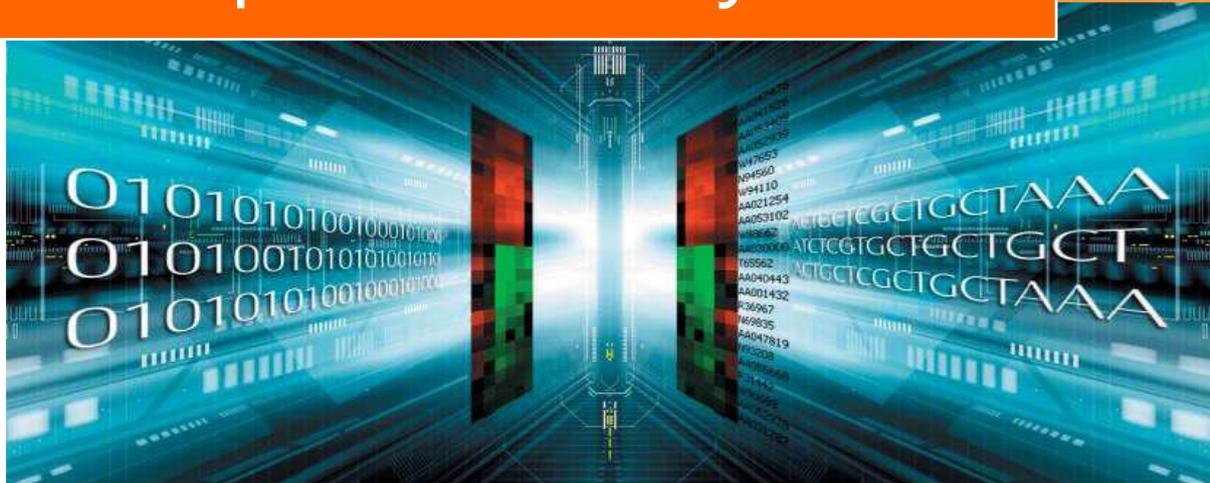


2017

miRNA-Seq Data Analysis



(주)이바이오젠

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

service@e-biogen.com

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

< 목 차 >

1. 엑셀기반 miRNA-Seq data 분석 (ExDEGA)

1-1. miRNA Category

1-2. Differential expression analysis

2. Functional category별 miRNA 분석

3. 후보 miRNA의 target gene 분석

4. DAVID tool을 이용한 target gene set 분석

5. MeV Software (free) 이용 Clustering 분석

1. 엑셀기반 miRNA-Seq data 분석 (ExDEGA)

㈜이바이오젠에서는 miRNA-Seq data 를 엑셀 기반에서 쉽게 분석할 수 있도록 ExDEGA (Excel based Differentially Expressed Gene Analysis) tool 을 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 업데이트될 예정이다.

이바이오젠에서 제공하는 Microarray data 와 RNA-Seq data (엑셀 데이터)를 열기전에 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 설치전에 실행 중인 엑셀 파일이 있으면 종료시킨 후 다시 실행해야 ExDEGA 를 사용할 수 있다.

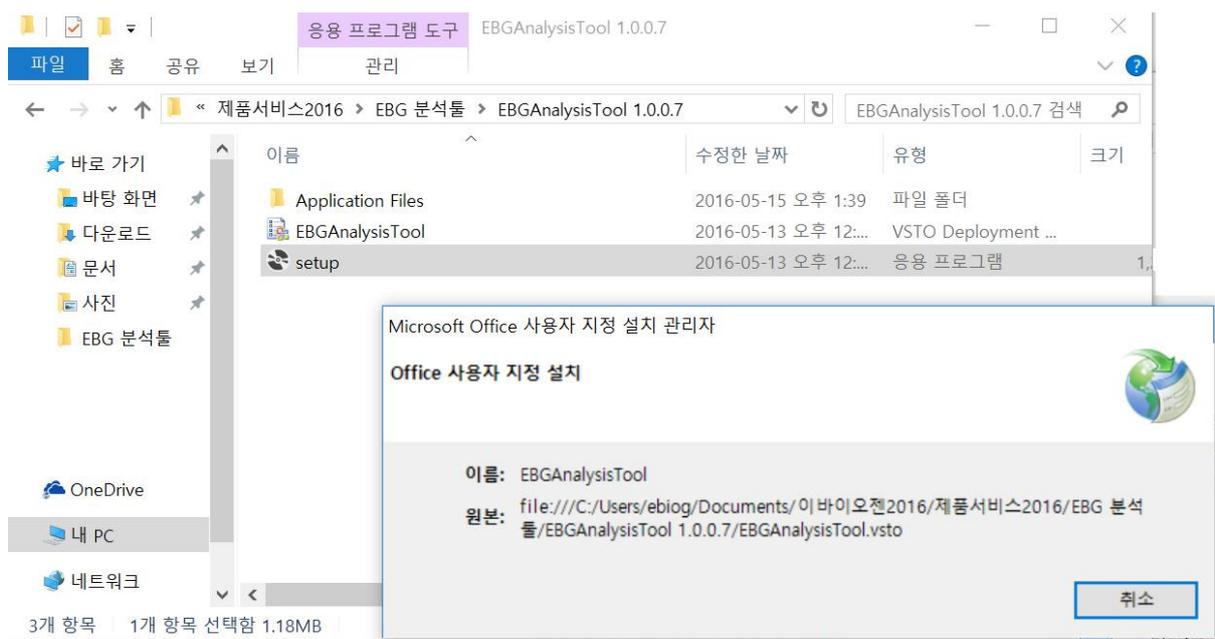


그림 1-1. ExDEGA set up

miRNA-Seq Report???.xls 파일을 열면 왼쪽에 miRNA Category 분석 창과 가운데에 miRNA-Seq data, 오른쪽에 Differential expression 분석 창이 나온다(그림 1-2).

GO 분석 창에서는 기본 설정된 GO 와 사용자가 원하는대로 GO 를 구성하여 분석할 수 있고 DEG 분석과 함께 연동하여 데이터를 쉽게 얻을 수 있다. DEG 분석 창에서는 Fold change, Read Count(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 Venn Diagram 및 Gene expression graph, Gene search 기능도 이용할 수 있어 연구자가 RNA-Seq data 를 쉽게 활용할 수 있다.

1-1. miRNA Category

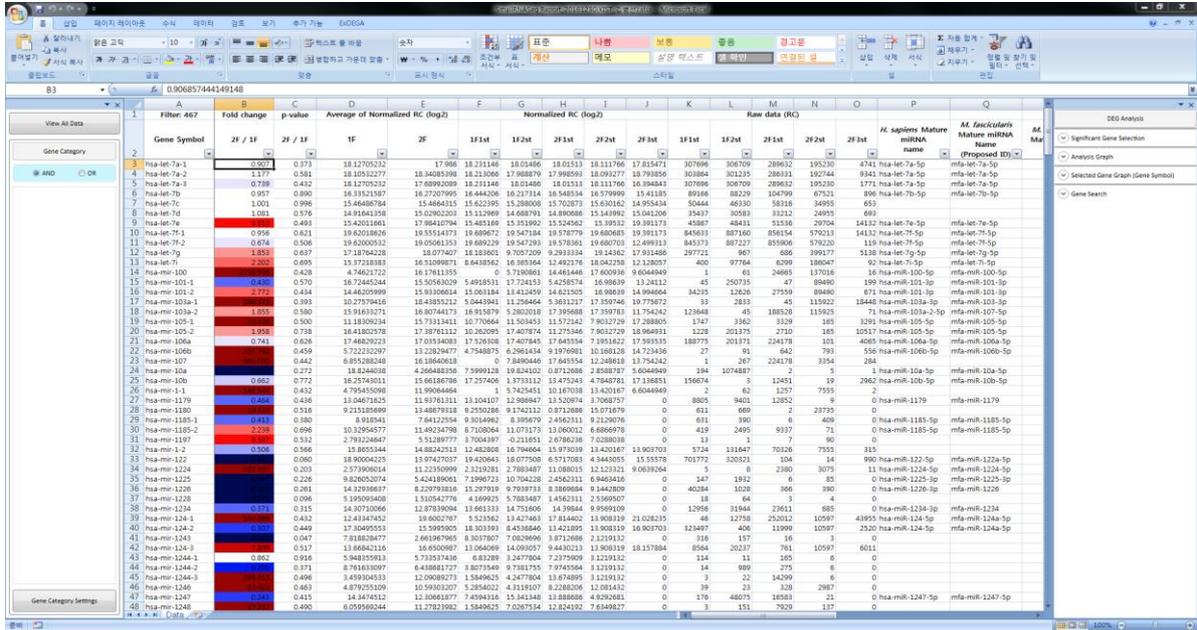


그림 1-2. miRNA-Seq data format made in E-Biogen

miRNA-Seq data 는 수 천개의 miRNA 를 포함하기 때문에 miRNA 를 한 개씩 분석하기 보다 기능별로 그룹을 지어 분석을 하는 것이 용이하다. 이를 위해 ExDEG 는 많은 연구자들에게 miRNA Category 기능을 제공한다. miRNA Category 는 비슷한 기능의 유전자들을 조절하는 miRNA 들을 묶어놓은 그룹이라고 생각하면 이해하기 쉽다.

miRNA Category 를 설정하기 위해서는 화면 좌측하단의 miRNA Category Settings 을 클릭한다(그림 1-3).

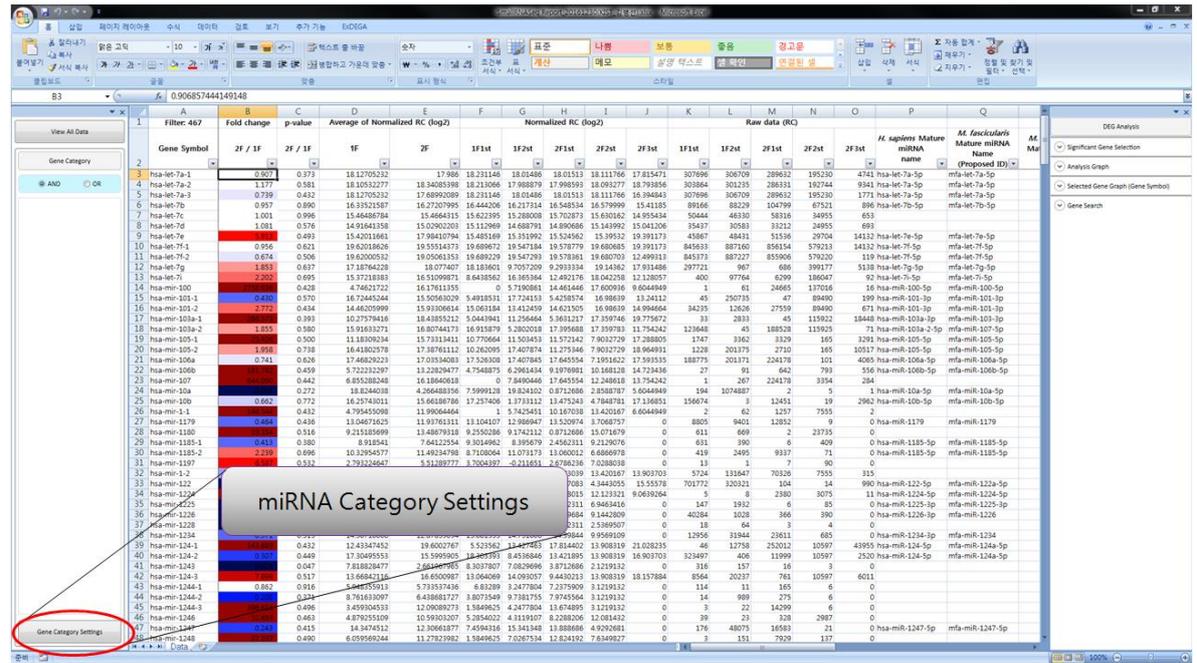


그림 1-3. miRNA Category Settings

miRNA Category Settings 을 클릭하면 (그림 1-4)와 같은 miRNA Category Settings 창이 나타난다.

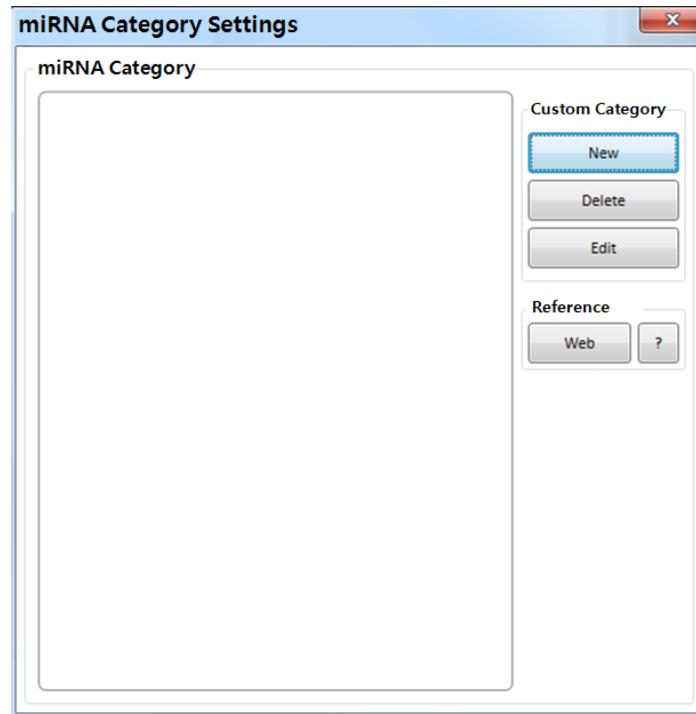


그림 1-4 miRNA Category Settings

Reference의 Web을 클릭하면 (그림 1-5)과 같은 웹페이지(<http://zmf.umm.uni-heidelberg.de/apps/zmf/mirwalk2/holistic.html>)로 연결된다. 아래의 웹페이지에서 miRNA Category를 위한 다양한 자료를 다운로드 받을 수 있다. 또한, "?" 버튼을 누르면 miRNA Category를 추가하는 방법이 자세히 설명되어 있다.

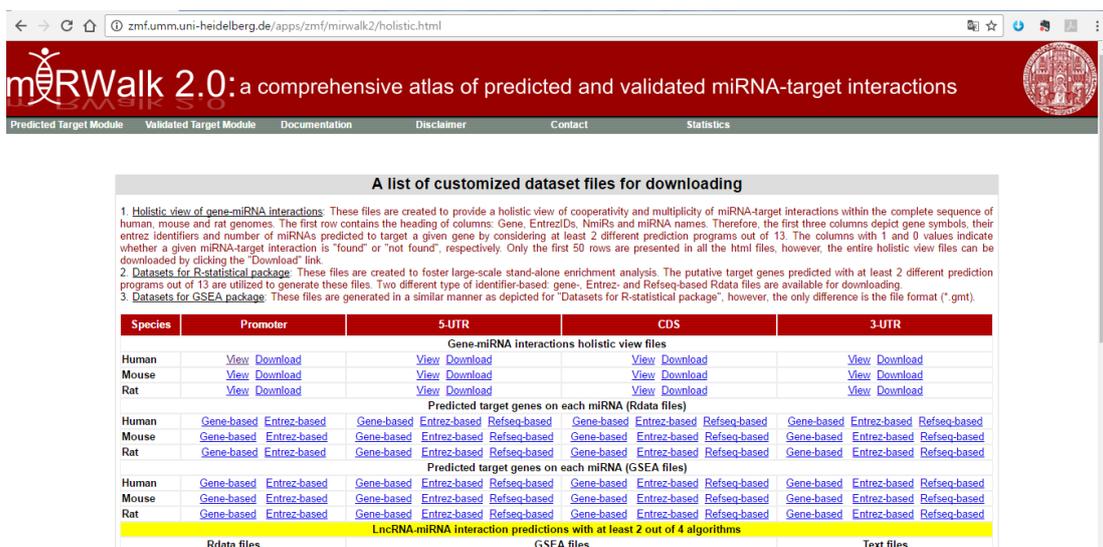


그림 1-5 miRWalk 2.0 website

여기서는 immune response 에 관계되는 miRNA Category 를 입력하는 방법을 설명한다. miRWalk 2.0 에서 다운로드 받은 자료 또는 이바이오젠에서 제공하는 파일을 열어 그림 1-6 과 같이 [immune response]를 포함하도록 필터링하여 miRNA name 을 확보한다.

