

**Integrated multi-omic analysis of HCC formation in the  
*SRF-VP16<sup>iHep</sup>* mouse model.**

**Dissertation**

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Abhishek Thavamani

aus

Dombivli, Indien

Tübingen

2016

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation: 23.12.2016

Dekan: Prof. Dr. Wolfgang Rosenstiel

1. Berichterstatter: Prof. Dr. Alfred Nordheim

2. Berichterstatter: Prof. Dr. Boris Macek

---

## COLLABORATORS

My special thanks to all collaborators who were involved in the project:

<b>Department</b>	<b>Institute/ Company</b>	<b>Experiment/ Contribution</b>	<b>People</b>
Genome Centre	Max Planck Institute for Developmental Biology, Tuebingen	RNAseq measurement	Dr. Lanz, Prof. Dr. Weigel
Proteome Center Tuebingen (PCT)	University of Tuebingen	LC-MS/MS analysis	Dr. Matic Prof. Dr. Macek
Quantitative Biology Center (QBiC)	University of Tuebingen	Bioinformatics support Data management	Dr. Nahnsen, Dr. Codrea, Dr. Czemmell Dr. Kenar
Department of Human Genetics	University Clinic, Tuebingen	Microarray analysis	Dr. Bonin
Centre for Bioinformatics, Tuebingen	University of Tuebingen	Clustering analysis	Dr. Römer Prof. Dr. Zell
Genomics Core Facility	DKFZ, Heidelberg	Whole exome sequencing (WES)	NA
Computational Oncology	DKFZ, Heidelberg	Bioinformatic analysis of WES data	Dr. Werner Dr. Schlesner
Pathology department	University Clinic Heidelberg	Pathology, Human HCC patient data	Prof. Dr. Schirmacher Prof. Dr. Longerich
			(to be continued...)

## COLLABORATORS

---

<b>Department</b>	<b>Institute/ Company</b>	<b>Experiment/ Contribution</b>	<b>People</b>
Epigenomics and cancer risk factors	DKFZ, Heidelberg	DNA methylation analysis	Dr. Daniel Lipka Dr. Christopher Plass
NA	Johannes Gutenberg University	Human HCC patient data	Prof. Dr. Marquardt
Animal Facility	Verfügungsgebäude, University of Tuebingen	Animal Care (mice)	Team of animal caretakers
NA	MetaSysX GmbH, Potsdam	Metabolomic analysis	NA
NA	LGC Genomics GmbH	Sanger sequencing	NA



## DECLARATION

---

The work described in the dissertation has been carried out under the supervision of Prof. Dr. Alfred Nordheim during the period of August 2012 - November 2016 at the Interfaculty Institute of Cell Biology, University of Tuebingen. Results from the following collaborative efforts are included in this thesis.

1. The *SRF-VP16<sup>iHep</sup>* mouse model was generated by Dr. Stefan Griesbach at Nordheim lab, Interfaculty Institute of Cell Biology, University of Tuebingen.
2. Microarray gene expression profiling was done by Dr. Michael Bonin at the University Hospital, University of Tuebingen. The data mining resulted in figures 3-11 to 3-13 and tables 3-4 to 3-6, 3-11.
3. Sanger sequencing of PCR-amplified fragments was performed at LGC Genomics GmbH. Figure 3-10 was generated based on these sequences.
4. DNA methylation analysis was performed by Dr. Daniel Lipka at DKFZ, Heidelberg. Figure 3-16a was generated by Dr. Daniel Lipka. Figure 3-16 was generated by including this methylation data.
5. RNAseq analysis was performed by Dr. Christa Lanz at the Genome Center, Max Planck Institute for Developmental Biology (MPI), Tuebingen.
6. Downstream bioinformatics analysis of the RNAseq data was assisted by Dr. Marius Codrea and Dr. Stefan Czernel, Quantitative Biology Center (QBiC), University of Tuebingen. The figure 3-18 was generated by QBiC. Subsequent data mining was performed by myself based on the generated RNAseq data (Figures 3-27 to 3-31, 3-33 to 3-35, 3-42; Tables 3-14A, 3-15 to 3-18).
7. Proteomics and phosphoproteomics measurements were performed by Dr. Katarina Matic at Proteome Center Tübingen (PCT), University of Tuebingen. Figures 3-26 and 3-32 were generated by Dr. Katarina Matic. Subsequent data mining was performed by myself based on the generated proteomics and phosphoproteomics data (Figures 3-27 to 3-31, 3-33 to 3-35, 3-42; Table 3-10).

## DECLARATION

---

8. Whole exome sequencing (WES) was performed at the Genomics Core Facility, DKFZ, Heidelberg.
  9. The bioinformatics analysis of the WES reads was processed by Dr. Johannes Werner from Dr. Matthias Schlesner's group at DKFZ, Heidelberg. The mining of the generated data resulted in the figures 3-36 to 3-41.
  10. The dataset of the Heidelberg human HCC cohort was generated by Dr. Thomas Longerich at DKFZ, Heidelberg. Figure 3-14 was generated based on this data. Clustering analysis performed using this data resulted in figure 3-15 and table 3-9.
  11. The dataset of the Mainz human HCC cohort was generated by Dr. Jens Marquardt at Department of Medicine I, Johannes Gutenberg University, Mainz. Clustering analysis using this data is shown in figure 3-24.
  12. Clustering analysis was performed using an algorithm developed by Dr. Michael Römer from the group of Prof. Dr. Andreas Zell, Department for Cognitive Systeme, University of Tuebingen.
  13. The publicly accessible Cancer Genome Atlas (TCGA) was used to obtain human HCC expression data. Clustering analysis using this data is shown in figure 3-25.
  14. The metabolomics measurements were performed by MetaSysX GmbH, Potsdam. The bioinformatics analysis of the metabolomics data was done by Dr. Erhan Kenar from QBiC, University of Tuebingen (Table 3-20).
-

## ACKNOWLEDGEMENT

---

I thank my supervisor Prof. Dr. Alfred Nordheim for his constant support. Thank you sir for everything you have taught me.

I thank the IMPRS program for giving me this opportunity, and my special thanks to the coordinator Dr. Sarah Danes, for organizing everything very well, right from the interviews.

I thank my thesis advisory committee members Prof. Dr. Detlef Weigel and Prof. Dr. Boris Macek, for the great discussions and suggestions that helped shape the project in the right ways.

I thank DKFZ and DKTK for funding the projects and financing the experiments. I thank Dr. Jürgen Frank for his guidance through the DKTK official processes and for assistance with the orderings.

I specially thank all the past and present collaborators for the valuable contribution towards the project. It was a pleasure working with you all.

I thank Prof. Dr. Tassula Proikas-Cezanne and her group for the wonderful journal club and research report discussions and feedbacks.

I thank Dr. Stefan Griesbach, former post doc at Nordheim lab, for his guidance and orientation during the beginning of my project. I also thank him for having developed the *SRF-VP16<sup>iHep</sup>* mouse model, which my project is entirely based upon.

My special thanks to Heidemarie Riehle, Heiderose Kraus and Elena Kullmann, for the wonderful support all along. Vielen vielen Dank... Ihr seid die Besten...

I also thank Dr. Siegfried Alberti for his constant guidance and help at times of need. I thank Katja Kleemann and Anke Biedermann for their timely assistance in times of need.

## ACKNOWLEDGEMENT

---

I would like to thank former post doc members of the Nordheim lab - Dr. Christine Weinl, Dr. Melissa Cox and Dr. Christine Stritt, for always being helpful and approachable.

Special thanks to former doctoral colleagues – Dr. Dong Jeong Park and Dr. Daniela Bakula, for their special friendship. I also thank my current doctoral colleagues, especially Ivana Akrap for being a constant motivation and also for the scientific discussions.

I thank the former lunch gang – Mario Graeve, Kshitij Mohan, Tobias, Max, Natalie, Katrin and Michael for the fun we all had together. I also thank Marco Seehawer for his contributions to the project during his internship time at the Nordheim lab.

I thank Georg Tiedemann for his help in solving technical snags and also for promptly distributing the posts.

I am specially grateful to the animal care-takers who took proper care of the mice, which is very important in order to have unbiased results.

I cannot thank my family enough, my parents and sisters and everyone, for being pillars of support in all situations.

I thank my lovely wife, who has been the perfect companion, patient and supportive, and my son for bringing joy and happiness.

# Table of Contents

COLLABORATORS.....	
DECLARATION.....	
ACKNOWLEDGEMENT.....	
TABLE OF CONTENTS.....	
ZUSAMMENFASSUNG .....	<b>1</b>
ABSTRACT.....	<b>2</b>
<b>1 INTRODUCTION.....</b>	<b>3</b>
<b>1.1 Cancer .....</b>	<b>3</b>
1.1.1 Definition .....	3
1.1.2 Mortality and morbidity.....	3
1.1.3 Recent advancements in cancer research – ‘omics’ approaches.....	4
1.1.4 Liver cancer.....	10
1.1.5 Types of liver cancer.....	12
<b>1.2 Hepatocellular carcinoma (HCC) .....</b>	<b>13</b>
1.2.1 Etiological factors .....	13
1.2.2 Histopathology of HCCs .....	19
1.2.3 HCC types.....	19
1.2.4 HCC treatment modalities .....	21
1.2.5 Molecular mechanisms of hepato-carcinogenesis .....	23
1.2.6 Mouse models for HCC.....	31
<b>1.3 Serum Response Factor .....</b>	<b>37</b>
<b>1.4 Objective of this study .....</b>	<b>47</b>
<b>2 MATERIALS AND METHODS.....</b>	<b>50</b>
<b>2.1 Animal Experiments .....</b>	<b>50</b>
2.1.1 Maintenance of experimental mice .....	50
2.1.2 Mouse breeding .....	51
2.1.3 Mouse ear marking.....	52

## TABLE OF CONTENTS

---

2.1.4	Genotyping.....	53
2.1.5	Mouse health monitoring.....	62
2.1.6	Mouse sacrifice .....	62
<b>2.2</b>	<b>DNA analysis .....</b>	<b>63</b>
2.2.1	DNA extraction .....	63
2.2.2	Mutation detection .....	63
<b>2.3</b>	<b>RNA analysis.....</b>	<b>66</b>
2.3.1	RNA extraction.....	66
2.3.2	Reverse transcription.....	66
2.3.3	Quantitative PCR for RNA expression analysis.....	67
2.3.4	Mutation detection .....	69
2.3.5	Microarray analysis.....	71
<b>2.4</b>	<b>Protein analysis .....</b>	<b>72</b>
2.4.1	Protein extraction .....	72
<b>2.5</b>	<b>Combined omics analysis .....</b>	<b>73</b>
2.5.1	Whole exome sequencing.....	74
2.5.2	Transcriptomics and proteomics.....	76
2.5.3	RNAseq analysis .....	78
2.5.4	LC-MS/MS analysis .....	79
<b>2.6</b>	<b>Metabolomics analysis .....</b>	<b>80</b>
2.6.1	Samples.....	80
<b>2.7</b>	<b>Comparison to human HCCs .....</b>	<b>81</b>
2.7.1	Heidelberg cohort.....	81
<b>2.8</b>	<b>Detailed procedures .....</b>	<b>85</b>
2.8.1	RNA extraction.....	85
2.8.2	DNA extraction .....	86
2.8.3	Protein extraction .....	88
<b>2.8.4</b>	<b>RNAseq library preparation .....</b>	<b>89</b>
<b>2.8.5</b>	<b>MS sample preparation .....</b>	<b>101</b>
<b>2.8.6</b>	<b>Circular DNA synthesis.....</b>	<b>105</b>
2.8.7	PCR product purification.....	106
2.8.8	Agarose gel electrophoresis.....	107
<b>2.9</b>	<b>Softwares / online tools used .....</b>	<b>109</b>

---

<b>2.10</b>	<b>Primer sequences .....</b>	<b>110</b>
<b>3</b>	<b>RESULTS .....</b>	<b>113</b>
<b>3.1</b>	<b>The <i>SRF-VP16<sup>iHep</sup></i> mouse model .....</b>	<b>113</b>
3.1.1	<i>Genotype.....</i>	113
3.1.2	<i>Phenotypic observations .....</i>	117
3.1.3	<i>Liver weight to body weight ratio.....</i>	118
3.1.3	<i>Hepatocellular carcinomas in liver .....</i>	119
<b>3.2</b>	<b>Mutation detection .....</b>	<b>121</b>
3.2.1	<i>Frequency of mutations.....</i>	121
3.2.2	<i>Cttnb1 mutation associated gene expression changes.....</i>	124
<b>3.3</b>	<b>Gene expression profiling - microarray analysis .....</b>	<b>125</b>
3.3.1	<i>Sub-lists – microarray data .....</i>	126
<b>3.4</b>	<b>Comparison to human HCCs .....</b>	<b>131</b>
3.4.1	<i>Heidelberg cohort.....</i>	131
3.4.2	<i>Oncofetal liver genes.....</i>	134
	<b>Combined-omics approach .....</b>	<b>137</b>
<b>3.5</b>	<b>Transcriptome and proteome analysis .....</b>	<b>138</b>
<b>3.5.1</b>	<b>Transcriptomics - RNAseq .....</b>	<b>139</b>
3.5.1.1	<i>SRF-VP16 expression.....</i>	139
3.5.2	<i>Principal Component Analysis (PCA).....</i>	140
3.5.3	<i>Transcriptomic profile - RNAseq .....</i>	141
3.5.4	<i>Comparison of RNAseq data to Microarray data.....</i>	143
3.5.5	<i>Comparison to human HCCs – RNAseq data.....</i>	146
<b>3.6</b>	<b>Proteomics analysis .....</b>	<b>148</b>
3.6.1	<i>Principal Component Analysis.....</i>	150
3.6.2	<i>Altered protein levels.....</i>	150
3.6.3	<i>Comparison to RNAseq.....</i>	152
<b>3.7</b>	<b>Phospho-proteomics analysis .....</b>	<b>154</b>
3.7.1	<i>Principal Component Analysis.....</i>	154
3.7.2	<i>Differentially phosphorylated peptides .....</i>	155
<b>3.8</b>	<b>Whole exome sequencing.....</b>	<b>157</b>
3.8.1	<i>Single Nucleotide Variations .....</i>	158

## TABLE OF CONTENTS

---

3.8.1	<i>Insertions/ Deletions</i> .....	160
3.8.3	<i>Copy number alterations</i> .....	162
3.8.4	<i>Tumor heterogeneity</i> .....	163
<b>3.9</b>	<b>Molecular mechanisms involved in HCC formation</b> .....	<b>165</b>
3.9.1	<i>Wnt signaling pathway</i> .....	165
3.9.2	<i>SRF-target gene regulation</i> .....	166
3.9.3	<i>Cell cycle regulation</i> .....	167
3.9.4	<i>Metabolic reprogramming</i> .....	168
<b>3.10</b>	<b>Metabolomics analysis</b> .....	<b>171</b>
<b>4</b>	<b>DISCUSSION</b> .....	<b>174</b>
4.1	<b>Insights obtained from the present study</b> .....	<b>174</b>
4.2	<b>Future Perspectives</b> .....	<b>182</b>
<b>5</b>	<b>REFERENCES</b> .....	<b>183</b>
<b>6</b>	<b>ABBREVIATIONS DIRECTORY</b> .....	<b>196</b>
<b>7</b>	<b>TABLES DIRECTORY</b> .....	<b>202</b>
<b>8</b>	<b>FIGURES DIRECTORY</b> .....	<b>204</b>
<b>9</b>	<b>RELEVANT TRAINING</b> .....	<b>208</b>
<b>10</b>	<b>POSTER PRESENTATIONS</b> .....	<b>209</b>
<b>11</b>	<b>CURRICULUM VITAE</b> .....	<b>211</b>
<b>12</b>	<b>CONTRIBUTION TO DISSERTATION</b> .....	<b>212</b>



---

# Zusammenfassung

---

## Integrierte multi-Omics Analyse der Leberkarzinogenese im *SRF-VP16<sup>iHep</sup>* Mausmodell

*SRF-VP16<sup>iHep</sup>* Mäuse zeigen Aktivität des SRF-VP16 Proteins, eines konstitutiven Transkriptionsfaktors und Aktivators von CAAT-box-regulierten Zielgenen, im hepatozellulären Mosaikmuster. Die dadurch ausgelöste Hyperproliferation von Hepatozyten führt zur Ausbildung von prä-malignen Knoten („nodules“), aus denen sich Hepatozelluläre Karzinome (murine HCCs, mHCCs) entwickeln (Ohrnberger\* and Thavamani\* et al., 2015). Die vorliegende Arbeit konzentriert sich auf die Charakterisierung der HCC Tumorgewebe, sowie die bei deren Entstehung durchlaufenen Progressionsstadien, indem verschiedene genomweite Omics-Analyseplattformen zum Einsatz gebracht werden. Dabei war es Ziel dieser Arbeit, essentielle molekulare und zelluläre Ereignisse bei der HCC Tumorentstehung zu identifizieren.

Die eingesetzten Omics-Analyseplattformen waren Transkriptomik (RNAseq), Proteomik/Phospho-Proteomik, und Exom-Sequenzierung. Genomische Sanger-Sequenzierung von Kandidaten-Loci zeigte die Anwesenheit von *Cttnb1* und *Trp53* Mutationen in zwei verschiedenen Untergruppen der mHCCs. Die Transkriptom- und Proteomprofile von *Cttnb1*-Wildtyp und *Cttnb1*-mutierten mHCCs zeigten signifikante Unterschiede. Vergleichende Clusteranalyse der Transkriptionsprofile der *SRF-VP16<sup>iHep</sup>*-Proben mit humanen HCC Tumorproben zeigte signifikante Überlappung der *SRF-VP16<sup>iHep</sup>* HCCs mit der G1/G2 Untergruppe humaner HCCs. Dabei wurden gemeinsame Gensignaturen identifiziert, z. B. bezüglich der erhöhten Expression von „oncofetal liver genes“. Eine vollständige Exom-Sequenzierung (WES) von prä-malignen Knoten und mHCCs der *SRF-VP16<sup>iHep</sup>* Mäuse zeigte die inhärente Heterogenität der mHCCs. Die durchgeführte multi-Omics Analyse gibt Hinweise auf erfolgte metabolische Reprogrammierung in mHCC Geweben, speziell auf die Aktivierung der aeroben Glycolyse und des mitochondrialen  $\alpha$ -Metabolismus.

## Abstract

---

### **Integrated multi-omic analysis of HCC formation in the *SRF-VP16<sup>iHep</sup>* mouse model.**

*SRF-VP16<sup>iHep</sup>* mice activate the constitutive transcription factor SRF-VP16 in a hepatocyte-specific mosaic fashion and consequently, develop hepatocellular carcinoma (mHCC) (Ohrnberger\* and Thavamani\*, et al., 2015). Using the *SRF-VP16<sup>iHep</sup>* mouse model, this thesis focuses on the characterization of mHCC formation from pre-malignant liver nodules. The objective of the current study was to identify carcinogenic driver events responsible for malignant transformation and tumor formation in *SRF-VP16<sup>iHep</sup>* livers. To this end, an integrated multi-omics analysis approach was undertaken, which involved the investigation of the genome (exome), transcriptome and proteome at the premalignant and advanced stages of HCC development. Sanger sequencing of genomic candidate loci revealed the presence of *Cttnb1* and *Trp53* mutations in two different subsets of mHCCs, respectively. Transcriptome and proteome profiles of *Cttnb1* mutant and *Cttnb1* wildtype mHCCs showed significant differences between the two subsets. *SRF-VP16<sup>iHep</sup>* tumor transcriptomes showed striking similarities to that of a distinct subset of human HCCs. This identified *SRF-VP16<sup>iHep</sup>* to resemble the G1/G2 subtype of human HCCs, as was indicated by common gene expression signatures, for example activated oncofetal genes. Whole exome sequencing (WES) analysis of mHCCs and pre-malignant nodular tissues revealed an inherent heterogeneity of each mHCC, even when originating within one individual liver. In addition, integrated omics analysis of mHCCs from *SRF-VP16<sup>iHep</sup>* mice indicated the occurrence of metabolic reprogramming, for example the activation of aerobic glycolysis and of the mitochondrial 1C-metabolism.

# 1 INTRODUCTION

## 1.1 Cancer

### 1.1.1 Definition

Malignant neoplasm, also known as cancer, is a collective term used to denote hyper-proliferative diseases, in which abnormal cells divide uncontrollably and can invade neighboring tissues and thereby spread to other parts of the body through the blood and lymph systems.

### 1.1.2 Mortality and morbidity

Cancer is the second highest cause of death worldwide, accounting for more than 8 million deaths in the year 2012 (Figure 1-1). That is, one among every seven persons in the world dies of cancer-related causes.

	Worldwide		
	Rank	Deaths	%
Cardiovascular diseases	1	17,513	31%
Malignant neoplasms	2	8,204	15%
Infectious and parasitic diseases	3	6,431	12%
Respiratory diseases	4	4,040	7%
Unintentional injuries	5	3,716	7%

*Figure 1-1: The top 5 causes of death worldwide in 2012. The numbers of deaths represented are in thousands. Source: GLOBOCAN 2012 (Ferlay et al., 2015).*

The International Agency on Cancer (IARC) estimates that in 2012 there were 14.1 million new cancer cases diagnosed worldwide. The comparable estimate for the year 2008 had been 12.7 million new cases. The yearly incidence of cancer in the year 2030 is expected to rise to 21.7 million and the annual deaths to 13 million. However the World Health Organization cautions that actual numbers in 2030 might be drastically larger, owing to the people around the world increasingly adopting lifestyle habits that can significantly increase the risk of cancer occurrence (Ferlay et al., 2015).

### **1.1.3 Recent advancements in cancer research – ‘omics’ approaches**

#### **1.1.3.1 Human genome project**

The invention of DNA sequencing technology by Frederick Sanger spurred numerous developments in the field of genome sequencing. Further advancements in this area led to the complete automation of the process. This achievement encouraged scientists to undertake the sequencing of the complete human genome. Eventually, in the year 1988 there was a Memorandum of Understanding (MoU) signed between the National Institute of Health (NIH) and the United States Department of Energy (DOE) to work together towards the common goal of unraveling the nucleotide sequence composition of the entire human genome. This joint initiative, which started in 1990, was later known as the Human Genome Project.

In June 2000, it was announced that the entire human genome had been sequenced and the first publication regarding the sequenced human genome was published in the journal *Nature*, in February 2001. Some of the techniques used in the process were DNA sequencing, Restriction Fragment-Length Polymorphism (RFLP) analysis, use of Yeast Artificial Chromosomes (YAC) or of Bacterial Artificial Chromosomes (BAC), Polymerase Chain Reaction (PCR), and electrophoresis (Lander et al., 2001).

The unlocking of the human genome sequence revolutionized cancer research, effectively giving rise to the ‘omics’ era. This coincided with the technological developments enabling the high-throughput analysis of genes, transcripts, proteins, and metabolites therefore resulting in the various omics fields namely, genomics, transcriptomics, epigenomics, proteomics, and metabolomics (Rochfort, 2005). Omics-based cancer research has resulted in a vast expansion of knowledge regarding the genomic basis of cancer and also helped in the application of this knowledge for better treatment.

### 1.1.3.2 Genomics

High-throughput determination of the cancer genome sequences has resulted in the discovery of numerous oncogenes and tumor suppressor genes that play major roles in the progression of different tumors. Mutations could be detected in tumor suppressor genes thereby impairing their vital functions such as cell cycle regulation (*TP53*, *CDK1*, *CDK2*) or DNA mismatch repair (*MMR*), and thus assisting the tumor in its incessant proliferation (Weinert and Lydall, 1993) (Hsieh and Yamane, 2008). Alternatively there were also mutations found in various oncogenes, resulting in a gain of function in these genes. These included growth factor receptors (*EGFR*, *VEGFR*) and other genes that promote cellular proliferation and invasion (*MYC*, *RAS*) (Hanahan and Weinberg, 2000). These studies have also resulted in the development of treatment modalities targeting these genes in order to inhibit cancer growth in a specific manner (Kandula et al., 2013; Nan et al., 2017). The past few decades have seen significant advances in the field of Next Generation Sequencing (NGS) which has enabled the sequencing of tumor samples at a high resolution that could be compared against the human genome using various tools (Gu et al., 2013; Li and Durbin, 2009). Comparison of the consensus sequences of reads which together help detecting genomic variations including the following.

- ❖ Genomic point mutations
- ❖ Genomic insertions and deletions
- ❖ Genomic copy number variation
- ❖ Genomic loss of heterozygosity (LOH)
- ❖ Gene translocation or gene fusion
- ❖ Genome and chromatin modifications, including CpG methylation and histone modification.

This information can give significant insight about the genomic profiles of the cancer cells and thereby help us to understand the molecular mechanisms involved in carcinogenesis. The NGS technology could be used in different approaches, namely

whole genome sequencing (WGS), whole exome sequencing (WES) and whole transcriptomic sequencing (RNAseq) analyses, depending on the part of the genome or transcriptome studied.

### **1.1.3.3 Whole Exome Sequencing**

A majority of the cancer mutations occurred in the coding region of the genome, which comprises about 1% of the entire genome. In order to focus more on this region, various methods were developed to enrich these regions during the initial sample library preparation steps. Probe hybridization methods were used to select these nucleic acid fragments, from the pool of random sheared DNA fragments (Gnirke et al., 2009; Hodges et al., 2007; Turner et al., 2009). Given the constraints of sequencing depth and costs, the WES approach turned out to be an effective way of analyzing cancer genome alterations (Ng et al., 2010; Ng et al., 2009). This was mainly due to the 200 times more coverage and about 100 times more read depth that could be achieved with the same amount of sequencing power compared to that of whole genome sequencing (WGS). The technique is still being improved, particularly regarding the more efficient and equal capturing of exons throughout the genome (Bamshad et al., 2011).

### **1.1.3.4 Transcriptomics**

Initially, analyses of the RNA content of cells (transcriptomics) had been carried out using the method known as the serial analysis of gene expression (SAGE), which involved the manual sequencing of 9 base or 13 base tags of cDNAs (Tarasov et al., 2007). But this technique was superseded by the microarray technology. The microarray technology is based on the principle of hybridization of the RNAs with complementary DNA microarrays (Bumgarner, 2013). The DNA microarrays were designed to cover the entire complement of RNAs present in a cell or tissue sample, i.e. the transcriptome. Microarrays continued to be the method of choice for transcriptome profiling until the advancement of Next-Generation Sequencing (NGS) technology, which was based on short-read sequencing of cDNA fragments of about 100 bp in size. NGS can be applied

to sequence analysis of genomes and of RNA populations (the latter is referred to as RNAseq). However, the potential of RNAseq was greatly increased with the invention of high-throughput techniques for short-read sequencing by Illumina Inc./Solexa, whose Genomic Analyzers enabled reading of gigabase lengths of sequence in a single run (Davies, 2010). RNAseq analysis facilitates the measurement of total gene expression levels, the detection of RNA splicing events, the identification of mutations at the mRNA level and also for the identification of hitherto unknown coding regions. Notably, RNAseq analyses have revealed previously unknown translocations and gene fusions associated with cancer (*ALK-EML4*, *ERG-TMPRSS2*, *BCR-ABL*) (Salesse and Verfaillie, 2002; Soda et al., 2007; Tomlins et al., 2005). These analyses have also led to the development of treatment possibilities targeting the fusion gene products (Wang et al., 2013). Recently, there have been improvements in the RNAseq technology to identify miRNAs, snoRNAs, and translated RNAs (Ribo-seq, CLIP-seq). These developments help greatly in advancing the field of cancer research (McGettigan, 2013).

### **1.1.3.5 Proteomics**

Given the significant advancements in the genomics field, the proteomics also had its share of substantial developments in the past few decades. The term proteome itself was coined in the year 1995 by Ian Humphery-Smith and colleagues to denote the entire protein content being expressed from a genome (Wilkins et al., 1996). In line with the omics development, the term proteomics encompasses the study of protein expression, protein-protein interactions and protein post-translational modifications (Prabakaran et al., 2012). The initial proteomic methods were based on 2-dimensional gel electrophoresis (O'Farrell, 1975) followed by identification of the separated proteins via mass spectrometry (MS). 2D-gel electrophoresis had its own limitations such as low coverage depth, low range of detection and improper protein separation and so on (Gygi et al., 2000). However, with a variety of factors coming into play such as improved genomics technologies, improved mass spectrometry instruments, and also better protein and peptide separation techniques, a new high-throughput proteomic

technology could be developed. It is referred to as the shotgun proteomics. It involves a combination of top down and bottom up approaches for protein analysis. In this approach, protein mixtures were subjected to digestion by proteolytic enzymes such as trypsin or Lys C and the resulting peptide mixtures were analyzed (Tyers and Mann, 2003). In this technique, proteins isolated from cells and fragmented into peptides, which are separated by liquid chromatography (LC) and identified by coupled mass spectrometry (LC-MS). This includes the fractionation of selected peptides using tandem mass spectrometry (LC-MS/MS) (Hu et al., 2007). The obtained spectra are then analyzed for both peptide identification and quantification purposes. This quantification could be done in various ways such as label-free quantification, or by using stable isotope labeling (Asara et al., 2008; Zhu et al., 2010). In recent years there have been numerous improvements in the different steps of the LC-MS/MS method, which have helped in enabling characterization and comparison of the proteomic profiles, absolute quantification of proteins, as well as the determination of dynamic and spatial characteristics of proteins in the cells.

Moreover, there have been significant advances in the field of identifying post-translational modifications (PTM) as well. In recent years, various strategies have been developed to address the identification of various PTMs. For example, in case of protein phosphorylation, new methods have been developed such as phospho-peptide enrichments using immobilized metal ion chromatography (IMAC), titanium dioxide chromatography, phosphotyrosine immunoprecipitation, or a combination of these approaches (De Corte et al., 1999; Macek et al., 2009; Pinkse et al., 2004; Villen and Gygi, 2008). These technologies have contributed to our understanding of the signaling pathways involved in cancer. The accurate quantification of phospho-peptides has helped to delineate the role of kinases and phosphatases in cancer progression.



### **1.1.3.6 Metabolomics**

Metabolomics denotes analysis of the complete spectrum of metabolites present in a cell or tissue. Dynamic analysis of the cellular metabolome, in response to different treatments at different time-points, helps in the elucidation of the physiologic responses of cells to drug treatments or any other external modulation (Kaddurah-Daouk et al., 2008). This can be used in many ways in cancer research since tumor cells have to adjust their physiology to the specific needs of a malignant tumor tissue. Animal studies have also demonstrated the advantages of metabolomic analyses in the pursuit of toxicological studies (van Ravenzwaay et al., 2007). Metabolic pathways in cancer frequently relate to altered requirements of tumor cells for energy (ATP) and anabolic building blocks (nucleotides, amino acids, membrane components, etc.). Newer fields such as lipidomics as well are catching up, in order to look at the wide spectrum of chemical components involved in tumor proliferation (Shevchenko and Simons, 2010). Metabolomic analyses can also be carried out using mass spectrometry-based analytical tools. Developments in LC-MS/MS analyses have greatly improved the sensitivity, coverage and efficiency of metabolomic studies (Menni et al., 2013; Mittelstrass et al., 2011).

### **1.1.3.7 Combined 'omics' analysis**

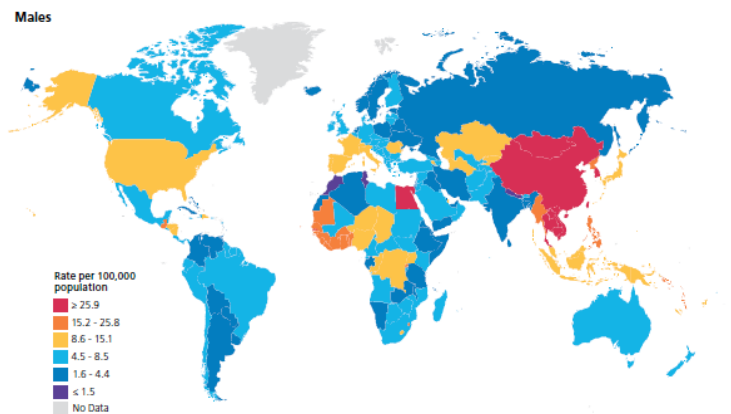
In order to better understand the molecular mechanisms underlying the special characteristics of cancer cells, various combinatorial approaches can be employed using different omics strategies. The combined omics approaches have already been successfully used to generate significant insights into human disease biology. One study, for example, reported the simultaneous analysis of different omics technologies, including transcriptomics, proteomics, genomics and metabolomics of a single individual patient for a particular period of time; the study showed various phenotypic insights, which ultimately helped in revealing various medical risks of the patient (Chen et al., 2012). Another example demonstrating advantages of such combined omics

approaches in cancer research investigated ovarian cancer. In this study, ovarian tumor samples collected as part of The Cancer Genome Atlas (TCGA) program were characterized based on their genomic copy number variations, proteomic profile and posttranslational modifications. This study revealed various pathways that could be associated with patient survival (Zhang et al., 2016a). It is well known that the cancer cells can acquire resistance to certain drugs, often resulting in recurrent aggressive tumors. Such resistance is known to be acquired by mutations in key genes involved in regulating cancer cell physiology. In one such example, mutations in the EGFR gene rendered acquired resistance of lung cancer cells against the tyrosine kinase inhibitor Gefitinib (Ma et al., 2011). Combined omics approaches can effectively identify such survival mechanisms activated by the cancer cell at genomic, proteomic or metabolomic levels. This may lead to strategies of personalized medicine, where a combined omics analysis of tumor tissue from an individual cancer patient can help decide the best possible treatment modality. Such therapeutic decisions could be made in an individualistic manner, using combined multi-omics approaches.

Various databases exist, such as TCGA (The Cancer Genome Atlas) and ICGC (International Cancer Genome Consortium), which have compiled large-scale multi-omics data sets of human cancer patients. This encourages further research on the alterations occurring at different omics levels in various cancers.

### **1.1.4 Liver cancer**

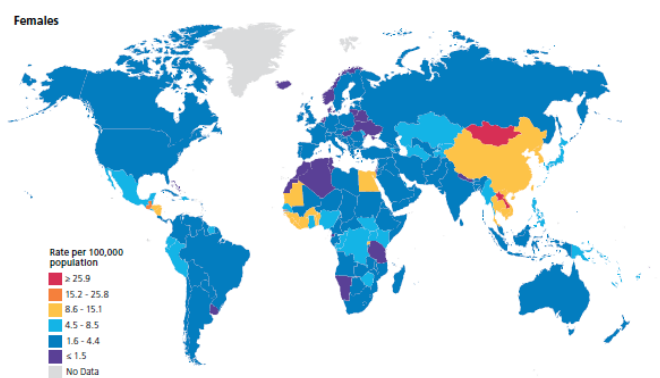
Cancer of the liver is the sixth most frequently occurring cancer worldwide and is the second highest cause of cancer-related deaths. It is more prevalent among men than women. The geographical distribution of liver cancer occurrences among men and women for the year 2012 is shown in (Figure 1-2a and 1-2b). Liver cancer is more common in the regions of sub-Saharan Africa and Southeast Asia. This is likely due to the higher prevalence of Hepatitis B and Hepatitis C virus infections in these regions. Among the 782,000 new liver cancer incidences diagnosed worldwide in 2012, about half were registered in China (Ferlay et al., 2015).



\*Per 100,000, age standardized to the World Standard Population. Source: GLOBOCAN 2012.

Figure 1-2a: Global incidence map of liver cancers in males.

Source : (Ferlay et al., 2015).



\*Per 100,000, age standardized to the World Standard Population. Source: GLOBOCAN 2012.

Figure 1-2b: Global incidence map of liver cancers in females.

.Source : (Ferlay et al., 2015).

The mortality of liver cancer is the second highest among all cancers. Liver cancer has the worst mortality-to-incidence ratio among the top 8 common cancers around the world (Figure 1-3) (Ferlay et al., 2015).

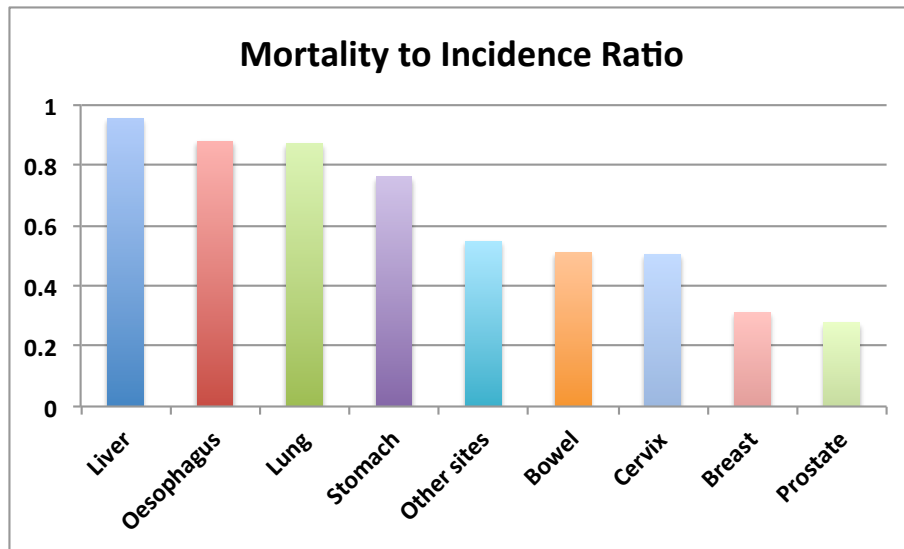


Figure 1-3: Mortality-to-incidence ratios of cancers of different organs. Liver cancer had the worst mortality-to-incidence ratio in the year of 2012. Source: Globocan 2012

### 1.1.5 Types of liver cancer

There are different types of liver cancer depending on their cellular origin.

➤ **Hepatocellular carcinoma (HCC)**

More than 70% of liver cancers are of this sub-type. They arise from the hepatocytes of the liver. HCCs can be further classified into different sub-categories such as scirrhous HCC, fibrolamellar carcinoma, combined HCC-cholangiocarcinoma (HCC-CC), sarcomatoid HCC, lymphoepithelioma-like HCC, clear cell HCC, diffuse cirrhosis-like HCC, and steatohepatic HCC (Shafizadeh and Kakar, 2013).

➤ **Intrahepatic cholangiocarcinoma (ICC)**

These comprise about 10% - 20% of all liver cancers. They originate from the cholangiocytes, which are cells lining the bile ducts inside the liver.

➤ **Angiosarcoma and hemangiosarcoma**

This sub-type of liver cancer arises from the endothelial cells, lining the blood vessels within the liver. The major risk factor associated with these cancers is the exposure to vinyl chloride and thorium dioxide (Thorostrast). It can also be caused by

exposure to radium or arsenic. Individuals having the condition of hereditary hemochromatosis are known to be more prone to develop this cancer.

➤ **Hepatoblastoma**

This is a rare type of liver cancer affecting mainly children younger than 4 years of age. The morphology of the cells of this tumor closely resembles that of the fetal liver cells.

➤ **Secondary liver cancer (Metastatic liver cancer)**

Secondary liver cancer occurs mainly from primary cancers of the pancreas, colon, stomach, breast and lung. It is more predominant in Western countries compared to Africa and Asia.

## **1.2 Hepatocellular carcinoma (HCC)**

HCC is the most common form of primary liver cancer. It is the third most leading cause of cancer-related deaths worldwide (Waller et al., 2015). In the United States alone, HCC is the fastest rising cause of cancer mortality. The US incidence of HCC has increased three-fold in the past few decades. This may be attributed to the increasing Hepatitis C virus infections among the population as well as dietary habits. The etiological factors associated with HCC are as follows.

### **1.2.1 Etiological factors**

#### **1.2.1.1 Cirrhosis**

Cirrhosis is one of the main risk factors associated with HCC. Cirrhosis could be caused by different sources such as Hepatitis virus infection or inflammation. Prolonged inflammation or any other disease in the liver might affect proliferative activity of hepatocytes., Hepatocytes are among the cells of the human body of highest proliferation capacity, contributing to the liver's unique regenerative ability (Delhaye et al., 1996). Increased fibrosis in the liver, followed by cirrhosis, is associated with liver

## INTRODUCTION

---

cancer progression, enabling malignant transformation of hepatocyte and hence HCC formation (Caillot et al., 2009).

The mechanisms causing cirrhotic livers to develop HCCs have been investigated widely. Two of the key factors accelerating this process are telomere dysfunction and changes in the liver micro- and macro-environment (El-Serag and Rudolph, 2007).

The telomere size of the cell is maintained by the enzyme telomerase, which also maintains the chromosomal stability of proliferating cells (Wege and Brummendorf, 2007). Shortening of telomere lengths as a consequence of elevated cell division activity has been directly correlated with enhanced fibrosis and is often observed in chronic diseases of the liver (Wiemann et al., 2002). Telomere dysfunction is always accompanied by other factors such as *p53* mutations, accelerating the progression of cirrhotic livers into cancer stages (Sanyal et al., 2010).

Other factors involved in fibrosis and cirrhosis include the activation of stellate cells, often observed in cirrhotic livers, which results in the production of cytokines and growth factors, as well as stiffening extracellular matrix structures, all of which assist in the development of liver tumors. The milieu of a cirrhotic liver may also cause the activation and transformation of liver stem cells, as suggested by RNA expression patterns found in human HCCs that are typical of stem cells (Sanyal et al., 2010). Other molecular alterations that might trigger the malignant transformation from cirrhotic tissue are those affecting the DNA damage checkpoints such as impaired *p53* function, inhibition of *p27* cell cycle regulator, down-regulation of *p16* cell-cycle inhibitor, and loss of heterozygosity of the Insulin-like growth factor 2-receptor (*IGF2R*) locus (Sanyal et al., 2010).

### 1.2.1.2 Hepatitis B virus (HBV) infection

HBV infection accounts for more than 50% of HCC occurrences worldwide. About 90% of HBV infections are acute infections. Increased levels of HBV DNA has been found to predispose to elevated HCC incidence in patients (Farazi and DePinho, 2006). HBV is a partially double-stranded DNA virus belonging to the *Hepadnaviridae* family. The HBV genome codes for the following components: reverse transcriptase and DNA polymerase (pol), capsid protein called the hepatitis B core antigen (HBcAg), the envelope proteins which are associated with the endoplasmic reticulum membrane, and the x protein (HBx). The integration of the HBV genome in the host cell DNA results in a variety of immediate responses in the cell, such as regulation of cancer-related signaling pathways and activation of relevant genes, namely telomerase reverse transcriptase (*TERT*), platelet-derived-growth-factor receptor- $\beta$  (*PDGFR- $\beta$* ), *PDGF- $\beta$*  and mitogen activated protein kinase 1 (*MAPK1*) (Farazi and DePinho, 2006). HBV infection can also trigger genomic instability by causing chromosomal rearrangements (Farazi and DePinho, 2006). The protein HBx can alter the expression of several growth-control genes and can also bind and inactivate the p53 tumor suppressor protein. This causes an increase in cellular proliferation at the cost of efficient DNA-damage checkpoints. The protein HBx alone was sufficient to induce HCCs in mice at a high infectivity of 90% (Sanyal et al., 2010).

There are multiple ways of tumor induction upon HBV infection. One involves the immune response of the host cell, resulting in T-cell activation, hepatocyte necrosis and liver inflammation, which leads to constant regeneration and ultimately liver cancer (Farazi and DePinho, 2006). Another mechanism involves the interaction of the virus with the endoplasmic reticulum (ER), resulting in ER stress and subsequently oxidative stress, leading to the formation of free radical-induced mutations and also stellate cell activation. HBV infection can result in HCCs either directly or by causing pre-cancerous cirrhosis. HBV is also known to cause liver diseases other than HCCs (Farazi and DePinho, 2006).

### 1.2.1.3 Hepatitis C virus (HCV) infection

HCV infection accounts for about one-third of liver cancers worldwide. HCV infection always results in cirrhosis of the liver before progressing into cancers. About 60 – 80% of HCV infections are chronic infections. HCV is a positive-stranded RNA virus belonging to the *Flaviviridae* family. The genome of the HCV virus encodes a host of non-structural core proteins (NS2, NS3, NS4A, NS5A and NS5B), which interact with the ER of the host cell and produces the viral replicase and viral envelope proteins (Farazi and DePinho, 2006). The HCV core protein can also promote oxidative stress, resulting in activation of the MAPK and the NF $\kappa$ B signaling pathways. This triggers cytokine production and then inflammation of the liver, which leads to tumor formation (Sanyal et al., 2010). HCV infection has been found to result in hepatic steatosis of the liver and induction of reactive oxygen species, subsequently causing HCC development (Farazi and DePinho, 2006)

The HCV is known for its elaborate immune evasion mechanisms, either by preventing T-cell activation by inhibiting dendritic cell functions, or by interacting with various factors such as Tumor Necrosis factor  $\alpha$  (TNF $\alpha$ ) and interferon- $\alpha$  (IFN $\alpha$ ). HCV core proteins have been observed to directly interact with components of the MAPK signaling pathway and activate it. The core protein NS5A has been reported to inactivate p53 function by sequestering it to the perinuclear membrane (Farazi and DePinho, 2006). HCV infection has also been associated with fibrosis and type-2 diabetes in patients, mainly due to the induction of insulin resistance (IR) in the patients. Mice containing the HCV core protein have been shown to develop IR and also an increase in levels of the pro-inflammatory factor TNF $\alpha$  and IL-6, probably secreted by the Kupffer cells in response to oxidative stress caused by the HCV core protein (Sanyal et al., 2010) This suggests a complex mechanism involved in the HCV induced carcinogenesis of the liver.



#### **1.2.1.4 Alcohol consumption**

Alcohol is one of the most important life style-associated risk factors of HCC. Similar to consequences upon HCV infection, chronic alcohol exposure can result in the activation of Kupffer cells and monocytes and subsequently cause the release of cytokines and chemokines such as  $\text{TNF}\alpha$ ,  $\text{IL-1}\beta$  and  $\text{IL-6}$ , which greatly affects hepatocyte survival. This sets up a mechanism of continuous destruction and regeneration of hepatocytes, which leads to stellate cell activation and ultimately liver cancer (Farazi and DePinho, 2006). Alcohol exposure also leads to oxidative stress, which can result in fibrosis of the liver, by induction of the activated stellate cells into synthesizing more collagen, and enhanced cell proliferation. The fibrotic condition can subsequently develop into cirrhosis and ultimately support the formation of HCC. This was demonstrated in PDGF transgenic mice (Campbell et al., 2005). Oxidative stress may also cause hepatocyte damage by inhibiting the  $\text{STAT1}$ -directed activation of  $\text{IFN}\gamma$ , and also by causing oncogenic mutations. It may also cause chromosomal instability by accelerating the shortening of telomeres, which activates liver cirrhosis (Farazi and DePinho, 2006).

Alcohol can also act synergistically with HCV infections in causing HCCs. Alcohol can enhance HCV replication and also induce changes in the hyper-variable region of the viral genome. The mitochondrial and ER stress caused due to HCV infection could be further aggravated upon alcohol exposure, which further leads to increased ROS production and glutathione oxidation (Sanyal et al., 2010). The association of alcohol and HCV infection has the maximum effect on induction of oxidative stress in the liver, thereby pushing the organ into fibrosis and cirrhosis.

### **1.2.1.5 Aflatoxin B<sub>1</sub>**

Aflatoxin B<sub>1</sub> is a fungal toxin associated with elevated incidences of HCC formation. Aflatoxin B<sub>1</sub> exposure and HBV infection often co-exist in HCC patients. Aflatoxin B<sub>1</sub> is known to cause activating mutations in the HRAS oncogene and also inactivating mutations in the p53 tumor suppressor gene (Farazi and DePinho, 2006).

### **1.2.1.6 Non-alcoholic fatty liver disease (NAFLD)**

About one-fifth of the population of Western countries suffers from this condition. NAFLD is a high risk factor associated with HCC occurrence. NAFLD is histologically similar to alcoholic hepatitis although it is not caused due to alcohol exposure. Some of the variations observed in NAFLD livers include features such as hepatic steatosis, hepatocyte injury, and cytological ballooning. This implies an underlying complexity in terms of pathological conditions existing in these livers, such as excessive fat, inflammation, ballooning degeneration and nonalcoholic steatohepatitis (NASH) (Sanyal et al., 2010). NAFLD has been closely associated with type 2 diabetes and obesity. The three conditions of NAFLD, NASH, and HCC are likely to be co-occurring. NAFLD is postulated to be driven by insulin resistance (IR), which increases the fatty acid uptake of the liver by increasing peripheral lipolysis. This results in fat deposition in the liver coupled with simultaneous increase in lipogenesis (de novo) and reduction of lipoprotein secretion from the liver, which causes accumulation of triglycerides in the liver. Increased levels of intrahepatic fatty acids can result in oxidative stress and the subsequent production of ROS, which oxidize the fat deposits into lipid peroxidation products, leading to oxidative damage to the mitochondrial genome and impairment of the respiratory chain. This can also induce the production of various cytokines, activation of stellate cells and subsequently progression to HCC (Sanyal et al., 2010).

## **1.2.2 Histopathology of HCCs**

Histological examination of tumor biopsies is one of the most important methods in HCC diagnostics. The morphological features of HCCs include tumor vascularization accompanied by a trabecular phenotype. Other prominent features include acinar patterns, cytologic atypia, mitotic activity, active vascular invasion and the loss of Kupffer cells and reticulin network. It has been observed that the morphological features of HCCs can differ significantly from one patient to another. This may be due to the different stages of progression and dedifferentiation of the tumor, discernable by the different growth patterns in the tumor mass. It had led to the postulate that an initial well-differentiated small lesion undergoes step-wise dedifferentiation into larger tumors of lower differentiation states, completely replacing the lesion by a nodule-like appearance. Other hypotheses postulate a multi-centric origin of HCCs, mainly due to the advanced HCCs showing morphological features such as encapsulation and well-established neovascularization consisting of unpaired arteries, and at the same time containing the other typical histological patterns of HCC (Kojiro, 1998). These tumors can also give rise to satellite nodules in the vicinity of the progressed primary tumor. The unpaired arteries of these primary tumors were found positive for smooth muscle actin (SMA) and CD34 (Nakashima et al., 1995). Based on the histological observations, HCCs with unique features can be classified into different types (see below), which guides the choice of subsequent therapeutic strategies to be applied.

## **1.2.3 HCC types**

### **➤ Fibrolamellar HCC**

Fibrolamellar HCC is one of the less frequent types of HCC, observed in young patients. It is not associated with liver cirrhosis or any other predisposing risk factors. The histopathological features of fibrolamellar HCC consist of the tumor cells growing in trabeculae and are separated by collagen fibers, visible by their lamellar pattern (Ganeshan et al., 2014).

### ➤ **Sarcomatous HCC**

This type of HCC can also be found within classical HCC. Histologically, sarcomatous HCC are recognizable by their spindle-shaped appearance and the presence of various anaplastic figures. They bear resemblance to leiomyosarcoma and fibrosarcoma, when occurring alone, and not along with other types of HCC (Yamaguchi et al., 1997).

### ➤ **Scirrhou HCC**

This sub-type of HCC is characterized by fibrosis occurring in response to antitumor treatments and rarely in untreated tumors. Reports have shown that a majority of these tumors have observable CD8+ lymphocyte infiltration, along with trabecular atrophy (Kojiro, 1998).

### ➤ **Clear-cell variant of HCC**

This variant of HCC was found arranged in trabecular pattern and the cells of these tumors are characterized by presence of fat vesicles in the cytoplasm, containing glycogen. It is also found predominant in males rather than in females (Yang et al., 1996).

### ➤ **HCC with lymphoid stroma**

As the name suggests, this sub-type is characterized by inflammatory cells such as lymphocytes, macrophages, plasma cells, neutrophils and other giant cells. There were also found in the tumor tissues, CD4+ T cells and CD3+ T cells were also found (Park et al., 2009).

## **1.2.4 HCC treatment modalities**

### **1.2.4.1 Surgical resection**

Removal of the tumor-containing part of the liver is the favored method of treatment, provided the HCC is not associated with cirrhosis and the lesions are far early in their stage of progression. In cirrhotic livers, resection is performed only if other criteria, such as the tumor size and bilirubin levels are conducive. Surgical resection is the preferred method of treatment for HBV-related HCCs, due to these livers usually developing no cirrhosis (Llovet et al., 2005).

### **1.2.4.2 Liver transplantation**

For the patients who are unsuited to be treated with surgery, liver transplantation is the preferred therapeutic strategy (Ryder and British Society of, 2003). Owing to practical and technical difficulties associated with liver transplantation, stringent rules and criteria have been developed, which assist in the selection of patients for this treatment. The Milan criteria have been applied worldwide to categorize the suitability of patients for this treatment. Additionally, the MELD scores have been assigned to patients, to decide on the priority of patients on the waiting list to receive donor organs. When patients do not meet the respective criteria for liver transplantation, they are planned for other forms of therapy. There have been different targeted therapies such as mTOR inhibitors, which are prescribed to patients after liver transplants, to prevent the recurrence of tumors, with considerable success (Cholongitas et al., 2014)

### **1.2.4.3 Local ablation**

The patients deemed unsuitable for liver transplantation are treated with local ablation of the tumor. This can be done by Percutaneous Ethanol Injection (PEI) or by the more modern Radiofrequency Ablation (RFA). The RFA method has been shown to

be more successful in the removal of smaller sized tumors, compared to larger sized tumors. Comparing to surgical resection, RFA has not shown any significant advantage in terms of recurrence-free survival or overall survival of the patient (N'Kontchou et al., 2009).

