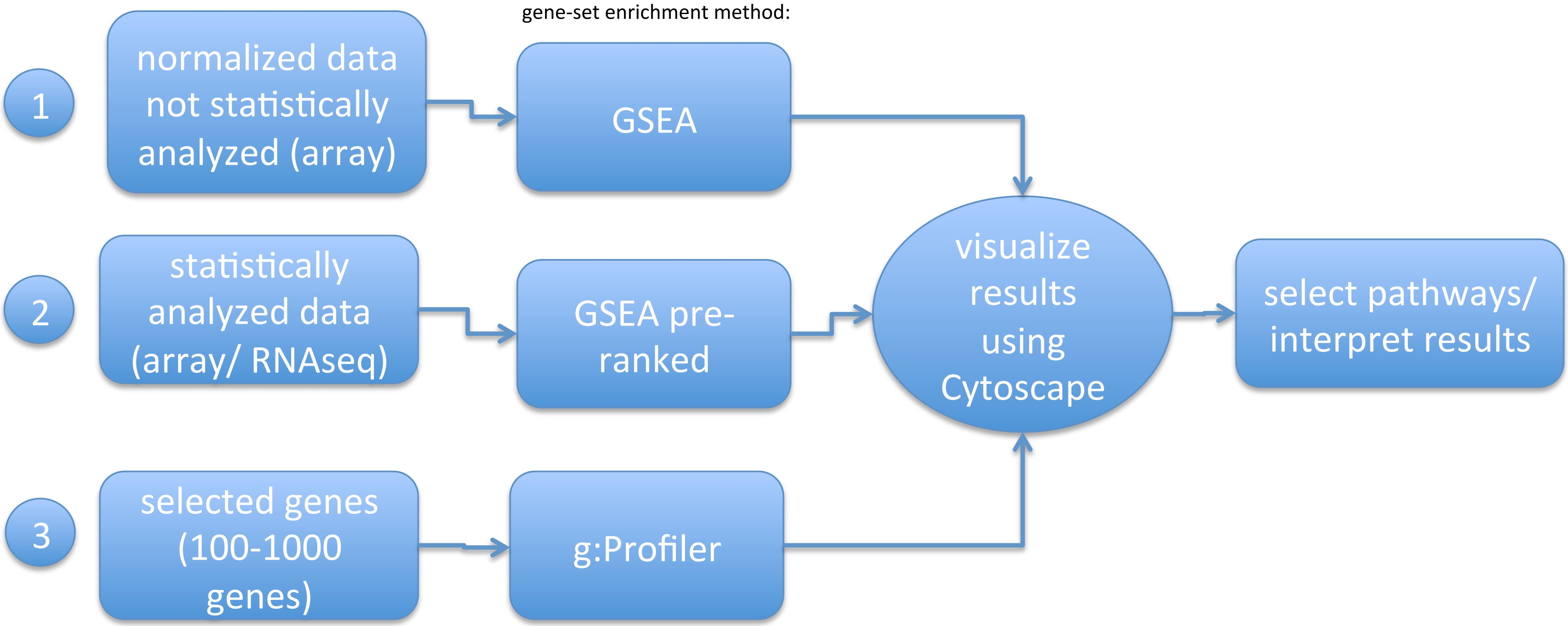
Bader lab workflow for gene set enrichment analysis of gene expression data



gene-set enrichment method:

1 normalized data not statistically analyzed (array) → GSEA

2 statistically analyzed data (array/ RNAseq) → GSEA pre-ranked

3 selected genes (100-1000 genes) → g:Profiler

visualize results using Cytoscape → select pathways/ interpret results

Tutorials:
Workflow 1) http://baderlab.org/Software/EnrichmentMap/Tutorial
Workflow 2) See next slides
Workflow 3) http://baderlab.org/Software/EnrichmentMap/GProfilerTutorial

Alternative workflows: DAVID, BiNGO, generic format: go to http://baderlab.org/Software/EnrichmentMap

Links

GSEA download: http://www.broadinstitute.org/gsea/downloads.jsp
→ try 1Gb first then 2Gb if out of memory problem

GSEA format: http://www.broadinstitute.org/cancer/software/gsea/wiki/index.php/Data_formats

Bader lab gene-set file: http://download.baderlab.org/EM_Genesets/
→ Use current release: choose Human or Mouse, choose Gene Symbol or EntrezGene and GO_AllPathways_no_GO_iea

Cytoscape download: http://www.cytoscape.org/download.php
→ Use Cytoscpae 2.8.3

Enrichment map plugin download: http://baderlab.org/Software/EnrichmentMap
→ Use EM 1.3DEV most recent release (today: 07/02/2013)

Wordcloud download: http://baderlab.org/WordCloud

EnrichmentMap tutorials: http://baderlab.org/Software/EnrichmentMap

Software installation

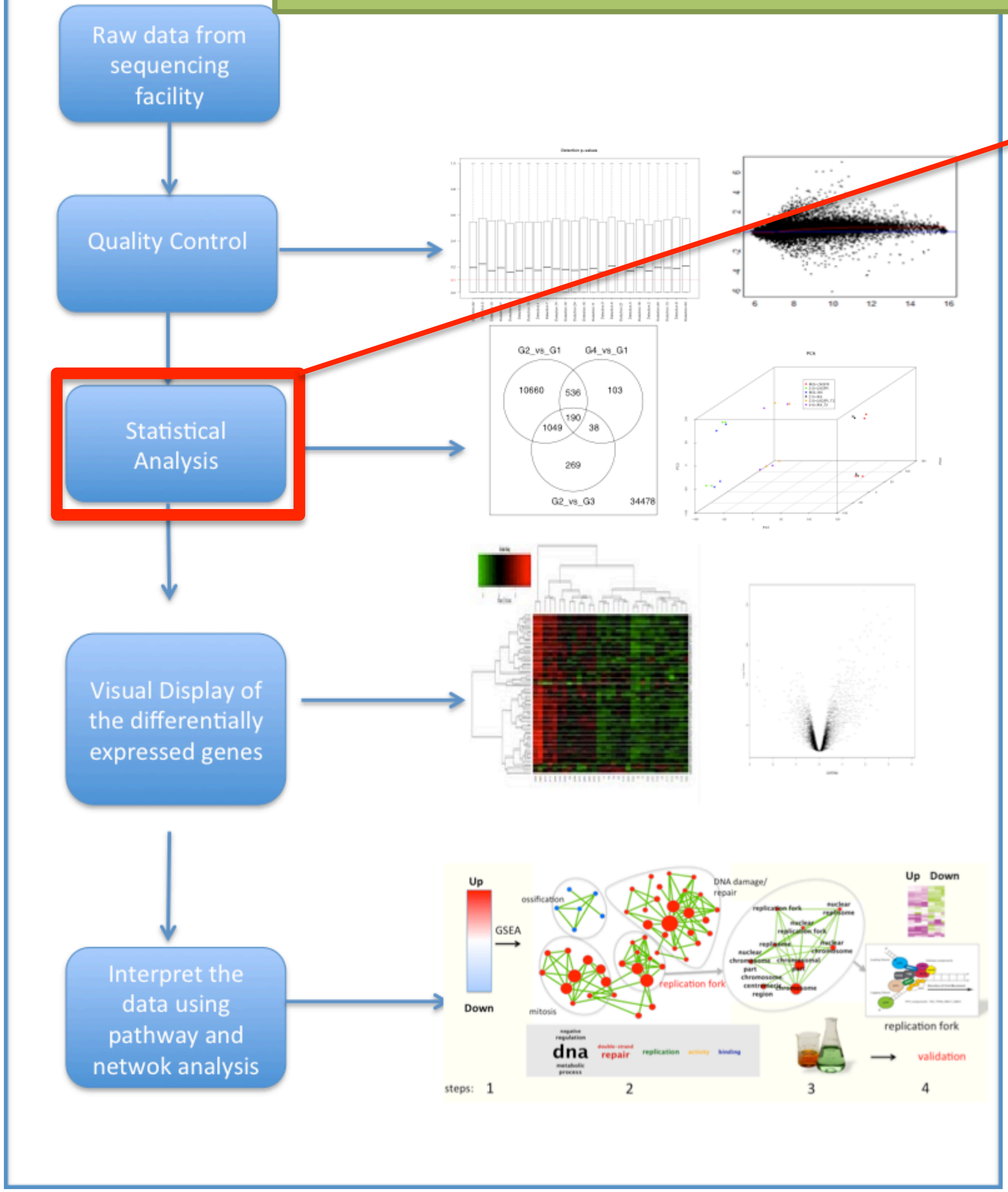
1) Download GSEA
2) Download Cytoscape
3) Download the EnrichmentMap and Wordcloud jar files