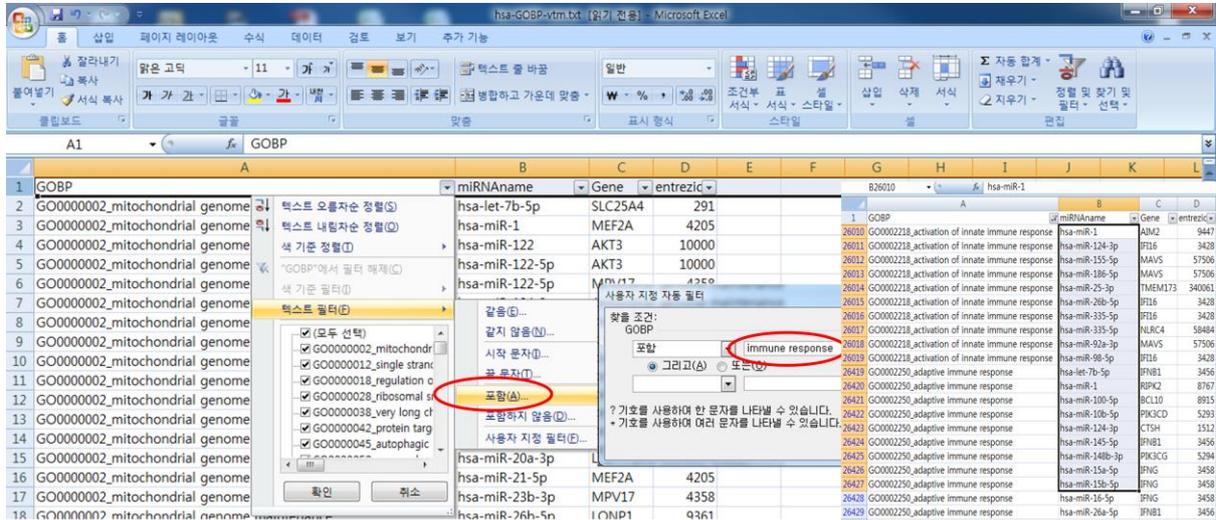


그림 1-6. miRNA Category list from miRWalk 2.0

해당하는 모든 miRNA 의 name 을 복사하여 miRNA Category settings 창에서 [New]를 클릭하고, Category 이름을 입력하고 miRNA name 을 붙여넣은 후, [OK] 버튼을 클릭하여 설정을 마친다(그림 1-7).

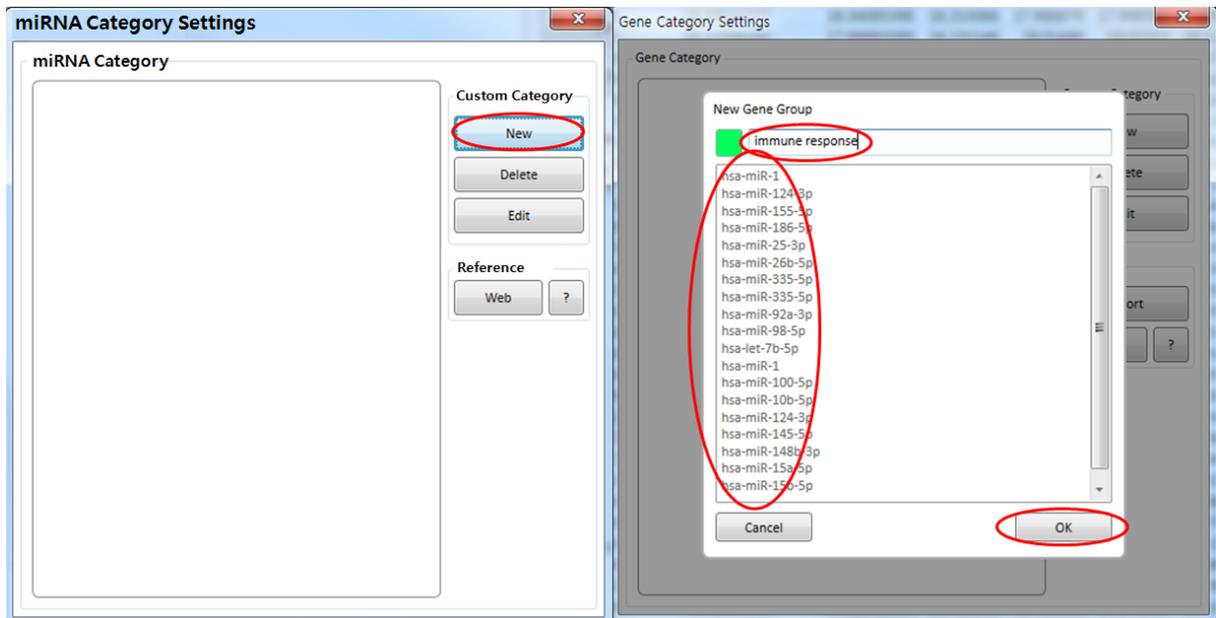


그림 1-7. miRNA Category set-up

연구자의 목적에 맞는 miRNA Category 를 설정한 이후, Aging 관련 유전자만 분석을 원할 경우, miRNA Category 창에서 Aging 을 선택하면 관련된 miRNA 가 필터링 된다(그림 1-8).

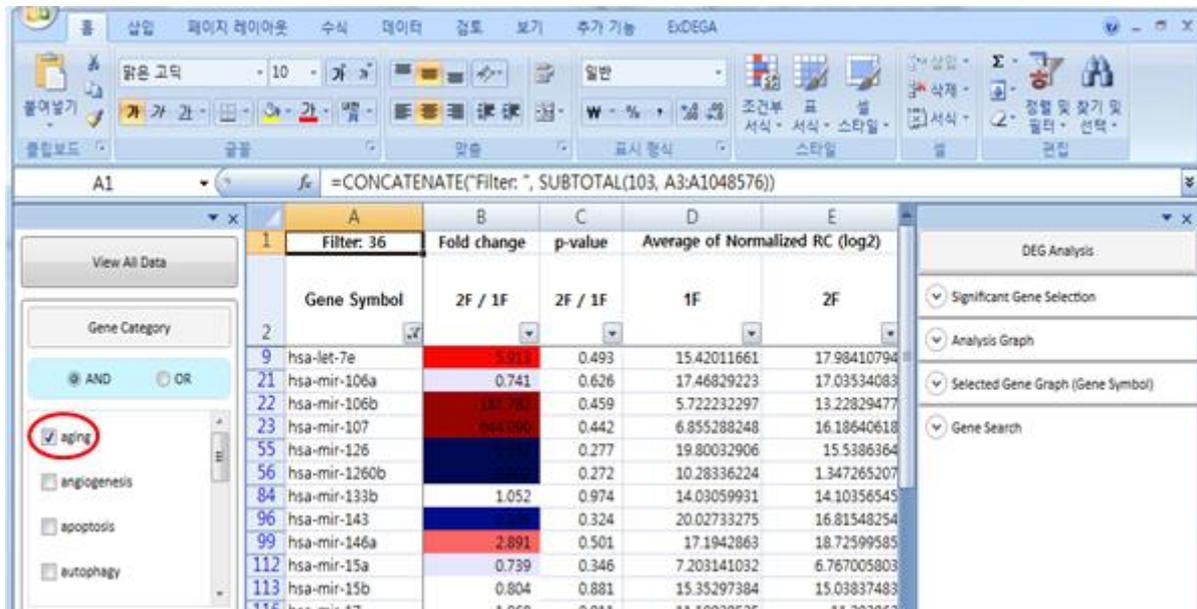


그림 1-8. miRNA Category (Aging) selection

miRNA Category 창은 중복 선택이 가능하다. And 를 선택한 상태에서 Aging 과 Angiogenesis 를 선택하면 Aging 에도 관련이 있고, Angiogenesis 에도 관련 있는 miRNA 가 필터링 된다(그림 1-9).

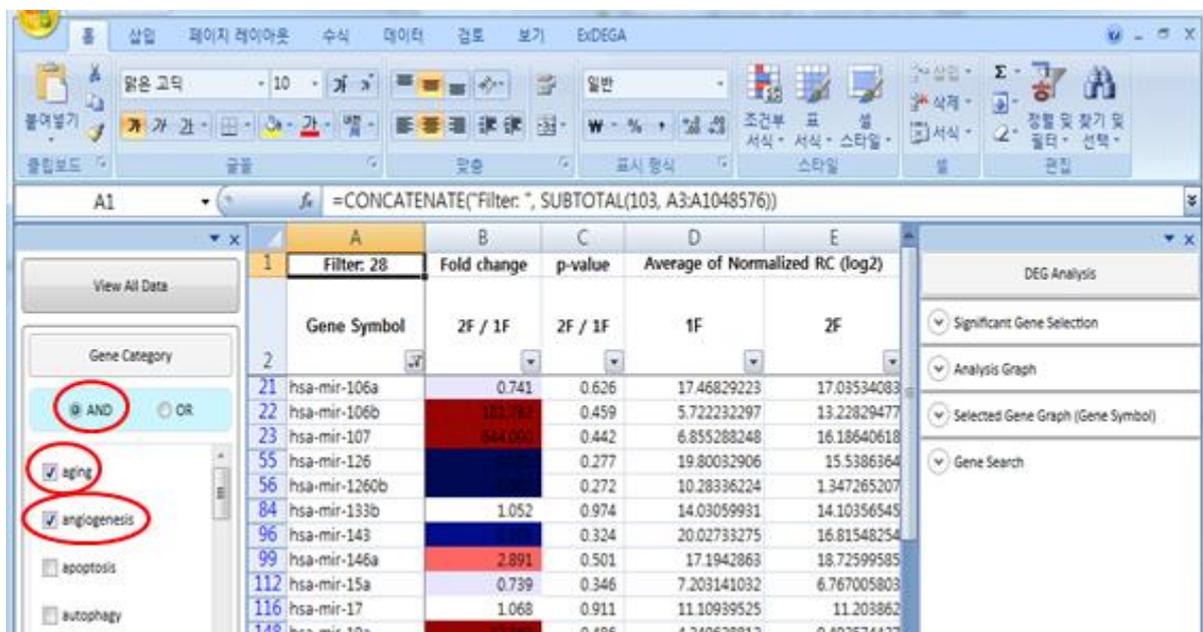


그림 1-9. miRNA Category (Aging and Angiogenesis) selection

Or 를 선택한 상태에서 Aging 과 Angiogenesis 를 선택하면 Aging 또는 Angiogenesis 에 관련된 miRNA 가 필터링 된다(그림 1-10).

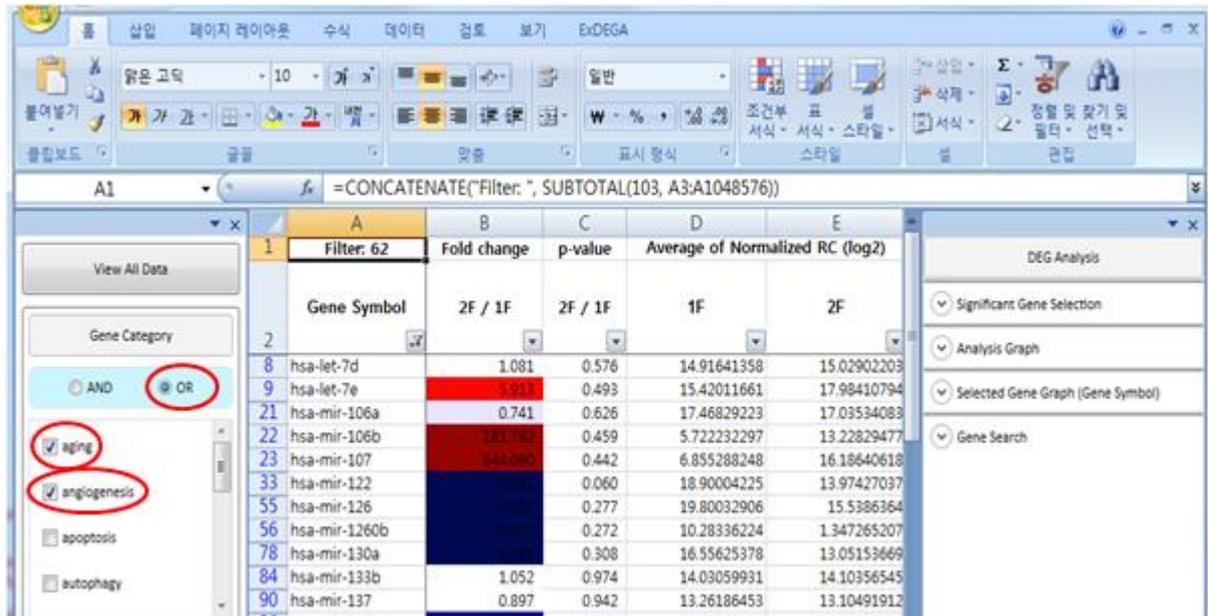


그림 1-10. miRNA Category (Aging or Angiogenesis) selection

'View All Data' 버튼을 누르면 필터를 해제하여 다시 전체 결과를 볼 수 있다.

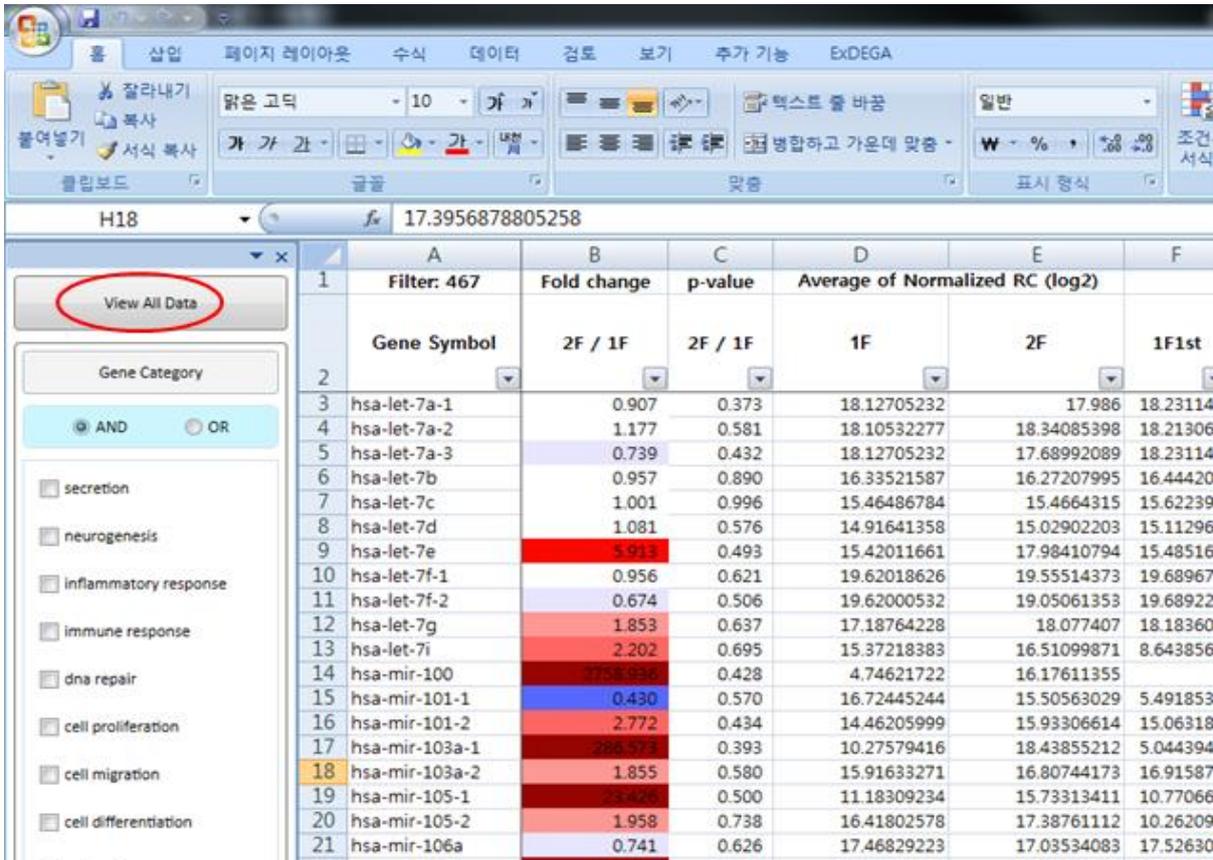


그림 1-11. 데이터 전체 다시보기

1-2. Differential expression analysis

Significant miRNA Selection 창은 전체 결과 중 control 과 test 를 비교한 결과에서 유의하게 발현 차이가 나는 miRNA 를 필터링 할 수 있도록 만들어 놓은 것이다. 예를 들어, control (A) 기준으로 test (B)에서 발현이 2 배 이상 증가 또는 감소하고, t-test 결과 p-value 값이 0.05 이하이고, normalized RC(log)값이 6 이상인 miRNA(반복 실험한 데이터의 경우)를 선택하고자 하면 그림과 같이 옵션 값을 설정해주면 된다(그림 1-12).

