Version 3.0

RNA-Seq/Microarray DEG Analysis



(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. 엑셀기반 DEG 분석 (ExDEGA v.1.6.0)

2. Web 기반 Gene Set Enrichment 분석

- 2-1. DAVID tool을 이용한 Functional Annotation 분석
- 2-2. String-db tool을 이용한 gene set분석
- 2-3. MSigDB기반 GSEA 분석
- 3. KEGG DB 기반 Pathway 분석
- 4. MeV Software 이용 Clustering Heatmap 작성

1. 엑셀기반 DEG 분석 (ExDEGA v.1.6.0)

㈜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

??? 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.



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

1-1. 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 15개를 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.

| * X | | A | В | C | D | E | F | G | н | | J | K | L | M | N | 0 |
|------------------------|------|-------|-------------|-----------|-----------|-------|-------|-----------|-----------|-------|-------|-----------|---------|-----------------|----------|-------|
| | 1 | 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 |
| Casa Catagoni | 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 | 10.23 |
| Gene Category | 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 | 11.44 |
| ✓ 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 | 3.00 |
| 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 | 3.70 |
| | 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 | 9.29 |
| 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 | 0.49 |
| Cell orcle | 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 | 7.19 |
| | 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 | 12.33 |
| 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 | 13.045 | 11.417 | 10.32 |
| 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 | 5.54 |
| 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 | 8.57 |
| | 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 | 1.78 |
| 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 | 12.03 |
| Inflammatory 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 | 11.71 |
| RNA splicing | 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 | 5.31 |
| | 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 | 10.68 |
| 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 |
| cell growth | 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 | 14.48 |
| | 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 | 9.91 |
| AND OD | 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 |
| IND 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 | Carvhn | 0.810 | 1.226 | 1.513 | 1.309 | 0.017 | 0.044 | 0.201 | 0.124 | 8.967 | 8.664 | 9.261 | 11.525 | 11.91 |

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.

| ▼ × | | Gene Category Settings | | · · · · · · · · · · · · · · · · · · · | н | - I | J | К | L | N |
|------------------------|------|------------------------|--------|---------------------------------------|-----------|-----------|--------|------------|-------------|----|
| | 1 | | | | alized RC | | | Normalized | d RC (log2) | |
| View All Data | 2 | Aging | ^ | Custom Category |) C | • A1 • | A2 💌 | B1 💌 | B2 💌 | c |
| Cons Catogony | 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 |
| O AND O OR | 242 | Apoptotic process | | Delete | 11.5 | 11.088 | 11.273 | 11.088 | 11.046 | 11 |
| | 252 | | | Edit | 8.2 | 50 8.538 | 8.433 | 7.950 | 8.399 | 8 |
| ^ | 270 | Cell cycle | | Luit | 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 | | | 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 | 7.001 | 7.006 | 6.358 | 6.137 | 6 |
| | 864 | Immune response | | | 11.6 | 12.093 | 12.001 | 11.653 | 12.085 | 11 |
| Cell proliferation | 884 | | | | 9.1 | 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 | X | ▼ × | | 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 | 0 | 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 O'CK | 966 | NEXN-AS1 | 0.435 | 1.108 | 0.43 |
| Cell migration | | | 974 | IFI44 | 2.001 | 5.216 | 2.00 |
| Cell proliferation | 이바이오젠 | | 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 |
| | 에(위) 아디보(N) | | 1436 | PSMD4 | 2.018 | 0.991 | 2.01 |
| | | Inflammatory response | 1721 | LOC101928372 | 0.470 | 0.338 | 0.47 |
| Inflammatory response | | Neurogenesis | 1999 | LHX9 | 0.355 | 1.667 | 0.35 |
| | | | 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 | | Secretion | 2410 | NLRP3 | 2.629 | 1.474 | 2.62 |
| Garretian | | | 2486 | LINC00704 | 2.777 | 3.115 | 2.77 |
| | | New category | 2490 | AKR1C1 | 2.498 | 0.923 | 2.49 |
| New category | | | 2506 | IL15RA | 2.450 | 1.717 | 2.45 |
| | V | Gene Category Settings | 2562 | OLAH | 2.323 | 1.012 | 2.32 |
| | | | 2717 | CXCL12 | 0.498 | 4.046 | 0.49 |

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

1-2. 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.

| | A | В | С | D | E | F | G | н | I | J | К | L | M | N | 0 | P | | | |
|------|------------|---------------|-----------|-----------------|--------|---------|-----------|-----------|-------|-------|---------|------------|---------------|--------|--------|----------|--------|---------------|-----------|
| 1 | Filter: 95 | | | Fold cf | iange | | | p-val | Je | | | Average of | Normalized RC | (log2) | | | | | * |
| 2 | ID ,T | 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 Analy | ysis |
| 76 | 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 | | |
| 80 | 5476 | Cyp2b9 | i actor | 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 | Signit | hcant Gene Se | ection |
| 176 | 5519 | Cyp3a11 | 0.006 | 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 | | | |
| 895 | 13488 | Mup17 | 0.348 | diam'n a starte | 0,225 | 0.414 | 0.002 | 0.001 | 0.003 | 0.500 | 4.129 | 2.606 | 0.456 | 1.271 | 0.000 | 4.29 | Fold | change | |
| 1024 | 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 | ÷ |
| 1079 | 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 | | | |
| 1407 | 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 | Nor | nalized RC (I | 002) |
| 1636 | 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 | 1401 | manized ree (| oge/ |
| 1913 | 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 | • |
| 2845 | 8350 | Gm13279 | 0.358 | 0.484 | 1.350 | 0.148 | 0.006 | 0.186 | 0.047 | 0.493 | 5.828 | 4.347 | 4.780 | 2.757 | 0.000 | 5.92 | | | |
| 2849 | 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-va | lue | |
| 2909 | 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 | | 0.050 | |
| 3124 | 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 | | 010.00 | |
| 3247 | 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 | Com | pla / Control | |
| 3510 | 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 | Salli | pie / control | |
| 3655 | 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 | e |
| 3718 | 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 | 177 | | |
| 4097 | 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 | ~ | A/Control | S |
| 4154 | 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 | |
| 4188 | 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 | 100 | D.(A | |
| 4840 | 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 | | B/A | |
| 4940 | 8353 | Gm13285 | 0.445 | 0.281 | 0.632 | dames - | 0.032 | 0.011 | 0.146 | 0.500 | 7.408 | 6.240 | 5.579 | 4.665 | 0.000 | 7.33 | | D/C | |
| 4964 | 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 | | | |
| 5090 | 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 | | | |
| 5115 | 19817 | Spon2 | 0.457 | 1.025 | 2.242 | 0.147 | 0.029 | 0.938 | 0.033 | 0.031 | 10.478 | 9.349 | 10.514 | 8.538 | 5.774 | 10.03 | | | |
| 5412 | 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 | | | |
| 5475 | 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 | | | |
| 5478 | 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 | (| AND O | OR |
| 5974 | 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 | | | |
| 6038 | 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 | | | |
| 6328 | 11637 | 00100861978 | 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 * | | | |
| 4 | je - | Data info | + | | | | | | | | 4 | | | | | Þ. | Filter | Gene Catego | bry Chart |

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 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

1-3. 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 |
|---------|------------------------------|
| • |) Significant Gene Selection |
| \odot |) Analysis 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 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

1-4. 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 Select에서 Heatmap에 표현할 비교그룹을 체크하여 "Data Export"를 클릭한 후 "???.txt"로 저장하면 된다.

