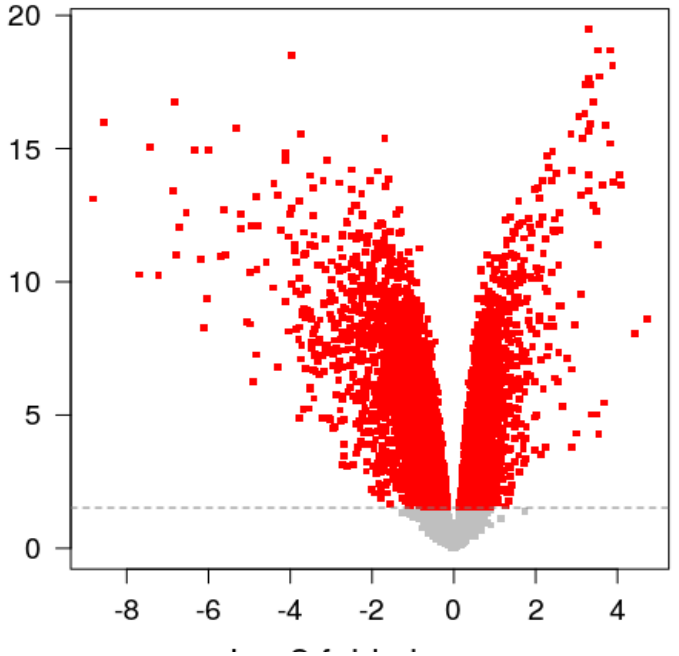
Analysis of Differential Gene Expression on Liver Hepatocellular Carcinoma



Mariona Torrens
Ariadna Marín
Alejandro Varela

Introduction

- Hepatocellular carcinoma is the most common primary type of liver cancer.
- It is also the second leading cause of cancer deaths worldwide.
- Risk factors:
  - cirrhosis
  - chronic hepatitis B or C.
  - diabetes and obesity
- Cure only possible in early stages, so medical treatment is usually palliative

- Need for biomarkers that detect it sooner → motivation of this project.

Materials and Methods

- **Data source:**

  -We used the Liver Hepatocellular Carcinoma set provided at The Cancer Genome Atlas (TCGA) Data Portal

- **Data manipulation and normalization:**

  -SummarizedExperiment package to manipulate gene expression data

  -edgeR package to normalize the raw reads

  -$\log_2$ scale to stabilize variability. Prior counts of 0.5.

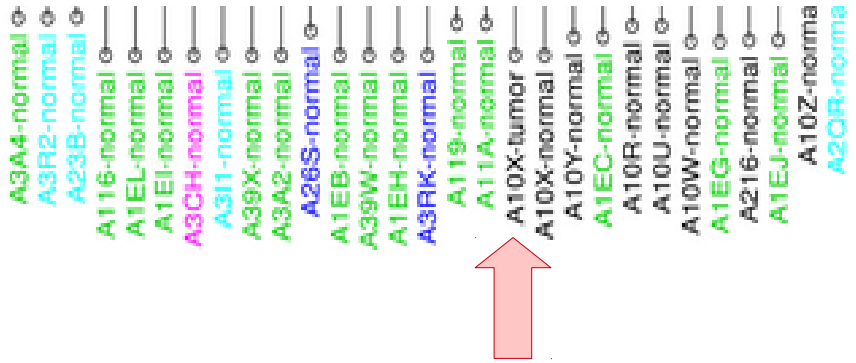
- **Data filtering:**

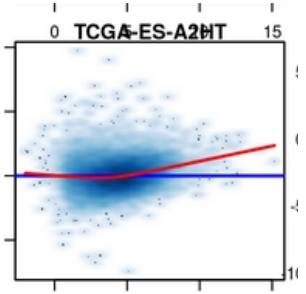
  Paired data and coverage greater than 40 million reads per sample.

  -Initial data set of 30 normal and 30 tumour.