Gene Symbol	Fold change		p-value		Average of Normalized RC(log2)			
	B/A	D/C	B/A	D/C	A	B	C	D
20 hsa-mir-103a-1	2.421	1.871	0.0187744	0.3766344	11.9875	8.77547	3.6614	12.749
52 hsa-mir-105-1	0.492	3.096	0.0312472	0.6696446	9.14907	8.12679	6.30496	8.6397
88 hsa-mir-106b	12.788	0.933	2.12E-04	0.1144528	13.2646	16.8951	17.0833	16.983
107 hsa-mir-107	0.747	0.748	8.73E-06	0.0598366	13.6655	16.9505	17.3513	16.931
152 hsa-mir-10a	0.487	2.421	0.0250858	0.645718	13.7203	12.682	12.1922	13.467
176 hsa-mir-1-1	0.206	2.697	0.0195234	0.7914111	11.668	9.38705	9.82516	11.256
181 hsa-mir-1179	0.390	1.888	0.0053004	0.4281095	8.35773	6.99831	6.88047	7.7973
187 hsa-mir-1180	0.372	2.296	0.0212439	0.6404833	11.1527	9.72622	9.61862	10.817
193 hsa-mir-1185-1	0.464	2.160	0.0316842	0.685645	11.0571	9.94986	9.73884	10.849
200 hsa-mir-1185-2	0.591	0.708	0.0076187	0.9678769	14.6354	11.8139	10.1206	14.648
208 hsa-mir-1197	2.239	0.728	0.0011345	0.0674716	8.73315	9.92767	10.5724	10.110
215 hsa-mir-122	0.591	0.708	0.0082426	0.6693572	10.5773	7.42637	3.4649	9.758
379 hsa-mir-1224	0.428	1.671	0.0119818	0.3236517	9.8022	8.5763	8.00558	9.8816

그림 1-12. Significant miRNA selection

miRNA category selection 과 Significant miRNA selection 은 연동 가능하다. 그림 1-12 상태에서 miRNA Category selection 의 Cell cycle 을 선택하면 해당 miRNA 가 필터링 된다(그림 1-11). 선별된 miRNA 는 cell cycle 관련 유전자들 중 유의하게 발현이 증가 또는 감소한 miRNA 를 의미한다(그림 1-13).

Gene Symbol	Fold change		p-value		Average of Normalized RC(log2)			
	B/A	D/C	B/A	D/C	A	B	C	D
1571 hsa-mir-1246	0.676	0.863	1.49E-08	0.0629123	16.1358	19.3805	19.8216	19.61
4230 hsa-mir-1247	7.308	0.783	1.21E-05	0.0674095	14.4294	17.3751	17.876	17.52
4314 hsa-mir-1248	4.253	0.840	0.0475577	0.0814806	5.94168	8.03005	8.203	7.950
4591 hsa-mir-1249	7.220	0.688	1.27E-05	0.0097965	13.8326	16.4846	16.9744	16.43
5519 hsa-mir-1251	0.676	0.863	0.877	4.71E-07	0.0757178	13.4867	16.8267	17.1606
8288 hsa-mir-1258	4.311	1.364	0.0071489	0.2261625	4.38359	7.04156	6.22701	6.675
8653 hsa-mir-125a	6.538	1.522	0.0014571	0.5862774	11.4451	14.0675	13.7751	14.38
9528 hsa-mir-126	4.859	1.425	0.0013916	0.6165391	8.81015	11.0303	10.8702	11.38
13252 hsa-mir-1260b	0.676	0.863	1.001	2.13E-05	0.2714542	10.6399	14.046	14.2345
15737 hsa-mir-1264	3.709	1.205	9.31E-06	0.4953291	11.7812	15.0694	15.0802	15.3
15863 hsa-mir-1268a	0.676	0.863	0.0038226	0.4756645	15.8024	12.9768	13.0331	15.90
18212 hsa-mir-127	2.976	0.737	0.0332078	0.1283121	5.35408	6.92742	7.34692	6.907
19098 hsa-mir-1271	0.747	0.748	1.77E-06	0.6210015	13.1238	16.1496	16.8209	16.42
19100 hsa-mir-1273e	1.276	0.988	0.001259	0.945631	11.1101	14.0572	14.555	14.53
20804 hsa-mir-1273f	1.173	0.901	0.0448439	0.20393	5.14408	6.80906	7.42062	7.270
24547 hsa-mir-1273g	1.210	0.929	4.30E-05	0.7795307	17.8087	20.2174	20.8463	20.73

그림 1-13. Significant miRNAs related to cell cycle

실험 결과에 따라 발현 변화값 (fold change), p-value, normalized RC(log2) 기준을 조정할 수 있다. 반복 실험이 아닌 경우에는 p-value 가 없다.

'Gene Ontology Graph' 버튼을 누르면 각 Category 관련 miRNA 중 발현이 유의하게 차이나는 miRNA 의 %와 수가 그래프로 그려진다. 본 분석을 통해 어떤 Category 의 miRNA 들이 상대적으로 많은 발현 변화가 있었는지를 확인할 수 있다. Significant miRNA Selection 에서 선택한 조건에 따라 GO graph 가 변화된다. 그래프의 각 영역을 마우스로 올리면 해당 miRNA 수를 확인할 수 있고 클릭하면 해당 miRNA 들이 Data sheet 에 필터링 된다. 예를 들어 Pie 그래프의 특정영역을 클릭하면 해당 리스트가 필터링 되고 Cell cycle 관련 miRNA 중 up 이 되는 miRNA 그래프를 클릭하면 Data sheet 에 해당 miRNA 가 필터링 된다(그림 1-14).

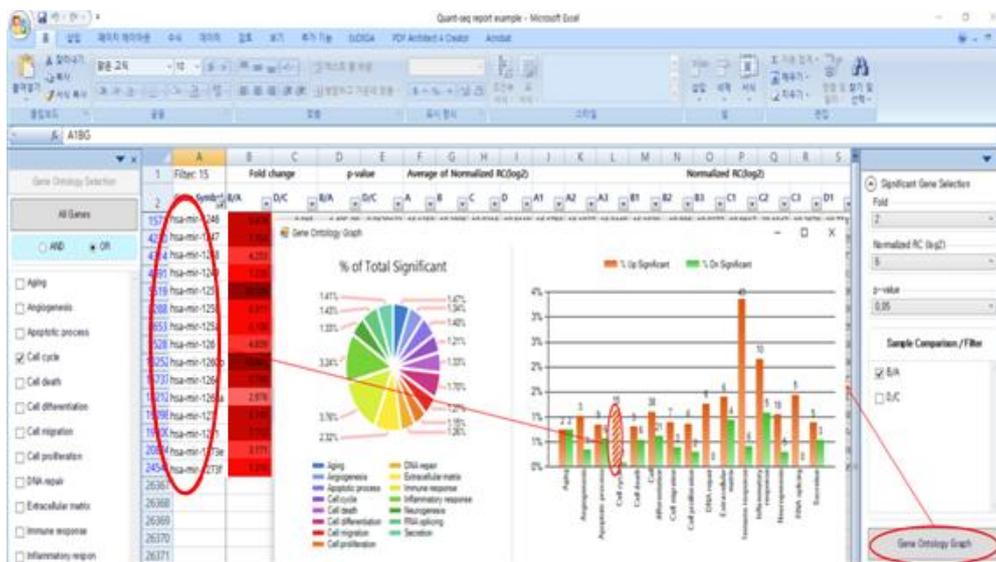


그림 1-14. Up-regulated genes related to aging

선별한 miRNA 의 ID 를 복사하여 Selected gene graph 창에 붙여 넣고 'Graph View'를 누르면 normalized RC(log2) 값, fold change 값으로 line graph 가 그려진다(그림 1-15).

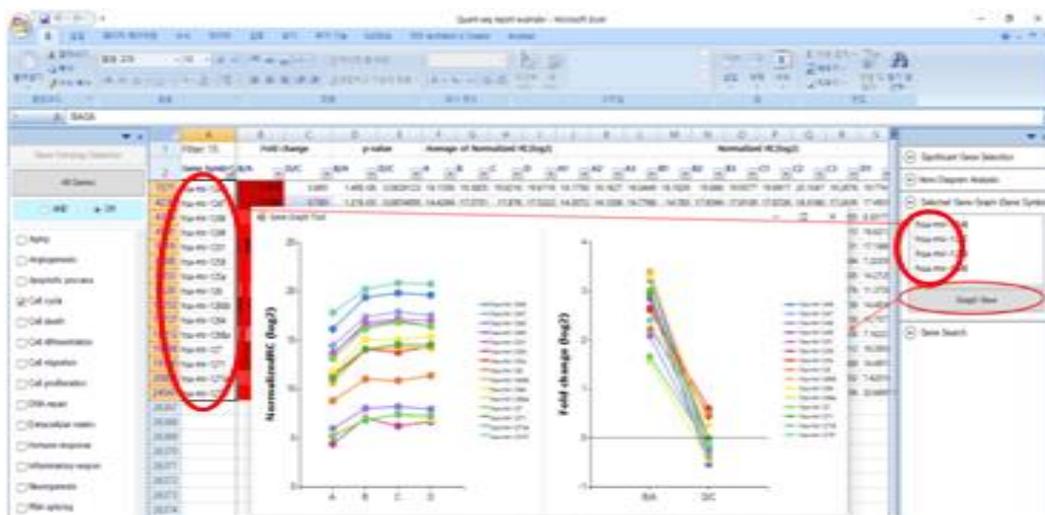


그림 1-15. Gene graph

특정 keyword 관련 miRNA 를 검색하고 싶을 때는 gene search 창을 이용한다. 예를 들어 'insulin'을 검색하면 insulin 관련 miRNA 가 필터링 된다(그림 1-16).

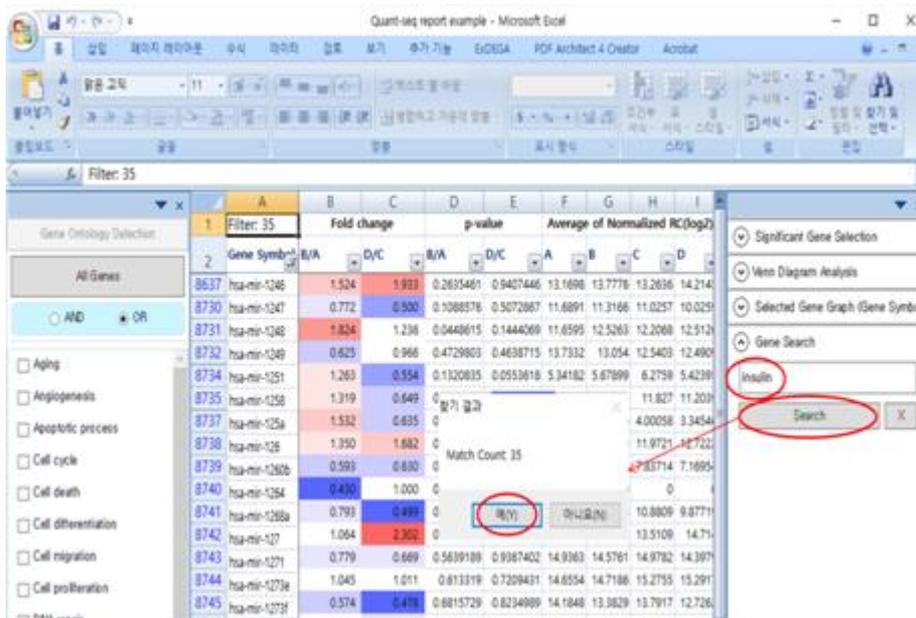


그림 1-16. Genes related to insulin

ExDEGA 에서는 Venn Diagram analysis 를 통해 Venn Diagram 을 작성할 수 있다. Significant miRNA Selection 부분의 화살표를 눌러 최소화 한 후, Venn Diagram analysis 의 Gene View 를 클릭한다(그림 1-17).

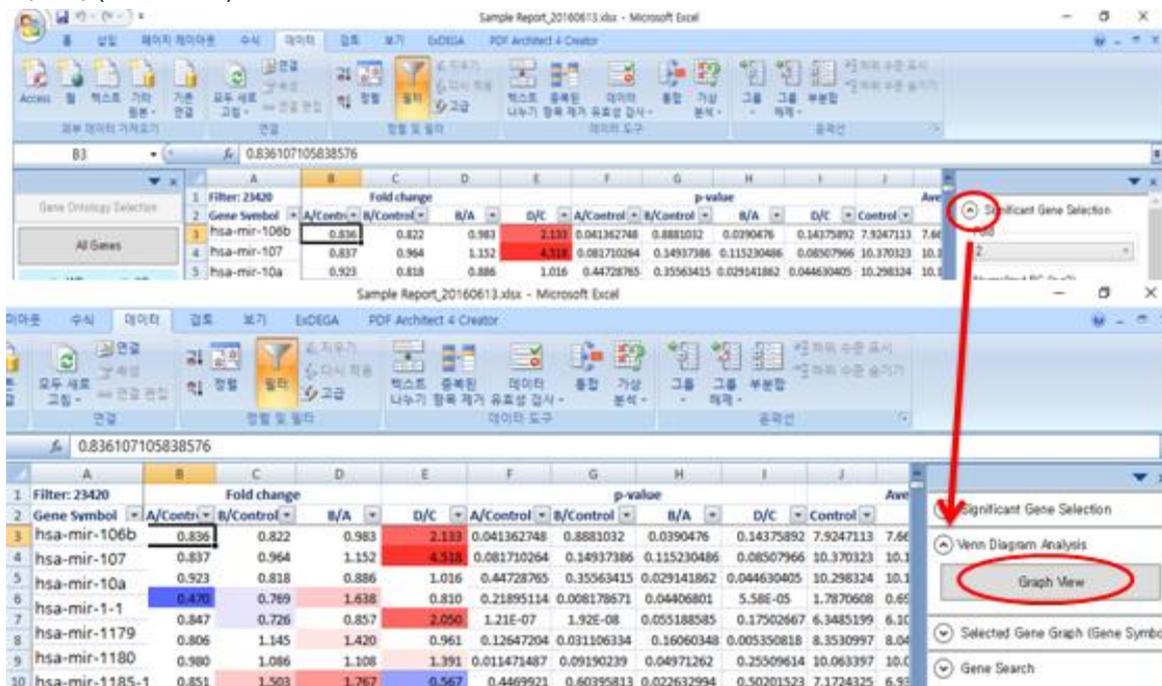


그림 1-17. Select Venn Diagram Analysis

Venn Diagram 을 그릴 그룹과 fold change(fold 부분), Normalized RC(log2), p-value 를 선택 후, Diagram View 를 클릭하면 결과를 확인할 수 있으며, 그룹은 최대 4 그룹까지 선택 가능하다(그림 1-18).

