	0.05		7.060	12.793	12.819	6.0	B/Control
Inflammatony rosponso	766	7659	Fst	0.688	0.653	0.950		0.220	0.192	0.04	D/C	7.947	10.499	6.936	8.9	B/A
Innammatory response	803	8031	Gli2	1.313	1.138	0.867	2.171	0.145	0.821	0.03		8.845	13.873	14.991	8.6	10.0%
Neurogenesis	835	8355	Gm13287	0.506	0.330	0.653		0.085	0.005	0.19		3.116	2.782	0.000	4.7	U/C
DNA	928	9286	Gmnc	0.630	0.370	0.587		0.199	0.035	0.23		6.209	3.861	0.000	7.6	
KINA splicing	952	9523	Gpx1	1.026	1.150	1.121	0.378	0.461	0.151	0.23		14.278	15.515	14.112	14.0	
Secretion	988	9879	Hdgfrp3	1.066	1.020	0.957		0.166	0.801	0.04		9.842	9.814	7.461	9.8	
	994	9947	Hhex	1.048	0.849	0.811	0.465	0.038	0.031	0.05		10.897	12.621	11.516	11.0	
cell growth	995	9948	Hhip	1.153	0.833	0.722	0.279	0.122	0.032	0.21		10.540	8.910	7.068	10.8	
	1034	1 10339	Id2	1.263	1.234	0.977		0.024	0.846	0.72		12.934	11.246	8.868	12.7	
AND OR	1070	2 10700	Irs2	1.237	0.827	0.669	0.316	0.087	0.469	0.26		10.963	11.489	9.828	11.3	
	1074	9 10747	Itgam	1.414	1.336	0.945		0.004	0.047	0.66	and the second second	11.793	8.114	5.945	11.4	
	1075	5 10753	Itgb2	1.530	1.674	1.094		0.008	0.073	0.66	Data Export	12.840	6.906	3.178	11.9	Data Export
Gene Category Settings	1077	8 10776	Itokb	1.037	0.959	0.925	0.193	0.513	0.044	0.74	100111 17.117 17.11	19.717	9.656	7.286	12.7 *	
			Data info	(+)							1 4				•	0

그림 1-16. Clustering Heatmap Support

7. Selected Gene Plot & Gene Search 사용 방법

ExDEGA 의 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

- A	Α	В	с	D	E	F	G	н	1	J	K	L	M	N	0	Р	Q	R	S	
1	Filter: 36			Fold ch	nange			p-va	lue		A	verage of	Normalized	RC (log2)						* ×
2	IC ,7	Gene symb 🖕	B /Cor	E232-25 /Con *	E305-100 /Con *	LPS /Co 🗸	B/Co 🗸	E232-25 /Cor *	E305-100 /Con *	LPS /C	Con 👻	8 💌	E232-2*	E305-1**	LPS 🚽	Con-	B831-	232-2	E305-100- 1 *	DEG Analysis
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	<u> </u>
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	Significant Gene Selection
458	456	Irs1	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	 Analysis Graph
1999	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	\frown
2484	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	Clustering Heatmap Support
2485	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	
3535	3533	lgf2bp1	1.487	9.216	0.705	14:831	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	Selected Gene Plot (ID input)
635	3633	Igfbp4	3.957	Director C.	0.511	0.445	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	() court court
361	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	2.883	2.577	2.098	1.296	Gene search
417	7415	lgfbp6	1.422	4.096	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	
702	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	insulin
3264	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	
\$472	8470	Igfals	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	Search X
0355	10353	Insl6	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	Jearen
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	0.415	0.825	0.414	0.414	0.413	1.278	0.996	0.016	0.013	0.010	1.929	1.580	0.000	0.000	
2094	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	0.655	0.924	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	

그림 1-18. Genes related to insulin