Plugin installation

1) Copy the enrichmentmap.jar and the wordcloud.jar into your Cytoscape plugins folder.
2) Open Cytoscape

Workflow 2: GSEA pre-ranked

Typical workflow for gene expression data from Illumina beadchips or Affymetrix arrays



Raw data from sequencing facility

Quality Control

Statistical Analysis

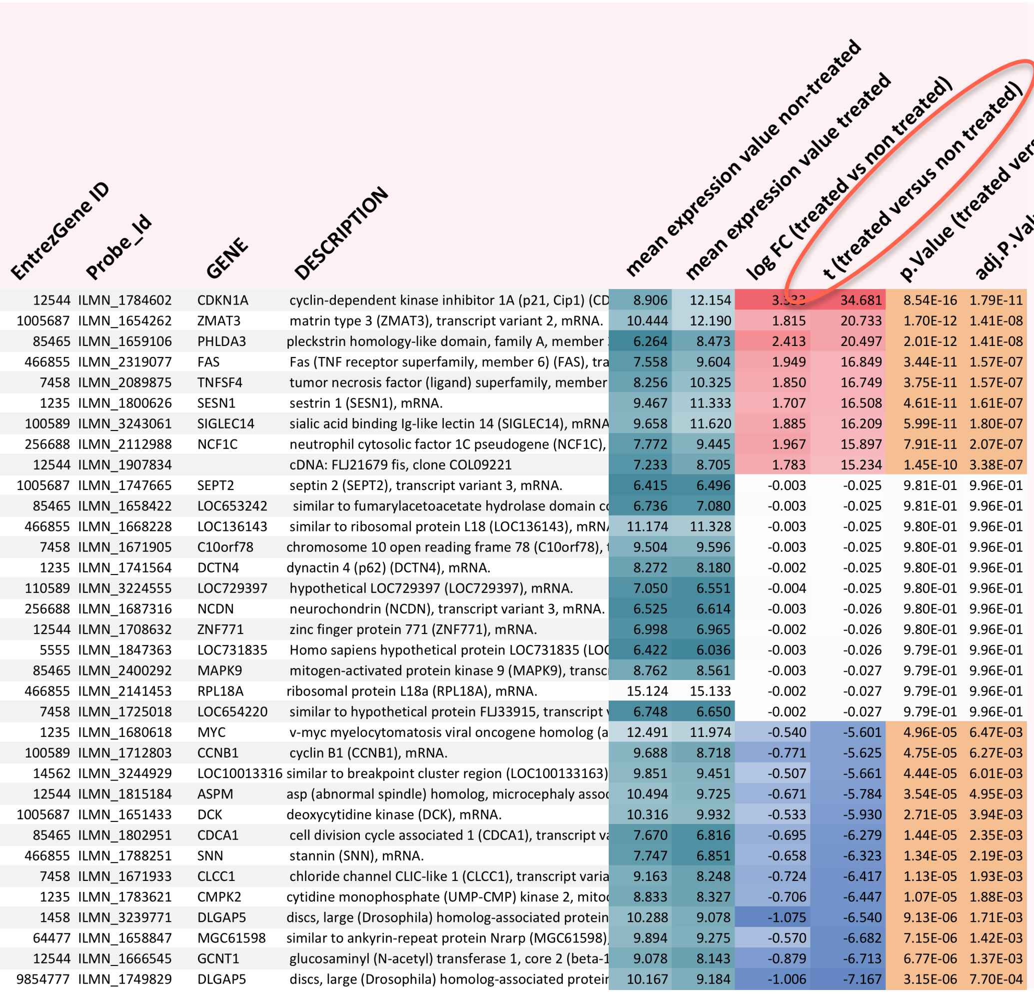
Visual Display of the differentially expressed genes

Interpret the data using pathway and netwok analysis

The typical output from a differential expression analysis of microarray is a spreadsheet containing all the probes on the array as rows. It will include as columns values to estimate differential expression (p-value, FDR, logFC), and also raw normalized data, and gene identifiers.

From this spreadsheet, we need to create 2 files: the rank file (.rnk) and the expression file (.GCT)

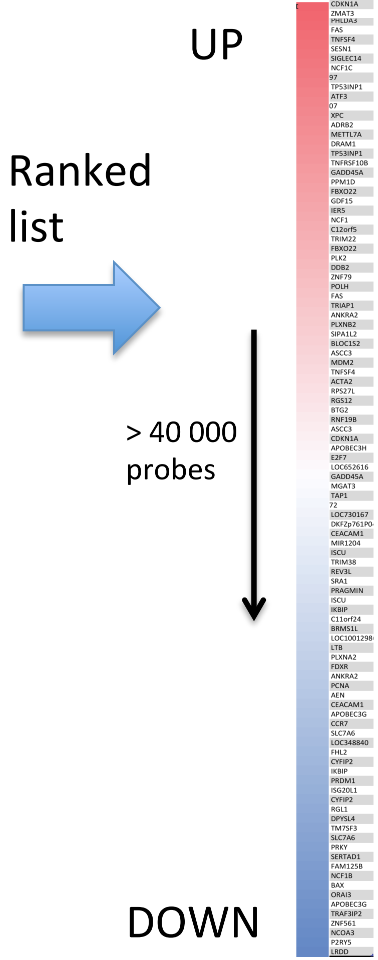
Rank the genes using a differential expression value



EntrezGene ID	Probe_Id	GENE	DESCRIPTION	mean expression value non-treated	mean expression value treated	log FC (treated vs non treated)	t (treated versus non treated)	p.Value (treated versus non treated)	adj.P.Value (treated versus non treated)
12544	ILMN_1784602	CDKN1A	cyclin-dependent kinase inhibitor 1A (p21, Cip1) (CD	8.906	12.154	3.333	34.681	8.54E-16	1.79E-11
1005687	ILMN_1654262	ZMAT3	matrin type 3 (ZMAT3), transcript variant 2, mRNA.	10.444	12.190	1.815	20.733	1.70E-12	1.41E-08
85465	ILMN_1659106	PHLDA3	pleckstrin homology-like domain, family A, member	6.264	8.473	2.413	20.497	2.01E-12	1.41E-08
466855	ILMN_2319077	FAS	Fas (TNF receptor superfamily, member 6) (FAS), tra	7.558	9.604	1.949	16.849	3.44E-11	1.57E-07
7458	ILMN_2089875	TNFSF4	tumor necrosis factor (ligand) superfamily, member	8.256	10.325	1.850	16.749	3.75E-11	1.57E-07
1235	ILMN_1800626	SESN1	sestrin 1 (SESN1), mRNA.	9.467	11.333	1.707	16.508	4.61E-11	1.61E-07
100589	ILMN_3243061	SIGLEC14	sialic acid binding Ig-like lectin 14 (SIGLEC14), mRNA	9.658	11.620	1.885	16.209	5.99E-11	1.80E-07
256688	ILMN_2112988	NCF1C	neutrophil cytosolic factor 1C pseudogene (NCF1C),	7.772	9.445	1.967	15.897	7.91E-11	2.07E-07
12544	ILMN_1907834		cDNA: FLJ21679 fis, clone COL09221	7.233	8.705	1.783	15.234	1.45E-10	3.38E-07
1005687	ILMN_1747665	SEPT2	septin 2 (SEPT2), transcript variant 3, mRNA.	6.415	6.496	-0.003	-0.025	9.81E-01	9.96E-01
85465	ILMN_1658422	LOC653242	similar to fumarylacetoacetate hydrolase domain c	6.736	7.080	-0.003	-0.025	9.81E-01	9.96E-01
466855	ILMN_1668228	LOC136143	similar to ribosomal protein L18 (LOC136143), mRN	11.174	11.328	-0.003	-0.025	9.80E-01	9.96E-01
7458	ILMN_1671905	C10orf78	chromosome 10 open reading frame 78 (C10orf78),	9.504	9.596	-0.003	-0.025	9.80E-01	9.96E-01
1235	ILMN_1741564	DCTN4	dynactin 4 (p62) (DCTN4), mRNA.	8.272	8.180	-0.002	-0.025	9.80E-01	9.96E-01
110589	ILMN_3224555	LOC729397	hypothetical LOC729397 (LOC729397), mRNA.	7.050	6.551	-0.004	-0.025	9.80E-01	9.96E-01
256688	ILMN_1687316	NCDN	neurochondrin (NCDN), transcript variant 3, mRNA.	6.525	6.614	-0.003	-0.026	9.80E-01	9.96E-01
12544	ILMN_1708632	ZNF771	zinc finger protein 771 (ZNF771), mRNA.	6.998	6.965	-0.002	-0.026	9.80E-01	9.96E-01
5555	ILMN_1847363	LOC731835	Homo sapiens hypothetical protein LOC731835 (LO	6.422	6.036	-0.003	-0.026	9.79E-01	9.96E-01
85465	ILMN_2400292	MAPK9	mitogen-activated protein kinase 9 (MAPK9), transc	8.762	8.561	-0.003	-0.027	9.79E-01	9.96E-01
466855	ILMN_2141453	RPL18A	ribosomal protein L18a (RPL18A), mRNA.	15.124	15.133	-0.002	-0.027	9.79E-01	9.96E-01
7458	ILMN_1725018	LOC654220	similar to hypothetical protein FLJ33915, transcript	6.748	6.650	-0.002	-0.027	9.79E-01	9.96E-01
1235	ILMN_1680618	MYC	v-myc myelocytomatosis viral oncogene homolog (a	12.491	11.974	-0.540	-5.601	4.96E-05	6.47E-03
100589	ILMN_1712803	CCNB1	cyclin B1 (CCNB1), mRNA.	9.688	8.718	-0.771	-5.625	4.75E-05	6.27E-03
14562	ILMN_3244929	LOC10013316	similar to breakpoint cluster region (LOC100133163)	9.851	9.451	-0.507	-5.661	4.44E-05	6.01E-03
12544	ILMN_1815184	ASPM	asp (abnormal spindle) homolog, microcephaly asso	10.494	9.725	-0.671	-5.784	3.54E-05	4.95E-03
1005687	ILMN_1651433	DCK	deoxycytidine kinase (DCK), mRNA.	10.316	9.932	-0.533	-5.930	2.71E-05	3.94E-03
85465	ILMN_1802951	CDCA1	cell division cycle associated 1 (CDCA1), transcript v	7.670	6.816	-0.695	-6.279	1.44E-05	2.35E-03
466855	ILMN_1788251	SNN	stannin (SNN), mRNA.	7.747	6.851	-0.658	-6.323	1.34E-05	2.19E-03
7458	ILMN_1671933	CLCC1	chloride channel CLIC-like 1 (CLCC1), transcript varia	9.163	8.248	-0.724	-6.417	1.13E-05	1.93E-03
1235	ILMN_1783621	CMPK2	cytidine monophosphate (UMP-CMP) kinase 2, mito	8.833	8.327	-0.706	-6.447	1.07E-05	1.88E-03
1458	ILMN_3239771	DLGAP5	discs, large (Drosophila) homolog-associated protein	10.288	9.078	-1.075	-6.540	9.13E-06	1.71E-03
64477	ILMN_1658847	MGC61598	similar to ankyrin-repeat protein Nrarp (MGC61598)	9.894	9.275	-0.570	-6.682	7.15E-06	1.42E-03
12544	ILMN_1666545	GCNT1	glucosaminyl (N-acetyl) transferase 1, core 2 (beta-1	9.078	8.143	-0.879	-6.713	6.77E-06	1.37E-03
9854777	ILMN_1749829	DLGAP5	discs, large (Drosophila) homolog-associated protei	10.167	9.184	-1.006	-7.167	3.15E-06	7.70E-04