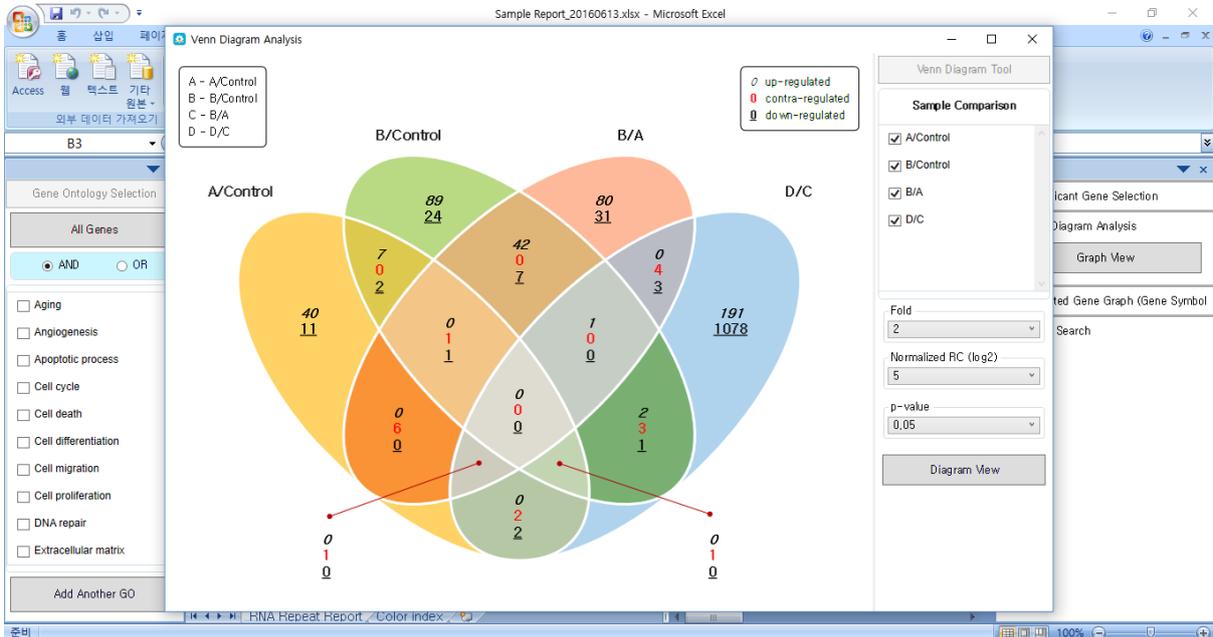


그림 1-18. Result of 4 groups Venn Diagram Analysis

아래의 그림은 A/Control, B/Control, B/A 결과 중, 2fc 이상 up,down 되면서 Normalized RC(log2)는 5 이상, p-value 는 0.05 이하인 list 를 가지고 Venn Diagram 을 작성한 결과이다(그림 1-19). 해당 기준들은 연구자가 보고자 하는 기준에 따라 언제든지 변경할 수 있다.

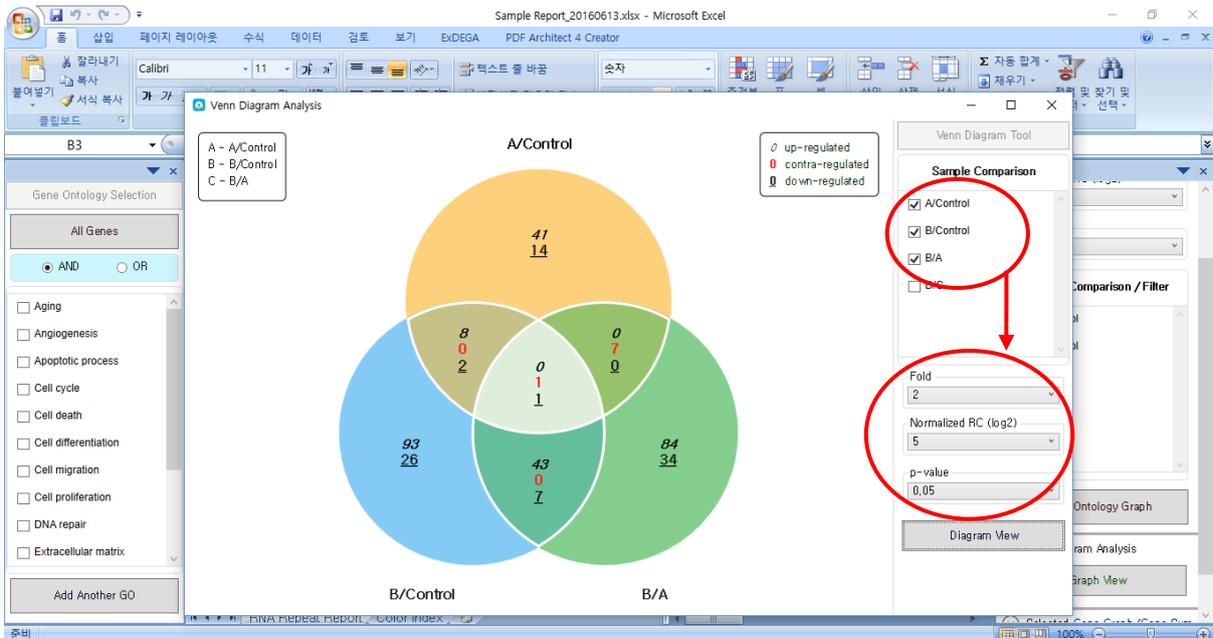


그림 1-19. Venn Diagram View

Venn Diagram 결과에서 표시되는 형식은 다음과 같다(그림 1-20).

1. **기울어진 숫자** : 2fold 이상 up-regulated 된 miRNA 수
(A/Control 과 B/Control 에서 공통적으로 2fold 이상 up 되는 miRNA **8** 개)
2. **빨간색 숫자** : regulation 이 대조되는 miRNA 수
(A/Control 에서는 2fold 이상 up 되고, B/Control 에서 2fold 이상 down 되면서
A/Control 에서는 2fold 이상 down 되고, B/Control 에서는 2fold 이상 up 되는 miRNA **0** 개)
3. **밑줄 친 숫자** : 2fold 이상 down-regulated 된 miRNA 수
(A/Control 과 B/Control 에서 공통적으로 2fold 이상 down 되는 miRNA **2** 개)

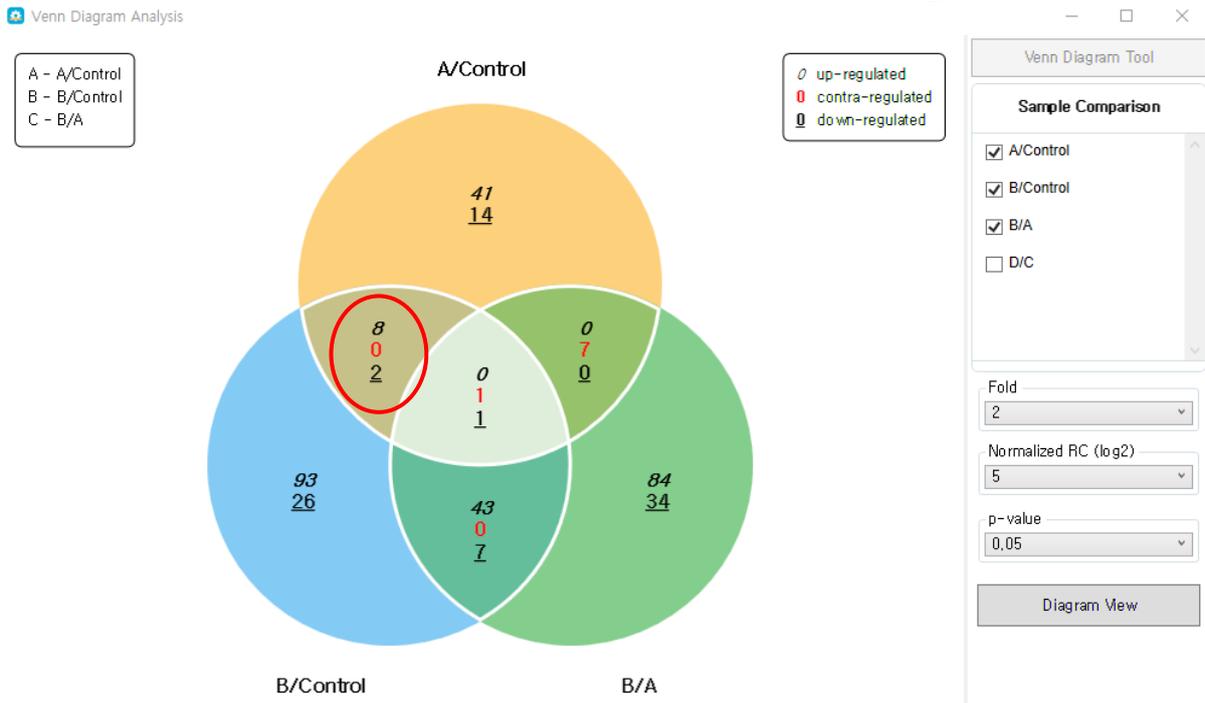


그림 1-20. For example of up,down,contra-regulated in Venn Diagram

Venn Diagram 이미지를 오른쪽 클릭하면 Venn Diagram 각 영역에 어떤 miRNA 들이 있는지 확인할 수 있다. 예를 들어, A/Control 에서만 2fold up 되면서 Normalized RC(log2)는 5 이상, p-value 는 0.05 이하가 되는 miRNA 를 보고 싶으면, Venn Diagram 에서 A/Control 에서만 해당되는 영역을 찾아 마우스 오른쪽 클릭 하면 2 fold up 된 miRNA list 가 엑셀 sheet 에 filter 된다(그림 1-21).

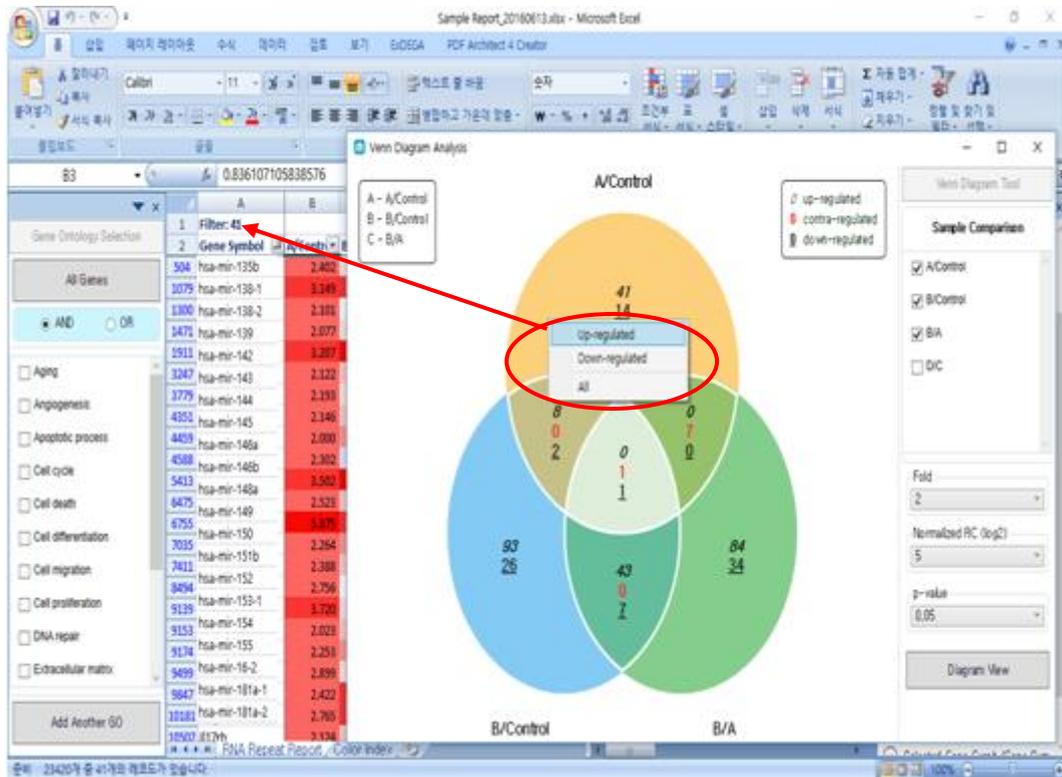


그림 1-21. Filtering 2fold up-regulated miRNA list in Venn Diagram

2. Functional category별 miRNA 분석

miRWalk 2.0 (<http://zmf.umm.uni-heidelberg.de/apps/zmf/mirwalk2/index.html>)을 이용하면 functional category 별로 관련 miRNA를 분석 할 수 있다. Predicted Target Module에서 gene ontology (GO), pathway, gene class 등의 관련 miRNA를 찾을 수 있다(그림2-1).

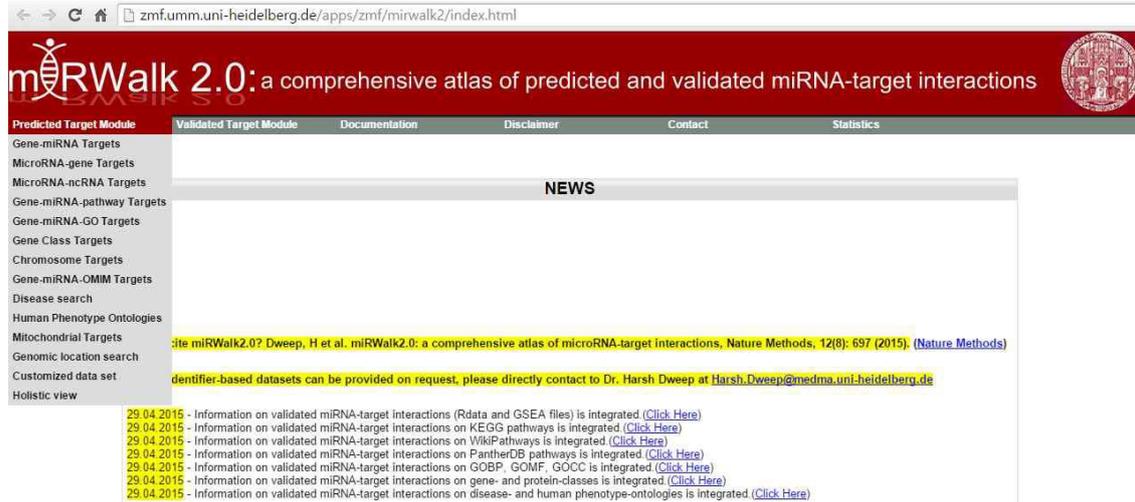


그림 2-1. miRWalk 2.0 site

Pathway 관련 miRNA를 검색하고 싶으면 gene-miRNA-pathway targets 항목을 선택한다. 종 (species)과 pathway database를 선택하고 관심있는 pathway를 선택한다(그림2-2). 다른 miRNA database 기반으로도 pathway 관련 miRNA를 분석하고 싶으면 other database에서 원하는 database를 선택한다.

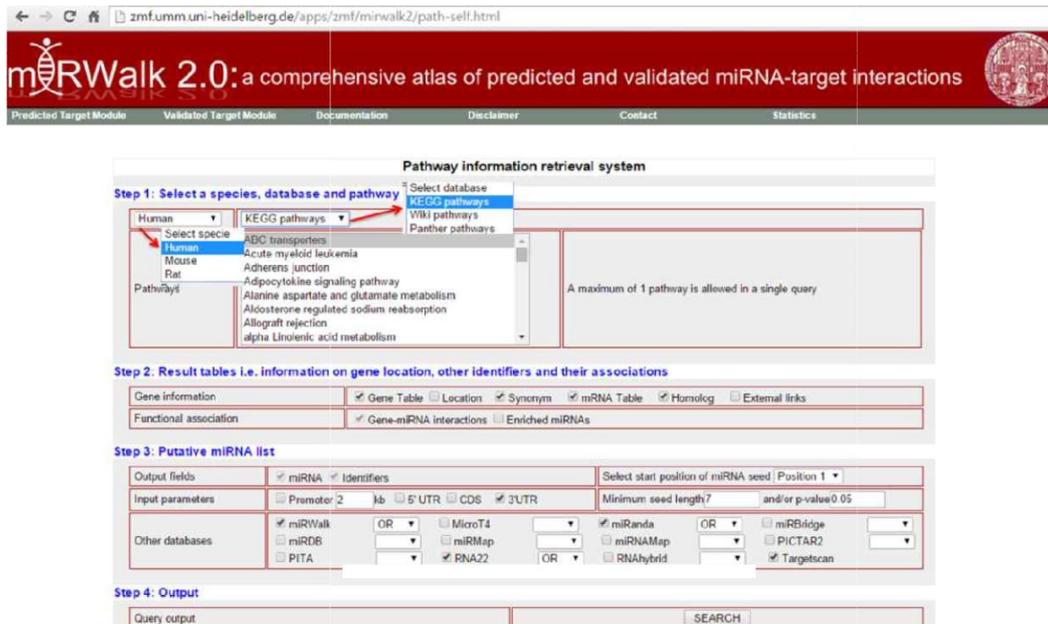


그림 2-2. pathway miRNA search in miRWalk 2.0 site

Search를 누르면 결과 창이 뜬다. Gene table에서는 해당 pathway의 관련 유전자 정보를 확인할 수 있다(그림2-3).

