

# Comprehensive Analysis of Disease Related Nuclear and Mitochondrial Genes in Hepatocellular Carcinoma



Inaugural Dissertation

zur

Erlangung des Doktorgrades

Dr. nat. med.

der Medizinischen Fakultät

und

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

**Wafa Amer**

aus Misurata, Libya

Köln 2016

Berichtersteller/Berichterstellerin: Prof. Dr. Tobias Goeser  
Prof. Dr. Thomas Wunderlich

Tag der letzten mündlichen Prüfung:.....**28.10.2016**.....

# Table of content

<b>Summary .....</b>	<b>IV</b>
<b>Zusammenfassung.....</b>	<b>VI</b>
<b>1 Introduction .....</b>	<b>1</b>
1.1 HCC aetiology and pathophysiology .....	1
1.2 Molecular pathogenesis of HCC.....	2
1.2.1 Altered signalling during HCC development.....	2
1.2.2 Mitochondrial genomic alteration.....	4
1.3 HCC management.....	5
1.3.1 Eliminating chronic hepatitis-related cirrhosis .....	6
1.3.2 HCC surveillance .....	7
1.3.3 HCC histopathology.....	7
1.3.4 Surgical intervention .....	8
1.3.5 Systemic targeted therapy.....	9
<b>2 Aim of study .....</b>	<b>10</b>
<b>3 Materials and Methods .....</b>	<b>11</b>
3.1 Materials .....	11
3.1.1 Reagents and kits.....	11
3.1.2 Primers .....	14
3.1.3 Clinical samples.....	21
3.1.4 Buffers .....	22
3.1.5 Cell lines .....	23
3.1.6 Devices .....	23
3.2 Methods.....	24
3.2.1 Cell culture.....	24
3.2.1.1 Passage of cells.....	24
3.2.1.2 Cryopreservation of cells .....	24
3.2.2 HCC sample collection from cryoconserved liver resectates for genomic HCC analysis .	25
3.2.3 HCC sample collection from cryoconserved liver resectates for mt-genome analysis .....	26
3.2.4 COX\SDH enzyme histochemical staining of dysplastic and HCC nodules.....	26

3.2.5	Macrodissection of HCC nodules .....	26
3.2.6	Genomic DNA extraction .....	27
3.2.7	Analysis of DNA extracts from HCC samples .....	27
3.2.7.1	Quality control of DNA by microfluidic electrophoresis.....	27
3.2.7.2	DNA quantification by fluorescence absorption and qPCR.....	27
3.2.8	Target enrichment of HCC related nuclear gene using multiplex PCR .....	28
3.2.8.1	Design of HCC hotspot loci .....	29
3.2.8.2	Multiplex PCR for HCC target enrichment.....	30
3.2.8.3	Library construction of HCC related hotspot gene loci.....	31
3.2.9	Target enrichment of HCC related mt-genome .....	33
3.2.9.1	Primer design for entire mitochondrial genome.....	33
3.2.9.2	Singleplex PCR for mt-target enrichment.....	33
3.2.9.3	Library construction of HCC related mt-gene loci.....	35
3.2.10	Quality assessment and quantification of the constructed libraries .....	37
3.2.11	Illumina sequencing.....	38
3.2.12	Data analysis and software .....	39
3.2.12.1	Data analysis of HCC related hotspot gene loci.....	39
3.2.12.2	Data analysis of HCC related mt-gene loci .....	40
3.2.13	Variant Confirmation .....	40
<b>4</b>	<b>Results.....</b>	<b>42</b>
4.1	Characterisation of the genomic landscape of HCC .....	42
4.1.1	Primer design of specific genomic HCC related regions.....	43
4.1.2	Mutational analysis of hotspot regions in HCC nodules.....	45
4.1.3	Variant validation by Sanger sequencing .....	51
4.1.4	Significantly mutated genes in HCC .....	51
4.2	Evolution analysis of HCC by ultra-deep sequencing of the mt-genome.....	55
4.2.1	Characterisation of HCC nodules by histoenzymatic stains (COX/SDH).....	55
4.2.2	Establishment of the mt-primer design.....	56
4.2.3	Data analysis revealed an extensive and ultra-deep analysis, targeting the entire mt-genome .....	57
4.2.4	Data filtering and accurate mt-gene annotation .....	58
4.2.5	High frequency of mt-mutation located in D-LOOP and respiratory chain subunits.....	59

---

4.2.6	Accumulation of mt-mutations in non-tumour areas, reflecting the pre-tumour machinery of mtDNA.....	60
4.2.7	Analysis of HCC clonality by mt-genome .....	62
4.3	Establishment of entire mt-genome deep-sequencing for tumour-tracking and clonality analysis of FFPE material .....	63
<b>5</b>	<b>Discussion.....</b>	<b>65</b>
5.1	Massive sequencing of HCC-relevant nuclear gene targets .....	65
5.1.1	β-Catenin and TP53 mutations in HCC .....	66
5.2	Ultra-deep sequencing of the entire mt-genome.....	71
5.2.1	The mt-genome alteration as an early marker of HCC development.....	72
5.2.2	The mt-genome alterations as tools for tumour tracking analysis of HCC .....	75
<b>6</b>	<b>References.....</b>	<b>78</b>
<b>7</b>	<b>Appendix.....</b>	<b>87</b>
7.1	Primer sequences of HCC gene panel.....	87
7.2	Sequencing data.....	102
	<b>Glossary .....</b>	<b>119</b>
	<b>Curriculum Vitae .....</b>	<b>123</b>
	<b>Acknowledgements.....</b>	<b>125</b>
	<b>Erklärung.....</b>	<b>127</b>
	<b>Publications .....</b>	<b>128</b>

## Summary

Hepatocellular carcinoma (HCC) is the third most lethal cancer due to late detection, high recurrence and limited therapeutics. Although genetic alterations have been recently studied by whole genome or whole exome next generation sequencing, a comprehensive analysis of HCC relevant genes on lesions with low and high grade of dedifferentiation is missing. Herein, I aimed to study the nuclear and mitochondrial genomic alterations in order to characterize HCC development and clonality.

First, for target enrichment of HCC relevant genes and mt-genome, gene loci of relevance had to be selected and primer sets designed. Subsequent ultra-deep sequencing revealed that in accordance with previous studies, the  $\beta$ -catenin gene (*CTNNB1*) was shown to be the most frequently mutated oncogene, whereas the *TP53* and *AXIN1* genes were the most frequently mutated tumour suppressor genes. Interestingly, *CTNNB1* mutations were detected in lesions with early as well as advanced HCC stage, confirming its role in hepatocarcinogenesis. In contrast, *TP53* mutations were only detected in nodules with advanced HCC stage. In addition, non-protein damaging mutations in the *AXIN1* gene were detected in higher frequency in non-tumour than in tumour lesions; indicating the loss of heterozygosity (LOH) in some tumour samples. Since *AXIN1* alterations effect as well as  $\beta$ -catenin mutations the WNT signalling, my study supports new therapeutical strategies targeting components of the WNT pathway.

Furthermore, this study addressed the intra-tumour clonal structure by ultra-deep sequencing of the entire mitochondrial (mt) genome for better understanding how HCC originate, develop and progress. Since, the mt-genome is highly susceptible to DNA alterations due to the lack of protective histones and a limited DNA repair system, the mutations of the mtDNA are ideal targets to be used for follow up of tumour progression and analysis of tumour clonality. Therefore, I established a NGS approach for a rapid and sensitive mutation screening analysis of the entire mt-genome as a novel tool for tumour cell tracking.

Beside a high mt-mutation rate in tumour areas, also frequent mt-mutations were observed in peri-tumour area suggesting that mt-genome is susceptible at earliest stage of hepatocarcinogenesis. Furthermore, most HCC nodules of individual sample have identical mt-mutations, indicates the monoclonal HCC origin. Interestingly, the

increasing numbers and frequency of a particular mt-hotspot mutation refer to the progression of the HCC dedifferentiation.

In summary, I established a fast pipeline of mutation analysis for simultaneous testing of a tumour specific, hotspot gene panel covering diagnostic relevant loci of HCC, and could assist in the selection of currently available treatments likely to be most effective for HCC patients. Notably, effective targeting of the Wnt signalling pathway is considered as a potential target for pharmacological therapy that is eagerly awaited.

Additionally, our mt-genome screening based approach representing rapid and sensitive molecular tool and provide novel insights in cancer diagnostics and therapeutic strategies.

## Zusammenfassung

Das hepatozelluläre Karzinom (HCC) ist die dritthäufigste tödliche Krebsart, da es spät erkannt wird und es nur limitiert therapierbar ist. Obwohl in jüngster Vergangenheit genetische Veränderungen im HCC durch neue Sequenzieransätze des gesamten Exoms und Genoms analysiert wurden, fehlen umfangreiche Kenntnisse über tumortreibende Mutationen in HCC in Abhängigkeit der Dedifferenzierung. In der vorliegenden Arbeit sollte daher der Mutationsstatus in der Entwicklung von HCC untersucht und genomische Unterschiede zwischen frühen und fortgeschrittenen Tumoren betrachtet werden.

Zunächst wurden HCC relevante Genloci für die Studie festgelegt und ein entsprechendes Primerset für die Anreicherung und nachfolgende Analyse der Zieldomänen durch Tiefensequenzierung entworfen. Übereinstimmend mit früheren Daten konnte in der vorliegenden Studie dargelegt werden, dass das  $\beta$ -Catenin-Gen (*CTNNB1*) das am häufigsten mutierte Onkogen in HCC ist, während *TP53* und *AXIN1* die meist mutierten Tumorsuppressorgene sind. Hierbei konnte herausgestellt werden, dass *CTNNB1*-Mutationen insbesondere bereits in frühen HCC-Stadien nachgewiesen werden konnten, was die tumortreibende Rolle von  $\beta$ -Catenin in der Hepatokarzinogenese unterstützt. *TP53*-Mutationen wurden dagegen vermehrt in HCC mit hohem Dedifferenzierungsgrad gefunden. Daneben wurden Veränderungen im *AXIN1*-Gen nachgewiesen, die aufgrund der Frequenzunterschiede zwischen peritumoralen und Tumorgeweben in einigen Fällen auch auf einen Verlust der Heterozygotie (LOH) schließen lassen. Da *AXIN1* ebenfalls wie  $\beta$ -Catenin im WNT-Signalweg anzusiedeln ist, unterstützen die vermehrten *AXIN1*-Veränderungen in der Hepatokarzinogenese, dass neue Therapieansätze besonders auf den Wnt-Signalweg zielen sollten.

Darüber hinaus adressierte diese Studie die intratumorale klonale Struktur des HCC. Aufgrund der fehlenden Histone und einem begrenzten Reparatursystem ist das mitochondriale (mt) Genom sehr anfällig für DNA-schädigende Agenzien und daraus hervorgehende Mutationen. Das mt-Mutationsmuster kann daher genutzt werden, um die Tumorevolution zu verfolgen und die Klonalität zu beurteilen. Hierfür wurde die Anreicherung und Ultra-Tiefensequenzierung ganzer mitochondrialer (mt) Genome als neuer Ansatz der Tumorbeurteilung etabliert.

## 1 Introduction

Hepatocellular carcinoma (HCC) is the most common malignancy of the liver and represents the fifth most common cancer worldwide. HCC ranks as one of the most critical global health problems due to its increasing incidence and mortality rate over the years. In 2000, HCC was estimated to be the third most common cause of cancer deaths, with 500,000 deaths worldwide each year [1, 2]. Due to its increasing incidence, it was estimated in 2012 to be the second most important cause of cancer deaths, and caused approximately 750,000 deaths annually [3]. The incidence of HCC has striking global variations; it is particularly high in developing countries (East Asia and sub-Saharan Africa), and lower in developed regions (North America and most of Europe), though here the incidence is increasing [4, 5]. These regional differences mainly depend on the complex aetiology of HCC and associated risk factors. The most prominent factors associated with HCC are chronic viral hepatitis B (HBV) and hepatitis C (HCV), alcohol abuse, aflatoxin-B1 consumption, and virtually all cirrhosis-inducing conditions. These factors drive HCC incidence rates in different regions [6]. In addition to regional tendencies of HCC incidence, HCC is notably more prevalent in males, with an average ratio of between 2:1 and 4:1 male patients to female patients [5].

### 1.1 HCC aetiology and pathophysiology

HCC is a heterogeneous disease and is associated with various risk factors. 80%-90% of all HCC cases are associated with liver cirrhosis, which represents the major single risk factor. Chronic infections of HBV and HCV are also main risk factors [5, 7]. HBV infections are responsible for 50%-80% of HCC cases worldwide, whereas 10%-25% of cases are thought to be a result of HCV infection. The mechanisms by which HBV contributes to HCC carcinogenesis are unclear. Chromosomal integration of HBV-DNA in a critical location within the host genome can promote cellular proliferation and an increased risk of HCC [8]. In addition, the viral expression of the transcriptional activator, the HBV Xgene (HBx), is suggested to promote an important cellular pathway that is implicated in hepatocarcinogenesis as the RAS-RAF-MAPK kinase pathway [9, 10], while the carcinogenic role of the HCV infection in HCC is thought to be linked to the viral

transcriptional activator NS5A [11], and also to the viral core protein [12]. Furthermore, the oncogenic role of aflatoxin B1 is linked to somatic mutation in the *TP53* tumour suppressor gene [13].

HBV-related HCC and non-alcoholic liver fatty liver disease (NAFLD) can occur in the absence of cirrhosis, while almost all cases of HCV-related HCC occur with cirrhosis [14], but mostly HCC is based on chronic inflammation and cirrhosis. Cirrhosis is characterised by an increase in fibrous tissue and a destruction of liver cells, which stimulates cellular proliferation and in turn leads to the development of precancerous nodules.

Besides of viral infections, also non-viral risk factors occur, which include chronic alcohol abuse, aflatoxin intake, diabetes, obesity, or certain hereditary conditions such as hemochromatosis [15].

Both diabetes and obesity are implicated in the development of non-alcoholic steatohepatitis (NASH), which can progress to cirrhosis and its related complications, including HCC [16]. Hereditary hemochromatosis, mostly due to homozygous HFE-C282Y mutations, is associated with an increased risk for HCC as a result of increased iron stores stimulating the carcinogenesis [17].

## **1.2 Molecular pathogenesis of HCC**

### **1.2.1 Altered signalling during HCC development**

Hepatocarcinogenesis is a multi-step process that is mostly linked to chronic liver damage as described above. Accumulated genetic and epigenetic alterations result in the activation of tumour enhancer genes (proto-oncogenes) and their mitogenic signalling pathways, as well as the inactivation of tumour suppressor genes, leading to autonomous cellular proliferation. Main pathways contributing to HCC development include the Wnt/ $\beta$ -catenin, insulin growth factor (IGF), PI3K/AKT/mTOR, and MAPK pathways and growth factor-regulated angiogenic signalling.

The Wnt/ $\beta$ -catenin pathway is suggested to be one of the main drivers of hepatocarcinogenesis [18].  $\beta$ -catenin is a structural protein that acts as a regulator in the cadherin-mediated cell-cell adhesion system and in the Wingless/Wnt signal transduction pathway [19, 20]. In HCC,  $\beta$ -catenin accumulation mainly results from

the mutation of the phosphorylation domain targeted by glycogen synthase kinase 3b (GSK3b) and subsequent inhibition of ubiquitin-proteasome degradation. Furthermore, mutations in the genes whose products activate  $\beta$ -catenin signalling, such as *APC* and *AXIN1*, lead to  $\beta$ -catenin accumulation [18, 21-23].

Accumulated  $\beta$ -catenin translocates to the nucleus and forms a complex with T cell-factor (TCF)\lymphoid enhancer-factor (LEF), activating transcription activity of target genes such as cyclin D1, c-Jun, and c-Myc [21, 24]. The expression of these potential  $\beta$ -catenin targets play a critical role in regulating cellular growth, differentiation, and apoptosis in liver cancer [25, 26].

Another important pathway that is involved in HCC is the IGF signalling. This pathway is reactivated in HCC and mainly occurs at the level of IGF-II expression. Activation of the IGF-signalling pathway is triggered by receptor phosphorylation and ultimately leads to the activation of MAPK, PI3K\AKT\mTOR pathways [27]. The IGF-signalling pathway regulates several cellular processes, including proliferation, motility, and inhibition of apoptosis [28].

The PI3K\AKT\mTOR pathway plays a significant role in HCC progression by promoting neoangiogenesis [29]. Activated AKT phosphorylates several cytoplasmic proteins, such as mTOR and BCL-2 associated death promoter (BAD), leading to increased cellular proliferation, decreased apoptosis, and increased cell survival [30]. This pathway is negatively regulated by phosphatase and tensin homolog (PTEN), which dephosphorylate the lipid products of PI3K [31]. Therefore, any anomalies in *PTEN* function result in over activation of the PI3K\AKT\mTOR pathway in HCC.

The ERK\MAPK pathway is a ubiquitous signal transduction pathway that regulates crucial cellular processes, including proliferation, differentiation, angiogenesis, and survival [32]. Activation of this pathway usually occurs by oncogenic mutations within the *RAS* gene: up to 30% of *N-RAS* mutation has been reported in HCC [33]. Furthermore, *B-RAF* activation resulting from dysregulated overexpression of growth factors and their receptors has been demonstrated [32].

A number of angiogenic growth factors, including members of the vascular endothelial growth factor (VEGF) family, have been detected to be upregulated at the gene expression and plasma protein levels in HCC patients [34]. These growth

factors induce the angiogenic signalling through the activation of previous above signalling pathways [31].

Furthermore, epigenetic changes are suspected to occur early in the pre-neoplastic stage, leading to quantitative alteration of gene expression in the absence of detected structural changes of genes or chromosomes [35]. Elevated expression of DNA methyltransferases has been detected in higher levels in HCC than it has in chronic hepatitis and cirrhosis [36]. In addition, microsatellite instability has been observed in some chronic hepatitis, cirrhosis, and HCC [36, 37]. Other structural genetic alterations, such as amplification, deletion, and mutation, occur in some genes and chromosomal loci during the early pre-neoplastic stage, and increase markedly in dysplastic lesions and HCCs [38]. Tumour suppressor genes, such as *TP53* and *RB*, are often located in frequently deleted chromosomal regions, and are inactivated by deletion or mutation during HCC. Likewise, some of the areas of chromosome regional gain contain oncogenes, like the *MYC* gene, which is overexpressed in most HCCs [39].

### **1.2.2 Mitochondrial genomic alteration**

The mitochondria are the main intracellular generator of ATP, which has a crucial role in the regulation of cellular function, metabolism, and apoptosis [40, 41]

Human mitochondrial DNA (mtDNA) is a small circular, self-replicating chromosome of approximately 16.6 Kb, which encodes 13 essential subunits of oxidative phosphorylation (OXPHOS) complexes (complexes I, III, IV, and V), as well as two rRNA and 22 tRNA genes [42]. Due to the high rate of Reactive Oxygen Species (ROS) generation as well as a lack of protective histones and inefficient DNA repair activities, mtDNA are predisposed to a high mutational rate in comparison to nuclear DNA (nDNA) [43]. Mutated mtDNA copies may confer a selective growth advantage and are likely to survive through selection during cellular development. Mutated mtDNA may expand to all intracellular mtDNA copies (homoplasmy) or only to a proportion of them (heteroplasmy) [44, 45].

In 1924, Warburg hypothesised that cancer cells cause a defect in mitochondria and impair aerobic respiration, leading to glycolytic metabolism regardless of the availability of oxygen; this metabolism is referred to as 'aerobic glycolysis' or 'the

Warburg effect' [46]. Several published studies have addressed the role of mtDNA mutation in cancer initiation and metastasis, including for liver cancer [47], breast cancer [48], prostate cancer [49], and other solid tumours [50-52].

HCC is usually preceded by chronic inflammation, repeated destruction, and regeneration of liver tissue, leading to increased oxidative stress and ROS generation. ROS overproduction results in mtDNA mutations and subsequent mitochondrial respiratory defect [53, 54]. A study by Nishikawa et al. found that accumulated mtDNA mutations were markedly increased in cancerous and noncancerous tissue of HCC patients [55]. In addition, the frequency of mtDNA mutations correlated with the degree of tumour dedifferentiation [55]. In HCC, decreased the mtDNA copy number was shown to significantly correlate with large tumour size, liver cirrhosis, and poor five-year survival [56]. Furthermore, mtDNA-deplete hepatoma cells displayed chemoresistance and cell migration mediated by nuclear-dependent pathways [57-59]. These findings support the contribution of mtDNA mutation and mitochondrial dysfunction to malignant progression of HCC.

Due to the resulting high mutation rate, combined with the presence of hundreds of mtDNA copies within each cell and the homoplasmic nature of somatic mtDNA mutations [44, 45], the mitochondrial genome (mt-genome) is an ideal target to use for tumour cell tracking [47-49, 52, 60-62]. Since clonal expansion and the dynamic evolution of cancer are areas of unmet need for development of precise anti-tumour therapy [63], the mt-genome plays a major role in examining the evolutionary history of human cancer.

### **1.3 HCC management**

The Barcelona Clinic Liver Cancer (BCLC) staging classification summarizes the criteria for HCC management. The BCLC classification includes five stages: 0, A, B, C, and D. These stages are determined according to tumour factors (size, number, vascular invasion, lymph node, and extrahepatic metastasis), liver function (Child-Pugh score), and patient health status [64]. Briefly, patients at stage 0, the very early stage, have a single tumour nodule which is less than <2cm in diameter, and are optimal candidates for a hepatic resection. Patients at stage A, the early stage, have a single tumour nodule >2cm in diameter or three tumour nodules <3cm in diameter, and are candidates for radical therapies (resection,

liver transplantation, or percutaneous treatments). Patients at stage B, the intermediate stage, have multinodular tumours and may benefit from chemoembolisation. Patients at stage C, the advanced stage, have lymph node and extrahepatic metastasis, and may receive targeted therapy (sorafenib). Patients at stage D, the end stage, have tumour-related disability, and may receive symptomatic and palliative treatment [64].

HCC is a therapy-refractory tumour. Early diagnosis and transplantation will result in the best treatment options, but transplantations are mostly not possible due to donor shortage. Furthermore, despite successful hepatic resection, the remnant pre-neoplastic cirrhotic liver frequently develops new HCC lesions. Sorafenib is a palliative therapy, and is the only systemic targeted treatment available for HCC patient who can no longer be treated with potentially more effective therapies. Therefore, there is an urgent need to develop an efficient radical management strategy for HCC, including eliminating the HCC risk factors, enhancing early screening strategies, and developing tolerable molecularly targeted agents.

### **1.3.1 Eliminating chronic hepatitis-related cirrhosis**

Antiviral therapy significantly decreases the risk of HCC in patients infected with viral hepatitis. HBV vaccination programs have also been started in many countries and dramatically reduce the rates of mother-to-infant transmission of HBV, thereby reducing the rate of HCC among children [65]. Furthermore, antiviral therapy against HBV has a beneficial role in decreasing the serum level of HBV-DNA, and reducing the risk of HCC development among patients with chronic HBV [8].

Previous advances have been made in managing chronic HCV infections with effective protease inhibitors. In the past few years, interferon and ribavirin were the only two drugs approved by the Food and Drug Administration (FDA) for HCV treatment. Recently, however, the FDA granted a new anti-HCV drugs called direct-acting antivirals (DAAs) as a gold standard for HCV infection treatment [66]. The DAAs are once-daily pills, used for 12 weeks with a cure rate greater than 90% [66, 67]. These new, effective, and well tolerated drugs induce a dramatic improvement in clinical outcome and reduce the mortality rate of HCV-related HCC [67].

In other efforts to decrease the risk of HCC, health education programs may have a prime impact on diabetes, obesity, and NASH, which are becoming increasingly important HCC risk factor in the United States and Europe.

### **1.3.2 HCC surveillance**

The determination of serum alpha-fetoprotein (AFP) for defining HCC risk lacks adequate sensitivity and specificity for efficient surveillance [68]. Ultrasound examination is an optimal tool for HCC surveillance and highly recommended to be performed in an interval of six months. The HCC diagnosis should be based on imaging techniques and biopsy. The application of dynamic imaging criteria should be applied only to patients with cirrhosis of any aetiology and to patients with chronic HBV who may not have fully developed cirrhosis or who have regressed cirrhosis. Interpretation of biopsies and distinction between high-grade dysplastic nodules (H-DN) and HCC is challenging. Further diagnostic techniques are reinforced by staining for glypican 3, heat shock protein 70, and glutamine synthetase, because positivity for two of these three stains confirms HCC [69].

### **1.3.3 HCC histopathology**

Advances in imaging techniques and development of surveillance protocols for those patients at high-risk of HCC development, result in early detection of small hepatic nodules. However, the histopathology of these nodules has a wide range of diagnostic entities, some benign and some with malignant potential, that needs to distinguish for clinical management.

Macroscopically HCC forms soft mass of nodular or invasive pattern, with a heterogeneous macroscopic appearance, and foci of haemorrhage or necrosis. HCC could be a single or multiple nodules ranging from less than <1 cm to over >30 cm in size. Microscopically, HCC shows different histological patterns: the trabecular pattern, the acinar or pseudoglandular pattern, and the solid or compact pattern. Cytologically, tumoural hepatocytes are polygonal, displaying an eosinophilic granular cytoplasm, rounded nuclei and prominent nucleoli.

In 1995, an International Working Party (IWP) of the World Congresses of Gastroenterology classified the hepatic nodules that found in chronic liver disease

into large regenerative nodule, low grade dysplastic nodule (L-DN), high-grade dysplastic nodule (H-DN), and HCC [70]. The IWP defined a small HCC as a tumour measuring less than <2 cm. Small HCC have been classified into early HCC that is well differentiated, and progressed HCC that is moderately differentiated with evidence of microvascular invasion [71]. The diagnostic challenge was to differentiate the H-DNs from early HCC. Detection of stromal invasion is a hallmark for diagnosis of well-differentiated HCC [72]. Furthermore, appearance of nodule-in-nodule lesion suggests the presence of HCC.

Edmondson and Steiner system divided HCC into four grades based on the histological differentiation [73]. In grade I (G1), the tumour cells are small and arranged in thin trabeculae. In grade II (G2), cells are larger with abnormal nuclei and glandular structures may be present. In grade III (G3), nuclei are large and hyperchromatic, with an eosiphilic granular cytoplasm. In grade IV (G4), tumor cells are much less differentiated with hyperchromatic nuclei and loss of trabecular pattern. In fact, most of HCC present as G2 or G3. Due to heterogeneity of cell differentiation inside a tumour, a three-scale system including well-, moderately and poorly differentiated HCC was frequently applied [74].

#### **1.3.4 Surgical intervention**

Hepatic resection and transplantation remain the first option and the most effective treatment for early localised HCC. Hepatic resection for small HCC in normal liver is the best treatment of choice. Recently, surgical outcome of hepatic resection have improved due to a better understanding of liver segmental anatomy, and the improvements in surgical techniques. The reported overall five-year survival rate is 30%-60%. However, only 10%-30% of HCCs are amenable to cure by surgical resection at the time of diagnosis [75, 76].

Whereas, liver transplantation (LT) is the most effective therapeutic option in cirrhotic liver and unresectable HCC, because LT removes both the tumour and the underlying diseased liver. Theresults from Mazzaferro et al in Milan, demonstrated that patients with small tumors (i.e. a single nodule <5 cm or two or three nodules each <3 cm without vascular invasion)had a 4-year overall survival rate of 85% and a tumour-free survival rate of 92% [77]. Application of the Milan criteria improves the overall and disease-free survival rates [77].

Due to the shortage of donor livers and the consequent increase in waiting time for a transplantation, hepatic resection prior to liver transplantation is recommended for patients with small HCC and preserved liver function[78, 79].

Various local therapies are widely used in patients with small localised HCC, when surgical treatment is no longer suitable due to poor general condition or because of compromised liver function [80]. These modalities include i) local ablative therapy: radiofrequency ablation (RFA), microwave coagulation therapy (MCT), and percutaneous ethanol injection (PEI); and ii) trans-arterial techniques: trans-arterial embolisation (TAE), trans-arterial chemotherapy (TAC), trans-arterial chemoembolisation (TACE), and trans-arterial radioembolisation (TARE). These techniques result in a local area of coagulative necrosis within and around the tumour.

### **1.3.5 Systemic targeted therapy**

Sorafenib is a multi-kinase inhibitor considered to be a first-line treatment in patients with unresectable HCC. The antitumor activity of sorafenib in HCC is mediated by blocking the cell proliferation through targeting the RAF\MEK\ERK signalling pathway, and inhibits angiogenesis by targeting tyrosine kinases (TKs), VEGFR-2, VEGFR-3, and PDGFR of endothelial cells [81, 82]. Sorafenib is the only targeted therapy approved by the FDA for clinical use, and shows significant improvement in overall survival and doubles the time to progression among patients with advanced HCC [83].

## **2 Aim of study**

HCC is one of the most lethal cancers, worldwide. In order to better understand the genomic alterations characterising HCC nodules of different differentiation grades, a collective of well differentiated and poorly differentiated HCC nodules should be studied by ultra-deep sequencing of HCC-related genes. First a primer set had to be designed and applied to the enrichment of HCC relevant targets. Mutants should then be identified by ultra-deep sequencing of DNA from defined and well characterised macrodissected tumour areas. Furthermore, this study addressed the tumour history and clonality by analysis of the mtDNA mutation pattern. Therefore, the analysis of the entire mt-genome by ultra-deep sequencing had to be established. Next, the novel technology should be applied to HCC lesions in comparison to matching non-tumour tissues.

### 3 Materials and Methods

#### 3.1 Materials

##### 3.1.1 Reagents and kits

The names of all reagents and kits used in the experiments and their manufacturing companies are listed below (Table 1-8). All reagents for NGS are listed in the table 2 to 6.

**Table 1: Reagents for COX\SDH enzyme histochemical staining**

Reagent	Company	Catalog Number
<b>3,3'-diaminobenzidine tetrahydrochloride (DAB)</b>	Vector laboratories, Burlingame, USA	SK-4105
<b>Cytochrome C</b>	Sigma-Aldrich, Heidelberg, GER	C2506 CAS: 9007-43-6
<b>Bovine catalase</b>	Sigma-Aldrich, Heidelberg, GER	C1345CAS: 9001-05-2
<b>Nitroblue tetrazolium (NBT)</b>	Sigma-Aldrich, Heidelberg, GER	N6876 CAS: 298-83-9
<b>Sodium succinate</b>	Sigma-Aldrich, Heidelberg, GER	S2378 CAS: 6106-21-4
<b>Phenazine methosulfate (PMS)</b>	Sigma-Aldrich, Heidelberg, GER	P9625 CAS: 299-11-6
<b>Sodium azide</b>	Merck, Darmstadt , GER	1.06688.0100

**Table 2: Reagents and kits for quantification of DNA templates used for target enrichment**

Reagent	Reagent/Kit	Company
<b>Fluorescence Absorbance</b>		
<b>QuantiFluor</b>	QuantiFluor™ dsDNA System	Promega, Heidelberg, GER
<b>qPCR</b>		
<b>Master mix</b>	GoTaq® qPCR Master Mix	Promega, , Heidelberg , GER
<b>Primer</b>	<i>HFE</i> primer forward 5'ATG GAT GCC AAG GAG TTC GAA CC	Eurofins Genomics, Eberstadt, GER
	<i>HFE</i> primer reverse 5'GCC ATA ATT ACC TCC TCA GGC AC	

**Table 3: Reagents and kits used for NGS library construction of HCC related nuclear gene loci**

<b>Step</b>	<b>Reagent/Kit</b>	<b>Company</b>
<b>Multiplex PCR for target enrichment</b>	2x Ion AmpliSeq™ Primer Pool Panel 1 IAD40623 (pool1: 153 primer pairs, pool 2: 147 primer pairs; 100nM each)	Life Technologies, Darmstadt, GER
	2x Ion AmpliSeq™ Primer Pool Panel 2 IAD41957 (pool1: 49 primer pairs, pool 2: 45 primer pairs; 400nM each)	Life Technologies, Darmstadt, GER
	5x Ion AmpliSeq™ HiFi Master Mix (Ion AmpliSeq™ Library Kit 2.0)	Life Technologies, Darmstadt, GER
<b>3' Adenylation</b>	NEXTflex™ Adenylation Mix	Bioo Scientific, Austin, USA
<b>Primer Digestion</b>	FuPa reagent (Ion AmpliSeq™ Library Kit 2.0)	Life Technologies, Darmstadt, GER
<b>Purification and Size Selection</b>	Agencourt® AMPure® XP	Beckman Coulter, Krefeld, GER
<b>Adapter Ligation</b>	NEXTflex™ DNA barcodes – 48	Bioo Scientific, Austin, USA
	Switch Solution (Ion AmpliSeq™ Library Kit 2.0)	Life Technologies, Darmstadt, GER
	DNA Ligase (Ion AmpliSeq™ Library Kit 2.0)	Life Technologies, Darmstadt, GER
<b>Adapter-Specific PCR</b>	Platinum® PCR Super Mix (Ion AmpliSeq™ Library Kit 2.0)	Life Technologies, Darmstadt, GER
	NEXTflex™ Primer Mix (NEXTflex™ DNA Barcode Kit)	Bioo Scientific, Austin, USA

Table 4: Reagents and kits used for NGS library construction of HCC related mt-gene loci

Step	Reagent/Kit	Company
PCR amplification for mt-target enrichment	<b>Singleplex PCR:</b> PCR GoTaq® Colorless Master Mix	Promega, Heidelberg , GER
	<b>Multiplex PCR:</b> GeneRead DNaseq Panel PCR Puffer, 5x (Mix-n-Match Panel V2, Qiagen) GeneRead HotStar Taq DNA Polymerase (6U/μl) (Mix-n-Match Panel V2, Qiagen)	Qiagen, Hilden, GER
End Repair	End-Repair Buffer, 10x (Core Kit, Qiagen)	Qiagen, Hilden, GER
	End-Repair Enzyme Mix (Core Kit, Qiagen)	
3' Adenylation	A-Addition Buffer, 10x (Core Kit, Qiagen)	Qiagen, Hilden, GER
	Klenow Fragment (3' → 5' exo-) (Core Kit, Qiagen)	
Purification and Size Selection	Agencourt® AMPure® XP	Beckman Coulter, Krefeld, GER
Adapter Ligation	NEXTflex™ DNA barcodes – 48	Bioo Scientific, Austin, USA
	Ligation Buffer, 2x (Core Kit, Qiagen)	Qiagen, Hilden, GER
	T4 DNA Ligase (Core Kit, Qiagen)	
Adapter-Specific PCR	HiFi PCR Master Mix, 2x (Amp Kit, Qiagen)	Qiagen, Hilden, GER
	Primer Mix (10μM each) (Amp Kit, Qiagen, Hilden, GER)	

Table 5: Reagents and kits used for quality assessment and quantification of the generated libraries

Step	Reagent/Kit	Company
<b>Quality Assessment</b>		
Microfluidic Electrophoresis	High Sensitivity DNA Kit	Agilent Technologies, Heidelberg, GER
<b>Quantification</b>		
qPCR of Adapter-Carrying Amplicons	Illumina Adapter Primer Forward 5'AAT GAT ACG GCG ACC ACC GAG ATC TAC AC	Eurofins Genomics, Eberstadt, GER

Step	Reagent/Kit	Company
qPCR of Adapter-Carrying Amplicons	Illumina Adapter Primer Reverse	Eurofins Genomics, Eberstadt, GER
	5'CAA GCA GAA GAC GGC ATA CGA GAT	
qPCR of Adapter-Carrying Amplicons	GoTaq® qPCR Master Mix	Promega, Heidelberg, GER
	PhiX Control v3	Illumina, San Diego, USA
Fluorescence Absorbance	QuantiFluor™ dsDNA System	Promega, Heidelberg, GER

Table 6: Reagents and kits used for NGS on the MiSeq platform

Step	Reagent/Kit	Company
Next Generation Sequencing	MiSeq Reagent Kit v2 (300 cycles)	Illumina, San Diego, USA
	PhiX Control v3	Illumina, San Diego, USA

Table 7: Reagents and kits used for Sanger validation

Reagent\Kit	Company
RedTaq®-DNA polymerase	Sigma-Aldrich, Taufkirchen, GER
BigDye terminator sequencing kit v.3.1	Applied Biosystems, Darmstadt, GER

Table 8: Reagents used for cell culture work

Reagent	Company
Dulbecco's Modified Eagle Medium	Sigma-Aldrich, Taufkirchen, GER
FCS	Sigma-Aldrich, Taufkirchen, GER
Trypan blue solution (0.5 %)	Biochrom AG, Berlin, GER
Trypsin	GibcoBRL, Karlsruhe, GER

### 3.1.2 Primers

All the oligonucleotides used for mt-genome amplification, and for validation were ordered from Eurofins MWG Operon (Ebersberg, GER).

**Table 9: 108-primer sets for whole mitochondrial genome amplification.** Primer name\*: Previously published primers.

Amplicon	Primer Name	Primer Sequence	Targeted Region	References
1	F1	GATCACAGGTCTATCACCT	1-162	
	R162	GATAAATAATAGGATGAGGCAGG		
2	F155*	TATTTATCGCACCTACGTTTC	155-285	[84]
	R285*	GTTATGATGTCTGTGTGGAA		[84]
3	F265	TTCCACACAGACATCATAAC	265-401	
	R401	TAAAATTTGAAATCTGGTTAGGC		
4	F402*	ATCTTTTGGCGGTATGCACTTT	402-599	[85]
	R599*	TTGAGGAGGTAAGCTACATA		[84]
5	F567	ACCCCCACAGTTTATGTAGC	567-757	
	R757	ATGCTTGTTCCTTTTGATCGTGG		
6	F741	ATCAAAAGGAACAAGCATCAAGCACG	741-878	
	R878*	CCAACCCTGGGGTTAGTATAGC		[85]
7	F866	ACCCAGGGTTGGTCAATTTTC	866-979	
	R979	TCAGGTGAGTTTTAGCTTTATTGGG		
8	F980	TCACCTGAGTTGTAAAAAATC	980-1136	
	R1136*	GGCGAGCAGTTTTGTTGATT		[85]
9	F1135*	CCAGAACACTACGAGCCACA	1335-1332	[85]
	R1332	TTGACCTAACGTCTTTACGTGG		
10	F1320*	GACGTTAGGTCAAGGTGTAGCCC	1320-1510	[85]
	R1510	AGTATACTTGAGGAGGGTGACGG		
11	F1510	TCAAAGGACATTTAACTAAAACCCC	1510-1676	
	R1676	TTTAGCTCAGAGCGGTCAAG		
12	F1657*	CTTGACCGCTCTGAGCTAAAC	1657-1769	[86, 87]
	R1769*	GCCAGGTTTCAATTTCTATCG		[86, 87]
13	F1766	TGGCGCAATAGATATAGTACCG	1766-1924	
	R1924*	AGGTAGCTCGTCTGGTTTCG		[85]
14	F1921	ACCTAAGAACAGCTAAAAGAGCAC	1921-2115	
	R2115	AGCTGTTCCCTTTGGACTAAC		
15	F2105*	GAGGAACAGCTCTTTGGACAC	2105-2216	[86, 87]
	R2216*	TGTTGAGCTTGAACGCTTTCTT		[85]
16	F2213	AACACCCACTACCTAAAAAATCC	2213-2338	
	R2338	TTATGCCGAGGAGAATGTTTTTC		
17	F2333*	GCATAAGCCTGCGTCAGAT	2333-2439	[85]
	R2439*	ATGCCTGTGTTGGGTTGAC		[85]
18	F2432	ACAGGCATGCTCATAAGGA AAGG	2432-2630	
	R2630	ATACAGGTCCCTATTTAAGGAACAAG		
19	F2625*	CTGTATGAATGGCTCCACGAG	2625-2818	[85]
	R2818*	GCCCCAACCGAAATTTTAAAT		[85]
20	F2803	AAATTTTCGGTTGGGGCGACC	2803-2940	

Amplicon	Primer Name	Primer Sequence	Targeted Region	References
	R2940	TGTTATCCCTAGGGTAACTTGTTTC		
21	F2932*	GGGATAACAGCGCAATCCT AT	2932-3109	[85]
	R3109	AANGTAGATAGAAAACCGACCTG		
22	F3108	TTCAAATTCCTCCCTGTACGAAA GG	3108-3259	
	R3259	TATGCGATTACCGGGCTCTGC		
23	F3241*	AGAGCCCGGT AATCGCATAA	3241-3417	[85]
	R3417*	GGGGCCTTTGCGTAGTTGTA		[85]
24	F3424	AACGTTGTAGGCC CCTACGG	3424-3521	
	R3521	ATGTAGAGGGTGATGGTAGATGTG		
25	F3518	ACATCACCGCCCC GACCTTAG	3518-3632	
	R3632*	GAGGTGGCTAGAATAAATAGGAGGC		[85]
26	F3633	TAGCCTAGCCGTTTACTCAATC	3633-3825	
	R3825*	CAGAGGTGTTCTTGTGTTGTGAT		[85]
27	F3823	TGATTACTCCTGCCATCATGACC	3823-3987	
	R3987	TTCGGCTATGAAGAATAGGGCG		
28	F3964	TCGCCCTATTCTTCATAGCCG	3964-4162	
	R4162*	TGAGTTGGTCGTAGCGGAATC		[86, 87]
29	F4163	TACACCTCCTATGAAAAACTTCC	4163-4334	
	R4334	TCCTAGAAATAAGGGGGTTTAAG		
30	F4335	CTATGAGAATCGAACCCATCCCT	4335-4479	
	R4479*	GGGGATTAATTAGTACGGGAAGG		[85]
31	F4459	TTCCCGTACTAATTAATCCCCTG	4459-4608	
	R4608	TTTTGGTTAGAACTGGAATAAAAGCTAG		
32	F4609*	AAATAAACCCCTCGTTCCACAGA	4609-4676	[85]
	R4676*	GATTATGGATGCGGTTGCTT		[85]
33	F4675	TCCTTCTAATAGCTATCCTCTTCAAC	4575-4860	
	R4860	GAAGAAGCAGGCCGGATGT		
34	F4859	TCTCACATGACAAAACTAGCC	4859-5034	
	R5034*	ATCCTATGTGGGTAATTGAGGA		[85]
35	F5033	ATGAATAATAGCAGTTCTACCGTAC	5033-5210	
	R5210*	GGTGGATGGAATTAAGGGTGT		[85]
36	F5189	ACACCCTTAATTCCATCCACC	5189-5325	
	R5325*	TGATGGTGGCTATGATGGTG		[85]
37	F5318*	CACCATCACCTCCTTAACC	5318-5500	[86, 87]
	R5500	AGTATAAAAGGGGAGATAGGTAGG		
38	F5487	TCCCCTTTTATACTAATAATCTTATAG	5487-5681	
	R5681*	GTGGGTTTAAGTCCATTGGT		[85]
39	F5664*	AATGGGACTTAAACCCACAAA	5664-5799	[85]
	R5799*	TGCAAATTCGAAGAAGCAG		[85]
40	F5800	ATTCAATATGA AAATCACCTCGGAGC	5800-5994	
	R5994*	TGCCTAGGACTCCAGCTCAT		[88]
41	F5994	ACAGCTCTAAGCCTCCTTATTCG	5994-6184	
	R6184	AACGCCATATCGGGGCAC		

<b>Amplicon</b>	<b>Primer Name</b>	<b>Primer Sequence</b>	<b>Targeted Region</b>	<b>References</b>
42	F6178	TGGCGTTTCCCCGCATAAAC	6178-6359	
	R6359	TAGGTGTAAGGAGAAGATGGTTAG		
43	F6351	ACACCTAGCAGGTGTCTC	6351-6444	
	R6444*	TTTGGTATTGGGTTATGGCAG		
44	F6443	ACGCCCCCTTCGTCTGATC	6443-6642	
	R6642	TAAGAATATAAACTTCAGGGTGACCG		
45	F6636*	ATTCTTATCCTACCAGGCTTCG	6636-6832	[85]
	R6832	ATGGTAGCGGAGGTGAAATATG		
46	F6831	ATAATCATCGCTATCCCCACCG	6831-6899	
	R6899*	CATATTGCTTCCGTGGAGTGTG		
47	F6899	GAAATGATCTGCTGCAGTGCTC	6899-7076	
	R7076	CCTATGATGGCAAATACAGCTCC		
48	F7075	AGGCTTCATTCACTGATTTCCC	7075-7248	
	R7248*	TGGTGTATGCATCGGGGTAGT		
49	F7248	ACATGAAACATCCTATCATCTGTAGG	7248-7365	
	R7365	GTTCTTCTACTATTAGGACTTTTCGC		
50	F7366*	CCTCCATAAACCTGGAGTGA	7366-7489	[85]
	R7489	TGGCTTGAA ACCAGCTTTG		
51	F7481	TCAAGCCAACCCCATGGCC	7481-7680	
	R7680	AAAATGATTATGAGGGCGTGATCATG		
52	F7679	TTCCTTATCTGCTTCCTAGTCC	7679-7842	
	R7842	ATGTAAAGGATGCGTAGGGATG		
53	F7821*	CATCCCTACGCATCCTTTACAT	7821-7980	[85]
	R7980	TCGCCTGGTCTAGGAATAATG		
54	F7960	ATTATTCCTAGAACCAGGCG	7960-8141	
	R8141*	CGGTGAAAGTGGTTTGGTTTA		
55	F8129*	ACCACTTTCACCGCTACACG	8129-8269	[85]
	R8269	CTATAGGGTAAATACGGGCC		
56	F8256	TATTTACCCTATAGCACCCC	8256-8378	
	R8378*	TTAGTTGGGGCATTTCCTGT		
57	F8377	AAATACTACCGTATGGCCCACC	8377-8566	
	R8566	TTGTGGGGCAATGAATGAAGC		
58	F8563	ACAATCCTAGGCCTACCCG	8563-8640	
	R8640*	GATGAGATATTTGGAGGTGGG		
59	F8641	AACAACCGACTAATCACCACC	8641-8837	
	R8837	ATGGCTAGGTTTATAGATAGTTGG		
60	F8836	ATGGCCATCCCCTTATGAGC	8836-9031	
	R9031*	GGTGGCCTGCAGTAATGTTAG		
61	F9032	TACTCATGCACCTAATTGGAAGCG	9032-9146	
	R9146	GCGACAGCGATTTCTAGGATAG		
62	F9143	TCGCCTTAATCCAAGCCTACG	9143-9275	
	R9275	TGAGAGGGCCCCGTTAG		

<b>Amplicon</b>	<b>Primer Name</b>	<b>Primer Sequence</b>	<b>Targeted Region</b>	<b>References</b>
63	F9272*	CTCAGCCCTCCTAATGACCTC	9272-9376	[85]
	R9376*	CATTGGTATATGGTTAGTGTGTTGG		[85]
64	F9377	ATGGCGCGATGTAACACGAG	9377-9568	
	R9568	ATGCCTGTTGGGGGCCAG		
65	F9568	TCACCCCGCTAAATCCCCTAG	9568-9758	
	R9758	TACTCTGAGGCTTGTAGGAGGG		
66	F9757	TACTTCGAGTCTCCCTTCACC	9757-9940	
	R9940	ACAAAATGCCAGTATCAGGCGG		
67	F9940	TAGATGTGGTTTGACTATTTCTG	9940-10110	
	R10110	TTAGTAGTAAGGCTAGGAGG		
68	F10093	TCCTAGCCTTACTACTAATAATTATTAC	10093-10294	
	R10294*	AGGGCTCATGGTAGGGGTAA		[85]
69	F10294	TACAAACAACCTAACCTGCCACT	10294-10487	
	R10487	TATGTAAATGAGGGGCATTTGG		
70	F10488	AATATTATACTAGCATTACCATCTC	10488-10667	
	R10667	AAGACTAGTATGGCAATAGGC		
71	F10661	TCTTTGCCGCCTGCGAAG	10661-10942	
	R10942*	TAGGGGGTCGGAGGAAAAG		[85]
72	F10943	ACAACCCCCCTCCTAATAC	10943-1110	
	R1110	GATAAGTGTGGTTTCGAAGAAG		
73	F11108	TTATCCCCACCTTGGCTATC	11108-11318	
	R11318	ATAGTTCTTGGGCAGTGAGAG		
74	F11319*	CAAACCTCTGAGCCAACAACCTT	11319-11428	[85]
	R11428*	GGCTTCGACATGGGCTTT		[85]
75	F11424	AAGCCCCCATCGCTGGG	11424-11627	
	R11627	AAGAGTATGCAATGAGCGATTTTAGG		
76	F11626	TCAATCAGCCACATAGCCC	11626-11768	
	R11768*	TGCGTTCGTAGTTTGAGTTTG		[85]
77	F11760*	ACGAACGCACTCACAGTCG	11760-11955	[86, 87]
	R11955	TTGAGTCCTGTAAGTAGGAGAG		
78	F11952	TCAACATACTAGTCACAGCCC	11952-12089	
	R12089*	TGGGGGATAGGTGTATGAACA		[85]
79	F12089	ATTCTCCTCCTATCCCTCAACC	12089-12210	
	R12210	TCGGTAAATAAGGGGTCGTAAGC		
80	F12194*	CCCCTTATTTACCGAGAAAGC	12194-12302	[85]
	R12302*	GCCTAAGACCAATGGATAGCT		[85]
81	F12288	CCATTGGTCTTAGGCCCAA	12288-12470	
	R12470	TAATAAAGGTGGATGCGACAATGG		
82	F12452*	TTGTTCGCATCCACCTTATT	12452-12648	[85]
	R12648	TGAGAATTCTATGATGGACC		
83	F12648	ACTGTGATATATAAACTCAGACC	12648-12766	
	R12766*	AGCCGATGAACAGTTGGAATA		[85]
84	F12741*	CAACCTATTCCAACCTGTTTCATCG	12741-12936	[85]

