

- Keywords were selected as “expression, liver, tumor”, and dataset[1] belongs to species: mus musculus shown in **Figure 1** was chosen from NCBI-GEO.

Cirrhosis is malfunctioning of the liver tissue due to long-term (unknown) damage, and the factors behind this which predisposes Hepatocellular Carcinoma (HCC) has not clearly shown, yet[1]. Pathogen-free C57BL/6 mice with transgenic expression of Platelet-Derived -Growth Factor-c (PDGFC-Tg) inducing mice were used (control: Wild Type mice), which shares similar microenvironmental conditions with HCC, and mouse is a good model to study this disease mimicking the human version. In this study, researchers tried to find out the mechanism behind HCC in PDGFC-Tg mice.

The screenshot displays the NCBI GEO Dataset Browser interface. At the top, the NCBI logo and 'CURATED DATASET BROWSER' are visible. A search bar contains 'GDS5320[ACCN]'. Below the search bar, the dataset record for GDS5320 is shown, including tabs for 'Expression Profiles', 'Data Analysis Tools', and 'Sample Subsets'. The dataset title is 'Platelet-derived growth factor C transgenic model of hepatocellular carcinoma: liver stromal cells'. The summary describes the analysis of liver stroma from 8.8-week-old PDGF-C transgenics. The organism is 'Mus musculus'. The platform is 'GPL6246: [MoGene-1.0-st] Affymetrix Mouse Gene 1.0 ST Array [transcript (gene) version]'. Citations are provided for Wright JH, Johnson MM, Shimizu-Albergine M, Bauer RL et al. (2014) and Lee JJ, Wright JH, Johnson MM, Bauer RL et al. (2016). The reference series is GSE38199, with a sample count of 16 and a series published date of 2014/05/12. The value type is 'transformed count'. On the right, there is a 'Cluster Analysis' section with a heatmap and a 'Download' section with links to various file formats. At the bottom, the 'Data Analysis Tools' section is visible, including a 'Find genes' button and a search box for gene names or symbols.

Figure 1: Detailed information about the dataset which was used for GEO2R analysis.

My hypothesis is, if this is the case, I expect similar genes, and so pathways to be affected in human HCC by using DEGs of RNA-seq based results in human, which are more trustful than microarray (by using cBioPortal TCGA datasets [3]), and test those affected pathways based on DEGs on Panther/GO[4] . I will check human-mouse disease relationships at Mouse Genome Informatics (mouseDB)[2]. In addition, as a next step, it would be great to see similarity of related miRNAs (on miRnet tool[5]), because Feed Forward Loop (FFLs)-like trio structures of Transcription Factors (TFs), miRNAs and joint target genes might be used for therapeutic options (if available)(Supp Figure 1[7])[6].