Pathway-gene:miRNA interaction information retrieval system

Gene information
[Gene Table](#) | [Synonymous](#) | [Refseq Table](#) | [Homologous Table](#)

Putative miRNA-target interaction information on selected kegg pathway(s)

Putative miRNA binding sites predicted by miRWalk algorithm within mRNA selected regions

Putative miRNA binding sites predicted by chosen algorithms within mRNA selected regions

Gene Information Table

[Download Table](#)

Pathway	Accession	Gene (EntrezID)	Taxonomy
ABC transporters	hsa02010	ABCA1(19) , ABCA10(10349) , ABCA12(26154) , ABCA13(154564) , ABCA2(20) , ABCA3(21) , ABCA4(24) , ABCA5(23461) , ABCA6(23460) , ABCA7(10347) , ABCA8(10351) , ABCA9(10350) , ABCB1(5243) , ABCB10(23456) , ABCB11(8647) , ABCB4(5244) , ABCB5(340273) , ABCB6(10058) , ABCB7(22) , ABCB8(11194) , ABCB9(23457) , ABCC1(4363) , ABCC10(69845) , ABCC11(85320) , ABCC12(94160) , ABCC2(1244) , ABCC3(8714) , ABCC4(10257) , ABCC5(10957) , ABCC6(365) , ABCC6(6833) , ABCC7(10060) , ABCC11(215) , ABCC12(25) , ABCC3(5825) , ABCC4(5826) , ABCC11(9619) , ABCC2(5429) , ABCC4(64137) , ABCC5(64240) , ABCC6(64241) , CFTR(1080) , TAP1(6890) , TAP2(5891)	9606

그림 2-3. Genes related to ABC transporter pathway search in miRWalk 2.0 site

Putative miRNA binding sites predicted by miRWalk algorithm within mRNA selected regions 표의 3'UTR을 누르면 miRWalk 분석 결과로 나온 해당 pathway의 관련 miRNA 정보를 확인할 수 있다(그림2-4). Download complete table을 눌러 결과를 텍스트 파일로 저장한다. 이 텍스트 파일은 엑셀에서 열 수 있다.

Pathway-gene:miRNA interaction information retrieval system

Gene information
[Gene Table](#) | [Synonymous](#) | [Refseq Table](#) | [Homologous Table](#)

Putative miRNA-target interaction information on selected kegg pathway(s)

Putative miRNA binding sites predicted by miRWalk algorithm within mRNA selected regions

Putative miRNA binding sites predicted by chosen algorithms within mRNA selected regions

Putative miRNA binding site predictions within 3'-UTR region from position 1 in kegg pathways

*The first 50 interactions are displayed in the below table, however, all the interactions can be downloaded by clicking on the below links.

[Download Gene-miRNA Pairs Table](#) | [Download Refseq-miRNA Pairs Table](#) | [Download Complete Table](#) | [A comparative view from position 1 to 5](#)

SPMS: 'SL' 'SeedS' and 'SeedE' stand for starting position of a miRNA seed, seed length, seed start and end positions.

Pathway	Gene	EntrezID	RefseqID	miRNA	MIMATid	SPMS	SL	SeedS	SeedE	pvalue
ABC transporters	ABCA2	20	NM_225332	hsa-miR-c726-5p	MIMAT0019645	1	16	7914	7695	0.0000
ABC transporters	ABCA2	20	NM_001606	hsa-miR-d726-5p	MIMAT0019645	1	16	7922	7907	0.0000
ABC transporters	ABCB5	344273	NM_178559	hsa-miR-5095	MIMAT0020603	1	15	3829	3615	0.0000
ABC transporters	ABCB5	344273	NM_001163941	hsa-miR-5095	MIMAT0020603	1	15	4855	4841	0.0000
ABC transporters	ABCC9	344273	XM_005249737	hsa-miR-5095	MIMAT0020603	1	15	3775	3761	0.0000
ABC transporters	ABCA13	154664	XM_005249624	hsa-miR-4459	MIMAT0018581	1	13	5274	9262	0.0000
ABC transporters	ABCC1	4363	XM_005255326	hsa-miR-4794-3p	MIMAT0027489	1	13	5671	5655	0.0000
ABC transporters	ABCA5	23461	NM_019872	hsa-miR-4482-3p	MIMAT0020558	1	13	9026	9014	0.0000
ABC transporters	ABCA5	23461	NM_172232	hsa-miR-4482-3p	MIMAT0020558	1	13	7964	7942	0.0000
ABC transporters	ABCA13	154664	NM_194701	hsa-miR-4459	MIMAT0018581	1	13	16303	16291	0.0000
ABC transporters	ABCC1	4363	NM_004896	hsa-miR-4794-3p	MIMAT0027489	1	13	6723	5711	0.0000
ABC transporters	ABCC1	4363	XM_005255329	hsa-miR-4794-3p	MIMAT0027489	1	13	5464	5452	0.0000
ABC transporters	ABCC1	4363	XM_005255328	hsa-miR-4794-3p	MIMAT0027489	1	13	5563	5491	0.0000
ABC transporters	ABCA13	154664	XM_005249622	hsa-miR-4459	MIMAT0018581	1	13	9367	9355	0.0000
ABC transporters	ABCG4	64137	NM_001142525	hsa-miR-4613-5p	MIMAT0027526	1	13	2064	2072	0.0000
ABC transporters	ABCC1	4363	XM_005255327	hsa-miR-4794-3p	MIMAT0027489	1	13	5515	5503	0.0000
ABC transporters	ABCG4	64137	NM_022169	hsa-miR-4613-5p	MIMAT0027526	1	13	2368	2296	0.0000
ABC transporters	ABCG2	9429	NM_004827	hsa-miR-3124-3p	MIMAT0019200	1	12	3779	3768	0.0001
ABC transporters	ABCA1	19	XM_005261776	hsa-miR-3614-3p	MIMAT0017591	1	12	7129	7128	0.0002
ABC transporters	ABCA8	10351	XM_005260941	hsa-miR-312-3p	MIMAT0000763	1	12	3967	3896	0.0000

그림 2-4. miRNAs related to ABC transport pathway search in miRWalk 2.0 site

Putative miRNA binding sites predicted by chosen algorithms within mRNA selected regions 표의 3'UTR을 누르면 여러 miRNA database (miRanda, RNA22, Targetscan 등)를 기반으로 해당 pathway의 관련 miRNA 정보를 확인할 수 있다(그림2-5).

miRWalk 2.0: a comprehensive atlas of predicted and validated miRNA-target interactions

Pathway-gene:miRNA interaction information retrieval system

Gene information
[Gene Table](#) [Synonymous](#) [Refseq Table](#) [Homologous Table](#)

Putative miRNA-target interaction information on selected kegg pathway(s)

Putative miRNA binding sites predicted by miRWalk algorithm within mRNA selected regions

Not selected (Promoter)	Not selected (5'UTR)	Not selected (CDS)	3'UTR
-------------------------	----------------------	--------------------	--------------

Putative miRNA binding sites predicted by chosen algorithms within mRNA selected regions

Not selected (Promoter)	Not selected (5'UTR)	Not selected (CDS)	3'UTR
-------------------------	----------------------	--------------------	--------------

Comparative platform of miRNA binding site predictions on kegg pathways within 3'-UTR region resulting from different algorithms

the first 50 interactions are displayed in the below table, however, all the interactions can be downloaded by clicking on the below link
[Download Complete Table](#)

Pathway	Gene	EntrezID	RefseqID	miRNA	miRWalk	miRanda	RNA22	Targetscan	SUM
ABC transporters	ABCA1	19	NM_005502	hsa-miR-449b-5p					4
ABC transporters	ABCA1	19	NM_005502	hsa-miR-1257					4
ABC transporters	ABCA1	19	NM_005502	hsa-miR-450b-5p					4
ABC transporters	ABCA1	19	NM_005502	hsa-miR-1264					4
ABC transporters	ABCA1	19	NM_005502	hsa-miR-1265					4
ABC transporters	ABCA1	19	NM_005502	hsa-miR-1265p					4
ABC transporters	ABCA1	19	NM_005502	hsa-miR-1267					4
ABC transporters	ABCA1	19	NM_005502	hsa-miR-1271-5p					4
ABC transporters	ABCA1	19	NM_005502	hsa-miR-1273e					4
ABC transporters	ABCA1	19	NM_005502	hsa-miR-498					4
ABC transporters	ABCA1	19	NM_005502	hsa-miR-499a-5p					4
ABC transporters	ABCA1	19	NM_005502	hsa-miR-515-3p					4
ABC transporters	ABCA1	19	NM_005502	hsa-miR-518a-5p					4
ABC transporters	ABCA1	19	NM_005502	hsa-miR-1283					4
ABC transporters	ABCA1	19	NM_005502	hsa-miR-519d-3p					4
ABC transporters	ABCA1	19	NM_005502	hsa-miR-520a-3p					4
ABC transporters	ABCA1	19	NM_005502	hsa-miR-520b					4

그림 2-5. miRNAs related to ABC transporter pathway search in other miRNA database

3. miRNA target gene 분석

miRWalk 2.0 (<http://zmf.umm.uni-heidelberg.de/apps/zmf/mirwalk2/index.html>)을 이용하여 한번에 20 개까지 후보 miRNA 들을 입력하여 target gene 분석을 빠르고 쉽게 할 수 있다. Predicted Target Module 에서 MicroRNA-gene Targets 항목을 클릭하면 된다.(그림 3-1).

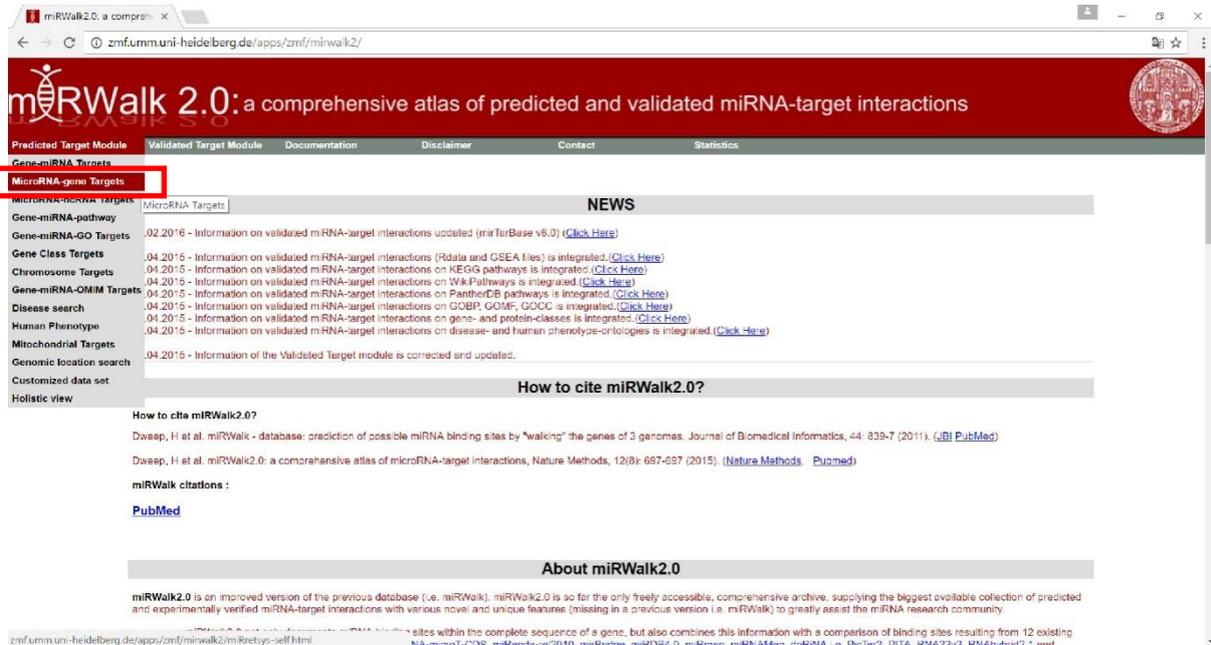


그림 3-1. miRWalk 2.0 site and MicroRNA-gene Targets start

아래 microRNA information retrieval system 에서 species 를 선택하고 database 선택을 miRBase 로 하고 "Paste identifier" 창에 mature miRNA acc. No 를 복사붙여넣기를 한다. 한번에 입력할 수 있는 miRNA 개수는 최대 20 개 까지 가능하다.

그리고 Step 2 ~ Step 5 를 수행하면 되는데 일반적으로 기본값으로 수행하면 된다.

모든 설정이 마무리 되면 "SEARCH" 항목을 클릭한다.[그림 3-2]

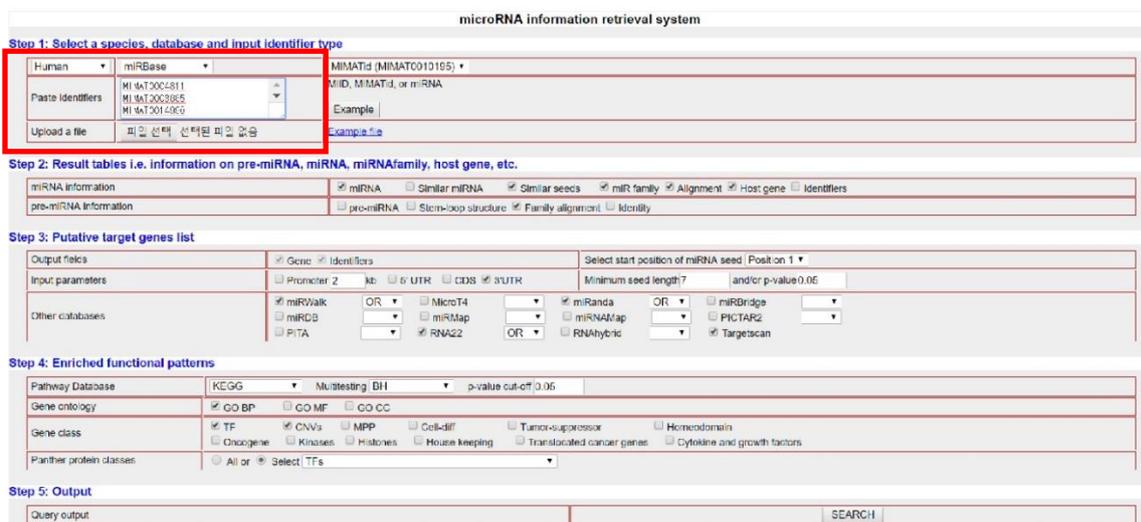


그림 3-2. microRNA input and target gene search

아래 결과 창이 나오면 “Putative target genes tables predicted by chosen algorithms within mRNA selected regions” 에서 “3UTR” 링크를 클릭하면 결과 페이지를 확인할 수 있다. [그림 3-3]

microRNA information retrieval system

Information on mature miRNAs

miRNA Table	Not selected	Similar Seeds
miRNA Family	miRNA Alignment	Host gene Not selected

Information on pre-miRNA

Not selected	Alignment	Not selected
--------------	-----------	--------------

Putative target genes tables

Putative target genes predicted by miRWalk algorithm within mRNA selected regions			
Promoter	5UTR	CDS	3UTR
Not selected	Not selected	Not selected	3UTR

Putative target genes predicted by chosen algorithms within mRNA selected regions			
Promoter	5UTR	CDS	3UTR
Not selected	Not selected	Not selected	3UTR

Enriched functional patterns information

KEGG pathways			
Pathways	Not selected	Not selected	Not selected 3UTR Download

Gene Ontologies			
GO BP	Not selected	Not selected	Not selected 3UTR Download

Gene Classes			
Transcription factors (TF)	Not selected (Promoter)	Not selected (5'-UTR)	Not selected (CDS) 3UTR Download
Copy number variants (CNVs)	Not selected (promoter)	Not selected (5'-UTR)	Not selected (CDS) 3UTR Download

Panther protein classification			
Protein classes	Not selected	Not selected	Not selected 3UTR Download

그림 3-3. microRNA target links

아래 결과 페이지에서 “Download Complete Table”을 클릭하여 전체 결과를 텍스트 파일을 저장한다. 그리고 저장한 파일은 엑셀에서 확인할 수 있다.[그림 3-4]

Comparative platform of miRNA binding site predictions within 3'-UTR region resulting from different algorithms

the first 50 interactions are displayed in the below table, however, all the interactions can be downloaded by clicking on the below links.

