Abstract
Hepatocellular carcinoma (HCC) is the third leading cause of cancer related death worldwide. Over the past years many efforts have been made to elucidate the molecular basis of the initiation and progression of (HCC). However, most of these studies have been focused on individual genes or single type of data which may lack the power to detect the complex mechanisms of cancer formation by overlooking the interaction networks. To gain insight into the molecular driver events, the combined effect of mRNA and miRNA expression was investigated in relation to cancer staging. We utilised the raw data of miRNA- and RNA- sequencing datasets from The Cancer Genome Atlas (TCGA). We performed filtering of the data with q-value normalization, scaling normalization and log2 transformation for the use of linear modeling. We then identified differentially expressed genes/miRNAs through three comparisons: Stage I/Normal, Stage II/ Normal and Stage III/ Normal. Genes and miRNAs with adjusted p-value < 0.05 and log2 fold change > 2 and 1 respectively, were considered significant. For each comparison type, we found the negative correlations between differentially expressed genes and miRNAs and the corresponding p-values. We combined Targetscan and microCosm databases in order to obtain a curated list of potential miRNA-miRNA interaction pairs. For each pair, the corresponding p-value was computed, adjusted for multiple testing using the BH method, using an adjusted p-value < 0.05 to identify and retain only the significant interactions. A pathway enrichment analysis, using KEGG annotation, revealed that up-regulated genes during the HCC progression are significantly correlated with the Cell cycle, Oocyte meiosis and p53 signaling pathways. Down-regulated genes are respectively correlated with Retinol metabolism and the Drug metabolism-cytochrome P450. Furthermore, we performed survival analyses with log rank test of patients of each stage in order to assess the relationship between differential gene/miRNA expression and overall survival. In the future, we will be interesting to correlate the differential miRNA/miRNA expression profiles with DNA methylation events, somatic mutations and CNVs of each stage in order to elucidate both genetic and epigenetic mechanisms underlying the development of HCC.

Aim of the study
There is great genetic and genomic diversity in HCC and in the need to gain insight into its molecular alterations, we performed
- Identification of differentially expressed genes and miRNAs profiles distinctive of disease staging
- Construction of miRNA-miRNA interaction networks in relation to disease staging
- Assessment of the clinical significance of newly identified expression profiles during HCC progression

Methodology
Data Workflow
RNAsq2, miRNAsq and clinical data of Liver Hepatocellular Carcinoma (LiCH) patients were retrieved from the TCGA data portal (http://cancergenome.nih.gov/). The full clinical dataset were downloaded (up to July 2015) and checked for the assessment of eligibility. The exclusion criteria were set to histologic diagnosis not being HCC. Overall, a total of 342 HCC patients were included in our study with the corresponding clinical data including age, gender, race, the American Joint Committee on Cancer staging system (also be called AJCC TNM staging system), tumor status, vital status, risk factors, vascular invasion, Child-Pugh classification and surgical types. Among the 367 patients (Cohort T), the adjacent non-tumor tissues were retrieved from 50 subjects (Cohort N). RNA-Seq version 2 was used in this meta-analysis generated by Illumina HighSeq. Normal samples were defined as pathologically non-cancerous tissue samples.

Results
Figure 1. Cluster analysis. Heatmaps of differentially expressed mRNAs (a, b, c) and miRNAs (d, e, f) of 3 comparison types: Stage I/Normal, Stage II/ Normal and Stage III/ Normal. Genes/miRNAs with adj p-value < 0.05 and logFC > 2/logFC > 1 were considered differentially expressed. Multiple testing correction was performed using Benjamini – Hochberg (BH) method. Hierarchical clustering analysis was performed using the ward.D2 linkage and Euclidean distance method. Each column represents a specimen and each row represents a gene. Green color indicates genes that are up-regulated and red color indicates genes that are down-regulated. Black indicates genes whose expression is unchanged in tumors compared to normal. Blue and orange sample labels indicate normal and tumor samples respectively.

Figure 2. Network analysis of 100 most significant predicted interactions between miRNAs and miRNAs in HCC development. For each comparison type: (a) Stage I/Normal, (b) Stage II/ Normal and (c) Stage III/ Normal negative correlations were computed between differentially expressed genes and miRNAs with corresponding p-values. Targetscan and microCosm databases were utilised in order to obtain a curated list of potential miRNA-miRNA interaction pairs. For each pair p-value was computed and adjusted for multiple testing using BH method, only the interactions with adjusted p-value < 0.05 were considered significant. Red-colored diamonds and blue-colored circles indicate miRNAs and miRNAs respectively. Node size represents node degree (number of edges connected to the node).

Figure 3. Pathway enrichment analysis. (a) Gene signature A and (b) B based on KEGG database. Gene signature A and B consists of genes with down- and up-regulation during HCC progression. Barplots represent the 5 more enriched KEGG terms of each signature with color coding: red indicates high enrichment, blue indicates low enrichment. The length of the bars represent the percentage of each row (KEGG category). The analysis was performed using the hypergeometric test and the adjustment for the estimated significance level to account for multiple hypothesis testing.

Figure 4. Kaplan-Meier survival analysis of patient samples. Patients were segmented in (a, b) Stage I, (c, d) Stage II and (e, f) Stage III were classified into high- or low-expression groups based on whether the expression of gene signature A (upper lane) and B (bottom lane) was greater than the median expression level. The P values from Logrank tests comparing two KM curves are shown in the bottom right side of each figure.

Conclusions:
(1) We established differential expression profiles of miRNAs/miRNAs distinctive of each HCC clinical stage (2) We identified significant ‘hubs’ in HCC disease progression (3) We identified prognostic expression signatures during HCC progression.