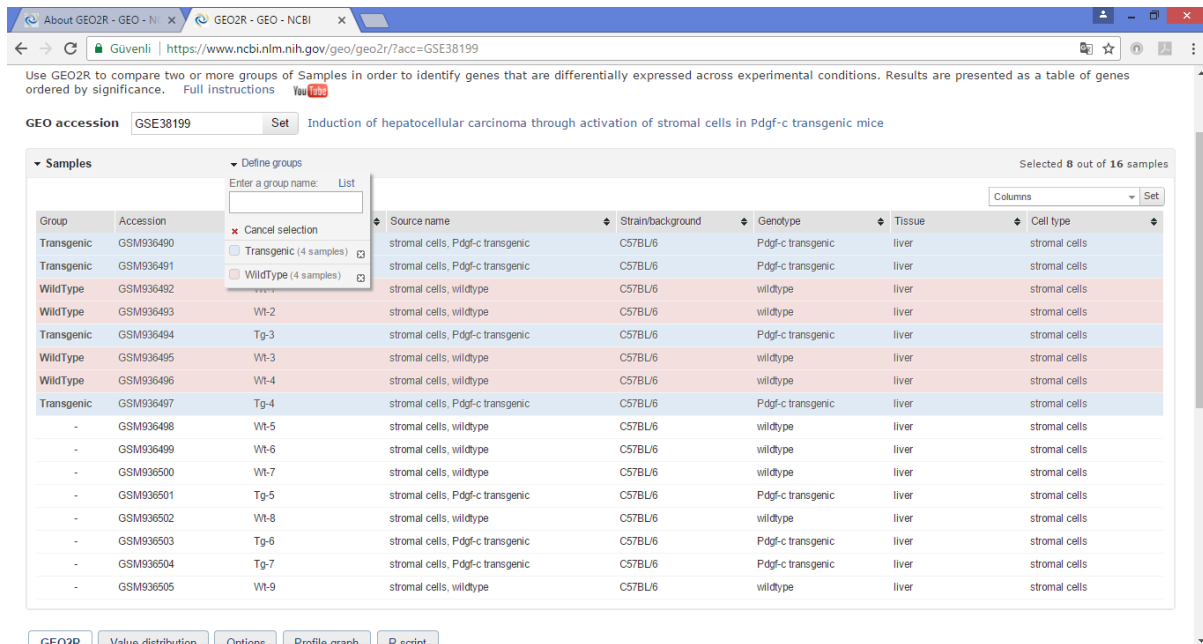


Figure 2: Groups for DEG analysis were defined based on wt and transgenic mice on GEO2R (4 for each).

DEGs Analysis was done on GEO2R with accession number GSE38199 (#reference series, **Figure 1**). 4 wild-type and 4 inductions of HCC by stromal cells of pdgf-transgenic mice were compared to figure out differentially expressed upregulated/downregulated genes (**Figure 2**). GEO2R uses limma or GEOquery packages from Bioconductor (R) for this purpose. It is available under the “R script” tab on GEO2R. By changing options, it enables researchers to set p-value options, log-transformation and annotation platform for genes. Because it was optimal for my study, default settings were used (**Figure 3**). In fact, under the GEO2R section, “select columns” options have changed, to see GO processes, GO processes was selected as one of the columns available under GEO2R menu.

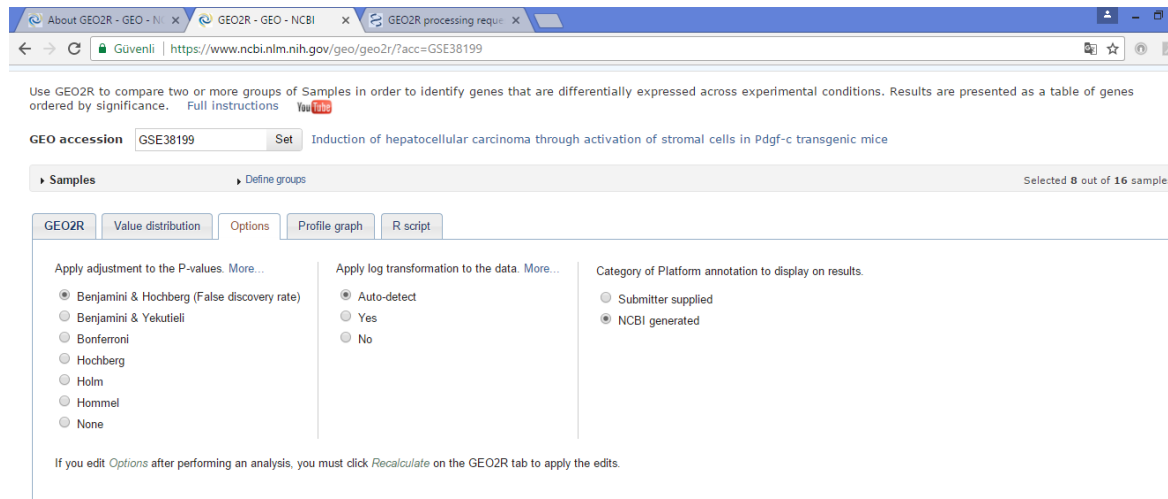


Figure3: GEO2R analysis tabs with default options.

GEO2R Top 250 Analysis gave us the adjusted P-value (based on False Discovery Rate/BH), p-value, t-statistic based on two different samples (in our case, wt vs. tg), log-odds (B), logFC (log-2-fold change between two groups), and t test for all pairwise comparisons for two groups (F), which was useful to test significance of your hypothesis result (shown in **Figure 4**). It is also available DEG analysis results for all the genes by clicking the “save all results” option. In addition, under the “Value Distribution” menu, it enables to see value distribution across different samples (selected wt and tg) (**Figure 5, left**). For my samples, it has almost no change (median-level) across wt and tg mice for

sample-wise distribution. However, it is possible to see the expression differences of specific gene across/in between (selected) samples (**Figure 5, right**) .