### **1.2.4.4 Chemoembolization**

This treatment method is recommended for patients of stage B classification according to the Barcelona clinical liver cancer staging recommendation, and to patients deemed unfit for other of the above-mentioned treatment modalities. It involves the catheterization of the hepatic artery to deliver chemotherapeutic drugs directly to the tumors. With the infusion of radioactive substances, such as Iodine-131, the response has been shown to be better (European Association for Study of et al., 2012). There have been more and more developments to this method, to reduce side effects and improve efficacy (Lencioni et al., 2013).

### **1.2.4.5 Targeted therapy**

With the increasing insights into the different signaling pathways being activated in HCCs, numerous targeted drugs have been developed, which inhibit these different cellular pathways and reduce proliferation of the cancer.

Sorafenib is an FDA approved targeted therapy drug for HCC. Sorafenib has been known to inhibit a host of cellular receptors including the VEGF receptors (VEGFR-1 , VEGFR-2, VEGFR-3), PDGFR $\beta$ , RET, and c-KIT. Sorafenib is known to inhibit the Ras/MAPK pathway (Wilhelm et al., 2006). The SHARP clinical trial has shown an overall survival benefit of 3 months for patients on Sorafenib compared to placebo treatment (Rimassa and Santoro, 2009). Sorafenib has also been used as neoadjuvant, along with chemotherapeutic agents such as doxorubicin, ocreotide, oxaliplatin, etc. Sorafenib has been administered in combination with all methods of treatment for HCC, such as chemoembolization and resection, in order to improve therapeutic efficiency (Ziogas and Tsoulfas, 2017).

There have been other targeted drugs subjected to clinical trials, including Sunitinib, Linifanib, Brivanib, Tivantinib, Everolimus Erlotinib, Bevacizumab and Brivanib (Llovet et al., 2015; Thillai et al., 2016). In the context of the therapeutic potential yet to be realized by targeted therapies, it is imperative to study the various underlying mechanisms of hepatocarcinogenesis, in order to identify new molecular targets.

### **1.2.5 Molecular mechanisms of hepatocarcinogenesis**

Different molecular mechanisms have been identified to be involved in the process of hepatocarcinogenesis. The most predominant of those is the inactivation of p53 in HCC patients with etiology of HBV/ HCV infection or aflatoxin B<sub>1</sub>-exposure (Hussain et al., 2007; Tornesello et al., 2013). Further, the MAPK pathway is activated by HBV and HCV infections. The underlying mechanisms of carcinogenesis might be different among different patients, due to different etiological factors. Recent advances in various genome analysis technologies have helped to obtain significant insight into some of these mechanisms.

#### **1.2.5.1 Genetic and epigenetic alterations**

The constant damage and regeneration of hepatocytes due to various etiological factors provide a platform for further genetic and epigenetic alterations to be acquired by these cells that ultimately assists in the malignant transformation process. Molecular analysis of tumors from HCC patients has unraveled different genomic alterations acquired during the step-wise progression of HCC from dysplastic nodules.

#### **p53 tumor suppressor**

Inactivating mutations in the p53 gene are the most frequently found mutations among HCC patients (Hussain et al., 2007). It has been observed in HBx transgenic mice that functional inactivation of p53 occurred in advanced tumors and not in early

nodules, thereby suggesting an involvement of this mechanism late in the process of malignant transformation (Feitelson et al., 2002). This observation has been corroborated by studies performed in human tumors, which also showed p53 inactivating mutations to be more predominant in HBV, or HCV infected HCC patients and also found more commonly in advanced tumors compared to recurrent pre-malignant nodules (Minouchi et al., 2002). Considering that most of the aetiological factors associated with HCC incidence cause oxidative stress, telomere shortening, or hepatocyte damage, the additional burden of p53 mutations might establish a situation wherein the cells are forced into continuous proliferation despite the presence of DNA damaging agents. This, along with telomere shortening, can lead to chromosomal instability in the tumors. This phenomenon was also shown in p53 heterozygous mice carrying shorter telomeres with chronic liver damage, which went on to develop HCCs (Farazi et al., 2006).

### **β-Catenin activation**

β-Catenin is a downstream effector of the Wnt signaling pathway. Activating mutations in the β-catenin gene (*CTNNB1*) is often observed in HCC patients and results in the stabilization of the protein. The associated accumulation of the β-catenin protein and its subsequent nuclear localization result in transcriptional activation of various cancer-related genes, such as *MYC*, *CCND1*, *MMP7*, etc (Ishizaki et al., 2004). Overexpression of the wild type β-catenin gene has also been reported in HCCs by various studies (Thorgeirsson and Grisham, 2002). In rare cases, β-catenin mutations have been found in conjunction with other mutations, such as in the p53 gene, or in HCCs with chromosomal instability. This apparent mutual exclusivity of β-catenin activation with other mechanisms suggests an alternative progression mechanism involving the Wnt signaling pathway harnessed by HCCs (Calvisi et al., 2004).

### **ErbB receptors overexpression**

The ErbB receptor tyrosine kinase family, comprising the proteins ErbB1, ErbB2, ErbB3, and ErbB4, has been found to be overexpressed in multiple HCC patients.



Particularly ErbB<sub>1</sub> and ErbB<sub>3</sub> overexpression was associated with more aggressive cancers (Yamamoto et al., 2001). Inhibition of these receptors by drugs such as Gefitinib and Erlotinib showed significant responses in HCC patients. Also human cell-lines were affected by these drugs *in vitro*, where they seemed to induce apoptosis (Hopfner et al., 2004). Furthermore, various studies involving one of the ligands of the ErbB family of receptors, demonstrated Transforming Growth Factor  $\alpha$  (TGF $\alpha$ ) to display a close link between elevated expression and HCC induction, progression and poor survival. Lack of TGF $\alpha$  in transgenic mice reversed the pro-carcinogenic effects of TGF $\alpha$ . This emphasized the importance of the TGF $\alpha$ -ErbB receptor pathways in hepatocarcinogenesis (Jakubczak et al., 1997).

### **Epigenetic regulation**

Epigenetic alterations have been reported in a majority of human HCCs, both in early and late stages of tumor development (Lee et al., 2003). Hyper-methylation of CpG positions in selected genes, such as p16 (*INK4A*), E-cadherin (*CDH1*), COX2, apoptosis-associated speck-like protein (*ASC*) and deleted-in-liver-cancer 1 (*DLC1*), have been observed. In one of the studies, treatment of tumor cells with demethylating agents resulted in over-expression of p16 (*INK4a*), thus implying a robust epigenetic silencing mechanism in order to suppress genes such as p16 (*INK4A*), encoding an inhibitor of cellular proliferation (Maeta et al., 2005). There have also been reports of hypomethylation events, which resulted in the loss of imprinting mechanisms and causing the over-expression of genes such as Insulin-like growth factor 2 (*IGF2*) that promote cellular proliferation and tumor progression.

### **1.2.5.2 Genomic instability**

Genomic Instability is a hallmark feature of cancer, including HCC, caused by multiple mechanistic cues.

### **Telomere shortening**

Telomere shortening has been widely observed in chronic liver disease. The various etiological factors contributing to hepatocyte proliferation, coupled to telomere shortening associated with chronic liver disease, can ultimately result in genomic instability in the cells (Urabe et al., 1996). Genomic instability can also assist in tumor progression by causing alterations in the genome with each successive cycle of cell division. Studies performed on transgenic mice have also proved the link between chromosomal instability and telomere shortening ultimately resulting in tumor initiation (Farazi et al., 2003).

### **Telomerase overexpression**

Over-expression of the *hTERT* gene has been found in a vast majority (about 90%) of human cancers (Zhou et al., 2016). Further evidence suggesting the important role of the TERT gene came from the observation that the HBV frequently integrates in the *TERT* locus of liver cells of human HCC patients (Ferber et al., 2003). Also, telomerase re-activation has been observed in recurrent HCCs after resection (Kobayashi et al., 2001). These findings, combined with the previous insights about telomere shortening, suggest a mechanism wherein telomere shortening triggers the malignant transformation in liver cells, which is followed by telomerase reactivation that assists in the progression of the tumors. This was supported by a study on *Tert* null transgenic mice where telomerase dysfunction aided the induction of HCCs, whereas complete lack of telomerase activity inhibited HCC progression (Farazi et al., 2003).

### **1.2.5.3 Metabolic reprogramming**

Changes in the metabolic preferences of cancer cells has been widely studied, ever since the landmark theory by Otto Warburg in 1924 also known as aerobic glycolysis or “Warburg effect”. There has been increasing interest of late in this theory, after recent findings emphasized the various metabolic alterations that might assist the progression of tumors (Figure 1-4). The metabolism-associated *HIF1* gene has been

found to assist in HCC progression. *HIF1* regulation is known to be controlled by other genes involved in determining metabolic activity of a tumor cell (Griffiths et al., 2002). There have been additional reports identifying metabolic changes in the tumor and stromal cells, thereby directly affecting endothelial cells and tumor neo-angiogenesis (Fraisl et al., 2008). Recently the gene *PARP14* was reported to be an important driver of the Warburg effect in HCCs (Iansante et al., 2015).

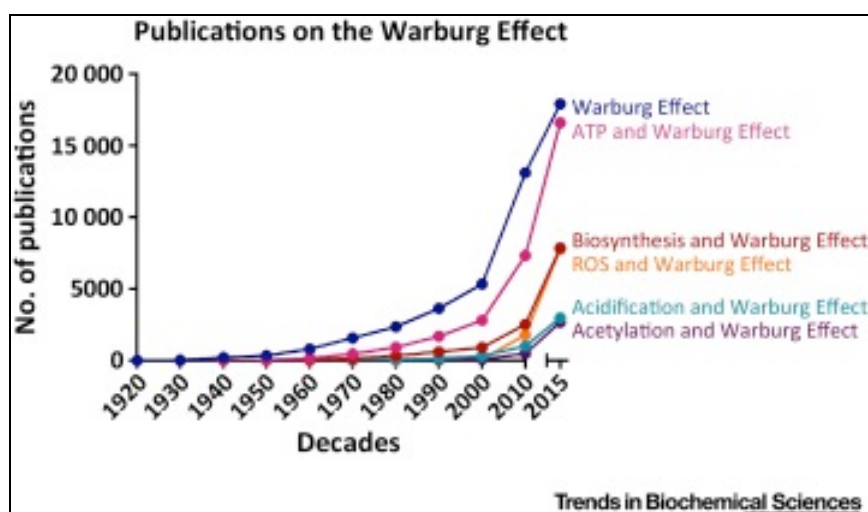


Figure 1-4: Publications characterizing the Warburg effect. A strong increase in publications analyzing the Warburg effect is evident in the past decades.

Picture from: (Liberti and Locasale, 2016)

#### 1.2.5.4 Tumor neo-angiogenesis

During the step-wise progression of HCC in human patients, it is noted that early dysplastic nodules are difficult to detect by the existing methods. Unlike the progressed HCCs, which can be visualized using ultrasound imaging due to the ability of HCCs to display arterial phase uptake upon contrast imaging. It has been postulated that the reason behind the above-mentioned observation rests on the formation of new blood vessels in the advanced HCCs. This has been considered a crucial milestone in the progression of HCC, after which the prognosis to all types of treatment modalities turns significantly worse (Takayama et al., 1998). This crucial event has been called the “angiogenic switch”. This has been further substantiated by the overexpression of VEGF and its regulator *HIF1* in many human tumors. Moreover, one of the most successful

chemotherapeutic drugs against HCC, Sorafenib, has been identified to inhibit VEGF, among other molecules. This further underscores the importance of angiogenesis in HCC progression.

### **1.2.5.5 Epithelial-to-mesenchymal transition**

During the progression of HCCs, some of the tumor cells undergo a phenomenon known as epithelial-to-mesenchymal transition (EMT). It is the process, which brings about a transient de-differentiation of epithelial-type tumor cells to a mesenchymal phenotype. This is accompanied by changes in gene expression patterns of the tumor cell, wherein mesenchymal marker genes, such as N-cadherin (*CDH2*),  $\alpha$ -smooth muscle actin ( *$\alpha$ SMA*) and fibroblast-specific protein (*FSP-1*) are over-expressed. The EMT is controlled by transcription factors including Snail (*SNA1*), Slug (*SNA2*), Twist and ZEB. Their expression levels are also increased in the context of EMT. The EMT leads to elevated motility of tumor cells, thereby contributing to the metastatic spread of cancers. The mesenchymal cells, however, can be reversed to their epithelial phenotype by a process called mesenchymal-to-epithelial transition (MET) (Giannelli et al., 2016). MET is frequently observed once tumor cells extravasate from the vessel system and engage in the formation of metastatic tumors at locations distant from the primary tumor.

### **1.2.5.6 Altered regulation of gene expression**

Multiple mechanisms are involved in the modulation of gene expression in tumor cells. Below, some molecular mechanisms are described, by which cancer cells engage in altered gene expression.

#### **MicroRNAs**

MicroRNAs (miRNAs) are short RNA molecules of about 22 nucleotides length that suppress the expression of their target genes by binding to the 3' region of target transcripts, subsequently triggering either target mRNA degradation or impairment of

target mRNA translation. This enables microRNAs to have both oncogenic and tumor suppressive function, depending on the gene targeted. In HCC, several miRNAs have been found differentially regulated. Some miRNAs, like miR-221, miR-195, miR-128-2, miR-331-3p and miR-371-3 cluster, are found over-expressed in tumors, thus inferring an oncogenic function, whereas the miRNAs miR-100, let-7a-2, miR-125b-1 and miR-148a have been found down-regulated in tumors probably due to their tumor suppressive functions (Cairo et al., 2010) (Xu et al., 2009). In recent years, our understanding of the roles of miRNAs in HCC formation has significantly increased. Figure 1-5 depicts the tremendous increase in Pubmed listings of publications dealing with this area of research .

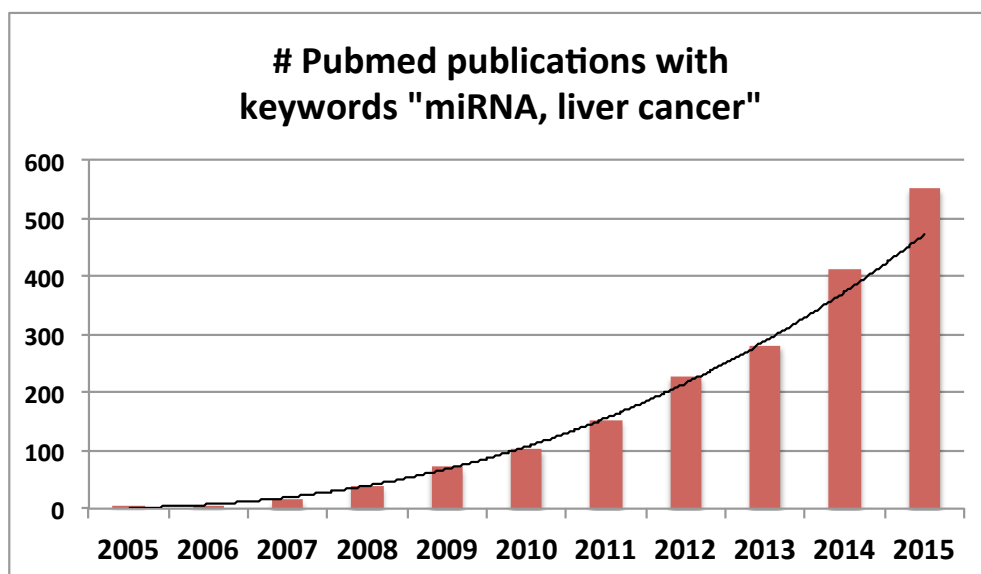


Figure 1-5: Pubmed publications with keywords “miRNA, liver cancer”. The number of publications linking miRNAs and cancer have further increased in the past years. Data source: Pubmed

### Transcription factors

Transcription factors are the special type of cellular proteins that mediate extra- and intra-cellular stimuli into changes in cellular gene expression. Hence they are one of the chief modulators of cellular functions in normal and tumor cells. Numerous transcription factors have been found differentially regulated in human HCCs, such as

Far upstream element binding proteins (FBPs), Yes-associated protein (YAP), signal transducer and activator of transcription (STAT3), SMADs, Nuclear Factor of kappaB (NF- $\kappa$ B), Forkhead box M1B (FOXO1), or ATF5. However, the underlying role and oncogenic functions of many of the transcription factors are not been completely understood yet. Nevertheless, numerous studies on human and mouse tumors have helped to unravel the specific functions of certain transcription factors in tumor progression.

➤ **v-myc avian myelocytomatosis viral oncogene homolog (c-Myc)**

c-Myc is a proto-oncogene predicted to regulate genes involved in various cellular processes such as growth, metabolism, cell-cycle, apoptosis, cell adhesion, DNA repair and microRNA regulation. c-Myc-dependent target gene expression signatures and c-Myc-associated miRNA signatures revealed overexpression in human HCCs (Cairo et al., 2010). The human genomic locus 8q24.21 containing the *c-Myc* gene was found amplified in 60 % of human HCCs, which could account for the over-expression of the protein in multiple human HCCs.

➤ **Activator protein-1 (AP-1)**

The AP-1 transcription factor is a hetero-dimer of different combinations of the proteins Jun, Fos and ATF. They are activated by multiple stimuli in the cell and they influence vital cellular processes such as proliferation, apoptosis, differentiation and consequently - tumorigenesis (Eferl and Wagner, 2003). The specific functions of this transcription factor complex depend largely upon the relative abundance and phosphorylation status of each of its sub-unit proteins. Altered activity of the proteins has been observed even in early stages of malignant transformation (Liu et al., 2002).

➤ **Hypoxia-inducible factor-1 (HIF-1)**

HIF-1 is a transcription factor mainly activated by hypoxia, as the name suggests. HIF-1 is a heterodimeric protein comprising of the subunits HIF-1 $\alpha$  and HIF-1 $\beta$ . The HIF-1 transcription factor is known to activate genes associated with a variety of cellular

functions such as angiogenesis, metabolism, and cell survival. A hypoxic tumor microenvironment is often present in advanced tumors and - due to uncontrolled tumor growth accompanied by high vascularization - these tumors are often poorly differentiated and contribute to very poor patient prognosis. These tumors have been found to contain high amounts of HIF-1 transcription factor and, interestingly, the neighboring non-tumorous liver tissues also expressed high amounts of HIF-1 $\alpha$ . This implies a paracrine effect of the tumor on the surrounding cells in the microenvironment, which might even indirectly assist tumor growth (Simon et al., 2010).

#### ➤ **E2F transcription factors**

E2F denotes a family of transcription factor proteins, including E2F1 through E2F8. They are regulated by the retinoblastoma tumor suppressor protein (Rb). Rb protein, when hypo-phosphorylated, binds and inactivates E2Fs. E2Fs are released and thereby activated upon phosphorylation of Rb and removal of phosphorylated Rb from the chromatin. Other regulatory mechanisms acting on E2F1 include miR-195-based mechanisms, and genomic amplification and deletion mutations of the E2F genes (Choi et al., 2001). Many genes belonging to the E2F family are found up-regulated in HCCs (Zeng et al., 2014).

### **1.2.6 Mouse models for HCC**

Experimental animal models have been assisting cancer research for the past couple of decades. Considering that HCC is a cancer displaying a complex multi-step process of malignant transformation, the need for animal models to study various aspects of HCC is apparent. This need has occasioned the development of numerous mouse models to study HCC.

### **Types of HCC mouse models**

The mouse models currently available, for the study of HCCs can be grouped into the following main categories.

- ❖ Chemically induced mouse models
- ❖ Implantation-based models
- ❖ Transgenic knock-in and knock-out mouse models
- ❖ Spontaneous mouse models

#### **1.2.6.1 Chemically Induced mouse models**

Several chemical agents have been identified that can cause liver cancer in mice when administered at a particular dose for a particular amount of time. The chemical agents include genotoxic substances that can cause changes in the DNA composition and structure, as well as tumor promoting compounds, which by themselves do not cause HCC but they enhance the carcinogenic effects of other compounds.

##### **a. Diethylnitrosamine**

N-nitrosodiethylamine (DEN) is one of the most widely used carcinogenic agents to cause HCCs. DENs can alkylate chromosomal DNA within hepatocytes. The mechanism of carcinogenesis by DEN administration involves the oxygen- and NADPH-dependent hydroxylation of DEN mediated by cytochrome P450 enzyme. This leads to the formation of an electrophilic ethyldiazonium ion upon acetaldehyde cleavage, which causes damage to the DNA. The cytochrome P450 enzyme activity can also cause the release of ROS, which can cause oxidative stress in the cells, which is also a major risk factor for hepatocarcinogenesis (Fausto and Campbell, 2010). Since DEN works by causing DNA damage, DEN-induced HCCs are largely dose-dependent. The duration and extent of DEN exposure also influence the age at which the treated mice develop tumors. DEN-triggered tumours have also been shown to metastasize to the lungs. Additionally, the DEN – induced model can also be combined with exposure to certain other tumor-promoting compounds in order to enhance the carcinogenic potential of



DEN. One such compound is Phenobarbital (PB) (Waxman and Azaroff, 1992). PB is often used in combination with DEN due to its ability to activate the cytochrome P450 enzyme. PB treatment is also known to trigger mutations in the *Cttnb1* gene. This method of cancer induction in mice by DEN is also affected by certain hormones, which results – depending on the mouse strain - in male mice being more prone to cancer than female mice. This gender bias is also observed during PB induced tumor promotion. Collectively, DEN administration is a robust way of HCC induction in mice, which resembles the oxidative stress-induced HCCs in humans (Heindryckx et al., 2009).

### **b. Aflatoxin B1**

The hepatotoxin Aflatoxin B<sub>1</sub> (AFB<sub>1</sub>) is a known risk factor for hepatocarcinogenesis in humans. Hence AFB<sub>1</sub>-induced murine tumors are one of the mouse models that closely resemble the human scenario. AFB<sub>1</sub> is produced in nature is produced by the fungus *Aspergillus flavus* and is found mostly in China and West Africa, where there is already high prevalence of hepatitis virus infections. These etiological factors together result in one of the predominant causes of HCCs around the world. AFB<sub>1</sub> is converted in the liver to an intermediate known as the exo-8,9-epoxide, that can bind to guanine residues in the DNA and cause mutations. These mutations often trigger other genetic or chromosomal aberrations ultimately resulting in HCCs (Heindryckx et al., 2009)

### **c. Carbon tetrachloride**

Carbon tetrachloride (CCl<sub>4</sub>) is a hepatotoxin that acts similar to DEN, whereby CCl<sub>4</sub> metabolizes cytochrome P450 to form free radicals inside the cell. These radicals, cause lipid peroxidation and DNA damage. As a response, Kupffer cells induce an inflammatory response, resulting in the secretion of numerous pro-inflammatory agents. These pro-inflammatory factors also attract monocytes, neutrophils and other immune cells, which can cause liver tissue damage. This repeated process of inflammation and repair leads to fibrosis and then to cancer (Heindryckx et al., 2009).

### **1.2.6.2 Implantation-based mouse models**

Implantation models are one of the most widely used forms of mouse models for HCC, due to their simple and cost-effective mechanism of hepatocarcinogenesis. They can be grouped into different sub-groups based on the site of implantation and cellular origin of the implant, such as xenografts, allografts, ectopic, and orthotopic.

#### **a. Xenograft models**

These are the most common of all types of mouse model. In these mouse models, human tumor cells are ectopically or orthotopically introduced into immunodeficient mice. Due to their simplicity, these models are the preferred choice for testing of therapeutic HCC drugs candidates for treatment of HCCs. The ectopic method, which involves the sub-cutaneous injection of cancer cells in the mouse, is a straightforward method for drug testing and other investigations concerning treatment modalities. The increase in number of human HCC cell-lines and the advancement in technologies, has resulted this method to be the most successful one. Moreover, the xenograft model enables the measurement of several tumor characteristics, such as tumor length and width.(He et al., 2015)

The orthotopic xenograft models are initiated when tumor tissues of patients, diced into a cubic millimeter size, are implanted into livers of immune deficient mice livers. This approach is somewhat laborious and complicated, owing to the implantation into the liver tissue of mice. This method does not allow the direct measurement of parameters like tumor volume, length and breadth. However, this method is preferable in cases where the tumor microenvironment represents a significant aspect of the study. Newer technologies have been developed to implant the tumors and also monitor the progress of the tumors in the mouse liver. Technologies such as magnetic resonance imaging (MRI) and positron emission tomography (PET) can be used in the study of these tumors. Such techniques can also aid in the monitoring of metastasis of the tumors into other mouse organs. There have been great advancements lately in the

orthotopic xenograft models, which includes the invention of synthetic fibres or tubes to assist in the implantation of the tissues in the mouse livers (He et al., 2015).

### **b. Allograft models**

With the recent advancements in immune therapy against cancer, the allograft models for HCC have been developed, in order to facilitate studies of the action of the immune system. Hence these mouse models are predominantly on non-immune compromised mice, and the tumor cells are also derived from mice. This method denies the use of human HCC cell-lines or primary human cancer cells, due to the danger of immune reactions that are elicited upon exposure of mice to cells from a foreign species. Hence this model is limited to murine cancer cells implanted into other mice. However, this method has resulted in significant findings regarding the role of immune modulation in cancer, and has opened up various avenues of cancer therapy.

### **1.2.6.3 Transgenic mouse models**

Transgenic mouse models use genetically engineered mice, used for assessing the effects of specific oncogenes or a combination of various oncogenes and other factors in hepatocarcinogenesis. This type of mouse models has an additional emphasis on the genetic alterations associated with carcinogenesis. Using this type of mouse model, genes carrying specific mutations can also be inserted into the mouse genome. In the past few decades, numerous transgenic mouse models have been developed for HCC studies. Significant insight has also been attained about HCC pathophysiology using these transgenic mouse models.

#### **a. HBV & HCV transgenic mice**

Given that the major cause of HCC worldwide is HBV infection, transgenic mice expressing either the whole genome or parts of the genome of HBV have been established. The HBx gene, is known to regulate genes such as *c-fos*, *c-myc* and *ICAM-1* that are associated with cell adhesion and proliferation. The *HBx* gene has been genetically engineered into mice. *HBx* overexpressing transgenic mice showed unique cellular morphologies such as an increased nuclear/cytoplasmic ratio, megalocytosis,

nuclear pleomorphism and hyperchromatism. Tumors were observable after the mice were about one year old. Pre-natal complications and decreased birthrates were observed in these mice. *HBx* transgenic mice, upon DEN administration, were more prone to acquire liver tumors compared to the wild-type mice. Thus the *HBx* transgenic mice are a significant tool to study HBV-induced human HCC formation (Heindryckx et al., 2009; Li et al., 2011).

Similar to the *HBx* transgenic mice, the HCV transgenic mice, which expressed the core protein of HCV expressed in mice, developed liver tumors. These mice however were less prone than the HBV transgenic mice. This might be due to influences exerted by the genetic background of the mice or due to molecular mechanism under control of HCV core proteins in the cell. The HBV core protein activates peroxisome proliferators in hepatocytes, which interferes with lipid metabolism and subsequently causes lipid accumulation in hepatocytes. HCCs were formed in HCV-expressing transgenic mice, albeit much later than in HBV transgenic mice. HCV transgenic mice also showed a synergistic effect with DEN in the context of tumor formation (Wu et al., 2009).

### **b. Oncogene over-expressing transgenic mice**

Various transgenic mouse strains have been developed, which over-express certain oncogenes in order to study oncogene functions in hepatocarcinogenesis.

*c-Myc* is a transcription factor that regulates the expression of numerous genes by directly binding to regulatory DNA elements. The *Myc* transgenic mice developed tumors, which resemble closely those human HCCs that have a good prognosis. *c-Myc* interacts with multiple other proteins and influences various pathways of oncogenesis. It was also found that a small subset of *c-myc* transgenic mice displayed overexpression of  $\beta$ -Catenin protein (Heindryckx et al., 2009).

$\beta$ -Catenin transgenic mice are also available, which display overexpression of this protein or a mutant form thereof. The *CTNNB1* gene, along with *TP53*, constitutes the most frequently mutated gene in human HCC, with *CTNNB1* alone contributing to about 30% of the mutations. *Ctnnb1* mutations in mice are not sufficient to induce

tumors alone. *Cttnb1* mutations are often coupled with mutations of the *H-Ras* genes, which rapidly trigger aggressive tumors in mice (Heindryckx et al., 2009).

### **c. Transgenic mice expressing growth factors**

Transforming growth factor- $\alpha$  (TGF- $\alpha$ ) is an important molecule for the stimulation of hepatocarcinogenesis. Liver tumors occurring in TGF- $\alpha$  over-expressing mice resemble closely those human HCCs that are associated with very poor prognosis. These TGF- $\alpha$  overexpressing transgenic mice have a HCC penetrance of about 50%. These mice, when bred with *c-myc* transgenic, developed aggressive liver tumors in a short period of time. This is indicative of an underlying cross-talk to exist between different pathways for hepatocarcinogenesis. Similarly, Epidermal Growth Factor (EGF) transgenic mice show far more aggressive phenotypes when coupled with other transgene overexpressing mice (Heindryckx et al., 2009).

Fibroblast growth factor 19 (FGF19) transgenic mice were generated, which do not have a liver-specific expression of the transgene. *FGF19* expression in the skeletal muscle cells resulted in the induction of tumors in the liver at the age of about 1 year. Interestingly, it was reported that the female *FGF19* transgenic mice showed over-expression of  $\beta$ -Catenin while the male transgenic mice did not (Heindryckx et al., 2009).

The various mouse models displaying HCC formation have helped improve the understanding of HCC pathogenesis and the molecular mechanisms and transcription factors involved in malignant transformation.

## **1.3 Serum Response Factor**

Serum Response Factor (SRF) is a transcription factor involved in the transcriptional regulation of numerous target genes, which include the Immediate Early Genes (IEGs). Its discovery was initially triggered by the observation that addition of serum to quiescent cells *in vitro* results in immediate induction of *c-fos* gene transcription (Greenberg and Ziff, 1984). Further studies identified a specific sequence of nucleotides, located about 300bp upstream of the transcription start site (TSS) of the

c-fos gene, which was called the Serum Response Element (SRE) (Treisman, 1986). The sequence, also known as the CArG box, consisted of an A/T rich core flanked by an inverted repeat, CC(A/T)<sub>6</sub>GG. SRF was found to be the transcription factor that bound to the CArG box and controlled the expression of the corresponding gene. SRF is known now to regulate the transcription of approximately 1000 target genes (Esnault et al., 2014).

### 1.3.1 SRF Gene

The human SRF gene sequence is 10213 bp long and is located on chromosome 6p21.1, oriented in the plus strand. The SRF gene is highly conserved, and orthologs have been identified in other species such as *Mus musculus*, *Gallus gallus*, *Anolis carolinesis*, *Xenopus laevis* and *Danio rerio*. The human SRF gene contains 7 exons and transcribes to a 4202 base long transcript that encodes a protein of 508 amino acids in length. The SRF protein has a molecular mass of 51593 Da, and contains the following domains.

- Nuclear Localisation Sequence
- DNA binding domain
- Transcriptional activation domain

The DNA binding domain, spanning the amino acids 133 through 222 of the protein, is also responsible for dimerization and interaction of SRF with other co-factors, essential for the function of the protein. This domain is also known as the MADS box, and all transcription factors containing this domain have been grouped together as the MADS box family of transcription factors (Shore and Sharrocks, 1995).

### 1.3.2 DNA binding

The SRF protein binds to the CArG box DNA sequence in order to modulate the transcription of various target genes. The composition of the CArG box sequence is shown in the Figure 1-6.

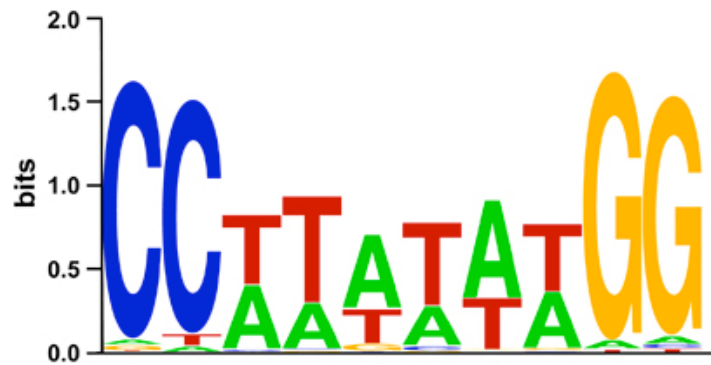


Figure 1-6: The CArG box sequence. The Y-axis denotes the conservation of the corresponding nucleotide stacks, measured in 'bits'. The height of each nucleotide at a given position represents its frequency. Picture source: (Miano, 2010)

The terminal G residues of the element have been shown to be the most significant nucleotides in the CArG box sequence, since here direct molecular interaction with SRF takes place. Mutations in these G residues could effectively deactivate the CArG elements. Insertions or deletions in the inner (A/T)<sub>6</sub> core of the CArG box sequence has also been able to prevent SRF binding, due to the disruption of the planar topology of the G residues, which normally reside in the major groove separated from each other by one helical turn, that enable the SRF monomer to be on the same planar face of the DNA, thereby enabling efficient binding (Miano, 2003). Substitutions in the (A/T)<sub>6</sub> core also seem to interrupt SRF binding, due to the presence of critical binding points in the (A/T)<sub>6</sub> core with the SRF protein, which are found within the narrowed minor groove of DNA. Moreover, an acute bend has also been observed in the DNA upon SRF binding (Pellegrini et al., 1995). This signifies the importance of the specific features of the CArG box in terms of its length and composition and structure, which bears importance for the action of SRF and its co-factors.

### 1.3.3 SRF co-factors

Apart from their binding to DNA, SRF is also known to bind to different co-factors. Depending upon the SRF co-factor involved, different cohorts of target genes are regulated. The best-studied interaction partners of SRF are the Ternary Complex Factors (TCFs) and the Myocardin-related transcription factors (MRTFs).

The TCFs were the first SRF cofactors discovered and they support SRF to regulate *c-fos* transcription (Shaw et al., 1989). They comprise of the proteins Ets like gene 1 (Elk1), SRF accessory protein 1 (SAP-1/Elk4) and Net (Elk3) (Dalton et al., 1993). They are activated by the MAPK signaling pathway. Each of the TCF proteins contains the ETS-DNA binding domain at the N terminal, close to the SRF binding domain. The highly conserved ETS-DNA binding domain is known to bind DNA at an ets-binding site (EBS), which is found 5' to the CArG box sequence of the *c-fos* promoter. The EBS and the CArG box jointly form the Serum Response Element (SRE).

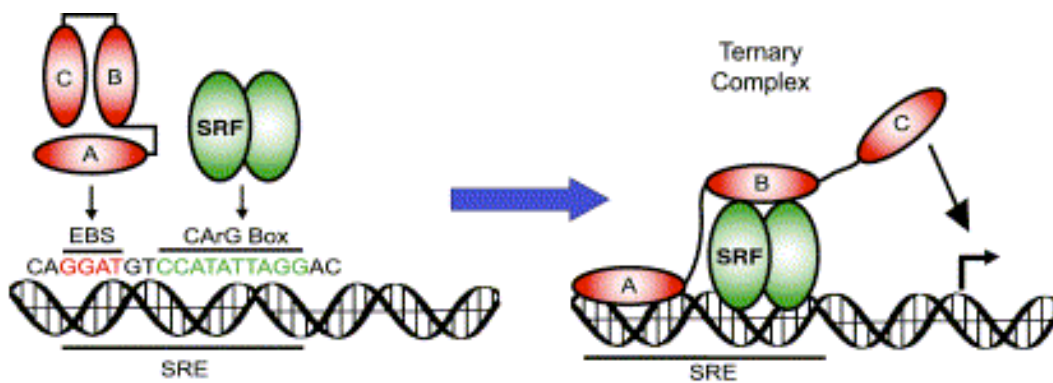


Figure 1-7 : SRF-TCF binding to DNA. The TCFs bind with DNA using the ETS Binding domain(A), simultaneously bound to SRF using the SRF binding domain (B), meanwhile the Transactivation domain (C) initiates transcription. Picture source: (Buchwalter et al., 2004)

The MRTF protein family consists of the proteins MRTF-A and MRTF-B. The MRTF proteins contain in their N-terminal region, a consensus sequence known as the RPEL-Motiv, which mediates MRTF interaction with globular actins (G-actins). MRTFs also contain an SRF-binding domain that is Glutamine-rich, and a Leucin Zipper



domain, which assists in the homo- or hetero-dimerization of the proteins. The MRTFs, themselves do not bind directly to the DNA, and hence they do not contain any DNA binding domain of their own.

In addition, other interaction partners of SRF have also been identified, which include GATA protein family of Zincfinger transcription factors, the homeodomainfactor Nkx2.5, the LIM-domain-containing Cysteine Rich Proteins (CRP), FHL2, Homeobox-Protein (HOP), Sp1, and ATF6, among others. These proteins are known to have either activating or inhibiting effects on SRF.

### **1.3.4 SRF activation and regulation**

SRF has been known to be activated by different agents such as serum, lysophosphatidic acid (LPA), anisomycin, mitogens, lipopolysaccharides, 12-O-tetradecanoylphorbol-13-acetate (TPA), cytokines, TNF $\alpha$ , Hepatitis B virus activator protein pX, activated oncogenes *v-src*, *v-ras*, *c-raf*, and also anti-oxidants, and UV light. However, the SRE regulating function of SRF is modulated by two major signaling pathways as shown in the Figure 1-9.

- MAPK signaling pathway
- Rho-actin signaling pathway

#### **1.3.4.1 MAPK signaling pathway**

The TCF-dependent transcription regulation by SRF is mainly activated through the MAPK cascade, which involves various protein groups such as the extracellular signal regulated kinases (ERK), comprising the c-Jun N-terminal kinases (JNK), the p38 kinases, and the ERK kinases. The TCFs are needed to be phosphorylated before binding to SRF and forming a complex, and then together they regulate the cohort of target genes, which have the SRE sequence 5' to their TSS. The TCFs can be phosphorylated by the kinases ERK, JNK, or p38 kinases in response to different upstream cues, as denoted in Figure 1-8.

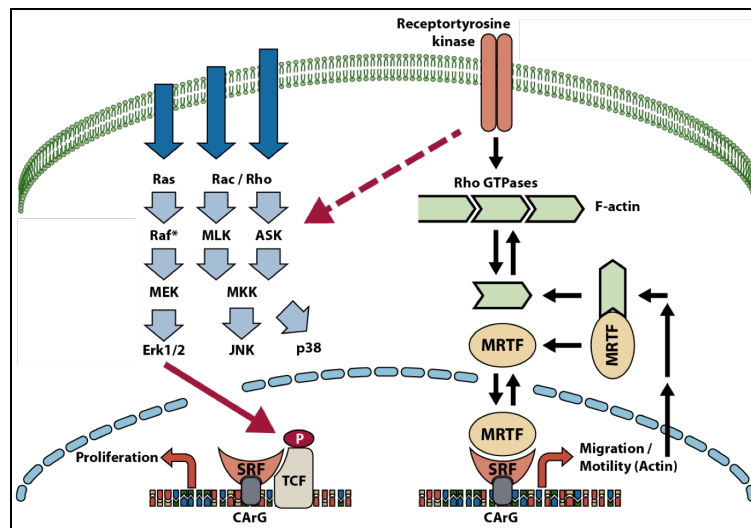


Figure 1-8: Activation of SRF. The figure shows the two major signaling pathways that stimulate SRF via activation of SRF partner proteins. (Picture from: Prof. Dr. Alfred Nordheim)

### 1.3.4.2 Rho-Actin signaling pathway

The MRTF-dependent transcriptional regulation by SRF is activated through the Rho-Actin signaling pathway, as shown in the Figure 1-8. This pathway is largely modulated by the amount of G-actin present in the cytoplasm. G-actin is known to bind MRTFs in the cytoplasm. It has been observed that a decrease in the amount of cytoplasmic G-actin in the cytoplasm, as generated in response to an increased synthesis of F-actin in the cytoplasm, results in increased nuclear localization of MRTFs (Miralles et al., 2003). This implies that the dissociation of G-actin from the MRTFs enables MRTF nuclear localization. Following nuclear localization, MRTFs immediately bind to SRF and regulate a specific cohort of target genes, which are dependent on this pathway for regulation (Olson and Nordheim, 2010).

### 1.3.5 SRF biological function

Various genes associated with proliferation and differentiation are regulated by SRF. The role of SRF in the transcriptional regulation of the IEGs has been proven by a study where SRF-deficient Embryonic Stem (ES) cells did not stimulate IEG activation

upon serum induction (Schratt et al., 2001). However a similar experiment in rat smooth muscle cells showed an increase in proliferation (Kaplan-Albuquerque et al., 2005). In human cells, the inhibition of SRF resulted in reduced proliferation (Werth et al., 2010). Thus, roles of SRF might vary among different cell types and organisms. Constitutive SRF-knockout showed that murine embryos lacking SRF failed to develop mesoderm and hence did not proceed beyond the gastrulation stage (Arsenian et al., 1998). This supported earlier studies where an inhibition of the SRF function in vitro showed a complete blockage of the differentiation of myoblasts to myotubes, which underscored the importance of SRF during differentiation, in a non-cell autonomous manner (Soulez et al., 1996). SRF knockout in ES cells was also shown to affect ES cell adhesion and migration potential, which are dynamic events essential during gastrulation, which might explain the previous observations of embryo death at the gastrulation stage (Schratt et al., 2001)

SRF depletion in cardiomyocytes has been observed to have lethal effects on mice, at both embryonal and adult stages, thus signifying the importance of SRF in both heart development and function (Parlakian et al., 2005; Parlakian et al., 2004). Interestingly, an overexpression of SRF through specific cardiac transgenesis resulted in hypertrophic cardiomyopathy and expression of antisense SRF improved the mouse cardiac performance (Chai and Tarnawski, 2002).

SRF depletion in the neurons of mice showed migration defects and motor dysfunction (Ataxie), ultimately resulting in the death of the mice within 3 weeks after birth (Alberti et al., 2005). Other studies have shown the importance of SRF in the development and function of mouse retina, smooth muscle cells, skeletal muscle cells, immune cells, pancreas, brain and liver. A list of cell-type specific SRF knockout experiments in mice and the corresponding phenotypic outcome is provided in Table 1-1.

## INTRODUCTION

*Table 1-1: Murine cell type-specific Srf knockouts and corresponding phenotypic effects. Table from: (Miano, 2010)*

<b>Target cell type</b>	<b>Phenotype</b>
Cardiomyocyte/SMC	Dilated heart/reduced SMC recruitment and embryo demise at day e10.5
Embryonic cells	Early embryonic lethality without developing to term
Adult cardiomyocytes	Heart failure and death 10 weeks after KO
Endothelial cell	Aneurysm, hemorrhage, EC junctional/actin cytoskeletal defects, with death at day e14.5
Skeletal muscle cell	Perinatal orthopnea with thinned myofibers
Adult skeletal muscle	Muscle hypotrophy and poor regeneration, with 50% mortality by 6 weeks of age
Hepatocyte	Reduced body size and chronic liver regeneration
All pancreatic cells	Pancreatitis, loss of exocrine pancreas
Adult SMC	Chronic intestinal pseudo-obstruction and death by 3 weeks after deletion of SRF
Forebrain neurons	Defective neuronal migration to hippocampus, impaired actin dynamics and axonal guidance. Attenuated activity-dependent immediate-early gene expression and long-term synaptic potentiation.
Hippocampal neurons	Defective layering of neurons and dendritic branching
Keratinocytes	Psoriasis-like condition with epidermal hyperproliferation, cytoskeletal and cell–cell contact perturbations Thickened epidermis, loss of barrier function and cell–cell contacts due to abnormal cytoskeleton
T lymphocytes	Loss in circulating T lymphocytes
B lymphocytes	Decrease in marginal zone B cells, low IgM
Megakaryocyte	Thrombocytopenia and macrothrombocytopenia

### 1.3.6 SRF and Cancer

Considering the fact that SRF regulates the transcription of genes related to cellular proliferation and differentiation, the role of SRF in carcinogenesis has been studied with intense interest. Different studies reported increased levels of SRF in certain human cancers (Choi et al., 2009; Park et al., 2007). However, whether the SRF overexpression in human tumors is a cause or a consequence of the progression of the cancer, was not clear. A role of SRF in the regulation of Hippo signaling pathway, through TAZ in breast cancer cells has been reported (Liu et al., 2016). There have been other reports about the role of SRF in enhancing the invasiveness of prostate cancer cells. Inhibition of SRF resulted in decreased proliferation in those cells (Hermann et al., 2016). Angiogenesis and increased cell migration during metastasis of certain cancers have also been attributed to SRF (Franco et al., 2013; Kircher et al., 2015). SRF and Notch signaling pathways were found to be active in highly motile melanoma cells.(Manning et al., 2015). MRTF-depleted and SRF-depleted breast cancer cells were reported to be less invasive and the xenografts in mice were unable to metastasize to the lung, thereby implying a fundamental role of SRF in cancer metastasis and invasiveness (Medjkane et al., 2009).

Another study described liver-specific downregulation of the microRNA miR-122. The downregulation of miR-122 corresponded with SRF overexpression in these tumors (Bai et al., 2009). This observation was further substantiated by the fact that SRF is a target of miR-122 and introduction of miR-122 in the cancer cells resulted in reduced tumorigenesis. Moreover, a truncated SRF gene lacking the miRNA-binding site, when introduced in the miR-122-expressing cells, the anti-proliferative effect had disappeared. This study showed a clear correlation between SRF and HCC (Bai et al., 2009).

SRF has also been associated with induction of EMT in HCC (Park et al., 2007). The fact that the Hepatitis B virus core protein HBx can activate SRF suggests a fundamental role of SRF in liver carcinogenesis.

### 1.3.7 SRF-VP16

SRF-VP16 is a fusion protein, where the transcription activation domain (TAD) of the human *SRF* cDNA, spanning exons 412 to 508 is replaced with the sequence coding for the viral protein VP16, derived from the Hepes Simplex virus. The scheme of the SRF-VP16 sequence is shown in Figure 1-9.

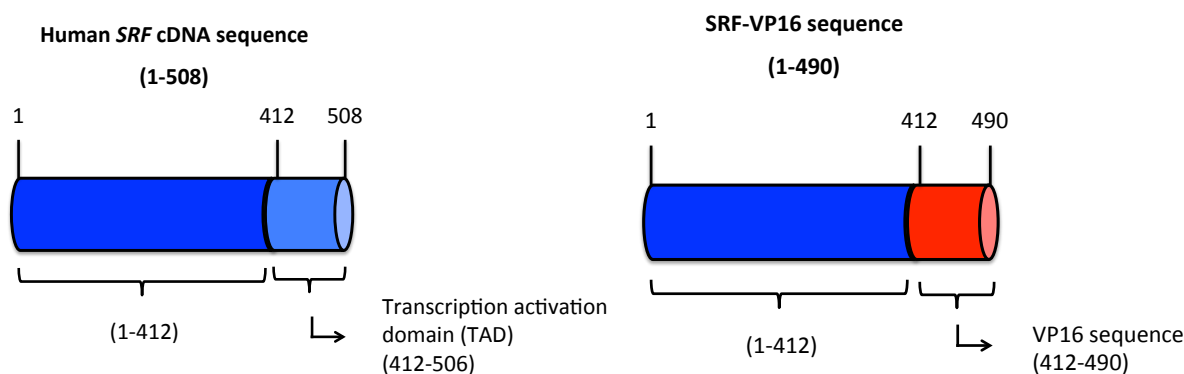


Figure 1-9: SRF-VP16 sequence. The figure shows the replacement of the terminal human SRF transcription activation domain (TAD) with the TAD of the VP16 protein (Picture from: Prof. Dr. Alfred Nordheim).

The protein SRF-VP16, due to the presence of the transactivation domain of the VP16 protein can perform the transcription function of SRF even without the upstream cues otherwise necessary to activate SRF. Similarly, SRF-VP16 function does not depend upon the interaction with partner proteins like TCFs or MRTFs. The SRF-VP16 fusion protein had been used widely in the study of the SRF function (Dalton and Treisman, 1992). The SRF-VP16 protein was shown to induce morphological changes in *Srf* (-/-) embryonic stem cells, and induce the transcription of *SRF* target genes (Schratt et al., 2002). SRF-VP16 also was shown to mimic SRF function in the neuronal cells by mediating their survival (Stern et al., 2012). Thus, SRF-VP16 serves as a constitutive active form of SRF carrying out the transcription regulatory function of SRF in different cell types.

### 1.3.7 *SRF-VP16<sup>iHep</sup>* mouse model

The *SRF-VP16<sup>iHep</sup>* mouse model was generated by Dr. Stefan Ohrnberger in the group of Prof. Dr. Alfred Nordheim, University of Tuebingen. The *SRF-VP16<sup>iHep</sup>* mice show mosaic activation of SRF-VP16 in hepatocytes. The expression of SRF-VP16 subsequently results in hepatocarcinogenesis in the livers of *SRF-VP16<sup>iHep</sup>* mice (Ohrnberger\*, Thavamani\* et al., 2015). The SRF-VP16 triggered HCC formation is accompanied by induction of SRF target genes and hyper proliferation of SRF-VP16-expressing hepatocytes.

## 1.4 Objective of this study

The specific aims of the current study were to characterize the SRF-VP16 triggered murine hepatocellular carcinoma (HCC) formation, using a combined multi-omics approach. It was thus the intention of identifying crucial events during the malignant transformation and subsequent HCC tumor progression, as occurring with 100% penetrance in *SRF-VP16<sup>iHep</sup>* mice. To address this, the following strategy was undertaken.

- 1) Identification of the genomic changes contributing to hepatocarcinogenesis was attempted by detecting
  - a. mutations in the genomic hotspot regions of mutations, using Sanger's sequencing, and
  - b. genome-wide nucleotide variations, using whole exome sequencing (collaboration with DKFZ, Heidelberg).
- 2) Charting alterations at the transcriptome level contributing to cancer formation in liver by using RNAseq technology and microarray technology (in collaboration with MPI, Tuebingen and UKT, University of Tuebingen respectively, bioinformatic support by QBiC). Individual candidate genes were to be tested using quantitative real time PCR.

- 3) Determine protein expression changes and posttranslational protein modifications (phosphorylation) by using LC-MS/MS proteomics (in collaboration with the Proteome Center Tuebingen(PCT); Prof. Macek.
- 4) Search for metabolic alterations by analyzing the metabolome and lipidome (in collaboration with MetaSysX GmbH, Potsdam).
- 5) Comparison of the obtained data to each other and to human HCCs.
- 6) The ultimate goal being to deduce mechanistic principles involved in hepatocarcinogenesis in *SRF-VP16<sup>iHep</sup>* mice, as revealed by different omics technologies.

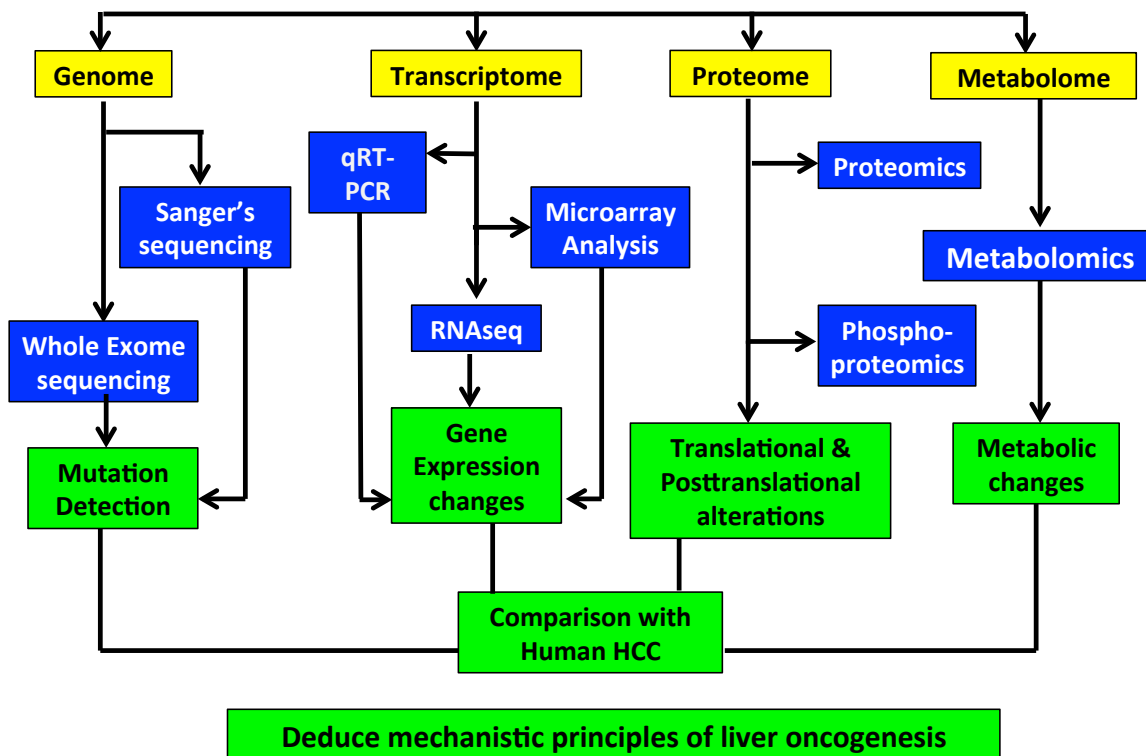


Figure 1-10: Objective of study described. The yellow boxes represent the components to be investigated. Blue boxes represent the various experimental techniques to be used. The green colored boxes represent the scientific insights aimed to be obtained from the experiments.