| | 1 | A | В | C | D | E | F | G | н | 1 7 | Clustering Heatmap Support | м | N | 0 | P . | _ |
|------------------------|-------|------------|-------------|-----------|---------|-------|---------|-----------|-----------|------|----------------------------|------------------|--------|--------|----------|----------------------------|
| * × | 1 | Filter: 64 | | | Fold ch | ange | | | p-val | lue | | of Normalized RC | (log2) | | | |
| View All Data | 2 | ID J | Gene Symbo* | A/Conti 🗧 | B/Conti | в/А 🖕 | D/C | A/Contr 🗧 | B/Contr 🖕 | B/A | Туре | в | C 🔍 | D | Control | DEG Analysis |
| Gene Category | 1794 | 1792 | Abcb1a | 0.956 | 0.798 | 0.835 | | 0.008 | 0.239 | 0.14 | Foldchange | 10.914 | 11.582 | 8.003 | 11.3 | |
| Gene category | 2534 | 16330 | Pitx2 | 0.334 | 0.484 | 1.450 | 0.340 | 0.085 | 0.352 | 0.03 | Version of the second | 4.679 | 9.506 | 7.949 | 5.5 | Clustering Heatmap Support |
| Aging | 2553 | 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 | 3406 | 3404 | Bcl2l1 | 0.802 | 0.989 | 1.233 | 0.301 | 0.172 | 0.852 | 0.04 | | 12.839 | 12.485 | 10.754 | 13.0 | Type |
| Anglogenesis | 3458 | 3456 | Bglap2 | 0.709 | 1.186 | 1.672 | - Admir | 0.014 | 0.075 | 0.03 | • Z-Score | 4.572 | 5.958 | 0.922 | 5.4 | Eoldshange |
| Apoptotic process | 3509 | 3507 | Bmp7 | 0.868 | 0.851 | 0.980 | | 0.479 | 0.400 | 0.04 | O Madian | 7.551 | 10.560 | 8.551 | 7.7 | © Polucitarige |
| Cell cycle | 3903 | 3901 | Card11 | 1.604 | 1.444 | 0.900 | | 0.055 | 0.083 | 0.03 | O ivieulari | 10.679 | 4.941 | 2.086 | 9.9 | C Raw Data (RC) |
| | 4233 | 4231 | Cd180 | 1.481 | 1.279 | 0.864 | | 0.046 | 0.917 | 0.03 | | 9.144 | 5.385 | 2.233 | 8.6 | 0 |
| Cell death | 4274 | 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 |
| Cell differentiation | 4303 | 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 |
| | 4739 | 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 | 4861 | 4859 | Clu | 0.846 | 0.784 | 0.927 | | 0.240 | 0.067 | 0.03 | Name | 13.747 | 12.194 | 10.204 | 14.2 | |
| Call proliferation | 5320 | 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 | 5788 | 5786 | Ddit4 | 0.974 | 1.177 | 1.209 | | 0.012 | 0.028 | 0.09 | Avcontrol | 9.920 | 11.008 | 7.720 | 8.8 | Nama |
| DNA repair | 6052 | 6050 | DIx5 | 0.491 | 2.239 | 4.563 | | 0.223 | 0.039 | 0.31 | B/Control | 5.726 | 7.699 | 5.161 | 4.5 | |
| Extracollular matrix | 6545 | 6543 | Egfr | 0.886 | 0.691 | 0.781 | | 0.441 | 0.206 | 0.34 | | 10.863 | 10.452 | 6.652 | 11.5 | A/Control |
| | 6772 | 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 | 7478 | 7476 | Fign11 | 1.309 | 2.002 | 1.530 | 1.019 | 0.119 | 0.008 | 0.05 | | 7.060 | 12.793 | 12.819 | 6.0 | U byconnor |
| Inflammatony recoonse | 7661 | 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 |
| | 8033 | 8031 | GI12 | 1.313 | 1.138 | 0.867 | 2.171 | 0.145 | 0.821 | 0.03 | | 8.845 | 13.873 | 14.991 | 8.6 | D/C |
| Neurogenesis | 8357 | 8355 | Gm13287 | 0.506 | 0.330 | 0.653 | | 0.086 | 0.005 | 0.19 | | 3.116 | 2.782 | 0.000 | 4.7 | 0/0 |
| PNA rolicino | 9288 | 9286 | Gmnc | 0.630 | 0.370 | 0.587 | | 0.199 | 0.035 | 0.23 | | 6.209 | 3.861 | 0.000 | 7.6 | |
| | 9525 | 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 | 9881 | 9879 | Hdgfrp3 | 1.066 | 1.020 | 0.957 | 0.196 | 0.166 | 0.801 | 0.04 | | 9.842 | 9.814 | 7.461 | 9.8 | |
| Call arouth | 9949 | 9947 | Hhex | 1.048 | 0.849 | 0.811 | 0.465 | 0.038 | 0.031 | 0.05 | | 10.897 | 12.621 | 11.516 | 11.0 | |
| | 9950 | 9948 | Hhip | 1.153 | 0.833 | 0.722 | 0.279 | 0.122 | 0.032 | 0.21 | | 10.540 | 8.910 | 7.068 | 10.8 | |
| S. 110 O 00 | 10341 | 10339 | Id2 | 1.263 | 1.234 | 0.977 | | 0.024 | 0.846 | 0.72 | | 12.934 | 11.246 | 8.868 | 12.7 | |
| • AND O OR | 10702 | 10700 | Irs2 | 1.237 | 0.827 | 0.669 | 0.316 | 0.087 | 0.469 | 0.26 | | 10.963 | 11.489 | 9.828 | 11.3 | |
| | 10749 | 10747 | Itgam | 1.414 | 1.336 | 0.945 | | 0.004 | 0.047 | 0.66 | Data Carat | 11.793 | 8.114 | 5.945 | 11.4 | Data Supert |
| | 10755 | 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 | 10778 | 10776 | Itokb | 1.037 | 0.959 | 0.925 | (1, 193 | 0.513 | 0.044 | 0.24 | INFIT DATA DATA | 19.919 | 9.656 | 7.286 | 12.7 * | |
| | | | Data info | (+) | | | | | | | 1 | | | | F | 0 |

그림 1-16. Clustering Heatmap Support

1-5. 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

| 1 | Α | В | С | D | E | F | G | н | 1 | J | K | L | M | N | 0 | P | Q | R | S | * | | |
|-------|---------------|-------------|--------|-------------------|--------------------|---------|--------|-------------------|--------------------|--------|--------|------------|------------|-------------|--------|--------|--------|--------|------------------|------------|------------------|--------|
| 1 | Filter: 36 | | | Fold ch | nange | | | p-va | lue | | 1 | Average of | Normalized | l 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 👻 | E232-25 | E305-1** | LPS 🖵 | Con-f | B831- | E232-2 | E305-100- 1 * | DE | G 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 | ~ | | |
| 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 | Significan | t Gene Selecti | ion |
| 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 | 0 | | |
| 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 C | iraph | |
| 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 | Chustering | Linetonen Com | |
| 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 | nearmap out | pport |
| 2485 | 2483 | Igfbp3 | 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 | C Calastad | Conc Dict (ID) | inout) |
| \$535 | 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 | U selected | serie Plot (ID I | mputy |
| \$635 | 3633 | Igfbp4 | 3.957 | | 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 | Gene Sea | ch | |
| 1361 | 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 | O dene bed | c.n | |
| 417 | 7415 | Igfbp6 | 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 | | - 1 |
| 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 | Seat | ch | 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 | | •••• | |
| 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

2. Web 기반 Gene Set Enrichment 분석

2-1. DAVID tool을 이용한 Functional Annotation 분석

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



DAVID 에서는 3 천 개 이상의 유전자는 분석할 수 없으므로 3 천 개 이하로 유전자를 선별해야 한다. mRNASeq 결과에서 significant gene 을 선별하여 DAVID 분석을 한다. DAVID 홈페이지 (<u>http://david.abcc.ncifcrf.gov/</u>)에 접속하여 "Functional Annotation"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 에서선별한 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 3 에서 "Gene List"를 체크하고 Step 4 에서 "Submit List"를 누른다. Gene Symbol 을 넣은 경우,"multiple species have been detected in your gene list"라는 창이 뜨면"확인"을 누른다.



그림 2-3. DAVID tool : Step 1 ~ Step 4

실험한 종을 선택하고 "Select Species"를 누르면 해당 종의 유전자를 대상으로 다시 분석된다. 예시에서는 160개의 유전자 리스트를 넣었지만 데이터베이스에서 기능이 밝혀진94개이기에 최종 94개 유전자를 대상으로 Functional Annotation 분석이 완료되었다(그림 2-4).