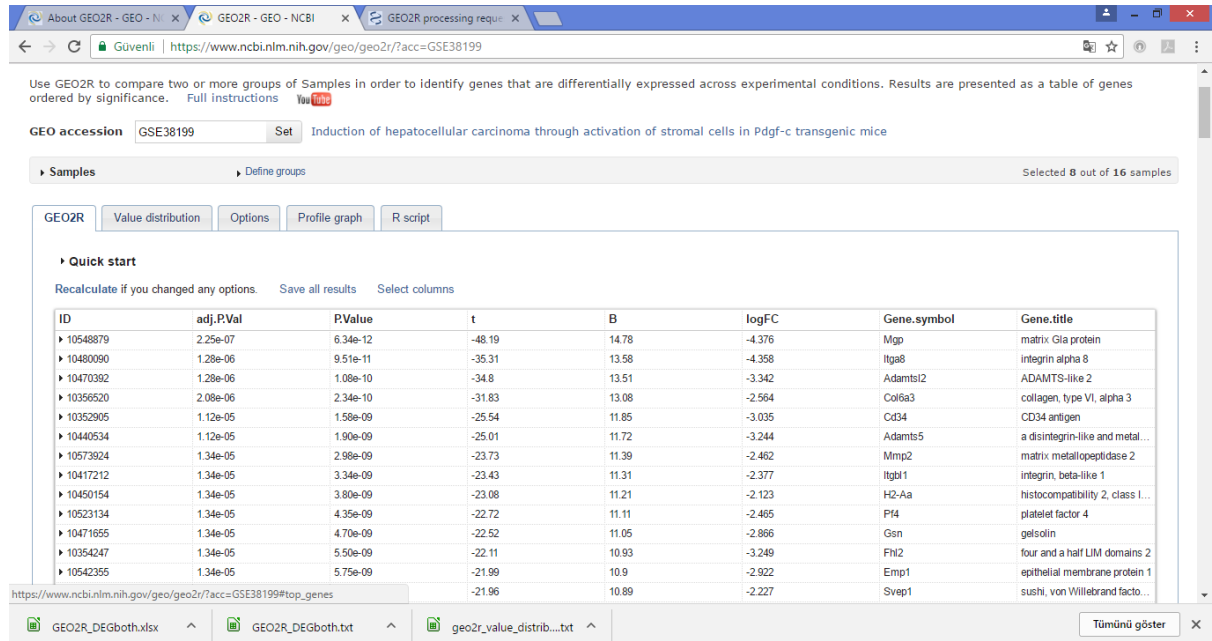


Figure 4 : GEO2R Top 250 Analysis result (based on limma package).