Up-regulated in treated

Down-regulated in treated



UP

Ranked list

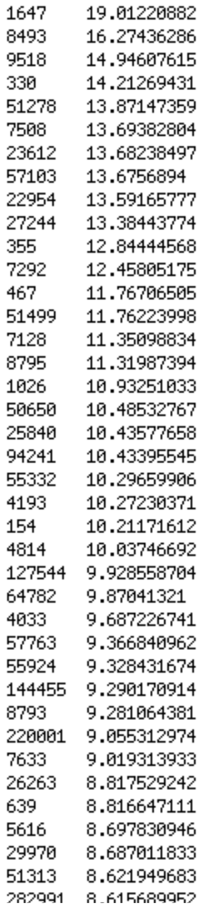
> 40 000 probes

DOWN

CDKN1A
ZMAT3
PHLDA3
FAS
TNFSF4
SESN1
SIGLEC14
NCF1C
97
TP53INP1
ATF3
07
XPC
ADRB2
METTL7A
DRAM1
TP53INP1
TNFRSF10B
GADD45A
PPM1D
FBXO22
GDF15
IER5
NCF1
C12orf5
TRIM22
FBXO22
PLK2
DDB2
ZNF79
POLH
FAS
TRIAP1
ANKRA2
PLXNB2
SIPA1L2
BLOC1S2
ASCC3
MDM2
TNFSF4
ACTA2
RPS27L
RGS12
BTG2
RNF19B
ASCC3
CDKN1A
APOBEC3H
E2F7
LOC652616
GADD45A
MGAT3
TAP1
72
LOC730167
DKFZp761P0
CEACAM1
MIR1204
ISCU
TRIM38
REV3L
SRA1
PRAGMIN
ISCU
IKBIP
C11orf24
BRMS1L
LOC1001298
LTB
PLXNA2
FDXR
ANKRA2
PCNA
AEN
CEACAM1
APOBEC3G
CCR7
SLC7A6
LOC348840
FHL2
CYFIP2
IKBIP
PRDM1
ISG20L1
CYFIP2
RGL1
DPYSL4
TM7SF3
SLC7A6
PRKY
SERTAD1
FAM125B
NCF1B
BAX
ORAI3
APOBEC3G
TRAF3IP2
ZNF561
NCOA3
P2RY5
LRDD

Files you need to prepare to run GSEA pre-ranked

Rank file

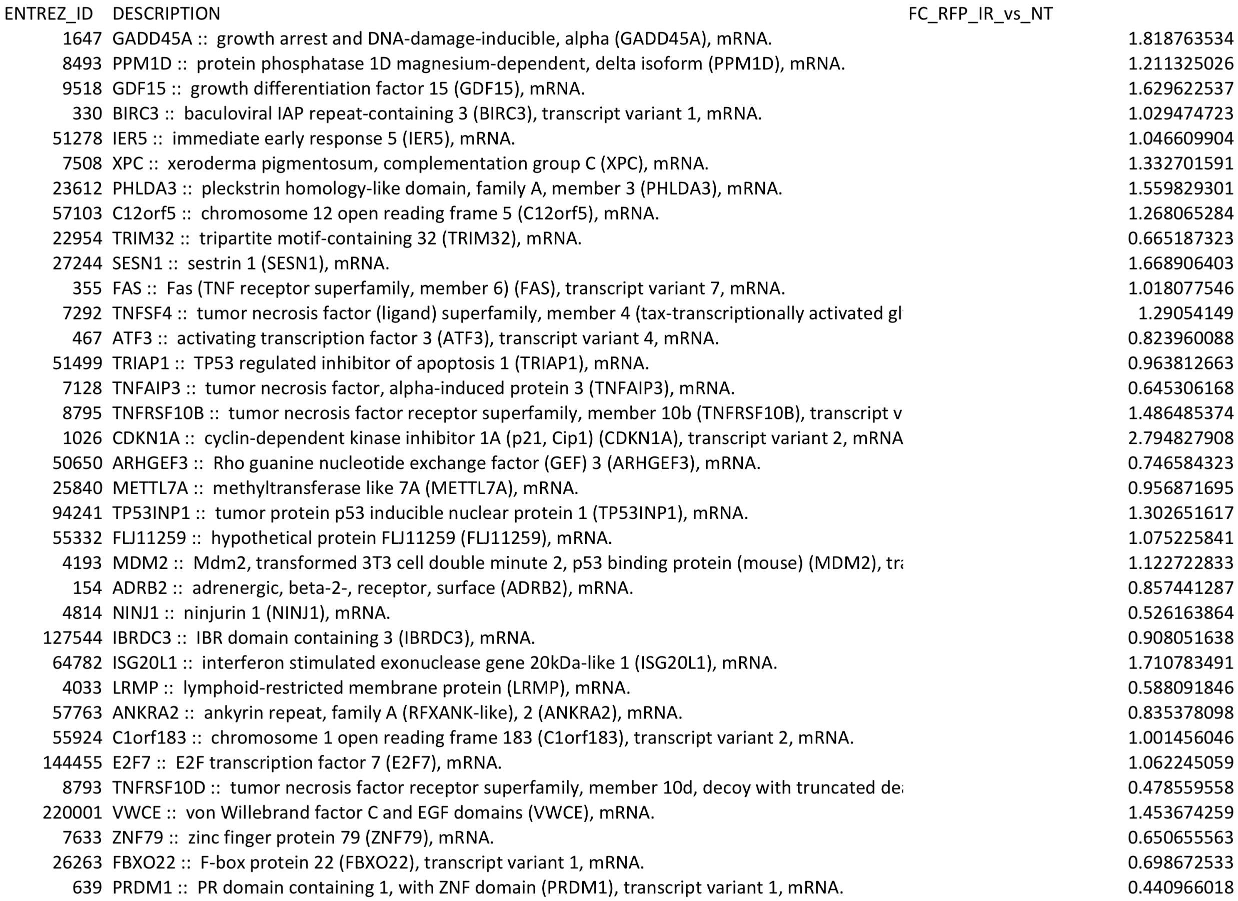


```
1647	19.01220882
8493	16.27436286
9518	14.94607615
330	14.21269431
51278	13.87147359
7508	13.69382804
23612	13.68238497
57103	13.6756894
22954	13.59165777
27244	13.38443774
355	12.84444568
7292	12.45805175
467	11.76706505
51499	11.76223998
7128	11.35098834
8795	11.31987394
1026	10.93251033
50650	10.48532767
25840	10.43577658
94241	10.43395545
55332	10.29659906
4193	10.27230371
154	10.21171612
4814	10.03746692
127544	9.928558704
64782	9.87041321
4033	9.687226741
57763	9.366840962
55924	9.328431674
144455	9.290170914
8793	9.281064381
220001	9.055312974
7633	9.019313933
26263	8.817529242
639	8.816647111
5616	8.697830946
29970	8.687011833
51313	8.621949683
282991	8.615689952
```