## MATERIALS AND METHODS

---

## 2 MATERIALS AND METHODS

---

### 2.1 Animal Experiments

#### 2.1.1 Maintenance of experimental mice

Mice were kept according to the regulations of legal animal protection laws in the Animal House present in the Interfaculty Institute for Cell Biology (Auf der Morgenstelle 15, University of Tübingen). The recommendations of the “Society for Laboratory Animal Science” (Gesellschaft für Versuchstierkunde) with regard to the holding of laboratory mice mentioned in the version-dated 20.03.2007 have been duly followed. The recommendations were based on the principles mentioned in Appendix A of Europarates ETS 123, entitled - “To the European Convention for the Protection of Vertebrate Animals used for Experimental and other Scientific Purposes. Guidelines for Accommodation and Care of Animals“.

The day/night rhythm followed a 12-hour cycle, which lasted, from 6:00 AM to 6:00 PM of ‘daytime’ with light and 6:00 PM to 6:00 AM of ‘night-time’ without any light. The mice had continuous access to water and dried food pellets. The humidity of the rooms was maintained at 55%(± 10%) , the temperature at 22 °C (± 2 °C) and the air flow at 10 times the room-volume per hour. The animals were maintained in groups of three to five in Type II long cages. Matings were also set up in Type II long cages, and were checked frequently. All mice were transferred to new cages, once in every week. The bottom of each of the cages was covered with wooden chips. Additionally the mice cages were supplied with cellulose material in the form of recycled paper for nest building. In all the cages, and additional play toy, shaped as a tunnel, made out of wood or plastic was furnished, in order to keep the mice occupied. Once in every six months, hygiene control tests were performed on randomly chosen mice, where the presence of different mouse pathogens was checked.

### 2.1.1.1 Permissions for mouse experiments

The mouse experiments carried out as part of this dissertation were duly approved by the corresponding authorities at the district administration (Regierungspräsidium) of Tübingen.

- I. The first part of the experiments were performed in accordance to the permission as stated in Paragraph 4 of the Animal protection laws in the version announced on 25<sup>th</sup> May 1988 (BGBl. S: 1105), which enables the sacrifice of mice, without any prior experimental treatments for organ preparation only.
- II. Upon changes in the animal protection laws that became active starting from 4<sup>th</sup> July 2013 the newer version regarding usage of laboratory animal for experimental purposes, from 12<sup>th</sup> August 2013 (BGBl. I S. 3125), a new application IM 1/14 was made on 18.11.2014 for “Experimental analysis at the RNA and protein levels in hepatocytes of the SRF-VP16<sup>iHEP</sup> mouse model“ and was approved on 05.01.2015.
- III. Abhishek Thavamani, was added to the application for animal experimentation IM 1/14 as experimenter, upon successful completion of the Animal handling course – “Versuchstierkundlichen Privatissimums” in the summer semester 2015.

### 2.1.2 Mouse breeding

Breeding of the transgenic mice was carried out as described below. *SRF-VP16(+/-):Alfp-CreER<sup>T2</sup>(-/-)* mice were mated with *SRF-VP16(-/-):Alfp-CreER<sup>T2</sup>(+/-)* mice. The mice were put together in Type II long cages for mating. The mating cages were held in rows separated from the other cages. The mating pairs were at least 6-7 weeks of age. The first morning after the setting up of the mating, the mice were checked for compatibility. In case, there were found bloodstains in either of the mice, or that they were continually fighting with each other, the pairs were separated and mated with other mice.

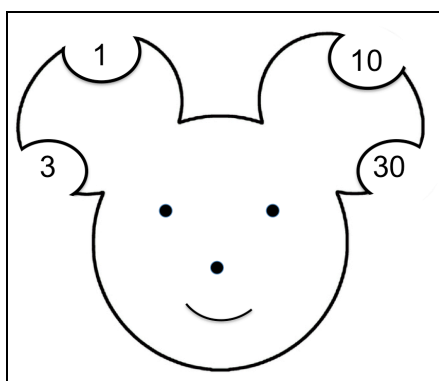
The mating pairs were monitored every morning for the presence of vaginal plugs. When a litter was born, the pups were monitored each day for the first few days to ensure the mothering instincts and proper nursing of the pups.

When the pups were 3 weeks old, they were weaned by transferring them into new cages, males and females separately.

Mating pairs were allowed to mate for about 5-6 months, until they had given birth to about 5-6 litters. Subsequently, mating pairs were separated into different cages, and a new set of animals was chosen as mating pairs. After every third generation of crossing, transgenic mice were crossed with wildtype C57BL/6 mice.

### 2.1.3 Mouse ear marking

When the pups had been weaned and separated into new cages, the ear marking and genotyping was done to uniquely identify them. The young mice were gently taken by their tails, and held by the skin behind their necks to make them motionless. Following this, a small hole was made in their ear based on a pre-determined scheme to designate the mouse with a unique code (Figure 2-1). The piece of skin thereby removed from the ear was immediately incubated in lysis buffer, and the DNA used for genotyping PCR. The mice were subsequently weighed and then carefully let into their cages back again. The unique identification codes corresponding to the position of ear punch is shown in Figure 2-1.



*Figure 2-1: Ear Marking Scheme*

## 2.1.4 Genotyping

### 2.1.4.1 Sample Lysis

#### Materials Used

Buffer/Instrument	Company/ Composition
Lysis Buffer	10 mM Tris-HCL, pH 8.3 200 mM NaCl 5mM EDTA 0.2% SDS and 200 µg/mL Proteinase K (freshly added each time before use)
Heating Block	Eppendorf GmbH

The ear skin samples of the mice were incubated in 100 µL of Lysis Buffer overnight at 55 °C. After that, they were briefly vortexed and then heated up at 99 °C for 10 minutes to inactivate the Proteinase K enzyme. The samples were then centrifuged at high speed for 10 minutes. The DNA released into the supernatant was taken as template for the genotyping PCR reactions.

### 2.1.4.2 Genotyping PCR

The DNA template was mixed with the PCR cocktail and set up in the thermocycler for PCR amplification using the corresponding primers. The amplified PCR products were subsequently analysed using Agarose Gel Electrophoresis.

**Materials used**

<b>Buffer/Instrument</b>	<b>Company/ Composition</b>
10x Polymerase Buffer	Thermo Scientific™
Primers	Sigma-Aldrich
DMSO	Sigma-Aldrich
40mM dNTPs	Promega GmbH (dATP,dCTP,dGTP and dTTP 10mM each)
Taq Polymerase (2 units)	Thermo Scientific™
Thermocycler	Biometra

**2.1.4.3 Genotyping of *SRF-VP16***

The presence of the *SRF-VP16* sequence at the *Rosa26* locus in the genome of the mice was checked using the following primers during genotyping PCR. The locations of the primers are shown in the Figure 2-2.

**Primers**

A multiplex PCR reaction using 3 primers was used for genotyping the *SRF-VP16* transgene insertion. The locations of the primers are shown in Figure 2-2. The primers used in the multiplex were as follows.

- i. *Rosa26* forward primer
- ii. *Rosa26* reverse primer
- iii. PURO reverse primer

*Rosa26* forward primer and the *Rosa26* reverse primer were specific to the wild type *Rosa26* locus of the mouse genome. They both combine to amplify a 275 bp fragment corresponding to the murine wild type *Rosa26* gene locus. The PURO reverse

primer is specific to the *Puromycin resistance gene* present in the STOP cassette of the inserted *SRF-VP16* transgene construct. The PURO reverse primer along with the Rosa26 forward primer amplify a fragment of 500 bp size, which occurs only upon the presence of *SRF-VP16* transgene construct in the mouse genome, which implies a successful insertion of the transgene construct at the Rosa26 locus. The primer mix was prepared by mixing 5  $\mu$ L of 100M Rosa26 forward primer, 10  $\mu$ L of 100  $\mu$ M Rosa26 reverse primer and 10  $\mu$ L of 100  $\mu$ M PURO reverse primer. The sequences of the primers are given in section 2.10.

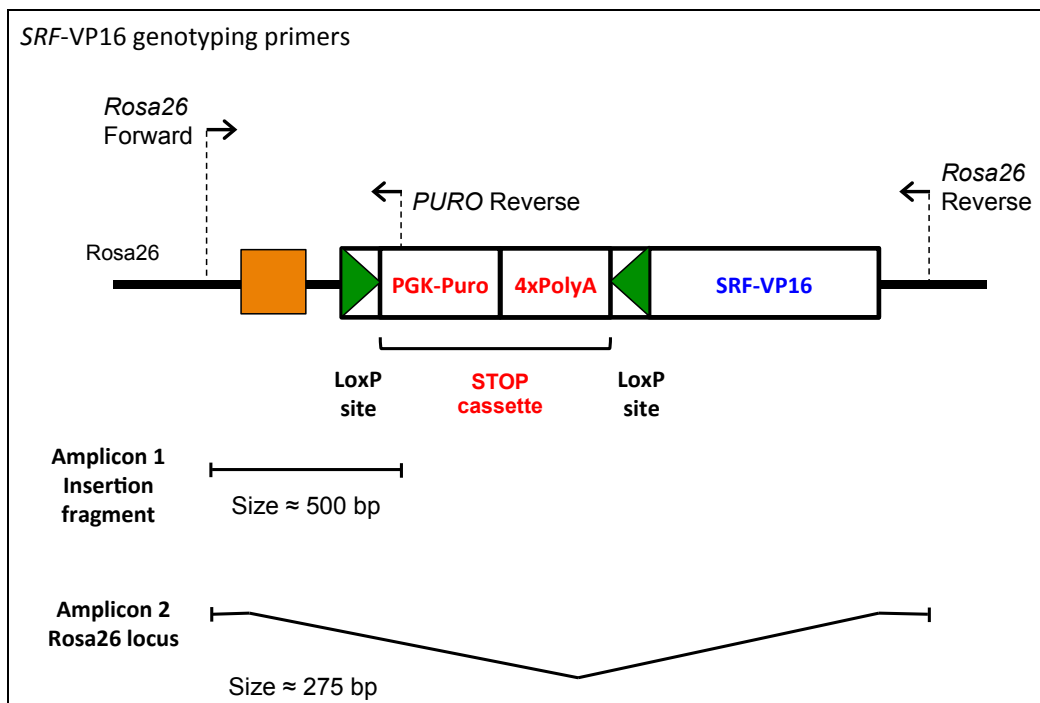


Figure 2-2: *SRF-VP16* genotyping primers. The figure shows the positions of the forward and reverse primers corresponding to the 2 different amplicons of the *SRF-VP16* genotyping PCR. The genomic sequence of the *SRF-VP16* transgene is explained in section 3.1.1.1. The primer sequences are described in section 2.10.

**PCR reaction mix**

The PCR cocktail for each genotyping reaction consisted of following:

<b>Buffer</b>	<b>Concentration</b>	<b>Volume (µL)</b>
Polymerase Buffer	10x	5
Primer mix	100x	0.5
DMSO	8x	4
dNTP mix	40mM	2
Taq Polymerase	2 units	0.5
Water		35.5
Template DNA		1
Total Volume		50

The above cocktail was mixed gently and then placed in the thermocycler and the *SRF-VP16 genotyping* PCR program was run.

***SRF-VP16* genotyping PCR program**

<b>Step</b>	<b>Temperature</b>	<b>Time</b>	<b>Repeat</b>
Lid pre-heating	100 °C		
Initial denaturation	98 °C	3 mins	
Denaturation	98 °C	30 secs	35 times
Annealing	62 °C	35 secs	
Extension	72 °C	70 secs	
Final Extension	72 °C	15 mins	
Storage	4°C	Hold	



#### 2.1.4.4 Genotyping of *Alfp-CreER<sup>T2</sup>*

The presence of the *Alfp-CreERT2* sequence in the genome of the mice was checked using the following primers during genotyping PCR. The locations of the primers are shown in the Figure 2-3.

##### Primers

A multiplex PCR reaction using 4 primers was used for genotyping the *Alfp-CreER<sup>T2</sup>* transgene insertion. The primers used in the multiplex were as follows.

- i. *Alfp-CreER<sup>T2</sup>* Forward
- ii. *Alfp-CreER<sup>T2</sup>* Reverse
- iii. PCR<sub>3</sub>
- iv. PCR<sub>4</sub>

*Alfp-CreER<sup>T2</sup>* forward primer and the *Alfp-CreER<sup>T2</sup>* reverse primer were specific to the *Alfp-CreER<sup>T2</sup>* transgene. They both combine to amplify a 200 bp fragment corresponding to the *Alfp-CreER<sup>T2</sup>* transgene insert. The presence of the 200 bp amplicon denoted the presence of the *Alfp-CreER<sup>T2</sup>* transgene in the mouse genome. In addition to the *Alfp-CreER<sup>T2</sup>* transgene primers, the primer pair PCR<sub>3</sub>/ PCR<sub>4</sub> was also taken in the multiplex, which amplifies a random fragment located elsewhere in the mouse genome, as an internal control. The PCR<sub>3</sub>/PCR<sub>4</sub> primer pair resulted in an amplicon of 500 bp size. The presence of this 500 bp amplicon ensured that the PCR reaction had been successful. The primer mix was prepared by mixing 10  $\mu$ L of 100 $\mu$ M *Alfp-CreER<sup>T2</sup>* forward primer, 10  $\mu$ L of 100 $\mu$ M *Alfp-CreER<sup>T2</sup>* reverse primer, 10  $\mu$ L of 100 $\mu$ M PCR<sub>3</sub> primer and 10  $\mu$ L of 100 $\mu$ M PCR<sub>4</sub> primer. The sequences of the primers are given in section 2.10

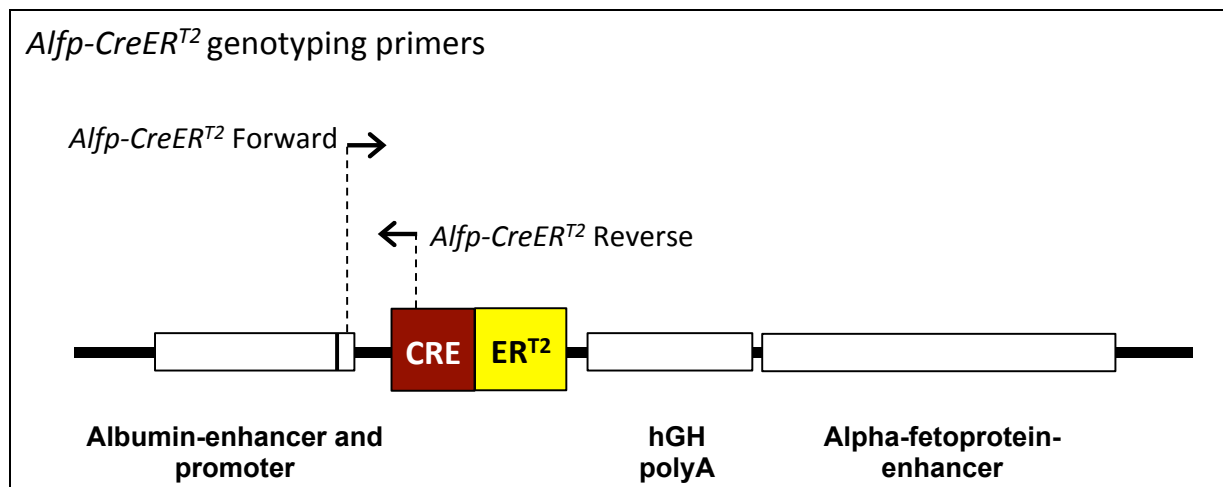


Figure 2-3: *Alfp-CreER<sup>T2</sup> genotyping primers.* The figure shows the positions of the forward and reverse primers used in the detection of *Alfp-CreER<sup>T2</sup> genotyping PCR.* The genomic sequence of the *Alfp-CreER<sup>T2</sup> transgene* is explained in section 3.1.1.2. The primer sequences are described in section 2.10.

### PCR reaction mix

The PCR cocktail for each genotyping reaction consisted of following:

Buffer	Amount/ Concentration	Volume ( $\mu$ L)
Polymerase Buffer	10x	5
Primer mix	100x	0.5
DMSO	5x	2.5
dNTP mix	40mM	1.5
Taq Polymerase	3 units	0.65
Water		38.85
Template DNA		1
Total Volume		50

The cocktail was mixed gently and then placed in the thermocycler and the *Alfp-CreER<sup>T2</sup> genotyping PCR* program was run.

***Alfp-CreER<sup>T2</sup>* genotyping PCR program**

Step	Temperature	Time	Repeat
Lid pre-heating	100 °C		
Initial denaturation	98 °C	1 min	
Denaturation	98 °C	10 secs	35 times
Annealing	55 °C	30 secs	
Extension	72 °C	20 secs	
Final extension	72 °C	10 mins	
Storage	4 °C	Hold	

**2.1.4.5 Genotyping of *Srf-flex1* transgene**

The presence of the *Srf-flex1* transgene, which contains two LoxP sites within the murine *Srf* gene, was checked using the following primers during genotyping PCR. The locations of the primers are shown in the Figure 2-4.

**Primers**

A multiplex PCR reaction using 3 primers was used for genotyping the *Srf-flex1* transgene insertion. The locations of the primers are shown in Figure 2-4. The primers used in the multiplex were as follows.

- i. Srf-KO-8 forward primer
- ii. Srf-KO-8 reverse primer
- iii. Lox2 primer

Srf-KO-8 forward primer and the Srf-KO-8 reverse primer were designed to amplify a fragment within the intron 1 of *Srf* gene in the mouse genome. In the wild type situation, the primers combine to produce an amplicon of 190 bp in length. In the *Srf-flex1* transgene, there exists a LoxP site in intron 1 of *Srf* gene sequence within the

## MATERIALS AND METHODS

amplifying region of the Srf-KO-8 forward and reverse primers. This results in a larger amplicon of 244 bp in length. Hence, during genotyping PCR, the presence of the 244 bp sized amplicon implies the insertion of the *Srf-flex1* transgene. The third primer in the multiplex the lox2 primer, is located on the other loxP site present in the *Srf-flex1* transgene. The second LoxP site is present at the exon 1 region of *Srf* sequence. In the event of recombination-mediated deletion between the LoxP sites, the lox2 primer combines with Srf-KO-8 reverse primer to generate an amplicon of 110 bp in length. The primer mix was prepared by mixing 15  $\mu$ L of Srf-KO-8 forward primer, 5  $\mu$ L of Srf-KO-8 reverse primer and 10  $\mu$ L of Lox2 primer. The sequences of the primers are given in section 2.10 .

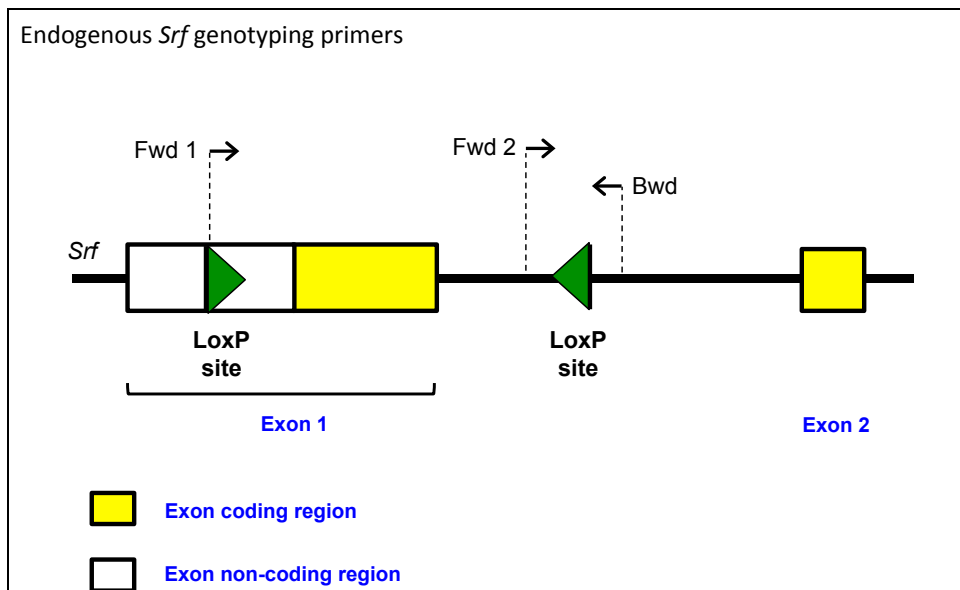


Figure 2-4: *Srf-flex1* transgene genotyping primers. The figure shows the positions of the forward and reverse primers used in the detection of *Srf-flex1* genotyping PCR. The genomic sequence of the *Srf-flex1* transgene is explained in section 3.1.1.3. The primer sequences are described in section 2.10.

### PCR reaction mix

The PCR cocktail for each genotyping reaction consisted of following:

Buffer	Concentration	Volume ( $\mu$ L)
Polymerase Buffer	10x	5
Primer mix	100x	0.5
DMSO	7x	3.5
dNTP mix	40mM	2
Taq Polymerase	5 units	1
Water		37
Template DNA		1
Total Volume		50

The cocktail was mixed gently and then placed in the thermocycler and the *Srf-flex1* genotyping PCR program was run.

### *Srf-flex1* genotyping PCR program

Step	Temperature	Time	Repeat
Lid pre-heating	100 °C		
Initial denaturation	98 °C	2 mins	
Denaturation	98 °C	30 secs	35 times
Annealing	60 °C	30 secs	
Extension	72 °C	45 secs	
Final extension	72 °C	10 mins	
Storage	4 °C	Hold	

### **2.1.4.6 Agarose gel electrophoresis**

In order to analyze the genotyping PCR products, 2% agarose gels were cast. The agarose gel electrophoresis was performed as described in the Detailed Procedure section 2.8.8. The images were then analyzed to determine the genotype of the mice.

### **2.1.5 Mouse health monitoring**

The experimental mice of the desired genotype were monitored closely, to ensure their proper health during the entire experimental period. Animal weight was measured every week, until mice reached the age of twenty weeks. Thereon, they were weighed twice a week until the age of thirty weeks. Thereafter, they were weighed every day until the day of sacrifice. Mice were also checked every day for any abnormalities in their posture, movement or eating and drinking habits.

### **2.1.6 Mouse sacrifice**

The experimental animals were sacrificed using CO<sub>2</sub> gas. Efforts were taken to ensure that the animal experiences the least stress during the entire procedure. Hence, the food, water and cage conditions of the mouse were maintained as mentioned in Section 2.1.1.

The cage containing the mouse was placed into a gas chamber and the lid was closed properly. CO<sub>2</sub> gas was released gently into the gas chambers. When the mouse stopped breathing, the CO<sub>2</sub> gas input was maintained for about 1 minute, to ensure that the mouse had died. Immediately, the body was taken for organ preparation.

#### **2.1.6.1 Mouse liver preparation**

The body weight of the mouse was measured before proceeding with the organ preparation. The corpse was sterilized with 70 % Ethanol, following which it was carefully dissected and the liver removed. The removed liver was rinsed in PBS and then the liver weight was measured. In case of tumor-bearing mice, the tumors were sliced

out of the liver and stored. The hyper proliferative nodule part of the liver was also stored. The liver weight to body weight ratios of the mice was then calculated.

### **2.1.6.2 Sample storage**

The samples collected from the harvested liver were stored appropriately for subsequent downstream analyses. For RNA extraction, the samples were stored in RNAlater and snapfrozen in liquid nitrogen. Samples required for protein extraction and for metabolomics analysis were placed in microcentrifuge tubes and snapfrozen immediately in liquid nitrogen. All the samples were later transferred to a -80 °C freezer for storage.

## **2.2 DNA analysis**

### **2.2.1 DNA extraction**

DNA extraction from murine liver tissue was carried out in conjunction with RNA extraction. The Qiagen Blood and Tissue kit was used for DNA extraction. From the tissue samples being processed for RNA extraction, the flow-through from the first centrifugation step of RNA extraction was used for DNA extraction. The detailed procedure is described in the Detailed Procedure section 2.8.2.

### **2.2.2 Mutation detection**

The extracted DNA was used for detection of mutations in the *Cttnb1* and *Trp53* gene. To this end, the genomic hotspot regions of mutations in the *Cttnb1* and *Trp53* genes were amplified first.

#### **Primer sequences**

The primer sequences used for the amplification of the corresponding amplicons are as follows. The primer sequences were obtained from Dr. Albert Braeuning.

**Materials used**

<b>Buffer/Instrument</b>	<b>Company/ Composition</b>
5X Phusion Buffer	Thermo Scientific™
Primers	Sigma-Aldrich
dNTPs	Promega GmbH (dATP,dCTP,DGTP,dTTP)
Phusion DNA polymerase	Thermo Scientific™
Thermocycler	Biometra

**PCR reaction mix**

The reaction mix for each PCR reaction consisted of following:

<b>Buffer</b>	<b>Amount/ Concentration</b>	<b>Volume (<math>\mu</math>L)</b>
Polymerase Buffer	5x	10
Forward Primer	0.2 $\mu$ M	2.5
Reverse Primer	0.2 $\mu$ M	2.5
dNTP mix	40mM	1
Taq Polymerase	2.5 units	0.5
Water		31
Template DNA		2.5
Total Volume		50

The cocktail was mixed gently and then placed in the thermo cycler and the corresponding PCR program was run.



***Ctnnb1* and *Trp53* DNA hotspot region amplification program**

Step	Temperature	Time	Repeat
Lid pre-heating	100 °C		
Initial denaturation	98 °C	30 secs	
Denaturation	98 °C	10 secs	30 times
Annealing	60 °C for <i>Trp53</i>	20 secs	
	60 °C for <i>Ctnnb1</i>		
Extension	72 °C	30 secs	
Final extension	72 °C	5 mins	
Storage	4 °C	Hold	

**2.2.2.1 Agarose gel electrophoresis**

A part (10 µL) of the amplification reaction from each sample was mixed with 6x (2 µL) of loading dye and was loaded into 2 % agarose gels. The agarose gel electrophoresis was performed as described in detail in the Detailed Procedure section 2.8.8.

**2.2.2.2 PCR product purification**

The PCR amplicons were then purified using the GeneJET PCR purification kit by following the instructions provided in the kit handbook. The detailed procedure is described in Detailed Procedure section 2.8.7.

**2.2.2.3 Sanger sequencing**

The purified amplicons were then sent to LGC Genomics GmbH for sequencing. 150 ng of the purified PCR amplicons was sent along with 5 pmols of the sequencing

primer (4  $\mu$ L). The results obtained were analyzed using Chromas Lite software to identify mutations in the loci.

## 2.3 RNA analysis

### 2.3.1 RNA extraction

Tissue samples stored earlier in RNAlater were subjected to RNA extraction, which was performed using the Qiagen “RNEasy Mini” kit, as per the manufacturer’s instructions. The detailed procedure is described in the Detailed Procedure section 2.8.1. Briefly, the tissue samples were lysed and RNA was extracted from the lysate using the spin columns provided by the kit. The RNA bound to the columns was washed and eluted in Elution Buffer (EB) provided by the kit. The concentration of the eluted RNA was measured using the Nanodrop spectrophotometer instrument.

### 2.3.2 Reverse transcription

#### Circular DNA synthesis

1  $\mu$ g of RNA was used for cDNA synthesis for further downstream processes. The Promega Reverse Transcription kit was used for cDNA synthesis. The procedure is described in the Detailed Procedure section 2.8.6. Briefly, the RNA was incubated with the cDNA synthesis cocktail, consisting of random hexamers, reverse transcriptase enzyme, RNase inhibitor, enzyme buffer, and dNTPs at optimum temperature for a specified amount of time according to the protocol.

### 2.3.3 Quantitative PCR for RNA expression analysis

#### Materials used

Buffer/Instrument	Company/ Composition
SYBr Green Mastermix	Roche
Primers	Sigma-Aldrich
Real Time PCR system	Applied Biosystems

#### 2.3.3.1 Primer design

Expression analysis of candidate genes was performed using quantitative real time PCR (qPCR). To this end, primers targeting the transcripts of each of the candidate genes were designed using the PrimerBlast online tool. The default parameters of primer designing, suggested by the tool were unchanged. Primers were also designed for house keeping genes for normalization of the qPCR reactions later on.

#### 2.3.3.2 qPCR reaction mix

The quantitative RT-PCR was performed based on SYBR green chemistry. The reaction mixture consisted of 2.5 pmols each of the forward and reverse primer, 50ng of cDNA template and 2x SyBr Green mastermix (Roche). The qPCR cocktail was mixed and added into 96-well plates and subjected to the following program in an ABI 7500 Fast RealTime PCR system.

**qPCR program**

Step	Temperature	Time	Repeat
Lid pre-heating	100 °C		
Sample pre-heating	50 °C	2 mins	
Initial denaturation	95 °C	10 mins	
Denaturation	95 °C	30 secs	40 times
Annealing	60 °C	15 secs	

**Data analysis**

The ABI 7500 Fast RealTime PCR system instrument measured fluorescence emitted after each cycle of the qPCR program. Based on the amplification curves of each of the sample, the cycle at which the fluorescence emitted by each sample exceeds a particular threshold is computed. This cycle number was known as the Threshold Cycle (Ct) for each sample. This number was used for further calculations as described below. This Ct value was obtained for each gene across all the different samples.

The Ct values corresponding to the house keeping genes was subtracted from the Ct values corresponding to the candidate genes, for each sample. The ensuing value was known as the  $\Delta$ Ct. The relative mRNA expression values for each sample were calculated using the following formula.

$$\text{Relative mRNA level} = 2^{-\Delta\text{Ct}}$$

The Relative mRNA level of the tumor and nodular samples for each candidate gene were normalized over the corresponding relative mRNA level of the control sample for the same gene. The ratio obtained herewith was the fold-change regulation of the candidate gene in the tumor and nodular samples compared to the control sample.

### 2.3.4 Mutation detection

The cDNA was also used for detection of mutations in the *Ctnnb1* and *Trp53* gene loci, by amplifying the transcriptome hotspot region of mutations and sequencing the amplicons.

#### Materials used

Buffer/Instrument/ Kit	Company/ Composition
5X Phusion Buffer	Thermo Scientific™
Primers	Sigma-Aldrich
dNTPs	Promega GmbH (dATP,dCTP,DGTP, and dTTP 10mM each)
Phusion DNA Polymerase	Thermo Scientific™
Thermocycler	Biometra
GeneJet purification kit	Thermo Scientific™

#### 2.3.4.1 Primer design

The primers for amplifying the hotspot of mutations in the *Ctnnb1* and *Trp53* gene locus had been designed by Stefan Ohrnberger using the PrimerExpress software from Applied Biosystems. The parameters considered were melting temperature (57°C to 63°C), primer length (18-22 bases long) and GC content (45 % – 55 %). The primer sequences used for this experiment are shown in section 2.10

#### 2.3.4.2 PCR reaction

The reaction mix for each PCR reaction consisted of following table.

## MATERIALS AND METHODS

---

<b>Buffer</b>	<b>Concentration</b>	<b>Volume (<math>\mu</math>L)</b>
Polymerase Buffer	5x	10
Forward Primer	0.2 $\mu$ M	2.5
Reverse Primer	0.2 $\mu$ M	2.5
dNTP mix	40mM	1
Taq Polymerase	2.5 units	0.5
Water		31
Template DNA		2.5
Total Volume		50

The cocktail was mixed gently and then placed in the thermo cycler and the corresponding PCR program was run.

### PCR program

<b>Step</b>	<b>Temperature</b>	<b>Time</b>	<b>Repeat</b>
Lid pre-heating	100 °C		
Initial denaturation	98 °C	30 secs	
Denaturation	98 °C	10 secs	30 times
Annealing	67 °C	20 secs	
Extension	72 °C	30 secs	
Final extension	72 °C	5 mins	
Storage	4 °C	Hold	

After the PCR program had completed, 10  $\mu$ L of the amplification reaction was subjected to agarose gel electrophoresis, to check for proper amplification of the loci.

### **2.3.4.3 Agarose gel electrophoresis**

A part (10  $\mu\text{L}$ ) of the amplification reaction from each sample was mixed with 6x (2  $\mu\text{L}$ ) of loading dye and was loaded into 2 % agarose gels. The agarose gel electrophoresis was performed as described in detail in section 2.8.8.

### **2.3.4.4 PCR Product purification**

The remaining PCR amplicons were then purified using the GeneJET PCR purification kit by following the instructions provided in the kit handbook. The detailed procedure is described in section 2.8.7

### **2.3.4.5 Sanger sequencing**

The purified amplicons were then sent to LGC Genomics GmbH for sequencing. 150 ng of the purified PCR amplicons was sent along with 5 pmols of the sequencing primer (4  $\mu\text{L}$ ). The results obtained were analyzed using Chromas Lite software to identify mutations in the loci.

## **2.3.5 Microarray analysis**

The microarray analysis of *SRF-VP16<sup>iHep</sup>* mouse tumor and nodule samples was performed in collaboration with Dr. Michael Bonin, Human Genetics, UKT, University of Tuebingen. The chip used for the microarray gene expression profiling was the Mouse Exon 1.0 ST kit from Affymetrix. The coverage of the Mouse Exon 1.0 ST kit is 26,166 transcripts of the mouse genome (mm9 mouse genome build).

The procedure was as follows. The RNA of each sample was hybridized with predesigned probes specific for each of the 26,166 mouse transcripts. The hybridization causes fluorescence to be emitted. The emitted fluorescence is directly proportional to the amount of transcript bound to each of the probes. The fluorescence signal for each probe for every sample was measured. The intensity of signal is proportional to the

## MATERIALS AND METHODS

---

amount of the corresponding transcript present in the sample (Schinke-Braun and Couget, 2007). The samples taken for the study are shown in Table 2-1.

*Table 2-1: List of samples taken for microarray analysis. The litter id, sample id, gender, the LWBW ratio and the sample type are shown in the table. LWBW stands for Liver Weight to Body Weight Ratio at time of sample collection. HCC<sub>A</sub> denotes Ctnnb1 wt tumors, while HCC<sub>B</sub> denotes Ctnnb1 mutant tumors and Nodules stand for pre-malignant, hyper-proliferative nodular liver tissue. Control refers to the samples taken from normal livers of littermate control mice,*

<b>Litter Id</b>	<b>Sample Id</b>	<b>Age (weeks)</b>	<b>Gender</b>	<b>LWBWR (%)</b>	<b>Sample type</b>
15	15-2C	18	M	4.8	Control
24	24-2C	30	F	4.4	Control
16	16-1C	25	M	5	Control
15	15-1N	18	M	12.4	Nodule
17	17-1N	17	F	11.7	Nodule
43	43-2T	17	M	10.4	Nodule
2	2-2T	32	M	24.3	HCC A
25	25-1T1	25	M	41.2	HCC A
24	24-1T2	30	F	25.3	HCC B
4	4-1T1	32	F	26.3	HCC B
26	26-3T1	31	F	26.8	HCC B
25	25-4T2	32	F	30.1	HCC B

## 2.4 Protein analysis

### 2.4.1 Protein extraction

The protein extraction was performed as described in Section 2.8.3. The murine liver tissue preparation was used. The flow-through from the RNA extraction column was subjected to protein precipitation using Acetone. The precipitated proteins were dissolved in urea buffer. The concentration of the proteins was measured using Bradford's reagent.



## 2.5 Combined omics analysis

Tumor samples harvested from the *SRF-VP16<sup>iHep</sup>* mouse liver were subjected to the following high-throughput omics analyses.

- Whole Exome Sequencing
- Transcriptomics analysis - RNAseq
- Proteomics analysis using – LC-MS/MS
- Phosphoproteomics analysis – LC-MS/MS
- Metabolomics analysis – LC-MS/MS

The results were then compared with the human HCCs, in order to derive at meaningful conclusions about hepatocarcinogenesis using the *SRF-VP16<sup>iHEP</sup>* mouse model. The strategy undertaken is summarized in Figure 2-5.

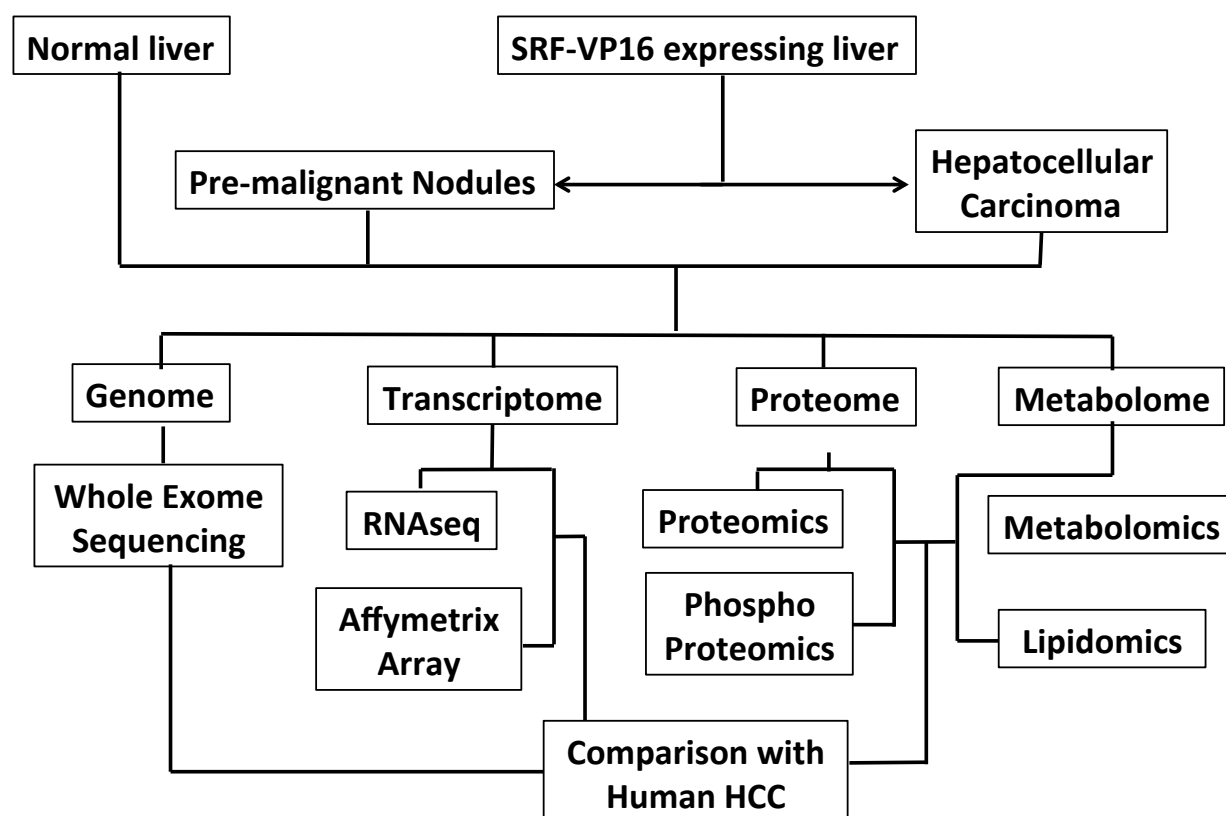


Figure 2-5: Combined-omics analysis strategy showing the usage of the different sample groups for combined omics approach. Affymetrix array denotes the Microarray analysis.

## 2.5.1 Whole exome sequencing

Whole Exome Sequencing (WES) was performed at the Genomics Unit of DKFZ, Heidelberg. 44 murine samples coming from 14 different mice, were subjected to Whole Exome sequencing analysis. The samples used for WES analysis are listed in Table 2-2.

*Table 2-2: List of samples taken for WES analysis. The litter id, sample id, gender, the LWBW ratio and the sample type are shown in the table. LWBW stands for Liver Weight to Body Weight Ratio at time of sample collection. HCC<sub>A</sub> denotes Ctnnb1 wt tumors, while HCC<sub>B</sub> denotes Ctnnb1 mutant tumors and Nodules stand for pre-malignant, hyper-proliferative nodular liver tissue. Control refers to the samples taken from normal livers of the littermate control mice.*

Litter Id	Sample Id	Age (weeks)	Gender	LWBWR (%)	Sample type
4	4-2C	32	F	4.8	Control
25	25-5C	32	F	5.1	Control
26	26-4C	31	F	4.2	Control
49	49-1C	52	M	4.9	Control
57	57-2S	65	F	NA	Control*
60	60-2S	42	F	NA	Control*
38	38_1S	68	M	NA	Control*
59	59-1S	30	M	NA	Control*
2	2-2N	32	M	24.3	Nodule
4	4-1N	32	F	26.3	Nodule
25	25-4N	32	F	30.1	Nodule
26	26-3N	31	F	26.8	Nodule
60	60-2N	42	F	29.5	Nodule
38	38_1N	68	M	18.1	Nodule
57	57-2T3	65	F	42.2	HCC
57	57-2T7	65	F	42.2	HCC
57	57-2T9	65	F	42.2	HCC
60	60-2T1	42	F	29.5	HCC
60	60-2T2	42	F	29.5	HCC
60	60-2T3	42	F	29.5	HCC
60	60-2T4	42	F	29.5	HCC
60	60-2T5	42	F	29.5	HCC
38	38_1T	68	M	18.1	HCC
32	32_1T1	30	F	33.6	HCC
47	47_2T1	56	F	36.9	HCC
47	47_2T2	56	F	36.9	HCC
47	47_2T3	56	F	36.9	HCC

Litter Id	Sample Id	Age (weeks)	Gender	LWBWR (%)	Sample type
54	54-2T2	41	F	22.2	HCC
54	54-1T4	62	F	26.7	HCC
54	54-1T6	62	F	26.7	HCC
54	54-1T7	62	F	26.7	HCC
58	58-1T1	71	F	8.7	HCC
58	58-1T5	71	F	8.7	HCC
2	2-2T	32	M	24.3	HCC <sub>A</sub>
4	4-1T2	32	F	26.3	HCC <sub>A</sub>
49	49-2T3	52	M	16.4	HCC <sub>A</sub>
59	59-1T4	30	M	23.4	HCC <sub>A</sub>
4	4-1T1	32	F	26.3	HCC <sub>B</sub>
24	24-1T2	30	F	25.3	HCC <sub>B</sub>
25	25-4T1	32	F	30.1	HCC <sub>B</sub>
26	26-3T1	31	F	26.8	HCC <sub>B</sub>
26	26-3T2	31	F	26.8	HCC <sub>B</sub>
59	59-1T2	30	M	23.4	HCC <sub>B</sub>
59	59-1T3	30	M	23.4	HCC <sub>B</sub>

\* For these control samples, the DNA was extracted from the skin of the SRF-VP16<sup>iHep</sup> mice, which had tumor-bearing livers.

DNA was extracted, as described in section 2.2.1 and was sent to the Genomics Unit of DKFZ, where the subsequent library preparation and sequencing was done. Briefly, the DNA was fragmented and then the exomes were specifically captured using specially synthesized biotinylated beads, which were specific for each of the exons of the genome. The bound DNA fragments were washed and then eluted. The adapter sequences were fused to the ends of the DNA fragments and then they were sequenced (Teer and Mullikin, 2010). The bioinformatics analysis of the sequenced reads was carried out by Dr. Johannes Werner from the group of Dr. Mathias Schlesner at DKFZ, Heidelberg.

## 2.5.2 Transcriptomics and proteomics

Transcriptomics and proteomics analysis were performed according to the following strategy. The protein and RNA extraction were done concurrently and they were subjected to these different omics analyses in collaboration with the following institutes/ research groups.

- Genome Center, Max Planck Institute (MPI), Tuebingen
- Dr. Katarina Matic, group of Prof. Dr. B. Macek, PCT, Tuebingen
- Quantitative Biology Center (QBiC), Tuebingen

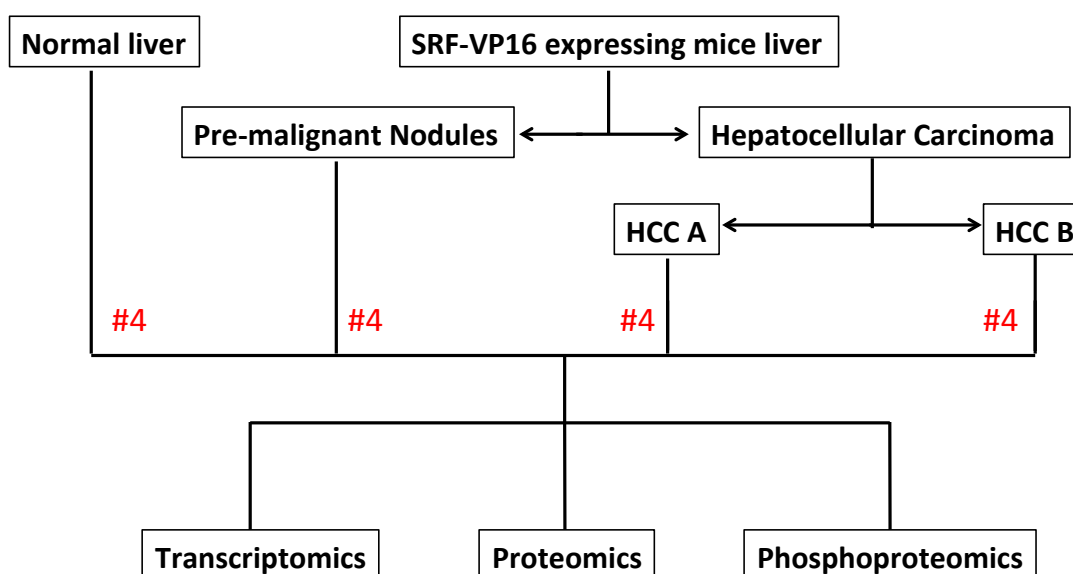


Figure 2-6: Transcriptomic and proteomic Strategy. The numbers mentioned in red denote the numbers of separate murine samples that were taken from each category for the downstream analyses.

A total of 16 samples, derived from 10 different mice, were taken and processed for RNAseq and mass spectrometric analysis. The samples used are described in Table 2-3.

*Table 2-3: List of samples taken for transcriptomic and proteomic analysis. The litter id, sample id, gender, the LWBW ratio and the sample type are shown in the table. LWBW stands for Liver Weight to Body Weight Ratio at time of sample collection. HCC<sub>A</sub> denotes Ctnnb1 wt tumors, while HCC<sub>B</sub> denotes Ctnnb1 mutant tumors and Nodules stand for pre-malignant, hyper-proliferative nodular liver tissue. Control refers to the samples taken from normal livers of the littermate mice.*

<b>Litter Id</b>	<b>Sample Id</b>	<b>Age (weeks)</b>	<b>Gender</b>	<b>LWBWR (%)</b>	<b>Sample type</b>
4	4-2C	32	F	4.8	Control
25	25-5C	32	F	5.1	Control
26	26-4C	31	F	4.2	Control
49	49-1C	52	M	4.9	Control
2	2-2N	32	M	24.3	Nodule
4	4-1N	32	F	26.3	Nodule
25	25-4N	32	F	30.1	Nodule
59	59-1N*	30	M	23.4	Nodule
2	2-2T	32	M	24.3	HCC <sub>A</sub>
4	4-1T2	32	F	26.3	HCC <sub>A</sub>
49	49-2T3	52	M	16.4	HCC <sub>A</sub>
59	59-1T1*	30	M	23.4	HCC <sub>A</sub>
25	25-4T1	32	F	30.1	HCC <sub>B</sub>
26	26-3T2	31	F	26.8	HCC <sub>B</sub>
59	59-1T2	30	M	23.4	HCC <sub>B</sub>
59	59-1T3	30	M	23.4	HCC <sub>B</sub>

- *The two samples were excluded from proteomics and phosphoproteomics downstream analysis due to low quality.*

### 2.5.3 RNAseq analysis

RNAseq analysis was performed in collaboration with Genome Center, Max Planck Institute, Tuebingen and QBiC. RNAseq profiling was carried out on 16 different mouse samples. The major steps involved in the RNAseq analysis are as follows:

- Library preparation
- Sequencing
- Bioinformatic processing
  - Data storage
  - Data analysis
  - Data mining

#### 2.5.3.1 Library preparation

The Truseq library preparation kit from Illumina Inc. was used for the preparation of sample libraries. The detailed procedure is described in section 2.8.4. Briefly, the RNA molecules were purified and fragmented. The RNA fragments were subjected to PCR amplification and cDNA synthesis. The adapter sequences were then ligated to the ends of the fragments. The adapter-ligated fragments from different samples were pooled together before loading into the Illumina Inc. sequencer.

#### 2.5.3.2 Sequencing

Sequencing of the sample libraries was performed on an Illumina Inc. HiSeq 2000 instrument by technicians at the Genome Center, Max Planck Institute, Tuebingen.

#### 2.5.3.3 Bioinformatics

The raw data files containing the sequenced reads were transferred to the Quantitative Biology Centre (QBiC), Tuebingen. Subsequent bioinformatics processing of the data was performed by QBiC. The data was also stored at the QBiC facility.

## 2.5.4 LC-MS/MS analysis

The protein samples were analyzed for variations in their expression levels by Mass spectrometry in collaboration with Dr. Katarina Matic from Prof. Dr. Boris Macek's group at the Proteome Center Tuebingen.

Mass spectrometric measurement of the protein peptides and phospho- enriched peptides involved the following steps.

- Sample preparation
- Measurement of spectra
- Bioinformatic processing

### 2.5.4.1 Sample preparation

The sample preparation was performed as described in section 2.8.5. Briefly, the protein samples were digested into peptides. The peptides were purified using StageTips (for proteomics) or Sep-Pak 1cc C<sub>18</sub> Cartridges (for Phosphoproteomics). In case of phosphoproteomics, the samples were enriched for phosphopeptides using TiO<sub>2</sub> beads and then purified on StageTips.

### 2.5.4.2 LC-MS analysis

The LC-MS Analysis was performed by Dr. Katarina Matic at the Proteome Center Tuebingen. The peptide digests were measured on EASY-nLC II nano-LCs (Proxeon Biosystems) coupled to an LTQ Orbitrap XL instrument in case of proteomics. For phosphoproteomic measurement, EASY-nLC II nano-LCs coupled to an LTQ Orbitrap Elite (Thermo Fisher Scientific) was used.

### 2.5.4.3 Bioinformatic processing

The Bioinformatic processing of the measurements was done by Dr. Matic (PCT, University Tuebingen) using the MaxQuant software.

## 2.6 Metabolomics analysis

Metabolomics analysis was performed on mouse tumor liver tissue in collaboration with the company MetaSysX. To this end, 50 mg of liver tissue was snap frozen in liquid nitrogen and was stored in  $-80\text{ }^{\circ}\text{C}$  and later sent in dry ice to the MetaSysX GmbH laboratories for subsequent processing.

### 2.6.1 Samples

The following samples were used for metabolomics analysis.

*Table 2-4: List of samples taken for Metabolomics analysis. The litter Id, sample Id, gender, the LWBW ratio and the sample type are shown in the table. LWBW stands for Liver Weight to Body Weight Ratio at time of sample collection. HCC denotes the tumor samples collected from SRF-VP16<sup>iHep</sup> mice livers, while Nodules stand for pre-malignant, hyper-proliferative nodular liver tissue. Control refers to the samples taken from normal livers of the littermate mice.*

Litter Id	Sample Id	Age (weeks)	Gender	LWBWR (%)	Sample type
61	61-2C	25	M	5.73	Control
62	62-3C	31	F	3.74	Control
63	63-2C	24	M	5	Control
64	64-2C	19	M	4.6	Control
61	61-1N	25	M	26.7	Nodule
62	62-1N	31	F	24.54	Nodule
62	62-2N	27	F	29.85	Nodule
63	63-1N	24	M	22.14	Nodule
64	64-1N	28	M	28.1	Nodule
61	61-1T1	25	M	26.7	HCC
61	61-1T2	25	M	26.7	HCC
61	61-1T3	25	M	26.7	HCC
62	62-1T1	31	F	24.54	HCC
62	62-2T1	27	F	29.85	HCC
62	62-2T2	27	F	29.85	HCC
62	62-2T3	27	F	29.85	HCC
62	62-2T4	27	F	29.85	HCC
62	62-2T6	27	F	29.85	HCC
62	62-2T7	27	F	29.85	HCC
62	62-2T8	27	F	29.85	HCC
64	64-1T1	28	M	28.1	HCC
64	64-1T2	28	M	28.1	HCC



## 2.7 Comparison to human HCCs

The high-throughput omics data, especially those of the transcriptome, were compared to existing human HCC datasets. The human HCC datasets used for the comparison are as follows.

### 2.7.1 Heidelberg cohort

#### 2.7.1.1 Patient data

The Heidelberg Cohort consisted of 40 human HCC patients. The characteristic features of the patients are described in Table 2-5. Dr. Thomas Longerich kindly shared the Microarray analysis data and selected methylation analysis data of these tumor samples

#### 2.7.1.2 Boyault classification

Based on the gene expression values, the tumor samples were classified into 6 different groups as recommended by Boyault et.al, 2007 (Boyault et al., 2007). To this end, the expression values of the classifier genes were taken and the values were subjected to the given formula according to Boyault et.al, 2007. The values generated by the formula were then used to assign the tumors to the respective sample group that resembled the sample the closest. The Boyault classification was performed in order to be able to better compare the gene expression profiling and other molecular characteristics of these tumors with those of the *SRF-VP16<sup>iHep</sup>* mice.

#### 2.7.1.3 Unsupervised hierarchical clustering

Unsupervised hierarchical clustering of the samples was performed using two different ways. One was by using the software Cluster 3.0 originally written by Michael Eisen. the version usable for Mac was written by Michiel de Hoon (Eisen et al., 1998).

## MATERIALS AND METHODS

---

The obtained cluster files were visualized using the Treeview software, also hosted by the same group.