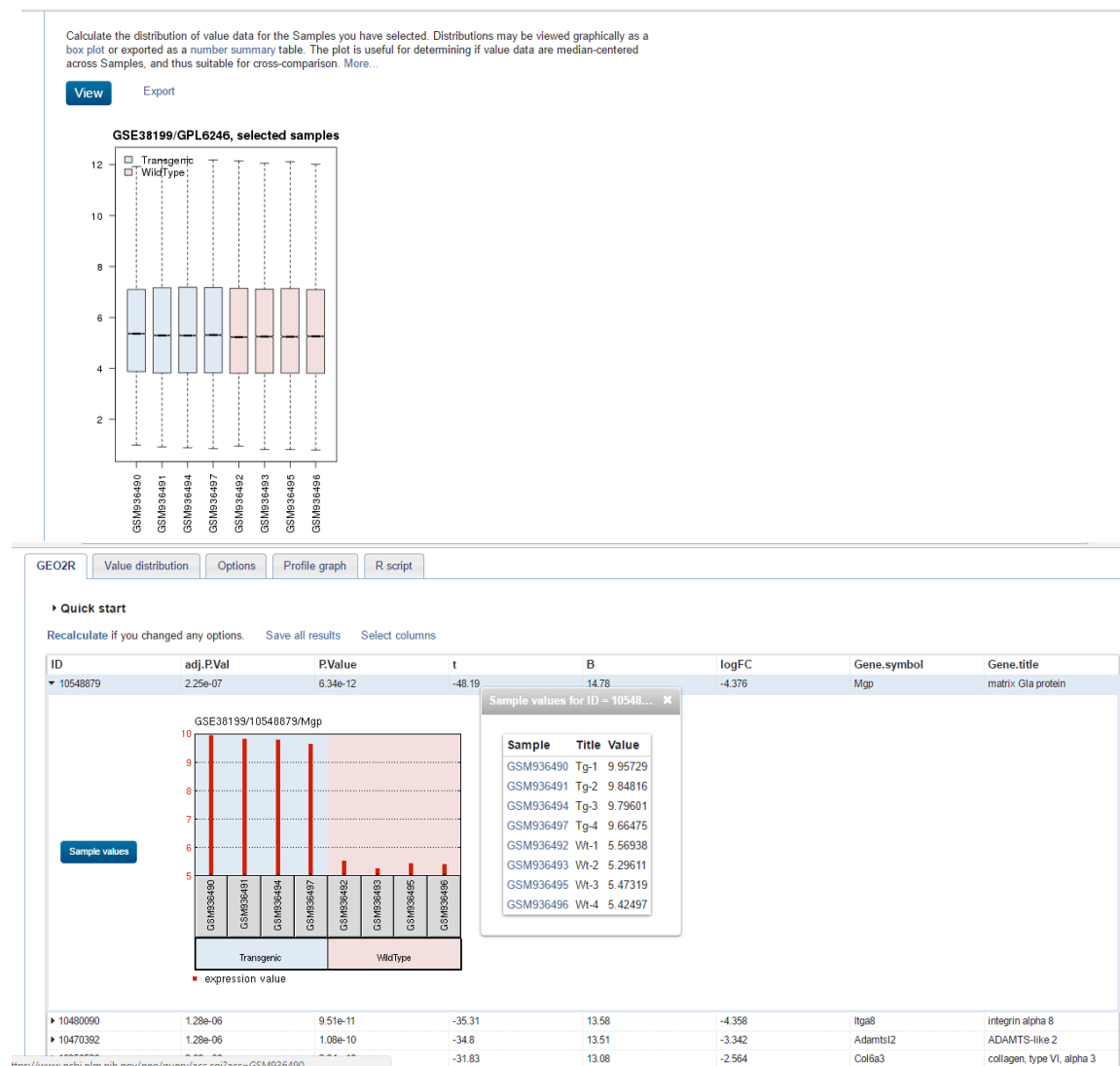


Figure 5: Value distribution across selected samples (left), and for specific genes (right).

In addition to GEO2R, GEO page also provides some statistical approaches and clustering options for the data available for that accession number (**Supplementary Figure 2,3,4**).

Based on log-2-fold change and adjusted p-values, top 200 and down 200 genes are selected and recorded in an excel file (available in final.zip). Then, these values are sorted, and genes without gene symbols and adjusted p-value > 0.05 for both down and up regulated genes are hidden on the excel files (up and down regulated gene files are available, separately in final.zip). Then remaining genes with gene symbols were uploaded to Panther/GO (**Supplementary Figure 5**).

2) a1) GO/Panther upload: See the pathways, processes related to DEGs for mice and human

a2) cBioPortal to compare expression profile human-liver disease relationship of selected genes (based on MGI genes filter applied for immune system & liver) with each other based on gene symbols

a3) Mouse Gene Informatics tool will help me to find interactions of disease between human and mice, and confirm the pathways once more

b) will be kept in the excel files and will be explained in the summary part

GO/Panther provided most related affected pathways of DEGs, which are shown in **Supplementary Figure 8**: mostly the immune system and integration mechanism of a cell were affected. Mouse Genome Informatics database will give the mouse-human disease connections based on the genes (DEGs from GEO2R) uploaded into the database, and most of them are related to immune system, homeostasis of the body, which might be related to liver at the end (**Figure 6** and **Supplementary Figure 6**). I have also checked the mouse strains used for specific research purposes in Mouse Tumor

Trm112 Igfals Agpat9 Ugt3a2///Ugt3a1 Creb3i3 Sic17a2 Akr1c14

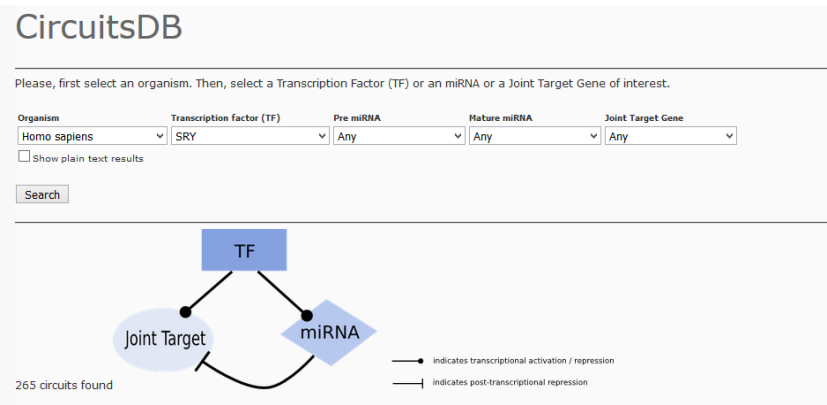
Gene Homologs x Phenotypes/Diseases (246 x 50)