그림 2-4. DAVID tool : Select Species

분석 결과를 확인하기 위해 예로 Gene Ontology 중 Biological Process 를 확인한다."Gene_Ontology"의 "+" 표시를 클릭하여 결과 창을 열고 "GOTERM_BP_FAT"의 "Chart"를 누르면 94 개 유전자들이 관여하는 Biological Process 에 속하는 GO를 확인할 수 있다(그림 2-5). 관심 GO를 클릭하면 QuickGO 데이터베이스로 연결되어 각 GO의 정보를 확인할 수 있다. GO의 Gene 막대를 클릭하면 해당 GO 관련 유전자들을 확인할 수 있다.

| unctional Categories (3 sele | cted) | | DAVID: Database for J | enotation, visualization, and integrated Discovery | Laboratory of Immunopa - Windows inte | net explorer | |
|------------------------------|-----------|--|------------------------|--|---------------------------------------|--------------|---|
| Sene_Ontology (3 selected) | , | | M nep // avid abce net | dr.gov/cranoleport/sprannots25 | | | - |
| GOTERM_BP_1 | 69.1% 65 | Chart | Sublist Categ | enr d | Iem | C RT Gene | |
| COTERM BR 2 | 69.1% 65 | Chart | GOTERM_BP | FAT cell-matrix adhesion | | R | 5 |
| | | (Chart) | GOTERM_BP | FAT cell-substrate adhesion | | RI F | |
| GOTERM_BP_3 | 68.1% 64 | Criart | GOTERM_BP | FAT cell adhesion | | RT 🚃 | |
| GOTERM_BP_4 | 64.9% 61 | Chart = | GOTERM_BP | FAT biological adhesion | | 1 = | |
| COTERM BR 5 | 59.6% 56 | Chart _ | GOTERM_BP | FAT inflammatory response | | RI = | |
| | | Chart A | GOTERM_BP | _FAT integrin-mediated signaling pathway | | RI = | |
| GOTERM_BP_ALL | 69.1% 65 | Crian | | PAT cellular cation nomeostasis | | | |
| GOTERM_BP_FAT | 63.8% 60 | Chart | GOTERM BE | FAT cation homeostasis | | | |
| GOTERM CC 1 | 73.4% 69 | Chart _ | GOTERM_BP | FAT kidney development | | RI = | |
| | 77 496 60 | Chart | GOTERM_BP | FAT cellular metal ion homeostasis | | RT = | |
| GOTERM_CC_2 | 73.470 05 | | GOTERM_BP | FAT metal ion homeostasis | | BI 🚍 | |
| GOTERM_CC_3 | 73.4% 69 | Chart | GOTERM_BP | FAT urogenital system development | | RT 🔳 | |
| GOTERM_CC_4 | 67.0% 63 | Chart | GOTERM_BP | FAT metanephros development | × × | <u>RI</u> | |
| GOTERM_CC_5 | 62.8% 59 | Gene R | eport | | | | |
| GOTERM CC ALL | 73.4% 69 | | | | | | |
| | (3 AV (4 | Current Ge | ne List: List_1 | | | | |
| GOTERM_CC_FAT | 03.8% 00 | Current Ba | ckground: Hom | o sapiens | | | |
| GOTERM_MF_1 | 71.3% 67 | 94 DAVID | US . | | | | |
| GOTERM MF 2 | 71.3% 67 | 6 record(s |) | | | | |
| 1 | 67 996 60 | OFFICIAL | _GENE_SYMBOL | GEN | ENAME | | R |
| GOTERM_MF_3 | 03.8% 00 | AGT | | angiotensingen (sergin pentidase inhit | ator, clade A. member 8) | RG | |
| | | ITGB18P1 | | integrin beta 1 binding protein 1 | | RG | |
| | | ITGA8 | | integrin, alpha 8 | | RG | |
| | | LAMAS | | laminin, alpha 5 | | RG | |
| | | CONCERNMENT OF THE OWNER OWN | | | | | |

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

이와 같은 방법으로 Pathway 결과를 확인해 보면 KEGG_PATHWAY database 에서 주요 Pathway가 나온다(그림 2-6).각 pathway 를 누르면 pathway 그림을 확인할 수 있다. pathway 그림에서 별 표시가 되어 있는 유전자가 input 유전자(160 개) 중 해당 pathway 에 관여하는 유전자이다. 유전자를 클릭하면 유전자 정보를 자세히 알 수 있다.



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

DAVID 분석은 input 한 유전자들이 유의하게 관련되는 GO, pathway 등을 분석하는 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 조정할 수 있다. David 분석 결과의 각 항목은 DAVID 홈페이지의 Help and Tool Manual 에 자세히 설명되어 있다(그림 2-7).

| Annotation Summary Result | S | Hala and Taal Manual |
|---|--------------|----------------------|
| Current Gene List: List_1 Current Background: Homo sapiens | 94 DAVID IDs | Clear All |
| Disease (1 selected) Functional Categories (3 selected) | | |
| Gene_Ontology (3 selected) | | |
| General Annotations (0 selected) Literature (0 selected) | | |



그림 2-7. DAVID Help and Tool Manual

2-2. String-db tool을 이용한 gene set분석

String-db tool 은 Protein-Protein Interaction 데이터 베이스를 기반으로 유전자의 상관관계를 통계적으로 분석하여 유전자의 주요 기능을 예측하고 Network 을 build 해 주는 분석툴이다. 분석과정은 그림 2-2-1과 같다.



String-db 에서는 500 개 이하의 유전자를 input 하는 것을 권장하고 있고 여러 public ID 중 EntrezGeneID 사용이 좀더 편리하다. mRNA-Seq 결과에서 significant gene 을 선별하고 String-db 홈페이지 (http://string-db.org/)에 접속하여 "Multiple proteins"을 클릭하고 "List of names" 입력창에 유전자 리스트를 복사한다.그리고 "Organism" 입력창에 해당 species 학명을 입력하고 "Search"를 클릭한다(그림 2-2-2).

| \leftrightarrow \rightarrow C \odot string-o | lb.org Version: 10.0 | | LOCIN RECIST | 略 ☆ : |
|--|---------------------------|---|---|-------|
| | 🕸 STRING | | Search Download Help My Da | ta |
| | Protein by name | > | SEARCH | |
| | Protein by sequence | > | Multiple Proteins by Names / Identifiers | |
| | Multiple proteins | > | | |
| | Multiple sequences | > | List Of Names: (one per line, examples: #1 #2 #3) | |
| | Organisms | > | Fix Serpina7 Cidn2 | |
| | Protein families ('COGs') | > | Kartr Adarg2 Frdr1 | |
| | Examples | > | or, upload a file: | |
| | Random entry | > | Browse | |
| | | | Organism: | |
| | | | Mus musiculus | |
| | | | | |
| | | | | |
| | | | SEARCH | |

그림 2-2-2. Multiple proteins search

"Search" 결과 중간에 아래 그림과 같은 유전자 확인 단계가 있고 별 이상이 없으면 "continue"를 클릭하여 계속 진행한다(그림 2-2-3).

| C C ctrin = all | | | |
|-----------------|---|---|--|
| U string-db | Version: 10.0 | LOGIN REGISTER | |
| | STRING Search Download | Help My Data | |
| | The following proteins in <i>Mus musculus</i> appear to match your input. Please review the list, then click 'Continue' to proceed. | CONTINUE -> | |
| | Sox17: | | |
| | Sax17 - SRV-bax containing gene 17, Acts as transcription regulator that binds target promoter NA and bends the DNA. Binds to the se or 5'-AACAAG-3'. Modulates transcriptional regulation via WNT3A. Inhibits Wirt signaling. Promotes degradation of activated CTNNB1. regulation of embryonic development. Required for normal looping of the embryonic heart tube. Required for normal development endoderm. Probable transcriptional activator in the premelotic germ cells. Isoform 2 (T_SOX17) shows no DNA-binding activity | quences 5- AACAAT-3 Plays a key role in the t of the definitive gut | |
| | Rev1': | | |
| | Rect - REVI homolog (S. cerevisiae); Decrycytidyl transferase involved in DNA repair. Transfera a dOMP residue from dCTP to the 3'emit temptate-dependent reaction. May assist in the first step in the bypass of abasic leations by the insertion of a nucleotide opposite the leation induction of mutations by physical and chemical agents. | d of a DNA primer in a on. Required for normal | |
| | Mad212 - MAD2 mitotic arrest deficient-like 2, Adapter protein able to interact with different proteins and involved in different biological pr interaction between the error-prone DNA polymerase zeta catalytic autouris REV3L and the inserter polymerase REV3, thereby mediating up switching in translesion DNA synthesis. Translesion DNA synthesis releases the replication blockade of replicative polymerases, stall lesions. May also regulate another aspect of cellular response to DNA damage through regulation of the JNK-mediated phosphoryla [] | ocesses. Mediates the he second polymerase ed in presence of DNA | |
| | Feep20 - RIKEN cDNA 2610002J02 gene; Component of the Fanconi anemia (FA) complex required to recruit the FA complex to DNA (ICLs) and promote ICLs repair. Following DNA damage recognizes and binds 'Lys-63'-linked ubiquitin generated by RNF8 at ICLs and reor of the FA complex. Promotes translesion synthesis via interaction with REVI (By similarity) | interstrand cross-links uits other components | |
| | Mars2: | | |
| | Mars2 - methionine tRNA synthetase 2 (mitochondrial) | | |
| | Igfbp2': | | |
| | ✓ Igftpp2 - insulin-like growth factor binding protein 2: Inhibits IGF-mediated growth and developmental rates (By similarity). IGF-binding protein a final disc of the IGFs and have been shown to either inhibit or stimulate the arrowth normation effects of the IGFs and have been shown to either inhibit or stimulate the arrowth. | oteins prolong the half- interaction of IGEs with | |

