

Basic analysis for Erny results

Filtering of data

- 50% of row < 12

- remove genes that had no gene symbol

- remove genes that I failed to attach entrez ID

- remove all genes of Pcdhgb1 family leave 1 behind.

- Keep PCDHGA1 from all family

Noise reduction

- A constant of 24 was added to the row data before LOG 2

Removal of column

- Disease 6 and healthy 5 were removed because they look different from other columns

Calculating pValue for TTEST

Separating to groups cut by TTEST pValue 0.01 and 100% fold change (after noise)

Top genes Disease Low

Symbol	Entrez ID	Description
MBOAT1	218121	membrane bound O-acyltransferase domain containing 1
E030024N20RIK	595139	RIKEN cDNA E030024N20 gene
CD2AP	12488	CD2-associated protein
LRRC51	69358	transmembrane O-methyltransferase
NOS1AP	70729	nitric oxide synthase 1 (neuronal) adaptor protein
TMCO6	71983	transmembrane and coiled-coil domains 6
FLI1	14247	flightless I homolog (Drosophila)
ELMO1	140580	engulfment and cell motility 1
NLRP1C	627984	NLR family, pyrin domain containing 1C, pseudogene
WNK1	232341	WNK lysine deficient protein kinase 1
SLC12A6	107723	solute carrier family 12, member 6
STARD3NL	76205	STARD3 N-terminal like
RPS4X	20102	ribosomal protein S4, X-linked
XIST	213742	inactive X specific transcripts
ACRBP	54137	proacrosin binding protein
IVNS1ABP	117198	influenza virus NS1A binding protein
LY86	17084	lymphocyte antigen 86
OPHN1	94190	oligophrenin 1
CCDC90A	67137	mitochondrial calcium uniporter regulator 1
HLTF	20585	helicase-like transcription factor
CTNND1	12388	catenin (cadherin associated protein), delta 1
MALAT1	72289	metastasis associated lung adenocarcinoma transcript 1 (non-coding RNA)

Top genes Disease High

Symbol	Entrez ID	Description
DDIT4	74747	DNA-damage-inducible transcript 4
TSC22D3	14605	TSC22 domain family, member 3
SERPINE2	20720	serine (or cysteine) peptidase inhibitor, clade E, member 2
SULT1A1	20887	sulfotransferase family 1A, phenol-preferring, member 1
PER1	18626	period circadian clock 1
NOTCH4	18132	notch 4

Separating to groups cut by TTEST pValue 0.01 and 50% fold change (after noise)

Results are two groups

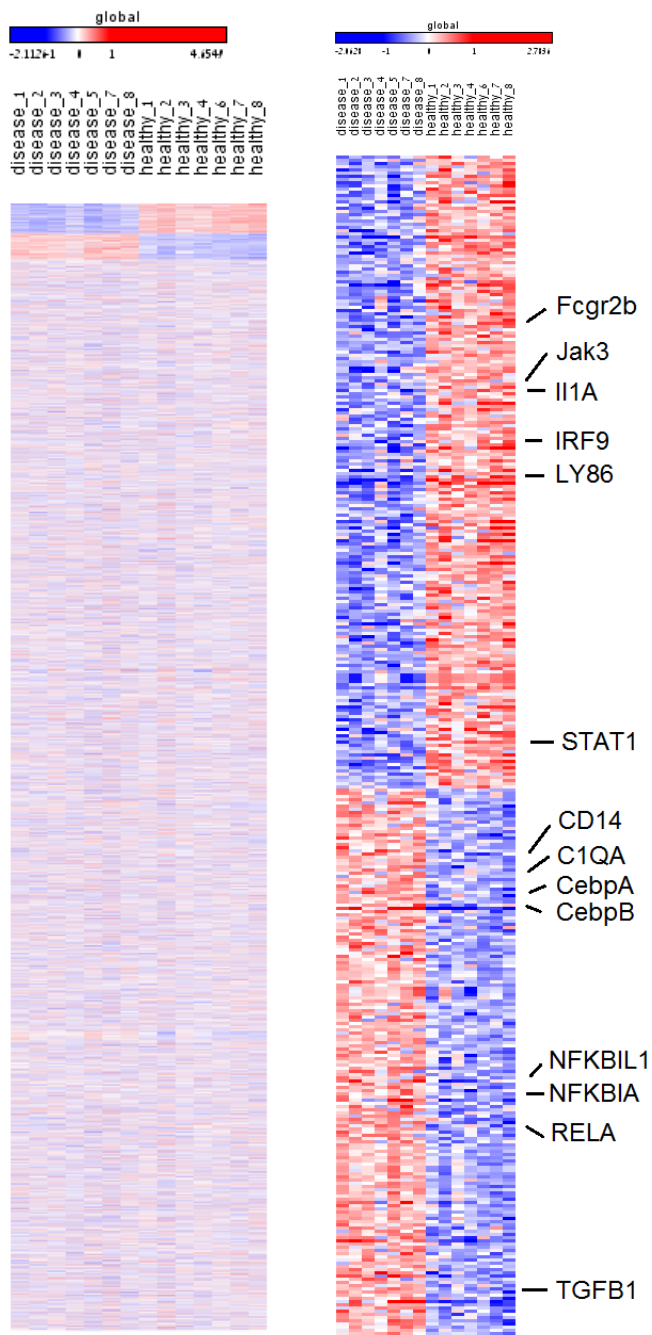
- 1) Disease Low with 198 genes
- 2) Disease High with 173 genes

