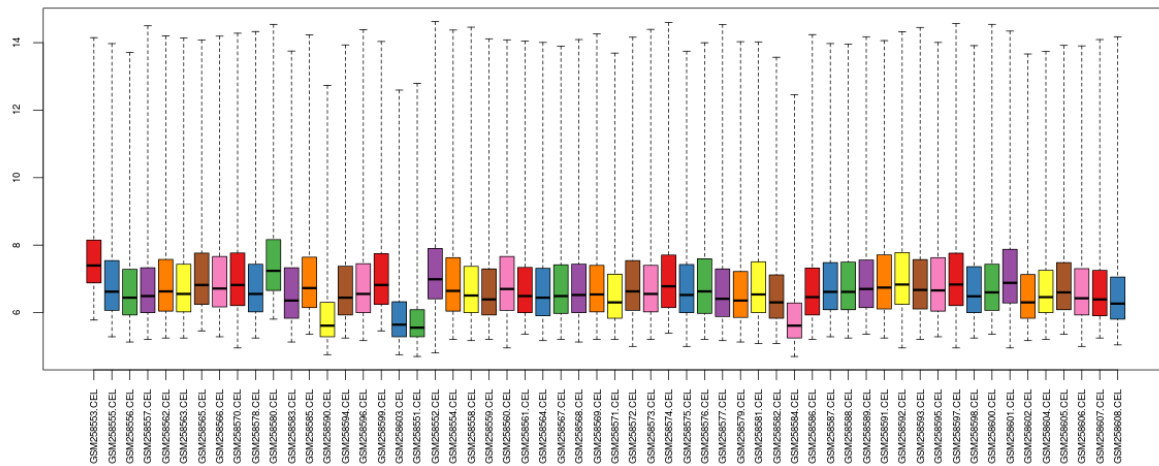
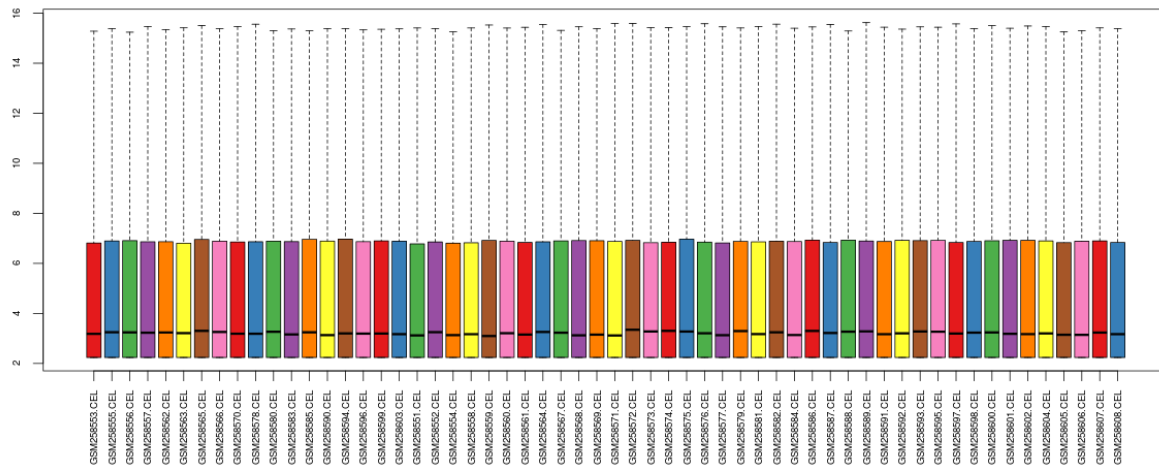


1. Box Plots



Box Plot for unnormalized data



Box Plot for normalized data(GC-RMA)