No column header!

tab delimited file (.txt)
No column names
First column: entrez gene ID or official gene symbol
Second column: differential expression value used to rank the gene list

Expression file



ENTREZ_ID	DESCRIPTION	FC_RFP_IR_vs_NT
1647	GADD45A :: growth arrest and DNA-damage-inducible, alpha (GADD45A), mRNA.	1.818763534
8493	PPM1D :: protein phosphatase 1D magnesium-dependent, delta isoform (PPM1D), mRNA.	1.211325026
9518	GDF15 :: growth differentiation factor 15 (GDF15), mRNA.	1.629622537
330	BIRC3 :: baculoviral IAP repeat-containing 3 (BIRC3), transcript variant 1, mRNA.	1.029474723
51278	IER5 :: immediate early response 5 (IER5), mRNA.	1.046609904
7508	XPC :: xeroderma pigmentosum, complementation group C (XPC), mRNA.	1.332701591
23612	PHLDA3 :: pleckstrin homology-like domain, family A, member 3 (PHLDA3), mRNA.	1.559829301
57103	C12orf5 :: chromosome 12 open reading frame 5 (C12orf5), mRNA.	1.268065284
22954	TRIM32 :: tripartite motif-containing 32 (TRIM32), mRNA.	0.665187323
27244	SESN1 :: sestrin 1 (SESN1), mRNA.	1.668906403
355	FAS :: Fas (TNF receptor superfamily, member 6) (FAS), transcript variant 7, mRNA.	1.018077546
7292	TNFSF4 :: tumor necrosis factor (ligand) superfamily, member 4 (tax-transcriptionally activated gl	1.29054149
467	ATF3 :: activating transcription factor 3 (ATF3), transcript variant 4, mRNA.	0.823960088
51499	TRIAP1 :: TP53 regulated inhibitor of apoptosis 1 (TRIAP1), mRNA.	0.963812663
7128	TNFAIP3 :: tumor necrosis factor, alpha-induced protein 3 (TNFAIP3), mRNA.	0.645306168
8795	TNFRSF10B :: tumor necrosis factor receptor superfamily, member 10b (TNFRSF10B), transcript v	1.486485374
1026	CDKN1A :: cyclin-dependent kinase inhibitor 1A (p21, Cip1) (CDKN1A), transcript variant 2, mRNA	2.794827908
50650	ARHGEF3 :: Rho guanine nucleotide exchange factor (GEF) 3 (ARHGEF3), mRNA.	0.746584323
25840	METTL7A :: methyltransferase like 7A (METTL7A), mRNA.	0.956871695
94241	TP53INP1 :: tumor protein p53 inducible nuclear protein 1 (TP53INP1), mRNA.	1.302651617
55332	FLJ11259 :: hypothetical protein FLJ11259 (FLJ11259), mRNA.	1.075225841
4193	MDM2 :: Mdm2, transformed 3T3 cell double minute 2, p53 binding protein (mouse) (MDM2), tra	1.122722833
154	ADRB2 :: adrenergic, beta-2-, receptor, surface (ADRB2), mRNA.	0.857441287
4814	NINJ1 :: ninjurin 1 (NINJ1), mRNA.	0.526163864
127544	IBRDC3 :: IBR domain containing 3 (IBRDC3), mRNA.	0.908051638
64782	ISG20L1 :: interferon stimulated exonuclease gene 20kDa-like 1 (ISG20L1), mRNA.	1.710783491
4033	LRMP :: lymphoid-restricted membrane protein (LRMP), mRNA.	0.588091846
57763	ANKRA2 :: ankyrin repeat, family A (RFXANK-like), 2 (ANKRA2), mRNA.	0.835378098
55924	C1orf183 :: chromosome 1 open reading frame 183 (C1orf183), transcript variant 2, mRNA.	1.001456046
144455	E2F7 :: E2F transcription factor 7 (E2F7), mRNA.	1.062245059
8793	TNFRSF10D :: tumor necrosis factor receptor superfamily, member 10d, decoy with truncated de	0.478559558
220001	VWCE :: von Willebrand factor C and EGF domains (VWCE), mRNA.	1.453674259
7633	ZNF79 :: zinc finger protein 79 (ZNF79), mRNA.	0.650655563
26263	FBXO22 :: F-box protein 22 (FBXO22), transcript variant 1, mRNA.	0.698672533
639	PRDM1 :: PR domain containing 1, with ZNF domain (PRDM1), transcript variant 1, mRNA.	0.440966018

Column Header

tab delimited file (.txt)
with column names
First column: entrez gene ID or official gene symbol (need to match with the rank file
Second column: gene name /description
Additional columns: log2 normalized data of each replicate

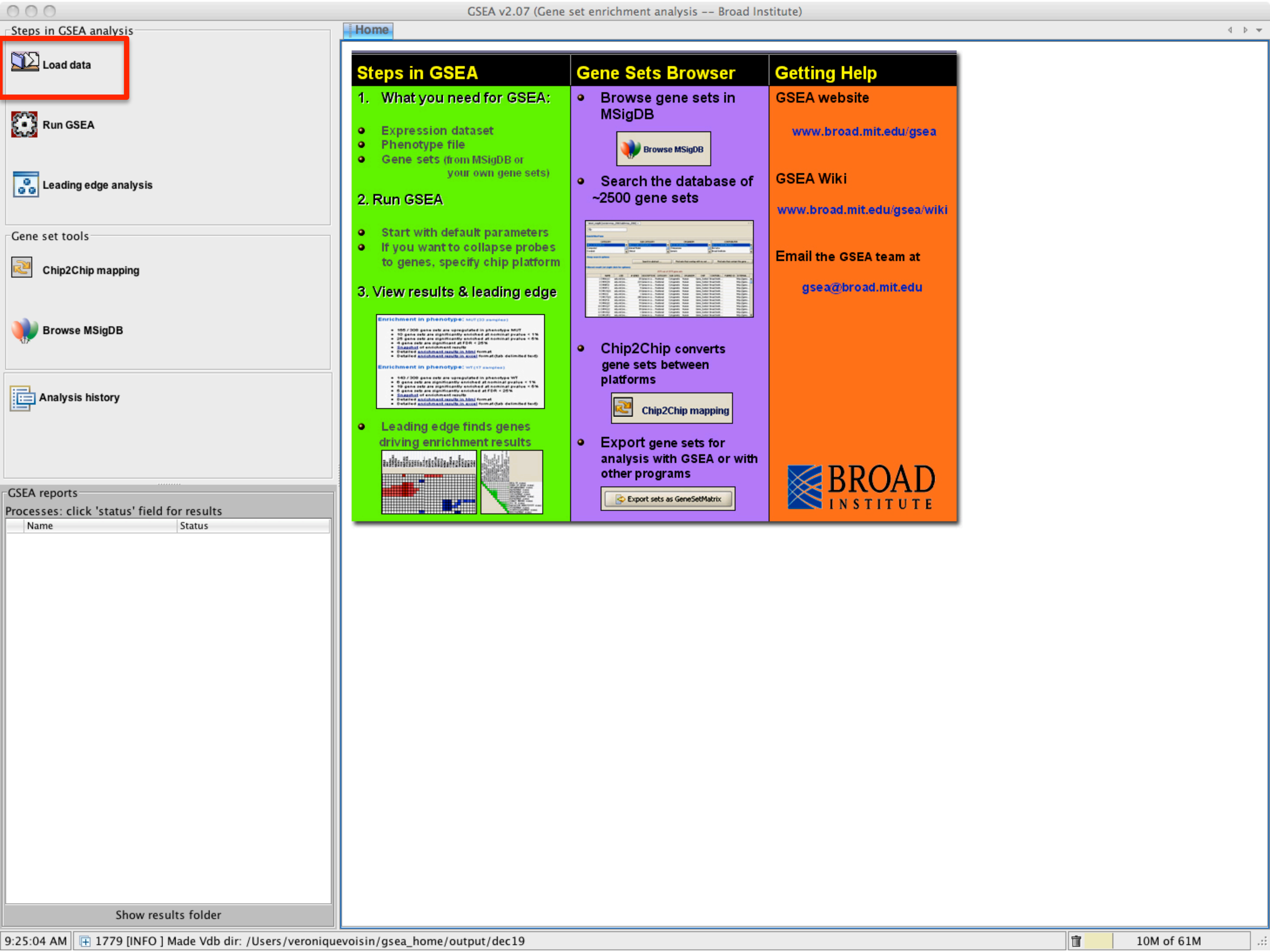
GSEA format:
http://www.broadinstitute.org/cancer/software/gsea/wiki/index.php/Data_formats