그림 2-2-3. Gene confirmation step

분석이 완료되면 그림 2-2-4와 같이 String DB 기반 Network 결과를 확인할 수 있고 "Analysis" 탭을 클릭하면 "Functional enrichments in your network" 결과를 확인할 수 있다(그림 2-2-5). 각 Functional DB 결과의 오른쪽 하단에 "more"를 클릭하면 FDR<0.05 이하에 해당하는 항목을 모두 볼 수 있다.



그림 2-2-4. String network result

| | | 0.2277/ # | | | And the second second second | | | | |
|----------------|-----------------------------|--------------------------------------|---------------------------------|------------------------------|------------------------------------|------------------------------|--|---------------|---------------------------|
| | | @ Legend > | Data Settings | © View Settings ≥ | III Tables / Exports 💙 | Fvidence | E Analysis Y | De | |
| Network Sto | ta | | | | | | | | |
| | | a see bar a | factor: DEF | | | a second as | mbar of originary 105 | | |
| | | number o | of edges: 272 | | | PP enri | thment p-value: D | | |
| | 0.03 | average node local clustering cor | efficient: 0.347 | | | you network has so | <mark>gnillicantiy more interac</mark> Jubat dasa task mosali | 16.02 | |
| | | 1 | | | | in an expected | THE SECOND CONTRACTOR | 6 | |
| Functional e | nrichments in your netw | ork | | | | | | | |
| | | | | Biological Process (| CO) | | | | |
| pathway iD | pathway description | | | | | | | eount in gene | set talae diadavery rate |
| GO:0001676 | long-chain fatty acid met | tabolic process | | | | | | 1 | 2.65e-06 |
| GD:0019575 | regulation of cell differen | way | | | | | | 41 | 7.41e-05 |
| GO:0042/38 | exogenous drug cataboli | ic process | | | | | | 7 | 0.000116 |
| GO:0050793 | regulation of developme | ntal process | | | | | | 50 | 0 000116 |
| | | | | | | | | | (nove) |
| | | | | Molecular Function | (GO) | | | | |
| pathway iD | pathway description | | | | | | | count in gene | set faise discovery rate |
| GO.0016712 | oxidoreductase activity, | acting on paired do | nors, with incorporation or rec | luction of molecular oxyges, | reduced flavin or flavoprotein a | s one donor, and incorp | cration of one atom of | oxygen 11 | 2.3e-09 |
| CO-0070.30 | monorygenate activity | | | | | | | | 1.106-00 |
| GD:0008392 | arachidonic acid eporyc | enace activity | | | | | | 8 | 1.99e-07 |
| GO.0020037 | herrie binding | | | | | | | 15 | 1.23e-06 |
| | | | | | | | | | (more) |
| | | | | KEGG Pathways | | | | | |
| pathway iD | pathway description | | | | | | | count in gene | set. Telse discovery rate |
| 00870 | Retinol metabolism | | | | | | | 15 | 9 DSe-12 |
| 00530 | Sterold pormone plosunt | theele | | | | | | | 3.200 00 |
| 00591 | Lincleic acid metabolism | | | | | | | 8 | 4.72e-06 |
| 0.52.54 | Chemical carcinogenesis | 5 | | | | | | 10 | 5.16e-06 |
| | | | | | | | | | (more) |
| | | | | PFAM Protein Donu | uins | | | | |
| pathway iD | pathway descoption | | | | | | | count in gene | set feise discovery rate |
| 11 00067 | Cytoenrome (1460 | | | | | | | 1 | 0.00745 |
| | | | 1 | NTERPRO Protein Domains | and Features | | | | |
| pathway ID | pathway description | | | | | | | count in gene | set Teise discovery rate |
| 12R001128 | Cytochrome P450 | nund atta | | | | | | 7 | P (00019 |
| IPR002401 | Cytochrome P150, E-c as | Es, group I | | | | | | 6 | 0.0238 |
| Statistical be | ockground | | | | | | | | |
| | | | | | | | | | |
| I or the above | e enrehment analysis, | | | | Maole Gener | | | LIDE ATTR | |
| is assumed: | y stati sticer backgroond | | | | whole Genu | ne v j | | UP CLAIR E | - |
| Save / Expor | rt | | | | | | | | |
| | | | Biological Process (| bcolnwob (00 | 49 GU terma argniticentiv enni | ched: tile format: tab dokr | mica | | |
| | | | Molecular I unction (| download (00 | 12 GO terma argnificently enric | ched; tile format, tab dolv | nned | | |
| | | | KEGG Pathw | oya download | 8 politivaya significantly crinici | hed five format tab delim | itcd | | |
| | | | P AM Protein Doma | ina <u>download</u> | one single domain is enriched | tile tarmat: tab delimited | | | |
| | | INTERT | 10 Protein Domains and Leatu | irca <u>download</u> | 3 domains significantly princh | ed file formal: tab delivriv | tod | | |

그림 2-2-5. Functional enrichments result

관심 있거나 중요한 Function을 클릭하면 Network상에서 해당 유전자들이 붉은색으로 표시되고 (그림 2-2-6) 관심 있는 유전자를 클릭하면 해당 유전자의 자세한 정보를 추가로 얻을 수 있다(그 림 2-2-7).



그림 2-2-7. Gene selection on your network

 "Legend" 탭에서는 Node, Edge, Input 유전자의 설명을 자세히 볼 수 있고(그림 2-2-8)"Tables/Exports" 탭에서는 Network와 유전자 정보를 파일로 저장할 수 있다.(그림 2-2-9)