2. Probesets with $|\log FC| > 2, 3, 4$

$|\log FC| > 2 \rightarrow 480$

$|\log FC| > 3 \rightarrow 150$

$|\log FC| > 4 \rightarrow 67$

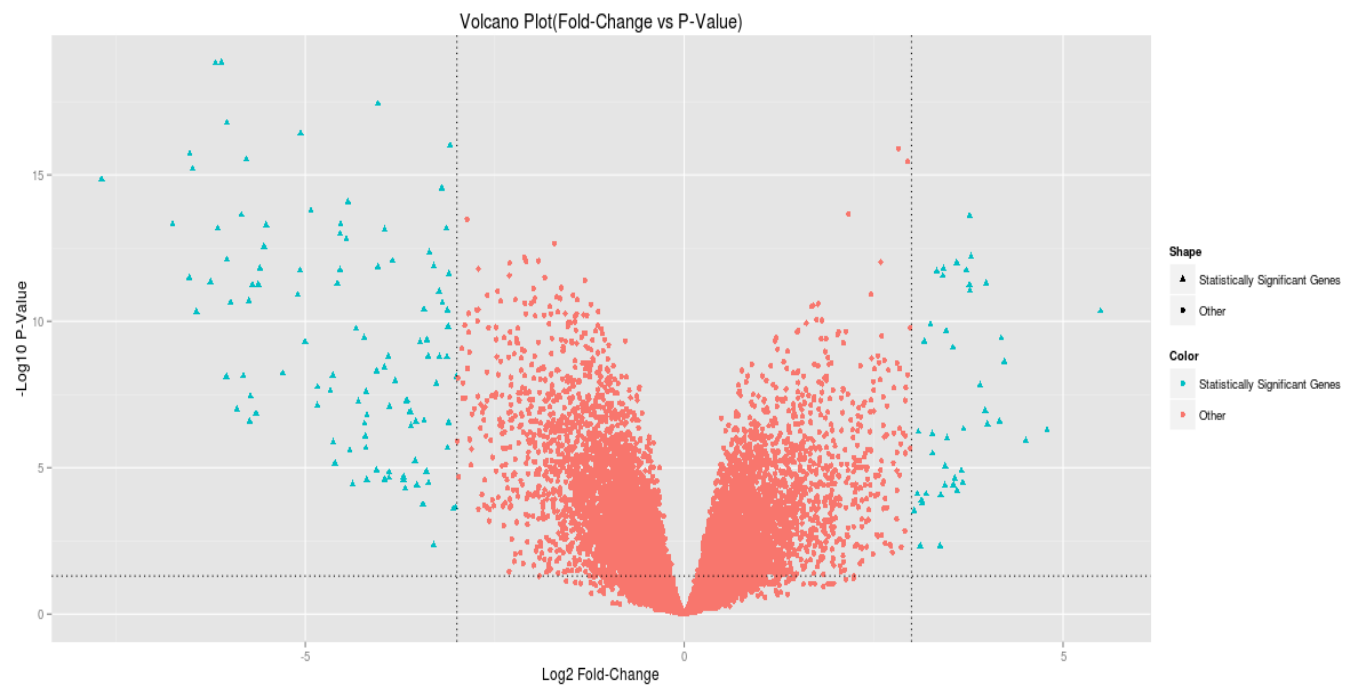
3. Top 10 up-regulated genes in Adenocarcinoma

	"logFC"	"AveExpr"	"t"	"P.Value"	"adj.P.Val"	"B"	"Symbols"
<hr/>							
"244056_at"	5.49	9.05	8.03	4.48e-11	1.34e-08	15.04	"SFTA2"
"206239_s_at"	4.78	8.91	5.61	5.33e-07	2.14e-05	5.98	"SPINK1"
"228979_at"	4.50	9.72	5.40	1.21e-06	4.04e-05	5.18	"SFTA3"
"231315_at"	4.22	7.87	7.01	2.38e-09	3.21e-07	11.20	NA
"203757_s_at"	4.18	11.37	7.48	3.81e-10	7.23e-08	12.97	"CEACAM6"
"206754_s_at"	4.16	6.55	5.80	2.62e-07	1.21e-05	6.66	"CYP2B6"
"220393_at"	4.00	5.56	5.75	3.19e-07	1.41e-05	6.47	"LGSN"
"229030_at"	3.98	5.93	8.59	5.05e-12	2.43e-09	17.15	"CAPN8"
"204437_s_at"	3.97	8.79	6.03	1.10e-07	6.21e-06	7.49	"FOLR1"
"211024_s_at"	3.90	8.67	6.54	1.52e-08	1.32e-06	9.41	"NKX2-1"
"209173_at"	3.78	11.58	9.13	6.22e-13	5.48e-10	19.18	"AGR2"