Spreadsheet:
Remove the duplicates= choose the best probe for each gene:

1) Rank the values using the absolute value of the expression value (t value, logFC, or –log10(pvalue)) in a decreasing order (largest to smallest)
2) Remove gene identifiers duplicates (official gene symbol or Entrez gene IDs)
3) Rank the list using the expression value in a decreasing order (largest to smallest)

Rank file:
Select the gene identifier and expression value and save it as tab delimited text file (.rnk)

Expression file:
Select the gene identifier, the gene full name, the raw normalized data for each sample and save it as a tab delimited file (.txt)



GSEA v2.07 (Gene set enrichment analysis -- Broad Institute)

Steps in GSEA

1. What you need for GSEA:

- Expression dataset
- Phenotype file
- Gene sets (from MSigDB or your own gene sets)

2. Run GSEA

- Start with default parameters
- If you want to collapse probes to genes, specify chip platform

3. View results & leading edge

- Leading edge finds genes driving enrichment results

Gene Sets Browser

- Browse gene sets in MSigDB

- Search the database of ~2500 gene sets
- Chip2Chip converts gene sets between platforms

- Export gene sets for analysis with GSEA or with other programs

Getting Help

GSEA website

www.broad.mit.edu/gsea

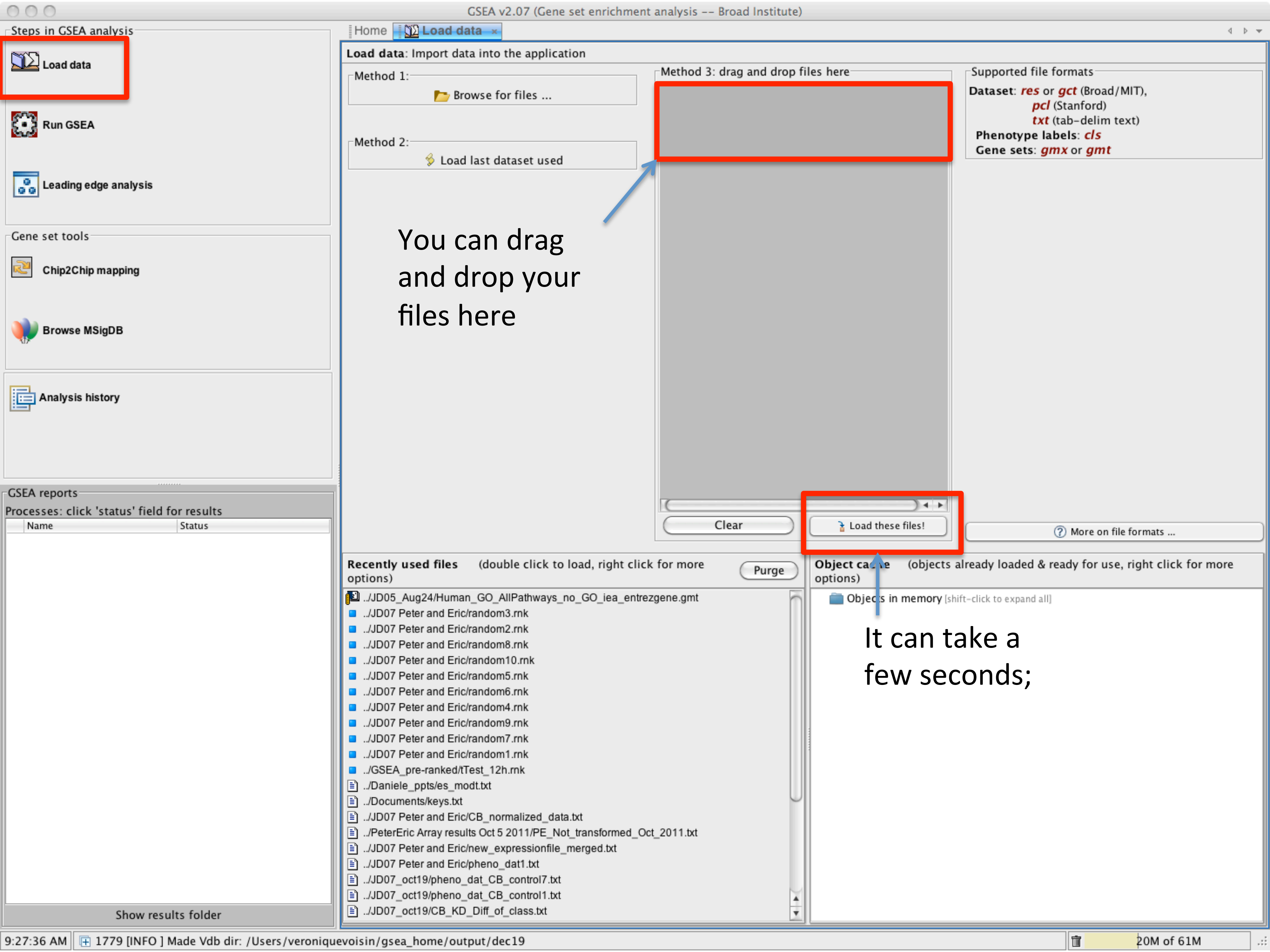
GSEA Wiki

www.broad.mit.edu/gsea/wiki

Email the GSEA team at

gsea@broad.mit.edu

```
9:25:04 AM  1779 [INFO ] Made Vdb dir: /Users/veroniquevoisin/gsea_home/output/dec19
```



Load data: Import data into the application

Dataset: *res* or *gct* (Broad/MIT),
pcl (Stanford)
txt (tab-delim text)
Phenotype labels: *cls*
Gene sets: *gmx* or *gmt*

Processes: click 'status' field for results

Recently used files (double click to load, right click for more options)

- ../JD05_Aug24/Human_GO_AllPathways_no_GO_iea_entrezgene.gmt
- ../JD07 Peter and Eric/random3.rnk
- ../JD07 Peter and Eric/random2.rnk
- ../JD07 Peter and Eric/random8.rnk
- ../JD07 Peter and Eric/random10.rnk
- ../JD07 Peter and Eric/random5.rnk
- ../JD07 Peter and Eric/random6.rnk
- ../JD07 Peter and Eric/random4.rnk
- ../JD07 Peter and Eric/random9.rnk
- ../JD07 Peter and Eric/random7.rnk
- ../JD07 Peter and Eric/random1.rnk
- ../GSEA_pre-ranked/tTest_12h.rnk
- ../Daniele_ppts/es_modt.txt
- ../Documents/keys.txt
- ../JD07 Peter and Eric/CB_normalized_data.txt
- ../PeterEric Array results Oct 5 2011/PE_Not_transformed_Oct_2011.txt
- ../JD07 Peter and Eric/new_expressionfile_merged.txt
- ../JD07 Peter and Eric/pheno_dat1.txt
- ../JD07_oct19/pheno_dat_CB_control7.txt
- ../JD07_oct19/pheno_dat_CB_control1.txt
- ../JD07_oct19/CB_KD_Diff_of_class.txt

Object cache (objects already loaded & ready for use, right click for more options)

```
9:27:36 AM   1779 [INFO ] Made Vdb dir: /Users/veroniquevoisin/gsea_home/output/dec19
```