<b>Amplicon</b>	<b>Primer Name</b>	<b>Primer Sequence</b>	<b>Targeted Region</b>	<b>References</b>
	R12936	TTGTTGTGGGTCTCATGAGTTG		
85	F12936	AATAGCCCTTCTAAACGCTAATC	12936-13025	
	R13025*	TGGAGACCTAATTGGGCTGA		[85]
86	F13025	ACCCCTGACTCCCCTCAG	13025-13215	
	R13215	AAGGGCGCAGACTGCTGC		
87	F13203*	AGTCTGCGCCCTTACACAAA	13203-13390	[85]
	R13390*	TGTTAAGGTTGTGGATGATGGA		[85]
88	F13381	AACCTTAACAATGAACAAGATATTCG	13381-13559	
	R13559*	GCTCAGGCGTTTGTGTATGAT		[85]
89	F13557	AGCCCTATCTATTACTCTCATC	13557-13740	
	R13740	AATGAGAAATCCTGCGAATAGG		
90	F13741	ACTAACAACATTTCCCCCGCATCC	13741-13855	
	R13855*	GGTAGTTGAGGTCTAGGGCTGTT		[85]
91	F13835*	CAGCCCTAGACCTCAACTACC	13835-13924	[85]
	R13924	GGTAGAATCCGAGTATGTTGGA		
92	F13921	ACCCTAGCATCACACACC	13921-14077	
	R14077	TTTGGGTTGAGGTGATGATG		
93	F14058*	CATCATCACCTCAACCCAAA	14058-14118	[85]
	R14118*	TGGGAAGAAGAAAGAGAGGAAG		[86, 87]
94	F14114	TCCCACTCATCCTAACC	14114-14312	
	R14312	TAGGAAGCTGAATAATTTATGAAGG		
95	F14271	TCAACCCTGACCCTCTCC	14271-14448	
	R14448*	GAGGAGTATCCTGAGGCATGG		[85]
96	F14431*	TGCCTCAGGATACTCCTCAAT	14431-14609	[85]
	R14609	AGCCTTCTCCTATTTATGGGG		
97	F14609	TTAGAAGAAAACCCACAAAC	14609-14721	
	R14721*	CGATGGTTTTTTCATATCATTGG		[85]
98	F14715	ACCATCGTTGTATTTCAACTACAAGAA	14715-14902	
	R14902*	GGCTAGGAATAGTCCTGTGGTG		[85]
99	F14903	ATGCACTACTCACCAGACGCC	14903-15083	
	R15083	AGTTTCTGAGTAGAGAAATGATCCG		
100	F15083	TGAAACATCGGCATTATCCTCC	15083-15279	
	R15279	AATCGTGTGAGGGTGGGAC		
101	F15278	TTCTTTACCTTTCACCTCATCTTGC	15278-15396	
	R15396*	TTATCGGAATGGGAGGTGATTC		[85]
102	F15396	AAATCACCTTCCACCCTTAC	15396-15585	
	R15585*	ATTGTGTAGGCGAATAGGAAATA		[85]
103	F15584	ATTCTCCGATCCGTCCTAAC	15584-15728	
	R15728*	GGAGTCAATAAAGTGATTGGCTTAG		[85]
104	F15729	TAGCCGCAGACCTCCTCATTCC	15729-15896	
	R15896	TACAAGGACAGGCCATTTGAG		
105	F15896	AGTATAAACTAATACACCAGTCTTG	15896-16042	

Amplicon	Primer Name	Primer Sequence	Targeted Region	References
	R16042*	CTGCTTCCCATGAAAGAAC		[85]
106	F16043	ATTTGGGTACCACCAAGTATTG	16043-16230	
	R16230	TAGTTGAGGGTTGATTGCTGTAC		
107	F16227	ACTATCACACATCAACTGCAACTC	16227-16410	
	R16410*	GAGGATGGTGGTCAAGGGA		[84]
108	F16409	TCCGTGAAATCAATATCCCGC	16409-16569	
	R16569	CATCGTGATGTCTTATTTAAGGG		

primer name\*: Previously published primers

Table 10: PCR primer of Sanger validation

Oligonucleotide	Species	Sequence
APC-F-E1317Q	Human	GATGTAATCAGACGACACAGGAA
APC-R-E1317Q	Human	AATCCAGCAGACTGCAGGG
ARID1A-1326+1327F	Human	CCTGAGGGAAACATGAGCACT
ARID1A-1326+1327R	Human	TTCCTATGGCAATCAGTTCTCCA
ARID1A-F-G1293fs	Human	GGTCCTTATGACAGAGTGAGGTA
ARID1A-R-G1293fs	Human	ATGCCTTCCAACCCAGACTC
ARID1A-F-G242-A243	Human	CCAACCACCAGTACAACCTCCTA
ARID1A-R-G242-A243	Human	TCGCTCAGCAGCGCTTCG
ARID2-998+999F	Human	CATCACCTGTCCCAGCTACTAA
ARID2-998+999R	Human	TTCAGCAGCCCAACAAGTAC
ARID2-F-Q835*	Human	CAGAGTACAGAACATACCAGCAT
ARID2-R-Q835*	Human	TACTGTTATCATAGCACCCCCAC
AXIN1- F-P278L	Human	CAGGACATGGATGAGGACGAT
AXIN1- R-P278L	Human	AAGGCAGAGAGTTCAGGTGAG
AXIN1-F-G650S	Human	GGATGCGGAGAAGAACCAGAAA
AXIN1-F-T402M	Human	CCGCGTGGAGCCTCAGAA
AXIN1-F-T402M	Human	ATGTCCTGATGTTCTTCATGGGG
AXIN1-R-G650S	Human	TTCAGTGTGTGTGGCGAGTGT
CDKN2A-F-A148T	Human	GAGCTGGGCCATCGCGAT
CDKN2A-R-A148T	Human	TTCCAGCACAGAAAAGTTCAGCC
CTNNB1-F-L753P	Human	GGTATGGACCCCATGATGGAA
CTNNB1-R-L753P	Human	ATCAGCTGGCCTGGTTTGATAC
PTEN-138F	Human	CCAATGGCTAAGTGAAGATGACAA
PTEN-138R	Human	TCCTTTCTCTTCTGGATCTG
PTEN-F-L182V	Human	CTGTCCACCAGGGAGTAACTA
PTEN-R-L182V	Human	AATGTTTCAGTGGCGGAACCTGC
TP53-207F	Human	GGCCTCTGATTCCTCACTGATT
TP53-207R	Human	AACTGGGGTCTCTGGGAGG
TP53-F-R337L	Human	GGACCAGACCAGCTTTCAAAAA
TP53-R-R337L	Human	CAGGCTAGGCTAAGCTATGAT

### 3.1.3 Clinical samples

**Table 11: Clinical features of HCC samples included in this study.\*** Aetiology refers to the frequent risk factors: Hepatitis B (HBV), Hepatitis C (HCV), Alcohol intake(AI.), Hemochromatosis(HM), Hepatic adenoma (HA), Steatohepatitis (SH), Underlying HCC are not defined (n.d).

Sample ID	Age	Gender	Aetiology*	Fibrosis	Diagnosis	Tumour Grade	TNM
H01	49Yrs	M	HCV	F4	HCC	G2-G3	T3NxMx
H02	57Yrs	M	Toxic	F0-F1	HCC	G2-G3	T3NxMx
H03	68Yrs	M	n.d	F0-F1	HCC	G1-G2	T2NxMx
H04	55Yrs	M	HCV	F4	HCC	G2	T3NxMx
H05	68Yrs	M	n.d	F0-F1	HCC	G1+G3-4	T2NxMx
H06	41Yrs	M	HCV	F4	HCC+DN	G2	T3NxMx
H07	58Yrs	M	Toxic	F4	HCC	G2-G3	T2NxMx
H08	50Yrs	M	HA	F4	HCC	G2	T1NxMx
H09	65Yrs	M	HBV	F4	HCC+DN	G2+G3	T4N1Mx
H10	70Yrs	F	n.d	F0-F1	HCC	G3	T3NxMx
H11	76Yrs	M	HCV	F4	HCC	G3	T3NxMx
H12	49Yrs	M	HBV	F4	HCC+DN	G1	T1NxMx
H13	73Yrs	F	HBV	F4	HCC	G3	T3NxMx
H14	53Yrs	F	HM	F0-F1	HCC	G1	T3NxMx
H15	68Yrs	M	n.d	F0-F1	HCC	G3	T3NxMx
H16	69Yrs	M	Toxic	F4	HCC	G2	T3NxMx
H17	65Yrs	M	HBV	F4	HCC	G4	T1NxMx
H18	40Yrs	M	HBV	F4	HCC	G2	T1NxMx
H19	82Yrs	M	n.d	F0-F1	HCC	G1	T1NxMx
H20	58Yrs	M	HBV	F4	HCC	G1-G2	T1NxMx
H21	75Yrs	F	n.d	F0-F1	HCC	G2-G3	T2NxMx
H22	64Yrs	M	HCV	F4	HCC	G2-G3	T3NxMx
H23	73Yrs	F	HM	F0-F1	HCC	G2	T3NxMx
H24	44Yrs	M	HCV	F4	HCC	G2	T1NxMx
H25	57Yrs	F	HBV,HDV	F4	HCC	G1	T1NxMx
H26	67Yrs	M	HCV	F4	HCC	G2	T1NxMx
H27	55Yrs	M	SH	F0-F1	HCC	G2	T1NxMx
H28	72Yrs	M	SH	F0-F1	HCC	G2	T3NxMx
H29	73Yrs	M	SH	F4	HCC	G1	T1NxMx
H30	65Yrs	M	n.d	F0-F1	HCC	G2	T1NxMx
H31	65Yrs	M	n.d	F4	HCC	G2	T1NxMx
H32	69Yrs	M	HCV	F4	HCC	G2	T1NxMx
H33	69Yrs	M	HCV	F4	HCC	G2	T2NxMx
H34	69Yrs	M	HCV	F4	HCC	G2	T2NxMx
H35	67Yrs	M	n.d	F0-F1	HCC	G2	T1NxMx
H36	76Yrs	M	AI	F4	HCC	G2	T1NxMx
H37	74Yrs	F	n.d	F0-F1	HCC	G2	T1NxMx

Sample ID	Age	Gender	Aetiology*	Fibrosis	Diagnosis	Tumour Grade	TNM
H38	74 Yrs	M	n.d	F0-F1	HCC	G2	T1NxMx
H39	53Yrs	M	HBV	F4	HCC	G2	T1NxMx
H40	53Yrs	M	Toxic	F4	HCC	G2	T1NxMx
H41	45Yrs	M	HBV	F2	HCC	G3	T3NxMx
H42	61Yrs	M	n.d	F4	HCC	G2	T3NxMx
H43	66Yrs	M	HCV	F4	HCC	G2	T2NxMx
H44	67Yrs	M	HBV	F0-F1	HCC	G2-G3	T3NxMx

\* Aetiology refers to the frequent risk factors: Hepatitis B (HBV), Hepatitis C (HCV), Alcohol intake(Al.), Hemochromatosis(HM), Hepatic adenoma (HA), Steatohepatitis (SH), Underlying HCC are not defined (n.d).

### 3.1.4 Buffers

All buffers were prepared using completely desalted millipore purified water (Millipore-Q Plus, Millipore, (Molsheim, GER)).

**Table 12: Buffer solutions used in various experiments**

Buffer	Concentrations and Dilutions	Substance/Company
<b>Lysogeny broth (LB):</b>	1.0 % [w/v]	Tryptone (Fluka, GER)
<b>Medium:</b>	0.5 % [w/v]	Bacto-yeast extracts (Difco, USA)
	0.8 % [w/v]	Sodium chloride (NaCl) (adjust pH value to 7.6 with NaOH)
<b>LB agar:</b>	15 g /l	LB medium Bacto-Agar (Difco)
<b>2 x TSS:</b>	20 % [w/v]	Polyethylene glycol (PEG) 8000 (Sigma, GER)
	10 % [v/v]	Di-methyl sulfoxide (DMSO) (Sigma, GER)
	70.0 mM	Magnesium chloride (MgCl <sub>2</sub> ) in LB, pH 6,5
Buffer	Concentrations and Dilutions	Substance/Company
<b>Phosphate-buffered</b>	74 g	NaCl
<b>Saline (PBS) 10x</b>	14.2 g	Di-sodium hydrogen phosphate (Na <sub>2</sub> HPO <sub>4</sub> x H <sub>2</sub> O)
	3.62 g	NaH <sub>2</sub> PO <sub>4</sub> x 2H <sub>2</sub> O
	Add to 1L	Distilled water (dH <sub>2</sub> O) mg, pH 7.4
<b>TAE buffer</b>	40 mM	Tris-acetate, pH 7.8
	5 mM	Sodium acetate (CH <sub>3</sub> COONa)
	1 mM	EDTA, pH 8.0

### 3.1.5 Cell lines

Hepatoma cell lines are used as experimental models for human liver cancer.

**Table 13: Hepatoma cell lines**

Cell line	Origin	Characteristics	Source
<b>HUH7</b>	Human	Hepato cellular carcinoma cell line	U. Protzer
<b>Pop10</b>	Human	Hepato cellular carcinoma cell line	U. Protzer
<b>HepG2</b>	Human	Hepato cellular carcinoma cell line	ATCC
<b>Hep3BG*</b>	Human	Hepato cellular carcinoma cell line	Prof. Jonel Trebicka
<b>HEK-293</b>	Human	Human embryonic kidney cells	ATCC
*Hep3B cells were kindly provided by Prof. Jonel Trebicka, University Hospital Bonn, Germany.			

### 3.1.6 Devices

All devices used in this study are listed below.

**Table 14: Devices that used in the experiments and their manufacturing companies**

Device	Company
Microscope Eclipse TE 300	Nikon, Düsseldorf, GER
MJ Research PTC-200 Peltier Thermal PCR Cycler	GMI, Minnesota, USA
Multiscan Ascent (photometer)	Thermo Scientific, Bonn, GER
Olympus Vanos-S AH2 (Fluorescence microscope)	Olympus, Hamburg, GER
Precellys 24 Homogenisator	Peqlab, Erlangen, GER
BioRad CFX96 Real-time PCR Cycler	Bio-Rad, Munich, GER
Roche Lightcycler 480	Roche, Mannheim, GER
Eppendorf BioPhotometer	Eppendorf, Hamburg, GER
Water bath	Dr. Hirtz & Co, Cologne, GER

## 3.2 Methods

### 3.2.1 Cell culture

The human cell lines were cultured in Dulbecco's modified Eagles medium (DMEM) (Sigma, Heidelberg, GER) supplemented with 10% fetal calf serum (FCS) (Sigma, Heidelberg, GER) and maintained at fully humidified atmosphere at 37°C and 5%CO<sub>2</sub>.

#### 3.2.1.1 Passage of cells

Medium was aspirated from the monolayer of adherent cell lines and were washed with 1 x PBS to remove serum which inhibits trypsin digestion. The cells were treated with trypsin (0.25 % trypsin / 0.02 % (w/v) EDTA in 1 x PBS) and after 10 seconds, most of the trypsin was aspirated. The cells were incubated at 37°C for 1-5 minutes until the adherence was abolished. Trypsin action was stopped by adding fresh medium to the plates. Cell clumps were broken up by vigorous pipetting. The cells were split and further cultivated in new cell culture plates with 10% FCS DMEM.

#### 3.2.1.2 Cryopreservation of cells

Monolayers of 80-90 % confluency were washed twice with 1 x PBS, trypsinized, suspended in fresh medium. Suspended cells were centrifuged at 800 rpm / 4 °C for 5 minutes in a Beckman GPR centrifuge (Beckman, Krefeld, GER). Afterwards, the cell pellet was resuspended in 500 µl fresh 10% FCS supplemented DMEM medium and mixed with DMSO to a final concentration of 10 %. The cells were slowly frozen at -70 °C over 24 h and finally stored in liquid nitrogen.

In order to cultivate cryopreserved cells, frozen cells were rapidly thawed and incubated in a culture dish with 9 ml fresh medium. The medium was changed after 12 hours to remove any residual DMSO.

### 3.2.2 HCC sample collection from cryoconserved liver resectates for genomic HCC analysis

One hundred fifteen resected specimens with HCC or matching non-tumorous areas (kindly provided by Professor Dr. Schirmacher, Institute for Pathology of the University of Heidelberg) were used from 44 patients. In addition, four hepatoma cell lines (Huh7, Hep3p, HepG2, and Pop10) were enrolled in this study.

The characteristics of the patients are shown in Table 15. The mean age was 63 years and most patients were male. Most of tumours showed no evidence of lymph node or distal metastasis, and the majority of tumours showed moderately differentiated histology.

**Table15: Patient characteristic**

	#	%
<b>Median age</b>		
Years	63.11	
Range	40-82	
<b>Gender</b>		
Male	37	84
Female	7	16
<b>TNM</b>		
T1NxMx	20	45
T2NxMx	7	16
T3NxMx	16	36
T4N1Mx	1	2
<b>Differentiation</b>		
Well	7	16
Moderately	23	52
Poorly	14	32
<b>Cirrhosis</b>		
Yes	28	64
No	16	36
<b>Aetiology</b>		
HCV	11	25
HBV	10	23
Toxic	4	9
Other*	19	43
<b>Total</b>	<b>44</b>	
Other*: Alcoholic intake (1), Hepatic adenoma (1), Hemochromatosis (2), the aetiology non determined (n.d) (12), Steatohepatitis (3).		

### **3.2.3 HCC sample collection from cryoconserved liver resectates for mt-genome analysis**

In total 48 HCC nodules, representing different tumor grades, and adjacent noncancerous regions from ten patients out of 44, and three control liver samples from individuals with liver disease other than HCC (including colorectal, kidney and breast cancer with liver metastasis) were investigated.

### **3.2.4 COX\SDH enzyme histochemical staining of dysplastic and HCC nodules**

A dual histochemical assay was applied for investigating mitochondrial dysfunction according to Ross et al. 2011 protocol [89]. Briefly, frozen sections were cut at a thickness of 15µm, and allowed to air dry at room temperature for 1 hour. Cytochrome C staining medium (COX) was prepared according to the protocol and 150-200µl of medium dropped to each slide, and spread onto all section by pipette tip. Then, the slides incubated at 37°C for 1 hour. Next, excess solution was removed and the slides rinsed well with 0.1 M PBS. Afterward, 150-200µl of prepared succinate dehydrogenase (SDH) medium was added to each slide, and spread by pipette tip. Then, the slides incubated for 40 minutes at 37 °C. Afterward, excess solution was removed and slides rinsed with PBS. Next the slides dehydrated in the ascending concentrations of ethanol, and placed in xylene for 10 minutes. Then, the slides mounted with entellan and coverslip, and allowed the slides to dry overnight in a ventilated area (Table 1).

### **3.2.5 Macrodissection of HCC nodules**

COX\SDH histochemical staining allows to clearly visualizing tumor nodules, and easily scraped off the individual tumor nodule. Manual macrodissection was then performed on fresh frozen sections (15µm) using a sterile disposable scalpel blade. Gross tumor nodules were scraped off with a scalpel and collected individually into 1,5ml Eppendorf tube which contains 300µl lysis buffer (Maxwell® 16 LEV Blood DNA Kit, Promega, Heidelberg, GER).

### **3.2.6 Genomic DNA extraction**

Genomic DNA was extracted from samples automatically by the Maxwell® 16 automate using the Maxwell® 16 LEV Blood DNA Kit (Promega, Heidelberg, GER) according to the manufacturer's instructions.

### **3.2.7 Analysis of DNA extracts from HCC samples**

#### **3.2.7.1 Quality control of DNA by microfluidic electrophoresis**

The quality of DNA was determined by parallel capillary electrophoresis using the Fragment Analyzer (Advanced Analytical, Heidelberg, GER) according to the manufacturer's protocol (DNF-474 High Sensitivity NGS Fragment Analysis Kit (1-6,000bp)).

#### **3.2.7.2 DNA quantification by fluorescence absorption and qPCR**

Two different quantification methods were used to detect the amount of input DNA. Genomic DNA was quantified by fluorescence absorbance using the QuantiFluor™ dsDNA System (Promega, Heidelberg, GER) and the QuantiFluor™-ST fluorometer (Promega, Heidelberg, GER) according to the manufacturer's instructions. Alternatively, the extracted genomic DNA was measured quantitatively by qPCR using the human hemochromatosis gene (*HFE*) as amplifying target (234 bp). DNA samples were quantified from genomic DNA standard curves that were prepared from native DNA extracted from human embryonic kidney cells (HEK-293, obtained from the American Type Culture Collection ATCC) (Table 16). Dilutions of HEK-293 cell DNA in a range of 0.195 to 50 ng/μl were used. All samples were measured in duplicates.

**Table 16: Set up of HFE qPCR and amplification conditions**

<b>A) HFE qPCR setup</b>		
<b>Component</b>	<b>Volume [<math>\mu</math>l]</b>	
Nuclease-free water	7.4	
<i>HFE</i> primer forward (10 $\mu$ M) <sup>*1</sup>	0.8	
<i>HFE</i> primer reverse (10 $\mu$ M) <sup>*2</sup>	0.8	
GoTaq® qPCR Master Mix (Promega, Heidelberg, GER)	10.0	
Total	19.0	
gDNA (10 ng)	1.0	
<b>Total</b>	<b>20.0</b>	
*1 5'ATG GAT GCC AAG GAG TTC GAA CC		
*2 5'GCC ATA ATT ACC TCC TCA GGC AC		
<b>B) HFE qPCR amplification conditions</b>		
<b>Stage</b>	<b>Temperature</b>	<b>Time</b>
Hold	94°C	3 Minutes
55 Cycles	94°C	30 Seconds
	60°C	30 Seconds
	72°C	30 Seconds
Plate read and melting curve		

### 3.2.8 Target enrichment of HCC related nuclear gene using multiplex PCR

HCC target panels were designed using the Ion AmpliSeq™ algorithm of Life Technologies, target enrichment and library preparation followed the instructions of the “Ion AmpliSeq™ Library Kit 2.0” (Life Technologies, Darmstadt, GER) and the “NEXTflex™ DNA Sequencing Kit, Manual V11.12” (Bioo Scientific, Austin, USA) are briefly described below.

### 3.2.8.1 Design of HCC hotspot loci

To select hotspot regions of genes for the target specific HCC panel, first different comprehensive databases as COSMIC (catalogue of somatic mutations in cancer) [90], BioGPS [91], UniProt [92] and EMBOSS (The European and Molecular Biology Open Software Suite, 2000) were screened [93]. In addition, most recent published data on HCC was involved [18, 94-96]. Based on the previous inquiry, two panels were designed (Figure 1). A first panel contained 300 primers in two primer pools with targets covering the whole coding regions of the eight genes: *ARID1A*, *ARID2*, *AXIN1*, *CDKN2A*, *CTNNB1*, *PTEN*, *RPS6KA3*, and *TP53*. A second panel covering 94 hotspot regions in two more primer pools with the following 32 genes: *APC*, *ASXL1*, *ATM*, *BRAF*, *CDH1*, *CSF1R*, *EGFR*, *ERBB2*, *EZH2*, *FZD1*, *GNAS*, *HNF1A*, *HRAS*, *IL6ST*, *IRF2*, *JAK2*, *KDM6A*, *KRAS*, *MET*, *MLL3*, *MMP9*, *NF1*, *NF2*, *NFE2L2*, *NOTCH2*, *NOTCH3*, *NAS*, *PDGFRA*, *PIK3CA*, *RB*, *SMAD2*, and *SMAD4*. Both two panels were designed using the Ion AmpliSeq™ algorithm of Life Technologies (Darmstadt, GER). In total, the panels contained primer for 394 target regions and were delivered in four primer pools (Supplemental Table S1 and S2).

HCC Targeted Gene Panel					
APC	ASXL1	ATM	BRAF	CDH1	CSF1R
EGFR	ERBB2	EZH2	FZD1	GNAS	HNF1A
HRAS	IL6ST	IRF2	JAK2	KDM6A	KRAS
MET	MLL3	MMP9	NF1	NF2	NFE2L2
NOTCH2	NOTCH3	NRAS	PDGFRA	PIK3CA	RB
SMAD2					
HCC Complete Gene Panel					
ARID1A	ARID2	AXIN1	CDKN2A	CTNNB1	PTEN
RPS6KA3	TP53				

**Figure 1: Designed HCC gene panels for multiplex target enrichment.** Two HCC gene panels of most relevant HCC genes. HCC targeted gene panel involving the hotspot regions of the gene. And HCC complete gene panel involving the whole coding sequences of the gene.

### 3.2.8.2 Multiplex PCR for HCC target enrichment

A total of 40ng genomic DNA was amplified in four separate multiplex PCR reactions per sample. Table 17 showed the PCR setup and amplification conditions.

**Table 17: Multiplex PCR reaction mix and amplification conditions**

<b>A) PCR setup</b>		
	<b>20µl Approach</b>	<b>10µl Approach</b>
<b>Component</b>	<b>Volume [µl]</b>	<b>Volume [µl]</b>
5x Ion AmpliSeq™ HiFi Master Mix (Ion AmpliSeq™ Kit 2.0, Life Technologies, Darmstadt, GER)	4	2
2x Ion AmpliSeq™ primer pool* (Life Technologies, Darmstadt, GER)	10	5
gDNA, ~10ng	Y	Y
Nuclease-free water	6-Y	3-Y
<b>Total</b>	<b>20</b>	<b>10</b>
*IAD40623 pool1: 153 primer pairs, pool 2: 147 primer pairs; 100nM each		
*IAD41957 pool1: 49 primer pairs, pool 2: 45 primer pair; 400nM each		
<b>B) PCR amplification conditions</b>		
<b>Stage</b>	<b>Temperature</b>	<b>Time</b>
Hold	99°C	2 Minutes
30 Cycles	99°C	15 Seconds
	60°C	4 Minutes
Hold	10°C	∞

Then, two of the four PCR reactions per sample were pooled and subjected to enzymatic digestion of the primer sequences according to the manufacturer's protocol (Life Technologies, Darmstadt, GER) (Table 18).

**Table 18: Protocol for enzymatic digestion of the amplicon primer sequences**

<b>Temperature</b>	<b>Time</b>
50°C	10 Minutes
55°C	10 Minutes
65°C	20 Minutes
10°C	For up to 1 hour

### 3.2.8.3 Library construction of HCC related hotspot gene loci

After multiplex-PCR-based target enrichment and enzymatic digestion, each reaction per sample was pooled and amplification products were purified from half of the reaction volume. All purification and size selection steps were performed in a 96 well format using Agencourt® AMPure® XP magnetic beads and a Biomek® FXp workstation (Beckman Coulter, Krefeld, GER). Each purification and size selection step contained two washing steps with 200 µl 80% ethanol. Purification step I was done with 1.6-fold reaction volume of magnetic beads to remove fragments <100 bp. DNA was eluted with 20 µl nuclease-free water. Subsequently, 10-fold diluted samples were adenylated (Table 19) and ligated to NEXTflex™ DNA barcodes-48 (Bioo Scientific, Austin, USA) (Table 20).

**Table 19: Adenylation reaction mix and conditions**

Component	Volume [µl]
Purified amplicons	17.0
NEXTflex™ Adenylation Mix (Bioo Scientific, Austin, USA)	4.0
<b>Total</b>	<b>21.0</b>
Incubation for 30 minutes at 37°C	

**Table 20: Adapter ligation reaction mix**

A) Components for adapter ligation	Volume [µl]
3' Adenylated DNA	20.0
NEXTflex™ DNA Barcode Adapter 1-48 (Bioo Scientific, Austin, USA) (12.5 µM)*	2.0
Switch Solution (Ion AmpliSeq™ Kit 2.0, Life Technologies, Darmstadt, GER)	4.0
DNA Ligase (Ion AmpliSeq™ Kit 2.0, Life Technologies, Darmstadt, GER)	2.0
Nuclease-free water	2.0
<b>Total</b>	<b>30.0</b>
*5'AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T * 5'GAT CGG AAG AGC ACA CGT CTG AAC TCC AGT CAC XXX XXX ATC TCG TAT GCC GTC TTC TGC TTG	

<b>B) Conditions</b>	
<b>Temperature</b>	<b>Time</b>
22°C	30 Minutes
72°C	10 Minutes
10°C	Hold

Further, purification step II was done with 1.8-fold reaction volume of magnetic beads to remove fragments <100 bp. DNA was eluted with 40 µl nuclease-free water. To enrich amplicons with library sizes between 200 and 400 bp size selection was performed with 0.8-fold and 0.2-fold reaction volume of magnetic beads. The DNA was eluted with 23 µl nuclease-free water. Amplicon adapter ligated libraries were enriched by a final PCR step (Table 21) and purified afterwards.

**Table 21: PCR parameters for the final PCR amplification of adapter ligated DNA libraries;A) PCR components; B) PCR conditions**

<b>A) Final PCR setup</b>		
<b>Component</b>	<b>Volume</b>	
Platinum® PCR SuperMix High Fidelity (Ion AmpliSeq™ Kit 2.0, Life Technologies, Darmstadt, GER)	25.0	
NEXTflex™ Primer Mix (Bioo Scientific, Austin, USA) (12.5µM)*	2.0	
DNA library	23.0	
<b>Total</b>	<b>50.0</b>	
*5'AAT GAT ACG GCG ACC ACC GAG ATC TAC AC *5'CAA GCA GAA GAC GGC ATA CGA GAT		
<b>B) Final PCR amplification parameters</b>		
<b>Stage</b>	<b>Temperature</b>	<b>Time</b>
Hold	98°C	2 Minutes
10 Cycles	98°C	15 Seconds
	60°C	60 Seconds
Hold	10°C	∞

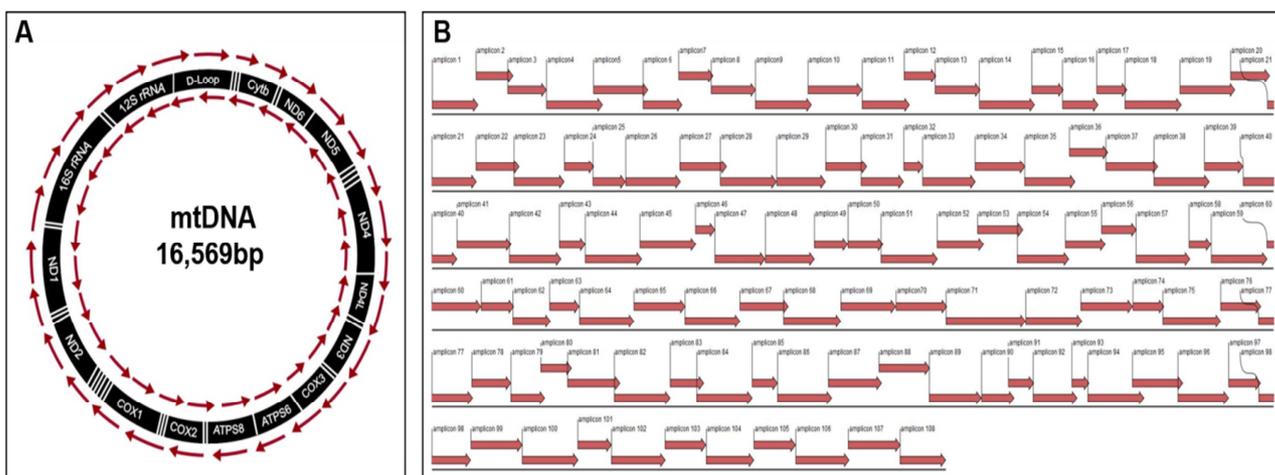
The final purification step III was done with 1-fold reaction volume of magnetic beads and elution of DNA with 20 µl nuclease-free water.

### 3.2.9 Target enrichment of HCC related mt-genome

Target enrichment of whole mt-genome performed by singleplex PCR based approach. One hundred-eight primer pairs were designed manually, and ordered from Eurofins MWG Operon (Eberstadt, GER). Library preparation followed the instructions of the GeneRead DNaseq Panel PCR kit(QIAGEN Inc. Hilden, GER) and the “NEXTflex™ DNA Sequencing Kit, Manual V11.12” (Bioo Scientific, Austin, USA) are briefly described below.

#### 3.2.9.1 Primer design for entire mitochondrial genome

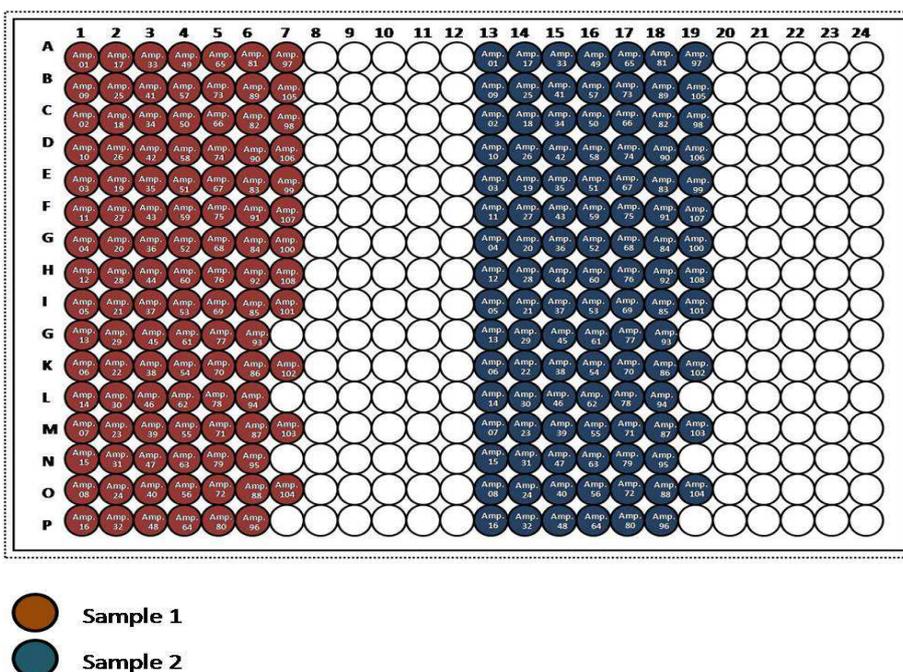
One hundred-eight primer sets spanning the whole mtDNA were designed manually. Each primer pairs are resulting in amplicons of around 60-200 bp in length. (Figure 2, Table 9, Section 3.1.2).



**Figure 2: Scheme of primer design of mt-genome.** Primer sets (Table 9) were designed (A) generating 108 amplicons spanning the whole mitochondrial genome (B).

#### 3.2.9.2 Singleplex PCR for mt-target enrichment

Automatic single PCR set-up was performed for enrichment of whole mitochondrial genome using the Biomek® FXp workstation (Beckman Coulter, Krefeld, GER). A total of 10ng genomic DNA was amplified in each 108 singleplex PCR reactions per sample. PCR reactions were performed in total volume of 6µl (Table 22). The 108 PCRs were conducted in 384-well plate (Figure 3).



**Figure 3: Amplification plate layout.** Simultaneous whole mt-genome amplification of two samples in a 384-well plate.

**Table 22: Singleplex PCR reaction mix and amplification conditions**

A) Singleplex PCR setup		
Component	Volume [ $\mu$ l]	
GoTaq® Colorless Master Mix	3	
Forward primer (10 $\mu$ M)	0,75	
Reverse primer (10 $\mu$ M)	0,75	
Genomic DNA (10ng)	X	
Nuclease-freewater	1,5-X	
<b>Total</b>	6	
B) Singleplex PCR amplification parameters		
Stage	Temperature	Time
Hold	94°C	3 Minutes
30 Cycles	94°C	30 Seconds
	60°C	4 Minutes
Extension	72°C	30 Seconds

Then, 2 $\mu$ l of each 108 PCR reactions per sample were pooled and subjected to the library construction.

### 3.2.9.3 Library construction of HCC related mt-gene loci

After singleplex-PCR-based target enrichment, each reaction per sample was pooled and amplification products were purified. All purification and size selection steps were performed in a 96 well format using Agencourt® AMPure® XP magnetic beads and a Biomek® FXp workstation. Each purification and size selection step contained two washing steps with 200 µl 80% ethanol. Purification step I was done with 1.8-fold reaction volume of magnetic beads. DNA was eluted with 25 µl nuclease-free water. Afterward, the quantity and quality of size selected and purified amplicons were assessed by Quantifluor (Promega, Heidelberg, GER) and Fragment Analyzer (Advanced Analytical Technologies, Heidelberg, GER), subsequently, according to the manufacturer's protocols. Next, 50ng enriched targets of each sample were adenylated (Table 23 and 24) and ligated to NEXTflex™ DNA barcodes-48 (Bioo Scientific, Austin, USA) (Table 25).

**Table 23: End repair reaction mix and conditions**

Component	Volume
DNA (10-200ng)	14.0µl
End-Repair Buffer, 10x (Core Kit, Qiagen, Hilden, GER)	1.7µl
End-Repair Enzyme Mix (Core Kit, Qiagen, Hilden, GER)	1.3µl
<b>Total</b>	<b>17.0µl</b>
Incubation for 30 minutes at 25°C, followed by 20 minutes at 75°C	

**Table 24: Adenylation reaction mix**

Component	Volume
End-repaired DNA	12.5µl
A-Addition Buffer, 10x (Core Kit, Qiagen, Hilden, GER)	1.5µl
Klenow Fragment (3' → 5' exo-) (Core Kit, Qiagen, Hilden, GER)	1.5µl
<b>Total</b>	<b>15.5µl</b>
Incubation for 30 minutes at 37°C, followed by 10 minutes at 75°C	

**Table 25: Adaptor ligation reaction mix**

Component	Volume
Adenylated DNA	15.5µl
Ligation Buffer, 2x (Core Kit, Qiagen, Hilden, GER)	22.5µl
T4 DNA Ligase (Core Kit, Qiagen, Hilden, GER)	2.0µl
DNase-free Water	3.0µl
Adaptor (NEXTflex, 25µM) (Bioo Scientific, Austin, USA)	2.0µl
<b>Total</b>	<b>45.0µl</b>
Incubation for 10 minutes at 25°C, with cold lid	

Further, purification step II was done with 1.8-fold reaction volume of magnetic beads. DNA was eluted with 80 µl nuclease-free water. To enrich amplicons with library sizes between 200 and 400 bp, size selection was performed with 0.8-fold reaction volume of magnetic beads. The DNA was eluted with 24 µl nuclease-free water. Amplicon adapter ligated libraries were enriched by a final PCR step (Table 26) and purified afterwards.

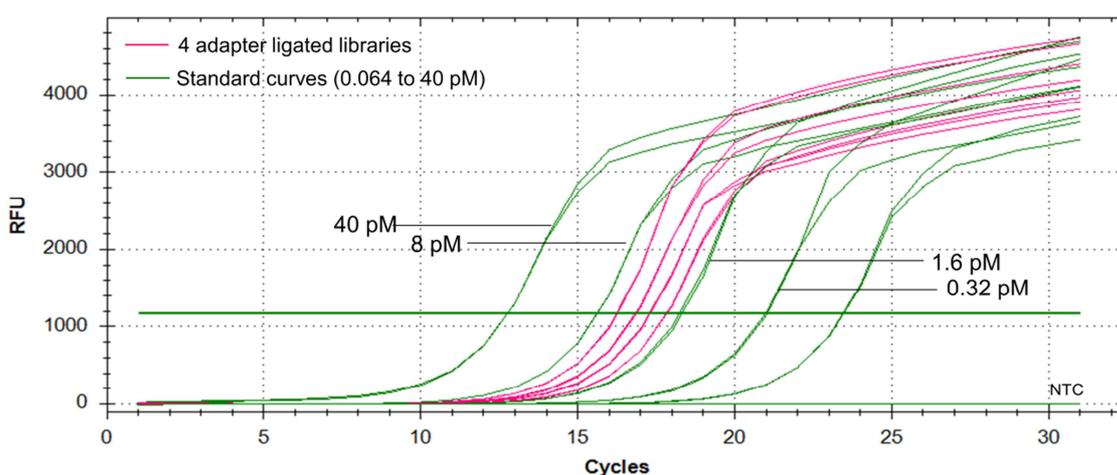
**Table26: PCR parameters for the final PCR amplification of adapter ligated DNA libraries**

<b>A) Final PCR setup</b>		
Component	Volume	
Adapter-ligated DNA	17.0µl	
HiFi PCR Master Mix, 2x (Amp Kit, Qiagen, Hilden, GER)	25.0µl	
Primer Mix (10µM each) (Amp Kit, Qiagen, Hilden, GER)	1.5µl	
RNase-free Water	6.5µl	
<b>Total</b>	<b>50µl</b>	
<b>B) Final PCR amplification parameters</b>		
Stage	Temperature	Time
<b>Denaturation</b>	98°C	2 Minutes
<b>12 Cycles</b>	98°C	20 Seconds
	60°C	30 Seconds
	72°C	30 Seconds
<b>Final extension</b>	72°C	1 Minutes
<b>Hold</b>	10°C	∞

The final purification step III was done with 1.2-fold reaction volume of magnetic beads and elution of DNA with 25  $\mu$ l nuclease-free water.

### 3.2.10 Quality assessment and quantification of the constructed libraries

The quality of the enriched libraries was analysed by capillary electrophoresis using a High Sensitivity NGS Fragment Analysis Kit on Fragment Analyzer (Advanced Analytical, Heidelberg, GER) according to the manufacturer's instructions (Section 3.2.7). Quantitative assessment of amplicon library was determined by qPCR using primers covering the Illumina adapter sequences. Five 5-fold dilutions of PhiX Control V3 (Illumina, San Diego, USA) in a range from 0.064 pM up to 40 pM served as reference standard (Figure 4).



**Figure 4:** Amplification curves of the established standard series (0.064pM-40pM) (green) and four different adapter ligated libraries (pink)

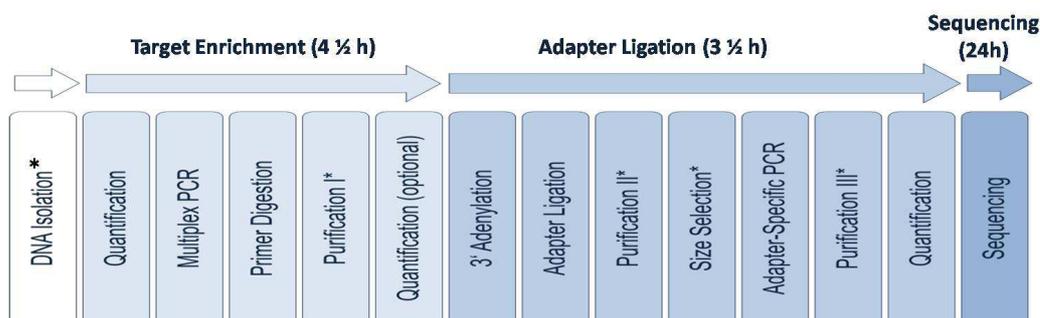
The library samples were diluted 1:4,000 and measured in duplicates (Table 27).

**Table 27: PCR parameters for quantification of the constructed libraries by qPCR; A) PCR components; B) PCR conditions**

<b>A) qPCR setup</b>		
<b>Component</b>	<b>Volume [<math>\mu</math>l]</b>	
Nuclease-free water	7.4	
Forward primer (10 $\mu$ M) <sup>*1</sup>	0.8	
Reverse primer (10 $\mu$ M) <sup>*2</sup>	0.8	
GoTaq® qPCR Master Mix (Promega, Heidelberg, GER)	10.0	
<b>Total</b>	<b>19.0</b>	
DNA library (diluted 1:4,000)	1.0	
<b>Total</b>	<b>20.0</b>	
*1 5'AAT GAT ACG GCG ACC ACC GAG ATC TAC AC		
*2 5'CAA GCA GAA GAC GGC ATA CGA GAT		
<b>B) qPCR amplification parameters</b>		
<b>Stage</b>	<b>Temperature</b>	<b>Time</b>
Hold	94°C	3 Minutes
30 Cycles	94°C	30 Seconds
	60°C	30 Seconds
	72°C	30 Seconds

### 3.2.11 Illumina sequencing

Ultra-deep sequencing of HCC related-nuclear and –mitochondrial genes were carried out by synthesis technology using MiSeq NGS platform. The samples were pooled in an equimolar ratio according to results of the qPCR. 12-15 pM library pools including 1%-2.5% PhiX Control V3 library were prepared for sequencing according to MiSeq System User Guide (Illumina, San Diego, USA). Finally, paired-end sequencing was carried out on a MiSeq instrument (Illumina) using the v2 chemistry and 300 cycle cartridges as recommended by the manufacturer.



**Figure 5:** The flow chart shows an overview of the developed workflow for targeted next generation sequencing of HCC samples; \*automated process.

### 3.2.12 Data analysis and software

#### 3.2.12.1 Data analysis of HCC related hotspot gene loci

Quality metrics analysis of illumina sequencing output can be assessed by Sequencing Analysis Viewer software (SAV) (Illumina, San Diego, USA). FASTQ files generated by the MiSeq Reporter Software were analysed with an in-house developed bioinformatics pipeline based on our general cancer genome analysis tool, which was further optimized for the diagnostic workflow [97]. Before aligning raw reads to the reference genome, adaptor sequences were trimmed. The resulting data was then aligned to the coordinates of the target amplicons according to hg19 reference genome (NCBI build 37) using the Burrows-Wheeler alignment tool [98]. Sequencing errors were estimated from known germline polymorphisms and each sequenced genomic position was analysed for variants that exceed this background rate. Detected variants were finally annotated and filtered to remove germline variants by using the databases dbSNP (single nucleotide polymorphism database; <http://www.ncbi.nlm.nih.gov/SNP/>) and the exome variant server (<http://evs.gs.washington.edu/EVS/>).

Obtained variants with an allelic frequency below 5% and synonymous variants were removed. Additional, visual analysis of called variants was performed by means of the Integrative Genomic Viewer (IGV, Broad Institute, Cambridge, MA, USA). False positive variants, particularly in repetitive or highly homologous regions of the genome, variants in high background noise, as well as single strand

variants, were either eliminated when they were clearly recognizable as artifacts, or were further re-assessed by Sanger sequencing.

Furthermore, variants were analysed for their functional impact on the protein by MutationAssessor (<http://mutationassessor.org>, release 2) [99] and implementation of the ANNOVAR algorithm [100], which combines the bioinformatics tools SIFT (sorting tolerant from intolerant) [101], PolyPhen2 (polymorphism phenotyping 2) [102] and MutationTaster [103].

### **3.2.12.2 Data analysis of HCC related mt-gene loci**

FASTQ files generated by the Illumina platform were analysed by means of the Biomedical Genomics Workbench 2.5.1 (QIAGEN Inc. Hilden, GER) ([www.qiagenbioinformatics.com](http://www.qiagenbioinformatics.com)). In order to determine run performance and off-target reads, the FASTQ sequences were mapped against the whole human reference genome hg19. For variant calling and annotation, the mt-genome (Genebank; accession no. [NC\\_012920](https://www.ncbi.nlm.nih.gov/nuccore/NC_012920)) served as a reference. Using the workflow tool of the Biomedical Genomics Workbench 2.5.1 software in batch mode ensured successive and identical analysis of all samples. The minimum read depth was set to 30, and the minimum variant frequency to 5%. Furthermore, variant calling was restricted to loci with a balanced forward-backward performance (<0.2). Polymorphisms were recognized using the MITOMAP (<http://www.mitomap.org/bin/view.pl/MITO-MAP/HumanMitoSeq>), dbSNP-v138 (<http://www.ncbi.nlm.nih.gov/SNP/id=138>) and HAPMAP\_phase\_3 ([http://hapmap.ncbi.nlm.nih.gov/hapmap3r3\\_B36/](http://hapmap.ncbi.nlm.nih.gov/hapmap3r3_B36/)) databases. Furthermore, sequencing error including sequences in the homopolymer tracts or repeat regions were ruled out, and identified variants were subsequently validated by conventional Sanger sequencing.