Genes (683) Diseases (98)

Legend: Terms are annotated to genes in **human** **mouse** [show more](#)

Showing 15 of 246 rows; scroll down to see more.

Supplementary Figures



Supplementary Figure 1: General FFL structure [7].



Supplementary Figure 2: Options at GEO page for the selected datasets.

Gene ontology

Customize ...

Differential expression

Updown genes

DataSet keyword

Customize ...

GEO accession

Customize ...

Clear all

Show additional filters

☐

Gm2889 - Platelet-derived growth factor C transgenic model of hepatocellular carcinoma: liver stromal cells

Annotation: Gm2889, predicted gene 2889

Organism: Mus musculus

Reporter: GPL6246, 10344614 (ID_REF), GDS5320, AK145513, AK145782, chr1:3054233-3054733 (SPOT ID)

DataSet type: Expression profiling by array, transformed count, 16 samples

ID: 122024701

GEO DataSets Gene Profile neighbors

☐

Lyp1a1 - Platelet-derived growth factor C transgenic model of hepatocellular carcinoma: liver stromal cells

Annotation: Lyp1a1, lysophospholipase 1

Organism: Mus musculus

Reporter: GPL6246, 10344624 (ID_REF), GDS5320, NM_008966, XM_006495471, AK050549, AK167231, BC013536, BC0552848, CT010201, U89352, AK016021, AK034851, chr1:4807862-4846736 (SPOT ID)

DataSet type: Expression profiling by array, transformed count, 16 samples

ID: 122024706

GEO DataSets Gene UniGene Profile neighbors Chromosome neighbors Homologous neighbors

☐

Pcmtd1 - Platelet-derived growth factor C transgenic model of hepatocellular carcinoma: liver stromal cells

Annotation: Pcmtd1, protein-L-isoaspartate (D-aspartate) O-methyltransferase domain containing 1

Organism: Mus musculus

Reporter: GPL6246, 10344707 (ID_REF), GDS5320, NM_183028, XM_006495535, XR_373188, AK161921, BC052741, BC064801, BC110360, BC117941, BC117942, chr1:7089150-7173628 (SPOT ID)

DataSet type: Expression profiling by array, transformed count, 16 samples

ID: 122024714

GEO DataSets Gene UniGene Profile neighbors Chromosome neighbors Homologous neighbors

☐

Ahcy - Platelet-derived growth factor C transgenic model of hepatocellular carcinoma: liver stromal cells

Profile pathways

Find pathways

Find related data

Database: Select

Find items

Recent activity

Turn Off Clear

GDS5320(ACCN) (3557)

GEO Profiles

Platelet-derived growth factor C transgenic model of hepatocellular carcinoma GDS5320

(GDS5320(ACCN)) AND GDS[filter] (1)

GDSBrowser

expression liver tumor (4324)

GEO DataSets

Identification of Chemical Inhibitors of β -Catenin-Driven Liver Tumorigenesis in ...

See more...