The second method was by running a DOS based script written mainly for this purpose by Dr. Michael Römer, from the group of Prof. Dr. Zell, University of Tuebingen. The added advantage of this method was that for each cluster generated, there was also a Silhouette score generated, which is a measure of distance between each of the clusters.

*Table 2-5: Patient information of the Heidelberg cohort consisting of 40 human HCC patients. The etiology, age, gender of the patients and the maximum tumor size, tumor grade and cirrhosis status of the liver are mentioned. The mutation status of the patients in the context of their CTNNB1 gene locus is also given. HBV and HCV etiology stands for Hepatitis B Virus and Hepatitis C virus infection respectively. The tumor grade is mentioned in numbers, where 1 stands for well-differentiated HCC, 2 for moderately differentiated HCC and 3 for poorly differentiated HCC. There was one patient with grade 4, which denotes aggressive and invasive HCC at very advanced stage. The presence of cirrhosis in the liver is denoted by 1 and the absence by 0. In the last column, the wild type CTNNB1 locus is denoted by WT and a mutant genotype is denoted by Mutant.*

Sample	Etiology	age	sex	max. size (cm)	Grading	cirrhosis	CTNNB1 mutation
HCC001	HBV	78	m	4.5	2	1	WT
HCC003	HCV	74	f	3.5	2	1	Mutant
HCC005	HCV	48	m	2.5	2	1	WT
HCC006	HBV	40	m	7.0	2	0	WT
HCC007	HBV	50	f	4.0	2	1	WT
HCC008	HBV	73	f	6.0	3	1	WT
HCC010	HCV	72	m	3.0	2	0	WT
HCC011	HBV	61	f	12.0	2	0	WT
HCC015	HBV	68	m	5.1	2	0	WT
HCC018	HCV	54	m	2.5	1	1	WT
HCC019	HCV	54	m	2.5	1	1	WT
HCC022	HCV	63	f	4.0	2	1	WT
HCC023	alcohol	70	m	7.0	3	1	Mutant
HCC025	cryptogenic	47	m	x	2	0	WT
HCC027	cryptogenic	38	f	29.0	2	0	WT
HCC028	alcohol	57	m	2.0	1	1	WT
HCC029	alcohol	57	m	3.0	2	1	Mutant
HCC035	alcohol	54	m	15.0	2	1	Mutant

Sample	Etiology	age	sex	max. size (cm)	Grading	cirrrhosis	CTNNB1 mutation
HCC037	cryptogenic	58	m	2.8	2	1	WT
HCC038	alcohol	57	m	3.5	2	1	WT
HCC039	cryptogenic	58	m	1.8	1	1	WT
HCC040	HFE	65	m	2.3	2	0	Mutant
HCC041	alcohol	52	f	12	2	1	WT
HCC042	cryptogenic	73	m	11.0	2	1	WT
HCC043	HCV	35	m	6.0	2	0	WT
HCC044	cryptogenic	16	m	11.5	2	0	Mutant
HCC045	HCV	51	m	2.8	2	1	WT
HCC046	HBV	53	m	5.0	4	1	WT
HCC048	cryptogenic	55	m	13.0	2	0	WT
HCC050	cryptogenic	54	m	8.0	1	0	Mutant
HCC051	cryptogenic	78	m	4.0	2	0	WT
HCC053	HCV	38	f	6.5	2	0	WT
HCC056			m	16.0	2	0	Mutant
HCC057	HBV	38	m	x	3	0	WT
HCC058	HFE	unbekannt	f	x	2	1	WT
HCC061	alcohol	55	m	4.0	2	1	WT
HCC063	cryptogenic	69	m	4.5	2	0	WT
HCC071	alcohol	68	m	7.0	2	0	WT
HCC072	cryptogenic	55	m	3.5	2	1	Mutant
HCC074	cryptogenic	69	f	9.0	3	0	Mutant

#### 2.7.1.4 Mainz cohort

The dataset, which was based on RNAsq was shared by Dr. Jens Marquardt from the Johannes Gutenberg University in Mainz. The sample details are as described in Figure 2-7 .

Briefly, this dataset consisted of 28 samples collected from 8 different HCC patients. From each of these patients, multiple samples had been collected, corresponding to different stages of tumor development such as low-grade dysplastic nodules, high grade dysplastic nodules, early HCC and progressed HCC. The advantage of this approach is to be able to study the early driver events that have occurred during the early malignant transformation process itself (Marquardt et al., 2014).

## MATERIALS AND METHODS

**Table 1. Clinicopathological information and HCC marker expression.**

Patient No.	Clinicopathological information						Marker expression				
	Sex	Age (yr)	Intrahepatic metastasis	Vascular invasion	Lesion	Size (cm)	GPC3	HSP70	GS	K19	AFP
1	M	68	No	No	LGDN	0.8 x 0.8	Neg	Neg	Neg	Neg	Neg
					HGDN	1.5 x 1.4	Neg	Neg	Neg	Neg	Neg
2	M	54	No	No	HGDN	0.8 x 0.8	Neg	Neg	Neg	Neg	Neg
					HGDN	0.8 x 0.7	Neg	Neg	Neg	Neg	Neg
					eHCC	1.2 x 1.0	Neg	Neg	Pos	Neg	Neg
					pHCC	3.0 x 2.0	Neg	Pos <sup>a</sup>	Neg	Neg	Neg
3	F	42	No	No	HGDN	1.0 x 0.9	Neg	Weak <sup>a</sup>	Neg	Neg	Neg
4	M	64	No	No	HGDN	1.5 x 1.3	Neg	Neg	Weak <sup>a</sup>	Neg	Neg
5	M	61	No	No	LGDN	1 x 0.8	Neg	Neg	Neg	Neg	Neg
					LGDN	0.8 x 0.8	Neg	Neg	Neg	Neg	Neg
					LGDN	1.2 x 1.1	Neg	Neg	Neg	Neg	Neg
					HGDN	1.0 x 1.0	Neg	Neg	Neg	Neg	Neg
					HGDN	1.3 x 0.9	Neg	Neg	Pos	Neg	Neg
					HGDN	1.0 x 1.1	Neg	Weak <sup>a</sup>	Pos	Neg	Neg
					eHCC	1.6 x 1.4	Weak <sup>a</sup>	Pos <sup>a</sup>	Pos	Neg	Neg
pHCC	3.8 x 3.2	Weak <sup>a</sup>	Pos <sup>a</sup>	Pos	Neg	Neg					
6	M	61	No	No	eHCC	2 x 1.9	Neg	Weak <sup>a</sup>	Neg	Neg	Neg
7	M	50	No	No	eHCC	1 x 1.3	Neg	Weak	Neg	Neg	Neg
8	F	60	No	No	HGDN	1.0 x 1.0	Neg	Neg	Neg	Neg	Neg
					eHCC	1.1 x 1.0	Pos	Pos <sup>a</sup>	Pos	Neg	Pos <sup>a</sup>
					pHCC	1.2 x 1.0	Pos	Pos <sup>a</sup>	Pos	Neg	Neg

Figure 2-7: Patient information of the Mainz cohort of 8 human HCC patients. The age, gender of the patients and the tumor size, intrahepatic metastasis status and vascular invasion status of the livers are mentioned. The type of lesions derived from each sample is also denoted. LGDN stands for low-grade dysplastic nodule. HGDN stands for high-grade dysplastic nodule. eHCC and pHCC stand for early HCC and progressed HCC respectively. The expressions of marker proteins of the samples analysed by immunohistochemistry are also given. Neg denotes negative and Pos denotes a strong positive, whereas Weak denotes a partial positivity of the samples for the corresponding marker proteins. <sup>a</sup> This symbol denotes focal positivity of the samples to the corresponding genes. Source: (Marquardt et al., 2014)

### 2.7.1.5 TCGA dataset

The Cancer Genome Atlas is a large-scale collaborative effort, with a main focus on understanding the basic molecular features of different cancers by using advanced large-scale genomic sequencing technologies. The initiative began as a pilot project in 2006 by National Cancer Institute and the National Human Genome Research Institute. The data was made freely available to the public and now consists of data covering more than 33 different types of cancers. In the TCGA database, data is available for 373 liver cancer patients, along with 51 control liver samples., which was used for clustering analyses (Ho et al., 2015).

## 2.8 Detailed procedures

### 2.8.1 RNA extraction

RNA was extracted from the tissues using the Qiagen RNEasy Mini kit based on the instructions provided in the handbook. The procedure consisted of the following steps.

#### Reagents Used

S. No.	Item	Supplier
1.	70% Ethanol	Sigma-Aldrich
2.	RLT Buffer	Qiagen
3.	RW1 Buffer	Qiagen
4.	RPE Buffer	Qiagen
5.	Elution Buffer	Qiagen

30 mg of tissue sample was taken for the extraction.

1. 600  $\mu$ L of RLT buffer was added to the tissue slices for lysis.
2. The tissue samples were disrupted using a douncer.
3. Tissue samples were homogenized using needles and syringes. Each of the disrupted tissue in lysis solution was passed through a needle for a couple of times.
4. Homogenized lysate was centrifuged at 14,000 rpm for 3 minutes.
5. The supernatant was transferred to a new tube and to it was added 1 volume of 70 % Ethanol and mixed gently.

6. The mixture was added onto the RNeasy spin columns provided in the kit.
7. The columns were then centrifuged at 10,000 rpm for 1 minute. The flow through was collected and stored for subsequent DNA and protein synthesis.
8. 700  $\mu$ L of Buffer RW<sub>1</sub> was added onto the columns. The columns were centrifuged at 10,000 rpm for 1 minute. The flow-throughs were discarded.
9. 500  $\mu$ L of Buffer RPE was added onto the columns. The columns were centrifuged at 10,000 rpm for 1 minute. The flow-throughs were discarded.
10. The previous step was repeated once more, this time the columns were centrifuged for 2 minutes.
11. The empty columns were centrifuged once again at 13,000 rpm for 3 minutes to remove any residual ethanol remaining in the column.
12. The RNA was eluted by adding 50  $\mu$ L of Elution Buffer to the columns placed on microcentrifuge tubes and centrifuging at 10,000 rpm for 1 minute.
13. The RNA concentration was measured using Qbit or the Nanodrop.

### **2.8.2 DNA extraction**

DNA was extracted from the tissues using the Qiagen Blood & Tissue kit based on the instructions provided in the handbook. The procedure consisted of the following steps.

**Reagents Used**

S. No.	Item	Supplier
1.	AL Buffer	Qiagen
2.	AW1 Buffer	Qiagen
3.	AW2 Buffer	Qiagen
4.	AE Buffer	Qiagen

1. 200  $\mu$ L of flow-through from the first centrifugation step of RNA extraction was taken for the sake of DNA preparation.
2. 200  $\mu$ L of Buffer AL was added to the solution and mixed thoroughly.
3. 200  $\mu$ L of absolute ethanol was added and the solution was mixed thoroughly once again.
4. The mixture was then loaded immediately into a DNeasy Mini spin column, and was centrifuged at 8000 rpm for 1 minute.
5. The flow-through was saved for protein extraction later on.
6. 500  $\mu$ L of Buffer AW<sub>1</sub> was added onto the column and centrifuged at 8000 rpm for 1 minute.
7. 500  $\mu$ L of Buffer AW<sub>2</sub> was added onto the column and centrifuged at 14,000 rpm for 3 minute.
8. The column was placed in a microcentrifuge tube.
9. The DNA was eluted from the column by adding 200  $\mu$ L of buffer AE and then centrifuged at 8000 rpm for 1 minute.
10. The concentration of the eluted DNA was measured using the Nanodrop apparatus.

### 2.8.3 Protein extraction

Proteins were extracted from the tissues using the Acetone precipitation method. The procedure involved the following steps.

#### Reagents Used

S. No.	Item	Supplier/ Composition
1.	Urea Buffer	6M Urea, 2M Thiourea, 10mM Tris, pH 8 + 1% NOG
2.	Acetone	Sigma-Aldrich

1. The entire flow-through from the first centrifugation step of DNA extraction was taken for the sake of protein preparation.
2. 2.5 volumes of Acetone was added to the solution and mixed gently.
3. The mixture was incubated at -20 °C overnight.
4. The next day, the solution was centrifuged at 14,000 rpm for 15 minutes.
5. The supernatant was carefully removed and discarded.
6. The pellet was then resuspended in Urea buffer for subsequent processing.
7. The concentration of the proteins was measured using Bradford's reagent.



## 2.8.4 RNAseq library preparation

The RNAseq sample libraries were prepared using the Truseq® RNA sample preparation kit based on the manufacturers instructions. The protocol consists of the following steps.

- I. Purification and Fragmentation of mRNA
- II. First Strand cDNA synthesis
- III. Second strand cDNA synthesis
- IV. Ends Repair
- V. Adenylation of 3' ends
- VI. Ligation of adapters
- VII. Enrichment of DNA fragments by PCR Amplification
- VIII. Validation of prepared sample libraries
- IX. Normalization and pooling of libraries

### I. Purification and fragmentation of mRNA

#### Reagents

S. No.	Item	Supplier
3.	Bead Binding Buffer	Illumina Inc.
4.	Bead Washing Buffer	Illumina Inc.
5.	Elute, Prime, Fragment Mix	Illumina Inc.
6.	Elution Buffer	Illumina Inc.
7.	Resuspension Buffer	Illumina Inc.
8.	RNA Purification Beads	Illumina Inc.

## MATERIALS AND METHODS

---

The above reagents were thawed at room temperature, vortexed and centrifuged before proceeding with the protocol.

1. Diluted 1  $\mu\text{g}$  of RNA with water to a final volume of 50  $\mu\text{l}$ .
2. Added 50  $\mu\text{l}$  RNA Purification Beads to each well of the RNA containing plate in order to bind the polyA RNA to the oligo-dT beads. Gently mixed the solution by pipetting up and down.
3. mRNA were denatured by incubating the samples at 65  $^{\circ}\text{C}$  for 5 minutes in a thermal cycler apparatus.
4. Removed the samples from the thermal cycler when it cooled down to 4  $^{\circ}\text{C}$ .
5. Samples were incubated at room temperature for 5 minutes to allow the binding of RNA to the beads.
6. Samples were placed on a magnetic stand at room temperature for 5 minutes to separate the RNA bound beads from the solution.
7. Discarded the supernatant from each well. Samples were removed from the magnetic stand
8. Washed the beads by adding 200  $\mu\text{l}$  of Bead Washing Buffer to each well. Gently mixed the solution by pipetting up and down. This removed the unbound RNA from the beads.
9. Samples were placed on a magnetic stand at room temperature for 5 minutes.
10. Discarded the supernatant mainly consisting of ribosomal and other non-messenger RNA. Samples were removed from the magnetic stand.
11. Added 50  $\mu\text{l}$  Elution Buffer to each well and the solution was gently mixed by pipetting up and down.

12. Samples were incubated at 80 °C for 2 minutes in a thermal cycler, in order to elute the mRNA and the contaminant rRNA that had been bound to the beads.
13. Samples were removed from the thermal cycler when they reached room temperature.
14. Added 50 µl of Bead Binding Buffer to each well, to enable the mRNA alone to rebind to the beads. Gently mixed the solution by pipetting up and down.
15. Incubated the samples at room temperature for 5 minutes.
16. Samples were placed on the magnetic stand at room temperature for 5 minutes.
17. Discarded the supernatant from each well and removed the samples from the magnetic stand.
18. Washed the beads with 200 µl of Bead Washing Buffer and gently mixed the solution by pipetting up and down.
19. Samples were placed on the magnetic stand at room temperature for 5 minutes.
20. Discarded the supernatant containing the residual rRNA and then removed the samples from the magnetic stand.
21. Added 19.5 µl of Elute, Prime, Fragment Mix (containing random hexamers) to the samples and mixed the solution by pipetting up and down.
22. Incubated the samples at 94°C for 8 minutes in a thermal cycler.
23. The samples were removed from the thermal cycler once it had cooled down to 4°C.

## II. Synthesis of first strand cDNA

### Reagents

S. No.	Item	Supplier
1.	First Strand Master Mix	Illumina Inc.

The above reagent was thawed at room temperature, vortexed and centrifuged before proceeding with the protocol. To the First Strand Master Mix tube was added 50  $\mu$ L of SuperScript II and mixed thoroughly.

1. Samples were placed on the magnetic stand at room temperature for 5 minutes.
2. Transferred 17  $\mu$ l of the supernatant consisting of fragmented and primed mRNA to new wells.
3. Added 8  $\mu$ l of pre-mixed First Strand Master Mix to each well and mixed
4. The samples were placed in a thermal cycler and the following program was run.
  - a. Lid pre-heating – 100 °C
  - b. 25°C for 10 minutes
  - c. 42°C for 50 minutes
  - d. 70°C for 15 minutes
  - e. Hold at 4 °C
5. Samples removed from thermal cycler once it had cooled to 4 °C.

### III. Synthesis of second strand cDNA

#### Reagents

S. No.	Item	Supplier
1.	Second Strand Master Mix	Illumina Inc.
2.	Resuspension Buffer	Illumina Inc.
3.	AMPure XP beads	Beckman Coulter GmbH
4.	Freshly prepared 80% Ethanol	Sigma-Aldrich

The above reagents 1 & 2 were thawed at room temperature, vortexed and centrifuged before proceeding with the protocol. The AMPure XP beads were mixed well before use.

1. Added 25  $\mu$ l of Second Strand Master Mix to the samples and mixed well.
2. Samples were incubated at 16°C for 1 hour in a thermal cycler.
3. Samples were removed from thermal cycler and let to stand at the bench to bring them to room temperature.
4. Added 90  $\mu$ l of well-mixed AMPure XP beads to the samples and mixed well
5. Incubated the mixture at room temperature for 15 minutes.
6. Samples were placed on magnetic stand at room temperature for 5 minutes.
7. Discarded 135  $\mu$ l of the supernatant from each well.
8. Beads were washed twice with 200  $\mu$ l of freshly prepared 80% ethanol for 30 seconds each.
9. The samples were left at room temperature for the beads to air dry.
10. Samples were removed from the magnetic stand.
11. Added 52.5  $\mu$ l of Resuspension Buffer to each well and mixed thoroughly by pipetting up and down.

## MATERIALS AND METHODS

---

12. Incubated the samples at room temperature for 2 minutes.
13. Samples were placed on magnetic stand at room temperature for 5 minutes.
14. Transferred 50  $\mu$ L of supernatant to a new well.

### IV. Ends Repair

#### Reagents

S. No.	Item	Supplier
1.	End Repair Mix	Illumina Inc.
2.	Resuspension Buffer	Illumina Inc.
3.	AMPure XP beads	Beckman Coulter GmbH
4.	Freshly prepared 80% Ethanol	Sigma-Aldrich

The above reagents 1 & 2 were thawed at room temperature, vortexed and centrifuged before proceeding with the protocol. The AMPure XP beads were mixed well before use.

1. Added 10  $\mu$ l of Resuspension Buffer to each well.
2. Added 40  $\mu$ l of End Repair Mix to the samples and mixed well by pipetting up and down.
3. Incubated the samples at 30 °C for 30 minutes in a pre-heated thermal cycler.
4. Removed the samples from the thermal cycler
5. Added 160  $\mu$ l of AMPure XP beads to the samples and mixed thoroughly by pipetting up and down.
6. Incubated the samples at room temperature for 15 minutes.
7. Samples were placed on magnetic stand at room temperature for 5 minutes.
8. Discarded the supernatant.

9. Beads were washed twice with 200  $\mu$ l of freshly prepared 80% ethanol for 30 seconds each.
10. Samples were removed from the magnetic stand.
11. The samples were left at room temperature for the beads to air dry.
12. Added 17.5  $\mu$ l of Resuspension Buffer to the samples.
13. Incubated the samples at room temperature for 2 minutes.
14. Samples were placed on magnetic stand at room temperature for 5 minutes.
15. Transferred 15  $\mu$ L of supernatant to a new well.
16. Stored the samples at -20 °C overnight.

## V. Adenylation of 3' ends

### Reagents

S. No.	Item	Supplier
1.	A-Tailing mix	Illumina Inc.
2.	Resuspension Buffer	Illumina Inc.

The above reagents were thawed at room temperature, vortexed and centrifuged before proceeding with the protocol.

1. The samples were thawed at room temperature and kept on ice.
2. Added 2.5  $\mu$ l of Resuspension Buffer to each well.
3. Added 12.5  $\mu$ l A-Tailing mix to the samples and mixed well.
4. The samples were placed in a thermal cycler and the following program was run.
  - a. Lid pre-heating – 100 °C
  - b. 37°C for 30 minutes
  - c. 70°C for 5 minutes
  - d. Hold at 4 °C
5. Samples removed from thermal cycler once it had cooled to 4 °C.

## VI. Ligation of adapters

### Reagents

S. No.	Item	Supplier
1.	Ligation Mix	Illumina Inc.
2.	Resuspension Buffer	Illumina Inc.
3.	Stop Ligation Buffer	Illumina Inc.
4.	AMPure XP beads	Beckman Coulter GmbH
5.	Freshly prepared 80% Ethanol	Sigma-Aldrich

The above reagents 1, 2 & 3 were thawed at room temperature, vortexed and centrifuged before proceeding with the protocol. The AMPure XP beads were mixed well before use.

1. Added 2.5 µl of Resuspension Buffer to each well.
2. Added 2.5 µl RNA Adapter Index to the samples and mixed well by pipetting up and down.
3. The samples were placed in a pre-heated thermal cycler and incubated at 30°C for 10 minutes.
4. Samples were removed from thermal cycler.
5. Added 5 µL of Stop Ligation Buffer to the samples and mixed to inactivate ligation.
6. Added 42 µl of AMPure XP beads to the samples and mixed thoroughly by pipetting up and down.
7. Incubated the samples at room temperature for 15 minutes.
8. Samples were placed on magnetic stand at room temperature for 5 minutes.
9. Discarded the supernatant.



10. Beads were washed twice with 200  $\mu$ l of freshly prepared 80% ethanol for 30 seconds each.
11. The samples were left at room temperature for the beads to air dry.
12. Samples were removed from the magnetic stand.
13. Added 52.5  $\mu$ l of Resuspension Buffer to the samples.
14. Incubated the samples at room temperature for 2 minutes.
15. Samples were placed on magnetic stand at room temperature for 5 minutes.
16. Transferred 50  $\mu$ L of supernatant to a new well.
17. Added 50  $\mu$ l of AMPure XP beads to the samples and mixed thoroughly by pipetting up and down.
18. Incubated the samples at room temperature for 15 minutes.
19. Samples were placed on magnetic stand at room temperature for 5 minutes.
20. Discarded the supernatant.
21. Beads were washed twice with 200  $\mu$ l of freshly prepared 80% ethanol for 30 seconds each.
22. The samples were left at room temperature for the beads to air dry.
23. Samples were removed from the magnetic stand.
24. Added 22.5  $\mu$ l of Resuspension Buffer to the samples.
25. Incubated the samples at room temperature for 2 minutes.
26. Samples were placed on magnetic stand at room temperature for 5 minutes.
27. Transferred 20  $\mu$ L of supernatant to a new well.

## VII. Enrichment of DNA fragments

### Reagents

S. No.	Item	Supplier
1.	PCR Master Mix	Illumina Inc.
2.	PCR Primer Cocktail	Illumina Inc.
3.	Resuspension Buffer	Illumina Inc.
4.	AMPure XP beads	Beckman Coulter GmbH
5.	Freshly prepared 80% Ethanol	Sigma-Aldrich

The above reagents 1, 2 & 3 were thawed at room temperature, vortexed and centrifuged before proceeding with the protocol. The AMPure XP beads were mixed well before use.

1. Added 5  $\mu$ l of PCR Primer Cocktail to each well.
2. Added 25  $\mu$ l PCR Master Mix to the samples and mixed well.
3. The samples were placed in a thermal cycler and the following program was run.
  - a. Lid pre-heating – 100 °C
  - b. 98 °C for 30 seconds
  - c. 15 cycles of
    - i. 98 °C for 10 seconds
    - ii. 60 °C for 30 seconds
    - iii. 72 °C for 30 seconds
  - d. 72 °C for 5 minutes
  - e. Hold at 10 °C

4. Samples removed from thermal cycler once it had reached 10 °C.
5. Added 50 µl of AMPure XP beads to the samples and mixed thoroughly by pipetting up and down.
6. Incubated the samples at room temperature for 15 minutes.
7. Samples were placed on magnetic stand at room temperature for 5 minutes.
8. Discarded the supernatant.
9. Beads were washed twice with 200 µl of freshly prepared 80% ethanol for 30 seconds each.
10. The samples were left at room temperature for the beads to air dry.
11. Samples were removed from the magnetic stand.
12. Added 32.5 µl of Resuspension Buffer to the samples.
13. Incubated the samples at room temperature for 2 minutes.
14. Samples were placed on magnetic stand at room temperature for 5 minutes.
15. Transferred 30 µL of supernatant to a new well.

### VIII. Validation of sample libraries

#### Reagents

S. No.	Item	Supplier
1.	Bioanalyzer	Agilent
2.	Agilent DNA 1000 chip	Agilent

1. Loaded 1 µl of sample into Agilent DNA 1000 chip and measured using the Agilent 2100 bioanalyzer instrument.

## IX. Normalization and pooling

### Reagents

S. No.	Item	Supplier
1.	Tris-HCl 10mM, pH 8.5 with 0.1 % Tween 20	Sigma-Aldrich

1. Transferred 10  $\mu$ L of sample into new wells.
2. Normalized the concentration of samples to 10 nM using Tris-HCl 10mM, pH 8.5 with 0.1% Tween 20.
3. Mixed thoroughly the normalized samples by pipetting up and down.
4. Pooled 16 different samples each containing unique adapter sequences ligated to them.
5. The pooled samples were processed further and loaded into the Illumina Inc. HiSeq 3000 sequencer by technicians.

## 2.8.5 MS sample preparation

1mg of protein was taken and processed before LC-MS/MS measurement. The protocol consists of the following steps.

- 1) Digestion of proteins
- 2) Digested peptides split for proteomics and phosphoproteomics
  - a. Phosphoproteomic samples
    - i. Acidification of peptides for phosphoproteomic analysis
    - ii. Phosphopeptide enrichment using TiO<sub>2</sub> beads
    - iii. Proceeded to purification of enriched Phosphopeptides on StageTips
  - b. Proteomic samples
    - i. Proceeded to purification of digested peptides on StageTips
- 3) Purification on StageTips
- 4) LC-MS/MS measurement of purified peptides.

### 1. Digestion of proteins

#### Reagents

S. No.	Item	Supplier
1.	Dithiothreitol	Amresco
2.	Iodoacetamide	Merck KGaA
3.	Endoproteinase Lys-C	WAKO
4.	Trypsin (sequence modified)	Promega GmbH

The above reagents were vortexed and centrifuged before proceeding with the protocol.

## MATERIALS AND METHODS

---

1. 1 mg of protein sample taken for processing.
2. Protein samples reduced with 1mM Dithiothreitol for 1 hour at room temperature.
3. Reduced protein sample were alkylated using 5.5 mM of Iodoacetamide for 1 hour at room temperature.
4. Protein samples pre-digested with endoproteinase Lys-C for 3 hours at room temperature.
5. Pre-digested samples diluted with 4 volumes of sterile water.
6. Samples digested overnight with trypsin.
7. Digested samples were split into two, for proteomics and phosphoproteomics.

### 2. Phosphoproteomic samples

#### a. Acidification of digested peptides

##### Reagents

S. No.	Item	Supplier
2.	Sep-Pak Vac 1cc C18 Cartridge columns	Waters
3.	Methanol	Merck KGaA
4.	Acetonitrile (ACN)	Merck KGaA
5.	Trifluoroacetic acid (TFA)	Merck KGaA
6.	Solvent A* - 2% ACN / 1% TFA	self-made by PCT
7.	Solvent A – 0.5% ACN	self-made by PCT

1. Sep-Pak Vac 1cc C18 cartridge columns were activated using Methanol.
2. Cartridge columns equilibrated with Solvent A\*.
3. Digested peptides loaded onto cartridge columns.
4. Washed the columns with Solvent A.
5. Eluted the peptides with 80% ACN / 6% TFA.

### b. Enrichment of phosphopeptides

#### Reagents

S. No.	Item	Supplier
5.	TiO <sub>2</sub> beads	ZirChrom
6.	Acetonitrile (ACN)	Merck KGaA
7.	Trifluoroacetic acid (TFA)	Merck KGaA
8.	Acetic Acid (AA)	Carl Roth GmbH
9.	NH <sub>3</sub> .H <sub>2</sub> O, pH 11.0	Merck KGaA
10.	Solvent B - 80% ACN / 0.5% AA	self-made by PCT

1. Eluted peptides were incubated with twice volume of TiO<sub>2</sub> beads for 10 mins.
2. The supernatant was transferred to a new batch of TiO<sub>2</sub> beads.
3. The initial batches of TiO<sub>2</sub> beads were washed with Solvent B.
4. The enriched phosphopeptides were eluted from the TiO<sub>2</sub> beads with 5% NH<sub>3</sub>.H<sub>2</sub>O, pH 11.0 and then 80% ACN / 2% TFA into 20% TFA.
5. Pooled all the eluted phosphopeptides and proceeded to purification.

### 3. Purification on StageTips

#### Reagents

S. No.	Item	Supplier
5.	C18 StageTips	C18 discs: 3M ("Empore Solid Phase Extraction disk"), StageTips: self-made by PCT
6.	Methanol	Merck KGaA
7.	Acetonitrile (ACN)	Merck KGaA
8.	Trifluoroacetic acid (TFA)	Merck KGaA
9.	Solvent A* - 2% ACN / 1% TFA	self-made by PCT
10.	Solvent A – 0.5% ACN	self-made by PCT

1. Activated C18 StageTips with of Methanol
2. Activated C18 StageTips with of Solvent A\*
3. Peptides were loaded onto equilibrated StageTips.
4. StageTips were washed with Solvent A.
5. Peptides were eluted with of 80% ACN / 6% TFA.
6. Proceeded to LC – MS/MS measurements.



## 2.8.6 Circular DNA synthesis

### Materials Used

S. No.	Item	Supplier
1.	Reverse Transcriptase enzyme	Promega GmbH
2.	Ribolock	Promega GmbH
3.	Random Hexamers	Sigma-Aldrich
4.	dNTPs	Promega GmbH
5.	RT Enzyme Buffer	Promega GmbH
6.	Ribolock Enzyme buffer	Promega GmbH
7.	Heating Block	Eppendorf GmbH

1. 1  $\mu\text{g}$  of RNA was taken for cDNA synthesis.
2. The volume of the RNA was adjusted to 8  $\mu\text{L}$  by adding sterile distilled water.
3. 5  $\mu\text{L}$  of random hexamers were added to the solution to prime the cDNA synthesis.
4. The mixture was incubated at 70  $^{\circ}\text{C}$  for 10 minutes.
5. The cDNA synthesis cocktail was added to the sample mixture.
6. They were incubated at room temperature for 10 minutes, after being mixed..
7. The samples were immediately transferred to a different heating block and

incubated at 42 °C for 45 minutes.

8. The samples were then incubated at 99 °C for 3 minutes.
9. The samples were then transferred to ice. The concentration of the cDNA was adjusted by adding 30 µL of sterile water.

### 2.8.7 PCR product purification

The amplified PCR fragments were purified, in order for them to be sequenced. The GeneJET PCR purification kit was used for this purpose.

#### Reagents Used

S. No.	Buffer/ Instrument	Company/Composition
1.	Binding Buffer	Thermo Scientific™
2.	Isopropanol	Thermo Scientific™
3.	Wash Buffer	Thermo Scientific™
4.	Elution Buffer	Thermo Scientific™

1. The PCR amplicons were mixed with equal volume of Binding Buffer.
2. When the size of the PCR amplicons was less than 500 bp, half the volume of isopropanol was added to the above mixture.
3. Transferred the mixture into purification columns.
4. The columns were centrifuged at 13,000 rpm for 1 minute. The flow-through was discarded.
5. 700 µL of Wash buffer was added to the columns.

6. The columns were centrifuged at 13,000 rpm for 1 minute. The flow-through was discarded.
7. The purified amplicons were eluted by adding 30  $\mu$ L of Elution buffer to the purification columns, which were placed on microcentrifuge tubes.
8. The concentration of the eluted amplicons was measured using the Nanodrop.

### 2.8.8 Agarose gel electrophoresis

The amplified PCR fragments or the freshly extracted DNA/RNA were subjected to agarose gel electrophoresis to check for quality.

#### Reagents Used

S. No.	Buffer/ Instrument	Company/Composition
1.	Agarose	Geneaxxon bioscience
2.	TAE Buffer	242.28 g Tris 18.61g EDTA 60.05 g Acetic acid filled upto 1L with water,pH 8.5
	Ethidium Bromide	Applichem

1. The required amount of agarose was weighed in glass containers and dissolved in the corresponding amount of TAE Buffer by boiling. The amounts of agarose required for different gel percentages are given below.

**Agarose amount table**

% Gel	Agarose (g)	TAE Buffer (mL)	EtBr ( $\mu$ L)
2	2.4	120	2.5
1.5	1.5	120	2.5
1	1.2	120	2.5
0.8	1	120	2.5

2. The agarose solution dissolved in TAE buffer was allowed to cool down at RT.
3. 2.5  $\mu$ L of Ethidium Bromide was added to the solution.
4. The solution was then poured carefully into the preset gel casting plates containing the gel combs. The gel casting setup was maintained at 4 °C.
5. The agarose was allowed to completely solidify.
6. The gel combs were carefully removed.
7. The agarose gels were transferred to the gel running chambers containing TAE Buffer.
8. The samples premixed with one-fifth the volume of loading dye were loaded into each of the wells. DNA ladder was also loaded into one of the wells, to serve as size marker.
9. The samples were allowed to separate in the gel chambers at an electric field of 100 V for 45 minutes to 1 hour until the samples had covered more than three quarters of the length of the gel.
10. Once the samples had separated, the images of the separation patterns were captured using the Gel documentation apparatus.

## 2.9 Softwares / online tools used

S.No.	Software/Tool	Purpose/Usage	Platform
1.	cBio Portal	Access TCGA and other human cancer datasets	Online Portal
2.	TCGA Assmebler	Download of TCGA data	R Package
3.	DAVID	Functional analysis of gene sets	Online Tool
4.	GSEA	Gene set enrichment analysis	Software
5.	KEGG Pathways	Pathway analysis using gene candidates	Online tool
6.	ABI 7500 software	Reatlime PCR analysis	Software
7.	Chromas Lite	Sangers sequencing analysis	Software
8.	4 Peaks	Sangers sequencing analysis	Software
9.	Biomart	Conversion of ids	Online Tool
10.	Cluster 3.0	Hierarchical clustering	Software
11.	Treeview	Clustering analysis - visualization	Software
12.	PROVEAN/SIFT	Amino Acid prediction of mutations	Online Tool
13.	IGV viewer	Visualisation of Reads	Software
14.	Primer BLAST	Primer designing	Online Tool
15.	Promotion	CArG box prediction	Online Tool

## 2.10 Primer sequences

Primer	Sequence (5' -> 3')	Purpose
<b>Genotyping</b>		
Rosa26 forward	CACCAGGTTAGCCTTTAAGCCTGCCC	SRF-VP16 genotyping
Rosa26 reverse	GGAGGCAGGAAGCACTTGCTCTCC	SRF-VP16 genotyping
PURO reverse	GAACGAGATCAGCAGCCTCTGGGTCCAC	SRF-VP16 genotyping
Alfp-CreER <sup>T2</sup> forward	GCAAACATACGCAAGGGAT	Alfp-CreER <sup>T2</sup> genotyping
Alfp-CreER <sup>T2</sup> reverse	CACAGTCAGCAGGTTGGAGA	Alfp-CreER <sup>T2</sup> genotyping
PCR3	AAGAAGGGTCCGGCCCCGAAGATGCTGGGC	Alfp-CreER <sup>T2</sup> genotyping
PCR4	CTGGATGCCCTCTCCTTCCCCGGAGCCCTG	Alfp-CreER <sup>T2</sup> genotyping
Srf-KO-8- forward	CCGGGAAATATGGGGAGAGGGGAGAT	<i>Srf-flex1</i> genotyping
Srf-KO-8- reverse	CTTCGCGCACACCAGGACACAGAGGAT	<i>Srf-flex1</i> genotyping
Lox2 forward	GCTCGCAGCGGCGGCCAGATCTATAAC	<i>Srf-flex1</i> genotyping
(continued ... )		

<b>Mutation Detection</b>		
<i>Ctnnb1</i> forward	GTCGTAGATGGCTTCTTCAGG	<i>Ctnnb1</i> genomic amplification
<i>Ctnnb1</i> reverse	TCCTTCAACCCACTTGTGC	
<i>Trp53</i> forward	GTCACCTGTAGTGAGGTAGG	<i>Trp53</i> genomic amplification
<i>Trp53</i> reverse	ACTCGTGGAACAGAAACAGG	
<i>Ctnnb1_cDNA_</i> forward	GTCAGCTCGTGCCTGTGAA	<i>Ctnnb1</i> transcriptome amplification
<i>Ctnnb1_cDNA_</i> reverse	CAGTGTCGTGATGGCGTAGA	
<i>Trp53_cDNA_</i> forward	GGCCCTCATCCTCCTCCTTC	<i>Trp53</i> transcriptome amplification
<i>Trp53_cDNA_</i> reverse	ACTGGCCCTTCTTGGTCTTCA	
<b>Quantitative PCR</b>		
<i>Ctnnb1_forward</i>	GTGCAATTCCTGAGCTGACA	<i>Ctnnb1</i> qPCR amplification
<i>Ctnnb1_reverse</i>	CTTAAAGATGGCCAGCAAGC	

## MATERIALS AND METHODS

---



## 3 Results

### 3.1 The *SRF-VP16<sup>iHep</sup>* mouse model

#### 3.1.1 Genotype

The genome of *SRF-VP16<sup>iHep</sup>* mice consists of three transgenes inserted into their wildtype genome. The transgenes are as follows.

- *SRF-VP16* transgene (referred also as *SRF-VP16* allele)
- *Alfp-CreER<sup>T2</sup>* transgene (referred also as Cre allele)
- *Srf-flex1* allele, replacing the endogenous *Srf* wt allele

##### 3.1.1.1 *SRF-VP16* transgene

The *SRF-VP16* transgene comprises of the human SRF cDNA sequence, who's transcriptional transactivation domain (TAD) has been partly replaced with a transcriptional activator of the human Herpes Simplex virus (HSV). This transcriptional activator protein is called Viral Protein 16 (VP16). The cDNA of the HSV VP16 fused to the truncated human SRF cDNA sequence generates the *SRF-VP16* transgene, which encodes the fusion protein SRF-VP16 (Ohrnberger et al., 2015). The transgene construct containing the *SRF-VP16* coding sequence is shown in Figure 3-1.

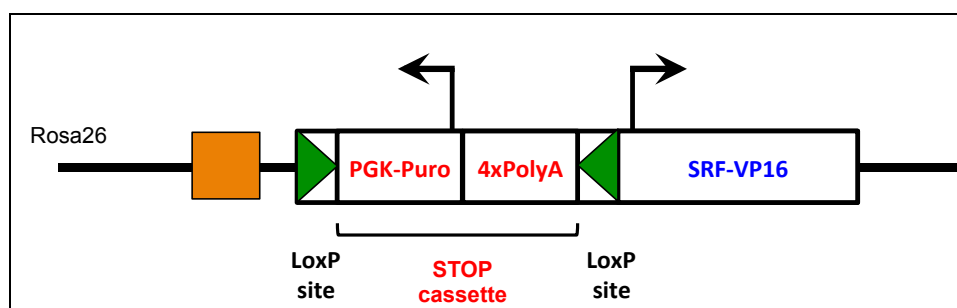


Figure 3-1: The *SRF-VP16* transgene construct. The *SRF-VP16* transgene construct is inserted in the *Rosa26* locus of the mouse genome. The construct consists of a STOP cassette flanked by LoxP sites, which prevents the spontaneous expression of *SRF-VP16* transgene. The black arrows denote the translational start sites and directions of transcription. The LoxP sites are represented in green triangles (Ohrnberger\*, Thavamani\* et al., 2015).

### 3.1.1.2 *Alfp-CreER<sup>T2</sup>* transgene

The *Alfp-CreER<sup>T2</sup>* transgene is responsible for the expression of the Cre recombinase enzyme in *SRF-VP16<sup>iHep</sup>* mice. Expression and nuclear localization of the Cre enzyme is essential for recombination-mediated deletion between the LoxP sites flanking the STOP cassette of the *SRF-VP16* transgene construct. This recombination causes the deletion of the STOP cassette, resulting in activation of *SRF-VP16* transgene expression. In the *CreER<sup>T2</sup>* construct, the Cre recombinase is fused to a mutant Estrogen Receptor protein, known as ER<sup>T2</sup>, which – upon tamoxifen application – translocates from the cytoplasm to the nucleus, and thus completes the regulation mechanism of tamoxifen-induced activation of *SRF-VP16* transcription. However, the observed leakiness of Cre activity in hepatocytes of *SRF-VP16<sup>iHep</sup>* mice enables it to enter the nucleus even in the absence of tamoxifen, upon rare circumstances. This results in a mosaic activation of *SRF-VP16* expression in a hepatocyte-specific manner in *SRF-VP16<sup>iHep</sup>* mice. The *Alfp-CreER<sup>T2</sup>* transgene construct contains the Albumin enhancer and promoter sequences that promote and regulate expression of the *Alfp-CreER<sup>T2</sup>* transgene exclusively in hepatocytes. The construct containing the *Alfp-CreER<sup>T2</sup>* transgene is represented in the Figure 3-2.

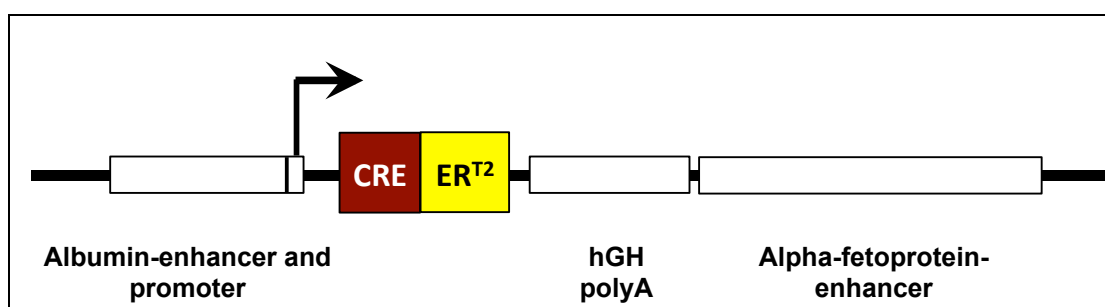


Figure 3-2: The *Alfp-CreER<sup>T2</sup>* transgene construct (Ohrnberger\*, Thavamani\* et al, 2015). The black arrow represents the transcription start site and the direction of transcription.

*SRF-VP16<sup>iHep</sup>* mice contain both the *Alfp-CreER<sup>T2</sup>* and *SRF-VP16* transgenes. These animals go on to develop HCCs in the liver after the age of about 20-25 weeks. There is a third transgene present in the genome *SRF-VP16<sup>iHep</sup>* mice, namely *Srf-flex1*, representing a floxed allele of the endogenous *Srf* gene.

### 3.1.1.3 *Srf-flex1* allele

The wildtype SRF protein can competitively inhibit the SRF-VP16 fusion protein, by occupying the CARG-box binding sites of SRF-VP16 on target DNA. The genomic replacement of the wildtype *Srf* allele with the *Srf-flex1* allele effectively eliminates wildtype SRF protein activity. The *Srf-flex1* allele consists of two LoxP sites; the first LoxP site is present within the 5' untranslated region of the *Srf* gene, whereas the second LoxP site lies within the intron 1 of the *Srf* gene. The Cre-mediated recombination results in a deletion of the complete coding region of *Srf* exon 1 (Wiebel et al., 2002). A homozygous deletion effectively nullifies SRF protein activity in the cells. Figure 2-3 describes the *Srf-flex1* allele.

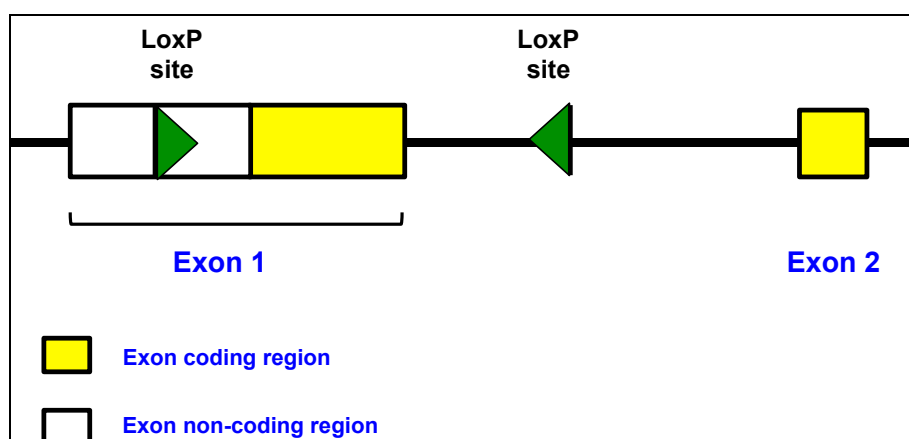


Figure 3-3: *Srf-flex1* allele. The LoxP sites are represented as green triangles. The yellow boxes refer to the coding regions of *Srf* exons and the white boxes the non-coding regions. The first LoxP site is present within the non-coding region of exon 1 and the second LoxP site is present within intron 1 of the *Srf* gene.

The presence of the *SRF-VP16* and the *Alfp-CreER<sup>T2</sup>* transgenes is necessary for hepatocarcinogenesis in *SRF-VP16<sup>iHep</sup>* mice, whereas the presence of the *Srf-flex1* allele ensures the minimum interference to SRF-VP16 activity by wildtype SRF protein. *SRF-VP16<sup>iHep</sup>* mice develop HCCs in their liver irrespective of the presence or absence of the *Srf-flex1* allele, however homozygous replacement of wildtype *Srf* by *Srf-flex1* accelerates HCC formation.

### 3.1.1.4 *SRF-VP16<sup>iHep</sup>* genotype

The genotype of *SRF-VP16<sup>iHep</sup>* mice encompasses the heterozygous presence of the *SRF-VP16* and *Alfp-CreER<sup>T2</sup>* transgenes, associated with replacement of endogenous *Srf* alleles by the *Srf-flex1* allele.

In the breeding scheme used, the *Srf-flex1* allele could either be homozygous, heterozygous, or even completely absent in the *SRF-VP16<sup>iHep</sup>* genome. Thus, the *SRF-VP16<sup>iHep</sup>* mouse comprises of the following three possible genotypes.

*SRF-VP16* (+/-) : *Alfp-CreER<sup>T2</sup>* (+/-) : *Srf-flex1* (-/-)

*SRF-VP16* (+/-) : *Alfp-CreER<sup>T2</sup>* (+/-) : *Srf-flex1* (+/-)

*SRF-VP16* (+/-) : *Alfp-CreER<sup>T2</sup>* (+/-) : *Srf-flex1* (+/+)

### 3.1.1.5 *SRF-VP16<sup>iHep</sup>* mating scheme

The mating for the *SRF-VP16<sup>iHep</sup>* mice was carried out according to the scheme described in Figure 3-4.

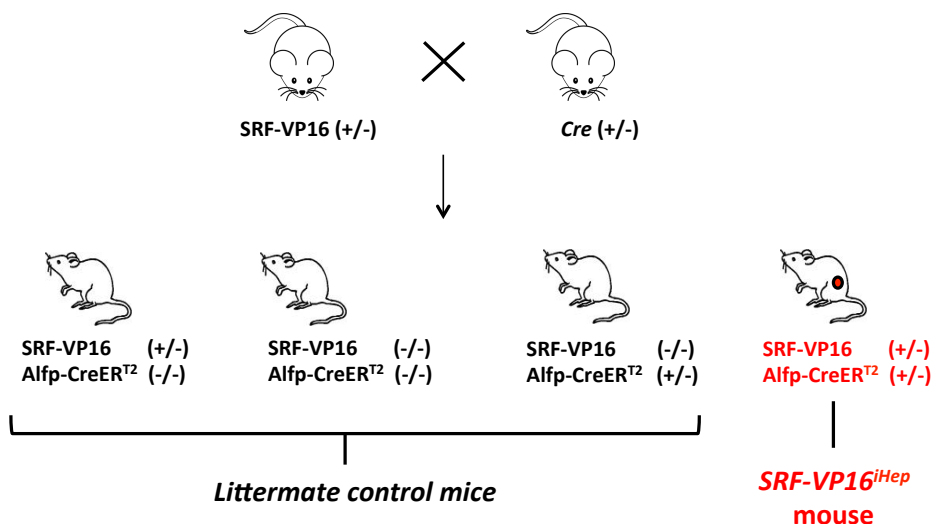


Figure 3-4: Mating scheme of *SRF-VP16<sup>iHep</sup>* mice. Mice heterozygous for the *SRF-VP16* transgene were crossed with mice heterozygous for the *Alfp-CreER<sup>T2</sup>* transgene. The probability of a mouse being born as heterozygote for both of the transgenes is 25%. The genotype of interest is obtained when both the *SRF-VP16* and *Alfp-CreER<sup>T2</sup>* transgenes are present as heterozygous alleles in the genome.

The *Srf-flex1* transgene genotype in the  $SRF-VP16^{iHep}$  mice could be homozygous, heterozygous or null.

$SRF-VP16^{iHep}$  mice containing the transgenes *SRF-VP16* and *Alfp-CreER<sup>T2</sup>* were separated into different Type II long cages, along with the corresponding litter-mate control mice. To keep the mouse-line replenished, healthy littermate control mice that are heterozygous for either of the transgenes *SRF-VP16* or *Alfp-CreER<sup>T2</sup>* were chosen and bred with the mice carrying the complementary genotype. For each of the experiments using the  $SRF-VP16^{iHep}$  mouse strain, at least one littermate control mouse was taken. Figure 3-5 describes the usage of the mice for HCC characterization.

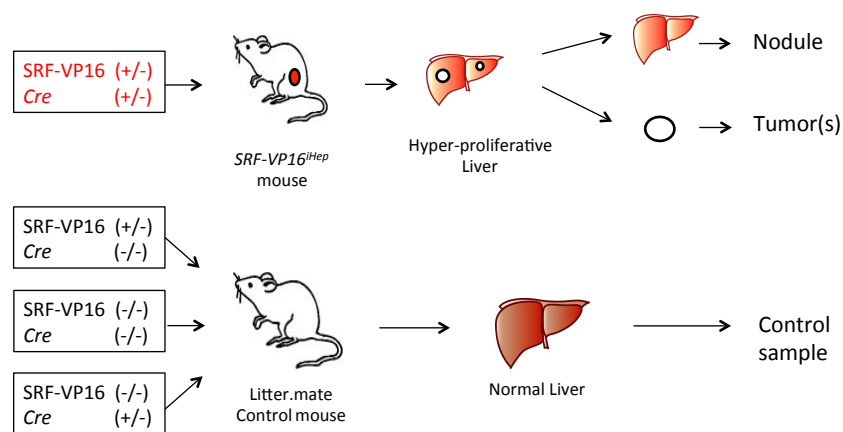


Figure 3-5: Experimental Mice genotypes and their usage for sample collection. Mice with the genotype of interest (red) are let to grow until liver tumors are developed in the  $SRF-VP16^{iHep}$  mice, at which time they are sacrificed for organ harvesting. The livers, containing tumors and hyper-proliferative nodular foci are taken for subsequent experiments. At the same time, the littermate control mice were also sacrificed and their livers were collected as control tissue samples.

### 3.1.2 Phenotypic observations

$SRF-VP16^{iHep}$  mice did not show any significant phenotypic abnormalities until the age of about 20 to 30 weeks, when the belly started bulging owing to the increasing liver size. However, the general health status of the mice was not compromised. The health of the mice was monitored by observing their body weight, movement activities and posture. The mice were examined regularly for any health impairments. The weight gain pattern of one  $SRF-VP16^{iHep}$  mouse, compared to that of a littermate control, is

## RESULTS

---

shown in figure 3-6. The weight gain pattern of the control mice correlates to that of the SRF-VP16 expressing mice until about the age of 20-weeks.

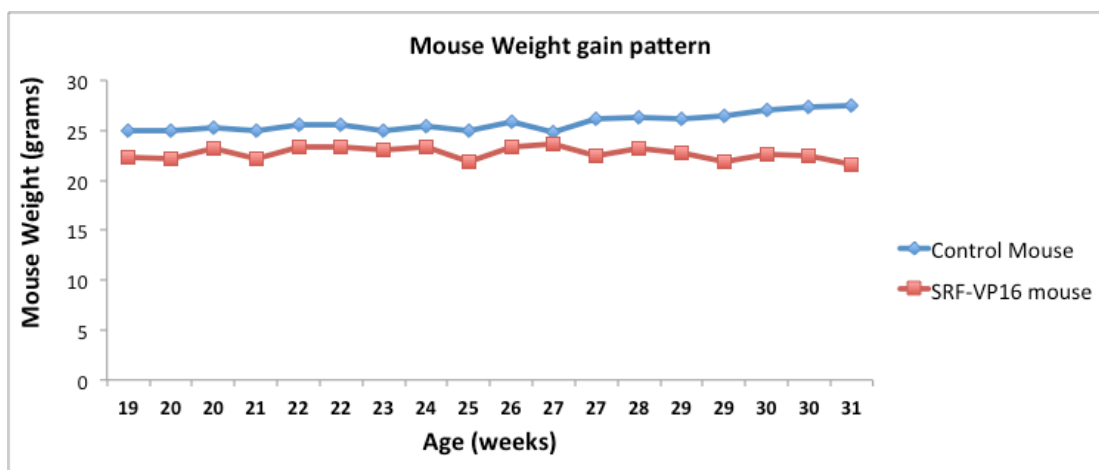


Figure 3-6: Weight gain patterns of individual mice. The weight gain patterns of SRF-VP16<sup>iHep</sup> and control mice show very little differences, until a time point starting at about 27 weeks of age.