View in Gene-E

Subtract average of value from each row to enhance the variety inside each row

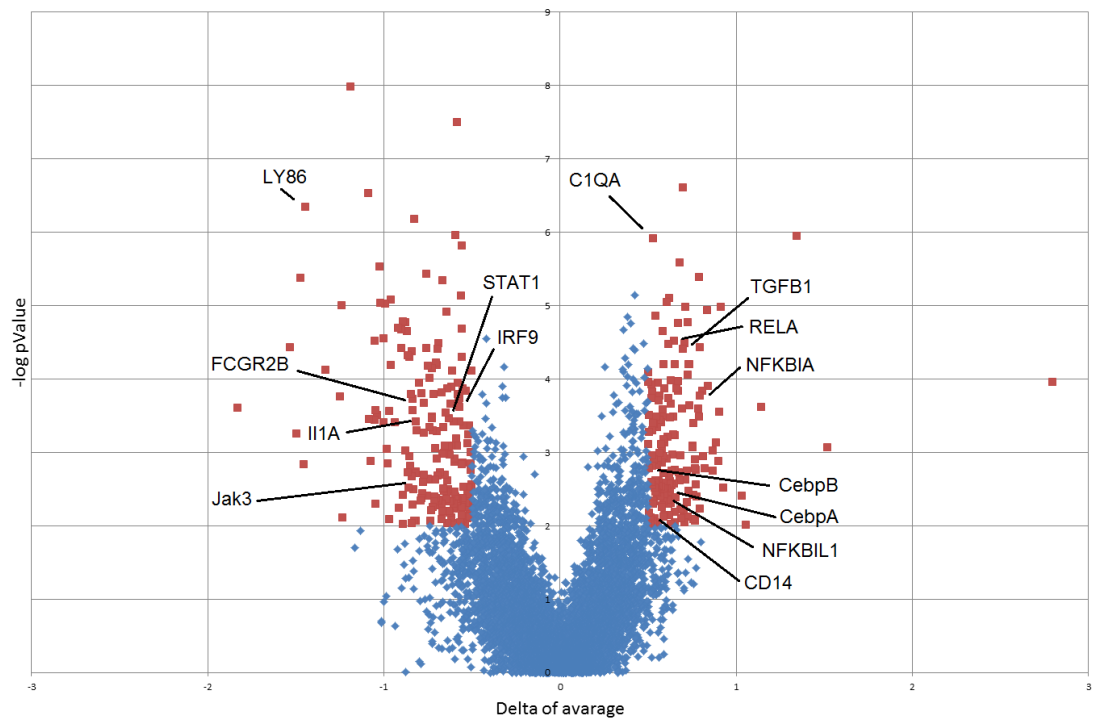
Left – the ratio of the changing groups to all genes