Important Links

GEO Home

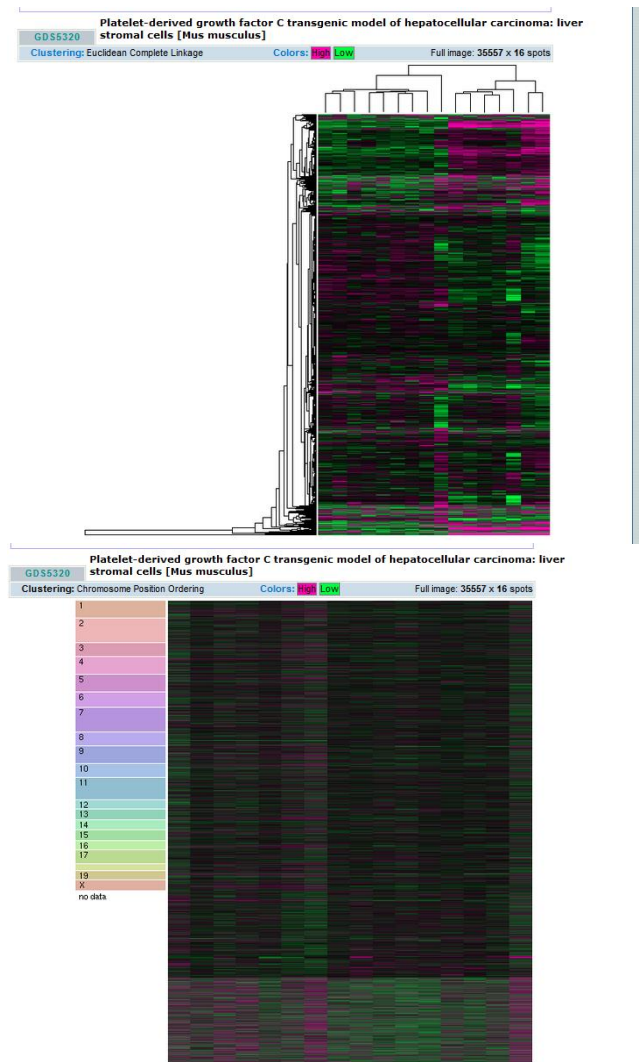
Profile GDS5320 / 10344614
Title Platelet-derived growth factor C transgenic model of hepatocellular carcinoma: liver stromal cells
Organism Mus musculus

GDS5320 / 10344614
 genotype/variation
 ■ transformed count
 ■ percentile rank within the sample

[Graph caption help](#)

Sample	Title	Value	Rank
GSM936490	Tg-1	5.93442	59
GSM936491	Tg-2	6.16761	63
GSM936494	Tg-3	5.89914	59
GSM936497	Tg-4	5.89726	59
GSM936501	Tg-5	6.10963	62
GSM936503	Tg-6	6.08324	62
GSM936504	Tg-7	5.87764	59

Supplementary Figure 3: Expression profile for selected genes with selected statistical approach (two-tailed t-test for this case) across samples at GEO.



Supplementary Figure 4: Hierarchical clustering heatmap (left) and heatmap based on chromosomal location of the genes (right). k-means clustering option is also available.

geneontology.org

Gene Ontology Consortium Home Documentation Downloads Community Tools About

Enrichment analysis

Vstm4
Ctsk
Ctss
Adamts9
Ccl19
D17H6S56E-5

biological process

Mus musculus

Submit

Help
Powered by PANTHER

Statistics

Search GO data

Search for terms and gene products...

Search

Ontology

Filter classes
Download ontology

Gene Ontology: the framework for the model of biology. The GO defines concepts/classes used to describe gene function, and relationships between these concepts. It classifies functions along three aspects:

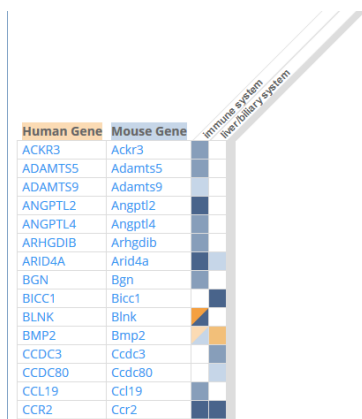
- molecular function**
molecular activities of gene products
- cellular component**
where gene products are active
- biological process**

Annotations

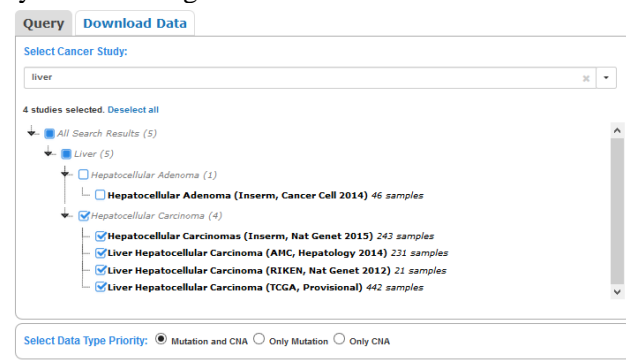
Download annotations (standard files)
Filter and download (customizable files <100k lines)

GO annotations: the model of biology. Annotations are statements describing the functions of specific genes, using concepts in the Gene Ontology. The simplest and most common annotation links one gene to one function, e.g. FZD4 + Wnt signaling pathway. Each statement is based on a specified piece of evidence. [more](#)