  -One subject was eliminated due to intensity-dependent biases (detected with MA-plot)

  -Another subject was discarded because the tumour sample clustered with normal samples in hierarchical clustering





Materials and Methods

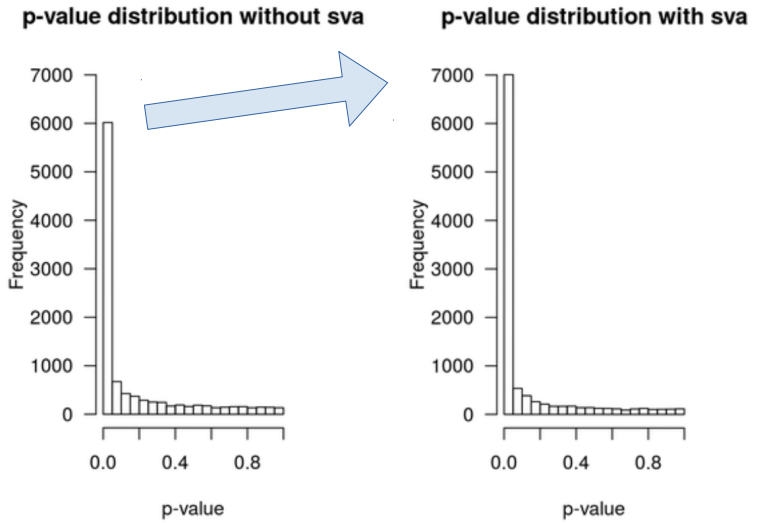
- **Batch analysis:**

  -Batch effect detection was carried out using multidimensional scaling (plotMDS function from edgeR package). We did not detect any batch effect.

- **Linear regression analysis:**

  -We looked for DE genes using a linear regression model which accounted for paired data design.

  -Surrogate variable analysis was employed to correct for unknown confounding factors, increasing statistical power.



  -Adjustment for mean-variance relationship was also carried out (voom function).

  -limma and sva packages were used.

Materials and Methods

- **Functional annotation and enrichment:**

  -Functional annotation of genes to GO terms → GOstats package.

  -Enriched GO terms were extracted and ranked according to p-value and Odds ratio.

  -Redundancies due to hierarchical GO's were filtered

  -Gene sets with less than 5 genes were filtered out.

- **Gene Set Enrichment Analysis:**

  -We used the GSVAdata and GSEABase packages

  -We studied the pathways from KEGG, REACTOME and BIOCARTA subcollections.

  -The analysis was conducted by z-score test and by Chi Squared Test (using Category)

  -Sets with significant adjusted p-values in at least one of the tests were considered as differently enriched pathways in HCC.

  -Gene sets with less than 10 genes or more than 200 were not considered.

Materials and Methods

- **Significance level:**

  -Differential expression of genes and sets was considered significant for p-values under 0.05, adjusted using FDR correction.