### **3.2.13 Variant Confirmation**

A subset of variants, including variants with less than 100 reads, was confirmed by conventional Sanger sequencing using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Life Technologies, Darmstadt, GER) (Table 28). Variants that could not be confirmed were excluded from further analysis.

**Table 28: PCR reaction mix and conditions for Sanger sequencing; A1 and 2) components and conditions for amplification of target regions by PCR I; B1 and 2) components and conditions for Sanger sequencing of target regions by PCR II**

<b>A1) PCR setup</b>			<b>B1) Sanger PCR setup</b>		
<b>Component</b>		<b>Volume [µl]</b>	<b>Component</b>		<b>Volume [µl]</b>
Nuclease-free water		11.5-Y	Nuclease-free water		13.5
Primer forward (10µM)		0.5	Primer forward or reverse (10 µM)		0.5
Primer reverse (10µM)		0.5	5x Big Dye sequencing buffer (Life Technologies, Darmstadt, GER)		4.5
Multiplex PCR Master Mix (Qiagen, Hilden, GER)		12.5	Big Dye® Terminator v3.1 Cycle (Life Technologies, Darmstadt, GER)		0.5
Total		24.0	Total		24.0
gDNA (10ng)		Y	DNA amplicons		1.0
Total		25.0	Total		20.0
<b>A2) PCR amplification parameters</b>			<b>B2) Sanger PCR amplification parameters</b>		
<b>Stage</b>	<b>Temperature</b>	<b>Time</b>	<b>Stage</b>	<b>Temperature</b>	<b>Time</b>
Hold	94°C	15 Minutes	Hold	95°C	1 Minutes
40 Cycles	94°C	30 Seconds	34Cycles	96°C	30 Seconds
	X°C*	90 Seconds		X°C*	60 Seconds
	72°C	60 Seconds		60°C	4 Minutes
Hold	72°C	10 Minutes	Hold	60°C	10 Minutes
* Annealing temperature of specific primer pair					

## 4 Results

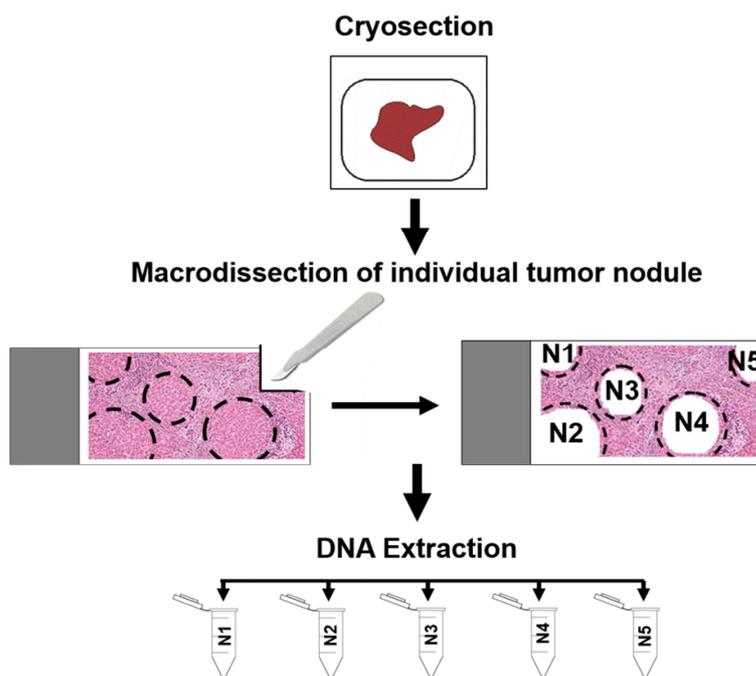
In the present study, the nuclear and mitochondrial genomic mutation profiling of HCC was successfully performed by multiplex and singleplex PCR target enrichment linked to next generation sequencing (NGS). First, nuclear gene panels were defined, targeting either HCC relevant mutation hotspots or complete genes that are often mutated in HCC. The appropriate primer sets for HCC-specific gene enrichment were designed. The designed primer sets were then applied in a deep sequencing study for the identification of mutations of different nodules in HCC tumours.

In order to better understand the complex dynamics of HCC, next I designed a unique method based on ultra-deep sequencing of the entire mt-genome as a novel molecular tool for characterisation of the tumour clonality. Primer sets for the entire mt-genome amplification were designed and used for the singleplex PCR linked to NGS.

Finally, mt-ultra-deep sequencing technology was tested on FFPE archival material. Since DNA of FFPE biopsies is not available in large quantities and it is highly fragmented, here a novel multiplex assay was established for target enrichment, followed by NGS [104].

### 4.1 Characterisation of the genomic landscape of HCC

One of the four serial tissue sections from cryoconserved specimens was used for dual COX/SDH enzyme histochemical staining (see Section 3.2.4). Adjacent non-tumour areas and tumour nodules encompassing the HCC, as defined by histology, were then subjected to macrodissection from the unstained tissue slides to significantly enrich tumour cells from the surrounding microenvironment, as depicted in Figure 6.



**Figure 6:** Schematic representation of tumour cell enrichment by macrodissection of a frozen tissue sample. Every tumour nodule was macrodissected individually and its DNA was extracted into a separate tube.

This material was then applied to comprehensive analysis of HCC relevant genes and the mt-genome.

#### 4.1.1 Primer design of specific genomic HCC related regions

In order to further characterise the genomic landscape, the mutational profile of HCC was assessed using two panels of 40 genes covered by 394 amplicons, which were chosen according to their HCC relevance by searching the literatures as well as COSMIC and other comprehensive databases as described in Material and Methods (see Section 3.2.8.1). Screening gene loci with HCC relevance resulted in the design of two panels (Supplemental Table S1 and S2), which covered the coding sequence of HCC hotspot target regions or the entire genes. Both panels were then used in four primer pools, resulting in a median coverage rate of 96.41% (Table 29).

Table 29: Overview of the genes covered by the HCC panel

Gene Symbol	Gene Description	Exons	Transcript ID	n Amplicons
APC	Adenomatosis Polyposis Coli	14, 16	NM_000038	10
ARID1A	AT Rich Interactive Domain 1A	Complete (20)	NM_006015	75
ARID2	AT Rich Interactive Domain 2	Complete (21)	NM_152641	71
ASXL1	Additional Sex Combs Like Transcriptional Regulator 1	13	NM_015338	2
ATM	ATM Serine/Threonine Kinase	11, 20, 50, 59, 62	NM_000051	10
AXIN1	Axin 1	Complete (11)	NM_003502	31
BRAF	B-Raf Proto-Oncogene	15	NM_004333	2
CDH1	Cadherin 1	4	NM_004360	1
CDKN2A	Cyclin-Dependent Kinase Inhibitor 2A	Complete (3)	NM_058195	11
CSF1R	Colony Stimulating Factor 1 Receptor	7	NM_001288705	1
CTNNB1	Catenin (Cadherin-Associated Protein)	Complete (15)	NM_001904	33
EGFR	Epidermal Growth Factor Receptor	20, 28	NM_005228	6
ERBB2	Erb-b2 Receptor Tyrosine Kinase 2	21	NM_001005862	1
EZH2	Enhancer of Zeste 2 Polycomb Repressive Complex 2 Subunit	17	NM_004456	3
FZD1	Frizzled Class Receptor 1	1	NM_003505	1
GNAS	GNAS Complex Locus	8	NM_080425	2
HNF1A	HNF1 Homeobox A	1, 2, 4	NM_000545	7
HRAS	Harvey Rat Sarcoma Viral Oncogene Homolog	3	NM_176795	1
IL6ST	Interleukin 6 Signal Transducer	6	NM_002184	2
IRF2	Interferon Regulatory Factor 2	5	NM_002199	2
JAK2	Janus Kinase 2	8	NM_004972	2
KDM6A	Lysine (K)-specific Demethylase 6A	11, 29	NM_021140	2
KRAS	Kirsten Rat Sarcoma Viral Oncogene Homolog	2, 3	NM_004985	5
MET	MET Proto-Oncogene	17, 19	NM_000245	2
MLL3	Lysine (K)-specific Methyltransferase 2C	38	NM_170606	3
MMP9	Matrix Metalloproteinase 9	4	NM_004994	1
NF1	Neurofibromin 1	26	NM_001042492	2
NF2	Neurofibromin 2	8	NM_000268	2
NFE2L2	Nuclear factor, Erythroid 2-like 2	2	NM_006164	3
NOTCH2	Notch 2	4	NM_024408	1
NOTCH3	Notch 3	5	NM_000435	1
NRAS	Neuroblastoma RAS Viral (v-ras) Oncogene Homolog	3	NM_002524	2
PDGFRA	Platelet-Derived Growth Factor Receptor	18, 22	NM_006206	3
PIK3CA	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase	3, 10, 21	NM_006218	7
PTEN	Phosphatase and Tensin Homolog	Complete (9)	NM_000314	18
RB	RB1-Inducible Coiled-Coil 1	2, 12	NM_014781	2
RPS6KA3	Ribosomal Protein S6 Kinase	Complete (22)	NM_004586	42
SMAD2	SMAD Family Member 2	10	NM_005901	1
SMAD4	SMAD Family Member 4	9, 1	NM_005359	4
TP53	Tumour Protein p53	Complete (12)	NM_001126114	19
<b>Total number of amplicons</b>				<b>394</b>

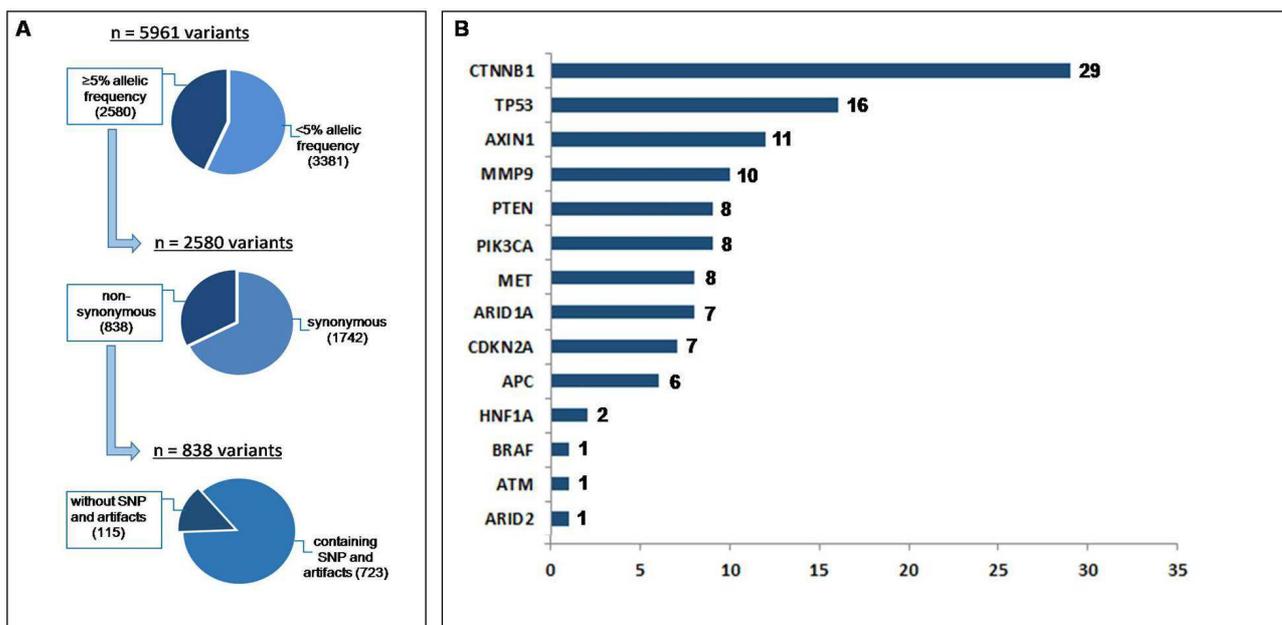
#### 4.1.2 Mutational analysis of hotspot regions in HCC nodules

NGS of 115 samples was conducted in four NGS runs based on a multiplex PCR approach for HCC-specific gene enrichment (see Section 3.2.8). In total, a throughput of  $25.14 \times 10^6$  reads and an output of 7.03 billion bases (Gb) were generated. The bridge amplification of the target regions resulted in a cluster density of 1,334 thousands per  $\text{mm}^2$  (K/ $\text{mm}^2$ ) in each run with 90.48% passing the quality filter parameters of the instrument. From these reads, 96% had a quality score equal to or greater than Q30. A quality score of Q30 means that one base call in 1,000 is predicted to be incorrect (Table 30).

**Table30: Run parameter of the four sequencing runs**

Parameter	Run 1	Run 2	Run 3	Run 4	Average
Entity	HCC	HCC	HCC	HCC	
Samples	28+4 Hepatoma Cell Lines	25	25	37	
Amplicons [n]	394	394	394	394	
Cluster Density [K/ $\text{mm}^2$ ]	1,444	1,261	1,432	1,199	1,334
Parameter	Run 1	Run 2	Run 3	Run 4	Average
Cluster Passed Filter	88.49%	93.48%	91.36	88.60%	90.48%
Reads	27.22M	24.38M	27.50M	22.53M	25.14M
Reads Passed Filter	24.02M	22.79M	25.12M	19,69M	22.91M
Total Yield	7.3Gb	7.0Gb	7.7Gb	6.1Gb	7.03Gb
$\geq$ Q30 Score	95.36%	97.10%	96.20%	95.00%	96.00%
Q30 Score= 0.1% chance of wrong base call; M=million; Gb=gigabases					

The HCC patient samples produced a total of 2,580 variants after raw data alignment and first (automated) background removal. Exclusion of sequencing artefacts, synonymous variants, and polymorphisms led to the determination of 115 significant variants with putatively deleterious-, neutral-protein coding variants, or not yet determined (ND) (91 missense, 11 splice site, 9 indels (7 insertions, 2 deletions), 4 nonsense) in 15 genes (Figure 7, Table 31).



**Figure 7: Filtering the NGS data and annotating the reliable variants from HCC cohorts.** NGS output data were filtered by removing the allelic frequency below 5%, synonymous variants, and reading error, which resulted in 120 final reliable variants (A). The variant count per gene is represented in the bar chart (B).

**Table31: A comprehensive annotation of non-synonymous variants from 44 HCC samples.** FI... functional impact; fs... frame shift; ND... not determined; OT... non tumour; TN... tumour nodule

X1... variants were confirmed by Sanger sequencing.

Pt ID	Gene Name	Position	Variant Type	cDNA Change	Protein Change	Allelic Frequency	FI (MutationAssessor&P ROVEAN based analysis)	
H41-OT	AXIN1	chr16:347063	missense	c.1948G>A	p.G650S <sup>1</sup>	98.90%	Neutral	Tolerated
H10-OT	AXIN1	chr16:347063	missense	c.1948G>A	p.G650S <sup>1</sup>	51.50%	Neutral	Tolerated
H10-TN1	AXIN1	chr16:347063	missense	c.1948G>A	p.G650S <sup>1</sup>	14.90%	Neutral	Tolerated
H43-TN1	AXIN1	chr16:354308	missense	c.1250G>A	p.R417H <sup>1</sup>	29.40%	Neutral	Damaging
H36-TN1	AXIN1	chr16:354353	missense	c.1205C>T	p.T402M <sup>1</sup>	58.40%	Neutral	Damaging
H36-OT	AXIN1	chr16:354353	missense	c.1205C>T	p.T402M <sup>1</sup>	77.20%	Neutral	Damaging
H11-TN1	AXIN1	chr16:396193	missense	c.833C>T	p.P278L <sup>1</sup>	50.70%	Neutral	Tolerated
H11-TN2	AXIN1	chr16:396193	missense	c.833C>T	p.P278L <sup>1</sup>	50.90%	Neutral	Tolerated
H11-TN3	AXIN1	chr16:396193	missense	c.833C>T	p.P278L <sup>1</sup>	50.60%	Neutral	Tolerated
H11-OT	AXIN1	chr16:396193	missense	c.833C>T	p.P278L <sup>1</sup>	47.70%	Neutral	Tolerated
Hep3B	AXIN1	chr16:396590	nonsense	c.436C>T	p.R146* <sup>1</sup>	99.30%	ND	ND

Pt ID	Gene Name	Position	Variant Type	cDNA Change	Protein Change	Allelic Frequency	FI (MutationAssessor & PROVEAN based analysis)	Pt ID
H37-TN1	TP53	chr17:7574017	missense	c.1010G>T	p.R337L	87.80%	Deleterious	Damaging
H37-TN2	TP53	chr17:7574017	missense	c.1010G>T	p.R337L	81.30%	Deleterious	Damaging
H37-TN3	TP53	chr17:7574017	missense	c.1010G>T	p.R337L	82.20%	Deleterious	Damaging
H33-TN1	TP53	chr17:7577534	missense	c.747G>T	p.R249S <sup>1</sup>	53.30%	Deleterious	Damaging
H41-TN1	TP53	chr17:7577534	missense	c.747G>T	p.R249S <sup>1</sup>	11.90%	Deleterious	Damaging
H41-TN2	TP53	chr17:7577534	missense	c.747G>T	p.R249S <sup>1</sup>	11.90%	Deleterious	Damaging
H41-TN3	TP53	chr17:7577534	missense	c.747G>T	p.R249S <sup>1</sup>	22.10%	Deleterious	Damaging
H41-TN5	TP53	chr17:7577534	missense	c.747G>T	p.R249S <sup>1</sup>	7.40%	Deleterious	Damaging
H11-TN1	TP53	chr17:7577539	missense	c.742C>T	p.R248W <sup>1</sup>	60.10%	Deleterious	Damaging
H11-TN2	TP53	chr17:7577539	missense	c.742C>T	p.R248W <sup>1</sup>	56.00%	Deleterious	Damaging
H11-TN3	TP53	chr17:7577539	missense	c.742C>T	p.R248W <sup>1</sup>	46.30%	Deleterious	Damaging
H23-TN1	TP53	chr17:7578176	splice	c.672_splice	e6+1	66.10%	splice	splice
Huh7	TP53	chr17:7578190	missense	c.659A>G	p.Y220C <sup>1</sup>	98.20%	Deleterious	Damaging
H07-TN1	TP53	chr17:7578218	missense	c.631A>C	p.T211P	31.70%	Deleterious	Damaging
H15-TN1	TP53	chr17:7578493	nonsense	c.437G>A	p.W146* <sup>1</sup>	16.10%	ND	ND
H15-TN2	TP53	chr17:7578493	nonsense	c.437G>A	p.W146* <sup>1</sup>	38.20%	ND	ND
H42-OT	CDKN2A	chr9:21970916	missense	c.442G>A	p.A148T	50.60%	Neutral	Damaging
H42-TN1	CDKN2A	chr9:21970916	missense	c.442G>A	p.A148T	47.90%	Neutral	Damaging
H12-TN1	CDKN2A	chr9:21970916	missense	c.442G>A	p.A148T	51.60%	Neutral	Damaging
H12-TN2	CDKN2A	chr9:21970916	missense	c.442G>A	p.A148T	50.00%	Neutral	Damaging
H12-OT	CDKN2A	chr9:21970916	missense	c.442G>A	p.A148T	49.70%	Neutral	Damaging
H17-TN1	CDKN2A	chr9:21970916	missense	c.442G>A	p.A148T	50.90%	Neutral	Damaging
H17-OT	CDKN2A	chr9:21970916	missense	c.442G>A	p.A148T	51.60%	Neutral	Damaging
H04-TN1	ARID1A	chr1:27023621	fs-insertion	c.726_727ins GCG	p.G242_A243insA	48.50%	ND	ND
H04-TN2	ARID1A	chr1:27023621	fs-insertion	c.726_727ins GCG	p.G242_A243insA	51.80%	ND	ND
H04-TN3	ARID1A	chr1:27023621	fs-insertion	c.726_727ins GCG	p.G242_A243insA	46.50%	ND	ND
H04-TN4	ARID1A	chr1:27023621	fs-insertion	c.726_727ins GCG	p.G242_A243insA	47.30%	ND	ND

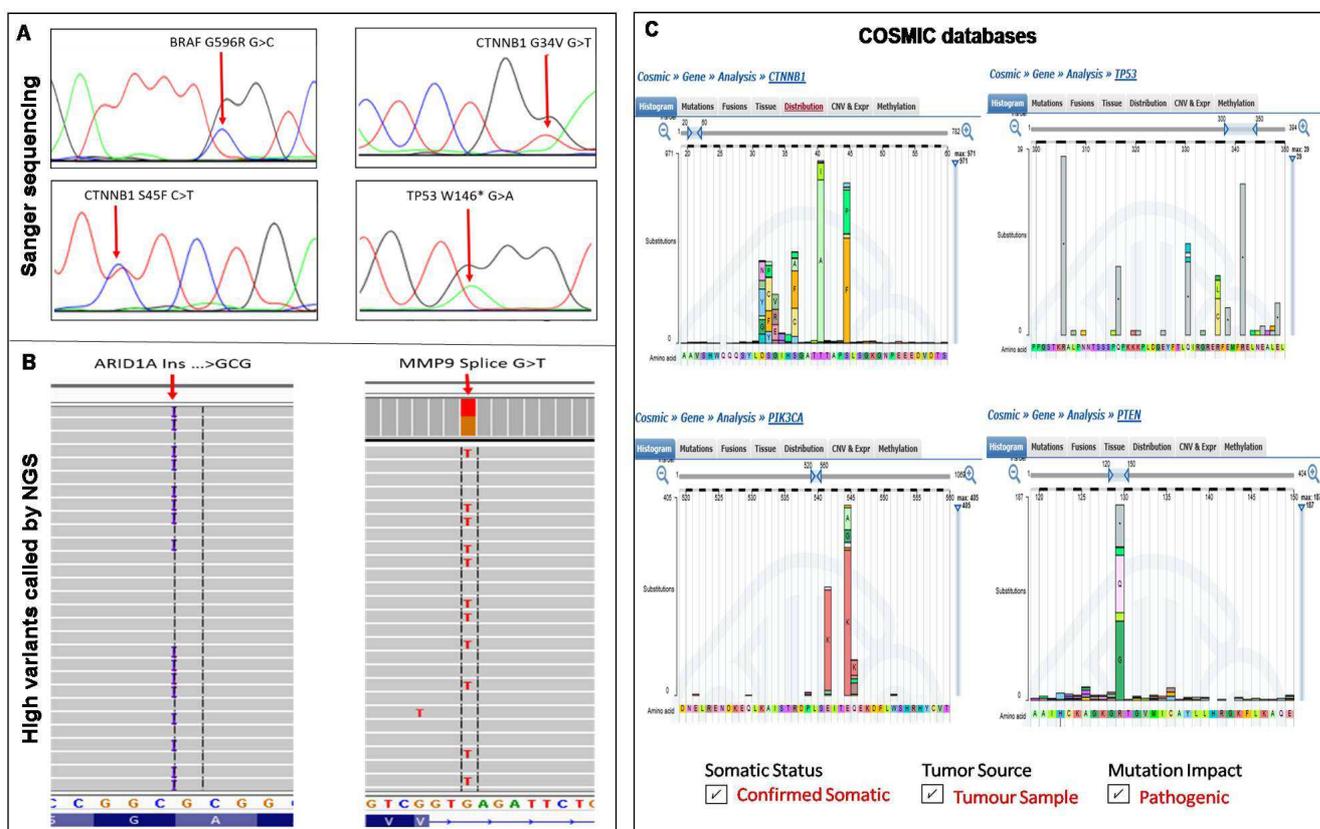
Pt ID	Gene Name	Position	Variant Type	cDNA Change	Protein Change	Allelic Frequency	FI (MutationAssessor & PROVEAN based analysis)	Pt ID
H04-TN5	ARID1A	chr1:27023621	fs-insertion	c.726_727ins GCG	p.G242_A243insA	51.60%	ND	ND
H04-TN6	ARID1A	chr1:27023621	fs-insertion	c.726_727ins GCG	p.G242_A243insA	39.50%	ND	ND
H04-OT	ARID1A	chr1:27023621	fs-insertion	c.726_727ins GCG	p.G242_A243insA	19.80%	ND	ND
H41-TN1	CTNNB1	chr3:41266097	missense	c.94G>T	p.D32Y <sup>1</sup>	9.70%	Deleterious	Damaging
H41-TN2	CTNNB1	chr3:41266097	missense	c.94G>T	p.D32Y <sup>1</sup>	9.70%	Deleterious	Damaging
H41-TN3	CTNNB1	chr3:41266097	missense	c.94G>T	p.D32Y <sup>1</sup>	17.50%	Deleterious	Damaging
H41-TN5	CTNNB1	chr3:41266097	missense	c.94G>T	p.D32Y <sup>1</sup>	7.10%	Deleterious	Damaging
H34-TN1	CTNNB1	chr3:41266098	missense	c.95A>G	p.D32G <sup>1</sup>	17.00%	Deleterious	Damaging
H38-TN1	CTNNB1	chr3:41266098	missense	c.95A>G	p.D32G <sup>1</sup>	29.20%	Deleterious	Damaging
H38-TN2	CTNNB1	chr3:41266098	missense	c.95A>G	p.D32G <sup>1</sup>	29.40%	Deleterious	Damaging
H38-TN3	CTNNB1	chr3:41266098	missense	c.95A>G	p.D32G <sup>1</sup>	32.80%	Deleterious	Damaging
H38-TN4	CTNNB1	chr3:41266098	missense	c.95A>G	p.D32G <sup>1</sup>	26.80%	Deleterious	Damaging
H20-TN1	CTNNB1	chr3:41266100	missense	c.97T>C	p.S33P <sup>1</sup>	46.70%	Deleterious	Damaging
H25-TN1	CTNNB1	chr3:41266104	missense	c.101G>T	p.G34V <sup>1</sup>	36.60%	Deleterious	Damaging
H36-TN1	CTNNB1	chr3:41266104	missense	c.101G>T	p.G34V <sup>1</sup>	29.30%	Deleterious	Damaging
H26-TN1	CTNNB1	chr3:41266124	missense	c.121A>G	p.T41A <sup>1</sup>	40.10%	Deleterious	Damaging
H18-TN1	CTNNB1	chr3:41266134	fs-deletion	c.131_133del CTT	p.P44_S45delinsP	22.50%	Deleterious	Damaging
H40-TN1	CTNNB1	chr3:41266136	missense	c.133T>C	p.S45P	36.10%	Deleterious	Damaging
H40-TN2	CTNNB1	chr3:41266136	missense	c.133T>C	p.S45P	37.20%	Deleterious	Damaging
H40-TN3	CTNNB1	chr3:41266136	missense	c.133T>C	p.S45P	30.70%	Deleterious	Damaging
H40-TN4	CTNNB1	chr3:41266136	missense	c.133T>C	p.S45P	22.60%	Deleterious	Damaging
H40-TN5	CTNNB1	chr3:41266136	missense	c.133T>C	p.S45P	35.50%	Deleterious	Damaging
H40-TN6	CTNNB1	chr3:41266136	missense	c.133T>C	p.S45P	22.20%	Deleterious	Damaging
H22-TN1	CTNNB1	chr3:41266136	missense	c.133T>C	p.S45P	29.40%	Deleterious	Damaging
H22-TN2	CTNNB1	chr3:41266136	missense	c.133T>C	p.S45P	32.90%	Deleterious	Damaging
H30-TN1	CTNNB1	chr3:41266136	missense	c.133T>C	p.S45P	21.00%	Deleterious	Damaging
H33-TN1	CTNNB1	chr3:41266137	missense	c.134C>G	p.S45C <sup>1</sup>	74.50%	Deleterious	Damaging

Pt ID	Gene Name	Position	Variant Type	cDNA Change	Protein Change	Allelic Frequency	FI (MutationAssessor & PROVEAN based analysis)	Pt ID
H32-TN1	CTNNB1	chr3:41266137	missense	c.134C>A	p.S45C <sup>1</sup>	37.60%	Deleterious	Damaging
H32-TN2	CTNNB1	chr3:41266137	missense	c.134C>A	p.S45C <sup>1</sup>	37.50%	Deleterious	Damaging
H32-TN3	CTNNB1	chr3:41266137	missense	c.134C>A	p.S45C <sup>1</sup>	37.70%	Deleterious	Damaging
H29-TN1	CTNNB1	chr3:41266137	missense	c.134C>T	p.S45C <sup>1</sup>	48.40%	Deleterious	Damaging
H02-TN2	CTNNB1	chr3:41280745	missense	c.2258T>C	p.L753P	5.60%	Neutral	Tolerated
H35-TN1	MMP9	chr20:44639692	splice	c.649_splice	e4+3	51.40%	splice	splice
H26-OT	MMP9	chr20:44639692	splice	c.649_splice	e4+3	50.30%	splice	splice
H21-OT	MMP9	chr20:44639692	splice	c.649_splice	e4+3	52.90%	splice	splice
H21-TN1	MMP9	chr20:44639692	splice	c.649_splice	e4+3	28.00%	splice	splice
H23-TN1	MMP9	chr20:44639692	splice	c.649_splice	e4+3	49.80%	splice	splice
H23-OT	MMP9	chr20:44639692	splice	c.649_splice	e4+3	50.40%	splice	splice
H41-OT	MMP9	chr20:44639692	splice	c.649_splice	e4+3	50.40%	splice	splice
H03-OT	MMP9	chr20:44639692	splice	c.649_splice	e4+3	49.00%	splice	splice
H09-OT	MMP9	chr20:44639692	splice	c.649_splice	e4+3	48.70%	splice	splice
H34-OT	MMP9	chr20:44639692	splice	c.649_splice	e4+3	99.30%	splice	splice
H18-TN1	ARID2	chr12:46244409	nonsense	c.2503C>T	p.Q835* <sup>1</sup>	22.10%	ND	ND
H37-TN3	PTEN	chr10:89692941	fs-deletion	c.425delG	p.R142fs	37.80%	ND	ND
H33-TN1	PTEN	chr10:89711926	missense	c.544T>G	p.L182V	5.10%	Neutral	Tolerated
H21-TN1	PTEN	chr10:89711926	missense	c.544T>G	p.L182V	17.40%	Neutral	Tolerated
H42-OT	PTEN	chr10:89711926	missense	c.544T>G	p.L182V	5.10%	Neutral	Tolerated
H24-TN1	PTEN	chr10:89711926	missense	c.544T>G	p.L182V	5.90%	Neutral	Tolerated
H31-TN1	PTEN	chr10:89711926	missense	c.544T>G	p.L182V	7.10%	Neutral	Tolerated
H17-OT	PTEN	chr10:89711926	missense	c.544T>G	p.L182V	5.50%	Neutral	Tolerated
H24-TN1	PTEN	chr10:89711938	missense	c.556C>G	p.L186V <sup>1</sup>	8.00%	Neutral	Tolerated
H43-TN1	ATM	chr11:108122592	missense	c.1636C>G	p.L546V <sup>1</sup>	41.50%	Neutral	Damaging
H06-OT	APC	chr5:112175122	missense	c.3831A>T	p.L1277F	9.20%	Deleterious	Damaging
H11-OT	APC	chr5:112175122	missense	c.3831A>T	p.L1277F	5.40%	Deleterious	Damaging

Pt ID	Gene Name	Position	Variant Type	cDNA Change	Protein Change	Allelic Frequency	FI (MutationAssessor & PROVEAN based analysis)	Pt ID
H17-TN1	APC	chr5:112175122	missense	c.3831A>T	p.L1277F	5.20%	Deleterious	Damaging
H27-OT	APC	chr5:112175240	missense	c.3949G>C	p.E1317Q <sup>1</sup>	45.60%	Neutral	Tolerated
H08-TN1	APC	chr5:112175240	missense	c.3949G>C	p.E1317Q <sup>1</sup>	43.30%	Neutral	Tolerated
H08-OT	APC	chr5:112175240	missense	c.3949G>C	p.E1317Q <sup>1</sup>	45.30%	Neutral	Tolerated
H33-TN1	MET	chr7:116423405	missense	c.3734G>A	p.R1245K <sup>1</sup>	9.90%	Deleterious	Damaging
H21-TN1	MET	chr7:116423405	missense	c.3734G>A	p.R1245K <sup>1</sup>	11.40%	Deleterious	Damaging
H30-TN1	MET	chr7:116423405	missense	c.3734G>A	p.R1245K <sup>1</sup>	7.80%	Deleterious	Damaging
H42-TN1	MET	chr7:116423405	missense	c.3734G>A	p.R1245K <sup>1</sup>	7.80%	Deleterious	Damaging
H24-OT	MET	chr7:116423405	missense	c.3734G>A	p.R1245K <sup>1</sup>	10.00%	Deleterious	Damaging
H24-TN1	MET	chr7:116423405	missense	c.3734G>A	p.R1245K <sup>1</sup>	11.20%	Deleterious	Damaging
H31-TN1	MET	chr7:116423405	missense	c.3734G>A	p.R1245K <sup>1</sup>	10.70%	Deleterious	Damaging
H43-TN1	MET	chr7:116423405	missense	c.3734G>A	p.R1245K <sup>1</sup>	16.20%	Deleterious	Damaging
H26-TN1	HNF1A	chr12:121432091	missense	c.838A>G	p.K280E <sup>1</sup>	34.80%	Deleterious	Damaging
H26-TN1	HNF1A	chr12:121432100	missense	c.847A>G	p.M283V <sup>1</sup>	38.70%	Neutral	Damaging
H27-TN1	BRAF	chr7:140453149	missense	c.1786G>C	p.G596R <sup>1</sup>	41.10%	Deleterious	Damaging
H42-TN1	PIK3CA	chr3:178936029	missense	c.1571G>A	p.R524K	20.00%	Neutral	Tolerated
H33-TN1	PIK3CA	chr3:178936092	missense	c.1634A>C	p.E545A	6.70%	Deleterious	Damaging
H21-TN1	PIK3CA	chr3:178936092	missense	c.1634A>C	p.E545A	5.60%	Deleterious	Damaging
H30-TN1	PIK3CA	chr3:178936092	missense	c.1634A>C	p.E545A	13.90%	Deleterious	Damaging
H42-TN1	PIK3CA	chr3:178936092	missense	c.1634A>C	p.E545A	5.90%	Deleterious	Damaging
H24-OT	PIK3CA	chr3:178936092	missense	c.1634A>C	p.E545A	8.20%	Deleterious	Damaging
H24-TN1	PIK3CA	chr3:178936092	missense	c.1634A>C	p.E545A	7.00%	Deleterious	Damaging
H31-TN1	PIK3CA	chr3:178936092	missense	c.1634A>C	p.E545A	9.50%	Deleterious	Damaging
FI... functional impact; fs... frame shift; ND... not determined; OT... non tumour; TN... tumour nodule								
X <sup>1</sup> ... variants were confirmed by Sanger sequencing								

### 4.1.3 Variant validation by Sanger sequencing

The significant variants were validated using conventional Sanger sequencing (56\115), and the rest were confirmed somatic mutations in tumour samples with pathogenic impact by COSMIC databases, or high confidence somatic mutations called by NGS (Figure 8 and Table 31). Furthermore, obtained variants were analysed for their functional impact on the protein using Mutation Assessor (<http://mutationassessor.org/>; release 2) and the PROVEAN-based analysis software (<http://provean.jcvi.org/index.php>).



**Figure 8: Evaluation of annotated variants.** The significant 115 variants were selected based on Sanger sequencing (A), reliable somatic mutations called by NGS were visually identified in the Integrative Genomic Viewer (IGV, Broad Institute) (B), and somatic mutations were confirmed in tumour samples with pathogenic impact according to COSMIC database (C),

### 4.1.4 Significantly mutated genes in HCC

The highest frequencies of mutated genes were observed in the *CTNNB1*, *TP53*, and *AXIN1* genes (Figures 7B and 9). All defined variants are summarised in Table 31.

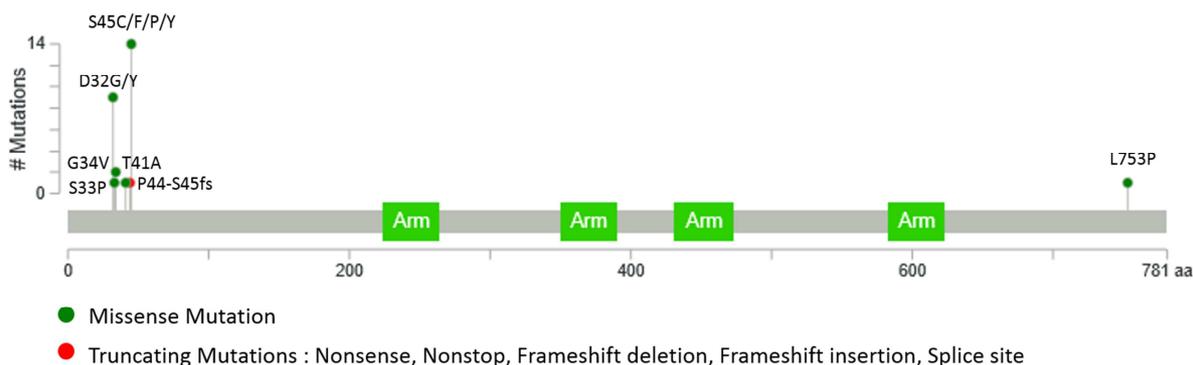
Sample ID	Grade	Cirrhosis	Aetiology	Gene														
				APC	ARID1A	ARID2	ATM	AXIN1	BRAF	CDKN2A	CTNNB1	HNF1A	MET	MMP9	PIK3CA	PTEN	RPS6KA3	TP53
H25	G1	+	HBV/HDV								G34V							
H08	G1	+	HA	E1317Q														
H29	G1	+	Toxic								S45F							
H12	G1	+	HBV							A148T								
H03	G1-2	-	n.d										e4+3					
H33	G2	+	HCV								S45C	R1245K		E545A	L182V			R249S
H32	G2	+	HCV								S45Y							
H26	G2	+	n.d								T41A	K280E/M283V		e4+3				
H36	G2	+	AI					T402M			G34V							
H34	G2	+	HCV								D32Y			e4+3				
H24	G2	+	HCV										R1245K		E545A	L182V	L186V	
H31	G2	+	n.d										R1245K		E545A	L182V		
H43	G2	+	HCV				L546V	R417H					R1245K					
H06	G2	+	HCV	L1277F														
H18	G2	+	HBV					Q835*					P44_S45 delinsP					
H04	G2	+	HCV		G242_A2 43 insA	G1293fs												
H40	G2	-	Toxic								S45P							
H35	G2	-	n.d										e4+3					
H27	G2	-	Toxic	E1317Q					G596R									
H37	G2	-	n.d													R142fs		R337L
H20	G2	-	HBV								S33P						I38S	
H23	G2	-	HC										e4+3					e6+1
H30	G2	-	n.d								S45P		R1245K		E545A			
H42	G2	-	n.d						A148T				R1245K		R524K/E545A	L182V		
H38	G2	-	n.d								D32Y							
H07	G2-3	+	Toxic															T211P
H09	G2-3	+	HBV											e4+3				
H22	G2-3	+	HCV								S45P							
H44	G2-3	-	HBV		S614A											V133I		
H02	G2-3	-	Toxic								L753P							
H21	G2-3	-	Toxic										R1245K	e4+3		E545A	L182V	
H11	G3	+	HCV	L1277F					P278L									R248W
H15	G3	-	n.d															W146*
H41	G3	-	HBV					G650S			D32Y			e4+3				R249S
H10	G3	-	n.d					G650S										
H17	G4	+	HBV	L1277F						A148T						L182V		
Hep3B	\	\	\						R146*									
Huh7	\	\	\															Y220C

**Figure 9:** Global annotated genomic variants in HCC nodules of different grades and in Hep3B and Huh7 hepatoma cell lines. The highest frequencies of potential pathogenic mutations were observed in the  $\beta$ -catenin, TP53 and AXIN1 genes.

The Wnt/ $\beta$ -catenin pathway was the most frequently altered, either due to activating mutations in the *CTNNB1* gene with a frequency of 25.2% (29/115) or inactivating mutations in *AXIN1* (11/115, 9.5%) or *APC* (6/115, 5.2%).

Most *CTNNB1* mutations strongly predicted the impairment of the function of the corresponding encoded proteins, and were predominantly detected in the N-terminus of  $\beta$ -catenin (amino acids 32 to 45), which responsible for  $\beta$ -catenin phosphorylation (Figure 10).

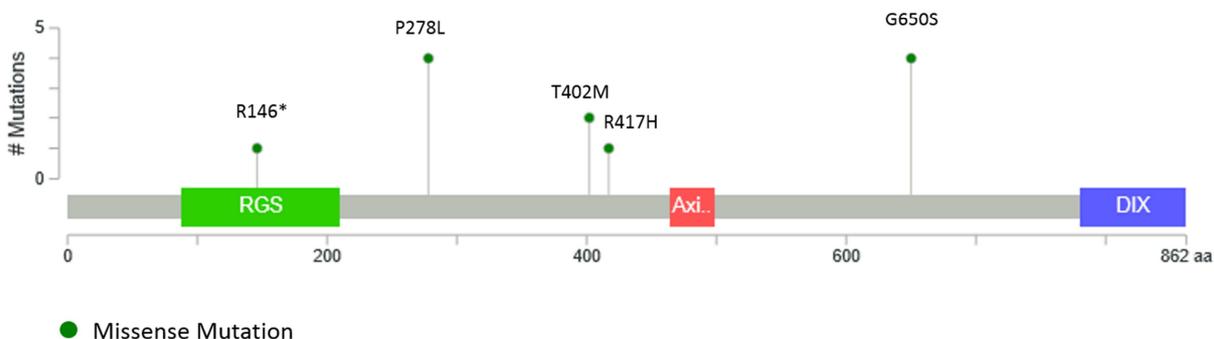
$\beta$ -catenin mutations were detected in 25% of HCCs; most of these were considered to cause amino acid substitutions (28 of 29), while one was interstitial deletion.

**CTNNB1**

**Figure10:** Sites of identified mutations in the *CTNNB1* gene. Arm: Armadillo/beta-catenin-like repeats (224-264/ 350-390/ 431-473/ 583-623).

Notably, *CTNNB1* mutations were only identified in the tumour sample, and were absent in corresponding non-tumour tissue. In contrast, *AXIN1* and *APC* mutations occurred in both the tumour and matched normal samples. Furthermore, point mutations in the *AXIN1* gene, e.g. at p.G650S and p.T402M, showed a higher frequency in the non-tumour sample (51.5% and 77.2%, respectively) than in the tumour sample (14.9% and 58.4%, respectively). This indicates the loss of heterozygosity (LOH) in the tumour samples, suggesting a tumour suppressor function of this gene.

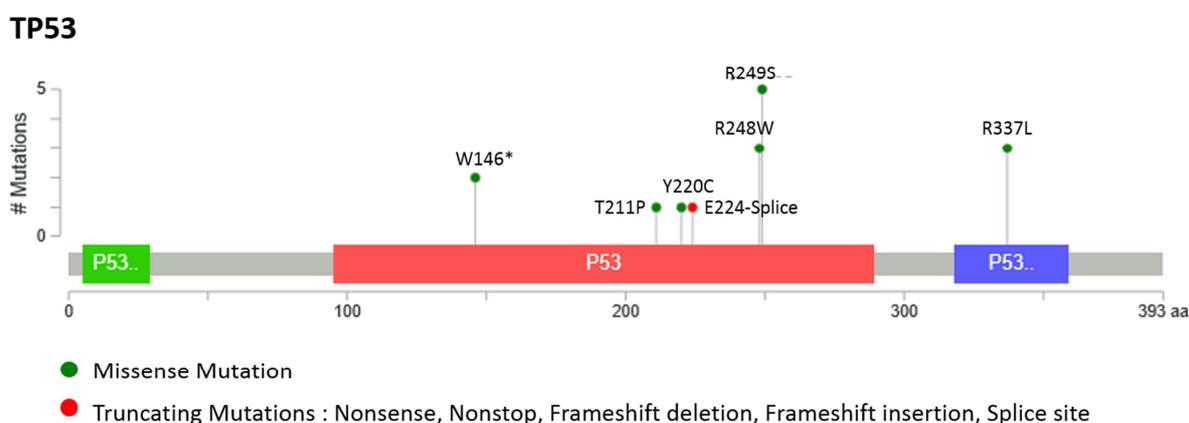
Four missense mutations of the *AXIN1* gene, p.G650S, p.R417H, p.T402M, and p.P278L, occurred in 10 samples, and one nonsense mutation p.R146\* was detected in the Hep3B cell line (9.5%) (Figure 11).

**AXIN1**

**Figure 11:** *AXIN1* variants and their detected locations. RGS: Regulator of G protein signalling domain (88-210); Axi: Axin beta-catenin binding domain (464-498); DIX: DIX domain (780-862).

Somatic changes of the *APC* gene were two missense mutations, p.E1317Q and p.L1277F, in six samples (5%).

The second most frequently altered pathway in HCC was the TP53 pathway. This resulted from either *TP53* inactivating mutations (16/115, 14%), or mutations in the *CDKN2A* gene (7/115, 6%), which is the downstream cellular outcome of TP53 and is important for the tumour suppressive ability of wild-type TP53. All except one *TP53* mutation occurred in the DNA-binding domain of the tumour suppressor and were identified to disrupt the TP53 DNA-binding function (Figure 12).



**Figure12:** Somatic mutations in the *TP53* gene. P53\_TAD pfama: P53 transactivation motif (5-29); P53 pfama: P53 DNA-binding domain (95-289); P53\_tetramer pfama: P53 tetramerisation motif (318-359).

Most of *TP53* mutations were missense mutations (p.T211P, p.Y220C, p.R248W, p.R249S, and p.R337L) and were predicted to cause a dysfunctional protein (Table 31).

Furthermore, one missense mutation was detected in *PICK3A* (p.E545A), *MET* (p.R1245K), *BRAF* (p.G596R), and *HNF1A* (p.K280E) each; these were determined to be functionally damaging mutations. Other mutations exhibited non-functional impact, and were demonstrated in both tumour and non-tumour tissue. This indicates a polymorphic variant, or these tumours exhibiting loss of heterozygosity.

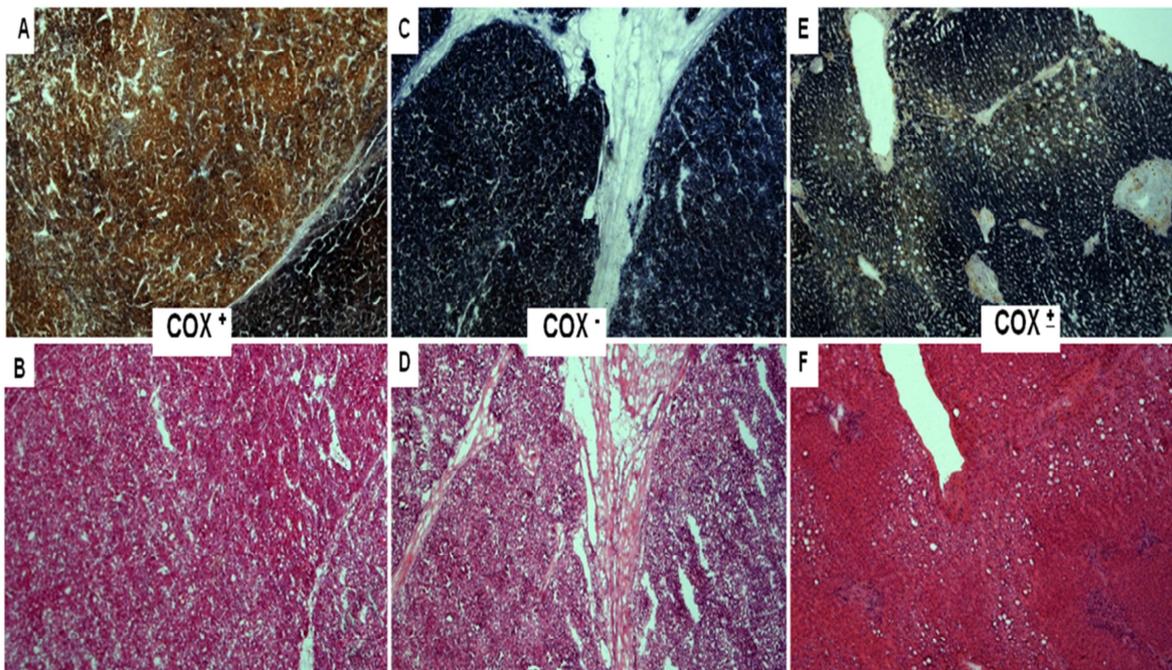
All non-synonymous defined variants with an allelic frequency  $\geq 5\%$  are summarised in Table 31.