After the age of 27 weeks, depending upon the individual animal, increasing proliferation in the liver results in a decrease in weight gain, accompanied by other obvious symptoms. This can be checked by observing behavioral features of the mouse in context of its movement and also its stature during movement, possibly suggesting physical discomfort or pain. When two or more of these symptoms begin to be displayed by an individual SRF-VP16<sup>iHep</sup> animal, it is ready to be sacrificed in order to prevent further discomfort.

### 3.1.3 Liver weight to body weight ratio

At the endpoint of the study, individual SRF-VP16<sup>iHep</sup> mice were sacrificed and their livers harvested. Animal weight before dissection and liver weight immediately upon harvested were scored. One of the most obvious and earliest observations were the increased sizes of the livers due to SRF-VP16 induced hyper-proliferation of hepatocytes. In these livers, large tumors could be seen and also multiple foci that consisted of hyper-proliferative cells. From such hyper-proliferative liver tissues full-

blown tumors could develop. The increased liver sizes also corresponded to the increased liver weights, which could be compared to the weights of the individual mice, in order to compute the liver weight to body weight (LWBW) ratios. Figure 3-7 shows a graph plotting the LWBW ratios of *SRF-VP16<sup>iHep</sup>* mice, compared to the LWBW ratios of corresponding littermate control mice.

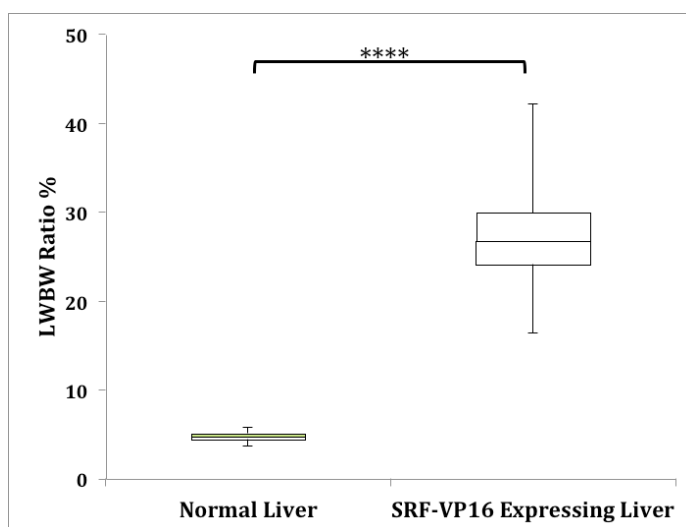
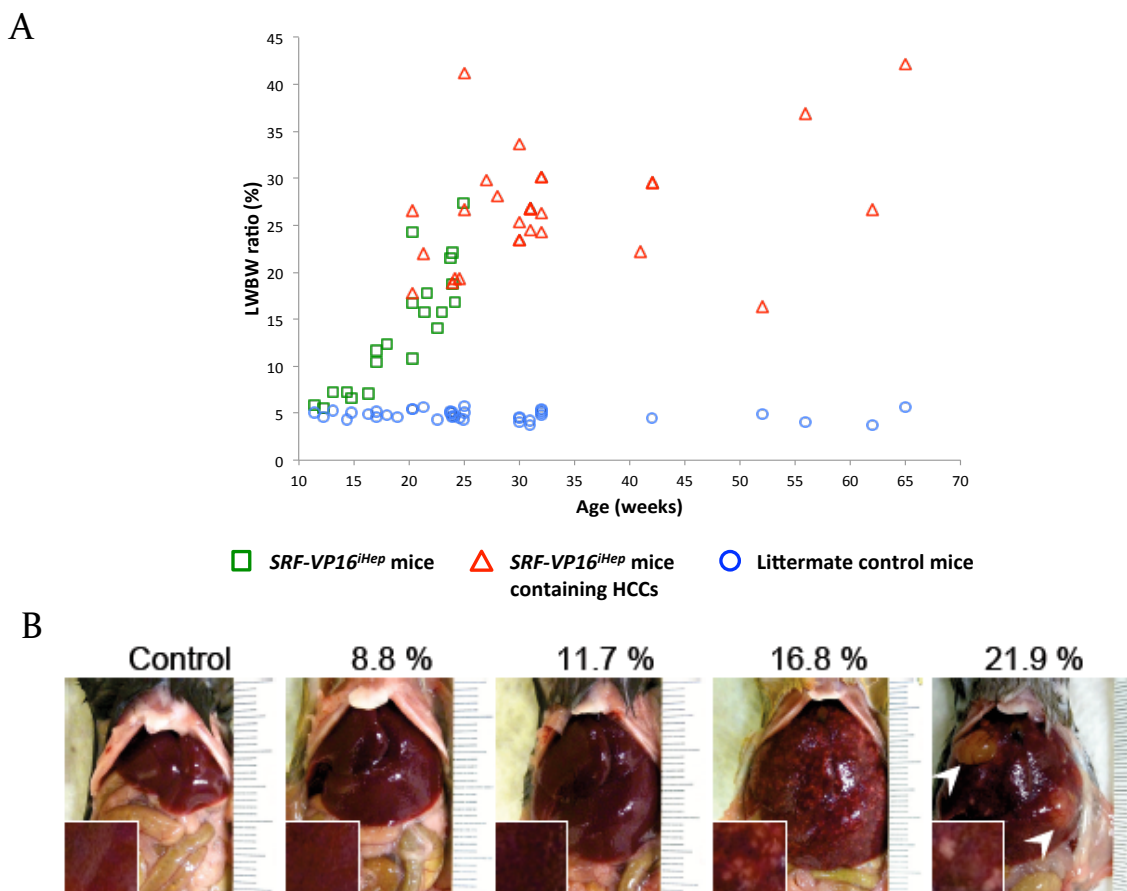


Figure 3-7: Comparison of Liver weight to body weight (LWBW) ratios. LWBW ratios of mice above the age of 25 weeks and manifested HCCs in their livers are plotted against the LWBW ratios of normal liver sample. The graph shows the difference in the liver weight to body weight ratios of *SRF-VP16<sup>iHep</sup>* mice (right) in comparison control mice (left).  $N = 42$ , for each genotype.

### 3.1.3 Hepatocellular carcinomas in liver

In *SRF-VP16<sup>iHep</sup>* mice the multiple hyper-proliferative foci ultimately result in the formation of tumors that could be classified by liver pathologists (Profs. Schirmacher and Longerich, Pathology University Clinic Heidelberg) as hepatocellular carcinoma (HCC). The number of tumors found in a particular liver can vary with each mouse, ranging from 1 to 5. The sizes of these tumors were also variable but increased with age.

## RESULTS



*Figure 3-8: Age-dependent progression toward HCC formation. A) The continuous increase of the liver to body weight ratios with advancing age can be noticed from the graph. The green squares denote the LWBW ratios of SRF-VP16<sup>iHep</sup> mice and the red triangles denote the LWBW ratios of SRF-VP16<sup>iHep</sup> mice that contained HCCs in their livers. The blue circles denote the LWBW ratios of the littermate control mice. B) The images show the simultaneous increase in numbers of hyper-proliferative pre-malignant foci, with increasing LWBW ratios. The LWBW ratios in percentages are displayed above each image. In the far right image, the ultimate development of liver tumors can be seen. The white arrows point to the tumors in the liver. Picture Source: Ohrnberger\*, Thavamani\* et al, 2015.*

The Figure 3-8 shows the age-dependent progression towards HCC formation observable by the gradually increasing LWBW ratios due to the increasing number of hyper-proliferative foci. However, the littermate control mice did not show any increase in their LWBW ratios, thus emphasizing the role of SRF-VP16 in triggering this malignant transformation in SRF-VP16<sup>iHep</sup> livers leading to HCCs. In Figure 3-8a are also included the LWBW ratios of mice that had been sacrificed by Dr. Stefan Ohrnberger (former postdoc, Nordheim Lab, University of Tuebingen). The frequency of occurrence



of HCCs in *SRF-VP16<sup>iHep</sup>* mice after the age of 25 weeks is significantly higher than at any other earlier time-point. This can be seen, in the Kaplan-Meier plot of tumor occurrence shown in Figure 3-9. The tumors and nodules were collected for further downstream experiments.

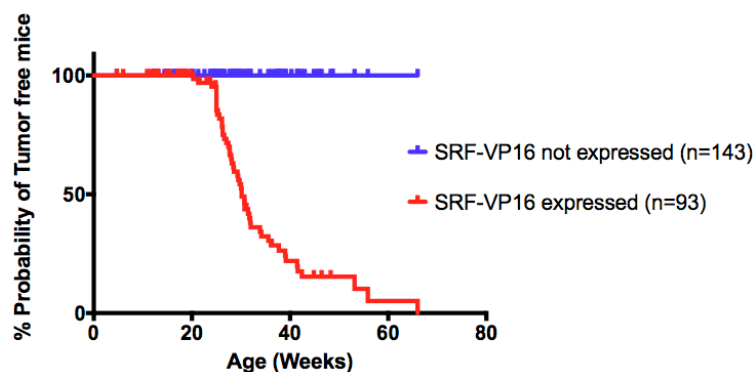


Figure 3-9: HCC Penetrance in *SRF-VP16<sup>iHep</sup>* mice. The Kaplan Meier graph shows the % probability of mice to have no HCCs in its liver at the time of sacrifice, at increasing ages. The HCC penetrance was found to be 100 % in the *SRF-VP16<sup>iHep</sup>* mice beyond the age of about 25 weeks. Litter mate control mice did not show any signs of tumor formation. Picture Source: Ohrnberger\*, Thavamani\* et al, 2015.

## 3.2 Mutation detection

To investigate the potential manifestation of genomic mutations in HCC tumor tissues, my initial investigation focused on the analysis of mutations in an oncogene (*Cttnb1*) and a tumor suppressor gene (*Trp53*), known to be frequently mutated in the genomes of murine and human liver cancers.

### 3.2.1 Frequency of mutations

Genomic hotspot regions of mutations in the *Cttnb1* and *Trp53* genes were amplified and sequenced. Mutations were detected in both the gene loci. The frequency of mutations in *SRF-VP16<sup>iHep</sup>* mice is shown in the Table 3-1.

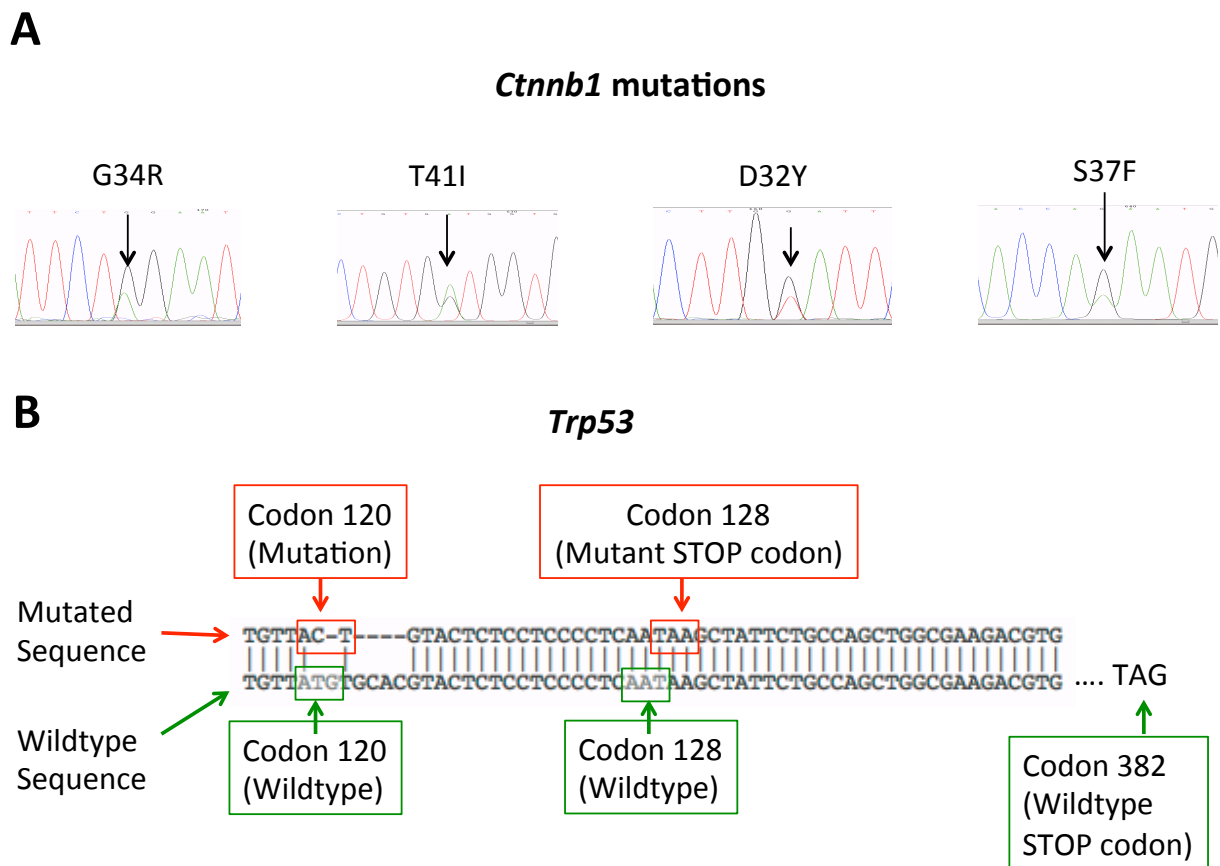
## RESULTS

---

Table 3-1: Mutation status of tumors from SRF-VP16<sup>iHep</sup> mice at *Cttnb1* and *Trp53* gene loci. In the table, the number of samples found to be mutant (Mt) or wild type (Wt) for the corresponding loci are shown. In more than 50 samples tested (#), *Cttnb1* mutations were found at a frequency of about 28 %. *Trp53* mutations were occurring at a frequency of 18%.

<b>Gene</b>	<b><i>Cttnb1</i></b>		<b><i>Trp53</i></b>	
<b>Status</b>	<b>Mt</b>	<b>Wt</b>	<b>Mt</b>	<b>Wt</b>
Samples #	15	39	9	42
Total #	54		51	
(%)	(28%)		(18%)	

Mutations observed in the *Cttnb1* locus were more frequent than those in the *Trp53* locus. In the *Cttnb1* gene locus, different mutations were detected. The observed point mutations caused the following coding changes: G<sub>34</sub>R, T<sub>41</sub>I, D<sub>32</sub>Y and S<sub>37</sub>F. These mutations can be predicted to cause the loss of  $\beta$ -Catenin protein phosphorylation by the GSK-3 $\beta$  kinase. This can be predicted to result in the protein not undergoing degradation via the ubiquitin-proteasome pathway (Aberle et al., 1997), leading to stabilization of the  $\beta$ -Catenin protein within the cell. Hence these mutations can be expected to serve as activating mutations of the  $\beta$ -Catenin protein and hence the Wnt signaling pathway (de La Coste et al., 1998).



*Figure 3-10: Ctnnb1 and Trp53 mutations. A) The different mutations observed in the Ctnnb1 locus, affecting the codons 32, 34, 37 and 41. Chromatograms of the sequencing results show heterozygous non-synonymous mutations in the above 4 codons. B) Insertion/deletion mutation in codon 120 of the Trp53 gene, compared to the wildtype sequence. The ensuing frameshift in the mutant allele results in a premature STOP codon at position 128. In the wildtype sequence the STOP codon occurs at position 382.*

As show in Figure 3-10, at the *Trp53* gene locus, an in-frame deletion was detected, which could result in a frame shift during transcription that can result in the formation of a premature stop codon. The ensuing truncated protein might result in the inactivation of the tumor suppressor P53 protein. This can be expected to further assist in HCC tumor progression.

### 3.2.2 *Ctnnb1* mutation associated gene expression changes

Quantitative real time PCR analysis of the *Ctnnb1* mutant and *Ctnnb1* wildtype tumor samples was performed, in order to evaluate mRNA expression levels of *Ctnnb1* target genes.

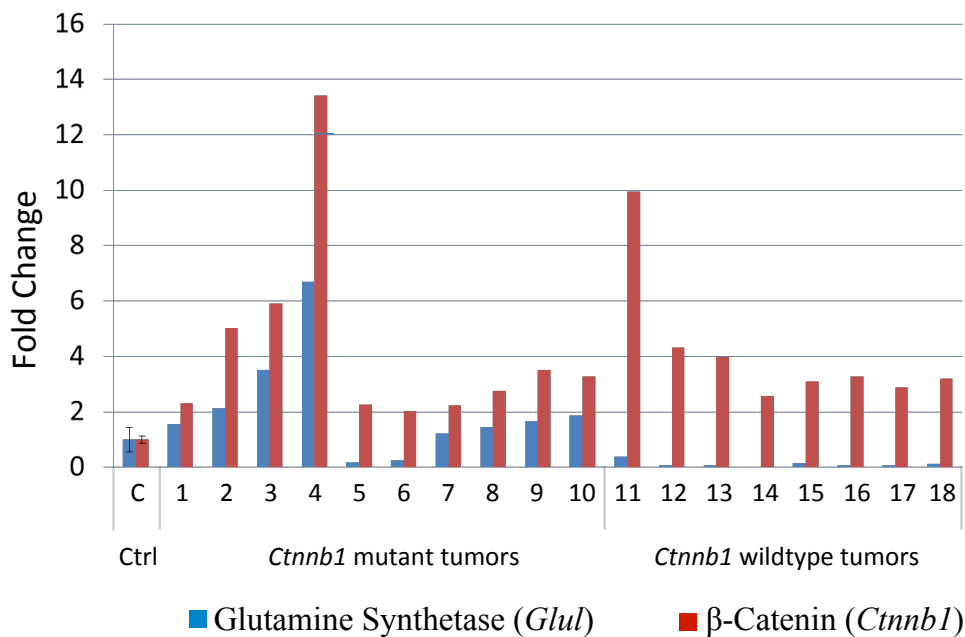


Figure 3-11: *Glu1* and *Ctnnb1* expression levels. Plotted are the fold-change expression values of different tumor samples for the genes *Glu1* and *Ctnnb1*, normalized against the control liver expression. The gene *Glu1*, encoding glutamine synthetase, was found specifically up-regulated in the *Ctnnb1* mutant tumors. However, the *Ctnnb1* expression itself was found to be variable between tumor samples.

It can be deduced from Figure 3-11 that mutations in the *Ctnnb1* gene were paralleled by enhanced expression of *Glu1*. It can also be noted that in the *Ctnnb1* wild type tumors there was reduced expression of *Glu1*. This could be due to the changes in cellular polarity and differentiation caused by *Ctnnb1*-stimulated malignant transformation. These observations warranted a genome-wide investigation of gene expression profiles. Henceforth, the *Ctnnb1* mutant tumors were classified as HCC<sub>A</sub> and the *Ctnnb1* wildtype tumors were classified as HCC<sub>B</sub>.

### 3.3 Gene expression profiling - microarray analysis

In order to investigate at a genome-wide level the mRNA expression levels of genes, microarray analysis was performed. 12 different samples (of different sample types) were analyzed in collaboration with Dr. Michael Bonin, Human Genetics, UKT, University of Tuebingen. The demographics of the samples analyzed are given below.

*Table 3-2: Sample Demographics - Microarray analysis.  $HCC_A$  denotes *Cttnb1* wt tumors, while  $HCC_B$  denotes *Cttnb1* mutant tumors. LWBW stands for Liver Weight to Body Weight Ratio at time of sample collection. Nodules stand for pre-malignant, hyper-proliferative nodular liver tissue.*

Category	Type	Sample #	Mouse #
Gender	Male	6	6
	Female	6	6
Sample type	Controls	3	3
	Nodules	3	3
	$HCC_A$	3	3
	$HCC_B$	3	3
Age	Median Age (weeks)	27.5 weeks	
LWBW ratio	Average LWBW ratio % (Tumors & Nodules)	23.2%	

For the downstream analysis of the microarray expression data, the p-value cut-off was set at 0.05 and the fold-change cut-off was set at 2. Hence only the transcripts that displayed a difference in expression levels by at least 2 fold, compared to the controls, with a high significance (p-value <0.05), were considered for the downstream analyses. About 3000 genes were found differentially expressed (compared to control liver tissue) across all the different sample categories (Ohrnberger et al., 2015). The number of differentially regulated transcripts and their overlaps across different groups are shown in Figure 3-12.

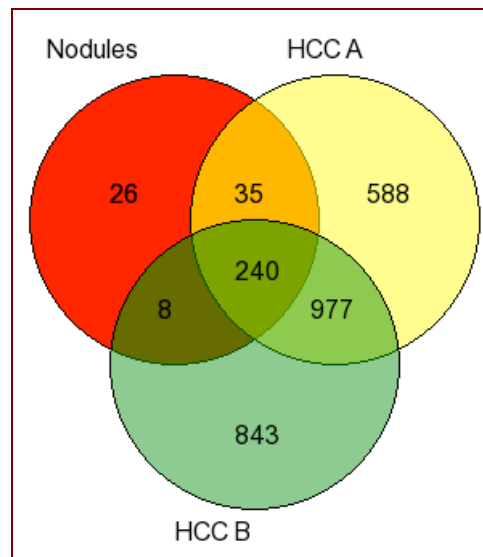


Figure 3-12: *Dysregulated Transcripts – Microarray analysis.* The overlap between the different sample groups is shown.  $HCC_A$  denotes *Cttnb1* wt tumors, while  $HCC_B$  denotes *Cttnb1* mutant tumors. Nodules stand for pre-malignant, hyper-proliferative nodular liver tissue.

### 3.3.1 Sub-lists – microarray data

#### 3.3.1.1 Early driver alterations

The following are some of the interesting gene lists found differentially expressed in all the different sample groups, probably implying their involvement in tumor formation, starting at the early pre-malignant nodular stages of tumor development. There were interesting candidates in the top 5 list of genes up-regulated across all sample categories, including *Igf2*, *H19*, and *CD63* among others. These genes are also known as the oncofetal liver genes, which are highly overexpressed in a specific subset of human HCC patients (Ohrnberger et al., 2015).

#### 3.3.1.2 Late driver alterations

Another group of genes was noticed, which were highly deregulated in the tumors tissues but not, or only slightly, altered at nodular stages.

**Table 3-3: The 5 most dramatically up- or down-regulated genes.** The highly deregulated samples across all groups of tissue analyzed by microarray expression analysis of mRNAs.  $HCC_A$  denotes *Cttnb1* wt tumors, while  $HCC_B$  denotes *Cttnb1* mutant tumors. Nodules stand for pre-malignant, hyper-proliferative nodular liver tissue.

Gene Symbol	Nodules vs Control Log2 Fold Change	HCC A vs Control Log2 Fold Change	HCC B vs Control Log2 Fold Change
<i>H19</i>	5.1	6.2	6.4
<i>Vil1</i>	3.2	4.6	4.3
<i>Cd63</i>	4.3	5.0	5.1
<i>Igf2</i>	3.7	6.7	6.7
<i>Cdh1</i>	3.1	3.2	3.3
<i>Cyp4a14</i>	-4.5	-6.6	-4.3
<i>Inhbe</i>	-2.0	-3.1	-5.0
<i>Oat</i>	-2.6	-4.5	-3.3
<i>Elovl3</i>	-2.3	-5.2	-5.4
<i>Sucnr1</i>	-2.3	-4.4	-4.8

Legend

Log2 Fold Change values				
<-3	-3 to -1	-1 to 1	1 to 3	>3

**Table 3-4: Genes dysregulated only in HCCs and not in Nodules– Microarray analysis.** Selected genes which were up- or down-regulated in tumors, with little or no dysregulation during early (nodular) stages of HCC progression are shown here.  $HCC_A$  denotes *Cttnb1* wt tumors, while  $HCC_B$  denotes *Cttnb1* mutant tumors. The color legend is the same as that used in Table 3-3.

Gene Symbol	Nodules vs Control Log2 Fold Change	HCC A vs Control Log2 Fold Change	HCC B vs Control Log2 Fold Change
<i>Tspan8</i>	1.6	6.0	4.2
<i>A2m</i>	1.1	5.2	4.2
<i>Bex1</i>	2.1	6.1	6.4
<i>Timp1</i>	0.7	5.6	4.0
<i>Cpe</i>	0.7	5.5	5.7
<i>Mup3</i>	-0.2	-3.9	-7.3
<i>Es31</i>	0.1	-5.5	-6.7
<i>Gls2</i>	-0.6	-3.4	-5.3
<i>Ces1</i>	-0.6	-3.1	-4.2
<i>Apoa5</i>	-0.4	-4.5	-3.9

## RESULTS

The genes that are specifically deregulated in the tumor stage reflect very likely changes in gene expression that were acquired later during progression toward malignant transformation. Interestingly, the genes *Tspan8*, *Timp1* and *Cpe* have been shown by different studies to be over-expressed in various cancers (Wei et al., 2015) (Toricelli et al., 2017) (Murthy et al., 2013). The gene *Bex1* is a member of the group of oncofetal liver genes, which have been found overexpressed abundantly in nodules and tumors of *SRF-VP16<sup>iHep</sup>* mice. The gene *Timp1* has been reported to be overexpressed in different cancers, but has also been reported in liver cancers, where it has been shown to mediate the communication between tumors and their respective microenvironment, especially between the cancer cells and the hepatic stellate cells via the TGF- $\beta$ -dependant FAK signaling pathway (Park et al., 2015).

### 3.3.1.3 Genes over-expressed in pre-malignant nodular tissue only

Specific sets of genes could be identified, which were up-regulated in nodular tissue but were down-regulated or unchanged in the tumor tissues. Listed in Table 3-5 are genes with such an expression pattern.

Table 3-5: Genes up-regulated in nodular tissues– Microarray analysis. Alterations representing early stages of HCC progression. *HCC<sub>A</sub>* denotes *Cttnb1* wt tumors, while *HCC<sub>B</sub>* denotes *Cttnb1* mutant tumors. The color legend is the same as that used in Table 3-3.

Gene Symbol	Nodules vs Control Log2 Fold Change	HCC A vs Control Log2 Fold Change	HCC B vs Control Log2 Fold Change
<i>Moxd1</i>	3.3	-0.4	-0.4
<i>Slco1a1</i>	1.3	-3.5	-4.2
<i>Saa2</i>	1.7	0.6	-3.6
<i>Clec2h</i>	1.3	-1.8	-1.9
<i>Saa1</i>	1.6	0.8	-2.9
<i>C8a</i>	1.0	0.2	-2.9
<i>Krt23</i>	1.0	0.7	-1.2



These genes are potential candidate genes to have tumor suppressive functions and hence they are down regulated at the tumor stages. The genes *Slco1a1*, *Clec2h* and *Saa2* were found differentially regulated in a HCC mouse model based on primary insulin resistance as well (Hines et al., 2011). This implies strong metabolic alterations as milestone events during the switch from hyper-proliferative nodular stages towards HCCs.

### 3.3.1.4 Genes over-expressed in HCC<sub>B</sub> only

Of further interest, genes could be identified, which were over-expressed specifically in the *Ctnnb1* mutant HCCs.

Table 3-6: Genes up-regulated in *Ctnnb1* mutant (HCC<sub>B</sub>) tumors – Microarray analysis. HCC<sub>A</sub> denotes *Ctnnb1* wt tumors, while HCC<sub>B</sub> denotes *Ctnnb1* mutant tumors. The color legend is the same as that used in Table 3-3.

Gene Symbol	Nodules vs Control Log2 Fold Change	HCC A vs Control Log2 Fold Change	HCC B vs Control Log2 Fold Change
Cyp2c39	-1.3	-1.8	3.1
Pgm5	0.0	-1.1	3.3
Car4	-0.1	0.0	4.9
Sema3c	0.2	-0.3	4.3
Amy2a5	0.0	-0.1	4.1
Aldh1a2	0.4	0.9	4.1
Airn	0.6	0.9	3.7
Gpc3	0.3	-0.1	3.1
Slc1a2	-1.2	-3.6	1.6
Lgr5	-0.7	-1.6	3.0
Avpr1a	0.2	-1.7	2.5
Pdk4	-0.8	-1.2	2.7
Rcan2	-1.7	-0.7	1.4

It is interesting to note that, indeed, some of the known target genes of  $\beta$ -Catenin, such as *Lgr5*, were found up-regulated only in HCC<sub>B</sub> tumors. The gene *Airn* encodes for a lncRNA and is known to be regulated by an imprinting mechanism (Latos

## RESULTS

et al., 2012). The gene *Gpc3* is a marker of HCC progression and is associated with poor prognosis in human HCC patients (Shirakawa et al., 2009). The genes *Sema3c* has also been shown to be overexpressed in various human cancers (Xu et al., 2017). The Cytochrome P450 genes have been shown to have unique gene expression profiles in different tumors (Braeuning, 2008). This can be observed in *SRF-VP16<sup>iHep</sup>* mice as well. *Cyp2c39* is found under-expressed in HCC<sub>A</sub> but is found overexpressed in HCC<sub>B</sub>.

### 3.3.1.5 Tumor heterogeneity

One of the interesting aspects of the tumors of the *SRF-VP16<sup>iHep</sup>* mice was their heterogeneity regarding gene expression profiles, which is a classical feature known for human HCCs. Below two graphs are shown, plotting the expression profiles of specific gene list subsets for all three different sample groups (nodular tissue, HCC<sub>A</sub>, and HCC<sub>B</sub>).

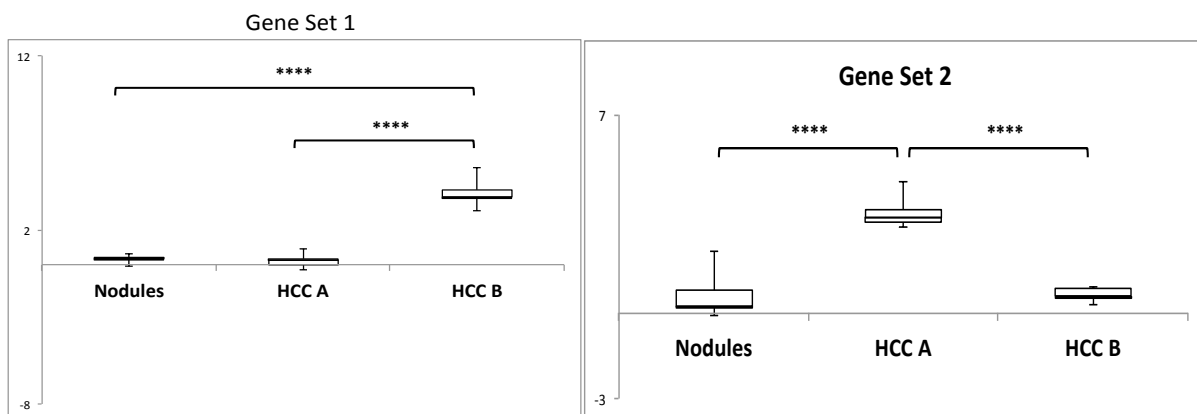


Figure 3-13: Expression profiles of specific gene sets - Microarray analysis. The upper panel shows the unique dysregulation pattern of HCC<sub>B</sub> for gene set 1 and lower panel shows the unique dysregulation pattern of HCC<sub>A</sub> for gene set 2.  $p < 0.0001$

For gene set 1, there is a clear up-regulation in HCC<sub>B</sub> samples, compared to the HCC<sub>A</sub> and nodular samples. Whereas, the gene set 2 represents a different group of genes that's expression profiles show a specific up-regulation in HCC<sub>A</sub> not observed in HCC<sub>B</sub> or the nodular samples. Such gene expression patterns hint toward the heterogeneity that exists among the *SRF-VP16<sup>iHep</sup>* tumors.

---

## 3.4 Comparison to human HCCs

The Microarray data obtained our mouse model of HCC formation was compared to different expression profiling datasets from human HCCs. This was done, in order to determine the relevance of these tumors to human HCCs, given the observation that the *Ctnnb1* gene mutations found in HCC<sub>B</sub> tumors are identical to mutations found in a subset of human HCCs as well.

### 3.4.1 Heidelberg cohort

The Microarray gene expression data of 40 human HCC patients was shared by Dr. Thomas Longerich, then working at the Department of Pathology, University Clinic Heidelberg. The sample demographics are as described in Table 3-7.

#### 3.4.1.1 Boyault classification

Boyault classification is a method used to classify human tumors into different sub-categories, using the expression profiles of a set of classifier genes (Boyault et al., 2007). The Boyault classification was made for the Heidelberg cohort, to determine the molecular sub-category of each of the included human tumors. Seven of the tumors belonged to the G6 sub-category of HCC, which usually harbors mutations in the *CTNNB1* gene locus. Eight of the human tumors were classified to belong to the G1 and G2 sub-categories that are categorized by overexpression of oncofetal genes (among other features).

#### 3.4.1.2 SRF expression

The expression of the *SRF* gene of the human tumors belonging to the HD 40 cohort was analyzed. The *SRF* transcript levels of the tumors were compared to those of the control human individuals. The expression levels of *SRF* in these human HCC samples was correlated with other molecular features of the patient samples, such as Boyault class, *DLC1* status, *IGF2* expression and *CTNNB1* mutation status, among others.

## RESULTS

The results are shown in Figure 3-14. The genomic mutation status of the *DLC1* gene of the human patients had been shared by Dr. Thomas Longerich, Heidelberg. The *DLC1* protein can be considered to be an upstream inhibitor of SRF (via Rho inhibition) and its loss implies an activation of SRF activity through stimulation Rho/actin signaling.

*Table 3-7: Patient demographics of Heidelberg cohort of human HCC patients.*

<b>N = 40</b>	<b>Variables</b>	<b># Patients</b>	<b># SRF Overexpressing Patients</b>
Median Age	56	40	19 (48%)
Gender	Male (30)	30	12 (40%)
	Female (10)	10	7 (70%)
Grading	Grade 1	5	1 (20%)
	Grade 2	30	14 (47%)
	Grade 3 or above	5	4 (80%)
Cirrhosis	Yes	22	8 (36%)
	No	18	11 (61%)
CTNNB1 mutation	Yes	10	6 (60%)
	No	30	13 (43%)
DLC1 status (genomic)	Loss	24	14 (58%)
	Balanced	13	4 (31%)
	Gain	3	1

*The SRF expression levels of the patients were also noted.*

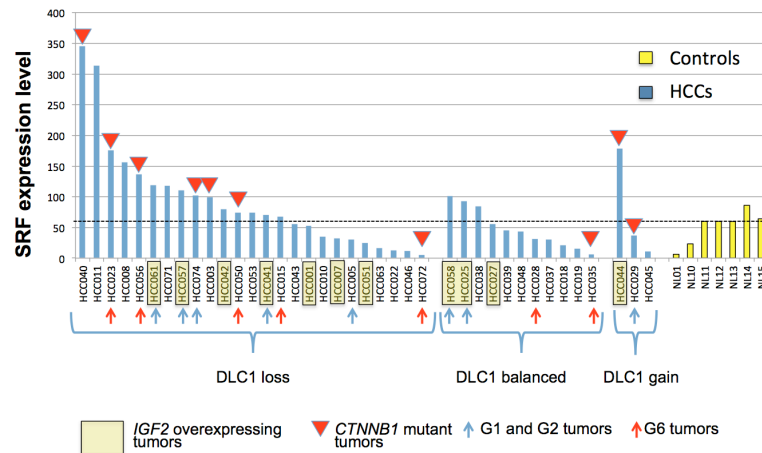


Figure 3-14: *SRF* expression levels in Heidelberg cohort of human HCC patients. Features such as *CTNNB1* mutation status, Boyault sub-category, and the *DLC1* mutation status are plotted. The figure indicates a clear correlation between *SRF* over-expression and loss of *DLC1* in a majority of human patients (Published in Ohrnberger\*, Thavamani\* et al., 2015).

### 3.4.1.3 Unsupervised hierarchical clustering

The microarray-generated expression profiles of the human and the murine *SRF-VP16<sup>iHep</sup>* tumors were subjected to comparison by unsupervised hierarchical clustering, using the Cluster 2.0 software. The unsupervised hierarchical clustering analysis identified a subset of human tumors, designated as the SC<sub>10</sub>, which show the maximum similarity to the mouse tumors. Interestingly they also show an enrichment of samples over-expressing *IGF2* and *SRF* and also belong to the G<sub>1</sub>/G<sub>2</sub> sub-class of Boyault classification, which holds true for the murine HCCs as well (Figure 3-15).

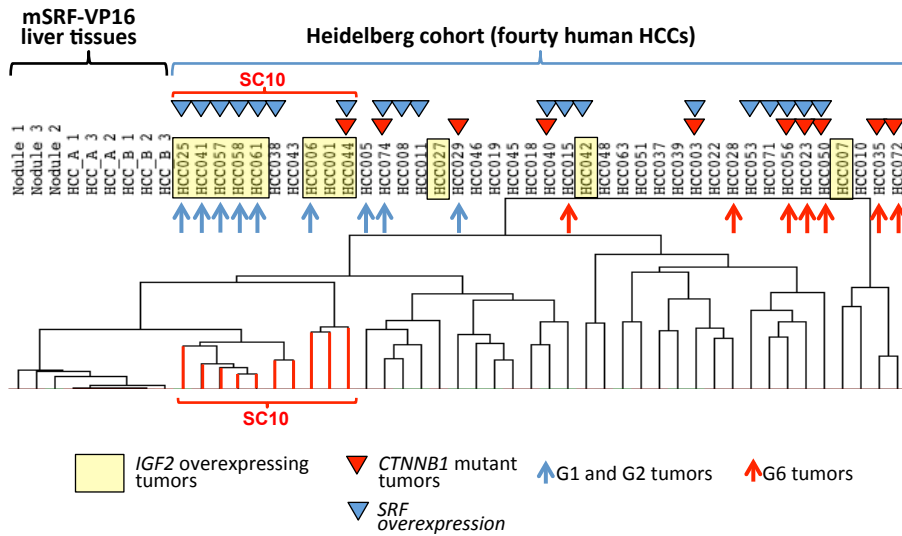


Figure 3-15: Unsupervised hierarchical clustering of mouse and human HCCs -  $SRF-VP16^{iHep}$  tumors versus Heidelberg human HCC cohort. Added to the clustering tree are the various annotations of the human tumors regarding their Boyault sub-class (Blue and red arrows pointing upward), their SRF and *Igf2* expression profiles (blue triangles and yellow boxes, and their *CTNNB1* mutation status (red triangles) (Published in Ohrnberger\*, Thavamani\* et al., 2015).

### 3.4.2 Oncofetal liver genes

The oncofetal liver genes are a set of unique genes, which are active in the embryonic liver, but not in adult liver. In case of liver cancer, however they have been shown to be frequently overexpressed (Yong et al., 2013). Overexpression of oncofetal liver genes has been described as one of the characteristic feature of a distinct sub-type of human cancers, the G<sub>1</sub>/G<sub>2</sub> subtypes of human HCCs, according to the Boyault classification (Boyault et al., 2007). Some of the oncofetal liver genes were found overexpressed in the murine  $SRF-VP16^{iHep}$ -derived tumors as well.

#### 3.4.2.1 Clustering analysis based on oncofetal liver genes

Unsupervised hierarchical clustering of the murine HCCs was performed not only comparison to the Heidelberg cohort (HD 40), but also to two other datasets obtained from the Gene Expression Omnibus data portal. The clustering was done based on the expression profiles of oncofetal liver genes. For this, 11 different oncofetal liver genes were chosen. The clustering analyses are summarized below. This part of the

clustering analysis was performed in collaboration with Dr. Michael Römer from the group of Dr. Zell, University of Tuebingen.

*Table 3-8: Oncofetal liver gene expression in SRF-VP16<sup>iHep</sup>-derived tissues. HCC<sub>A</sub> denotes Ctnnb1 wt tumors, while HCC<sub>B</sub> denotes Ctnnb1 mutant tumors. The color legend is the same as used in Table 3-3.*

Gene Symbol	Nodules vs Control Log2 Fold Change	HCC A vs Control Log2 Fold Change	HCC B vs Control Log2 Fold Change
Igf2	3.7	6.7	6.7
H19	5.1	6.2	6.4
Vil1	3.2	4.6	4.3
Bex1	2.1	6.1	6.4
Psat1	2.8	5.3	5.1
Peg3	1.2	3.9	2.3
Afp	1.0	2.7	2.7
Sox9	1.5	2.2	1.6

*Table 3-9: Summary of clustering analyses comparing oncofetal liver gene expression in three different data sets*

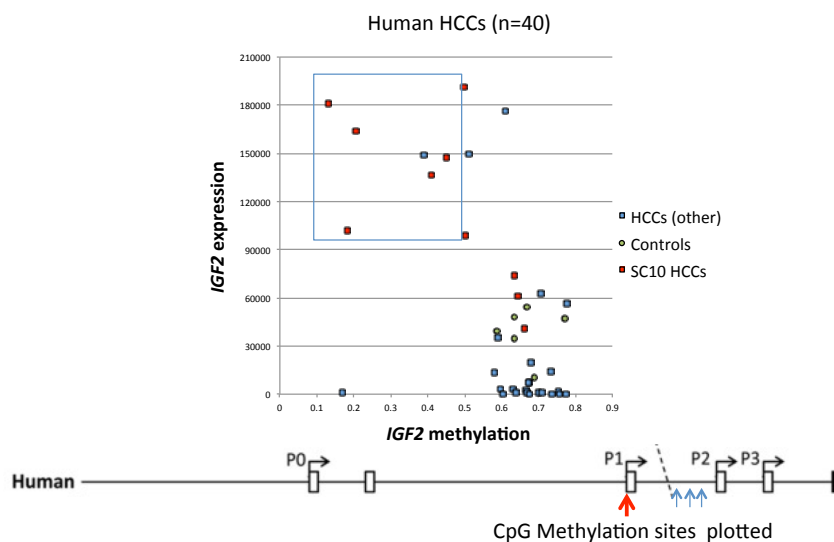
Dataset Id	Source	Total human patients in cohort	Total # human tumor samples clustering close to mouse tumors (and %)
HD 40	Dr. Thomas Longerich	40	7 (17.5 %)
ST 91	Thorgerirsson <i>et.al</i>	91	9 (10%)
NCI 247	National Cancer Institute	247	91 (37%)

Among the oncofetal liver genes, *Igf2* and *H19*, were known to be regulated by means of an imprinting mechanism, which involves the methylation of genomic CpG islands present in the CTGF protein binding site that is in close proximity to both of the genes. Epigenetic alteration in this locus, characterized by loss of imprinting, is a known characteristic feature of the G<sub>1</sub> sub-type of Boyault classified human HCCs (Boyault et al., 2007).

### 3.4.2.2 Genomic CpG methylation analysis

Genomic CpG methylation data was already available for the human HD 40 cohort of HCCs, and was shared by Dr. Longerich. The methylation status of the genomic CpG sites located close to one of the promoters of the *Igf2* gene were investigated (in collaboration with Prof. Dr. C. Plass, DKFZ, Heidelberg). A clear relationship could be observed between hypo-methylation at 3 specific genomic CpG sites close to the P2 promoter of the human *IGF2* gene and the over-expression of its transcript levels.

Interestingly, as shown in the Figure 3-16, a subset of tumors displays a close correlation between genomic hypo-methylation and elevated transcript levels. A majority of those tumors were also the ones that clustered close to the mouse HCCs during hierarchical clustering. This raised the hypothesis of a possible epigenetic CpG methylation mechanism existing in the mouse for regulation of oncofetal liver gene expression, similar to the specific subset of human patients.



*Figure 3-16: Transcript expression levels vs genomic methylation patterns – IGF2. Transcript expression levels of IGF2, plotted against genomic methylation levels of 3 CpG sites close to the P2 promoter of the IGF2 gene. The bottom panel shows the genomic positions of the tested CpG sites within the IGF2 gene locus. (Published in Ohrnberger\*, Thavamani\* et al., 2015).*



Genomic methylation analysis of the orthologous murine locus was done in collaboration with Dr. Daniel Lipka, group of Dr. C. Plass, DKFZ Heidelberg.

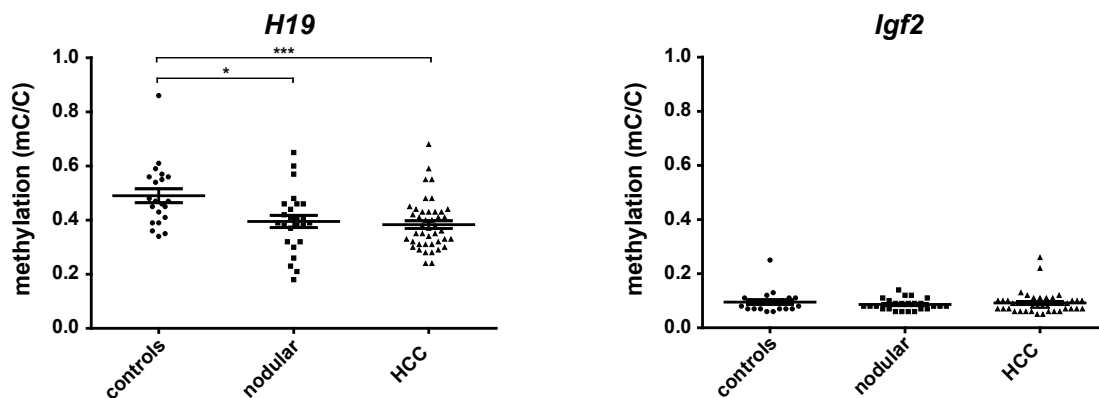


Figure 3-16a. Genomic methylation patterns of *IGF2* and *H19* genes in *SRF-VP16<sup>iHep</sup>* tumors (Published in Ohrnberger\*, Thavamani\* et al., 2015).

There was significant hypo-methylation observed in the *H19* gene locus, but not in the *Igf2* gene locus. This hypo-methylation might assume a significant role in determining the expression level changes of the same gene.

## Combined-omics approach

A combined omics approach for investigation of HCCs derived from *SRF-VP16<sup>iHep</sup>* murine livers was achieved in collaboration with the following institutes/research groups.

- Genome Center, Max Planck Institute for Developmental Biology (MPI), Tuebingen
- Dr. Katarina Matic, group of Prof. Dr. B. Macek, PCT, Tuebingen
- Quantitative Biology Center, Tuebingen
- Dr. Johannes Werner, group of Dr. M. Schlesner, Computational Oncology group, DKFZ, Heidelberg
- High Throughput Sequencing unit, DKFZ, Heidelberg
- MetaSysX GmbH, Potsdam

### 3.5 Transcriptome and proteome analysis

The combined transcriptomics and proteomics approach for RNA and protein analysis were carried out on samples collected from liver tissue samples of *SRF-VP16<sup>iHep</sup>* and littermate control mice. The demographics of the samples are described below in Table 3-10.

*Table 3-10: Demographics of samples used for transcriptomic and proteomic analysis. HCC<sub>A</sub> denotes Ctnnb1 wt tumors, while HCC<sub>B</sub> denotes Ctnnb1 mutant tumors. Nodules stand for pre-malignant hyper-proliferative liver tissue. LWBW stands for Live Weight to Body Weight Ratio.*

Category	Type	Sample #	Mouse #
Gender	Male	8	4
	Female	8	6
Sample type	Controls	4	4
	Nodules	4*	4
	HCC <sub>A</sub>	4*	4
	HCC <sub>B</sub>	4	4
Age	Median Age (weeks)	32 weeks	
LWBW ratio	Average LWBW ratio (Tumors & Nodules)	24.85%	

*\*During proteomic and phosphoproteomic data processing, one sample from Nodules group and one sample from the HCC<sub>A</sub> group were excluded in downstream analyses owing to inadequate quality.*

### 3.5.1 Transcriptomics - RNAseq

The transcriptome profile of the tumor samples was analyzed using the RNAseq method in collaboration with the Genome Center, Max Planck Institute for Developmental Biology, Tuebingen. Measurements were taken on an Illumina Inc. HiSeq 2000 sequencing instrument. The bioinformatics processing of the obtained data was carried out by QBiC, University of Tuebingen.

#### 3.5.1.1 *SRF-VP16* expression

The *SRF-VP16* transgene expression in the nodular and tumor samples was measured by counting the number of reads mapping to the *SRF-VP16* sequence among all different samples. This mapping and counting of reads corresponding to *SRF-VP16* transgene was performed by QBiC. The results are shown in Figure 3-17.

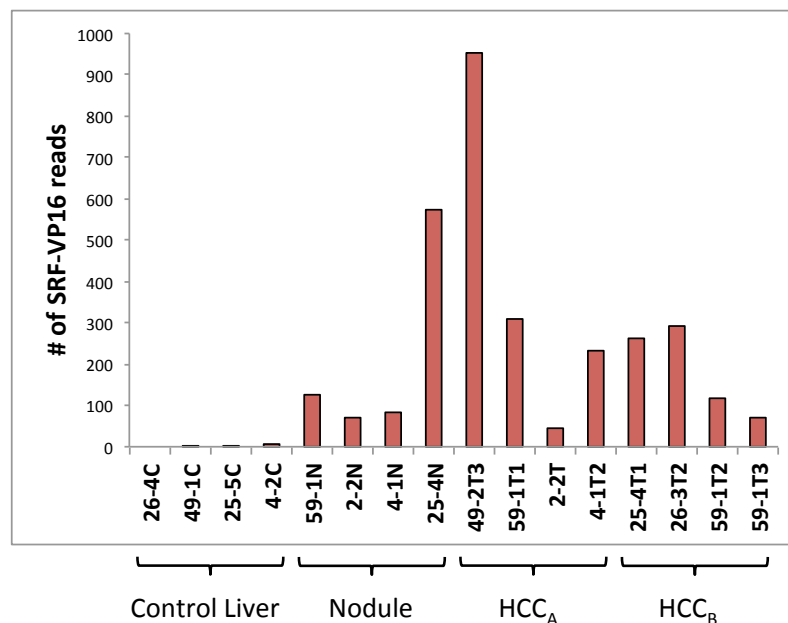


Figure 3-17. *SRF-VP16* transgene expression. The graph showing the expression of *SRF-VP16* sequence in the different samples, as determined by read counts.  $HCC_A$  denotes *Cttnb1* wt tumors, while  $HCC_B$  denotes *Cttnb1* mutant tumors. The *SRF-VP16* reads in the four groups of samples Control Liver, pre-malignant nodular tissue (Nodules),  $HCC_A$  and  $HCC_B$ , were plotted. The *SRF-VP16* reads were found only in the Nodular and HCC samples, but not in the control samples.

The number of reads of SRF-VP16 sequence was found at different levels in the different samples. The presence of SRF-VP16 sequence already in the Nodular samples implies functional expression of SRF-VP16 early in the malignant transformation and in tumor progression.

### 3.5.2 Principal Component Analysis (PCA)

Principle Component Analysis (PCA) of the different samples was performed by QBiC. PCA employs a mathematical algorithm to convert the gene expression values for all genes, of the different sample categories into two-dimensional principal components (PC) namely PC1 and PC2. Each sample group is assigned a pair of values corresponding to each principal component PC1 and PC2. The distance of separation between the sample groups in terms of their principal component values are plotted in a two dimensional graph to visualize the magnitude differences in gene expression patterns of all the sample groups. The PCA plot for transcriptomics data (differentially expressed mRNAs) is shown in Figure 3-18.

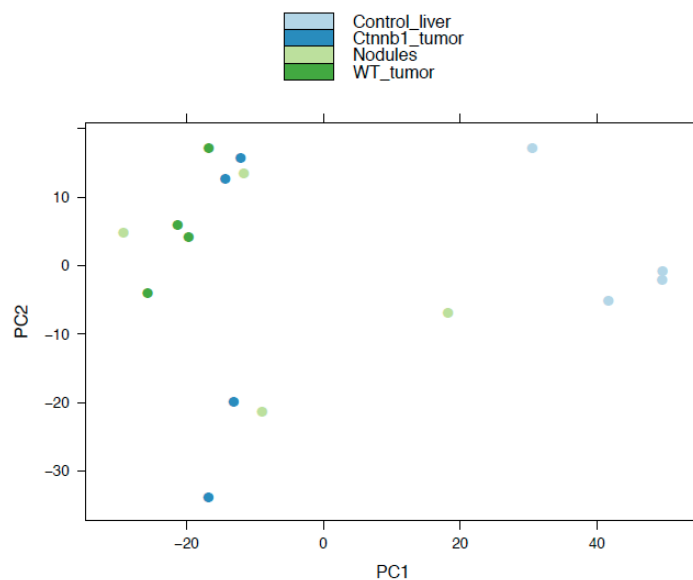


Figure 3-18. PCA plot of transcriptomics data for all samples. The plot shows the relative distance between the samples in terms of their gene expression profiles. Source : QBiC

The control samples cluster separately from all the other samples. The nodular samples and the tumor samples cluster separate from the control samples. This implies a significant difference between the controls and the rest of the samples in terms of their transcriptomic profiles.

### 3.5.3 Transcriptomic profile - RNAseq

#### 3.5.3.1 Number of Deregulated Transcripts

A large number of transcripts were found differentially regulated, compared to control tissue, in the samples that correspond to different stages of malignant transformation in *SRF-VP16<sup>iHep</sup>* murine livers. The total numbers of deregulated transcripts in each of the sample groups are shown in Figure 3-19.