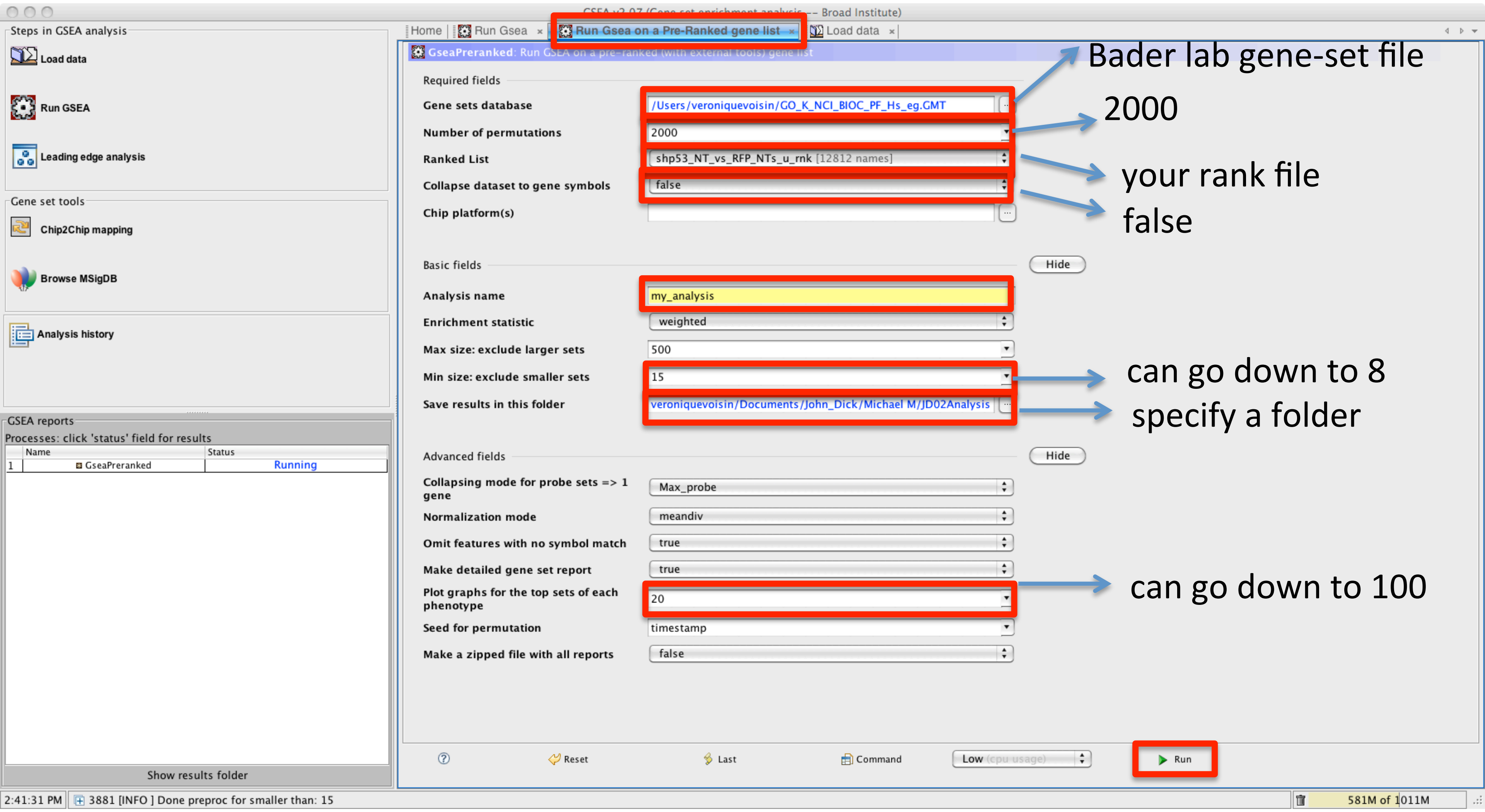
You can drag and drop your files here

It can take a few seconds;

Use the GSEA pre-ranked option to use GSEA with your ranked gene list (recommended)

1) Load Data -> Browse for files

2) Tools (menu bar) -> GSEAPreRanked



Processes: click 'status' field for results

	Name	Status
1	GseaPreranked	Running

GseaPreranked: Run GSEA on a pre-ranked (with external tools) gene list

Required fields

- Gene sets database: /Users/veroniquevoisin/GO_K_NCI_BIOC_PF_Hs_eg.GMT
- Number of permutations: 2000
- Ranked List: shp53_NT_vs_RFP_NTs_u_rnk [12812 names]
- Collapse dataset to gene symbols: false
- Chip platform(s):

Basic fields

- Analysis name: my_analysis
- Enrichment statistic: weighted
- Max size: exclude larger sets: 500
- Min size: exclude smaller sets: 15
- Save results in this folder: veroniquevoisin/Documents/John_Dick/Michael M/JD02Analysis

Advanced fields

- Collapsing mode for probe sets => 1 gene: Max_probe
- Normalization mode: meandiv
- Omit features with no symbol match: true
- Make detailed gene set report: true
- Plot graphs for the top sets of each phenotype: 20
- Seed for permutation: timestamp
- Make a zipped file with all reports: false

2:41:31 PM 3881 [INFO] Done preproc for smaller than: 15

Bader lab gene-set file

2000

your rank file

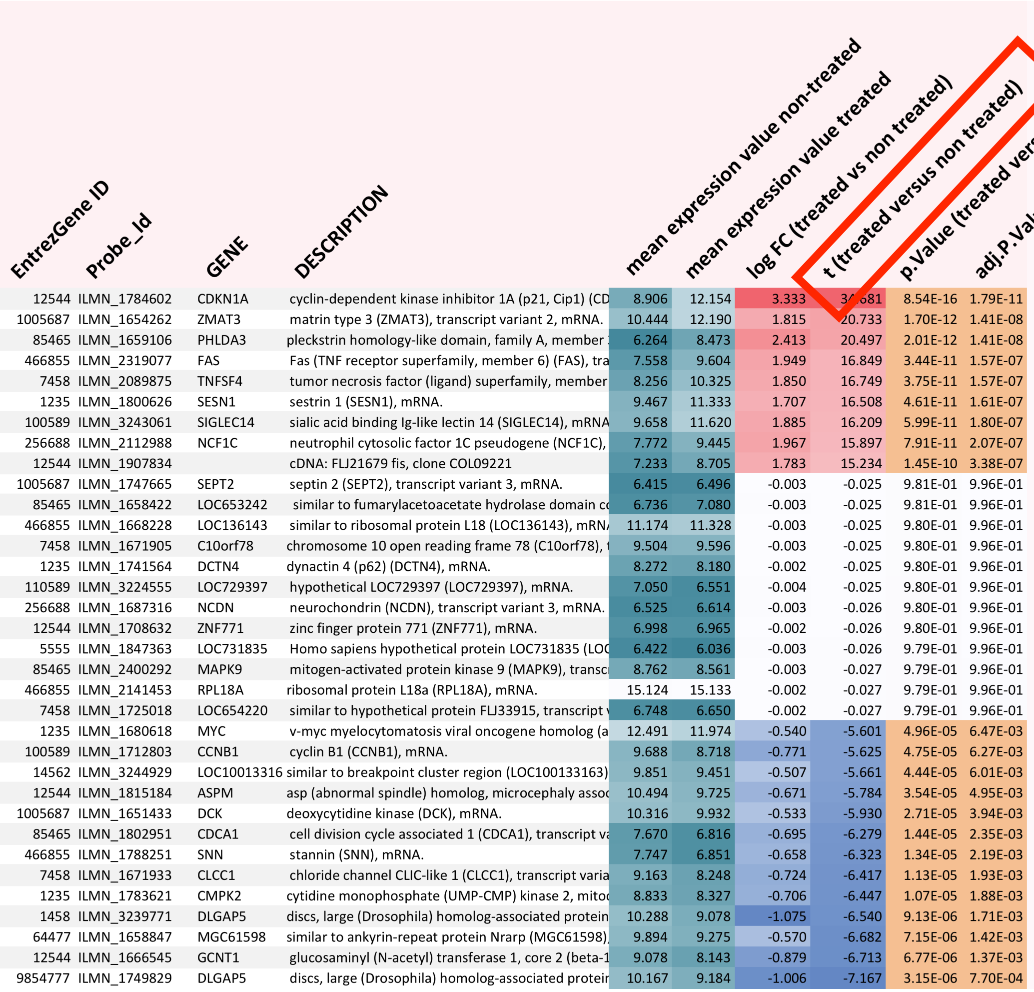
false

can go down to 8

specify a folder

can go down to 100

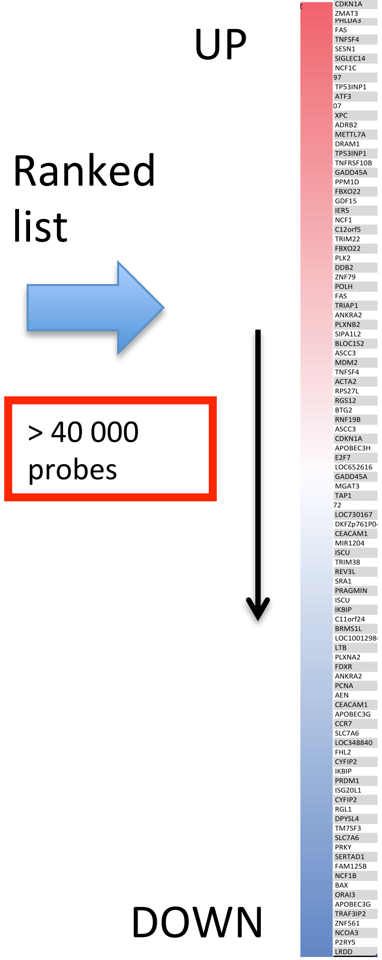
Rank the genes using a differential expression value