Right – Same two groups without other genes



Volcano plot

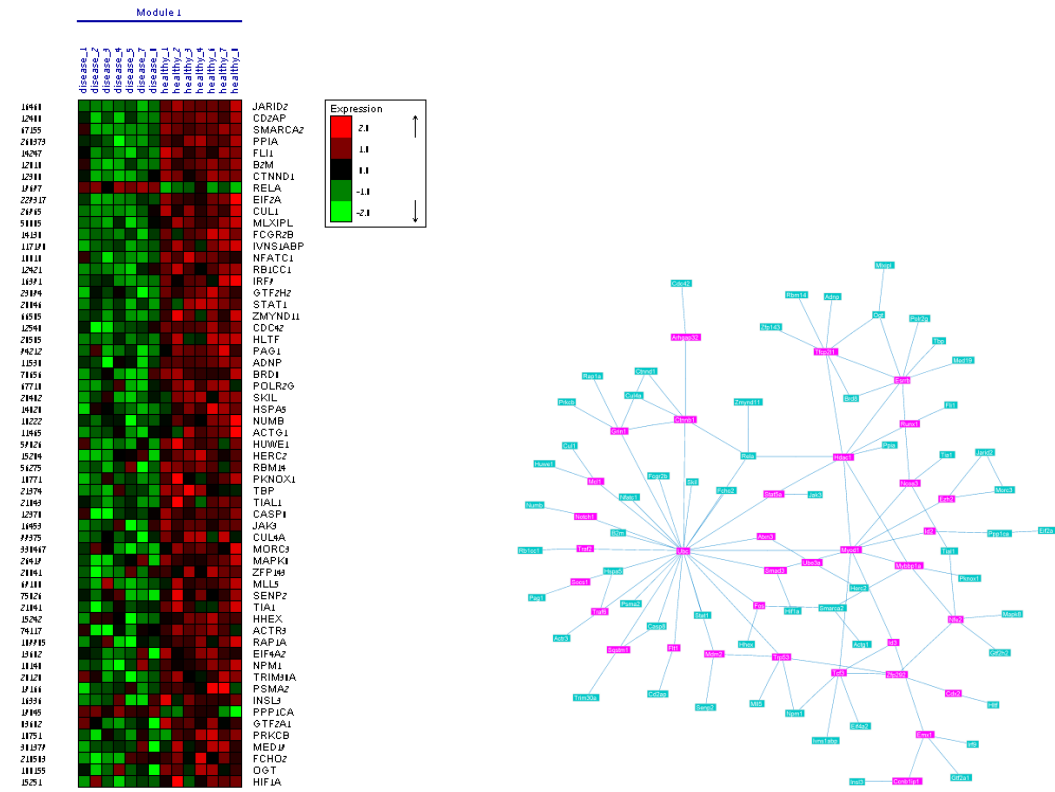
All data are seen in this plot in red are the genes picked in GENE-E



Enrichment results

Clustering in Expander for down regulated disease data.

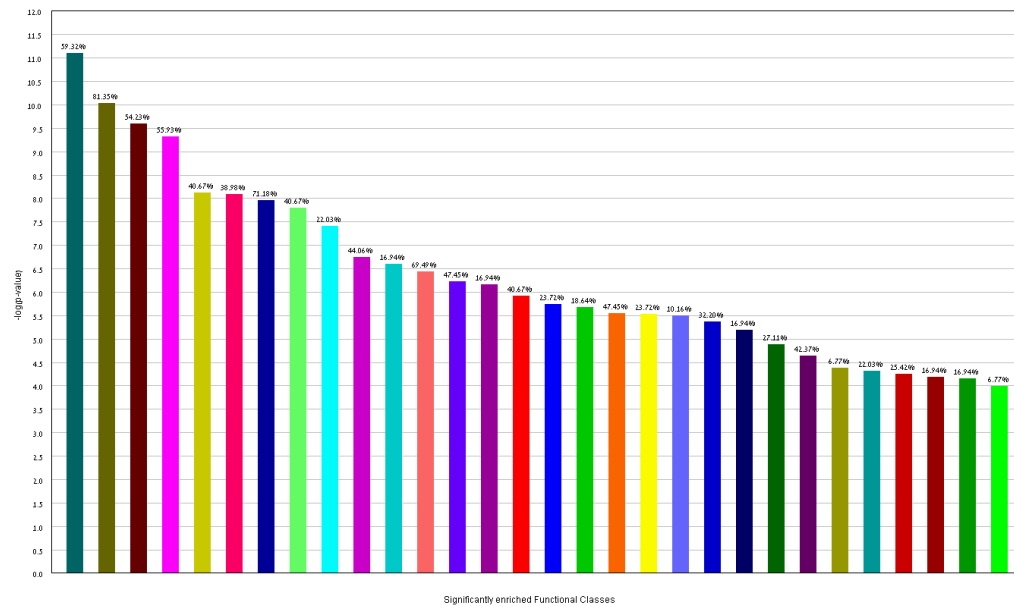
Degas algorithm with significance 0.05 and protein-protein networks.