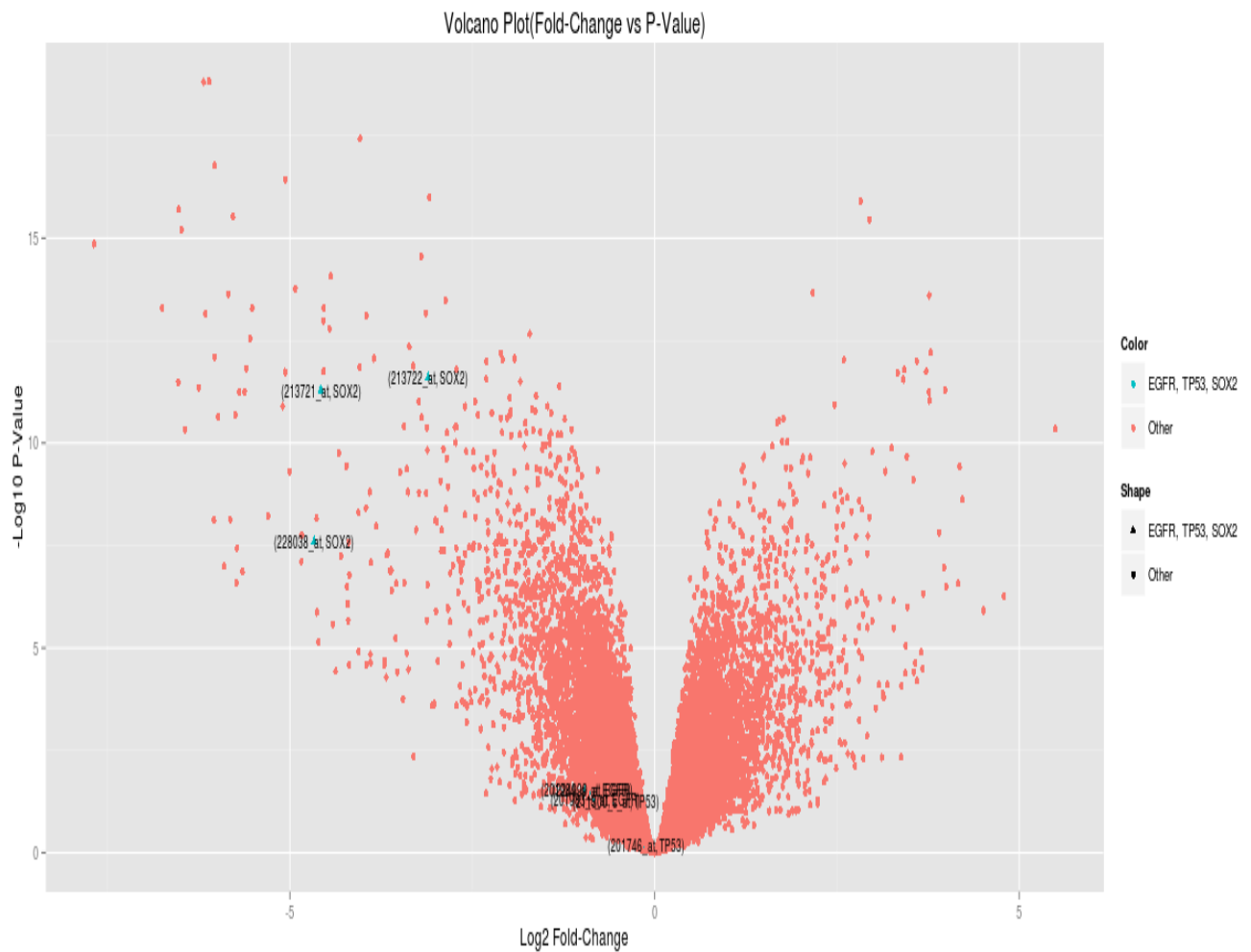
4. Top 10 down-regulated genes in Adenocarcinoma

	"logFC"	"AveExpr"	"t"	"P.Value"	"adj.P.Val"	"B"	"Symbols"
<hr/>							
"201820_at"	-7.68	5.65	-10.75	1.38e-15	3.14e-12	25.05	"KRT5"
"235075_at"	-6.75	4.79	-9.79	4.98e-14	6.28e-11	21.61	"DSG3"
"232082_x_at"	-6.53	4.99	-8.70	3.26e-12	1.68e-09	17.58	"SPRR3"
"206165_s_at"	-6.52	4.51	-11.29	1.94e-16	6.63e-13	26.94	"CLCA2"
"217528_at"	-6.48	4.79	-10.97	6.18e-16	1.53e-12	25.83	"CLCA2"
"213796_at"	-6.43	5.15	-8.02	4.63e-11	1.37e-08	15.01	"SPRR1A"
"209863_s_at"	-6.24	5.96	-8.62	4.45e-12	2.21e-09	17.27	"TP63"
"206032_at"	-6.18	4.72	-13.33	1.53e-19	2.09e-15	33.73	"DSC3"
"221796_at"	-6.15	5.05	-9.71	6.84e-14	7.79e-11	21.30	"NTRK2"
"206033_s_at"	-6.10	4.33	-13.35	1.44e-19	2.09e-15	33.79	"DSC3"