EntrezGene ID	Probe_Id	GENE	DESCRIPTION	mean expression value non-treated	mean expression value treated	log FC (treated vs non treated)	t (treated versus non treated)	p.Value (treated versus non treated)	adj.P.Value (treated versus non treated)
12544	ILMN_1784602	CDKN1A	cyclin-dependent kinase inhibitor 1A (p21, Cip1) (CD	8.906	12.154	3.333	34.681	8.54E-16	1.79E-11
1005687	ILMN_1654262	ZMAT3	matrin type 3 (ZMAT3), transcript variant 2, mRNA.	10.444	12.190	1.815	20.733	1.70E-12	1.41E-08
85465	ILMN_1659106	PHLDA3	pleckstrin homology-like domain, family A, member	6.264	8.473	2.413	20.497	2.01E-12	1.41E-08
466855	ILMN_2319077	FAS	Fas (TNF receptor superfamily, member 6) (FAS), tra	7.558	9.604	1.949	16.849	3.44E-11	1.57E-07
7458	ILMN_2089875	TNFSF4	tumor necrosis factor (ligand) superfamily, member	8.256	10.325	1.850	16.749	3.75E-11	1.57E-07
1235	ILMN_1800626	SESN1	sestrin 1 (SESN1), mRNA.	9.467	11.333	1.707	16.508	4.61E-11	1.61E-07
100589	ILMN_3243061	SIGLEC14	sialic acid binding Ig-like lectin 14 (SIGLEC14), mRNA	9.658	11.620	1.885	16.209	5.99E-11	1.80E-07
256688	ILMN_2112988	NCF1C	neutrophil cytosolic factor 1C pseudogene (NCF1C),	7.772	9.445	1.967	15.897	7.91E-11	2.07E-07
12544	ILMN_1907834		cDNA: FLJ21679 fis, clone COL09221	7.233	8.705	1.783	15.234	1.45E-10	3.38E-07
1005687	ILMN_1747665	SEPT2	septin 2 (SEPT2), transcript variant 3, mRNA.	6.415	6.496	-0.003	-0.025	9.81E-01	9.96E-01
85465	ILMN_1658422	LOC653242	similar to fumarylacetoacetate hydrolase domain c	6.736	7.080	-0.003	-0.025	9.81E-01	9.96E-01
466855	ILMN_1668228	LOC136143	similar to ribosomal protein L18 (LOC136143), mRN	11.174	11.328	-0.003	-0.025	9.80E-01	9.96E-01
7458	ILMN_1671905	C10orf78	chromosome 10 open reading frame 78 (C10orf78),	9.504	9.596	-0.003	-0.025	9.80E-01	9.96E-01
1235	ILMN_1741564	DCTN4	dynactin 4 (p62) (DCTN4), mRNA.	8.272	8.180	-0.002	-0.025	9.80E-01	9.96E-01
110589	ILMN_3224555	LOC729397	hypothetical LOC729397 (LOC729397), mRNA.	7.050	6.551	-0.004	-0.025	9.80E-01	9.96E-01
256688	ILMN_1687316	NCDN	neurochondrin (NCDN), transcript variant 3, mRNA.	6.525	6.614	-0.003	-0.026	9.80E-01	9.96E-01
12544	ILMN_1708632	ZNF771	zinc finger protein 771 (ZNF771), mRNA.	6.998	6.965	-0.002	-0.026	9.80E-01	9.96E-01
5555	ILMN_1847363	LOC731835	Homo sapiens hypothetical protein LOC731835 (LO	6.422	6.036	-0.003	-0.026	9.79E-01	9.96E-01
85465	ILMN_2400292	MAPK9	mitogen-activated protein kinase 9 (MAPK9), transc	8.762	8.561	-0.003	-0.027	9.79E-01	9.96E-01
466855	ILMN_2141453	RPL18A	ribosomal protein L18a (RPL18A), mRNA.	15.124	15.133	-0.002	-0.027	9.79E-01	9.96E-01
7458	ILMN_1725018	LOC654220	similar to hypothetical protein FLJ33915, transcript	6.748	6.650	-0.002	-0.027	9.79E-01	9.96E-01
1235	ILMN_1680618	MYC	v-myc myelocytomatosis viral oncogene homolog (a	12.491	11.974	-0.540	-5.601	4.96E-05	6.47E-03
100589	ILMN_1712803	CCNB1	cyclin B1 (CCNB1), mRNA.	9.688	8.718	-0.771	-5.625	4.75E-05	6.27E-03
14562	ILMN_3244929	LOC10013316	similar to breakpoint cluster region (LOC100133163)	9.851	9.451	-0.507	-5.661	4.44E-05	6.01E-03
12544	ILMN_1815184	ASPM	asp (abnormal spindle) homolog, microcephaly asso	10.494	9.725	-0.671	-5.784	3.54E-05	4.95E-03
1005687	ILMN_1651433	DCK	deoxycytidine kinase (DCK), mRNA.	10.316	9.932	-0.533	-5.930	2.71E-05	3.94E-03
85465	ILMN_1802951	CDCA1	cell division cycle associated 1 (CDCA1), transcript v	7.670	6.816	-0.695	-6.279	1.44E-05	2.35E-03
466855	ILMN_1788251	SNN	stannin (SNN), mRNA.	7.747	6.851	-0.658	-6.323	1.34E-05	2.19E-03
7458	ILMN_1671933	CLCC1	chloride channel CLIC-like 1 (CLCC1), transcript varia	9.163	8.248	-0.724	-6.417	1.13E-05	1.93E-03
1235	ILMN_1783621	CMPK2	cytidine monophosphate (UMP-CMP) kinase 2, mito	8.833	8.327	-0.706	-6.447	1.07E-05	1.88E-03
1458	ILMN_3239771	DLGAP5	discs, large (Drosophila) homolog-associated protein	10.288	9.078	-1.075	-6.540	9.13E-06	1.71E-03
64477	ILMN_1658847	MGC61598	similar to ankyrin-repeat protein Nrarp (MGC61598)	9.894	9.275	-0.570	-6.682	7.15E-06	1.42E-03
12544	ILMN_1666545	GCNT1	glucosaminyl (N-acetyl) transferase 1, core 2 (beta-1	9.078	8.143	-0.879	-6.713	6.77E-06	1.37E-03
9854777	ILMN_1749829	DLGAP5	discs, large (Drosophila) homolog-associated protei	10.167	9.184	-1.006	-7.167	3.15E-06	7.70E-04