- **Data Availability:**

  -GitHub: alejandrovr/Hepatocarcinoma-RNA-seq-analysis

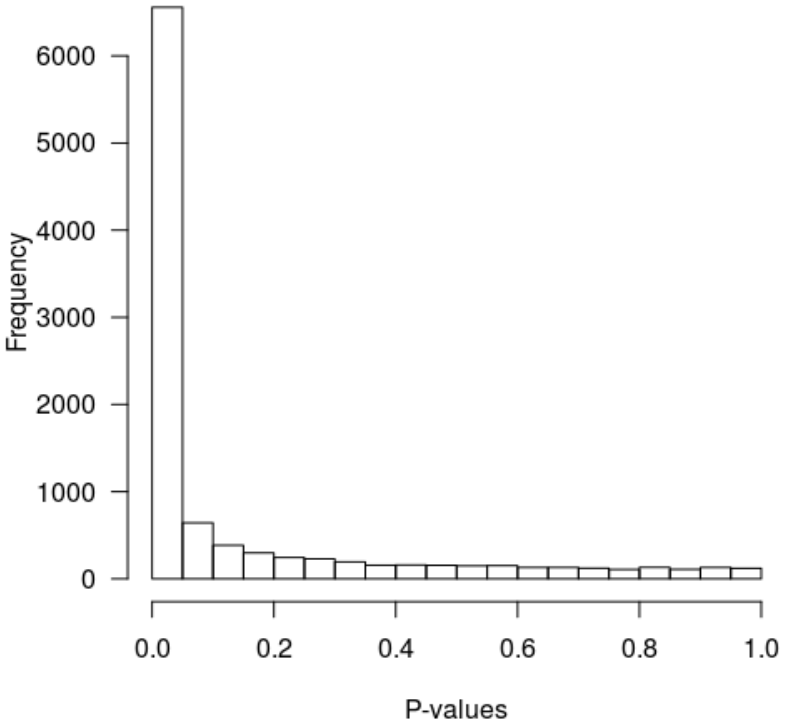
Results

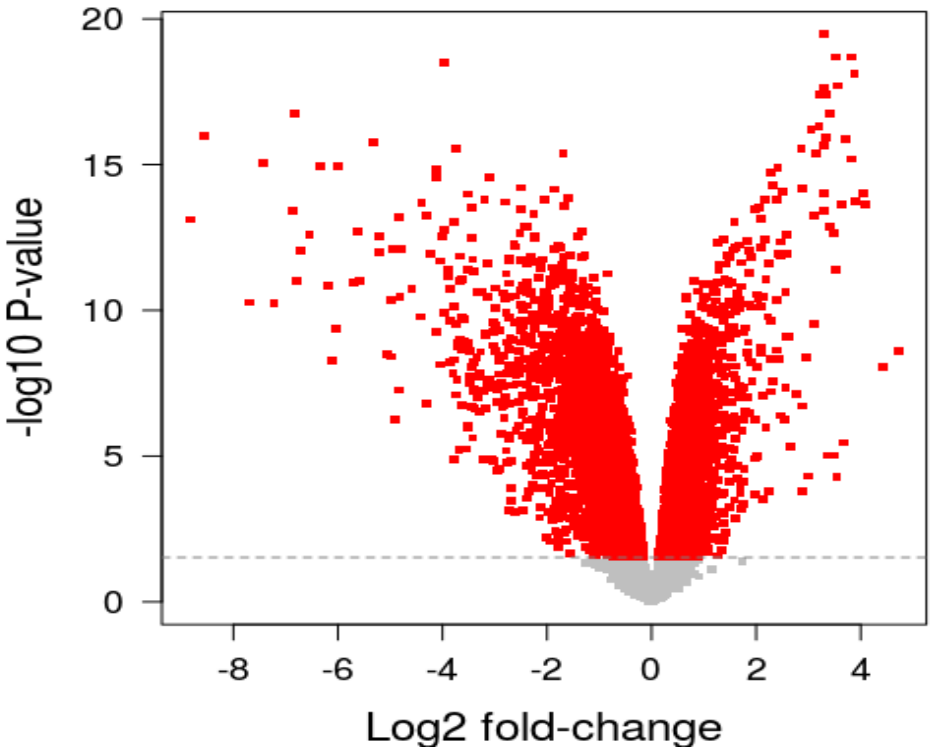
- **DE genes:**

-Average p-values corresponding to genes with increased transcription are smaller than those of genes with decreased expression.

-It is important to note that the genes with highest absolute FC are not the ones with smallest p-value