## **4.2 Evolution analysis of HCC by ultra-deep sequencing of the mt-genome**

Tumour tracking and evolution analysis are of particular interest for identifying the intra-tumour clonal structure and heterogeneity, affecting overall survival rates and therapeutic outcomes of many cancer types. Since HCCs manifest by a nodule-in-nodule development, in molecular pathology it is an important issue to identify whether the secondary tumour is a recurrent tumour or a new primary tumour that has a more favourable prognosis and overall survival rate. In the present study, a novel approach was designed for analysing the clonal characteristics of the HCC nodules by screening variations of the mt-genome[104].

### **4.2.1 Characterisation of HCC nodules by histoenzymatic stains (COX/SDH)**

In order to assess the clonal architecture of HCC nodules, first histoenzymatic stains were performed. Cytochrome C oxidase/succinate dehydrogenase (COX/SDH) double-labelling histochemistry stains are considered to be the histologic marker of the mitochondrial disorders. COX activity reflects the activity of complex IV of the mitochondrial respiratory chain and is mainly encoded by mtDNA, whereas SDH represents the activity of complex II and is encoded exclusively by nuclear DNA (nDNA). Sequential staining for COX followed by SDH was performed in this study. Deficient COX activity resulted in a reduction of nitro-blue tetrazolium chloride (NBT) due to SDH function, and to the formation of black or blue formazan (COX<sup>-</sup>). Brown insoluble compound resulted at the site of the COX activity (COX<sup>+</sup>) due to the use of 3,3' di-amino benzidine (DAB). Partial deficient COX activity resulted in brown and blue staining (COX<sup>±</sup>), as shown in Figure 13.

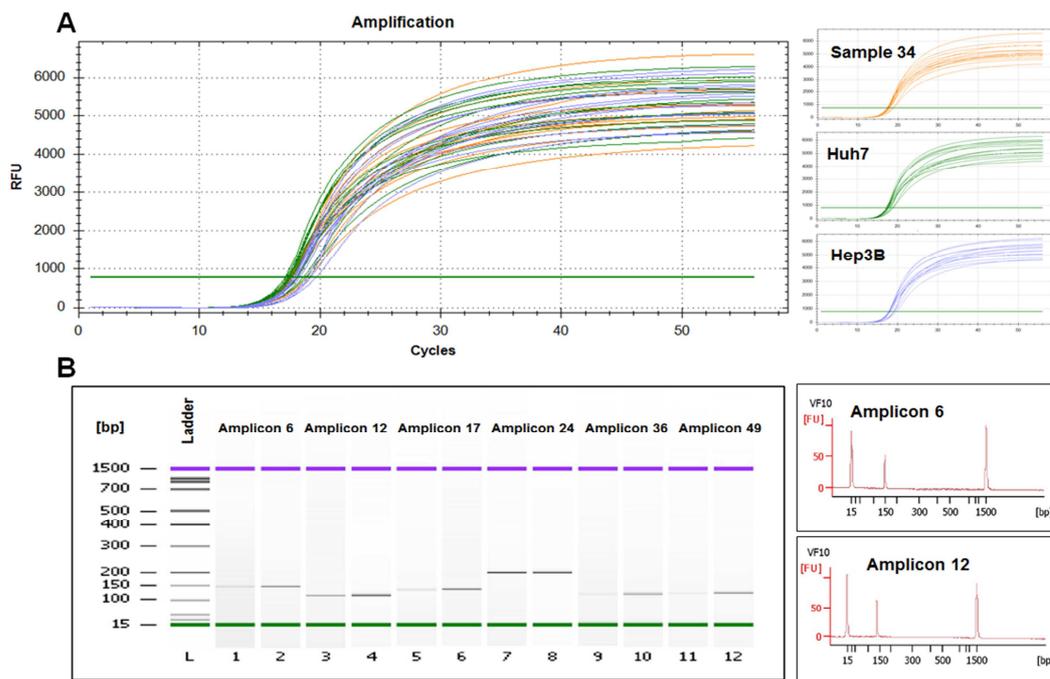


**Figure 13: COX/SDH staining in HCC lesions showing three different nodule types.** Use of DAB resulted in a brown insoluble compound at the site of COX activity, indicating that this nodule is COX positive (COX<sup>+</sup>) (A). Deficient COX activity resulted in the reduction of NBT and in the formation of dark blue formazan, demonstrating that this nodule is COX deficient (COX<sup>-</sup>) (C). Furthermore, intermediate activity of COX resulted in phenotypically mixed nodule (COX<sup>±</sup>) (E). (B, D, F) represent the corresponding haematoxylin and eosin staining.

#### 4.2.2 Establishment of the mt-primer design

For molecular tumour tracking, a PCR-based NGS of the entire mt-genome was established. In order to generate amplicons of low size (around 60-200bp), 108 primer sets, spanning the entire mtDNA (see Section 3.2.9.1), were designed according to the mt-sequence of accession no. [NC\\_012920](#) or taken from previously published primer sets (Figure 2, Table 9 of primer seq.).

The designed mitochondrial primers were tested on hepatoma cell lines Huh7 and Hep3B, as well as on HCC samples by qPCR and microfluidic analysis, showing an efficient and specific amplification (Figure 14).



**Figure 14: Efficient PCR amplification of mt-regions by designed mt-primers.** The average Ct value obtained during the amplification was 18 (A), and the PCR products resolved by microfluidic-based electrophoresis showed a single sharp band in the range of the molecular weight, estimated from the respective primer sets (B).

#### 4.2.3 Data analysis revealed an extensive and ultra-deep analysis, targeting the entire mt-genome

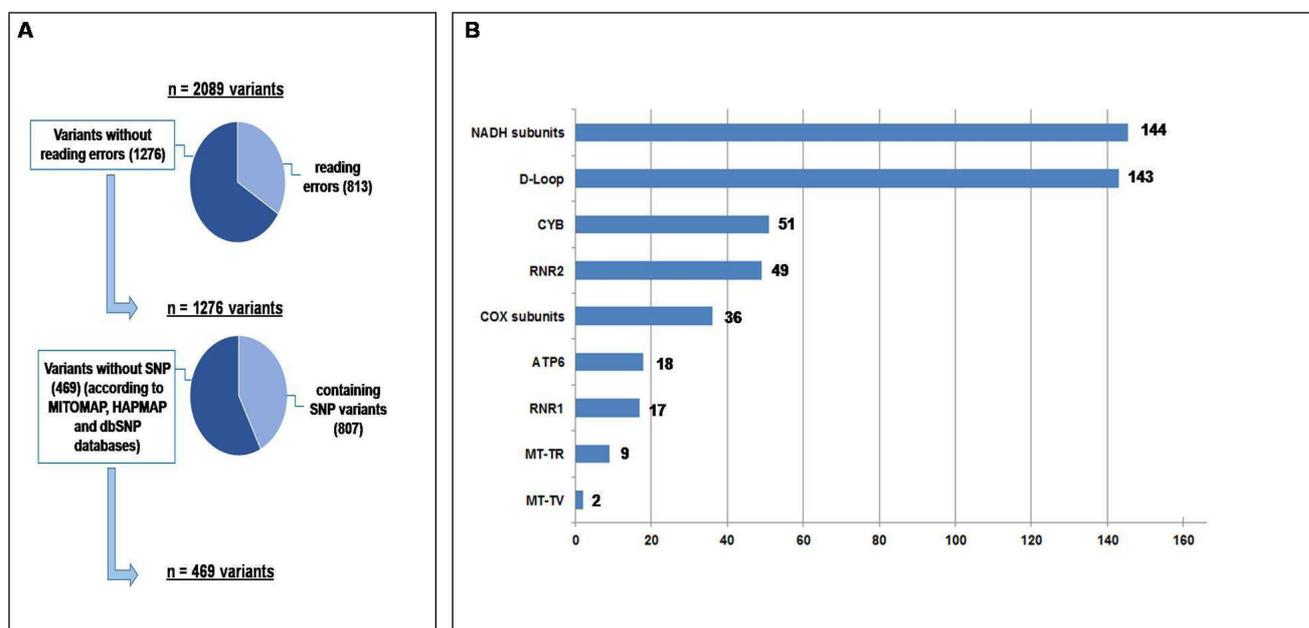
Forty-eight HCC nodules from 10 samples were studied for nodule heterogeneity. FASTQ files generated by the entire mitochondrial sequence of 48 samples were applied to bioinformatics analysis. The sequences were mapped against the entire human reference genome hg19 to exclude the possibility that nuclear pseudogenes, which have a high homology to the parts of the mt-genome, were recognised. For variant calling and annotation, a workflow was used. As a threshold for variant calling, a minimum read depth was set to 30, and the minimum variant frequency to 5%. Polymorphisms were recognised using the MITOMAP ([www.mitomap.org](http://www.mitomap.org)), dbSNP-v138 ([www.ncbi.nlm.nih.gov/SNP](http://www.ncbi.nlm.nih.gov/SNP)), and HAPMAP\_phase\_3 ([hapmap.ncbi.nlm.nih.gov](http://hapmap.ncbi.nlm.nih.gov)) databases.

The data analysis output generated an average of  $199 \times 10^4$  reads per sample. A total of  $194 \times 10^4$  reads were mapped to the mtDNA reference sequence, demonstrating ca. 98% run specificity. Thus, only 2.5% of all mapped reads were detected in out-target

regions, mainly affecting chromosome 1 (1.6%), thus indicating the efficiency of the designed primer sets and a good run performance for the analysis of the entire mt-genome.

#### 4.2.4 Data filtering and accurate mt-gene annotation

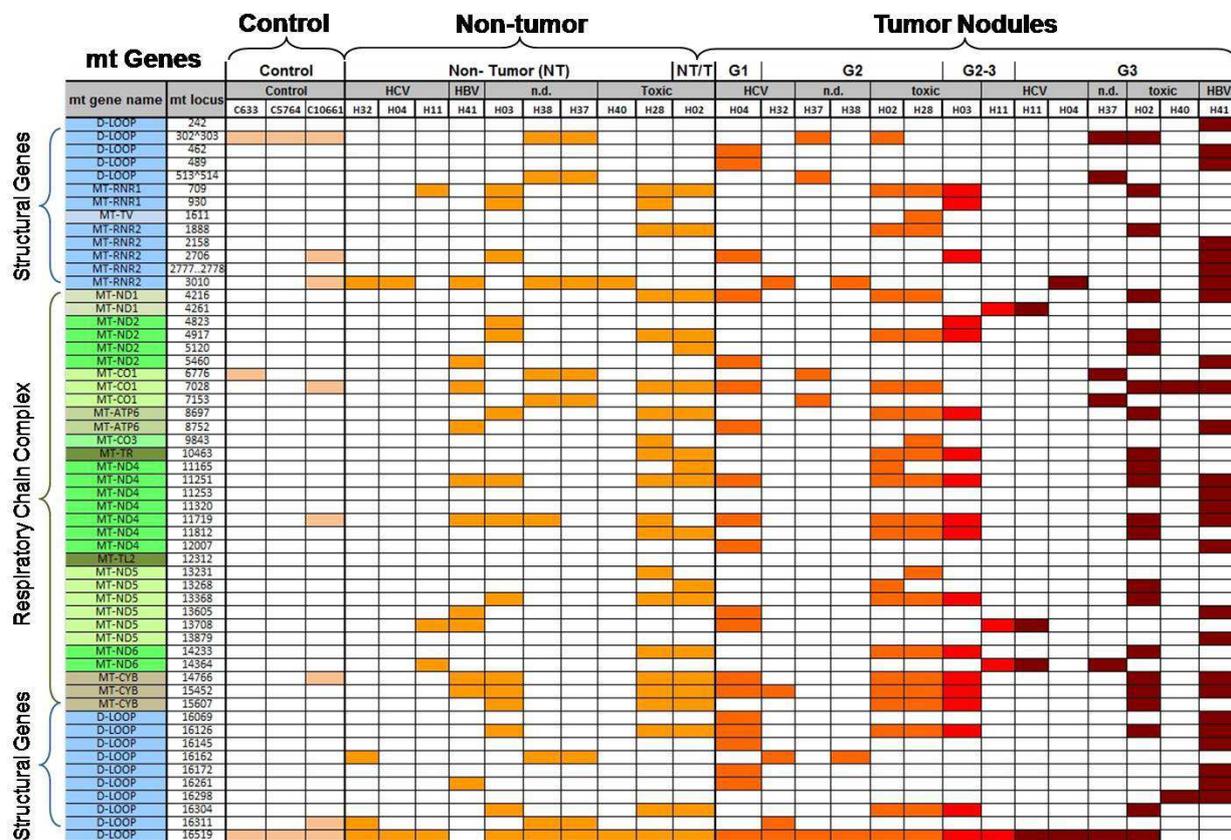
In total, the 48 HCC patient samples and three liver specimens of the control produced a total of 2,089 variants. Seven variants (263A, 311C-315C, 750A, 1438A, 4769A, 8860A, and 15326A) were excluded, as they are considered to be polymorphisms according to the MITOMAP database. Furthermore, the \*3107del variant, announced by 'N' annotating the wobble nucleotide in the reference, and 18 additional variants 16296C, 16294C, 16163A, 15928G, 10916T, 8995G, 295C, 235A, 207G, 204T, 200A, 199T, 195T, 189A, 186C, 182C, 93A, and 73A, recognised by the HAPMAP\_phase\_3 and dbSNP\_v138 databases, were removed. Next, sequencing errors, mostly occurring in homopolymer tracts, sequence repeats, and at the amplicon ends, were excluded from the data, resulting in 469 confident variants (Figure 15, Supplemental Table S3). The most relevant variants (430/469, 92%) were validated using Sanger sequencing.



**Figure 15: Workflow of NGS data interpretation and mt-variant frequencies.** After deletion of the reading errors and dbSNP variants as described in MITOMAP, dbSNP-v138, and HAPMAP\_phase\_3 databases from total sequencing variant output, 469 reliable variants remain (A). The most frequent mt-variants are located in the genes of respiratory chain, and at the D-LOOP regulatory site for replication and expression (B).

### 4.2.5 High frequency of mt-mutation located in D-LOOP and respiratory chain subunits

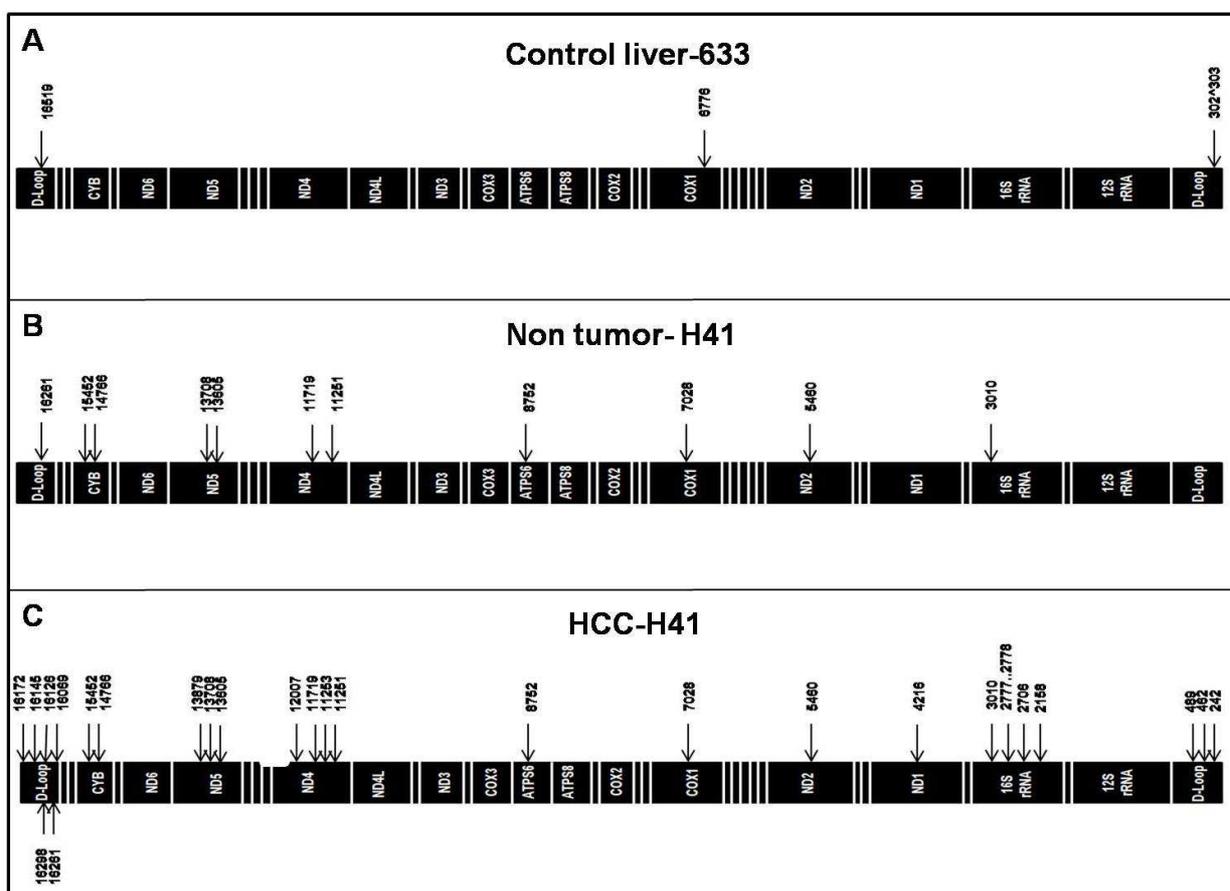
NGS analysis revealed a wide spectrum of mt-alterations. In agreement with previous reports[44, 45, 105], most of the mutations detected in the present study were homoplasmic, and the highest frequency was observed in the genes of respiratory chain complexes (144/469, 31%), and in the D-LOOP regulatory site for replication and expression of mtDNA (143/469, 30.5%). These mutations could lead to an abnormal metabolism, as well as to altered apoptosis. Furthermore, most of the samples contained a T>C transition at nucleotide 16519 (Figure 16). The defined mutations are listed in the Supplemental Table S3.



**Figure 16: Global annotated mt-variants in the HCC cohort and control liver.** A high frequency of mt-mutation was observed in the D-LOOP region and in the genes of respiratory chain complexes. The accumulation of mt-mutations in non-tumour lesions is indicative of early genetic alteration of the mt-genome.

#### 4.2.6 Accumulation of mt-mutations in non-tumour areas, reflecting the pre-tumour machinery of mtDNA

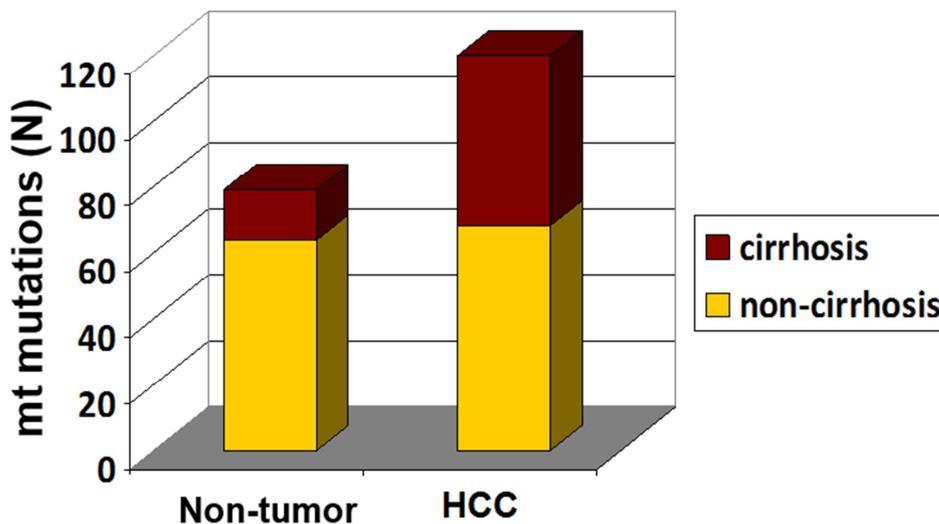
Interestingly, highly frequent mt-mutations were observed in both tumour and non-tumour liver specimens, compared to the control liver. The mtDNA sequence obtained from HCC nodules (e.g. Sample-H41) showed accumulated mt-mutations compared to non-tumour lesions of individual patients, and to the control liver, as depicted in Figure 17. This indicates that the mtDNA is susceptible at the earliest stage of hepatocarcinogenesis.



**Figure 17: Annotated mt-mutations in HCC sample and in control liver.** Numerous mt-mutations were detected in HCC nodules, here shown exemplary in the HCC sample H41 (26 variants) (A), and some of these mutations were apparent in the corresponding non-tumour, but cirrhotic liver specimen of the same patient (11 variants) (B). Only three variants were observed in the liver specimen of a control tissue derived from the non-tumorous part of CRC metastasized liver (C).

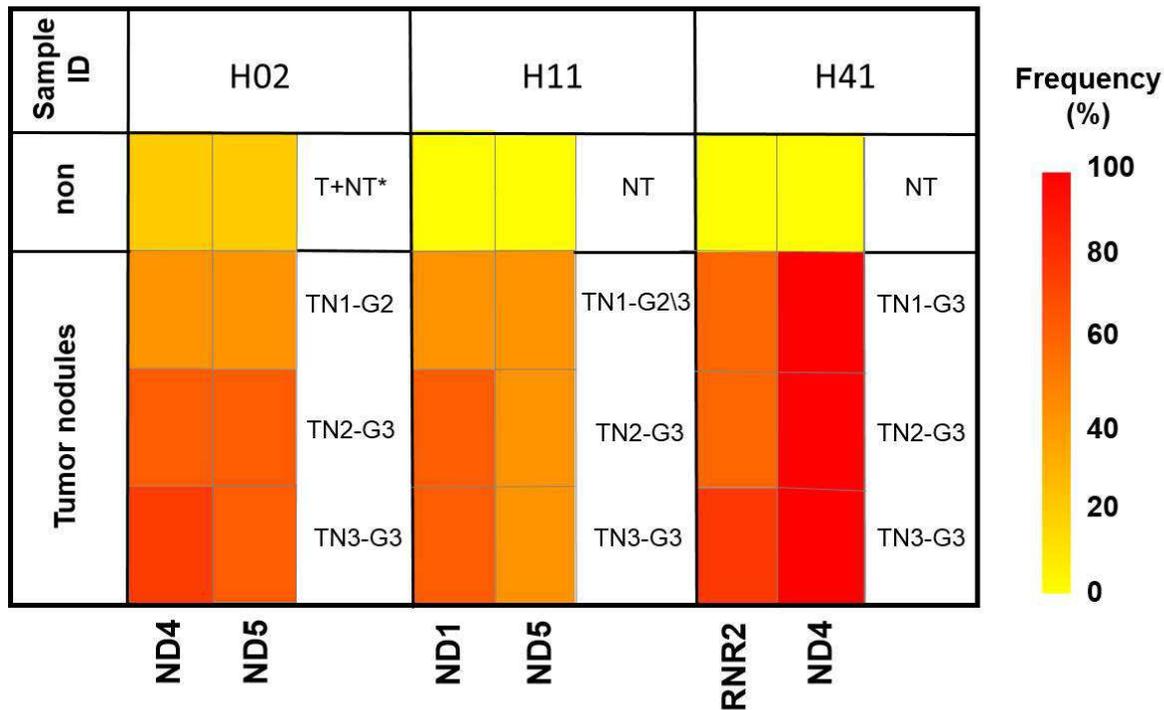
In addition, there were differences in the overall frequency of alteration between samples without cirrhosis and samples with cirrhosis. A high mt-mutation rate was

observed in non-cirrhotic, non-tumour liver lesions. However, in cirrhotic liver disease, mt-mutation frequencies were markedly increased after HCC development (Figure 18).



**Figure 18: The high rate of mt-mutation in HCC nodules is not based on advanced fibrosis or cirrhosis.** Quantitative analysis of the number of mt-mutations [N] in cirrhotic and non-cirrhotic pre-tumorous tissue showed a high mt-mutant accumulation in non-cirrhotic, non-tumour lesions, and occurred in nearly equal frequency in the HCC nodules.

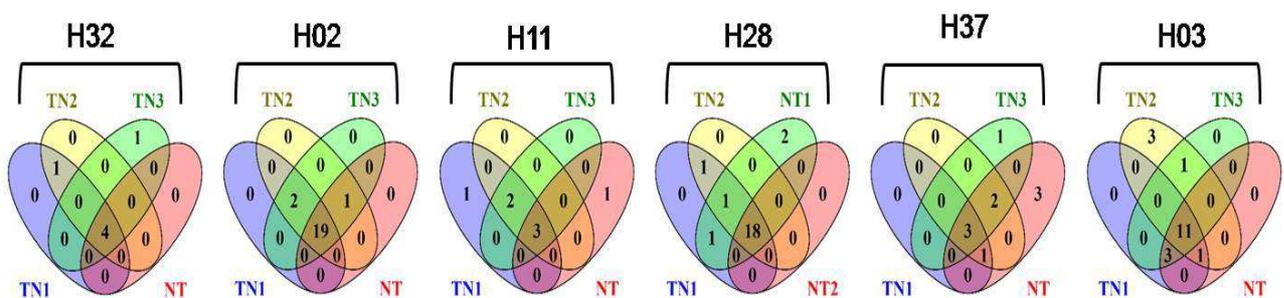
When comparing the mt-mutations present in tumour and non-tumour lesions of individual patients, some mt-mutations were found to exhibit a homoplasmic conversion in tumour nodules, while the heteroplasmic mutations present in non-tumour lesions became homoplasmic mutations in tumour nodules. These observations support the suggestions made in previous studies [106, 107]. Furthermore, some mt-mutations were detected only in tumour nodules and their frequencies correlated with tumour dedifferentiation, suggesting that the accumulation of mtDNA mutations in HCC tissue reflected the degree of malignancy (Figure 19).



**Figure 19: Frequency of mt-mutations in HCC nodules with different tumour grades.** The heat map shows the increasing frequency of mt-mutations linked to advanced tumour grades. (TN: tumour nodule; NT: non-tumour; T+NT\*: sample obtained at tumour border with mixed tumour and normal liver tissue).

#### 4.2.7 Analysis of HCC clonality by mt-genome

Screenings of mt-mutations are well-established clonal markers for cell lineage tracing. The present results show that the HCC nodules and non-tumour lesions harboured identical mt-mutations, indicating a common cell of origin (Figure 20, Supplemental Table S3).

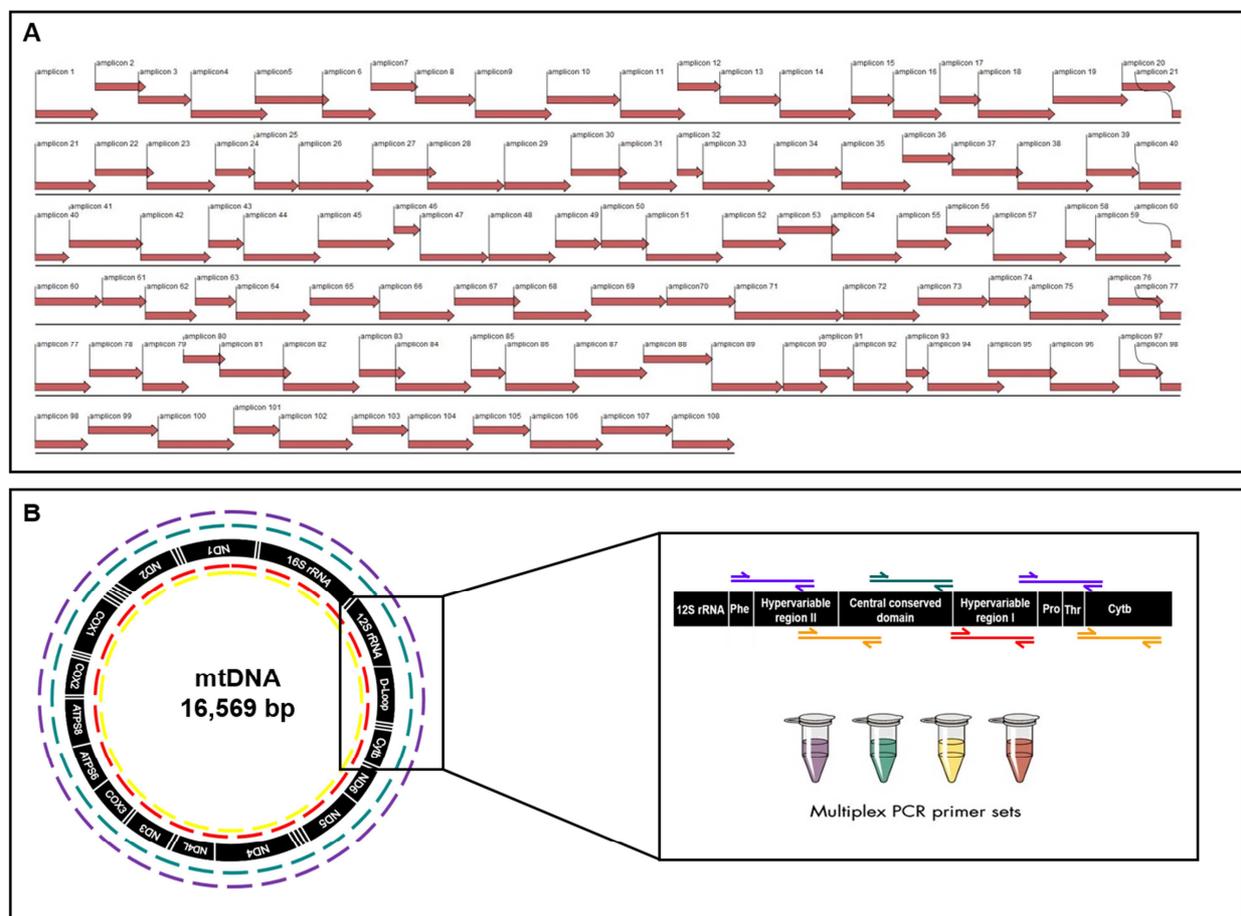


**Figure 20: Venn diagrams illustrating common mt-variants (intersection area) among the HCC nodules (TN) and non-tumour lesions (NT) of individual patients (H32, H02, H11, H28, H37, H03).**

### 4.3 Establishment of entire mt-genome deep-sequencing for tumour-tracking and clonality analysis of FFPE material

In order to use mt-genome analysis for tumour clonality analysis and evaluation of tumour history on a wide panel of samples which are mostly available as FFPE material, the technology of ultra-deep sequencing was adapted to the low DNA quantity and quality of FFPE biopsy material.

First, mt-genome target enrichment was established by multiplex PCR instead of singleplex PCR amplification, reducing the DNA input amount by a factor of 25 (Figure 21).



**Figure 21: Scheme of primer design and the multiplex PCR-based approach of mt-genome enrichment.** Designed primer sets generating 108 amplicons spanning the whole mitochondrial genome (A), were pooled in four primer pools (each 27 primer sets) for multiplex PCR (B), in order to ensure that overlapping amplicons are generated in separate reaction mixes.

Next, the method was tested on tumour DNA from FFPE material. Notably, the coverage of the mt-genome was 98 % and the out-off target sequencing was only 2 % as described by Amer et al. [104]. The method was then primarily used on various lesions of non-small-cell lung cancer (NSCLC). The well histologically classified, heterogeneous NSLCC lesions were studied for their tumour history by ultra-deep sequencing of the mt-genome. Application of this novel technology was successfully established on this approach, as described in my paper Amer et al. 2016 [104].

## 5 Discussion

HCC is an aggressive cancer of the liver with poor prognosis, and there is no effective treatment against it in spite of aggressive conventional therapy [108]. Genetic alterations in HCC have been identified previously using whole genome or whole exome sequencing [18, 94,95]. In the present study, the genetic alterations in different nodules of HCC, using a nuclear and mt-genome analysis by ultra-deep sequencing were addressed. The genomic HCC assay was first applied on 115 matched HCC tumour and non-tumour tissue areas of a total of 44 HCC patients, using newly designed HCC-relevant target panels. Furthermore, a novel approach of whole mtDNA analysis was used as a molecular tool for tumour cell tracking. Interestingly, the data from the mt-genome analysis revealed that the mt-genome alterations precede hepatocellular oncogenic transformation and might be used as an early marker of hepatocarcinogenesis.

### 5.1 Massive sequencing of HCC-relevant nuclear gene targets

In the present study, a novel HCC-relevant panel of genomic targets was first designed for ultra-deep sequencing using the information about HCC-relevant genomic alterations from previous whole exome data analyses [18, 94-96], as well as updated data documented in the COSMIC database. Next, the primer sets were configured in order to enrich the coding exomes of potential HCC-relevant genes by multiplex PCR assays.

In previous studies, whole genome (WGS) and whole exome sequencing (WES) were applied to identify the causative variations that might be critical in the development of HCC. Although these approaches provided information about HCC typical genomic alterations and might be used as a primary screening method, they do not present ultra-deep sequence data, which are needed to find rare variants and to evaluate tumours of early dedifferentiation.

Another benefit of target sequencing is that it enables simultaneous testing for a broad spectrum of point mutations and indels with high accuracy and sensitivity [109]. Furthermore, many variants detected by previous approaches are prone to high rates of false positives and false negatives. The false positives are due to high sequence reads, which are associated with sequencing errors and

misalignment [110], whereas most false negatives result from low coverage, or aggressive data filtering obtained from WGS [111].

The targeted gene panels that were designed for this thesis encompassed hotspots and commonly mutated regions in HCC, to detect the genomic changes resulting in oncogene activation and tumour suppressor gene inactivation, which are essential for maintenance of the malignant phenotype. Gene panel sequencing provides efficient target enrichment, analysis of several genes in one comprehensive test, high coverage in targeted regions, and sequences with high uniformity. Therefore, efficient strategies for selectively sequencing targeted gene panels would be preferable over whole exome sequencing for determining disease-causative mutations [112]. Notably, the genomic data generated from gene panel sequencing can help in the development of novel targeted therapies, and can assist in the selection of currently available treatments likely to be most effective.

By using the current protocol in this study, the simultaneous testing of 394 disease-specific targets was possible. Optimal sequencing depth of targeted regions is sufficient to detect low frequent mutation at every single nucleotide of a genomic region [113].

### 5.1.1 $\beta$ -Catenin and TP53 mutations in HCC

After a good NGS run performance, the data analysis revealed that the  $\beta$ -catenin gene (*CTNNB1*) was most frequently mutated. These data confirm previous studies that demonstrate that the Wnt/ $\beta$ -catenin pathway is the main and early events in hepatocarcinogenesis [18, 24,114]. Since the Wnt/ $\beta$ -catenin pathway is implicated in cell proliferation, migration, and invasiveness of cancer cells, activation of this pathway is associated with an aggressive HCC phenotype [114, 115]. In HCC the Wnt/ $\beta$ -catenin pathway is mainly activated by *CTNNB1* exon 3 mutations [18, 116].

In the present study, *CTNNB1* mutations were detected in 25% of HCCs, and were mostly considered to cause amino acid substitutions, while one mutation was an interstitial deletion. Most of the  $\beta$ -catenin mutations were located in the exon 3 encoding the N-terminus with amino acids 32 to 45. This location is representing

four presumptive GSK3b phosphorylation sites, and is responsible for  $\beta$ -catenin phosphorylation. These data confirm previous work [18, 116], demonstrating that the N-terminus is highly mutated, in this way affecting  $\beta$ -catenin phosphorylation, and thereby phosphorylated  $\beta$ -catenin is rapidly degraded by the ubiquitin-proteasome pathway [117].  $\beta$ -catenin mutations were detected exclusively in low- and high-grade tumour samples but were absent in the corresponding non-tumour tissue, indicating  $\beta$ -catenin's contribution in HCC initiation and development [22, 116]. Importantly,  $\beta$ -catenin was not related to a certain risk factor, such as intoxication or viral infection. However, other studies have demonstrated that the Wnt/ $\beta$ -catenin pathway is frequently dysregulated in HCV- and HBV-related HCC [118, 119]. Furthermore, Cieply et al. demonstrate that  $\beta$ -catenin mutations are more often observed in HCCs that are not based on cirrhosis. The data obtained in the present study show that  $\beta$ -catenin mutations occurred in some HCC lesions in the absence of cirrhosis. This indicates that  $\beta$ -catenin mutations themselves might be the tumour driving force, independent to the cirrhotic features [115]. Effective targeting of the Wnt signalling pathway is considered as a potential target for pharmacological therapy that is eagerly awaited.

Furthermore, the dysregulated signalling cascade of the Wnt/ $\beta$ -catenin pathway resulted from mutations in genes that regulate  $\beta$ -catenin signalling. Thus inactivated mutations were found in 9.5% of *AXIN1* and 5.2% of *APC*. Interestingly, novel *AXIN1* variants T402M and G650S were found that occur in the non-tumour tissue with a higher frequency (77% and 51%, respectively) than in tumour samples (58% and 15%, respectively), suggesting a germline polymorphisms of the *AXIN1* gene. The lower frequency in the tumour lesion indicates that one allele of the tumour suppressor *AXIN1* gene is lost in the tumour as a hallmark of hepatocarcinogenesis [120]. *AXIN1* is a master scaffolding protein critical for  $\beta$ -catenin degradation [121]. It acts as a tumour suppressor by negatively controlling the Wnt pathway and as a positive regulator of JNK, TGF $\beta$ , and Tp53 signalling in the liver [122, 123]. Targeting of the *AXIN1* gene is suggested to be an effective therapeutic molecule for tumour growth suppression in HCC [124].

Two *APC* gene mutations were detected in tumour and non-tumour samples. The p.E1317Q mutation seems to be a germline variant, although in the COSMIC

database this mutation is documented as a somatic mutation in various types of cancer. However, the average of 45% frequency in all nodules suggests that, in this respective case, the p.E1317Q is a germline variant that might have caused an increased individual risk of cancer. Furthermore, the p.L1277F mutation was predicted to have a functional impact on the APC protein, and its occurrences in pre-neoplastic cirrhotic lesions, suggesting its role in hepatocarcinogenesis.

In concordance with previous studies, the  $\beta$ -catenin gene was the most frequently mutated oncogene and the *TP53* was the most frequently mutated tumour suppressor gene (14% frequency). More than 50% of all human cancers exhibit mutations in the *TP53* gene [125, 126]. The vast majority of *TP53* mutations are missense mutations, single base-pair substitutions resulting in an exchange of the encoding amino acid. The majority of these missense mutations is located within the DNA-binding domain, the most conserved region of *TP53*[127]. These mutations generate mutant TP53 proteins with novel activities or gain-of-function activities that actively promote tumour growth by dominant-negative effects of the respective allelic TP53 wild type protein [128-130]. Interestingly, inactivated missense mutations in *TP53* were detected in patients with advanced-stage HCC. The *TP53* mutations have been reported to be associated with a high recurrence rate of HCC, suggesting that the mutant TP53 gene is correlated with tumour aggressiveness and metastatic potential [131]. Furthermore, the *TP53* hotspot mutation, R248W, is associated with a significantly poorer prognosis compared to other missense mutations [132]. In addition, the R249S mutation, a hotspot residue linked to aflatoxin B exposure, was observed in patients known to be infected with HCV and HBV, suggesting that aflatoxin B exposure could be a second risk factor for HCC development [9]. Villar et al. found that the R249S mutation occurs independent of cirrhosis in a context of HBV chronic infection [133]. One missense mutation was observed outside the DNA-binding domain in the TP53 tetramerisation domain, amino acid 337 arginine to leucine. This mutation was detected previously by others using yeast assay, which the colonies with R337L mutation exhibit a resistance to temperature, perhaps due to destabilisation of the tetramer [134, 135].

In accordance with previous studies, *TP53* mutations mainly occurred in HCC of a high grade (G3), which might be due to its late mutation [136-138]. Aberration in

*TP53* occurring late in tumourigenesis has already been shown in other tumours such as prostate cancer [139, 140], bladder cancer [141], pancreatic cancer [142], and breast cancer [132].

In addition, altered *TP53* signalling pathway may be caused by *CDKN2A* mutations. The *CDKN2A* encodes the p16 cell cycle regulating protein, downstream acting to *TP53*. One *CDKN2A* variant (p.A148T) was detected; it was a common germline mutant associated with familial cancer predisposition, e.g. melanoma, glioblastoma, and pancreatic cancer [143, 144]. However, this mutant was also previously described as a common polymorphism without a major impact on p16 function [145, 146]. In spite of these controversial reports, recent studies have also considered *CDKN2A* p.A148T to be a melanoma susceptibility variant [147, 148]; however, the influence of the p.A148T *CDKN2A* variant in HCC is not yet known. Nevertheless, it is assumed that it will be also be associated with cell cycle dysregulation and promotion of the hepatocarcinogenesis risk.

In 7% of HCC lesions, a p.R1245K of the *MET* gene occurred. The proto-oncogene *MET* is a protein coding gene that encodes for the receptor tyrosine kinase (RTK), MET, or the hepatocyte growth factor receptor (HGFR). The MET receptor is expressed on the surface of epithelial and endothelial cells. It is a multi-domain receptor that after binding of the HGF ligand, leads to receptor phosphorylation and activation of different downstream signalling pathways, including activation of mitogen-activated protein kinase (MAPK) cascade and the phosphatidylinositol 3 kinase (PI3K)-AKT signal transduction [149, 150]. HGF/MET signalling mediates an increase in cell proliferation and motility, resulting in enhanced tumour invasion and angiogenesis but inhibited apoptosis [151, 152]. MET overexpression was demonstrated in HCC and associated with poor prognosis and an aggressive phenotype [153].

One *BRAF* mutation (p.G596R) was detected in a single NASH-related HCC sample. The three *RAF* genes code for cytoplasmic serine/threonine kinases that are regulated by RAS binding and driving an uncontrolled process of cell division. This mutation occurs within the kinase domain, and has functional impact on the *BRAF* protein. According to the COSMIC database, mutations within the *BRAF* gene are extremely rare in HCC. Furthermore, p.G596R point mutation in *BRAF* has been detected previously in colorectal [154], melanoma [155], and bladder

cancers [156], and this is the first time that it has been detected in HCC. Sorafenib, the BRAF inhibitor, was the first clinically useful agent for the treatment of advanced unresectable HCC [83].

Somatic mutations in *PIK3CA* were found to carry the activating point mutation p.E545A in exon 9. This hotspot mutation occurs within the highly conserved helical domain, and has been previously reported in HCC [157-160] and several other cancers [161-164]. The functional impact of this mutation is linked to increased oncogenic potential by an increased catalytic activity resulting in enhanced downstream signalling [165]. HCC patients with the hotspot *PIK3CA* mutation (p.E545A) can benefit from inhibitors effective against the kinase activity of this mutant [166].

The coding sequence mutations of the *PTEN* gene were found in patients with advanced-stage HCC. These mutations have non-functional impact and were detected in two non-tumour samples, suggesting the importance of these mutations in tumorigenesis. The *PTEN* gene is a negative regulator of the PI3K/AKT/mTOR pathway by dephosphorylating of the lipid products of PI3K [167]. The PI3K counteracts activity of PTEN, resulting in inhibition of cell cycle progression, induction of cell death, modulation of the arrest signal, and stimulation of angiogenesis [168]. This result is consistent with that of a previous study that demonstrated that the decreased PTEN expression was correlated with tumour progression and poor HCC prognosis [169].

A putative splice site mutation was detected in the *MMP9* gene in tumour and non-tumour lesions with equal frequency, suggesting that this mutation may be a polymorphic variant. The *MMP9* gene mutation may lead to protein alteration and to an increase in the MMP9 expression. The *MMP9* gene plays a key role in cancer invasion and metastasis by degrading the extracellular matrix. The MMP9 was observed to have significant expression in HCC with capsular infiltration, suggesting a role of *MMP9* mutation in tumour invasion and metastatic potential [170, 171].

Two *HNF1A* missense mutations were observed in the same HCCs sample: one functional and one non-functional mutation were located in the DNA-binding domain. *HNF1A* inactivation was observed in the hepatocellular adenomas and in

well-differentiated HCC. The inactivation of both *HNF1A* alleles is required, indicating the tumour suppressor activity of the *HNF1A* gene [172].

Additionally, *ARID1A* and *ARID2* are genes that encode chromatin remodelling complex components. They have recently been shown to be inactivated in several tumour types [173, 174] and also in HCC [94, 96,175]. They are important tumour suppressor genes and a promising therapeutic target for HCC [96, 175,176]. Decreased expression of *ARID1A* has previously been observed in HCC, and has been associated with tumour progression, metastasis, and reduced overall survival. In the present study, frame-shifting insertions were found in tumour suppressors *ARID1A* located in exon 1. Furthermore, there was a nonsense mutation of the chromatin remodelling gene *ARID2*, located in proline-and glutamine-rich region (GLN) in one case. Mutations within subunits of SWI/SNF chromatin remodelling complex are predicted to truncate the *ARID1A* and *ARID2*proteins, leading to their inactivation.

## 5.2 Ultra-deep sequencing of the entire mt-genome

In the light of previous reports on accumulated mutations of the mtDNA in different types of cancer [48, 49, 51, 52]and in HCC [47, 55, 56, 177], an aim of this study was to determine the mtDNA mutations in different HCC grades and its correlation with tumour dedifferentiation. Therefore, the entire mtDNA was screened for the mutations in a cohort of 48 HCCs and the corresponding non-tumour tissues.

Since the mt-genome is a small circular DNAof approximately 16.6 Kb, 108 primer sets were designed to span the entire mt-genome. Conversely, others have focused on the analysis of the control region (D-LOOP), which encompasses a hotspot region but only covers 7% of the total mt-genome [47, 48, 52].

The primer design generated 108 amplicons of 160-bp in average, guaranteeing that these primers were applicable on FFPE material of the pathology routine processing[104], which is characterised by a high degree of DNA fragmentation [178]. Previous NGS-based approaches of mtDNA analysis have used long-range PCR amplicons, which only works on native material [107, 179]. To amplify the entire mt-genome per sample, first 108 singleplex PCRs were successfully performed. However, dealing with many reactions and samples increases the risk

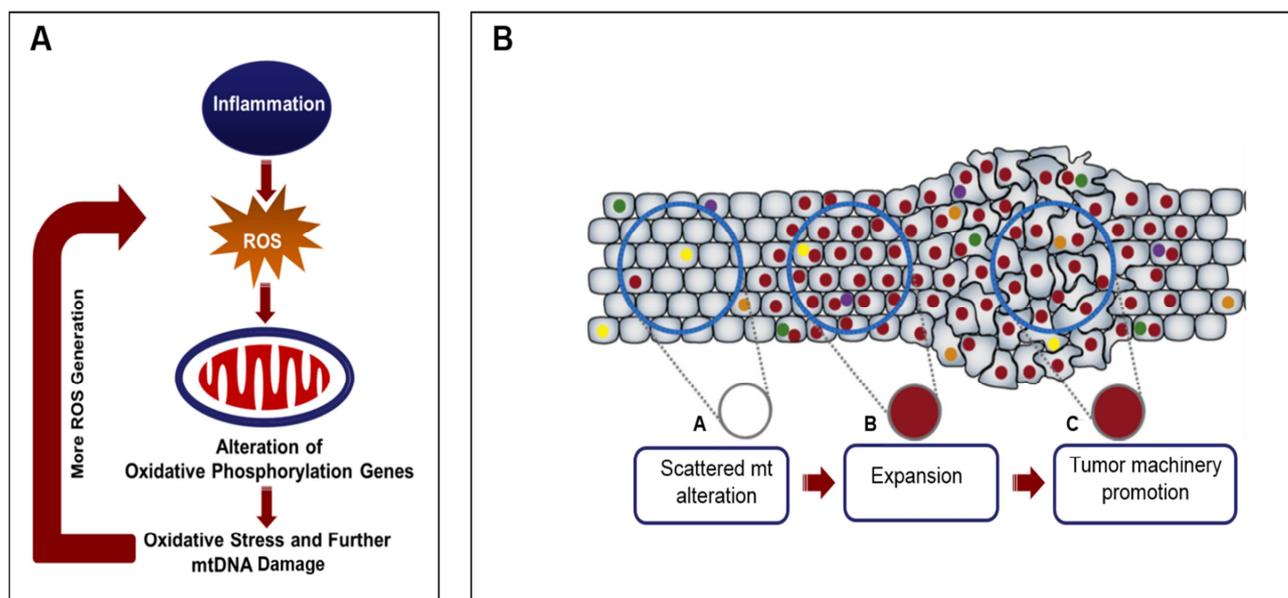
of contamination and error-prone pipetting; in addition, the process is time-consuming. Therefore, an automatic singleplex PCR processing was established in 384 well-layouts. Next, singleplex PCR-based ultra-deep sequencing was conducted.