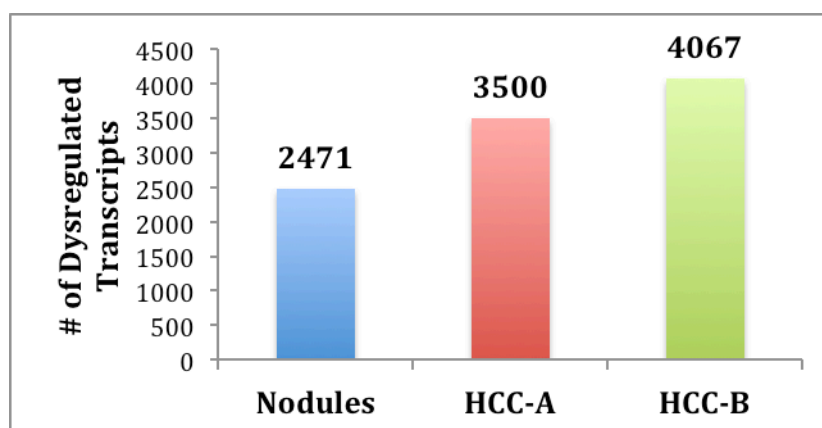


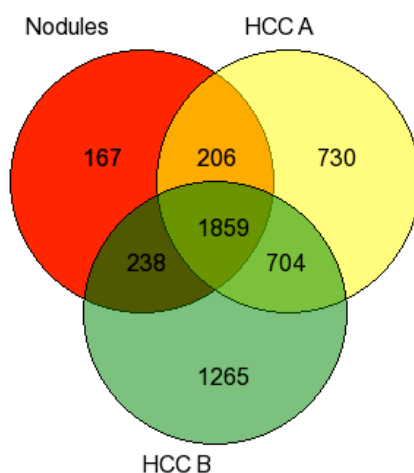
Figure 3-19. # Deregulated transcripts in each sample group. The plot shows the number of differentially regulated transcripts in pre-malignant nodular tissue (Nodules), HCC<sub>A</sub>, and HCC<sub>B</sub>, when compared to the control liver samples. HCC<sub>A</sub> denotes *Cttnb1* wt tumors, while HCC<sub>B</sub> denotes *Cttnb1* mutant tumors. Nodules denote premalignant hyper-proliferative liver tissues.

As shown in Figure 3-19, the HCC<sub>B</sub> sub-group consisting of tumors harboring mutations in the *Cttnb1* gene showed the maximum number of dysregulated transcripts. This could be due to a more advanced stage of malignant progression in this category, due to the modulation of additional signaling pathways such as the Wnt-signaling pathway contributing to tumor growth.

## RESULTS

---

There was also a strong overlap observed between the transcripts deregulated in the different sample groups as shown in Figure 3-20.



*Figure 3-20: Deregulated Transcripts – RNAseq analysis. The overlap between the different sample groups is shown. HCC A denotes Ctnnb1 wt tumors, while HCC B denotes Ctnnb1 mutant tumors, and Nodules denotes pre-malignant nodular tissue.*

In Figure 3-20, it can be noted that, 1859 transcripts were found commonly deregulated in all the different sample groups. More than 200 transcripts were found commonly deregulated in the nodular stage and in either of the HCC sub-types. Interestingly, more than 704 deregulated transcripts were found, which were common to both HCC<sub>A</sub> and HCC<sub>B</sub>. These might be genes involved in critical signaling pathways responsible for the progression from pre-malignant nodules to tumors. A large number of transcripts were exclusively deregulated in the tumors HCC<sub>A</sub> and HCC<sub>B</sub>, respectively. These transcripts are likely those which were altered late during the malignant transformation process as the tumor developed further within the livers. This also indicates a step-wise increase comparing the pre-malignant nodular tissue, versus HCC<sub>A</sub> and HCC<sub>B</sub>. Shown in Figure 3-21 are the numbers of transcripts whose expression are altered exclusively in each category of samples.

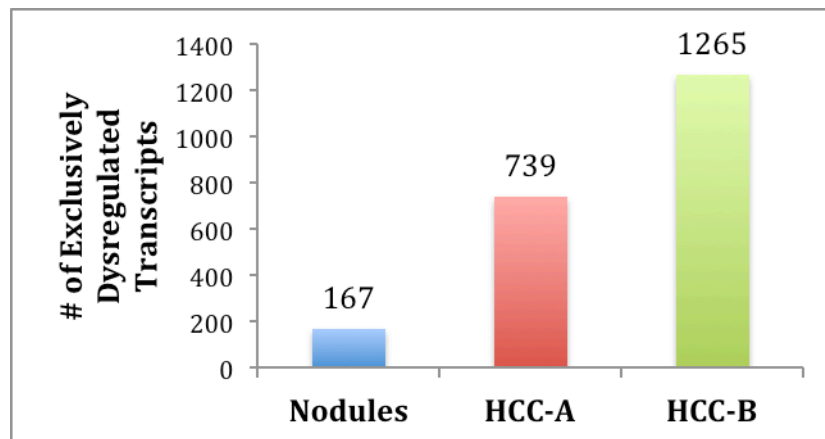


Figure 3-21: Exclusively Deregulated Transcripts – RNAseq analysis. The number of deregulated transcripts unique for each sub group.  $HCC_A$  denotes *Cttnb1* wt tumors, while  $HCC_B$  denotes *Cttnb1* mutant tumors. Nodules stand for pre-malignant hyper-proliferative liver tissue.

The nodules show the least number of uniquely deregulated transcripts. Among the 167 transcripts, the subgroup of upregulated transcripts likely encodes tumor suppressive functions, which are absent at later stages of tumor progression. In Figure 3-21 the step-wise increase of gene-expression changes is evident. The  $HCC_B$  shows the maximum number of exclusively deregulated transcripts. A majority of these could be the genes associated directly or indirectly with the Wnt signaling pathway, owing to the fact that  $HCC_B$  samples are those that harbor gain of function mutations in the *Cttnb1* gene.

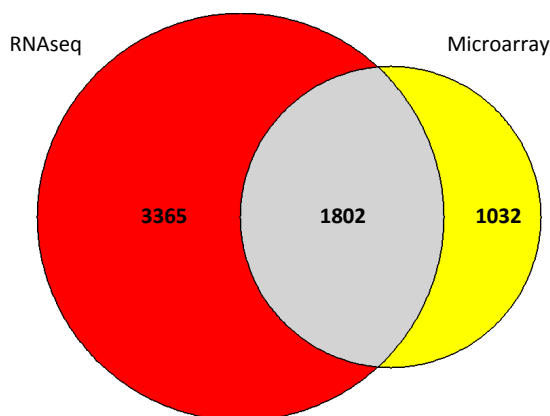
### 3.5.4 Comparison of RNAseq data to Microarray data

The results obtained from this RNAseq expression analysis complement the earlier microarray expression analysis. A comparison of the two datasets showed a significant overlap regarding the expression patterns of a majority of genes. Since different mouse samples were subjected to each of the experimental techniques, a direct comparison of both may not be prudent. However, the overlapping expression trends

## RESULTS

---

obtained from both technologies indicate genes vital for *SRF-VP16*-triggered hepatocarcinogenesis .



*Figure 3-22: Comparison of experimental strategies – RNAseq versus Microarray. The numbers displayed within the circles denote the number of transcripts found differentially regulated in each of the analytical platforms and the number of transcripts commonly detected by both the technologies, as denoted in the intersection of the two circles (grey area).*

It can be deduced from Figure 3-22 that a majority of the deregulated transcripts detected by the microarray experiment were also detected by RNAseq. In addition, the RNAseq experiment was able to detect 50% more of deregulated transcripts. More than thousand transcripts (a total of 1032) detected by the Microarray were not detected in the RNAseq experiment. These numbers could be well accounted by the fact that both these technologies made use of different tissue samples (from different mice). However to compare the expression profiles of the 1802 transcripts found deregulated in both technology platforms, the log<sub>2</sub> fold-change values compared to the control samples were calculated for each gene across the different sample groups from both RNAseq and microarray analysis.



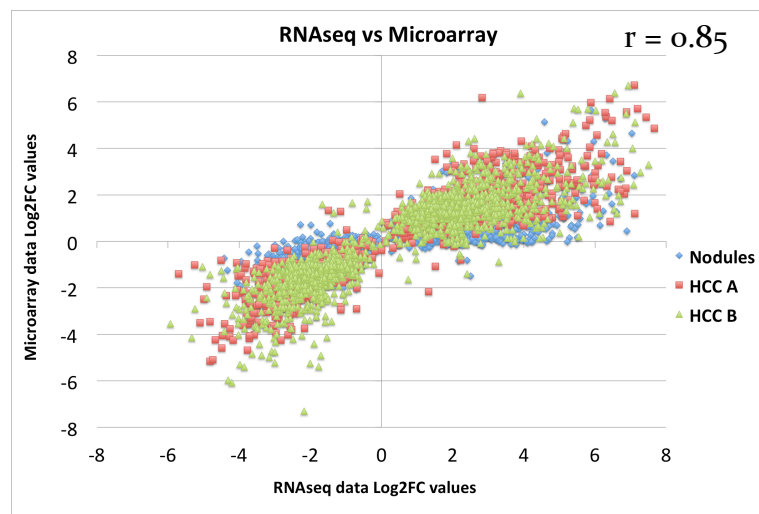


Figure 3-23: Comparison of experiments – RNAseq and Microarray. The  $\log_2$  fold-change values of genes found deregulated in the different experiments were plotted against each other.  $HCC_A$  denotes *Cttnb1* wt tumors, while  $HCC_B$  denotes *Cttnb1* mutant tumors. Nodules stand for pre-malignant hyper-proliferative liver tissue.

Gene expression levels of selected transcripts from among the 1802 deregulated transcripts, which displayed deregulation in both the RNAseq and microarray platforms, are shown in Table 3-11. Among many other examples, the oncofetal liver genes *Igf2*, *Vil1*, *H19*, and *Gldn* that were found overexpressed by microarray analysis were also found overexpressed in the RNAseq platform. Direct target genes of the SRF transcription factor were also among the 1802 transcripts to be commonly dysregulated in both platforms. Interestingly, among the down-regulated transcripts, the nodular tissue samples from the RNAseq experiment showed a greater fold-change difference, when compared to the nodular tissue sample group of the microarray experiments.

## RESULTS

Table 3-11: Comparison of experiments – RNAseq and Microarray. The changes in expression levels expressed as log2 fold-change values compared to the control samples for each gene from the two experiments are shown.  $HCC_A$  denotes *Cttnb1* wt tumors, while  $HCC_B$  denotes *Cttnb1* mutant tumors. Nodules stand for pre-malignant hyper-proliferative liver tissue. The color legend used was the same as that used in Table 3-3.

gene_id	RNAseq	RNAseq	RNAseq	Microarray	Microarray	Microarray
	Nodules vs Control Log2 Fold Change	HCC A vs Control Log2 Fold Change	HCC B vs Control Log2 Fold Change	Nodule vs Control Log2 Fold change	HCC A vs Control Log2 Fold change	HCC B vs Control Log2 Fold change
Cd63	5.2	5.7	5.4	4.3	5.0	5.1
Cdh1	4.4	4.7	4.7	3.1	3.2	3.3
Gldn	6.3	6.3	6.9	5.3	5.4	5.5
Igf2	6.3	7.1	6.9	3.7	6.7	6.7
Lpl	3.8	4.2	4.1	3.2	4.1	3.7
Vil1	4.8	6.0	6.6	3.2	4.6	4.3
H19	4.6	2.8	3.9	5.1	6.2	6.4
Acss3	-3.4	-4.7	-5.3	-2.8	-4.2	-4.1
Lipg	-4.0	-4.8	-5.9	-1.9	-3.5	-3.5
Suox	-3.1	-3.8	-3.6	-1.6	-3.6	-3.6
Inhbe	-3.0	-2.4	-3.2	-2.0	-3.1	-5.0
Raet1d	-4.1	-2.2	-3.1	-1.8	-1.6	-4.7
Slc15a5	-4.3	-4.4	-3.7	-1.2	-2.3	-2.3

### 3.5.5 Comparison to human HCCs – RNAseq data

The RNAseq data of the mouse samples were compared with the RNAseq data of human HCC patients.

#### 3.5.5.1 Mainz cohort

The RNAseq data of 8 HCC patients was obtained with thanks from Dr. Jens Marquardt from the Johannes Gutenberg University in Mainz. The dataset included human liver cancer samples that comprised of low-grade dysplastic nodules, high-grade dysplastic nodules, early HCC and progressed HCC (Table 3-12).

Table 3-12: Sample overview – Mainz cohort of human HCC patient samples.

Liver Cancer stage	# samples
Low-grade dysplastic nodule	3
High grade dysplastic nodule	10
Early HCC	5
Progressed HCC	3

Unsupervised hierarchical clustering of the murine *SRF-VP16<sup>iHep</sup>* samples was performed along with the human HCC samples, in order to compare the expression profiles of the two datasets. Clustering was done using the data of 7521 orthologous genes between 21 human samples and 12 mouse samples.

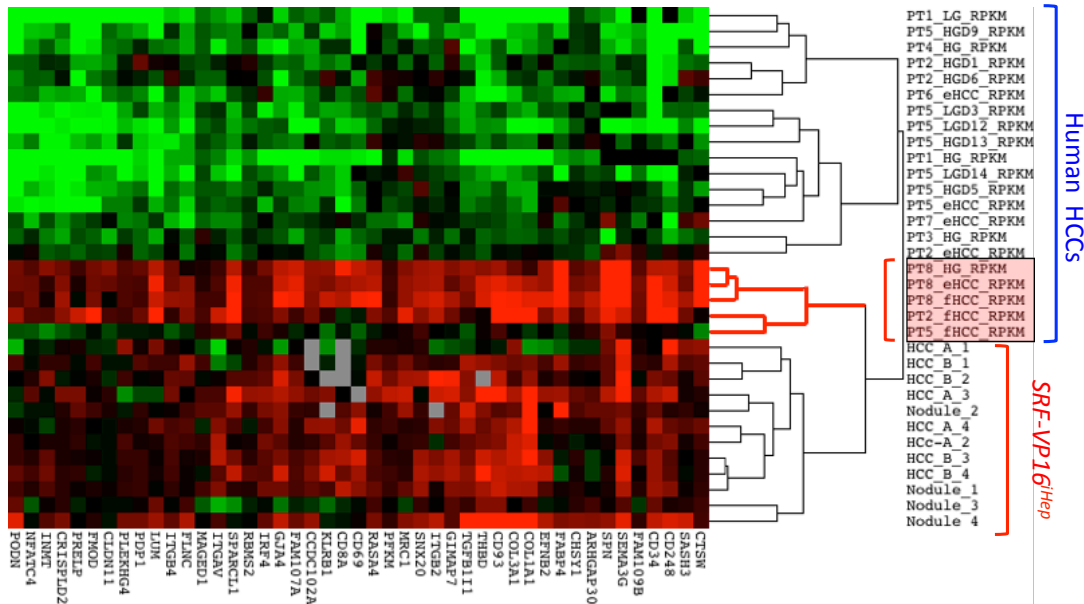


Figure 3-24: Clustering analysis: *SRF-VP16<sup>iHep</sup>* vs Mainz cohort. The human samples are highlighted by the blue line and label. The *SRF-VP16<sup>iHep</sup>* samples are labeled in red. Among the 21 human HCC and dysplastic nodule samples, 5 of those clustered along with the *SRF-VP16<sup>iHep</sup>* samples. The 5 samples are highlighted in a red colored box. 7500 orthologous genes were taken for the clustering analysis.

Upon clustering analysis, it was evident that the 3 progressed HCC samples from 3 different patients clustered closer to the murine samples. Moreover, the early HCC stage sample and the high-grade dysplastic nodular sample of one of the patients also clustered closer to the mice tumors. This suggests that the *SRF-VP16<sup>iHep</sup>* tumors represent a subset of human HCCs that show aggressive and advanced stage phenotype.

### 3.5.5.2 Comparison with The Cancer Genome Atlas (TCGA)

The transcriptome data for HCC patients from the Cancer Genome Atlas, made publicly available by the National Cancer Institute and the National Human Genome Research Institute, was downloaded using the R package TCGA-Assembler. The downloaded data consisted of the read counts of all transcripts for 373 HCC patient samples and 50 corresponding normal human liver samples. The average read count of each gene in the normal liver sample group was used for normalization. The log<sub>2</sub> fold-change values of gene expression changes in the HCC samples were calculated in comparison to the average of the 50 normal liver sample gene expression levels. These values were used for unsupervised hierarchical clustering analysis. For the unsupervised hierarchical clustering analysis, the expression level changes of 25 oncofetal liver genes from the human dataset and the *SRF-VP16<sup>iHep</sup>* dataset were taken (Figure 3-25).

## 3.6 Proteomics analysis

The proteomic analysis of the samples was performed at the Proteome Center Tuebingen. The data acquisition, followed by the data processing and the subsequent statistical analysis was done by Dr. Katarina Matic from Prof. Dr. Boris Macek's group at PCT, Tuebingen.

The samples used for the proteomic analysis were identical to those, which had been used for the RNAseq analyses. The samples have been described in Table 3-10.

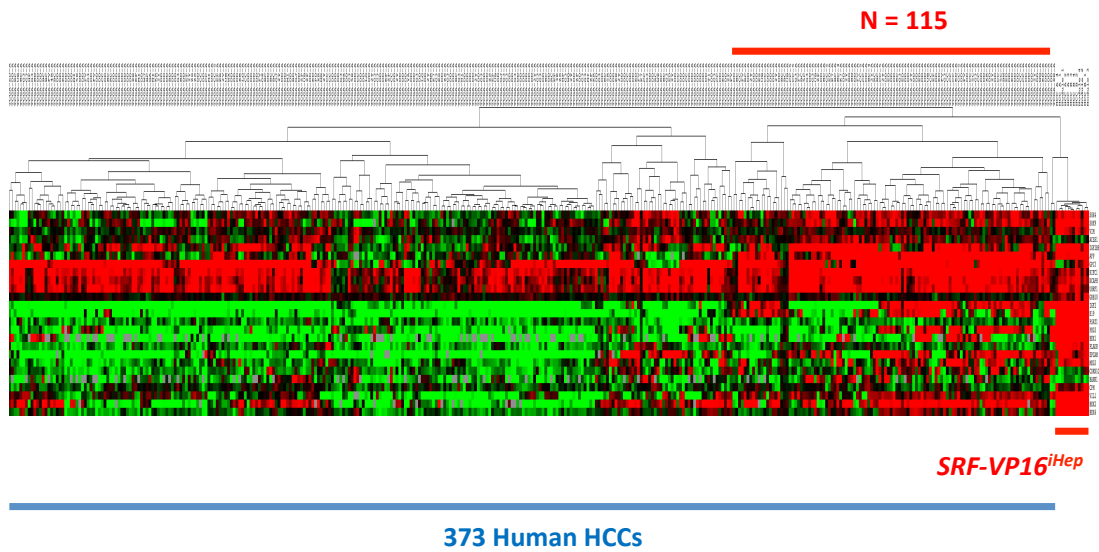


Figure 3-25: Clustering analysis: SRF-VP16<sup>iHep</sup> vs TCGA. The human samples are highlighted by the blue line and label. The SRF-VP16<sup>iHep</sup> samples are labeled in red. Among the 373 human HCC, 115 of those clustered along with the SRF-VP16<sup>iHep</sup> samples. The 115 samples are highlighted in a red line above. 25 orthologous oncofetal liver genes were taken for the clustering analysis.

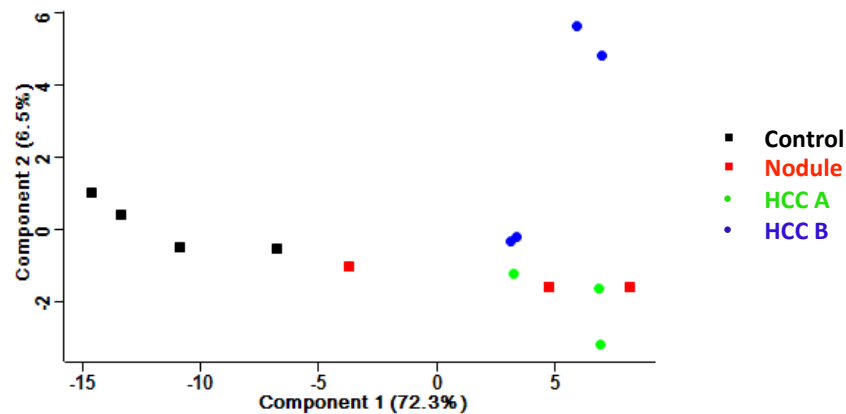


Figure 3-26: PCA plot – Proteomics. The plot shows the relative distance between the samples in terms of their proteomic profiles. HCC<sub>A</sub> denotes *Ctnnb1* wt tumors, while HCC<sub>B</sub> denotes *Ctnnb1* mutant tumors. Nodules stand for pre-malignant hyper-proliferative liver tissue. Source : Dr. Matic (PCT, University Tuebingen)

### 3.6.1 Principal Component Analysis

The Principle Component Analysis (PCA) of the proteome data was done by Dr. Matic.

The PCA plots show a clear segregation between the control samples and the other sample groups (Figure 3-26). Notably, there was found one particular nodular sample, which showed a higher proximity to the control samples when compared to the other samples. The same phenomenon had been observed in the PCA plots of RNAseq data as well. The nodules and tumor samples showed higher overlap, barring a subset of the *Ctnnb1* mutant HCCs

### 3.6.2 Altered protein levels

661 proteins were found in *SRF-VP16<sup>iHep</sup>* tissues that displayed a significant difference in their expression levels, when compared to control liver tissues. The analysis according to the variance statistical test had been applied to test the significance of the protein expression changes. A cut-off for p-value was set at 0.5, above which the protein-level changes were considered not significant.

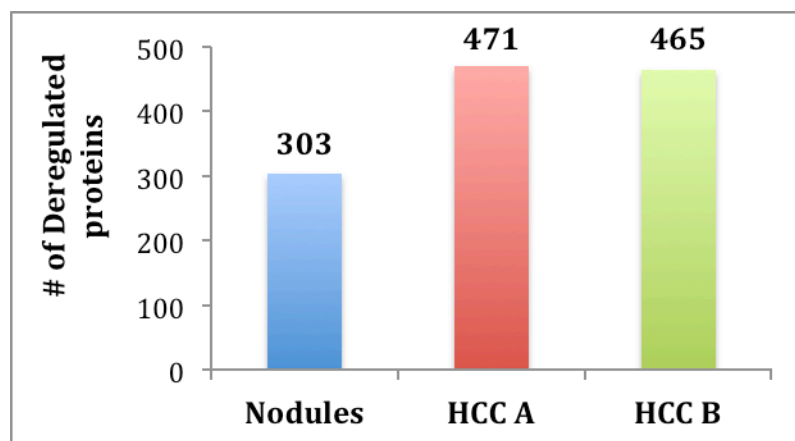
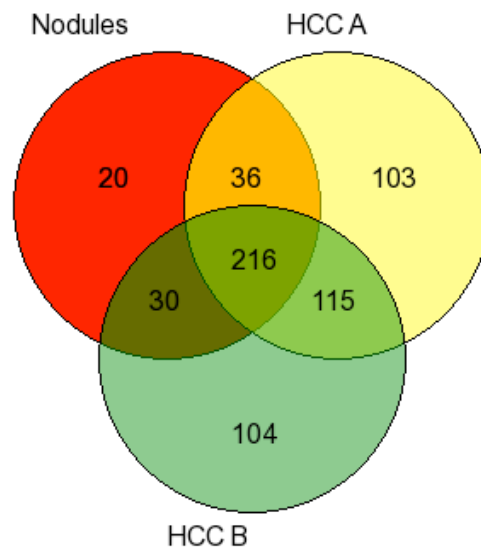


Figure 3-27: Total deregulated Proteins – The plot shows the number of differentially regulated proteins in Nodules, HCC<sub>A</sub>, and HCC<sub>B</sub>, when compared to the control liver samples. HCC<sub>A</sub> denotes *Ctnnb1* wt tumors, while HCC<sub>B</sub> denotes *Ctnnb1* mutant tumors. Nodules denote premalignant hyper-proliferative liver tissues.

At the nodule stage, 303 proteins were found altered in their expression levels. In both tumor stages, more than 450 proteins were found differentially expressed (Figure 3-27). Unlike our deduction from RNAseq data, the *Cttnb1* mutation status doesn't seem to have a huge difference in terms of the number of differentially expressed proteins. The overlap of deregulated proteins, between the different sample groups was analyzed. The results are shown in Figure 3-28.

More than 200 proteins were found, which were already altered in the nodule stage and which continued to be altered also in the tumor stages. These are proteins that might have a significant role in maintaining the high proliferative potential of the tumors. 115 proteins were found altered only in the tumor samples and not in the nodular samples. They might include oncogenes and other proteins involved in the tumor progression.



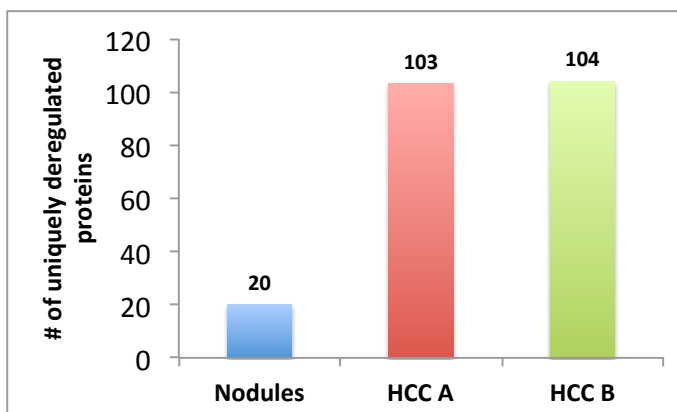
*Figure 3-28: Venn diagram representation of deregulated Proteins – The number of altered proteins across different sample groups are shown in the Venn diagram. HCC<sub>A</sub> denotes *Cttnb1* wt tumors, while HCC<sub>B</sub> denotes *Cttnb1* mutant tumors. Nodules denote premalignant hyper-proliferative liver tissues.*

However, upon plotting the number of proteins, which were found altered exclusively in either of the sample groups, the difference between tumors and nodules was much larger. The graph is shown in Figure 3-29. The number of proteins significantly altered exclusively in the tumor stages, compared to that of the early

## RESULTS

---

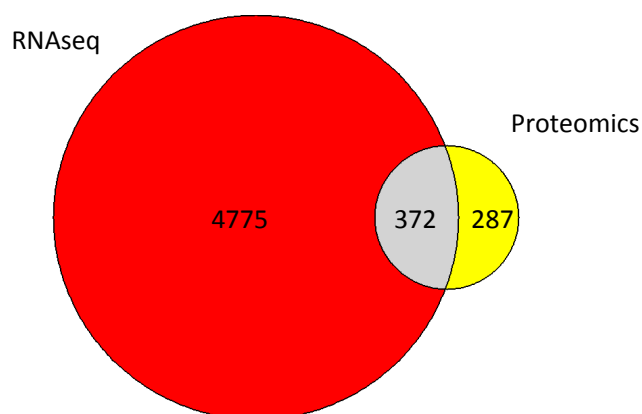
nodular stages is about five-fold higher. This gives a glimpse at the extent of protein synthetic machinery actively functioning in the tumors, compared to the nodules. There were also a large number of proteins, which overlapped between the tumors and the nodules in their expression patterns.



*Figure 3-29: Uniquely deregulated Proteins – The number of exclusively altered proteins across different sample groups are shown in the figure.  $HCC_A$  denotes *Ctnnb1* wt tumors, while  $HCC_B$  denotes *Ctnnb1* mutant tumors. Nodules denote premalignant hyper-proliferative liver tissues.*

### 3.6.3 Comparison to RNAseq

The proteomic data was compared to the RNAseq-derived transcriptomics data. The comparison is shown in Figure 3-30.



*Figure 3-30: Comparison of omics data – RNAseq and proteomics I. The numbers displayed within the circles denote the number of transcripts/proteins found differentially regulated in each of the omics analyses. The number of transcripts/proteins commonly identified by both the omics technologies, is denoted in the intersection of the two circles.*



As observed in Figure 3-30, the coverage of the proteomics experiment in terms of the number of peptides detected was lower than the coverage of the RNAseq experiment. However, a majority of the proteins identified to be differentially regulated were also identified by the RNAseq experiment to be differentially expressed at the transcript level. This implies the presence of a robust regulation mechanism at both the transcript and the protein levels, involved in hepatocarcinogenesis of murine *SRF-VP16<sup>iHep</sup>* livers.

To analyze the 372 commonly identified transcripts and proteins in terms of their regulation patterns, their corresponding log<sub>2</sub> fold changes were compared to the normal liver samples, calculated from both the RNAseq and proteomics experiment, and were plotted in the graph shown in Figure 3-31. The correlation coefficients between the two datasets across the different sample groups were above 0.86. This showed that the changes in RNA expression of the genes were reflected in the change in protein levels of the corresponding proteins. It also shows that the results obtained by either of the experiments were not completely random events. This emphasizes the importance of integrated multi omics approaches for studying changes in the cell.

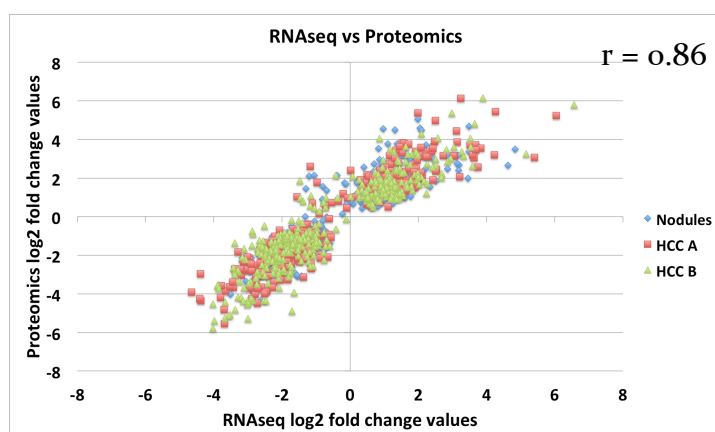


Figure 3-31: Comparison of omics data – RNAseq and proteomics II. The log<sub>2</sub> fold-change values of genes found deregulated in the different experiments were plotted against each other. HCC<sub>A</sub> denotes *Cttnb1* wt tumors, while HCC<sub>B</sub> denotes *Cttnb1* mutant tumors. Nodules stand for pre-malignant hyper-proliferative liver tissue.

## 3.7 Phospho-proteomics analysis

### 3.7.1 Principal Component Analysis

The Principle Component Analysis of the phospho-proteome data was performed by Dr. Matic (PCT, Tübingen).

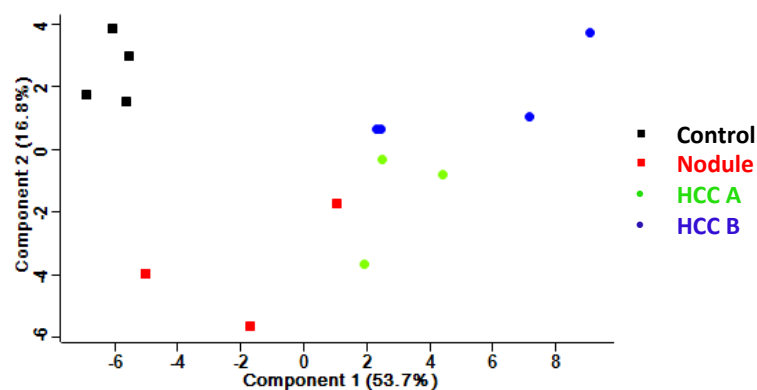
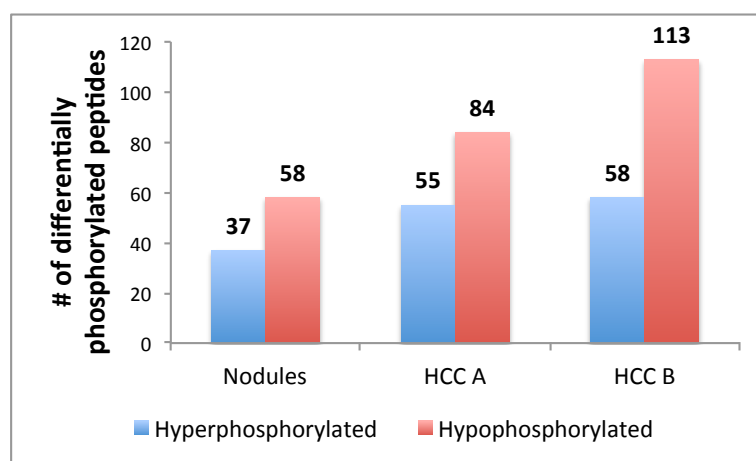


Figure 3-32: PCA plot – Phospho-proteomics. The plot shows the relative distance between the samples in terms of their phospho-proteomic profiles.  $HCC_A$  denotes *Cttnb1* wt tumors, while  $HCC_B$  denotes *Cttnb1* mutant tumors. Nodules stand for pre-malignant hyper-proliferative liver tissue. Source: Dr. Matic (PCT, University of Tuebingen).

Similar to the PCA plot of the proteomics data, the phospho-proteomics data also shows a clear segregation between the control samples and the experimental sample groups. In the PCA plots of RNAseq and proteomics data, one nodular sample was observed that had a very shorter distance from the control samples even in the component 1 segregation. However, in the phospho-proteomic data the separation between the control group and the nodule group is even more pronounced, especially in the separation based on component 2 (Figure 3-32).

### 3.7.2 Differentially phosphorylated peptides

Differential phosphorylation among the peptides was calculated by measuring the number of phospho-enriched peptides found in each sample, compared to the number of phospho-enriched peptides found in the control sample. The difference in number of phospho-enriched peptides in each category was representative of the increased or decreased phosphorylation events in the respective samples compared to the control sample.



*Figure 3-33: Total number of differentially phosphorylated peptides – The plot shows the number of differentially phosphorylated peptides in Nodules, HCC<sub>A</sub>, and HCC<sub>B</sub>, compared to the control liver samples. HCC<sub>A</sub> denotes Ctnnb1 wild type tumors, while HCC<sub>B</sub> denotes Ctnnb1 mutant tumors. Nodules denote premalignant hyper-proliferative liver tissues.*

Figure 3-33 shows the total number of differential phosphorylation events increased at different stages of hepatocarcinogenic progression. Moreover, it was noticed that there were more dephosphorylation events when compared to phosphorylation events. This is evident by the overall higher number of phosphopeptides that were found in lower amounts compared to control samples after phospho-enrichment, thereby implying an increasing number of peptides being hypophosphorylated in the nodules and tumors compared to the controls.

## RESULTS

The types of differentially phosphorylated peptides were also compared. As evident from Figure 3-34 there were more phosphorylation events associated with serine phosphosites compared to events associated with threonine phosphorylation events.

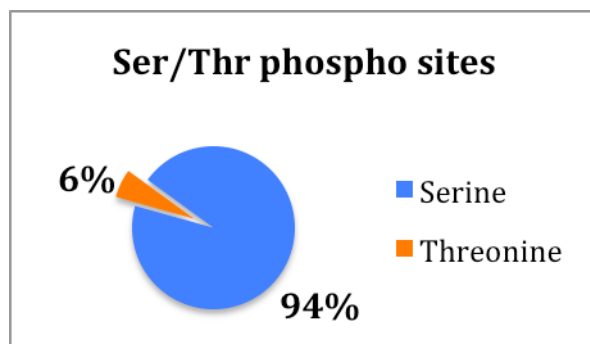


Figure 3-34: Total number of differentially modified serine and threonine phosphosites – The plot shows the number of differentially phosphorylated peptides which had serine or threonine as kinase targets. More than 90% of the observed differential phosphorylation events comprised serine phosphorylation.

The phosphopeptides that were found differentially modified in the nodules, HCC<sub>A</sub> and HCC<sub>B</sub> were compared against each other to check for overlaps between the sample groups. The comparison is shown in Figure 3-35.

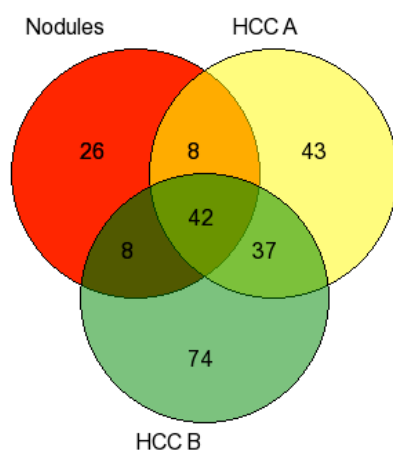


Figure 3-35: Differentially phosphorylated peptides –The plot shows the number of differentially modified phospho peptides in Nodules, HCC<sub>A</sub>, and HCC<sub>B</sub>, compared to the control liver samples. HCC<sub>A</sub> denotes *Ctnnb1* wild type tumors, while HCC<sub>B</sub> denotes *Ctnnb1* mutant tumors. Nodules denote premalignant hyper-proliferative liver tissues.

---

Among the phosphopeptides that were found differentially modified, 42 of them were observed to be commonly modified across all the sample groups. 37 of them were found only in the tumor samples (HCC<sub>A</sub> and HCC<sub>B</sub>) to be differentially phosphorylated. 26 phosphopeptides were differentially modified in the early nodular stage alone and not in the tumors. The tumors HCC<sub>A</sub> and HCC<sub>B</sub> also showed unique alterations in their posttranslational modification profiles regarding phosphorylation, 43 phosphorylation events were scored in HCC<sub>A</sub> and 74 such events in HCC<sub>B</sub>. Some of the interesting protein phosphorylation events were as follows. The HCC<sub>B</sub> tumors showed significant hypo phosphorylation in a Serine at position 675. This position might also be a probable GSK-3 $\beta$  kinase site, whose dephosphorylation might be involved in the stabilization of the protein. The protein encoded by the gene *Acacb* was hypo phosphorylated across all different sample groups including the nodules, thereby implying that it may be involved early on during the malignant transformation process. Interestingly the protein Acetyl-CoA carboxylase 2 is known to be the rate-limiting step in fatty acid biosynthesis in the mitochondria. Decreased phosphorylation of this enzyme early on during carcinogenesis emphasizes on the role of changes in the fatty acid metabolism in causing tumor growth and development. Among the hyper-phosphorylated proteins the Rho GTPase activating protein 32 is a notable candidate due to its involvement in the SRF-signalling pathway.

### 3.8 Whole exome sequencing

Whole exome sequencing analysis (WES) was performed on 30 HCCs and 6 nodules from *SRF-VP16<sup>iHep</sup>* mice, along with 8 control samples. The WES analysis was performed in collaboration with the Genomics Unit of DKFZ, Heidelberg. The bioinformatics analysis of sequenced reads was carried out by Dr. J. Werner at the group of Dr. Mathias Schlesner, DKFZ Heidelberg. The sample demographics are described in Table 3-13.

## RESULTS

---

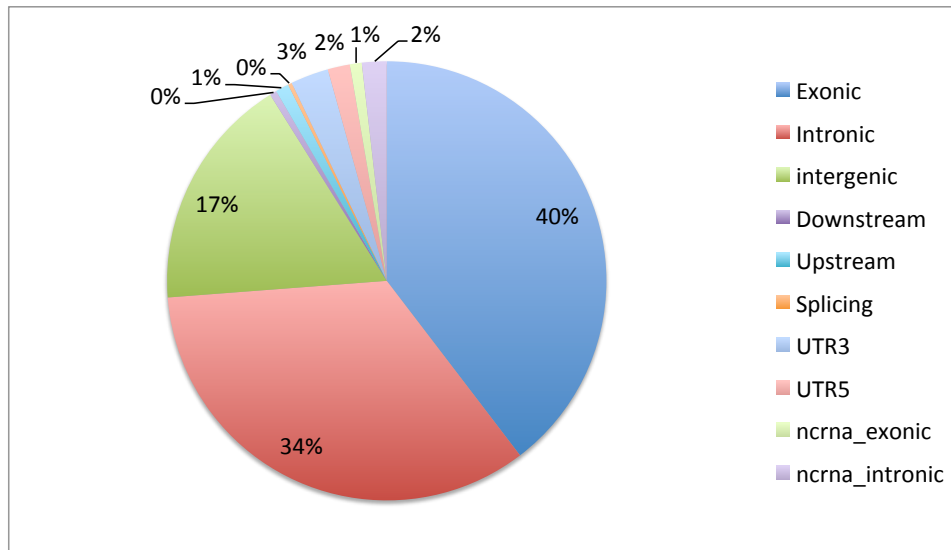
Table 3-13: Demographics of samples taken for WES analysis.  $HCC_A$  denotes *Cttnb1* wt tumors, while  $HCC_B$  denotes *Cttnb1* mutant tumors. Nodules stand for pre-malignant hyper proliferative liver tissue. HCC denote the tumors whose *Cttnb1* mutation status has not been determined. LWBW stands for Liver Weight to Body Weight Ratio.

Category	Type	Sample #	Mouse #
Gender	Male	11	4
	Female	33	10
Sample type	Controls	8	8
	Nodules	6	6
	HCC A	4	4
	HCC B	7	4
	HCC	19	8
Age	Median Age (weeks)	42 weeks	
LWBW ratio	Average LWBW ratio (HCCs & Nodules)	27.3%	

### 3.8.1 Single Nucleotide Variations

The total numbers of single nucleotide variations (SNVs) detected specifically, singly or multiply, in the nodules and tumor samples at different locations of the genome are shown in Figure 3-36.

The exonic variants form the majority of mutations detected in the samples. They comprise about 40 % of all SNVs detected. Interestingly, there was a significantly high number of SNVs present in the intergenic and intronic regions of the genome, considering the fact that the exons were specifically enriched during library preparation, before sequencing. This could be due to incomplete fragmentation of the genomic DNA, resulting in the intergenic and intronic regions to be captured along with the exons. For further downstream analyses, only the exonic mutations were considered.

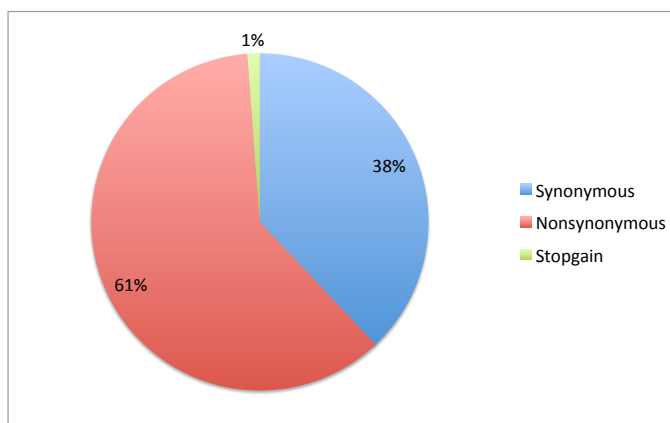


*Figure 3-36: Total single nucleotide variations detected – The location of different point mutations observed in the nodules and tumor samples are shown in the figure.*

Among the mutations (SNVs) detected within the exonic region of the genome, three types of alterations were observed, when considering the effect of a mutation on the transcriptome. Synonymous mutations are SNVs, which do not bring about a change in the amino acid sequence encoded by the transcript. These mutations are usually at the third base of the codon and do not result in altered codon information. The non-synonymous mutations are the ones that bring about a change in the amino acid coding sequence and hence are more likely to have a deleterious effect on the proteins. The stop-gain mutations, as the name suggests, result in a gain of a stop codon thereby causing termination of the translation process and resulting in truncated or stunted proteins, which are highly likely to be dysfunctional. Figure 3-37 shows the relative number of mutations of each sub-type detected in the samples.

## RESULTS

---



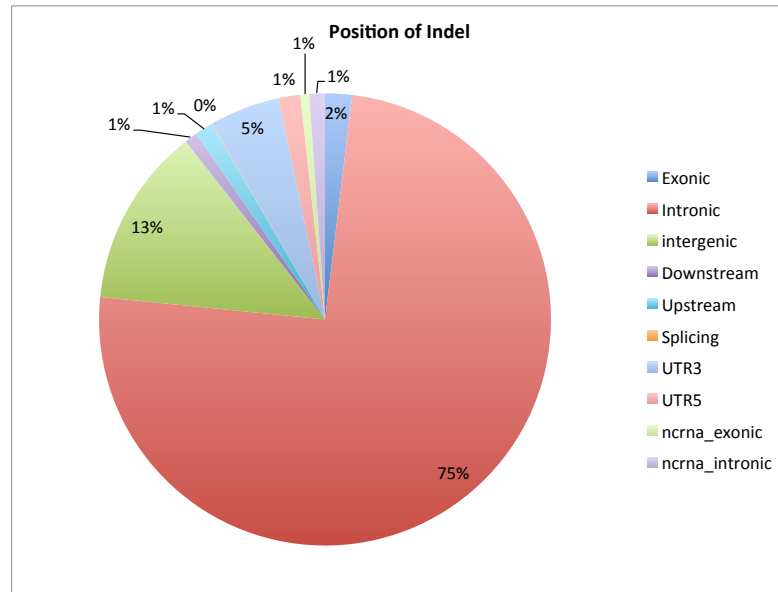
*Figure 3-37: Single nucleotide variations detected in exonic regions – The type of SNV detected in the exonic region of the genomes. Non-synonymous mutations comprise the majority of introduced SNV and synonymous mutations representing the second most frequent type of SNV. Also, a small number of stop-gain mutations were detected.*

### 3.8.1 Insertions/ Deletions

Apart from the SNVs, the exome sequencing analysis also revealed the presence of, insertions/deletions (indels) in the tumor samples.. Indels are nucleotide variations that involve more than one single nucleotide at a given locus. They may involve insertions or deletions of more than a single nucleotide. Occasionally they both occur together in the genome as well. The relative locations of the indels in the genome, regarding functional considerations, are shown in Figure 3-38.

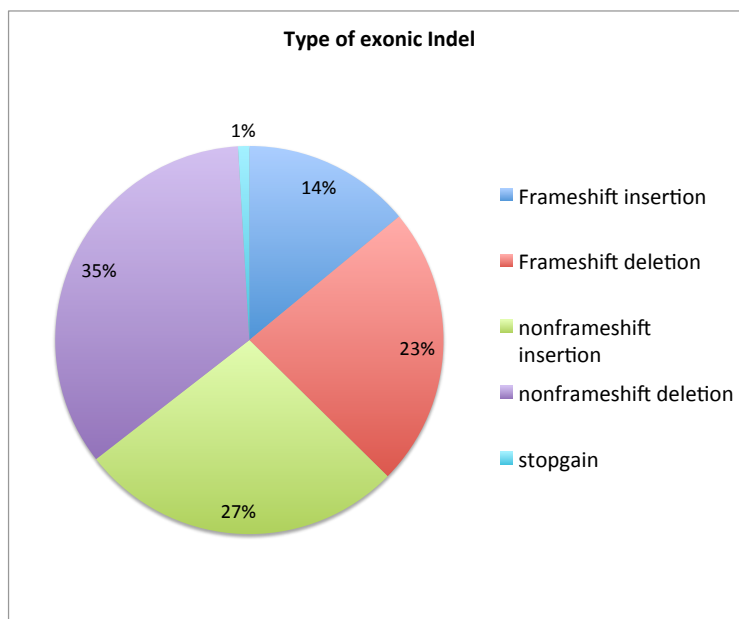
The intronic indels represented the major indels detected by WES analysis. The second most frequent were the intergenic indels. Interestingly, the number of indels mapping within exons was not high as in the case of SNVs. However, only indels mapping to the exonic region were chosen for further analyses.





*Figure 3-38: Total insertions/Deletions – The position of various indels detected in the samples are shown in the figure. Exonic indels are not the majority in the samples.*

There are many potential types of effects caused by indels in the exonic region of the genome. The major effect caused by indels is a frameshift in the coding capacity of a gene, resulting in completely different transcripts and translational products. This can be expected to lead to deleterious effects on the protein. The indels that do not cause frameshifts are known as non-frameshift indels. There might also be stop-gains that occur upon indel formation. The latter represent a small proportion of the different types of indels. The different types of indels detected in our WES analysis are shown in Figure 3-39.



*Figure 3-39: Insertions/deletions detected in Exonic regions – The type of indels detected in the exonic region of the genomes. Frameshift-causing insertions/deletions and non-frameshift causing insertions/deletions are being considered. Also a small number of indels were found that resulted in gain of premature stop codons (stop-gain) in the corresponding genes.*

### 3.8.3 Copy number alterations

Alterations in the copy number of a local DNA segment can lead to differential expression of here-encoded genes. The copy number variations observed across the different samples are showed in Figure 3-40. The copy number gains or losses were considered at the level of entire chromosomes. The copy number variation analysis was done by Dr. J. Werner (group of Dr. Mathias Schlesner, DKFZ, Heidelberg). Briefly, the number of reads mapping to the respective chromosomes are considered and compared to that of the control sample. It can be noted that chromosomes 15 and 18 showed the highest frequency of gains in the mouse genomes, while chromosomes 12 and 13 showed the highest frequency of losses.

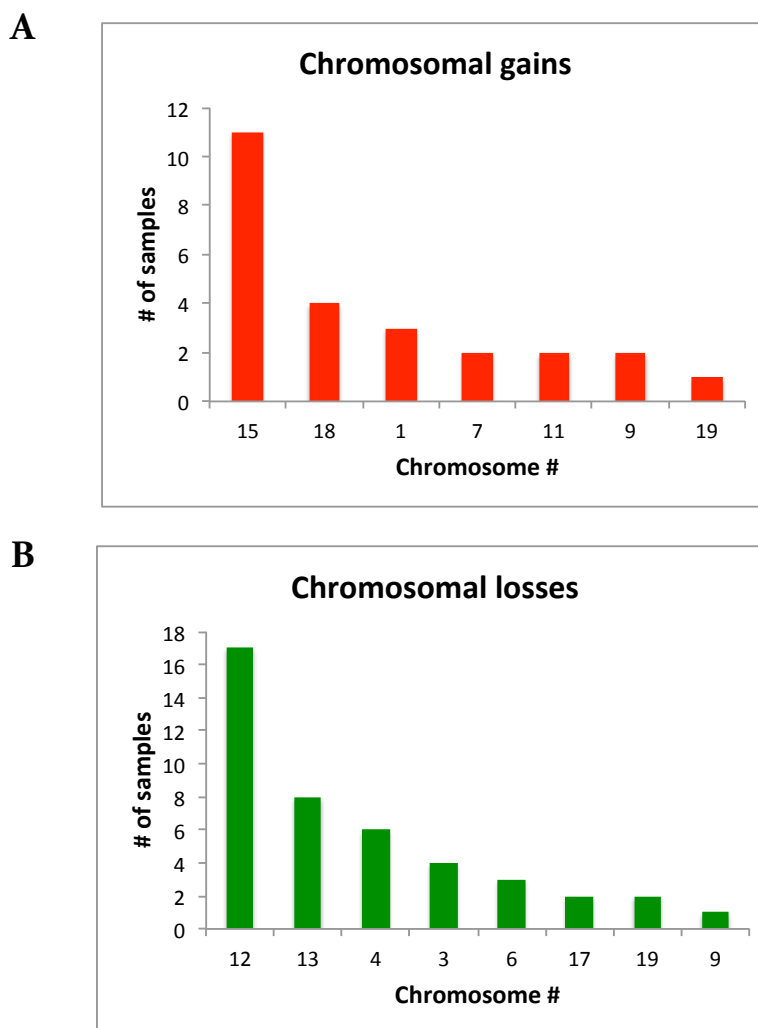


Figure 3-40: Copy number alterations observed in WES analysis – A) The number of samples showing gain in the corresponding chromosomes are shown. B) The number of samples showing loss in the corresponding chromosomes are shown.

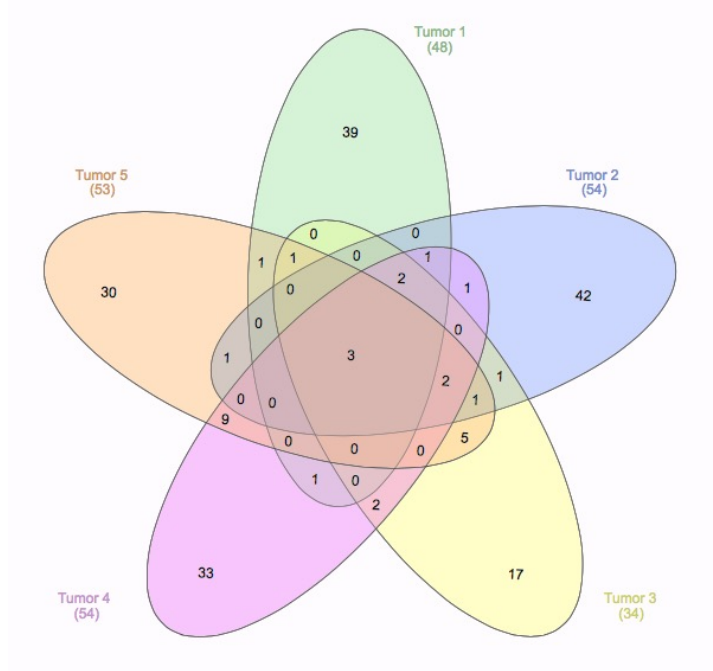
### 3.8.4 Tumor heterogeneity

The heterogeneity of SRF-VP16<sup>iHep</sup> tumors could be addressed using the WES analysis. The mutation profiles of individual tumors were compared against each other. In the cohort of samples taken for WES analysis, there were several tumors, which had originated from the same liver. The mutation profiles of these tumor samples were compared with each other in order to determine the degree of heterogeneity existing among the tumors, despite originating from the same liver.