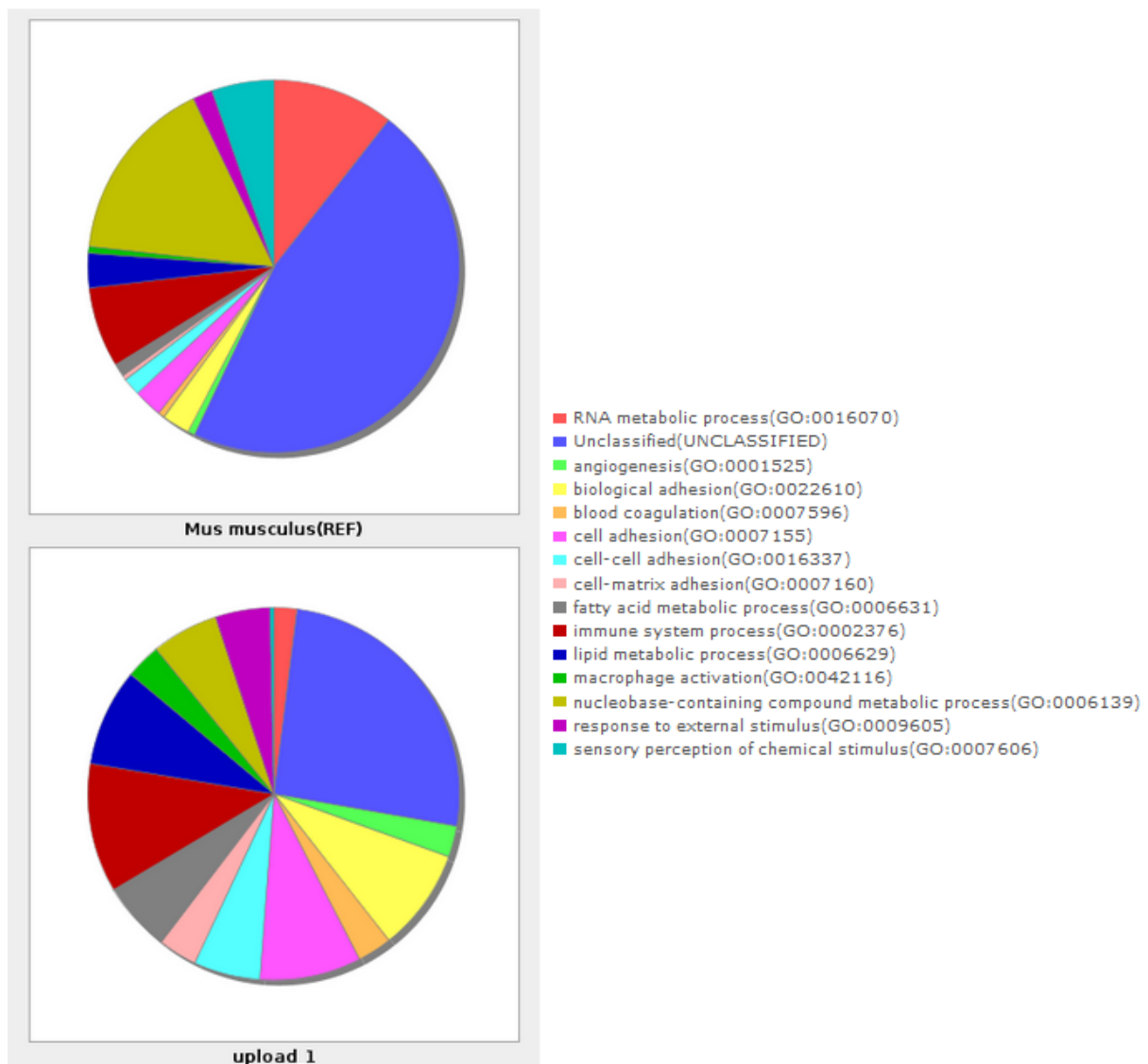
Supplementary figure 5: GO submission of top up and down regulated genes (<0.05 p-value)