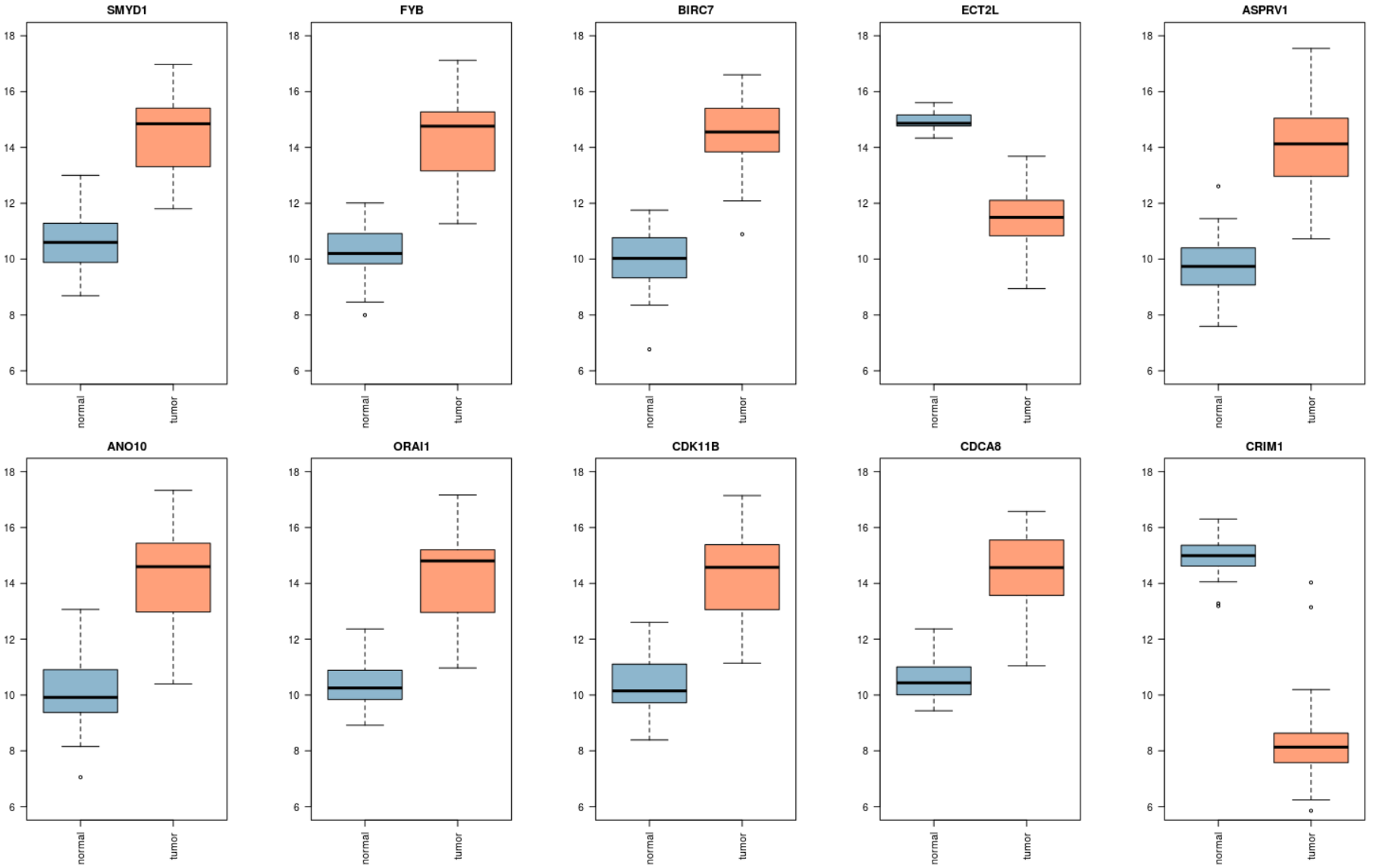
	FC < -2	FC > 2
# DE genes	335	85





Results

Top 10 DE genes

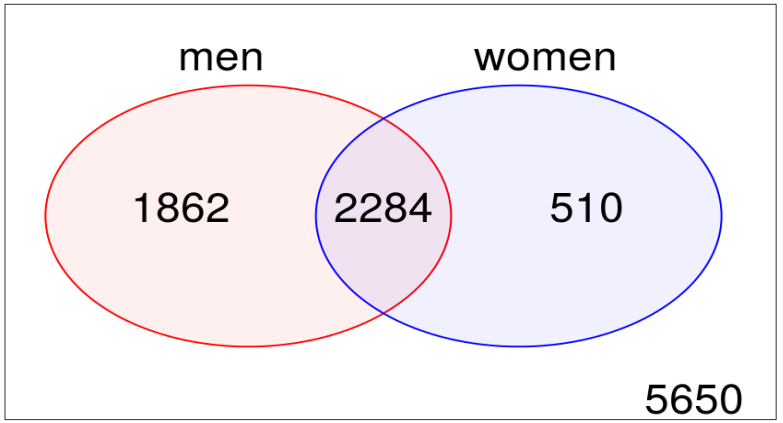


Results

- **Sexual dimorphism in DE genes in hepatocellular carcinoma:**

  -Only half of these genes are differentially expressed in both men and women.

  -Moreover, the number of male-specific genes is approximately three-fold that of female-specific genes



Results

- **GO terms:**

  -Most of them are related to normal liver function.

  -Some of them (not all included here) are cancer-related.