5. Volcano Plot for probe-sets with genes with |logFC| > 3 and p-value < 0.05 highlighted



6. Volcano Plot for probe-sets with genes EGFR, TP53 and SOX2 labelled



Labeled probe-sets:

(228038_at, SOX2)
 (213721_at, SOX2)
 (213722_at, SOX2)
 (201984_s_at, EGFR)
 (224999_at, EGFR)
 (201983_s_at, EGFR)
 (211300_s_at, TP53)
 (201746_at, TP53)

7. Enrichment Analysis

- Top 5 most enriched GO Biological Process terms

ID	Name	Source	pValue	FDR B&H	FDR B&Y	Bonferroni	Genes from Input	Genes in Annotation
1	GO:0060429	epithelium development	4.387E-18	1.615E-14	1.419E-13	1.615E-14	62	1110
2	GO:0008544	epidermis development	1.152E-16	2.121E-13	1.864E-12	4.242E-13	32	316
3	GO:0043588	skin development	6.127E-16	7.517E-13	6.606E-12	2.255E-12	33	358
4	GO:0009888	tissue development	2.225E-14	2.048E-11	1.800E-10	8.191E-11	74	1794
5	GO:0030855	epithelial cell differentiation	1.632E-12	1.202E-9	1.056E-8	6.008E-9	37	589

Corresponding genes in the input:

1. epithelium development(62 from input)

SOX2,CRLF1,SPRR1A,SPRR1B,SPRR3,E2F7,VSIG1,IRF6,SCEL,FOXA2,PTHLH,PHGDH,ADM,S100A7,BNC1,BMP7,SERPINB5,GJB5,CLIC5,GPC3,HPN,UPK1B,JAG1,CGN,COBL,DLG1,DLX5,DLX6,FGFR2,MMP12,FST,MLPH,FOXD1,SFN,FOXO1,KRT5,KRT6B,COL7A1,HNF1B,SHROOM3,KRT14,KRT15,KRT16,WNT5A,AGR2,SOSTDC1,CPS1,AQP3,TFAP2A,SYNE4,VANG1,GSTM3,ZNF750,DMRT2,NKX2-1,TP63,SNAIL,CSTA,GRHL3,IL1A,IRX2,CTSV

2. epidermis development(32 from input)

SPRR1A,SPRR1B,SPRR3,IRF6,SCEL,PTHLH,S100A7,BNC1,GJB5,CLIC5,JAG1,FGFR2,FST,SFN,FOXO1,KRT5,KRT6B,COL7A1,KRT14,KRT15,KRT16,WNT5A,SOSTDC1,AQP3,TFAP2A,VANG1,ZNF750,TP63,CSTA,GRHL3,IL1A,CTSV

3. skin development(33 from input)

SPRR1A,SPRR1B,SPRR3,IRF6,SCEL,PTHLH,S100A7,BNC1,GJB5,CLIC5,JAG1,FGFR2,FST,SFN,FOXO1,KRT5,KRT6B,COL7A1,KRT14,KRT15,KRT16,WNT5A,SOSTDC1,AQP3,TFAP2A,CLDN1,VANG1,ZNF750,TP63,CSTA,GRHL3,IL1A,CTSV

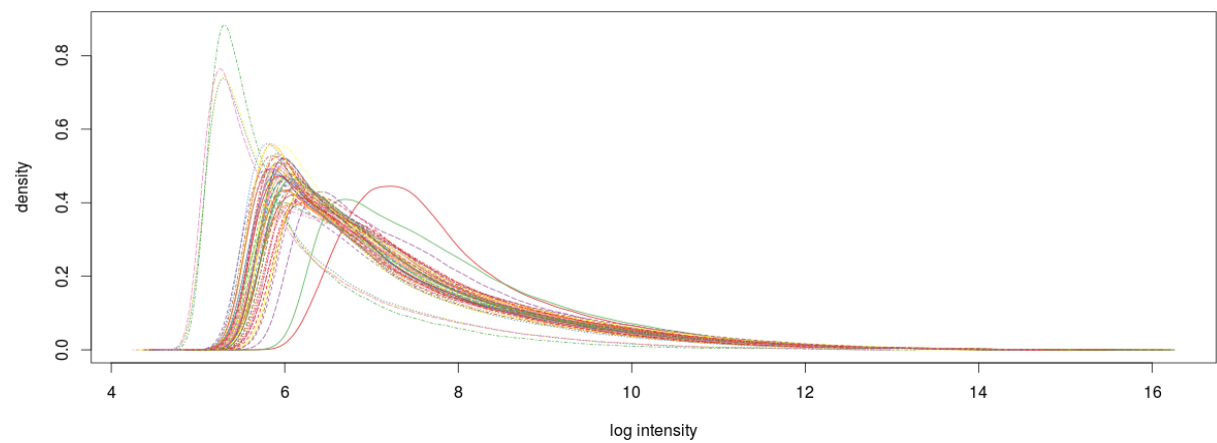
4. tissue development(74 from input)