Notably, the eukaryotic nuclear genome harbours pseudogenes with more than 60% homology to the mt-genome, the so-called nuclear mitochondrial DNA (Numt) as documented in the MITOMAP database. However, the possibility of accidental amplification of these sequences in reactions directed to mtDNA amplification is unlikely when sufficient mtDNA is present in a sample, due to the presence of many copies of mtDNA per cell [180]. Intriguingly, after mapping the sequencing reads against the entire human reference genome hg19, the generated data showed that the out-target annotation was in total only 2.5%, mainly affecting chromosome 1 (1.6%)[104]. These findings indicate the specificity and efficiency of the designed primer for mtDNA target enrichment. A high quality and output of data generated from 48 paired tumour/non-tumour samples demonstrated that the majority of mtDNA were homoplasmic or highly heteroplasmic, as has already been reported in previous studies [44, 105]. In addition, mt-mutations may acquire a selective replicative advantage during cellular development and become dominant, developing a homoplasmic mutant in the clonal cell population [44, 45,181].

### **5.2.1 The mt-genome alteration as an early marker of HCC development**

In agreement with previous reports, a majority of the mutations were homoplasmic, and most were T-C or G-A transitions [44, 45,105]. The highest frequency was observed in the D-LOOP region and in genes of respiratory chain complexes. This might lead to an impairment of replication and transcription of mtDNA, mitochondrial respiration, and Oxidative phosphorylation (OXPHOS) function, which might ultimately result in enhanced reactive oxygen species (ROS) production. Furthermore, a high number of mtDNA mutations were detected in tumour and corresponding non-tumour lesions, and were markedly greater than those in control liver tissue, which is consistent with a previous report [55]. This can be explained by the fact that during chronic liver injury, the parenchymal tissue undergoes repeated destruction and regeneration of hepatocytes, resulting in

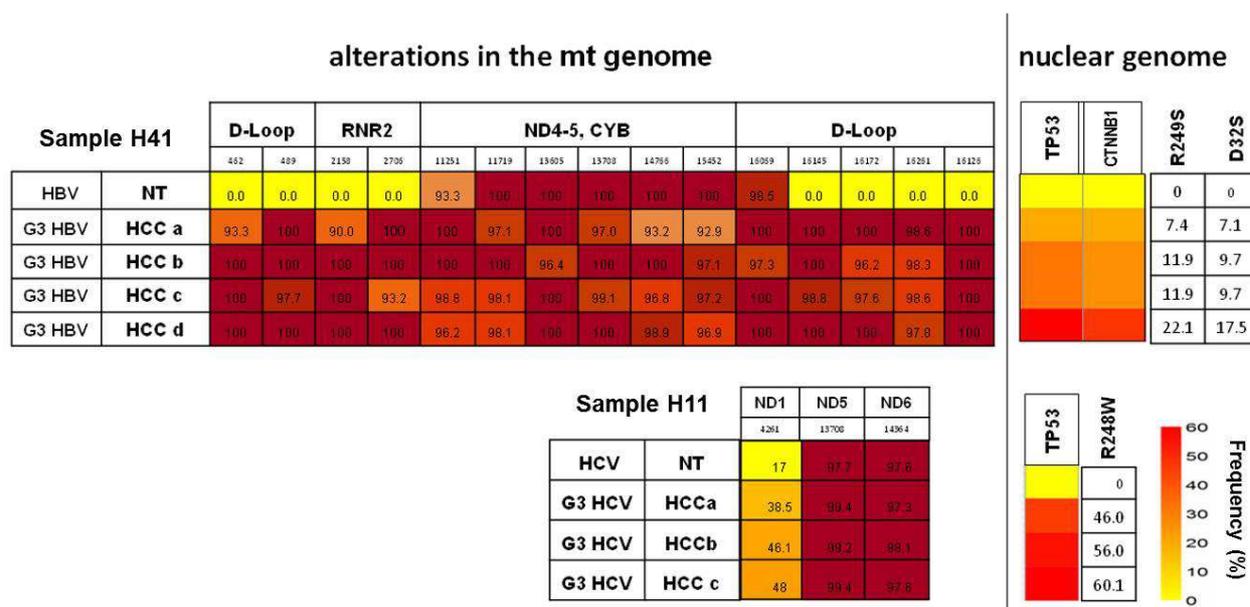
mtDNA alteration and subsequent mitochondrial respiratory dysfunction. Furthermore, a decline in mitochondrial respiratory function and oxidative phosphorylation led to an increase in mitochondrial production of ROS, which in turn led to further mtDNA damage. Repetitive oxidative damage and oxidative stress as a result of additional mitochondrial dysfunction may extend to oxidative damage of nuclear DNA (nDNA), contributing to the mutagenic transformation of hepatocytes [182-185]. Thus, concurrent accumulation of mtDNA mutations in pre-neoplastic lesions may possess a critical role in pre-tumour machinery and may contribute to hepatocarcinogenesis (Figure 22).



**Figure 22: Mechanism of mtDNA mutation and its contribution to tumourigenesis.** (A) Schematic representation of repetitive mtDNA damage by ROS leading to mtDNA mutations (c.f. main text). (B) Hypothetical model whereby mtDNA damage accumulates within the cell during the life [A] and confers a growth/survival advantage likely to become intra- and intercellular dominant. [B] Accumulation of mutant mtDNA causes further oxidative stress and an increase in ROS generation; this may in turn trigger nDNA alterations, which promote the tumour machinery [C](Figure 1B adapted from Salk et al. [186], with modification).

Some homoplasmic mutations were restricted to the tumour lesions, whereas others exhibited a homoplasmic conversion in the tumour nodules. Here, the homoplasmic conversion in tumour lesions means that the heteroplasmic mutation presents in non-tumour lesion become homoplasmic in tumour nodules. The homoplasmic conversions have been reported by others [44, 106, 107], and it has been indicated that a single tumour cell with a mutant mt-genome may acquire a

selective growth advantage during tumour evolution, allowing it to become dominant at the intracellular and intercellular levels [44]. Furthermore, the present study detected some heteroplasmic mutations with a different mutation load in the HCC lesions and the corresponding non-tumour samples. The heteroplasmic mutations can occur as a result of *de novo* mutations, which are not yet homoplasmic or which were intolerant homoplasmic mutations that underwent shift segregation [187]. In addition, heteroplasmic mutations are considered to be pathogenic mutations when the level of mutant mtDNA is higher in clinically abnormal tissues than in normal tissues [187]. Since most mtDNA point mutations are homoplasmic, it can be difficult to prove the pathogenicity of these mutations; therefore the application of parallel genomic nuclear analysis may be helpful. In the present study, nuclear and mitochondrial data aggregation revealed that mt-alterations precede nuclear alterations: e.g in sample21743, mt-mutations of respiratory chain subunits were observed with high frequency in non-tumour lesions, but the nuclear alterations of *TP53* (p.R249S) and *CTNNB1* (p.D32S) were undetected in non-tumour lesions, and increased in frequency with tumour initiation and progression (Figure 23).



**Figure 23: mt-alterations precede nuclear alterations in HCC.** Nuclear genes (*TP53* and *CTNNB1*) were not evident in non-tumour lesions, whereas the homoplasmic mtDNA mutations were accumulated in non-tumour nodules.

Taken together, our findings illustrate that the mtDNA is susceptible at the earliest stage of hepatocarcinogenesis, and that it has a critical role as an early marker of HCC development. Many studies support this result that somatic mt-mutations are specific biomarkers for cancer diagnostics due to; homoplasmic nature of mt-mutations, combined with the presence of hundreds of copies of mtDNA within each cell, 500-1000 fold more numerous in the cell than nDNA [105, 188, 189]. In a study by He et al., the tumour specificity of mutant mtDNA was determined; mt-mutations have been detected in the plasma of colorectal cancer patients after surgical removal of the cancer, while nDNA mutations were not evident. Additionally, the plasma concentration of mutant mtDNA was lower after surgery than before it [107].

Interestingly, in the present study a higher number of mt-mutations were observed in non-cirrhotic non-tumour lesions compared to in cirrhotic non-tumour lesions. This suggests that the exposure to strong carcinogens can induce genetic alterations, particularly in mtDNA, through alterations in cell cycle regulation, oxidative stress, and activated mitogenic signalling pathways that promote the development of HCC without underlying cirrhosis [190, 191]. However, in cirrhotic liver disease, mt-mutation frequencies are markedly increased after HCC development.

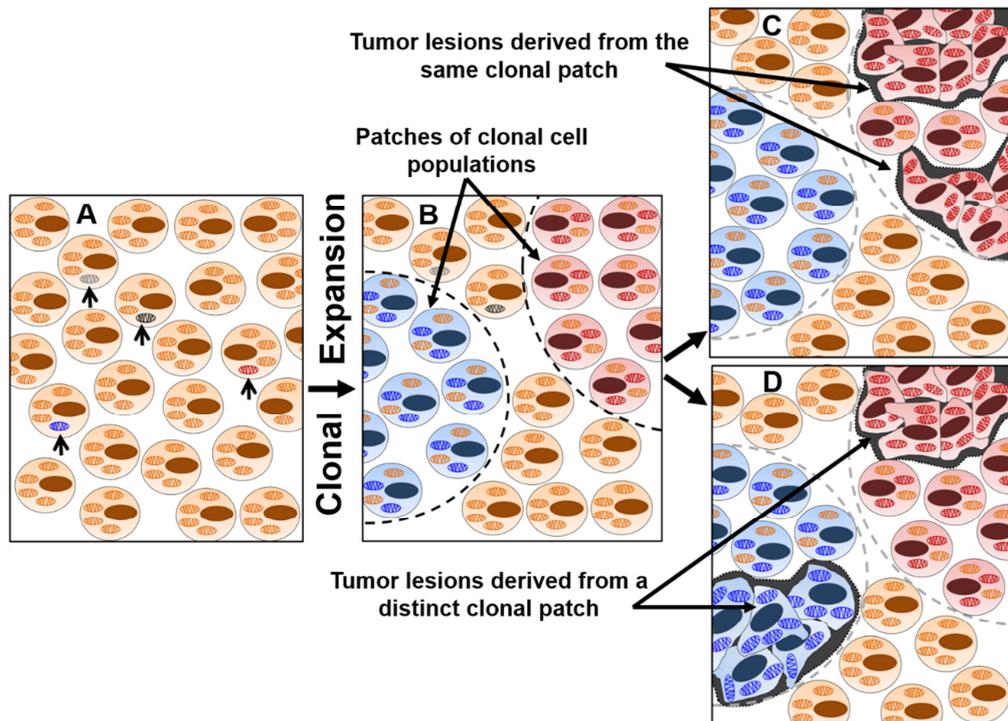
### **5.2.2 The mt-genome alterations as tools for tumour tracking analysis of HCC**

The mt-genome is a useful molecular tool with which to assess precise cell lineages and to track tumour history [47-49, 52]. The present study addressed the evolution of tumours by screening the mt-mutations in tumour/non-tumour nodules of an individual patient with multi-focal HCC. Previous studies have already alluded to the monoclonal nature of HCCs based on methylation patterns of the X chromosome genes [192-194]. However, these studies often ignore the distribution of X chromosome-inactivated cells. Because X-inactivation occurs relatively early in embryogenesis, the progeny of a single X-inactivated embryonic cell may be clustered together, forming a patch and sharing a common X-inactivation pattern. HCC nodules arising in the middle of a patch could be mono-phenotypic for an X-linked marker, but could be either monoclonally or polyclonally derived.

Furthermore, the polyclonality can be demonstrated only at the border of a patch; therefore the patch size is crucial in determining the chance of demonstrating polyclonality, which has rarely been mentioned in previous studies. Another limitation of these studies is that CpG methylation changes over time, especially in cancer, which is often associated with an increase or a decrease of the DNA-methylation pattern. Therefore, these studies might be not informative.

Using the mtDNA mutation pattern as a diagnostic tool, it was found in the present study that all HCC samples were monoclonal, because all HCC nodules and case-matching non-tumour lesions harboured identical mt-mutations, suggesting their single cell origin. Monoclonality of cirrhotic regenerative nodules of human liver have been characterized in the past [195]. Since the cirrhotic regenerative nodules represent pre-neoplastic stages of HCC development, it is likely that HCC lesions are also monoclonal.

Remarkably, mtDNA constantly undergoes mutation, with the loss or expansion of mutated mtDNA copies over several generations [45, 106]. Dominant mt-mutation during tumour development leads to the evolution of a homoplasmic status in clonal cell population [44, 45, 181]. Clonal homoplasmic expansion develops, as previously shown, by crypt fission, forming large tumour patches with identical mt-mutations [62, 196]. The present data demonstrate that tumour nodules arise within the clonal patch harbouring identical mt-mutations (Figure 24).



**Figure 24: Clonal mt-mutation expansion of HCC development.** A stochastic mt-mutation (open arrows) arises at the germline or somatic level (A) and confers a cellular growth advantage. It is likely to become dominant, leading to clonal patch formation (dashed boundaries) (B). Monoclonal tumour nodules, arising from a single patch, carry an identical mt-mutation pattern of homoplasmic and heteroplasmic mutations (red mitochondria) (C). Tumour nodules, which arise from different clonal patches, contain a distinct pattern of homoplasmic and heteroplasmic mt-mutations (blue mitochondria) and are polyclonal (D).

## 6 References

1. Okuda, K., *Hepatocellular carcinoma*. J Hepatol, 2000. **32**(1 Suppl): p. 225-37.
2. Parkin, D.M., et al., *Estimating the world cancer burden: Globocan 2000*. Int J Cancer, 2001. **94**(2): p. 153-6.
3. Ferlay, J., et al., *Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012*. Int J Cancer, 2015. **136**(5): p. E359-86.
4. Venook, A.P., et al., *The incidence and epidemiology of hepatocellular carcinoma: a global and regional perspective*. Oncologist, 2010. **15** Suppl 4: p. 5-13.
5. Nordenstedt, H., D.L. White, and H.B. El-Serag, *The changing pattern of epidemiology in hepatocellular carcinoma*. Dig Liver Dis, 2010. **42** Suppl 3: p. S206-14.
6. Badvie, S., *Hepatocellular carcinoma*. Postgrad Med J, 2000. **76**(891): p. 4-11.
7. Llovet, J.M., *Updated treatment approach to hepatocellular carcinoma*. J Gastroenterol, 2005. **40**(3): p. 225-35.
8. Di Bisceglie, A.M., *Hepatitis B and hepatocellular carcinoma*. Hepatology, 2009. **49**(5 Suppl): p. S56-60.
9. Kew, M.C., et al., *Mutant woodchuck hepatitis virus genomes from virions resemble rearranged hepadnaviral integrants in hepatocellular carcinoma*. Proc Natl Acad Sci U S A, 1993. **90**(21): p. 10211-5.
10. Muroyama, R., et al., *Nucleotide change of codon 38 in the X gene of hepatitis B virus genotype C is associated with an increased risk of hepatocellular carcinoma*. J Hepatol, 2006. **45**(6): p. 805-12.
11. Kato, N., et al., *Hepatitis C virus nonstructural region 5A protein is a potent transcriptional activator*. J Virol, 1997. **71**(11): p. 8856-9.
12. Moriya, K., et al., *The core protein of hepatitis C virus induces hepatocellular carcinoma in transgenic mice*. Nat Med, 1998. **4**(9): p. 1065-7.
13. Turner, P.C., et al., *The role of aflatoxins and hepatitis viruses in the etiopathogenesis of hepatocellular carcinoma: A basis for primary prevention in Guinea-Conakry, West Africa*. J Gastroenterol Hepatol, 2002. **17** Suppl: p. S441-8.
14. El-Serag, H.B. and K.L. Rudolph, *Hepatocellular carcinoma: epidemiology and molecular carcinogenesis*. Gastroenterology, 2007. **132**(7): p. 2557-76.
15. Block, T.M., et al., *Molecular viral oncology of hepatocellular carcinoma*. Oncogene, 2003. **22**(33): p. 5093-107.
16. Caldwell, S.H., et al., *Obesity and hepatocellular carcinoma*. Gastroenterology, 2004. **127**(5 Suppl 1): p. S97-103.
17. Kowdley, K.V., *Iron, hemochromatosis, and hepatocellular carcinoma*. Gastroenterology, 2004. **127**(5 Suppl 1): p. S79-86.
18. Guichard, C., et al., *Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma*. Nat Genet, 2012. **44**(6): p. 694-8.
19. Gumbiner, B.M., *Signal transduction of beta-catenin*. Curr Opin Cell Biol, 1995. **7**(5): p. 634-40.
20. Isaeva, A.V., et al., *[beta-Catenin: Structure, Function and Role in Malignant Transformation of Epithelial Cells]*. Vestn Ross Akad Med Nauk, 2015(4): p. 475-83.
21. Vilchez, V., et al., *Targeting Wnt/beta-catenin pathway in hepatocellular carcinoma treatment*. World J Gastroenterol, 2016. **22**(2): p. 823-32.
22. Miyoshi, Y., et al., *Activation of the beta-catenin gene in primary hepatocellular carcinomas by somatic alterations involving exon 3*. Cancer Res, 1998. **58**(12): p. 2524-7.
23. Farazi, P.A. and R.A. DePinho, *Hepatocellular carcinoma pathogenesis: from genes to environment*. Nat Rev Cancer, 2006. **6**(9): p. 674-87.
24. Waisberg, J. and G.T. Saba, *Wnt/-beta-catenin pathway signaling in human hepatocellular carcinoma*. World J Hepatol, 2015. **7**(26): p. 2631-5.
25. Lin, C.P., et al., *Targeting c-Myc as a novel approach for hepatocellular carcinoma*. World J Hepatol, 2010. **2**(1): p. 16-20.

26. Yuen, M.F., et al., *Expression of c-Myc, c-Fos, and c-jun in hepatocellular carcinoma*. Cancer, 2001. **91**(1): p. 106-12.
27. Pollak, M.N., E.S. Schernhammer, and S.E. Hankinson, *Insulin-like growth factors and neoplasia*. Nat Rev Cancer, 2004. **4**(7): p. 505-18.
28. Alexia, C., et al., *An evaluation of the role of insulin-like growth factors (IGF) and of type-I IGF receptor signalling in hepatocarcinogenesis and in the resistance of hepatocarcinoma cells against drug-induced apoptosis*. Biochem Pharmacol, 2004. **68**(6): p. 1003-15.
29. Li, W., et al., *Activation of Akt-mTOR-p70S6K pathway in angiogenesis in hepatocellular carcinoma*. Oncol Rep, 2008. **20**(4): p. 713-9.
30. Avila, M.A., et al., *New therapies for hepatocellular carcinoma*. Oncogene, 2006. **25**(27): p. 3866-84.
31. Roberts, L.R. and G.J. Gores, *Hepatocellular carcinoma: molecular pathways and new therapeutic targets*. Semin Liver Dis, 2005. **25**(2): p. 212-25.
32. Gollob, J.A., et al., *Role of Raf kinase in cancer: therapeutic potential of targeting the Raf/MEK/ERK signal transduction pathway*. Semin Oncol, 2006. **33**(4): p. 392-406.
33. Scharovsky, O.G., et al., *Inhibition of ras oncogene: a novel approach to antineoplastic therapy*. J Biomed Sci, 2000. **7**(4): p. 292-8.
34. Mas, V.R., et al., *Angiogenesis soluble factors as hepatocellular carcinoma noninvasive markers for monitoring hepatitis C virus cirrhotic patients awaiting liver transplantation*. Transplantation, 2007. **84**(10): p. 1262-71.
35. Thorgeirsson, S.S. and J.W. Grisham, *Molecular pathogenesis of human hepatocellular carcinoma*. Nat Genet, 2002. **31**(4): p. 339-46.
36. Saito, Y., et al., *Expression of mRNA for DNA methyltransferases and methyl-CpG-binding proteins and DNA methylation status on CpG islands and pericentromeric satellite regions during human hepatocarcinogenesis*. Hepatology, 2001. **33**(3): p. 561-8.
37. Kondo, Y., et al., *Genetic instability and aberrant DNA methylation in chronic hepatitis and cirrhosis--A comprehensive study of loss of heterozygosity and microsatellite instability at 39 loci and DNA hypermethylation on 8 CpG islands in microdissected specimens from patients with hepatocellular carcinoma*. Hepatology, 2000. **32**(5): p. 970-9.
38. Maggioni, M., et al., *Molecular changes in hepatocellular dysplastic nodules on microdissected liver biopsies*. Hepatology, 2000. **32**(5): p. 942-6.
39. Buendia, M.A., *Genetics of hepatocellular carcinoma*. Semin Cancer Biol, 2000. **10**(3): p. 185-200.
40. Copeland, W.C., et al., *Mitochondrial DNA alterations in cancer*. Cancer Invest, 2002. **20**(4): p. 557-69.
41. Modica-Napolitano, J.S. and K.K. Singh, *Mitochondrial dysfunction in cancer*. Mitochondrion, 2004. **4**(5-6): p. 755-62.
42. Pakendorf, B. and M. Stoneking, *Mitochondrial DNA and human evolution*. Annu Rev Genomics Hum Genet, 2005. **6**: p. 165-83.
43. Lightowlers, R.N., et al., *Mammalian mitochondrial genetics: heredity, heteroplasmy and disease*. Trends Genet, 1997. **13**(11): p. 450-5.
44. Polyak, K., et al., *Somatic mutations of the mitochondrial genome in human colorectal tumours*. Nat Genet, 1998. **20**(3): p. 291-3.
45. Taylor, R.W., et al., *Mitochondrial DNA mutations in human colonic crypt stem cells*. J Clin Invest, 2003. **112**(9): p. 1351-60.
46. Vander Heiden, M.G., L.C. Cantley, and C.B. Thompson, *Understanding the Warburg effect: the metabolic requirements of cell proliferation*. Science, 2009. **324**(5930): p. 1029-33.
47. Nomoto, S., et al., *Mitochondrial D-loop mutations as clonal markers in multicentric hepatocellular carcinoma and plasma*. Clin Cancer Res, 2002. **8**(2): p. 481-7.
48. Masuda, S., et al., *Analysis of gene alterations of mitochondrial DNA D-loop regions to determine breast cancer clonality*. Br J Cancer, 2012. **107**(12): p. 2016-23.

49. Gaisa, N.T., et al., *Clonal architecture of human prostatic epithelium in benign and malignant conditions*. J Pathol, 2011. **225**(2): p. 172-80.
50. Geurts-Giele, W.R., et al., *Mitochondrial D310 mutation as clonal marker for solid tumors*. Virchows Arch, 2015.
51. Ha, P.K., et al., *Mitochondrial C-tract alteration in premalignant lesions of the head and neck: a marker for progression and clonal proliferation*. Clin Cancer Res, 2002. **8**(7): p. 2260-5.
52. Kirches, E., et al., *Mitochondrial DNA as a clonal tumor cell marker: gliomatosis cerebri*. J Neurooncol, 2003. **61**(1): p. 1-5.
53. Kawanishi, S., et al., *Oxidative and nitrative DNA damage in animals and patients with inflammatory diseases in relation to inflammation-related carcinogenesis*. Biol Chem, 2006. **387**(4): p. 365-72.
54. Klaunig, J.E. and L.M. Kamendulis, *The role of oxidative stress in carcinogenesis*. Annu Rev Pharmacol Toxicol, 2004. **44**: p. 239-67.
55. Nishikawa, M., et al., *Somatic mutation of mitochondrial DNA in cancerous and noncancerous liver tissue in individuals with hepatocellular carcinoma*. Cancer Res, 2001. **61**(5): p. 1843-5.
56. Yamada, S., et al., *Correlation between copy number of mitochondrial DNA and clinico-pathologic parameters of hepatocellular carcinoma*. Eur J Surg Oncol, 2006. **32**(3): p. 303-7.
57. Park, S.Y., et al., *Resistance of mitochondrial DNA-depleted cells against cell death: role of mitochondrial superoxide dismutase*. J Biol Chem, 2004. **279**(9): p. 7512-20.
58. Li, C.H., et al., *Chloramphenicol-induced mitochondrial stress increases p21 expression and prevents cell apoptosis through a p21-dependent pathway*. J Biol Chem, 2005. **280**(28): p. 26193-9.
59. Chang, C.J., et al., *Mitochondrial dysfunction-induced amphiregulin upregulation mediates chemo-resistance and cell migration in HepG2 cells*. Cell Mol Life Sci, 2009. **66**(10): p. 1755-65.
60. Walther, V. and M.R. Alison, *Cell lineage tracing in human epithelial tissues using mitochondrial DNA mutations as clonal markers*. Wiley Interdiscip Rev Dev Biol, 2015.
61. Nicholson, A.M., et al., *Barrett's metaplasia glands are clonal, contain multiple stem cells and share a common squamous progenitor*. Gut, 2012. **61**(10): p. 1380-9.
62. Gutierrez-Gonzalez, L., et al., *The clonal origins of dysplasia from intestinal metaplasia in the human stomach*. Gastroenterology, 2011. **140**(4): p. 1251-1260 e1-6.
63. Yates, L.R. and P.J. Campbell, *Evolution of the cancer genome*. Nat Rev Genet, 2012. **13**(11): p. 795-806.
64. Pons, F., M. Varela, and J.M. Llovet, *Staging systems in hepatocellular carcinoma*. HPB (Oxford), 2005. **7**(1): p. 35-41.
65. Lavanchy, D., *Hepatitis B virus epidemiology, disease burden, treatment, and current and emerging prevention and control measures*. J Viral Hepat, 2004. **11**(2): p. 97-107.
66. Alric, L. and D. Bonnet, *Grazoprevir + elbasvir for the treatment of hepatitis C virus infection*. Expert Opin Pharmacother, 2016. **17**(5): p. 735-42.
67. Maan, R. and A.J. van der Meer, *Recent advances in managing chronic HCV infection: focus on therapy in patients with severe liver disease*. F1000Res, 2016. **5**.
68. Lok, A.S., et al., *Des-gamma-carboxy prothrombin and alpha-fetoprotein as biomarkers for the early detection of hepatocellular carcinoma*. Gastroenterology, 2010. **138**(2): p. 493-502.
69. *Pathologic diagnosis of early hepatocellular carcinoma: a report of the international consensus group for hepatocellular neoplasia*. Hepatology, 2009. **49**(2): p. 658-64.
70. International Working, P., *Terminology of nodular hepatocellular lesions*. Hepatology, 1995. **22**(3): p. 983-93.

71. Kojiro, M. and O. Nakashima, *Histopathologic evaluation of hepatocellular carcinoma with special reference to small early stage tumors*. Semin Liver Dis, 1999. **19**(3): p. 287-96.
72. International Consensus Group for Hepatocellular Neoplasia The International Consensus Group for Hepatocellular, N., *Pathologic diagnosis of early hepatocellular carcinoma: a report of the international consensus group for hepatocellular neoplasia*. Hepatology, 2009. **49**(2): p. 658-64.
73. Edmondson, H.A. and P.E. Steiner, *Primary carcinoma of the liver: a study of 100 cases among 48,900 necropsies*. Cancer, 1954. **7**(3): p. 462-503.
74. Paradis, V., *Histopathology of hepatocellular carcinoma*. Recent Results Cancer Res, 2013. **190**: p. 21-32.
75. Belghiti, J. and R. Kianmanesh, *Surgical treatment of hepatocellular carcinoma*. HPB (Oxford), 2005. **7**(1): p. 42-9.
76. Li, H.Y., et al., *Salvage liver transplantation in the treatment of hepatocellular carcinoma: a meta-analysis*. World J Gastroenterol, 2012. **18**(19): p. 2415-22.
77. Mazzaferro, V., et al., *Liver transplantation for the treatment of small hepatocellular carcinomas in patients with cirrhosis*. N Engl J Med, 1996. **334**(11): p. 693-9.
78. Belghiti, J., et al., *Resection prior to liver transplantation for hepatocellular carcinoma*. Ann Surg, 2003. **238**(6): p. 885-92; discussion 892-3.
79. Poon, R.T., et al., *Long-term survival and pattern of recurrence after resection of small hepatocellular carcinoma in patients with preserved liver function: implications for a strategy of salvage transplantation*. Ann Surg, 2002. **235**(3): p. 373-82.
80. Lau, W.Y., et al., *Percutaneous local ablative therapy for hepatocellular carcinoma: a review and look into the future*. Ann Surg, 2003. **237**(2): p. 171-9.
81. Wilhelm, S., et al., *Discovery and development of sorafenib: a multikinase inhibitor for treating cancer*. Nat Rev Drug Discov, 2006. **5**(10): p. 835-44.
82. Wilhelm, S.M., et al., *BAY 43-9006 exhibits broad spectrum oral antitumor activity and targets the RAF/MEK/ERK pathway and receptor tyrosine kinases involved in tumor progression and angiogenesis*. Cancer Res, 2004. **64**(19): p. 7099-109.
83. Llovet, J.M., et al., *Sorafenib in advanced hepatocellular carcinoma*. N Engl J Med, 2008. **359**(4): p. 378-90.
84. Irwin, J.A., et al., *Development and expansion of high-quality control region databases to improve forensic mtDNA evidence interpretation*. Forensic Sci Int Genet, 2007. **1**(2): p. 154-7.
85. Lyons, E.A., et al., *A high-throughput Sanger strategy for human mitochondrial genome sequencing*. BMC Genomics, 2013. **14**: p. 881.
86. Levin, B.C., et al., *Comparison of the complete mtDNA genome sequences of human cell lines--HL-60 and GM10742A--from individuals with pro-myelocytic leukemia and leber hereditary optic neuropathy, respectively, and the inclusion of HL-60 in the NIST human mitochondrial DNA standard reference material--SRM 2392-I*. Mitochondrion, 2003. **2**(6): p. 387-400.
87. Levin, B.C., H. Cheng, and D.J. Reeder, *A human mitochondrial DNA standard reference material for quality control in forensic identification, medical diagnosis, and mutation detection*. Genomics, 1999. **55**(2): p. 135-46.
88. Parsons, T.J., *Mitochondrial DNA genome sequencing and SNP assay development for increased power of discrimination*, in *NIJ Grant Final Report* 2006.
89. Ross, J.M., *Visualization of mitochondrial respiratory function using cytochrome c oxidase/succinate dehydrogenase (COX/SDH) double-labeling histochemistry*. J Vis Exp, 2011(57): p. e3266.
90. Forbes, S.A., et al., *The Catalogue of Somatic Mutations in Cancer (COSMIC)*. Curr Protoc Hum Genet, 2008. **Chapter 10**: p. Unit 10 11.
91. Wu, C., et al., *BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources*. Genome Biol, 2009. **10**(11): p. R130.
92. UniProt Consortium, *Activities at the Universal Protein Resource (UniProt)*. Nucleic Acids Res, 2014. **42**(Database issue): p. D191-8.

93. Rice, P., I. Longden, and A. Bleasby, *EMBOSS: the European Molecular Biology Open Software Suite*. Trends Genet, 2000. **16**(6): p. 276-7.
94. Fujimoto, A., et al., *Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators*. Nat Genet, 2012. **44**(7): p. 760-4.
95. Huang, J., et al., *Exome sequencing of hepatitis B virus-associated hepatocellular carcinoma*. Nat Genet, 2012. **44**(10): p. 1117-21.
96. Li, M., et al., *Inactivating mutations of the chromatin remodeling gene ARID2 in hepatocellular carcinoma*. Nat Genet, 2011. **43**(9): p. 828-9.
97. Peifer, M., et al., *Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer*. Nat Genet, 2012. **44**(10): p. 1104-10.
98. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform*. Bioinformatics, 2009. **25**(14): p. 1754-60.
99. Reva, B., Y. Antipin, and C. Sander, *Predicting the functional impact of protein mutations: application to cancer genomics*. Nucleic Acids Res, 2011. **39**(17): p. e118.
100. Wang, K., M. Li, and H. Hakonarson, *ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data*. Nucleic Acids Res, 2010. **38**(16): p. e164.
101. Kumar, P., S. Henikoff, and P.C. Ng, *Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm*. Nat Protoc, 2009. **4**(7): p. 1073-81.
102. Adzhubei, I.A., et al., *A method and server for predicting damaging missense mutations*. Nat Methods, 2010. **7**(4): p. 248-9.
103. Schwarz, J.M., et al., *MutationTaster evaluates disease-causing potential of sequence alterations*. Nat Methods, 2010. **7**(8): p. 575-6.
104. Amer, W.T., C.; Vassella, E.; Koitzsch, U.; Meinrath, J.; Eischeid, H.; Arens, A.; Huang, J.; Adam, A.; Scheel, A.; Buettner, R.; Schaefer, S.C.; Odenthal, M., *Evolution analysis of heterogeneous non-small cell lung carcinoma by ultra-deep sequencing of the mitochondrial genome*. The Journal of Pathology: Clinical Research, 2016.
105. Fliss, M.S., et al., *Facile detection of mitochondrial DNA mutations in tumors and bodily fluids*. Science, 2000. **287**(5460): p. 2017-9.
106. Collier, H.A., et al., *High frequency of homoplasmic mitochondrial DNA mutations in human tumors can be explained without selection*. Nat Genet, 2001. **28**(2): p. 147-50.
107. He, Y., et al., *Heteroplasmic mitochondrial DNA mutations in normal and tumour cells*. Nature, 2010. **464**(7288): p. 610-4.
108. Hertl, M. and A.B. Cosimi, *Liver transplantation for malignancy*. Oncologist, 2005. **10**(4): p. 269-81.
109. Nikiforova, M.N., et al., *Targeted next-generation sequencing panel (ThyroSeq) for detection of mutations in thyroid cancer*. J Clin Endocrinol Metab, 2013. **98**(11): p. E1852-60.
110. MacArthur, D.G., et al., *A systematic survey of loss-of-function variants in human protein-coding genes*. Science, 2012. **335**(6070): p. 823-8.
111. Goldfeder, R.L., et al., *Medical implications of technical accuracy in genome sequencing*. Genome Med, 2016. **8**(1): p. 24.
112. Park, J.Y., et al., *Clinical exome performance for reporting secondary genetic findings*. Clin Chem, 2015. **61**(1): p. 213-20.
113. Sims, D., et al., *Sequencing depth and coverage: key considerations in genomic analyses*. Nat Rev Genet, 2014. **15**(2): p. 121-32.
114. Suzuki, T., et al., *Beta-catenin expression in hepatocellular carcinoma: a possible participation of beta-catenin in the dedifferentiation process*. J Gastroenterol Hepatol, 2002. **17**(9): p. 994-1000.
115. Cieply, B., et al., *Unique phenotype of hepatocellular cancers with exon-3 mutations in beta-catenin gene*. Hepatology, 2009. **49**(3): p. 821-31.

116. de La Coste, A., et al., *Somatic mutations of the beta-catenin gene are frequent in mouse and human hepatocellular carcinomas*. Proc Natl Acad Sci U S A, 1998. **95**(15): p. 8847-51.
117. Wands, J.R. and M. Kim, *WNT/beta-catenin signaling and hepatocellular carcinoma*. Hepatology, 2014. **60**(2): p. 452-4.
118. Giles, R.H., J.H. van Es, and H. Clevers, *Caught up in a Wnt storm: Wnt signaling in cancer*. Biochim Biophys Acta, 2003. **1653**(1): p. 1-24.
119. Suarez, M.I., et al., *Wnt/beta-catenin signaling pathway in hepatocellular carcinomas cases from Colombia*. Ann Hepatol, 2015. **14**(1): p. 64-74.
120. Thiagalingam, S., et al., *Loss of heterozygosity as a predictor to map tumor suppressor genes in cancer: molecular basis of its occurrence*. Curr Opin Oncol, 2002. **14**(1): p. 65-72.
121. Zucman-Rossi, J., et al., *Differential effects of inactivated Axin1 and activated beta-catenin mutations in human hepatocellular carcinomas*. Oncogene, 2007. **26**(5): p. 774-80.
122. Furuhashi, M., et al., *Axin facilitates Smad3 activation in the transforming growth factor beta signaling pathway*. Mol Cell Biol, 2001. **21**(15): p. 5132-41.
123. Zhang, Y., et al., *Dimerization choices control the ability of axin and dishevelled to activate c-Jun N-terminal kinase/stress-activated protein kinase*. J Biol Chem, 2000. **275**(32): p. 25008-14.
124. Satoh, S., et al., *AXIN1 mutations in hepatocellular carcinomas, and growth suppression in cancer cells by virus-mediated transfer of AXIN1*. Nat Genet, 2000. **24**(3): p. 245-50.
125. Petitjean, A., et al., *Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: lessons from recent developments in the IARC TP53 database*. Hum Mutat, 2007. **28**(6): p. 622-9.
126. Vogelstein, B., D. Lane, and A.J. Levine, *Surfing the p53 network*. Nature, 2000. **408**(6810): p. 307-10.
127. Petitjean, A., et al., *TP53 mutations in human cancers: functional selection and impact on cancer prognosis and outcomes*. Oncogene, 2007. **26**(15): p. 2157-65.
128. Bossi, G., et al., *Mutant p53 gain of function: reduction of tumor malignancy of human cancer cell lines through abrogation of mutant p53 expression*. Oncogene, 2006. **25**(2): p. 304-9.
129. Liu, K., S. Ling, and W.C. Lin, *TopBP1 mediates mutant p53 gain of function through NF-Y and p63/p73*. Mol Cell Biol, 2011. **31**(22): p. 4464-81.
130. Peart, M.J. and C. Prives, *Mutant p53 gain of function: the NF-Y connection*. Cancer Cell, 2006. **10**(3): p. 173-4.
131. Chen, G.G., et al., *Mutation of p53 in recurrent hepatocellular carcinoma and its association with the expression of ZBP-89*. Am J Pathol, 2003. **162**(6): p. 1823-9.
132. Olivier, M., et al., *The clinical value of somatic TP53 gene mutations in 1,794 patients with breast cancer*. Clin Cancer Res, 2006. **12**(4): p. 1157-67.
133. Villar, S., et al., *Aflatoxin-induced TP53 R249S mutation in hepatocellular carcinoma in Thailand: association with tumors developing in the absence of liver cirrhosis*. PLoS One, 2012. **7**(6): p. e37707.
134. Chappuis, P.O., et al., *Prognostic significance of p53 mutation in breast cancer: frequent detection of non-missense mutations by yeast functional assay*. Int J Cancer, 1999. **84**(6): p. 587-93.
135. Lomax, M.E., et al., *Two functional assays employed to detect an unusual mutation in the oligomerisation domain of p53 in a Li-Fraumeni like family*. Oncogene, 1997. **14**(15): p. 1869-74.
136. Cohen, C. and P.B. DeRose, *Immunohistochemical p53 in hepatocellular carcinoma and liver cell dysplasia*. Mod Pathol, 1994. **7**(5): p. 536-9.
137. Murakami, Y., et al., *Aberrations of the tumor suppressor p53 and retinoblastoma genes in human hepatocellular carcinomas*. Cancer Res, 1991. **51**(20): p. 5520-5.
138. Oda, T., et al., *p53 gene mutation spectrum in hepatocellular carcinoma*. Cancer Res, 1992. **52**(22): p. 6358-64.

139. Kubota, Y., et al., *Tumor suppressor gene p53 mutations in human prostate cancer*. Prostate, 1995. **27**(1): p. 18-24.
140. Schlomm, T., et al., *Clinical significance of p53 alterations in surgically treated prostate cancers*. Mod Pathol, 2008. **21**(11): p. 1371-8.
141. Uchida, T., et al., *p53 mutations and prognosis in bladder tumors*. J Urol, 1995. **153**(4): p. 1097-104.
142. Hruban, R.H., et al., *Progression model for pancreatic cancer*. Clin Cancer Res, 2000. **6**(8): p. 2969-72.
143. Amato, E., et al., *Targeted next-generation sequencing of cancer genes dissects the molecular profiles of intraductal papillary neoplasms of the pancreas*. J Pathol, 2014. **233**(3): p. 217-27.
144. Bakos, R.M., et al., *The CDKN2A p.A148T variant is associated with cutaneous melanoma in Southern Brazil*. Exp Dermatol, 2011. **20**(11): p. 890-3.
145. Lilischkis, R., et al., *Cancer-associated mis-sense and deletion mutations impair p16INK4 CDK inhibitory activity*. Int J Cancer, 1996. **66**(2): p. 249-54.
146. Ranade, K., et al., *Mutations associated with familial melanoma impair p16INK4 function*. Nat Genet, 1995. **10**(1): p. 114-6.
147. Pastorino, L., et al., *CDKN2A mutations and MC1R variants in Italian patients with single or multiple primary melanoma*. Pigment Cell Melanoma Res, 2008. **21**(6): p. 700-9.
148. Puig, S., et al., *Role of the CDKN2A locus in patients with multiple primary melanomas*. J Clin Oncol, 2005. **23**(13): p. 3043-51.
149. Peschard, P. and M. Park, *From Tpr-Met to Met, tumorigenesis and tubes*. Oncogene, 2007. **26**(9): p. 1276-85.
150. Trusolino, L., A. Bertotti, and P.M. Comoglio, *MET signalling: principles and functions in development, organ regeneration and cancer*. Nat Rev Mol Cell Biol, 2010. **11**(12): p. 834-48.
151. Sierra, J.R., et al., *Tumor angiogenesis and progression are enhanced by Sema4D produced by tumor-associated macrophages*. J Exp Med, 2008. **205**(7): p. 1673-85.
152. Yi, S. and M.S. Tsao, *Activation of hepatocyte growth factor-met autocrine loop enhances tumorigenicity in a human lung adenocarcinoma cell line*. Neoplasia, 2000. **2**(3): p. 226-34.
153. Kaposi-Novak, P., et al., *Met-regulated expression signature defines a subset of human hepatocellular carcinomas with poor prognosis and aggressive phenotype*. J Clin Invest, 2006. **116**(6): p. 1582-95.
154. Yuen, S.T., et al., *Similarity of the phenotypic patterns associated with BRAF and KRAS mutations in colorectal neoplasia*. Cancer Res, 2002. **62**(22): p. 6451-5.
155. Ni, S., et al., *c-kit gene mutation and CD117 expression in human anorectal melanomas*. Hum Pathol, 2012. **43**(6): p. 801-7.
156. Guo, G., et al., *Whole-genome and whole-exome sequencing of bladder cancer identifies frequent alterations in genes involved in sister chromatid cohesion and segregation*. Nat Genet, 2013. **45**(12): p. 1459-63.
157. Colombino, M., et al., *BRAF and PIK3CA genes are somatically mutated in hepatocellular carcinoma among patients from South Italy*. Cell Death Dis, 2012. **3**: p. e259.
158. Jhunjhunwala, S., et al., *Diverse modes of genomic alteration in hepatocellular carcinoma*. Genome Biol, 2014. **15**(8): p. 436.
159. Kim, H., et al., *PIK3CA mutations in hepatocellular carcinoma in Korea*. Yonsei Med J, 2013. **54**(4): p. 883-7.
160. Schulze, K., et al., *Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets*. Nat Genet, 2015. **47**(5): p. 505-11.
161. Akagi, I., et al., *Overexpression of PIK3CA is associated with lymph node metastasis in esophageal squamous cell carcinoma*. Int J Oncol, 2009. **34**(3): p. 767-75.
162. Baohua, Y., et al., *Mutations of the PIK3CA gene in diffuse large B cell lymphoma*. Diagn Mol Pathol, 2008. **17**(3): p. 159-65.

163. Kang, S., et al., *Mutual exclusiveness between PIK3CA and KRAS mutations in endometrial carcinoma*. Int J Gynecol Cancer, 2008. **18**(6): p. 1339-43.
164. Levine, D.A., et al., *Frequent mutation of the PIK3CA gene in ovarian and breast cancers*. Clin Cancer Res, 2005. **11**(8): p. 2875-8.
165. Kang, S., A.G. Bader, and P.K. Vogt, *Phosphatidylinositol 3-kinase mutations identified in human cancer are oncogenic*. Proc Natl Acad Sci U S A, 2005. **102**(3): p. 802-7.
166. Tanaka, H., et al., *The selective class I PI3K inhibitor CH5132799 targets human cancers harboring oncogenic PIK3CA mutations*. Clin Cancer Res, 2011. **17**(10): p. 3272-81.
167. Maehama, T. and J.E. Dixon, *The tumor suppressor, PTEN/MMAC1, dephosphorylates the lipid second messenger, phosphatidylinositol 3,4,5-trisphosphate*. J Biol Chem, 1998. **273**(22): p. 13375-8.
168. Sansal, I. and W.R. Sellers, *The biology and clinical relevance of the PTEN tumor suppressor pathway*. J Clin Oncol, 2004. **22**(14): p. 2954-63.
169. Hu, T.H., et al., *Expression and prognostic role of tumor suppressor gene PTEN/MMAC1/TEP1 in hepatocellular carcinoma*. Cancer, 2003. **97**(8): p. 1929-40.
170. Aii, S., et al., *Overexpression of matrix metalloproteinase 9 gene in hepatocellular carcinoma with invasive potential*. Hepatology, 1996. **24**(2): p. 316-22.
171. El Samanoudy, A., et al., *Matrix metalloproteinase-9 gene polymorphism in hepatocellular carcinoma patients with hepatitis B and C viruses*. Genet Mol Res, 2014. **13**(3): p. 8025-34.
172. Bluteau, O., et al., *Bi-allelic inactivation of TCF1 in hepatic adenomas*. Nat Genet, 2002. **32**(2): p. 312-5.
173. Wu, J.N. and C.W. Roberts, *ARID1A mutations in cancer: another epigenetic tumor suppressor?* Cancer Discov, 2013. **3**(1): p. 35-43.
174. Manceau, G., et al., *Recurrent inactivating mutations of ARID2 in non-small cell lung carcinoma*. Int J Cancer, 2013. **132**(9): p. 2217-21.
175. Zhao, H., et al., *ARID2: a new tumor suppressor gene in hepatocellular carcinoma*. Oncotarget, 2011. **2**(11): p. 886-91.
176. He, F., et al., *Decreased expression of ARID1A associates with poor prognosis and promotes metastases of hepatocellular carcinoma*. J Exp Clin Cancer Res, 2015. **34**: p. 47.
177. Yin, P.H., et al., *Somatic mutations of mitochondrial genome in hepatocellular carcinoma*. Mitochondrion, 2010. **10**(2): p. 174-82.
178. Gillio-Tos, A., et al., *Efficient DNA extraction from 25-year-old paraffin-embedded tissues: study of 365 samples*. Pathology, 2007. **39**(3): p. 345-8.
179. Zhang, W., H. Cui, and L.J. Wong, *Comprehensive one-step molecular analyses of mitochondrial genome by massively parallel sequencing*. Clin Chem, 2012. **58**(9): p. 1322-31.
180. Goios, A., et al., *Specificity of mtDNA-directed PCR-influence of Nuclear MTDNA insertion (NUMT) contamination in routine samples and techniques*. Int J Legal Med, 2008. **122**(4): p. 341-5.
181. Khrapko, K., et al., *Cell-by-cell scanning of whole mitochondrial genomes in aged human heart reveals a significant fraction of myocytes with clonally expanded deletions*. Nucleic Acids Res, 1999. **27**(11): p. 2434-41.
182. Lee, H.C. and Y.H. Wei, *Oxidative stress, mitochondrial DNA mutation, and apoptosis in aging*. Exp Biol Med (Maywood), 2007. **232**(5): p. 592-606.
183. Cha, M.Y., D.K. Kim, and I. Mook-Jung, *The role of mitochondrial DNA mutation on neurodegenerative diseases*. Exp Mol Med, 2015. **47**: p. e150.
184. Cui, H., Y. Kong, and H. Zhang, *Oxidative stress, mitochondrial dysfunction, and aging*. J Signal Transduct, 2012. **2012**: p. 646354.
185. Ma, Y.S., et al., *Response to the increase of oxidative stress and mutation of mitochondrial DNA in aging*. Biochim Biophys Acta, 2009. **1790**(10): p. 1021-9.
186. Salk, J.J. and M.S. Horwitz, *Passenger mutations as a marker of clonal cell lineages in emerging neoplasia*. Semin Cancer Biol, 2010. **20**(5): p. 294-303.

187. Chinnery, P.F., et al., *Mitochondrial DNA analysis: polymorphisms and pathogenicity*. J Med Genet, 1999. **36**(7): p. 505-10.
188. Jones, J.B., et al., *Detection of mitochondrial DNA mutations in pancreatic cancer offers a "mass"-ive advantage over detection of nuclear DNA mutations*. Cancer Res, 2001. **61**(4): p. 1299-304.
189. Wong, L.J., et al., *Detection of mitochondrial DNA mutations in the tumor and cerebrospinal fluid of medulloblastoma patients*. Cancer Res, 2003. **63**(14): p. 3866-71.
190. Evert, M. and F. Dombrowski, [*Hepatocellular carcinoma in the non-cirrhotic liver*]. Pathologe, 2008. **29**(1): p. 47-52.
191. Gaddikeri, S., et al., *Hepatocellular carcinoma in the noncirrhotic liver*. AJR Am J Roentgenol, 2014. **203**(1): p. W34-47.
192. Kawai, S., et al., *Clonality in hepatocellular carcinoma: analysis of methylation pattern of polymorphic X-chromosome-linked phosphoglycerate kinase gene in females*. Hepatology, 1995. **22**(1): p. 112-7.
193. Ochiai, T., et al., *Clonal expansion in evolution of chronic hepatitis to hepatocellular carcinoma as seen at an X-chromosome locus*. Hepatology, 2000. **31**(3): p. 615-21.
194. Okuda, T., et al., *Clonal analysis of hepatocellular carcinoma and dysplastic nodule by methylation pattern of X-chromosome-linked human androgen receptor gene*. Cancer Lett, 2001. **164**(1): p. 91-6.
195. Lin, W.R., et al., *The histogenesis of regenerative nodules in human liver cirrhosis*. Hepatology, 2010. **51**(3): p. 1017-26.
196. Greaves, L.C., et al., *Mitochondrial DNA mutations are established in human colonic stem cells, and mutated clones expand by crypt fission*. Proc Natl Acad Sci U S A, 2006. **103**(3): p. 714-9.