| lones: | | | | | | | | | | |
|--|---|---|---|---|---|--|--|--|--|--|
| Network nodes n | epresent proteins | Node Size | | Node Color | | | | | | |
| splice isoforms or are collopsed, i.e. proteins produced focus. | post-transfational modifications sach node represente all the hy a single, protein-coding gene | ernali nodes: protein of unknown large nodes: some 3D structure i | ID structure koown or predicted | eolored nodea: query proteins and first shell of interactors white nucles: second shell of interactors | | | | | | |
| dges: | | | | | | | | | | |
| Edges represent associations are n mesoingful, i.e. pro- shared function; fi they are physically | protein-protein associations neant to be specific and mems jointly contribute to a is does not necessarily mean binding each other. | Known Interactions | Predicted I nes C-O nined C-O O-O | nteractions gene neighborhood gene finsions gene co-occurrence | Others Lextmining Construction protein homology | | | | | |
| our Input: | | | | | | | | | | |
| 🖶 Pparg | peroxisome proliferator a Once activated by a figan controls the peroxisomal a key regulator of the tiss mediated proinfia [] (60 | peroxieomo proliferator activated receptor gammo; Receptor that binds peroxiaomo proliferatore such as hypolipidemic druge and fatty acids Once activated by a figand, the receptor binds to a promoter element in the gene for acyl-CoA axidase and activates its transcription. It theref controls the peroxicomal bota oridiation pathway of fatty acids. Key regulator of adiposyste differentiation orid glucose homeostasis. ARF6 ac a key regulator of the bissue-specific adipocyle P2 (aP2) enhancer. Acts as a critical regulator of gul homeostasis by suppressing NF-kappa-5 medicade prohifid_1j (60 sol) | | | | | | | | |
| 🖶 Pdgíra | platelet derived growth (PUR) C and plays an essu context, promotes or inlu mesenchymal stem cells development of the mass | plaielet derived growth factor receptor, alpha polypeptide, "prosine-protein Kinase linat acts as a cell-surface receptor for PDGPA, PDGPB and FULLI C and plays an essential role in the regulation of embrycoic development, cell proliferation, survival and chemotaxis. Thepending on the context, promotes to tablets cell proliferation and cell magnation. Plays an important role in the differentiation of bone marrow-derived mesenchymal stein cells. Its guined for contral skeletics development and rephalic closure during embrycoic development. Hequied for normal development of the maxims into all 1. (FORM 9 as) | | | | | | | | |
| e Ifreil | interferon-related develop Induced by NGF. May bo | interteron-related developmental regulator I; Could play a role in regulating gene activity in the profilerative and/or ditterentiative pathways Induced by NGF. May be an outocrine foctor that attenuates or amplifies the Initial ligand induced algonal (449 aa) | | | | | | | | |
| Rec8 | REC8 homolog (yeast); R on chromosomo arms by centromeres during anap | o (300 BB) equired during meiosis for separation soparin during anophase i allows for hase II allows for sister chromatid se | of sister chrometids ar homologous chromoso paration in meiosis II (5 | nd homologous chromos une ocparation in meloa 191 aa) | iomes. Proteolytic cleavage of REC8 Is i and cleavage of REC8 on | | | | | |
| 👹 Tiam1 | L cell lymphome invesior activities: Acts as a GDP Activates (IAG1, CIIC42, i | 1 cell Jumphoms available and meteritaries 17 Modulates the activity of 10 IT-blue proteins and connects extrace/blue signals to cytotabeletal articlifies. Acts are a OD-disacebalue stimulate protein but stimulates the OD-OTP activity or SH-D-Be Theorem and activities them. Activities 14.01, ITIDED, and a deserver event fulling fly antimulary. Affects investmenses of Lymphome activity (1997 and). | | | | | | | | |
| Cyp2c29 | eviochrome P450. family ae) | cyrachrome P450, family 2, subfamily 6, polypoptide 29; Metabolizos arachidonic aeld to produce 14,15 ele opoxyetecaatrienole acid (EET) (490 ael | | | | | | | | |
| 🛢 Срхб | glistathione peroxidase 5 | (221 AB) | | | | | | | | |
| A complete | evtochrome P450, famliv | cytochrome P480, family 2, subfamily b, polypeptide 18 (491 aa) | | | | | | | | |
| U Cyp2013 | | | | | | | | | | |

그림 2-2-7. Legend of your network

| | as a bitmar | image download file | format is 'ENG' portable netwo | ark areahic | | |
|------------------|---------------|--|--|---|--|-------|
| as a hio | h-resolution | bitman download san | e PNG format, but regaintion : | nt 400 doi | | |
| | as a vector | graphic download SV/ | ecalable ventor graphic - can | be opened and edited in litretrator Dr | collicate Dia ato | |
| as simula | abular tex | output download TSL | tab appointed values - can be | ononod to Exact | | |
| a | s an XMI su | mmary download stru | ctured XML interaction data a | coording to the 'PSI-MI' data standard | | |
| | aburch com | diustas download of | t file format describing the se | ardinatan and colora of noden in the n | attuo de | |
| | protain con | unates, download and | tene tornat acounting the con | eremates and colors or nodes in the m | build a | |
| | protein seq | derices. download war | t multi-rasta tormat - containii 5 defeedad Ele deseeddae des | ng me aminoacid sequences in the net | work | |
| | protein a niu | rations, <u>download</u> a ta | e-delimited the accorpting the l | names, comans ano annotareo tanoti | ons of the herwork proteins | |
| Browse Inter | actions in t | abular form: | | | | |
| Anode1 | node2 | nodel accession | node2 accession | node1 annotation | node2 annotation | score |
| Acot1 | Cyp4a10 | ENSMUSP0000012644 | 8 ENSMUSP0000061126 | acyl-CoA thioesterase 1; Acyl-CoA | cytochrome P450, family 4, subfam. | 0.561 |
| Apot1 | Cyp4a14 | ENSMUSP0000012644 | 8 ENSMUSP0000030487 | acyl-CoA thlocstcrase 1; Acyl-CoA | cytochrome P450, family 4, subfam. | 0.475 |
| Apot1 | Rhbg | ENSMUSP0000012644 | 8 ENSMUSP00000130767 | acyl-CoA thloesterase 1; Acyl-CoA | Rhesus blood group-associated B | 0.473 |
| Acot2 | Cyp4a10 | ENSMUSP000002164 | 9 FNSMUSP0000061126 | acyl-CoA thloestcrase 2: Acyl-CoA | cytochrome P450, family 4, subfam. | 0.432 |
| Adrb3 | Adrbk2 | ENSMUSP000008016 | 2 FNSMUSP00000070445 | adrenergie receptor, beto 3; Beta-a | adrenergie receptor kinase, beta 2: | 0.649 |
| Adrb3 | Mc1r | ENSMUSP0000008016 | 2 ENSMUSP0000095929 | adrenergie receptor, beto 3, Beta-a | melanocortin 1 receptor; Receptor | 0.900 |
| Adrb3 | Pparg | ENSMUSP000008016 | 2 ENSMUSP0000000450 | adronorgio receptor, beto 3, Beta-a | регохівоте proliferator activated г | 0.791 |
| Adrb3 | Slo2a4 | FNSMUSP000008016 | 2 ENSMUSP0000018710 | adrenergie receptor, beto 3; Beta-a | solute carrier family 2 (focilitated g. | 0.493 |
| Adrbk2 | Adrb3 | ENSMUSP0000007044 | 5 ENSMUSP0000080162 | adrenergie receptor kinase, beta 2; | adrenergie receptor, beta 3; Beta-a | 0.649 |
| Adrbk2 | Ncald | ENSMUSP000007044 | 5 FNSMUSP0000087611 | adrenergie receptor kinase, beta 2; | neurocalotn delta; May be Involved | 0.660 |
| Aldh3a2 | Cyp4a10 | ENSMUSP0000007376 | 4 FNSMUSP0000061126 | aldehyde dehydrogenase family 3, | cytochrome P450, family 4, subfam. | 0.943 |
| Aldh3a2 | Cyp4a12b | ENSMUSP000007376 | 4 ENSMUSP0000092487 | aldehyde dehydrogenase family 3, | cytochrome P450, family 4, subfam. | 0.916 |
| Aldh3a2 | Cyp4a14 | ENSMUSP0000007376 | 4 ENSMUSP0000030487 | aldehyde dehydrogenase family 3 | cytochrome P450, family 4, subfam. | 0.947 |
| Aldh3a2 | Cyp4a32 | ENSMUSP000007376 | 4 ENSMUSP0000081369 | aldehyde dehydrogenase family 3 | cytochrome P450, family 4, subfam. | 0.904 |
| Apol10a | Apol11a | ENSMUSP000006065 | 0 FNSMUSP00000132565 | apolipoprotoin L 10A | opolipoprotein L 11a | 0.900 |
| Apol11a | Apol10a | ENSMUSP0000013256 | 5 FNSMUSP0000060650 | apolipoprotein L 11a | opolipoprotein L 19A | 0.900 |
| Aqp4 | Avpr1a | ENSMUSP000007808 | 8 ENSMUSP00000020323 | aquaporin 4; Forms a water-specifi. | orginine vocopressin receptor 1/; | 0.430 |
| Aqp4 | Sle1a2 | ENSMUSP0000007808 | 8 FNSMUSP00000079100 | aquaporin 4; Forms a water-specifi | solute carrier family 1 (glial high af | 0.630 |
| Aqp4 Arhgap10 | Trpv4 Rac3 | ENSMUSP0000007808 ENSMUSP0000007565 | 8 ENSMUSP00000071859 8 ENSMUSP00000018156 | aquaporin 4; Forma a water-specifi Rho GTPase activating protein 10; | transient receptor potential cation RAS-related C3 botulinum aubstrat | 0.910 |
| | e 1 of 28 ⊯ | н | | | | |

그림 2-2-8. Tables/Exports of your network

2-3. MSigDB기반 GSEA 분석

GSEA 분석은 MSigDB 기반으로 유전자의 상관관계를 통계적으로 분석하여 입력한 유전자 셋의 주요 기능을 예측하고 각 유전자가 어떤 기능들에 포함되는지 overlap 분석을 제공해 준다. 분석과정은 그림 2-3-1과 같다.