GO term	p-value	odds ratio	exp counts	counts	size	function
GO:0019373	0.00	Inf	7.22	12	12	epoxygenase P450 pathway
GO:0003094	0.01	Inf	6.02	10	10	glomerular filtration
GO:0009303	0.01	Inf	6.02	10	10	rRNA transcription
GO:0071941	0.01	Inf	6.02	10	10	nitrogen cycle metabolic process
GO:0006699	0.00	10.63	10.23	16	17	bile acid biosynthetic process
GO:0032369	0.01	8.63	8.42	13	14	negative regulation of lipid transport
GO:0042738	0.01	8.63	8.42	13	14	exogenous drug catabolic process
GO:1902622	0.01	8.63	8.42	13	14	regulation of neutrophil migration
GO:0003016	0.02	7.30	7.22	11	12	respiratory system process
GO:0001977	0.03	6.63	6.62	10	11	renal system process, regulation of blood volume
GO:0035640	0.03	6.63	6.62	10	11	exploration behavior
GO:0045909	0.03	6.63	6.62	10	11	positive regulation of vasodilation

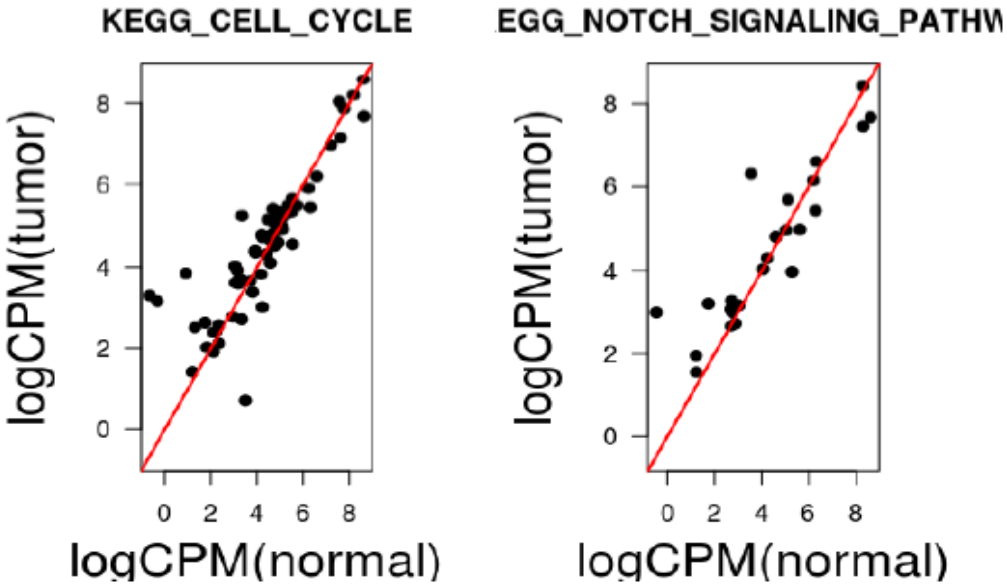
Results

- **Gene set analysis:**

  -These sets refer mostly to specific cell cycle transitions, mitosis and pathways involved in differentiation (such as Notch).

  -All ten sets correspond to biological functions that are up-regulated in tumour.

Gene set	DB	num.
Cell cycle	KEGG	128
Notch signaling pathway	KEGG	47
Pathogenic *E. coli* infection	KEGG	59
Hypertrophic cardiomyopathy	KEGG	85
Cell cycle mitotic	Reactome	306
Centrosome maturation	Reactome	72
G2-M checkpoints	Reactome	43
G2-M transition	Reactome	84
Loss NLM mit. centrosomes	Reactome	62
S phase	Reactome	103



Discussion

- Through examination of individual genes, we have found that they are involved in functions related to cell cycle control, migration and calcium signaling.

- Results from GSEA show similar functions enriched, with an emphasis in progression of the cell cycle, from signaling pathways to phase transitions.

- On the other hand, GO term enrichment of differentially expressed genes highlights alterations of the biological functions that are carried out by a healthy liver, such as lipid metabolism and bile acid production, and, therefore, are specific to HCC

  and other liver malignancies

- Our analysis has also shown differences in gene expression between male and female HCC samples.

- These same conclusions have been often reported in the literature.