[Download Gene symbol Table](#)
 [Download RefseqID Table](#)
 [Download Complete Table](#)

miRNA	Gene	EntrezID	RefseqID	miRWalk	miRanda	RNA22	Targetscan	SUM
hsa-miR-3127-5p	ADCYAP1R1	117	NM_001136935					4
hsa-miR-3127-5p	ADORA1	134	XM_005244301					4
hsa-miR-3127-5p	XIAF	331	NM_0011167					4
hsa-miR-3127-5p	AQP2	358	NM_000486					1
hsa-miR-3127-5p	ARE3	377	XM_005288966					4
hsa-miR-3127-5p	ALDH7A1	501	NM_001132					4
hsa-miR-3127-5p	ALDH7A1	501	NM_001201377					4
hsa-miR-3127-5p	ATP1B1B2	626	NM_001883					1
hsa-miR-3127-5p	BCL2	596	NM_000833					4
hsa-miR-3127-5p	BCL2L1	598	NM_138578					4
hsa-miR-3127-5p	BCL2L2	599	NM_004050					4
hsa-miR-3127-5p	BDKR2	624	XM_005287980					1
hsa-miR-3127-5p	CACNA1E	777	XM_005245477					4
hsa-miR-3127-5p	CAMK2B	816	NM_001220					4
hsa-miR-3127-5p	TPP1	1200	NM_000391					4
hsa-miR-3127-5p	DCX	1841	NM_000566					1
hsa-miR-3127-5p	DIO3	1735	NM_001362					4
hsa-miR-3127-5p	SLC26A2	1836	NM_000112					4
hsa-miR-3127-5p	E2F2	1974	NM_001950					4
hsa-miR-3127-5p	CELSR2	1962	NM_001408					1
hsa-miR-3127-5p	EGFR	1966	XM_005271748					4
hsa-miR-3127-5p	E1F4G2	1982	NM_001172705					4
hsa-miR-3127-5p	ERBB2	2064	XM_005287139					4
hsa-miR-3127-5p	FRG	2078	NM_001243128					1
hsa-miR-3127-5p	ETS1	2113	NM_005238					4
hsa-miR-3127-5p	ETV5	2119	XM_005247210					4
hsa-miR-3127-5p	EVC	2121	NM_153717					4
hsa-miR-3127-5p	EZR	2149	NM_001982					1

그림 3-4. microRNA target result page

Target gene 분석결과를 엑셀에서 [그림 3-5]와 같이 확인할 수 있는데 여기서 관심 miRNA 와 Target DB 를 선택하면 된다.

	A	B	C	D	E	F	G	H	I	J	K
1	miRNA	MIMATid	Gene	Entrezl	Refseq	miRwa	miRand	RNA22	Targets	SUM	
2	hsa-miR-3127-5p	MIMAT0014990	ACVR1B	91	NM_02032	1	1	1	1	4	
3	hsa-miR-3127-5p	MIMAT0014990	ADCYAP1F	117	NM_00119	1	1	1	1	4	
4	hsa-miR-3127-5p	MIMAT0014990	ADORA1	134	XM_00524	1	1	1	1	4	
5	hsa-miR-3127-5p	MIMAT0014990	XIAP	331	NM_00116	1	1	1	1	4	
6	hsa-miR-3127-5p	MIMAT0014990	AQP2	359	NM_00048	1	1	1	1	4	
7	hsa-miR-3127-5p	MIMAT0014990	ARF3	377	XM_00526	1	1	1	1	4	
8	hsa-miR-3127-5p	MIMAT0014990	ALDH7A1	501	NM_00118	1	1	1	1	4	
9	hsa-miR-3127-5p	MIMAT0014990	ALDH7A1	501	NM_00120	1	1	1	1	4	
10	hsa-miR-3127-5p	MIMAT0014990	ATP6V1B2	526	NM_00169	1	1	1	1	4	
11	hsa-miR-3127-5p	MIMAT0014990	BCL2	596	NM_00063	1	1	1	1	4	
12	hsa-miR-3127-5p	MIMAT0014990	BCL2L1	598	NM_13857	1	1	1	1	4	
13	hsa-miR-3127-5p	MIMAT0014990	BCL2L2	599	NM_00405	1	1	1	1	4	
14	hsa-miR-3127-5p	MIMAT0014990	BDKRB2	624	XM_00526	1	1	1	1	4	
15	hsa-miR-3127-5p	MIMAT0014990	CACNA1E	777	XM_00524	1	1	1	1	4	
16	hsa-miR-3127-5p	MIMAT0014990	CAMK2B	816	NM_00122	1	1	1	1	4	
17	hsa-miR-3127-5p	MIMAT0014990	TPP1	1200	NM_00035	1	1	1	1	4	
18	hsa-miR-3127-5p	MIMAT0014990	DCX	1641	NM_00055	1	1	1	1	4	
19	hsa-miR-3127-5p	MIMAT0014990	DIO3	1735	NM_00136	1	1	1	1	4	
20	hsa-miR-3127-5p	MIMAT0014990	SLC26A2	1836	NM_00017	1	1	1	1	4	
21	hsa-miR-3127-5p	MIMAT0014990	E2F4	1874	NM_00195	1	1	1	1	4	
22	hsa-miR-3127-5p	MIMAT0014990	CELSR2	1952	NM_00140	1	1	1	1	4	
23	hsa-miR-3127-5p	MIMAT0014990	EGFR	1956	XM_00527	1	1	1	1	4	
24	hsa-miR-3127-5p	MIMAT0014990	EIF4G2	1982	NM_00117	1	1	1	1	4	
25	hsa-miR-3127-5p	MIMAT0014990	ERBB2	2064	XM_00525	1	1	1	1	4	
26	hsa-miR-3127-5p	MIMAT0014990	ERG	2078	NM_00124	1	1	1	1	4	
27	hsa-miR-3127-5p	MIMAT0014990	ETS1	2113	NM_00523	1	1	1	1	4	
28	hsa-miR-3127-5p	MIMAT0014990	ETV5	2119	XM_00524	1	1	1	1	4	

그림 3-5. microRNAtarget gene complete table

엑셀의 필터기능을 적용하여 [그림 3-6]과 같이 “miRNA” 항목에서 원하는 miRNA 를 선택하고 “SUM”항목에서 DB 의 숫자를 선택한다. 여기서 DB 의 숫자는 기본 4 개의 DB 에서 target gene 을 포함하는 DB 의 숫자를 의미한다.

The screenshot shows the Excel spreadsheet with filter menus open for columns A and J. Column A (miRNA) has a dropdown menu with 'hsa-miR-3127-5p' selected. Column J (SUM) has a dropdown menu with '4' selected. A red box highlights the 'miRNA' dropdown, and another red box highlights the 'SUM' dropdown.

그림 3-6. microRNAtarget gene complete table

4. DAVID tool을 이용한 target gene set 분석

DAVID 는 다양한 데이터 베이스를 기반으로 유전자의 상관관계를 통계적으로 분석하여 유전자의 주요 기능을 예측하는 analysis tool 이다. 분석과정은 그림 4-1 과 같다.

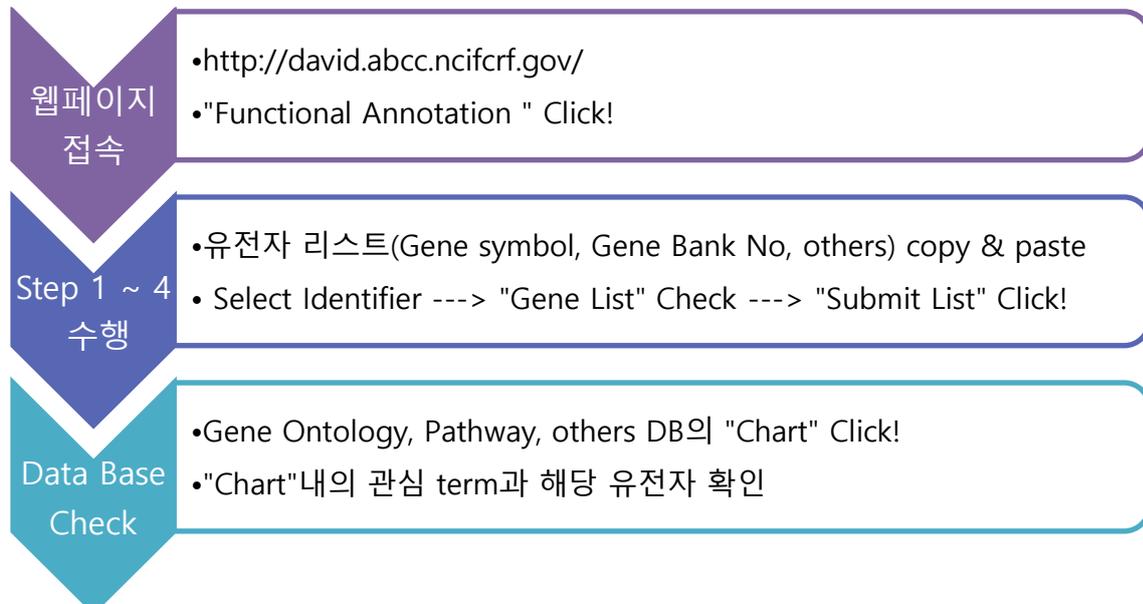


그림 4-1. DAVID tool analysis process

DAVID 에서는 3 천 개 이상의 유전자는 분석할 수 없으므로 3 천 개 이하로 유전자를 선별해야 한다. MiRNA-Seq 결과에서 significant gene 을 선별하여 DAVID 분석을 한다. DAVID 홈페이지 (<http://david.abcc.ncifcrf.gov/>)에 접속하여 "Functional Annotation"을 클릭한다(그림 4-2).

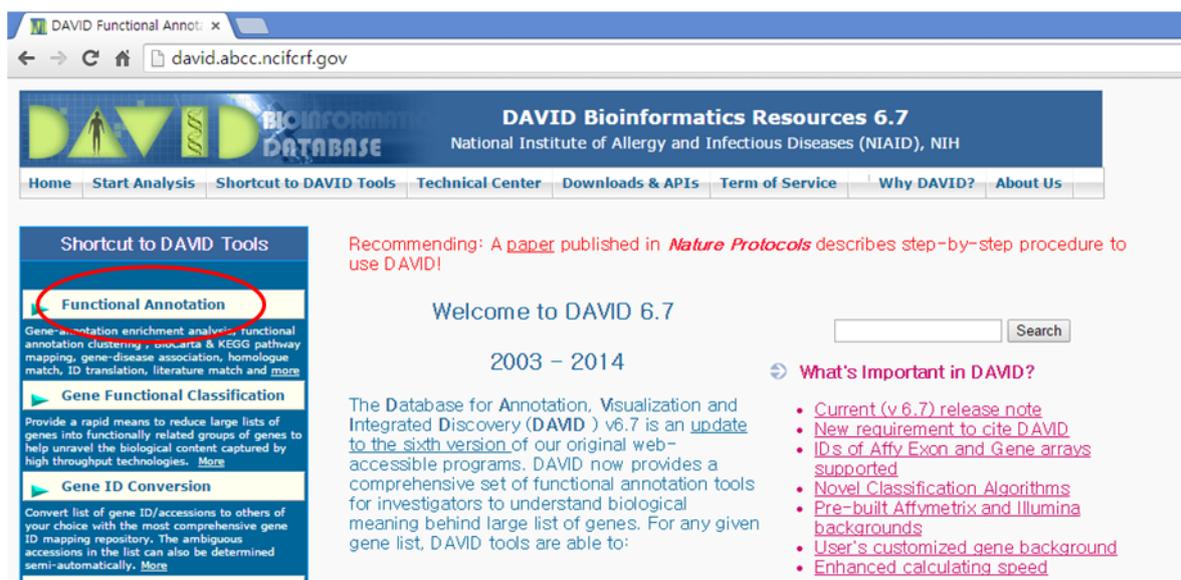


그림 4-2. DAVID tool webpage

“Upload” 탭에서 Step 1 에서 Step 4 까지 수행한다(그림 4-3). Step 1 에서 선별한 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 4-3. DAVID tool : Step 1 ~ Step 4

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

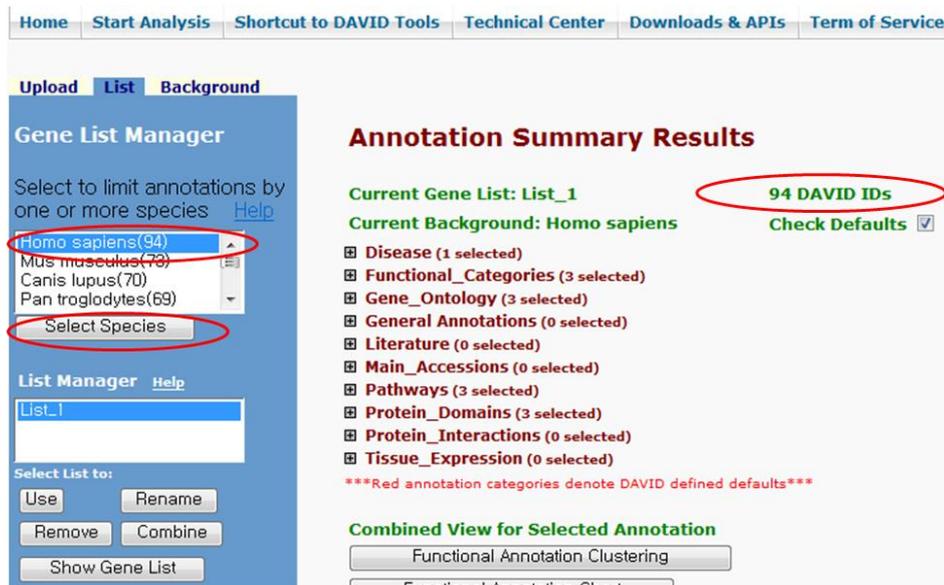


그림 4-4. DAVID tool : Select Species

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 GO 를 확인할 수 있다(그림 4-5). 관심 GO 를 클릭하면 QuickGO 데이터베이스로 연결되어 각 GO 의 정보를 확인할 수 있다. GO 의 Gene 막대를 클릭하면 해당 GO 관련 유전자들을 확인할 수 있다.