그림 2-3-1. Web based GSEA tool analysis process

MSigDB에 접속하여 "Investigate gene sets"을 클릭하고 등록한 이메일을 입력하여 로그인을 수행한다.(그림 2-3-2).만약 등록이 필요할 시 "Click here"을 클릭하여 등록을 진행하면 된다.(그림



| 🗱 GSEA Login 🛛 🗙 | σ |
|---|--------------------------|
| → C ③ software.broadinstitute.org/gsea/login.jsp | 20 众 |
| | |
| CSFA | |
| COLA | antation Portact |
| Galar nome powniejska moleculor arginiculta, powniejska pocul | andown contract |
| | |
| Login to GSEA/MSigDB | |
| Login | |
| Click here to register to view the MSigDB gene sets and/or download the GSEA software. This helps us track and better | erve our user community. |
| If you have already registered for GSEA or MSigDB please enter your registration email address below. | |
| Items marked with # are required. | |
| | |
| | |
| Email: * | |

그림 2-3-3. GSEALogin page

"Gene Identifiers"입력창에 유전자 리스트(Gene Symbol, EntrezGeneID 또는 public ID)를 입력하고 "Compute Overlaps"에 원하는 DB 를 클릭한 후 맨 아래 "compute overlaps" 버튼을 클릭한다.(그림 2-3-4).DB 선택시 DB 명 앞의 파란색 글자를 누르면 해당 DB 정보를 확인할 수 있다.



그림 2-3-4. GSEAAnalysis

분석이 완료되면 그림 2-3-5 와 그림 2-3-6 과 같이 통계적으로 유의한 Gene Set List 와 Gene/Gene-set Overlap Matrix 결과를 확인할 수 있다.



그림 2-3-6. GSEAAnalysisResult(Gene/Gene-set Overlap Matrix)

3. KEGG DB 기반 Pathway 분석

mRNA-Seq 분석 결과에서 up/down-regulated 유전자들이 어떤 Pathway에 속하는지 확인하고자 한다면 KEGG에서 제공하는 KEGG Mapper를 이용하면 된다. 사용방법은 그림 3-1과 같은 순서로 진행된다.



그림 3-1. KEGG Mapper tool analysis process

그림 3-2는 mRNA-Seq report에서 2fold, normalized RC(log2)>6을 기준으로 선별한 유전자를 KEGG 분석하는 과정이다.

*KEGG input 값은 excel 파일의 Annotation 항목 앞에 제작되어 있다.

오른쪽 필터에서 Fold change와 Normalized RC (반복실험의 경우 p-value) 값을 지정하고, 확인하 고자 하는 Fold change 조합을 선택하여 필터를 적용 한다.

필터를 적용하여 선별 된 유전자의 KEGG input [Entrez ID, FC Color(#숫자,black)] cell을 복사하여, KEGG 분석에 사용할 것이다.

| | ※ 잘라내기 ···································· | Calibri | • 11 · | ר א ר | = = = | # ¹ /** | 📑 텍스트 줄 | 들 바꿈 | 숫자 | | • | 1 | - 🖹 | | 통 합계 · 🛃 | A |
|------|--|-----------------|---------------|--------------|-------------|--------------------|---------|-------------|-------|------------|------------------------|-------------------|----------|------|----------|--------------------|
| 붙이 | 1넣기 🗸 🛷 서식 복사 | · 가가가 가 | · 🖽 • 🎒 • 💾 | 내철 | | | 韓 병합하고 | 가운데 맞춤 | • ₩• | % , | 8,000 조건부 표 서식 × 서성 | 트 셸 샵(닉 * 스타) | | | | 같기 및 선택 ▼ |
| | 클립보드 | 5 | 글꼴 | 5 | | <u>,</u> | !ē | | 5 | 표시 형식 | 5 <u></u> | 4 <u>8</u> | L | | C | |
| | B3 | - (• | fx 1.48037036 | 5202988 | | | | | | | | |] ľ | (EQ | U | |
| | А | В | С | D | E | F | G | Н | 1 | J | К | L | | | | - |
| 1 | Filter: 1039 | Fold c | nange | Norm | alized RC (| log2) | Ra | w data (RC) | | | KEGG mapper | nput | • | | 01 | Inalysis |
| 2 | Gene symbol | B/A 💌 | C/A 💌 | Α 🔽 | B | c 🔽 | Α 🔽 | В | C 🔽 | Entrez_In_ | B/A | C/A | In | put | Ť | election |
| 724 | EIF2B3 | 2.021 | 1.579 | 9.681 | 8.706 | 10.560 | 936 | 405 | 1416 | 3891 | #FF6347,black | #FFA07A,black | | | | |
| 751 | UQCRH | 3.086 | 1.929 | 11.872 | 10.406 | 13.375 | 4277 | 1318 | 9966 | 7388 | #FF6347,black | #FFA07A,black | | -+-1 | | |
| 947 | TYW3 | 2.051 | 1.946 | 10.192 | 9.584 | 11.211 | 1334 | 745 | 2224 | 127253 | #FF6347,black | #FFA07A,black | | 시 | | |
| 966 | NEXN-AS1 | 0.435 | 1.108 | 8.972 | 8.607 | 7.677 | 572 | 378 | 191 | 374987 | #87CEEB,black | | | • | |) |
| 1001 | LOC646626 | 0.498 | 0.586 | 8.060 | 7.484 | 6.778 | 303 | 173 | 102 | 546626 | #87CEEB,black | #B0E0E6,black | Ļ | | | v |
| 1023 | GBP3 | 2.319 | 2.148 | 7.301 | 6.399 | 8.231 | 179 | 81 | 281 | 2635 | #FF6347,black | #FF6347,black | GBP3 | 7 | p-value | |
| 1184 | LAMTOR5 | 2.047 | 1.839 | 11.397 | 10.069 | 12.313 | 3076 | 1043 | 4775 | 10542 | #FF6347,black | #FFA07A,black | LAMTOR5 | HE | - | |
| 1271 | CD101 | 0.454 | 2.828 | 4.446 | 6.170 | 4.004 | 24 | 69 | 14 | 9398 | #87CEEB,black | #FF6347,black | CD101 | EV | | |
| 1331 | TXNIP | 2.506 | 3.261 | 6.363 | 5.797 | 7.349 | 93 | 53 | 152 | 10628 | #FF6347,black | #FF6347,black | TXNIP | ES | Sample C | omparison / Filter |
| 1403 | MRPS21 | 2.122 | 1.345 | 9.250 | 8.035 | 10.133 | 694 | 254 | 1053 | 54460 | #FF6347,black | #FFE4B3,black | MRPS21 | MI | - | |
| 1409 | ADAMTSL4 | 0.481 | 0.591 | 10.394 | 10.734 | 8.768 | 1535 | 1655 | 408 | 54507 | #87CEEB,black | #B0E0E6,black | ADAMTSL4 | AE | ✓ B/A | <u>^</u> |
| 1428 | SEMA6C | 0.440 | 0.343 | 9.007 | 9.265 | 7.518 | 586 | 597 | 171 | 10500 | #87CEEB,black | #87CEEB,black | SEMA6C | m- | C/A | |
| 1436 | PSMD4 | 2.018 | 0.991 | 12.673 | 12.047 | 13.476 | 7455 | 4111 | 10689 | 5710 | #FF6347,black | | PSMD4 | AF | | |
| 1441 | SELENBP1 | 0.168 | 2.248 | 6.289 | 5.086 | 3.102 | 88 | 32 | 7 | 3991 | #00BFFF,black | #FF6347,black | SELENBP | | | |
| 1465 | HRNR | 0.391 | 0.935 | 10.512 | 10.985 | 9.284 | 1666 | 1969 | 584 | 388697 | #87CEEB,black | | HRNR | | 21-1-1 | |
| 1550 | SHE | 0.386 | 1.612 | 6.004 | 5.289 | 4.553 | 72 | 37 | 21 | 126669 | #87CEEB,black | #FFA07A,black | SHE | | ᆡ아 | |
| 1570 | EFNA3 | 0.481 | 0.398 | 10.079 | 9.817 | 8.449 | 1233 | 876 | 327 | 1944 | #87CEEB,black | #87CEEB,black | EFNA3 | | | |
| 1575 | TRIM46 | 0.430 | 0.437 | 8.955 | 9.185 | 7.081 | 565 | 565 | 126 | 30128 | #87CEEB,black | #87CEEB,black | TRIM46 | | | |
| 1632 | CRABP2 | 0.379 | 1.222 | 11.041 | 10.557 | 9.249 | 2403 | 1463 | 570 | 1382 | #87CEEB.black | | CRABP2 | | - 1 | |
| | | | | | | | | | | | | | | | FOI | d |
| | | | | | | | | | | | | | | | | - |
| | | | | | | | | | | | | | | cha | ang | e 항 |
| | | | | | | | | | | | | | | 5 | ᆚᄷ | IEH |