## 7 Appendix

### 7.1 Primer sequences of HCC gene panel

**Table S1:** Primer sets of complete coding sequence of HCC-related genes

Chr	Gene	Ion_AmpliSeq Fwd_Rev_Primer*	Amplicon Start	Amplicon Stop
chr1	ARID1A-1F	GTCTCAGCAGTCCCAGCAAA	27058041	27058167
	ARID1A-1R	GTTCTCATGACACTAACCCCAAG		
chr1	ARID1A-2F	AGAGATATTAGTGAGTTGCTAGTGAGT	27089404	27089577
	ARID1A-2R	GCTGTCCATGCATTTGACCTC		
chr1	ARID1A-3F	CGCCTCCTCGTCGTCTT	27023668	27023796
	ARID1A-3R	GACGTGAGCAGTTGGTTGAG		
chr1	ARID1A-4F	CAGACTCCCCAGTCAACCAG	27099017	27099188
	ARID1A-4R	AAGGAGTCCCATGCACTTATCTT		
chr1	ARID1A-5F	GGCCTTATGGCCCTAACATGG	27089641	27089808
	ARID1A-5R	CACTGACACCCTCTCTGCTT		
chr1	ARID1A-6F	GACATCTTTGCAGCTGCTGATTC	27097757	27097926
	ARID1A-6R	ACAGATCCTTGGCATATCCTGTT		
chr1	ARID1A-7F	ACATTCCAGAAGCGGAATTCCA	27099355	27099512
	ARID1A-7R	GAGCACAAGTTCAAATAGCAATCAGAT		
chr1	ARID1A-8F	GTGCCATTGCAGTGCAGAAG	27106937	27107087
	ARID1A-8R	CGCCGCATCATGTCCACA		
chr1	ARID1A-9F	GCTGTTGGACATCTCGGTATCA	27107160	27107310
	ARID1A-9R	AGTCAGTTTCTAAGTTCTCCACACAC		
chr1	ARID1A-10F	GCCCCCTCCATCTAACTACCA	27101416	27101567
	ARID1A-10R	CTGCCTTCTGCATTTTCATCCC		
chr1	ARID1A-11F	AGGGTTGCTAGAGCTCCTTGTA	27105519	27105662
	ARID1A-11R	GGAGCTGGACTAGACACCTTG		
chr1	ARID1A-12F	ATGCTCCGGGAAAACACCTT	27106552	27106682
	ARID1A-12R	TCAGCTGAAGGGCAAACCTG		
chr1	ARID1A-13F	AACCAGGCGGGAGATATACCT	27097564	27097725
	ARID1A-13R	CATTCAAAGGCATAGAGACACTGGATA		
chr1	ARID1A-14F	GGTCTTGGAACCCCTCAGCAA	27106740	27106872
	ARID1A-14R	CGGGTTCTTTGCGTCACTGA		
chr1	ARID1A-15F	GAATTCCCAGGCTTAGCCATGATA	27087823	27087953
	ARID1A-15R	GCAATGCTTGATTGGTTCATGGAAG		
chr1	ARID1A-16F	CACTGAGCATATCCAGACCCA	27105942	27106070
	ARID1A-16R	TGGTCTGTTGTCCCTGGTGTGA		
chr1	ARID1A-17F	CCCAGGATAAAGGATGGAGAGCAT	27088569	27088739
	ARID1A-17R	GTGTGTATCTGTCCTCCGGAAG		
chr1	ARID1A-18F	CTCATCATCAGTGCATAGCTTCTCA	27057553	27057724
	ARID1A-18R	GGTGAGGACTTTGCTGGTTGTA		
chr1	ARID1A-19F	CTCAACCACCACAGCTCCA	27057809	27057981
	ARID1A-19R	CTGCTGCTGGTAAGGAGACT		
chr1	ARID1A-20F	GCAGAGAGCAGATGGGCATTA	27102118	27102244
	ARID1A-20R	AGAAGATCCCAAACCTCTCAATCT		
chr1	ARID1A-21F	CCCCCTCAATGACCTCCAGT	27059207	27059348
	ARID1A-21R	CTAACTAGTTTGCAATGGTTTCTCTCT		
chr1	ARID1A-22F	GGGCTAGGAAATGTGGCGAT	27099920	27100051
	ARID1A-22R	CAGATTGAGCCCACTATAGCTTCAG		
chr1	ARID1A-23F	CCGCTGGGAAAGGAGCTG	27023096	27023254
	ARID1A-23R	CGGCTCCGTGAGGTTATTGTT		
chr1	ARID1A-24F	TGGTTGTAGCCATCTTGGCAT	27100737	27100911
	ARID1A-24R	CCCTGCCAGTGCTGTA		

Chr	Gene	Ion_AmpliSeq Fwd_Rev_Primer*	Amplicon Start	Amplicon Stop
chr1	ARID1A-25F	GGCGCCTCCTCACTCAG	27023293	27023427
	ARID1A-25R	CTTTGTTGTCCGCCATGTTGT		
chr1	ARID1A-26F	CCTTCCCCTCAGCAAGATGTAT	27100979	27101109
	ARID1A-26R	ACACGGTCTCGGCCAAAC		
chr1	ARID1A-27F	CATACAGGCATCAGCTGAGGTT	27101170	27101332
	ARID1A-27R	CCTCTGATCTGTGTGCAGCATT		
chr1	ARID1A-28F	ACCTTCCAGAAATCCAGTTCTTCTACT	27094272	27094446
	ARID1A-28R	AGAGGTCCAGAGGTTTCTACC		
chr1	ARID1A-29F	GGGATGTAAAGTTAACTCCAGCCA	27092981	27093135
	ARID1A-29R	ACTCTCTTTTCTTTTCTTCCCTATCAC		
chr1	ARID1A-30F	TGGTTCAATAGATGACCTCCCAT	27087355	27087485
	ARID1A-30R	GGGAGGTATGAGGAGAGAAAGG		
chr1	ARID1A-31F	CTTGACCGAGGATGGAGCTAA	27106152	27106305
	ARID1A-31R	GTCCAGAAGGGTACACAGTGG		
chr1	ARID1A-32F	GTCCAATACCATTGCAAGCCTGT	27106341	27106514
	ARID1A-32R	CTCACCCCTTGGTCCTGTTC		
chr1	ARID1A-33F	TGATGGGAAGTGGACCTCCT	27092747	27092899
	ARID1A-33R	TGACTCTTGAAGAAATCCCTAGTGAGAA		
chr1	ARID1A-34F	ATTCGGCGGGATATCACCTTC	27101621	27101775
	ARID1A-34R	GAGGGAACAATGGACAAGATGGAAA		
chr1	ARID1A-35F	GCCCCAGCAGAACTCTCA	27023485	27023612
	ARID1A-35R	CCGGAGTGCCACCTCTC		
chr1	ARID1A-36F	GAGGCCATCAAAGCTCAGGTT	27056097	27056257
	ARID1A-36R	GGGTACCCATGTCTTGCTG		
chr1	ARID1A-37F	AAGAGGAAGTAGTTGAAAATGATGAGGA	27105722	27105892
	ARID1A-37R	GCACACGCCCAAGCTTATCT		
chr1	ARID1A-38F	ATGCCTTCCAACCCAGACTC	27100132	27100302
	ARID1A-38R	GGAATCATGTCTATAGGAGGCAAAGAAA		
chr1	ARID1A-39F	CCTGAAGCTATAGTGGGCTCAA	27100026	27100176
	ARID1A-39R	GTAGCGGCTAGGAGAATACATCC		
chr1	ARID1A-40F	CCCAGAGTAAGAAGCTTTAACT	27094173	27094324
	ARID1A-40R	AACTTGGTGATCTTCTCATTGGTTGT		
chr1	ARID1A-41F	GACTCCTGCGTGTCTTTGTTA	27099229	27099398
	ARID1A-41R	TGATACCCAGGGTTTGAGTCA		
chr1	ARID1A-42F	GCTGGAGGCATAAACCCCAT	27089529	27089684
	ARID1A-42R	CCCAACCTGAGGTGGCATATTG		
chr1	ARID1A-43F	CCCACCCTTTGAGCCAACTA	27107049	27107206
	ARID1A-43R	GTGAAACCAATGAGTTCATCAACGG		
chr1	ARID1A-44F	CAAGTCTCCATTCTGCACTCT	27101524	27101661
	ARID1A-44R	GCTTCAACAGAGCCAGGTG		
chr1	ARID1A-45F	ACTCTGCCTCTCCCAACTGATA	27105402	27105565
	ARID1A-45R	CAATCAGGCATCGTCGGAAATATTC		
chr1	ARID1A-46F	GCACCACTAAGTATGAAAAGGAGGAG	27106468	27106594
	ARID1A-46R	CCGAGATGTTGGCGAGTGTA		
chr1	ARID1A-47F	CTGGACGGACTCCTACTG	27106642	27106782
	ARID1A-47R	ATTGTTGTCTGGATGCTGAGT		
chr1	ARID1A-48F	GCCAGCTCCTTGA AAAAGCAGTA	27097676	27097800
	ARID1A-48R	ATCTTGGGCTGGGACTTCTTG		
chr1	ARID1A-49F	TTGTATAGCACTATGGTGCCTT	27106828	27106978
	ARID1A-49R	CCAGGAGGTTGCCGATACTG		
chr1	ARID1A-50F	AAGAATGATCCATTTGTGGTGGACT	27105844	27105984
	ARID1A-50R	CAGCTCTGTCTTGCTCTCGAA		
chr1	ARID1A-51F	AGTCGGACAGCATCATGCATC	27087908	27088051
	ARID1A-51R	GGTCAAACAGCTCTCCAAAAGCTA		

Chr	Gene	Ion_AmpliSeq Fwd_Rev_Primer*	Amplicon Start	Amplicon Stop
chr1	ARID1A-52F	CAGCCTTATCTCCGCGTCA	27088697	27088850
	ARID1A-52R	CTCTCACCCGTATCTACTCCTAACT		
chr1	ARID1A-53F	TCAACAGGGCCAGACTCCATAT	27057681	27057849
	ARID1A-53R	GGAGTATGGAGGCTGAGAGGAC		
chr1	ARID1A-54F	GTGGATAGACGACATGGAGGTT	27102038	27102165
	ARID1A-54R	ATCATACAGCAGGATGTTGATGGTATC		
chr1	ARID1A-55F	CCCCCTACTCACAGCCACA	27057938	27058084
	ARID1A-55R	GGAAGCGCTGCTGGGAATA		
chr1	ARID1A-56F	CATGACAAATCTGCCTGCTGTG	27094403	27094559
	ARID1A-56R	AGCCAGTGAGTACCTAGAAAAGG		
chr1	ARID1A-57F	GAGAGACAGTCCATAACCCTTTC	27059076	27059250
	ARID1A-57R	GTTTCATATCTTCTTGCCCTCCCTT		
chr1	ARID1A-58F	TGTTACCCGCTTGCTTTTCTA	27099809	27099966
	ARID1A-58R	CCATAGGGATAGTGCTGTGCTG		
chr1	ARID1A-59F	GCCTTAGGAAGAACTTTCCTCAAAGAG	27098883	27099057
	ARID1A-59R	CCTCCTTCTGCCATGGAAGT		
chr1	ARID1A-60F	CAACAGGGACCTCCGTCAG	27056214	27056388
	ARID1A-60R	AACAAGGTCAAGTAATCACAATCA		
chr1	ARID1A-61F	CACTGCCACCCTAATCCT	27100253	27100427
	ARID1A-61R	AACCTGCATGCCAGTCCAT		
chr1	ARID1A-62F	CAGGTAAGTGCTAGTCATTCTCACT	27092855	27093029
	ARID1A-62R	CCATCTGCCTTGTGTTTCATTTTGG		
chr1	ARID1A-63F	CCGGACCTGAAGAACTCGAA	27023186	27023352
	ARID1A-63R	GGTTGCCCGAAGCCGTA		
chr1	ARID1A-64F	GGCGTGAACCGAACAGATGA	27101291	27101464
	ARID1A-64R	AATGTGATTCTGCATGCTTGGTG		
chr1	ARID1A-65F	GGAGATGTACAGCGTGCCA	27100876	27101021
	ARID1A-65R	AGGCATTGCCATACTGGTTGT		
chr1	ARID1A-66F	CCTACGCGCTGAGCTCC	27023577	27023701
	ARID1A-66R	CTGCTGAGCGAAGGACGAA		
chr1	ARID1A-67F	GCCCCAGAACCAATTTCCATT	27101067	27101211
	ARID1A-67R	CACATGGTGCCTTGCTGAG		
chr1	ARID1A-68F	TGTGTGATGTATTTGCTCTTGGTTGT	27087229	27087403
	ARID1A-68R	AGGACTCAGAGCTCCTTCTGTC		
chr1	ARID1A-69F	AGCAGAGTAATCCAGCTCAGTCT	27087441	27087613
	ARID1A-69R	CCCTAAAGCCCAGCACTTGA		
chr1	ARID1A-70F	CTACTGGATCCTGGGAGGTTCA	27105619	27105772
	ARID1A-70R	TGTCCTTGCTGAAAAGGCTATC		
chr1	ARID1A-71F	GGAAGCATGTGACAACAGCAGA	27106025	27106195
	ARID1A-71R	TCTCCTTGATGGCCTCTGAACT		
chr1	ARID1A-72F	CGAACCCACAGTAAGGATGAGA	27106260	27106385
	ARID1A-72R	AAGTCATTGCCTGGCACAAATG		
chr1	ARID1A-73F	ACAGCACTATTTGGCTCCAGTT	27092631	27092794
	ARID1A-73R	CAGCCATACTATTAATCCCTTGCCATA		
chr1	ARID1A-74F	GCCGTCTTCCACCAACAACAT	27023393	27023542
	ARID1A-74R	GTAGTAGGAGTTGTACTGGTGGTTG		
chr1	ARID1A-75F	CCCGGACCCCTCAGGTA	27024018	27024154
	ARID1A-75R	GCAGCTCCCCAGCAGTG		
chr12	ARID2-1F	CAGCAACTGAACATGGAGGAAAAG	46285598	46285762
	ARID2-1R	ACGTCAAAATCTTTAATGCAGTCAAAATGA		
chr12	ARID2-2F	CAACTCTGGAAGAAGCATAGCTCA	46231055	46231217
	ARID2-2R	CTTCAGTGAAAACCTCACCTCTCATCTTT		
chr12	ARID2-3F	ACGTATGCACTCTCCTATCAAATGAAAAG	46211567	46211691
	ARID2-3R	ACATTCAGCTTAAAACCTTACTGTGCTCA		

Chr	Gene	Ion_AmpliSeq Fwd_Rev_Primer*	Amplicon Start	Amplicon Stop
chr12	ARID2-4F	ACAGTCGCAGGAATTCCAAATAAAGTA	46246482	46246638
	ARID2-4R	GCTTTTGCTGAACAGCACTTTCTATTG		
chr12	ARID2-5F	AAATGGGAGAAAGTACAGTGACTCAAG	46245788	46245952
	ARID2-5R	ATATCTTGATGTCTATTTGTTCTGAGTT		
chr12	ARID2-6F	CATCCGACTAACAGCTGCCTTAATAT	46287442	46287611
	ARID2-6R	TTGATGAGAAAAACAATGAAGAAAATACAGTTA		
chr12	ARID2-7F	CCAAATCATACAGTGAAGAGAGTGGGA	46243373	46243545
	ARID2-7R	GCAGGAGATACATCAGGAACAGAAT		
chr12	ARID2-8F	AATGAGAAAACCTGGACAGAACTTCA	46254682	46254807
	ARID2-8R	GATGACAGAACTTAAATATAAAAGCATTCCAA		
chr12	ARID2-9F	GAAAAGGAAAACAAATATGTGGTGAGAG	46205115	46205289
	ARID2-9R	CAATAGGAAGCTGTGGCTTTGG		
chr12	ARID2-10F	TTAAAGCTGTGTTCTGTATTCAAGGG	46211367	46211538
	ARID2-10R	TCCAGATAACAGTGAAAGCACCAATTTA		
chr12	ARID2-11F	AAGTTCGTGACCTCATTCTGACA	46230400	46230569
	ARID2-11R	GGTGGATGAAATAAAGACTCCCAA		
chr12	ARID2-12F	AGTGTTCATGTCACAACTCTGCAT	46254544	46254669
	ARID2-12R	TGACTGGATCCACTGAATGGAGAA		
chr12	ARID2-13F	AGCTTTAGGATCCTTTTCCACTGTATTTG	46215201	46215331
	ARID2-13R	TCTATTGTTGGCTTAGCTGCATGTAA		
chr12	ARID2-14F	TTCTTTAGCTTCCAGAGCAGTTGTAG	46240631	46240772
	ARID2-14R	CTTTGCTTTACTCCTCCCAAGGATAA		
chr12	ARID2-15F	TTTGCAAAGCAGAAGATAATGGTGTFTT	46231302	46231429
	ARID2-15R	GTATAGCACCTCGAGTGTTGAGATTAC		
chr12	ARID2-16F	GGAGATGTTGCTTGACAAAAATTG	46231445	46231617
	ARID2-16R	CCTGAGGAGGTAATGACTCTTCAAAAATA		
chr12	ARID2-17F	GGTTCTCATTTAAGCAAAAACATTCCA	46246011	46246184
	ARID2-17R	TGAACCACCCAAAGTTAATACAGGTC		
chr12	ARID2-18F	GGTACTTCAGATTGCAGTGATTTTGAGA	46230606	46230780
	ARID2-18R	GCTTACCTCAGCTGCAATATTTCC		
chr12	ARID2-19F	TTCTACTTGGGATTTTTCAGTCATTTTCAT	46243758	46243926
	ARID2-19R	CACCTTGTACAGGAAAAGACATTTGTG		
chr12	ARID2-20F	ATTTGCTTAATGGACCTCTAGCTTCAA	46246243	46246412
	ARID2-20R	GTCAGATGATAGGGCAGGAGAATG		
chr12	ARID2-21F	CACTGATCTCAAATAGCCCAGCA	46245367	46245527
	ARID2-21R	CTGCGTTCGGCTCATGGTA		
chr12	ARID2-22F	TCCTCTACCCCTCAATCACAGG	46244877	46245047
	ARID2-22R	TGCATTCGACTGTTGAGGTTGA		
chr12	ARID2-23F	AATGCTCCAGTGACTGTCATTCAA	46243995	46244166
	ARID2-23R	CAGGTTGTGGCCCTGTACTATG		
chr12	ARID2-24F	CTGATGCACTAGCTGCGGTA	46233157	46233322
	ARID2-24R	ACACTCGTAACCACACTGAAAAAGAA		
chr12	ARID2-25F	CTACTCCCCATTCAAAGGTGAT	46245604	46245739
	ARID2-25R	GTGTTTGTGGCTCCTCGTC		
chr12	ARID2-26F	TGTTTCTCGAGCAGAAATGTATTCTGA	46242655	46242808
	ARID2-26R	AAGAGATGTTGTAAGTGACGTAATGGA		
chr12	ARID2-27F	GGTCCCGGCTGACAAGT	46123787	46123914
	ARID2-27R	GAATCCGCCTAAAGTAGTGACTCTG		
chr12	ARID2-28F	CTTCAACACCAGGTGGTAAGCTTAT	46245115	46245288
	ARID2-28R	GAACATTTTGCTGCCCAACCAA		
chr12	ARID2-29F	TCATAGCACCCCCACAGTATGT	46244446	46244582
	ARID2-29R	CCCTACCCTTGAGGCTTGTG		
chr12	ARID2-30F	GGCGCTTTTAAAAACACCGATCT	46123567	46123711
	ARID2-30R	CCTCTGCTGTGGTGGAAGTCTG		

Chr	Gene	Ion_AmpliSeq Fwd_Rev_Primer*	Amplicon Start	Amplicon Stop
chr12	ARID2-31F	TTCTTAGATTGTTACAGTTCAGACAACCT	46124954	46125100
	ARID2-31R	CACCGCAAGTAATACTGTTTTAAAGCAA		
chr12	ARID2-32F	CCATCTGTAATTCCACAGCAGTCTC	46244202	46244375
	ARID2-32R	GGTGATTAAGTGTGACCCTGTGAA		
chr12	ARID2-33F	CAACCCAACAAAGCGTAGTGATT	46244668	46244821
	ARID2-33R	GAAGATGGAGTTATGTTAGGTTGTCCAG		
chr12	ARID2-34F	AGTGCTAGCCATTAGTAACATGGAAG	46298747	46298912
	ARID2-34R	TGTGAAATGTGGCTGACTTTGAGT		
chr12	ARID2-35F	CTCTTCTCAAAAAGGATAAGCACTGT	46285779	46285950
	ARID2-35R	CGTTTCATTTTCTTGTGTTTTGCCTAGT		
chr12	ARID2-36F	CAAGCTCAACTCCTAGAGCACA	46287223	46287360
	ARID2-36R	ACTTTTACAGATAAGGGTTTCAAGGCAT		
chr12	ARID2-37F	ATATTAAGATGTAGTTTTGTTGTTTTTCCAAGC	46215097	46215253
	ARID2-37R	CAGTCTTCTCTTCCATTCTTCTCCAAA		
chr12	ARID2-38F	CCATGGACTTTAATTGCGCAAATGATT	46211480	46211619
	ARID2-38R	TTTTTCAAGTTGCATGACGTGCT		
chr12	ARID2-39F	GGCCAGCCTAACAGTTCTCA	46254623	46254750
	ARID2-39R	AATAAAATGCCATTACTTTTTACAAGACTGC		
chr12	ARID2-40F	GAGAATGGAGCATCTGGGAAACA	46245900	46246072
	ARID2-40R	TCACCATTTCTACATGATTGGAAGTTT		
chr12	ARID2-41F	CAGATCCAAGCACTGTAGCTAAAGTAG	46246585	46246720
	ARID2-41R	CAGAAATTTATAAGAGAAATGTGATCATGGAA		
chr12	ARID2-42F	GTCATGGAGAACCCGCTCTG	46245699	46245838
	ARID2-42R	TTCCCTGAGTTTGAAGGAGGTAGA		
chr12	ARID2-43F	AAGCATGCCTTGAAACCTTATCT	46287329	46287503
	ARID2-43R	TGCGACCACATTCTGAATATTTACCAAT		
chr12	ARID2-44F	CTGGTTCATACATAAGCAACAGTAAACA	46243252	46243421
	ARID2-44R	CCTGCCATTGCTACTGGAATC		
chr12	ARID2-45F	CAGTTTCTACTTCTGTTGTTCTGTGTTG	46243494	46243668
	ARID2-45R	GCTACTGGAGACACAAAAGTTATACAT		
chr12	ARID2-46F	GCTTTTACTGGACCCTGTTGATTTT	46231103	46231229
	ARID2-46R	AGAAAGTAAATACTTCAGTGAAAACACCTC		
chr12	ARID2-47F	TGTGTTACCAGTGATGGCAT	46233042	46233201
	ARID2-47R	TGGAACCTGGGTGTTCAATGAGTTT		
chr12	ARID2-48F	AGAAGTTGTTCTAACGCTGCCT	46125051	46125197
	ARID2-48R	AACTGTTTGGTGTGCAAGGTTT		
chr12	ARID2-49F	ACCTAGCCTGTGTGTGACTATTTCT	46230291	46230458
	ARID2-49R	TTTTGCTTTCACTAACTTACCATGAGACT		
chr12	ARID2-50F	GATGAGGTACCACCAGGCAAT	46205247	46205371
	ARID2-50R	GCGGACATATATCCTGTAAATTTGCAGT		
chr12	ARID2-51F	GCCAAAGGTGATCAACTAGAAAAGATTT	46246125	46246293
	ARID2-51R	GCTGAGGCACATCTGAATTCAAAC		
chr12	ARID2-52F	CCCGACTCAGGATCAAAAAGTATCC	46246365	46246537
	ARID2-52R	GGGTCACTGATTGTAACAATTCTAACTC		
chr12	ARID2-53F	GGTATTCATTGTGTTAATGGGTTCAATATCC	46240538	46240687
	ARID2-53R	ACTATTCCTGGAGGTGGAGCAA		
chr12	ARID2-54F	TGTAGAAGGTACATCAGGAGAATGGAT	46230519	46230657
	ARID2-54R	ATTGCCCTCCTCAAAGGAAAGATT		
chr12	ARID2-55F	TCATCTCACTTTACCTGATGTGCTG	46231375	46231501
	ARID2-55R	TCCAGTCTTACCTATGCTCTTTTCTACT		
chr12	ARID2-56F	AGGCAATTAGGCCTTGACACATTA	46230733	46230860
	ARID2-56R	GGCAACTATAAAATGCCACATCCTAAAA		
chr12	ARID2-57F	CAAACCTCAGCTCAGCAGCTA	46245246	46245412
	ARID2-57R	CCAGAAGTCCCTTGAAAATGGT		

Chr	Gene	Ion_AmpliSeq Fwd_Rev_Primer*	Amplicon Start	Amplicon Stop
chr12	ARID2-58F	GCTGGTCAGACAGTTTCAGCTAA	46244772	46244920
	ARID2-58R	TGACTGACAGTAGGAGGTGGTC		
chr12	ARID2-59F	CAACCAATCTTCAAATCTGACTGCAA	46243874	46244044
	ARID2-59R	ACTTCACAAGGAATTGGAGCTTACT		
chr12	ARID2-60F	GCTCATTGCTCCAGCAGGAAT	46245488	46245654
	ARID2-60R	CTTCCTCCTCCTTTTGGCAAATTATTTT		
chr12	ARID2-61F	CCCCAACAAGTACAGATGCAAGT	46245003	46245160
	ARID2-61R	GGAATCTGTGGAGCTGGGAGA		
chr12	ARID2-62F	GAACATACCAGCATGTACTTCTACAGT	46244324	46244495
	ARID2-62R	TGAGACAATATTGGATGCAGAAGTTGTT		
chr12	ARID2-63F	CATTGCCTCCACATGGCTTTAG	46242534	46242708
	ARID2-63R	GCTAATTTACTGCAAGTCGAGAGGTAT		
chr12	ARID2-64F	GGGAAGGAGCTGGATCTTCAC	46123858	46123985
	ARID2-64R	GGGTCAAAGGAGACTTTTTGGA		
chr12	ARID2-65F	CTCCACAGAGTTCTGTTGTTTCAAGT	46244119	46244250
	ARID2-65R	GTCCTGGTACCAGTGTGTGTAATG		
chr12	ARID2-66F	GACTCGCTTTCCTGGACGA	46123666	46123812
	ARID2-66R	AAGCCCCGCACTTGTCA		
chr12	ARID2-67F	TTACTATTGCTGGTGTCCAAGTC	46244539	46244710
	ARID2-67R	TGTTGAGCTGGCTGGCTTAC		
chr12	ARID2-68F	CTGTGACCTTGGAGACTGACA	46298673	46298803
	ARID2-68R	TCATAAAGGCATTTGGCAAGGGT		
chr12	ARID2-69F	GTTGGCAGGGTGGTCATAGTT	46287122	46287267
	ARID2-69R	TGGGATGATTCACAATGGCCTT		
chr12	ARID2-70F	TGTGAGCCTTTTCAGCGACA	46285654	46285827
	ARID2-70R	CTGCAAGTAGGGCATCCTTTGA		
chr12	ARID2-71F	TACTTGAATACATCCAAGTGCTACTACTTTTA	46285520	46285646
	ARID2-71R	CAAAGACACTGCCCTGGATATACAT		
chr16	AXIN1-1F	CCGCTCCTCCAGCTTCTC	354322	354467
	AXIN1-1R	GTGCCTCACAGTCCTGTCC		
chr16	AXIN1-2F	CCTGACTTGGGTACGTGCTT	347944	348106
	AXIN1-2R	CTGAGAGCATCCTGGACGAG		
chr16	AXIN1-3F	GCGGGAAGTGGTGCCAA	348161	348292
	AXIN1-3R	GCTGGTGCTGAGAGGTGATG		
chr16	AXIN1-4F	CGGCTCTACGATGGGACCT	341112	341241
	AXIN1-4R	CGGTGCTGCACGTGGTA		
chr16	AXIN1-5F	TCCGTGAGGGACTGGGTATC	364508	364667
	AXIN1-5R	GAGCCAGTCAACCCCTATTATGTC		
chr16	AXIN1-6F	CCGTTTTTCCCCTGAAGACCT	343387	343540
	AXIN1-6R	CGACGTCTGGAGGAGGAAGAA		
chr16	AXIN1-7F	GGTGGTACACCCAACACTGTT	338005	338156
	AXIN1-7R	AGAAGATCATCGGCAAAGTGGAG		
chr16	AXIN1-8F	GTCTTGGATGAAGAGGTGGGA	343598	343751
	AXIN1-8R	CCACAGTCACAACCTTGTCTCTTG		
chr16	AXIN1-9F	CATGTCCTGGTCACACTTCCA	396267	396436
	AXIN1-9R	CCTATCCCTCCTTCTTAAGTCTGATAT		
chr16	AXIN1-10F	CTGATCTCCTTTTCCCCCTCA	347080	347253
	AXIN1-10R	GGCGACTCGCAGCTCTT		
chr16	AXIN1-11F	GGCTTATCCCATCTTGGTCATCC	396725	396898
	AXIN1-11R	GTTTCTGCTCCGGGAAAGGT		
chr16	AXIN1-12F	GCAGCGCCAACACTCTCT	347752	347905
	AXIN1-12R	ACGTCCACCACCACGTC		
chr16	AXIN1-13F	CAGGATCGATCAGCTGCTTCA	396488	396638
	AXIN1-13R	TGTGACTCGAACGAGGAGAAGA		

Chr	Gene	Ion_AmpliSeq Fwd_Rev_Primer*	Amplicon Start	Amplicon Stop
chr16	AXIN1-14F	GAAACTTGCTCCGAGGTCCAA	396978	397118
	AXIN1-14R	TGGTTTCAACAGGACAGATTGATTCA		
chr16	AXIN1-15F	CGTTGGGCACCCACATACTC	339385	339559
	AXIN1-15R	CAGCATCGTTGTGGCGTACTA		
chr16	AXIN1-16F	AGGACATCCGGTGTGGGTTA	396041	396180
	AXIN1-16R	CAGTAGACGGTACAGCGAAGG		
chr16	AXIN1-17F	ACCAGCCTATCAGTCCACCTT	338113	338277
	AXIN1-17R	TCACTGCCCTCCTCACCT		
chr16	AXIN1-18F	GTCCTCAGCACACGCTGTA	348064	348206
	AXIN1-18R	CCGTGTCAAGCTGCCT		
chr16	AXIN1-19F	GCACGTACTCTGTCTCGGA	341182	341322
	AXIN1-19R	ACCTTTGACGCGGGTTGTT		
chr16	AXIN1-20F	TGAGGAGGTTCTGGCCTTCT	347648	347788
	AXIN1-20R	CAAGGTCCCGAGGCTACTC		
chr16	AXIN1-21F	AGGGCATAGCCGGCATT	364627	364752
	AXIN1-21R	GAAAATGTGAAATAAATGAGAGGCTGTT		
chr16	AXIN1-22F	GCTCGGCTGGCTCTCTT	343503	343646
	AXIN1-22R	CCCTCAGCTCCGGACCT		
chr16	AXIN1-23F	TCGGACTCACCTGAACTCTCT	396138	396312
	AXIN1-23R	GCCGACCTTAAATGAAGATGAGGAA		
chr16	AXIN1-24F	GATGGCTCTCGCCAGCTT	396594	396767
	AXIN1-24R	GAGTCACTGCATTCCCTGCT		
chr16	AXIN1-25F	ACTCGCCACACACACTGAAG	346958	347125
	AXIN1-25R	ACCAGAAAATCATGCAGTGGATCAT		
chr16	AXIN1-26F	CCGGGCATCCCCATGAA	354211	354384
	AXIN1-26R	CTCATCCACCGCCTGGAG		
chr16	AXIN1-27F	GAGCCTGTCTCGTATATTCCAAATAA	396382	396530
	AXIN1-27R	AGCTTCATAAAGGGCTGCATCA		
chr16	AXIN1-28F	TCGAAGTCTCACCTTTAATGCCAA	396854	397025
	AXIN1-28R	ATGAATATCCAAGAGCAGGGTTTCC		
chr16	AXIN1-29F	GCCTCCACCTGCTCCTT	347851	348014
	AXIN1-29R	CACGTGGCCAAGATGCC		
chr16	AXIN1-30F	ATGGGTTCCCCGAGAAG	339521	339650
	AXIN1-30R	CTCAGTCCAAAACCAGGTACCA		
chr16	AXIN1-31F	GAGGACGATGGGCTGAGG	359935	360102
	AXIN1-31R	CTGAGTTAACGGCTGCCTCTT		
chr9	CDKN2A-1F	ACCAGCGTGTCCAGGAAG	21971074	21971239
	CDKN2A-1R	CCTGGCTCTGACCATTCTGTT		
chr9	CDKN2A-2F	CCTGTAGGACCTTCGGTGACT	21968162	21968300
	CDKN2A-2R	TGTGCCACACATCTTTGACCT		
chr9	CDKN2A-3F	CCTTCTGAAAACCTCCCCAGGAA	21974543	21974708
	CDKN2A-3R	CACCGAATAGTTACGGTCCGGA		
chr9	CDKN2A-4F	GCGGGATGTGAACCACGA	21994246	21994395
	CDKN2A-4R	GCTCACCTCTGGTGCCAAA		
chr9	CDKN2A-5F	TTTGGAAGCTCTCAGGGTACAA	21970856	21971023
	CDKN2A-5R	GTCTGCCCGTGGACCTG		
chr9	CDKN2A-6F	GTCAGCCGAAGGCTCCAT	21974785	21974952
	CDKN2A-6R	GGCTCTTCCGCCAGCAC		
chr9	CDKN2A-7F	GACCCTCTACCCACCTGGAT	21974663	21974810
	CDKN2A-7R	GGAGCAGCATGGAGCCTT		
chr9	CDKN2A-8F	CGGACTTTTCGAGGGCCTT	21994113	21994280
	CDKN2A-8R	GCCGCGAGTGAGGGTTT		
chr9	CDKN2A-9F	CCAGCCAGCTTGCGATAAC	21968655	21968814
	CDKN2A-9R	AGAAGCCAGAGCACATGAATAAATGA		

Chr	Gene	Ion_AmpliSeq Fwd_Rev_Primer*	Amplicon Start	Amplicon Stop
chr9	CDKN2A-10F	CCCCTTCAGATCTTCTCAGCA	21974427	21974592
	CDKN2A-10R	GCTTCGATTCTCCGAAAAAGG		
chr9	CDKN2A-11F	GATGGCCCAGCTCCTCAG	21970988	21971139
	CDKN2A-11R	CGACCCCGCCACTCTCA		
chr3	CTNNB1-1F	AAAACACTGTGGACCACAAGCA	41268764	41268888
	CTNNB1-1R	CTGCAGATGCTATACACAAGACTCA		
chr3	CTNNB1-2F	CCCAGAATGCAGTTCGCCTT	41275276	41275419
	CTNNB1-2R	CAGCTAGAGATGCTTCATTTTCAATTC		
chr3	CTNNB1-3F	GTACCGGAGCCCTTCACATC	41277249	41277375
	CTNNB1-3R	TTTATGGACATAAAACCTAGAACACTTCACTT		
chr3	CTNNB1-4F	GTCCTCTGTGAACCTTGCTCAGG	41277885	41278027
	CTNNB1-4R	ACACATCTGCTAAAGGCTTTGGTT		
chr3	CTNNB1-5F	TCTAGGTGGAATGCAAGCTTTAGG	41274827	41274963
	CTNNB1-5R	CCATGGCACCAGTTTACTCAGAA		
chr3	CTNNB1-6F	GTGACATTTAACAGGTATCCAGTGA	41265426	41265578
	CTNNB1-6R	ACAAACCTTGAGTAGCCATTGTCC		
chr3	CTNNB1-7F	TCAGAATGTCTACCCAATACCAGTACTT	41266761	41266927
	CTNNB1-7R	GTACAATAGCAGACACCATCTGAGG		
chr3	CTNNB1-8F	GGTTGCCTTGCTCAACAAAACAAA	41267262	41267392
	CTNNB1-8R	CAATGCTCCATGAAAACACATAAAGAA		
chr3	CTNNB1-9F	CTTGTTCACTTCTGGGTTTCAAGT	41275047	41275220
	CTNNB1-9R	CTCAGTGATGTCTTCCCTGTCA		
chr3	CTNNB1-10F	CGACATATGCAGCTGCTGTTT	41278079	41278248
	CTNNB1-10R	ACCTGCTCTAGGAACTACCAGATAAAT		
chr3	CTNNB1-11F	TGCTGAACTTTGGATGCCCTAA	41280517	41280683
	CTNNB1-11R	GTCCATACCCAAGGCATCCTG		
chr3	CTNNB1-12F	GGTGCCATTCCACGACTAGTTC	41275696	41275833
	CTNNB1-12R	GCATTCCCTTTTAGATAGCCAGGTATCAC		
chr3	CTNNB1-13F	CTCATGGATGGGCTGCCT	41280771	41280920
	CTNNB1-13R	CCAACCAAGTCTTTCTGAAGTTCTGTA		
chr3	CTNNB1-14F	GTGAGGGCTTACTGGCCATC	41267003	41267171
	CTNNB1-14R	AACACAGAATCCACTGGTGAACCT		
chr3	CTNNB1-15F	GTATGCAATGACTCGAGCTCAGA	41266458	41266620
	CTNNB1-15R	CAAGTTTACAACCTGCATGTTTCAAGCA		
chr3	CTNNB1-16F	CCTTCTCTGAGTGGTAAAGGCAATC	41266133	41266285
	CTNNB1-16R	ATTCTGACTTTTCAAGGCAATGAA		
chr3	CTNNB1-17F	ATTTCAATGGGTCATATCACAGATTCTT	41265928	41266093
	CTNNB1-17R	GTAAGACTGTTGCTGCCAGTG		
chr3	CTNNB1-18F	CCTCAAACCTTTACAGAGGAGAATGCC	41277160	41277290
	CTNNB1-18R	GGTTGTGAACATCCCGAGCTA		
chr3	CTNNB1-19F	AATTGGGAATGTTTGCACCACAG	41277755	41277929
	CTNNB1-19R	AATAGCTTCTGCAGCTTCCCTTGT		
chr3	CTNNB1-20F	AGACAGAAAAGCGGCTGTTAGT	41266051	41266181
	CTNNB1-20R	AGGTATCCACATCCTCTTCTCAG		
chr3	CTNNB1-21F	CATACAACCTGTTTTGAAAATCCAGCGT	41265528	41265681
	CTNNB1-21R	CCTGACAAGTAAGCAGGGAGAGA		
chr3	CTNNB1-22F	TTTGCTTTCTATTCTTCTTCTTGTG	41279466	41279640
	CTNNB1-22R	AACCGGCTCTTCTGATAAAAGTTTTAGA		
chr3	CTNNB1-23F	AGACACGCTATCATGCGTTCT	41266882	41267045
	CTNNB1-23R	GCAGGAATGCCTCCAGACTTAAA		
chr3	CTNNB1-24F	CGTTCTTTTCACTCTGGTGGATATGG	41280636	41280810
	CTNNB1-24R	CCAGCTGATTGCTGTACCT		
chr3	CTNNB1-25F	GCTGCAACTAAAACAGGTAAATTTCTGAGT	41274921	41275095
	CTNNB1-25R	CTGCACAGGTGACCACATTTATATC		

Chr	Gene	Ion_AmpliSeq Fwd_Rev_Primer*	Amplicon Start	Amplicon Stop
chr3	CTNNB1-26F	TTGTGGTGAAGAAAAGAGAGTAATAGCA	41266334	41266501
	CTNNB1-26R	GGAACATAGCAGCTCGTACCC		
chr3	CTNNB1-27F	GTTGTATGCCAGTTCTTCCTTCTGT	41275598	41275738
	CTNNB1-27R	GATGTGCACGAACAAGCAACT		
chr3	CTNNB1-28F	TGTGCGTACTGTCCTTCGG	41275175	41275322
	CTNNB1-28R	CTTAACCACAACCTGGTAGTCCATAGTG		
chr3	CTNNB1-29F	CGTTTGGCTGAACCATCACAGA	41266573	41266731
	CTNNB1-29R	GCAGACTAAAAAGCTAGCTATGTCATTG		
chr3	CTNNB1-30F	CCCTGGTGAAGTGTGGGTAA	41267045	41267200
	CTNNB1-30R	GGTTGTGGAGAGTTGTAATGGCATA		
chr3	CTNNB1-31F	ACTCTAGGAATGAAGGTGTGGGT	41277970	41278122
	CTNNB1-31R	CTTGTCTCAGACATTCGGAACA		
chr3	CTNNB1-32F	CTTAGGTGAAAGGAAGTAACTCTGGAA	41274707	41274873
	CTNNB1-32R	GACTTGGATCTGTCAGGTGAAGT		
chr3	CTNNB1-33F	GGCTCTTCTCAGACATGTGATCAA	41268635	41268808
	CTNNB1-33R	ACAGATAGCACCTTCAGCACTC		
chr10	PTEN-1F	ATTTTCTTTCTCTAGGTGAAGCTGTACTTC	89725029	89725175
	PTEN-1R	ATCAGAGTCAGTGGTGTGAGAATATCTA		
chr10	PTEN-2F	GCCATCTCTCTCCTCTTTTCTT	89624184	89624346
	PTEN-2R	GTGACAGAAAGGTAAGAGGAGCA		
chr10	PTEN-3F	TTGACTTTTTGCAAATGTTTAAACATAGGT	89720535	89720708
	PTEN-3R	AGGTTTCTCTGGTCTGGTAT		
chr10	PTEN-4F	CTACGACCCAGTTACCATAGCAATTTA	89711785	89711923
	PTEN-4R	GGTAGCTATAATAACACATAGCGCCTC		
chr10	PTEN-5F	GACAGTTAAAGGCATTTCTGTGA	89717556	89717717
	PTEN-5R	GTAACGGCTGAGGGAACCTCAA		
chr10	PTEN-6F	AACTTTTCTTTTAGTTGTGCTGAAAGACATT	89690789	89690943
	PTEN-6R	TCACTCGATAATCTGGATGACTCATT		
chr10	PTEN-7F	TAACCCACCACAGCTAGAACTTATC	89692795	89692946
	PTEN-7R	TGCCCCGATGTAATAAATATGCACAT		
chr10	PTEN-8F	TCCAAACATTATTGCTATGGGATTTCTT	89653789	89653959
	PTEN-8R	ATCTTTTCTGTGGCTTAGAAATCTTTTC		
chr10	PTEN-9F	AAGAAATCGATAGCATTGTCAGTATAGA	89720742	89720866
	PTEN-9R	TGGAGAAAAGTATCGGTTGGCTTT		
chr10	PTEN-10F	AGTATTCTTTTAGTTTGATTGCTGCATATTTCA	89653719	89653843
	PTEN-10R	CCTGTATACGCCTTCAAGTCTTTCT		
chr10	PTEN-11F	TTACAGTGCTTAAAAATTAATATGTTTCATCTGC	89724916	89725085
	PTEN-11R	ATTTGACGGCTCCTCTACTGTTTT		
chr10	PTEN-11F	GTTAGCTCATTTTTGTTAATGGTGGCT	89685218	89685388
	PTEN-11R	ACAAGCAGATAACTTTCACTTAATAGTTGTT		
chr10	PTEN-12F	TCACTTTTGGGTAAATACATTCTTCATACCA	89720662	89720794
	PTEN-12R	ATATTCCTTGTGATTATCTGCACGC		
chr10	PTEN-13F	GAAAGGGACGAACCTGGTGAATGAT	89692896	89693061
	PTEN-13R	ATAAATCTCAGATCCAGGAAGAGGAAAG		
chr10	PTEN-14F	GCAGTTCAACTTCTGTAACACCAG	89725096	89725268
	PTEN-14R	TATTTTCATGGTGTGTTTATCCCTCTTGA		
chr10	PTEN-15F	CGACGGGAAGACAAGTTCATGTA	89717672	89717821
	PTEN-15R	TGTCCTTATTTTGGATATTTCTCCAATG		
chr10	PTEN-16F	TGCAACATTTCTAAAGTTACCTACTTGT	89692686	89692845
	PTEN-16R	TGGTCAAGATCTTCACAAAAGGGTTT		
chr10	PTEN-17F	CCAGGGAGTAACTATTCCAGTCA	89711871	89712045
	PTEN-17R	ATGGAAGGATGAGAATTTCAAGCA		
chrX	RPS6KA3-1F	CTCCTCCTCTCCCATAGGTTTCAT	20252891	20253065
	RPS6KA3-1R	GTGGTTCTACTTCATTAGAGCCCTTTAG		

Chr	Gene	Ion_AmpliSeq Fwd_Rev_Primer*	Amplicon Start	Amplicon Stop
chrX	RPS6KA3-2F	AACTGGTAGTATTTCTTCTCACTATGGA	20179685	20179859
	RPS6KA3-2R	GTCCTGATGATACACCAGAGGAAA		
chrX	RPS6KA3-3F	GCTTTAGTTCAAACAGGATGCATGTAA	20212250	20212399
	RPS6KA3-3R	TTTCTTCATTTCAGCTTTTCAAACACTGAAGG		
chrX	RPS6KA3-4F	GCAGGTTTAAATGGCGGATGAATT	20194430	20194560
	RPS6KA3-4R	GACTGGAATGTAAGTATAGAATGAAAACCTGC		
chrX	RPS6KA3-5F	AGCTTACACCATGGTACCAAATATCTC	20173477	20173651
	RPS6KA3-5R	TTCTTTTGTTTCAGGGTGCCA		
chrX	RPS6KA3-6F	GCCTGGCCCTAATACAAGTTTTTAAT	20206486	20206640
	RPS6KA3-6R	TGAAGAAGGTCACATCAAGTTAACAGG		
chrX	RPS6KA3-7F	AAAAGCACACACTCATGACTTACCT	20227383	20227554
	RPS6KA3-7R	CCTGTTTGCTCCTCTTACGTTTATCTC		
chrX	RPS6KA3-8F	AGTCTGGAAAAGATTTTCTGATGAAACA	20194287	20194424
	RPS6KA3-8R	GCAGGCCTGAAGATACATTCTATTTTGA		
chrX	RPS6KA3-9F	TGAACAAATGCTTAGGTGCTTAGAACA	20181039	20181172
	RPS6KA3-9R	TGCACTTTTTCTAGGTTTTAAAAAGACAAGG		
chrX	RPS6KA3-10F	TTTCTATCTAATTCTAGCTCAAACATAAAGGAA	20185632	20185758
	RPS6KA3-10R	TGCTGTCTCTGTTCACTATAACTAAAACC		
chrX	RPS6KA3-11F	TGGCCTCTCGTTCAGAGAAAAAAT	20185760	20185924
	RPS6KA3-11R	TTCGTTTTGGGCACATATAGTTTCA		
chrX	RPS6KA3-12F	TTTAACATTCACTGCTGATCCTCACTTA	20205875	20206046
	RPS6KA3-12R	CATATTCTTTTTGTGGAAGTGTGGAGT		
chrX	RPS6KA3-13F	ATCCTTTAGAGTGATAATGTTTGGATGCT	20187520	20187693
	RPS6KA3-13R	AATACCCAAGGTCTGTAGCTATAAACAA		
chrX	RPS6KA3-14F	ACTAATACTGCAAGCAAACCTCTCTCAC	20182984	20183139
	RPS6KA3-14R	GTGGATGAATCTGGTAATCCGGAAT		
chrX	RPS6KA3-15F	GGGAGTTTCATGAAGCCACTTGAT	20211519	20211682
	RPS6KA3-15R	CTACTTGGCTGAACTTGCACCTTG		
chrX	RPS6KA3-16F	ACCTCTACCAAGATATCACGTTCCA	20213215	20213346
	RPS6KA3-16R	TTAACATGGGCACCAAAAAGAATTTGT		
chrX	RPS6KA3-17F	ACATCCACTCACCTGAACAATTGA	20193270	20193399
	RPS6KA3-17R	TGGCATTCCACCTAGTGCTAATG		
chrX	RPS6KA3-18F	GTGTGGCCTTCTTCAATACCTTCA	20222146	20222293
	RPS6KA3-18R	GCAGGGTCTATTTTAAACAAGGATATAGTTTAT		
chrX	RPS6KA3-19F	GAGGACCTGTGGAACACAGTGA	20174187	20174359
	RPS6KA3-19R	GTCAAAGATGCTTCATGTAGACCTT		
chrX	RPS6KA3-20F	AACAAATATCTACTCCCGCTAAAAACA	20195040	20195201
	RPS6KA3-20R	GCCAAACTTGGAAATGCCACA		
chrX	RPS6KA3-21F	CTTCACTGCAAACCTCCATGTTTGT	20190864	20191029
	RPS6KA3-21R	GGCATTATGGGACTCTTACCAATTTAT		
chrX	RPS6KA3-22F	ATCAGAACGGGCCGAGTG	20284585	20284730
	RPS6KA3-22R	GGACCCGTGGCAGAAGATG		
chrX	RPS6KA3-23F	AGGGATGAGGTCAAGGGATGAT	20213139	20213277
	RPS6KA3-23R	ACAATCTTTTATAGTTCGAGACCGAGTTC		
chrX	RPS6KA3-24F	GCCCAGCCTGAAGCATTTATTTTAA	20204352	20204519
	RPS6KA3-24R	TGACTAACACTTCTGATTATTTGCCTTCT		
chrX	RPS6KA3-25F	GACGATTAACACTTCTGGAGCCATAT	20205993	20206120
	RPS6KA3-25R	ATGTAGGATATTTATGATATTTGTTTTGGCAGA		
chrX	RPS6KA3-26F	CCAGTCTATCGTTGAGAAAAATGAATGTC	20194555	20194694
	RPS6KA3-26R	CAAGTATTGAAAACAAAGGCTGAGCAA		
chrX	RPS6KA3-27F	AGTAGCATGATTCTTACTTTTCTGGTTTT	20211587	20211711
	RPS6KA3-27R	GTGATGTTACAGAAGAAGATGTCAAAT		
chrX	RPS6KA3-28F	TGCAAAGCCAAAATCACAAATTCGAATA	20183086	20183257
	RPS6KA3-28R	CTTGACAGTTTGTGAATAAAGAACTTCA		