SOX2,LPCAT1,CRLF1,GCNT3,SPRR1A,SPRR1B,SPRR3,E2F7,VSIG1,IRF6,EYA2,SCEL,FOXA2,PTHLH,PHGDH,ADM,S100A7,BNC1,BMP7,SERPINB5,GJB5,FBN2,CLIC5,GPC3,HOXD10,HPN,UPK1B,JAG1,CGN,PITX1,PLA2G4A,COBL,DLG1,DLX5,DLX6,FGFR2,FGL1,MMP12,FST,MLPH,FOXD1,SFN,FOXO1,GPC1,KRT5,KRT6B,COL7A1,HNF1B,SHROOM3,KRT14,KRT15,KRT16,WNT5A,DUSP4,AGR2,SOSTDC1,ARTN,CPS1,AQP3,TFAP2A,SYNE4,VANG1,GSTM3,ZNF750,DMRT2,NKX2-1,TP63,SNAIL,CSTA,RORC,GRHL3,IL1A,IRX2,CTSV

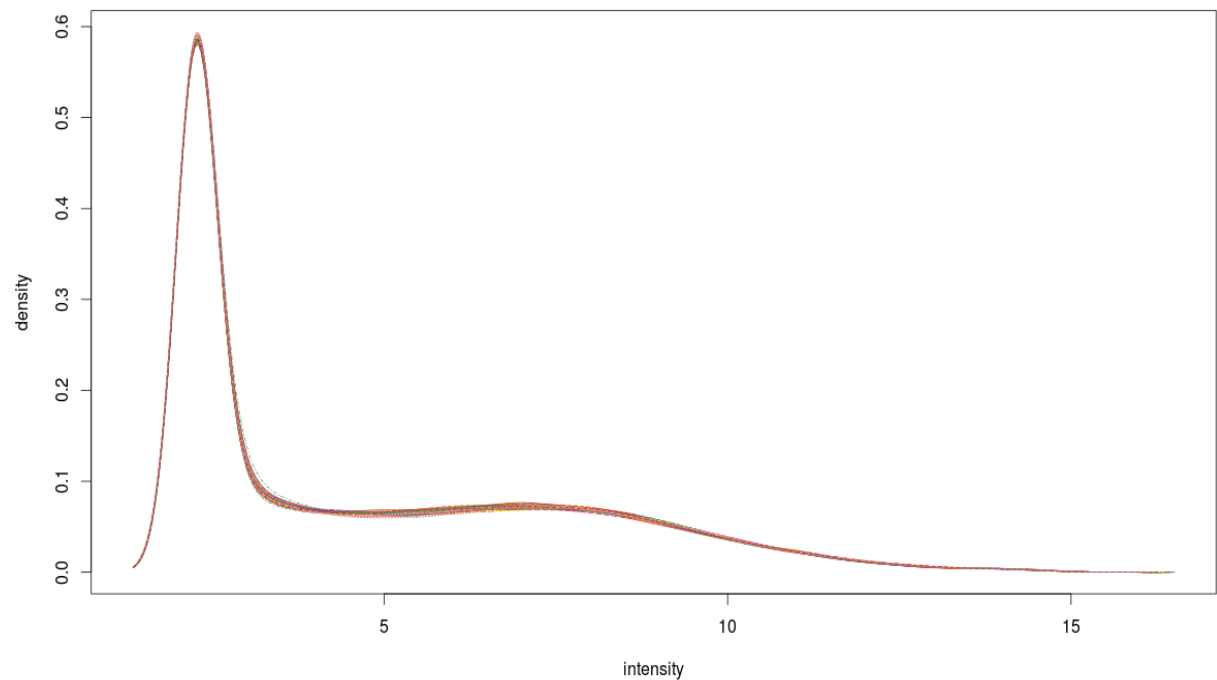
5. epithelial cell differentiation(37 from input)