그림 3-2. KEGG Mapper tool analysis process

그림 3-3과 같이 KEGG Mapper 웹페이지(<u>http://www.genome.jp/kegg/tool/map_pathway2.html</u>)에 접속하고 Search & Color pathway 링크에 들어가면 아래와 같은 화면이 보여진다. 분석하고자 하 는 유전자의 species를 선택하고, 'primary ID'는 KEGG identifiers로 선택한 뒤 'Enter objects one per line followed bgcolor, fgcolor' 창에 엑셀에서 준비해 놓은 Entrez ID, Color 항목을 복사-붙여넣 기를 한다. 마지막으로 "Include aliases"와 "Use uncolored diagram" 항목에 체크를 한 후 Exec 버 튼을 누른다.

| KEGG KEGG | Mapper – Search&Color Pathway | |
|--|---|--|
| About KEGG Mapper | Search agains: hsa | org |
| Search Pathway Search&Color Pathway Color Pathway Color Pathway WebGL | Primary ID: KEGG identifiers | pecific nathways only) |
| Search Brite Search&Color Brite Join Brite Join Brite Table | 633 # <u>87CEEB</u> ,black ▲ 105373383 # <u>87CEEB</u> ,black 1852 # <u>87CEEB</u> ,black 10134 # <u>87CEEB</u> ,black 2157 # <u>87CEEB</u> ,black 283981 # <u>87CEEB</u> ,black | Examples Select Find organism - Chrome |
| Search Module Search&Color Module | 1438 # <u>87CEEB</u> ,black 9189 # <u>87CEEB</u> ,black 114758 # <u>87CEEB</u> ,black | Find three- or four-letter KEGG organism code |
| Search Disease Reconstruct Pathway | Alternatively, enter the file name containing the data: 파일 선택 선택된 파일 없음 | human Select Cancel human Homo sapiens (human) [hsa] human body Jouse) Pediculus humanus corporis (human body Jouse) [ph |
| Reconstruct Brite Reconstruct Module Map Taxonomy | If necessary, change default bgcolor: pink | numan body louse) realcalus numanus corports (numan body louse) (pric |
| Convert ID | ✓ Use uncolored diagrams | * |
| Annotate Sequence BlastKOALA | Display objects not found in the search | |
| KEGG Atlas KEGG | Search pathways containing all the objects (AND search) Exec Clear | |

그림 3-3. KEGG Mapper tool analysis process

분석결과, 입력한 유전자들이 관여하는 pathway list가 나온다(그림 3-4). pathway 이름 옆에 있는 괄호 안 숫자는 입력한 유전자 중 각 pathway에 관여하는 유전자의 수이다. 괄호 안 숫자를 클릭 하면 해당 유전자 목록을 볼 수 있다. pathway 이름을 클릭하면 해당 pathway chart가 열리고 입 력한 유전자의 발현 up/down (red/green)이 색으로 표시되어 있다. Pathway 이미지는 "다른 이름 으로 저장"이 가능하고 "html"으로 저장하면 이미지에 링크된 항목을 그대로 유지해서 저장이 가 능하다.



그림 3-4. KEGG Mapper tool analysis result

4. MeV Software 이용 Clustering Heatmap 작성

MeV 소프트웨어는 미국의 Dana-Farber Cancer Institute에서 개발한 Microarray, mRNA-Seq 전용 분석 프로그램으로 연구자들에게 무료로 공급하고 있다. 주로 clustering 분석과 통계분석(Kmeans clustering, Hierarchical clustering, t-test, Significance Analysis of mRNA-Seqs, Gene Set Enrichment Analysis, EASE)을 할 수 있는 프로그램이다. 아래 웹페이지에 접속하면 최신의 업데이 트된 프로그램과 매뉴얼을 다운받을 수 있다.

<u>http://www.tm4.org</u> >> 오른쪽 Browse 항목내 "TM4 MeV Stand-Alone Client" 클릭 프로그램을 다운받아 압축을 풀고, MeV 또는 TMEV를 클릭해서 프로그램을 실행시킨다(그림4-1).MEV프로그램을 실행시키면 세 개의 창이 나타난다(그림4-2). 분석창은 프로그램창의 메뉴에서 file->New multiple array viewer를 통해 여러개를 생성할 수 있고 데이터 분석은 분석창을 통해 진 행한다.







그림 4-2. MeV program windows

본 자료에서는 MeV 프로그램을 이용하여 Clustering 분석 방법을 설명한다. 우선 MeV 프로그램 에 input할 데이터를 엑셀에서 파일 양식에 맞춰 저장해야 한다. 엑셀에 clustering 하고자 하는 유전자 이름과 fold change 또는 발현값(intensity)를 정리한다(그림 4-3). 그리고 '텍스트 (탭으로 분리)'파일 형식으로 저장해야 MeV에 upload 할 수 있다. MeV에서는 2만 개 이상의 유전자는 clustering 분석을 할 수 없으므로 2만 개 이하로 유전자를 선별해야 한다.

| 0 | 1 | (° -) = | | | | | | | | | | | test | - Micro | soft Exc | cel | | |
|-----|-----------|------------|------------|----------|--------------------------------|--------------------|---------|----------|-------------------|------------|--------------|------------|-----------|------------|-----------|---|-----|--------------------|
| | · · · · | 남입 페이 | 지 레이아웃 | 수식 | 데이터 경 | 김토 | 보기 | 추가 | 기능 | Acrobat | | | | | | | | |
| Acc | | 스트 기타 원본 - | 기존 모 연결 | 두 새로 이 이 | 건결 속성 건결 편집 ^역 | 다 <mark>공</mark> 회 | | | 우기 시 적용 1급 | 텍스트 나누기 | 중복된 항목 제가 | 데이터 유효성 검사 | - 통합 - | 가상 분석 • | **** # | * 전 전 전 전 전 전 전 전 전 전 전 전 전 전 전 전 전 전 전 | ·분합 | 이 문 이 위 이 문 하 위 |
| | 4440 | 14 101201 | 6 | C a | | | 08: | × 르미 | | 10 | | 데이터 포구 | | | - | | 271 | |
| - | G4 | • | 6 | Jx | | | | 0 | - | | | | 14 | | | | | |
| 1 | A | B | CIA | D | E /A | F | | G | - | н | - | J | K | | L | M | | N |
| 2 | Gene symt | B/A 2.52 | C/A 2 241 | D/A | E/A | 8 | | | | | 다른 | 이름으로 저 | 장 | | | | | × |
| 2 | RRDMS2 | 0.482 | 22 152 | 0.002 | 0.438 | E | | · ↑ 🔳 | 바탕 호 | 타면 > | | | C | 바탕 화 | 면 검색 | | , | P |
| 4 | NELL2 | 0.402 | 2 098 | 2 254 | 10 377 | | | | | | | | | | | 1714 | | |
| 5 | ORCE | 0.401 | 3,659 | 2.607 | 0.306 | 7 | 성 • | 새 쓸더 | | | | | | | | (i)- • | | |
| 6 | VKORC1 | 2.166 | 2.332 | 29.065 | 5.463 | | 🗼 다운 | 로드 | ^ | 3 | | | | | | | | ^ |
| 7 | ABCC6 | 3.455 | 4.018 | 4.935 | 0.5 | | 비망 바탕 | 화면 | | | 홍 그를 | | | | | | | |
| 8 | TBCD | 0.466 | 0.191 | 2.789 | 2.56 | | 웹 최근 | 위지 | | _ | | | | | | | | |
| 9 | PNMT | 0.275 | 4.544 | 2.396 | 0.373 | 1.00 | 1 E 73 | | | 0 | ebg | | | | | | | |
| 10 | SERPINF1 | 13.766 | 2.581 | 2.217 | 2.473 | | 0 = -1 | | | 10 | | | | | | | | |
| 11 | PMP22 | 2.011 | 0.194 | 4.287 | 2.779 | 14 | I LI PC | | ~ | | 10.00 | | | | | | | ~ |
| 12 | GIT1 | 0.341 | 0.142 | 2.056 | 4.914 | | TL 01 | | tart | | TH BC | | | | | | | |
| 13 | KRT38 | 8.543 | 2.753 | 0.257 | 0.311 | | | 이금((14). | LESI | - | - | | | | | | | Ť |
| 14 | KRTAP9-3 | 0.236 | 3.14 | 2.553 | 0.21 | | 파일 | 영식(0: | 텍스트 | (법으로 분 | 리) | > | | | | | | ~ |
| 15 | KRT34 | 0.146 | 0.428 | 2.655 | 2.068 | | | 만든 이; | ebg | | | E | H그: 태 | 그 추가 | | | | |
| 16 | COPZ2 | 6.842 | 0.345 | 5.132 | 0.487 | | | | | | | | | | | | | |
| 17 | GRP | 2.008 | 2.532 | 0.394 | 3.52 | | 문더 수 | 7171 | | | | 도구 | (L) - | 저장 | ł(S) | 취 | 소 | |
| 18 | SFRP1 | 0.371 | 14.466 | 2.145 | 0.125 | | e -1 6 | | | | | | | - | | | | |
| 19 | VSIG10L | 0.293 | 0.209 | 2.052 | 8.236 | | | | | | | | | | | | | |
| 20 | ATG4D | 0.487 | 2.069 | 0.31 | 8.464 | | | | | | | | | | | | | |