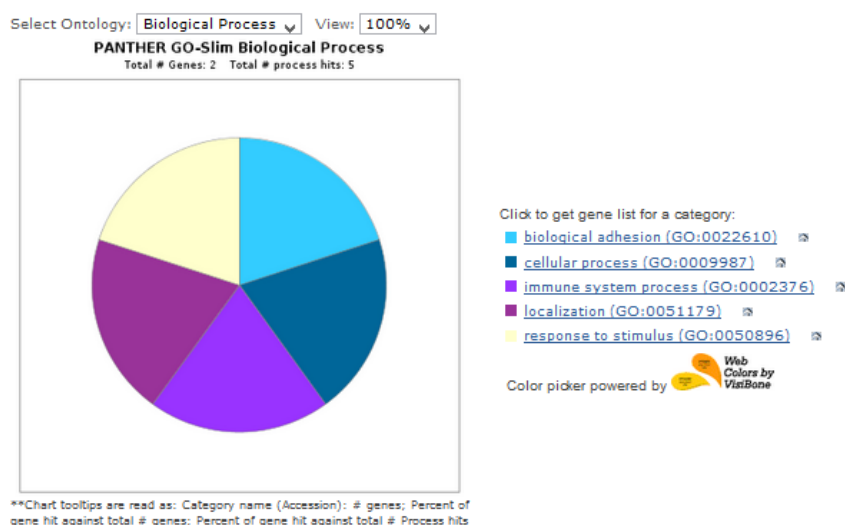
Supplementary Figure 6: MGI human-mouse disease connection results after applying liver-immune system related genes.



Supplementary Figure 7 : Datasets which were chosen from cBioPortal. p53 signaling related genes were checked.



Supplementary Figure 8 : Panther-Go-Slim Biological Processes Multiple Pie Chart based on DEGs.



Supplementary Figure 9: Significance for co-occurrence of CCDC3 and CLEC4G were the highest in cBioPortal. Figure shows the biological processes related to these two genes. Some of them did not show any hit on the Panther Biological Processes(data was not shown).

References:

- [1] Wright, J. H., Johnson, M. M., Shimizu-Albergine, M., Bauer, R. L., Hayes, B. J., Surapisitchat, J., ... Campbell, J. S. (2014). Paracrine activation of hepatic stellate cells in platelet-derived growth factor C transgenic mice; evidence for stromal induction of hepatocellular carcinoma. *International Journal of Cancer. Journal International Du Cancer*, 134(4), 778–788. <http://doi.org/10.1002/ijc.28421>
- [2] Blake JA, Eppig JT, Kadin JA, Richardson JE, Smith CL, Bult CJ, and the Mouse Genome Database Group. 2017. Mouse Genome Database (MGD)-2017: community knowledge resource for the laboratory mouse. *Nucl. Acids Res.* 2017 Jan. 4;45 (D1): D723-D729.
Bult CJ, Krupke DM, Begley DA, Richardson JE, Neuhauser SB, Sundberg JP, Eppig JT. Mouse Tumor Biology (MTB): a database of mouse models for human cancer. *Nucleic Acids Res.* 2015 Jan 28;43(Database issue):D818-24.
- [3] Cerami et al. The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discovery*. May 2012 2; 401.
Gao et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* 6, p11 (2013)
- [4] Paul D. Thomas, Michael J. Campbell, Anish Kejariwal, Huaiyu Mi, Brian Karlak, Robin Daverman, Karen Diemer, Anushya Muruganujan, Apurva Narechania. 2003. **PANTHER: a library of protein families and subfamilies indexed by function.** *Genome Res.*, 13: 2129-2141.
- [5] Fan, Y., Siklenka, K., Arora, S. K., Ribeiro, P., Kimmins, S., & Xia, J. (2016). miRNet - dissecting miRNA-target interactions and functional associations through network-based visual analysis. *Nucleic Acids Research*, 44(Web Server issue), W135–W141. <http://doi.org/10.1093/nar/gkw288>
- [6] Hong-Mei Zhang, Shuzhen Kuang, Xushen Xiong, Tianliuyun Gao, Chenglin Liu, An-Yuan Guo; Transcription factor and microRNA co-regulatory loops: important regulatory motifs in biological processes and diseases. *Brief Bioinform* 2015; 16 (1): 45-58. doi: 10.1093/bib/bbt085
- [7] Friard, O., Re, A., Taverna, D., De Bortoli, M., & Corá, D. (2010). CircuitsDB: a database of mixed microRNA/transcription factor feed-forward regulatory circuits in human and mouse. *BMC Bioinformatics*, 11, 435. <http://doi.org/10.1186/1471-2105-11-435>