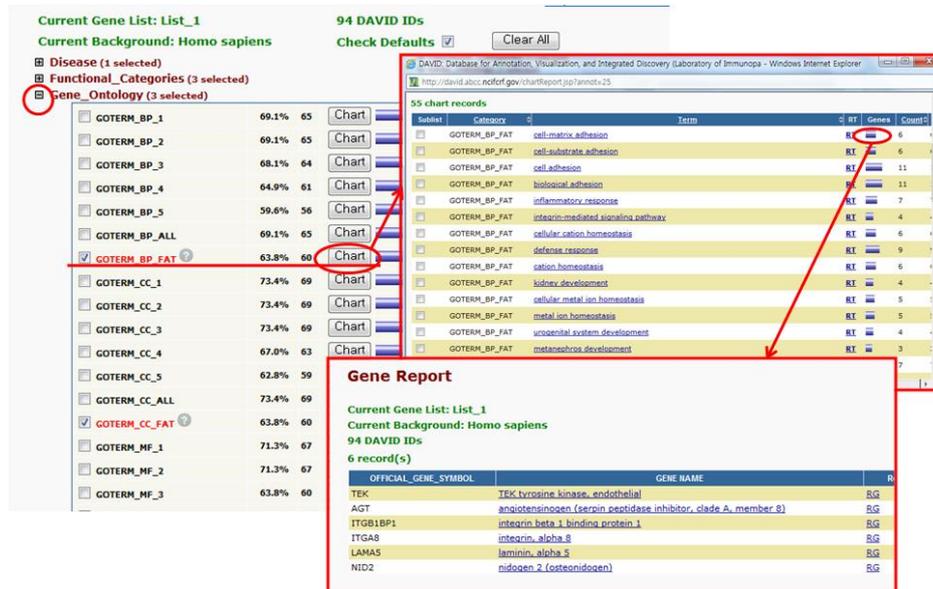


그림 4-5. DAVID tool : exploring Gene Ontology analysis result

이와 같은 방법으로 Pathway 결과를 확인해 보면 KEGG_PATHWAY database 에서 주요 Pathway 가 나온다(그림 4-6). 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.

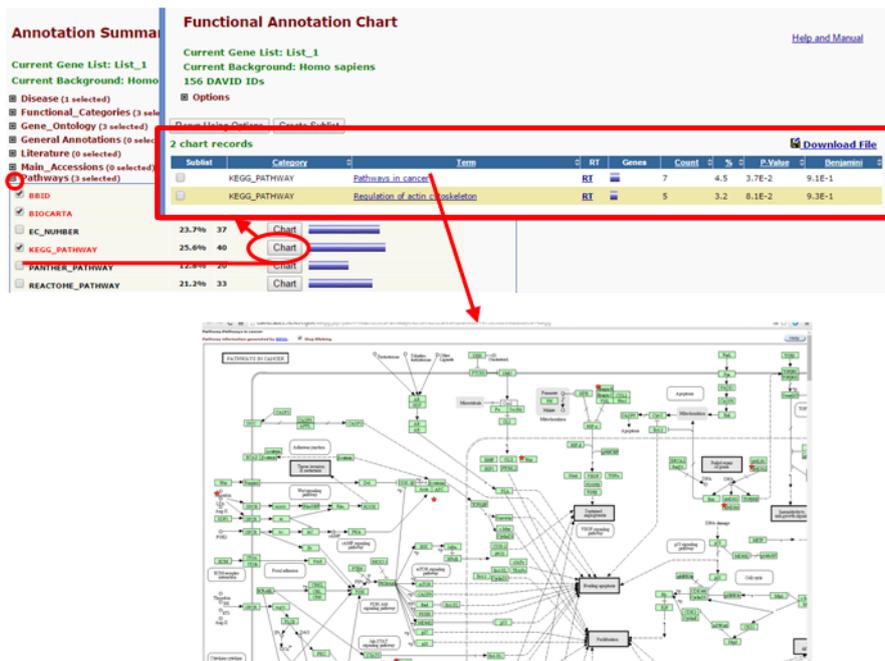


그림 4-6. DAVID tool : exploring Pathway analysis result

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 4-7).

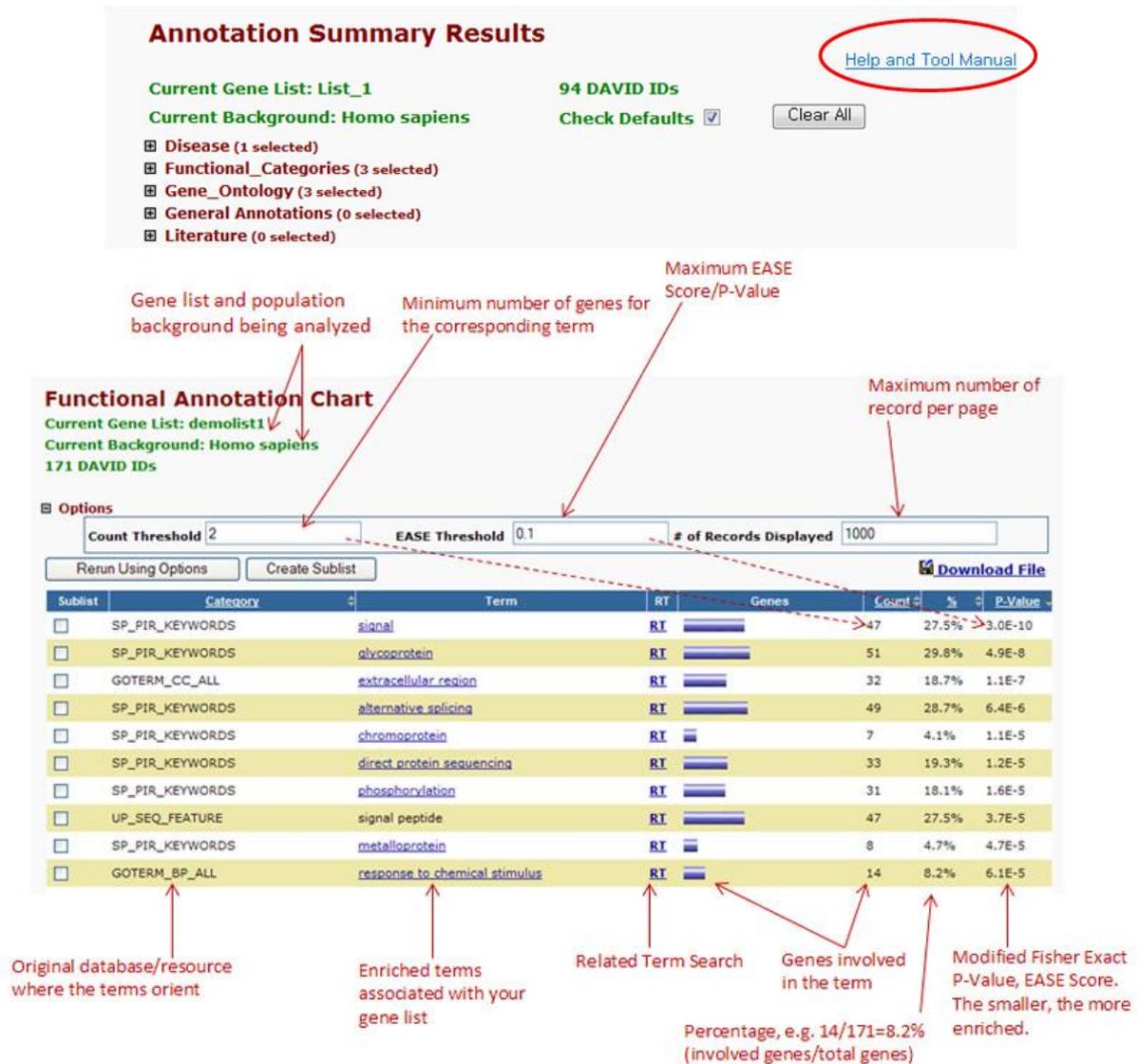


그림 4-7. DAVID Help and Tool Manual

5. MeV Software (free) 이용 Clustering 분석

MeV (MultiExperiment Viewer) 소프트웨어는 미국의 Dana-Farber Cancer Institute에서 개발한 MiRNA-Seq 전용 분석 프로그램으로 연구자들에게 무료로 공급하고 있다. 주로 clustering 분석과 통계분석 (K-means clustering, Hierarchical clustering, t-test, Significance Analysis of MiRNA-Seqs, Gene Set Enrichment Analysis, EASE)을 할 수 있는 프로그램이다. 아래 웹 페이지에 접속하면 최신의 업데이트된 프로그램과 매뉴얼을 다운받을 수 있다.

<http://www.tm4.org/mev.html>

프로그램을 다운받아 압축을 풀고, MeV 또는 TMEV를 클릭해서 프로그램을 실행시킨다(그림5-1). MEV 프로그램을 실행시키면 세 개의 창이 나타난다(그림5-2). 분석 창은 프로그램창의 메뉴에서 file -> New multiple array viewer를 통해 여러 개를 생성할 수 있고 데이터 분석은 분석 창을 통해 진행한다.

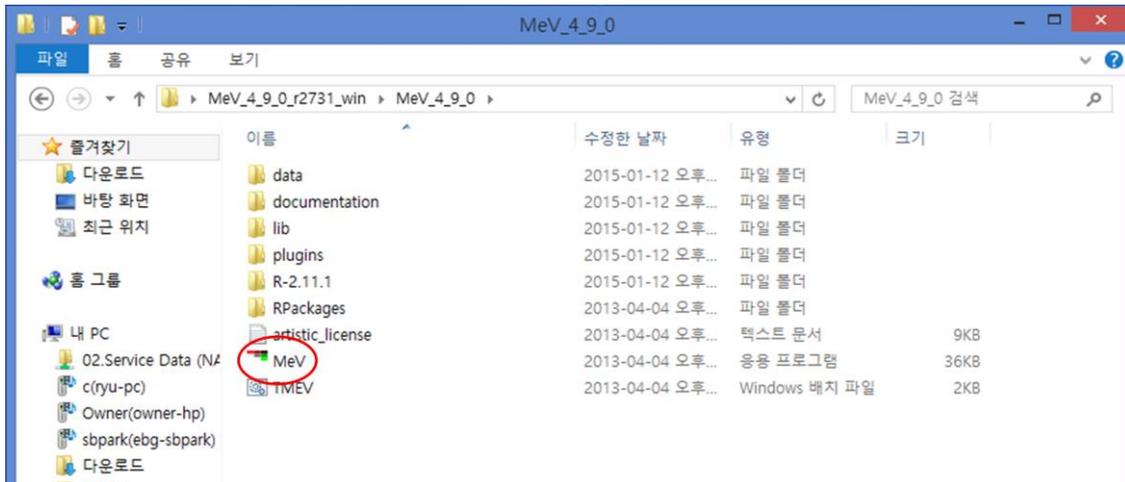


그림 5-1. MeV program folder and files

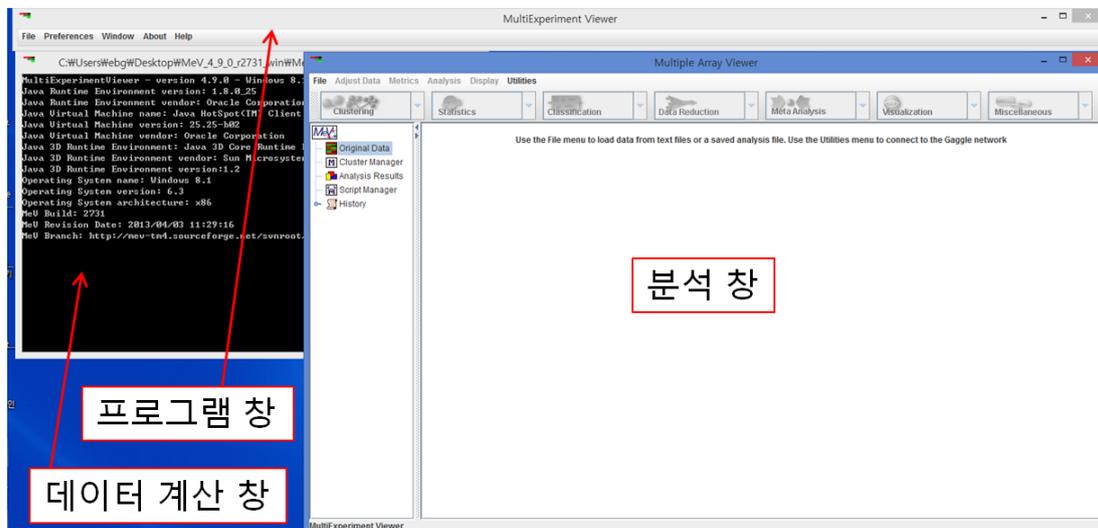


그림 5-2. MeV program windows

본 자료에서는 MeV 프로그램을 이용하여 Clustering 분석 방법을 설명한다. 우선 MeV 프로그램에 input할 데이터를 엑셀에서 파일 양식에 맞춰 저장해야 한다. 엑셀에 clustering 하고자 하는 유전자 이름과 fold change 또는 발현값(intensity)를 정리한다(그림 5-3). 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.

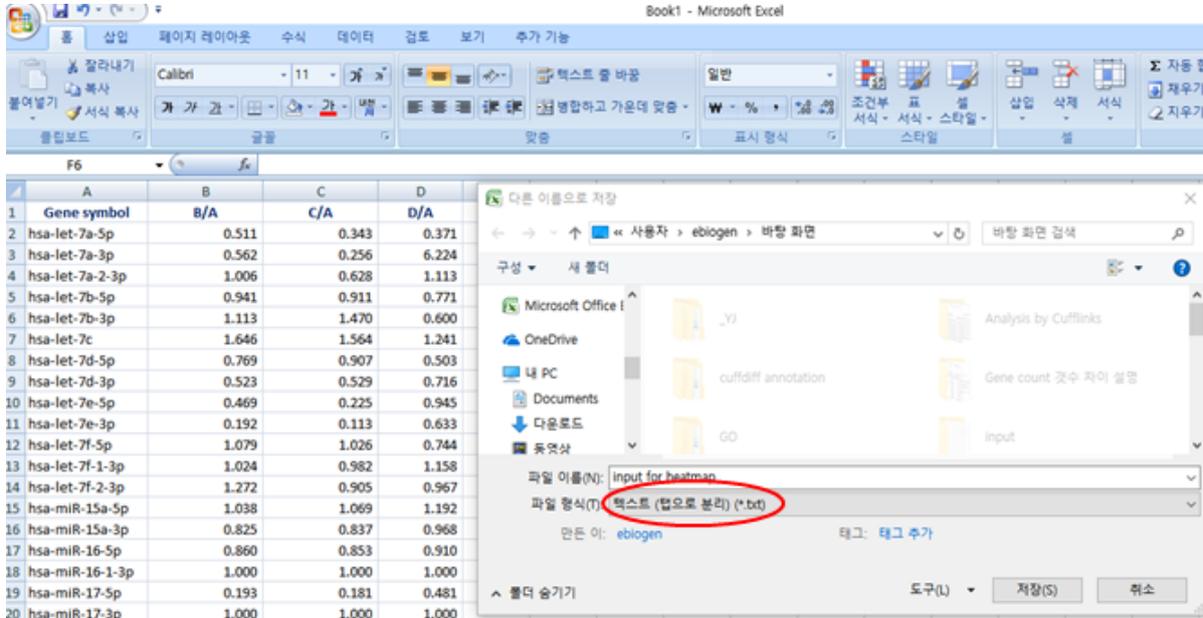


그림 5-3. Data format example

input 데이터 저장에 완료되면 MeV 프로그램의 분석 창에서 file -> load data를 실행한다(그림 5-4). Browse를 클릭하여 input 데이터를 선택한다. 데이터가 fold change인 경우 "Two-color Array"로 체크하고 데이터가 intensity인 경우는 "Single-color Array"에 체크한다. 마우스로 데이터가 시작되는 부위를 클릭한 후 load를 누른다.