Chr	Gene	Ion_AmpliSeq Fwd_Rev_Primer*	Amplicon Start	Amplicon Stop
chrX	RPS6KA3-29F	ACTCCATATATCACAAAGCAGCATCATAG	20181114	20181242
	RPS6KA3-29R	TTGGAATATATAAATTCAACTCATAGCCTTACAA		
chrX	RPS6KA3-30F	CTTACCCTTGTGCGTGAAGATATT	20185702	20185826
	RPS6KA3-30R	CTTATGAAAGGAGGTGAATTGCTGGATA		
chrX	RPS6KA3-31F	CTCCGCTGAGCAAGAGTAGAG	20173550	20173705
	RPS6KA3-31R	TTTTGTATTTGTGTACGTGTGACTATCCA		
chrX	RPS6KA3-32F	GCTGGTAACAGGAACCTATTCAG	20222022	20222196
	RPS6KA3-32R	GCTCTGATGCTAGGCAGCTTTATG		
chrX	RPS6KA3-33F	TGATGAGTCTTTGTCTACCACATGAT	20187420	20187571
	RPS6KA3-33R	TGAAATTCCTTCGTTATGGACAGCA		
chrX	RPS6KA3-34F	CCGCTACCTATTCGTGCCAAT	20179814	20179957
	RPS6KA3-34R	CTGAATTAGAGTACTTTTGAGATTACCAACCA		
chrX	RPS6KA3-35F	TTGCTCACTTGAAGTCAGTAGAAAAAGA	20193188	20193317
	RPS6KA3-35R	GCTATGCAGACAGTTGGTGTACAT		
chrX	RPS6KA3-36F	CCCCGAAAAAGCTGATGTG	20193357	20193489
	RPS6KA3-36R	AGTGAAAAAGGTCATAGTGTCAACAGTT		
chrX	RPS6KA3-37F	ACATTTTCTCCTTACCTTTGGGAGTT	20194353	20194479
	RPS6KA3-37R	CCCCTCCAGAACTGTATAGAAGAGA		
chrX	RPS6KA3-38F	GAGCAGCAGTCAGTCTCTGATG	20174313	20174468
	RPS6KA3-38R	CAGTATCACAGAAGCTAAGGAAGAATGT		
chrX	RPS6KA3-39F	AACAGAAAACATTGCTGGTAGGAAGA	20227284	20227432
	RPS6KA3-39R	GTATTAGGGCAGGGATCATTGGAA		
chrX	RPS6KA3-40F	ACCTGTAACTTGATGTGACCTTCTTC	20206613	20206743
	RPS6KA3-40R	TGGTGATCTGATGCTTAAGGATGTAATTT		
chrX	RPS6KA3-41F	GGGAGTGTAATTTTAACTGGTACTGA	20190786	20190924
	RPS6KA3-41R	CTCCTACTCTGTTTGAAGAGATGTATA		
chrX	RPS6KA3-42F	TGTA AAAATGATTACTCTGCTCAGCTGAA	20252784	20252941
	RPS6KA3-42R	TTTTACAGAATGGACAGCAAATTTATGG		
chr17	TP53-1F	CTCTGGCATTCTGGGAGCTT	7579483	7579618
	TP53-1R	TCTGACTGCTCTTTTCACCCATC		
chr17	TP53-2F	GCCACTGACAACCACCCTTAA	7578121	7578258
	TP53-2R	GGAAGGAAATTTGCGTGTGGAGTA		
chr17	TP53-3F	GGGCCAGACCTAAGAGCAATC	7578281	7578453
	TP53-3R	CATGGCCATCTACAAGCAGTCA		
chr17	TP53-4F	GGGCCTGCCCTTCCAAT	7579810	7579941
	TP53-4R	GCAGCCAGACTGCCTTCC		
chr17	TP53-5F	GTGCTTCTGACGCACACCTA	7572811	7572957
	TP53-5R	GTTCAAGACAGAAGGGCCTGA		
chr17	TP53-6F	GGGATGTGATGAGAGGTGGAT	7577372	7577532
	TP53-6R	CCATCCTCACCATCATCACACTG		
chr17	TP53-7F	TCAGGCAAAGTCATAGAACCATTTTCA	7576590	7576714
	TP53-7R	TGGTTGTAGCTAACTAACTTCAGAACAC		
chr17	TP53-8F	CATGAAGGCAGGATGAGAATGGAA	7573839	7573988
	TP53-8R	GCCTTGGAACCTCAAGGATGCC		
chr17	TP53-9F	GGATACGGCCAGGCATTGAAG	7579258	7579407
	TP53-9R	TCATCTTCTGTCCCTTCCCAGA		
chr17	TP53-10F	AAGAGGCAAGGAAAGGTGATAAAAAG	7576933	7577087
	TP53-10R	CAGAGGAAGAGAATCTCCGCAAG		
chr17	TP53-11F	CAAGAAGTGGAGAATGTCAGTCTGAG	7572911	7573063
	TP53-11R	CAGACCCTCTCACTCATGTGATG		
chr17	TP53-12F	GGCTCCTGACCTGGAGTCTT	7577489	7577636
	TP53-12R	CTCATCTTGGGCCTGTGTTATCT		
chr17	TP53-13F	CTGGAGTGAGCCCTGCT	7573927	7574060
	TP53-13R	ACTTCTCCCCCTCCTCTGTTG		

Chr	Gene	Ion_AmpliSeq Fwd_Rev_Primer*	Amplicon Start	Amplicon Stop
chr17	TP53-14F	GTCGAAAAGTGTCTGTGCATCCAAA	7578209	7578382
	TP53-14R	CAGATAGCGATGGTGAGCAG		
chr17	TP53-15F	GGGACTGTAGATGGGTGAAAAGA	7579587	7579750
	TP53-15R	GGGACTGACTTTCTGCTCTTG		
chr17	TP53-16F	CGTAGCTGCCCTGGTAGGTTT	7579365	7579523
	TP53-16R	CTGAAGACCCAGGTCCAGATG		
chr17	TP53-17F	ACAACCTCCGTCATGTGCTG	7578412	7578585
	TP53-17R	TTTCAACTCTGTCTCCTTCCTCTT		
chr17	TP53-18F	GTGGTGAGGCTCCCTTT	7577047	7577186
	TP53-18R	CCTCTTGCTTCTCTTTTCCTATCCT		
chr17	TP53-19F	TGGAACTTTCCACTTGATAAGAGG	7576812	7576978
	TP53-19R	TGCAGTTATGCCTCAGATTCACCTTTTAT		

**Table S2:** Primer sets of target coding sequence of HCC-related genes

Chr.	Gene	Ion_AmpliSeq Fwd_Rev_Primer*	Amplicon Start	Amplicon Stop
chr5	APC-ex14F	AGGTTATTGCGAGTGTGAGGAA	112164551	112164675
	APC-ex14R	AGGTACCTTTTTAACTTCTAAAGCACAT		
chr5	APC-ex14F	TTGTCTTGGCGAGCAGATGT	112164577	112164750
	APC-ex14R	AAATCTCATGGCTAAAAGAAGGCA		
chr5	APC-ex16F	GCAAATCCTAAGAGAGAACAACCTGTCT	112173250	112173422
	APC-ex16R	TGAGGTTCTTGAGCATGCTAACTG		
chr5	APC-ex16F	AAAGATTGGAAGTAGGTCAGCTGAAG	112175218	112175362
	APC-ex16R	TCCTGAAGAAAATTCAACAGCTTTGTG		
chr5	APC-ex16F	TCATTATCATCTTTGTGCATCAGCTGAA	112175117	112175264
	APC-ex16R	CTGGAACCTCGCTCACAGGAT		
chr5	APC-ex16F	GTGGTGCTCAGACACCCAAAA	112175382	112175550
	APC-ex16R	GGGCTTATAATGCCACTTACCAT		
chr5	APC-ex16F	TACTCCAGATGGATTTTCTTGTTTCATC	112175776	112175917
	APC-ex16R	AGGCTGCTCTGATTCTGTTTCATT		
chr5	APC-ex16F	CCAGTTCAGGAAAATGACAATGGG	112175870	112176043
	APC-ex16R	TGGCATGGCAGAAATAATACATTCTTCT		
chr5	APC-ex16F	AGCAGCCTAAAAGAAATCAAATGAAAACC	112175910	112176084
	APC-ex16R	GCAGTCTGGGCTGGCTTTTTT		
chr20	ASXL1-ex13F	GGAAGCAGCCCAGTTCCTTA	31024005	31024178
	ASXL1-ex13R	TGTAATGGGATTTGTCACTGGATGG		
chr11	ATM-ex11F	GCCAGGCACTGTCTGATAGA	108122485	108122616
	ATM-ex11R	TTCTTGGAATATACTGGTGGTCA		
chr11	ATM-ex11F	CAGTATGCTGTTGACTTTGGCA	108122569	108122734
	ATM-ex11R	GTGCTATTTCTAAGTCACCCTCTAACT		
chr11	ATM-ex20F	ATGTTCTTGAACCTCTGAAACCACTAT	108141846	108142017
	ATM-ex20R	ACAAACATCTTGGTCACGACGA		
chr11	ATM-ex20F	CCACAGCAATGTGTGTTCTTTGTA	108141972	108142127
	ATM-ex20R	GCTCCAATTACTGTAAGAACTGTCCTT		
chr11	ATM-ex50F	TCATTAAATGTTGTATATCATGTGTGATTTTGT	108200855	108200990
	ATM-ex50R	GCAGGGCTAATTCATCCAACCTC		
chr11	ATM-ex50F	ATACACAGTAAAGTTTCAGCGAGAG	108200941	108201083
	ATM-ex50R	CCCACATATCATGTTCTTCTCCACTTA		
chr11	ATM-ex59F	TCAGTGGTCTTAATTGAAATTATGGCT	108217921	108218075

Chr.	Gene	Ion_AmpliSeq Fwd_Rev_Primer*	Amplicon Start	Amplicon Stop
	ATM-ex59R	AGTTCGCTGACTGCTCATTATCAA		
chr11	ATM-ex59F	GGACTTGGTGATAGACATGTACAGAAT	108218020	108218159
	ATM-ex59R	GCAAACAACATTCCATGATGACCAAATA		
chr11	ATM-ex62F	CATCAACTACCATGTGACTGGCTTA	108235700	108235841
	ATM-ex62R	GGTCCAGTCAAAGAGTGGATCATATAG		
chr11	ATM-ex62F	TTAATACATATGTTCTCTGTTTAGGTCCTT	108235782	108235911
	ATM-ex62R	TCAGAGTAGGGTGAAGTCAGTTT		
chr7	BRAF-ex15F	AAATAGCCTCAATTCTTACCATCCACAA	140453056	140453221
	BRAF-ex15R	CTGTTTTCTTTACTTACTACACCTCAGA		
chr16	CDH1-ex4F	CCTGAAGTATCCGTCTTGAATTGTCTTA	68842271	68842444
	CDH1-ex4R	CTTTTTCATTTTTCTGGGCAGCTGA		
chr5	CSF1R-ex7F	GGGTAGGCCTCCACCATGA	149452960	149453134
	CSF1R-ex7R	TGGGTGGATGACAAAATGGACAAATAA		
chr7	EGFR-ex20F	GGTCTTTTGACAGGCACAGCTTT	55248838	55249000
	EGFR-ex20R	CATCACGTAGGCTTCCTGGAG		
chr7	EGFR-ex20F	CCACACTGACGTGCCTCTC	55248956	55249123
	EGFR-ex20R	GTCTTTGTGTTCCCGGACATAGT		
chr7	EGFR-ex20F	ACGCAGCTCATGCCCTT	55249070	55249237
	EGFR-ex20R	GCGCAGACCCGATGTGA		
chr7	EGFR-ex28F	CCGAGTATCTCAACACTGTCCA	55273084	55273210
	EGFR-ex28R	GGAAAGAAGTCCCTGCTGGTAGTC		
chr7	EGFR-ex28F	GCTCTGTGCAGAATCCTGTCTA	55272985	55273130
	EGFR-ex28R	GAATGTGCTGTTGACACAGGT		
chr17	ERBB2-ex21F	TCGGAACGTGCTGGTCAAG	37881352	37881525
	ERBB2-ex21R	TCAGAGTTCTCCCATGGGCTA		
chr7	EZH2-ex17F	CCCAGCTCTGAAACATACCATTGT	148507407	148507558
	EZH2-ex17R	AGGCTTGATCACCTTTATCCAAAAGAAT		
chr7	EZH2-ex17F	CCCCATGCCTCTAAGGAGAT	148507294	148507453
	EZH2-ex17R	GTGCAGCTTTCTGTTCAACTTGA		
chr7	FZD1-ex1F	CTGCTGCTGCTGCTTTGG	90894358	90894532
	FZD1-ex1R	GGTCCGGGACGGAGATG		
chr20	GNAS-ex8F	CCAGACCTTTGCTTTAGATTGGCA	57484325	57484499
	GNAS-ex8R	AAAAGGTAACAGTTGGCTTACTGGAA		
chr12	HNF1A-ex1F	GCAGCCGAGCCATGGTT	121416561	121416721
	HNF1A-ex1R	GCAGGACTCCCCCTTGTC		
chr12	HNF1A-ex1F	CTACCTCCTGGCTGGAGAAG	121416676	121416817
	HNF1A-ex1R	CGTGAAGTCTTCCCACATCGT		
chr12	HNF1A-ex2F	CCTGGGCTCCATAACTGCTTT	121426502	121426670
	HNF1A-ex2R	ACTTGACCATCTTCGCCACAC		
chr12	HNF1A-ex2F	GTCGATACCACTGGCCTCAA	121426709	121426874
	HNF1A-ex2R	GCCCTCCAGGGAAGATG		
chr12	HNF1A-ex2F	CTGAGCAGATCCCGTCTTG	121426605	121426774
	HNF1A-ex2R	CTTCATGGGAGTGCCCTTGTT		
chr12	HNF1A-ex4F	CTGAGCCTGGCCTGGAG	121431926	121432092
	HNF1A-ex4R	TTGTGCCGGAAGGCTTCTT		
chr12	HNF1A-ex4F	GCTATTTCTGCAGGGCGGAAT	121431954	121432119
	HNF1A-ex4R	GGCCCCTGTACGTGTC		
chr11	HRAS-ex3F	CTTGGTGTGTTGATGGCAAACA	533792	533966
	HRAS-ex3R	CCTGAGCCCTGTCCCTCT		
chr5	IL6ST-ex6F	AAACGACTACAGTGTCAAATAAACTCT	55259926	55260095
	IL6ST-ex6R	CACCTCATGCACTGTTGATTATTCTACT		

Chr.	Gene	Ion_AmpliSeq Fwd_Rev_Primer*	Amplicon Start	Amplicon Stop
chr5	IL6ST-ex6F	CCCAGACTTCAATGTTGACAAAATACAC	55260040	55260185
	IL6ST-ex6R	TCAGATGTGTTTTTAATGACTGATTCTTCAATT		
chr4	IRF2-ex5F	CCTGTGACTCATCTTTACAAGGCA	185339237	185339409
	IRF2-ex5R	CCCTTCGTGAATTAATCCAAAACACTACAT		
chr9	JAK2-ex8F	AATGGCTCTGTAAATTCTACCGTTT	5055617	5055756
	JAK2-ex8R	CAACTCGGCTTTCATTTGAACCC		
chr9	JAK2-ex8F	TGTCAGTATTAAGCAAGCAAACCAAGA	5055707	5055847
	JAK2-ex8R	AATACACATTTCTGAGAACTATAAAATGAAACC		
chrX	KDM6A-ex29F	TTTTATGTTGCATAGCAGGCTGTTG	44970493	44970661
	KDM6A-ex29R	CAATATCAAGATGAGGCGGATGGTAA		
chrX	KDM6A-ex29F	CCTTACTTTTTATTTTCAGGCTCCTCCA	44970609	44970748
	KDM6A-ex29R	GCCTTACGAGATAGCTACAAAAACCA		
chr12	KRAS-ex2F	AAAGAATGGTCTGCACCAGTAA	25398161	25398331
	KRAS-ex2R	AGGCCTGCTGAAAATGACTGAATATAA		
chr12	KRAS-ex2F	CCAGCTCCAACCTACCACAAGT	25398284	25398410
	KRAS-ex2R	CTGGTGGAGTATTTGATAGTGTATTAACCTT		
chr12	KRAS-ex3F	TTCAATTTAAACCCACCTATAATGGTGA	25380152	25380286
	KRAS-ex3R	ACAGCAGGTCAAGAGGAGTACA		
chr12	KRAS-ex3F	ATGTA CTGGTCCCTCATTGCAC	25380243	25380372
	KRAS-ex3R	TCCAGACTGTGTTTCTCCCTTCT		
chr7	MET-ex17F	GCCTGCGAAGTGAAGGGT	116418926	116419098
	MET-ex17R	TTTTTAAAGACTCAGAGCAGGCCTATTT		
chr7	MET-ex19F	GTTGCTGATTTTGGTCTTGCCA	116423383	116423557
	MET-ex19R	AAGAGGAGAAACTCAGAGATAACCA		
chr7	MLL3-ex38F	GAACCAGAGTTTTGTTTTCTTGTTCCTT	151874122	151874249
	MLL3-ex38R	ACTTGACATGGGAGATAAGAAAAGCAT		
chr7	MLL3-ex38F	AGTTTCACATTTGGATTCCACCTTAGA	151874021	151874179
	MLL3-ex38R	AATCAGTGTGTATCTGTTGAACCAAAAA		
chr20	MMP9-ex4F	GGGTATCCCTTCGACGGGAA	44639572	44639746
	MMP9-ex4R	GGTGATGAGTGGGAGAGAATGAAG		
chr17	NF1-ex26F	CCATTCACACCATGCACATATGATT	29559619	29559786
	NF1-ex26R	CCTGTTTGCGCACTTTCATCTTC		
chr17	NF1-ex26F	CCTTTTGAATGACTGCAGTGAAGTT	29559739	29559884
	NF1-ex26R	GACCACTGTCTACGTTGGCATT		
chr2	NF2-ex8F	GACTGCTTCTTTCTGCCTGAGA	30057087	30057231
	NF2-ex8R	CTGGCTACTGCTACTTGGCAT		
chr2	NF2-ex8F	AGGTTTTGTGGGCTTGACACA	30057190	30057317
	NF2-ex8R	GTA CTTCCCCTAGAATTTTCTGCCT		
chr2	NFE2L2-ex2F	TTAAACCTGCCATAACTTTCCCAAGA	178098691	178098827
	NFE2L2-ex2R	AGTTACA ACTAGATGAAGAGACAGGTGA		
chr2	NFE2L2-ex2F	CTCTTTCCGTCGCTGACTGA	178098910	178099040
	NFE2L2-ex2R	AAACATGAGCTCTCTCCTTCTTTTT		
chr2	NFE2L2-ex2F	GGCTGGCTGAATTGGGAGAAAT	178098778	178098917
	NFE2L2-ex2R	GGAAAGAGTATGAGCTGGAAAAACAGAA		
chr1	NOTCH2-ex4F	AGGTGCCCTCCATTGACACAA	120539665	120539839
	NOTCH2-ex4R	AAATGTGAGACTGATGTCAATGAGTGT		
chr19	NOTCH3-ex15F	ACACAGATGCAGCAGTCCA	15302476	15302640
	NOTCH3-ex15R	TGTCCAGGACACCGATGTCT		
chr4	NRAS-ex3F	GACAGAAA ACTAGGTCAATAGGTCTTGG	115256475	115256619
	NRAS-ex3R	CAGGCTAGAAGAAAGGCAGATACAAG		
chr4	NRAS-ex3F	TATAATCCTTTCAGGACTGACTGTATGA	115256381	115256530

Chr.	Gene	Ion_AmpliSeq Fwd_Rev_Primer*	Amplicon Start	Amplicon Stop
	NRAS-ex3R	CCTACACAATGCCAGATTCTAGAAAACA		
chr4	PDGFRA-ex18F	GAAGGCCAGCCCTTTATATCCAG	55151881	55152009
	PDGFRA-ex18R	CACTGCATGGAAAAGGAAGAAATGA		
chr4	PDGFRA-ex18F	GCTACAGATGGCTTGATCCTGAG	55151962	55152135
	PDGFRA-ex18R	CGTACACTGCCTTTTCGACACATAG		
chr4	PDGFRA-ex22F	AATGCATACATTGGTGTACCTACA	55156555	55156727
	PDGFRA-ex22R	AGCTACCTGTGTCTGTTCTCT		
chr3	PIK3CA-ex10F	GAGTAACAGACTAGCTAGAGACAATGA	178935997	178936129
	PIK3CA-ex10R	CACTTACCTGTGACTCCATAGAAAATCT		
chr3	PIK3CA-ex10F	CGAGATCCTCTCTCTGAAATCACTGA	178936067	178936202
	PIK3CA-ex10R	ACATGCTGAGATCAGCCAAATT		
chr3	PIK3CA-ex21F	TCGACAGCATGCCAATCTCTT	178951918	178952091
	PIK3CA-ex21R	CCATGATGTGCATCATTCATTTGTTTCA		
chr3	PIK3CA-ex21F	ATCATGGTGGCTGGACAACAA	178952085	178952218
	PIK3CA-ex21R	CTGCTGAGAGTTATTAACAGTGCAGT		
chr3	PIK3CA-ex21F	TGGAATGCCAGAACTACAATCTTTTGA	178951969	178952129
	PIK3CA-ex21R	TTGTGTGGAAGATCCAATCCATT		
chr3	PIK3CA-ex3F	ATTGTGATCTTCCAAATCTACAGAGTTC	178917376	178917540
	PIK3CA-ex3R	AGTCCTGTACTTCTGGATCTTTAACCAT		
chr3	PIK3CA-ex3F	GCATGCCAGTGTGTGAATTTGAT	178917490	178917614
	PIK3CA-ex3R	TGCTCTACTATGAGGTGAATGAGGT		
chr10	PTEN-ex2F	AGTATTCTTTTAGTTTGATTGCTGCATATTTCA	89653719	89653843
	PTEN-ex2R	CCTGTATACGCCTTCAAGTCTTTCT		
chr10	PTEN-ex2F	TCCAAACATTATTGCTATGGGATTTCCCT	89653789	89653959
	PTEN-ex2R	ATCTTTTTCTGTGGCTTAGAAATCTTTTC		
chr10	PTEN-ex4F	AACTTTCTTTTAGTTGTGCTGAAAGACATT	89690789	89690943
	PTEN-ex4R	TCACTCGATAATCTGGATGACTCATT		
chr10	PTEN-ex5F	GAAAGGGACGAACTGGTGAATGAT	89692896	89693061
	PTEN-ex5R	ATAAATTCTCAGATCCAGGAAGAGGAAAG		
chr10	PTEN-ex5F	GCACAATATCCTTTTGAAGACCATAAC	89692772	89692946
	PTEN-ex5R	TGCCCCGATGTAATAAATATGCACAT		
chr10	PTEN-ex7F	CTGTGTGTGGTGATACAAAGTAGAGT	89717718	89717892
	PTEN-ex7R	ACACCTGCAGATCTAATAGAAAACAAATT		
chr10	PTEN-ex7F	ATTTAACCATGCAGATCCTCAGTTT	89717596	89717769
	PTEN-ex7R	AGCATCTTGTCTGTTTGTGGAAGA		
chr10	PTEN-ex8F	AAGAAATCGATAGCATTTCAGTATAGA	89720742	89720866
	PTEN-ex8R	TGGAGAAAAGTATCGGTTGGCTTT		
chr10	PTEN-ex8F	TCACTTTTGGGTAAAATACATTCTTCATACCA	89720662	89720794
	PTEN-ex8R	ATATTCCTTGTCAATTATCTGCACGC		
chr13	RB-ex12F	TTTCTCCCTTCATTGCTTAACACATTTTC	48947493	48947619
	RB-ex12R	GGAAATCAGATTTTCTGAAGTTGATCAC		
chr13	RB-ex2F	TCATTTGGTAGGCTTGAGTTGAAGAAA	48881405	48881565
	RB-ex2R	AAACGTTTTAAGAAAATCCTTACCAATACTC		
chr18	SMAD2-ex10F	ACTCCAGAATATGCAAGAATGCAATGA	45371636	45371809
	SMAD2-ex10R	GGCTCAGTCTGTTAATCAGGGT		
chr18	SMAD4-ex10F	TCAATATTAAGCATGCTATACAATCTGAACTAA	48593301	48593445
	SMAD2-ex10R	ACCAAACATCACCTTACCTTTAC		
chr18	SMAD4-ex10F	GCAAAGGTGTGCAGTTGGAAT	48593400	48593535
	SMAD2-ex10R	ATCTTATGAACAGCATCTCCAGGTG		
chr18	SMAD4-ex9F	GTTTTGGGTGCATTACATTTCCATCT	48591661	48591830
	SMAD4-ex9R	CATTTCAAAGTAAGCAATGGAACACCAA		

Chr.	Gene	Ion_AmpliSeq Fwd_Rev_Primer*	Amplicon Start	Amplicon Stop
chr18	SMAD4-ex9F	CCCATTTATTTCTATAGCTCCTGAGTA	48591775	48591948
	SMAD4-ex9R	GGACATTGGAGAGTTGACCCA		

## 7.2 Sequencing data

**Table S3:** A comprehensive data of mt-genome sequencing. X<sup>1</sup>: represents the variants which validated by Sanger sequencing

Sample ID	Aetiology	Cirrhosis	Gene Cards	Region	Reference	Allele	Zygoty	Frequency
H41-TN1	HBV	-	D-LOOP	242 <sup>1</sup>	C	T	Heterozygous	95.2
H41-TN3	HBV	-	D-LOOP	242 <sup>1</sup>	C	T	Homozygous	100.0
H04-TN3	HCV	+	D-LOOP	462 <sup>1</sup>	C	T	Homozygous	100.0
H41-TN1	HBV	-	D-LOOP	462 <sup>1</sup>	C	T	Heterozygous	93.3
H41-TN2	HBV	-	D-LOOP	462 <sup>1</sup>	C	T	Homozygous	100.0
H41-TN3	HBV	-	D-LOOP	462 <sup>1</sup>	C	T	Homozygous	100.0
H41-TN5	HBV	-	D-LOOP	462 <sup>1</sup>	C	T	Homozygous	100.0
H04-TN3	HCV	+	D-LOOP	489 <sup>1</sup>	T	C	Homozygous	100.0
H41-TN1	HBV	-	D-LOOP	489 <sup>1</sup>	T	C	Homozygous	100.0
H41-TN2	HBV	-	D-LOOP	489 <sup>1</sup>	T	C	Homozygous	100.0
H41-TN3	HBV	-	D-LOOP	489 <sup>1</sup>	T	C	Homozygous	97.7
H41-TN5	HBV	-	D-LOOP	489 <sup>1</sup>	T	C	Homozygous	100.0
H02-T+OT	Toxic	-	RNR1	709 <sup>1</sup>	G	A	Homozygous	98.6
H02-TN1	Toxic	-	RNR1	709 <sup>1</sup>	G	A	Homozygous	98.7
H02-TN2	Toxic	-	RNR1	709 <sup>1</sup>	G	A	Homozygous	99.1
H02-TN3	Toxic	-	RNR1	709 <sup>1</sup>	G	A	Homozygous	99.4
H03-OT	n.d	-	RNR1	709 <sup>1</sup>	G	A	Homozygous	100.0
H03-TN1	n.d	-	RNR1	709 <sup>1</sup>	G	A	Homozygous	100.0
H03-TN2	n.d	-	RNR1	709 <sup>1</sup>	G	A	Homozygous	98.3
H03-TN3	n.d	-	RNR1	709 <sup>1</sup>	G	A	Heterozygous	94.1
H11-OT	HCV	+	RNR1	709 <sup>1</sup>	G	A	Heterozygous	17.0
H28-OT	SH	-	RNR1	709 <sup>1</sup>	G	A	Homozygous	99.2
H28-TN1	SH	-	RNR1	709 <sup>1</sup>	G	A	Homozygous	99.2

## Appendix

Sample ID	Aetiology	Cirrhosis	Gene Cards	Region	Reference	Allele	Zygoty	Frequency
H28-TN2	SH	-	RNR1	709 <sup>1</sup>	G	A	Homozygous	99.3
H28-TN3	SH	-	RNR1	709 <sup>1</sup>	G	A	Homozygous	99.1
H03-OT	n.d	-	RNR1	930 <sup>1</sup>	G	A	Homozygous	100.0
H03-TN1	n.d	-	RNR1	930 <sup>1</sup>	G	A	Homozygous	100.0
H03-TN3	n.d	-	RNR1	930 <sup>1</sup>	G	A	Homozygous	100.0
H28-TN3	SH	-	RNR1	930 <sup>1</sup>	G	A	Homozygous	99.3
H28-TN1	SH	-	TV	1611 <sup>1</sup>	G	A	Heterozygous	50.3
H28-TN2	SH	-	TV	1611 <sup>1</sup>	G	A	Heterozygous	5.3
H02-T+OT	Toxic	-	RNR2	1888 <sup>1</sup>	G	A	Homozygous	97.2
H02-TN1	Toxic	-	RNR2	1888 <sup>1</sup>	G	A	Homozygous	98.0
H02-TN2	Toxic	-	RNR2	1888 <sup>1</sup>	G	A	Homozygous	98.5
H02-TN3	Toxic	-	RNR2	1888 <sup>1</sup>	G	A	Homozygous	98.9
H28-OT	SH	-	RNR2	1888 <sup>1</sup>	G	A	Homozygous	99.1
H28-TN1	SH	-	RNR2	1888 <sup>1</sup>	G	A	Homozygous	98.7
H28-TN2	SH	-	RNR2	1888 <sup>1</sup>	G	A	Homozygous	97.5
H28-TN3	SH	-	RNR2	1888 <sup>1</sup>	G	A	Homozygous	98.8
H41-TN1	HBV	-	RNR2	2158 <sup>1</sup>	T	C	Heterozygous	90.0
H41-TN2	HBV	-	RNR2	2158 <sup>1</sup>	T	C	Homozygous	100.0
H41-TN3	HBV	-	RNR2	2158 <sup>1</sup>	T	C	Homozygous	100.0
H41-TN5	HBV	-	RNR2	2158 <sup>1</sup>	T	C	Homozygous	100.0
C10661	Control	-	RNR2	2706 <sup>1</sup>	A	G	Heterozygous	74.0
H03-TN2	n.d	-	RNR2	2706 <sup>1</sup>	A	G	Homozygous	97.1
H03-TN3	n.d	-	RNR2	2706 <sup>1</sup>	A	G	Heterozygous	92.3
H04-TN3	HCV	+	RNR2	2706 <sup>1</sup>	A	G	Homozygous	100.0
H40-TN3	Toxic	-	RNR2	2706 <sup>1</sup>	A	G	Homozygous	100.0
H41-TN1	HBV	-	RNR2	2706 <sup>1</sup>	A	G	Homozygous	100.0
H41-TN2	HBV	-	RNR2	2706 <sup>1</sup>	A	G	Homozygous	100.0
H41-TN3	HBV	-	RNR2	2706 <sup>1</sup>	A	G	Heterozygous	93.2
H41-TN4	HBV	-	RNR2	2706 <sup>1</sup>	A	G	Homozygous	100.0

## Appendix

Sample ID	Aetiology	Cirrhosis	Gene Cards	Region	Reference	Allele	Zygoty	Frequency
H41-TN5	HBV	-	RNR2	2706 <sup>1</sup>	A	G	Homozygous	100.0
C10661	Control	-	RNR2	3010 <sup>1</sup>	G	A	Heterozygous	34.0
H04-OT	HCV	+	RNR2	3010 <sup>1</sup>	G	A	Homozygous	100.0
H04-TN2	HCV	+	RNR2	3010 <sup>1</sup>	G	A	Homozygous	99.0
H04-TN3	HCV	+	RNR2	3010 <sup>1</sup>	G	A	Homozygous	99.0
H04-TN4	HCV	+	RNR2	3010 <sup>1</sup>	G	A	Heterozygous	95.2
H04-TN5	HCV	+	RNR2	3010 <sup>1</sup>	G	A	Homozygous	100.0
H04-TN6	HCV	+	RNR2	3010 <sup>1</sup>	G	A	Heterozygous	96.6
H32-OT	HCV	+	RNR2	3010 <sup>1</sup>	G	A	Homozygous	97.4
H32-TN1	HCV	+	RNR2	3010 <sup>1</sup>	G	A	Homozygous	99.2
H32-TN2	HCV	+	RNR2	3010 <sup>1</sup>	G	A	Homozygous	98.1
H32-TN3	HCV	+	RNR2	3010 <sup>1</sup>	G	A	Homozygous	98.5
H37-OT	n.d	-	RNR2	3010 <sup>1</sup>	G	A	Heterozygous	28.0
H38-OT	n.d	-	RNR2	3010 <sup>1</sup>	G	A	Heterozygous	58.0
H38-TN1	n.d	-	RNR2	3010 <sup>1</sup>	G	A	Homozygous	98.6
H38-TN2	n.d	-	RNR2	3010 <sup>1</sup>	G	A	Homozygous	97.9
H38-TN3	n.d	+	RNR2	3010 <sup>1</sup>	G	A	Heterozygous	96.9
H38-TN4	n.d	-	RNR2	3010 <sup>1</sup>	G	A	Homozygous	98.5
H40-OT	Toxic	-	RNR2	3010 <sup>1</sup>	G	A	Homozygous	99.2
H41-OT	HBV	-	RNR2	3010 <sup>1</sup>	G	A	Homozygous	100.0
H41-TN1	HBV	-	RNR2	3010 <sup>1</sup>	G	A	Homozygous	99.1
H41-TN2	HBV	-	RNR2	3010 <sup>1</sup>	G	A	Homozygous	97.2
H41-TN3	HBV	-	RNR2	3010 <sup>1</sup>	G	A	Homozygous	100.0
H41-TN5	HBV	-	RNR2	3010 <sup>1</sup>	G	A	Homozygous	97.1
H02-T+OT	Toxic	-	ND1	4216 <sup>1</sup>	T	C	Homozygous	98.8
H02-TN1	Toxic	-	ND1	4216 <sup>1</sup>	T	C	Homozygous	99.5
H02-TN2	Toxic	-	ND1	4216 <sup>1</sup>	T	C	Homozygous	99.6
H02-TN3	Toxic	-	ND1	4216 <sup>1</sup>	T	C	Homozygous	99.9
H04-TN3	HCV	+	ND1	4216 <sup>1</sup>	T	C	Homozygous	100.0
H28-OT	STE	-	ND1	4216 <sup>1</sup>	T	C	Homozygous	99.1

## Appendix

Sample ID	Aetiology	Cirrhosis	Gene Cards	Region	Reference	Allele	Zygoty	Frequency
H28-TN1	SH	-	ND1	4216 <sup>1</sup>	T	C	Homozygous	99.7
H28-TN2	SH	-	ND1	4216 <sup>1</sup>	T	C	Homozygous	98.3
H28-TN3	SH	-	ND1	4216 <sup>1</sup>	T	C	Homozygous	99.0
H41-TN1	HBV	-	ND1	4216 <sup>1</sup>	T	C	Homozygous	100.0
H41-TN3	HBV	-	ND1	4216 <sup>1</sup>	T	C	Homozygous	100.0
H11-TN1	HCV	+	ND1	4261	T	C	Heterozygous	38.5
H11-TN2	HCV	+	ND1	4261	T	C	Heterozygous	48.0
H11-TN3	HCV	+	ND1	4261	T	C	Heterozygous	46.1
H03-OT	n.d	-	ND2	4823 <sup>1</sup>	T	C	Homozygous	100.0
H03-TN1	n.d	-	ND2	4823 <sup>1</sup>	T	C	Homozygous	100.0
H03-TN2	n.d	-	ND2	4823 <sup>1</sup>	T	C	Homozygous	100.0
H02-T+OT	Toxic	-	ND2	4917 <sup>1</sup>	A	G	Homozygous	98.1
H02-TN1	Toxic	-	ND2	4917 <sup>1</sup>	A	G	Homozygous	98.5
H02-TN2	Toxic	-	ND2	4917 <sup>1</sup>	A	G	Homozygous	99.0
H02-TN3	Toxic	+	ND2	4917 <sup>1</sup>	A	G	Homozygous	99.4
H03-OT	n.d	-	ND2	4917 <sup>1</sup>	A	G	Homozygous	100.0
H03-TN1	n.d	-	ND2	4917 <sup>1</sup>	A	G	Homozygous	100.0
H03-TN2	n.d	-	ND2	4917 <sup>1</sup>	A	G	Homozygous	98.0
H03-TN3	n.d	-	ND2	4917 <sup>1</sup>	A	G	Homozygous	100.0
H28-OT	STE	-	ND2	4917 <sup>1</sup>	A	G	Homozygous	98.2
H28-TN1	SH	-	ND2	4917 <sup>1</sup>	A	G	Homozygous	99.1
H28-TN2	SH	-	ND2	4917 <sup>1</sup>	A	G	Homozygous	97.1
H28-TN3	SH	-	ND2	4917 <sup>1</sup>	A	G	Homozygous	99.4
H02-T+OT	Toxic	-	ND2	5120 <sup>1</sup>	A	G	Heterozygous	22.7
H02-TN2	Toxic	-	ND2	5120 <sup>1</sup>	A	G	Heterozygous	41.2
H02-TN3	Toxic	-	ND2	5120 <sup>1</sup>	A	G	Heterozygous	37.0
H04-TN3	HCV	+	ND2	5460 <sup>1</sup>	G	A	Homozygous	98.4
H41-OT	HBV	-	ND2	5460 <sup>1</sup>	G	A	Homozygous	97.6
H41-TN2	HBV	-	ND2	5460 <sup>1</sup>	G	A	Homozygous	97.6
H41-TN3	HBV	-	ND2	5460 <sup>1</sup>	G	A	Heterozygous	96.1

## Appendix

Sample ID	Aetiology	Cirrhosis	Gene Cards	Region	Reference	Allele	Zygoty	Frequency
H41-TN5	HBV	-	ND2	5460 <sup>1</sup>	G	A	Homozygous	98.2
C633	Control	-	CO1	6776 <sup>1</sup>	T	C	Homozygous	99.5
H37-OT	n.d	-	CO1	6776 <sup>1</sup>	T	C	Heterozygous	78.3
H37-TN1	n.d	-	CO1	6776 <sup>1</sup>	T	C	Homozygous	97.9
H37-TN2	n.d	-	CO1	6776 <sup>1</sup>	T	C	Homozygous	98.9
H37-TN3	n.d	-	CO1	6776 <sup>1</sup>	T	C	Homozygous	98.7
H38-OT	n.d	-	CO1	6776 <sup>1</sup>	T	C	Heterozygous	49.4
C10661	Control	-	CO1	7028 <sup>1</sup>	C	T	Heterozygous	71.5
H02-T+OT	Toxic	-	CO1	7028 <sup>1</sup>	C	T	Homozygous	97.7
H02-TN1	Toxic	-	CO1	7028 <sup>1</sup>	C	T	Homozygous	98.1
H02-TN2	Toxic	-	CO1	7028 <sup>1</sup>	C	T	Homozygous	98.3
H02-TN3	Toxic	-	CO1	7028 <sup>1</sup>	C	T	Homozygous	99.1
H04-TN3	HCV	+	CO1	7028 <sup>1</sup>	C	T	Homozygous	99.1
H28-OT	STE	-	CO1	7028 <sup>1</sup>	C	T	Homozygous	98.2
H28-TN1	SH	-	CO1	7028 <sup>1</sup>	C	T	Homozygous	98.5
H28-TN2	SH	-	CO1	7028 <sup>1</sup>	C	T	Homozygous	98.0
H28-TN3	SH	-	CO1	7028 <sup>1</sup>	C	T	Homozygous	98.6
H40-TN1	Toxic	-	CO1	7028 <sup>1</sup>	C	T	Heterozygous	93.3
H40-TN2	Toxic	-	CO1	7028 <sup>1</sup>	C	T	Homozygous	97.3
H40-TN3	Toxic	-	CO1	7028 <sup>1</sup>	C	T	Homozygous	99.2
H40-TN4	Toxic	-	CO1	7028 <sup>1</sup>	C	T	Homozygous	98.2
H40-TN5	Toxic	-	CO1	7028 <sup>1</sup>	C	T	Homozygous	99.4
H40-TN6	Toxic	-	CO1	7028 <sup>1</sup>	C	T	Homozygous	97.3
H41-OT	HBV	-	CO1	7028 <sup>1</sup>	C	T	Homozygous	98.1
H41-TN1	HBV	-	CO1	7028 <sup>1</sup>	C	T	Homozygous	99.3
H41-TN2	HBV	-	CO1	7028 <sup>1</sup>	C	T	Homozygous	97.1
H41-TN3	HBV	-	CO1	7028 <sup>1</sup>	C	T	Homozygous	99.0
H41-TN4	HBV	-	CO1	7028 <sup>1</sup>	C	T	Homozygous	98.9
H41-TN5	HBV	-	CO1	7028 <sup>1</sup>	C	T	Homozygous	99.4

Sample ID	Aetiology	Cirrhosis	Gene Cards	Region	Reference	Allele	Zygoty	Frequency
H37-OT	n.d	-	CO1	7153 <sup>1</sup>	T	C	Heterozygous	9.1
H37-TN2	n.d	-	CO1	7153 <sup>1</sup>	T	C	Heterozygous	15.0
H37-TN3	n.d	-	CO1	7153 <sup>1</sup>	T	C	Heterozygous	10.7
H38-OT	n.d	-	CO1	7153 <sup>1</sup>	T	C	Heterozygous	6.8
H02-T+OT	Toxic	-	ATP6	8697 <sup>1</sup>	G	A	Homozygous	98.2
H02-TN1	Toxic	-	ATP6	8697 <sup>1</sup>	G	A	Homozygous	98.9
H02-TN2	Toxic	-	ATP6	8697 <sup>1</sup>	G	A	Homozygous	98.8
H02-TN3	Toxic	-	ATP6	8697 <sup>1</sup>	G	A	Homozygous	99.3
H03-OT	n.d	-	ATP6	8697 <sup>1</sup>	G	A	Homozygous	100.0
H03-TN1	n.d	-	ATP6	8697 <sup>1</sup>	G	A	Homozygous	100.0
H03-TN2	n.d	-	ATP6	8697 <sup>1</sup>	G	A	Homozygous	99.0
H03-TN3	n.d	-	ATP6	8697 <sup>1</sup>	G	A	Homozygous	100.0
H28-OT	STE	-	ATP6	8697 <sup>1</sup>	G	A	Homozygous	98.5
H28-TN1	SH	-	ATP6	8697 <sup>1</sup>	G	A	Homozygous	99.2
H28-TN2	SH	-	ATP6	8697 <sup>1</sup>	G	A	Homozygous	98.8
H28-TN3	SH	-	ATP6	8697 <sup>1</sup>	G	A	Homozygous	98.8
H04-TN3	HCV	+	ATP6	8752 <sup>1</sup>	A	G	Homozygous	97.4
H41-OT	HBV	-	ATP6	8752 <sup>1</sup>	A	G	Homozygous	100.0
H41-TN1	HBV	-	ATP6	8752 <sup>1</sup>	A	G	Homozygous	97.8
H41-TN2	HBV	-	ATP6	8752 <sup>1</sup>	A	G	Homozygous	98.2
H41-TN3	HBV	-	ATP6	8752 <sup>1</sup>	A	G	Homozygous	98.1
H41-TN5	HBV	-	ATP6	8752 <sup>1</sup>	A	G	Homozygous	98.2
H28-OT	STE	-	CO3	9843 <sup>1</sup>	A	G	Homozygous	98.8
H28-TN1	SH	-	CO3	9843 <sup>1</sup>	A	G	Homozygous	98.9
H28-TN2	SH	-	CO3	9843 <sup>1</sup>	A	G	Homozygous	97.7
H28-TN3	SH	-	CO3	9843 <sup>1</sup>	A	G	Homozygous	98.7
H02-T+OT	Toxic	-	TR	10463 <sup>1</sup>	T	C	Heterozygous	96.2
H02-TN1	Toxic	-	TR	10463 <sup>1</sup>	T	C	Homozygous	100.0
H02-TN2	Toxic	-	TR	10463 <sup>1</sup>	T	C	Homozygous	98.5
H02-TN3	Toxic	-	TR	10463 <sup>1</sup>	T	C	Homozygous	99.9

## Appendix

Sample ID	Aetiology	Cirrhosis	Gene Cards	Region	Reference	Allele	Zygoty	Frequency
H03-TN2	n.d	-	TR	10463 <sup>1</sup>	T	C	Homozygous	100.0
H28-OT	SH	-	TR	10463 <sup>1</sup>	T	C	Homozygous	100.0
H28-TN1	SH	-	TR	10463 <sup>1</sup>	T	C	Homozygous	100.0
H28-TN2	SH	-	TR	10463 <sup>1</sup>	T	C	Homozygous	100.0
H28-TN3	SH	-	TR	10463 <sup>1</sup>	T	C	Homozygous	99.4
H02-T+OT	Toxic	-	ND4	11165 <sup>1</sup>	T	C	Heterozygous	28.7
H02-TN1	Toxic	-	ND4	11165 <sup>1</sup>	T	C	Heterozygous	39.0
H02-TN2	Toxic	-	ND4	11165 <sup>1</sup>	T	C	Heterozygous	41.3
H02-TN3	Toxic	-	ND4	11165 <sup>1</sup>	T	C	Heterozygous	56.2
H02-T+OT	Toxic	-	ND4	11251 <sup>1</sup>	A	G	Homozygous	98.7
H02-TN1	Toxic	-	ND4	11251 <sup>1</sup>	A	G	Homozygous	99.2
H02-TN2	Toxic	-	ND4	11251 <sup>1</sup>	A	G	Homozygous	98.8
H02-TN3	Toxic	-	ND4	11251 <sup>1</sup>	A	G	Homozygous	99.4
H03-OT	n.d	-	ND4	11251 <sup>1</sup>	A	G	Homozygous	100.0
H03-TN1	n.d	-	ND4	11251 <sup>1</sup>	A	G	Heterozygous	94.1
H03-TN2	n.d	-	ND4	11251 <sup>1</sup>	A	G	Homozygous	100.0
H03-TN3	n.d	-	ND4	11251 <sup>1</sup>	A	G	Homozygous	100.0
H04-TN3	HCV	+	ND4	11251 <sup>1</sup>	A	G	Homozygous	99.2
H28-OT	STE	-	ND4	11251 <sup>1</sup>	A	G	Homozygous	99.0
H28-TN1	SH	-	ND4	11251 <sup>1</sup>	A	G	Homozygous	99.2
H28-TN2	SH	-	ND4	11251 <sup>1</sup>	A	G	Homozygous	98.1
H28-TN3	SH	-	ND4	11251 <sup>1</sup>	A	G	Homozygous	99.1
H41-OT	HBV	-	ND4	11251 <sup>1</sup>	A	G	Heterozygous	93.3
H41-TN1	HBV	-	ND4	11251 <sup>1</sup>	A	G	Homozygous	100.0
H41-TN2	HBV	-	ND4	11251 <sup>1</sup>	A	G	Homozygous	100.0
H41-TN3	HBV	-	ND4	11251 <sup>1</sup>	A	G	Homozygous	98.8
H41-TN5	HBV	-	ND4	11251 <sup>1</sup>	A	G	Heterozygous	96.2
H41-TN3	HBV	-	ND4	11253 <sup>1</sup>	T	C	Heterozygous	13.7
H41-TN5	HBV	-	ND4	11253 <sup>1</sup>	T	C	Heterozygous	10.3