## RESULTS

---

Figure 3-41 shows the mutations existing in each tumor belonging to a single mouse liver.



*Figure 3-41: Heterogeneity among tumors – The figure shows a Venn diagram representation of the mutations present in each of the 5 tumors that had originated from the same mouse liver. It can be noted that the majority of mutations present in the samples are unique to them, with little or no overlap between the samples.*

The mutation profiles of the samples, as highlighted in Figure 3-37, indicates that the tumors originating within the same mouse liver could still have their own mutation profile, thus implying that the HCCs occurring in SRF-VP16<sup>iHep</sup> mice are heterogeneous in nature, as reflected by their genomic mutation profiles.

## 3.9 Molecular mechanisms involved in HCC formation

### 3.9.1 Wnt signaling pathway

Given that the gain-of-function (i.e. activating) *Ctnnb1* mutations observed in the HCC<sub>B</sub> subset of tumors likely render the  $\beta$ -catenin protein to be stabilized, thereby resulting in the Wnt signaling pathway to be hyper-activated. This mechanism is indeed suggested by our RNAseq expression profiling to be at work in the *SRF-VP16<sup>iHep</sup>* tumors of HCC<sub>B</sub> type. Listed below are the expression profiles of some of the known *Ctnnb1* target genes.

*Table 3-14: A) Expression profiles of Ctnnb1 target genes– RNAseq data. It shows an upregulation of Wnt signaling target genes that are found selectively overexpressed in HCC<sub>B</sub> samples harboring gain-of-function mutations in the Ctnnb1 gene. B) Phosphorylation status of  $\beta$ -catenin protein. The phospho-proteome analysis shows a decrease in phosphorylation status of the beta-Catenin protein that results in its stabilization. HCC<sub>A</sub> denotes Ctnnb1 wild type tumors, while HCC<sub>B</sub> denotes Ctnnb1 mutant tumors. Nodules denote premalignant hyper-proliferative liver tissues. The color legend is the same as that used in Table 3-3.*

*Table 3-14 A) Expression profiles of Ctnnb1 target genes– RNAseq data.*

gene_id	Nodules vs Control Log2 Fold Change	HCC A vs Control Log2 Fold Change	HCC B vs Control Log2 Fold Change
Afp	5.11	4.55	5.61
Cdh1	4.38	4.75	4.72
Igf2	6.32	7.11	6.93
Itga6	3.35	3.98	3.55
Timp1	5.52	6.88	5.21
Tspan8	5.75	5.88	4.70
Axin2	-1.63	-2.41	1.16
Dnmt1	0.36	0.11	1.04
Lgr5	-1.46	-2.42	2.08

*Table 3-14 B) Phosphorylation status of  $\beta$ -catenin protein*

Gene.names	Protein.names	Log2FC Nodule vs Con	Log2FC HCC_A vs Con	Log2FC HCC_B vs Con
Ctnnb1	Catenin beta-1	-0.4	-0.2	-2.9

## RESULTS

In the Table 3-14, some of the target genes are over-expressed only in the *Cttnb1* mutant tumor samples (HCC<sub>B</sub>), thereby strongly inferring the underlying mechanism involved in the cellular proliferation of these tumor cells. The genes *Axin2*, *Dnmt1* and *Lgr5*, were found unchanged or under-expressed in the nodules and in HCC<sub>A</sub>, but they are found overexpressed in HCC<sub>B</sub>. The additional evidence from the phospho-proteomics analysis showing the hypo-phosphorylation of  $\beta$ -catenin protein proves that the Wnt signaling pathway is active in *SRF-VP16*<sup>*iHep*</sup> murine tumors.

### 3.9.2 SRF-target gene regulation

As expected, a large number of known *SRF* target genes were found differentially regulated in the tumor and nodular samples, owing to the expression of SRF-VP16 in the cells. Shown below in Table 3-15 are the expression profiles of selected *SRF* target genes.

*Table 3-15: Expression profiles of selected SRF target genes– RNAseq data. Up-regulation of Srf target genes is evidenced in samples across all three categories. HCC<sub>A</sub> denotes Cttnb1 wild type tumors, while HCC<sub>B</sub> denotes Cttnb1 mutant tumors. Nodules stand for premalignant hyper-proliferative liver tissues. The color legend is the same as that used in Table 3-3.*

Gene Symbol	Nodules vs Control Log2 Fold Change	HCC A vs Control Log2 Fold Change	HCC B vs Control Log2 Fold Change
Ankrd1	4.2	3.2	3.0
Egr1	5.9	6.1	6.2
Egr2	4.3	4.0	3.5
Fhl3	4.3	4.7	4.7
Fos	4.8	3.6	3.4
Ier3	4.2	4.3	4.6
Tes	4.0	3.9	3.5
Thbs1	4.1	3.0	5.8
Atf3	3.9	3.9	3.0
Hk1	2.2	3.1	2.4
Acsl4	2.1	2.1	2.6
Actb	1.3	1.9	1.9
Ahnak	2.1	2.5	3.0
Bcl2	2.9	2.9	2.9

---

The classical target genes of *SRF*, namely *Fos*, *Ier3*, *Thbs1*, *Egr1* and *Egr2*, were found highly up regulated in all the sample groups. The up-regulation of the abovementioned genes also implies cellular hyper proliferation in the *SRF-VP16<sup>iHep</sup>* liver tissues. Among all known *SRF* target genes, more than 100 transcripts were found significantly over-expressed in the *SRF-VP16<sup>iHep</sup>* tissue. This suggests that the role of *SRF-VP16* protein during malignant transformation is not limited to the early stages of triggering hyper proliferation. The up-regulation of *SRF* target genes throughout the different stages of tumor progression emphasizes the role of *SRF* in hepatocarcinogenesis of the *SRF-VP16<sup>iHep</sup>* animal model. *SRF* could play a major role in hepatocarcinogenesis by modulating the metabolic pathways as well. The gene *Hki* encodes for the enzyme *Hexokinase 1* which catalyzes the first step of glycolysis. Considering the general insight that cancer cells tend to feed off from glycolysis more than normal cells, it may be speculated that *SRF* could play a major role in various aspects of cancer development, including cancer cell metabolism.

### 3.9.3 Cell cycle regulation

*SRF-VP16<sup>iHep</sup>* tumors and nodules show high up-regulation of genes associated with the cell cycle. Mitosis-related genes such as *Aurka*, *Cdk1*, *Ccnb1* and *Ccnb2* are all found up regulated consistently across all sample groups. The expression patterns of a selected list of mitosis-related genes observed in the RNAseq analysis are shown in Table 3-16. The up-regulation of kinases, namely *Ttk*, *Limk2*, *Plk1* and *Cdk1*, is mirrored in multiple human tumors as well. The human TCGA data set was curated for expression patterns of mitosis-related genes in various human cancers. The data were downloaded using TCGA Assembler and the log<sub>2</sub> fold-changes were calculated for each dataset, by normalizing against the control samples. Different human cancer subtypes including breast adenocarcinoma, lung adenocarcinoma, cholangiocarcinoma, hepatocellular carcinoma and glioblastoma multiforme, showed over-expression of mitosis-related genes in a majority of the patient samples. The highest was in

## RESULTS

cholangiocarcinoma, whereas the cancer showing lowest frequency of mitosis-related genes overexpression was prostate adenocarcinoma.

*Table 3-16 Expression profiles of mitosis-related genes – RNAseq data. There is found a significant up regulation of mitosis related genes in samples across all groups. HCC<sub>A</sub> denotes Ctnnb1 wild type tumors, while HCC<sub>B</sub> denotes Ctnnb1 mutant tumors. Nodules stand for premalignant hyper-proliferative liver tissues. The color legend is the same as that used in Table 3-3.*

gene_id	Nodules vs Control Log2 Fold Change	HCC A vs Control Log2 Fold Change	HCC B vs Control Log2 Fold Change
Cenpf	3.25	1.94	3.31
Ccnb1	3.46	2.59	3.14
Ttk	2.99	2.98	3.17
Ccna2	2.64	2.64	2.48
Limk2	1.14	1.20	1.58
Bub1	3.13	2.07	2.20
Plk1	3.23	2.77	2.93
Aurka	1.42	1.73	1.37
Cdk1	3.34	2.79	3.20
Cdkn3	3.45	2.92	3.31
Ccnb2	4.28	3.60	3.91

### 3.9.4 Metabolic reprogramming

Alterations in the regulation of numerous metabolic pathways were observed in *SRF-VP16<sup>iHep</sup>* samples.

#### 3.9.4.1 Glycolysis

Glycolysis pathway was one of the highly overexpressed pathways in the *SRF-VP16<sup>iHep</sup>* samples. Most of the proteins involved in the glycolytic pathway were found over expressed in the proteomics data. This observation was complemented well by the RNAseq data where a majority of the transcripts coding for enzymes or proteins involved in the glycolysis pathway were found up regulated. The Table 3-17 shows a list of selected genes belonging to the glycolytic pathway whose protein levels were found up regulated.



Table 3-17: Glycolysis-related genes expression. Selected genes belonging to the glycolytic pathway found over-expressed in the *SRF-VP16<sup>iHep</sup>* tumors. The majority of entries in the table denote the protein expression levels whereas the entries for the genes *Slc2a1* and *Eno2* denote their transcript levels. *HCC<sub>A</sub>* denotes *Cttnb1* wild type tumors, while *HCC<sub>B</sub>* denotes *Cttnb1* mutant tumors. Nodules stand for premalignant hyper-proliferative liver tissues. The color legend is the same as that used in Table 3-3.

Gene.names	Nodules vs Control Log2 Fold Change	HCC A vs Control Log2 Fold Change	HCC B vs Control Log2 Fold Change
Hk1	2.03	2.69	2.21
Pfkp	1.87	3.12	2.58
Aldoa	1.63	1.80	1.24
Slc2a1*	1.73	2.05	0.88
Eno2*	3.96	4.41	3.61
Gapdh	0.74	1.22	1.11
Tpi1	0.73	1.42	0.91
Pgk1	0.55	0.72	0.43
Pgam1	0.43	0.26	0.26
Acs14	3.29	3.38	4.15
Ldha	1.14	1.76	1.52

The higher representation of the glycolysis pathway members in *SRF-VP16<sup>iHep</sup>* samples denotes that the *SRF-VP16<sup>iHep</sup>* HCCs depend a lot on glycolysis for the break down of glucose and energy synthesis. One of the end products of the glycolysis pathway is the metabolite pyruvate, which is either converted to lactate, or is transferred to the citric acid cycle

### 3.9.4.2 Citric acid cycle

The Citric acid cycle (TCA cycle) is one of the major metabolic hubs, functioning in the mitochondria. It is the main acceptor of carbon units in the mitochondria, and in turn generates high-energy electrons (for ATP synthesis) and building blocks for e.g. fatty acid biosynthesis. The glycolysis pathway, glutaminolysis pathway and other amino acids catabolism pathways, feed into the TCA cycle (Zong et al., 2016). The TCA cycle is essential for the mitochondrial electron transport chain, as it provides electrons that facilitate the transfer of protons throughout the electron transport chain. In the

## RESULTS

*SRF-VP16*<sup>Hep</sup> mice samples, there is a decrease in the transcript and proteins involved in the TCA cycle. The genes encoding the succinate dehydrogenase enzymes, fumarate hydratase, and isocitrate dehydrogenase, were under expressed at the transcript level.

It must be noted that, not only the TCA cycle, but also a majority of mitochondria-associated components are also found down regulated in the RNAseq and proteomics datasets. Gene set Enrichment analysis using the RNAseq and proteomics data revealed this phenomenon (Figure 3-42).

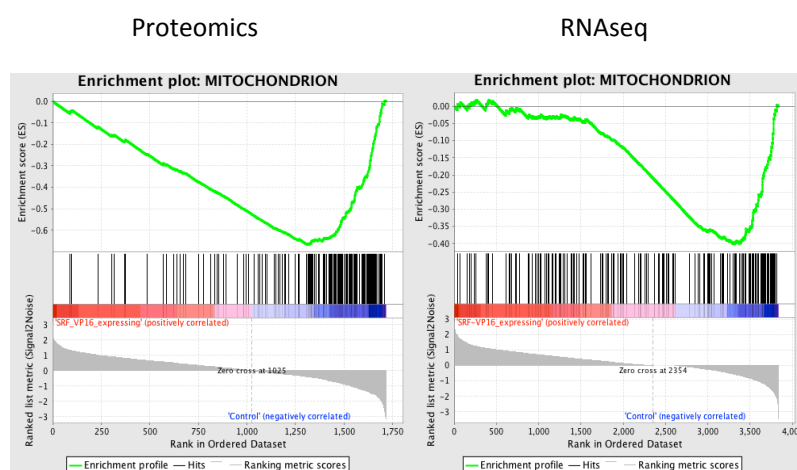


Figure 3-42: Gene set enrichment analysis (GSEA) of proteomics and RNAseq data – The GSEA analysis showed a significant down-regulation of transcripts and proteins associated with the mitochondria.

The majority of proteins and transcripts of the mitochondria are down regulated except the ones belonging to one particular pathway, the one-carbon metabolism pathway where the proteins and transcripts of the mitochondrial compartment are found over-expressed.

### 3.9.4.3 One-carbon metabolism

The one-carbon metabolism (1C-metabolism) is an important metabolic pathway as it is involved in purine and thymidine biosynthesis. The 1C metabolic pathway occurs primarily in the cytoplasm. However, in case of certain cancers it has been noticed that the mitochondrial components of the pathway are highly over-expressed (Nilsson et al,

2014). Serine is the major fuel for the 1C-metabolic pathway, however, the mitochondrial compartment of the pathway can derive 1C-metabolites from glycine as well (Locasale, 2013). The mitochondrial 1C-metabolism pathway consists of 4 main proteins, encoded by the genes *Shmt2*, *Mthfd1l*, *Mthfd2*, and *Aldh1l2*. They were all found up-regulated in *SRF-VP16<sup>iHep</sup>* samples.

*Table 3-18: Expression profiles of 1 C metabolism-related genes – RNAseq data. HCC<sub>A</sub> denotes Ctnnb1 wild type tumors, while HCC<sub>B</sub> denotes Ctnnb1 mutant tumors. Nodules stand for premalignant hyper-proliferative liver tissues. The corresponding percentage of human HCC patients showing a similar pattern of deregulation is also shown. The color legend is the same as that used in Table 3-3.*

Gene_Id	Nodules vs Control Log2 Fold Change	HCC A vs Control Log2 Fold Change	HCC B vs Control Log2 Fold Change	Percentage TCGA Human HCCs showing deregulation %
Mthfd1	-1.87	-2.39	-2.21	67.56
Mthfd1l	2.58	3.10	3.28	67
Mthfd2	2.86	4.44	3.45	17
Shmt1	-1.79	-1.52	-2.08	70.24
Shmt2	-1.45	-1.29	-1.56	24.66
Aldh1l2	2.22	3.78	3.07	29

### 3.10 Metabolomics analysis

The demographics of the samples subjected to metabolomics analysis are given in Table 3-19. Metabolomics analysis was done in collaboration with MetaSysX GmbH (Potsdam). Additional bioinformatic analysis of the metabolomics data was done by QBIC, Tuebingen.

## RESULTS

Table 3-19: Sample Demographics - Metabolomics analysis.

Category	Type	Sample #	Mouse #
Gender	Male	11	6
	Female	11	6
Sample type	Controls	3	4
	Nodules	3	5
	HCC	3	3
Age	Median Age (weeks)	27 weeks	
LWBW ratio	Average LWBW ratio (HCCs & Nodules)	27.84%	

Various metabolites and lipids were found differentially expressed in the tumors and nodules when compared to the control liver samples. A list of up-regulated and down-regulated metabolites in the nodules and/or the tumor samples is displayed in Table 3-20.

The metabolite S-Adenosyl-L-homocysteine (SAH) was found down regulated in the *SRF-VP16<sup>iHep</sup>* samples. In certain liver cancer mouse models, SAH was found highly over-expressed. The ratio between SAH and S- Adenosyl-Methionine (SAM) is a known marker for methylation profile changes associated with various diseases, including cancer and cardiovascular diseases (Caudill et al., 2001; Zhang et al., 2016b). The under-expression of glutamine is also interesting, given the context of glutaminolysis feeding into the TCA cycle, which is found down-regulated in *SRF-VP16<sup>iHep</sup>* samples.

Functional classification and further scrutiny of the metabolites is required to reveal potentially interesting insights into metabolic changes associated with hepatocarcinogenesis in *SRF-VP16<sup>iHep</sup>* mice.

*Table 3-20: Up-regulated and down-regulated metabolites. Metabolomics analysis. Indicated are metabolites that were found differentially displayed in the nodules and/or tumors.*

<b>Up regulated Metabolites</b>	<b>Down regulated Metabolites</b>
Fructose	Adenosine 5prime-phosphate
Xanthine	N-Acetylglutamic acid
Pantothenic acid (Vitamine B5)	Xanthosine
L-Tyrosine	Uridine 5prime-monophosphate
Adenosine-5-monophosphate	Uric acid
O-Acetyl-L-carnitine	L-Glutamine
Citric acid	Glycerophosphate
Cytidine	Xanthine
Idose	Cysteine
Indole-3-lactic acid	Cystine
Cytidine 3prime-phosphate	Pyridoxal-5prime-monophosphate
4-Guanidinobutyric acid	S-Adenosyl-L-homocysteine
O-Acetyl-L-homoserine	Deoxyguanosine diphosphate

## 4 DISCUSSION

---

The *SRF-VP16<sup>iHep</sup>* mouse strain proved a highly useful model for the study of hepatocellular carcinoma. In the present study, the HCCs occurring in *SRF-VP16<sup>iHep</sup>* mice have been characterized, along with premalignant hyper-proliferative nodular tissues. To this end, a combined multi-omics approach has been undertaken, in order to study the various molecular events, as detected by different omics technology platforms, collectively involved in hepatocarcinogenesis in *SRF-VP16<sup>iHep</sup>* murine livers.

### 4.1 Insights obtained from the present study

#### 4.1.1 Mosaic hepatocyte expression of *SRF-VP16* establishes a cellular platform for initiation and progression of HCC formation in mice

*SRF-VP16<sup>iHep</sup>* mice carry the *SRF-VP16* and *Alfp-CreER<sup>T2</sup>* transgenes in their genome, along with the *Srf-flex1* transgene. Activation of the Cre recombinase induces *SRF-VP16* transgene expression upon recombination-mediated deletion of a STOP cassette flanked by LoxP sites. The presence of the STOP cassette prevents spontaneous expression of *SRF-VP16* transgene in the absence of active Cre. However, as reported earlier (Ohrnberger\*, Thavamani\*, et al., 2015), leakiness of Cre recombinase activity results in sporadic *SRF-VP16* activation in a small number of hepatocytes in the absence of the inducer tamoxifen. This triggers a cascade of events, including hyper-proliferation of *SRF-VP16*-expressing hepatocytes. This ultimately leads to HCC formation in the livers of *SRF-VP16<sup>iHep</sup>* mice. It must be noted that SRF was first identified as a transcription factor that regulates the expression of various proliferation-associated genes, including the family of immediate early response genes (IEGs) (Chai and Tarnawski, 2002). Thus, it can be concluded that – at least in hepatocytes – SRF has a vital proliferative function. The knockout of *Srf* in early embryonic stages was shown to hinder developmental processes (Schratt et al., 2001). In contrast, upon *SRF-VP16*

---

expression, numerous genes associated with cell proliferation were found activated. These include the *Egr1*, *Egr2* and *c-fos* genes. Activation of these genes likely causes a continuous proliferative stimulus to be exerted, resulting in hepatocyte hyper-proliferation. Proliferative stress can result in the accumulation of replication errors, DNA damage, and genomic instability. Furthermore, hyper-proliferation of *SRF-VP16<sup>iHep</sup>* hepatocytes resulted in changes in the expression levels of genes regulating the cell cycle. This situation can additionally favor the occurrence of mutations in the cellular genome. The observed alterations in gene expression profiles, in association with the multitude of mutations identified in our WES analysis, are best explained by a cancer-promoting scenario that is based on hepatocyte hyper-proliferation. We also found a significant overexpression of SRF target genes, even in the early nodular samples. The expression levels of the SRF target genes continued to remain high even at the terminal tumor stages (HCC<sub>A</sub> and HCC<sub>B</sub>). This indicated the importance SRF-VP16 transcriptional activity throughout the hepatocarcinogenesis process, in order to provide a platform for cancer development via accumulating additional pro-carcinogenic mechanisms.

It must be noted that similar scenarios of transcription factor-driven carcinogenesis have been reported previously. A transgenic mouse was reported, in which expression of the *Myc* gene gave rise to HCCs (Murakami et al., 1993). The heterozygous *Myc* transgenic mice were crossed with homozygous *TGF $\alpha$*  transgenic mice, the *TGF $\alpha$*  transgene being driven by the metallothionein promoter. These double transgenic mice were found to develop HCCs at a penetrance of 20% at 5 months of age and a penetrance of 100 % at 8 to 10 months of age (Fausto and Campbell, 2010).

Transcription factor-induced liver carcinogenesis has been reported before (Heindryckx et al., 2009). However, *SRF-VP16<sup>iHep</sup>* mice are unique in their ability to cause tumors with a 100% penetrance. This strongly suggests a more general role of SRF and SRF-directed target genes in cancer formation.

#### 4.1.2 The step-wise progression of HCCs starting from early stages can be studied in detail using *SRF-VP16<sup>iHep</sup>* mice.

The chosen multi-omic approach to study the *SRF-VP16<sup>iHep</sup>* mice has enabled the investigation of early, pre-malignant stages of HCC development. The nodular samples present a unique sample set that enables the study of the early events occurring during liver carcinogenesis. In the current study, numerous events were identified at the genomic, transcriptomic and proteomic levels that were exclusive to the early nodular stages. The number of deregulation events exclusively observed in the nodular stage was far lower than those observed in the tumors. This indicated a principle of accumulation of transformation-inducing events at successive stages of tumor progression. Dysregulation of certain genes, e.g. *Moxd1*, *Slco1a1*, *Saa1*, *Saa2*, *Clec2h*, *C8a* and *Krt23*, was observed at the nodular stage only. Subsequently, these genes were not differentially expressed in the tumors; in some cases they were even down-regulated in the tumors, as was the case for *Slco1a1*. The *Slco1a1* gene belongs to the Solute carrier organic anion transporter family. It has been shown to be over-expressed in certain disease conditions in the liver. It has been associated with acute liver congestion (Shimizu et al., 2015).

Other genes showed similar deregulated expression patterns at early stages of hyper-proliferation, which might have tumor suppressive functions in the *SRF-VP16<sup>iHep</sup>* liver. The *SRF-VP16<sup>iHep</sup>* mice provide a tool for studying these early stage events of HCC progression. The value of such premalignant nodular samples for studying the pathogenesis of the liver cancer has been realized by others. In one such example, 8 human HCC patients from Mainz (Germany) were characterized by the group of Dr. Jens Marquardt (Fausto and Campbell, 2010; Marquardt et al., 2014). The data generated in this study was compared with the *SRF-VP16<sup>iHep</sup>* data. The advanced stage tumors from 3 patients clustered closer to the mHCCs. Interestingly the nodule samples of the mouse clustered close to the advanced HCC samples of human patients. This implies that the



---

nodule stages of *SRF-VP16<sup>iHep</sup>* mice already show features similar to those of advanced stages in humans. The presence of samples from early malignancy stages could result in the identification of biomarkers that might be used for early detection of liver cancer in human patients. Moreover, the premalignant samples provide a tool for the development of therapeutic modalities at early stages of the disease when it is more tractable compared to the later stages. Thus, the pre-malignant nodular samples can serve as an experimental tool for studies regarding detection of early biomarkers and also regarding development of therapeutic modalities for early-detected cancer patients.

### **4.1.3 Multiple mechanisms are involved in the progression of HCCs in *SRF-VP16<sup>iHep</sup>* murine livers**

The step-wise progression of HCCs in the *SRF-VP16<sup>iHep</sup>* livers was paralleled by cumulative activation of molecular mechanisms supporting tumor growth. A synergism of various mechanisms appears to be involved in the progression of *SRF-VP16<sup>iHep</sup>* tumors. Some of the different mechanisms suggested by the present study are discussed in the following.

#### **4.1.3.1 Mitosis**

Genes and proteins associated with cell cycle control and regulation of mitosis were found to be up-regulated in the samples obtained from *SRF-VP16<sup>iHep</sup>* livers. The mitosis-associated genes were also found up-regulated in various human cancers (TCGA data), such as lung adenocarcinoma, breast adenocarcinoma, brain cancer and liver cancers. The mitosis-associated genes that are found up-regulated in the *SRF-VP16<sup>iHep</sup>* HCCs include *Cenpf*, *Ccnb1*, *Ccnb2*, *Ttk*, *Limk2*, *Bub1*, *Plk1*, *Aurka* and *Cdk1* among others. Interestingly, a few of the genes are considered direct target genes of SRF.

#### **4.1.3.2 Wnt signaling pathway**

The Wnt signaling pathway is also found activated in some *SRF-VP16<sup>iHep</sup>* tumors.  $\beta$ -Catenin is one of the downstream effector proteins of Wnt signaling pathway. The gene coding for the  $\beta$ -Catenin protein, the *Ctnnb1* gene, is found mutated in a subset of

*SRF-VP16<sup>iHep</sup>* tumors; such tumors were labeled HCC<sub>B</sub>. The observed mutations can be considered activating (gain-of-function) mutations, resulting in the stabilization of the  $\beta$ -Catenin protein. This subsequently results in the Wnt signaling pathway being consistently active. RNAseq analysis on the *SRF-VP16<sup>iHep</sup>* tumors revealed that, in HCC<sub>B</sub> tumors, indeed numerous target genes of the Wnt signaling pathway were over-expressed. Thus, the Wnt signaling pathway is found to be actively contributing to the tumor growth in a subset of *SRF-VP16<sup>iHep</sup>* tumors.

### 4.1.3.3 Hippo signaling pathway

The Hippo signaling pathway is a major regulator of vital cellular functions such as proliferation, differentiation, growth and cell death. The pathway is basically a cascade of kinase activities resulting ultimately in the transcription regulation of target genes, via activation of the transcription factors TEAD<sub>1</sub> to TEAD<sub>4</sub>, as well as their partner proteins YAP/TAZ. In the *SRF-VP16<sup>iHep</sup>* tumors, a large proportion of Hippo signaling pathway target genes were found over-expressed, including such as *Loxl4*, *Ccnb2*, *Foxm1*, *Psat1*, *Dusp9*, and *Tmem171*, among others. The over-expression was also observed at the proteomics level, with a significant number of proteins, encoded by Hippo target genes, being overexpressed such as the proteins encoded by *Myh10*, *Myh9*, *Aldoa*, *Bub3*, *Fus*, *Ranbp1*, *Rangap1*, *Ywhaz* and *Crip1* among others. Interestingly, it has been speculated that the proteins YAP/TAZ, which are the major modulators of the Hippo signaling pathway, can also interact with SRF to regulate the transcription of various target genes (Liu et al., 2016). The results from our current study also suggest that some of the target genes of Hippo signaling may indeed be regulated by SRF.

### 4.1.3.4 MAPK signaling pathway

The MAPK signaling pathway is an activator of SRF-mediated transcriptional activation. The MAPK signaling pathway is activated in a variety of tumors. The constitutively active SRF-VP16 protein can regulate target genes of MAPK signaling in the absence of partner protein activation. Thus, the expression profile of *SRF-VP16<sup>iHep</sup>*

---

tumors has shown considerable overlap with the gene expression profiles of ras-mediated tumors.

#### 4.1.3.5 Nanog signaling pathway

Nanog is a transcription factor, which is known to be associated with stem cell features in cells. Nanog itself is known to be a stem cell marker protein (Jeter et al., 2015). Our RNAseq analysis revealed that a large number of Nanog target genes were found over-expressed in the *SRF-VP16<sup>iHep</sup>* tumors. Examples include *Acsl4*, *Foxj1*, *Cdh1*, *Rab15*, *Anxa1*, *Alcam*, *Nfe2l3* and *Bcas1* among others. However, Nanog itself was not differentially regulated. This suggests the presence of alternative regulatory mechanisms for Nanog target genes in *SRF-VP16<sup>iHep</sup>* tumors. It is known that Nanog expression generally is inversely proportional to the  $\beta$ -Catenin expression, wherein Nanog is known to stimulate  $\beta$ -Catenin phosphorylation and subsequent ubiquitin-mediated proteasomal degradation (Cheng et al., 2015). In *SRF-VP16<sup>iHep</sup>* tumors, both these signaling pathways seem to be active. This highlights the inherent heterogeneity of *SRF-VP16<sup>iHep</sup>*-derived tumors.

#### 4.1.4 *SRF-VP16<sup>iHep</sup>* HCCs show heterogeneous molecular features

The heterogeneity of liver tumors, especially HCCs, continues to be a major challenge for determining treatment modalities specifically suited for an individual patient. *SRF-VP16<sup>iHep</sup>* tumors also show a high degree of heterogeneity. WES analysis of different *SRF-VP16<sup>iHep</sup>* tumors, occurring in the same liver tissue, showed highly varying mutation profiles. Significant molecular differences were displayed by the different tumors, although they arose within the same liver. This indicates that each tumor may have had its unique path towards malignant transformation during tumor progression. This observation also indicates that, at the genomic level, the various alterations observed were not directly caused by *SRF-VP16* expression. The heterogeneity of the HCCs characterized reflects the characteristic feature of human HCCs also displaying high molecular heterogeneity. Due to their heterogeneity, human HCCs were classified

into sub-categories (Boyault et al., 2007) in order to achieve a better understanding of the etiologies and therapeutic responses of individual HCCs.

### 4.1.4.1 Classification of human HCCs

There have been significant efforts to classify human HCCs. One such attempt to classify human HCCs was put forth by Boyault et al. in 2007. This classification was based on gene expression profiles of the tumor samples. The classification consisted of 6 different sub-categories namely G<sub>1</sub> – G<sub>6</sub>. In this study, the *SRF-VP16<sup>iHep</sup>* tumors were subjected to this classification and were found to overlap most closely with the G<sub>1</sub>/G<sub>2</sub> sub-class of human HCC. The *CTNNB1* mutant human tumors were classified under the G<sub>6</sub> category. In the *SRF-VP16<sup>iHep</sup>* tumors, we found molecular features corresponding to both G<sub>1</sub>/G<sub>2</sub> and G<sub>6</sub>. This indicates molecular heterogeneity of the *SRF-VP16<sup>iHep</sup>* tumors. Recent advances in genome-wide technologies can be expected in the future to assist a more refined characterization of human HCCs. Our multi-omic analysis of the *SRF-VP16<sup>iHep</sup>* mouse model might support the efforts currently being taken in the human cancer clinic. Comprehensive datasets, such as the human cancer genome atlas (TCGA), which contain information at multiple omics levels, will aid the formulation of the best possible tumor classification strategy. Mouse models, such as the *SRF-VP16<sup>iHep</sup>* line, can also be expected to support these efforts.

### 4.1.5 *SRF-VP16<sup>iHep</sup>* tumors closely resemble a subset of human HCCs

Comparative, unsupervised hierarchical clustering of *SRF-VP16<sup>iHep</sup>* mouse and human tumor was performed. This revealed that the *SRF-VP16<sup>iHep</sup>* samples clustered closely to a subset of human HCCs. For example, in the dataset from the HD-40 cohort of human patients, a subset of 10 HCCs clustered closely with our mouse samples (SC 10). These 10 HCCs were further characterized and similarities were observed in terms of the molecular characteristics of both human and murine tumors. For example, most of the SC 10 tumors were over-expressing the *IGF2* gene. The murine ortholog *Igf2* was found highly over-expressed in *SRF-VP16<sup>iHep</sup>* tumors as well. Further clustering analyses

---

with multiple other datasets also showed the presence of specific subsets of human HCCs that bore striking molecular similarities to the *SRF-VP16<sup>iHep</sup>* tumors.

#### **4.1.6 Key metabolic pathways are altered in *SRF-VP16<sup>iHep</sup>* HCCs**

Numerous metabolic pathways were found altered in the *SRF-VP16<sup>iHep</sup>* tumors. One example is the glycolysis pathway. Genes and proteins of the glycolysis pathway were found highly overexpressed in the *SRF-VP16<sup>iHep</sup>* tumors. The initiating enzyme of the pathway Hexokinase 1 is encoded by the gene *Hk1*, which is an SRF target gene. The enzymes involved in subsequent steps of glycolysis were all found up-regulated. This indicates aerobic glycolysis to be at work in *SRF-VP16<sup>iHep</sup>* tumors, a phenomenon called “Warburg effect”. Glycolysis generates pyruvate, which feeds the mitochondrial TCA cycle. However in the *SRF-VP16<sup>iHep</sup>* tumors, the TCA cycle enzymes were found highly down-regulated. This is in line with pyruvate being converted into lactate, as is the case in aerobic glycolysis. Many other mitochondrial enzymes were also found down regulated at the mRNA or protein levels. As a notable exception, the genes encoding mitochondrial enzymes of the one-carbon metabolism (1C-metabolism) were found over-expressed, at the expense of their cytoplasmic counterparts. This seems to indicate that the *SRF-VP16<sup>iHep</sup>* tumors efficiently engage in mitochondrial 1C-metabolism. In support, given that serine is required for the 1C-metabolic pathway, the enzymes involved in the serine synthesis pathway are also found up-regulated (at both the transcriptomic and proteomic levels). Thus, considerable metabolic reprogramming is occurring in *SRF-VP16<sup>iHep</sup>* HCCs.

## 4.2 Future Perspectives

### 4.2.1 Usage of *SRF-VP16<sup>iHep</sup>* mouse line

The *SRF-VP16<sup>iHep</sup>* mouse model promises to represent an apt tumor model for the study of human HCC. This may be particularly true for the G<sub>1</sub>/G<sub>2</sub> subtype of human HCC. The *SRF-VP16<sup>iHep</sup>* mouse line can be used for the study of other molecular features of liver tumors, such as cellular interactions within the tumor microenvironment, epigenetic regulation of cancer formation, metabolic reprogramming, and others. The mouse line could be used for the testing of various drugs targeting specific pathways that are highly dysregulated in liver cancer, specifically HCC.

### 4.2.2 Drug development for HCC treatment

There is an urgent need for new drugs in targeted therapy of human HCC. The use of mouse models such as *SRF-VP16<sup>iHep</sup>* promise to suit such needs. The multi-omics characterization of the *SRF-VP16<sup>iHep</sup>* mouse tumors has revealed significant insights into molecular mechanisms involved in hepatocarcinogenesis at various stages of tumor progression. These insights will support the identification of new therapeutic targets and of drugs affecting such target activities.

### 4.2.3 Effect of *SRF-VP16* activation in other cell types and organs.

Having observed the effect upon expression of *SRF-VP16* in hepatocytes has given rise to a new scope of developing mouse models for various cancers in different cell types. By following approaches comparable to *SRF-VP16<sup>iHep</sup>* tumor formation, *SRF-VP16<sup>iPan</sup>* mice could be generated, which express the *SRF-VP16* transgene in fibroblasts of the pancreas. This could help unravel the role of SRF in carcinogenesis of other tumor entities.

---

## 5 REFERENCES

---

- Aberle, H., Bauer, A., Stappert, J., Kispert, A., and Kemler, R. (1997). beta-catenin is a target for the ubiquitin-proteasome pathway. *EMBO J* 16, 3797-3804.
- Alberti, S., Krause, S.M., Kretz, O., Philippar, U., Lemberger, T., Casanova, E., Wiebel, F.F., Schwarz, H., Frotscher, M., Schutz, G., *et al.* (2005). Neuronal migration in the murine rostral migratory stream requires serum response factor. *Proc Natl Acad Sci U S A* 102, 6148-6153.
- Arsenian, S., Weinhold, B., Oelgeschlager, M., Ruther, U., and Nordheim, A. (1998). Serum response factor is essential for mesoderm formation during mouse embryogenesis. *EMBO J* 17, 6289-6299.
- Asara, J.M., Christofk, H.R., Freemark, L.M., and Cantley, L.C. (2008). A label-free quantification method by MS/MS TIC compared to SILAC and spectral counting in a proteomics screen. *Proteomics* 8, 994-999.
- Bai, S., Nasser, M.W., Wang, B., Hsu, S.H., Datta, J., Kutay, H., Yadav, A., Nuovo, G., Kumar, P., and Ghoshal, K. (2009). MicroRNA-122 inhibits tumorigenic properties of hepatocellular carcinoma cells and sensitizes these cells to sorafenib. *J Biol Chem* 284, 32015-32027.
- Bamshad, M.J., Ng, S.B., Bigham, A.W., Tabor, H.K., Emond, M.J., Nickerson, D.A., and Shendure, J. (2011). Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet* 12, 745-755.
- Boyault, S., Rickman, D.S., de Reynies, A., Balabaud, C., Rebouissou, S., Jeannot, E., Herault, A., Saric, J., Belghiti, J., Franco, D., *et al.* (2007). Transcriptome classification of HCC is related to gene alterations and to new therapeutic targets. *Hepatology* 45, 42-52.
- Braeuning, A. (2008). Beta-Catenin and Ha-ras - Master Regulators of Zonal Gene Expression in Mouse Liver? (Logos Verlag Berlin).
- Buchwalter, G., Gross, C., and Wasylyk, B. (2004). Ets ternary complex transcription factors. *Gene* 324, 1-14.
- Bumgarner, R. (2013). Overview of DNA microarrays: types, applications, and their future. *Curr Protoc Mol Biol Chapter 22, Unit 22 21.*
- Caillot, F., Derambure, C., Bioulac-Sage, P., Francois, A., Scotte, M., Gorla, O., Hiron, M., Daveau, M., and Salier, J.P. (2009). Transient and etiology-related transcription regulation in cirrhosis prior to hepatocellular carcinoma occurrence. *World J Gastroenterol* 15, 300-309.
- Cairo, S., Wang, Y., de Reynies, A., Durore, K., Dahan, J., Redon, M.J., Fabre, M., McClelland, M., Wang, X.W., Croce, C.M., *et al.* (2010). Stem cell-like micro-RNA

signature driven by Myc in aggressive liver cancer. *Proc Natl Acad Sci U S A* 107, 20471-20476.

Calvisi, D.F., Factor, V.M., Ladu, S., Conner, E.A., and Thorgeirsson, S.S. (2004). Disruption of beta-catenin pathway or genomic instability define two distinct categories of liver cancer in transgenic mice. *Gastroenterology* 126, 1374-1386.

Campbell, J.S., Hughes, S.D., Gilbertson, D.G., Palmer, T.E., Holdren, M.S., Haran, A.C., Odell, M.M., Bauer, R.L., Ren, H.P., Haugen, H.S., *et al.* (2005). Platelet-derived growth factor C induces liver fibrosis, steatosis, and hepatocellular carcinoma. *Proc Natl Acad Sci U S A* 102, 3389-3394.

Caudill, M.A., Wang, J.C., Melnyk, S., Pogribny, I.P., Jernigan, S., Collins, M.D., Santos-Guzman, J., Swendseid, M.E., Cogger, E.A., and James, S.J. (2001). Intracellular S-adenosylhomocysteine concentrations predict global DNA hypomethylation in tissues of methyl-deficient cystathionine beta-synthase heterozygous mice. *J Nutr* 131, 2811-2818.

Chai, J., and Tarnawski, A.S. (2002). Serum response factor: discovery, biochemistry, biological roles and implications for tissue injury healing. *J Physiol Pharmacol* 53, 147-157.

Chen, R., Mias, G.I., Li-Pook-Than, J., Jiang, L., Lam, H.Y., Chen, R., Miriami, E., Karczewski, K.J., Hariharan, M., Dewey, F.E., *et al.* (2012). Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell* 148, 1293-1307.

Cheng, P., Sun, X., Yin, D., Xu, F., Yang, K., Qin, L., Dong, Y., Guo, F., Chen, A., Zhang, W., *et al.* (2015). Nanog down-regulates the Wnt signaling pathway via beta-catenin phosphorylation during epidermal stem cell proliferation and differentiation. *Cell Biosci* 5, 5.

Choi, B.H., Choi, M., Jeon, H.Y., and Rho, H.M. (2001). Hepatitis B viral X protein overcomes inhibition of E2F1 activity by pRb on the human Rb gene promoter. *DNA Cell Biol* 20, 75-80.

Choi, H.N., Kim, K.R., Lee, J.H., Park, H.S., Jang, K.Y., Chung, M.J., Hwang, S.E., Yu, H.C., and Moon, W.S. (2009). Serum response factor enhances liver metastasis of colorectal carcinoma via alteration of the E-cadherin/beta-catenin complex. *Oncol Rep* 21, 57-63.

Cholongitas, E., Mamou, C., Rodriguez-Castro, K.I., and Burra, P. (2014). Mammalian target of rapamycin inhibitors are associated with lower rates of hepatocellular carcinoma recurrence after liver transplantation: a systematic review. *Transpl Int* 27, 1039-1049.

Dalton, S., Marais, R., Wynne, J., and Treisman, R. (1993). Isolation and characterization of SRF accessory proteins. *Philos Trans R Soc Lond B Biol Sci* 340, 325-332.

Dalton, S., and Treisman, R. (1992). Characterization of SAP-1, a protein recruited by serum response factor to the c-fos serum response element. *Cell* 68, 597-612.

Davies, K. (2010). 13 years ago, a beer summit in an English pub led to the birth of SOLEXA. In *BioIT World*.



- De Corte, V., Demol, H., Goethals, M., Van Damme, J., Gettemans, J., and Vandekerckhove, J. (1999). Identification of Tyr438 as the major in vitro c-Src phosphorylation site in human gelsolin: a mass spectrometric approach. *Protein Sci* 8, 234-241.
- de La Coste, A., Romagnolo, B., Billuart, P., Renard, C.A., Buendia, M.A., Soubrane, O., Fabre, M., Chelly, J., Beldjord, C., Kahn, A., *et al.* (1998). Somatic mutations of the beta-catenin gene are frequent in mouse and human hepatocellular carcinomas. *Proc Natl Acad Sci U S A* 95, 8847-8851.
- Delhaye, M., Louis, H., Degraef, C., Le Moine, O., Deviere, J., Gulbis, B., Jacobovitz, D., Adler, M., and Galand, P. (1996). Relationship between hepatocyte proliferative activity and liver functional reserve in human cirrhosis. *Hepatology* 23, 1003-1011.
- Eferl, R., and Wagner, E.F. (2003). AP-1: a double-edged sword in tumorigenesis. *Nat Rev Cancer* 3, 859-868.
- Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 95, 14863-14868.
- El-Serag, H.B., and Rudolph, K.L. (2007). Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. *Gastroenterology* 132, 2557-2576.
- Esnault, C., Stewart, A., Gualdrini, F., East, P., Horswell, S., Matthews, N., and Treisman, R. (2014). Rho-actin signaling to the MRTF coactivators dominates the immediate transcriptional response to serum in fibroblasts. *Genes Dev* 28, 943-958.
- European Association for Study of, L., European Organisation for, R., and Treatment of, C. (2012). EASL-EORTC clinical practice guidelines: management of hepatocellular carcinoma. *Eur J Cancer* 48, 599-641.
- Farazi, P.A., and DePinho, R.A. (2006). Hepatocellular carcinoma pathogenesis: from genes to environment. *Nat Rev Cancer* 6, 674-687.
- Farazi, P.A., Glickman, J., Horner, J., and Depinho, R.A. (2006). Cooperative interactions of p53 mutation, telomere dysfunction, and chronic liver damage in hepatocellular carcinoma progression. *Cancer Res* 66, 4766-4773.
- Farazi, P.A., Glickman, J., Jiang, S., Yu, A., Rudolph, K.L., and DePinho, R.A. (2003). Differential impact of telomere dysfunction on initiation and progression of hepatocellular carcinoma. *Cancer Res* 63, 5021-5027.
- Fausto, N., and Campbell, J.S. (2010). Mouse models of hepatocellular carcinoma. *Semin Liver Dis* 30, 87-98.
- Feitelson, M.A., Sun, B., Satiroglu Tufan, N.L., Liu, J., Pan, J., and Lian, Z. (2002). Genetic mechanisms of hepatocarcinogenesis. *Oncogene* 21, 2593-2604.
- Ferber, M.J., Montoya, D.P., Yu, C., Aderca, I., McGee, A., Thorland, E.C., Nagorney, D.M., Gostout, B.S., Burgart, L.J., Boix, L., *et al.* (2003). Integrations of the hepatitis B virus (HBV) and human papillomavirus (HPV) into the human telomerase reverse transcriptase (hTERT) gene in liver and cervical cancers. *Oncogene* 22, 3813-3820.

- Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D., and Bray, F. (2015). Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer* 136, E359-386.
- Fraisl, P., Baes, M., and Carmeliet, P. (2008). Hungry for blood vessels: linking metabolism and angiogenesis. *Dev Cell* 14, 313-314.
- Franco, C.A., Blanc, J., Parlakian, A., Blanco, R., Aspalter, I.M., Kazakova, N., Diguët, N., Mylonas, E., Gao-Li, J., Vaahtokari, A., *et al.* (2013). SRF selectively controls tip cell invasive behavior in angiogenesis. *Development* 140, 2321-2333.
- Ganeshan, D., Szklaruk, J., Kundra, V., Kaseb, A., Rashid, A., and Elsayes, K.M. (2014). Imaging features of fibrolamellar hepatocellular carcinoma. *AJR Am J Roentgenol* 202, 544-552.
- Giannelli, G., Koudelkova, P., Dituri, F., and Mikulits, W. (2016). Role of epithelial to mesenchymal transition in hepatocellular carcinoma. *J Hepatol* 65, 798-808.
- Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E.M., Brockman, W., Fennell, T., Giannoukos, G., Fisher, S., Russ, C., *et al.* (2009). Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol* 27, 182-189.
- Greenberg, M.E., and Ziff, E.B. (1984). Stimulation of 3T3 cells induces transcription of the c-fos proto-oncogene. *Nature* 311, 433-438.
- Griffiths, J.R., McSheehy, P.M., Robinson, S.P., Troy, H., Chung, Y.L., Leek, R.D., Williams, K.J., Stratford, I.J., Harris, A.L., and Stubbs, M. (2002). Metabolic changes detected by in vivo magnetic resonance studies of HEPA-1 wild-type tumors and tumors deficient in hypoxia-inducible factor-1beta (HIF-1beta): evidence of an anabolic role for the HIF-1 pathway. *Cancer Res* 62, 688-695.
- Gu, S., Fang, L., and Xu, X. (2013). Using SOAPaligner for Short Reads Alignment. *Curr Protoc Bioinformatics* 44, 11.11.11-17.
- Gygi, S.P., Corthals, G.L., Zhang, Y., Rochon, Y., and Aebersold, R. (2000). Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology. *Proc Natl Acad Sci U S A* 97, 9390-9395.
- Hanahan, D., and Weinberg, R.A. (2000). The hallmarks of cancer. *Cell* 100, 57-70.
- He, L., Tian, D.A., Li, P.Y., and He, X.X. (2015). Mouse models of liver cancer: Progress and recommendations. *Oncotarget* 6, 23306-23322.
- Heindryckx, F., Colle, I., and Van Vlierberghe, H. (2009). Experimental mouse models for hepatocellular carcinoma research. *Int J Exp Pathol* 90, 367-386.
- Hermann, M.R., Jakobson, M., Colo, G.P., Rognoni, E., Jakobson, M., Kupatt, C., Posern, G., and Fassler, R. (2016). Integrins synergise to induce expression of the MRTF-A-SRF target gene ISG15 for promoting cancer cell invasion. *J Cell Sci* 129, 1391-1403.
- Hines, I.N., Hartwell, H.J., Feng, Y., Theve, E.J., Hall, G.A., Hashway, S., Connolly, J., Fecteau, M., Fox, J.G., and Rogers, A.B. (2011). Insulin resistance and metabolic

- hepatocarcinogenesis with parent-of-origin effects in AxB mice. *Am J Pathol* 179, 2855-2865.
- Ho, D.W., Kai, A.K., and Ng, I.O. (2015). TCGA whole-transcriptome sequencing data reveals significantly dysregulated genes and signaling pathways in hepatocellular carcinoma. *Front Med* 9, 322-330.
- Hodges, E., Xuan, Z., Balija, V., Kramer, M., Molla, M.N., Smith, S.W., Middle, C.M., Rodesch, M.J., Albert, T.J., Hannon, G.J., *et al.* (2007). Genome-wide in situ exon capture for selective resequencing. *Nat Genet* 39, 1522-1527.
- Hopfner, M., Sutter, A.P., Huether, A., Schuppan, D., Zeitz, M., and Scherubl, H. (2004). Targeting the epidermal growth factor receptor by gefitinib for treatment of hepatocellular carcinoma. *J Hepatol* 41, 1008-1016.
- Hsieh, P., and Yamane, K. (2008). DNA mismatch repair: molecular mechanism, cancer, and ageing. *Mech Ageing Dev* 129, 391-407.
- Hu, L., Ye, M., Jiang, X., Feng, S., and Zou, H. (2007). Advances in hyphenated analytical techniques for shotgun proteome and peptidome analysis--a review. *Anal Chim Acta* 598, 193-204.
- Hussain, S.P., Schwank, J., Staib, F., Wang, X.W., and Harris, C.C. (2007). TP53 mutations and hepatocellular carcinoma: insights into the etiology and pathogenesis of liver cancer. *Oncogene* 26, 2166-2176.
- Iansante, V., Choy, P.M., Fung, S.W., Liu, Y., Chai, J.G., Dyson, J., Del Rio, A., D'Santos, C., Williams, R., Chokshi, S., *et al.* (2015). PARP14 promotes the Warburg effect in hepatocellular carcinoma by inhibiting JNK1-dependent PKM2 phosphorylation and activation. *Nat Commun* 6, 7882.
- Ishizaki, Y., Ikeda, S., Fujimori, M., Shimizu, Y., Kurihara, T., Itamoto, T., Kikuchi, A., Okajima, M., and Asahara, T. (2004). Immunohistochemical analysis and mutational analyses of beta-catenin, Axin family and APC genes in hepatocellular carcinomas. *Int J Oncol* 24, 1077-1083.
- Jakubczak, J.L., Chisari, F.V., and Merlino, G. (1997). Synergy between transforming growth factor alpha and hepatitis B virus surface antigen in hepatocellular proliferation and carcinogenesis. *Cancer Res* 57, 3606-3611.
- Jeter, C.R., Yang, T., Wang, J., Chao, H.P., and Tang, D.G. (2015). Concise Review: NANOG in Cancer Stem Cells and Tumor Development: An Update and Outstanding Questions. *Stem Cells* 33, 2381-2390.
- Kaddurah-Daouk, R., Kristal, B.S., and Weinshilboum, R.M. (2008). Metabolomics: a global biochemical approach to drug response and disease. *Annu Rev Pharmacol Toxicol* 48, 653-683.
- Kandula, M., Ch, K.K., and Ys, A.R. (2013). Molecular Mechanism and Targeted Therapy Options of Triple-Negative (ER, PgR, HER-2/neu) Breast Cancer: Review. *World J Oncol* 4, 137-141.

Kaplan-Albuquerque, N., Van Putten, V., Weiser-Evans, M.C., and Nemenoff, R.A. (2005). Depletion of serum response factor by RNA interference mimics the mitogenic effects of platelet derived growth factor-BB in vascular smooth muscle cells. *Circ Res* 97, 427-433.

Kircher, P., Hermanns, C., Nossek, M., Drexler, M.K., Grosse, R., Fischer, M., Sarikas, A., Penkava, J., Lewis, T., Prywes, R., *et al.* (2015). Filamin A interacts with the coactivator MKL1 to promote the activity of the transcription factor SRF and cell migration. *Sci Signal* 8, ra112.

Kobayashi, T., Kubota, K., Takayama, T., and Makuuchi, M. (2001). Telomerase activity as a predictive marker for recurrence of hepatocellular carcinoma after hepatectomy. *Am J Surg* 181, 284-288.

Kojiro, M. (1998). Pathology of early hepatocellular carcinoma: progression from early to advanced. *Hepatology* 45 Suppl 3, 1203-1205.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860-921.

Latos, P.A., Pauler, F.M., Koerner, M.V., Senergin, H.B., Hudson, Q.J., Stocsits, R.R., Allhoff, W., Stricker, S.H., Klement, R.M., Warczok, K.E., *et al.* (2012). Airn transcriptional overlap, but not its lncRNA products, induces imprinted *Igf2r* silencing. *Science* 338, 1469-1472.

Lee, S., Lee, H.J., Kim, J.H., Lee, H.S., Jang, J.J., and Kang, G.H. (2003). Aberrant CpG island hypermethylation along multistep hepatocarcinogenesis. *Am J Pathol* 163, 1371-1378.

Lencioni, R., Petruzzi, P., and Crocetti, L. (2013). Chemoembolization of hepatocellular carcinoma. *Semin Intervent Radiol* 30, 3-11.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754-1760.

Li, Y., Tang, Z.Y., and Hou, J.X. (2011). Hepatocellular carcinoma: insight from animal models. *Nat Rev Gastroenterol Hepatol* 9, 32-43.

Liberti, M.V., and Locasale, J.W. (2016). The Warburg Effect: How Does it Benefit Cancer Cells? *Trends Biochem Sci* 41, 211-218.