그림 4-3. Data format example

input 데이터 저장이 완료되면 MeV 프로그램의 분석창에서 file -> load data를 실행한다(그림 4-4). Browse를 클릭하여 input 데이터를 선택한다. 데이터가 fold change인 경우 "Two-color Array" 로 체크하고 데이터가 intensity인 경우는 "Single-color Array"에 체크한다. 마우스로 데이터가 시 작되는 부위를 클릭한 후 load를 누른다.

| <u>ی</u> | | | | Expressio | on File Loade | er | | | - 🗆 🗙 |
|----------------|-------------------|-----------------|-------------------|----------------------|---------------|------------|--------|------------------|-------------------|
| Select File Lo | ader <u>H</u> elp | , | | | | | | | |
| -File (Tab D | elimited Mu | Itinie Samnle (| (* *)) | | | | | | |
| 1110 (100.00 | chinica ma | nuple Sumple | | | | | | | \frown |
| Select expre | ecion data | file C:WUsers | #ebg#Desktop | #test.txt | | | | | Browse |
| Selectexpre | 551011 uata | ine | | | | | | | |
| Selected file | s | C:#Users | #ebg#Desktop | #test.txt | | | | | |
| Two- | color Array | | | | Single | e-color Ar | тау | | |
| Load Annota | tion Data | | ~ | | | ~ | < < | | |
| Load Annota | uon Data | | Fold chan | ae | | | | intensity | |
| | | L | | 9- | | | | | |
| Automatic | tically dowr | nload | Load from loc | al file | | | Loa | ad Annotation | |
| Choose an | organism | - N | | | | | Please | choose an array | and species name. |
| | | - | | Choose | File | | | | |
| | | | | | | | _ | | |
| Expression | Table | | Load 전 | 데이터 | 시작 위치 | click | | | |
| gene symbo | B/A | C/A | D/A | E/A | | | | | |
| FAM174B | 2.520 | 3.341 | 0.062 | 2.436 | | | | | ^ |
| RBPMS2 | 0.482 | 22.152 | 0.437 | 0.438 | | | | | |
| NELL2 | 0.427 | 2.098 | 2.254 | 10.377 | | | | | |
| ORC6L | 0.401 | 3.659 | 2.607 | 0.306 | | | | | |
| VKORC1 | 2.166 | 2.332 | 29.065 | 5.463 | | | | | |
| ABCC6 | 3.455 | 4.018 | 4.935 | 0.500 | | | | | |
| TBCD | 0.466 | 0.191 | 2.789 | 2.560 | | | | | |
| PNMT | 0.275 | 4.544 | 2.396 | 0.373 | | | | | |
| SERPINE1 | 13.766 | 2.581 | 2.217 | 2.473 | | | | | |
| PMP22 | 2.011 | 0.194 | 4.287 | 2.119 | | | | | |
| | | | | | | | | | |
| Click the upp | per-leftmos | t expression v | alue. Click the L | oad button | to finish. | | | | |
| | | | ? Me∖ | • MultiExp Viewer | periment C | ancel | Load | \triangleright | |

그림 4-4. Data uploading method

데이터가 열리면 Adjust Data -> Log Transformation -> Log2 Transform을 선택하여 fold change는 log2(fold change)로, intensity는 log2(intensity)로 바꿔준다(그림 4-5). 왼쪽 메뉴의 Original Data - > Expression image를 보면 log2 값으로 바뀌어 색이 변한 것을 확인할 수 있다.



그림 4-5. Log2 transformation

Analysis-> Clustering-> HCL을 선택하여 Clustering 분석을 시작한다(그림 4-6).



Clustering 분석 시 다양한 옵션을 선택할 수 있다(그림 4-7). Gene tree를 선택하면 fold change 또는 intensity가 유사한 유전자끼리 clustering한 결과가 나온다. Sample tree를 선택하면 발현이 유사한 샘플끼리 clustering한 결과가 나온다.당사에서 clustering 분석을 할 때 Distance Metric는 Euclidean Distance로 Linkage Method Selection은 Average linkage clustering으로 설정한다. 다른 옵션을 선택해도 된다. 옵션을 선택하고 OK를 누른다.

| HCL: Hierarchical Clustering |
|---|
| MeV |
| Tree Selection |
| Gene Tree Sample Tree |
| Ordering Optimization |
| Optimize Gene Leaf Order Optimize Sample Leaf Order |
| (Leaf ordering optimization will increase the calculation time) |
| Distance Metric Selection |
| Current Metric: Euclidean Distance |
| (The default distance metric for HCL is Pearson Correlation) |
| Use Absolute Distance |
| Linkage Method Selection |
| <u>Average linkage clustering</u> |
| Complete linkage clustering |
| Single linkage clustering |
| Validation |
| Use Validation (Requires MeV+R) |
| ? MeV•MultiExperiment Reset Cancel OK |

그림 4-7. Hierarchical Clustering Method

clustering이 완료되면 왼쪽 메뉴에 Analysis Results에 HCL 결과가 생긴다. HCL -> HCL tree를 클 릭하면 clustering 결과가 화면에 나온다(그림 4-8). 위의 tree는 sample clustering 결과이고 왼쪽 tree는 gene clustering 결과이다.각 tree에는 distance scale bar가 있어서 tree의 길이를 가늠할 수 있다. tree의 길이는 distance이며, distance가 짧을수록 유전자 간 또는 샘플 간의 발현이 비슷한 것, 길수록 발현이 다른 것이다.



그림 4-8. Hierarchical Clustering Result

clustering 결과는 이미지의 크기와 색상을 조절하여 원하는 형태의 이미지를 만들 수 있다(그림 4-9,4-10)



그림 4-9. Clustering image size control

Display -> Set Color Scale Limits을 누르면 color scale bar의 최소값, 중간값, 최대값을 설정할 수 있다. 보통 log2(fold change)는 최소값과 최대값은 같은 크기에 부등호만 바꿔주고(예: min:-3, max:3) 중간값은 0으로 설정해 준다(그림 4-10). 이렇게 하면 up-regulated genes은 red, down-regulated genes은 green으로 나타나게 된다.



그림 4-10. Clustering image color setting

원하는 이미지 조절이 완료되면 File -> Save image를 눌러 이미지를 저장한다. 이때 파일 이름에 파일 확장자명(예: .jpg)을 꼭 기입하여야 이미지 파일로 저장이 된다(그림 4-11).



그림 4-11. Clustering image save