## Appendix

Sample ID	Aetiology	Cirrhosis	Gene Cards	Region	Reference	Allele	Zygoty	Frequency
C10661	Control	-	ND4	11719 <sup>1</sup>	G	A	Heterozygous	64.3
H02-TN1	Toxic	-	ND4	11719 <sup>1</sup>	G	A	Homozygous	99.0
H02-TN2	Toxic	-	ND4	11719 <sup>1</sup>	G	A	Homozygous	99.0
H02-TN3	Toxic	-	ND4	11719 <sup>1</sup>	G	A	Homozygous	99.5
H03-OT	n.d	-	ND4	11719 <sup>1</sup>	G	A	Homozygous	98.4
H03-TN1	n.d	-	ND4	11719 <sup>1</sup>	G	A	Homozygous	98.3
H03-TN3	n.d	-	ND4	11719 <sup>1</sup>	G	A	Homozygous	100.0
H04-TN3	HCV	+	ND4	11719 <sup>1</sup>	G	A	Homozygous	97.1
H28-TN1	SH	-	ND4	11719 <sup>1</sup>	G	A	Homozygous	99.4
H28-TN3	SH	-	ND4	11719 <sup>1</sup>	G	A	Homozygous	98.8
H41-OT	HBV	-	ND4	11719 <sup>1</sup>	G	A	Homozygous	100.0
H41-TN1	HBV	-	ND4	11719 <sup>1</sup>	G	A	Homozygous	97.1
H41-TN2	HBV	-	ND4	11719 <sup>1</sup>	G	A	Homozygous	100.0
H41-TN3	HBV	-	ND4	11719 <sup>1</sup>	G	A	Homozygous	98.1
H41-TN5	HBV	-	ND4	11719 <sup>1</sup>	G	A	Homozygous	98.1
H02-T+OT	Toxic	-	ND4	11812 <sup>1</sup>	A	G	Homozygous	98.3
H02-TN1	Toxic	-	ND4	11812 <sup>1</sup>	A	G	Homozygous	99.1
H02-TN2	Toxic	-	ND4	11812 <sup>1</sup>	A	G	Homozygous	98.6
H02-TN3	Toxic	-	ND4	11812 <sup>1</sup>	A	G	Homozygous	98.7
H03-TN2	n.d	-	ND4	11812 <sup>1</sup>	A	G	Homozygous	100.0
H28-OT	SH	-	ND4	11812 <sup>1</sup>	A	G	Homozygous	98.0
H28-TN1	SH	-	ND4	11812 <sup>1</sup>	A	G	Homozygous	98.3
H28-TN2	SH	-	ND4	11812 <sup>1</sup>	A	G	Homozygous	98.7
H28-TN3	SH	-	ND4	11812 <sup>1</sup>	A	G	Homozygous	98.2
H04-TN3	HCV	+	ND4	12007	G	A	Homozygous	97.2
H41-TN1	HBV	-	ND4	12007	G	A	Homozygous	98.0
H41-TN2	HBV	-	ND4	12007	G	A	Homozygous	97.6
H41-TN3	HBV	-	ND4	12007	G	A	Homozygous	99.1
H41-TN5	HBV	-	ND4	12007	G	A	Homozygous	98.0
H28-TN1	SH	-	ND5	13231 <sup>1</sup>	A	-	Heterozygous	72.7

## Appendix

Sample ID	Aetiology	Cirrhosis	Gene Cards	Region	Reference	Allele	Zygoty	Frequency
H28-TN2	SH	-	ND5	13231 <sup>1</sup>	A	-	Heterozygous	43.3
H28-TN3	SH	-	ND5	13231 <sup>1</sup>	A	-	Heterozygous	8.4
H02-T+OT	Toxic	-	ND5	13268 <sup>1</sup>	G	A	Heterozygous	21.8
H02-TN1	Toxic	-	ND5	13268 <sup>1</sup>	G	A	Heterozygous	53.2
H02-TN2	Toxic	-	ND5	13268 <sup>1</sup>	G	A	Heterozygous	50.8
H02-TN3	Toxic	-	ND5	13268 <sup>1</sup>	G	A	Heterozygous	37.0
H02-T+OT	Toxic	-	ND5	13368 <sup>1</sup>	G	A	Homozygous	99.3
H02-TN1	Toxic	-	ND5	13368 <sup>1</sup>	G	A	Homozygous	98.9
H02-TN2	Toxic	-	ND5	13368 <sup>1</sup>	G	A	Homozygous	99.7
H02-TN3	Toxic	-	ND5	13368 <sup>1</sup>	G	A	Homozygous	99.7
H03-OT	n.d	-	ND5	13368 <sup>1</sup>	G	A	Homozygous	100.0
H03-TN1	n.d	-	ND5	13368 <sup>1</sup>	G	A	Heterozygous	90.0
H03-TN2	n.d	-	ND5	13368 <sup>1</sup>	G	A	Homozygous	98.9
H03-TN3	n.d	-	ND5	13368 <sup>1</sup>	G	A	Homozygous	100.0
H28-OT	SH	-	ND5	13368 <sup>1</sup>	G	A	Homozygous	99.5
H28-TN1	SH	-	ND5	13368 <sup>1</sup>	G	A	Homozygous	99.6
H28-TN2	SH	-	ND5	13368 <sup>1</sup>	G	A	Homozygous	99.5
H28-TN3	SH	-	ND5	13368 <sup>1</sup>	G	A	Homozygous	99.4
H04-TN3	HCV	+	ND5	13605 <sup>1</sup>	C	T	Homozygous	100.0
H41-OT	HBV	-	ND5	13605 <sup>1</sup>	C	T	Homozygous	100.0
H41-TN1	HBV	-	ND5	13605 <sup>1</sup>	C	T	Homozygous	100.0
H41-TN2	HBV	-	ND5	13605 <sup>1</sup>	C	T	Heterozygous	96.4
H41-TN3	HBV	-	ND5	13605 <sup>1</sup>	C	T	Homozygous	100.0
H41-TN5	HBV	-	ND5	13605 <sup>1</sup>	C	T	Homozygous	100.0
H04-TN3	HCV	+	ND5	13708 <sup>1</sup>	G	A	Homozygous	100.0
H11-OT	HCV	+	ND5	13708 <sup>1</sup>	G	A	Homozygous	97.7
H11-TN1	HCV	+	ND5	13708 <sup>1</sup>	G	A	Homozygous	99.4
H11-TN2	HCV	+	ND5	13708 <sup>1</sup>	G	A	Homozygous	99.4
H11-TN3	HCV	+	ND5	13708 <sup>1</sup>	G	A	Homozygous	99.2
H41-OT	HBV	-	ND5	13708 <sup>1</sup>	G	A	Homozygous	100.0

## Appendix

Sample ID	Aetiology	Cirrhosis	Gene Cards	Region	Reference	Allele	Zygoty	Frequency
H41-TN1	HBV	-	ND5	13708 <sup>1</sup>	G	A	Homozygous	97.0
H41-TN2	HBV	-	ND5	13708 <sup>1</sup>	G	A	Homozygous	100.0
H41-TN3	HBV	-	ND5	13708 <sup>1</sup>	G	A	Homozygous	99.1
H41-TN5	HBV	-	ND5	13708 <sup>1</sup>	G	A	Homozygous	100.0
H41-TN1	HBV	-	ND5	13879	T	C	Homozygous	100.0
H41-TN2	HBV	-	ND5	13879	T	C	Homozygous	100.0
H02-T+OT	Toxic	-	ND6	14233 <sup>1</sup>	A	G	Homozygous	97.4
H02-TN1	Toxic	-	ND6	14233 <sup>1</sup>	A	G	Homozygous	98.1
H02-TN2	Toxic	-	ND6	14233 <sup>1</sup>	A	G	Homozygous	98.8
H02-TN3	Toxic	-	ND6	14233 <sup>1</sup>	A	G	Homozygous	99.2
H03-TN2	n.d	-	ND6	14233 <sup>1</sup>	A	G	Homozygous	100.0
H28-OT	SH	-	ND6	14233 <sup>1</sup>	A	G	Homozygous	99.0
H28-TN1	SH	-	ND6	14233 <sup>1</sup>	A	G	Homozygous	99.2
H28-TN2	SH	-	ND6	14233 <sup>1</sup>	A	G	Homozygous	98.1
H28-TN3	SH	-	ND6	14233 <sup>1</sup>	A	G	Homozygous	97.2
H11-OT	HCV	+	ND6	14364 <sup>1</sup>	G	A	Homozygous	97.6
H11-TN1	HCV	+	ND6	14364 <sup>1</sup>	G	A	Homozygous	97.3
H11-TN2	HCV	+	ND6	14364 <sup>1</sup>	G	A	Homozygous	97.6
H11-TN3	HCV	+	ND6	14364 <sup>1</sup>	G	A	Homozygous	98.1
H37-TN3	n.d	-	ND6	14364 <sup>1</sup>	G	A	Heterozygous	25.0
C10661	Control	-	CYB	14766 <sup>1</sup>	C	T	Heterozygous	83.6
H02-T+OT	Toxic	-	CYB	14766 <sup>1</sup>	C	T	Homozygous	98.0
H02-TN1	Toxic	-	CYB	14766 <sup>1</sup>	C	T	Homozygous	98.6
H02-TN2	Toxic	-	CYB	14766 <sup>1</sup>	C	T	Homozygous	98.7
H02-TN3	Toxic	-	CYB	14766 <sup>1</sup>	C	T	Homozygous	99.2
H03-OT	n.d	-	CYB	14766 <sup>1</sup>	C	T	Homozygous	100.0
H03-TN1	n.d	-	CYB	14766 <sup>1</sup>	C	T	Homozygous	100.0
H03-TN2	n.d	-	CYB	14766 <sup>1</sup>	C	T	Homozygous	98.4
H03-TN3	n.d	-	CYB	14766 <sup>1</sup>	C	T	Homozygous	100.0

Sample ID	Aetiology	Cirrhosis	Gene Cards	Region	Reference	Allele	Zygoty	Frequency
H04-TN3	HCV	+	CYB	14766 <sup>1</sup>	C	T	Homozygous	100.0
H28-OT	SH	-	CYB	14766 <sup>1</sup>	C	T	Homozygous	99.2
H28-TN1	SH	-	CYB	14766 <sup>1</sup>	C	T	Homozygous	99.1
H28-TN2	SH	-	CYB	14766 <sup>1</sup>	C	T	Homozygous	98.9
H28-TN3	SH	-	CYB	14766 <sup>1</sup>	C	T	Homozygous	98.8
H41-OT	HBV	-	CYB	14766 <sup>1</sup>	C	T	Homozygous	100.0
H41-TN1	HBV	-	CYB	14766 <sup>1</sup>	C	T	Heterozygous	93.2
H41-TN2	HBV	-	CYB	14766 <sup>1</sup>	C	T	Homozygous	100.0
H41-TN3	HBV	-	CYB	14766 <sup>1</sup>	C	T	Heterozygous	96.8
H41-TN5	HBV	-	CYB	14766 <sup>1</sup>	C	T	Homozygous	98.9
H02-T+OT	Toxic	-	CYB	15452 <sup>1</sup>	C	A	Homozygous	98.7
H02-TN1	Toxic	-	CYB	15452 <sup>1</sup>	C	A	Homozygous	98.0
H02-TN2	Toxic	-	CYB	15452 <sup>1</sup>	C	A	Homozygous	98.6
H02-TN3	Toxic	-	CYB	15452 <sup>1</sup>	C	A	Homozygous	97.5
H03-OT	n.d	-	CYB	15452 <sup>1</sup>	C	A	Heterozygous	96.2
H03-TN1	n.d	-	CYB	15452 <sup>1</sup>	C	A	Homozygous	100.0
H03-TN2	n.d	-	CYB	15452 <sup>1</sup>	C	A	Homozygous	100.0
H03-TN3	n.d	-	CYB	15452 <sup>1</sup>	C	A	Homozygous	100.0
H04-TN3	HCV	+	CYB	15452 <sup>1</sup>	C	A	Homozygous	98.4
H28-OT	SH	-	CYB	15452 <sup>1</sup>	C	A	Homozygous	98.4
H28-TN1	SH	-	CYB	15452 <sup>1</sup>	C	A	Homozygous	98.6
H28-TN2	SH	-	CYB	15452 <sup>1</sup>	C	A	Homozygous	98.4
H28-TN3	SH	-	CYB	15452 <sup>1</sup>	C	A	Homozygous	98.6
H32-TN1	HCV	+	CYB	15452 <sup>1</sup>	C	A	Heterozygous	37.5
H32-TN2	HCV	+	CYB	15452 <sup>1</sup>	C	A	Heterozygous	46.7
H41-OT	HBV	-	CYB	15452 <sup>1</sup>	C	A	Homozygous	100.0
H41-TN1	HBV	-	CYB	15452 <sup>1</sup>	C	A	Heterozygous	92.9
H41-TN2	HBV	-	CYB	15452 <sup>1</sup>	C	A	Homozygous	97.1
H41-TN3	HBV	-	CYB	15452 <sup>1</sup>	C	A	Homozygous	97.2
H41-TN5	HBV	-	CYB	15452 <sup>1</sup>	C	A	Heterozygous	96.9

Sample ID	Aetiology	Cirrhosis	Gene Cards	Region	Reference	Allele	Zygoty	Frequency
H02-T+OT	Toxic	-	CYB	15607 <sup>1</sup>	A	G	Homozygous	98.5
H02-TN1	Toxic	-	CYB	15607 <sup>1</sup>	A	G	Homozygous	98.8
H02-TN2	Toxic	-	CYB	15607 <sup>1</sup>	A	G	Homozygous	99.0
H02-TN3	Toxic	-	CYB	15607 <sup>1</sup>	A	G	Homozygous	99.5
H03-OT	n.d	-	CYB	15607 <sup>1</sup>	A	G	Homozygous	99.5
H03-TN1	n.d	-	CYB	15607 <sup>1</sup>	A	G	Homozygous	99.4
H03-TN2	n.d	-	CYB	15607 <sup>1</sup>	A	G	Homozygous	99.4
H03-TN3	n.d	-	CYB	15607 <sup>1</sup>	A	G	Homozygous	98.9
H28-OT	SH	-	CYB	15607 <sup>1</sup>	A	G	Homozygous	99.3
H28-TN1	SH	-	CYB	15607 <sup>1</sup>	A	G	Homozygous	99.3
H28-TN2	SH	-	CYB	15607 <sup>1</sup>	A	G	Homozygous	98.0
H28-TN3	SH	-	CYB	15607 <sup>1</sup>	A	G	Homozygous	99.1
H04-TN3	HCV	+	D-LOOP	16069	C	T	Homozygous	100.0
H41-TN1	HBV	-	D-LOOP	16069	C	T	Homozygous	100.0
H41-TN2	HBV	-	D-LOOP	16069	C	T	Homozygous	97.3
H41-TN3	HBV	-	D-LOOP	16069	C	T	Homozygous	100.0
H41-TN5	HBV	-	D-LOOP	16069	C	T	Homozygous	100.0
H02-T+OT	Toxic	-	D-LOOP	16126 <sup>1</sup>	T	C	Heterozygous	96.3
H02-TN1	Toxic	-	D-LOOP	16126 <sup>1</sup>	T	C	Homozygous	98.3
H02-TN2	Toxic	-	D-LOOP	16126 <sup>1</sup>	T	C	Homozygous	98.7
H02-TN3	Toxic	-	D-LOOP	16126 <sup>1</sup>	T	C	Homozygous	99.0
H03-OT	n.d	-	D-LOOP	16126 <sup>1</sup>	T	C	Homozygous	100.0
H03-TN1	n.d	-	D-LOOP	16126 <sup>1</sup>	T	C	Homozygous	100.0
H03-TN2	n.d	-	D-LOOP	16126 <sup>1</sup>	T	C	Homozygous	99.1
H03-TN3	n.d	-	D-LOOP	16126 <sup>1</sup>	T	C	Homozygous	97.1
H04-TN3	HCV	+	D-LOOP	16126 <sup>1</sup>	T	C	Homozygous	98.4
H28-OT	SH	-	D-LOOP	16126 <sup>1</sup>	T	C	Homozygous	97.8
H28-TN1	SH	-	D-LOOP	16126 <sup>1</sup>	T	C	Homozygous	98.8
H28-TN2	SH	-	D-LOOP	16126 <sup>1</sup>	T	C	Homozygous	97.7

## Appendix

Sample ID	Aetiology	Cirrhosis	Gene Cards	Region	Reference	Allele	Zygoty	Frequency
H28-TN3	SH	-	D-LOOP	16126 <sup>1</sup>	T	C	Homozygous	98.1
H41-TN1	HBV	-	D-LOOP	16126 <sup>1</sup>	T	C	Homozygous	100.0
H41-TN2	HBV	-	D-LOOP	16126 <sup>1</sup>	T	C	Homozygous	100.0
H41-TN3	HBV	-	D-LOOP	16126 <sup>1</sup>	T	C	Homozygous	100.0
H41-TN5	HBV	-	D-LOOP	16126 <sup>1</sup>	T	C	Homozygous	100.0
H04-TN3	HCV	+	D-LOOP	16145	G	A	Heterozygous	96.8
H41-TN1	HBV	-	D-LOOP	16145	G	A	Homozygous	100.0
H41-TN2	HBV	-	D-LOOP	16145	G	A	Homozygous	100.0
H41-TN3	HBV	-	D-LOOP	16145	G	A	Homozygous	98.8
H41-TN5	HBV	-	D-LOOP	16145	G	A	Homozygous	100.0
H32-OT	HCV	+	D-LOOP	16162 <sup>1</sup>	A	G	Homozygous	97.5
H32-TN1	HCV	+	D-LOOP	16162 <sup>1</sup>	A	G	Homozygous	99.0
H32-TN2	HCV	+	D-LOOP	16162 <sup>1</sup>	A	G	Homozygous	97.7
H32-TN3	HCV	+	D-LOOP	16162 <sup>1</sup>	A	G	Homozygous	97.7
H37-OT	n.d	-	D-LOOP	16162 <sup>1</sup>	A	G	Heterozygous	25.9
H38-OT	n.d	-	D-LOOP	16162 <sup>1</sup>	A	G	Heterozygous	47.6
H38-TN1	n.d	-	D-LOOP	16162 <sup>1</sup>	A	G	Homozygous	97.9
H38-TN2	n.d	-	D-LOOP	16162 <sup>1</sup>	A	G	Homozygous	97.6
H38-TN3	n.d	-	D-LOOP	16162 <sup>1</sup>	A	G	Homozygous	98.2
H38-TN4	n.d	-	D-LOOP	16162 <sup>1</sup>	A	G	Homozygous	98.7
H04-TN3	HCV	+	D-LOOP	16172	T	C	Heterozygous	95.2
H41-TN1	HBV	-	D-LOOP	16172	T	C	Homozygous	100.0
H41-TN2	HBV	-	D-LOOP	16172	T	C	Heterozygous	96.2
H41-TN3	HBV	-	D-LOOP	16172	T	C	Homozygous	97.6
H41-TN5	HBV	-	D-LOOP	16172	T	C	Homozygous	100.0
H04-TN3	HCV	+	D-LOOP	16261 <sup>1</sup>	C	T	Homozygous	99.4
H41-OT	HBV	-	D-LOOP	16261 <sup>1</sup>	C	T	Homozygous	98.5
H41-TN1	HBV	-	D-LOOP	16261 <sup>1</sup>	C	T	Homozygous	98.6

Sample ID	Aetiology	Cirrhosis	Gene Cards	Region	Reference	Allele	Zygoty	Frequency
H41-TN2	HBV	-	D-LOOP	16261 <sup>1</sup>	C	T	Homozygous	98.3
H41-TN3	HBV	-	D-LOOP	16261 <sup>1</sup>	C	T	Homozygous	98.6
H41-TN5	HBV	-	D-LOOP	16261 <sup>1</sup>	C	T	Homozygous	97.8
H40-TN1	Toxic	-	D-LOOP	16298	T	C	Homozygous	97.3
H40-TN2	Toxic	-	D-LOOP	16298	T	C	Heterozygous	96.8
H40-TN3	Toxic	-	D-LOOP	16298	T	C	Homozygous	97.2
H40-TN4	Toxic	-	D-LOOP	16298	T	C	Homozygous	97.8
H40-TN5	Toxic	-	D-LOOP	16298	T	C	Heterozygous	94.0
H40-TN6	Toxic	-	D-LOOP	16298	T	C	Heterozygous	86.9
H41-TN4	HBV	-	D-LOOP	16298	T	C	Homozygous	97.9
H02-T+OT	Toxic	-	D-LOOP	16304 <sup>1</sup>	T	C	Homozygous	98.3
H02-TN1	Toxic	-	D-LOOP	16304 <sup>1</sup>	T	C	Homozygous	98.9
H02-TN2	Toxic	-	D-LOOP	16304 <sup>1</sup>	T	C	Homozygous	98.6
H02-TN3	Toxic	-	D-LOOP	16304 <sup>1</sup>	T	C	Homozygous	99.2
H03-OT	n.d	-	D-LOOP	16304 <sup>1</sup>	T	C	Homozygous	99.5
H03-TN1	n.d	-	D-LOOP	16304 <sup>1</sup>	T	C	Homozygous	99.0
H03-TN2	n.d	-	D-LOOP	16304 <sup>1</sup>	T	C	Homozygous	98.0
H03-TN3	n.d	-	D-LOOP	16304 <sup>1</sup>	T	C	Homozygous	97.5
H28-OT	SH	-	D-LOOP	16304 <sup>1</sup>	T	C	Homozygous	98.9
H28-TN1	SH	-	D-LOOP	16304 <sup>1</sup>	T	C	Homozygous	99.1
H28-TN2	SH	-	D-LOOP	16304 <sup>1</sup>	T	C	Homozygous	98.4
H28-TN3	SH	-	D-LOOP	16304 <sup>1</sup>	T	C	Homozygous	98.8
C10661	Control	-	D-LOOP	16311 <sup>1</sup>	T	C	Heterozygous	52.6
H32-OT	HCV	+	D-LOOP	16311 <sup>1</sup>	T	C	Homozygous	97.3
H32-TN1	HCV	+	D-LOOP	16311 <sup>1</sup>	T	C	Homozygous	99.1
H32-TN2	HCV	+	D-LOOP	16311 <sup>1</sup>	T	C	Homozygous	98.7
H32-TN3	HCV	+	D-LOOP	16311 <sup>1</sup>	T	C	Homozygous	98.4
H37-OT	n.d	-	D-LOOP	16311 <sup>1</sup>	T	C	Heterozygous	22.7

## Appendix

Sample ID	Aetiology	Cirrhosis	Gene Cards	Region	Reference	Allele	Zygoty	Frequency
H38-OT	n.d	-	D-LOOP	16311 <sup>1</sup>	T	C	Heterozygous	22.6
C10661	Control	-	D-LOOP	16519 <sup>1</sup>	T	C	Heterozygous	63.4
C5764	Control	-	D-LOOP	16519 <sup>1</sup>	T	C	Homozygous	99.7
C633	Control	-	D-LOOP	16519 <sup>1</sup>	T	C	Homozygous	99.7
H02-T+OT	Toxic	-	D-LOOP	16519 <sup>1</sup>	T	C	Homozygous	99.8
H02-TN1	Toxic	-	D-LOOP	16519 <sup>1</sup>	T	C	Homozygous	99.8
H02-TN2	Toxic	-	D-LOOP	16519 <sup>1</sup>	T	C	Homozygous	99.9
H02-TN3	Toxic	-	D-LOOP	16519 <sup>1</sup>	T	C	Homozygous	99.9
H03-OT	n.d	-	D-LOOP	16519 <sup>1</sup>	T	C	Homozygous	100.0
H03-TN1	n.d	-	D-LOOP	16519 <sup>1</sup>	T	C	Homozygous	100.0
H03-TN2	n.d	-	D-LOOP	16519 <sup>1</sup>	T	C	Homozygous	98.5
H03-TN3	n.d	-	D-LOOP	16519 <sup>1</sup>	T	C	Homozygous	100.0
H04-OT	HCV	+	D-LOOP	16519 <sup>1</sup>	T	C	Homozygous	97.0
H04-TN2	HCV	+	D-LOOP	16519 <sup>1</sup>	T	C	Homozygous	100.0
H04-TN4	HCV	+	D-LOOP	16519 <sup>1</sup>	T	C	Homozygous	100.0
H04-TN6	HCV	+	D-LOOP	16519 <sup>1</sup>	T	C	Homozygous	100.0
H11-OT	HCV	+	D-LOOP	16519 <sup>1</sup>	T	C	Homozygous	99.0
H11-TN1	HCV	+	D-LOOP	16519 <sup>1</sup>	T	C	Homozygous	99.8
H11-TN2	HCV	+	D-LOOP	16519 <sup>1</sup>	T	C	Homozygous	99.3
H11-TN3	HCV	+	D-LOOP	16519 <sup>1</sup>	T	C	Homozygous	99.8
H28-OT	Toxic	-	D-LOOP	16519 <sup>1</sup>	T	C	Homozygous	100.0
H28-TN1	SH	-	D-LOOP	16519 <sup>1</sup>	T	C	Homozygous	99.8
H28-TN2	SH	-	D-LOOP	16519 <sup>1</sup>	T	C	Homozygous	99.2
H28-TN3	SH	-	D-LOOP	16519 <sup>1</sup>	T	C	Homozygous	99.6
H32-OT	HCV	+	D-LOOP	16519 <sup>1</sup>	T	C	Homozygous	100.0
H32-TN1	HCV	+	D-LOOP	16519 <sup>1</sup>	T	C	Homozygous	99.6
H32-TN2	HCV	+	D-LOOP	16519 <sup>1</sup>	T	C	Homozygous	100.0
H32-TN3	HCV	-	D-LOOP	16519 <sup>1</sup>	T	C	Heterozygous	92.9

Sample ID	Aetiology	Cirrhosis	Gene Cards	Region	Reference	Allele	Zygoty	Frequency
H37-OT	n.d	-	D-LOOP	16519 <sup>1</sup>	T	C	Homozygous	99.9
H37-TN1	n.d	-	D-LOOP	16519 <sup>1</sup>	T	C	Homozygous	100.0
H37-TN2	n.d	-	D-LOOP	16519 <sup>1</sup>	T	C	Homozygous	100.0
H37-TN3	n.d	-	D-LOOP	16519 <sup>1</sup>	T	C	Homozygous	99.9
H38-OT	n.d	-	D-LOOP	16519 <sup>1</sup>	T	C	Homozygous	99.8
H38-TN1	n.d	-	D-LOOP	16519 <sup>1</sup>	T	C	Homozygous	99.7
H38-TN2	n.d	-	D-LOOP	16519 <sup>1</sup>	T	C	Homozygous	100.0
H38-TN3	n.d	-	D-LOOP	16519 <sup>1</sup>	T	C	Homozygous	99.6
H38-TN4	n.d	-	D-LOOP	16519 <sup>1</sup>	T	C	Homozygous	100.0
H40-OT	Toxic	-	D-LOOP	16519 <sup>1</sup>	T	C	Homozygous	100.0
H11-TN1	HCV	+	ND5	12384 <sup>1</sup> 2385	-	C	Heterozygous	37.7
H11-TN2	HCV	+	ND5	12384 <sup>1</sup> 2385	-	C	Heterozygous	36.2
H11-TN3	HCV	+	ND5	12384 <sup>1</sup> 2385	-	C	Heterozygous	37.5
H41-TN1	HBV	-	RNR2	2777..2 778	GT	-	Heterozygous	62.5
H41-TN2	HBV	-	RNR2	2777..2 778	GT	-	Heterozygous	66.7
H41-TN3	HBV	-	RNR2	2777..2 778	GT	-	Heterozygous	69.0
H41-TN5	HBV	-	RNR2	2777..2 778	GT	-	Heterozygous	47.4
C10661	Control	-	D-LOOP	302 <sup>1</sup> 303	-	C	Heterozygous	17.5
C5764	Control	-	D-LOOP	302 <sup>1</sup> 303	-	CC	Heterozygous	7.8
C5764	Control	-	D-LOOP	302 <sup>1</sup> 303	-	C	Heterozygous	78.2
C633	Control	-	D-LOOP	302 <sup>1</sup> 303	-	CC	Heterozygous	5.4
C633	Control	-	D-LOOP	302 <sup>1</sup> 303	-	C	Heterozygous	79.2
H02-TN1	Toxic	-	D-LOOP	302 <sup>1</sup> 303	-	C	Homozygous	97.2
H02-TN2	Toxic	-	D-LOOP	302 <sup>1</sup> 303	-	C	Heterozygous	95.5
H02-TN3	Toxic	-	D-LOOP	302 <sup>1</sup> 303	-	C	Heterozygous	95.1
H37-OT	n.d	-	D-LOOP	302 <sup>1</sup> 303	-	C	Heterozygous	6.9
H37-OT	n.d	-	D-LOOP	302 <sup>1</sup> 303	-	CC	Heterozygous	47.4
H37-TN2	n.d	-	D-LOOP	302 <sup>1</sup> 303	-	C	Heterozygous	8.4

## Appendix

Sample ID	Aetiology	Cirrhosis	Gene Cards	Region	Reference	Allele	Zygoty	Frequency
H37-TN2	n.d	-	D-LOOP	302 <sup>1</sup> 303	-	CC	Heterozygous	90.4
H37-TN3	n.d	-	D-LOOP	302 <sup>1</sup> 303	-	CC	Heterozygous	90.3
H38-OT	n.d	-	D-LOOP	302 <sup>1</sup> 303	-	C	Heterozygous	5.5
H38-OT	n.d	-	D-LOOP	302 <sup>1</sup> 303	-	CC	Heterozygous	29.8
H37-OT	n.d	-	D-LOOP	513 <sup>1</sup> 514	-	CA	Heterozygous	67.5
H37-TN1	n.d	-	D-LOOP	513 <sup>1</sup> 514	-	CA	Homozygous	100.0
H37-TN2	n.d	-	D-LOOP	513 <sup>1</sup> 514	-	CA	Heterozygous	96.0
H37-TN3	n.d	-	D-LOOP	513 <sup>1</sup> 514	-	CA	Heterozygous	93.1
H38-OT	n.d	-	D-LOOP	513 <sup>1</sup> 514	-	CA	Heterozygous	44.7

X<sup>1</sup> : represents the variants which validated by Sanger sequencing

## Glossary

Gene names are written italic

<b>Abbreviation</b>	<b>Explanation</b>
°C	degrees celsius
µg	microgram
µl	microliter
µM	micromolar
AFP	alpha-fetoprotein
AI	Alcohol intake
<i>APC</i>	<i>Adenomatous Polyposis Coli</i>
<i>ARID1A</i>	<i>AT-Rich Interactive Domain-containing protein 1A</i>
<i>ARID2</i>	<i>AT-Rich Interactive Domain-containing protein 2</i>
<i>ASXL1</i>	<i>Additional Sex Combs Like Transcriptional Regulator 1</i>
<i>ATM</i>	<i>Ataxia Telangiectasia Mutated</i>
ATP	Adenosine triphosphate
<i>AXIN1</i>	<i>Axin 1</i>
BAD	BCL-2 associated death promoter
BCL-2	B-Cell CLL/Lymphoma 2
BCLC	Barcelona Clinic Liver Cancer
bp	basepairs
<i>BRAF</i>	<i>v-Raf murine sarcoma viral oncogene homolog B</i>
<i>CDH1</i>	<i>Cadherin-1</i>
<i>CDKN2A</i>	<i>Cyclin-Dependent Kinase Inhibitor 2A</i>
cDNA	complementary DNA
COSMIC	Catalogue Of Somatic Mutations in Cancer
COX	Cytochrome C
<i>CSF1R</i>	<i>Colony Stimulating Factor 1 Receptor</i>
<i>CTNNB1</i>	<i>Catenin (Cadherin-Associated Protein), Beta 1</i>
DAAs	Direct-Acting Antivirals
DAB	3,3'-diaminobenzidine tetrahydrochloride
dbSNP	Single Nucleotide Polymorphism database
DMEM	Dulbecco's modified Eagles medium
DMSO	dimethyl sulfoxide
DNA	DeoxyriboNucleic Acid
DNMT	DNA methyltransferase
dsDNA	double-stranded DNA
EDTA	Ethylene Diamine Tetraacetic Acid
EGFR	Epidermal Growth Factor Receptor
EMBOSS	European and Molecular Biology Open Software Suite
<i>ERBB2</i>	<i>Erb-B2 Receptor Tyrosine Kinase 2</i>
<i>EZH2</i>	<i>Enhancer of zeste homolog 2</i>
FCS	Fetal Calf Serum
FDA	Food and Drug Administration

Gene names are written italic

<b>Abbreviation</b>	<b>Explanation</b>
<i>FZD1</i>	<i>Frizzled Class Receptor 1</i>
g	gram
<i>GNAS</i>	<i>GNAS complex locus</i>
<i>GSK3b</i>	<i>Glycogen Synthase Kinase 3b</i>
h	hour
HA	Hepatic adenoma
HBV	hepatitis B
HBx	HBV X gene, viral transcriptional activator
HCC	HepatoCellular Carcinoma
HCV	hepatitis C
H-DN	High-grade dysplastic nodules
HEK cells	Human Embryonic Kidney cells
HFE	Hemochromatosis gene
Hg19	Human genome version 19
<i>HGF</i>	<i>Hepatocyte growth factor</i>
HM	Hemochromatosis
<i>HNF1A</i>	<i>Hepatocyte nuclear factor 1 homeobox A</i>
<i>HRAS</i>	<i>Harvey Rat Sarcoma Viral Oncogene Homolog</i>
indels	the insertion or the deletion
<i>IGF</i>	<i>insulin-like growth factor</i>
<i>IGF-I</i>	<i>insulin-like growth factor 1</i>
<i>IGF-II</i>	<i>Insulin-Like Growth Factor 2</i>
IGV	Integrative Genomic Viewer
IL-6	interleukin 6
<i>IL6ST</i>	<i>Interleukin 6 Signal Transducer</i>
Indels	Insertions/deletions
<i>IRF2</i>	<i>Interferon regulatory factor 2</i>
IWP	International Working Party
<i>JAK2</i>	<i>Janus Kinase 2</i>
<i>JNK</i>	<i>c-Jun N-terminal kinase</i>
<i>JUN</i>	<i>Proto-Oncogene</i>
Kb	Kilobase
<i>KDM6A</i>	<i>Lysine (K)-Specific Demethylase 6A</i>
<i>KRAS</i>	<i>Kirsten Rat Sarcoma Viral Oncogene Homolog</i>
LB	Lysogeny broth
L-DN	low-grade dysplastic nodules
LEF	lymphoid enhancer-factor
LT	liver transplantation
<i>MAPK 1</i>	<i>Mitogen-Activated Protein Kinase 1</i>
MCT	microwave coagulation therapy
<i>MET</i>	<i>Proto-Oncogene, Receptor Tyrosine Kinase</i>
mg	milligram
ml	milliliter
<i>MLL3</i>	<i>Mixed-lineage leukemia protein 3</i>

Gene names are written italic

<b>Abbreviation</b>	<b>Explanation</b>
<i>MMP9</i>	<i>Matrix Metallopeptidase 9</i>
mt	mitochondria
mtDNA	mitochondrial DNA
<i>MYC</i>	<i>V-Myc Avian Myelocytomatosis Viral Oncogene Homolog</i>
n.d	Underlying HCC are not defined
NAFLD	Non-alcoholic fatty liver disease
NASH	nonalcoholic steatohepatitis
NBT	Nitroblue tetrazolium
nDNA	nuclaeer DNA
<i>NF1</i>	<i>Neurofibromin 1</i>
<i>NF2</i>	<i>Neurofibromin 2 (Merlin)</i>
<i>NFE2L2</i>	<i>Nuclear Factor, Erythroid 2-Like 2</i>
ng	nanogram
NGS	next generation sequencing
nm	nanometer
<i>NOTCH1</i>	<i>Notch homolog 1</i>
<i>NOTCH2</i>	<i>Notch homolog 2</i>
<i>NOTCH3</i>	<i>Notch homolog 3</i>
<i>NRAS</i>	<i>Neuroblastoma RAS Viral (V-Ras) Oncogene Homolog</i>
NS5A	Nonstructural protein 5A
nt	nucleotides
OXPPOS	Oxidative phosphorylation
PBS	Phosphate-Buffered Saline
PCR	Polymerase Chain Reaction
<i>PDGF</i>	<i>the platelet-derived growth factors</i>
<i>PDGFRA</i>	<i>Platelet-Derived Growth Factor Receptor, Alpha Polypeptide</i>
PEI	Percutaneous Ethanol Injection
<i>PIK3CA</i>	<i>Phosphatidylinositide-3-kinase, catalytic, alpha polypeptide</i>
PMS	Phenazine methosulfate
<i>PTEN</i>	<i>Phosphatase and tensin homolog</i>
qPCR	quantitative real-time polymerase chain reaction
<i>RB</i>	<i>Retinoblastoma</i>
RFA	RadioFrequency Ablation
RNA	ribonucleic acid
ROS	Reactive Oxygen Species
<i>RPS6KA3</i>	<i>Ribosomal Protein S6 Kinase, 90kDa, Polypeptide 3</i>
rRNA	ribosomal RNA
SAV	Sequencing Analysis Viewer
SDH	Succinate dehydrogenase
SH	Steatohepatitis
SIFT	sorting tolerant from intolerant
<i>SMAD2</i>	<i>SMAD Family Member 2</i>
<i>SMAD4</i>	<i>SMAD Family Member 4</i>
TAC	Trans-Arterial Chemotherapy

---

Gene names are written italic

<b>Abbreviation</b>	<b>Explanation</b>
TACE	Trans-Arterial ChemoEmbolisation
TAE	Trans-Arterial Embolisation
TARE	Trans-Arterial RadioEmbolisation
TCF	T Cell-Factor
TE buffer	Tris EDTA buffer
TGF- $\beta$	Transforming Growth Factor beta
TKs	Tyrosine Kinases
<i>TP53</i>	<i>Tumor protein p53</i>
tRNA	transfer RNA
<i>VEGF</i>	<i>Vascular Endothelial Growth Factor</i>
<i>VEGFR-2</i>	<i>Vascular Endothelial Growth Factor Receptor-2</i>
<i>VEGFR-3</i>	<i>Vascular Endothelial Growth Factor Receptor-3</i>

## Curriculum Vitae

### Personal Information

Name	<b>Wafa Amer</b>
Date of birth	14 February 1985
Family	<u>Husband</u> : Dr. Mohamed Abudabbous, Senior consultant and co-director for obstetrics and gynaecology at Korbach Hospital, Korbach, Germany  <u>Son</u> : Khaled (4 Ys), <u>Daughter</u> : Hawa (3 Ms)
Address (Home)	Lessingstr. 6, 37671 Hörter
Address (official)	Pathology Institute, University of Cologne, Kerpener Str. 62, 50937 Köln
Phone (mobile)	+218 (0)91 3215439
Phone (official)	+49 (0)221 478-6351
E-mail	<a href="mailto:wata.amer@uk-koeln.de">wata.amer@uk-koeln.de</a> <a href="mailto:w.amer85@yahoo.com">w.amer85@yahoo.com</a>
Nationality (original)	Libyan
Other Nationality	German

### School Education and Experience

<u>From April 2012</u>	PhD- Thesis at the Institute of Pathology, University Hospital Cologne, Cologne, Germany: <i>“Comprehensive analysis of disease related nuclear and mitochondrial genes in hepatocellular carcinoma”</i>
<u>January 2011–March 2012</u>	Master-Thesis at the Institute for Pathology at the University Hospital Cologne, Cologne, Germany: <i>“Studies in molecular pathology”</i>
<u>April 2010–December 2010</u>	Came to Germany to learn state-of-the-art techniques, and started German language courses, Abendakademie, Mannheim, Germany
<u>February 2009–February 2010</u>	Clerkship training in Misurata Central Hospital, Misurata, Libya

<u>2008–2009</u>	1 <sup>st</sup> session Exam in 2008, approved in January 2009, with general average grade very good (82.00%). MBCh degree awarded, Misurata Central Hospital, Misurata, Libya
<u>2002–2008</u>	Studies in Medicine at Misurata University, Faculty of medicine, Misurata, Libya
<u>1998–2001</u>	Secondary school, Misurata, Libya
<u>1991–1997</u>	Primary and preparatory school, Misurata, Libya

### Conference Contributions

<u>May 2016</u>	2nd International Conference on Hepatology, Chicago, USA ( <b>Oral Presentation, Co-Chair</b> )
<u>January 2016</u>	Falk Symposium (Communication and System Relevance in Liver Damage and Regeneration), Düsseldorf, Germany ( <b>Poster Presentation</b> )
<u>April 2015</u>	50 <sup>th</sup> The International Liver Congress, EASL 2015, Vienna, Austria ( <b>Two Poster Presentations</b> )
<u>January 2015</u>	31 <sup>st</sup> Annual Meeting, German Association of the Study of the Liver, GASL 2015, Munich, Germany ( <b>Oral Presentation</b> )
<u>February 2014</u>	1 <sup>st</sup> IPMM-PhD Days, Cologne, Germany ( <b>Poster Presentation</b> )
<u>January 2013</u>	29 <sup>th</sup> Annual Meeting, German Association of the Study of the Liver, GASL 2013, Hannover, Germany ( <b>Poster Presentation</b> )
<u>January 2013</u>	Falk Symposium (Targeted Therapies in Hepatology), Hannover, Germany ( <b>Poster Presentation</b> ).

### Awards

<u>Falk Symposium Travel Grant</u>	(Falk Workshop 2013, Hannover, Germany)
<u>Falk Symposium Travel Grant</u>	(Falk Workshop 2016, Düsseldorf, Germany)
<u>Registration Bursary</u>	(50 <sup>th</sup> The International Liver Congress, EASL 2015, Vienna, Austria)
<u>EASL Membership</u>	(50 <sup>th</sup> The International Liver Congress, EASL 2015, Vienna, Austria)

## Acknowledgements

First and foremost, I would like to express my sincere gratitude to my advisor **Professor Dr. Margarete Odenthal**, who provided me an opportunity to join her team and work in her lab. To work under your supervision has been a real pleasure to me. You have been a steady influence throughout my PhD career; you have oriented and supported me with promptness and care. You have always been patient and encouraging in times of new ideas and difficulties; you have listened to my ideas and discussions with you frequently led to key insights. Your ability to select and to approach compelling research problems, your high scientific standards, and your hard work set an example. I am very grateful for your motivation, enthusiasm, and immense knowledge. Your guidance helped me in all the time of research and writing of this thesis. You have been a tremendous mentor for me. I could not have imagined having a better advisor and mentor for my PhD study. Above all, you have always been available to help and advice for life difficulties. You made me feel of non-alienation, which I appreciate from my heart. For this, I cannot thank you enough. I am forever grateful.

I want to thank **Professor Dr. R. Buettner**, Director, for the academic support and the facilities provided to carry out the research work at the Institute.

I would also like to thank my IPMM tutors, **Professor Dr. Tobias Goeser**, and **Professor Dr. Thomas Wunderlich**. Their insightful comments and suggestions are greatly appreciated. Thank you.

I am particularly thankful to **Dr. Debora Grosskopf-Kroiher**, Center for Molecular Medicine University of Cologne, for your kind encouragement and support that extended to all IPMM-PhD students. You are a valuable source of motivation, and you set a good example of scientific mother, who building successful career while also building family.

My sincere thanks also goes to my colleagues especially; **Hannah Eischeid**, and **Ulrike Koitzsch**, your professional work in scheduling and relaxed demeanor made for a good working relationship. Your technical excellence and tremendous grasp of experimental issues had a great impact on me. I admire your ability to make a hard concentrated work, heaps of fun and excitement. Also, I would like to thank **Sojung, Iris, Priya**, and **Xiaojie** for sharing their knowledge, and providing

a great work environment. My time in the lab was made enjoyable with my colleagues in last five years, all of you have been there to help and support me. Thank you.

My gratitude is also extended to **Dr. Marion Kopp** who has known the answer to every question I've ever asked her regarding my work and life. You were one of the first friendly faces to greet me when I began this doctoral program and has always been a tremendous help no matter the task or circumstance.

Many of the results on next generation sequencing of mitochondrial genome would not have been possible without **Dr. Anne Arens**. Thank you for teaching me how to interpret the NGS-data by Biochemical Genomics Workbench.

I owe a lot to my parents, specially my father Emhemed Amer, who is not with me to share this joy. You encouraged and helped me at every stage of my personal and academic life, and longed to see this achievement come true. I am deeply missing you Dad. Words cannot express how grateful I am to my mother for all of the sacrifices that you've made on my behalf. Your prayer for me was what sustained me thus far. My Dad and my Mom, I am proud of you and love you very much.

I am very much indebted to my family, my husband Dr. Mohamed Abudabbous, and my kids (Khaled & Hawa), you supported me in every possible way to see the completion of this work. You are always cheering me up, I love you very much

## Erklärung

Ich versichere, dass ich die mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit - einschließlich Tabellen, Karten, und Abbildungen - , die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie – abgesehen von unten angegebenen Teilpublikationen – noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde\*.

Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Margarete Odenthal (Institut für Pathologie) betreut worden.

Übersicht der Publikationen:

1. Analysis of microsatellite instability by microfluidic-based electrophoresis.  
Natalia Elfimova, **Wafa Amer**, and Margarete Odenthal. Methods Mol Biol. 2013;919:287-96. DOI: 10.1007/978-1-62703-029-8\_25.
2. Promoter regulation of the miR-29a/b1 gene by transforming growth factor- $\beta$  in hepatic stellate cells and in hepatoma cells  
Jia Huang, **Wafa Amer**, Hans Peter Dienes, Reinhard Büttner, and Margarete Odenthal. Z Gastroenterol 2015, 53-A1\_41. DOI: 10.1055/s-0034-1397082

Ich versichere, dass ich alle Angaben wahrheitsgemäß nach bestem Wissen und Gewissen gemacht habe und verpflichte mich, jedmögliche, die obigen Angaben betreffenden Veränderungen, dem Promotionsausschuss unverzüglich mitzuteilen.

\*Hiermit möchte ich darauf hinweisen, dass die nachfolgende Arbeit als Publikation beim Journal of Pathology Clinical Research im März eingereicht wurde.

.....

Datum

.....

Unterschrift

---

## Publications

1. *Analysis of microsatellite instability by microfluidic-based electrophoresis.*

Natalia Elfimova, **Wafa Amer**, and Margarete Odenthal. *Methods Mol Biol.* 2013;919:287-96. DOI: 10.1007/978-1-62703-029-8\_25.

2. *Promoter regulation of the miR-29a/b1 gene by transforming growth factor- $\beta$  in hepatic stellate cells and in hepatoma cells*

Jia Huang, **Wafa Amer**, Hans Peter Dienes, Reinhard Büttner, and Margarete Odenthal. *Z Gastroenterol* 2015, 53-A1\_41. DOI: 10.1055/s-0034-1397082

## In Review

3. *Tumor tracking and evolution analysis of heterogeneous non-small cell lung carcinoma by ultra-deep sequencing of the entire mitochondrial genome*

**Wafa Amer**, Csaba Toth, Erik Vassella, Ulrike Koitzsch, Hannah Eischeid, Jeannine Meinrath, Anne Arens, Jia Huang, Alexander Adam, Andreas Scheel, Reinhard Buettner, Stephan C. Schaefer, and Margarete Odenthal.

## In Preparation

4. *mtDNA mutations may promote the tumor machinery in chronic liver disease*

**Wafa Amer**, Alexander Quass, Anne Arens, Ulrike Koitzsch, Hannah Eischeid, Reinhard Buettner, and Margarete Odenthal.