Many results among them ...

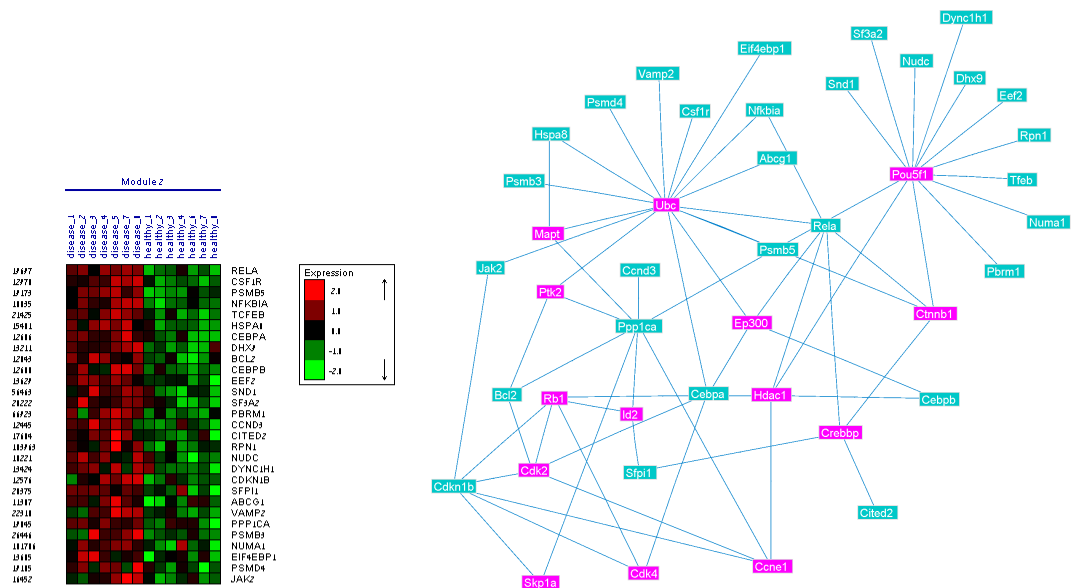
Term	#Protein from list	Hyper Raw P-Value
cell differentiation - GO:0030154	INSL3, JARID2, RELA, ADNP, CTNND1, B2M, ACTR3, ACTG1, CDC42, HHEX, MLL5, HIF1A, PKNOX1, HUWE1, PPIA, TIAL1, CASP8, NUMB, MAPK8, JAK3, SKIL, SMARCA2, NFATC1	8.09E-09
regulation of immune system process - GO:0002682	TRIM30A, MLL5, HIF1A, FCGR2B, RELA, JAK3, OGT, PAG1, B2M, PRKCB	6.58E-05
immune system process - GO:0002376	JARID2, RELA, STAT1, B2M, PRKCB, HHEX, MLL5, HIF1A, PKNOX1, FCGR2B, CASP8, JAK3, SKIL	4.92E-05