Liu, C.Y., Chan, S.W., Guo, F., Toloczko, A., Cui, L., and Hong, W. (2016). MRTF/SRF dependent transcriptional regulation of TAZ in breast cancer cells. *Oncotarget* 7, 13706-13716.

Liu, P., Kimmoun, E., Legrand, A., Sauvanet, A., Degott, C., Lardeux, B., and Bernuau, D. (2002). Activation of NF-kappa B, AP-1 and STAT transcription factors is a frequent and early event in human hepatocellular carcinomas. *J Hepatol* 37, 63-71.

Llovet, J.M., Schwartz, M., and Mazzaferro, V. (2005). Resection and liver transplantation for hepatocellular carcinoma. *Semin Liver Dis* 25, 181-200.

- Llovet, J.M., Villanueva, A., Lachenmayer, A., and Finn, R.S. (2015). Advances in targeted therapies for hepatocellular carcinoma in the genomic era. *Nat Rev Clin Oncol* 12, 436.
- Locasale, J.W. (2013). Serine, glycine and one-carbon units: cancer metabolism in full circle. *Nat Rev Cancer* 13, 572-583.
- Ma, C., Wei, S., and Song, Y. (2011). T790M and acquired resistance of EGFR TKI: a literature review of clinical reports. *J Thorac Dis* 3, 10-18.
- Macek, B., Mann, M., and Olsen, J.V. (2009). Global and site-specific quantitative phosphoproteomics: principles and applications. *Annu Rev Pharmacol Toxicol* 49, 199-221.
- Maeta, Y., Shiota, G., Okano, J., and Murawaki, Y. (2005). Effect of promoter methylation of the p16 gene on phosphorylation of retinoblastoma gene product and growth of hepatocellular carcinoma cells. *Tumour Biol* 26, 300-305.
- Manning, C.S., Hooper, S., and Sahai, E.A. (2015). Intravital imaging of SRF and Notch signalling identifies a key role for EZH2 in invasive melanoma cells. *Oncogene* 34, 4320-4332.
- Marquardt, J.U., Seo, D., Andersen, J.B., Gillen, M.C., Kim, M.S., Conner, E.A., Galle, P.R., Factor, V.M., Park, Y.N., and Thorgeirsson, S.S. (2014). Sequential transcriptome analysis of human liver cancer indicates late stage acquisition of malignant traits. *J Hepatol* 60, 346-353.
- McGettigan, P.A. (2013). Transcriptomics in the RNA-seq era. *Curr Opin Chem Biol* 17, 4-11.
- Medjkane, S., Perez-Sanchez, C., Gaggioli, C., Sahai, E., and Treisman, R. (2009). Myocardin-related transcription factors and SRF are required for cytoskeletal dynamics and experimental metastasis. *Nat Cell Biol* 11, 257-268.
- Menni, C., Zhai, G., Macgregor, A., Prehn, C., Romisch-Margl, W., Suhre, K., Adamski, J., Cassidy, A., Illig, T., Spector, T.D., *et al.* (2013). Targeted metabolomics profiles are strongly correlated with nutritional patterns in women. *Metabolomics* 9, 506-514.
- Miano, J.M. (2003). Serum response factor: toggling between disparate programs of gene expression. *J Mol Cell Cardiol* 35, 577-593.
- Miano, J.M. (2010). Role of serum response factor in the pathogenesis of disease. *Lab Invest* 90, 1274-1284.
- Minouchi, K., Kaneko, S., and Kobayashi, K. (2002). Mutation of p53 gene in regenerative nodules in cirrhotic liver. *J Hepatol* 37, 231-239.
- Miralles, F., Posern, G., Zaromytidou, A.I., and Treisman, R. (2003). Actin dynamics control SRF activity by regulation of its coactivator MAL. *Cell* 113, 329-342.
- Mittelstrass, K., Ried, J.S., Yu, Z., Krumsiek, J., Gieger, C., Prehn, C., Roemisch-Margl, W., Polonikov, A., Peters, A., Theis, F.J., *et al.* (2011). Discovery of sexual dimorphisms in metabolic and genetic biomarkers. *PLoS Genet* 7, e1002215.

Murakami, H., Sanderson, N.D., Nagy, P., Marino, P.A., Merlino, G., and Thorgeirsson, S.S. (1993). Transgenic mouse model for synergistic effects of nuclear oncogenes and growth factors in tumorigenesis: interaction of c-myc and transforming growth factor alpha in hepatic oncogenesis. *Cancer Res* 53, 1719-1723.

Murthy, S.R.K., Dupart, E., Al-Sweel, N., Chen, A., Cawley, N.X., and Loh, Y.P. (2013). Carboxypeptidase E promotes cancer cell survival, but inhibits migration and invasion. *Cancer Lett* 341, 204-213.

N'Kontchou, G., Mahamoudi, A., Aout, M., Ganne-Carrie, N., Grando, V., Coderc, E., Vicaut, E., Trinchet, J.C., Sellier, N., Beaugrand, M., *et al.* (2009). Radiofrequency ablation of hepatocellular carcinoma: long-term results and prognostic factors in 235 Western patients with cirrhosis. *Hepatology* 50, 1475-1483.

Nakashima, O., Sugihara, S., Kage, M., and Kojiro, M. (1995). Pathomorphologic characteristics of small hepatocellular carcinoma: a special reference to small hepatocellular carcinoma with indistinct margins. *Hepatology* 22, 101-105.

Nan, X., Xie, C., Yu, X., and Liu, J. (2017). EGFR TKI as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer. *Oncotarget* 8, 75712-75726.

Ng, S.B., Buckingham, K.J., Lee, C., Bigham, A.W., Tabor, H.K., Dent, K.M., Huff, C.D., Shannon, P.T., Jabs, E.W., Nickerson, D.A., *et al.* (2010). Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* 42, 30-35.

Ng, S.B., Turner, E.H., Robertson, P.D., Flygare, S.D., Bigham, A.W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E.E., *et al.* (2009). Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 461, 272-276.

O'Farrell, P.H. (1975). High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* 250, 4007-4021.

Ohrnberger, S., Thavamani, A., Braeuning, A., Lipka, D.B., Kirilov, M., Geffers, R., Autenrieth, S.E., Romer, M., Zell, A., Bonin, M., *et al.* (2015). Dysregulated serum response factor triggers formation of hepatocellular carcinoma. *Hepatology* 61, 979-989.

Park, H.S., Jang, K.Y., Kim, Y.K., Cho, B.H., and Moon, W.S. (2009). Hepatocellular carcinoma with massive lymphoid infiltration: a regressing phenomenon? *Pathol Res Pract* 205, 648-652.

Park, M.Y., Kim, K.R., Park, H.S., Park, B.H., Choi, H.N., Jang, K.Y., Chung, M.J., Kang, M.J., Lee, D.G., and Moon, W.S. (2007). Expression of the serum response factor in hepatocellular carcinoma: implications for epithelial-mesenchymal transition. *Int J Oncol* 31, 1309-1315.

Park, S.A., Kim, M.J., Park, S.Y., Kim, J.S., Lim, W., Nam, J.S., and Yhong Sheen, Y. (2015). TIMP-1 mediates TGF-beta-dependent crosstalk between hepatic stellate and cancer cells via FAK signaling. *Sci Rep* 5, 16492.

Parlakian, A., Charvet, C., Escoubet, B., Mericskay, M., Molkenstin, J.D., Gary-Bobo, G., De Windt, L.J., Ludosky, M.A., Paulin, D., Daegelen, D., *et al.* (2005). Temporally

- controlled onset of dilated cardiomyopathy through disruption of the SRF gene in adult heart. *Circulation* 112, 2930-2939.
- Parlakian, A., Tuil, D., Hamard, G., Tavernier, G., Hentzen, D., Concordet, J.P., Paulin, D., Li, Z., and Daegelen, D. (2004). Targeted inactivation of serum response factor in the developing heart results in myocardial defects and embryonic lethality. *Mol Cell Biol* 24, 5281-5289.
- Pellegrini, L., Tan, S., and Richmond, T.J. (1995). Structure of serum response factor core bound to DNA. *Nature* 376, 490-498.
- Pinkse, M.W., Uitto, P.M., Hilhorst, M.J., Ooms, B., and Heck, A.J. (2004). Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal Chem* 76, 3935-3943.
- Prabakaran, S., Lippens, G., Steen, H., and Gunawardena, J. (2012). Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding. *Wiley Interdiscip Rev Syst Biol Med* 4, 565-583.
- Rimassa, L., and Santoro, A. (2009). Sorafenib therapy in advanced hepatocellular carcinoma: the SHARP trial. *Expert Rev Anticancer Ther* 9, 739-745.
- Rochfort, S. (2005). Metabolomics reviewed: a new "omics" platform technology for systems biology and implications for natural products research. *J Nat Prod* 68, 1813-1820.
- Ryder, S.D., and British Society of, G. (2003). Guidelines for the diagnosis and treatment of hepatocellular carcinoma (HCC) in adults. *Gut* 52 Suppl 3, iii1-8.
- Salesse, S., and Verfaillie, C.M. (2002). BCR/ABL: from molecular mechanisms of leukemia induction to treatment of chronic myelogenous leukemia. *Oncogene* 21, 8547-8559.
- Sanyal, A.J., Yoon, S.K., and Lencioni, R. (2010). The etiology of hepatocellular carcinoma and consequences for treatment. *Oncologist* 15 Suppl 4, 14-22.
- Schinke-Braun, M., and Couget, J.A. (2007). Expression profiling using affymetrix genechip probe arrays. *Methods Mol Biol* 366, 13-40.
- Schratt, G., Philippar, U., Berger, J., Schwarz, H., Heidenreich, O., and Nordheim, A. (2002). Serum response factor is crucial for actin cytoskeletal organization and focal adhesion assembly in embryonic stem cells. *J Cell Biol* 156, 737-750.
- Schratt, G., Weinhold, B., Lundberg, A.S., Schuck, S., Berger, J., Schwarz, H., Weinberg, R.A., Ruther, U., and Nordheim, A. (2001). Serum response factor is required for immediate-early gene activation yet is dispensable for proliferation of embryonic stem cells. *Mol Cell Biol* 21, 2933-2943.
- Shafizadeh, N., and Kakar, S. (2013). Hepatocellular Carcinoma: Histologic Subtypes. *Surg Pathol Clin* 6, 367-384.
- Shaw, P.E., Schroter, H., and Nordheim, A. (1989). The ability of a ternary complex to form over the serum response element correlates with serum inducibility of the human c-fos promoter. *Cell* 56, 563-572.

Shevchenko, A., and Simons, K. (2010). Lipidomics: coming to grips with lipid diversity. *Nat Rev Mol Cell Biol* 11, 593-598.

Shimizu, A., Kobayashi, A., Motoyama, H., Sakai, H., Yamada, A., Yoshizawa, A., Momose, M., Kadoya, M., and Miyagawa, S. (2015). Features of acute liver congestion on gadoxetate disodium-enhanced MRI in a rat model: Role of organic anion-transporting polypeptide 1A1. *J Magn Reson Imaging* 42, 828-836.

Shirakawa, H., Suzuki, H., Shimomura, M., Kojima, M., Gotohda, N., Takahashi, S., Nakagohri, T., Konishi, M., Kobayashi, N., Kinoshita, T., *et al.* (2009). Glypican-3 expression is correlated with poor prognosis in hepatocellular carcinoma. *Cancer Sci* 100, 1403-1407.

Shore, P., and Sharrocks, A.D. (1995). The MADS-box family of transcription factors. *Eur J Biochem* 229, 1-13.

Simon, F., Bockhorn, M., Praha, C., Baba, H.A., Broelsch, C.E., Frilling, A., and Weber, F. (2010). Deregulation of HIF1-alpha and hypoxia-regulated pathways in hepatocellular carcinoma and corresponding non-malignant liver tissue--influence of a modulated host stroma on the prognosis of HCC. *Langenbecks Arch Surg* 395, 395-405.

Soda, M., Choi, Y.L., Enomoto, M., Takada, S., Yamashita, Y., Ishikawa, S., Fujiwara, S., Watanabe, H., Kurashina, K., Hatanaka, H., *et al.* (2007). Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature* 448, 561-566.

Soulez, M., Rouviere, C.G., Chafey, P., Hentzen, D., Vandromme, M., Lautredou, N., Lamb, N., Kahn, A., and Tuil, D. (1996). Growth and differentiation of C2 myogenic cells are dependent on serum response factor. *Mol Cell Biol* 16, 6065-6074.

Stern, S., Sinske, D., and Knoll, B. (2012). Serum response factor modulates neuron survival during peripheral axon injury. *J Neuroinflammation* 9, 78.

Takayama, T., Makuuchi, M., Hirohashi, S., Sakamoto, M., Yamamoto, J., Shimada, K., Kosuge, T., Okada, S., Takayasu, K., and Yamasaki, S. (1998). Early hepatocellular carcinoma as an entity with a high rate of surgical cure. *Hepatology* 28, 1241-1246.

Tarasov, K.V., Brugh, S.A., Tarasova, Y.S., and Boheler, K.R. (2007). Serial Analysis of Gene Expression (SAGE): a useful tool to analyze the cardiac transcriptome. *Methods Mol Biol* 366, 41-59.

Teer, J.K., and Mullikin, J.C. (2010). Exome sequencing: the sweet spot before whole genomes. *Hum Mol Genet* 19, R145-151.

Thillai, K., Ross, P., and Sarker, D. (2016). Molecularly targeted therapy for advanced hepatocellular carcinoma - a drug development crisis? *World J Gastrointest Oncol* 8, 173-185.

Thorgeirsson, S.S., and Grisham, J.W. (2002). Molecular pathogenesis of human hepatocellular carcinoma. *Nat Genet* 31, 339-346.

Tomlins, S.A., Rhodes, D.R., Perner, S., Dhanasekaran, S.M., Mehra, R., Sun, X.W., Varambally, S., Cao, X., Tchinda, J., Kuefer, R., *et al.* (2005). Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* 310, 644-648.



- Toricelli, M., Melo, F.H.M., Hunger, A., Zanatta, D., Strauss, B.E., and Jasiulionis, M.G. (2017). Timp1 Promotes Cell Survival by Activating the PDK1 Signaling Pathway in Melanoma. *Cancers (Basel)* 9.
- Tornesello, M.L., Buonaguro, L., Tatangelo, F., Botti, G., Izzo, F., and Buonaguro, F.M. (2013). Mutations in TP53, CTNNB1 and PIK3CA genes in hepatocellular carcinoma associated with hepatitis B and hepatitis C virus infections. *Genomics* 102, 74-83.
- Treisman, R. (1986). Identification of a protein-binding site that mediates transcriptional response of the c-fos gene to serum factors. *Cell* 46, 567-574.
- Turner, E.H., Lee, C., Ng, S.B., Nickerson, D.A., and Shendure, J. (2009). Massively parallel exon capture and library-free resequencing across 16 genomes. *Nat Methods* 6, 315-316.
- Tyers, M., and Mann, M. (2003). From genomics to proteomics. *Nature* 422, 193-197.
- Urabe, Y., Nouse, K., Higashi, T., Nakatsukasa, H., Hino, N., Ashida, K., Kinugasa, N., Yoshida, K., Uematsu, S., and Tsuji, T. (1996). Telomere length in human liver diseases. *Liver* 16, 293-297.
- van Ravenzwaay, B., Cunha, G.C., Leibold, E., Looser, R., Mellert, W., Prokoudine, A., Walk, T., and Wiemer, J. (2007). The use of metabolomics for the discovery of new biomarkers of effect. *Toxicol Lett* 172, 21-28.
- Villen, J., and Gygi, S.P. (2008). The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nat Protoc* 3, 1630-1638.
- Waller, L.P., Deshpande, V., and Pysopoulos, N. (2015). Hepatocellular carcinoma: A comprehensive review. *World J Hepatol* 7, 2648-2663.
- Wang, Q., Xia, J., Jia, P., Pao, W., and Zhao, Z. (2013). Application of next generation sequencing to human gene fusion detection: computational tools, features and perspectives. *Brief Bioinform* 14, 506-519.
- Waxman, D.J., and Azaroff, L. (1992). Phenobarbital induction of cytochrome P-450 gene expression. *Biochem J* 281 ( Pt 3), 577-592.
- Wege, H., and Brummendorf, T.H. (2007). Telomerase activation in liver regeneration and hepatocarcinogenesis: Dr. Jekyll or Mr. Hyde? *Curr Stem Cell Res Ther* 2, 31-38.
- Wei, L., Li, Y., and Suo, Z. (2015). TSPAN8 promotes gastric cancer growth and metastasis via ERK MAPK pathway. *Int J Clin Exp Med* 8, 8599-8607.
- Weinert, T., and Lydall, D. (1993). Cell cycle checkpoints, genetic instability and cancer. *Semin Cancer Biol* 4, 129-140.
- Werth, D., Grassi, G., Konjer, N., Dapas, B., Farra, R., Giansante, C., Kandolf, R., Guarnieri, G., Nordheim, A., and Heidenreich, O. (2010). Proliferation of human primary vascular smooth muscle cells depends on serum response factor. *Eur J Cell Biol* 89, 216-224.

Wiebel, F.F., Rennekampff, V., Vintersten, K., and Nordheim, A. (2002). Generation of mice carrying conditional knockout alleles for the transcription factor SRF. *Genesis* 32, 124-126.

Wiemann, S.U., Satyanarayana, A., Tsahuridu, M., Tillmann, H.L., Zender, L., Klempnauer, J., Flemming, P., Franco, S., Blasco, M.A., Manns, M.P., *et al.* (2002). Hepatocyte telomere shortening and senescence are general markers of human liver cirrhosis. *FASEB J* 16, 935-942.

Wilhelm, S., Carter, C., Lynch, M., Lowinger, T., Dumas, J., Smith, R.A., Schwartz, B., Simantov, R., and Kelley, S. (2006). Discovery and development of sorafenib: a multikinase inhibitor for treating cancer. *Nat Rev Drug Discov* 5, 835-844.

Wilkins, M.R., Pasquali, C., Appel, R.D., Ou, K., Golaz, O., Sanchez, J.C., Yan, J.X., Gooley, A.A., Hughes, G., Humphery-Smith, I., *et al.* (1996). From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnology (N Y)* 14, 61-65.

Wu, L., Tang, Z.Y., and Li, Y. (2009). Experimental models of hepatocellular carcinoma: developments and evolution. *J Cancer Res Clin Oncol* 135, 969-981.

Xu, T., Zhu, Y., Xiong, Y., Ge, Y.Y., Yun, J.P., and Zhuang, S.M. (2009). MicroRNA-195 suppresses tumorigenicity and regulates G<sub>1</sub>/S transition of human hepatocellular carcinoma cells. *Hepatology* 50, 113-121.

Xu, X., Zhao, Z., Guo, S., Li, J., Liu, S., You, Y., Ni, B., Wang, H., and Bie, P. (2017). Increased semaphorin 3c expression promotes tumor growth and metastasis in pancreatic ductal adenocarcinoma by activating the ERK1/2 signaling pathway. *Cancer Lett* 397, 12-22.

Yamaguchi, R., Nakashima, O., Yano, H., Kutami, R., Kusaba, A., and Kojiro, M. (1997). Hepatocellular carcinoma with sarcomatous change. *Oncol Rep* 4, 525-529.

Yamamoto, H., Iku, S., Itoh, F., Tang, X., Hosokawa, M., and Imai, K. (2001). Association of trypsin expression with recurrence and poor prognosis in human esophageal squamous cell carcinoma. *Cancer* 91, 1324-1331.

Yang, S.H., Watanabe, J., Nakashima, O., and Kojiro, M. (1996). Clinicopathologic study on clear cell hepatocellular carcinoma. *Pathol Int* 46, 503-509.

Yong, K.J., Chai, L., and Tenen, D.G. (2013). Oncofetal gene SALL4 in aggressive hepatocellular carcinoma. *N Engl J Med* 369, 1171-1172.

Zeng, X., Yin, F., Liu, X., Xu, J., Xu, Y., Huang, J., Nan, Y., and Qiu, X. (2014). Upregulation of E2F transcription factor 3 is associated with poor prognosis in hepatocellular carcinoma. *Oncol Rep* 31, 1139-1146.

Zhang, H., Liu, T., Zhang, Z., Payne, S.H., Zhang, B., McDermott, J.E., Zhou, J.Y., Petyuk, V.A., Chen, L., Ray, D., *et al.* (2016a). Integrated Proteogenomic Characterization of Human High-Grade Serous Ovarian Cancer. *Cell* 166, 755-765.

- 
- Zhang, H., Liu, Z., Ma, S., Zhang, H., Kong, F., He, Y., Yang, X., Wang, Y., Xu, H., Yang, A., *et al.* (2016b). Ratio of S-adenosylmethionine to S-adenosylhomocysteine as a sensitive indicator of atherosclerosis. *Mol Med Rep* 14, 289-300.
- Zhou, X.U., Lu, J., and Zhu, H. (2016). Correlation between the expression of hTERT gene and the clinicopathological characteristics of hepatocellular carcinoma. *Oncol Lett* 11, 111-115.
- Zhu, W., Smith, J.W., and Huang, C.M. (2010). Mass spectrometry-based label-free quantitative proteomics. *J Biomed Biotechnol* 2010, 840518.
- Ziogas, I.A., and Tsoulfas, G. (2017). Evolving role of Sorafenib in the management of hepatocellular carcinoma. *World J Clin Oncol* 8, 203-213.
- Zong, W.X., Rabinowitz, J.D., and White, E. (2016). Mitochondria and Cancer. *Mol Cell* 61, 667-676.

## 6 ABBREVIATIONS DIRECTORY

---

<i>μL</i>	<i>Microliters</i>
<i>A2m</i>	<i>alpha-2-macroglobulin</i>
<i>Acsl4</i>	<i>acyl-CoA synthetase long-chain family member 4</i>
<i>Acss3</i>	<i>acyl-CoA synthetase short-chain family member 3</i>
<i>Actb</i>	<i>actin, beta</i>
<i>Afp</i>	<i>alpha fetoprotein</i>
<i>Ahnak</i>	<i>AHNAK nucleoprotein (desmoyokin)</i>
<i>Aldh1a2</i>	<i>aldehyde dehydrogenase family 1, subfamily A2</i>
<i>Aldh1l2</i>	<i>aldehyde dehydrogenase 1 family, member L2</i>
<i>Alfp</i>	<i>Alpha fetoprotein</i>
<i>Ankrd1</i>	<i>ankyrin repeat domain 1 (cardiac muscle)</i>
<i>Ap 1</i>	<i>Activator protein -1</i>
<i>ASC</i>	<i>apoptosis associated speck like protein</i>
<i>Atf3</i>	<i>activating transcription factor 3</i>
<i>ATF5</i>	<i>Activating Transcription Factor 5</i>
<i>ATP</i>	<i>Adenosine Tri phosphate</i>
<i>Aurka</i>	<i>aurora kinase A</i>
<i>Axin2</i>	<i>axin2</i>
<i>BAC</i>	<i>Bacterial Artificial Chromosomes</i>
<i>Bcl2</i>	<i>B cell leukemia/lymphoma 2</i>

---

<i>Bex1</i>	<i>brain expressed gene 1</i>
<i>bp</i>	<i>base pairs</i>
<i>Bub1</i>	<i>budding uninhibited by benzimidazoles 1 homolog (S. cerevisiae)</i>
<i>c-fos</i>	<i>FBJ murine osteosarcoma</i>
<i>CCl<sub>4</sub></i>	<i>Carbon tetrachloride</i>
<i>Ccn<sub>2</sub></i>	<i>cyclin A2</i>
<i>Ccnb<sub>1</sub></i>	<i>cyclin B1</i>
<i>Ccnb<sub>2</sub></i>	<i>cyclin B2</i>
<i>CCND<sub>1</sub></i>	<i>Cyclin D1</i>
<i>Cd63</i>	<i>CD63 antigen</i>
<i>cdh<sub>1</sub></i>	<i>E-cadherin</i>
<i>Cdh<sub>2</sub></i>	<i>N-cadherin</i>
<i>Cdk<sub>1</sub></i>	<i>cyclin-dependent kinase 1</i>
<i>Cdkn<sub>3</sub></i>	<i>cyclin-dependent kinase inhibitor 3</i>
<i>cDNA</i>	<i>circular DNA</i>
<i>Cenpf</i>	<i>centromere protein F</i>
<i>cox<sub>2</sub></i>	<i>Cytochrome c oxidase 2</i>
<i>Cpe</i>	<i>carboxypeptidase E</i>
<i>DEN</i>	<i>N-nitrosodiethylamine</i>
<i>DLC<sub>1</sub></i>	<i>deleted in liver cancer 1</i>
<i>DMSO</i>	<i>Dimethyl sulfoxide</i>
<i>DNA</i>	<i>Deoxyribonucleic acid</i>
<i>Dnmt<sub>1</sub></i>	<i>DNA methyltransferase (cytosine-5) 1</i>

## ABBREVIATIONS DIRECTORY

---

<i>dNTPs</i>	<i>Nucleoside tri phosphates</i>
<i>EDTA</i>	<i>Ethylendiaminetetraacetate</i>
<i>Egr1</i>	<i>early growth response 1</i>
<i>Egr2</i>	<i>early growth response 2</i>
<i>Elk-1</i>	<i>Ets like gene 1</i>
<i>EMT</i>	<i>Epithelial-to-mesenchymal transition</i>
<i>Eno2</i>	<i>enolase 2, gamma neuronal</i>
<i>ER</i>	<i>Estrogen Receptor</i>
<i>FBPs</i>	<i>Far upstream element binding proteins</i>
<i>fgf19</i>	<i>fibroblast growth factor 19</i>
<i>Fhl3</i>	<i>four and a half LIM domains 3</i>
<i>flex1</i>	<i>floxed exon 1</i>
<i>Fos</i>	<i>FBJ osteosarcoma oncogene</i>
<i>FOXM1</i>	<i>Forkhead box M1B</i>
<i>FSP</i>	<i>fibroblast specific protein</i>
<i>Gldn</i>	<i>gliomedin</i>
<i>Gls2</i>	<i>glutaminase 2 (liver, mitochondrial)</i>
<i>H19</i>	<i>NA</i>
<i>HBV</i>	<i>Hepatitis B virus</i>
<i>HCC</i>	<i>Hepatocellular carcinoma</i>
<i>HCl</i>	<i>Hydrochloric acid</i>
<i>HCV</i>	<i>Hepatitis C virus</i>
<i>HIF1</i>	<i>Hypoxia-inducible factor 1</i>

---

<i>Hk1</i>	<i>hexokinase 1</i>
<i>HOP</i>	<i>Homeobox Protein</i>
<i>IARC</i>	<i>International Agency on Cancer</i>
<i>ICC</i>	<i>Intrahepatic cholangiocarcinoma</i>
<i>IEGs</i>	<i>Immediate early genes</i>
<i>Ier3</i>	<i>immediate early response 3</i>
<i>IFN <math>\gamma</math></i>	<i>Interferon <math>\gamma</math></i>
<i>IFN<math>\alpha</math></i>	<i>interferon <math>\alpha</math></i>
<i>Igf2</i>	<i>insulin-like growth factor 2</i>
<i>IMAC</i>	<i>immobilized metal ion chromatography</i>
<i>Inhbe</i>	<i>inhibin beta E</i>
<i>IR</i>	<i>insulin resistance</i>
<i>Itga6</i>	<i>integrin alpha 6</i>
<i>LC</i>	<i>Liquid Chromatography</i>
<i>Lgr5</i>	<i>leucine rich repeat containing G protein coupled receptor 5</i>
<i>Limk2</i>	<i>LIM motif-containing protein kinase 2</i>
<i>Lipg</i>	<i>lipase, endothelial</i>
<i>LoxP</i>	<i>Locus of X-over P1</i>
<i>Lpl</i>	<i>lipoprotein lipase</i>
<i>LWBW</i>	<i>Liver Weight to Body Weight ratio</i>
<i>MAPK</i>	<i>Mitogen activated protein kinase</i>
<i>MET</i>	<i>Mesenchymal-to-epithelial transition</i>
<i>miRNA</i>	<i>MicroRNA</i>

## ABBREVIATIONS DIRECTORY

---

<i>Mt</i>	<i>Mutant</i>
<i>Mthfd1</i>	<i>methylenetetrahydrofolate dehydrogenase (NADP+ dependent), methenyltetrahydrofolate cyclohydrolase, formyltetrahydrofolate synthase</i>
<i>Mthfd1l</i>	<i>methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 1-like</i>
<i>Mthfd2</i>	<i>methylenetetrahydrofolate dehydrogenase (NAD+ dependent), methenyltetrahydrofolate cyclohydrolase</i>
<i>Mup3</i>	<i>major urinary protein 3</i>
<i>Peg3</i>	<i>paternally expressed 3</i>
<i>Pfkp</i>	<i>phosphofructokinase, platelet</i>
<i>Plk1</i>	<i>polo-like kinase 1</i>
<i>Psat1</i>	<i>phosphoserine aminotransferase 1</i>
<i>Raet1d</i>	<i>retinoic acid early transcript delta</i>
<i>Shmt1</i>	<i>serine hydroxymethyltransferase 1 (soluble)</i>
<i>Shmt2</i>	<i>serine hydroxymethyltransferase 2 (mitochondrial)</i>
<i>Slc15a5</i>	<i>solute carrier family 15, member 5</i>
<i>Slc2a1</i>	<i>solute carrier family 2 (facilitated glucose transporter), member 1</i>
<i>SNA1</i>	<i>Snail</i>
<i>SNA2</i>	<i>Slug</i>
<i>Sox9</i>	<i>SRY (sex determining region Y)-box 9</i>
<i>Suox</i>	<i>sulfite oxidase</i>
<i>Tes</i>	<i>testis derived transcript</i>
<i>TGFβ</i>	<i>transforming growth factor β</i>
<i>Thbs1</i>	<i>thrombospondin 1</i>



*Timp1*    *tissue inhibitor of metalloproteinase 1*

*Tspan8*    *tetraspanin 8*

*Ttk*        *Ttk protein kinase*

*Vil1*        *villin 1*

*Wt*         *Wild type*

## 7 TABLES DIRECTORY

Table #	Chapter	Title	Page #
1-1	Introduction	SRF Knockouts and effects	44
2-1	Materials and Methods	List of samples taken for Microarray analysis	72
2-2	Materials and Methods	Sample Demographics - WES analysis	74-75
2-3	Materials and Methods	List of samples taken for transcriptomic and proteomic analysis	77
2-4	Materials and Methods	List of samples taken for metabolomics analysis	80
2-5	Materials and Methods	Patient information of the Heidelberg cohort consisting of 40 HCC patients.	82-83
3-1	Results	Mutations Status – <i>Cttnb1</i> and <i>Trp53</i>	122
3-2	Results	Sample Demographics - Microarray analysis	125
3-3	Results	The 5 most up or down-regulated genes	127
3-4	Results	Genes dysregulated only in HCCs and not in nodules – Microarray analysis.	127
3-5	Results	Genes up-regulated in nodular tissues – Microarray analysis	128
3-6	Results	Gene up-regulated in <i>Cttnb1</i> mutant ( <i>HCC<sub>B</sub></i> ) tumors – Microarray analysis	129
3-7	Results	Patient Demographics of Heidelberg cohort of human HCC patients	132

3-8	Results	Oncofetal liver genes expression	135
3-9	Results	Summary of clustering analyses	135
3-10	Results	Sample Demographics of samples used for transcriptomic and proteomic analysis.	138
3-11	Results	Comparison of experiments – RNAseq and Microarray	146
3-12	Results	Sample overview – Mainz cohort of human HCC patients	147
3-13	Results	Demographics of samples taken for WES analysis	158
3-14 A	Results	<i>Expression profiles of Ctnnb1 target genes– RNAseq data.</i>	165
3-14 B	Results	<i>Phosphorylation status of <math>\beta</math>-catenin protein</i>	165
3-15	Results	Expression profiles of selected SRF target genes– RNAseq data.	166
3-16	Results	Expression profiles of mitosis-related genes – RNAseq data.	168
3-17	Results	Glycolysis-related genes expression	169
3-18	Results	Expression profiles of 1 C metabolism-related genes – RNAseq data	171
3-19	Results	Sample Demographics - Metabolomics analysis.	172
3-20	Results	Up-regulated and down-regulated metabolites	173

## 8 FIGURES DIRECTORY

Figure #	Chapter	Title	Page #
1-1	Introduction	The top 5 causes of death worldwide in 2012	3
1-2	Introduction	Global incidence map of liver cancers	11
1-3	Introduction	Mortality to Incidence ratio	12
1-4	Introduction	Publications on the Warburg effect	27
1-5	Introduction	Pubmed publications with keywords “miRNA, liver cancer”	29
1-6	Introduction	The CArG box sequence	39
1-7	Introduction	SRF-TCF binding to DNA	40
1-8	Introduction	Activation of SRF	42
1-9	Introduction	SRF-VP16 sequence	46
1-10	Introduction	Objective of study	48
2-1	Materials and Methods	Ear Marking Scheme	52
2-2	Materials and Methods	SRF-VP16 genotyping primers	55
2-3	Materials and Methods	Alfp-CreERT2 genotyping primers	58
2-4	Materials and Methods	Srf-flex1 transgene genotyping primers	60
2-5	Materials and Methods	Multi-omics analysis strategy	73

2-6	Materials and Methods	Transcriptomic and Proteomic Strategy	76
2-7	Materials and Methods	Patient information of the Mainz cohort of human HCC patients	84
3-1	Results	The SRF-VP16 transgene construct	113
3-2	Results	The Alfp-CreER <sup>T2</sup> transgene construct	114
3-3	Results	<i>Srf-flex1</i> allele	115
3-4	Results	Mating scheme of <i>SRF-VP16<sup>iHep</sup></i> mice	116
3-5	Results	Experimental Mice genotypes and their usage for sample collection	117
3-6	Results	Weight gain patterns of individual mice	118
3-7	Results	Comparison of Liver weight to body weight (LWBW) ratios	119
3-8	Results	Age-dependent progression toward HCC formation	120
3-9	Results	HCC Penetrance in <i>SRF-VP16<sup>iHEP</sup></i> mice	121
3-10	Results	<i>Ctnnb1</i> and <i>Trp53</i> mutations	123
3-11	Results	<i>GS</i> and <i>Ctnnb1</i> expression levels	124
3-12	Results	Dysregulated Transcripts – Microarray analysis	126
3-13	Results	Expression profiles of specific gene sets - Microarray analysis	130
3-14	Results	SRF expression levels in Heidelberg cohort of human HCC patients.	133
3-15	Results	Unsupervised hierarchical clustering of mouse and human HCCs.	134

## FIGURES DIRECTORY

3-16	Results	Transcript expression levels vs genomic methylation patterns – <i>IGF2</i>	136
3-16a	Results	Genomic methylation patterns of <i>IGF2</i> and <i>H19</i> genes in <i>SRF-VP16<sup>iHep</sup></i> tumours	137
3-17	Results	SRF-VP16 transgene expression - RNAseq	139
3-18	Results	PCA plot of transcriptomics data for all samples.	140
3-19	Results	# Dysregulated transcripts in each sample group	141
3-20	Results	Dysregulated Transcripts – RNAseq analysis	142
3-21	Results	Exclusively Dysregulated Transcripts – RNAseq analysis	143
3-22	Results	Comparison of experimental strategies – RNAseq and Microarray 1	144
3-23	Results	Comparison of experiments – RNAseq and Microarray 2	145
3-24	Results	Clustering analysis: <i>SRF-VP16<sup>iHep</sup></i> vs Mainz cohort	147
3-25	Results	Clustering analysis: <i>SRF-VP16<sup>iHep</sup></i> vs TCGA	149
3-26	Results	PCA plot – Proteomics	149
3-27	Results	Total Deregulated Proteins	150
3-28	Results	Venn diagram representation of deregulated Proteins	151
3-29	Results	Uniquely Deregulated Proteins	152
3-30	Results	Comparison of omics data – RNAseq and Proteomics I	152

3-31	Results	Comparison of omics data – RNAseq and Proteomics II	153
3-32	Results	PCA plot – Phospho-proteomics	154
3-33	Results	Total number of differentially phosphorylated peptides	155
3-34	Results	Total number of differentially modified serine and threonine phosphosites	156
3-35	Results	Differentially phosphorylated peptides	156
3-36	Results	Total single nucleotide variations detected	159
3-37	Results	Single nucleotide variations detected in exonic regions	160
3-38	Results	Insertions/Deletions	161
3-39	Results	Insertions/deletions detected in Exonic regions	162
3-40	Results	Copy number alterations observed in WES analysis	163
3-41	Results	Heterogeneity among tumors	164
3-42	Results	Gene set enrichment analysis (GSEA) of proteomics and RNAseq data	170

## 9 RELEVANT TRAINING

---


- ❖ EMBL Advanced Course: Analysis and Integration of Transcriptome and Proteome Data, 9-13 February 2015, EMBL Advanced Training Centre, Heidelberg, Germany
- ❖ Course: Journal Club – “Current topics in proteome research”, Winter semester 2013-14. Proteome Center Tübingen (PCT).
- ❖ Versuchstierkundlichen Privatissimum, Summer semester 2015.




# 10 POSTER PRESENTATIONS

## 1. Virtual Liver Network retreat, Poster presentation

EBERHARD KARLS  
UNIVERSITÄT  
TÜBINGEN



Faculty of Science  
Interfaculty Institute of Cell Biology



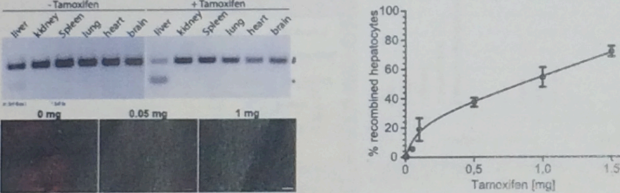
### Dysregulated SRF in hepatocytes induces hyperproliferative nodules, cell depolarization and altered liver architecture *in vivo*.

Ohrnberger, S.<sup>1</sup>, Von Recklinghausen, I.<sup>2</sup>, Hammad, S.<sup>2</sup>, Thavamani, A.<sup>1</sup>, Hengstler, J.<sup>2</sup> and Nordheim, A.<sup>1</sup>

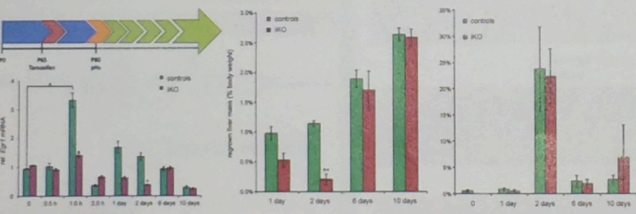
<sup>1</sup> Department of Molecular Biology, University of Tuebingen, Auf der Morgenstelle 15, 72076 Tuebingen  
<sup>2</sup> Leibniz Research Centre for Working Environment and Human Factors (IfAdo), Dortmund, Germany.

**Abstract:**  
 Serum Response Factor (SRF) is a highly conserved transcription factor, which regulates gene expression in cooperation with two alternative sets of partner proteins, the ternary complex factors (TCFs) and the myocardin-related transcription factors (MRTFs). TCFs are targets of MAPK signaling, while MRTFs respond to changes in cytoskeletal actin dynamics. TCF/SRF-dependent transcription regulates immediate early genes (IEGs) involved in cell proliferation and MRTF/SRF-dependent transcription regulates cytoskeletal genes. To analyze the role of SRF in liver architecture we used the Alfp-CreERT<sup>2</sup> mouse line that we generated in collaboration with the group of Prof. Schütz (DKFZ, Heidelberg), enabling the genetic disruption of the *Srf* gene specifically in hepatocytes in a time controlled manner. To further test the relevance under pathological conditions, we performed partial hepatectomy and analyzed the progression of regeneration. In addition, our laboratory generated a new mouse line permitting the controlled expression in hepatocytes of a constitutively active variant of SRF, the SRF-VP16 fusion protein. In this mouse line, individual hyperproliferative liver nodules were induced, in which distinct gene expression alterations were observed. The normal liver micro-architecture was compromised in these nodules and hepatocellular polarity was lost. Later, hepatocellular carcinomas (HCC) developed.

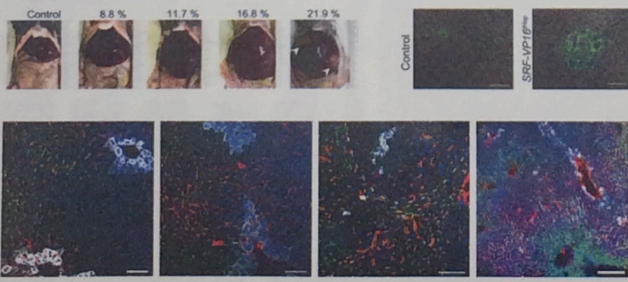
**Characterisation of the Alfp-CreERT<sup>2</sup> mouse line**

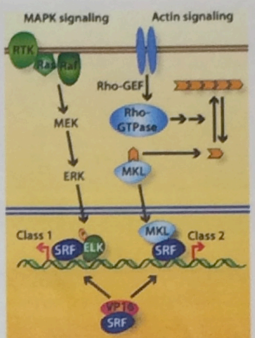


**Liver regeneration after pHx in hepatocyte-specific SRF deficient mice**



**SRF-VP16**





**Conclusion**  
 The Alfp-CreERT<sup>2</sup> mouse line enables time defined hepatocyte specific recombination. The proportion of recombined cells correlated directly with the amount of applied tamoxifen. Without induction a small population of 0.4% recombines until the age of 10 weeks.

**Conclusion**  
 Hepatocyte specific SRF depletion results in altered IEG induction after pHx. The regeneration is moderately delayed at 2 days after pHx. Overall the difference in the regeneration process is close to the control mice. Therefore modelling was not further addressed.

**Conclusion**  
 Stochastic induction of the constitutively active fusion protein SRF-VP16 results in massive alterations in the liver. Due to uncontrolled proliferation, pre-malignant nodules are formed, that in late stages transform into HCC. Liver architecture is greatly compromised. This model system enables modelling of changes in liver architecture in situations where individual cells start to hyper-proliferate.

© Universität Tübingen - Faculty of Science  
 Interfaculty Institute of Cell Biology  
 Auf der Morgenstelle 15 72076 Tübingen Germany  
 Telefon +49 7071 29-78893 - Telefax +49 7071 29-5359 <http://www.molbio.uni-tuebingen.de>

# 2. International Max Planck Research School symposium, Poster presentation



INTERNATIONAL MAX PLANCK RESEARCH SCHOOL

"From Molecules to Organisms"



MAX-PLANCK-GESELLSCHAFT

## Murine Hepatocellular Carcinoma (HCC) triggered by dysregulated Serum Response Factor (SRF)

Abhishek Thavamani, Stefan Ohrnberger, Alfred Nordheim\*

Interfaculty Institute for Cell Biology

\*alfred.nordheim@uni-tuebingen.de

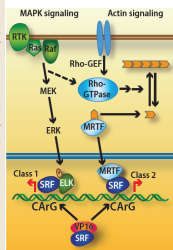
### Abstract

Hepatocellular carcinoma (HCC) is one of the most lethal types of human carcinoma. One of the remarkable features of HCC progression is the accumulation of numerous genomic alterations, owing to various mutagenic insults leading to HCC (e.g. viral infection, liver cirrhosis and chemical exposure). However, the initiating driver events that instigate malignant transformation and the subsequent tumor progression driver events are not well-defined as yet. Here, I propose to track the initiation and progression of HCC using the *SRF-VP16<sup>Hep</sup>* mouse line, which displays controlled expression of a constitutive active version of the transcription factor Serum Response Factor (SRF-VP16). SRF-VP16 expression in murine hepatocytes causes HCC formation. This mouse cancer model allows the identification of genomic alterations associated with oncogenic transformation and progression to malignant liver cancer.

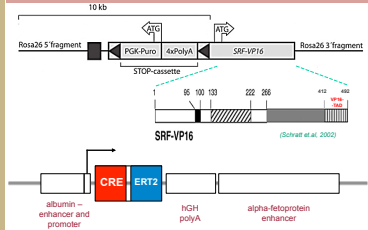
### Introduction

#### Serum Response Factor (SRF)

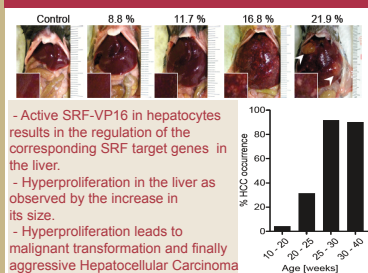
SRF is a transcription factor, which is activated by different stimuli including two major signalling pathways – the MAPK pathway and the Rho-dependant actin dynamics pathway. SRF subsequently binds with other partner proteins such as Ternary Complex factors (TCFs) or the Myocardin-related transcription factors (MRTFs) and regulates cytoskeletal genes and also the immediate early genes (IEGs).



#### SRF-VP16<sup>Hep</sup> mouse line



### Previous Work

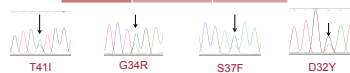


- Active SRF-VP16 in hepatocytes results in the regulation of the corresponding SRF target genes in the liver.
- Hyperproliferation in the liver as observed by the increase in its size.
- Hyperproliferation leads to malignant transformation and finally aggressive Hepatocellular Carcinoma

### Results

#### Mutation Analysis

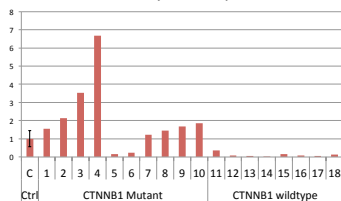
Gene	CTNNB1		TP53	
	Mt	Wt	Mt	Wt
Status	10	8	0	18
Samples	10	8	0	18
Total	18		18	



- CTNNB1 gene mutations detected in 55% of tumours upon sequencing by Sanger's method
- No mutations detected in TP53 gene
- All detected mutations – comparable to human HCC

#### Expression Analysis

##### Glutamine Synthetase Expression



- Glutamine Synthetase gene specifically up-regulated in CTNNB1 mutant HCCs

#### Affymetrix Array

Category	Control	Nodular	HCC A (CTNNB1 wt)	HCC B (CTNNB1 mt)
Number of Samples	3	3	3	3

#### Mouse Gene 1.0 ST Array

Relative Profile	Nodular vs Control	HCC A vs Control	HCC B vs Control
No. of genes		3719	

Min. Fold change > 1.5

Ctrl	Nodular	HCC A	HCC B	Relative Profile	Number of Genes	Total No. of Genes
0	↑			N up	644	1632
0	↓			N down	388	
0		↑		HCC A up	1400	2539
0		↓		HCC A down	1139	
0			↑	HCC B up	1458	2454
0			↓	HCC B down	1196	
0		↑	↑	HCC A and B up	314	
0		↓	↓	HCC A and B down	493	807

- Comparison with human HCC data revealed high degree of correlation

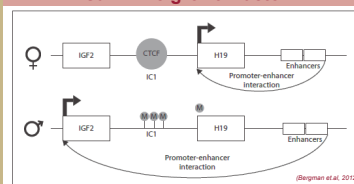
#### Gene Ontology Clustering Analysis: -

Rank	Molecular Function	Cellular Component
1	Protein binding	Membranous
2	Metal-ion binding	Integral to membrane
3	Catalytic activity	Cytoplasm
4	Transferase activity	Nucleus
5	Hydrolase activity	Plasma membrane

#### Top 5 genes highly up-regulated in the HCCs

Gene Symbol	Fold-Change (Nodular vs. Control)	Fold-Change (HCC_A vs. Control)	Fold-Change (HCC_B vs. Control)
Igf2	13.1	106.2	104.2
Bex1	4.1	70.4	83.8
H19	35.4	72.7	82.7
Tspan8	3.1	62.8	18.6
Klrb1a	11.0	52.6	19.5

#### Insulin-like growth factor 2

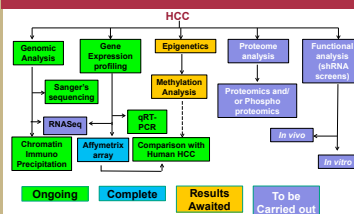


- Both Igf2 and H19 are highly up-regulated in our dataset and also have SRF-binding sites close to their promoters.
- Both genes are regulated inversely to each other due to the presence of an imprinting loci in between. Methylation at this site switches gene regulation.
- Igf2 and H19 have a canonical and non-canonical SRF-binding sequence near to their promoters respectively.

#### Implications

- Activations of SRF or other downstream transcription factors could be the initiation driver event in tumors could explain drug resistance and tumor recurrence too.
- Mutations and other genomic alterations could be progressive driver events acquired later on to assist in proliferation.
- Considering the presence of SRF binding sequence in genes like Igf2, Ins2, and Igf2als, SRF could have implications in diabetes and a wide range of other diseases as well.
- Results could infer a much greater role of metabolism in cancer development and progression.
- The *SRF-VP16<sup>Hep</sup>* mouse line could serve as a platform to study the mechanism behind the various characteristics of tumours

### Future Plans



### Reference

- Phillips, T.H., L., *Transcription factors and transcriptional control in eukaryotic cells*. Nature Education, 2008, 1(1).
- Lee, T.J. and R.A. Young, *Transcription of eukaryotic protein-coding genes*. *Annu Rev Genet*, 2000, 34: p. 77-137.
- Miano, J.M., *Role of serum response factor in the pathogenesis of disease*. *Lab Invest*, 2010, 90(9): p. 1274-84.
- Powers, G. and R. Treisman, *Actin together: serum response factor, its cofactors and the link to signal transduction*. *Trends Cell Biol*, 2006, 16(11): p. 588-96.
- Ohson, E.N. and A. Nordheim, *Linking actin dynamics and gene transcription to drive cellular motile functions*. *Nat Rev Mol Cell Biol*, 2010, 11(5): p. 353-65.
- Behnen, M., et al., *Immediate-early gene induction by the stressors umbonycin and arsenite in human osteosarcoma cells involves MAPK cascade*.
- Bergman, D., et al., *Insulin-like growth factor 2 in development and disease: a mini-review*. *Gerontology*, 2013, 59(3):240-9.
- Schmitt, C., et al., *Serum response factor is crucial for actin cytoskeletal organization and focal adhesion assembly in embryonic stem cells*. *J Cell Biol*, 2002 Feb 18; 156(4): 737-750.

# 11 CURRICULUM VITAE

---

## Abhishek Thavamani

AG Nordheim, Department of Molecular Biology,

### Education:

- 2012-2016 Ph.D. in Biology, University of Tuebingen, Germany  
Advisor: Prof. Dr. Alfred Nordheim
- 2008-2011 Master (M.Tech ) in Molecular Biology and Human Genetics  
Bharathiar University, Department of Zoology, India  
Advisor: Dr. Rita Mulherkar
- 2005-2008 Bachelor in Biochemistry, Madurai Kamaraj University, India

### Research Experience:

- 2011-2012 Research Co-ordinator, Tata Memorial Hospital, India
- 2010-2011 Master dissertation, Mulherkar Lab, ACTREC, India
- 2009 Summer trainee, Institute of Mathematical Sciences, India

### Publications:

- Ohrnberger, S., **Thavamani, A.**, Braeuning, A., Lipka, D.B., Kirilov, M., Geffers., R, Autenrieth, S.E., Roemer, M., Zell, A., Bonin, M., Schwarz, M., Schuetz, G., Schirmacher, P., Plass, C., Longrich, T., Nordheim, A. (2014) Dysregulated SRF triggers formation of hepatocellular carcinoma. *Hepatology* 61(3): 979-89
- Noronha, V., Prabhash, K., **Thavamani, A.**, Chougule, A., Purandare, N., Joshi, A., Sharma, R., Desai, S., Jambekar, N., Dutt, A., Mulherkar, R. (2013) EGFR mutations in Indian lung cancer patients: clinical correlation and outcome to EGFR targeted therapy. *PLoS ONE* 8(4): e61561.
- Akrap I, **Thavamani A**, Nordheim A., (2016) Vps4A-mediated tumor suppression upon exosome modulation? *Ann Transl Med.* 4(9):180
- Constanze Hermanns, Veronika Hampl, Achim Aigner, Josef Penkava, Natalie Frank, Dietmar E. Martin, Kerstin C. Maier, Margarete Goppelt-Struebe, Kerstin Holzer, Stephan Singer, **Abhishek Thavamani**, Alfred Nordheim, Thomas Gudermann, Susanne Muehlich. (2017) The novel MKL target gene myoferlin modulates expansion and senescence of hepatocellular carcinoma., *Oncogene* 36:3464–3476

## 12 Contribution to Dissertation

The following experiments were performed solely by Abhishek Thavamani. for this dissertation project work.

<b>S.No.</b>	<b>Experiment</b>	<b>Thesis section</b>
1.	Mouse breeding	2.1.2
2.	Mouse ear marking	2.1.3
3.	Genotyping	2.1.4
4.	Mouse sacrifice	2.1.6
5.	DNA, RNA & protein extraction	2.2.1, 2.3.1, & 2.4-1
6.	Mutation Detection	2.2.2
7.	Reverse transcription & cDNA synthesis	2.3.2
8.	Quantitative real time PCR	2.3.3
9.	RNAseq library preparation (supervised by Dr. Christa Lanz, MPI Tuebingen)	2.8.4
10.	MS sample preparation (supervised by Dr. Katarina Matic, PCT Tuebingen)	2.8.5
11.	Data Mining – RNAseq / Microarray	3.3 & 3.5.1
12.	Data Mining – Proteomics & WES analysis	3.6, 3.7 & 3.8
14.	Unsupervised Hierarchical clustering	3.4
15.	Multi-omics comparison	3.5