Up-regulated in treated

Down-regulated in treated



UP

Ranked list

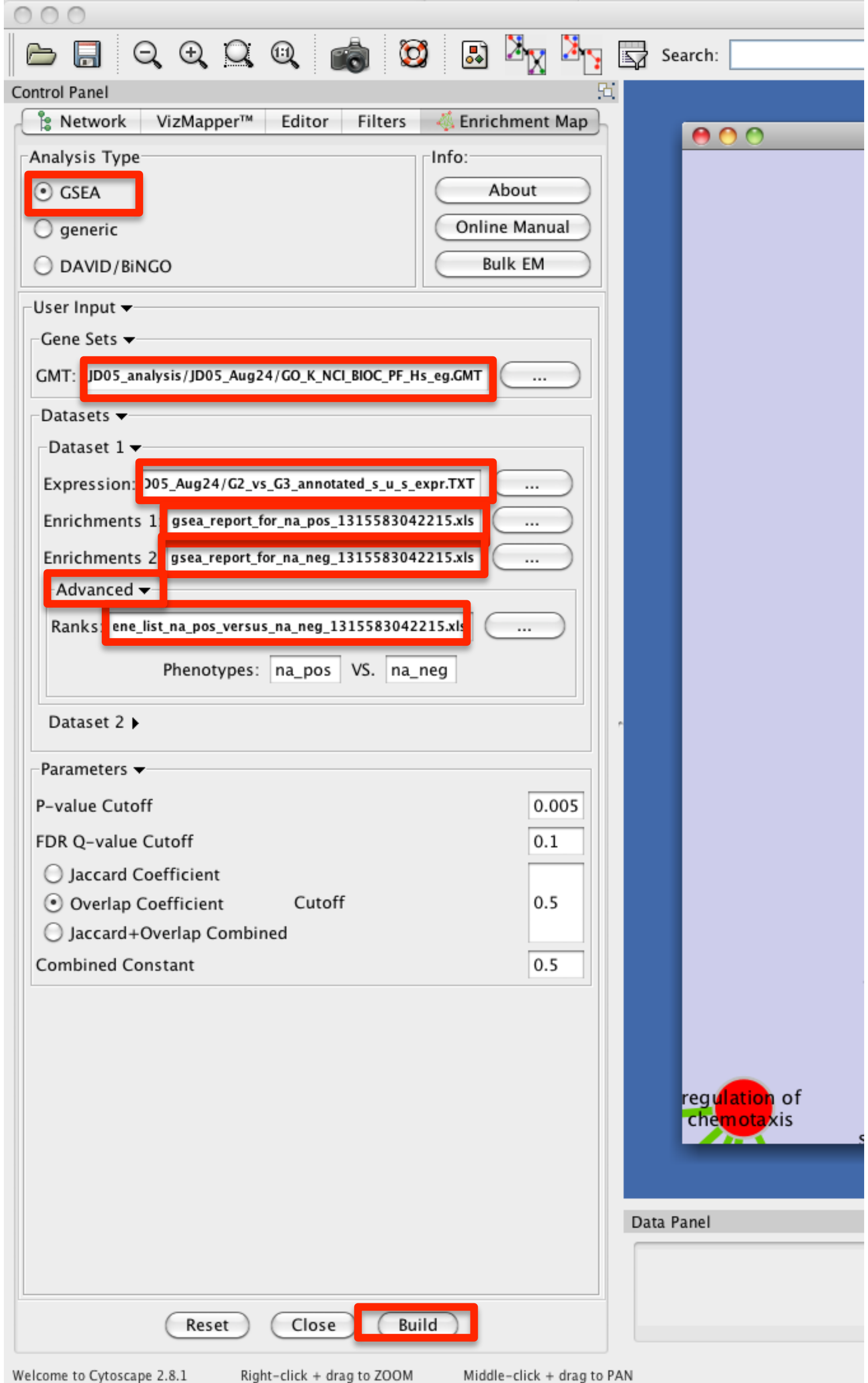
> 40 000 probes

DOWN

Cytoscape
-> Plugins (menu bar)
-> Enrichment Map
-> Control Panel
-> Load Enrichment Results

To run enrichment map , you need:

1) GMT file .GMT : provided (Baderlab gene-set file)
2) Expression file .TXT
3) GSEA reports (use the .rpt file in the GSEA folder to upload the rank file and the enrichment files)
4) rank file (from GSEA folder) or ulpoaded automatically if .rpt file has been used



Analysis Type
● GSEA
○ generic
○ DAVID/BiNGO

GMT: JD05_analysis/JD05_Aug24/GO_K_NCI_BIOC_PF_Hs_eg.GMT

Expression: D05_Aug24/G2_vs_G3_annotated_s_u_s_expr.TXT

Enrichments 1: gsea_report_for_na_pos_1315583042215.xls

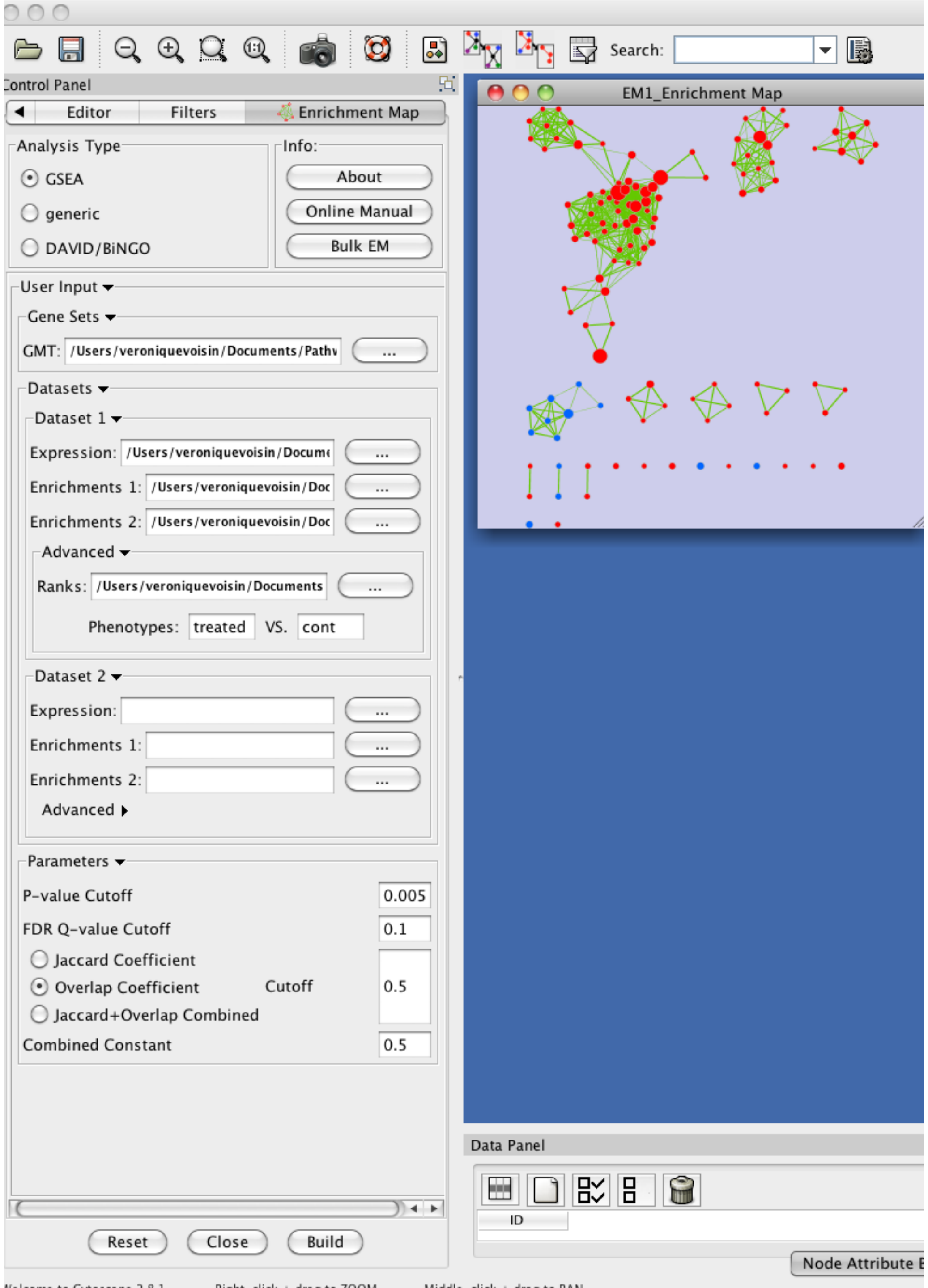
Enrichments 2: gsea_report_for_na_neg_1315583042215.xls

Ranks: ene_list_na_pos_versus_na_neg_1315583042215.xl

Phenotypes: na_pos VS. na_neg

P-value Cutoff 0.005
FDR Q-value Cutoff 0.1
○ Jaccard Coefficient
● Overlap Coefficient Cutoff 0.5
○ Jaccard+Overlap Combined
Combined Constant 0.5

Once the map is created:
Go to menu → view show graphics details
Go to Control panel → Vizmapper → Node label → EM1_GS_DESCR



Analysis Type
● GSEA
○ generic
○ DAVID/BiNGO

User Input

Gene Sets

GMT: /Users/veroniquevoisin/Documents/Pathv

Datasets

Dataset 1

Expression: /Users/veroniquevoisin/Docume

Enrichments 1: /Users/veroniquevoisin/Doc

Enrichments 2: /Users/veroniquevoisin/Doc

Advanced

Ranks: /Users/veroniquevoisin/Documents

Phenotypes: treated VS. cont

Dataset 2

Expression:

Enrichments 1:

Enrichments 2:

Advanced

Parameters

P-value Cutoff	0.005
FDR Q-value Cutoff	0.1
○ Jaccard Coefficient ● Overlap Coefficient ○ Jaccard+Overlap Combined	Cutoff 0.5
Combined Constant	0.5