SOX2,SPRR1A,SPRR1B,SPRR3,E2F7,VSIG1,IRF6,SCEL,FOXA2,PTHLH,S100A7,BMP7,CLIC5,UPK1B,JAG1,CGN,DLX5,DLX6,FGFR2,MLPH,SFN,KRT6B,HNF1B,SHROOM3,KRT14,WNT5A,AGR2,CPS1,AQP3,TFAP2A,SYNE4,GSTM3,NKX2-1,TP63,CSTA,IL1A,CTSV

b. Histograms before and after normalization



Histogram before Normalization



Histogram after Normalization

c. Other significantly enriched things relevant to the study

ID	Name	Source	pValue	FDR B&H	FDR B&Y	Bonferroni	Genes from Input	Genes in Annotation	
1	KRT	Keratins	genenames.org	2.354E-7	1.036E-5	4.528E-5	1.036E-5	8	61
2	CLDN	Claudins	genenames.org	7.944E-5	1.748E-3	7.642E-3	3.495E-3	4	22
3	ABC	ATP-binding cassette transporters	genenames.org	2.646E-3	3.881E-2	1.697E-1	1.164E-1	4	54

The table above shows the gene family that's most significantly expressed in non-small cell lung cancer.

As we can see from the table before, Epithelium Development seems to be the most significant biological process related to the NSCLC and Keratin filaments are present in epithelial cells. So it verifies the fact that Keratins play a major role in the study.

Genetic silencing of ABCA3 in the NSCLC cell line models A549, NCI-H1650 and NCI-H1975 significantly increased tumor cell susceptibility to the cytostatic effects of both cisplatin (in all cell lines) and paclitaxel (in two of three cell lines). Taken together, ABCA3 emerges as a modulator of NSCLC cell susceptibility to cytostatic therapy.

The claudins, members of a large family of adherent junction proteins, regulate the integrity and function of tight junctions and claudin gene expression is frequently altered in several human cancers. CLDN-1, -4 and -5 have especially known to have different expressions in NSCLC.

8. Table for top 3 most enriched Pathways

	ID	Name	Source	pValue	FDR B&H	FDR B&Y	Bonferroni	Genes from Input	Genes in A
1	M15394	Acute Myocardial Infarction	MSigDB C2: BioCarta	1.993E-8	2.025E-5	1.519E-4	2.025E-5	7	
2	M15997	Intrinsic Prothrombin Activation Pathway	MSigDB C2: BioCarta	6.043E-8	3.070E-5	2.303E-4	6.140E-5	7	
3	M15422	Vitamin C in the Brain	MSigDB C2: BioCarta	5.076E-7	1.719E-4	1.289E-3	5.157E-4	5	
4	730306	Assembly of collagen fibrils and other multimeric structures	BioSystems: REACTOME	2.670E-6	6.782E-4	5.087E-3	2.713E-3	8	
5	P00050	Plasminogen activating cascade	PantherDB	6.269E-6	1.274E-3	9.556E-3	6.370E-3	5	

Genes in the input corresponding to the pathways:

1. Acute Myocardial Infraction

FGA,FGB,FGG,COL4A3,COL4A4,COL4A5,COL4A6

2. Intrinsic Prothrombin Activation Pathway

FGA,FGB,FGG,COL4A3,COL4A4,COL4A5,COL4A6

3. Vitamin C in the Brain

COL4A3,COL4A4,COL4A5,COL4A6,SLC2A1

4. Assembly of collagen fibrils

DST,MMP3,COL4A3,COL4A4,COL4A5,COL4A6,COL7A1,CTSV

5. Plasminogen activating cascade

FGA,FGB,FGG,MMP3,SERPINB2

References:

1. <http://www.ncbi.nlm.nih.gov/pubmed/23689165> - Intracellular ATP-binding cassette transporter A3 is expressed in lung cancer cells and modulates susceptibility to cisplatin and paclitaxel
2. <http://www.ncbi.nlm.nih.gov/pubmed/19231096> - Diagnostic utility of expression of claudins in non-small cell lung cancer: different expression profiles in squamous cell carcinomas and adenocarcinomas
3. <http://en.wikipedia.org/wiki/Keratin>
4. <http://www.cookbook-r.com/Graphs/>
5. <http://www.statmethods.net/advgraphs/ggplot2.html>
6. Bioconductor Case Studies - Florian Hahne, Wolfgang Huber, Robert Gentleman, Seth Falcon