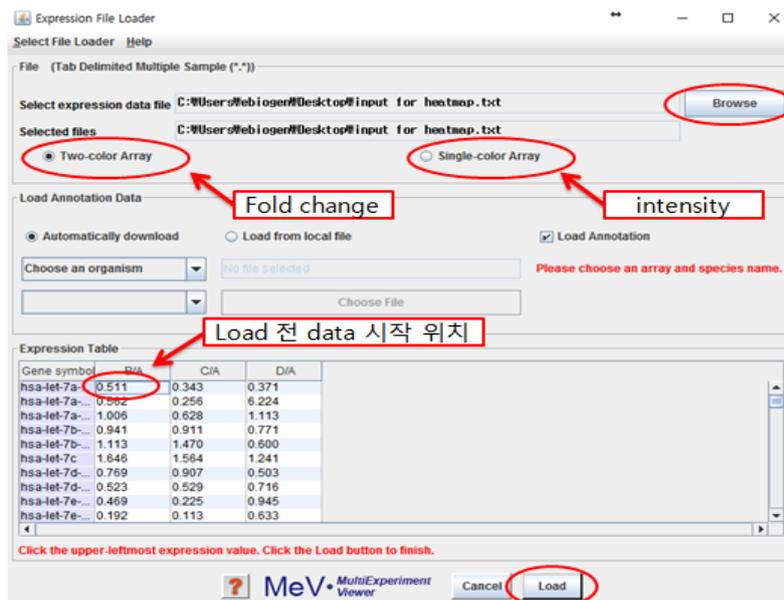


그림 5-4. Data uploading method

데이터가 열리면 Adjust Data -> Log Transformation -> Log2 Transform을 선택하여 fold change는 $\log_2(\text{fold change})$ 로, intensity는 $\log_2(\text{intensity})$ 로 바꿔준다(그림 5-5). 왼쪽 메뉴의 Original Data -> Expression image를 보면 \log_2 값으로 바뀌어 색이 변한 것을 확인할 수 있다.

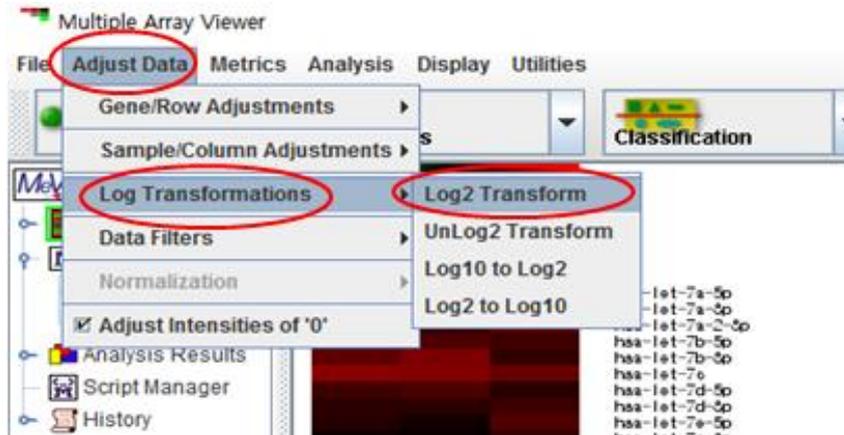


그림 5-5. Log2 transformation

Analysis-> Clustering-> HCL을 선택하여 Clustering 분석을 시작한다(그림 5-6).

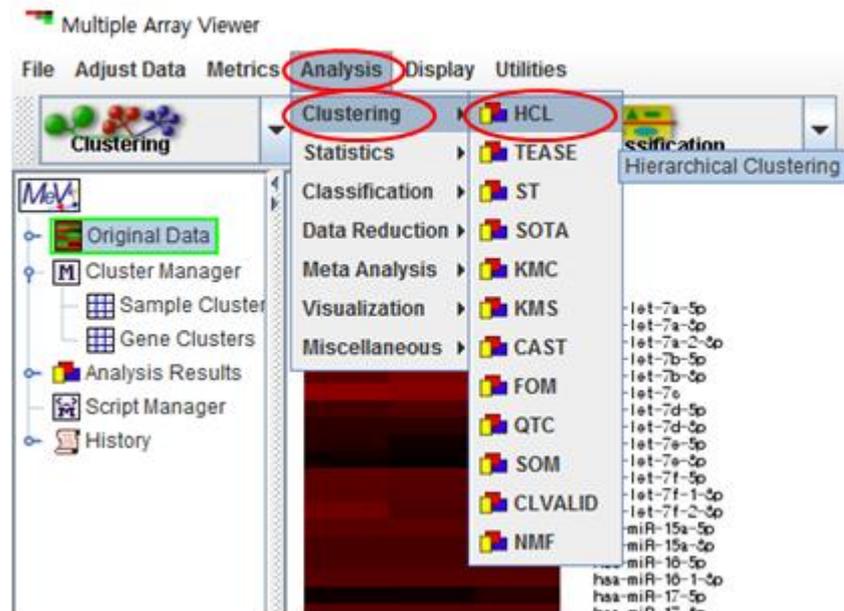


그림 5-6. Hierarchical Clustering Selection

Clustering 분석 시 다양한 옵션을 선택할 수 있다(그림 5-7). 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.

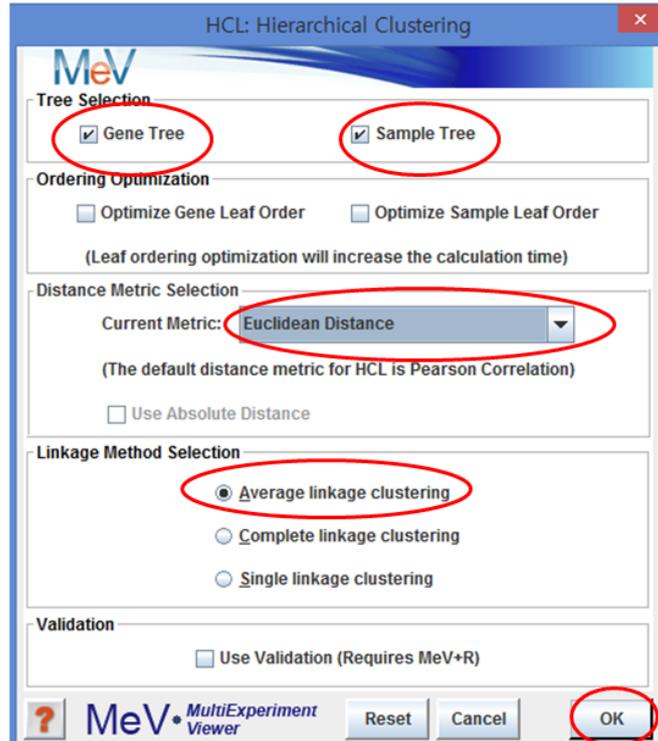


그림 5-7. Hierarchical Clustering Method

clustering이 완료되면 왼쪽 메뉴에 Analysis Results에 HCL 결과가 생긴다. HCL -> HCL tree를 클릭하면 clustering 결과가 화면에 나온다(그림 5-8). 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.

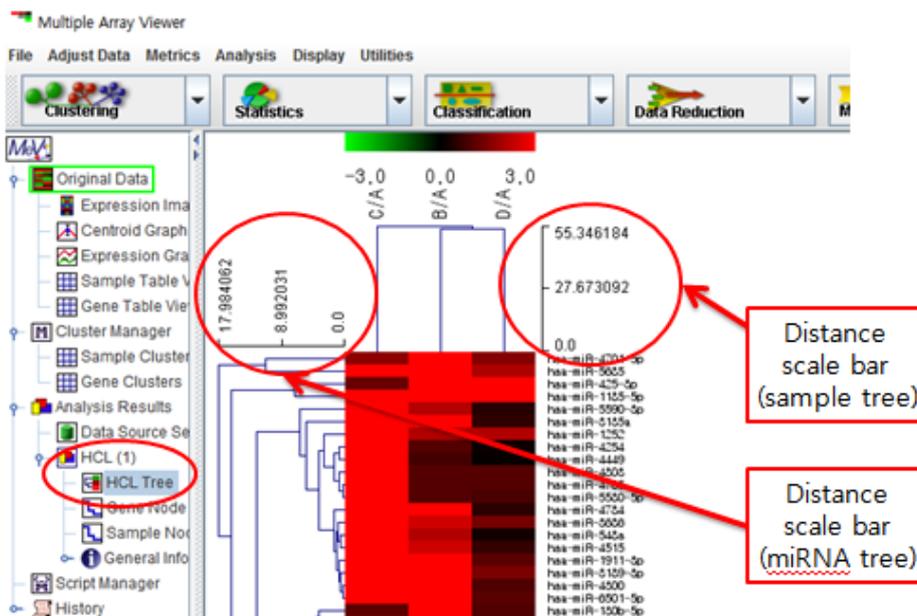


그림 5-8. Hierarchical Clustering Result

clustering 결과는 이미지의 크기와 색상을 조절하여 원하는 형태의 이미지를 만들 수 있다(그림 5-9, 5-10)

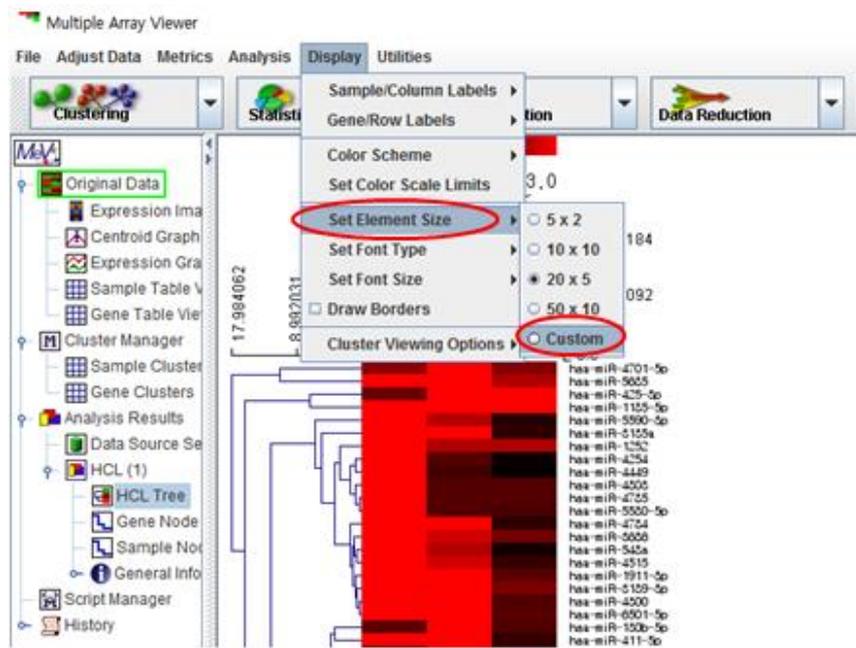


그림 5-9. Clustering image size control

Display -> Set Color Scale Limits를 누르면 color scale bar의 최소값, 중간값, 최대값을 설정할 수 있다. 보통 $\log_2(\text{fold change})$ 는 최소값과 최대값은 같은 크기에 부등호만 바꿔주고(예: min:-3, max:3) 중간값은 0으로 설정해 준다(그림 5-10). 이렇게 하면 up-regulated genes은 red, down-regulated genes은 green으로 나타나게 된다.

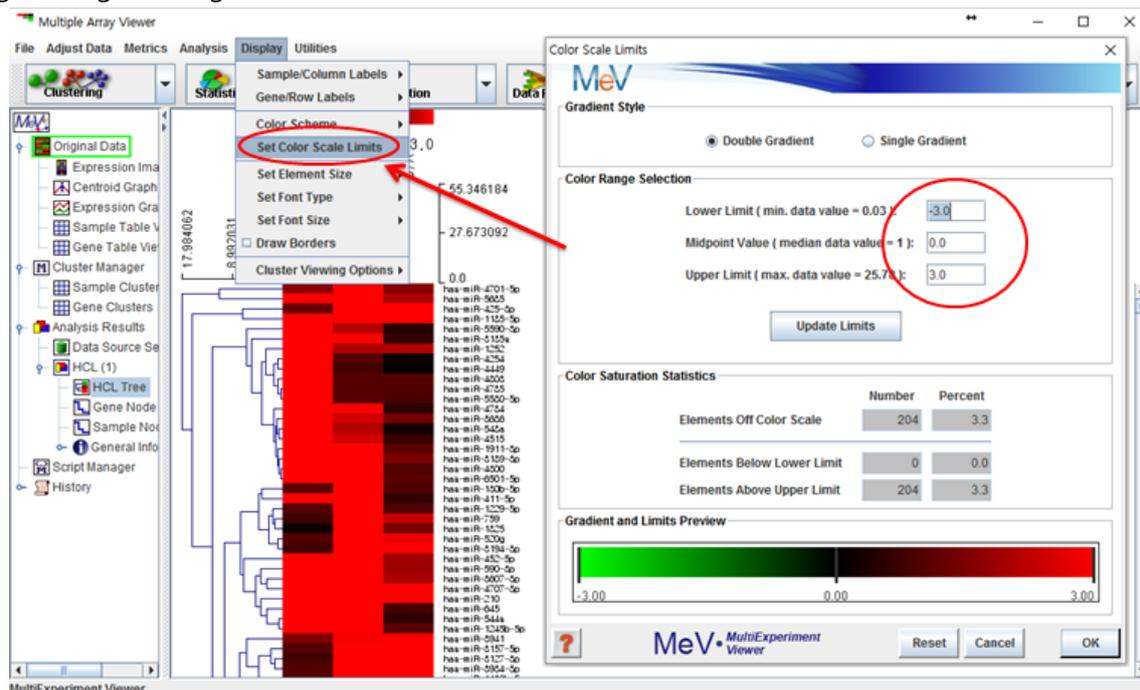


그림 5-10. Clustering image color setting

원하는 이미지 조절이 완료되면 File -> Save image를 눌러 이미지를 저장한다. 이때 파일 이름에 파일 확장자명(예: .png)을 꼭 기입하여야 이미지 파일로 저장이 된다(그림 5-11).

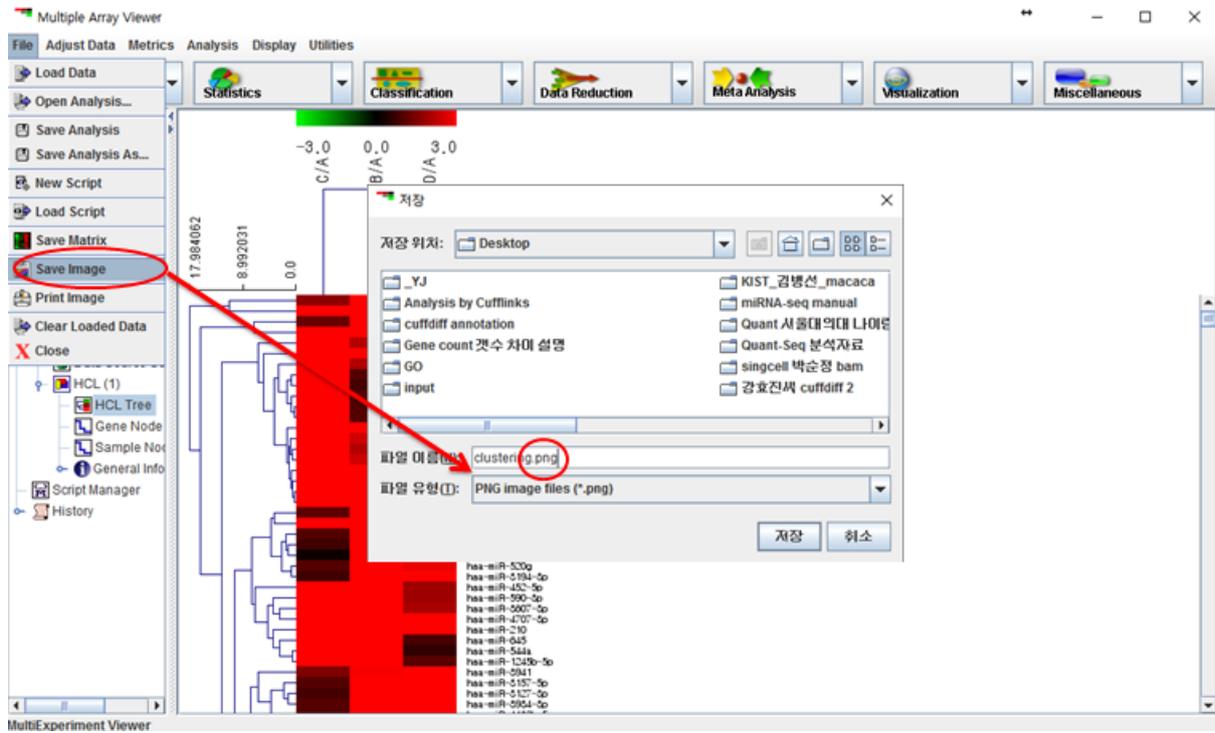


그림 5-11. Clustering image save