All GO found



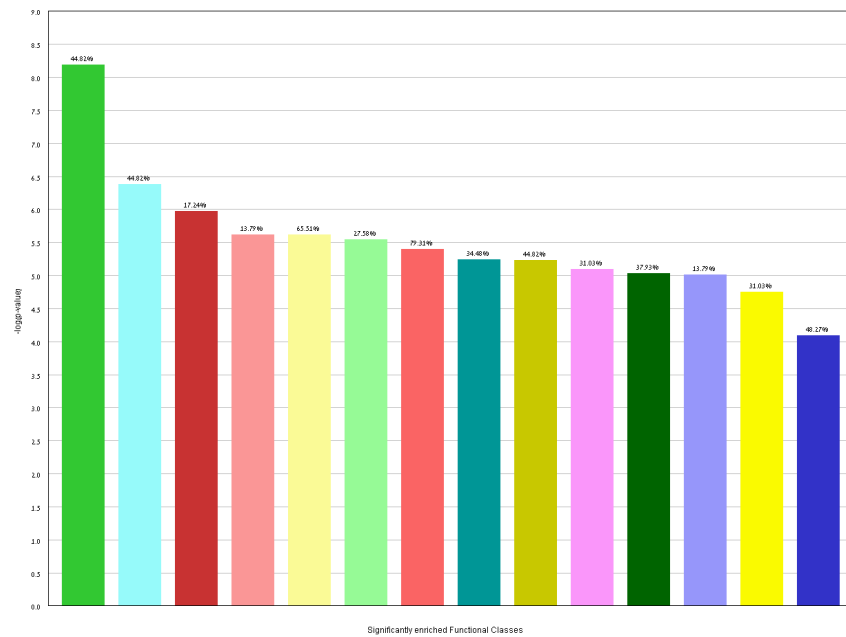
- regulation of multicellular organismal development - GO: 2000026
- liver development - GO: 0001889
- signaling - GO: 0023052
- regulation of cell proliferation - GO: 0042127
- positive regulation of biological process - GO: 0048518
- negative regulation of signaling - GO: 0023057
- DNA binding - GO: 0003677
- positive regulation of developmental process - GO: 0051094
- enzyme binding - GO: 0019899
- positive regulation of metabolic process - GO: 0009893
- negative regulation of cellular process - GO: 0048523
- regulatory region DNA binding - GO: 0000975
- cellular macromolecule metabolic process - GO: 0044260
- chromatin organization - GO: 0006325
- regulation of immune system process - GO: 0002682
- repressing transcription factor binding - GO: 0070491
- transcription from RNA polymerase II promoter - GO: 0006366
- immune system process - GO: 0002376
- positive regulation of cellular component organization - GO: 0051130
- regulation of protein metabolic process - GO: 0051246
- developmental process - GO: 0032502
- regulation of cellular process - GO: 0050794
- cellular component organization - GO: 0016043
- regulation of cellular biosynthetic process - GO: 0031326
- cell differentiation - GO: 0030154
- protein metabolic process - GO: 0019538
- cellular response to stimulus - GO: 0051716
- positive regulation of neurogenesis - GO: 0050769
- transcription, DNA-dependent - GO: 0006351
- protein binding - GO: 0005515

Matisse algorithm with pre filter for the most differential genes, limit module size between 100-30



Many results among them ...

Term	#Protein from list	Hyper Raw P-Value
transcription factor binding - GO:0008134	CEBPA, DHX9, CEBPB, RELA, BCL2, NFKBIA, SFPI1, CITED2	2.83E-06
immune system process - GO:0002376	CEBPA, CCND3, RELA, BCL2, NFKBIA, JAK2, SFPI1, TCFEB, CITED2, CSF1R	5.73E-06



- protein binding - GO:0005515
- regulation of cell proliferation - GO:0042127
- positive regulation of metabolic process - GO:0009893
- regulation of macromolecule metabolic process - GO:0060255
- immune system process - GO:0002376
- regulation of cell differentiation - GO:0045595
- organ development - GO:0048513
- myeloid leukocyte differentiation - GO:0002573
- transcription factor binding - GO:0008134
- RNA polymerase II distal enhancer sequence-specific DNA binding transcription factor
- negative regulation of biological process - GO:0048519
- regulation of protein metabolic process - GO:0051246
- response to organic substance - GO:0010033
- placenta development - GO:0001890